

**UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de Marília**

FELIPE STEFANI CORREIA DOS SANTOS

**Ferramenta de suporte à avaliação de performance e tomada de decisão no
processo de ensino-aprendizagem de EAD**

Felipe Stefani Correia dos Santos

Ferramenta de suporte à avaliação de performance e tomada de decisão no processo de ensino-aprendizagem de EAD

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação como parte dos requisitos para a obtenção do título de mestre em Ciência da Informação pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Campus de Marília.

Área de Concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Leonardo Castro Botega

Financiadora: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

Marília
2023

S237f

Santos, Felipe Stefani Correia dos

Ferramenta de suporte à avaliação de performance e tomada de decisão no processo de ensino-aprendizagem de EAD / Felipe Stefani Correia dos Santos. -- Marília, 2023

113 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília

Orientador: Leonardo Castro Botega

1. Moodle. 2. Aprendizado de máquina. 3. LMS. 4. Processamento da informação. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Felipe Stefani Correia dos Santos

**Ferramenta de suporte à avaliação de performance e tomada de decisão no
processo de ensino-aprendizagem de EAD**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação como parte dos requisitos para a obtenção do título de mestre em Ciência da Informação pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Campus de Marília.

Área de concentração: Informação, Tecnologia e Conhecimento
Linha de pesquisa: Informação e Tecnologia

Banca Examinadora

Prof. Dr. Leonardo Castro Botega
UNESP – Câmpus de Marília
Orientador

Prof. Dr. Caio Saraiva Coneglian
UNESP – Câmpus de Marília

Prof. Dr. Ricardo Araújo Rios
UFBA

Marília, 01 de Março de 2023.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

O contexto do atual modelo educacional se apresenta em constante mudança devido a necessidade de adaptação ao novo cenário super conectado dos dias de hoje. Torna-se um desafio para os professores de Educação a Distância (EAD) se adaptar aos Learning Management System (LMS) e desenvolver novas técnicas de análise e avaliação do processo de ensino-aprendizagem. Este trabalho aborda o LMS Moodle, o qual gera inúmeros registros de atividades realizadas no sistema chamados de LOGs. Porém, a falta de acompanhamento desses registros no ambiente, mais especificamente quanto às interações do estudante com a plataforma, pode produzir um déficit informacional quando se trata de análise e avaliação de performance dos mesmos. Essa falta de acompanhamento pode ser justificada pela complexidade e volume de dados gerados pela plataforma. Desta maneira, este trabalho apresenta um modelo de processamento da informação, denominado Performi (Performance estudantil dirigida por processamento informacional), que tem como objetivo guiar a criação de uma ferramenta que forneça subsídios informacionais de cunho quali-quantitativo, referente à performance e interações de estudantes no LMS Moodle, utilizando do arcabouço teórico da Ciência da Informação, para estabelecer processos, etapas ou fluxos sistematizados e estruturados. Para tal, são utilizadas técnicas de aprendizado de máquina como instrumento de processamento de dados, possibilitando a profissionais de educação o usufruto de informações enriquecidas. A natureza deste trabalho é quali-quantitativa, de finalidade teórico-aplicada e de tipo exploratória, sendo sua metodologia uma pesquisa exploratória com prova de conceito. Para validar o modelo, foi desenvolvida uma prova de conceito guiada pelo modelo Performi. Os resultados indicam que o modelo extrai dados quali-quantitativos dos registros de LOG e os transforma em informações contextualizadas, além de organizar, descrever e guiar a interpretação da informação, gerando subsídios informacionais para tomada de decisão.

Palavras-chave: Moodle, Aprendizado de máquina, LMS, Processamento da informação

ABSTRACT

The current educational model context is constantly changing due to the need to adapt to today's new, super-connected scenario. It becomes a challenge for distance education teachers to adapt to the Learning Management System (LMS) and develop new techniques for analysis and evaluation of the teaching-learning process. This paper discusses the LMS Moodle, which generates numerous records of activities performed in the system called LOGs. However, the lack of monitoring of these records in the environment, more specifically regarding student interactions with the platform, can produce an informational deficit when it comes to analysis and evaluation of their performance. This lack of follow-up can be justified by the complexity and volume of data generated by the platform. Therefore, this paper presents an information processing model, called Performi (Student Performance Driven by Informational Processing), which aims to guide the creation of a tool that provides quali-quantitative informational support regarding the performance and interactions of students on the LMS Moodle, using the theoretical foundation of Information Science, to establish systematized and structured processes, steps, or flows. To do this, machine learning techniques are used as a data processing tool, allowing education professionals to enjoy enriched information. The nature of this work is quali-quantitative, of theoretical-applied purpose and of exploratory type, being its methodology exploratory research with proof of concept. To validate the model, a proof of concept guided by the Performi model was developed. The results indicate that the model extracts qualitative-quantitative data from LOG records and transforms them into contextualized information, besides organizing, describing, and guiding the interpretation of information, generating informational subsidies for decision making.

Keywords: Moodle, Machine learning, LMS, Information processing

LISTA DE FIGURAS

Figura 1 – Fluxo metodológico	29
Figura 2 – Organização metodológica.....	30
Figura 3 – Preparação para coleta de dados	32
Figura 4 – Dificuldades dos docentes.....	40
Figura 5 – Satisfação com o ambiente.....	41
Figura 6 – Uso de ferramentas analíticas	41
Figura 7 – Uso de dados educacionais	42
Figura 8 – Dados mais importantes	42
Figura 9 – Uso de registros de Log	43
Figura 10 – Motivo do não uso dos Logs	43
Figura 11 – Lista de eventos de um curso Moodle.....	49
Figura 12 – Estrutura dos dados	50
Figura 13 – Gráfico de frequência de arquivos enviados	51
Figura 14 – Correlação de dados entre eventos.....	52
Figura 15 – Correlação de dados entre eventos mais relevantes	53
Figura 16 – Aprendizagem de máquina no modelo performi.....	61
Figura 17 – Cross-Validation	63
Figura 18 – Camada de apresentação	65
Figura 19 – Diagrama completo de funcionamento do modelo Performi	69
Figura 20 – Diagrama de produção de dados	72
Figura 21 – Logs do ambiente Moodle.....	74
Figura 22 – Diagrama de processamento de dados	76
Figura 23 – Dados orientados ao registro	77
Figura 24 – Dados orientados ao usuário	78
Figura 25 – Diagrama de análises descritivas	79
Figura 26 – Diagrama de análises preditivas.....	81
Figura 27 – Exemplificação de modelo preditivo	82
Figura 28 – Exemplificação de informação servida de forma direta.....	84
Figura 29 – Exemplificação de informação servida de forma indireta	84
Figura 30 – Organização da informação: Estrutura de dicionário.....	86
Figura 31 – Dashboard de exemplo de apresentação da informação	88
Figura 32 – Uso dos dados do modelo	89
Figura 33 – Dados de LOG	92
Figura 34 – Notas finais	92
Figura 35 – Base de dados direcionada ao aluno	93
Figura 36 – Relevância das features	96
Figura 37 – Tradução da resposta do algoritmo	97
Figura 38 – Dashboard de análise de performance	98
Figura 39 – Comparação entre aluno e histórico	100

LISTA DE QUADROS

Quadro 1 – Feature e relevância.....83

LISTA DE TABELAS

Tabela 1 – Comparação de desempenho dos algoritmos.....	95
--	----

LISTA DE ABREVIATURAS E SIGLAS

LMS	Learning Management System
EaD	Educação à Distância
IES	Instituição de Ensino Superior
RL	Revisão de Literatura
Moodle	Modular Object Oriented Distance Learning

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Problemática.....	16
1.2 Objetivos	16
1.2.1 Objetivos Específicos	17
1.3 Trabalhos Relacionados	17
2 METODOLOGIA.....	26
2.1 Natureza, Abordagem e Classificação.....	26
2.2 Método de Pesquisa.....	27
2.2.1 Proposta Metodológica.....	28
2.3 Organização Metodológica.....	30
3 O PROCESSO DE AVALIAÇÃO DE PERFORMANCE EM SISTEMAS DE GESTÃO DA APRENDIZAGEM	34
3.1 Processo de Tomada de Decisão Pedagógica	34
3.2 Uso de Ferramentas Para Auxiliar a Tomada de Decisão em Ambiente Educacional	36
3.3 Avaliação de Performance do Estudante em Ambiente Lms.....	38
3.4 Levantamento e Discussão Sobre o Uso de Dados em LMS Para Tomada de Decisão Pedagógica	39
4 PROCESSAMENTO DA INFORMAÇÃO NO AMBIENTE LMS MOODLE	45
4.1 Fontes de Dados.....	45
4.2 Estruturação dos Dados	48
4.3 Tomada de Decisão Docente Dirigida por Dados.....	53
4.4 Transformação dos Dados em Informação.....	56
4.5 Aprendizado de Máquina Como Ferramenta de Análise Preditiva no Contexto da Educação a Distância	57
4.6 Artefatos Informacionais e Camada de Visualização no Contexto Educacional	63
5 MODELO DE PROCESSAMENTO DA INFORMAÇÃO PARA AVALIAÇÃO DE PERFORMANCE DE DISCENTES EM AMBIENTE LMS.....	67
5.1 Introdução ao Modelo Performí	67
5.1.1 Processo Completo de Avaliação de Performance Estudantil	70
5.1.2 Operação Cíclica e Auditável.....	71
5.2 Processos Internos do Modelo Performí	72
5.2.1 Produção dos Dados	72

5.2.2 Processamento de Dados	76
5.2.3 Enriquecimento da Informação	78
5.2.4 Organização da Informação	85
5.2.5 Apresentação da Informação	87
5.2.6 Tomada de Decisão	88
6 PROVA DE CONCEITO.....	91
6.1 Processo e Resultado da Prova de Conceito	91
6.2 Artefatos Informacionais.....	100
6.3 Discussão.....	101
7 CONCLUSÕES.....	103
7.1 Atendimento aos Objetivos de Pesquisa	103
7.2 Limitações da Pesquisa.....	106
7.3 Contribuições da Pesquisa	107
7.4 Trabalhos Futuros	107
REFERÊNCIAS.....	109

1 INTRODUÇÃO

Atualmente a internet tem guiado o modelo educacional a mudanças e necessidades de adaptação a novas realidades, instituições educacionais e seus gestores precisam sempre estar atualizados e informados, esta tendência tem favorecido a utilização de recursos tecnológicos, tanto em cursos presenciais, como em cursos totalmente a distância (DIAS JR, 2017). Neste contexto, a falta de informações e recursos reflete diretamente em questões estruturais e gerenciais de organizações públicas e privadas, exigindo dos profissionais uma proatividade em suas tomadas de decisões nos ambientes de gerenciamento.

Muitas dessas mudanças são devidas aos avanços nas Tecnologias da Informação e Comunicação (TICs), abrangendo um vasto conjunto de ferramentas destinadas à sistematização, tratamento, organização, acesso e disseminação da informação, modificando a forma com que a sociedade, indivíduos e organizações tratam a informação, assim desbravando novas formas de acesso e aquisição de conhecimento. Le Coadic (1996) associa as TICs aos fenômenos de explosão da quantidade de informação.

Com a exposição ao grande volume de informação e a informação assumindo papéis cada vez mais importantes nas organizações, a Ciência da Informação tendo a informação como objeto de estudo se posiciona como uma área de estudos preocupada com os princípios e técnicas de processamento da informação e sua aplicabilidade. Além disso, a Ciência da Informação, aliada à tecnologia da informação, busca estabelecer processos, etapas ou fluxos sistematizados e estruturados a fim de se obter os resultados almejados e gerar novas informações utilizando como intermediador sistemas de informação, algoritmos, ferramentas e processos computacionais.

Definir um fluxo informacional é um importante passo para manter a gestão do processo educacional organizada e otimizada, pois a utilização desta ferramenta possibilita estabelecer etapas de obtenção, tratamento, armazenamento, distribuição, disseminação e uso da informação, no contexto informacional. Os autores Oliveira e Bertucci (2005, p 8) destacam "o gerenciamento da informação tornou-se um instrumento estratégico necessário para controlar e auxiliar decisões, através de melhorias no fluxo da informação, do controle, análise e consolidação da informação para os usuários", essa melhoria visa, neste trabalho, utilizando como intermédio a tecnologia, entregar aos gestores subsídios informacionais de qualidade, processados com auxílio computacional, para auxiliar no processo de tomada de decisão, elucidando que as organizações que com eficiência administrarem suas informações, terão um recurso estratégico fundamental para maximizar a qualidade do processo decisório.

Em complemento, torna-se notável e explícita a importância das TICs na atual sociedade, as transformações não ocorrem apenas em decorrência da utilização e popularização de novas tecnologias, mas também, em decorrência de mudanças sociais, econômicas e culturais. Neste contexto torna-se imprescindível a busca por conhecimentos e qualificação das pessoas. SILVA, ANDRADE e BARTOLOMEI SILVA (2011) ressaltam que a utilização das tecnologias, está cada vez mais presentes no cotidiano, criando relações, novos conhecimentos e novas maneiras de aprender e de pensar.

A partir do supracitado e com o desenvolvimento de novas tecnologias e alinhada às necessidades da sociedade, as instituições de ensino começaram a utilizar meios de comunicação para melhorar o processo de ensino-aprendizagem. A partir disso, surgem modalidades de educação a distância, iniciando em 1910 com cursos de correspondência, passando por outras tecnologias, como as divulgações de materiais por rádio e telecursos nos canais televisivos culminando, a partir de 1990 com o advento da internet e dos computadores, na educação online, com o intuito de atender demandas educacionais antes não atendidas e novas necessidades geradas pela globalização e popularização dos computadores, sejam eles portáteis, móveis ou pessoais, promovendo também um processo de transformação do espaço e do tempo. (Belloni,2006)

No cerne desta nova modalidade, surgem em meados da década de 90 os Learning Management System (LMS) ou Sistema de Gestão da Aprendizagem, com o objetivo de apoiar as atividades de ensino e aprendizagem remoto, com acesso flexível e novas formas de comunicação e avaliação para professores e alunos, baseados na internet. O surgimento dessas novas tecnologias mudou a forma que os alunos aprendem, possibilitando aos professores disponibilizarem recursos educacionais sem a necessidade de desenvolver habilidades técnicas para isso.

Neste contexto e com a atual crescente de popularidade do ensino online e a amplificação da abordagem de ensino online propiciada pela pandemia causada pela SARS-CoV-2 (COVID-19), os ambientes de ensino online vem ganhando popularidade como uma forma das pessoas continuarem as suas graduações. Com o aumento do interesse por ensino online, a procura por LMS também cresceu e a necessidade das instituições e gestores por informações de qualidade nesses ambientes também, sendo assim necessária a geração de novas formas de geração de informação a partir desses ambientes.

No universo das instituições de ensino superior brasileiras, pautando-se apenas nos ambientes LMS, um dos grandes problemas enfrentados na gestão dos cursos de graduação é a evasão dos alunos. Uma vez que existem várias condições que propiciam a evasão, este

problema se torna cada vez mais complexo. Quando se trata de ambiente de Educação a Distância (EaD), características como localização geográfica, horário de aula e tempo de aula, são menos relevantes para avaliar a evasão de um aluno, devido ao material poder ser consultado a todo instante há qualquer hora do dia, otimizando o acesso ao acervo informacional e a produção cultural e intelectual.

Adicionalmente neste contexto existem vários LMS disponíveis para utilização, um dos mais populares e utilizados é o Modular Object Oriented Distance Learning (Moodle), que é uma ferramenta de código aberto e livre. Estes pontos incentivam que muitas instituições de ensino o utilizem, e permite que elas façam as personalizações necessárias para atender suas necessidades educacionais específicas.

De acordo com Pulino Filho, destacam-se algumas vantagens da utilização do Moodle:

"Muitos administradores de ambientes de aprendizagem têm declarado sua adesão ao Moodle principalmente em virtude de ser ele um sistema aberto, baseado em uma forte filosofia educacional, com uma comunidade de usuários crescente dia a dia que contribui para o desenvolvimento e apoio a novos usuários (PULINO FILHO, 2005, p. 5). "

No Moodle há diversas áreas para a apresentação de conteúdo, atividades de verificação da aprendizagem e espaços de interação síncrona e assíncrona, possibilitando a interação dos alunos entre si, com os professores e com o conteúdo. Esta plataforma armazena informações de praticamente todos os registros das interações e recursos utilizados pelos usuários que aconteceram na plataforma e permite gerar arquivos de registro de eventos relevantes chamados de logs onde é possível identificar data, hora e quais os recursos e atividades que foram acessadas pelo aluno, notas das avaliações sobre o desempenho em determinada disciplina.

Em decorrência da evolução dos recursos tecnológicos e a inserção de ferramentas tecnológicas para o auxílio da aprendizagem e da gestão do conhecimento, contribuindo para esta evolução, novos processos e técnicas de apoio à educação surgem, possibilitando aprendizagem mais assertivas, personalizadas e sob demanda, em especial em ambientes de educação online, que tem por objetivo contribuir com a interrelação entre todas as pessoas envolvidas (DIAS JR, 2017).

Contudo, apesar da evolução apresentada, ainda é observado algumas limitações nos ambientes de educação online, principalmente nas ferramentas e informações para auxílio à tomada de decisão. Estas limitações são pautadas em questionamentos geralmente relacionados a baixa interação humana e feedback, elevado índice de evasão dos alunos, além da carência de avaliações efetivas sobre a eficiência pedagógica (DIAS JR, 2017; MARGARYAN; BIANCO;

LITTLEJOHN, 2015; PAIVA, 2017).

De acordo com o Censo EAD 2019, as taxas de evasão ainda é um ponto preocupante mesmo apresentando uma redução, frente aos anos anteriores. Num contexto em que em 2018 havia 16750 cursos totalmente à distância, o volume de matrículas foi de 2.358.934 e taxa média de evasão de 62,2%, temos em torno de 1.467.256 de alunos evadidos e este se apresenta como um dos principais desafios da Educação a Distância.

Corroborando com o contexto supracitado Branco, Conte e Habowski (2020) comentam que a taxa média de evasão em cursos EAD é de 26,3% e que 85% dos estudantes evadem no início do curso. O autor destaca o fato de que em alguns cursos a taxa se eleva consideravelmente. Autores como Paiva (2017), Dias Jr (2017) e Branco, Conte e Habowski (2020) apontam motivos que levam os estudantes a não concluir os cursos em ambientes de educação online, sendo esses motivos: apontam: 1) o estudante não tem o interesse de concluir o curso; 2) o estudante não tem tempo para estudar; 3) o estudante não tem o conhecimento básico necessário; 4) falta de suporte dos professores.

Como elucidado, não obstante ao contexto, o autor Paiva (2017) comenta sobre a falta de apoio tecnológico apropriado a professores e tutores que possibilitem lidar com os dados oriundos de ambientes LMS, gerando um impedimento da integração da prática institucional com a tecnologia, dificultando o uso das tecnologias e a promoção da aprendizagem significativa e engajadora.

Devido a esta falta de apoio tecnológico apropriado agregada a grande quantidade de dados produzidos por ambientes LMS, muita informação não é gerada nesse ambiente, trazendo prejuízos tanto para os alunos quanto para os professores. Para evitar esse desperdício de de informação, softwares são utilizados para processar a informação, porém para Paiva (2017) não é recomendado que as tomadas de decisões sejam realizadas unicamente por software, elucidando essencialidade do uso de apoio computacional para fornecer subsídios à tomada de decisão em ambientes de educação online.

Posta esta contextualização e tendo em vista o atual cenário educacional, com o aumento da necessidade de utilização de ambiente remotos, a popularização dos cursos ofertados em ambientes de ensino online além da demanda por informações de melhor qualidade, justifica-se a importância em realizar um trabalho que possa enriquecer o fluxo informacional e dar suporte a tomada de decisões educacional no ensino superior, a fim de fornecer insumos informacionais para possibilitar ações preventivas que resultem na diminuição da evasão de alunos de um curso.

1.1 Problemática

A Ciência da informação é uma disciplina que investiga as propriedades e o comportamento da informação, como dito por H. Borko (1968) uma de suas linhas investigativas traz como força regente o estudo dos meios de processamento para a otimização do acesso e uso da informação, dentro do ambiente informacional deste trabalho esta força está relacionada a um corpo de conhecimento que se importa com a transformação, interpretação e utilização da informação, com enfoque em uma transformação que leva em consideração a necessidade de contexto da informação para uma melhor interpretação e utilização da informação.

A falta de acompanhamento dos registros de LOGs do ambiente LMS, no caso deste trabalho, o ambiente Moodle, é problemática pois estes dados permitem o acompanhamento das atividades dos alunos, fornecendo assim subsídios para estudos que relacionem a atividade exercida na plataforma com o seu desempenho, não acompanhar as atividades e não entender o comportamento dos estudantes na plataforma dificulta o processo de ensino-aprendizagem. Visto os números apresentados pelo CensoEAD 2019 quanto a evasão, corroborado pelos autores Branco, Conte e Habowksi (2020), além da grande quantidade de dados gravados no ambiente do Moodle, através da análise e processamento destes dados se torna possível gerar informações e análises de qualidade, relevantes e enriquecidas, com mais qualidade informacional e também características preditivas em relação ao aluno.

É de suma importância ressaltar o caráter quantitativo dos dados provenientes do ambiente Moodle, visto que, os dados disponíveis são espúrios, ou seja, espalhados, bagunçados e que não servem para mapear os problemas de performance e evasão que possam vir a emergir no LMS. Além disso, é impossível prever qualquer situação com o atual conjunto de dados disponível nos registros de atividades sem uma ferramenta intermediadora. Contudo, a tarefa de realizar uma análise nesse ambiente é necessária, e a informação gerada necessita de qualidade, porém é inviável de ser realizada manualmente, tornando necessária a intermediação tecnológica para auxiliar nesse processo e facilitar a coleta e análise dessa grande base de dados de forma ágil e confiável.

1.2 Objetivos

O objetivo deste trabalho é desenvolver um modelo de aplicação de aprendizagem de máquina em ambiente LMS para guiar a criação de ferramentas para os gestores educacionais

(professores e tutores), que forneça subsídios informacionais de cunho quali-quantitativo, referente à performance e interações de estudantes em ambiente LMS, utilizando o processamento de dados brutos, fornecidos pelos registros de log do ambiente LMS Moodle, com o auxílio computacional de algoritmos aprendizagem de máquina supervisionados, preditivos e interpretáveis, utilizando métodos de classificação, a fim de auxiliar a tomada de decisão pedagógica provendo dados enriquecidos.

1.2.1 Objetivos Específicos

Para atingir esse objetivo geral esta pesquisa a seguinte ordem de objetivos específicos foi estabelecida:

- Definir um processo informacional a partir do LMS Moodle. que descreva a estrutura dos dados, as transformações e manipulações que serão feitas e gerar uma informação contextualizada de uma análise preditiva;
- Desenvolver uma prova de conceito com o LMS Moodle para validação da informação produzida;

1.3 Trabalhos Relacionados

O ambiente Moodle dispõe, além da possibilidade de instalação e edição do ambiente para se adaptar às demandas particulares, da possibilidade de instalar plugins listados em uma biblioteca de plugins, que são compatíveis com o sistema e é alimentada pela comunidade apoiadora do Moodle, Dentre essas ferramentas existem várias ferramentas de todos os objetivos, sejam eles, integração com sistemas de videoconferência, armazenamento, análises entre outras funcionalidades para enriquecer o ambiente, este trabalho é focado em analisar somente os plugins que utilizam alguma forma de análise preditiva, que utiliza os logs como fonte de dados, além de trabalhos acadêmicos correlatos os quais realizam análises preditivas com dados educacionais de ambientes LMS.

Dentro do ambiente moodle a partir da versão 3.4 um dos plugins mais utilizados é o plugin inspire que traz, por meio de análise preditiva, se o estudante está ou não em risco de reprovação baseado no engajamento dos estudantes, analisando os logs e mostrando em frente ao nome do usuário uma mensagem informando se há ou não o risco de reprovação ou abandono do curso. Em complemento este plugin está a cinco anos sem nova versão, além de fornecer

somente a métrica classificatória, o plugin também mostra os dados que são levados em consideração na análise e sua completude, mas não utiliza os dados históricos para avaliação somente a relação entre o que o estudante fez em relação a quantidade de atividades emitidas.

Adicionalmente, outro plugin é o IADLearning, que visa conceder acesso ao conteúdo de aprendizagem adaptável criado, oferecendo ao aluno conteúdos e recomendações a partir de suas preferências e aos professores um conjunto de análises preditivas que visam permitir a ação proativa nos processos de aprendizagem em andamento. Em complemento este plugin também está a cinco anos sem nova versão, além de fornecer somente a métrica classificatória, mostrando muitos dados quantitativos e gráficos sobre os estudantes, mas não um dado que contextualiza o porquê deste estudante estar em risco de reprovação.

Outro plugin utilizado é o Students at risk of missing assignment due dates, cujo tem o objetivo específico de prever, por meio de um modelo de aprendizagem de máquina, os alunos que irão perder os prazos de entrega de suas tarefas. O modelo utiliza de dados sobre as atividades para avaliá-las e utilizando dados históricos dos logs como, atividade não iniciada, tarefas em curso não iniciada, tarefas ocultas para os alunos, entre outros. Em complemento o modelo traz uma boa visão sobre como a informação é gerada e como deve ser interpretada, mas está há três anos sem manutenção, entretanto é um modelo que por ser extremamente específico agrega bastante o processo de tomada de decisão quando se trata de tarefas.

Adicionalmente, outro plugin disponível é o Dropout Detective™ que visa também demonstrar por intermédio de algoritmo de aprendizagem de máquina os alunos que correm risco de abandonar ou reprovar nos cursos. Este plugin, ao contrário dos outros, tem atualizações mais recentes, sendo a última dezessete meses atrás e traz inúmeras informações quantitativas e descritivas, assim como os outros plugins, sobre os discentes, o mesmo disponibiliza meios de notificação automática o que facilita a comunicação da informação para o professor, mas não contextualiza como este risco é calculado, ou a partir de qual visão.

Em complemento a isso, o autor AVELLA, John T. et al (2016), em seu trabalho realiza uma revisão sistemática de literatura sobre os benefícios, desafios e métodos do learning analytics na educação superior. Neste estudo os autores abordam a Learning Analytics (LA) que é considerada parte da ciência dos dados educacionais e o seu propósito apresentado por Pinnell (2017) e Filatro (2019) é identificar o perfil dos alunos, coletar e qualificar o maior número possível de dados e informações referente suas interações com os ambientes de aprendizagem, possibilitando compreender o progresso destes alunos e sua performance, construir práticas pedagógicas satisfatórias, identificar situações de risco e avaliar fatores que afetam a conclusão e o sucesso dos estudos. Os autores ressaltam que a LA trata de percepções

sobre fatores como: a capacidade, o comportamento de aprendizagem, a previsibilidade das preocupações de aprendizagem, a nutrição de aspectos cognitivos, entre outros.

Os autores dividem os resultados em seções descritivas de métodos utilizados em Learning Analytics, os benefícios e por fim os desafios mais emergentes no cenário. Em sua seção de métodos lista técnicas como, clusterização de dados, visualização de dados, processamento de dados, algumas ferramentas disponíveis e utilizadas, entre outros métodos analíticos e estatísticos, ao fim desta seção o autor comenta também sobre descobertas através de modelos o que traz um caráter informacional ao trabalho por ressaltar não diretamente a importância e a diferença entre o dado de um resultado da aplicação de uma técnica e a informação contextualizada por um modelo contextualizado.

Em sua seção de benefícios o autor trabalha uma série de melhorias a qual a análise desse cenário de big data poderia trazer, como a identificação de cursos mais relacionados ao objetivo do estudante, a melhoria do currículo do estudante, a melhoria referente a performance do estudante, processo educacional e comportamento do discente perante aos seus estudos, ressalta a importância e o benefício do aprendizado personalizado, a melhoria não só do aluno mas sim dos instrutores também, por fim aborda fator empregabilidade pós formação,

Os desafios citados pelo autor são inúmeros, dentre eles o mapeamento de dados, a disponibilidade dos dados, o processo de avaliação e análise desses dados, a melhoria do ambiente de aprendizado, as emergentes tecnologias que podem aumentar o potencial do LA e por fim a ética e a privacidade dos dados e informações geradas que é uma discussão emergente é extremamente importante no universo de dados não só educacionais.

Dito isso o trabalho traz em sua discussão evidentes pontos positivos a favor do LA, porém traz indiretamente em suas preocupações a forma como todos esses dados serão manipulados para tornar-se informação, quem são esses profissionais os quais realizarão tal tarefa e como esta área tem abertura para inúmeras contribuições interdisciplinares não somente a partir da ótica dos benefícios do big data e do LA para a educação mas também da ótica das partes interessadas no processo educacional para descobrir novos caminhos neste processo.

KRISHNAN, Reshmy et al (2022). desenvolve um trabalho relacionado a análise de dados dos estudantes a partir da ótica das interações realizadas nas atividades da plataforma moodle, onde o trabalho apresenta uma visão de quantos trabalhos foram submetidos e quantos questionários foram submetidos a tempo, por meio de plugins adicionados a um específico curso na plataforma. A distribuição dos resultados é também fornecida em gráficos, a fim de facilitar o entendimento dos dados produzidos, assim com insights sobre as preferências de atividades dos discentes analisados, o número de download das ferramentas utilizadas nos

curso, estudantes ativos em específicos horários entre outros dados que contextualizam o ambiente o qual a turma se encontra.

Estas análises supracitadas trazem em si um caráter descritivo o qual tem o intuito de informar sobre a situação atual de forma ilustrativa e rápida, é uma análise a qual contextualiza o tomador de decisão referente ao estado atual da turma perante ao curso, trata-se de uma análise generalista da turma, um visão geral do comportamento da turma em relação às atividades propostas. Estas análises são uma ótima forma de contextualizar o ambiente, porém não abrange os discentes individualmente, somente o grupo, tornando difícil a identificação de situações que demandem maior atenção por parte dos professores e tutores, o que faria com que, após analisar tais dados, o tomador de decisão análise caso a caso para poder identificar pontos que necessitem de maior atenção.

SANTANA, Marcelo A.; DOS SANTOS NETO, Balduino Fonseca; DE BARROS COSTA, Evandro (2014) avaliam o desempenho, o uso e a eficácia das ferramentas educacionais no ambiente virtual de aprendizagem Moodle, para isso utilizam técnicas de mineração de dados educacionais e um questionário para os tutores online, ressaltando que toda a pesquisa foi realizada na Universidade Federal do Alagoas e que todos os dados utilizados são referentes somente a este escopo.

Em seu estudo os autores afirmam que as ferramentas do Moodle estão sendo subutilizadas quando se trata da questão de aproveitamento das mesmas para melhoria do desempenho dos discentes da Universidade Federal do Alagoas, onde deixa implícito os tutores por mais que conheçam e utilizem as ferramentas, seu aproveitamento da plataforma ainda é baixo e sempre que encontram formas de utilizar alguma ferramenta externa a qual está mais habituado a usar, vai usar, como no caso do uso de e-mail para comunicação mesmo a plataforma disponibilizando recurso de mensagem.

