

# Uma análise dos algoritmos de classificação com base na evasão dos estudantes dos cursos técnicos integrados ao Ensino Médio do Campus Ceres do IF Goiano

Ana Luiza G. de Souza<sup>1</sup>, Adriano H. Braga<sup>1</sup>

<sup>1</sup>Instituto Federal Goiano – Campus Ceres (IF GOIANO)  
Caixa Postal 51 – 76.300-000 – Ceres – GO – Brasil

gs.analuiza@gmail.com, adriano.braga@ifgoiano.edu.br

**Abstract.** *This work analyze the data mining algorithms that best fit the conditions and data of the integrated technical courses to the high school of the Ceres Campus of IF Goiano. Using the KDD methodology, it was possible to work with a sample of 1,478 enrollments. The Weka tool, compare the following algorithms: J48, Naive Bayes, Logistic, Multilayer Perceptron, IBk, and LibSVM. Among them, LibSVM proved to be the best predictor, achieving the best result for three of the five metrics considered in the study.*

**Resumo.** *Objetivou-se a análise de algoritmos de mineração de dados que melhor se adequem às condições e dados dos cursos técnicos integrados ao Ensino Médio do Campus Ceres do IF Goiano. A partir da metodologia KDD, foi possível trabalhar com uma amostra de 1.478 matrículas. Com a ferramenta Weka, pôde ser feita a comparação dos algoritmos: J48, Naive Bayes, Logistic, Multilayer Perceptron, IBk e LibSVM. Dentre eles, o LibSVM se mostrou o melhor preditor, alcançando o melhor resultado de três das cinco métricas consideradas no estudo.*

## 1. Introdução

O decreto 7.566/1909 criou 19 Escolas Aprendizes Artífices com a finalidade de oferecer ensino profissionalizante, que posteriormente foram transformadas na década de 1930, para Liceus Industriais (FERNANDES, 2013). Já pela Lei Orgânica do Ensino Industrial em 1942, as mesmas começaram a se chamar Escolas Industriais e Técnicas e em 1959, Escolas Técnicas Federais (VIDOR *et al.*, 2011). Somente em 2008, com a criação da Lei 11.892/08 escolas como: Escolas Vinculadas às Universidades Federais, Centros Federais de Educação e as Escolas Agrotécnicas Federais se transformaram em Rede Federal de Educação Profissional, Científica e Tecnológica (RFEPCT) (OTRANTO, 2013) e ofertando prioritariamente cursos técnicos integrados ao Ensino Médio.

Apesar das melhorias e evoluções passadas pela educação no Brasil, as instituições de ensino ainda enfrentam desafios no caminho para o ensino de qualidade, uma delas é a evasão escolar (SILVA; NUNES, 2015). Conforme dados apresentados pela Plataforma Nilo Peçanha - PNP (2020), 11,9% dos estudantes do centro-oeste abandonaram o curso técnico integrado ao Ensino Médio onde realizaram suas matrículas no ano de 2018 e 10,3% no ano de 2019. Ao analisar os dados do ensino médio regular, nos anos de 2018 e 2019, a região centro-oeste registrou uma taxa de abandono de 5,2% e 4,8%, respectivamente (INEP, 2020). Embora tenha tido redução

nas taxas de ambos níveis de estudo apresentados, o problema da evasão escolar continua bastante significativo (BRASIL, 2020).

Essa prática acontece quando o aluno antes de concluir o curso em que está matriculado o abandona, perdendo o vínculo com a instituição (PNP, 2020). Inúmeros fatores podem levar estudantes a evadirem. Variáveis como: tempo na escola, reprovações, localização da escola, condições socioeconômicas, notas baixas, entre outras, podem ser consideradas determinantes para evasão (RUMBERGER; LIM, 2008; SILVA FILHO; ARAÚJO, 2017).

