

**UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC**  
**CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT**  
**TRABALHO DE CONCLUSÃO DE CURSO – CIÊNCIA DA COMPUTAÇÃO**

**DANIELLA MARTINS VASCONCELLOS**

**DETECÇÃO DE PADRÕES EM RELAÇÃO À PRESENÇA DOS ESTUDANTES NO**  
**ENSINO BÁSICO BRASILEIRO**

**JOINVILLE**

**2023**

**DANIELLA MARTINS VASCONCELLOS**

**DETECÇÃO DE PADRÕES EM RELAÇÃO À FREQUÊNCIA DOS ESTUDANTES NO  
ENSINO BÁSICO BRASILEIRO**

Trabalho de Conclusão de Curso I submetido à  
Universidade do Estado de Santa Catarina como  
parte dos requisitos para a obtenção do grau de  
Bacharel em Ciência da Computação

Orientadora: Isabela Gasparini

Coorientadora: Elaine Harada Teixeira de Oli-  
veira (UFAM)

**JOINVILLE**

**2023**

Para gerar a ficha catalográfica de teses e  
dissertações acessar o link:  
<https://www.udesc.br/bu/manuais/ficha>

Martins Vasconcellos, Daniella

Detecção de padrões em relação à frequência dos  
estudantes no ensino básico brasileiro / Daniella  
Martins Vasconcellos. - Joinville, 2023.

48 p. : il. ; 30 cm.

Orientadora: Isabela Gasparini.

Coorientadora: Elaine Harada Teixeira de Oliveira.

Trabalho de Conclusão de Curso - Universidade  
do Estado de Santa Catarina, Centro de Ciências  
Tecnológicas, Programa de Graduação em Ciência da  
Computação, Joinville, 2023.

1. Frequência escolar. 2. Educação básica. 3.  
Mineração de dados educacionais. 4. Análise de padrões.  
5. Evasão. I. Gasparini, Isabela . II. Teixeira de  
Oliveira, Elaine Harada . III. Universidade do Estado  
de Santa Catarina, Centro de Ciências Tecnológicas,  
Programa de Graduação em Ciência da Computação. IV.  
Detecção de padrões em relação à presença dos estudantes  
no ensino básico brasileiro.

**DANIELLA MARTINS VASCONCELLOS**

**DETECÇÃO DE PADRÕES EM RELAÇÃO À FREQUÊNCIA DOS ESTUDANTES NO  
ENSINO BÁSICO BRASILEIRO**

Trabalho de Conclusão de Curso I submetido à  
Universidade do Estado de Santa Catarina como  
parte dos requisitos para a obtenção do grau de  
Bacharel em Ciência da Computação

Orientadora: Isabela Gasparini

Coorientadora: Elaine Harada Teixeira de Oli-  
veira (UFAM)

**BANCA EXAMINADORA:**

Isabela Gasparini – Doutorado em Computação  
Universidade do Estado de Santa Catarina

Membros:

Elaine Harada Teixeira de Oliveira – Doutorado em Informática na Educação  
Universidade Federal do Amazonas

Debora Cabral Nazario – Doutorado em Engenharia e Gestão do Conhecimento  
Universidade do Estado de Santa Catarina

Rebeca Schroeder Freitas – Doutorado em Informática  
Universidade do Estado de Santa Catarina

Joinville, 6 de julho de 2023

Aos meus amigos, pela inspiração e pela  
confiança.

## **AGRADECIMENTOS**

Agradeço às minhas orientadoras, Isabela Gasparini e Elaine Harada Teixeira de Oliveira, por aceitarem conduzir o meu trabalho de pesquisa.

A todos os meus amigos que estiveram ao meu lado durante essa longa trajetória, me dando forças para continuar seguindo em frente.

Ao meu grupo de iniciação científica, Grupo de Pesquisa em Informática na Educação (GPIE), em especial aos colegas também orientados pela professora Isabela Gasparini, que tiveram grande impacto na construção da minha produção científica, a qual se inclui esse trabalho de conclusão de curso.

Deixo um agradecimento especial à minha orientadora, Isabela Gasparini, pelo incentivo, pela dedicação do seu escasso tempo ao meu projeto de pesquisa, e principalmente pela confiança que depositou em mim durante os últimos anos.

*“I never am really satisfied that I understand anything; because, understand it well as I may, my comprehension can only be an infinitesimal fraction of all I want to understand about the many connections and relations which occur to me, how the matter in question was first thought of or arrived at.” (ADA LOVELACE)*

## RESUMO

A educação deveria ser uma prioridade no Brasil, e o Ministério da Educação (MEC) vem implementando políticas visando aprimorar o ensino no país. A coleta de dados dos cidadãos que utilizam serviços educacionais é crucial para o gerenciamento eficiente de recursos e atendimento às necessidades da população. Esses dados são amplamente utilizados por pesquisadores para compreender o sistema educacional e desenvolver programas para melhorar a qualidade da educação. A ciência de identificação de padrões é uma abordagem valiosa para analisar esses dados complexos, permitindo previsões precisas e decisões em evidências. Nesse contexto, o objetivo deste estudo é identificar padrões de presença em sala de aula de estudantes do ensino básico no Brasil, abrangendo a pré-escola, ensino fundamental e médio. Usando aprendizado de máquina e análise de séries temporais, serão identificados padrões de frequência escolar ao longo do tempo. Espera-se que essa análise forneça perspectivas sobre fatores relacionados à evasão escolar e contribua para o desenvolvimento de políticas e programas eficazes.

**Palavras-chave:** Frequência escolar. Educação básica. Mineração de dados educacionais. Reconhecimento de padrões. Evasão.



## **ABSTRACT**

Education should be a priority in Brazil, and the Ministry of Education (MEC) has been implementing policies to enhance education in the country. The collection of data from citizens who use educational services is crucial for efficient resource management and meeting the population's needs. Researchers widely utilize this data to understand the educational system and develop programs to improve the quality of education. The science of pattern identification is a valuable approach for analyzing these complex data, enabling accurate predictions and informed decision-making. In this context, the objective of this study is to identify patterns of classroom attendance among students in basic education in Brazil, including preschool, elementary, and high school. Using machine learning and time series analysis, patterns of school attendance over time will be identified. It is expected that this analysis will provide insights into factors related to school dropout and contribute to the development of effective policies and programs.

**Keywords:** School attendance. Basic education. Educational data mining. Pattern analysis. Evasion.

## **LISTA DE ILUSTRAÇÕES**

Figura 1 – Histórico de frequência de estudantes no Brasil entre os meses de fevereiro e março, e de abril e maio de 2019. . . . .	36
--	----

## **LISTA DE TABELAS**

Tabela 1 – Planejamento de tarefas para o TCC 2 . . . . .	42
---	----

## **LISTA DE ABREVIATURAS E SIGLAS**

DM	Data Mining
EDM	Educational Data Mining
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Educação e Pesquisa Anísio Teixeira
LA	Learning Analytics
MEC	Ministério da Educação
LMS	Learning Management Systems
NEES	Núcleo de Excelência em Tecnologias Sociais
PNE	Plano Nacional de Educação
ProUni	Portal Único de Acesso ao Ensino Superior
TCC	Trabalho de Conclusão de Curso
UDESC	Universidade do Estado de Santa Catarina
UFAL	Universidade Federal de Alagoas
UFAM	Universidade Federal do Amazonas

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>13</b>
1.1	TED 11476 . . . . .	15
1.2	OBJETIVO GERAL . . . . .	16
1.3	OBJETIVOS ESPECÍFICOS . . . . .	16
1.4	METODOLOGIA . . . . .	17
1.5	ESTRUTURA DO TRABALHO . . . . .	17
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>18</b>
2.1	PRESENÇA E FREQUÊNCIA . . . . .	18
2.2	EDUCATIONAL DATA MINING (EDM) . . . . .	19
<b>2.2.1</b>	<b>Learning Analytics (LA) . . . . .</b>	<b>21</b>
<b>2.2.2</b>	<b>Conexão entre EDM e LA . . . . .</b>	<b>23</b>
2.3	RECONHECIMENTO DE PADRÕES . . . . .	24
<b>2.3.1</b>	<b>Técnicas . . . . .</b>	<b>25</b>
<b>2.3.2</b>	<b>Visualização da informação . . . . .</b>	<b>27</b>
2.4	DISCUSSÕES SOBRE VALORES HUMANOS RELACIONADOS . . . . .	28
<b>2.4.1</b>	<b>Ética em <i>Learning Analytics</i> . . . . .</b>	<b>29</b>
<b>2.4.2</b>	<b>Coleta de dados por sistemas de frequência . . . . .</b>	<b>29</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>32</b>
3.1	INICIATIVAS NO EXTERIOR . . . . .	33
3.2	INICIATIVAS NO BRASIL . . . . .	35
3.3	ANÁLISE COMPARATIVA . . . . .	38
<b>4</b>	<b>PROPOSTA . . . . .</b>	<b>40</b>
4.1	CONCLUSÃO . . . . .	41
	<b>REFERÊNCIAS . . . . .</b>	<b>43</b>

## 1 INTRODUÇÃO

O sistema educacional brasileiro tem sido alvo de várias políticas públicas nas últimas décadas, visando garantir o acesso, a permanência e a qualidade do ensino em todas as etapas e modalidades de ensino. Entre as políticas mais relevantes estão o Plano Nacional de Educação (PNE) (BRASIL, 2001), o Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (FUNDEB) (BRASIL, 2020), e o Programa Nacional de Alimentação Escolar (PNAE) (BRASIL, 1955). O Ministério de Educação (MEC) é o órgão responsável por coordenar e executar as políticas públicas voltadas para a educação no país. Em sua história, o ministério tem atuado de forma significativa na definição das políticas educacionais, na formulação de diretrizes curriculares, na elaboração de programas e materiais didáticos, no estabelecimento de critérios de avaliação e na capacitação de professores.

Como os programas governamentais do MEC têm como objetivo o fornecimento de serviços públicos essenciais para atender a variadas necessidades educacionais da população, é fundamental que se colete informações e dados sobre os cidadãos que usufruem dos serviços para melhor utilização de recursos e retorno à população. Por exemplo, o Portal Único de Acesso ao Ensino Superior (ProUni) acessa dados socioeconômicos dos estudantes que estão em bases governamentais para poder analisar possíveis dispensas de apresentação de documentos (como comprovação de renda familiar mensal bruta ou a situação da pessoa com deficiência) (BRASIL, 2005). Dentre outras políticas públicas implementadas pelo governo brasileiro que estão ligadas diretamente ao MEC, o Bolsa Família (BRASIL, 2021) se destaca por ser um programa de transferência de renda que tem como objetivo promover a redução da pobreza e da desigualdade social. O programa foi criado em 2004 (BRASIL, 2004), e tem como uma das principais condições para o recebimento do benefício a frequência escolar dos filhos. Ou seja, para receber o benefício, as famílias precisam se comprometer a manter seus filhos na escola e garantir a frequência escolar mínima estabelecida pelo MEC, que varia de acordo com a faixa etária dos alunos. A coleta da frequência escolar é realizada atualmente pelas escolas em que os alunos estão matriculados, e os dados são repassados ao MEC para fins de monitoramento e avaliação do programa através do Sistema Presença<sup>1</sup>.

A coleta de dados de frequência escolar do Bolsa Família é importante porque permite ao governo monitorar a frequência dos alunos e garantir que as famílias estejam cumprindo seus compromissos. Além disso, essa coleta de dados é uma ferramenta importante para as escolas, para as secretarias de Educação (estaduais e municipais) e para o MEC, pois permite a identificação de problemas de abandono escolar e o desenvolvimento de novas políticas públicas baseadas em evidências para reduzi-los. Esses dados também são utilizados para aferir a qualidade do ensino e a efetividade das já existentes políticas públicas voltadas para a educação (SANTOS et al., 2019b; MONTEIRO; FERREIRA; TEIXEIRA, 2009). A partir desses dados, o MEC pode identificar regiões do país, estados, cidades e escolas com baixa frequência escolar e

---

<sup>1</sup> <https://presenca.mec.gov.br/>

direcionar esforços para melhorar a qualidade do ensino nessas áreas.

Apesar de sua importância, a coleta de dados de frequência escolar do Bolsa Família ainda enfrenta diversos desafios. Dentre os principais se encontram fragilidades no sistema educacional brasileiro, como a dualidade entre escolas públicas e privadas. Escolas particulares tendem a ser mais equipadas, melhor localizadas e ter professores mais preparados se comparadas às escolas públicas. A diferença de estrutura das escolas para coletar e processar os dados, a falta de treinamento dos professores e gestores escolares para lidar com as informações e a falta de recursos financeiros para implementar sistemas de coleta de dados que permitiriam um processo mais ágil e preciso das informações (DUTRA-THOMÉ; PEREIRA; KOLLER, 2016). Para superar essa dificuldade, o MEC tem buscado parcerias com instituições de tecnologia e com governos locais para garantir o acesso à internet e aos equipamentos necessários para a coleta e processamento dos dados.

Diante desses desafios, o MEC tem adotado medidas para aprimorar a coleta de dados de frequência escolar do Bolsa Família. Entre as iniciativas em andamento estão a implementação de sistemas de coleta de dados mais eficientes e o desenvolvimento de programas de capacitação para professores e gestores escolares. Apesar dos desafios, a coleta de dados de presença escolar do Bolsa Família tem sido fundamental para a efetividade do programa e para o monitoramento das políticas públicas voltadas para a educação. Com a coleta e análise desses dados, é possível identificar os locais com maior índice de abandono escolar, desenvolver estratégias para melhorar a qualidade do ensino e garantir o acesso à educação de qualidade para todos os alunos brasileiros.

Os pesquisadores têm utilizado os dados do MEC para diversos fins, como por exemplo, para avaliar a eficácia de programas educacionais (VARGAS; ZUCCARELI, 2021), para identificar as necessidades de formação de professores, ou para analisar as desigualdades educacionais no país e desenvolver políticas públicas que buscam melhorar a qualidade da educação (SANTOS et al., 2019b). Esses dados são essenciais para a elaboração de políticas e programas educacionais eficazes e para o monitoramento contínuo da qualidade da educação brasileira, permitindo aos pesquisadores, gestores e educadores entenderem melhor a realidade educacional do país.