Complementar ao supracitado os autores concluem que necessitam de análises de outros escopos fora da universidade para melhor avaliação e que mesmo com a subutilização das ferramentas por parte dos tutores, há viabilidade de realizar inferências quanto ao uso de ferramentas disponíveis, podendo ser úteis para professores no sentido de auxílio no processo de avaliação e participação dos discentes a fim de motivá-los.

DA SILVA, Danilo Garcia; ALONSO, Kátia Morosov; MACIEL, Cristiano (2012) analisam por meio da trajetória de aprendizagem sob a ótica da participação e interação os relatórios de atividades do Moodle buscando entender a finalidade e a relevância dos recursos de gestão de aprendizagem empregados no âmbito pedagógico que permita contribuir com o professor quanto a visualização e levantamento de informações, avaliação e definição de

estratégias. Este trabalho realiza uma análise, entrevistando profissionais da educação que trabalham com o ambiente EAD, onde questionam sobre a utilização dos relatórios, presentes no ambiente, quanto a participação e interação dos alunos com as atividades.

Em meio aos resultados os pesquisadores afirmam que grande parte dos docentes, sabem da existência, sabem como adquirir os relatórios, mas o fazem de forma operacional, apenas para cumprir protocolo, o entendimento dos dados não reflete o mesmo percentual trazendo a tona o problema com a forma de representação dos dados e ressaltando a diferença entre dados e informação. O texto também ressalta a importância da utilização de dados como esses em práticas pedagógicas, pois tais dados podem direcionar os docentes para melhores entendimentos, melhores decisões, descobertas informacionais e até novas direções pedagógicas, adotando ações e direções que abordam dados como subsídio e guiam a trajetória dos alunos nesses ambientes.

HOODA, Monika; RANA, Chhavi (2020), em seu trabalho busca trazer reflexões sobre a melhora da educação superior a partir da lente do Learning Analytics (LA), buscando por meio de uma revisão de literatura, entender o atual cenário e o impacto do LA na melhora da educação superior. As pesquisas não conseguem apontar de forma clara e recente o cenário de LA no domínio da educação superior, porém demonstra o crescimento e os investimentos da área ao redor do mundo e aponta que isso vai melhorar a performance dos estudantes e professores nas instituições, apontando para um futuro em que as decisões são mais assertivas e dirigidas por dados.

O presente estudo responde algumas questões pertinentes aos desafios da área, atual estado e crescimento, qual o propósito das instituições adotando tais métodos, também fala sobre como as técnicas podem ser empregadas e sobre os desafios de se implementar tal área nas instituições, devido aos inúmeros fatores, como escalabilidade, inclusão, privacidade, ética entre outros. Tais questões norteadoras afloram interessantes reflexões não explícitas no texto sobre como é volátil e sensível o tema, pois depende de inúmeros fatores para seu sucesso e até a localidade onde está sendo desenvolvido pode interferir nos resultados, a cultura ou mesmo questões menos sensíveis como o volume de dados trafegando por este ambiente, todos esses pontos são de suma importância para manter o ecossistema vivo e atualizado de forma a trazer informações que sempre refletem a realidade.

NAMOUN, Abdallah; ALSHANQITI, Abdullah (2020) abordam em seu texto utilizando o método de revisão sistemática da literatura os modelos de predição de performance de estudantes utilizando mineração de dados e técnicas de Learning Analytics, esta revisão visa entender o cenário atual, demonstrar as limitações, alcance através da análise de alguns critérios

como, presença de implicações e recomendações nos textos, resultados suficientemente detalhados, limitações e desafios presentes nos textos, entre outros critérios para garantir uma avaliação de qualidade dos modelos.

Entre os desafios e limitações encontradas pelos autores, a falta de validação com mais de uma base de dados é uma das mais presentes nos trabalhos analisados e isto está fortemente ligado a como o mesmo modelo pode se comportar de forma diferente dependendo do cenário onde é aplicado, o que traz a tona outro desafio listado pelos autores, o desafio de evoluir uma análise preditiva para uma análise explicativa, onde o modelo leva em consideração o ambiente e o contexto visando gerar uma informação que não só prediz mas contextualiza o usuário do atual cenário e dá ao mesmo uma subsídio informacional sobre o curso em andamento. Poucos trabalhos também relatam a validade do modelo, ou sequer quais os fatores que ameaçam os modelos quando se trata do fator tempo de vida do modelo.

SANTOS, Jose Roberto (2019) analisam o escopo onde no ambiente de gestão do aprendizado moodle existem muitos dados e o desenvolvimento de aplicações de e Learning Analytics Dashboards (LAD) tem se desenvolvido para apoiar as análises de aprendizagem. Em seu estudo relata a baixa adesão dos educadores e busca analisar se este problema surge por ausência de ferramentas ou pelo desconhecimento das mesmas e para isso apresenta um estudo comparativo entre plugins de LAD para o ambiente Moodle relacionados a funcionalidades de acompanhamento do processo de aprendizagem pelo professor e pelo estudante.

O trabalho faz análise de cinco plugins do moodle sendo eles, analytics graphs, GISMO, completion progress, heatmap, forum graph, para avaliar estes plugins elenca sete questões a serem respondidas pelas análises dos recursos oferecidos, sendo elas questões sobre, acompanhamento e monitoramento de presença, atividades e interações em ferramentas de comunicação, também sobre a possibilidade de disparar alertas e mensagens automáticas e não automáticas, e se as informações estão disponíveis para professores e alunos. Ao fim das análises um quadro comparativo é desenvolvido a fim de mapear as ferramentas mais completas, tal quadro demonstra falta de preocupação com o disparo de alertas e mensagens e destaca que a maior parte da informação produzida está voltada para o professor e não para o estudante, ele destaca também a importância de alertas automáticos configuráveis para alertar os professores de casos específicos como, baixo número de acessos.

Apesar do trabalho ser focado em análises descritivas e carecer de análises preditivas e classificatórias, este trabalho traz a tona uma questão de importante sobre a temporalidade da informação, onde ressalta implicitamente a importância do professor saber das informações assim que acontecem pois o fator tempo pode estar intrinsecamente relacionado a assertividade

da decisão, principalmente em casos delicados como o de um estudante em curso que não se sabe o momento o qual o problema terá início.

BOGNÁR, László; FAUSZT, Tibor (2020) analisam em trabalho 16 diferentes modelos de aprendizagem de máquina para prever o sucesso de 57 estudantes de tempo integral do curso de estatística aplicada da universidade de Dunaújváros na Hungria, a partir da ótica das atividades cognitivas realizadas pelos estudantes. Para realização do estudo foram utilizados os seguintes parâmetros do ambiente Moodle os números de visualizações de leituras em pdf, vídeos, livros de exercícios resolvidos, minitab vídeos (que são vídeos de resolução de problemas com softwares de estatística) e quizzes e as notas dos quizzes foram utilizadas como fator de avaliação de sucesso. O trabalho avalia a qualidade dos modelos a partir de 3 métricas, a acurácia dos modelos, o F1 score que é uma métrica composta e complexa para medir a precisão de modelos, e por fim o Phi coefficient ou Matthews's correlation coefficient (MCC) que se trata de uma métrica de avaliação de qualidade de modelos de classificação binários, ou seja, modelos que tem dois resultados.

Os resultados são promissores do ponto de vista estatístico e trazem uma ferramenta capaz de realizar o propósito do trabalho, as avaliações de performance que levam em consideração o impacto de cada atributo e combinação na performance do modelo são providenciais para entendimento do cenário, porém apesar de um trabalho completo do ponto de vista estatístico carece de informação, os dados gerados pelos modelos são de qualidade mas a informação gerada é pobre e descontextualizada, onde um modelo de alta performance para o cenário especificado gera dados sobre os estudantes mas entrega ao tomador de decisão uma informação baseada em apenas no ponto de vista de uma segmentação de dados, a qual é assertiva mas não necessariamente traz ao tomador de decisão clareza sobre qual decisão tomar, tornando assim um dado o qual para o tomador de decisão serve como alerta mas não como direção, uma informação que sinaliza um possível acontecimento mas peca ao se contextualizar e ao ambientar o usuário dos possíveis fatores que induzem de tal acontecimento.

SALLOUM, Said A. (2020) realiza uma revisão sobre os trabalhos envolvendo mineração de dados na educação, em seu trabalho enumera diversas áreas que abordam esse tema na educação. Uma das áreas abordadas e a que condiz com este trabalho é a de predição de performance estudantil, onde ele ressalta a importância do uso de aprendizagem de máquina e o seu uso para “aprender” sobre os estudantes, ressaltando também sobre como estas abordagens podem encurtar alguns caminhos na descoberta de como melhorar a experiência dos estudantes. O autor também ressalta como o uso de aprendizagem de máquina pode ajudar, nesse ambiente diferente da classe tradicional, a encontrar pontos destoantes do normal que

seriam dificilmente encontrados por outros meios devido a grande diferena de ambiente e ao pouco tempo de exist4ncia desses novos ambientes.

OKUBO, Fumiya (2017) prop4e em seu trabalho um modelo de predio de notas finais de estudantes utilizando redes neurais recorrentes ou em ingl4s Recurrent Neural Network (RNN) com dados dos logs gerados pelo ambiente Moodle. O autor utiliza dos dados de logs baseado nas atividades realizadas pelos estudantes na plataforma como insumo de dados para alimentar um modelo de aprendizagem de m4quina, utiliza clusters para separar as notas finais dos estudantes m4trica a qual mede o aproveitamento sem a necessidade de precis4o num4rica decimal e utiliza de recurso temporal para avaliar a qualidade do modelo em uma tabela onde mostra a acur4cia do modelo implementado avaliando os estudantes a cada semana cursada, s4o 108 estudantes avaliados e o algoritmo consegue a partir da sexta semana uma precis4o acima de 90%.

O estudo apresenta um resultado 4timo o qual pode fornecer para os tomadores de decis4o subs4dios informacionais confi4veis desde que contextualizados na plataforma, apresentando assim aos usu4rios a origem dos dados e explicitando o objetivo do software al4m da forma a qual estes dados devem ser interpretados, devido ao uso de clusters para avaliar as notas dos estudantes, pode ser erroneamente interpretado pelo usu4rio.

CHAUHAN, Nupur (2019) os autores utilizam 5 algoritmos de aprendizagem de m4quina para predizer as notas finais dos estudantes, o estudo tem como objetivo encontrar os melhores modelos de aprendizagem de m4quina a fim de ajudar alunos a se destacarem em todas as etapas do curso. Utilizando os modelos K-nearest neighbor (kNN), Decision Tree, Support Vector Machines (SVM), Random Forest e Multiple Linear Regression, o autor avalia a performance dos algoritmos, utilizando os mesmos dados com mesma organizao e estrutura, a partir das m4tricas, Mean square error, root mean square error, mean absolute error, R-Squared e define que o modelo mais perform4tico para este cen4rio 4 o modelo Multiple Linear Regression.

O estudo se destaca por utilizar mais de um modelo de aprendizagem de m4quina avaliando assim a performance do modelo e n4o da infer4ncia de dados para determinar o melhor modelo. Outro ponto importante e de destaque 4 que o trabalho tem o estudante como persona e n4o o tomador de decis4o, o trabalho visa auxiliar os estudantes a melhorar em todos os est4gios do seu desenvolvimento, o que demonstra uma clareza na informao pois o estudo tem bem definido quem vai consumir a informao e como ela vai ser ingerida pela persona e para qual objetivo essa informao foi gerada.

Em complemento o ponto negativo do estudo 4 que n4o descreve a estrutura dos dados,

o fluxo dos dados e nem quais dados são utilizados, pecando na contextualização da origem dos dados, para o usuário, apesar de ter contexto do como utilizar e qual o objetivo, pode parecer uma informação gerada sem base sólida.

Adicionalmente, este trabalho se diferencia dos demais por trazer um modelo de ferramenta de processamento da informação para suporte à avaliação de performance e tomada de decisão no processo de ensino-aprendizagem de EAD, que visa subsidiar a tomada de decisão com informações quali-quantitativas contextualizadas e interpretáveis sobre a performance do aluno.

O modelo proposto tem como pilar a exploração dos dados espúrios dos registros de LOG da plataforma moodle, trazendo como contribuição análises dos padrões de comportamento dos alunos e informações contextualizadas sobre o passado do curso o presente a possível performance do aluno utilizando aprendizagem de máquina para auxiliar o processo de geração da informação.

Com intuito de atingir os objetivos e manter o rigor científico, além da relevância da pesquisa, é necessário seguir métodos científicos rigorosos. Para obter tal resultado, o próximo capítulo apresenta os procedimentos metodológicos os quais esta pesquisa utilizará durante seu processo de execução.

2 METODOLOGIA

Neste capítulo serão apresentados os procedimentos metodológicos e suas respectivas execuções para que seja possível a execução de proposta apresentada, entende-se como “metodologia” um conjunto de técnicas de averiguação ou verificação de determinada disciplina, com o objetivo de garantir o uso eficaz das técnicas selecionadas e gerar valor científico.

2.1 Natureza, Abordagem e Classificação

A natureza que mais se enquadra com a proposta e execução deste, é a pesquisa aplicada, descritiva e exploratória de cunho quali-quantitativa. Para Mascarenhas (2012, p. 45): “Baseia-se na quantificação para coletar e, mais tarde, tratar os dados obtidos”. De acordo com o apresentado esta pesquisa se enquadra em uma abordagem dirigida por dados e assim seguirá.

De acordo com Gil (1999), é possível agrupar as pesquisas científicas em 3 grandes grupos de objetivos: pesquisas descritivas, explicativas e exploratórias. Segundo o autor, as pesquisas descritivas têm como objetivo a descrição das características de determinada população, fenômeno ou o estabelecimento de relações entre variáveis. As pesquisas exploratórias, por sua vez, têm como finalidade desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses passíveis de verificação para estudos posteriores.

Este trabalho é caracterizado como uma pesquisa exploratória de cunho interdisciplinar, que se encaixa entre áreas consolidadas sendo elas a Ciência da Informação, Ciência da Computação e Inteligência Artificial. A associação das três áreas cria um cenário bastante específico a ser desenvolvido, sendo assim, a forma de pesquisa exploratória permite aproximá-las de forma a conhecer e entender os fatos e fenômenos do cenário.

Marconi e Lakatos (2012), nos trazem a pesquisa, ou estudos exploratórios como investigações, pesquisas empíricas, onde o principal objetivo é alcançar a definição de questionamentos ou problemas, ou seja, criar indagações que não seriam notáveis sem a devida profundidade em um determinado tema. Esses questionamentos notados, irão auxiliar a construir, corroborar ou modificar: hipóteses, familiaridade com o domínio, pesquisas futuras e conceitos.

O levantamento bibliográfico deste trabalho definiu como melhor abordagem a utilização do método de pesquisa aplicada dirigida por uma pesquisa exploratória e prova de

conceito, com o intuito de definir um recorte de um cenário específico que guiará o desenvolvimento do trabalho, colaborando também para avanços futuros no mesmo domínio.

2.2 Método de Pesquisa

Este trabalho se propõe a utilizar em seu desenvolvimento o método de pesquisa exploratória e uma prova de conceito. Onde utiliza uma abordagem de caráter empírica, para investigar um fenômeno dentro de um contexto contemporâneo e possibilita o conhecimento do mesmo, além de responder questões do tipo “como?” e “por que?”, baseando-se na análise de um limitado número de casos.

Para alcançar este objetivo, serão utilizados alguns conceitos do método estudo de caso. Yin (2015), diz que o estudo de caso é um método bem amplo, que abrange desde a fase de planejamento e levantamento de requisitos, até a coleta e análise de dados. Em complemento o autor traz ressaltos sobre a abrangência do método e a capacidade de trabalho com um, ou múltiplos casos, ou seja, associar diversos cenários e casos dentro de uma mesma pesquisa ou estudo, assim como diferentes abordagens, quantitativas ou qualitativas. Corroborando com o autor Voss, Tsikriktsis e Frohlich (2002) diz que, apesar de pouco estruturado, o estudo de caso promove o desenvolvimento de novas teorias e de novos insights.

Diante dos conceitos apresentados, esta pesquisa tem como objetivo analisar os dados gerado pelos alunos durante o uso do ambiente educacional MOODLE, utilizando o auxílio computacional dos algoritmos de aprendizado de máquina, para assim fornecer dados qualitativos sobre a performance dos mesmos aos tomadores de decisão, buscando diminuir a possibilidade de evasão do curso fornecendo subsídio informacional para que possam identificar previamente casos passíveis de evasão.

Como fora supracitado, o objetivo desta pesquisa é gerar informações qualitativas sobre os discentes para que a possibilidade de evasão dos alunos no ambiente EAD MOODLE possa ser identificada de forma antecipada auxiliando tomadores de decisão, conforme dito por Schramm (1971) a tendência principal nos estudos de caso, é que ela tenta esclarecer uma decisão ou decisões a partir dos motivos pelo qual foram tomadas, como foram implementadas e com quais resultados.

No contexto da pesquisa, esta tenta auxiliar a decisão dos gestores facilitando a análise de uma grande quantidade de dados, com intuito de esclarecer uma possibilidade de evasão baseada em dados, utilizando a implementação de algoritmos computacionais de aprendizagem de máquina para gerar resultados pertinentes aos gestores.

Em complemento, outra característica presente nesta pesquisa é o aspecto interdisciplinar presente na pesquisa, o qual pode ser observado a partir da necessidade de articulação entre as áreas de Ciência da Informação, Ciência da Computação e Pedagogia. A partir disso, este trabalho busca entender os fenômenos que acontecem em ambientes EAD, de maneira a gerar dados e informações que ofereçam subsídios para o apoio às tomadas de decisões pedagógicas.

A partir do contexto apresentado, existe a necessidade de se obter métodos de pesquisa e uma base epistemológica que contribuam para o contexto estabelecido, para isso, serão utilizadas algumas etapas do método estudo de caso, aliada à análise exploratória e a prova de conceito, a fim de alcançar o objetivo proposto com o devido rigor científico.

2.2.1 Proposta Metodológica

Este trabalho, refere-se ao estudo detalhado de um caso específico, cujo objeto de pesquisa é o ambiente MOODLE. Yazan e De Vasconcelos (2016) entendem que o método estudo de caso supõe o conhecimento de um fenômeno a partir da exploração de um caso único e pode ser aplicado no âmbito da Ciência da Informação.

Adicionalmente, a partir disso, algumas etapas do método de estudo de caso serão utilizadas, pois por se tratar de uma investigação empírica, a qual envolve métodos de coleta e análise de dados e, permite o pesquisador focar em um caso e reter uma perspectiva completa sobre um determinado fenômeno.

Para isto é necessário que, previamente, o pesquisador realize um planejamento rigoroso auxiliado de um abrangente referencial teórico.

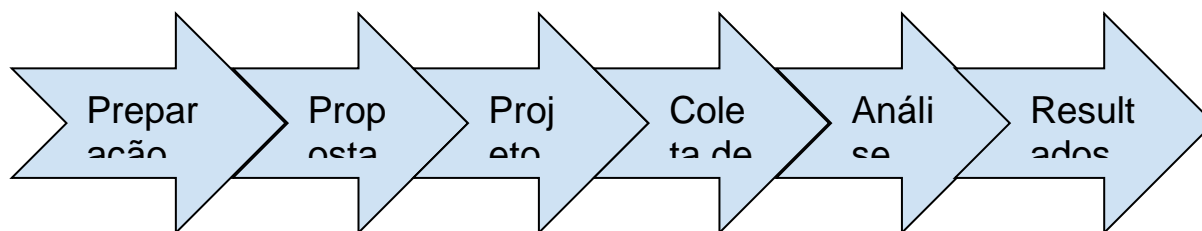
Porém para Yin (2015) obter uma definição clara acerca do estudo de caso é difícil e para isso apresenta em sua obra uma definição de cinco componentes necessários para um projeto, o qual este trabalho irá utilizar como base metodológica, embora o estado atual da ciência não forneça orientação detalhada sobre os últimos dois.

Este projeto considera o seguinte fluxo metodológico, preparação teórica por parte do pesquisador, proposta de trabalho como um todo, projeto que será realizado no trabalho, coleta dos dados necessários, análise dos dados obtidos e apresentação dos resultados do desenvolvimento do trabalho, como representado pela Figura 1.

A preparação teórica é o primeiro passo para o desenvolvimento da pesquisa e consiste na familiarização com o tema a ser pesquisado, e a delimitação do objeto ou artefato a ser estudado. Nesta etapa uma revisão de literatura será realizada com intuito de realizar o

embasamento teórico e a ampliação de conhecimento sobre o tema pesquisado.

Figura 1 – Fluxo metodológico



Fonte: Autoria própria

Após a revisão de literatura o alinhamento entre a proposta do trabalho e a literatura acerca dele será realizado, com intuito de alinhar os procedimentos metodológicos e expressar os conhecimentos adquiridos na etapa anterior e delimitar o escopo da pesquisa, indicar onde procurar evidência e fornecer insumos para preparação do projeto.

O projeto consiste na definição das unidades de análise e os métodos os quais serão aplicados a elas. A definição das abordagens a serem feitas e como serão feitas seguindo as seguintes etapas:

- Definir um processo informacional a partir do LMS Moodle, que descreva a estrutura dos dados, as transformações e manipulações que serão feitas e gerar uma informação contextualizada proveniente de uma análise preditiva;
- Especificar as limitações identificadas no processo de tomada de decisão educacional quanto ao uso de dados históricos para auxílio na avaliação de performance no LMS Moodle;
- Identificar técnicas de aprendizado de máquina em contextos educacionais já utilizadas;
- Avaliar e aplicar os algoritmos e técnicas de aprendizado de máquina a fim de identificar os modelos que melhor se adaptam ao problema.

Após o projeto a coleta de dados é iniciada, ela deve seguir em duas fases, primeiro a preparação para a coleta dos dados, que visa garantir o rigor do trabalho e qualidade dele, em seguida a execução da coleta, a qual deve assegurar a qualidade dos dados e dados suficientes para viabilizar a análise. Para que a coleta de dados obtenha sucesso em seu processo, a preparação para a coleta deve consistir na realização de processos que buscam garantir o sucesso da execução do trabalho, projetando a coleta de dados para que seja executada da

melhor forma e traga melhores insumos.

No processo de coleta de dados, este trabalho utilizará de uma fonte de dados, fonte a qual contém uma variedade de dados sobre o objeto estudado. A princípio esta pesquisa visa utilizar os LOGS de acesso fornecidos pelo ambiente EAD MOODLE, para garantir a neutralidade e a reprodução desta pesquisa, assim como proteger dados pessoais de terceiros, esta pesquisa utiliza dados de um ambiente aberto com dados que reproduzem a realidade de um ambiente EAD sem comprometer a integridade dos dados de terceiros.

As etapas seguintes referentes a análise dos dados e resultados, percorrerá ao longo da pesquisa conforme a coleta de dados. O estudo quali-quantitativo realizado nesta fazer requer a utilização de técnicas para compreensão dos dados e devido ao grande volume de dados técnicas computacionais serão utilizadas para o auxílio dessa tarefa.

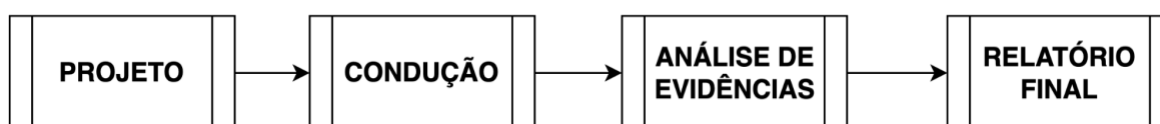
Para viabilizar uma melhor compreensão dos dados técnicas de visualização dos dados, técnicas de pré-processamento de dados, algoritmos de aprendizado de máquina e inteligência artificial serão utilizados nesse processo. Os resultados então serão revisados e relatados, assim poderá ser realizada a verificação das proposições e seguir para a conclusão.

2.3 Organização Metodológica

Assim como citado anteriormente este trabalho utiliza algumas etapas método estudo. Yin (2015), em seu trabalho propõe quatro etapas para a realização de um estudo de caso, sendo elas: 1) projeto do estudo de caso, 2) condução do estudo, 3) analisando as evidências do estudo de caso 4) composição do relatório final.

Estas etapas se enquadram nas quatro últimas etapas do fluxo metodológico deste trabalho, estando presentes nas partes de projeto, com a primeira etapa de estruturação projeto sendo a coleta de dados, com a etapa de condução do estudo, também presente na parte de análise de dados com a etapa 3 analisando as evidências e por fim nos resultados com a etapa 4, composição do relatório final.

Figura 2 – Organização metodológica



Fonte: Autoria própria

A primeira etapa a ser definida é o projeto e para isso são definidas as as questões de estudo, onde é definido por Yin (2015) como a forma da questão - em termos de "quem", "o quê", "onde", "como" e "por que" - fornecendo uma chave importante para se estabelecer a estratégia de pesquisa mais relevante a ser utilizada. Sendo mais apropriadas às questões do tipo “como” e “por que”, assim tornando-se necessário ter clareza sobre a natureza das questões do estudo.

Após definidas as questões a serem realizadas, o projeto necessita de um objeto de estudo, ou seja, algo para ser a base do projeto, podendo esse ser um indivíduo, um evento, entidade, decisões, ambiente, entre outros. Neste estudo o ambiente EAD MOODLE será analisado a partir dos LOGs de acesso gerados pelos usuários. E para isso o capítulo quatro contextualiza as etapas que serão conduzidas do ponto de vista informacional.

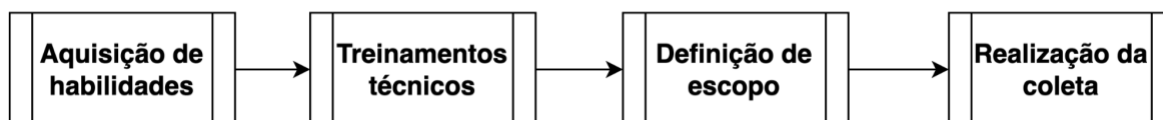
As duas últimas etapas do projeto são, a lógica que une os dados e as proposições e os critérios para se interpretar as descobertas. Para isso o capítulo cinco traz informações de como o projeto foi de fato desenvolvido demonstrando as técnicas de análise de dados, como a Análise Exploratória de Dados (EDA) cuja definição é uma abordagem para a exploração dos dados que se deseja analisar, posteriormente os algoritmos computacionais de aprendizado de máquina e inteligência artificial aplicados para que então os resultados sejam interpretados sob a visão das proposições e dos critérios propostos.

Após a conclusão do projeto e suas minúcias, inicia-se a etapa de condução do projeto, onde é dividida em duas etapas, sendo elas: 1) preparação para a coleta de dados e 2) coleta de evidências.

A primeira etapa consiste na preparação para a coleta de dados, cujo envolve as habilidades adquiridas previamente por parte do pesquisador, os treinamentos realizados e a preparação para o início do projeto.

Dito isso, se faz necessário por parte do pesquisador realizar sessões intensivas de treinamento, com intuito de desenvolver e aprimorar habilidades técnicas para execução do trabalho.

Para realização desta etapa então foi elencado 4 etapas, sendo essas: 1) aquisição de habilidades por parte do pesquisador, 2) treinamentos técnicos, 3) definição de escopo e 4) realização da coleta, como ilustrado na Figura 3, estas etapas se apresentam no fluxo metodológico inseridas nas partes de preparação teórica, proposta e projeto, por se tratarem de partes que envolvem capacitação e preparação, além disso quando se trata da organização metodológica estas etapas se encaixam na parte de projeto, onde a preparação para a realização do trabalho é realizada.

Figura 3 – Preparação para coleta de dados

Fonte: Autoria própria

Após realizada a preparação, a segunda etapa de coleta de evidências tem início, nesta etapa, são considerados aspectos como, as fontes de evidência, sendo elas os dados disponíveis no ambiente Moodle de acordo com sua documentação oficial. O processo de coleta de dados para o estudo de caso é mais complexo que os processo utilizado em outras estratégias de pesquisa, diz Yin(2015), pois o pesquisador deve ser versátil metodologicamente e deve obedecer certos procedimentos formais para garantir o controle de qualidade.

Para garantir a qualidade dos dados, o autor sugere dois princípios sendo eles: 1) utilizar várias fontes de evidência e 2) manter o encadeamento de evidências, o ambiente Moodle, por ser um ambiente projetado para a alta replicabilidade, fornece todos estes fatores em sua plataforma.

Estes dois princípios mantêm a qualidade dos dados, a segurança dos dados, a replicabilidade do experimento e a confiabilidade das informações, porém o autor não limita o uso destes dois princípios pontuando que podem ser usados mais ou menos princípios caso necessário, justificando sua fala sobre a versatilidade do pesquisador.

Finalizada a coleta de dados, a etapa de análise de evidências, a qual consiste em examinar, categorizar, classificar, recombinar as evidências tendo em vista as proposições do estudo.

De acordo com o autor Yin (2015) analisar as evidências é uma tarefa árdua, pois não existem estratégias ou técnicas para todos os casos, cada caso tem sua peculiaridade e demanda do pesquisador habilidades específicas.

Para isso o autor propõe o início das análises com uma estratégia analítica geral e no decorrer da análise estabelecer prioridades do que deve ser analisado. Quatro estratégias são propostas como exemplos efetivos, sendo elas, adequação ao padrão, construção de explanação, análise de séries temporais e modelos lógicos de programa, porém o autor não limita a somente essas e as deixa como exemplo de aplicações.

Concluindo o estudo vem o relatório final, cujo trata da exposição dos resultados obtidos. Para isso devem ser seguidos processos de composição, sendo eles: identificar o

público almejado para o relatório, desenvolver uma estrutura de composição e adotar certos procedimentos.

O relatório final não segue uma fórmula pré-definida dando liberdade pro autor discutir e apresentar os resultados da forma que mais convém, o que justifica o rigor ao identificar o público, sua estrutura deve ser pensada para transmitir o todas as informações podendo então conter textos, imagens e afins.

Em complemento, neste capítulo buscou-se abordar o problema de pesquisa e os procedimentos metodológicos que serão utilizados, discutindo algumas de suas características considerando a interdisciplinaridade entre as áreas da Ciência da Informação, da Ciência da Computação e da Pedagogia, em seus pontos mais relevantes.

Este trabalho visa obter como resultados responder às questões: 1) Como dados provenientes de algoritmos de aprendizado de máquina, vão efetivamente compor melhores informações referente a performance dos discentes? 2) Estes dados podem influenciar a tomada de decisão? 3) Quais informações podem ser geradas para auxiliar o tomador de decisão quanto a performance do aluno? 4) Quais dados estão disponíveis e podem ser utilizados? e 5) Estas informações podem gerar algum tipo de discriminação?