Tendo isso em vista, entender os principais motivos que podem levar os alunos a abandonarem seus estudos se faz cada vez mais necessário. Pois, quando se entende o padrão da evasão escolar, as políticas públicas podem focar naqueles estudantes que se encaixam nesse agrupamento de risco desde início de sua jornada, a fim de reverter os índices (BEZERRA *et al.*, 2016). Uma das formas de se realizar isso, é através da mineração de dados educacionais.

Sendo assim, este trabalho busca, através de uma abordagem quantitativa e exploratória, analisar diversos algoritmos de classificação já conhecidos em mineração de dados, no âmbito da evasão escolar a partir de informações de estudantes dos cursos Técnicos Integrados ao Ensino Médio do Campus Ceres. A fim de escolher um que melhor possa servir de auxílio na diminuição dos índices de evasão da Instituição.

Para isso, são apresentados na seção seguinte alguns trabalhos relacionados ao presente estudo, seguido dos materiais e métodos utilizados para concepção da pesquisa. Sucessivamente, na seção 3 há a exploração dos resultados e discussões a partir dos mesmos e na seção 4, encontram-se as considerações finais acerca do estudo.

## 2. Trabalhos Relacionados

Bitencourt e Ferrero (2019) aplicando mineração de dados com os algoritmos de árvore de decisão *Random Forest*, *Decision Trees*, além do algoritmo *XGBoost*, propõem o uso de descoberta de conhecimento em bases de dados para extrair informações de 772 instâncias de alunos de cursos técnicos que realizaram matrícula entre 01/2018 e 07/2019 no Instituto Federal de Santa Catarina. Ao final dos experimentos foi possível recuperar 25% dos potenciais evasores, com precisão de 86%.

Já em um panorama de graduação, Brito Junior *et al.* (2019) procuraram identificar fatores que pudessem influenciar a evasão no curso de Sistemas de Informação. Com uma amostra de 196 instâncias, os autores executaram os algoritmos *K-Means* e *J48*. Após realização de três experimentos, com precisão de aproximadamente 90 e 95%, constataram quatro diferentes perfis de alunos tendenciosos a evasão, sendo eles: escolares que reprovaram nas quatro disciplinas base do curso; escolares com idade de 26 anos ou mais; escolares que excederam o tempo comum de conclusão do curso e não participam de nenhum projeto; e escolares que reprovaram na disciplina de algoritmos, depois de reprovar na disciplina de introdução a informática.

Outro estudo sobre evasão foi feito por Tasnim, Paul e Sattar (2019). As informações foram obtidas em um repositório de conjuntos de dados denominado *UCI Machine Learning Repository*. Os autores aplicaram algoritmos de máquina de vetores de suporte (do inglês, *Support Vector Machine*, SVM), redes bayesianas e regressão

logística. Após separação do *dataset* em dois conjuntos, A e B, conseguiram alcançar uma precisão de 97% no conjunto A e 100% no conjunto B, ambos com o método de mineração de dados SVM.

Em sua dissertação de mestrado, Bezerra (2019) utilizou a metodologia CRISP-DM para aplicação de mineração de dados. Com algoritmos de árvore de decisão e redes bayesianas, e com dados acadêmicos e socioeconômicos, do sistema acadêmico do Instituto Federal de Rondônia - Campus Ji-Paraná, pôde identificar dois perfis de discentes com chances de evadir. Entre eles, alunos do primeiro ano letivo, dos turnos matutino e vespertino e alunos que utilizam ônibus coletivo como transporte escolar, possuem baixa renda familiar e que recebem auxílio estudantil.

Em um estudo comparativo de classificadores de árvore de decisão, SVM, regressão logística, redes neurais e *k*-ésimo vizinho mais próximo, na tarefa de predição de dados, Ramos *et al.* (2018), usaram 18.094 instâncias do ensino a distância (EaD) da graduação de licenciaturas em Pedagogia e Ciências Biológicas da Universidade de Pernambuco (UPE). Com atributos relacionados à interação dos alunos com a plataforma utilizada no EaD e manuseio da ferramenta *RStudio*, conseguiram alcançar 89,6% de acurácia no preditor de *k*-ésimo vizinho mais próximo e uma classificação correta de 71,9% dos estudantes que evadem com algoritmos de regressão logística.

