

EVASÃO NO ENSINO SUPERIOR: INVESTIGAÇÃO DAS CAUSAS VIA MINERAÇÃO DE DADOS

Deivison Lamonica Barreto¹

Mariane Rangel de Matos²

Henrique Rego Monteiro da Hora³

Aline Pires Vieira de Vasconcelos⁴

Recebido em: abril/2019

Publicado em: dezembro/2019

Resumo

Um grande problema observado em cursos de diferentes níveis é a evasão de alunos. Isto afeta tanto instituições de ensino como estudantes. O presente trabalho visa a identificação do comportamento de alunos evadidos de cursos superiores em uma instituição de ensino federal. Esta identificação pode auxiliar a assimilar alunos que possuem as mesmas características, ou seja, que podem ser considerados como estudantes em risco de evasão. O trabalho foi realizado em cinco etapas: obtenção e seleção de dados, pré-processamento dos dados, transformação de dados, *data mining* e interpretação e análise dos resultados. A identificação do comportamento foi obtida através de técnicas de mineração de dados, utilizando o algoritmo de classificação J48, que gera árvores de decisão. Foram identificados alguns comportamentos com alta propensão à evasão e destes, podem-se destacar jovens com até 21 anos de idade e maiores que 27 anos. Foi observado maior impacto na evasão para os cursos de tecnologia em que o aluno possui renda familiar de até dois salários mínimos. De posse dessas informações, a instituição poderá adotar medidas que busquem a redução da evasão dos seus alunos. A identificação precoce desse problema permite que a questão seja trabalhada pela instituição, trazendo benefícios não somente para ela, mas também para a região, para os alunos e para a sociedade.

Palavras-chave: mineração de dados; evasão; ensino superior.

¹ Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Rio de Janeiro, Brasil. E-mail: deivisonlb@gmail.com

² Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Rio de Janeiro, Brasil. E-mail: mariane.rmatos@gmail.com

³ Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Rio de Janeiro, Brasil. E-mail: dahora@gmail.com

⁴ Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Rio de Janeiro, Brasil. E-mail: alineprofcefet@gmail.com

EVASION IN HIGHER EDUCATION: INVESTIGATION OF CAUSES VIA DATA MINING

Abstract

A major problem observed in courses at different levels is student evasion. This affects both educational institutions and students. The present work aims at identifying the behavior of students evaded from higher education in a federal educational institution. This identification can help to assimilate students who have the same characteristics, in other words, they can be considered as students at risk of evasion. The work was performed in five stages: data acquisition and selection, data preprocessing, data transformation, data mining and interpretation and analysis of results. The behavior identification was obtained through data mining techniques, using the classification algorithm J48, which generates decision trees. Some behaviors with high propensity to evasion were identified and from these, young people up to 21 years old and people older than 27 years can be highlighted. A greater impact on evasion was observed for technology courses where the student has a family income up to two minimum wages. With this information, the institution may adopt measures that seek to reduce the evasion of its students. The early identification of this problem allows the issue to be worked by the institution, bringing benefits not only for it, but also for the region, the students and the society.

Keywords: data mining; evasion; higher education.

INTRODUÇÃO

A evasão de estudantes no ensino superior é uma questão crítica nas universidades brasileiras, tendo como consequências a formação de profissionais abaixo da capacidade instalada (o que limita, portanto, o impacto social das instituições), frustração dos alunos que não conseguem concluir sua graduação e significativo desperdício de recursos (HIPÓLITO, 2011).

Conforme exposto por Viana e Lima (2010), o nível educacional de uma nação influencia sua economia de diversas formas e em geral, existe uma correlação positiva entre o nível de escolarização de um país, sua produtividade e os ganhos salariais de sua população. Assim, a evasão no ensino superior pode influenciar negativamente alguns aspectos da economia de um país (ASSIS, 2017). Além de representar um problema de ordem econômica, a evasão atinge diretamente o aluno. Hegde e Prageeth (2018) destacam que o estudante possui diversos sonhos e expectativas ao ingressar em um curso. Quando essas expectativas não são satisfeitas, pode ocorrer a motivação de evadir, a qual pode ser influenciada por fatores individuais e institucionais.

Márquez *et al.* (2016) afirmam que quanto mais cedo forem identificados os estudantes propensos a evadir de um curso, maiores são as chances de sucesso da política de permanência

escolar. Estudantes cujo risco de evasão é conhecido podem ser ajudados de forma personalizada, tanto pela família quanto pela instituição de ensino superior (IES).

