UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS

Martin Douglas Brownrigg

APRENDIZADO DE MÁQUINA APLICADO À ANÁLISE DE TENDÊNCIAS NO CONTEXTO DE CAPITAL DE RISCO NO BRASIL

FLORIANÓPOLIS
2022

MARTIN DOUGLAS BROWNRIGG

APRENDIZADO DE MÁQUINA APLICADO À ANÁLISE DE TENDÊNCIAS NO CONTEXTO DE CAPITAL DE RISCO NO BRASIL

Trabalho de Conclusão de Curso submetido à Universidade Federal de Santa Catarina, como requisito necessário para obtenção do grau de Bacharel em Engenharia Elétrica

Florianópolis, dezembro de 2022

UNIVERSIDADE FEDERAL DE SANTA CATARINA

MARTIN DOUGLAS BROWNRIGG

Este Trabalho de Conclusão de Curso foi julgado adequado para a obtenção do Título de Bacharel em Engenharia Elétrica e aceito, em sua forma final, pelo Curso de Graduação em Engenharia Elétrica.

Documento assinado digitalmente
Miguel Moreto
Data: 21/12/2022 12:22:28-0300
CPF: ***.850.100-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Miguel Moreto
Coordenador do Curso de Graduação em Engenharia Elétrica

Banca Examinadora:

Documento assinado digitalmente
Eduardo Luiz Ortiz Batista
Data: 21/12/2022 12:33:29-0300
CPF: ***.521.889-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Eduardo Luiz Ortiz Batista
Universidade Federal de Santa Catarina

Documento assinado digitalmente
Richard Demo Souza
Data: 21/12/2022 11:42:40-0300
CPF: ***.267.379-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Richard Demo Souza
Universidade Federal de Santa Catarina

Documento assinado digitalmente
MILTON BIAGE
Data: 22/12/2022 13:29:47-0300
CPF: ***.070.831-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Milton Biage
Universidade Federal de Santa Catarina

Florianópolis, 16 de dezembro de 2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Brownrigg, Martin Douglas
Aprendizado de máquina aplicado à análise de tendências no contexto de capital de risco no Brasil / Martin Douglas Brownrigg ; orientador, Eduardo Luiz Ortiz Batista, 2022.
51 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia Elétrica, Florianópolis, 2022.

Inclui referências.

1. Engenharia Elétrica. 2. Capital de risco. 3. Aprendizado de máquina. 4. Séries temporais. I. Batista, Eduardo Luiz Ortiz. II. Universidade Federal de Santa Catarina. Graduação em Engenharia Elétrica. III. Título.

Agradecimentos

Agradeço aos meus pais, Franky e Mariana, e à minha irmã, Elisa, pela minha formação e apoio ao longo de toda minha vida acadêmica.

À minha namorada, Bianca, pelo companheirismo ao longo desta e de outras etapas de minha vida acadêmica e pessoal.

Aos amigos e colegas de curso que estiveram comigo nos momentos mais difíceis da graduação.

Aos colegas da Invest Jr., por despertar em mim o interesse sobre finanças e investimentos.

Aos colegas de trabalho, Rodolfo, Eduardo e Vitor, pelo apoio no trabalho realizado e motivação também do interesse no mercado de capital de risco.

Ao professor Eduardo Batista, pelo acompanhamento e excelente orientação ao longo da realização do trabalho.

Resumo

Em plena emergência ao longo dos últimos anos, o mercado brasileiro de capital de risco carece de estudos aprofundados que permitam a análise de tendências desta modalidade de investimentos no país. Ainda que com uma amostra limitada por conta do estágio inicial em que se encontra o mercado do Brasil em comparação aos pioneiros desta prática como os Estados Unidos, fez-se possível propor métodos qualitativos e quantitativos por meio da utilização de modelos de aprendizado de máquina (VAR, Autorregressor Floresta Aleatória e Classificador XGBooster) de maneira a se analisar o panorama geral deste ramo do mercado financeiro e as tendências particulares aos entes atuantes no mercado.

Palavras-chave: capital de risco, aprendizado de máquina, séries temporais, causalidade de Granger, ciência de dados, macroeconomia, mercado brasileiro.

Abstract

In full emergence over the last few years, the Brazilian Venture Capital market lacks in-depth studies and robust analyses of trends regarding this asset class. Despite dataset size limitations due to the current stage of the Brazilian market compared to Venture Capital pioneers such as the United States, qualitative and quantitative methods were successfully proposed through machine learning models (VAR, Random Forest Regressor and XGBooster Classifier) towards an analysis of both the overall scenario and particular trends regarding players involved in the Brazilian Venture Capital market.

Keywords: venture capital, machine learning, time series, Granger causality, data science, macroeconomics, Brazilian market.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

NVCA: *National Venture Capital Association*

PaaS: *Platform as a Service*

ABVCAP: Associação Brasileira de Private Equity e Venture Capital

CVM: Comissão de Valores Mobiliários

VC: *Venture Capital*

IPO: *Initial Public Offering*

VAR: Modelo autorregressivo vetorial

AIC: *Akaike Information Criterion*

BIC: *Bayesian Information Criterion*

MAE: *Mean Absolute Error*

IBGE: Instituto Brasileiro de Geografia e Estatística

INPC: Índice Nacional de Preços ao Consumidor

PIM-PF: Pesquisa Industrial Mensal - Produção Física

IPCA: Índice Nacional de Preços ao Consumidor Amplo

SELIC: Sistema Especial de Liquidação e de Custódia

PIB: Produto Interno Bruto

EMBI: *Emerging Markets Bond Index*

IBOVESPA: Índice da Bolsa de Valores de São Paulo

Sumário

1 Introdução

Capital de Risco, do inglês *Venture Capital*, é uma modalidade de investimentos na qual se aplicam recursos em empresas de estágios iniciais, comumente chamadas de *startups*. Por conta do perfil estratégico de companhias desta categoria, liquidez dos aportes realizados e dispersão das taxas de retorno, estes investimentos podem ser caracterizados como de alto risco e também elevado potencial de rentabilidade (Distrito, 2022).

De maneira semelhante a outras formas de investimento e práticas capitalistas, a modalidade *Venture Capital* teve os Estados Unidos como país pioneiro. Toma-se como referência as décadas de 1960 e 1970 para os primeiros registros de aplicações em capital de risco como uma classe de ativos, acompanhada também da fundação de grandes gestoras, fundos de investimento e da Associação Nacional de Venture Capital - da sigla NVCA, em inglês (Gold, 2022).

Embora os marcos históricos da indústria de capital de risco a nível global sejam datados de mais de 5 décadas atrás, o mercado brasileiro teve emergência recente. Segundo relatório da Distrito, pioneira no quesito PaaS (*Platform as a Service*) para inteligência de dados acerca do mercado de capital de risco no país, os marcos correspondentes aos supracitados em território brasileiro ocorreram em 1994 e 2000, referentes à criação da Associação Brasileira de Private Equity e Venture Capital (ABVCAP) e à regulamentação desta classe de ativos na Comissão de Valores Mobiliários (CVM), respectivamente.

A pouca idade da modalidade no Brasil traz consigo a raridade de estudos aprofundados sobre o cenário de *Venture Capital* no país. A mencionada ABVCAP produz relatórios periódicos acerca do panorama nacional de investimentos em *startups* - estes,

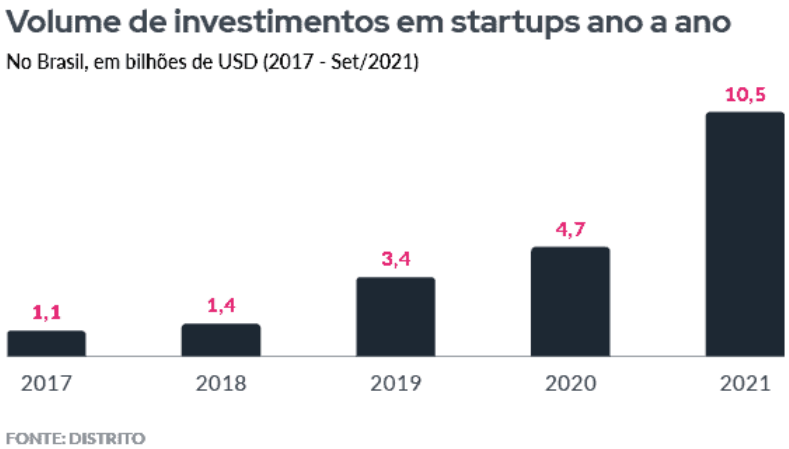

Figura 1 – Histórico recente do mercado brasileiro

Fonte: Distrito [3]

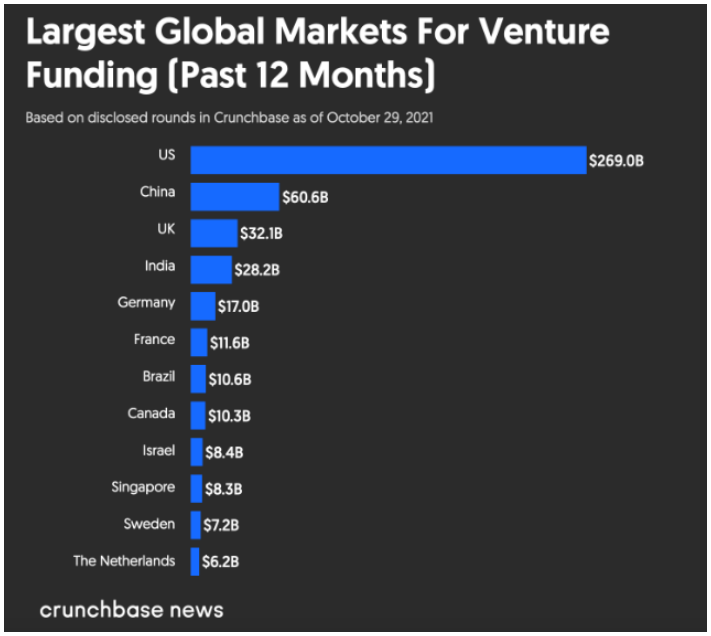

Figura 2 – Principais mercados de capital de risco a nível global (2021)

Fonte: Crunchbase [4]

entretanto, possuem fins informativos e carecem de análises minuciosas sobre tendências passíveis de identificação. Cabe mencionar, nesse contexto, a reduzida riqueza dos dados disponibilizados como fatores limitantes de tais estudos e outros semelhantes, motivada pela sensibilidade envolvida na divulgação de informações privadas de empresas investidas e da estratégia de alocação de recursos dos atuantes do mercado.

Àqueles que analisam a situação econômica do Brasil, surge a importância de compreender as tendências desse mercado, dada a rápida ascensão do volume investido (Figura 1). Da parte dos gestores de recursos e tomadores de decisão, cita-se a relevância do entendimento de como tais tendências podem se correlacionar com a performance de seu portfólio e estratégia de alocação de recursos - cujo impacto mostra-se amplificado pelas características de risco e retorno desta classe de ativos. Ainda neste contexto, uma análise crítica acerca de estratégias e comportamento dos entes envolvidos na indústria de *Venture Capital* no Brasil será discutida posteriormente.

