



**MINISTÉRIO DA EDUCAÇÃO
INSTITUTO FEDERAL DO ESPÍRITO SANTO**

Campus Colatina

Av. Arino Gomes Leal, 1.700 - Santa Margarida - 29700-558 - Colatina - Espírito Santo - (27) 3723 – 1500

FOLHA DE APROVAÇÃO

**Edynelton Luiz de Carvalho Rio Branco
Gisele Vieira Luiz**

**A IMPORTÂNCIA ESTRATÉGICA DA MINERAÇÃO DE DADOS
EDUCACIONAIS - UMA ANÁLISE DE PESQUISAS RELEVANTES DE 2010 A 2022**

Trabalho de Conclusão de Curso apresentado como requisito parcial para conclusão do curso de Pós-graduação *Lato Sensu* em Conectividade e Tecnologias da Informação.

Aprovado em 25/07/2022.

COMISSÃO EXAMINADORA

Prof M.Sc Vanderson José Ildfonso Silva
Instituto Federal do Espírito Santo – Campus Colatina
Orientador

Prof D.Sc Alextian Bartholomeu Liberato
Instituto Federal do Espírito Santo – Campus Colatina
Membro Interno

Prof D.Sc Renan Osório Rios
Instituto Federal do Espírito Santo – Campus Colatina
Membro Interno



Emitido em 24/08/2022

FOLHA DE APROVAÇÃO-TCC N° 5/2022 - COL-CCGSI (11.02.21.01.08.02.03)

(N° do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 25/08/2022 06:07)

ALEXTIAN BARTHOLOMEU LIBERATO
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO COL -
CTMSI (11.02.21.01.08.02.12)
Matricula: 2630839

(Assinado digitalmente em 24/08/2022 17:58)

RENAN OSORIO RIOS
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO COL -
CCTI (11.02.21.01.08.02.08)
Matricula: 1828039

(Assinado digitalmente em 24/08/2022 21:28)

VANDERSON JOSÉ IDELFONSO SILVA
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO COL -
CTMSI (11.02.21.01.08.02.08)
Matricula: 1173010

Para verificar a autenticidade deste documento entre em <https://sipac.ifes.edu.br/documentos/> informando seu número: 5, ano: 2022, tipo: FOLHA DE APROVAÇÃO-TCC, data de emissão: 24/08/2022 e o código de verificação: 6ca514f3f9

A IMPORTÂNCIA ESTRATÉGICA DA MINERAÇÃO DE DADOS EDUCACIONAIS - UMA ANÁLISE DE PESQUISAS RELEVANTES DE 2010 A 2022

THE STRATEGIC IMPORTANCE OF EDUCATIONAL DATA MINING - AN ANALYSIS OF RELEVANT RESEARCH FROM 2010 TO 2022

Edynelton Luiz de Carvalho Rio Branco e Gisele Vieira Luiz.
Vanderson José Ildefonso Silva (orientador)

Instituto Federal do Espírito Santo. Campus Colatina

Resumo: A Mineração de Dados Educacionais é uma ferramenta tecnológica que tem por objetivo auxiliar em diversas soluções de problemas através de análises feitas em grandes bases de dados. Através dessas análises podemos gerar um conhecimento que leva a identificar e analisar os pontos fortes e fracos e ajudam na tomada de decisões estratégicas para a solução de problemas. Neste artigo buscamos analisar e mostrar os resultados obtidos através do uso da Mineração de Dados Educacionais para resolução de problemas de evasão escolar, nortear melhorias no sistema educacional brasileiro, definir os principais fatores que influenciam nas notas dos alunos, montar modelos avaliativos de desempenho dos alunos, analisar técnicas de aprendizagem de máquina e aprendizagem profunda.

Palavras-chave: mineracao; dados; educacionais; revisao; analise.

Abstract: Educational Data Mining is a technological tool that aims to help problem solving through comparisons in several databases. Through approaches, we can generate knowledge that leads to the identification and analysis of strengths and weaknesses and helps in making strategic decisions to solve problems. In this article, we seek to analyze and show the results obtained with Educational Data Mining to solve truancy problems, guide improvements in the Brazilian education system, and define the main factors that influence student grades, create evaluation models of student performance and analyze machine learning and deep learning techniques.

Keywords: mining; data; educational; revision; analysis.

1 INTRODUÇÃO

A mineração de dados é definida em termos de esforços para descoberta de padrões em bases de dados. A partir dos padrões descobertos, têm-se condições de gerar conhecimento útil para um processo de tomada de decisão. Trata-se, portanto, da aplicação de técnicas, implementadas por

meio de algoritmos computacionais, capazes de receber, como entrada, um conjunto de fatos ocorridos no mundo real e devolver, como saída, um padrão de comportamento, o qual pode ser expresso, por exemplo, como uma regra de associação, uma função de mapeamento ou a modelagem de um perfil. (AUGUSTO et al., 2016).

A mineração de dados pode ser útil para tomadas de decisão. Por exemplo, ao minerar um estoque de loja de games poderia –se descobrir qual categoria de game ou marca se vende mais, portanto, um gerente que obtém uma “nova informação” poderia planejar a compra de estoque e aumentar a quantidade de game ou marca que vende mais. A Mineração tem sido aplicada em várias áreas do conhecimento, como por exemplo, informática, saúde e educação. Com a expansão de cursos a distância, a área da Informática voltada na Educação tem mostrado muito interesse em utilizar mineração de dados para investigar perguntas científicas na área de educação. Dentre as ferramentas disponíveis para Mineração de Dados, uma que mais se destaca é a *Waikato Environment for Knowledge Analysis* (WEKA), que de acordo com Abernethy (2010) trata – se de um software de código aberto e gratuito que possibilita a transformação de dados em conhecimento útil.

Apesar das aplicações de Mineração de Dados serem implementadas em diversos setores, o foco deste trabalho é a Educação, onde a Mineração de Dados atua em desenvolver e adaptar algoritmos já existentes para compreender melhor os dados de contexto educacional.

Um dos principais objetivos de qualquer sistema educacional é equipar alunos com o conhecimento e as habilidades que precisam para ter sucesso na carreira que escolherem dentro de um período específico.

Diversos trabalhos vêm sendo desenvolvidos a partir de Mineração de Dados Educacionais (MDE) para fins de tomada de decisão (Rodrigues et al. 2013).

A aplicação da MDE tem como objetivo a descoberta de informações que ajudem na proposta educacional, no melhoramento das condições de infraestrutura escolar, no processo ensino-aprendizagem, na previsão de desempenho dos alunos, além de outros fatores que influenciam a aprendizagem (Baker et al.

2011), dentre os quais pode-se destacar a reprovação e a evasão escolar.

2 REFERENCIAL TEÓRICO

2.1 MINERAÇÃO DE DADOS: DEFINIÇÃO, FUNÇÃO

Mineração de dados (em inglês, data mining) é o processo de encontrar anomalias, padrões e correlações em grandes conjuntos de dados para prever resultados. Através de uma variedade de técnicas, você pode usar essas informações para aumentar a renda, cortar custos, melhorar o relacionamento com os clientes, reduzir riscos e mais.

Por ser considerada multidisciplinar, as definições acerca do termo Mineração de Dados variam de acordo com o campo de atuação dos autores. Consideradas como de maior expressão dentro da Mineração de Dados: Estatística, Aprendizado de Máquina e Banco de Dados. Em Zhou, é feita uma análise comparativa sobre três perspectivas:

- Em Hand et al., a definição é dada de uma perspectiva estatística: "Mineração de Dados é a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados".
- Em Cabena et al., a definição é dada de uma perspectiva de banco de dados: "Mineração de Dados é um campo interdisciplinar que junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados".

- Em Fayyad et al., a definição é dada da perspectiva do aprendizado de máquina: "Mineração de Dados é um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados."

Apesar das definições sobre a Mineração de Dados levar a crer que o processo de extração de conhecimento se dá de uma forma totalmente automática, sabe-se hoje que de fato isso não é verdade. Mesmo encontrando diversas ferramentas que nos auxiliam na execução dos algoritmos de mineração, os resultados ainda precisam de uma análise humana.

Ainda assim, a mineração contribui de forma significativa no processo de descoberta de conhecimento, permitindo aos especialistas concentrarem esforços apenas em partes mais importantes e substanciais dos dados.

2.2 MINERAÇÃO DE DADOS EDUCACIONAIS

A área emergente de Mineração de Dados Educacionais procura desenvolver ou adaptar métodos e algoritmos de mineração existentes, de tal modo que se prestem a compreender melhor os dados em contextos educacionais, produzidos principalmente por estudantes e professores, considerando os ambientes nos quais eles interagem, tais como AVAs, Sistemas Tutores Inteligentes (STIs), entre outros. Com tais métodos visa-se, por exemplo, entender melhor o estudante no seu processo de aprendizagem, analisando sua interação com o ambiente.

Assim, há a necessidade, por exemplo, de adequação dos algoritmos de mineração de dados existentes para lidar com especificidades inerentes aos dados educacionais, tais como a não independência estatística e a hierarquia dos

dados. Por outro lado, há uma necessidade significativa e urgente no provimento de ambientes computacionais apropriados para mineração de dados educacionais, oferecendo facilidades de uso para cada um dos atores envolvidos, notadamente ao professor.

2.3 ETAPAS DA MINERAÇÃO DE DADOS

O contínuo desenvolvimento e progresso, atualmente verificados, em áreas como a tecnologia da informação, Internet, sistemas gerenciadores de bancos de dados, entre outros, tem proporcionado o armazenamento de dados em heterogêneas e robustas base de dados (RISSINO, 2009).