Este trabalho se difere dos demais apresentados, por além de ter como público-alvo especificamente estudantes de cursos técnicos integrados ao Ensino Médio, utiliza de um número maior de atributos, incluindo até mesmo informações anteriores ao ingresso do estudante à instituição, como: nota do Processo Seletivo (PS), reserva de vagas e escola de origem.

### 3. Materiais e Métodos

O presente estudo adota a metodologia KDD (*Knowledge Discovery in Databases*) para nortear a predição dos dados. Escolhido por satisfazer as características de retratar com precisão o conteúdo do banco de dados, gerar resultados interessantes e ser eficiente (FRAWLEY; PIATETSKY-SHAPIRO; MATHEUS, 1993). Essa técnica contempla as etapas principais de seleção, pré-processamento, transformação, mineração dos dados e avaliação.

A etapa de seleção dos dados se deu a partir da obtenção da base de dados do sistema acadêmico e do PS do Campus Ceres do IF Goiano. A partir disso, foi preciso desempenhar um estudo acerca dos dados para entender como filtrar as variáveis desejadas. Quanto ao sistema acadêmico que possui uma enorme quantidade de tabelas e relacionamentos que não são necessários para a pesquisa, pelo fato de ser adaptado para uso de instituições diversas. Desta forma, foram criadas *Views* para simplificar a integração das informações contidas na base do PS e assim uma equidade de dados como: nome, ano de ingresso e curso, dos alunos devidamente matriculados.

A base de dados do sistema Acadêmico do IF Goiano foi adquirida diretamente com a Instituição de forma anonimizada, já que se trata de dados sensíveis. Estes foram disponibilizados por meio de exportação em arquivos na linguagem SQL (*Structured Query Language*), única e exclusivamente para fins de pesquisa. Por outro lado, as informações do PS foram obtidas através dos resultados dos mesmos no site oficial da instituição e precisaram ser convertidos de arquivo PDF (*Portable Document Format*) para CSV (*Character-separated Values*).

Os dados dos alunos contemplados na pesquisa foram os do curso técnico integrado ao ensino médio do Campus Ceres, do IF Goiano, que se matricularam entre os anos de 2015 e 2019. O *dataset* final foi extraído no dia 01 de outubro de 2020, totalizando 1.769 alunos, que assim passaram por um processo de anonimização a fim de preservar a identidade dos envolvidos, removendo todas informações sensíveis. Os atributos advindos do *dataset* estão elencados na Tabela 1.

Para facilitar o pré-processamento, os registros foram inseridos em um banco de dados com o Sistema de Gerenciamento de Banco de Dados (SGBD) *MySQL* 8.0<sup>1</sup>. Assim, foi executada a limpeza, remoção de dados repetidos, *outliers* ou fora do escopo da pesquisa. Outrossim, foram retirados os registros referentes às disciplinas em curso, já que as mesmas não possuíam notas.

**Tabela 1. Atributos selecionados para a fase de transformação dos dados**

<b>Atributo</b>	<b>Tipo de dado</b>	<b>Possíveis valores</b>
Ano do último registro	Numérico	Valores de 2015 a 2020.
Cidade	Catégorico	90 valores: Ceres, Rialma, entre outros.
Coefficiente de rendimento	Numérico	Valores decimais de 0,0 a 10,0.
Cotas	Catégorico	10 valores: AC, RI, RS, entre outros.
Curso	Catégorico	3 valores: Agropecuária, Informática, entre outros.
Data de nascimento	Data	Valores no formato AAAA-MM-DD.
Escola de origem	Catégorico	3 valores: pública, estadual e privada.
Nota das disciplinas	Numérico	Valores decimais de 0,0 a 10,0.
Nota de ingresso	Numérico	Valores inteiros de 0 a 60.
Quantidade de falta	Numérico	Valores inteiros e positivos.
Quantidade de reprovação	Numérico	Valores inteiros e positivos.
Série	Numérico	Valores inteiros 1 a 3.
Sexo	Catégorico	2 valores: F e M.
Situação de matrícula	Catégorico	8 valores: concluído, evasão, entre outros.