Uma das várias formas de análise de dados é a ciência do reconhecimento de padrões, que na computação é uma área interdisciplinar que envolve conhecimentos de matemática, estatística e inteligência artificial. Esse campo de estudo busca identificar padrões e regularidades em conjuntos de dados complexos, permitindo que sejam feitas previsões e tomadas de decisão mais precisas e eficazes (PAOLANTI; FRONTONI, 2020). O histórico da detecção de padrões na computação remonta à década de 1950 com estudos focados em padrões de imagens (SELFIDGE, 1955), e desde então muitos outros métodos foram conquistados, como a análise de séries temporais (GURALNIK; SRIVASTAVA, 1999), aprendizagem de máquina e mineração de dados.

Hoje em dia, a detecção de padrões tem se mostrado uma área de grande relevância para o desenvolvimento de tecnologias inovadoras, como carros autônomos (TSELENTIS; PAPADIMITRIOU, 2023), assistentes virtuais (SAYED; ASHOUR, 2021), reconhecimento

facial (CHEN et al., 2021), sistemas de recomendação (AHMED; IMTIAZ; KHAN, 2018) e até para a medicina personalizada, em que a análise de grandes conjuntos de dados pode ajudar a identificar tratamentos mais eficazes e personalizados para cada paciente. Sendo assim, é uma das áreas mais promissoras da computação e tem contribuído significativamente para o avanço de diferenciadas tecnologias e diversos campos de estudo. No contexto da educação brasileira, o grande volume da coleta de dados do MEC pode permitir que padrões de comportamento e aprendizado sejam identificados, por exemplo, por meio da análise de dados de desempenho dos estudantes em diferentes disciplinas, que pode inferir padrões de dificuldades em determinadas matérias, indicando a necessidade de ajustes na metodologia de ensino ou na grade horária.

Sendo assim, é possível inferir que a detecção de padrões possa ser aplicada para a análise da frequência dos estudantes em sala de aula, e ainda de variados modos, como por reconhecimento facial (BUDIMAN et al., 2023) ou por dispositivos *wearables* (FERREIRA; RODRIGUEZ; MOTTI, 2020). A coleta de dados de presença dos estudantes pode ser utilizada para detectar padrões de ausências, indicando a necessidade de intervenções para melhorar a participação dos estudantes nas aulas, já que a baixa frequência de estudantes está relacionada com maiores índices de evasão escolar. O absentismo, por sua vez, tem distintos e complexos fatores de risco, incluindo vulnerabilidade social, abuso de substâncias, baixo envolvimento dos pais com a escola, a falta de correspondência entre as abordagens educacionais e as necessidades dos estudantes, a carência de iniciativas governamentais efetivas e a exploração de trabalho infantil (NASCIMENTO et al., 2020; GUBBELS; PUT; ASSINK, 2019). Neste contexto, a análise dos dados referentes a presença escolar dos estudantes pode ser uma ferramenta útil, permitindo o aprimoramento de práticas pedagógicas com enfoque em retenção escolar. É importante destacar, no entanto, que a coleta de dados dos estudantes deve ser realizada de forma ética e responsável, respeitando a privacidade e os direitos dos estudantes.

## 1.1 TED 11476

O presente trabalho encontra-se no contexto do projeto celebrado entre o Núcleo de Excelência em Tecnologias Sociais (NEES), coordenado pela Universidade Federal de Alagoas (UFAL), e o Ministério da Educação (MEC) por meio do Plano de Trabalho do Termo de Execução Descentralizada (TED) nº 11476/2022, que possui como meta geral o desenvolvimento de uma solução de gestão inteligente para controle de frequência de alunos do ensino básico, com alto potencial de adesão das escolas, visando o enfrentamento do abandono e da evasão escolar. Essa plataforma deve ser baseada em uma coleta de dados semi-automática de baixo custo; uma infraestrutura robusta capaz de armazenar, gerenciar e permitir o monitoramento de uma quantidade massiva de dados; e o uso de algoritmos de Inteligência Artificial para gestão de usuários e alerta automatizados. Além disso, a solução automatizada também deve ser capaz de corrigir textos no contexto da plataforma. O objetivo é ampliar a atuação da solução automatizada para melhorar a qualidade da educação básica no Brasil, sendo que o responsável final pelo



acompanhamento da execução do objeto do TED é a Secretaria de Educação Básica, podendo utilizar relatórios periódicos, visitas técnicas ou outras formas de monitoramento.

Para que o TED consiga atingir seu objetivo, o projeto conta com uma equipe técnica composta por professores e pesquisadores com alto grau de qualificação na área de tecnologia e educação, e com liderança científica na área de Inteligência Artificial na Educação, sendo divididos em equipes como Pesquisa, User Experience, Desenvolvimento, Arquitetura de Software, e assim por diante. O NEES trabalha para reduzir as desigualdades educacionais através de políticas públicas, programas educacionais e inovação, e é considerado um dos mais importantes e mais qualificados grupos de pesquisa em Informática na Educação do país, tendo mais de 30 prêmios na área. O objetivo do grupo é gerar conhecimento científico de vanguarda e transferir os resultados para a sociedade, impactando mais de 7.000 escolas, 500.000 alunos e 20.000 professores em mais de 1.500 municípios. Este trabalho de conclusão de curso está associado diretamente com o núcleo do NEES da UFAL, e é uma parceria da UDESC com o NEES.

Ademais, o TED 11476 também está associado ao MEC, mais especificamente ao Fundo Nacional de Desenvolvimento da Educação (FNDE), que é a unidade descentralizadora responsável pelo TED. Há uma atuação conjunta com a Secretaria de Tecnologia da Informação e Comunicação (STIC) do MEC para avaliar os aspectos de conectividade das escolas e suas limitações, além de preparar a infraestrutura da STIC para o armazenamento dos dados coletados. Ou seja, o MEC está envolvido no projeto como um parceiro estratégico.

Por fim, é importante reforçar que o TED 11476 teve início no ano de 2022 e tem previsão para ter seu prazo encerrado em 2024. Este trabalho de conclusão de curso apoia o projeto ao contribuir de maneira efetiva para o desenvolvimento da solução de gestão inteligente, uma vez que a detecção de padrões na frequência dos alunos é fundamental para compreender os fatores que podem influenciar a evasão escolar, tendo a possibilidade de auxiliar na tomada de decisões dentro do projeto.

## 1.2 OBJETIVO GERAL

Encontrar padrões de características de estudantes do ensino básico brasileiro com base na presença em sala de aula.

## 1.3 OBJETIVOS ESPECÍFICOS

Com base no objetivo geral, os seguintes objetivos específicos são definidos:

- Investigar sistemas de presença na literatura;
- Realizar levantamento bibliográfico sobre algoritmos de reconhecimento de padrões relacionados à mineração de dados;

- Estudar as bases de dados disponibilizadas pelo NEES sobre a presença dos estudantes do ensino básico brasileiro;
- Propor um modelo para a análise de dados sobre estudantes do ensino básico;
- Descobrir padrões existentes nas bases de dados.

#### 1.4 METODOLOGIA

Este trabalho de conclusão de curso possui natureza aplicada, e é caracterizado como um trabalho de carácter experimental, pois será necessária a aplicação de técnicas que serão analisadas no decorrer da pesquisa. A abordagem será na forma quantitativa, já que serão estudados dados estatísticos que posteriormente serão traduzidos em visualizações.

A pesquisa iniciará com um levantamento bibliográfico para investigar como é realizada a detecção e cobrança de presença de alunos no ensino básico brasileiro atualmente, assim como técnicas de reconhecimento de padrões dentro da área de aprendizado de máquina.

Após ter sido realizada a fundamentação teórica, será feita a análise dos dados obtidos, aplicando as técnicas encontradas de reconhecimento de padrões para encontrar modelos semelhantes de tipos de estudantes.

#### 1.5 ESTRUTURA DO TRABALHO

O trabalho está organizado da seguinte forma: no Capítulo 2 é dado foco na fundamentação teórica das técnicas a serem aplicadas no trabalho. No Capítulo 3 são apresentados os trabalhos relacionados e apontados suas diferenças com o atual trabalho de conclusão de curso. Por fim, no Capítulo 4 são apresentadas as conclusões parciais do trabalho e a proposta do que será feito no TCC 2.

## 2 FUNDAMENTAÇÃO TEÓRICA

Com o progresso tecnológico e a crescente importância de monitorar a presença dos alunos, têm surgido sistemas de registro de frequência cada vez mais avançados em várias nações ao redor do globo. No contexto brasileiro, a adoção de tais sistemas também tem se tornando uma prática amplamente adotada nas escolas, buscando assegurar a presença dos alunos e mitigar a taxa de evasão escolar. Através desses sistemas, é possível obter dados precisos sobre a assiduidade dos estudantes, permitindo uma análise mais aprofundada dos padrões de comparecimento e oferecendo informações valiosas para o desenvolvimento de estratégias educacionais eficazes. Além disso, esses sistemas possibilitam um acompanhamento mais individualizado e um monitoramento mais consistente do engajamento dos alunos, contribuindo para a identificação precoce de possíveis problemas e permitindo uma intervenção mais ágil por parte das instituições de ensino. Dessa forma, a utilização de sistemas de frequência representa um importante recurso no esforço contínuo de melhorar a eficiência e a qualidade do sistema educacional brasileiro.

### 2.1 PRESENÇA E FREQUÊNCIA

Antes de tratar sobre a análise de dados, é necessário definir a diferença entre as palavras presença e frequência, já que, apesar de semelhantes, os termos não devem ser confundidos pelas etimologias de suas palavras acarretarem significados diferentes. Ou seja, frequência escolar não deve ser confundida com a presença escolar.

A palavra “frequência” tem sua origem no latim “*frequentia*” (NASCENTES, 1966), que significa “ser frequente” ou “repetir-se”. O termo “*frequentia*” foi incorporado ao latim para expressar a ideia de repetição ou ocorrência frequente de algo. Posteriormente, o termo passou a ser usado na matemática e na física para descrever a quantidade de vezes que um evento se repete em um determinado intervalo de tempo. A partir dessa noção, a palavra “frequência” foi adotada em outros contextos, como nas ciências, na linguagem cotidiana e nas tecnologias, para indicar a medida da quantidade, intensidade ou regularidade de algum fenômeno ou evento. Sendo assim, detecção de frequência em uma sala de aula refere-se à identificação da presença física dos indivíduos naquele espaço. Ela busca determinar se os alunos estão fisicamente presentes na sala de aula, independentemente do nível de participação ou envolvimento. Pode ser realizada por meio de sistemas de monitoramento, como câmeras ou registros de entrada e saída, ou ainda de forma manual, pela anotação de uma pessoa por meio da lista de chamada.

Já a palavra “presença” tem sua origem no latim “*praesentia*” (NASCENTES, 1966), que deriva do termo “*praesens*”, que significa “presente” ou “estar ao lado”. “*Praesens*” é formado pela junção do prefixo “*prae-*”, que indica “antes” ou “à frente”, e do verbo “*esse*”, que significa “ser” ou “estar” (PEREIRA, 2011). Portanto, a palavra “presença” remete à ideia de estar à frente ou ao lado de alguém. Sendo assim, é possível inferir que a detecção de presença em uma sala de aula refere-se não apenas à medição da regularidade com que um aluno participa

das atividades educacionais, mas também acompanhar o nível de envolvimento ativo do aluno durante as aulas, a forma com que ele participa, interage com o professor e colegas, realiza tarefas, entre outros aspectos. A detecção de presença busca compreender a consistência com que o estudante se engaja nas atividades ao longo do tempo, podendo ser realizada por meio de sistemas de monitoramento, como câmeras ou sensores de movimento, ou com sensores vestíveis que podem detectar outras informações do estudante, como batimentos cardíacos.

Dessa forma, é possível concluir que a frequência escolar consiste no número de vezes que o aluno frequenta as atividades escolares, enquanto a presença escolar refere-se à participação efetiva do aluno nas atividades escolares. Isso significa que, para ser considerado presente, o aluno precisa não apenas comparecer à escola, mas também participar ativamente das atividades propostas pelos professores, cumprir tarefas, fazer perguntas, entre outras ações que demonstram seu engajamento no processo de aprendizagem. Dessa forma, a presença escolar é considerada um indicador mais abrangente e efetivo do que a frequência escolar, pois leva em conta não apenas a quantidade, mas também a qualidade da participação do aluno nas atividades escolares. Todavia, ainda é importante que a frequência escolar seja acompanhada de perto pelos educadores, juntamente com outros indicadores de desempenho, como as notas e avaliações do aluno, pois, além de ser mais facilmente medida, é ela o critério legal para a manutenção da matrícula escolar (BRASIL, 1996). Neste trabalho de conclusão de curso, trataremos principalmente da coleta da frequência escolar.

Também é importante destacar que o engajamento e a participação ativa dos alunos nas atividades escolares são fatores determinantes para o sucesso acadêmico. Apenas a presença física na sala de aula não é suficiente para avaliar o envolvimento do aluno com o processo de ensino-aprendizagem, sendo necessário considerar outros indicadores, como a qualidade das tarefas realizadas e a interação com colegas e professores. Nesse sentido, a implementação de sistemas de presença escolar mais sofisticados, capazes de coletar dados precisos sobre a situação do aluno em sala de aula poderia ser uma solução promissora para o aprimoramento das políticas públicas voltadas para a educação (FERREIRA; RODRIGUEZ; MOTTI, 2020).

A compreensão da diferença entre presença e frequência é um aspecto crucial para uma análise precisa dos dados educacionais, sendo fundamental para a compreensão do conceito de pesquisas na área de *Educational Data Mining*. Ao entender a diferença entre presença e frequência e o papel do EDM na análise de dados educacionais, é possível explorar como a aplicação desses conceitos e técnicas pode contribuir para aprimorar o sistema educacional, promovendo o engajamento dos estudantes e implementando estratégias eficazes de ensino e aprendizagem.

