



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS JURÍDICAS CCJ  
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO

Bruno Cassol da Silva

**Inteligência Artificial Explicável e Decisões Judiciais:** Experimentações com o método SHAP

Florianópolis/SC  
2024

Bruno Cassol da Silva

**Inteligência Artificial Explicável e Decisões Judiciais: Experimentações com o método SHAP**

Dissertação submetida ao Programa de Pós-Graduação em Direito da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre em Direito Estado e Sociedade.

Orientador: Prof. Dr. Aires J. Rover

Florianópolis/SC  
2024

**Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.  
Dados inseridos pelo próprio autor.**

SILVA, BRUNO CASSOL  
Inteligência Artificial Explicável e Decisões Judiciais  
: Experimentações com o método SHAP / BRUNO CASSOL SILVA ;  
orientador, AIRES JOSÉ ROVER, 2024.  
201 p.

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Centro de Ciências Jurídicas, Programa de Pós  
Graduação em Direito, Florianópolis, 2024.

Inclui referências.

Aprendizado de Máquina. 4. Previsão de Decisões  
Judiciais.  
5. Direito. I. ROVER, AIRES JOSÉ. II. Universidade Federal  
de Santa Catarina. Programa de Pós-Graduação em Direito.  
III. Título.

Bruno Cassol da Silva

Inteligência Artificial Explicável e Decisões Judiciais: Experimentações com o método SHAP

O presente trabalho em nível de mestrado foi avaliado e aprovado, em 14/03/2024 às 10:00 da manhã através de sessão virtual via google meet, contando com a presença dos membros:

Prof. Dr. Aires José Rover  
UFSC

Prof. Dr. Fabiano Hartmann Peixoto  
UniMontreal

Prof. Dr. Sérgio Pádua  
UniBrasil

Certificamos que esta é a versão original e final do trabalho.

---

Coordenação do Programa de Pós-Graduação em Direito em UFSC

---

Prof. Dr. Aires J. Rover

Florianópolis/SC, 2024

## RESUMO

Segundo o relatório "Justiça em Números" do Conselho Nacional de Justiça (CNJ), houve um acréscimo de 10% no número de novas ações judiciais em 2022, chegando a um total impressionante de 31,5 milhões. Diante desse cenário, este trabalho tem como objetivo principal examinar como os conhecimentos técnico-simbólicos podem ser empregados para proporcionar explicabilidade à Inteligência Artificial (IA) no contexto do Direito, especificamente através do método SHAP. Para isso, realizou-se uma revisão bibliográfica sobre o potencial da IA e da Inteligência Artificial Explicável (XAI), uma revisão sistemática de IA e XAI em fontes acadêmicas (Scopus, Web of Science, SciELO, IEEE Xplore, ACM Digital Library e Google Scholar), experimentos para entender a interação de modelos preditivos com o conhecimento jurídico estruturado, experimentos com o modelo explicativo SHAP para prever decisões de IA em relação a decisões judiciais envolvendo ações de companhias aéreas, e uma validação da abordagem proposta com especialistas da área e operadores do direito. Os resultados indicaram que o método SHAP é capaz de gerar rotas explicativas de decisões algorítmicas na representação de conhecimento jurídico, contribuindo para a transparência e a compreensão da IA jurídica.

**Palavras-chave:** Inteligência Artificial Explicável; Direito; Método SHAP; Aprendizado de Máquina; Previsão de Decisões Judiciais.

## ABSTRACT

According to the “Justice in Numbers” report by the National Council of Justice (CNJ), there was a 10% increase in the number of new lawsuits in 2022, reaching an impressive total of 31.5 million. In light of this scenario, this study aims to examine how technical-symbolic knowledge can be employed to provide explainability to Artificial Intelligence (AI) in the context of Law, specifically through the SHAP method. To this end, a bibliographic review was conducted on the potential of AI and Explainable Artificial Intelligence (XAI), a systematic review of AI and XAI in academic sources (Scopus, Web of Science, SciELO, IEEE Xplore, ACM Digital Library, and Google Scholar), experiments to understand the interaction of predictive models with structured legal knowledge, experiments with the SHAP explanatory model to predict AI decisions regarding judicial decisions involving airline companies’ lawsuits, and a validation of the proposed approach with experts in the field and legal practitioners. The results indicated that the SHAP method is capable of generating explanatory pathways for algorithmic decisions in the representation of legal knowledge, contributing to the transparency and understanding of legal AI.

**Keywords:** Explainable Artificial Intelligence; Law; SHAP Method; Machine Learning; Prediction of Judicial Decisions.

## LISTA DE FIGURAS

Figura 1 Ilustração do Explain Model.....	7
Figura 2 Ilustração do Explain Prediction.....	7
Figura 3 Ilustração do Fluxo de Trabalho do Aprendizado de Máquina .....	24
Figura 4 Ilustração da preparação de dados para Aprendizado de Máquina.....	25
Figura 5 Ilustração do treinamento de IA explicável (XAI).....	31
Figura 6 Ilustração do Orange Canvas .....	32
Figura 7 Imagem de falha na segurança da informação em Ashley Madison .....	36
Figura 8 Processo de Revisão Sistemática de Literatura.....	49
Figura 9 - Número de publicações por ano entre 2014 e 2024.....	56
Figura 10 - Número de documentos por autor .....	57
Figura 11- Número de documentos por país.....	57
Figura 12 - Nuvem de Palavras mais citadas nos artigos científicos .....	58
Figura 13 Ilustração do treinamento de IA explicável (XAI).....	85
Figura 14 Ilustração do Orange Canvas .....	88
Figura 15 Ilustração do treinamento de IA explicável (XAI) em Orange Canvas .....	91
Figura 16 Ilustração do treinamento de IA explicável (XAI).....	93
Figura 17 Ilustração do AdaBoost .....	94
Figura 18 Ilustração do AdaBoost – Base de Aprendizado .....	95
Figura 19 Ilustração do AdaBoost – Base de Testes.....	96
Figura 20 Ilustração do AdaBoost – Fluxo de trabalho.....	96
Figura 21 Ilustração do AdaBoost – Test and Score.....	97
Figura 22 Ilustração do AdaBoost – Explain Model.....	98
Figura 23 Ilustração do AdaBoost – Explain Model.....	98
Figura 24 Ilustração do AdaBoost – Explain Model.....	99
Figura 25 Ilustração do AdaBoost – Base de Testes.....	100
Figura 26 Ilustração do AdaBoost – Explain Model.....	100
Figura 27 Ilustração do AdaBoost – Explain Model.....	101
Figura 28 Ilustração do AdaBoost – Explain Model.....	102
Figura 29 Ilustração do AdaBoost – Explain Model.....	102
Figura 30 Ilustração do AdaBoost – Base de Testes.....	103
Figura 31 Ilustração do AdaBoost – Tabela de dados .....	104
Figura 32 Ilustração do AdaBoost – Explain Prediction .....	104
Figura 33 Ilustração do AdaBoost – Planilha de dados CSV .....	105
Figura 34 Ilustração do AdaBoost – Planilha de dados CSV .....	105
Figura 35 Ilustração do AdaBoost – Tabela de dados/Explain Prediction .....	106
Figura 36 Ilustração do AdaBoost – Planilha de dados CSV .....	107
Figura 37 Ilustração do AdaBoost – Planilha de dados CSV .....	107
Figura 38 Ilustração do AdaBoost – Tabela de dados/Explain Prediction .....	108
Figura 39 Ilustração do AdaBoost – Planilha de dados CSV .....	109
Figura 40 Ilustração do AdaBoost – Explain Prediction .....	110
Figura 41 Ilustração do AdaBoost – Planilha de dados CSV .....	111
Figura 42 Ilustração do Gradient Boosting .....	112
Figura 43 Ilustração do Gradient Boosting – Fluxo de Trabalho.....	113
Figura 44 Ilustração do Gradient Boosting – Base de Aprendizado .....	113
Figura 45 Ilustração do Gradient Boosting – Base de Testes.....	114

<b>Figura 46 Ilustração do Gradient Boosting – Fluxo de Trabalho.....</b>	<b>115</b>
<b>Figura 47 Ilustração do Gradient Boosting – Test and Score .....</b>	<b>115</b>
<b>Figura 48 Gradient Boosting –Explain Model.....</b>	<b>116</b>
<b>Figura 49 Ilustração do Gradient Boosting – Explain Model .....</b>	<b>116</b>
<b>Figura 50 Ilustração do Gradient Boosting – Explain Prediction .....</b>	<b>117</b>
<b>Figura 51 Ilustração de Redes Neurais – Base de Aprendizado .....</b>	<b>118</b>
<b>Figura 52 Ilustração de Redes Neurais – Base de Testes .....</b>	<b>119</b>
<b>Figura 53 Ilustração de Redes Neurais – Seletor do Atributo.....</b>	<b>120</b>
<b>Figura 54 Ilustração de Redes Neurais – Fluxo de Trabalho.....</b>	<b>120</b>
<b>Figura 55 Ilustração de Redes Neurais – Algoritmo .....</b>	<b>121</b>
<b>Figura 56 Ilustração de Redes Neurais – Test and Score .....</b>	<b>122</b>
<b>Figura 57 Ilustração de Redes Neurais – Explain Model .....</b>	<b>122</b>
<b>Figura 58 Ilustração de Redes Neurais – Explain Model .....</b>	<b>123</b>
<b>Figura 59 Ilustração de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>124</b>
<b>Figura 60 Ilustração de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>124</b>
<b>Figura 61 Ilustração de Planilha de Dados em CSV .....</b>	<b>125</b>
<b>Figura 62 Ilustração de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>126</b>
<b>Figura 63 Ilustração de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>126</b>
<b>Figura 64 Ilustração de Planilha de Dados em CSV .....</b>	<b>127</b>
<b>Figura 65 Ilustração de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>127</b>
<b>Figura 66 Ilustração de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>128</b>
<b>Figura 67 Ilustração de Redes Neurais – Fluxo de Trabalho.....</b>	<b>128</b>
<b>Figura 68 Ilustração de Redes Neurais – Base de Testes .....</b>	<b>129</b>
<b>Figura 69 Ilustrações de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>130</b>
<b>Figura 70 Ilustração de Planilha de Dados em CSV .....</b>	<b>131</b>
<b>Figura 71 Ilustrações de Redes Neurais – Tabela de Dados e Explain Prediction .....</b>	<b>132</b>
<b>Figura 72 Ilustrações de Planilha em CSV .....</b>	<b>133</b>
<b>Figura 73 Ilustração de Regressão Logística - Base de Aprendizado .....</b>	<b>134</b>
<b>Figura 74 Ilustração de Regressão Logística - Base de Aprendizado.....</b>	<b>135</b>
<b>Figura 75 Ilustração de Regressão Logística – Seletor do Atributo .....</b>	<b>136</b>
<b>Figura 76 Ilustração de Regressão Logística – Fluxo de Trabalho .....</b>	<b>137</b>
<b>Figura 77 Ilustração de Regressão Logística – Algoritmo.....</b>	<b>137</b>
<b>Figura 78 Ilustração de Regressão Logística – Test and Score .....</b>	<b>138</b>
<b>Figura 79 Ilustração de Regressão Logística – Explain Model.....</b>	<b>139</b>
<b>Figura 80 Ilustração de Regressão Logística – Explain Model.....</b>	<b>140</b>
<b>Figura 81 Ilustração de Regressão Logística - Tabela de Dados e Explain Prediction.....</b>	<b>141</b>
<b>Figura 82 Ilustração de Regressão Logística – Tabela de Dados e Explain Prediction ....</b>	<b>141</b>
<b>Figura 83 Ilustração de Planilha em CSV .....</b>	<b>142</b>
<b>Figura 84 Ilustração de Regressão Logística – Tabela de Dados e Explain Prediction ...</b>	<b>142</b>
<b>Figura 85 Ilustração de Planilha em CSV .....</b>	<b>143</b>
<b>Figura 86 Ilustração de Regressão Logística – Tabela de Dados e Explain Prediction ....</b>	<b>143</b>
<b>Figura 87 Ilustração de Planilha em CSV.....</b>	<b>144</b>
<b>Figura 88 Ilustração de Regressão Logística – Fluxo de Trabalho.....</b>	<b>145</b>
<b>Figura 89 Ilustração de Regressão Logística – Base de Testes.....</b>	<b>145</b>
<b>Figura 90 Ilustrações de Regressão Logística – Tabela de Dados e Explain Prediction...</b>	<b>146</b>
<b>Figura 91 Ilustração de Planilha em CSV .....</b>	<b>147</b>
<b>Figura 92 Ilustração de Regressão Logística – Tabela de Dados e Explain Prediction ...</b>	<b>148</b>
<b>Figura 93 Ilustração de Planilha em CSV .....</b>	<b>149</b>



<b>Figura 94 Ilustração do CoCounsel da Casetext. ....</b>	<b>154</b>
<b>Figura 95 Ilustração da LIVIA. ....</b>	<b>155</b>
<b>Figura 96 Ilustração de Questionário de Validação com servidora do JEC/UFSC. ....</b>	<b>160</b>
<b>Figura 97 - Ilustração de Questionário de Validação com especialista em IA.....</b>	<b>162</b>

## LISTA DE TABELAS

<b>Tabela 1 - Procedimentos e Técnicas.....</b>	<b>4</b>
<b>Tabela 2 - Artigos Identificados Inicialmente.....</b>	<b>51</b>
<b>Tabela 3 - Artigos Selecionados .....</b>	<b>53</b>

## LISTA DE ABREVIACOES E ACRONIMOS

ANPD	Autoridade Nacional de Proteo de Dados
CF/88	Constituio Federal de 1988
CNJ	Conselho Nacional de Justia
COMPAS	<i>Correctional Offender Management Profiling For Alternative Sanctions</i>
EPSRC	<i>Engineering &amp; Physical Sciences Research Council</i>
GPDR	<i>General Data Protection Regulation</i>
IA	Inteligncia Artificial
JEC	Juizado Especial Cvel
JEF	Juizado Especial Federal
LGPD	Lei Geral de Proteo de Dados
LLMs	<i>Large Language Models</i> ou Modelo de linguagem grande
MDPs	Processos de Deciso de Markov
MRF	Campos Aleatrios de Markov
ODRs	<i>Online Dispute Resolutions</i> ou Resoluo de Disputas <i>Online</i> .
ROC	<i>Receiver Operating Characteristic</i>
SHAP	<i>Shapley Additive Explanations</i>
SLR	Reviso Sistemtica de Literatura
STM	Superior Tribunal Militar
STF	Supremo Tribunal Federal
STJ	Superior Tribunal de Justia
SVM	Support Vector Machine
TJs	Tribunais de Justia
TREs	Tribunais Regionais Eleitorais
TRFs	Tribunais Regionais Federais
TRTs	Tribunais Regionais do Trabalho
TSE	Tribunal Superior Eleitoral
TST	Tribunal Superior do Trabalho
XAI	<i>Explainable Artificial Intelligence</i> ou IA Explicvel

## SUMÁRIO

1-	INTRODUÇÃO.....	2
1.1	Contexto e motivação de pesquisa.....	2
1.2	Problema de pesquisa, hipótese e objetivos .....	2
1.3	Metodologia.....	4
1.4	Originalidade .....	8
1.5	Estrutura do trabalho .....	8
2-	REVISÃO BIBLIOGRÁFICA .....	9
2.1	Definição, evolução história da IA e dilemas éticos .....	9
2.2	Aprendizado de Máquina ( <i>Machine Learning</i> ) .....	19
2.2.1	<i>Inteligência Artificial Explicável (XAI)</i> .....	27
2.2.2	<i>Proteção De Dados Pessoais</i> .....	34
2.3	Revisão Sistemática de Literatura .....	48
3-	ESTADO DA ARTE DE IA E XAI NO DIREITO .....	59
4-	EXPERIMENTOS.....	85
4.1	Juizado Especial Cível.....	85
4.2	Base de dados e estruturação .....	87
4.3	Etapas do aprendizado de máquina e experimento com método <i>SHAP</i> .....	90
4.4	Análise e classificação de algoritmos de aprendizado de máquina em sua acurácia associado ao método explicativo <i>SHAP</i> .....	92
4.4.1	<i>Adaptive Boosting</i> .....	93
4.4.2	<i>Gradientboost</i> .....	112
4.4.3	Redes Neurais ( <i>Neural Network</i> ).....	117

4.4.4	Regressão Logística ( <i>Logistic Regression</i> ).....	134
5-	PROPOSTA E VALIDAÇÃO .....	150
5.1	Proposta de aplicação de método explicativo <i>SHAP</i> para modelos preditivos de IA no judiciário brasileiro.....	150
5.2	Validação, resultados e discussão.....	157
5.2.1	Validação com especialistas ao longo do desenvolvimento dos experimentos de XAI.....	158
5.2.2	Questionários de validação com expert em IA e JEC/UFSC .....	158
5.2.2.1	Resultados da validação com a servidora do JEC/UFSC .....	159
5.2.2.2	Resultados da validação com a expert em IA.....	161
6-	CONCLUSÃO.....	164
6.1	Considerações finais e contribuições.....	164
6.2	Limitações .....	166
6.3	Trabalhos futuros.....	168
	REFERÊNCIAS .....	171
	APÊNDICE A – QUESTIONÁRIO DE PESQUISA COM EXPERT E JEC/UFSC .....	180
	ANEXO A – (JEC/UFSC) AUTORIZAÇÃO DE PESQUISA RELACIONADA A DADOS ANONIMIZADOS DE SENTENÇA DO JEC/UFSC .....	188

# 1- INTRODUÇÃO

## 1.1 Contexto e motivação de pesquisa

Este experimento investiga a aplicação da Inteligência Artificial Explicável (*XAI*) no campo do Direito, utilizando o método *SHAP* para gerar explicações sobre as previsões de IA em decisões judiciais. A motivação de pesquisa surge da necessidade de compreender e explicar como a IA funciona e como ela afeta os direitos e deveres das pessoas, especialmente no contexto jurídico, que envolve questões de legalidade, legitimidade e justiça. A *XAI* busca fornecer transparência, confiança e responsabilidade aos sistemas de IA, contribuindo para o aprimoramento do sistema jurídico brasileiro.

O experimento utiliza dados extraídos de decisões judiciais envolvendo ações de companhias aéreas, que representam uma parcela significativa dos processos no Brasil. Segundo o Relatório Justiça em Números 2020, o Brasil possui mais de 77 milhões de processos em tramitação, sendo que a matéria de direito do consumidor representa 10,4% do total. O método *SHAP* é uma técnica que atribui valores de importância às características de entrada de um modelo de IA, indicando como elas influenciam na previsão do modelo. O objetivo é representar o conhecimento jurídico de forma a possibilitar a geração de rotas explicativas de decisões algorítmicas.

## 1.2 Problema de pesquisa, hipótese e objetivos

Como representar conhecimento jurídico de modo a contribuir para a justificar adequadamente as previsões de IA em decisões judiciais? Atualmente, um dos grandes debates em torno do Direito e da IA é o das “black boxes”, isto é, inteligências artificiais que apesar da sua eficácia e acurácia em suas previsões não explicam de forma coerente o resultado e, tampouco, os procedimentos que levaram ao resultado.

Desta feita, para que o uso de IA no Judiciário seja feito de uma forma coerente com as demandas de transparência, investiga-se como o método SHAP pode ser utilizado para tornar a IA jurídica mais transparente e compreensível. O método SHAP (SHapley Additive exPlanations) é uma técnica de *XAI* que atribui valores de importância a cada característica de entrada de um modelo de aprendizado de máquina, baseado na teoria dos jogos cooperativos. Esses valores representam a contribuição de cada característica para

a predição do modelo, permitindo assim identificar os fatores que mais influenciam a decisão da IA. A delimitação do tema ocorre no âmbito da representação de conhecimento jurídico por meio do método SHAP, visando a construção de modelos preditivos e explicáveis de IA.

A hipótese subjacente é que seja possível representar o conhecimento jurídico de maneira a contribuir para a justificação adequada das predições de IA em decisões judiciais, por meio da aplicação do método SHAP em conjunto com dados extraídos de decisões judiciais. Para testar essa hipótese, definimos os seguintes objetivos:

- **Objetivo geral:** Examinar como os conhecimentos técnico-simbólicos podem ser empregados para proporcionar explicabilidade à IA no contexto do Direito, especificamente através do método SHAP.
- **Objetivo específico 1:** Fazer uma revisão bibliográfica que destaque o potencial da IA e da XAI no Direito, bem como os desafios e as limitações existentes. Isso vai permitir situar o problema de pesquisa no contexto teórico e prático da área, além de identificar as lacunas do conhecimento que a pesquisa pretende preencher.
- **Objetivo específico 2:** Levantar dados relativos a IA e XAI em fontes acadêmicas, como Web of Science, Scopus e Scielo, para mapear a abordagem e as aplicações práticas da XAI no Direito, assim como também locais onde seria possível obter uma base de dados confiável para realização de experimentos. Isso vai permitir selecionar os dados mais relevantes e adequados para a análise e a experimentação.
- **Objetivo específico 3:** Aplicar o método SHAP em experimentos com IA na predição de resultado de sentenças judiciais de Tribunais e ou Juizados Especiais. Isso vai permitir verificar a eficácia e a explicabilidade do método SHAP na representação do conhecimento jurídico e na justificação das predições de IA.
- **Objetivo específico 4:** A partir do diagnóstico dos experimentos realizados com XAI, propor o seu uso e aplicação em grau de testes nos Tribunais e Juizados Especiais. Isso vai permitir avaliar a viabilidade e o impacto da proposta na melhoria da transparência e da confiabilidade da IA no Direito.
- **Objetivo específico 5:** A validação da abordagem proposta por especialistas da área e operadores do direito, com o intuito de determinar sua eficácia e aceitação.

Isso vai permitir obter feedbacks e sugestões para aprimorar a proposta e garantir sua qualidade e relevância

### 1.3 Metodologia

Nossa abordagem metodológica é guiada por três componentes principais: (i) perspectiva de mundo; (ii) método de abordagem; (iii) procedimentos e técnicas. Adotamos uma perspectiva de mundo sistêmica/autopoiética, conforme proposto por Maturana e Varela (2011) e Luhmann (1983). Isso nos permite ver o problema de pesquisa de uma maneira holística e interconectada. O método de abordagem que empregamos é indutivo, permitindo-nos formular perguntas a partir do problema de pesquisa de uma maneira única e inovadora. Quanto aos procedimentos e técnicas, utilizou-se uma abordagem multicamadas. A primeira camada envolveu uma revisão bibliográfica abrangente sobre a Inteligência Artificial Explicável (*XAI*) e suas aplicações na Administração Pública globalmente. Em seguida, realizamos a coleta de dados sobre o estado da arte por meio de fontes acadêmicas, além de conduzir experimentos para entender a interação de modelos preditivos com o conhecimento jurídico estruturado. Finalmente, com base nesses dados, realizamos experimentos com o modelo explicativo *SHAP* para prever decisões de IA em relação a decisões judiciais envolvendo ações de companhias aéreas. Para cada objetivo específico, aplicamos diferentes procedimentos e técnicas, conforme detalhado na Tabela 1.

**Tabela 1 - Procedimentos e Técnicas**

<b>Objetivo específico</b>	<b>Procedimento</b>	<b>Técnicas</b>
1	Revisão bibliográfica	Pesquisa Descritiva
2	Revisão Sistemática	Pesquisa Descritiva
3	Experimentação com <i>Machine Learning</i>	Experimentação
4	Proposta de aplicação de métodos explicativos	Experimentação
5	Validação da abordagem proposta	Questionário



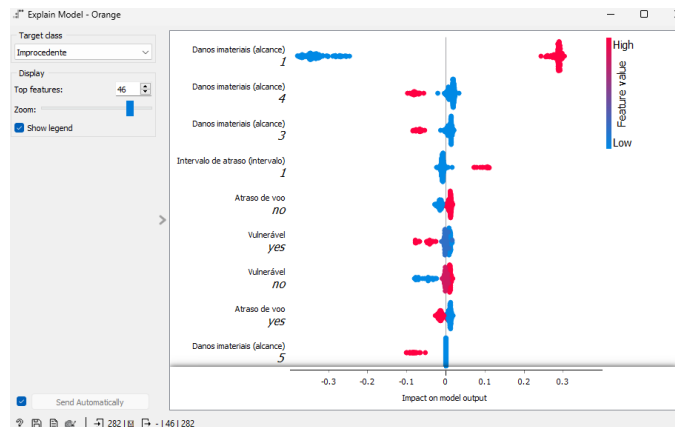
- **Revisão bibliográfica:** Nesta etapa, realizamos uma pesquisa descritiva sobre a IA e suas aplicações na Administração Pública globalmente, buscando compreender o conceito, os benefícios, os desafios e os exemplos de uso da IA em diferentes contextos e domínios. A revisão bibliográfica foi feita a partir de fontes primárias e secundárias, como livros, artigos, relatórios, teses e dissertações, que abordam o tema da IA de forma teórica ou prática. A revisão bibliográfica nos permitiu obter uma visão geral e atualizada sobre a IA e suas implicações para o Direito.
- **Revisão Sistemática de Literatura (SLR):** Nesta etapa, realizamos uma análise das fontes acadêmicas sobre IA, com foco na IA e no Direito. O objetivo foi identificar as principais tendências, lacunas e oportunidades de pesquisa na área, bem como os métodos, as técnicas e as ferramentas utilizadas pelos pesquisadores. A análise foi feita a partir de bases de dados científicas, como *Web of Science*, *Scopus* e *Scielo*, utilizando palavras-chave e critérios de seleção adequados. A análise da revisão sistemática de literatura nos permitiu situar o nosso estudo no cenário acadêmico e definir o nosso diferencial e contribuição.
- **Experimentação com *Machine Learning*:** Nesta etapa, realizamos experimentos para entender a interação de modelos preditivos com o conhecimento jurídico estruturado. O objetivo foi construir e avaliar modelos de IA capazes de prever o resultado de decisões judiciais envolvendo ações de companhias aéreas, utilizando dados extraídos de sentenças do Tribunal de Justiça de Santa Catarina. Os experimentos foram feitos utilizando o software *Orange* e a linguagem de programação *Python*, ambos de código-fonte aberto. Os experimentos envolveram técnicas de pré-processamento, representação, agrupamento, associação, classificação e regressão de dados, utilizando algoritmos de *Machine Learning* e *Natural Language Processing*. A experimentação com *Machine Learning* nos permitiu obter modelos de IA com alto desempenho e precisão.
- **Proposta de aplicação de métodos explicativos:** Nesta etapa, realizamos experimentos com o modelo explicativo *SHAP* para prever decisões de IA em relação a decisões judiciais envolvendo ações de companhias aéreas. O objetivo foi gerar explicações sobre as predições de IA, utilizando o método *SHAP*, que é

uma técnica que atribui valores de importância às características de entrada de um modelo de IA, indicando como elas influenciam na predição do modelo. O método *SHAP* é baseado na teoria dos jogos cooperativos, e tem como vantagens a consistência, a localidade e a aditividade. Os experimentos foram feitos utilizando o software *Orange* e a linguagem de programação *Python*, ambos de código-fonte aberto. A proposta de aplicação de métodos explicativos nos permitiu obter explicações claras, coerentes e compreensíveis sobre as predições de IA.

- Validação da abordagem proposta: Nesta etapa, realizamos a avaliação da nossa abordagem proposta com especialistas na área, com o intuito de verificar a sua viabilidade e eficácia. O objetivo foi obter feedbacks sobre a qualidade, a relevância e a utilidade das explicações geradas pelo método *SHAP*, bem como sobre as limitações e as sugestões de melhoria da nossa abordagem. A avaliação foi feita por meio de observação participante e questionário de pesquisa, aplicados a um grupo de profissionais do Direito, como juízes, advogados e professores. A validação da abordagem proposta nos permitiu aprimorar a nossa proposta e validar os nossos resultados.

A seguir, apresentamos alguns exemplos e ilustrações das explicações geradas pelos experimentos realizados utilizando o método SHAP, para que o leitor possa visualizar melhor como elas funcionam e como elas contribuem para a justificação das predições de IA. A Figura abaixo, mostra um exemplo de uma explicação gerada pelo método SHAP para uma predição de IA em relação a uma decisão judicial envolvendo uma ação de companhia aérea. Nessa explicação, em específico, podemos ver que as características mais importantes para a predição são o resultado procedente ou improcedente. Cada característica tem um valor de importância associado, que indica o quanto ela influencia na predição do modelo. Além disso, cada característica tem uma cor que indica se ela aumenta ou diminui a probabilidade da predição. Por exemplo, nesse caso, os valores em vermelho influenciam muito no resultado, o que significa que ele aumenta a probabilidade da predição, que é de procedência da ação.

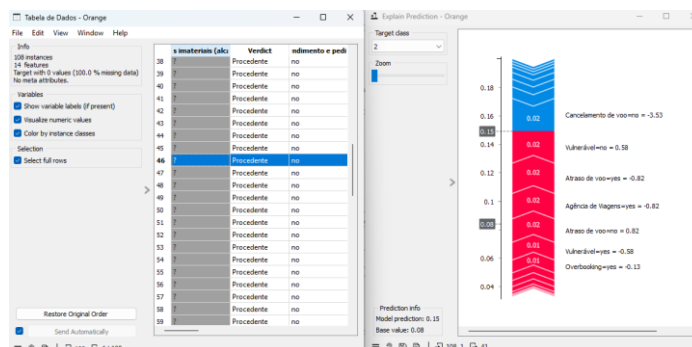
**Figura 1 Ilustração do Explain Model**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

A Figura a seguir, por sua vez, mostra outro exemplo de uma explicação gerada pelo método SHAP para uma previsão de IA em relação a uma decisão judicial envolvendo uma ação de companhia aérea. Nessa explicação, podemos ver que as características mais importantes para a previsão são os intervalos de valor da causa. Além disso, cada característica tem uma cor que indica se ela aumenta ou diminui a probabilidade da previsão. Por exemplo, nesse caso, o tipo de intervalo de valor de causa é vermelho, o que significa que ele aumenta a probabilidade da previsão, que é de ser aquele intervalo o definido ao final da ação.

**Figura 2 Ilustração do Explain Prediction**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Por fim, discutimos as limitações e as sugestões de melhoria da nossa abordagem, embora nem todas tenham encontrado tempo hábil para sua implantação, com base nos feedbacks que recebemos dos especialistas e dos profissionais do Direito. Algumas das

limitações e sugestões de melhoria da nossa abordagem são: (i) ampliar a base de dados utilizada para os experimentos, incluindo mais processos de diferentes tribunais e juizados; (ii) testar outros métodos explicativos além do SHAP, como o LIME e o Anchors, para comparar os resultados e as explicações; (iii) avaliar o impacto das explicações geradas pelo método SHAP na confiança e na satisfação dos usuários da IA jurídica, como as partes, os advogados e os juízes.

#### **1.4 Originalidade**

Este estudo se propõe a contribuir para o avanço do conhecimento sobre a *XAI* e sua aplicação no campo jurídico, por meio de uma abordagem inovadora e interdisciplinar. A originalidade da pesquisa reside na utilização do método *SHAP* para representar o conhecimento jurídico e dar explicabilidade à IA, utilizando dados de decisões judiciais de tribunais. A aplicação desse método no contexto jurídico visa tornar a IA mais transparente e compreensível, fornecendo justificativas para as suas previsões e facilitando a sua avaliação e validação. Dessa forma, o estudo busca preencher uma lacuna na literatura sobre a *XAI* e o Direito, oferecendo uma solução prática e eficiente para os desafios da explicabilidade da IA.

#### **1.5 Estrutura do trabalho**

Este trabalho está organizado em seis capítulos, além dos elementos pré-textuais e pós-textuais. O primeiro capítulo apresenta a introdução, contendo o contexto e a motivação da pesquisa, o problema de pesquisa, a hipótese, os objetivos, a metodologia, a originalidade e a estrutura do trabalho. O segundo capítulo faz uma revisão bibliográfica e sistemática sobre a IA, a *XAI* e a proteção de dados pessoais, abordando as definições, as evoluções históricas e os dilemas éticos desses conceitos. O terceiro capítulo analisa o estado da arte de IA e *XAI* no Direito, explorando as fontes acadêmicas e os estudos de caso existentes. O quarto capítulo descreve os experimentos realizados com o método *SHAP* para representar o conhecimento jurídico e dar explicabilidade à IA, utilizando dados de decisões judiciais de tribunais. O quinto capítulo propõe e valida uma metodologia para a aplicação de métodos explicativos de previsões de IA no judiciário brasileiro, utilizando o método *SHAP* e avaliando a proposta com especialistas. O sexto

capítulo apresenta a conclusão, contendo as considerações finais, as contribuições, as limitações e os trabalhos futuros da pesquisa.

## **2- REVISÃO BIBLIOGRÁFICA**

### **2.1 Definição, evolução história da IA e dilemas éticos**

De início, cabe considerar que o campo multidisciplinar da Inteligência Artificial (IA) que se dedica ao desenvolvimento de sistemas capazes de realizar tarefas que normalmente requerem inteligência humana. Contudo, cabe considerar que a definição e evolução histórica da IA são fundamentais para compreensão do contexto em que essa área de estudo se desenvolveu e quais as contribuições de pesquisadores ao longo do tempo. Para tanto, as contribuições de Stuart Russell e Peter Norvig são relevantes e extremamente pertinentes para entendermos os marcos da evolução da IA.

Russel e Norvig, em sua obra “Inteligência Artificial: Uma Abordagem Moderna”, definem a inteligência como que “relacionada principalmente a uma ação racional. No caso ideal, um agente inteligente adota a melhor ação possível em uma situação”<sup>1</sup>. E, mais além, a IA, através de inúmeros conceitos correlatos apresentados é aquela que não tem por base o ser humano, não tem base biológica. É a inteligência feita por seres humanos para computadores, e como tal, a IA é a ideia de que uma máquina seja capaz de elaborar respostas fora do padrão repetitivo, e se transformar em uma criação do ser humano.

Essa definição enfatiza a capacidade dos sistemas de agirem de forma autônoma, adaptativa e orientada a objetivos, características fundamentais da IA. Para tanto, os autores propõem que a evolução histórica da IA possa ser dividida em algumas fases históricas principais. No início, durante a década de 1950, os primeiros trabalhos nessa área foram influenciados pela ideia de criar máquinas capazes de imitar a inteligência humana<sup>2</sup>. Nesse período, resta impossível ignorar os estudos de Alan Turing, que propôs

---

<sup>1</sup> Russell, Stuart J. 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 55.

<sup>2</sup> [...] Na IBM, Nathaniel Rochester e seus colegas produziram alguns dos primeiros programas de IA. Herbert Gelernter (1959) construiu o Geometry Theorem Prover, que podia demonstrar teoremas que seriam considerados bastante complicados por muitos alunos de matemática. A partir de 1952, Arthur Samuel escreveu uma série de programas para jogos de damas que eventualmente aprendiam a jogar em um nível amador elevado. Ao mesmo tempo, ele contestou a ideia de que os computadores só podem realizar as atividades para as quais foram programados: seu programa aprendeu rapidamente a jogar melhor

o famoso teste de Turing para avaliar a capacidade de uma máquina exibir comportamento inteligente.

Alan Turing quebrou a máquina de código nazista e antes de morrer fez a ousada pergunta: Como eu sei que estou falando com uma máquina? Para tanto, considera o experimento de estar em uma sala isolada, buscando interagir com alguém na sala ao lado, onde não se sabe se existe um computador e/ou um humano. Em seguida, considera alcançado o objetivo da IA quando o indivíduo não consegue distinguir quando está falando com um ser humano, ou com a máquina.<sup>3</sup>

Em seguida, nas décadas de 1960 e 1970, em meio ao entusiasmo inicial e grandes expectativas surgiram os primeiros sistemas de IA baseados em regras lógicas, conhecidos como sistemas especialistas. Já nesse período, os sistemas eram capazes de armazenar conhecimento em forma de regras e aplicá-lo para resolver problemas específicos em domínios restritos<sup>4</sup>. Um exemplo notável dessa época foi o sistema DENDRAL<sup>5</sup>, desenvolvido para realizar análise química.

Contudo, Russel e Norvig, revelam que apesar de ousados nos prognósticos de seus sucessos futuros, os pesquisadores dessa época tiveram que enfrentar uma boa dose de realidade. Como sabemos, atualmente a tradução automática de texto, embora imperfeita, é amplamente utilizada em documentos técnicos, comerciais, governamentais e da Internet. No entanto, logo após o lançamento do Sputnik, em 1957, o National Research Council dos Estados Unidos, amargou a dor de não conseguir obter sucesso nos primeiros esforços de tradução automática, dado o fato de que a tradução exata dos termos

---

que seu criador (Russell, Stuart J. 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 43).

<sup>3</sup> [...] O teste de Turing, proposto por Alan Turing (1950), foi projetado para fornecer uma definição operacional satisfatória de inteligência. O computador passará no teste se um interrogador humano, depois de propor algumas perguntas por escrito, não conseguir descobrir se as respostas escritas vêm de uma pessoa ou de um computador (Russell, Stuart J. 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 25).

<sup>4</sup> [...] O quadro de resolução de problemas que havia surgido durante a primeira década de pesquisas em IA foi o de um mecanismo de busca de uso geral que procurava reunir passos elementares de raciocínio para encontrar soluções completas. Tais abordagens foram chamadas métodos fracos porque, embora gerais, não podiam ter aumento de escala para instâncias de problemas grandes ou difíceis. A alternativa para métodos fracos é usar um conhecimento mais amplo e específico de domínio que permita passos de raciocínio maiores e que possam tratar com mais facilidade casos que ocorrem tipicamente em especialidades estritas. Podemos dizer que, para resolver um problema difícil, praticamente é necessário já saber a resposta (Russell, Stuart J. (Stuart Jonathan), 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 47).

<sup>5</sup> DENDRAL é um sistema especialista e um projeto pioneiro em inteligência artificial, que começou a ser desenvolvido em 1965, na Universidade de Stanford (IBIDEM).

exigiria conhecimento profundo do assunto para solucionar ambiguidades e estabelecer o conteúdo da sentença<sup>6</sup>.

Ainda assim, na década de 1980, a IA se torna uma indústria, com o primeiro sistema especialista comercial bem-sucedido, o R1, que ajudou a configurar pedidos de novos sistemas de computadores<sup>7</sup>. Inclusive, no mesmo período, a abordagem da IA simbólica ganhou destaque, buscando representar o conhecimento e o raciocínio humano por meio de símbolos e regras formais. Surgiram linguagens de programação dedicadas à IA, como o Prolog<sup>8</sup>, e foram desenvolvidas técnicas de representação do conhecimento, como redes semânticas e frames.

A partir da década de 1990, com o avanço da capacidade computacional e a disponibilidade de grandes quantidades de dados, a tradução automática retorna, assim como o desenvolvimento de redes neurais levou ao surgimento do paradigma do aprendizado de máquina (*machine learning*) passou a dominar a área<sup>9</sup>. Basicamente, o aprendizado de máquina se baseia no treinamento de modelos estatísticos a partir de dados, permitindo que eles aprendam padrões e tomem decisões com base nesses padrões. Com efeito, tivemos também o surgimento da tecnologia de mineração de dados (*data mining*), que hoje se vê amplamente utilizada em indústria, comércio e serviços<sup>10</sup>.

Mais adiante, o aprendizado de máquina para se adaptar a novas perguntas e conclusões, associado ao processamento de linguagem natural, representação de conhecimento para armazenar o que se sabe ou ouve; assim como o raciocínio automatizado permitem que aos poucos a IA se assemelhe ao comportamento humano<sup>11</sup>.

De acordo com a análise realizada por PÁDUA<sup>12</sup>, a implementação de tecnologias

---

<sup>6</sup> Russell, Stuart J. (Stuart Jonathan), 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 46.

<sup>7</sup> Russell, Stuart J. (Stuart Jonathan), 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 49.

<sup>8</sup> Ibidem.

<sup>9</sup> Russell, Stuart J. (Stuart Jonathan), 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 52.

<sup>10</sup> STEIW. Leandro. SAIBA COMO FUNCIONA A MINERAÇÃO DE DADOS (OU DATA MINING). Insper. 2022. Disponível em: <<https://www.insper.edu.br/noticias/mineracao-de-dados-ou-data-mining/>>.

<sup>11</sup> Russell, Stuart J. (Stuart Jonathan), 1962- Inteligência artificial / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Pág. 25.

<sup>12</sup> [...] A elaboração de decisões judiciais pode se beneficiar de diversas tecnologias de inteligência artificial (já desenvolvidas ou a serem implementadas), uma vez que essa é a fronteira a ser perseguida na relação entre campos eminentemente jurídicos (teoria da interpretação e teoria da decisão) e inteligência artificial (a qual é dotada de subcampos próprios de estudo) (PÁDUA, Sergio. O Juiz Ciborgue: Inteligência Artificial e Decisão Judicial. CONPEDI. 2020. P.13. Disponível

de inteligência artificial, como o machine learning, na elaboração de decisões judiciais representa uma fronteira a ser explorada na interseção entre os campos jurídicos, como a teoria da interpretação e da decisão, e a inteligência artificial, que possui áreas de estudo específicas.

Nesse sentido, SABO E ROVER<sup>13</sup> argumentam que o aprendizado de máquina pode ser fundamental para a identificação e tratamento de casos jurídicos repetitivos, oferecendo benefícios como a classificação de textos jurídicos para identificar padrões, o gerenciamento eficiente de demandas repetitivas, a aplicação e distinção de padrões em casos similares, e a identificação de processos afetados por questões jurídicas específicas. Essa abordagem, na perspectiva dos autores, possibilita uma maior eficiência, celeridade e precisão no sistema judiciário, contribuindo para uma melhor gestão de casos repetitivos e uma aplicação mais consistente de precedentes legais

PEIXOTO<sup>14</sup>, de igual modo, destaca que a Inteligência Artificial (IA) pode beneficiar o Direito de diversas formas, como reconhecimento de padrões, identificação de consistências, melhoria na gestão de informações, organização de ações estratégicas, garantia de registros confiáveis e apoio à jurisdição. A IA pode ser aplicada em áreas como análise de documentos, pesquisa jurídica, predição de decisões, *compliance* e otimização de diligências, contribuindo para a eficiência, segurança jurídica e qualidade dos serviços jurídicos prestados.

É inegável que, nos últimos anos, com o crescimento exponencial do poder

---

em:<[https://www.academia.edu/98912842/O\\_JUIZ\\_CIBORGUE\\_INTELIG%C3%8ANCIA\\_ARTIFICIAL\\_E\\_DECISAO\\_JUDICIAL?uc-sb-sw=9525500](https://www.academia.edu/98912842/O_JUIZ_CIBORGUE_INTELIG%C3%8ANCIA_ARTIFICIAL_E_DECISAO_JUDICIAL?uc-sb-sw=9525500)>).

<sup>13</sup> [...] A IA não pretende substituir o papel do julgador, do advogado ou dos serventuários da Justiça. Ao contrário, as tecnologias a ela inerentes, uma vez aplicadas aos instrumentos da jurisdição, intenta auxiliar esses papéis no sentido de efetivar o que foi imposto pelo CPC de 2015, elevando-se a funcionalidade e a efetividade da norma, bem como a sua legitimidade. A própria celeridade processual é, como dito, o âmago do processo eletrônico, sem a qual o acesso à Justiça sequer se concretizaria, dado o aumento exponencial de demandas em trâmite no Poder Judiciário (SABO, Isabela Cristina; ROVER, Aires José. Observância de precedentes e gestão de demandas repetitivas por meio do aprendizado de máquina. Revista Opinião Jurídica, v. 18, n. 28, p. 19, 2020. Disponível em:<<https://www.redalyc.org/journal/6338/633868859003/633868859003.pdf>>).

<sup>14</sup> [...] Para o Direito, sistemas de IA podem ser desenvolvidos ou utilizados para comodidade e incremento de desempenho. Desde análises e revisões de documentos; reunião e organização de informações estratégicas; pesquisa jurídica e predição de decisões para determinada tese; compliance e planejamentos sobre passivos, análise e otimização de diligências; ampliação de canais de comunicação e inserção profissional são só algumas áreas do Direito que já relataram experiências interessantes para o apoio da IA (PEIXOTO, Fabiano Hartmann. Direito e inteligência artificial: referenciais básicos com comentários à resolução CNJ 332/2020. Brasília, DF: Ed do autor: DR.IA., 2020. E-book. P.23-24. DOI: 10.29327/521174. Disponível em: [https://drive.google.com/file/d/1eqeHvPFT\\_4OnBMnXUkCFYxcCRcbp\\_Hr2/view](https://drive.google.com/file/d/1eqeHvPFT_4OnBMnXUkCFYxcCRcbp_Hr2/view). Acesso em: 13 abr. 2022).



computacional e o desenvolvimento de algoritmos mais avançados, como redes neurais profundas, a IA tem alcançado resultados impressionantes em diversas áreas, como visão computacional, processamento de linguagem natural e jogos estratégicos.

Nesse sentido, é válido mencionar que a evolução da IA não se restringe apenas a avanços tecnológicos, mas também envolve debates éticos, sociais e legais. A questão da ética no desenvolvimento e uso da IA, bem como a preocupação com a transparência e explicabilidade dos sistemas, são temas cada vez mais relevantes e que precisam ser considerados no contexto atual.

Contudo, justamente nesse ponto, talvez resida um dos maiores obstáculos a serem superados, isso porque ao considerar a abordagem da ética a partir da ótica aristotélica, mais especificamente, através da obra *Ética à Nicômaco*, um dos textos mais influentes da tradição clássica e greco-romana, resta claro que Aristóteles estabelece a ética como uma prática e não um dom<sup>15</sup>.

Em outras palavras, a ética não vem porque nascemos. A ética é uma prática a partir de virtudes, como justiça e temperança. É a base para a felicidade individual e pública<sup>16</sup>. Ou seja, a felicidade política, a felicidade da *polis*.<sup>17</sup> Ela vem da capacidade humana de ter consciência destas virtudes e a capacidade de exercê-las frequentemente, do aperfeiçoamento do ser humano neste aspecto.

É a ética que rege a escolha pelos valores, pelo bom e pelo belo. Diferente de uma ética pela moral, pois a moral seria uma reflexão religiosa. Por que não matar? Na perspectiva cristã, talvez alguém se pronuncie a favor do quinto mandamento, contudo Aristóteles considera admitir que independente de existir um Deus que observe, ou de existir inferno ou paraíso, devemos assumir o compromisso com a vida de outra pessoa e que, portanto, tão ação constitui uma violação das práticas das suas virtudes<sup>18</sup>. Ou seja, ética diz respeito a prática humana. O *Ethos*, o comportamento, diz respeito a prática

---

<sup>15</sup> CULTURAL, Nova. *Ética a Nicômaco*. Aristóteles: Obras incompletas, 1996. Disponível em: <https://abdet.com.br/site/wp-content/uploads/2014/12/%C3%89tica-a-Nic%C3%B4maco.pdf>.

<sup>16</sup> CULTURAL, Nova. *Ética a Nicômaco*. Aristóteles: Obras incompletas, 1996. P.235-237. Disponível em: <https://abdet.com.br/site/wp-content/uploads/2014/12/%C3%89tica-a-Nic%C3%B4maco.pdf>.

<sup>17</sup> Em toda pólis grega encontrava-se as mesmas instituições: assembleia, conselho e magistrados; mas apenas as cidades democráticas punham a tomada de decisões referentes ao conjunto da comunidade nas mãos da assembleia reunindo todos os cidadãos, e apenas nelas o poder judiciário era exercido por juizes escolhidos dentre o conjunto dos cidadãos (MOSSÉ, C. *Dicionário da Civilização Grega*. Trad. Carlos Ramallete. Rio de Janeiro: Jorge Zahar Editor, 2004, p. 240).

<sup>18</sup> CULTURAL, Nova. *Ética a Nicômaco*. Aristóteles: Obras incompletas, 1996. P.45-48. Disponível em: <https://abdet.com.br/site/wp-content/uploads/2014/12/%C3%89tica-a-Nic%C3%B4maco.pdf>.

humana. Logo, no campo das máquinas, da I.A, da Internet das Coisas, dos robôs, será preciso perguntar algo que Aristóteles não fez, a ética fora do campo humano.

O ponto crucial para o caro leitor é compreensão de que ética humana é fundada na busca por fazer a escolha correta; ao passo que a ética da inteligência artificial é desenvolvida para não existir escolha. Enquanto que, na perspectiva do ser humano a ética é formada para dirigir sua liberdade para uma regra social; no contexto da inteligência artificial, tais reflexões são direcionadas para impedir que exista essa liberdade, e mais ainda que ela se transforme em uma ação autônoma e efetiva.

Dito de outro modo, a ética humana gira em torno do fato de que os seres humanos são aleatórios, instáveis, e que, portanto, seja impossível prever do que um ser humano é capaz, pois as pessoas tomam decisões erradas e surtam, são impulsivas e agressivas, tem síndromes e ataques de pânico e, nem sempre, agem de acordo com a razão. Os robôs, por outro lado, devem ser programados para seguir certos princípios e limites, ou seja, para se comportarem de forma consistente e previsível.

Um dos grandes problemas da atualidade é exatamente entender e lidar com avanço capacidade humana de fazer inteligência mais sofisticada associada as máquinas automatizadas como a IA. Essa busca por máquinas que sejam cada vez mais seres humanos melhorados do ponto de vista da inteligência e cada vez menos seres humanos comuns do ponto de vista da moral nos leva ao seguinte apontamento: Seria a IA, de fato, semelhante ao ser humano, embora tolhida de liberdades? Ou seja, despidas de nossas características impulsivas e puramente inteligentes no sentido de tomar decisões sábias, equilibradas e éticas?

Muito embora, as três leis da robótica sejam uma fantasia dos filmes e textos de ficção científica. Asimov imaginou algo que não existe: O cérebro positrônico. Em sua obra “Eu robô”, expõe uma realidade de robôs que estavam condicionados de tal forma por essas regras que nunca poderiam quebrá-las<sup>19</sup>.

---

<sup>19</sup> [...] Temos o seguinte. A primeira: um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano venha a ser ferido.–Certo!–A segunda –continuou Powell –um robô deve obedecer às ordens dadas por seres humanos, exceto nos casos em que tais ordens entrem em conflito com a Primeira Lei.–Certo!–E a terceira: um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira ou com a Segunda Lei” (ASIMOV, Isaac. Eu, robô. Aleph. 2014, p. 65. Disponível em: <[https://books.google.com.br/books?hl=pt-BR&lr=&id=-8-uCgAAQBAJ&oi=fnd&pg=PT4&dq=asimov&ots=IKENCjC--T&sig=\\_KRWvSGvMJmN2LygxDdh8KJLSVQ#v=onepage&q=asimov&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=-8-uCgAAQBAJ&oi=fnd&pg=PT4&dq=asimov&ots=IKENCjC--T&sig=_KRWvSGvMJmN2LygxDdh8KJLSVQ#v=onepage&q=asimov&f=false)>).

Para fins de esclarecimento, a palavra robô deriva da língua Tcheca e quer dizer escravo<sup>20</sup>. Ela surge no século XX, como associada a um autômato, um escravo, que não tem a sua própria vontade. Ocorre que, o ser humano tem que garantir que a inteligência seja tão grande dessa máquina, que essa máquina consiga fazer coisas complexas, que o indivíduo sozinho não consiga fazer, mas não consiga decidir, pois isso cabe ao ser humano.

Dito de outro modo, são humanos melhorados na força física e na inteligência, ao mesmo tempo que são humanos piorados na liberdade e na autonomia. Nesse sentido, Yuval Harari, acredita que cada vez mais será necessário a existência de filósofos para que as indústrias e, a humanidade em geral, possa decidir o que vem a ser inteligência. Porque, a base de todo processo ético é a liberdade. Não existe ética, onde não existe a liberdade<sup>21</sup>.

O ser humano é basicamente alguém que carece de um código ético, porque o ser humano desde o início da sua história foi caracterizado pela liberdade. É o fato de que o indivíduo pode tomar uma atitude impensada, não ética, não previsível, não racional, não adequada, que faz com que seja preciso de ética. A ética é sinônimo de liberdade. As reflexões aristotélicas clássicas e ética estoica<sup>22</sup>, que vem depois, assim como a ética

---

<sup>20</sup> DUARTE, Marcelo. Qual é a origem da palavra “robô”? O Guia dos Curiosos. 2021. Disponível em: <<https://www.guiadoscuriosos.com.br/lingua-portuguesa/qual-e-a-origem-da-palavra-roboto/>>.

<sup>21</sup> “Tudo isso não significa que as fake news não sejam um problema sério, ou que políticos e sacerdotes tenham liberdade total para mentir descaradamente. Também seria errado concluir que tudo são apenas fake news, que toda tentativa de descobrir a verdade está destinada ao fracasso e que não existe diferença entre jornalismo sério e propaganda” (HARARI, Yuval Noah. 21 lições para o século 21. Editora Companhia das Letras, 2018. P.259. Disponível em: <[https://edisciplinas.usp.br/pluginfile.php/7667521/mod\\_resource/content/5/21-Li%C3%A7%C3%B5es-Para-o-S%C3%A9culo-21-by-Yuval-Noah-Harari.pdf](https://edisciplinas.usp.br/pluginfile.php/7667521/mod_resource/content/5/21-Li%C3%A7%C3%B5es-Para-o-S%C3%A9culo-21-by-Yuval-Noah-Harari.pdf)>).

<sup>22</sup> [...] A partir disso, podemos dizer que é preciso compreender a afirmação epictetiana da dupla natureza humana (ou seu duplo parentesco) não platonicamente, mas a partir da doutrina estoica da oikeiosis, significando uma dupla possibilidade para cada ser humano: manter-se no nível da oikeiosis eklektike, própria dos animais irracionais, tornando-se egoísta e incapaz de uma visão abrangente, amorosa e comunitária da humanidade e do Cosmos, ou alçar-se à oikeiosis hairesike, pela qual tem acesso às ações convenientes próprias a cada relação, tornando-se sociável, comunitário e cidadão cósmico. As relações são estabelecidas através de impulsos provenientes dos desdobramentos da oikeiosis e, portanto, oriundos da natureza individual. (DINUCCI, Aldo; JULIEN, Alfredo. O encheirídion de Epicteto. São Cristóvão: Editora UFS, 2012, p. 85-86).

medieval de base moral<sup>23</sup>; ética moderna, especialmente com Spinoza<sup>24</sup>, e a ética de Kant<sup>25</sup>, além da ética pragmática do século XIX e XX<sup>26</sup>; todas elas dialogam com o mesmo princípio liberdade. A liberdade que a humanidade tem de agir que pressupõe um código, com leis instáveis, para tentar conduzir o ser humano de Aristóteles até hoje, para um comportamento mais previsível, mais eficaz e mais próximo de um comportamento útil para o grupo.

Todavia, o fato é de que as máquinas não são nem humanas e nem abelhas. Isso faz com que esse código ético tenha que ser colocado por seres humanos que não tem ética clara e que traduzam em suas máquinas a ética que eles próprios não possuem e assim façam das máquinas, humanos melhorados, mas inferiores no sentido de não terem essa liberdade.

Dessa maneira, quando o ser humano produz um comportamento que seja aceitável, está na verdade dialogando com o grande medo, em primeiro lugar, de que se perca o controle sobre as próprias obras. Em segundo lugar, que perdendo o controle das obras, das máquinas, computadores e inteligência artificial, seja perdido o controle sobre

---

<sup>23</sup> [...] A ética medieval tem a função de ordenar a vida pessoal e social pelo mesmo conjunto de virtudes, em busca da paz e da felicidade. Uma, a paz, entendida como modo de viver em comunidade, remetendo para a dimensão política; outra, a felicidade, como fim do ser humano, valorando a ação como transcendência. A prática das virtudes solicita a aquisição de hábitos para moldar o caráter virtuoso e dotar o de dignidade, responsabilidade e autonomia, apontando-o para o horizonte da paz e da felicidade (SANTOS, Maria Tereza. A ética nos estudos acerca do medievo. *Acta Scientiarum. Education*, v. 42, 2020. P.10. Disponível em: <educa.fcc.org.br/scielo.php?script=sci\_arttext&pid=S2178-52012020000100115#:~:text=A%20ética%20medieval%20tem%20a,valorando%20a%20ação%20como%20transcendência>).

<sup>24</sup> [...] chamo de livre uma coisa que é e age pela exclusiva necessidade de sua natureza; coagida, aquela que está determinada por uma outra a existir e a agir de uma certa maneira determinada. Deus, por exemplo, existe livremente, ainda que necessariamente, porque existe pela exclusiva necessidade de sua natureza. [...] Vós bem o vedes, não faço consistir a liberdade em um decreto, mas numa livre necessidade. Mas desçamos às coisas criadas, que são todas determinadas a existir por causas exteriores e a agir de um modo determinado. Para tornar isso claro e inteligível, concebamos uma coisa muito simples: uma pedra, por exemplo, recebe de uma causa exterior, que a empurra, uma certa quantidade de movimento e, tendo cessado o impulso da causa exterior, ela continuará a se mover necessariamente. Essa persistência da pedra no movimento é uma coação, não porque seja necessária, mas porque se define pelo impulso de uma certa causa externa. [...] toda coisa singular é, necessariamente, determinada por uma causa exterior a existir e a agir de uma certa maneira determinada. (ESPINOSA, Baruch de. *Breve Tratado: de Deus, do homem e do seu bem-estar*. Tradução: Emanuel Angelo da Rocha Fragoso e Luís César Guimarães Oliva. Belo Horizonte: Autêntica, 2014, p. 242-243).

<sup>25</sup> “Age de tal forma que tu queiras que a tua máxima se converse em lei universal” .KANT, Emmanuel. *Fundamentação da Metafísica dos Costumes*. Trad. Paulo Quintela. Lisboa: Edições 70. p.70.

<sup>26</sup> [...] Seguir o caminho da moralidade social resulta forçosamente no temperamento, senão na prática do espírito democrático, pois implica essa experiência humana diversificada resultante da simpatia, que são o fundamento e a garantia da democracia. [...] Temos a obrigação moral de escolher nossas experiências. [...] Sabemos instintivamente que se desprezarmos nossos semelhantes e limitarmos conscientemente nossas relações a certos tipos de pessoas que decidimos respeitar anteriormente, não apenas circunscrevem enormemente nosso âmbito de vida, mas limitamos o escopo de nossa ética. (ADDAMS, Jane. *Charity and Social Justice*. *The North American Review*, v. 192, n. 656, p. 77- 78, jul. 1910).

a própria criação, e que essas máquinas tirem a liberdade dos seres humanos, de acordo com diversas fantasias e distopias em que as máquinas tornam os seres humanos escravos delas mesmas, e os seres humanos são submetidos a essas máquinas.

Evidente que o debate sobre a inteligência artificial tem ganhado cada vez mais relevância, dado que questiona se o progresso tecnológico vai ser acompanhado de maturidade ética, ou seja, se vai ser acompanhando de capacidade e ponderação. Se o ser humano é capaz de criar vida autônoma eticamente, ao ponto de que sua produção não contrarie, não violente, não ataque, não entre em choque com ele mesmo.

TEGMARK em sua obra LIFE 3.0, todavia, foi capaz de traduzir exatamente o ser humano na era da inteligência artificial. O autor faz perguntas muito importantes, tais como: Humanos, vocês estão preparados para não serem a espécie mais inteligente do planeta? Estão preparados para competir com máquinas criadas por seres humanos, que são mais rápidas, tem mais volume de informação, e processa dados de forma mais eficaz do que nós mesmos conseguimos pensar?<sup>27</sup>

Reforça ainda que o esforço ético durante toda a história da espécie humana, partiu do pressuposto que o nosso protagonismo como seres inteligentes era o único. Não há na história da espécie humana, nenhum outro ser que tenha rivalizado com ser humano em inteligência. Somos desde a origem na pré-história até hoje os seres mais inteligentes do planeta, e, não obstante, graças a isso neste momento que esse ser humano que aliás é mais fraco que a maior parte dos animais, menos eficaz que a maior parte das sociedades animais, como a sociedade das abelhas ou das formigas, consegue desenvolver uma máquina inteligente<sup>28</sup>.

Ou seja, temos que atualmente a inteligência artificial (IA) é uma força poderosa que pode transformar o futuro da humanidade, tanto para o bem quanto para o mal. Além disso, a IA é um campo vasto que abrange diferentes paradigmas e abordagens, cada uma com suas características e métodos específicos. Contudo, a compreensão desses

---

<sup>27</sup> [...] Nós humanos construímos nossa identidade sendo Homo sapiens, as entidades mais inteligentes por aqui. À medida que nos preparamos para sermos humilhados por máquinas cada vez mais inteligentes, eu sugiro que nos reinventemos como Homo sentiens (TEGMARK, Max. Life 3.0. Being human in the age of artificial intelligence. New York: Penguin, 2017. p. 399. Disponível em: <[https://brill.com/view/journals/rt/26/1-2/article-p169\\_8.xml](https://brill.com/view/journals/rt/26/1-2/article-p169_8.xml)>).

<sup>28</sup> [...] Tradicionalmente, nós humanos temos fundado nosso valor próprio na ideia de excepcionalismo humano: a convicção de que somos as entidades mais inteligentes do planeta e, portanto, únicas e superiores. O surgimento da IA nos forçará a abandonar isso e nos tornarmos mais humildes.” (TEGMARK, Max. Life 3.0. Being human in the age of artificial intelligence. New York: Penguin, 2017. p. 398. Disponível em: <[https://brill.com/view/journals/rt/26/1-2/article-p169\\_8.xml](https://brill.com/view/journals/rt/26/1-2/article-p169_8.xml)>).

paradigmas e abordagens é fundamental para explorar a diversidade de técnicas e abordagens empregadas no desenvolvimento de sistemas inteligentes.

Um dos mais conhecidos, talvez seja o paradigma simbólico, também conhecido como abordagem baseada em conhecimento. Ele enfatiza a representação e manipulação de símbolos e regras lógicas. Essa abordagem busca modelar o conhecimento humano utilizando linguagens formais e sistemas lógicos. Isso porque, o objetivo é representar o raciocínio humano por meio de símbolos e realizar inferências lógicas a partir disso, como ocorre com os sistemas especialistas e ontologias.

