



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO
CONHECIMENTO

Breno de Almeida Biagiotti

Framework conceitual para análise de Big Data Educacional em MOOCs

Florianópolis

2021

Breno de Almeida Biagiotti

Framework conceitual para análise de Big Data Educacional em MOOCs

Tese submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Engenharia e Gestão do Conhecimento.

Orientadora: Profa. Dra. Maria José Baldessar

Co-orientador: Prof. Dr. Francisco A.P. Fialho

Florianópolis

2021

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

Biagiotti, Breno de Almeida
Framework conceitual para análise de Big Data
Educativa em MOOCs / Breno de Almeida Biagiotti ;
orientador, Maria José Baldessar, coorientador, Francisco
Antônio Pereira Fialho, 2021.
228 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em ,
Florianópolis, 2021.

Inclui referências.

1. . 2. Big data educacional. 3. Mídia e conhecimento.
4. MOOCs. 5. MOODLE. I. Baldessar, Maria José . II.
Fialho, Francisco Antônio Pereira. III. Universidade
Federal de Santa Catarina. Programa de Pós-Graduação em .
IV. Título.

Breno de Almeida Biagiotti

Framework conceitual para análise de Big Data Educacional em MOOCs

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Márcio Vieira de Souza, Dr.
Universidade Federal de Santa Catarina

Prof.(a) Araci Hack Catapan, Dra.
Universidade federal de Santa Catarina

Prof.(a) Roseli Zen Cerny, Dra.
Universidade federal de Santa Catarina

Prof. Paulo Jorge Canas Rodrigues, Dr.
Universidade federal da Bahia

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia e Gestão do Conhecimento.

Prof. Dr. Roberto Carlos dos Santos Pacheco

Profa. Dra. Maria José Baldessar
Orientadora

Florianópolis, 2021.

Este trabalho é dedicado à minha esposa Gislaine, pela força e incentivo, à minha filha Joanna, pelos sorrisos sinceros, à Lucy e Tina, pela companhia na escrita e aos meus familiares e amigos que acompanham e torcem pelo meu crescimento pessoal.

AGRADECIMENTOS

Enfim a tese nasceu. O parto não foi fácil! Foi uma gestação longa, cerca de 5 anos, com períodos de grandes felicidades (congressos, aulas, confraternizações) e algumas complicações (prorrogação em virtude da pandemia, atrasos no cronograma, stress). Assim como a gestação, a tese tem seu tempo de amadurecimento. Faço essa analogia pois justamente passei pela experiência de ser pai durante o período da pesquisa e da escrita. E posso afirmar que, apesar das dificuldades, o resultado vale a pena! Afinal, a tese, assim como um filho(a), será sempre seu motivo de orgulho.

E por mais que seja uma caminhada muitas vezes solitária, não se faz uma tese sozinho. Portanto devo reconhecer com gratidão a ajuda de pessoas maravilhosas que passaram pelo meu caminho nesse período. À minha esposa Gi, que nunca duvidou do meu potencial e me levantava quando eu desanimava. À minha filha Joanna, que chegou para mostrar o que é o amor incondicional. À minha família que sempre perguntava como ia o doutorado, mesmo sem saber o tema do trabalho. Aos colegas do curso de auriculoterapia, sempre dispostos a ajudar, respondendo os questionários, disponibilizando os materiais. Agradecimento especial ao professor e amigo Lúcio Botelho, por autorizar o uso dos dados do curso de auriculoterapia. Ao professor Luiz Alberto Peregrino Ferreira, por me colocar no caminho da pós-graduação, incentivando minha inscrição no mestrado.

Agradeço também à Professora Lucy Parucker e os colegas do Telelab. Especialmente agradeço ao professor Marcos José Machado pela cordialidade e ajuda nas análises quantitativas!

Agradeço de coração minha orientadora Maria José Baldessar, pessoa que me “adotou” no mestrado e sempre estendeu a mão nos momentos que eu mais precisava. Gratidão também aos professores dessa banca, por aceitarem o convite e por contribuírem com seus conhecimentos.

Me contaram e eu esqueci. Vi e entendi. Fiz e aprendi.
Confúcio

RESUMO

Big Data é uma área em plena expansão no mundo atual, pois possibilita coletar uma enorme quantidade de dados, analisá-los e, a partir do resultado obtido, extrair *insights* para melhorar o seu negócio. Ao identificar tendências, padrões e inconsistências, o Big Data proporciona uma capacidade de entender o passado, monitorar o presente e antecipar o futuro. No âmbito da educação, tamanha capacidade de produção de dados, especialmente os *logs* registrados pelos sistemas de gestão de conteúdo como o Moodle, pode resultar em diversas oportunidades e melhorias para alunos e professores. Contudo, essa imensa quantidade de informações, se não for bem trabalhada e processada adequadamente, corre o risco de se tornar irrelevante. Por isso, é tão importante estudar formas de utilizar o Big Data educacional dentro das instituições de ensino, já que não é mais possível ignorar a sua capacidade de gerar vantagens competitivas. Ainda mais quando falamos da educação a distância e os cursos massivos online (MOOCs), cada vez mais frequentes e necessários na propagação de conhecimentos nas mais variadas áreas. Os procedimentos metodológicos utilizados neste trabalho se iniciaram por uma revisão narrativa, com objetivo de buscar a fundamentação teórica sobre três temas principais: Big Data Educacional, MOOCs e *Learning Analytics*. A segunda etapa foi uma revisão bibliográfica sistemática que buscou entender os principais *gaps* na pesquisa sobre big data educacional, os potenciais e os desafios da área, e também analisar alguns *frameworks* existentes na literatura. De uma forma geral, percebeu-se que a grande dificuldade encontrada para realizar uma análise satisfatória, é dominar as questões técnicas. Por exemplo: o professor que deseja melhorar seu curso, mas não sabe onde procurar os dados educacionais e nem quais métodos de análise utilizar; ou o cientista de dados que domina as técnicas analíticas, porém não entende de pedagogia e o funcionamento dos MOOCs; ou o design instrucional que conhece tudo sobre educação a distância, mas não entende nada de estatística e ciência de dados. Esses três exemplos são cenários comuns. Por se tratar de uma área interdisciplinar que engloba campos tão distintos do conhecimento, percebeu-se a necessidade de criar uma espécie de guia para orientar os pesquisadores. Para entender na prática como funciona essa análise de dados educacionais, realizou-se um estudo de caso de um MOOC sobre Auriculoterapia da UFSC, o qual capacitou milhares de alunos nos últimos anos. Foram realizadas pesquisas qualitativas com os autores e a equipe técnica do curso, além de uma análise da Big Data Educacional fornecida pelo Moodle do curso. Como resultado é apresentado um framework conceitual que tem como objetivo orientar pesquisadores no que se refere ao manejo da Big Data Educacional de MOOCs. Ele é composto de cinco etapas que abrangem desde a determinação dos objetivos e métricas da pesquisa, passando pela identificação e localização dos dados necessários e a indicação dos métodos mais apropriados para cada tipo de análise. O bom uso dos dados educacionais se torna uma vantagem competitiva importante para melhorar a aprendizagem e a experiência dos alunos nos cursos online.

Palavras-chave: Big Data Educacional. MOOCs. Learning Analytics. Framework conceitual

ABSTRACT

Big Data is an area in full expansion in today's world, as it makes it possible to collect a huge amount of data, analyze it and, from the result obtained, extract insights to improve your business. By identifying trends, patterns and inconsistencies, Big Data provides an ability to understand the past, monitor the present and anticipate the future. In the field of education, such data production capacity, especially the logs registered by content management systems such as Moodle, can result in several opportunities and improvements for students and teachers. However, this immense amount of information, if not well worked and processed properly, runs the risk of becoming irrelevant. Therefore, it is so important to study ways to use educational Big Data within educational institutions, since it is no longer possible to ignore its ability to generate competitive advantages. Even more when we talk about distance education and massive open online courses (MOOCs), which are increasingly frequent and necessary in the dissemination of knowledge in the most varied areas. The methodological procedures used in this work began with a narrative review, with the aim of seeking theoretical foundations on three main themes: Educational Big Data, MOOCs and Learning Analytics. The second stage was a systematic literature review that sought to understand the main gaps in research on educational big data, the potentials and challenges in the area, and also to analyze some existing frameworks in the literature. In general, it was noticed that the great difficulty found to carry out a satisfactory analysis is to master the technical issues. For example: the teacher who wants to improve his course, but does not know where to look for educational data or which analysis methods to use; or the data scientist who has mastered analytical techniques but does not understand pedagogy and how MOOCs work; or instructional design that knows all about distance education but understands nothing about statistics and data science. These three examples are common scenarios. As this is an interdisciplinary area that encompasses such different fields of knowledge, the need to create a kind of guide to guide researchers was perceived. To understand in practice how this analysis of educational data works, a case study of a MOOC on Auriculotherapy at UFSC was carried out, which has trained thousands of students in recent years. Qualitative research was carried out with the authors and the course's technical team, in addition to an analysis of the educational big data provided by the course's Moodle. As a result, a conceptual framework is presented that aims to guide researchers regarding the management of Educational Big Data of MOOCs. It consists of five steps that range from determining the research objectives and metrics, through identifying and locating the necessary data and indicating the most appropriate methods for each type of analysis. Good use of educational data becomes an important competitive advantage to improve the learning and experience of students in online courses.

Keywords: *Educational Big Data. MOOCs. Learning Analytics. Conceptual framework*

LISTA DE FIGURAS

Figura 1 - Crescimento do mercado de Big Data	19
Figura 2 - Segmentos afetados pelo crescimento da Big Data	20
Figura 3 - Principais linhas de pesquisas sobre MOOCs	25
Figura 4 - Framework de literacia de dados e suas competências.....	34
Figura 5 - infográfico “Extraindo valor de mercado dos 4 Vs da Big Data	36
Figura 6 - Sequência de tomada de decisões para análise de dados	39
Figura 7 - Diferenças entre os dados educacionais estáticos e dinâmicos	41
Figura 8 - Categorias dos dados educacionais.....	42
Figura 9 - Tipos de dados educacionais na educação superior.....	43
Figura 10-Visão macro dos níveis de dados educacionais	45
Figura 11 - Ciclo de curadoria de dados.....	50
Figura 12 - categorização dos dados educacionais.....	47
Figura 13 - Principais fontes de dados educacionais.....	49
Figura 14 - Diretrizes elaboradas pela ENISA.....	55
Figura 15 - Aspectos éticos e qualitativos dos dados educacionais	57
Figura 16 - Interseção de áreas de conhecimento que compõe o LA.....	60
Figura 17 - Principais objetivos do LA	64
Figura 18 - Modelo referencial de LA.....	65
Figura 19 - Ciclo de vida do LA.....	67
Figura 20 - categorias de análise dos dados	70
Figura 21 - Etapas do manejo dos dados educacionais	72
Figura 22 - Exemplo de fórum comum em MOOCs.....	83
Figura 23 - imagem da rede de interações dos alunos.....	87
Figura 24 - Tipos comuns de gráficos usados na visualização de dados.....	90
Figura 25 - Técnicas de ML mais utilizadas	93
Figura 26 - Principais técnicas utilizadas na gestão de dados educacionais	94
Figura 27 - Modelo de funcionamento da ferramenta de LA do Moodle	100
Figura 28 - Framework conceitual da comunidade de Investigação	102
Figura 29 - Evolução da tradição em pesquisa educacional.....	107
Figura 30 - Nuvem de palavras geradas através da RBS	110
Figura 31 - Protótipo de dashboard elaborado pelo MOOC Research Project.....	117
Figura 32 - Modelo de dashboard gerado pelo Google Big Query	118

Figura 33 - Modelo ADDIE revisitado para ambientes digitais.....	121
Figura 34 - Etapas do estudo	125
Figura 35 - Gráfico de respostas sobre a abordagem pedagógica do curso	133
Figura 36 - árvore de decisão para determinação do conceito de evasão.....	138
Figura 37 - <i>Dashboard</i> para análise das notas dos alunos	143
Figura 38 - Visualização das médias das notas e o número de tentativas	144
Figura 39 - Comparativo aprovados X reprovados X evadidos	141
Figura 40 - Tabela com dados de alunos que abandonaram o curso na etapa final .	145
Figura 41 – Etapa 1 do framework - Objetivos / Métricas	151
Figura 42 - Etapa 2 do framework: Big Data Educacional	153
Figura 43 - Etapa 3 do framework: Tabelas Moodle.....	156
Figura 44 - Etapa 4 do Framework: Tipo de Análise	159
Figura 45 - Indicações numéricas que relacionam os temas do framework.....	161
Figura 46 - Etapa 5 do Framework: Métodos de Análise.....	162
Figura 47 - Framework Conceitual para análise de BDE em MOOCs	163

LISTA DE QUADROS

Quadro 1 - Aspectos positivos e negativos do LA	27
Quadro 2 - Stakeholders dos MOOCs e suas características.....	28
Quadro 3 - Quadro temático de mestrado e doutorado do EGC/UFSC	30
Quadro 4 - Pesquisas do EGC com temáticas correlatas ao tema desta teses	31
Quadro 5 - Dimensões do framework	35
Quadro 6 - Características de cada área de pesquisa	37
Quadro 7 - Lista de dados fornecidos pelo LMS Moodle	47
Quadro 8 - Lista de plugins úteis na análise de dados educacionais	48
Quadro 9 - Analytics e suas variações.....	61
Quadro 10 - Métodos de visualização de dados	89
Quadro 11 - Panorama das técnicas utilizadas em LA	93
Quadro 12 - Os 3 perfis do moodle e suas atribuições	97
Quadro 13 - Lista de recursos e atividades do Moodle	98
Quadro 14 - Procedimentos metodológicos da pesquisa.....	105
Quadro 15 - Diferenças entre os formatos de MOOCs	113
Quadro 16 - Objetivo da fase 1 – Análise	122
Quadro 17 - Objetivo da fase 2 - Design.....	123
Quadro 18 - Objetivo da fase 3 - Desenvolvimento	123
Quadro 19 - Objetivo da fase 4 - Implementação.....	124
Quadro 20 - Objetivo da fase 5 - Avaliação	124
Quadro 21 - Estrutura do curso de auriculoterapia.....	126
Quadro 22 - Tabelas do Moodle utilizadas neste estudo.....	136
Quadro 23 - Cronograma das atividades de análise	139

LISTA DE TABELAS

Tabela 1 - Evasão de alunos por região.....	142
Tabela 2 - Número de alunos aprovados e reprovados por região	146

LISTA DE ABREVIATURAS

MPB - Música Popular Brasileira
GPS - *Global Position System*
LMS - *Learning Management System*
BDE - Big Data Educacional
MOOC - *Massive Open Online course*
LA - Learning Analytics
MDE - Mineração de Dados Educacionais
ENISA - Agência de Segurança da Rede e de Informações da União Europeia
MIT - *Massachusetts Institute of Technology*
EGC - Engenharia e Gestão do Conhecimento
GC - Gestão do Conhecimento
KDD - *Knowledge Discovery in Databases*
BI - *Business Intelligence*
RBS - Revisão Bibliográfica Sistemática
EDX - Provedora de MOOCs americana
TIC - Tecnologias de Informação e Comunicação
ML - *Machine Learning*
IA - Inteligência Artificial
SNA - *Social Network Analysis*
GDPR – *General Data Protection Regulation*
LGPDP – Lei Geral de Proteção de Dados Pessoais
OCR – Reconhecimento ótico de caracteres
RG – Registro geral
CPF – Cadastro de pessoa física
AM – Aprendizagem de máquina
SNA – *Social Network Analysis*
ML – *Machine Learning*
MOODLE – *Modular Object-Oriented Dynamic Learning Environment*
PIC – Práticas integrativas e complementares
MTC – Medicina tradicional chinesa
ESF – Estratégia de saúde da família

SUS – Sistema único de saúde

SCORM – *Sharable Content Object Reference Model*

EAD – Educação a distancia

SETIC - Superintendência de Governança Eletrônica e Tecnologia da Informação e
Comunicação

SUMÁRIO

1	INTRODUÇÃO	18
1.1	OBJETIVO GERAL	23
1.2	Objetivos específicos	23
1.3	Justificativa e relevância	24
1.4	Delimitação e escopo da pesquisa	28
1.5	Aderência ao PPGE GC	29
2	BIG DATA EDUCACIONAL	33
2.1	Características da big data	35
2.2	Pesquisa educacional x pesquisa em big data educacional	37
2.3	Importância dos dados educacionais	38
2.4	Características dos dados educacionais	39
2.5	Aonde procurar por dados educacionais?	46
2.6	Curadoria de dados educacionais	49
2.6.1	Etapa 1 - organização dos dados	51
2.6.2	Etapa 2 - reuso dos dados	52
2.7	Cuidados com a gestão dos dados	53
2.8	Qualidade dos dados	55
2.9	Direitos individuais dos que cederam os dados	58
3	<i>LEARNING ANALYTICS</i>	59
3.1	Objetivos do la	62
3.2	Frameworks existentes	64
3.3	Métricas	68
3.4	Ciência de dados e educacao	71
3.5	Métodos estatísticos	72
3.6	Mineração de dados	74
3.7	Aprendizagem de máquina (<i>machine learning</i>)	76
3.8	Análise qualitativa	81

3.9	<i>Open learning analytics</i>	83
3.10	Análise de redes sociais	85
3.11	Apresentação dos dados	88
3.12	Técnicas mais utilizadas	91
4	MOODLE	96
4.1	Recursos e atividades do Moodle	98
4.2	Base de dados	99
4.3	<i>Inspire Analytics</i>	100
5	PROCEDIMENTOS METODOLÓGICOS	105
5.1	A quarta tradição metodológica da pesquisa educacional	106
5.2	Revisão Bibliográfica Sistemática	108
5.2.1	Discussão dos resultados da RBS	110
5.2.2	<i>Machine learning</i>	111
5.2.3	Aprendizagem colaborativa	112
5.2.4	Engajamento e evasão	114
5.2.5	Comportamento e padrões de aprendizagem	115
5.2.6	Performance de aprendizagem	116
5.2.7	Apresentação e <i>dashboards</i>	117
5.2.8	Personalização	118
5.2.9	Temas secundários	119
6	ESTUDO DE CASO: MOOC AURICULOTERAPIA	125
6.1	o contexto do curso	126
6.1.1	Etapa presencial	128
6.1.2	Processo de certificação	128
6.2	Avaliação do curso pelos alunos	129
6.2.1	Coleta de dados qualitativos: questionário com integrantes do projeto	130
6.3	Coleta de dados quantitativos: extraindo dados educacionais do moodle	136

6.4	Análise de dados	138
7	<i>FRAMEWORK</i> CONCEITUAL DE BIG DATA EDUCACIONAL	148
7.1	Proposição do Framework	150
7.2	Objetivos / métricas	150
7.3	Big Data Educacional	153
7.4	Tabelas Moodle	155
7.5	Tipo de análise	158
7.6	Métodos de análise	161
8	CONCLUSÕES E TRABALHOS FUTUROS	165
	APÊNDICE 1	169
	APÊNDICE 2	179
	APÊNDICE 3	187
	APÊNDICE 4	194
	APÊNDICE 5	199
	APÊNDICE 6	202
	APÊNDICE 7	203
	APÊNDICE 8.....	210
9	REFERÊNCIAS BIBLIOGRÁFICAS	221

1 INTRODUÇÃO

Você já parou para pensar que nos dias atuais praticamente tudo o que fazemos em nossas rotinas deixa rastros? Tome como exemplo um cidadão comum em um dia normal de trabalho. Ele acorda e vai preparar o seu café. Enquanto isso navega pelos principais portais de notícias, que já direcionam o conteúdo de acordo com seu interesse. No caminho do trabalho, coloca no Spotify uma música da Cássia Eller, e o programa já recomenda outros artistas similares da MPB, todos que agradam seu gosto musical. Nisso ele acaba descobrindo um novo artista ou uma nova música.

Ao parar no posto de combustível, realiza o pagamento com o cartão de crédito e aproveita para pontuar no programa de fidelidade. Ao chegar no trabalho, acessa a internet e cada clique no mouse, cada página visitada, fica arquivado no histórico os seus rastros de navegação. Como suas férias estão chegando, ele aproveita para pesquisar destinos para sua próxima viagem. Sendo assim os algoritmos já começam a oferecer pacotes de viagem, oferta de hotéis, produtos de viagens, malas, seguros etc...

No retorno para casa, para em uma padaria, compra alguns produtos e, ao sair, seu celular pede para avaliar o local. Como o trânsito está muito congestionado, o GPS sugere uma rota alternativa, já que sabe que seu destino será sua residência.

Antes de dormir você resolve ver uma série no Netflix, que sugere inúmeras atrações de acordo com seu perfil. Você escolhe uma comédia romântica com base em um site de avaliação que outros usuários contribuíram, afinal hoje em dia não temos tempo para ver filmes ruins...

Essa situação narrada, comum a muitas pessoas, nos leva a uma reflexão de que fornecemos dados a todo momento. Mesmo inconscientemente, nossos atos deixam rastros que podem ser detectados e explorados pelas empresas detentoras dessas informações, que criam anúncios e publicidades cada vez mais assertivas. A sociedade atual enfrenta um desafio crescente imposto pela Big Data¹.

¹ Big Data apresenta um conceito relativo, porém se refere a uma quantidade exponencial de dados que supera a habilidade dos softwares de captura, armazenamento, gerenciamento e análise (Manyika *et al.*, 2011, p.1).

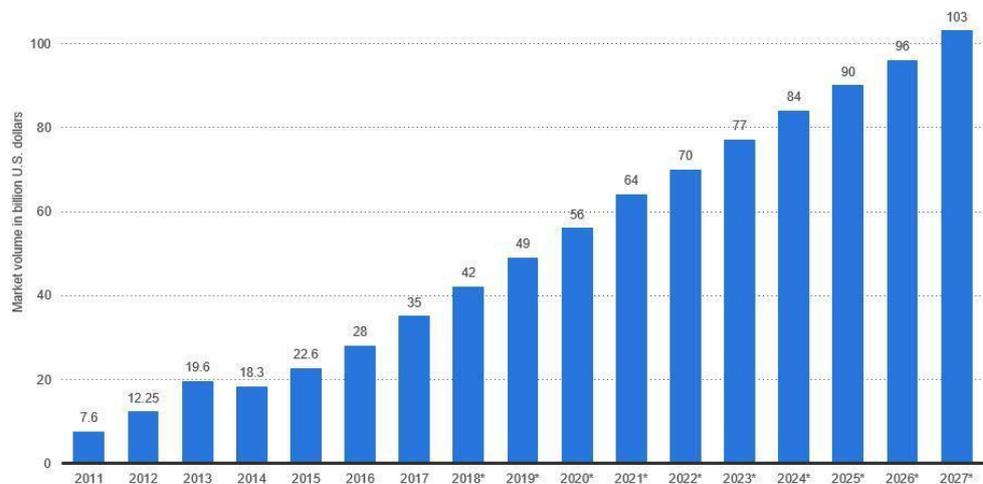
Nesse contexto, quem tem a habilidade de extrair informações relevantes a partir dessa massa enorme de dados possui uma vantagem competitiva em relação aos concorrentes. Isso porque a Big Data pode ser útil na melhoria de processos nas empresas, melhoria no relacionamento com clientes, pois é capaz de identificar padrões e criar tecnologias que melhoram o nosso cotidiano.

Nos capítulos seguintes será abordado com mais profundidade o conceito e diretrizes da Big Data e sua relação com os aspectos educacionais, que compõem o cerne dessa tese. Por hora, se faz necessário entender a importância desses dados e sua relevância no mercado. A Figura 1 mostra uma previsão de receita e investimentos em Big Data, de 2011 a 2027, na qual nota-se claramente o crescimento de investimentos nesse mercado.

Figura 1 - Crescimento do mercado de Big Data²

Forecast Revenue Big Data Market Worldwide 2011-2027

Big Data Market Size Revenue Forecast Worldwide From 2011 To 2027
(in billion U.S. dollars)



Fonte: Wikibon (2019)

Percebe-se um crescimento exponencial do mercado de Big Data ano após ano. Segundo Liebowitz (2013), o crescimento do volume de dados em setores como saúde e educação crescem cerca de 35% ao ano. Isso por que o mundo da Big Data envolve diversos

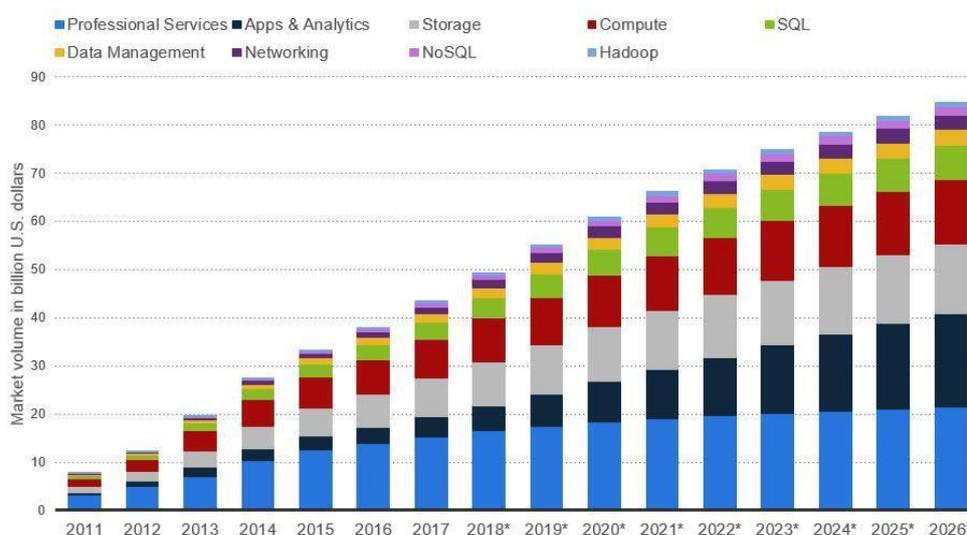
²<https://www.forbes.com/sites/louiscolombus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/#62f9587d2926>

setores, ou seja, tem um caráter interdisciplinar. A Figura 2 mostra os segmentos que serão mais impactados com o crescimento da Big Data:

Figura 2 - Segmentos afetados pelo crescimento da Big Data³

Big Data Market Worldwide Segment Revenue Forecast 2011-2026

**Big Data Market Forecast Worldwide from 2011 to 2026, by segment
(in billion U.S. dollars)**



Fonte: Wikibon (2019)

Os dois principais segmentos que mais crescerão nos próximos anos são o mercado de profissionais hábeis a trabalhar com Big Data, assim como o desenvolvimento de aplicativos e conhecimento em técnicas analíticas. Para entender a Big Data, é necessário considerar três aspectos, conhecidos como os 3Vs: volume, velocidade e variedade. O volume nada mais é do que a quantidade de dados que possuímos; velocidade é o tempo que os dados são obtidos e variedade são os tipos de dados, originários de diversas fontes diferentes. Os profissionais analíticos utilizam o conhecimento sobre os 3Vs para cruzar informações e gerar *insights* para a resolução de problemas e proposição de inovações nas mais diversas áreas. Alguns autores utilizam a classificação de 4Vs, conforme veremos mais adiante.

Corroborando com essa tendência do mercado, o *Horizon Report - Higher Education Edition*, em sua versão do ano de 2018, identificou e descreveu as tendências e desafios da

³ <https://www.forbes.com/sites/louiscolombus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/#62f9587d2926>

tecnologia educacional para o ensino superior. Na categoria "Impulsos para adoção de tecnologia" para os próximos dois anos, o foco na **medição de aprendizagem** e no **re-design dos ambientes de aprendizagem** receberam destaque:

Essa tendência descreve o interesse na validação de uma ampla variedade de métodos e ferramentas que os educadores usam para avaliar, medir e documentar a prontidão acadêmica, progresso de aprendizagem, aquisição de habilidades, entre outras necessidades educacionais dos estudantes. (...) A proliferação de *softwares* de mineração de dados e seu desenvolvimento na educação on-line, o aprendizado móvel e os sistemas de gerenciamento de aprendizado (LMS) estão se unindo aos ambientes de aprendizagem que se aproveitam e entregam análises e visualização de dados de aprendizado, de uma maneira multidimensional e portátil (Becker *et al.*, 2018)

Na categoria "Importantes Desenvolvimentos Tecnológicos", as **tecnologias analíticas** receberam destaque, com implementação no curto prazo. Em um período de dois a três anos, **tecnologias de aprendizagem adaptativa** serão realidade:

Cientistas de dados e plataformas de coleta de dados agora são capazes de organizar computacionalmente *petabytes* e *exabytes* de dados para que seja fácil analisar e identificar padrões que antes não eram detectáveis. Com a complexidade em torno de conjuntos tão grandes e diversos de dados, a visualização da informação é crucial para o sucesso. Ela se utiliza de infográficos, representação gráfica de dados técnicos projetados para serem compreendidos de maneira rápida e fácil. Na educação, a mineração de dados já está sendo utilizada para atingir os alunos com necessidades, personalizar a aprendizagem e criar caminhos flexíveis para o sucesso. À medida que as instituições educacionais se tornam mais hábeis em trabalhar com a interpretação da Big Data, elas podem tomar decisões mais assertivas que refletem nas reais necessidades dos alunos. No entanto, ao mesmo tempo, há uma necessidade de melhores ferramentas e de uma exploração mais profunda sobre como coletar e analisar esses dados. Se nós queremos que os alunos, educadores e gestores usem melhor os dados disponíveis, eles precisam de ferramentas e treinamento para fazer isso. Os alunos devem ter bons métodos para avaliar e gerenciar informações. Entender como usar novas ferramentas de dados e potencializar as habilidades analíticas, incluindo "alfabetização de dados", pensamento computacional e programação, são essenciais para que tanto o corpo docente como os alunos avancem na compreensão e uso da Big Data. (Becker *et al.*, 2018)

Portanto há uma clara necessidade de aprender a trabalhar com esses dados para conseguir alcançar resultados significativos. A análise de dados educacionais, entretanto, apesar de parecer um fenômeno recente, já foi alvo de pesquisas em 1979. Naquela ocasião, Mcintosh (1979) realizou estudos longitudinais na Inglaterra, ao monitorar o progresso de milhares de alunos ao longo de dez anos. Foram analisados diversos cursos e, mesmo naqueles tempos remotos, o autor já alertava para uma "explosão de dados", com riqueza e profusão disponíveis, atuando como uma barreira ao seu uso:

Por definição, BDE se refere à grande quantidade de dados educacionais eletrônicos que são tão complexos que se torna difícil (ou impossível) de administrar com softwares e hardwares tradicionais, mas que podem ser gerenciados com ferramentas comuns de gestão de dados e métodos apropriados. BDE é uma área relativamente nova e inexplorada. Os recursos da Big Data dos MOOCs geralmente são compostos de três partes: usuários do MOOC, materiais dos cursos e comportamento dos usuários. (ZHENG & YIN, 2015)

O cenário atual da educação, impulsionado pelos avanços tecnológicos e uma demanda crescente de educação a distância potencializou a geração e o armazenamento massivo de dados de milhares de alunos espalhados pelo mundo. Segundo Daniel (2015), a BDE é um novo paradigma de pesquisa pois estimula novas formas de elaborar hipóteses, de estruturar estudos, analisando e visualizando dados.

A popularização de diversas formas de tecnologias digitais para aprendizagem, como é o caso dos *Massive Open Online Courses* (MOOCs) contribuiu para a aquisição de um grande volume de dados dos alunos. Muitos desses são similares com os dados de aulas presenciais: materiais de aprendizagem, informações demográficas, dados de participação em atividades, avaliações e notas. Corroborando com isso, Zheng & Yin (2015) afirmaram:

Os comportamentos dos usuários também são uma grande parte dos dados massivos do MOOC. Por exemplo, segundo a segundo, dados de interação do usuário em vídeos do MOOC, como o abandono de vídeos, sessões de re-assistir; o fluxo de clique comportamental dos alunos, como quantos vídeos do curso eles assistiram, quantos textos do curso eles leram, quantas atividades do curso eles completaram; *log* completo de atividades estudantis, como o tempo gasto em recursos específicos do curso; comportamento de postagem em fóruns de discussão do MOOC; comportamento de conversa em salas de bate-papo do MOOC. Todos estes dados despertam interesse em muitos pesquisadores. (Zheng & Yin, 2015 p.682)

Novas formas de integrar essa vasta quantidade de dados e métodos com objetivo de extrair alguma informação útil delas vem sendo estudadas (Cook & Das, 2012). Duas áreas que estão em pleno desenvolvimento e ganham destaque nesse contexto de utilizar Big Data na educação são o *Learning Analytics* (LA) e a Mineração de dados educacionais (MDE). Apesar de apresentarem objetivos parecidos, elas diferem em suas origens, técnicas aplicadas e tipos de descoberta (Chatti *et al.*, 2012). Segundo Papamitsiou e Economides (2014), o LA tem uma característica mais holística, ou seja, procura compreender todo o sistema em sua complexidade, enquanto o MDE adota uma visão reducionista e positivista, procurando por padrões e modificando algoritmos através de métodos computadorizados. Ambos têm como objetivo implementar a experiência de ensino e aprendizado nos cursos online.

Diante dessa realidade, nos últimos anos houve um aumento de pesquisas e projetos com foco no desenvolvimento de ferramentas, sistemas e estratégias para implementar iniciativas de sucesso que realmente melhorassem o desempenho dos alunos. A essa mistura da necessidade de implementar a aprendizagem com o uso inteligente dos dados, surgiu o que a literatura chama de LA. Desde sua primeira menção, em meados de 2012, LA vem chamando a atenção da comunidade acadêmica e, por consequência, ganhando relevância. Muitos pesquisadores propuseram definições, mas basicamente todas abordam LA como a "medição, coleta, análise e reporte de dados sobre estudantes e seus contextos, com o propósito de entender e otimizar a aprendizagem e o ambiente que ela ocorre (Elias, 2011). Alejandro Peña-Ayala, pesquisador mexicano, ressaltou a importância do LA no prefácio de um livro sobre o tema:

Neste contexto, a análise de aprendizagem (LA) surge como uma disciplina emergente que busca melhorar o ensino e a aprendizagem por meio de uma avaliação crítica dos dados brutos e a geração de padrões que caracterizam os hábitos do aluno, prever as respostas e fornecer *feedback* oportuno. Além disso, LA suporta tomada de decisões, adequa o conteúdo, facilita avaliações realistas e fornece supervisão pessoal do progresso do aluno. O objetivo é dimensionar a exploração em tempo real de LA por alunos, acadêmicos e sistemas educacionais baseados em computador para melhorar as realizações dos alunos no curso (Alejandro Peña-Ayala, 2017 p.5)

Desde então foram desenvolvidos alguns *frameworks*, estudos empíricos e até mesmo congressos internacionais sobre o tema, com a finalidade de trocar conhecimentos e experiências nessa área. Apesar dos avanços nas pesquisas, a maioria dos *frameworks* propostos são muito genéricos. Abordam essas questões com focos diferentes e, geralmente, não aprofundam o estudo dos métodos analíticos. Também não foi encontrado um *framework* específico que analise a Big Data Educacional (BDE) de MOOCs. Dessa forma, pergunta-se: **como propor um *framework* conceitual de *Learning Analytics* que oriente educadores e gestores na gestão da big data educacional de MOOCs?**

1.1 OBJETIVO GERAL

Elaborar um *framework* conceitual de *learning analytics*, que oriente educadores e gestores na gestão da big data educacional de MOOCs.

1.2 OBJETIVOS ESPECÍFICOS

- Levantar os principais dados educacionais utilizados;
- Analisar a estrutura da big data educacional de MOOCs que utilizam o Moodle;

- Identificar os principais métodos para análise da big data educacional;
- Aplicar os métodos analíticos na big data educacional de um MOOC;

1.3 JUSTIFICATIVA E RELEVÂNCIA

A escolha pelo tema dessa tese se deu pelo interesse do autor a respeito do tema, tanto de forma acadêmica como profissional. Em meados de 2015 iniciou-se um grande projeto de extensão na UFSC que visava capacitar massivamente profissionais da área da saúde para o uso da auriculoterapia na atenção básica, por meio da modalidade de educação a distância. Através de pesquisas em bases de dados científicas, buscava-se um formato apropriado de curso online que comportasse essa demanda. Nessa mesma época diversos artigos sobre MOOCs estavam surgindo, com relatos de experiências promissoras.

Renomadas universidades, como Stanford, MIT e Harvard, foram pioneiras ao investirem milhões de dólares na criação de suas plataformas, impulsionando os MOOCs e atraindo muitos pesquisadores interessados. O tema foi gradativamente se tornando mais relevante e presente nas pesquisas.

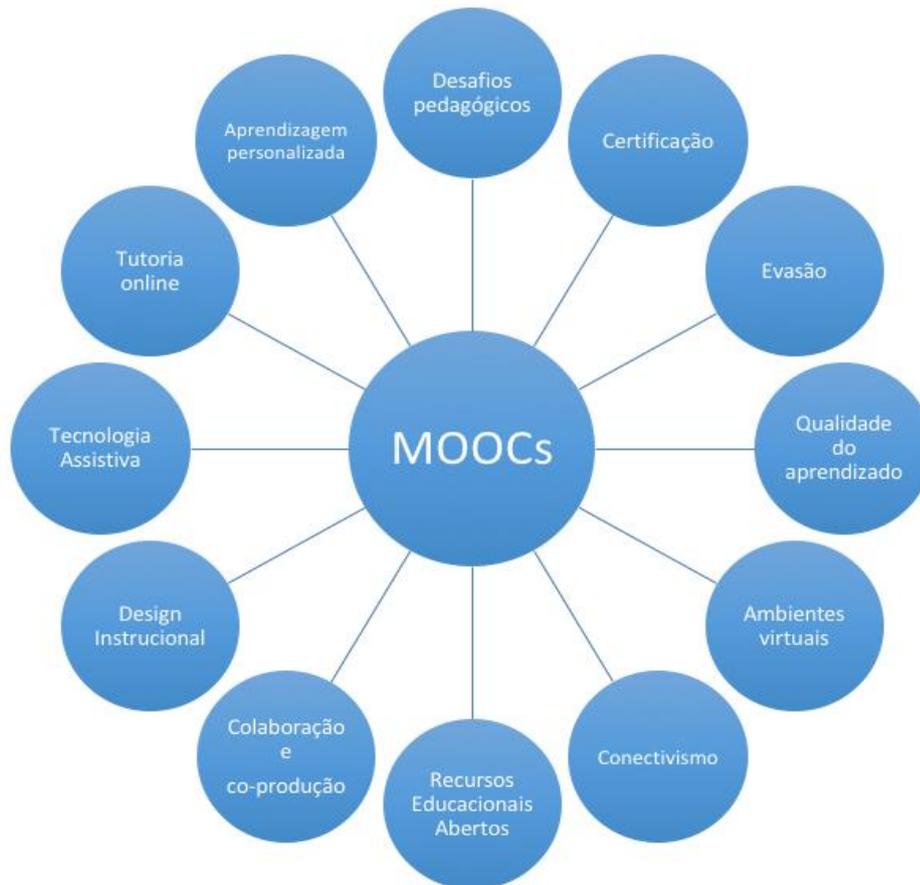
A medida que novas publicações sobre MOOCs iam surgindo na literatura, novas linhas de pesquisa sobre o tema iam despontando. Logo apareceram várias críticas sobre esse formato, onde o aluno tinha total liberdade para acessar o conteúdo, autonomia na escolha de seus caminhos de aprendizagem (abordagem construtivista) e questionamentos pedagógicos sobre esse formato foram levantados. Os métodos de certificação também não tinham um padrão e notou-se que as taxas de evasão desses cursos eram bem elevadas.

Em 2014, o autor desse trabalho iniciou seu mestrado na UFSC, no Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento e aprofundou-se na pesquisa sobre os MOOCs, utilizando como estudo de caso o MOOC Telelab. Através de questionários com alunos e gestores educacionais, juntamente com análise dos bancos de dados fornecidos pelo sistema Moodle, foi elaborada a dissertação sobre a avaliação da qualidade da informação em MOOCs, a qual foi defendida em 2016⁴.

⁴ Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/167690>

Desde então, novas linhas de pesquisa sobre MOOCs foram surgindo paulatinamente, permeando aspectos pedagógicos, tecnológicos e até mesmo culturais. A Figura 3 mostra as principais linhas de pesquisas em MOOCs nos últimos anos.

Figura 3 - Principais linhas de pesquisas sobre MOOCs



Fonte: Elaborado pelo autor (2019)

Com o passar do tempo e uma maior maturidade nas pesquisas sobre os MOOCs, muitos dados fornecidos por esses cursos começaram a ser utilizados, servindo de indicadores para pesquisas científicas. Ficou mais fácil, por exemplo, entender os motivos das altas taxas de evasão, ou quais metodologias agradaram mais os alunos de um determinado curso. Percebeu-se que a BDE fornecida por esses cursos poderia ser útil para prever alguns comportamentos e auxiliar na implementação de melhorias.

A ciência de dados não é uma linha de pesquisa exclusiva dessa área, muito menos uma novidade no campo científico. A busca por informações relevantes em base de dados é antiga, mas a aplicação de novas técnicas e a finalidade de implementar a aprendizagem, isso sim, deu origem ao campo de LA.

Quando se associa o uso de técnicas qualitativas de LA em cursos massivos, o principal fator que se apresenta é a dificuldade de obter o *feedback* personalizado dos alunos, devido à grande quantidade de envolvidos. Nesses casos a recomendação é aplicar as técnicas em pequenos grupos para facilitar o monitoramento e o suporte durante o processo (Fournier *et al.*, 2011). Outros autores também corroboram com essa ideia:

Tradicionalmente, professores (e tutores) são capazes de suportar um número limitado de estudantes, com *feedback* personalizado. Este problema tem sido compensado pela inclusão de assistentes de ensino relativamente baratos em cursos, mas com matrículas na casa da centena de milhares se torna impossível manter um suporte apropriado aos alunos em função dos recursos limitados e orçamento (Drachler e Kalz, 2016 p.281).

Segundo Zengh e Yin (2015), o fluxo de geração de dados em cursos massivos nunca foi tão grande (velocidade) e o acúmulo é contínuo, gerando um grande volume de dados. Esses dados, entretanto, são de natureza estruturada, semiestruturada e não-estruturada, combinando fatores que tornam a tarefa de aplicar LA nos MOOCs um grande desafio aos pesquisadores e gestores de cursos. A diversidade estrutural e a complexidade do Big Data são atribuídas ao surgimento de novas formas de dados gerados por redes de sensores, aplicativos de mídia social e outros dispositivos móveis e ubíquos (Manyika *et al*, 2011; Snijders, Matzat, & Reips, 2012; Ward & Barker, 2013).

Notou-se, entretanto, um *gap* nas pesquisas sobre LA e MOOCs, uma vez que havia necessidade de mais estudos empíricos, compartilhamento de experiências e técnicas exitosas e, logicamente, estudos estruturados com rigor científico. Muitos artigos integravam técnicas de análise de dados de uma forma confusa, misturando conceitos e metodologias, dificultando o entendimento.

Nesse contexto, existem basicamente duas abordagens principais para a análise de dados educacionais: Learning Analytics e Mineração de dados educacionais (MDE). Enquanto o LA abrange uma série de várias ferramentas e possibilidades de trabalhar com grandes quantidades de dados complexos de estudantes, levando em conta o contexto que eles ocorrem (Greer & Mark, 2016), MDE é mais focado na descoberta de conhecimento por meio de coleta e análise de dados de forma automatizada. Embora os dois tenham objetivos em comum (melhorar a aprendizagem), LA parece ser mais adequado para análise de dados tão heterogêneos dos MOOCs e oportuniza *insights* para a melhoria do processo educacional. Os benefícios, entretanto, são diversos para quem consegue converter esses dados em informações.

No Quadro 1 podemos analisar alguns aspectos positivos e negativos da utilização do LA:

Quadro 1 - Aspectos positivos e negativos do LA

Aspectos Positivos	Aspectos Negativos
<ul style="list-style-type: none">● Grande volume de dados educacionais● Múltiplas interpretações para auxiliar os alunos● Descobrir pontos críticos e padrões de aprendizagem● Gerar insights e estratégias de aprendizado	<ul style="list-style-type: none">● Má interpretação dos resultados (falha de interpretação humana)● Viés na análise dos dados● Fonte de dados é heterogênea e não-amigável;● Predominantemente quantitativo. Métodos qualitativos não se provaram efetivos;● Questões éticas (privacidade de dados)● Transparência (alunos não sabem que estão fazendo parte dessas experiências)● Armazenamento (custo com equipamentos e quantidade de dados que aumentam exponencialmente)

Fonte: adaptado de Papamitsiou e Economides (2014) e Kahlil *et al.* (2016)

Pensando justamente nessa dificuldade de trabalhar com esse grande volume de dados, e também a importância de saber analisá-los corretamente, percebeu-se a necessidade de elaborar um *framework* conceitual que facilitasse o uso de técnicas analíticas para os educadores e gestores dos MOOCs com base em LA. Acredita-se que com diretrizes bem descritas, baseadas em experiências bem sucedidas, a tarefa de analisar os dados e implementar a aprendizagem em MOOCs se torne mais simples e com resultados positivos.

1.4 DELIMITAÇÃO E ESCOPO DA PESQUISA

O *framework* proposto nessa tese apresenta algumas delimitações que precisam ser bem explicitadas. Primeiramente, todo o trabalho foi pensado para a análise dos dados dos MOOCs. Por serem cursos massivos, que capacitam em larga escala, a BDE é mais rica para as análises propostas, uma vez que os alunos desses cursos apresentam uma heterogeneidade maior, tanto em nível de escolaridade, objetivos educacionais, localização geográfica, entre outros.

Outro ponto importante é que o *framework* será baseado em MOOCs que utilizem o Moodle como sistema de gestão de aprendizagem. Esse sistema é *open source*, ou seja, tem código aberto e é gratuito. É utilizado principalmente num contexto de *e-learning* ou *b-learning*⁵. Possui uma enorme gama de desenvolvedores que o utilizam e é altamente difundido nos ambientes de educação a distância, está disponível em 75 línguas diferentes, conta com mais de 25 mil websites registrados em mais de 175 países. Como será um *framework* baseado em evidências, todas as análises da BDE serão feitas utilizando os dados contidos no Moodle do MOOC Auriculoterapia UFSC, que desde 2015 já capacitou milhares de alunos.

Outro recorte que se faz necessário diz respeito ao público alvo do *framework*. Quando se trata de MOOCs, existem quatro principais *stakeholders*, conforme pode ser visto no Quadro 2.

Quadro 2 - Stakeholders dos MOOCs e suas características

<i>Stakeholder</i>	Objetivos, benefícios e perspectivas
Aluno	Ajuda os alunos com feedbacks personalizados, recomendações de acordo com suas necessidades com foco na melhoria de aprendizagem.
Educador	Ajuda e entender como funciona o processo de aprendizagem dos alunos, auxilia na performance dos métodos de ensino, ajuda a compreender os aspectos comportamentais, cognitivos e sociais.
Pesquisador	Ajuda a desenvolver técnicas de Mineração de dados que resolvam certos problemas, como a

⁵ A modalidade de aprendizagem híbrida, o *blended learning* ou *b-learning*, busca combinar práticas pedagógicas do ensino presencial e do ensino a distância, com o objetivo de melhorar o desempenho dos alunos em ambos os ensinos

	avaliação efetiva da aprendizagem em diferentes formas.
Administrador ou Gestor educacional	Avaliação dos recursos institucionais e ofertas educacionais.

Fonte: Romero e Ventura (2013) traduzido pelo autor.

Apesar de todos terem papéis importantes, cada qual com suas peculiaridades, somente o educador e o administrador (gestor) têm poder de decisão para implementar mudanças que realmente impactem na aprendizagem. Por isso eles serão público-alvo. O framework objetiva facilitar a gestão de dados educacionais para a tomada de decisões. Cabe aos alunos, portanto, a alimentação da base de dados, com os logs gerados através de seu acesso à plataforma.

1.5 ADERÊNCIA AO PPGEGC

A proposta da pesquisa está inserida na área de concentração de Mídia e Conhecimento do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. O presente trabalho propõe um *framework* que busca extrair conhecimentos através de *logs*, experiências e bases de dados de cursos online, uma vez que “Conhecimento é conteúdo ou processo efetivado por agentes humanos ou artificiais em atividades de geração de valor científico, econômico, social ou cultural” (Pacheco, 2014)⁶

O conceito de MOOC engloba o foco central do programa que visa disseminar e compartilhar o conhecimento. Aborda aspectos pertinentes da Educação a distância e inclusão digital, uma vez que os MOOCs proporcionam capacitação massiva de alunos. A proposta desse *framework* também busca facilitar a visualização de informação e conhecimento, ao mostrar graficamente como tirar proveito desses dados que ficam "escondidos" em bases de dados e Tabelas.

Essa preocupação em transformar dados em conhecimento, com a finalidade de melhorar o processo de aprendizagem para alunos, gestores e professores pode ser relacionado com temas da gestão do conhecimento (GC), se pensarmos num MOOC como uma organização que capacita milhares de pessoas, e que precisa constantemente de atualização e inovação. A GC na educação trabalha na retenção e transferência desses conhecimentos. A relação entre GC e *analytics* fica evidente quando percebemos uma tendência crescente no campo educacional: a "aprendizagem baseada em dados". Dessa forma, a Faculdade de Educação de Hong Kong

⁶Disponível em: <http://www.inf.ufsc.br/~tite/LODBrasil/Abertura/DadosEGovernoAbertoNaSocConh.pdf>, acessado em 03 de fevereiro de 2019.

criou recentemente o periódico *Knowledge Management & E-Learning*⁷ buscando o aprofundamento de pesquisas nessa área.

Relacionando com o campo da Engenharia do Conhecimento, a temática de *Analytics* e *Big Data* aproxima-se da área de *Data mining*, KDD & BI, já que busca extrair conhecimento de base de dados a fim de gerar vantagem competitiva e inovações educacionais:

Embora o Big Data na educação seja um fenômeno novo, com a disponibilidade de grande quantidade de dados educacionais armazenados em bancos de dados institucionais (por exemplo, dados obtidos de mídias sociais e sistemas de gerenciamento de aprendizado), a pesquisa educacional provavelmente se tornará um campo de uso intensivo de dados, utilizando métodos e técnicas da Ciência de dados, cujo objetivo principal é o desenvolvimento e uso de ferramentas, bem como extrair e discernir conhecimento valioso de dados complexos (Leek, 2013; Waller & Fawcett, 2013).

Essa aderência se mostra mais clara ao analisar o Quadro de temáticas de mestrado e doutorado do EGC (Quadro 3), no qual estão destacadas as linhas de pesquisa dessa tese:

Quadro 3 - Quadro temático de mestrado e doutorado do EGC/UFSC

Gestão do Conhecimento	Engenharia do Conhecimento	Mídia e Conhecimento
Governança corporativa	EC, metodologias e métodos	Educação a distância
Governança pública	Sistemas de conhecimento	Visualização do conhecimento
Aprendizagem Organizacional	Engenharia de ontologias	<i>Design Thinking</i>
Memória organizacional	Web semântica	Ciência cognitiva
Inteligência coletiva	Governo eletrônico e semântico	Design e usabilidade
GC & liderança	Linked data	Inovação social
GC & inovação	KDD, <i>Data mining</i> e BI	Mídias sociais
GC & empreendedorismo	Representação do conhecimento	Inclusão digital
Parques tecnológicos	<i>Grid computing</i>	Comunicação corporativa
Observatórios urbanos	Experimentação remota	Semiótica
Sustentabilidade	<i>Smart Cities</i>	Redes sociais

Fonte: EGC

Uma análise dos trabalhos publicados no banco de teses e dissertações do PPGEGC revela alguns trabalhos com temáticas correlatas, dos quais destacam-se:

⁷ <http://www.kmel-journal.org/ojs/>

Quadro 4 - Pesquisas do EGC com temáticas correlatas ao tema desta teses

Tema: pesquisas sobre MOOCs
BLEICHER, Sabrina. Processos flexíveis para a produção de materiais didáticos para a Educação a Distância: recomendações pautadas na perspectiva interdisciplinar. Tese, 2015
EBONE, Denise Santin. Avaliação e Seleção de Plataforma para Cursos Online Abertos e Massivos em Instituições de Ensino Superior. Dissertação, 2015.
BIAGIOTTI, Breno de Almeida. Avaliação da Qualidade da Informação de Cursos Massivos: Um Estudo de Caso do Telelab. Dissertação, 2016.
MENDES, Vanessa Nascimento. Análise da usabilidade e experiência do usuário: um estudo de caso de um curso da plataforma tim Tec. Dissertação, 2018.
NASCIMENTO, Heluiza Ormelez de Almeida. Acessibilidade e Usabilidade em Plataformas MOOC: Indicadores de Boas Práticas. Dissertação, 2019.
Tema: pesquisas que abordam a Aprendizagem adaptativa:
PALOMINO, Cecilia Estela Giuffra. Aplicação de um modelo adaptativo de tutores inteligentes para disseminação do conhecimento em ambientes virtuais de ensino-aprendizagem. Tese, 2017.
AMARAL, Marília A. Modelo RHA – Retroalimentação em Hipermídia Adaptativa. Tese, 2008.
OBREGON, Rosane de Fátima A. Validação de Instrumento de Identificação do Perfil de Usuário Através de Ícones Representativos das Inteligências Múltiplas. Dissertação, 2009.
Tema: pesquisa sobre Frameworks para gestão em ead:
NUNES, Carolina Schmitt. Gestão em educação a distância: um framework baseado em boas práticas. Tese, 2017.

Fonte: elaborado pelo autor

Não foram encontradas teses e dissertações que tratassem de Big Data Educacional nem trabalhos que abordassem *Learning Analytics* especificamente. Na área da Engenharia do Conhecimento foram encontradas teses e dissertações que utilizam algumas técnicas de mineração de dados e texto, porém fora do contexto dessa pesquisa.

1.6 ESTRUTURA DA TESE

Após os elementos introdutórios expostos no capítulo 1, o qual apresentam o tema desta pesquisa, o trabalho teve em sua sequência as seguintes abordagens teóricas.

No capítulo 2 foi abordado o contexto da Big Data Educacional e a importância de desenvolver a literacia de dados. Observou-se os principais tipos de dados educacionais, suas

características, categorias, cuidados em sua gestão, curadoria e reuso, além dos aspectos éticos inerentes da área.

Na sequência, no capítulo 3, tratou-se dos conceitos de *Learning Analytics*, contextualizando esse área promissora na pesquisa educacional, seu pontos fortes e fracos, seus principais objetivos e métricas, e também apresentou-se alguns frameworks de utilização dessa técnica. Além disso foram elencadas as principais técnicas analíticas desde métodos de análise estatístico a métodos utilizados na ciência de dados, fato que determina o caráter interdisciplinar e holístico dessa técnica.

No capítulo 4 foi apresentado o Moodle, um dos mais conhecidos gerenciadores de conteúdo para educação online no mercado atualmente. Foram listados seus recursos e atividades, além de seus relatórios. Evidenciou-se a complexidade de sua base de dados, composta por mais de 400 Tabelas, e também sua eficiência no registro dos *logs* de tudo o que o aluno faz dentro do ambiente virtual de aprendizagem, o que proporciona a BDE.

A seguir, no capítulo 5, são apontados os procedimentos metodológicos, divididos nas seguintes fases: definição do tema da pesquisa; revisão narrativa (fundamentação teórica); Revisão bibliográfica sistemática; Estudo de caso de um MOOC; Elaboração do *framework* (metodologia ADDIE); verificação com os *stakeholders* e por fim o refinamento do *framework* conceitual.

No capítulo 6 é apresentada a construção das 5 etapas do *framework* conceitual, com base em todo arcabouço teórico obtido nas etapas anteriores. Cada categoria é explicada individualmente até a completa proposição do *framework* conceitual

Por fim, o capítulo 7 apresenta as conclusões quanto aos objetivos da pesquisa, além de recomendações para trabalhos futuros.

2 BIG DATA EDUCACIONAL

A pesquisa na área da educação vem passando por diversos melhoramentos pedagógicos nos últimos tempos. Os ambientes de aprendizagem virtual evoluíram, assim como o papel dos alunos na aprendizagem online. Novos métodos como a sala de aula invertida, têm demandado diversas atividades baseadas na internet. O uso de *smartphones* na aprendizagem já é uma realidade e tudo isso é suportado por sistemas de gerenciamento de conteúdo, que registram todos os rastros dos usuários, dando origem a uma enorme massa de dados armazenados. Essa quantidade de dados acaba por se tornar tão grande que as técnicas tradicionais de processamento não dão conta. Essa limitação de processamento impulsionou as instituições a explorarem novas tecnologias para conseguirem processar essa BDE.

Conforme mostrado anteriormente, a BDE, se for bem explorada, proporciona um grande campo de pesquisa. O profissional que domine as habilidades de acessar os bancos de dados, extrair os dados coerentes e analisar corretamente esses dados, consegue obter informações relevantes, que auxiliam na tomada de decisão para a melhoria da aprendizagem.

Essa “literacia de dados” é reconhecida internacionalmente como um conjunto chave de competências que gera vantagens competitivas para obter melhores resultados no ensino e aprendizagem online. Para Risdale *et al.* (2015), literacia de dados educacionais compreende a habilidade de coletar, gerir, avaliar e aplicar dados de uma maneira crítica. Love (2013) define como a habilidade de observar de forma precisa, analisar e responder a uma grande variedade de dados com o propósito de melhorar continuamente o ensino e a aprendizagem na sala de aula e na escola. Corroborando com essas definições, Mandinach e Grummer (2013) definiram essa literacia como a habilidade de entender o uso efetivo dos dados para tomada de decisões, composta por habilidades específicas e conhecimento para transformar dados em informações e, por fim, em conhecimento prático.

Considerada como um novo paradigma de pesquisa, a Big Data Educacional estimula novas maneiras de elaborar e planejar pesquisas educacionais através da análise e visualização de dados (Daniel, 2015).

Segundo o MOOC Learn2Analyze⁸, esses dados educacionais podem revelar *insights*, tanto quanto o design do curso como também sobre as práticas educacionais, proporcionando

⁸ Learn2 Analyze é uma aliança de conhecimento entre academia e indústria para aprimorar as competências dos profissionais de treinamento on-line (designers instrucionais e e-instrutores) em análise de dados educacionais, que é co-financiada pela Comissão Europeia por meio do Programa Erasmus + da União Europeia. Mais informações em: <http://www.learn2analyze.eu/>

uma visão holística da aprendizagem do passado, presente e futuro. Esses dados ajudam a desenvolver uma visão aprofundada das atividades de aprendizagem, comportamentos e preferências, gerando como resultados, intervenções personalizadas que realmente ajudam os alunos a atingirem seus objetivos.

