

Aprendizado de Máquina Aplicado à Análise de Evasão no Ensino Superior

Mariana Andressa Luna Pinheiro¹, Josenildo Costa da Silva², Bruno Feres de Souza¹

¹Coordenação de Engenharia da Computação
Universidade Federal do Maranhão
São Luís – MA – Brasil

²Departamento Acadêmico de Informática
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão
São Luís – MA – Brasil

mariana.luna@aluno.ecp.ufma.br,bruno.feres@ufma.br,jcsilva@ifma.edu.br

Abstract. *Student dropout is a major problem and it is present in all educational contexts, causing great harm to the student, the society and the country. In order to combat it, one must first find out which students are most likely to dropout. In order to deal with this difficulty, in this work, it is sought, through the application of Machine Learning techniques, to identify the profile of the quitter. To do so, three learning algorithms were applied in a database with information from students of a Higher Education Institution in Brazil. The results of the experiments conducted here in the classification context were analyzed and discussed, indicating the potential of the use of Machine Learning in the identification of possible quitter.*

Resumo. *Evasão estudantil é um problema importante e presente em todos os contextos educacionais, causando grandes prejuízos ao aluno, à sociedade e ao país. Para combatê-lo, é preciso inicialmente descobrir quais discentes são mais propensos a evadir. A fim de lidar com esta dificuldade, neste trabalho, busca-se, por meio da aplicação de técnicas de Aprendizado de Máquina, identificar qual o perfil do aluno evasor. Para tanto, foram aplicados três algoritmos de aprendizagem em uma base de dados com informações de alunos de uma Instituição de Ensino Superior do Brasil. Os resultados dos experimentos aqui conduzidos no contexto de classificação foram analisados e discutidos, indicando o potencial da utilização de Aprendizado de Máquina na identificação de possíveis alunos evasores.*

1 - Introdução

A Educação, segundo o filósofo René Hubert, é um conjunto de ações que buscam inserir o indivíduo no contexto que o cerca. É por meio dela que se dão a transmissão e o aprimoramento do conhecimento, que a tecnologia se desenvolve e que é possível eliminar ou diminuir problemas como o trabalho infantil e a marginalização [Souza 2008].

Dentre os níveis de escolaridade, é no ensino superior que o indivíduo se profissionaliza e se torna atrativo ao mercado de trabalho. Tendo isto em vista, entre 2003 e 2014,

o Governo Federal realizou investimentos para melhorar as condições das Instituições Federais de Educação Superior, incrementando as universidades e os institutos já existentes e instaurando novos *campi* no interior dos estados [Gemaque e Souza 2016].

No entanto, a evasão têm representado um sério entrave ao avanço da Educação. Tal fenômeno configura-se como a decisão própria do discente de desvincular-se de seus estudos formais. De maneira mais geral, ela pode ser entendida como a interrupção do ciclo de estudos do aluno [Gaioso 2005]. Esta descontinuação pode ser via abandono de curso, desligamento institucional ou desistência completa do nível de ensino iniciado.

O problema da evasão não é recente e está presente em todos os contextos educacionais, em maior ou menor grau. Por este motivo, os estudos acerca do tema vêm se intensificando desde 1995, quando foi fundada a Comissão Especial de Estudos sobre Evasão, pelo Ministério da Educação. O interesse do Governo Federal em políticas de combate à evasão ocorre devido ao grande desperdício de recursos que este problema gera, incluindo professores, funcionários, equipamentos e espaço físico, entre outros.

Considerando que as instituições de ensino vêm armazenando eletronicamente dados acadêmicos e socioeconômicos de seus alunos, é possível realizar diversas investigações em busca de padrões e informações relacionadas com a condição de evasão. Como o processo exige a extração de conhecimento em grande volume de dados, o Aprendizado de Máquina (AM) apresenta-se como uma alternativa viável para a realização das análises devidas [Bishop 2006].