Por outro lado, temos o paradigma conexionista também chamado de abordagem baseada em redes neurais, dado que se inspira no funcionamento do cérebro humano. Válido ressaltar que, nessa abordagem, são utilizadas redes neurais artificiais compostas por unidades de processamento interconectadas. Cada unidade representa um neurônio artificial, e as conexões entre as unidades são responsáveis pelo processamento paralelo e distribuído de informação. Basicamente, as redes neurais são treinadas para aprender a partir de exemplos e reconhecer padrões complexos nos dados.

Como se não bastasse, existe ainda o paradigma probabilístico que emprega técnicas estatísticas e teoria de probabilidade para lidar com a incerteza presente em muitos problemas de IA. Para tanto, faz-se uso de modelos probabilísticos para representar e inferir informações incertas. Diversas técnicas, como redes Bayesianas, processos de decisão de Markov (MDPs) e campos aleatórios de Markov (MRFs), são utilizadas para modelar e resolver problemas complexos que envolvem incerteza e tomada de decisão.

Ademais, é importante ter em mente que tais paradigmas, muitas vezes, são combinados para abordar problemas específicos. Ou seja, a escolha do paradigma ou abordagem adequados depende das características do problema, dos recursos disponíveis e das metas desejadas. Inclusive, a combinação desses diferentes paradigmas e abordagens contribui para a diversidade e riqueza da área de IA permitindo a criação de sistemas inteligentes adaptáveis e eficazes em diferentes áreas, incluindo, o direito.

Até então, abordamos alguns conceitos gerais e questões éticas envolvendo a inteligência artificial, como sua definição, história, aplicações, desafios e dilemas. Vimos que a IA é um campo multidisciplinar e complexo, que requer uma reflexão crítica e responsável sobre seus impactos na sociedade e no meio ambiente. No próximo

subcapítulo, vamos nos aprofundar em um dos aspectos mais importantes da IA: os modelos de aprendizado de máquina, que são algoritmos capazes de aprender a partir de dados e realizar tarefas específicas. Veremos como eles funcionam, quais são seus tipos, vantagens e limitações, e como avaliá-los.

## 2.2 Aprendizado de Máquina (*Machine Learning*)

O aprendizado de máquina, também conhecido como *machine learning*<sup>29</sup>, é um ramo da Inteligência Artificial que se dedica ao desenvolvimento de algoritmos e técnicas capazes de permitir que os computadores aprendam a partir dos dados, sem serem explicitamente programados. Essa abordagem permite que as máquinas adquiram conhecimento, identifiquem padrões, façam previsões e tomem decisões com base nas informações disponíveis.

Inclusive, o processo de aprendizado ocorre por meio da análise e interpretação dos dados. Os algoritmos são projetados para reconhecer padrões e extrair informações relevantes, permitindo que os sistemas automatizados realizem tarefas específicas ou forneçam insights valiosos. Esses algoritmos são alimentados com dados de treinamento, que consistem em exemplos ou informações previamente rotuladas.

Nesse sentido, existem diferentes tipos de aprendizado de máquina com abordagens que se aplicam aos diferentes contextos e objetivos. O aprendizado supervisionado é uma abordagem em que os algoritmos são treinados usando dados de entrada e saída rotulados. Ou seja, o conjunto de treinamento contém exemplos em que as respostas corretas são fornecidas<sup>30</sup>.

O objetivo é que o algoritmo aprenda a mapear os dados de entrada para as saídas

---

<sup>29</sup> Arthur Samuel, pioneiro americano da era do computador, trabalhou inicialmente na IBM. O artigo em que descreve seu trabalho sobre jogo de damas foi o primeiro a usar o termo aprendizado de máquina, embora Alan Turing já tivesse falado em “máquina que pode aprender com a experiência”, em 1947. Arthur Samuel, “Some studies in machine learning using the game of checkers”. (STUART, Russel. Inteligência Artificial a nosso favor: como manter o controle sobre a tecnologia. São Paulo: Companhia das Letras, 2021. p.457 *In* IBM Journal of Research and Development, v. 3, pp. 210-29, 1959).

<sup>30</sup> [...] A forma mais comum de aprendizado de máquina é chamada de aprendizado supervisionado. Um algoritmo de aprendizado supervisionado recebe uma coleção de exemplos para treinamento, cada um deles rotulado com o output correto, e deve produzir uma hipótese sobre qual é a regra correta. Normalmente, um sistema de aprendizado supervisionado busca otimizar a concordância entre as hipóteses e os exemplos para treinamento (STUART, Russel. Inteligência Artificial a nosso favor: como manter o controle sobre a tecnologia. São Paulo: Companhia das Letras, 2021. p.440).

corretas, permitindo a previsão de novos exemplos. Por exemplo, em um problema de classificação, o algoritmo aprende a atribuir rótulos a diferentes classes com base nas características dos dados. Alguns algoritmos comuns de aprendizado supervisionado incluem regressão linear<sup>31</sup>, regressão logística<sup>32</sup> e árvores de decisão<sup>33</sup>.

Por outro lado, no aprendizado não supervisionado, os algoritmos são aplicados a conjuntos de dados não rotulados. Diferentemente do aprendizado supervisionado, não há rótulos fornecidos, e o objetivo é descobrir padrões, estruturas ou agrupamentos nos dados<sup>34</sup>. Os algoritmos de aprendizado não supervisionado são usados principalmente para explorar os dados e obter insights valiosos. Por exemplo, algoritmos de *clusterização*<sup>35</sup> agrupam os dados com base em sua similaridade, enquanto algoritmos de redução de dimensionalidade procuram reduzir a complexidade dos dados, preservando suas características essenciais, como ocorre, por exemplo com o *K-means*<sup>36</sup>.

---

<sup>31</sup> A análise de regressão linear é usada para prever o valor de uma variável com base no valor de outra. A variável que deseja prever é chamada de variável dependente. A variável que é usada para prever o valor de outra variável é chamada de variável independente (Regressão Linear. IBM. 2023. Disponível em: <<https://www.ibm.com/br-pt/analytics/learn/linear-regression>>).

<sup>32</sup> [...] Esse tipo de modelo estatístico (também conhecido como modelo logit) é frequentemente usado para classificação e análise preditiva. A regressão logística estima a probabilidade de ocorrência de um evento, como um voto, com base em um determinado conjunto de dados de variáveis independentes. Como o resultado é uma probabilidade, a variável dependente é limitada entre 0 e 1 (O que é regressão logística. IBM. 2023. Disponível em: <<https://www.ibm.com/br-pt/topics/logistic-regression>>).

<sup>33</sup> [...] Esta abordagem, às vezes conhecida como indução de regra, tem várias vantagens. Primeiro, o processo de raciocínio por trás do modelo é evidente quando navegar pela árvore. Isso está em contraste com outras técnicas de modelagem caixa preta em que a lógica interna pode ser difícil de trabalhar. Segundo, o processo inclui automaticamente em sua regra apenas os atributos que realmente forem importantes para uma tomada de decisão. Os atributos que não contribuírem com a precisão da árvore são ignorados. Isso pode produzir informações muito úteis sobre os dados e também ser utilizado para reduzir os dados para campos relevantes antes de treinar outra técnica de aprendizado, como uma rede neural (Modelos de Árvore de Decisão. IBM. 2023. Disponível em: <<https://www.ibm.com/docs/pt-br/spss-modeler/18.4.0?topic=trees-decision-tree-models>>).

<sup>34</sup> Aprendizado não supervisionado, também conhecido como *machine learning* não supervisionado, usa algoritmos de *machine learning* para analisar e agrupar conjuntos de dados não rotulados. Esses algoritmos descobrem padrões ocultos ou agrupamentos de dados sem a necessidade de intervenção humana. Sua capacidade de descobrir semelhanças e diferenças nas informações o torna a solução ideal para análise exploratória de dados, estratégias de vendas cruzadas, segmentação de clientes e reconhecimento de imagem. (O que é aprendizado não supervisionado. IBM. 2023. Disponível em: <<https://www.ibm.com/br-pt/topics/unsupervised-learning>>).

<sup>35</sup> [...] Cluster, em inglês, significa grupo. Portanto, clusterizar nada mais é do que agrupar. Esse agrupamento, por sua vez, pode ser de um conjunto de dados, de clientes, de computadores ou o que mais for necessário. Assim, o termo é utilizado com mais frequência por desenvolvedores, profissionais de marketing, TI ou cientistas de dados, os quais recorrem à clusterização como forma de organizar dados ou segmentá-los (Clusterização: o que é, importância e aplicações. Fiveacts. 2021. Disponível em: <<https://www.fiveacts.com.br/clusterizacao/>>).

<sup>36</sup> No algoritmo conhecido como *K-means*, clusters diferentes podem ser dispostos em um gráfico hipotético. Nele, o centro de cada cluster chama-se centroide, no qual é calculada a média dos valores para um cluster. Desse modo, o algoritmo buscará o centroide mais perto, usando métricas de distância e atribuindo um ponto encontrado ao cluster mais próximo dele (Clusterização: o que é, importância e aplicações. Fiveacts. 2021. Disponível em: <<https://www.fiveacts.com.br/clusterizacao/>>).



Outrossim, temos também o aprendizado por reforço<sup>37</sup>, no qual os algoritmos aprendem a tomar decisões em um ambiente dinâmico para maximizar uma recompensa. Nessa abordagem, um agente interage com um ambiente e recebe *feedback* na forma de recompensas ou punições com base em suas ações. O objetivo do agente é aprender a tomar ações que levem a recompensas acumuladas ao longo do tempo. Não raro, o aprendizado por reforço é aplicado em *games* e robótica.

Como é de se esperar, a maneira como ensinamos a IA afeta a forma como ela aprende. Por fim, não menos importante, o aprendizado semi-supervisionado<sup>38</sup> que combina elementos do aprendizado supervisionado e não supervisionado. Nessa abordagem, o conjunto de treinamento contém tanto exemplos rotulados quanto não rotulados. A ideia é aproveitar a informação contida nos dados não rotulados para melhorar o desempenho do algoritmo de aprendizado. Isso é especialmente útil quando obter rótulos para todos os dados é custoso ou impraticável. Algoritmos de aprendizado semi-supervisionado buscam aprender tanto a estrutura subjacente dos dados quanto as relações entre os rótulos conhecidos.

Esses são os principais tipos de aprendizado de máquina. Cada abordagem tem suas características e aplicações específicas, e a escolha do tipo de aprendizado depende da natureza do problema e dos dados disponíveis. O aprendizado supervisionado é adequado quando há um conjunto de dados rotulados para treinar o algoritmo e prever saídas para novos exemplos. O aprendizado não supervisionado é útil quando se deseja explorar e descobrir padrões nos dados não rotulados. O aprendizado por reforço é

---

<sup>37</sup> A Aprendizagem Por Reforço é o treinamento de modelos de aprendizado de máquina para tomar uma sequência de decisões. O agente aprende a atingir uma meta em um ambiente incerto e potencialmente complexo. No aprendizado por reforço, o sistema de inteligência artificial enfrenta uma situação. O computador utiliza tentativa e erro para encontrar uma solução para o problema. Para que a máquina faça o que o programador deseja, a inteligência artificial recebe recompensas ou penalidades pelas ações que executa. Seu objetivo é maximizar a recompensa total. Embora o Cientista de Dados ou o Engenheiro de IA defina a política de recompensa – isto é, as regras do jogo – ele não dá ao modelo nenhuma dica ou sugestão de como resolver o jogo. Cabe ao modelo descobrir como executar a tarefa para maximizar a recompensa, começando com testes totalmente aleatórios e terminando com táticas sofisticadas (O Que é Aprendizagem Por Reforço. Data Science Academy. 2022. Disponível em: <<https://www.deeplearningbook.com.br/o-que-e-aprendizagem-por-reforco/>>).

<sup>38</sup> [...] Essa nova forma de aprendizado, denominada aprendizado semi-supervisionado é bastante útil quando apenas um pequeno número de exemplos rotulados encontra-se disponível, o que ocorre na grande maioria dos casos. O aprendizado semi-supervisionado pode ser utilizado tanto em tarefas de classificação, quando os exemplos rotulados são utilizados no processo de rotulação de exemplos, em tarefas de *clustering*, sendo os exemplos rotulados responsáveis por auxiliar o processo de formação de *clusters* (SANCHES, Marcelo Kaminski. Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados. 2003. Tese de Doutorado. Universidade de São Paulo. p.71. Disponível em: <[https://teses.usp.br/teses/disponiveis/55/55134/tde-12102003-140536/publico/Dissertacao\\_MKS.pdf](https://teses.usp.br/teses/disponiveis/55/55134/tde-12102003-140536/publico/Dissertacao_MKS.pdf)>).

apropriado para situações em que o agente aprende a tomar decisões com base em recompensas obtidas por meio da interação com o ambiente. Já o aprendizado semi-supervisionado é vantajoso quando apenas alguns exemplos estão rotulados e é possível aproveitar informações adicionais dos dados não rotulados.

Frisa-se que esses tipos de aprendizado de máquina não são exclusivos e muitas vezes diferentes abordagens podem ser combinadas para abordar problemas mais complexos. Além disso, a evolução contínua do campo do aprendizado de máquina tem levado ao surgimento de novas técnicas e abordagens, ampliando ainda mais as possibilidades de aplicação e os métodos disponíveis.

Ademais, é feito uso de diversos algoritmos no campo do aprendizado de máquina que podem fornecer respostas e ajudar a resolver uma variedade de problemas. As árvores de decisão são algoritmos de aprendizado supervisionado que utilizam estruturas em forma de árvore para representar decisões e suas possíveis consequências<sup>39</sup>. Esses algoritmos dividem os dados com base em diferentes características, criando uma estrutura hierárquica de decisões. Cada nó interno da árvore representa uma característica e cada folha representa uma decisão ou resultado final. As árvores de decisão são fáceis de entender e interpretar, além de serem úteis para lidar com problemas de classificação e regressão.

Igualmente, as máquinas de vetores de suporte, ou *support vector machines* (*SVM*), são algoritmos de aprendizado supervisionado que mapeiam os dados em um espaço de alta dimensão e encontram um hiperplano que separa as diferentes classes de forma otimizada<sup>40</sup>. Esses algoritmos são eficazes em problemas de classificação binária e podem lidar com dados não linearmente separáveis por meio do uso de funções de kernel. Válido mencionar ainda que as SVMs são amplamente utilizadas em áreas como

---

<sup>39</sup> As árvores de decisão são modelos mais complexos do que os drivers unidirecionais e bidirecionais. Elas ampliam a sequência, assim como os modelos de combinação. A principal diferença é que as árvores de decisão suportam a descoberta de interação entre vários preditores e, portanto, oferecem insights mais profundos do que os drivers (Árvore de decisão. IBM. 2023. Disponível em:<<https://www.ibm.com/docs/pt-br/cognos-analytics/11.1.0?topic=analytics-decision-tree>>).

<sup>40</sup> Support Vector Machine (SVM) é uma classificação e regressão robusta técnica que maximiza a acurácia preditiva de um modelo sem sobreajustar os dados de treinamento. O SVM é particularmente adequado para analisar dados com números muito grandes (por exemplo, milhares) de campos preditores. O SVM tem aplicações em muitas disciplinas, incluindo relacionamento com o cliente gerenciamento (CRM), reconhecimento facial e outras imagens, bioinformática, conceito de mineração de texto extração, detecção de intrusão, predição de estrutura de proteínas e reconhecimento de voz e fala (Sobre o SVM. IBM. 2023. Disponível em:<<https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-about-svm>>).

reconhecimento de padrões, bioinformática e processamento de imagens.

Para além dos acima mencionados, temos também as redes neurais artificiais, que são modelos inspirados no funcionamento do cérebro humano. Elas consistem em redes de unidades interconectadas, chamadas de neurônios artificiais, que processam e transmitem informações<sup>41</sup>. Esses algoritmos são capazes de aprender a partir dos dados, ajustando os pesos das conexões entre os neurônios. As redes neurais profundas, também conhecidas como Deep Learning, são um tipo de rede neural com múltiplas camadas ocultas, que têm alcançado resultados impressionantes em áreas como visão computacional, processamento de linguagem natural e reconhecimento de fala.

Outro algoritmo notoriamente conhecido e utilizado é o algoritmo K-Nearest Neighbors, que funciona através de uma abordagem de aprendizado supervisionado que classifica os dados com base na proximidade em relação aos exemplos de treinamento<sup>42</sup>. Ele atribui uma classe aos novos exemplos com base na maioria das classes dos “k” vizinhos mais próximos. O valor de “k” é um parâmetro definido pelo usuário. O KNN é simples de implementar e pode ser utilizado tanto para problemas de classificação quanto de regressão.

Por fim, temos a regressão logística que é um algoritmo amplamente utilizado em IA para resolver problemas de classificação, dado que é uma técnica estatística que permite prever a probabilidade de um evento binário ocorrer com base em variáveis

---

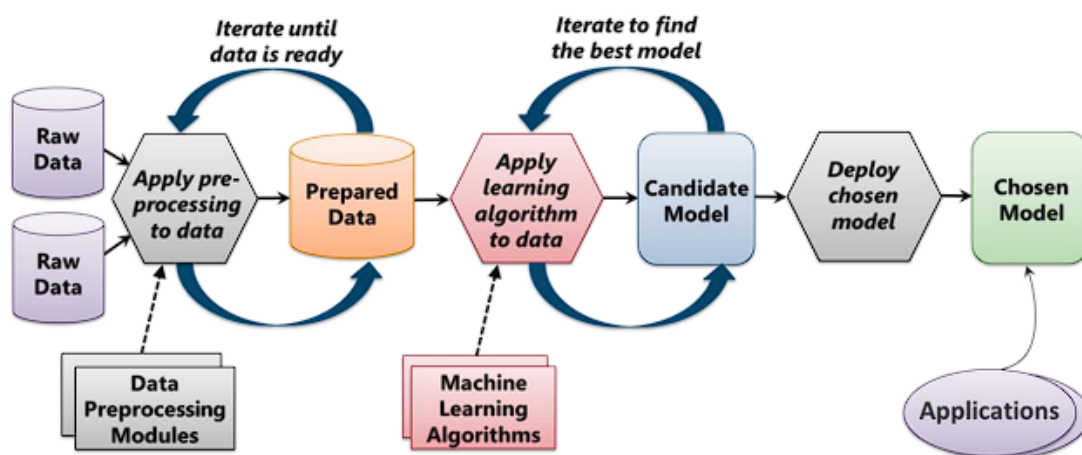
<sup>41</sup> As redes neurais, também conhecidas como redes neurais artificiais (ANNs) ou redes neurais simuladas (SNNs), são um subconjunto de machine learning e estão no cerne dos algoritmos de deep learning. Seu nome e estrutura são inspirados no cérebro humano, imitando a maneira como os neurônios biológicos enviam sinais uns para os outros. As redes neurais artificiais (ANNs) são compostas por camadas de um nó, contendo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada nó, ou neurônio artificial, conecta-se a outro e tem um peso e um limite associados. Se a saída de qualquer nó individual estiver acima do valor do limite especificado, esse nó será ativado, enviando dados para a próxima camada da rede. Caso contrário, nenhum dado será transmitido junto à próxima camada da rede (O que são Redes Neurais? IBM. 2023. Disponível em:<<https://www.ibm.com/br-pt/topics/neural-networks>>).

<sup>42</sup> O algoritmo K-Nearest Neighbors, também conhecido como KNN ou k-NN, é um classificador de aprendizagem supervisionado não paramétrico, que usa a proximidade para fazer classificações ou previsões sobre o agrupamento de um ponto de dados individual. Embora possa ser usado para problemas de regressão ou classificação, ele é normalmente usado como um algoritmo de classificação, trabalhando com a suposição de que pontos semelhantes podem ser encontrados próximos uns dos outros. Para problemas de classificação, um rótulo de classe é atribuído com base em um voto da maioria, ou seja, o rótulo que é mais frequentemente representado em torno de um determinado ponto de dados é usado. Embora isso seja tecnicamente considerado "voto plural", o termo "voto majoritário" é mais comumente usado na literatura. A distinção entre essas terminologias é que o "voto majoritário" tecnicamente requer uma maioria superior a 50%, o que funciona principalmente quando há apenas duas categorias (What is the k-nearest neighbors algorithm? IBM. 2023. Disponível em:<<https://www.ibm.com/topics/knn>>).

independentes<sup>43</sup>. O funcionamento da regressão logística envolve o uso da função sigmoide para modelar a probabilidade de um evento binário. Durante o treinamento do modelo, os coeficientes são ajustados para minimizar a diferença entre as probabilidades preditas pelo modelo e as classes reais no conjunto de dados de treinamento. Uma vez que o modelo tenha sido treinado, ele pode ser usado para fazer previsões em novos dados.

Esses são apenas alguns exemplos dos principais algoritmos de aprendizado de máquina. Evidente que existem muitos outros algoritmos com diferentes características e propriedades que podem ser aplicados de acordo com a natureza do problema e dos dados disponíveis. Contudo, a escolha do algoritmo adequado depende das características dos dados, do tipo de problema e dos objetivos da aplicação.

**Figura 3 Ilustração do Fluxo de Trabalho do Aprendizado de Máquina**



Fonte: CHAPPELL, David. *Introducing azure machine learning. A guide for technical professionals*, sponsored by microsoft corporation, 2015. p.5. Disponível em:<[http://www.davidchappell.com/writing/white\\_papers/Introducing-Azure-ML-v1.0--Chappell.pdf](http://www.davidchappell.com/writing/white_papers/Introducing-Azure-ML-v1.0--Chappell.pdf)>.

Ademais, o processo de treinamento e avaliação de modelos de aprendizado de máquina envolve várias etapas interligadas. Desse modo, antes de treinar um modelo, é necessário pré-processar os dados<sup>44</sup>. Isso envolve a limpeza dos dados, lidando com

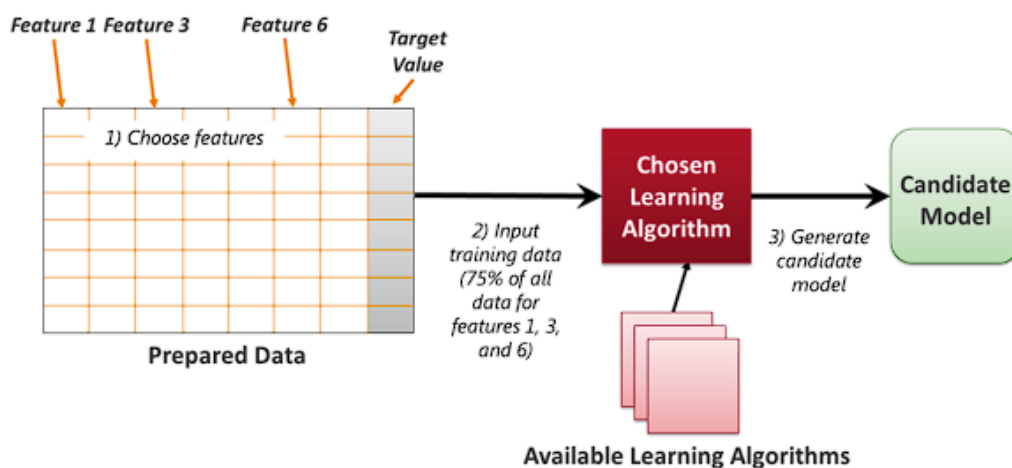
<sup>43</sup> A regressão logística é útil para situações nas quais você deseja poder prever a presença ou ausência de uma característica ou resultado com base em valores de um conjunto de variáveis preditoras. É semelhante a um modelo de regressão linear, mas é adequado para modelos em que a variável dependente é dicotômica. Os coeficientes de regressão logística podem ser usados para estimar razões de chances para cada uma das variáveis independentes no modelo. A regressão logística é aplicável a um intervalo mais amplo de situações de pesquisa do que de análise discriminante (Regressão Logística. IBM. 2023. Disponível em:<<https://www.ibm.com/docs/pt-br/spss-statistics/saas?topic=regression-logistic>>).

<sup>44</sup> [...] Sua aplicação é necessária, pois as bases de dados em geral são muito grandes (gigabytes ou mais) e contêm registros que comprometem a qualidade dos dados, como por exemplo, registros inconsistentes,

valores ausentes, tratando dados categóricos, normalizando ou padronizando os dados, entre outras técnicas. Essa etapa é essencial para garantir que os dados estejam em um formato apropriado para o treinamento do modelo.

Superada a etapa inicial, temos que os dados devem ser divididos em conjuntos de treinamento e teste. O conjunto de treinamento é usado para treinar o modelo, enquanto o conjunto de teste é usado para avaliar o desempenho do modelo em dados não vistos anteriormente. Igualmente, a proporção de dados para treinamento e teste pode variar dependendo do tamanho do conjunto de dados e da natureza do problema.

**Figura 4 Ilustração da preparação de dados para Aprendizado de Máquina**



Fonte: CHAPPELL, David. *Introducing azure machine learning. A guide for technical professionals*, sponsored by microsoft corporation, 2015. p.12. Disponível em:<[http://www.davidchappell.com/writing/white\\_papers/Introducing-Azure-ML-v1.0--Chappell.pdf](http://www.davidchappell.com/writing/white_papers/Introducing-Azure-ML-v1.0--Chappell.pdf)>.

Em seguida, surge a necessidade então de escolha do algoritmo e treinamento do modelo, isto é, com os dados de treinamento preparados, é hora de escolher o algoritmo adequado e treinar o modelo. Evidente que, como vimos anteriormente, a escolha do algoritmo depende do tipo de problema e dos objetivos da aplicação, levando em

---

falta de informação (registros faltantes), registros duplicados, outliers (valores discrepantes), assimetria, transformação entre outros. Nesse sentido, as técnicas de pré-processamento podem ser divididas em: a) Limpeza dos dados (data cleaning); Integração de dados (data integration); Transformação de dados (data transformation); Redução de dados (data reduction). (SCHMITT, Jeovani et al. *Pré-processamento para a mineração de dados: uso da análise de componentes principais com escalonamento ótimo*. 2005. p.25 in KAMBER, Jiawei Han Micheline; 范明. *数据挖掘概念与技术*. Data Mining Concepts and Techniques, v. 8, 2001, p.105. Disponível em:<<http://repositorio.ufsc.br/handle/123456789/101803>>).

consideração as características dos dados. Durante o treinamento, o modelo aprende a partir dos dados de treinamento, ajustando seus parâmetros ou pesos para encontrar padrões e fazer previsões.

Encaminhando-se para o final, temos a avaliação do modelo que se dá após o treinamento, com nítido interesse de avaliar o desempenho do modelo. Tal avaliação é feita utilizando o conjunto de teste, que contém dados não vistos anteriormente pelo modelo. Para tanto, pode ser feito o uso de métricas como acurácia, precisão, *recall*<sup>45</sup>, *F1-score*<sup>46</sup> e medir a qualidade das previsões do modelo. Igualmente, gráficos de curvas de precisão-*recall* ou curvas ROC<sup>47</sup> podem ser utilizados para analisar o desempenho do modelo em diferentes limiares de decisão.

Por fim, após o treinamento e a avaliação, o modelo pode ser implantado em um ambiente de produção, onde é utilizado para fazer previsões em tempo real. Ainda assim, é importante realizar o monitoramento contínuo de desempenho do modelo e fazer ajustes conforme necessário. Inclusive, novos dados podem ser coletados e utilizados para retrainar ou atualizar o modelo periodicamente. Em outras palavras, à medida que novos dados se tornam disponíveis e os objetivos da aplicação mudam, o modelo pode precisar ser reavaliado, reajustado ou até mesmo substituído por um algoritmo mais adequado.

Inclusive, é importante ressaltar que o processo de treinamento e avaliação de modelos de aprendizado de máquina requer cuidado e atenção para evitar problemas

---

<sup>45</sup> [...] O recall é a porcentagem de dados classificados como positivos comparado com a quantidade real de positivos que existem em nossa amostra. (PÁDUA. Mateus. Machine Learning -Métricas de avaliação: Acurácia, Precisão e Recall, F1-score. MEDIUM. 2020. P.3-4. Disponível em:<<https://medium.com/@mateuspdua/machine-learning-m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-e-recall-d44c72307959>>).

<sup>46</sup> [...] F1-Score é a métrica que une precisão e recall a fim de trazer um número único que determine a qualidade geral do modelo. (PÁDUA. Mateus. Machine Learning -Métricas de avaliação: Acurácia, Precisão e Recall, F1-score. MEDIUM. 2020. P.3-4. Disponível em:<<https://medium.com/@mateuspdua/machine-learning-m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-e-recall-d44c72307959>>).

<sup>47</sup> [...] Há uma série de estimadores estatísticos que são empregados na análise do desempenho de modelos classificatórios, e um dos mais utilizados é a curva ROC (receiver operating characteristic), que consiste em uma representação gráfica da performance de um modelo de dados quantitativos segundo sua taxa de sensibilidade (fração dos verdadeiros positivos) e a fração dos falsos positivos (1-especificidade), segundo diferentes valores do teste (POLO, Tatiana Cristina Figueira; MIOT, Hélio Amante. Aplicações da curva ROC em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*, v. 19, p. 1-2, 2020. in Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J*. 2017;34(6):357-9. <http://dx.doi.org/10.1136/emered-2017-206735>. PMID:28302644; . Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8(4):283-98. [http://dx.doi.org/10.1016/S0001-2998\(78\)80014-2](http://dx.doi.org/10.1016/S0001-2998(78)80014-2). PMID:112681. Disponível:<<https://www.scielo.br/j/jvb/a/8S8Pfqnz8csmQJVqwgZT8gH/>>).

como *overfitting*<sup>48</sup> ou *underfitting*<sup>49</sup>. O *overfitting* ocorre quando o modelo se ajusta excessivamente aos dados de treinamento e não generaliza bem para novos dados. Já o *underfitting* ocorre quando o modelo não é capaz de capturar os padrões dos dados de treinamento e apresenta um desempenho fraco tanto nos dados de treinamento quanto nos dados de teste.

Em resumo, o processo de treinamento e avaliação de modelos de aprendizado de máquina envolve a preparação dos dados, a escolha do algoritmo adequado, o treinamento do modelo, a avaliação do desempenho, o ajuste de parâmetros e monitoramento após a sua implementação. Essas etapas são fundamentais para o desenvolvimento de modelos de aprendizado de máquina eficazes, que sejam capazes de fazer previsões precisas e generalizáveis em diferentes contextos e aplicações.

### 2.2.1 Inteligência Artificial Explicável (XAI)

A inteligência artificial explicável, ou *explainable artificial intelligence (XAI)*<sup>50</sup>, refere-se à capacidade de compreender e explicar os processos de tomada de decisão dos modelos de inteligência artificial (IA). Enquanto os modelos de IA têm demonstrado um desempenho impressionante em várias tarefas complexas, como reconhecimento de padrões, tradução automática e diagnóstico médico, muitas vezes eles são considerados

---

<sup>48</sup> [...] Um cenário de *overfitting* ocorre quando, nos dados de treino, o modelo tem um desempenho excelente, porém quando utilizamos os dados de teste o resultado é ruim. Podemos entender que, neste caso, o modelo aprendeu tão bem as relações existentes no treino, que acabou apenas decorando o que deveria ser feito, e ao receber as informações das variáveis preditoras nos dados de teste, o modelo tenta aplicar as mesmas regras decoradas, porém com dados diferentes esta regra não tem validade, e o desempenho é afetado. É comum ouvirmos que neste cenário o modelo treinado não tem capacidade de generalização (*Underfitting* e *Overfitting*. Didática Tech. 2023. p.2. Disponível:<<https://didatica.tech/underfitting-e-overfitting/>>).

<sup>49</sup> [...] Neste cenário o desempenho do modelo já é ruim no próprio treinamento. O modelo não consegue encontrar relações entre as variáveis e o teste nem precisa acontecer. Este modelo já pode ser descartado, pois não terá utilidade. Vamos entender estes conceitos com exemplos práticos e visuais. Na figura abaixo vemos uma distribuição de dados entre as variáveis x e y. Consideremos que estes são os dados que temos a nossa disposição para treinar e testar o modelo (*Underfitting* e *Overfitting*. Didática Tech. 2023. p.2. Disponível:<<https://didatica.tech/underfitting-e-overfitting/>>).

<sup>50</sup> A inteligência artificial explicável (*XAI*) é um conjunto de processos e métodos que permite aos usuários humanos compreender e confiar nos resultados e saídas criados por algoritmos de aprendizado de máquina. A IA explicável é usada para descrever um modelo de IA, seu impacto esperado e potenciais vieses. Ele ajuda a caracterizar a precisão do modelo, a justiça, a transparência e os resultados na tomada de decisões baseadas em IA. A IA explicável é crucial para uma organização na construção de confiança ao colocar modelos de IA em produção. A explicabilidade da IA também ajuda uma organização a adotar uma abordagem responsável para o desenvolvimento de IA (O que é IA explicável (*XAI*)?. IBM. 2023. p.1. Disponível em:<<https://www.ibm.com/watson/explainable-ai>>).

caixas-pretas, ou seja, são capazes de fornecer respostas precisas, mas sem explicar como chegaram a essas respostas.

Com efeito, temos que a explicabilidade é crucial em muitos contextos, especialmente em áreas como o direito, medicina e finanças, onde as decisões tomadas pelos modelos de IA podem ter impactos significativos na vida das pessoas. Inclusive, a capacidade de entender e explicar como um modelo de IA chegou a uma determinada decisão é fundamental para a transparência, responsabilidade e confiança na aplicação desses modelos.

Nesse sentido, a explicabilidade da IA tem várias vantagens e benefícios. Em primeiro lugar, proporciona uma compreensão mais profunda dos processos internos dos modelos de IA permitindo a detecção e correção de possíveis vieses, erros ou decisões injustas. Isso é especialmente importante quando se trata de aplicações críticas, como o sistema judiciário, onde a equidade e imparcialidade são valores fundamentais.

Além disso, a *XAI* permite que os usuários compreendam as previsões e recomendações feitas pelos modelos de IA tornando-as mais confiáveis e aceitáveis. Isso é essencial para promover a adoção e aceitação dessas tecnologias, tanto por profissionais quanto pelo público em geral. A capacidade de fornecer explicações claras e compreensíveis também permite que os usuários detectem possíveis erros ou problemas nos modelos, melhorando sua confiabilidade e segurança<sup>51</sup>.

Outro benefício da *XAI* é a capacidade de verificar a conformidade dos modelos de IA com regulamentações e políticas específicas. Em setores como saúde e finanças, onde existem diretrizes e leis que devem ser seguidas, a explicabilidade dos modelos de IA auxilia na identificação de possíveis violações e na garantia de que as decisões estejam em conformidade com as normas estabelecidas.

Em suma, a capacidade de entender e explicar os processos de tomada de decisão dos modelos de IA permite a detecção de vieses, melhoria da confiabilidade, garantia de

---

<sup>51</sup> [...] À medida que a IA se torna mais avançada, os humanos são desafiados a compreender e retrair como o algoritmo chegou a um resultado. Todo o processo de cálculo é transformado no que é comumente chamado de "caixa preta" que é impossível de interpretar. Esses modelos de caixa preta são criados diretamente dos dados. E, nem mesmo os engenheiros ou cientistas de dados que criam o algoritmo conseguem entender ou explicar o que exatamente está acontecendo dentro deles ou como o algoritmo de IA chegou a um resultado específico (O que é IA explicável (*XAI*)?. IBM. 2023. p.1. Disponível em:<<https://www.ibm.com/watson/explainable-ai>>).



conformidade com regulamentações e, em última análise, promove a aceitação e confiança nas aplicações de IA<sup>52</sup>. A busca por técnicas e abordagens que promovam a explicabilidade dos modelos de IA é uma área de pesquisa em crescimento, visando tornar a IA mais acessível, justa e confiável para a sociedade como um todo.

Contudo, em atenção ao mundo dos fatos, resta claro que a falta de explicabilidade em modelos de IA apresenta diversos desafios e implicações que podem limitar a confiança, aceitação e adoção desses modelos. No tocante à falta de transparência, temos que, atualmente, modelos de IA complexos, como redes neurais profundas, muitas vezes tem seus processos de tomada de decisão como difíceis de serem compreendidos e explicados. Isso resulta em falta de transparência, impedindo que os usuários entendam como o modelo chegou a uma determinada conclusão ou previsão.

Destarte, a falta de explicabilidade pode levar a decisões injustas ou enviesadas. Os modelos de IA podem aprender a partir de conjuntos de dados com vieses embutidos, o que pode levar a discriminação em relação a certos grupos ou indivíduos. Sem a capacidade de entender como essas decisões são tomadas, é difícil identificar e corrigir esses vieses, resultando em práticas discriminatórias e injustas<sup>53</sup>.

Igualmente, a explicabilidade é um componente essencial para a confiança e aceitação dos modelos de IA. Os usuários podem relutar em confiar em um modelo que não consegue fornecer uma justificativa ou explicação para suas decisões. Não raro, isso pode levar à resistência à adoção dessas tecnologias e à preferência por métodos mais tradicionais.

Por fim, a falta de explicabilidade dificulta a atribuição de responsabilidade quando ocorrem erros ou resultados indesejados. Sem compreender como um modelo chegou a uma determinada decisão, é difícil determinar quem é responsável por suas ações. E, válido ressaltar que, na falta de explicabilidade também é difícil garantir que os modelos de IA estejam em conformidade com as regulamentações acerca do uso e

---

<sup>52</sup> [...] A IA explicável também ajuda a estimular a confiança do usuário final, a auditabilidade do modelo e o uso produtivo da IA. Ela também minimiza os riscos de conformidade, jurídicos, de segurança e de reputação da IA de produção (O que é IA explicável (XAI)?). IBM. 2023. p.2. Disponível em:<<https://www.ibm.com/watson/explainable-ai>>).

<sup>53</sup> As redes neurais usadas no deep learning são algumas das mais difíceis para um humano entender. Viés, muitas vezes com base em raça, gênero, idade ou localização, tem sido um risco de longa data ao treinar modelos de IA. Além disso, o desempenho do modelo de IA pode variar ou degradar porque os dados de produção diferem daqueles de treinamento (O que é IA explicável (XAI)?). IBM. 2023. p.2. Disponível em:<<https://www.ibm.com/watson/explainable-ai>>).

proteção de dados.

Para enfrentar esses desafios, tímidos esforços veem se desenhando no horizonte da Inteligência Artificial explicável (*XAI*) que, por sua vez, tem se concentrado no desenvolvimento de técnicas e abordagens que visam fornecer explicações compreensíveis e confiáveis para os modelos de IA. Essas técnicas incluem desde a utilização de modelos mais interpretáveis, como árvores de decisão, até abordagens mais avançadas, como a geração de explicações baseadas em evidências ou a interpretação de pesos e relevâncias dos atributos.

A busca por explicabilidade em modelos de IA é essencial para garantir a transparência, a justiça e a confiabilidade desses sistemas. A superação dos desafios ante a falta de explicabilidade é fundamental para o avanço e a aplicação ética da IA em diversas áreas, permitindo que os usuários compreendam as razões por trás das decisões tomadas pelos modelos de IA.

De fato, a pesquisa em *XAI* tem avançado no desenvolvimento de métodos e técnicas para fornecer explicações compreensíveis. Como veremos mais adiante, inclui abordagens como a geração de regras de decisão, a interpretação de pesos de atributos, a visualização de processos de aprendizado e muito mais<sup>54</sup>. Essas técnicas visam capacitar os usuários a compreender como os modelos de IA chegam a suas conclusões, facilitando assim a confiança, a aceitação e o uso ético desses sistemas.

Na área de Inteligência Artificial explicável (*XAI*), existem diversas abordagens e técnicas que visam fornecer explicações compreensíveis sobre as decisões e predições feitas por modelos de IA. Essas abordagens têm como objetivo principal aumentar a transparência e a interpretabilidade dos sistemas de IA permitindo que os usuários entendam como as decisões foram tomadas e confiem nos resultados obtidos.

Uma das abordagens mais comuns em *XAI* é a utilização de regras<sup>55</sup>. Nesse caso,

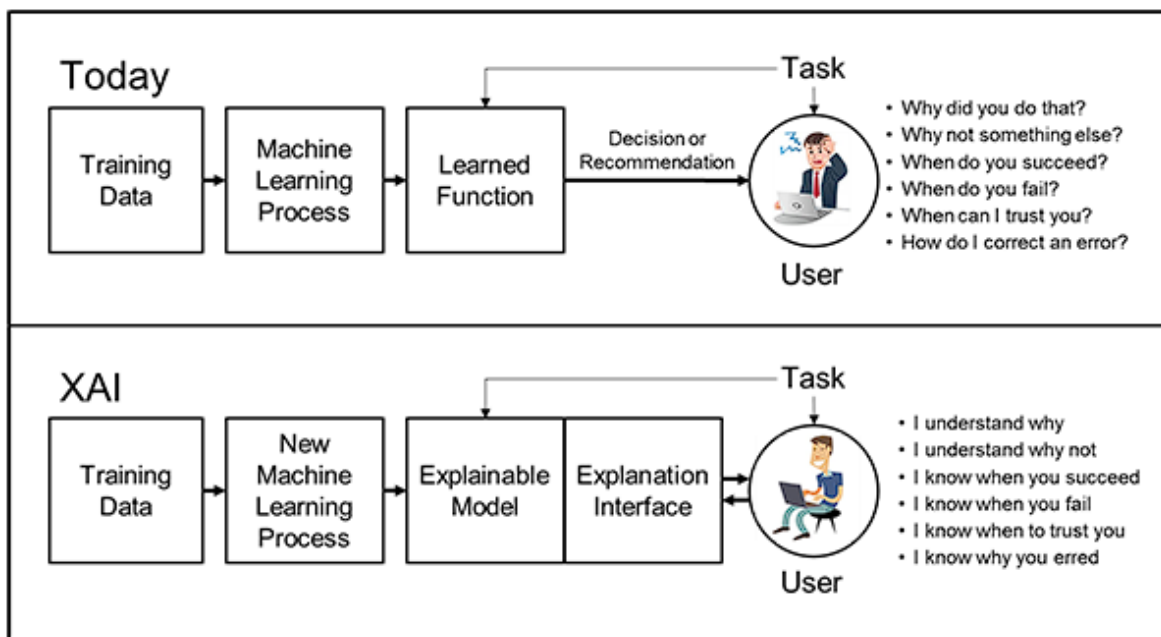
---

<sup>54</sup> [...] Por exemplo, se treinar um algoritmo de ML utilizando um conjunto de dados financeiros para determinar se deve aprovar um pedido de empréstimo, a *XAI* pode entregar a resposta e ainda, informar como e porque obteve tal conclusão. Esse processo é diferente dos modelos caixa-preta, mais comuns no universo do ML, onde alguns casos são mais interpretáveis do que outros. Podemos citar a árvore de decisão que gera resultados facilmente traduzíveis, enquanto redes neurais são mais opacas. Por isso, muitas vezes há um desequilíbrio entre acurácia e explicabilidade (Guia essencial para IA explicável (*XAI*). Alteryx. 2023. p.1. Disponível em: <<https://www.alteryx.com/pt-br/resources/whitepaper/essential-guide-to-explainable-ai>>).

<sup>55</sup> [...] Técnicas como a LIME e a *SHAP* oferecem respostas matemáticas muito literais para essa questão, e os resultados dessa matemática podem ser apresentados a cientistas de dados, gerentes, reguladores e

o modelo de IA é capaz de gerar regras explícitas que descrevem as decisões tomadas. Essas regras podem ser facilmente compreendidas pelos usuários e ajudam a elucidar o processo de raciocínio do modelo. Além disso, a interpretabilidade de modelos é outra técnica amplamente utilizada. Ela envolve o uso de métodos que permitem analisar os pesos e a importância dos atributos utilizados pelo modelo, de forma a entender quais características influenciam mais nas previsões.

**Figura 5 Ilustração do treinamento de IA explicável (XAI)**



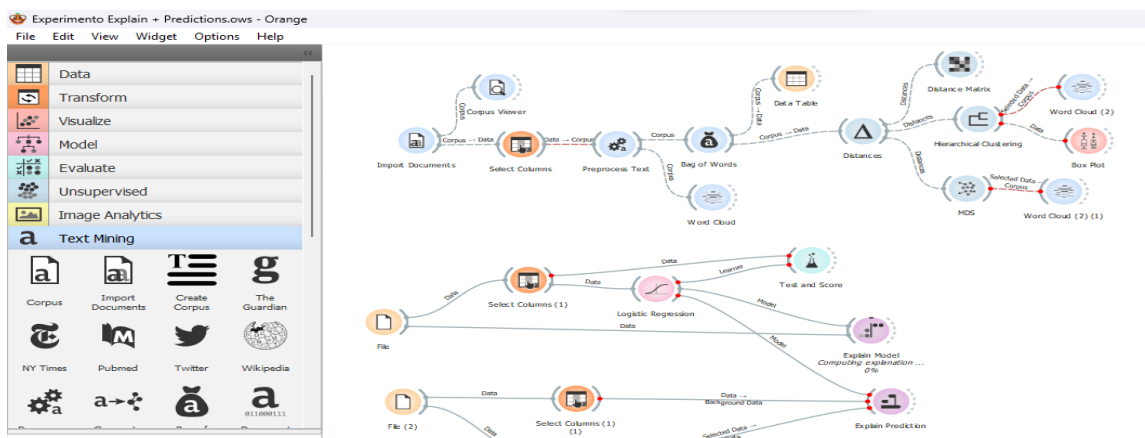
Fonte: ASHLEY. John. O que é IA explicável. Nvidia. 2021. Disponível em:<<https://blog.nvidia.com.br/2021/08/02/o-que-e-ai-explicavel/>>.

De igual modo, a visualização de dados também desempenha um papel importante na XAI. Por meio de gráficos, diagramas ou outras representações visuais, é possível mostrar de forma intuitiva como os dados estão sendo processados e como as decisões são tomadas. Essas visualizações auxiliam na compreensão dos padrões identificados pelo modelo e permitem que os usuários acompanhem o raciocínio do sistema.

---

consumidores. Para alguns dados, como imagens, áudios e textos, resultados semelhantes podem ser visualizados com o uso da “atenção” nos modelos, forçando o próprio modelo a mostrar seu processo (ASHLEY. John. O que é IA explicável. Nvidia. 2021. Disponível em:<<https://blog.nvidia.com.br/2021/08/02/o-que-e-ai-explicavel/>>).

Figura 6 Ilustração do Orange Canvas



Fonte: Orange Data Mining. 2023. Disponível em: < <https://orangedatamining.com/download/#windows> >.

Além dessas abordagens, existem outras técnicas em desenvolvimento na área de *XAI*, como o uso de modelos de interpretação, que são modelos mais simples e compreensíveis que buscam representar o comportamento dos modelos de IA complexos. Inclusive, a *XAI* também pode ser aplicada na automatização de tarefas rotineiras e repetitivas no campo do Direito, como a revisão de contratos, a análise de documentos legais e a predição de resultados de casos judiciais. Os modelos de IA podem aprender a partir de exemplos históricos e, com base nesses aprendizados, realizar tarefas de forma rápida e precisa. Isso libera os profissionais do Direito para se concentrarem em atividades mais complexas e estratégicas.

A *XAI* pode ser útil também na área de *compliance*<sup>56</sup> e na detecção de fraudes. Os modelos de IA podem ser treinados para identificar padrões suspeitos em transações financeiras, contratos ou outras atividades empresariais, auxiliando na prevenção e investigação de práticas ilegais. A explicabilidade desses modelos permite que as

<sup>56</sup> A princípio, ao realizar uma tradução literal do termo *compliance*, podemos entender que se trata de estar em conformidade, todavia, se realmente significasse isso, não existiria nenhuma novidade. Isso porque, em um Estado Democrático de Direito, todos os cidadãos têm o dever de cumprir as normas. Desse modo, *compliance* não é estar em conformidade com as leis somente. A ideia de *compliance* diz respeito também a uma questão de organização de pessoas.

[...] *Compliance* é um conjunto normativo que assegura o cumprimento de regras de determinado setor. O significado tem como objetivo identificar possibilidades de infrações, falta de cumprimento de uma norma legal ou atividades que se configurem como atos de corrupção. (ANTONIK, Luis Roberto. *Compliance, Ética, Responsabilidade Social e Empresarial-Uma visão prática*. Alta Books Editora, 2016. P.50).

autoridades e as empresas compreendam os critérios e as evidências que levaram a uma determinada conclusão, fortalecendo o processo de investigação e a responsabilização.

Frisa-se que a *XAI* no campo do Direito não busca substituir a atuação humana, mas sim complementá-la e fornece suporte para a tomada de decisões informadas. Ainda no âmbito das decisões judiciais, a *XAI* pode ser aplicada para explicar como determinadas conclusões foram alcançadas. Isso é especialmente relevante em casos complexos, nos quais os modelos de IA podem identificar padrões e relações difíceis de serem percebidos pelos seres humanos. A explicabilidade dos modelos permite que as partes envolvidas entendam as razões por trás das decisões, aumentando a confiança nas resoluções judiciais.

A *XAI* também pode desempenhar um papel importante no fortalecimento da *accountability*<sup>57</sup> no sistema jurídico. Ao fornecer explicações compreensíveis sobre as decisões automatizadas, os modelos de IA tornam-se mais passíveis de escrutínio e avaliação crítica. Isso contribui para a identificação de eventuais vieses ou falhas nos algoritmos, bem como para a correção e aprimoramento contínuo dos sistemas de IA utilizados no Direito.

Conclui-se, portanto, que a *XAI* pode trazer benefícios significativos para o campo do Direito, proporcionando transparência, eficiência e automação de tarefas. A explicabilidade dos modelos de IA, não representa a solução de todos os problemas, porém permite uma compreensão clara e justificável das decisões tomadas, promovendo a confiança e a equidade no sistema de justiça. Com a *XAI*, o Direito pode se beneficiar da aplicação dessa abordagem para promover um sistema jurídico mais justo, transparente e eficiente.

A seguir, veremos um ponto importante no cenário da segurança da informação e da governança de dados que é o da proteção de dados pessoais, tanto na LGPD do Brasil,

---

<sup>57</sup> A *accountability* pode ser entendida como a “responsabilização permanente dos gestores públicos em razão dos atos praticados em decorrência do uso do poder que lhes é outorgado pela sociedade”, desde que se estabeleça que a responsabilização se dê mediante algum tipo de sanção. Portanto, a *accountability* se realizará “quando qualquer ação, independente da sua origem, venha a representar alguma forma de sanção, seja ela legal ou moral, e que se reflita em constrangimento ou embaraço efetivo à atividade do agente público.” (ROCHA, Arlindo Carvalho. O processo orçamentário brasileiro como instrumento de *accountability*. Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração, 2008. p. 3-4).

quanto na GDPR<sup>58</sup> da Europa, e demais proteções a esse que é um superativo de todos os indivíduos e das empresas.

### 2.2.2 *Proteção De Dados Pessoais*

No contexto brasileiro, em especial, estamos enfrentando diversas crises relacionadas ao aspecto organizacional, não somente de segurança de informação. Mais recentemente, tivemos casos relacionados a própria questão da pandemia de Covid-19, e um pouco antes questões ambientais envolvendo deslizamentos, como o caso de Mariana, e outros contextos que envolvem a questão de crise e, obviamente envolvem o aspecto de informação.

Outrossim, ainda existem pessoas que acreditam que segurança da informação é assunto exclusivo da área de informática, o que não procede, pois, a segurança da informação é um assunto muito importante em nossas vidas. Desde o contexto doméstico até o contexto empresarial, como o de profissionais dentro de uma organização. Além disso, os sistemas tradicionais de segurança já não conseguem atender aos crescentes demandas de invasões digitais e as brechas acabam ocorrendo dentro da rede de computadores, se tornando uma ciência de dados.

A sua existência precede o século XX, tanto que a primeira criptografia que se tem registro na história foi a Cifra de César<sup>59</sup>, que teve sua importância para que a comunicação de Roma alcançasse as extremidades de seu Império e os Generais Romanos, sem gerar boato ao longo do caminho. No entanto, a criptografia é somente uma das muitas áreas da segurança da informação.

Atualmente, a criptografia é usada diariamente através das comunicações por *Whatsapp* de ponta a ponta, ou também em compras online, onde a comunicação entre o servidor do comerciante e nossas máquinas estão criptografadas de ponta a ponta. Ocorre que existem outros pontos não tecnológicos que são mais perigosos, ou não é dada a devida importância. E nesse elo da cadeia de comunicação, o elo mais fraco é o ser

---

<sup>58</sup> GDPR: O regulamento da (UE) 2016/679, também conhecido como Regulamento Geral sobre a Proteção de Dados, é aplicável a todos os indivíduos na União Europeia e no Espaço Econômico Europeu. Diz respeito à privacidade e à proteção de dados pessoais, bem como à livre circulação desses dados.

<sup>59</sup> Cifra de César é o código criptográfico mais antigo do mundo, trata-se de um tipo de cifra de substituição na qual cada letra do texto é substituída por outra, localizada no alfabeto abaixo dela um número fixo de vezes.

humano e esse talvez seja um dos maiores problemas, tanto por questões culturais, quanto pela própria essência do ser humano.

De acordo com a *National Cyber Security Alliance*<sup>60</sup>, a cibersegurança começa em casa, isso porque a principal técnica de ataque social é, na verdade, a engenharia social, isto é o uso de informações para ludibriar ou tentar obter algo de alguém, especialmente no momento de necessidade. E isso acaba entrando em um contexto que não é mais presencial.

No que diz respeito a engenharia social aplicada aos adultos, é recorrente o recebimento de *e-mails* contendo informações no estilo “seu CPF foi cancelado”, ou “você foi premiado” e, tomados por uma emoção, as pessoas clicam em um link para tentar resolver a questão e acabam instalando um *software* indevido, e esse software indevido acaba instalando um vírus, um *malware*<sup>61</sup>, ou *backdoor*<sup>62</sup>. E, com isso, buscando extrair informações pessoais, por vezes financeiras, ou monitorar o computador para ter acesso a ele, ou também sequestrar o processamento do computador e utilizar sua memória para minerar moedas virtuais<sup>63</sup>.

Já no que diz respeito ao campo das organizações públicas e privadas, ao final de 2020, a *Trustwave*<sup>64</sup> divulgou um estudo sobre tendências em cibersegurança<sup>65</sup>, que diz que a principal forma de ataque é a engenharia social, cuja forma de proteção é o treinamento. Informa que 100% dos aplicativos da *web* possuem ao menos uma vulnerabilidade. Ao passo que a extorsão sexual cresce a ritmo galopante na *internet*,

---

<sup>60</sup> *National Cyber Security Alliance* foi fundada em 2001, é a principal parceria público-privada sem fins lucrativos dos Estados Unidos, e promove a educação e a conscientização sobre segurança cibernética e privacidade.

<sup>61</sup> *Malware* é a abreviação de “*malicious software*”. Trata-se de um programa de computador destinado a infiltrar-se em um sistema, causando danos, alterações ou roubo de informações.

<sup>62</sup> *Backdoor* é um método, geralmente secreto, para escapar da autenticação ou criptografia em um sistema computacional. Os *backdoors* costumam ser usados para proteger o acesso remoto a um computador e ter acesso a textos simples em sistemas criptográficos.

<sup>63</sup> Cryptojacking Malware Skyrocks: Um aumento acentuado ano a ano de 1.250% foi observado no malware de cryptojacking. Usado para colocar secretamente mineradores de moedas JavaScript legítimos em sites ou infectar roteadores de nível de operadora, o malware de cryptojacking extrai ilegalmente criptomoedas para cibercriminosos usando os recursos de computação de vítimas desavisadas. 2019 Trustwave Global Security Report, AT&T Cybersecurity / Insights Report V8.5. Telefonica Global Security Report. Disponível em: <http://www.portaldaprivacidade.com.br/wp-content/uploads/2019/05/2019-trustwave-global-security-report.pdf>

<sup>64</sup> Trustwave: Criada em 1995, é uma empresa americana e marca de cibersegurança da companhia de telecomunicações cingapuriana Singtel Group Enterprise. A Trustwave Holdings é dedicada à segurança de rede e de internet de empresas

<sup>65</sup> 2019 Trustwave Global Security Report, AT&T Cybersecurity / Insights Report V8.5. Telefonica Global Security Report. Disponível em: <http://www.portaldaprivacidade.com.br/wp-content/uploads/2019/05/2019-trustwave-global-security-report.pdf>

como ocorreu no caso da empresa *Ashley Madison*, em que hackers tiveram acesso a informações pessoais da empresa e as utilizaram para extorquir os clientes:

**Figura 7 Imagem de falha na segurança da informação em Ashley Madison**



Fonte: ZETTER, Kim. *Hackers Finally Post Stolen Ashley Madison Data*. Wired. 2018. Disponível em: <https://www.wired.com/2015/08/happened-hackers-posted-stolen-ashley-madison-data/>

O estudo feito pela *Trustwave*, também informa que as organizações são mais vulneráveis do que pensam, no sentido de que falta treinamento e sensibilização, principalmente com o fator humano; bem como que não estão orientadas para avaliar o risco plenamente. Por fim, também ressalta que o Adobe<sup>66</sup> é o vetor importante de malware, principalmente com arquivos PDFs.

Em Relatório da *McAfee*<sup>67</sup>, o cibercrime custa US\$ 1 trilhão por ano, enquanto que os danos globais de *ransomware* giram em torno de US\$ 20 bilhões, pois existe todo um retrabalho para construir o que foi perdido. E, por outro lado, o custo de um kit de ferramentas para cibercrime na *DarkWeb* é US\$ 1.

Se tem a oferta do ataque, é porque tem muitos indivíduos e organizações que caem no ataque. Sendo assim, é fundamental ter uma visão analítica da segurança da informação, isto é, o que ela possui de ativos, de vulnerabilidades e de ameaças. É preciso segurança e governança de TI sabendo que se a organização vive de dados, toma decisão com base em dados, coleta dados, então a segurança de dados é um assunto de nível

<sup>66</sup> ADOBE INC. Fundada em dezembro de 1982 por Charles Geschke e John Warnock, é uma multinacional desenvolvedora de softwares como o Photoshop, o Premiere e o Acrobat Reader. Em 2018, a companhia obteve um faturamento de mais de nove bilhões de dólares.

<sup>67</sup> MCAFEE: Fundada em 1986 por John McAfee, é uma empresa americana de informática e desenvolvedora de softwares de segurança. Com mais de cinco bilhões de dólares em ativos, a organização oferece diversos serviços de segurança.



estratégico. Outrossim, é fundamentação um plano de contingência para se preparar para desastres e também um plano de continuidade, caso perca sua estrutura física, por exemplo, ou um servidor digital.

A mudança sem precedentes para o trabalho remoto também deixou as organizações globais despreparadas para monitorar ou identificar ameaças internas devido ao acesso remoto não autorizado, senhas fracas, redes não protegidas e o uso indevido de dispositivos pessoais, que constituem ameaças internas. Ocorre que, de acordo com a *Forrester*, brechas de ameaças internas eram estimadas a crescer 8% em 2021 e responder por 33% de todos os incidentes de cibersegurança.<sup>68</sup> Acresce-se a isso o fato de muitos sistemas acabarem indo para a *Cloud*, onde existe diversos riscos, tais como: Armazenamento em *cloud* mal configurado, controle e visibilidade reduzidos, deleção de dados incompleta, *apps* em cloud vulneráveis.

Conforme a ISO 27.001<sup>69</sup>, a segurança da informação é um planejamento contínuo e para tanto necessita de um desenho e gestão do programa de segurança (*Security Policy Framework & Enterprise Security Architecture*); gestão do risco, identificando, medindo, reportando, e escolhendo como controlar os riscos (*Security Risk Assessment, Security Health Check, Security Workshop, Information Security Framework*); e, por fim, não menos importante, serviços de privacidade e conformidade com processos de aderência às decisões e políticas (*Privacy workshop, Privacy Controls Assessment, Privacy Strategy & Implementation, PCI expertise*).

A partir disso, é possível se verificar a dimensão e complexidade da governança de dados e segurança da informação. Por tudo isso, a *ISO 27001*, prevê como boas práticas da segurança da informação: 1- Política de Segurança da informação, 2- Segurança organizacional, 3- Classificação e controle dos ativos de informação, 4- Segurança em pessoas, 5- Segurança Física e Ambiental (controle de incêndio, controle da parte elétrica, controle contra inundação), 6- Gerenciamento das operações e comunicações, 7- Controle de acesso (físico e digital), 8- Desenvolvimento de Sistemas e Manutenção (tem de estar

---

<sup>68</sup> Forrester Paper: Predictions 2021: Cybersecurity. Disponível em: <<https://www.code42.com/resources/reports/forrester-paper-predictions-2021-cybersecurity>>.

<sup>69</sup> ISO 27001: É um padrão para sistemas de gestão da segurança da informação publicado em outubro de 2005 pelo International Organization for Standardization pelo International Electrotechnical Commission. A norma foi elaborada para estabelecer, implementar, operar, monitorar, analisar criticamente, manter e melhorar um Sistema de Gestão de Segurança da Informação (SGSI).

de acordo com a política de segurança da empresa), 9- Gestão da continuidade do negócio , 10- Conformidade (Questão de adequação a legislação e boas práticas).

Antes da LGPD houve um vácuo legislativo no Brasil no que concerne ao detalhamento da proteção aos dados pessoais. No Código Civil, tínhamos algumas regras com relação ao uso de dados, mas que estavam dentro de um guarda-chuva da privacidade, isto é, não se permite violar a intimidade a privacidade das pessoas, mas não se detalhava como é que isso na prática se colocava ao uso de dados. No Código de Defesa do Consumidor, também existe algumas regras do consumidor, regras de inserção no Serasa, no SPC, de banco de dados, de consumo, mas só isso não avançava para outros usos dos nossos dados. Outrossim, o Marco Civil da Internet prevê tão somente a proteção de dados coletados por provedores de aplicação ou de conexão. Ocorre que faltava uma norma que trouxesse um detalhamento das expectativas e das boas práticas nesse campo.

