Análise de Agrupamentos para Caracterização de Indicadores de Evasão

Daniel Victor Saraiva, Silas S. L. Pereira, Reinaldo B. Braga, Carina T. de Oliveira

[1] Laboratório de Redes de Computadores e Sistemas (LAR)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

victordvs@hotmail.com, {silas,reinaldo,carina}@lar.ifce.edu.br

Abstract. *Understanding the behavior of students at risk of dropout is essential for actions to encourage their permanence and success. In this context, this paper presents a proposal that performs data clustering to identify student profiles with risk of dropout, considering academic and socioeconomic information of students. A case study is presented taking into account a database of a Technical Informatics course. The results show that variables such as family income, study shift, student sex, age group and mother's level of education can influence students' academic performance.*

Resumo. *Entender o comportamento de estudantes em situação de risco de evasão é essencial para que possam ser realizadas ações de incentivo à permanência e êxito. Diante dessa problemática, este trabalho apresenta uma proposta que utiliza agrupamento de dados para identificar perfis de estudantes com risco de evasão a partir de informações acadêmicas e socioeconômicas dos estudantes. Um estudo de caso é apresentado considerando uma base de dados de um curso Técnico em Informática. Os resultados demonstram que, a partir da análise descritiva dos dados, variáveis como renda familiar, turno de estudo, sexo do estudante, faixa etária e grau de instrução da mãe podem influenciar no rendimento acadêmico dos estudantes.*

1. Introdução

Na Constituição Federal de 1988 [BRASIL 1988], a educação aparece como um direito social e universal, sendo dever de todas as instâncias do poder público promover e regulamentar os meios de acesso à educação. Em seu artigo 205, a educação está relacionada ao preparo dos sujeitos para o exercício da cidadania e qualificação para o trabalho.

Entretanto, mesmo diante das previsões constitucionais [BRASIL 1988] e legais (Lei de Diretrizes e Bases da Educação Nacional - LDB) [BRASIL 1996], que estabelecem meios para o acesso à educação, permanência e êxito estudantil, a evasão na educação emerge como um problema real. Diante disso, a evasão de estudantes representa um grande desafio para as instituições, desde a educação básica até o nível superior [IBGE 2020, INEP 2021].

Nesse cenário, analisar dados de estudantes em bases de dados educacionais para mapear as causas e motivos que levam os estudantes a evadirem pode auxiliar no planejamento de ações de intervenção na redução dos índices de evasão. Desse modo, técnicas de *mineração de dados* usando métodos de *aprendizagem de máquina* podem ser utilizadas no campo educacional para ampliar a compreensão do processo de aprendizagem, identificando e avaliando as variáveis que interferem no desempenho dos estudantes [Zhang and Li 2018, Hegde and Prageeth 2018].

Neste contexto, este trabalho apresenta uma proposta que utiliza *mineração de dados* e *análise de agrupamentos* para particionar uma base de dados real do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) com informações socioeconômicas e acadêmicas de estudantes entre diferentes perfis. O objetivo é identificar padrões ao encontrar grupos em que os estudantes compartilhem características ou propriedades similares entre eles. A técnica de *aprendizagem de máquina* não supervisionada *k-means* é utilizada para o agrupamento dos estudantes com base em seus atributos. A partir de uma *análise descritiva dos dados* de cada grupo formado, pode-se discriminar e caracterizar os perfis de estudantes que evadem e que se formam ou concluem o curso. Os resultados da *análise descritiva* mostram que variáveis como renda familiar, turno de estudo, sexo do estudante, faixa etária e grau de instrução da mãe podem influenciar no rendimento acadêmico dos estudantes. Com os resultados, o objetivo é que o conhecimento gerado possa auxiliar na tomada de decisões pedagógicas que visem propor ações de intervenções para controlar, acompanhar e conter a evasão estudantil baseadas nos principais indicadores de cada grupo de estudantes.

2. Mineração de Dados e Descoberta de Conhecimentos

A mineração de dados é um processo *automático* ou *semiautomático* de exploração para descoberta de padrões relevantes em bases de dados [Silva et al. 2016]. Segundo Faceli *et al.* (2011), técnicas de Aprendizagem de Máquina (AM) estão entre as mais empregadas no processo de mineração.

