

Evasão Acadêmica e suas Causas em Cursos de Bacharelado em Ciência da Computação: Um Estudo de Caso na UFERSA

Leonardo Torres Marques¹, Bruno Torres Marques¹, Rayana Souza Rocha², Lenardo Chaves e Silva², Paulo Gabriel Gadelha Queiroz², Angélica Félix de Castro²

¹Universidade Federal do Ceará (UFC)
Caixa Postal 15.064 – 91.501-970 – Quixadá – CE – Brasil

²Programa de Pós-Graduação em Ciência da Computação (UFERSA/UERN) –
Universidade Federal Rural do Semi-Árido (UFERSA)
Caixa Postal 137 - 59625 - 900 – Mossoró – RN – Brasil

leonardo.torresmarques@gmail.com, brunotorresepj@gmail.com

rayana.rocha@alunos.ufersa.edu.br, lenardo@ufersa.edu.br

pgabriel@ufersa.edu.br, angelica@ufersa.edu.br

Abstract. *Academic dropout is an important, complex problem faced by many higher education institutions. In such a context, this study unravels its causes in the Computer Science course at UFERSA. Specially, we use a clustering technique over a database extracted from the Integrated Management System for Academic Activities. There are two clusters: one to represent characteristics of the evaded students, and another of the students who concluded the course. The main features of the second group include: living in the academic housing, having previous experience in the area prior beginning the course, being white, and other privileges (non-divorced parents, private high schools, high monthly income).*

Resumo. *A evasão acadêmica é um problema importante e complexo enfrentado por muitas instituições de ensino superior. Nesse contexto, este estudo desvenda suas causas no curso de Ciência da Computação em UFERSA. Em especial, utilizamos uma técnica de agrupamento sobre um banco de dados extraído do Sistema de Gestão Integrado de Atividades Acadêmicas. São dois clusters: um para representar as características dos alunos evadidos e outro dos alunos que concluíram o curso. As principais características do segundo grupo são: residir em moradia acadêmica, ter experiência anterior na área antes do início do curso, ser branco e outras privilégios (pais não divorciados, ensino médio particular, renda mensal elevada).*

1. Introdução

A evasão acadêmica nos cursos de graduação das universidades brasileiras representa um problema complexo que atinge inúmeras instituições. Existe uma preocupação dos governos e das instituições em atenuar os índices de evasão acadêmica dos cursos por parte dos estudantes universitários [Manhães *et al.* 2011]. Uma pesquisa realizada no

Brasil mostrou que cerca de quarenta por cento (40%) dos alunos da rede pública de ensino superior estão abandonando os cursos [Pessoal 2018]. Este grave problema resulta no desperdício de dinheiro público, na não assimilação do conhecimento necessário às ciências e aos ofícios e na abdicação de uma conquista individual [Souza 2008].

O caso é ainda mais grave nos cursos da área de Computação. No Brasil, setenta e três por cento (73%) dos alunos de bacharelado em Ciência da Computação (CC) desistem da faculdade precocemente, enquanto nos cursos de Sistemas de Informação (SI), o percentual é de sessenta e sete virgula três por cento (67,3%) [Hoed 2016].

Além disso, é importante reforçar a relevância do problema da evasão, sob a perspectiva de que ele pode desencadear diversos outros malefícios para as instituições, tais como, problemas inerentes às áreas econômica e social, além do descumprimento da função política gerencial da instituição.

Observa-se ainda que a verba para manutenção das universidades públicas está diretamente associada a quantidade de alunos com matrícula ativa [Manhães *et al.* 2011]. Dessa forma, o orçamento universitário sofre muitas perdas, dificultando a gestão institucional, pois o número de docentes, técnicos administrativos, serviços terceirizados e a estrutura institucional continuam os mesmos, independentemente do número de alunos.

Quando a universidade não consegue manter o aluno até o final do curso, pode-se denominar de fracasso institucional, que inclui desde o docente que não conseguiu exercer o papel enquanto educador, até os programas e planos estabelecidos pelas Instituições de Ensino Superior (IES) por não cumprir a missão institucional de formar o seu alunado [do Nascimento *et al.* 2018].

Embora existam diversos fatores relacionados a evasão acadêmica, eles podem ser divididos entre fatores internos e externos à instituição. Os fatores internos são ligados ao curso, e podem ser subdivididos em: infraestrutura, corpo docente e a assistência sócio educacional. Os fatores externos à instituição estão relacionados ao aluno, e são exemplificados em: aspectos vocacionais, aspectos socioeconômicos e problemas pessoais [da Silva Zanato *et al.* 2018].

