

COMPORTAMENTO DE ESTUDANTES EVADIDOS DE CURSOS TÉCNICOS: UM ESTUDO UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Renata Gomes Cordeiro¹

Munir de Sá Mussa²

Henrique Rego Monteiro da Hora³

RESUMO

Dada a problemática da evasão e a quantidade de dados armazenados pelas instituições de ensino, através da utilização da mineração de dados torna-se possível extrair características que descrevem os alunos. O objetivo do presente trabalho é aplicar a mineração de dados para identificação de comportamento dos alunos evadidos em cursos de nível técnico. Para o cumprimento do objetivo é realizado um estudo de caso no Instituto Federal Fluminense. São executadas as etapas de obtenção, tratamento, higienização e mineração dos dados. Na última etapa é utilizado o método de classificação J48. A partir das árvores geradas pelo método foi possível identificar que em sua maioria, os alunos evadidos são descritos pelo desempenho ou frequência nas disciplinas dos primeiros módulos do curso. Os resultados são considerados satisfatórios por terem permitido realizar análises que de fato descrevem comportamentos de alunos que evadiram na amostra analisada.

Palavras-chave: Mineração de dados; Evasão; Ensino Técnico.

ABSTRACT

Given the problem of evasion and the amount of data stored by educational institutions, through the use of data mining in the databases of these institutions it is possible to extract characteristics that describe the students. The objective of the present work is to apply the data mining to identify the behavior of the students evaded in courses of technical level. To accomplish the objective, a case study is carried out at the Federal Fluminense Institute. The steps of obtaining, treating, sanitizing and mining the data are performed. In the last step the classification method J48 is used. From the trees generated by the method it was possible to identify that the majority of the students of the evaded profiles are described by the performance or frequency in the subjects of the

¹ Mestrado em Sistemas Aplicados à Engenharia e Gestão (SAEG) do Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF). E-mail: renata.cordeiro@iff.edu.br

² Mestrado em Sistemas Aplicados à Engenharia e Gestão (SAEG) do Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF). E-mail: munirmussa@gmail.com

³ Mestrado em Sistemas Aplicados à Engenharia e Gestão (SAEG) do Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF). E-mail: dahora@gmail.com

first modules of the course. The results are considered satisfactory because they allowed to carry out analyzes that in fact describe profiles of students that evaded in the sample analyzed.

Keywords: Data mining; Dropout; Technical education.

INTRODUÇÃO

A informatização na área da educação, assim como em diversas outras áreas, deve ser um aliado não só na eficiência dos serviços, mas, também, na eficácia de planos e políticas que buscam aprimoramento e melhores resultados. Na educação, além do seu objetivo final de levar conhecimento e ensinar, busca-se cada vez mais o desenvolvimento de políticas focadas na redução da evasão escolar. Sendo os dados armazenados sobre toda a trajetória escolar dos alunos uma fonte de informação relevante para o desenvolvimento da gestão escolar.

De acordo com Souza (2014), a evasão é um tema discutido em todo sistema educacional brasileiro, sendo necessário propor estratégias que garantam uma educação inclusiva e que busquem fatores propícios a conclusão do curso por parte do aluno. Cruz (2013) afirma que entre os diversos aspectos apresentados para a evasão dos estudantes, estão fatores que têm ligação com a complexidade da vida pessoal, familiar e financeira. Além disso, existem as instituições responsáveis pela educação e as políticas sociais que nem sempre atendem às necessidades dos estudantes.

Dada a problemática da evasão e a quantidade de dados armazenados pelas instituições de ensino, através da utilização da mineração de dados nas bases de dados dessas instituições, torna-se possível extrair características que descrevem os alunos. E apesar da questão da evasão ser multifatorial, a identificação dessas características é capaz de contribuir para o direcionamento da gestão escolar.

Figueiredo e Salles (2017) ressaltam que indicadores de permanência e evasão podem fornecer informações relevantes às pesquisas que visam avaliar a eficiência e a eficácia de programas governamentais que buscam a ampliação da oferta de cursos técnicos. Porém os dados estatísticos não são suficientes, é necessário extrair informações qualitativas que tragam um mapeamento dos alunos e uma realidade social mais complexa. Os autores afirmam que a evasão é um problema enfrentado pela maioria das instituições de ensino, oriundo de diversos fatores e que quanto mais precoce for identificado o risco da evasão, maiores são as possibilidades de sucesso das

políticas de permanência escolar. E para a identificação do risco de evasão, a exposição dos fatores que podem levar a isso torna-se uma forte aliada.

A partir das transformações econômicas e, conseqüentemente, no mercado de trabalho, a Educação Profissional tem estado cada vez mais em evidência. Tendo governos e empresas constituindo alianças no sentido de criar e manter cursos que, de alguma maneira, possam suprir a demanda de mão de obra para o desenvolvimento do país, contribuindo, igualmente, para a elevação do nível de escolarização dos trabalhadores (FIGUEIREDO; SALLES, 2017).

O ensino técnico é uma alternativa, principalmente aos jovens, com objetivo de inserir-se no mercado de trabalho como mão de obra qualificada de forma mais rápida. Porém, assim como cresce a demanda por cursos técnico, cresce também o índice de evasão. Souza (2014) acredita que ao evadir, o aluno está desistindo do que por alguma razão não atendeu às suas expectativas, e que isso tem um significado forte na vida do educando e conseqüentemente, da escola.

De acordo com Araújo e Santos (2012), a baixa formação qualificada e falta de habilitação profissional que atualmente existe no mercado de trabalho se deve significativamente ao problema do acesso e da permanência do cidadão em instituições que proporcionem formação qualificada e isso constitui um problema de ordem democrática. As autoras afirmam ainda que por consistir em um desestímulo aos estudos, para compreender a evasão é necessário um estudo profundo em várias perspectivas: aluno, escola e sociedade.

Em seu trabalho Figueiredo e Salles (2017) afirmam que estudos têm revelado a insuficiência de esforços, oriundos das mais diversas esferas de atuação das escolas, no sentido de pensar projetos e desenvolver ações que favoreçam a permanência dos estudantes nos cursos. Mesmo sabendo que será impossível impedir a evasão na sua totalidade, sendo esta com origem na própria estrutura do sistema econômico em que estamos inseridos.