Assim, neste trabalho, busca-se estudar a problemática da evasão no contexto universitário através da utilização de técnicas de AM. Para tanto, foram utilizados dados de alunos de uma Instituição de Ensino Superior (IES) do Brasil. Os atributos extraídos dos dados deram origem a três bases distintas, sobre as quais foram aplicadas algoritmos de classificação capazes de distinguir alunos evasores de alunos não-evasores. O objetivo é verificar a adequação da aplicação de AM ao domínio, a fim de determinar precocemente quais alunos são mais propensos a evadir. Com isso, espera-se auxiliar a administração superior da instituição a tomar providências para evitar que estes alunos de fato evadam.

Este artigo se organiza como segue. Na Seção 2, alguns conceitos básicos da área de AM são apresentados. Na Seção 3, trabalhos relevantes ao tema são revisados. Na Seção 4, são descritos a base de dados e os algoritmos utilizados neste trabalho, assim como os experimentos conduzidos. Na Seção 5, são apresentados e analisados os resultados obtidos. Por fim, a Seção 6 apresenta as considerações finais do estudo e indica possíveis trabalhos futuros.

2 - Aprendizado de Máquina

Aprendizado de máquina é uma subárea da Inteligência Artificial que busca prover a um dado sistema a habilidade de aprender. Em outras palavras, é o estudo de como possibilitar uma máquina melhorar seu desempenho em determinada tarefa a partir de experiências, adquirindo assim conhecimento de forma automática [Facelli et al. 2011].

Dentre as estratégias disponíveis para o desenvolvimento de abordagens de AM, aquelas baseadas no conceito de indução estão entre as mais utilizadas. O aprendizado indutivo consiste em determinar relações gerais presentes nos dados analisando observações particulares dos mesmos. Neste contexto, duas vertentes de AM tem sido mais comu-

mente estudadas: supervisionada e não-supervisionada, ilustradas na Figura 1, extraída de [Facelli et al. 2011].



Figura 1. Hierarquia de AM.

No aprendizado supervisionado, exemplos rotulados são utilizados para aprender um modelo que consiga indicar o conceito subjacente aos dados. Com isso, pretende-se determinar o rótulo de exemplos não vistos anteriormente. Algoritmos de classificação consideram rótulos discretos nessa tarefa, enquanto algoritmos de regressão operam com rótulos contínuos. No aprendizado não-supervisionado, não há rótulos disponíveis para guiar o aprendizado. Ele deve, portanto, basear-se nos padrões implícitos dos dados a fim de explicar as observações. Algoritmos de agrupamento, de associação e de sumarização são abordagens dessa categoria.

Neste trabalho, o interesse é realizar a distinção entre alunos evasores e não-evasores. Portanto, a tarefa de aprendizado é de classificação. Na Seção 4, os algoritmos de classificação utilizados são descritos.

3 - Análise de dados educacionais

De acordo com o Censo da Educação Superior 2014, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira¹, há uma discrepância entre a quantidade de alunos ingressantes em Instituições de Ensino Superior no Brasil e a quantidade de alunos concluintes. Tal fato é ilustrado na Tabela 3, que fornece estatísticas gerais sobre a Educação Superior do Brasil. Percebe-se que a lacuna aumenta quando se analisa a relação entre matrículas ativas e concluintes em instituições federais.

Tabela 1. Estatísticas gerais da Educação Superior do Brasil- 2014

Estatísticas Básicas	Total geral	Pública				Privada
		Total	Federal	Estadual	Municipal	
Cursos	32 878	11 036	6 177	3 781	1 078	21 842
Matrículas	7 828 013	1 961 002	1 180 068	615 849	165 085	5 867 011
Ingresso Total	3 110 848	548 542	346 991	148 616	52 935	2 562 306
Concluintes	1 027 092	241 765	128 084	89 602	24 079	785 327

Para melhor entender cenários como o apresentado na Tabela 3, alguns estudos tem aplicado AM no contexto educacional (vide [Santos et al. 2016] e referências ali contidas). Como exemplo, os autores de [Brito et al. 2014] estudaram a relação entre a nota

¹Disponível em <http://portal.inep.gov.br>

de ingresso na universidade e o desempenho do estudante no primeiro período. Para isso, foram utilizados cinco classificadores de diferentes tipos, desde vizinhos mais próximos a redes neurais. Os resultados foram satisfatórios, mostrando que é possível estimar o desempenho dos alunos já no momento do ingresso, e assim tomar medidas para combater reprovações e, conseqüentemente, a evasão.

