

UNIVERSIDADE EVANGÉLICA DE GOIÁS – UniEVANGÉLICA
BACHARELADO EM ENGENHARIA DE SOFTWARE

Iago Marques dos Santos
André Silva Fernandes

**EVASÃO UNIVERSITÁRIA: APLICANDO A MINERAÇÃO
DE DADOS PARA A PREDIÇÃO DE UM PERFIL DE RISCO**

Anápolis - GO

2022

Iago Marques dos Santos

André Silva Fernandes

EVASÃO UNIVERSITÁRIA: APLICANDO A MINERAÇÃO DE DADOS PARA A PREDIÇÃO DE UM PERFIL DE RISCO

Projeto de Pesquisa apresentado como requisito parcial para a conclusão da disciplina de Trabalho de Curso II do curso de Bacharelado em Engenharia de Software da Universidade Evangélica de Goiás – UniEVANGÉLICA.

Orientador (a): Prof. Eduardo Ferreira de Souza

Anápolis - GO

2022

Iago Marques dos Santos

André Silva Fernandes

**EVASÃO UNIVERSITÁRIA: APLICAÇÃO DA MINERAÇÃO DE
DADOS PARA CONSTRUÇÃO DE UM PERFIL DE RISCO**

Trabalho de Curso II apresentado à
Universidade Evangélica de Goiás para a
obtenção do título de Bacharel em Engenharia
de Software, sob a orientação do Prof.
Eduardo Ferreira de Souza.

Aprovado (a) pela banca examinadora em 14 de D e z e m b r o de 2022, composta por:

Prof. Eduardo Ferreira de Souza

Prof^a Pollyana dos Reis Pereira Fanstone

Prof. William Pereira dos Santos Junior

Resumo

A evasão universitária ainda é um dos desafios a ser combatido no ambiente das instituições de ensino, devido ao fato de ser um problema que possui características locais, de instituição para instituição, população, ambiente estudantil, entre outros, podendo variar bastante, o que requer estudos constantes. A evasão é um problema que tem gerado prejuízos tanto para o país, como para alunos e instituições de ensino. Tendo em vista esse cenário, o objetivo deste trabalho é utilizar as técnicas de Mineração de Dados, e aplicar para prever os perfis acadêmicos com risco à evasão na UniEVANGÉLICA. O tratamento de dados é realizado pelo *software WEKA*, seguindo a metodologia DCBD. Assim, é possível analisar e prever os acadêmicos que possuem risco à evasão, baseado nos conhecimentos obtidos durante o processo. Espera-se que este trabalho possa servir como ferramenta extremamente oportuna e viável para identificar possíveis fatores que podem levar os estudantes a evadirem, e a partir dos conhecimentos obtidos, embasar a criação de uma estratégia e prevenção de evasão para as instituições de ensino.

Palavras chaves: Mineração de Dados; Evasão Universitária; IES.

LISTA DE SIGLAS E ABREVIATURAS

ARFF	Formato de Arquivo de Relação de Atributo
CNTE	Confederação Nacional dos Trabalhadores em Educação
DCBD	Descoberta de Conhecimento em Base de Dados
EDM	<i>Educational Data Mining</i>
IES	Instituições de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas
KDD	<i>Knowlegde Discovery in Database</i>
LGPD	Lei Geral de Proteção de Dados
MD	Mineração de Dados
ML	<i>Machine Learning</i>
SEMESP	Sindicato das Entidades Mantenedoras do Ensino Superior do
Estado de São Paulo	
SGBD	Sistema Gerenciador de Banco de Dados
TI	Tecnologia da Informação
UniEVANGÉLICA	Universidade Evangélica de Goiás

SUMÁRIO

1 INTRODUÇÃO	7
2 FUNDAMENTAÇÃO TEÓRICA	9
2.1 EVASÃO UNIVERSITÁRIA.....	9
2.2 MINERAÇÃO DE DADOS	10
2.3 MINERAÇÃO DE DADOS EDUCACIONAIS	11
2.4 PROCESSOS DE KDD (<i>Knowledge Discovery in Database</i>).....	12
3 TRABALHOS RELACIONADOS.....	13
4 METODOLOGIA	14
5 RESULTADOS ALCANÇADOS	16
6 ANÁLISE E DISCUSSÃO DOS RESULTADOS	17
6.1 LEVANTAMENTO DE DADOS.....	17
6.2 PREPARAÇÃO DOS DADOS.....	18
6.3 APLICAÇÃO DOS ALGORITMOS DE MINERAÇÃO DE DADOS	19
7 CONCLUSÕES E TRABALHOS FUTUROS	19
REFERÊNCIAS BIBLIOGRÁFICAS	20

1 INTRODUÇÃO

O Sistema Educacional Brasileiro possui um problema complexo, que atinge inúmeras instituições de graduação, que é a evasão e retenção nos cursos. Os índices de evasão de estudantes podem chegar a níveis alarmantes, segundo o Governo Federal (BRASIL, 2007). Esse problema atinge tanto as Instituições de Ensino Superior (IES) públicas quanto privadas. Diferentes motivos ocasionam tal circunstância.