A evasão no ensino técnico tem se tornado um tema em evidência a partir da percepção de que a avaliação estatística de oferta e demanda por esses cursos não são suficientes para analisar o seu resultado para a sociedade. É necessário investigar a proporção de alunos que se forma e as causas da não formação. Como afirmado por Veiga e Bergiante (2016), o desperdício gerado pela evasão escolar não é apenas de

ordem financeira, mas também social, por levar o indivíduo a um profundo processo de exclusão.

O objetivo do presente trabalho é aplicar a mineração de dados para identificação do comportamento dos alunos evadidos em cursos de nível técnico. Sua importância se dá no fato de aplicar a técnica de mineração de dados para extrair potenciais informações da grande quantidade de dados armazenados pela instituição com o intuito de explicitar características comuns aos alunos que evadem, permitindo, por parte da instituição de ensino a criação de políticas focadas em alunos que possuam as características identificadas.

METODOLOGIA DA PESQUISA

1.1. População e amostra

A pesquisa é realizada no Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFFluminense). De acordo com o portal institucional (*iff.edu.br*), o IFFluminense está presente em 11 municípios, é composto por 12 *campi*, um Polo de Inovação, um Centro de Referência em Tecnologia, Informação e Comunicação na Educação e a Reitoria. Conta ainda com polos de Educação a Distância em 4 municípios.

Atualmente, o IFFluminense oferta cursos técnicos nas modalidades: subsequente, para os que já concluíram o ensino médio, concomitante, para os que estão cursando o ensino médio e o integrado com o ensino médio.

São considerados como amostra os cursos técnicos na modalidade subsequente e concomitante. São utilizados os dados dos alunos que estão matriculados, concluíram ou evadiram o curso e que se matricularam a partir de 2014, pois nesse ano foi iniciada a utilização do sistema de inscrição *online*. Nesse sistema, qualquer cidadão que queira participar de algum processo seletivo para estudar no IFFluminense deve realizar o cadastro no sistema de inscrições, informando dados pessoais e respondendo ao questionário socioeconômico, sendo um cadastro único por pessoa que pode ser atualizado a qualquer momento.

1.2. Procedimentos técnicos

A metodologia da presente pesquisa baseia-se nas macroetapas consolidadas no trabalho de Cordeiro, Mussa e Hora (2017). Para o cumprimento do objetivo desse trabalho são seguidas as etapas apresentadas na Figura 1.

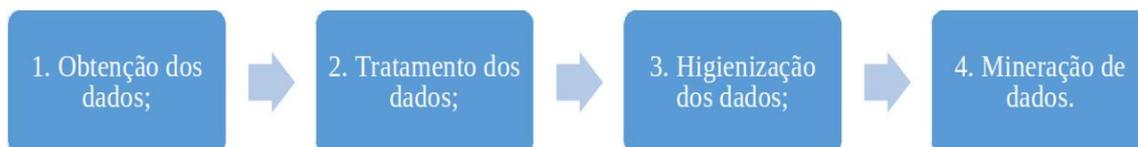


Figura 1 - Etapas da pesquisa (CORDEIRO; MUSSA; HORA, 2017).

1.3. Obtenção de Dados: Na primeira etapa é realizada obtenção dos dados. Nesse trabalho os dados são oriundos de duas bases distintas: a base do sistema acadêmico e a base do sistema de inscrições *online*. O Quadro 1 apresenta os dados retirados de cada uma das bases de dados utilizadas.

Quadro 1 - Dados retirados da base de dados do sistema acadêmico e do sistema de inscrições.

Dados do sistema acadêmico	Dados do sistema de inscrições
<ul style="list-style-type: none"> • Matrícula • Data da matrícula • Situação da matrícula • CPF • Data de nascimento • Grau de instrução • Descrição do curso • Modalidade do curso • Turno • <i>Campus</i> • Nota • Quantidade de faltas • Disciplina • Quantidade de aulas 	<ul style="list-style-type: none"> • CPF • Sexo • PCD (pessoas com deficiência visual, auditiva, física ou intelectual) • Estado civil • Cor • Tipo de escola que cursou o nível escolar anterior • Turno em que cursou o nível escolar anterior • Situação do curso superior • Motivo de escolha do curso • Exercício de atividade remunerada • Renda mensal familiar • Participação na economia familiar • Utilização de computador • Data de preenchimento do formulário • Data de atualização do formulário

Fonte: Elaboração própria.

Como apresentado no Quadro 1, foram extraídos 14 atributos da base de dados do sistema acadêmico e 15 da base do sistema de inscrições. Da base de dados do sistema acadêmico, foram retirados dados pessoais e relacionados ao desempenho acadêmico. Do sistema de inscrições, foram retirados dados pessoais que não constavam para todos os registros retirados do sistema acadêmico e também aqueles referentes ao questionário socioeconômico que os alunos respondem ao se cadastrarem no sistema de inscrições. Após obter os dados das duas bases, foi necessário uni-los em uma base única, e para isso, o atributo CPF foi utilizado como referência. As etapas seguintes são realizadas nessa nova base.

1.4 Tratamento de Dados: Na segunda etapa, são realizados tratamentos na base de dados. Os tratamentos foram:

- Cálculo da proporção de ausência em cada disciplina, dividindo o número de faltas pelo número de aulas ministradas;
- Cálculo da idade através da data de nascimento;
- Cálculo do intervalo entre a data de matrícula e data da última atualização do questionário socioeconômico.

Neste último item são excluídos os registros que possuem um intervalo maior que dois anos. Como o questionário reflete dados variáveis da vida pessoal do aluno, como por exemplo: renda mensal e participação na economia familiar, avaliou-se que restringindo esse intervalo a dois anos haveria menos discrepância da realidade do aluno entre o momento em que atualizou o questionário pela última vez e o momento da matrícula.

Nessa etapa a base de dados também é dividida em outras bases, separando os dados por *campus* e pela modalidade do curso. A Figura 2 apresenta as bases formadas após a divisão.