Os algoritmos de AM podem ser organizados de acordo com diferentes critérios. Um deles diz respeito ao paradigma de aprendizado a ser adotado para lidar com a sua tarefa. Essas tarefas podem ser divididas em: *preditivas* e *descritivas*. Em tarefas *preditivas* (aprendizado supervisionado), o objetivo é encontrar um modelo ou hipótese a partir dos dados de treinamento que possa ser utilizado para prever, com base no seus atributos de entrada, um rótulo ou valor que caracterize um novo exemplo. Em tarefas *descritivas* (aprendizado não supervisionado), o objetivo é explorar ou descrever um conjunto de dados no qual não se fazem uso de atributos de saída [Faceli et al. 2011].

Uma das divisões das *tarefas descritivas*, foco deste trabalho, é a *análise de grupos*, também conhecida como *agrupamento de dados*. Esse termo é usado para caracterizar métodos numéricos de análise de dados multivariados com objetivo de descobrir grupos homogêneos de objetos. Além disso, o agrupamento representa uma forma conveniente de organizar grandes bases de dados, permitindo realizar tarefas mais sofisticadas como na tomada de decisão em processos críticos [Castro and Ferrari 2016].

O algoritmo *k-means* é um dos métodos mais conhecidos para agrupamento [Castro and Ferrari 2016]. Em seu funcionamento, o algoritmo recebe como entrada um conjunto de dados $X = \{x_1, x_2, ..., x_m\}$ e um número k de agrupamentos a serem obtidos. O passo inicial é a designação dos k centroides iniciais $\mu_1, \mu_2, \dots, \mu_K$ para formação dos agrupamentos iniciais. Iterativamente, cada exemplo $x_i \in X$ é associado ao centroide mais próximo c_i, a partir de uma métrica para mensurar a distância entre os exemplos.

A segunda etapa envolve o cálculo dos centroides a partir dos novos agrupamentos formados. Este passo se repete até que não haja mais movimentação entre os centroides [Wu et al. 2008]. A função objetivo é descrita pela expressão $fc = \sum_{i=1}^{k} \sum_{x \in g_i} d(x, c_i)$, onde fc é a função de custo da base, x é um objeto qualquer da base X, c_i é o centroide do grupo g_i, $d(x, c_i)$ é a distância entre o objeto e o centroide do grupo.

3. Trabalhos Relacionados

A evasão apresenta-se como um problema educacional que preocupa muitas instituições de ensino. Para essas instituições, as maiores preocupações estão em identificar as características e situações que contribuem para a evasão dos estudantes.

A Tabela 1 apresenta e compara trabalhos nacionais e internacionais que utilizam mineração de dados e algoritmos de aprendizagem de máquina em bases de dados educacionais para classificar estudantes com risco de evasão. Os trabalhos apresentados na Tabela 1 são baseados em paradigmas distintos de aprendizagem para escolha do melhor algoritmo em classificar corretamente um estudante entre *Egresso com êxito* e *Egresso sem êxito*.

Diferente dos trabalhos apresentados na Tabela 1, o objetivo do presente trabalho é aplicar o algoritmo de agrupamento de dados *k-médias* para criar perfis diferentes de estudantes, proporcionando uma previsão mais aprofundada e eficiente da situação dos perfis e, então, apontar possíveis indicadores que levam à evasão.

Tabela 1. Trabalhos Relacionados.

Cenário	Trabalho	Qte registros	Algoritmo(s) de Classificação	Melhor Desempenho
Nacional	Maria, Damiani e Pereira (2016)	666	Rede Bayesiana	Rede Bayesiana (85,60%)
Nacional	Paz e Cazella (2017)	4.601	J48	J48 (91,42%)
Nacional	Lanes e Alcântara (2018)	916	J48	J48 (90,70%)
Nacional	Gonçalves, Silva e Cortes (2018)	574	Naive Bayes, SVM, e J48	J48 (98,08%)
Internacional	Hegde e Prageeth (2018)	50	Naive Bayes	Naive Bayes (72,00%)
Internacional	Solis et al. (2018)	15.720	SVM, MLP, *Random Forest* e Regressão Logística	Random Forest (85,00%)
Internacional	Perez, Castellanos e Correal (2018)	802	Naive Bayes, Árvore de decisão e Regressão logística	Árvore de decisão (94,00%)
Internacional	Dharmawan, Ginardi e Munif (2018)	103	SVM, KNN e Árvore de Decisão	SVM e Árvore de Decisão (66,00%)