Enquanto a revisão de literatura era desenvolvida, notou-se a importância do estudo de caso para pesquisas em Ciências da Informação e para pesquisas interdisciplinares. Apresentando um rigor metodológico e fornecendo maneiras de se integrar várias áreas de conhecimento em um só ambiente, podendo explorá-lo ao máximo, buscando produzir conhecimento para apoiar solução de problemas de fenômenos contemporâneos inseridos em algum contexto da vida real. Vale ressaltar que na estruturação da pesquisa, as escolhas metodológicas levaram em conta a classificação descritiva e exploratória, a natureza quanti e qualitativa e a abordagem dirigida por dados.

No próximo capítulo será abordado o processo de avaliação de performance em sistemas de gestão de aprendizagem, visando a partir da ótica informacional, identificar os principais problemas informacionais do processo a fim de compor uma melhor informação final.

3 O PROCESSO DE AVALIAÇÃO DE PERFORMANCE EM SISTEMAS DE GESTÃO DA APRENDIZAGEM

Neste capítulo serão abordadas as etapas do processo de avaliação de performance que compõe o processo de tomada de decisão pedagógica, a partir da ótica informacional, estas etapas são de suma importância para entender o processo e identificar os problemas com intuito de compor a informação final em acordo com a necessidade real e melhor modelar a solução proposta. Entende-se, em meio ao apresentado, que o processo de análise de performance do discente é muito amplo, logo este trabalho aborda somente a análise de performance do discente no sistema de gestão da aprendizagem MOODLE.

3.1 Processo de Tomada de Decisão Pedagógica

A tomada de decisão pedagógica refere-se a um processo complexo que envolve diversas etapas e variados níveis de conhecimento referente aos estudantes. Estas decisões podem ser realizadas por meio de decisões intuitivas, analíticas e deliberativas, de forma que possibilitem aos tomadores de decisão a incumbência de gerar impacto positivo na experiência da aprendizagem, para tanto, a efetividade dessas informações é gerada por meio de dados educacionais produzidos, entre outras fontes, a partir dos Sistemas de Gerenciamento da Aprendizagem (LMS – Learning Management System) (PAIVA, 2017).

No ensino presencial, o professor tem a sua disposição recursos de acompanhamento que apoiam os processos de tomada de decisões pedagógicas, como a utilização de técnicas de observação em sala de aula, entre outros. Esses recursos favorecem a verificação do aproveitamento, das aptidões e do ajustamento pessoal e social dos alunos. No contexto da educação à distância, encontra-se dificuldades ou até impossibilidades de fazer uso dessas técnicas, dentre os motivos destaca-se a baixa interação humana e feedback, sendo estas algumas das limitações apresentadas em relação a utilização dos LMS, que precisam ser superadas para que se obtenha maior eficácia nos processos de decisões pedagógicas (DIAS JR, 2017; MARGARYAN, BIANCO, LITTLEJOHN, 2015; PAIVA, 2017).

DA SILVA, Danilo Garcia; ALONSO, Kátia Morosov; MACIEL, Cristiano (2012) apontam em seu estudo, o uso operacionalizado das ferramentas disponibilizadas no sistema de gestão da aprendizagem MOODLE, sendo utilizada pelos professores, mas carente de entendimento das ferramentas como um subsídio informacional para auxílio em sua tomada de decisão, muitas vezes pelo desconhecimento de tal abordagem ou por dificuldade com o uso da

plataforma.

Conclui-se que há espaço para uma gama de oportunidades para aplicação de inúmeras ferramentas para processamento da informação com intuito de elucidar a informação aos usuários de sistemas de gestão da aprendizagem, como dito por LOPES e SCHIEL:

"percebe-se claramente que um longo caminho pode ser construído, e um dos grandes problemas é articular as inúmeras vantagens criadas pelas tecnologias e criar modelos de acompanhamento do aprendizado que efetivamente contenham alguma possibilidade de mudança qualitativa (LOPES; SCHIEL, 2004, p. 1). "

Paiva (2017), Dias Jr (2017) e Branco, Conte e Habowski (2020) também salientam que para ofertar o suporte requerido aos alunos, é necessário que os profissionais da educação analisem dados decorrentes das interações dos estudantes, porém para realizar tais análises os profissionais da educação, professores, tutores, gestores, necessitam de treinamento para desenvolver tais habilidades, que muitas vezes, principalmente em um ambiente como o MOODLE, não são habilidades relacionadas somente à educação e sim a manipulação de dados utilizando ferramentas computacionais, este é um esforço interdisciplinar entre a matemática, a computação, a informação e a educação, o que geralmente foge ao propósito do educador, caso o educador tenha tais habilidades ou atue no ensino de áreas correlatas é entendível que o mesmo possa saber utilizar e manipular tais ferramentas, mas isso é um caráter de exceção e não deve ser o foco do educador.

Em complemento, tais ferramentas quando entregues à mão de profissionais os quais não estão preparados para manipulá-las, ou não estão ambientados ao contexto dos dados apresentados pelas ferramentas, tornam-se apenas dados vazios, sem informação. Dados gerados por essas ferramentas sem o devido cuidado informacional de organizá-los para que sejam corretamente interpretados, não se tornarão informações, serão apenas dados sem significado ou utilidade.

Como dito por Araújo, C. A. Á. (2009) o usuário precisa ter conhecimento de sua ausência de conhecimento, para que a informação possa ser vista como algo útil na perspectiva do sujeito, ou seja, ao desenvolver tais processos informacionais, ou utilizar tais ferramentas, é necessário levar em consideração a possível ausência de conhecimento do usuário, isto é, junto ao contexto de como, com qual, para que e para quem essa informação foi gerada, tornando o dado gerado por software uma informação direcionada ao usuário, com intuito de resolver um problema e extinguir a possível ausência de conhecimento.

3.2 Uso de Ferramentas Para Auxiliar a Tomada de Decisão em Ambiente Educacional

Como supracitado o uso de ferramentas para analisar dados provenientes dos alunos é uma forma de conseguir transformar uma grande quantidade de dados em informações úteis, a partir dessas informações provenientes de ferramentas externas, os tomadores de decisão adquirem um maior entendimento do ambiente e do momento em que estão antes de tomar sua decisão.

Este processo todo de tomar conhecimento sobre a situação é chamado de Consciência Situacional (do inglês *Situational Awareness - SAW*) e refere-se a um conceito fundamentado no processo de tomada de decisão de cada indivíduo. Um indivíduo com SAW bem estruturada e definida está ciente das informações ao seu redor, compreende-se que este tem entendimento de como a informação ao seu redor será utilizada em todos os períodos temporais, passado, presente e futuro, tornando o processo decisório mais eficaz e assertivo, de acordo com Endsley (1988).

Pode-se identificar tal estado em todos os cenários, sendo eles informacionalmente simples, ou complexos, onde há informações presentes e uma tomada de decisão a ser feita o SAW está presente, seja em um simples ato de desviar de outros pedestres ao caminhar em um lugar populoso ou em um complexo cenário onde um piloto deve observar todas as condições aéreas, comunicar-se com a torre de comando e utilizar da técnica e ordem correta de ações para pousar a aeronave, como dito por Kokar e Endsley (2012).

Dito isso uma ferramenta que auxilia o processamento da informação e gera insumos informacionais para os usuários, de forma a qual este, por meio desses novos insumos pode ter um melhor entendimento do ambiente em que está situado e compreende como essas informações podem interferir em sua tomada de decisão, este tem uma SAW bem estruturada e definida, podendo assim ter maior assertividade em sua decisão.

Os autores WANG, Amy et al. em seu estudo sobre decisões dirigidas por dados, comparou-se cenários onde, para alguns grupos é oferecido a professores, dashboards explicativos com dados analíticos sobre os alunos e para outros grupos não é oferecida essa ferramenta. Após acompanhamento destes grupos notou-se uma melhora de performance dos alunos de baixo rendimento em comparação ao grupo que não disponibilizavam de dados analíticos para avaliar os discentes, corroborando com o supracitado, onde, quando é considerada a possível ausência de conhecimento por parte do usuário e se desenvolve uma ferramenta intuitiva, tomando a perspectiva do sujeito como necessária, a informação gerada torna-se útil e acessível, aumentando a SAW dos professores e tornando a tomada de decisão

mais rica de informação e mais assertiva.

Os autores SANTANA, Marcelo A.; DOS SANTOS NETO, Balduino Fonseca; DE BARROS COSTA, Evandro (2014) em sua obra relatam que as ferramentas disponibilizadas são subutilizadas, ou seja, não são exploradas em sua totalidade. Corroborando com esta afirmação os autores DA SILVA, Danilo Garcia; ALONSO, Kátia Morosov; MACIEL, Cristiano (2012) relatam que os professores tem acesso, tem conhecimento sobre as ferramentas, sabem como encontrá-las e utilizá-las, mas o fazem de forma operacional, ou seja utilizam por convenção e não por conhecimento ou por esta ser relevante em seu processo. Para que ferramentas sejam úteis os usuários precisam entender a relevância delas, caso contrário se tornará apenas burocracia e será utilizada de forma operacional.

Em complemento, para que ferramentas de análise de aprendizagem sejam eficazes, é necessário ter utilidade e relevância para educadores. O processamento de dados deve se traduzir em feedback em tempo real, intervenção mais rápida e instrução individualizada e personalizada. Professores precisam de autonomia suficiente para gerenciar suas turmas, pois são eles que estão mais familiarizados com as necessidades de seus alunos. No entanto, para serem capazes de usar tecnologias de forma eficaz, novas competências precisam ser assimiladas pelos professores, tais como crítica, analítica e habilidades para interpretar dados, não deve ser confundido com necessidade de ter conhecimento técnico de como adquirir, processar e gerar esses dados, mas saber como utilizar e qual a relevância deles.

Adicionalmente, com o objetivo de desenvolver tais ferramentas e auxiliar a geração de subsídios informacionais aos professores, alunos e tomadores de decisão, áreas como Learning Analytics surgem para criar profissionais e pesquisadores que se concentrem em avançar os estudos sobre os dados educacionais, apoiados pelos profissionais da educação estes profissionais são responsáveis por gerar a informação e satisfazer as necessidades informacionais destes educadores, a partir de técnicas de ciência de dados, ciência da computação, matemática e estatística, apoiados em problemas reais do ambiente educacional, como identificar o perfil do aluno, coletar e qualificar dados, relacionar dados, classificar interações entre outras funções.

Pinnell (2017) e Filatro (2019) dizem que a área de Learning Analytics é considerada parte da ciência de dados educacionais e o seu propósito é identificar o perfil dos alunos, coletar e qualificar o maior número possível de dados e informações referente suas interações com os ambientes de aprendizagem, de forma a possibilitar a compreensão do progresso destes alunos e sua performance, construir práticas pedagógicas satisfatórias, identificar situações de risco e avaliar fatores que afetam a conclusão e o sucesso dos estudos.

Não obstante ao contexto, Schildkamp (2019) corrobora e acrescenta que a disponibilidade de dados educacionais favorece a procura por padrões sem ter hipóteses definidas, podendo descobrir insights que nunca foram pensados.

3.3 Avaliação de Performance do Estudante em Ambiente Lms

A avaliação de performance do estudante é crucial na tomada de decisão, pois esta performance não se trata somente do quão bem o aluno vai em um exame formal, como uma prova ou um trabalho, a performance permeia campos como a motivação do estudante, seus estados mentais, sua absorção de conhecimento, suas habilidades de resolver problemas, entre outras capacidades que o discente realiza.

U. Jayasinghe, A. Dharmaratne and A. Atukorale, separa, de acordo com o sistema tradicional de educação, a avaliação de performance dos estudantes em duas categorias, sendo elas a Avaliação formal e a Avaliação informal. A avaliação formal trata de exames, provas, trabalhos, testes, entre outros processos que avaliam e retornam ao estudante uma nota. A avaliação informal trata de outras formas de avaliação não quantitativas, ou seja, avaliações emocionais, comportamentais, entre outras avaliações que o professor pode fazer do estudante pelo convívio, conversa, comportamento, entre outras.

Em complemento, as avaliações informais, são de suma importância para os professores em sua tomada de decisão, pois elas que trazem a informação pro professor de forma antecipada, ou seja, antes do aluno realizar uma avaliação formal e reprovar, o professor por meio de avaliações informais consegue entender a situação e direcionar o estudante para o caminho da aprovação.

Adicionalmente, em um ambiente presencial, o professor tem contato direto com os alunos. logo, é natural que o próprio professor crie formas de avaliação do aluno, como a interação com as aulas, o interesse demonstrado, analisar o comportamento dos alunos durante as aulas, entre outras formas que vão auxiliar o professor a entender a dimensão da avaliação informal. Porém, em um ambiente online. onde não se tem essa proximidade e esse convívio social. é mais difícil para o professor desenvolver essas técnicas de avaliação naturalmente, já que ele não tem tantas informações, ao qual ele já está acostumado a trabalhar e a interpretar, tendo somente uma quantidade muito grande de dados e muitas vezes sem o conhecimento necessário para explorar e interpretar esses dados.

Para resolver a carência supracitada, novas ferramentas precisam ser desenvolvidas, ou seja, novas abordagens que transformem esses dados inoperantes em dados que possam ser

interpretados pelos professores. Com isso inúmeros trabalhos relacionados a criação de ferramentas surgiram, alguns utilizando Internet Of Things (IOT) para coletar mais dados sobre os discente e gerar informações semelhantes às do ensino presencial como, emoções dos alunos, motivação, entre outras, já outras ferramentas focam nos dados já obtidos e produzidos pela plataforma utilizando de análises estatísticas, técnicas de aprendizagem de máquina e outros recursos computacionais para prover informações relevantes, mas todas abordagens com a mesma motivação de trazer melhores informações para os professores.

Nesse sentido, dentro da área de Learning Analytics uma de suas ramificações busca estudar e melhorar o entendimento dos professores quanto a performance dos alunos, por meio do processamento dos dados da plataforma, gerando subsídios informacionais para o professor, com técnicas de mineração de dados, aprendizagem de máquina, análises descritivas, análises quantitativas, análises preditivas, entre outras formas para tentar prover melhores dados aos professores, como o estudo realizado por CHAUHAN, Nupur (2019) onde a aprendizagem de máquina é ferramenta para produção da análise preditiva da performance do estudante.

Adicionalmente, SHAHIRI, Amirah Mohamed et al, em seu trabalho revisando as técnicas de mineração de dados usadas para prever a performance dos estudantes, o autor reafirma a importância de prever a performance dos estudantes para auxiliar os professores e os estudantes no processo de ensino e aprendizado, finaliza dizendo que o desenvolvimento dessas ferramentas em inúmeros ambientes vão ajudar a desenvolver o sistema educacional no quesito de monitoramento dos estudantes de forma sistemática quanto a sua performance.

Estas afirmações corroboram e contribuem com a evolução dos sistemas de gestão da aprendizagem, para cada vez mais se aproximar do ambiente tradicional quando se trata das avaliações informais e de entender o ambiente e a situação de cada aluno para melhorar seu desempenho.

3.4 Levantamento e Discussão Sobre o Uso de Dados em LMS Para Tomada de Decisão Pedagógica

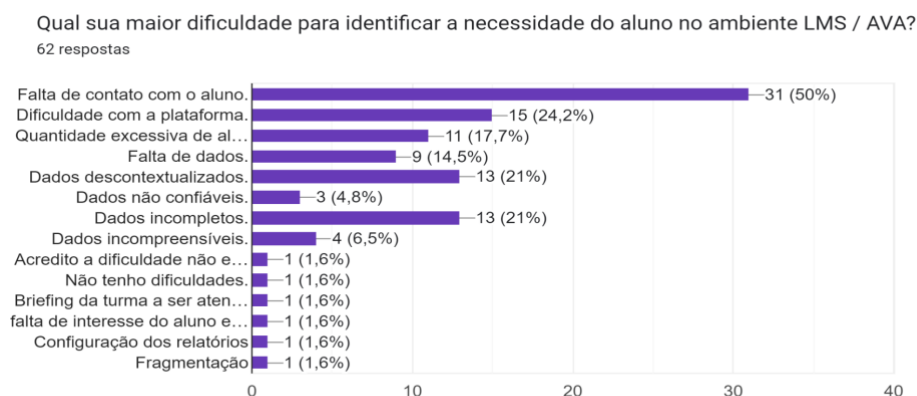
O Grupo de Interação Humano-Computador (GIHC), por meio da realização de um questionário direcionado a professores que utilizam sistemas de gestão da aprendizagem no seu dia-dia, com a contribuição de 62 respostas, levantou dados pertinentes a este trabalho quando se trata da tomada de decisão pedagógica em ambientes de gestão da aprendizagem como Moodle.

Em complemento, este levantamento teve como objetivo entender as reais necessidades

dos professores quando se trata da utilização de ambientes de gestão do aprendizado para tomada de decisão pedagógica, partindo do pressuposto da tomada de decisão dirigida por dados e baseado em relatos de pesquisas relacionadas onde apontam o uso operacionalizado das ferramentas em ambiente LMS, ou seja o uso sem ter o entendimento do pra que ou por que como relatado pelos autores SANTANA et al (2014) onde relata a subutilização das ferramentas disponíveis, corroborado por DA SILVA et al (2012) onde diz que os professores conhecem as ferramentas mas utilizam de modo operacionalizado, SANTOS, Jose Roberto et al complementa dizendo que há uma baixa adesão de plugins analíticos por parte dos professores. Em suma, este levantamento tem como objetivo também analisar as maiores dificuldades no uso, entendimento e manipulação dos dados disponíveis na plataforma, de forma a mapear diretamente com o usuário final suas necessidades, ressaltando o caráter anônimo da pesquisa e para fins exclusivamente acadêmicos.

Neste levantamento, ao se deparar com déficit informacional no processo de de identificar as necessidades dos alunos os professores relataram suas maiores dificuldades ao tentar interpretar a situação do estudante quando, sob a ótica da tomada de decisão pedagógica, encontram-se em uma situação onde precisam realizar uma avaliação do estudante usando as técnicas e ferramentas que tem a disposição para aumentar seu entendimento sobre a situação do estudante em relação ao curso e ao ambiente em que ele está situado, sendo as dificuldades as mais relatadas a falta de contato com o aluno, quantidade excessivas de alunos, as dificuldades com a plataforma e falta de dados e dados descontextualizado somaram a maior parte das reclamações dos professores em relação ao ambiente LMS, como descrito na Figura 4.

Figura 4 – Dificuldades dos docentes



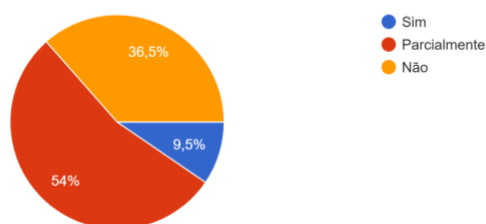
Fonte: Autoria própria

Em complemento, uma forma de amenizar tais problemas supracitados é com a utilização de ferramentas, estas podendo ser nativas do ambiente ou anexadas, os plugins. A fim de tentar identificar o uso dessas ferramentas, já que como fora constatado no capítulo anterior o índice de utilização de ferramentas do Moodle tem aumentado, mas ainda são subutilizadas, neste levantamento os professores foram questionados sobre a eficiência dessas ferramentas, e como resultado foi apontado que os ambientes carecem de informações adicionais para acompanhamento do desenvolvimento do aluno como descrito na Figura 5.

Figura 5 – Satisfação com o ambiente

Em sua opinião, o LMS que você utiliza oferece um sistema de acompanhamento do desenvolvimento do aluno de forma eficiente?

63 respostas



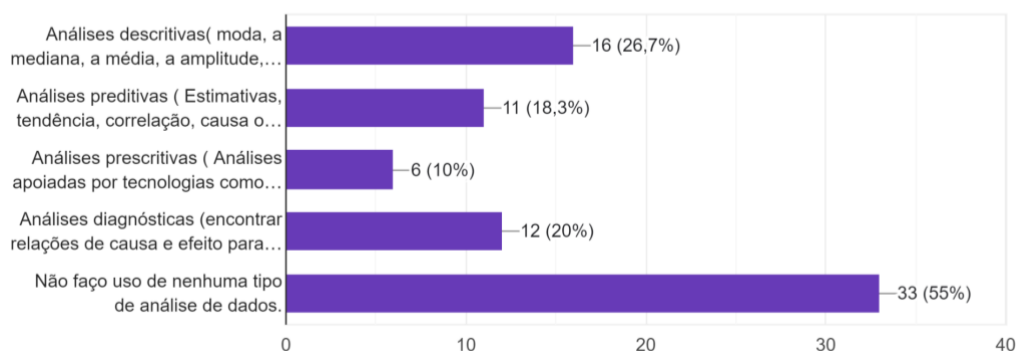
Fonte: Autoria própria

Adicionalmente, por mais que ambientes como Moodle forneçam uma série de ferramentas próprias para auxiliar no acompanhamento do desenvolvimento do aluno e tenha uma comunidade que desenvolve e fornece plugins para agregar ao sistema novas funcionalidades, ainda não é difundida e esta mesma pesquisa verificou que há pouco uso de ferramentas analíticas neste ambiente, como demonstrado na Figura 6.

Figura 6 – Uso de ferramentas analíticas

Você faz uso de alguma ferramenta analítica* sobre os alunos?

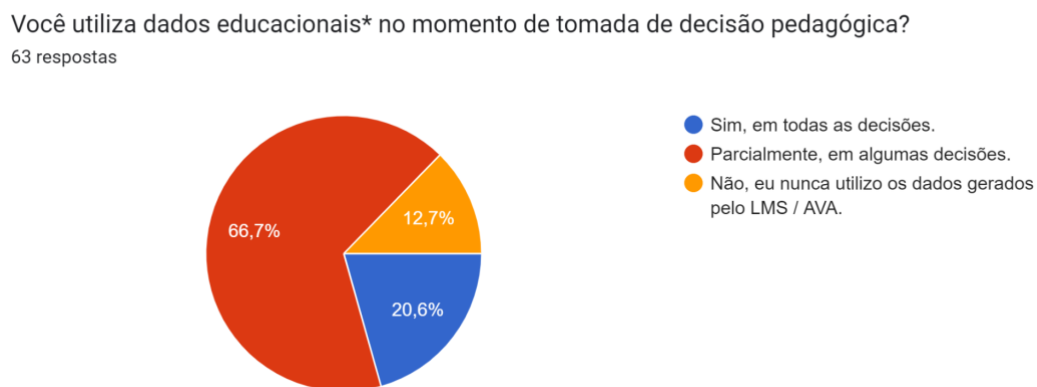
60 respostas



Fonte: Autoria própria

Em complemento mesmo com a baixa adesão de ferramentas analíticas, os professores ainda demonstram um apoio ao uso de dados educacionais, mesmo que de forma parcial, como demonstrado pela Figura 7.

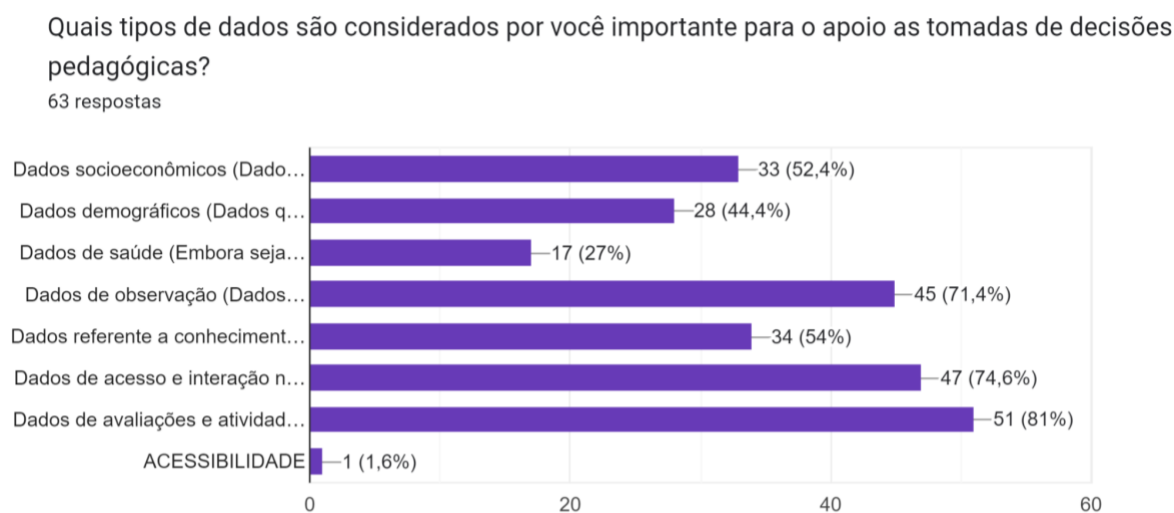
Figura 7 – Uso de dados educacionais



Fonte: Autoria própria

Adicionalmente, apesar da aderência ao uso de dados ser parcial, quando se trata da importância dos dados para tomada de decisão, os professores veem, neste ambiente específico a distância, em sua maioria os dados referentes à interação, observação, acesso e a avaliação formal em si, como os dados mais importantes para basear suas decisões como descrito na Figura 8.

Figura 8 – Dados mais importantes

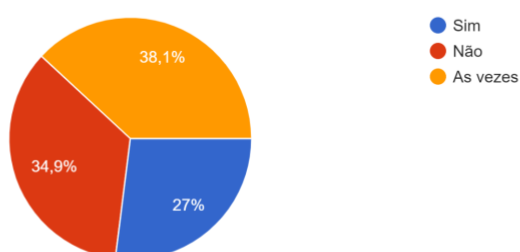


Fonte: Autoria própria

Em complemento, é de suma importância constatar que estes dados os quais os professores consideram mais importantes são de característica descritiva comportamental, ou seja descrevem uma ação, descrevem um comportamento, demonstrando claramente um espaço para o desenvolvimento de ferramentas analíticas e também corroborando para a utilização dos dados de Log, que são dados com as mesmas características e que apesar de ricos em informação, principalmente por conter em seus registros os dados históricos dos alunos, como visto na Figura 9 estes dados também são parcialmente utilizados ou não utilizados no caso dos 34,9% que além de não utilizar, uma parcela majoritária destes que não utilizam informam que desconheciam essa opção ou tem dificuldade em seu manejo como demonstra a Figura 10.

Figura 9 – Uso de registros de Log

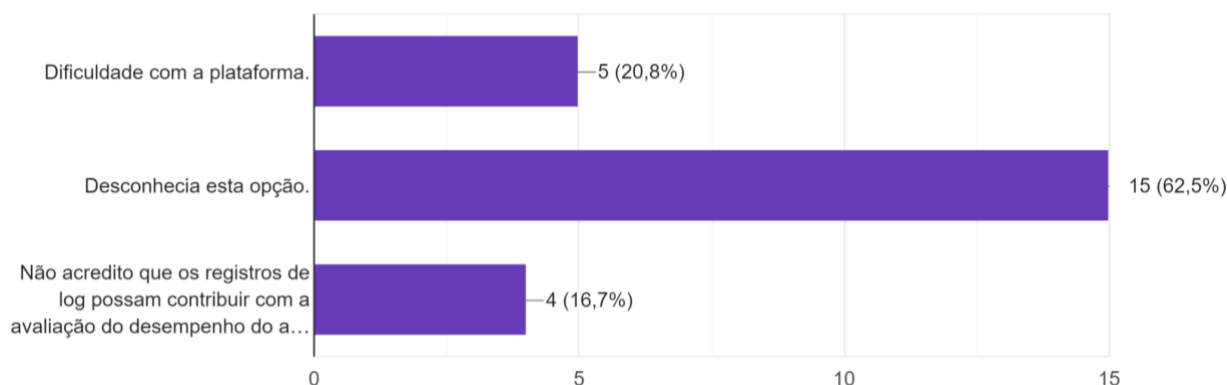
Você faz uso dos registros de "log" gerados pelo LMS / AVA no momento de tomadas de decisões pedagógicas?
63 respostas



Fonte: Autoria própria

Figura 10 – Motivo do não uso dos Logs

Caso sua resposta anterior tenha sido negativa, você não utiliza por qual motivo?
24 respostas



Fonte: Autoria própria

Em conclusão, é visível a abertura por parte dos professores quanto a utilização de dados para tomada de decisão pedagógica, essa abertura unida a constatação de que os professores consideram importantes os dados descritivos comportamentais dos alunos, trazem à tona a relevância de se criar ferramentas analíticas mais intuitivas e contextualizadas para que os professores passem a utilizar os dados de forma ativa e não operacional, tais ferramentas precisam extrair mais dados úteis aos professores, o ambiente tem inúmeras formas de prover dados mas necessita de intervenção para processá-los e torná-los úteis.

No próximo capítulo será abordado o processamento da informação no ambiente moodle onde será descrito suas fontes de dados, como estruturar esses dados, como fazer transformação dos dados em informação, a utilização de aprendizagem de máquina para realizar análises preditivas, os artefatos informacionais e a camada de visualização, visando trazer um melhor entendimento da informação e dos processos que a rodeiam.

4 PROCESSAMENTO DA INFORMAÇÃO NO AMBIENTE LMS MOODLE

Neste capítulo serão abordadas as etapas presentes no processamento da informação no ambiente Lms moodle, desde da obtenção de dados bruto, manipulação e adequação dos dados, as necessidades informacionais dos usuários, transformação do dado em informação e as métricas de qualidade utilizadas. Este capítulo visa trazer um melhor entendimento da informação e dos processos que a rodeiam, a fim de trazer um melhor contexto de como a informação é coletada e produzida nesse ambiente e quais técnicas são utilizadas.

4.1 Fontes de Dados

O Modular Object-Oriented Dynamic Learning Environment ou Moodle é um projeto de software aberto que visa ajudar educadores a criar um ambiente educacional online de qualidade, sendo um software gratuito e altamente versátil e podendo ser adaptado para cada necessidade particular como dito por BRANDL, Klaus. O Moodle é um ambiente que é alimentado por sua comunidade de usuários e vem evoluindo conforme os anos, sendo muito utilizado por ser uma plataforma gratuita que fornece uma gama de funcionalidades já nativas da plataforma além de uma comunidade gigante que sempre cria novas funcionalidades e novas ferramentas para conctar com este ambiente, os chamados plugins.

Apesar de ser uma ferramenta que detém uma grande comunidade que cria inúmeros plugins, o Moodle por si só em sua versão inalterada provém de inúmeros recursos para o auxílio da gestão e acompanhamento do ensino aprendizagem em sua plataforma. Em sua documentação oficial, que é disponibilizada em seu site oficial moodle.org, encontram-se inúmeros tutoriais, visões gerais, explicações entre outros documentos sobre a plataforma, sua totalidade no idioma inglês.

Em meio a sua documentação o Moodle oferece uma interface atualizada e moderna desenvolvida para ser acessível, responsiva e de fácil navegação tanto em computadores quanto em dispositivos mobile como telefones e tablets. Nesta interface encontra-se, calendários para acompanhamento do cronograma acadêmico, cursos, reuniões, eventos pessoais, entre outras coisas, o gerenciamento de arquivos também é disponibilizado e pode ser integrado com outros serviços de armazenamento como dropbox, google drive, entre outros, o ambiente também disponibiliza em seus recursos alguns dashboards personalizados, notificações, um editor de texto intuitivo, mas as ferramentas mais importantes e que mais geram dados referente aos discentes, são as ferramentas de acompanhamento de progresso e de atividades colaborativas.