As condições que podem motivar a evasão são classificadas em três agrupamentos: O Econômico, que impossibilita o aluno de permanecer no curso por questões socioeconômicas; O Vocacional onde o aluno não se identifica com o curso; O Institucional tendo abandono por fracasso nas disciplinas iniciais, deficiências prévias de conteúdos anteriores, inadequação aos métodos de estudo, dificuldades de relacionamento com colegas, ou com membros da instituição (BARROSO; FALCÃO, 2004).

Outro fator que influenciou o aumento de evasão dos estudantes no último ano, foi à pandemia do COVID-19. Uma pesquisa feita pelo Sindicato das Entidades Mantenedoras do Ensino Superior do Estado de São Paulo (SEMESP) mostra que cursos de graduação tiveram um aumento de evasão de 5,9% na modalidade presencial e de 5% na modalidade à distância no ano de 2020, comparado ao ano de 2019 (LIN, 2021).

Pesquisas mostram que fazendo uso apenas das primeiras notas semestrais dos calouros, é possível identificar com precisão de até 80% a situação final de um aluno no curso (BARROS *et al.*, 2019).

Como aplicar a mineração de dados, para traçar os possíveis alunos com risco de evasão?

Um método utilizado para o estudo das causas da evasão escolar é o uso da descoberta de conhecimento em base de dados, por meio de técnicas de Mineração de Dados (MD), denominado *Educational Data Mining* (EDM) (ROMERO; VENTURA, 2007). A descoberta do conhecimento em dados educacionais será adotada neste trabalho, no qual será utilizado dados dos alunos da Universidade Evangélica de Goiás (UniEVANGÉLICA) para identificar quais cursos possuem maior índice de evasão.

Sendo assim, o presente trabalho tem como temática apresentar os recursos da Mineração de Dados para predizer perfis acadêmicos com risco de evasão conforme seu histórico acadêmico, no qual utiliza-se a base de dados para a descoberta de conhecimento. Salienta-se que, os dados disponibilizados para este trabalho com o aval do Pró-Reitoria Acadêmico junto aos gestores que realizam a manutenção dos sistemas acadêmicos desta IES.

Com isso, busca-se auxiliar o (a) gestor (a) da instituição na tomada de decisão e na elaboração de políticas para reduzir o índice de evasão.

Este projeto tem como objetivo apresentar as técnicas de Mineração de Dados e verificar a disponibilidade em aplicação para prever os perfis acadêmicos com risco à evasão na UniEVANGÉLICA com base em seus registros acadêmicos da seguinte forma:

1. Levantamento bibliográfico sobre a utilização de mineração de dados para a predição escolar.
2. Delimitar os dados pertinentes para a execução de algoritmos de predição.
3. Mapear técnicas para a aplicação da Mineração de Dados.
4. Desenvolver proposta de relatório para identificação.

Muitas empresas, segundo o Instituto Semesp, hoje usam da mineração de dados para encarar diversos problemas e em diversas áreas, seja no setor de serviços, comércio ou na indústria, ela ajuda no enfrentamento da dificuldade da empresa de se adaptar às constantes mudanças. Com o aumento constante da evasão no ensino superior, principalmente no período da pandemia, que aumentou de 30% para 35% de 2019 para 2020 (SEMESP, 2021).

A mineração de dados, usando diversas técnicas, pode fazer previsões de comportamentos futuros que incluem demandas, riscos e novos mercados. As empresas podem reduzir seus riscos, aumentar sua renda, melhorar o relacionamento com o cliente além de vários outros benefícios.

Como a mineração de dados é um processo em que a tecnologia é utilizada para localizar padrões em uma grande quantidade de dados, ele pode facilmente identificar o padrão presente nos casos de evasão nas universidades. Não só em casos de alunos já evadidos, ela pode analisar o desempenho de acadêmicos com dados, como também proporcionar uma porcentagem de alunos com situações que podem levar a evasão.

Resultará com a aplicação da mineração de dados, uma facilidade em organizar a rotina de Gestão Acadêmica da IES, para o controle e entendimento dessa alta quantidade de dados. Com isso, acarretará criações de rotinas semestrais para a obtenção de dados e campanhas de retenção. Além de benefícios como, uma melhor gestão de tempo, gerando uma melhor usabilidade, que facilitará o manuseio do (a) Gestor (a) Acadêmico (a) da Instituição para que tenha total confiabilidade em utilizar os recursos do sistema após sua implementação. Desse modo, a Mineração de Dados proporciona aos gestores desta IES, informações privilegiadas, que haveria maior dificuldade para serem geradas sem a utilização

desse recurso.