Já [Adeodato et al. 2014] avaliaram a qualidade do ensino secundário privado no Brasil. Para tal, basearam-se nas notas do Enem e do Censo Escolar e utilizaram uma árvore de decisão para gerar regras. Este estudo mostrou que aspectos econômico-financeiros são os fatores mais influentes, direta ou indiretamente, em uma boa qualidade na educação.

Buscando alertar professores sobre potenciais evasores no contexto de educação à distância, [Kampff et al. 2014] propuseram gerar regras de classificação. Para tanto, foi aplicado o algoritmo *RuleLearner* sobre dados pessoais e acadêmicos dos estudantes. Assim, alertas com base nas regras notificavam o professor, identificando o grupo de alunos e a regra específica. Foi possível então intervir em auxílio dos estudantes. Como resultado, houve o aumento dos índices de aprovação e redução nos índices de evasão.

Ainda, no trabalho de [Gonçalves 2015], três algoritmos de classificação foram comparados para identificar o perfil de estudantes mais propensos a evadir. Os dados utilizados foram de estudantes de um Instituto Federal de Educação, Ciência e Tecnologia do Brasil. Com este estudo, mostrou-se que, quando se trata deste tipo de dado, uma etapa muito importante é o pré-processamento, que se faz essencial para adequar os dados para utilização de classificadores.

O presente trabalho encorpa a literatura nacional no assunto, apresentando conjuntos de atributos originais para a caracterização dos alunos e mostrando resultados promissores em dados de uma IES.

4 - Experimentos

Os experimentos conduzidos durante este trabalho consistem da aplicação de algoritmos de AM do paradigma supervisionado a uma base de dados de alunos de uma IES. O objetivo dos experimentos é verificar se classificadores gerados por esses algoritmos são capazes de distinguir entre alunos que concluem seus cursos de ensino superior (rótulo *Formado*) e alunos propensos a evadir (rótulo *Evasão*), considerando informações acadêmicas, sociais e econômicas dos mesmos.

Bases de dados

Consideraram-se inicialmente os dados brutos utilizados em trabalho anterior [Gonçalves 2015], oriundos do Sistema Acadêmico do Instituto Federal de Educação Ciência e Tecnologia do Maranhão - IFMA. Estes dados contêm informações acerca de aspectos pessoais, socioeconômicos e acadêmicos dos alunos no período de 1992 a 2013. A fim de adequar os dados às análises feitas neste trabalho, foram selecionados apenas os alunos em situação de *Evasão* ou *Formado*, que são as classes de interesse. Desconsideraram-se, ainda, alunos com notas não cadastradas no sistema. Assim, obteve-se um conjunto de 966 estudantes, sendo 70% evadidos e 30% formados². A fim de anali-

²Os dados utilizados aqui podem ser obtidos por comunicação pessoal com os autores deste trabalho.

sar diferentes visões dos dados, três conjuntos de atributos foram manualmente definidos, originando bases distintas, conforme descrito a seguir.

A primeira base é composta por 5 atributos extraídos no momento da matrícula do aluno. São eles:

- Estado civil (**At01**): este atributo foi considerado devido a indivíduos em diferentes estados civis disporem de formas diversas de aproveitamento do tempo, o que pode influenciar no seu desempenho acadêmico.
- Forma de ingresso (**At02**): esta informação foi incluída porque indivíduos que ingressam na universidade de forma distinta podem estar preparados de forma diferente para a vida acadêmica. No caso foram considerados Seleção, Transferência ex officio, Transferência Externa, Processo Seletivo, Obtenção de Novo Título, Vestibular/Enem, Outras Formas de Ingresso, Reopção de Curso, Certificação Enem, Reingresso Matrícula de Graduado
- Turno do curso (**At03**): em turnos distintos ou integral, os alunos devem ser capazes de gerir seu tempo de forma diferente ou mesmo se deslocar para a instituição em condições mais ou menos acessíveis, o que pode influenciar em seu desempenho acadêmico.
- Sigla do curso (**At04**): este atributo foi incluído devido a diferentes cursos podem apresentar diferentes taxas de evasão.
- Sexo (**At05**): este atributo foi incluído para verificar se há relação entre o gênero do aluno e a condição de evasão.