As atividades colaborativas, são atividades em que o discente interage com a plataforma a fim de um resultado, cada atividade tem seu objetivo e pode ser personalizada de acordo com a demanda. O moodle oferece quinze atividades padrões as quais se pode ativar ou desativar como o ambiente necessitar.

Dentre essas atividades estão as atribuições, que permite que o professor avalie e comente as atividades enviadas pelos alunos, ou seja, trabalhos dissertativos, planilhas, arquivos em geral que foram demandados por uma certa atividade criada pelo professor. Muito próxima a esta atividade, o ambiente fornece também a opção de criar questões de múltipla escolha facilitando a correção e a avaliação por parte do docente e trazendo outra possibilidade de avaliação formal.

Outra funcionalidade é a de videoconferências ao vivo dentro do ambiente do moodle, onde o professor pode transmitir e realizar aulas ao vivo, palestras, treinamentos entre outras atividades. Junto a essa ideia de sincronidade a funcionalidade do chat ajuda os alunos e professores a discutir por texto de forma direta, ou em grupo, sincronamente dentro do ambiente. Caso o docente não queira realizar de forma síncrona os processos da discussão o professor pode enviar de forma assíncrona um feedback para o aluno, respondendo sua questão ou avaliando seu desempenho, participação ou o que achar relevante para informar o estudante. Os estudantes e professores também podem manter discussões ativas de forma assíncrona e com registros para as próximas turmas por meio dos Fóruns, onde pode-se deixar registradas discussões sobre o curso, dúvidas, entre outras informações que serão moderadas pelos gerenciadores da plataforma, professores e tutores.

Os discentes também podem criar glossários para os alunos, que são uma lista de definições como um dicionário para auxiliar o entendimento de certos conceitos, atividades, siglas, ambientes entre outras informações cujo docente acredita ser relevante para o aprendizado do estudante. Não obstante a este o docente pode criar WIKIS que são uma coleção de páginas web que todos podem contribuir, adicionando conteúdos e informações para enriquecer o conhecimento e compartilhar informações que serão moderadas pelos gerenciadores da plataforma, professores e tutores.

Os professores também podem criar Quizes que se corrigem automaticamente, no caso de respostas de múltipla escolha, e dão o feedback para o estudante assim que finalizam o teste. Outra ferramenta disponível com o mesmo objetivo de avaliação formal, são as lições que permite o professor disponibilizar conteúdo para os alunos de formas mais flexíveis, criando tarefas interativas com vídeos, questões abertas ou de múltipla escolha entre outras funções. Não obstante a isso, a plataforma também disponibiliza atividades de workshops que são uma

atividade avaliativa por pares, onde os alunos enviam suas respostas e elas são distribuídas entre seus colegas para que possam dar notas em uma escala definida pelo professor.

Existem várias formas de acompanhar o progresso do aluno no Moodle, mas focando em um curso específico do ponto de vista do professor o ambiente oferece, primeiramente, o acompanhamento de notas dos alunos, sendo esse essencial para a avaliação formal do estudante, pois não trata somente de notas finais das disciplinas, mas de cada, atividade, quiz, entre outras formas de avaliação formal.

A plataforma disponibiliza também um recurso de competências o qual tem como objetivo descrever o nível de proficiência ou aprendizado de um certo conjunto de conhecimentos relacionados, competências e outros fatores que demonstram habilidades relacionadas à graduação, que o aluno demonstrou. Outra funcionalidade similar é a de Badges, que traz um simbolismo para motivar o estudante, com um selo de competência de uma atividade concluída ou habilidade adquirida, que pode ser adquirida em diferentes estágios do curso.

O Moodle também oferece, tanto para discente quanto para o docente, um quadro de acompanhamento das atividades completadas, onde o professor e o aluno podem acompanhar as atividades enviadas, com um informativo dizendo quais foram e quais não foram entregues, que pode ser marcado manualmente pelo professor ou automaticamente caso o estudante conclua os critérios de aceite.

Comparável com esta ferramenta, mas em um nível maior, o ambiente disponibiliza um quadro de acompanhamento dos cursos, onde tanto o estudante quanto os professores, conseguem visualizar os cursos completos pelo aluno, neste quadro os professores podem marcar manualmente que o aluno finalizou o curso ou caso os estudantes concluam os critérios de aceite, o próprio ambiente sinaliza o curso como finalizado.

Além disso, o Moodle oferece relatórios sobre os alunos, referente aos cursos, as atividades, entre outras interações com a plataforma. Dentre estes relatórios também existem os chamados de LOGS que são registros de atividade dentro da plataforma, ou seja, qualquer interação com a plataforma é registrada neste relatório, sendo registrado desde a abertura de uma página, o login na plataforma, a realização de uma atividade, a tentativa de execução de questionários, visualização de um comentário, criação de um comentário, entre outras interações com a plataforma que permitem segmentar os dados para além dos usuários, como a visualização dos cursos, tipos de eventos, o curso o qual este log se refere, permitindo o cruzamento dos dados entre interação e visualização, são dados ricos que permitem análises diversas porém brutos e espúrios.

A partir de sua versão 3.4 o Moodle oferece também uma ferramenta chamada analytics, que disponibiliza uma detalhada descrição e prescrição dos relatórios para auxiliar os professores no melhor entendimento dos relatórios, são ferramentas simples, porém robustas, que provém ao professor uma visão melhor do ambiente que está inserido, trazendo dados sobre os alunos, risco de evasão e até uma ferramenta de aprendizagem de máquina para predição do risco de evasão.

Mas essas ferramentas são limitadas, no próprio site do Moodle informam que, primeiramente os modelos de aprendizagem de máquina precisam ser treinados com os dados históricos de um curso já existente, para que então possa ser utilizado e avisa que estes modelos necessitam de modelagem para as prioridades educacionais da instituição, deixando claro que não é um modelo genérico que pode ser simplesmente ativado e utilizado em qualquer instituição, necessitando de modelagem, moderação e análise por parte de alguém capacitado.

4.2 Estruturação dos Dados

Com intuito de entender como a informação é construída, é necessário ter o problema como senso diretor, percorrer todo o processo de geração da informação, partindo da captação de dados, transformação, exploração até a informação final. Portanto, partindo dessa premissa, utilizando a linguagem de programação python aliada a biblioteca pandas para manipulação e análise de dados, formulou-se um passo a passo visando extrair informações quali-quantitativas destes dados através de análises preditivas, com modelos de aprendizagem de máquina interpretáveis e auditáveis e análises descritivas como supracitado. Este passo a passo foi organizado a partir de algumas etapas, sendo elas: organização dos dados, limpeza e padronização; de forma a facilitar a manipulação para realizar a análise dos dados descrita anteriormente nos processos denominados "processamento de dados" e "enriquecimento da informação", observando sempre o formato dos dados e colocando o aluno como pivô ao centro da solução para ter dados alinhados a informação, que deve estar alinhada a necessidade do ambiente.

Em complemento utilizando esses dados sobre os discentes, onde está organizado de forma a ser uma base quantitativa das atividades dos discentes na plataforma, uma lista com as atividades mais realizadas entre os discentes foi gerada, para entender quais os passos mais comuns entre os discentes, a partir desta sumarização, o tipo de atividade realizada pode ser definido sendo esse tipo dividido entre atividades de visualização de algum material ou página e atividades de interação com a plataforma como a entrega de um relatório ou resposta de um

questionário como demonstrado na Figura 11.

Figura 11 – Lista de eventos de um curso Moodle

Módulo do curso visualizado	10397
Curso visto	9170
O status da submissão foi visualizado.	2482
Tentativa do questionário visualizada	2139
Conclusão da atividade do curso atualizada	1027
Formulário de submissão visualizado.	729
Tentativa do questionário revisada	687
Questões avaliadas manualmente	421
Usuário recebeu nota	401
Relatório de notas do usuário visualizado	366
Um arquivo foi enviado.	343
O envio foi avaliado.	333
O stado do workflow foi atualizado.	324
Tabela de notas visualizada	296
Submissão criada.	293
Um envio foi submetido.	288
Comentário visualizado	269
Formulário de confirmação de submissão visualizado.	248
Formulário de notas visualizado	232

Fonte: Autoria própria

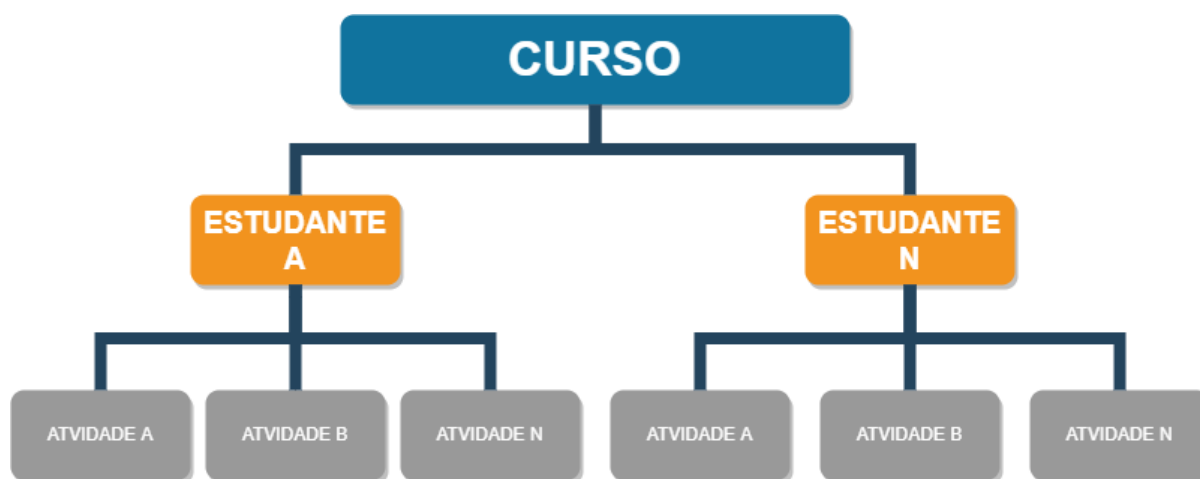
A partir dessa lista é possível segmentar as atividades realizadas no sistema em visualizações e interações, visualizações são um dado de maior volume principalmente por ser uma ferramenta web logo cada navegação dentro das páginas geram registros, já as interações são dados menos volumosos por ser derivado de uma interação direta com alguma atividade proposta ou seja é limitado aos recursos disponíveis e a forma com que o curso foi pensado para ser executado dentro da plataforma, porém são tão importantes quanto as visualizações pois estes registram atuação concreta do estudante na plataforma.

Em complemento às duas abordagens, tanto de visualização quanto de interação viabilizam o estudo, que visa a análise da performance acadêmica do estudante, já que ambos descrevem o comportamento dos discentes na plataforma, de forma a qual padrões podem ser traçados a partir dos dados gerados por cada discente, mas esses dados no formato e estrutura que estão não descrevem a realidade de cada discente, só sumarizam os comportamentos mais realizados.

Para realizar essas transformações uma estrutura foi montada visando descrever o aluno

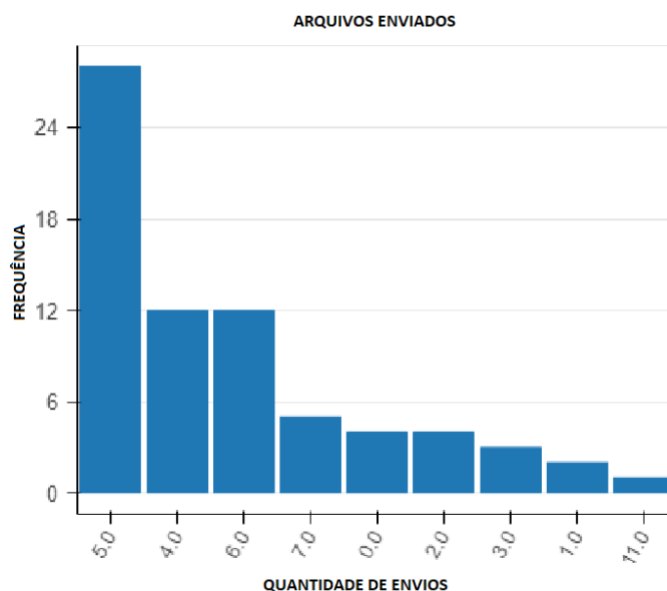
de acordo com os eventos que ele realiza, para isso os dados foram remodelados de forma a estar separado por estudante e contabilizando cada ação do estudante dentro da plataforma, como ilustrado na Figura 12.

Figura 12 – Estrutura dos dados



Fonte: Autoria própria

A partir dessa estrutura temos os dados já mostram de forma quantitativa as ações realizadas pelos estudantes, e isso já cria um contexto limitado pelo escopo para os dados, no sentido de que estes dados estão adequados ao totalização das ações de cada estudante neste curso, podendo assim gerar médias, correlações, entre outras análises sobre cada uma dessas atividades realizadas, como os dados são históricos e não apenas das turma atual, logo uma análise comportamental pode ser realizada para entender por exemplo a frequência com que os alunos submetem um arquivo, como mostra a Figura 13.

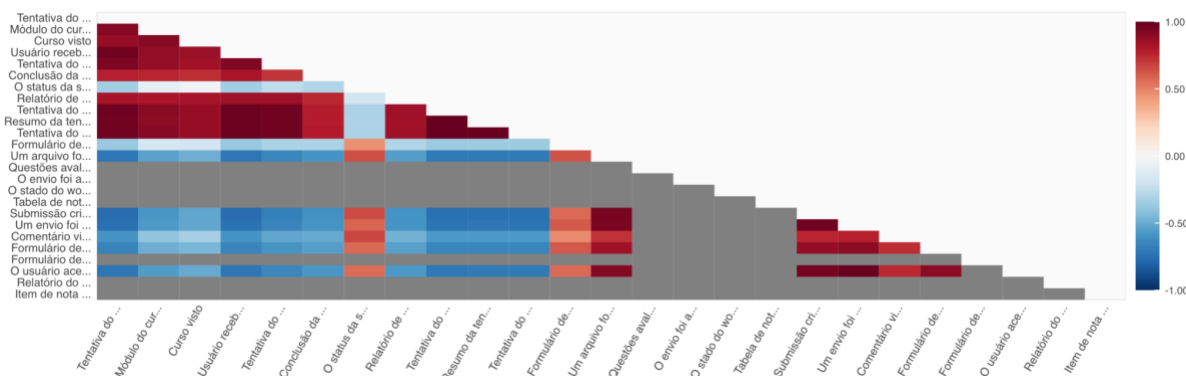
Figura 13 – Gráfico de frequência de arquivos enviados

Fonte: Autoria própria

Em complemento, com tantas informações disponíveis sobre as atividades dos discentes, uma importante tarefa é identificar quais atividades não impactam em nenhuma outra ou têm menor impacto ou qual o comportamento deste dado em meio aos outros, sob a visão da performance, logo para isso os dados históricos das avaliações dos alunos foram adicionado ao conjunto de dados, já que a única forma de avaliação de performance no Moodle é através das notas finais dos estudantes este dado foi escolhido como parâmetro de performance. Para isso a técnica de correlacionar os dados é uma opção de análise descritiva dos dados, como dito por Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009), uma técnica muito utilizada como critério de otimização, para derivar diferentes filtros de redução de ruído nos dados é o coeficiente de correlação de Pearson onde, caso a correlação resulte em um valor menor que zero, significa que é algo negativo uma para a outra, isto é, que as variáveis são inversamente relacionadas, caso positivo significa que a correlação é positiva, isto é, que as variáveis estão diretamente correlacionadas, mas quando a correlação é igual a zero, significa que não é possível determinar qualquer senso de covariação. Então a partir dessa técnica um gráfico de correlação foi gerado como demonstrado pela Figura 14, a fim de determinar se as atividades selecionadas têm alguma correlação e quais deveriam ser descartadas nas próximas análises e com isso foram identificadas atividades desnecessárias para o estudo, que não trariam nenhum benefício a solução do problema e, ao estudá-las mais a fundo, aplicando-as ao modelos entendendo sua origem e avaliando-as, ao removê-las o gráfico ficou muito mais compreensível e as correlações

mais claras como demonstrado na Figura 15.

Figura 14 – Correlação de dados entre eventos

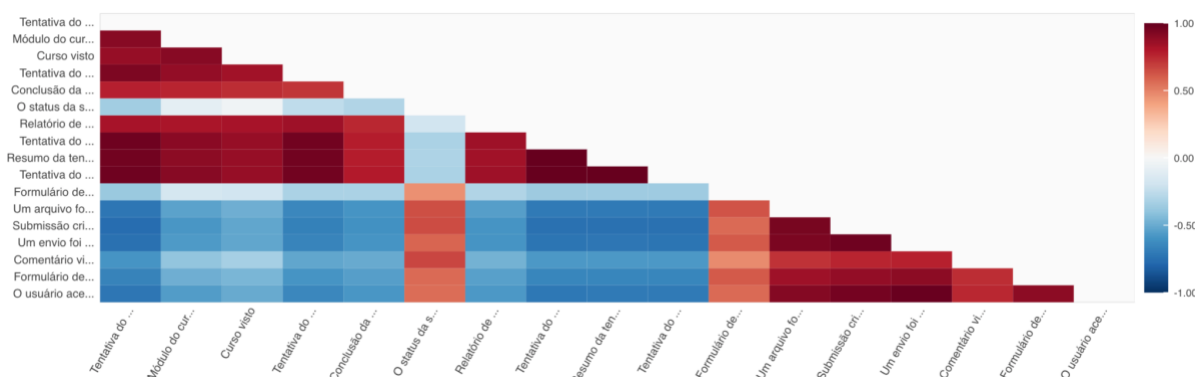


Fonte: Autoria própria

Adicionalmente, a Figura 14 tem os seguintes dados "Tentativa do questionário visualizada", "Módulo do curso visualizado", "Curso visto", "Usuário recebeu nota", "Tentativa do questionário revisado", "Conclusão da atividade do curso atualizada", "O status da submissão foi visualizado", "Relatório de notas do usuário visualizado", "Tentativa do questionário iniciada", "Resumo da tentativa do questionário visualizada", "Tentativa do questionário entregue", "Formulário de submissão visualizado", "Um arquivo foi enviado", "Questões avaliadas manualmente", "O envio foi avaliado", "O estado do workflow foi atualizado", "Tabela de notas visualizada", "Submissão criada", "Um envio foi submetido", "Comentário visualizado", "Formulário de confirmação de submissão visualizado", "Formulário de notas visualizado", "O usuário aceitou o acordo da tarefa", "Relatório do questionário visualizado", "Item de nota atualizado", nesta ordem.

Adicionalmente, a Figura 15 atualizada após a remoção de tem os seguintes dados "Tentativa do questionário visualizada", "Módulo do curso visualizado", "Curso visto", "Usuário recebeu nota", "Tentativa do questionário revisado", "Conclusão da atividade do curso atualizada", "O status da submissão foi visualizado", "Relatório de notas do usuário visualizado", "Tentativa do questionário iniciada", "Resumo da tentativa do questionário visualizada", "Tentativa do questionário entregue", "Formulário de submissão visualizado", "Um arquivo foi enviado", "Submissão criada", "Um envio foi submetido", "Comentário visualizado", "Formulário de confirmação de submissão visualizado", "O usuário aceitou o acordo da tarefa", nesta ordem.

Figura 15 – Correlação de dados entre eventos mais relevantes



Fonte: Autoria própria

Adicionalmente, após a identificação destas variáveis os dados foram removidos para otimização das futuras análises. Com o intuito de trazer mais opções de contexto para a informação, os dados foram separados em três diferentes contextos visando estudá-los e avaliá-los para melhorar a performance e melhorar a assertividade das análises preditivas que serão realizadas, esses contextos então foram separados por, visualizações, onde busca encontrar padrões olhando para onde o estudante caminhou na plataforma, interações, visando buscar padrões a partir das interações realizadas entre aluno e sistema e por fim mesclados em interações e visualizações juntos, para entender se os contextos juntos melhoram enriquecem as análises e a assertividade da informação.

Em destarte, com os dados organizados, limpos e os contextos criados, os dados finalmente estão alinhados com o ambiente do problema e a proposta de solução, sendo um modelo que visa analisar padrões de comportamento em relação às atividades realizadas por cada discente, baseado nos dados históricos do curso, com os contextos alinhados às necessidades informacionais dos atores, visando por meio de análises preditivas e descritivas entender a como está a performance do estudante para promover antecipadas intervenções que melhorem a performance auxiliando na redução de casos de baixa performance, também indiretamente, não como objetivo do modelo, mas podendo aumentar as aprovações, melhorar o aprendizado e evitar a evasão.

4.3 Tomada de Decisão Docente Dirigida por Dados

A tomada de decisão pedagógica é complexa e envolve inúmeras etapas e níveis de conhecimento referente aos estudantes, as quais são realizadas por meio de decisões intuitivas,

analíticas ou deliberativas, as quais possibilitem e devem possibilitar aos tomadores de decisão a incumbência de gerar impacto positivo na experiência da aprendizagem, para tanto, a efetividade dessas informações é gerada por meio de dados educacionais produzidos, entre outras fontes, a partir dos Sistemas de Gerenciamento da Aprendizagem (LMS – Learning Management System) (PAIVA, 2017).

Paiva (2017) aponta alguns exemplos de tomadas de decisões pedagógicas que podem ser melhor definidas a partir da análise de dados educacionais, são: necessidades de intervenções pedagógicas em tempo hábil; análise do desempenho do aluno ou turma; análise do perfil de aluno ou turma; atribuição de professor / tutor a uma turma; definição de metodologias e táticas pedagógicas adequadas; avaliação e adequação da didática; identificação de lacunas de aprendizagem; entre outras.

Este trabalho foca na necessidade de análise de desempenho do aluno, para prover ao professor um maior entendimento do cenário e uma informação útil para auxiliar na melhora da performance do estudante, fornecendo um modelo para a criação de ferramentas que forneçam subsídios informacionais para que o professor se mantenha focado nas decisões e práticas pedagógicas pois como dito por PAIVA, 2017 e DIAS JR, 2017 para ofertar o suporte requerido aos alunos, é necessário que professores, tutores e gestores analisem dados decorrentes das interações dos estudantes e para que isto ocorra, é fundamental possuir ambientes com fluxos informacionais capazes de apoiar o acesso de vários usuários simultaneamente e que promova a interação entre estes usuários e o ambiente.

Porém tais análises demandam habilidades muitas vezes de áreas não relacionadas às áreas de estudo dos professores e desenvolver estas habilidades para realizar estas análises desviaria o foco do professor dos processos pedagógicos, visto que, não necessariamente essas ferramentas se tornam melhores por serem desenvolvidas pelos professores, os professores devem participar do processo e devem fornecer aos profissionais que vão desenvolver tais ferramentas informações relevantes sobre o seu atual problema, para que as ferramentas possam ser desenvolvidas em torno deste problema e se tornar um artefato útil ao professor. Como dito por Araújo, C. A. Á. (2009), utilizar este pensamento como modelo para desenvolvimento da ferramenta informacional, enfatiza as percepções dos usuários em relação à sua própria ausência de conhecimento, mas o usuário pode auxiliar nos passos trilhados para sanar essa ausência de conhecimento, auxiliando na forma como a informação será utilizada para a execução de determinada tarefa ou problema e a partir disso, a informação gerada passa a ser vista como algo útil na perspectiva do sujeito e deixa de ser uma ferramenta misteriosa que apenas gera dados descontextualizados.

Corroborando com isto, na Figura 7 podemos verificar que em uma pesquisa realizada com 63 professores que utilizam sistemas de gestão de aprendizagem, apenas 20.6% dos professores utilizam de dados pedagógicos em todas suas decisões pedagógicas e 66.7% utilizam parcialmente os dados fornecidos pela plataforma. Mesmo com o elevado número de professores que utilizam dados pedagógicos em algumas decisões, verificou-se neste mesmo questionário que 55% dos professores não utilizam nenhuma ferramenta que faça análise desses dados como visto na Figura 7.

Corroborando com o supracitado, a Figura 6 demonstra espaço nos ambientes para trazer ferramentas mais robustas para os professores, já que 36.5% informam que os ambientes não dispõem de ferramentas eficientes para o acompanhamento do desenvolvimento do aluno e 54% informam que o sistema oferece parcialmente, o que indica insatisfação com o atual modelo, devido que apenas 9.5% afirmam que o sistema oferece ferramentas eficientes para o acompanhamento do desenvolvimento do estudante.

Tal insatisfação está amplamente ligada à falta de ferramentas contextualizadas, que transformem os dados do sistema em informações pertinentes e úteis, alinhadas com as demandas informacionais dos professores, sem isso torna-se extremamente difícil a tomada de decisão neste ambiente, principalmente pelo fato que os alunos geram uma grande quantidade de dados dentro do sistema os quais não são possíveis de se analisar sem uma ferramenta robusta e o devido conhecimento técnico, como dito por PAIVA (2017) em meio a grande quantidade de dados educacionais produzidos, o seu processamento manual fica impossibilitado.

Adicionalmente, o autor ressalta que não é recomendado que as tomadas de decisões sejam realizadas unicamente por softwares educacionais, corroborando com o supracitado, onde se explicita a necessidade de ferramentas que sejam desenvolvidas para auxiliar a tomada de decisão, trazendo para os professores os insumos necessários para isso de acordo com cada necessidade.

Neste sentido o professor necessita de mais informações úteis e contextualizadas, mais ferramentas analíticas para todos os cenários, este trabalho foca em uma ferramenta analítica que a partir dos dados históricos de outras turmas, a fim de prover uma análise preditiva e contextualizada sobre a performance dos estudante, contendo a origem dos dados, a fonte de informação e quais padrões estão sendo analisados, uma ferramenta complexa e que provê uma informação útil para o professor, guiando seu entendimento sobre os dados, auxiliando a tomada de decisão.

4.4 Transformação dos Dados em Informação

De acordo com o autor Ackoff, R. L. (1989) informações são uma quantidade de dados úteis e em seu trabalho exemplifica sua fala, onde diz que os coletores do censo recolhem dados, o escritório do censo, no específico caso brasileiro o IBGE, processa esses dados, convertendo os dados em informação que é representada em números tabelas e publicadas em resumos estatísticos.

Não obstante a isto no presente cenário podemos dizer que o sistema de gerenciamento do aprendizado recolhe os dados dos estudantes e armazena estes dados em um banco de dados, o próprio sistema em alguns casos, disponibiliza as transformações necessárias para que informações prontas possam ser apresentadas ao professor de forma visual simplificada pronta para utilização.

Porém existem dados os quais o sistema não realiza os devidos tratamentos para transformá-los em informação, como os dados de logs, que por mais que estejam disponíveis em tabelas, reports, não são úteis ao professor, não caracterizando informação. Diante disto é importante ressaltar que em contextos diferentes e a partir de visões diferentes os mesmos dado pode ou não ser uma informação, para um estatístico um gráfico de dispersão ou diagrama de dispersão tem muito mais informação que para um leigo no assunto.

Para que o sistema use esses dados e transforme eles em dados são necessárias ferramentas que consigam processar esses dados e gerar análises, descrições, predições, entre outras técnicas que sejam úteis para os usuários. Mas para gerar tais informações, é necessário entendimento do dado, seu tipo, sua origem, seu objetivo, o contexto, para então selecionar as ferramentas corretas para a realização das análises. No presente caso dos logs, os dados são dados originários da interação do usuário com o sistema, gerando um histórico de processos realizados na plataforma de cada usuário e até do próprio sistema.

A realização dessas transformações dos logs pode ser feitas de inúmeras maneiras, todas dependem do contexto e do objetivo da análise, o primeiro passo para um dado se transformar em informação é uma demanda, um cenário onde o usuário precisa de um informação o qual não está explícita nos dados. Para retirar essas informações implícitas, é necessário o uso de ferramentas e técnicas como análise descritiva, análise exploratória, análise preditiva, prescritiva ou diagnóstica, cada uma delas tem seu objetivo e essas técnicas podem ser aplicadas utilizando linguagens de programação como python, R, java, que podem gerar ferramentas mais robustas e personalizadas e também empregar modelos robustos de aprendizagem de máquina e inteligência artificial, ou para visualizações e análises mais descritivas os softwares como

Power BI, tableau, Excel, entre outros, cada abordagem com a sua devida aplicação e complexidade.

Estas ferramentas por si só não são suficientes para transformar um dado em informação, pois como dito pelo autor Ackoff, R. L. (1989), a diferença entre o dado e a informação não é estrutural e sim funcional, em destarte alterar a estrutura dos dados não necessariamente os transforma em informação, mas ao transformar os dados de forma a que se adeque a função que ele precisa exercer, então caso úteis, os dados se transformam em informações.

No contexto educacional de análise de performance do estudante, podemos dizer que os dados dos logs não são informações, são muitos dados e muito dispersos, mas caso transformados e reestruturados com auxílio computacional através do uso e combinações de técnicas analíticas, transformações essas dirigidas pela função ao qual esses dados precisam cumprir, então a informação é gerada.

4.5 Aprendizado de Máquina Como Ferramenta de Análise Preditiva no Contexto da Educação a Distância

As Tecnologias da Informação e Comunicação estão presentes nos mais diversos segmentos da sociedade moderna e a sua presença tem propiciado uma ampla distribuição e acesso, em tempo real, de informações. Nesse sentido, tem-se observado a aplicação de esforços no desenvolvimento de recursos automatizados, construídos com o objetivo de favorecer a utilização de artefatos informacionais e tornar a informação mais acessível e tais recursos têm ampla utilização no contexto educacional, principalmente em ambientes digitais, tal qual o ambiente de educação a distância.

Nesse contexto, Araújo (2009) destaca novas possibilidades no campo de representação da informação ao considerar a aplicação de ontologias, web semântica, folksonomia, aspectos relacionados à interação humano computador, a Inteligência Artificial (IA), e outras, como temáticas relevantes. Imerso nesse cenário, Neves (2020) afirma que a Ciência da Informação necessita discutir, verticalmente, a respeito da Inteligência Artificial, em especial a Computação Cognitiva, ao considerar que o uso de dispositivos inteligentes vem avançando paulatinamente nos espaços de construção do conhecimento, assim como tem proporcionado novas formas de interação entre sujeitos, o que é um dos grandes desafios da educação a distância, visto que a modalidade é jovem em comparação com o ensino presencial há poucas formas ainda de interação entre os sujeitos.

Adicionalmente, para Neves (2020) e Mueller e Massaron (2019) a computação

cognitiva, também conhecida como aprendizado de máquina (Machine Learning - ML), é o que se tem de mais avançado em tecnologia computacional e apresenta componentes da engenharia e científicos.