Portanto, os recursos da Mineração de Dados neste trabalho, procurou-se predizer os padrões de evasão de alunos da UniEVANGÉLICA, utilizando a base de dados acadêmicos disponibilizada por esta IES, nas quais foram definidos rendimento escolar, curso, sexo, cidade, tipo de instituição em que se concluiu o Ensino Médio, situação financeira dentre outros dados relevantes para análise.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 EVASÃO UNIVERSITÁRIA

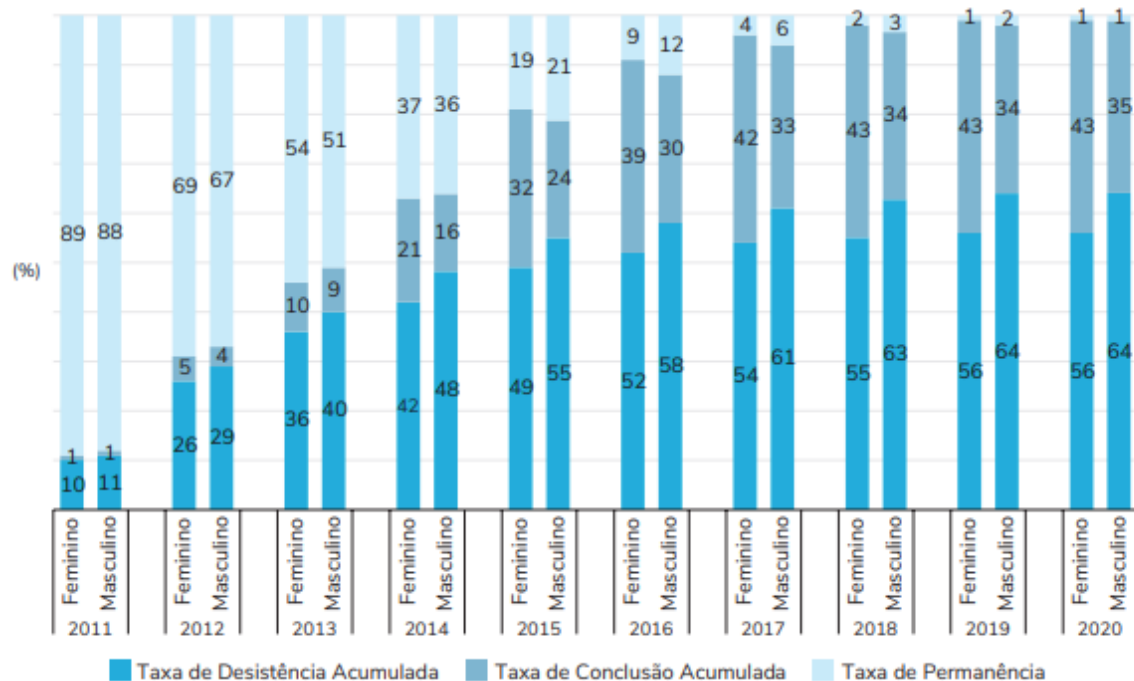
A evasão no ensino superior é um problema nacional que afeta as instituições e o sistema educacional do país. Ainda assim, segundo Silva Filho *et al.* (2007), a grande maioria das IES do Brasil não possui nenhum programa de combate à evasão.

Ainda de acordo com Silva Filho *et al.* (2007), o Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia realizou uma coleta de dados entre os anos de 2000 e 2005 com os cursos de Bacharelados Presenciais no Brasil, com base nos dados do Instituto Nacional de Estudos e Pesquisas (INEP), e teve como resultado uma média de 22% nas IES do Brasil.

A pesquisa tinha como objetivo identificar as principais causas desse fenômeno, no qual pode ser caracterizado em fatores internos e externos. Os fatores internos seriam questões relacionadas à estrutura e dinâmica de todos os cursos, e os fatores externos, seriam questões voltadas para o setor econômico, social e até mesmo individual podendo intervir na vida acadêmica gerando essa grande quantidade de evasões.

Com relação aos dados de evasão pelo Censo do Ministério da Educação (MEC) 2020 obtém-se a evolução da média dos indicadores de ingressantes em Cursos de Graduação por sexo, na qual é notório o aumento gradativo relacionado à taxa de evasão entre os anos de 2011 a 2020, segundo gráfico.

Gráfico 1: Evolução da Média dos Indicadores de Trajetória dos Ingressantes em Cursos da Graduação, por sexo – Brasil – 2011 – 2020.



Fonte: Censo da Educação Superior 2020 – MEC.

Após análise no gráfico 1, conclui-se que a taxa de ingressantes e permanência no ensino superior em 2020 diminuiu consideravelmente, com relação ao levantamento realizado pelas IES nos anos anteriores perante o Censo da Educação Superior – MEC, tendo assim, um aumento significativo na taxa de desistência/evasão entre os anos de 2011 – 2017 e sendo estabilizado nos anos subsequentes até 2020.

2.2 MINERAÇÃO DE DADOS

Desde o surgimento dos sistemas computacionais, um dos principais objetivos das organizações é o armazenamento de dados. Nas últimas décadas essa tendência ficou ainda mais em evidência devido à queda nos custos de aquisição de *hardware*, tornando possível armazenar dados cada vez maiores.

Com todo esse aumento em armazenamento de dados, as técnicas tradicionais de exploração de dados não eram mais adequadas para tratamento de dados na maioria dos repositórios. Com isso, no final da década de 80 foi proposto por Usama Fayyad a Mineração de Dados (*Data Mining*, termo em inglês), com a finalidade de resolver o problema encontrado na época.