Em suma, apresentam-se como fatores relevantes à oportunidade observada e consequente motivação do trabalho realizado:

- Posicionamento do Brasil como mercado relevante de capital de risco (Figura 2), embora caracterizado como economia emergente;
- Incertezas sob o aspecto macroeconômico no contexto da instabilidade política e crise de COVID-19;
- Escassez de análises aprofundadas e robustas de maneira análoga a estudos existentes

acerca de mercados mais tradicionais como a bolsa de valores.

1.1 Objetivo

1.1.1 Objetivo Geral

Como objetivo geral, estipula-se a aplicação de métodos de aprendizado de máquina e ciência de dados para a identificação de padrões no mercado de investimentos em capital de risco no Brasil.

1.1.2 Objetivos Específicos

A partir do objetivo geral traçado para este trabalho, menciona-se como objetivos específicos:

- Estruturar os modelos VAR e autorregressor tipo floresta aleatória, de maneira a identificar relações de causalidade entre o volume de investimentos em capital de risco no Brasil e os indicadores macroeconômicos do respectivo mês;
- Analisar o impacto da crise de COVID-19 nos investimentos em capital de risco no Brasil, e comparar o comportamento e recuperação desta em relação à bolsa de valores de maneira qualitativa;
- Realizar uma investigação acerca de fatores que indicam sucesso de empresas em meio à dinâmica de mercado mencionada no primeiro item.

2 Fundamentação teórica

2.1 Financiamento de startups

A categorização de empresas em estágios iniciais como *startups* tem evoluído desde a emergência dos exemplos mais famosos, como Google e Apple; no período da popularização da internet, no início dos anos 2000, chegou-se a associar o termo a qualquer negócio com potencial inovador envolvido no ambiente *online*. No contexto atual em que o mercado de investimentos em capital de risco encontra-se em consolidação, todavia, investidores e estudiosos do meio têm convergido para uma nova definição acerca deste modelo de negócios: "uma *startup* é um grupo de pessoas à procura de um modelo de negócios repetível e escalável, trabalhando em condições de extrema incerteza" (SEBRAE, 2014).

As consequências da condição supracitada de extrema incerteza da operação de empresas como estas estão diretamente ligadas ao quesito financeiro. Em meio ao processo da busca por um produto ou serviço que atinja ao público-alvo de forma escalável, é comum entre *startups* o resultado negativo em seus estágios iniciais, comumente conhecido como "queima de caixa". Neste contexto, surge a necessidade de se recorrer a entes capazes de não somente auxiliar com o financiamento para o plano de ação dos fundadores da companhia, mas também participar do planejamento estratégico visando a escalabilidade do modelo de negócios idealizado. Neste modelo de financiamento em que se procura o sócio-investidor capaz de contribuir com visão de negócios, governança, *networking* no mercado, entre outros tipos de apoio, denomina-se o aporte como *smart money*, ou "capital intelectual", em uma tradução da semântica do termo.

Ao contrário de outras modalidades de investimentos como o mercado de ações ou de criptomoedas, as participações societárias adquiridas por meio da dinâmica previamente descrita não costumam ser negociadas de maneira especulativa ou recorrente no mercado de *Venture Capital*. Eventos de liquidez, portanto, posicionam-se como o objetivo final dos sócios-investidores - a saída e valorização do investimento destes entes pode se dar de três formas:

- compra da participação por parte de um fundo de investimentos voltado a empresas de capital fechado em estágios mais avançados - modalidade conhecida como *Private Equity*;

- aquisição total ou do controle da *startup* por parte de um concorrente ou empresa consolidada no meio em que atua;

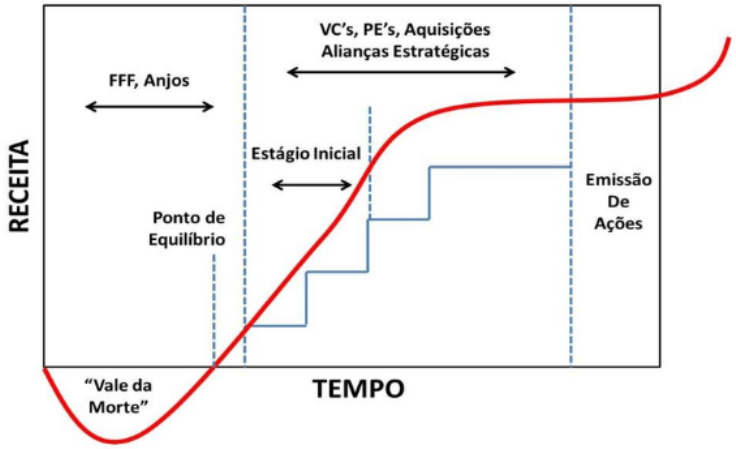

Figura 3 – Ciclo de financiamento de uma *startup*

Fonte: Slap Law [6]

Estágio Investimento	Valor aporte	Equity	Valuation
Aceleradoras	R$ 100.000,00	8%	R$ 800.000,00
Anjos	R$ 180.000,00	12%	R$ 2.160.000,00
Pré-Seed	R$ 300.000,00	10%	R$ 3.000.000,00
Seed	R$ 2.000.000,00	15%	R$ 13.000.000,00
Série A	R$ 8.000.000,00	20%	R$ 40.000.000,00
Série B	R$ 27.000.000,00	30%	R$ 90.000.000,00

Figura 4 – Rodadas de financiamento de uma *startup*

Fonte: Blog João Kepler [7]

- abertura do capital da empresa na bolsa de valores, conhecida como IPO (da sigla em inglês *Initial Public Offering*).

Findando atingir os eventos de liquidez mencionados, o ciclo padrão (Figura 3) percorrido por uma *startup* envolve o processo de repetidas captações de recursos, cujos eventos são conhecidos como rodadas de investimentos. A avaliação (*valuation*), o valor aportado e a diluição para cada tipo de rodada (Figura 4) podem variar de acordo com fatores como o perfil do sócio-investidor, setor de atuação da empresa, momento de mercado e indicadores de sucesso da companhia no momento da captação. Ainda nesta linha de raciocínio, cabe ressaltar que, dada a estratégia comum de alocação agressiva de recursos em busca de escalabilidade para o modelo de negócios, o fato da *startup* passar por subsequentes rodadas de financiamento significa o sucesso da empresa a caminho de um evento de liquidez, que interessa aos investidores de estágios mais iniciais.

2.1.1 Estratégias de investimento

Conforme citado na introdução do presente estudo, investimentos em capital de risco são caracterizados como de alto risco mas também de alto potencial de retorno. Em

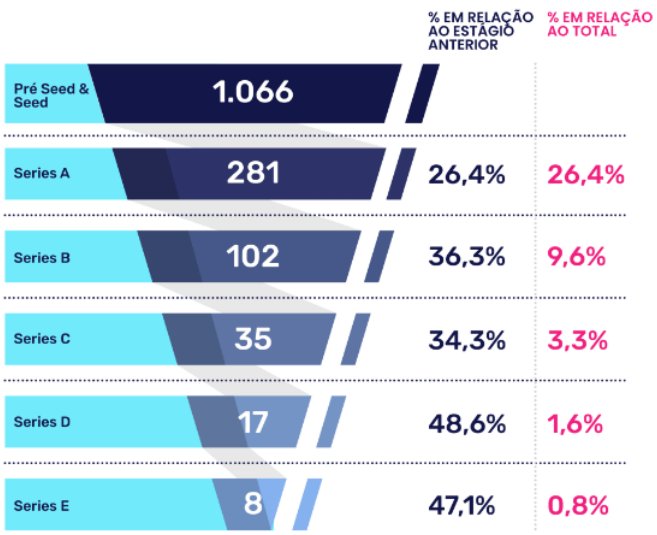

Figura 5 – Taxa de sobrevivência de *startups* brasileiras
Fonte: Distrito [8]

relação à primeira característica, tem-se como evidência o estudo realizado pela plataforma Distrito com sua base de *startups* brasileiras entre 2011 e 2020 (Figura 5) e sua respectiva capacidade de manter ativa sua operação ao longo do percurso padrão de captações de rodada de investimento. Em concordância com o foco principal do presente estudo, nota-se a baixa taxa de sobrevivência das empresas principalmente em estágios iniciais. Por outro lado, evidencia-se na Figura 6 a forte tendência de desempenho superior a longo prazo de fundos de investimento em capital de risco em relação a índices da bolsa de valores.

De maneira a se obter os retornos observados em meio a tal dinâmica deste ramo do mercado financeiro, a estratégia de alocação de recursos de fundos de investimento em capital de risco varia. Uma delas é denominada "*spray and pray*", que consiste em alocar recursos em grande quantidade de empresas com a visão de que a minoria do portfólio construído tende a obter sucesso capaz de compensar os investimentos sem sucesso e resultar em uma taxa de retorno global interessante ao portfólio.

A estratégia alternativa para a alocação de recursos, sem nomenclatura específica, é a de se selecionar empresas para investimento com processo de avaliação mais longo e minucioso. Ao contrário de uma prática comum na estratégia "*spray and pray*", costuma-se definir, para cada investimento feito na composição do portfólio, a participação societária e volume de financiamento ideais.

Embora adotada pela Bossanova, uma das gestoras de recursos mais premiadas em território brasileiro, a estratégia de investimentos em *startups* a um grande volume e com seleção de investimentos menos criteriosa é motivo de divergência entre especialistas do ramo (Manzoni Jr., 2018). Por este motivo, o presente trabalho visa abordar uma

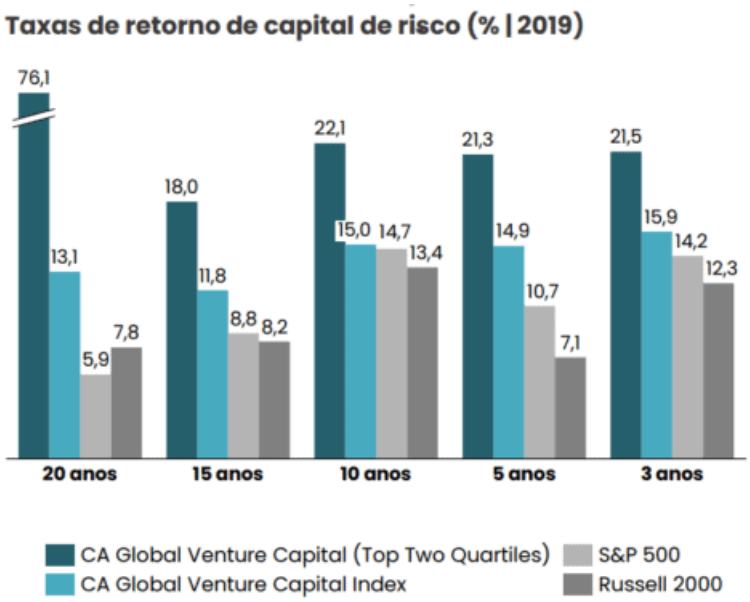

Figura 6 – Retorno de fundos de *Venture Capital* em relação a índices da bolsa de valores estadounidense

Fonte: G2D Investments [9]

análise estatística sobre a correlação entre o volume de empresas investidas e o número de investimentos de sucesso no portfólio de entes atuantes no mercado de capital de risco no Brasil.