4. Proposta

As etapas utilizadas nesse trabalho para a geração de agrupamentos com diferentes perfis de estudantes são apresentadas de forma simplificada na Figura 1. Este fluxo é baseado na metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [Wirth and Hipp 2000], comumente utilizada em projetos de mineração de dados. Para que a proposta possibilite resultados adequados, retornos a etapas anteriores para ajustes e verificações nos procedimentos adotados podem ser necessários (setas em verde na Figura 1). Além da interação entre as etapas, o processo ainda pode conter *loops* entre duas ou mais etapas para alinhar os resultados com os objetivos da pesquisa.

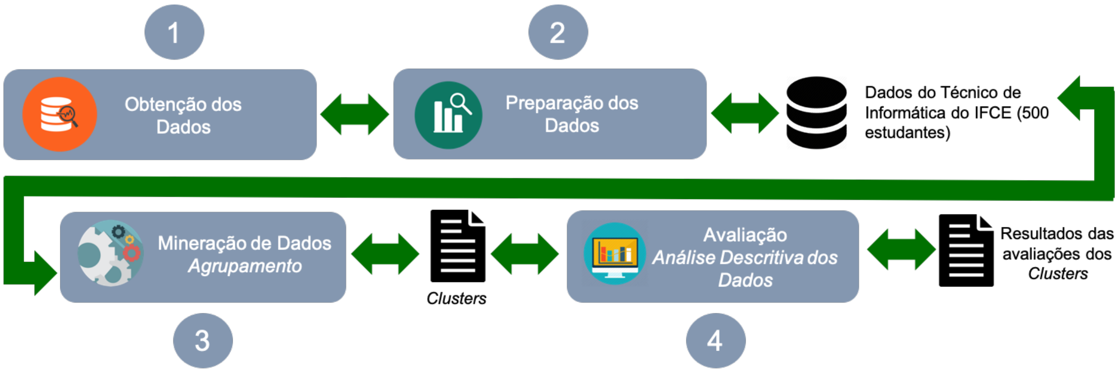


Figura 1. Etapas do processo de descoberta de conhecimento.

4.1. Etapa 1 - Obtenção dos Dados

A partir de estudos e discussões sobre a evasão, o Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) [IFCE 2017] analisa a evasão sob a ótica do *curso* por considerar que é a granularidade que mais permite se aproximar de suas causas, possibilitando uma análise mais completa acerca das possibilidades de saída do estudante da instituição, *com êxito*, por meio da conclusão do curso ou *sem êxito* por meio da evasão. Seguindo tal orientação, a proposta deste trabalho analisa a evasão de estudantes por *curso*.

Nesse cenário, os dados selecionados são de um curso Técnico em Informática do IFCE que possui alta taxa de evasão. Do total de 16 semestres avaliados, em 10 as porcentagens de egressos sem êxito passam dos 50% em relação ao total de ingressantes. A base contém dados de 710 estudantes (*Egressos com êxito* e *Egressos sem êxito*) entre os semestres 2010/2 e 2018/2. Os dados foram disponibilizados em arquivo no formato *Comma-Separated Values* (CSV) com informações acadêmicas e socioeconômicas dos estudantes da instituição.

4.2. Etapa 2 - Preparação dos Dados

Em seguida, técnicas de pré-processamento de dados (exclusão de atributos irrelevantes, exclusão de instâncias com muitos valores ausentes e correção de inconsistências) são aplicadas na etapa de *Preparação dos Dados* para melhorar a qualidade da base. Também é realizada a *padronização* dos dados com o objetivo de resolver as diferenças de unidades e escalas dos dados, como também tratar possíveis *outliers* na base. Ao final dessa etapa, é gerada uma base de dados com 500 registros e 17 atributos com informações sobre os estudantes.

Os atributos com informações socioeconômicas e acadêmicas dos estudantes são: idade ao entrar no curso; sexo; renda familiar; etnia; estado civil do estudante; grau de instrução dos estudante antes de entrar no curso; estado civil dos pais; grau de instrução da mãe; grau de instrução do pai; forma de ingresso no curso; Índice de Rendimento Acadêmico (IRA); turno do estudante; total de períodos aprovados no curso; total de períodos aprovados parcialmente no curso[1]; total de semestres aprovados com dependência[2]; total de semestres reprovados e total de períodos trancados.

