

Análise e predição de evasão dos alunos de um curso de Graduação em Sistemas de Informação por meio da mineração de dados educacionais

Edson Noetzold

Universidade Federal de Santa Maria Campus Frederico Westphalen
edsonversusnoetzold@gmail.com
<https://orcid.org/0000-0002-3491-501X>

Solange de L. Pertile

Universidade Federal de Santa Maria Campus Frederico Westphalen
solange.pertile@ufsm.br
<https://orcid.org/0000-0001-9922-5506>

Resumo: Os elevados índices de evasão escolar constituem uma realidade presente em diversos cursos superiores ofertados no Brasil, o que evidencia a necessidade de investigação dessa problemática, responsável por perdas econômicas nas instituições e impactos no cenário global da educação. Este artigo tem como proposta desenvolver um estudo sobre os padrões da evasão escolar no ensino superior, com base na análise de dados fornecidos pelo curso de Sistemas de Informação da Universidade Federal de Santa Maria - Campus Frederico Westphalen. Esses dados passaram por uma sistemática de tratamento de dados, a fim de apontar indicadores relacionados a fatores que classifiquem possíveis evasões. Foram gerados resultados por meio de árvores de decisão, que apontaram dados referentes ao aluno e seu desempenho acadêmico como fatores importantes para identificação da evasão escolar no ensino superior.

Palavras-chave: Evasão Escolar. Mineração de Dados. Ensino Superior.

Analysis and prediction of student dropout from an Undergraduate Information Systems course through educational data mining

Abstract: *The high school dropout rates are a reality present in several higher education courses offered in Brazil, which highlights the need to investigate this issue, which is responsible for economic losses in institutions and impacts on the global education scenario. This paper aims to develop a study on school dropout patterns in higher education, based on the analysis of data provided by the Information Systems course of an Institution. These data went through a data processing system, in order to point out indicators related to factors that classify possible evasions. Results were generated through decision trees, which pointed to data regarding the student and his academic performance as important factors for school dropout in higher education.*

Keywords: *School Dropout. Data Mining. Higher Education.*

1. Introdução

De acordo com Favero (2006), denomina-se a evasão escolar como o processo de desistência do ensino pelo discente de determinado curso, indiferentemente da

porcentagem de participação do aluno nas aulas. Já Almeida e Kappel (2020) apontam que tal problemática encontra-se emergente no cenário atual, como observado nas preocupantes taxas de evasão. Segundo Silva Filho *et al.* (2007), esse impasse vem impactando até mesmo o cenário internacional, afetando expressivamente os resultados dos sistemas educacionais e causando perdas nas instituições públicas e privadas. Já Johann (2012) define como evasão escolar a situação na qual o aluno, por qualquer motivo, rompe o vínculo estabelecido com a instituição de ensino, não renovando sua matrícula no curso.

Assim, a motivação para a delimitação da temática deste artigo reside em analisar o problema apresentado, bem como identificar o perfil de alunos propensos a evadir no curso de Sistemas de Informação, utilizando a mineração de dados como ferramenta. A escolha do curso a ser analisado se deu devido a uma análise dos dados dos alunos fornecidos pela coordenação do curso de Sistemas de Informação, a qual constatou uma taxa de evasão superior a 40% no período entre 2010 (início do curso) e 2019. Essa taxa foi calculada a partir do número de estudantes em situação de abandono, transferência interna/externa e cancelamento da matrícula em relação ao total de alunos ingressantes no mesmo período. A partir dos resultados obtidos neste trabalho, espera-se auxiliar os gestores na predição de alunos propensos a evadir, de modo a permitir a adoção de medidas preventivas que possam minimizar esse problema.

2 Dados e causas da evasão escolar no ensino superior

Para compreender melhor a questão da evasão, é importante observá-la sob dois panoramas diferentes: a evasão anual média e a evasão total. A evasão anual média visa medir o percentual de estudantes matriculados em um determinado curso de uma instituição de ensino superior que, ainda não formado, não efetivou sua matrícula para o semestre/ano subsequente. O cálculo dessa porcentagem utiliza o total de alunos vinculados ao curso, ou seja, o objetivo é descobrir quantos alunos evadiram em certo período em relação à quantidade total de estudantes (Silva Filho *et al.*, 2007).