Foram efetuadas transformações nos dados para adequá-los aos algoritmos. O cálculo da idade se deu a partir da data de nascimento. O atributo “cidade” foi usado para verificar se o estudante era originário da mesma localidade da Instituição. Dessa maneira, a informação foi transformada em valores booleanos: “sim” para estudantes da mesma cidade e “não”, para estudantes residentes de diferentes cidades.

Os cursos, por sua vez, foram agrupados em 3 modalidades. “AGRO” contendo o curso de Técnico em Agropecuária Integrado ao Ensino Médio, “INFO” abrangendo os cursos de Técnico em Informática Integrado ao Ensino Médio e Técnico em Informática para Internet Integrado ao Ensino Médio e “MA” relativo ao curso de Técnico em Meio Ambiente Integrado ao Ensino Médio.

Quanto às notas do ensino médio, foram calculadas médias aritméticas por disciplina de núcleo comum cursada para cada aluno, independente da série. Já para as disciplinas de nível técnico, a média foi feita por série independente da disciplina, ou

<sup>1</sup> <https://dev.mysql.com/downloads/mysql/>

seja, cada série cursada possuía a média das notas obtidas nas disciplinas técnicas da mesma. Além disso, à variável principal, situação de matrícula, foi atribuído “sim” para aqueles que evadiram e “não” para concludentes, matriculados ou que se transferiram. Devido ao detalhamento das notas, ao final, totalizaram 28 atributos.

Na fase de mineração de dados, foram utilizados os algoritmos para indução de modelos preditivos: J48, *Naive Bayes*, *Logistic*, *Multilayer Perceptron*, *IBk* e *LibSVM*. Estes podem ser encontrados na ferramenta Weka<sup>2</sup> (*Waikato Environment for Knowledge Analysis*), desenvolvida pela Universidade de Waikato na Nova Zelândia, contendo bibliotecas de algoritmos e suporte no processo de mineração de dados (HALL *et al.*, 2009).

Com apenas 79 evasões em um total de 1.400 matrículas, observou-se que a amostra estava desbalanceada. Um dos motivos para esse fenômeno é a falta de atualização dos registros por parte da instituição gerenciadora. Por essa razão, optou-se pela utilização da técnica SMOTE (*Synthetic Minority Oversampling Technique*) (CHAWLA *et al.*, 2002) de balanceamento para equilibrar artificialmente um conjunto de dados, já que a taxa de evasores se revelou consideravelmente menor que a taxa de concluintes e matriculados.

Algoritmos de mineração de dados normalmente trazem consigo parâmetros que ajudam a construir e moldar seus resultados. Estes, por sua vez, não possuem um padrão próprio e correto de valores para cada *dataset*. Uma das estratégias usadas para a definição desses parâmetros baseia-se no teste de diferentes combinações de valores até que se chegue no que mais se encaixa aos dados e por conseguinte, gera bons resultados.

Partindo desse pressuposto, as configurações pré-estabelecidas pelo Weka foram as que obtiveram os melhores resultados para os algoritmos J48, *LibSVM* e *Logistic*. No preditor *Naive Bayes* foi usado o estimador de *kernel* para atributos numéricos em vez de uma distribuição normal, ativando o parâmetro use *Kernel Estimator*. Já no *Multilayer Perceptron* o número de épocas para treinar, *Training Time*, foi aumentado para 750. Por fim, no *IBk*, a função usada para calcular a distância das instâncias foi a *Filtered Distance*, pois esta aplica o filtro antes de chamar a função de distância.