[1]Atributo que representa que o estudante ficou reprovado em, pelo menos, uma disciplina do período.
[2]Atributo que representa que o estudante ficou reprovado em até duas disciplinas do período letivo.

4.3. Etapa 3 - Mineração de Dados

Na etapa de *Mineração de Dados*, a utilização de um modelo de agrupamento é a etapa central de todo o processo envolvido na criação dos *clusters*, em que o algoritmo *k-means* (Seção 2) é aplicado aos dados para identificação de *clusters*. Essa etapa também avalia os resultados gerados e deve determinar se os *clusters* são significativos ao definir o número apropriado para a base de dados analisada.

4.4. Etapa 4 - Avaliação

A última etapa da Figura 1 é a *Avaliação* dos agrupamentos gerados por meio da *Análise Descritiva de Dados*. Essa etapa engloba a compreensão dos *clusters* a partir dos quais se deseja descobrir algum tipo de conhecimento. Para uma melhor interpretação dos dados existentes nos *clusters*, é utilizado o software *Tableau*[3] (versão 2019.2), uma solução para visualização de dados interativos focada em *Business Intelligence* (BI) que permite a criação de filtros, imagens, painéis interativos, entre outras funcionalidades, que facilitam a exploração e a análise de dados.

5. Resultados e Discussões

Esta seção apresenta os principais indicadores que levam os estudantes de um curso de Técnico em Informática do IFCE a evadirem conforme resultados alcançados com o processo detalhado na Seção 4.

5.1. Agrupamento dos dados

No primeiro passo do processo de agrupamento, é determinada a medida de *similaridade* para os objetos. Na execução, é utilizada a *distância euclidiana* como medida de *similaridade* por ser a mais comumente utilizada para representar a distância física entre pontos em um espaço m-dimensional [Castro and Ferrari 2016].

Para realização dos agrupamentos, é utilizado o algoritmo *k-means* que está disponível no *sklearn*[4]. Além disso, o método de inicialização dos centroides do algoritmo *k-means++* [Arthur and Vassilvitskii 2007] é escolhido por selecionar os centroides iniciais de forma mais distribuída nos dados, diferente da escolha aleatória de centroides. Outros dois parâmetros do algoritmo *k-means* são os critérios de parada: o primeiro é baseado na função de custo que interrompe o algoritmo quando não há mudança no posicionamento dos centroides *(tol= 1e-4)*; o segundo é o número máximo de iterações do algoritmo caso a função de custo ainda apresente pequenas variações ($max\_iter = 500$).

Para definir o valor final de k *clusters* para a fase de *análise descritiva dos dados*, o trabalho avalia a qualidade dos *clusters* gerados utilizando uma medida *externa*. Neste contexto, é realizada a avaliação dos *clusters* gerados em relação à base original classificada entre *Egressos com êxito* e *sem êxito* (atributo classe *Situação da Matrícula*).

Neste trabalho, optou-se por usar um número de $k = 3$. Essa decisão é tomada após a realização de execuções com valores de k até 5 *clusters*, onde o valor de $k = 3$ apresentou os melhores perfis de estudantes entre os demais valores de k. De acordo com a Figura 2, para o número de $k = 3$, ao analisar a classe *Sit. da Matrícula* por *cluster*

[3]https://www.tableau.com

[4]A biblioteca está disponível em: http://scikit-learn.sourceforge.net.

gerado, o *cluster 1* possui a maioria dos estudantes *Egressos com êxito* (83,82% do total de 241 estudantes), o *cluster 2* possui a maioria dos estudantes *Egressos sem êxito* (97,99% do total de 199 estudantes), e o *cluster 0* possui um equilíbrio entre o total de *Egressos com êxito* (40,00% do total de 60 estudantes) e *Egressos sem êxito* (60,00% do total de 60 estudantes), gerando assim, três perfis de estudantes diferentes.