Um *framework* elaborado pela equipe do Learn2Analyse, demonstra de forma bem didática todas as variáveis e as etapas a serem percorridas para obter essa “literacia de dados”, e iniciar as pesquisas com a Big Data Educacional, conforme mostra a Figura 4:

Figura 4 - Framework de literacia de dados e suas competências



Fonte: Learn2Analyse (2019)

Esse *framework*, elaborado por professores, apresenta 6 dimensões principais e 17 afirmações, que podem ser vistos no Quadro 5 a seguir:

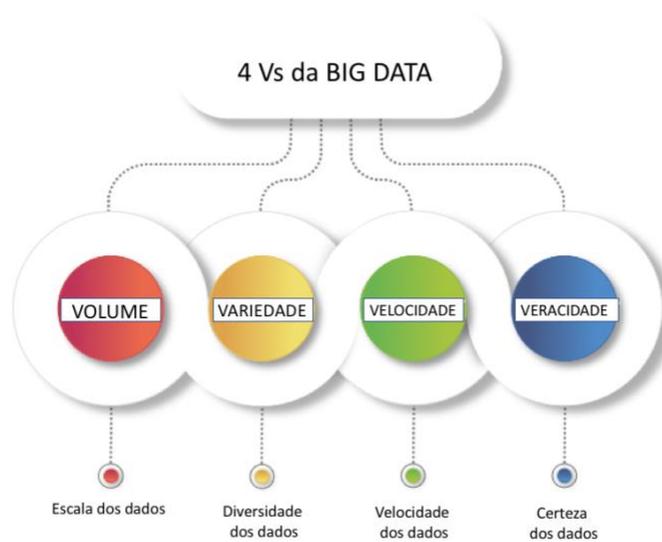
Quadro 5 - Dimensões do framework

Coleta de dados	<ul style="list-style-type: none"> ● Obter, acessar e reunir dados apropriados / fontes de dados. ● Conhecer as limitações dos dados e suas medidas de qualidade (validade, confiabilidade, acurácia, vieses nos dados...).
Gestão dos dados	<ul style="list-style-type: none"> ● Aplicar os métodos de processamento de dados (limpar, organizar e estruturar os dados). ● Aplicar a descrição dos dados (metadados). ● Fazer a curadoria dos dados. ● Usar tecnologias para preservar esses dados.
Análise dos dados	<ul style="list-style-type: none"> ● Aplicar análise de dados e métodos de modelagem (estatística descritiva, análise exploratória, mineração de dados) ● Aplicar métodos de apresentação dos dados (visualização por gráficos, mapas e outras formas didáticas)
Compreensão e interpretação dos dados	<ul style="list-style-type: none"> ● Interpretar as propriedades dos dados (desvios, erros, discrepâncias) ● Interpretar métodos estatísticos convencionais (tendência, desvio padrão, média, mediana) ● Obter insights através da análise (padrões, hipóteses, tendências) ● Processar essas informações e insights a favor da aprendizagem.
Aplicação dos dados	<ul style="list-style-type: none"> ● Tomar decisões através da análise dos dados. ● Avaliar essas decisões em relação à aprendizagem.
Ética dos dados	<ul style="list-style-type: none"> ● Usar o formulário de consentimento. ● Garantir a privacidade, a integridade, confidencialidade e a segurança dos dados. ● Aplicar autoria, propriedade, governança e compartilhamento dos dados.

2.1 CARACTERÍSTICAS DA BIG DATA

Big Data é considerado como o termo que descreve um incrível crescimento em volume, estrutura e velocidade na qual os dados estão sendo gerados. Com isso em vista, a IBM propôs um infográfico denominado “Extraindo valor de mercado dos 4 Vs da Big Data, conforme mostra a Figura 5:

Figura 5 - infográfico “Extraindo valor de mercado dos 4 Vs da Big Data



Fonte: IBM (traduzido pelo autor)

Geralmente, encontra-se na literatura pesquisas com classificações além dos três núcleos (3Vs). Há autores que utilizam o termo 4Vs e até mesmo 6Vs (Daniel, 2015):

- **Volume** - usado para descrever uma grande quantidade de informações que geralmente é um desafio para armazenar, processar, transferir, analisar e apresentar. O tamanho dos dados disponíveis tem crescido a uma taxa exponencial. Com volumes de dados mais altos, você pode ter uma visão mais holística do passado, presente e provável futuro do seu assunto.
- **Velocidade** - termo associado ao aumento da taxa na qual os fluxos de informação de uma organização (por exemplo, instituições que lidam com informações financeiras e relacionam isso aos recursos humanos e à produtividade). Os fluxos de dados são criados a uma velocidade sem precedentes. Em velocidades de dados mais altas, você pode fundamentar suas decisões em dados continuamente atualizados em tempo real.
- **Veracidade** - refere-se aos vieses, ruído e anormalidade nos dados gerados a partir de várias fontes dentro de uma organização. A veracidade também cobre questões de confiança e incerteza associada à coleta, processamento e utilização de dados. Não é apenas a precisão ou a veracidade de um conjunto de dados, mas também a confiabilidade da fonte, tipo e processamento dos dados. À medida que a veracidade

dos dados aumenta, você pode ter certeza de que está trabalhando com os dados mais verdadeiros, limpos e consistentes.

- **Variedade** - referente a dados apresentados em formato diversificado tanto estruturado quanto não-estruturado. Com variedades mais amplas de dados, você pode ter uma visão mais sutil do assunto em mão.
- **Verificação** - refere-se à corroboração e segurança de dados.
- **Valor** - refere-se à capacidade dos dados em gerar informações úteis, benefícios, processos empresariais, etc., dentro de uma instituição.

O uso de dados para tomadas de decisão, entretanto não é novidade. Geralmente os relatórios eram elaborados através de dados estruturados e organizados, coletados através de questionários e censos. A Big Data trouxe um expressivo aumento na quantidade desses dados, proporcionando um rastro digital em tudo o que fazemos no ambiente digital. Essa nova realidade nos obrigou a mudar as ferramentas e estratégias analíticas, buscando padrões, fazendo previsões e tomando decisões.

2.2 PESQUISA EDUCACIONAL X PESQUISA EM BIG DATA EDUCACIONAL

Embora a Big Data ofereça várias oportunidades para a educação, BDE e Pesquisa Educacional são duas áreas separadas de investigação, exigindo diferentes conjuntos de habilidades e conhecimentos (Quadro 5). Enquanto a pesquisa educacional está amplamente preocupada com a investigação de vários aspectos da educação, como a aprendizagem dos alunos, métodos de ensino, aprendizagem aprimorada pela tecnologia, a BDE lida com a análise de dados grandes e complexos, usando técnicas de Data Science.

Quadro 6 - Características de cada área de pesquisa

Pesquisa Educacional	Big Data Educacional
Contexto dos dados conhecidos pelos pesquisadores	Contexto de dados pode ser desconhecido para pesquisadores
O pesquisador pode estar envolvido na coleta de dados	O pesquisador pode usar dados já coletados
Epistemologia e ontologia focalizada	Epistemologia e ontologia emergente
Protocolos éticos claros e responsabilização	A responsabilidade ética pode ser desconhecida
	Precisa de conhecimento adicional de Ciência de Dados

Requer conhecimentos em educação e métodos de pesquisa	Estruturas de dados grandes e complexas (medidas em <i>Terabytes, Petabytes, Exabyte</i>)
Dados de tamanho de amostra limpos e geralmente pequenos / gerenciáveis (medidos em Megabyte, Gigabyte)	Pode empregar análise em tempo real
Não requer análise em tempo real	Sistemas de arquivos altamente distribuídos (HDFS) ou nosql
Os dados são armazenados dentro dos limites / possibilidades dos mecanismos de armazenamento disponíveis	Usa o Hadoop, sistemas mapreduce, aplicativos de mineração da web, redes de sensores, monitoramento de tráfego
A análise é manual ou usando sistemas de software independentes, como (SPSS, nvivo, STATA)	

Fonte: Daniel (2017)

Dessa forma, a BDE deriva da grande quantidade de dados educacionais armazenados em bancos de dados institucionais, obtidos através das mídias sociais ou dos gerenciadores de aprendizagem (*Learning Management Systems - LMS*) como Moodle, *blackboard*, entre outros. A pesquisa educacional, portanto, está se tornando um campo de dados intensivo, ou seja, cada vez mais utilizando técnicas e métodos da ciência de dados, campo que desenvolve o uso de ferramentas e processos para lidar com grandes quantidades de dados, a fim de extrair conhecimento válido desses dados complexos (Leek, 2013; Waller e Fawcett, 2013)

2.3 IMPORTÂNCIA DOS DADOS EDUCACIONAIS

Ao se optar pela utilização de uma abordagem analítica de dados, é recomendado sistematizar o passo a passo desde o início. Diante dos desafios inerentes à manipulação de grandes quantidades de dados, Mandinach (2012) ressaltou a importância do uso da tomada de decisão orientada por dados. Segundo o autor, realizar uma coleta sistemática, análise, exame e interpretação de dados para informar práticas e políticas em contextos educacionais, facilita na obtenção de resultados consistentes, componente essencial na prática educacional, já que fundamenta as decisões com base em dados e evidências.

Figura 6 - Sequência de tomada de decisões para análise de dados



Fonte: Learn2Analyze

O desenvolvimento de competências para essa abordagem eficaz é essencial para os profissionais da educação. Tais competências requerem "transformar efetivamente informações em conhecimentos e práticas acionáveis, coletando, analisando e interpretando todos os tipos de dados". (Ridsdale *et al.*, 2015)

Essa tomada de decisões pode apresentar diversos objetivos, como um diagnóstico individual ou sistêmico dos alunos, formulando previsões, intervenções e recomendações que resultem em melhoria da aprendizagem. Também podem ser usadas para personalização de conteúdos e até mesmo para reflexão do processo como um todo.

A análise de dados refere-se a métodos e ferramentas para analisar grandes conjuntos de diferentes tipos de dados de diversas fontes, que visam apoiar e melhorar a tomada de decisões. A análise de dados é uma tecnologia madura atualmente aplicada nos sistemas financeiro, comercial e de saúde da vida real. No entanto, segundo o *Horizon Report*, apenas recentemente é que a análise de dados foi considerada na educação - primeiro no ensino superior e mais recentemente no ensino escolar (Bienkowski, *et al.*, 2012).

No capítulo sobre *Learning Analytics* essas questões serão esmiuçadas e explicadas, baseadas em frameworks existentes na literatura

2.4 CARACTERÍSTICAS DOS DADOS EDUCACIONAIS

Os dados educacionais se originam de uma variedade de fontes e apresentam diversos formatos. Saber identificar, armazenar e utilizar esses dados é imprescindível para desenvolver competências nessa área. Os métodos de ensino tradicionais, presenciais ou online, são responsáveis por uma gama de dados, que estão armazenados em diversas fontes diferentes. Inicialmente iremos abordar os principais tipos de dados educacionais existentes e, posteriormente esses serão categorizados e divididos para uma melhor compreensão.

Na educação superior, por exemplo, muitas universidades já fazem uso de técnicas analíticas de dados para melhorar seus processos. Grandes avanços nessa área puderam ser

vistos nos últimos anos. A implementação de sistemas de informação e ambientes virtuais já é realidade na maioria das instituições. Entretanto Shacklock (2016) ressalta que os setores que lidam com esses dados ainda precisam implementar suas capacidades e procedimentos de gestão de dados para aumentar a abrangência das análises.

O autor relata a diferença entre os dados estáticos e os dados fluidos (ou dinâmicos). No primeiro caso, os dados são armazenados nas instituições e, historicamente, sempre foram utilizados em estratégias institucionais e governamentais, auxiliando na tomada de decisões e gerando vantagens competitivas. Esses dados geralmente são relacionados às atividades dos alunos que ocorrem durante o processo de aprendizagem e são coletados pelos tutores online, professores, basicamente através dos ambientes virtuais de aprendizagem.

Já os dados fluidos, tem como característica prover um panorama acertado sobre a performance dos alunos, mas são mais difíceis de captar, armazenar e processar. Esses dados surgem a cada interação do estudante, seja com seus colegas, a sua rotina no campus, ou o acesso aos ambientes virtuais de aprendizagem. Esses dados apresentam um enorme potencial para melhorar a experiência de aprendizagem dos alunos, de forma personalizada inclusive, mas demandam técnicas analíticas mais modernas e avançadas.

Para Shacklock (2016), as instituições na Inglaterra (e de forma geral) ainda não estão preparadas para utilizar o potencial completo que esses dados oferecem. Para conhecer um pouco mais sobre a diferença entre essas duas classificações de dados, observe a Figura 7 a seguir:

Figura 7 - Diferenças entre os dados educacionais estáticos e dinâmicos



Fonte: Learn2Analyze (2019). Traduzido pelo autor.

Essa categorização já nos auxilia na ideia da amplitude que os dados educacionais representam. Quanto mais completo for a captura desses dados, melhores serão as chances de proporcionarem relatórios assertivos futuramente. Além dos dados acadêmicos e financeiros, conseguir armazenar informações sobre a rotina do aluno se torna um grande desafio para as instituições. Dados de transporte (como os alunos chegam até a escola), dados alimentares (o que os alunos têm consumido), quais livros têm pego na biblioteca, dados de necessidade especiais (em caso de deficiências) podem ajudar a identificar problemas socioeconômicos que estejam afetando sua performance acadêmica, por exemplo.

Para facilitar ainda mais o entendimento sobre dados educacionais, Lai and Schildkamp (2013) criaram quatro categorias de acordo com os objetivos de cada dado, conforme pode ser visto na Figura 8:

Figura 8 - Categorias dos dados educacionais



Fonte: Lai e Schildkamp (2013, p. 11-12)

Em uma abordagem semelhante, Deloitte (2014) propôs uma categorização em 12 etapas, com uma lista abrangente de dados, desde o controle de horários, dados de pesquisa, dados administrativos até históricos e dados de ex-alunos, como pode ser visto na Figura 9:

Figura 9 - Tipos de dados educacionais na educação superior



Fonte: (Deloitte, 2014)

Os dados educacionais também podem ser categorizados como qualitativos ou quantitativos. Eles são mais efetivos quando são usados juntos, de forma complementar. Os dois tipos são igualmente importantes nos processos analíticos.

Os dados qualitativos são aqueles que não são facilmente reduzidos a números. Em muitos casos eles surgem de observações, conversas, questionários, entrevistas e documentos escritos. Esses dados são bons para localizar tendências ao longo do tempo, entender padrões e diferenças entre grupos de alunos.

Já os dados quantitativos são aqueles que se apresentam em forma de números. Podem ser feitos cálculos de média, mediana, soma etc. São úteis para explorar problemas emergentes, pois proporcionam uma descrição rica do contexto que está sendo estudado. São importantes para explicar tendências e padrões.

Diante dessa complexidade e quantidade de variáveis que englobam a Big Data Educacional, desenvolver a literacia de dados é fundamental. Segundo Pettey (2018), o pesquisador deve aprender a falar a “língua do dado”, que é igual a qualquer outra linguagem.

Começa pela compreensão dos termos básicos e descrição dos conceitos chave. Uma das formas de começar a estruturar a pesquisa em dados educacionais é saber que dados devemos coletar.

Para isso basta responder as seguintes perguntas:

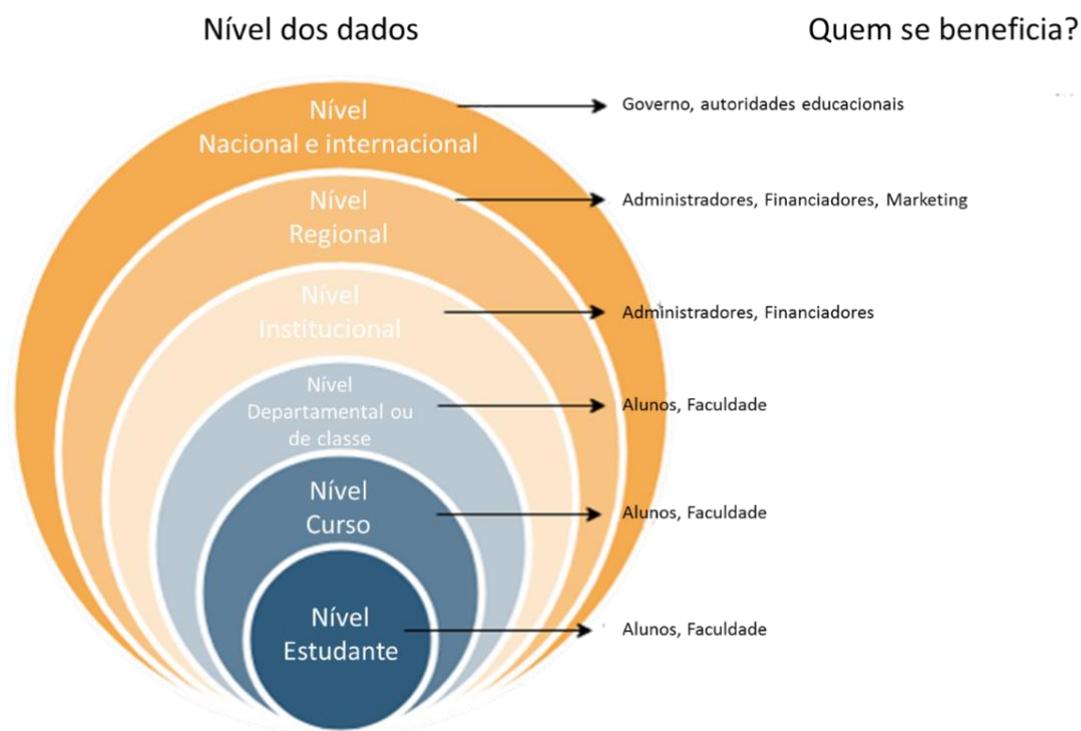
- Por quê os dados são necessários?
- Que dados são necessários?
- Quando os dados serão coletados?
- Onde esses dados estão localizados?
- Como os dados serão coletados?

Saber também quais os objetivos e estratégias serão adotadas ajuda na hora de desenvolver o processo de criação do curso. Saber qual é o seu público alvo, quais as necessidades instrucionais serão necessárias, quais são as características do ambiente virtual de aprendizagem e quais critérios serão usados para avaliar os alunos. Tudo isso já pode ser pensado e projetado de forma a facilitar o acesso posterior aos dados.

Mas nem sempre temos a oportunidade de criar um curso do zero. E mesmo com os cursos já em funcionamento, questionamentos podem ser levantados: como se dá o fluxo de aprendizagem dos alunos? Como é realizado o suporte aos alunos? Tem tutores para estimular o engajamento? A relação entre tutoria e performance acadêmica é positiva? Esses são apenas alguns questionamentos que podem surgir ao analisar esses cursos. O importante é saber que dados podem auxiliar na busca por essas respostas.

Em uma abordagem mais macro, podemos observar que uma gestão de dados educacionais bem planejada pode beneficiar desde o aluno até mesmo o próprio governo (Figura 10).

Figura 10-Visão macro dos níveis de dados educacionais



Fonte: MOOC Learn2Analyze

Tão importante quanto categorizar os dados que serão armazenados, é a organização e cuidado na gestão desses dados. Muitas vezes encontramos barreiras que dificultam o trabalho com a gestão de dados. As Figuras apresentadas anteriormente mostram um panorama dos tipos de dados que devemos procurar e armazenar, mas não transparece a dificuldade encontrada muitas vezes para acessar esse conteúdo. Obter dados de diferentes fontes, tanto internas quanto externas ao sistema escolar pode ser, em muitos casos, uma tarefa árdua. A adoção tardia de uma cultura de dados na instituição pode afetar a eficiência das intervenções analíticas.

Nesse contexto, muitos dados podem estar esquecidos em pastas de arquivos, guardadas no fundo de armários, juntando poeira. Ou até mesmo, se já forem dados digitalizados, provavelmente estão em computadores antigos, com sistemas obsoletos. Em muitos casos esses dados apresentam uma baixa qualidade, estão incompletos ou corrompidos. O trabalho de recuperação desse material pode se tornar uma tarefa que consome muito tempo e recursos humanos.

2.5 AONDE PROCURAR POR DADOS EDUCACIONAIS?

Como foi abordado anteriormente, os dados educacionais se encontram espalhados nas mais diferentes fontes e apresentam os mais diferentes formatos. Mas com um bom planejamento e meta, sabendo-se quais dados realmente são necessários na pesquisa, podemos identificar e localizar os dados mais comumente utilizados.

Uma das principais fontes de dados educacionais são os sistemas de informação da instituição: geralmente esses dados se encontram de forma digitalizada e disponível. Nessa categoria se incluem as notas e avaliações dos alunos, dados socioeconômicos, dados sobre cidadania e status de imigração, habilidades prévias, domínio de idiomas, dados financeiros, histórico escolar etc..

Outra fonte repleta de dados educacionais se encontra nos softwares de gestão da aprendizagem (LMS). Esses dados são originários dos registros que os alunos deixam ao utilizar essa plataforma durante suas atividades online.

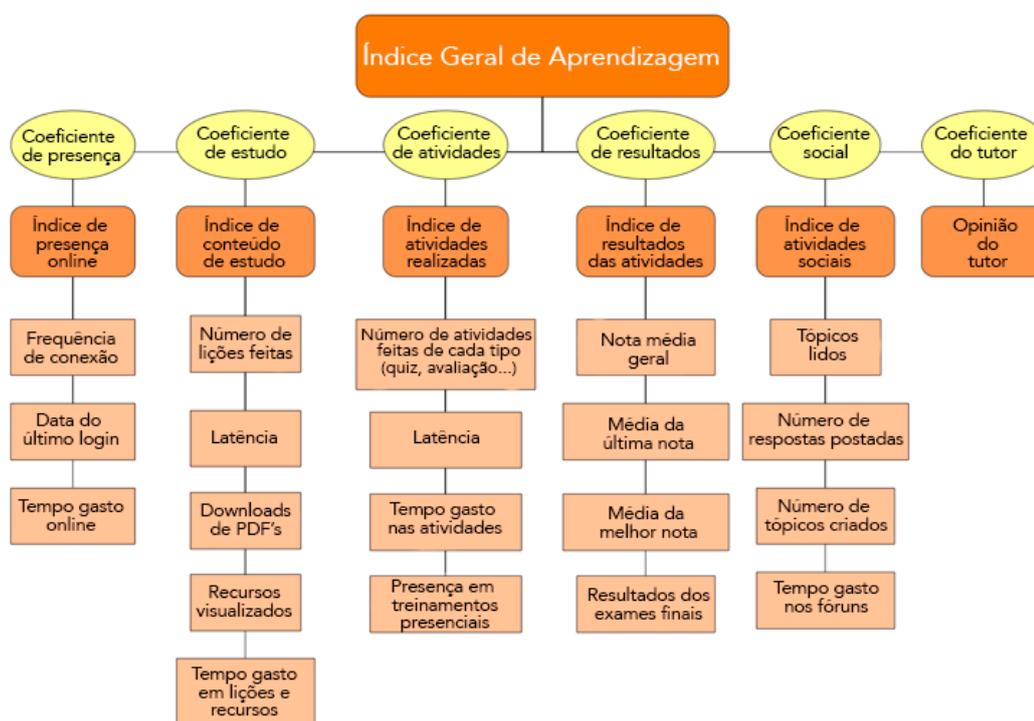
Os rastros de aprendizado podem ser divididos em duas categorias: métricas que registram os resultados e métricas que registram comportamentos (que podem ou não levar a resultados).

As métricas de resultados podem conter dados com pontuações de competência avaliadas por humanos, como notas para tarefas ou exames, pontuações de competência avaliadas por computador para trabalhos em andamento, avaliações de metas de aprendizagem, pesquisas de eficácia de educadores, pesquisas de eficácia de cursos, notas para participação, notas para presença e outras métricas.

Já as Métricas comportamentais podem incluir dados como duração da pausa na digitação, velocidade de digitação, velocidade de rolagem em uma página da *web*, movimentos oculares, calor corporal ou outras informações biométricas, frequência diária, atenção, perguntas feitas, linguagem usada em fóruns ou em sala de aula, tempo e duração da estadia em sites de cursos ou recursos online, uso de recursos de laboratório entre outros.

Bovo *et al.* (2013) propôs uma outra classificação dos dados, categorizados em 6 principais coeficientes: Coeficiente de presença; Coeficiente de estudo; Coeficiente de atividades; Coeficiente de resultados; Coeficiente Social e coeficiente de tutoria. A Figura 12 mostra essas categorias com mais detalhes:

Figura 11 - categorização dos dados educacionais



Fonte: Bovo *et al.* (2013) Traduzido pelo autor.

De forma geral, os gerenciadores de conteúdos educacionais fornecem os seguintes dados em seus sistemas:

Quadro 7 - Lista de dados fornecidos pelo LMS Moodle

Número total de cliques	Número de sessões online
Tempo total online	Número de páginas visitadas
Irregularidade do tempo de estudo	Irregularidade do intervalo de estudo
Maior período de inatividade	Tempo desde a primeira atividade
Média de tempo por sessão	Número de recursos vistos
Número de links clicados	Número do conteúdo de páginas vistas
Número de visualizações de discussões em fóruns	Número de postagem em fóruns
Número de inícios em avaliações	Número de tentativas por avaliação
Número de avaliações aprovadas	Número de avaliações submetidas
Número de edições em wiki	Número de visualizações wiki
Média de notas das avaliações	

Fonte: elaborado pelo autor

O Moodle oferece diversos plug-ins que auxiliam na coleta de dados educacionais. O Quadro 7 a seguir sumariza algumas das principais alternativas que podem ser instaladas no LMS:

Quadro 8 - Lista de plugins úteis na análise de dados educacionais

Plugins	Descrição
GISMO	Sistema de rastreamento interativo criado para o Moodle que exibe dados através de uma interface gráfica
SNAPP	Ferramenta da Web que fornece análise de redes sociais
Meerkat-ED	Ferramenta da Web que analisa participantes e suas interações em fóruns de discussão
MOC Log	É uma soma de ferramentas usadas para a análise e apresentação de dados no Moodle
Excel Pivot <i>Tables</i>	É uma ferramenta que pode ser usada para a produção de estatísticas de aprendizado provenientes do Moodle
<i>Analytics and Recommendations</i>	É uma ferramenta para a visualização do envolvimento dos alunos em cada atividade de um curso on-line, bem como uma ferramenta de consulta que pode recomendar atividades aos alunos para que eles possam melhorar seu desempenho.
Desire2Learn	Plataforma integrada de aprendizado que aborda desafios com engajamento, retenção e resultados de aprendizado

Fonte: elaborado pelo autor

Também podemos obter dados relevantes através de sistemas que analisam a discussão online em fóruns e redes sociais ou mesmo pela aplicação de questionários ao longo do curso. Uma boa alternativa é utilizar um questionário pré-curso e outro ao final, podendo, dessa forma, comparar as expectativas dos alunos com suas opiniões pós-curso. A Figura 13 a seguir sumariza as principais fontes de dados disponíveis em um curso:

Figura 12 - Principais fontes de dados educacionais



Fonte: Learn2Analyze

2.6 CURADORIA DE DADOS EDUCACIONAIS

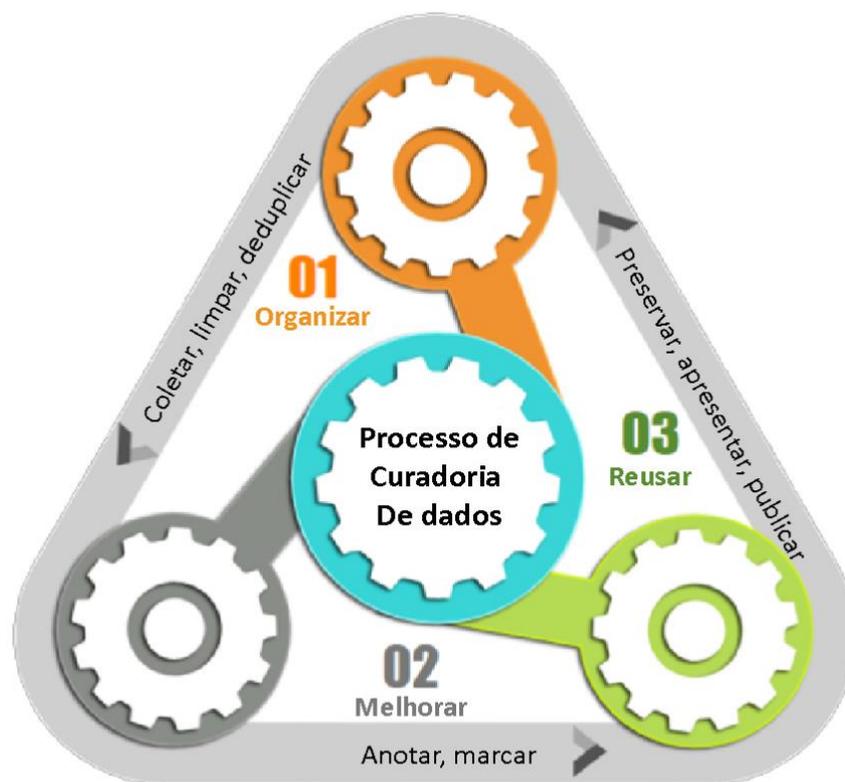
Em todo processo de análise de dados é preciso uma certa organização para facilitar o trabalho com os dados. A literatura ressalta a importância da curadoria desses dados, tanto no momento de prepara-los para análise, quanto no momento de armazenar e preservar esses dados. Sabemos que, ao trabalhar com a BDE alguns desafios precisam ser levados em consideração.

- **Armazenamento** - Embora tenha áreas que a quantidade de dados supera a casa dos *Exabytes*, a BDE ainda não chegou nesse patamar. Apesar de ser um campo em expansão, a preocupação com o armazenamento deve demandar uma atenção especial. Preferivelmente os dados devem ser armazenados em bancos de dados que sejam facilmente consultados. Além disso é recomendável utilizar tecnologias de compressão para dados em repouso.
- **Análise** - uma das dificuldades encontradas ao trabalhar com a BDE é a natureza dos dados que podem ser estruturados ou não-estruturados. Como existem diversos sistemas de aprendizagem, com vários tipos de dados, essa tarefa na análise pode consumir muito tempo e recursos. Dividir os dados em partes menores e utilizar computadores em rede para processar pode ser uma alternativa.
- **Relatórios** - Devido à grande quantidade de dados envolvida, novas formas de apresentar os relatórios devem ser pensadas. Somente mostrar dados estatísticos em

Tabelas, na forma de números, não facilitam a leitura. Técnicas de visualização, *dashboards* e gráficos interativos são as melhores interfaces nesses casos.

Para organizar esse processo, Stonebracker (2014) definiu o termo curadoria de dados como sendo o processo de transformar fontes de dados criadas de forma independente (dados estruturados e semiestruturados) em conjuntos de dados unificados prontos para análise, usando especialistas em domínio para orientar o processo. Esse processo passa pela identificação das fontes de dados, verificação, limpeza, transformação, integração e deduplicação (remoção de dados duplicados). A Figura 11 mostra em detalhes essas etapas:

Figura 13 - Ciclo de curadoria de dados



Fonte: Learn2Analyze

2.6.1 ETAPA 1 - ORGANIZAÇÃO DOS DADOS

Como mencionado anteriormente, os dados educacionais aparecem nos mais variados formatos e fontes. Podem se apresentar como dados dos sistemas de gerenciamento de aprendizagem (Moodle, Blackboard, etc...), podem ser questionários, testes, dados demográficos, entre outros. Para traçar panoramas assertivos acerca dos processos de aprendizagem, há necessidade de juntar esses dados para conseguir obter o potencial que eles proporcionam. E é justamente nessa etapa que começam as dificuldades.

Esses dados apresentam formatos variados, inconsistências, discrepâncias, dados faltantes e/ou duplicados. Para usar esse material há necessidade de um processo de limpeza de dados. Nessa etapa são identificados esses erros, os quais são descartados. Romero e Ventura (2014) sugerem algumas abordagens (existem muitos métodos de imputação) para lidar com os dados faltantes:

- Usar “nulo” ou “?” Nos campos que faltam;
- Usar um valor substituto como a média ou a moda;
- Para determinar o valor provável que falta, usar métodos de regressão;
- Em alguns casos mais extremos, remover esses dados da análise.

Outro acontecimento comum são os valores considerados “fora da curva”. Isso ocorre quando um valor se destaca dos demais, muito para cima ou para baixo. Pode ser causado por erros de mensuração ou até mesmo erros de digitação. Nestes casos de dados com possíveis “*outliers*”, existem métodos estatísticos robustos que podem ajudar. Entretanto, esses valores que se destacam podem sim estar corretos e indicar um fenômeno de interesse. O pesquisador deve estar atento a esse tipo de fenômeno. Um exemplo desse tipo de erro pode ser exemplificado no tempo que o aluno passa logado no ambiente virtual. Digamos que a média de tempo da turma é de 15hs, enquanto determinado aluno ficou 60hs. Isso pode indicar que o aluno realmente passou esse tempo estudando, mas é um forte indício também que ele pode ter se esquecido o ambiente virtual aberto e não realizou o *logoff*. Nesses casos é necessário um conhecimento do contexto que está sendo estudado, para refutar ou não esses dados discrepantes.

Dados inconsistentes também são comuns nessa etapa de limpeza. Eles aparecem quando um grupo de dados é muito diferente de outro grupo sem nenhuma razão aparente.

Como exemplo, podemos citar dados de datas em formatos diferentes (padrão americano e europeu), ou dados com padrões métricos diferentes (metros e polegadas). Muitas vezes podem surgir dados abreviados, em letras maiúsculas, que geram inconsistências que afetam análises futuras. Nesses casos é preciso estipular uma padronização de formato.

Outro cuidado a ser tomado é localizar instâncias duplicadas nos dados. Em alguns casos, uma variável pode estar presente mais de uma vez, já que os dados são coletados de diversas fontes. A esse processo de localizar e eliminar dados em duplicidade se dá o nome de deduplicação.

Esse processo de limpeza e organização de dados é uma das etapas que mais consomem tempo dos cientistas de dados. Fazer a padronização das bases de dados e conferir valores é uma etapa imprescindível para, posteriormente, obter insights desses dados.

2.6.2 ETAPA 2 - REUSO DOS DADOS

Após o processo de limpeza e categorização dos dados, chega o momento de botar em prática os cuidados com a preservação e apresentação desses dados. Diversas atitudes podem ser tomadas para garantir a segurança e o rápido acesso ao material quando necessário. Nessa hora, alguns questionamentos costumam surgir:

- Armazenar em um servidor interno ou externo?
- Quais dados armazenar?
- Por quanto tempo os dados ficarão armazenados?
- Qual o formato que os dados serão armazenados?
- Existe possibilidade de degradação dos dados (físicos e digitais)?
- Quais são os custos para armazenar esse material?
- Quais profissionais serão necessários para lidar com esse material?
- Qual a capacidade de armazenamento será preciso?

Essas são questões comuns enfrentadas por profissionais que trabalham com grandes quantidades de dados. Seguindo algumas simples orientações propostas no manual sobre preservação digital⁹, os riscos relacionados à preservação de dados se tornam bem reduzidos.

“A preservação não é mais apenas uma preocupação para a memória das instituições a longo prazo, mas para todos os interessados em usar e acessar materiais digitais. Quanto maior a importância dos materiais digitais, maior a necessidade de sua preservação: a preservação digital protege o investimento, captura potencial e transmite oportunidades para as gerações futuras e para a nossa. Os materiais digitais - e as oportunidades que eles criam - são frágeis ”(Manual de Preservação Digital, 2015)

2.7 CUIDADOS COM A GESTÃO DOS DADOS

Um dos principais cuidados ao manipular dados educacionais é a questão da ética e do sigilo. Há sempre a responsabilidade de manter a privacidade e a segurança dos dados dos alunos, de tal forma que eles não sejam identificáveis. Para isso, o *Digital Preservation Handbook* (2015) listou algumas técnicas para manter a segurança da informação:

- **Criptografia:** é uma técnica que protege o material digital, convertendo em um formato “embaralhado”. Para utilizar esses dados é necessário convertê-los ao seu formato original, através de uma chave.
- **Controle de acesso:** um administrador especifica quem pode acessar o material, assim como o tipo de acesso (somente leitura, ou gravação).
- **De-identificação:** remoção ou substituição de informações confidenciais, usando anonimização ou pseudonimização.

De acordo com a *General Data Protection Regulation* (GDPR), lei europeia que entrou em vigor em maio de 2018 e com a lei Geral de Proteção de Dados Pessoais (LGPD) nº13.709/2018, lei brasileira, devemos entender que, no quesito privacidade, existem diversas categorias de dados:

- **Dados Pessoais:** Nome; RG; CPF; atributos físicos, psicológicos, genético, mental, econômico, cultural ou de identidade social.

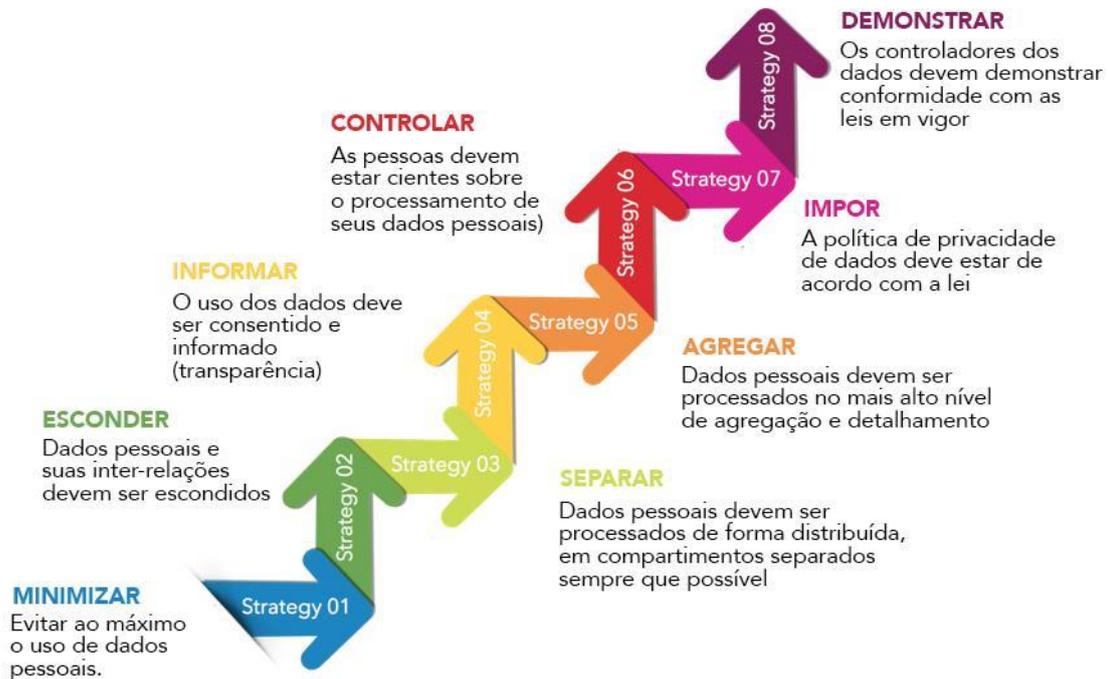
⁹ Disponível em <https://www.dpconline.org/handbook/getting-started>

- **Informações anônimas:** Informações que não se relacionam e não identificam o indivíduo. Podem ser os dados pessoais, trabalhados de uma maneira que não os tornem identificáveis.
- **Dados sensíveis:** dados pessoais que revelam a raça, etnia, crenças religiosas e filosóficas. Dados genéticos e biométricos que identifiquem o ser humano; Dados de saúde; Dados relativos aos hábitos e orientação sexual.
- **Dados “mascarados”:** dados processados de tal forma que não podem ser atribuídos à sua origem sem o uso de uma informação adicional; Dados criptografados.

Ainda segundo as leis de proteção de dados, estes "só devem ser processados quando não for razoavelmente possível realizar o processamento de outra maneira. Sempre que possível, é preferível usar dados anônimos. Onde os dados pessoais são necessários, eles devem ser adequados, relevantes e limitados ao necessário para a finalidade ('minimização de dados')." E se tratando de dados educacionais, todo o cuidado na gestão se faz extremamente relevante e necessário.

Embora existam diversas formas de “maquiar” esses dados, alguns autores garantem que não existe uma técnica 100% eficaz. Em um grau maior ou menor de dificuldade, experts podem reverter esse processo, cruzando esses dados com outras fontes de informações. Tendo isso em vista, todo cuidado deve ser tomado nesse processo de anonimização. A Figura 14 mostra as diretrizes da Agência de Segurança da Rede e de Informações da União Europeia (ENISA)

Figura 14 - Diretrizes elaboradas pela ENISA.



Fonte: D' Acquisto *et al.*. (2015). Traduzido pelo autor.

2.8 QUALIDADE DOS DADOS

Como toda atividade com uso intensivo de dados, os cuidados com a qualidade devem ser tomados, uma vez que influenciam diretamente no resultado esperado. Dados pobres, inconsistentes, refletem isso nos resultados das análises, e o inverso também se aplica. No caso dos dados educacionais, como visto anteriormente, a heterogeneidade de formatos (dados não estruturados), as múltiplas fontes e o grande volume são características que tendem a dificultar o processo de LA.

Tendo isso em vista, Mihăiloaie (2015) elencou alguns aspectos que devem ser observados durante a manipulação de dados:

- **Completude:** ocorre quando não existem “buracos” nos dados coletados, ou seja, não há diferença entre o que você espera coletar e o que realmente coletou. Sem valores faltantes, uma base de dados pode ser considerada completa.

- **Consistência:** ocorre quando os tipos de dados estão alinhados e compatíveis com o esperado. Não apresentam contradições e não violam nenhuma restrição de integridade.
- **Precisão:** também chamada de acurácia, ocorre quando a coleta de dados representa o que deveria com exatidão e rigor.
- **Oportunidade:** os dados devem ser recebidos no tempo esperado para que as informações sejam utilizadas com eficiência.
- **Validade:** ocorre quando determinado dado se mostra bem fundamentado e corresponde exatamente ao mundo real.
- **Exclusividade:** ocorre quando não há duplicidade de dados na base.

A falta dessas dimensões nos dados coletados afeta diretamente os resultados. No caso do LA, dados imprecisos levam à elaboração de métricas sem qualidade que, por fim, resultam em ações que não atingem o objetivo principal de todo o trabalho: melhorar a qualidade da aprendizagem.

Mas a qualidade dos dados não se refere somente aos aspectos técnicos. A forma como conduzimos nosso relacionamento com os dados também importa. Os aspectos éticos também devem ser considerados, juntamente como indicadores de qualidade. Ter cuidado com as questões de privacidade, consentimento, propriedade e transparência refletem na qualidade da pesquisa em si, conforme mostra a Figura 15:

Figura 15 - Aspectos éticos e qualitativos dos dados educacionais



Fonte: Learn2Analyse (2019)

As principais considerações éticas citadas na Figura referem-se à:

- **Privacidade:** o modo como os dados pessoais estão sendo informados e distribuídos, devendo ser protegidos contra possíveis invasões e devidamente anonimizados (de-identificados).
- **Propriedade:** o ato de ter direitos legais e completo controle sobre os dados disponíveis, através de uma política de aquisição e distribuição previamente definida.
- **Consentimento:** documentação que descreva os processos de coleta e análise dos dados, com as devidas explicações sobre o uso dos mesmos.
- **Transparência:** deixar bem claro a finalidade e o objetivo do uso dos dados, de forma que mantenha a confidencialidade e a integridade dos personagens envolvidos.

2.9 DIREITOS INDIVIDUAIS DOS QUE CEDERAM OS DADOS

O uso de dados e informações devem ser consentidos, conforme comentamos anteriormente. Entretanto, caso o cedente dos dados mude de ideia, ele pode reivindicar os seus direitos relativos ao uso de seus dados. Na União Europeia essa questão tem sido levada a sério, dando oportunidade aos cidadãos controlarem melhor o uso de seus dados pessoais, além de propor diretrizes para as organizações que farão uso desses dados. Alguns desses direitos são:

- **Direito de ser esquecido:** solicitar para que seus dados sejam apagados após o processamento ou quando não forem mais necessários;
- **Direito de portabilidade:** solicitar seus dados em formatos comuns (lidos por computadores) ou enviá-los para outro controlador;
- **Direito de restringir o processamento:** solicitar que seus dados não sejam mais utilizados em determinados casos;
- **Direito de acesso:** saber de seus dados pessoais estão sendo utilizados;
- **Direito de retificação:** solicitar correções em caso de dados incompletos ou errados;
- **Direito de objeção:** optar por não ceder os dados para finalidade de marketing ou situações que não te agradam;
- **Direitos relacionados ao processamento computadorizado de dados:** direito de contestar os resultados obtidos através de processamento computadorizado de dados;

Esse último tópico é bem pertinente em relação ao trabalho desenvolvido nessa tese. A utilização de *Learning Analytics* envolve o processamento de dados tanto por humanos quanto por máquinas. As principais técnicas avançadas de ciência de dados são realizadas através de algoritmos previamente definidos, que tem a capacidade de processar uma enorme massa de dados, o que seria humanamente impossível.

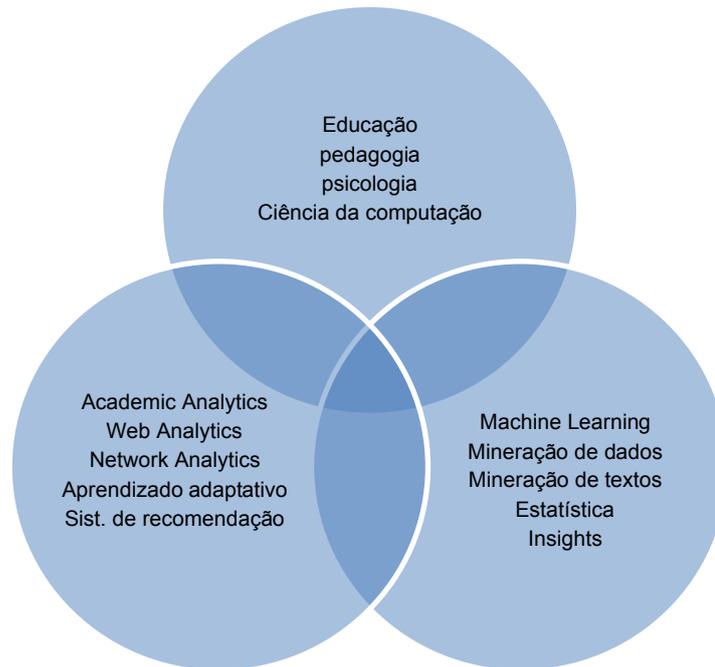
3 *LEARNING ANALYTICS*

LA é um domínio que está no cruzamento entre análise de Big Data e MOOCs. Embora o termo *Learning Analytics* seja relativamente novo, vale mencionar que a pesquisa nessa área começou muito antes. Em 2009, Ryan S.J.D. Baker e Kalina Yacef apresentaram uma revisão de última geração no campo da Mineração de Dados Educacionais (MDE). Esse trabalho mostrou que o LA tem suas origens na área de MDE, que usa métodos para mineração de dados, estatística e visualização e mineração da *Web* (clusterização, classificação, padrões de mineração, mineração de texto). Em 2005, essas técnicas foram abrigadas sob o termo LA, definido pela *Society for Learning Analytics Research* (solar, 2011) da seguinte forma: “*Learning Analytics* é a medição, coleta, análise e relato de dados sobre aprendizagem e seus contextos, para fins de compreender e otimizar a aprendizagem e o ambiente em que ela ocorre.” (Fernández e Mariño 2015).

Segundo Baker e Siemens (2009), os pesquisadores de MDE focam mais nos métodos computadorizados que buscam por padrões nos dados, enquanto os pesquisadores de LA estão mais interessados nos aspectos humanos inerentes aos dados educacionais. Pesquisadores em MDE buscam formas de adaptação automatizada, na qual um sistema identifica as necessidades dos alunos e automaticamente propõe soluções. Do outro lado, pesquisadores de LA buscam formas de informar e empoderar alunos, instrutores, professores sobre novas formas de tratar as questões educacionais, elaborando uma visão global e holística como um todo.

LA não é uma nova área genuína de pesquisa. Ela reflete um campo que intercede inúmeras disciplinas acadêmicas (por exemplo, ciência da aprendizagem, pedagogia, psicologia, ciência da Web, ciências da computação) (Dawson *et al.*, 2014). E empresta conceitos de uma variedade de campos relacionados (análise acadêmica, análise de ações, mineração de dados, sistemas de recomendação, aprendizado adaptativo personalizado) e sintetiza várias técnicas existentes (por exemplo, *machine learning (ML)*, mineração de dados, regressão, estatística e visualização) (Chatti *et al.*, 2014; Ferguson, 2012).

Figura 16 - Interseção de áreas de conhecimento que compõe o LA.



Fonte: Elaborado pelo autor.

LA existe como um subconjunto do campo *Analytics*, focada especificamente nos campos da educação e da aprendizagem. Embora existam propriedades compartilhadas entre as técnicas, LA tem suas próprias características únicas que merecem atenção especial. (Chatti, *et al.*, 2012). Preocupa-se com a medição, coleta, análise e relato de dados sobre alunos e seus contextos, para fins de compreensão e otimização do aprendizado nos ambientes em que ela ocorre (Siemens e Long, 2011). LA é realizada mais a nível de ensino e aprendizagem institucional e está amplamente preocupada em melhorar o sucesso (Jones, 2012).

Analytics é um novo ramo que vem ganhando interesse, pois tenta coletar dados valiosos, examinar informações de maneira inteligente e descobrir conhecimento útil para apoiar a tomada de decisão. Existem algumas variações nessa área:

- **Web analytics:** “a coleta, análise e relatório do uso do site pelos visitantes e clientes para entender a eficácia do online...” (mcfadden, 2015).
- **Data Analytics:** a aplicação sistemas computacionais para a análise de grandes conjuntos de dados para o apoio na tomada de decisões.

- **Big Data Analytics:** tem o objetivo analisar conjuntos de dados em massa, que ocorrem cada vez mais em negócios na web, distribuindo o processamento e utilizando maciços sistemas paralelos de análise (Saecker e Markl 2013);
- **Business Analytics:** é a prática de analisar dados quantitativos para suportar a tomada de decisão, baseado em uma gama de métodos de análise de dados, incluindo visualização de dados e compreensão de relatórios (Shmueli *et al.*, 2016).

O Quadro 8 a seguir mostra alguns domínios relacionados no campo do Analytics:

Quadro 9 - Analytics e suas variações

	Domínio	Dados	Objetivo
<i>Web Analytics</i>	Uso para web	Sistemas operacionais	Melhoria do website
<i>Visual Analytics</i>	Tomada de decisões	Tudo	Raciocínio analítico
<i>Business Analytics</i>	Negócios	Sistemas operacionais	Melhoria de performance
<i>Educational data mining</i>	Educação	Sistemas operacionais	Detecção de padrões
<i>Academic analytics</i>	Educação	Sistemas de gerenciamento da aprendizagem (LMS)	Performance organizacional
<i>Action analytics</i>	Educação	Sistemas de gerenciamento da aprendizagem (LMS)	Performance organizacional
<i>Learning analytics</i>	Aprendizagem	Dados de usuários, sistemas operacionais	Performance da aprendizagem

Fonte: Wild (2016). Traduzido pelo autor.

Ainda existem outros campos como *Ubiquitous Analytics* (Elmqvist e Irani 2013), *Social Network Data Analytics* (Aggarwal, 2011), *predictive analytics* (Siegel 2013), *Advanced Analytics* (Ryza et al. 2015) e *Text analytics* (Xiang 2015).

Entretanto, em todos esses processos citados anteriormente, podemos encontrar os quatro pilares essenciais que norteiam essas práticas: coleta de dados; análise; reporte dos dados e tomada de decisões (ações). Essas quatro etapas estão presentes em todos os *frameworks* que veremos daqui para frente e fazem parte do caminho percorrido por pesquisadores que utilizam o LA.

O dado é o ativo primário, matéria prima que será transformada em informação útil. Como vimos no capítulo anterior, quando falamos de dados no contexto educacional, nos

referimos aos processos de aprendizagem, interações dos alunos, notas, desempenho acadêmico, entre outros.

A análise se dá após a coleta e por meio das métricas e objetivos propostos, os dados são “lapidados” (por meio de técnicas computadorizadas, algoritmos, estatística). Desse processo são aplicadas técnicas de limpeza e modelagem nesses dados com objetivo de extrair informações relevantes que nos auxiliem na tomada de decisões e ações que melhorem o processo de ensino e aprendizagem.

A próxima etapa é a apresentação dos resultados da análise e a confecção do relatório. Para isso é necessário sumarizar as informações encontradas de forma apropriada e didática. Definir quando é melhor utilizar gráficos ou Tabelas, escolher formas que facilitem a visualização da informação para auxiliar a próxima etapa: a tomada de decisões.

Todo processo de LA deve culminar em ações. Nessa etapa os professores e gestores devem aplicar intervenções no processo educacional. Pode ser uma alteração no modo como se dá as aulas online, ou a mudança de um método de avaliação que não foi bem aceito por determinada turma. O resultado dessas intervenções é o que determinará se o processo de LA tinha um objetivo claro e se foi executado com sucesso. E por ser um processo cíclico, essas etapas estão em constante evolução.

3.1 OBJETIVOS DO LA

A utilização do LA pode ser útil em diversos contextos educacionais, tanto em aulas presenciais, *blended learning* e em aulas a distância. Pode ser utilizado em todos os níveis, do ensino médio à pós-graduação. Uma vez entendendo como funciona e se aplica essas técnicas, o uso pode ser customizado para atender os mais diversos objetivos.

Em uma sala de aula presencial, por exemplo, nem sempre é fácil identificar as virtudes e fraquezas de determinados alunos, suas deficiências e necessidades. Dessa forma, desenvolver ações que favoreçam um ensino personalizado para determinados alunos, com recomendação de tarefas personalizadas ou tratamento especial pode resultar em uma tarefa extremamente difícil. Ainda mais se levar em conta que essa é apenas uma turma entre tantas outras que o professor é responsável.

No ambiente online, por exemplo, é muito comum que os alunos não concluam suas atividades e as taxas de evasão costumam ser altas. Para o tutor é muito difícil identificar as

emoções dos alunos, suas habilidades e disponibilidade de socializar. Como prever se determinado aluno irá concluir determinada tarefa? Como mantê-lo engajado até o final?

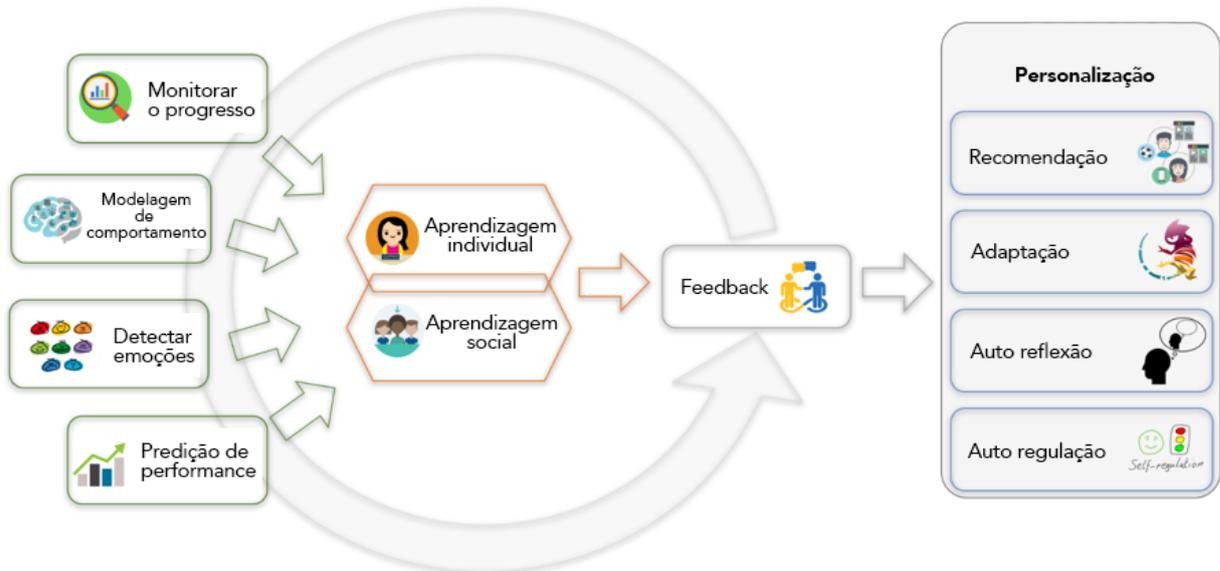
E em um curso que utiliza o formato *blended learning*? O aluno pode ter dificuldade de desenvolver sua própria aprendizagem e acaba por procrastinar demais. Como o professor pode monitorar o progresso desse aluno e ajudá-lo com *feedbacks* assertivos?

As técnicas de LA podem ser usadas de várias maneiras na análise de aprendizado conforme lista elaborada por Charlton *et al.* (2013):

- **Previsão de desempenho** - O desempenho do aluno pode ser previsto analisando a interação do aluno em um ambiente de aprendizado com outros alunos e professores.
- **Deteção de risco de evasão** - Ao analisar o comportamento do aluno, o risco de alunos abandonarem os cursos pode ser detectado e medidas podem ser implementadas no início do curso para manter os alunos.
- **Visualização de dados** - Os relatórios sobre dados educacionais tornam-se cada vez mais complexos à medida que esses dados crescem em tamanho. Os dados podem ser visualizados usando técnicas de visualização para identificar facilmente as tendências e relações nos dados apenas observando os relatórios visuais.
- **Feedback inteligente** - Os sistemas de aprendizagem podem fornecer feedback inteligente e imediato aos alunos em resposta às suas contribuições, o que melhora a interação e o desempenho dos alunos.
- **Recomendação de cursos** - Novos cursos podem ser recomendados aos alunos com base nos interesses dos alunos identificados através da análise de suas atividades. Isso garantirá que os alunos não sejam equivocados na escolha de campos nos quais possam não ter interesse.
- **Estimativa de habilidade do aluno** - Estimativa das habilidades adquiridas pelo aluno durante o curso.
- **Deteção de comportamentos** - Deteção de comportamentos de alunos em atividades ou jogos baseados na comunidade que ajudam no desenvolvimento de um modelo de estudante.

A Figura 17 abaixo ajuda a entender como a aplicação do LA pode colaborar na rotina educacional:

Figura 17 - Principais objetivos do LA



Fonte: Learn2Analyze (2019)

Conhecer o potencial que o LA pode proporcionar é o primeiro passo para sua utilização. Ter bem claro os objetivos os quais se pretende melhorar é determinante para o sucesso. Desenvolver estratégias baseadas em dados tendem a trazer resultados mais precisos e assertivos. Conhecer bem a rotina educacional e os recursos que são utilizados é uma forma de elencar quais dados serão prioritários na análise. Por isso, desenvolver a literacia de dados (explicada no capítulo anterior) se torna uma ferramenta extremamente útil para os *stakeholders* envolvidos no processo educacional como um todo.