Como apresentado na Figura 2, a partir dos dados disponíveis, foram formadas bases de dados para 10 *campi*, sendo que para os *campi* Campos Centro e Quissamã foram formadas duas bases uma com dados dos alunos dos cursos concomitantes e outra com dados dos alunos dos cursos subsequentes.

Outro importante tratamento realizado nesta etapa é a normalização dos dados que descrevem as disciplinas cursadas e a proporção de ausência em cada disciplina. A nota

de cada disciplina é transformada em um atributo assim como a proporção de ausência em cada disciplina.

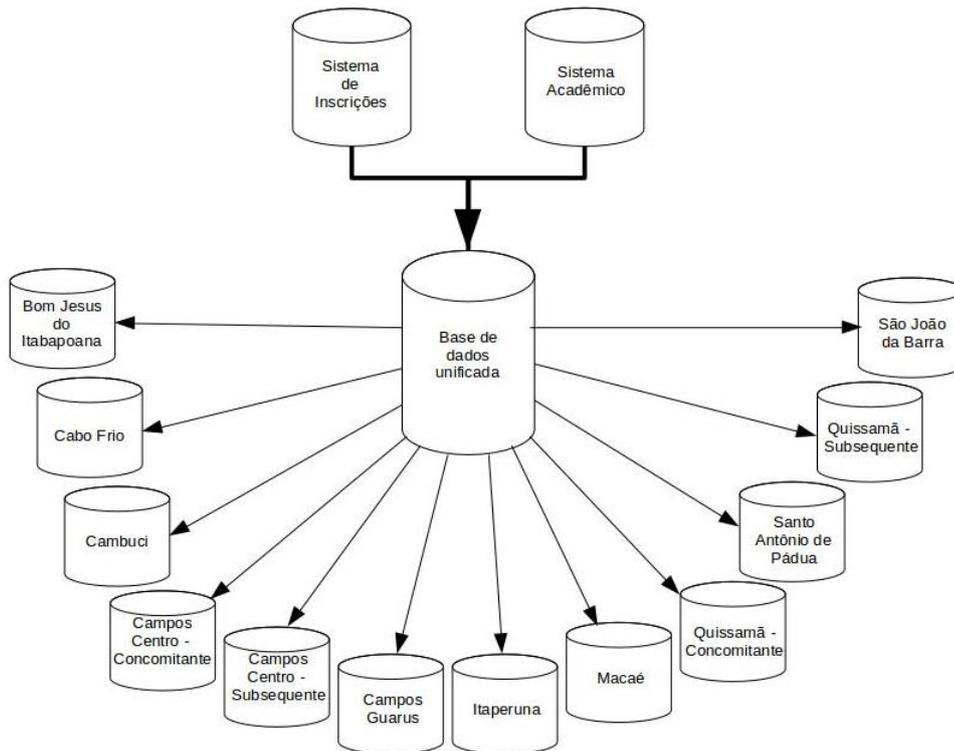


Figura 2 - Composição do banco de dados do estudo proposto. Fonte: Elaboração própria.

Ao final da segunda etapa, o banco de dados resultante contém os dezessete atributos, entre esses dados está o dado classificador que é aquele que descreve a situação do aluno, podendo ser matriculado, concluído ou evadido. Nesta etapa, a base ainda contém a modalidade do curso: concomitante ou subsequente; a nota e a proporção de ausência obtida em cada disciplina. Sendo assim, a quantidade de atributos para cada *campus* variou de acordo com a quantidade de disciplinas.

1.5. Higienização dos Dados: Na terceira etapa ocorre o pré-processamento dos dados que inclui a higienização dos mesmos. A higienizações realizadas foram:

- Exclusão dos registros com dados faltantes
- Exclusão de dados discrepantes:
 - idades menores que 15;

- grau de escolaridade igual a ensino fundamental completo, ensino fundamental incompleto, não declarado ou alfabetizado.

1.6. Mineração dos Dados: Na quarta etapa também foram realizadas adaptações nos dados, transformando alguns dados nominais em numéricos. Essas alterações foram feitas para três atributos: grau de instrução, renda mensal familiar e exercício de atividade remunerada. Visto que as opções de valores para esses atributos são graduais, a transformação para numérico permite que o método classificador da etapa de mineração gere resultados traçando valores mínimos ou máximos para esses atributos nas classificações.

Na etapa de mineração de dados é utilizado o método de classificação J48. Sendo esse um método de árvore de decisão implementado com base no algoritmo C4.5 proposto por Quinlan (1987). Nesse método, as folhas são as classes e os nós são os atributos. Os ramos da árvore são gerados de acordo com os valores possíveis de cada atributo. Dessa forma, a classificação de uma instância é realizada a partir da folha alcançada percorrendo a árvore de acordo com os valores de seus atributos. Os nós determinam o atributo que será avaliado e o percurso no nível seguinte da árvore é definido a partir do valor do atributo do nó na instância analisada.

Para análise de confiabilidade dos resultados, a etapa de classificação é realizada duas vezes, utilizando métodos diferentes em cada uma delas. Além do método J48 é utilizado para o reteste o método JRip, baseado no algoritmo *Ripper (Repeated Incremental Pruning to Produce Error Reduction* ou Poda Incremental Repetida para Produzir Redução de Erro) proposto por Cohen (1995). O método JRip foi escolhido por também ser amplamente utilizado em trabalhos que investigam a evasão a partir da mineração de dados educacionais (DEKKER; PECHENIZKIY; VLEESHOUWERS, 2009; JIMÉNEZ-GÓMEZ et al., 2015; MÁRQUEZ-VERA et al., 2016; PRADEEP; DAS; KIZHEKKETHOTTAM, 2015).

Cordeiro, Mussa e Hora (2017) apontam os métodos J48 (árvore de decisão) e JRip (regras de indução) como os mais utilizados na mineração de dados educacionais, então para garantir a confiabilidade dos dados, é adotada a técnica de teste-reteste (GUTTMAN, 1945), onde o mesmo objeto é submetido à dois tratamentos e seus resultados são comparados. O JRip nesta pesquisa não tem outra função senão apenas

aferir a confiabilidade da base de dados, uma vez que o J48 é adotado como algoritmo para gerar os resultados a serem analisados.