Já a evasão total refere-se ao número de alunos que, após sua entrada na instituição de ensino, não concluiu a formação em um número de anos. Essa avaliação é comumente relacionada ao índice de titulação – que mede a porcentagem de formandos na totalidade de ingressantes de um determinado ano. Portanto, se cinquenta alunos ingressaram no ano “x” e vinte e cinco se formaram, tem-se um índice de titulação de 50%, com um índice de evasão também de 50% por consequência (Silva Filho *et al.*, 2007).

De acordo com Silva (2007), é imprescindível a investigação das causas de evasão nas instituições de ensino superior. Isso porque, embora já existam pesquisas apontando os agentes causadores da evasão escolar, é perceptível a ausência de homogeneidade nos graus de evasão para os diferentes cursos, impossibilitando a construção de um padrão universal.

A Plataforma Nilo Peçanha (PNP) é um ambiente virtual de cunho federal iniciado em 2017, desenvolvido com o intuito de unificar e disseminar as estatísticas oficiais da Rede Federal de Educação Profissional, Científica e Tecnológica (Rede Federal). Nesse contexto, o ambiente virtual disponibiliza ao usuário dados sobre as unidades que fazem parte da Rede Federal, como seus cursos ofertados, corpo docente, discente e técnico-administrativo, além de relatórios anuais de análise dos indicadores.

A PNP publicada no ano de 2020, a qual utilizou 2019 como ano base, apontou V. 19 N° 1, julho, 2021 RENOTE
DOI: <https://doi.org/10.22456/1679-1916.118525>

índices de evasão escolar no ensino superior das Instituições Federais de aproximadamente 12%, correspondentes principalmente à evasão por abandono (6,51%), seguido de desligamento (4,69%), cancelamento (0,23%), transferência externa (0,22%), reprovação (0,10%) e transferência interna (0,02%) (Peçanha, 2020). Tal percentual soma mais de trinta mil matrículas evadidas, evidenciando os elevados índices de evasão presentes na Rede Federal. A PNP publicada com base nos dados do ano de 2018 indicou cerca de 29.000 matrículas evadidas, permitindo concluir que a evasão no ensino superior está aumentando nas instituições federais (Peçanha, 2019).

Ao se abordar a questão das causas da evasão, é importante citar que se constitui de uma problemática de diversas causas, as quais podem ser classificadas como causas psicológicas, sociológicas, organizacionais, interacionais e econômicas (Schargel e Smink, 2002). A categoria psicológica abrange, de maneira geral, o conjunto de causas ligadas ao comportamento do indivíduo, que por sua vez, interferirá nos seus resultados acadêmicos. Nesse contexto, é possível citar primeiramente a influência das reprovações excessivas na motivação do estudante, que tende a diminuir conforme elas ocorrem. Para Negrine (1994), a motivação é um requisito indispensável ao processo de aprendizagem, visto que ela resulta em um comprometimento verdadeiro com o compromisso acadêmico.

Silva e Dessen (2003) também frisam a importância da família ao afirmarem que as interações estabelecidas no microsistema familiar têm um impacto mais expressivo no desenvolvimento individual, embora outros fatores sociais – como a escola e os outros círculos sociais também exerçam certa influência. Por fim, a imaturidade e rebeldia constituem fatores influentes no desempenho do aluno (Gaioso, 2005).

No campo da sociologia, é possível citar a falta de orientação vocacional – recurso importante no que tange o autoconhecimento do aluno e que possibilita a escolha profissional mais assertiva (Pimenta, 1981). Além disso, para Kirby et al. (2004), os estudos acabam surgindo como uma fonte de estresse no seio familiar, especialmente se tratando de estudantes mais velhos que já construíram suas próprias famílias, tornando possível apontar a família como um fator causador da evasão. Por fim, as dificuldades enfrentadas no ensino superior como consequência da educação básica deficiente também atuam como aspectos causadores da evasão no ensino válidos de serem declarados (Schargel e Smink, 2002; Gaioso, 2005).

Já a categoria organizacional, a qual reflete a influência da instituição de ensino sobre o discente, cita como causa para o problema em questão o desconhecimento do aluno para com a metodologia implementada na instituição onde estuda (Scharge e Smink, 2002). Conforme Gaioso (2005), a concorrência entre as diferentes instituições de ensino também interferem negativamente nesse problema.