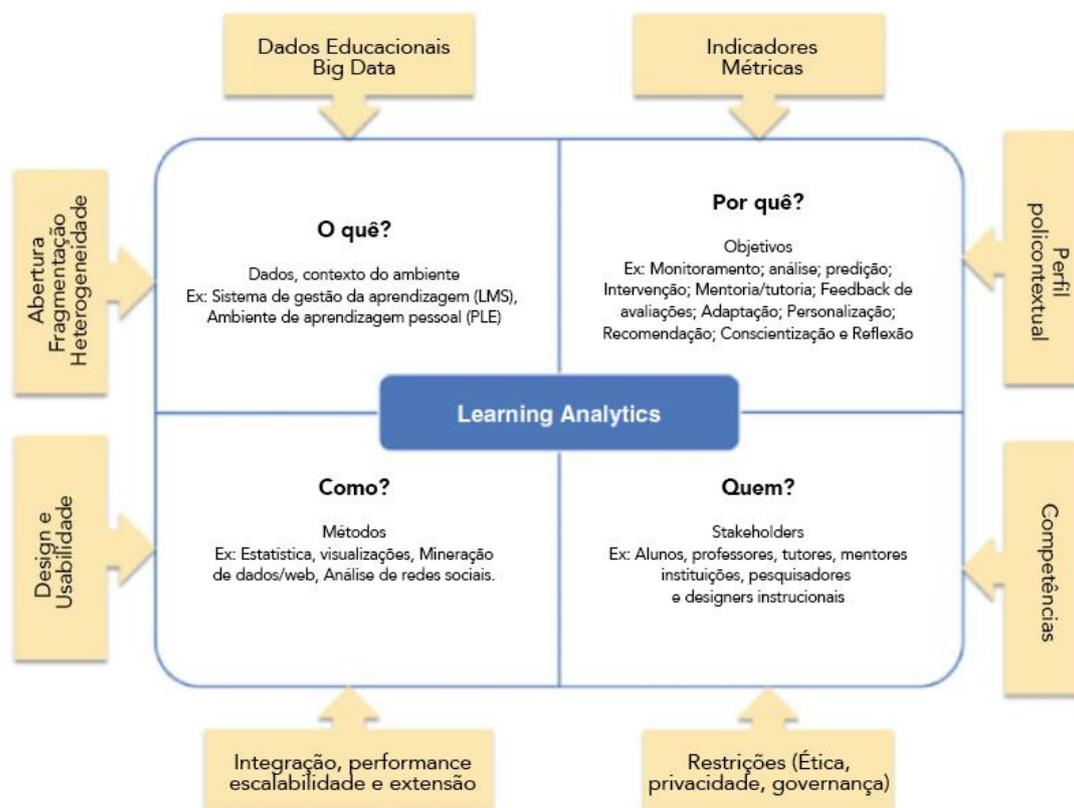
A seguir serão apresentados alguns *frameworks* de LA que orientam sua aplicação, de acordo com seus objetivos.

3.2 FRAMEWORKS EXISTENTES

A introdução generalizada de sistemas de gestão de aprendizagem (LMS), tais como Blackboard e Moodle resultaram em conjuntos cada vez maiores de dados. Todos os dias, esses sistemas acumulam quantidades crescentes de dados de interação dos alunos, dados pessoais,

informações do sistema e informações acadêmicas (Romero et al., 2008). OS LMS mantêm o registro de "ações-chave" dos alunos. Esses dados também podem incluir informações mais detalhadas sobre o conteúdo das publicações, suas escolhas e progresso através de tarefas interativas, suas preferências particulares e hábitos etc... (Friesen, 2013). Usando LA, essas informações podem ser usadas para entender o comportamento, o ambiente de aprendizagem, a eficácia do ensino, bem como o ambiente em que ele é realizado. Para compreender melhor esse contexto, Chatti *et al.* (2014) elaboraram um modelo conceitual:

Figura 18 - Modelo referencial de LA



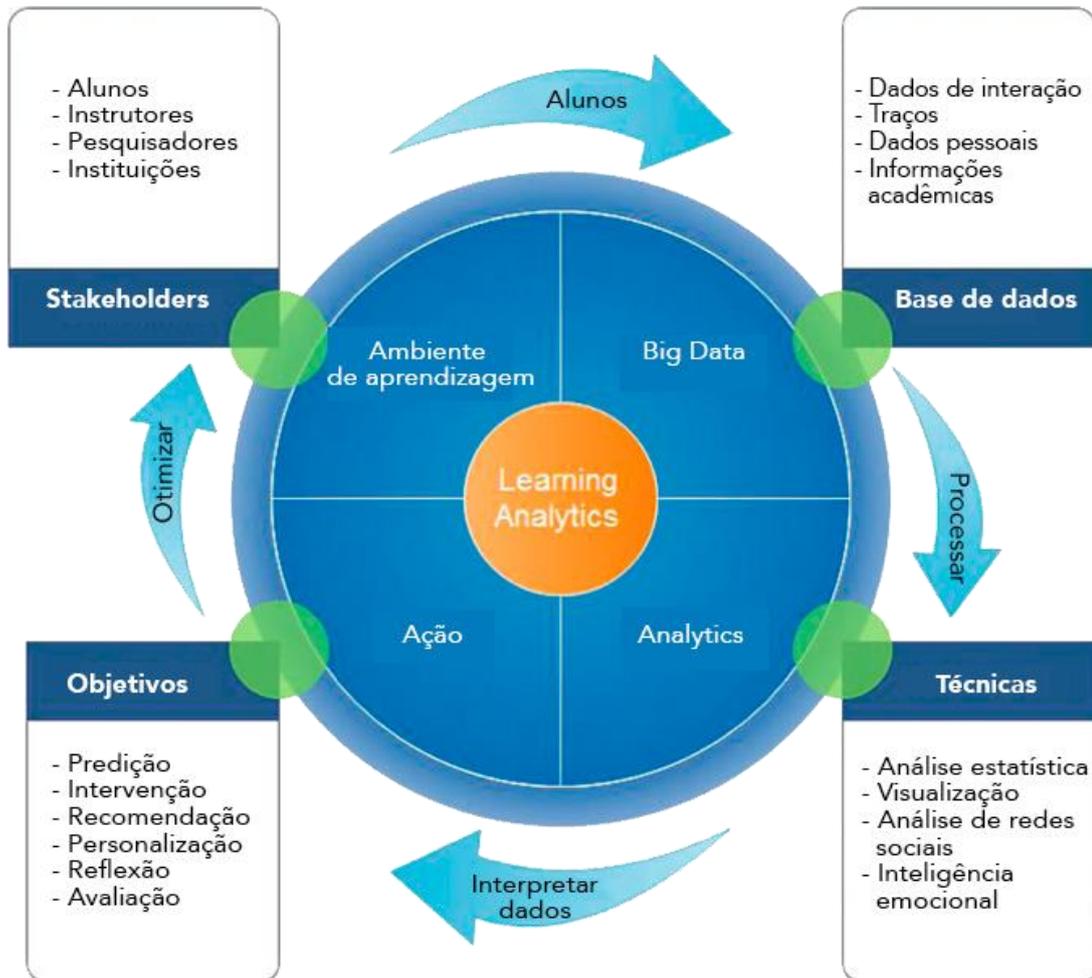
Fonte: (Chatti *et al.*, 2014)

- **O quê?** - Que tipo de dados o sistema coleta, gerencia e usa para a análise? Essa dimensão se refere aos dados usados na tarefa LA. Também se refere aos ambientes e contextos em que a aprendizagem ocorre. Os dados educacionais vêm de canais de aprendizagem formais e informais. Ele também pode vir em diferentes formatos, distribuídos pelo espaço, tempo e mídia.

- **Quem?** - Quem é alvo da análise? A aplicação de LA pode ser orientada para diferentes partes interessadas, incluindo estudantes, professores (inteligentes) tutores / mentores, instituições educacionais (administradores e gerentes do corpo docente), pesquisadores e projetistas de sistemas com diferentes perspectivas, objetivos e expectativas do exercício de LA.
- **Por quê?** - Por que o sistema analisa os dados coletados? Há muitos objetivos em LA de acordo com o ponto de vista particular dos diferentes *stakeholders*. Possíveis objetivos da AL incluem monitoramento, análise, previsão, intervenção, orientação/tutoria, avaliação, *feedback*, adaptação, personalização, recomendação, conscientização e reflexão.
- **Como?** Como o sistema realiza a análise dos dados coletados? LA aplica métodos diferentes para detectar padrões interessantes escondidos em conjuntos de dados educacionais. Os métodos possíveis incluem estatísticas, visualização de informações (IV), mineração de dados (DM) e análise de redes sociais (SNA).

Khalil e Ebner, pesquisadores precursores no campo de LA, propuseram um ciclo de vida do LA, elencando as principais técnicas, objetivos, *stakeholders* e dados, de forma visual (Figura 19), facilitando o entendimento do funcionamento do processo:

Figura 19 - Ciclo de vida do LA



Fonte: Khalil e Ebner (2015)

Geralmente os *frameworks* sobre LA podem ser considerados como uma nova abordagem sob a perspectiva dos dados, de ciclos de aprendizagem propostos lá na década de 80, como o ciclo de aprendizagem de Kolb (1984)¹⁰. Isso porque há correspondências diretas entre as teorias: dados sobre as ações dos alunos (Experiência concreta); dados gerados (observação); métricas (conceptualização abstrata); ações e intervenções (experimentação ativa).

Os diversos *frameworks* propostos se complementam e certamente ajudam o processo de entendimento geral sobre o tema. Mostram que a preocupação com *analytics* deve ser considerada com o um processo contínuo, por isso a maioria deles são cíclicos. Porém esses

¹⁰ Teoria amplamente descrita no livro *Experiential learning: experience as the source of learning and development*.

frameworks vão guiando os usuários até o determinado momento que precisam tomar uma decisão. E é aí que surgem as dúvidas: quais técnicas mais apropriadas para processar a *Big Data*? Quais métricas utilizar? Como e quais dados garimpar? Como extrair informações relevantes dessa massa de dados? Os *frameworks* sugerem o uso de determinadas técnicas, mas isso é insuficiente para o pesquisador, principalmente da área da educação, prosseguir com suas análises.

3.3 MÉTRICAS

Toda abordagem que utilize LA deve começar com a seguinte premissa: a melhoria educacional. Sendo assim, algumas indagações se apresentam: quais as virtudes e fraquezas do meu curso? Quais objetivos educacionais eu gostaria de alcançar? Como aumentar o engajamento dos alunos? Como melhorar a motivação dos alunos? Essa são apenas algumas perguntas comuns que costumam surgir. Antes de qualquer análise, é necessário entender que dados você precisa coletar e como será feita a coleta, com base nos objetivos e metas propostos.

Por isso, a elaboração das métricas deve ser uma etapa inicial. Um bom panorama atual da situação educacional já ajuda a visualizar quais passos devem ser tomados. Os dados em si não oferecem nenhuma informação válida. Eles precisam ser associados com objetivos e métricas.

O capítulo anterior, sobre BDE, abordou os diversos tipos de dados e suas respectivas fontes. Para conseguir traçar um perfil de determinado aluno, basta recorrer ao seu histórico escolar, seus dados pessoais, aplicar um questionário ou entrevista. Para monitorar sua performance, basta acessar os traços deixados no ambiente virtual e suas avaliações. Para identificar aspectos sociais desse aluno, pode-se acessar suas redes sociais, por exemplo. Entretanto, dados individuais de alunos não são úteis, de forma isolada, quando o objetivo é implementar uma política de educação mais ampla na instituição. Para isso, a análise deve ser feita em grupos, turmas, ou seja, uma amostra mais representativa. LA é um processo baseado em contexto bem específico. Dados de instituições diferentes, analisados da mesma forma, podem demonstrar resultados completamente diferentes. A análise de dados trabalha em conjunto com a observação e os aspectos qualitativos.

De acordo com Robbins (2017), de forma geral as métricas associadas ao LA englobam três áreas principais: eficiência, efetividade e resultado.

- **Eficiência:** é relacionada com atividades centradas na aprendizagem: número de alunos, tempo gasto em atividades, frequência de *downloads*, pontuações de questionários, utilização de recursos, entre outros.
- **Efetividade:** é relacionado com atividades centradas na avaliação: qualidade dos resultados, aquisição de conhecimento, colaboração, progresso, desempenho.
- **Resultado:** é relacionado ao resultado final. Uma validação das suas métricas anteriores.

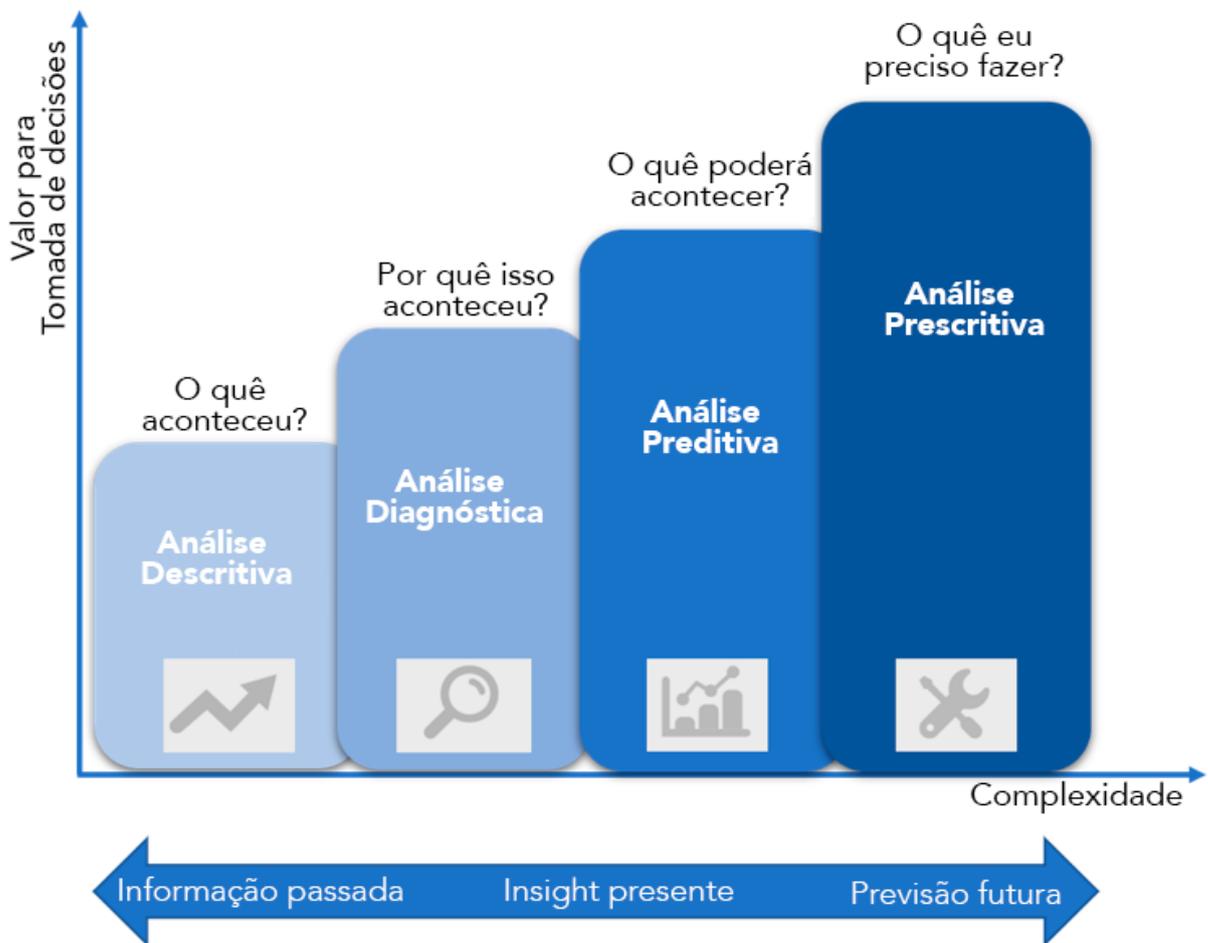
Como vemos, o nível de análise das métricas vai se aprofundando. Quando a métrica é a eficiência, a análise é mais rasa. Os objetivos são analisar como o aluno tem usado o ambiente virtual, se tem utilizado os recursos, respondido as tarefas propostas. Enfim, dá *insights* sobre a funcionalidade dos recursos. Ainda não há uma preocupação com a aprendizagem em si. Perguntas que podem sugerir essa métrica: Os alunos estão utilizando os recursos? Estão participando das atividades? Estão motivados com o conteúdo apresentado?

Na métrica da efetividade, os objetivos de análise já são um pouco mais ambiciosos. Há uma necessidade de entendimento de construtos de aprendizagem mais complexos. Há uma vontade de entender a efetividade da aprendizagem, o desenvolvimento de como funciona o engajamento e a colaboração dos alunos. Perguntas que podem sugerir essa métrica: Os alunos estão aprendendo bem? Estão colaborando para criar conhecimento? Estão engajados nas atividades propostas? Estão tirando notas boas? O conteúdo está apropriado para os alunos?

E por fim, a métrica do resultado, que combina as duas anteriores e trata do processo educacional de uma forma mais geral. Perguntas dessa métrica: o curso ou aula atingiu seus objetivos? Realmente contribuiu para o conhecimento dos alunos? Agregou experiência para a vida dos estudantes? O que pode ser melhorado na próxima oferta desse curso?

Dependendo do objetivo desejado, os métodos de LA aumentam seu grau de complexidade. De acordo com a Figura 20, elaborada pela equipe do MOOC Learn2Analyze, com base nos estudos de (Lang et al.. 2017; Soltanpoor & Sellis, 2016; Scapin, 2015), podemos observar quatro categorias de análise, que vão avançando em complexidade:

Figura 20 - categorias de análise dos dados



Fonte: Learn2Analyze

- A **análise descritiva** é útil para obter *insights* de informações passadas. Analisar dados de cursos que já ocorreram. É ideal para a elaboração de relatórios. As principais técnicas aplicadas para trabalhar com esses dados são a estatística, mineração de dados e agregação de dados.
- A **análise diagnóstica** utiliza métodos mais avançados como mineração de dados, correlação e a descoberta de dados (coleta de várias fontes em busca de detectar padrões e anormalidades). Pode ser aplicada em dados de cursos que estão em andamento, visando a melhoria de desempenho de processos.
- A **análise preditiva** já apresenta uma maior abrangência de métodos mais complexos e sofisticados (*machine Learning*, inteligência artificial e mineração

de dados). É útil para detectar padrões de comportamento e prever possíveis acontecimentos futuros.

- **Análise prescritiva** se baseia nos comportamentos passados e presentes, sugerindo o que poderá acontecer e fornecendo subsídios através de alertas, notificações e sistemas de recomendação. É útil, por exemplo, para analisar padrões de aprendizagem de alunos e sugerir conteúdos personalizados, a fim de evitar a evasão do curso.

3.4 CIÊNCIA DE DADOS E EDUCACAO

A utilização de técnicas de LA visa suprir alguns objetivos comuns que profissionais da educação enfrentam diariamente em suas rotinas. Como melhorar o progresso dos meus alunos? Como modelar o comportamento de aprendizagem? Como detectar os alunos que estão descontentes? Como prever a performance dos meus alunos? Como diminuir as taxas de evasão dos cursos online? Como posso gerar feedback personalizado para tantos alunos? Como posso fazer recomendações personalizadas?

Sistematizar essas dúvidas é o primeiro passo para a análise, se ela será descritiva, diagnóstica, preditiva ou prescritiva. Isso é determinante pois através dessa definição, serão determinados os dados e os métodos utilizados. Por exemplo: se o objetivo for analisar um curso já ofertado, dados passados, o uso de estatística descritiva e técnicas de visualização simples já atendem esse objetivo. Se for realizar uma análise diagnóstica, as ferramentas são outras: inferência estatística, correlações e visualizações podem ajudar.

Se o objetivo for prever comportamentos futuros, baseados nos dados do presente e passado, os métodos de regressão são os mais indicados. Por fim, se o objetivo for gerar recomendações, os métodos de análise precisam ser mais sofisticados, com o emprego de *machine learning* e heurísticas, os quais necessitam um conhecimento aprofundado em ciência de dados. Por isso, quanto maior a complexidade dos objetivos, mais importante a necessidade de montar uma equipe multidisciplinar para orientar as análises. Apesar dos métodos serem computadorizados, sempre haverá a necessidade de intervenção humana e de um especialista em educação.

A seguir abordaremos brevemente as principais técnicas utilizadas, suas virtudes e pontos fracos, assim como as indicações de quando devem ser utilizadas no manejo dos dados educacionais, conforme mostrado na Figura 21.

Figura 21 - Etapas do manejo dos dados educacionais



Fonte: Learn2Analyze

3.5 MÉTODOS ESTATÍSTICOS

Dentre os métodos utilizados na análise de dados educacionais, a estatística é certamente o mais utilizado. Trata-se de uma ciência que inclui métodos de coleta, organização e análise, de tal forma que auxilia na formulação de conclusões. Nessa área, duas categorias se destacam: estatística descritiva e inferência estatística.

A estatística descritiva lida com dados processados, sem traçar nenhuma inferência deles, ao contrário da inferência estatística que usa ferramentas matemáticas para fazer previsões e generalizações para uma população maior, através dos dados analisados.

Por definição a estatística descritiva¹¹ é um ramo que aplica diversas técnicas para descrever e sumarizar um conjunto de dados. Se utiliza de medidas de tendência central e medidas de variabilidade (ou dispersão). Algumas medidas de tendência central são:

- **Média:** é definida como o valor que demonstra a concentração dos dados de uma distribuição, como o ponto de equilíbrio das frequências em um histograma.
- **Mediana:** é o valor que separa a metade maior e a metade menor de uma amostra, uma população ou uma distribuição de probabilidade. Em termos mais simples, mediana pode ser o valor do meio de um conjunto de dados ordenados.
- **Moda:** é o valor mais comum para uma variável particular.

¹¹ https://pt.wikipedia.org/wiki/Estat%C3%ADstica_descritiva

Já algumas medidas de variabilidade ou dispersão são:

- **Desvio padrão:** indica uma medida de dispersão dos dados em torno de média amostral. Um baixo desvio padrão indica que os pontos dos dados tendem a estar próximos da média ou do valor esperado.
- **Variância:** é uma medida da sua dispersão estatística, indicando "o quão longe" em geral os seus valores se encontram do valor esperado.
- **Valor máximo e mínimo:** é o maior e o menor valor encontrado para uma amostra.

Portanto, se o objetivo educacional for apenas analisar o que aconteceu no passado e reportar os resultados em um relatório, a estatística descritiva já proporciona boas ferramentas para essa finalidade. Entretanto, essa seria uma análise mais superficial, com dados claramente disponíveis. Se o que se pretende é entender mais a fundo não somente o que aconteceu, mas como aconteceu, será necessário realizar inferências nos dados, e para isso, será necessário adotar outras ferramentas proporcionadas pela inferência estatística, que é um ramo da Estatística cujo objetivo é fazer afirmações a partir de um conjunto de valores representativo (amostra) sobre um universo (população). Essa já é uma análise mais complexa que procura por diferenças significantes entre as variáveis e os grupos (populações).

- **Correlação:** procura descrever a natureza de um relacionamento entre duas variáveis, como forte, positivo negativo, fraco ou estatisticamente significativo. Se uma correlação for encontrada, ela indica um relacionamento ou padrão, mas lembre-se de que não indica ou implica causalidade.
- **Análise de variância (ANOVA):** é a técnica estatística que permite avaliar afirmações sobre as médias de populações. A análise visa, fundamentalmente, verificar se existe uma diferença significativa entre as médias e se os fatores exercem influência em alguma variável dependente.
- **Regressão:** é uma técnica que permite explorar e inferir a relação de uma variável dependente com variáveis independentes específicas.

No Apêndice 1, localizado no final dessa tese, há uma lista de artigos selecionados que utilizaram métodos estatísticos em dados educacionais.

3.6 MINERAÇÃO DE DADOS

Para realizar análises futuras mais aprofundadas é necessário o uso da aprendizagem indutiva. A indução é a inferência de conhecimento a partir dos dados. Ocorre através da criação de modelos que procuram por padrões e tendências em uma base de dados. Esses modelos geralmente buscam dados com características similares e os agrupam em classes. Esse tipo de aprendizagem pode ser dividido em dois grupos:

- **Aprendizagem supervisionada:** É realizada através de exemplos, onde o analista sugere as classes e ajuda a construir o modelo. O sistema tem que determinar a descrição para cada classe, ou seja, o conjunto de propriedades comuns nos exemplos que lhe são fornecidos. Estando a descrição determinada, é possível formular a regra de classificação que pode ser utilizada para prever a classe de um objeto que não tenha sido considerado aquando da aprendizagem.
- **Aprendizagem não supervisionada:** Essa é realizada com base na observação e descoberta. Não há definições prévias de classes. O sistema define a melhor forma de trabalhar os dados. Como resultado é gerado um conjunto de descrições de classes, o qual o analista tem que interpretar.

As aplicações de mineração de dados podem ser classificadas por conjuntos de problemas que possuem características semelhantes, nos vários domínios de aplicação. Ou seja, podemos aplicar os mesmos modelos para atividades com comportamentos semelhantes, porém de áreas distintas. A diferença reside apenas na forma como os modelos são parametrizados, isto é, que atributos específicos do domínio, armazenados no repositório de dados, são utilizados e como são utilizados. Cada técnica de mineração possui suas peculiaridades, e o que as difere são seus conjuntos de aproximações algorítmicas que são utilizadas para extrair as relações relevantes nos dados. Estas aproximações diferem em função do tipo de problema que visam solucionar. As técnicas mais utilizadas são as seguintes:

- **Classificação e regressão (aprendizagem supervisionada):** Consiste em prever o valor de um atributo categórico (a classe) com base nos valores de outros atributos (os atributos de previsão). Um algoritmo de pesquisa é usado para induzir um classificador a partir de um conjunto de instâncias de dados classificados corretamente, chamado de conjunto de treinamento. Outro conjunto de instâncias de dados classificados corretamente, conhecido como conjunto de testes, é usado para medir a qualidade do classificador obtido após o processo de aprendizado. Diferentes tipos de modelos podem ser usados para representar classificadores, e há uma grande variedade de algoritmos disponíveis para induzir classificadores a partir de dados (Romero et al., 2013).
Aplicação em dados educacionais: descobrir grupos potenciais de estudantes com características e reações semelhantes a uma estratégia pedagógica específica; agrupar estudantes com as mesmas dificuldades e propor soluções personalizadas a esses grupos; identificar alunos com baixa motivação e encontrar ações corretivas para reduzir as taxas de abandono escolar; prever/classificar os alunos ao usar sistemas de tutoria inteligentes, entre outras aplicações.
- **Clusterização (Aprendizagem não supervisionada):** é um método de segmentação de dados que partilham tendências e padrões semelhantes. Este tipo de estudo não tem uma variável dependente. Os resultados emergem da análise de dados sem nenhuma intenção previamente definida. As técnicas de clusterização procuram semelhanças e diferenças num conjunto de dados e agrupam os registros semelhantes em segmentos ou clusters, de uma forma automática, de acordo com algum critério ou métrica. Não é necessário definir os grupos nem os atributos que devem ser utilizados para segmentar o conjunto de dados. Geralmente é uma das primeiras técnicas aplicadas na massa de dados, a fim de obter os primeiros insights e guiar futuras análises. Aplicação em dados educacionais: identificar perfis de aprendizagem semelhantes; analisar o comportamento de um grupo de alunos e seu desempenho escolar; descobrir padrões de comportamento que favorecem o engajamento ou a evasão, entre outros.
- **Associação (aprendizagem não supervisionada):** São ocorrências ligadas a um único evento. Através dessa técnica, pode-se associar, por exemplo, que alunos com notas mais baixas em determinadas matérias são aqueles que possuem uma rotina de trabalho

mais exaustiva, e não têm tempo para se dedicar mais aos estudos. Cabe aos gestores e professores tomarem atitudes e flexibilizar as atividades para esse público.

- **Análise de sequenciação** (aprendizagem não supervisionada): Nessa situação os eventos estão ligados ao longo do tempo. Por exemplo: um estudo de modelos educacionais pode descobrir que determinados alunos que finalizam um curso sobre educação financeira, 60% se inscreveram, em seguida, em outro curso sobre renda variável. Com essas informações, os gerentes podem tomar decisões mais acertadas pois aprenderam a respeito do interesse dos alunos.
- **Prognóstico** (aprendizagem supervisionada): Variante do problema de classificação, que envolve a geração de uma pontuação para cada registo. Por exemplo, em vez de se atribuir uma classificação binária a cada ação do aluno (bom ou mau), esta aproximação gera uma pontuação do interesse em se conceder novos materiais, tendo por base um conjunto de treino pré-definido. Dessa forma, dá para direcionar materiais mais avançados para alunos mais experientes, e vice-versa.

3.7 APRENDIZAGEM DE MÁQUINA (*MACHINE LEARNING*)

O aprendizado de máquina é um subcampo da ciência da computação que se preocupa em construir algoritmos que, para serem úteis, dependem de uma variedade de exemplos de algum fenômeno. Esses exemplos podem vir da natureza, serem artesanais por seres humanos ou gerados por outro algoritmo. O aprendizado de máquina também pode ser definido como o processo de solução de um problema prático: 1) reunindo um conjunto de dados e 2) construindo algoritmicamente um modelo estatístico com base nesse conjunto de dados. Presume-se que esse modelo estatístico seja usado de alguma forma para resolver o problema prático (Burkov, 2019)

O objetivo principal é detectar padrões e aplicar regras para categorizar pessoas ou coisas, prever comportamentos, identificar relações entre variáveis e detectar comportamentos inesperados. O interessante é que os algoritmos não são estáticos, ou seja, conforme o sistema for alimentado com novos dados, os algoritmos mudam, se adaptam, fazendo com que a performance das análises aumente ao longo do tempo.

O uso de modelos analíticos permite que pesquisadores, cientistas de dados, engenheiros, e analistas possam "produzir decisões e resultados confiáveis e repetíveis" e descobrir os "insights escondidos" através do aprendizado das relações e tendências históricas nos dados.

Sua aplicação nos ambientes virtuais é amplamente difundida, nos mais diversos setores e atividades: filtragem de spam, reconhecimento ótico de caracteres (OCR), processamento de linguagem natural, motores de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional, entre outros. Na área educacional, vem sendo utilizado para aumentar o engajamento de alunos, diminuir taxas de evasão, implementando sistemas de recomendação de materiais didáticos e cursos, além de estratégias de marketing digital para venda de cursos. Convém lembrar que o resultado da análise está intimamente ligado à escolha do algoritmo correto. Dessa forma, quando há mudança nos dados, há mudança no resultado da análise.

Segundo descrito no e-book da SAS (2017), empresa líder de mercado no ramo de *analytics*, o aprendizado de máquina é particularmente útil para resolver problemas onde:

- Associações ou regras aplicáveis podem ser criadas, mas não são facilmente codificadas ou descritas por regras lógicas simples.
- Potenciais resultados ou ações são definidos, mas que ação a ser tomada depende de diversas condições que não podem ser previstas ou identificadas antes que um evento aconteça.
- Precisão dos resultados é mais importante que sua interpretação.
- Os dados não são propícios para as técnicas analíticas tradicionais. Especificamente, quando o conjunto de dados apresenta uma grande quantidade de atributos em cada registro ou os dados são altamente correlacionados (dados com valores semelhantes ou intimamente relacionados) podem confundir os métodos analíticos tradicionais.

Assim como a mineração de dados, o aprendizado de máquina também engloba uma série de técnicas analíticas que podem ser supervisionadas, não supervisionadas, semi supervisionadas e por reforço. As definições a seguir pertencem ao ebook da SAS (2017), disponível no site da empresa¹².

¹² https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/machine-learning-primer-108796.pdf

A aprendizagem supervisionada é realizada através de exemplos, onde o analista sugere as classes e ajuda a construir o modelo. O uso de dados históricos, por exemplo, pode guiar a análise futura. Estando a descrição determinada, é possível formular a regra de classificação que pode ser utilizada para prever a classe de um objeto que não tenha sido considerado quando da aprendizagem. Sua utilização é adequada em sistemas de detecção de fraudes bancárias, segmentação de público, reconhecimento de texto, personalização de interações, avaliação de riscos, entre outras. As Técnicas mais comuns são:

- **Estatística Bayesiana:** consiste na avaliação de hipóteses pela máxima verossimilhança, uma decorrência imediata da fórmula de Bayes, e é fundamental para métodos computacionais relacionados à inteligência, mineração de dados, ou linguística, sejam eles métodos bayesianos de aprendizado de máquina (AM) ou não-bayesianos.
 - **Árvores de decisões:** é uma representação de uma Tabela de decisão sob a forma de uma árvore, porém pode haver outras aplicações. Tem a mesma utilidade da Tabela de decisão. Trata-se de uma maneira alternativa de expressar as mesmas regras que são obtidas quando se constrói a Tabela.
 - **Redes neurais:** método baseado no entendimento da arquitetura dos neurônios do cérebro. Utiliza critérios cada vez mais discretos e abstratos (também conhecidos como camadas) para relacionar entradas a saídas discretas.
 - **Análise de regressão:** é um conjunto de processos estatísticos para estimar os relacionamentos entre uma variável dependente (geralmente chamada de 'variável de resultado') e uma ou mais variáveis independentes (geralmente chamadas de preditores). A forma mais comum de análise de regressão é a regressão linear, na qual um pesquisador encontra a linha (ou uma combinação linear mais complexa) que mais se ajusta aos dados de acordo com um critério estatístico chamado método dos mínimos quadrados.
 - **Florestas aleatórias:** são um método de aprendizado conjunto para classificação, regressão e outras tarefas que operam construindo várias árvores de decisão no momento do treinamento e gerando a classe que é o modo das classes (classificação) ou previsão
-

média (regressão) das árvores individuais. As florestas de decisão aleatória corrigem o hábito de adaptação das árvores de decisão ao seu conjunto de treinamento.

A aprendizagem não supervisionada: Essa é realizada com base na observação e descoberta. Não há definições prévias de classes. O sistema define a melhor forma de trabalhar os dados. Como resultado é gerado um conjunto de descrições de classes, o qual o analista tem que interpretar. As técnicas mais comuns são:

- **Análise de afinidade:** método que identifica quando acontecimentos (eventos ou pessoas) tipicamente ocorrem simultaneamente.
- **Clusterização:** é um método que agrupa coisas (eventos, pessoas, objetos) por proximidade ou similitude de atributos. Um dos mais utilizados é o K-Means: é um método que agrupa coisas (eventos, pessoas, objetos) por proximidade a um ponto “K” previamente definido pelo pesquisador e, posteriormente, refinado pelo algoritmo. Outro método comum é o Mapeamento de vizinho mais próximo: o qual agrupa coisas (eventos, pessoas, objetos) baseado no que o grupo possui de mais prevalente na área que está sendo mapeada.
- **Decomposição de valor singular:** um método para redução de dados altamente variáveis e dimensionais (incluindo dados altamente correlacionados) em menos dimensões para permitir que a estrutura e os relacionamentos sejam mais facilmente compreendidos.

A aprendizagem semi-supervisionada é usada para resolver problemas semelhantes aos da supervisionada. No entanto, no aprendizado semi-supervisionado, a máquina recebe alguns dados com a resposta definida (também conhecida como rotulada) junto com dados adicionais que não estão identificados com a resposta. Em outras palavras, alguns dos dados de entrada são marcados com a saída desejada (resposta) enquanto outros não estão marcados. A aprendizagem semi-supervisionada é usada nos casos em que há muitos dados ou variações sutis para poder fornecer um conjunto abrangente de exemplos. Nesse caso, as entradas e saídas fornecidas fornecem as informações gerais padrão que a máquina pode extrapolar e aplicar aos dados restantes. Exemplos práticos desse tipo de aprendizado são os aplicativos com reconhecimento de fala, reconhecimento e classificação de imagens (cibersegurança) e classificação de sites.

Na aprendizagem por reforço é fornecido à máquina um conjunto de ações permitidas, regras e estados finais em potencial. Em outras palavras, as regras do jogo são definidas. Aplicando as regras, explorando diferentes ações e observando reações resultantes, a máquina aprende a explorar as regras para criar o desejado resultado. Assim, determinar quais séries de ações, em que circunstâncias, serão levar a um resultado ótimo ou otimizado. O aprendizado por reforço é o equivalente a ensinar alguém a jogar um jogo. As regras e objetivos estão claramente definidos. No entanto, o resultado de qualquer jogo único depende do julgamento do jogador que deve ajustar sua abordagem em resposta ao ambiente histórico, habilidade e ações de um determinado oponente. As aplicações práticas desse tipo de aprendizagem são na área de jogos, robótica e na tecnologia embutida nos GPS, nos quais o sistema determina a melhor rota baseado no fluxo de trânsito, condições climáticas etc. As técnicas mais comuns são:

- **Redes Neurais artificiais:** são modelos computacionais inspirados pelo sistema nervoso central de um animal (em particular o cérebro) que são capazes de realizar o aprendizado de máquina bem como o reconhecimento de padrões. Redes neurais artificiais geralmente são apresentadas como sistemas de "neurônios interconectados, que podem computar valores de entradas", simulando o comportamento de redes neurais biológicas.
- **Autômato de aprendizagem:** é um tipo de algoritmo de aprendizado de máquina estudado desde os anos 70. Os autômatos de aprendizado selecionam suas ações atuais com base em experiências passadas do ambiente. Ele se enquadra no aprendizado de reforço se o ambiente for estocástico e um processo de decisão de Markov (MDP) for usado.
- **Processo de decisão de Markov (MDP):** é um processo de controle estocástico de tempo discreto. Ele fornece uma estrutura matemática para modelar a tomada de decisão em situações em que os resultados são parcialmente aleatórios e parcialmente sob o controle de um tomador de decisão. Os mdps são úteis para estudar problemas de otimização resolvidos por meio de programação dinâmica e aprendizado por reforço.
- **Q-Learning:** é um algoritmo de aprendizado por reforço sem modelo para aprender uma política que diz a um agente que ação tomar em que circunstâncias.

Ele não requer um modelo (daí a conotação "livre de modelo") do ambiente e pode lidar com problemas com transições e recompensas estocásticas, sem exigir adaptações.

3.8 ANÁLISE QUALITATIVA

Para obter uma visão mais holística do sistema educacional como um todo, há uma necessidade de integrar aspectos da análise quantitativa assim como qualitativa. Esse é justamente o diferencial do *Learning Analytics*. Os *insights* gerados através de análise da BDE dos alunos são muito valiosos, mas podem ser mal interpretados quando o analista não conhece bem as peculiaridades do fenômeno que está sendo estudado. Por exemplo, um cientista de dados que domine diversas técnicas de mineração de dados, aprendizado de máquina e métodos estatísticos, pode tirar conclusões precipitadas através de dados corretos. Ao analisar, por exemplo, as taxas de evasão dos alunos de um determinado MOOC, as quais apresentam taxas de cerca de 50%, ele pode concluir que esse curso foi um fracasso, pois metade dos alunos não obteve o diploma. Entretanto, qualquer profissional da educação que conheça a dinâmica de um MOOC, sabe que esses cursos possuem essa característica e esses números estão dentro da normalidade. Segundo Perna *et al.* (2013), em um estudo sobre evasão em 32 cursos ofertados pela Coursera, obteve-se uma taxa de 4% de concluintes. O mesmo estudo demonstrou que a taxa de alunos que concluíram os cursos da EDX em 2014 chegou a 7%, e em 2015 a 8%. Tais números demonstram que, mesmo nas grandes universidades, os MOOCs apresentam números altos de evasão.

Por melhor que seja o MOOC, com conteúdo de qualidade, ou chancelado por universidades renomadas mundo afora, o número de alunos que concluem o curso é sempre menor do que o número de alunos que se matriculam. Isso comprova que os dados nem sempre são suficientes em uma análise de atividade educacional, que envolve tantos aspectos subjetivos.

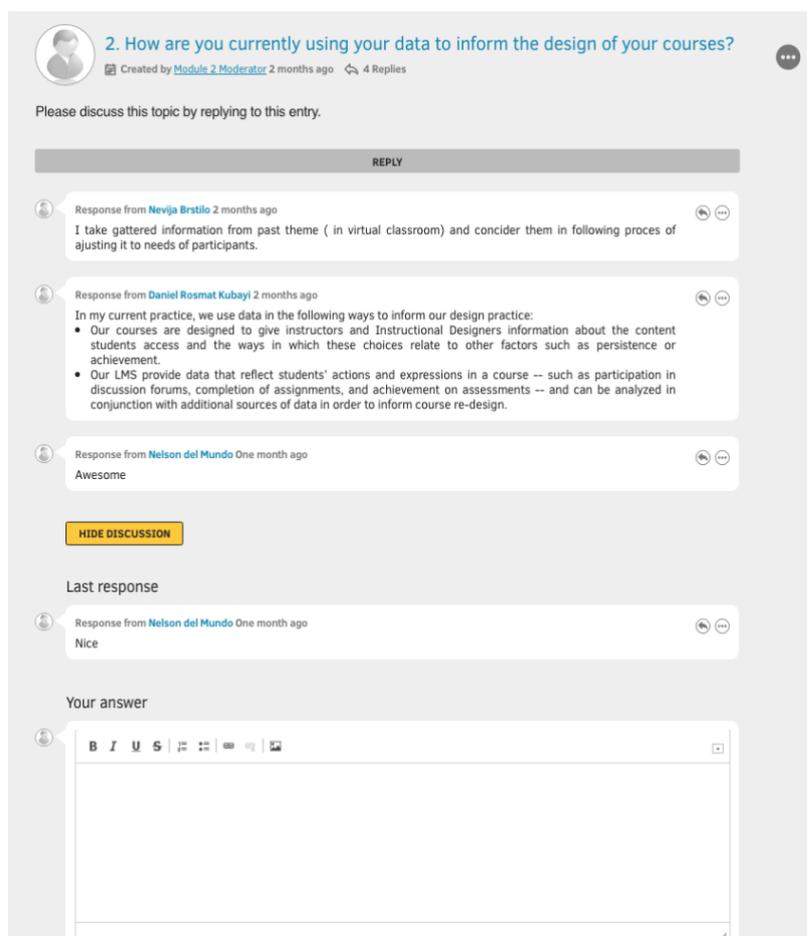
E, justamente, essas peculiaridades estão sendo pesquisadas no mundo todo, com professores, gestores educacionais, tutores, educadores, enfim, todos interessados em entender o porquê desses comportamentos. Inúmeros artigos com estudos de casos publicados, buscam a compreensão de fenômenos que os dados não conseguem explicar. Nesses casos, os aspectos qualitativos ganham destaque e as discussões geradas em congressos e fóruns ganham relevância no entendimento dessas questões.

A verdade é que a coleta de dados qualitativos deve ser pensada já no momento da criação do curso, na etapa do design instrucional do material. A inserção de um pré-questionário, aplicado no início do curso, abordando as perspectivas esperadas pelos alunos, e um questionário pós-curso, por exemplo, já proporciona uma boa análise de expectativa X realidade. Dessa forma dá para comparar o que os alunos esperam do curso e o que eles realmente acharam. Isso também gera insights para futuras melhorias e aperfeiçoamentos no material didático.

Uma técnica interessante é a utilização de pequenos questionários durante a apresentação do material. Esse simples recurso já fornece bons *insights* para formular uma ideia do nível de conhecimento que a turma possui sobre determinados assuntos.

Outra ferramenta muito útil para obter dados qualitativos são os fóruns. Através deles, o tutor pode direcionar questionamentos e avaliar as respostas e interações dos alunos, tendo um bom subsídio para avaliar o nível de conhecimento e maturidade de alunos específicos e até mesmo do grupo como um todo. Além disso o fórum tende a aumentar o engajamento e a socialização entre alunos e tutores, gerando um material rico para análise, além do que os dados podem proporcionar, conforme mostra a Figura 22:

Figura 22 - Exemplo de fórum comum em MOOCs



Fonte: Learn2Analyze

3.9 OPEN LEARNING ANALYTICS

A comunidade de pesquisadores de LA aponta para a necessidade cada vez maior dessa troca de conhecimentos para avançar nas pesquisas nos ambientes educacionais. Isso engloba uma série de procedimentos metodológicos que devem ser devidamente explicitados e publicados. Há necessidade de mais estudos sobre os aspectos qualitativos que envolvam o agrupamento e colaboração de estudantes; análise de redes sociais; desenvolvimento de mapas conceituais; construção do material didático e no planejamento e agendamento de atividades.

Apesar do entusiasmo dos pesquisadores a respeito do potencial do LA para o avanço das pesquisas na área, muitos acham que o campo ainda não tem maturidade suficiente, necessitando de estudos mais aprofundados e resultados mais concretos. Segundo González-Torres, García-Peñalvo, & Therón (2013), estudos nessa área devem focar na apresentação de visualizações habilidosamente projetadas para facilitar os resultados analíticos.

Um aspecto central dessa discussão é o conceito de *Open Learning Analytics (OLA)*. Siemens *et al.* (2011) fornecem uma proposta inicial expressando a importância de uma plataforma integrada e modularizada para integrar análises de aprendizagem heterogêneas técnicas. O conceito de OLA representa uma mudança significativa em direção a um novo modelo analítico de aprendizagem que leve em consideração a “abertura”. Isso leva a perguntas sobre como “aberto” deve ser interpretado em relação à análise de aprendizado? Quais são os desafios da análise de aprendizagem aberta? Quais são os componentes de um ecossistema de análise de aprendizado aberto? Quais são os requisitos para uma abertura efetiva plataforma de análise de aprendizagem? Quais são os detalhes técnicos (por exemplo, arquitetura e módulos) de uma plataforma de análise de aprendizagem aberta?

Segundo Shum Buckingham (2012), para o efetivo avanço da pesquisa sobre LA, há necessidade da convergência dos três tipos de análise que o LA abrange:

- **Macro Análise:** procura possibilitar análise entre instituições, por exemplo, através pesquisas de "maturidade" sobre as atuais práticas institucionais, ou melhorar o acesso aos dados, padronizando condutas de avaliação dos alunos. A macro análise tende a se tornar cada vez mais em tempo real, incorporando mais dados dos níveis meso (de granularidade mais fina), e se beneficiar das metodologias de *benchmarking* e integração de dados desenvolvidas em setores não educacionais.
- **Meso Análise:** operar em nível institucional. Na medida em que as instituições educacionais compartilham processos de negócios comuns a setores que já se beneficiam do *Business Intelligence (BI)*, elas podem ser vistas como um novo setor de mercado de BI, que pode apropriadamente usar ferramentas para integrar silos de dados em armazéns corporativos, otimizar fluxos de trabalho, gerar painéis, extrair dados não estruturados, prever melhor 'rotatividade de clientes' e mercados futuros e assim por diante. É essencial que o BI otimize os processos de negócios que motivam parcialmente os esforços para criar “análises acadêmicas” em nível institucional, assim vemos comunidades de prática especificamente para o BI nas organizações educacionais, que têm suas próprias culturas e legado tecnológico.
- **Micro Análise:** apoiar o rastreamento e a interpretação dos dados no nível do processo para alunos individuais e grupos. Esses dados são de interesse primário dos próprios alunos e dos responsáveis pelo seu sucesso, pois podem fornecer o melhor nível de

detalhe sobre o desempenho acadêmico do aluno. Esses dados mostram o caminho escolhido pelo aluno, seu rastro de aprendizagem e podem divulgar a atividade online clique por clique, além de atividades físicas como geolocalização, empréstimos para bibliotecas, compras e dados interpessoais, como redes sociais. Os pesquisadores estão adaptando técnicas de campos como gamificação, mineração de dados educacionais, aprendizado colaborativo suportado por computador, sistemas de recomendação, sistemas de tutoria inteligente / hiperídia adaptativa, visualização de informações, linguística e argumentação computacional e análise de redes sociais.

De forma geral, o que a comunidade acadêmica apregoa é a padronização de metodologias de análise que gerem bons impactos na aprendizagem. E isso só será possível quando houver uma ampla divulgação de conhecimentos e melhores práticas, devidamente documentados. Uma iniciativa que exemplifica bem essa cultura de compartilhamento de conhecimento é o consórcio entre universidades europeias que originou o MOOC Learn2Analyze. Ele contempla seis universidades de cinco países (Grécia, Noruega, Alemanha, Itália e Irlanda) e conta com a parceria de pesquisadores australianos também. O resultado foi a criação de um conteúdo muito rico, proporcionando avanços na pesquisa sobre LA, capacitação de profissionais para atuarem na área da educação a distância e movimentando o mercado dessa área.

3.10 ANÁLISE DE REDES SOCIAIS

A análise de redes sociais (SNA), também chamada de Análise de Rede Organizacional em ambientes corporativos, torna visível as estruturas e a dinâmica das redes interpessoais, para entender como as pessoas desenvolvem e mantêm essas relações. No contexto educacional ela permite, por exemplo, ver onde estão os alunos incorporados na rede, sua influência estrutural, nível de engajamento e perfis de usuários. As pessoas podem formar "laços" de diferentes tipos, variando de interação direta e prolongada, que reflete laços significativos, a laços mais indiretos.

A SNA pode ser utilizada para análise de redes sociais e fóruns, observando como se dá a construção de conhecimento nesses ambientes. Algumas pesquisas começam a demonstrar que as conexões que os alunos estabelecem entre si e as estruturas grupais resultantes podem

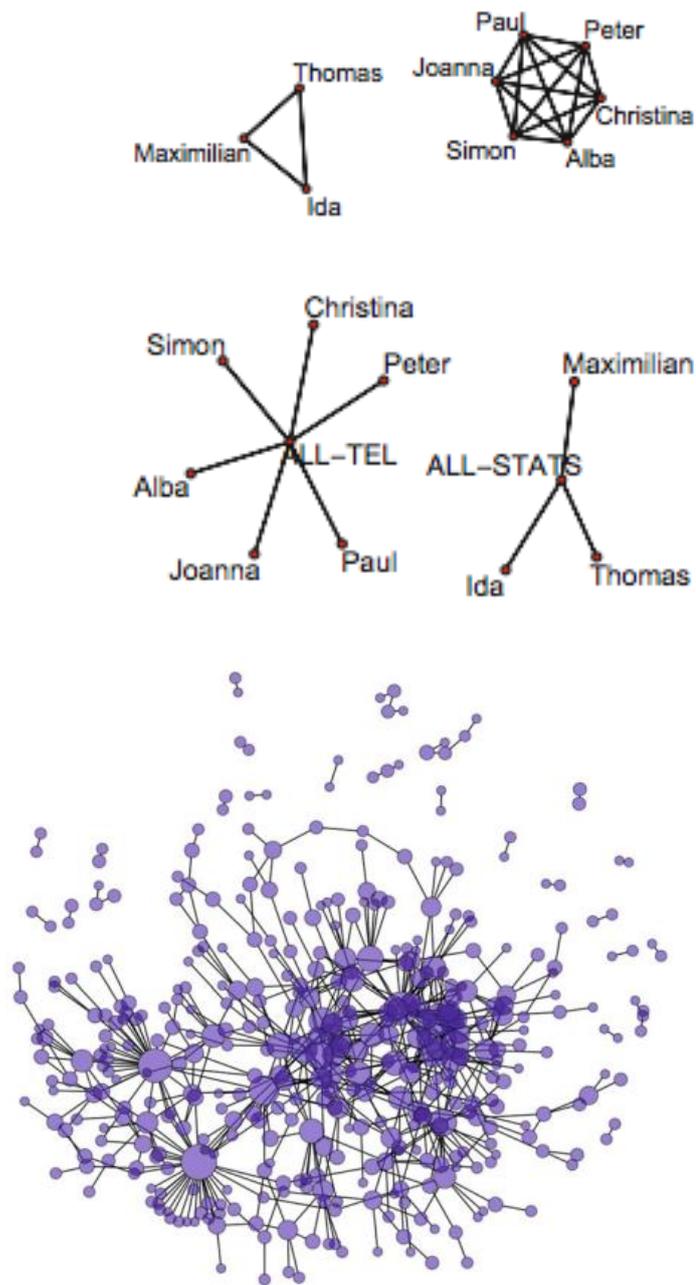
se correlacionar com um aprendizado mais ou menos eficaz (Haythornthwaite, 2010). Corroborando com esses resultados, Dawson (2010) ressalta:

O SNA baseia-se em vários conceitos da teoria dos grafos e da teoria estrutural para avaliar propriedades de rede como densidade, centralidade, conectividade, intermediação e graus. Essas medidas fornecem uma estrutura para interpretar e desenvolver uma compreensão dos padrões observados de trocas que ocorrem entre os atores sociais. No aprendizado on-line, os dados dos alunos podem ser coletados sobre vários tipos de comunicação, incluindo registros de bate-papo, postagens em fóruns de discussão, postagens em blogs e comentários. Há uma riqueza potencial para mineração de dados dessas redes de alunos.

O objetivo principal desse tipo de análise é destacar de antemão problemas potenciais, sobre grupos de estudantes cujos padrões de participação são incomuns. Por exemplo, tendo o conhecimento prévio dos alunos que demonstram sinais de desistência do curso, os instrutores podem atingir esses grupos de alunos, fornecendo-lhes suporte especializado baseado no estilo de aprendizagem, tempo disponível e objetivos pessoais. Por outro lado, uma indicação prévia sobre alunos que estão demonstrando sinais de envolvimento ativo no curso discussões, pode permitir que os instrutores forneçam a eles incentivos para continuar compartilhando seus conhecimentos.

No livro *Learning Analytics in R with SNA, LSA, and MPIA* (Wild, 2016), o autor apresenta a teoria da Análise de Interação Propositiva Significativa (MPIA), que combina a análise de rede social (SNA) com a análise semântica latente (LSA) para ajudar a criar e analisar um cenário de aprendizado significativo a partir dos traços digitais deixados por uma comunidade de aprendizagem na co-construção de conhecimento. Destaca-se a imagem gerada das redes analisadas, das quais é possível ver os nós e as interações entre os alunos, conforme mostra a Figura 23:

Figura 23 - Imagem da rede de interações dos alunos



Fonte: Wild (2016)

Segundo Haythornthwaite (2010), quando aplicada no contexto da aprendizagem online, a SNA possui quatro objetivos principais:

- Usar a evidência dos padrões encontrados na rede social para gerar intervenções que melhorem as atividades de aprendizado online;
- Descobrir os antecedentes das atividades de aprendizagem;

- Descobrir e utilizar métricas da SNA para prever novos objetivos de aprendizagem;
- Entender a natureza e o significado dos laços de aprendizagem.

Esse campo de pesquisa vem evoluindo rapidamente nos últimos anos. Prova disso é a criação de diversos softwares para essa finalidade. O “Enterprise 2.0¹³” pode ser usado para identificar os usuários mais ativos de uma rede online e aqueles que provavelmente têm maior influência na atividade de terceiros. Outra ferramenta projetada especificamente para redes de aprendizado é o SNAPP¹⁴, que publica postagens em fóruns de discussão como um diagrama de rede, identifica alunos desconectados ou visualiza como o apoio ao professor é empregado na rede. Outro é o NAT, projetado para visualizar a rede de tal forma a identificar relações entre pessoas que estão interagindo em torno dos mesmos tópicos, além de fornecer um sociograma baseado na interação de alunos em fóruns.

Vale ressaltar que a SNA é uma técnica relativamente nova e que precisa de um amadurecimento. Ainda não apresenta um sólido *framework* que ajude os pesquisadores na interpretação dos resultados. Essa falta de padronização na aplicação dessa técnica pode levar os pesquisadores ao risco de obterem resultados e conclusões imprecisas.

3.11 APRESENTAÇÃO DOS DADOS

Ao final do processo de análise dos dados inicia-se a etapa de apresentação ou, que consiste na representação gráfica de informações e dados. É o momento no qual são disponibilizados os resultados da análise de uma maneira simples e didática. Nessa fase, os métodos mais comuns incluem o uso de Tabelas, gráficos, mapas, diagramas, entre outros. Essas tecnologias são essenciais para analisar a BDE e facilitar a tomada de decisões impulsionadas pelos dados, que é o objetivo final do processo de LA.

Para criar uma apresentação de dados efetiva é fundamental equilibrar a estética e a funcionalidade. Os dados e elementos visuais devem funcionar em conjunto, de tal forma que informem de forma fácil e atrativa. Segundo o site tableau.com¹⁵, há vários métodos de apresentar os dados de forma eficaz e interessante:

¹³ <http://mzinga.com/omnisocial-learning-extended-enterprise/>

¹⁴ <http://research.uow.edu.au/learningnetworks/seeing/snapp>

¹⁵ <https://www.tableau.com/pt-br/learn/articles/data-visualization>

Quadro 10 - Métodos de apresentação de dados

Tipos comuns de apresentação de dados:	
• Tabelas	• Gráficos
• Mapas	• Diagramas
• Painéis	• Infográficos

Exemplos mais específicos de métodos de apresentação de dados:	
• Gráfico de barras	• Gráfico de caixa
• Nuvem de bolhas	• Gráfico de marcador
• Cartogramas	• Exibição de círculos
• Mapa de distribuição de pontos	• Gráfico de Gantt
• Mapa de variações	• Tabela de destaque
• Histograma	• Matriz
• Rede	• Área polar
• Árvore radial	• Gráfico de dispersão (2D ou 3D)
• Gráfico de fluxo	• Tabelas de texto
• Linha do tempo	• Mapa de árvore
• Gráfico de segmentos	• Nuvem de palavras
• Gráfico de área	• E qualquer combinação de gráficos em um painel.

Fonte: Tableau.com

O mais comum é o gráfico com barras retangulares e comprimento proporcional aos valores que ele representa. As barras podem ser desenhadas verticalmente ou horizontalmente. Um eixo do gráfico mostra especificamente o que está sendo comparado enquanto o outro eixo representa valores discretos. Geralmente é usado quando se quer mostrar como algumas quantidades variam entre um conjunto particular de itens.

O gráfico de linha exibe informações com uma série de pontos de dados chamados de marcadores, ligados por segmentos de linha reta. Muitas vezes ele é usado para visualizar uma tendência nos dados em intervalos de tempo (uma série de tempo). Assim, a linha é muitas vezes atraída por ordem cronológica.

Outra forma comum de apresentar os dados é através do gráfico de setores (comumente conhecido como gráfico de pizza). Consiste em um diagrama circular em que os valores de cada categoria estatística representada são proporcionais às respectivas medidas dos ângulos.

O histograma (ou distribuição de frequências) é a representação gráfica em colunas ou em barras (retângulos) de um conjunto de dados previamente tabulado e dividido em classes uniformes ou não uniformes.