Para a execução do método é utilizado o método *k-fold cross validation*. Nesse método, os dados são divididos em k conjuntos mutuamente exclusivos. A validação é realizada k vezes, de forma que a cada validação um conjunto diferente é utilizado como teste e os outros k-1 conjuntos são utilizados para treinamento. O resultado final é obtido a partir da média de todas as k validações.

RESULTADOS

2.1. A evasão no IFFluminense

As Figuras 3 e 4 mostram o quantitativo de alunos concluintes e evadidos que ingressaram a partir de 2014 no IFFluminense nos cursos concomitantes e subsequentes, respectivamente. Não foram considerados os dados do ano de 2017 por não estarem disponíveis. Os dados foram retirados do Sistema Unificado de Administração Pública (SUAP), utilizado no IFFluminense. Nas Figuras 3 e 4, no eixo das abscissas são apresentados os cursos e o quantitativo de alunos evadidos e concluintes.

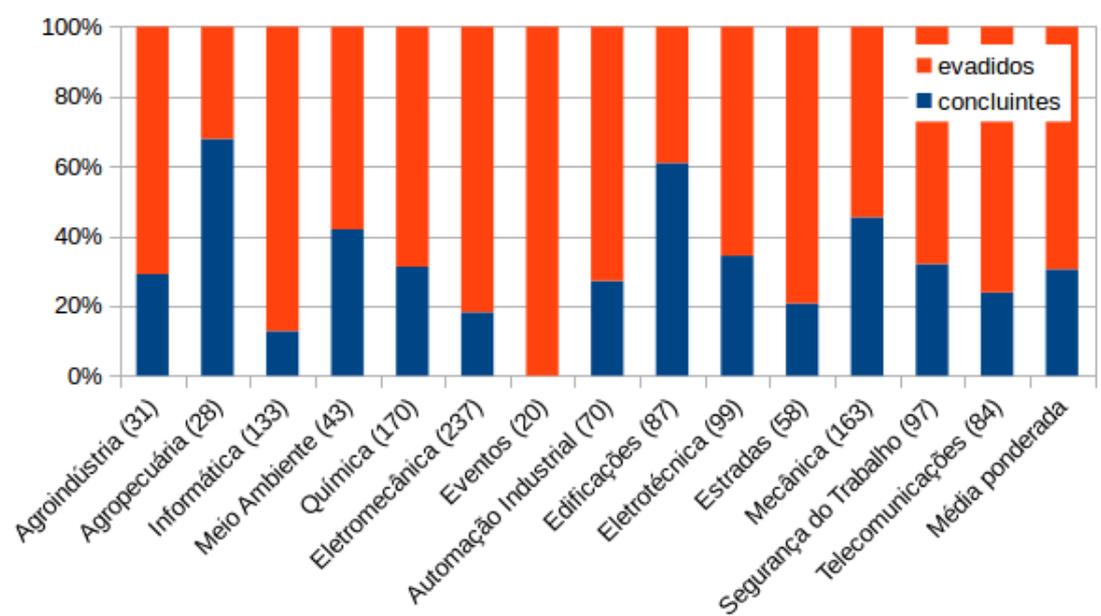


Figura 3 - Quantitativo de alunos concluintes e evadidos entre 2014 e 2016 nos cursos concomitantes. Fonte: Elaboração própria.

Na Figura 3, é possível observar que apenas nos cursos técnicos de Agropecuária e Edificações o número de concluintes foi superior ao de evadidos. O curso de Eventos foi o curso com o maior número de evasões, seguido pelos cursos de Eletromecânica e Informática.

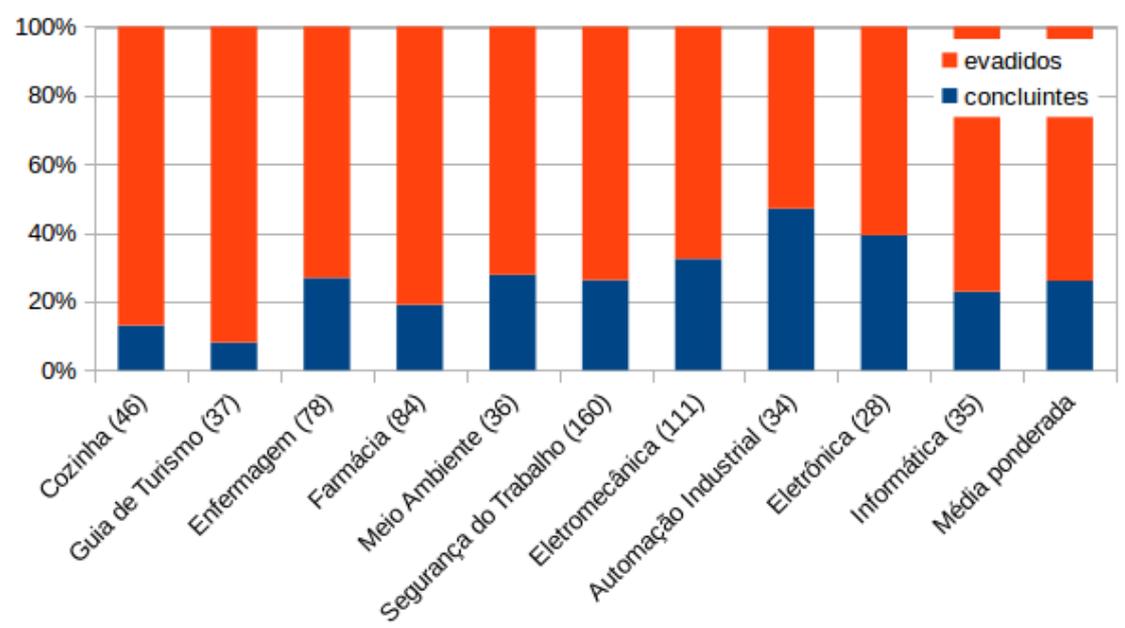


Figura 4 - Quantitativo de alunos concluintes e evadidos entre 2014 e 2016 nos cursos subsequentes. Fonte: Elaboração própria.

Na Figura 4, relacionada aos cursos da modalidade subsequente, observa-se que nenhum curso possui o quantitativo de concluintes superior ao de evadidos. Sendo o curso de Segurança do Trabalho o com maior proporção de evadidos, e o curso de Automação com menor proporção de evadidos.