A segunda base é composta por 12 atributos obtidos durante a vida acadêmica do aluno. São eles:

- Coeficiente de rendimento - CR (**At06**): este atributo foi incluído por representar o desempenho geral do aluno em sua vida acadêmica.
- Tempo de curso (**At07**): atributo calculado pela diferença entre o ano em que o aluno ingressou no IF e seu ano atual no sistema. Esta informação foi incluída pela possibilidade de haver um período mais provável de evasão.
- Média de reprovações por falta (**At08**): para a obtenção deste atributo, foram calculadas as médias de reprovações por falta em relação à quantidade de disciplinas cursadas, por ano. Destas médias, foi calculada a média geral, representando as reprovações por falta durante todo o tempo de curso. Foi considerado por ser um indicativo do desinteresse do estudante no curso.
- Desvio padrão das reprovações por falta (**At09**): calculado sobre as quantidades médias de reprovação por falta em cada ano. O atributo foi considerado para identificar médias de reprovação por falta altas ou baixas isoladas, que podem não fazer parte do perfil do aluno, sendo provavelmente devidas a fatores externos.
- Média de reprovação por nota (**At10**): análogo ao cálculo de reprovação por falta. Este atributo foi incluído por ser um indicativo de dificuldade do estudante em acompanhar o curso.
- Desvio padrão de reprovação por nota (**At11**): análogo ao desvio padrão de reprovações por falta. Este atributo foi incluído por ser um indicativo de dificuldade do estudante em acompanhar certas disciplinas do curso.
- Média de presença (**At12**): da mesma forma que as reprovações, esta média foi calculada sobre as médias de cada ano. Este atributo foi incluído por indicar interesse do aluno em seguir com o curso.

- Média dos desvios padrões das presenças (**At13**): para a obtenção deste atributo foram calculados os desvios padrões das presenças em cada ano e, sobre estes, a média foi computada. Este atributo foi considerado por poder denunciar estudantes que não são assíduos em todas as disciplinas, indicando desinteresse em determinadas disciplinas do curso.
- Média de disciplinas (**At14**): este atributo consiste na quantidade média de matérias cursadas a cada ano pelo estudante. Foi considerado pela possibilidade de haver uma quantidade adequada de disciplinas cursadas por ano para um bom acompanhamento do curso.
- Desvio padrão da quantidade de disciplinas (**At15**): este atributo foi considerado para discriminar estudantes que não possuem regularidade na quantidade de disciplinas que cursa a cada ano, podendo indicar falta de comprometimento com o curso.
- Média das notas (**At16**): cálculo análogo ao das presenças. Foi considerado pelo mesmo motivo que o CR.
- Média dos desvios padrões das notas (**At17**): cálculo e interesse análogos ao da média dos desvios padrões das presenças.

A terceira base de dados constitui-se da junção das duas bases acima descritas, o que resulta num total de 17 atributos, mais a classe.

Algoritmos

Para a classificação das bases consideradas, foram utilizados os seguintes algoritmos de AM [Bishop 2006]: *Naive Bayes*, *Árvore de Decisão* e *Support Vector Machines*, implementados na linguagem R utilizando os pacotes `e1071` e `rpart`, com os parâmetros padrão. A seguir, cada uma dessas abordagens é brevemente comentada.

Naive Bayes (NB) é um classificador de abordagem probabilística, motivo pelo qual foi escolhido para este trabalho. Utiliza o teorema de Bayes para determinar a probabilidade de a instância pertencer a cada classes e supõe que os atributos são independentes entre si. Trata-se de um algoritmo robusto a ruídos isolados e capaz de lidar com valores faltantes e atributos irrelevantes. Pode apresentar deterioração de desempenho caso a correlação entre os atributos seja alta.