Ao considerar a engenharia Mueller e Massaron (2019) argumenta que é possível quantificar suas tecnologias com base em teorias fundamentadas. No contexto científico, possibilita que sejam testados novos conceitos e ideias e permite realizar descobertas, muitas vezes não óbvias, em sistemas informacionais, de forma abrangente e ágil.

Adicionalmente, é nesse contexto que a Ciência da Informação deve se fazer presente e auxiliar na descoberta de novas informações, em ambientes como o ambiente de educação a distância, onde a falta de informação sobre os alunos é um recorrente problema para os professores devido a falta de interação com os mesmos e a falta de técnicas para obter informações por ser uma área jovem ainda em desenvolvimento.

Em complemento, Navarro, Coneglian e Segundo (2018) identificam a área de aprendizado de máquina como um subconjunto da Inteligência Artificial, na qual são utilizados softwares capazes de aprender de forma autônoma, a partir de um conjunto de dados e informações.

Monard e Baranauskas (2003) evidenciaram que o desenvolvimento de recursos computacionais, assim como sistemas capazes de aprender automaticamente, firmam o objetivo desta área. Neste sentido, o ambiente de educação a distância, que em geral é um ambiente em que se mantém em um sistema de gerenciamento de aprendizagem, cujo é um software que produz diariamente centenas de dados sobre os estudantes, são uma grande fonte de dados que podem conter informações ricas dentro deles aguardando para serem exploradas.

Adicionalmente, o campo aprendizado de máquina teve sua origem nos anos de 1950, após a criação do teste de Turing, com os primeiros softwares que conseguiram aprender a partir de dados de entrada e hoje encontra-se presente em diversos segmentos, por exemplo a capacidade de falar com dispositivos e obter resultados a partir de execuções de acordo com o interesse do usuário ou obter recomendações com base em históricos de interações realizadas e obtenção de dados de fontes diversas (NAVARRO; CONEGLIAN; SEGUNDO, 2018; MUELLER; MASSARON, 2019).

Para Russel, S., e Norvig, P. (2013) o aprendizado de máquina é utilizado para se adaptar a novas circunstâncias e detectar e extrapolar padrões. A sua base é a matemática, na qual, os algoritmos interpretam dados e os processam de maneiras específicas para gerar novos dados baseadas nos padrões encontrados.

Como supracitado os algoritmos de aprendizagem de máquina são adaptáveis de acordo

com os dados nele inserido, em destarte, o mesmo algoritmo pode ser utilizado em vários contextos, já que se os dados inseridos apresentarem padrões diferentes, em seu treinamento o modelo irá detectar e extrapolar os padrões nestes dados encontrados.

Neste sentido, o ambiente educacional, que é um ambiente de inúmeros contextos, visto que cada área de estudo é diferente, traz um cenário onde pode ser amplamente explorado por esse tipo de ferramenta, principalmente em contextos como o da educação a distância, contexto que carece de informação e excede dados de inúmeras naturezas, que caso sejam explorados de forma convencional por um ser humano levaria um tempo inviável para ser processado. É de suma importância ressaltar que caso o cenário permaneça, mas o objetivo mude, o algoritmo utilizado precisa ser reavaliado para atender a nova demanda.

Em complemento, a pesquisa de aprendizagem de máquina avança em direção ao desenvolvimento de sistemas computacionais que melhoram sua performance conforme adquirem experiência, como dito por MITCHELL, Tom et al.

Nesse sentido os algoritmos devem ser selecionados de acordo com o escopo visando sempre a melhor informação, uma análise preditiva deve ser feita utilizando dados históricos e atualizada de acordo com os novos dados a fim de corrigir e dar mais experiência para o modelo, porém caso o objetivo venha a mudar para uma análise descritiva, então o modelo precisa ser repensado, recriado no sentido que a experiência que o modelo adquiriu não será mais útil. Por mais que a aprendizagem de máquina tenha em sua base o aprendizado baseado na forma como os humanos aprendem, os algoritmos são ferramentas específicas que visam resolver um problema somente por vez e ser o mais eficiente possível nele, não acumulando experiências diferentes e nunca um algoritmo sozinho torna-se multifuncional.

Devido a essas características, algoritmos de aprendizagem de máquina são amplamente utilizados para resolver problemas extremamente específicos os quais um ser humano levaria muito tempo para resolver sozinho, se tornando uma ótima ferramenta para operacionalizar e automatizar alguns trabalhos, principalmente trabalhos que precisam reconhecer padrões. Esta característica corrobora com a ideia supracitada de que para gerar informação é necessário transformar os dados de forma a que se adeque a função que ele precisa exercer para que então sejam úteis como dito por Ackoff, R. L. (1989). Neste sentido, essas características se enquadram muito bem no ambiente de educação a distância, pois como supracitado, é um ambiente que excede dados e esses dados seriam muito trabalho para um agente humano sozinho processar.

Adicionalmente, os algoritmos de aprendizagem de máquina são uma ponte entre o dado e a informação, já que para usá-lo é necessário entender o problema e a função que esse dado

vai exercer, o que contextualiza o desenvolvedor a ter uma visão da informação que ele precisa, posto isso os dados precisam ser estruturados a fim de ter o formato correto para responder ao problema corretamente, precisando estar adequado ao problema e ao algoritmo para que realize sua função, então os dados começam a tomar forma para se tornar informação e por fim quando ingeridos pelo algoritmo de aprendizagem de máquina, que está nesse processo para realizar um trabalho especializado de reconhecer dentro desses dados os padrões e aprender com eles, para então trazer uma forma de reconhecer o comportamento desse dado e descrever uma situação para o usuário, no caso deste trabalho descrever o comportamento do aluno no ambiente moodle visando sua performance.

Em complemento um algoritmo de aprendizagem de máquina com fim preditivo, teria sua funcionalidade descrita, de modo a estar no sistema, com intuito único e específico de reconhecer os padrões de comportamento dos dados inseridos, cujo o qual já foram previamente selecionados e padronizados, provendo, de acordo com os dados anteriores e a partir de modelos estatísticos de predição, uma informação sobre o um futuro dado que poderá ser gerado. Em suma, cria-se um ambiente, selecionando os dados os quais deseja observar, um contexto o qual deseja entender, no caso um contexto futuro, sob uma perspectiva a qual deseja obter informações previamente. Toda essa ambientação, torna o dado de saída do modelo um dado rico, pois pode ser descrito pelo do processo que foi gerado e atender a sua função, caso seja útil ao usuário torna-se então informação de fato.

Adicionalmente, a análise preditiva como dito por Eckerson, W. W. (2007) é uma seção de análises dentro do conjunto de tecnologias de Business Intelligence (BI) que revela padrões e relações dentro de um grande volume de dados que pode ser usado para prever cenários e eventos, ao contrário de outras análises a análise preditiva é uma visão futura que usando o passado uma possível prévia do futuro.

Nesse sentido, este tipo de análise se encaixa perfeitamente com a tarefa específica e trabalhosa de analisar os padrões do estudante para inferir algo, como no caso deste trabalho a performance do estudante, para este caso a aprendizagem de máquina pode melhorar e automatizar, pois esse tipo de análise contém todos os fatores que a aprendizagem de máquina aborda, um grande volume de dados e uma tarefa específica e o reconhecimento de padrões.

Adicionalmente, como dito pelos autores Samanpour, A.R., Ruegenberg, A., Ahlers, R. (2018) a aprendizagem de máquina, oferece previsões automatizadas a partir de dados computacionais e as converte em um formato útil para os seres humanos.

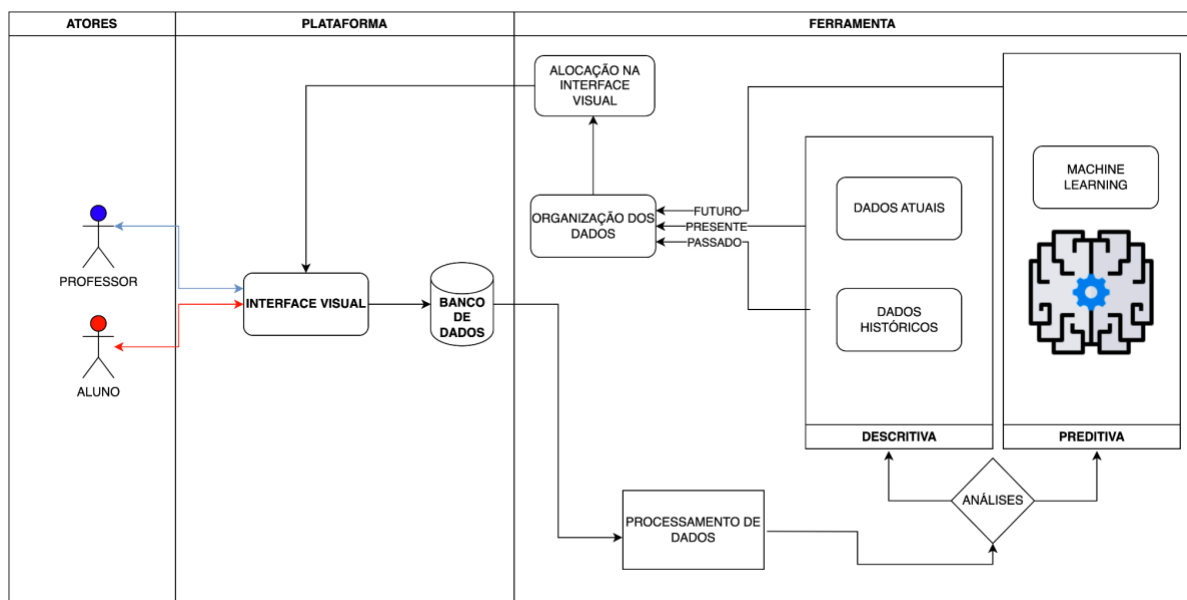
Modelos de aprendizagem de máquina supervisionados, por exemplo, fazem previsões baseadas nos chamados valores empíricos ou probabilidades. No entanto, o processo de

aprendizagem é crucial para que um sistema seja capaz de fazer previsões, este processo sempre segue a seguinte sequência começa, o modelo inicia em branco e este é treinado usando dados de amostra, a partir desses dados ele realiza cálculos de uma determinada tarefa, e o resultado é comparado com o resultado desejado, a diferença alvo comparado ao valor real é então retornada ao modelo e recalculada de acordo com um procedimento adequado (por exemplo, o gradient descent do algoritmo de backpropagation) para reduzir o erro a um mínimo pré-definido.

Em complemento, a aprendizagem de máquina, neste trabalho, no processo de geração da informação, encontra-se no centro do processo, sendo o último gerador de dados antes do contato do usuário com a informação, como supracitado, no trabalho dos autores Samanpour, A.R., Ruegenberg, A., Ahlers, R. (2018) gerando dados em formatos úteis, estando somente atrás da organização e apresentação dos dados, como demonstrado de forma simplificada na Figura 16.

Utilizando o modelo proposto por este trabalho como exemplo, modelo performi que é descrito na próxima seção, o modelo de aprendizagem de máquina recebe os dados processados e com ele faz seu treinamento, por meio dos dados inseridos nele definirá a melhor regra para avaliar a performance do estudante, ou seja, cursos diferentes que têm perfis de comportamento diferentes, geram resultados diferentes, o modelo aprende com os dados e fornece a melhor solução para aquele contexto naquele ambiente.

Figura 16 – Aprendizagem de máquina no modelo performi



Fonte: Autoria própria

A partir dessas definições, a fim de exemplificar uma escolha de algoritmo de aprendizagem de máquina, com o objetivo de selecionar os melhores algoritmos de aprendizagem de máquina, foram selecionados oito algoritmos para comparar sua performance na geração da informação. Os algoritmos foram selecionados a partir do escopo desse trabalho que utiliza algoritmos classificadores, supervisionados e interpretáveis consolidados que suportam múltiplas entradas de dados e apenas uma saída porém com múltiplas classes, com uma ótima performance que se adaptam bem a estrutura de dados proposta como entrada e saída, estes algoritmos são One vs rest (E E Wu, W., Gao, X., & Gao, S. (2006)), random forest (Breiman, L. (2001)), logistic regression(Cramer, J. S. (2003)), k-nearest neighbor (Fix, E., & Hodges Jr, J. L. (1952)), ridge (Saunders, C., Gammerman, A., & Vovk, V. (1998)), decision tree (Quinlan, J. R. (1986)), gaussian naive bayes (Bayes, T. (1968)), XGBoost (Chen, T., & Guestrin, C. (2015)), support vector (Cortes, C., Vapnik, V (1995)) da biblioteca python scikit-learn.

Em complementos, os algoritmos funcionam de forma parecida, suas maiores diferenças estão na forma como trabalham os dados, mas o objetivo é o mesmo, sendo assim sua diferença é a técnica de aprendizado por trás do modelo e sua forma de avaliação durante o treinamento. Todos os modelos citados acima, aceitam a mesma estrutura proposta a sua avaliação, sendo os dados inseridos os mesmos e todos produzem os mesmos tipos de dado, sendo esse dado um número dentro do conjunto de 4 classes variando de 0 a 3, onde são interpretados como as notas "A", "B", "C" e "D".

Adicionalmente, os treinamentos realizados em todos os algoritmos passando por validação cruzada para evitar enviesamento, validação essa que auxilia na avaliação da capacidade de generalização do modelo utilizando o método k-fold. Em complemento, os dados são separados em clusters de treinamento e avaliação de forma randômica em K subconjuntos onde K é definido previamente com aproximadamente a mesma quantidade de amostras em cada um deles.

A partir dessas definições, os clusters são iterados, como demonstrado na Figura 17, e a cada iteração o subconjunto utilizado como teste é avaliado por métricas próprias para modelos de aprendizagem de máquina classificatórios multi-classes, sendo as escolhidas: Acurácia balanceada (do inglês balanced accuracy) e F1 ponderado (do inglês f1 weighted); disponíveis também na biblioteca scikit learn.

Adicionalmente, estas métricas foram selecionadas visando um melhor entendimento da performance do modelo, logo, por se tratar de um modelo com dados desbalanceados, visto que existem menos alunos com notas mais baixas como D comparado a média.

Em complemento, a fim de garantir que cada subconjunto será testado em algum momento da avaliação. A partir dos resultados obtidos a avaliação dos modelos será realizada utilizando as métricas de qualidade de modelo, como acurácia, precisão e métricas mistas como F-score (Piotroski, J. D. (2000)).

Figura 17 – Cross-Validation

TESTE	TREINO		
TREINO	TESTE	TREINO	
TREINO		TESTE	TREINO
TREINO			TESTE
TREINO			TESTE

Fonte: Autoria própria

Adicionalmente, com essas métricas avaliativas dos modelos podemos avaliar o comportamento de cada modelo com os dados, não somente os erros e acertos ou a proximidade dos dados com a realidade, mas também a consistência do modelo, em relação aos resultados. Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019) afirma que incluindo essas métricas que auxiliam na medição da interpretabilidade e nas definições de contexto de diferentes casos de uso, indicando em qual direção os modelos devem seguir.

Em complemento, com os dados de avaliação dos modelos, é possível definir a viabilidade dos modelos de aprendizagem de máquina quanto a seus acertos e a qualidade do dado produzido, em referência a precisão e a confiabilidade desse dado em relação aos dados históricos, onde a partir disso o modelo mais performático é selecionado como modelo o qual será utilizado na ferramenta para gerar as análises neste trabalho podendo em casos de aplicação deste modelo a outros casos, os modelos, constantemente ser avaliados de acordo com a mudança de dados de entrada, a partir do processo de descarte e recuperação de dado do ciclo de vida dos dados, a fim de sempre manter um modelo de aprendizagem de máquina saudável e com maior tempo de vida sem perder precisão e performance.

4.6 Artefatos Informacionais e Camada de Visualização no Contexto Educacional

Todo este processo supracitado de transformação dos dados, análises preditivas e

descritivas, geram artefatos informacionais que no caso deste trabalho são, de natureza estruturada, numérica e contínua no caso das análises descritivas e um artefato de natureza estruturada, categórica e ordinal no caso da análise preditiva. Principalmente o dado gerado pelo modelo de aprendizagem de máquina da análise preditiva, que se trata de uma classificação da possível nota do estudante, esse dado não pode ser apresentado ao professor como um dado sem uma descrição para contextualizá-lo, pois nesse caso seria um dado não útil e pobre. Como dito por Rendón Rojas (2005) a estrutura da informação é entendida como a união de dois extremos, os elementos objetivos, ou seja, os dados de fato, aliado aos elementos subjetivos, ou seja as estruturas interpretativas do sujeito.

Em complemento, no contexto educacional onde levamos em consideração que o ambiente informacional é o sistema de gerenciamento de aprendizado e o sujeito é o professor, que está analisando o contexto do estudante, todo o processo de processamento de dados pode ser caracterizado como uma forma de gerar elementos para enriquecer o processo informacional da tomada de decisão pedagógica, no âmbito deste trabalho, avaliação de performance, esses elementos subjetivos precisam ainda se aliar as estruturas interpretativas do sujeito, para a partir dessa união se tornar informação. Para auxiliar na interpretação dos elementos objetivos uma camada de visualização que contextualiza o sujeito pode ser estruturada como um guia para o agente informacional consumidor da informação e auxiliar na tomada de decisão.

Adicionalmente, uma forma de guiar o sujeito a entender e interpretar os elementos objetivos de forma mais clara é por meio de referências conhecidas pelo agente consumidor da informação, como dito por Araújo, C. A. Á. (2018), os sujeitos agem diante de outros sujeitos, com ações referenciadas, e tais ações estão ligadas aos contextos concretos que acontecem. Com este contexto entende-se que o dado gerado pelo modelo de aprendizagem de máquina por si só não é um dado de fácil referência, ao ver este dado por si só não trará naturalmente uma boa interpretabilidade intrínseca, este dado é classificado como um contexto concreto e precisa se ligar a uma ação referenciada, para isso, no âmbito de estrutura interpretativa, precisamos de mais elementos objetivos presentes no ambiente em que o agente consumidor está habituado e entendi, trazendo a referência para o sujeito, ou sua interpretação pode vir a ser superficial e o dado não será útil não chegando a se tornar uma informação, tornando a ferramenta sem utilidade ou o tornando seu uso operacionalizado sem um objetivo de promover o progresso e o pensamento crítico do professor.

Em complemento, como supracitado para que a ferramenta tenha uma fácil interpretabilidade deve conter uma estrutura que contextualize os dados e traga referências para o sujeito, logo as informações descritivas sobre o processo que gerou essa informação devem

ser demonstrados ao professor, como em uma tabela ou gráfico, como demonstrado na Figura 18, demonstrando os parâmetros atuais do estudante, comparado com a média histórica das outras turmas, para que o professor ao ver esses dados, o qual ele tem conhecimento da origem, possa discernir sobre a informação preditiva que ele vê sobre em seguida e possa referenciar esses dados conhecidos entendendo a situação do estudante, a partir disso utilizar essa análise preditiva com base em dados históricos para auxiliar a decisão e conseguir discernir sobre a margem de melhora, ou até se é necessária uma decisão interventoria antecipada, uma mudança nas abordagens avaliativas entre outras decisões que podem ser pautadas nessas informações.

Figura 18 – Camada de apresentação



Fonte: Autoria própria

Adicionalmente, a Figura 18 traz uma exemplificação de um dashboard interativo criado pela ferramenta powerBI onde traz elementos objetivos para que o professor crie referências e tenha de forma visual elementos de fácil compreensão sobre os dados históricos da de uma turma, os atuais dados de um estudante, sua nota atual, que ainda não foi consolidada por uma avaliação formal finalizada, para ilustrar o estudante que ainda não realizou exames formais de

avaliação e uma possível nota que traz de forma simples e intuitiva um dado complexo que fora construído utilizando como subsídio dos outros dados mostrados na tela.

No próximo capítulo será apresentado o modelo de processamento da informação Performi (Performance estudantil dirigida por processamento informacional), seus módulos e processos informacionais, visando a gerar subsídios informacionais sobre a performance do discente para auxiliar na tomada de decisão do professor.

5 MODELO DE PROCESSAMENTO DA INFORMAÇÃO PARA AVALIAÇÃO DE PERFORMANCE DE DISCENTES EM AMBIENTE LMS

O processo de tomada de decisão em ambientes como o ambiente educacional, é desafiador, pois demandam uma gama de conhecimentos sobre: o cenário do curso, os alunos, o ambiente de aprendizado, o material, o desenvolvimento da turma, entre outras informações que precisam estar consolidadas e disponíveis ao tomador de decisão. Todos esses ambientes, junto a dinamicidade dos ambientes de aprendizado frente a grande gama de áreas de estudos e a diferença entre cada turma, traz a necessidade do processamento dessas informações como forma de auxiliar o processo de tomada de decisão.

Neste capítulo é apresentado o modelo Performi (Performance estudantil dirigida por processamento informacional), um modelo de processamento da informação para auxílio na avaliação de performance educacional (Figura 19). Este modelo visa entregar informações processadas e contextualizadas aos tomadores de decisão de ambiente de educação a distância como moodle. Para tal o modelo apresenta módulos de: aquisição, exploração, processamento e limpeza dos dados, alinhamento dos dados com a necessidade informacional, análises preditivas e descritivas dos dados, a escolha e uso dos modelos de aprendizagem de máquina, as métricas de qualidade do modelo, organização e disponibilização das informações contextualizadas e os artefatos informacionais gerados.

5.1 Introdução ao Modelo Performi

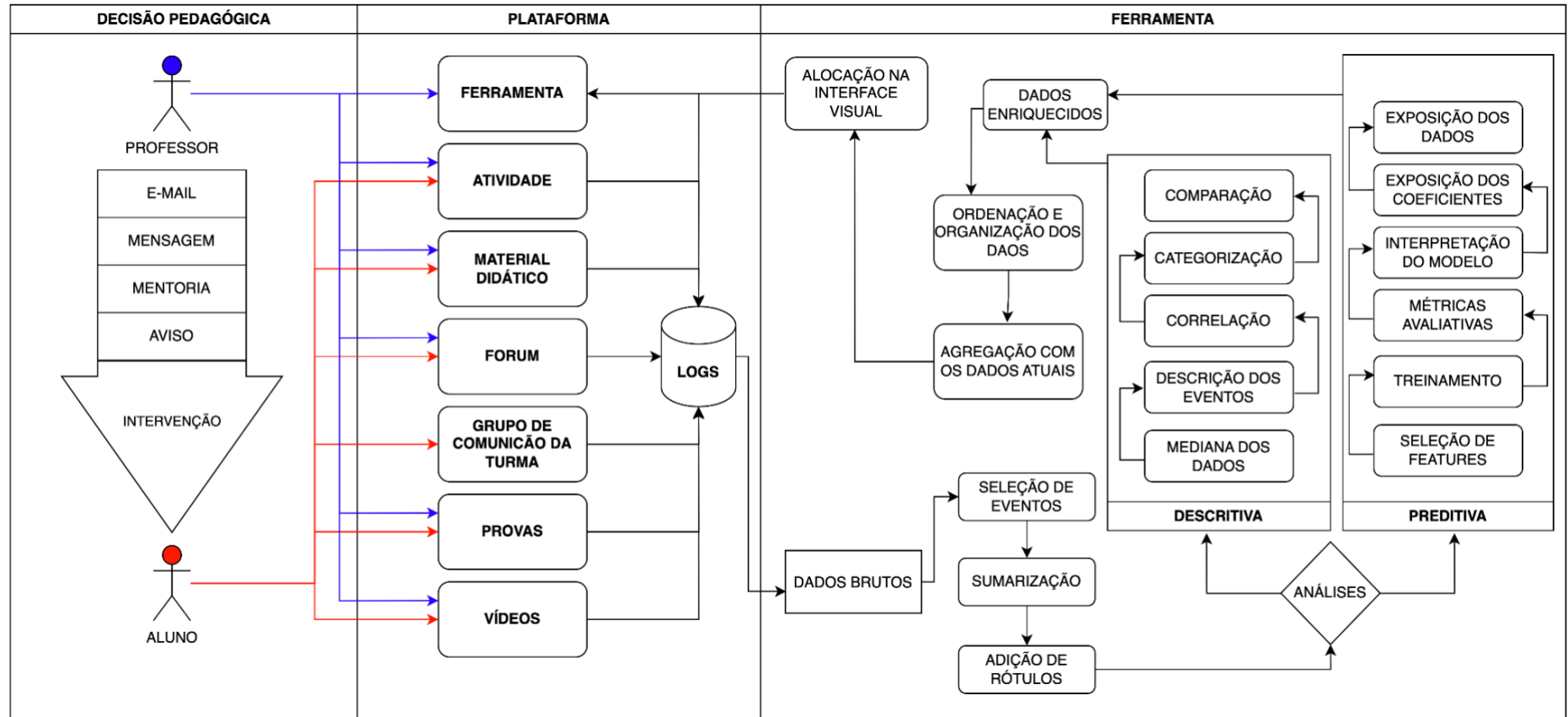
O modelo performi visa apoiar o processo decisório de tomada de decisão pedagógica a partir da avaliação de performance estudantil, onde se liga de forma cíclica a dois atores de um processo informacional, os estudantes e os professores. Em complemento, o modelo visa propor um processo a ser acoplado ao processo decisório como uma forma de enriquecer as informações necessárias para a tomada de decisão. Provendo informações sobre o ambiente e sobre o ator estudante para auxiliar na tomada de decisão do ator professor, como descrito na Figura 19.

Adicionalmente, o modelo consiste em seis módulos, nomeados: “produção dos dados”, “processamento de dados”, “enriquecimento da informação”, “Organização da informação”, “Apresentação da informação”, “tomada de decisão”.

Em complemento, estes processos podem ser vistos contendo três tipos de agentes informacionais, sendo o aluno e o professor os agentes humanos que interferem no

comportamento e na maneira a qual o sistema funciona diretamente, em segundo plano o sistema que pode ser visto como agente computacional, cujo a função é realizar tarefas operacionais a partir dos dados e inferências gerados pelos agentes humanos sob a diretriz do terceiro agente, a decisão, cujo o sistema é apoiador, criando um ciclo onde o agente aluno gera dados, o sistema processa os dados gera a informação guiado pela decisão, o professor consome essa informação interpreta e toma sua decisão, que retorna ao aluno e o ciclo se repete como demonstrado na Figura 19. Ressalta-se o caráter especial do último processo nomeado "tomada de decisão", onde este é tido como um processo cujo a execução é dada exclusivamente pelo ator professor sem interferência direta de nenhum agente não humano presente no modelo, apesar da decisão ser considerada um agente do modelo, é um agente que utiliza o ator professor como motor para seu funcionamento.

Figura 19 – Diagrama completo de funcionamento do modelo Performi



Fonte: Autoria própria.

Especificamente, o modelo tem como objetivo orientar o desenvolvimento de sistemas de apoio à tomada de decisão pedagógica, dedicados a avaliação de performance dos alunos no sistema moodle, especialmente onde há dificuldade na aquisição de informação para subsidiar a decisão.

Dentro das principais capacidades do modelo, estão a capacidade de:

- Extrair informação histórica, atual e preditiva de dados espúrios.
- Suportar que professores e tomadores de decisão possam usufruir de informações contextualizadas sobre a performance dos alunos.
- Adaptabilidade aos contextos educacionais distintos de cada curso.

O modelo compreende:

- Um processo completo de avaliação de performance estudantil, sobre a situação histórica do curso e a atual e futura do discente, com as etapas de aquisição, exploração, processamento, alinhamentos com a necessidade informacional, análises preditivas e descritivas, organização e disponibilização das informações contextualizadas a partir de dados espúrios, provenientes da base de logs.
- Acompanhamento da performance do aluno.
- Análises preditivas para auxílio em intervenções antecipadas.
- Análises descritivas para contextualização da informação.
- Dados históricos relativos ao curso.

Nas próximas subseções serão detalhadas as características gerais do modelo.

5.1.1 Processo Completo de Avaliação de Performance Estudantil

O modelo suporta ambientes educacionais reais e complexos que geram dados continuamente, podendo ser aplicado mais de uma vez de forma paralela, a fim de se adaptar aos inúmeros contextos de cursos heterogêneos, adaptando-se constantemente aos novos padrões baseando-se em ciclos de melhoria contínua, promovendo uma melhora contínua na informação produzida.

Um processo completo de avaliação de performance estudantil acomoda tais ciclos de forma a prover subsídios ao modelo para que este possa, adquirir, processar e representar a

informação, que pode vir a impactar a tomada de decisão estudantil.

O modelo se dedica a suportar a tomada de decisão fornecendo subsídios informacionais, em operação cíclica e auditável de aquisição e processamento de dados. Para isso os dados espúrios contidos nas bases de logs do ambiente moodle é explorado de forma se tornar uma base de dados com foco nos padrões de comportamento do aluno. Este módulo é descrito com mais detalhes na seção 5.2.1.

Em continuidade às informações processadas são enviadas ao módulo de enriquecimento da informação, que está relacionado às análises e inferências realizadas pelo modelo, para gerar dados que serão contextualizados a partir dos módulos de contextualização e apresentação da informação e oferecidas ao tomador de decisão para que este consuma a informação e por fim realize a tomada de decisão. Estes módulos são descritos com mais detalhes nas seções 5.2.2, 5.2.3, 5.2.4, 5.2.5 e 5.2.6.

As informações podem mudar a qualquer instante, porém o cenário educacional torna-se crítico não pela sua alta demanda de informação em tempo real, mas pelo impacto da tomada de decisão frente a sociedade. Dito isto o modelo visa uma análise cíclica sob demanda, onde os dados de entrada podem ser continuamente inseridos e serão absorvidos assincronamente, enquanto novas análises podem ser disponibilizadas, atualizadas e mantidas separadas do modelo.

5.1.2 Operação Cíclica e Auditável

A operação do modelo Performi se baseia no conceito cíclico de reavaliação do próprio modelo e da qualidade dos dados produzidos. Semelhante ao ciclo de Boyd (Plehn, 2000), o modelo Performi fornece módulos auditáveis e incentiva a audição constante dos módulos, a fim de fornecer dados para sustentar mudanças na implementação do modelo, ou seja, identificar melhorias em cada parte do processo informacional.

Adicionalmente, apesar de oferecer módulos auditáveis, os módulos do modelo Performi tem dados interdependentes, ou seja, um módulo depende dos dados produzidos e padronizados pelos módulos anteriores e os dados produzidos podem interferir futuramente nos processos anteriores, logo apesar de cada módulo ter a possibilidade de ser auditado e melhorado individualmente, a modificação de um módulo acarreta em mudanças nos módulos subsequentes, tornando o modelo reativo a mudanças em qualquer parte de seu processo.