Em ambas as Figuras é possível observar a discrepância entre a quantidade de concluintes e evadidos, demonstrando que este é um problema enfrentado pela instituição.

2.2. Mineração de dados na identificação do comportamento de alunos evadidos

O Quadro 2 demonstra as taxas de acertos obtidas em cada uma das bases de dados para o método de classificação J48.

Para a obtenção dos resultados apresentados no Quadro 2, o método J48 foi executado no *software Weka*. Sendo o valor para o parâmetro de número mínimo igual a

2, que significa o número mínimo de instâncias para geração de um ramo na árvore. Também foi realizado o teste para o valor do parâmetro igual a 10% da quantidade de instâncias da classe mais rara. Porém, em todos os casos, o melhor resultado foi para o valor igual a 2.

Quadro 2 - Taxas de acerto para as bases de dados utilizando o método J48

<i>Campus</i>	Modalidade	Taxa de acerto
Bom Jesus do Itabapoana	Concomitante	73,01 %
Cabo Frio	Concomitante	57,85%
Cambuci	Concomitante	61,11%
Campos Centro	Concomitante	65,96%
	Subsequente	47,06%
Campos Guarus	Subsequente	63,16%
Itaperuna	Concomitante	74,26%
Macaé	Subsequente	61,39%
Santo Antônio de Pádua	Concomitante	77,78%
Quissamã	Concomitante	66,24%
	Subsequente	58,82%
São João da Barra	Concomitante	66,67%

Fonte: Elaboração própria.

Como demonstrado no Quadro 2, o melhor resultado obtido foi 77,78% para o *campus* Santo Antônio de Pádua. E o pior resultado foi 47,06% para o *campus* Campos Centro na modalidade subsequente. Os demais resultados variaram entre 74,26% e 57,85%.

A seguir serão apresentadas, para algumas bases, os desfechos de evadido nas árvores obtidas a partir do método J48. Serão analisadas as árvores cuja taxa de acerto é maior que 65%, o nível máximo, ou profundidade da árvore, é 50, e em que houve taxa de acerto para a classe de alunos evadidos.

2.2.1. CAMPUS BOM JESUS DO ITABAPOANA

Como demonstrado na Figura 5, no *campus* Bom Jesus do Itabapoana, onde o atributo principal foi o turno, houve ocorrência de evasão entre os alunos que estudavam no período vespertino. Casos de evasão também foram registrados entre os que

estudavam no período noturno, eram do curso técnico em Meio Ambiente, não registraram ausência na disciplina de ecologia mas tenham obtiveram uma nota inferior a 6,2 na disciplina de educação ambiental. Ambas as disciplinas pertencem ao 1º módulo do curso.



Figura 5 - Árvore com desfecho evadido do campus Bom Jesus do Itabapoana. Fonte: Elaboração própria.

2.2.2. CAMPUS ITAPERUNA

Na árvore do *campus* Itaperuna, demonstrada na Figura 6, o primeiro atributo foi a idade seguido pelo desempenho na disciplina matemática aplicada. Estão entre os evadidos os que possuem mais de 17 anos e apresentaram um desempenho igual ou inferior a 5,3 na disciplina matemática aplicado, essa disciplina está presente no 1º módulo dos cursos Eletromecânica, Mecânica, Eletrotécnica e Química. Para os que obtiveram nota superior a 5,3 em matemática aplicada, o próximo atributo é o curso. No curso Técnico em Química, evadiram os que não tiveram ausência nas aulas de química geral aplicada, que compõe o 1º módulo, mas obtiveram nota igual ou inferior a 4,6 em microbiologia II, que pertence ao 2º módulo. Evadiram também os que não apresentaram 100% de presença na disciplina química geral aplicada e apresentaram rendimento igual ou inferior a 6,2 nessa mesma disciplina.

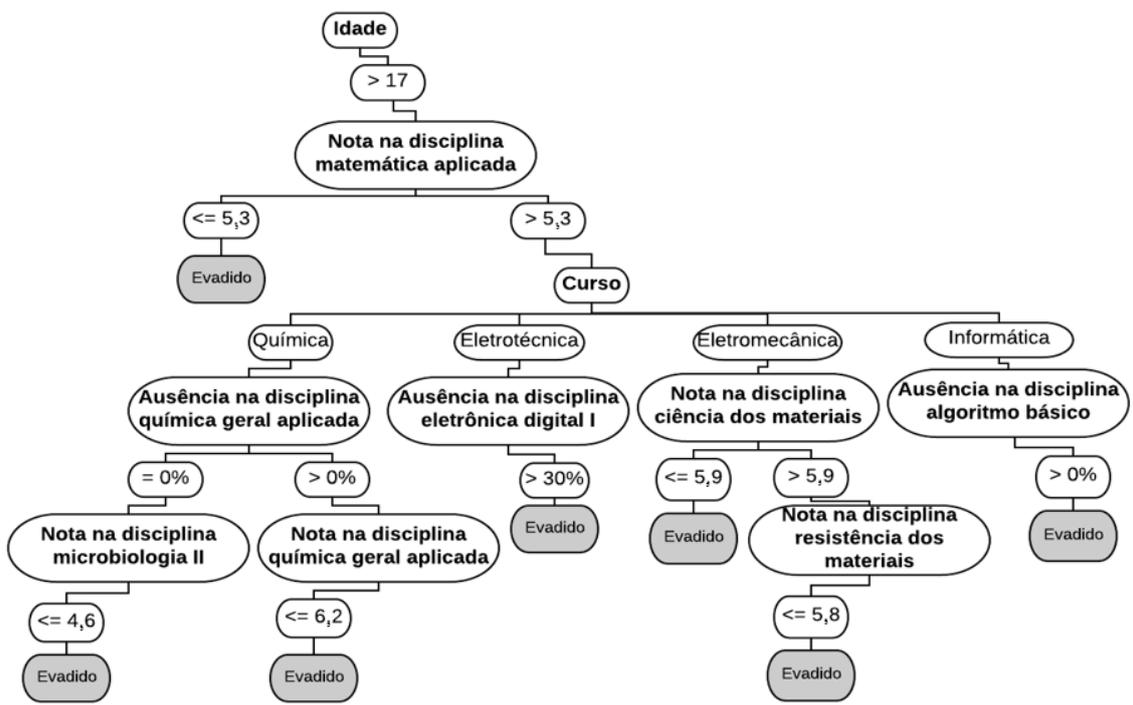


Figura 6 - Árvore com desfecho evadido do campus Itaperuna. Fonte: Elaboração própria.