Support Vector Machines (SVMs) são classificadores do paradigma estatístico. Sua abordagem consiste em encontrar a fronteira ótima de separação das classes. Para isso, são utilizados os vetores de suporte, que são os exemplos mais próximos à borda de separação. Este algoritmo foi escolhido por apresentar desempenho considerado adequado em diversos domínios.

Árvores de Decisão (AD) são uma das técnicas de classificação mais utilizadas. Este classificador atua montando uma árvore de decisão a partir dos dados, estrutura que se define recursivamente como um nó folha ou um nó decisão: o primeiro contém a classe correspondente aos exemplos e o segundo, um teste sobre algum atributo. A técnica de geração é simples e pode ser aplicada em qualquer domínio. Como a classificação é obtida da própria estrutura do modelo, sua compreensão é fácil, e por isso este classificador foi escolhido.

Metodologia

Após a realização do pré-processamento dos dados, restaram 966 exemplos para o estudo. Sobre esses dados foi aplicada a técnica de validação cruzada balanceada para a mensurar a qualidade preditiva dos modelos gerados pelos algoritmos. Utilizaram-se $k = 10$ partições de treinamento e teste, para cada uma das três bases de dados consideradas. Posteriormente, reuniram-se os resultados de todas as partições, montando novamente o conjunto inicial, e compararam-se os resultados com a classes reais dos exemplos. Com isso, foi possível medir o desempenho dos algoritmos utilizando as métricas acurácia, sensibilidade, especificidade e precisão. Nota-se que aqui a classe *Evasão* é a positiva, enquanto a classe *Formado* é a negativa.

5 - Resultados e discussão

Como ilustrado na Tabela 2, somente com as informações da matrícula dos indivíduos, nenhum dos classificadores chegou à acurácia de 80%. No entanto, todos se saíram melhor do que apenas prever a classe majoritária dos dados, cuja acurácia corresponderia a 70%. Observa-se também que a acurácia e a sensibilidade da árvore de decisão foram superiores às dos demais classificadores. Em outras palavras, a Árvore de Decisão classificou mais alunos corretamente, dentre os quais identificou melhor os discentes evasores. Entretanto, sua especificidade se mostrou inferior. Isto indica que a AD considerou evasores estudantes concludentes em maior proporção que os demais métodos.

Tabela 2. Desempenhos dos algoritmos considerando a base de dados com atributos da matrícula.

Métrica	SVM	NB	AD
Acurácia	76.71%	75.67%	77.54%
Especificidade	79.44%	68.29%	55.40%
Sensibilidade	75.55%	78.79%	86.89%
Precisão	89.69%	85.46%	82.17%

Analisando como um todo os resultados de classificação para a base de dados com atributos obtidos no momento da matrícula, observa-se que os algoritmos SVM e NB foram melhores, proporcionalmente em relação a AD, em identificar discentes que não fazem parte do grupo de risco de evasão.

Já com os dados acadêmicos, houve um acréscimo geral na capacidade preditiva para os classificadores considerados (Tabela 3), indicando que o desempenho do estudante durante o curso é mais relevante para seu sucesso do que as condições no momento de matrícula. O SVM apresentou os melhores resultados, para todas as métricas. Quanto aos demais algoritmos, percebe-se que o NB e a AD estão empatados quanto a quantidade de métricas em que foram superiores em relação ao outro. Desta forma, o usuário de sistemas baseados em AM no contexto educacional deve definir quais métricas são mais relevantes para seu caso.

Como discutido anteriormente, é preciso determinar o máximo de alunos propensos a evadir, para que medidas contrárias a isto sejam tomadas. Por outro lado, não é interessante intervir em estudantes que muito provavelmente vão concluir o curso. Portanto,

Tabela 3. Desempenhos dos algoritmos considerando a base de dados com atributos acadêmicos.