Na União Europeia essas normas já vão se sucedendo desde a década de 80, onde lá na Alemanha, por exemplo, já se tinha por ocasião de 1995 uma previsão à proteção de dados de seus cidadãos. No entanto, em 2016, foi aprovado o GDPR, ou regulamento europeu de proteção de dados pessoais na norma, talvez a mais dura lei ao nível global nessa área, muito rigorosa, que colocou os requisitos em um outro patamar no nível de detalhamento nunca antes visto, para que se pudesse usar e tratar dados pessoais.

Essa norma além de impactar a União Europeia, ela teve uma repercussão global e por uma razão muito simples, porque essa lei e regulamento, trouxe junto uma regra que se o tratamento dos dados está em um outro território em outro país, em outro continente, e quer tratar os cidadãos europeus, para fazer isso, é necessário ter condições de mostrar que o país destino tem um nível de proteção equivalente ao que foi estabelecido para União Europeia. E isso fez com que legislações, mundo a fora, começassem a ser revisadas, para que pudessem alcançar esse reconhecimento de adequação, para que não houvesse uma perda comercial nas relações com a União Europeia. E assim, surgiu então a LGPD, para permitir ter acesso a esse fluxo de dados, para permitir que o Brasil tenha operações que tratem de dados dos cidadãos europeus também e naturalmente para proteger os dados pessoais dos cidadãos brasileiros, nosso direito fundamental.

A LGPD é uma lei que se preocupa com dados pessoais. Isso significa que talvez essa lei não devesse ser chamada de LGPD, mas “LGPDP”, porque ela não é uma lei que proteja dados em geral, ela é uma lei que protege dados pessoais, ou seja, aquelas

informações que consiga associar a um indivíduo em especial, e não, por exemplo, um dado que seja de mercado e que não se conecte a uma pessoa, tal qual, a quantidade de pessoas vacinadas em São Paulo ou Rio de Janeiro. Esses últimos são dados que não importam a LGPD, porque não se referem a um indivíduo em especial, mas sim a uma coletividade não identificável nos seus componentes.

Da mesma forma, dados relacionados a uma empresa ou organização qualquer, como o faturamento e os resultados num determinado período. Eles não são relevantes para fins de LGPD, porque na prática elas não se referem a informações que permitem associação a um indivíduo em especial. Logo seu uso segue sendo livre tal, qual era até pouco tempo o uso de dados pessoais, no sentido de não ter uma norma com tanto detalhamento das práticas permitidas e das práticas não são mais aceitas.

A LGPD também exige que seja pública a informação de quais os tipos de dados que serão coletados e sobre a forma como esses dados serão utilizados e os procedimentos para o exercício dos direitos dos titulares, isto é, até mesmo a forma como o responsável pela criança pode exercer esses direitos. E mais ainda, tem algumas vedações, uma delas, por exemplo, que aquele que captura o dado (controlador) da criança e do adolescente, não pode condicionar a participação em jogos e aplicações de internet e demais atividades; ao fornecimento de seus dados pessoais, salvo, se for estritamente necessário para a dinâmica e funcionamento daquela ferramenta, mas outros dados que não sejam necessários, eles não podem ser requisitos para o acesso a um jogo ou alguma ferramenta de entretenimento, isto é, que estimule a criança a se expor além do que for estritamente necessário da expectativa dos seus próprios pais.

Então a primeira capacidade importante para caminharmos em busca da conformidade, é justamente a capacidade de identificar os processos nos quais vai precisar atuar. E nesses processos, o que a LGPD vem trazer de regras e requisitos é para o que a chamamos de tratamento de dados pessoais, ou seja, ela vem ditar como deve-se dar o tratamento de dados daqui para frente.

A Lei diz que tratar dados é toda e qualquer operação realizada com dados pessoais, desde<sup>70</sup>: 1- Coleta, 2- Produção, 3- Recepção, 4- Classificação, 5- Utilização, 6- Acesso, 7- Reprodução, 8- Transmissão, 9- Distribuição, 10- Processamento, 11-

---

<sup>70</sup> Art. 5º da LGPD. Disponível em:< [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

Arquivamento, 12- Armazenamento, 13- Eliminação, 14- Avaliação, 15- Controle, 16- Modificação, 17- Comunicação, 18- Transferência, 19- Difusão, 20- Extração.

Ou seja, praticamente qualquer coisa que seja feita com dado pessoal. Qualquer um desses processos, para ele poder estar adequado, para ele poder ser realizado, é preciso de um enquadramento de adequação, de uma conformidade com os requisitos desta lei.

Por outro lado, o tratamento lícito, ou seja, o tratamento correto é aquele que atende ao que a lei mesmo estabelece, o que a LGPD mesmo estabelece, principalmente aos seus princípios que são seus valores mais importantes<sup>71</sup>: 1-Boa-fé, 2- Finalidade, 3- Adequação, 4- Necessidade, 5- Livre Acesso, 6- Qualidade dos dados, 7- Transparência, 8- Segurança, 9- Prevenção, 10- Não-discriminação, 11- Responsabilização e Prestação de Contas.

Em outras palavras, teremos o tratamento lícito de dados se atender todos os princípios legais e se enquadra em ao menos uma das hipóteses legais. Já o tratamento irregular de dados é aquele que não observa a legislação (viés jurídico), ou que não fornece a segurança que o titular dos dados pode esperar (viés técnico).

Então, temos uma lista pré-definida das hipóteses que permitem o tratamento de dados pessoais, que não são aquelas dos dados pessoais sensíveis<sup>72</sup>: 1- Obrigação legal ou regulatória; 2- Execução de contrato; 3- Proteção do Crédito; 4- Interesses legítimos do controlador ou de terceiro; 5- Exercício regular de direitos em processo judicial, administrativo ou arbitral; 6- Realização de estudos por órgão de pesquisa; 7- Pela Administração pública, para execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres; 8- Proteção da vida ou da incolumidade física do titular ou de terceiros; 9- Tutela da saúde, em procedimento realizado por profissionais da área da saúde ou por entidades sanitárias; 10- Mediante o fornecimento de consentimento pelo titular.

O consentimento tem que ser por escrito e dar para finalidades específicas, ele tem que ser uma manifestação livre, informada (saber para que será usado), inequívoca (sem

---

<sup>71</sup> Art. 6º da LGPD. Disponível em:< [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

<sup>72</sup> Art. 7º da LGPD. Disponível em:< [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

dúvidas), pela qual o titular concorda com o tratamento de seus dados pessoais. Além disso, pode ser revogado a qualquer momento.

Na prática o consentimento é a última opção, pois se torna bastante complexo lidar com consentimentos, e é bastante custoso. Imagine uma organização que tem dados de dezenas de milhões de cidadãos. O consentimento exigiria dela saber quem concordou com o que e quando para cada uso de dados pessoais. Ou seja, a cada novo uso daquele dado, seria necessário um novo consentimento e se a pessoa se arrependeu e revogou, a empresa deve eliminar esse dado pessoal.

A organização precisa fazer então esse confronto entre o processo que ela executa que usa dados pessoais e essas hipóteses em que a lei autoriza o tratamento de dados. Agora sim, se estivermos diante de um dado pessoal sensível a regra vai ser que, para tratar o dado, será necessário o consentimento, com exceção de sete cenários, a dizer<sup>73</sup>: 1- Obrigação Legal ou regulatória; 2- Exercício regular de direitos em contrato, processo judicial, administrativo ou arbitral; 3- Realização de estudos por órgão de pesquisa; 4- Pela administração pública, para execução de políticas públicas previstas em lei, regulamentos, ou respaldadas em contratos, convênios ou instrumentos congêneres; 5- Proteção da vida ou da incolumidade física do titular ou de terceiros; 6- Tutela da saúde, em procedimento realizado por profissionais da área da saúde ou por entidades sanitárias; 7- Prevenção à fraude e à segurança do titular, nos processos de identificação e a cadastro em sistemas eletrônicos (por exemplo, a biometria).

Outrossim, durante a vida do tratamento de dados, é preciso verificar se os dados coletados sejam realmente necessários para essa finalidade. E aqui está uma segunda abordagem de um projeto de conformidade que é identificar se não há dados excessivos e se não há um desvio de finalidade.

A partir do momento que o tratamento de dados é descontinuado, acabou a finalidade do tratamento, e, portanto, aqueles dados devem ser excluídos ou eliminados. Por fim, ainda para caracterizar a LGPD, temos de apresentar os seus personagens, a dizer: 1- Titular, 2- Agente de Tratamento, 3- Autoridade Nacional de Proteção de Dados.

Não é como uma propriedade tradicional, mas o titular é o dono do dado, é a pessoa natural a quem se referem os dados pessoais que são objeto de tratamento. Ele

---

<sup>73</sup> Art. 11º da LGPD. Disponível em: < [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

também possui alguns direitos de acordo com a LGPD como<sup>74</sup>: 1- Titularidade do dado, 2- Informações completas sobre o tratamento (Se vai para o exterior ou não vai, por quanto tempo será usado), 3- Cópia dos dados, 4- Correção, 5- Exclusão de dados, 6- Portabilidade, 7- Revisar decisões automatizadas.

Até então, acreditava-se que as empresas e organizações eram os verdadeiros donos desses dados que tinham como um dos seus principais ativos numa fusão, ou aquisição de negócios, esses dados. Ocorre que, a LGPD deixa claro que, na verdade, o dono do dado não é a empresa, mas sempre a pessoa a quem o dado se refere. Quanto ao procedimento para o exercício de direitos. A pessoa pode<sup>75</sup>: 1- Realizar requerimento expresso ao agente de tratamento, que deve, então, fornecer em 15 dias; 2- Peticionar contra o controlador perante a ANPD, sujeitando a fiscalização e multas; 3- Mover ação judicial.

O segundo personagem é a empresa/pessoa que trata o dado, aquele que coletou os dados. O agente de tratamento é uma pessoa natural ou jurídica, de direito público ou privado a quem competem as decisões referentes ao tratamento de dados pessoais. Esse agente de tratamento a pessoa física ou jurídica, ele pode ter dois perfis, ele pode ser o que a lei chama de controlador do dado ou o operador do dado.

O controlador é aquele que toma decisões com relação ao tratamento de dados é aquele que define, por exemplo, em um formulário, quais serão os campos daquele formulário. Se vai ser perguntado gênero, renda ou não. É quem define como os dados serão protegidos, aonde vão ser armazenados e por quanto tempo vão ser guardados. Então naturalmente o controlador é quem tem maiores responsabilidades nesse processo.

O operador do dado é alguém contratado pelo controlador, normalmente para fins de executar alguma tarefa relacionada ao tratamento de dados. Imagine que seja contratado um serviço de armazenamento de dados na nuvem da *Dropbox*<sup>76</sup>. A *Dropbox*, é o prestador de serviço, isto é o operador. E esse operador, não tem controle de que tipo de dados serão inseridos, se são dados sensíveis, ou não são dados sensíveis. Por um contrato ele se obriga a guardar aqueles dados sob alguma forma de cuidado. Isso que vai

---

<sup>74</sup> Art. 19º da LGPD. Disponível em: < [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

<sup>75</sup> Ibidem.

<sup>76</sup> DROPBOX: Criado em 2007 por Drew Houston e Arash Ferdowsi, é um serviço de armazenamento e compartilhamento de arquivos em nuvem. Utilizado por mais de 400 milhões de pessoas no mundo, a Dropbox Inc, foi avaliada, no ano de 2014, em US\$ 10 bilhões.

delimitar a responsabilidade dele. Ele não pode descumprir aquilo que se obrigou com o controlador, mas ele não vai responder por decisões que são do próprio controlador com relação a quais dados estão sendo tratados e como se dá esse tratamento.

Do mesmo modo, no que concerne ao agente de tratamento, ele também possui uma série de deveres. A maioria diz respeito a adotar medidas de segurança, técnicas e administrativas, bem como fiscalizar a observância das normas sobre proteção de dados pelo operador, e, enquadrar operações de tratamento de dados nas hipóteses legais: 1- Manter registro das operações de tratamento de dados pessoais<sup>77</sup>, 2- Comunicar a ocorrência de incidente de segurança que possa acarretar risco ou dano relevante aos titulares, 3- Indicar encarregado pelo tratamento de dados pessoais: (profissional de referência, chamado *data protection officer*, que vai ser um ponto de referência nessa matéria dentro da empresa)<sup>78</sup>; 4- Privacidade *by design e by default*.

Por *by design*, quer dizer que cada produto na origem deve levar em conta a proteção de dados e da privacidade como uma etapa do desenvolvimento. Ainda que respeite a LGPD, é preciso demonstrar que isso foi uma etapa do desenvolvimento. Por *by default*, temos que não pode presumir que as pessoas abriram mão de qualquer direito, tal como uma opção de termos de uso previamente marcada ao acessar um *site*.

Por fim, chegamos ao terceiro, e último personagem. A Autoridade Nacional de Proteção de Dados (ANPD)<sup>79</sup> que é um órgão da Administração Pública Federal, que tem como papel fiscalizar se as empresas estão cumprindo com seus deveres e respeitando os direitos dos titulares dos dados. Caso estejam desrespeitando, a ANPD é quem aplica sanções e penalidades. Além disso, se espera que a ANPD traga os parâmetros das melhores práticas, do que é esperado, que ajude a traduzir todos esses requisitos que a lei prevê. As sanções são variadas<sup>80</sup>: I- Advertência; II- Multa de até 2% do faturamento da empresa, grupo ou conglomerado no Brasil no seu último exercício, excluídos os tributos, limitada, no total, a R\$ 50 milhões por infração; III- Multa diária limitada, no total a R\$ 50 milhões por infração; IV- Publicização da infração; V- Bloqueio dos dados pessoais a

---

<sup>77</sup> Art. 37 da LGPD. Disponível em: < [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

<sup>78</sup> Art. 41 da LGPD. Disponível em: < [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

<sup>79</sup> ANPD: Criada no dia 8 de julho de 2019, logo após a tramitação da Medida Provisória n.869, a Autoridade Nacional de Proteção de Dados possui atribuições relacionadas à proteção de dados pessoais e à privacidade, devendo fiscalizar o cumprimento da Lei Geral de Proteção de Dados Pessoais.

<sup>80</sup> Art. 52 da LGPD. Disponível em: < [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>.

que se refere a infração; VI- Eliminação dos dados pessoais a que se refere a infração; VII- Indenizações, sejam elas individuais ou coletivas.

O compartilhamento de dados é um ponto que traz muita atenção para a LGPD, porque sempre que se retira o dado da empresa e coloca esse dado numa outra empresa, num parceiro, num fornecedor; se amplia e muito os riscos desse tratamento. A razão é óbvia. Dentro de seu ambiente, é possível ter condições de atestar que adota as medidas necessárias, que faz o dever de casa, ou ainda que analisou e mapeou os processos. Agora é muito mais difícil, dizer o mesmo da casa do vizinho, da casa de um parceiro, onde não se tem esse controle. Então, a LGPD traz algumas exigências para o compartilhamento de dados pessoais. Quando se tratar de dados que são baseados no consentimento, depende de consentimento adicional e específico. Também exige a transparência, que geralmente se resolve através das políticas de privacidade, onde as empresas informam com quem elas costumam compartilhar os dados. Ademais, o titular tem direito de obter informação das entidades públicas e privadas com as quais o controlador realiza o uso compartilhado de dados.

Outrossim, deve-se dar o cumprimento de direitos do titular. O agente de tratamento deve informar, de maneira imediata, aos demais com os quais tenha realizado uso compartilhado de dados a correção, a eliminação, a anonimização ou o bloqueio dos dados, para que repitam idêntico procedimento, exceto nos casos em que esta comunicação seja comprovadamente impossível ou implique esforço desproporcional.

No que tange ao Poder Público, temos que o uso compartilhado de dados pessoais pelo Poder Público deve atender a finalidades específicas de execução de políticas públicas e atribuição legal pelos órgãos e pelas entidades públicas, respeitados os princípios de proteção de dados sensíveis.

Além disso, no caso da comunicação ou o uso compartilhado de dados pessoais de pessoa jurídica de direito público, a pessoa de direito privado será informada pela autoridade nacional e dependerá de consentimento do titular. E, também, a autoridade nacional poderá estabelecer normas complementares para as atividades comunicação e de uso compartilhado de dados pessoais.

Acerca dos dados sensíveis, temos que o compartilhamento de dados pessoais sensíveis entre controladores com objetivo de obter vantagem econômica poderá ser objeto de vedação ou de regulamentação por parte da autoridade nacional. Também é



vedada à comunicação ou o uso compartilhado entre controladores de dados pessoais sensíveis referentes à saúde com objetivo de obter vantagem econômica, exceto nas hipóteses relativas a prestação de serviços de saúde, de assistência farmacêutica e de assistência à saúde, em benefício dos interesses dos titulares de dados.

No entanto, o foco principal do compartilhamento de dados, está na relação com parceiros e fornecedores. Isso porque, é o mais comum. Dificilmente uma empresa consegue entregar isolada, sem que para tanto precise do apoio de outras empresas, outros negócios, numa cadeia produtiva. É muito usual que se tenha essas colaborações. E as vezes elas são em grande número.

A LGPD estabelece o que chamamos de responsabilidade solidária<sup>81</sup>. O operador responde solidariamente pelos danos causados pelo tratamento quando descumprir as obrigações da LGPD ou quando não tiver seguido as instruções lícitas do controlador, hipótese em que o operador se equipara ao controlador. Ademais, os controladores que estiverem diretamente envolvidos no tratamento no qual decorreram danos ao titular dos dados respondem solidariamente. E, também, aquele que reparar o dano ao titular tem direito de regresso contra os demais responsáveis, na medida de sua participação no evento danoso.

Em geral, o que está acontecendo é que a adequação à LGPD passou a ser um critério adicional de seleção de parceiros e fornecedores. As empresas têm criado normas internas na seleção de parceiros onde esse item é considerado, e aonde dá preferência para empresas que tenham condições de mostrar que estão conformes. Há uma preocupação toda especial com os contratos desses fornecedores e com a inclusão de cláusulas que protejam que resguardem a empresa/organização desse enquadramento na LGPD. E muitas vezes, tem-se exigido desses fornecedores declarações de conformidade, onde o fornecedor garanta que está atendendo a tudo aquilo que a LGPD determina.

Em termos de contratos, temos vários tipos de contratos hoje em dia que envolvem dados e pode ser entre dois controladores, um controlador com um operador, dois operadores, ou também de um operador para um controlador. Nesses contratos, ainda, se inserem cláusulas que ganham muita relevância. Um capítulo novo de proteção de dados,

---

<sup>81</sup> Responsabilidade Solidária: De acordo com o Tribunal de Justiça do Distrito Federal e dos Territórios “havendo pluralidade de devedores, o credor pode cobrar o total da dívida de todos ou apenas do que achar que tem mais probabilidade de quitá-la”.

tal qual, temos um capítulo anticorrupção, um capítulo relacionado a obrigações trabalhistas, agora temos um capítulo para proteção de dados, onde ali vai constar a finalidade do tratamento, por que se está compartilhando dados. Então é preciso deixar claro no contrato, o que que é o objeto daquele compartilhamento.

Normalmente, o fornecedor tem que aderir as políticas do contratante, sobre privacidade, segurança e códigos de ética e de conduta. E para evitar de ter que rediscutir pontos já superados nos contratos vigentes, tem se aplicado muitas vezes a declaração de conformidade. Essa declaração é encaminhada após emissão de comunicado da empresa que está fazendo um trabalho de conformidade e quem trabalha ou presta serviços para essa empresa também precisa estar em conformidade.

Por fim, outro ponto polêmico é o da transferência internacional de dados, onde a preocupação da legislação de proteção de dados é muito maior. Inclusive é a razão de edição da nossa lei, reflexo de uma regra da GDPR, dado o supertráfego de dados que desconhece fronteiras.

As grandes empresas nos segmentos de hospedagem nem sempre são empresas nacionais, por vezes, são empresas que provem de outros países, e que, por sua vez, tem servidores ainda em outros mais. Começamos a ver que existe uma certa limitação, talvez no que se resolve internamente com base nas nossas leis do país, pois a circulação dos dados desconhece as fronteiras físicas.

Como solução para essa dificuldade prática de circulação global de dados lá na União Europeia, se elegeu a GDPR. Essa lei é criada para controlar a relação de dados dos cidadãos que estão naquele território, mas que em primeiro lugar, ela vai se aplicar também mesmo se quem coleta e trata esses dados está em outro território. O mesmo vale aqui para a LGPD.

Ou seja, é uma regra que pode suscitar o que chamamos de aplicação extraterritorial da lei. Não importa aonde está, se está tratando dados de cidadãos que estão aqui território brasileiro, não importa aonde está servidor, terá de estar sujeito a essa a LGPD. Mas mais do que isso essas leis novas de proteção de dados têm também alguma preocupação e trazem regras específicas quando o dado em si, transita fora das suas fronteiras, não só quando quem está conectando está em outro território.

A grande preocupação aqui que vai se ter, é se esse outro país que está recepcionando nosso dado, será que ele vai ter um regime de proteção tão rigoroso quanto o nosso? Essa é a principal análise que é feita a fim de se permitir o uso de fornecedores que tenham servidores ou que estejam estabelecidos no exterior. A avaliação consiste em analisar se o dado lá no território desse parceiro/fornecedor vai estar seguro tanto quanto vai estar o dado sendo tratado aqui no Brasil ou no caso da União Europeia dentro do contexto onde surgiu essas regras.

Então as legislações na verdade criam alguns mecanismos, algumas ferramentas à disposição de quem se trata o dado para que ele possa avaliar a segurança e a possibilidade da adoção do fornecedor e do envio dos dados para outros países. Vejam que a LGPD, nesse sentido é expressa, ela diz que a transferência internacional de dados somente é permitida nas hipóteses do art.33 da referida lei<sup>82</sup>.

Verifica-se, portanto que na governança de dados e segurança da informação existem pontos cruciais que devem ser atacados em uma análise de risco, tais como: O uso excessivo de dados (sensíveis, principalmente); o compartilhamento de dados (principalmente, o compartilhamento internacional); a diversidade de banco de dados; a ausência de banco de dados; a ausência de hierarquização de acessos; o tratamento de dados de crianças e adolescentes e o nível de maturidade para atendimento dos novos direitos e deveres para titulares e controladores.

---

<sup>82</sup> Art. 33. A transferência internacional de dados pessoais somente é permitida nos seguintes casos: I - para países ou organismos internacionais que proporcionem grau de proteção de dados pessoais adequado ao previsto nesta Lei; II - quando o controlador oferecer e comprovar garantias de cumprimento dos princípios, dos direitos do titular e do regime de proteção de dados previstos nesta Lei, na forma de: a) cláusulas contratuais específicas para determinada transferência; b) cláusulas-padrão contratuais; c) normas corporativas globais; d) selos, certificados e códigos de conduta regularmente emitidos; III - quando a transferência for necessária para a cooperação jurídica internacional entre órgãos públicos de inteligência, de investigação e de persecução, de acordo com os instrumentos de direito internacional; IV - quando a transferência for necessária para a proteção da vida ou da incolumidade física do titular ou de terceiros; V - quando a autoridade nacional autorizar a transferência; VI - quando a transferência resultar em compromisso assumido em acordo de cooperação internacional; VII - quando a transferência for necessária para a execução de política pública ou atribuição legal do serviço público, sendo dada publicidade nos termos do inciso I do caput do art. 23 desta Lei; VIII - quando o titular tiver fornecido o seu consentimento específico e em destaque para a transferência, com informação prévia sobre o caráter internacional da operação, distinguindo claramente está de outras finalidades; ou IX - quando necessário para atender as hipóteses previstas nos incisos II, V e VI do art. 7º desta Lei. Parágrafo único. Para os fins do inciso I deste artigo, as pessoas jurídicas de direito público referidas no parágrafo único do art. 1º da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), no âmbito de suas competências legais, e responsáveis, no âmbito de suas atividades, poderão requerer à autoridade nacional a avaliação do nível de proteção a dados pessoais conferido por país ou organismo internacional. LGPD. Disponível em:<[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm)>.

A par dos fatores de risco, é preciso também priorizar o atendimento desses riscos, e entender que a LGPD também é gestão de riscos, tendo em vista que nem sempre teremos 100% (cem por cento) de conformidade. Como vimos, existem boas práticas e a governança de dados que auxiliam nesse processo através de: a) Programa de Governança de Privacidade, b) Relatório de Impacto à Proteção de Dados Pessoais, c) Plano de Resposta a Incidentes, d) Políticas Claras e Acessíveis, e) Códigos de Conduta/Certificações, f) Medidas de Conscientização, Treinamento, Fiscalização e Sanções Disciplinares.

A questão de um plano de continuidade dos negócios, seja ele estatal ou não, para retomada de operações em situações de crise, com menor impacto possível, preservando processos críticos é crucial para a resiliência, isto é, rápida recuperação após a crise. Não significa estar imune ao vazamento de dados, mas a capacidade de lidar com tais situações fazendo uso das boas práticas e governança, seja através de um comitê de crise<sup>83</sup>, seja através do encarregado (*data protection officer*).

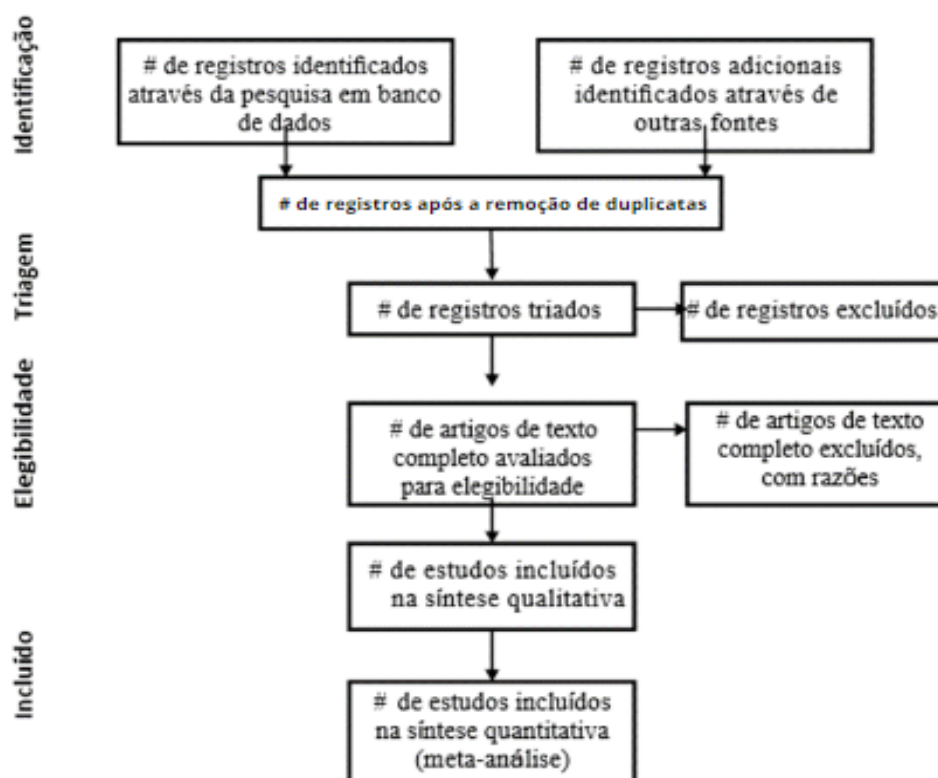
### **2.3 Revisão Sistemática de Literatura**

Uma revisão sistemática (*SLR*), assim como outras revisões de literatura, é um tipo de pesquisa em que usamos publicações/registros como fonte de dados. Esta investigação fornece o estado da arte sobre um determinado problema, aplicando métodos de pesquisa explícitos, avaliando criticamente os resultados e, em seguida, resumindo as informações selecionadas. Torna-se útil para identificar tópicos que precisam de novos estudos, ajudando a orientar futuras pesquisas. A Figura 8 ilustra o caminho metodológico de uma SLR.

---

<sup>83</sup> Comitê de Crise: Profissionais treinados, de acordo com o risco, prontos para atuar imediatamente. Foi popularizado com a pandemia, onde Estados e Municípios tiveram que criar os seus comitês de crise, mas é algo que é necessário também para as empresas frente aos vazamentos de dados digitais, onde esse comitê é um grupo de profissionais treinados, preparados para atuar, mas treinados de acordo com a natureza do risco. Prontos para atuar imediatamente, isso implica também existirem backups, existirem alternativas para eventual indisponibilidade de um e outro, mas o interessante é que são profissionais com competências complementares. Um com habilidades jurídicas, outro com habilidades de comunicação, outro com habilidades de segurança e tecnologia da informação e assim por diante.

**Figura 8 Processo de Revisão Sistemática de Literatura**



Fonte: Imagem do Autor.

Com base nessa metodologia, estruturamos esta *SLR* em: (1) definição da questão de pesquisa e termos de pesquisa; (2) bancos de dados e filtros; (3) critérios de inclusão e exclusão; (3) resultados e artigos selecionados; (5) análise quantitativa; e (6) análise qualitativa. Realizamos pesquisas durante o período de 12/03/2023 a 12/05/2023. Em seguida, repetimos e atualizamos a *SLR* durante o período de 21/08/2023 a 23/11/2023.

### **1. Definição da questão de pesquisa e termos de pesquisa**

Partindo da questão de pesquisa “Como podemos representar o conhecimento jurídico de maneira a contribuir para a justificação adequada das predições de IA em decisões judiciais?” definimos termos de pesquisa para cada domínio de conhecimento abordado.

- A) Inteligência Artificial Explicável no Direito: “*XAI*”, “Inteligência Artificial Explicável”, “*Explainable artificial intelligence*”.
- B) Inteligência Artificial no Direito: “Inteligência Artificial no Direito”, “*Artificial Intelligence on Law*”, “IA no Direito” e “*AI on Law*”.

## 2. Base de dados e filtros

- Bancos de dados: Inicialmente, escolhemos os bancos de dados Scopus, Web of Science e SciELO como repositórios de interesse devido à interdisciplinaridade da pesquisa e por indexar revistas com alto fator de impacto. Também adicionamos o Google Scholar para procurar quaisquer registros interessantes não vistos nos três primeiros bancos de dados. Posteriormente, na atualização da Revisão Sistemática de Literatura, incluímos o IEEE Xplore e a ACM Digital Library, devido ao maior interesse em literatura específica que abrange computação e tecnologia da informação.
- Campos do documento: Limitamos os campos do documento por título, resumo e palavras-chave, esperando obter artigos assertivos em relação ao cruzamento de assuntos. Excepcionalmente no Google Scholar, os campos são reduzidos por título devido às limitações deste banco de dados.
- Tipo de documento: Filtramos o tipo por artigos de periódicos, com a intenção de retornar artigos mais robustos e completos ao nível da experimentação (artigos de conferência geralmente dizem respeito a fases de pesquisa em andamento). Observamos que no Google Scholar não é possível escolher este filtro e para o ACM aplicamos um equivalente, denominado por eles como ‘artigo de pesquisa’.
- Idioma, ano e editora: Não restringimos a pesquisa por idioma, ano ou editora. No entanto, devido ao aspecto interdisciplinar dos artigos selecionados, conseqüentemente todos eles estão escritos em inglês.

## 3. Critério de Inclusão e Exclusão

- **Inclusão:** Selecionamos documentos que apresentam experimentos, estudos de caso ou sistemas acerca de IA aplicada ao Judiciário, que detalham as técnicas utilizadas, considerando que estamos buscando a aplicação de IA nos diferentes sistemas jurídicos no mundo.

- **Exclusão:** Artigos essencialmente teóricos, ou que discutem tendências na área, ou com temas desconexos da área jurídica.

#### 4. Resultados e Artigos Selecionados

A Tabela 2 indica as estratégias de pesquisa utilizadas em cada banco de dados e os resultados obtidos.

**Tabela 2 - Artigos Identificados Inicialmente**

Base de Dados	Termos pesquisados	Campos do Documento	Tipo de Documento	RT	RD	R S	NR	N S
Scopus	“Inteligência Artificial no Direito” (“Artificial Intelligence on Law”, “IA no Direito” e “AI on Law”) e “XAI” (“Inteligência Artificial Explicável”, “Explainable artificial intelligence”)	Título, resumo e palavra-chave	Artigo Científico	0	3	0	0	0
Web of Science	“Inteligência Artificial no Direito” (“Artificial Intelligence on Law”, “IA no Direito” e “AI on Law”) e “XAI” (“Inteligência Artificial Explicável no Direito”, “Explainable artificial intelligence on Law”)	Título, resumo e palavra-chave	Artigo Científico	0	0	0	0	0

SciELO	“Inteligência Artificial no Direito” (“Artificial Intelligence on Law”, “IA no Direito” e “AI on Law”) e “XAI” (“Inteligência Artificial Explicável”, “Explainable artificial intelligence”)	Título, resumo e palavra-chave	Artigo Científico	1	16	12	0	0
IEEE Xplore	“XAI” (“Inteligência Artificial Explicável no Direito”, “Explainable artificial intelligence on Law”)	Título, resumo e palavra-chave	Artigo Científico	29	23	6	0	0
ACM Digital Library	“XAI” (“Inteligência Artificial Explicável”, “Explainable artificial intelligence”)	Título, resumo e palavra-chave	Artigo Científico	82.282	82.281	1	13	0
Google Scholar	“Inteligência Artificial no Direito” (“Artificial Intelligence on Law”, “IA no Direito” e “AI on Law”) e “XAI” (“Inteligência Artificial Explicável”, “Explainable artificial intelligence”)	Tudo	Tudo	4.360.000	4.359.999	1	38.500	0

\*TR: resultados totais; DR: resultados divergentes; SR: resultados selecionados; NR: novos resultados (atualização da SLR); NS: novos resultados selecionados (atualização da SLR).

Após aplicar os critérios e avaliar a qualidade metodológica de cada manuscrito, selecionamos 29 documentos, que listamos na Tabela 3.



**Tabela 3 - Artigos Selecionados**

<b>Título</b>	<b>Autores</b>	<b>Afiliação</b>	<b>Ano</b>	<b>Tipo de Documento:</b>
<i>Tecnología de reconocimiento facial y sus riesgos en los derechos humanos.</i>	SANABRIA MOYANO, Jesús E.; ROA AVELLA, Marcela del Pilar; LEE PÉREZ, Oscar Iván.	Revista Criminalidad, v. 64, n. 3, p. 61-78.	<b>2022</b>	<b>Artigo Científico</b>
Processo Penal e Inteligência Artificial: Rumo a um Direito (Processual) Penal da Segurança Máxima?	SANTOS, Hugo Luz dos.	Revista Brasileira de Direito Processual Penal, v. 8, p. 767-821.	<b>2022</b>	<b>Artigo Científico</b>
<i>Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos.</i>	AVELLA, Marcela del Pilar Roa; SANABRIA-MOYANO, Jesús Eduardo; DINAS-HURTADO, Katherin	Revista Brasileira de Direito Processual Penal, v. 8, p. 275-310.	<b>2022</b>	<b>Artigo Científico</b>
<i>Inteligencia artificial y tutela cautelar. Especial referencia a la prisión provisional.</i>	PENA, Ana María Neira.	Revista Brasileira de Direito Processual Penal, v. 7, p. 1897-1933.	<b>2022</b>	<b>Artigo Científico</b>
<i>Análisis de la inteligencia artificial en las relaciones laborales.</i>	GRANADOS FERREIRA, Jackeline.	Revista CES Derecho, v. 13, n. 1, p. 111-132.	<b>2022</b>	<b>Artigo Científico</b>
A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais.	DOURADO, Daniel de Araujo; AITH, Fernando Mussa Abujamra.	Revista de Saúde Pública, v. 56, 2022	<b>2022</b>	<b>Artigo Científico</b>
<i>Artificial Intelligence Applied to Robotics in Armed Conflict. Debates on Autonomous Lethal Weapons Systems and the (In) Sufficiency of the Protocols of International Humanitarian Law.</i>	PORCELLI, Adriana Margarita.	<i>Estudios Socio-Juridicos</i> , v. 23, p. 483.	<b>2021</b>	<b>Artigo Científico</b>
<i>Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisiones judiciales.</i>	CÁRDENAS, Erick Rincón; MOLANO, Valeria Martinez.	Revista Direito GV, v. 17.	<b>2021</b>	<b>Artigo Científico</b>

<i>Purposes of education along with artificial intelligence developments.</i>	BARRIOS-TAO, Hernando; DÍAZ, Vianney; GUERRA, Yolanda M	Cadernos de Pesquisa, v. 51	2021	Artigo Científico
Métodos de resolução digital de controvérsias: estado da arte de suas aplicações e desafios.	MOULIN, Carolina Stange Azevedo.	Revista Direito GV, v. 17.	2021	Artigo Científico
<i>Legal tech: analytics, inteligência artificial e as novas perspectivas para a prática da advocacia privada.</i>	ANDRADE, Mariana Dionísio de; ROSA, Beatriz de Castro; PINTO, Eduardo Régis Girão de Castro.	Revista Direito GV, v. 16.	2020	Artigo Científico
O uso da inteligência artificial na aplicação do direito público: o caso especial da cobrança dos créditos tributários-um estudo objetivado nos casos brasileiro e português.	ABRAHAM, Marcus; CATARINO, João Ricardo.	e-Pública: Revista Eletrônica de Direito Público, v. 6, n. 2, p. 188-219.	2019	Artigo Científico
<i>La Mediación Mercantil Ante los Nuevos Retos de la Inteligencia Artificial: Una Visión Comparada Entre los Ordenamientos Jurídicos Español e Italiano.</i>	CARDONA, Noemí Jiménez.	Revista Internacional Consinter de Direito, p. 493-505.	2022	Artigo Científico
<i>Can Judges be Replaced by Machines? The Brazilian Case.</i>	LIMBERGER, Têmis; GIANNAKOS, Demétrio Beck da Silva; SZINVELSKI, Mártin M.	Mexican law review, v. 14, n. 2, p. 53-81.	2022	Artigo Científico
Identificação de evasão fiscal utilizando dados abertos e inteligência artificial.	XAVIER, Otávio Calaça et al.	Revista de Administração Pública, v. 56, p. 426-440.	2022	Artigo Científico
<i>Enhancing the value of patents as corporate assets in South Africa: How can artificial intelligence (AI) assist?</i>	NDLOVU, Lonias.	Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, v. 24, n. 1.	2021	Artigo Científico

<i>The regulation and use of artificial intelligence and 5G technology to combat cybercrime and financial crime in South African banks.</i>	CHITIMIRA, Howard; NCUBE, Princess.	Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, v. 24, n. 1.	2021	Artigo Científico
<i>Ciencia Ficción como Fuente de Principios Jurídicos para regular la Inteligencia Artificial.</i>	SÁNCHEZ MORALES, Santiago.	Revista IUS, v. 15, n. 48, p. 55-76.	2021	Artigo Científico
<i>No es solo un robot: consideraciones en torno a una nueva personalidad jurídica y el redimensionamiento de las relaciones interpersonales</i>	CHÁVEZ VALDIVIA, Ana Karin.	<i>Ius et Praxis</i> , v. 26, n. 2, p. 55-77.	2020	Artigo Científico
<i>Cyborgs and the future of the human spirit.</i>	MATTHEE, Machdel.	Tydskrif vir Geesteswetenskappe, v. 53, n. 4, p. 530-557.	2013	Artigo Científico
<i>A layer-wise theoretical framework for deep learning of convolutional neural networks</i>	NGUYEN, Huu-Thiet; LI, Sitan; CHEAH, Chien Chern	<b>IEEE Access</b> , v. 10, p. 14270-14287.	2022	Artigo Científico
<i>Assessing Trustworthy AI in Times of COVID-19: Deep Learning for Predicting a Multiregional Score</i>	ALLAHABADI, Himanshi et al.	<b>IEEE transactions on technology and society</b> , v. 3, n. 4, p. 272-289,	2022	Artigo Científico
<i>Conveying the Degree of Lung Compromise in COVID-19 Patients</i>				
<i>Explaining individual and collective programming students' behavior by interpreting a black-box predictive model</i>	PEREIRA, Filipe Dwan et al.	<b>IEEE Access</b> , v. 9, p. 117097-117119	2021	Artigo Científico
<i>Interpretable Multi-criteria ABC analysis based on semi-supervised clustering and Explainable Artificial Intelligence.</i>	QAFFAS, ALAA ASIM et al.	<b>IEEE Access</b>	2023	Artigo Científico
<i>The robustness of counterfactual explanations over time</i>	FERRARIO, Andrea; LOI, Michele.	<b>IEEE Access</b> , v. 10, p. 82736-82750.	2022	Artigo Científico
<i>Algorithmic Unfairness through the Lens of EU Non-Discrimination Law: Or Why the Law is not a Decision Tree.</i>	WEERTS, Hilde et al.	<b>Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency</b>	2023	Artigo Científico

*Explainable artificial intelligence applications in cyber security: State-of-the-art in research*

ZHANG, Zhibo et al

IEEE Access

2022

Artigo Científico

*Explainable AI and Law: An Evidential Survey*

RICHMOND, Karen  
McGregor et al.

Digital Society, v. 3, n. 1,  
p. 1

2024

Artigo Científico

## 5. Análise Quantitativa

A análise quantitativa desta revisão avalia o ano com a maior incidência de artigos, os autores e seus respectivos países que mais produziram sobre o tema em discussão. A Figura 9 mostra que nos últimos dez anos, o intervalo de 2020 a 2023 teve o maior número de publicações. Esse intervalo coincide com o início e final da pandemia de Covid-19, o que nos leva a crer que se a pandemia trouxe a fusão de trabalho e local de trabalho, a IA se tornou também um ramo muito mais atrativo, até mesmo para o judiciário brasileiro em termos de informatização do processo e incentivo ao uso de novas tecnologias. Isso sugere que esses movimentos são globais no ambiente judicial.

**Figura 9 - Número de publicações por ano entre 2014 e 2024.**

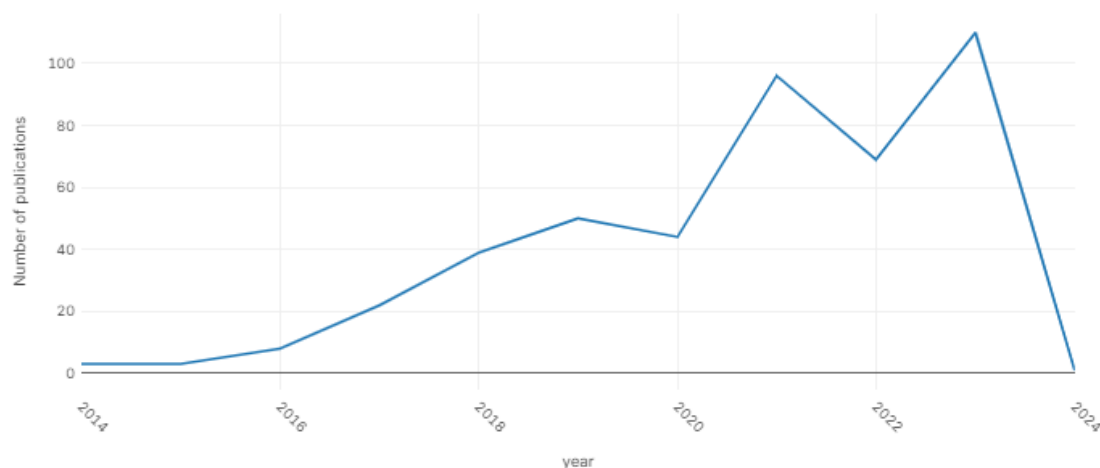
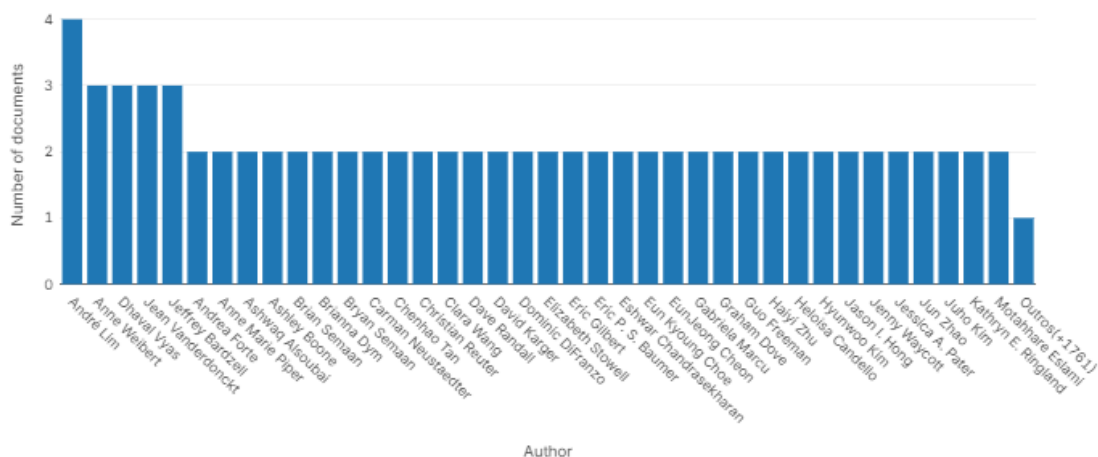


Figura 90 – Número de publicações por ano entre 2014 e 2024 (SLR).

As Figuras 10 e 11, por sua vez, mostram uma variedade de pesquisadores de diferentes universidades ao redor do mundo que trabalharam em pesquisas associadas a IA no Poder Judiciário de diferentes países do mundo. O país líder em número de

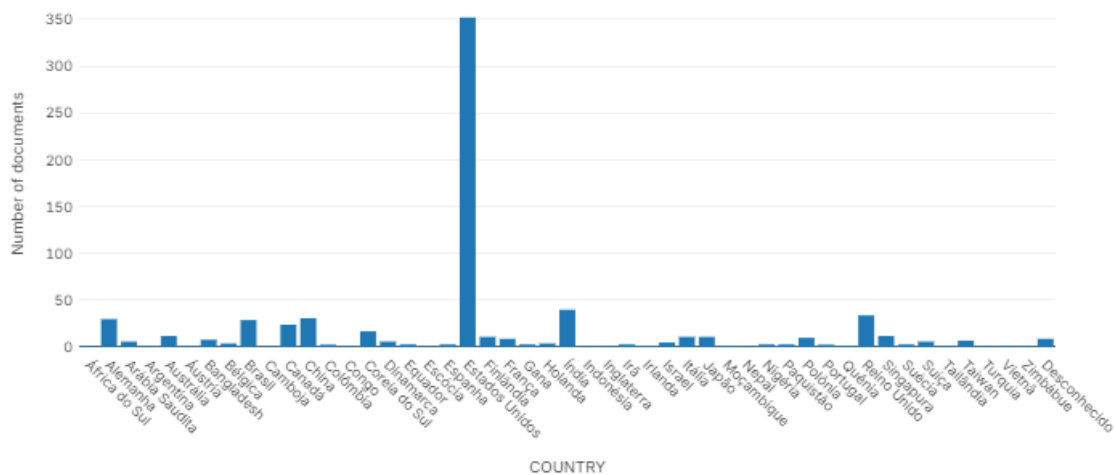
pesquisas produzidas é sem dúvidas os EUA (352), seguido de Índia (40), Reino Unido (34), China (31), Alemanha (30), Brasil (29) e Canadá (24). Entre os autores que mais produziram obras nesse campo de estudo temos André Lim, Anne Weibert, Dhaval Vyas, Jean Vanderdonckt e Jeffrey Bardzell, como pode ser visto na Figura 10.

**Figura 10 - Número de documentos por autor**



**Figura 91 - (SLR) Número de documentos por autor.**

**Figura 11- Número de documentos por país**



**Figura 92 - (SLR) Número de documentos por país**

A análise foi atualizada, com a inclusão de dados da *ACMLibrary* e *IEEEExplorer*, assim como pela a inclusão de novos artigos. O que de modo algum, mudou o cenário interpretativo, qual seja, de que devido à pandemia de Covid-19 e ao crescimento

exponencial das transações online, podemos ver que a pesquisa e o investimento em ODR e IA no judiciário foram se tornando uma constante global.

## 6. Análise Qualitativa

A análise qualitativa desta revisão apresenta uma visão geral das técnicas de IA usadas para avançar o estado da arte no domínio da IA e *XAI* no Direito. A Figura 12 mostra uma nuvem de palavras formada com as palavras-chave dos artigos, nas quais podemos notar termos destacados em seu centro.

Figura 12 - Nuvem de Palavras mais citadas nos artigos científicos



Figura 93 - (SLR) Nuvem de Palavras mais citadas nos artigos em Revisão Sistemática de Literatura.

Após a leitura e avaliação dos artigos, observamos que na maioria deles, as técnicas de IA servem para representar o conhecimento jurídico nos sistemas ODR para uma melhor interação com as partes em disputa (como os termos Ontologia, Sistemas Especialistas). A arquitetura de *SHAP values* também é amplamente utilizada para treinamento de programação, com o intuito de buscar a interpretabilidade de modelos caixa preta (*black-boxes*), sendo citada, inclusive, para fins educacionais de programação.

Isso se torna ainda mais relevante, dado que permite a melhor identificação e correção de vieses discriminatórios de um modelo de IA.

Além disso, a ideia de utilizar o método *SHAP* para atingir a explicabilidade da IA esteve presente em diversos trabalhos, os quais foram detalhadamente explicados na seção 3. No entanto, parte considerável dos estudos também se destinam a ressaltar a importância de uma fixação legal dos parâmetros éticos relacionados ao uso de novas tecnologias pelo Judiciário. Na primeira Revisão Sistemática de Literatura (*SLR*), encontramos apenas um artigo que aborda superficialmente *XAI* no Direito, indicando uma lacuna de conhecimento, ou seja, um problema a ser pesquisado. Então, com a atualização, mais sete foram adicionados, o que pode sugerir uma direção da comunidade científica para este campo, especialmente com o fim da pandemia de Covid-19, o surgimento do *ChatGPT*, concorrentes e derivados. Consequentemente, há uma crescente necessidade de resolver problemas com o apoio da tecnologia hoje em dia.

### **3- ESTADO DA ARTE DE IA E XAI NO DIREITO**

Neste capítulo, vamos explorar o universo fascinante da Inteligência Artificial (IA) e da Inteligência Artificial Explicável (*XAI*) no Direito. Para isso, vamos consultar as principais plataformas de pesquisa científica, como *Web of Science*, *Scopus* e *SciELO*, e mapear as constelações de conhecimento que elas nos revelam. No entanto, nessas plataformas, não encontramos artigos sobre o tema de Inteligência Artificial Explicável no Direito, usando os termos “*XAI on Law*”, “*Explainable Artificial Intelligence*” e “Inteligência Artificial Explicável no Direito”.

Por isso, recorreremos às plataformas *ACM Library* e *IEEEExplore*, que nos apresentaram milhares de resultados, sendo 82.282 na primeira e 29 na última. Desses, apenas 8 (oito) artigos eram relevantes para o nosso estudo. Já sobre inteligência artificial no direito, encontramos um tesouro de 66 artigos, usando as palavras-chave “Inteligência Artificial no Direito”, “*Artificial Intelligence on Law*”, “IA no Direito” e “*AI on Law*”. Desses, selecionamos 21 artigos que realmente tratam do assunto, e que serão analisados neste capítulo.

Agora, prepare-se para mergulhar nas profundezas desses artigos selecionados. Vamos explorar os aspectos mais sombrios e controversos da IA na sociedade, antes de

emergir para apreciar os aspectos positivos e intrigantes. Inicialmente, abordaremos os artigos científicos relativos a *XAI* e o Direito seguido pelos artigos e estudos mais amplos referentes a IA e seu impacto no Direito. Então, caros leitores, ajustem seus óculos de leitura e preparem-se para uma imersão profunda neste tema fascinante e crucial.

Inicialmente, temos NGUYEN *et al*, que na obra intitulada “*A layer-wise theoretical framework for deep learning of convolutional neural networks*”<sup>84</sup>, propõem um framework teórico para o aprendizado profundo de redes neurais convolucionais (*CNNs*), que são modelos de inteligência artificial amplamente usados em diversas aplicações, mas cuja compreensão teórica ainda é limitada. O *framework* visa fornecer uma base matemática para o treinamento e a análise das *CNNs*, usando uma abordagem de aprendizado em camadas, onde cada camada é treinada de forma independente usando operações de convolução e *pooling*.

O framework também visa melhorar a robustez e a confiabilidade das *CNNs*, tornando-as mais explicáveis e transparentes, o que é importante para áreas onde a tomada de decisão de alto risco é envolvida, como na medicina ou na segurança. Ademais, também menciona a *XAI* e como a falta de compreensão teórica das *CNNs* pode limitar sua aplicação em tomadas de decisão de alto risco<sup>85</sup>.

NGUYEN *et al* argumentam que a falta de compreensão teórica das *CNNs* torna esses modelos caixas pretas (*black boxes*), o que pode limitar sua aplicação em áreas onde a transparência e a explicabilidade são essenciais, como em diagnósticos médicos ou em sistemas de segurança críticos. O framework proposto visa abordar essa questão, fornecendo uma compreensão mais profunda do processo de aprendizado em *CNNs* e melhorando sua confiabilidade e robustez. Portanto, o artigo destaca a importância da *XAI* e como o *framework* proposto pode contribuir para tornar as *CNNs* mais explicáveis e confiáveis, permitindo que elas sejam usadas em áreas onde a transparência e a explicabilidade são essenciais<sup>86</sup>.

ALLAHABADI, Himanshi *et al*. em artigo intitulado "*Assessing Trustworthy AI in Times of COVID-19: Deep Learning for Predicting a Multiregional Score Conveying*

---

<sup>84</sup> NGUYEN, Huu-Thiet; LI, Sitan; CHEAH, Chien Chern. A layer-wise theoretical framework for deep learning of convolutional neural networks. **IEEE Access**, v. 10, p. 14270-14287, 2022. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9698045/>>.

<sup>85</sup> *Ibidem*, p.11.

<sup>86</sup> *Ibidem*, p.17.



*the Degree of Lung Compromise in COVID-19 Patients*<sup>87</sup> apresentam uma autoavaliação pós-hoc de um sistema de inteligência artificial (IA) que usa *deep learning* para prever o grau de comprometimento pulmonar em pacientes com *COVID-19* a partir de radiografias de tórax. O sistema foi desenvolvido e aplicado em um hospital público durante a pandemia, seguindo as diretrizes da UE para IA confiável. O objetivo do artigo é avaliar a confiabilidade do sistema de IA e discutir as questões éticas, técnicas e específicas do domínio que surgem no uso da IA na área da saúde em um contexto de crise.

Para isso, os autores usam uma metodologia chamada *Z-Inspection*®, que consiste na análise de cenários socio técnicos, na definição de mapeamentos para o quadro de IA confiável e na identificação de questões-chave e recomendações<sup>88</sup>. Os resultados mostram que o sistema de IA é confiável e pode auxiliar os radiologistas na estimativa e comunicação da gravidade do dano pulmonar em pacientes com *COVID-19*. No entanto, os autores também apontam algumas questões que precisam ser abordadas para garantir a confiabilidade do sistema, como a transparência e explicabilidade do modelo de IA, a privacidade dos dados e a usabilidade do sistema<sup>89</sup>.

PEREIRA, Filipe Dwan *et al* na obra "*Explaining individual and collective programming students' behavior by interpreting a black-box predictive model*"<sup>90</sup> exploram o comportamento de estudantes de programação usando um sistema de inteligência artificial (IA) que prevê o seu desempenho a partir de dados de um juiz online, o *Jutge.org*<sup>91</sup>. Ademais, também apresentam uma abordagem para explicar as decisões do sistema de IA usando um framework baseado em teoria dos jogos (*SHAP*).

Muito embora, os dados foram analisados usando um modelo preditivo de caixa preta, que permitiu aos pesquisadores identificar padrões de comportamento entre os estudantes de programação; no estudo, os pesquisadores utilizaram técnicas de *XAI* para interpretar as decisões do modelo preditivo de caixa preta e entender como ele chegou a

---

<sup>87</sup> ALLAHABADI, Himanshi et al. Assessing Trustworthy AI in Times of COVID-19: Deep Learning for Predicting a Multiregional Score Conveying the Degree of Lung Compromise in COVID-19 Patients. **IEEE transactions on technology and society**, v. 3, n. 4, p. 272-289, 2022. Disponível em:< <https://ieeexplore.ieee.org/abstract/document/9845195/>>.

<sup>88</sup> *Ibidem*, p. 276.

<sup>89</sup> *Ibidem*, p. 278.

<sup>90</sup> PEREIRA, Filipe Dwan et al. Explaining individual and collective programming students' behavior by interpreting a black-box predictive model. **IEEE Access**, v. 9, p. 117097-117119, 2021. Disponível em:< <https://ieeexplore.ieee.org/abstract/document/9517104/>>.

<sup>91</sup> JUTGE.ORG. 2023. Disponível em:< <https://jutge.org/>>.

suas previsões<sup>92</sup>. Isso permitiu que eles identificassem padrões de comportamento entre os estudantes de programação e fornecessem informações úteis para os professores e programas de suporte ao estudante.

O estudo reconhece a importância da explicabilidade em inteligência artificial (*XAI*) para tornar o sistema de IA mais transparente e compreensível para os usuários. Isso porque, descobrem como a *XAI* pode ajudar a identificar os padrões de comportamento dos estudantes de programação e a fornecer informações úteis para a educação em programação<sup>93</sup>.

QAFFAS, ALAA ASIM *et al* na obra “*Interpretable Multi-criteria ABC analysis based on semi-supervised clustering and Explainable Artificial Intelligence*”<sup>94</sup> propõem uma abordagem semi-supervisionada e explicável para a análise ABC de múltiplos critérios, que é uma técnica de gerenciamento de estoques e otimização da cadeia de suprimentos.

Basicamente, a análise ABC classifica os itens em três classes (A, B e C) de acordo com sua importância para a atividade e receita total, seguindo o princípio de Pareto. A abordagem utiliza o *clustering* semi-supervisionado para agrupar os dados em *clusters* com base em suas características, e depois classifica os *clusters* em classes ABC<sup>95</sup>. A abordagem também utiliza o *framework SHAP (Shapley Additive Explanations)* para fornecer explicações detalhadas sobre os critérios utilizados para a classificação em cada classe ABC, tornando o processo mais transparente e interpretável para os tomadores de decisão<sup>96</sup>.

Ferrario *et al.* na obra intitulada “*The robustness of counterfactual explanations over time*”<sup>97</sup> propõem um método para melhorar a robustez das explicações contrafactuais ao longo do tempo em contextos de aprendizado de máquina. As explicações contra

---

<sup>92</sup> PEREIRA, Filipe Dwan et al. Explaining individual and collective programming students' behavior by interpreting a black-box predictive model. **IEEE Access**, v. 9, p. 117112, 2021. Disponível em:<<https://ieeexplore.ieee.org/abstract/document/9517104/>>.

<sup>93</sup> *Ibidem*, p. 117.114.

<sup>94</sup> QAFFAS, ALAA ASIM et al. Interpretable Multi-criteria ABC analysis based on semi-supervised clustering and Explainable Artificial Intelligence. **IEEE Access**, 2023. Disponível em:<<https://ieeexplore.ieee.org/abstract/document/10114405/>>.

<sup>95</sup> p. 43.781.

<sup>96</sup> p. 43.782.

<sup>97</sup> FERRARIO, Andrea; LOI, Michele. The robustness of counterfactual explanations over time. **IEEE Access**, v. 10, p. 82736-82750, 2022. Disponível em:<<https://ieeexplore.ieee.org/abstract/document/9851645/>>.

factuais são uma forma de explicar os resultados de modelos de aprendizado de máquina, mostrando como os dados de entrada poderiam ser alterados para obter um resultado diferente.

Ademais, também relata um estudo empírico no qual o método proposto é aplicado para mitigar a perda de robustez das explicações contra factuais ao longo do tempo, fornecendo *insights* valiosos sobre a sua eficácia. Inclusive, discutem os resultados do estudo empírico e possíveis direções futuras de pesquisa, destacando a importância de aprimorar a robustez das explicações contra factuais em contextos dinâmicos<sup>98</sup>.

De maneira resumida, o estudo se insere no campo da explicabilidade em inteligência artificial (*XAI*), que visa compreender e justificar as decisões tomadas por modelos de aprendizado de máquina. Novamente, assim como os demais, ressalta a importância da *XAI* em múltiplos contextos, incluindo o jurídico, nos quais é necessário estabelecer a conformidade com legislações e requisitos legais<sup>99</sup>.

WEERTS, Hilde *et al.* na obra intitulada "*Algorithmic Unfairness through the Lens of EU Non-Discrimination Law: Or Why the Law is not a Decision Tree*"<sup>100</sup> apresentam uma análise crítica da discriminação algorítmica e da legislação antidiscriminação da União Europeia (UE). Os autores abordam o conceito de discriminação algorítmica, definindo-o como a situação em que um algoritmo produz resultados que discriminam um grupo de pessoas em relação a outro grupo, com base em características protegidas pela lei, como gênero, raça e idade.

Destarte, explica que a discriminação algorítmica pode ocorrer de várias maneiras, como a seleção de dados de treinamento tendenciosos, a escolha de recursos discriminatórios e a otimização de métricas de justiça que perpetuam a discriminação. Além disso, também apresenta exemplos de algoritmos que produzem resultados discriminatórios em diferentes domínios, como recrutamento, crédito e saúde<sup>101</sup>.

Igualmente, o estudo descreve a legislação antidiscriminação da UE, que proíbe a discriminação direta e indireta, bem como a discriminação por associação, por percepção

---

<sup>98</sup> *Ibidem*, p. 82.744-82.747.

<sup>99</sup> *Ibidem*, p. 82.748.

<sup>100</sup> WEERTS, Hilde *et al.* Algorithmic Unfairness through the Lens of EU Non-Discrimination Law: Or Why the Law is not a Decision Tree. In: **Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**. 2023. p. 805-816. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/3593013.3594044>>.

<sup>101</sup> *Ibidem*, p. 810.

e por múltiplas características. Inclusive, discute a jurisprudência da Corte de Justiça da UE, que interpreta e aplica a legislação antidiscriminação em casos concretos. Nesse sentido, compara os casos de discriminação algorítmica com os casos de discriminação jurídica e mostra as semelhanças e diferenças entre as duas abordagens<sup>102</sup>.