Para cada modelo, foi adotada a estratégia de validação cruzada dividida em 10 subconjuntos. Posteriormente, foi feita a comparação do desempenho dos algoritmos. Como forma de avaliação, utilizou-se as métricas: acurácia, que diz respeito a taxa de acerto comparado ao desempenho geral; pontuação F1 (*F1-Score*), que é a média harmônica entre precisão e revocação e permite qualificar o modelo de um modo geral; Curva Característica de Operação do Receptor (Curva ROC) que avalia o desempenho do modelo; taxas de falsos positivos e verdadeiros positivos que identificam a frequência das classes no modelo.

Além disso, os resultados encontrados nas matrizes de confusão também foram utilizados como métrica de avaliação dos algoritmos. Esta é composta pela quantidade de erros e acertos de classificação do sistema, de modo a visualizar as amostras confundidas pelo mesmo (SOARES *et al.*, 2020). Para tanto, temos que: Verdadeiro Positivo (VP) é a quantidade de alunos que evadiram e o modelo classificou como

---

<sup>2</sup> <https://www.cs.waikato.ac.nz/ml/weka/>

evadido, Falso Positivo (FP) é a quantidade de alunos que não evadiram, mas o modelo classificou como evadido, Verdadeiro Negativo (VN) é a quantidade de alunos que não evadiram e o modelo classificou como não evadido e Falso Negativo (FN) é a quantidade de alunos que evadiram, mas o modelo classificou como não evadido.

#### 4. Resultados e Discussões

A amostra final para a execução dos algoritmos foi de 1.478 matrículas, possuindo dados de estudantes dos cursos de Agropecuária, Informática, Informática para Internet e Meio Ambiente, todos integrados ao ensino médio ofertados no Instituto Federal Goiano - Campus Ceres. Dessas matrículas, 41% são de estudantes do sexo feminino e 59% do masculino. Em sua maioria, os estudantes não são da cidade de Ceres e o ensino fundamental foi concluído em escola estadual. A média das notas no processo seletivo para entrada na Instituição, do coeficiente de rendimento e das disciplinas técnicas são de aproximadamente 31 de 60, 5,92 de 10 e 7,56 de 10, respectivamente. O número de estudantes evadidos da amostra foi de 158 e está presente em todos os cursos (Figura 1).