Já os gráficos de dispersão são representações de dados de duas (tipicamente) ou mais variáveis que são organizadas em um gráfico. O gráfico de dispersão utiliza coordenadas cartesianas para exibir valores de um conjunto de dados. Muito útil para representar correlações existente entre os dados e o quão forte são essas correlações. A Figura 24 mostra uma representação de cada tipo de gráfico abordado anteriormente.

Figura 24 - Tipos comuns de gráficos usados na visualização de dados



Fonte: Learn2Analyze

Convém lembrar que muitas vezes a visualização de dados e a análise estatística descritiva serve para sumarizar as principais características de uma base de dados, dando um panorama rápido sobre o fenômeno estudado. É um passo inicial que promove uma aproximação e um maior conhecimento dos dados antes de começar a fazer inferências e insights mais elaborados.

O MOOC Learn2Analyze elencou uma série de ferramentas de visualização que consolidam informações educacionais importantes, facilmente disponibilizadas na web. Uma delas é o *dashboard* LAPA (*Learning Analytics for Prediction and Action*), dos autores (Park

& Jo, 2015), cujo objetivo é informar visualmente o comportamento dos alunos para os tutores poderem promover uma aprendizagem mais personalizada. Outra ferramenta para identificar estudantes em risco de evasão é o LADA (*Learning Analytics Dashboard for Advisers*). Através de gráficos, o sistema mostra para o aluno como está o seu desempenho em relação às metas do curso e um comparativo com outros alunos do curso (Gutiérrez et al. 2018). Com uma proposta semelhante ao LADA, Charleer et al.2018) desenvolveram o LISSA (*Learning dashboard for Insights and Support during Study Advice*) que, através das médias de performance dos alunos, elaborar histogramas para o aluno ver sua posição em relação aos colegas.

Já o smartclass¹⁶ é uma ferramenta integrada ao Moodle que gera gráficos e Tabelas que ajudam os tutores e professores observarem de forma fácil o desempenho dos seus alunos. A ferramenta identifica quais alunos realizaram ou não determinadas tarefas, identifica padrões de aprendizagem e habilidades mais desenvolvidas dos estudantes.

Apesar de fornecerem muita informação válida sobre a performance dos alunos, esses dados obtidos através dos logs ou traços deixados pelos alunos, através dos ambientes virtuais, não medem os aspectos qualitativos da aprendizagem. Saber quantas horas o aluno ficou online no Moodle, ou quantas atividades ele realizou, por exemplo, não dá condição de medir o quanto ele aprendeu efetivamente, mas pode demonstrar dados comportamentais (engajamento, esforço, força de vontade) que determinam um perfil do estudante E isso pode ajudar tutores e professores no planejamento de estratégias de estudos.

A ciência de dados pode influenciar substancialmente a compreensão do aprendizado em ambientes on-line e híbridos. Isso, é claro, implica uma mudança no papel típico dos educadores, de tutores e facilitadores para a execução de algumas das tarefas que os analistas de dados costumam realizar. Eles precisam ser capazes de descobrir os padrões nos dados e transmitir o significado em termos educacionais, ou seja, interpretar os resultados da análise em esquemas de aprendizado significativos.

3.12 TÉCNICAS MAIS UTILIZADAS

Pesquisadores pioneiros no campo do LA, Baker e Siemens (2009) elencaram quatro tipos de análises que podem ser feitas e recomendaram os respectivos métodos mais apropriados:

¹⁶ https://moodle.org/plugins/local_smart_klass

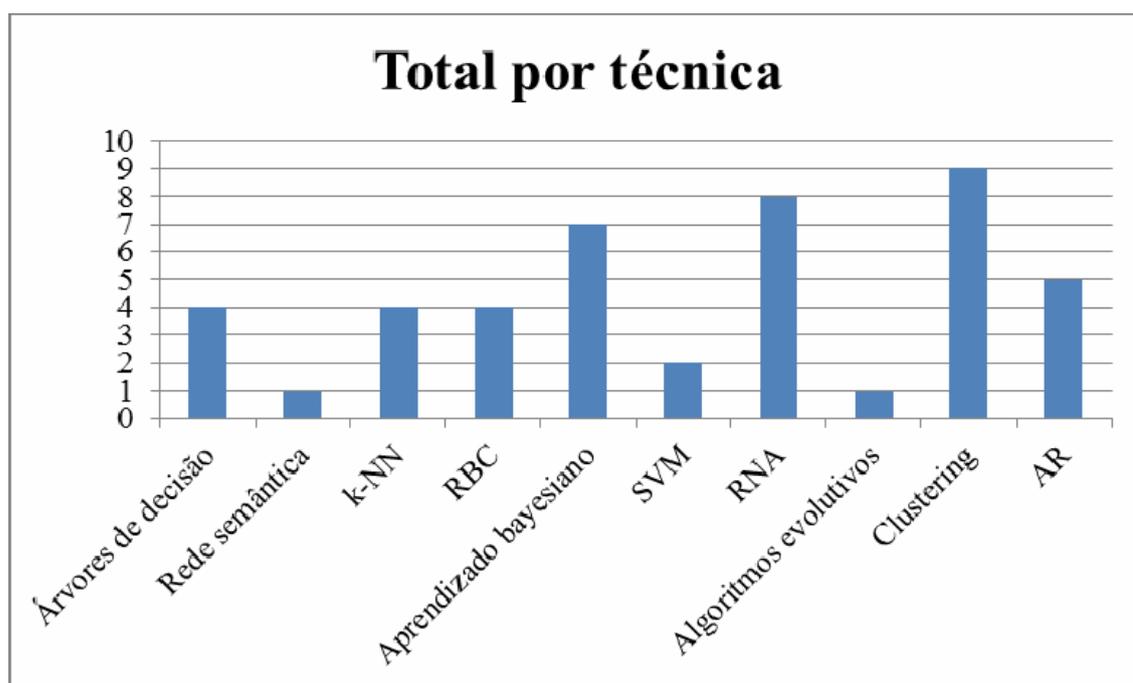
1. Predição – Há a tentativa de modelar uma variável previamente escolhida. Para isso se utilizam técnicas de classificação (árvore de decisões, floresta aleatória, regras de decisão, regressão *stepwise* e regressão logística) e Estimativa do conhecimento latente (rastreamento de conhecimento Bayesiano e Análise fatorial de performance)
2. Descoberta de estruturas – Tentativa de achar alguma informação relevante que emerge dos dados, sem um objetivo previamente definido. Para isso são utilizados técnicas de Clusterização, análise fatorial, descoberta domínios estruturados e análise de redes.
3. Mineração de relação – Objetiva descobrir relações entre as variáveis de um banco de dados com muitas variáveis. As principais técnicas utilizadas são a mineração correlacional, mineração de padrões sequenciais, mineração de dados causais
4. Destilação de dados para gestão humana.

Definidas as métricas e os objetivos da pesquisa, é preciso delimitar o procedimento mais apropriado para realizar a análise dos dados. Essa escolha é fundamental para conseguir extrair o potencial máximo que os dados podem oferecer. Observar os parâmetros citados anteriormente, organizar seus dados de forma consistente, já é o primeiro caminho para uma análise mais assertiva.

A definição do procedimento de LA segue a lógica da ciência de dados, que nada mais é que uma mistura de diversas ferramentas, algoritmos e princípios de aprendizagem de máquina (*machine learning*) que nos auxiliam na descoberta de padrões escondidos nos dados brutos (Sharma, 2019).

Um artigo de Beserra et al. (2014), publicado na Revista RENOTE, da Universidade Federal do Rio Grande do Sul, realizou um mapeamento sistemático para identificar quais as técnicas de *machine learning* mais utilizadas em objetos de aprendizagem. Para tanto eles analisaram 38 artigos de em um período de nove anos e constataram que as principais práticas utilizadas foram as de clusterização, conforme mostra a Figura 25.

Figura 25 - Técnicas de ML mais utilizadas



Fonte: Beserra *et al.* (2014)

O Quadro 10 revela as principais técnicas utilizadas para a análise da BDE, conforme Leitner (2013).

Quadro 11 - Panorama das técnicas utilizadas em LA

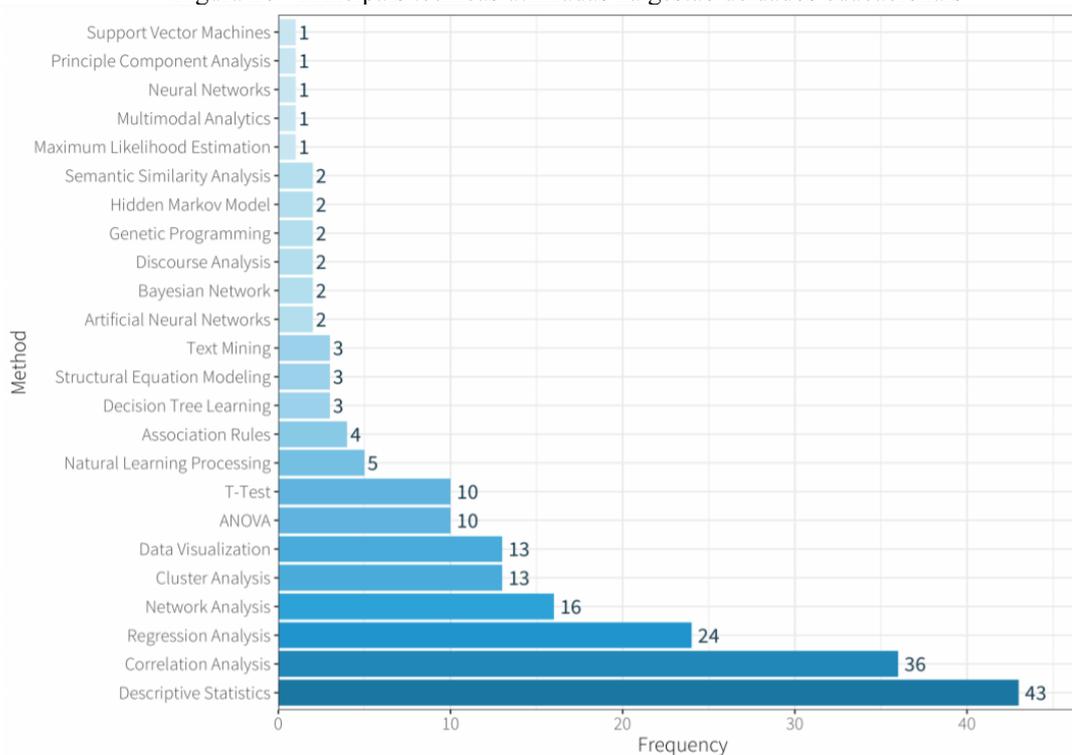
Técnicas	Aplicação
Predição	Prever a performance de estudantes Detectar comportamentos
Clusterização	Agrupar materiais similares ou alunos com base em sua aprendizagem e padrões de interação
Detecção	Detectar de alunos com dificuldades ou com processos irregulares de aprendizagem
Mineração relacional	Identificar relações nos padrões de comportamento dos alunos Diagnosticar dificuldade dos alunos
Análise de Redes Sociais	Interpretar a estrutura e as relações nas atividades colaborativas e interações com ferramentas de comunicação
Mineração de processos	Refletir sobre o comportamento do aluno em relação às tarefas de avaliação, frequência de acesso, caminho de aprendizagem.
Mineração de textos	Analisar o conteúdo de fóruns, chats e documentos.
Destilação de dados para julgamento humano	Ajudar instrutores a visualizar e analisar as atividades em curso do estudantes e o uso da informação
Descoberta por modelos	Identificar relações entre comportamentos e características dos alunos ou variáveis contextuais. Integrar estruturas de modelagem psicométrica em modelos de aprendizado de máquina

Gamificação	Incluir possibilidades para aprender jogando, ajudando a manter a motivação. Integração de realizações, pontos de experiência ou emblemas como indicadores de sucesso
Machine Learning	Encontrar automaticamente insights ocultos nos dados (com base em modelos que Estão expostos a novos dados e se adaptam independentemente)
Estatística	Analisar e interpretar dados quantitativos para a tomada de decisões

Fonte: Leitner *et al.* (2013)

Outro estudo mais recente de Misiejuk e Wasson (2017), elencou as principais técnicas utilizadas na análise de dados educacionais, conforme mostra a Figura 26:

Figura 26 - Principais técnicas utilizadas na gestão de dados educacionais



Fonte: Misiejuk e Wasson (2017)

Nota-se que o uso de estatística descritiva foi o método mais comum, seguido por análise correlacional, análise de regressão e análise de redes. Pode-se, portanto, afirmar que a maioria dos estudos sobre LA ainda focam na análise descritiva e análise diagnóstica. Esse resultado acontece em virtude da dificuldade da aplicação de técnicas mais complexas, que na maioria das vezes, foge do conhecimento do pesquisador. Por isso a formação de equipes

multidisciplinares é recomendada nesse tipo de pesquisa. Aliar os conhecimentos da pedagogia com a ciência de dados pode ser o caminho para análises mais aprofundadas sobre esses fenômenos.

Cada método tem suas vantagens e desvantagens. Muitas vezes procuramos respostas simples para questões pontuais. Isso não quer dizer que o uso de estatística descritiva resulte em análises superficiais. Esse é o método mais propício para analisar o uso do ambiente virtual, por exemplo, se sua métrica for focada na **eficiência**.

Caso a métrica escolhida seja a **efetividade**, o uso da análise de variância, ANOVA e Teste-T são mais adequados para explicar construtos mais complexos como a relação entre tempo online, número de visitas, frequência de participação, engajamento, uso dos materiais do curso etc...

Entretanto, se a métrica escolhida foca nos **objetivos**, o uso de técnicas de mineração de dados (classificação e clusterização) pode ajudar na explicação da performance acadêmica do aluno.

4 MOODLE

O Moodle (*Modular object-orientes Dynamic Learning Enviroment*) é uma plataforma de aprendizagem online criada para auxiliar educadores, administradores e alunos, baseada em um sistema robusto, seguro e integrado, que dá possibilidades de criar ambientes virtuais de aprendizagem personalizados.¹⁷

Atualmente, o Moodle é a plataforma de aprendizagem mais utilizada em escala global (230 países), com uma comunidade de mais de 213 milhões de usuários e cerca de 90 mil sites registrados, desde escolas, colégios, universidades e até mesmo empresas do setor público e organizações privadas.

O projeto foi criado em 2001 pelo educador e cientista computacional Martin Dougiamas, com o objetivo de auxiliar educadores e o processo de aprendizagem de forma geral, através de uma plataforma simples e extremamente customizável, permitindo que sua configuração mudasse de acordo com as necessidades de cada escola.

Além disso o Moodle também constitui-se de um sistema de administração de atividades educacionais para a criação de comunidades online com grande ênfase na aprendizagem colaborativa em ambientes virtuais.

Uma das grandes vantagens do Moodle é o fato dele ser uma ferramenta gratuita e de código aberto (*Open Source*), a qual se beneficia através de uma larga e ativa comunidade online de desenvolvedores que tiram dúvidas e dão suporte para os usuários. Isso inclui a criação de diversos plug-ins, que aumentam a funcionalidade do Moodle e facilitam diversas tarefas tanto para alunos quanto para professores. Essa comunidade do Moodle, por exemplo, traduziu o programa para mais de 120 idiomas, aumento o alcance da ferramenta em nível mundial. Outra vantagem de ser uma plataforma aberta, é que as instituições não ficam presas a um determinado *software*, que pode ser descontinuado a qualquer momento.

Além disso, o Moodle possui uma alta capacidade de ser escalável para qualquer tamanho, ou seja, oferece uma boa experiência desde turmas pequenas até cursos com milhões de usuários. Por isso nesta tese optamos pela análise de dados educacionais de MOOCs que utilizam o Moodle.

¹⁷ https://docs.moodle.org/39/en/About_Moodle

Os três principais utilizadores do Moodle são: administradores, professores e alunos. O Quadro 11 mostra quais são os papéis de cada um desses grupos:

Quadro 12 - Os 3 perfis do moodle e suas atribuições

ADMINISTRADOR	PROFESSOR	ALUNO
Gerir utilizadores	Configuração da disciplina	Recursos
Definir modelos de autenticação	Gestão de alunos	Atividades
Programar cópias de segurança automáticas	Gestão de grupos	Bloco administração
Gerir disciplinas e as suas categorias	Gestão de cópias de segurança	
Gerir idiomas	Análise de relatórios	
Gerir módulos (atividades e blocos)	Gestão de escala de notas	
Gerir página inicial	Análise de notas dos alunos	
Gerir aparência do site	Gestão de sistema de arquivos/ficheiros	
Aceder a relatórios	Acesso a fórum de professores	
Instalar novos blocos de atividades	Acesso a tarefas efetuadas pelos alunos	
Editar aparência dos temas		
Atualizar a versão do Moodle		

Fonte: moodle.org

Sua capacidade modular permite que os cursos sejam configurados em diversos formatos, entre eles:

- Formato Social – Muito utilizado em MOOCs conectivistas, onde a criação do conhecimento se dá através da interação dos alunos nos fóruns e chats. Nesse formato, o professor publica um tema e há uma ampla discussão em torno desse assunto, com intervenções de tutores e professores. Pode ser integrado com o uso de redes sociais e ferramentas de comunicação, videoconferências etc..
- Formato semanal – é um modelo de curso mais fechado, com conteúdo programático dividido em semanas e com datas de início e fim bem definidas. Nesse formato, os alunos vão avançando gradativamente conforme os novos conteúdos forem liberados.

- Formato em tópicos – onde cada assunto elaborado pelo professor representa um tópico, que fica à disposição dos alunos sem um tempo pré-definido. Nesse formato não há turmas, os alunos vão se cadastrando livremente. Esse formato é muito utilizado em xMOOCs, cursos de educação permanente e cursos auto instrucionais com grande quantidade de alunos.

Na verdade, o Moodle oferece uma gama de possibilidades na montagem de um curso online, cabendo à equipe de designers instrucionais e professores utilizarem a criatividade na elaboração das aulas. Isso reflete diretamente no aumento de motivação e engajamento dos alunos. O fato de permitir a integração com páginas da web e outras ferramentas de comunicação, faz com o que o Moodle suporte a comunicação síncrona ou assíncrona, contribuindo para um padrão elevado tanto no ensino presencial, quanto no ensino a distância.

4.1 RECURSOS E ATIVIDADES DO MOODLE

No Moodle, o material do curso pode ser um recurso ou uma atividade. Recurso é um item que o estudante vê, escuta, lê ou faz o *download*. Já uma atividade é um item que o estudante interage, ou possibilita a interação com o professor ou outros estudantes. O Quadro 12 mostra os principais recursos e atividade do Moodle.

Quadro 13 - Lista de recursos e atividades do Moodle

RECURSOS	ATIVIDADES
Base de dados	Arquivos para download
Chat	Livro
Diário	Pasta
Escolha	Página web
Fórum	Link
Glossário	Vídeos
Laboratório de avaliação	
Lição	
Pesquisa	
Questionário	
Tarefa	
Wiki	

Fonte: Rice (2015)

Atualmente o Moodle se encontra em sua versão 3.8.3, atualizada em junho de 2020, e como principal inovação, apresentou uma significativa melhora em sua experiência mobile, assim como ferramentas de *learning analytics*, melhorias nos fóruns, integração H5P e uso de emojis.

4.2 BASE DE DADOS

A interação dos alunos com esses recursos e atividades ficam armazenados em Tabelas que compõem o banco de dados do Moodle. Praticamente todo clique que o aluno dá no ambiente virtual pode ser monitorado pelo sistema e fica guardado como um log de acesso. Multiplicando esses cliques pelo número de alunos, chegamos a uma massa de dados muito ampla e complexa de ser analisada. A complexidade do sistema pode ser vista no fluxograma (Apêndice 6) elaborado por Marcus Green, que contempla todos os módulos e itens que formam a estrutura da base de dados do Moodle.

Para facilitar a extração de dados do Moodle a plataforma fornece uma série de relatórios que podem ser baixados. A seguir apresentaremos algumas dessas funcionalidades:

- **Relatórios de logs** - O relatório Logs fornece ao administrador do site a capacidade de gerar um relatório que inclui uma lista de todos os usuários no site Moodle, ou apenas um em particular, e detalhes de suas interações com o site durante um determinado período de tempo.
- **Estatísticas** - Se as estatísticas estiverem ativadas para o site por meio de Recursos avançados no menu Administração do site Moodle, gráficos e Tabelas de atividade do usuário poderão ser gerados. Os gráficos e Tabelas de estatísticas mostram quantas ocorrências ocorreram em várias partes do seu site durante vários períodos de tempo.
- **Relatórios Live Logs** - O relatório *Live logs* fornece ao administrador do site a capacidade de visualizar de maneira muito simples e rápida as interações recentes (Criar, Visualizar, Atualizar e Excluir) dos usuários do site Moodle.
- **Regras de monitoramento de eventos**: fornece ao administrador do site a capacidade de gerar notificações automáticas quando um evento específico

ocorre no site Moodle. Isso é feito criando uma regra, que gera uma notificação quando o evento ocorre.

Através desses relatórios, já é possível localizar e monitorar o progresso dos alunos, por meio de gráficos e Tabelas geradas pelo sistema. Entretanto, esses dados são úteis para realizar análises descritivas e diagnósticas, sobre acontecimentos passados. Para realizar análises preditivas e prescritivas, as estratégias adotadas são mais complexas e envolvem técnicas avançadas de ciência de dados.

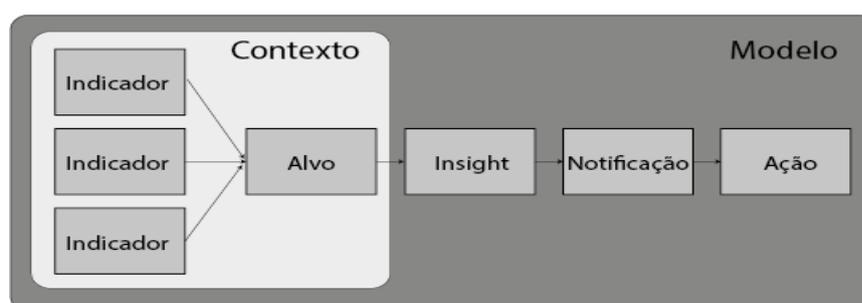
4.3 INSPIRE ANALYTICS

O Moodle passou a disponibilizar em sua última versão uma nova ferramenta chamada *Inspire Analytics*, que permite definir modelos de predição que combina indicadores e um alvo. O Alvo é o evento que queremos prever e os indicadores são os dados que podem proporcionar uma predição acertada do alvo.

O sistema pode avaliar esses modelos e, se a precisão da previsão for alta o suficiente, o Moodle treina internamente um algoritmo de aprendizado de máquina usando cálculos com base nos indicadores definidos nos dados do site. Uma vez que novos dados que correspondem aos critérios definidos pelo modelo estão disponíveis, o Moodle começa a prever a probabilidade de o evento alvo ocorrer.

A ideia é notificar os administradores e tutores quando um comportamento foge do padrão esperado. Dessa forma eles podem intervir e ajudar esses alunos. O modelo de funcionamento é bem semelhante aos frameworks de Learning Analytics apresentados nos capítulos anteriores:

Figura 27 - Modelo de funcionamento da ferramenta de LA do Moodle



Fonte: Learn2Analyze (2019)

Como exemplo de indicadores podemos citar:

- Número de postagens do aluno no fórum
- Número de recursos vistos pelo aluno
- Número de atividades realizadas pelo aluno
- Média de tentativas para passar em uma prova
- Frequência de logins na plataforma

Por ser um recurso novo no Moodle, o *Inspire Analytics* apresenta apenas dois modelos prontos: um para localizar estudantes com risco de evasão e outro para detectar a ausência de atividade de ensino. Porém o sistema pode adicionar futuramente novos modelos, de acordo com a necessidade dos usuários.

Convém lembrar, entretanto, que para criar um modelo efetivo, ele deve ser treinado previamente com dados históricos, ou seja, com dados de cursos anteriores. Através dos dados armazenados no seu banco de dados do Moodle, o sistema realiza um processo de treinamento de dados e, quando alcança uma certa maturidade, começa a fazer as previsões para futuros cursos.

Para a criação desses modelos, os desenvolvedores utilizaram os preceitos da teoria do Modelo de comunidade de investigação de engajamento dos alunos, proposta pelos filósofos C.S Pierce e John Dewey, a qual enfatiza que o conhecimento é necessariamente incorporado em um contexto social e, portanto, requer um acordo intersubjetivo entre os envolvidos no processo de investigação de legitimidade. Em 1996 Randy Garrison, da Universidade de Alberta, expandiu os conceitos dessa teoria para o contexto da aprendizagem online, elaborando uma espécie de guia heurístico para educadores que utilizam o meio online. Esse guia se baseia em três pilares principais: presença cognitiva, presença social e presença de ensino, conforme mostra a Figura 28 teoria essa que vem sendo implementada ao longo dos anos.

Figura 28 - Framework conceitual da comunidade de Investigação



Fonte: Garrison *et al.* (2015)

Nesse *framework*, o domínio “presença social” refere-se à habilidade dos participantes se identificarem com os outros integrantes do grupo de estudo, estabelecendo uma boa comunicação e até mesmo laços afetivos. No ambiente virtual do Moodle, essa capacidade é avaliada examinando as oportunidades que o curso oferece aos alunos para desenvolverem essa comunicação. Para mensurar, foi criada uma escala de 1 a 5, que gradua os níveis de interação:

1. O aluno não interagiu com nenhum outro participante (ex: apenas leu a página)
2. O aluno interagiu com pelo menos 1 participante (ex: deu um feedback através de um quizz)
3. O aluno interagiu com múltiplos participantes (ex: postou em um fórum, wiki ou base de dados)

4. O aluno interagiu com outros participantes, estabelecendo uma comunicação mais aprofundada, com várias perguntas e respostas.
5. O aluno interagiu com outros participantes em atividades extraclasse (ex: formando uma comunidade de prática)

Já o domínio “Presença cognitiva”, busca medir até que ponto os participantes em qualquer configuração específica de uma comunidade de investigação são capazes de construir significado através da comunicação sustentada. Para isso são analisados os tipos de atividades oferecidas nos cursos e o nível de engajamento cognitivo nessas atividades. Uma escala de 1 a 5 também foi criada para poder mensurar o construto, onde:

1. O aluno viu a atividade em detalhes.
2. O aluno inseriu algum conteúdo na atividade.
3. O aluno viu algum *feedback* de um instrutor ou outro participante do curso.
4. O aluno deu um *feedback* para um instrutor ou outro participante do grupo.
5. O aluno realizou uma revisão e reenviou conteúdos em uma atividade.

Por ser uma nova ferramenta, o *Inspire Analytics* apresenta algumas limitações em seu uso:

- Esse modelo de previsão pressupõe que os cursos tenham datas fixas de início e término e não foi projetado para ser usado com cursos de inscrição sem interrupção (característica de muitos MOOCs e cursos de educação permanente)
- Esse modelo requer o uso de seções nos cursos, para dividir todas as atividades em intervalos de tempo.
- Cursos com datas de início e término com mais de um ano de intervalo não serão utilizados.
- Este modelo requer uma certa quantidade de dados no Moodle para fazer previsões. Atualmente, apenas as atividades principais do Moodle estão incluídas no conjunto de indicadores. Os cursos que não incluem várias atividades principais do Moodle por "intervalo de tempo" terão pouco suporte preditivo nesse modelo.

Dado o exposto, para realmente fazer uma boa análise da Big Data Educacional dos cursos em Moodle, somente a utilização do *Inspire Analytics* em seu atual estágio de desenvolvimento, não chega perto da capacidade analítica proporcionada por outros métodos de ciência de dados. Por isso, nesta pesquisa, utilizaremos outras ferramentas complementares.

5 PROCEDIMENTOS METODOLÓGICOS

Para atender o objetivo geral que é elaborar um *framework* conceitual de *Learning Analytics*, baseado em evidências, que oriente alunos, educadores, pesquisadores e gestores de MOOCs a escolherem as intervenções mais assertivas para melhoria da aprendizagem, serão utilizados diferentes procedimentos metodológicos, que podem ser agrupados conforme as seguintes etapas:

Quadro 14 - Procedimentos metodológicos da pesquisa

Etapas	Procedimento	Resultado
1: Definição do problema de pesquisa e metodologia	Elaborado a partir da vivência e reflexão do autor	Delimitação da pesquisa
2: Revisão narrativa	Seleção, análise e interpretação de artigos e livros sobre o tema	Fundamentação teórica
3: Revisão Bibliográfica Sistemática (RBS)	Seleção e categorização das principais obras sobre o tema, contidas nas principais bases de dados acadêmicas.	Análise dos <i>frameworks</i> existentes e levantamento de lacunas na bibliografia
4: Estudo de caso: MOOC Auriculoterapia	Análise da estrutura do curso e dos dados qualitativos e quantitativos por meio de técnicas de LA	Proposição de melhorias para o curso e insights para a elaboração do <i>framework</i> conceitual
5: Elaboração do <i>Framework</i> beta	Sistematização dos resultados encontrados nas etapas anteriores, utilização da metodologia ADDIE e entrevista com <i>stakeholders</i>	Primeira versão do <i>framework</i>
5: Verificação e validação com <i>stakeholders</i>	Trabalho em conjunto com <i>stakeholders</i> e cientistas de dados	Considerações para melhoria e refinamento do <i>framework</i>
6: Refinamento do <i>Framework</i>	Correção e implementação das considerações levantada na etapa anterior	Versão final do <i>framework</i>

Fonte: Elaborado pelo autor (2019)

Na primeira etapa (definição do problema) levou-se em conta o interesse e as experiências profissionais e acadêmicas do pesquisador. Foi a continuidade de uma pesquisa iniciada no mestrado e que acompanhou a evolução das pesquisas sobre o tema proposto.

Na etapa 2 (Revisão narrativa) realizou-se uma revisão de literatura sobre os três pilares conceituais da tese: MOOCs, *Learning Analytics* e Big Data Educacional. Foi uma etapa necessária, uma vez que não é possível interpretar, explicar e compreender uma determinada realidade sem um referencial teórico (Trivinos, 2010). Ela se difere da etapa 3 pois não

apresenta um grande aprofundamento no tema, constitui-se apenas uma análise básica da literatura, interpretação e análise crítica do pesquisador, com propósito de descrever o estado da arte de um assunto específico, sob o ponto de vista teórico (Rother, 2007 p.9)

5.1 A QUARTA TRADIÇÃO METODOLÓGICA DA PESQUISA EDUCACIONAL

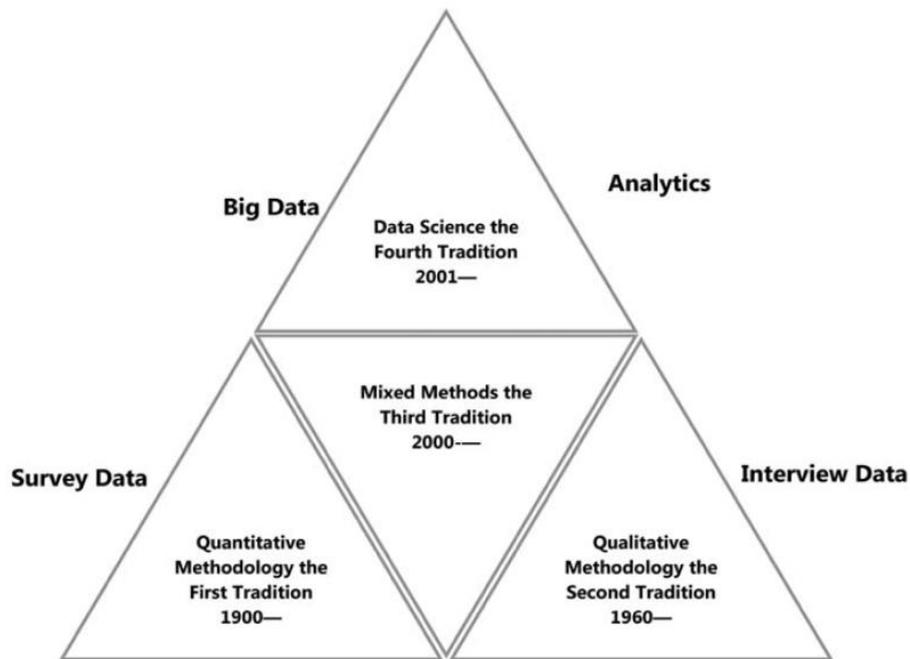
Determinar os procedimentos metodológicos mais adequados para alcançar os objetivos propostos nessa pesquisa não foi uma tarefa fácil. O fato de trabalhar com bancos de dados e métodos estatísticos dá a entender, em um primeiro momento, que a pesquisa será de cunho quantitativo. Entretanto, esses dados fazem parte de um escopo maior de análise que é a melhoria da aprendizagem. E quando se trata de educação e aprendizagem, não há como deixar os aspectos qualitativos de lado.

Baseado na literatura, buscou-se em trabalhos correlatos novas formas de analisar esse cenário, através de artigos e livros de autores que pesquisam esse tema com a utilização de uma metodologia mais apropriada. O artigo "Preparando a próxima geração de pesquisadores de educação para a Big Data no ensino superior" levantou algumas reflexões pertinentes:

A pesquisa em ciências sociais, educação, psicologia e humanas ainda é dominada por metodologias de pesquisa que basicamente dividem o mundo em abordagens qualitativas ou quantitativas. Este kit de ferramentas relativamente pequeno para entender fenômenos complexos no mundo limita a próxima geração de pesquisadores em educação quando se deparam com o aumento da disponibilidade decorrente da Big Data (Gibson & Ifenthaler, 2017 p.29)

Ben Daniel, pesquisador renomado na área de Big Data e educação também alertou para as mudanças metodológicas inerentes dessa vasta quantidade de dados existentes. O autor classificou a ciência de dados como a quarta tradição de metodologia de pesquisa educacional, conforme mostra na Figura 29.

Figura 29 - Evolução da tradição em pesquisa educacional



Fonte: Daniel (2015)

Segundo Daniel, a ciência de dados faz parte de uma evolução natural dos métodos de pesquisas educacionais, que iniciaram nos anos 1900 com ênfase quantitativa e positivista, com dados coletados por questionários. Por volta de 1960, essas pesquisas passaram a ter uma abordagem mais qualitativa, com entrevistas em profundidade para ajudar a entender melhor os temas. A partir dos anos 2000, o uso de métodos mistos ganhou destaque, de forma que os pesquisadores começaram a diversificar as formas de coletar e analisar dados educacionais. Com o advento da banda larga e a disseminação das tecnologias de informação e comunicação (TIC), houve um aumento exponencial da oferta de cursos a distância. Essa mudança de paradigma originou o fenômeno da BDE e também alterou as formas de pesquisar a educação em si. Da mesma forma a aquisição de dados foi facilitada, sua análise, entretanto, ficou muito mais complexa e além das capacidades humanas de codificar tamanha variedade e quantidade. Trabalhar com BDE requer um entendimento universal das teorias científicas para inferências indutivas (Frické, 2015). Implica abraçar novas formas de empirismo que transcendem as tradições qualitativas e quantitativas (Kitchin, 2014).

Uma das promessas do Big Data em ambientes educacionais é permitir um novo nível de pesquisa baseada em evidências sobre aprendizagem e instrução e tornar possível obter visão altamente detalhada sobre o desempenho dos alunos e suas trajetórias de

aprendizagem como necessário para personalizar e adaptar o currículo e a avaliação (Shum & Ferguson, 2011).

Essa nova forma de metodologia educacional, portanto, se utiliza da exploração analítica dos dados e é confrontada com a hipótese do pesquisador. As duas formas devem ser complementares e podem gerar um resultado, onde a primeira fase (quantitativa) propõe um modelo e a segunda abordagem (qualitativa) faz a validação. Segundo Mayer-Schönberger e Cukier (2013), a análise correlacional da pesquisa em Big Data pode muitas vezes gerar conexões úteis para o desenvolvimento das intervenções, mesmo na ausência de causalidade. No entanto, tratar erroneamente a correlação como causalidade pode resultar na escolha de intervenções ineficazes, mesmo que tal resultado seja baseado na análise de um grande conjunto de dados. Isso ressalta a importância de dosar bem os métodos, não acreditando cegamente no que os números mostram, nem confiando plenamente nas inferências do pesquisador, a fim de evitar que a pesquisa seja comprometida:

O que estamos defendendo é um equilíbrio do impulso criativo com validação externa, ambos com maior engajamento da comunidade profissional global e estabelecimento de pesquisas que sejam mais abertas, transparentes e passíveis de escrutínio, atendendo aos critérios de reprodutibilidade e generalização (Gibson & Ifenthaler, 2017 p.39)

Aderindo esse procedimento metodológico, respeita-se o rigor científico e a reprodutibilidade imposta pelos dados, mas também se consideram as subjetividades e individualidades inerentes do processo de educação:

A próxima geração de pesquisadores precisa entender essa mudança e suas implicações para significância em novas pesquisas e abraçar uma “terceira via” de pensamento sobre a integração e fortalecimento de ambas perspectivas (qualitativa e quantitativa) de um programa de pesquisa por meio de modelagem intensiva computacional, métodos de visualização e análise exploratória de dados. (Gibson & Ifenthaler, 2017 p.31)

5.2 REVISÃO BIBLIOGRÁFICA SISTEMÁTICA

Na etapa 3 realizou-se uma revisão bibliográfica sistemática (RBS) baseada no RBS *Roadmap* proposto por Conforto, Amaral e Silva (2011) e no Protocolo para Busca Sistemática de Literatura, da UFSC (2012). A revisão sistemática utiliza métodos rigorosos e explícitos para identificar, selecionar, coletar dados e descrever as contribuições relativas à pesquisa (Cordeiro

et al., 2007). Trata-se de uma revisão analítica transparente e reprodutível, que compreende as seguintes etapas: a) planejamento; b) execução; c) relatório de pesquisa (Tranfield, Denver e Smart, 2003).

Na etapa de planejamento foram estabelecidos os critérios de busca e inclusão dos artigos, como tópicos de pesquisa, tipo de documento, período de tempo, idioma. A partir dos tópicos foi definido o *string* de pesquisa. Na etapa de execução, os resultados decorrentes da busca passaram por três filtros:

- 1) leitura título, resumo e palavras-chave para verificar pertinência ao tema, e disponibilidade de acesso (nesta ordem);
- 2) leitura da introdução e conclusão;
- 3) leitura completa dos artigos.

Por fim, o relatório de pesquisa apresenta o estado da arte enumerando os principais temas encontrados, os procedimentos relatados nos artigos e suas principais características.

Optou-se por analisar artigos científicos, revisões científicas, capítulos de livros, livros e trabalhos publicados em eventos científicos indexados na base de dados SCOPUS¹⁸. Como os MOOCs surgiram em 2012, foram considerados apenas artigos publicados a partir desse período até o presente. Foram considerados os documentos publicados nos idiomas espanhol, inglês e português e, nenhuma restrição de território de publicação. Os seguintes tópicos foram delimitados para a busca: *MOOCs*; *Learning Analytics*. A seguinte estratégia de busca foi elaborada:

"MOOCS*" AND *learning analytics*

Delimitou-se os campos para a pesquisa a fim de encontrar de forma mais específica e objetiva os documentos que abordassem o conteúdo em questão, sendo estes: título, resumo e palavras-chave.

Na busca realizada no dia 25 de fevereiro de 2019, foram encontrados 252 (duzentos e cinquenta e dois) documentos para os critérios estabelecidos: 110 trabalhos publicados em

¹⁸ www.scopus.com

O termo MOOC foi o mais aparente, uma vez que é o tema central da pesquisa. Todos os artigos escolhidos na revisão tratam de cursos massivos online. O termo "análise de dados" também aparece com destaque, já que envolve o escopo da tese que é analisar a Big Data Educacional de MOOCs. *Learning Machine* também aparece com grande destaque e Figura como a principal técnica utilizada para trabalhar com os dados. A preocupação com o aspecto social dos MOOCs também emerge, já que diversos trabalhos buscam entender como se dá a aprendizagem através dos fóruns e redes sociais. O termo "evasão" também aparece com bastante regularidade, assim como "padrões" e "sucesso". Com menos relevância, surgem os termos "prever", "propor", "detectar", "analisar", "relação" e "avaliar". Curiosamente estas são ações que o LA oferece na análise de dados educacionais.

5.2.2 *MACHINE LEARNING*

Mesmo com a elaboração da nuvem, foi realizada uma mineração de palavras a fim de encontrar os temas e as linhas de pesquisas mais abordadas nos últimos anos. O *machine Learning* foi o conjunto de técnicas mais utilizada pelos pesquisadores. ML também é sinônimo de inteligência computacional e é uma área emergente e interdisciplinar que possibilitou avanços em diversas áreas como robótica, redes sociais, comércio eletrônico, medicina e saúde e também na área educacional. Através dos inúmeros algoritmos de aprendizagem, os dados são processados e os resultados apresentam um grau maior ou menor de confiabilidade. Existem milhares de algoritmos o que dificulta a tarefa de escolher o algoritmo mais apropriado. Na área educacional, esses algoritmos estão sendo utilizados para identificar padrões de comportamento de alunos de MOOCs, por meio de técnicas de clusterização. O computador processa os dados de alunos com comportamentos similares e os divide em categorias. Outra aplicação é para identificar alunos com tendência a abandonar o curso ou descobrir padrões de engajamento de alunos. As aplicações são inúmeras e cada vez mais surgem pesquisas empíricas demonstrando o potencial dessa técnica. A grande dificuldade, entretanto, é saber qual algoritmo utilizar, uma vez que existem milhares de opções. Segundo o prof. Pedro Domingos¹⁹, da Universidade de Washington, é preciso ter em mente que todo algoritmo de ML é composto de somente três componentes:

¹⁹ <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>

- **Representação:** o fenômeno a ser estudado deve ser representado em uma linguagem formal, ou seja, interpretável por um computador. Quando se escolhe um algoritmo, significa que está sendo escolhido um conjunto de modelos que podem ser aprendidos a partir daquela representação. Esse conjunto é denominado espaço de hipótese.
- **Avaliação:** essa função é útil para distinguir modelos bons de modelos ruins. Elas podem ser funções internas e externas.
- **Otimização:** método para encontrar, dentre os diversos modelos disponíveis, aquele que maximize o seu desempenho.

Essa dificuldade de encontrar o algoritmo mais assertivo é muito comum e acaba por dificultar o avanço das pesquisas em LA. Segundo George Siemens, muitos pesquisadores da área são educadores (75%), e isso implica que quando se deparam com o uso de tecnologias da ciência da computação, ou até mesmo estatística avançada, os pesquisadores encontram barreiras. Isso é um sinal da fragmentação da comunidade acadêmica e indica uma necessidade clara de ampliar esforços em busca da interdisciplinaridade. Atualmente as pesquisas em LA que apresentam mais resultados concretos são justamente conduzidas por pesquisadores das áreas da ciência da computação, justamente pela facilidade que eles têm de aplicar técnicas de *machine learning*, redes neurais, inteligência artificial, entre outras. Eles não dominam as questões pedagógicas, mas conhecem profundamente as ferramentas tecnológicas e os meios de aplicá-las na Big Data.

5.2.3 APRENDIZAGEM COLABORATIVA

Outro aspecto bem frequente nas pesquisas sobre LA é a busca pelo entendimento da aprendizagem colaborativa em MOOCs, por meio de fóruns e redes sociais. Os MOOCs são divididos em duas categorias: xMOOCs e cMOOCs, conforme o Quadro 14.

Quadro 15 - Diferenças entre os formatos de MOOCs

XMOOCs		CMOOCs
Escalabilidade de oferta	Massivo	Comunidades e conexões
Acesso aberto licença restrita	Aberto	Acesso e licença aberta
Aprendizado individual em única plataforma. Uso de vídeos, manuais e resolução de problemas.	Online	Aprendizado em rede através de múltiplas plataformas e serviços (<i>blogs</i> , posts, imagens, diagramas e vídeos)
Adquirir conhecimentos e habilidades para o currículo. Avaliação e certificados.	Curso	Desenvolver práticas compartilhadas, conhecimento e compreensão

Fonte: adaptado de Yuan *et al.* (2014)

Grande parte das pesquisas que utilizam LA buscam traçar padrões de aprendizagem por meio do conteúdo e engajamento de alunos participantes dos fóruns dos MOOCs. Lembrando que nos MOOCs conectivistas o conhecimento emerge da interação e do desenvolvimento de práticas compartilhadas, os autores buscam utilizar ferramentas tecnológicas para entender como isso acontece. Sendo assim, alguns trabalhos tentam identificar padrões de comportamentos de alunos, baseados no conteúdo que eles disponibilizam e da assiduidade das participações. Os dados gerados são minerados, através de técnicas de mineração de palavras, por exemplo. As metodologias de estudo nessa área são as mais diversas.

Os primeiros artigos com aplicações de LA para análise de fóruns e redes sociais surgiram em 2015. Chounta *et al.* (2015) utilizou métricas e análise de redes para entender como acontece a troca de conhecimentos nesses ambientes. Klusener e Fortenbacher (2015) e Milligan (2015) utilizaram algoritmos de ML para determinar o engajamento dos alunos. No mesmo ano, Ezen-Can (2015) também sustentou suas pesquisas utilizando ML e complementando a pesquisa com análise qualitativa.

Em 2016, Atapattu e Falkner (2016) elaboraram um framework para geração automática de tópicos em fóruns de cursos massivos, através de técnicas de ML. Gašević *et al.*

(2019) focaram na análise do conteúdo propagado nas redes sociais, *Social Network Analysis* (SNA) e também utilizaram insights para complementar o estudo. Já O'Riordan *et al.* (2016) propuseram 4 métodos para avaliar de conteúdo em fóruns.

Cohen *et al.* (2019) mudou a estratégia e utilizou a mineração de dados educacionais para entender como acontece a aprendizagem colaborativa em fóruns. Estudos mais recentes na área apresentam uma abordagem diferente. Moreno-Marcos *et al.* (2018) elaboraram uma técnica de análise do sentimento através de algoritmos de ML. Essa metodologia foi denominada 3S (*Social, sentiments e skills*). Wu (2018) propôs uma análise através da taxonomia comportamental, minerando dados dos fóruns e associando a sentimentos.

Apesar das diversas técnicas aqui apresentadas, estudos nessa área ainda estão em pleno avanço e movimentam a comunidade acadêmica que pesquisa sobre conectivismo, MOOCs e que buscam novas forma mais assertivas de entender a aprendizagem nesse contexto.

5.2.4 ENGAJAMENTO E EVASÃO

Outro tema que tem despertado bastante interesse dos pesquisadores é a questão da evasão e do engajamento dos alunos. Uma das principais características negativas dos cursos massivos foi, e ainda é, justamente os altos índices de alunos que abandonam o curso. As taxas de alunos que concluem todas as atividades e obtém os certificados é muito pequeno em relação ao número de alunos que se inscrevem para começar o curso. Portanto, diversas linhas de pesquisa tentam achar explicações para esse comportamento, seja implementando práticas pedagógicas, como monitorando os alunos, enviando comunicados e *emails* para atrair sua atenção e fazer com que ele retorne para a plataforma. Clow (2013) já chamava a atenção para as altas taxas de evasão dos MOOCs e a necessidade de implementar técnicas de LA para entender o fenômeno.

Já em 2014, Coffrin *et al.* (2014) iniciou um debate, relatando uma série de abordagens de LA para analisar a performance e o engajamento de alunos de MOOCs. Tseng *et al.* (2015) aplicaram de técnicas de clusterização e encontraram três padrões de engajamento de alunos em MOOCs. Sinclair e Kalvala (2015) elaboraram um *framework* para medir o engajamento, o desenvolvimento e avaliação da aprendizagem. Ainda nesse ano, Ferguson e Clow (2015) também utilizaram análise de clusters para estabelecer uma relação entre a pedagogia aplicada nos cursos e o nível de engajamento dos alunos. Além disso, os pesquisadores aplicaram

técnicas de clusterização para traçar sete padrões de engajamento dos alunos tanto em xMOOCs quanto cMOOCs.

As pesquisas progrediram, até que Wachtler *et al.* (2016) introduziu o uso de vídeos interativos em MOOCs buscando aumentar o engajamento dos alunos. No mesmo anos, Sinclair e Kalvala (2016) elaboraram um sistema que faz recomendações aos professores/gestores dos MOOCs, sobre o engajamento de seus alunos, a fim de prever o comportamento. Liang *et al.* (2016) publicou um artigo no qual analisa a predição de evasão nos cursos da EDX, utilizando algoritmos de classificação de *machine learning*.

Vitiello *et al.* (2017) apresentou uma nova abordagem ao buscar entender as causas da evasão de alunos em MOOCs e diferenciar os alunos que intencionalmente abandonam o curso e os alunos que tem dificuldades para continuar os estudos, ao passo que Srilekshmi *et al.* (2017) buscava diferentes técnicas de *machine learning* para prever a evasão de alunos.

Em 2018, Khalil e Ebner (2018) pesquisaram sobre a correlação entre engajamento e evasão assim como Sunar *et al.* (2018) focou as pesquisas no engajamento social dos alunos. Isidro *et al.* (2018) e Alexandron *et al.* (2018) continuavam a busca pelo algoritmo mais assertivo para prever a evasão. Gardner e Brooks (2018) propuseram um modelo para diminuição da evasão, baseado em um estudo onde foram aplicadas técnicas de LA, testes estatísticos e algoritmos específicos. Itani *et al.* (2018) publicaram um estudo sobre a evasão de alunos em MOOCs, que utiliza técnicas de *machine learning* como algoritmos preditivos e algoritmos explicativos para gerar intervenções automatizadas e personalizadas.

Apesar das pesquisas serem muito similares em seus objetivos, elas gradativamente consolidam o uso de técnicas como MDE e ML na busca por padrões de alunos, oferecendo à comunidade acadêmica a descrição dos métodos utilizados e publicando resultados para o crivo de outros pesquisadores, demonstrando o rigor científico necessário para poder determinar quais algoritmos são adequados para análise de BDE de MOOCs.

5.2.5 COMPORTAMENTO E PADRÕES DE APRENDIZAGEM

Outras linhas de pesquisa bastante relevante são as que abordam técnicas para descobrir comportamentos e padrões de aprendizagem dos alunos de MOOCs. Autores como Mukala *et al.* (2015), Liu *et al.* (2017), Gupta e Sabitha (2018) e Thankachan (2018) utilizaram técnicas de mineração de dados educacionais para prever comportamentos e propor soluções, como sistemas de tutoria inteligente.

Wang *et al.* (2017) buscou formas de usar o LA para entender o comportamento dos alunos em MOOCs, dando ênfase nos dados coletados da web social dos estudantes.

Shridharan *et al.* (2018) e Lau *et al.* (2018) focaram suas pesquisas na análise comportamental dos alunos ao assistirem as videoaulas, monitorando os dados através de técnicas de regressão.

Van den Beemt *et al.* (2018) explorou a relação entre o comportamento do aluno e o seu desempenho ao longo do curso. Por meio das técnicas de clusterização e mineração de processos, descobriram quatro perfis comportamentais de alunos de MOOCs. Chen *et al.* (2018) elaboraram um sistema de visualização analítica (Viseq) com objetivo de resolver a perda de informações, melhorar a visualização da sequência de aprendizagem de diferentes grupos e entender as razões por trás dos comportamentos de aprendizagem em MOOCs.

Outros artigos apresentaram a utilização de métodos mistos. Xu *et al.* (2018) fizeram uma revisão de um *framework* de aprendizagem online. A partir de um estudo qualitativo e quantitativo, utilizando técnicas de LA, os autores identificaram quatro tipos de comportamentos de aprendizagem que podem servir como preditores para a aprendizagem em MOOCs. McLaren *et al.* (2018) apresentaram uma pesquisa com 957 alunos de um MOOC na área da saúde, e através de métodos mistos e uso de LA, buscaram entender o comportamento dos alunos durante o curso.

5.2.6 PERFORMANCE DE APRENDIZAGEM

A performance de aprendizagem dos alunos também foi alvo de pesquisas ao longo dos últimos anos. Já em 2014, Coffrin *et al.* (2014) publicaram um artigo no qual debatiam uma série de abordagens de LA para analisar a performance e o engajamento de alunos de MOOCs. Com o passar dos anos, mais estudos empíricos foram publicados, descrevendo técnicas mais apuradas e resultados mais conclusivos. Conijn *et al.* (2018) utilizaram correlações, regressões múltiplas e mineração de processos para entender quais dados dos MOOCs podem ser usados para implementar a performance de aprendizagem.

Tan *et al.* (2018) utilizaram técnicas de LA (*k-means cluster analysis e stepwise linear regressions*) e identificaram 3 perfis de aprendizagens em um MOOCs de língua inglesa, com potencial de melhoria na performance dos alunos. Bystrova *et al.* (2018) utilizaram um algoritmo de LA para prever a performance dos alunos em MOOCs. Rahmani e Demmans

Epp (2018) analisaram a relação entre a personalidade dos alunos e sua performance acadêmica em MOOCs. Por meio de algoritmos de *machine learning*, os autores buscaram relacionar a colaboração e fatores sociais no impacto na aprendizagem.

5.2.7 APRESENTAÇÃO E *DASHBOARDS*

Não basta uma boa análise de dados se não conseguirmos apresentá-los corretamente. Sendo assim, o desenvolvimento de ferramentas de visualização e *dashboards*²⁰ também foi uma área que chamou atenção dos pesquisadores. Cobos *et al.* (2016) apresentaram uma proposta de *dashboard* de visualização para MOOCs, com objetivo de facilitar a interpretação de dados de alunos e professores. A Figura 31 mostra um protótipo de *dashboard* para MOOCs:

Figura 31 - Protótipo de dashboard elaborado pelo MOOC Research Project



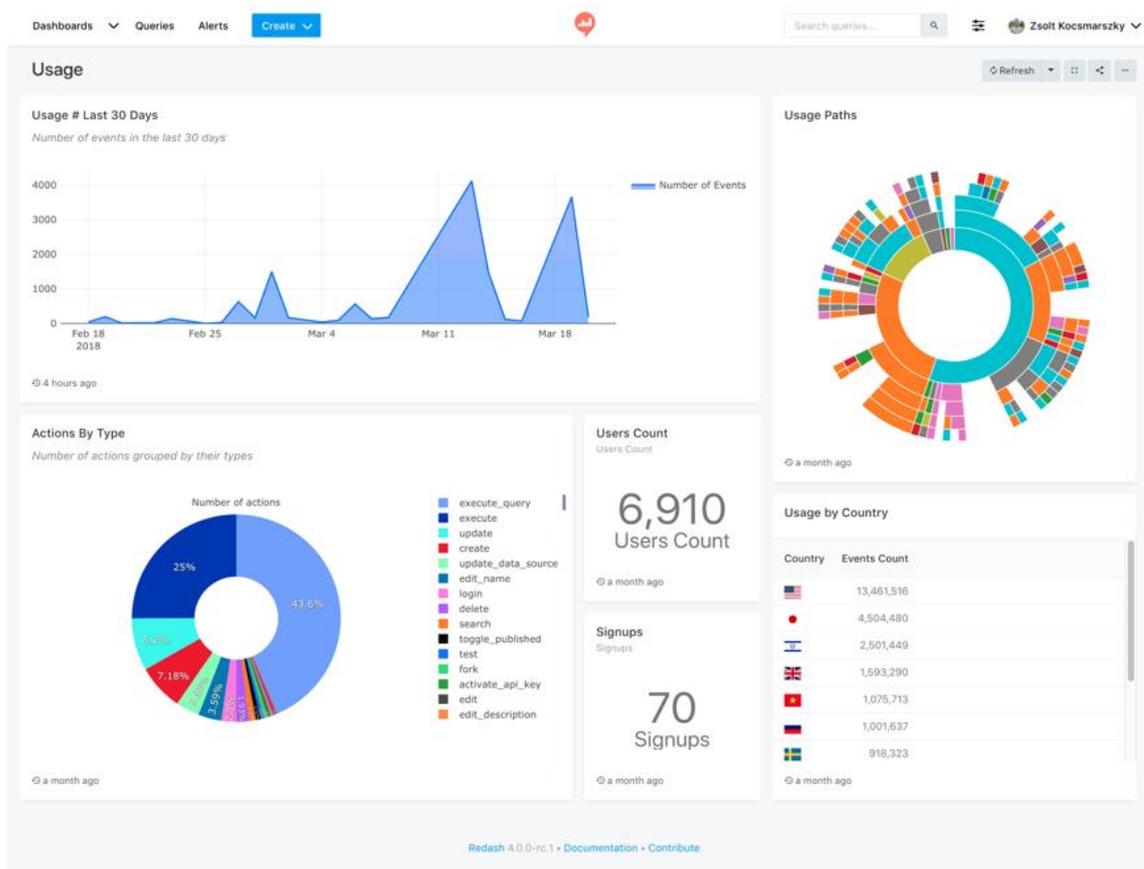
Fonte: Kristin Stephens-Martinez²¹

Lopez *et al.* (2017) detalham o funcionamento de uma ferramenta de LA chamada bigquery²², fornecida pelo Google. Essa ferramenta analisa enorme quantidade de dados e oferece *dashboards* de visualização, conforme mostra a Figura 32.

²⁰ *Dashboards*, também conhecidos como painéis, têm um propósito muito específico de exibir as informações mais importantes necessárias para alcançar um ou mais objetivos; consolidados e organizados em uma única tela para que as informações possam ser monitoradas rapidamente.

²¹ <https://users.cs.duke.edu/~ksm/MOOC-dashboard.html>

²² <https://cloud.google.com/bigquery/>

Figura 32 - Modelo de *dashboard* gerado pelo Google Big Query

Fonte: Google

León-Urrutia e Tang (2017) apresentam o funcionamento de um *dashboard* de visualização de dados dos MOOCs da Future-Learn, em tempo real. Ruiperez-Valiente *et al.* (2017) sugerem que as ferramentas de LA para MOOCs precisam ser melhoradas e ajustadas. Para isso ele propuseram um sistema chamado ANALYSE, e aplicaram em dados de grandes provedores de MOOCs, dando diretrizes de como trabalhar com grandes quantidades de dados, facilitando a visualização.

5.2.8 PERSONALIZAÇÃO

A massividade, uma das características principais dos MOOCs, é responsável pelo alcance global desses cursos. Entretanto, ao tratar da mesma forma todos os alunos, sem levar em consideração suas peculiaridades e experiências anteriores, a massividade pode se tornar um problema. Desde o surgimento dos MOOCs essa foi uma questão levantada pela

comunidade acadêmica. A utilização dos dados gerados pelos alunos pode gerar insights interessantes para a personalização de conteúdos, de tal forma que o MOOC se adapte melhor ao estudante.

Khribi *et al.* (2015) apresentaram a ideia de um framework para sistemas de recomendações personalizados baseados nos dados gerados pelos alunos. Parfenov e Zaporozhko (2018) sumarizaram as abordagens e técnicas utilizadas para promover a personalização de MOOCs através do processamento de dados de big data, mineração de dados educacionais e LA. El Mawas *et al.* (2018) abordaram a necessidade de personalização de conteúdos para aumentar as taxas de conclusão dos MOOCs, e para isso utilizaram técnicas de LA.