Com a informatização dos sistemas escolares tem-se um grande volume de dados armazenados na área educacional que requerem a utilização de técnicas que propiciem a interpretação e análise dos mesmos, com objetivo de agregar valor à gestão e aos educadores. A esta busca por informações em banco de dados é dada o nome de Processo de Descoberta de Conhecimentos em Banco de Dados (do inglês, *Knowledge Discovery in Database – KDD*).

Cardoso e Machado (2008) consideram o KDD como o processo de descoberta de conhecimento útil de um banco de dados, enquanto a mineração de dados está inserida como um dos principais passos necessários para a realização do processo de KDD. Esta é definida sucintamente como um conjunto de técnicas e procedimentos para identificação e representação de padrões relevantes a partir de conjuntos de dados (WITTEN; FRANK; HALL, 2011).

Silva Filho *et al.* (2007) afirmam que o estudo interno, realizado por uma IES com base em seus dados, pode ser muitas vezes mais detalhado porque é possível institucionalizar-se um mecanismo de acompanhamento da evasão, registrando os diversos casos, agrupando e analisando subgrupos, ou diferentes situações.

Desse modo, para diminuir a evasão nas instituições de ensino superior é necessário entender seu comportamento. Assim, o objetivo do presente trabalho é identificar o comportamento dos alunos evadidos de nível superior, tendo como foco uma instituição de ensino federal. A realização deste estudo foi feita através de técnicas de mineração de dados. Conhecendo o perfil desses alunos a questão da evasão poderá ser trabalhada pela instituição, que poderá adotar medidas que busquem a redução da evasão, trazendo benefícios para ela própria, para os alunos e para a sociedade.

REFERENCIAL TEÓRICO

Evasão no ensino superior

De acordo com Silva e Santos (2017), a evasão é a saída definitiva do estudante do seu curso. Entretanto, essa definição contém limitações, visto que um estudante pode evadir de um curso e se tornar concluinte em outro. Com o objetivo de esclarecer essa limitação, o Ministério da

Educação (BRASIL, 1996) propõe como formas de evasão: a evasão de curso (quando o estudante se desliga do curso superior em situações como transferência e abandono); a evasão da instituição (quando o estudante se desliga da instituição em que está matriculado) e a evasão do sistema (quando o estudante abandona o ensino superior de forma definitiva ou temporária).

De acordo com Schmitt *et al.* (2012), a maioria dos estudos brasileiros ligados à evasão está voltada para a educação básica, de modo que os estudos sobre evasão no ensino superior ainda são incipientes no país e apenas começaram a crescer em quantidade na década de 2000. A pesquisa de Schmitt foi feita considerando artigos A1, A2, B1 e B2, segundo as normas de classificação postuladas pelo sistema *Qualis* da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CLOAPES), na área da educação com o tema de evasão no ensino superior no contexto brasileiro. Foram encontrados apenas sete artigos que atendiam à especificidade do tema. A busca foi feita considerando os seguintes parâmetros: evasão, educação superior e seus correlatos como: ensino superior, abandono, permanência e fracasso.

A evasão no Ensino Superior tem sido tema de pesquisas acadêmicas que, na sua maioria, buscam compreender as causas do aumento significativo do número de alunos que entram no Ensino Superior e evadem, muitas vezes após o primeiro semestre do curso (SOUZA, SILVA e GESSINGER, 2017).

Mineração de dados e KDD

A mineração de dados educacionais é uma área emergente em que muitos trabalhos buscam a análise de dados com foco na evasão (MÁRQUEZ *et. al.*, 2016). Geralmente, os trabalhos que utilizam mineração de dados para estudos de evasão escolar se baseiam nas etapas definidas pelo KDD (MEHTA; BUCH, 2016). *Data mining*, além de ser o nome dado a uma etapa do processo de KDD, é também o nome de uma área interdisciplinar da ciência, que desperta o interesse de pesquisadores de diversas outras áreas, onde se procura aplicar algoritmos e técnicas específicas com objetivo de extrair informações e relacionamentos de informações em bancos de dados (ASSEISS, 2017).

Lam-On e Boongoen (2014) observaram que os bancos de dados tendem a incluir muitas variáveis redundantes (principalmente pela criação de novas variáveis derivadas a partir dos dados originais) e, em tais condições, alguns métodos de classificação têm seu desempenho degradado. Os

autores propõem uma etapa de transformação dos dados antes da aplicação dos métodos de classificação, a fim de eliminar a redundância e aumentar a acurácia dos classificadores.