2.2 Aprendizado de máquina

Aprendizado de máquina é uma área da ciência de dados em que se utiliza algoritmos baseados em conceitos de matemática, estatística e inteligência artificial de maneira a se realizar tarefas como tomada de decisão, identificação de padrões, previsão e classificação. Conforme sugere o nome da técnica, um dos princípios por trás dos modelos de aprendizado de máquina é o de que estes passem por um processo de aprendizado ao processar grandes quantidades de informação, assim como sua capacidade de adaptação a diferentes tipos de conjuntos de dados fornecidos.

Existem duas subáreas principais em que se dividem os modelos de aprendizado de máquina:

- Classificação: utiliza-se dados de entrada para se definir critérios implícitos de categorização da variável de interesse. A saída é dada pela classificação de uma amostra dentre as limitadas possibilidades fornecidas ao modelo;
- Regressão: utiliza-se dados de entrada para se prever o valor numérico da variável de interesse. A saída é dada por um número, sem que haja limitações impostas para tal

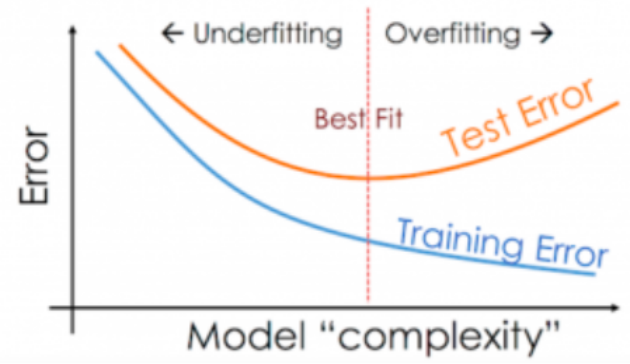

Figura 7 – Ajuste em modelos de aprendizado de máquina

Fonte: Data Analytics [11]

previsão.

Em ambas as subáreas, os algoritmos utilizados contam com parâmetros e hiperparâmetros em sua estrutura. Conquanto ambos sejam influentes na performance dos modelos, sua distinção se dá pelo processo de ajuste: enquanto os parâmetros são ajustados de forma autônoma ao se treinar o modelo, hiperparâmetros são variáveis ajustáveis previamente ao treinamento. Uma vez efetuado o processo de treinamento, faz-se possível a otimização de hiperparâmetros tendo em vista a potencialização do desempenho do modelo - a ferramenta *Grid Search*, que realiza de maneira iterativa combinações de hiperparâmetros fornecidos para posterior avaliação e armazenamento destes como ideais para o modelo treinado, é comumente utilizada para este fim.

Além da performance dos algoritmos de aprendizado de máquina, outro aspecto vital a ser ajustado por meio da otimização de hiperparâmetros é o ajuste. Diz-se que um modelo é sub-ajustado (fenômeno de *underfitting*) quando o desempenho é insatisfatório após o treinamento, sendo o algoritmo incapaz de determinar padrões e estabelecer critérios para regressão ou classificação a partir dos dados fornecidos. Por outro lado, um modelo é dito sobreajustado (fenômeno de *overfitting*) quando a performance é satisfatória para o conjunto de treinamento, mas insuficiente no momento do teste; o fenômeno de *overfitting* ocorre, portanto, quando se observa a incapacidade de generalização, causada principalmente pela exacerbada complexidade dada ao algoritmo ou reduzida quantidade de amostras no conjunto de treinamento. Tal relação entre complexidade do modelo e performance para os conjuntos de treino e teste é representada na Figura 7.

2.2.1 Autorregressor tipo floresta aleatória

Em modelos de aprendizado de máquina em que se visa à previsão de valores para a variável de interesse a partir de valores passados desta, utiliza-se o conceito de autorregressão. Para problemas deste tipo, pode-se utilizar a classe chamada *ForecasterAutoreg*,

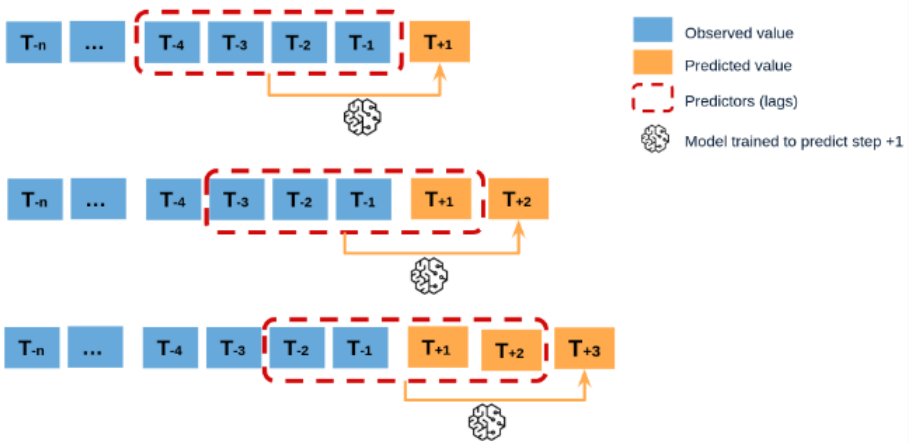

Figura 8 – Modelo de previsão baseado no conceito de autorregressão recursiva

Fonte: Ciencia de Datos [12]

proveniente da biblioteca *Skforecast* da linguagem *Python*, de forma com que se realize uma autorregressão chamada recursiva, em que cada nova previsão se baseia em outras também geradas pelo mesmo modelo (Figura 8).

Dentro da classe supracitada, faz-se ainda necessário eleger um modelo específico de regressão para endereçar o problema proposto. Neste contexto, um dos modelos mais amplamente utilizados é o regressor do tipo Floresta Aleatória (do inglês *Random Forest*), em que se "cria várias árvores de decisão e as combina para obter uma predição com maior acurácia e mais estável" (Silva, 2018). Como vantagens deste tipo de modelo de aprendizado de máquina, cita-se como relevantes: a mitigação do sobreajuste, por conta da criação de subconjuntos aleatórios e consequente construção e combinação de árvores de decisão reduzidas (Figura 9), e a possibilidade de se trabalhar com variáveis exógenas, em que se conhece valores futuros das variáveis de entrada.

Dentre os hiperparâmetros chave a serem otimizados em modelos de autorregressão usando florestas aleatórias, estão incluídos:

- *n_estimators*, que se refere ao número de árvores de decisão combinadas no treinamento do modelo;
- *max_depth*, que se refere à profundidade máxima das árvores incluídas na floresta.

2.2.2 Modelo autorregressivo vetorial (VAR)

Outro modelo útil para se trabalhar com a previsão de séries temporais é o autorregressivo vetorial. Segundo Melo (2020), "Modelo autorregressivo vetorial (VAR) é um algoritmo de previsão usado quando duas ou mais séries influenciam-se mutuamente. Na prática, é um modelo de regressão que trata todas as variáveis como endógenas e permite que cada uma delas dependam de p valores de *lags* passados, esses valores são da

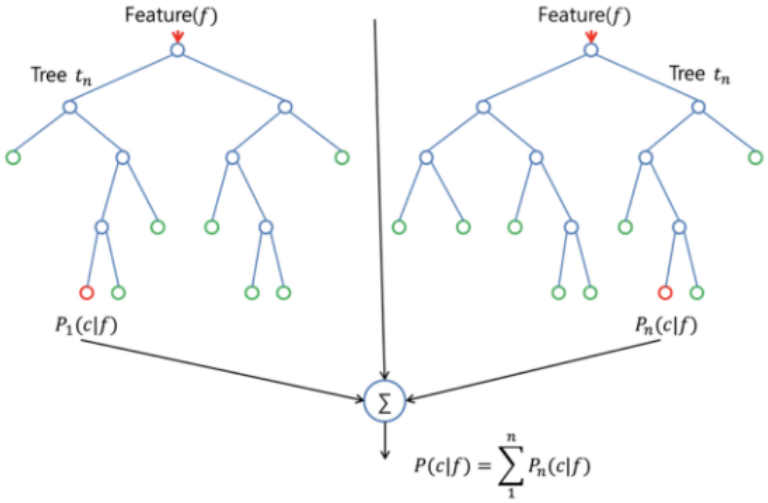

Figura 9 – Modelo de aprendizado de máquina do tipo floresta aleatória

Fonte: Machina Sapiens [13]

própria série e das outras séries". Além do requisito da influência mútua entre múltiplas séries, estas devem também ser estacionárias para se trabalhar com este modelo - o conceito de estacionariedade será abordado adiante.

Do ponto de vista matemático, o modelo em questão pode ser representado pela seguinte equação:

$$y_t = a_0 + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + u_t \tag{2.1}$$

Sendo:

- p a ordem do modelo, referente ao número de *lags* passados;
- y_t um vetor Nx1, com N variáveis endógenas;
- a_0 um vetor Nx1 de constantes;
- *A1, A2, ... Ap* matrizes pNxN de coeficientes autorregressivos;
- u_t um vetor Nx1 que representa um ruído branco.

Do ponto de vista da avaliação do modelo autorregressivo vetorial, há dois critérios amplamente utilizados: *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC). As expressões matemáticas para estes critérios são as que seguem:

$$AIC = 2k - 2\ln(L) \tag{2.2}$$

$$BIC = k \ln(n) - 2 \ln(L) \tag{2.3}$$

Sendo:

- k o número de parâmetros do modelo;
- n o número de observações no conjunto de dados;
- L o valor máximo da função de verossimilhança.

Acerca da seleção do modelo ideal, escolhe-se aquele com o menor valor para ambos os casos (AIC e BIC). Fazendo-se uma análise crítica das expressões matemáticas acima, é possível observar que o critério AIC penaliza modelos mais complexos, ao passo em que dá preferência a modelos com maior desempenho no conjunto de treinamento. O critério BIC, que compartilha das características citadas, tende ainda a valorizar modelos treinados com menor número de observações, resultando em maior penalização da complexidade dos modelos analisados.

Para fins de avaliação do VAR, uma das métricas amplamente utilizadas é o erro médio absoluto (da sigla em inglês MAE - *Mean Absolute Error*), cuja expressão matemática encontra-se descrita abaixo. Esta métrica também é aplicável para avaliação do modelo descrito na seção anterior e demais modelos de regressão de maneira geral.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |yi - \overline{y}| \tag{2.4}$$

Sendo:

- N o número de observações (amostras);
- yi o valor predito de y;
- $\overline{y}$ o valor médio de y.