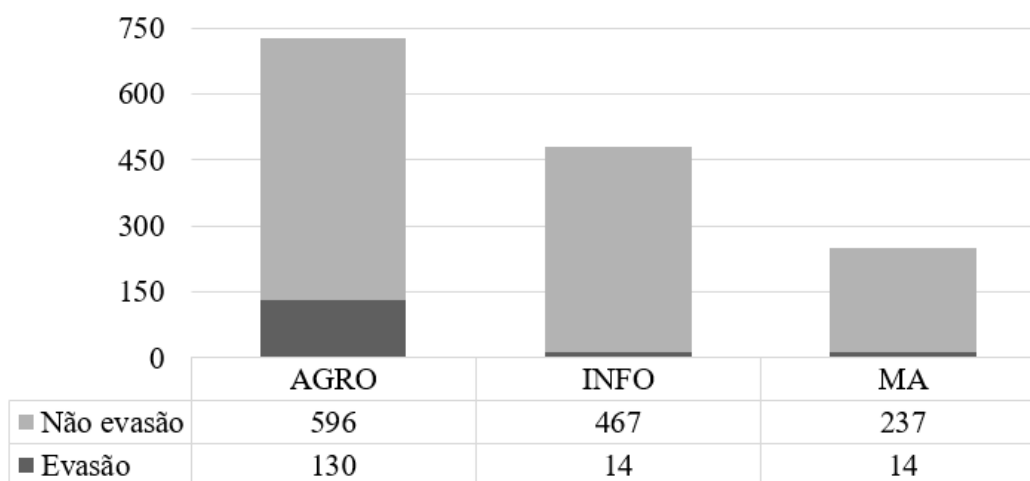


Figura 1. Quantitativo de matrículas por tipo de curso nos anos de 2015 a 2020

Como métrica de avaliação dos resultados, as matrizes de confusão foram calculadas e estão apresentadas na Figura 2. Levando-se em consideração esses valores, a melhor taxa de recuperação (VP) foi de 82,9% com o modelo *Naive Bayes*, que se sobressaiu em relação aos outros modelos. No entanto, ele precisou errar quase a metade (49%) das previsões a respeito do valor real (FP) para chegar a essa taxa. Em contrapartida, o classificador *LibSVM* obteve a menor taxa de erro daqueles que não evadiram (12,7%), mas conseguiu prever apenas 56,3% sobre aqueles que realmente evadiram.

Na Tabela 2 são apresentados os resultados experimentais dos modelos preditivos para o conjunto de dados utilizado. Pode-se observar que a acurácia dos classificadores se manteve acima de 92%, com exceção do algoritmo de redes bayesianas que obteve 89,6%. Desta forma, o *LibSVM* pode ser considerado como o modelo de melhor desempenho, pois teve o melhor resultado de três das cinco métricas aferidas no estudo.

Marquez-Vera, Morales e Soto (2013) obtiveram valores de desempenho semelhantes, executando os algoritmos *JRip*, *NNge*, *OneR*, *Prism*, *Ridor*, *ADTree*, *J48*,

*RandomTree*, *REPTree* e *SimpleCart*. A maior acurácia foi de 93% e a amostra utilizada era de apenas 670 alunos do primeiro ano do ensino médio. Demais trabalhos, assim como de Couto e Santana (2017), apesar de possuírem métodos similares, as variáveis não são compatíveis com a deste trabalho.

<b>Modelo</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	VP	FP
Não Evadiu	FN	VN

<b>J48</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	102	42
Não Evadiu	56	1278

<b>IBk</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	118	45
Não Evadiu	40	1275

<b>LibSVM</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	89	13
Não Evadiu	69	1307

<b>Logistic</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	103	55
Não Evadiu	50	1270

<b>Multilayer Perceptron</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	104	40
Não Evadiu	54	1280

<b>Naive Bayes</b>	<b>Valor real</b>	
<b>Previsto</b>	Evadiu	Não Evadiu
Evadiu	131	126
Não Evadiu	27	1194

**Figura 2. Matrizes de confusão de cada algoritmo executado na pesquisa**

Já Maria, Damiani e Pereira (2016) desenvolveram um modelo preditivo com redes bayesianas para classificar a evasão em cursos técnicos não integrados ao ensino médio, obtendo 85,6% de acurácia. O modelo não foi escolhido por meio de comparações com outros preditores e, portanto, não foi possível saber o desempenho de outros modelos com o mesmo conjunto de dados utilizado.

**Tabela 2. Resultados de classificação das métricas consideradas no estudo**

<b>Algoritmo</b>	<b>Acurácia</b>	<b>F1-Score</b>	<b>Taxa VP</b>	<b>Taxa FP</b>	<b>Curva ROC</b>
<i>IBk</i>	0,942	<b>0,943</b>	0,942	0,230	0,867
J48	0,934	0,932	0,934	0,320	0,923
<i>LibSVM</i>	<b>0,945</b>	0,939	<b>0,945</b>	<b>0,391</b>	0,905
<i>Logistic</i>	0,929	0,928	0,929	0,315	0,940
<i>Multilayer Perceptron</i>	0,936	0,935	0,936	0,308	<b>0,949</b>
<i>Naive Bayes</i>	0,896	0,907	0,896	0,163	0,930

A metodologia deste estudo vai de encontro com os resultados obtidos na pesquisa de Colpo, Primo, Pernas e Cechinel (2020). A começar pelas técnicas de mineração de dados com a escolha dos algoritmos. Além da utilização de dados até então pouco explorados na literatura do Congresso Brasileiro de Informática na Educação. Ademais, os resultados apresentados possuem grande relevância a alguns disponíveis na literatura, pois analisou maior quantitativo de dados comparando diversos algoritmos.