No curso Técnico em Eletrotécnica houve evadidos entre os que apresentaram mais de 30% de ausência na disciplina eletrônica digital I no 1º módulo do curso. No curso Técnico em Eletromecânica, evadiram alunos que obtiveram nota igual ou inferior a 5,9 em ciência dos materiais, sendo essa uma disciplina do 1º módulo, ou obtiveram nota maior que 5,9 nessa mesma disciplina, mas em resistência dos materiais, que pertence ao 2º módulo, obtiveram uma nota igual ou inferior a 5,8. Já no Curso Técnico em Informática houve evasão entre os que não foram a todas as aulas da disciplina de algoritmo básico.

2.2.3. CAMPUS SANTO ANTÔNIO DE PÁDUA

Para o *campus* Santo Antônio de Pádua, as duas disciplinas que compõem a árvore apresentada na Figura 7, pertencem ao 1º módulo do curso Técnico em Mecânica. Entre os evadidos estão aqueles que obtiveram nota inferior ou igual a 5,4 na disciplina introdução aos processos de fabricação e faltaram pelo menos a uma aula de usinagem I.

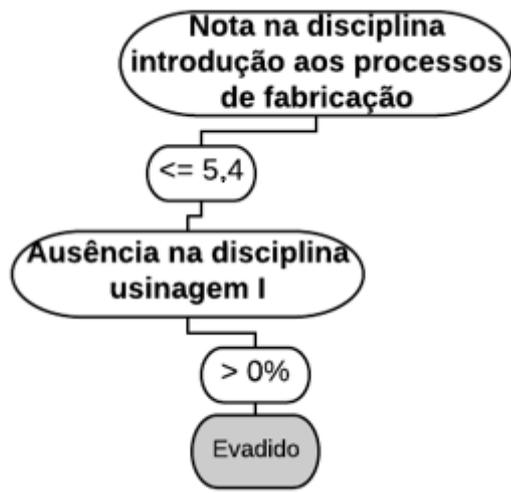


Figura 7 - Árvore com desfecho evadido do campus Santo Antônio de Pádua. Fonte: Elaboração própria.

2.2.4. CAMPUS QUISSAMÃ

No *campus* Quissamã, em que a base de dados da modalidade concomitante foi composta apenas por alunos do curso Técnico em Eletromecânica, o primeiro atributo que aparece na árvore, apresentada na Figura 8, é o desempenho na disciplina informática. Entre os alunos evadidos, estão aqueles que obtiveram nota igual ou inferior a 6,9 em informática e também um índice máximo de ausência de 23% nas aulas de eletrotécnica I. Também houve evadidos entre os que obtiveram nota superior a 6,9 na disciplina informática, sendo que parte desses obtiveram nota igual ou inferior a 8,5 na disciplina desenho técnico I, renda mensal máxima de três salários mínimos e meio e ensino médio incompleto. Os evadidos que obtiveram nota superior a 6,9 na disciplina informática e superior a 8,5 na disciplina desenho técnico I também apresentaram alguma ausência nas aulas de eletrotécnica I. Todas as disciplinas que aparecem na árvore compõem o 1º módulo do curso (Figura 8).

Analisando as quatro árvores apresentadas, de início é possível observar que em todas elas o atributo nota foi considerado em alguma disciplina. Em 4 delas foi considerada também a porcentagem de ausência nas aulas. Outro ponto interessante é que todas as disciplinas que aparecem nas árvores pertencem ao 1º ou 2º módulo dos cursos. Os outros atributos considerados foram: turno, renda mensal familiar, grau de instrução, curso e idade.

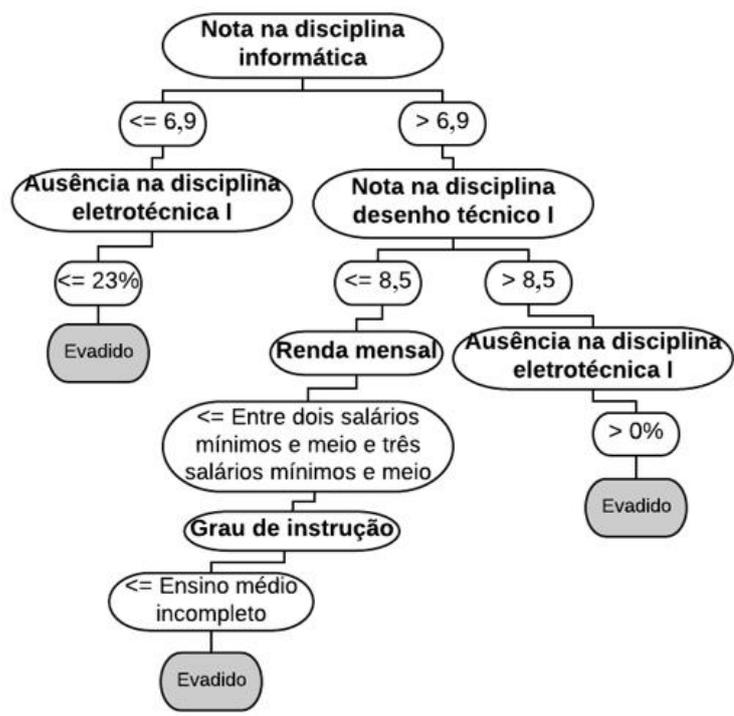


Figura 8 - Árvore com desfecho evadido do campus Quissamã na modalidade concomitante.
Fonte: Elaboração própria.

Para as bases de dados compostas por cursos da modalidade subsequente, como apresentado no Quadro 2, foram obtidas taxas de acerto mais baixas. Será analisada a árvore cuja taxa de acerto foi superior a 60% e em que houve taxa de acerto para a classe de alunos evadidos.