O estudo das causas da evasão acadêmica e a tomada de medidas preventivas estão fortemente ligados ao contexto de cada instituição de ensino. Entretanto, observa-se que nem todas as instituições conseguem tomar medidas efetivas, pois segundo Manhães *et al.* (2011), a identificação dos fatores que influenciam à evasão acadêmica e atribuição de uma ordem de importância para esses fatores é um trabalho complexo que está diretamente ligado a análise do conjunto de alunos. Vale destacar que cada instituição deve identificar as causas do fenômeno em seu ambiente educacional [Nagai e Cardoso 2017].

Atualmente, o processo de identificação das causas da evasão acadêmica e retenção dos alunos da Universidade Federal Rural do Semi-Árido (UFERSA) campus Mossoró/RN é manual, subjetivo, empírico e sujeito a falhas, pois depende primordialmente da experiência acadêmica e do envolvimento, tanto dos docentes, como dos discentes. Geralmente, os docentes desempenham inúmeras atividades, além das atividades de sala de aula. Portanto, é difícil acompanhar e reconhecer as necessidades de cada aluno e identificar aqueles alunos que apresentam risco de evasão acadêmica. Dessa forma, a adoção de mecanismos automatizados para viabilizar a detecção precoce de grupos de alunos com risco de evasão acadêmica pode fornecer resultados promissores no combate

à evasão acadêmica.

Nesse contexto, buscou-se neste estudo identificar os motivos das altas taxas de evasão acadêmica dos alunos do curso de bacharelado em CC da UFERSA campus Mossoró/RN. Ressalta-se que foi utilizada uma base de dados extraída do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) fornecida pela instituição, sendo utilizado como elemento principal para análise, o percurso do estudante no curso antes da evasão acadêmica.

O restante deste trabalho está organizado da seguinte forma: na Seção 2, apresentam-se os trabalhos relacionados; na Seção 3, são apresentados os procedimentos, métodos e ferramentas utilizadas neste trabalho; na Seção 4, são discutidos os resultados obtidos; e por fim, na Seção 5, são apresentadas as conclusões, limitações e perspectivas de trabalhos futuros.

2. Trabalhos Relacionados

Nos últimos anos, diversos trabalhos têm explorado os benefícios que a Mineração de Dados Educacionais (MDE) pode trazer ao ambiente educacional e principalmente para desvendar causas da evasão acadêmica. Em Freitas Júnior *et al.* (2019), destaca-se que uma solução propícia para o estudo das causas da evasão acadêmica é o uso da descoberta de conhecimento, por meio de técnicas de bancos de dados e Mineração de Dados (MD). Enquanto isso, em Marques *et al.* (2019), os autores buscaram elencar ferramentas, técnicas e fatores indutores utilizados para prever as causas do problema (evasão escolar) nos últimos dez (10) anos.

Do Nascimento *et al.* (2018) apontou o baixo desempenho e falta de conhecimento nas disciplinas como motivos da evasão escolar dos alunos de ensino fundamental. Para isso, foram utilizados modelos de regressões linear. Os resultados indicam que a regressão robusta obteve melhores resultados na estimação das variáveis elencadas nesse estudo.

No trabalho de Martinho *et al.* (2013), os autores propõem um sistema inteligente que visa a possibilidade de identificar, de forma proativa, contínua e precisa, os alunos da educação presencial tradicional, propensos à evasão acadêmica no ensino superior. Foi utilizada uma Rede Neural *ARTMAP-Fuzzy*, uma das técnicas de Inteligência Artificial, com possibilidade de aprendizagem continuada. Os resultados mostraram que o sistema proposto é satisfatório. Contudo, os autores propõem a aplicação de outras técnicas para comparação de resultados.

Resumidamente, os autores destes trabalhos observaram que o emprego de algoritmos de MD em dados educacionais para a previsão da situação acadêmica é um campo de investigação ainda não estabilizado, demandando investigações complementares tanto na definição das variáveis a serem utilizadas quanto nas técnicas de MD empregadas. Sendo assim, este estudo se diferencia dos trabalhos apresentados pelo emprego de novas variáveis na exploração de indicadores de evasão escolar e por mostrar como essas variáveis estão correlacionadas.