## 5. Considerações finais

Um dos diferenciais da RFEPCT é a oferta dos cursos técnicos integrados ao Ensino Médio. Desta forma, é importante que sejam feitas pesquisas para análise destes índices, principalmente pelo fato que esta oferta de cursos possui menos de 15 anos de concepção.

Com o intuito de contribuir na compreensão da evasão deste público-alvo em específico, este trabalho pode ser considerado inovador justamente pela população que foi objeto de estudo e, além disso, por ter entre as suas variáveis dados dos estudantes antes do ingresso na Instituição, ou seja, levando em consideração a nota obtida por meio da realização de uma verificação de aprendizagem como seleção para o ingresso ao curso.

Conclui-se que, os modelos preditivos aqui comparados são de total relevância para que juntamente de políticas de acompanhamento da Instituição possa ser possível diminuir os números de evasão. O melhor dos casos analisados, o algoritmo *LibSVM*, pode auxiliar na predição com baixa taxa de falsos positivos. O que leva a considerar que dependendo do foco da administração e recursos disponíveis, é mais vantajoso ter o modelo que não as levam a aplicar recursos em alunos que não vão evadir.

Apesar do objetivo deste trabalho não ser a indicação de quais estudantes evadem, foi verificado uma maior evasão nos estudantes, em ordem de significância, com: coeficiente de rendimento menor que 5,6; estudantes no primeiro ano de curso; estudantes com mais de 18 anos; estudantes do curso técnico em agropecuária integrado ao ensino médio; estudantes faltosos; e estudantes que obtiveram nota menor ou igual a 25 na prova do processo seletivo. Carecendo de maiores análises quanto a variáveis socioeconômicas e extrínsecas, estratificando por grupos de estudantes.

Para trabalhos futuros, sugere-se a adoção de métricas ainda não investigadas e a inclusão de outras variáveis que não foram de escopo deste estudo. A exemplo disso seriam as atividades realizadas pelos estudantes durante o curso na Instituição e os fatores externos a ele, que muitas vezes podem influenciar a decisão dos mesmos sobre sua permanência.

É possível ainda utilizar a mesma metodologia empregada neste trabalho para a análise dos dados dos cursos superiores da Instituição, comparando assim com a maioria dos estudos citados como revisão e ainda ampliar a amostra com a inclusão de dados de outros campi do IF Goiano ou da RFEPCT.

## Referências

- Bezerra, C., Scholz, R., Adeodato, P., Lucas, T. e Ataíde, I. (2016). Evasão Escolar: Aplicando Mineração de Dados para Identificar Variáveis Relevantes. In *Anais do Simpósio Brasileiro de Informática na Educação (CBIE 2016)*, p. 1096.
- Bezerra, J. H. da S. (2019). *Análise da Evasão Escolar do Instituto Federal de Rondônia – Campus Ji-Paraná – Utilizando Técnicas de Mineração de Dados*. Instituto Politécnico do Porto.
- Bitencourt, P. B. de e Ferrero, C. (2019). Predição de Risco de Evasão de Alunos Usando Métodos de Aprendizado de Máquina em Cursos Técnicos. In *Anais dos*



*Workshops do VIII Congresso Brasileiro de Informática na Educação (CBIE 2019), p. 149.*