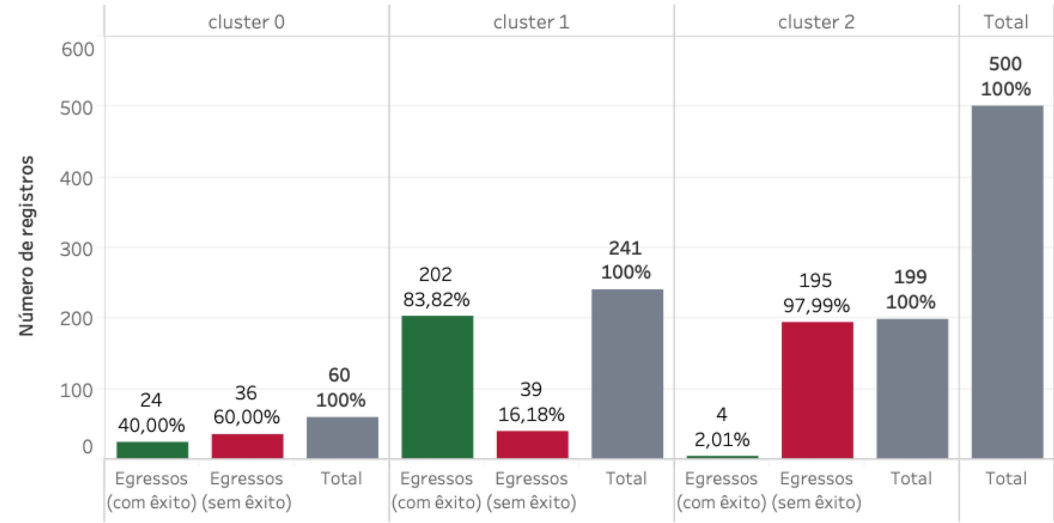
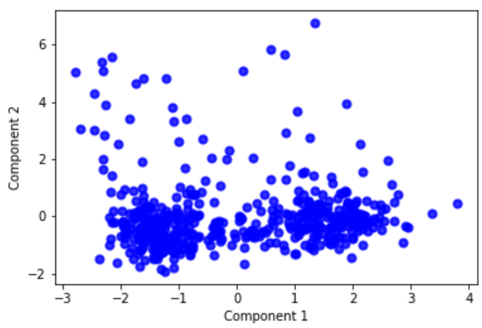
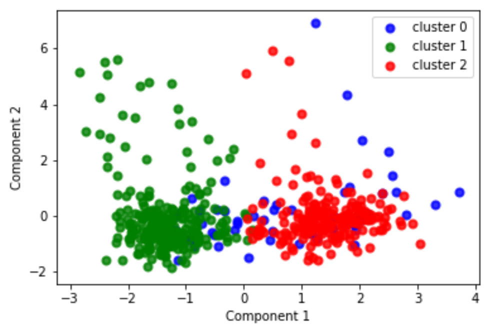


Figura 2. *Situação da Matrícula* por *cluster*.

Na Figura 3 é utilizada a técnica de *Análise de Componentes Principais* (PCA) disponível no *sklearn* para auxiliar na visualização de como os dados estão estruturados, ao possibilitar a redução da dimensionalidade da base de dados com uma perda mínima de informação. A Figura 3(a) apresenta a base original com 500 registros dispostos em um plano bidimensional. Neste gráfico, a base composta pelos dezessete atributos selecionados é reduzida para dois componentes principais. Já a Figura 3(b), apresenta a base com 500 registros dispostos em um plano bidimensional com o valor final de $k = 3$.



(a) Base original



(b) Base para $k = 3$

Figura 3. Análise das componentes principais.

5.2. Análise Descritiva dos Dados

Entre os atributos avaliados, são apresentados nesta seção os resultados das análises de três dos atributos que indicam relação com o desempenho dos estudantes: Índice de Rendimento Acadêmico (IRA); Renda Familiar e Turno de Estudo.

5.2.1. Índice de Rendimento Acadêmico (IRA)

Para os três *clusters*, ao analisar a média do IRA (atributo que mede o desempenho acumulado dos estudantes no curso) por *cluster* gerado (Figura 4), tem-se o *cluster 0* com média de 5,06 para um total de 60 estudantes, representando estudantes com **desempenho médio** no curso, o *cluster 1* com média de 7,71 para um total de 241 estudantes, representando estudantes com **bom desempenho** no curso e *cluster 2* com média de 2,14 para um total de 199 estudantes, representando estudantes com **desempenho ruim**.