Outra possibilidade para a evasão é referente a estrutura oferecida pela instituição, cuja qualidade pode ser questionada pelo aluno. Nessa perspectiva, Dalrymple e Parsons (2003) salientaram a importância da gestão da qualidade de serviços das instituições, que devem considerar todos os recursos e suportes necessários ao aluno para o bom rendimento acadêmico – como bibliotecas, secretarias, laboratórios, entre outros.

Os aspectos econômico-financeiros também possuem certo grau de ligação para com a evasão discente, como as divergências no concílio entre os horários de trabalho com a carga horária do curso e a ausência de vantagens imediatas promovidas pela titulação (Jacob, 2000). Para Gomes (1999), o interesse em cursar o ensino superior está diretamente vinculado à possibilidade de ascensão econômica e social que, quando não

ocorre rapidamente, pode levar à frustração do aluno e, conseqüentemente, à evasão no ensino.

A universidade, diante das transformações econômicas, políticas e culturais recentes que influenciam a educação, enfrenta a necessidade de repensar e transformar seus vínculos com a sociedade. É preciso corrigir alguns fatores e eliminar outros para que os acadêmicos possam ter, além do acesso à universidade, a garantia da conclusão do curso (Amaral, 2016).

Em um cenário de constantes mudanças econômicas, políticas e culturais que afetam direta ou indiretamente a educação, é de suma importância que cada instituição de ensino pesquise e desenvolva estratégias de transformação de seus vínculos com a sociedade. É necessário oferecer recursos para que o estudante possa não apenas ingressar no ensino superior, como também concluir seu curso escolhido, e deve-se combater os fatores divergentes a essa perspectiva (Amaral, 2016).

3. Trabalhos Relacionados

Paz e Cazella (2017) procuraram desenvolver um estudo de caso de cunho exploratório acerca da identificação do perfil de alunos propensos à evasão escolar. O estudo teve enfoque em uma Universidade Comunitária, de onde foram coletados dados referentes aos alunos regularmente matriculados no segundo semestre de 2016. Para o desenvolvimento do estudo de caso, foram utilizadas técnicas de mineração de dados. Especificamente, a mineração ocorreu por meio da tarefa de classificação dos dados na ferramenta WEKA, juntamente com a técnica de construção de árvores de decisão a partir do algoritmo J48. Os materiais utilizados na mineração foram cedidos pelos cursos de graduação de uma Universidade Comunitária do Rio Grande do Sul providos pelo setor de Tecnologia da Informação (TI). De acordo com os resultados da mineração, foi possível aos autores concluir que alunos de semestres iniciais sem auxílio (financeiro, por exemplo) possuem uma alta incidência de evasão. A hipótese levantada inicialmente pelos autores de que morar fora da cidade onde se situa o campus seria um fator para a evasão foi descartada. Por último, foi constatado que há uma tendência geral de evasão nos semestres iniciais.

Manhães et al. (2012) investigaram os problemas relacionados a alunos evadidos em uma Universidade Federal, o que foi desenvolvido a partir de diversas técnicas de mineração de dados. A realização do projeto teve como base um estudo de caso envolvendo alunos de graduação da Universidade Federal do Rio de Janeiro (UFRJ), cujas informações foram mineradas por técnicas de mineração, visando localizar relações que direcionassem o desempenho acadêmico do discente. As técnicas utilizadas foram avaliadas a fim de medir sua acurácia quando aplicadas a dados estudantis. Nesse contexto, o algoritmo *Naive Bayes* foi selecionado, devido a seu modelo interpretável e à capacidade de conversão gráfica de seus resultados numéricos, o que permite a análise quantitativa dos resultados posteriormente. A execução do algoritmo selecionado se deu no software WEKA, ferramenta escolhida para a realização da mineração devido às suas diversas versões de algoritmos implementados, sua modalidade *open source* e as opções de recursos estatísticos para a análise de dados. Com os resultados da mineração, foi possível concluir que alunos evadidos apresentavam no mínimo uma disciplina reprovada por frequência e média e pelo menos uma disciplina reprovada por média no primeiro semestre, redução no número de disciplinas cursadas e aprovações por semestre e média inferior às dos demais alunos ao final do primeiro semestre. Quanto aos alunos que vieram a concluir o curso, foi mostrado que costumam manter o número

de disciplinas cursadas por semestre, possuem alto índice de aprovações, suas médias semestrais se mantêm próximas do coeficiente de rendimento acadêmico acumulado até em torno do oitavo semestre e seu número de disciplinas reprovadas por média se mantêm próximo de zero. Em suma, foi observado que os alunos que concluíram o curso apresentaram um comportamento homogêneo entre os semestres que se sucederam, com uma tendência geral de aumento de reprovações por média ao final do curso – possivelmente em função de estágios curriculares ou outras atividades.