Institutos Federais e Educação Profissional e Tecnológica (EPT)

Os Institutos Federais de Educação, Ciência e Tecnologia foram criados em 2008 por meio da Lei nº. 11.892, absorvendo os Centros Federais de Educação Tecnológica (CEFETs) e as Escolas Técnicas remanescentes (BRASIL, 2011). De acordo com o MEC (BRASIL, 2019a), os Institutos Federais são instituições pluricurriculares e multicampi, possuindo em sua estrutura Reitoria, *campus*, *campus* avançado, Polos de Inovação e Polos de Educação a Distância. Possuem como especialidade a oferta de Educação Profissional e Tecnológica (EPT) em todos os seus níveis e formas de articulação com os demais níveis e modalidades da Educação Nacional.

De acordo com o MEC (BRASIL, 2019b), a EPT é uma modalidade educacional prevista na Lei de Diretrizes e Bases da Educação Nacional (LDB) que abrange cursos de qualificação, habilitação técnica e tecnológica, e de pós-graduação, organizados de maneira que permite o aproveitamento contínuo e articulado dos estudos. Prevê a integração com os diferentes níveis e modalidades da Educação e às dimensões do trabalho, da ciência e da tecnologia. Envolve, por exemplo, a modalidade da educação de jovens e adultos e a educação básica no nível do ensino médio, na forma integrada, concomitante e na forma subsequente.

METODOLOGIA

A presente pesquisa foi aplicada no Instituto Federal de Educação, Ciência e Tecnologia Fluminense *campus* Campos Centro. Em relação aos cursos de nível superior oferecidos por este *campus*, foi estudada a evasão em dez cursos, em todos os turnos: matutino, vespertino, noturno e integral. Os cursos foram agrupados de acordo com o seu tipo. Para os cursos de Bacharelado, têm-se: Arquitetura e Urbanismo, Engenharia de Controle e Automação e Sistemas de Informação. Para os cursos de Licenciatura, têm-se os seguintes cursos: Ciências da Natureza, Geografia, Letras e Matemática. Já para os cursos de Tecnologia, têm-se os cursos de Design Gráfico, Manutenção Industrial e Sistemas de Telecomunicações.

Os cursos de Licenciatura em Teatro, Licenciatura em Educação Física, Bacharelado em Engenharia Elétrica e Bacharelado em Engenharia de Computação não possuíam alunos concluintes, apenas alunos evadidos. Essa situação ocorreu porque nesses cursos ainda não houve formação da primeira turma. Esse tipo de situação impede que algoritmo utilizado consiga classificar os dados, uma vez que existe apenas uma classe. Por essa razão, os registros referentes aos alunos desses cursos não foram considerados.

Processo de KDD

Neste trabalho, foi utilizado o processo de KDD que é composto por cinco etapas: etapa de obtenção e seleção de dados, etapa de pré-processamento dos dados, etapa de transformação de dados, etapa de *data mining* e etapa de interpretação e análise dos resultados. Todas estas etapas são importantes, pois preparam os dados e fornecem um procedimento organizado para a descoberta de novos conhecimentos, sendo que se destaca a importância da etapa de *data mining*, em que algoritmos específicos são aplicados aos conjuntos de dados, o que permite a detecção de padrões e relacionamentos importantes para a descoberta de conhecimento. Estas atividades são descritas com detalhes abaixo.

Obtenção e seleção de dados

Para a obtenção e seleção dos dados brutos foi utilizado o Sistema Acadêmico de uma instituição de ensino, em que foram obtidas três planilhas no formato *comma-separated values* (CSV) referentes a todos os alunos que estudaram nesta instituição nos seguintes semestres: primeiro semestre de 2017 (2017.1), segundo semestre de 2017 (2017.2) e primeiro semestre de 2018 (2018.1). As planilhas possuíam inicialmente treze atributos: bairro; cidade; curso; endereço; escola de origem (privada, pública e não informado); instituição; naturalidade; nascimento; renda familiar; sexo; situação matrícula; situação período e turno (diurno, integral, matutino, noturno e vespertino).

As planilhas possuíam informações de todos os alunos cadastrados no sistema acadêmico, incluindo estudantes que pertenciam aos cursos de línguas, assim como alunos do ensino médio. A seleção inicial de dados foi feita excluindo os alunos do ensino médio, dos cursos de línguas e dos

cursos de pós-graduação, mantendo assim apenas os alunos de cursos de graduação. Outro atributo que também foi eliminado foi a instituição, visto que essa informação era a mesma para todos os alunos.