## 2.2 EDUCATIONAL DATA MINING (EDM)

*Data Mining* (DM), ou mineração de dados, é um processo que envolve a descoberta de informações valiosas e *insights* a partir de conjuntos de dados grandes e complexos. Seus

conceitos combinam técnicas estatísticas, de banco de dados e de inteligência artificial para explorar os dados e identificar padrões, tendências e relacionamentos ocultos. A mineração de dados permite extrair conhecimentos úteis e significativos a partir de grandes volumes de dados, que podem ser provenientes de diferentes fontes, como bancos de dados corporativos, registros de transações, mídias sociais, registros de compras e muito mais. O objetivo é descobrir padrões consistentes e informações relevantes que possam ser utilizadas para tomar decisões estratégicas, identificar oportunidades de negócio, prever comportamentos futuros e melhorar a eficiência operacional.

O processo de mineração de dados envolve várias etapas, como a coleta de dados de diferentes fontes, o armazenamento em um banco de dados, o pré-processamento dos dados para garantir sua qualidade e consistência, a aplicação de técnicas de mineração de dados para descobrir padrões e relacionamentos, e a interpretação dos resultados obtidos. A mineração de dados desempenha um papel crucial em diversas áreas, tais como marketing, finanças, saúde, varejo, segurança, entre outras. Ela permite identificar segmentos de clientes, prever demandas futuras, otimizar processos, detectar fraudes, personalizar recomendações, melhorar a qualidade do atendimento ao cliente, entre muitas outras aplicações.

A *Educational Data Mining*, ou mineração de dados educacionais, é uma disciplina emergente que é preocupada com o desenvolvimento de métodos para explorar os dados únicos e cada vez mais em grande escala que vêm de ambientes educacionais e usar esses métodos para entender melhor os alunos e os ambientes em que eles aprendem (KOÇ, 2017). Sendo assim, EDM é uma área de pesquisa interdisciplinar que combina técnicas de mineração de dados, aprendizado de máquina e estatística para analisar grandes conjuntos de dados educacionais. Seu principal objetivo é analisar dados educacionais para resolver problemas da pesquisa educacional, descobrindo padrões, tendências e relações nos dados coletados a fim de obter conclusões valiosas para melhorar a eficácia do ensino e da aprendizagem.

Através da aplicação de técnicas de mineração de dados e análise estatística, o EDM permite extrair conhecimento a partir de dados educacionais, que podem incluir registros acadêmicos, históricos de notas, interações do aluno com sistemas de aprendizagem online, dados de questionários, entre outros. Esses dados são processados e analisados para identificar padrões e relações que possam ajudar a compreender o desempenho dos alunos, suas dificuldades, preferências de aprendizagem e outros aspectos relevantes.

Alguns exemplos de aplicações do EDM incluem, mas não estão limitados a:

1. Previsão do desempenho dos alunos: Ao analisar dados de sistemas educacionais, como informações dos alunos e notas, além de dados de sistemas que não estão diretamente ligados a uma universidade ou escola, como dados de vestibular do Sistema de Seleção Unificada (Sisu) ou Exame Nacional do Ensino Médio (Enem), os pesquisadores podem identificar padrões associados ao sucesso ou fracasso dos estudantes, é possível processar esses dados e identificar padrões e correlações que podem ser usados para prever o

desempenho dos estudantes (WILL; KEMCZINSKI; PARPINELLI, 2019). A previsão do desempenho dos alunos pode trazer diversos benefícios para as instituições de ensino, como a possibilidade de oferecer uma educação personalizada aos alunos de risco e o oferecimento de bolsas para alunos que se destacam.

2. Identificação de alunos em situação de risco: Ao analisar dados sobre o comportamento e desempenho dos estudantes, os pesquisadores podem identificar fatores associados a resultados acadêmicos insatisfatórios. Por exemplo, eles podem descobrir que os alunos que faltam às aulas ou não concluem as tarefas no prazo têm maior probabilidade de abandonar a escola. Essas informações podem ser utilizadas para desenvolver sistemas de alerta precoce que informam os educadores quando um aluno está em risco de abandonar os estudos, permitindo que intervenham antes que seja tarde demais (RAMOS et al., 2020).
3. Acompanhar automaticamente atividades de estudantes: Ao analisar dados de registro provenientes de plataformas LMS, os pesquisadores podem obter percepções sobre como os estudantes interagem com os materiais do curso e identificar áreas em que possam estar enfrentando dificuldades. Essas informações podem ser utilizadas para personalizar os ambientes de aprendizagem, fornecendo *feedback* direcionado ou recomendações de recursos adicionais (SANTOS et al., 2019a).
4. Recomendar recursos de aprendizagem aos alunos com base em seu desempenho e preferências: Ao analisar dados sobre o comportamento e desempenho dos estudantes, os pesquisadores podem desenvolver algoritmos que recomendam recursos de aprendizagem específicos de acordo com as necessidades e interesses individuais de cada aluno (OLIVEIRA; BERNARDINI; VITERBO, 2021).

Contudo, para se realizar essas pesquisas deve-se coletar e analisar dados de estudantes que podem ser classificados como sensíveis, e portanto há importantes preocupações quanto à privacidade e segurança dos dados, uma vez que a pesquisa, caso feita sem cuidado, pode potencialmente colocar em risco informações pessoais dos estudantes. Além disso, a necessidade de habilidades especializadas para analisar grandes conjuntos de dados pode criar uma barreira para alguns educadores ou instituições que não tenham recursos ou conhecimentos especializados para usar efetivamente essa tecnologia. Outra possível desvantagem é a possibilidade de reforçar vieses existentes nos dados, o que pode perpetuar preconceitos e desigualdades ou limitar oportunidades para determinados grupos de estudantes.

### 2.2.1 Learning Analytics (LA)

*Learning Analytics*, ou analíticas de aprendizado, é uma área de estudo que utiliza técnicas e ferramentas analíticas para coletar, analisar e interpretar dados relacionados ao processo de aprendizagem. Envolve a aplicação de métodos quantitativos e qualitativos para obter informações úteis sobre o desempenho dos estudantes, padrões de comportamento, eficácia

do ensino, entre outros aspectos relacionados ao aprendizado. O objetivo principal do *Learning Analytics* é melhorar a eficácia e a eficiência do processo de aprendizagem. Ao coletar dados relevantes, como desempenho acadêmico, interações em plataformas de aprendizado online, tempo gasto em atividades, feedback dos alunos, entre outros, as instituições educacionais podem identificar padrões e tendências que ajudam a entender melhor o progresso dos alunos e oferecer suporte personalizado. As técnicas de *Learning Analytics* podem ser aplicadas em diversos contextos educacionais, desde escolas e universidades até plataformas de ensino online. Com base nas informações coletadas, é possível tomar decisões informadas para aprimorar o planejamento curricular, desenvolver intervenções pedagógicas, identificar necessidades de alunos em risco, personalizar a experiência de aprendizagem e avaliar a eficácia de métodos de ensino específicos.

Segundo Aldowah, Al-Samarraie e Fauzy (2019), há determinadas etapas necessárias para se conduzir um estudo de LA, as quais incluem:

1. Definir o problema de pesquisa: O primeiro passo é identificar o problema de pesquisa que se deseja abordar. Isso pode envolver a identificação de uma questão específica relacionada ao ensino e aprendizagem que possa ser respondida por meio da análise de dados.
2. Coletar dados: O próximo passo é coletar os dados necessários para responder à questão de pesquisa. Isso pode envolver a coleta de dados de várias fontes, como sistemas de gerenciamento de aprendizagem (LMS), registros acadêmicos, questionários de alunos e outros dados relevantes.
3. Pré-processar os dados: Depois que os dados são coletados, eles precisam ser pré-processados para garantir que estejam limpos e prontos para análise. Isso pode envolver a limpeza de dados ausentes ou inconsistentes, a normalização de dados e a seleção de recursos relevantes.
4. Analisar os dados: O próximo passo é analisar os dados usando técnicas de aprendizado de máquina, mineração de dados e outras técnicas de análise de dados. Isso pode envolver a identificação de padrões e tendências nos dados, a criação de modelos preditivos e a realização de outras análises relevantes.
5. Interpretar os resultados: Depois que os dados são analisados, os resultados precisam ser interpretados para responder à questão de pesquisa original. Isso pode envolver a identificação de *insights* relevantes, a validação dos resultados e a comunicação dos resultados para as partes interessadas relevantes.
6. Tomar medidas: Com base nos resultados da análise, as partes interessadas podem tomar medidas para melhorar o ensino e a aprendizagem. Isso pode envolver a implementação de intervenções específicas para alunos ou grupos de alunos, a revisão de materiais de ensino ou a melhoria de práticas de ensino.

É notável que no processo de *Learning Analytics* há grande envolvimento da coleta de dados de diversas fontes, como sistemas de gerenciamento de aprendizagem, registros de atividades dos alunos, interações em fóruns e redes sociais, entre outros. Esses dados são então analisados para identificar padrões e tendências que possam ajudar a entender melhor o comportamento dos alunos e o desempenho do sistema educacional como um todo. Uma das principais características de LA é a sua abordagem holística, que considera o sistema educacional como um todo. Isso significa que a análise de dados não se limita apenas ao desempenho dos alunos, mas também leva em conta fatores como o ambiente de aprendizagem, a qualidade do conteúdo, a eficácia dos métodos de ensino, entre outros.

Outra característica importante é a ênfase na intervenção e participação humana. Embora a análise de dados seja realizada por meio de ferramentas computacionais, a interpretação dos resultados e a tomada de decisões são feitas por professores, gestores e outros profissionais da educação (CAMPOS et al., 2020). Isso significa que a tecnologia é usada como uma ferramenta para apoiar a tomada de decisões, e não como um substituto para o julgamento humano. Isso é uma consideração importante pois os alunos estão em constante mudança de perfil, o que exige novas e eficazes ferramentas para auxiliar nesse processo. Sendo assim, as instituições de ensino vêm produzindo e armazenando grande quantidade de dados ao longo do tempo.

### **2.2.2 Conexão entre EDM e LA**

Como visto nas seções anteriores, tanto LA quanto EDM são áreas específicas que são usadas para representar o uso e a aplicação de mineração de dados em diversos contextos educacionais. Elas estabelecem um ecossistema que pode consecutivamente coletar, processar, reportar e trabalhar com dados continuamente para melhorar o processo educacional. Contudo, as etapas de ambos os processos são diferentes.

Ambas as áreas têm como objetivo a melhoria de processos de ensino e aprendizagem por meio da análise de dados em larga escala, de maneira sistematizada, que possam auxiliar na ampliação de processos de avaliação, compreensão de problemas e planejamento de intervenções. Segundo Moissa, Gasparini e Kemczinski (2015), tanto EDM e LA têm como objetivo analisar dados educacionais para entender o processo de ensino e aprendizagem e melhorá-lo, incluindo o uso de algoritmos de aprendizado de máquina, a análise de dados em tempo real e a aplicação de técnicas de análise de redes sociais. Seus resultados e análises podem ser aplicados no desenvolvimento de sistemas de recomendação personalizados para estudantes e professores.

Uma das principais vantagens em se beneficiar de ambas as áreas é que isso permite aos tomadores de decisão avaliar os processos de aprendizagem dos estudantes e atender às suas diferentes necessidades de aprendizagem com base em seus comportamentos e preferências reais. Isso pode ser alcançado por meio do uso de técnicas de mineração de dados que analisam grandes conjuntos de dados para identificar padrões e tendências no comportamento dos alunos, como seu envolvimento com materiais do curso, seu desempenho em avaliações e suas interações com colegas e instrutores (CEREZO et al., 2016a). Outro benefício do uso da mineração de dados



educacionais e de analíticas de aprendizado é que podem fornecer um suporte inestimável ao processo de tomada de decisão. Por exemplo, a análise preditiva pode ser usada para identificar alunos que correm o risco de desistir ou fracassar em um curso, permitindo que os instrutores intervenham precocemente e forneçam suporte direcionado. Da mesma forma, as aplicações de mineração de dados podem ser usadas para identificar áreas em que os materiais do curso ou as estratégias instrucionais possam precisar ser revisados ou aprimorados (ALDOWAH; AL-SAMARRAIE; FAUZY, 2019).

Apesar disso, existem algumas distinções entre as duas áreas. A EDM dá mais ênfase na análise de dados históricos e a identificação de padrões, geralmente se concentrando em dados históricos ou estruturados, como notas, frequência e dados demográficos dos alunos, se tornando mais voltada para a pesquisa acadêmica. Enquanto isso, a LA ao buscar coletar, medir, analisar e relatar os dados e seus contextos, é mais focada na análise em tempo real e na personalização da aprendizagem, podendo incluir dados não estruturados, como interações em fóruns de discussão e registros de atividades em plataformas de aprendizagem online. Pode ainda fornecer informações em tempo real aos educadores, permitindo que eles tomem decisões sobre como personalizar a aprendizagem para atender às necessidades individuais dos alunos; ou seja, apresenta etapas e ciclos em que é considerada a intervenção humana (CAMPOS et al., 2021).

Independente das diferenças, as abordagens têm o potencial de oferecer suporte valioso ao processo de tomada de decisão e contribuir para o aprimoramento da educação.

## 2.3 RECONHECIMENTO DE PADRÕES

A mineração de dados abrange diversos sistemas para pré-processamento, análise e interpretação de dados. Essas estratégias podem ser divididas principalmente em duas áreas: reconhecimento de padrões e aprendizado de máquina. A detecção de padrões em análise de dados desempenha um papel crucial na extração de informações valiosas a partir de conjuntos de dados. Essa abordagem analítica permite identificar regularidades, tendências ou estruturas subjacentes, revelando relações complexas entre variáveis e eventos (VELURI et al., 2022). O reconhecimento de padrões é realizado por meio de uma combinação de técnicas estatísticas e algoritmos de aprendizado de máquina. Esses algoritmos são treinados em conjuntos de dados previamente coletados, nos quais são expostos a exemplos de diferentes padrões e classes. Ao analisar esses exemplos, os algoritmos aprendem a identificar e generalizar padrões, permitindo a aplicação desses conhecimentos a novos dados para classificação, previsão ou tomada de decisões.