2.2.3 Classificador XGBooster

No que tange a modelos de classificação, o classificador do tipo *XGBooster* destaca-se por seu desempenho e flexibilidade devida ao elevado número de hiperparâmetros passíveis de otimização (Melo, 2019). Tendo o nome proveniente de *Extreme Gradient Boosting*, esta categoria de algoritmos combina conceitos de árvores de decisão e aumento de gradiente de maneira a otimizar o modelo treinado (Figura 10).

Observa-se que o conceito matemático por trás deste tipo de classificador vai ao encontro da floresta aleatória mencionada acima; ainda segundo Melo, “o princípio do

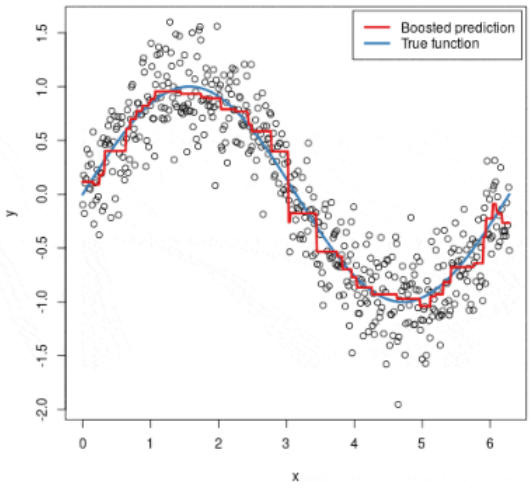

Figura 10 – Exemplificação de modelos do tipo XGBoost

Fonte: Sigmoidal AI [15]

Gradient Boosting é a capacidade de combinar resultados de muitos classificadores 'fracos', tipicamente árvores de decisão, que se combinam para formar algo parecido com um comitê forte de decisão".

A avaliação da performance geral de modelos que endereçam problemas de classificação mostra-se menos complexa do aspecto matemático. A precisão de um modelo, que determina o percentual de assertividade dentre as classificações positivas, pode ser calculada como:

$$Prec. = \frac{VP}{(VP + FP)} \tag{2.5}$$

Sendo:

- VP o número de observações classificadas corretamente como "positivas";
- FP o número de observações classificadas incorretamente como "positivas".

2.3 Séries temporais

No que tange à resolução de problemas de econometria, as práticas modernas amplamente difundidas usam como princípio a modelagem e manipulação de séries temporais (Maddala e Lahiri, 2009). Séries temporais (Figura 11) são definidas como conjuntos de dados indexados da data de cada observação. Seja uma amostra de tamanho "T", $\{y_t\}_{t=1}^{T} = \{y_1, y_2, \ldots, y_T\}$, a série temporal $\{y_t\}_{t=1}^{T}$ é dada pela descrição de seu t-ésimo elemento.

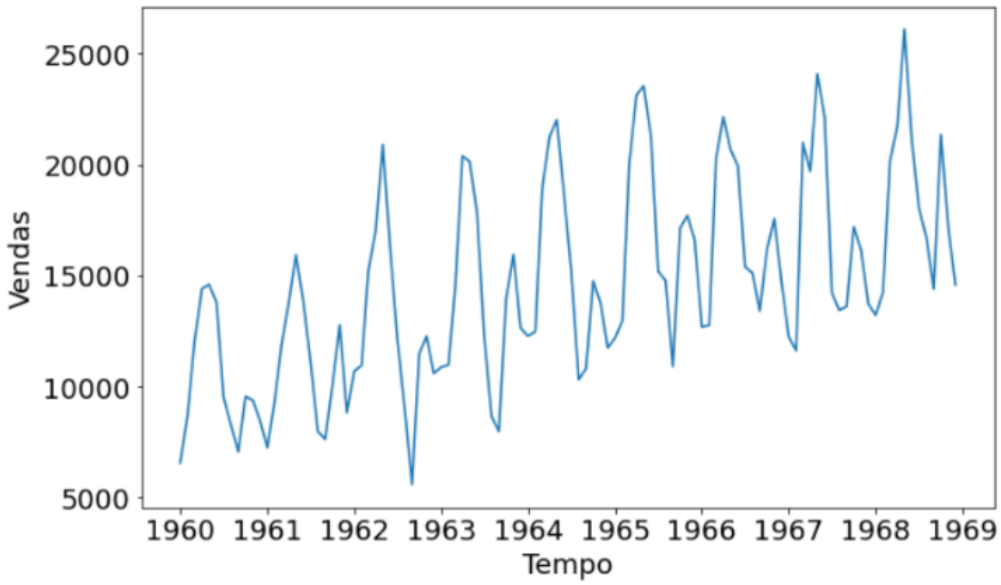

Figura 11 – Exemplificação de série temporal

Fonte: Alura [17]

Conforme a tendência recente de estudos no ramo da econometria apontam, séries temporais e sua interpretação podem ser de grande utilidade para a identificação de padrões, interdependência entre dados adjacentes e previsão de comportamentos futuros.

2.3.1 Séries estacionárias

Como requisitos para se trabalhar com séries temporais em determinados modelos de aprendizado de máquina, cita-se a estacionariedade da série manipulada. Segundo Maddala e Lahiri (2009), "uma série temporal é estacionária quando suas características estatísticas (média, variância, autocorrelação, ...) são constantes ao longo do tempo. É uma série que se desenvolve aleatoriamente no tempo, em torno de uma média constante, refletindo alguma forma de equilíbrio estatístico estável".

Em relação à observação da sazonalidade ou tendência de uma série temporal e sua consequente estacionariedade, esta costuma ser perceptível de maneira visual (Figura 12), embora existam ferramentas - a serem discutidas posteriormente - capazes de testar esta característica de séries temporais.

2.3.2 Séries diferenciais

Séries temporais não estacionárias podem ser tratadas de diferentes maneiras para que se tornem estacionárias. Uma das técnicas utilizadas para este fim é a diferenciação; seja a série temporal $\{y_t\}_{t=1}^{T}$ descrita acima, a correspondente série diferencial se dá por:

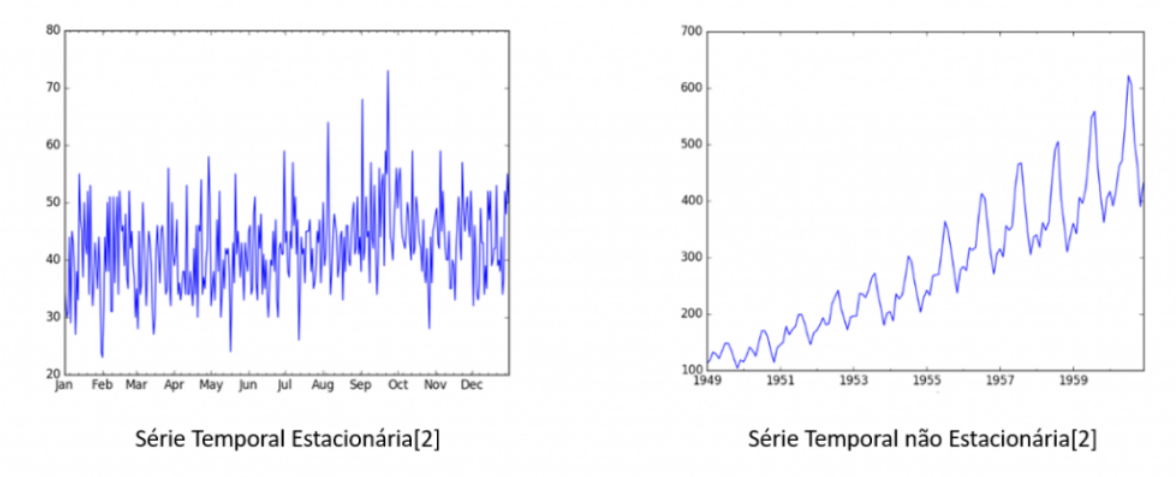

Figura 12 – Estacionariedade de séries temporais

Fonte: BI4All [18]

$$y'_t = y_t - y_{t-1}, \forall t \in [2, T] \tag{2.6}$$

Sendo esta a série diferencial de primeira ordem. Nota-se que o intervalo de amostras passa a ser $[2, T]$ pelo fato de não haver a possibilidade de se calcular o y_1 sem que haja um termo y_0. Cabe ressaltar, ainda, que uma série temporal pode ser diferenciada N vezes até que se obtenha uma série estacionária passível de utilização em modelos de aprendizado de máquina.

3 Pré-processamento

3.1 Base de dados

A base de dados com a qual se trabalhou no presente estudo é proveniente da plataforma Crunchbase, uma das referências mundiais no que tange à inteligência de dados referentes ao mercado de capital de risco (*Venture Capital*). Cada amostra do conjunto total de 4655 representa uma rodada de financiamento de *startups* sediadas no Brasil, com informações como:

- tamanho da rodada (recursos levantados);
- número de rodadas pela qual a empresa já passou até o presente momento (retirada dos dados em março de 2022);
- investidor líder (maior parte dos recursos aportados);
- indústria à qual pertence a *startup*;
- data do evento de captação.

Tendo-se traçado o objetivo de analisar e prever o volume mensal de investimentos em *Venture Capital* no Brasil, as rodadas foram agrupadas conforme seu respectivo mês de anúncio. Ao se deparar com a base de dados importada para o arquivo *Python* em que se trabalhou, foi notada a presença de amostras não referentes a rodadas de captação de recursos em linha com o ciclo de financiamento de *startups* citado anteriormente. Por este motivo, fez-se uma filtragem de maneira a se excluir do estudo aquelas rodadas referentes a eventos de liquidez, bem como duplicidades.

Como dados de entrada para os modelos a serem aplicados, foram trazidos os principais indicadores macroeconômicos fornecidos pelo portal do Instituto Brasileiro de Geografia e Estatística - IBGE, além de outras variáveis que refletem a dinâmica do mercado financeiro como um todo (Tabela 1). Cabe pontuar que as variáveis obtidas também foram discretizadas mensalmente.