Por fim, Itani, *et al.* (2018) abordou a evasão de alunos em MOOCs, utilizando técnicas de *machine learning* como algoritmos preditivos, algoritmos explicativos para gerar intervenções automatizadas e personalizadas.

5.2.9 TEMAS SECUNDÁRIOS

Outros temas que emergiram na RBS também servem como indicadores para compor o estado da arte da relação entre MOOCs e LA. Prinsloo e Slade (2015) realizaram uma análise profunda sobre a privacidade e as questões éticas impostas pelo uso do LA. Abordaram a questão do consentimento dos alunos e a necessidade de manter a impessoalidade.

A padronização dos dados e o debate sobre quais técnicas de LA são mais assertivas também originou a publicação de diversos artigos. Inclusive há uma série de pesquisadores que acreditam que esse campo de pesquisa só irá avançar se houver uma padronização de técnicas e de dados. Para isso cunharam o termo *Open Learning Analytics*, que é uma comunidade de pesquisadores que levantam essa bandeira. Pardos *et al.* (2016) apresentou uma plataforma com objetivo de integrar o LA de diversos MOOCs.

Um fator interessante na pesquisa sobre LA é justamente o caráter interdisciplinar que envolve o tema. Nota-se claramente quando as pesquisas são lideradas por educadores ou por profissionais da área tecnológica. As abordagens são completamente diferentes. Enquanto que os educadores buscam entender os aspectos pedagógicos dos MOOCs, pesquisadores de TI apresentam estudos muito mais "complexos" em termos de metodologia e ferramentas. Há estudos que utilizam Redes Neurais, ontologias, web semântica, sistemas de recomendação e até mesmo inteligência artificial. Esses artigos propõem sistemas, *frameworks*, elaboração de

algoritmos de ML, com objetivo de apresentar resultados mais conclusivos. É interessante ver como as diversas áreas de conhecimento estudam o mesmo fenômeno, e a aproximação cada vez maior dos cientistas de dados na área educacional certamente impulsionará a qualidade dessas pesquisas.

Alguns estudos relacionando LA com o design instrucional dos cursos também apareceram na revisão sistemática. Eradze e Tammets (2017) apresentaram um estudo de caso da aplicação de um sistema de LA que oferece *feedback* instantâneo para alunos e professores e sugere recomendações para mudanças no design instrucional (DI) de MOOCs, baseados nos dados fornecidos pelo sistema. Alonso-Ramos, *et al.* (2018) também relacionam como o LA podem ajudar na elaboração de um DI mais interessante para os alunos.

Para finalizar a lista de temas que emergiram nessa RBS, Muñoz-Merino *et al.* (2015) apresentaram o único artigo que relacionava o uso de LA para analisar as atividades avaliativas (provas e questionários) de MOOCs. Rohloff *et al.* (2018) apresentaram um estudo quantitativo sobre a presença de dispositivos móveis em MOOCs, ao analisar os dados de cursos baseados em plataformas ubíquas através de técnicas de LA.

5.3 ESTUDO DE CASO

Após a fundamentação teórica e a revisão sistemática, o próximo passo foi analisar na prática os dados educacionais de um MOOC de auriculoterapia, produzido na UFSC e, para isso, foi utilizada a metodologia do estudo de caso. Essa modalidade de pesquisa vem sendo cada vez mais utilizada nas Ciências Sociais e caracteriza-se por ter como objeto de estudo uma ou mais unidades que são analisadas profundamente permitindo amplo conhecimento (GIL, 2002). Segundo Yin (1994), nessa metodologia examina-se o “caso” (ou um pequeno número de “casos”) em detalhe, em profundidade, no seu contexto natural, reconhecendo-se a sua complexidade e recorrendo-se para isso todos os métodos que se revelem apropriados. Corroborando com essa ideia, Triviños (2006) ressalta que o estudo de caso objetiva a descrição de determinada realidade e fornece conhecimento aprofundado da realidade do ambiente estudado, uma vez que os resultados encontrados podem ajudar na formulação de proposições e teorias para o estudo em outras localidades. Em estudos de caso onde uma organização é a unidade de análise, o pesquisador deve partir do conhecimento que existe sobre a unidade que deseja estudar. Esse conhecimento pode estar em arquivos, publicações e pessoas.

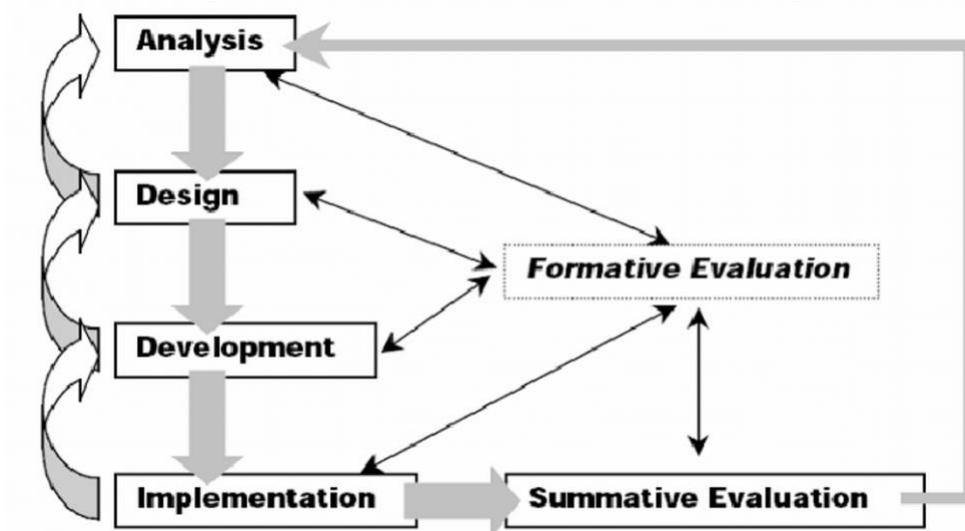
Dessa forma, diversas técnicas foram utilizadas para conhecer e entender os problemas pelo qual o curso vem enfrentando. Além do acesso ao banco de dados, para a realização das análises quantitativas, entrevistas com gestores, autores de conteúdo e equipe técnica foram realizadas. Também foram aplicados questionários específicos para cada área, a fim de localizar potenciais pontos críticos.

Todos os procedimentos adotados, assim com os resultados dessa análise podem ser encontrados no capítulo 6 desta tese, onde estão devidamente descritos e explicados.

5.4 ELABORAÇÃO DO FRAMEWORK CONCEITUAL

Na etapa 5, referente à elaboração do framework conceitual, utilizou-se como base a metodologia ADDIE revisitada (Drljaca *et al.*, 2017), que consiste em uma abordagem de cinco fases para construir soluções de aprendizagem efetivas voltadas para os ambientes eletrônicos (virtuais). Essas etapas abrangem o processo todo, desde de sua análise, design, desenvolvimento, implementação e avaliação. Tendo isso em vista, foi necessário juntar uma equipe multidisciplinar que conhecem profundamente o MOOC analisado nessa pesquisa. Essa equipe é formada por médicos, enfermeiros, pedagogos, professores, técnicos, profissionais de TI e gestores educacionais com experiência em projetos.

Figura 33 - Modelo ADDIE revisitado para ambientes digitais



Fonte: Drljaca *et al.* (2017)

Durante a fase de análise, ocorreu um grupo focal os profissionais envolvidos, no qual foram identificados os problemas instrucionais que estão acontecendo atualmente, **sempre**

referentes ao contexto de ensino e aprendizagem. Nessa fase, cada profissional identifica os pontos forte e pontos fracos, gaps causados por falta de conhecimentos e habilidades, e também relatam os objetivos desejáveis para o sucesso do curso. Segundo Drljaca *et al.*, (2017) esse é um estágio contemplativo, no qual é necessário pensar em uma abordagem centrada no aluno, tanto no design do curso quanto dos materiais. Essa fase é fundamental para conseguir começar a construir um panorama do MOOC, identificando o que dá para ser melhorado sob o ponto de vista de cada profissional.

Nessa fase é definido o público-alvo (informações educacionais dos alunos, idade, nacionalidade, experiências prévias e interesses); São definidos os objetivos e necessidades de aprendizagem; características do ambiente de aprendizagem (virtual, ao vivo, presencial, *blended learning*); conhecimentos e habilidades necessárias dos alunos; são escolhidos os objetivos instrucionais, de ensino e aprendizagem; são definidos os recursos necessários (técnico, humano, tempo, habilidades).

Dentro dessa fase de análise, temos 3 objetivos principais, denominados A1, A2 e A3.

Quadro 16 - Objetivo da fase 1 – Análise

A1	Identificar quais intervenções são necessárias para os problemas encontrados.	Identificação de problemas
A2	Visa captar onde a intervenção educacional será implementada.	Análise contextual
A3	Objetiva analisar para quem a intervenção educacional será projetada.	Análise dos alunos

Dessa forma, pode-se concluir que a fase de análise visa obter uma visão do contexto educacional que está sendo estudado, levantando informações úteis para a etapa seguinte: a fase do design.

Nessa fase 2 são estipulados os objetivos educacionais a serem alcançados, selecionada a abordagem mais apropriada para atingir esses objetivos, assim como os métodos de avaliação apropriados. Assim como na fase anterior, a fase do Design também apresenta subdivisões, apresentados a seguir:

Quadro 17 - Objetivo da fase 2 - Design

DES1	Definição do objetivo geral e dos objetivos específicos	Definição dos objetivos educacionais
DES2	Seleção da abordagem apropriada para a aprendizagem dos alunos e mapeamento dessas atividades em relação ao objetivos.	Seleção da estratégia de aprendizagem
DES3	Criação/verificação dos métodos de avaliação de acordo com os objetivos	Seleção dos métodos de avaliação

Portanto, o principal objetivo fase de design é conhecer e traçar um plano detalhado do fluxo e descrição das atividades de aprendizagem.

A fase 3 é a etapa do desenvolvimento, onde ocorre a seleção dos materiais educacionais para a entrega apropriada dos objetivos traçados na fase anterior. Hora de garantir que a equipe toda esteja envolvida no processo, garantir os prazos de preparação e criação dos materiais. Com as informações coletadas até o presente momento, já é possível colocar em prática um teste piloto e determinar, na prática, quais estratégias estão funcionando bem, através dos feedbacks e observações coletadas.

Quadro 18 - Objetivo da fase 3 - Desenvolvimento

DEV1	Desenvolvimento de recursos para o suporte das atividades de avaliação	Desenvolvimento/seleção dos recursos educacionais
DEV2	Organização da entrega apropriada de acordo com os objetivos do curso	Desenvolvimento/seleção entrega apropriada

O objetivo principal da fase de desenvolvimento é selecionar os meios mais apropriados para a produção de materiais de forma que eles atinjam os objetivos estipulados anteriormente.

A fase 4 é a de implementação. O trabalho das fases anteriores toma forma e é colocado em prática para os alunos. Verificar o feedback dos professores e alunos e preparar alternativas para possíveis erros durante a implementação. O ambiente de aprendizagem é preparado. Devem ser observados os mecanismos de suporte de aprendizagem do grupo e de forma individual também. Os cuidados para manter o engajamento dos alunos são fundamentais. Dados previamente estipulados devem ser coletados para futura análise e tomada de decisão.

Quadro 19 - Objetivo da fase 4 - Implementação

11	Entrega e publicação dos materiais fornecidos nas etapas anteriores	Entrega do material
12	Acompanhamento das ações de professores e alunos através da coleta de dados educacionais.	Monitoramento

Portanto, o objetivo principal da fase 4 é oferecer suporte e monitoramento para os alunos alcançarem os objetivos educacionais propostos. Período para analisar quais processos precisam de melhorias e quais estão funcionando corretamente.

A fase 5 é a avaliação. É o momento para a ver o processo de ensino e aprendizagem como um todo. É a hora de avaliar quais os objetivos educacionais foram alcançados. Definir quais categorias obtiveram bom desempenho, definir a política de coleta de dados (estratégias, frequência, ferramentas), configurar as ferramentas para garantia de qualidade, elaborar um relatório final das atividades.

Quadro 20 - Objetivo da fase 5 - Avaliação

E1	Processo contínuo de avaliação de todas as etapas do processo a fim de maximizar a eficácia pedagógica.	Avaliação formativa
E2	Processo contínuo de avaliação de todas as etapas do processo e sua eficiência de implementação (redução de tempo/custos)	Avaliação somativa

Nessa fase de avaliação o principal objetivo é identificar problemas e necessidades que precisam ser mudadas. Propor a melhoria do processo em si, através da identificação dos erros. Aprender com os erros e não voltar a repeti-los futuramente.

6 ESTUDO DE CASO: MOOC AURICULOTERAPIA

Para a realização desta pesquisa escolhemos como objeto de análise o Curso de formação em auriculoterapia para profissionais da atenção básica, pois ele preenche todos os requisitos: é um MOOC (objetiva a capacitação massiva e é gratuito), utiliza o Moodle, apresenta um bom histórico de dados (está na 7ª edição), além de possuir todo o processo documentado, desde sua criação. Dessa forma, esse curso mostrou um bom potencial de análise e um bom grau de amadurecimento. As etapas compreendidas neste estudo de caso percorreram o seguinte caminho:

Figura 34 - Etapas do estudo



Fonte: Elaborado pelo autor (2020)

6.1 O CONTEXTO DO CURSO

O curso de formação em auriculoterapia é um convênio entre a Universidade Federal de Santa Catarina (UFSC) e o Ministério da Saúde (MS), mais especificamente pela equipe de Práticas Integrativas e complementares (PIC). A proposta da elaboração desse curso surgiu em 2014, em virtude do destaque e a larga experiência da UFSC em ensino a distância e a relativamente exitosa, embora recente, experiência da secretaria municipal de saúde de Florianópolis, no estímulo à prática e na formação em PIC (Santos e Tesser, 2012).

Uma equipe multidisciplinar foi formada, a maioria com experiência prévia e ou com atuação na atenção básica. Essa equipe idealizou e propôs a estrutura e a concepção do curso, definindo seus conteúdos e suas estratégias. Juntamente com professores da UFSC, auxiliando nos processos pedagógicos, o ano de 2015 foi marcado por inúmeras reuniões para elaboração do conteúdo, que ficou dividido da seguinte forma:

Quadro 21 - Estrutura do curso de auriculoterapia

ETAPA A DISTÂNCIA					ETAPA
MÓDULO 1	MÓDULO 2	MÓDULO 3	MÓDULO 4	MÓDULO 5	PRESENCIAL
Introdução ao curso e à auriculoterapia	Reflexologia da orelha	Medicina tradicional chinesa e auriculoterapia	Evidências, teorias biomédicas e efeitos adversos	Usos da Auriculoterapia na atenção básica	Aula Prática

Fonte: Relatório do curso de auriculoterapia/UFSC

Algumas opções pedagógicas foram assumidas pela equipe:

- Houve opção por uma estratégia de aproximar ao máximo o curso do espírito e do ideário da educação permanente em saúde.
- A definição do público-alvo do curso foi indicada pelo Ministério da Saúde: profissionais de saúde de nível superior da atenção básica, incluindo equipes do NASF, consultórios de rua, de atendimento domiciliar, academia da saúde e similares.
- Opção por contextualização da auriculoterapia na medicina tradicional chinesa (MTC), na reflexologia e na biociência (neurofisiologia, evidências, sinais de alarme e efeitos adversos).

- Opção por contextualizar a auriculoterapia no processo difícil, desafiador e lento, mas progressivo, de inserção das PIC no ambiente da atenção básica, especialmente na ESF.
- Opção por um formato totalmente auto instrucional na etapa EAD, em que o aluno estuda sozinho por meio dos materiais didáticos disponibilizados sem necessidade de aulas, tutorias, tarefas ou aulas por web conferência em horários predeterminados durante a etapa EAD.

Como objetivos de aprendizagem, a equipe pretendia alcançar as seguintes metas:

- Promover a expansão da auriculoterapia no SUS, de forma participativa e multidisciplinar;
- Capacitar profissionais que compreendam o indivíduo como sujeito do processo de promoção e cuidado de sua saúde, e visualizá-lo como agente corresponsável no processo de cuidado e promoção da saúde;
- Conhecer princípios fundamentais da MTC, especialmente da Auriculoterapia, e integrar esses conhecimentos de forma aos cuidados biomédicos;
- Dominar a técnica de auriculoterapia e utilizá-la como ferramenta terapêutica em diversos Quadros clínicos comuns na atenção primária;
- Incluir a auriculoterapia como terapia complementar nos atendimentos individuais, atendimentos em grupos, acolhimento e autocuidado da equipe;
- Discutir com a equipe e planejar a inserção da auriculoterapia na prática de sua unidade de saúde;
- Contribuir de forma proativa para a expansão sustentável das PIC no SUS.

A primeira edição do curso foi disponibilizada para os alunos através do Moodle. Os cinco módulos (de 15hs cada) possuía uma apostila e um conteúdo em telas (no formato SCORM) com narração e recursos interativos. Havia um apresentador virtual que ia apresentado os materiais. Ao final do processo havia uma avaliação com questões múltipla escolha que liberavam o certificado, em caso de aprovação (nota 7 ou superior). Com essa estratégia, cerca de 1500 alunos foram capacitados, mas inúmeras mensagens de alunos relatando alguns problemas como material didático fizeram com que a equipe do projeto revisasse o conteúdo.

Para a segunda edição, todo o material didático foi reformulado. Passou-se a utilizar como principal recurso videoaulas de 5 a 12 minutos, produzidas com uma linguagem jornalística mais dinâmica. Algumas melhorias nas apostilas também foram realizadas. O Moodle também foi a ferramenta utilizada para disponibilizar os materiais, que eram liberados conforme a aprovação do aluno. Ao final de cada módulo, havia uma prova de 10 questões, onde a nota 6 ou superior, permitia o avanço para a etapa seguinte.

6.1.1 ETAPA PRESENCIAL

Terminada as 5 etapas a distância, iniciava-se a etapa presencial: 5hs de aula, com exercícios práticos e troca de conhecimentos entre os alunos. As etapas presenciais das várias edições do curso ocorreram em polos regionais e foram ministradas por preceptores previamente capacitados pela equipe gestora do curso, sempre oferecida nas semanas posteriores ao término da Etapa EAD, conforme viabilidade e características de cada local. As turmas eram compostas de, no máximo, 60 alunos, divididos em grupos menores. Esse número decorreu de que tais aulas foram conduzidas por dois professores do curso e um monitor (na maioria das vezes, um acupunturista ou auriculoterapeuta gentilmente disponibilizado pelo município polo), e sua dinâmica foi organizada de modo que cada um dos três orientava sete grupos de três alunos na realização, uns nos outros, das atividades práticas da aula. As aulas foram conduzidas conjuntamente pelos dois professores do curso que se deslocavam até o polo regional.

Devido à necessidade logística identificada durante a 1ª edição, antes da 2ª edição houve uma expansão do número de professores das etapas presenciais por meio de uma capacitação presencial de 16 horas de profissionais acupunturistas e ou auriculoterapeutas realizada na UFSC em 16 e 17 de março de 2017, que aumentou de 8 para aproximadamente 24 os professores capacitados. Tal capacitação envolveu a organização de um evento na UFSC, em que foi realizada uma oficina de auriculoterapia com alunos de graduação, aberta aos estudantes de todos os cursos da saúde (com 160 vagas, todas preenchidas, divididas em 8 turmas de 20 graduandos).

6.1.2 PROCESSO DE CERTIFICAÇÃO

A certificação dos participantes aconteceu de forma automática, de acordo com as regras do curso. Os participantes precisavam ter êxito na realização dos cinco módulos da etapa EAD e obrigatoriamente participar da aula presencial. Os participantes que atenderam estes requisitos, quando iniciou o período de certificação, receberam seu certificado por meio de e-mail. A qualquer momento o certificado pode ser acessado por meio do sistema de certificados (<http://certificados.ufsc.br>), utilizando as mesmas credenciais necessárias para efetuar a inscrição e realizar os estudos na etapa EAD.

6.2 AVALIAÇÃO DO CURSO PELOS ALUNOS

Ao final da etapa presencial de cada edição do curso, como última atividade desse encontro, foi aplicada uma ficha de avaliação do curso a ser preenchida pelos alunos, de forma voluntária e sem identificação, que permitiu um *feedback* sobre o curso no ato da sua conclusão.

Os comentários e manifestações nas questões abertas da avaliação ao final da aula presencial, ao mesmo tempo em que foram elogiosos no geral, registraram sugestões, demandas e algumas reclamações dos estudantes. Cabe destacar que as manifestações dos alunos refletiram a evolução e melhoria na qualidade da etapa EAD no seu material didático e a reformulação do ambiente virtual de aprendizagem. Os resultados dessa pesquisa podem ser visualizados no Apêndice 2.

Além desse questionário ao final do curso, a equipe do projeto realizou uma pesquisa online com 5793 alunos matriculados, sendo que 2982 (52%) responderam. Dessa pesquisa, que originou o artigo “Capacitação em auriculoterapia para profissionais do sus em 2016-2017: perfil dos participantes, do seu envolvimento no curso e percepção sobre a prática”²³ pôde-se traçar um perfil dos alunos capacitados, o que ajuda na elaboração das métricas de *Learning Analytics*.

Do total de respondentes, 86% eram mulheres, com idade média de 36,8 anos. Sobre o perfil profissional, 35% eram enfermeiras, 14% fisioterapeutas, 8% psicólogas e 7% médicas. A taxa de conclusão do curso foi de 95% e cerca de 73% colocaram em prática em sua rotina profissional os conhecimentos obtidos no curso. Ou seja, 3/4 dos alunos começaram a praticar a auriculoterapia, um número bastante satisfatório (Tesser, C. D. & et al., 2020)

²³ <https://www3.ufrb.edu.br/seer/index.php/revise/article/view/1769/1113>

Note, entretanto, que nos anos posteriores ao artigo (2018 e 2019), mais de 5 mil novos profissionais foram capacitados pelo curso de auriculoterapia, número que vem aumentando exponencialmente a cada ano que passa. Dessa forma, novas pesquisas são necessárias para avaliar o impacto e o significado da inserção em larga escala da auriculoterapia na prática dos profissionais (Tesser, C. D. & et al., 2020)

6.2.1 COLETA DE DADOS QUALITATIVOS: APLICAÇÃO DO QUESTIONÁRIO COM INTEGRANTES DO PROJETO

Uma etapa importante para a elaboração de uma estratégia de LA é conhecer em detalhes o curso que será analisado, entender como funcionam as rotinas, quantos profissionais são envolvidos em sua criação, quais papéis cada um deles desempenha, detectar pontos fracos e fortes do curso. De forma geral, elabora-se um panorama do curso, aplicando uma espécie de engenharia reversa para entender como a estratégia de ensino foi montada. Em conversa com a equipe do projeto, ficou claro que existiam 3 áreas principais de atuação: área técnico-pedagógica, área tecnológica e área administrativa.

EQUIPE PEDAGÓGICA

A área técnico-pedagógica é composta por professores universitários com larga experiência em saúde pública e profissionais com experiência prática em auriculoterapia. A equipe era bem multidisciplinar, tendo médicos, enfermeiros e profissionais com perfil acadêmico. A elaboração do material didático dos 5 módulos do curso foi dividido entre os profissionais de acordo com suas áreas de expertise. Portanto, esse grupo foi o responsável pela elaboração dos materiais instrucionais, apostilas, videoaulas e avaliações.

Para entender quais demandas deram certo e o que precisa melhorar nos cursos, aplicou-se um questionário online (Apêndice 3), com questões qualitativas e quantitativas. Diversas informações relevantes emergiram.

Os principais desafios na elaboração do curso foi justamente sintetizar todo o conhecimento de auriculoterapia, agregar o conteúdo e deixar em uma linguagem simples para os alunos. Segundo o um dos autores, “Por se tratar de um curso absolutamente original, equipe com pouca experiência em EAD, tema delicado e com pouca institucionalização (fora do

ambiente acadêmico), nos exigiu muito rigor e simplicidade na elaboração do material didático”. Outro desafio foi integrar a equipe multidisciplinar de autores a fim de chegar em um consenso sobre o material. “As apostilas foram mais tranquilas de produzir, mas as videoaulas foram difíceis, sobretudo na interação com o roteirista”.

De forma geral, para os autores o conteúdo tem correspondido aos objetivos de ensino. Houve algumas correções pontuais no material, mas o feedback dos alunos tem sido muito positivo. Entretanto há uma clara percepção de que há necessidade de preparar novos materiais para os alunos egressos, com mais embasamento científico, para aumentar a legitimidade acadêmica do tema.

Alguns pontos críticos foram citados nas questões referentes ao conhecimento pedagógico, segundo relatou um dos autores: “Acho que o conhecimento pedagógico sobre ensino a distância podia ser melhorado, pois o pouco que há foi desenvolvido empiricamente pela equipe sem estudo específico teórico ou prático sobre esse tema”. Outro autor sugeriu materiais complementares para a fixação do conteúdo:

“Como um recurso complementar, poderia dispor de algum vídeo da aula para que os alunos possam acessar, caso tenham alguma dúvida posterior à aula prática. Vejo que as dúvidas que eles possuem sobre o conteúdo escrito, depois são sanadas na aula prática. A orelha interativa também é um instrumento de suporte que ajuda a tirar as dúvidas quanto à localização dos pontos. Outras dúvidas seriam em relação aos pontos utilizados para o tratamento das diversas condições. Este material já está sendo produzido nesta nova etapa do projeto”.

Questões sobre o aspecto tecnológico do curso também emergiram. Um dos autores relatou dificuldades de comunicação com a equipe de TI, dificultando as melhorias propostas. De forma geral, o conteúdo tem sido acessado facilmente pelos alunos, todos os recursos têm funcionado bem, com raras reclamações pontuais: “Em geral, o *feedback* dos alunos tem sido que os materiais didáticos estão bons e acessíveis. Há algumas poucas reclamações sobre acessibilidade, mas parecem estar relacionadas mais à internet do país e dos serviços de saúde em que os alunos do curso trabalham do que à plataforma do curso”.

Outra questão importante foi acerca dos *feedbacks* dos professores para os alunos. Desde seu início o curso foi projetado para ser massivo e autoinstrucional:

O curso é totalmente autoinstrucional - não há *feedbacks* personalizados por tutores, nem há fóruns de discussão. Há apenas na avaliação dos 5 módulos auto instrucionais da parte EAD do curso *feedbacks* sobre as questões erradas nas provas ao final de cada módulo, orientando que os alunos estudem novamente trechos específicos das apostilas. Esta opção de o curso ser 100% autoinstrucional foi uma opção inicialmente decorrente da larga escala que era almejada (e aos poucos vem sendo alcançada -

quase 10 mil formados já) e da pequenez da equipe de conteudista e da gestão/administração do curso.

Mesmo sendo um formato consolidado e unânime entre a equipe de autores, alguns alunos solicitaram um tutor durante as etapas EAD para ajudar no processo de aprendizagem. Para isso, foram criados grupos de *whatsapp* (extraoficiais) para atender algumas demandas pontuais de alunos: “A aula prática (no final) foi elaborada para que houvesse uma maior interação e que fosse possível sanar as dúvidas em relação ao conteúdo e fornecer maior segurança aos profissionais para atenderem em suas unidades. Muitos participantes/locais fizeram grupo de *whatsapp* e trocam experiências entre eles e tiram dúvidas”.

Essa criação de grupos de *whatsapp* por parte dos alunos se mostrou um fato curioso. Transpareceu que há interesse dos alunos na formação de uma rede para troca de informações e construção de conhecimentos. Segundo um dos autores, uma possível solução seja a criação de um fórum aberto, sem tutoria, para viabilizar essa troca entre os alunos, ou a formação de grupos regionais sob a supervisão de um articulador local.

Sobre o processo de avaliação, os autores consideram o processo apropriado, de forma que os alunos chegam preparados para a aula prática final. Não houve relatos sobre a necessidade de melhorias nesse sentido.

A etapa presencial é realizada em apenas um dia, com cerca de 5hs de duração. Segundo os autores, os principais desafios têm sido o pouco tempo para a realização da prática (em virtude de o grupo de trabalho ser grande) e também os aspectos de logística:

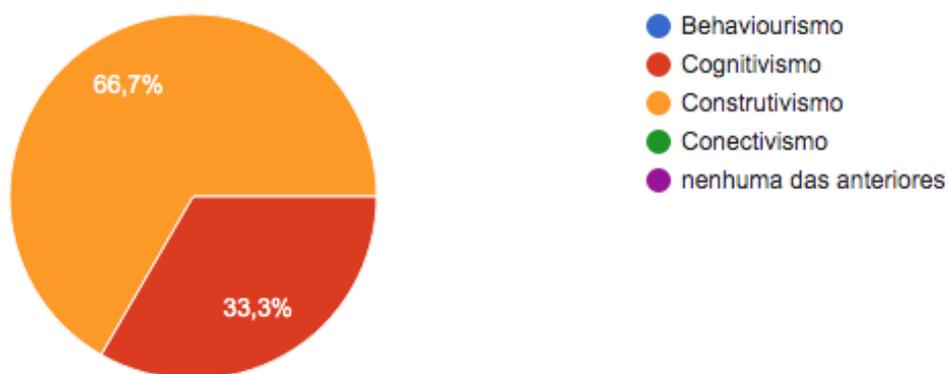
Os principais desafios são logísticos, dado que as aulas são espalhadas pelo país. A aula é relativamente tranquila. Ela consiste de uma revisão muito sintética de tópicos essenciais da parte teórica e é bem mais centrada na aquisição de habilidades psicofísicas de observação, percepção e seleção de pontos e sua aplicação, para que os alunos aprendam a fazer o mesmo.

Para os autores, a divisão em grupos menores de trabalho (três pessoas por instrutor) é fundamental para aumentar a qualidade do aprendizado na aula prática.

No quesito questões éticas sobre o armazenamento dos dados educacionais, consentimento e transparência, os autores relataram que os dados são de posse do MS e cabe a ele decidir o que fazer ao final do curso. Atualmente os principais dados se encontram no setor de TI da UFSC e somente a coordenação do curso tem acesso. Portanto, para o uso desses dados, para pesquisas acadêmicas e afins, é necessária a autorização prévia dos coordenadores.

Sobre a abordagem pedagógica adotada no curso, 66,7% dos autores alegaram que o curso de auriculoterapia utiliza estratégias do construtivismo, enquanto que 33,3% acham que o curso apresenta aspectos do cognitivismo.

Figura 35 - Gráfico de respostas sobre a abordagem pedagógica do curso



Fonte: elaborado pelo autor (2019)

Questionados sobre o que teriam feito de forma diferente no processo de criação do curso, os autores apresentaram respostas bem diversas: “Pelas avaliações dos participantes, vejo que estão confiantes e utilizando a prática no seu dia a dia na unidade de saúde. Talvez repensar sobre a MTC ampliando a abordagem. Na parte prática é onde os participantes apresentam maior dúvida pela complexidade do conteúdo”. Outro autor considera a aula prática o ponto forte do curso, e sugere inserir mais videoaulas com professores na etapa EAD. Já outro autor sugere mudanças no curso e no aspecto de gestão do grupo em geral:

Eu repensaria a questão de ser puramente autoinstrucional, ainda que apenas após as primeiras edições. De igual eu faria o esforço para que o curso tivesse um esquema de gestão e decisão o mais colegiado possível, pois apesar dos efeitos adversos (maior tempo para tomada de decisões, maior chance de voltar atrás em decisões etc) me parece que vale a pena.

Sobre o processo de avaliação e implementação de melhorias no curso, os autores não chegaram em um consenso. Para alguns, houve melhorias e correção de materiais no início: “Tivemos melhorias em relação aos vídeos após a primeira edição. O acesso às aulas também melhorou, assim como o processo de inscrição. Melhoramos a aula prática também com o passar das edições”. Outro autor corroborou:

“A avaliação é feita em reuniões e a partir também das avaliações dos participantes. Houve melhorias e correções do material, mas para essa edição de 2020 não houve alterações. Houve evolução ao longo das edições. Teve uma maior procura por parte dos polos e mais inscrições dos profissionais. Ampliou o Quadro de preceptores do curso”.

Entretanto, outro autor foi mais taxativo em relação às melhorias: “Não. A equipe se acomodou com a qualidade considerada boa do curso em depoimentos informais em congressos e por vias informais, e não tem havido melhorias a cada edição”.

Perguntados sobre a política na tomada de decisões baseadas em dados, a equipe concordou. Utilizam análise estatística, questionários aos egressos, além de relatórios semanais da equipe de TI enquanto está acontecendo a etapa EAD. Esses dados são úteis e auxiliam a equipe na tomada de decisões. Note que nenhum dado educacional do Moodle foi utilizado pela equipe ao longo dos anos!

Por fim, como sugestão, um dos autores comentou sobre a possibilidade de criar uma ferramenta que avalie o efeito do curso através dos atendimentos em auriculoterapia nos serviços de saúde onde foi oferecido o curso.

EQUIPE DE TECNOLOGIA DA INFORMAÇÃO (TI)

Outra importante parte do curso é composta pela equipe de TI, responsável pela infraestrutura e suporte tecnológico. Eles cuidam dos sistemas do curso, a base de dados, o Moodle, o sistema de inscrição dos alunos, a hospedagem do site, além de proverem os dados para os relatórios gerenciais. O questionário pode ser visto no Apêndice 4.

Questionados acerca de diversos processos de suporte tecnológico, a equipe considerou boa as seguintes atividades: Sistema de inscrição, matrícula e homologação, acesso ao Moodle, comunicação com os alunos, suporte aos alunos, suporte mobile e acesso aos dados. A avaliação *online*, a hospedagem do site e a certificação foram considerados ótimos em sua avaliação. Entretanto a tarefa de geração de relatórios foi apenas satisfatória.

Sobre as fontes de dados educacionais do curso de auriculoterapia, um dos integrantes da equipe explica:

Os dados são obtidos por diversos sistemas da UFSC e são integrados para evitar a duplicação da informação, sendo que cada sistema possui sua responsabilidade sobre os dados, por exemplo, os dados das pessoas só podem ser criados ou alterados no Sistema de Cadastro de Pessoas, ou a inscrição em um curso só pode ser realizada no sistema de Inscrições, o certificado só pode ser emitido pelo Sistema de Certificados,

no entanto só o Moodle poderá informar se o estudante possui os atributos necessários para a certificação.

Segundo os profissionais, o Moodle tem atendido plenamente as demandas do curso pois ele integra os diversos setores da UFSC e diversos sistemas computacionais. Além disso ele faz o registro de todos os logs dos alunos, gerando um conteúdo bem rico para geração de relatórios. A maioria dos dados armazenados atualmente correspondem às atividades realizadas pelos alunos e dados de avaliação (número de tentativas e notas das avaliações).

O Moodle utilizado pela equipe não utiliza nenhum *plugin* adicional, entretanto a equipe está buscando soluções para facilitar sua rotina de trabalho: “Alguns procedimentos poderiam ser automatizados, como a geração da lista de presença”. De forma geral a equipe considera que o curso está bem estruturado, não havendo necessidade de melhorias no momento.

EQUIPE ADMINISTRATIVA

São os responsáveis pelas rotinas administrativas do curso: logísticas, pagamentos, emissão de passagens, secretaria executiva e demais tarefas. Também auxiliam na elaboração dos relatórios gerenciais para o MS. O questionário utilizado com esses profissionais pode ser visto no Apêndice 5.

Quando questionados sobre a utilização dos dados para a tomada de decisões, os profissionais consideraram que os utilizam de forma consistente e apropriada. A principal fonte de dados são os registros dos alunos, que são solicitados sob demanda para a equipe de TI. Essa solicitação é de acordo com as demandas do coordenador do projeto. Os dados mais utilizados são os dados de inscrição dos alunos, alunos homologados (que cumprem todos os requisitos para realizar o curso) e os dados sobre certificação.

Cabe à equipe administrativa fazer a filtragem dos alunos homologados e efetivarem ou não sua matrícula no curso. Esse trabalho é realizado de forma manual atualmente, o que aumenta demais a demanda de trabalho nos períodos de matrícula. Segundo uma integrante da equipe, “seria interessante a elaboração de uma ferramenta que automatizasse esse processo”.

Os profissionais garantem que dão o melhor para dar assistência durante todos os processos do curso, de tal forma que não visualizam pontos de melhoria nesse momento, além dos relatados.

6.3 COLETA DE DADOS QUANTITATIVOS: EXTRAINDO DADOS EDUCACIONAIS DO MOODLE

Para realizar a análise dos dados do Moodle, inicialmente foi solicitado para a equipe do projeto uma autorização para essa finalidade. O contato foi feito diretamente com a equipe de TI da Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação, responsável pela gestão de dados dos projetos da UFSC. Foi fornecido, então, um *dump* da base de dados, ou seja, um arquivo contendo toda a base de dados do moodle do projeto.

Esse *dump* é um arquivo gigante (MYSQL) com todo o backup do curso. Fazendo uma analogia, o *dump* seria uma rocha, que precisa ser quebrada e dividida em pedaços menores. Esses pedaços são divididos de acordo com seu tamanho e importância, para posteriormente serem lapidados. E para realizar essa triagem dos dados que precisaríamos buscar, foi solicitada a ajuda de uma empresa especializada no manejo de dados educacionais, liderada pelo pesquisador e entusiasta do Moodle Daniel Neis.

Nessa fase foram realizadas diversas videoconferências para alinhar os objetivos da pesquisa e buscar incluir os dados que realmente fossem necessários. Foi definido que o trabalho seria realizado em duas frentes: extrair dados do *dump1* (referente ao projeto piloto do curso de auriculoterapia) e comparar com os dados do *dump 2* (estrutura atual do curso). Esse *dump 2* era bem mais extenso pois compreendia os dados de 86 turmas realizadas em polos espalhados por todo o Brasil.

As consultas realizadas no banco de dados podem ser vistas no Apêndice 7. Como resultado final desse processo de seleção de dados, 21 Tabelas .csv foram criadas:

Quadro 22 - Tabelas do Moodle utilizadas neste estudo

Grupos_auriculo_tentativas_por_quiz_por_curso.csv	Grupos_auriculo_cliques.csv
Auriculo_notas_cursos_1.csv	Auriculo_tentativas_por_quiz.csv
Grupos_auriculo_notas_cursos_2.csv	Auriculo_notas_nao_aprovadas.csv
Auriculo_notas_quiz.csv	Auriculo_notas_aprovadas.csv
Grupos_auriculo_notas_aprovadas.csv	Auriculo_modulos.csv

Grupos_auriculo_modulos_por_curso.csv	Grupos_auriculo_notas_nao_aprovadas.csv
Grupos_auriculo_media_tentativas_por_quiz_por_curso.csv	Auriculo_media_tentativas.csv
Grupos_auriculo_media_quiz_por_curso.csv	Auriculo_media_notas.csv
Grupos_auriculo_notas_quiz.csv	Auriculo_cursos.csv
Grupos_auriculo_cursos.csv	Auriculo_cliques.csv
Grupos_auriculo_logs.csv	

Fonte: Elaborado pelo autor

Essas Tabelas compreendem

- Dados da estrutura dos cursos (tipos de exercícios, módulos, aulas),
- Dados de avaliação dos alunos (notas e número de tentativas),
- Dados de rastreamento (*logs*).

Com isso, a etapa de extração e coleta foi concluída. O próximo passo foi organizar esses dados, limpar e de-duplicar (eliminar duplicidades). Logo no início percebeu-se que alguns dados como cargo/função e a idade dos alunos não contemplavam a base de dados do moodle. Com isso, qualquer correlação entre profissão e idade não poderia ser realizada. A localização de variáveis de dados faltantes já é um resultado positivo para a melhoria do curso, e uma indicação de ação que precisa ser implementada para as próximas edições.

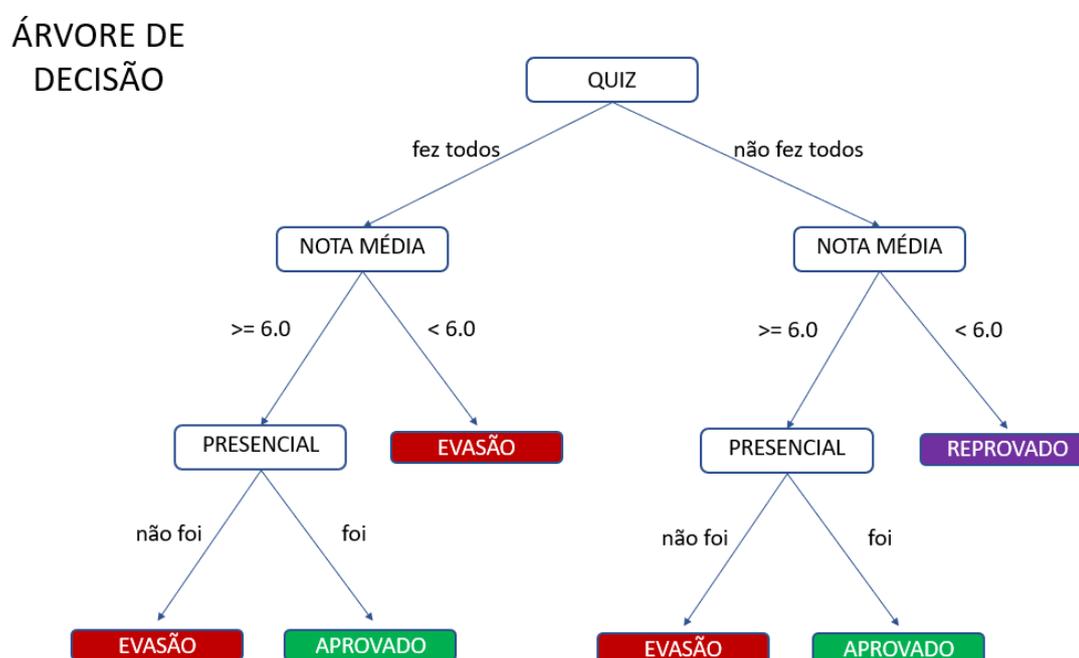
Durante a organização, outras questões foram aparecendo: itens duplicados, formatos diferentes, escalas de avaliação fora do padrão (por exemplo: notas de 0 a 10 em um curso e 0 a 100 no outro). Nessa etapa foi descoberta uma inconsistência em relação à nota final para aprovação. Para ser considerado aprovado, o aluno deveria obter, no mínimo, nota 6 em cada uma das atividades do curso. O que foi detectado, entretanto, é que o sistema estava calculando errado essa nota. O aluno era considerado aprovado caso a média das notas fosse maior que 6. Isso ficou claro ao observar que diversos alunos que tiraram notas 10 nas primeiras avaliações, nem chegaram a realizar as últimas, pois já haviam alcançado a média mínima (36 pontos). A inconsistência encontrada nesse cenário indicou a necessidade de reformular a maneira de calcular as notas do curso.

Outro ponto importante foi a determinação do conceito de evasão. Quando podemos considerar que um aluno abandonou o curso? Para isso foi definido a seguinte regra. É considerado abandono quando o aluno:

- * para curso com nota final no campo como NULL
- * NÃO fez etapa presencial (mesmo com média acima de 6)
- * se fez etapa presencial, mas não fez as provas dos módulos

Essa definição no conceito de evasão foi determinante para obter resultados mais precisos acerca desse tema que é tão comum em MOOCs. Para facilitar o entendimento dessas definições, foi criada uma árvore de decisões, conforme mostra a Figura 36 abaixo:

Figura 36 - árvore de decisão para determinação do conceito de evasão



Fonte: Autor (2021)

A unificação da base de dados em apenas uma tabela foi um trabalho complexo porém necessário. Somente com os dados organizados dessa forma foi possível iniciar a análise propriamente dita. Para isso foi utilizado o *software* Power BI, uma excelente ferramenta de análise e que possui uma interface amigável para a visualização dos resultados.

6.4 ANÁLISE DE DADOS

Com os dados devidamente organizados e limpos, estipulou-se um cronograma de 4 semanas para a realização dessa etapa. Optou-se por realizar uma análise exploratória, com uso

de técnicas estatísticas, pois se tratavam de dados de edições passadas do curso. A ideia era buscar entender bem o funcionamento dos cursos, ver se houve uma evolução entre a etapa piloto (*Dump 1*) e a etapa atual (*Dump 2*), encontrar possíveis pontos críticos que pudessem prejudicar o andamento do curso e, a partir disso, ter insights para a proposição de melhorias. A divisão dessa etapa foi feita em 4 semanas, conforme mostra o Quadro 22 a seguir:

Quadro 23 - Cronograma das atividades de análise

	<p>SEMANA 1</p> <p>Objetivo: Procurar por padrões de presença. Será que os alunos mais engajados concluem antes o curso? Será que eles possuem notas melhores?</p> <p>Tarefas: Identificar alunos que mais acessaram; Identificar os que menos acessaram; Identificar o tempo que o aluno levou para concluir o curso Determinar 2 grupos (Precoces X Atrasados) Comparar essas variáveis no <i>Dump1</i> e no <i>Dump2</i> Verificar se houve diferenças entre os cursos</p>
	<p>SEMANA 2</p> <p>Objetivo: verificar o desempenho dos alunos. Se há relação entre o número de recursos acessados com a nota final.</p> <p>Tarefas: Listar o número de recursos acessados por cada aluno Listar número de cliques Listar número de atividades realizadas Determinar 2 grupos (alunos dedicados X aluno desmotivados) Comparar os grupos e suas notas finais Comparar essas variáveis no <i>Dump1</i> e no <i>Dump2</i></p>

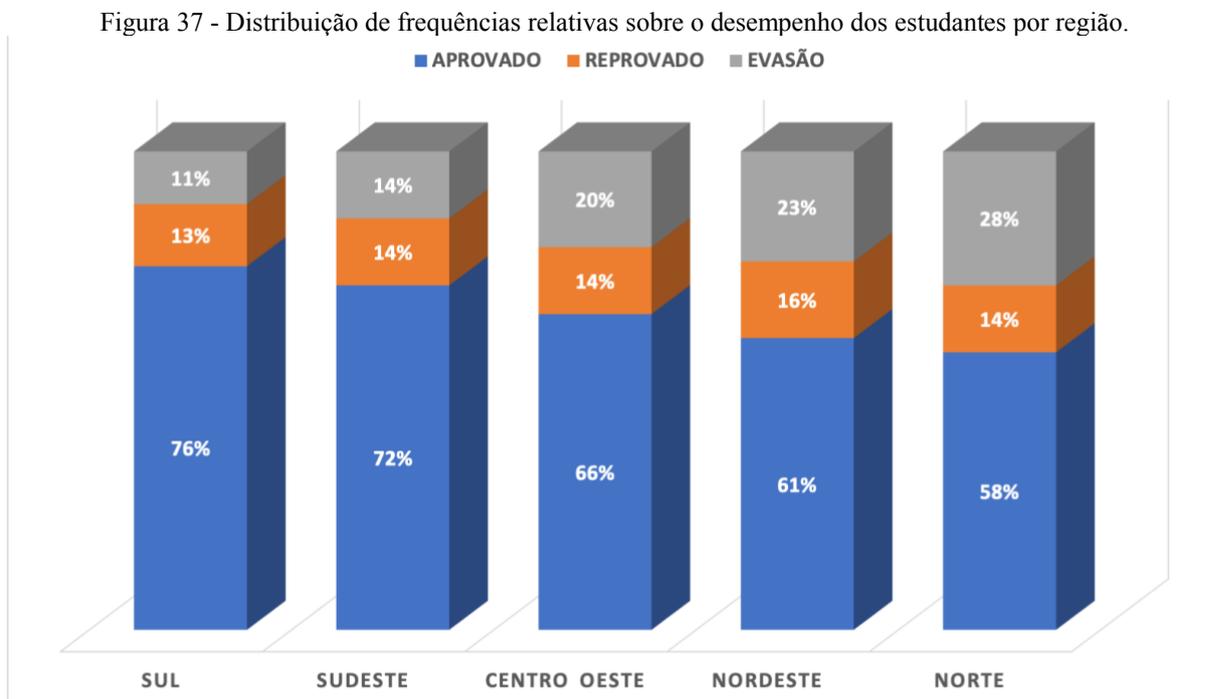
	<p>SEMANA 3</p> <p>Objetivo: verificar o desempenho dos alunos nas atividades do curso. Identificar atividades com erros, com níveis muito difíceis ou fáceis.</p> <p>Tarefas: Listar as atividades de cada curso Verificar o acesso dos alunos em cada atividade Identificar o número de tentativas na avaliação Comparar essas variáveis no <i>Dump1</i> e no <i>Dump2</i>.</p>
	<p>SEMANA 4</p> <p>Objetivo: verificar o desempenho da nota final dos alunos e sua relação com o caminho que eles trilharam no curso</p> <p>Tarefas: Identificar possíveis questões com problemas Verificar média de notas Identificar o número de tentativas na avaliação Comparar essas variáveis no <i>Dump1</i> e no <i>Dump2</i>.</p>

Fonte: Elaborado pelo autor

Pelo fato de ser difícil determinar o tempo que o aluno está logado, ou o tempo que ele demora para realizar determinadas atividades, essas variáveis temporais foram desconsideradas no estudo. Isso porque não há uma precisão e nem certeza que o aluno realmente está ali presente realizando a atividade. Muitas vezes o aluno pode logar no AVA e deixar ele aberto, mesmo não estando em frente ao computador. É muito difícil conseguir essas variáveis temporais no ensino a distância.

Para analisar os dados, foram geradas algumas visualizações no Power BI. Seguindo o cronograma proposto, o foco inicial da análise foi nos dados referentes à presença dos alunos.

De uma forma geral os alunos apresentaram bons índices de participação e taxas de evasão razoáveis, ainda mais se tratando de um MOOC, que costuma apresentar taxas de evasão muito altas. A Figura 37 mostra a distribuição de frequências relativas sobre o desempenho dos estudantes por região onde houve turmas.



Fonte: elaborado pelo autor

Chama a atenção a taxa de 28% da região norte (mesmo sendo a menor em números absolutos) e 23% da região nordeste. Índices muito superiores aos da região sul e sudeste. Conforme mostra a Tabela 1. Aqui fica um ponto que vale a investigação da equipe do projeto. Por que essas duas regiões apresentaram índices de evasão tão superiores ao restante do país? Ainda constatou-se que cerca de 16% dos alunos (2207) não chegaram a realizar nenhum dos *quizzes* propostos no curso.

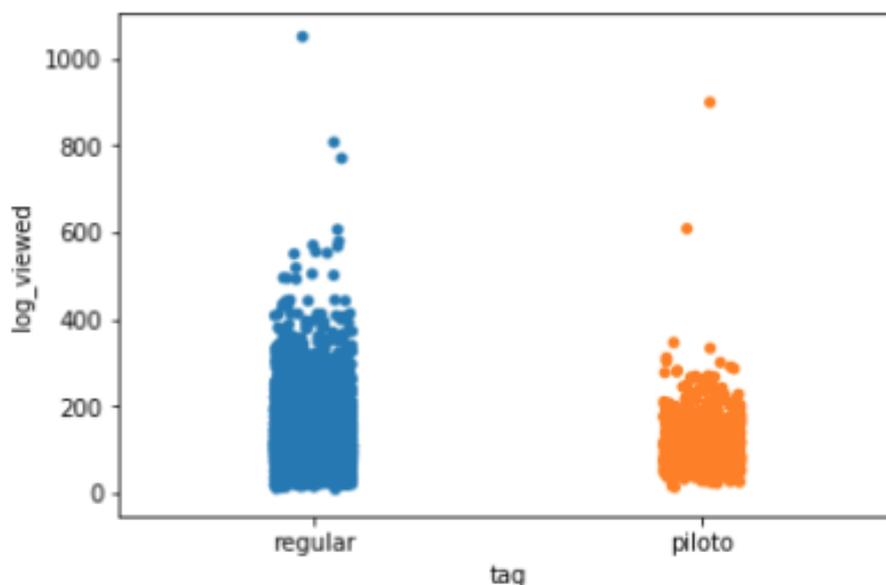
Tabela 1 - Evasão de alunos por região

	regiao	evasao	n_evasao	total	taxa
0	Centro Oeste	223	868	1091	0.2044
1	Nordeste	513	1752	2265	0.2265
2	Norte	218	551	769	0.2835
3	Sudeste	733	4317	5050	0.1451
4	Sul	520	4103	4623	0.1125
5	Total Geral	2207	11591	13798	0.1600

Fonte: elaborado pelo autor

Outro dado interessante que pôde ser observado é o número total de cliques. Fazendo um comparativo entre o curso piloto (dump1) e o curso atual (dump2) pôde-se observar que os alunos do curso atual clicam muito mais nos recursos e atividades do que os alunos da turma piloto (Figura 38). Além disso, o número de logins também é superior no curso atual, o que demonstra um maior engajamento dos alunos tanto nas atividades, quanto no acesso ao AVA.

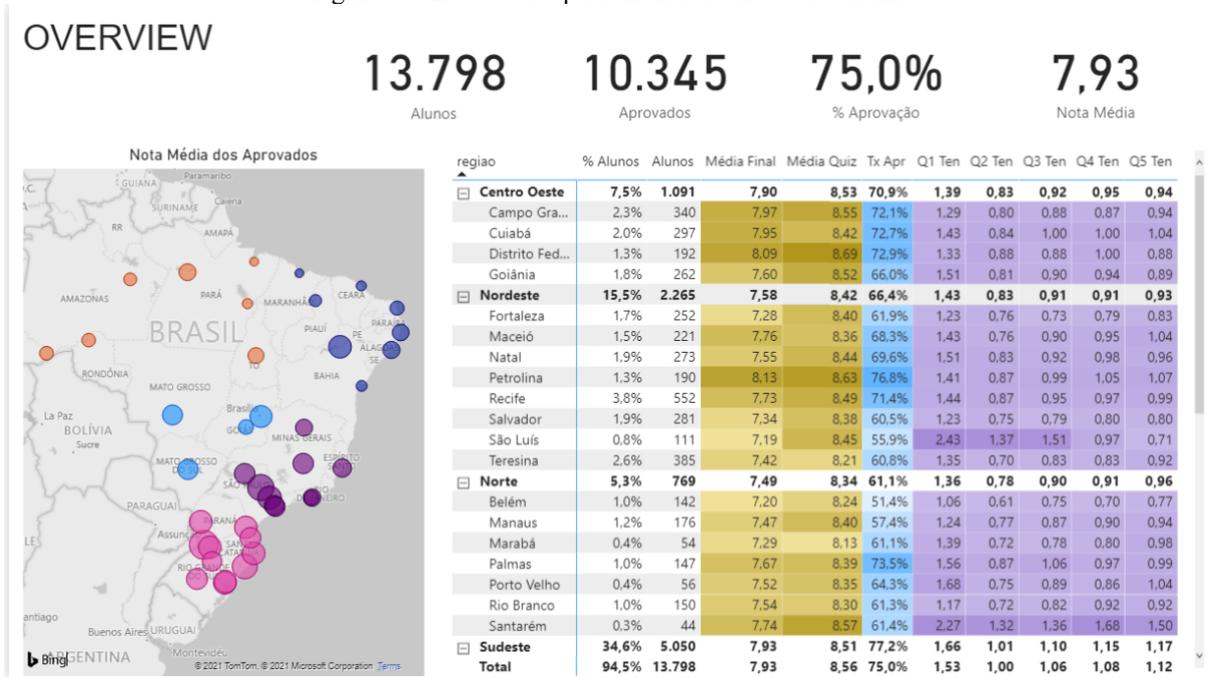
Figura 38 Número de recursos visualizados



Fonte: elaborado pelo autor

Um *dashboard* com dados das notas dos alunos foi criado, e incluído um mapa com a distribuição dos polos onde foram realizados os cursos. Segundo a Figura 38, pode-se observar o número total de alunos aprovados, o índice de aprovação e a nota média dos alunos.

Figura 39 - *Dashboard* para análise das notas dos alunos



Fonte: elaborado pelo autor

Analisando os resultados dos alunos, ainda nesse *dashboard*, pôde-se visualizar o número de tentativas dos alunos em cada um dos *quizzes* do curso, e suas notas médias. Percebe-se que o *quiz* 1 foi o que apresentou maior dificuldade entre os alunos (Figura 40)

Figura 40 - Visualização das médias das notas e o número de tentativas

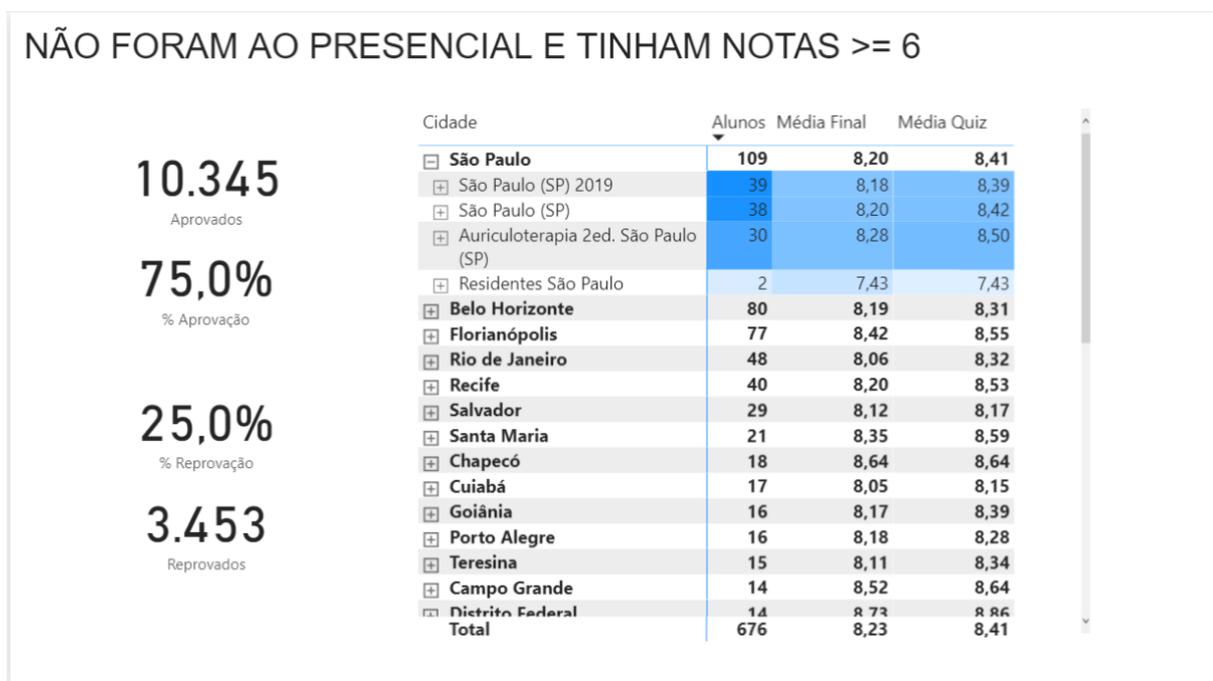


Fonte: elaborado pelo autor

Alguns testes de correlação e seus respectivos gráficos podem ser vistos no Apêndice 8 deste trabalho. O que chamou atenção é que o número de tentativas de realização de quiz e o número de recursos acessados no curso não mostra correlação com a nota final.