A Mineração de Dados é uma das tecnologias mais promissoras da atualidade. Um dos

fatores de sucesso é o fato de dezenas, e muitas vezes centenas de milhões serem gastos por companhia na coleta de dados e, no entanto, nenhuma informação relevante é identificada.

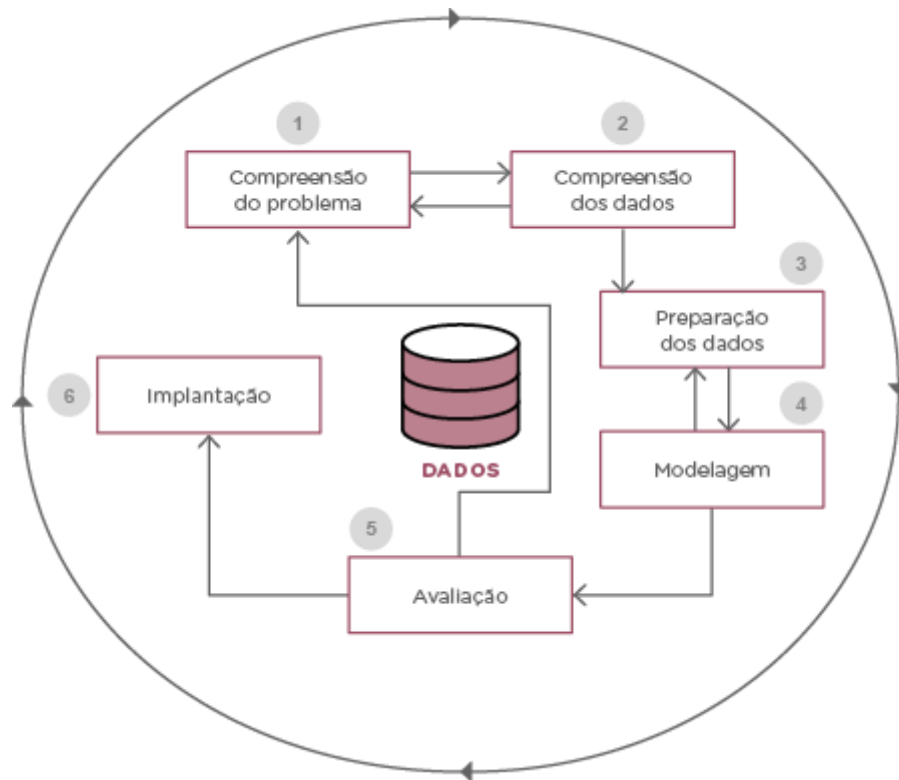
Witten e Frank (2005), Olson e Delen (2008) e Bramer (2007), apresentam algumas áreas onde a Mineração de Dados é aplicada de forma satisfatória:

- Retenção de Clientes: Identificar perfis de determinados produtos, venda cruzada;
- Bancos: Identificar padrões para auxiliar no gerenciamento de relacionamento com o cliente;
- Cartão de Crédito: Identificar segmentos no mercado, identificar padrões de rotatividade;
- Cobranças: Detecção de fraudes;
- Eleitoral: Identificação de um perfil para possíveis votantes;
- Medicina: Indicação de diagnóstico mais preciso.

2.3 MINERAÇÃO DE DADOS EDUCACIONAIS

A *Educational Data Mining* (EDM) ou Mineração de Dados Educacionais é um método desenvolvido para explorar dados únicos provenientes de ambientes educacional para melhor entendimento. A EDM utiliza de vários recursos providos pela MD para conversão de dados brutos dos sistemas digitais em informações úteis, na qual pode impactar pesquisas educacionais. Alguns dos métodos mais utilizados são as clusterizações, a mineração de textos e relacionamentos, dentre outros.

O processo da EDM possui uma estreita semelhança com o processo de KDD, na qual consiste com a compreensão do problema, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação. Para maiores entendimentos, veja detalhadamente o modelo do projeto na figura abaixo:

Figura 2: Etapas do Processo de Mineração de Dados Educacionais

Fonte: Filatro Andrea, 2020.