3. Procedimentos e Métodos

Quanto ao tipo de estudo, este trabalho, caracteriza-se como descritivo e aplicado, validado por meio de estudo de caso. Em relação à abordagem desta pesquisa, classifica-se

como predominante quantitativa. Os principais dados primários coletados dizem respeito a fatores e índices de evasão acadêmica do curso de CC da UFERSA campus Mossoró/RN.

A tarefa de análise automática de dados é baseada em um processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, objetivando melhorar o entendimento de um problema ou apoiar uma tomada de decisão [Fayyad *et al.* 1996]. Esse processo envolve preparação de dados, aplicação de algoritmos de MD, análise e interpretação de padrões e avaliação do conhecimento extraído. Os dados devem representar casos, cenários, exemplos ou instâncias representativas do domínio a ser tratado [Webber *et al.* 2013]. Neste trabalho, utilizou-se as etapas do processo de *Knowledge Discovery in Databases*¹ (KDD) para exploração, processamento e análise dos dados.

Na averiguação inicial do experimento se previu algumas atividades preliminares. Primeiro, os dados passaram por um pré-processamento (limpeza dos dados e normalização). Após essa etapa, foi realizada a adequação dos dados na ferramenta (*RStudio*) ajustando de acordo com os parâmetros do algoritmo (*K-Means*). Com os dados ajustados, o algoritmo foi executado oito (8) vezes para obtenção de dados confiáveis. A cada execução foi verificada a acurácia dos resultados, e finalmente, quando os resultados se estabilizaram foi considerada a última iteração.

Na etapa de análise se utilizou a técnica de clusterização (grupos). Essa técnica se vale de modelos estatísticos multivariados tencionando formar grupos de objetos homogêneos e similares entre si. Notadamente as técnicas de agrupamento buscam formar grupos maximizando à similaridade intragrupos e minimizando as similaridades intergrupos [Webber *et al.* 2013]. Além disso, a análise de grupos permite identificar objetos que não apresentam similaridades com nenhum grupo.

Os grupos podem ser específicos (cada instância pertence a apenas um grupo) ou amontoados (uma instância pode pertencer a mais de um grupo). Os métodos específicos podem ser baseados em distâncias (as instâncias são particionadas conforme medidas de similaridade e de distância) ou hierárquicos (as instâncias são estruturadas em níveis hierárquicos partindo de um alto nível geral até o nível de instâncias). Os métodos que permitem amontoados de instâncias e grupos são probabilísticos (uma instância pode pertencer a cada grupo com uma certa probabilidade) ou *fuzzy* (uma instância pode pertencer a cada grupo em diferentes graus de inclusão). Em geral a escolha do melhor método de agrupamento não é fácil, devendo ser guiada pela natureza dos dados e pelos resultados esperados.

Preliminarmente foram realizados testes com três (3) métodos de agrupamento: *AGNES*, *DBSCAN* e *K-Means*. Após os testes iniciais, decidiu-se continuar a análise utilizando o *K-Means* para se aplicar na etapa de MD deste estudo, por determinar melhores padrões do que os demais, com alto nível de acurácia. Conhecido como *K-Medias*, o algoritmo *K-Means* é um algoritmo de agrupamento simples e eficiente, que fornece uma classificação baseada em análises e comparações dos resultados presentes nos dados, fornecendo uma classificação automática sem nenhuma pré-classificação existente.

Foram analisados os dados de quatrocentos e cinquenta (450) alunos que concluíram ou evadiram do curso de CC da UFERSA campus Mossoró/RN. Aproximada-

¹Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>

mente sessenta por cento (60%), dos dados são de alunos evadidos e quarenta por cento (40%) dos dados são de concludentes do curso. Esses dados foram analisados estatisticamente por meio de medidas descritivas como frequências e por meio da utilização do algoritmo *K-Means* na análise de agrupamento.

Os grupos formados pelo algoritmo selecionado foram avaliados pelos critérios de homogeneidade e separação. O critério de homogeneidade visa averiguar se os dados mais similares possíveis foram colocados no mesmo grupo, por isso quanto menor o resultado obtido mais similar é o grupo. O critério de separação avalia se os grupos estão suficientemente diferentes uns dos outros, ou seja, quanto maior o resultado, maior a distância entre os grupos [Webber *et al.* 2013].