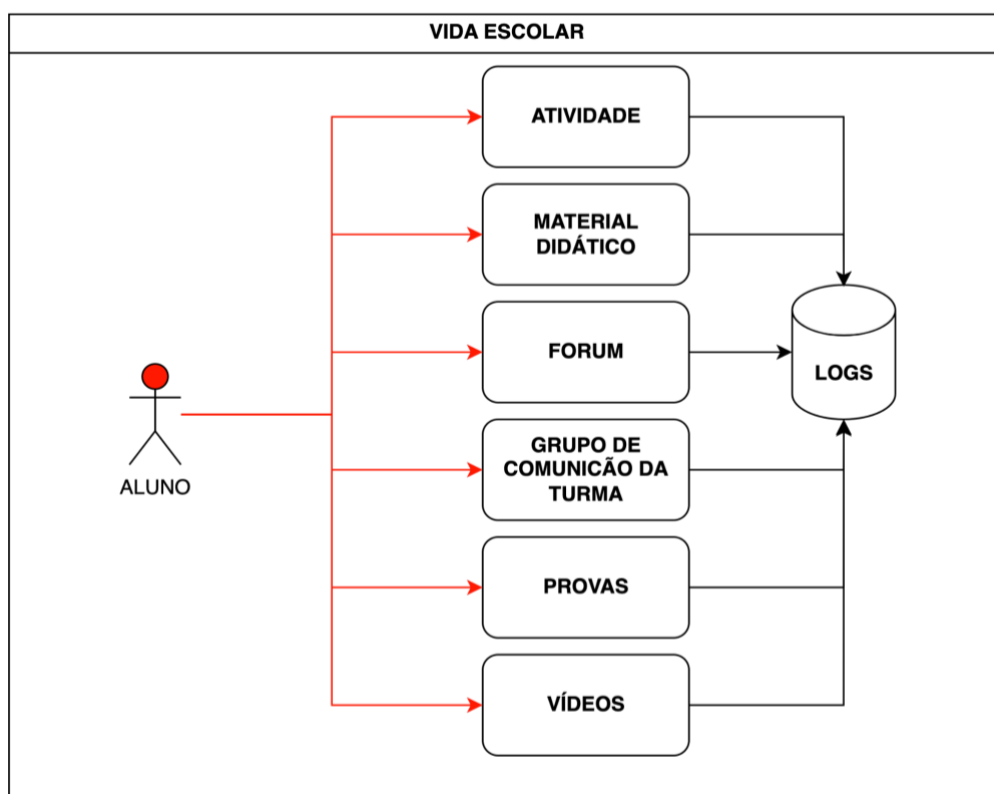
5.2 Processos Internos do Modelo Performi

Nesta seção os processos internos do modelo serão descritos, sendo estes separados em seis módulos: “produção dos dados”, “processamento de dados”, “enriquecimento da informação”, “Organização da informação”, “Apresentação da informação”, “tomada de decisão”.

5.2.1 Produção dos Dados

A partir da organização supracitada, o módulo nomeado "produção de dados" está intrinsecamente relacionado ao ator aluno, tendo o ator como fonte geradora de dados, ou seja, a informação gerada pela ferramenta é dirigida pelos padrões do estudante dentro do escopo do curso. Estes padrões são definidos a partir da interação do ator aluno com a plataforma, ou seja, a partir da vida escolar do aluno dentro a da plataforma moodle, realizando toda e qualquer tarefa, sem exceções, a plataforma registra cada interação com a plataforma em um banco de dados em uma tabela de logs, como demonstrado na Figura 20.

Figura 20 – Diagrama de produção de dados



Fonte: Autoria própria

Como demonstrado na Figura 20 acima, o ator aluno em sua vida escolar, interage diretamente com vários ambientes, realizando atividades, consultando material, fazendo provas, se comunicando, entre outras formas de interação com a plataforma. Tais interações geram dentro da plataforma registros de log, não somente registros formalmente enviados pelos alunos, mas registros de acesso, como por exemplo, o caso do aluno que acessou o fórum disponível na plataforma e fez uma busca, o aluno não enviou nada formalmente a plataforma, mas nos registros de log constam a passagem do aluno x pelo ambiente fórum e o registro de busca realizada pelo aluno x no fórum.

Em complemento, este trabalho é pautado em uma prova de conceito, que utiliza de dados para gerar informação a partir de registros de atividades de discente de educação a distância (EAD) da plataforma moodle, os chamados LOGS, como demonstrado na Figura 21, que se trata de uma expressão utilizada para descrever o processo de registro de eventos relevantes num sistema computacional.

A partir deste escopo acima delimitado, entende-se que há limitações em meio aos dados, são dados de difícil acesso e manipulação por serem espúrios. Porém ainda há uma vasta quantidade de informações contidas nestes dados que necessitam de exploração.

Em complemento, estes dados, assim como em um ambiente presencia, um professor observa os padrões de comportamento, interações e ações dos alunos para entender a atual situação dos alunos, esses registros de LOGS são uma fonte de dados que registram padrão de comportamento dos alunos no sistema, contendo as interações e ações do aluno dentro do sistema.

Adicionalmente, esses dados podem ser utilizados como insumo para gerar informação para o professor, com intermédio de técnicas de análise de dados, esses dados podem gerar informações sobre a atual situação dos alunos, do curso e da turma.

Tais dados são gerados e armazenados em um banco de dados e podem ser facilmente recuperados por inúmeros meios, seja conseguindo permissão de acesso ao banco de dados, ou no próprio sistema é possível exportar os dados. Esses dados gerados, são dados espúrios, cujo o qual tem finalidade única de registro, então contém inúmeras informações sobre datas, qual curso, qual o usuário, quem foi afetado por essa ação, qual evento foi realizado, qual componente do sistema, o que foi feito devidamente, uma descrição sobre o ocorrido, qual a plataforma que gerou esse registro, seja web ou mobile, e o endereço de IP de quem realizou a ação, como mostra a Figura 21.

Figura 21 – Logs do ambiente Moodle

Time	User full name	Affected user	Event context	Component	Event name	Description	Origin	IP address
24 January 2023, 2:48:16 AM	Guest user	-	Course: History: Russia in Revolution	System	Course viewed	The user with id '1' viewed the course with id '56'.	web	181.67.205.129
24 January 2023, 2:48:09 AM	Guest user	-	Assignment: Assignment: Causes of the October 1917 Revolution	Assignment	The status of the submission has been viewed.	The user with id '1' has viewed the submission status page for the assignment with course module id '573'.	web	181.67.205.129
24 January 2023, 2:48:09 AM	Guest user	-	Assignment: Assignment: Causes of the October 1917 Revolution	Assignment	Course module viewed	The user with id '1' viewed the 'assign' activity with course module id '573'.	web	181.67.205.129
24 January 2023, 2:47:56 AM	Guest user	-	Assignment: Assignment:	Assignment	Course module viewed	The user with id '1' viewed the 'assign' activity with course	web	181.67.205.129

Fonte: Documentação oficial do Moodle¹

O sistema é montado de forma hierárquica, tendo cada nível de hierarquia uma permissão oferece um nível de informação, logo os professores conseguem ver os dados de seus alunos e do próprio professor, o aluno só vê os próprios registros e o administrador do sistema consegue ver tanto do aluno quanto do professor quanto do sistema. Como o moodle é um sistema open source, ele permite a alteração do formato desses logs, podendo adicionar ou remover dados deste documento, alterar a forma como são registrados, inferir mais informações específicas, tornando-se extremamente customizável a dispor de quem utilizará, porém, este trabalho se atém aos dados disponibilizados por padrão pela plataforma.

Por padrão a plataforma ainda provê alguns simples relatórios quantitativos sobre as atividades realizadas pelo estudante, como a quantidade de vezes que abriu os materiais, se fez ou não download de material, mostra também as participações do estudante em fóruns e quais fóruns, mostra também em seu relatório, quais foram as notas das atividades realizadas, quais as atividade o estudante fez, quais feedback recebeu, entre outros dados quantitativos e registros de interação do estudante com a plataforma, para que o professor avalie a atual situação do estudante. Outro relatório padrão da plataforma é o de participação, onde mostra uma lista com quem participou das atividades dispostas, e quantas vezes, este relatório pode ser filtrado por visualização e publicação e pode ser visto individualmente também a participação de cada aluno.

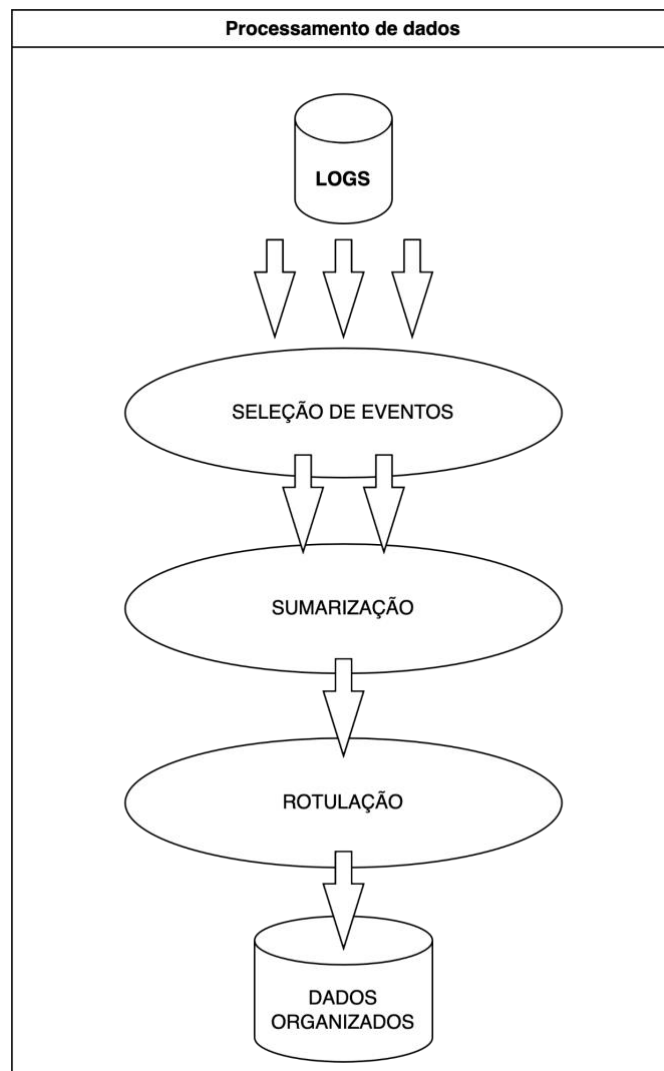
Esses relatórios contêm inúmeros dados sobre os estudantes, mas ainda são dados que somente descrevem o atual momento do estudante, não existe nenhum dado comparativo, para realizar isso o professor precisa iterar entre todos os estudantes e realizar essa análise manualmente, coletando todos os dados espúrios e sumarizando-os para, caso tenha as permissões para visualizar os dados de turmas anteriores, que não necessariamente eram de sua responsabilidade, ainda necessitaria de habilidades analíticas e estatísticas necessárias, computacionais e lógicas, além de entender a atual situação desta turma e mesmo assim poderia ficar sem dados das outras turmas para entender se existe algum padrão de comportamento ou comum dificuldade em algum ponto específico, ficando refém de sua própria experiência com turmas anteriores para tentar identificar algum padrão ou modo de agir.

Em complemento, estes dados são como um mapa marcando toda a trajetória do aluno dentro da plataforma até atingir seu objetivo, contendo todas as experiências que este obteve durante seu período letivo. Porém a plataforma gera dados sobre toda a plataforma e todos os atores do ambiente, gerando assim dados espúrios, desorganizados, sem outro propósito que não o registro de interações. Estes dados carregam cargas valiosas de informação, mas precisam ser processados para que se tornem dados úteis.

5.2.2 Processamento de Dados

A partir dos dados recuperados da plataforma, o segundo módulo nomeado “processamento de dados” é inserido no ciclo, com o intuito de formatar esses dados para que sejam utilizados nos módulos subsequentes, logo, a ferramenta realiza o tratamento e a organização desses dados como uma peneira, para que estes dados sejam tratados e organizados de acordo com a necessidade informacional do processo decisório abordado, o processo de avaliação de performance educacional. Esta peneira, como demonstrado na Figura 22, visa separar os dados indesejados, dos dados úteis ao problema, como uma forma de reorganizar as informações e retirar impurezas.

Figura 22 – Diagrama de processamento de dados



Fonte: Autoria própria

Em complemento, este processo visa criar uma base de dados alinhada à demanda informacional, criando uma base voltada ao aluno onde se tem insumos para o processamento da informação, trazendo dados direcionados, organizados, tabulados e padronizados. A partir disso, como demonstrado na Figura 23, o dado passa por uma seleção de eventos, ou seja, os eventos contidos nos logs são selecionados, então, retira desses registros quais eventos foram registrados utilizando o nome dos eventos. Com esses eventos selecionados, então os dados são sumarizados em torno do aluno, ou seja, coloca o aluno como centro e realiza a contagem dos eventos que esse aluno realizou na plataforma. Por fim os dados são rotulados, com o objetivo da avaliação, ou seja, agrega-se nos dados dos alunos as notas que esses alunos obtiveram.

Exemplificando o processo, seria como obter uma enorme quantidade de pontos em um mapa e traçar uma reta até o destino final. Logo, utilizando um raciocínio lógico simplificado, pode-se afirmar que, o dado "A" que descrevia que o aluno "X" acessou a plataforma no horário "Y" e realizou o evento "Z", tornou-se parte de um novo dado que diz que o aluno "X" realizou n vezes o evento "Z" e obteve a nota "B", como demonstrado na Figura 23 e a Figura 24 estes dados descritivos de difícil interpretação, tornam-se dados quantitativos mais fáceis de serem manipulados e compreendidos. Este processo, apesar de ser um processo realizado por agente computacional, é um processo de curadoria de dados e deve ser frequentemente revisado, pois é de suma importância para a informação final.

Figura 23 – Dados orientados ao registro

	Hora	Nome completo	Usuário afetado	Contexto do Evento	Componente	Nome do evento	Descrição	Origem	endereço IP
200	2/06/2021 10:04	Fake name	-	Curso: 2021	Sistema	Curso visto	The user with id '1234' viewed the course with id '2021'.	web	000.0.000.000
201	2/06/2021 10:04	Fake name	-	Curso: 2021	Sistema	Curso visto	The user with id '1234' viewed the course with id '2021'.	web	000.0.000.000
202	2/06/2021 10:03	Fake name 2	-	URL: Videoconferência gravada 11/05/2021	URL	Módulo do curso visualizado	The user with id '1236' viewed the 'url' activity with course module id '2021123'.	web	000.0.000.000
203	2/06/2021 10:03	Fake name 3	-	Curso: 2021	Sistema	Curso visto	The user with id '1235' viewed the course with id '2021'.	web	000.0.000.000
204	2/06/2021 10:03	Fake name	-	URL: Videoconferência gravada 18/05/2021	URL	Módulo do curso visualizado	The user with id '1234' viewed the 'url' activity with course module id '2021123'.	web	000.0.000.000

Fonte: Autoria própria

Figura 24 – Dados orientados ao usuário

	Nome completo	Tentativa do questionário visualizada	Módulo do curso visualizado	Curso visto	Usuário recebeu nota	Tentativa do questionário revisada
0	fake name 1	903.0	259.0	280.0	138.0	172.0
1	fake name 2	885.0	437.0	355.0	95.0	109.0
2	fake name 3	694.0	304.0	284.0	46.0	10.0
3	fake name 4	1108.0	555.0	345.0	134.0	261.0
4	fake name	675.0	296.0	215.0	122.0	96.0

Fonte: Autoria própria

Em complemento, como dito anteriormente, este processo necessita de constante revisão, visto que esta é a base bruta para a extração da informação, estes dados alteram diretamente os processos a seguir, ressalta-se a dinamicidade nos ambientes educacionais, frente ao vasto número de disciplinas que podem: ser ministradas e curadas pelos professores e estudadas pelos alunos; Diferente da base de logs esta nova base tem um objetivo que vai além de somente armazenamento, logo necessita de cuidados diferentes.

Adicionalmente, como dito por SANTANA, R. C. G (2013), a partir do primeiro registro de informações em suportes, o objetivo desses suportes é propiciar acesso aos dados, de forma a propiciar uma percepção contínua sobre o estado atual dos inventários. Porém propiciar essa percepção necessita conhecimentos e técnicas que cobrem todas as fases de acesso aos dados. Dito isso, a Ciência da Informação traz um arcabouço teórico para apoiar essas técnicas onde pode-se alinhar os dados aos objetivos informacionais visando uma informação final mais trabalhada. Esta parte do processo necessita e se encaixa perfeitamente neste cenário, como parte crucial desse processo é necessária a constante revisão do ciclo de vida dos dados.

Em complemento, este ciclo de vida permeia todo o modelo, com as etapas descritas como, coleta, armazenamento e recuperação, esta etapa específica abordou a coleta, a partir dos dados vindos dos logs, e o armazenamento em uma nova base com um objetivo diferente, em meio a essas duas etapas existe uma etapa adicional, que sofre influência direta dos demais processos, esta etapa adicional decide entre o descarte e a utilização do dado, visando descartar dados tidos como desnecessários pelos próximos processos.

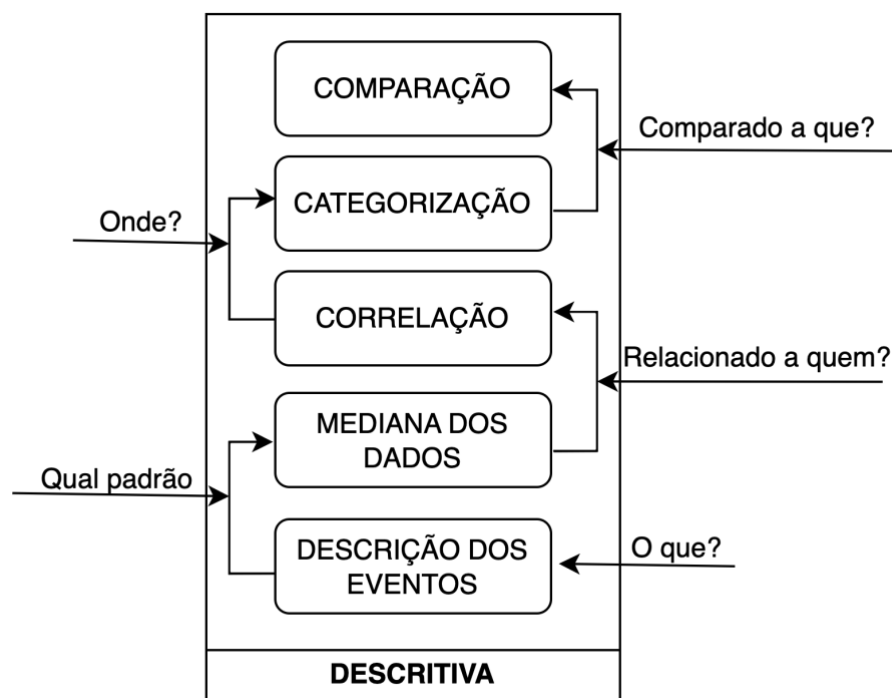
5.2.3 Enriquecimento da Informação

Em continuidade ao processo, após a finalização do processamento de dados e de todos os devidos cuidados necessários na preparação dos dados, visando processar a informação contida nesta nova base de dados voltada ao aluno, o modelo inicia o processo denominado "enriquecimento da informação", cujo objetivo é consumir os dados existentes e agregar valor

informacional a eles, ou seja, trazer mais utilidade para esse dados por meio de análises e inferências estatísticas.

Em complemento, para realizar tal objetivo o módulo é desenvolvido utilizando análises descritivas, ou seja, análises que utilizam técnicas e conceitos estatísticos para descrever os dados, como médias, medianas, contagens, como exemplificado na Figura 25 utilizando uma sequência de técnicas alinhadas visando gerar melhores insumos informacionais descritivos para o agente consumidor, o professor, além de em meio a isso conseguir realizar gerar esses insumos para ambos os ambientes existentes, divididos entre dados históricos e dados atuais.

Figura 25 – Diagrama de análises descritivas



Fonte: Autoria própria

Utilizando os dados históricos sobre as turmas anteriores, fornecido pelo módulo anterior, como dados de entrada deste módulo, primeiramente é necessário saber o que é que será analisado, ou seja, qual evento, logo em sequência a primeira técnica utilizada é a mediana, visando encontrar a mediana de cada evento disponível, com intuito de traçar o padrão de acessos da maioria dos alunos, ou seja, entender como a maioria se comportou na ferramenta, este dado somado a descrição dos eventos, trazem uma boa visão de como a maioria dos alunos utilizam a ferramenta.

Um exemplo aplicado ao cenário exposto, seria entender que a mediana de acessos ao material didático é de 240 acessos no período letivo, ou seja, a maioria dos alunos acessam em uma margem próxima a 240 vezes o material didático.

Em complemento, apenas essa informação seria um dado vazio, quase sem utilidade se sozinho, porém ao ter medianas de outras ações pode-se traçar correlações entre elas, visando entender o quão impactante essa ação é perante a performance. Exemplificando, um curso que tem um dado histórico de média de acessos ao material didático de 240 e uma média de avaliação 3 em uma escala de 1 a 4, se a turma atual tem uma média de acessos abaixo desse valor, ela pode vir a ter uma média menor, caso note-se uma correlação direta entre os dois valores.

A partir desses dados o modelo visa categorizar os dados gerados, neste caso, em duas categorias, históricos, que são os dados derivados das turmas anteriores que já finalizaram o curso e atuais, visando, a partir dessas categorias, promover uma comparação entre os dois cenários, a fim de guiar o professor a entender o cenário do curso como um todo, criando uma descrição do passado e do presente.

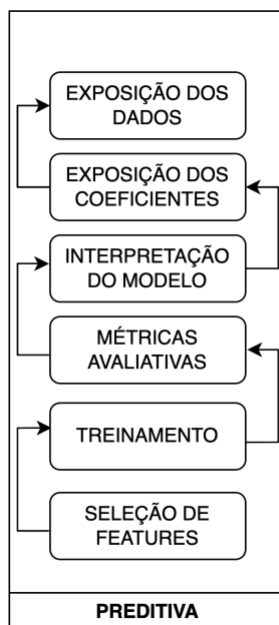
Estas análises resultam em um maior entendimento dos dados, tanto para o agente consumidor, ou seja, o professor, quanto para o agente produtor, ou seja, o próprio modelo, podendo ser utilizada como um meio de entender os dados utilizados pelo modelo e separar, insumos que agregam e melhoram a informação gerada pelo modelo, de insumos irrelevantes ou que podem gerar confusão, algum viés desnecessário para o ambiente ou de cunho discriminatório ou excludente.

Em complemento, aliada a análise descritiva também visando o enriquecimento da informação, como supracitado, visando trazer uma ambientação não somente sobre o atual estado dos estudantes e o padrão histórico do curso o modelo adiciona uma análise preditiva em seu processo denominado "enriquecimento da informação" de forma paralela e complementar a análise descritiva, utilizando técnicas de aprendizagem de máquina para reconhecer os padrões dos estudantes, a fim de além de demonstrar o estado atual do curso e o estado passado do curso, como supracitado com as análises descritivas, mostrar também o estado futuro do curso para o tomador de decisões.

Adicionalmente, como demonstrado na Figura 26, a análise preditiva é realizada selecionando os eventos mais relevantes como insumos para seu processo, visando sempre a melhor informação possível, esta etapa colabora e se complementa da análise descritiva, apesar de realizadas em paralelo, os resultados e o aprendizado retirado das análises descritivas, são extremamente úteis para o entendimento de quais eventos utilizar como insumo para o modelo

e para guiar o modelo a ter dados mais relevantes para o objetivo proposto.

Figura 26 – Diagrama de análises preditivas



Fonte: Autoria própria

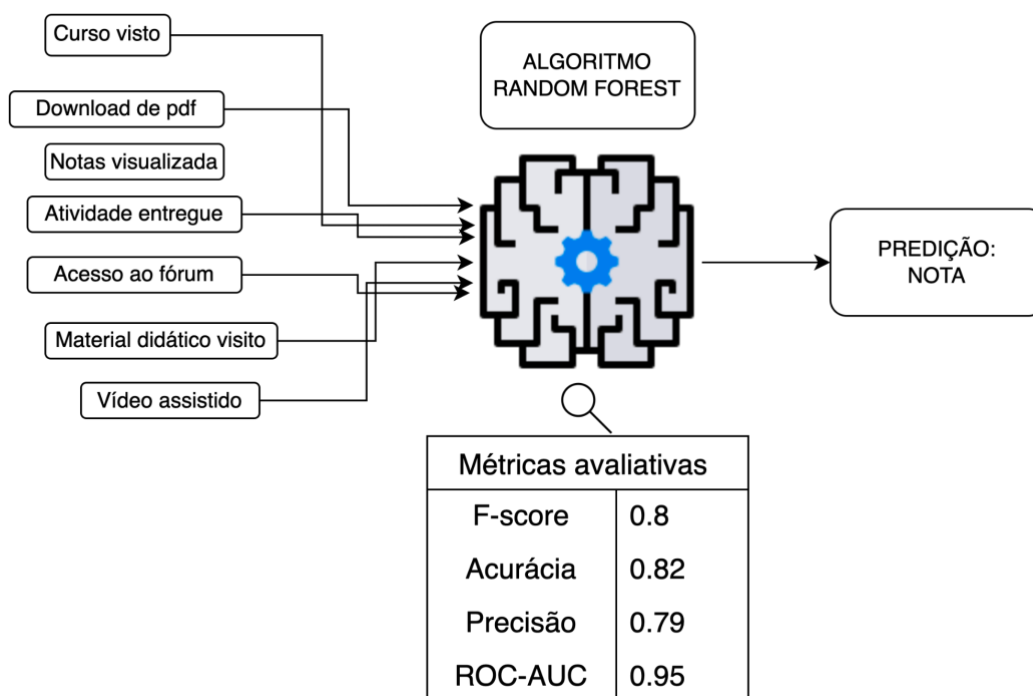
Em complemento, antes do início do treinamento do algoritmo de aprendizagem de máquina, é necessária a seleção dos dados que vão servir de insumo para o algoritmo, pois esses dados serão responsáveis pela performance do algoritmo, logo melhores insumos melhores resultados. Como exemplo, um ambiente pode ter mais de 20 eventos diferentes, mas ao inseri-los no algoritmo e analisar os resultados é identificado que apenas 10 desses eventos são úteis ao modelo e ao refazer o treinamento e as análises nota-se que a performance melhora ao utilizar somente esses 10 eventos, logo, apesar de ter mais insumos disponíveis, é melhor utilizar somente os 10, assim como é exemplificado na Figura 27 onde o evento notas visualizada, não é utilizado pelo algoritmo. Em contraponto, o oposto é válido, ao retirar features o algoritmo pode ter sua performance prejudicada logo, mesmo caso ela tenha baixa relevância é importante mantê-la no algoritmo visando uma melhor performance.

Visando a melhor performance e um resultado mais alinhado com o contexto, os eventos selecionados como insumos para o treinamento do algoritmo devem ser reavaliados de acordo com os resultados das próximas etapas deste módulo.

A partir dos já selecionados e padronizados dados, o treinamento do algoritmo de aprendizagem de máquina escolhido é iniciado, como visto na Figura 27, neste caso em

específico, como já supracitado, um modelo supervisionado classificatório e interpretável. Visando sempre uma melhor performance, os resultados deste treinamento são submetidos a métricas avaliativas, tal qual como uma avaliação formal no plano educacional, onde tais métricas são aplicadas sobre o modelo visando entender, de forma matemática, o quão confiáveis são os dados gerados pelo modelo treinado em comparação a realidade dentro de um universo já conhecido, assim como pode ser visto no exemplo da Figura 27, onde demonstra numericamente algumas métricas do algoritmo random forest. Estas métricas são utilizadas para avaliar não somente a performance do modelo, mas sua saúde, sua usabilidade, sua vida útil como um todo e deve sempre ser reavaliada para evitar modelos arcaicos, imprecisos, tendenciosos e que gerem informações irrelevantes.

Figura 27 – Exemplificação de modelo preditivo



Fonte: Autoria própria

Adicionalmente, após avaliado o modelo, se faz necessária a compreensão dos resultados obtidos, visto que os resultados obtidos pelo modelo são diretos, ou seja, é um resultado numérico, por exemplo, no caso do modelo podendo ser de 0 a 4, onde 4 seria a nota "A" assim sequencialmente decrescendo até a nota "D" que seria o 0. Esse resultado não basta ser apenas traduzido de um número para uma nota pelo modelo, mas também é necessária a

contextualização dessa nota para o agente consumidor da informação.

Em complemento, para realizar esta contextualização para o agente consumidor da informação, a etapa de interpretação do modelo traz uma forma de avaliar os dados consumidos de forma a trazer a relevância de cada feature selecionada na tomada de decisão do modelo, como demonstrado pela Tabela 1, inspecionando o modelo e tornando o modelo interpretável, assim possibilitando a criação de um ambiente onde o agente consumidor está ciente dos motivos para tal informação ser servida a ele.

Quadro 1 – Feature e relevância

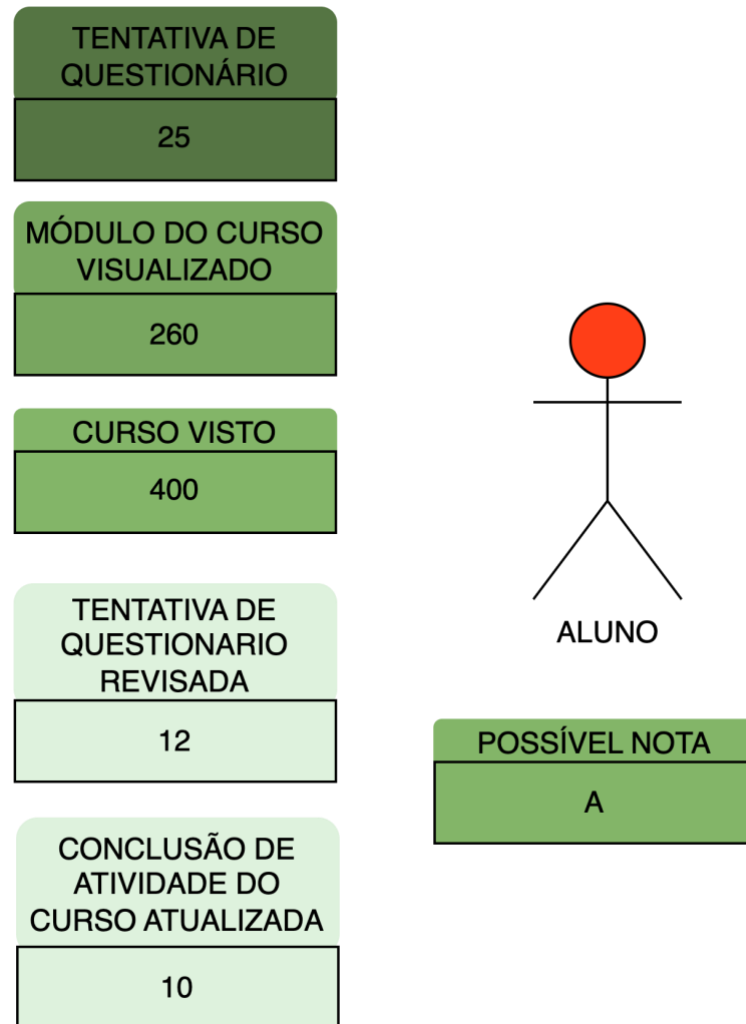
Feature	Relevância
Tentativa do questionário visualizada	0.1529279279279279
Módulo do curso visualizado	0.20450450450450447
Curso visto	0.24707207207207205
Tentativa do questionário revisada	0.020495495495495463
Conclusão da atividade do curso atualizada	0.10135135135135132

Fonte: Autoria própria

Adicionalmente, utilizando estes dados sobre a relevância de cada feature do modelo, seja de forma direta, informando ao agente consumidor quais os graus de importância de cada feature analisada por meio de uma visualização, como exemplificado em um esquema de baixo nível de fidelidade na Figura 28, ou de forma indireta ordenando a informação para que seja consumida de forma sequencial por importância cada dado sobre a decisão como exemplificado em um esquema de baixo nível de fidelidade na Figura 29.