2.4 PROCESSOS DE KDD (*Knowledge Discovery in Database*)

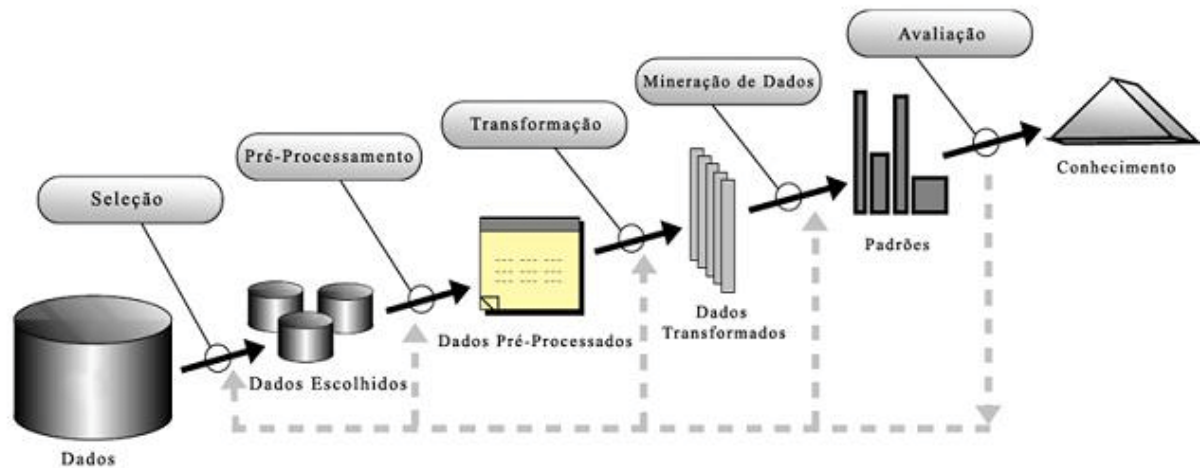
Com o avanço da tecnologia, informações são geradas em altas quantidades e mudam rapidamente. Sendo assim, é necessário armazenar esses dados em grande base de dados para que seja possível transformá-las em conhecimento no qual é utilizado em tomada de decisões e estratégias.

A área da Tecnologia da Informação (TI) vem avançando bastante nos últimos anos, trazendo consigo a modernização de tecnologias existentes, tais como: computadores, celulares, Sistema Gerenciador de Banco de Dados (SGBD) entre outros. Esses recursos facilitam a propagação de bases, fazendo com que esses dados precisem ser analisados para que sejam transformados em informações relevantes que serão utilizadas para gerar conhecimento. O processo para análise, interpretação de dados e estudos desses dados seriam impossível para o ser humano sem assistência de ferramentas pertinentes, gerando assim, a necessidade de desenvolvimento das Técnicas de Mineração de Dados para auxiliar o homem em estratégias dentro de seu contexto.

O processo de KDD foi normatizado em 1989 com objetivo em extração de conhecimento a partir dessas bases de dados. Para a exploração dessa informação é preciso

seguir uma sequência de processos que devem ser obedecidas, tendo início na coleta de informação e, por fim, na extração de conhecimento em seu resultado final. As etapas operacionais deste processo são compostas conforme ilustrado na figura 1.

Figura 2: Etapas do Processo de KDD/DCBD



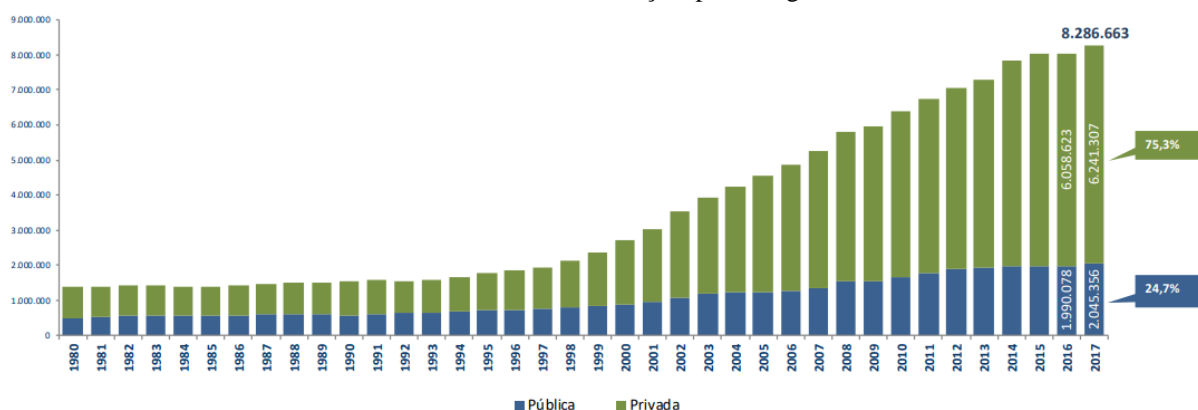
Fonte: Fayyad *et al.* (1996) *apud* Camilo e Silva (2009)

3 TRABALHOS RELACIONADOS

Neste tópico, será mostrado os trabalhos relacionados ao tema de Evasão Universitária: Aplicando a Mineração de Dados para a predição de um perfil de risco.

O principal propósito de uma instituição de ensino superior, além de entregar a melhor formação possível para seus alunos, é evitar a evasão. Por isso, a permanência de alunos é a meta constante de quase toda IES e um ponto que todas têm em comum e o que impossibilita atingir essa meta é o alto índice de evasão.

Desde o início da década de 90, o Brasil tem assistido a um aumento considerável do contingente de ingressantes em IES, da mesma forma que a quantidade de estabelecimentos e cursos também em apresentado um aumento importante. De acordo com os dados obtidos pelo Censo da Educação Superior em 2017, em 1991, foram registradas cerca de 1.500.000 matrículas em IES privadas, e em 2007, esse número se aproximava de 5.000.000. Em 2011, o número total de matrículas foi de 6.739.689, segundo gráfico.