2.2.5. CAMPUS CAMPOS GUARUS

Na árvore da Figura 9, gerada para a base de dados do *campus* Guarus ocorreram cinco desfechos para a classe evadido. O principal atributo é a nota na disciplina metodologia científica. Ocorreu evasão entre os que obtiveram nessa disciplina nota inferior ou igual a 5 e eram do curso de Enfermagem, Eletromecânica ou Farmácia. Sendo que, no curso Técnico em Enfermagem, onde ocorreu o primeiro desfecho, houve evasão entre os que obtiveram nota igual ou inferior a 5,3 na disciplina informática básica mesmo apresentando uma ausência igual ou inferior a 20% das aulas de fundamentos de enfermagem I. Desses, houve evasão entre os alunos do sexo feminino, e do sexo masculino para os que obtiveram nota igual a 0 na disciplina anatomia e fisiologia humana. Todas as disciplinas citadas compõem o 1º módulo do curso.

No curso Técnico de Eletromecânica, foi observada evasão entre os que obtiveram nota igual ou inferior a 4,6 na disciplina de informática, que pertence ao 1º módulo. Já no curso Técnico de Farmácia apenas a nota inferior ou igual a 5 na disciplina de metodologia científica foi decisiva para a evasão. Já entre os alunos evadidos que obtiveram nota superior a 5 em metodologia científica, o comportamento identificado foi daqueles que obtiveram nota igual 0 na disciplina enfermagem em paciente crítico, nota inferior ou igual a 0,06 em primeiros socorros, nota superior a 5 na disciplina farmacologia aplicada à enfermagem, cursavam no turno matutino e obtiveram nota igual ou inferior a 5 em enfermagem. Para esse último desfecho, todas as disciplinas são do curso de enfermagem, sendo que a disciplina enfermagem com paciente crítico pertence ao 3º módulo, primeiros socorros e farmacologia aplicada à enfermagem pertencem ao 2º módulo e psicologia aplicada à enfermagem ao 1º módulo do curso.

A partir da árvore analisada para a base de dados composta por cursos na modalidade subsequente, é possível perceber que, diferenciado das árvores para os cursos na modalidade concomitante, além de disciplinas dos 1º e 2º módulo dos cursos, esteve presente também o atributo de nota em uma disciplina do 3º módulo.

DISCUSSÃO DOS RESULTADOS

A partir dos resultados obtidos analisa-se que em geral os atributos que mais descrevem os comportamentos dos alunos evadidos no ensino técnico estão relacionados à ausência e às notas obtidas nas disciplinas do curso. Corroborando com o estudo de Jiménez-Gómez *et al.* (2015), foi vislumbrado que os comportamentos podem ser identificados a partir de características detectáveis já nos módulos iniciais do curso.

No estudo realizado por Souza (2014), busca-se a identificação de fatores que levam a evasão em cursos técnicos subsequentes no Instituto Federal do Rio Grande do Norte através da aplicação de questionários e realização de entrevistas. A autora constatou que os principais fatores destacados pelos alunos estão relacionados ao currículo, disciplinas de difícil compreensão e metodologia dos professores. Assim como na presente pesquisa, em que, em sua maioria as regras incluem o desempenho ou frequência em alguma disciplina.



Figura 9 - Árvore com desfecho evadido do campus Campos Guarus. Fonte: Elaboração própria.

Bastos e Gomes (2017), que investigaram fatores que levam a evasão no ensino técnico, identificaram que a maioria dos alunos evade no segundo módulo do curso, e que em geral as disciplinas que determinam o comportamento dos alunos evadidos ocorrem no primeiro módulo. Um ponto em comum com os resultados obtidos nesta pesquisa, onde a maioria dos atributos que compõem as árvores estão relacionados às disciplinas do 1º ou 2º módulo dos cursos.

Cruz (2013) investiga as principais causas de evasão de cursos profissionalizantes gratuitos. Em relação à análise dos alunos evadidos, o autor constata que a maioria são jovens entre 18 e 27 anos e renda mensal de até R\$ 2040,00. O resultado relacionado ao *campus* Itaperuna apresentou como principal atributo a idade, onde os alunos que evadiram possuíam no mínimo 17 anos, englobando alunos jovens assim como no

trabalho de Cruz (2013). Em relação a renda mensal, no *campus* Quissamã também foi obtido resultado compatível.

Bastos e Gomes (2017), Cruz (2013) e Veiga e Bergiante (2016) identificaram como determinante para a evasão o fato do aluno ter que conciliar os estudos com o trabalho. O presente estudo não observou a mesma realidade, já que não foram obtidos resultados onde o atributo decisivo foi aquele que descreve se o aluno exerce atividade remunerada.

No trabalho de Cunha, Moura e Analide (2016), além de fatores relacionados ao perfil socioeconômico do aluno, foi identificado que o rendimento dos alunos em determinadas disciplinas também foi um fator em comum entre os alunos evadidos. Os autores verificaram que os alunos evadidos demonstraram dificuldades em disciplinas relacionadas à lógica e à matemática assim como disciplinas da área técnica. Na presente pesquisa os atributos que mais descreveram o comportamento dos alunos evadidos também estão relacionados às disciplinas técnicas, e no *campus* Itaperuna a primeira disciplina da árvore é matemática aplicada, corroborando com os resultados de Cunha, Moura e Analide (2016).

CONCLUSÃO

A partir de dados armazenados em sistemas utilizados pelas instituições de ensino e de técnicas de mineração de dados que propiciem a extração de informações relevantes desses dados é possível explorar o problema de evasão escolar. Esta exploração se dá no sentido de oferecer informações que contribuem para a elaboração de políticas estudantis que busquem reduzir a evasão.

De acordo com Cruz (2013), pode-se concluir que a análise da evasão no ensino é uma tarefa desafiadora, por buscar compreender que o modo como o estudante enfrenta e reage às novas tarefas, às mudanças e às vivências escolares pode influenciar no sucesso do processo de ensino-aprendizagem, diplomação, permanência e evasão escolar.