No contexto do *Learning Analytics*, o reconhecimento de padrões desempenha um papel fundamental na compreensão do processo de aprendizagem. Ao examinar padrões de reprovação em um curso ou disciplina, é possível identificar características, comportamentos e fatores que estão correlacionados com um maior risco de insucesso dos alunos. Essas informações podem incluir variáveis como desempenho prévio, engajamento, participação em atividades ou

interações com o material didático.

Além disso, o reconhecimento de padrões em análise de dados pode revelar correlações complexas e sutis que não seriam facilmente identificadas por meio de métodos tradicionais. Padrões de desempenho, por exemplo, podem estar relacionados a fatores como estilo de aprendizagem, habilidades específicas ou até mesmo fatores externos, como ambiente familiar ou condições socioeconômicas dos estudantes.

### 2.3.1 Técnicas

Técnicas de reconhecimento de padrões são métodos computacionais que permitem identificar regularidades ou padrões em dados. Essas técnicas são usadas para analisar e extrair informações úteis a partir de dados complexos, como imagens, sinais de áudio, séries temporais e conjuntos de dados multidimensionais. As técnicas de reconhecimento de padrões são amplamente utilizadas em diversas áreas, como visão computacional, processamento de sinais, bioinformática e análise de dados em geral (BISHOP, 2007). Estas técnicas de reconhecimento de padrões podem ser associada a *Learning Analytics* pois elas podem ser usadas para extrair informações úteis dos grandes conjuntos de dados educacionais e ajudar a entender e otimizar todo o processo de aprendizagem.

Algumas delas incluem, mas não são limitadas a:

1. *Clustering*: processo de agrupar um conjunto de objetos físicos ou abstratos em classes de objetos semelhantes. Um *cluster* é uma coleção de objetos de dados que são semelhantes entre si dentro do mesmo *cluster* e dissimilares em relação aos objetos em outros *clusters*. Um *cluster* de objetos de dados pode ser tratado coletivamente como um grupo e, portanto, considerado uma forma de compressão de dados. Um dos algoritmos de *clustering* mais populares é o *K-Means*, onde o usuário define inicialmente o número de clusters ( $k$ ) que deseja criar. Em seguida, o algoritmo seleciona aleatoriamente  $k$  objetos do conjunto de dados como centróides iniciais. Depois disso, cada objeto de dados é atribuído ao centróide mais próximo com base em uma medida de distância, geralmente a distância euclidiana. Após a atribuição de todos os objetos de dados a um centróide, os centróides são recalculados como a média dos objetos de dados atribuídos a eles. Esse processo é repetido até que os centróides não mudem mais ou até que um número máximo de iterações seja atingido. O resultado final é um conjunto de  $k$  *clusters*, onde cada *cluster* é representado por seu centróide. Em contextos educacionais, o *clustering* pode ser utilizado para agrupar alunos com base em seus desempenhos acadêmicos ou outras características relevantes. Por exemplo, um professor pode usar o *k-means* para agrupar alunos com base em suas notas em diferentes disciplinas ou em seus estilos de aprendizagem. Isso pode ajudar o professor a personalizar o ensino para cada grupo de alunos e melhorar o desempenho geral da turma. Além disso, o *k-means* também pode ser usado para agrupar

cursos com base em suas características, como nível de dificuldade, carga horária, etc. (ALI; KADHUM, 2017).

2. *Association Rules*: técnica de mineração de dados que busca identificar relações entre diferentes variáveis em um conjunto de dados, sendo frequentemente usada em educação para identificar padrões em dados de desempenho do aluno e comportamento do aluno. No contexto educacional, Cerezo et al. (2016b) discutem que as regras de associação podem ser usadas para identificar padrões de comportamento e desempenho de estudantes, já que podem ser usadas para identificar relacionamentos entre diferentes variáveis, como dados demográficos, materiais do curso e resultados acadêmicos. Regras de associação podem ser usadas, por exemplo, para identificar quais materiais do curso estão mais fortemente associados ao sucesso ou fracasso do aluno, ou para desenvolver recomendações personalizadas para alunos com base em suas necessidades e preferências individuais.
3. *Time series*: técnica estatística que analisa dados coletados ao longo do tempo. O reconhecimento de padrões em séries temporais envolve o uso de algoritmos de reconhecimento de padrões que mapeiam automaticamente uma representação de entrada para uma categoria de saída. Existem três tipos principais de algoritmos de reconhecimento de padrões: supervisionado, não supervisionado e semi-supervisionado. No aprendizado supervisionado, um modelo funcional é usado para mapear entradas observadas para categorias de saída. Várias técnicas de construção de modelos foram desenvolvidas para esse fim, incluindo árvores de decisão, indução de regras, redes Bayesianas, raciocínio baseado em memória, máquinas de vetores de suporte (SVMs) e redes neurais. No aprendizado não supervisionado, um padrão de entrada é atribuído a uma classe desconhecida, sendo útil quando não há rótulos de classe disponíveis para o conjunto de dados. No aprendizado semi-supervisionado, um padrão de entrada é atribuído a uma das classes pré-definidas, usando tanto dados rotulados quanto não rotulados. (LIN et al., 2012) No contexto educacional, pode ser usado para analisar o desempenho acadêmico dos alunos ao longo do tempo e identificar tendências ou padrões sazonais.
4. *Sequential Pattern analysis*: técnica de análise de dados que consiste em descobrir padrões interessantes em um conjunto de sequências. Esses padrões podem ser medidos em termos de vários critérios, como frequência de ocorrência, comprimento e lucro. Dois algoritmos que envolvem essas técnicas são o *Generalized Sequential Pattern* (GSP) e o *Sequential Pattern Discovery using Equivalence classes* (SPADE) (FOURNIER-VIGER et al., 2017). O algoritmo GSP é um algoritmo de busca em profundidade que gera candidatos a padrões de tamanho  $k + 1$  a partir de padrões de tamanho  $k$ , enquanto o algoritmo SPADE usa uma estrutura de árvore de prefixo para gerar candidatos a padrões. No contexto educacional, pode ser usado para identificar sequências frequentes de comportamentos ou ações dos alunos durante o processo de aprendizagem.

### 2.3.2 Visualização da informação

Tendo estabelecida a importância da coleta de dados, é necessário também discutir sobre como esses dados serão apresentados para garantir que não fiquem ambíguos ou confusos. A área da visualização de dados desempenha um papel fundamental nesse aspecto. A visualização de dados busca representar informações de forma gráfica e intuitiva, permitindo que os usuários compreendam e interpretem os dados de maneira mais eficiente.

A percepção visual é uma ferramenta poderosa para os seres humanos, pois ajuda no reconhecimento de padrões complexos em dados. É possível identificar detalhes, distinguir formas e padrões com maior facilidade a partir de informações visuais do que a partir de grandes quantidades de dados numéricos. A mineração de dados simbólicos é uma área complexa que envolve a agregação de grandes quantidades de dados clássicos em uma forma mais compacta. No entanto, devido à sua estrutura complexa, a visualização desse tipo de dado enfrenta desafios adicionais. Diferente dos dados numéricos, que podem ser representados e visualizados de maneira direta, os dados simbólicos exigem técnicas especiais para serem visualizados de forma eficaz. A transformação de dados simbólicos em representações visuais compreensíveis é um desafio, pois é necessário preservar a essência dos dados enquanto se simplifica sua estrutura. No entanto, superar esses desafios é crucial, pois a visualização de dados simbólicos oferece benefícios significativos, permitindo que os analistas compreendam melhor os padrões e relacionamentos ocultos nos dados, facilitando a tomada de decisões informadas (UMBLEJA; ICHINO; YAGUCHI, 2020).

Uma das maneiras de enfrentar a complexidade da visualização de dados é utilizar técnicas de codificação visual, como o *Zoom Star* e o *shape encoding*. A solução *Zoom Star* é uma ferramenta poderosa para visualizar grandes conjuntos de dados de forma agregada. Ela utiliza gráficos radiais para representar objetos simbólicos em 2D ou 3D, permitindo comparar diferentes objetos simbólicos lado a lado ou sobrepostos em um mesmo gráfico. A representação em gráficos radiais é particularmente útil para dados simbólicos, que são aqueles que não podem ser medidos diretamente, mas podem ser descritos por meio de categorias ou intervalos. Por exemplo, na análise de dados sobre imóveis, a solução *Zoom Star* pode ser utilizada para representar diferentes imóveis em um mesmo gráfico, com cada variável (como número de quartos, tamanho do terreno, preço, etc.) representada por um eixo radial. Os pontos no gráfico representam cada imóvel, e o tamanho dos pontos pode ser usado para representar a frequência de cada categoria ou intervalo (NOIRHOMME-FRAITURE, 2002). Sendo assim, a ferramenta é útil para visualizar informações de reconhecimento de padrões, especialmente quando se trata de dados simbólicos, pois com ela é possível representar diferentes objetos em um mesmo gráfico, com cada variável representada por um eixo radial, e utilizar diferentes cores e transparências para destacar diferentes objetos ou variáveis.

Já o *shape encoding* é uma técnica de codificação visual que mapeia valores de dados multidimensionais em formas geométricas. Cada variável é atribuída a uma forma e a escala da

forma é usada para representar a magnitude da variável. Por exemplo, uma variável pode ser atribuída a um círculo e a escala do círculo pode ser aumentada para representar valores mais altos da variável. Os valores dos dados são mapeados em cores para fornecer uma representação visual adicional. Valores baixos são mapeados em azul, valores médios em vermelho e valores altos em verde. Valores ausentes são mapeados em preto. Essa técnica permite que os operadores de banco de dados acessem rapidamente padrões promissores dentro e entre registros ou amostras (BEDDOW, 1990). Como o *shape encoding* é uma técnica simples, flexível e capaz de explorar grandes conjuntos de dados de forma eficiente, é uma boa técnica a ser utilizada para identificar padrões e tendências em uma base de dados. A técnica de codificação visual mapeia valores de dados multidimensionais em formas geométricas, permitindo que seja visível como as variáveis se relacionam entre si, com a codificação de cores ajudando a identificar rapidamente valores extremos e a simplificar o campo visual. Conforme Livingston e Decker (2011), o *shape encoding* é considerado uma técnica simples porque utiliza formas geométricas básicas e atributos visuais para representar dados multivariados. Os autores afirmam que o *shape encoding* é uma técnica fácil de entender e interpretar, pois as formas geométricas são familiares e intuitivas para a maioria das pessoas.

Torna-se evidente que a visualização de dados desempenha um papel fundamental na compreensão e interpretação de dados complexos, principalmente quando se pretende identificar padrões entre eles, atuando como uma ponte entre a análise de dados e a tomada de decisões informadas. Os algoritmos e técnicas discutidos fornecem ferramentas poderosas para representar informações de maneira clara, concisa e impactante, possibilitando uma visualização mais intuitiva e acessível.

## 2.4 DISCUSSÕES SOBRE VALORES HUMANOS RELACIONADOS

Mesmo que a coleta de dados tenha se tornado uma parte essencial do mundo moderno e que as técnicas descritas anteriormente se beneficiem com grandes volumes de dados detalhados, à medida que coletamos cada vez mais informações sobre indivíduos e sociedades, surge a questão dos valores humanos relacionados a essa prática. A coleta de dados, quando mal administrada, pode levantar preocupações éticas, como a invasão de privacidade e o uso indevido das informações pessoais. Portanto, é crucial considerar os valores humanos fundamentais, como privacidade, autonomia e transparência, ao lidar com a coleta e o uso de dados.

Sendo assim, para a condução dessa pesquisa, é crucial não apenas assegurar que a coleta de dados não perpetue ou amplie desigualdades existentes, como vieses discriminatórios ou exclusão de determinados grupos, mas também discorrer sobre as implicações éticas para se ter certeza que os dados coletados para o trabalho foram coletados de forma ética. Devemos buscar a equidade no acesso aos dados e garantir que as informações coletadas sejam utilizadas de forma justa e imparcial, considerando os diferentes contextos e necessidades dos indivíduos e comunidades.

### 2.4.1 Ética em *Learning Analytics*

Para discutir os desafios éticos e de privacidade relacionados à *Learning Analytics*, é necessário primeiro compreender melhor ambos os conceitos e sua relação um com o outro. Ética pode ser definida como um código moral de normas e convenções que existe na sociedade, externo à pessoa, enquanto a privacidade é uma parte intrínseca da identidade e integridade de uma pessoa (DRACHSLER; GRELLER, 2016). A compreensão do que constitui um comportamento ético varia ao longo do tempo e entre culturas. Por outro lado, a privacidade é contextual, ou seja, estabelece os limites da pessoa ou identidade em relação às outras entidades. Assim, a compreensão da privacidade pode variar consideravelmente, mesmo entre pessoas que pertencem à mesma cultura, mas vivem em diferentes arranjos familiares.

No uso cotidiano e no pensamento popular, há uma sobreposição significativa entre ética e privacidade, o que às vezes leva a confusões ao discutir os efeitos da *Learning Analytics* em relação a ambos. Ambos os conceitos podem ser abordados de várias perspectivas, especialmente sociológica, psicológica, filosófica e religiosa. Eles se manifestam de maneira diferente em um contexto legal e estão sujeitos a mudanças ao longo do tempo.

A ética é a filosofia moral que envolve a sistematização, defesa e recomendação de conceitos de conduta correta e incorreta. Em contraste, a privacidade é um conceito vivo construído por meio de negociações contínuas dos limites pessoais com o ambiente ético ao redor. A ética de pesquisa tornou-se um tema relevante nos últimos anos, principalmente a partir de discussões sobre códigos de conduta nas ciências biomédicas, como o genoma humano, e mais recentemente, por meio da *Responsible Research and Innovation* (RRI, ou Pesquisa e Inovação Responsável) promovida pela Comissão Europeia<sup>1</sup>.