Limpeza e pré-processamento dos dados

Nesta etapa, removeu-se valores ou informações distorcidos, tratou-se ruídos e valores ausentes nos dados. O pré-tratamento dos dados foi feito com a abertura dos arquivos com extensão CSV no software livre LibreOffice Calc. Este programa foi escolhido devido ter uma melhor interface com o software a ser utilizado na mineração de dados. Um dos ajustes feitos foi que, ao importar as informações para o LibreOffice Calc, por exemplo, retirou-se todos os caracteres aspas e apóstrofo. Verificou-se também se não houve nenhum dado que foi deslocado a informação para a coluna errada. Outro ajuste realizado foi a complementação de informações faltantes. Por exemplo, foi realizado o preenchimento do campo “Bairro” em registros onde se tinham apenas o nome da rua e cidade de residência dos estudantes.

Transformação de dados

Para cada vínculo de aluno ao curso existe um atributo que indica a situação desse aluno no curso. É através da situação do aluno no curso que se faz possível, neste trabalho, determinar quais alunos são considerados evadidos e quais não são. Segue abaixo os status da situação considerados neste trabalho:

- Cancelado -> aluno que no semestre de referência da obtenção do dado está com a matrícula cancelada;
- Concluído -> aluno que foi aprovado em todas as disciplinas cursadas no período de referência, mas que ainda não realizou o procedimento de colação de grau;
- Egresso -> aluno que além de ter sido aprovado em todas as disciplinas cursadas no período de referência e que já realizou o procedimento de colação de grau;
- Evasão -> aluno que no semestre de referência da obtenção do dado não possui vínculo com o curso por motivos de abandono, desligamento, entre outros.

Os status acima foram subdivididos em Evadido para aqueles que possuíam a situação de matrícula como Cancelado e Evasão; e Não Evadido para os alunos que estavam com a situação de matrícula como Egresso e Concluído. Para este estudo, buscou-se encontrar padrões para os casos em que o aluno evadiu ou teve a matrícula cancelada.

Foram escolhidos alguns atributos para mostrar suas relações com a classe de interesse, a evasão. Além disso, foram criados outros atributos a partir dos originais para também possuir dados mais homogêneos possíveis. Os dados criados foram:

- idade → dado necessário para conhecer a idade dos alunos. Esse item foi adicionado para verificar se a idade dos alunos possuiria relação com a evasão. Se fosse utilizado o dado original, que continha a data de nascimento, os algoritmos não seriam capazes de encontrar nenhum padrão relacionado a esse dado, tendo em vista a quantidade de diferentes datas de nascimento presentes nas bases de dados trabalhadas. Além disso, a idade é um valor numérico, podendo ser utilizado em cálculos e comparações realizados pelos algoritmos.

- tipo de curso → tendo em vista a grande quantidade de cursos de nível superior, os mesmos foram divididos por tipos, sendo estes cursos de bacharelado, licenciatura e tecnologia. No processamento do algoritmo, foi utilizado o campo “tipo de curso” ao invés do campo “curso”.

- situação → situação do aluno, ou seja, se o mesmo é evadido ou não evadido.

Data Mining

Neste trabalho foi utilizado um software livre que possui uma coleção de algoritmos para mineração de dados chamado *Waikato Environment for Knowledge Analysis* – WEKA (WEKA, 2018). Os recursos padrões disponíveis neste programa de computador são: pré-processamento, classificação, regressão, clusterização e regras de associação.

Foi feito inicialmente uma redução de dados, etapa que consiste na descoberta de atributos úteis ao se aplicar métodos de tentativa e inclusão de dados transformados. A etapa tem o objetivo de encontrar as características nos dados que sejam mais representativas ao objetivo do processo.

Foram realizados experimentos de seleção de atributos, isto é, identificação de quais as características mais relevantes para se prever a probabilidade do aluno evadir e experimentos de classificação, considerando dezoito informações: escola de origem, renda familiar, sexo, turno, idade, distância, tipo de curso, situação, região, situação período, situação matrícula, bairro, cidade,

endereço, naturalidade, data de nascimento, curso e período. O resultado desta redução foi o uso de apenas oito atributos, a saber: escola de origem, renda familiar, sexo, turno, idade, distância, tipo de curso e situação.

Foi realizada a mineração de dados propriamente dita, em que se buscou padrões de interesse em uma forma de representação específica, como a árvore de decisão utilizando o método de classificação J48.

O algoritmo J48, de acordo com Lorenzett e Teloken (2016), consiste na criação de uma árvore de decisão a partir de uma base de dados para obtenção de conhecimento, e com isto realizar uma tomada de decisão, onde o atributo mais significativo é considerado a raiz da árvore. No presente trabalho, foi utilizada a quantidade mínima de 5% do total de registros de cada semestre para geração de um ramo na respectiva árvore.