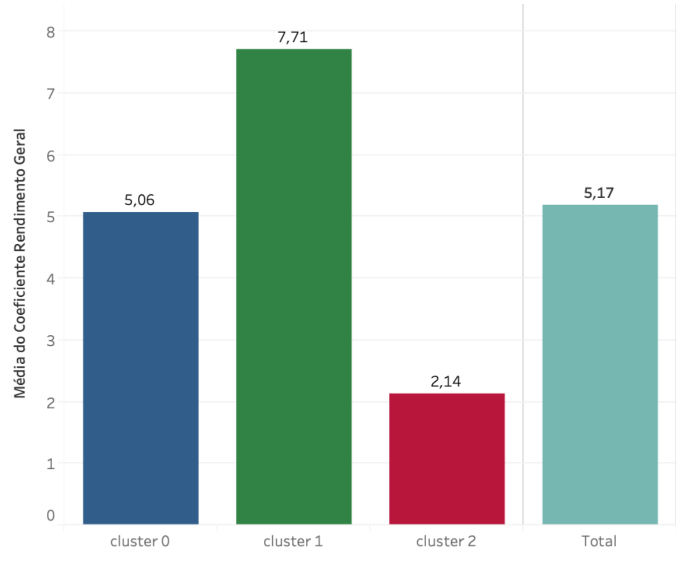


Figura 4. Média do Índice de Rendimento Acadêmico (IRA) por *cluster*.

5.2.2. Renda Familiar

Em relação à renda familiar, de acordo com a Figura 5, os melhores desempenhos acadêmicos são atribuídos aos estudantes com melhor renda. As faixas de renda de *1 a 2 salários* até a faixa *acima de 2 salários* possuem mais de 55% dos estudantes por faixa presentes no *cluster 1*, que possui média do IRA de 7,71 (Figura 4). Além disso, o curso possui maioria de estudantes de renda familiar até um salário.

5.2.3. Turno de Estudo

Como mostra a Figura 6, os estudantes que estão matriculados no período noturno possuem uma maior tendência de ser um *Egresso com êxito*, ou seja, concluir o curso. Para o turno da noite, do total de 234 estudantes, 58,97% estão presentes no *cluster* dos estudantes com melhor desempenho (*cluster 1)* e 41,03% estão presentes no *cluster* com pior desempenho médio em relação ao IRA (*cluster 2)*. Já o turno vespertino, a maioria dos estudantes estão presentes no *cluster* com pior desempenho médio em relação ao IRA (*cluster 2)*. Em relação ao turno matutino, do total de 50 estudantes, 56% estão presentes no *cluster* dos estudantes com desempenho médio em relação ao IRA (*cluster 0)*.

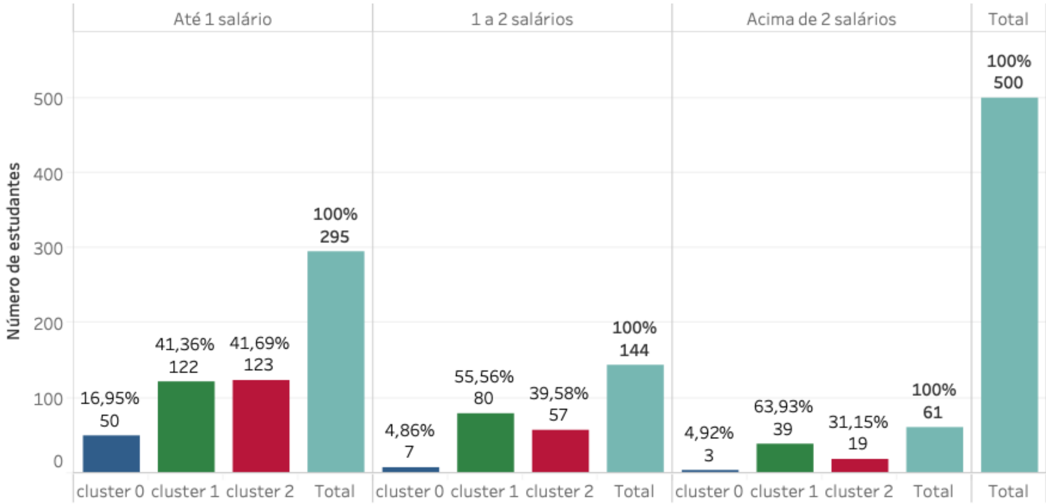


Figura 5. Renda familiar por *cluster*.