Mais adiante, explora as bases normativas das métricas de justiça e intervenções técnicas que visam mitigar a discriminação algorítmica. Os autores demonstram como as suposições normativas muitas vezes permanecem implícitas em ambas as abordagens e expõe as limitações da prática atual de inteligência artificial e do direito antidiscriminação.

Conclui-se, portanto, que existem implicações para os reguladores e desenvolvedores de inteligência artificial, sugerindo que os reguladores devem garantir que os algoritmos sejam transparentes e explicáveis, que os desenvolvedores de inteligência artificial sejam responsabilizados por resultados discriminatórios e que os algoritmos sejam desenvolvidos em conformidade com a legislação antidiscriminação da UE. WEERTS, Hilde *et al.* também sugere que os desenvolvedores de inteligência artificial devem adotar uma abordagem multidisciplinar e participativa, que envolva as partes interessadas e os afetados pela inteligência artificial, e que considere os valores e os direitos fundamentais da UE<sup>103</sup>.

ZHANG, Zhibo *et al.* em “*Explainable Artificial Intelligence Applications in Cyber Security*”<sup>104</sup> apresentam uma revisão abrangente da literatura sobre aplicações de Inteligência Artificial Explicável (XAI) em segurança cibernética. Eles mostram como técnicas de Inteligência Artificial, como Aprendizado de Máquina (ML) e Aprendizado Profundo (DL), são usadas para lidar com ataques cibernéticos, mas também como elas podem ser opacas e difíceis de entender.

Os autores propõem a *XAI* como uma solução para tornar os modelos de IA mais transparentes e compreensíveis, aumentando a confiança e a responsabilidade dos usuários e especialistas em segurança. Ademais, explicam como a *XAI* pode fornecer

---

<sup>102</sup> *Ibidem*, p. 807.

<sup>103</sup> *Ibidem*, p. 814.

<sup>104</sup> ZHANG, Zhibo *et al.* Explainable artificial intelligence applications in cyber security: State-of-the-art in research. **IEEE Access**, 2022. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9875264/>>.

explicações claras e compreensíveis sobre o funcionamento e o raciocínio dos modelos de IA, reduzindo o viés e melhorando a interpretabilidade<sup>105</sup>.

Do mesmo modo, também discutem a relação da XAI com o Direito, destacando a importância de cumprir com regulamentações e leis de Proteção de Dados, como o GDPR. Eles enfatizam a necessidade de considerar questões de justiça, responsabilidade social e mitigação de riscos ao aplicar a XAI na cibersegurança, pois essa área envolve problemas sociais sérios, e não apenas cálculos de custo-benefício, como ocorre com os dados da justiça criminal<sup>106</sup>.

Resumidamente, o estudo oferece uma visão geral das aplicações de XAI em segurança cibernética, e propõe uma estrutura conceitual para classificar e comparar diferentes abordagens de XAI nesse contexto. Os autores também apresentam alguns exemplos de casos de uso e aplicações práticas de XAI em segurança cibernética.

Por fim, temos o estudo de RICHMOND, Karen McGregor et al. intitulado “*Explainable AI and Law: An Evidential Survey*”<sup>107</sup>. Ele aborda a interseção entre IA e o campo jurídico, explorando os desafios e as oportunidades da implementação de sistemas de IA para tarefas de tomada de decisão automatizada no contexto legal. Os autores destacam a importância da explicabilidade da IA (XAI) para compreender e tornar interpretáveis e explicáveis as decisões tomadas por sistemas de IA no domínio jurídico, que é composto por um conjunto complexo de regras, decisões e meta-regras.

O estudo se concentra em discutir diferentes métodos utilizados para modelar o raciocínio jurídico, incluindo raciocínio jurídico baseado em regras, raciocínio jurídico baseado em casos, raciocínio jurídico baseado em argumentação e raciocínio jurídico baseado em evidências, além de abordagens híbridas<sup>108</sup>. Eles também identificam vários desafios e oportunidades para a implementação de sistemas de IA no contexto jurídico, incluindo a necessidade de lidar com questões de justiça algorítmica, viés e discriminação, bem como a necessidade de desenvolver métricas claras para avaliar a explicabilidade da IA<sup>109</sup>.

---

<sup>105</sup> Ibidem, p. 93.114.

<sup>106</sup> Ibidem, p. 93.128.

<sup>107</sup> RICHMOND, Karen McGregor et al. *Explainable AI and Law: An Evidential Survey*. **Digital Society**, v. 3, n. 1, p. 1, 2024. Disponível em: < <https://link.springer.com/article/10.1007/s44206-023-00081-z> >.

<sup>108</sup> Ibidem, p. 13.

<sup>109</sup> Ibidem, p. 7.

Frisa-se que os autores também abordam a natureza de "caixa preta" (*black-boxes*) de algumas variantes de IA, como o aprendizado profundo, que impacta a compreensão das decisões das máquinas no domínio jurídico, tornando difícil para os seres humanos entenderem como as decisões são tomadas. Isso pode ser problemático no contexto jurídico, onde a transparência e a explicabilidade são fundamentais para garantir a justiça adequada. Com efeito, destacam a importância da XAI para tornar as decisões de IA mais interpretáveis e explicáveis para os seres humanos, especialmente no contexto jurídico.

Por fim, o estudo apresenta a colaboração interdisciplinar entre especialistas jurídicos e de IA neste documento de pesquisa, que fornece várias conclusões e insights importantes. Os autores estabelecem uma taxonomia de categorias explicativas comuns encontradas na literatura sobre XAI e Direito, que inclui raciocínio jurídico baseado em regras, raciocínio jurídico baseado em casos, raciocínio jurídico baseado em argumentação e raciocínio jurídico baseado em evidências, além de abordagens híbridas. Eles também destacam a importância de compreender as diferentes formas de raciocínio jurídico em diferentes subdomínios jurídicos e como isso pode afetar a escolha de modelos de IA apropriados<sup>110</sup>.

Superada a análise dos artigos relacionados a XAI no Direito, percebemos que vários deles utilizam o método *SHAP* para alcançar a explicabilidade de suas previsões. Mais do que isso, sua utilização transcende a aplicação no âmbito jurídico, afetando também diversos aspectos da administração, economia, educação e também da saúde.

Agora sim, no que tange os estudos referentes a Inteligência Artificial no Direito temos o artigo “Tecnologia de reconhecimento facial e seus riscos aos direitos humanos”<sup>111</sup> de SANABRIA MOYANO & ROA AVELLA discute o uso de sistemas de inteligência artificial no reconhecimento facial na aplicação da lei<sup>112</sup>. Eles argumentam

---

<sup>110</sup> Ibidem, p. 27.

<sup>111</sup> SANABRIA MOYANO, Jesús E.; ROA AVELLA, Marcela del Pilar; LEE PÉREZ, Oscar Iván. Tecnología de reconocimiento facial y sus riesgos en los derechos humanos. *Revista Criminalidad*, v. 64, n. 3, p. 61-78, 2022. Disponível em: < [http://www.scielo.org.co/scielo.php?pid=S1794-31082022000300061&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S1794-31082022000300061&script=sci_arttext) >.

<sup>112</sup> [...] *Departamento de policía de Londres tecnología Neoface. En enero de 2020, el Departamento de Policía MET (Metropolitan Police), que tiene una jurisdicción importante en Londres, emitió un comunicado acerca de la implementación del proyecto de reconocimiento facial, ejecutado por la compañía japonesa NEC (Perrigo, 2020). Este sistema de vigilancia tiene el objetivo primordial de identificar personas que se encuentran en una base de datos y que señala que requieren control por parte de la policía. Esto se hace con la finalidad de prevenir la reincidencia o la comisión de conductas contrarias a la ley. Dicha tecnología está basada en IA, con funciones de vigilancia y reconocimiento en tiempo real y cuenta con 627.707 cámaras CCTV. Sin embargo, algunos han resaltado la inexactitud con*

que, apesar dos benefícios potenciais para a investigação criminal, essas tecnologias apresentam riscos aos direitos humanos, especialmente durante a fase de investigação. As ferramentas de reconhecimento facial, que consistem em algoritmos capazes de armazenar grandes quantidades de dados, têm uma taxa de imprecisão de cerca de 81%, de acordo com dados do Departamento de polícia de Londres com tecnologia *Neoface*<sup>113</sup>, o que pode prejudicar os direitos processuais das pessoas envolvidas em processos penais.

Os autores defendem a necessidade de equilibrar o direito à propriedade intelectual das empresas desenvolvedoras e o direito ao devido processo, dada a presença de vieses e informações incorretas ou imprecisas nas decisões dos algoritmos. O artigo destaca que países como Inglaterra e Estados Unidos implementaram sistemas de reconhecimento facial com maior controle ético e jurídico.

Para evitar violações dos direitos humanos decorrentes do uso de tecnologias de reconhecimento facial, os autores sugerem que qualquer medida que aplique essas ferramentas deve estar prescrita pela lei e se ajustar aos mandatos legais. Além disso, defendem a realização de uma avaliação rigorosa dos algoritmos de reconhecimento facial para identificar e corrigir possíveis vieses, o estabelecimento de limites claros sobre o uso da tecnologia de reconhecimento facial no sistema penal e a implementação de medidas de transparência e prestação de contas<sup>114</sup>.

No artigo “Processo Penal e Inteligência Artificial: Rumo a um Direito (Processual) Penal da Segurança Máxima?”<sup>115</sup>, HS SANTOS discute a transformação do processo penal pelo uso da inteligência artificial e da internet das coisas. O autor enfatiza a necessidade de equilibrar o uso da tecnologia com os direitos fundamentais dos acusados, especialmente o direito à privacidade mental. Ele alerta que o uso excessivo da inteligência artificial pode resultar em uma ditadura digital da inteligência artificial,

---

*la que empezó a funcionar este sistema, según Jee (2019), el 81% de las decisiones que toma este algoritmo es errado, lo que genera gran desconfianza y preocupación. Ensayos realizados previamente desde agosto de 2016 en el Carnaval de Notting Hill y posteriormente en lugares como Leicester Square, Westfield Stratford y Whitehall en 2017; (Manthorpe & Martin, 2019) concluyeron que cuatro de cada cinco personas identificadas como sospechosas por el algoritmo eran inocentes (SANABRIA MOYANO, Jesús E.; ROA AVELLA, Marcela del Pilar; LEE PÉREZ, Oscar Iván. Tecnología de reconocimiento facial y sus riesgos en los derechos humanos. Revista Criminalidad, v. 64, n. 3, p. 66-67, 2022 Disponível em:< [http://www.scielo.org.co/scielo.php?pid=S1794-31082022000300061&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S1794-31082022000300061&script=sci_arttext) >).*

<sup>113</sup> Ibidem, p.66.

<sup>114</sup> Ibidem, p.72.

<sup>115</sup> SANTOS, Hugo Luz dos. Processo Penal e Inteligência Artificial: Rumo a um Direito (Processual) Penal da Segurança Máxima?. Revista Brasileira de Direito Processual Penal, v. 8, p. 767-821, 2022. Disponível em:<<https://www.scielo.br/j/rbdpp/a/fg4rvTXgNxBPH6VjknC5d9C/>>.

comprometendo a identidade humanista e antropocentrista do processo penal.

SANTOS defende que a inteligência artificial pode ser usada para melhorar a eficiência do processo penal, mas é importante equilibrar essas melhorias com os direitos fundamentais do arguido. Ele destaca que um dos principais desafios é proteger a privacidade mental do arguido, que deriva do princípio *nemo tenetur se ipsum accusare*<sup>116</sup>. A inteligência artificial pode coletar e analisar grandes quantidades de dados, incluindo informações pessoais e privadas do arguido, o que pode violar esse direito fundamental.

Por fim, o autor sugere que a robotização do ser humano e a hibridização do pensamento humano podem privar o arguido de seu direito à privacidade mental<sup>117</sup>. No entanto, o texto não fornece uma resposta clara sobre como proteger a privacidade mental do arguido em um sistema de justiça criminal cada vez mais automatizado.

Mais adiante, o artigo “Uso do algoritmo COMPAS no processo penal e riscos aos direitos humanos”<sup>118</sup>, AVELLA discute o uso de algoritmos preditivos de risco, como o COMPAS, na administração da justiça e processos penais. O autor destaca que, embora esses sistemas possam melhorar a eficiência na tomada de decisões, eles também apresentam riscos aos direitos humanos devido à falta de transparência e à presença de vieses nos algoritmos.

O COMPAS, uma ferramenta que avalia o risco de reincidência do acusado e as possibilidades de sucesso na reabilitação, é usado para apoiar decisões relacionadas à liberdade, reabilitação ou serviços sociais<sup>119</sup>. No entanto, o uso desses sistemas pode impedir a identificação do processo pelo qual se chega à classificação de nível de risco, colocando em risco direitos como o devido processo e a liberdade pessoal.

O autor conclui que é necessário um design ético e regulamentação precisa para o uso adequado desses algoritmos. No entanto, o artigo não especifica medidas concretas para garantir que os direitos humanos não sejam violados no uso do algoritmo

---

<sup>116</sup> Ibidem, p.768.

<sup>117</sup> Ibidem, p.805.

<sup>118</sup> AVELLA, Marcela del Pilar Roa; SANABRIA-MOYANO, Jesús Eduardo; DINAS-HURTADO, Katherin. Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. Revista Brasileira de Direito Processual Penal, v. 8, p. 275-310, 2022. Disponível em:<<https://www.scielo.br/j/rbdpp/a/6W9b8CHYbXcsc6qczDxCSfr/?lang=es>>.

<sup>119</sup> Ibidem, p.281.



COMPAS<sup>120</sup>.

No artigo “Inteligência artificial e tutela cautelar. Especial referência à prisão preventiva”<sup>121</sup>, PENA explora o uso da Inteligência Artificial (IA) na tomada de decisões cautelares, especificamente na prisão preventiva. O estudo detalha como os juízes avaliam os pressupostos para essa decisão e como a IA pode potencialmente melhorar esses processos. PENA examina os vieses dos juízes e como as ferramentas de IA podem aprimorar as decisões cautelares. Além disso, o estudo reflete sobre os riscos que a introdução da IA pode representar para os direitos do investigado. A pesquisa conclui que o uso da IA pode aumentar a eficácia das decisões cautelares, mas é necessário adotar precauções para garantir o respeito aos direitos fundamentais do investigado<sup>122</sup>.

As principais salvaguardas propostas por PENA incluem a precisão legal do uso da IA, a transparência do funcionamento dos algoritmos, a explicitação na motivação judicial do uso desses instrumentos e o uso do resultado algorítmico apenas como complemento às decisões judiciais. Além disso, PENA enfatiza a necessidade de considerar que os algoritmos e os dados usados pela IA são baseados em decisões humanas. Isso pode levar à replicação e até exacerbação dos vieses humanos nos resultados do algoritmo. Portanto, é crucial garantir que a IA não perpetue ou exacerbe esses vieses ao ser implementada no sistema de justiça penal<sup>123</sup>.

Em “Análise da inteligência artificial nas relações de trabalho”<sup>124</sup>, Granados Ferreira explora a interseção entre inteligência artificial (IA) e relações trabalhistas. Ela enfatiza a necessidade de estudar a evolução da IA e seu impacto no mundo do trabalho. GRANADOS FERREIRA observa que a incorporação da IA nas empresas tem provocado mudanças significativas na estrutura convencional das relações de trabalho, o que pode ter implicações para os direitos dos trabalhadores<sup>125</sup>.

GRANADOS FERREIRA adverte que a IA pode representar uma ameaça aos

---

<sup>120</sup> Ibidem, p.293.

<sup>121</sup> PENA, Ana María Neira. Inteligencia artificial y tutela cautelar. Especial referencia a la prisión provisional. Revista Brasileira de Direito Processual Penal, v. 7, p. 1897-1933, 2022. Disponível: <<https://www.scielo.br/j/rbdpp/a/BkS7hyMvvTf96xfCY4CfbLz/?lang=es>>.

<sup>122</sup> Ibidem, p.1905.

<sup>123</sup> Ibidem, p.1926.

<sup>124</sup> GRANADOS FERREIRA, Jackeline. Análisis de la inteligencia artificial en las relaciones laborales. Revista CES Derecho, v. 13, n. 1, p. 111-132, 2022. Disponível em: <[http://www.scielo.org.co/scielo.php?pid=S2145-77192022000100111&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S2145-77192022000100111&script=sci_arttext)>.

<sup>125</sup> Ibidem, p.126.

direitos dos trabalhadores, pois as decisões que afetam a qualidade de vida dos trabalhadores são frequentemente tomadas com base em algoritmos. Além disso, a coleta de dados pelo empregador por meio de meios digitais pode influenciar as decisões que afetam o emprego, levantando questões sobre possíveis violações de direitos.

Ela destaca que as empresas devem buscar um equilíbrio entre o avanço tecnológico e o respeito aos direitos dos trabalhadores. GRANADOS FERREIRA sugere que as empresas implementem a proteção necessária para o trabalho na era digital e encontrem um modelo que garanta o desenvolvimento da IA e a proteção dos trabalhadores e seus direitos<sup>126</sup>.

Por fim, a autora menciona que as empresas devem ter uma estratégia competitiva para a devida incorporação da IA e demonstrar que a aplicação é eficaz na geração de confiança aos consumidores. Ela conclui que a proteção dos direitos humanos é fundamental para garantir o bem-estar individual e social dos trabalhadores, e que a regulamentação do direito trabalhista deve ser adaptada para lidar com essas mudanças.

Igualmente interessante, DOURADO em artigo intitulado " A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais"<sup>127</sup>, aponta que, as novas tecnologias de inteligência artificial são mais promissoras para a área da saúde em geral, mas em particular o subtipo *deep learning* deve ter um impacto significativo nos próximos anos na prática clínica, na gestão de sistemas de saúde e na relação entre os pacientes e a rede assistencial – ao permitir que eles processem seus próprios dados para promover a saúde<sup>128</sup>.

O autor considera que a regulação da inteligência artificial na saúde é um tema emergente e essencial, pois o uso de inteligência artificial tem potencial para modificar a forma de prestação da assistência à saúde e do autocuidado, além de influenciar a organização dos sistemas de saúde<sup>129</sup>. A regulação pode ajudar a garantir a segurança e a eficácia das tecnologias de inteligência artificial, bem como a proteção dos dados pessoais dos pacientes. Além disso, a regulação pode ajudar a definir responsabilidades e a

---

<sup>126</sup> Ibidem, p.125.

<sup>127</sup> DOURADO, Daniel de Araujo; AITH, Fernando Mussa Abujamra. A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. Revista de Saúde Pública, v. 56, 2022. Disponível em: <<https://www.scielo.br/j/rsp/a/k38jGvJdbQSYN4MpzGZpfXw/?lang=pt>>.

<sup>128</sup> Ibidem.

<sup>129</sup> Ibidem.

estabelecer padrões éticos para o uso da inteligência artificial na saúde.

Destarte, o artigo indica que a transparência é um princípio-chave para a inteligência artificial na saúde. Isso significa que informações suficientes sobre as tecnologias de IA devem ser documentadas antes da implantação, de modo a facilitar consulta pública e entendimento sobre seu funcionamento no mundo real. Espera-se que os sistemas sejam inteligíveis e explicáveis para desenvolvedores, profissionais de saúde, pacientes, usuários e reguladores, de acordo com a capacidade de compreensão de cada grupo e mesmo de cada indivíduo. Instrumentalizar a transparência dos algoritmos é necessário para que outros princípios-chave do uso da IA na saúde tenham eficácia. Isso pode ser feito por meio de normas e leis específicas que estabeleçam padrões éticos e de segurança para o uso da inteligência artificial na saúde, bem como por meio de mecanismos de auditoria e revisão de decisões automatizadas.

PORCELLI, por sua vez, em artigo intitulado "*La inteligencia artificial aplicada a la robótica en los conflictos armados. Debates sobre los sistemas de armas letales autónomas y la (in)suficiencia de los estándares del derecho internacional humanitario*"<sup>130</sup>, aborda um tema preocupante: o uso de armas autônomas letais em conflitos armados, popularmente conhecidas como "robôs assassinos". O autor do artigo investiga se essas armas podem ser reguladas pelo Direito Internacional Humanitário e pelos Direitos Humanos e se é necessário um marco jurídico e ético para proibir sua implementação preventivamente<sup>131</sup>.

Inclusive, traz à baila o posicionamento de *Jody Williams*, prêmio Nobel da Paz em 1997, que chamou as armas letais autônomas de "robôs assassinos" e se referiu a elas como um problema principal nos conflitos armados<sup>132</sup>. No entanto, o texto não fornece informações adicionais sobre a posição específica de *Jody Williams* sobre as armas letais autônomas.

A pesquisa foi conduzida através do método científico de pesquisa bibliográfica e concluiu que, embora muitos protótipos dessas armas sejam testados em laboratórios,

---

<sup>130</sup> PORCELLI, Adriana Margarita. *Artificial Intelligence Applied to Robotics in Armed Conflict. Debates on Autonomous Lethal Weapons Systems and the (In) Sufficiency of the Protocols of International Humanitarian Law. Estudios Socio-Juridicos*, v. 23, p. 483, 2021. Disponível em: <[https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/estscj23&section=23](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/estscj23&section=23)>.

<sup>131</sup> Ibidem, p.486.

<sup>132</sup> Ibidem, p.483.

ainda não há consenso internacional sobre sua regulamentação<sup>133</sup>. Enquanto alguns argumentam que essas armas podem reduzir baixas humanas e ajudar na reconstrução objetiva dos fatos, outros afirmam que elas não possuem empatia ou sentimentos de compaixão e podem violar os princípios do direito humanitário. Além disso, o texto conclui que, considerando o risco potencial dessas armas, uma proibição preventiva pode ser necessária antes que seja tarde demais para mudar de rumo. Portanto, é essencial que a comunidade internacional se envolva nesse debate e estabeleça medidas para garantir que essas armas não sejam implementadas em conflitos futuros<sup>134</sup>.

Finalmente, após uma análise profunda dos aspectos mais preocupantes e alarmantes sobre a inteligência artificial abordados nos sete artigos anteriores, é hora de olharmos para os aspectos positivos e curiosos que essa tecnologia pode nos proporcionar. Isso porque, a IA pode ser aplicada em diversas áreas, como saúde, transporte, educação, indústria, assim como direito, trazendo inovações e soluções para problemas complexos.

CÁRDENAS E MOLANO, em sua obra intitulada ‘Um estudo sobre a possibilidade de aplicar a inteligência artificial nas decisões judiciais’<sup>135</sup>, exploram a viabilidade da aplicação da inteligência artificial (IA) no campo judicial. Eles argumentam que o objetivo não é substituir os juízes, mas sim fornecer um sistema de apoio que disponibilize todos os elementos normativos e jurisprudenciais aplicáveis em uma decisão. A pesquisa acadêmica e teórica conduzida pelos autores sugere que a participação de um sistema jurídico especializado na tomada de decisões judiciais pode resultar em um aumento da segurança jurídica no sistema judicial. No entanto, eles enfatizam que a responsabilidade final por essas decisões sempre recairá sobre o ser humano, especificamente o juiz.

Inicialmente, o texto analisa a viabilidade da aplicação da IA em decisões judiciais. Além disso, os autores indicam que é necessário um estudo detalhado para determinar até que ponto é viável alocar parte do orçamento do ramo judicial para a implementação de sistemas especializados<sup>136</sup>. Eles defendem que a principal implicação da participação de um sistema especializado jurídico na tomada de decisões judiciais é o

---

<sup>133</sup> Ibidem, p.486.

<sup>134</sup> Ibidem.

<sup>135</sup> CÁRDENAS, Erick Rincón; MOLANO, Valeria Martinez. *Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisiones judiciales*. Revista Direito GV, v. 17, 2021. Disponível em: <<https://www.scielo.br/j/rdgv/a/vZDXYYPRrcwgsgJDWQf97QG/>>.

<sup>136</sup> Ibidem, p.14.

aumento da segurança jurídica no sistema judicial, resultante de um aumento na confiança no mesmo. Eles argumentam que a aplicação da IA nas decisões judiciais pode aumentar a confiança que a cidadania deposita na administração da justiça, o que seria uma implicação importante para o ramo judicial e as decisões judiciais, pois aumentaria a segurança jurídica depositada neste ramo<sup>137</sup>.

Adicionalmente, os autores apontam que a participação de um sistema especializado jurídico na tomada de decisões judiciais pode aumentar a eficiência da administração da justiça, o que justificaria os custos de sua implementação. Para defender essa tese, eles citam o exemplo da cidade chinesa de *Hengezhou*, onde foi realizado o primeiro tribunal cibernético em agosto de 2017. Este tribunal, competente em assuntos como operações em rede, comércio eletrônico e propriedade intelectual, introduziu um software que utiliza IA para auxiliar na tomada de decisões em procedimentos legais ordinários, como a abertura e o processamento de casos<sup>138</sup>. Este software ajuda os juízes humanos em questões repetitivas, permitindo que se concentrem em questões jurídicas fundamentais.

Em “Propósitos da educação juntamente com o desenvolvimento da inteligência artificial”<sup>139</sup>, BARRIOS-TAO, DÍAZ & GUERRA discutem o impacto da IA na educação. Eles propõem uma interpretação equilibrada dos benefícios e desafios da IA na formação educacional, considerando aspectos humanísticos, autonomia e equidade.

Os autores destacam que a IA pode ser usada para promover o pensamento criativo e crítico, bem como uma filosofia de aprendizagem contínua. No entanto, expressam preocupações com a estabilidade do emprego dos funcionários, implicações éticas em relação à confidencialidade dos dados e à privacidade dos atores educacionais, bem como a possível redução do papel dos professores a meros motivadores ou facilitadores da aprendizagem<sup>140</sup>.

Mais especificamente, quanto a equidade e formação integral na educação, temos que a IA pode ser usada para apoiar a aprendizagem em países em desenvolvimento por meio de telepresença colaborativa e sistemas de tutoria inteligentes como o *Computer*

---

<sup>137</sup> Ibidem, p.7.

<sup>138</sup> Ibidem, p.13.

<sup>139</sup> BARRIOS-TAO, Hernando; DÍAZ, Vianney; GUERRA, Yolanda M. Purposes of education along with artificial intelligence developments. *Cadernos de Pesquisa*, v. 51, 2021. Disponível em: <<https://www.scielo.br/j/cp/a/4xLrQkM5v36QqnQRP8ZmMPC/?lang=en>>.

<sup>140</sup> Ibidem, p.7.

*Assisted Learning*. Nitidamente, existem possibilidades de melhoria das capacidades dos atores educacionais e promoção de uma distribuição mais eficiente dos recursos. Outrossim, a IA pode ser usada para criar sistemas de dados inclusivos e de qualidade, que podem ajudar a monitorar e melhorar a equidade dos sistemas educacionais<sup>141</sup>.

Portanto, embora existam oportunidades para os sistemas de IA fortalecerem os propósitos educacionais e melhorarem algumas práticas educacionais, os autores alertam para os riscos associados à comercialização da educação, à possível manipulação e governança dos sistemas de IA e ao risco de reconfiguração do humano com valores obscuros. Eles concluem que são necessárias novas investigações para compreender completamente o impacto da IA na educação.

No artigo “Métodos de Resolução Digital de Controvérsias: Estado da Arte de suas Aplicações e Desafios”<sup>142</sup>, MOULIN explora o uso da tecnologia na prática jurídica através dos métodos de resolução digital de controvérsias, ou *ODRs (Online Dispute Resolution)*. O artigo visa apresentar uma visão geral sobre o conceito, as modalidades de ODR, suas principais aplicações práticas e os desafios para sua regulamentação.

Segundo o autor, existem duas modalidades principais de softwares ODR: os sistemas instrumentais e os principais. Os sistemas instrumentais são plataformas virtuais especializadas que facilitam a comunicação e o compartilhamento de informações entre as partes. Eles representam a abordagem mais simples da integração tecnológica à prática jurídica e incluem ferramentas de pesquisa de jurisprudência, aplicativos de mensagens e videoconferência, sistemas de gestão eletrônica de processos, programas de edição compartilhada de documentos, softwares de elaboração automatizada de textos jurídicos.

Os sistemas principais, por outro lado, são softwares que auxiliam na tomada de decisões, seja em um processo judicial ou extrajudicial. Eles são subdivididos em sistemas expert e de suporte. Os sistemas expert utilizam IA para analisar dados e fornecer soluções para o conflito, enquanto os sistemas de suporte auxiliam as partes a chegar a um acordo, mas não fornecem soluções prontas.

---

<sup>141</sup> Ibidem, p.14.

<sup>142</sup> MOULIN, Carolina Stange Azevedo. Métodos de resolução digital de controvérsias: estado da arte de suas aplicações e desafios. Revista Direito GV, v. 17, 2021. Disponível em: <<https://www.scielo.br/j/rdgv/a/K6Td7TJ6fcMtpyRDWVdzbPN/?lang=pt>>.

A IA pode ser utilizada em sistemas expert de ODR para analisar dados e fornecer soluções para o conflito, aumentando a capacidade de análise de dados e alcance de soluções para o conflito, tornando o processo mais eficiente e preciso<sup>143</sup>. O texto menciona dois exemplos concretos de sistemas de apoio à decisão que utilizam IA em ODR: o *Adjusted Winner*, aplicado no direito de família australiano, e o *UMCourt*, implementado na Universidade do Minho, em Portugal.

MOULIN também alerta para o fato de que a regulação transnacional dos mecanismos de ODR está sendo feita por entidades como a União Europeia (UE) e a Comissão das Nações Unidas para o Direito Comercial Internacional (UNCITRAL), bem como por entidades privadas, como a Internet Corporation for *Assigned Names and Numbers (ICANN)*<sup>144</sup>.

Por fim, o artigo conclui que os mecanismos de ODR têm um potencial muito maior do que o inicialmente previsto, sendo mais apropriados para a resolução de disputas oriundas de relações consumeristas eletrônicas massificadas, de baixo valor e transnacionais. Sistemas de apoio à decisão têm sido implementados com sucesso em áreas como direito de propriedade intelectual, família, sucessões e administrativo.

O estudo de ANDRADE, ROSA & PINTO, intitulado '*Legal tech: analytics, inteligência artificial e as novas perspectivas para a prática da advocacia privada*'<sup>145</sup>, explora a aplicação da inteligência artificial e da plataforma *Litigation Analytics* na advocacia privada. A pesquisa adota uma abordagem teórico-analítica para examinar a interseção entre a automação no Poder Judiciário brasileiro e a prática da advocacia privada.

Os autores argumentam que o uso de ferramentas inteligentes pode aprimorar consultas e otimizar o tempo gasto com pesquisas por jurisprudência, identificando rapidamente possíveis divergências, tendências decisórias de tribunais e argumentos compatíveis. Eles sugerem que a inteligência artificial pode promover maior

---

<sup>143</sup> Ibidem, p.20.

<sup>144</sup> Ibidem, p.21.

<sup>145</sup> ANDRADE, Mariana Dionísio de; ROSA, Beatriz de Castro; PINTO, Eduardo Régis Girão de Castro. Legal tech: analytics, inteligência artificial e as novas perspectivas para a prática da advocacia privada. Revista Direito GV, v. 16, 2020. Disponível em:<<https://www.scielo.br/j/rdgv/a/xL839bvvvK4QgvZfxwR6b4J/?lang=pt>>.

previsibilidade na advocacia privada, aprofundando um campo de pesquisa ainda incipiente no Brasil.

Além disso, os autores enfatizam que as novas tecnologias permitem que os advogados trabalhem com mais eficiência, expandam suas áreas de especialização e inaugurem uma nova era na condução das demandas - mais eficiente e transparente<sup>146</sup>. O estudo também destaca que a automação no cotidiano do Poder Judiciário brasileiro pode potencializar consultas e otimizar o tempo gasto com pesquisas por jurisprudência. Os avanços tecnológicos fornecem as ferramentas necessárias para que os escritórios de advocacia se mantenham atualizados na evolução técnica da informação jurídica<sup>147</sup>.

A conclusão do estudo é que softwares e plataformas inteligentes, derivados da jurimetria em uma compreensão algorítmica do Direito, estão irreversivelmente presentes no cenário da análise jurídica. Eles tendem a impactar vários aspectos do trabalho dos advogados, incluindo tarefas que historicamente dependiam do esforço humano especializado, como a tentativa de previsão de resultados judiciais. Portanto, a automação pode impactar positivamente a atuação profissional dos advogados, permitindo que eles trabalhem com mais eficiência e precisão, além de possibilitar a ampliação de suas áreas de especialização.

Em seguida, temos ABRAHAM & CATARINO na obra "O uso da inteligência artificial na aplicação do direito público – o caso especial da cobrança dos créditos tributários – um estudo objetivado nos casos brasileiro e português"<sup>148</sup>. Esse é um estudo abrangente sobre o impacto da inteligência artificial (IA) no campo do direito, com foco especial na cobrança de créditos tributários, tanto no âmbito judicial quanto na fiscalização tributária. O estudo apresenta exemplos práticos de como a IA tem sido aplicada em questões jurídicas e fiscais em diferentes partes do mundo, incluindo o Brasil e Portugal. Além disso, o artigo discute os desafios éticos que surgem com o uso da IA em relação à moralidade humana e algorítmica.

---

<sup>146</sup> Ibidem, p.17.

<sup>147</sup> Ibidem, p.17.

<sup>148</sup> ABRAHAM, Marcus; CATARINO, João Ricardo. O uso da inteligência artificial na aplicação do direito público: o caso especial da cobrança dos créditos tributários-um estudo objetivado nos casos brasileiro e português. e-Pública: Revista Eletrônica de Direito Público, v. 6, n. 2, p. 188-219, 2019. Disponível em:<<https://e-publica.pt/api/v1/articles/34335-o-uso-da-inteligencia-artificial-na-aplicacao-do-direito-publico-o-caso-especial-da-cobranca-dos-creditos-tributarios-um-estudo-objetivado-nos-cas.pdf>>.



Entre as formas pelas quais a inteligência artificial pode ser usada para melhorar a cobrança de créditos tributários, destaca-se a localização dos devedores e seus bens, sugestões de medidas de cobrança mais adequadas diante das circunstâncias fáticas identificadas, bem como a intermediação de uma solução amigável antes do ajuizamento da ação<sup>149</sup>. Igualmente, considera que a IA pode ser usada para analisar grandes quantidades de dados fiscais e jurídicos, identificar padrões e tendências e fornecer *insights* valiosos para os órgãos de fiscalização tributária<sup>150</sup>.

Entretanto, o estudo destaca alguns desafios e limitações da aplicação da IA no Direito Público. Um dos principais desafios é garantir que a IA seja usada de maneira ética e justa, sem violar os direitos humanos ou discriminar grupos vulneráveis<sup>151</sup>. Existe também limitações práticas, como a necessidade de grandes quantidades de dados de alta qualidade para treinar os algoritmos de aprendizado de máquina; bem como também garantir a transparência e a explicabilidade dos sistemas de IA, para que os usuários possam entender como as decisões são tomadas e contestá-las se necessário.

Além disso, os autores entendem que a IA não deve substituir completamente a tomada de decisão humana do Direito Público. Sendo assim, os profissionais do Direito precisam estar preparados para trabalhar com a tecnologia de maneira eficaz e responsável. Por fim, apesar de o estudo discutir as experiências brasileira e portuguesa com IA na cobrança de impostos, mas não faz uma comparação direta entre elas.

Ainda assim, o estudo destaca que ambos os países têm enfrentado desafios semelhantes na aplicação da IA no Direito Público, como a necessidade de garantir a ética e a transparência dos sistemas de IA, assim como também a importância de treinar os profissionais do Direito para trabalhar com a tecnologia de maneira eficaz. ABRAHAM & CATARINO concluem com uma reflexão sobre o futuro das máquinas e sua relação com os seres humanos, lembrando o relato bíblico da Torre de Babel e o dilema entre o criador e a criatura. Eles enfatizam que apenas o tempo pode responder aos rumos deste "admirável mundo novo"<sup>152</sup>.

Sob outra perspectiva de análise, temos CARDONA no artigo intitulado "*La Mediación Mercantil Ante los Nuevos Retos de la Inteligencia Artificial: Una Visión*

---

<sup>149</sup> Ibidem, p. 199.

<sup>150</sup> Ibidem, p. 191.

<sup>151</sup> Ibidem, p. 196.

<sup>152</sup> Ibidem, p. 216.

*Comparada Entre los Ordenamientos Jurídicos Español e Italiano*”<sup>153</sup>. O artigo faz uma análise sobre a aplicação de ferramentas baseadas em inteligência artificial e big data no processo de mediação, especialmente em conflitos comerciais, comparando as legislações da Itália e da Espanha.

Através da análise doutrinária e jurisprudencial, conclui-se que a incorporação da inteligência artificial na mediação pode fornecer uma ferramenta adicional de suporte ao mediador, sem substituí-lo. Inclusive, o autor entende que o uso da inteligência artificial pode ajudar na seleção de mediadores mais adequados para um caso específico e na avaliação da conveniência de um processo de mediação. No entanto, é necessário garantir o correto funcionamento dessas ferramentas.

Destarte, dado o fato de que a mediação comercial tem se tornado cada vez mais comum no ambiente empresarial e jurídico, CARDONA compreende que com a evolução tecnológica, a mediação tem passado por três grandes etapas: a utilização de tecnologia meramente instrumental, a incorporação de sistemas algorítmicos e a utilização da inteligência artificial através de árbitros e mediadores robôs<sup>154</sup>. A incorporação da inteligência artificial pode ser abordada sob duas perspectivas: assistencial e decisória. Na primeira, a inteligência artificial seria uma ferramenta de suporte para o mediador, auxiliando na seleção de mediadores e na avaliação da conveniência de um processo de mediação. Já na segunda perspectiva, a inteligência artificial seria concebida como decisória, em que robôs resolveriam a controvérsia levantada.

Outro artigo que inova na escolha do tema é o de LIMBERGER, GIANNAKOS & SZINVELSKI, intitulado ‘*Can Judges be Replaced by Machines? The Brazilian Case*’<sup>155</sup>. Muito embora não estabeleça as condições necessárias para a substituição de

---

<sup>153</sup> CARDONA, Noemí Jiménez. La Mediación Mercantil Ante los Nuevos Retos de la Inteligencia Artificial: Una Visión Comparada Entre los Ordenamientos Jurídicos Español e Italiano. Revista Internacional Consinter de Direito, p. 493-505, 2022. Disponível: <<https://revistaconsinter.com/index.php/ojs/article/view/24>>.

<sup>154</sup> [...] *De hecho, en orden a la imbricación de la mediación con los recursos electrónicos, que ha llegado para quedarse, es factible individualizar tres grandes etapas que marcan la evolución tecnológica de nuestros días: una primera, en que la utilización de la tecnología es meramente instrumental, caracterizada por la aparición de sistemas de arbitraje y mediación en línea; una segunda, presidida por la incorporación de sistemas algorítmicos que permiten decidir, por ejemplo, acerca de la conveniencia o no, en el caso concreto, de desarrollar una actividad de negociación o seguir una mediación; y una tercera, más reciente en el tiempo, en la que ya empieza a hablarse de la incorporación de la inteligencia artificial mediante sistemas computacionales y estructuras de cómputo, con árbitros y mediadores robot que dirigen y resuelven, por sí mismos, la controversia planteada* (IBIDEM, p. 501).

<sup>155</sup> LIMBERGER, Têmis; GIANNAKOS, Demétrio Beck da Silva; SZINVELSKI, Martín M. Can Judges be Replaced by Machines? The Brazilian Case. Mexican law review, v. 14, n. 2, p. 53-81, 2022. Disponível: <[https://www.scielo.org.mx/scielo.php?pid=S1870-05782022000100053&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S1870-05782022000100053&script=sci_arttext)>.

juízes por máquinas no judiciário brasileiro, destaca a existência de ferramentas para a previsão de resultados judiciais usando IA.

O estudo sugere que a IA no sistema judicial do Brasil oferece uma oportunidade para reduzir os custos do Poder Judiciário, permitindo que tarefas judiciais relativamente simples e frequentes sejam resolvidas mais rapidamente, mantendo a responsabilidade final pela tomada de decisão judicial com o juiz<sup>156</sup>.

Os autores discutem os potenciais benefícios e desvantagens de substituir os juízes por máquinas, incluindo a redução dos custos do poder judiciário, o aumento da eficiência do processo judicial e a redução do potencial de erro humano. Eles destacam o ‘Victor’, uma IA criada pelo Supremo Tribunal Federal (STF), que lê todos os recursos extraordinários que chegam ao STF e identifica quais estão relacionados a temas recorrentes de repercussão geral.

No entanto, o estudo também considera as potenciais desvantagens de substituir os juízes por máquinas, incluindo a possibilidade de a IA não ser capaz de replicar totalmente o processo de tomada de decisão humana, levando a decisões tendenciosas ou injustas. Há também preocupações sobre a transparência e responsabilidade da tomada de decisão da IA, o potencial de a IA ser hackeada ou manipulada, e questões éticas sobre o uso da IA no sistema judicial, incluindo o papel do julgamento humano e o potencial da IA perpetuar os desequilíbrios de poder existentes<sup>157</sup>.

O artigo conclui que a introdução gradual da IA na rotina forense é recomendada, mas os profissionais do direito devem se adaptar à nova realidade. No entanto, a ideia de substituição de juízes por IA ainda é ambígua devido às incertezas em torno das aplicações e das bases de dados usadas pela máquina e à dificuldade de formalizar as regras, princípios e valores éticos que informam toda a tomada de decisão judicial<sup>158</sup>

---

<sup>156</sup> [...] *Our conclusion is that AI will reduce the cost of the judicial machinery by allowing many relatively simple and frequently occurring judicial tasks to be resolved more quickly, while leaving ultimate responsibility for judicial decision-making with the judge. Replacement of the judge by automated algorithmic tools would require specific, necessary conditions, even if it were to promote efficiency and cost reduction. These conditions are (i) the adequacy and efficiency of AI generated results, and (ii) the ability of computerized routines to adapt and improve the quality of their application of the law by accommodating and incorporating human corrections of the automated decisions. We also diagnose the philosophical issue that replacing judges with machines might signal a possible “end of interpretation,” even though ongoing human inspection and correction of the computerized systems would be necessary* (Ibidem, p.53).

<sup>157</sup> Ibidem, p.67.

<sup>158</sup> Ibidem, p.81.

No estudo “Identificação de evasão fiscal utilizando dados abertos e inteligência artificial”<sup>159</sup>, XAVIER, PIRES, MARQUES & SOARES propõem uma solução inovadora para detectar a evasão fiscal. Utilizando dados abertos e públicos de diferentes bancos de dados governamentais, eles desenvolveram três modelos de inteligência artificial que, após aperfeiçoamentos, atingiram uma precisão superior a 98% na previsão do perfil dos sonegadores.

A evasão fiscal, que corresponde a 8% do PIB brasileiro, tem um impacto significativo na economia do país, reduzindo a arrecadação de impostos e contribuições que poderiam ser investidos em áreas como saúde, educação e infraestrutura<sup>160</sup>. A solução proposta utiliza dados abertos disponíveis no Portal dados.gov.br, embora o artigo não especifique quais dados foram utilizados.

Os modelos de Random Forest, Redes Neurais e Grafos foram treinados com dados abertos da Receita Federal e do Conselho Administrativo Tributário do Estado de Goiás, entre outros. Após melhorias, o modelo clássico Random Forest obteve uma melhora de quase 8% de acurácia e foi escolhido para a construção da solução final<sup>161</sup>.

A contribuição científica deste trabalho reside em mostrar a viabilidade do uso de dados públicos para abordar o problema da evasão fiscal. Os resultados superaram as expectativas dos autores e auditores fiscais, e o sistema já está em uso.

O artigo científico intitulado "*Enhancing the Value of Patents as Corporate Assets in South Africa: How can Artificial Intelligence (AI) Assist?*"<sup>162</sup> de NDLOVU explora como a inteligência artificial (IA) pode melhorar a qualidade das patentes na África do Sul, transformando-as em ativos corporativos valiosos. A IA pode ajudar a mitigar as fraquezas na legislação de patentes sul-africana, garantindo o arquivamento de patentes de qualidade que atendam aos critérios de patenteabilidade prescritos e realizar buscas de arte anterior de forma mais rápida e precisa.

---

<sup>159</sup> XAVIER, Otávio Calaça et al. Identificação de evasão fiscal utilizando dados abertos e inteligência artificial. Revista de Administração Pública, v. 56, p. 426-440, 2022. Disponível:<<https://www.scielo.br/j/rap/a/5q38f9RdbQYSrZXF8zfDJqv/?format=html&lang=pt&stop=next>>

<sup>160</sup> Ibidem, p.426.

<sup>161</sup> Ibidem, p.437.

<sup>162</sup> NDLOVU, Lonias. Enhancing the value of patents as corporate assets in South Africa: How can artificial intelligence (AI) assist?. Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, v. 24, n. 1, 2021. Disponível em:<<https://www.ajol.info/index.php/pej/article/view/235800>>.

A IA tem o potencial de melhorar a qualidade das patentes no contexto sul-africano, permitindo que resultados precisos de busca de arte anterior sejam obtidos mais rapidamente. A integração da IA nos sistemas de patentes, especialmente nas buscas e exames de patentes, pode ajudar a eliminar muitos dos problemas identificados no sistema de patentes sul-africano. No entanto, a posição estatutária atual na África do Sul não menciona explicitamente buscas de arte anterior gerados por IA.

NDLOVU considera que a abordagem proposta é positiva, envolvendo o uso de tecnologias de IA, como aprendizado de máquina, processamento de linguagem natural, aprendizado profundo e análise de PI, para fortalecer o sistema de PI da África do Sul e criar valor de ativos para as empresas. Apesar de a IA ainda não ter sido amplamente explorada no contexto da propriedade intelectual, há um nítido e crescente interesse na relação entre patentes e IA. É necessário um estudo multidisciplinar para explorar novas maneiras de utilizar a IA para melhorar o cenário de propriedade intelectual na África do Sul e em outros países<sup>163</sup>.

Por outro lado, artigo intitulado "A Regulação e o Uso da Inteligência Artificial e da Tecnologia 5G para Combater Crimes Cibernéticos e Financeiros em Bancos Sul-Africanos"<sup>164</sup>, de CHITIMIRA & NCUBE explora o uso da Inteligência Artificial (IA) e da tecnologia 5G para combater crimes financeiros e cibernéticos em bancos sul-africanos. A IA, em suas diversas formas - funcional, interativa, textual, visual e analítica - é vista como uma ferramenta potencial para melhorar a competitividade e expansão dos negócios das instituições financeiras.

A IA e o 5G são apresentados como meios eficazes de detectar e combater crimes cibernéticos. A IA pode identificar atividades suspeitas, enquanto o 5G permite uma rápida conexão à internet para monitoramento e resposta em tempo real<sup>165</sup>. No entanto, o artigo também alerta para o potencial abuso dessas tecnologias por criminosos para cometer crimes financeiros, como lavagem de dinheiro e negociação com informações privilegiadas.

---

<sup>163</sup> Ibidem, p. 4.

<sup>164</sup> CHITIMIRA, Howard; NCUBE, Princess. The regulation and use of artificial intelligence and 5G technology to combat cybercrime and financial crime in South African banks. Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, v. 24, n. 1, 2021. Disponível em: <<https://www.ajol.info/index.php/pelj/article/view/235606>>.

<sup>165</sup> Ibidem. p. 3-4.

Diante disso, o artigo defende a adoção de tecnologias como a 5G e a IA pelos bancos para aprimorar a detecção e o combate a crimes cibernéticos na África do Sul. Além disso, enfatiza a necessidade de uma revisão cuidadosa do Projeto de Lei de Crimes Cibernéticos pelos formuladores de políticas, incluindo o uso da IA e da 5G na detecção e combate à cibercriminalidade<sup>166</sup>.

Por fim, o artigo se insere em um contexto mais amplo de discussões sobre a relação entre a IA e o direito, abordando temas como ficção científica, espiritualidade e a possibilidade de um novo sujeito de direito associado à IA.

SÁNCHEZ MORALEZ, em obra diversa intitulada "Ciência Ficção como Fonte de Princípios Jurídicos para Regulação da Inteligência Artificial"<sup>167</sup>, explora a ficção científica como fonte de princípios jurídicos para regular a Inteligência Artificial (IA). Com o rápido avanço da tecnologia, o direito precisa se adaptar e reavaliar seus princípios. A ficção científica pode ser uma fonte de inspiração para prever possíveis dilemas jurídicos relacionados à IA, permitindo antecipar conflitos éticos e jurídicos e buscar soluções<sup>168</sup>.

Os autores argumentam que o Direito deve reavaliar suas práticas e princípios para se adequar às mudanças que a IA está gerando em nossa sociedade. Propõem que os princípios jurídicos devam ser adaptados à nova realidade tecnológica, pois a contextualidade e a praticidade são características compartilhadas por todos os elementos da IA.

O artigo sugere que é tarefa do Direito resolver as disjuntivas que surgem diante das decisões tomadas pela IA. Isso requer uma nova interpretação dos princípios jurídicos que permita a continuação de sua função, fazendo inferências e resolvendo ambiguidades diante dos casos concretos que se apresentam. A filosofia da experiência jurídica sugere que a ficção científica pode ser uma fonte valiosa para analisar dilemas teóricos sobre a IA que afetam a esfera jurídica humana<sup>169</sup>.

---

<sup>166</sup> Ibidem, p.23.

<sup>167</sup> SÁNCHEZ MORALES, Santiago. *Ciencia Ficción como Fuente de Principios Jurídicos para regular la Inteligencia Artificial*. Revista IUS, v. 15, n. 48, p. 55-76, 2021. Disponível em: <[https://www.scielo.org.mx/scielo.php?pid=S1870-21472021000200055&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S1870-21472021000200055&script=sci_arttext)>.

<sup>168</sup> Ibidem, p. 55.

<sup>169</sup> Ibidem, p. 65.

Por fim, o artigo conclui que a ficção científica pode ser uma ferramenta valiosa para o direito diante do avanço tecnológico, permitindo aos juristas encontrar soluções para casos concretos e adaptar a profissão às mudanças em curso. A experiência jurídica será a que limitará, incentivará, governará ou corrigirá o que significa direito no futuro<sup>170</sup>.

CHÁVEZ VALDIVIA, em sua obra “*No es solo un robot: consideraciones en torno a una nueva personalidad jurídica y el redimensionamiento de las relaciones interpersonales*”<sup>171</sup>, explora a necessidade de uma nova categoria de sujeitos jurídicos para lidar com a emergente geração de robôs sociais. Estes robôs, projetados para interagir de forma personalizada e humana, estão provocando uma reavaliação das formas tradicionais do Direito e trazendo à tona mudanças sociais significativas.

A concessão de personalidade jurídica a robôs sociais implicaria em um processo de transição que variaria entre países e níveis de aceitação. Os robôs sociais têm o potencial de melhorar nossas relações interpessoais, pois são capazes de se comunicar, interagir, compreender e se relacionar conosco de forma pessoal. Além disso, eles devem ser capazes de se entender e nos entender em termos sociais, e nós, por nossa vez, devemos poder entendê-los nos mesmos termos<sup>172</sup>.

A autora também menciona medidas tomadas para regular o uso de robôs, como os “Princípios da Robótica: regulando robôs no mundo real” publicados pelo EPSRC no Reino Unido, e as normas para o projeto e uso de robôs apresentadas pela Comissão de Ética da Robótica e da Inteligência Artificial nos Estados Unidos. Em suma, Chávez Valdivia defende que a criação de um novo ramo do direito, o “Direito dos Andróides”, é essencial diante das qualidades únicas dos robôs e da inteligência artificial e suas implicações no campo jurídico<sup>173</sup>.

Encerrando, temos o artigo “*Cyborgs and the future of the human spirit*”<sup>174</sup>, onde MATTHEE explora a influência dos ciborgues, máquinas pensantes, na espiritualidade

---

<sup>170</sup> Ibidem, p. 76.

<sup>171</sup> CHÁVEZ VALDIVIA, Ana Karin. *No es solo un robot: consideraciones en torno a una nueva personalidad jurídica y el redimensionamiento de las relaciones interpersonales*. *Ius et Praxis*, v. 26, n. 2, p. 55-77, 2020. Disponível em: <[https://www.scielo.cl/scielo.php?pid=S0718-00122020000200055&script=sci\\_arttext&tlng=pt](https://www.scielo.cl/scielo.php?pid=S0718-00122020000200055&script=sci_arttext&tlng=pt)>.

<sup>172</sup> Ibidem, p. 69.

<sup>173</sup> Ibidem, p. 73.

<sup>174</sup> MATTHEE, Machdel. *Cyborgs and the future of the human spirit*. *Tydskrif vir Geesteswetenskappe*, v. 53, n. 4, p. 530-557, 2013. Disponível: <[http://www.scielo.org.za/scielo.php?pid=S0041-47512013000400005&script=sci\\_arttext](http://www.scielo.org.za/scielo.php?pid=S0041-47512013000400005&script=sci_arttext)>.

humana. A autora destaca a visão de Kurzweil, que vê os ciborgues como não apenas uma simulação da cognição humana, mas também da consciência e da espiritualidade, o que poderia levar a uma melhora radical da consciência e espiritualidade humana.

MATTHEE concebe o ciborgue como uma máquina projetada para imitar características e capacidades humanas, com potencial para afetar nosso futuro de várias maneiras. Isso inclui a superação de limitações humanas, como a morte, e o aprimoramento de nossa capacidade de comunicação e aprendizagem. A autora argumenta que essas máquinas podem nos forçar a nos redefinir à luz das novas possibilidades que oferecem.

O futuro dos ciborgues, segundo MATTHEE, é incerto, com possibilidades que vão desde um super-humano que exercerá a pura vontade de poder até uma pessoa com total liberdade e ingenuidade. No entanto, ela alerta para o perigo do desumano que pode surgir do crescimento e progresso tecnológico e enfatiza a importância de uma espiritualidade humana viva e rica para enfrentar esse desafio<sup>175</sup>.

Por fim, MATTHEE ressalta que os seres humanos, como criadores da tecnologia, devem assumir a responsabilidade de sua liberdade para se defender contra sua desumanidade. Ela argumenta que a suposição de que a cognição pode ser mecanizada leva à desencarnação da inteligência e do pensamento, tornando os seres humanos estranhos a si mesmos e aos outros. A autora encerra com um apelo à luta pelo reencantamento do espírito humano e destaca a importância das ciências sociais e dos líderes na busca por soluções para os desafios impostos pela tecnologia.

Assim sendo, concluindo a análise dessa série de artigos que examina o impacto da inteligência artificial no âmbito jurídico, podemos identificar que em boa parte dos artigos, principalmente nos primeiros, os autores enfatizam a importância de se estabelecer regulamentações éticas e precisas para garantir o uso adequado da tecnologia e proteger os direitos fundamentais das pessoas.

Igualmente, entre os artigos analisados neste texto, alguns apresentam diversas perspectivas sobre o impacto da inteligência artificial no campo do direito e da educação. Enquanto alguns artigos se concentram em questões específicas, como a aplicação da IA na tomada de decisões judiciais e na resolução de disputas online, outros discutem os

---

<sup>175</sup> Ibidem, p. 532-533.



aspectos positivos e negativos da IA na educação.

No entanto, todos os artigos enfatizam a importância de equilibrar o potencial da tecnologia com os valores humanísticos, a autonomia e a equidade. Nesse sentido, podemos verificar que as pesquisas apresentadas mostram que a aplicação da IA pode fornecer informações valiosas para a tomada de decisões importantes no campo jurídico, mas ressaltam a importância da responsabilidade humana na tomada de decisões.

Incrivelmente, podemos ver que já é aplicada a inteligência artificial (IA) no judiciário brasileiro, tendo papel decisivo na detecção de evasão fiscal. Além disso, também proporcionou um melhor aproveitamento das patentes e mais eficiência na luta contra crimes financeiros e cibernéticos na África do Sul. Em resumo, a IA e o desenvolvimento de sua explicabilidade, pode ser uma ferramenta poderosa para melhorar a eficiência e o senso de justiça no campo jurídico, muito embora, deva ser usada de maneira responsável e com consciência dos impactos em potencial.

## **4- EXPERIMENTOS**

### **4.1 Juizado Especial Cível**

De acordo com a CF/88, o Judiciário Brasileiro é composto por: (a) Supremo Tribunal Federal (STF); (b) Superior Tribunal de Justiça (STJ); (c) Tribunais Regionais Federais (TRFs) e juízes federais; (d) Tribunal Superior do Trabalho (TST), Tribunais Regionais do Trabalho (TRTs) e juízes do trabalho; (e) Tribunal Superior Eleitoral (TSE), Tribunais Regionais Eleitorais (TREs) e juízes eleitorais; (f) Superior Tribunal Militar (STM) e juízes militares; (g) Tribunais de Justiça Estaduais (TJs) e juízes estaduais. Além disso, os Tribunais Federais e Estaduais podem, dentro de sua jurisdição, criar os Juizados Especiais (JECs e JEFs), que são responsáveis por julgar casos locais menos complexos.<sup>176</sup> A Figura 6 ilustra essa organização.

### **Figura 13 Ilustração do treinamento de IA explicável (XAI)**

---

<sup>176</sup> BRASIL. [Constituição (1988)]. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Presidente da República, [2016]. Disponível em:<[https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm)>

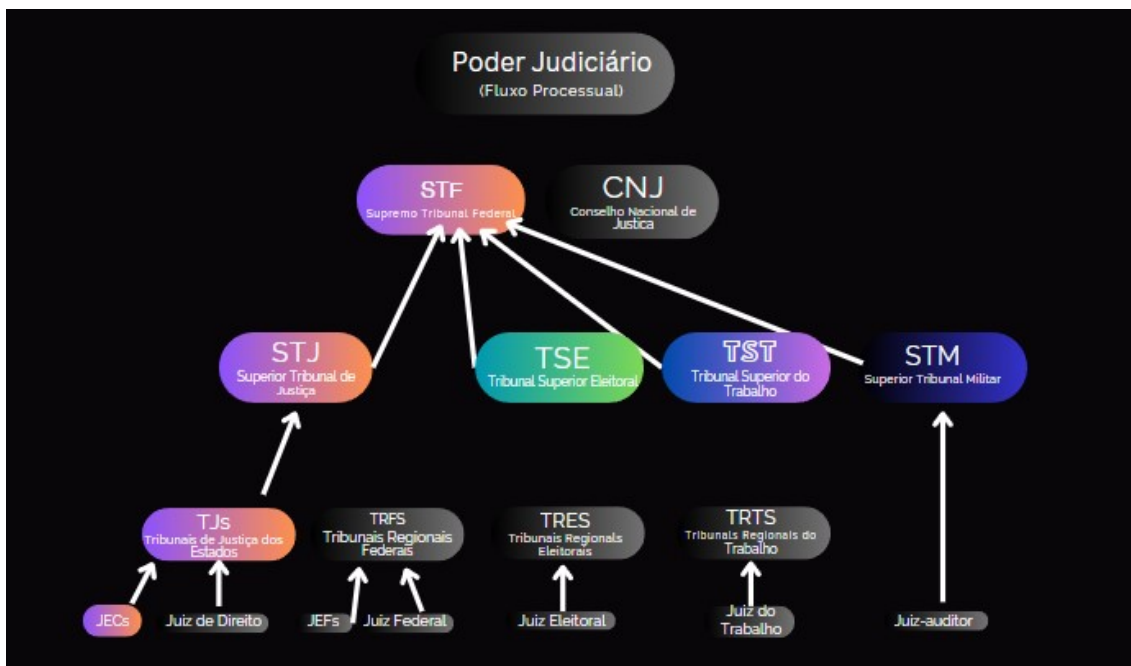


Figura 6 - Organograma do Judiciário Brasileiro.

Os Juizados Especiais Cíveis (JECs) são órgãos do Poder Judiciário Estadual que foram criados em 1984 pela Lei nº 7.244<sup>177</sup>, com o objetivo de facilitar o acesso do cidadão à justiça, especialmente nas causas de menor complexidade. Os JECs oferecem aos cidadãos a possibilidade de resolverem seus problemas sem custas processuais, com procedimentos simplificados e com incentivo à conciliação entre as partes litigantes. Os JECs atendem às demandas jurídicas do cidadão comum, que está envolvido em conflitos diários e menores, como questões de consumo, família e vizinhança.

Segundo um estudo realizado pelo Conselho Nacional de Justiça (CNJ) em 2020, os JECs receberam mais de 7 milhões de processos em 2019, sendo a maioria de competência cível. O tempo médio de tramitação de um processo nos JECs foi de um ano e seis meses, enquanto nos demais processos de 1º Grau foi de três anos e sete meses. O índice de conciliação nos JECs estaduais foi de 21%, enquanto nos JECs federais foi de 18%. O estudo também apontou alguns desafios para o funcionamento dos JECs, como a falta de estrutura adequada, a sobrecarga de trabalho, a baixa qualidade das petições

<sup>177</sup> BRASIL. LEI Nº 7.244, DE 7 DE NOVEMBRO DE 1984. Dispõe sobre a criação e o funcionamento do Juizado Especial de Pequenas Causas. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/leis/1980-1988/l7244.htm#:~:text=L7244&text=LEI%20N%C2%BA%207.244%2C%20DE%207%20DE%20NOVEMBRO%20DE%201984.&text=Disp%C3%B5e%20sobre%20a%20cria%C3%A7%C3%A3o%20e,Juizado%20Especial%20de%20Pequenas%20Causas.](https://www.planalto.gov.br/ccivil_03/leis/1980-1988/l7244.htm#:~:text=L7244&text=LEI%20N%C2%BA%207.244%2C%20DE%207%20DE%20NOVEMBRO%20DE%201984.&text=Disp%C3%B5e%20sobre%20a%20cria%C3%A7%C3%A3o%20e,Juizado%20Especial%20de%20Pequenas%20Causas.)>.

iniciais, a resistência à conciliação, a judicialização excessiva e a falta de padronização dos procedimentos<sup>178</sup>.