Gonçalves, Da Silva e Cortes (2018) utilizaram a mineração de dados como recurso para identificar alunos propensos à evasão escolar no ensino superior do Instituto Federal do Maranhão (IFMA). O projeto foi desenvolvido no formato de estudo de caso, o qual foi composto pela identificação do problema, escolha das técnicas de mineração para a adequação dos dados, escolha dos algoritmos e escolha dos parâmetros para validação. Para a identificação do problema, foram construídos levantamentos teóricos acerca da evasão escolar no ensino superior e da aplicação da mineração de dados no contexto educacional, o que permitiu um discernimento maior por parte do grupo quanto à problemática. Em seguida, a segunda etapa se deu com o pré-processamento dos dados fornecidos à equipe, que foram adaptados às exigências dos algoritmos utilizados. Para isso, foram escolhidas as técnicas de pré-processamento *Information Gain* e *Correlation Based Feature Selection*. Ainda, foi definida uma terceira técnica manual para pré-processar os dados em consonância com as técnicas automatizadas. Por fim, foram escolhidos três algoritmos para a etapa de mineração de dados: o *Naive Bayes*, *Support Vector Machine* e *J48*. A mineração de dados foi feita a partir do software WEKA e, por fim, foram escolhidas as métricas para comparar os resultados, o que ocorreu posteriormente com a utilização da interface *Experimenter* do WEKA. Os resultados foram avaliados na etapa final, buscando validar a utilidade das informações obtidas por meio da mineração. Os resultados apontaram para uma tendência de evasão entre acadêmicos com tempo de curso inferior a três semestres. As taxas de acertos observadas alcançaram 94% para o *Naive Bayes*, 96% para o *Support Vector Machine* e 97% para o *J48*. O coeficiente de rendimento também recebeu destaque, sendo indicado pelo algoritmo *J48* a propensão à evasão por alunos com rendimento menor ou igual a nota 5.0, exceto quando esse aluno permanece no curso por mais de nove semestres. O algoritmo classificado como mais eficiente foi o *J48*.

4. Solução Proposta

O procedimento de descoberta do conhecimento ocorreu com base nos dados de alunos do curso de Sistemas de Informação, e foi dividido em cinco etapas principais, as quais são descritas a seguir:

- **Seleção dos Dados**

Primeiramente, o conjunto de dados foi obtido a partir das informações dos alunos disponíveis em um sistema de gestão acadêmica. A seleção de dados abrangeu quatro planilhas contendo os dados referentes a todos os alunos que mantiveram vínculo com o curso entre o segundo semestre de 2010 (início do curso) e o segundo semestre de 2019, fornecendo um total de 479 registros (alunos). Visando garantir a privacidade dos estudantes, seus códigos de identificação (ID) contidos em todas as planilhas utilizadas foram substituídos por um código fornecido pela coordenação do curso. Esse código foi inserido em todas as planilhas utilizadas nas etapas seguintes. As planilhas contém dados como: ano e semestre de ingresso, semestre de evasão, sexo, data de nascimento, naturalidade, endereço atual, ano e semestre de cada disciplina cursada, situação do

aluno (status de formado, evadido ou regular), média de notas por semestre/ano, percentual de frequência por semestre/ano.

- **Pré-processamento e Limpeza dos Dados**

A seguinte etapa teve por objetivo eliminar os dados considerados incompletos ou irrelevantes para o estudo, além de formar novos dados de importância com base na relação entre os já presentes. Foram considerados como evadidos os alunos que evadiram do curso por meio do abandono, transferência ou cancelamento de matrícula.