Vale salientar, que os dados utilizados foram de alunos que ingressaram no curso de CC da UFERSA campus Mossoró/RN, entre os períodos de 2012.1 e 2018.2. As variáveis que foram avaliadas neste trabalho foram: situação do aluno (evadido ou concludente), morou na vila acadêmica, experiência prévia, relação conjugal dos pais ao ingressar no curso, cor da pele, estado de origem, tipo de escola que o aluno estudou no ensino médio (pública ou privada), forma de ingresso (ação afirmativa), Índice de Rendimento Acadêmico (IRA), total de reprovações e renda familiar do aluno.

É importante destacar que algumas destas variáveis utilizadas não foram exploradas nos trabalhos relacionados apresentados na seção 2, tais como: morou na vila acadêmica, experiência prévia, relação conjugal dos pais ao ingressar no curso e tipo de escola que o aluno estudou no ensino médio (pública ou privada). Essas variáveis foram definidas com o auxílio da psicóloga e do núcleo de atendimento ao alunado da instituição. Uma vez definidas as variáveis, estas foram disponibilizadas (em forma de perguntas) por meio de um questionário no SIGAA, assim os alunos tiveram acesso ao questionário e puderam respondê-lo.

Ressalta-se, que tanto a limpeza e normalização quanto a análise dos dados foram realizadas com uso da ferramenta *RStudio* (*RStudio Team* 2015²) que é um ambiente de software livre para Computação Estatística, que fornece uma ampla variedade de técnicas estatísticas e visualização de dados por meio de gráficos [Zhao 2011].

4. Resultados e Discussão

Nesta seção são apresentados e analisados os principais resultados das manipulações dos dados. As análises são descritas em duas etapas: primeiro uma perspectiva para cada variável deste estudo (estatística descritiva), em seguida, são expostos os resultados da aplicação do algoritmo *K-Means* sobre os dados (análise de agrupamento).

4.1. Variáveis Exploradas

Evidencia-se que o índice de evasão no decorrer do curso é alarmante em detrimento aos que concluíram, pois, cerca de sessenta por cento (60%) dos alunos que responderam o questionário desistiu no período. A diferença de alunos que não residiram na vila acadêmica em relação aqueles que residiram é discrepante, pois apenas quinze por cento (15%) dos alunos residiram na vila acadêmica. Destaca-se que cerca de setenta por cento (70%) dos alunos não tinham experiência prévia na área (Computação) ao ingressar

²Disponível em: <https://rstudio.com/>

no curso, deve-se destacar que moram na vila acadêmica alunos que apresentam vulnerabilidade socioeconômica. Quanto a relação conjugal dos pais dos alunos ao ingressar no curso, evidencia-se que aproximadamente sessenta e três por cento (63%) dos alunos responderam que seus pais moravam separados.

São destacadas as características biológicas dos alunos evadidos e concludentes na Figura 1. Observa-se proeminência para os pardos, que representam metade de todas as características biológicas declaradas pelos alunos. E em seguida, os auto declarados brancos, aparecem com cerca de quarenta por cento (40%). Verifica-se um percentual menos significativo de negros com sete por cento (7%) e amarelos com dois por cento (2%).

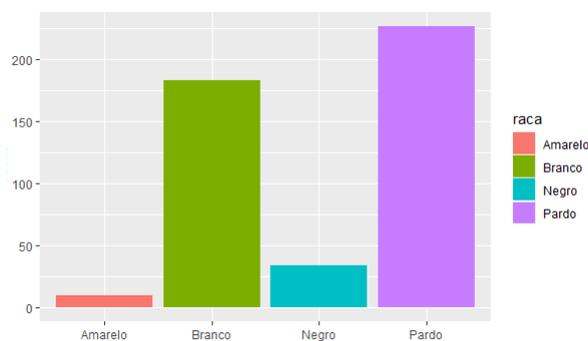


Figure 1. Cor da Pele.

Os estados que apareceram com maiores números de ingressantes são o Rio Grande do Norte (RN) com mais de trezentos (300) ingressantes, e o estado do Ceará (CE), com mais de cem (100) ingressantes, esses números em relação aos estados de origem dos alunos se observa na Figura 2. Quanto ao tipo de escola que o aluno estudou no ensino médio, acentua-se que aproximadamente sessenta e oito por cento (68%) cursaram integralmente o ensino médio em escolas públicas, enquanto isso, vinte e quatro por cento (24%) cursaram integralmente o ensino médio em escolas privadas e oito por cento (8%) cursaram o ensino médio de forma híbrida, parte em escolas públicas e outra parte em escolas privadas.