As partes não precisam de um advogado para entrar nos JECs. Quando uma parte entra no Juizado Especial, é emitida uma convocação para a outra com a data da audiência de conciliação. Ambas as partes são obrigadas a comparecer. A conciliação pode ser conduzida por: (1) o juiz titular (que é responsável pelo tribunal e julga os processos); ou (2) o juiz leigo (uma pessoa que tem diploma de direito, mas não é investida no cargo); ou (3) o conciliador (uma pessoa que tem diploma de direito e é qualificada para conciliar). Uma vez aberta a sessão, quem a conduz deve explicar às partes as vantagens da conciliação, bem como os riscos e consequências em prosseguir com o processo<sup>179</sup>.

Por fim, antes de darmos continuidade, importante destacar que esse é um experimento de explicabilidade com IA no Direito, que dá continuidade aos esforços empregados por pela Profa. Dra Isabela Sabo<sup>180</sup> e Prof. Dr. Aires Rover. Isso porque, ao longo de sua tese de doutorado defende justamente o emprego de IA para auxiliar nos processos conciliatórios do JEC/UFSC, sem, contudo, ter tido tempo hábil para atingir a explicabilidade almejada nesse trabalho. A unidade que estudaremos é o Juizado Especial Cível localizado na Universidade Federal de Santa Catarina (JEC/UFSC).

## 4.2 Base de dados e estruturação

Os experimentos com *XAI* possuem como objetivo geral avaliar maneiras de proporcionar explicabilidade de predições de IA na representação de conhecimento jurídico. Nesse sentido, cabe aqui mencionar os procedimentos e técnicas utilizadas, para melhor compreensão do leitor.

---

<sup>178</sup> BANDEIRA, R. Estudo revela realidade e desafios dos Juizados Especiais. CNJ. 2020. Disponível em: <<https://www.cnj.jus.br/estudo-revela-realidade-e-desafios-dos-juizados-especiais/>>.

<sup>179</sup> BRASIL. L.9099. LEI Nº 9.099, DE 26 DE SETEMBRO DE 1995. Dispõe sobre os Juizados Especiais Cíveis e Criminais e dá outras providências. Disponível em: <[<sup>180</sup> SABO, Isabela Cristina et al. A machine learning-based model for judgement results prediction and support in Brazilian Special Court? s conciliation hearings. 2022. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/241026>>.](https://www.planalto.gov.br/ccivil_03/leis/19099.htm#:~:text=L9099&text=LEI%20N%C2%BA%209.099%2C%20DE%2026%20DE%20SETEMBRO%20DE%201995.&text=Disp%C3%B5e%20sobre%20os%20Juizados%20Especiais%20C%C3%ADveis%20e%20Criminais%20e%20d%C3%A1%20outras%20provid%C3%Aancias.></a>></p></div><div data-bbox=)

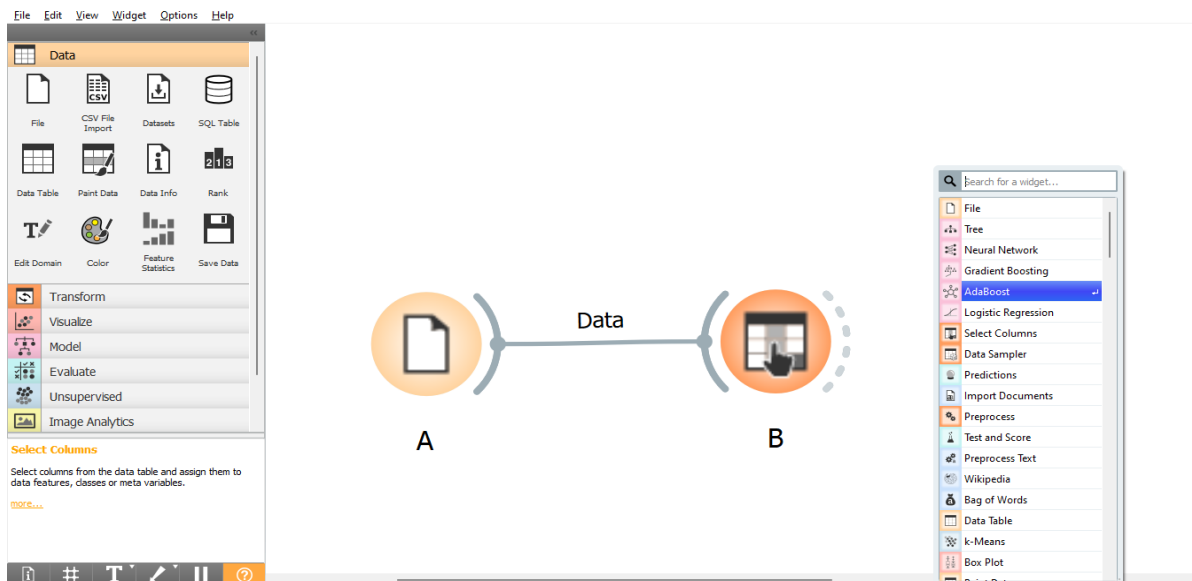
Inicialmente, adotou-se o procedimento de pesquisa descritiva e explicativa mediante a confecção da revisão bibliográfica sobre IA, bem como sobre sua explicabilidade, em que se fez o uso da técnica de análise bibliográfica e documental de produções científicas sólidas sobre IA e *XAI*. Os mesmos procedimentos e técnicas de pesquisa são adotados para o capítulo de Estado da Arte, no qual se levantam dados relativos à IA e *XAI*, através de plataformas como *Web of Science*, *Scopus* e *Scielo*.

A posteriori, faz-se uso do procedimento de pesquisa instrumental e avaliativa através de experimentação com a aplicação de modelos explicativos de IA através do método *SHAP* as previsões de IA no Judiciário, mediante uso de técnicas de análise documental e estatística. Isso se aplica tanto a fase de estruturação dos dados para aprendizado de máquina, como das análises de previsões e extração de sua explicabilidade.

Especialmente, nesse último ponto é válido mencionar que a principal ferramenta utilizada para trabalhar com algoritmos de aprendizado de máquina foi o *Orange Canvas*, todavia, isso não descarta a possibilidade de obtenção dos mesmos resultados ou até melhores com ferramentas distintas.

O verdadeiro motivo da escolha de tal ferramenta se deve pela a facilidade de interação devido ao caráter intuitivo, bem como também se tratar de *software Open Source*. Basicamente, cada componente representa uma etapa na análise de dados, como leitura e pré-processamento de dados, visualização de dados, construção de modelos de aprendizado de máquina e avaliação de desempenho.

**Figura 14 Ilustração do *Orange Canvas***



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Destarte, os usuários podem importar conjuntos de dados em vários formatos, como planilhas ou arquivos *CSV*, e então realizar diversas operações, como filtragem, transformação, seleção de recursos e criação de gráficos. Uma vez que o fluxo de trabalho tenha sido construído, o *Orange Canvas* permite executar a análise de dados e visualizar os resultados de forma interativa.

Portanto, podemos entender o *Orange Canvas* como uma ferramenta versátil e acessível para análise de dados e aprendizado de máquina, que permite aos usuários explorar dados, construir modelos preditivos e obter insights valiosos sem a necessidade de conhecimento avançado em programação.

Outrossim, os experimentos tiveram como base um conjunto de 390 (trezentos e noventa) processos relativos a falhas de companhias aéreas entre os anos de 2014 (dois mil e quatorze) e 2020 (dois mil e vinte). Esses processos foram obtidos mediante concessão autorizada dos registros do JEC e serviram como ponto de partida para a análise. Outrossim, dentre os 390 (trezentos e noventa) processos, 108 (cento e oito) foram selecionados para compor a base de teste, enquanto que 282 (duzentos e oitenta e dois) restantes foram utilizados como base de aprendizado de máquina.

Destarte, para a obtenção da explicabilidade da inteligência artificial (*XAI*), foram aplicados quatro modelos distintos de aprendizado de máquina em plataforma *Open Source* chamada *Orange Canvas*. Os modelos utilizados foram: a) *Adaboost*; b) *Gradient*

*Boost*; c) *Neural Network*; d) *Logistic Regression*. Esses modelos foram escolhidos devido à sua relevância e eficácia em tarefas de classificação e previsão. Inclusive, cada um deles foi treinado utilizando a base de aprendizado de máquina composta pelos 282 (duzentos e oitenta e dois) processos relacionados a falhas de companhias aéreas estruturados em arquivos CSV.

Para além dos modelos de aprendizado de máquina no campo das previsões, foram utilizadas ferramentas de *Explain Model* e *Explain Prediction* no *Orange Canvas*, que compõe um modelo explicativo conhecido com método *SHAP*, em homenagem a Lloyd Shapley, economista ganhador do Nobel em 2012, no contexto da teoria dos jogos. Brevemente, com o *Explain Model*, é possível analisar as características mais importantes para a tomada de decisão do modelo, enquanto o *Explain Prediction* permite entender os fatores que influenciam uma previsão em específico.

#### **4.3 Etapas do aprendizado de máquina e experimento com método *SHAP***

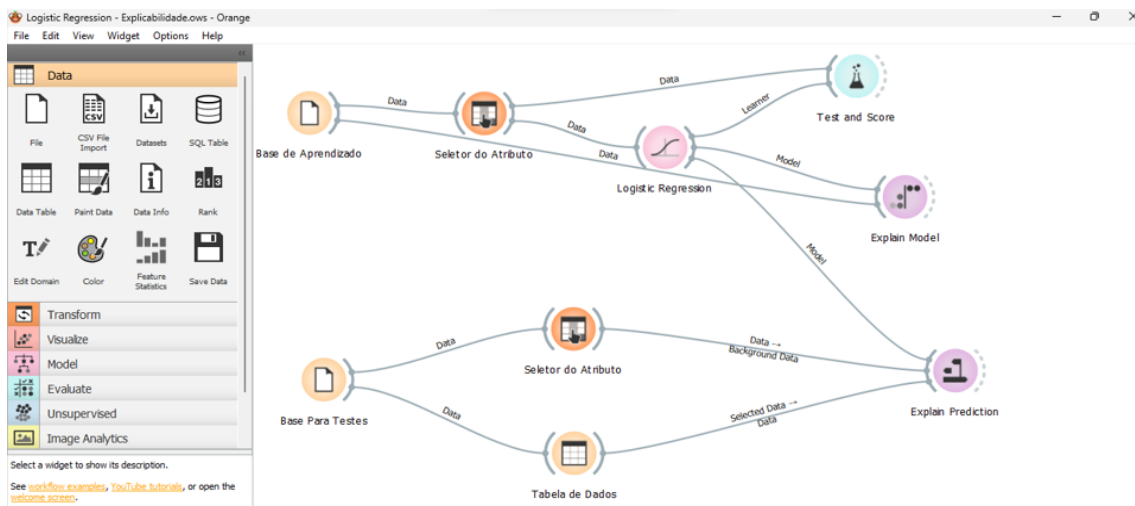
De fato, no contexto do Direito, é essencial garantir a segurança das decisões judiciais, assim como a sua eficácia. Por conta disso, a IA tem se tornado cada vez mais presente no Judiciário Brasileiro, trazendo consigo benefícios e desafios. Logo, com o objetivo de explorar e compreender as decisões judiciais sobre falhas de companhias aéreas, um estudo de caso foi realizado no Juizado Especial Cível (JEC) do Fórum Norte da Ilha, localizado em Florianópolis/SC.

Já sabemos que o aprendizado de máquina é um ramo da inteligência artificial que permite que as máquinas aprendam com dados e façam previsões ou decisões. Como mencionado nos capítulos anteriores, existem diferentes tipos de aprendizado de máquina, como supervisionado, não supervisionado e por reforço, que dependem do tipo e da disponibilidade dos dados e do objetivo do problema.

Do mesmo modo, foi mencionado que o *Orange Canvas* é um software de código aberto que facilita o processo de aprendizado de máquina, oferecendo uma interface gráfica intuitiva e interativa. Nesse sentido, o *Orange Canvas* permite que o usuário crie fluxos de trabalho de aprendizado de máquina arrastando e soltando widgets, que são componentes que realizam diferentes funções, como carregar dados, visualizar gráficos, aplicar modelos, testar e avaliar resultados.

Para fins de elucidar seu funcionamento, temos um fluxo de trabalho de aprendizado de máquina que envolve várias etapas, como coleta, preparação, modelagem, treinamento, avaliação e predição dos dados. Cada etapa requer ferramentas e técnicas adequadas para garantir a qualidade e a eficiência dos resultados, conforme exemplo a seguir:

Figura 15 Ilustração do treinamento de IA explicável (XAI) em *Orange Canvas*



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

O fluxo de trabalho no *Orange Canvas* exemplificado acima, foi utilizado para realizar uma regressão logística, que é um tipo de modelo de aprendizado de máquina supervisionado que permite prever uma variável binária (sim ou não) a partir de variáveis independentes. O fluxo de trabalho segue as seguintes etapas:

- i. Carregar os dados de um arquivo CSV usando o widget “Base de Aprendizado”.
- ii. Selecionar os atributos (variáveis) relevantes para o modelo usando o widget “Seletor do Atributo”.
- iii. Aplicar o modelo de regressão logística usando o widget “Logistic Regression”.
- iv. Testar e avaliar o desempenho do modelo usando o widget “Test and Score”.
- v. Explicar o funcionamento do modelo usando o widget “*Explain Model*”.
- vi. Carregar os dados de teste de outro arquivo CSV usando o widget “Base Para Testes”.

- vii. Selecionar os atributos (variáveis) relevantes para a predição usando o widget “Seletor do Atributo”.
- viii. Visualizar os dados de teste em uma tabela usando o widget “Tabela de Dados”.
- ix. Fazer a predição do modelo usando o widget “*Explain Prediction*”.

Portanto, podemos entender o *Orange Canvas* como uma ferramenta versátil e acessível para análise de dados e aprendizado de máquina. Isso porque, permite aos usuários explorar dados, construir modelos preditivos e obter insights valiosos sem a necessidade de conhecimento avançado em programação.

#### **4.4 Análise e classificação de algoritmos de aprendizado de máquina em sua acurácia associado ao método explicativo *SHAP*.**

Os experimentos a seguir abordam a aplicação de algoritmos de aprendizado de máquina em processos judiciais, incluindo experimentos com regressão logística e redes neurais para prever resultados de ações judiciais. Para tanto, utilizam uma base de aprendizado com atributos relevantes, analisando métricas de desempenho para avaliar a eficácia dos modelos. Frisa-se a importância da base de testes para avaliar a capacidade de adaptação dos modelos a diferentes situações do mundo real.

Ademais, os experimentos com diferentes algoritmos compartilham algumas características comuns:

- Base de aprendizado: Todos os experimentos utilizam uma base de aprendizado contendo informações relevantes para a construção da sentença, com atributos selecionados para análise.
- Exclusão de atributos: Os experimentos excluem atributos específicos, como o nome do juiz, tipo de juiz e informações do número da decisão em arquivo *TXT*, visando direcionar o aprendizado de máquina para os dados mais relevantes.
- Análise de métricas: Após a aplicação dos algoritmos, são analisadas métricas de desempenho, como *AUC*, acurácia, *F1*, precisão e *recall*, para avaliar a eficiência dos modelos. Essas características comuns refletem a abordagem sistemática e criteriosa adotada nos experimentos com diferentes algoritmos.



Em suma, os experimentos apresentados demonstram a viabilidade da aplicação de algoritmos de aprendizado de máquina em processos judiciais, com destaque para a regressão logística e redes neurais. Os algoritmos *Adaptive Boosting* e *Gradient Boosting*, por outro lado, não foram tão bem-sucedidos, deixando a desejar principalmente no quesito explicabilidade e acurácia.

#### 4.4.1 *Adaptive Boosting*

O *AdaBoost* (*Adaptive Boosting*) é um algoritmo de aprendizado de máquina que é utilizado para melhorar o desempenho de modelos de classificação. Ele foi proposto por Yoav Freund e Robert Schapire em 1996. Nesse sentido, é um algoritmo de aprendizado de conjunto (*ensemble learning*), o que significa que combina a saída de vários modelos de aprendizado de máquina simples para criar um modelo mais poderoso.

A cada iteração, o *AdaBoost* aumenta o peso das instâncias que foram classificadas incorretamente nas iterações anteriores, para que o próximo modelo se concentre nessas instâncias mais difíceis. O resultado final é um modelo forte que combina as previsões dos modelos fracos ponderados pela sua acurácia. O *AdaBoost* se adapta aos exemplos mais difíceis durante o treinamento, o que o torna um algoritmo robusto e eficiente para melhorar a classificação. A fórmula matemática que representa o *AdaBoost* é:

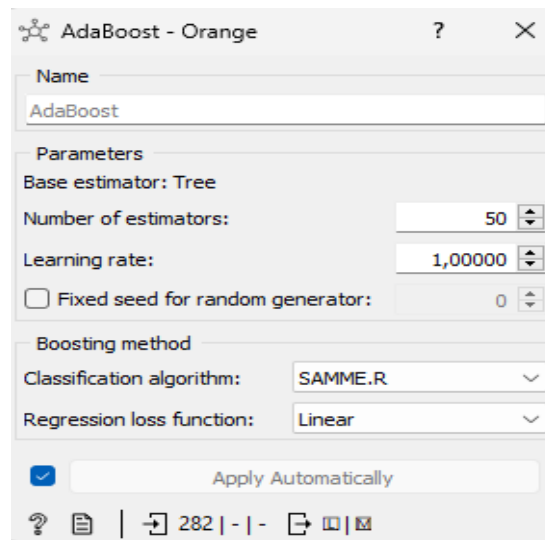
**Figura 16 Ilustração do treinamento de IA explicável (XAI)**

$$f(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right)$$

Nota: Imagem da fórmula matemática que representa o *AdaBoost*.

Onde  $f(x)$  é o modelo forte,  $h_t(x)$  é o modelo fraco na iteração  $t$ ,  $\alpha_t$  é o peso do modelo fraco na iteração  $t$ , e *sign* é a função que retorna o sinal de uma expressão. No *Orange Canvas*, o *AdaBoost* está disponível como um dos algoritmos de aprendizado de máquina pré-implementados. A figura a seguir ilustra o processo do *AdaBoost*:

**Figura 17 Ilustração do *AdaBoost***

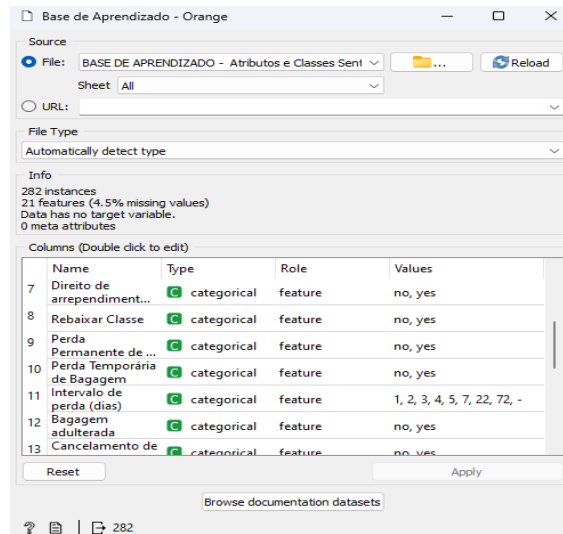


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

O resultado final do *AdaBoost* é um modelo conjunto que combina as previsões dos modelos fracos ponderados de acordo com sua precisão. Sua capacidade de adaptar-se aos exemplos mais difíceis durante o treinamento o torna um algoritmo poderoso para melhorar o desempenho da classificação.

No caso em tela, temos, contudo, uma base de aprendizado de máquina, que se constituiu como um conjunto de dados utilizados para treinar e avaliar modelos de aprendizado de máquina. Em outras palavras, é uma coleção de instâncias, onde cada instância representa uma observação ou exemplo e é composta por um conjunto de características (atributos) e um valor de destino (rótulo) que se deseja prever.

**Figura 18 Ilustração do *AdaBoost* – Base de Aprendizado**

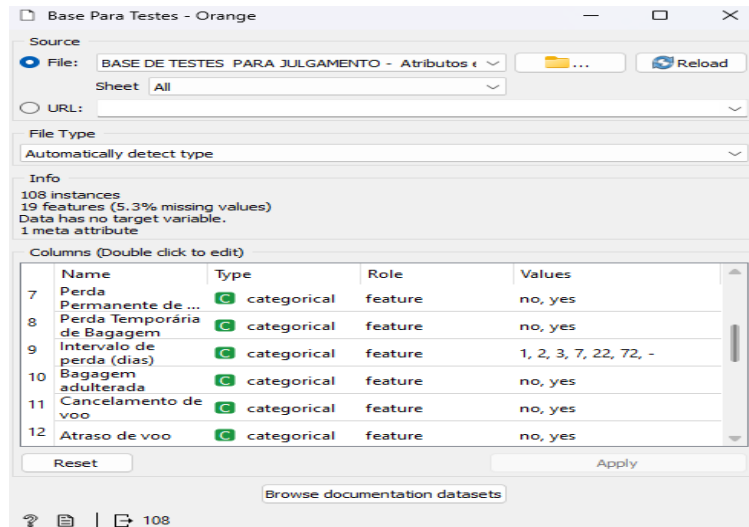


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

A ilustração acima, extraída do software apresenta alguns dados relevantes, como o número de 282 processos analisados e estruturados no arquivo CSV, nos quais temos 21 diferentes atributos. Válido ressaltar que, mesmo após o treinamento, a base de aprendizado também é usada para avaliar o desempenho do modelo. Igualmente, é possível visualizar e explorar os dados da base de aprendizado, realizar pré-processamento, criar e treinar modelos de aprendizado de máquina, assim como também avaliar o desempenho dos modelos.

Por outro lado, temos ainda uma base de testes que é um conjunto de dados separado da base de aprendizado utilizado para avaliar o desempenho de um modelo de aprendizado de máquina, após o treinamento. Obviamente que a base de testes é composta por instâncias que não foram usadas durante o treinamento do modelo e, portanto, representa dados não vistos anteriormente.

Figura 19 Ilustração do *AdaBoost* – Base de Testes

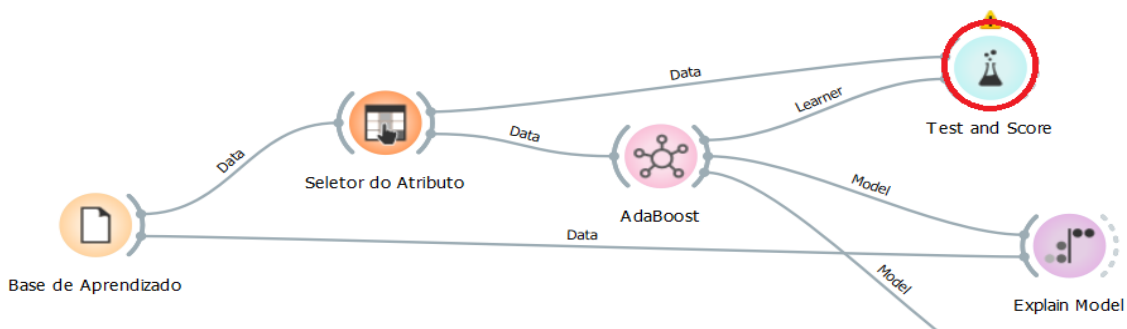


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

De igual modo, temos aqui 108 processos analisados e estruturados em arquivo CSV, com 19 atributos analisados. Destarte, a base de testes é importante para verificar como o modelo generaliza e se comporta em novos dados, ou seja, como ele é capaz de fazer previsões precisas em situações reais. É uma forma de avaliar a capacidade do modelo de se adaptar a diferentes padrões e de medir sua eficácia em relação ao mundo real. Ademais, ao avaliar o desempenho do modelo na base de testes, é possível utilizar métricas de avaliação, como acurácia, precisão e *recall*.

Destarte, no *Orange Canvas*, temos também o *Test and Score*, que é um componente que permite avaliar o desempenho de um modelo de aprendizado de máquina. Ele é usado para medir a capacidade do modelo em fazer previsões precisas com base nos dados fornecidos.

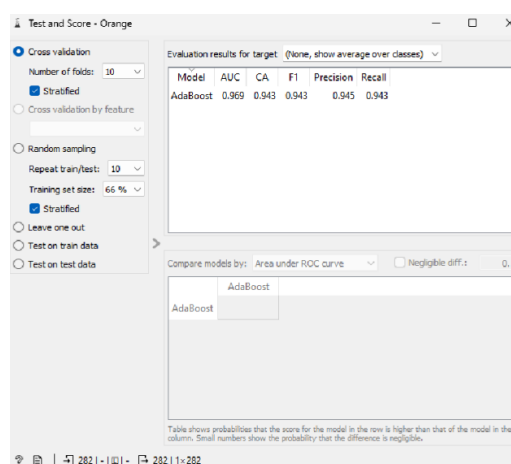
Figura 20 Ilustração do *AdaBoost* – Fluxo de trabalho



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Em seguida, uma vez que o *Test and Score* está configurado corretamente, será executado o modelo treinado nos dados de teste e gerará várias métricas de avaliação. Essas métricas podem incluir medidas como precisão, revocação, *F1-score*, matriz de confusão, curva ROC, entre outras, dependendo do tipo de problema e do tipo de modelo utilizado.

**Figura 21 Ilustração do *AdaBoost* – Test and Score**

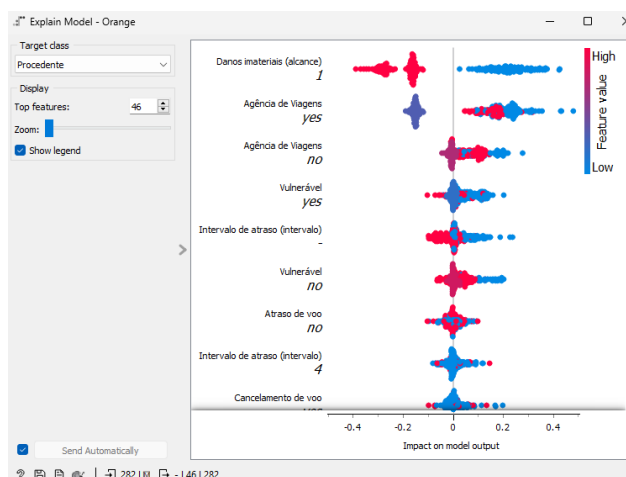


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Assim sendo, com o modelo de aprendizado de máquina *Adaboost* foi possível identificar através da base de aprendizado anteriormente mencionada, as seguintes métricas: a) *AUC*: 0.969; b) *CA*: 0.943; c) *F1*: 0.943; d) *Precision*: 0.945; e) *Recall*: 0.943. Essas métricas são positivas e ajudam a entender o quão bem o modelo está performando, assim como comparar com os demais modelos de aprendizado de máquina.

Outrossim, para a verificação da explicabilidade do modelo, foi utilizado o *Explain Model* no *Orange Canvas*, que, de modo simplificado, é uma funcionalidade que permite entender e interpretar o comportamento e as decisões de um modelo de aprendizado de máquina. Ela auxilia na exploração dos fatores e características que influenciam as previsões feitas pelo modelo.

Figura 22 Ilustração do *AdaBoost – Explain Model*

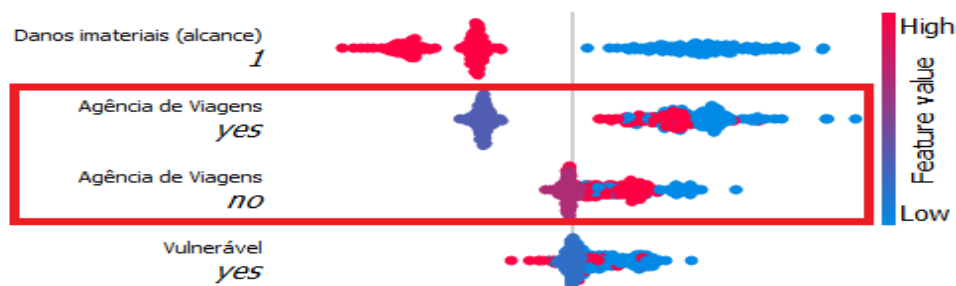


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Na ilustração acima, temos uma imagem extraída do *software* que identifica os principais atributos que são levados em consideração para a predição quando o valor-alvo (*target*) é o das decisões judiciais procedentes. Quanto mais alto o atributo estiver na sequência de 46 *features*, mais importante ele é para o resultado. Destarte, os pontos azuis representam valores menores daquele atributo, ao passo que os pontos vermelhos representam valores altos. Entretanto, o que é mais importante, é a sua localização, isto é a direita da origem (impacto negativo no modelo), ou a esquerda (impacto positivo no modelo).

Assim sendo, temos que valores baixos de danos imateriais no intervalo 1 (um), influenciam positivamente para a procedência do pedido, assim como a ausência de agência de viagens representada por pontos em azul no atributo 2 (dois) e/ou pontos em vermelho no atributo 3 (três):

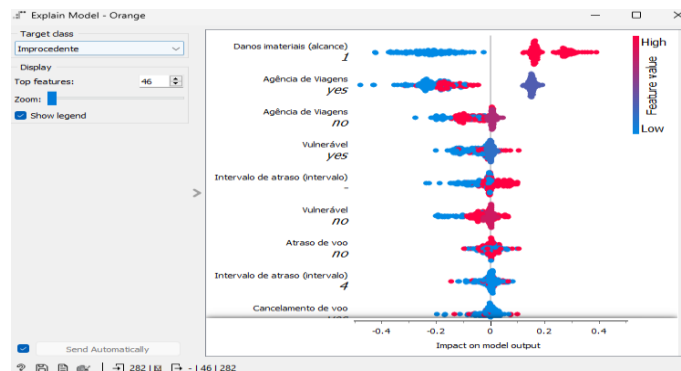
Figura 23 Ilustração do *AdaBoost – Explain Model*



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Ademais, quando invertemos o valor-alvo (*target*) de procedente para improcedente, podemos notar que a ordem de relevância dos atributos é mantida. De fato, existem alterações, porém estas residem tão somente na coloração e sentido de alguns dos pontos apresentados para cada um dos atributos, como se observa na ilustração a seguir:

**Figura 24 Ilustração do *AdaBoost – Explain Model***

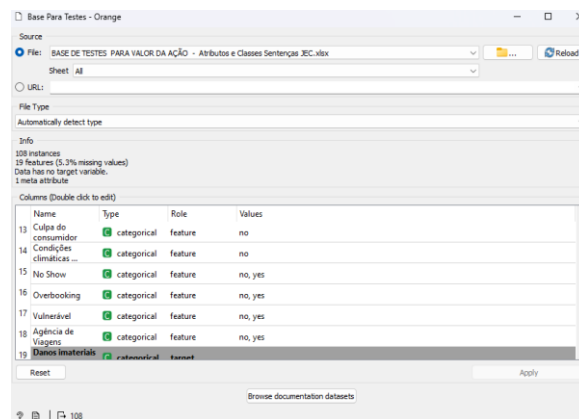


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Entretanto, tenhamos em mente que o que se busca é também a explicabilidade da previsão de resultado do valor de ação. Ou seja, para além do resultado da sentença, buscou-se definir uma explicabilidade para previsibilidade em valor de ação mantida nos intervalos de: a) (1) R\$ 0,00 ; b) (2) R\$ 1,00 - R\$ 2.000,00; c) (3) R\$ 2.001,00 - R\$ 5.000,00; d) (4) R\$ 5.001,00 - R\$ 8.000,00; e) (5) R\$ 8.001,00 - R\$ 25.000,00.

Para tanto, se faz uso de uma segunda base de testes, na qual ao invés de suprimir o resultado procedente ou não da sentença em relação ao pedido; o que se busca justamente é suprimir o valor da ação para que, justamente, o modelo de predição possa descobrir através dos dados e evidências dispostos nos demais atributos, conforme trecho a seguir:

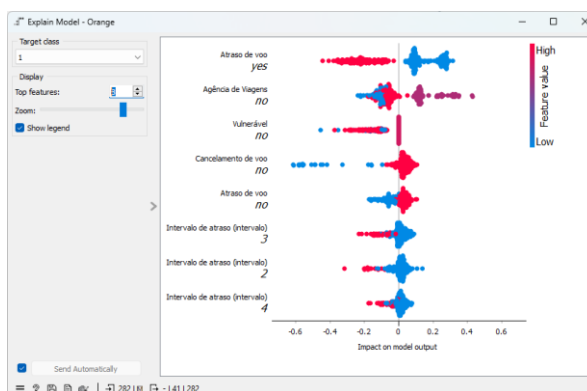
Figura 25 Ilustração do *AdaBoost* – Base de Testes



Nota: Imagem do Autor. Extraída de *Orange Canvas*

Evidente, que a ferramenta do *Explain Model*, diferentemente do *Explain Prediction*, analisa o modelo preditivo como um todo. Com efeito, temos que para o intervalo de valor 1 (um), isto é, quando não existe quantia a ser percebida, o modelo preditivo elenca alguns atributos que estão presentes nessas situações, através do aprendizado de máquina na base de dados de aprendizado, somados ao da base de testes, conforme ilustração a seguir:

Figura 26 Ilustração do *AdaBoost* – *Explain Model*



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

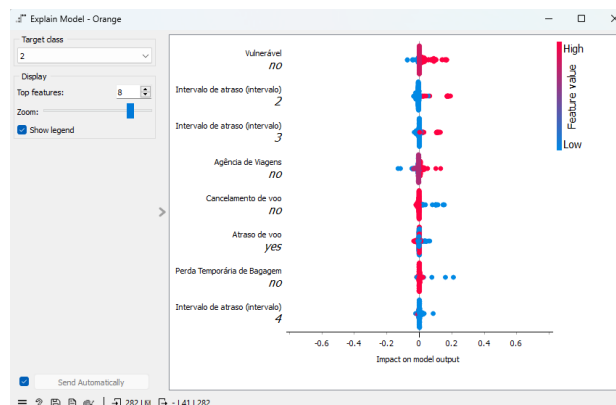
A primeira variável, isto é a mais importante para o modelo considerar a indenização no intervalo 1 (um) é o atraso de voo. Entretanto, percebe-se que pela



coloração dos pontos vermelhos ao lado esquerdo da origem, que, de fato o atraso de voo contribui negativamente para a predição da indenização no intervalo 1(um). Existem outras variáveis, todavia elas permanecem com disposições mais próximas da origem, como é o caso da agência de viagens, presença de vulnerável, cancelamento de voo e intervalo de atraso.

Antes de prosseguir, cumpre salientar que do mesmo modo que o valor da ação foi dividido em intervalos, o mesmo ocorreu com o tempo de atraso do voo, a dizer: (1) 0:00:00 - 4:00:00; (2) 4:01:00 - 8:00:00; (3) 8:01:00 - 12:00:00; (4) 12:01:00 - 24:00:00; (5) 24:01:00 - 72:00:00. Isso é importante, pois como veremos o intervalo de atraso é um dos principais atributos levados em consideração pelo modelo preditivo, ao analisa casos em que se dá a indenização no intervalo 2 (dois) em que se percebe valores entre R\$ 1,00 (um real) e R\$ 2.000,00 (dois mil reais):

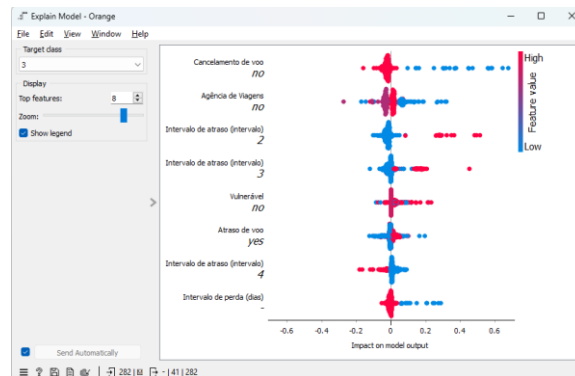
**Figura 27 Ilustração do AdaBoost – Explain Model**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Percebe-se que, através dos 282 casos apresentados em base de aprendizado, o modelo identificou 41 casos em que o valor-alvo (*target*) está no intervalo 2 (dois). O atributo da vulnerabilidade se destaca nesse intervalo, isto é, considera que a ausência de vulnerável contribui positivamente esse resultado. Contudo, é possível identificar que os valores ficaram muito divididos, em que pouco é possível distinguir, dado que existe notável mescla de pontos vermelhos e azuis dispostos próximos a origem, ou zero. O que não acontece, por exemplo, com o intervalo 3 (três) de R\$ 2.001,00 (dois mil e um reais) à R\$ 5.000,00 (cinco mil reais):

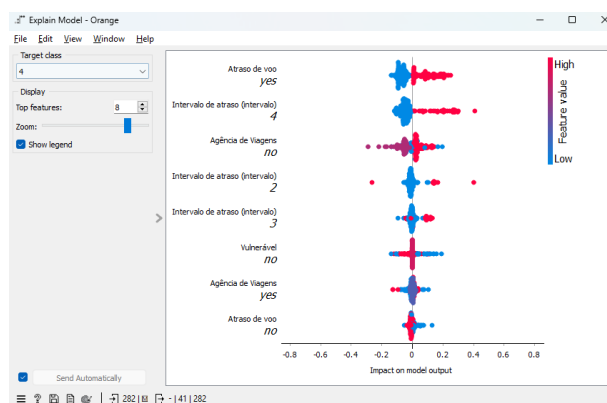
Figura 28 Ilustração do *AdaBoost – Explain Model*



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Como se percebe, já existe um espaçamento mais entre os pontos vermelhos e azuis. O primeiro atributo é o do cancelamento de voo. Apesar da existência de alguns pontos azuis, que confirmam casos em que a indenização está no intervalo 3 (três), boa parte dos pontos vermelhos relativos ao atributo, estão do lado esquerdo da origem, o que demonstra que a ausência de cancelamento de voo, contribui negativamente para a previsão de tal resultado. A seguir temos o *Explain Model* para o intervalo de valor de ação 4 (quatro) de R\$ 5.001,00 (cinco mil e um reais) à R\$ 8.000,00 (oito mil reais):

Figura 29 Ilustração do *AdaBoost – Explain Model*

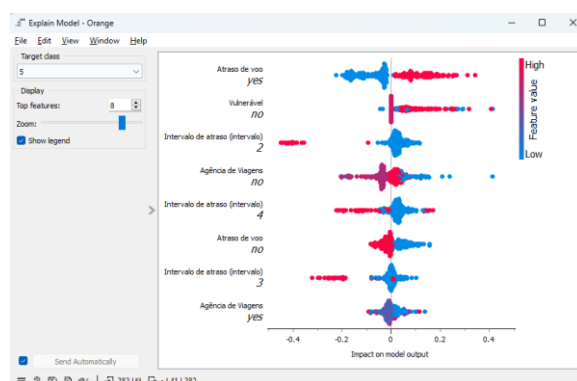


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Aqui percebe-se que, a confirmação de atraso de voo e do intervalo de atraso 4 (quatro) de 12:01:00 (doze horas e um minuto) à 24:00:00 (vinte quatro horas) são os

principais atributos para cognição da máquina. Todavia, o atributo da agência de viagens, encontra-se mesclado próximo da origem de pontos azuis e vermelhos, o que indica que esse atributo, pouco contribui para a predição do resultado. Outros atributos são apresentados, como é o caso da vulnerabilidade e outros intervalos de atraso. Por fim, temos a seguir a ilustração do *Explain Model* para o intervalo de valor de ação 5 (cinco) de R\$ 8.001,00 (oito mil e um reais) à R\$ 25.000,00 (vinte cinco mil reais):

**Figura 30 Ilustração do *AdaBoost* – Base de Testes**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

É possível identificar, nesse caso, que boa parte dos pedidos de valor de indenização nesse intervalo, considera como principais atributos para a predição do resultado o atraso de voo, vulnerabilidade, intervalo de atraso e agência de viagens. Entretanto, muito embora compreenda que o atraso de voo é positivo para a predição do resultado, também compreende que a ausência de vulneráveis é, igualmente, positiva. Resumidamente, o *Explain Model* funciona como uma visão panorâmica do modelo, ao passo que o *Explain Prediction*, é a lente que dá zoom e permite ver detalhes intrínsecos.

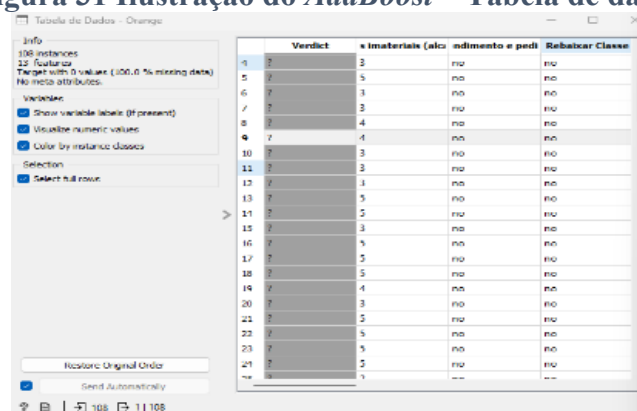
Dito de outra maneira, a ferramenta *Explain Prediction* no *Orange Canvas* é uma funcionalidade que permite compreender e explicar individualmente as previsões feitas por um modelo de aprendizado de máquina. Ela ajuda a entender os fatores específicos que influenciam uma previsão em particular e fornece uma visão detalhada sobre como o modelo chegou a essa decisão.

Ao usar a *Explain Prediction* no *Orange Canvas*, é possível explorar uma previsão específica do modelo e investigar quais características dos dados foram consideradas relevantes para essa previsão em particular. Isso é especialmente útil quando se deseja

entender o raciocínio do modelo e as justificativas para suas previsões.

De modo geral, a *Explain Prediction* pode ser aplicada a vários tipos de modelos, incluindo árvores de decisão, modelos lineares, modelos de aprendizado profundo, entre outros. Além disso, também oferece recursos visuais para ajudar a interpretar as explicações, como gráficos de importância das características, gráficos de contribuição individual e representações gráficas dos padrões identificados pelo modelo para uma previsão específica.

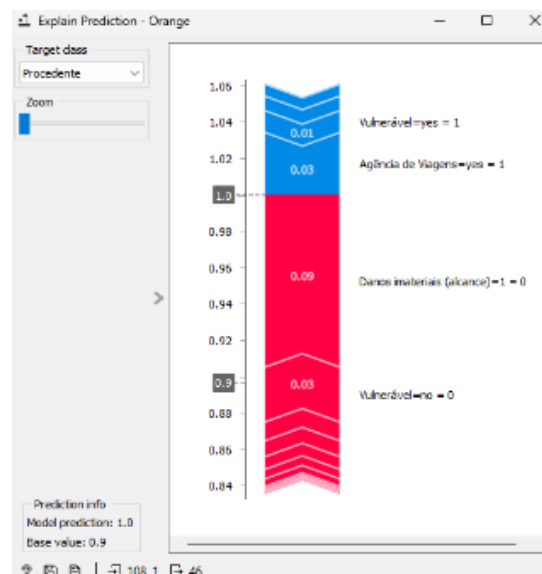
Figura 31 Ilustração do *AdaBoost* – Tabela de dados



	Verdict	Danos materiais (altura)	ndimento e pedr	Reabscar Classe
1	?	3	no	no
5	?	5	no	no
6	?	3	no	no
7	?	3	no	no
8	?	4	no	no
9	?	4	no	no
10	?	3	no	no
11	?	3	no	no
12	?	3	no	no
13	?	5	no	no
14	?	5	no	no
15	?	3	no	no
16	?	5	no	no
17	?	5	no	no
18	?	5	no	no
19	?	4	no	no
20	?	3	no	no
21	?	5	no	no
22	?	5	no	no
23	?	5	no	no
24	?	5	no	no
25	?	5	no	no

Nota: Imagem do Autor. Extraída de *Orange Canvas*

Figura 32 Ilustração do *AdaBoost* – *Explain Prediction*



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Em análise interpretativa do valor-alvo (*target*) procedente, temos que para o caso número 9 (nove), disposto na base de testes para julgamento, a *XAI* aponta que exista de 90% (noventa por cento) à 100% (cem por cento de chance) de que aquela ação judicial tenha como veredicto a procedência do pedido.

Novamente, cumpre salientar que a informação sobre a procedência ou não da ação é omitida na base de testes para julgamento, justamente com o intuito de realizar a previsão do caso pelo *machine learning* e otimizar o seu resultado, conforme se verifica em trecho extraído do documento CSV, estruturado para o *Orange Canvas*:

**Figura 33 Ilustração do *AdaBoost* – Planilha de dados CSV**

	A	B
1	TXT file	Danos imateriais
8	655	3
9	656	4
10	657	4
11	658	3
12	659	3
13	660	3

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Igualmente, em outro arquivo é possível realizar a checagem das informações, onde consta a informação de que o *TXT file* 656 teve o veredicto “*partly\_founded*”, isto é parcialmente procedente, conforme se observa no trecho a seguir:

**Figura 34 Ilustração do *AdaBoost* – Planilha de dados CSV**

1	TXT file	Verdict	Verdict (2)	Immaterial dam	Immaterial dam
658	653	partly_founded	-	RS 4.600,00	3
659	654	partly_founded	-	RS 4.500,00	3
660	655	partly_founded	-	RS 5.000,00	3
661	656	partly_founded	-	RS 7.000,00	4
662	657	partly_founded	-	RS 8.000,00	4
663	658	partly_founded	-	RS 3.500,00	3
664	659	partly_founded	-	RS 3.000,00	3
665	660	partly_founded	-	RS 5.000,00	3
666	661	partly_founded	-	RS 10.000,00	5

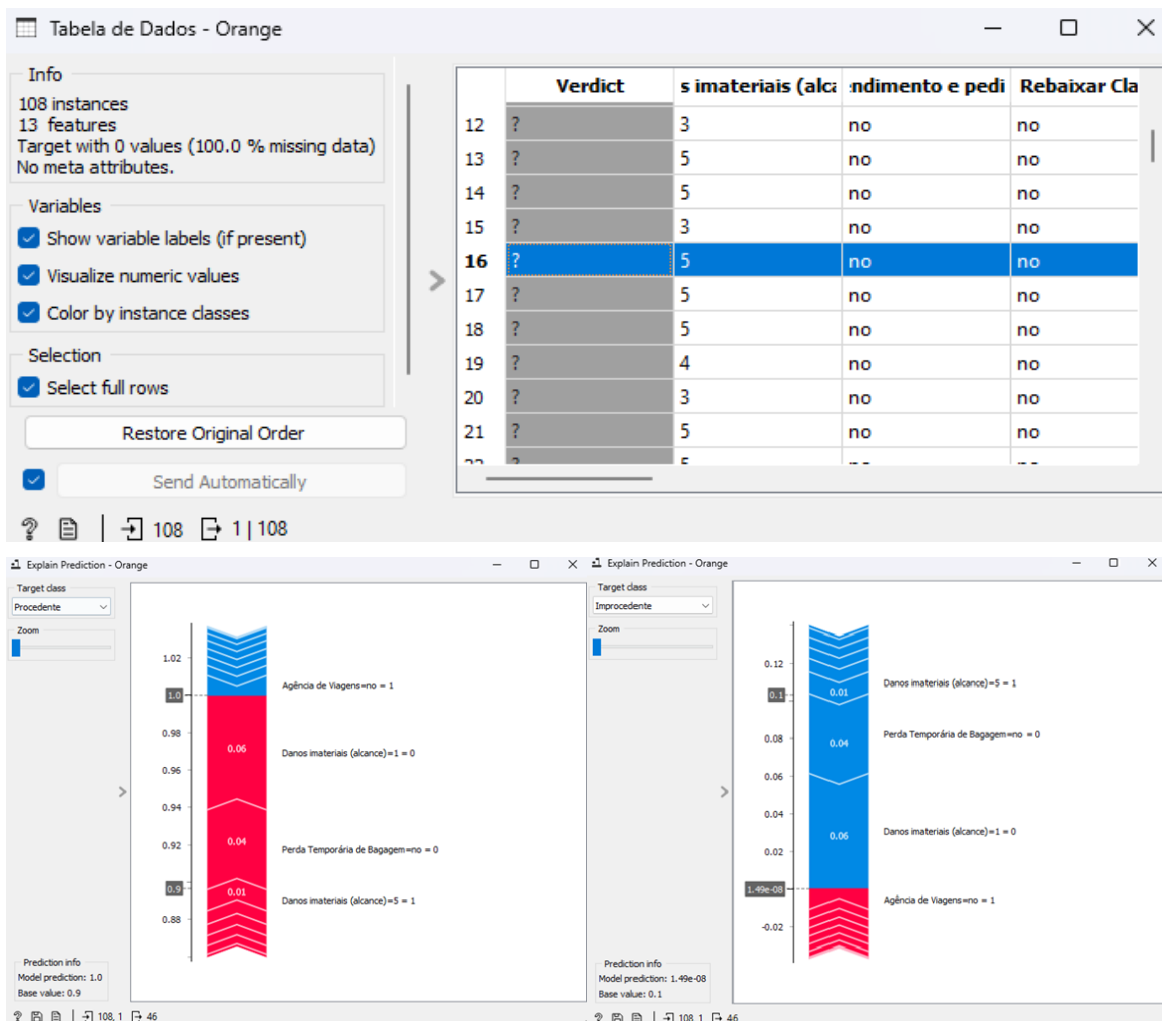
Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Contudo, dado o fato de que foi dado apenas dois *targets* para a IA, (a) procedente e (b) improcedente, temos que o resultado foi razoavelmente atingido e explicado, dado

que elenca tanto atributos favoráveis como contrários a interpretação.

Mais adiante, temos outro exemplo analisado, dessa vez com o caso número 16 (dezesseis) da base de testes para julgamento, em que novamente se busca definir a chance de procedência do julgamento. Todavia, dessa vez temos também a chance de improcedência da ação:

Figura 35 Ilustração do *AdaBoost* – Tabela de dados/*Explain Prediction*



Nota: Imagem do Autor. Extraída de *Orange Canvas*

Através das imagens apresentadas, destaca-se o fato de que o modelo define que a chance da ação número 16 (dezesseis) ser procedente está entre 90% (noventa por cento) e 100% (cem por cento), ao passo que a chance de ser improcedente está abaixo de 10% (dez por cento). Além disso é capaz de identificar que dados que influenciam

positivamente na previsão de procedência como a presença de pedido de danos imateriais, assim como os que influenciam negativamente como a agência de viagens.

Mais uma vez, cumpre salientar que a informação sobre a procedência ou não da ação é omitida na base de testes para julgamento, justamente com o intuito de realizar a previsão do caso pelo *machine learning* e otimizar o seu resultado, conforme se verifica em trecho extraído do documento CSV, estruturado para o *Orange Canvas*:

Figura 36 Ilustração do *AdaBoost* – Planilha de dados CSV

	A	B
1	TXT file	Danos imateriais
15	662	5
16	663	3
17	665	5
18	667	5
19	668	5

Nota: Imagem do Autor. Extraída de *Orange Canvas*

Assim também, em outro arquivo é possível realizar a checagem das informações, onde consta a informação de que o *TXT file* 656 teve o veredicto “*partly\_founded*”, isto é parcialmente procedente, conforme se observa no trecho a seguir:

Figura 37 Ilustração do *AdaBoost* – Planilha de dados CSV

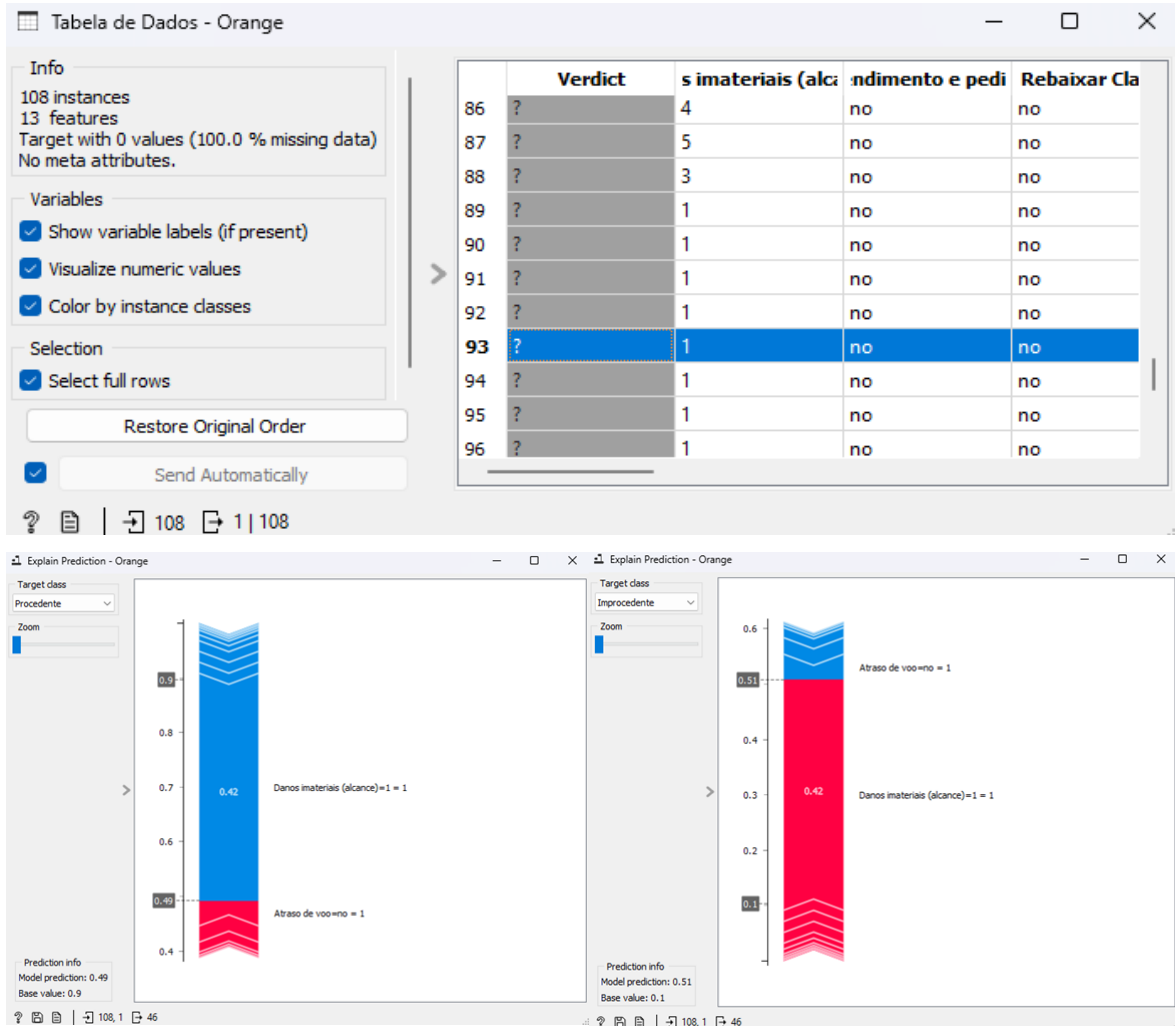
	A	B	C	D
1	TXT file	Verdict	Verdict (2)	Immaterial damage
664	659	partly_founded	-	R\$ 3.000,00
665	660	partly_founded	-	R\$ 5.000,00
666	661	partly_founded	-	R\$ 10.000,00
667	662	well_founded	-	R\$ 15.000,00
668	663	partly_founded	-	R\$ 5.000,00
669	664	not_founded	-	R\$ 0,00
670	665	partly_founded	-	R\$ 10.000,00
671	666	partly_founded	-	R\$ 0,00
672	667	partly_founded	-	R\$ 11.000,00

Nota: Imagem do Autor. Extraída de *Orange Canvas*

Embora fantástico e fascinante tais casos temos, por outro lado, situações em que o resultado não é tão facilmente observável. Uma dessas situações acontecem com o caso número 93 (noventa e três) em que a *XAI* não consegue realizar a previsibilidade com

uma acurácia tão precisa quanto aos demais casos:

Figura 38 Ilustração do *AdaBoost* – Tabela de dados/*Explain Prediction*



Nota: Imagem do Autor. Extraída de *Orange Canvas*

Como podemos identificar nas imagens, o modelo de *XAI*, nesse caso em específico, afirma que a chance de procedência do pedido está abaixo dos 49%, na mesma medida em que a chance de improcedência está entre 10% e 51%. Essa ligeira diferença, contudo, não é suficiente para considerar que o resultado da ação seja improcedente.

Todavia, em cumprimento da ritualística estabelecida, temos a seguir tanto trecho extraído do documento CSV, estruturado para o *Orange Canvas*, quanto outro arquivo em que é possível realizar a checagem das informações, onde consta a informação do *TXT*



file:

Figura 39 Ilustração do *AdaBoost* – Planilha de dados CSV

	A	B
1	TXT file	Danos imateriais
91	1140	1
92	1141	1
93	1142	1
94	1143	1
95	1108	1

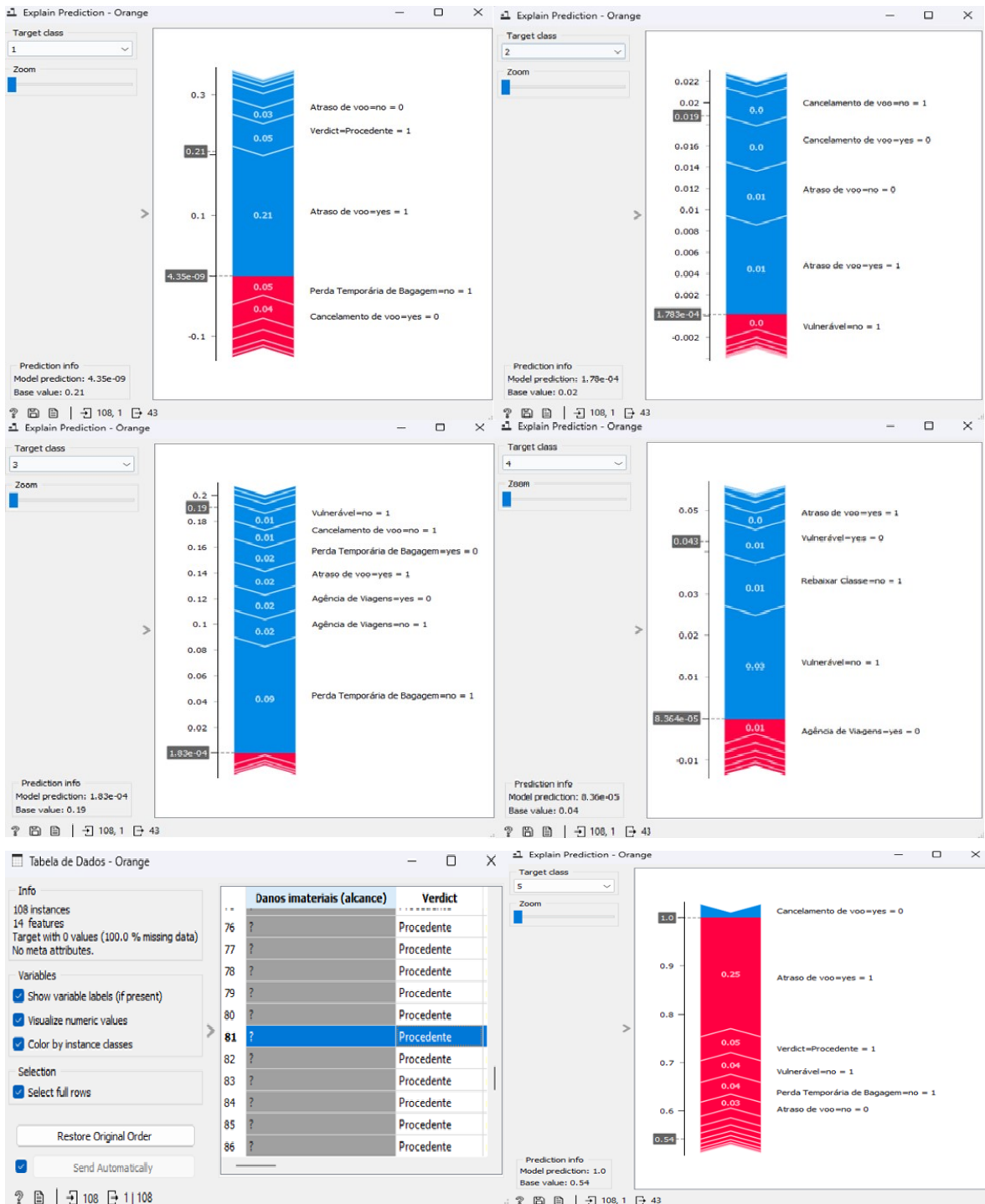
	A	B	C	D	E
1	TXT file	Verdict	Verdict (2)	Immaterial dama:	Immaterial dama:
1146	1138	partly_founded	-	R\$ 4.000,00	3
1147	1139	not_founded	-	R\$ 0,00	1
1148	1140	not_founded	-	R\$ 0,00	1
1149	1141	not_founded	-	R\$ 0,00	1
1150	1142	not_founded	-	R\$ 0,00	1
1151	1143	not_founded	-	R\$ 0,00	1
1152	1144	well_founded	-	R\$ 0,00	1
1153	1145	partly_founded	-	R\$ 2.500,00	3
1154	1146	well_founded	-	R\$ 0,00	1
1155	1147	partly_founded	-	R\$ 15.000,00	5
1156	1148	partly_founded	-	R\$ 9.000,00	5
1157	1149	partly_founded	-	R\$ 1.000,00	2

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Conclui-se, portanto, que o *TXT file* 1141 (mil cento e quarenta e um), que corresponde ao caso número 92 (noventa e dois), teve como resultado “*not\_founded*”, isto é, a sua improcedência com relação ao pedido. No caso em tela, já não é possível levar em consideração a predição da máquina dado o nível de incerteza apresentado.

Em seguida, temos o *Explain Prediction*, porém dessa vez aplicado a explicabilidade da previsão sobre o valor a ser indenizado em ação judicial. Para tanto, além da base de aprendizado, foi feito o uso de uma base de testes que omite o valor da ação, assim como foi feito anteriormente com o veredito.