Com relação ao evento presencial, cerca de 84,8% dos alunos (8829) participaram dessa etapa. Para poder participar dessa etapa final do curso era obrigatório o aluno ser aprovado na etapa online, com nota igual ou superior a seis. Entretanto, um número chamou atenção. 676 alunos aptos ao evento presencial não compareceram (7,2%).

Figura 41 - Tabela com dados de alunos que abandonaram o curso na etapa final



Fonte: elaborado pelo autor

Além disso, cerca de 170 alunos que não obtiveram nota mínima na etapa online participaram do encontro presencial. Aqui ficou claro que houve uma falha na verificação dos requisitos mínimos para a participação. Ao indagar a equipe responsável, eles afirmaram que esses 170 alunos realmente realizaram a etapa presencial, porém não obtiveram o certificado de conclusão do curso. Tais dados mostraram a necessidade de haver um maior rigor e cuidado nessa seleção dos alunos aprovados.

Analisando as taxas de aprovação e reprovação do curso, constatou-se que 9676 alunos foram aprovados (83,5%) e 1915 alunos foram reprovados (16,5%). A Tabela a seguir mostra os números absolutos por região:

Tabela 2 - Número de alunos aprovados e reprovados por região

	regiao	resultado	qtd	taxa
0	Centro Oeste	APROVADO	714	0.8226
1	Centro Oeste	REPROVADO	154	0.1774
2	Nordeste	APROVADO	1.381	0.7882
3	Nordeste	REPROVADO	371	0.2118
4	Norte	APROVADO	442	0.8022
5	Norte	REPROVADO	109	0.1978
6	Sudeste	APROVADO	3.618	0.8381
7	Sudeste	REPROVADO	699	0.1619
8	Sul	APROVADO	3.521	0.8582
9	Sul	REPROVADO	582	0.1418
10	Total Geral	APROVADO	9.676	0.8348
11	Total Geral	REPROVADO	1.915	0.1652

Fonte: elaborado pelo autor

Outras questões que emergiram das análises qualitativas realizadas com a equipe do curso se referiram à mudanças na estrutura atual do curso, que é autoinstrucional. Para alguns membros, a presença de um tutor para auxiliar os alunos durante o curso seria uma alternativa interessante. Assim como a possibilidade de aumentar a interação entre os alunos, abrindo alguns tópicos no formato fórum. Essas implementações aumentaria a interatividade dos cursos, mas também aumentaria a demanda por respostas e moderação das conversas do fórum. Diante dessas demandas, talvez fosse interessante adotar um sistema de tutoria inteligente, baseado em inteligência artificial e aprendizado de máquina. Com isso, as perguntas mais frequentes seriam respondidas sempre que necessário, mitigando dúvidas dos alunos.

Outro ponto levantado seria a elaboração de uma nova metodologia para a escolha dos alunos homologados. No formato atual do curso, as vagas são limitadas. Assim que se abre o período de inscrição, quando o número de vagas é preenchido, há uma lista de espera. Como os alunos devem, obrigatoriamente, serem profissionais do SUS, muitas vezes essa triagem é feita de forma manual, o que demanda um trabalho minucioso a cada edição do curso. Como

sugestão, a adoção de um sistema baseado em *machine learning*, que analisa o perfil dos alunos com boas taxas de participação e aprovação, que possa recomendar os melhores alunos com maior potencial de sucesso que estão na lista de espera. Essa medida visa diminuir as taxas de evasão, por meio de uma escolha mais assertiva de alunos.

Além disso, outra demanda da equipe seria a configuração de um sistema que pudesse monitorar os alunos formados, fazendo uma correlação com a melhora dos atendimentos de auriculoterapia nos seus devidos municípios. Para isso, seria interessante realizar uma integração com o sistema do DATA SUS, órgão responsável pelos dados do sistema único de saúde do Brasil (dados abertos inclusive) para fazer uma correlação entre a aplicação da auriculoterapia e a melhora do índice de tratamento das enfermidades que essa técnica trata.

De uma forma geral pode-se afirmar que o curso atual supera em diversos quesitos o curso da etapa remota. O formato, baseado em uma variada quantidade de recursos e vídeos, engajou muito mais os alunos que o curso no formato SCORM, com telas e narração de slides.

7 FRAMEWORK CONCEITUAL DE BIG DATA EDUCACIONAL

Os problemas enfrentados quando lidamos com dados educacionais (e educação a distância de uma forma geral) podem ser descritos de várias formas. Não há uma receita de bolo, um tutorial ou uma cartilha que, se for seguida rigorosamente, replica os resultados de sucesso. A educação trata de indivíduos e a aprendizagem é um fenômeno inerente de cada pessoa. Cada aluno desenvolve a melhor forma de aprendizado que lhe atenda.

Mesmo assim, analisando a Big Data Educacional, podemos encontrar perfis semelhantes, alunos que pensam parecido, agem da mesma forma. O comportamento humano em ambientes de aprendizagem gera padrões que podem ser analisados. Sendo assim, muitos pontos em comum surgem, e quando aparecem com certa frequência, evidenciam algumas características e merecem atenção.

Segundo SHEHABUDDEEN et al. (2000), “processo”, “procedimento”, “técnica” e “ferramenta” são palavras frequentemente associadas a abordagens preocupadas com a resolução prática de problemas nesse campo (ou seja, aplicações), “sistema”, “framework” e “modelo” são termos comumente usados para representar preocupações com sua conceituação (ou seja, representações). Por isso o *framework* conceitual foi escolhido para compor o escopo teórico dessa tese, pois ajuda a entender a dinâmica do aluno nesse ambiente, auxilia nos processos decisórios e na resolução de problemas.

Por definição, *Framework* conceitual é um conjunto de conceitos usado para resolver um problema de um domínio específico, não se trata de um software executável, mas sim de um modelo de dados para um domínio. Por definição, "*Framework* é um conjunto de classes que colaboram para realizar uma responsabilidade para um domínio de um subsistema da aplicação" (Fayad e Schmidt, 1997)

Os frameworks podem variar em finalidade e estilo. Em termos de finalidade, os frameworks podem tanto descrever como um objetivo específico pode ser atingido (*know-how*) quanto retratar uma situação específica (*know-what*). Em termos de estilo, podem ser monocamadas, quando apresentam uma única camada de análise, ou multicamadas, quando apresentam duas ou mais camadas de análise. Nas ciências exatas os frameworks costumam ter conotação prática, “normalmente materializados em artefatos aplicáveis a casos concretos, como é o caso de um framework para o desenvolvimento de softwares executáveis”, enquanto nas ciências sociais essa terminologia costuma ser utilizada de forma mais abstrata, buscando,

em geral, representar e compreender um conjunto de conceitos que versam sobre um assunto específico (DELGADO, 2016).

O Framework Conceitual deve evidenciar os principais conceitos e construtos adotados na pesquisa e as relações entre eles. Essa construção deve se dar em formato de texto, mas pode ser auxiliada por um desenho ou por Quadro esquemático. É importante considerar que o Framework Conceitual indica ao leitor em que conceitos a pesquisa se baseia, de que modo ela deve ser lida, com quais pesquisas anteriores ela dialoga e quais aspectos do fenômeno estão sendo analisados. Eles correspondem à visão atualizada de um pesquisador sobre uma determinada questão de pesquisa e evoluem à medida que o pesquisador explora essa questão. Seu resultado pode ser gráfico ou narrativo; entretanto, são frequentemente representados por meio de gráficos ou diagramas acompanhados de elementos textuais (DELGADO, 2016).

A forma de um framework varia de acordo com o seu propósito. O framework conceitual apresentado nesta tese tem o propósito de **auxiliar professores e pesquisadores que querem melhorar a experiência educacional de seus MOOCs através da gestão de dados educacionais armazenados no LMS Moodle**. Todos os elementos contidos no framework são frutos de pesquisa teórica e empírica, formulados através da experiência de dezenas de pesquisadores da área que compartilharam suas experiências, seus êxitos e fracassos, sempre em busca das melhores práticas.

Um framework conceitual efetivo precisa ser de fácil leitura e entendimento. Diferentemente de um fluxograma, o *framework* conceitual não deve ser usado como uma fonte de soluções ou manual de instruções. Ele deve apresentar todos os aspectos conceituais do tema, a fim de orientar as pessoas acerca do assunto pesquisado. Essa visão do todo é o que fornece suporte e insights para os usuários do *framework*.

Vale ressaltar que esse framework não visa ensinar a utilização de técnicas de Data Science, nem tampouco ensinar a mexer na base de dados do Moodle. Essas são tarefas extremamente técnicas e que precisam de profissionais capacitados para desempenhá-las. Se esse framework conceitual conseguir elucidar as dúvidas de seus usuários, através da demonstração de alguns caminhos que podem ser percorridos para facilitar a concretização de seus objetivos, já pode ser considerado um *framework* de sucesso.

7.1 PROPOSIÇÃO DO FRAMEWORK

Diferentemente dos *frameworks* de LA e Big Data, apresentados nos capítulos anteriores, o framework conceitual desta tese não apresenta um formato cíclico. Entendemos que a análise de dados e a busca por melhorias nos cursos deve ser constante, entretanto o caminho trilhado não será sempre o mesmo. Para alcançar objetivos diferentes é preciso pensar e agir diferente. Portanto, esse *framework* será composto por núcleos, temas e subtemas, e cada nível desse respeita uma hierarquia de importância.

Além dessa divisão, há também indicações de setas e ícones, que conectam temas com muita correlação. Por exemplo: se seu objetivo for analisar dados passados, os métodos estatísticos são os mais indicados para isso. Mas se você quiser prever um comportamento, o indicado é utilizar *Machine learning*, para simular dados futuros, com base no que já aconteceu. Logicamente cada pessoa é livre para utilizar o *framework* da forma que quiser, dependendo do seu objetivo e nível de entendimento do assunto.

Esse *framework* foi elaborado para auxiliar as pessoas que, de alguma forma, têm um envolvimento com cursos a distância e se interessam pelo potencial que os dados educacionais podem proporcionar. Podem ser alunos, gestores, professores, pedagogos, designers instrucionais, tutores, enfim, uma série de profissionais, cada um com objetivos e metas diferentes.

7.2 OBJETIVOS / MÉTRICAS

Dessa forma, o *framework* começa com a definição dos objetivos ou métricas de pesquisa. Elencamos aqui as que mais apareceram na revisão sistemática realizada, e que são as mais comuns quando tratamos de pesquisas em dados educacionais. Esses objetivos são divididos em 3 categorias macro, propostas por Robbins (2017): Eficiência (centrada nas atividades de aprendizagem), Efetividade (foco na avaliação e aquisição de conhecimento) e Resultado (união das duas categorias anteriores). A Figura 41 mostra os principais objetivos da pesquisa em BDE:

Figura 42 – Etapa 1 do framework - Objetivos / Métricas



*Eficiência: é relacionada com atividades centradas na aprendizagem.

**Efetividade: é relacionado com atividades centradas na avaliação, qualidade dos resultados, aquisição de conhecimento.

***Resultado: é relacionado ao resultado final. Uma validação das duas métricas anteriores.

Fonte: elaborado pelo autor

Quando tratamos da métrica "eficiência", o monitoramento do comportamento dos alunos durante o curso é um dos objetivos mais comuns na pesquisa de dados educacionais. O Moodle proporciona uma vasta gama de dados de logs, os quais é possível identificar o caminho que o aluno trilhou através dos recursos acessados e atividades realizadas. Pesquisadores utilizam esses dados para identificar fatores que ajudem na diminuição das taxas de evasão, no acompanhamento do desempenho e do engajamento dos alunos. Outra possibilidade é monitorar o acesso aos materiais didáticos (recursos) e verificar se eles estão mais alinhados com a proposta pedagógica do curso. O Moodle disponibiliza Tabelas exclusivas para essa finalidade. Outra forma de garantir a qualidade da aprendizagem é monitorar o *feedback* dos alunos, tentando sempre corrigir possíveis problemas.

Outros pesquisadores focam sua pesquisa na métrica "efetividade". Há bastante interesse em análises sobre o design instrucional dos cursos, assim como sua proposta pedagógica. Descobrir quais recursos educacionais foram utilizados, se eles tiveram êxito ou não, se proporcionaram engajamento dos alunos. Também há muita pesquisa sobre a proposta pedagógica desses cursos, se são behavioristas, cognitivistas, conectivistas ou construtivistas. Se utilizam aulas assíncronas ou síncronas, se utilizam fórum e chats, entre outros. Os pesquisadores que optam por esses objetivos de pesquisa, muito provavelmente são desenvolvedores de cursos, nas mais variadas áreas, e procuram nas experiências mais exitosas

uma inspiração para criar um modelo de curso de sucesso. Eles não têm tanto interesse no desempenho individual de um aluno, e sim na experiência educacional como um todo.

Uma vantagem do Moodle é que ele oferece acesso aos dados em tempo real. Isso pode proporcionar, por exemplo, o envio de alertas e notificações para alunos, durante a vigência do curso. Essa análise em tempo real pode gerar insights e propor melhorias. Muitas vezes pode ser identificado um problema estrutural no curso, algum material com problema, um link quebrado, ou até mesmo um grupo de alunos que não acessou o conteúdo ainda. Por meio do uso de algoritmos de *machine learning*, por exemplo, pode ser criado um modelo que identifique e notifique esses alunos.

Os dados educacionais também podem ser analisados com o objetivo de realizar uma reflexão sobre o curso, identificando prós e contras, propiciando melhorias em uma eventual nova oferta desse curso. Basicamente como se cada edição gerasse insights que vão sendo "lapidados" com objetivo de torná-lo cada vez melhor. Com o crescimento exponencial de cursos online e da educação a distância de uma forma geral (efeito positivo da pandemia de Covid-19), garantir a qualidade de um curso online é fundamental. E muitos pesquisadores chegam até comparar estratégias de ensino de cursos, em busca das mais assertivas.

A terceira métrica (Resultado) é uma validação das duas métricas anteriores. Pesquisas sob essa ótica englobam a avaliação de aprendizagem e experiência. E nesse campo uma das áreas que mais vem ganhando destaque é a personalização de conteúdo. Por mais que os MOOCs sejam cursos massivos e muitas vezes englobam um público alvo heterogêneo, oferecer conteúdos personalizados de acordo com a necessidade dos alunos vem se mostrando uma iniciativa acertada. Por meio dos dados educacionais e da ciência de dados, os pesquisadores têm conseguido identificar grupos de alunos que preferem assistir vídeos, ou que preferem materiais para leitura em PDF. Há também o grupo dos alunos que preferem ouvir *podcasts*. Conseguir essa flexibilidade de formatos e conteúdos que atendam as necessidades e preferências dos alunos pode contribuir para o sucesso de um curso online, aumentando o engajamento e o interesse de forma geral.

7.3 BIG DATA EDUCACIONAL

A segunda categoria do *Framework* se refere aos tipos de dados educacionais que permeiam a pesquisa de BDE, dos quais muitos são fornecidos pelo Moodle. A Figura 42 mostra esses dados.

Figura 43 - Etapa 2 do framework: Big Data Educacional



Fonte: Autor (2021)

Eles podem ser divididos em duas categorias: estáticos e dinâmicos. Os dados estáticos são aqueles que identificam o aluno e não costumam mudar. São importantes para análises qualitativas e geralmente são usados na definição de bolsas escolares e cotas. Nas pesquisas de BDE em MOOCs esses dados não costumam ter tanta relevância.

Já os dados dinâmicos, esses aparecem em maior quantidade, uma vez que o Moodle registra cada clique do aluno dentro de um ambiente virtual de aprendizagem. Esse registro completo, multiplicado pelo número de alunos, é o que forma a BDE de um MOOC. Para facilitar a categorização desses dados, adotou-se a divisão proposta por Bovo et al. (2013).

No coeficiente de presença, encontram-se os dados referentes ao nível de participação dos alunos, engajamento, motivação, frequência de conexão e último *login*. Esses dados, se analisados da maneira correta, oferecem um panorama do nível de interesse e interação do aluno ao longo do curso. Esses dados costumam ser frequentemente pesquisados na busca pelo entendimento do comportamento dos alunos no ambiente virtual.

O coeficiente estudo abrange os dados referentes ao número de lições feitas, latência e número de download de materiais. Esses dados buscam demonstrar o caminho percorrido pelo aluno, se ele acessou todo o conteúdo didático elaborado pelos professores, se baixou o material complementar, ou simplesmente se o aluno se matriculou e não realizou nenhuma ação. Nos MOOCs, as taxas de evasão costumam ser muito altas e existem inúmeras pesquisas para tentar entender por que isso acontece, assim como desenvolver estratégias para aumentar o engajamento do aluno.

Já o coeficiente atividades engloba os dados referentes aos recursos visualizados, páginas visitadas e visualizações *wiki*. Ele difere um pouco do coeficiente estudo pois não mede as atividades avaliativas do curso, e sim a participação do aluno em todas as atividades propostas pelo design instrucional do curso. É muito comum encontrar perfis de alunos que não acessam os conteúdos e atividades e já vão direto para as atividades avaliativas, em busca de um diploma rápido, sendo que esse não é um tipo de comportamento exemplar, pois não há aquisição de conhecimento. Isso é muito comum em cursos de capacitação que possibilitam a progressão de carreira.

O Coeficiente resultados foca no desempenho do aluno. Nessa categoria podemos encontrar dados das avaliações formativas e somativas, o número de tentativas do aluno em determinadas provas, o número de avaliações submetidas, sua nota geral, sua nota média em relação aos outros alunos, entre outros. Esses dados são úteis para avaliar o desempenho geral do curso, mas não devem ser utilizados de forma unilateral, uma vez que a aquisição de conhecimento não é um fator simplesmente quantitativo, e uma nota alta nem sempre reflete que o aluno realmente aprendeu.

Por fim, o coeficiente Social, que muitas vezes é negligenciado nos cursos. Esse coeficiente compreende dados que medem o número de tópicos lidos, o número de respostas e tópicos criados nos fóruns do curso. Uma característica dos MOOCs é o número muito alto de alunos matriculados, o que gera uma grande quantidade de posts e mensagens nos fóruns e chats. Isso demanda uma constante tutoria para avaliação desse conteúdo e o fornecimento do

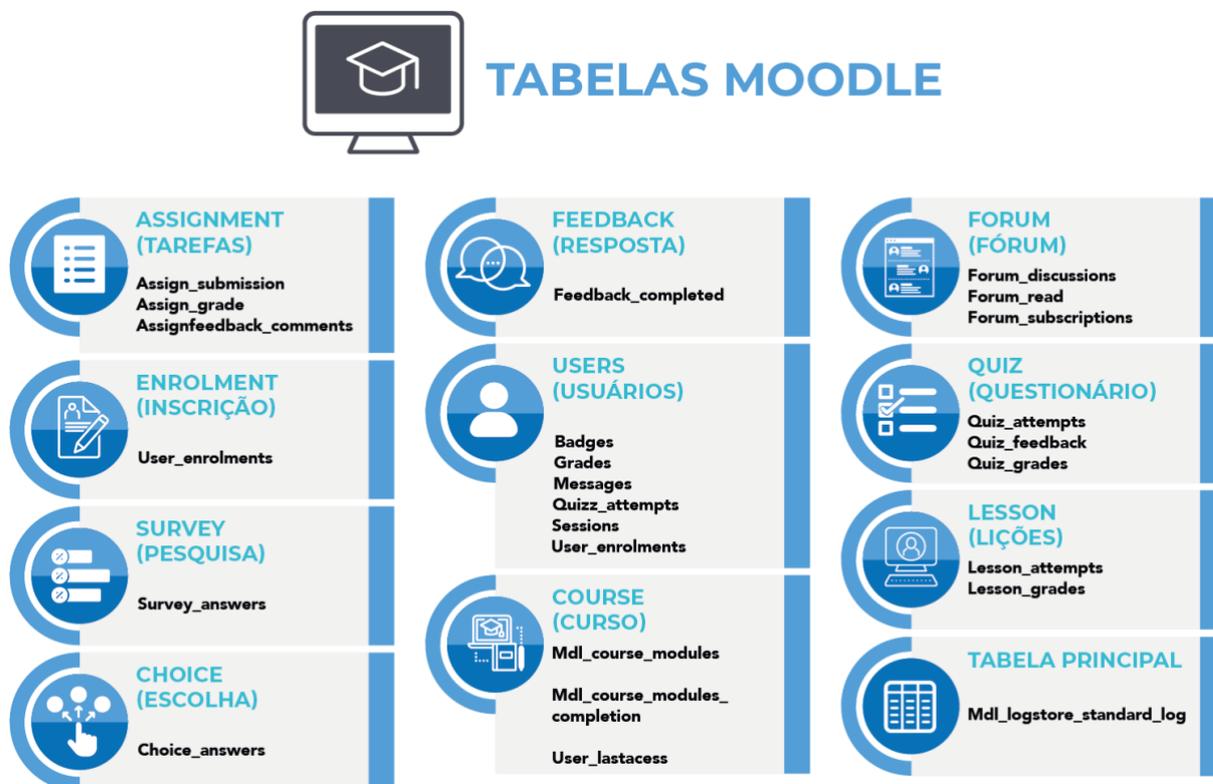
devido *feedback*. Pesquisadores de BDE têm buscado formas de desenvolver técnicas de feedback inteligente automatizado, localizando as perguntas mais frequentes e termos mais comentados. É extremamente importante manter a discussão dos tópicos em alto nível nos fóruns, uma vez que a troca de conhecimentos entre os alunos, mediados pelo tutor, é o principal fator de construção de conhecimento, inclusive esse é o cerne da teoria conectivista.

7.4 TABELAS MOODLE

Uma das maiores contribuições dessa tese está relacionada justamente com a busca de um maior entendimento acerca dos dados educacionais contidos na base de dados do Moodle. Apesar de saber que estão ali, encontrar os dados necessários nem sempre é uma tarefa fácil. O Moodle apresenta uma infraestrutura complexa de Tabelas e nem sempre seus relatórios padrões suprem a demanda dos pesquisadores. São 419 Tabelas e 3843 colunas, que compilam uma enorme quantidade de dados.

Nesse *framework* foram selecionadas somente as Tabelas mais relevantes para a pesquisa em BDE e que abrangem os dados necessários para a obtenção dos objetivos propostos. Outras Tabelas com dados sistêmicos e de cunho técnico, como backup do sistema, e variáveis de configurações foram excluídas do *framework*. A Figura 44 mostra as Tabelas do moodle incluídas no estudo.

Figura 44 - Etapa 3 do framework: Tabelas Moodle



Fonte: Autor (2021)

A Tabela *Assignment* registra os dados relacionados às tarefas existentes no curso. A atividade do tipo Tarefa permite a atribuição de um professor para comunicar tarefas, recolher o trabalho e fornecer notas e comentários. Os estudantes podem apresentar qualquer conteúdo digital (arquivos), como documentos de texto, planilhas, imagens ou áudio e vídeos. Dessa forma, a Tabela *Assign_submission* registra o envio das tarefas, *Assign-grades* registra a nota da atividade submetida e a Tabela *Assignfeedback_comments* registra os feedbacks em texto das tarefas submetidas pelos alunos.

A Tabela *user_enrollment* registra os dados referentes à inscrição do aluno em cursos. Ideal para saber um pouco mais sobre o histórico escolar dos estudantes.

A Tabela *survey_answers* guarda todas as respostas relacionadas aos questionários aplicados ao longo do curso, assim como a Tabela *choice answers* guarda os dados da atividade *Choice*. Esse é um recurso relativamente novo no Moodle que permite que o professor faça

uma pergunta e configure botões de opção nos quais os alunos podem clicar para fazer uma seleção a partir de uma série de respostas possíveis. Eles podem escolher uma ou mais opções e podem atualizar sua seleção. As escolhas podem ser úteis como uma pesquisa rápida para estimular o pensamento sobre um tópico; para permitir que a classe vote em uma direção para o curso ou para avaliar o progresso.

A Tabela *feedback_completed* registra todos os feedbacks que o aluno forneceu durante o curso. Geralmente ao final das atividades o Moodle disponibiliza uma caixa de texto para o aluno fazer algum comentário. Esses dados podem fornecer uma boa análise qualitativa sobre o ponto de vista dos alunos sobre as atividades do curso, e até mesmo pode servir como auxílio para localizar inconsistências em atividades.

No grupo de Tabelas *User*, a Tabela *Badges* registra as "medalhas" que os alunos conquistaram ao longo do curso (no caso desse recurso de gamificação estiver ativado), a Tabela *grades* registra as notas desse aluno, assim como a Tabela *messages* registra todas as mensagens enviadas pelo aluno, na comunicação entre alunos, professores e tutores. Já a Tabela *quiz_attempts* registra o número de tentativas do aluno até conseguir obter a nota mínima em determinada atividade. Essa é uma boa métrica para verificar a persistência e o engajamento do aluno. Já a Tabela *Sessions* registra o número de vezes que o aluno acessou o ambiente virtual de aprendizagem. Se o aluno fica um tempo inativo, o moodle desconecta o aluno automaticamente, obrigando-o a realizar uma nova sessão em seu futuro acesso.

O grupo de Tabelas *Course* registra os dados referentes aos módulos instalados no curso. A Tabela *course_modules* indica quais os módulos estão instalados no curso, assim como a Tabela *course_modules_completion* registra o status dos alunos, se completaram ou não os recursos contidos no módulo. Outra Tabela importante é a *user_lastaccess* que indica o último acesso do aluno no curso (muito útil para saber quais alunos estão mais engajados com o curso, e também serve como um indicador de evasão do aluno, sugerindo uma possível intervenção para recuperar esse aluno).

Já as principais Tabelas sobre o Fórum do Moodle são: *forum_discussions*, que registra as discussões geradas entre alunos e tutores, *Forum_read*, que mostra os posts que cada aluno leu e *forum_subscriptions*, que monitora quais alunos estão inscritos em diversos fóruns. Em MOOCs conectivistas, as análises de dados dos fóruns são imprescindíveis, principalmente para avaliar a aquisição de conhecimento nesse meio.

As Tabelas *Quiz* formam um grupo importante de dados úteis para verificar o desempenho dos alunos. No Moodle, quiz é uma atividade de avaliação que visa atender a as

necessidades de ensino, e podem ser apresentadas na forma de testes de conhecimento simples e de múltipla escolha até tarefas complexas de autoavaliação com *feedback* detalhado. A Tabela *Quiz_attempts* monitora o número de tentativas dos alunos nos *Quizzes* do curso. *Quiz_feedbacks* registra os feedbacks dados para cada quiz submetido pelos alunos, e a Tabela *Quiz-grades* que grava todos as notas dos alunos nessas atividades.

Ainda falando de avaliações, o Moodle também registra os dados das atividades de Lição. Esse modelo de atividade permite que os professores criem exercícios de 'ramificação' onde os alunos são apresentados a conteúdos e, em seguida, dependendo de suas respostas, são direcionados para páginas específicas. O conteúdo pode ser texto ou multimídia. As principais Tabelas desse grupo são *Lesson_attempts* (que registra o número de tentativas dos alunos) e a Tabela *lesson_grade* (que mostra as notas dos alunos nessa atividade).

Por fim, uma das Tabelas mais importantes do Moodle: *Logstore_standard_log*. Essa Tabela é responsável por armazenar todas as informações dos acessos e interações realizadas, ou seja, os *logs* de utilização, dentro do AVA. Ideal para identificar o caminho trilhado pelo aluno durante o acesso ao curso.

7.5 TIPO DE ANÁLISE

Essa categoria é de suma importância para o pesquisador de dados educacionais, pois ela une os objetivos de pesquisa com o método de análise mais adequado. E nem sempre essa escolha fica clara para o pesquisador. A Figura 45 mostra essa categoria que compreende 4 tipos de análise: Descritiva, Diagnóstica, Preditiva e Prescritiva. Esses tipos de análise já foram explicados nos capítulos anteriores, mas retomaremos alguns conceitos para facilitar o entendimento.

Figura 45 - Etapa 4 do Framework: Tipo de Análise



Fonte: Autor (2021)

Seguindo o fluxo proposto neste framework conceitual, o pesquisador que chegou neste ponto já sabe o objetivo de sua pesquisa, já sabe os dados que ele vai precisar e já tem uma noção de quais Tabelas do Moodle ele precisa acessar. Agora é preciso definir qual caminho irá seguir: ele pode analisar os dados passados (de cursos que já ocorreram) para elaborar relatórios e tentar entender o que aconteceu nesses cursos (Análise descritiva). Ele também pode ter identificado algum fator que não funcionou tão bem e precisa melhorar para a próxima oferta do curso. Dessa forma ele vai buscar nos dados o que causou esse problema (análise diagnóstica), e com isso tentará fazer as correções necessárias.

Mas também há aqueles profissionais que tentam antecipar os acontecimentos, e para isso analisam os dados para propor melhorias e realizar testes (análise preditiva). Esse tipo de análise utiliza dados históricos para tentar identificar a probabilidade de resultados futuros. Esse tipo de análise serve para:

- Otimizar processos;
- Detectar perfis de alunos;
- Melhorar a aprendizagem;

- Reduzir risco de evasão;
- Prevenir comportamentos impróprios;

E, por fim, há pesquisadores que buscam nos dados insights para realizar a melhoria de seus cursos. Dessa forma, utilizam métodos computadorizados complexos para extrair informações e antecipar tendências (análise preditiva). Esse tipo de análise traz alguns benefícios como:

- Otimização de processos e estratégias;
- Reduz custos de produção do material didático, sem afetar o desempenho;
- Aumenta a probabilidade de planejamento e aperfeiçoamento dos cursos.

Vale ressaltar que o nível de complexidade dos métodos de análise vai aumentando quando passamos de uma análise descritiva para uma prescritiva. Também é correto afirmar que há uma hierarquia envolvendo esses processos, sendo que um depende do outro para compor uma estratégia analítica eficiente. Quando tratamos de BDE, muitas vezes o uso de computadores para processamento dos dados é a única alternativa, uma vez que é humanamente impossível realizar essas tarefas. Cabe ao pesquisador, portanto, saber analisar essas informações que as máquinas trazem, levando em conta todos os aspectos do curso.

O próximo passo foi relacionar o tipo de análise com o método mais apropriado. E para auxiliar nessa tarefa, o framework sinaliza com números quais os métodos de análise são mais indicados, conforme mostra a Figura 46:

Figura 46 - Indicações numéricas que relacionam os temas do framework



Observe a Figura 46 e veja alguns exemplos: o número 1 corresponde a análise descritiva. Ao relacionar com os métodos de análise, percebe-se que para realizar uma análise descritiva é recomendado o uso de estatística, mineração de dados ou análise qualitativa. Seguindo essa lógica, para realizar uma análise diagnóstica (número 2), é indicado o uso de estatística, mineração de dados, *machine learning* ou análise qualitativa. E assim por diante.

Entendida essa parte, a próxima categoria do *framework* conceitual compreende os métodos de análise.

7.6 MÉTODOS DE ANÁLISE

Essa última categoria apresenta os principais métodos de análise de BDE. O pesquisador que chegou até essa etapa já definiu seus objetivos, já sabe os dados que irá utilizar, sabe localizá-los no banco de dados do Moodle e sabe que tipo de análise irá fazer. Com base nisso, a escolha do método é primordial para trazer os resultados esperados.

Os principais métodos de análise são estatística, mineração de dados, *machine learning*, análise qualitativa, análise de redes sociais e métodos de visualização, conforme mostra a Figura 47.

Figura 47 - Etapa 5 do Framework: Métodos de Análise



Fonte: Autor (2021)

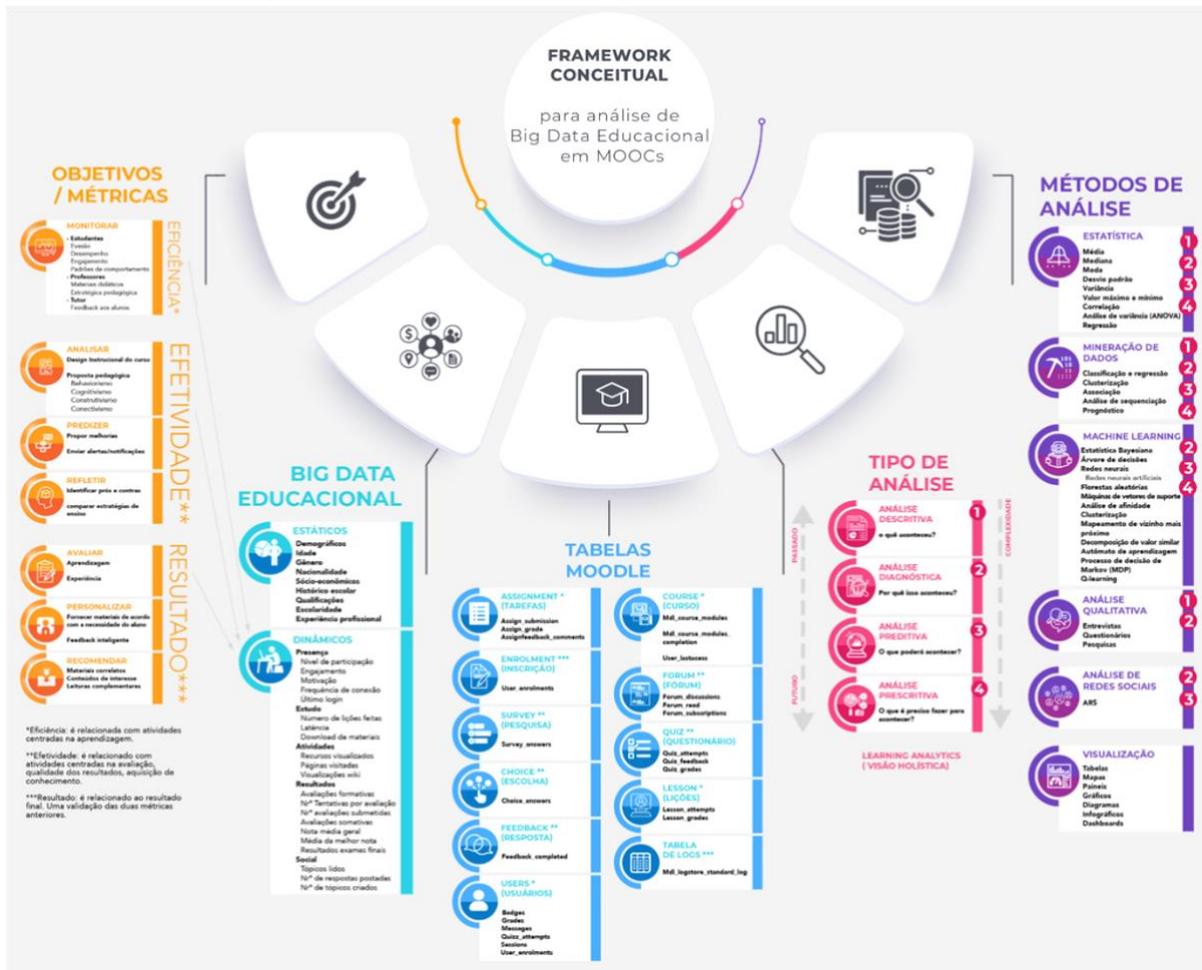
Conforme citado anteriormente, este *framework* não tem objetivo de ensinar a utilização desses métodos, e sim indicar os que são mais apropriados segundo consta na literatura. Atualmente há no mercado diversos *softwares* (gratuitos e pagos) capazes de realizar essas análises mais complexas de dados. Empresas especializadas em ciência de dados podem auxiliar pesquisadores se estes já tiverem bem claros e definidos as metas e objetivos da pesquisa.

Uma possibilidade é a utilização das técnicas de visualização no início da pesquisa. Na maioria dos casos, os pesquisadores utilizam essas ferramentas no final do trabalho, para apresentar os dados em relatórios, artigos e apresentações. Entretanto, a literatura apresenta alguns relatos interessantes do uso das técnicas de visualização para realizar a análise

exploratória dos dados. Nesses casos, o pesquisador cria gráficos e diagramas com base nos dados brutos e busca *insights* para iniciar sua pesquisa através dessas visualizações.

Apresentadas todas as categorias de forma individual, agora é hora de ver o *framework* conceitual completo (Figura 48)

Figura 48 - Framework Conceitual para análise de BDE em MOOCs



24

Fonte: Autor (2021)

Esse *framework* conceitual foi pensado para ser lido da esquerda para direita, entretanto, ele não é um caminho fechado e único. Entendemos que a pesquisa com dados educacionais apresenta muitas nuances. Para quem está começando uma pesquisa do início, certamente o *framework* simboliza um pontapé inicial. Mas também há casos de pesquisadores que têm objetivos claros, mas não sabem manipular os dados. O contrário também acontece, quando o

24 O framework pode ser visto em alta resolução no seguinte link: https://drive.google.com/file/d/1uPWlcmG6uboWqhwW4n9ui163q7p6_NR/view?usp=sharing

pesquisador sabe todos os dados que possui, mas não sabe o que fazer com eles. Independente do estágio da pesquisa e da habilidade dos pesquisadores, o *framework* acima sempre servirá como uma diretriz.

8 CONCLUSÕES E TRABALHOS FUTUROS

O desenvolvimento dos MOOCs ao longo da última década transformou a área da educação a distância no Brasil e no mundo. Nunca na história a busca pelo conhecimento foi tão democrática e facilitada. Inúmeros são os cursos abertos disponíveis na internet, com os mais variados temas e formatos. Essa diversidade é muito positiva pois proporciona conhecimento gratuito e auxilia muitas pessoas a obterem uma profissão, a se capacitarem, ou até mesmo aprenderem um novo *hobby*. O perfil dos alunos que procuram esses cursos é muito variado. Dessa forma, o mercado da educação a distância se tornou muito concorrido e rentável. A facilidade proporcionada pelas ferramentas de criação possibilitam a inserção de cursos na internet com poucos cliques. Isso provocou uma enxurrada de cursos online com as mais variadas qualidades. E como se destacar num cenário de extrema competitividade?

Focar na qualidade do conteúdo é uma das formas. Proporcionar uma boa experiência de aprendizagem também ajuda. Entender o perfil dos alunos, suas necessidades e preferências pode ser um fator determinante do sucesso de um curso. Essas implementações podem ser realizadas com base em uma abordagem centrada em dados. E quando começamos a adentrar o mundo da Big Data Educacional, conseguimos vislumbrar inúmeras possibilidades de análise e melhorias. Parece que tudo fica mais claro. Mas é justamente nesse momento que muitos (como eu) esbarram nas dificuldades inerentes da ciência de dados.

A ciência de dados é um campo de atuação espetacular. Os profissionais da área têm uma visão analítica do mundo e enxergam padrões e possibilidades onde a maioria das pessoas enxergam apenas números desconexos. Ao tentar entender como funciona essa ciência, percebemos a necessidade de criar uma cultura direcionada à dados, ou seja, criar elementos que possam ser analisados, que gerem um rastro. Por mais que pareça besteira, isso faz toda a diferença. Resumindo, se você quer fazer uma boa análise de dados, primeiramente você precisa armazenar uma boa variedade de dados, e de preferência que eles tenham qualidade.

Trazendo isso para o mundo dos MOOCs, ao criar um curso novo, eleja uma plataforma que garanta uma boa curadoria de dados, que seja intuitiva e segura, afinal ali serão armazenados muitos dados relevantes dos alunos. Garanta que todo esse banco tenha qualidade e que você realmente armazene os dados que serão relevantes para análise. Por padrão, cursos criados no Moodle já apresentam toda infraestrutura de dados necessária para a realização de uma boa análise. Não é à toa que o Moodle é o LMS mais utilizado e difundido no mundo. Por detrás das telas há uma enorme comunidade de usuários entusiastas, programadores e

profissionais de diversas áreas que implementam o Moodle através da criação de módulos e plugins que visam facilitar a experiência educacional. Mas a falta de conhecimento, aliada à falta de interesse acaba por deixar na maioria das vezes a BDE ociosa e esquecida, minando qualquer possibilidade de melhoria dos cursos.

As instituições, de forma geral, são lentas na aplicação de projetos de BDE. Mesmo sabendo do enorme potencial e oportunidades que os dados possibilitam, elas não são imediatamente acessíveis aos pesquisadores educacionais, a menos que oportunidades de desenvolvimento profissional sejam fornecidas (Daniel, 2017). É importante capacitar profissionais para atuarem nessa área.

O framework proposto nesta tese foi elaborado com base nas lacunas encontradas na bibliografia, mas também foi uma motivação pessoal, uma tentativa de entender esse universo complexo. Segundo Johann Ari Larusson e Brandon White, no capítulo de introdução do livro *Learning Analytics: from research to practice*, simplesmente disponibilizar a tecnologia não é garantia de sucesso. É preciso explorar efetivamente as comunidades de aprendizagem a fim de entender as complexas interações que estão ocorrendo. Quais são os métodos de análise disponíveis? Qual é a unidade de análise? Como as organizações podem organizar seus dados de forma eficaz? Como os dados são coletados? Como podemos impactar positivamente práticas pedagógicas? Como implementar estratégias eficientes que possam ser replicadas com sucesso posteriormente? Penso ter avançado no caminho de conseguir respostas a essas perguntas através do framework conceitual aqui proposto.

Ao realizar a revisão sistemática englobando esses temas, logo percebe-se que o Brasil ainda está bem atrás nas pesquisas. São poucos os trabalhos que abordam essa temática, e por muitas vezes os estudos são bem superficiais. Interessantes linhas de pesquisa têm surgido na Europa (mais especificamente na Espanha), mas definitivamente os Estados Unidos e a China são os países com o maior número de produção acadêmica nessa área. A China, por exemplo, tem publicado artigos muito interessantes com o foco nas metodologias utilizadas para encontrar os melhores resultados nas análises de dados educacionais. São relatos do uso de técnicas computadorizadas muito complexas que vem apresentando ótimos resultados na melhoria e detecção de problemas ou comportamentos inadequados nos cursos.

Porém vale ressaltar que o tecnicismo não é garantia de melhores resultados. Analisar a BDE e toda sua complexidade não se limita ao uso de algoritmos e cálculos. O fator humano tem relevante papel no entendimento da dinâmica desses cursos. Há quem diga que o uso

excessivo da BDE é apenas uma forma de corrigir problemas (um conserto técnico) ao invés de uma forma de capacitar pesquisadores para realizar melhores pesquisas educacionais (Eynon,2013). É um ponto de vista interessante, porém não compactuo com esse pensamento. Acredito que o *Learning Analytics*, e seu caráter holístico, faz com que os pesquisadores melhorem sua pesquisa educacional, ao tentar encontrar os métodos mais assertivos para alcançar seus objetivos.

É importante ressaltar que para realizar uma pesquisa da BDE de MOOCs deve-se levar em conta questões como infraestrutura, ferramentas e equipe capacitada para realizar a coleta, a limpeza, a análise e a visualização dessa grande quantidade de dados. É preciso compreender os dados como um todo, analisar as motivações dos alunos, tentar entender suas dificuldades, usar a empatia. Por isso é extremamente importante haver uma equipe interdisciplinar para realizar essa análise da BDE. O professor (especialista no assunto do curso) em conjunto com um pedagogo (especialista em educação), um design instrucional (fazendo a mediação entre educação, tecnologia, comunicação, produção e gestão) e um cientista de dados (especialista em análise). Essa equipe trabalhando em conjunto certamente traria ótimos resultados.

Mas como a realidade do mercado de trabalho nem sempre é a desejável, o *framework* conceitual desta tese já serve como uma orientação para conseguir entender como funciona esse processo de análise e implementação de melhorias. Cada curso tem suas peculiaridades e, apesar de servir como guia, o framework não é uma garantia de sucesso. Deve-se ter em mente a importância de achar as ferramentas e métodos que se adequem aos seus dados e seus objetivos de pesquisa. Nesse ramo não existe uma receita de bolo pronta, que dê sempre certo e possa ser replicada facilmente.

Saber lidar com imprevistos e frustrações também é um fator importante na pesquisa em BDE. Nem sempre todos os dados estarão disponíveis. Nem sempre eles estarão no formato correto. O *backup* pode falhar (problemas técnicos acontecem). O importante nessas situações é tentar fazer o melhor possível dentro das possibilidades de cada curso. E vale lembrar que todo esse processo de análise de dados educacionais é contínuo, nunca para! A busca pelas melhorias educacionais devem ser constantes, uma vez que esses cursos online estão intimamente ligados com a tecnologia, e essa, avança a passos largos.

Como indicação de trabalhos futuros, pretende-se colocar em prática as melhorias propostas para o curso de auriculoterapia (analisadas com base no *framework* deste trabalho), e aprofundar mais as análises preditivas e prescritivas do curso. Isso demanda uma maior

maturidade no uso do *framework* e também uma certa credibilidade frente aos gestores do curso, que consigam enxergar as possibilidades que os dados educacionais proporcionam.

APÊNDICE 1

Quadro com os artigos selecionados na revisão bibliográfica sistemática.

Título do artigo	Autores	Abordagem
SENS: Network analytics to combine social and cognitive perspectives of collaborative learning	Gašević, D., Joksimović, S., Eagan, B.R., Shaffer, D.W. (2019)	Os autores sugerem uma nova abordagem para análise da aprendizagem colaborativa em MOOCs, por meio dos dados de análise epistêmica de redes sociais.
A Learning Analytics Tool for Predictive Modeling of Dropout and Certificate Acquisition on MOOCs for Professional Learning	Cobos, R., Olmos, L. (2019)	O artigo aborda o uso de uma ferramenta de modelagem preditiva por meio de algoritmos de <i>Machine Learning(ML)</i> , para prever duas situações em MOOCs: descobrir alunos com tendência a abandono do curso e detectar alunos que desejam concluir a certificação (obter certificado).
Active learners' characterization in MOOC forums and their generated knowledge	Cohen, A., Shimony, U., Nachmias, R., Soffer, T. (2018)	Os autores utilizam métodos de mineração de dados educacionais e <i>Learning Analytics</i> para analisar as relações interpessoais entre alunos em fóruns do MOOC Coursera.
Dropout detection in MOOCs: An exploratory analysis	Isidro, C., Carro, R.M., Ortigosa, A. (2018)	O artigo aborda o uso de diferentes técnicas de <i>machine learning</i> testadas em dados reais de MOOCs da EDX e UAM para prever a evasão de alunos.
Innovative Social Approach in the Nuclear Sector: A MOOC on Nuclear Safety Culture Within H2020 ANNETTE Project	Alonso-Ramos, M., Sanchez-Elvira, A., Gozalo, J.S., Murias, T.F., Castro, M. (2018)	Os autores relatam estudos de casos e o uso intensivo de learning analytics para auxiliar na elaboração do design instrucional de um MOOC sobre segurança nuclear.
Towards a Better Understanding of Mobile Learning in MOOCs	Rohloff, T., Bothe, M., Renz, J., Meinel, C. (2018)	O artigo apresenta um estudo quantitativo sobre a presença de dispositivos móveis em MOOCs, ao analisar os dados de cursos baseados em plataformas ubíquas através de técnicas de <i>Learning Analytics</i> .
Adaptive Intelligence for Massive Online Open Course	Thankachan, K. (2018)	O autor discute o design de um sistema de tutoria inteligente, que ofereça suporte pedagógico em MOOCs. Mineração de dados educacionais e técnicas de <i>Learning Analytics</i> foram utilizados para entender o comportamento dos alunos.
Predicting student performance in a blended MOOC	Conijn, R., Van den Beemt, A., Cuijpers, P. (2018)	O artigo busca entender quais dados dos MOOCs podem ser usados para implementar a performance de aprendizagem. Para isso foram utilizadas correlações, regressões múltiplas e mineração de processos.

Re-examining the community of inquiry framework from the perspective of learning analytics	Xu, S., Luo, H., Tan, Y. (2018)	O artigo faz uma revisitação de um framework de aprendizagem online. A partir de de um estudo qualitativo e quantitativo, utilizando técnicas de LA, os autores identificaram 4 tipos de comportamentos de aprendizagem que podem servir como preditores para a aprendizagem em MOOCs.
Learning Profiles, Behaviors and Outcomes: Investigating International Students' Learning Experience in an English MOOC	Tan, Y., Zhang, X., Luo, H., Sun, Y., Xu, S. (2018)	Utilizando técnicas de LA (<i>k-means cluster analysis e stepwise linear regressions</i>) os autores identificaram 3 perfis de aprendizagens em um MOOCs de língua inglesa, com potencial de melhoria na performance dos alunos.
Sentiment analysis in MOOCs: A case study	Moreno-Marcos, P.M., Alario-Hoyos, C., Munoz-Merino, P.J., Estevez-Ayres, I., Kloos, C.D. (2018)	O artigo aborda a utilização de técnicas de um algoritmo de <i>machine learning</i> que analisa o sentimento dos alunos, através das mensagens trocadas nos fóruns do MOOC.
Predictive learning analytics for video-watching behavior in MOOCs	Shridharan, M., Willingham, A., Spencer, J., Yang, T.-Y., Brinton, C. (2018)	Os autores desenvolveram algoritmos para prever o comportamento do aluno ao assistir as videoaulas de um MOOC, através dos dados gerados pelos cliques dos estudantes.
Discovery and temporal analysis of latent study patterns in MOOC interaction sequences	Boroujeni, M.S., Dillenbourg, P. (2018)	Os autores utilizam técnicas de clusterização para descobrir padrões de aprendizagem em MOOCs.
Understanding online interaction in language MOOCs through learning analytics	Martín-Monje, E., Castrillo, M.D., Mañana-Rodríguez, J. (2018)	O artigo propõe a utilização de técnicas de mineração de dados para identificar quais objetos de aprendizagem são os mais utilizados, quais interações de alunos se relacionam com o sucesso de completar o curso e quais perfis de alunos são os mais prováveis a ter sucesso em um MOOC de língua estrangeira.
Using learning analytics to evaluate a video-based lecture series	Lau, K.H.V., Farooque, P., Leydon, G., (...), Sadler, R.M., Moeller, J.J. (2018)	Os autores utilizaram análise regressiva para avaliar os comportamento dos alunos ao assistirem as videoaulas de um MOOC na área da saúde.
Implementation of learning analytics in MOOC by using artificial unintelligence	Yulianto, B., Prabowo, H., Kosala, R., Hapsara, M. (2018)	O artigo aborda a utilização de métodos de LA e inteligência artificial (IA) para identificar e enviar mensagens para potenciais alunos que tem o perfil de abandonar o curso.
Learning analytics in massive open online courses as a tool for predicting learner performance	Bystrova, T., Larionova, V., Sinitsyn, E., Tolmachev, A. (2018)	Os autores utilizam um algoritmo de LA para prever a performance dos alunos em MOOCs.
Automated quality assurance in MOOCs through learning analytics	Renz, J., Rohloff, T., Meinel, C. (2018)	Este artigo descreve como os dados do LA podem ser usados para criar um sistema com garantia de qualidade automatizada para MOOCs.

Mapping the factors influencing success of Massive Open Online Courses (MOOC) in higher education	Albelbisi, N., Yusop, F.D., Salleh, U.K.M. (2018)	Os autores propõem 3 dimensões (Presságio, Processo e Produto) no qual divide os dados dos MOOCs em 3 categorias com objetivo de entender os fatores que influenciam no sucesso de um MOOC.
Evaluating the Robustness of Learning Analytics Results Against Fake Learners	Alexandron, G., Ruipérez-Valiente, J.A., Lee, S., Pritchard, D.E. (2018)	O objetivo do artigo é avaliar a robustez dos resultados encontrados por LA quando os dados indicam um número considerável de "alunos falsos". Aplicou-se técnicas de <i>machine learning</i> para comparar situações.
Can learning analytics find success in didactical measurements? Results from a MOOC case study	Khalil, M., Ebner, M. (2018)	O artigo apresenta um estudo de caso de da aplicação de técnicas de LA em um MOOC na área da saúde, priorizando a análise de evasão e engajamento dos alunos.
Modelling MOOC learners' social behaviours	Sunar, A.S., Abbasi, R.A., Davis, H.C., White, S., Aljohani, N.R. (2018)	O artigo utiliza técnicas de LA na busca por padrões para medir o engajamento social dos alunos de MOOCs.
Analysing structured learning behaviour in Massive Open Online Courses (MOOCs): An approach based on process mining and clustering	Van den Beemt, A., Buys, J., van der Aalst, W. (2018)	O objetivo do artigo é explorar a relação entre o comportamento do aluno e o seu desempenho ao longo do curso. Por meio da técnica de clusterização e mineração de processos, emergiram 4 perfis de alunos de MOOCs.
A Learning Analytics System for Cognition Analysis in Online Learning Community	Wu, Y., Wu, W. (2018)	O artigo apresenta um modelo analítico que quantifica o ganho de aprendizagem usando uma taxonomia comportamental e considerando fluxos de conhecimento. Através de técnicas de LA os autores buscam entender como se dá a aquisição de conhecimento em fóruns de discussão de MOOCs
Developing SMART educational cloud environment on the basis of adaptive massive open online courses	Parfenov, D., Zaporozhko, V. (2018)	O artigo sumariza as abordagens e técnicas utilizadas para promover a personalização de MOOCs. Aborda o processamento de dados de big data, mineração de dados educacionais e LA.
A learning analytics methodology for understanding social interactions in MOOCs	Moreno-Marcos, P.M., Alario-Hoyos, C., Munoz-Merino, P.J., Estevez-Ayres, I., Delgado Kloos, C. (2018)	O artigo propõe uma metodologia 3s(<i>Social, sentiments and Skills</i>) para analisar as interações sociais em fóruns e ver como se dá o processo de aprendizagem nesses ambientes.
Towards personalized content in massive open online courses	El Mawas, N., Gilliot, J.-M., Garlatti, S., Euler, R., Pascual, S. (2018)	O estudo aborda a necessidade de personalização de conteúdos para aumentar as taxas de conclusão dos MOOCs, e para isso utilizam técnicas de LA.

Learning analytics suggest a positive experience: a descriptive analysis of a care and compassion MOOC	Mclaren, J., Donaldson, J., Smith, S. (2018)	O artigo apresenta uma pesquisa com 957 alunos de um MOOC na área da saúde, e através de métodos mistos e uso de LA, os autores buscam entender o comportamento dos alunos durante o curso.
Deciphering the attributes of student retention in massive open online courses using data mining techniques	Gupta, S., Sabitha, A.S. (2018)	O artigo foca na retenção de alunos em MOOCs, utilizando técnicas de mineração de dados para detectar padrões de comportamentos de alunos.
Communication at scale in a MOOC using predictive engagement analytics	Le, C.V., Pardos, Z.A., Meyer, S.D., Thorp, R. (2018)	Através de técnicas de LA os autores buscam estabelecer uma forma de comunicação com os perfis de alunos de MOOCs para diminuir taxas de evasão.
Dropout model evaluation in MOOCs	Gardner, J., Brooks, C. (2018)	Os autores propõem um modelo para diminuição da evasão, baseado em um estudo onde foram aplicadas técnicas de LA, testes estatísticos e algoritmos específicos.
Viseq: Visual Analytics of Learning Sequence in Massive Open Online Courses	Chen, Q., Yue, X., Plantaz, X., Pong, T., Qu, H. (2018)	Os autores elaboraram um sistema de visualização analítica (Viseq) com objetivo de resolver a perda de informações, melhorar a visualização da sequência de aprendizagem de diferentes grupos e entender as razões por trás dos comportamentos de aprendizagem em MOOCs.
Understanding Learner's Drop-Out in MOOCs	Itani, A., Brisson, L., Garlatti, S. (2018)	Estudo sobre a evasão de alunos em MOOCs, que utiliza técnicas de <i>machine learning</i> como algoritmos preditivos, algoritmos explicativos para gerar intervenções automatizadas e personalizadas.
The Effect of Personality and Course Attributes on Academic Performance in MOOCs	Rahmani Hanzaki, M., Demmans Epp, C.(2018)	O artigo analisa a relação entre a personalidade dos alunos e sua performance acadêmica em MOOCs. Por meio de algoritmos de <i>machine learning</i> , os autores buscam relacionar a colaboração e fatores sociais e seu impacto na aprendizagem.
Scaling to Massiveness with ANALYSE: A Learning Analytics Tool for Open edx	Ruiperez-Valiente, J.A., Munoz-Merino, P.J., Gascon-Pinedo, J.A., Kloos, C.D. (2017)	Os autores sugerem que as ferramentas de LA para MOOCs precisam ser melhoradas e ajustadas. Para isso, propuseram um sistema chamado ANALYSE, e aplicaram em dados de grandes provedores de MOOCs, dando diretrizes de como trabalhar com grandes quantidades de dados, facilitando a visualização.
Videomark: A video-based learning analytic technique for MOOCs	Huang, N.-F., Hsu, H.-H., Chen, S.-C., Ou, P.-W., Tzeng, J.-W. (2017)	O artigo investiga a interação dos alunos com os vídeos, propondo uma nova metodologia de LA para analisar essa relação.
Implementation of learning analytics framework for MOOCs using state-of-The-Art in-memory computing	Laveti, R.N., Kuppili, S., Ch, J., Pal, S.N., Babu, N.S.C. (2017)	Os autores desenvolveram um framework de LA, utilizando dados de 39 cursos do MOOC EDX. Foram utilizados métodos de análise estatística e <i>machine learning</i> .