Importadas as séries temporais referentes aos indicadores mencionados, obteve-se um novo conjunto de dados com o histórico do volume mensal de investimentos em capital de risco no Brasil (variável de interesse) e os respectivos determinantes a serem testados. Dada a emergência recente da modalidade de investimentos no país e as lacunas observadas em determinados períodos, filtrou-se a base de dados de forma a se trabalhar com uma janela de aproximadamente 10 anos, entre janeiro de 2012 e fevereiro de 2022. O tamanho

Tabela 1 – Dados de entrada para previsão do volume mensal de investimentos

Variável	**Fonte**	**Observação**
Taxa de desemprego	IBGE	Desocupação mensal no Brasil (%)
INPC	IBGE	Índice Nacional de Preços ao Consumidor - Variação (%)
PIM-PF	IBGE	Pesquisa Industrial Mensal - Produção Física - Variação (%)
IPCA	IBGE	Índice geral de inflação - Variação (%)
SELIC	ADVFN	Histórico da taxa mensal básica de juros (%)
PIB	Ipeadata	Interpolação a partir dos dados do IBGE (R$)
EMBI+	Ipeadata	Risco-Brasil referente ao primeiro dia do mês (Pontos)
IBOVESPA	BR Investing	Fechamento do índice no primeiro dia do mês (Pontos)
Dólar	BR Investing	Cotação da moeda no primeiro dia do mês (Valor real)

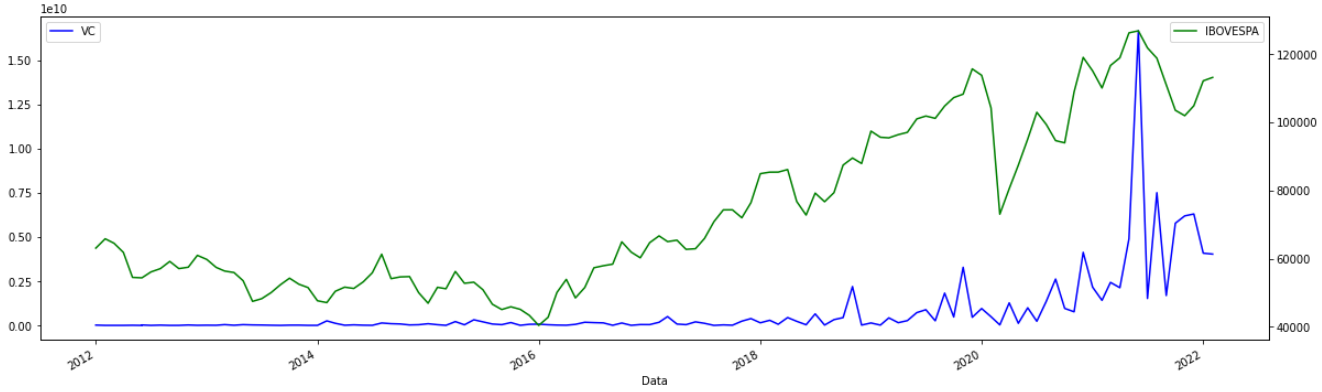

Figura 13 – IBOVESPA e volume mensal de investimentos em capital de risco no Brasil, 2012-2022

da amostra para se atingir o primeiro objetivo específico citado no capítulo 1, portanto, consolidou-se em 122 observações.

3.2 Pré-processamento

3.2.1 Investigações iniciais

Para fins de investigação inicial de possíveis relações de causalidade entre a evolução dos indicadores e a do volume de investimentos em capital de risco, foram elaborados gráficos comparativos entre as curvas de determinados indicadores e a variável de interesse.

As comparações traçadas (Figuras 13 e 14) evidenciam a aparente correlação existente entre algumas das séries temporais, conforme o seguinte racional: a volatilidade do índice da bolsa de valores se assemelha ao perfil de alto risco dos investimentos em *Venture Capital*, enquanto outras modalidades de investimento tendem a ser favorecidas

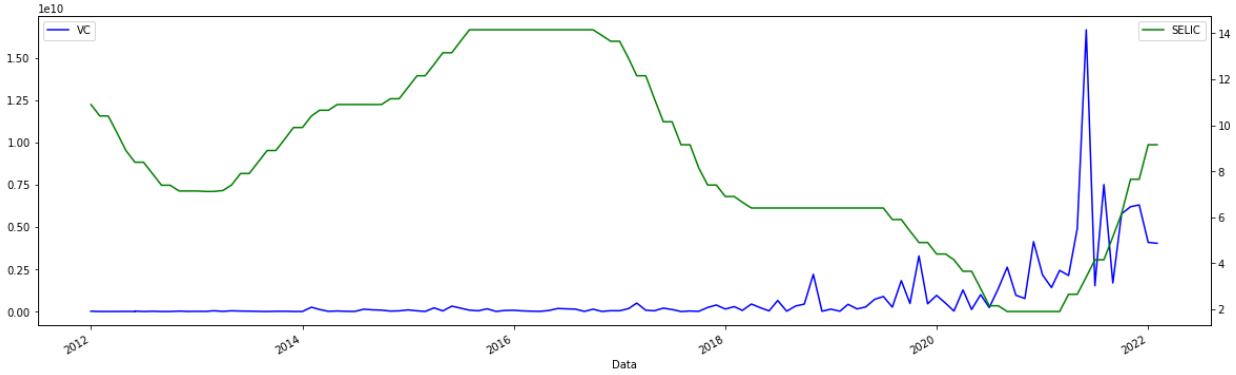

Figura 14 – Taxa SELIC e volume mensal de investimentos em capital de risco no Brasil, 2012-2022

em períodos de alta das taxas de juros e inflação.

3.2.2 Estacionariedade e causalidade

Identificadas as aparentes relações de causalidade na subseção anterior, buscou-se embasamento estatístico para a verificação de tais relações. Devido ao fato de a comprovação empírica da real causalidade entre os indicadores não pertencer ao escopo do presente trabalho, sobretudo em um setor complexo como a macroeconomia de um país das dimensões do Brasil, utilizou-se do conceito de causalidade de Granger.

Proposto em 1969 pelo matemático Clive Granger, o teste de causalidade de Granger é utilizado para determinar a relação de causalidade entre duas séries temporais. Ao se escolher duas variáveis para teste, a causalidade de Granger permite inferir a utilidade de se utilizar os valores passados da primeira para se prever o comportamento futuro da série temporal referente à segunda variável. Em linha com o que foi descrito no parágrafo anterior, o teste de Granger atesta estritamente acerca da causalidade temporal de duas séries, sendo a comprovação da real relação de causa e efeito entre variáveis inviável através desta prática.

De maneira análoga ao requisito encontrado nos supracitados modelos de aprendizado de máquina, o teste de causalidade de Granger requer a estacionariedade das séries cuja causalidade venha a ser testada. Findando averiguar tal característica das séries temporais referentes aos indicadores macroeconômicos e ao volume total de investimentos em capital de risco, utilizou-se do teste *Augmented Dickey-Fuller*. Comumente adotado em se tratando de séries temporais complexas para abordagem de problemas de econometria, este teste tem como princípio a verificação da existência de uma raiz unitária, a qual indica a tendência de um processo estocástico e a consequente não-estacionariedade de uma série temporal.

Verificada a não estacionariedade de algumas das séries temporais ao se aplicar o

	lags=1	lags=2	lags=3	lags=4	lags=5	lags=6	lags=7	lags=8	lags=9	lags=10	lags=11	lags=12
Taxa de desemprego (%)	0.021100	0.027100	0.034300	0.060000	0.073600	0.042400	0.099900	0.152600	0.076500	0.108700	0.058600	0.042700
IPCA	0.034500	0.231000	0.148900	0.251700	0.363200	0.033600	0.019100	0.031300	0.105700	0.041700	0.033500	0.044900
SELIC	0.002000	0.032700	0.005300	0.005000	0.060200	0.060300	0.029900	0.050400	0.037400	0.102600	0.161200	0.054800
PIB	0.000000	0.001300	0.004100	0.000800	0.002400	0.001800	0.001900	0.003700	0.013700	0.003600	0.011800	0.002900
EMBI + Risco-Brasil	0.389700	0.660700	0.719300	0.897400	0.937400	0.891900	0.694800	0.729600	0.454700	0.556300	0.496400	0.310200
IBOVESPA	0.000000	0.003300	0.013000	0.054500	0.035700	0.051700	0.078800	0.056600	0.018000	0.005000	0.007800	0.002100
Dólar	0.000000	0.002800	0.008100	0.024500	0.044800	0.075700	0.083400	0.065100	0.305800	0.366000	0.395400	0.184600
INPC	0.025100	0.163500	0.161000	0.315100	0.398700	0.025500	0.016100	0.031700	0.128000	0.053800	0.051000	0.075100
PIM-PF	0.093900	0.381400	0.603500	0.791900	0.931200	0.922700	0.930300	0.907200	0.865400	0.624200	0.498100	0.014300

Figura 15 – Resultado do teste de Granger entre os indicadores macroeconômicos e o volume de investimentos em capital de risco

método de Dickey-Fuller, utilizou-se do processo de diferenciação e repetição do teste até que se verificasse a estacionariedade de todas as séries temporais referentes aos indicadores macroeconômicos e à variável de interesse.

Uma vez dadas as séries manipuladas como estacionárias, fez-se possível a aplicação do teste de causalidade de Granger. Por meio de um processo iterativo ao longo do número de máximo de *lags* a serem testados (12), foi calculado o valor-p para determinar a causalidade de Granger de cada indicador sobre a variável de interesse. Definido como "a probabilidade de se observar um valor da estatística de teste maior ou igual ao encontrado" (Ferreira e Patino, 2015), o valor-p testou as seguintes hipóteses em cada instância:

- H0: hipótese nula - com o número de *lags* testado, pode-se inferir a causalidade de Granger entre as séries temporais;
- H1: hipótese alternativa - com o número de *lags* testado, há probabilidade significante de se obter uma melhor descrição da causalidade temporal entre as duas séries, que não possuem, portanto, causalidade de Granger entre elas.

Definido o valor de corte para o teste das hipóteses com o valor-p como sendo 0.05, observou-se a ausência de causalidade de Granger para algumas das séries temporais em relação ao volume mensal de investimentos em capital de risco (Figura 15). De maneira a potencializar o desempenho dos algoritmos de aprendizado de máquina, portanto, foram retiradas do conjunto de dados tais variáveis insignificantes do ponto de vista da previsão da variável de interesse.

Julgando-se o conjunto de dados como propriamente manipulado e preparado para a aplicação dos modelos de previsão, foi realizada a separação dos conjuntos de treino e de teste. Devido ao fato de se tratar de uma base amostral pequena, à emergência recente do volume total de investimentos e às técnicas visadas para a redução do sobreajuste, foi definido o tamanho do conjunto de teste como de 5% sobre o total de observações, resultando em 6 amostras.

4 Processamento e Resultados

4.1 Estudo de mercado

Nesta seção, serão discutidos os procedimentos e resultados obtidos acerca das tendências do mercado brasileiro de capital de risco como um todo.

4.1.1 Volume mensal de investimentos

Para fins de previsão do volume mensal de recursos aportados em *startups* brasileiras no período de 2012 a 2022, foram utilizados dois diferentes modelos de aprendizado de máquina, conforme descrito a seguir.

4.1.1.1 Modelo Autorregressivo Vetorial (VAR)

A partir dos resultados obtidos por meio do teste de causalidade de Granger (Figura 15), observou-se a tendência da não causalidade temporal entre a maioria das séries para os *lags* de 7 a 12. Sendo assim, para a aplicação do algoritmo VAR, foram testados os modelos de ordem 1 a 6 (Figura 16).