Através deste trabalho buscou-se aplicar as etapas e métodos mais utilizados, de acordo com Cordeiro, Mussa e Hora (2017), para extração de características de alunos evadidos no ensino técnico. O estudo de caso foi realizado no Instituto Federal

Fluminense, mais especificamente, utilizando dados de alunos de cursos técnicos nas modalidades concomitante e subsequente.

O fracasso escolar produz marcas no aluno, favorecendo a baixa estima e o desenvolvimento de um processo depreciativo que promove a desmotivação com os estudos e com a escola de modo geral. Porém, as marcas no desenvolvimento humano são mais profundas, uma vez que o aluno passa a ter comprometidas suas potencialidades e habilidades (ARAÚJO; SANTOS, 2012). Isso reforça a necessidade de estudos no sentido de entender a evasão e acredita-se que este trabalho contribui para o alcance desses objetivos.

A partir dos resultados obtidos, conclui-se que, de forma geral, não há diferença entre os comportamentos dos alunos evadidos de cursos técnicos na modalidade concomitante e na modalidade subsequente. Em ambos os casos, as disciplinas dos primeiros módulos que traçam os comportamentos. Em sua maioria, são disciplinas técnicas e os fatores determinantes são o desempenho e a ausência nas aulas.

Apesar das taxas de acertos obtidas não terem ultrapassado 80%, ainda assim, os resultados são considerados satisfatórios por terem permitido realizar análises que de fato descrevem os comportamentos de alunos que evadiram na amostra analisada. Ainda que não representem um comportamento considerado genérico, no universo dos cursos técnicos são resultados pertinentes ao estudo de caso realizado.

Acredita-se que a principal contribuição dessa pesquisa é explicitação dos fatores que descrevem os comportamentos de alunos evadidos no ensino técnico. Além da utilização de técnicas de mineração em dados educacionais, área identificada como emergente, foram identificados os atributos que descrevem o comportamento dos alunos evadidos no ensino técnico, sendo este um nível de ensino pouco explorado na mineração de dados educacionais.

Entre as dificuldades encontradas nesse trabalho está o fato do sistema acadêmico não ser constantemente atualizado em todos os *campi*. Isso resulta em dados faltantes e consequentemente a exclusão de alunos da base de dados analisada.

Como trabalhos futuros propõe-se aplicar a seleção de atributos previamente a fase de mineração de dados, com o intuito de verificar se houve melhora nas taxas de acerto obtidas. Além disso, com o mesmo intuito, poderá ser investigada a utilização de outros métodos de classificação. Futuramente, sugere-se ainda analisar o

comportamento dos alunos concluintes e aplicar a mesma metodologia em bases de dados contendo apenas dados socioeconômicos com o intuito de observar como esses dados descrevem o comportamento do aluno evadido.

REFERÊNCIAS

ARAÚJO, C. F. DE; SANTOS, R. A. DOS. **A educação profissional de nível médio e os fatores internos/externos às instituições que causam a evasão escolar.** The 4th international congress university industry cooperation. **Anais.** In: CONGRESSO INTERNACIONAL DE COOPERAÇÃO UNIVERSIDADE- INDÚSTRIA. Taubaté, SP: 2012

BASTOS, O. G. A.; GOMES, C. F. S. **A evasão escolar no Ensino Técnico: entendendo e enfrentando as dificuldades - Um estudo de caso do CEFET-RJ.** In: CONGRESSO NACIONAL EM EXCELÊNCIA EM GESTÃO. 2017

COHEN, W. W. **Fast Effective Rule Induction.** In Proceedings of the Twelfth International Conference on Machine Learning. **Anais...**Morgan Kaufmann, 1995

CORDEIRO, R. G.; MUSSA, M. DE S.; HORA, H. M. R. DA. **Mineração de dados educacionais com foco na evasão: Uma revisão sistemática.** In: ENCONTRO FLUMINENSE DE ENGENHARIA DE PRODUÇÃO. Nova Iguaçu, RJ: ENFEPro, 2017

CRUZ, A. P. DA. **Evasão nos cursos técnicos profissionalizantes: uma análise das principais causas e identificação de perfil dos alunos evadidos do Senac Sete Lagoas.** Pedro Leopoldo: Fundação Pedro Leopoldo, 2013.

CUNHA, J. A.; MOURA, E.; ANALIDE, C. Data mining in academic databases to detect behaviors of students related to school dropout and disapproval. **Advances in Intelligent Systems and Computing**, v. 445, p. 189–198, 2016.

DEKKER, G. W.; PECHENIZKIY, M.; VLEESHOUWERS, J. M. **Predicting students drop out: A case study.** In: EDM'09 - EDUCATIONAL DATA MINING 2009: 2ND INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING. 2009

FIGUEIREDO, N. G. DA S.; SALLES, D. M. R. Educação Profissional e evasão escolar em contexto: motivos e reflexões. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 25, n. 95, p. 356–392, abr. 2017.

GUTTMAN, L. A basis for analyzing test-retest reliability. **Psychometrika**, v. 10, n. 4, p. 255–282, 1945.

JIMÉNEZ-GÓMEZ, M. A. et al. **Discovering clues to avoid middle school failure at early stages.** . In: ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES. 2015

MÁRQUEZ-VERA, C. et al. Early dropout prediction using data mining: A case study with high school students. **Expert Systems**, v. 33, n. 1, p. 107–124, 2016.

PRADEEP, A.; DAS, S.; KIZHEKKETHOTTAM, J. J. **Students dropout factor prediction using EDM techniques**. . In: PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON SOFT-COMPUTING AND NETWORK SECURITY, ICSNS 2015. 2015

QUINLAN, J. R. Simplifying Decision Trees. **Int. J. Man-Mach. Stud.**, v. 27, n. 3, p. 221–234, set. 1987.

SOUZA, J. **Permanência e evasão escolar: Um estudo de caso em uma instituição de ensino profissional**. Dissertação (Mestrado) Juiz de Fora: Universidade Federal de Juiz de Fora, 2014.

VEIGA, C.; BERGIANTE, N. **Fatores predominantes da evasão escolar no ensino médio profissional: uma revisão de literatura**. In: CONGRESSO NACIONAL DE EXCECELÊNCIA EM GESTÃO. 2016.