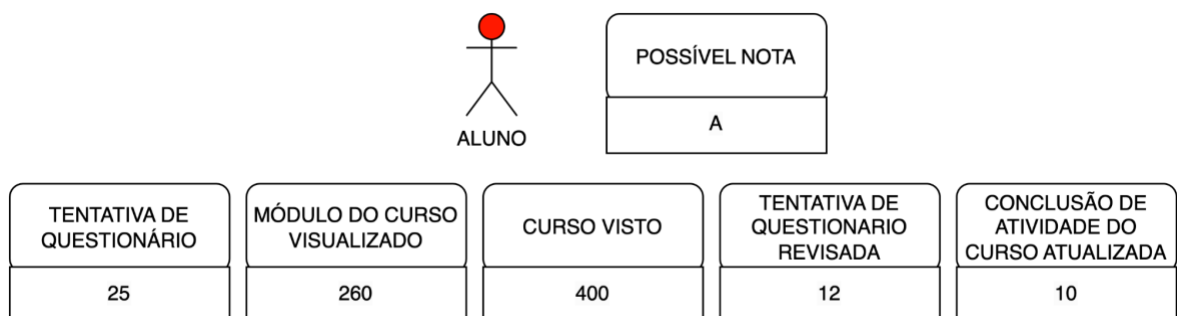
Como exemplo deste processo de enriquecimento, pode-se considerar os dados numéricos mostrados nas Figuras 28 e 29 como dados sobre o comportamento do aluno na plataforma que serão utilizados como dados de entrada, e a partir da associação deles com os dados de saída das análises, como exemplificado no caso da análise preditiva pelo dado "Possível nota" presente nas Figuras 28 e 29 e ainda associados a ordem criada pela relevância, como uma informação final enriquecida e contextualizada, podendo esses dados ainda serem associados aos dados históricos providos pelas análises descritivas a fim de enriquecer ainda mais a informação.

Figura 28 – Exemplificação de informação servida de forma direta



Fonte: Autoria própria

Figura 29 – Exemplificação de informação servida de forma indireta



Fonte: Autoria própria

De forma complementar, além de servir como um direcionamento para o modo como a informação será servida ao consumidor, facilitando a compreensão e dando relevância a cada feature, este método também se faz útil na seleção de features irrelevantes para o modelo, pois, se uma feature apresenta relevância zero na avaliação feita pelo modelo, ela pode ser retirada, tornando o modelo mais leve, mais preciso muitas vezes e de menor custo computacional. Ressalta-se a importância da avaliação da vida útil desses dados, descritos no trabalho de SANTANA, R. C. G (2013), visto que estes permanecem armazenados em outra base é válido a reavaliação a fim de melhorar e aumentar a vida útil do modelo com os melhores parâmetros disponíveis, sendo representada no ciclo de vida dos dados como a etapa de recuperação, visando recuperar dados que por sua vez não tinha utilidade no devido momento em que foi descartado, mas devido a um novo contexto se faz útil.

5.2.4 Organização da Informação

Após a geração dos dados e o enriquecimento destes pelos módulos anteriores, mesmo com o dado disponível, a informação ainda não está completamente instaurada, para isso é necessário processo de ordenação e organização dos dados, para que os dados possam ser corretamente consumidos e interpretados. Para isso o modelo utiliza o módulo denominado "organização da informação", que entra como uma forma de organizar os dados gerados pelas análises, para uma melhor interpretação do agente consumidor, o professor, selecionando os dados enriquecidos, os tornando disponíveis e de fácil entendimento, agregando os dados históricos aos dados atuais e os contextualizando ao cenário.

Em complemento, este processo é um processo de curadoria da informação e deve ser adaptado levando em consideração as habilidades do agente consumidor da informação, neste caso o professor, esses dados gerados precisam ser associados a informações descritivas sobre estes em forma estruturada, sempre moldados de acordo com a necessidade do próximo módulo, que será detalhado na seção seguinte, a camada de apresentação da informação, logo o modelo deve gerar e organizar essa informação visando o agente consumidor da informação e seguindo as diretrizes do módulo de apresentação da informação. Exemplificando, uma forma de ordenação e organização dos dados visando o agente consumidor e dirigida pela camada de visualização, é a estrutura de dicionários, onde você pode organizar e dar rótulos aos dados de acordo com cada necessidade, como demonstrado na Figura 30, os dados podem ser organizados de acordo com cada seção que será apresentada como, a seção dados históricos, que apresenta os dados das turmas passadas, seguido dos dados da turma atual e a seção gráficos

que tem dados específicos para os gráficos que serão apresentados.

Figura 30 – Organização da informação: Estrutura de dicionário

```
{
  "dados_historicos":{
    "Curso visto":{
      "Mediana": 670
    },
    "Tentativa do questionário visualizada":{
      "Mediana": 60
    },
    "Módulo do curso visualizado":{
      "Mediana": 400
    },
    "Tentativa do questionário revisada":{
      "Mediana": 50
    },
    "Conclusão da atividade do curso atualizada":{
      "Mediana": 34
    }
  },
  "turma_atual":{
    "Curso visto":{
      "Mediana": 300
    },
    "Tentativa do questionário visualizada":{
      "Mediana": 20
    },
    "Módulo do curso visualizado":{
      "Mediana": 349
    },
    "Tentativa do questionário revisada":{
      "Mediana": 12
    },
    "Conclusão da atividade do curso atualizada":{
      "Mediana": 10
    }
  },
  "graficos":{
    "notas por visualização":{
      "visualização":[100,223,333,412,652,789,847],
      "notas":["A","B","C","D"],
    },
    "trabalhos enviados":{
      "envios":[23,25,12,20,15,13],
      "frequencia":[0,5,10,15,20,25,30,35]
    }
  }
}
```

Fonte: Autoria própria

5.2.5 Apresentação da Informação

Em continuidade, com os dados devidamente processados e organizados, em torno da demanda informacional, a fim de prover uma melhor interpretação, o módulo da camada de apresentação denominado "apresentação da informação", torna-se uma ponte entre o agente consumidor da informação e o sistema, esta camada é importante para tornar o processo todo mais intuitivo, neste processo os dados são interativos e explicados utilizando uma interface gráfica de apresentação, como um site, um relatório, um dashboard com gráficos tabelas, filtros, descrições entre outros artifícios para tornar a informação mais fácil de ser consumida pelo ator.

Adicionalmente, este processo é de suma importância e deve ser executado de forma a tornar os dados produzidos dados de fácil acesso, visto que os ambientes como supracitados são dinâmicos e divergentes em sua natureza, precisam ser apresentados considerando as capacidades interpretativas do agente consumidor, visando guiar o entendimento do consumidor para evitar informações dubitativas, que podem atrapalhar ou induzir o tomador de decisão a não discernir sobre as demais informações úteis disponíveis, que não fazem parte do escopo deste modelo, mas que podem interferir num momento de tomada de decisão, visto a grandeza e complexidade do ambiente e do processo de tomada de decisão pedagógico.

Como exemplificado pela Figura 31, onde traz os dados históricos de um curso exemplo, e forma de gráficos e tabela, mostrando nos gráficos de dados históricos o valor mínimo, o valor máximo daquele campo e a mediana dos dados como métrica principal, além de mostrar a quantidade de estudantes contidos nesse relatório, as performances desses alunos, a quantidade de alunos reprovados e um gráfico mostrando a relação de alguns eventos com a nota final do aluno. Todas essas informações e visualizações tem um único objetivo em comum que é transmitir uma mensagem de forma mais clara e mais direta ao agente consumidor, para que este possa interpretar mais rápido e absorver esta informação para sua tomada de decisão

Figura 31 – Dashboard de exemplo de apresentação da informação



Fonte: Autoria própria

5.2.6 Tomada de Decisão

Por fim, o último processo é a tomada de decisão em si, este processo depende somente do ator professor, onde este deve fazer a interpretação das informações fornecidas pelo sistema e discernir sobre a sua importância em sua tomada de decisão, que pode intervir ou não no processo de aprendizado do estudante. Esta interferência é uma interferência que no sistema pois pode mudar a forma com que o estudante interage com a plataforma e criar novos padrões com o tempo, por isso o modelo deve ser utilizado de forma cíclica, sendo revisado visando ser adaptável e treinar os novos padrões para utilizá-los em seus próximos ciclos, aumentando o tempo de vida do modelo e dos dados presentes no sistema, como supracitado.

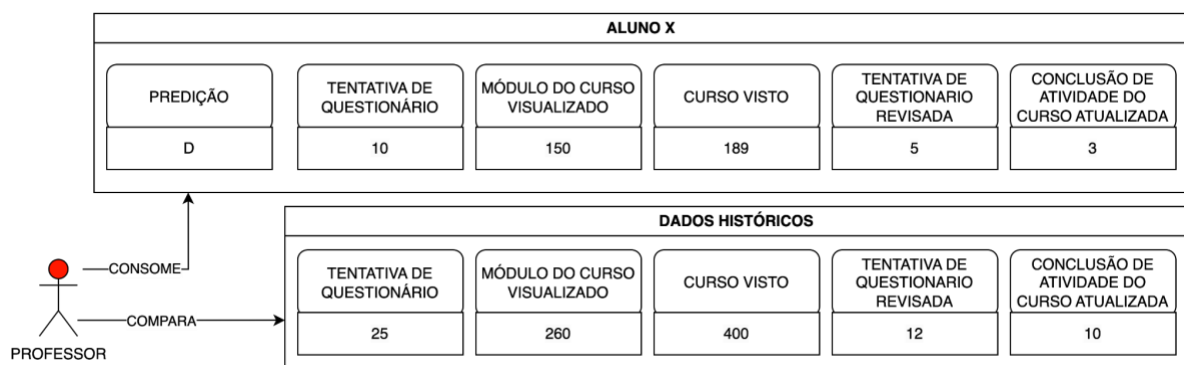
Em complemento, visando exemplificar a utilização deste modelo no processo de tomada de decisão, é necessário ressaltar que este modelo visa apoiar a tomada de decisão

pedagógica com informações referentes a performance estudantil, porém este processo de tomada de decisão é muito vasto, logo, a exemplificação se limita a apenas alguns casos, mas a utilização do modelo não se limitando a somente estes. Os processos de tomada de decisão cujo este modelo visa apoiar, são os relacionados a turma em si, ou seja, avaliar a atual situação da turma ou do aluno, logo, o professor ao se ver sem entender a atual situação da turma pode utilizar os dados do modelo como uma forma de comparar o desempenho com as turmas anteriores e avaliar se a turma está bem ou se necessita de uma intervenção, o mesmo vale ao avaliar um específico aluno, e o comparar com as performances das turmas anteriores.

Para isso o professor pode comparar tanto os dados descritivos, sobre os padrões do curso e os atuais dados do aluno, mas também pode utilizar a predição da nota, aliada aos dados descritivos auxilia na identificação dos alunos que fogem ao padrão e que podem ter uma baixa performance.

Em complemento, o professor também pode utilizar as informações produzidas pelo modelo como uma forma de acompanhar o progresso da turma, ou seja, verificar se a turma está se comportando como o esperado, utilizando esses momentos também para identificar os alunos passíveis de reprovação, como demonstrado na Figura 32, o ator professor consome a informação preditiva do modelo, auxiliada dos dados descritivos que contextualizam a informação preditiva, a partir desses dados o professor tem insumos para avaliar o aluno mesmo que este ainda não tenha uma avaliação formal, estes momentos podem ser utilizados para a intervenção antecipada e aumentar a porcentagem de aprovação do curso, já que mesmo em caso de um aluno que venha a ter uma performance que o aprove mesmo sem intervenção, um maior cuidado pode melhorar a performance desse aluno.

Figura 32 – Uso dos dados do modelo



Fonte: Autoria própria

Visto que os ambientes educacionais, como supracitados, são dinâmicos e divergentes em sua natureza, este trabalho não visa gerar uma regra generalizada para todos os ambientes, mas sim um modelo trazendo a tona as reais necessidades e preocupações contidas nessas etapas que devem ser devidamente tratadas em cada ambiente e em cada situação onde o modelo for aplicado.

No próximo capítulo, será apresentada uma prova de conceito utilizando dados de um único curso, objetivando a aplicação do modelo performi para avaliação do modelo e dos resultados obtidos.

6 PROVA DE CONCEITO

Neste capítulo é apresentada e detalhada uma prova de conceito onde, o modelo Performi anteriormente detalhado, é aplicado e dados da vida real do ambiente de educação a distância moodle, são submetidos cujo resultado visa obter uma informação final contextualizada e útil com suporte do modelo Performi.

A prova de conceito baseia-se na avaliação da informação produzida, mais especificamente de uma turma de um curso do ambiente moodle. Para realização de tal prova de conceito, foi utilizada a linguagem de programação python junto às bibliotecas pandas e scikit learn como ferramentas para construção de um ambiente orientado ao modelo proposto, sendo implementados cada módulo separado visando a avaliação de cada etapa do processo.

6.1 Processo e Resultado da Prova de Conceito

O processo de avaliação e construção da informação na avaliação de performance estudantil, desde a aquisição dos dados, até a apresentação da informação, dirigida pelo modelo Performi é apresentado nessa seção.

Adicionalmente, esta prova de conceito irá trabalhar em cima de um único curso e uma única turma, com dados baseados em turmas reais anonimizadas e mantidos somente o essencial para este trabalho. Apesar da estrutura de um LOG geralmente conter mais informações sobre o acesso, visando manter somente o mínimo de informações necessárias foram removidas da estrutura original do LOG os seguintes dados "Hora", "Usuário afetado", "Descrição", "Origem", "endereço IP", "Contexto do Evento", "Componente" e o componente "Nome completo" foi substituído por um ID único para cada usuário visando manter preservada a identidade do aluno, sendo assim, como esta prova de conceito utiliza somente os nomes dos eventos e os identificadores dos alunos, somente essas duas informações foram mantidas.

Considere o curso X, no ambiente moodle, que tem uma base de dados de registro de logs ao curso, ex alunos e alunos cursando, como fonte de dados, ou seja o que representaria o módulo nomeado "produção dos dados", onde a base de dados foi submetida ao modelo Performi em estrutura de tabela, no formato csv, com a estrutura fora acima descrita e demonstrada na Figura 33, com a coluna "Nome completo" sendo o identificador único do aluno e os eventos sendo listados na coluna "Nome do evento".

Figura 33 – Dados de LOG

	Nome completo	Nome do evento
120357	43	Usuário recebeu nota
120358	43	Usuário recebeu nota
120359	43	Usuário recebeu nota
120360	43	Usuário recebeu nota
120361	43	Módulo do curso visualizado
120362	43	Curso visto
120363	43	Relatório de notas do usuário visualizado
120364	43	Relatório de notas do usuário visualizado
120365	43	Curso visto
120366	43	Módulo do curso visualizado
120367	43	Curso visto
120368	43	Relatório de notas do usuário visualizado
120369	43	Tentativa do questionário revisada
120370	43	Curso visto
120371	43	Curso visto

Fonte: Autoria própria

Adicionalmente, visando a necessidade futura de rotulação dos dados, o produtor de dados também fornece em mesma estrutura e formato as notas dos ex alunos atreladas aos seus identificadores, como demonstrado na Figura 34, com as colunas Nome completo e Resultado.

Figura 34 – Notas finais

	Nome completo	Resultado
101	30	B
102	31	D
103	32	B
104	33	C
105	34	B
106	37	B
107	38	C
108	39	C
109	40	B
110	41	B

Fonte: Autoria própria

Em continuidade, com os dados devidamente carregados, o módulo nomeado "processamento de dados" é iniciado, onde esses dados serão sumarizados e transformados em uma base de dados com o aluno como centro da informação. Como demonstrado na Figura 35, onde agora cada linha da tabela é referente a um único aluno e cada evento tornou-se uma coluna na tabela que contém a quantidade de vezes que estes alunos realizaram tal ação, tornando-se uma base quantitativa de registros de cada aluno.

Figura 35 – Base de dados direcionada ao aluno

	Nome completo	Tentativa do questionário visualizada	Módulo do curso visualizado	Curso visto	Usuário recebeu nota	Tentativa do questionário revisada	Conclusão da atividade do curso atualizada	O status da submissão foi visualizado.	Relatório de notas do usuário visualizado	Tentativa do questionário iniciada	...
0	10	903.0	259.0	280.0	138.0	172.0	32.0	7.0	34.0	55.0	...
1	11	885.0	437.0	355.0	95.0	109.0	35.0	21.0	36.0	45.0	...
2	12	694.0	304.0	284.0	46.0	10.0	47.0	8.0	6.0	23.0	...
3	13	1108.0	555.0	345.0	134.0	261.0	37.0	13.0	58.0	70.0	...
4	15	675.0	296.0	215.0	122.0	96.0	53.0	21.0	12.0	56.0	...
5	16	741.0	478.0	209.0	120.0	147.0	91.0	23.0	71.0	56.0	...
6	17	1242.0	519.0	436.0	162.0	291.0	58.0	22.0	70.0	81.0	...
7	18	997.0	343.0	388.0	252.0	167.0	119.0	3.0	72.0	61.0	...
8	19	28.0	11.0	11.0	4.0	1.0	3.0	0.0	1.0	1.0	...
9	20	864.0	443.0	369.0	147.0	177.0	82.0	8.0	27.0	69.0	...

Fonte: Autoria própria

Em complemento, essa tabela é gerada por uma função iteradora que conta a quantidade de eventos por aluno e cria uma coluna nomeada com o nome do evento, com esses valores e a cada iteração adiciona na tabela o evento como uma coluna extra. Em sequência visando rotular esses dados, para a utilização no próximo módulo, os dados então são unidos pelo identificador único do usuário com as notas finais, tornando-se base para o próximo módulo de enriquecimento da informação.

Em continuidade, criada esta base de dados, inicia-se o módulo de enriquecimento da informação com uma análise para determinar qual algoritmo de aprendizagem de máquina multiclass melhor desempenharia a função preditiva dentro do modelo. Para isso foram selecionados alguns algoritmos, sendo eles: One vs rest (Wu, W., Gao, X., & Gao, S. (2006)), random forest (Breiman, L. (2001)), logistic regression (Cramer, J. S. (2003)), k-nearest neighbor (Fix, E., & Hodges Jr, J. L. (1952)), ridge (Saunders, C., Gammerman, A., & Vovk, V. (1998)), decision tree (Quinlan, J. R. (1986)), Gaussian naive bayes (Bayes, T. (1968)),

XGBoost (Chen, T., & Guestrin, C. (2015).), support vector (Cortes, C., Vapnik, V (1995)) da biblioteca python scikit-learn.

A partir dessa seleção todos os algoritmos foram avaliados sob as mesmas métricas, com os mesmos dados de entrada e todos utilizando a parametrização padrão fornecida pela biblioteca scikit-learn.

Por se tratar de dados desbalanceados onde as notas A, B, C e D possuem, respectivamente, 138, 110, 32, 29 registros totalizando 309 registros, foi utilizada a biblioteca imblearn e um pipeline para balancear os dados tanto com o SMOTE (Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002)), que faz um oversample dos dados, ou seja, os dados das classes minoritárias são aumentados para ter mais exemplos dele para o algoritmo treinar, quando utilizando a função RandomUnderSampler, que faz o oposto da oversample, ou seja, os dados das classes majoritária são reduzidos para que tenha uma quantidade melhor balanceada de exemplos para o algoritmo treinar, após isso os algoritmos são submetidos todos ao mesmo método de cross validation, da biblioteca scikit-learn.

A partir destes dados a pipeline criada, resultou em um balanceamento onde todas as notas contém a mesma quantidade de 117 registros.

Em continuidade, às métricas as quais os algoritmos foram submetidos foram Acurácia balanceada (ou em inglês balanced accuracy) e f1 ponderado (ou em inglês f1 weighted), visando encontrar o algoritmo com as métricas mais balanceadas e mais altas, para garantir um melhor desempenho.

Adicionalmente, as métricas foram selecionadas visando um melhor entendimento da performance do modelo, logo, por se tratar de um modelo com dados desbalanceados, por existirem poucos alunos com notas D e C comparado as notas A e B, as métricas comuns poderiam gerar uma falsa sensação de performance, onde, por exemplo, de dez alunos, se oito alunos tiveram notas A e B e dois tiveram notas C e D, caso o modelo só classifique A e B teria uma performance reduzida de somente 20% mesmo excluindo 50% das possibilidades.

Para evitar tal situação, as métricas de acurácia balanceada e f1 ponderado foram selecionadas pois, a acurácia balanceada trabalha com a média da soma de revocação (do inglês recall) de cada classe do modelo o que evita o tipo de situação supracitada pois cada elemento tem sua relevância considerada. Adicionalmente o f1 ponderado segue a mesma lógica, porém utiliza o f1 de cada classe multiplicado por sua proporção no volume dados.

Em complemento, tabela da Figura 36 foi criada para facilitar a avaliação de modo comparativo entre as métricas, a partir desses dados expostos na tabela, o algoritmo que apresentou melhores métricas e mais balanceadas foi o random forest, que foi escolhido como

o algoritmo para esta prova de conceito em específico.

Tabela 1 – Comparação de desempenho dos algoritmos

	Acurácia balanceada	Desvio padrão acurácia	F1 ponderado	Desvio padrão F1
logistic regression	0.748	0.092	0.693	0.102
random forest	0.934	0.114	0.921	0.126
knn	0.851	0.079	0.827	0.086
ridge	0.73	0.056	0.67	0.087
xgboost	0.918	0.077	0.92	0.075
decision tree	0.903	0.054	0.916	0.056
gaussian naive bayes	0.696	0.092	0.672	0.14
support vector machines	0.688	0.08	0.676	0.139
one vs rest	0.781	0.74	0.724	0.93

Fonte: Autoria própria

Ressalta-se a importância como supracitado no capítulo anterior da avaliação dos modelos, pois devido a ampla diversidade nos múltiplos ambientes educacionais com mudança nos padrões dos estudantes os valores dessas métricas podem ser alterados e outro modelo poderá performar melhor.

Em continuidade, visando uma melhoria da performance do modelo, utilizando a função `permutation_importance` também da biblioteca `scikit-learn`, foi extraída a relevância de cada feature inserida no algoritmo `random forest`, como demonstrado na Figura 37, criando uma forma de entender quais os eventos mais impactantes para a tomada de decisão do algoritmo, facilitando a compreensão do resultado do algoritmo.

Figura 36 – Relevância das features

9	Resumo da tentativa do questionário visualizada	-0.003239
16	Tabela de notas visualizada	0.000000
15	O stado do workflow foi atualizado.	0.000000
14	O envio foi avaliado.	0.000000
13	Questões avaliadas manualmente	0.000000
23	Relatório do questionário visualizado	0.000000
8	Tentativa do questionário iniciada	0.000000
24	Item de nota atualizado	0.000000
20	Formulário de confirmação de submissão visuali...	0.000000
21	Formulário de notas visualizado	0.000000
3	Usuário recebeu nota	0.000000
17	Submissão criada.	0.001215
18	Um envio foi submetido.	0.002024
2	Curso visto	0.004049
5	Conclusão da atividade do curso atualizada	0.004453
22	O usuário aceitou o acordo da tarefa.	0.004858
19	Comentário visualizado	0.009717
7	Relatório de notas do usuário visualizado	0.010526
11	Formulário de submissão visualizado.	0.013360
12	Um arquivo foi enviado.	0.025101
0	Tentativa do questionário visualizada	0.028340
4	Tentativa do questionário revisada	0.031579
1	Módulo do curso visualizado	0.037652
10	Tentativa do questionário entregue	0.072065
6	O status da submissão foi visualizado.	0.075304

Fonte: Autoria própria

Adicionalmente, a partir desses novos dados gerados e do modelo treinado, os dados históricos começam a ser explorados, para facilitar essa exploração esses dados foram inseridos em um software de business intelligence, neste caso o Google Data Studio ou também conhecido como Looker studio. Este software foi escolhido pela simplicidade no uso e por ser gratuito, mas poderia ser facilmente substituído por outra solução como Power bi, tableau ou até uma página web integrada ao ambiente moodle.

Para conseguir trabalhar esses dados, da melhor forma, uma base de dados separada foi criada no mesmo formato dos dados históricos para a turma atual. A partir dessa base separada, foi realizada a predição utilizando o algoritmo random forest, obtendo os resultados demonstrados na Figura 38, que foram traduzidos para as possíveis notas de acordo com a rotulação, onde a nota "A" equivale ao número 0, a nota "B" equivale ao número 1, a nota "C" equivale ao número 2 e a nota "D" equivale ao número 3. Este resultado traduzido dá início ao módulo de organização da informação, e esse dado é inserido na tabela da atual turma com o nome de "possível nota" visando diferenciar de um resultado final real como os que constam nos dados históricos.

Figura 37 – Tradução da resposta do algoritmo

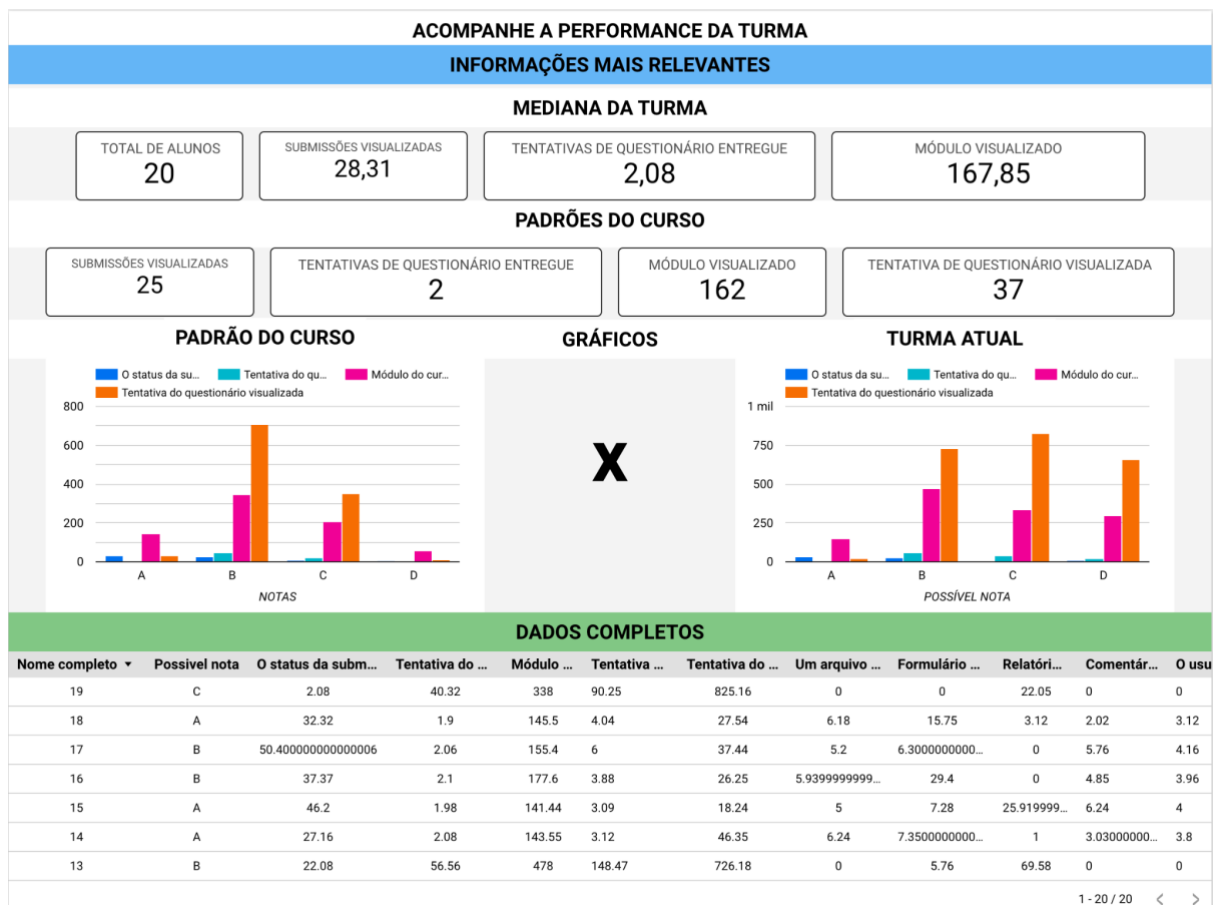
saída do modelo	tradução do resultado do modelo	
0	0	A
1	1	B
2	0	A
3	1	B
4	3	D
5	0	A
6	1	B
7	1	B
8	1	B
9	0	A
10	1	B
11	0	A
12	0	A
13	1	B
14	0	A
15	0	A
16	1	B
17	1	B
18	0	A
19	2	C

Fonte: Autoria própria

Em complemento, com os dados formatados os dados são organizados por nível de importância de acordo com a tabela gerada pelo resultado da função de `permutation_importance`

do algoritmo random forest. Em continuidade, passamos para o módulo de apresentação da informação onde é criada a camada de visualização na plataforma Google Data Studio no formato de um dashboard. Este dashboard foi montado como uma forma de dar uma visão macro do curso, onde visa comparar os padrões da turma atual com os padrões históricos das turmas anteriores, como demonstrado pela Figura 39.