Segundo Mathias (2015) o termo *KDD* – *Knowledge Discovery in Databases* (Descoberta de Conhecimento em Bases de Dados), foi cunhado em 1989, com o objetivo de representar todo o processo de busca e extração de conhecimento que, em seu nível mais operacional, inclui a aplicação de técnicas e algoritmos de mineração de dados (*datamining*) para manipular e encontrar indícios de correlação ou de implicação em grandes volumes de dados e que envolve para tal fim, a aplicação de diversas áreas do conhecimento humano, como: estatística, matemática, inteligência artificial, etc.

Segundo Castanheira (2008), para se iniciar um processo de *KDD* é necessário o entendimento do domínio da aplicação e dos objetivos finais a serem atingidos.

O *KDD* é empregado em vários campos de pesquisa com finalidades matemáticas, como por exemplo, estatísticas, aprendizado de máquinas, base de dados, reconhecimento de padrões, econômicas, desenvolvimento para técnicas e soluções para viabilizar a extração de dados de conhecimento em grandes bases de dados (FAYYAD et al., 1996a).

Sem o auxílio do processo de *KDD* em uma base de dados com um grande volume, um usuário passaria décadas para achar uma informação em um arquivo específico que

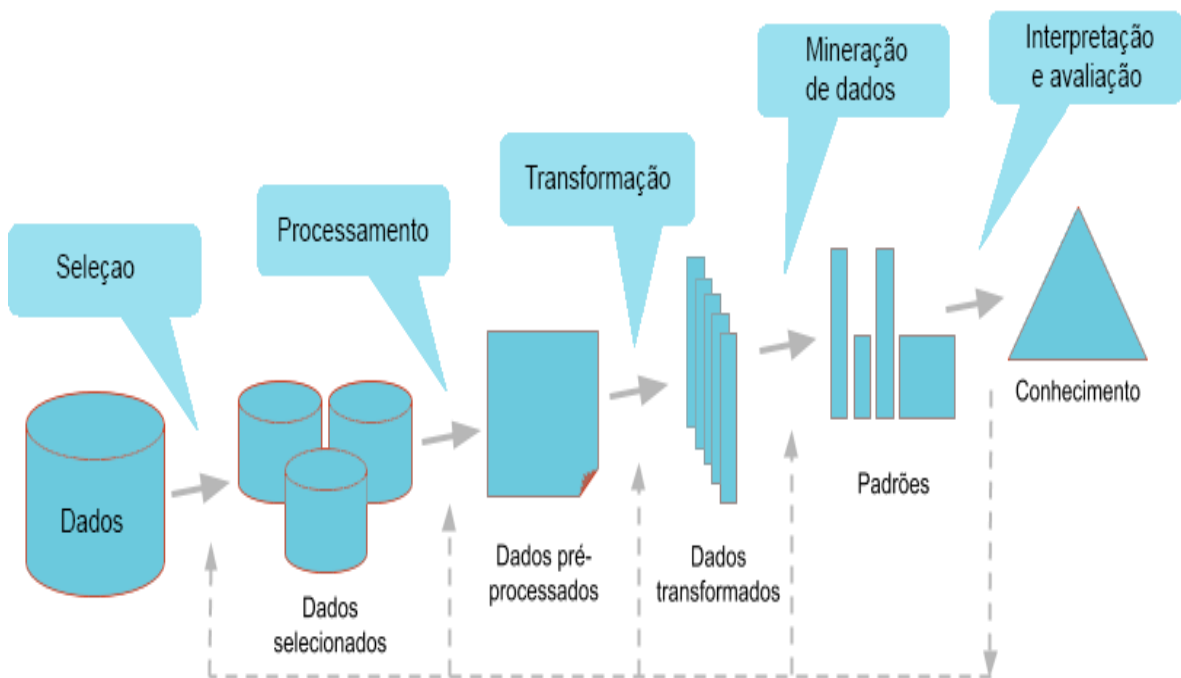
tenha sido enviado em apenas um dia, já com o auxílio do KDD, nessa mesma base de dados, essa informação seria facilmente encontrada em questão de segundos (FRAWLEY et al., 1992).

Atualmente, empresas que geram milhares de dados por dia necessitam de ferramentas automatizadas e inteligentes, na tarefa de interpretar, analisar e relacionar os dados da forma que possa ser

desenvolvida uma estratégia para cada aplicação (GOLDSCHIMIDT et al., 2005).

O processo de KDD ilustrado na Figura 1 envolve a avaliação e interpretação dos resultados com o objetivo de aplicar decisões sobre o que foi extraído envolvendo um processo que visa a tomada de decisões sobre as definições do conhecimento.

Figura 1: Etapas do processo de *KDD - Knowledge Discovery in Databases*



Fonte: Adaptado de FAYYAD et al. (1996)

A busca por conhecimento é um processo iterativo e interativo, que envolve três etapas principais, nas quais cada uma requer a capacidade de análise do usuário e de tomada de decisões. As principais etapas do processo de *KDD* são: pré-processamento, ligada a todas as funções relacionadas à organização, captação e ao tratamento dos dados; mineração de dados, na qual é realizada a busca por conhecimentos úteis na aplicação do *KDD*; pós-processamento, no qual é realizado o tratamento do conhecimento obtido através da mineração de dados com o objetivo de facilitar o entendimento do usuário em

relação ao conhecimento descoberto (FAYYAD et al., 1996).

2.3.1 ETAPAS DA DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

O processo de *KDD* é composto das seguintes etapas:

1. Consolidação de dados: esta etapa tem como objetivo coletar os dados e consolidá-los para o processo de extração do conhecimento. Na mesma são definidos os atributos relevantes para o processo, são definidos os problemas e a forma de

tratamento para os dados que não possuem valores e, se necessário, são removidas as exceções (WATSON, R. T., 2004).

2. Seleção de dados: Nesta parte inicial do processo, é feita a seleção e a identificação das informações dos dados nas bases de dados que serão utilizadas no processo de *KDD* e as informações selecionadas serão utilizadas durante todo o processo do *KDD*. (GOLDSCHMIDT et al., 2005);

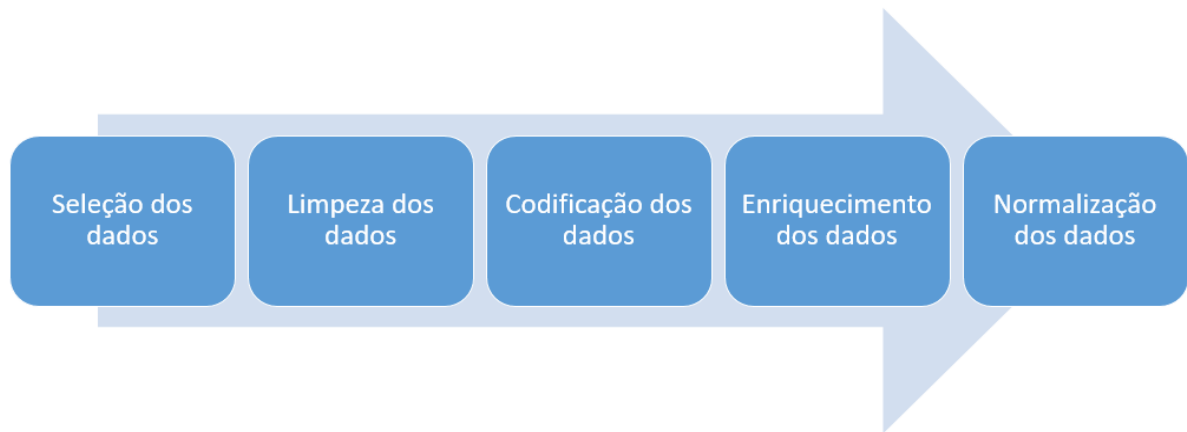
3. Pré-processamento: o objetivo desta fase é limpeza dos dados, tanto quanto

possível e adequá-los da melhor forma para sua utilização no processamento com as ferramentas de mineração de dados.

Os tipos de pré-processamento incluem tradicional projeção do banco de dados, seleção e mapeamento do valor e das funções de classificação e limpeza de dados (WILLIAMS, G. J e HUANG., 1996).

Nesta fase as etapas a serem seguidas são: limpeza, codificação, enriquecimento e normalização, conforme ilustrado na Figura 2;

Figura 2: Etapas do Pré-processamento em *KDD - Knowledge Discovery in Databases*



Fonte: Adaptado de WILLIAMS, G. J E HUANG (1996).

4. Mineração de dados: a mineração de dados é o processo de descoberta de informações úteis em grandes depósitos de dados (TAN et al., 2009). As técnicas de mineração de dados são planejadas, elaboradas e organizadas para atuar sobre os dados com o intuito de identificar e identificar padrões úteis e recentes que poderiam de outra forma, permanecer ignorados;

5. Pós-processamento: nesta etapa existem várias medidas para tratar o conhecimento descoberto; como por exemplo a visualização, análise e interpretação gerada pela etapa de mineração de dados. Envolve a depuração e as sínteses dos padrões nos volumes de dados.

O volume de conhecimento dos dados é muito extenso, praticamente é um conjunto

de informações relevantes ao processo que auxilia na análise das tomadas de decisões. Para auxiliar, nesse processo são utilizadas técnicas como padrões redundantes, relações irrelevantes, entre outros (CALIL et al., 2008).