Alunos que foram selecionados para ingresso mas que não efetivaram a matrícula foram excluídos, uma vez que não haviam dados sobre esses alunos que pudessem ser utilizados neste estudo. Como resultado, teve-se 409 registros, sendo 67 registros referentes a alunos formados, 207 a evadidos e 135 a alunos regulares, totalizando 70 registros excluídos por falta de matrícula.

No contexto de integração, os dados de interesse para a proposta da mineração foram integrados em três grupos, cada qual criado com o objetivo de avaliar um dos três períodos definidos, 2010 a 2014, 2015 a 2019 e 2010 a 2019. O objetivo dessa separação foi investigar uma possível diminuição da evasão a partir de 2015, ano em que a instituição a que o curso pertence passou por uma significativa melhora em sua infraestrutura.

As informações referentes à situação do aluno em cada semestre foram padronizadas de modo que houvesse apenas três possibilidades para o aluno: regular, formado ou evadido. Dessa forma, o status de cancelamento de matrícula, transferência e abandono foram todos classificados como “evadido”.

Uma das suspeitas levantadas durante o estudo acerca da evasão escolar no ensino superior foi a possível relação entre a evasão escolar no curso de Sistemas de Informação com a distância entre a Instituição e a moradia do estudante. Nesse contexto, tomou-se por objetivo calcular essa distância aproximada em quilômetros, o que foi feito por meio de uma interface de programação de aplicações (API) do *Google Maps*.

Outra etapa importante do pré-processamento foi a discretização dos dados em conceitos, com o objetivo de padronizar dados muito diversos para sua utilização durante a mineração de dados. De maneira geral, as informações foram classificadas nos conceitos “MUITO BAIXO”, “BAIXO”, “MEDIO”, “ALTO” e “MUITO ALTO”, de acordo com a fórmula abaixo, responsável por estabelecer a amplitude dos intervalos que definiriam cada conceito.

$$\text{média} = (\text{máximo} - \text{mínimo}) / \text{número de conceitos}$$

As fórmulas utilizadas na discretização dessas distâncias estão contidas no Quadro 1, as quais também foram utilizadas na conceituação de outros atributos.

Quadro 1. Fórmulas utilizadas para conceituar o atributo x

| Conceito | Fórmula utilizada |
|-------------|--|
| MUITO BAIXO | $\text{mínimo} \geq x \leq (\text{mínimo} + \text{média})$ |
| BAIXO | $(\text{mínimo} + \text{média}) > x \leq (\text{mínimo} + 2 * \text{média})$ |
| MEDIO | posição mediana |
| ALTO | $(\text{máximo} - \text{média} * 2) > x \leq (\text{máximo} - \text{média})$ |
| MUITO ALTO | $(\text{máximo} - \text{média}) > x \leq (\text{máximo})$ |

Fonte: elaborado pelo autor (2020).

Nas fórmulas utilizadas, “x” pode assumir o valor referente à distância entre o aluno e a Instituição; a porcentagem de aprovações e reprovações e a porcentagem de presenças e ausências semestrais. Já as médias de notas semestrais foram convertidas em conceitos representados pelas letras do alfabeto latino “A”, “B”, “C” e “D”, utilizando como base os critérios de conversão definidos pela Instituição. É possível observar os intervalos utilizados para discretizar as médias dos alunos no Quadro 2.

Quadro 2. Intervalos definidos para conceituar as médias dos alunos

| Conceito | Intervalo |
|----------|-------------------------|
| A | média ≥ 9 |
| B | $7 \leq$ média ≤ 9 |
| C | $5 \leq$ média < 7 |
| D | média ≤ 5 |

Fonte: elaborado pelo autor (2020).

Ao final, foram obtidos 12 atributos, os quais são descritos no Quadro 3. Note que cada atributo foi coletado por aluno/semestre. Por exemplo, um aluno formado repetirá na base de dados por 9 vezes (número de semestres para conclusão do curso) e para cada semestre os valores dos atributos serão atualizados de acordo com o semestre cursado.