Edx-MAS: Model analyzer system	Cobos, R., Palla, V.M. (2017)	Os autores apresentam uma ferramenta para predição chamada EDX-MAS. Essa ferramenta engloba técnicas de diferentes algoritmos de <i>machine learning</i> .
Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks	Yang, T.-Y., Brinton, C.G., Joe-Wong, C., Chiang, M. (2017)	O artigo apresenta um novo modelo de predição para MOOCs, analisando através de técnicas de <i>machine learning</i> e redes neurais e propondo um algoritmo assertivo para melhorar o desempenho e a aprendizagem.
Undirected bipartite networks as an alternative methodology to probabilistic exploration: Online interaction and academic attainment in MOOC	Martínez-Cerdá, J.-F., Torrent-Sellens, J. (2017)	O artigo explora como técnicas de análise gráfica podem implementar o processo de LA em MOOCs, realizando experimentos com 2353 alunos de 6 países.
Mining individual learning topics in course reviews based on author topic model	Liu, S., Ni, C., Liu, Z., Peng, X., Cheng, H.N.H. (2017)	O estudo apresenta uma abordagem teórica sobre LA em MOOCs, com foco na aprendizagem não supervisionada e aprendizagem adaptativa.
Machine learning approaches to predict learning outcomes in Massive open online courses	Al-Shabandar, R., Hussain, A., Laws, A., Lunn, J., Radi, N. (2017)	O artigo relata o uso de diversas técnicas de <i>machine learning</i> em MOOCs, a fim de achar algum que apresentasse resultados mais assertivos.
Learning Analytics	Korhonen, A., Multisilta, J. (2017)	O capítulo de livro apresenta uma abordagem teórica do uso correto do LA em MOOCs, alertando que para o sucesso desses cursos, eles devem ser projetados já pensando na análise de dados.
Google bigquery for education: Framework for parsing and analyzing edx MOOC data	Lopez, G., Seaton, D.T., Ang, A., Tingley, D., Chuang, I. (2017)	O artigo detalha o funcionamento de uma ferramenta de LA chamada bigquery, fornecida pelo Google. Essa ferramenta analisa enorme quantidade de dados e oferece <i>dashboards</i> de visualização.
Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories	Khalil, M., Ebner, M. (2017)	O artigo busca, através de técnicas de clusterização, definir categorias de estudantes de um MOOC e seu comprometimento com o curso.
Workshop on integrated learning analytics of MOOC post-course development	Wang, Y., Davis, D., Chen, G., Paquette, L. (2017)	Os autores buscam entender como usar LA para entender o comportamento dos alunos em MOOCs, dando ênfase nos dados coletados da web social dos estudantes.
Cooking with learning analytics recipes	Jaakonmäki, R., Dietze, S., Drachsler, H., Kickmeier-Rust, M., Marenzi, I. (2017)	O artigo aborda que pesquisas sobre LA estão na moda, entretanto não apresentam resultados concretos. Sendo assim, os autores propõe o LA4S, projeto dedicado para relatar soluções práticas, fornecer ferramentas e publicar "receitas" de como aplicar o LA.
Learning Analytics to Identify Students At-risk in MOOCs	Srilekshmi, M., Sindhumol, S., Chatterjee, S., Bijlani, K. (2017)	O estudo busca entender os motivos da evasão dos alunos de MOOCs, fazendo predições para identificar esse alunos e tomar medidas e feedbacks para evitar o abandono do curso.

Learning analytics in MOOCs: EMMA case	Eradze, M., Tammets, K. (2017)	O artigo apresenta um estudo de caso da aplicação de um sistema de LA que oferece feedback instantâneo para alunos e professores e apresenta recomendações para mudanças no design instrucional de MOOCs, baseados nos dados fornecidos.
Statistical approaches to the model comparison task in learning analytics	Gardner, J., Brooks, C. (2017)	O estudo mostra uma comparação de performance de métodos preditivos em MOOCs, analisando quais abordagens são mais efetivas, utilizando técnicas de <i>machine learning</i> .
Mining learning behavioral patterns of students by sequence analysis in cloud classroom	Liu, S., Hu, Z., Peng, X., Cheng, H.N.H., Sun, J. (2017)	Os autores buscam por padrões comportamentais de estudantes de MOOCs, utilizando técnicas de mineração de dados e análise sequenciais.
MOOC learner behaviour: Attrition and retention analysis and prediction based on 11 courses on the TELESCOPE platform	Vitiello, M., Gütl, C., Amado-Salvatierra, H.R., Hernández, R. (2017)	O estudo busca entender as causas da evasão de alunos em MOOCs e diferenciar os alunos que intencionalmente abandonam o curso e os alunos que tem dificuldades para continuar os estudos.
The proof of the pudding: Examining validity and reliability of the evaluation framework for learning analytics	Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., Specht, M. (2017)	O artigo apresenta um framework (EFLA) que provê uma forma eficiente e efetiva de gerar insights para aplicações de LA em instituições. Os autores ainda apresentam alguns <i>widgets</i> , testados em big data de MOOCs que mostraram uma boa confiabilidade.
Community learning analytics support for audio-visual web-based learning contents: The CIDRE framework	Klamma, R., Spaniol, M. (2017)	O artigo apresenta um framework CIDRE para avaliar o conteúdo audiovisual de MOOCs. Segundo os autores, o conteúdo ministrado via videoaulas não são compatíveis com as técnicas de LA. Essa proposta de <i>framework</i> vem para suprir esse <i>Gap</i> .
The ethics of machine-based learning: Advancing without losing humanity	Willis, J.E., Strunk, V.A. (2017)	Os autores argumentam que os avanços tecnológicos estão ofuscando o papel humano do processo educacional. LA está alterando a forma como professores trabalham e o pensamento crítico está subjugado às análises de dados.
Predicting attrition from massive open online courses in futurelearn and edx	Cobos, R., Wilde, A., Zaluska, E. (2017)	Os autores utilizam diferentes técnicas de <i>machine learning</i> para prever atritos em grandes provedores de MOOCs: Future-Learn e EDX.
The university of Southampton MOOC observatory dashboard	León-Urrutia, M., Tang, D. (2017)	O artigo mostra o funcionamento de um <i>dashboard</i> de visualização de dados dos MOOCs da Future-Learn, em tempo real.
Machine learning based on big data extraction of massive educational knowledge	Hadioui, A., El Faddouli, N.-E., Touimi, Y.B., Mohammed, S.B. (2017)	O artigo busca traçar algumas diretrizes do uso de LA e <i>machine learning</i> em Big Data educacional, apresentando um algoritmo para produção de conhecimento.

Analysing event transitions to discover student roles and predict grades in MOOCs	Pérez-Lemonche, Á., Martínez-Muñoz, G., Pulido-Cañabate, E. (2017)	O artigo utiliza técnicas de LA para identificar os caminhos traçados pelos alunos de um MOOC. São utilizadas técnicas de <i>machine learning</i> como o algoritmos k-means, Random Forest e redes neurais.
Early prediction and variable importance of certificate accomplishment in a MOOC	Ruipérez-Valiente, J.A., Cobos, R., Muñoz-Merino, P.J., Andujar, Á., Kloos, C.D. (2017)	Os autores utilizam 4 algoritmos de <i>machine learning</i> para detectar com antecedência alunos com comportamento de abandonar o curso.
Towards a Learning Analytics Support for Intelligent Tutoring Systems on MOOC Platforms	Baneres, D., Caballe, S., Clariso, R. (2016)	O artigo apresenta o design de um sistema de LA que oferece ferramentas de assistência para MOOCs (Sistema de tutoria inteligente).
Analyzing navigation logs in MOOC: A case study	Mercado-Varela, M.A., García-Holgado, A., García-Peñalvo, F.J., Ramírez-Montoya, M.S. (2016)	Os autores analisam os rastros deixados através dos logs dos alunos nos sistemas de gestão de aprendizagem (LMS).
Big data application in education: Dropout prediction in edx MOOCs	Liang, J., Yang, J., Wu, Y., Li, C., Zheng, L.(2016)	O artigo analisa a predição de evasão em MOOCs da EDX, utilizando algoritmos de classificação de <i>machine learning</i> .
The MOOC and learning analytics innovation cycle (MOLAC): A reflective summary of ongoing research and its challenges	Drachsler, H., Kalz, M.(2016)	O artigo apresenta o framework para ciclo de inovação MOLAC, que possui 3 níveis de análises para o uso de LA em MOOCs.
Open-dlas: An open dashboard for learning analytics	Cobos, R., Gil, S., Lareo, Á., Vargas, F.A.(2016)	O artigo apresenta uma proposta de <i>dashboard</i> de visualização para MOOCs, a fim de facilitar a interpretação de dados de alunos e professores.
A framework for topic generation and labeling from MOOC discussions	Atapattu, T., Falkner, K. (2016)	Os autores criaram um framework para geração automática de tópicos para fóruns de discussão em MOOCs, gerados através de técnicas de <i>machine learning</i> e estatística.
MOOCrp: An open-source analytics platform	Pardos, Z.A., Whyte, A., Kao, K. (2016)	O artigo apresenta dois estudos de caso utilizando os dados de MOOCs, e foca na elaboração de modelos analíticos para facilitar a vida de pesquisadores. Também é apresentado o conceito de <i>Open Learning Analytics</i> e uma plataforma para integrar LA de diversos MOOCs.
Advancing MOOC and SPOC development via a learner decision journey analytic framework	Lei, C.-U., Hou, X., Kwok, T.T.O., Yeung, Y.-C.A., Lai, C.(2016)	O artigo elabora um framework para análise de MOOCs, entrevistando alunos pós-curso e aplicando técnicas de LA. Segundo os autores, o framework é cíclico e serve para implementar futuros cursos.

An overview of studies about students' performance analysis and learning analytics in MOOCs	Duru, I., Dogan, G., Diri, B. (2016)	O artigo traça um panorama sobre o uso de técnicas de LA em MOOCs, por meio de comparações de estudos de caso na área.
When learning analytics meets MOOCs - A review on imoox case studies	Khalil, M., Ebner, M. (2016)	O artigo ressalta o crescimento da área de LA e a importância de aproximar os cientistas de dados com a educação. Além disso, os autores aplicam técnicas de LA para entender o caminho que os alunos trilham nos MOOCs.
Visualization improvement in learning analytics using semantic enrichment	Fernández, G., Mariño, O. (2016)	As autoras analisaram os dados de um MOOC, utilizando diversas técnicas de LA e detalhando os procedimentos utilizados.
Student engagement in massive open online courses	Sinclair, J., Kalvala, S. (2016)	O artigo utiliza técnicas de LA para analisar o engajamento de alunos em MOOCs, fazendo recomendações para desenvolvedores de MOOCs.
On using learning analytics to track the activity of interactive MOOC videos	Wachtler, J., Khalil, M., Taraghi, B., Ebner, M.(2016)	O artigo apresenta uma análise específica sobre vídeos interativos em MOOCs. Um estudo quantitativo e qualitativo a respeito dessa prática, em busca de aumentar o engajamento de alunos.
How should we measure online learning activity?	O'Riordan, T., Millard, D.E., Schulz, J. (2016)	O artigo busca achar os melhores métodos para se medir a atividade de aprendizagem. Com uma abordagem qualitativa, foram utilizadas 4 abordagens diferentes para analisar o conteúdo dos fóruns de MOOCs.
Detecting cheaters in MOOCs using item response theory and learning analytics	Alexandron, G., Lee, S., Chen, Z., Pritchard, D.E. (2016)	O artigo busca padrões de alunos que buscam trapacear por meio de regressões logísticas e técnicas de mineração de dados.
Predicting students' success based on forum activities in MOOCs	Klusener, M., Fortenbacher, A.(2015)	O estudo utiliza técnicas de <i>machine learning</i> para detectar padrões de alunos e prever o sucesso, baseado em dados dos fóruns dos MOOCs.
Attentive learner: Adaptive mobile MOOC learning via implicit cognitive states inference	Xiao, X., Pham, P., Wang, J.(2015)	O artigo apresenta um sistema de LA que utiliza dados como o batimento cardíaco do aluno, para interpretar seu estado cognitivo e de atenção. Pode ser usado como um auxiliar de feedback para LA.
A STEM MOOC for school children - What does learning analytics tell us?	Khalil, M., Ebner, M.(2015)	Os autores aplicam técnicas de LA em um MOOC austríaco para identificar o percurso dos alunos na plataforma.
Assessment activities in massive open on-line courses assessment activities in MOOCs	Muñoz-Merino, P.J., Ruipérez-Valiente, J.A., Moreno, J.L.S., Kloos, C.D.(2015)	O artigo aborda o uso de LA para avaliar as atividades de avaliação nos MOOCs.
Comparing MOOC-based platforms: Reflection on pedagogical support, framework and learning analytics	Lebron, D., Shahriar, H.(2015)	O artigo faz um comparativo entre 3 provedores de MOOCs, analisando o suporte pedagógico e as possibilidades de aplicação de LA em cada um deles.

Crowd-sourced learning in MOOCs: Learning analytics meets measurement theory	Milligan, S.(2015)	O artigo busca entender como acontece o aprendizado de alunos em MOOCs conectivistas, ou seja, baseado em fóruns, confrontando técnicas de LA e da teoria da medição.
Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance	Kennedy, G., De Barba, P., Coffrin, C., Corrin, L.(2015)	O artigo busca entender se há relação entre o sucesso de um aluno que cursa um MOOC e o nível de seu conhecimento e habilidades anteriores. Para essa análise foram utilizadas técnicas de LA.
Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach	Ezen-Can, A., Boyer, K.E., Kellogg, S., Booth, S.(2015)	O estudo foca na análise dos conteúdos de fóruns de MOOCs. Por meio de técnicas de clusterização, os autores traçaram perfis e fizeram comparações com anotações de pesquisadores da área. Depois realizaram uma análise qualitativa desses dados a fim de entender como funciona a aprendizagem nesses ambientes.
Student privacy self-management: Implications for learning analytics	Prinsloo, P., Slade, S.(2015)	O artigo faz uma análise profunda sobre a privacidade e questões éticas impostas pelo uso do LA.
Examining engagement: Analysing learner subpopulations in massive open online courses	Ferguson, R., Clow, D.(2015)	Os autores pesquisam sobre o engajamento e evasão de alunos em MOOCs, analisando tanto xMOOCs quanto cMOOCs. Eles descrevem 7 padrões de engajamento obtidos por técnicas de clusterização.
Combining Learning Analytics and the Flipped Classroom in a MOOC of maths	Martínez, D.R., Muñoz-Merino, P.J., Ruipérez-Valiente, J.A., Díaz, H.J.P., Ruiz, J.S.(2015)	O artigo descreve uma configuração para o uso de LA em um MOOC de matemática que utiliza a metodologia da sala de aula invertida.
Learning analytics on coursera event data: A proceeb mining approach	Mukala, P., Buijs, J., Leemans, M., Van Der Aalst, W.(2015)	O objetivo do artigo é obter insights do comportamento dos alunos e sua performance em MOOCs, utilizando técnicas de mineração de dados e utilizando coeficientes de correlação estatística.
Exploring forum dynamics in a Chinese MOOC: A longitudinal probabilistic social network analysis	Zhang, J., Skryabin, M. (2015)	O artigo explorar as dinâmicas de um fórum de um MOOC chinês, buscando entender a aprendizagem social. A técnica utilizada foi a análise longitudinal probabilística de redes sociais.

Learning analytics in linked open online courses	Hover, K.M., Muhlhauser, M. (2015)	Aplicação de técnicas de ontologias e web semântica para analisar MOOCs.
Every answer has a question: Exploring communication and knowledge exchange in MOOCs through learning analytics	Chounta, I., Hecking, T., Hoppe, H.U. (2015)	O estudo utiliza métricas e análise de redes para entender como acontece a troca de conhecimentos em fóruns de MOOCs.
Moving through MOOCs: Pedagogy, learning design and Patterns of Engagement	Ferguson, R., Clow, D., Beale, R., Bayne, S., Woodgate, A.(2015)	Os autores utilizaram uma análise de clusters para entender a relação entre a pedagogia aplicada e o engajamento de alunos de MOOCs.
Engagement measures in massive open online courses	Sinclair, J., Kalvala, S.(2015)	O artigo utiliza técnicas de LA para medir o engajamento dos alunos de MOOCs. No final propõe um framework para desenvolvimento e avaliação da aprendizagem.
Who will pass? Analyzing learner behaviors in MOOCs	Tseng, S.-F., Tsao, Y.-W., Yu, L.-C., Chan, C.-L., Lai, K.R.(2015)	Através de técnicas de clusterização, os autores encontraram 3 padrões de engajamento de alunos em MOOCs.
Recommendation systems for personalized technology-enhanced learning	Khribi, M.K., Jemni, M., Nasraoui, O.(2015)	Os autores apresentam a ideia de sistemas de recomendações personalizados baseados nos dados gerados pelos alunos e apresentam um framework.
Towards the development of a learning analytics extension in open edx	Ruiz, J.S., Díaz, H.J.P., Ruipérez-Valiente, J.A., Muñoz-Merino, P.J., Kloos, C.D.(2014)	Os autores elaboraram um plano de implementação de ferramentas de LA para os MOOCs da EDX, a fim de entender melhor o processo de aprendizagem.
Visualizing patterns of student engagement and performance in MOOCs	Coffrin, C., Corrin, L., De Barba, P., Kennedy, G.(2014)	O artigo debate uma série de abordagens de LA para analisar a performance e o engajamento de alunos de MOOCs.
Where is research on massive open online courses headed? A data analysis of the MOOC research initiative	Gašević, D., Kovanović, V., Joksimović, S., Siemens, G.(2014)	O artigo relata um panorama da pesquisa envolvendo LA e MOOCs e traz dados interessantes sobre essa área emergente.
Big data and learning analytics: A potential way to optimize elearning technological tools	García, O.A., Secades, V.A.(2013)	O artigo relata o potencial do uso da big data e do LA para analisar e entender melhor as ferramentas de aprendizagem online (MOOCs)
MOOCs and the funnel of participation	Clow, D. (2013)	O autor já relatava na época as altas taxas de evasão dos MOOCs e a necessidade de implementar técnicas de LA para entender o fenômeno.

APÊNDICE 2

Pesquisa de satisfação com aluno do curso de auriculoterapia:

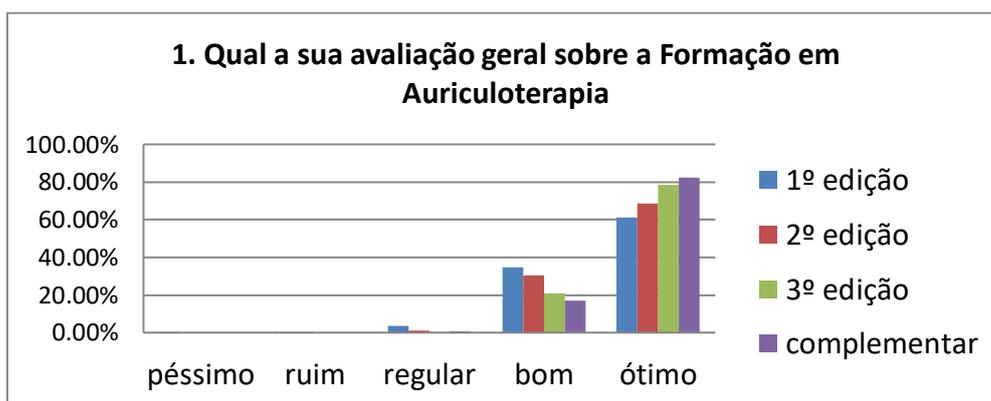
QUESTÕES FECHADAS: Foram 7 questões objetivas, respondidas numericamente na escala de um a cinco, correspondendo a péssimo, ruim, regular, bom e ótimo:

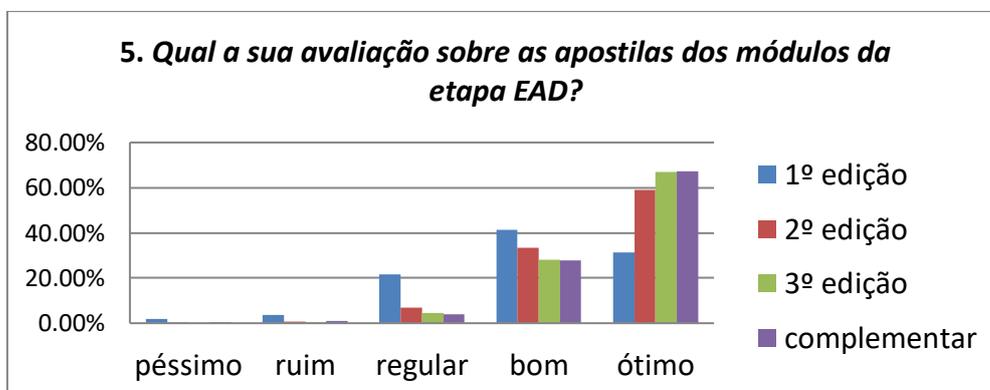
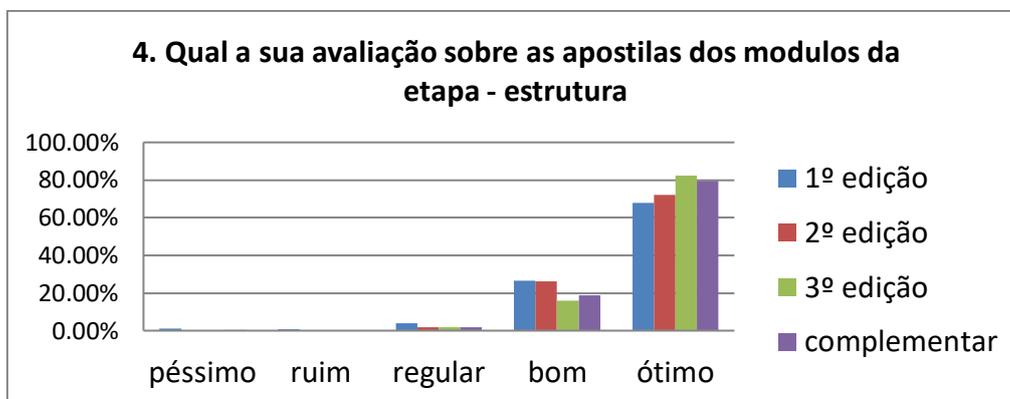
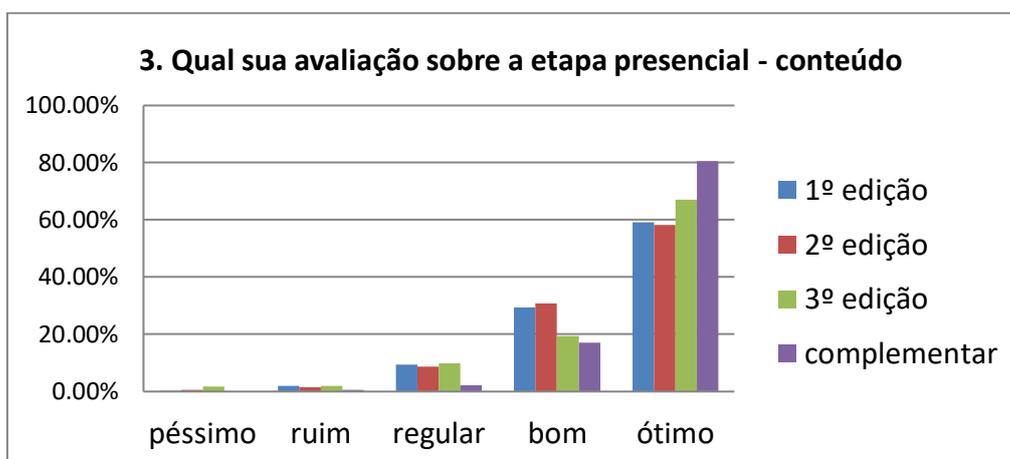
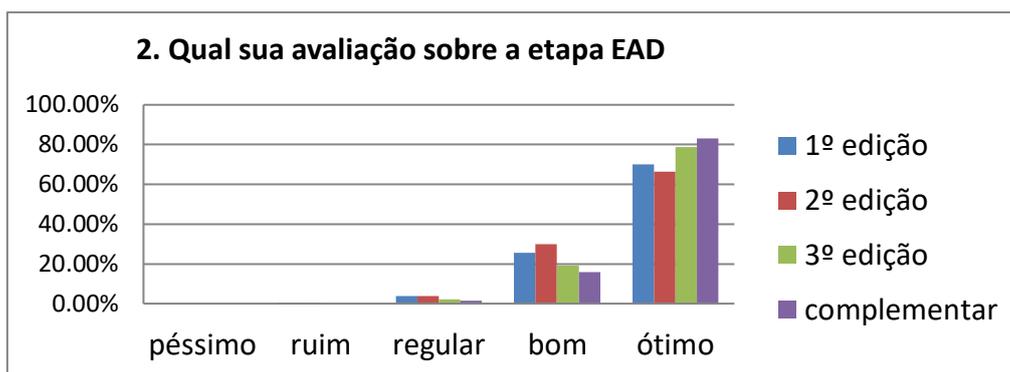
1. *Qual a sua avaliação geral sobre a Formação em Auriculoterapia?*
2. *Qual a sua avaliação sobre a Etapa Ead?*
3. *Qual sua avaliação sobre a etapa presencial – conteúdo?*
4. *Qual sua avaliação sobre a etapa presencial – estrutura?*
5. *Qual a sua avaliação sobre as apostilas dos módulos da etapa EAD?*
6. *Qual a sua autoavaliação sobre a plataforma de educação da etapa EAD?*
7. *Qual sua autoavaliação em relação a sua dedicação e aproveitamento nas atividades do curso?*

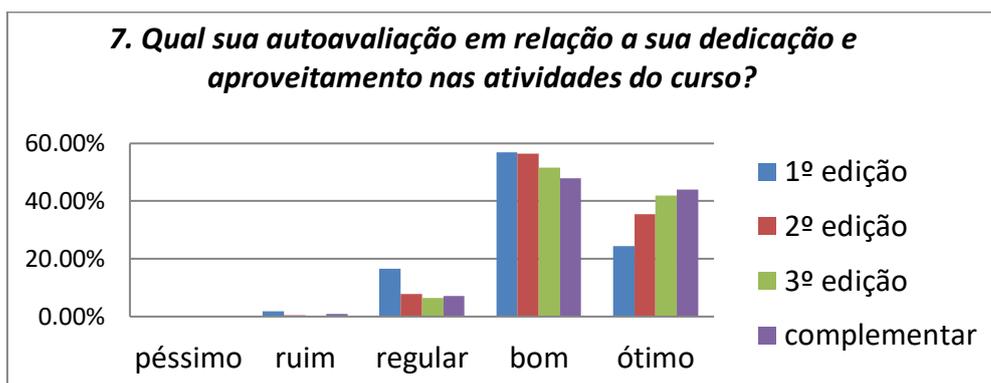
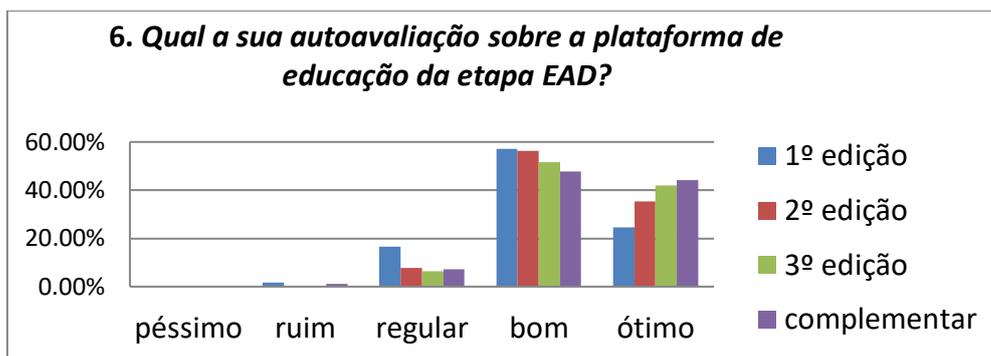
QUESTÕES ABERTAS: foram três questões abertas, que permitiram a livre dissertação dos alunos sobre suas impressões do curso:

7. *O que você mais gostou do curso?*
8. *Que sugestões você teria para melhorar o curso?*
9. *Outros comentários*

Uma descrição e análise das respostas dos alunos às questões fechadas segue abaixo em formato gráfico autoexplicativo, e a seguir apresentamos Quadros sintetizadores da análise das respostas abertas, agrupando-as por semelhança.







Respostas das questões abertas:

COMENTÁRIOS	O que você mais gostou do curso?			
	1ª EDIÇÃO	2ª EDIÇÃO	3ª EDIÇÃO	EDIÇÃO COMPL.
ALTA QUALIDADE DO MATERIAL DIDÁTICO	<p>“Material didático maravilhosos, linguagem fácil e acessível” (FO21)</p> <p>“O material em PDF é muito bom...” (Fpolis12)</p>	<p>“O material é excelente, ilustrado e as referências bibliográficas são fundamentais para aprofundar os estudos” (BH300)</p> <p>“Material didático excelente... Acrescentou muito as 3 teorias (MTC, Neuro e Reflexo)” (SP25)</p>	<p>“O conteúdo muito bom, material didático e aulas online muito boas. Parabéns!” (CU27)</p> <p>“O material teórico disponibilizado é excelente e também as ferramentas práticas à distancia” (FP41)</p>	<p>“Mais exercícios como os casos clínicos na etapa EAD (também localizando os pontos na orelha interativa” (CB136)</p> <p>“O material didático foi riquíssimo. Parabéns” (MA7)</p> <p>“Fiquei impressionada com a qualidade do material didático, organizado, bem estruturado e objetivo” (PL16)</p>

<p>DIFICULDADE DE ASSISTIR AS VÍDEOS-AULAS / BAIXA QUALIDADE DOS SLIDES</p>	<p>“Nas vídeo-aulas, os slides contém exatamente o texto da apostila, que é apenas lido, exatamente como está escrito. Isto torna as vídeo-aulas pouco atrativas Desse jeito era muito mais proveitoso ler a apostila que está muito boa. Sugiro então melhorar consideravelmente as vídeo-aulas.” (PO18) “Tornar vídeos mais dinâmicos, como menos textos e mais diálogos, imagens, dinâmicas” (REC65)</p>	<p>“Tive dificuldade para carregar os vídeos” (BH3) “Diminuir a resolução dos vídeos pois não consegui assistir” (CP163) “Melhorar em relação aos vídeos, pois não carregam” (CB21)</p>	<p>“Melhorar os vídeos pois são pesados para visualização (SA16) “Gostaria de baixar os vídeos mas não conseguimos” (SA19) “...]Não consegui assistir pois os mesmos não carregavam” (VII1)</p>	<p>“Não consegui baixar os vídeos, visto que no interior a internet não é de boa qualidade” (CB129)</p>
<p>PLATAFORMA EAD COM LIMITAÇÕES DE USO</p>	<p>“Gostei muito do curso, somente é necessário melhorar a plataforma.” (Fpolis02) “Não possibilita o uso em tablets e smartphones” (REC80) “Sugiro possuir mais facilidade no acesso, bem como layout melhorado” (FO34)</p>	<p>Sem registro</p>	<p>Sem registro</p>	<p>Sem registro</p>
<p>FALHAS NA COMUNICAÇÃO/ INFORMAÇÕES DA ETAPA EAD</p>	<p>“Melhorar comunicação e plataforma” (CAMP23) “Melhorar a publicização de inscrições/cursos” (PO2)</p>	<p>“Tive dificuldades de encontrar no site o local e o horário do presencial” (BHFO34)</p>	<p>“No link da terceira edição já deveria ter o link direcionado para o curso” (SA39) “Avisar sobre a abertura da plataforma por email” (VI71) “Melhorar a informação sobre a homologação. Fiz a inscrição, mas não recebi nenhum email avisando que estava homologado. Tive que ficar procurando” (FP131)</p>	<p>“Melhorar a comunicação com os alunos” (CB36) “Maior informação sobre datas de início, execução e finalização” (PL44)</p>
<p>NECESSIDADE DE TUTORIA OU</p>	<p>“Senti falta de um fórum de debates para</p>	<p>“Para melhor aprendizado, seria</p>	<p>Sem registro</p>	<p>“Só gostaria de sugerir o</p>

ALGUM APOIO A DISTÂNCIA	<i>tirar dúvidas ou de um tutor de grupo”(SP4) “A etapa EAD ser mais interativa.” (RJ34) “Comunicação mais permanente na plataforma a distância; talvez criação de fóruns durante o processo, criando condições dos profissionais do mesmo território se comunicarem e também, para quem quiser, fórum nacional” (REC103)</i>	<i>interessante sabermos o que erramos para estudar mais detalhadamente” (sobre exercício EAD) (SP79)</i>		<i>acompanhamento no ambiente virtual” (TE125)</i>
TEMPO DE REALIZAÇÃO INSUFICIENTE	<i>“Achei a etapa online muito curta”(CAMP32) “Gostei de tudo [...] Só achei pouco tempo para parte EAD”(CAMP 104)</i>	<i>“A duração do curso é algo curta. Acredito que uma a duas semanas a mais seria o ideal” (CP31)</i>	<i>“... Achei muito curto o prazo de tempo para leitura dos módulos, assistir aos vídeos e fazer avaliação... Poderiam ser 2 meses” (SA39) “Mais tempo para estudo na etapa à distância” (VI5)</i>	<i>“Achei pouco tempo para 5 apostilas, era bastante conteúdo para estudar e conciliar com o trabalho” (CB60) “Ter um tempo maior entre a liberação do curso e aula presencial” (PL46)</i>

COMENTÁRIOS	Que sugestões você teria para melhorar o curso?			
	1ª EDIÇÃO	2ª EDIÇÃO	3ª EDIÇÃO	EDIÇÃO COMPL.
TEMPO DE AULA INSUFICIENTE	<i>“Sugiro encontros “semestrais” para avaliar e comentar avanços, entraves, dificuldades, dúvidas” (JP60) “Mais aulas presenciais com carga horária maior” (MAC31)</i>	<i>“Ampliar a carga horária presencial, foi ótima mas muito acelerada” (CH196) “Acredito que a etapa presencial deveria ter mais horas para exercícios práticos e esclarecimentos de dúvidas” (BH58)</i>	<i>“... Ter pelo menos 8 horas” (CU3)</i>	<i>“que a aula presencial tenha um tempo maior” (MA11) “poderia ter uma maior carga horária prática-presencial” (TE49)</i>
ALTA QUALIDADE DA AULA E DOS PRECEPTORES	<i>“Aula prática sensacional, com uma didática maravilhosa através da revisão de todos os conteúdos” (FO22)</i>	<i>“Preceptores extremamente disponíveis e claros” (N54) “Parabéns por essa iniciativa. Curso extremamente proveitoso. Profissionais que</i>	<i>“Gostei muito do curso, mas a aula prática foi fantástica” (POA33)</i>	<i>“A etapa presencial foi bastante importante e trouxe muito aprendizado, além de sanar várias dúvidas” (MA55)</i>

	<p>“Conteúdo rico e profissionais competentes” (RJ54)</p> <p>“Transmissão do conhecimento em auriculoterapia de maneira simples e objetiva” (CAMP61)</p>	<p>ministram a etapa presencial são ótimos. Amei” (CP109)</p>		
DISPONIBILIZAÇÃO DO MATERIAL DA PRÁTICA P/ COMPRA	<p>“Poderia estar presente no curso alguém que comercializa os materiais para aquisição” (PO7)</p>	<p>“Informar locais para compra dos materiais” (N163)</p> <p>“Materiais podem ser vendidos na aula presencial” (CP145)</p>	<p>“Poderia convidar uma das lojas para podermos comprar o material” (POA14)</p>	<p>“Disponibilidade e das sementes, da pinça e do apalpador para entrega ou compra” (MA68)</p>
ESTRUTURA FÍSICA INADEQUADA E/OU APOIO LOCAL DEFICIENTE	<p>“Local da prática mais centralizado”(Fpolis 16)</p> <p>“Faltou água, sala com acústica ruim e calor, no banheiro sem pia para higiene das mãos” (REC56)</p> <p>“(…)Melhorar o local da aula prática e aumentar o tempo da aula prática” (JP59)</p>	<p>“Melhorar a estrutura física da aula” (N9)</p> <p>“Auxílio pra quem vem de fora (orientações, sugestões, hotéis...)” (CP87)</p>	<p>“Falta de estrutura e organização inicial” (POA66)</p> <p>“Local mais próximo ao centro da cidade” (FP138)</p>	<p>“(ter)Local físico da etapa presencial com estrutura de alimentação” (CB48)</p> <p>“que os colaboradores (apoiadores locais) fossem mais organizados, Polo de Porto Velho deixou a desejar... Dificultando até a atividade prática pelo atraso do material” (PV3)</p>
ALUNOS DE MESMO MUNICÍPIO EM TURMAS DIFERENTES	<p>Sem registro</p>	<p>“Flexibilidade em necessidade de mudança de período de prática quando devidamente justificado” (CH88)</p> <p>“Agrupar profissionais do mesmo município para o mesmo dia/horário” (CH132)</p>	<p>“Colocar os participantes do mesmo município no mesmo dia” (SA74)</p>	<p>Sem registro</p>
INFORMAÇÕES SOBRE ETAPA PRESENCIAL DEFICIENTES	<p>Sem registro</p>	<p>“O endereço do polo incorreto no site... E sem telefone de suporte” (CP17)</p>	<p>“Melhorar informações dos locais da aula prática (SA50)</p> <p>“Divulgação da data do encontro presencial via email ou</p>	<p>Sem registro</p>

			plataforma” (SA55)	
--	--	--	-----------------------	--

COMENTÁRIOS	Outros comentários			
	1ª EDIÇÃO	2ª EDIÇÃO	3ª EDIÇÃO	EDIÇÃO COMPLEMENTAR - EXEMPLOS
METODOLOGIA TEÓRICO-PRÁTICA ADEQUADA	Sem registro	“Gostei da metodologia teórico-prática” (N34) “Achei o curso excelente, desde o material didático até a plataforma muito fácil de usar. Filmes excelentes e achei muito fácil a absorção do conhecimento” (SP134)	“(gostei da) ... Possibilidade de fazer EAD e as dívidas sanadas na EP” (VI21) “(gostei da) ... Facilidade de fazer o curso, a maneira de passar o conhecimento pelas palestrantes e fácil aplicabilidade no ambiente de trabalho” (RJ104)	“... A forma como o curso está organizado foi eficiente e a etapa presencial potencializou a conclusão do curso” (MA62)
ALTO GRAU DE SATISFAÇÃO DO ALUNO	Parabéns pela iniciativa. Capacitar a auriculoterapia na Atenção Básica era um sonho que está tornando-se realidade. Esse curso é o que precisávamos para dar credibilidade e certificação para nós [...]. Por favor, continuem a realizar mais cursos como esse, voltado para práticas integrativas [...].” (REC10) O curso superou minhas expectativas, embora ter sido pouco tempo. EAD, o material e a forma de ensino foi de fácil compreensão. (SP 9)	“Amei o curso, estão de parabéns todos os envolvidos! Obrigada! (CH122) “Meus sinceros parabéns à equipe de profissionais envolvidos no curso por sua dedicação e extrema competência” (CH63) “Estou muito satisfeita com a iniciativa pois compreendo que é uma forma de universalidade e integralidade, amplia também o olhar dos profissionais de saúde e fortalece a própria PIC (política)” (BH156) “Quero dar aqui meus sinceros parabéns para toda a equipe, o curso é excelente” (N80) “Parabéns pela contribuição ao SUS e capacitação	“Foi muito bom o curso, fico grata pelo apoio de vcs” (VI45) “Estou plenamente satisfeita com o curso. Posso afirmar que acrescentou muito conhecimento tanto na minha vida pessoal como na profissional” (RJ40) “Parabéns pelo curso e pela iniciativa. Espero que o resultado seja um SUS mais humanizado e menos medicalizado” (FP43)	“Gostaria de parabenizar a todos que proporcionaram esse curso maravilhoso...” (CB93) “Agradecer ao MS e UFSC pela oportunidade de realização do curso, equipe técnica de excelência” (CB125) “Amei. Superou minhas expectativas. Gratidão pelo conhecimento compartilhado e por essa oportunidade” (TE29)

		<p><i>dos profissionais” (CP142)</i> <i>“O curso foi maravilhoso, estou super satisfeita e acho que valeu a pena o investimento de se deslocar 800Km. Parabéns à equipe! (CB52)</i></p>		
<p>APOIO E INCENTIVO À EQUIPE DO CURSO E PERSPECTIVAS DOS ALUNOS APÓS FINALIZAÇÃO DO CURSO</p>	<p><i>“Gostei da ideia de aplicar a auriculoterapia na AB, visto que ambas tem a preocupação com a prevenção de agravos e promoção de saúde. Reduzindo com isso o uso exagerado de medicação”(PO7)</i> <i>“A auriculoterapia é uma prática que irei levar para minha atuação como enfermeira na Unidade Básica de Saúde que trabalho, será muito rico este conhecimento para os pacientes atendidos no PSF uma excelente ideia capacitar os profissionais. Muito grata pelo conhecimento.” (REC27)</i> <i>“Adorei o curso, ele vai abrir outras portas para os profissionais que concluíram. Vai nos ajudar a trabalhar com outras formas de tratamento” (MAC01)</i></p>	<p><i>“... Me estimulou a procurar mais conhecimento, fiz outro curso online sobre PIC da MTC e comprei um livro de MTC” (CH122)</i> <i>“Já penso em um projeto no local de trabalho. Com certeza vou aplicar” (SP77)</i> <i>“Sugiro futuramente um curso de aprofundamento em auriculoterapia” (BH289)</i> <i>“Esse curso precisa continuar e também, quiçá, abranger outras pics” (N12)</i> <i>“Que se multiplique pelo Brasil inteiro cada vez mais” (N13)</i> <i>“Ótimo curso. Ter outros cursos de outras pics” (CP10)</i></p>	<p><i>“A equipe está de parabéns! Precisamos nos mobilizar para assegurar as conquistas de implantação das pics no SUS” (SA67)</i> <i>“Muito importante esse espaço como instrumento/estratégia para reorientação da prática em saúde de modo integral, democrático e equânime” (SA1)</i> <i>“Só tenho a agradecer pela oportunidade e em seguida já estarei colocando em prática” (RJ40)</i></p>	<p><i>“Amei o curso, aprendi muito, quero realizar troca de saberes com outros profissionais e multiplicar essa experiência ótima” (CB13)</i> <i>“O MS poderia realizar junto com a UFSC mais cursos voltados a outras pcs” (MN21)</i> <i>“Curso muito bom, será excelente colocar todos os aprendizados em prática” (TE25)</i> <i>“Aguardando os próximos cursos ansiosamente” (TE109)</i></p>

APÊNDICE 3

Modelo de questionário para área pedagógica

10/08/2020

Questionário sobre dados educacionais - Área Pedagógica / Conteúdo.

Questionário sobre dados educacionais - Área Pedagógica / Conteúdo.

Esse questionário faz parte da pesquisa de doutorado do aluno Breno Biagiotti (EGC/UFSC). O tema aborda o uso de técnicas de Learning Analytics para análise de dados educacionais em cursos online. Sua participação é fundamental para poder traçar um panorama assertivo do atual funcionamento do curso, com foco na captação de respostas qualitativas, para poder propor melhorias nas próximas edições.

Aspectos relativos ao conteúdo do curso.

Destinado a equipe pedagógica.

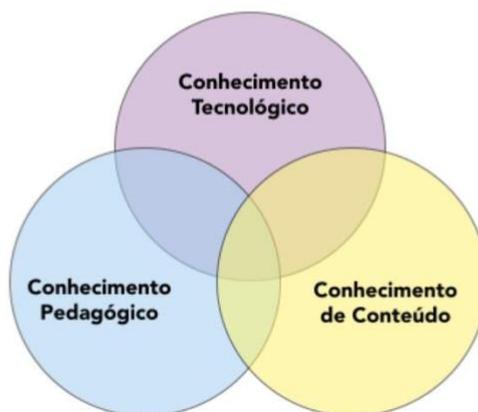
1. Quais foram os principais desafios na elaboração do conteúdo desse curso?

2. O material didático do curso tem correspondido com os objetivos de ensino? O que poderia ser feito para melhorar?

10/08/2020

Questionário sobre dados educacionais - Área Pedagógica / Conteúdo.

3. Uma boa aprendizagem online depende de 3 componentes principais: conhecimento tecnológico, conhecimento pedagógico e conhecimento de conteúdo, além da relação entre essas áreas (Mishra & Koehler, 2016). No curso de Auriculoterapia, qual sua opinião sobre o que pode ser melhorado em cada uma dessas áreas?



4. O design instrucional do curso online está bem projetado? Os alunos encontram facilidade no uso dos recursos? O que poderia melhorar?

- 5. Como funciona o suporte aos alunos? Há feedbacks personalizados? Há tutores para auxiliar o processo?

- 6. Há construção de conhecimento a partir da interação entre alunos? (fóruns, redes sociais, comunidades de prática)

- 7. O processo de avaliação está adequado? Os alunos chegam preparados na aula presencial?

- 8. Após a conclusão do curso online, os alunos participam de uma capacitação presencial. Quais são os principais desafios dessa etapa?

10/08/2020

Questionário sobre dados educacionais - Área Pedagógica / Conteúdo.

9. Já foi realizada alguma pesquisa de satisfação com os alunos do curso?

Marcar apenas uma oval.

Sim

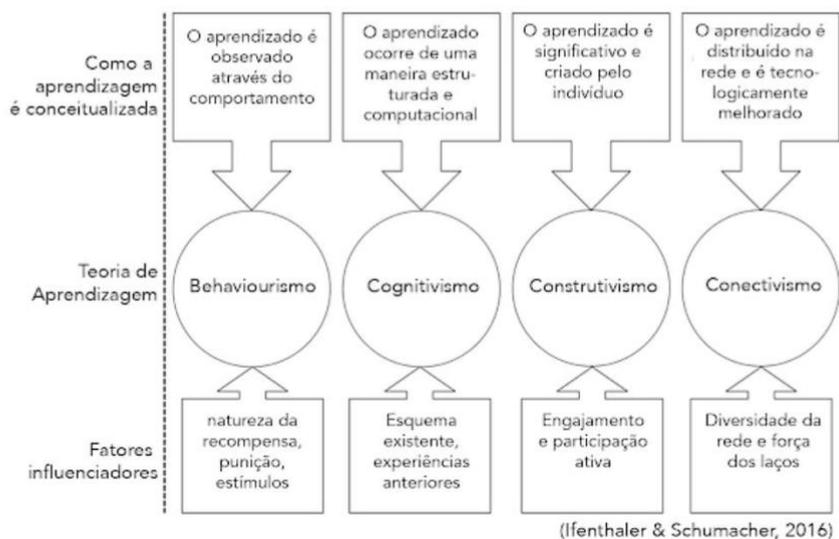
Não

10. Em relação às questões éticas referentes ao uso e armazenamento de dados dos alunos: há uma preocupação com a privacidade, propriedade, consentimento e transparência desses dados?

Aspectos Pedagógicos do curso.

11. Qual é a principal meta educacional do curso de Auriculoterapia?

12. Em sua opinião, o curso de Auriculoterapia se enquadra em qual dessas teorias de aprendizagem?



Marcar apenas uma oval.

- Behaviourismo
- Cognitívismo
- Construtívismo
- Conectívismo
- nenhuma das anteriores

10/08/2020

Questionário sobre dados educacionais - Área Pedagógica / Conteúdo.

13. Se pudesse voltar ao momento da criação do curso de Auriculoterapia, o que você teria feito de diferente? O que você julga que foi acertado e faria igual?

14. Como é feita atualmente a avaliação do curso por parte da equipe? Ocorre uma implementação de melhorias a cada nova edição? Notou alguma evolução ao longo das edições?

15. Há uma política de tomada de decisão baseada em dados nesse projeto? Costumam utilizar os dados armazenados para melhorar a gestão do curso?



16. Caso haja uma "cultura de dados" para elaboração de relatórios e tomada de decisões, quais métodos de análise são mais utilizados?



Marcar apenas uma oval.

- Estatística descritiva (média, mediana, moda, desvio padrão, variância...)
- Inferência estatística (Correlação, ANOVA, regressão)
- Mineração de dados (classificação, regressão, clusterização, associação...)
- Aprendizado de máquina (árvores de decisões, redes neurais, clusterização, K-Means...)
- Utilizo dados qualitativos dos alunos (Feedback, questionários)
- Análise de redes sociais
- Outro: _____

17. Deixe seu comentário ou sugestão de perguntas que possam contribuir com essa pesquisa!

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE 4

Modelo de questionário para área tecnológica (TI)

10/08/2020

Questionário sobre dados educacionais - Área técnica (TI)

Questionário sobre dados educacionais - Área técnica (TI)

Esse questionário faz parte da pesquisa de doutorado do aluno Breno Biagiotti (EGC/UFSC). O tema aborda o uso de técnicas de Learning Analytics para análise de dados educacionais em cursos online. Sua participação é fundamental para poder traçar um panorama assertivo do atual funcionamento do curso, com foco na captação de respostas qualitativas, para poder propor melhorias nas próximas edições. As respostas são sigilosas e nenhum nome será mencionado na pesquisa.

*Obrigatório

1. Classifique os processos de suporte tecnológico do curso.

Marcar apenas uma oval por linha.

	1- Ruim	2- Regular	3- Satisfatório	4- Bom	5- Ótimo
Sistema de inscrição dos alunos	<input type="radio"/>				
Matrícula/homologação	<input type="radio"/>				
Acesso ao Moodle	<input type="radio"/>				
Comunicação com os alunos	<input type="radio"/>				
Suporte aos alunos	<input type="radio"/>				
Avaliação online	<input type="radio"/>				
Emissão dos certificados	<input type="radio"/>				
Geração de relatórios	<input type="radio"/>				
Hospedagem do site	<input type="radio"/>				
Suporte Mobile	<input type="radio"/>				
Acesso aos dados	<input type="radio"/>				

2. Quais são as principais fontes para obtenção de dados nesse projeto? Como é feito o gerenciamento desses dados?

3. O Sistema de gerenciamento de aprendizagem utilizado (Moodle) tem atendido bem as expectativas e demandas do curso? Quais suas principais virtudes e limitações?

4. O Moodle está atualizado e devidamente configurado para o rastreamento das atividades (logs) dos alunos e também para geração de relatórios?

Marcar apenas uma oval.

- Sim
- Não, precisa de configuração/atualização

10/08/2020

Questionário sobre dados educacionais - Área técnica (TI)

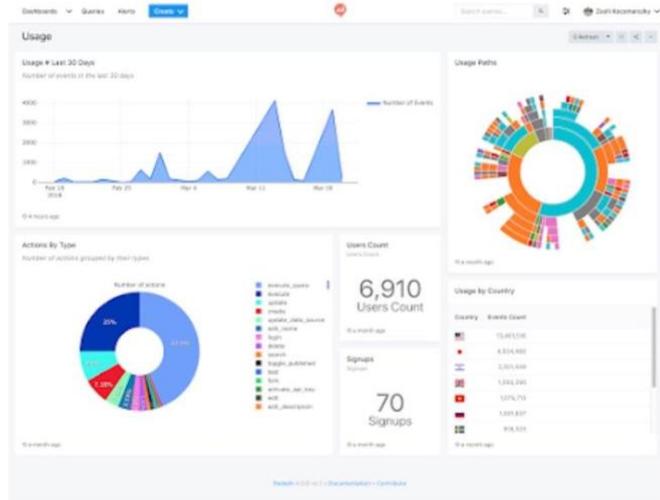
5. Que tipos de dados educacionais o sistema armazena atualmente? *



Marcar apenas uma oval.

- Dados demográficos do aluno
- Dados de presença online (logins, tempo online, frequência de acesso)
- Dados de atividades realizadas (logs)
- Dados de avaliação (provas)
- Dados de redes sociais (facebook, twitter, etc...)
- Dados qualitativos (fóruns, wikis, questionários...)
- Outro: _____

- 6. Vocês utilizam algum plugin de Analytics no Moodle? Como fazem a visualização dos dados?



- 7. Relate as principais dificuldades técnicas que surgiram ao longo do projeto. O que deu errado e precisou ser implementado? O que ainda precisa melhorar?

10/08/2020

Questionário sobre dados educacionais - Área técnica (TI)

8. Há uma política de tomada de decisão baseada em dados nesse projeto? Você considera eficiente o uso dos dados armazenados?



9. Levando em conta a sua área de atuação no projeto, o que poderia ser feito para melhorar o curso como um todo (focando na melhoria da aprendizagem dos alunos)?

10. Deixe seu comentário ou sugestão de perguntas que possam contribuir com essa pesquisa!

Este conteúdo não foi criado nem aprovado pelo Google.

APÊNDICE 5

Modelo de questionário para área administrativa

10/08/2020

Questionário sobre dados educacionais - Área Administrativa / Gestão.

Questionário sobre dados educacionais - Área Administrativa / Gestão.

Esse questionário faz parte da pesquisa de doutorado do aluno Breno Biagiotti (EGC/UFSC). O tema aborda o uso de técnicas de Learning Analytics para análise de dados educacionais em cursos online. Sua participação é fundamental para poder traçar um panorama assertivo do atual funcionamento do curso, com foco na captação de respostas qualitativas, para poder propor melhorias nas próximas edições. As respostas são sigilosas e nenhum nome será mencionado na pesquisa.

1. Há uma política de tomada de decisão baseada em dados nesse projeto? Você considera eficiente o uso dos dados armazenados?



2. Quais são as principais fontes para obtenção de dados nesse projeto? Como é feito o gerenciamento desses dados?

10/08/2020

Questionário sobre dados educacionais - Área Administrativa / Gestão.

3. Como é realizada atualmente a análise de dados dos cursos? Que tipos de dados são utilizados nos relatórios gerenciais?

4. Você utiliza algum tipo de dado educacional em seus relatórios?



Marcar apenas uma oval.

- Dados demográficos do aluno
- Dados de presença online (logins, tempo online, frequência de acesso)
- Dados de atividades realizadas (logs)
- Dados de avaliação (provas)
- Dados de redes sociais (facebook, twitter, etc...)
- Dados qualitativos (fóruns, wikis, questionários...)
- Outro: _____

5. Relate as principais dificuldades técnicas que surgiram ao longo do projeto. O que deu errado e precisou ser implementado? O que ainda precisa melhorar?

6. Levando em conta a sua área de atuação no projeto, o que poderia ser feito para melhorar o curso como um todo (focando na melhoria da aprendizagem dos alunos)?

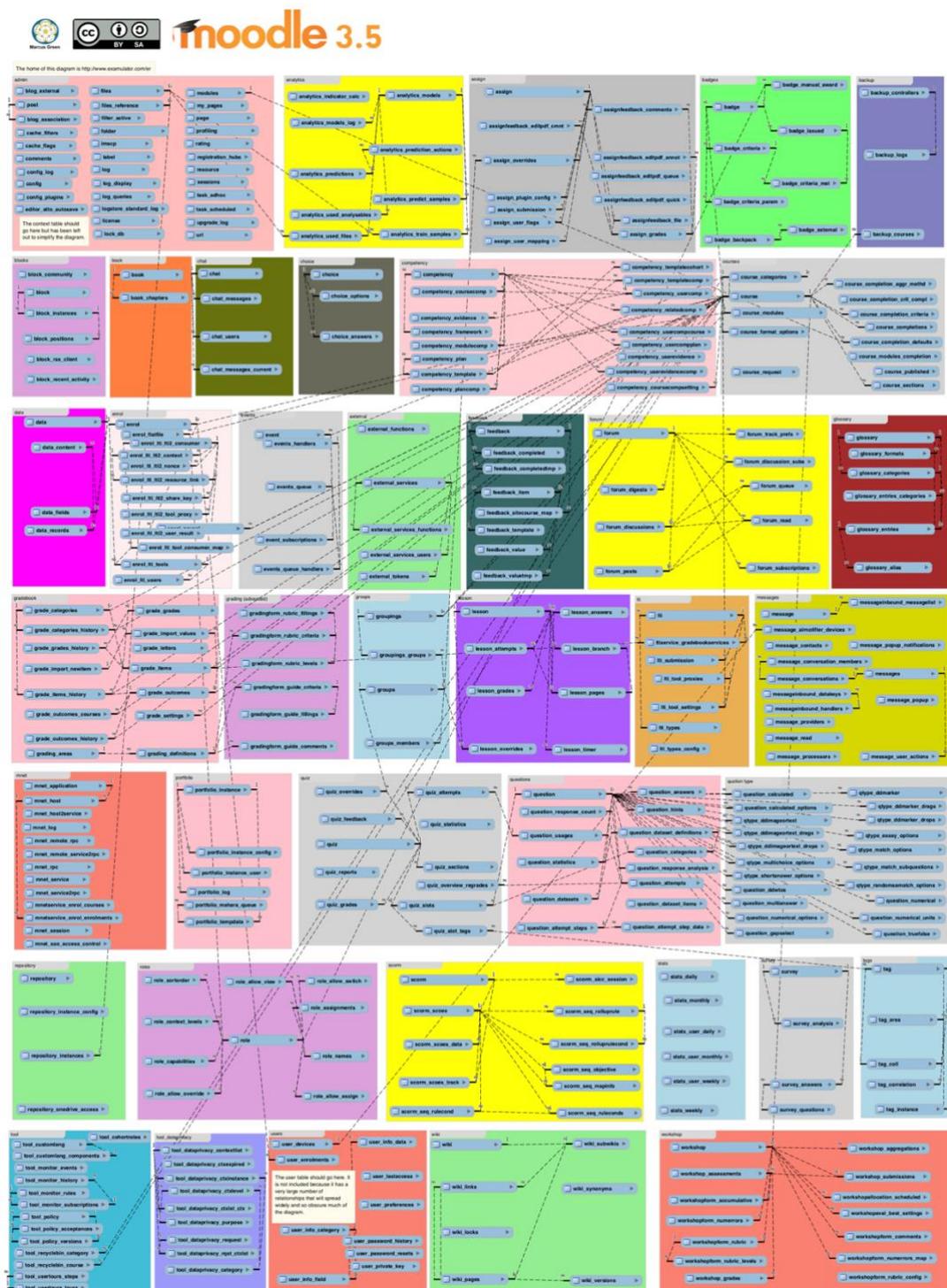
7. Deixe seu comentário ou sugestão de perguntas que possam contribuir com essa pesquisa!

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE 6

Fluxograma²⁵ da base de dados do Moodle 3.5



APÊNDICE 7

Consultas no banco de dados do Moodle, selecionando os dados com potencial para análise.

IDENTIFICANDO O "CURSO MOODLE" QUE TEMOS INTERESSE, ELE É O DE ID 2:

```
MARIADB [BRENO_AURICULO]> SELECT ID,FULLNAME FROM MOOD_COURSE;
```

```
+-----+-----+
| ID | FULLNAME |
+-----+-----+
| 1 | FORMAÇÃO EM AURICULOTERAPIA |
| 2 | FORMAÇÃO EM AURICULOTERAPIA PARA PROFISSIONAIS DE SAÚDE DA ATENÇÃO BÁSICA |
+-----+-----+
```

2 ROWS IN SET (0.001 SEC)

NÚMERO TOTAL DE CLIQUES: 433274

```
MARIADB [BRENO_AURICULO]> SELECT COUNT(*),COURSEID FROM MOOD_LOGSTORE_STANDARD_LOG GROUP BY COURSEID;
```

```
+-----+-----+
| COUNT(*) | COURSEID |
+-----+-----+
| 83528 | 0 |
| 36656 | 1 |
| 433274 | 2 |
+-----+-----+
```

3 ROWS IN SET (0.137 SEC)

NÚMERO DE LINKS CLICADOS, SERIA O NÚMERO DE CLIQUES, PROVAVELMENTE, UMA VEZ QUE A GENTE NÃO CONTA OS CLIQUES PARA CHEGAR NO AMBIENTE, QUE SÃO O DE COURSEID = 1.