Nessa etapa é muito comum ver o especialista do domínio da aplicação e o especialista de *KDD* trabalharem juntos, avaliando os resultados e os modelos de conhecimento e assim, definirem a melhor forma de utilizar os dados minerados.

2.3 ETAPAS DA MINERAÇÃO DE DADOS EDUCACIONAIS

A Mineração de Dados Educacionais (MDE) é uma área de

pesquisa interdisciplinar que lida com o desenvolvimento de métodos para explorar dados oriundos de contextos educacionais (Romero & Ventura, 2010; Paiva et al., 2012).

Os tipos de estudos desta área são classificados, segundo Romero & Ventura (2010) em:

- Educação *off-line*: para análises em dados de desempenho do aluno, comportamento, currículo e etc., ou seja, coletados em ambientes de sala de aula;
- Aprendizado eletrônico (*e-learning*) e Sistema de Gestão da Aprendizagem ou LMS (do inglês, *Learning Management System*) para análise de dados armazenados em sistemas LMS no formato de log e bases de dados;
- Sistemas Tutores Inteligentes ou ITS (do inglês *Intelligent Tutoring System*) e Sistemas Hipermídias Adaptativos Educacionais ou AEHS (do inglês, *Adaptive Educational Hypermedia System*) os quais aplicados sobre dados de sistemas que se adaptam a cada estudante em particular, aos cursos oferecidos que estão em forma de log, aos modelos de usuários e etc.

2.3.1 TAREFAS E ALGORITMOS DE MINERAÇÃO DE DADOS EDUCACIONAIS

Em sua grande parte, as técnicas utilizadas na área de MDE são provindas da área de mineração de dados (BAKER et al., 2011).

Entretanto, na maioria das vezes há a necessidade de adaptá-las devido às particularidades existentes em ambientes educacionais e seus dados.

As técnicas estão apresentadas conforme sua categorização nas subáreas de MDE, seguindo-se o que consta na taxonomia proposta por Baker et al. (2011), a seguir:

- Predição

Na tarefa de predição, a meta é desenvolver modelos que façam inferência sobre aspectos específicos dos dados (variáveis preditivas) por meio da análise e associação dos diversos aspectos encontrados nos dados (variáveis preditoras).

Segundo Baker et al. (2011), há dois benefícios relacionados à utilização da predição em MDE. Primeiro, os métodos de predição podem ser utilizados para estudar quais aspectos de um modelo são importantes para predição.

Esta estratégia é frequentemente utilizada em pesquisas que tentam, de forma direta, predizer os benefícios educacionais de determinadas técnicas e ferramentas para um conjunto de estudantes, isso sem considerar os fatores intermediários, como apresentado em (ROMERO et al., 2008).

Segundo os métodos de predição auxiliam a predizer o valor das variáveis utilizadas em um modelo. O intuito de utilizar essa abordagem é verificar quais dados são mais importantes para o modelo pois analisar todos os dados de um grande banco de dados para gerar um modelo é inviável, do ponto de vista financeiro e de tempo (BAKER et al., 2011).

Em MDE, são utilizados mais frequentemente dois tipos de técnicas de predição: classificação e regressão. Na classificação a variável preditiva é binária ou categórica e na regressão a variável preditiva é contínua. Em ambos os casos, as variáveis preditoras podem ser categóricas ou contínuas.

- Classificação

Na classificação, os algoritmos mais utilizados são árvores de decisão e máquina de vetores de suporte.

- Regressão

Em relação à regressão, os algoritmos mais populares são regressão linear, redes neurais e máquinas de vetores de suporte para regressão.

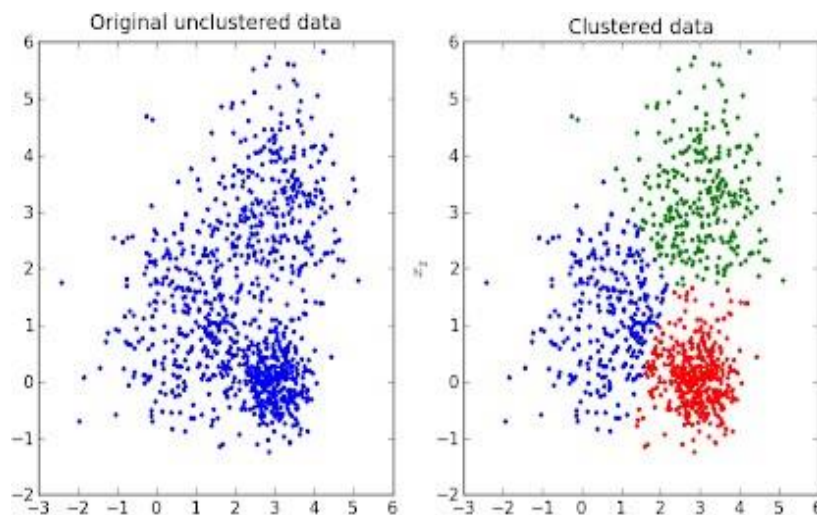
- Agrupamento

Em agrupamento, o objetivo é dividir o conjunto de dados em grupos, de forma que os objetos contidos nos dados fiquem agrupados naturalmente de acordo com a semelhança entre eles. Os algoritmos de agrupamento, são técnicas de aprendizado não- supervisionado, logo os grupos ou categorias, e até mesmo suas quantidades, não são conhecidos inicialmente.

- Algoritmo *K-Means*

O algoritmo *K-Means* é largamente utilizado para a tarefa de agrupamento. Em sua forma mais comum Lloyd, S. (1982), algumas vezes referida por Algoritmo de Lloyd. O número K de grupos que se deseja encontrar precisa ser informado de antemão. Tal como aponta Witten, I. H and Frank, E. (2005), este método de agrupamento é simples e efetivo como ilustrado na Figura 3.

Figura 3: Agrupamento de dados *K-means*



Fonte: Micreiros.com (2020).

- Mineração de Relações

Em mineração de relações, o objetivo é descobrir possíveis relações entre variáveis de um banco de dados com muitas variáveis, investigando quais variáveis estão mais fortemente relacionadas com uma determinada variável de interesse (ex.: a nota do aluno), ou investigando relações fortes entre quaisquer duas variáveis. Existem 4 tipos de mineração de relações:

1. Regras de associação;
2. Correlações;
3. Padrões sequenciais;
4. Causas.

- Destilação de dados para facilitar decisões humanas

Nesta área, o principal objetivo é apresentar os dados de forma mais legível e visual para facilitar a compreensão humana e assim apoiar decisões importantes baseadas nos dados. É uma área de interesse crescente dentro da mineração de dados educacional. Pois, a análise dos dados realizada por agente humanos, em sua maioria, só pode ser realizada se os dados forem apresentados de forma apropriada.

O principal método dessa área é o de visualização da informação. No entanto, a maioria dos métodos de visualização, normalmente, usados dentro da mineração de dados educacional são frequentemente diferentes do que aqueles mais frequentemente usados em problemas de

visualização da informação (HERSHKOVITZ e NACHMIAS., 2008; KAY et al., 2006).

Alguns exemplos dessas particularidades são destacados em Baker 2010a: Os dados são organizados em termos da estrutura do material de aprendizagem (habilidades, problemas, unidades, aulas) e da estrutura de contexto de aprendizagem (alunos, professores, pares de colaboração, classes e escolas).

A destilação dos dados para facilitar decisões humanas tem dois propósitos principais (Baker 2010b):

a) Identificação: os dados são apresentados de forma que humanos possam identificar os padrões mais facilmente, que são difíceis de expressar formalmente;

b) Classificação: a destilação de dados pode ser usada também para apoiar a modelos de predição. Neste caso, parte dos dados são exibidos para serem rotulados por humanos. Esses rótulos são utilizados como base para a construção desses modelos.

- Descobertas com modelos

Em descoberta com modelos, parte-se de um modelo gerado por um método de predição, tal como classificação, ou por um método de agrupamento, ou ainda manualmente, por meio de engenharia de conhecimento. Em seguida, esse modelo é utilizado como componente, ou ponto de partida, em outra análise com técnicas de predição ou mineração de relações.

- *Learning Analytics*

Learning analytics é utilizado para analisar a aprendizagem por intermédio da tecnologia.

O acesso e uso da tecnologia tem crescido em diversos ambientes devido à facilidade de interação entre plataformas digitais, devido a isso o *learning analytics* ou LA é usada para coletar e analisar os dados em estudos que exploram e mostram

as dificuldades de aprendizagem que os alunos possuem.

Quando surgiu, a LA tinha por objetivo lidar somente com estudantes da área de Ensino a Distância. No entanto, atualmente, é utilizada também em instituições que ofertam educação infantil, ensino fundamental e médio e também nas instituições de cursos superiores presenciais. Isso se deve devido a facilidade de acesso à tecnologia.

A partir dos resultados encontrados com o LA, os docentes podem criar estratégias de forma a melhorar a aprendizagem do aluno, sendo assim, após aplicadas as estratégias, pode-se reutilizar LA como forma de medir o impacto das mudanças nos alunos e de uma forma minimizada, verificar o mesmo em todo o ambiente escolar de uma forma mais ampla.

Para analisar o desempenho educacional dos alunos por meio de LA, é necessário ter os dados disponíveis para se trabalhar e esses dados são obtidos quando os alunos acessam vídeo-aulas, simulados ou até mesmo fazendo atividades e provas online, por exemplo.

Existem plataformas online de LA que coletam e disponibilizam esses dados, em *dashboards*, por exemplo, para facilitar o entendimento que as escolas terão ao analisar estas informações.