Os princípios éticos básicos para a pesquisa foram estabelecidos nos julgamentos de Nuremberg em 1949, onde foram usados para condenar médicos nazistas por suas atrocidades durante a Segunda Guerra Mundial. O chamado Código de Nuremberg é o primeiro manifesto de ética em pesquisa e contém dez princípios internacionalmente reconhecidos para a experimentação em seres humanos. Desde então, outros documentos como a Declaração de Helsinki (World Medical Association, 1964) e o Relatório Belmont (Department of Health, Education, and Welfare, 1978) foram desenvolvidos, fornecendo diretrizes éticas para pesquisas envolvendo seres humanos.

### 2.4.2 Coleta de dados por sistemas de frequência

Ao se discutir a ética em *Learning Analytics*, é fundamental abordar não apenas as implicações éticas das análises que serão realizadas ao decorrer deste trabalho de conclusão de curso, mas também a forma como os dados foram obtidos. A coleta de dados por sistemas de frequência desempenha um papel crítico nesse contexto, pois envolve a obtenção de informações sobre a presença e frequência dos alunos, que são dados sensíveis e pessoais.

<sup>1</sup> <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation>

A implementação de sistemas de frequência escolar pode gerar várias questões éticas em relação à proteção de dados dos alunos e professores, conforme estabelecido pela Lei Geral de Proteção de Dados (LGPD) no Brasil (BRASIL, 2019). Portanto, é importante que as escolas estabeleçam uma política clara de privacidade e proteção de dados, informando os alunos, professores e outros funcionários, incluindo diretores, coordenadores e secretários, sobre quais informações serão coletadas e como serão usadas e protegidas pelo sistema de frequência. Isso ajuda a garantir a transparência e a segurança na gestão dos dados dos envolvidos. Além disso, é importante haver uma discussão mais ampla na sociedade sobre quais dados podem ser acessados e por quem. Por exemplo, será que um professor poderia ter acesso aos dados de alunos de outra turma, ministrada por outro professor? Será que pais ou responsáveis podem ter acesso aos dados de estudantes maiores de idade? Uma escola poderia compartilhar informação com outras escolas? Quem poderia ter acesso a justificativas de faltas sensíveis, como casos de abuso e violência familiar? O acesso aos dados anonimizados podem ser utilizados para pesquisas científicas brasileiras?

A privacidade e segurança dos alunos e professores são preocupações essenciais quando se trata de coleta de dados de frequência escolar. Dessa forma, é fundamental que sejam estabelecidos limites claros sobre o acesso e uso desses dados, a fim de garantir que apenas pessoas autorizadas tenham acesso às informações coletadas pelo sistema de frequência, como professores, coordenadores pedagógicos e diretores escolares. É importante que essas pessoas tenham uma compreensão clara do propósito da coleta de dados e de suas responsabilidades em relação à proteção dos dados, tornando-se necessária a adoção de políticas e medidas de segurança adequadas, como a criptografia de dados e a autenticação de usuários. A literatura acadêmica destaca a importância da proteção dos dados pessoais em ambientes educacionais e a necessidade de medidas efetivas para garantir a privacidade e segurança desses dados (AMO et al., 2021).

Na mídia, há muitos casos de quebra de privacidade de estudantes, e ressalta-se cada vez mais o quão abusivas e invasivas são algumas técnicas utilizadas para controle de turmas. Um exemplo disso pode ser visto relatado por Jamal (2018): uma escola de ensino médio na China implementou um sistema de reconhecimento facial que analisa o comportamento dos alunos na sala de aula. A tecnologia de reconhecimento facial introduzida na escola registra as expressões faciais de todos os alunos enquanto estão nas salas de aula. O sistema escaneia a sala a cada 30 segundos e reconhece sete expressões diferentes, como neutra, feliz, triste, desapontada, assustada, brava e surpresa. O sistema é chamado de “Sistema de Gerenciamento de Comportamento em Sala de Aula Inteligente” e está sendo utilizado na *Hangzhou No. 11 High School*. Além de escanear as expressões faciais, o sistema tem a capacidade de analisar seis tipos de comportamento dos alunos, como ficar em pé, ler, escrever, levantar a mão, ouvir o professor e apoiar-se na mesa. As informações sobre como os alunos estão reagindo e se comportando durante as aulas são posteriormente enviadas aos professores para uma análise mais eficiente do comportamento dos estudantes. O sistema também é usado para monitorar a frequência dos

alunos.

Apesar de desafiar a privacidade dos alunos, a *Hangzhou No. 11 High School* pensa de forma diferente. Zhang Guanchao, vice-diretor da escola, afirmou que o sistema apenas coleta e analisa os resultados da análise de reconhecimento facial e os armazena em um banco de dados local, em vez de enviá-los para a nuvem. O vice-diretor acrescentou que o sistema não salva as imagens, a fim de evitar acesso não autorizado aos dados. Um estudante da escola de Hangzhou diz: “Antes, quando tinha aulas que não gostava muito, eu ficava preguiçoso e talvez cochilasse na mesa ou folheasse outros livros didáticos. Mas agora não me atrevo a me distrair depois que as câmeras foram instaladas nas salas de aula. É como se um par de olhos misteriosos estivesse me observando constantemente”.

No Brasil, a LGPD estabelece que os dados pessoais devem ser coletados de forma transparente e segura e que o armazenamento e processamento desses dados devem ser feitos com a devida proteção contra vazamentos e uso indevido, o cumprimento da lei requer que as escolas implementem medidas de segurança para proteger os dados pessoais dos estudantes, como a criptografia dos dados, o uso de senhas fortes e a realização de backups frequentes. Além disso, a LGPD estabelece que as escolas devem coletar dados pessoais de forma transparente e informar aos estudantes e seus responsáveis sobre como os dados serão utilizados e protegidos (FERREIRA et al., 2022). Dados de biometria, como no caso de reconhecimento facial, levantam questões ainda mais delicadas relacionadas à coleta e armazenamento de dados visuais (imagens ou representações latentes destas (WANG; DENG, 2021). É necessário um planejamento cuidadoso para garantir a proteção de dados pessoais na educação, especialmente com o aumento do uso de tecnologias digitais em sala de aula.



### 3 TRABALHOS RELACIONADOS

Após uma busca pela literatura, foi encontrada uma tese de doutorado escrita por Gray (2019), intitulada “*Learning Analytics Integrating Student Attendance Data*”, que é uma pesquisa que explora a aplicação de técnicas de análise de dados para melhorar a compreensão do comportamento dos alunos em relação à frequência às aulas. A tese examina como os dados de frequência podem ser coletados, analisados e usados para melhorar o desempenho acadêmico dos alunos. A tese argumenta que a análise dos dados de frequência pode fornecer informações valiosas sobre o comportamento dos alunos e ajudar a identificar padrões que possam ser usados para melhorar o desempenho acadêmico, e apresenta um estudo empírico que usa técnicas de aprendizado de máquina para analisar os dados de frequência dos alunos, mostrando ser possível prever com precisão o desempenho acadêmico dos alunos com base em seus padrões de frequência às aulas.

Durante o trabalho, a autora utilizou a metodologia de “*data-first*”, abordagem que prioriza os dados como o ponto central em todo o processo de tomada de decisões e desenvolvimento de estratégias, para isolar e explicar a discrepância nos dados de frequência dos alunos. Em vez de tentar encontrar um evento conhecido no conjunto de dados, ela tentou identificar eventos candidatos dentro dos dados e, em seguida, explicá-los como uma atividade secundária. Essa abordagem permitiu que ela explorasse os padrões de frequência dos alunos e identificasse possíveis eventos que poderiam estar afetando a frequência. A metodologia utilizada pela autora baseou-se nos resultados do modelo preditivo. Ela usou esses resultados para responder à seguinte pergunta de pesquisa: “Esse modelo pode ser usado para fornecer informações adicionais sobre os padrões de frequência dos alunos?”, identificando métodos e ferramentas apropriados para análise numérica e estatística. Cada candidato precisava ser correlacionado com um evento acadêmico, social ou físico. Para realizar sua pesquisa, a autora utilizou uma abordagem mista que combinava métodos quantitativos e qualitativos, coletando dados quantitativos sobre a frequência dos alunos usando o sistema de gerenciamento de aprendizagem da universidade. Também utilizou técnicas estatísticas avançadas para analisar os dados coletados. Ela usou análise descritiva para descrever as características básicas dos dados, como média, mediana e desvio padrão, e também uma análise multivariada para examinar as relações entre as variáveis e identificar possíveis padrões.

A autora realizou uma vasta revisão da literatura em várias áreas, incluindo ciência da computação, ciência de dados, psicologia, filosofia moral e educação. Essa revisão permitiu que a autora identificasse as melhores práticas e tendências atuais em análise de dados educacionais. Durante a tese foi investigado separadamente cada uma das ferramentas/ambientes/sistemas mais populares ou mais citados, usando fontes como publicações acadêmicas, folhetos de dados públicos ou literatura de produtos/demonstrações. Isto permitiu que ela avaliasse as diferentes ferramentas disponíveis para análise de dados educacionais e determinasse quais seriam mais adequadas para sua pesquisa. No entanto, é importante notar que essas fontes não produzem

uma avaliação completa em todas as situações, e a autora reconhece que essas fontes podem não fornecer uma avaliação completa das ferramentas em todas as situações e que pode haver outras ferramentas disponíveis que não foram incluídas em sua pesquisa.

Ao utilizar um modelo de aprendizado de máquina para identificar alunos que estavam em risco de se tornarem desengajados e faltosos e comparando os resultados do modelo com o método anteriormente utilizado pela universidade, a autora descobriu que o modelo de aprendizado de máquina identificou 62 alunos que não foram identificados pelo método anterior. Adicionalmente, a autora descobriu que a métrica de engajamento comportamental (BEM) era mais poderosa do que uma simples razão de sessões frequentadas, observando melhoria empírica de 12% usando a métrica BEM em comparação com a razão de engajamento (ER). No entanto, ela não forneceu análise numérica e teoria matemática para explicar por que isso pode ser verdade. Por fim, também foi descoberto um padrão de grandes variações na contagem de alunos acima e abaixo do valor médio BEM em cada semana entre as semanas 5 e 8 do primeiro semestre. Ela usou essa observação para tentar isolar possíveis causas ou explicações para o efeito da falta de engajamento dos alunos.

Há também a dissertação de Moissa (2016), que teve como o objetivo geral avaliar a influência das ferramentas de LA na interação, desempenho e satisfação dos alunos em ambientes virtuais de aprendizagem. Para alcançar esse objetivo, a autora propõe a realização de um experimento com usuários reais, a análise dos dados coletados e a comparação entre os alunos que tiveram acesso à ferramenta e os alunos que não tiveram, bem como entre os alunos que usaram a ferramenta e os que não usaram. Ainda que a autora teve foco específico em ambiente educacionais virtuais, o trabalho possui semelhança com o atual por buscar entender como as ferramentas de LA podem ser utilizadas para melhorar a experiência de aprendizagem de estudantes, compreendendo a influência dessas ferramentas para ajudar a desenvolver estratégias mais eficazes de ensino.

Para realizar a coleta dos dados, a autora desenvolveu ferramentas para auxiliar no processo de ensino-aprendizagem e realizou um experimento com usuários reais. Através desse experimento, foram coletados dados sobre a interação dos alunos com a ferramenta, seu desempenho e sua satisfação. Os dados foram analisados através de técnicas estatísticas e de Mineração de Dados, como análise de regressão, análise de correlação e análise de *cluster*. Além disso, a autora aplicou um questionário de satisfação para coletar a opinião dos alunos em relação ao minicurso realizado e à ferramenta utilizada, sendo o questionário composto por questões objetivas, questões discursivas e questões de múltipla escolha.

### 3.1 INICIATIVAS NO EXTERIOR

Estudos de ponta sobre sistemas de frequência escolar no exterior costumam ser realizados em países classificados como desenvolvidos, sempre com o objetivo de garantir que os estudantes estejam frequentando as aulas e participando ativamente das atividades escolares, o que contribui

para uma educação de qualidade e para a promoção da igualdade de oportunidades na sociedade. Isso pode significar não apenas a detecção da frequência por si só, mas também como o aluno está se comportando dentro da sala de aula, ou seja, se está realmente presente.

Uma das tecnologias mais utilizadas em sistemas de frequência de ponta é a detecção facial, seja por fotos ou por vídeos. Neste contexto, dois trabalhos são apresentados. O primeiro, (BHAT et al., 2020), utilizou a tecnologia de *Deep Learning* para a detecção facial em vídeos de até 33 *frames per second* (fps). O sistema proposto foi desenvolvido com base em uma abordagem de aprendizado de máquina conhecida como *One Shot Learning*, que envolve a capacidade de reconhecer objetos ou rostos com apenas uma única imagem de treinamento, o tornando capaz de identificar um grande número de alunos com apenas uma foto (*frame*) de referência. A precisão do sistema é superior a 90%, o que faz com que seja possível concluir que o sistema poderia ser útil para aplicação futura não apenas em escolas, mas também em empresas, possuindo baixo custo de implementação.

O segundo trabalho, de Chauhan et al. (2022), implementou um sistema de frequência baseado em reconhecimento facial, utilizando uma metodologia baseada na combinação de Análise de Componentes Principais (PCA) com Redes Neurais Convolucionais (CNN). O sistema é capaz de realizar uma verificação cruzada entre a imagem capturada e as imagens do banco de dados, e registrar a frequência do estudante caso haja correspondência. Os dados são então salvos em uma planilha do *Excel*.