Conforme discutido no capítulo 2, observa-se a priorização de modelos mais simplificados quando se trata da métrica BIC para avaliação de modelos autorregressivos vetoriais, enquanto aqueles julgados ideais por meio da métrica AIC tendem a ser mais complexos. Tendo em vista a também mencionada reduzida amostra com que se trabalha e a redução do sobrajuste, optou-se por se trabalhar com o modelo de ordem 2 (2 *lags*).

Para o modelo do tipo VAR escolhido, obteve-se um erro médio absoluto (VAR) de R$ 13.002.840.770,00 para a série diferenciada. Fazendo-se uma análise crítica da curva

VAR Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	85.55	85.75	1.430e+37	85.63
1	79.82	81.63	4.666e+34	80.56
2	74.80	78.21*	3.118e+32	76.18
3	73.27	78.30	7.123e+31	75.31
4	71.92	78.55	2.030e+31	74.61
5	70.72	78.96	7.240e+30	74.06
6	68.87*	78.72	1.496e+30*	72.87*

Figura 16 – Seleção do modelo ideal do tipo autorregressivo vetorial

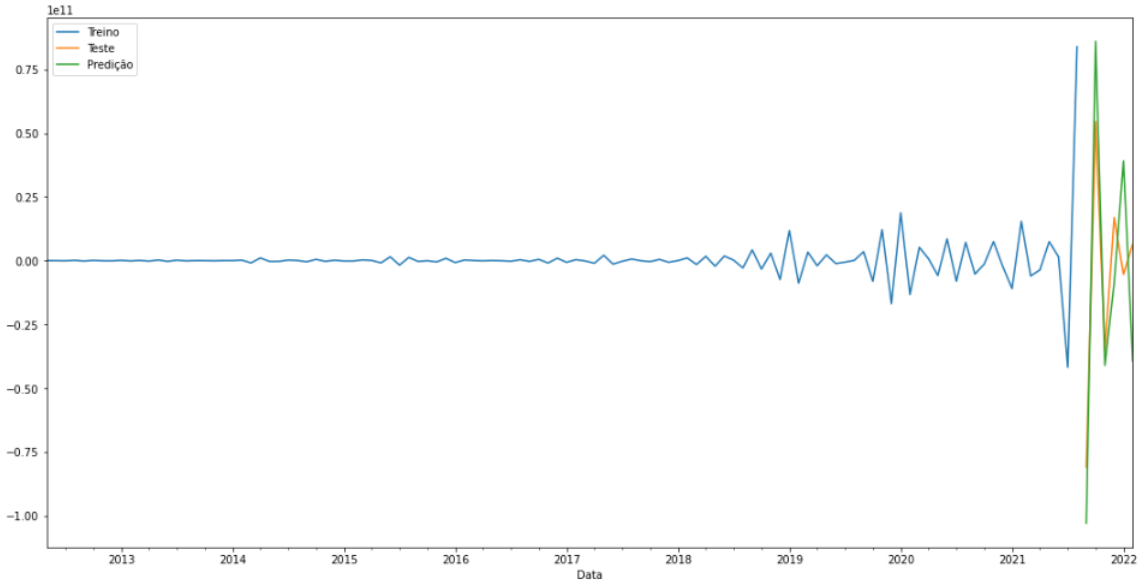

Figura 17 – Resultado da previsão do volume mensal de investimentos por meio do modelo VAR

dos resultados obtidos (Figura 17), é possível observar a previsão de fortes oscilações no volume mensal de investimentos, devida à característica do modelo referente ao tratamento de todas as variáveis como endógenas. Sendo assim, as fortes oscilações observadas ao final do período de treino tiveram forte influência na característica das previsões, sendo o comportamento de altas variações no volume de investimentos tomado como tendência para o futuro.

4.1.1.2 Autorregressor tipo floresta aleatória

Além do modelo VAR, foi ainda testado o autorregressor do tipo floresta aleatória para a previsão do volume mensal de investimentos do mercado de *Venture Capital* brasileiro. Conforme supracitado, este modelo permite a utilização de variáveis exógenas: os indicadores macroeconômicos para o período do conjunto de teste, portanto, são conhecidos ao se trabalhar com esta técnica.

Com um erro médio absoluto (MAE) de R$ 28.463.702.535,00, nota-se que o modelo discutido na presente subseção previu oscilações menores para o volume de aportes, ao contrário do modelo autorregressivo vetorial (Figura 18). Ao se observar as reduzidas variações proporcionais dos indicadores macroeconômicos em relação à variável de interesse no período de teste, observa-se que o modelo não tomou como padrão as fortes oscilações observadas do volume de investimentos ao final do período de teste.

Visando à otimização dos resultados obtidos, utilizou-se a ferramenta *Grid Search* para a otimização dos hiperparâmetros *n estimators* e *max_ depth*. De maneira análoga ao objetivo da redução do sobreajuste descrito anteriormente, foi também priorizado o teste

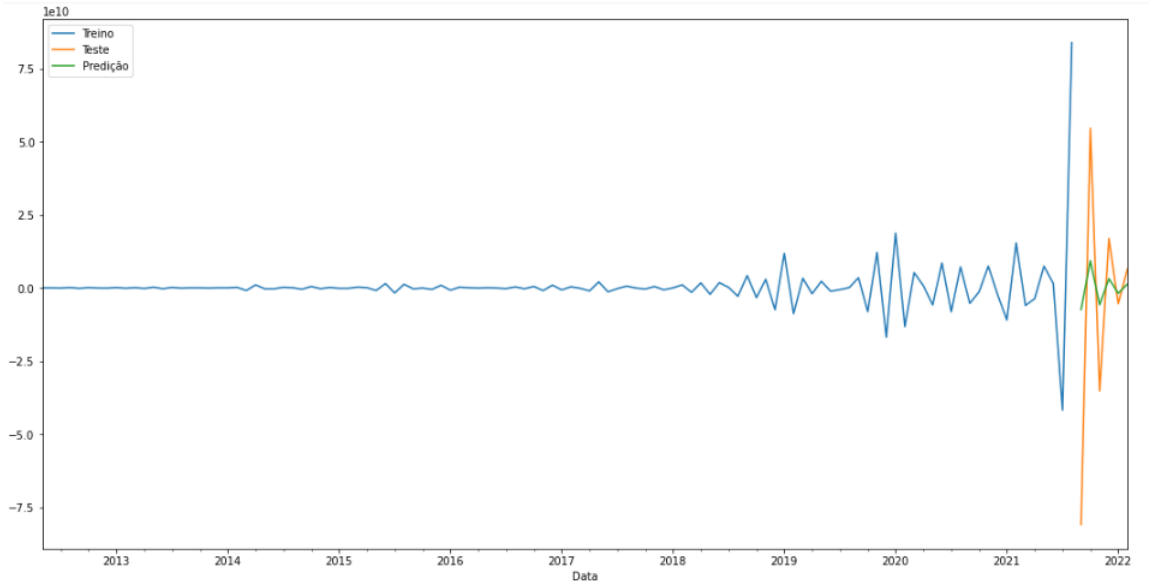

Figura 18 – Resultado da previsão do volume mensal de investimentos por meio do modelo autorregressor tipo floresta aleatória

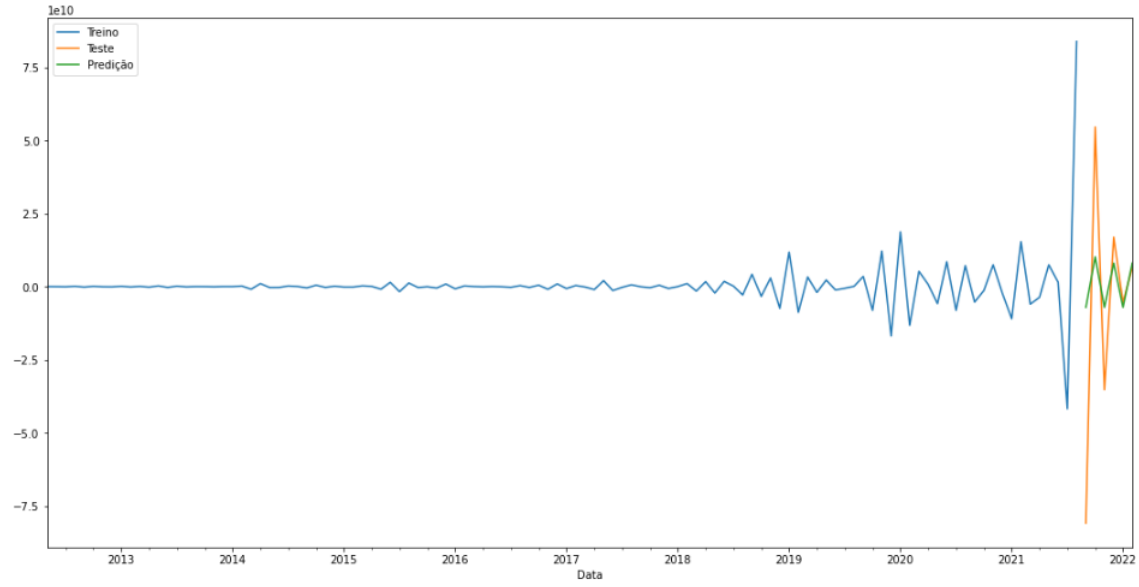

Figura 19 – Resultado da previsão do volume mensal de investimentos por meio do modelo autorregressor tipo floresta aleatória após otimização de hiperparâmetros

de modelos mais simplificados para o autorregressor do tipo floresta aleatória.

Após a otimização dos hiperparâmetros, obteve-se um erro médio absoluto de R$ 26.432.775.139,00, sendo este ainda superior ao erro observado no modelo autorregressivo vetorial. Ao contrário do modelo utilizado anteriormente, entretanto, faz-se possível notar a precisão deste em prever a oscilação positiva ou negativa do volume mensal de investimentos em capital de risco (Figura 19), ainda que com erro absoluto considerável para algumas das observações.