Um dos benefícios de utilizar LA nas escolas é que os alunos terão ciência de quais são os pontos fortes e a fracos em todas as disciplinas, facilitando assim sua aprendizagem.

Outro benefício é que a instituição poderá personalizar aulas e métodos para cada aluno ou turma e gerar estratégias educacionais que auxiliam na melhoria do desempenho dos estudantes.

Um exemplo prático da LA seria uma análise de uma turma do 3^a ano do ensino médio, pois estão se preparando para a prova do Enem. Eles passariam o ano letivo treinando redações nos moldes utilizados no exame do Enem.

Com a ferramenta de LA, é possível identificar quais notas cada aluno atingiria

nas 5 competências da redação do Enem. De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) são elas:

1. Domínio da escrita formal da língua portuguesa
2. Compreender o tema e não fugir do que é proposto
3. Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista
4. Conhecimento dos mecanismos linguísticos necessários para a construção da argumentação
5. Respeito aos direitos humanos

De acordo com os resultados encontrados após análise feita com LA, é possível identificar a média da turma e de cada aluno.

- Aprendizagem de Máquina (AM)

É um tipo de inteligência artificial que permite que seja tomada decisões com ajuda de algoritmos que reconhecem padrões na análise de dados. Com isso, pode ser capaz de fazer previsões ou tomadas de decisões e automatizar processos com mínimo ou se possível nenhuma intervenção humana.

Um exemplo de aprendizagem de máquina é o supercomputador *Watson* da IBM - *International Business Machines*, que foi utilizado nas olimpíadas de 2016 pelo jornal *Washington Post*, onde eles postaram notícias sobre a olimpíadas utilizando apenas análise de dados na internet sem intervenção humana.

Outro exemplo para deixar mais claro como aprendizagem de máquina é bastante utilizada, são os carros autônomos, onde através de algoritmos e técnica de AM, hoje é possível viagens de carro, seja a passeio ou a trabalho, sem intervenção humana.

- Aprendizagem profunda (AP)

Do inglês *Deep Learning* (DL) é um subgrupo de técnicas que faz parte da Aprendizagem de Máquina que utiliza algoritmos desde o básico ao mais avançado para simular o comportamento do cérebro

humano como reconhecimento de imagens, fala e linguagem natural, no intuito de analisar dados e produzir resultados mais rápidos e precisos, identificando padrões e tomando decisões com pouca intervenção humana possível.

Um exemplo foi a utilização do supercomputador *Watson* que é uma plataforma da IBM que utiliza aprendizagem de máquina, e através dela, pesquisadores japoneses utilizaram-se da AP para análise de documentos médicos, que conseguiu em 10 minutos diagnosticar uma paciente que tinha uma doença tão rara que não foi possível identificar com intervenção humana.

- *Naive Bayes*

É um classificador multinominal e um dos modelos mais utilizados no AM. Tendo como base a suposição de que as variáveis de um problema são independentes, o modelo *Naive Bayes* trabalha com uma classificação baseada na probabilidade de observações, caracterizando-as em classes pré-definidas.

É mais utilizado em análises de crédito, diagnósticos médicos ou busca por falhas em sistemas mecânicos, classificação textual tais como: documentos legais de políticas públicas ou estratégias de mercado até e-mails.

Um exemplo de uso é separar e-mails de spam, baseado em vários termos pré-definidos o classificador consegue fazer a separação destes e-mails.

- *K Nearest Neighbor* – *KNN* ou K-Vizinhos mais próximos

Faz parte dos algoritmos utilizados em AM, sendo o *KNN* o mais simples deles em termos de complexidade computacional, e é utilizado para tarefas de classificação e regressão, sendo mais usado para tarefas de classificação.

É uma das técnicas mais simples de previsão/classificação, os resultados dependem de como é feito o

dimensionamento das características, a medida da semelhança e quão grande K é definido, sendo K representante da quantidade de classes próximas que deverão ser escolhidas. Diferente do *Naive Bayes*, o *KNN*, trabalha com preditores em sua forma numérica.

- Máquina de vetores de suporte (*Support Vector Machine - SVM*)

É um algoritmo de AM supervisionado, utilizado para trabalhar com um conjunto de dados menores e principalmente na criação de modelos.

O *SVM* é um classificador linear binário não probabilístico e pode ser usado para análises de classificação ou regressão, sendo seu maior foco no treinamento e classificação de um conjunto de dados.

- *Randon Forest*

É um modelo de árvore de decisão que quando utilizado, decompõe todos os dados em grupo, formando uma árvore de decisão de modo bem aleatório, analisando diferentes variáveis que possui de cada grupo, para determinar como dividir os dados em nós, onde verifica em qual caminho uma condição irá passar, sempre levando ao próximo nó até a finalização da árvore.

- Classificador J48

Algoritmo classificador que permite a criação de árvores de decisão. Gera árvores em que cada nó avalia a efetividade ou importância de cada atributo individualmente.

Elas são geradas do topo para base através da escolha de atributos mais apropriados para cada situação, depois dividida em subgrupos até que cada parte

permaneça a uma única classe. É muito utilizado para comparar percentagem de acerto com outros algoritmos.

- *Multi-Layer Perceptron (MLP)*

Segundo Hinton, G. E (1989) é um algoritmo conectado por um conjunto de nós ou vértices, que se interconectam por unidades de processamento que imitam os neurônios humanos. O objetivo desse algoritmo é minimizar o erro entre os valores previsto pela *MLP* e os valores verdadeiros.

- *Hybrid Educational Data Mining Model (HEDM)*

É um modelo da MDE que utiliza a combinação do classificador *Naive Bayes* e a técnica do J48 para previsão de risco e fatores de qualidade e performance.

3 PROCESSOS METODOLÓGICOS/MATERIAIS E MÉTODOS

Neste artigo é feita uma análise dos resultados obtidos através do uso da MDE na busca por soluções em diversas áreas de estudo e também quais e como as técnicas são utilizadas neste processo de análise dos dados.

Foram analisados neste artigo trabalhos em português e publicados no Brasil, trabalhos em inglês e publicados na em outros países, autores brasileiros e publicados em inglês e autores estrangeiros traduzidos para publicação no Brasil. Na Tabela 1 a seguir, estão listados todos os artigos trabalhados.

Tabela 1 - Locais e idiomas dos artigos

Ano	Autor	Local de Publicação	Idioma da Publicação
2021	Vanessa Faria de Souza, Tony Carlos Bignardi dos Santos	Brasil	Português (Brasileiro)
2020	Kelly Joany de Oliveira Santos	Brasil	Português (Brasileiro)
2020	V.Ganesh Karthikeyan, P.Thangaras e S.Karthik	Índia	Inglês (Americano)
2019	Inamar Pererira de Brito, Humberto Rabelo, Ângela Maria Chuvas Naschold, Almir Miranda Ferreira, Aquiles Medeiros Filgueira Burlamaqui, Daniel Silva de Souza Rabelo, Ricardo Alexsandro de Medeiros Valentim	Brasil	Português (Brasileiro)
2017	Roger, Flavia, Juliana, Jaques e Eldipio	Brasil	Português (Brasileiro)
2017	Sarardeep Roy, Anchal Garg	Índia	Inglês (Americano)
2016	Stella Oggioni da Fonseca e Anderson Amendoeira Namen	Brasil	Português (Brasileiro)
2015	Marcio Aurelo dos Santos Alencar, Eulanda Miranda dos Santos e José Francisco de Magalhães Netto	Brasil	Inglês (Canadense)
2014	Sando J. Rigo, Wagner Vambuzzi e Jorge L. V. Barbosa	Brasil	Português (Brasileiro)

Fonte – Autores.

Este artigo não é apenas uma coleção de dados contida em artigos, mas, ao contrário, concentra-se na reflexão inovadora e analista de certos textos e dos conceitos levantados neles, no qual foi realizada uma consulta a livros, dissertações

e por artigos científicos selecionados através de busca nas seguintes bases de dados (livros, sites de banco de dados, etc....). Neste trabalho foram utilizadas as bases *Scielo* (Tabela 2), *Google Acadêmico* (Tabela 3), *ResearchGate* (Tabela 4).

Tabela 2: Critérios de pesquisa na base do Google Acadêmico

Data de Busca	09/01/2022
---------------	------------

Site de Busca	Google Acadêmico
URL	scholar.google.com.br/
Período	2010 - 2022
Chave de Busca 1	("educational data mining") and ("analysis")
Chave de Busca 2	("mineracao de dados educacionais") and ("analise")
Critério de Escolha	Quantidade de citações divulgados pela fonte de busca, informações disponibilizadas pelo texto sobre as ferramentas utilizadas.

Fonte – Autores.

Tabela 3: Critérios de pesquisa na base Scielo

Data de Busca	15/02/2022
Site de Busca	Base Scielo
URL	https://search.scielo.org/
Período	2010 - 2022
Chave de Busca 1	("educational data mining") and ("analysis")
Chave de Busca 2	("mineracao de dados educacionais") and ("analise")
Critério de Escolha	Quantidade de citações divulgados pela fonte de busca, informações disponibilizadas pelo texto sobre as ferramentas utilizadas.

Fonte: Autores.

Tabela 4: Critérios de pesquisa na base do researchgate

Data de Busca	18/02/2022
Site de Busca	Research Gate
URL	https://www.researchgate.net/
Período	2010 - 2022
Chave de Busca 1	("educational data mining") and ("analysis")
Chave de Busca 2	("mineracao de dados educacionais") and ("analise")
Critério de Escolha	Quantidade de citações divulgados pela fonte de busca

Fonte: Autores.