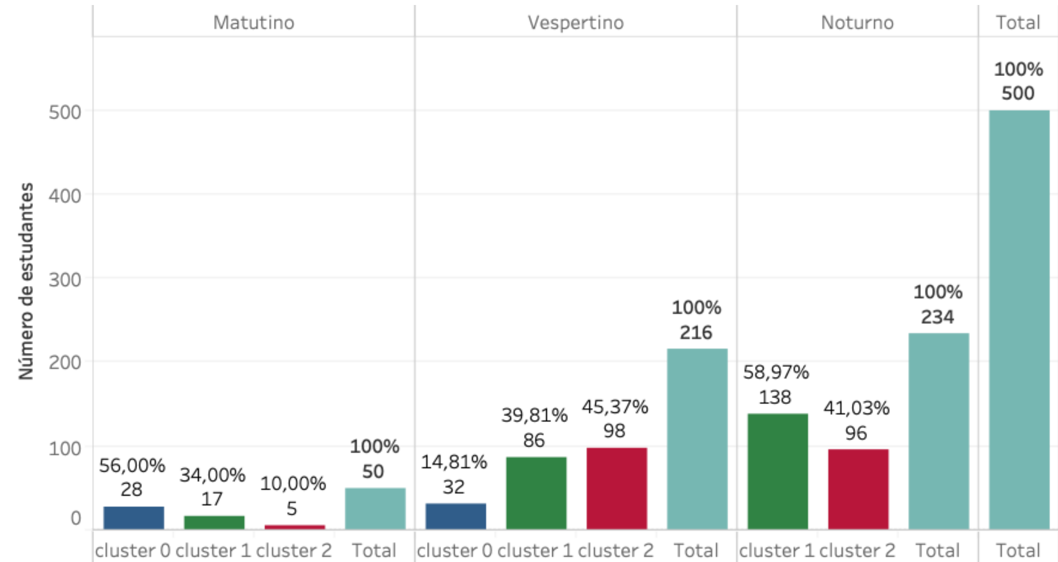


Figura 6. Turno de estudo por *cluster*.

5.3. Conclusões da Análise Descritiva dos Dados

A Tabela 2 mostra as avaliações dos principais atributos que indicam relação com o desempenho dos estudantes junto com os três atributos apresentados na Seção 5.2. Os estudantes com melhor desempenho (com IRA médio de 7,71) têm como atributos: renda familiar acima de 2 salários; turno de estudo noturno; sexo masculino; faixa etária acima de 20 anos e grau de instrução da mãe com o Ensino Médio Incompleto ou com o Ensino Médio Completo. Já para os estudantes com pior desempenho (com IRA médio de 2,14) têm como atributos: renda familiar até um salário; turno de estudo vespertino; sexo feminino; faixa etária até 19 anos e grau de instrução da mãe sem formação até o Ensino Médio Incompleto. Portanto, pode-se verificar um conjunto de indicadores que podem estar associados ao baixo desempenho acadêmico do estudante, como renda familiar, turno de estudo, sexo do estudante, faixa etária e grau de instrução da mãe.

Tabela 2. Resultado das avaliações dos agrupamentos.

Atributo	*Cluster 0*	*Cluster 1*	*Cluster 2*
IRA	IRA médio de 5,06	IRA médio de 7,71	IRA médio de 2,14
Renda Familiar	Até um salário	Acima de 2 salários	Até um salário
Turno de Estudo	Matutino	Noturno	Vespertino
Sexo do Estudante	Feminino	Masculino	Feminino
Faixa Etária	Até 19 anos	Acima de 20 anos	Até 19 anos
Grau de Instrução da Mãe	Sem formação até o Ensino Fundamental Completo	Ensino Médio Incompleto ou Ensino Médio Completo	Sem formação até o Ensino Fundamental Completo