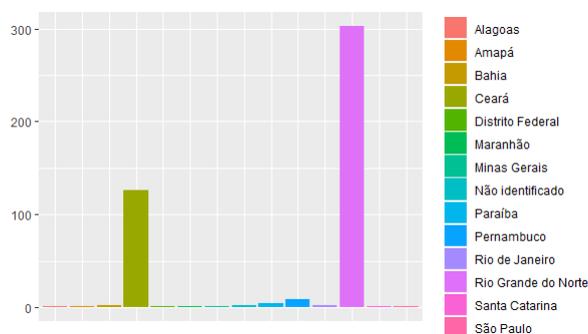


Figure 2. Estado de Origem.

Na Figura 3 são detalhadas as formas de ingresso (ação afirmativa) dos alunos na instituição, destacando aqueles que entraram por Ampla Concorrência (AC), bem como os tipos de cotas que alguns se encaixaram. É possível observar, que a grande maioria

dos alunos entraram por AC, com cerca de setenta e sete por cento (77%) das vagas preenchidas dessa forma. Em seguida, aparece a cota L06 com aproximadamente oito por cento (8%) de ingressantes. Essa cota é referente a candidatos autodeclarados pretos, pardos ou indígenas que, independente da renda, tenham cursado integralmente o ensino médio em escolas públicas. Vale ressaltar, que essas formas de ingresso incluem tanto os alunos que concluíram como aqueles que evadiram.

Outros tipos de cotas apareceram na seguinte ordem: a L02, representa candidatos autodeclarados pretos, pardos ou indígenas, com renda familiar bruta per capita igual ou inferior a 1,5 salário mínimo e que tenham cursado integralmente o ensino médio em escola pública; a L05 são candidatos que, independente da renda, tenham cursado integralmente o ensino médio em escolas públicas; e, por fim, a L01 representa candidatos com renda familiar bruta per capita igual ou inferior a 1,5 salário mínimo, que tenham cursado integralmente o ensino médio em escolas públicas.

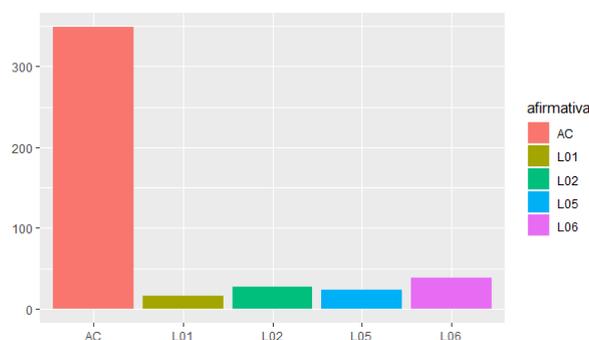


Figure 3. Formas de Ingresso na Instituição.

Vale realçar algumas características importantes das demais variáveis exploradas: ressalta-se o baixo aproveitamento com IRA's menores que dois (2), representando aproximadamente trinta e nove por cento (39%) das notas finais dos evadidos do curso em questão. Os demais aproveitamentos mantiveram proporções parecidas, com exceção daqueles com notas entre nove (9) e dez (10), com menos de um por cento (1%) do total de alunos. Em relação ao quantitativo de reprovações de disciplinas, enfatiza-se números precisos e intervalares, entre os quais se verifica que a maior proporção é de onze (11) a vinte (20) reprovações, por aluno, com aproximadamente vinte por cento (20%) do total. Em seguida, observa-se alunos com cinco (5) reprovações e os que tiveram de seis (6) a dez (10) reprovações, com aproximados dezoito por cento (18%) e dezoito por cento (18%), respectivamente. Outro aspecto que vale destacar, é que o percentual de alunos que não reprovaram nenhuma disciplina é, em torno, de dez por cento (10%).

Salienta-se, a distribuição de renda dos alunos, em que é realçado aqueles que ganham menos de quinhentos (500) reais mensais. Esse valor representa a renda de aproximadamente vinte e seis por cento (26%) dos alunos que ingressaram no curso, durante o período estudado. Outros valores com alta representatividade são os que ganham acima de três mil (3000) reais com (15%), entre quinhentos e um (501) e mil (1000) reais com (13%) e os que ganham entre mil e um (1001) e mil e quinhentos (1500) reais mensais com (12%), nessa ordem. Os demais valores são menos expressivos para o estudo.