Métrica	SVM	NB	AD
Acurácia	98.65%	91.72%	97.10%
Especificidade	98.95%	98.61%	97.56%
Sensibilidade	98.53%	88.81%	96.91%
Precisão	99.55%	99.34%	98.94%

a sensibilidade e a especificidade são as métricas que mais contribuem para a proposta do trabalho. No entanto, o NB é mais específico e a árvore de decisão mais sensível. Assim, como o objetivo da classificação é, principalmente, identificar estudantes evasores, escolhe-se o que tenha maior taxa de acertos positivos, no caso a AD.

Com o último conjunto de dados, que considera tanto os atributos da matrícula quanto os acadêmicos, no geral houve melhora dos desempenhos, como consta na Tabela 4. Como se observa das Tabelas 3 e 4, a AD apresentou uma pequena queda de desempenho com a junção de todos os dados em relação a somente os dados acadêmicos. Já o NB foi o algoritmo que apresentou maior melhora no resultado, pois todas as suas métricas apresentaram desempenho superior. A melhora no desempenho do NB quando da junção das duas bases pode indicar que os atributos da matrícula e os acadêmicos estão pouco correlacionados. Do contrário, esperaria-se queda no desempenho deste classificador.

Tabela 4. Desempenhos dos algoritmos considerando a base de dados com atributos da matrícula e acadêmicos.

Métrica	SVM	NB	AD
Acurácia	98.76%	92.86%	97.00%
Especificidade	99.65%	99.30%	97.21%
Sensibilidade	98.38%	90.13%	96.91%
Precisão	99.85%	99.67%	98.80%

A fim de discutir sobre os fatores que, potencialmente, mais influenciam na distinção entre alunos evadidos e formados, foram geradas Árvore de Decisão para cada um dos três conjunto de dados. Elas são exibidas nas Figuras 2(a), 2(b) e 2(c). Como mencionado anteriormente, a vantagem do algoritmo de AD é a legibilidade do modelo gerado, pois a explicação da classificação é a própria estrutura da árvore

Ao se analisar a AD da figura 2(a), percebe-se a falta de pureza nas folhas. Isto indica que as decisões tomadas não particionam muito bem os dados de classes distintas, como já esperado, pois somente informações como curso, turno ou sexo não devem ser suficientes para gerar regras de decisão confiáveis sobre evasão. Este resultado condiz com o apresentado na tabela 2.

Quando atributos relativos ao desempenho dos estudantes são utilizados, como consta na Figura 2(b), as folhas resultantes são mais puras, pois as decisões dividiram melhor os dados. Observa-se que o atributo de tempo de curso (At07) possui alto ganho de

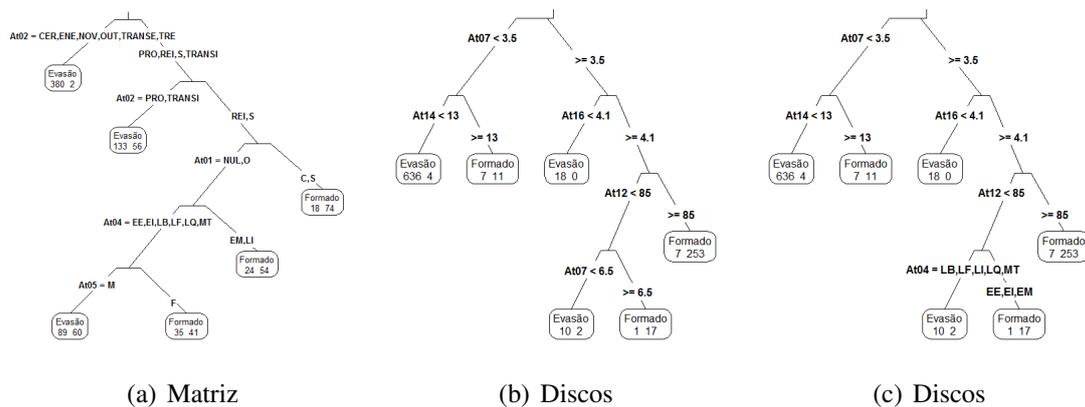


Figura 2. Árvores de decisão geradas utilizando diferentes conjuntos de atributos.

informação sobre a classe. Esta condição é esperada, pois para haver conclusão do curso é preciso que o aluno tenha concluído o tempo do seu curso. Outro atributo relevante é o CR (At16). A decisão de $CR < 4.1$ leva a uma folha pura. Assim, estudantes, mesmo que com um tempo de curso por volta da metade, que apresentam CR baixo são fortes candidatos à evasão.