Buscando pelos termos “data minig”, no Google Acadêmico foram encontrados 19.500 artigos científicos publicados entre 2014 e 2022, destes, apenas 254 artigos foram publicados em português, não sendo possível segregar por autoria ou coautoria de pesquisadores Brasileiros. A busca no Scielo retornou 854 artigos científicos onde foi possível fazer uma melhor seleção dos filtros, destes, apenas 220 artigos foram publicados com autoria ou coautoria de pesquisadores brasileiros. A busca no Research Gate resultou em um total de 11.278 artigos, não sendo possível segregar por região, autoria ou coautoria de pesquisadores brasileiros.

Neste trabalho foram incluídos os artigos na língua portuguesa e na língua inglesa pertinentes ao tema e que se enquadram nos critérios de busca definidos nas tabelas acima.

Os artigos que foram excluídos, não atendem a base de critérios citados nas tabelas acima ou não estavam em um idioma de fácil entendimento.

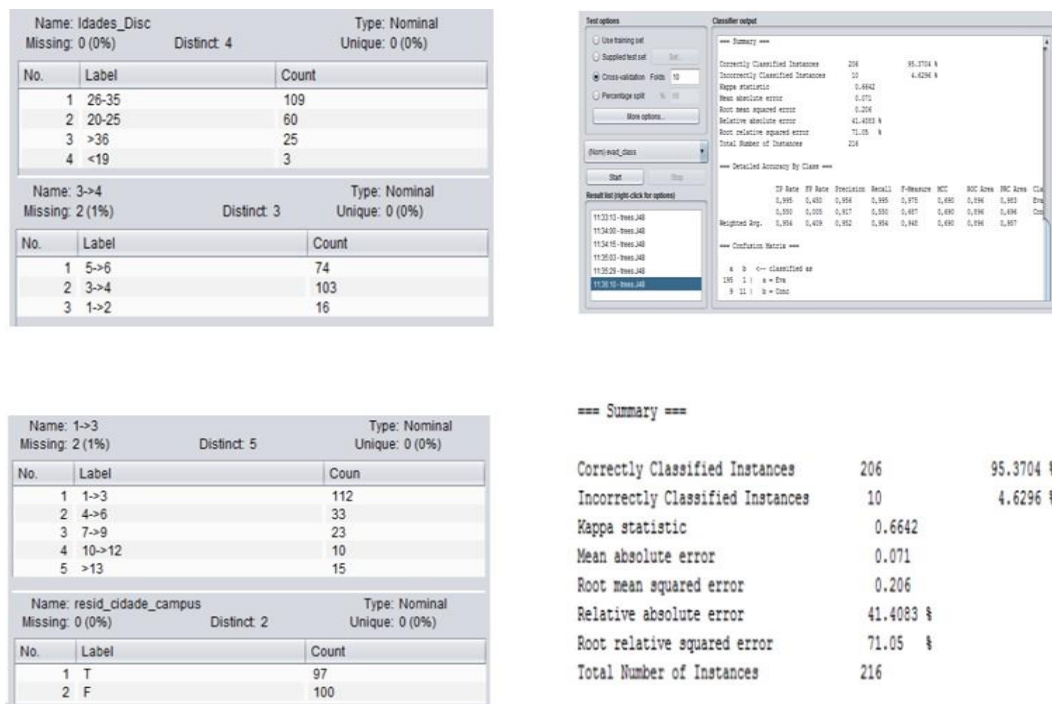
4 DISCUSSÃO

BRITO et al. (2019) propõem uma análise de dados sobre os históricos educacionais. Foram utilizados dados acadêmicos de um curso de Ensino Superior da Universidade Federal do Rio Grande do Sul (UFRN).

A biblioteca utilizada para análise de dados foi o *WEKA* para análise foram utilizados os algoritmos *K-means* e o *J48*.

Os resultados encontrados mostram uma precisão dos classificadores de aproximadamente 90 a 95%, esses valores foram determinados a partir da média geral de todos os experimentos feitos conforme ilustrado na Figura 4. Os autores deixaram evidente a importância desta pesquisa, que teve como objetivo combater a evasão através da identificação de perfis mais propícios a esse fenômeno, foi possível notar a grande importância o uso da mineração de dados neste contexto.

Figura 4: Resultado do processo de Análise utilizando J48



Fonte: BRITO et al. (2019).

Rigo et al. (2014) neste artigo podemos ver a utilização da mineração de dados alinhada ao *learning analytics*. Sendo adotadas na mitigação da evasão escolar, providenciando informações a respeito dos alunos que são utilizadas em intervenções pedagógicas.

Através da análise dos trabalhos sobre a implementação destas aplicações o artigo traz a descrição das técnicas empregadas e quão importante estes trabalhos demonstram ser.

Apresenta ainda os principais motivos da evasão escolar e descreve a utilização de um sistema para Mineração de Dados Educacionais (MDE) e *Learning Analytics* (LA).

Foram testes realizados durante 18 meses em cursos de graduação na modalidade de Educação a Distância.

Apresenta o estudo de aspectos da evasão escolar e descreve a aplicação e os resultados de uso de sistemas de MDE e LA em uma experiência prática de aplicação de recursos tecnológicos em busca do tratamento do problema de evasão escolar. Através da implementação do sistema pode-se chegar a um modelo conceitual que demonstra quais critérios foram utilizados para a análise dos perfis dos alunos.

Quanto a Mineração de Dados, esta pode ser entendida como uma das etapas do processo de Descoberta de Conhecimento em Base de Dados (DCBD). O processo de DCBD pode ser definido como sendo o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis presentes nos dados sendo analisados e tratados. O objetivo principal da etapa de Mineração de Dados constitui-se em extrair conhecimento implícito em bases de dados, através de diversas tarefas e com objetivos, em geral, de descrever ou então de prever a ocorrência de padrões úteis para atividades associadas.

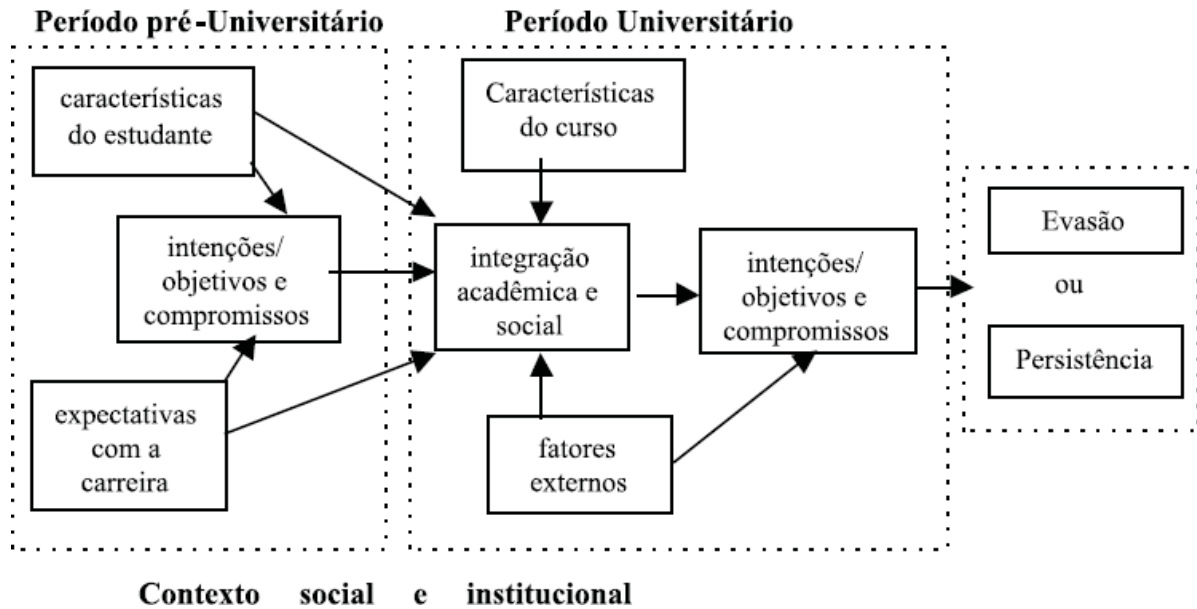
Souza et al. (2021) apresenta um estudo que tem como objetivo demonstra a aplicação do processo de MDE para realizar a previsão do desempenho de alunos, em um conjunto de dados públicos, e comparar as técnicas de Aprendizagem de Máquina (AM) e Aprendizagem profunda (AP). Baseada em comparar técnicas de AM e AP para indicar quais os principais atributos preditores para o desempenho dos alunos. O estudo foi composto basicamente por 3 fases:

- 1) Coleta de dados;
- 2) Extração de recursos e limpeza de dados (Pré-Processamento e Transporte)
- 3) Processamento analítico e algoritmos.

Os resultados deste estudo foram sistematizados de acordo com as 3 questões de pesquisa, com relação à previsão de desempenho as técnicas de MDE aplicadas foram adequadas, em que os resultados alcançados são os seguintes: *Naïve Bayes* com uma precisão de 66%; *Árvore de Decisão* com 87%; *Random Forest* com 83%; *Support Vector Machine* com 82%; e *RNAM* com 94% de precisão.