Figura 40 Ilustração do AdaBoost – Explain Prediction



Nota: Imagem do Autor. Extraída de Orange Canvas

Observa-se que para a XAI, a ação de número 81 (oitenta e um) tem mais chance de ser compatível com o intervalo 5 (cinco) de valor de ação, o qual corresponde a quantia

entre R\$ 8.001,00 (oito mil e um reais) à R\$ 25.000,00 (vinte e cinco mil reais). Nesse sentido, o modelo de *XAI* aponta baseado em fatores com procedência do pedido, pessoa em situação de vulnerabilidade, atraso de voo e alguns mais; que as chances de se obter indenização no intervalo 5 está entre 54% (cinquenta e quatro por cento) e 100% (cem por cento).

Para além da predição, em consulta a base de checagem, é possível identificar que a previsão não atingiu o objetivo, conforme se ilustra nas imagens a seguir:

**Figura 41 Ilustração do *AdaBoost* – Planilha de dados CSV**

	A	B	C	D	E
1	TXT file	Verdict	Judge	Type of judge	Direito de arrep:
77	753	Procedente Vânia Petermann	chief		no
78	754	Procedente Vânia Petermann	chief		no
79	755	Procedente Vânia Petermann	chief		no
80	757	Procedente Vânia Petermann	chief		no
81	758	Procedente Vânia Petermann	chief		no
82	759	Procedente Vânia Petermann	chief		no
83	760	Procedente Vânia Petermann	chief		no
84	761	Procedente Vânia Petermann	chief		no
85	763	Procedente Vânia Petermann	chief		no

	A	B	C	D	E
1	TXT file	Verdict	Verdict (2)	Immaterial damage	Immaterial damage
759	751	well_founded	-	R\$ 0,00	1
760	752	partly_founded	-	R\$ 6.500,00	4
761	753	partly_founded	-	R\$ 9.000,00	5
762	754	partly_founded	-	R\$ 3.000,00	3
763	755	partly_founded	-	R\$ 10.000,00	5
764	756	not_founded	-	R\$ 0,00	1
765	757	partly_founded	-	R\$ 13.500,00	5
766	758	partly_founded	-	R\$ 4.500,00	3
767	759	partly_founded	-	R\$ 5.000,00	3
768	760	partly_founded	-	R\$ 6.000,00	4
769	761	partly_founded	-	R\$ 4.000,00	3
770	762	partly_founded	-	R\$ 0,00	1

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Embora, exista nível de acurácia razoável para o resultado de procedência, ou não; o mesmo não acontece com a ação de número 81 (oitenta e um), dado que o intervalo correto seria o de número 3 (três), ao invés do 5 (cinco). Seria leviano, contudo, renegar o potencial latente da *machine learning*, ou ainda dizer que tal mecanismo não se aplica a explicabilidade de predições de IA em decisões judiciais.

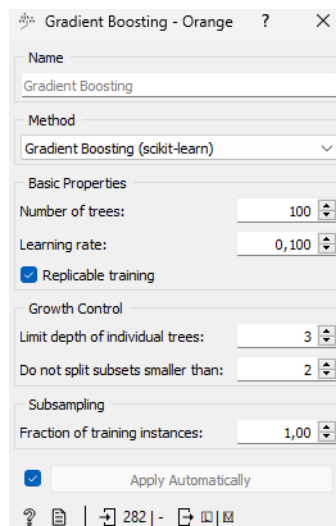
O que nos leva, inicialmente a considerar que talvez o modelo de aprendizagem *AdaBoost*, necessite de uma base de dados maior para poder ter um índice de acerto digno de uma posterior análise de explicabilidade de sua predição, ainda que tenha tido louvável êxito no tocante a previsão do resultado do julgamento, assim como na definição dos

atributos e pontos mais relevantes levados em consideração pela IA.

#### 4.4.2 Gradientboost

O *Gradient Boost* é um algoritmo de aprendizado de máquina que faz parte da família de métodos de *Gradient Boosting*, que também inclui o *AdaBoost*. No *Orange Canvas*, ele está disponível como um dos algoritmos de aprendizado de máquina pré-implementados:

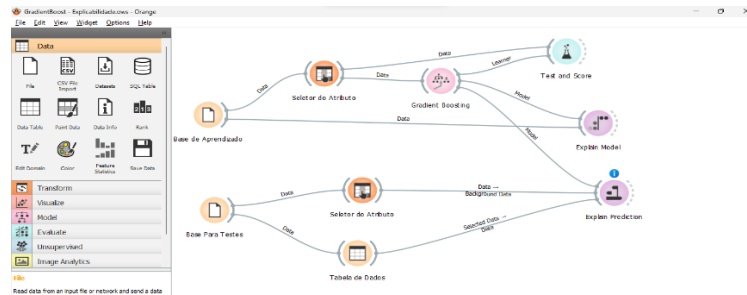
**Figura 42 Ilustração do Gradient Boosting**



Nota: Imagem do Autor. Extraída de *Orange Canvas*

O *Gradient Boost* é usado para problemas de regressão e classificação, onde o objetivo é prever um valor contínuo ou classificar instâncias em classes distintas. Assim como o *AdaBoost*, ele também utiliza um conjunto de modelos fracos, geralmente árvores de decisão simples, para criar um modelo mais poderoso.

**Figura 43 Ilustração do *GradientBoosting* – Fluxo de Trabalho**



Nota: Imagem do Autor. Extraída de *Orange Canvas*

A diferença fundamental entre o *AdaBoost* e o *Gradient Boost* está na forma como os modelos fracos são treinados e combinados. No *Gradient Boost*, os modelos fracos são treinados de forma sequencial, onde cada modelo subsequente é treinado para corrigir os erros do modelo anterior.

Nesse sentido, o *Gradient Boost* é conhecido por ser um algoritmo robusto e eficiente, capaz de lidar com problemas complexos de regressão e classificação. Ele tem sido amplamente utilizado em diversas aplicações, especialmente em competições de ciência de dados, devido à sua capacidade de obter resultados de alta qualidade.

Nesse segundo experimento, de igual modo é utilizada uma base de aprendizado, na qual estão contidas as informações referentes as ações envolvendo companhias aéreas, obviamente, estruturadas em atributos e classes. A base contém ao todo 282 (duzentos e oitenta e dois) casos para aprendizado de máquina, com 21 (vinte e um) atributos selecionados para análise, conforme se observa na ilustração a seguir:

**Figura 44 Ilustração do *Gradient Boosting* – Base de Aprendizado**

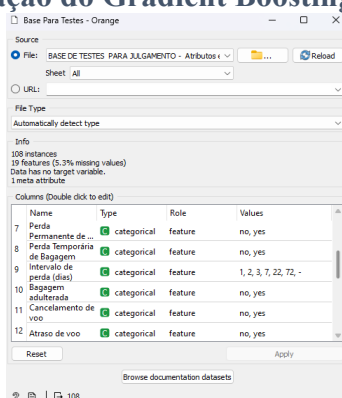
Name	Type	Role	Values
1 TXT file	numeric	skip	
2 Verdict	categorical	feature	Improcedente, Procedente
3 Verdict (2)	categorical	skip	-, Improcedente
4 Danos Imateriais (alcance)	categorical	target	
5 Judge	categorical	skip	Ana Luisa Schmidt Ramos, Bruna Luiza ...
6 Type of judge	categorical	skip	assistant, chief, voluntary
7 Direito de	categorical	feature	no_vet

Nota: Imagem do Autor. Extraída de *Orange Canvas*

De fato, são suprimidos os atributos do tipo de juiz, nome do juiz e informações do número da decisão em arquivo *TXT*. O objetivo, é o de simplesmente direcionar o aprendizado de máquina para os dados realmente relevantes para a construção da sentença judicial e, por conseguinte, da predição pela máquina de um possível resultado.

Ademais, como não poderia deixar de ser, cumpre salientar que, de igual modo, temos uma base de teste com 108 (cento e oito) processos analisados e estruturados em arquivo *CSV*. A diferença sua em relação a base de aprendizado, está no fato de ser composta por instâncias que não foram usadas durante o treinamento do modelo e, portanto, representa dados não vistos anteriormente.

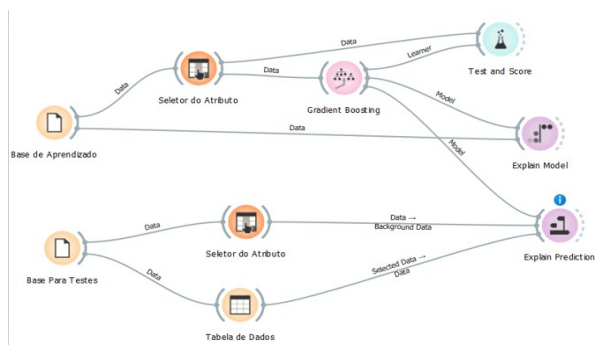
**Figura 45 Ilustração do Gradient Boosting – Base de Testes**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

A base de testes é fundamental para verificar como o modelo generaliza e se comporta com novos dados, isto é, como consegue realizar previsões precisas em situações reais. Em outras palavras, essa se constitui como uma forma de avaliar a capacidade do modelo de se adaptar a diferentes padrões, assim como também medir o desempenho mediante métricas de avaliação como acurácia, precisão e *recall*. Nesse sentido, temos o *Test and Score*, utilizado para medir a capacidade do modelo em fazer tais previsões precisas com base nos dados fornecidos.

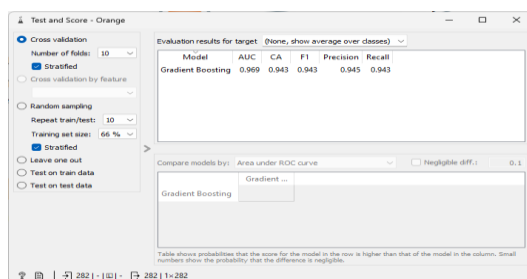
Figura 46 Ilustração do Gradient Boosting – Fluxo de Trabalho



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

A partir dessa sequência apresentada na ilustração acima, é possível identificar que o *Test and Score* será executado em um algoritmo de aprendizado de máquina, que no caso em tela é o *Gradient Boosting*. Em seguida, o modelo treinado gerará várias métricas de avaliação, tais como precisão, revocação, *F1-score*, matriz de confusão e curva ROC; dependendo do modelo utilizado e do problema em específico.

Figura 47 Ilustração do Gradient Boosting – Test and Score



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Portanto, conforme as informações extraídas do *Orange Canvas*, o modelo de aprendizado de máquina *Gradientboost* alcançou seguintes métricas: a) AUC: 0.969; b) CA: 0.943; c) F1: 0.943; d) Precision: 0.945; e) Recall: 0.943. A princípio, tais métricas são muito semelhantes ao do modelo de aprendizado *Adaboost*, igualmente positivas, no tocante a predição de resultados, todavia, faz-se mister ressaltar que o que se busca aqui é também a explicabilidade de tais predições.

As diferenças começam a surgir, justamente quando se observa os dados referentes ao *Explain Model*, onde se verifica a distribuição dos atributos e gradação da importância para predição. A seguir, temos a experimentação com o valor-alvo (*target-class*) procedente ao pedido da ação envolvendo companhias aéreas:

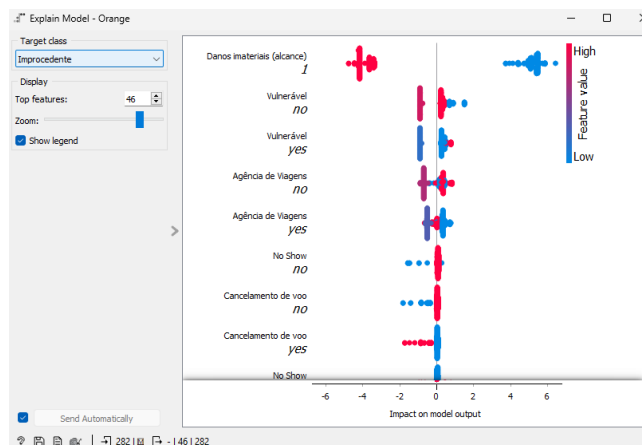
**Figura 48 Gradient Boosting – Explain Model**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Notadamente, verificamos que apesar de nos ser fornecido uma gradação de relevância para predição com o algoritmo de aprendizado *Gradient Boost*, apresenta nítida ausência de identificação da polaridade de impacto de tais atributos no modelo preditivo. Isto é, não consegue prever com precisão o que influencia positivamente e negativamente para o resultado ser procedente ao pedido da ação. O mesmo não ocorre quando o valor-alvo é o resultado improcedente da ação, conforme se observa no trecho a seguir:

**Figura 49 Ilustração do Gradient Boosting – Explain Model**



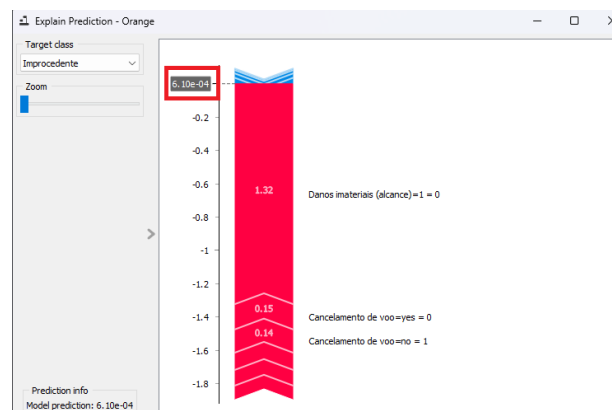
Nota: Imagem do Autor. Extraída de *Orange Canvas*.



No caso em tela, os atributos mais relevantes identificados para o resultado foram a presença de danos imateriais no intervalo 1 (R\$ 0,00), a vulnerabilidade e a agência de viagens. Ainda assim, os dados não parecem estar suficientemente claros para propiciar uma devida explicabilidade das predições, o que nos leva a crer que talvez esse não seja o melhor algoritmo para tratar do tema, dada a dimensão da base de dados e o tipo de resultado esperado.

Inclusive, no tocante a análise preditiva de resultado procedente ao pedido da ação, ou não; as predições foram ainda mais imprecisas. Isso porque, o modelo não soube aferir o valor-alvo apresentando inconsistências na sua tabela de atribuições, de acordo com a imagem a seguir:

**Figura 50 Ilustração do *Gradient Boosting – Explain Prediction***



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Conclui-se, portanto, que para fins de estudo de modelos explicativos de predições de IA aplicadas ao judiciário, que o algoritmo de aprendizado de máquina *Gradient Boost*, ao menos o que consta na plataforma do *Orange Canvas*, não pode apresentar resultados conclusivos acerca da explicabilidade de suas predições. A seguir, temos a análise de experimentos com outros dois algoritmos, a dizer: a) Redes Neurais (*Neural Network*) e b) Regressão Logística (*Logistic Regression*).

#### 4.4.3 Redes Neurais (*Neural Network*)

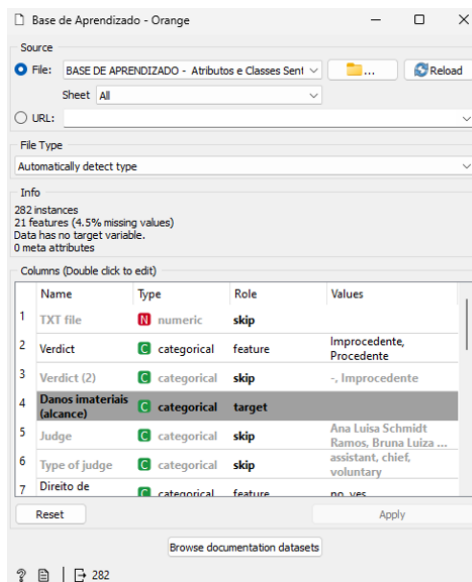
As redes neurais são sistemas de computação com nós interconectados que

funcionam como os neurônios do cérebro humano. Usando algoritmos, elas podem reconhecer padrões escondidos e correlações em dados brutos, agrupá-los e classificá-los, e – com o tempo – aprender e melhorar continuamente<sup>181</sup>.

No *Orange Canvas*, de igual modo, elas funcionam como um modelo de aprendizado de máquina baseado em uma rede de neurônios artificiais interconectados, inspirados no funcionamento do cérebro humano. Tal plataforma oferece suporte para criar, treinar e utilizar redes neurais em tarefas de classificação, regressão e outras aplicações.

De maneira resumida, cada neurônio é responsável por realizar cálculos e transferir informações para os neurônios nas camadas subsequentes. As redes neurais podem ter várias camadas, incluindo camadas de entrada, camadas ocultas e camadas de saída. Após o treinamento, é possível utilizar a rede neural para fazer previsões em novos dados. Isso sem contar o fato de que o *Orange Canvas* fornece recursos para avaliar o desempenho da rede neural, como métricas de precisão e gráficos de aprendizado.

**Figura 51 Ilustração de Redes Neurais – Base de Aprendizado**

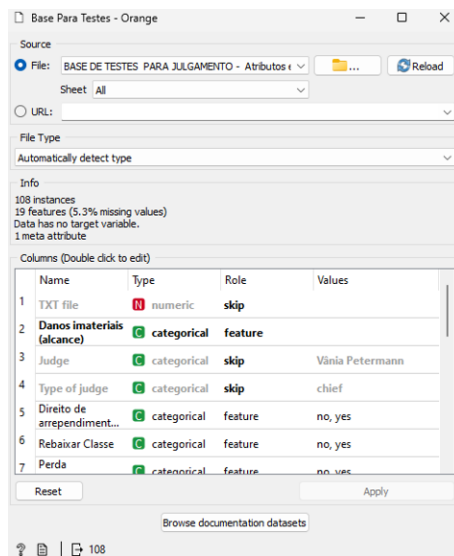


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

<sup>181</sup> Uma rede neural simples inclui uma camada de entrada, outra de saída (ou alvo) e, entre elas, uma camada oculta. As camadas são conectadas através de nós e essas conexões formam uma "rede" – a rede neural – de nós interconectados (Redes Neurais, O que são e qual sua importância? SAS. p.4-5. 2023. Disponível em: <[https://www.sas.com/pt\\_br/insights/analytics/neural-networks.html](https://www.sas.com/pt_br/insights/analytics/neural-networks.html)>).

Cabe mencionar que, para fins ritualísticos que se trata de uma base de aprendizado com 282 (duzentos e oitenta e dois processos), onde se exclui tão somente os atributos estruturados referentes ao nome do juiz ou juíza, tipo de juiz (assistente, chefe ou voluntário), assim como número da sentença contido em arquivo TXT. A seguir, temos a base de testes utilizada para a experimentação:

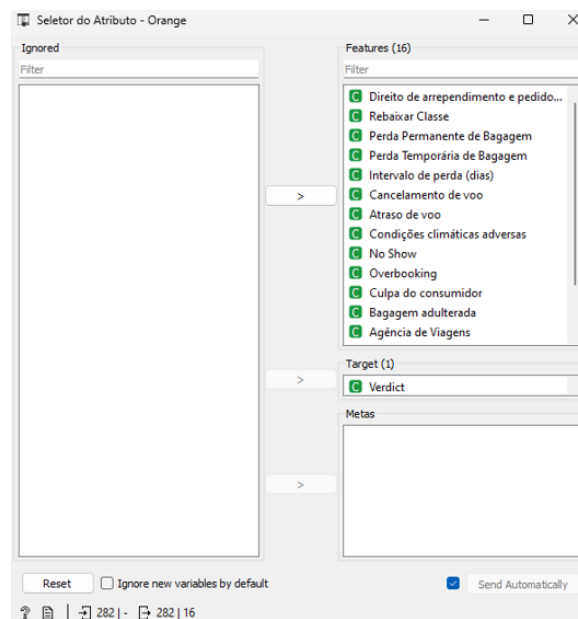
**Figura 52 Ilustração de Redes Neurais – Base de Testes**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Igualmente, temos uma base de teste para julgamento, onde se busca prever o resultado procedente, ou não, ao pedido da ação. Para além das informações anteriormente mencionadas, temos aqui também a omissão do resultado da sentença, que, por óbvio, se busca prever com exatidão. Essa base de testes, difere da base referente a valores da sentença, que ao contrário desta, omite o valor a ser percebido pela parte autora ao fim do processo.

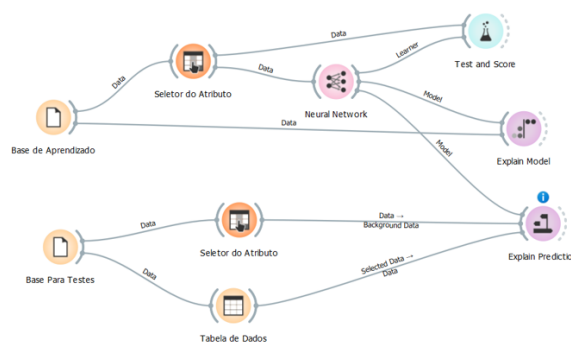
**Figura 53 Ilustração de Redes Neurais – Seletor do Atributo**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

A imagem acima, ilustra a ferramenta de seleção de atributos, ou *feature selection*, que no fluxo de trabalho seleciona qual o valor-alvo (*target*) que será analisado. Como mencionamos anteriormente, o que se busca aferir inicialmente é o resultado da sentença, isto é, procedente, ou não, ao pedido da ação. Para fins de esclarecimento, logo abaixo, temos uma melhor compreensão do fluxo de trabalho referente as redes neurais em *Orange Canvas*:

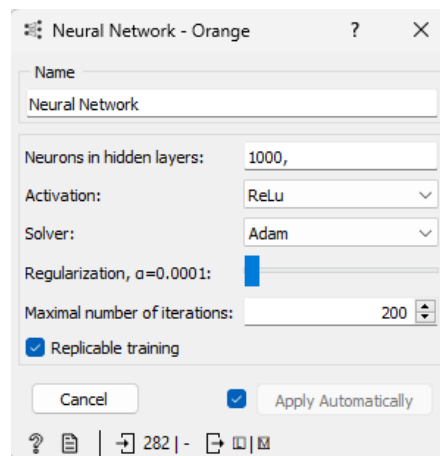
**Figura 54 Ilustração de Redes Neurais – Fluxo de Trabalho**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Logo após a seleção de atributo, temos a escolha do algoritmo de aprendizado de máquina, que no caso em tela é o das redes neurais. Apesar de parecer complexa, essa etapa é muito mais prática do que a estruturação de dados. A ferramenta fornece ainda opções de definição do usuário referente ao número de camadas, interações e conexões entre a rede neural artificial.

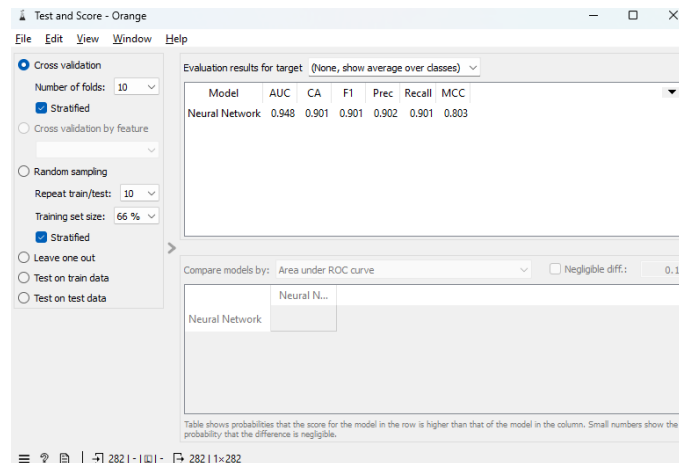
**Figura 55 Ilustração de Redes Neurais – Algoritmo**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Em seguida, é aplicada a observação do *Test and Score*, que permite verificar se as métricas de precisão, acurácia e *recall*. Tais informações permitem identificar se as margens de erro estão dentro de um intervalo razoável que permita prosseguir com os estudos e a experimentação. Os valores obtidos com o algoritmo de redes neurais para a predição de resultado da sentença, estão dispostos a seguir:

Figura 56 Ilustração de Redes Neurais – *Test and Score*

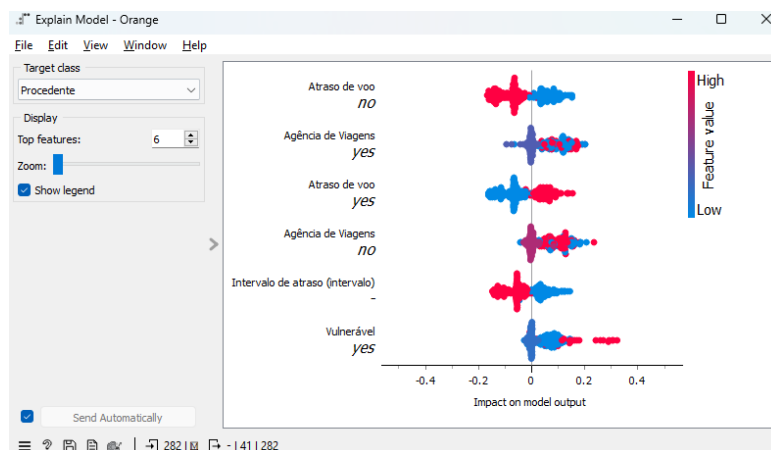


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Dessa maneira, temos que o modelo de redes neurais atingiu as seguintes métricas: a) *AUC*: 0.948; b) *CA* 0,901; c) *F1*: 0.901; d) *Precision*: 0.902; e) *Recall*: 0.901. A princípio, apesar de serem métricas boas, isso por si só, não garante a eficiência do modelo. Por isso torna-se fundamental a observação do seu comportamento diante de modelos explicativos.

Após a devida verificação das métricas de acurácia, temos no fluxo de trabalho das redes neurais a ferramenta de *Explain Model* que permite, inicialmente, a análise de principais atributos levando em consideração na predição de procedência, ou não do pedido da ação em base de testes.

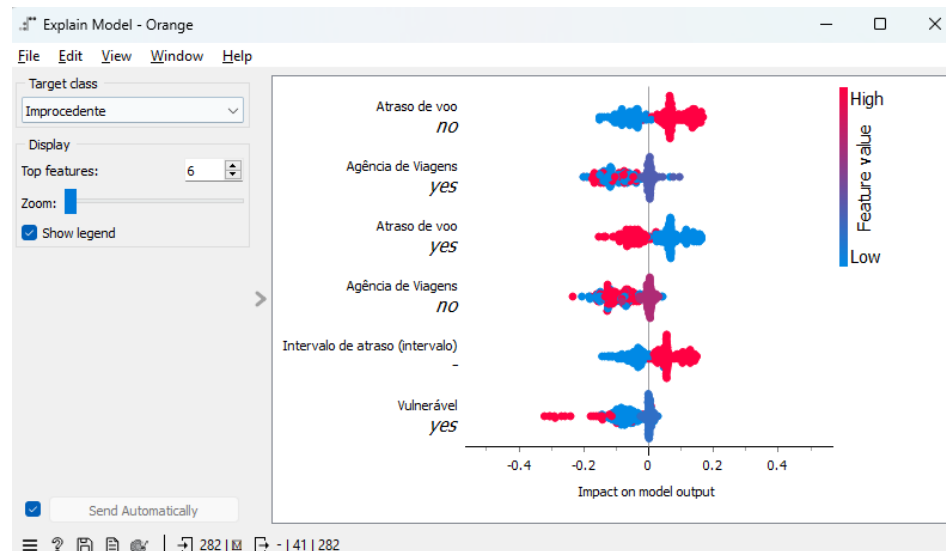
Figura 57 Ilustração de Redes Neurais – *Explain Model*



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Incrivelmente, apesar da repetição de alguns, os primeiros 4 (quatro) atributos tomados com referência são atraso de voo, agência de viagens, intervalo de atraso e vulnerabilidade. Percebe-se que a ausência de atraso de voo gera uma influência negativa para a predição do resultado procedente do pedido, indicada pelos pontos vermelhos à esquerda da origem. Por outro lado, a presença de agência de viagens, parece influenciar positivamente o modelo, diante da concentração de indistinta de pontos ao lado direito da origem. Por fim, a presença de vulneráveis também importa para a predição do resultado procedente do pedido. A seguir, temos a análise do valor-alvo de improcedência da ação:

**Figura 58 Ilustração de Redes Neurais – Explain Model**



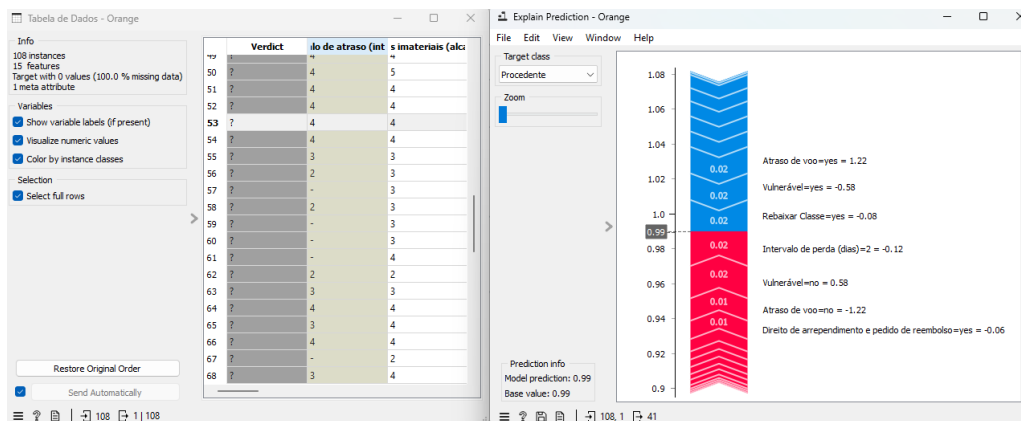
Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Nota-se aqui, que os principais atributos para a predição do resultado improcedente da ação são atraso de voo, agência de viagens, intervalo de atraso e vulnerabilidade. Todavia, os valores não se apresentam tão claros e interpretáveis como anteriormente. Isso porque, existem repetições de variáveis em suas vertentes positivas e negativas. Além disso, não existe uma predominância evidente de pontos em um dos lados da origem. Com efeito, ao contrário da análise anterior, esta não se apresenta tão interpretável.

Mais adiante, temos a análise da ferramenta de *Explain Prediction*, que traz

elementos explicativos de um valor-alvo (*target*) de predição (procedente ou improcedente). Seu funcionamento se dá mediante a escolha de um caso dentro os 108 (cento e oito) casos dispostos na base de testes do algoritmo de redes neurais. A seguir temos a imagem de uma tabela de dados à direita e a ferramenta de *Explain Prediction* à esquerda:

**Figura 59 Ilustração de Redes Neurais – Tabela de Dados e *Explain Prediction***

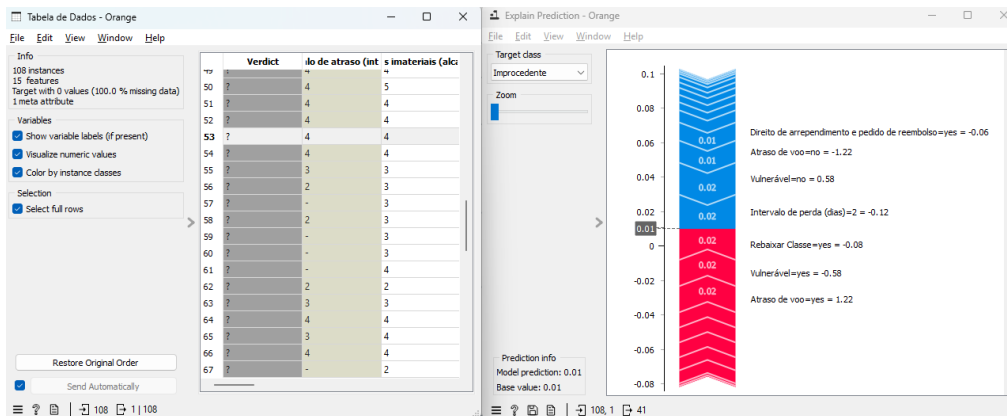


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Conforme verificado na ilustração acima, foi escolhido o caso número 53 (cinquenta e três) para ser verificado como valor-alvo as chances de procedência do pedido da ação. O que se identifica, logo ao lado é que o modelo identifica como principais variáveis o atraso de voo e intervalo de dias de atraso, a vulnerabilidade e a solicitação de pedido de reembolso. Destarte, considera que as chances de procedência do pedido estão em 99% (noventa e nove por cento). O resultado também é coerente com os dados obtidos com o valor-alvo invertido, isto é, quando se busca as chances de improcedência do pedido da ação:

**Figura 60 Ilustração de Redes Neurais – Tabela de Dados e *Explain Prediction***





Nota: Imagem do Autor. Extraída de *Orange Canvas*.

As chances de improcedência da ação, remanescem em 0,01% (zero virgula zero um por cento), ao passo que as variáveis que contribuem para a inferência do modelo permanecem as mesmas. Indo além, ao identificar a planilha em que se dispõe as informações não reveladas na base de teste, é possível aferir que a predição acerta no resultado, conforme se observa no trecho abaixo:

Figura 61 Ilustração de Planilha de Dados em CSV

1	TXT file		Danos imateriais
50		713	4
51		715	5
52		717	4
53		721	4
54		722	4

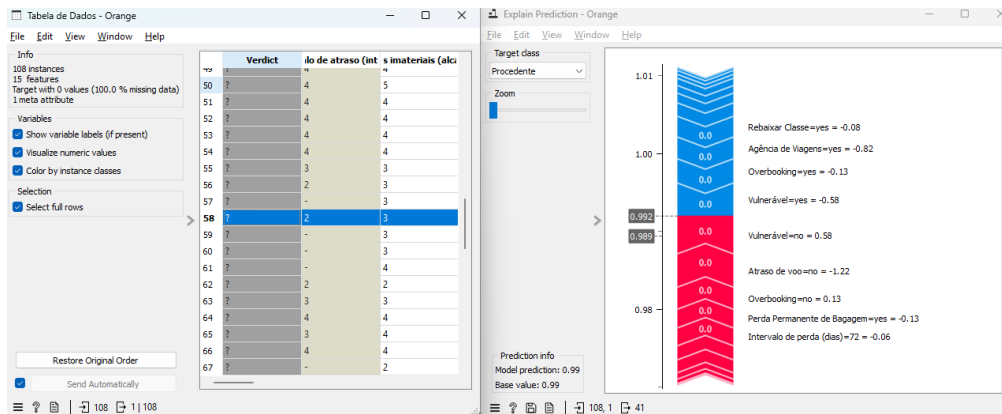
  

	A	B	C	D	E
1	TXT file	Verdict	Verdict (2)	Immaterial damage	Immaterial damage
724		717	partly_founded	-	R\$ 7.000,00
725		718	not_founded	-	R\$ 0,00
726		719	not_founded	-	R\$ 0,00
727		720	partly_founded	-	R\$ 0,00
728		721	partly_founded	-	R\$ 8.000,00
729		722	well_founded	-	R\$ 7.000,00
730		723	not_founded	-	R\$ 0,00
731		724	partly_founded	-	R\$ 7.000,00
732		725	partly_founded	-	R\$ 4.000,00

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Para fins de esclarecimento, cumpre informar que os conceitos de procedente (*well founded*) e parcialmente procedente (*partly founded*) foram unificados em um único atributo chamado procedente; ao passo que improcedente com e sem resolução de mérito foram unificados no atributo improcedente. A seguir temos a verificação do caso 58 junto a ferramenta de *Explain Prediction*:

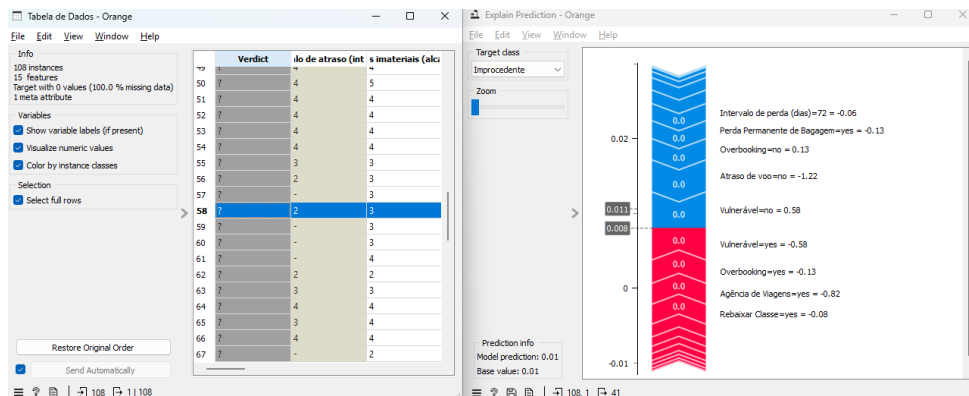
**Figura 62 Ilustração de Redes Neurais – Tabela de Dados e *Explain Prediction***



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Outrossim, as chances de procedência da ação são identificadas entre 99,2% (noventa e nove vírgula dois por cento) e 98,9% (noventa e oito vírgula nove por cento), tendo como principal variável contributiva para o resultado a vulnerabilidade, o atraso de voo, o *overbooking*, perda permanente de bagagem e o intervalo de dias de atraso. O resultado se consolida com a análise do valor-alvo improcedência do pedido da ação:

**Figura 63 Ilustração de Redes Neurais – Tabela de Dados e *Explain Prediction***



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Com efeito, temos que para a improcedência do caso número 58 (cinquenta e oito), as chances estão entre 0,011% (zero vírgula zero onze por cento) e 0,08% (zero vírgula oito por cento). Contribuindo negativamente para a improcedência do pedido da ação, permanece a variável da vulnerabilidade, *overbooking*, atraso de voo, perda permanente de bagagem. Ademais, em consulta à planilha em que se dispõe as informações não reveladas na base de teste, é possível aferir que, mais uma vez a predição acerta no



testes:

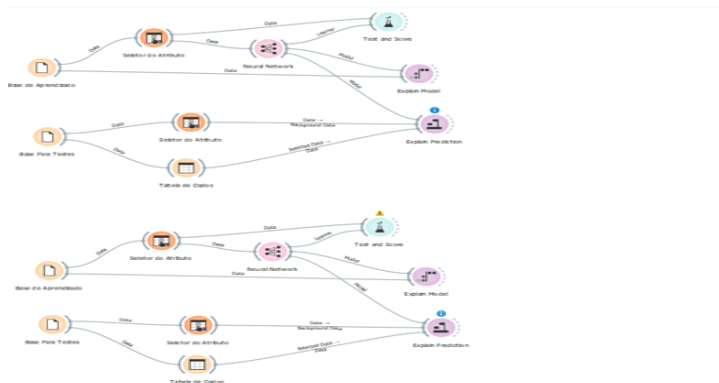
**Figura 66 Ilustração de Redes Neurais – Tabela de Dados e *Explain Prediction***

1	TXT file	Danos imateriai	A	B	C	D		
99	1139		1	1	TXT file	Verdict	Verdict (2)	Immaterial dam
100	1085		1	866	860	partly_founded	-	RS 4.500,00
101	864		1	867	861	partly_founded	-	RS 4.000,00
102	846		1	868	862	partly_founded	-	RS 6.000,00
103	847		1	869	863	partly_founded	-	RS 8.500,00
104	843		1	870	864	not_founded	-	RS 0,00
			1	871	865	partly_founded	-	RS 4.000,00
			1	872	866	partly_founded	-	RS 6.000,00

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Dessa vez, a predição foi incorreta quanto ao resultado de decisão judicial ao considerar maiores as chances de procedência, muito embora, a margem de erro tenha sido razoável. Por fim, temos a análise das predições e do modelo explicativo quanto ao valor de indenização a ser pago. Isto é, para além do resultado da sentença, buscou-se definir uma explicabilidade para previsibilidade em valor de ação mantida nos intervalos de: a) (1) R\$ 0,00 ; b) (2) R\$ 1,00 - R\$ 2.000,00; c) (3) R\$ 2.001,00 - R\$ 5.000,00; d) (4) R\$ 5.001,00 - R\$ 8.000,00; e) (5) R\$ 8.001,00 - R\$ 25.000,00.

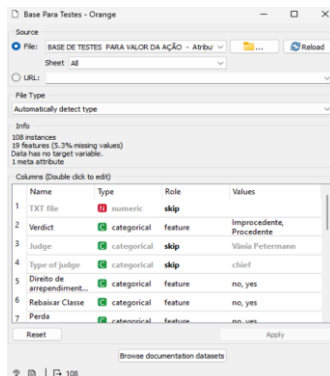
**Figura 67 Ilustração de Redes Neurais – Fluxo de Trabalho**



Nota: Imagens do Autor. Extraídas de *Orange Canvas*.

Destarte, para realização desse experimento, conforme mencionado anteriormente, foi criada uma base de testes específica para identificação do valor da ação. Ou seja, ao contrário da base de teste para o julgamento, aqui são omitidas informações quanto ao valor da indenização por danos morais, bem como demais informações que não contribuem para a elucidação como o nome do juiz ou juíza, tipo de juiz (assistente, chefe ou voluntário) e o número do processo em arquivo TXT:

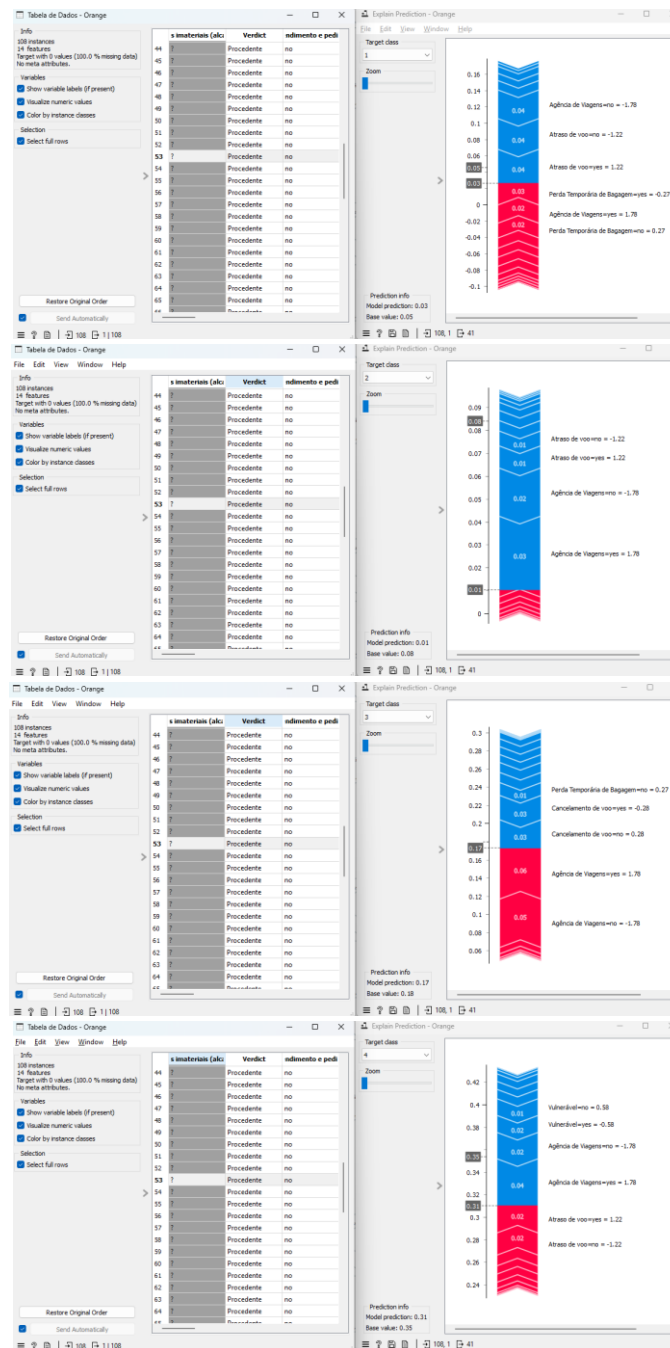
Figura 68 Ilustração de Redes Neurais – Base de Testes

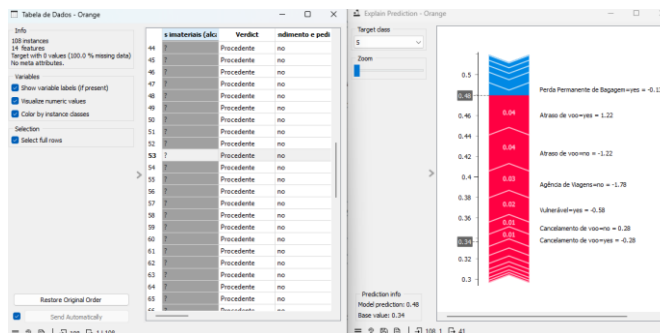


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

O fluxo de trabalho é idêntico ao utilizado para identificar o resultado julgamento, alterando tão somente a base de dados. A seguir temos o caso número 32 (trinta e dois) para a verificação através do *Explain Prediction*, da predição do resultado, assim como também dos principais atributos levados em consideração ao auferir a quantia de deverá ser paga em caso de procedência da ação:

Figura 69 Ilustrações de Redes Neurais – Tabela de Dados e *Explain Prediction*





Nota: Imagens do Autor. Extraídas de *Orange Canvas*.

Observa-se, que o modelo explicativo que fez o uso de redes neurais artificiais, identificou que em caso de procedência da ação o valor indenizatório possui 48% (quarenta e oito por cento) de chances de estar no intervalo 5 (R\$ 8.001,00 – R\$ 25.000,00), 31% (trinta e um por cento) de chances de estar no intervalo 4 (R\$ 5.001,00 – R\$ 8.000,00), seguido por 17% (dezessete por cento) de chances de estar no intervalo 3 (R\$ 2.001,00 – R\$ 5.000,00), 0.03% (zero vírgula zero três por cento) de chances de estar no intervalo 1 (R\$ 0,00) e, por fim, 0.01% (zero vírgula zero um por cento) de chances de estar no intervalo 2 (R\$ 1,00 – R\$ 2.000,00). Entre os principais atributos levados em consideração pelo modelo temos a existência de pessoa vulnerável, cancelamento e o atraso de voo. A seguir temos a planilha com as informações omitidas da base de testes, para fins de checagem do resultado obtido:

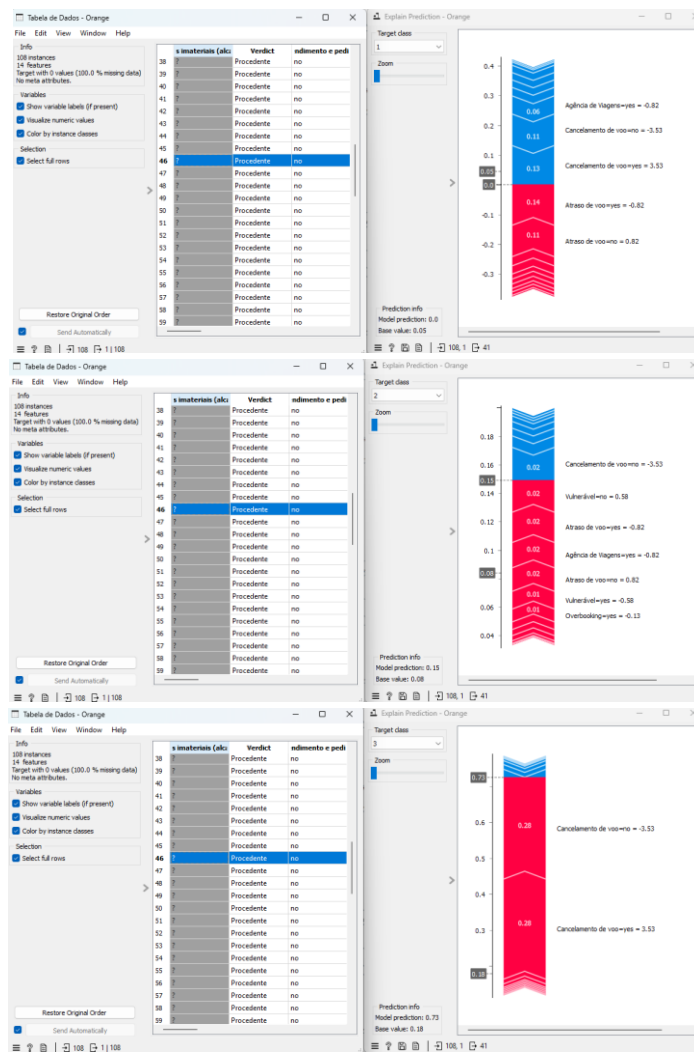
**Figura 70 Ilustração de Planilha de Dados em CSV**

	A	B		A	B	C	D
1	TXT file	Verdict	1	TXT file	Verdict	Verdict (2)	Immaterial dam
28	682	Procedente	690	684	partly_founded	-	R\$ 8.000,00
29	683	Procedente	691	685	partly_founded	-	R\$ 10.000,00
30	684	Procedente	692	686	partly_founded	-	R\$ 6.000,00
31	685	Procedente	693	687	partly_founded	-	R\$ 3.000,00
32	686	Procedente	694	688	partly_founded	-	R\$ 5.000,00
33	687	Procedente					

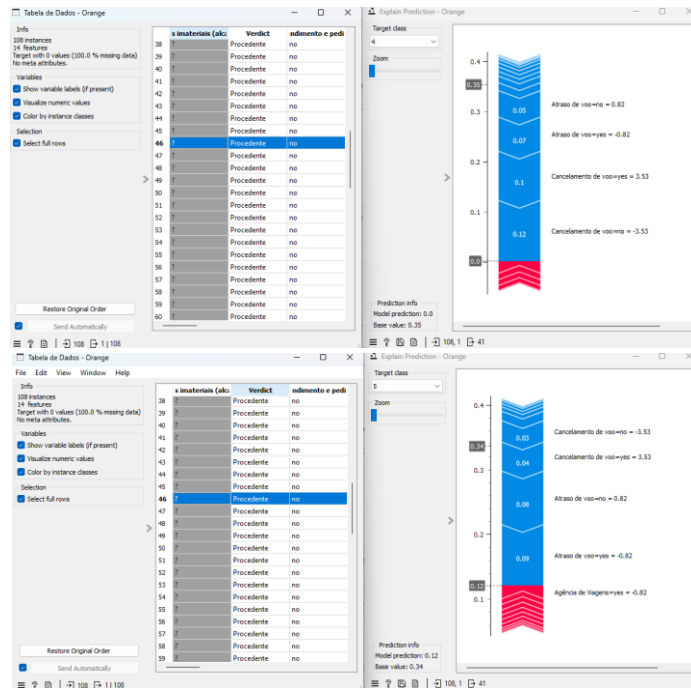
Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Apesar do modelo utilizado não ter conseguido identificar qual a maior chance dentre os cinco intervalos selecionados, acertou ao definir que a indenização estaria acima dos R\$ 5.000,00 (cinco mil reais), dado que se somado as chances dos intervalos 4 e 5, temos 79% (setenta e nove por cento) de chances do resultado ser um dos dois. Ainda assim, é preciso saber como se comporta em outro caso. Para tanto, a seguir temos a análise do caso número 46 (quarenta e seis), para os 5 (cinco) intervalos ou cenários:

Figura 71 Ilustrações de Redes Neurais – Tabela de Dados e *Explain Prediction*







Nota: Imagens do Autor. Extraídas de *Orange Canvas*.

Identifica-se como principais atributos levados em consideração para a predição à vulnerabilidade, agência de viagens, *overbooking*, o atraso e cancelamento de voo. Ao contrário do exemplo anterior, aqui temos uma predominância de intervalo de valores para a indenização em caso de procedência da ação. As chances da indenização estar no intervalo 3 (R\$ 2.001,00 – R\$ 5.000,00) estão 73% (setenta e três por cento), ao passo que as chances de estar no intervalo 2 (R\$ 1,00 – R\$ 2.000,00) são de 15% (quinze por cento), seguido pelo intervalo 5 (R\$ 8.001,00 - R\$ 25.000,00) com 12%. Todavia, é preciso realizar a consulta em planilha com as informações omitidas na base de testes para auferir o valor da ação:

**Figura 72 Ilustrações de Planilha em CSV**

1	TXT file	Verdict	A	B	C	D	E
43	704	Procedente	710	703	partly_founded	-	R\$ 7.000,00
44	705	Procedente	711	704	partly_founded	-	R\$ 5.000,00
45	706	Procedente	712	705	partly_founded	-	R\$ 7.000,00
46	708	Procedente	713	706	partly_founded	-	R\$ 6.000,00
47	710	Procedente	714	707	partly_founded	-	R\$ 0,00
			715	708	partly_founded	-	R\$ 8.000,00
			716	709	partly_founded	-	R\$ 10.000,00

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Apesar da incrível assertividade quanto ao resultado da procedência, ou não da ação; o mesmo não acontece com a predição do valor da ação que no caso número 46 (quarenta e seis) erra ao prever que o valor da indenização em caso de procedência do

pedido. O modelo prevê que o resultado da ação em caso de procedência do pedido deve ser o intervalo 3 (R\$ 2.001,00 – R\$ 5.000,00), contudo o julgamento definiu que o valor da indenização deve estar no intervalo 4 (R\$ 5.001,00 - R\$ 8.000,00).

Em resumo, podemos considerar que o algoritmo de redes neurais para aprendizado de máquina com a dimensão de base de dados atual, cumpre parcialmente o objetivo do estudo. Isso porque, atende com acurácia devida e apresenta explicações claras para a procedência, ou não da ação; todavia, não persiste tal acurácia no tocante ao valor indenizatório em caso de procedência da ação, como vimos nos exemplos anteriores. A seguir, temos a análise do experimento com Regressão Logística (*Logistic Regression*).

#### 4.4.4 Regressão Logística (*Logistic Regression*)

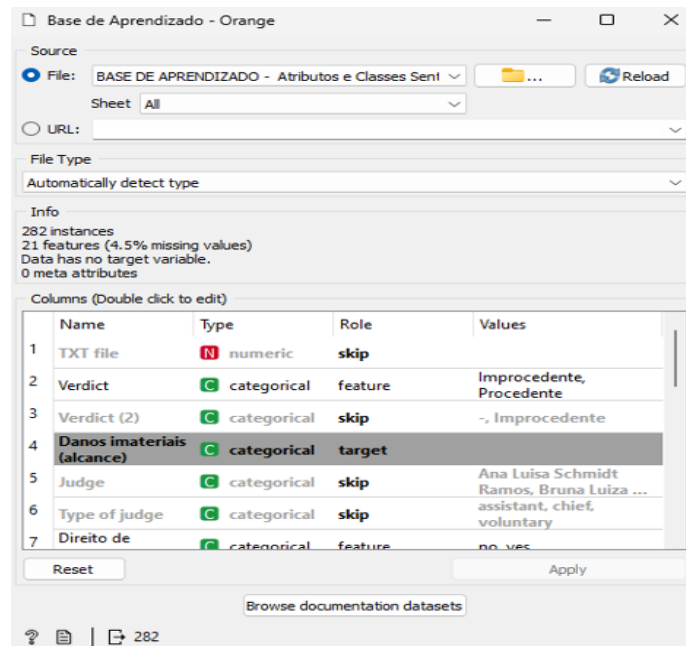
A Regressão Logística é um algoritmo de aprendizado de máquina utilizado para problemas de classificação binária<sup>182</sup>. No *Orange Canvas*, a Regressão Logística está disponível como um dos algoritmos de aprendizado de máquina pré-implementados. Ela é amplamente utilizada quando o objetivo é prever a probabilidade de um exemplo pertencer a uma das duas classes possíveis. Inclusive, utiliza uma função logística para modelar a relação entre as características dos dados e a probabilidade de pertencer a uma determinada classe.

Ou seja, a Regressão Logística é conhecida por ser um algoritmo interpretável e eficiente, que pode ser facilmente aplicado a problemas de classificação binária. É amplamente utilizado em várias áreas, como direito, medicina, finanças e ciência de dados, onde a interpretabilidade e a precisão são importantes.

#### **Figura 73 Ilustração de Regressão Logística - Base de Aprendizado**

---

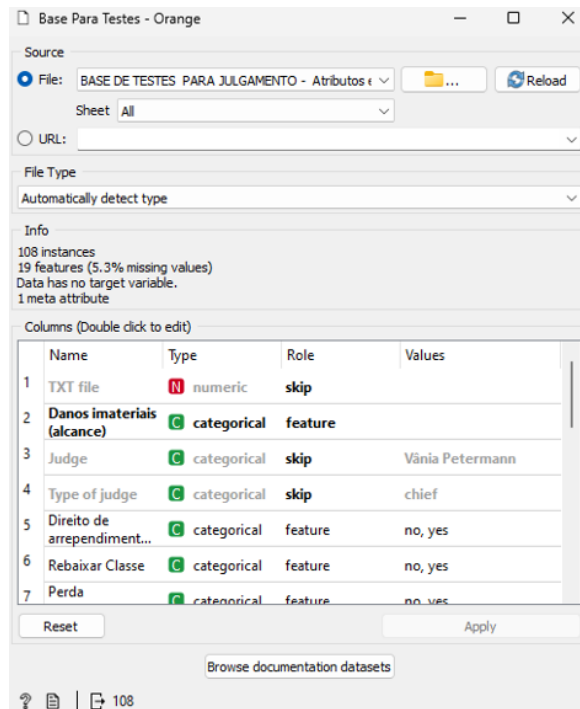
<sup>182</sup> Esse tipo de modelo estatístico (também conhecido como modelo logit) é frequentemente usado para classificação e análise preditiva. A regressão logística estima a probabilidade de ocorrência de um evento, como um voto, com base em um determinado conjunto de dados de variáveis independentes. Como o resultado é uma probabilidade, a variável dependente é limitada entre 0 e 1 (O que é regressão logística. IBM. 2023. p.1. Disponível em:<<https://www.ibm.com/br-pt/topics/logistic-regression>>).



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Destarte, são utilizados para o experimento do regressão logística dados coletados e estruturados de 282 (duzentos e oitenta e dois processos), em que se exclui da análise tão somente os atributos estruturados referentes ao nome do juiz ou juíza, tipo de juiz (assistente, chefe ou voluntário), assim como número da sentença contido em arquivo *TXT*. Para fins de esclarecimento, temos a seguir, a base de testes utilizada para a experimentação:

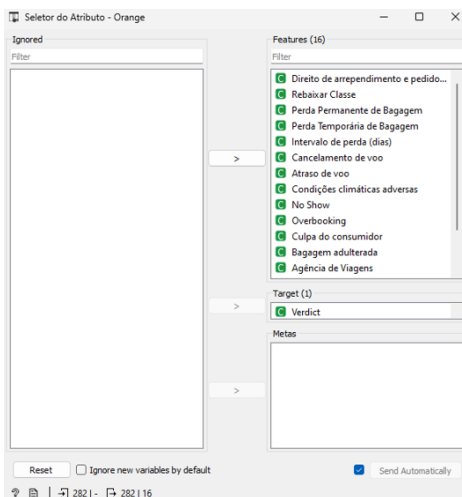
**Figura 74 Ilustração de Regressão Logística - Base de Aprendizado**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

É justamente através da base de teste para julgamento que se busca prever o resultado precedente, ou não, ao pedido da ação. Além disso, é excluída da base de teste, as informações do resultado da sentença para que possa ser avaliada a exatidão da predição realizada pela IA. Ademais, é importante mencionar que, a diferença entre a base de dados de aprendizado e a base de dados de testes reside no fato de que apenas a base de testes omite o resultado da sentença, isto é, precedente, ou não ao pedido da ação.

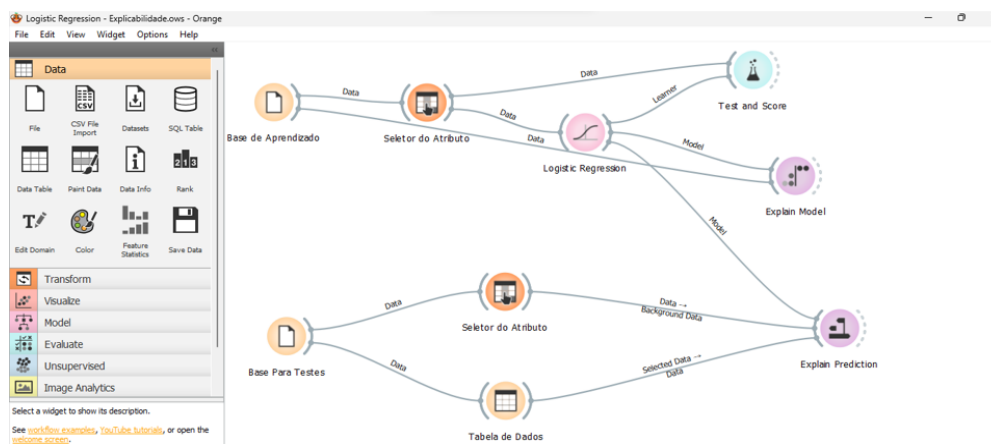
**Figura 75 Ilustração de Regressão Logística – Seletor do Atributo**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Como vimos, a ferramenta de seleção de atributos, ou *feature selection*, determina no fluxo de trabalho qual o valor-alvo (*target*) que será analisado. Nesse sentido, inicialmente, o que se busca aferir é o resultado da sentença. Isso se torna de mais fácil compreensão através fluxo de trabalho referente a regressão logística em *Orange Canvas*:

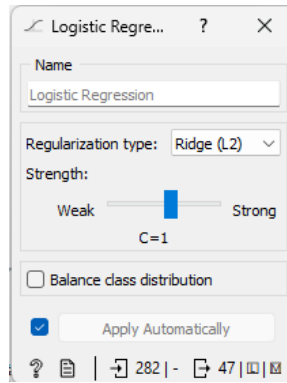
**Figura 76 Ilustração de Regressão Logística – Fluxo de Trabalho**



Nota: Imagem do Autor. Extraída de *Orange Canvas*

Em seguida, após a seleção de atributo, temos a escolha do algoritmo de aprendizado de máquina, que no experimento em questão é o da regressão logística. Válido mencionar que, essa etapa é muito mais prática do que a estruturação de dados. A ferramenta fornece ainda diversas opções de definição do usuário:

**Figura 77 Ilustração de Regressão Logística – Algoritmo**

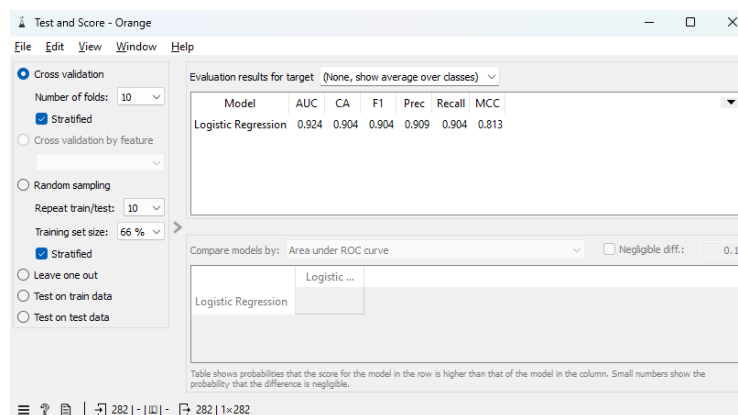


Nota: Imagem do Autor. Extraída de *Orange Canvas*

Nesse sentido, é possível ajustar diferentes parâmetros, como a regularização (para controlar o *overfitting*), a taxa de aprendizado e o tipo de regularização (L1 ou L2). Esses parâmetros permitem controlar a complexidade do modelo e ajustar o seu desempenho. Ademais, trata-se de um algoritmo linear, o que significa que assume uma relação linear entre as características dos dados e a probabilidade de classificação. No entanto, o *Orange Canvas* permite usar transformações e combinações de características para capturar relações não-lineares, aumentando a flexibilidade do modelo.