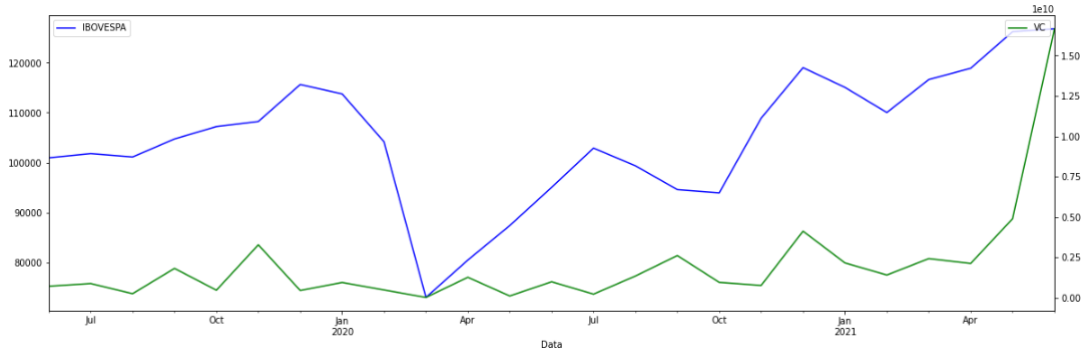

Figura 20 – Comparação do IBOVESPA e investimentos em capital de risco ao longo da pandemia de COVID-19

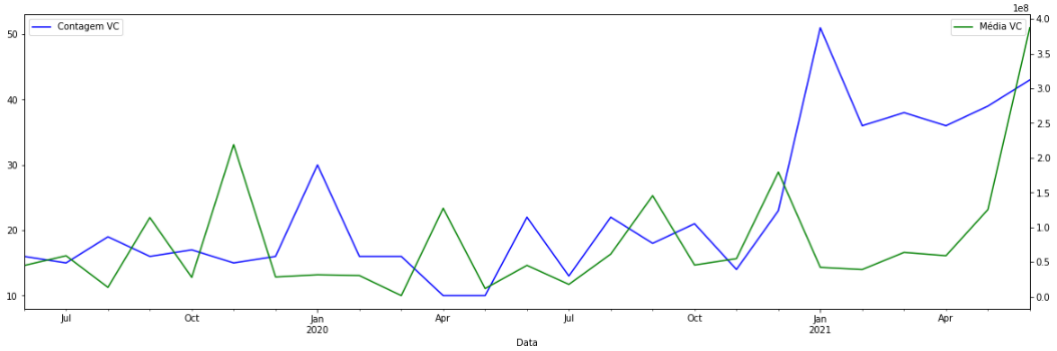

Figura 21 – Contagem e média dos aportes em *startups* no período da pandemia de COVID-19

4.1.2 Pandemia COVID-19

Sabendo-se dos efeitos globais da crise econômica causada pelo coronavírus, filtrou-se o conjunto de dados para observar o comportamento dos investimentos em capital de risco, bem como dos indicadores macroeconômicos no período entre junho de 2019 e junho de 2021.

De forma similar ao principal índice da bolsa de valores brasileira, o volume de investimentos em *Venture Capital* no país sofreu queda considerável a partir de janeiro de 2020 (Figura 20). Nota-se, porém, a rápida recuperação de ambas as modalidades de investimento nos meses subsequentes, além das altas históricas atingidas pelos aportes em *startups* em 2021.

Ainda neste contexto, outra observação a ser feita no período analisado é relacionada à mudança no perfil dos investimentos realizados. Conforme descrito nas investigações iniciais discutidas na seção 3.2.1, as baixas nas taxas de juros tendem a tornar os investimentos em renda fixa menos atrativos, ao passo em que modalidades de investimento mais voláteis e de mais alto risco tendem a prevalecer. A comparação do número e aporte médio das rodadas de investimento no período da pandemia (Figura 21) comprovam a elevada suscetibilidade a risco por parte dos investidores, com a tendência do maior número de

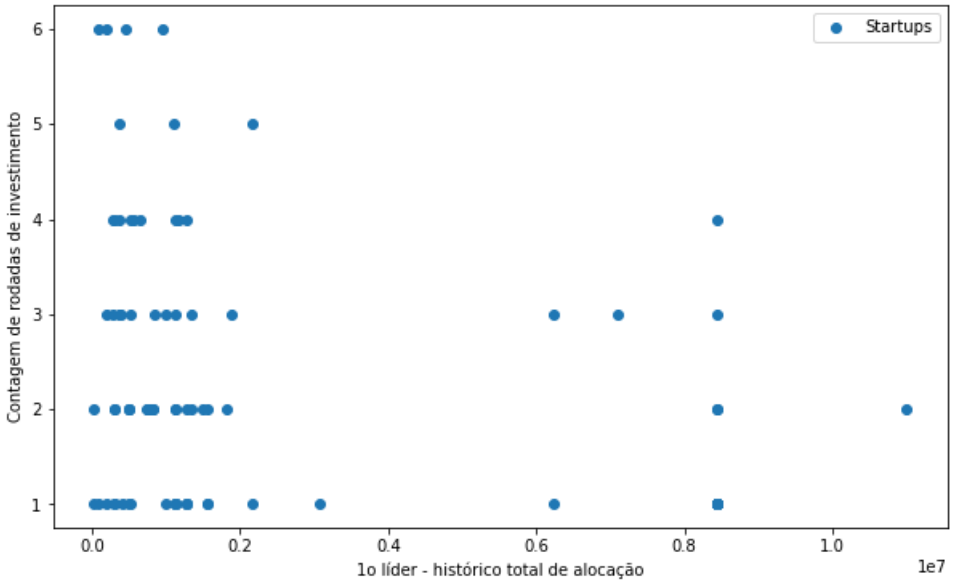

Figura 22 – Número de rodadas de investimento de *startups* brasileiras de acordo com o histórico de alocação do líder da primeira captação

aporte em empresas de estágios iniciais, com grau de incerteza elevado.

4.2 Tendências particulares

Com motivação proveniente das observações mencionadas na seção anterior, as análises a serem descritas a seguir referem-se às tendências no comportamento dos tomadores de decisão do mercado de capital de risco, bem como das *startups* capazes de captar recursos nesta modalidade.

4.2.1 Estratégias de investimento

Dado o elevado volume aportado em empresas no início do ciclo de financiamento ao longo do período da pandemia, surge o questionamento acerca da estratégia de alocação de recursos dos gestores de investimentos em capital de risco. Com o objetivo de analisar o desempenho do portfólio e das teses de investimento destes entes, portanto, filtrou-se a base para que se trabalhasse apenas com os primeiros eventos de captação de cada empresa, sendo estas rodadas do tipo "Pre-Seed" e "Seed", em sua ampla maioria.

Dentre as 127 empresas da base filtrada, nota-se visualmente (Figura 22) a ausência de correlação do histórico total de alocação em empresas *early-stage* do líder da primeira rodada de financiamento com o número de rodadas subsequentes. Embora não se possa atrelar o histórico de alocação do primeiro sócio-investidor ao sucesso da *startup*, entende-se a relevância de outras características da parceria para a potencialização do modelo de negócios, conforme o conceito de *smart money* citado no capítulo 2.

Além dos fatores intangíveis citados acima, entende-se que outro motivo influente nesta ausência de correlação é a tese de investimentos de cada um dos sócios-investidores analisados: em contraste com a estratégia *spray and pray*, há a possibilidade de se observar a maximização do sucesso das empresas após captarem recursos com investidores cuja estratégia se baseia em maior minuciosidade na seleção dos investimentos, com consequente menor volume histórico de alocação.

Para teste da hipótese mencionada no parágrafo anterior, definiu-se o sucesso do investimento como a capacidade de a *startup* passar por nova rodada de financiamento após o primeiro evento de captação em que participaram os investidores líderes analisados. Prática comum em estudos semelhantes acerca do mercado de *Venture Capital* (ARROYO, Javier, et al., 2019), este procedimento permitiu o contraste do número de investimentos de sucesso com a contagem de aportes realizados por parte dos investidores líderes.

Tabela 2 – Investidores com maior número de *cases* de sucesso como investidores líderes da primeira captação das empresas investidas

Investidor	**Investimentos de sucesso**	**Total de investimentos**
WOW Aceleradora	4	55
Ventiur Aceleradora	4	8
Bossa Nova Investimentos	3	3
Darwin Startups	3	3
ACE Startups	2	5

Embora não seja possível mensurar o retorno do portfólio dos entes estudados a partir da definição de sucesso dos investimentos neste e em outros estudos encontrados na literatura, considerando ainda a taxa de sobrevivência de *startups* observada na Figura 5, nota-se a presença na Tabela 2 de investidores renomados como Darwin Startups e ACE Aceleradora, previamente vencedores do prêmio Startup Awards Brasil, e da supracitada Bossa Nova Investimentos. Ademais, cabe a inferência da não correlação do sucesso do portfólio dos fundos de investimento com a estratégia *spray and pray* ou com sua alternativa.

4.2.2 Indicativos de sucesso

Ainda no contexto das observações feitas a partir do gráfico da Figura 21, surge também a oportunidade de se analisar os indicativos de sucesso das empresas passíveis de identificação no momento do primeiro evento de captação destas *startups*.

De maneira a investigar a influência do momento de mercado na capacidade de empresas com indicadores de performance inferiores de captarem recursos, foi calculada e inserida no conjunto de dados uma nova variável: a média das últimas 3 variações

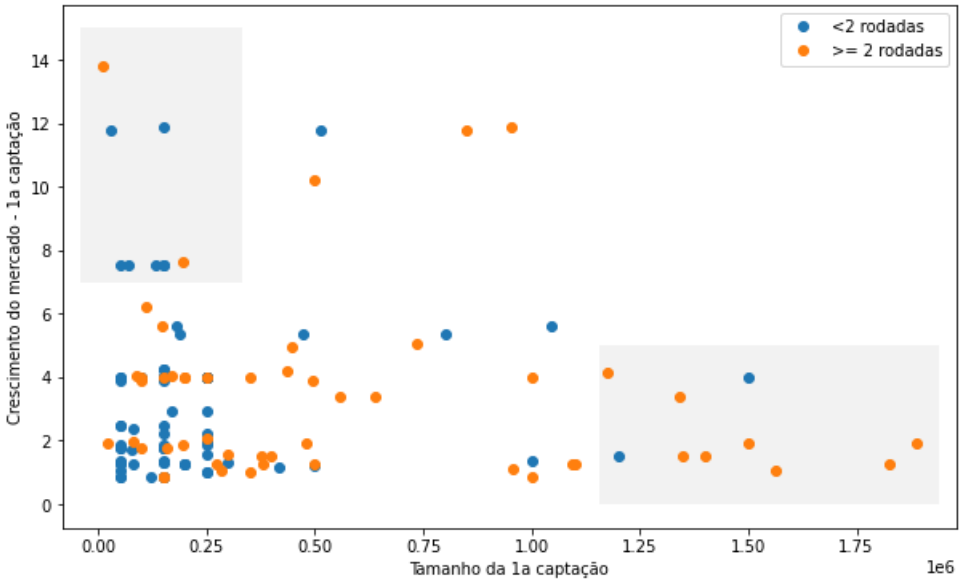

Figura 23 – Indicativos de sucesso de *startups* brasileiras no momento da primeira captação

percentuais do volume total de investimentos em empresas *early stage* no Brasil no momento da primeira captação.

Conforme destacado nas regiões acinzentadas do gráfico obtido (Figura 23), foram identificadas duas tendências no contraste do momento de mercado de capital de risco com a ordem da primeira captação de recursos: as *startups* capazes de levantar recursos da ordem de R$ 1.2 milhão a R$ 2 milhões em rodadas "Pre-Seed" e "Seed" em momentos de crescimento reduzido do mercado tendem a passar por rodadas subsequentes, enquanto aquelas que captam recursos de ordem menor em momentos de crescimento acelerado do mercado de capital de risco são mais propensas a serem incapazes de sobreviver ao ciclo de financiamento de uma *startup* a caminho de um evento de liquidez.