Ademais foi possível identificar quais são os principais atributos que dão melhor suporte na precisão de desempenho de alunos, além disso os resultados das avaliações dos algoritmos exibidos, podem do suporte na escolha de métodos mais eficazes para aplicação em conjuntos de dados educacionais, levando em consideração a Aprendizagem Profunda, que sem dúvida é uma técnica bastante relevante no contexto da Ciência de Dados.

Figura 5: Modelo da relação aprendizagem/permanência



Fonte: SOUZA et al. (2021).

Machado et al. (2017) teve como objetivo conhecer a evolução dos temas no mundo científico, incorporando suas correlações de forma a agregar o conhecimento da mineração de dados na evasão escolar universitária e buscar possíveis lacunas, através de um estudo bibliométrico nas bases de pesquisa *Scopus Web Of Science, ScienceDirect e Scielo*.

A bibliometria aborda a necessidade de investimento em pesquisas específicas no tema, trazendo a lacuna de conhecimento quando abordados trabalhos correlatos, o que apresenta abranger conhecimentos na área contribuindo cientificamente para o desenvolvimento dos temas deste trabalho.

Contudo, se buscou a identificação de trabalhos que abordam o problema da evasão com a utilização das técnicas de Mineração de Dados e a apresentação de uma visão da produção científica atual.

Visando comparar e analisar os resultados obtidos, a pesquisa foi

estruturada de modo que a busca pudesse ser replicada nas quatro bases de dados.

Para tanto, as palavras foram pesquisadas dentro dos campos “Título do Artigo” (*Article Title*), “Resumo” (*Abstract*) e “Palavras-chave” (*Keywords*).

Em todas as etapas da pesquisa, foram consultados apenas artigos científicos publicados nos últimos 10 anos, além dos artigos publicados até março do ano de 2015. A pesquisa permitiu encontrar 19 artigos científicos abordando o problema da evasão escolar com a utilização de técnicas de mineração de dados.

Com base nestes, é possível concluir que a evasão escolar é um tema global, uma vez que foram encontrados trabalhos de várias partes do mundo.

Tendo em vista que 36,84% dos trabalhos encontrados foram publicados entre 2013 e 2014.

Com base no presente estudo bibliométrico, dada a importância do problema da evasão frente ao crescente

interesse pelo tema, bem como a carência por trabalhos correlatos, tanto nacionais quanto internacionais, conclui-se que o presente trabalho traz contribuições científicas para embasar pesquisas que visem contribuir para a diminuição do problema evasão escolar universitária através da mineração de dados.

Karthikeyan et al. (2020) O estudo conduzido por V. Ganesh Karthikeyan1, P. Thangaraj, S. Karthik propôs uma aplicação e métodos para desenvolver o modelo de mineração de dados educacional híbrido (MDEH) para uma avaliação eficiente e precisa do desempenho dos alunos.

Muitas instituições de ensino utilizam a mineração de dados para manter os registros dos alunos, especificamente os desempenhos acadêmicos, que são mais significativos.

Os desempenhos acadêmicos dos alunos devem ser analisados para melhorar seus resultados e também os resultados gerais das instituições.

Com isso, este artigo desenvolve uma nova abordagem chamada modelo de mineração de dados educacional híbrido (MDEH) para analisar o desempenho do aluno para melhorar efetivamente a qualidade educacional dos alunos.

O modelo proposto avalia o desempenho dos alunos com base em fatores distintivos que proporcionam resultados adequados. Além disso, o modelo combina as eficiências da técnica de classificação *Naive Bayes* e do Classificador J48 para derivar os resultados e categorizar o desempenho dos alunos de maneira precisa.

O modelo é avaliado com o conjunto de dados de educação de referência que está disponível online no ambiente *WEKA*. Os resultados mostram que o modelo proposto supera os resultados de trabalhos existentes na avaliação do desempenho dos alunos em MDE.

Com base na seleção eficaz de atributos do conjunto de dados obtido, a operação de classificação é realizada e os alunos são categorizados em classes

especificadas. Ao combinar dois modelos de classificações, as listas de alunos são classificadas em classes como, *Excellent_List*, *Good_List*, *Average_List* e *Low_List*. O processo de avaliação é realizado em ambiente *WEKA* com os conjuntos de dados obtidos que são disponíveis online, com base nas métricas de avaliação, como taxa de precisão, *recall*, taxa de erro e medida F.

É mostrado nos resultados que o trabalho proposto produz maior taxas de exatidão e precisão com taxas de erro reduzidas.

Este faz com que os acadêmicos obtenham melhores resultados com base na análise de desempenho e tomar decisões eficientes para melhorar a qualidade do aluno, juntamente com a instituições.

Fonseca et al. (2014), de acordo com o artigo, embora existam estudos que abordam o conjunto de dados coletados pelo INEP, ainda é analisado se esse rico banco de dados tem sido utilizado pela comunidade para a extração de informações correlacionadas com o desempenho dos estudantes e se tem feito uma análise profunda desses dados.

Nesse contexto, o presente estudo explana a possibilidade de realização de análises profundas dos dados educacionais por meio de metodologias avançadas baseadas em inferência estatísticas. Em seguida são apresentadas as etapas do processo denominada de *KDD (Knowledge Discovery in Databases)* depois aplica-se o processo em base de dados do INEP provenientes da avaliação da educação básica, objetivando avaliar algumas questões referente aos docentes quem podem ou não influenciar no seu desempenho, em particular a matemática.

Por fim, são apresentados resultados provenientes da implementação do algoritmo *Naves Bayes* disponibilizada no software *WEKA*.

O Estudo releva que apesar do INEP organizar e disponibilizar uma ampla base de dados relacionada a educação, muitas

vezes, não são exploradas em todo seu potencial, devido à dificuldade inerente ao grande volume de dados envolvido.

Diante desse contexto, o desafio aqui apresentado foi fazer uso efetivo desses dados, analisando-os por meio de metodologias que propiciassem a extração de informações que oferecessem rotas para a definição de ações voltadas à melhoria dos resultados do processo de ensino-aprendizagem de Matemática na Educação Básica.

A alternativa apresentada foi a Mineração de Dados, uma tecnologia que se alicerça em conceitos estatísticos e de inteligência computacional. Diante desse campo abrangente de aplicação alternativa apresentada foi a utilização Mineração de Dados, uma tecnologia que se alicerça em conceitos estatísticos e de inteligência computacional.

Assim, seria possível comparar os fatores que vêm afetando o desempenho dos discentes ao longo dos anos e que possa servir de estímulo para o uso mais intenso de métodos de mineração de dados em bases de dados do INEP, com intuito de nortear melhorias no sistema educacional brasileiro.

Roy et al. (2017) apresenta um método para estudar a performance acadêmica do aluno implementando vários dados da mineração usando técnicas de algoritmos de classificação, assim, encontrando o melhor algoritmo de classificação, aquele que ajuda após uma avaliação da nota final de um estudante e encontra o mais apto para identificação de possíveis resultados, para que intervenções apropriadas sejam feitas.

Nesta pesquisa foram coletados dados de registros acadêmicos, atual e anterior, de um renomado curso de ensino superior relacionado a um conjunto de estudantes.

Os dados usados foram referentes a alunos de 15 a 16 anos que estão em fase de nota final e querem cursar o ensino superior.

Estes dados foram filtrados, limpos e processados para serem usados no modelo de técnicas de classificação para encontrar dentre os estudantes da base de dados, quem está em estado de alto risco ou em estado de baixo risco de perder a performance acadêmica, permitindo assim, aos educadores providenciar o aconselhamento adequado para aquele aluno em tempo hábil.

Nesse contexto, os autores deixam claro que, prever o desempenho acadêmico do aluno é uma grande preocupação em todos institutos de ensino. Isso ajuda a identificar as habilidades dos alunos, seus interesses e fraquezas.

O desempenho do aluno pode ser influenciado por diferentes tipos de atributos, como por exemplo: social, demográfico ou relacionado à escola.

O artigo utilizou de atributos que incluem notas dos alunos, dados demográficos, características sociais.

Usando o algoritmo *Naïve Bayes*, árvore de decisão J48 e *Multi-Layer Perceptron (MLP)* para prever o desempenho dos alunos, ficou evidente que sem as notas anteriores fica muito difícil prever as notas finais mais pode se analisar outros dados e descobrir quais outros atributos são importantes na previsão das notas finais.

Os principais fatores que influenciam na nota são:

- 1) O alto consumo de álcool no fim de semana;
- 2) A educação dos pais;
- 3) Um relacionamento romântico
- 4) O nível de educação de mãe é outro fator importante na determinação das boas pontuações.

Alencar et al. (2015) mostra que a quantidade de instituições de ensino que trabalham com cursos a distância é crescente. Como consequência, estudos têm mostrado que as taxas de evasão de alunos nesse tipo de sistema educacional também aumentaram.

Embora os Ambientes Virtuais de Aprendizagem (AVA) registrem todas as interações dos alunos ao longo do curso, as informações fornecidas pelo AVA não são suficientes para prever e evitar altas taxas de evasão de alunos.

Nesse contexto, o objetivo deste artigo foi apresentar um levantamento sobre abordagens e técnicas de mineração de dados educacionais a fim de apontar soluções que possam ser empregadas para prever e prevenir altas taxas de evasão de alunos. Para tal, foram utilizadas a base de dados da CETAM EAD, usado o sistema acadêmico integrado para Moodle (*Modular Object-Oriented Dynamic Learning Environment* ou Ambiente Virtual De Aprendizagem).

O objetivo deste estudo foi analisar a interações gravadas no CETAM EAD para cursos técnicos, usando técnicas de mineração de dados educacionais que sejam capazes de mostrar quais alunos tem uma chance de evasão escolar.