4.2. Aplicação do algoritmo *K-means* sobre os dados

Foram definidos dois *clusters*, o *cluster 1*, representa os concludentes e o *cluster 2*, os evadidos do curso de CC da UFERSA campus Mossoró/RN, salientando que o nível de acurácia ficou em torno de noventa e três por cento (93%). O algoritmo foi executado oito (8) vezes para obtenção de dados confiáveis. A cada execução foi verificada a acurácia dos resultados, e por fim, quando os resultados se estabilizaram foi considerada a última iteração.

Na Figura 4, é representada uma visão geral da proporcionalidade e explicação de cada grupo. Enfatiza-se a distribuição nos *clusters* daqueles que evadiram em que se denota uma maior aglomeração dos dados, consequentemente mostrando maior tendência de explicação das variáveis. Na Tabela 1, são descritas as características dos *clusters* encontrados.

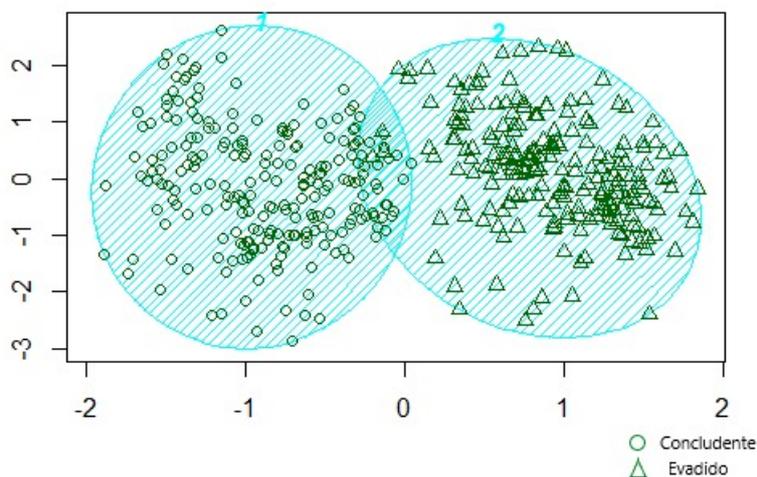


Figure 4. Representação dos *Clusters*.

Table 1. Descrição dos *Clusters*.

Variáveis	Cluster 1	Cluster 2
situação	concludente	evadido
morou_vila_acadêmica	sim	não
experiência_prévia	sim	não
relação_conjugal_pais	moram juntos	separados
cor_da_pele	branco	pardo
estado_origem	RN	RN
tipo_escola_ensino_médio	pública	privada
ação_afirmativa	AC	AC
IRA	entre 7 e 8	menos de 2
reprovações	5 reprovações	11 a 20 reprovações
renda	entre 1000 e 1500 R\$ mensal	menos de 500 R\$ mensal

O *cluster 1* representa os concludentes do curso de CC. Os alunos deste grupo caracterizam-se, por serem alunos que residiram na vila acadêmica, que tiveram experiências prévias na área (Computação) ao ingressar no curso, tinham seus pais morando

juntos no momento em que ingressaram no curso, tem cor da pele branca, naturais do estado do RN, cursaram integralmente o ensino médio em escola pública, ingressaram na instituição por meio de AC, concluíram com IRA entre sete (7) e oito (8), tiveram cinco (5) reprovações durante o curso e a renda mensal da família entre mil e um (1001) a mil e quinhentos (1500) reais.

O *cluster 2* caracteriza os evadidos do curso de CC. Os alunos deste grupo caracterizam-se, por serem alunos que não residiram na vila acadêmica, que não tiveram experiências prévias na área (Computação) ao ingressar no curso, seus pais eram separados no momento em que ingressaram no curso, tem cor da pele parda, naturais do estado do RN, cursaram integralmente o ensino médio em escola privada, ingressaram na instituição por meio de AC, concluíram com IRA menor que dois (2), tiveram em torno de onze (11) a vinte (20) reprovações durante o curso e a renda mensal da família menor que quinhentos (500) reais.