Na Figura 2(c), tem-se a representação da árvore gerada com os atributos da matrícula e acadêmicos. Observa-se que as folhas possuem a mesma quantidade de exemplos. No segundo e terceiro níveis, tratam-se dos mesmos estudantes classificados em 2(a). O que difere as duas árvores é a última decisão, que é baseada em atributos diferentes. Porém, é provável que se tratem dos mesmos exemplos, pois os cursos que levam à direita em 2(c) possuem tempo de conclusão maior, que é a decisão tomada pelo At07, em 2(b).

6 - Conclusão

De acordo com os resultados obtidos, é possível observar como os algoritmos de AM conseguem prever adequadamente casos de evasão. De posse desta informação, pretende-se integrar ao Sistema Acadêmico da IES um módulo baseado em aprendizado de máquina que auxilie coordenadores e professores a identificar potenciais evasores a fim que possam acompanhá-los de forma mais direta, buscando meios de mantê-los na instituição. Verifica-se ainda que informações acadêmicas foram mais relevantes para a classificação, mas em geral melhores desempenhos são obtidos quando estes dados se combinam com os da matrícula. Ressalta-se que a forma como os atributos são calculados é contribuição original deste trabalho.

Ademais, das árvores de decisão geradas para cada base de dados, observam-se desempenhos melhores que os obtidos pela classificação pela classe majoritária. Isto pode indicar que regras com boa capacidade preditiva estejam presentes no conjunto. Esta investigação também será objeto de trabalho futuro. Por fim, a fim de realizar uma comparação mais completa dos resultados, vai-se empreender uma análise estatística dos resultados, além de uma investigação qualitativa dos mesmos.

Referências

- Adeodato, P. J., Santos Filho, M. M., and Rodrigues, R. L. (2014). Predição de desempenho de escolas privadas usando o enem como indicador de qualidade escolar. Em *Simpósio Brasileiro de Informática na Educação-SBIE*, volume 25, páginas 891–895.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Brito, D. M., de Almeida Júnior, I. A., Queiroga, E. V., and Rego, T. G. (2014). Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. Em *Simpósio Brasileiro de Informática na Educação-SBIE*, volume 25, página 882.
- Facelli, K., Lorena, A. C., ao Gama, J., and de Carvalho, A. (2011). *Inteligência Artificial. Uma Abordagem de Aprendizado de Máquina*. LTC.
- Gaioso, N. P. L. (2005). O fenômeno da evasão escolar na educação superior no brasil. Dissertação de mestrado, Universidade Católica de Brasília, Brasília.
- Gemaque, L. S. B. and Souza, L. G. (2016). Diplomação, retenção e evasão: estudo com enfoque na evasão dos cursos de graduação na universidade federal do maranhão no período de 2008 a 2010. *Ensino e multidisciplinaridade*, 2(1):84–105.
- Gonçalves, T. C. (2015). Análise da evasão no ensino superior do instituto federal de educação, ciência e tecnologia do maranhão utilizando técnicas de mineração de dados. Monografia (Sistemas de Informação), Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, São Luís, Brasil.
- Kampff, A. J. C., Ferreira, V. H., Reategui, E. B., and Lima, J. V. d. (2014). Identificação de perfis de evasão e mau desempenho para geração de alertas num contexto de educação á distância. *RELATEC: Revista Latinoamericana de Tecnología Educativa*, 13(2):61–76.
- Santos, R., Pitangui, C., Vivas, A., and Assis, L. (2016). Análise de trabalhos sobre a aplicação de técnicas de mineração de dados educacionais na previsão de desempenho acadêmico. Em *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 5, páginas 960–969.
- Souza, C. R. T. (2008). A educação não-formal e a escola aberta. Em *Anais do VIII Congresso Nacional de Educação (EDUCERE)*. Editora PUCRS.