Um outro trabalho apresenta uma maneira de detecção de frequências utilizando Internet das Coisas (IOT). Alassery (2019) teve como objetivo desenvolver um sistema de frequência baseado em *Radio Frequency Identification* (RFID), ou Identificação de Rádio Frequência, com foco em segurança, portabilidade e prontidão para ser implantado em larga escala. O sistema fornece uma solução prática e eficiente para monitorar a frequência dos alunos, utilizando o sistema IoT para registrar e buscar dados em um servidor ou nuvem, tornando-os disponíveis para o usuário a qualquer momento e em qualquer lugar. O trabalho também propõe que os próprios alunos possam conferir os próprios dados a qualquer momento. O artigo de Alassery (2019) também utiliza de um sistema IoT, porém com a tecnologia *Wireless Sensor Networks* (WSNs), cuja proposta era criar cadeiras inteligentes, equipadas com sensores de peso e amplificadores que enviam sinais digitais para um receptor, permitindo a identificação da presença dos alunos durante o horário de aula. Também foi instalado um leitor de impressões digitais para aumentar a segurança e identificação dos alunos.

Outra solução para o problema de detecção de presenças são os dispositivos vestíveis, cujo mapeamento sistemático da literatura de Ferreira, Rodriguez e Motti (2020) aponta bons estudos da área. Segundo os autores, os países que possuem estudos na área são Estados Unidos, França, Itália, China, Japão, Eslovênia e Reino Unido, a maioria com crianças do ensino fundamental de entre 7 a 13 anos. O trabalho investigou o posicionamento dos sensores, a maneira de armazenamento de dados e a validação dos dados, trazendo à tona que alguns dos trabalhos também focavam em coletar dados sobre a saúde física dos estudantes. O trabalho também cita

que nenhum estudo brasileiro foi encontrado, trazendo a hipótese de que a baixa qualidade da infraestrutura de muitos contextos educacionais brasileiros poderia ser um impeditivo desse tipo de estudo em território nacional.

### 3.2 INICIATIVAS NO BRASIL

A Base de Dados do Cadastro Único para Programas Sociais (CadÚnico) é uma ferramenta utilizada pelo Governo Brasileiro para registrar informações sobre famílias em situação de vulnerabilidade social, além de ser uma importante fonte de dados sobre a escolaridade no Brasil, pois registra informações sobre a escolaridade dos membros das famílias cadastradas que são utilizadas para monitorar e avaliar as políticas públicas de educação, bem como para identificar as famílias que precisam de apoio para garantir o acesso e a permanência na escola (GARCIA; HILLESHEIM, 2017). Desde sua criação, a base de dados CadÚnico tem recebido diversos prêmios, como o prêmio de Inovação na Gestão Pública Federal, concedido em 2017 pelo Ministério do Planejamento, Desenvolvimento e Gestão, e o Prêmio Excelência em Governo Eletrônico, concedido em 2016 pela Associação Brasileira de Entidades Estaduais de Tecnologia da Informação e Comunicação.

No que diz respeito à unificação da base de frequência com o chamado Sistema Presença, o Brasil tem buscado implementar medidas para garantir o registro da frequência escolar de todos os estudantes. O Sistema Presença foi lançado em 2019 pelo MEC com o objetivo de padronizar as informações sobre a frequência escolar em todo o país. Antes da criação do sistema, as informações sobre a frequência dos estudantes eram registradas em diferentes sistemas pelos estados e municípios, o que dificultava o acompanhamento e a análise dessas informações. Vale destacar que a chave única de identificação dos alunos neste sistema é o Número de Identificação Social (NIS), que é gerado pelo CadÚnico. No entanto, a implementação do Sistema Presença tem enfrentado alguns desafios, como a necessidade de capacitação dos profissionais envolvidos no registro e análise dos dados, para que os responsáveis pela coleta e análise dessas informações estejam devidamente preparados para lidar com as novas tecnologias e sistemas de informação. Outro desafio é a questão de investimentos em infraestrutura tecnológica para garantir o funcionamento do sistema em todo o país, o que envolve desde a aquisição e manutenção de equipamentos e softwares até a disponibilidade de uma conexão de internet estável e rápida, especialmente em regiões mais remotas (ECHAZARRA; RADINGER, 2019). Por fim, a pandemia de COVID-19 trouxe novos desafios para a implementação do sistema, especialmente no que diz respeito à adaptação do sistema às novas modalidades de ensino remoto. Com a suspensão das aulas presenciais em muitas regiões do país, foi necessário buscar alternativas para registrar a frequência dos estudantes em aulas virtuais e atividades remotas. Isso exigiu uma adaptação rápida do sistema e dos profissionais envolvidos, bem como investimentos adicionais em tecnologias e recursos para permitir a coleta dessas informações.

Com o Sistema Presença, o acompanhamento da frequência dos estudantes tornou-se mais

efetivo, mesmo com os desafios de questões profissionais e monetárias, o que pode contribuir para a promoção de políticas públicas mais eficazes na área da educação. Em uma pesquisa realizada pelo MEC, em 2019, ano de implementação do sistema, a frequência escolar alcançou os melhores índices históricos, como é possível verificar pela Figura 1 (Ministério da Educação, 2019). As duas imagens representam o histórico de frequência entre os anos de 2007 (um ano após a implementação do Sistema Presença) e 2019, indicando que houve índices recordes de frequência escolar em todo o país; ou seja, a taxa de alunos beneficiários dentro da sala de aula, entre fevereiro e março de 2019, chegou a 90,31%, e entre abril e maio, em 89,81%.

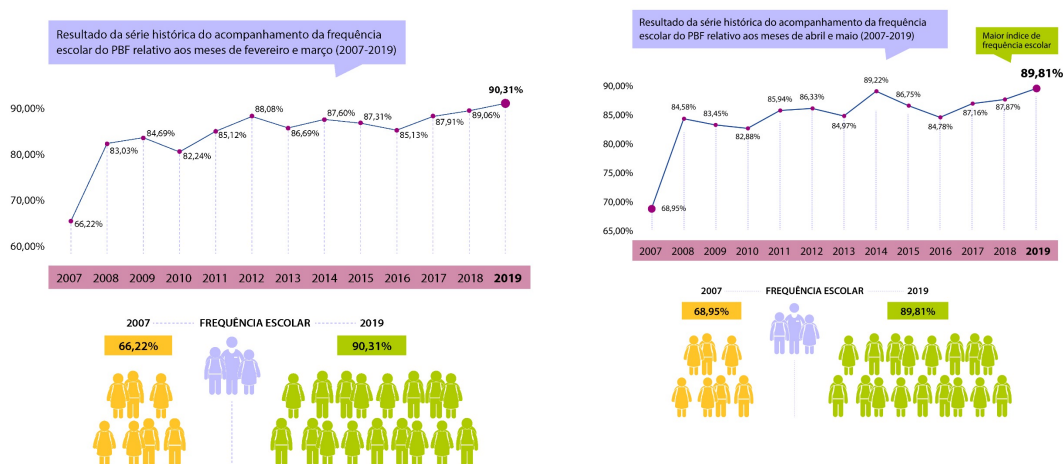


Figura 1 – Histórico de frequência de estudantes no Brasil entre os meses de fevereiro e março, e de abril e maio de 2019.

O Sistema Presença tem limitações importantes. Como o objetivo deste sistema é fornecer dados para o Bolsa Família, apenas alunos beneficiários deste programa têm sua frequência coletada. Além disso, a frequência é coletada com granularidade mensal. Isto impossibilita o desenvolvimento de ferramentas de alerta de evasão, baseadas em Inteligência Artificial, que sejam capazes de detectar padrões nos dados e permitam ações de combate à evasão em um estágio precoce. Por outro lado, não há a integração dos dados de frequência com outros dados relevantes do estudante, como dados de desempenho escolar. Embora o Sistema Presença ainda esteja em fase de implementação, desde o início dos anos 2000 houveram estudos foram conduzidos com o objetivo de resolver o problema de coleta de dados de frequência de forma mais eficiente (SOLIGO, 2010). É o caso do trabalho de Pereira et al. (2012), que apresenta uma proposta de monitoramento do estudante no momento de chegada e saída da instituição, sendo realizada em tempo real a comunicação aos responsáveis através do envio de SMS. Adicionalmente, será possível a geração de relatórios acerca da frequência escolar do aluno. A utilização da aplicação proposta visa tornar o controle de frequência mais rápido e eficiente, além de proporcionar a comunicação dos dados coletados aos responsáveis do discente, estabelecendo a sensação de segurança por parte da instituição. Além da frequência, há dispositivos que buscam

coletar outros tipos de dados de alunos em sala de aula, como demonstrado no projeto de Ferreira et al. (2022). Neste projeto, foram utilizados objetos vestíveis para detectar os movimentos das crianças do ensino fundamental, a fim de observar o comportamento dos alunos em sala de aula e compreender melhor a rotina escolar. A metodologia utilizada para coletar os dados de movimento das crianças foi a utilização de um dispositivo *wearable Actigraph GT9X Link*, que possui dois acelerômetros e um giroscópio, sendo estes os sensores de captação de movimento. As informações geradas tratam-se de registros em arquivo dos pontos dos eixos de (x, y, z) de cada um dos sensores para cada centésimo de segundo. Para ser traçado o perfil dos estudantes voluntários, foram enviados formulários aos responsáveis que envolviam perguntas como idade, peso, altura, sexo, mão com a qual escreve e se o aluno possui alguma deficiência diagnosticada. Durante a coleta de dados, foram feitas anotações específicas de todos os movimentos, atividades e atenção de cada aluno por um período de 30 minutos, e os demais movimentos no restante do tempo foram anotados apenas quando tratavam-se de levantar, sentar, andar e sair da sala. Por mais que o estudo não fosse conclusivo e não apresentasse resultados específicos da coleta de dados de movimento das crianças, foi concluído que a utilização de sensores vestíveis em sala de aula pode auxiliar o professor a melhorar a qualidade das palestras, especialmente se as informações forem fornecidas em tempo real, principalmente quando os sensores conseguem estimar o nível de engajamento dos alunos na aula e fornecem essa informação diretamente ao educador.

Contudo, é importante reforçar que a coleta de dados educacionais no Brasil apresenta desafios que vão além da infraestrutura de tecnologia da informação. Em regiões mais remotas do país, como comunidades ribeirinhas e áreas rurais, muitas escolas têm dificuldades para enviar e receber informações por meio de sistemas eletrônicos (COSTA, 2021). Além disso, a realidade das escolas que atendem adolescentes em conflito com a lei, como a Fundação Casa (antiga Febem), também apresenta dificuldades para a coleta de dados educacionais: Em primeiro lugar, muitos desses jovens chegam à instituição com histórico de evasão escolar e baixa escolaridade, o que dificulta o acompanhamento de sua trajetória educacional. A rotatividade de professores e funcionários nessas instituições, aliada a uma infraestrutura muitas vezes precária, pode dificultar o registro e a organização dos dados educacionais dos alunos. Por fim, há ainda a questão da própria natureza das atividades desenvolvidas nessas instituições, que podem ser muito diferentes das atividades escolares convencionais e, portanto, requerem uma avaliação mais específica e adaptada. O atendimento educacional dessas instituições muitas vezes é feito de forma precária e com poucos recursos, o que pode prejudicar a coleta de informações precisas sobre a frequência e o desempenho dos estudantes (MORAIS; FERREIRA, 2019). Essas diferenças no contexto educacional brasileiro tornam ainda mais desafiadora a tarefa de coletar e analisar dados educacionais para subsidiar políticas públicas efetivas.

Em suma, a pluralidade brasileira apresenta desafios significativos para a coleta e análise de dados educacionais. Embora iniciativas como o Cadastro Único para Programas Sociais e o Sistema Presença tenham sido criadas para padronizar as informações sobre a frequência escolar

e a vulnerabilidade social das famílias, ainda existem muitos obstáculos a serem superados. A infraestrutura de tecnologia da informação, por exemplo, é um desafio em regiões remotas, e a rotatividade de professores e funcionários em instituições como a Fundação Casa pode dificultar o registro e a organização dos dados educacionais dos alunos. Portanto, novas pesquisas podem ser realizadas para aprimorar essas iniciativas e desenvolver soluções tecnológicas que possam superar esses desafios e garantir a coleta de informações precisas sobre a educação no Brasil. Também é importante ressaltar que a avaliação das atividades educacionais em instituições específicas, como a Fundação Casa, requer uma abordagem mais específica e adaptada, uma vez que essas atividades podem ser muito diferentes das atividades escolares convencionais.

### 3.3 ANÁLISE COMPARATIVA

Nesta seção, será realizada uma análise comparativa entre os trabalhos relacionados e os dados coletados para a pesquisa de detecção de padrões em relação à presença dos estudantes no ensino básico brasileiro. Os trabalhos relacionados abordam diferentes aspectos da análise de dados educacionais e fornecem conclusões relevantes para a compreensão dos padrões de frequência dos alunos. Por mais que os trabalhos relacionados no exterior e no Brasil tenham grande enfoque na aplicação e otimização de sistemas de frequência, é importante que sejam estudados para que sejam percebidas quais as variáveis mais relevantes durante a coleta de frequência escolar. Sendo assim, é importante destacar que o trabalho atual não irá focar na aplicação de um novo sistema de frequência, e sim, em aplicar algoritmos de LA e EDM em dados que serão coletados futuramente, o que já se torna a diferença principal entre os trabalhos vistos nas seções 3.1 e 3.2. Ao comparar esses trabalhos relacionados com a pesquisa proposta, observa-se que todos eles compartilham o objetivo comum de compreender e melhorar a frequência dos alunos no ensino básico. No entanto, cada estudo e país possui suas peculiaridades e desafios específicos, o que demanda abordagens adaptadas às suas realidades.