Tendo-se identificados os indicativos de sucesso mencionados no parágrafo anterior, bem como a já discutida relevância do primeiro sócio-investidor no que tange à maximização do sucesso das companhias, buscou-se propor um modelo capaz de prever o sucesso de *startups* brasileiras com base nestas 3 variáveis referentes à primeira captação: investidor líder, crescimento do mercado de investimentos em *early stage* e recursos levantados. Sendo o sucesso definido como uma variável booleana, portanto, utilizou-se o modelo XGBooster para endereçar o problema de classificação.

Por conta do conjunto de dados reduzido (124 empresas), foram ajustados os hiperparâmetros do modelo a fim de mitigar o sobreajuste, tendo como exemplo *max_ depth*, análogo ao autorregressor tipo floresta aleatória, conforme discutido na fundamentação teórica do capítulo 2. Foram realizadas, ainda, 20 iterações do modelo com o sorteio aleatório de 20% da base de dados como sendo o conjunto de teste; os resultados descritos na tabela 3 comprovam a possibilidade de se prever com grau relevante de precisão a

Tabela 3 – Resultado da aplicação do modelo XGBooster para previsão do sucesso de *startups* brasileiras

Menor precisão	40,00%
Maior precisão	88,89%
Média das precisões	67,83%
Desvio padrão das precisões	11,67%

capacidade de *startups* passarem por novas rodadas de investimento após a primeira captação de recursos, mesmo sem se ter acesso a indicadores de performance ou de tendência do mercado em que tais empresas atuam.

5 Conclusão

Para efeitos de comparação com os resultados obtidos, cabe mencionar os estudos listados na Tabela 4:

Tabela 4 – Resultados de estudos análogos encontrados na literatura

Título	**Autor**	**Referência**	**Métrica**	**Resultado**
Assessment of ML Performance for Decision Support in VC Investments	J. Arroyo et al.	Startups com rodadas subsequentes	Precisão (melhor resultado)	0.64
The driving forces of venture capital investments	Y. Ning et al.	Volume trimestral de investimentos	R^2 ajustado (melhor resultado)	0.353
What Drives VC Investments in Europe? New Results from a Panel Data Analysis	M. Cherif et al.	Volume de investimentos *early stage*	R^2 ajustado (melhor resultado)	0.48

Em relação ao primeiro estudo representado na Tabela 4, nota-se o objetivo similar àquele proposto no uso do modelo XGBooster discutido no capítulo 3. Quando traçado o comparativo com a Tabela 3, é possível observar a obtenção de resultados de ordem semelhante ou superior aos de Arroyo et al. (2019), ainda que utilizado um conjunto de dados com menor riqueza de detalhes. Infere-se, portanto, a relevância de fatores menos granulares para a previsão da sobrevivência de *startups* às primeiras fases do ciclo de financiamento, como crescimento do mercado de capital de risco no momento da primeira captação e sócio-investidor líder referente a este mesmo evento.

Os demais estudos descritos na Tabela 4, por sua vez, endereçam problemas semelhantes àquele supracitado em que se utilizou modelos de aprendizado de máquina para a previsão da variação mensal do volume de investimentos. Pelo fato de não serem encontrados na literatura estudos que tratem do problema por meio da previsão de séries temporais, mas sim por meio do cálculo da correlação pura e de regressões lineares considerando indicadores macroeconômicos e investimentos em capital de risco, observa-se também a utilização de diferentes métricas para avaliação dos modelos.

Além de se notar o valor da métrica R^2 ajustado inferior àquele amplamente considerado ideal em problemas de regressão, cabe ainda a análise crítica em relação à utilidade dos resultados obtidos nestes estudos em comparação aos do presente trabalho: conforme mencionado no capítulo anterior, fez-se possível neste a proposição de diferentes modelos para que se denotem diferentes tendências do mercado de capital de risco, como a

oscilação positiva ou negativa do volume de investimentos ou a potencialização da previsão do valor desta mesma variável.

Ainda no contexto do cumprimento dos objetivos propostos, entende-se como atingidos os demais objetivos específicos descritos no Capítulo 1, incluindo a análise de tendências específicas do período da pandemia do coronavírus e de indicativos particulares de sucesso de *startups*. Como aspecto geral, foi concretizado o embasamento matemático das investigações iniciais feitas para cada problema proposto por meio de métodos quantitativos de aprendizado de máquina e ciência de dados.

5.1 Sugestões de melhoria

Como fatores limitantes do estudo conduzido, cabe mencionar a limitação dos dados não só pelo número de amostras, mas também sob o aspecto da riqueza de detalhes: entende-se que o acesso a dados específicos sobre as empresas analisadas como receita, perfil dos fundadores e presença nas mídias sociais permitiria a maior precisão dos modelos de classificação propostos, com a consequente ampliada utilidade aos tomadores de decisão envolvidos no mercado de capital de risco.

Com relação ao volume mensal de investimentos em *Venture Capital* no Brasil, nota-se o mesmo padrão sob o aspecto do histórico recente de forte aceleração do crescimento do mercado; julga-se que os modelos propostos podem servir como base para análises feitas em um futuro próximo, em que há a tendência de se ter uma base de dados ainda mais ampla para a aplicação de modelos de aprendizado de máquina para previsão de séries temporais.

Referências

DISTRITO. Venture Capital: o que é e como funciona?. Disponível em: <https://distrito.me/blog/venture-capital-o-que-e-e-como-funciona/>. Acesso em: 14 nov. 2022. Citado na página 19.

GOLD, Shaun. A Brief History of Venture Capital. 2022. Disponível em: <https://openvc.app/blog/history-of-venture-capital#1960s-1980s-vc-becomes-an-asset-class>. Acesso em: 15 nov. 2022. Citado na página 19.

DISTRITO. Histórico do venture capital no Brasil: do surgimento até hoje. 2021. Disponível em: <https://distrito.me/blog/historico-do-venture-capital-no-brasil/>. Acesso em: 10 nov. 2022. Citado na página 19.

GLASNER, Joanna. Esses países têm o maior investimento em startups para seu tamanho. 2021. Disponível em: <https://news.crunchbase.com/startups/countries-most-startup-investment/>. Acesso em: 15 nov. 2022. Citado na página 20.

SEBRAE. O que é uma startup? 2014. Disponível em: <https://www.sebrae.com.br/sites/PortalSebrae/artigos/o-que-e-uma-startup,6979b2a178c83410VgnVCM1000003b74010aRCRD>. Acesso em: 10 nov. 2022. Citado na página 23.

SUDBRACK, Gustavo. Rodada De Investimentos Das Startups: Conheça Essa Jornada. 2020. Disponível em: <https://slap.law/a-jornada-de-investimentos-das-startups/>. Acesso em: 19 nov. 2022. Citado na página 24.

KEPLER, João. Escada de Investimento em Startups no Brasil. Disponível em: <https://joaokepler.com.br/escada-de-investimento-em-startups-no-brasil/>. Acesso em: 12 nov. 2022. Citado na página 24.

DISTRITO. Rodada de investimento: entenda como é o seu funcionamento. Disponível em: <https://distrito.me/blog/rodada-investimento-seed-series-a/>. Acesso em: 14 nov. 2022. Citado na página 25.

G2DINVESTMENTS. Como a queda no mercado tech afeta o Venture Capital? 2022. Disponível em: <https://www.g2d-investments.com/queda-no-mercado-tech/>. Acesso em: 10 nov. 2022. Citado na página 26.

MANZONI JUNIOR, Ralphe. A Bossanova quer ser a "XP das startups" (e vai ter até um banco). Disponível em: <https://neofeed.com.br/blog/home/

a-bossanova-quer-ser-a-xp-das-startups-e-vai-ter-ate-um-banco/>. Acesso em: 12 nov. 2022. Citado na página 25.

KUMAR, Ajitesh. Overfitting & Underfitting in Machine Learning. Disponível em: <https://vitalflux.com/overfitting-underfitting-concepts-interview-questions/>. Acesso em: 15 dez. 2022. Citado na página 27.

RODRIGO, Joaquín Amat; ORTIZ, Javier Escobar. Skforecast: previsão de séries temporais com Python e Scikit-learn. 2022. Disponível em: <https://www.cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html>. Acesso em: 19 nov. 2022. Citado na página 28.

SILVA, Josenildo Costa da. Aprendendo em uma Floresta Aleatória. Disponível em: <https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleat%C3%B3ria-3545f6babdf8>. Acesso em: 18 nov. 2022. Citado 2 vezes nas páginas 28 e 29.

Referência: GOLFETTE, Bruno; MELO, Maisa Kely de. Tutorial do modelo autorregressivo vetorial em Python. 2020. Disponível em: <https://maisamelo.medium.com/tutorial-do-modelo-autorregressivo-vetorial-em-python-175bfb80e0d3>. Acesso em: 19 nov. 2022. Citado na página 28.

MELO, Carlos. XGBoost: aprenda este algoritmo de Machine Learning em Python. 2019. Disponível em: <https://sigmoidal.ai/xgboost-aprenda-algoritmo-de-machine-learning-em-python/>. Acesso em: 21 nov. 2022. Citado 2 vezes nas páginas 30 e 31.

MADDALA, G s; LAHIRI, Kajal. Introduction to Econometrics. Reino Unido: John Wiley Sons Ltd, 2009. Citado 2 vezes nas páginas 31 e 32.

SPADINI, Allan Segovia. Séries temporais e suas aplicações. 2021. Disponível em: <https://www.alura.com.br/artigos/series-temporais-e-suas-aplicacoes>. Acesso em: 21 nov. 2022. Citado na página 32.

SIMÕES, Miguel. Machine Learning na previsão de Séries Temporais. 2021. Disponível em: <https://www.bi4all.pt/noticias/blog/machine-learning-na-previsao-de-series-temporais/>. Acesso em: 21 nov. 2022. Citado na página 33.

FERREIRA, Juliana Carvalho; PATINO, Cecilia Maria. O que realmente significa o valor-p? 2015. Disponível em: <https://www.scielo.br/j/jbpneu/a/SWk5XsCsXTW7GBZq8n7mVMJ/?format=pdf&lang=pt>. Acesso em: 26 nov. 2022. Citado na página 38.

J. Arroyo, F. Corea; G. Jimenez-Diaz; J. A. Recio-Garcia. Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments, IEEE Access, Vol. 7 (2019). Citado 2 vezes nas páginas 44 e 47.

Ning, Y., Wang, W. Yu, B. The driving forces of venture capital investments. Small Bus Econ 44, 315–344 (2015). Citado na página 47.

Mondher, C. Kaouthar, G. What Drives Venture Capital Investments in Europe? New Results from a Panel Data Analysis. Journal of Applied Business and Economics, 12 (2011). Citado na página 47.