Quadro 3. Atributos selecionados para a mineração de dados

| Atributos | Descrição (por semestre) |
|------------------------|---|
| IDADE | Idade do aluno no semestre |
| SEXO | Gênero do aluno |
| DISTANCIA_CONCEITO | Conceito referente à distância entre a moradia do estudante e a Instituição. |
| MEDIA_NOTAS | Conceito referente a média de notas do estudante por semestre |
| PERCENTUAL_APROVAÇÕES | Conceito referente ao percentual de aprovação do estudante por semestre |
| PERCENTUAL_REPROVAÇÕES | Conceito referente ao percentual de reprovação do estudante por semestre |
| PERCENTUAL_PRESENÇA | Conceito referente ao percentual de presenças do estudante durante o semestre |
| PERCENTUAL_AUSENCIA | Conceito referente ao percentual de ausências do estudante durante o semestre |
| TRANCAMENTOS_TOTAL | Conceito referente à opção de trancamento total até o referido semestre |
| TRANCAMENTO_PARCIAL | Conceito referente ao número de trancamentos parciais por semestre |
| INGRESSO | Forma de ingresso na Instituição |
| SITUACAO | Situação do aluno por semestre |

Fonte: elaborado pelo autor (2020).

- **Mineração de Dados**

Terminada a seleção, pré-processamento e limpeza dos dados, estes foram inseridos na ferramenta WEKA com vistas em iniciar a etapa de mineração de dados, a qual foi

realizada com os três conjuntos de dados. A primeira etapa consistiu em minerar os dados referentes ao período de 2010 até 2014, enquanto a segunda avaliou os dados do intervalo entre 2015 e 2019. Ambas as etapas foram definidas de acordo com um interesse em comum, avaliar o impacto da melhora na infraestrutura da instituição nos índices de evasão escolares observadas. Por fim, a terceira etapa abrangeu a análise do conjunto total de dados, ou seja, todo o decorrer de tempo entre 2010 e 2019, com o objetivo de obter uma visão geral da problemática trabalhada. O algoritmo selecionado para efetuar a mineração de dados em ambos os conjuntos foi o J48, apontado em trabalhos relacionados como uma eficiente ferramenta para a proposta em questão (Gonçalves, Da Silva e Cortes, 2018).

Com o objetivo de balancear a quantidade de classes presentes no banco de dados exportado, optou-se por aplicar o filtro *Spread Subsample* da própria ferramenta Weka. Essa ação foi importante porque para cada semestre cursado o estudante recebe um *status*, até que o curso seja concluído ou o estudante tenha evadido. Nessa perspectiva, o filtro produziu uma amostra aleatória dos dados entre a classe mais rara e a mais comum, alterando a proporção de registros dos alunos de 1:1, podendo assim, contornar a discrepância entre os dados a serem minerados.

O primeiro conjunto de dados minerado pelo WEKA referente ao período entre 2010 e 2014 era composto por 669 registros de alunos com situação “REGULAR”, 64 registros como “EVADIDO” e 7 registros com a situação “FORMADO”. Após a aplicação do filtro *Spread Subsample*, os dados foram limitados a 7 alunos regulares, 7 formados e 7 em situação de evadido.

Em seguida, foi efetuada a etapa de mineração do conjunto de dados referente ao período de 2015 a 2019, que inicialmente contava com 1230 registros de alunos em situação “REGULAR”, 143 como “EVADIDO” e 60 como “FORMADO”. Com a aplicação do filtro *Spread Subsample*, o número de registros diminuiu para 60 alunos em cada uma das situações possíveis.

A terceira e última etapa da mineração foi a etapa relacionada aos dados pertencentes ao período de 2010 a 2019, composta inicialmente por 1899 registros de alunos com situação “REGULAR”, 207 com “EVADIDO” e 67 com “FORMADO”. As três situações diminuíram para 67 pela aplicação do filtro *Spread Subsample*, com o conjunto de dados final sendo formado por 201 instâncias.

5 Discussão dos resultados

É notável a baixa incidência de alunos formados no período entre 2010 e 2014, fato esse que pode ser explicado porque o período entre o segundo semestre de 2010 e o segundo semestre de 2014 corresponde a 9 semestres, tempo mínimo para a conclusão do curso.

No segundo período, entre 2015 a 2019, observa-se um aumento expressivo tanto no número de formados como de evadidos, o que pode ser esperado considerando o maior tempo de existência do curso, o qual possibilitou a conclusão do curso por alunos que não haviam se formado no período estudado anteriormente, bem como a conclusão por novos alunos. Por fim, o período de 2010 a 2019 apresentou a totalidade dos dados registrados desde a criação do curso até o final de 2019.