Gráfico 2: Número de Matrículas em Cursos de Graduação, por Categoria Administrativa – 1980 – 2017.

Fonte: Censo da Educação Superior 2017 – MEC.

Ressalta-se que uma parte das pessoas que ingressam em um curso superior, não o conclui. Do ponto de vista das instituições, números tão altos tendem a prejudicar o planejamento financeiro e acadêmico e, especialmente para as públicas, cada aluno que deixa de concluir o curso significa um desperdício de investimento público, segundo Silva Filho *et al.* (2007).

De acordo com Ruiz *et al.* (2007), um estudo realizado pela Confederação Nacional dos Trabalhadores em Educação (CNTE) apontou que a evasão nas universidades brasileiras é uma realidade que vem se intensificando no decorrer dos tempos. Em vários cursos do ensino superior as taxas (os índices) de evasão estão cada vez mais elevadas, acarretando assim (ou ao menos contribuindo e muito) para a escassez, e dependendo da região do Brasil, na falta de professores com formação adequada para ingressar no mercado de trabalho, seja nos níveis fundamental e médio.

4 METODOLOGIA

O levantamento bibliográfico para essa pesquisa foi o ponto inicial deste trabalho de curso, com base nele foram delineadas as técnicas e ferramentas que serão adotadas em seu desenvolvimento. Quanto aos objetivos, este trabalho inicializou uma pesquisa quantitativa exploratória, tratando-se de um projeto com o objetivo em obter um resultado exato.

Em primeiro instante, foi realizada uma solicitação perante a Pró-Reitoria Acadêmica da Universidade Evangélica de Goiás – UniEVANGÉLICA para a obtenção dos dados solicitados, tendo como prioridade seguir as normativas da Lei Geral de Proteção de Dados (LGPD) para que não seja possível identificar informações pessoais de tais dados dos envolvidos neste levantamento, visando assim, manter a integridade das informações fornecidas.

Na parte sequente, foi aplicado a clusterização que é uma Técnica de Mineração de Dados conhecida como Agrupamento de Dados. Esse é um processo no qual é realizado a identificação de dados semelhantes e não semelhantes entre si. Essa segmentação de dados é essencial para a seleção de grupos de dados e posteriormente a geração de *insights*.

Na próxima etapa, foram organizados os dados com base em suas estruturas e seu tipo, podendo realizar a transformação alinhando de acordo com os objetivos da mineração. Acarretando em um melhor tempo de resposta, tornando as informações mais fáceis de discernir os padrões finais.

Para a realização do tratamento na base de dados, utiliza-se o *software Weka*, para aplicação das técnicas de Mineração de Dados, uma ferramenta que proporciona facilidade em seu manuseio, contém uma biblioteca de algoritmos e classificação, geração e visualização de relatórios no qual pode-se aplicar as técnicas que serão utilizados neste trabalho (WEKA, 2017).

Por fim, com a importação da base de dados no *Weka Explorer*, logo após é possível tratar as informações e convertê-las para o formato ARFF (Formato de Arquivo de Relação de Atributo), para que os dados contidos naquela base possam ser lidos pela ferramenta e realize as comparações necessárias durante sua análise.

Após a obtenção da base de dados e entendimento no modelo sequencial de desenvolvimento, todas as etapas citadas foram organizadas sequencialmente em um fluxograma como descrito figura 2.



Fonte: Autores.

5 RESULTADOS ALCANÇADOS

Essa pesquisa acadêmica, busca-se aplicar as Técnicas de Mineração de Dados para predição de perfis acadêmicos com riscos de evasão na UniEVANGÉLICA com base nos registros acadêmicos desta IES.

Definir detalhadamente o problema supracitado neste trabalho foi essencial, não apenas para um bom resultado, mas também para as escolhas das técnicas que serão utilizados para a predição. Através do processo de MD, foram obtidos os resultados de melhor funcionamento de processo, melhora dos resultados em métricas e indicadores, desenvolvimento de uma inovação na área da evasão universitária.

É perceptível que a taxa de permanência no Ensino Superior em 2020, diminuiu consideravelmente, com relação ao levantamento realizado pelas IES nos anos anteriores (2011 – 2019) perante o Censo Superior, tendo assim, um aumento significativo na taxa de desistência/evasão ao longo dos anos.

O resultado obtido, sem dúvidas, foi à absorção de conhecimento durante a pesquisa para o desenvolvimento do referido trabalho. Os dados estão em todos os lugares, e com o pensamento lógico e estratégico, será utilizada a percepção adquirida durante o levantamento bibliográfico para a resolução do problema enfrentado nesta IES, conforme assunto tema.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste tópico, serão apresentados as metodologias que foram realizadas para a conclusão desse trabalho de curso. Na seção 6.1, é apresentado o levantamento dos dados da UniEVANGÉLICA. Na seção 6.2, é apresentado a etapa de preparação dos dados necessários para a aplicação das técnicas de MD. Na seção subsequente 6.3, é apresentada algoritmos de classificação de dados com o auxílio do *software Weka*.