O estudo realizado por Gray (2019) aborda a aplicação de técnicas de análise de dados para compreender o comportamento dos alunos em relação à frequência escolar. A autora utiliza técnicas de aprendizado de máquina para analisar os dados de frequência dos alunos e identificar padrões que possam contribuir para melhorar o desempenho acadêmico. A metodologia adotada incluiu a coleta de dados quantitativos sobre a frequência dos alunos e a aplicação de técnicas estatísticas avançadas para análise desses dados. Por outro lado, a dissertação de Moissa (2016) tem foco na avaliação do impacto das ferramentas de aprendizagem analítica (LA) na interação, desempenho e satisfação dos alunos em ambientes virtuais de aprendizagem. Embora o contexto seja diferente do estudo proposto, existe uma relevância em comum, que é entender como as ferramentas de LA podem ser utilizadas para aprimorar a experiência de aprendizagem dos estudantes. A autora conduziu um experimento com usuários reais, coletando dados sobre a interação dos alunos com a ferramenta e utilizando técnicas estatísticas e de Mineração de Dados para análise dos dados coletados. A abordagem mista, que combinou métodos quantitativos e qualitativos,

proporcionou uma compreensão mais abrangente dos resultados obtidos. No entanto, o presente trabalho tem um foco distinto, concentrando-se em outras áreas, como a aplicação da detecção de padrões e a visualização das informações dos resultados obtidos. Busca-se explorar como essas técnicas podem ser utilizadas para identificar padrões de frequência escolar e visualizar esses dados de forma compreensível e significativa. Dessa forma, pretende-se oferecer insights valiosos aos educadores e formuladores de políticas educacionais, permitindo uma tomada de decisão embasada na compreensão dos padrões de frequência dos alunos. Assim, apesar das semelhanças entre os estudos mencionados, o atual trabalho busca aprimorar o conhecimento na área, concentrando-se em aspectos específicos da análise de dados e visualização das informações sobre a frequência escolar, com o objetivo de contribuir para a melhoria da experiência educacional dos alunos.

Na seção 3.1, foram apresentadas iniciativas no exterior que abordam a detecção de frequência escolar por meio de tecnologias avançadas, como detecção facial, RFID e dispositivos vestíveis. Essas abordagens demonstram a diversidade de soluções utilizadas para coletar e analisar dados de frequência dos alunos. Os estudos apresentados destacam a precisão e eficiência das tecnologias utilizadas, além de considerar aspectos como segurança, portabilidade e disponibilidade dos dados. Já na seção 3.2, foram abordadas iniciativas no Brasil, como o Cadastro Único para Programas Sociais (CadÚnico) e o Sistema Presença. O CadÚnico é uma fonte de dados importante para monitorar a escolaridade no país, registrando informações sobre a frequência dos alunos. O Sistema Presença, por sua vez, busca padronizar as informações sobre a frequência escolar em todo o país. Ele é utilizado pelos gestores educacionais para registrar a presença dos alunos nas escolas, permitindo o monitoramento em tempo real da frequência e a identificação de possíveis problemas de evasão escolar. O sistema também auxilia na tomada de decisões e na formulação de políticas públicas para garantir a universalização do acesso à educação, e o *software* é o que servirá para aplicação do NEES.



## 4 PROPOSTA

O objetivo do presente trabalho de conclusão de curso é investigar os padrões de características dos estudantes do ensino básico, abrangendo a pré-escola, o ensino fundamental e o ensino médio, com base na sua presença em sala de aula. Especificamente, busca-se coletar informações dos estudantes, como notas, gênero, idades e registros de presenças em sala de aula, a fim de identificar possíveis correlações com a evasão escolar.

O projeto coordenado pelo NEES tem como objetivo desenvolver um software de coleta de dados voltado para dispositivos móveis de baixo custo. Esse software será utilizado para coletar dados dos alunos e armazená-los em um data lake, que será modelado durante o processo de desenvolvimento do software. O objetivo do NEES é garantir uma experiência de uso pouco intrusiva e flexível para garantir adesão na educação básica, considerando a diversidade das redes em suas diferentes modalidades, como a educação do campo, quilombola e educação especial, além das diferentes etapas de ensino. O software planejado permitirá que as escolas e instituições de ensino possam identificar problemas e desenvolver soluções para melhorar a qualidade da educação, e também pode ser utilizado para prevenir a evasão escolar, por meio de um sistema de alerta preventivo.

É importante reforçar que ainda não foram disponibilizados pelo NEES os dados referentes à frequência individual dos estudantes, pois será realizado um teste piloto do *software* no segundo semestre de 2023. Porém, o NEES disponibilizou outros dados de bases diversas que ajudarão a formar um panorama mais complexo e robusto da situação do ensino básico brasileiro. Entre essas bases, há dados do censo escolar do INEP, Sistema Universal de Saúde (SUS), Instituto Brasileiro de Geografia e Estatística (IBGE), etc. Os dados disponibilizados envolvem, mas não estão limitados a:

1. Estimativa da população por idade por município entre 2014 a 2021;
2. Projeção da população por idade de 2010 a 2060, realizada em 2018, por estado;
3. Quantidade de matriculados no ensino básico por municípios entre 2014 a 2021;
4. Quantidade de matriculados no ensino básico por idade e cor e raça;
5. Quantidade de turmas no ensino básico nos anos iniciais e nos anos finais;
6. Base dos indicadores de cada estado e município com relação às metas um, dois, três e seis do programa PNE Em Movimento<sup>1</sup>. As metas referentes são:
  - a) Meta 1: Universalizar, até 2016, a educação infantil na pré-escola para as crianças de 4 a 5 anos de idade e ampliar a oferta de educação infantil em creches de forma a atender, no mínimo, 50% das crianças de até 3 anos;

<sup>1</sup> <https://pne.mec.gov.br/assistencia-tecnica/programas-do-mec-metas>

- b) Meta 2: Universalizar o ensino fundamental de 9 anos para toda a população de 6 a 14 anos e garantir que pelo menos 95% dos alunos concluam essa etapa na idade recomendada;
  - c) Meta 3: Universalizar, até 2016, o atendimento escolar para toda a população de 15 a 17 anos e elevar, até o final do período de vigência deste PNE, a taxa líquida de matrículas no ensino médio para 85%;
  - d) Meta 6: as escolas públicas, de forma a atender, pelo menos, 25% dos (as) alunos (as) da educação básica.
7. Microdados do Sistema Nacional de Avaliação da Educação Básica (SAEB) dos anos de 2015, 2017 e 2019, que avaliam informações de questionários aplicados a alunos, professores, diretores e Secretários Municipais de Educação de cada escola em cada município;
  8. Tipo de localização da escola (urbana ou rural);
  9. Microdados do censo escolar de 2021 realizado pelo INEP (qualidade do acesso a internet, tipo de localização da escola, disponibilidade de bibliotecas, dentre outros);

Considerando essas variáveis e levando em consideração que as técnicas de reconhecimento de padrões serão aplicadas após a obtenção dos dados de frequência escolar, propõe-se neste trabalho uma análise da relação entre a quantidade de turmas, o número de matrículas e o quão cada escola da educação básica atingiu os por município no período de 2014 a 2021. Para esta análise serão considerados estes e outros dados, como a a variação da população dos municípios e estados, a faixa etária dos estudantes e a autodeclaração étnico-racial dos estudantes, bem como e as respostas dos questionários da SAEB. Essa análise de abordagem multidimensional permitirá traçar um panorama das condições das escolas no Brasil e examinar como houve melhorias ao longo dos últimos anos, ou identificar oportunidades para futuras melhorias. Essas informações serão essenciais para personalizar ainda mais a aplicação futura do software desenvolvido pelo NEES, de modo a atender de forma mais precisa às necessidades dos estudantes em diferentes ambientes educacionais, bem como obter uma compreensão mais abrangente do panorama educacional brasileiro, identificando tendências e padrões que possam direcionar políticas e ações educacionais para o aprimoramento contínuo das escolas e do sistema educacional como um todo.

#### 4.1 CONCLUSÃO

Para alcançar o objetivo proposto, serão investigados os dados educacionais dos estudantes do ensino básico, contemplando informações como notas obtidas nas disciplinas, gênero, idades e registros de frequências em sala de aula. Esses dados serão analisados por meio de técnicas de mineração de dados e análise estatística, visando identificar padrões e relações que

possam contribuir para a compreensão da evasão escolar. A participação desse trabalho no contexto do projeto celebrado entre o NEES e o MEC pelo TED 11476 proporcionará acesso a uma ampla gama de recursos e conhecimentos, além de uma colaboração interinstitucional que enriquecerá o desenvolvimento deste trabalho. A parceria entre o NEES-UFAL, UDESC e UFAM permitirá a utilização de metodologias e ferramentas avançadas, promovendo uma análise abrangente e rigorosa dos dados coletados.

Os resultados esperados deste estudo serão a identificação de padrões e características dos estudantes do ensino básico que possam estar correlacionados com a evasão escolar. Essas descobertas poderão contribuir para a formulação de políticas educacionais mais efetivas e intervenções direcionadas para prevenir e combater a evasão escolar, promovendo uma educação mais inclusiva e de qualidade.

Os itens definidos sobre a metodologia de condução da pesquisa do trabalho de conclusão de curso foram os seguintes:

1. Estudar sistemas atuais de checagem de frequência de estudantes em sala de aula;
2. Estudar diferentes algoritmos de reconhecimento de padrões para grandes bases de dados;
3. Fazer levantamento de trabalhos relacionados sobre detecção de presença estudantil em sala de aula;
4. Estudar as bases de dados dos estudantes do ensino básico disponibilizadas pelo NEES;
5. Definir algoritmo(s) de reconhecimento de padrões que seja(m) mais apropriado(s) para as bases de dados disponibilizadas pelo NEES;
6. Aplicar algoritmo(s) de reconhecimento de padrões;
7. Gerar visualizações;
8. Analisar os resultados;
9. Escrever o trabalho de conclusão de curso.

A pesquisa até esse momento concluiu as etapas 1, 2 e 3. Sendo assim, o planejamento de tempo das outras etapas é mostrado na Tabela 1.

Etapas	2023/2				
	Ago	Set	Out	Nov	Dez
4					
5					
6					
7					
8					
9					

Tabela 1 – Planejamento de tarefas para o TCC 2

## REFERÊNCIAS

- AHMED, Muyeed; IMTIAZ, Mir Tahsin; KHAN, Raiyan. Movie recommendation system using clustering and pattern recognition network. In: **2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)**. [S.l.: s.n.], 2018. p. 143–147. Citado na página 15.
- ALASSERY, Fawaz. A smart classroom of wireless sensor networks for students time attendance system. In: **2019 IEEE Integrated STEM Education Conference (ISEC)**. [S.l.: s.n.], 2019. p. 324–331. Citado na página 34.
- ALDOWAH, Hanan; AL-SAMARRAIE, Hosam; FAUZY, Wan Mohamad. Educational data mining and learning analytics for 21st century higher education: A review and synthesis. **Telematics and Informatics**, Elsevier BV, v. 37, p. 13–49, abr. 2019. Disponível em: <<https://doi.org/10.1016/j.tele.2019.01.007>>. Citado 2 vezes nas páginas 22 e 24.
- ALI, Huda H.; KADHUM, Lubna Emad. K- means clustering algorithm applications in data mining and pattern recognition. In: . [S.l.: s.n.], 2017. v. 6. Citado na página 26.
- AMO, Daniel et al. Local technology to enhance data privacy and security in educational technology. **International Journal of Interactive Multimedia and Artificial Intelligence**, Universidad Internacional de La Rioja, v. 7, n. 2, p. 262, 2021. Disponível em: <<https://doi.org/10.9781/ijimai.2021.11.006>>. Citado na página 30.
- BEDDOW, Jeff. Shape coding of multidimensional data on a microcomputer display. In: **Proceedings of the First IEEE Conference on Visualization: Visualization '90**. [S.l.: s.n.], 1990. p. 238–246. Citado na página 28.
- BHAT, Aruna et al. Deep-learning based group-photo attendance system using one shot learning. In: **2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)**. [S.l.: s.n.], 2020. p. 546–551. Citado na página 34.
- BISHOP, Christopher Michael. **Pattern Recognition and Machine Learning**. [S.l.]: Springer, 2007. Citado na página 25.
- BRASIL. **Decreto nº 37.106, de 31 de Março de 1955**. 1955. Diário Oficial da União. <<https://www2.camara.leg.br/legin/fed/decret/1950-1959/decreto-37106-31-marco-1955-332702-publicacaooriginal-1-pe.html>>. Citado na página 13.
- BRASIL. **Lei Nº 9.394, de 20 de dezembro de 1996**. 1996. Diário Oficial da União. <[http://www.planalto.gov.br/ccivil\\_03/leis/l9394.htm](http://www.planalto.gov.br/ccivil_03/leis/l9394.htm)>. Citado na página 19.
- BRASIL. **Lei Nº 010172, de 9 de janeiro de 2001**. 2001. Diário Oficial da União. <[https://www.planalto.gov.br/ccivil\\_03/leis/leis\\_2001/l10172.htm](https://www.planalto.gov.br/ccivil_03/leis/leis_2001/l10172.htm)>. Citado na página 13.
- BRASIL. **Lei Nº 10.836, de 9 de janeiro de 2004**. 2004. Diário Oficial da União. <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2004-2006/2004/lei/l10.836.htm](https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.836.htm)>. Citado na página 13.
- BRASIL. **Lei Nº 11.096, de 13 de janeiro de 2005**. 2005. Diário Oficial da União. <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2004-2006/2005/lei/l11096.htm](https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/lei/l11096.htm)>. Citado na página 13.

BRASIL. **Lei Nº 13.709, de 14 de agosto de 2018**. 2019. Diário Oficial da União. <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm)>. Citado na página 30.

BRASIL. **Lei Nº 14.113, de 25 de dezembro de 2020**. 2020. Diário Oficial da União. <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2020/lei/l14113.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l14113.htm)>. Citado na página 13.

BRASIL. **Lei Nº 14.284, de 29 de dezembro de 2021**. 2021. Diário Oficial da União. <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2023-2026/2023/mpv/mpv1164.htm](https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/mpv/mpv1164.htm)>. Citado na página 13.

BUDIMAN, Andre et al. Student attendance with face recognition (lbph or cnn): Systematic literature review. **Procedia Computer Science**, v. 216, p. 31–38, 2023. ISSN 1877-0509. 7th International Conference on Computer Science and Computational Intelligence 2022. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S187705092202186X>>. Citado na página 15.

CAMPOS, Aline de et al. Mineração de dados educacionais e learning analytics no contexto educacional brasileiro: um mapeamento sistemático. **Informática na educação: teoria & prática**, v. 23, n. 3 Set/Dez, dez. 2020. Disponível em: <<https://seer.ufrgs.br/index.php/InfEducTeoriaPratica/article/view/102618>>. Citado na página 23.