5. Considerações Finais

Buscou-se neste estudo, a verificação de que forma algumas variáveis influenciam no processo de evasão acadêmica no curso de CC da UFERSA, campus Mossoró/RN. As variáveis foram apresentadas na subseção 4.1. Desvendou-se por meio dos dois (2) grupos identificados, diferenças entre os alunos que conseguiram concluir e aquelas que evadiram no decorrer do curso. No segundo caso, há uma tendência de alunos que não residiram na vila acadêmica, não tinham experiências prévias na área (Computação) ao ingressar no curso, seus pais são separados, tendem a ter a cor da pele parda, são essencialmente naturais do estado do RN, cursaram integralmente o ensino médio em escolas públicas, tiveram muitas reprovações durante o curso e possuem uma renda mensal baixa.

Como exemplo de ação possível, propõe-se que, uma vez que os alunos sejam considerados de risco a evadir, eles sejam designados para um tutor, a fim de lhes fornecer apoio acadêmico e orientação para motivar e tentar evitar a evasão acadêmica. Finalmente, como próximo passo desta pesquisa, pretende-se realizar mais experimentos utilizando uma base de dados maior de alunos de Computação, e também, utilizando dados de alunos de outros cursos, para testar se os mesmos resultados de acurácia do algoritmo utilizado são obtidos com diferentes bases de dados.

Vale salientar as limitações do estudo, pois se identificou a falta de armazenamento de variáveis, tais como: as socioeconômicas e as longitudinais dos alunos, que poderiam ser mais explicativas para o problema da evasão acadêmica. Além disso, entre as informações que são arquivadas dos estudantes, constatou-se que existem dados ausentes.

Como trabalho futuro, destaca-se o seguinte: adicionar novas variáveis que possibilitem prever se o aluno irá evadir o mais rápido possível, pois, quanto mais cedo, melhor. Além disso, propor outras estratégias para ajudar os alunos identificados dentro do grupo de risco. Ademais, será desenvolvido um *plug-in* para ser integrado ao SIGAA da instituição, com o objetivo de alertar coordenadores de curso, docentes e os demais departamentos da instituição sobre os alunos que estão potencialmente em risco de desistir.

References

- Camilo, C. O. and Silva, J. C. d. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *Universidade Federal de Goiás (UFG)*, pages 1–29.
- da Silva Zanato, K. Y., Ventura, T. M., and Ribeiro, J. M. (2018). Análise da evasão de alunos da área de tecnologia da informação por meio de um banco de dados orientado a grafos. *Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação*, 1(8).
- do Nascimento, R. L. S., da Cruz Junior, G. G., and de Araújo Fagundes, R. A. (2018). Mineração de dados educacionais: Um estudo sobre indicadores da educação em bases de dados do inep. *RENOTE*, 16(1).
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37.
- Hoed, R. M. (2016). Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de computação. *Brasília, DF: Universidade de Brasília*.
- Júnior, O. d. G. F., Rodrigues, W. R. M., Barbirato, J. C. C., and de Barros Costa, E. (2019). Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de maceió. *RENOTE-Revista Novas Tecnologias na Educação*, 17(1):296–305.
- Manhães, L. M. B., Da Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., and Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)*, volume 1.
- Marques, L. T., Queiroz, P. G. G., De Castro, A. F., Marques, B. T., and Silva, J. C. P. (2019). Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um mapeamento sistemático da literatura. *RENOTE-Revista Novas Tecnologias na Educação*, 17(3):194–203.
- Martinho, V. R., Nunes, C., and Minussi, C. R. (2013). Prediction of school dropout risk group using neural network. In *2013 Federated Conference on Computer Science and Information Systems*, pages 111–114. IEEE.
- Nagai, N. P. and Cardoso, A. L. J. (2017). A evasão universitária: Uma análise além dos números. *Revista Estudo & Debate*, 24(1).
- PESSOAL, D. (2018). Instituto nacional de estudos e pesquisas educacionais anísio teixeira.
- Souza, S. (2008). *Evasão no ensino superior: um estudo utilizando a mineração de dados como ferramenta de gestão do conhecimento em um banco de dados referente à graduação de engenharia*. PhD thesis, Dissertação de Mestrado. Coppe–UFRJ, Rio de Janeiro.
- Webber, C. G., Zat, D., do Prado, M. d. F. W., et al. (2013). Utilização de algoritmos de agrupamento na mineração de dados educacionais. *RENOTE*, 11(1).
- Zhao, Y. (2011). R and data mining: Examples and case studies. *R and Data Mining*, pages 1–4.