6.1 LEVANTAMENTO DE DADOS

Nesse estudo, foram utilizados os dados histórico de estudantes dos cursos de Bacharelado em Direito, Educação Física e Administração. Os dados foram disponibilizados pela ProACAD da IES e contém registros acadêmicos com a situação atual do estudante até o ano de 2022. Ressalta-se que esse quantitativo de alunos é impossível de ser identificado, uma vez que, a identificação dos mesmos está sendo preservada. Os registros representam dados de alunos com sua situação de matrícula evadida, ou seja, que abandonaram o curso. A base de dados é composta por 3739 registros, na tabela 1 é apresentado o quantitativo desses registros por curso.

Tabela 1: Quantitativo de registros por curso.

Curso	Número de Evadidos
DIREITO	2233
EDUCAÇÃO FÍSICA	783
ADMINISTRAÇÃO	723

Fonte: Autores.

O aluno é classificado como evadido caso não realize matrícula em pelo menos 1 disciplina, ao longo dos períodos letivos seguintes. Por meio dessa análise, foi possível identificar e selecionar as informações de maior significância para o experimento realizado. Os atributos escolhidos para pesquisa foram:

- Curso: identifica em qual o curso foi regularmente matriculado.
- Turno: refere-se a qual período do dia o curso é ofertado.
- Data Ingresso: trata-se da data em que o estudante ingressou na IES.
- Unidade Física: local onde foi regularmente matriculado.
- Situação: refere-se ao status atual em que se encontra no sistema.

- Tipo de Instituição (Ensino Médio): refere-se ao tipo da instituição em que concluiu o ensino médio (pública ou privada).
- Disciplinas Matriculadas: Número de disciplinas em que ficou matriculado.
- Disciplinas Reprovadas: Número de disciplinas em que reprovou.
- Ano/Semestre: Ano e Semestre cursado.
- Município: local onde reside.
- Cor/Raça: identificação de qual tipo de cor/raça.
- Sexo: identificação do sexo.
- Situação Financeira: trata-se da identificação de que se o estudante está adimplente ou inadimplente.
- Possui Bolsa: identificação de que se o estudante possui algum tipo de custeio para pagamento de suas mensalidades.

6.2 PREPARAÇÃO DOS DADOS

A partir da obtenção dos dados, os mesmos podem ser armazenados em um Sistema de Gerenciamento de Banco de Dados (SGBD) em três tabelas, a primeira possuía 4 atributos, acerca de seus dados pessoais dos alunos composta por: município onde reside, cor/raça na qual se identifica, sexo e o tipo de instituição em que concluiu o ensino médio. A segunda tabela possuía 9 atributos com relação a situação acadêmica formado pelo curso, turno, data e tipo de ingresso, unidade física na qual está vinculado (a), situação, quantidade de disciplinas matriculadas e reprovadas, ano e semestre. Já a terceira tabela possuía informações referente ao seu resumo financeiro composta por situação financeira, possui bolsa e nome bolsa.

Para obter a extração, transformação e carregamento dos dados mapeados, pode ser utilizado a ferramenta *Power Query*, no qual é um dos recursos disponibilizados pelo *PoweBI* para execução da tarefa e exportação dos dados transformados em seu respectivo formato .CSV para importação no *software Weka* para dar início as análises e aplicações dos algoritmos de MD.

6.3 APLICAÇÃO DOS ALGORITMOS DE MINERAÇÃO DE DADOS

Nesta etapa pode realizar testes de classificação na base de dados apresentada na seção 6.1 utilizando algoritmos populares de *Machine Learning* (ML), no *software WEKA*, permitindo que as atividades de pré-processamento, mineração de dados e sua visualização de forma simples, além de facilitar a comparação entre diversos métodos e verificar o mais adequado ao problema.

Com a extensa biblioteca de algoritmos do *software WEKA*, pode ser designado para aplicação da MD 4 algoritmos inicialmente para testes. Os algoritmos selecionados foram: *Decision Tree*, *Naive Bayes*, *Support Vector Machine* (SVM), *Random Forest*. Nesta fase os dados são fragmentados, sendo um conjunto enviado para teste de predição do algoritmos classificador, e o outro remetido para o treinamento do algoritmo escolhido. Para medir o desempenho dos algoritmos na fase de classificação, considera-se a proporção em que a acurácia das instâncias foi classificada como correta e a taxa de revocação de frequência em que o classificador encontra os exemplos de uma classe.

7 CONCLUSÕES E TRABALHOS FUTUROS

Os resultado obtidos mostram que é viável aplicar a Mineração de Dados para identificar alunos que apresentam risco de evasão escolar. Neste estudo, propôs-se a aplicação da MD juntamente com 4 algoritmos de classificação no *software WEKA* para identificação de alunos com risco a evasão, com base em seus registros acadêmicos no decorrer de sua jornada universitária, tendo como início os alunos dos cursos de Direito, Educação Física e Administração na UniEVANGÉLICA, uma vez que, são os curso que mais possuem alunos com a situação de evadidos. Desse modo, espera fornecer um estudo que auxilie os gestores desta IES para a identificação dos estudantes que necessitam de apoio, contribuindo com sua tomada de decisão para mitigar à evasão escolar.