Superada a etapa anterior, chega a vez da inserção do mecanismo de *Test and Score* que auxilia revelando as métricas de precisão, acurácia e *recall*. Elas indicam se os erros são toleráveis para seguir com os estudos e a experimentação. O algoritmo de regressão logística para prever o resultado da sentença teve os valores abaixo:

**Figura 78 Ilustração de Regressão Logística – Test and Score**



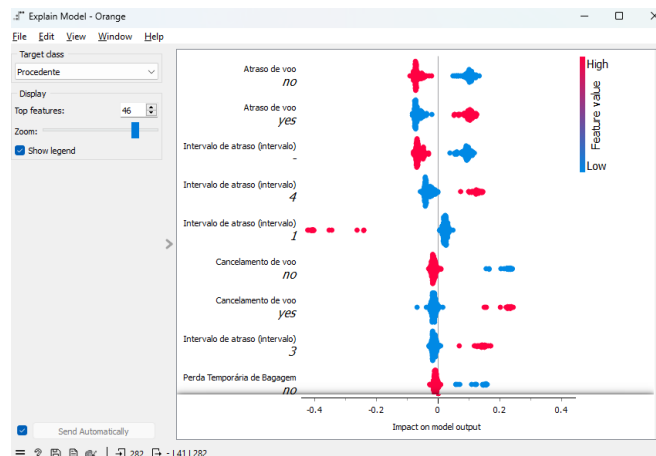
Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Portanto, temos que o modelo de regressão logística atingiu as seguintes métricas:

a) *AUC*: 0.924; b) *CA* 0,904; c) *FI*: 0.904; d) *Precision*: 0.909; e) *Recall*: 0.904. Essas métricas são boas, mas não provam a eficiência do modelo. Por isso, é essencial a observação do seu comportamento diante de modelos explicativos.

Sendo assim, após a devida verificação das métricas de acurácia, temos no fluxo de trabalho da regressão logística a ferramenta de *Explain Model*. Ela permite, inicialmente, a análise de atributos (*features*) relevantes levando em consideração na predição de procedência, ou não do pedido da ação em base de testes, conforme se verifica na ilustração a seguir:

**Figura 79 Ilustração de Regressão Logística – *Explain Model***

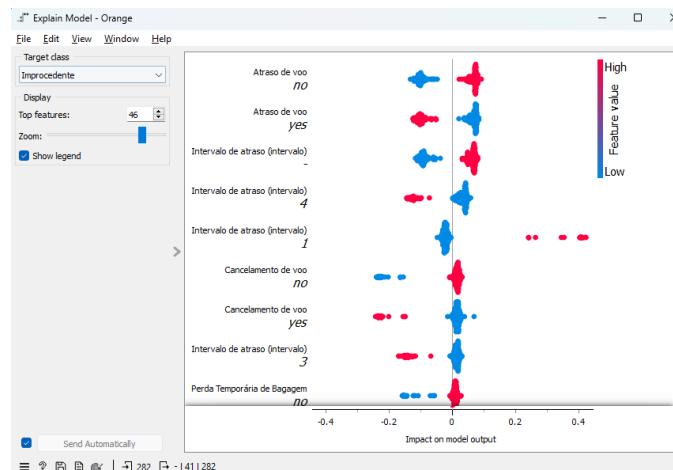


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Surpreendentemente, os primeiros três atributos considerados como referência são o atraso de voo, o intervalo de atraso e cancelamento de voo. A notável concentração de marcadores vermelhos à esquerda do gráfico de negativa de atraso de voo sugere que o modelo interpreta que situações como essa inferindo que influenciam negativamente para o resultado selecionado que é o de procedência da ação. O que acontece na segunda linha é tão somente uma replicação do cenário da primeira linha, porém de modo inverso.

Da mesma forma, podemos observar a presença de diversos pontos vermelhos à direita do gráfico no intervalo de atraso 4 (quatro) e cancelamento de voo. Isso indica que, por outro lado, intervalo de atraso 4 (quatro) e cancelamento de voo têm uma influência positiva no modelo, levando-o a considerar o resultado da ação como procedente. Essa interpretação é ainda mais reforçada quando analisamos o critério de improcedência da ação.

**Figura 80 Ilustração de Regressão Logística – Explain Model**



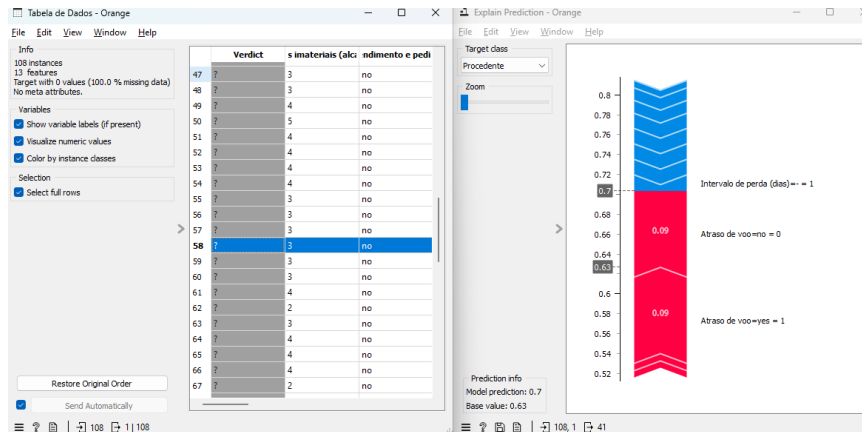
Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Neste ponto, observamos que os marcadores vermelhos associados a negativa de atraso de voo, que anteriormente estavam à esquerda do gráfico, agora se deslocam para a direita. Ao mesmo tempo, os marcadores vermelhos ligados aos casos positivos de atraso de voo se movem para a esquerda do gráfico. Isso reforça a mesma interpretação de que a ausência de atraso de voo tem uma influência positiva na improcedência do pedido da ação, enquanto que o intervalo de atraso 4 (quatro) tem uma influência negativa na improcedência.

Avançando no experimento, cumpre realizar uma análise detalhada da ferramenta *Explain Prediction*, que fornece informações esclarecedoras sobre a predição de resultado precedente, ou não, ao pedido da ação. Essa funcionalidade opera através da seleção de um caso a partir de um conjunto de 108 (cento e oito) casos disponíveis na base de testes do nosso algoritmo de regressão logística. Na sequência, temos uma imagem com uma tabela de dados à direita e a ferramenta de *Explain Prediction* à esquerda:



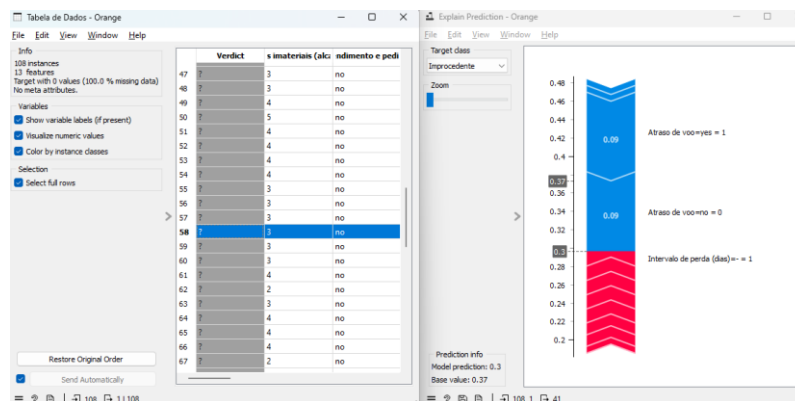
**Figura 81 Ilustração de Regressão Logística - Tabela de Dados e Explain Prediction**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Optou-se por analisar o caso de número 58, a fim de determinar as chances de procedência da petição em questão. Nota-se imediatamente que o modelo destaca como variável-chave os intervalos de perda de bagagem e atraso de voo para avaliar as probabilidades de sucesso, ou seja, de procedência do pedido, que variam de 63% a 70%. Este resultado é coerente mesmo quando investigamos as chances de improcedência do pedido da ação, ou seja, quando buscamos o oposto:

**Figura 82 Ilustração de Regressão Logística – Tabela de Dados e Explain Prediction**



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

As probabilidades de improcedência da ação permanecem entre 30% e 37%, enquanto as variáveis que influenciam a inferência do modelo permanecem inalteradas. Além disso, ao examinarmos a planilha que contém as informações não reveladas na base de testes, constatamos que a previsão se alinha com o resultado, como ilustrado abaixo:

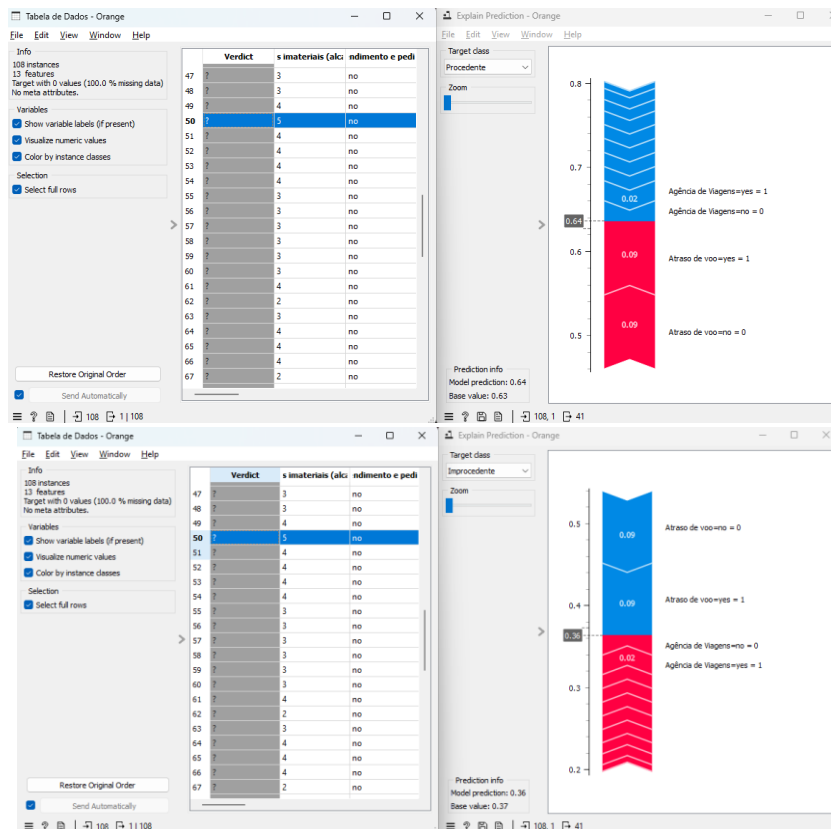
**Figura 83 Ilustração de Planilha em CSV**

1	TXT file	Danos imateriais	1	TXT file	Verdict	Verdict (2)	Immaterial damage
54	722	4	728	721	partly_founded	-	R\$ 8.000,00
55	724	4	729	722	well_founded	-	R\$ 7.000,00
56	725	3	730	723	not_founded	-	R\$ 0,00
57	726	3	731	724	partly_founded	-	R\$ 7.000,00
58	727	3	732	725	partly_founded	-	R\$ 4.000,00
59	729	3	733	726	partly_founded	-	R\$ 5.000,00
60	730	3	734	727	partly_founded	-	R\$ 5.000,00
			735	728	partly_founded	-	R\$ 5.000,00
			736	729	partly_founded	-	R\$ 3.000,00

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Válido destacar que os conceitos de procedente (*well founded*) e parcialmente procedente (*partly founded*) foram unificados em um único atributo chamado procedente; ao passo que improcedente com e sem resolução de mérito foram unificados no atributo improcedente. A seguir, temos a verificação do caso 50 junto a ferramenta de *Explain Prediction*:

**Figura 84 Ilustração de Regressão Logística – Tabela de Dados e *Explain Prediction***



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Novamente, aqui é possível verificar a constância da variável do atraso de voo e agência de viagens, tanto no que concerne ao valor-alvo (*target*) procedente, quanto para improcedente. Ainda assim, as chances de procedência são de 64%, ao passo que as chances de improcedência remanesçam são de 36%. Surpreende ainda mais, quando se realiza a checagem da planilha com as informações suprimidas da base de testes para julgamento:

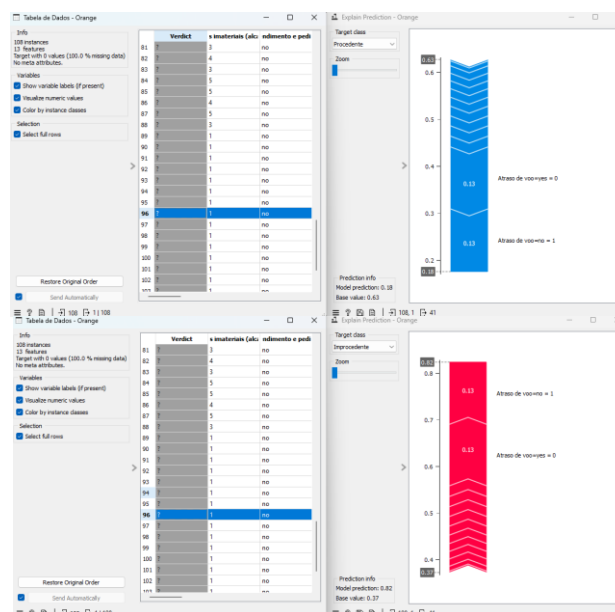
**Figura 85 Ilustração de Planilha em CSV**

1	TXT file	Danos imateriais	A	B	C	D	
47	710	2	1	TXT file	Verdict	Verdict (2)	Immaterial damage
48	711	3	720	713	partly_founded	-	R\$ 7.000,00
49	712	3	721	714	not_founded	-	R\$ 0,00
50	713	4	723	715	partly_founded	-	R\$ 11.000,00
51	715	5	724	716	not_founded	-	R\$ 0,00
52	717	4	725	717	partly_founded	-	R\$ 7.000,00
53	721	4	726	718	not_founded	-	R\$ 0,00
				719	not_founded	-	R\$ 0,00
				720	partly_founded	-	R\$ 0,00

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Até então, temos que considerar que mesmo diante de diversos atributos (*features*), o algoritmo de regressão logística tende a escolher um atributo como o principal fator contributivo para sua predição. Ainda assim, é preciso verificar como ele se comporta a partir de outros casos. Logo abaixo, temos o caso 96 da base de testes para julgamento:

**Figura 86 Ilustração de Regressão Logística – Tabela de Dados e *Explain Prediction***



Nota: Imagem do Autor. Extraída de *Orange Canvas*.

Inicialmente, a ferramenta de *explain prediction* aponta que as chances de procedência do pedido da ação estão entre 18% e 63%, enquanto que as chances de improcedência estão entre 37% até 82%. Outra vez, a variável-chave para a predição do modelo apontada é o atraso de voo sofrido pelo eventual lesado em ação judicial. No entanto, é preciso verificar de que maneira esse modelo se comporta diante as informações omitidas de sua base de testes de julgamento:

**Figura 87 Ilustração de Planilha em CSV.**

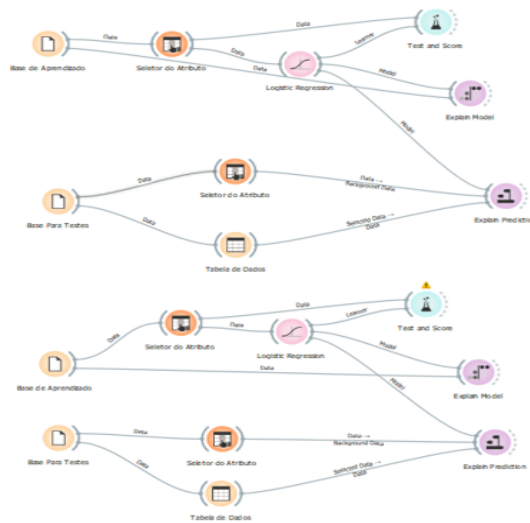
1	TXT file	Danos imateriais		A	B	C	D
92	1141	1	1	TXT file	Verdict	Verdict (2)	Immaterial damag
93	1142	1	1086	1079	partly_founded	-	R\$ 10.000,00
94	1143	1	1087	1080	partly_founded	-	R\$ 0,00
95	1108	1	1088	1081	partly_founded	-	R\$ 0,00
96	1084	1	1089	1082	partly_founded	-	R\$ 5.000,00
97	1037	1	1090	1083	partly_founded	-	R\$ 0,00
98	1038	1	1091	1084	not_founded	-	R\$ 0,00
			1092	1085	not_founded	-	R\$ 0,00
			1093	1086	partly_founded	-	R\$ 10.000,00

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Até aqui, não surpreende o índice de acerto da predição feita pelo algoritmo de regressão logística, mas sim o fato de que a ferramenta de *Explain Prediction*, realmente consegue destacar os critérios utilizados pelo aprendizado de máquina durante a cognição de sua predição.

Ou seja, apesar da predição assertiva ser algo louvável por si só, inova o resultado, quando fornece elementos utilizados pela IA na formação de suas respostas. No que tange à análise das previsões e ao modelo explicativo quanto ao valor da indenização a ser paga, buscou-se não apenas o resultado da sentença, mas também a definição de uma explicação para a previsibilidade do valor da ação, mantendo-se nos seguintes intervalos: a) (1) R\$ 0,00; b) (2) R\$ 1,00 - R\$ 2.000,00; c) (3) R\$ 2.001,00 - R\$ 5.000,00; d) (4) R\$ 5.001,00 - R\$ 8.000,00; e) (5) R\$ 8.001,00 - R\$ 25.000,00.

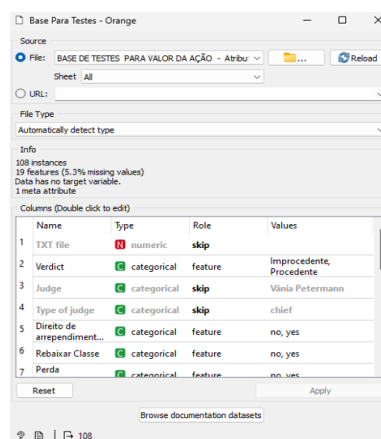
**Figura 88 Ilustração de Regressão Logística – Fluxo de Trabalho**



Nota: Imagens do Autor. Extraídas de *Orange Canvas*.

Destaca-se que, para conduzir este experimento, conforme mencionado anteriormente, foi criada uma base de testes específica para identificar o valor da ação. Diferentemente da base de testes para julgamento, aqui foram omitidas informações relacionadas ao valor da indenização por danos morais, bem como outras informações irrelevantes para a elucidação, como o nome e tipo do juiz (assistente, chefe ou voluntário) e o número do processo, que estão contidas em um arquivo *TXT*.

**Figura 89 Ilustração de Regressão Logística – Base de Testes**

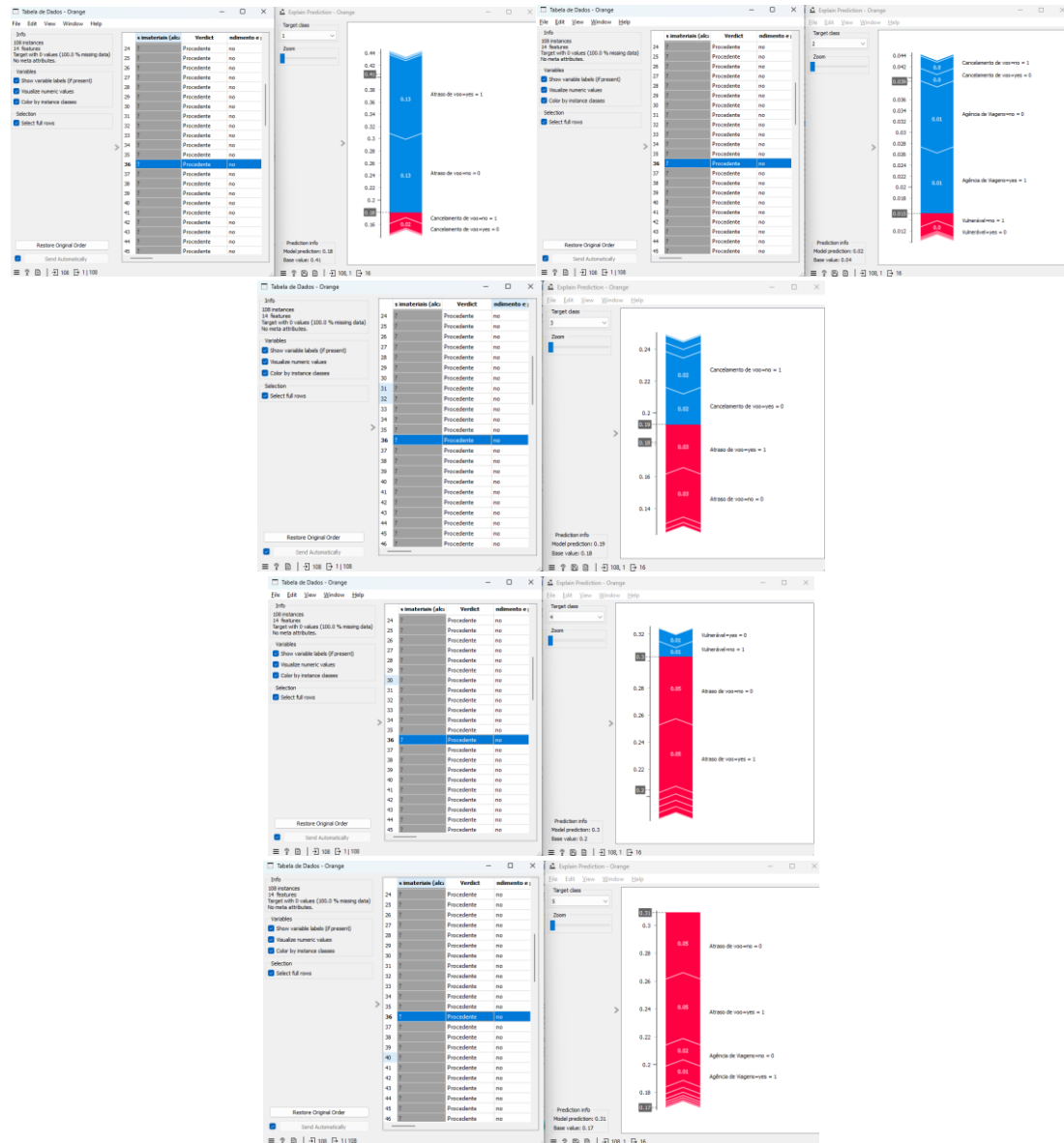


Nota: Imagem do Autor. Extraída de *Orange Canvas*.

O fluxo de trabalho permanece idêntico ao utilizado para identificar o resultado do julgamento, diferenciando-se apenas pela base de dados. A seguir, apresentamos o

caso número 36, no qual realizamos a verificação através do *Explain Prediction*, incluindo a previsão do resultado e os principais atributos considerados para determinar o montante a ser pago em caso de procedência da ação:

**Figura 90 Ilustrações de Regressão Logística – Tabela de Dados e *Explain Prediction***



Nota: Imagens do Autor. Extraídas de *Orange Canvas*.

Verifica-se que o modelo explicativo, baseado na regressão logística, identificou que em caso de procedência da ação, o valor indenizatório possui 30% (trinta por cento) de chances de estar no intervalo 4 (R\$ 5.001,00 – R\$ 8.000,00) e 31% (trinta e um por cento) de chances de estar no intervalo 5 (R\$ 8.001,00 – R\$ 25.000,00). Em seguida, temos 19% (dezenove por cento) de chances de estar no intervalo 3 (R\$ 2.001,00 – R\$

5.000,00), 18% (dezoito por cento) no intervalo 1 (R\$ 0,00) e, por fim, 0,15% (zero vírgula quinze por cento) no intervalo 2 (R\$ 1,00 – R\$ 2.000,00). Entre os principais atributos considerados pelo modelo estão a existência de uma pessoa vulnerável, agência de viagens, cancelamento e o atraso de voo. A planilha a seguir apresenta as informações omitidas da base de testes para verificar o resultado obtido.

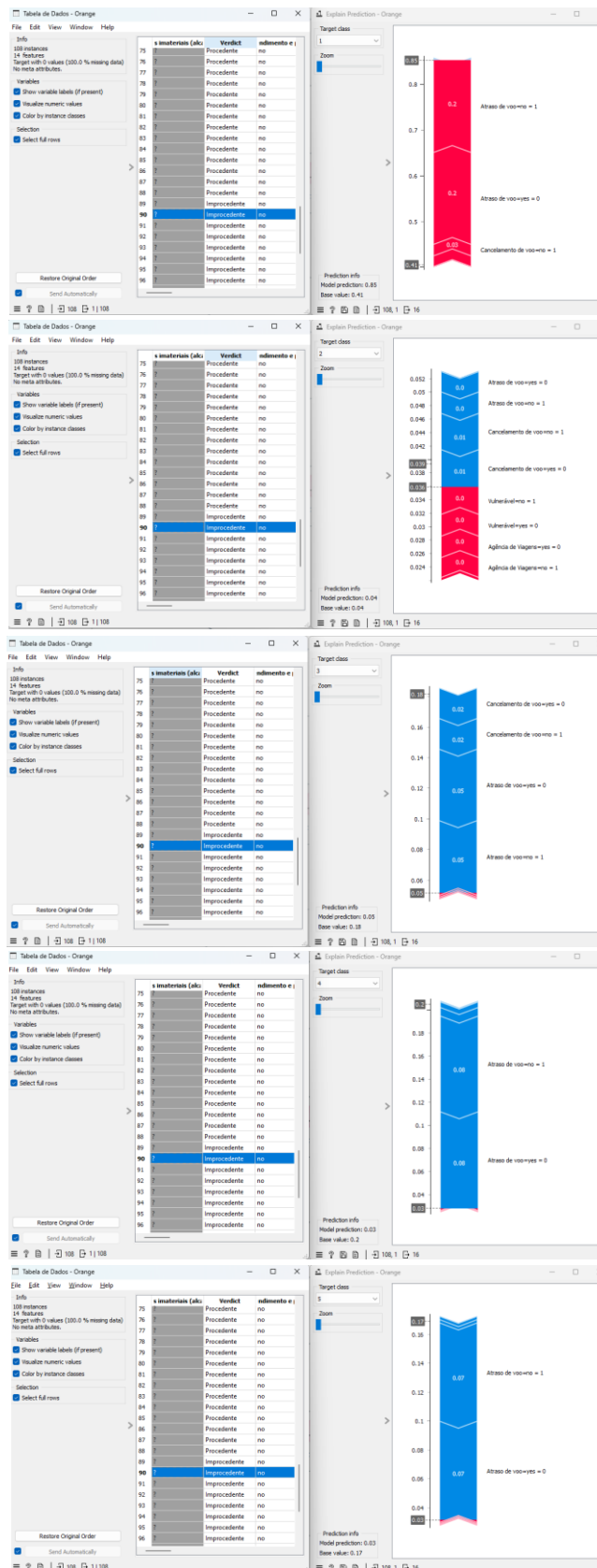
**Figura 91 Ilustração de Planilha em CSV**

	A	B		A	B	C	D	E
1	TXT file	Verdict	1	TXT file	Verdict	Verdict (2)	Immaterial dam	Immaterial dam
			691	685	partly_founded	-	R\$ 10.000,00	5
34	688	Procedente	692	686	partly_founded	-	R\$ 6.000,00	4
35	689	Procedente	693	687	partly_founded	-	R\$ 3.000,00	3
36	690	Procedente	694	688	partly_founded	-	R\$ 5.000,00	3
37	691	Procedente	695	689	partly_founded	-	R\$ 15.000,00	5
			696	690	partly_founded	-	R\$ 10.000,00	5
			697	691	partly_founded	-	R\$ 8.000,00	4

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Muito embora o modelo não tenha sido capaz de identificar inquestionavelmente o intervalo 5, acerca ao prever as maiores chances do mesmo. Aliás, ele também acertou ao prever que a indenização estaria acima de R\$ 5.000,00, dado que, ao somar as chances dos intervalos 4 e 5, obtemos 61% de probabilidade de o resultado estar em um deles. Contudo, é fundamental avaliar o comportamento em outros casos. Com efeito, temos análise do caso número 90 para os cinco intervalos ou cenários a seguir:

Figura 92 Ilustração de Regressão Logística – Tabela de Dados e *Explain Prediction*



Nota: Imagens do Autor. Extraídas de *Orange Canvas*.



Identificam-se como principais atributos considerados para a previsão a vulnerabilidade, agência de viagens, o atraso e o cancelamento de voo. Entretanto, em dissonância com o exemplo anterior, aqui também temos uma nítida diferença de chances entre os intervalos. Isso porque, as chances são de 85% para o intervalo 1 (R\$ 0,00), 0,36% para o intervalo 2 (R\$ 1,00 – R\$ 2.000,00), seguidas por 0,5% para o intervalo 3 (R\$ 2.001,00 – R\$ 5.000,00), 0,3% para o intervalo 4 (R\$ 5.001,00 – R\$ 8.000,00) e 0,3% para o intervalo 5 (R\$ 8.001,00 – R\$ 25.000,00). Prosseguindo com a metodologia, é necessário realizar a consulta em uma planilha com as informações não fornecidas na base de testes para o valor da ação:

**Figura 93 Ilustração de Planilha em CSV**

	A	B		A	B	C	D	E
1	TXT file	Verdict	1	TXT file	Verdict	Verdict (2)	Immaterial dam	Immaterial dam
85	763	Procedente	1042	1035	well_founded	-	R\$ 5.000,00	3
86	765	Procedente	1043	1036	partly_founded	-	R\$ 0,00	1
87	766	Procedente	1044	1037	not_founded	-	R\$ 0,00	1
88	767	Procedente	1045	1038	not_founded	-	R\$ 0,00	1
89	768	Procedente	1046	1039	not_founded	-	R\$ 0,00	1
90	1039	Improcedente	1047	1040	well_founded	-	R\$ 10.000,00	5

Nota: Imagem do Autor. Extraída de *Microsoft Excel*.

Apesar da notável assertividade, temos que ressaltar que o resultado do veredicto foi improcedente. Algo que poderia passar batido é o indicio de que, mesmo quando não fornecido o resultado da sentença, o valor da indenização no intervalo 1 (R\$ 0,00), pode indicar que o modelo não vê grandes chances de êxito na demanda. É bem verdade, que com uma base de dados maiores, os resultados também tendem a ser mais exitosos, ainda assim, esse é um resultado surpreendente dada a dimensionalidade e tempos reduzidos para a experimentação.

Desta feita, podemos considerar que o algoritmo de regressão logística para aprendizado de máquina, com a dimensão da base de dados atual, atende parcialmente ao objetivo do estudo. Isso ocorre porque ele alcança uma acurácia adequada e fornece explicações claras para a procedência ou não da ação. No entanto, essa acurácia não significa que o experimento não pode sofrer reparos e aprimoramentos contínuos, pois não que tange o desenvolvimento de IA, dificilmente teremos uma obra acabada em definitivo e sem atualizações. A seguir, apresentamos o capítulo referente à validação do estudo com um especialista.

## 5- PROPOSTA E VALIDAÇÃO

### 5.1 Proposta de aplicação de método explicativo *SHAP* para modelos preditivos de IA no judiciário brasileiro

O método explicativo *SHAP*<sup>183</sup> aplicado a *XAI* de predições de resultado de sentenças é uma forma de entender como um modelo de *Machine Learning* que classifica as sentenças como positivas ou negativas faz suas decisões. Ou seja, os experimentos realizados com a proposta de aplicação de tal método explicativo para modelos preditivos de IA no judiciário brasileiro envolveram duas etapas principais: a) A experimentação com *Machine Learning* e b) A proposta de aplicação de métodos explicativos.

Na primeira etapa, foram realizados experimentos para entender a interação de modelos preditivos com o conhecimento jurídico estruturado. O objetivo foi construir e avaliar modelos de IA capazes de prever o resultado de decisões judiciais envolvendo ações de companhias aéreas, utilizando dados extraídos de sentenças do Tribunal de Justiça de Santa Catarina. Os experimentos foram feitos utilizando o software *Orange* e a linguagem de programação *Python*, ambos de código-fonte aberto. Os experimentos envolveram técnicas de pré-processamento, representação, agrupamento, associação, classificação e regressão de dados, utilizando algoritmos de *Machine Learning* e *Natural Language Processing*.

Na segunda etapa, foram realizados experimentos com o modelo explicativo *SHAP* para prever decisões de IA em relação a decisões judiciais envolvendo ações de companhias aéreas. O objetivo foi gerar explicações sobre as predições de IA, utilizando o método *SHAP*, que é uma técnica que atribui valores de importância às características de entrada de um modelo de IA, indicando como elas influenciam na predição do modelo. Os experimentos foram feitos utilizando o software *Orange* e a linguagem de programação *Python*, ambos de código-fonte aberto.

Os experimentos mais bem sucedidos foram os com algoritmos de redes neurais

---

<sup>183</sup> [...] O *SHAP* é uma ferramenta muito útil para se obter importâncias de variáveis, independente do modelo utilizado. O método para a geração de suas importâncias, ou contribuições, empresta conceitos provenientes da teoria dos jogos, o que concede propriedades únicas aos valores obtidos (*SHAP*: O que é e por que usar. Big Data Brasil. 2020. Disponível em: <<https://medium.com/big-data-blog/shap-o-que-%C3%A9-e-por-que-usar-6b01d37ae592>>).

e regressão logística (com mais de 90% de acurácia), que utilizaram o método explicativo *SHAP* para gerar explicações sobre as predições de IA em relação a decisões judiciais envolvendo ações de companhias aéreas. Os atributos ou *features* consideradas determinantes para a predição, apresentadas no fluxo de trabalho pelo *Explain Model* e *Explain Prediction*, permitiram obter explicações claras, coerentes e até certo ponto compreensíveis sobre as predições de IA tornando-a mais transparente e compreensível, fornecendo justificativas para as suas predições e facilitando a sua avaliação e validação.

Analisando os possíveis efeitos da implantação do método SHAP e XAI no Judiciário brasileiro, temos que uma das implicações práticas do método SHAP é que ele pode auxiliar os operadores do direito na tomada de decisões baseadas em modelos de aprendizado de máquina. Um juiz pode usar o método SHAP para entender como um modelo de predição de risco de reincidência penal atribuiu um determinado valor para um réu, quais foram as variáveis mais relevantes para essa atribuição, e como elas se relacionam com as evidências e os argumentos apresentados no processo. Ou seja, o juiz pode ter mais confiança e transparência na sua decisão, bem como verificar se o modelo está alinhado com os princípios e normas jurídicas.

Ademais, ele pode garantir os direitos e deveres dos cidadãos que são afetados por decisões baseadas em modelos de aprendizado de máquina. Um cidadão que teve seu pedido de crédito negado por um modelo de análise de risco pode usar o método SHAP para solicitar uma explicação sobre os motivos dessa negativa, quais foram os fatores que influenciaram o resultado, e como ele pode contestar ou recorrer dessa decisão. Assim, o cidadão pode exercer seu direito à explicação, à informação, à defesa e ao contraditório, bem como cumprir seus deveres de boa-fé e cooperação.

Igualmente, existem implicações éticas do método SHAP é que ele pode evitar ou mitigar vieses e discriminações que podem estar presentes em modelos de aprendizado de máquina. A título de esclarecimento, temos que um pesquisador pode usar o método SHAP para identificar e quantificar a contribuição de variáveis potencialmente sensíveis, como gênero, raça, idade, religião e assim por diante, para a saída do modelo, e verificar se elas estão sendo usadas de forma justa e proporcional. Em outras palavras, o pesquisador pode avaliar e corrigir possíveis efeitos e vieses indesejados ou prejudiciais do modelo para determinados grupos ou indivíduos.

Por fim, o método SHAP pode facilitar a regulação e fiscalização de modelos de aprendizado de máquina que são usados no contexto jurídico. Um órgão regulador pode usar o método SHAP para verificar se um modelo de aprendizado de máquina que é usado para fins jurídicos atende aos requisitos de qualidade, segurança, eficiência, responsabilidade, que são exigidos pela legislação e pela jurisprudência. Desse modo, o órgão regulador pode garantir que o modelo está em conformidade com os padrões legais e éticos, e aplicar sanções ou medidas corretivas em caso de violação.

Conclui-se, portanto, que o método SHAP é uma ferramenta poderosa e útil para explicar e compreender modelos de aprendizado de máquina que são aplicados no contexto jurídico. Ele pode trazer benefícios práticos e éticos para os operadores do direito, os cidadãos, os pesquisadores e os órgãos reguladores, contribuindo para a melhoria da qualidade, da confiabilidade, da justiça e da transparência das decisões baseadas em inteligência artificial.

No entanto, além do método SHAP, existem outros modelos explicativos de IA que podem ser utilizados no direito, tais como:

- I- ***LIME (Local Interpretable Model-Agnostic Explanations)***: Um método que gera explicações locais para qualquer modelo de aprendizado de máquina, criando uma aproximação linear simples e interpretável em torno da predição de interesse. Enquanto o LIME ajusta um modelo linear mais simples a um modelo complexo, o SHAP tenta atribuir a contribuição relativa de cada recurso ao desvio da previsão da previsão média do modelo<sup>184</sup>.
- II-  ***Anchors***: O  *Anchors* foi desenvolvido por Marco Ribeiro, que também criou o LIME. Explicadores de alta precisão que usam métodos de aprendizagem por reforço para chegar a um conjunto de condições de características (chamadas  *Anchors*), que ajudarão a explicar a observação de interesse e também um conjunto de observações circundantes com alta precisão (o usuário é livre para escolher seu corte de precisão mínimo). Eles também não são computacionalmente onerosos como o SHAP, e possuem melhor generalização do que LIME.<sup>185</sup>.
- III- ***Counterfactual Explanations***: Podem ser usadas para fornecer  *insights* acionáveis sobre as previsões do modelo, permitindo-nos

---

<sup>184</sup> DIDUGU. C. Anchors: A simple Introduction. Medium. 2020. p.5. Disponível em:<<https://medium.com/swlh/ultimate-guide-to-model-explainability-anchors-2deab8239f57>>.

<sup>185</sup> Ibidem, p.2.

alterar instâncias individuais como um caminho para alcançar um resultado desejado.<sup>186</sup>.

Esses modelos explicativos de IA podem ser utilizados no direito para diversos fins, tais como:

- a) Avaliar a qualidade e a confiabilidade dos modelos de aprendizado de máquina usados no direito, comparando as explicações geradas por diferentes métodos e verificando se elas são consistentes e coerentes com os critérios jurídicos.
- b) Melhorar a comunicação e a compreensão entre os profissionais do direito e os desenvolvedores de modelos de aprendizado de máquina, usando as explicações como uma linguagem comum e acessível para discutir os objetivos, os requisitos, os limites e os resultados dos modelos.
- c) Promover a educação e a capacitação dos profissionais do direito sobre o uso e o impacto da IA no direito, usando as explicações como uma ferramenta pedagógica e interativa para demonstrar e explorar as possibilidades e os desafios da aplicação da IA no campo jurídico.

Desta feita, é razoável crer que, em um futuro próximo, a aplicação do método SHAP e de outros modelos explicativos no contexto jurídico podem tornar a IA mais transparente e compreensível, fornecendo justificativas para as suas previsões e facilitando a sua avaliação e validação.

Evidente, que como veremos a seguir, os experimentos passaram por ciclos validativos, que contribuíram para a sua maior robustez. Muito embora, apenas os algoritmos de redes neurais e regressão logística tenham tido um desempenho aceitável, é interessante verificar as possibilidades que estão no horizonte. Isso porque, ferramentas com essa potencializadas por uma equipe multidisciplinar podem ser determinantes para reduzir o tempo de procedimentos processuais, auxiliar na negociação e mediação entre as partes, assim como pode ajudar na restauração da confiança por parte dos cidadãos no Judiciário Brasileiro.

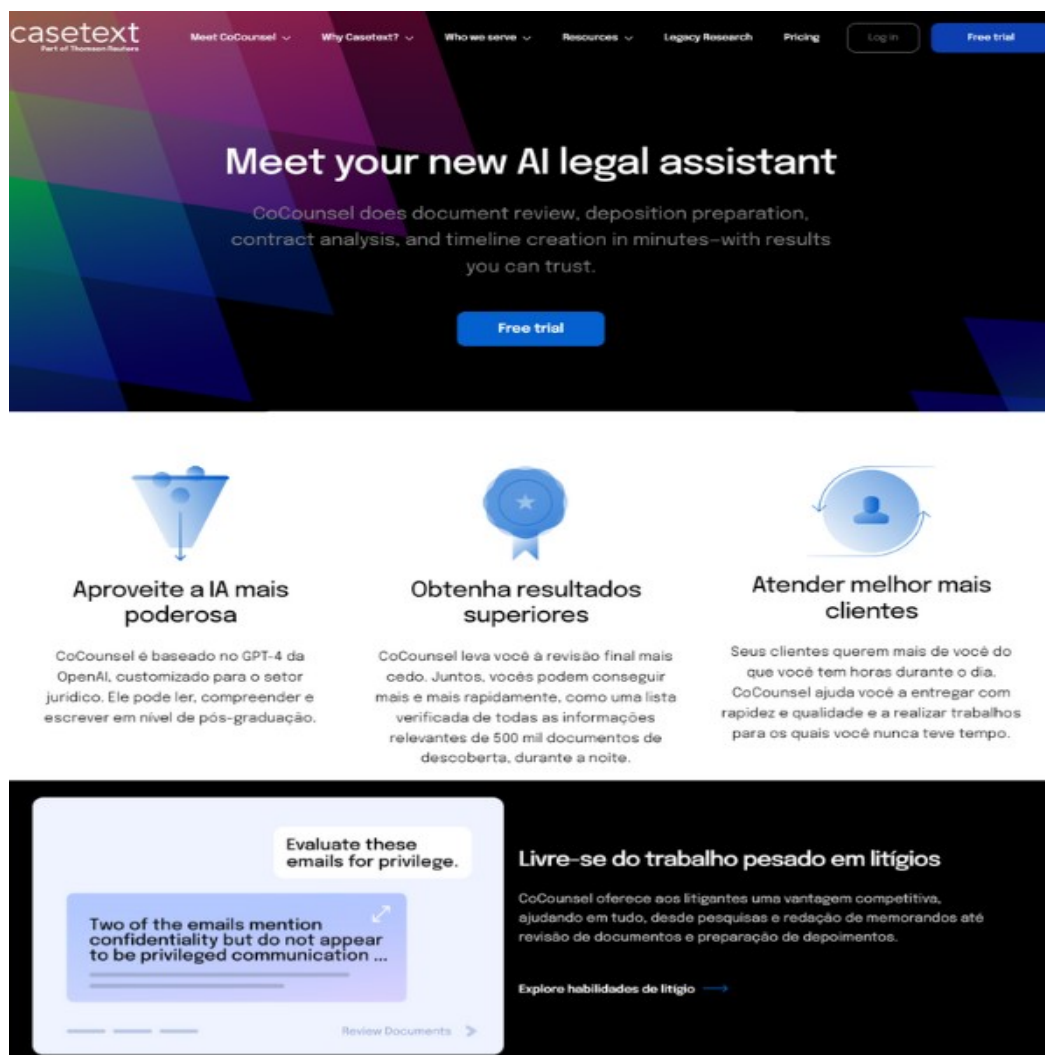
Mais recentemente, temos visto o surgimento de diversos modelos *LLMs* (*Language Model for Legal Text*), o que pode ser uma possibilidade interessante para aprimorar a capacidade de compreensão e análise de textos jurídicos por parte dos

---

<sup>186</sup> EUSEBIO. B. Counterfactual Explanations in Model Interpretations. Medium. 2020. p.10. Disponível em:<<https://towardsdatascience.com/counterfactual-explanations-in-model-interpretations-a73caec5b74b>>.

modelos de IA. O *ChatGPT*<sup>187</sup>, por exemplo, é um *LLM* recentemente lançado que pode ser utilizado para gerar respostas automáticas a perguntas jurídicas complexas. A aplicação de *LLMs* pode ser uma área de pesquisa promissora para melhorar a eficácia e precisão dos modelos de IA no contexto jurídico.

**Figura 94 Ilustração do *CoCounsel* da *Casetext*.**



Nota: Imagem do Autor. Extraída da plataforma *CoCounsel* desenvolvida pela *Casetext*<sup>188</sup>.

<sup>187</sup> The introduction of ChatGPT has set off an arms race amongst large companies and start-ups alike on creating the next best LLM. A key recurring theme has been that simply increasing the number of parameters in the model was no longer the sole method of improving LLM quality (LUK, Martin. *Generative AI: Overview, economic impact, and applications in asset management. Economic Impact, and Applications in Asset Management*, p.26, 2023. Available in: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4574814](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4574814)>).

<sup>188</sup> Casetext. *Cocounsel*. 2023. Disponível em: <<https://casetext.com/>>.

A ferramenta *CoCounsel* da *Casetext*, por exemplo, funciona como um assistente jurídico de IA que usa a tecnologia GPT-4 para realizar diversas tarefas legais, como revisão de documentos, preparação de depoimentos, análise de contratos, criação de cronogramas, entre outras funcionalidades. Ela é projetada para aumentar a eficiência, a precisão e a qualidade do trabalho jurídico, permitindo que, mais especificamente, os advogados se concentrem em aspectos mais estratégicos e criativos da prática.

**Figura 95 Ilustração da LIVIA.**

**Descubra o poder da LIVIA**

Consiga eficiência, precisão e insights inigualáveis — permitindo que você se concentre no aspecto humano do direito.

- Análise de Documentos**  
Faça uma pergunta sobre seus documentos e LIVIA irá examiná-los a fundo e responder com referências precisas. Localize facilmente informações essenciais, como depoimentos-chave em transcrições longas ou termos cruciais em contratos complexos.
- Preparação para Depoimentos**  
Crie um roteiro completo de depoimento rapidamente. Insira informações sobre o depoente e as questões do caso, e LIVIA identificará tópicos relevantes e formulará perguntas pertinentes.
- Exploração do Banco de Dados**  
Basta fazer uma pergunta e LIVIA pesquisará seus documentos, analisará os relevantes e fornecerá uma resposta. Simplifique o processo de localização de modelos adequados, precedentes ou conhecimento interno.
- Suporte à Pesquisa Jurídica**  
Apresente uma questão de pesquisa, especificando o máximo de detalhes desejado, e LIVIA recuperará rapidamente recursos relevantes, fornecendo uma resposta fundamentada complementada por fontes de apoio.
- Resumo de Documentos**  
Extraia rapidamente informações críticas de vários documentos - como contratos complexos ou opiniões extensas - sem sacrificar detalhes essenciais.
- Extração de Dados de Contratos**  
Obtenha respostas rápidas e uma lista abrangente de cláusulas relevantes de sua coleção de contratos, facilitando o acompanhamento de termos de negócio, valores monetários e prazos.
- Conformidade em Contratos**  
LIVIA detecta cláusulas não conformes em um conjunto de contratos, destaca possíveis riscos e sugere revisões necessárias.

Nota: Imagem do Autor. Extraída da plataforma LIVIA<sup>189</sup> desenvolvido pela INTELIJUS.

<sup>189</sup> LIVIA. INTELIJUS. 2023. Disponível em: <<https://www.intelijus.ai/>>.

Além disso, no mercado de IA jurídica, existem outras ferramentas semelhantes como a LIVIA, que é a primeira IA jurídica multilíngue, que pode entender e gerar textos em português, inglês e espanhol. Inclusive, ela pode auxiliar os operadores do direito em pesquisas, redação, revisão e tradução de documentos jurídicos, além de fornecer respostas e orientações sobre questões legais.

Em outras palavras, não apenas as experimentações individuais e coletivas na pesquisa científica, senão, o próprio mercado jurídico e suas dinâmicas apontam para uma maior interdependência entre os universos das tecnologias digitais e o direito. Por isso, é importante que as práticas jurídicas sejam mais associadas as atividades públicas ou privadas, não se dissociem da transparência e de seus princípios basilares. Ainda que, não seja fácil obter a explicabilidade e interpretabilidade de predições e resultados, isso não deve desmotivar a busca pela melhoria contínua e pelo alcance de uma maior transparência das ferramentas de IA disponíveis atualmente.

Dito de outra maneira, a transparência é um elemento fundamental se o que se busca é uma IA responsável, que respeite os direitos humanos, a ética e a democracia. No Brasil, como demonstrado na seção 2, temos a Estratégia Brasileira de Inteligência Artificial (EBIA), alinhada com os princípios da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Entre esses princípios, está a transparência e a explicabilidade<sup>190</sup>, que visam assegurar que as decisões e os resultados dos sistemas de IA sejam claros, acessíveis e úteis para os usuários e as partes interessadas.

Ademais, o Art. 20 da LGPD também prevê que o cidadão pode solicitar a revisão de decisões tomadas unicamente com base no tratamento automatizado de dados pessoais que afetem seus interesses, o que inclui a IA. É bem verdade, por outro lado, que a lei não define critérios claros para a transparência dessas decisões, nem como elas devem ser revisadas.

Como se não bastasse, o país também abriga o PL 2338/23, em tramitação no Congresso Nacional que visa regulamentar IA no Brasil. Entre os vários tópicos

---

<sup>190</sup> MACOHIN. A. Transparência: elemento fundamental para uma IA responsável. SERPRO. 2024. Disponível em: <<https://www.serpro.gov.br/menu/noticias/noticias-2024/transparencia-ia-responsavel>>.



abordados, está, coincidentemente ou não, a necessidade de transparência das soluções de IA, principalmente em casos de alto risco, ou seja, que possuem alta probabilidade de violar direitos e causar impactos negativos. O projeto também propõe a criação de um órgão regulador da IA.

Portanto, em se tratando da realidade brasileira, é preciso atentar ao fato de que uma mudança legislativa, pode provocar a necessidade de adaptação, *update*, atualização de IA's que ainda funcionem perante estruturas de *black-boxes*, onde se tem pouco ou nenhum conhecimento da sua estrutura operadora e os caminhos tomados até o resultado. A seguir, temos a verificação dos resultados apresentados junto a especialista em IA e também pelo Juizado Especial Cível (JEC/UFSC).

## **5.2 Validação, resultados e discussão**

Nesta seção, apresentamos e discutimos os resultados obtidos com a proposta de aplicação do método explicativo SHAP para modelos preditivos de IA no judiciário brasileiro, bem como validamos a abordagem proposta com especialistas na área. Para tanto, realizamos dois tipos de validação: a validação com especialistas ao longo do desenvolvimento dos experimentos de XAI e, posteriormente, os questionários de validação com expert em IA e JEC/UFSC.

A validação da abordagem foi fundamental para aprimorar a proposta e validar os resultados obtidos, bem como para demonstrar a importância, a relevância e a contribuição para a área de IA aplicada ao campo jurídico. Igualmente, se insere no cenário atual e futuro da utilização de IA no judiciário brasileiro, que apresenta desafios, oportunidades, expectativas e demandas para o sistema jurídico e para a sociedade.

Do mesmo modo, se alinha com iniciativas, projetos, normas e diretrizes que estão sendo desenvolvidas ou implementadas para regular, orientar e avaliar o uso de IA no judiciário, como o Programa de Inteligência Artificial do Governo Federal e A Lei Geral de Proteção de Dados (LGPD). Por fim, tais experimentos contribuem ainda para propositura e validação de uma abordagem baseada no método SHAP, que é um dos

métodos mais avançados e robustos de IA explicável, considerando o caso específico das ações envolvendo passagens aéreas no JEC/UFSC.

### 5.2.1 Validação com especialistas ao longo do desenvolvimento dos experimentos de XAI

A validação com especialistas ao longo do desenvolvimento dos experimentos de XAI foi feita por meio de observação participante do Grupo de pesquisa E-gov, principalmente na figura da Profa. Dra. Isabela Sabo e do Prof. Dr. Aires Rover. O objetivo foi obter *feedbacks* sobre a qualidade, a relevância e a utilidade das explicações geradas pelo método SHAP, bem como sobre as limitações e as sugestões de melhoria da nossa abordagem. Tal validação contínua permitiu aprimorar a proposta e validar os resultados obtidos.

Os especialistas acompanharam o processo de desenvolvimento dos experimentos de XAI, desde a definição do problema, a coleta e o tratamento dos dados, a construção e o treinamento do modelo preditivo de IA, a aplicação e a visualização do método SHAP, até a geração e a comunicação das explicações. Do mesmo modo, forneceram orientações, recomendações, críticas e elogios sobre cada etapa do processo, ajudando a qualidade e a relevância da abordagem.

Ademais, reconheceram o valor e a utilidade das explicações geradas pelo método SHAP, destacando como elas podem aumentar a transparência, a confiança e a responsabilidade da IA no campo jurídico. Evidente que apontaram as limitações e os desafios do método SHAP, como a complexidade computacional, a sensibilidade aos dados de treinamento e a variabilidade das explicações.

Por fim, sugeriram algumas melhorias na comunicação das explicações, como a revisão de termos técnicos, a discussão das explicações no resumo, e a inclusão de exemplos práticos. Conclui-se, portanto, que a validação com especialistas ao longo do desenvolvimento dos experimentos de XAI foi essencial para garantir a qualidade, a relevância e a utilidade do método SHAP, assim como também para identificar e superar limitações e desafios.

### 5.2.2 Questionários de validação com expert em IA e JEC/UFSC

No intuito de validar a aplicação do método explicativo SHAP para modelos preditivos de IA no campo jurídico, foram realizados dois questionários com especialistas, a saber: uma servidora do JEC/UFSC e um especialista em IA. Ambos os questionários foram elaborados com base nos objetivos específicos da validação, que eram: a) verificar a compreensão dos especialistas sobre o conceito e a utilidade da XAI no campo jurídico; b) avaliar o nível de interesse e confiança dos especialistas na aplicação do método SHAP para explicar as decisões de IA; e c) obter feedbacks e recomendações dos especialistas sobre a abordagem proposta.

Igualmente, os questionários foram compostos por perguntas fechadas e abertas, utilizando uma escala *Likert* de cinco pontos para medir o grau de concordância ou discordância dos especialistas com os questionamentos apresentados. As respostas foram coletadas e analisadas por meio da plataforma *Google Forms*, que permitiu calcular as médias, desvios padrões, extrair as principais ideias e citações das respostas qualitativas.

As respostas dos questionários de validação com a servidora do JEC/UFSC e o especialista em IA forneceram insights valiosos sobre a aplicação do método explicativo *SHAP* para modelos preditivos de IA no campo jurídico. Válido ressaltar que, ambos receberam para além dos questionários acesso aos experimentos e resumos didáticos. A seguir, são apresentados e discutidos os principais resultados obtidos com cada especialista.

#### 5.2.2.1 Resultados da validação com a servidora do JEC/UFSC

A servidora do JEC/UFSC demonstrou um conhecimento razoável sobre o conceito de IA e sua aplicação no campo jurídico. Ela expressou total concordância com a utilidade da IA para auxiliar os juízes na análise e decisão de processos judiciais, com uma média de 5 (cinco) e um desvio padrão de 0 (zero) na escala *Likert*.

Além disso, demonstrou grande interesse em entender como a IA chega a uma determinada predição ou decisão, destacando a importância de uma explicação clara e transparente sobre os fatores que influenciam a decisão da IA. Com base nos resultados do experimento apresentado com IA explicável, a servidora acredita que isso tem grandes

chances de ser utilizado para tornar o uso de IA, em um futuro próximo, mais confiável pelo Judiciário, com uma média de 4 (quatro) e desvio padrão de 1 (um) na escala *Likert*.

Acresce-se a isso o fato de que a servidora expressou um forte interesse em uma ferramenta que pudesse prever o resultado de uma sentença judicial baseada em dados de outras decisões do mesmo JEC, especialmente no caso de ações envolvendo passagens aéreas, com uma média de 4 (quatro) e um desvio padrão de 1 (um) na escala *Likert*. Isso indica uma necessidade real e uma oportunidade para a aplicação de IA no campo jurídico. A servidora também considerou os resultados apresentados pela experimentação como bem confiáveis, com uma média de 4 (quatro) e um desvio padrão de 1 (um) na escala *Likert*, o que reforça a eficácia do método *SHAP*.

**Figura 96 Ilustração de Questionário de Validação com servidora do JEC/UFSC.**

<p>1- Você conhece o conceito de inteligência artificial (IA) e como ela pode ser aplicada no campo jurídico?</p> <p><input type="radio"/> Não sei nada sobre o assunto. (1)</p> <p><input type="radio"/> Conheço pouco sobre o assunto. (2)</p> <p><input checked="" type="radio"/> Tenho conhecimento razoável sobre o assunto. (3)</p> <p><input type="radio"/> Conheço bem o assunto. (4)</p> <p><input type="radio"/> Conheço muito bem o assunto. (5)</p>	<p>5- A partir dos resultados do experimento apresentado com IA explicável, você acredita que isso possa ser utilizado para tornar o uso de IA, em um futuro próximo, mais confiável pelo Judiciário?</p> <p><input type="radio"/> Não acredito. (1)</p> <p><input type="radio"/> Acredito muito pouco. (2)</p> <p><input type="radio"/> Talvez seja possível. (3)</p> <p><input checked="" type="radio"/> Tem grandes chances de ser utilizada pelo Judiciário. (4)</p> <p><input type="radio"/> Acredito totalmente no potencial do experimento. (5)</p>
<p>2- Concorda que a IA pode ser útil para auxiliar os juizes na análise e na decisão de processos judiciais?</p> <p><input type="radio"/> Discordo totalmente. (1)</p> <p><input type="radio"/> Tenho pouca esperança. (2)</p> <p><input type="radio"/> Acredito razoavelmente. (3)</p> <p><input type="radio"/> Concordo que existe possibilidade. (4)</p> <p><input checked="" type="radio"/> Concordo totalmente. (5)</p>	<p>6- A partir de sua ótica, qual o nível de confiabilidade dos resultados apresentados pela experimentação?</p> <p><input type="radio"/> Não confio. (1)</p> <p><input type="radio"/> Pouco confiável. (2)</p> <p><input type="radio"/> Razoavelmente confiável. (3)</p> <p><input checked="" type="radio"/> Bem confiável. (4)</p> <p><input type="radio"/> Muito confiável. (5)</p>
<p>3- Você acredita no futuro de uma ferramenta que pudesse prever o resultado de uma sentença judicial baseada em dados de outras decisões do mesmo JEC, como no caso de ações envolvendo passagens aéreas?</p> <p><input type="radio"/> Não me sinto a vontade com o uso de IA no Judiciário. (1)</p> <p><input type="radio"/> Acredito pouco no futuro de uma ferramenta desse tipo. (2)</p> <p><input type="radio"/> Razoavelmente, a depender da segurança dos resultados e funcionamento. (3)</p> <p><input checked="" type="radio"/> Bastante, dada importância da implantação de ferramentas de IA explicável. (4)</p> <p><input type="radio"/> Totalmente, acredito que isso pode ajudar a tornar o Judiciário mais eficiente. (5)</p>	<p>7- Você tem alguma dúvida, sugestão ou crítica sobre o experimento realizado com a ferramenta de explicabilidade de IA desenvolvida para o JEC? Se sim, por favor, escreva abaixo:</p> <hr/>
<p>4- O quanto você gostaria de saber como a IA chegou a uma determinada predição ou decisão? Você considera importante ter uma explicação clara e transparente sobre os fatores que influenciaram a decisão da IA?</p> <p><input type="radio"/> Simplesmente não tenho interesse. Não é importante. (1)</p> <p><input type="radio"/> Tenho muito pouco interesse. Pode vir a ser importante algum dia. (2)</p> <p><input type="radio"/> Tenho um interesse razoável. De fato, tem a sua importância. (3)</p> <p><input type="radio"/> Tenho interesse considerável. Considero importante. (4)</p> <p><input checked="" type="radio"/> Tenho muito interesse. Considero importantíssimo. (5)</p>	

Nota: Imagem do Autor. Extraída de *Google Formulários*.

Em síntese, os resultados da validação com a servidora do JEC/UFSC mostram que ela reconhece e valoriza a aplicação do método explicativo SHAP para modelos preditivos de IA no campo jurídico, bem como confia e se interessa pelos resultados e pelas explicações geradas pelo método SHAP. Isso indica que a nossa abordagem atende às expectativas e às necessidades da servidora, que é uma potencial usuária ou afetada pela IA no campo jurídico.

#### 5.2.2.2 Resultados da validação com a expert em IA

O expert em IA demonstrou um alto conhecimento sobre o conceito de IA e sua aplicação no campo jurídico. Ele concorda totalmente que a IA pode contribuir para a melhoria da qualidade e da eficiência das decisões judiciais, com uma média 5 (cinco) e um desvio padrão 0 (zero) na escala *Likert*.

Ademais, o especialista também considera muito necessário que a IA forneça uma explicação clara e transparente sobre os fatores que influenciaram a sua predição ou decisão, com uma média de 5 (cinco) e um desvio padrão de 0 (zero) na escala *Likert*. Com base nos resultados do experimento apresentado com IA explicável, ele acredita que isso tem grandes chances de ser utilizado para tornar o uso de IA, em um futuro próximo, mais confiável pelo Judiciário, com uma média de 4 (quatro) e um desvio padrão de 1 (um) na escala *Likert*. Por fim, informou que sentiu falta de uma discussão acerca das explicações no resumo e sugeriu revisar alguns termos técnicos, como “valor de Shapley” e “valor de base”, que podem não ser familiares para o público de juristas.