É destacável que, dentre as três análises realizadas, a que apresentou os melhores resultados de classificação foi a referente ao período 2015-2019, alcançando a classificação correta de 70,56% das instâncias analisadas. A seguir, o segundo melhor resultado foi o referente a 2010 até 2019, o qual englobou todo o período analisado, apresentando 68,16% de instâncias corretamente classificadas. Por último, o período entre 2010 a 2014 apresentou resultados inferiores as demais avaliações, com 42,90% de instâncias corretamente classificadas. Como já foi mencionado, este conjunto teve

um número menor de instâncias devido ao tempo de existência do curso neste período.

Notou-se que, nos três períodos estudados, os atributos mais relevantes para a classificação no algoritmo foram referentes à porcentagem de aprovação e a idade do aluno. Alunos com baixa taxa de aprovação ou com média de notas muito baixa em um referido semestre são mais propensos a evadir. Além disso, identificou-se que mesmo com uma taxa de aprovação alta mas com idade menor de 19 anos também são propensos a evadir. Após análise dos dados, pode-se perceber que muitos dos estudantes que abandonam no início do curso, 1º ou 2º semestre são ingressantes por processo seletivo (Vestibular/SISU), sendo possivelmente alunos que saíram recentemente do ensino médio e possam ter dúvidas do curso que pretendem cursar.

5. Considerações Finais

Este estudo teve objetivo investigar as causas da evasão escolar no ensino superior de um curso de Sistemas de Informação utilizando a mineração de dados como principal ferramenta. Um dos objetivos propostos foi analisar a relação entre os dados socioeconômicos dos alunos com a tendência à evasão, o que não foi possível devido ao não fornecimento dos dados solicitados pelo órgão responsável. Os resultados obtidos foram considerados satisfatórios, tendo em vista a baixa quantidade de atributos, restritos ao desempenho acadêmico.

Em uma visão geral dos casos, o algoritmo acabou por classificar atributos relacionados a média, aprovação, idade e presença do aluno, ou seja, atributos ligados diretamente ao estudante e seu desempenho acadêmico, que se mostraram fortes indicadores de evasão. O estudo também apontou que a distância entre a moradia do estudante e a Instituição não atuou como um fator relevante para a evasão neste estudo. Além disso, pode-se observar que a melhoria na infraestrutura da instituição também não impactou nas taxas de evasão, pelo contrário, a taxa de evasão no conjunto de dados de 2010-2014 é menor em relação ao conjunto de 2015-2019.

Para trabalhos futuros, recomenda-se a obtenção de mais atributos que possam ampliar a extração de conhecimento desses conjuntos de dados, bem como dados socioeconômico e psicológicos dos alunos, que não foram fornecidos para este estudo. Essa metodologia também pode ser aplicada em dados de outros cursos da instituição, bem como de outros centros de ensino, com vista em expandir os conhecimentos acerca da evasão escolar no ensino superior.

Referências Bibliográficas

- DO AMARAL, Daniela. P. ; DE OLIVEIRA, F. B.. O Prouni e a conclusão do ensino superior: novas trajetórias pessoais e profissionais dos egressos. **Ensaio: avaliação e políticas públicas em educação**, v. 19, n. 73, p. 861-890, 2011.
- AMARAL, F.. **Aprenda mineração de dados: teoria e prática**. Alta Books Editora, 2016.
- DO AMARAL, J. B.. Evasão discente no ensino superior: estudo de caso no Instituto Federal de Educação, Ciência e Tecnologia do Ceará (Campus Sobral). 2013.
- ANDRIOLA, W. B.. Doze motivos favoráveis à adoção do Exame Nacional do Ensino Médio (ENEM) pelas instituições federais de ensino superior (IFES). 2011.
- APRILE, M. R.; BARONE, R. E. M.. Educação superior: políticas públicas para inclusão social. **Revista@mbienteeducação**, v. 2, n. 1, p. 39-55, 2018.
- DALRYMPLE, D. J.; PARSONS, L. J. **Introdução à administração de marketing**. Ltc, 2003.
- DA SILVA, A. M. P.. Processos de ensino-aprendizagem na era digital. 2009.