Figura 38 – Dashboard de análise de performance



Fonte: Autoria própria

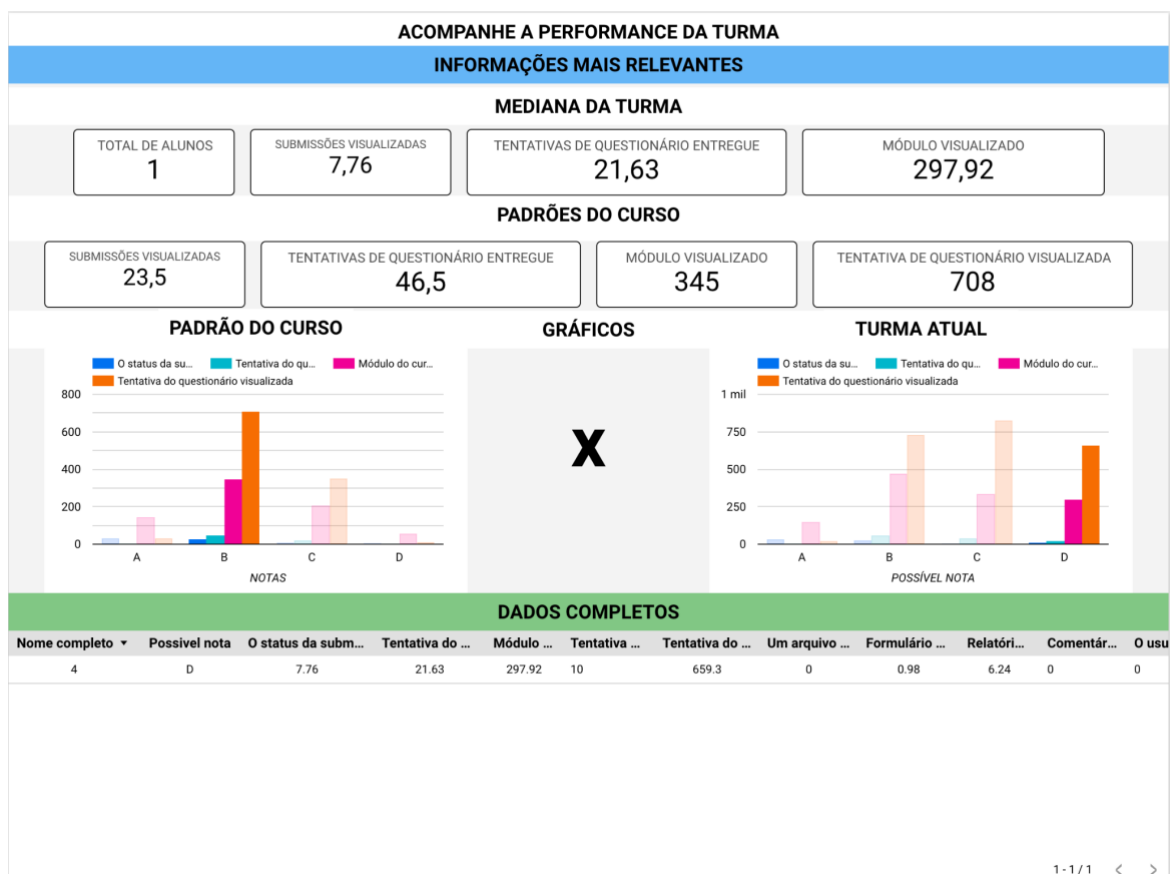
Em complemento, a Figura 39 traz um recorte das medianas dos da turma de acordo com os 3 mais relevantes eventos utilizados pelo algoritmo de aprendizagem de máquina, a partir disso traz gráficos que utilizam os dados gerados pela análise preditiva para que possa ser comparado com os dados históricos, ou seja, uma forma de comparar os eventos mais relevantes dos alunos com o padrão histórico e identificar as semelhanças e diferenças, além de por ser um gráfico dashboard interativo, traz a possibilidade de selecionar os alunos cujo a performance pode vir a ser baixa e ver os dados completos deles na tabela logo abaixo, dando visibilidade para o professor saber onde agir.

Adicionalmente, ressalta-se a importância de adaptar esta camada de acordo com a necessidade dos professores que irão consumi-la, dito que, ambientes educacionais tem diversos perfis de professores com necessidades informacionais diferentes, nem todos os professores podem estar familiarizados com esta forma de apresentação e isto deve ser levado em consideração no momento em que será estruturada a camada de apresentação, tanto para poder prover um melhor treinamento, quanto para a melhor absorção da informação e um uso mais assertivo da ferramenta.

Em continuidade, o último módulo do modelo é a tomada de decisão, este módulo é realizado unicamente pelo professor, a ferramenta tem como objetivo fornecer subsídios informacionais, no quesito de análise de performance estudantil, para que o professor realize a tomada de decisão. O objetivo deste trabalho não é limitar a utilização da ferramenta para apenas algumas decisões, mas fornecer informações enriquecidas para que o professor consiga ter mais subsídios para tomar sua decisão, logo, visando exemplificar o uso da ferramenta, algumas decisões as quais os dados gerados podem subsidiar são, o entendimento dos padrões de comportamento dos alunos do curso na plataforma, a identificação de alunos com possibilidade de baixa performance e a comparação entre o aluno ou um grupo de alunos com os padrões dos ex alunos a fim de identificar e intervir em alunos de baixa performance.

Adicionalmente, como supracitado, o professor pode utilizar essa ferramenta como uma forma de comparação direta entre o aluno e o padrão do curso visando um diagnóstico de onde agir, como exemplificado na Figura 40, os dados indicam que o aluno tem um certo déficit na questão de tentativas de questionário, o que pode vir a ser estudado pelo professor, entrando em contato com o aluno verificando se ele está conseguindo absorver todo o conhecimento, incentivar o aluno a revisar os conteúdos e até mesmo os próprios questionários, entre outras ações para tentar reverter o cenário a favor do estudante a aumentar sua performance.

Figura 39 – Comparação entre aluno e histórico



Fonte: Autoria própria

6.2 Artefatos Informacionais

O processo supracitado de transformação de dados, análise preditiva e descritiva, realizado pela ferramenta, resulta em artefatos informacionais de natureza estruturada, numérica e contínua, para o caso das análises descritivas e estruturada, categórica e ordinal, para o caso do modelo de aprendizagem de máquina. O dado gerado pelo modelo é uma classificação das possíveis notas dos estudantes. Nesse sentido, o contexto educacional onde levamos em consideração que o ambiente informacional é o sistema de gerenciamento de aprendizado e o agente é o professor, que realiza a análise do estudante a partir dos dados disponíveis, podemos considerar que, todo o processo de processamento de dados realizado pela ferramenta, gerou artefatos informacionais, no âmbito deste trabalho, avaliação de performance.

Adicionalmente, esses artefatos se aliados a interface gráfica, criam estruturas

interpretativas para o agente, a partir dessa união, interpretar a informação, interpretação essa que é guiada por intermédio da interface gráfica, visando deixar claro o objetivo da ferramenta de fornecer subsídios informacionais sobre os padrões de comportamento dos alunos do curso dentro da plataforma e prever a possível performance do aluno. Em complemento, os artefatos informacionais gerados são, explicitamente, as informações sobre os padrões históricos dos alunos, as informações sobre os padrões atuais do curso e a informação sobre a futura performance do aluno.

6.3 Discussão

A preocupação com a tomada de decisão acadêmica no quesito de avaliação de performance não se resume somente a ter mais subsídios informacionais. Os dados dentro do processo evoluem à medida que ele avança, mas também podem sofrer transformações ao longo do tempo. Desta maneira devido às mudanças contínuas nos ambientes educacionais, há necessidade de gestão, desde os dados coletados, da informação final e como ela será apresentada ao na camada de visualização até treinamento e capacitação dos profissionais que vão consumir essa informação

Adicionalmente, devido aos múltiplos cenários existentes no ambiente educacional, além de inúmeras inovações e customizações dos sistemas de gerenciamento da aprendizagem, seja por meio de gamificação ou a própria evolução da plataforma, como o moodle que está em sua versão 4.1.1 e tem uma ativa comunidade que cria inúmeras extensões para a plataforma, o modelo necessita ser adaptado para a necessidade do ambiente inserido, modificado para atender a necessidade do professor, respeitando o ambiente, ou seja, caso o ambiente forneça melhores dados para o modelo trabalhar, eles devem ser utilizados, mesmo que isso signifique modificar o modelo.

Em complemento, os módulos do modelo performi, apesar de ter dependência de dados e necessitar da execução na ordem proposta pelo modelo, cada parte do módulo fornecem dados que podem auxiliar na melhoria do próprio modelo, por se tratar de um modelo auditável onde todas as etapas podem ser revisadas e modificadas para atingir a necessidade informacional, durante a gestão do modelo podem surgir melhorias a partir de cada módulo, seja pelo decaimento da performance do modelo ou por um evento novo inserido que pode alterar os resultados tanto preditivos quanto trazer melhor entendimento nas análises preditivas.

Em contraponto, o modelo apesar de fornecer informações enriquecidas e análises descritivas e preditivas, ainda depende da capacitação do professor para conseguir interpretar

as informações produzidas. Apesar de ser um provedor de informações, sozinho ainda é limitado, por não compreender todo o fluxo informacional necessário para a tomada de decisão, mesmo no quesito análise de performance, outros fatores externos e muitas vezes sensíveis, ainda precisam ser levados em conta e estes não compreendem e não competem a proposta do modelo, apesar de influenciar indiretamente nos dados que o modelo utiliza para avaliação.

Em complemento, apesar do modelo trazer apenas dados baseados na interação com a plataforma e não ter dados sensíveis como, gênero, classe social, etnia ou outros dados sensíveis, a informação produzida é uma informação sem nenhum enviesamento discriminatório, mas esta informação ainda é um subsídio a tomada de decisão e deve ser tratada como tal, deve ser associada a outras informações visando tornar a decisão algo humano, ou seja, o agente humano professor deve discernir e interpretar as informações para a tomada de decisão. Ressalta-se a importância da avaliação da informação produzida nesse aspecto, caso os dados utilizados estejam gerando informações discriminatórias ou informações que associadas a outro dado gerem alguma forma de discriminação, o modelo deve ser revisto e os dados de treinamento devem ser repensados a fim de evitar quaisquer enviesamento.

Conclui-se que o modelo é um provedor de informação que atende ao objetivo que se propõe de fornecer subsídios informacionais para avaliação de performance, mas necessita de constante gerenciamento e capacitação para melhoria contínua e manutenção da informação produzida.

Adicionalmente, sob o ponto de vista do professor tomador de decisão, esta ferramenta altera seu fluxo informacional quando se trata de avaliação de performance discente, trazendo um ponto centralizador com inúmeros dados ricos sobre o discente e sobre o comportamento geral do curso, para que o professor consiga se ambientar sobre o passado o presente e entender antecipadamente qual pode ser o futuro da performance do discente e assim aprimorar o aproveitamento dos estudante por conseguir adquirir mais rápido a consciência situacional necessária para tomar a decisão.

No próximo capítulo serão apresentadas as conclusões do trabalho frente aos objetivos, hipóteses e contribuições do trabalho.

7 CONCLUSÕES

Os sistemas de gerenciamento do aprendizado como o Moodle são complexos ambientes que geram muitos dados para os tomadores de decisão, porém quando se trata de análise de performance os dados são espúrios. Por se tratar de um sistema nascido em um ambiente de tecnologias da informação e comunicação, naturalmente há mais dados estruturados a serem explorados, porém o desafio está em transformá-los em informações, devido a sua natureza dos dados e as habilidades necessárias para a interpretação dos dados, além do escopo crítico que a educação se encontra, em que os resultados afetam diretamente o futuro do aluno. O processamento da informação é considerado uma forma de se extrair informações de dados espúrios e de difícil entendimento, no caso deste trabalho, utilizando auxílio computacional e a estatístico como ferramentas para gerar informações úteis.

O estado da arte apresenta modelos de análise descritiva e de análise preditiva, que auxiliam na geração de mais dados para este ambiente, orientados a performance computacional ou a descrição de cenários educacionais específicos. A literatura apresenta poucos modelos de processamento da informação no contexto educacional, menos ainda com modelos orientados à necessidade informacional visando subsidiar a tomada de decisão pedagógica.

Este trabalho apresentou um modelo de ferramenta de processamento da informação para suporte à avaliação de performance e tomada de decisão no processo de ensino-aprendizagem de EAD, que visa subsidiar a tomada de decisão com informações quali-quantitativas sobre a performance do aluno. O modelo Performi "Performance estudantil fomentada por processamento informacional" tem como pilar a exploração dos dados espúrios dos registros de LOG da plataforma moodle, trazendo análises dos padrões de comportamento dos alunos e informações contextualizadas sobre o passado do curso o presente a possível performance do aluno de acordo com os cenários expostos.

Os resultados da prova de conceito indicaram que o modelo consegue extrair dados quali-quantitativos dos registros de LOG e transformá-los em informações contextualizadas e apresentá-las de forma organizada a descrever e guiar a interpretação da informação podendo ser utilizados como subsídio informacional para a tomada de decisão.

7.1 Atendimento aos Objetivos de Pesquisa

Os resultados indicaram o atendimento dos objetivos de pesquisa para as proposições:

- Definir um processo informacional a partir do LMS Moodle, que descreva a estrutura dos dados, as transformações e manipulações que serão feitas e gerar uma informação contextualizada proveniente de uma análise preditiva;

Neste sentido, o modelo Performi realiza o mapeamento de todo o processo necessário para transformar os dados em informação, descrevendo a aquisição dos dados, as estruturas deles, quais transformações devem ser feitas e qual direção os dados devem tomar, além de descrever como contextualizar e interpretar as informações geradas, o que garante um processo informacional bem definido.

O modelo performi segue esse mapeamento com em seus módulos, onde separa cada processo necessário para o processamento da informação e descreve as necessidades informacionais de cada etapa, visando guiar o desenvolvimento de uma ferramenta que compreenda todos os aspectos listados acima e descreve as necessidades informacionais visando aumentar a compreensão de cada processo para que possam ser aprimorados de acordo com o contexto.

- Especificar as limitações identificadas no processo de tomada de decisão educacional quanto ao uso de dados históricos para auxílio na avaliação de performance no LMS Moodle;

A fim de identificar as limitações quanto ao uso de dados históricos, a pesquisa realizada pelo grupo de pesquisa GIHC (Grupo de Interação Humano-Computador), descreve como a tomada de decisão utilizando os dados históricos é pouco utilizada ou utilizada de forma operacional pelos professores, evidenciando que a maior limitação quanto ao uso de dados históricos é em sua maior parte por falta de conhecimento ou falta de informação na plataforma.

Adicionalmente, o capítulo de trabalhos relacionados, traz uma série de relatos sobre a ferramentas subutilizadas, onde trazem motivos como, dificuldade com a plataforma, dados complexos e até operacionalização da ferramenta, onde os dados são utilizados, mas sem entendimento do motivo, ou do objetivo deles, somente sendo utilizado por hábito ou obrigação.

Conclui-se que, existem limitações tanto quanto a falta de informações, quanto ao processo em si que pode ter informações, mas são difíceis de se interpretar e consumir, até a ter a ferramenta, mas o utilizador não entender a relevância dela no seu processo decisório.

- Identificar técnicas de aprendizado de máquina em contextos educacionais já utilizadas;

A pesquisa de trabalhos correlatos, traz uma série de aplicações de algoritmos e técnicas

de aprendizagem de máquina no contexto educacional e os analisa visando entender o atual estado da arte e avançar no sentido da informação com as melhores ferramentas.

Neste sentido, após o levante bibliográfico, foram selecionados os algoritmos de aprendizagem de máquina do tipo supervisionado classificatórios, visando um processo mais facilmente interpretável e auditável, além de se alinharem melhor com o problema de gerar um dado categórico para cada aluno, sua nota.

- Avaliar e aplicar os algoritmos e técnicas de aprendizado de máquina a fim de identificar os modelos que melhor se adaptam ao problema;

A prova de conceito realizada, compara algoritmos de aprendizagem de máquina visando obter os algoritmos com a performance mais adequada para resolução do problema, utilizando métricas avaliativas iguais para todos os modelos e selecionando o algoritmo com os resultados mais equilibrados. A partir da análise dessas métricas foi selecionado o algoritmo random forest, por se tratar do algoritmo com o melhor resultado em relação aos dados utilizados.

- Desenvolver uma prova de conceito com o LMS Moodle para validação da informação produzida

A prova de conceito no capítulo 6 foi desenvolvida a fim de validar o modelo proposto e a avaliar a informação final produzida. Com a prova de conceito, a partir da transformação dos dados, foi possível comprovar a possibilidade de transformar a base de registros em uma base de dados orientada ao usuário onde é possível entender os padrões do aluno.

Em complemento, foi possível comprovar a efetividade do algoritmo de aprendizagem de máquina para prever a possível nota do aluno com um algoritmo interpretável, com a junção dos dados da base sobre o usuário com os dados preditivos do algoritmo de aprendizagem de máquina é possível comparar com os dados históricos e ter uma informação enriquecida e contextualizada.

Adicionalmente, estes dados produzidos podem ser anexados ao moodle como um relatório do curso, junto aos outros relatórios de curso para o professor, como um link para um dashboard criado exclusivamente para aquele curso visando dar subsídio ao professor sobre a performance dos alunos em geral naquele curso. Assim o professor teria mais um subsídio dentro da plataforma para acompanhar sua turma, a evolução, compará-la com turmas anteriores e dar um parecer mais claro e objetivo para o aluno sem a necessidade de ter habilidades técnicas para criar tal ferramenta.

Neste sentido, o professor ou gestor, fica responsável por criar e expor as suas necessidades informacionais, dando feedback ao time de gestão da ferramenta, para que então essas necessidades sejam traduzidas para a ferramenta e o time de gestão da ferramenta fica responsável pela parte técnica de gerir os dados e as informações geradas, adequando a informação final para que seja corretamente interpretada pelos professores e gestores.

Em contrapartida, não é possível apenas utilizar essa ferramenta sem uma gestão contínua, ou seja, sem nenhuma alteração ou sem alguém com habilidades técnicas para manter tal ferramenta, pois, como o ambiente educacional é muito amplo e diversificado, sem a devida adaptação ao cenário o qual ele será inserido e o devido acompanhamento, pode vir a ter uma má performance, fornecer dados incorretos e imprecisos. Nesse sentido se faz necessária uma equipe técnica para orquestrar e organizar os processos internos da ferramenta, seguindo as diretrizes criadas pelo modelo performi e atendendo as necessidades expostas pelos professores.

Adicionalmente, não foi possível confirmar a real efetividade da informação produzida em um ambiente educacional e qual seria o seu impacto em relação a melhoria de performance ao longo de turmas em que os professores tiveram acesso a ferramenta, em comparação com resultados de turmas sem o auxílio da ferramenta.

7.2 Limitações da Pesquisa

Este trabalho apresenta como limitações:

Como evidenciado no capítulo 5 o modelo busca avaliar a performance do aluno a partir dos padrões de comportamento do aluno dentro da plataforma e a partir desses padrões inferir uma possível performance do aluno e descrever o comportamento geral a partir desses dados, isso limita a quantidade de informações produzidas pelo modelo, logo, o modelo realiza apenas uma parte do fluxo informacional da tomada de decisão, fornecendo subsídios informacionais apenas sobre os padrões contidos nos registros de LOG.

Adicionalmente apesar do modelo ser adaptável e ter como objetivo fornecer subsídios informacionais, sem a devida gestão e o devido treinamento dos professores, os dados os quais o modelo fornece são dados complexos e podem gerar uma operacionalização do modelo, de forma a trazer um uso deste sem o devido entendimento, já que o modelo altera o fluxo de obtenção de informação do professor adicionando mais uma etapa antes da tomada de decisão e por se tratar de um modelo preditivo que traz uma informação confiável em sua maioria, sem a devida orientação pode vir a ser subutilizado ou utilizado da forma errada.

7.3 Contribuições da Pesquisa

Este trabalho traz como contribuições:

- Um modelo, implementação e arquitetura de uma ferramenta de processamento da informação, para guiar o desenvolvimento de ferramentas para análise de performance estudantil na plataforma moodle.
- Um processo informacional que descreve a estrutura dos dados e da informação no processo de avaliação de performance estudantil.
- Uma estrutura de dados padronizada e centrada no estudante, que mapeia os padrões de interação do estudante com a plataforma.
- Um modelo auditável que traz informações organizadas e ordenadas pela relevância definida pelo próprio modelo ao encontrar padrões.
- Um dashboard interativo que traz dados históricos e atuais das turmas, com comparativos e ordenados conforme os dados considerados mais relevantes pela análise preditiva do algoritmo de aprendizagem de máquina.
- Uma informação preditiva contextualizada por análises descritivas do cenário e transparente ao organizar os dados por relevância.
- Adiciona um ponto centralizador com inúmeros dados ricos sobre o discente e sobre o comportamento geral do curso, para que o professor consiga ter uma melhor visão do comportamento dos alunos e dar um parecer melhor ao aluno sobre sua performance.

7.4 Trabalhos Futuros

Como trabalhos futuros é possível:

- Criar um modelo de visualização dos dados para melhor compreensão dos resultados. Um modelo de visualização traria maior interpretabilidade para as informações disponibilizadas e aumentaria a probabilidade do sujeito de utilizar a ferramenta.
- Criar modelos que suportem outras decisões educacionais. Outros modelos podem ser associados a este e aumentar a qualidade e a quantidade dos subsídios informacionais no fluxo de tomada de decisão pedagógica.
- A evolução do modelo para contemplar outras formas de análise, como análise de sentimentos e a partir disso gerar insumos informacionais não somente sobre o

comportamento dos alunos, mas também sobre o interesse e a satisfação do aluno com o curso.

- Implementar o modelo diretamente conectado ao ambiente moodle, de forma a deixar dentro da plataforma o fluxo informacional completo partindo da aquisição de informação sobre a performance do aluno até a tomada de decisão.
- Medir a efetividade do modelo comparando a assertividade de casos de avaliação com e sem o modelo em ambiente reais, visando a melhoria dos processos e da informação, além da identificação de possíveis gaps dentro do processo de geração da informação.

REFERÊNCIAS

- ACKOFF, R. L. From data to wisdom. **Journal of applied systems analysis**, v. 16, n. 1, p. 3-9, 1989.
- ARAÚJO, C. A. A. Correntes teóricas da ciência da informação. **Ciência da informação**, v. 38, p. 192-204, 2009.
- AVELLA, J. T. et al. Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. **Online Learning**, v. 20, n. 2, p. 13-29, 2016.
- BAYES, T. Naive bayes classifier. **Article Sources and Contributors**, 1-9, 1968.
- BELLONI, M. L. **Educação a distância**. 4.ed. Campinas: Editora Autores Associados, 2006
- BENESTY, J. et al. Pearson correlation coefficient. In: **Noise reduction in speech processing**. Springer, Berlin: Heidelberg, 2009. p. 1-4.
- BOGNÁR, L.; FAUSZT, T. Different learning predictors and their effects for Moodle Machine Learning models. In: **2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)**. IEEE, p. 405-410, 2020.
- BRANCO, L. S. A.; CONTE, E.; HABOWSKI, A. C. Evasão na educação a distância: pontos e contrapontos à problemática. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 25, n. 1, p. 132–154, 2020.
- BRANDL, K. Review of are you ready to “Moodle”? **Language Learning & Technology**, v. 9, n. 2, p. 16-23, 2005.
- BREIMAN, L. Random forests. **Machine learning**, 45, p. 5-32, 2001.
- CARVALHO, D. V.; PEREIRA, E. M.; CARDOSO, J. S. Machine learning interpretability: A survey on methods and metrics. **Electronics**, v. 8, n. 8, p. 832, 2019.
- CHAUHAN, N. Prediction of student's performance using machine learning. In: **2nd International Conference on Advances in Science & Technology (ICAST)**. 2019.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, 16, 321-357, 2002.
- CHEN, T.; GUESTRIN, C. Xgboost: Reliable large-scale tree boosting system. In: **Proceedings of the 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA**, 2015, p. 13-17.
- CRAMER, J. S. The origins and development of the logit model. **Logit models from economics and other fields**, 2003, 1-19.
- DA SILVA, D. G.; ALONSO, K. M.; MACIEL, C. Trajetórias de Aprendizagem, Participação e Interação: Possibilidades dos Relatórios do Moodle. In: **SIED: EnPED-**

Simpósio Internacional de Educação a Distância e Encontro de Pesquisadores em Educação a Distância, 2012.

ECKERSON, W. W. Predictive analytics. Extending the Value of Your Data Warehousing Investment. **TDWI Best Practices Report**, v. 1, p. 1-36, 2007.

ENDSLEY, M. R. Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors Society annual meeting. Sage CA: Los Angeles, CA: **SAGE Publications**, p. 97-101, 1988.

FILATRO, A. Learning analytics: Análise e desempenho do ensino e aprendizagem. **Senac**, 2019.

FILATRO, A. Data Science na Educação: presencial, a distância e corporativa. **Saraiva Educação SA**, 2020.

FIX, E.; HODGES JR, J. L. **Discriminatory analysis-nonparametric discrimination**: Small sample performance. California: Univ Berkeley, 1952.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Editora Atlas, 1999.

HOODA, M.; RANA, C. Learning analytics lens: Improving quality of higher education. **International journal of emerging trends in engineering research**, v. 8, n. 5, 2020.

JAYASINGHE, U.; DHARMARATNE, A.; ATUKORALE, A.. Students' performance evaluation in online education system Vs traditional education system. In: **Proceedings of 2015 12th International Conference on Remote Engineering and Virtual Instrumentation (REV)**. IEEE, 2015. p. 131-135.

JÚNIOR, J. B. D. **META-MOOC: UMA FERRAMENTA PARA GERAÇÃO DE MOOCS ADAPTATIVOS E PERSONALIZÁVEIS**. [s.l.] Faculdade de Engenharia Elétrica – UFU, Uberlândia- MG, 2017.

KOKAR, M. M.; ENDSLEY, M. R. Situation awareness and cognitive modeling. **IEEE Intelligent Systems**, v. 27, n. 3, p. 91-96, 2012.

KRISHNAN, R. et al. Smart Analysis of Learners Performance Using Learning Analytics for Improving Academic Progression: A Case Study Model. **Sustainability**, v. 14, n. 6, p. 3378, 2022.

LE COADIC, Y. F. **A ciência da informação**. Brasília: Briquet de Lemos Livros, 1996

MARCONI, M. A.; LAKATOS, E. M. **Técnicas de pesquisa: planejamento e execução de pesquisa; amostragens e técnicas de pesquisa; elaboração, análise e interpretação de dados**, 2012.

MARGARYAN, A.; BIANCO, M.; LITTLEJOHN, A. Instructional quality of Massive Open Online Courses (MOOCs). **Computers and Education**, v. 80, p. 77–83, 2015.

MASCARENHAS, S. A. Metodologia Científica. São Paulo: **Pearson Education do Brasil**,

2012.

MITCHELL, T. et al. Machine learning. **Annual review of computer science**, v. 4, n. 1, p. 417-433, 1990.

MUELLER, J. P.; MASSARON, L.. **Aprendizado de Máquina para Leigos. Alta Books, 2019.**

NAMOUN, A.; ALSHANQITI, A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. **Applied Sciences**, v. 11, n. 1, p. 237, 2020.

NAVARRO, F. P.; CONEGLIAN, C. S.; SEGUNDO, J. E. S. Big data no contexto de dados acadêmicos: o uso de machine learning na construção de sistema de organização do conhecimento. **Informação & Tecnologia**, v. 5, n. 2, p. 181-200, 2018. DOI: 10.22478/ufpb.2358-3908.2018v5n2.45807

NEVES, B. C. Inteligência artificial e computação cognitiva em unidades de informação: conceitos e experiências. **Logeion: filosofia da informação**, v. 7, p. 186-205, 2020. DOI: 10.21728/logcion.2020v7n1.p186-205

OKUBO, F. et al. A neural network approach for students' performance prediction. In: **Proceedings of the seventh international learning analytics & knowledge conference**. 2017. p. 598-599.

OLIVEIRA, M.; BERTUCCI, M. G. E. S. A pequena e média empresa e a gestão da informação. **Informação e Sociedade**, João Pessoa, v.13, n.2., p. 1-16, 2003. Disponível em: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/91>.

PAIVA, R. O. A. **AUTORIA DE DECISÕES PEDAGÓGICAS INFORMADAS POR DADOS SOB A PERSPECTIVA DE UM MOOC**. [s.l.] Universidade Federal de Campina Grande - UFCG, 2017.

PINNELL, C. et al. Curricular and learning analytics: a big data perspective. In: **Big Data and Learning Analytics in higher education**. Springer, Cham, 2017. p. 125-145.

PIOTROSKI, J. D. Value investing: The use of historical financial statement information to separate winners from losers. **Journal of Accounting Research**, p. 1-41, 2000.

PULINO FILHO, A. R. Um sistema de gerenciamento de cursos. Brasília, DF: **UNB**, 2005. Disponível em: <http://www4.tce.sp.gov.br/sites/default/files/manual-completo-moodle.pdf>.

QUINLAN, J. R. Induction of decision trees. **Machine learning**, 1, 81-106, 1986.

RENDÓN ROJAS, M. A. Relación entre los conceptos: información, conocimiento y valor. Semejanzas y diferencias. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 52-61, maio/ago. 2005.

RUSSEL, S.; NORVIG, P. **Artificial intelligence: a modern approach** (Vol. 256). London: Pearson Education Limited, 2013.

SALLOUM, S. A. et al. Mining in educational data: review and future directions. In: **The International Conference on Artificial Intelligence and Computer Vision**. Springer, Cham, 2020. p. 92-102.

SANTANA, R. C. G. **Ciclo de vida dos dados e o papel da Ciência da Informação**, 2013
Disponível em: <http://hdl.handle.net/20.500.11959/brapci/185240>.

SANTANA, M. A.; DOS SANTOS NETO, B. F.; DE BARROS COSTA, E. Avaliando o uso das ferramentas educacionais no ambiente virtual de aprendizagem moodle. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**, 2014. p. 278.

SANTOS, J. R. et al. Estudo comparativo de plugins Moodle para Análise e Acompanhamento da Aprendizagem. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**, 2019. p. 189.

SAUNDERS, C., GAMMERMAN, A., VOVK, V. **Ridge regression learning algorithm in dual variables**, 1998.

SHAHIRI, A. M. et al. A review on predicting student's performance using data mining techniques. **Procedia Computer Science**, v. 72, p. 414-422, 2015.

SILVA, A. G.; ANDRADE, L.C.; BARTOLOMEI SILVA, M.. Educação à Distância: As novas tecnologias e o papel do tutor na perspectiva da construção do conhecimento. Relatório de Pesquisa. Ensino e Aprendizagem em EAD. **Congresso 2011**. Campo Grande -MS.2011. Disponível em: <http://www.abed.org.br/congresso2011/cd/46.pdf>.

SCHILDKAMP, K. Data-based decision-making for school improvement: Research insights and gaps. **Educational research**, v. 61, n. 3, p. 257-273, 2019.

SCHRAMM, W. **Notes on Case Studies of Instructional Media Projects**, 1971.

VOSS, C.; TSIKRIKTSIS, N.; FROHLICH, M. Case research in operations management. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 195-219, 2002.

WANG, A. et al. **Dismantling the Data Ivory Tower: Empowering Teachers with Data Driven Decision Making**.

WU, W.; GAO, X.; GAO, S. One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. In **2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (pp. 2387-2390)**, 2006.

YAZAN, B.; DE VASCONCELOS, I. C. O. Três abordagens do método de estudo de caso em educação: Yin, Merriam e Stake. **Meta: Avaliação**, v. 8, n. 22, p. 149–182, 2016.

YIN, R. K. **Estudo de Caso Planejamento e Métodos**. 5. ed. ed. Porto Alegre: Bookman, 2015.