## Figura 97 - Ilustração de Questionário de Validação com especialista em IA.

1- Você está familiarizado com o uso de IA no campo jurídico? \*

Não sei nada sobre o assunto. (1)

Já ouvi falar, porém desconheço a aplicação. (2)

Razoavelmente, tenho conhecimento de seu uso e aplicação. (3)

Sim, já tive contato com o uso de aplicação de IA no campo jurídico (4).

Sim, além de dominar o tema, trabalho com IA no campo jurídico (5).

---

2- Concorda que a IA pode contribuir para a melhoria da qualidade e da eficiência das decisões judiciais?

Discordo totalmente. (1)

Tenho pouca esperança. (2)

Acredito razoavelmente. (3)

Concordo que existe possibilidade. (4)

Concordo totalmente. (5)

---

3- Você acredita no futuro de uma ferramenta que pudesse prever o resultado de uma sentença judicial baseada em dados de outras decisões do mesmo JEC, como no caso de ações envolvendo passagens aéreas?

Não me sinto a vontade com o uso de IA no Judiciário. (1)

Acredito pouco no futuro de uma ferramenta desse tipo. (2)

Razoavelmente, a depender da segurança dos resultados e funcionamento. (3)

Bastante, considero essencial a implantação de ferramentas de IA explicável. (4)

Totalmente, acredito que isso pode ajudar a tornar o Judiciário mais eficiente. (5)

---

4- O quanto você considera necessário que a IA forneça uma explicação clara e transparente sobre os fatores que influenciaram a sua predição ou decisão?

Não é necessário. (1)

Pouco necessário. (2)

Razoavelmente necessário. (3)

Muito necessário. (4)

Extremamente necessário. (5)

---

5- A partir dos resultados do experimento apresentado com IA explicável, você acredita que isso possa ser utilizado para tornar o uso de IA, em um futuro próximo, mais confiável pelo Judiciário?

Não acredito. (1)

Acredito muito pouco. (2)

Talvez seja possível. (3)

Tem grandes chances de ser utilizada pelo Judiciário. (4)

Acredito totalmente no potencial do experimento. (5)

6- A partir de sua ótica, qual o nível de confiabilidade dos resultados apresentados pela experimentação?

Não confio. (1)

Pouco confiável. (2)

Razoavelmente confiável. (3)

Bem confiável. (4)

Muito confiável. (5)

---

7- Você tem alguma dúvida, sugestão ou crítica sobre o experimento realizado com a ferramenta de explicabilidade de IA desenvolvida para o JEC? Se sim, por favor, escreva abaixo.

Os resultados são bem interessantes! As explicações geradas pelo SHAP são bem claras. Mas no resumo, senti falta de uma discussão a cerca das explicações. Ou seja, faz sentido atributo X tem um peso maior que o atributo Y? Parabéns pelo trabalho! Sugiro revisar alguns termos técnicos.

Nota: Imagem do Autor. Extraída de *Google Formulários*.

De modo mais amplo, as respostas dos questionários de validação indicam que a aplicação do método explicativo *SHAP* para modelos preditivos de IA no campo jurídico é bem recebida e considerada confiável pelos especialistas. Isso reforça a relevância e a eficácia da nossa abordagem, contribuindo para o aprimoramento do sistema jurídico brasileiro. Ainda assim, também destacam a necessidade de melhorar a comunicação das explicações geradas pelo método *SHAP*, especialmente no que diz respeito à discussão das explicações no resumo e à revisão de termos técnicos. Esses insights serão valiosos para orientar futuras melhorias na nossa abordagem.

Muito embora, exista tópico destinado ao tema, uma das limitações encontradas durante o processo de validação, em específico, foi o tamanho reduzido da amostra de especialistas, que pode não ser representativa da população de potenciais usuários de IA explicável no campo jurídico. Além disso, o método *SHAP* também apresenta algumas limitações., como a complexidade computacional, a sensibilidade aos dados de

treinamento, e a variabilidade das explicações. Essas limitações podem afetar a qualidade e a consistência das explicações geradas pelo método SHAP, bem como a sua interpretação pelos usuários. Para minimizar essas limitações, foram adotadas algumas medidas, como a seleção criteriosa dos especialistas, a utilização de dados reais e atualizados, a aplicação de técnicas de otimização e a validação dos resultados. No entanto, essas medidas não eliminam completamente as limitações, e por isso, recomenda-se cautela e senso crítico ao utilizar e avaliar a IA explicável no campo jurídico.

Ao comparar os resultados da validação com a literatura existente (identificada na revisão sistemática de literatura), pode-se observar que a experimentação e pesquisa científica estão alinhadas com as tendências e os desafios da XAI no campo jurídico. De acordo com a literatura, a IA explicável, ou XAI, é uma área emergente e promissora, que visa aumentar a transparência, a confiança e a responsabilidade da IA no campo jurídico, bem como facilitar a interação e colaboração entre humanos e máquinas. No entanto, a literatura também aponta que a IA explicável enfrenta diversos obstáculos, como a falta de padrões, de métodos, de ferramentas, de avaliações e de regulamentações específicas para o campo jurídico.

De igual modo, a literatura existente também destaca que a IA explicável deve levar em conta as características, as necessidades, as expectativas e as preferências dos usuários, bem como o contexto, o domínio e o propósito da aplicação da IA no campo jurídico. Nesse sentido, a dissertação contribui para o campo científico ao propor e validar uma abordagem baseada no método SHAP, que é um dos métodos mais avançados e robustos de IA explicável, para modelos preditivos de IA no campo jurídico, considerando o caso específico das ações envolvendo passagens aéreas no JEC/UFSC.

Ademais, também podem ser consideradas contribuições as recomendações e *feedbacks* dos especialistas sobre a abordagem científica, que podem servir de base para futuros trabalhos na área de IA explicável (XAI) no campo jurídico. Obtida a validação da proposta e discussão dos resultados, temos a seguir a conclusão com as considerações, contribuições, limitações e opções de trabalhos futuros.

## 6- CONCLUSÃO

### 6.1 Considerações finais e contribuições

Após a farta análise de textos científicos e experimentação, temos que é possível considerar viável a implantação de Inteligência Artificial Explicável (*XAI*) no campo do Direito, utilizando o método *SHAP* para gerar explicações sobre as predições de IA em decisões judiciais, é uma abordagem viável e eficaz. A metodologia proposta permitiu obter explicações claras, coerentes e compreensíveis sobre as predições de IA, fornecendo transparência, confiança e responsabilidade aos sistemas de IA.

Destarte, a utilização do método *SHAP* para representar o conhecimento jurídico e dar explicabilidade à IA, utilizando dados de decisões judiciais de tribunais, destaca a originalidade e relevância deste estudo. Isto é, a aplicação desse método no contexto jurídico visa tornar a IA mais transparente e compreensível, fornecendo justificativas para as suas predições e facilitando a sua avaliação e validação.

A validação da abordagem proposta com especialistas na área, como engenheiros de IA e servidores do JEC/UFSC, permitiu aprimorar a metodologia e validar os resultados obtidos. Ademais, a avaliação foi feita por meio de observação participante e questionário de pesquisa, o que garantiu a qualidade, a relevância e a utilidade das explicações geradas pelo método *SHAP*, bem como sobre as limitações e as sugestões de melhoria da nossa abordagem.

Por fim, é importante destacar que este estudo contribui para o avanço do conhecimento sobre a *XAI* e sua aplicação no campo jurídico, por meio de uma abordagem inovadora e interdisciplinar. A utilização de métodos explicativos de predições de IA no judiciário brasileiro, utilizando o método *SHAP*, pode ser uma solução prática e eficiente para os desafios da explicabilidade da IA no âmbito jurídico, oferecendo transparência, confiança e responsabilidade aos sistemas de IA.

No entanto, frisa-se novamente que, além do método *SHAP*, existem outros modelos explicativos de IA que podem ser utilizados no direito, tais como:

- IV- ***LIME (Local Interpretable Model-Agnostic Explanations)***: Um método que gera explicações locais para qualquer modelo de



aprendizado de máquina, criando uma aproximação linear simples e interpretável em torno da predição de interesse. Enquanto o LIME ajusta um modelo linear mais simples a um modelo complexo, o SHAP tenta atribuir a contribuição relativa de cada recurso ao desvio da previsão da previsão média do modelo<sup>191</sup>.

- V- ***anchors***: O *anchors* foi desenvolvido por Marco Ribeiro, que também criou o LIME. Explicadores de alta precisão que usam métodos de aprendizagem por reforço para chegar a um conjunto de condições de características (chamadas *anchors*), que ajudarão a explicar a observação de interesse e também um conjunto de observações circundantes com alta precisão (o usuário é livre para escolher seu corte de precisão mínimo). Eles também não são computacionalmente onerosos como o SHAP, e possuem melhor generalização do que LIME.<sup>192</sup>.
- VI- ***Counterfactual Explanations***: Podem ser usadas para fornecer *insights* acionáveis sobre as previsões do modelo, permitindo-nos alterar instâncias individuais como um caminho para alcançar um resultado desejado.<sup>193</sup>.

Assim como o método SHAP utilizado no experimento, os demais modelos explicativos de IA podem ser utilizados no direito para diversos fins, tais como:

- a) Avaliar a qualidade e a confiabilidade dos modelos de aprendizado de máquina usados no direito, comparando as explicações geradas por diferentes métodos e verificando se elas são consistentes e coerentes com os critérios jurídicos.
- b) Melhorar a comunicação e a compreensão entre os profissionais do direito e os desenvolvedores de modelos de aprendizado de máquina, usando as explicações como uma linguagem comum e acessível para discutir os objetivos, os requisitos, os limites e os resultados dos modelos.
- c) Promover a educação e a capacitação dos profissionais do direito sobre o uso e o impacto da IA no direito, usando as explicações como uma ferramenta pedagógica e interativa para demonstrar e explorar as possibilidades e os desafios da aplicação da IA no campo jurídico.

Desta feita, é razoável crer que, em um futuro próximo, a aplicação do método SHAP e de outros modelos explicativos no contexto jurídico podem tornar a IA mais transparente e compreensível, fornecendo justificativas para as suas previsões e facilitando a sua avaliação e validação.

---

<sup>191</sup> DIDUGU. C. *anchors*: A simple Introduction. Medium. 2020. p.5. Disponível em:<<https://medium.com/swlh/ultimate-guide-to-model-explainability-anchors-2deab8239f57>>.

<sup>192</sup> Ibidem, p.2.

<sup>193</sup> EUSEBIO. B. *Counterfactual Explanations in Model Interpretations*. Medium. 2020. p.10. Disponível em:<<https://towardsdatascience.com/counterfactual-explanations-in-model-interpretations-a73caec5b74b>>.

Evidente que tais soluções tecnológicas não suprem todas as necessidades sociais e jurídica, senão que reforçam a relevância e a eficácia da nossa abordagem, contribuindo para o aprimoramento do sistema jurídico brasileiro. Ainda assim, também destacam a necessidade de melhorar a comunicação das explicações geradas pelo método *SHAP*, especialmente no que diz respeito à discussão das explicações e à revisão de termos técnicos.

Com o intuito de evitar um *looping* acadêmico científico, podemos inferir que a aplicação do método explicativo *SHAP* para modelos preditivos de IA no campo jurídico é bem recebida e considerada confiável pelos especialistas, o que reforça a relevância e a eficácia da nossa abordagem. Dito de outro modo, esse estudo preenche uma lacuna na literatura sobre a *XAI* e o Direito, oferecendo uma solução prática, não exclusiva e eficiente para os desafios da explicabilidade da IA.

## **6.2 Limitações**

No que tange as limitações, apesar dos resultados promissores e das contribuições significativas deste estudo, é importante reconhecer as limitações inerentes à metodologia e abordagem adotadas. Uma das principais limitações está relacionada à generalização dos resultados, uma vez que a aplicação do método *SHAP* para a explicabilidade da IA em decisões judiciais envolvendo ações de companhias aéreas pode não ser diretamente transferível para outros contextos jurídicos. Portanto, é crucial considerar a necessidade de adaptação e validação da abordagem em diferentes domínios do Direito.

Igualmente, outra limitação a ser destacada é a dependência da qualidade e disponibilidade dos dados judiciais utilizados no estudo. A precisão e representatividade das explicações geradas pelo método *SHAP* estão intrinsecamente ligadas à qualidade e abrangência dos dados de decisões judiciais. Dessa forma, a falta de dados completos ou representativos pode impactar a eficácia e confiabilidade das explicações geradas, limitando a aplicabilidade da metodologia em cenários com dados escassos ou enviesados. No caso em tela, boa parte dos dados já havia sido estruturados e estavam anonimizados.

Do mesmo modo, é importante ressaltar que a interpretação das explicações

geradas pelo método *SHAP* pode ser complexa e exigir um alto nível de expertise técnica e jurídica. A compreensão e análise das contribuições de cada característica de entrada para as predições do modelo de IA podem ser desafiadoras para os usuários finais, especialmente aqueles sem conhecimento especializado em ciência de dados ou direito. Portanto, destaca-se aqui a necessidade de simplificar e tornar as explicações mais acessíveis para diferentes públicos, isto é, deve ser considerada como uma limitação a ser superada.

Ademais, a validação da abordagem com especialistas, embora fundamental, pode apresentar limitações inerentes à subjetividade das opiniões e ao viés dos participantes. A diversidade de perspectivas e experiências dos especialistas pode influenciar a avaliação da viabilidade e eficácia da abordagem proposta, o que pode introduzir incertezas na interpretação dos resultados da validação. Portanto, é importante considerar a necessidade de abordagens complementares para a validação, a fim de mitigar possíveis vieses e garantir a robustez das conclusões.

Como nem tudo são flores, outra limitação importante está relacionada à dinâmica e evolução das práticas jurídicas e tecnológicas, que podem impactar a relevância e eficácia da abordagem proposta ao longo do tempo. O avanço da legislação, a evolução das práticas judiciais e o desenvolvimento de novas técnicas e ferramentas de IA podem influenciar a aplicabilidade e utilidade da metodologia baseada no método *SHAP*. Portanto, é crucial considerar a necessidade de atualização e adaptação contínua da abordagem para acompanhar as mudanças no ambiente jurídico e tecnológico.

Por fim, é importante reconhecer que a aplicação de métodos explicativos de predições de IA no contexto jurídico apresenta desafios éticos e sociais que vão além das considerações técnicas e metodológicas. A discussão sobre a responsabilidade, equidade e justiça na utilização da IA no sistema jurídico é complexa e multifacetada, e as limitações éticas e sociais associadas à explicabilidade da IA devem ser cuidadosamente consideradas. Como visto, na revisão sistemática de literatura (*SLR*), existe uma necessidade de abordar questões éticas e sociais mais amplas relacionadas à *XAI* no contexto jurídico. Assim sendo, também deve ser reconhecida como uma limitação e um desafio a ser enfrentado em futuras pesquisas e aplicações práticas.

### 6.3 Trabalhos futuros

A partir dos resultados e limitações identificadas nos experimentos com *XAI* e o método explicativo *SHAP* para previsões de IA acerca de decisões judiciais envolvendo companhias aéreas, há várias possibilidades de trabalhos futuros que podem contribuir para o avanço da *XAI* no contexto jurídico. Uma das principais áreas de pesquisa que pode ser explorada é o aprimoramento da metodologia de explicabilidade da IA, com o objetivo de tornar as explicações mais acessíveis e compreensíveis para diferentes públicos. Isso pode envolver o desenvolvimento de técnicas de visualização de dados mais intuitivas e aprimoradas, bem como a simplificação da linguagem técnica utilizada nas explicações.

Oportunidades também surgem para a aplicação de mecanismos de *fine tuning* para melhorar a precisão e eficácia dos modelos de IA utilizados no contexto jurídico. O *fine tuning* envolve o ajuste de parâmetros específicos do modelo de IA para melhorar sua capacidade de generalização e adaptação a diferentes contextos. Isso pode ser particularmente útil no contexto jurídico, onde a complexidade e a diversidade dos casos podem exigir modelos de IA mais sofisticados e adaptáveis.

Outrossim, a utilização de *LLMs* (*Language Model for Legal Text*) pode ser uma possibilidade interessante para aprimorar a capacidade de compreensão e análise de textos jurídicos por parte dos modelos de IA. O *ChatGPT*<sup>194</sup>, por exemplo, é um *LLM* recentemente lançado que pode ser utilizado para gerar respostas automáticas a perguntas jurídicas complexas. A aplicação de *LLMs* pode ser uma área de pesquisa promissora para melhorar a eficácia e precisão dos modelos de IA no contexto jurídico.

Nesse sentido, entre as ferramentas que mais tiveram destaque recentemente podemos citar o *CoCounsel* desenvolvido pela *CaseText*<sup>195</sup>, que é um assistente de redação jurídica baseado em inteligência artificial que foi mencionado no documento fornecido. Ele é descrito como um assistente de redação jurídica baseado no *GPT-4* da

---

<sup>194</sup> The introduction of ChatGPT has set off an arms race amongst large companies and start-ups alike on creating the next best LLM. A key recurring theme has been that simply increasing the number of parameters in the model was no longer the sole method of improving LLM quality (LUK, Martin. *Generative AI: Overview, economic impact, and applications in asset management. Economic Impact, and Applications in Asset Management*, p.26, 2023. Available in: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4574814](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4574814)>).

<sup>195</sup> *Casetext. Cocounsel*. 2023. Disponível em: <<https://casetext.com/>>.

*OpenAI*, que é adaptado para a área jurídica, com base no treinamento por advogados e especialistas em IA. Outro notável exemplo, é o *Robin AI*<sup>196</sup> que é uma plataforma que oferece assistência na criação e revisão de contratos legais, utilizando o modelo de linguagem *GPT-4* da *OpenAI*. A *Robin AI* é descrita como utilizando o modelo *Claude* da *Anthropic*, integrado para também auxiliar na criação e revisão de contratos, bem como na realização de pesquisas relacionadas a questões legais. Ambas, ilustram como a IA generativa está sendo utilizada para desenvolver soluções comerciais específicas para o setor jurídico, demonstrando o potencial impacto da IA na prática legal.

Além disso, a utilização de técnicas de interpretabilidade de modelos de IA pode ser uma área de pesquisa promissora para melhorar a transparência e confiabilidade dos modelos de IA no contexto jurídico. Isso pode envolver o desenvolvimento de técnicas de interpretabilidade mais avançadas, como a análise de sensibilidade e a decomposição de modelos, bem como a utilização de técnicas de interpretabilidade em conjunto com outras técnicas de explicabilidade.

Conforme identificado no início do estudo, outra possibilidade de trabalho futuro é a aplicação de técnicas de privacidade e segurança de dados para garantir a proteção dos dados pessoais e sensíveis utilizados nos modelos de IA no contexto jurídico. Isso pode envolver o desenvolvimento de técnicas de anonimização e criptografia de dados, bem como a utilização de técnicas de segurança cibernética para proteger os modelos de IA contra ataques maliciosos.

Por fim, não menos importante, a realização de novos estudos empíricos para avaliar a eficácia e impacto da *XAI* no contexto jurídico pode ser uma área de pesquisa promissora para entender melhor os desafios e oportunidades associados à utilização de IA no sistema jurídico. Isso pode envolver a realização de estudos de caso em diferentes domínios do Direito, bem como a avaliação da percepção e aceitação dos usuários finais em relação à *XAI* no contexto jurídico.

Conclui-se, portanto, que há várias possibilidades de trabalhos futuros que podem contribuir para o avanço da *XAI* no contexto jurídico, incluindo o aprimoramento da metodologia de explicabilidade da IA, a aplicação de mecanismos de *fine tuning* e *LLMs*, a utilização de técnicas de aprendizado por reforço e interpretabilidade, a aplicação de

---

<sup>196</sup> *RobinAi*. *RobinAi*™. 2023. Disponível em: <<https://www.robinai.com/>>.

técnicas de privacidade e segurança de dados, e a realização de estudos empíricos para avaliar a eficácia e impacto da *XAI* no sistema jurídico.

## REFERÊNCIAS

ABRAHAM, Marcus; CATARINO, João Ricardo. **O uso da inteligência artificial na aplicação do direito público: o caso especial da cobrança dos créditos tributários-um estudo objetivado nos casos brasileiro e português.** e-Pública: Revista Eletrônica de Direito Público, v. 6, n. 2, 2019. Disponível em: <<https://e-publica.pt/api/v1/articles/34335-o-uso-da-inteligencia-artificial-na-aplicacao-do-direito-publico-o-caso-especial-da-cobranca-dos-creditos-tributarios-um-estudo-objetivado-nos-cas.pdf>>. Acesso em: 16-01-2023.

ADDAMS, Jane. **Charity and Social Justice.** *The North American Review*, v. 192, n. 656, jul. 1910. Disponível em:< <https://www.jstor.org/stable/25106710>>. Acesso em: 05-03-2023.

ALLAHABADI, Himanshi et al. **Assessing Trustworthy AI in Times of COVID-19: Deep Learning for Predicting a Multiregional Score Conveying the Degree of Lung Compromise in COVID-19 Patients.** *IEEE transactions on technology and society*, v. 3, n. 4, p. 272-289, 2022. Disponível em:<<https://ieeexplore.ieee.org/abstract/document/9845195/>>. Acesso em: 17-01-2024.

ANDRADE, Mariana Dionísio de; ROSA, Beatriz de Castro; PINTO, Eduardo Régis Girão de Castro. **Legal tech: analytics, inteligência artificial e as novas perspectivas para a prática da advocacia privada.** *Revista Direito GV*, v. 16, 2020. Disponível em:<<https://www.scielo.br/j/rdgv/a/xL839bvvvK4QgvZfxwR6b4J/?lang=pt>>. Acesso em: 19-04-2023.

ANTONIK, Luis Roberto. **Compliance, Ética, Responsabilidade Social e Empresarial-Uma visão prática.** Alta Books Editora, 2016.

ARISTÓTELES. **Ética a Nicômaco.** In: ARISTÓTELES. *Obras incompletas.* São Paulo: Nova Cultural, 1996. (Coleção Os Pensadores).

**Árvore de decisão.** IBM. 2023. Disponível em:<<https://www.ibm.com/docs/pt-br/cognos-analytics/11.1.0?topic=analytics-decision-tree>>. Acesso em: 13-06-2022.

ASHLEY. John. **O que é IA explicável.** Nvidia. 2021. Disponível em:<<https://blog.nvidia.com.br/2021/08/02/o-que-e-ai-explicavel/>>. Acesso em: 21-04-2023.

ASIMOV, Isaac. **Eu, robô.** São Paulo: Aleph, 2014. Disponível em:<<https://books.google.com.br/books?id=-8-uCgAAQBAJ>>. Acesso em: 16-01-2023.

AVELLA, Marcela del Pilar Roa; SANABRIA-MOYANO, Jesús Eduardo; DINAS-HURTADO, Katherin. **Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos.** *Revista Brasileira de Direito Processual Penal*, v. 8, 2022. Disponível

em:<<https://www.scielo.br/j/rbdpp/a/6W9b8CHYbXcsc6qczDxCSfr/?lang=es>>. Acesso em: 22-04-2023.

BANDEIRA, R. **Estudo revela realidade e desafios dos Juizados Especiais**. CNJ. 2020. Disponível em:<<https://www.cnj.jus.br/estudo-revela-realidade-e-desafios-dos-juizados-especiais/>>. Acesso em: 06-02-2023

BARRIOS-TAO, Hernando; DÍAZ, Vianney; GUERRA, Yolanda M. **Purposes of education along with artificial intelligence developments**. Cadernos de Pesquisa, v. 51, 2021. Disponível em:<<https://www.scielo.br/j/cp/a/4xLrQkM5v36QqnQRP8ZmMPC/?lang=en>>. Acesso em: 22-04-2023

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidente da República, [2016]. Disponível em:<[https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm)>. Acesso em: 15-05-2023.

BRASIL. **LEI Nº 7.244, DE 7 DE NOVEMBRO DE 1984**. Dispõe sobre a criação e o funcionamento do Juizado Especial de Pequenas Causas. Disponível em:<[https://www.planalto.gov.br/ccivil\\_03/leis/1980-1988/17244.htm#:~:text=L7244&text=LEI Nº 7.244%2C DE 7 DE NOVEMBRO DE 1984.&text=Dispõe sobre a criação e,Juizado Especial de Pequenas Causas.>](https://www.planalto.gov.br/ccivil_03/leis/1980-1988/17244.htm#:~:text=L7244&text=LEI%20N%267%20DE%20NOVEMBRO%20DE%201984.&text=Disp%263e%20sobre%20a%20cria%27o%20e%20o%20funcionamento%20do%20Juizado%20Especial%20de%20Pequenas%20Causas.>)>. Acesso em: 19-05-2023.

BRASIL. L.9099. **LEI Nº 9.099, DE 26 DE SETEMBRO DE 1995**. Dispõe sobre os Juizados Especiais Cíveis e Criminais e dá outras providências. Disponível em:<[https://www.planalto.gov.br/ccivil\\_03/leis/19099.htm#:~:text=L9099&text=LEI Nº 9.099%2C DE 26 DE SETEMBRO DE 1995.&text=Dispõe sobre os Juizados Especiais Cíveis e Criminais e dá outras providências>](https://www.planalto.gov.br/ccivil_03/leis/19099.htm#:~:text=L9099&text=LEI%20N%269%20DE%2026%20DE%20SETEMBRO%20DE%201995.&text=Disp%263e%20sobre%20os%20Juizados%20Especiais%20C%263ives%20e%20Criminais%20e%20d%263a%20outras%20provid%263encias.>)>. Acesso em: 18-05-2023.

BRASIL. **LGPD**. Disponível em:<[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)>. Acesso em: 13-05-2023.

CÁRDENAS, Erick Rincón; MOLANO, Valeria Martinez. **Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisiones judiciales**. Revista Direito GV, v. 17, 2021. Disponível em:<<https://www.scielo.br/j/rdgv/a/vZDXYYPRrcwgsjJDWQf97QG/>>. Acesso em: 02-05-2023.

CARDONA, Noemí Jiménez. **La Mediación Mercantil Ante los Nuevos Retos de la Inteligencia Artificial: Una Visión Comparada Entre los Ordenamientos Jurídicos Español e Italiano**. Revista Internacional Consinter de Direito, 2022. Disponível:<<https://revistaconsinter.com/index.php/ojs/article/view/24>>. Acesso em: 03-05-2023.



CASETEXT. **Cocouncil**. 2023. Disponível em:<<https://casetext.com/>>. Acesso em: 20-01-2024.

CHÁVEZ VALDIVIA, Ana Karin. **No es solo un robot: consideraciones en torno a una nueva personalidad jurídica y el redimensionamiento de las relaciones interpersonales**. Ius et Praxis, v. 26, n. 2, 2020. Disponível em:<[https://www.scielo.cl/scielo.php?pid=S0718-00122020000200055&script=sci\\_arttext&tlng=pt](https://www.scielo.cl/scielo.php?pid=S0718-00122020000200055&script=sci_arttext&tlng=pt)>. Acesso em: 04-05-2023.

CHITIMIRA, Howard; NCUBE, Princess. **The regulation and use of artificial intelligence and 5G technology to combat cybercrime and financial crime in South African banks**. Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, v. 24, n. 1, 2021. Disponível em:<<https://www.ajol.info/index.php/pej/article/view/235606>>. Acesso em: 05-05-2023.

**Clusterização: o que é, importância e aplicações**. Fiveacts. 2021. Disponível em:<<https://www.fiveacts.com.br/clusterizacao/>>. Acesso em: 18-05-2023.

CULTURAL, Nova. **Ética a Nicômaco**. Aristóteles: Obras incompletas, 1996. Disponível em: <<https://abdet.com.br/site/wp-content/uploads/2014/12/Ética-a-Nicômaco.pdf>>. Acesso em: 16-01-2024.

DIDUGU. C. **Anchor: A simple Introduction**. Medium. 2020. p.2. Disponível em:<<https://medium.com/swlh/ultimate-guide-to-model-explainability-anchors-2deab8239f57>>.

DINUCCI, Aldo; JULIEN, Alfredo. **O encheirídon de Epicteto**. São Cristóvão: Editora UFS, 2012. Acesso em: 28-05-2023.

DOURADO, Daniel de Araujo; AITH, Fernando Mussa Abujamra. **A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais**. Revista de Saúde Pública, v. 56, 2022. Disponível em: <<https://www.scielo.br/j/rsp/a/k38jGvJdbQSYN4MpzGZpfXw/?lang=pt>>. Acesso em: 23-05-2023.

DUARTE, Marcelo. **Qual é a origem da palavra “robô”?** O Guia dos Curiosos, 2021. Disponível em: <<https://www.guiadoscuriosos.com.br/lingua-portuguesa/qual-e-a-origem-da-palavra>>. Acesso em: 17-05-2023.

ESPINOSA, Baruch de. **Breve Tratado: de Deus, do homem e do seu bem-estar**. Tradução: Emanuel Angelo da Rocha Frago e Luís César Guimarães Oliva. Belo Horizonte: Autêntica, 2014. Disponível em:<<https://search.proquest.com/openview/c54783434155d39e1e7b55319b9db4af/1?pq-origsite=gscholar&cbl=1206337>>. Acesso em: 02-05-2023.

EUSEBIO, B. **Counterfactual Explanations in Model Interpretations**. Medium. 2020. p.10. Disponível em:<<https://towardsdatascience.com/counterfactual-explanations-in-model-interpretations-a73caec5b74b>>.

FERRARIO, Andrea; LOI, Michele. **The robustness of counterfactual explanations over time**. IEEE Access, v. 10, p. 82736-82750, 2022. Disponível em:<<https://ieeexplore.ieee.org/abstract/document/9851645>>. Acesso em: 15-01-2024.

**Forrester Paper: Predictions 2021: Cybersecurity**. Disponível em:<<https://www.code42.com/resources/reports/forrester-paper-predictions-2021-cybersecurity>>. Acesso em: 19-05-2023.

GRANADOS FERREIRA, Jackeline. **Análisis de la inteligencia artificial en las relaciones laborales**. Revista CES Derecho, v. 13, n. 1, p. 111-132, 2022. Disponível em:<[http://www.scielo.org.co/scielo.php?pid=S2145-77192022000100111&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S2145-77192022000100111&script=sci_arttext)>. Acesso em: 13-05-2023.

**Guia essencial para IA explicável (XAI)**. Alteryx. 2023. p.1. Disponível em:<<https://www.alteryx.com/pt-br/resources/whitepaper/essential-guide-to-explainable-ai>>. Acesso em: 12-05-2023.

HARARI, Yuval Noah. **21 lições para o século 21**. São Paulo: Companhia das Letras, 2018. Disponível em:<[https://edisciplinas.usp.br/pluginfile.php/7667521/mod\\_resource/content/5/21-Lições-Para-o-Século-21-by-Yuval-Noah-Harari.pdf](https://edisciplinas.usp.br/pluginfile.php/7667521/mod_resource/content/5/21-Lições-Para-o-Século-21-by-Yuval-Noah-Harari.pdf)>. Acesso em: 03-05-2023.

**JUTGE.ORG**. 2023. Disponível em:<<https://jutge.org/>>. Acesso em: 18-01-2023.

KANT, Emmanuel. **Fundamentação da Metafísica dos Costumes**. Trad. Paulo Quintela. Lisboa: Edições 70. Disponível em:<[https://books.google.com.br/books?hl=pt-BR&lr=&id=1Jq3EAAAQBAJ&oi=fnd&pg=PA7&dq=KANT,+Emmanuel.+Fundamenta%C3%A7%C3%A3o+da+Metaf%C3%ADsica+dos+Costumes.+Trad.+Paulo+Quintela.Lisboa:+Edi%C3%A7%C3%B5es+70.&ots=30taBu\\_W4k&sig=73jAOkhICuRT8xdqOyKFFDqYFp0](https://books.google.com.br/books?hl=pt-BR&lr=&id=1Jq3EAAAQBAJ&oi=fnd&pg=PA7&dq=KANT,+Emmanuel.+Fundamenta%C3%A7%C3%A3o+da+Metaf%C3%ADsica+dos+Costumes.+Trad.+Paulo+Quintela.Lisboa:+Edi%C3%A7%C3%B5es+70.&ots=30taBu_W4k&sig=73jAOkhICuRT8xdqOyKFFDqYFp0)>. Acesso em: 26-05-2023.

LIMBERGER, Têmis; GIANNAKOS, Demétrio Beck da Silva; SZINVELSKI, Mártin M. **Can Judges be Replaced by Machines? The Brazilian Case**. Mexican law review, v. 14, n. 2, p. 53-81, 2022. Disponível em:<[https://www.scielo.org.mx/scielo.php?pid=S1870-05782022000100053&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S1870-05782022000100053&script=sci_arttext)>. Acesso em: 22-05-2023.

**LIVIA**. INTELIJUS. 2023. Disponível em:<<https://www.intelijus.ai/>>. Acesso em: 20-01-2024.

MACOHIN, A. **Transparência: elemento fundamental para uma IA responsável**. SERPRO. 2024. Disponível em:<<https://www.serpro.gov.br/menu/noticias/noticias-2024/transparencia-ia-responsavel>>. Acesso em: 20-01-2024.

MATTHEE, Machdel. **Cyborgs and the future of the human spirit**. Tydskrif vir Geesteswetenskappe, v. 53, n. 4, p. 530-557, 2013. Disponível:<[http://www.scielo.org.za/scielo.php?pid=S0041-47512013000400005&script=sci\\_arttext](http://www.scielo.org.za/scielo.php?pid=S0041-47512013000400005&script=sci_arttext)>. Acesso em: 30-05-2023.

MATURANA, Humberto; VARELA, Francisco. **A árvore do conhecimento**. São Paulo: Editorial Psy II. 1995. Disponível em:<<http://materiaapoioaotcc.pbworks.com/f/Arvore+do+Conhecimento+Maturana+e+Varela.pdf>>. Acesso em: 15-08-2022.

MATURANA, Humberto; PÖRKSEN, Bernhard. **Del ser al hacer**. Ed. JC Sáez, 2004. Disponível em: <[https://des-juj.infed.edu.ar/sitio/educacion-emocional-2019/upload/Maturana\\_Romesin\\_H\\_-\\_Del\\_Ser\\_Al\\_Hacer.pdf](https://des-juj.infed.edu.ar/sitio/educacion-emocional-2019/upload/Maturana_Romesin_H_-_Del_Ser_Al_Hacer.pdf)>. Acesso em: 21-08-2022.

MATURANA, Humberto. **La objetividad. Un argumento para obligar**. Santiago de Chile: Dolmen, 1997. Disponível em:< [https://des-juj.infed.edu.ar/sitio/educacion-emocional-2019/upload/Maturana\\_Humberto\\_-\\_La\\_Objektividad\\_Un\\_Argumento\\_Para\\_Obligar.PDF](https://des-juj.infed.edu.ar/sitio/educacion-emocional-2019/upload/Maturana_Humberto_-_La_Objektividad_Un_Argumento_Para_Obligar.PDF)>. Acesso em: 28-08-2022.

MOSSÉ, C. **Dicionário da Civilização Grega**. Trad. Carlos Ramallete. Rio de Janeiro: Jorge Zahar Editor, 2004. Disponível em:<[https://www.academia.edu/download/58913222/dicionario\\_da\\_civilizacao\\_grega.pdf](https://www.academia.edu/download/58913222/dicionario_da_civilizacao_grega.pdf)>. Acesso em: 23-02-2023.

MOULIN, Carolina Stange Azevedo. **Métodos de resolução digital de controvérsias: estado da arte de suas aplicações e desafios**. Revista Direito GV, v. 17, 2021. Disponível em: <<https://www.scielo.br/j/rdgv/a/K6Td7TJ6fcMtpyRDWVdzbPN/?lang=pt>>. Acesso em: 15-01-2023.

NDLOVU, Lonias. **Enhancing the value of patents as corporate assets in South Africa: How can artificial intelligence (AI) assist?**. Potchefstroom Electronic Law Journal/Potchefstroomse Elektroniese Regsblad, v. 24, n. 1, 2021. Disponível em:<<https://www.ajol.info/index.php/pelj/article/view/235800>>. Acesso em: 01-05-2023.

NGUYEN, Huu-Thiet; LI, Sitan; CHEAH, Chien Chern. **A layer-wise theoretical framework for deep learning of convolutional neural networks**. IEEE Access, v. 10, p. 14270-14287, 2022. Disponível em:<<https://ieeexplore.ieee.org/abstract/document/9698045/>>. Acesso em: 13-01-2024.

**O que é aprendizado não supervisionado**. IBM. 2023. Disponível em: <<https://www.ibm.com/br-pt/topics/unsupervised-learning>>. Acesso em: 03-05-2023.

**O que é IA explicável (XAI)?** IBM. 2023. p.1. Disponível em:<<https://www.ibm.com/watson/explainable-ai>>. Acesso em: 03-05-2023.

**O que é regressão logística.** IBM. 2023. Disponível em:< <https://www.ibm.com/br-pt/topics/logistic-regression>>. Acesso em: 05-05-2023.

**O que são Redes Neurais?** IBM. 2023. Disponível em:<<https://www.ibm.com/br-pt/topics/neural-networks>>. Acesso em: 04-05-2023.

PÁDUA, Mateus. **Machine Learning -Métricas de avaliação: Acurácia, Precisão e Recall, F1-score.** *MEDIUM.* 2020. Disponível em:<<https://medium.com/@mateuspdua/machine-learning-métricas-de-avaliação-acurácia-precisão-e-recall-d44c72307959>>. Acesso em: 17-09-2023.

PÁDUA, Sergio. **O Juiz Ciborgue: Inteligência Artificial e Decisão Judicial.** CONPEDI. 2020. Disponível em:<[https://www.academia.edu/98912842/O\\_JUIZ\\_CIBORGUE\\_INTELIG%C3%80NCIA\\_ARTIFICIAL\\_E\\_DECISAO\\_JUDICIAL?uc-sb-sw=9525500](https://www.academia.edu/98912842/O_JUIZ_CIBORGUE_INTELIG%C3%80NCIA_ARTIFICIAL_E_DECISAO_JUDICIAL?uc-sb-sw=9525500)>. Acesso em: 22-02-2024.

PEIXOTO, Fabiano Hartmann. **Direito e inteligência artificial: referenciais básicos com comentários à resolução CNJ 332/2020.** Brasília, DF: Ed do autor: DR.IA., 2020. E-book. DOI: 10.29327/521174. Disponível em: [https://drive.google.com/file/d/1eqeHvPFT\\_4OnBMnXUkCFYxcCRcbp\\_Hr2/view](https://drive.google.com/file/d/1eqeHvPFT_4OnBMnXUkCFYxcCRcbp_Hr2/view). Acesso em: 13 abr. 2022.

PENA, Ana María Neira. **Inteligencia artificial y tutela cautelar. Especial referencia a la prisión provisional.** Revista Brasileira de Direito Processual Penal, v. 7, p. 1897-1933, 2022. Disponível:<<https://www.scielo.br/j/rbdpp/a/BkS7hyMvvTf96xfCY4CfbLz/?lang=es>>. Acesso em: 11-04-2023.

PEREIRA, Filipe Dwan et al. **Explaining individual and collective programming students' behavior by interpreting a black-box predictive model.** IEEE Access, v. 9, p. 117097-117119, 2021. Disponível em:< <https://ieeexplore.ieee.org/abstract/document/9517104/>>. Acesso em: 09-01-2024.

PORCELLI, Adriana Margarita. **Artificial Intelligence Applied to Robotics in Armed Conflict. Debates on Autonomous Lethal Weapons Systems and the (In) Sufficiency of the Protocols of International Humanitarian Law.** Estudios Socio-Juridicos, v. 23, p. 483, 2021. Disponível em:<[https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/estscj23&section=23](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/estscj23&section=23)>. Acesso em: 13-04-2023.

POLO, Tatiana Cristina Figueira; MIOT, Hélio Amante. **Aplicações da curva ROC em estudos clínicos e experimentais.** Jornal Vascular Brasileiro, v. 19, p. 1-2, 2020. in Hoo ZH, Candlish J, Teare D. What is an ROC curve? Emerg Med J. 2017;34(6):357-9. <http://dx.doi.org/10.1136/emered-2017-206735>. PMID:28302644; . Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8(4):283-98. [http://dx.doi.org/10.1016/S0001-2998\(78\)80014-2](http://dx.doi.org/10.1016/S0001-2998(78)80014-2). PMID:112681.

Disponível:<<https://www.scielo.br/j/jvb/a/8S8Pfqnz8csmQJVqwgZT8gH/>>. Acesso em: 16-03-2023.

QAFFAS, ALAA ASIM et al. **Interpretable Multi-criteria ABC analysis based on semi-supervised clustering and Explainable Artificial Intelligence**. IEEE Access, 2023. Disponível em:< <https://ieeexplore.ieee.org/abstract/document/10114405/>>. Acesso em: 18-01-2024.

**Redes Neurais, O que são e qual sua importância?** SAS. p.4-5. 2023. Disponível em: <[https://www.sas.com/pt\\_br/insights/analytics/neural-networks.html](https://www.sas.com/pt_br/insights/analytics/neural-networks.html)>. Acesso em: 18-05-2023.

**Regressão Linear**. IBM. 2023. Disponível em:< <https://www.ibm.com/br-pt/analytics/learn/linear-regression>>. Acesso em: 05-05-2023.

RICHMOND, Karen McGregor et al. **Explainable AI and Law: An Evidential Survey**. Digital Society, v. 3, n. 1, p. 1, 2024. Disponível em:< <https://link.springer.com/article/10.1007/s44206-023-00081-z>>. Acesso em: 14-01-2024.

ROCHA, Arlindo Carvalho. **O processo orçamentário brasileiro como instrumento de accountability**. Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração, 2008. Acesso em: 09-09-2023.

**ROBÔ, o que é? Origem, significado e representações na cultura**. Segredos do Mundo. Disponível em: <<https://segredosdomundo.r7.com/robo-significado/>>. Acesso em: 08-09-2023.

RUSSELL, Stuart J. **1962- Inteligência artificial** / Stuart Russell, Peter Norvig; tradução Regina Célia Simille. – Rio de Janeiro: Elsevier, 2013. Acesso em: 03-08-2023.

SABO, Isabela Cristina et al. **A machine learning-based model for judgement results prediction and support in Brazilian Special Court? s conciliation hearings**. 2022. Disponível em:<<https://repositorio.ufsc.br/handle/123456789/241026>>. Acesso em: 29-08-2023.

SABO, Isabela Cristina; ROVER, Aires José. **Observância de precedentes e gestão de demandas repetitivas por meio do aprendizado de máquina**. Revista Opinião Jurídica, v. 18, n. 28, p. 69-93, 2020. Disponível em:<<https://www.redalyc.org/journal/6338/633868859003/633868859003.pdf>>. Acesso em: 20-02-2024.

SANCHES, Marcelo Kaminski. **Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados**. 2003. Tese de Doutorado. Universidade de São Paulo. p.71. Disponível em: < [https://teses.usp.br/teses/disponiveis/55/55134/tde-12102003-140536/publico/Dissertacao\\_MKS.pdf](https://teses.usp.br/teses/disponiveis/55/55134/tde-12102003-140536/publico/Dissertacao_MKS.pdf) >. Acesso em: 13-11-2023.

SANCHEZ MORALES, Santiago. **Ciencia Ficción como Fuente de Principios Jurídicos para regular la Inteligencia Artificial**. Revista IUS, v. 15, n. 48, p. 55-76, 2021. Disponível em:<[https://www.scielo.org.mx/scielo.php?pid=S1870-21472021000200055&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S1870-21472021000200055&script=sci_arttext)>. Acesso em: 30-05-2023.

SANTOS, Hugo Luz dos. **Processo Penal e Inteligência Artificial: Rumo a um Direito (Processual) Penal da Segurança Máxima?** Revista Brasileira de Direito Processual Penal, v. 8, p. 767-821, 2022. Disponível em:<<https://www.scielo.br/j/rbdpp/a/fg4rvTXgNxBPH6VjknC5d9C/>>. Acesso em: 23-05-2023.

SANTOS, Maria Tereza. **A ética nos estudos acerca do medievo**. Acta Scientiarum. Education, v. 42, 2020. P.10. Disponível em:<[educa.fcc.org.br/scielo.php?script=sci\\_arttext&pid=S2178-52012020000100115#:~:text=A%20ética%20medieval%20tem%20a,valorando%20a%20ação%20como%20transcendência](https://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S2178-52012020000100115#:~:text=A%20ética%20medieval%20tem%20a,valorando%20a%20ação%20como%20transcendência)>. Acesso em: 05-06-2023.

SANABRIA MOYANO, Jesús E.; ROA AVELLA, Marcela del Pilar; LEE PÉREZ, Oscar Iván. **Tecnología de reconocimiento facial y sus riesgos en los derechos humanos**. Revista Criminalidad, v. 64, n. 3, p. 61-78, 2022. Disponível em:<[http://www.scielo.org.co/scielo.php?pid=S1794-31082022000300061&script=sci\\_arttext](http://www.scielo.org.co/scielo.php?pid=S1794-31082022000300061&script=sci_arttext)>. Acesso em: 07-06-2023.

SABO, Isabela Cristina et al. **A machine learning-based model for judgement results prediction and support in Brazilian Special Court? conciliation hearings**. 2022. Disponível em:<<https://repositorio.ufsc.br/handle/123456789/241026>>. Acesso em: 13-11-2022.

**Sobre o SVM**. IBM. 2023. Disponível em:<<https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-about-svm>>. Acesso em: 11-05-2023.

SCHMITT, Jeovani et al. **Pré-processamento para a mineração de dados: uso da análise de componentes principais com escalonamento ótimo**. 2005. p.25 in KAMBER, Jiawei Han Micheline; 范明. 数据挖掘概念与技术. Data Mining Concepts and Techniques, v. 8, 2001, p.105. Disponível em:<<http://repositorio.ufsc.br/handle/123456789/101803>>. Acesso em: 19-06-2023.

STEIW. Leandro. **Saiba como funciona a mineração de dados (ou data mining)**. Insper. 2022. Disponível em:<<https://www.insper.edu.br/noticias/mineracao-de-dados-ou-data-mining/>>. Acesso em: 01-06-2023.

STUART, Russel. **Inteligência Artificial a nosso favor: como manter o controle sobre a tecnologia**. São Paulo: Companhia das Letras, 2021. Disponível em:<<https://www.companhiadasletras.com.br/livro/9786559213085/inteligencia-artificial-a-nosso-favor>>. Acesso em: 15-11-2022.

TEGMARK, Max. Life 3.0. **Being human in the age of artificial intelligence**. New York: Penguin, 2017. Disponível em: <[https://brill.com/view/journals/rt/26/1-2/article-p169\\_8.xml](https://brill.com/view/journals/rt/26/1-2/article-p169_8.xml)>. Acesso em: 22-07-2022.

**Underfitting e Overfitting**. Didática Tech. 2023. Disponível: <<https://didatica.tech/underfitting-e-overfitting/>>. Acesso em: 15-05-2023.

**What is the k-nearest neighbors algorithm?**. IBM. 2023. Disponível em: <<https://www.ibm.com/topics/knn>>. Acesso em: 23-05-2023.

XAVIER, Otávio Calaça et al. **Identificação de evasão fiscal utilizando dados abertos e inteligência artificial**. Revista de Administração Pública, v. 56, 2022. Disponível: <<https://www.scielo.br/j/rap/a/5q38f9RdbQYSrZXF8zfDJqv/?format=html&lang=pt>>. Acesso em: 07-06-2023.

**2019 Trustwave Global Security Report, AT&T Cybersecurity / Insights Report V8.5. Telefonica Global Security Report**. Disponível em: <<http://www.portaldaprivacidade.com.br/wp-content/uploads/2019/05/2019-trustwave-global-security-report.pdf>>. Acesso em: 08-06-2023.

WEERTS, Hilde et al. **Algorithmic Unfairness through the Lens of EU Non-Discrimination Law: Or Why the Law is not a Decision Tree**. In: **Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**. 2023. p. 805-816. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/3593013.3594044>>. Acesso em 20-01-2023.

ZHANG, Zhibo et al. **Explainable artificial intelligence applications in cyber security: State-of-the-art in research**. IEEE Access, 2022. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9875264/>>. Acesso em 20-01-2023.

**APÊNDICE A – QUESTIONÁRIO DE PESQUISA COM EXPERT E  
JEC/UFSC**



# Questionário de Validação de Experimento com IA explicável (XAI) em Juizado Especial Cível do Norte da Ilha

Programa de Pós-Graduação em Direito da UFSC

Grupo de Pesquisa: E-gov

Docente (Orientador): Professor Aires José Rover

Discente: Bruno Cassol da Silva (Mestrando em Direito pela UFSC)

## Objetivo do questionário:

Este questionário tem como objetivo validar o experimento realizado com explicabilidade de IA aplicada a decisões do Juizado Especial Cível do Norte da Ilha envolvendo ações de passagens aéreas. O experimento consistiu em utilizar o método SHAP (Shapley Additive Explanations), que permite a identificação dos fatores mais relevantes para a tomada de decisão de um modelo preditivo de IA. Desse modo, buscamos avaliar a aplicação de métodos explicativos de predições de IA no campo jurídico, mais especificamente no Judiciário Brasileiro.

## Instruções:

- O questionário é composto por 7 (sete) perguntas de múltipla escolha.
- As perguntas discursivas devem ser respondidas com suas próprias palavras, de forma clara e objetiva.
- As perguntas de múltipla escolha devem ser marcadas com um X na opção que melhor representa sua opinião.
- O questionário é anônimo e leva cerca de 10 minutos para ser concluído.
- Conjuntamente ao questionário, foi encaminhada uma amostragem da experimentação de inteligência artificial explicável em dados estruturados de decisões judiciais do Juizado Especial Cível do Norte da Ilha.
- Agradecemos sua colaboração e participação neste estudo.

1. 1- Você conhece o conceito de inteligência artificial (IA) e como ela pode ser aplicada no campo jurídico? \*

Marcar apenas uma oval.

- Não sei nada sobre o assunto. (1)
- Conheço pouco sobre o assunto. (2)
- Tenho conhecimento razoável sobre o assunto. (3)
- Conheço bem o assunto. (4)
- Conheço muito bem o assunto. (5)

2. 2- Concorda que a IA pode ser útil para auxiliar os juizes na análise e na decisão de processos judiciais? \*

Marcar apenas uma oval.

- Discordo totalmente. (1)
- Tenho pouca esperança. (2)
- Acredito razoavelmente. (3)
- Concordo que existe possibilidade. (4)
- Concordo totalmente. (5)

3. 3- Você acredita no futuro de uma ferramenta que pudesse prever o resultado de uma sentença judicial baseada em dados de outras decisões do mesmo JEC, como no caso de ações envolvendo passagens aéreas? \*

Marcar apenas uma oval.

- Não me sinto a vontade com o uso de IA no Judiciário. (1)
- Acredito pouco no futuro de uma ferramenta desse tipo. (2)
- Razoavelmente, a depender da segurança dos resultados e funcionamento. (3)
- Bastante, dada importância da implantação de ferramentas de IA explicável. (4)
- Totalmente, acredito que isso pode ajudar a tornar o Judiciário mais eficiente. (5)

4. 4- O quanto você gostaria de saber como a IA chegou a uma determinada predição ou decisão? Você considera importante ter uma explicação clara e transparente sobre os fatores que influenciaram a decisão da IA?

Marcar apenas uma oval.

- Simplesmente não tenho interesse. Não é importante. (1)
- Tenho muito pouco interesse. Pode vir a ser importante algum dia. (2)
- Tenho um interesse razoável. De fato, tem a sua importância. (3)
- Tenho interesse considerável. Considero importante. (4)
- Tenho muito interesse. Considero importantíssimo. (5)

5. 5- A partir dos resultados do experimento apresentado com IA explicável, você acredita que isso possa ser utilizado para tornar o uso de IA, em um futuro próximo, mais confiável pelo Judiciário?

Marcar apenas uma oval.

- Não acredito. (1)
- Acredito muito pouco. (2)
- Talvez seja possível. (3)
- Tem grandes chances de ser utilizada pelo Judiciário. (4)
- Acredito totalmente no potencial do experimento. (5)

6. 6- A partir de sua ótica, qual o nível de confiabilidade dos resultados apresentados pela experimentação?

Marcar apenas uma oval.

- Não confio. (1)
- Pouco confiável. (2)
- Razoavelmente confiável. (3)
- Bem confiável. (4)
- Muito confiável. (5)

7. 7- Você tem alguma dúvida, sugestão ou crítica sobre o experimento realizado com a ferramenta de explicabilidade de IA desenvolvida para o JEC? Se sim, por favor, escreva abaixo:

---

---

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

12/01/2024, 19:29

Questionário de Validação com Expert em IA sobre Experimento com IA explicável (XAI) em Juizado Especial Cível do Nort...

## Questionário de Validação com Expert em IA sobre Experimento com IA explicável (XAI) em Juizado Especial Cível do Norte da Ilha

Programa de Pós-Graduação em Direito da UFSC

Grupo de Pesquisa: E-gov

Docente (Orientador): Professor Aires José Rover

Discente: Bruno Cassol da Silva (Mestrando em Direito pela UFSC)

**Objetivo do questionário:**

Este questionário tem como objetivo validar o experimento realizado com explicabilidade de IA aplicada a decisões do Juizado Especial Cível do Norte da Ilha envolvendo ações de passagens aéreas. O experimento consistiu em utilizar o método SHAP (Shapley Additive Explanations), que permite a identificação dos fatores mais relevantes para a tomada de decisão de um modelo preditivo de IA. Desse modo, buscamos avaliar a aplicação de métodos explicativos de predições de IA no campo jurídico, mais especificamente no Judiciário Brasileiro.

**Instruções:**

- O questionário é composto por 7 (sete) perguntas de múltipla escolha.
- As perguntas discursivas devem ser respondidas com suas próprias palavras, de forma clara e objetiva.
- As perguntas de múltipla escolha devem ser marcadas com um X na opção que melhor representa sua opinião.
- O questionário é anônimo e leva cerca de 10 minutos para ser concluído.
- Conjuntamente ao questionário, foi encaminhada uma amostragem da experimentação de inteligência artificial explicável em dados estruturados de decisões judiciais do Juizado Especial Cível do Norte da Ilha.
- Agradecemos sua colaboração e participação neste estudo.

1. 1- Você está familiarizado com o uso de IA no campo jurídico? \*

*Marcar apenas uma oval.*

- Não sei nada sobre o assunto. (1)
- Já ouvi falar, porém desconheço a aplicação. (2)
- Razoavelmente, tenho conhecimento de seu uso e aplicação. (3)
- Sim, já tive contato com o uso de aplicação de IA no campo jurídico (4).
- Sim, além de dominar o tema, trabalho com IA no campo jurídico (5).

2. 2- Concorda que a IA pode contribuir para a melhoria da qualidade e da eficiência das decisões judiciais?

*Marcar apenas uma oval.*

- Discordo totalmente. (1)
- Tenho pouca esperança. (2)
- Acredito razoavelmente. (3)
- Concordo que existe possibilidade. (4)
- Concordo totalmente. (5)

3. 3- Você acredita no futuro de uma ferramenta que pudesse prever o resultado de uma sentença judicial baseada em dados de outras decisões do mesmo JEC, como no caso de ações envolvendo passagens aéreas?

*Marcar apenas uma oval.*

- Não me sinto a vontade com o uso de IA no Judiciário. (1)
- Acredito pouco no futuro de uma ferramenta desse tipo. (2)
- Razoavelmente, a depender da segurança dos resultados e funcionamento. (3)
- Bastante, considero essencial a implantação de ferramentas de IA explicável. (4)
- Totalmente, acredito que isso pode ajudar a tornar o Judiciário mais eficiente. (5)

4. 4- O quanto você considera necessário que a IA forneça uma explicação clara e transparente sobre os fatores que influenciaram a sua predição ou decisão?

*Marcar apenas uma oval.*

- Não é necessário. (1)
- Pouco necessário. (2)
- Razoavelmente necessário. (3)
- Muito necessário. (4)
- Extremamente necessário. (5)

5. 5- A partir dos resultados do experimento apresentado com IA explicável, você acredita que isso possa ser utilizado para tornar o uso de IA, em um futuro próximo, mais confiável pelo Judiciário?

*Marcar apenas uma oval.*

- Não acredito. (1)
- Acredito muito pouco. (2)
- Talvez seja possível. (3)
- Tem grandes chances de ser utilizada pelo Judiciário. (4)
- Acredito totalmente no potencial do experimento. (5)

6. 6- A partir de sua ótica, qual o nível de confiabilidade dos resultados apresentados pela experimentação?

*Marcar apenas uma oval.*

- Não confio. (1)
- Pouco confiável. (2)
- Razoavelmente confiável. (3)
- Bem confiável. (4)
- Muito confiável. (5)

- 7- Você tem alguma dúvida, sugestão ou crítica sobre o experimento realizado com a ferramenta de explicabilidade de IA desenvolvida para o JEC? Se sim, por favor, escreva abaixo:

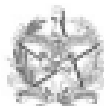
---

---

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

# ANEXO A – (JEC/UFSC) AUTORIZAÇÃO DE PESQUISA RELACIONADA A DADOS ANONIMIZADOS DE SENTENÇA DO JEC/UFSC



ESTADO DE SANTA CATARINA  
TRIBUNAL DE JUSTIÇA  
GABINETE DA PRESIDÊNCIA - NÚCLEO ADMINISTRATIVO

## DESPACHO

1. Trata-se de processo administrativo instaurado em razão da mensagem eletrônica enviada pelo Centro de Ciências Jurídicas da Universidade Federal de Santa Catarina, subscrita pelo coordenador do projeto de pesquisa científica, Professor Aires José Rover, objetivando autorização de acesso às sentenças proferidas pelo Juízo do Juizado Especial Cível e Criminal da Universidade Federal de Santa Catarina, a partir de 01/10/2020 até a data do término do projeto, para o projeto intitulado *"Inteligência Artificial e Justiça Catarinense: modelos preditos para gerar celeridade e transparência ao processo judicial"* (doc. n. 7500405).

Em anexo ao requerimento, foram colacionados os seguintes documentos: plano de trabalho (doc. n. 7500385), carta de anuência da Juíza de Direito titular do Juizado Especial Cível e Criminal da Universidade Federal de Santa Catarina (doc. n. 7500389), projeto de pesquisa (doc. n. 7500399), pedido de acesso e tratamento de dados de processos judiciais para projeto de pesquisa (doc. n. 7500405) e termo de responsabilidade e confidencialidade (doc. n. 75004120).

Ato contínuo, o Comitê Gestor de Proteção de Dados Pessoais - CGPDP elaborou parecer, opinando pelo deferimento do pedido com as seguintes considerações: *a) que o pedido de acesso às sentenças de processos ativos seja encaminhado ao juiz da causa; e b) sobre o pedido de acesso às sentenças de processos baixados, que sejam disponibilizadas as sentenças necessárias para o alcance da finalidade pretendida, sem o acesso ilimitado do requerente aos autos dos processos judiciais, a fim de observar o princípio da necessidade, segurança e prevenção (art. 6, III, VII e VIII, da LGPD)* (doc. n. 7515243).

2. Ressai dos autos que projeto de pesquisa pretende analisar sentenças de processos ativos e baixados proferidas no âmbito do Juizado Especial Cível e Criminal da UFSC (Fórum Norte da Ilha), tendo havido desde logo a anuência da Magistrada responsável pelos processos vinculados àquela unidade judiciária.

Com relação às sentenças de processos baixados, o requerente informa que *"o acesso será feito a partir da consulta a cada processo, por pesquisador da área do Direito, que identificará e armazenará o texto contido na sentença. Para tanto, é necessário o fornecimento, pelo Tribunal, dos números dos processos que interessam à base objeto da pesquisa, considerando os seguintes parâmetros das Tabelas Processuais Unificadas"*.

Entretanto, cumpre destacar que a coleta de dados e a delimitação do objeto da pesquisa acadêmica é uma tarefa atribuível exclusivamente ao pesquisador, não podendo o Poder Judiciário dispor de servidores para a atribuição de encargos que não são de sua competência.

Bem a propósito, o disposto no art. 12, incisos II e III, da Resolução CNJ n. 215/2015 dispõe que:

Art. 12. Não serão atendidos pedidos de acesso à informação:



[...]

II - desproporcionais ou desarrazoados;

III - **que exijam trabalhos adicionais de análise, interpretação ou consolidação de dados e informações, serviço de produção ou tratamento de dados que não seja de competência do órgão ou entidade.**

Assim, a busca por dados de feitos baixados exigiria um grupo considerável de servidores para realizar a consulta manual e individualizada a inúmeros processos judiciais, o que comprometeria diretamente o exercício das funções típicas e atípicas do Poder Judiciário, nos termos do referido artigo da Resolução CNJ supracitada, eximindo-se este Tribunal do fornecimento dos dados na forma requerida, sem prejuízo, por óbvio, da possibilidade de obtenção direta das informações pelo requerente via consulta pública do eproc.

Por outro lado, é consabido que o inteiro teor das decisões, sentenças, votos e acórdãos representam dados processuais básicos de livre acesso, nos termos do art. 2º, IV, da [Resolução n. 121, de 5 de outubro de 2010, do CNJ](#).

3. À vista do exposto, defiro o pedido para a pesquisa dos processo ativos, sob supervisão da juíza competente, assim como a dos processos baixados, desde que coletados pelo próprio acadêmico como usuário externo.

Cientifique-se o requerente, com cópia do presente despacho.

Após, conclua-se o processo.

Florianópolis, data da assinatura digital

Iolanda Volkmann  
Juíza Auxiliar da Presidência



Documento assinado eletronicamente por **Iolanda Volkmann, Juíza Auxiliar da Presidência**, em 10/10/2023, às 17:01, conforme art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida no site <http://sei.tjsc.jus.br/verificacao> informando o código verificador **7521644** e o código CRC **99FAC12B**.