- Brasil (2020). Evasão Escolar ou Abandono Escolar. Disponível em: <https://www.gov.br/mdh/pt-br/navegue-por-temas/crianca-e-adolescente/dados-e-indicadores/evasao-escolar-ou-abandono-escolar>. Acesso em: 17 set. 2020.
- Brito Junior, I., Rabelo, H., Naschold, A. M. C., Ferreira, A. M., Burlamaqui, A., Rabelo, D. S. d. S., e Valentim, R. (2019). Uso de Mineração de Dados Educacionais para a classificação e identificação de perfis de Evasão de graduandos em Sistemas de Informação. In *Anais dos Workshops do VIII Congresso Brasileiro de Informática na Educação (CBIE 2019)*, p. 159.
- Chawla, N. V., Bowyer, K. W., Hall, L. O. e Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357.
- Colpo, M. P., Primo, T. T., Pernas, A. M. e Cechinel, C. (2020). Mineração de Dados Educacionais na Previsão de Evasão: uma RSL sob a Perspectiva do Congresso Brasileiro de Informática na Educação. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)*, p. 1102–1111.
- Couto, D. da C. do e Santana, Á. L. de (2017). Mineração de dados educacionais aplicada à identificação de variáveis associadas à evasão e retenção. *Congresso sobre Tecnologias na Educação*, p. 333–344.
- Fernandes, M. R. da S. (2013). *O processo de verticalização da educação profissional e tecnológica e suas implicações na qualidade do trabalho dos docentes do Câmpus São Vicente do Sul do Instituto Federal Farroupilha*. Dissertação, Universidade Federal Rural do Rio de Janeiro.
- Frawley, W. J., Piatetsky-Shapiro, G. e Matheus, C. J. (1993). Knowledge Discovery in Databases: An Overview. *AI Magazine*, v. 13, n. 3, p. 57–70.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., e Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, v. 11, n. 1, p. 10–18.
- INEP (2020). Taxas de Rendimento. Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/taxas-de-rendimento>. Acesso em: 1 jul. 2021.
- Maria, W., Damiani, J. L. e Pereira, M. (2016). Rede Bayesiana para previsão de Evasão Escolar. In *Anais dos Workshops do V Congresso Brasileiro de Informática na Educação (CBIE 2016)*, p. 920.
- Marquez-Vera, C., Morales, C. R. e Soto, S. V. (2013). Predicting School Failure and Dropout by Using Data Mining Techniques. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, v. 8, n. 1, p. 7–14.
- Otranto, C. R. (2013). A reforma da educação profissional e seus reflexos na educação superior. *Revista Temas em Educação*, v. 22, n. 2, p. 122–135.
- PNP (2020). Plataforma Nilo Peçanha 2020. Disponível em: <http://plataformanilopecanha.mec.gov.br/2020.html>. Acesso em: 1 jul. 2021.

- Ramos, J. L. C., Silva, J., Prado, L., Gomes, A. e Rodrigues, R. (2018). Um estudo comparativo de classificadores na previsão da evasão de alunos em EAD. In *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*, p. 1463.
- Rumberger, R. e Lim, S. A. (2008). Why Students Drop Out of School: A Review of 25 Years of Research. *California Dropout Research Project, Policy Brief 15, University of California*,
- Silva, J. L. D. da e Nunes, I. D. (2015). Mineração de Dados Educacionais como apoio para a classificação de alunos do Ensino Médio. In *Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)*, p. 1112 – 1121.
- Silva Filho, R. B. e Araújo, R. M. de L. (2017). Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências. *Educação Por Escrito*, v. 8, n. 1, p. 35.
- Soares, L. C. C. P., Ronzani, R. A., Carvalho, R. L. De e Silva, A. T. R. da (2020). Aplicação de técnicas de aprendizado de máquina em um contexto acadêmico com foco na identificação dos alunos evadidos e não evadidos. *Revista Humanidades e Inovação*, v. 7, n. 8, p. 224–235.
- Tasnim, N., Paul, M. K. e Sattar, A. H. M. S. (2019). Identification of Drop Out Students Using Educational Data Mining. In *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, p.1-5.
- Vidor, A., Rezende, C., Pacheco, E. e Caldas, L. (2011). Institutos federais: Lei no 11.892 de 29/12/2008 – Comentários e reflexões. In: Pacheco, E. [Org.], *Institutos Federais: Uma Revolução Na Educação Profissional e Tecnológica*, p. 47–113. Editora Moderna, Brasília.