Como trabalhos futuros, sugere-se estender este estudo e iniciar a aplicação da MD nos cursos citados anteriormente, verificando a eficiência na predição a partir do desempenho nas disciplinas ao decorrer dos períodos. Ressalta-se também, a implementação de um *dashboard* interativo e dinâmico para análise de seus resultados e aplicações em tempo real. Evidencia que este trabalho será disponibilizado no Repositório Institucional para que possa contribuir com pesquisas futuras nesta mesma linha de pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

BARROS, R. P. Predição do rendimento dos alunos em lógica de programação com base no desempenho das disciplinas do primeiro período do curso de ciências e tecnologia utilizando técnicas de mineração de dados. 2019. VIII Congresso Brasileiro de Informática na Educação (CBIE 2019). **Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)**. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/viewFile/8882/6435>. Acesso em: 13 abr. 2022.

BARROSO, M. F.; FALCÃO, E. B. M. Evasão universitária: o caso do instituto de física da UFRJ. **IX Encontro Nacional de Pesquisa em Ensino de Física**, 2004. Disponível em: <https://www.if.ufrj.br/~carlos/pef/materiais/marta-epf2004-evasao-co12-2.pdf>. Acesso em: 26 abr. 2022.

BEZERRA, Camila et al. Evasão Escolar: Aplicando Mineração de Dados para Identificar Variáveis Relevantes. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, [S.l.], p. 1096, nov. 2016. ISSN 2316-6533. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/view/6795>. Acesso em: 03 jul. 2022.

BRAMER, M. **Principles of Data Mining - Undergraduate Topics in Computer Science**. 3 ed. Springer, 2007.

BRASIL. Ministério da Educação. **Diretrizes gerais do programa de apoio a planos de reestruturação e expansão das universidades federais (Reuni)**. Disponível em: <http://portal.mec.gov.br/sesu/arquivos/pdf/diretrizesreuni.pdf>. Acesso em: 24 abr. 2022.

CAMILO, O. C.; SILVA, C. J. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Instituto de Informática Universidade Federal de Goiás-UFG, 2009. 29 f. Disponível em: https://rozero.webcindario.com/disciplinas/fbmg/dm/RT-INF_001-09.pdf. Acesso em: 04 abr. 2022.

COLPANI, R. Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. **Informática na educação: teoria & prática**, Porto Alegre, v. 21, n. 3, 2018. DOI: 10.22456/1982-1654.87880. Disponível em: <https://seer.ufrgs.br/index.php/InfEducTeoriaPratica/article/view/87880>. Acesso em: 3 jul. 2022.

FILATRO, Andrea C. **Data science da educação**. [Digite o Local da Editora]: Editora Saraiva, 2020. E-book. ISBN 9786587958446. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786587958446/>. Acesso em: 24 ago. 2022.

LIN, N. **Evasão do ensino superior privado aumenta na pandemia**. Radioagência Nacional, São Paulo, 09 de jun. 2021. Disponível em: <https://agenciabrasil.ebc.com.br/radioagencia-nacional/educacao/audio/2021-06/evasao-do-ensino-superior-privado-aumenta-na-pandemia>. Acesso em: 07 abr. 2022.

MARTINS, R. R.; BERTUCI, M. H. .; PENIANI, L. P. . MINERAÇÃO DE DADOS NO COMBATE À EVASÃO ESCOLAR EM INSTITUIÇÕES DE ENSINO SUPERIOR BRASILEIRAS. **Revista Interface Tecnológica, [S. l.]**, v. 17, n. 2, p. 103–115, 2020.

Disponível em: <https://revista.fatectq.edu.br/interfacetecnologica/article/view/885>. Acesso em: 3 jul. 2022.

OLSON, D. L.; DELEN, D. **Advanced Data Mining Techniques**. Springer, 2008.

ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. **Expert systems with applications**, [S.l.], v. 33, n. 1, p. 135–146, 2007.

RUIZ, A. I.; RAMOS, M. N.; HINGEL, M. Escassez de professores no Ensino Médio: Propostas Estruturais e Emergenciais. Brasília, DF: Ministério da Educação.

SEMESP. Secretaria de Modalidades Especializadas de Educação. **Evasão durante a pandemia, o que as IES estão fazendo para mitigar a questão?** 2021. Disponível em: <https://www.semesp.org.br/eventos/evasao-durante-a-pandemia-o-que-as-ies-estao-fazendo-para-mitigar-a-questao/>. Acesso em: 13 abr. 2022.

SILVA FILHO, R. *et al.* A evasão no ensino superior brasileiro. Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia. **Cadernos de pesquisa**, v. 37, n. 132, p. 641-659, 2007.

WEKA. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 04/06/2022.

WITTEN, I. H; FRANK, E. **Data Mining – Practical Machine Learning Tools and Techniques**. Elsevier, 2005.