Santos et al. (2020) o principal propósito foi realizar através de algoritmos de mineração de dados educacionais, um estudo de caso que visa entender as razões que levam os estudantes a evadirem das unidades de ensino.

Para o contexto foi selecionado o âmbito da Universidade Federal de Sergipe cujo os dados dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia da Computação foram considerados para a aplicação específica do problema.

A abordagem para avaliar o desempenho de alunos em disciplinas busca apresentar uma análise aprofundada do contexto evolutivo do aluno por disciplina realizada através das técnicas de Mineração de Dados Educacionais.

É proposto uma abordagem computacional orientada a 5 perguntas que foram formuladas para refletir os principais problemas da instituição analisada.

Neste trabalho, foram escolhidos os principais classificadores encontrados nos

trabalhos de MDE: árvore de decisão, *Naive Bayes*, vizinho mais próximo - *KNN*, máquina de vetores de suporte (*SVM*) e a floresta aleatória - *Random Forest*. A escolha do algoritmo é muito importante para qualquer trabalho na área de mineração de dados (ATHANI et al., 2017).

Nesta pesquisa, explorou-se a transformação de dados, que estavam a priori armazenados em CSV's acadêmicos, em informações potencialmente úteis para apoiar a mitigação da evasão através da identificação de perfis evasivos com o auxílio de técnicas de mineração de dados.

5 RESULTADOS

Com base no estudo dos artigos encontrados identificamos os principais métodos de mineração de dados educacionais explorados e mais utilizados para tratar de informações referentes ao nível de aprendizagem do aluno, evasão escolar, ensino-aprendizagem dos docentes e ainda nortear melhorias no contexto da educação no Brasil.

A Tabela 5 a seguir lista as diferentes características de cada trabalho analisado.

Tabela 5: Visão geral dos artigos estudados

Ano	Autor	Ferramenta KDD	Técnica de Mineração ou Atributos utilizados
2021	Vanessa Faria de Souza, Tony Carlos Bignardi dos Santos	<i>TensorFlow</i>	Preditiva, Classificação
2020	Kelly Joany de Oliveira Santos	<i>KNIME</i>	Preditiva, Classificação
2020	V.Ganesh Karthikeyan, P.Thangaras e S.Karthik	<i>WEKA</i>	Classificação
2019	Inamar Pererira de Brito, Humberto Rabelo, Ângela Maria Chuvas Naschold, Almir Miranda Ferreira, Aquiles Medeiros Filgueira Burlamaqui, Daniel Silva de Souza Rabelo, Ricardo Alexsandro de Medeiros Valentim	<i>WEKA</i>	Preditiva, Classificação
2017	Roger, Flavia, Juliana, Jaques e Eldipio	Não especificada	Agrupar
2017	Sarardeep Roy, Anchal Garg	<i>WEKA</i>	Preditiva, Classificação
2016	Stella Oggioni da Fonseca e Anderson Amendoeira Namen	<i>PostgreSQL</i>	Classificação
2015	Marcio Aurelo dos Santos Alencar, Eulanda Miranda dos Santos e José Franscisco de Magalhães Netto	<i>WEKA</i>	Classificação
2014	Sando J. Rigo, Wagner Vambuzzi e Jorge L. V. Barbosa	Ferramenta própria desenvolvida em ASP.NET	<i>Learning Analytics (LA)</i>

Fonte: Autores.

6 CONCLUSÕES

Com a análise e interpretação dessas publicações, em destaque dos últimos 12 anos, foi possível compreender quais eram as principais técnicas, os dados utilizados para formação de bancos/bases e os resultados que foram alcançados, enfim, entender o motivo pelo qual a MDE tem destaque ao ser utilizada neste contexto de solução de problemas educacionais.

Além disso, podemos entender como ocorreu o processo de sua evolução, que teve início devido a dois fatores: 1) A adoção de grandes bases de dados na educação, principalmente com o surgimento de cursos e-learning ou on-line, e também como ferramentas de ensino como o *MOODLE*; 2) O avanço das tecnologias, que são indispensáveis para aplicação das técnicas de MDE.

Tais avanços, proporcionaram melhorias nas técnicas/ferramentas de MD já existentes, simplificando assim as tarefas dos pesquisadores e aperfeiçoando os resultados obtidos.

Dessa forma, tais técnicas/ferramentas puderam ser aplicadas e testadas ao grande volume de dados educacionais disponíveis, mostrando assim que a MDE é um importante conjunto de técnicas e estratégia de pesquisa sobre dados educacionais.

A partir dos resultados obtidos e das análises aqui apresentadas podemos perceber que a área de mineração de dados educacionais demonstra um crescente número de usuários e de plataformas que a estão utilizando como meio de obter respostas em uma análise de dados mais precisa, contribuindo assim para que a ferramenta estudada seja mais conhecida. A cada ano vem crescendo o número de artigos publicados sobre o tema.

Percebe-se que, com o passar do tempo, métodos distintos foram criados com base nas técnicas pesquisadas e cada vez mais o número de pesquisas publicadas na área vem aumentando, o que mostra a amplitude desta área de pesquisa.

Os artigos aqui estudados mostram um uso em comum das ferramentas MDE como o J48 e *Naive Bayes*; *KDD e learning analytics* e o software *WEKA*. Apesar de usarem estas ferramentas, a MD tem em seu catálogo diversas opções de ferramentas disponíveis.

É importante considerar como trabalhos futuros, a expansão desta análise da MDE, seja para ampliar a quantidade de artigos investigados ou para incluir um maior número de técnicas de mineração de dados educacionais, selecionar um

contexto específico e analisar quais os propósitos da MDE, identificar tendências e tópicos de estudos mais relevantes, encontrar quais as técnicas e algoritmos de MDE mais utilizados e ainda as principais oportunidades de pesquisa e os desafios mais relatados, formando um estudo amplo que apoie pesquisadores da área.

AGRADECIMENTOS

Gostaríamos de agradecer primeiramente a Deus, pela oportunidade de chegar no fim deste curso, estando em um período de pandemia e também a nossas famílias e amigos pelo apoio durante esta jornada.

Ao corpo docente do Instituto Federal Do Espírito Santo – Campus Colatina, pelo empenho e dedicação em encontrar soluções para o prosseguimento do curso e que em todo processo não nos deixaram desanimar. Ao nosso orientador Prof. Vanderson Silva, que nas reuniões sempre procurava nos tranquilizar e orientar.

A todos os colegas de curso que em algum momento nos ajudaram a superar os desafios das disciplinas e seguir para a próxima etapa. Nosso muito obrigado a todos!

REFERÊNCIAS

ABERNETHY - 2010 - **Mineração de dados com o WEKA, Parte 2** Classific2

<<https://pt.scribd.com/document/530009147/Abernethy-2010-Mineracao-de-dados-com-o-WEKA-Parte-2-Classific2>>.

Acessoem: 20 out. 2021.

ALENCAR, MÁRCIO AURÉLIO DOS SANTOS; SANTOS, EULANDA MIRANDA; NETTO, JOSÉ FRANCISCO

DE MAGALHÃES. **Identifying Students With Evasion Risk Using Data Mining. 2015.** 6 f. Trabalho de Conclusão de Curso (Pós Graduação) - Universidade Federal do Amazonas, [S. l.], 2015.

- ATHANI, S. S. et al. **Student performance predictor using multiclass support vector classification algorithm**. In: IEEE. 2017 International Conference on Signal Processing and Communication (ICSPC). [S.l.], 2017.
- BAKER, R. S., CORBETT, A. T., ROLL, I., AND KOEDINGER, K. R. (2008). **Developing a generalizable detector of when students game the system. User Modeling and User-Adapted Interaction**, 18(3):287–314.
- Baker, R. (2010a). **Data Mining for Education**. In McGaw, B., Peterson, P., Baker, E. (Eds.) **International Encyclopedia of Education (3rd edition)**, Elsevier, Oxford, UK.
- BAKER, R. (2010B). **Mining Data for Student Models**. In NKMABOU, R., MIZOGUCHI, R., & BOURDEAU, J. (Eds.) **Advances in Intelligent Tutoring Systems**, pp. 323-338. Secaucus, NJ: Springer.
- BAKER, R. (2010a). **Data Mining for Education**. In McGaw, B., PETERSON, P., BAKER, E. (Eds.) **International Encyclopedia of Education (3rd edition)**, Elsevier, Oxford, UK.
- BAKER, R.S.J.D., I. S. D. C. A. (2011). **Mineração de dados educacionais: Oportunidades para o brasil**. Revista Brasileira de Informática na Educação, 19(2).
- BAKER, R. AND YACEF, K. (2009). **The state of educational data mining in 2009: A review and future visions**. Journal of Educational Data Mining, 1(1):3–17.
- BAKER, R. S. J.; ISOTANI, S.; CARVALHO, A. **Mineração de dados educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação, 2011, v. 12, n. 2, p. 3 – 13.
- BRITO, INAMAR PERERIA et al. **Uso da Mineração de Dados Educacionais para a classificação e identificação de perfis de Evasão de graduandos em Sistemas de Informações da UFRN**. 2019. 11 p. Anais dos Workshops do VII Congresso Brasileiro de Informática na Educação (Pós Graduação) - Universidade Federal do Rio Grande do Norte, [S. l.], 2019. DOI 10.5753/cbie.wcbie.2019.159. Disponível em:<https://www.researchgate.net/publication/337528851_Uso_de_Mineraçao_de_Dados_Educacionais_para_a_classificacao_e_identificacao_de_perfis_de_Evasao_de_graduandos_em_Sistemas_de_Informacao.html>. Acesso em: 4 jan. 2022.
- CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; AND ZANASI, A. (1998). **Discovering data mining: from concept to implementation**. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.
- CALIL, L. A de A. et al. **Mineração de Dados e Pós Processamento em Padrões Descobertos**. UEPG Exact Earth Sci, Ponta Grossa, v 14, n 3. P. 207-215, 2008.
- CASTANHEIRA, L. G. **Aplicação de Técnicas de Mineração de Dados em Problemas de Classificadores de Padrões**. 2008. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais.
- COSTA, EVANDRO.; BAKER, RYAN S.J.D.; AMORIM, LUCAS; MAGALHÃES, JONATHAS.; MARINHO, TARSIS. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. Jornada de Atualização em Informática na Educação – JAIE 2012. Disponível em:<<https://www.br-ie.org/pub/index.php/pie/article/view/2341>>. Acesso em: 20 Out. 2021.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.;
- UTHURUSAMY, R., 1996, **Advances in knowledge discovery and data mining. Edition 1st edition**, March Publisher MIT Press, 1996. Acesso em: 20 dez. 2021.