CAMPOS, Aline De et al. Mineração de dados educacionais e learning analytics no contexto educacional brasileiro: um mapeamento sistemático. **Informática na educação: teoria & prática**, Universidade Federal do Rio Grande do Sul, v. 23, n. 3 Set/Dez, mar. 2021. Disponível em: <<https://doi.org/10.22456/1982-1654.102618>>. Citado na página 24.

CEREZO, Rebeca et al. Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. **Computers & Education**, Elsevier BV, v. 96, p. 42–54, maio 2016. Disponível em: <<https://doi.org/10.1016/j.compedu.2016.02.006>>. Citado na página 23.

CEREZO, Rebeca et al. Students' lms interaction patterns and their relationship with achievement: A case study in higher education. **Computers & Education**, v. 96, p. 42–54, 2016. ISSN 0360-1315. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360131516300264>>. Citado na página 26.

CHAUHAN, Vineet Kumar et al. Image-based attendance system using facial recognition. In: **2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)**. [S.l.: s.n.], 2022. p. 1487–1490. Citado na página 34.

CHEN, Weijun et al. Yolo-face: a real-time face detector. *Vis Comput*, n. 37, p. 805–813, 2021. Citado na página 15.

COSTA, Eliane Miranda. Escolas ribeirinhas e seus desafios: faces da educação do campo na amazônia marajoara. **Revista Teias**, v. 22, n. 66, p. 384–397, 2021. ISSN 1982-0305. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/revistateias/article/view/51951>>. Citado na página 37.

Department of Health, Education, and Welfare. **The Belmont Report**. 1978. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

<<https://web.archive.org/web/20111017133845/http://www.hhs.gov/ohrp/archive/documents/19790418.pdf>>. Citado na página 29.

DRACHSLER, Hendrik; GRELLER, Wolfgang. Privacy and analytics. In: **Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16**. ACM Press, 2016. Disponível em: <<https://doi.org/10.1145/2883851.2883893>>. Citado na página 29.

DUTRA-THOMÉ, Luciana; PEREIRA, Anderson Siqueira; KOLLER, Silvia Helena. O desafio de conciliar trabalho e escola: Características sociodemográficas de jovens trabalhadores e não-trabalhadores. **Psicologia: Teoria e Pesquisa**, FapUNIFESP (SciELO), v. 32, n. 1, p. 101–109, mar. 2016. Disponível em: <<https://doi.org/10.1590/0102-37722016011944101109>>. Citado na página 14.

ECHAZARRA, Alfonso; RADINGER, Thomas. Learning in rural schools: Insights from pisa, talis and the literature. OECD, 2019. Citado na página 35.

FERREIRA, Lamara et al. A panorama of the implementation of the general law for the protection of personal data (LGPD) in brazil: an exploratory survey. In: **2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)**. IEEE, 2022. Disponível em: <<https://doi.org/10.1109/ccwc54503.2022.9720879>>. Citado na página 31.

FERREIRA, Poliana; RODRIGUEZ, Carla; MOTTI, Vivian. Wearables para coleta de dados de estudantes em ambiente escolar: Mapeamento sistemático. In: **Anais do XXXI Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2020. p. 1353–1362. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12891>>. Citado na página 15.

FERREIRA, Poliana et al. Coleta de informações de movimento e atividades de crianças na escola com wearables: Relato de experiência. In: **Anais do XXXIII Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2022. p. 1102–1113. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/22485>>. Citado na página 37.

FERREIRA, Poliana Nascimento; RODRIGUEZ, Carla Lopes; MOTTI, Vivian Genaro. Wearables para coleta de dados de estudantes em ambiente escolar: Mapeamento sistemático. In: **Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)**. [S.l.]: Sociedade Brasileira de Computação, 2020. Citado 2 vezes nas páginas 19 e 34.

FOURNIER-VIGER, Philippe et al. A survey of sequential pattern mining. In: . [S.l.: s.n.], 2017. Citado na página 26.

GARCIA, Adir Valdemar; HILLESHEIM, Jaime. Pobreza e desigualdades educacionais: uma análise com base nos planos nacionais de educação e nos planos plurianuais federais. **Educar em Revista**, FapUNIFESP (SciELO), v. 33, n. spe.2, p. 131–147, set. 2017. Disponível em: <<https://doi.org/10.1590/0104-4060.51386>>. Citado na página 35.

GRAY, Cameron. **Learning Analytics Integrating Student Attendance Data**. Tese (Doutorado) — Profysgol Bangor University, 2019. Citado 2 vezes nas páginas 32 e 38.

GUBBELS, Jeanne; PUT, Claudia E. van der; ASSINK, Mark. Risk factors for school absenteeism and dropout: A meta-analytic review. **Journal of Youth and Adolescence**, v. 48, n. 48, 2019. ISSN 1637–1667. Disponível em: <<https://doi.org/10.1007/s10964-019-01072-5>>. Citado na página 15.

GURALNIK, Valery; SRIVASTAVA, Jaideep. Event detection from time series data. In: . [S.l.: s.n.], 1999. p. 33–42. Citado na página 14.

JAMAL, Mohammad. This school scans classrooms every 30 seconds through facial recognition technology. **TechJuice**, 2018. Disponível em: <<https://www.techjuice.pk/this-school-scans-classrooms-every-30-seconds-through-facial-recognition-technology/>>. Citado na página 30.

KOÇ, Mustafa. Learning analytics of student participation and achievement in online distance education: A structural equation modeling. **Educational Sciences: Theory & Practice**, Egitim Danismanligi ve Arastirmalari (EDAM), 2017. Disponível em: <<https://doi.org/10.12738/estp.2017.6.0059>>. Citado na página 20.

LIN, Jessica et al. Pattern recognition in time series. **Advances in Machine Learning and Data Mining for Astronomy**, v. 1, p. 617–645, 03 2012. Citado na página 26.

LIVINGSTON, Mark A.; DECKER, Jonathan W. Evaluation of trend localization with multi-variate visualizations. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2053–2062, 2011. Citado na página 28.

Ministério da Educação. **Frequência escolar de alunos bate mais um recorde da série histórica**. 2019.

<<http://portal.mec.gov.br/ultimas-noticias/12-aco-es-programas-e-projetos-637152388/77821-frequencia-escolar-de-alunos-bate-mais-um-recorde->da-serie-historica>>. Citado na página 36.

MOISSA, Barbara. **A Influência de Ferramentas de Learning Analytics na Interação, Desempenho e Satisfação dos Alunos**. Dissertação (Mestrado) — Universidade do Estado de Santa Catarina, Joinville, SC, 2016. Dissertação de Mestrado. Citado 2 vezes nas páginas 33 e 38.

MOISSA, Barbara; GASPARINI, Isabela; KEMCZINSKI, Avanilde. Educational data mining versus learning analytics: estamos reinventando a roda? um mapeamento sistemático. In: **Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)**. Sociedade Brasileira de Computação - SBC, 2015. Disponível em: <<https://doi.org/10.5753/cbie.sbie.2015.1167>>. Citado na página 23.

MONTEIRO, Doraliza Auxiliadora Abranches; FERREIRA, Marco Aurélio Marques; TEIXEIRA, Karla Maria Damiano. Determinantes da gestão do programa bolsa família: análise do índice de gestão descentralizada em minas gerais. **Saúde e Sociedade**, FapUNIFESP (SciELO), v. 18, n. 2, p. 214–226, jun. 2009. Disponível em: <<https://doi.org/10.1590/s0104-12902009000200005>>. Citado na página 13.

MORAIS, Juscislayne Bianca Tavares de; FERREIRA, Maria D'alva Macedo. Os desafios na escolarização de adolescentes em cumprimento de medida socioeducativa de liberdade assistida. **Revista de Políticas Públicas**, Universidade Federal do Maranhão, v. 23, n. 2, p. 621, dez. 2019. Disponível em: <<https://doi.org/10.18764/2178-2865.v23n2p621-639>>. Citado na página 37.

NASCENTES, Antenor. **Dicionário Etimológico Resumido**. [S.l.]: Instituto Nacional do Livro, 1966. Citado na página 18.

NASCIMENTO, João Carlos de Souza et al. Fracasso escolar e evasão no ensino médio no brasil: estado do conhecimento. **Revista Educar Mais**, v. 4, n. 2, p. 379–393, set. 2020. Disponível em: <<https://periodicos.ifsul.edu.br/index.php/educarmais/article/view/1823>>. Citado na página 15.

NOIRHOMME-FRAITURE, Monique. Visualization of large data sets: The zoom star solution. In: . [S.l.: s.n.], 2002. Citado na página 27.

OLIVEIRA, Thiago Nazareth De; BERNARDINI, Flavia; VITERBO, José. An overview on the use of educational data mining for constructing recommendation systems to mitigate retention in higher education. In: **2021 IEEE Frontiers in Education Conference (FIE)**. [S.l.: s.n.], 2021. p. 1–7. Citado na página 21.

PAOLANTI, Marina; FRONTONI, Emanuele. Multidisciplinary pattern recognition applications: A review. **Computer Science Review**, v. 37, p. 100276, 2020. ISSN 1574-0137. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1574013719300899>>. Citado na página 14.

PEREIRA, Ana Paula de Moura et al. Proposta de um sistema para acompanhamento automatizado da frequência escolar. **Anais do VII Congresso Norte-Nordeste de Pesquisa e Inovação (CONNEPI)**, 2012. Disponível em: <<https://propi.ifto.edu.br/ocs/index.php/connepi/vii/paper/view/3019>>. Citado na página 36.

PEREIRA, Zilma Maria. **Origem da Palavra: Presença**. 2011. <https://origemdapalavra.com.br/palavras/presenca/>. Citado na página 18.

RAMOS, Jorge et al. Crisp-edm: uma proposta de adaptação do modelo crisp-dm para mineração de dados educacionais. In: **Anais do XXXI Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2020. p. 1092–1101. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12865>>. Citado na página 21.

SANTOS, Jose Roberto et al. Estudo comparativo de plugins moodle para análise e acompanhamento da aprendizagem. In: **Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)**. Brazilian Computer Society (Sociedade Brasileira de Computação - SBC), 2019. Disponível em: <<https://doi.org/10.5753/cbie.sbie.2019.189>>. Citado na página 21.

SANTOS, Mariana Cristina Silva et al. Programa bolsa família e indicadores educacionais em crianças, adolescentes e escolas no brasil: revisão sistemática. **Ciência & Saúde Coletiva**, ABRASCO - Associação Brasileira de Saúde Coletiva, v. 24, n. Ciênc. saúde coletiva, 2019 24(6), p. 2233–2247, Jun 2019. ISSN 1413-8123. Disponível em: <<https://doi.org/10.1590/1413-81232018246.19582017>>. Citado 2 vezes nas páginas 13 e 14.

SAYED, J.; ASHOUR, M. W. Digital intelligent virtual assistant (diva) with natural speech and accent recognition. **IET Conference Proceedings**, Institution of Engineering and Technology, p. 170–175(5), 2021. Disponível em: <<https://digital-library.theiet.org/content/conferences/10.1049/icp.2022.0334>>. Citado na página 14.

SELFIDGE, O. G. Pattern recognition and modern computers. In: **Proceedings of the March 1-3, 1955, Western Joint Computer Conference**. New York, NY, USA: Association for Computing Machinery, 1955. (AFIPS '55 (Western)), p. 91–93. ISBN 9781450378567. Disponível em: <<https://doi.org/10.1145/1455292.1455310>>. Citado na página 14.



SOLIGO, Valdecir. Possibilidades e desafios das avaliações em larga escala da educação básica na gestão escolar. **Revista on line de Política e Gestão Educacional**, v. 9, n. 9, p. 1–15, jul 2010. Disponível em: <<https://periodicos.fclar.unesp.br/rpge/article/view/9275>>. Citado na página 36.

TSELENTIS, Dimitrios I.; PAPADIMITRIOU, Eleonora. Driver profile and driving pattern recognition for road safety assessment: Main challenges and future directions. **IEEE Open Journal of Intelligent Transportation Systems**, v. 4, p. 83–100, 2023. Citado na página 14.

UMBLEJA, Kadri; ICHINO, Manabu; YAGUCHI, Hiroyuki. Improving symbolic data visualization for pattern recognition and knowledge discovery. **Visual Informatics**, Elsevier BV, v. 4, n. 1, p. 23–31, mar. 2020. Disponível em: <<https://doi.org/10.1016/j.visinf.2019.12.003>>. Citado na página 27.

VARGAS, Hustana Maria; ZUCCARELI, Carolina. A nova face da docência: uma proposta de revisão do censo da educação superior. **Estudos em Avaliação Educacional**, scielo, v. 32, 00 2021. ISSN 0103-6831. Disponível em: <[http://educa.fcc.org.br/scielo.php?script=sci\\_arttext&pid=S0103-68312021000100103&nrm=iso](http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0103-68312021000100103&nrm=iso)>. Citado na página 14.

VELURI, Ravi Kishore et al. Learning analytics using deep learning techniques for efficiently managing educational institutes. **Materials Today: Proceedings**, Elsevier BV, v. 51, p. 2317–2320, 2022. Disponível em: <<https://doi.org/10.1016/j.matpr.2021.11.416>>. Citado na página 24.

WANG, Mei; DENG, Weihong. Deep face recognition: A survey. **Neurocomputing**, Elsevier, v. 429, p. 215–244, 2021. Citado na página 31.

WILL, Nilcimar Neitzel; KEMCZINSKI, Avaniilde; PARPINELLI, Rafael. Deep learning para previsão do desempenho do estudante: Um mapeamento sistemático da literatura. In: **Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)**. Brazilian Computer Society (Sociedade Brasileira de Computação - SBC), 2019. Disponível em: <<https://doi.org/10.5753/cbie.sbie.2019.1798>>. Citado na página 21.

World Medical Association. **World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects**. 1964. JAMA. <<https://jamanetwork.com/journals/jama/articlepdf/1760318/jsc130006.pdf>>. Citado na página 29.