TIPOS DE ATIVIDADES:

```
MARIADB [BRENO_AURICULO]> SELECT COUNT(*),M.NAME FROM MOOD_COURSE_MODULES CM JOIN MOOD_MODULES M ON CM.MODULE = M.ID WHERE COURSE = 2 GROUP BY MODULE;
```

```
+-----+-----+
| COUNT(*) | NAME |
+-----+-----+
| 1 | FORUM |
| 1 | LABEL |
| 5 | QUIZ |
+-----+-----+
```

```
| 7 | RESOURCE |
| 16 | SCORM |
+-----+-----+
5 ROWS IN SET (0.001 SEC)
```

QUANTIDADES DE TENTATIVAS POR QUIZ:

```
MARIADB [BRENO_AURICULO]> SELECT COUNT(*),Q.NAME FROM MOOD_QUIZ_ATTEMPTS QA JOIN MOOD_QUIZ Q
ON Q.ID = QA.QUIZ WHERE Q.COURSE = 2 GROUP BY Q.ID;
```

```
+-----+-----+
| COUNT(*) | NAME |
+-----+-----+
| 1722 | EXERCÍCIOS MÓDULO 1 |
| 1937 | EXERCÍCIOS MÓDULO 2 |
| 1668 | EXERCÍCIOS MÓDULO 3 |
| 1442 | EXERCÍCIOS MÓDULO 4 |
| 1918 | EXERCÍCIOS MÓDULO 5 |
+-----+-----+
5 ROWS IN SET (0.007 SEC)
```

MÉDIA DAS TENTATIVAS:

```
MARIADB [BRENO_AURICULO]> SELECT COUNT(*),Q.NAME,AVG(QA.SUMGRADES) FROM MOOD_QUIZ_ATTEMPTS
QA JOIN MOOD_QUIZ Q ON Q.ID = QA.QUIZ WHERE Q.COURSE = 2 GROUP BY Q.ID;
```

```
+-----+-----+-----+
| COUNT(*) | NAME | AVG(QA.SUMGRADES) |
+-----+-----+-----+
| 1722 | EXERCÍCIOS MÓDULO 1 | 3.492168740 |
| 1937 | EXERCÍCIOS MÓDULO 2 | 8.145423689 |
| 1668 | EXERCÍCIOS MÓDULO 3 | 8.191902703 |
| 1442 | EXERCÍCIOS MÓDULO 4 | 8.714550640 |
| 1918 | EXERCÍCIOS MÓDULO 5 | 7.369858570 |
+-----+-----+-----+
5 ROWS IN SET (0.028 SEC)
```

MÉDIA DAS NOTAS DOS QUESTIONÁRIOS (NOTA FINAL, NÃO A NOTA DE TODAS AS TENTATIVAS, SÓ DA TENTATIVA CONSIDERADA PARA NOTA DO QUESTIONÁRIO)

```
MARIADB [BRENO_AURICULO]> SELECT COUNT(*),Q.NAME,AVG(QG.GRADE) FROM MOOD_QUIZ_GRADES QG JOIN
MOOD_QUIZ Q ON Q.ID = QG.QUIZ WHERE Q.COURSE = 2 GROUP BY Q.ID;
```

```
+-----+-----+-----+
| COUNT(*) | NAME | AVG(QG.GRADE) |
```

```

+-----+-----+-----+
| 1187 | EXERCÍCIOS MÓDULO 1 | 9.554399452 |
| 1113 | EXERCÍCIOS MÓDULO 2 | 8.761311455 |
| 1054 | EXERCÍCIOS MÓDULO 3 | 8.937950654 |
| 1024 | EXERCÍCIOS MÓDULO 4 | 9.024983662 |
| 1019 | EXERCÍCIOS MÓDULO 5 | 8.080422217 |
+-----+-----+-----+
5 ROWS IN SET (0.017 SEC)

```

QUANTIDADE DE NOTAS FINAIS EM QUESTIONÁRIOS ONDE A NOTA DO ALUNO É MAIOR OU IGUAL A NOTA CONFIGURADA COMO NOTA DE APROVAÇÃO:

```

MARIADB [BRENO_AURICULO]> SELECT COUNT(*),Q.NAME FROM MOOD_QUIZ Q JOIN MOOD_GRADE_ITEMS GI ON
GI.ITEMINSTANCE = Q.ID AND ITEMMODULE = 'QUIZ' JOIN MOOD_GRADE_GRADES GG ON GG.ITEMID = GI.ID WHERE
GG.FINALGRADE >= GI.GRADEPASS GROUP BY Q.ID;

```

```

+-----+-----+
| COUNT(*) | NAME          |
+-----+-----+
| 1164 | EXERCÍCIOS MÓDULO 1 |
| 1088 | EXERCÍCIOS MÓDULO 2 |
| 1038 | EXERCÍCIOS MÓDULO 3 |
| 1016 | EXERCÍCIOS MÓDULO 4 |
| 953 | EXERCÍCIOS MÓDULO 5 |
+-----+-----+
5 ROWS IN SET (0.026 SEC)

```

QUANTIDADE DE NOTAS FINAIS EM QUESTIONÁRIOS ONDE A NOTA DO ALUNO É MENOR DO QUE A NOTA CONFIGURADA COMO NOTA DE APROVAÇÃO:

```

MARIADB [BRENO_AURICULO]> SELECT COUNT(*),Q.NAME FROM MOOD_QUIZ Q JOIN MOOD_GRADE_ITEMS GI ON
GI.ITEMINSTANCE = Q.ID AND ITEMMODULE = 'QUIZ' JOIN MOOD_GRADE_GRADES GG ON GG.ITEMID = GI.ID WHERE
GG.FINALGRADE < GI.GRADEPASS GROUP BY Q.ID;

```

```

+-----+-----+
| COUNT(*) | NAME          |
+-----+-----+
| 17 | EXERCÍCIOS MÓDULO 1 |
| 22 | EXERCÍCIOS MÓDULO 2 |
| 15 | EXERCÍCIOS MÓDULO 3 |
| 7 | EXERCÍCIOS MÓDULO 4 |
| 66 | EXERCÍCIOS MÓDULO 5 |
+-----+-----+
5 ROWS IN SET (0.023 SEC)

```

QUANTIDADE DE ACESSOS POR NOME DO EVENTO, COMPONENTE E AÇÃO:

```
MARIADB [BRENO_AURICULO]> SELECT COUNT(*),EVENTNAME,COMPONENT,ACTION FROM
MOOD_LOGSTORE_STANDARD_LOG WHERE COURSEID = 2 GROUP BY EVENTNAME,COMPONENT,ACTION;
```

COUNT(*)	EVENTNAME	COMPONENT	ACTION
1	\CORE\EVENT\COURSE_CREATED	CORE	CREATED
37	\CORE\EVENT\COURSE_MODULE_CREATED	CORE	CREATED
11	\CORE\EVENT\COURSE_MODULE_DELETED	CORE	DELETED
106	\CORE\EVENT\COURSE_MODULE_UPDATED	CORE	UPDATED
700	\CORE\EVENT\COURSE_RESOURCES_LIST_VIEWED	CORE	VIEWED
42	\CORE\EVENT\COURSE_SECTION_UPDATED	CORE	UPDATED
5	\CORE\EVENT\COURSE_UPDATED	CORE	UPDATED
90	\CORE\EVENT\COURSE_USER_REPORT_VIEWED	CORE	VIEWED
117901	\CORE\EVENT\COURSE_VIEWED	CORE	VIEWED
4	\CORE\EVENT\ENROL_INSTANCE_CREATED	CORE	CREATED
1	\CORE\EVENT\ENROL_INSTANCE_UPDATED	CORE	UPDATED
1833	\CORE\EVENT\GRADE_DELETED	CORE	DELETED
26	\CORE\EVENT\GROUP_CREATED	CORE	CREATED
2	\CORE\EVENT\GROUP_DELETED	CORE	DELETED
1109	\CORE\EVENT\GROUP_MEMBER_ADDED	CORE	ADDED
4	\CORE\EVENT\GROUP_MEMBER_REMOVED	CORE	REMOVED
8	\CORE\EVENT\GROUP_UPDATED	CORE	UPDATED
4	\CORE\EVENT\QUESTION_CATEGORY_CREATED	CORE	CREATED
1389	\CORE\EVENT\ROLE_ASSIGNED	CORE	ASSIGNED
15	\CORE\EVENT\ROLE_UNASSIGNED	CORE	UNASSIGNED
1406	\CORE\EVENT\USER_ENROLMENT_CREATED	CORE	CREATED
13	\CORE\EVENT\USER_ENROLMENT_DELETED	CORE	DELETED
47031	\CORE\EVENT\USER_GRADED	CORE	GRADED
3387	\CORE\EVENT\USER_LIST_VIEWED	CORE	VIEWED
1968	\CORE\EVENT\USER_PROFILE_VIEWED	CORE	VIEWED
50	\GRADEREPORT_GRADER\EVENT\GRADE_REPORT_VIEWED	GRADEREPORT_GRADER	VIEWED
1	\GRADEREPORT_HISTORY\EVENT\GRADE_REPORT_VIEWED	GRADEREPORT_HISTORY	VIEWED
3	\GRADEREPORT_OUTCOMES\EVENT\GRADE_REPORT_VIEWED	GRADEREPORT_OUTCOMES	VIEWED
2191	\GRADEREPORT_OVERVIEW\EVENT\GRADE_REPORT_VIEWED	GRADEREPORT_OVERVIEW	VIEWED
7	\GRADEREPORT_SINGLEVIEW\EVENT\GRADE_REPORT_VIEWED	GRADEREPORT_SINGLEVIEW	VIEWED
8679	\GRADEREPORT_USER\EVENT\GRADE_REPORT_VIEWED	GRADEREPORT_USER	VIEWED
1	\MOD_BOOK\EVENT\CHAPTER_CREATED	MOD_BOOK	CREATED

```

| 1 |\MOD_BOOK\EVENT\CHAPTER_VIEWED |MOD_BOOK |VIEWED |
| 3 |\MOD_FORUM\EVENT\COURSE_MODULE_VIEWED |MOD_FORUM |VIEWED |
| 118 |\MOD_FORUM\EVENT\USER_REPORT_VIEWED |MOD_FORUM |VIEWED |
| 31 |\MOD_QUIZ\EVENT\ATTEMPT_PREVIEW_STARTED |MOD_QUIZ |STARTED |
| 10258 |\MOD_QUIZ\EVENT\ATTEMPT_REVIEWED |MOD_QUIZ |REVIEWED |
| 8678 |\MOD_QUIZ\EVENT\ATTEMPT_STARTED |MOD_QUIZ |STARTED |
| 8345 |\MOD_QUIZ\EVENT\ATTEMPT_SUBMITTED |MOD_QUIZ |SUBMITTED |
| 10797 |\MOD_QUIZ\EVENT\ATTEMPT_SUMMARY_VIEWED |MOD_QUIZ |VIEWED |
| 96394 |\MOD_QUIZ\EVENT\ATTEMPT_VIEWED |MOD_QUIZ |VIEWED |
| 548 |\MOD_QUIZ\EVENT\COURSE_MODULE_INSTANCE_LIST_VIEWED |MOD_QUIZ |VIEWED |
| 21131 |\MOD_QUIZ\EVENT\COURSE_MODULE_VIEWED |MOD_QUIZ |VIEWED |
| 45 |\MOD_QUIZ\EVENT\EDIT_PAGE_VIEWED |MOD_QUIZ |VIEWED |
| 25 |\MOD_QUIZ\EVENT\REPORT_VIEWED |MOD_QUIZ |VIEWED |
| 21491 |\MOD_RESOURCE\EVENT\COURSE_MODULE_VIEWED |MOD_RESOURCE |VIEWED |
| 999 |\MOD_SCORM\EVENT\COURSE_MODULE_INSTANCE_LIST_VIEWED |MOD_SCORM |VIEWED |
| 32333 |\MOD_SCORM\EVENT\COURSE_MODULE_VIEWED |MOD_SCORM |VIEWED |
| 35 |\MOD_SCORM\EVENT\REPORT_VIEWED |MOD_SCORM |VIEWED |
| 33536 |\MOD_SCORM\EVENT\SCO_LAUNCHED |MOD_SCORM |LAUNCHED |
| 1 |\MOD_SCORM\EVENT\TRACKS_VIEWED |MOD_SCORM |VIEWED |
| 11 |\REPORT_LOGLIVE\EVENT\REPORT_VIEWED |REPORT_LOGLIVE |VIEWED |
| 31 |\REPORT_LOG\EVENT\REPORT_VIEWED |REPORT_LOG |VIEWED |
| 116 |\REPORT_LOG\EVENT\USER_REPORT_VIEWED |REPORT_LOG |VIEWED |
| 16 |\REPORT_OUTLINE\EVENT\ACTIVITY_REPORT_VIEWED |REPORT_OUTLINE |VIEWED |
| 87 |\REPORT_OUTLINE\EVENT\REPORT_VIEWED |REPORT_OUTLINE |VIEWED |
| 219 |\REPORT_PARTICIPATION\EVENT\REPORT_VIEWED |REPORT_PARTICIPATION |VIEWED |
+-----+-----+-----+-----+

```

57 ROWS IN SET (1.365 SEC)

AGORA PARA O DUMP MAIOR:

IDENTIFICANDO OS CURSOS MOODLE

```
MARIADB [BRENO_GRUPOS_AURICULO]> SELECT ID,FULLNAME FROM COURSE;
```

```

+----+-----+
| ID | FULLNAME |
+----+-----+
| 1 | GRUPOSAURICULO |
| 2 | AURICULOTERAPIA 2ED. BELO HORIZONTE (MG) |
| 24 | AURICULOTERAPIA 2ED. CAMPINAS (SP) |
| 27 | AURICULOTERAPIA 2ED. CHAPECÓ (SC) |
| 30 | AURICULOTERAPIA 2ED. SÃO LUÍS (MA) |

```

- | 31 | AURICULOTERAPIA 2ED. NATAL (RN) |
- | 32 | AURICULOTERAPIA 2ED. CUIABÁ (MT) |
- | 33 | AURICULOTERAPIA 2ED. SÃO PAULO (SP) |
- | 34 | AURICULOTERAPIA 3ED. SALVADOR (BA) |
- | 35 | AURICULOTERAPIA 3ED. CAMPO GRANDE (MS) |
- | 36 | AURICULOTERAPIA 3ED. VITÓRIA (ES) |
- | 37 | AURICULOTERAPIA 3ED. PORTO ALEGRE (RS) |
- | 38 | AURICULOTERAPIA 3ED. RIO DE JANEIRO (RJ) |
- | 39 | AURICULOTERAPIA 3ED. FLORIANÓPOLIS (SC) |
- | 40 | RESIDENTES FLORIPA |
- | 41 | AURICULOTERAPIA POLO CURITIBA |
- | 42 | AURICULOTERAPIA POLO MACEIÓ |
- | 43 | AURICULOTERAPIA POLO MANAUS |
- | 44 | AURICULOTERAPIA POLO PALMAS |
- | 45 | AURICULOTERAPIA POLO PORTO VELHO |
- | 46 | AURICULOTERAPIA POLO TERESINA |
- | 47 | RESIDENTES SÃO PAULO |
- | 48 | BELO HORIZONTE (MG) |
- | 49 | FLORIANÓPOLIS (SC) |
- | 50 | NITERÓI (RJ) |
- | 51 | SÃO JOSÉ DO RIO PRETO (SP) |
- | 52 | TERESINA (PI) |
- | 53 | BELÉM (PA) |
- | 54 | GOIÂNIA (GO) |
- | 55 | MANAUS (AM) |
- | 56 | MARABÁ (PA) |
- | 57 | MONTES CLAROS (MG) |
- | 58 | PALMAS (TO) |
- | 59 | PETROLINA (PE) |
- | 60 | RECIFE (PE) |
- | 61 | RIO BRANCO (AC) |
- | 62 | RIO DE JANEIRO (RJ) |
- | 63 | SANTA MARIA (RS) |
- | 64 | SANTARÉM (PA) |
- | 65 | SANTO ANDRÉ (SP) |
- | 66 | SÃO PAULO (SP) |
- | 67 | RESIDENTES FLORIPA 2018.1 |
- | 68 | AURICULOTERAPIA ESPECIAL MS |
- | 69 | AURICULOTERAPIA ESPECIAL DF |

70	AURICULOTERAPIA ESPECIAL CUNHA PORÃ
71	BELO HORIZONTE (MG) 2019
72	BRASÍLIA (DF) 2019
73	CAMPO GRANDE (MS) 2019
74	CANOAS (RS) 2019
75	CRICIÚMA (SC) 2019
76	CASCAVEL (PR) 2019
77	CUIABÁ (MT) 2019
78	FORTALEZA (CE) 2019
79	PASSO FUNDO (RS) 2019
80	RIO BRANCO (AC) 2019
81	SÃO CARLOS (SP) 2019
82	SÃO PAULO (SP) 2019
83	BELÉM (PA) 2019.2
84	FLORIANÓPOLIS (SC) 2019.2
85	JOINVILLE (SC) 2019.2
86	MANAUS (AM) 2019.2
87	PORTO ALEGRE (RS) 2019.2

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

62 ROWS IN SET (0.001 SEC)

APÊNCICE 8

Análise de dados exploratória realizada no Microsoft Power BI.

31/03/2021

Projeto Breno_30mar

Projeto Breno Biagiotti

Curso Aurículo

Análise Exploratória

In [1]:

```
import pandas as pd
import seaborn as sns
```

In [2]:

```
dados_originais = pd.read_csv('base_CSV.csv', sep = ';', encoding = 'latin-1', decimal = b',')
```

Dados - desconsidera os alunos que fizeram abandono do curso

In [3]:

```
dados_apr_rpr = dados_originais.loc[(dados_originais['qtd_quiz_realizado']>0)]
len(dados_apr_rpr)
```

Out[3]:

11591

Evasão

2.207 (16%) alunos NÃO fizeram nenhum quiz

In [4]:

```
dados_evasao = dados_originais.loc[(dados_originais['qtd_quiz_realizado']==0)]
len(dados_evasao)
```

Out[4]:

2207

In [5]:

```
dados_evasao_resumo = pd.read_csv('base_evasao_CSV.csv', sep = ';', encoding = 'latin-1', decimal = ',')
dados_evasao_resumo
```

Out[5]:

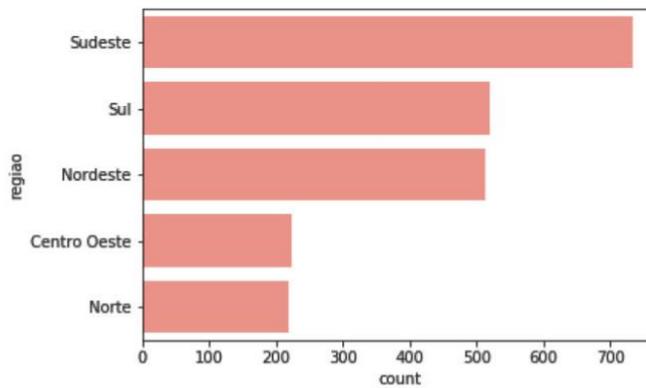
	regiao	evasao	n_evasao	total	taxa
0	Centro Oeste	223	868	1091	0.2044
1	Nordeste	513	1752	2265	0.2265
2	Norte	218	551	769	0.2835
3	Sudeste	733	4317	5050	0.1451
4	Sul	520	4103	4623	0.1125
5	Total Geral	2207	11591	13798	0.1600

In [6]:

```
# NÚMEROS ABSOLUTOS
sns.countplot(y='regiao', data=dados_evasao, color='salmon')
```

Out[6]:

<matplotlib.axes._subplots.AxesSubplot at 0x1c9bea74da0>



31/03/2021

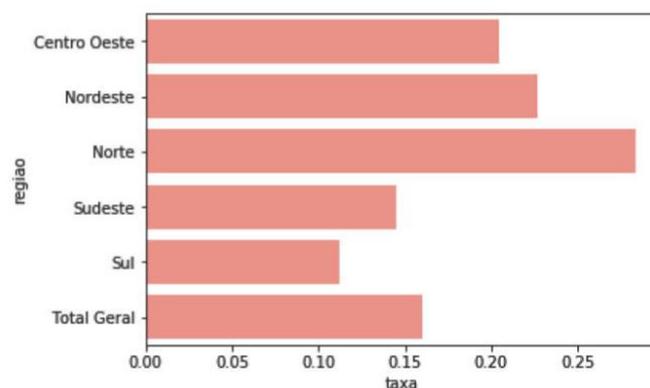
Projeto Breno_30mar

In [7]:

```
# PERCENTUAL  
sns.barplot(x='taxa', y='regiao', data=dados_evasao_resumo, color='salmon')
```

Out[7]:

<matplotlib.axes._subplots.AxesSubplot at 0x1c9beb8d438>



Evento Presencial

Considerado somente para cursos regulares (piloto nao tinha o evento)

base 10.400

foram ao evento 8.829 (84,8%)

NÃO foram ao evento e estavam aprovados 669 (7,2% dos com média >= 6)

foram no evento e estavam reprovados 170 (9,7% dos reprovados)

In [8]:

```
dados_presencial = dados_apr_rpr.loc[(dados_apr_rpr['tag']=='regular')]  
len(dados_presencial)
```

Out[8]:

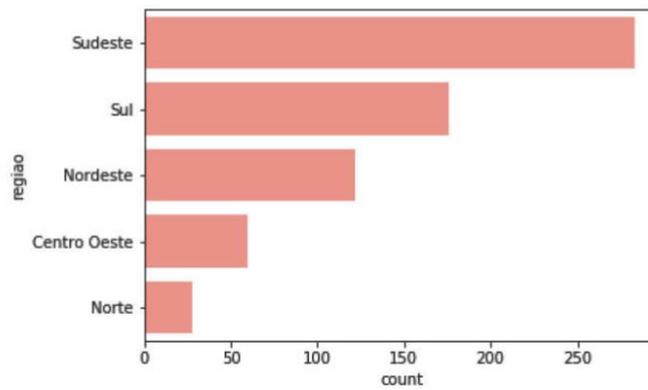
10400

In [9]:

```
# NÚMEROS ABSOLUTOS  
# MEDIA >= 6 e NÃO FOI AO EVENTO  
sns.countplot(y='regiao', data=dados_presencial.loc[(dados_presencial['graderfinal']>=6)  
& (dados_presencial['presencial2']==0)], color='salmon')
```

Out[9]:

<matplotlib.axes._subplots.AxesSubplot at 0x1c9bf2d5828>



In [10]:

```
len(dados_presencial.loc[(dados_presencial['graderfinal']>=6) & (dados_presencial['presencial2']==0)])
```

Out[10]:

669

Aprovados/ Reprovados

9.676 (83,5%) alunos APROVADOS 1.915 (16,5%) alunos REPROVADOS

31/03/2021

Projeto Breno_30mar

In [11]:

```
dados_apr_rpr_resumo = pd.read_csv('base_apr_rpr_CSV.csv', sep = ';', encoding = 'latin-1', decimal = ',')
dados_apr_rpr_resumo
```

Out[11]:

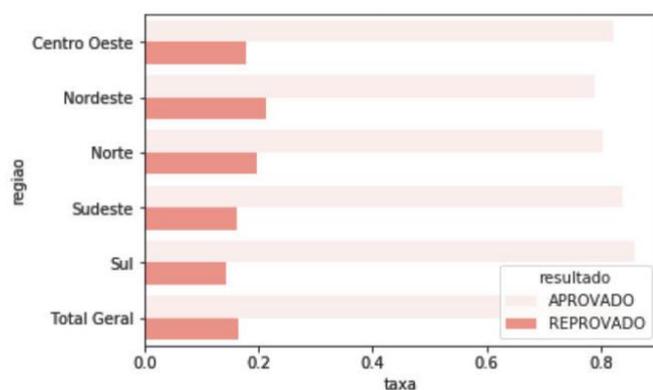
	regiao	resultado	qtd	taxa
0	Centro Oeste	APROVADO	714	0.8226
1	Centro Oeste	REPROVADO	154	0.1774
2	Nordeste	APROVADO	1.381	0.7882
3	Nordeste	REPROVADO	371	0.2118
4	Norte	APROVADO	442	0.8022
5	Norte	REPROVADO	109	0.1978
6	Sudeste	APROVADO	3.618	0.8381
7	Sudeste	REPROVADO	699	0.1619
8	Sul	APROVADO	3.521	0.8582
9	Sul	REPROVADO	582	0.1418
10	Total Geral	APROVADO	9.676	0.8348
11	Total Geral	REPROVADO	1.915	0.1652

In [12]:

```
# PERCENTUAL
sns.barplot(x='taxa', y='regiao', hue='resultado', data=dados_apr_rpr_resumo, color='salmon')
```

Out[12]:

<matplotlib.axes._subplots.AxesSubplot at 0x1c9bf59f550>



Pré-Aprovados que NÃO fizeram os 5 quiz, mas atingiram a média mínima de 6

156 (1,5%) alunos com MÉDIA maior ou igual a 6

In [13]:

```
len(dados_apr_rpr.loc[(dados_apr_rpr['gradefinal2']=='SIM')
                      & (dados_apr_rpr['qtd_quiz_realizado']>0)
                      & (dados_apr_rpr['qtd_quiz_realizado']<5)])
```

Out[13]:

156

Correlação - Quiz

In [14]:

```
del dados_apr_rpr['userid2']
del dados_apr_rpr['firstaccess']
del dados_apr_rpr['lastaccess']
del dados_apr_rpr['userid']
del dados_apr_rpr['courseid']
del dados_apr_rpr['chaveuser']
del dados_apr_rpr['chavecourse']
del dados_apr_rpr['chavecourseuser']
del dados_apr_rpr['resultado']
del dados_apr_rpr['presencial']
del dados_apr_rpr['quiz_media']
del dados_apr_rpr['qtd_quiz_realizado']
del dados_apr_rpr['qtd_quiz_NAO_realizado']
del dados_apr_rpr['quiz_media_1']
del dados_apr_rpr['quiz_media_2']
del dados_apr_rpr['quiz_media_3']
del dados_apr_rpr['quiz_media_4']
del dados_apr_rpr['quiz_media_5']
del dados_apr_rpr['log_downloaded']
```

In [16]:

```
dados_quiz = dados_apr_rpr.copy()
del dados_quiz['courseid2']
del dados_quiz['resultado2']
del dados_quiz['presencial2']
del dados_quiz['regiao']
del dados_quiz['uf']
del dados_quiz['cidade']
del dados_quiz['tag']
```

31/03/2021

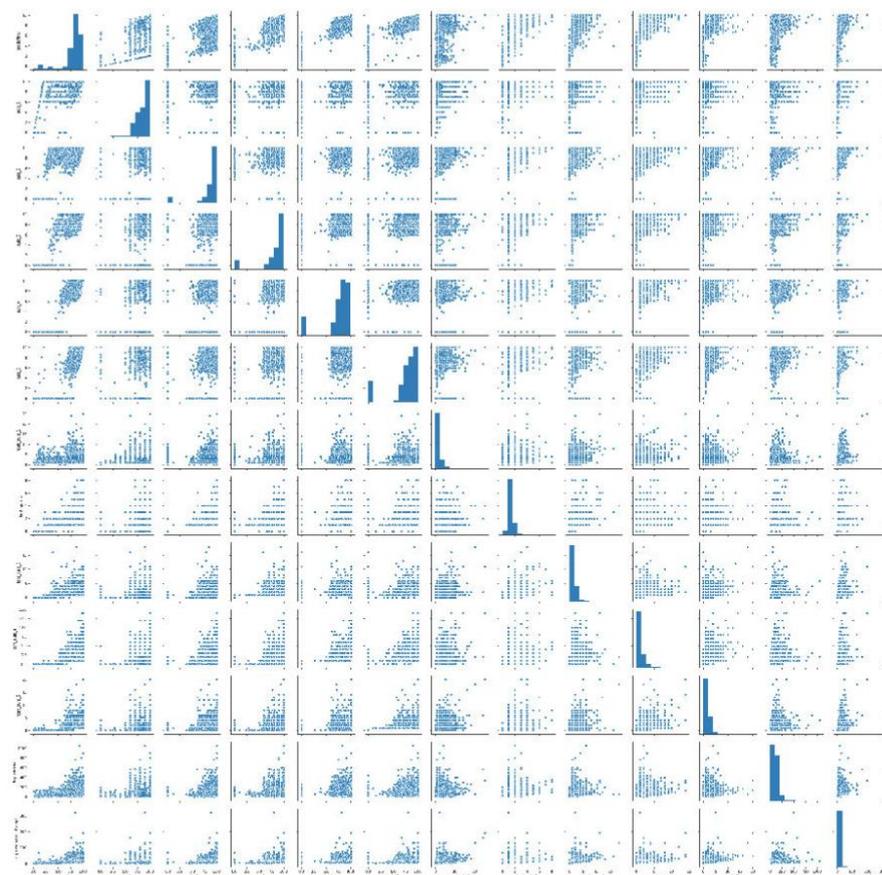
Projeto Breno_30mar

In [17]:

```
sns.pairplot(dados_quiz)
```

Out[17]:

```
<seaborn.axisgrid.PairGrid at 0x1c9bf634cc0>
```



In [18]:

```
dados_quiz.corr(method = 'spearman')
```

Out[18]:

	gradefinal	quiz_1	quiz_2	quiz_3	quiz_4	quiz_5	tent_1
gradefinal	1.000000	0.612055	0.595655	0.739914	0.740399	0.752429	-0.0
quiz_1	0.612055	1.000000	0.267756	0.332068	0.345676	0.264695	-0.0
quiz_2	0.595655	0.267756	1.000000	0.410137	0.376806	0.379208	-0.0
quiz_3	0.739914	0.332068	0.410137	1.000000	0.496045	0.465647	-0.0
quiz_4	0.740399	0.345676	0.376806	0.496045	1.000000	0.477018	-0.0
quiz_5	0.752429	0.264695	0.379208	0.465647	0.477018	1.000000	-0.0
tent_quiz_1	-0.088147	-0.082853	-0.040741	-0.079085	-0.055201	-0.042426	1.0
tent_quiz_2	0.334998	0.229157	0.317774	0.276113	0.296667	0.262257	0.2
tent_quiz_3	0.305590	0.138987	0.215991	0.301490	0.286031	0.274008	0.2
tent_quiz_4	0.396400	0.149105	0.270995	0.333913	0.378043	0.374482	0.2
tent_quiz_5	0.349749	0.110119	0.199588	0.259697	0.337366	0.357517	0.2
log_viewed	0.268070	0.069200	0.214905	0.237942	0.233927	0.271885	0.1
log_reviewed_attempt	0.216250	0.052429	0.159021	0.176117	0.216310	0.234818	0.5

31/03/2021

Projeto Breno_30mar

In [19]:

```
dados_quiz.corr(method = 'pearson')
```

Out[19]:

	gradefinal	quiz_1	quiz_2	quiz_3	quiz_4	quiz_5	tent_1
gradefinal	1.000000	0.455331	0.840965	0.918326	0.928890	0.897119	-0.0
quiz_1	0.455331	1.000000	0.315795	0.319571	0.307476	0.264392	-0.0
quiz_2	0.840965	0.315795	1.000000	0.756729	0.695885	0.647203	-0.0
quiz_3	0.918326	0.319571	0.756729	1.000000	0.826918	0.765686	-0.0
quiz_4	0.928890	0.307476	0.695885	0.826918	1.000000	0.853257	-0.0
quiz_5	0.897119	0.264392	0.647203	0.765686	0.853257	1.000000	-0.0
tent_quiz_1	-0.018779	-0.055886	-0.008857	-0.020833	-0.007709	-0.004583	1.0
tent_quiz_2	0.430041	0.247711	0.460410	0.380871	0.361927	0.328524	0.2
tent_quiz_3	0.372704	0.129037	0.317030	0.386546	0.342399	0.314473	0.3
tent_quiz_4	0.435826	0.164401	0.328718	0.388809	0.442399	0.413402	0.3
tent_quiz_5	0.384886	0.102947	0.275145	0.330436	0.388341	0.413100	0.3
log_viewed	0.366373	0.094014	0.318029	0.339657	0.346853	0.348248	0.1
log_reviewed_attempt	0.221436	0.052263	0.185174	0.200906	0.215663	0.216751	0.5

In [20]:

```
dados_quiz.corr(method = 'kendall')
```

Out[20]:

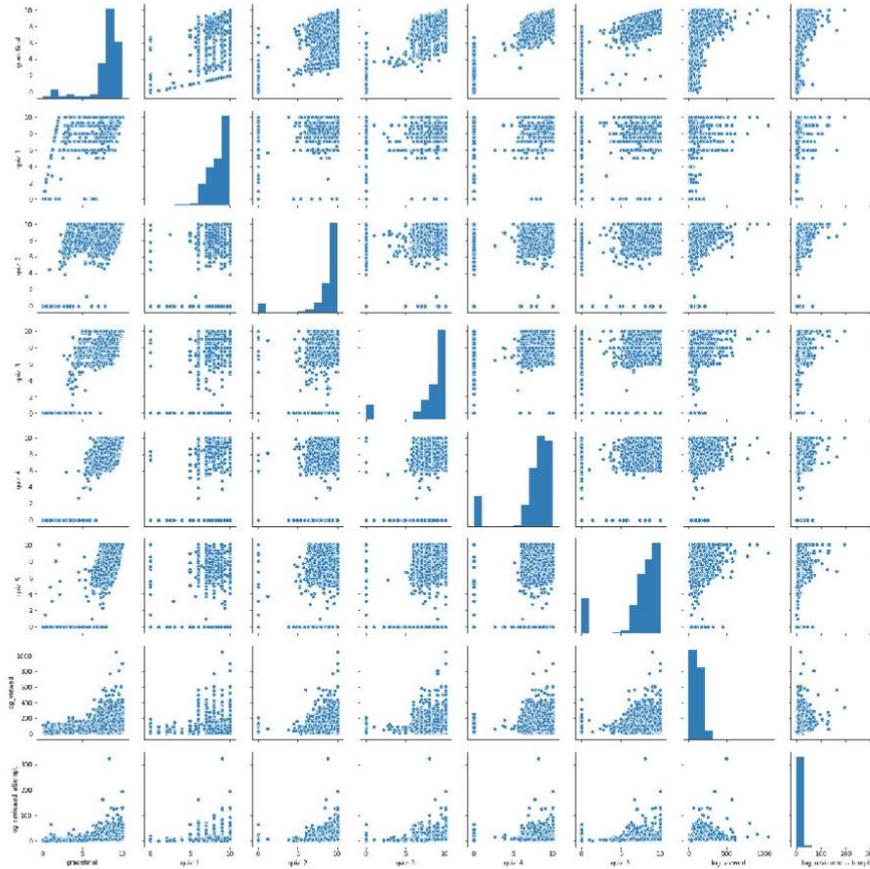
	gradefinal	quiz_1	quiz_2	quiz_3	quiz_4	quiz_5	tent_1
gradefinal	1.000000	0.485219	0.446533	0.594421	0.562721	0.583995	-0.0
quiz_1	0.485219	1.000000	0.207710	0.267012	0.262529	0.204149	-0.0
quiz_2	0.446533	0.207710	1.000000	0.319158	0.274586	0.283099	-0.0
quiz_3	0.594421	0.267012	0.319158	1.000000	0.380343	0.363686	-0.0
quiz_4	0.562721	0.262529	0.274586	0.380343	1.000000	0.350418	-0.0
quiz_5	0.583995	0.204149	0.283099	0.363686	0.350418	1.000000	-0.0
tent_quiz_1	-0.068486	-0.070022	-0.032795	-0.065792	-0.043064	-0.034003	1.0
tent_quiz_2	0.271108	0.200377	0.266196	0.237591	0.241049	0.216961	0.2
tent_quiz_3	0.240440	0.119528	0.175030	0.253785	0.226058	0.221082	0.2
tent_quiz_4	0.313172	0.127131	0.220776	0.281211	0.299333	0.304136	0.2
tent_quiz_5	0.268769	0.093036	0.158669	0.212431	0.262927	0.283664	0.2
log_viewed	0.183331	0.050567	0.151669	0.172795	0.158624	0.190126	0.1
log_reviewed_attempt	0.151080	0.039823	0.115236	0.129474	0.150851	0.168746	0.4

In [21]:

```
sns.pairplot(dados_quiz, vars=['gradefinal', 'quiz_1', 'quiz_2', 'quiz_3', 'quiz_4', 'quiz_5', 'log_viewed', 'log_reviewed_attempt'])
```

Out[21]:

```
<seaborn.axisgrid.PairGrid at 0x1c9c8e3b780>
```



In [22]:

```
del dados_quiz['tent_quiz_1']  
del dados_quiz['tent_quiz_2']  
del dados_quiz['tent_quiz_3']  
del dados_quiz['tent_quiz_4']  
del dados_quiz['tent_quiz_5']
```

Piloto x Regular: acesso

31/03/2021

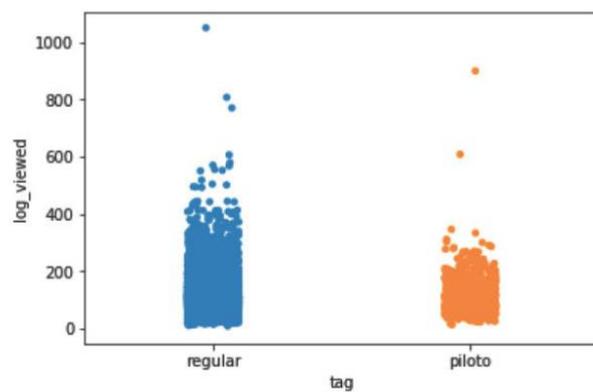
Projeto Breno_30mar

In [24]:

```
# ACESSO DOS ALUNOS POR GRUPO (PILOTO e REGULARES)
sns.stripplot(x='tag', y='log_viewed', data = dados_apr_rpr.loc[(dados_apr_rpr['gradefi
nal']>=6)], jitter = True)
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c9ce50f860>
```



9 REFERÊNCIAS BIBLIOGRÁFICAS

ALEXANDRON, Giora et al. Evaluating the robustness of learning analytics results against fake learners. In: **European Conference on Technology Enhanced Learning**. Springer, Cham, 2018. P. 74-87.

ALONSO-RAMOS, Mercedes et al. Innovative social approach in the nuclear sector: a MOOC on Nuclear Safety Culture within H2020 ANNETTE project. In: **2018 Learning With MOOCs (LWMOOCs)**. IEEE, 2018. P. 55-58.

ATAPATTU, Thushari; FALKNER, Katrina. A framework for topic generation and labeling from MOOC discussions. In: **Proceedings of the Third (2016) ACM Conference on Learning@ Scale**. 2016. P. 201-204.

BECKER, Samantha Adams et al. **Horizon report 2018 higher education edition brought to you by educause**. EDUCAUSE, 2018.

BESERRA, Clébia Alves et al. Aplicação de Técnicas de Aprendizagem de Máquina em Objetos de Aprendizagem baseado em Software: um Mapeamento Sistemático a partir das Publicações do SBIE. **RENOTE**, v. 12, n. 1, 2014.

BIENKOWSKI, Marie; FENG, Mingyu; MEANS, Barbara. Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. **Office of Educational Technology, US Department of Education**, 2012.

BOVO, Angela et al. Clustering moodle data as a tool for profiling students. In: **2013 Second international conference on E-Learning and E-Technologies in education (ICEEE)**. IEEE, 2013. P. 121-126.

BURKOV, Andriy. **The hundred-page machine learning book**. Quebec City, Can.: Andriy Burkov, 2019.

BYSTROVA, Tatiana et al. Learning analytics in massive open online courses as a tool for predicting learner performance. **Вопросы образования**, n. 4 (eng), 2018.

CHARLTON, Patricia; MAVRIKIS, Manolis; KATSIFLI, Demetra. The potential of learning analytics and big data. **Ariadne**, n. 71, 2013.

CHATTI, Mohamed Amine et al. A reference model for learning analytics. **International Journal of Technology Enhanced Learning**, v. 4, n. 5-6, p. 318-331, 2012.

CHEN, Qing et al. Viseq: Visual analytics of learning sequence in massive open online courses. **IEEE transactions on visualization and computer graphics**, v. 26, n. 3, p. 1622-1636, 2018.

CHOUNTA, Irene-Angelica; HECKING, Tobias; HOPPE, H. Ulrich. Every answer has a question: Exploring communication and knowledge exchange in MOOCs through learning

analytics. In: **CYTED-RITOS International Workshop on Groupware**. Springer, Cham, 2015. P. 140-147.

CLOW, Doug. MOOCs and the funnel of participation. In: **Proceedings of the third international conference on learning analytics and knowledge**. 2013. P. 185-189.

COBOS, Ruth et al. Open-dlas: An open dashboard for learning analytics. In: **Proceedings of the third (2016) ACM conference on learning@ scale**. 2016. P. 265-268.

COFFRIN, Carleton et al. Visualizing patterns of student engagement and performance in MOOCs. In: **Proceedings of the fourth international conference on learning analytics and knowledge**. 2014. P. 83-92.

COHEN, Anat et al. Active learners' characterization in MOOC forums and their generated knowledge. **British journal of educational technology**, v. 50, n. 1, p. 177-198, 2019.

CONFORTO, E., AMARAL, D., & SILVA, S. Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos. 8º Congresso Brasileiro de Gestão de Desenvolvimento de Produto. Porto Alegre. Setembro de 2011.

CONIJN, Rianne; VAN DEN BEEMT, Antoine; CUIJPERS, P. Predicting student performance in a blended MOOC. **Journal of Computer Assisted Learning**, v. 34, n. 5, p. 615-628, 2018.

COOK, Diane J.; DAS, Sajal K. Pervasive computing at scale: Transforming the state of the art. **Pervasive and Mobile Computing**, v. 8, n. 1, p. 22-35, 2012.

CORDEIRO, Alexander Magno et al. Revisão sistemática: uma revisão narrativa. **Revista do Colégio Brasileiro de Cirurgiões**, v. 34, p. 428-431, 2007.

D'ACQUISTO, Giuseppe et al. Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. **Arxiv preprint arxiv:1512.06000**, 2015.

DANIEL, Ben. Big Data and analytics in higher education: Opportunities and challenges. **British journal of educational technology**, v. 46, n. 5, p. 904-920, 2015.

DELGADO, A. A. S. Framework para caracterizar la innovación social sobre sus procesos. Orientador: João Bosco da Mota Alves. 2016. 245 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.

DELOITTE CONSULTING, L. L. P.; BY DELOITTE, Bersin. Global human capital trends 2014: Engaging the 21st-century workforce. 2014.

Digital Preservation Handbook, 2nd Edition, <https://www.dpconline.org/handbook>, Digital Preservation Coalition © 2015.

DRACHSLER, Hendrik; KALZ, Marco. The MOOC and learning analytics innovation cycle (MOLAC): A reflective summary of ongoing research and its challenges. **Journal of Computer Assisted Learning**, v. 32, n. 3, p. 281-290, 2016.

DRLJAČA, Dalibor et al. Addie model for development of e-courses. In: **Documento procedente de la International Scientific Conference on Information Technology and Data Related Research SINTEZA [Internet]**. 2017. P. 242-247.

EL MAWAS, Nour et al. Towards personalized content in massive open online courses. In: **10th International Conference on Computer Supported Education**. SCITEPRESS-Science and Technology Publications, 2018.

ELIAS, Tanya. Learning analytics. **Learning**, p. 1-22, 2011.

FERGUSON, Rebecca. Learning analytics: drivers, developments and challenges. **International Journal of Technology Enhanced Learning**, v. 4, n. 5-6, p. 304-317, 2012.

ERADZE, Maka; TAMMETS, Kairit. Learning analytics in MOOCs: EMMA case. In: **Data Science and Social Research**. Springer, Cham, 2017. P. 193-204.

EYNON, Rebecca. **The rise of Big Data: what does it mean for education, technology, and media research?**. 2013.

EZEN-CAN, Aysu et al. Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach. In: **Proceedings of the fifth international conference on learning analytics and knowledge**. 2015. P. 146-150.

FAYAD, Mohamed; SCHMIDT, Douglas C. Object-oriented application frameworks. **Communications of the ACM**, v. 40, n. 10, p. 32-38, 1997.

FERGUSON, Rebecca; CLOW, Doug. Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs). In: **Proceedings of the fifth international conference on learning analytics and knowledge**. 2015. P. 51-58.

FOURNIER, Hélène; KOP, Rita; SITLIA, Hanan. The value of learning analytics to networked learning on a personal learning environment. In: **Proceedings of the 1st international conference on learning analytics and knowledge**. 2011. P. 104-109.

FRICKÉ, Martin. Big data and its epistemology. **Journal of the Association for Information Science and Technology**, v. 66, n. 4, p. 651-661, 2015.

GARDNER, Joshua; BROOKS, Christopher. Dropout model evaluation in MOOCs. In: **Thirty-Second AAAI Conference on Artificial Intelligence**. 2018.

GARRISON, D. Randy. **Thinking collaboratively: Learning in a community of inquiry**. Routledge, 2015.

GAŠEVIĆ, Dragan et al. SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. **Computers in Human Behavior**, v. 92, p. 562-577, 2019.

GIBSON, David C.; IFENTHALER, Dirk. Preparing the next generation of education researchers for big data in higher education. In: **Big data and learning analytics in higher education**. Springer, Cham, 2017. P. 29-42.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. Ed. São Paulo: Atlas, 2002.

GREER, Jim; MARK, Mary. Evaluation methods for intelligent tutoring systems revisited. **International Journal of Artificial Intelligence in Education**, v. 26, n. 1, p. 387-392, 2016.

GUPTA, Shivangi; SABITHA, A. Sai. Deciphering the attributes of student retention in massive open online courses using data mining techniques. **Education and Information Technologies**, v. 24, n. 3, p. 1973-1994, 2019.

HANZAKI, Mahdi Rahmani; EPP, Carrie Demmans. The effect of personality and course attributes on academic performance in MOOCs. In: **European conference on technology enhanced learning**. Springer, Cham, 2018. P. 497-509.

ISIDRO, Cristina; CARRO, Rosa M.; ORTIGOSA, Alvaro. Dropout detection in MOOCs: An exploratory analysis. In: **2018 International Symposium on Computers in Education (SIEE)**. IEEE, 2018. P. 1-6.

ITANI, Alya; BRISSON, Laurent; GARLATTI, Serge. Understanding learner's drop-out in MOOCs. In: **international conference on intelligent data engineering and automated learning**. Springer, Cham, 2018. P. 233-244.

KHALIL, Mohammad; EBNER, Martin. Can learning analytics find success in didactical measurements? Results from a MOOC case study. In: **Digital Workplace Learning**. Springer, Cham, 2018. P. 211-225.

KHALIL, Mohammad; TARAGHI, Behnam; EBNER, Martin. Engaging Learning Analytics in MOOCs: the good, the bad, and the ugly. **Arxiv preprint arxiv:1606.03776**, 2016.

KHRIBI, Mohamed Koutheaïr; JEMNI, Mohamed; NASRAOUI, Olfa. Recommendation systems for personalized technology-enhanced learning. In: **Ubiquitous learning environments and technologies**. Springer, Berlin, Heidelberg, 2015. P. 159-180.

KITCHIN, Rob. Big Data, new epistemologies and paradigm shifts. **Big data & society**, v. 1, n. 1, p. 2053951714528481, 2014.

KLÜSENER, Marcus; FORTENBACHER, Albrecht. Predicting students' success based on forum activities in MOOCs. In: **2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)**. IEEE, 2015. P. 925-928.

KOLB, David A. Experience as the source of learning and development. **Upper Sadle River: Prentice Hall**, 1984.

LANG, Charles et al. (Ed.). **Handbook of learning analytics**. New York, NY, USA: SOLAR, Society for Learning Analytics and Research, 2017.

LAU, KH Vincent et al. Using learning analytics to evaluate a video-based lecture series. **Medical teacher**, v. 40, n. 1, p. 91-98, 2018.

Leek, J. Six types of analyses every data science should know. **Data scientist Insights blog**. Acessado em 2 dezembro de 2019, <http://datascientistinsights.com/2013/01/29/six-types-of-analyses-everydata-scientistshould-know/>. 2013.

LEÓN-URRUTIA, Manuel; TANG, Darron. The University of Southampton MOOC Observatory Dashboard. 2017.

LIANG, Jiajun et al. Big data application in education: dropout prediction in edx MOOCs. In: **2016 IEEE Second International Conference on Multimedia Big Data (bigmm)**. IEEE, 2016. P. 440-443.

LIEBOWITZ, Jay (Ed.). **Business analytics: An introduction**. CRC Press, 2013.

LIU, Sanya et al. Mining individual learning topics in course reviews based on author topic model. **International Journal of Distance Education Technologies (IJDET)**, v. 15, n. 3, p. 1-14, 2017.

LOPEZ, Glenn et al. Google bigquery for education: Framework for parsing and analyzing edx MOOC data. In: **Proceedings of the fourth (2017) ACM conference on learning@ scale**. 2017. P. 181-184.

LOVE, Nancy. **Data literacy for teachers**. Dude Publishing, an impint of National Professional Resources, Incorporated, 2013.

MANDINACH, Ellen B. A perfect time for data use: Using data-driven decision making to inform practice. **Educational Psychologist**, v. 47, n. 2, p. 71-85, 2012.

MANDINACH, Ellen B.; GUMMER, Edith S. A systemic view of implementing data literacy in educator preparation. **Educational Researcher**, v. 42, n. 1, p. 30-37, 2013.

MANYIKA, James et al. **Big data: The next frontier for innovation, competition, and productivity**. Mckinsey Global Institute, 2011.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big data: A revolution that will transform how we live, work, and think**. Houghton Mifflin Harcourt, 2013.

MCINTOSH, Naomi E. Barriers to implementing research in higher education. **Studies in Higher Education**, v. 4, n. 1, p. 77-86, 1979.

MCLAREN, Julie; DONALDSON, Jayne; SMITH, Stephen. Learning Analytics Suggest A Positive Experience. In: **17th European Conference on e-learning**. Academic Conference and Publishing International, 2018. P. 670-678.

MILLIGAN, Sandra. Crowd-sourced learning in MOOCs: learning analytics meets measurement theory. In: **Proceedings of the Fifth International Conference on Learning Analytics And Knowledge**. 2015. P. 151-155.

MORENO-MARCOS, Pedro Manuel et al. Sentiment analysis in MOOCs: A case study. In: **2018 IEEE Global Engineering Education Conference (EDUCON)**. IEEE, 2018. P. 1489-1496.

MUKALA, Patrick et al. Learning analytics on coursera event data: A process mining approach. In: **SIMPDA**. 2015. P. 18-32.

MUÑOZ-MERINO, Pedro J. Et al. Assessment activities in massive open on-line courses: Assessment activities in MOOCs. In: **Furthering higher education possibilities through massive open online courses**. IGI Global, 2015. P. 165-192.

O'RIORDAN, Tim; MILLARD, David E.; SCHULZ, John B. How should we measure online learning activity?. **Research in Learning Technology**, v. 24, p. 1-28, 2016.

PACHECO, RCDS. Dados e governo abertos na sociedade do conhecimento. **Florianópolis, SC: Linked Open Data-Brasil**, 2014.

PAPAMITSIOU, Zacharoula K.; ECONOMIDES, Anastasios A. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. **J. Educ. Technol. Soc.**, v. 17, n. 4, p. 49-64, 2014.

PARDOS, Zachary A.; KAO, Kevin. MOOCrp: An open-source analytics platform. In: **Proceedings of the Second (2015) ACM conference on learning@ scale**. 2015. P. 103-110.

PARFENOV, Denis; ZAPOROZHKO, Veronika. Developing SMART educational cloud environment on the basis of adaptive massive open online courses. In: **Conference Internationalization of Education in Applied Mathematics and Informatics for hightech Applications**. 2018. P. 35-41.

PERNA, L., RUBY, A., BORUCH, R., WANG, N., SCULL, J., EVANS, C. & AHMAD, S. (2013). **The life cycle of a million MOOC users**. Acessado em março de 2021. [Http://www.gse.upenn.edu/pdf/ahead/perna_ruby_boruch_MOOCs_dec2013.pdf](http://www.gse.upenn.edu/pdf/ahead/perna_ruby_boruch_MOOCs_dec2013.pdf)

PEÑA-AYALA, Alejandro. Learning Analytics: fundamentos, aplicaciones, and trends. **A view of the current state of the art to enhance e-learning**, 2017.

PRINSLOO, Paul; SLADE, Sharon. Student privacy self-management: implications for learning analytics. In: **Proceedings of the fifth international conference on learning analytics and knowledge**. 2015. P. 83-92.

RICE, William. **Moodle e-learning course development**. Packt Publishing Ltd, 2015.

RIDSDALE, Chantel et al. **Strategies and best practices for data literacy education: Knowledge synthesis report**. 2015.

Robbins, S. **Learning Analytics and the Value of Understanding L&D Metrics**. 2017

ROHLOFF, Tobias et al. Towards a better understanding of mobile learning in MOOCs. In: **2018 Learning With MOOCS (LWMOOCS)**. IEEE, 2018. P. 1-4.

ROMERO, Cristobal et al. Web usage mining for predicting final marks of students that use Moodle courses. **Computer Applications in Engineering Education**, v. 21, n. 1, p. 135-146, 2013.

ROMERO, Cristóbal; ROMERO, José Raúl; VENTURA, Sebastián. A survey on pre-processing educational data. In: **Educational data mining**. Springer, Cham, 2014. P. 29-64.

ROTHER, E. T. Revisão sistemática x revisão narrativa. **Acta paul. Enferm.**, São Paulo, v. 20, n. 2, jun. 2007

RUIPÉREZ-VALIENTE, José A. Et al. Scaling to massiveness with analyse: A learning analytics tool for open edx. **IEEE Transactions on Human-Machine Systems**, v. 47, n. 6, p. 909-914, 2016.

SCAPIN, Rafael. **Learning Analytics in Education: Using Student's Big Data to improve Teaching**. 2015.

SHEHABUDDEEN, N.; PROBERT, D.; PHAAL, R.; PLATTS, K. Management representations and approaches: exploring issues surrounding frameworks. In: **BRITISH ACADEMY OF MANAGEMENT CONFERENCE**, 2000, Edimburgo. Proceedings [...]. Londres: British Academy of Management, 2000.

SHRIDHARAN, Madhumitha et al. Predictive learning analytics for video-watching behavior in MOOCs. In: **2018 52nd Annual Conference on Information Sciences and Systems (CISS)**. IEEE, 2018. P. 1-6.

FERGUSON, Rebecca; SHUM, Simon Buckingham. Learning analytics to identify exploratory dialogue within synchronous text chat. In: **Proceedings of the 1st International Conference on Learning Analytics and Knowledge**. 2011. P. 99-103.

SINCLAIR, Jane; KALVALA, Sara. Engagement measures in massive open online courses. In: **International Workshop on Learning Technology for Education in Cloud**. Springer, Cham, 2015. P. 3-15.

SINCLAIR, Jane; KALVALA, Sara. Student engagement in massive open online courses. **International Journal of Learning Technology**, v. 11, n. 3, p. 218-237, 2016.

SNIJDERS, Chris; MATZAT, Uwe; REIPS, Ulf-Dietrich. " Big Data": big gaps of knowledge in the field of internet science. **International journal of internet science**, v. 7, n. 1, p. 1-5, 2012.

SOLTANPOOR, Reza; SELLIS, Timos. Prescriptive analytics for big data. In: **Australasian database conference**. Springer, Cham, 2016. P. 245-256.

SRILEKSHMI, M. Et al. Learning analytics to identify students at-risk in MOOCs. In: **2016 IEEE Eighth International Conference on Technology for Education (T4E)**. IEEE, 2016. P. 194-199.

SUNAR, Ayse Saliha et al. Modelling MOOC learners' social behaviours. **Computers in Human Behavior**, v. 107, p. 105835, 2020.

TAN, Yueying et al. Learning profiles, behaviors and outcomes: Investigating international students' learning experience in an English MOOC. In: **2018 International Symposium on Educational Technology (ISET)**. IEEE, 2018. P. 214-218.

TESSER, C. D. Et al. Capacitação em auriculoterapia. **Revista Revise**, v.05,fluxo contínuo (2020): Dossiê experiências de integração ensino serviço nas Práticas Integrativas e Complementares, p.1-18. 2020.

THANKACHAN, Karun. Adaptive Intelligence for Massive Online Open Course. In: **2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)**. IEEE, 2018. P. 1-7.

TRANFIELD, David; DENYER, David; SMART, Palminder. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. **British journal of management**, v. 14, n. 3, p. 207-222, 2003.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 2010.

TSENG, Shu-Fen et al. Who will pass? Analyzing learner behaviors in MOOCs. **Research and Practice in Technology Enhanced Learning**, v. 11, n. 1, p. 1-11, 2016.

UFSC – Universidade Federal de Santa Catarina. Biblioteca Universitária. Programa de Capacitação do usuário. **Protocolo para busca sistemática da literatura**. Florianópolis. 2012

VAN DEN BEEMT, Antoine; BUIJS, Joos; VAN DER AALST, Wil. Analysing structured learning behaviour in massive open online courses (MOOCs): an approach based on process mining and clustering. **International Review of Research in Open and Distributed Learning**, v. 19, n. 5, 2018.

VITIELLO, Massimo et al. MOOC learner behaviour: Attrition and retention analysis and prediction based on 11 courses on the TELESCOPE platform. In: **International Workshop on Learning Technology for Education in Cloud**. Springer, Cham, 2017. P. 99-109.

WACHTLER, Josef et al. On Using Learning Analytics to Track the Activity of Interactive MOOC Videos. In: **SE@ VBL@ LAK**. 2016. P. 8-17.

WALLER, Matthew A.; FAWCETT, Stanley E. **Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management**. 2013.

WANG, Yuan et al. Workshop on integrated learning analytics of MOOC post-course development. In: **Proceedings of the Seventh International Learning Analytics & Knowledge Conference**. 2017. P. 506-507.

WARD, Jonathan Stuart; BARKER, Adam. Undefined by data: a survey of big data definitions. **Arxiv preprint arxiv:1309.5821**, 2013.

WILD, Fridolin. Erratum to: Learning Analytics in R with SNA, LSA, and MPIA. In: **Learning Analytics in R with SNA, LSA, and MPIA**. Springer, Cham, 2016. P. E1-E1.

WU, Yinan; WU, Wenjun. A learning analytics system for cognition analysis in online learning community. In: **Asia-Pacific Web (apweb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data**. Springer, Cham, 2018. P. 243-258.

XU, Songer; LUO, Heng; TAN, Yueying. Re-examining the community of inquiry framework from the perspective of learning analytics. In: **2018 13th International Conference on Computer Science & Education (ICCSE)**. IEEE, 2018. P. 1-5.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. Ed. Porto Alegre: Bookman, 2010.

YUAN, Li; POWELL, S. J.; OLIVIER, Bill. **Beyond MOOCs: Sustainable online learning in institutions**. 2014.

ZHENG, Yanyan; YIN, Beibei. Big data analytics in MOOCs. In: **2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing**. IEEE, 2015. P. 681-686.