- DA SILVA, T. L. C. et al. Análise em Big Data e um Estudo de Caso utilizando Ambientes de Computação em Nuvem. 2013.
- DA SILVA, W. V. et al. Avaliação da escolha de um fornecedor sob condição de riscos a partir do método de árvore de decisão. **REGE Revista de Gestão**, v. 15, n. 3, p. 77-94, 2008.
- DE ALMEIDA, T. L. ; KAPPEL, M. A. A.. Aplicação de Técnicas de Aprendizado de Máquina para Predição de Risco de Evasão Escolar em Instituições Públicas de Ensino Superior no Brasil. **Revista Brasileira de Informática na Educação**, v. 28, p. 838-863, 2020.
- DOURADO, L. F.. Reforma do Estado e as políticas para a educação superior no Brasil nos anos 90. **Educação & Sociedade**, v. 23, n. 80, p. 234-252, 2002.
- FAVERO, R. V. M.. Dialogar ou evadir: Eis a questão!: um estudo sobre a permanência e a evasão na educação a distância. 2006.
- GAIOSO, N. P. De L.. O fenômeno da evasão escolar na educação superior no Brasil. **Brasília, DF: Universidade Católica de Brasília**, p. 20, 2005.
- GOMES, A. A.. Evasão e Evadidos: O discurso dos alunos sobre evasão escolar nos cursos de licenciatura. *Nuances: estudos sobre Educação*, v. 5, n. 5, 1999.
- GONÇALVES, T. C.; DA SILVA, J. C.; CORTES, O. A. C.. Técnicas de mineração de dados: um estudo de caso da evasão no ensino superior do Instituto Federal do Maranhão. **Revista Brasileira de Computação Aplicada**, v. 10, n. 3, p. 11-20, 2018.
- JACOB, C. A. R.. A evasão escolar e a construção do sujeito/profissional em curso de Ciências Econômicas. 2000.
- JOHANN, C. C. et al. Evasão escolar no Instituto Federal Sul-Rio-Grandense: um estudo de caso no campus Passo Fundo. 2012.
- KIRBY, P. G. et al. Adults returning to school: The impact on family and work. **The Journal of Psychology**, v. 138, n. 1, p. 65-76, 2004.
- LEMONS, A. C. M.. Uma visão psicopedagógica do bullying escolar. **Revista Psicopedagogia**, v. 24, n. 73, p. 68-75, 2007.
- MANHÃES, L. M. B. et al. Identificação dos fatores que influenciam a evasão em cursos de graduação através de sistemas baseados em mineração de dados: Uma abordagem quantitativa. **Anais do VIII Simpósio Brasileiro de Sistemas de Informação, São Paulo**, 2012.
- MANTOAN, M. T. E.; PRIETO, R. G.. Inclusão escolar: o que é. **Por quê**, p. 12, 2003.
- NEGRINE, A.. Aprendizagem e desenvolvimento infantil: perspectivas psicopedagógicas. Porto Alegre: Prodil, v. 2, 1994.
- NOGUEIRA, C. M. M. et al. Promessas e limites: o Sisu e sua implementação na Universidade Federal de Minas Gerais. **Educação em Revista**, v. 33, 2017.
- PAZ, F. ; CAZELLA, S.. Identificando o perfil de evasão de alunos de graduação através da Mineração de dados Educacionais: um estudo de caso de uma Universidade Comunitária. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2017. p. 624.
- PEÇANHA, Plataforma Nilo. PNP 2019 (ano base 2018). 2019. Disponível em: <<http://plataformanilopecanha.mec.gov.br/2019.html>>. Acesso em: 11 out. 2020.
- PEÇANHA, Plataforma Nilo. PNP 2020 (ano base 2019). 2020. Disponível em: <<http://plataformanilopecanha.mec.gov.br/2020.html>>. Acesso em: 11 out. 2020.
- PIMENTA, S. G.. Orientação e decisão: Estudo crítico da situação do Brasil. 1981.
- SILVA FILHO, Roberto Leal Lobo et al. A evasão no ensino superior brasileiro. **Cadernos de pesquisa**, v. 37, n. 132, p. 641-659, 2007.
- SILVA, N. L. P. ; DESSEN, M. A.. Crianças com síndrome de Down e suas interações familiares. **Psicologia: reflexão e crítica**, v. 16, n. 3, p. 503-514, 2003.