- FAYYAD, U. et al. **From Data Mining to Knowledge Discovery in Databases**. AI Magazine, California, v. 17, n. 3. P. 37-54, 1996a. The KDD Process for Extracting Useful Knowledge from Volumes of Data. In: Communications of the ACM, Nov. 1996b.
- FRAWLEY, W. J. et al. **Knowledge Discovery in Databases: An Overview**. AI Magazine, Califórnia, v. 13, n. 3. P. 57-70, 1992.
- FININ, T.; DING, L.; ZHOU, L.; AND JOSHI, A. **Social networking on the semantic web. The Learning Organization**, 12(5):418–435, 2005.
- FONSECA, STELLA OGGIONI; NAMEN, ANDERSON AMENDOEIRA. **Mineração em Base de Dados do INEP: Uma Análise Exploratória para Nortear Melhorias no Sistema Educacional Brasileiro**. 2015. 26f. Trabalho de Conclusão de Curso (Pós-Graduação em Informática na Educação) - Universidade do Estado do Rio de Janeiro, [S.l.], 2014. DOI <https://doi.org/10.1590/0102-4698140742>. Disponível em: <https://www.scielo.br/j/edur/a/4HycJMvexZccLVMDg6qyRsw/?lang=pt>. Acesso em: 21 fev. 2022.
- GOLDSCHIMIDT, R.; PASSOS, E. **Data Mining: um guia prático**. Rio de Janeiro: Elsevier, 2005.
- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN 0123814790.
- INEP, Portal: **Inep divulga dados inéditos sobre fluxo escolar na educação básica**. 20 jun. 2017. Notícias. Disponível em: http://portal.inep.gov.br/artigo//asset_publisher/B4AQV9zFY7Bv/content/inep-divulga-dados-ineditos-sobre-fluxo-escolar-na-educacao-basica/21206.
- Acesso em: 20 out. 2021.
- HERSHKOVITZ, A. AND NACHMIAS, R. (2008). **Developing a log based motivation-measuring tool**. In de Baker, R. S. J., Barnes, T., and Beck, J. E., editors, EDM, pages 226–233. www.educationaldatamining.org.
- KARTHIKEYAN, V.G; THANGARAJ, P; KARTHIK, S et al. **Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation**. 2020. 11 f. Department of Computer Science and Engineering, Adithya Institute of Technology Department of Computer Science and Engineering, Adithya Institute of Technology, Coimbatore, Tamilnadu 641107, India, 2020. DOI <https://doi.org/10.1007/s00500-020-05075-4>. Disponível em: <https://link.springer.com/article/10.1007/s00500-020-05075-4>. Acesso em: 17 fev. 2022.
- LLOYD, S. **Least squares quantization in pcm**. Information Theory, IEEE Transactions on, 28(2):129 – 137.
- MACHADO, ROGER DOUGLAS et al. **Lacuna para Pesquisas Científicas Sobre Técnicas de Mineração de Dados como Base para Auxiliar na Resolução de Problemas de Evasão Escolar**. 2017. 28 f. Trabalho de Conclusão de Curso (Pós-Graduação em Informática na Educação) - Universidade Federal do Rio Grande do Sul, UFRGS, 2017. DOI 10.14488/ENEGEP2017_TN_STP_245_419_31685. Disponível em: https://www.researchgate.net/publication/323381660_LACUNAS_PARA_PESQUISAS_CIENTIFICAS_SOBRE_TECNICAS_DE_MINERACAO_DE_DADOS_COMO_BASE_PARA_AUXILIAR_NA_RESOLUCAO_DE_PROBLEMAS_DE_EVASAO_ESCOLAR. Acessado em: 15 fev. 2022.
- MATHIAS, L. A. F. **Mineração De Dados Em Sistemas De Energia Elétrica**

Utilizando Algoritmos Fundamentados Em Lógica Paraconsistente Anotada - LPA. 2015. 70 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia Mecânica, Programa de Pós-Graduação em Engenharia Mecânica, Universidade Santa Cecília, Santos, 2015.

RIGO, SANDRO J. et al. **Aplicação de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios: Educational data mining and Learning analytics applications in dropout: opportunities and challenges.** 2014. 15 f. Trabalho de Conclusão de Curso (Pós Graduação em Computação Aplicada) - Universidade do Vale do Rio dos Sinos - Unisinos, Universidade Federal de Ciências da Saúde de Porto Alegre, Departamento de Educação e Informação em Saúde - UFCSPA, (S. 1.), 2013.

RODRIGUES, RODRIGO LINS et al. **Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem. Simpósio Brasileiro de Informática na Educação,** (S.L.), v. 0, n. 0, p. 1-10, 22 nov. 2013. Sociedade Brasileira de Computação. <http://dx.doi.org/10.5753/cbie.sbie.2013.607>. Acesso em: 19 ev. 2022

ROMERO, C., & VENTURA, S. (2010). **Educational data mining: a review of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and Reviews,** IEEE Transactions on, 40(6), 601-618.

ROY, SAGARDEEP; GARG, ANCHAL. **Predicting Academic Performance of Student Using Classification Techniques.** 2018. 26 f. International Conference on Electrical, Computer and Electronics (UPCON) (PhD) - Department of Computer Science and Engineering ASET, Amity University Noida, India, International Conference on Electrical, Computer and Electronics (UPCON), 2017. DOI 10.1109/UPCON.2017.8251112. Disponível em:

<<https://ieeexplore.ieee.org/document/8251112>>. Acesso em: 21 fev. 2022.

SANTOS, KELLY JOANY DE OLIVEIRA. **Education Data Mining Para Apoio à Gestão Estratégica da Identificação de Perfis Evasivos e Atenuação da Evasão Escolar no Ensino Superior.** Orientador: Dr. André Britto de Carvalho. 2020. 99 f. Dissertação (Mestrado) - Universidade Federal de Sergipe Centro de Ciências e Tecnologia, São Cristóvão, 2020.

SILVA, LEANDRO; PERES, SARAJANE; BOSCARIOLI, CLODIS. **Introdução à Mineração de Dados Com Aplicação em R** – Elsevier 2016. Disponível em: <<https://www.grupogen.com.br/introducao-a-mineracao-de-dados-com-aplicacoes-em-r>>. Acesso em: 20 Out. 2021

SOUZA, VANESSA FARIA; SANTOS, TONY C. B. et al. **Processo de Mineração de Dados Educacionais aplicado na Previsão do Desempenho de Alunos: Uma comparação entre as Técnicas de Aprendizagem de Máquina e Aprendizagem Profunda: Educational Data Mining Process applied to Student Performance Prediction: A comparison between Machine Learning and Deep Learning Techniques.** 2021. 28 f. Trabalho de Conclusão de Curso (Pós Graduação em Informática na Educação) – Universidade Federal do Rio Grande do Sul, UFRGS, 2021.

PAIVA, R., BITTENCOURT I. I., PACHECO H., SILVA A. P., JACQUES P., ISOTANI S. **“Mineração de dados e a gestão inteligente da aprendizagem: desafios e direcionamentos”,** XXXII Congresso da Sociedade Brasileira de Computação, 2012, Curitiba. Anais do DESafIE! - I Workshop de Desafios da Computação Aplicada à Educação, 2012.

RISSINO, S. das D. **Metodologia de Avaliação da Relevância de Atributos em Grandes Bases de Dados Incompletas Utilizando Conjuntos Aproximados e Lógica Paraconsistente.** 2009. Tese

(Doutorado em Engenharia Elétrica) -
Universidade Federal de Itajubá, Minas
Gerais.

TAN, P.; STEINBACH, M. e KUMAR, V.
**Introdução ao Datamining: Mineração
de Dados.** Rio de Janeiro: Ciência Moderna,
2009.

WATSON, R. T. **Data Management:
Banco de dados e Organizações.** 3^a.
Edição. Editora LTC. Rio de Janeiro. p.
100-155. 2004.

WILLIAMS, G. J.; HUANG,
Zhexue. Modelling. **Modelling the KDD
Process.** Data Mining Portfolio, Australia,
1996.

WITTEN, I. H. AND FRANK, E. (2005).
**Data Mining: Practical Machine
Learning Tools and Techniques.** Morgan
Kaufmann Series in Data.