6. Considerações finais

O uso de técnicas de mineração de dados e análise de agrupamentos em contextos educacionais oferece oportunidades para que educadores e pesquisadores tenham acesso a conhecimentos úteis, gerados a partir de conjuntos de dados de instituições de ensino. Acrescenta-se também que técnicas de visualização de dados podem ser uma ferramenta útil para descobrir padrões interessantes a partir de dados agrupados. Além disso, a partir da análise descritiva dos três *clusters* obtidos pelo algoritmo *k-means*, pode-se verificar um conjunto de indicadores que estão associados ao desempenho acadêmico do estudante, como renda familiar, turno de estudo, sexo do estudante, faixa etária e grau de instrução da mãe. Assim, por meio dos resultados, professores e gestores podem propor ações de intervenção baseadas nos indicadores gerados para controlar, acompanhar e conter a evasão estudantil.

Direcionamentos futuros desta pesquisa incluem o desenvolvimento de uma ferramenta para apoio à decisão do gestor educacional que possibilite a visualização amigável e maior compreensão dos indicadores resultantes da aplicação das técnicas mineração de dados e aprendizagem de máquina. Ademais, planeja-se ainda experimentos adicionais envolvendo outros *datasets* educacionais no intuito de avaliar a capacidade de outras abordagens de aprendizagem não supervisionada.

Referências

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *ACM-SIAM Symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics.

BRASIL (1988). *Constituição da República Federativa do Brasil*. Constituição (1988), Senado Federal.

BRASIL (1996). *Lei nº 9.394, de 1996, que estabelece as diretrizes e bases da educação nacional, e legislação correlata*. BRASIL. Lei de Diretrizes e Bases da Educação Nacional.

Castro, L. N. and Ferrari, D. G. (2016). *Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações*. Saraiva, São Paulo, 1 edition.

Dharmawan, T., Ginardi, H., and Munif, A. (2018). Dropout detection using non-academic data. In *International Conference on Science and Technology (ICST)*.

Faceli, K., Lorena, A. C., Gama, J., and de Carvalho, A. C. P. L. F. (2011). *Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina*. LTC, Rio de Janeiro.

Gonçalves, T., Silva, J., and Cortes, O. (2018). Técnicas de mineração de dados: um estudo de caso da evasão no ensino superior do instituto federal do maranhão. *Revista Brasileira de Computação Aplicada*, 10(3):11–20.

Hegde, V. and Prageeth, P. P. (2018). Higher education student dropout prediction and analysis through educational data mining. In *International Conference on Inventive Systems and Control (ICISC)*.

IBGE (2020). *Educação: 2019*. IBGE. Instituto Brasileiro de Geografia e Estatistica.

IFCE (2017). *Plano estratégico para permanência e êxito dos estudantes do IFCE*. IFCE. Instituto Federal de Educação, Ciência e Tecnologia do Ceará. Pró-reitoria de Ensino - PROEN, Fortaleza.

INEP (2021). *Resumo técnico do Censo da Educação Superior 2019*. INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília-DF.

Lanes, M. and Alcântara, C. (2018). Predição de alunos com risco de evasão: estudo de caso usando mineração de dados. In *Simpósio Brasileiro de Informática na Educação (SBIE)*.

Maria, W., Damiani, J. L., and Pereira, M. (2016). Rede bayesiana para previsão de evasão escolar. In *Congresso Brasileiro de Informática na Educação (CBIE)*.

Paz, F. J. and Cazella, S. C. (2017). Identificando o perfil de evasão de alunos de graduação através da mineração de dados educacionais: um estudo de caso de uma universidade comunitária. In *Congresso Brasileiro de Informática na Educação (CBIE)*.

Perez, B., Castellanos, C., and Correal, D. (2018). Applying data mining techniques to predict student dropout: A case study. In *IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*.

Silva, L. A., Peres, S. M., and Boscarioli, C. (2016). *Introdução à mineração de dados: com aplicações em R*. Elsevier, Rio de Janeiro, 1 edition.

Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., and Hernandez, M. (2018). Perspectives to predict dropout in university students with machine learning. In *IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*.

Wirth, R. and Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *International Conference on the Practical Applications of Kknowledge Discovery and Data Mining*.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37.

Zhang, L. and Li, K. F. (2018). Education analytics: Challenges and approaches. In *International Conference on Advanced Information Networking and Applications Workshops (WAINA)*.