Em relação à validação das regras encontradas, foi utilizada a técnica de validação cruzada (*cross-validation*), que consiste na divisão do conjunto de dados em subconjuntos, nos quais um é utilizado para treino e os demais para validação. Para o presente trabalho, o conjunto original de dados foi subdividido em dez subconjuntos.

O objetivo desta etapa é selecionar o método de busca por padrões nos dados e decidir quais modelos e parâmetros do método são apropriados.

Interpretação e análise dos resultados

Os resultados foram analisados considerando os atributos mais relevantes que levam os alunos a evadirem. Para efeitos de detecção de padrões, foi considerado como número mínimo de objetos a quantidade correspondente a 5% do total de registros de cada semestre estudado. A classificação foi avaliada de acordo com sua acurácia geral, isto é, a taxa de acertos do classificador ao estimar, com base nas informações de alguns atributos, se o aluno irá ou não abandonar o curso. Além desse indicador, também é apresentada a precisão do algoritmo para cada desfecho (Evadido e Não Evadido), pois um mesmo algoritmo pode possuir um melhor desempenho em um desfecho do que em outro. Assim é possível identificar padrões de interesse de acordo com o critério pretendido e que é interpretado para tornar o conhecimento extraído mais compreensível através de forma gráfica ou relatórios demonstrativos.

RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados os dados coletados para investigação com os resultados e discussões do presente trabalho. Eles estão organizados da seguinte forma: dados referentes ao quantitativo de alunos e apresentação de árvore de decisão gerada através do algoritmo J48 para cada semestre analisado. Por último, é apresentada a matriz de confusão da árvore, além de outras informações como quantidade de instâncias classificadas corretamente, quantidade de instâncias classificadas incorretamente e precisão, o que permite a análise dos resultados obtidos.

A pesquisa foi aplicada no Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF) *campus* Campos Centro, localizado na cidade de Campos dos Goytacazes – RJ. Esta instituição atende em sua maioria alunos do norte fluminense e do estado do Espírito Santo, sendo uma referência na formação de alunos na região.

Bases de dados

As bases de dados estudadas possuíam informações de todos os alunos da instituição e cadastrados no Sistema Acadêmico, incluindo estudantes que pertenciam aos cursos de línguas e alunos do ensino médio. A seleção inicial de dados foi feita excluindo os alunos do ensino médio, dos cursos de línguas e dos cursos de pós-graduação, mantendo assim apenas os alunos de cursos de graduação. Outro atributo que também foi eliminado foi a instituição, visto que essa informação era a mesma para todos os alunos. Com o resultado desta redução, foram geradas as novas bases dos semestres analisados e o quantitativo de alunos observados segue na Figura 1.

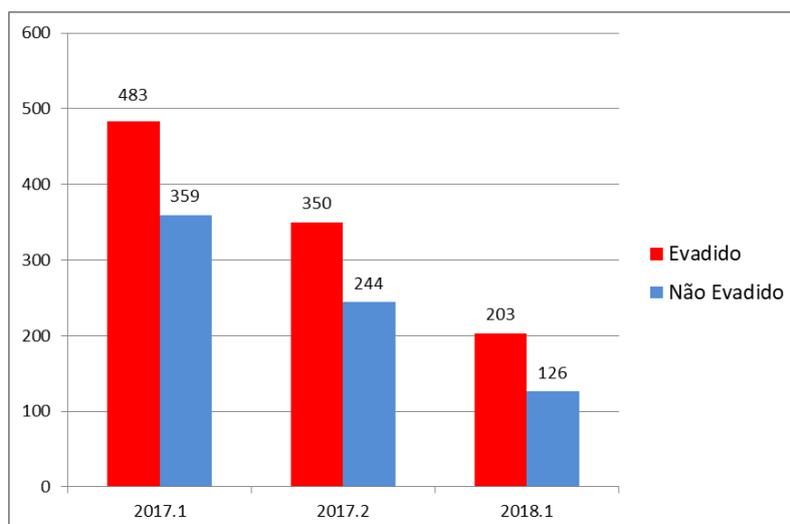


Figura 1: Quantitativo de registros das bases de dados trabalhadas.
Fonte: Elaboração Própria.

Árvore de Decisão para o Semestre 2017.1

A Figura 2 apresenta Árvore de decisão utilizando o algoritmo J48 para o semestre 2017.1:

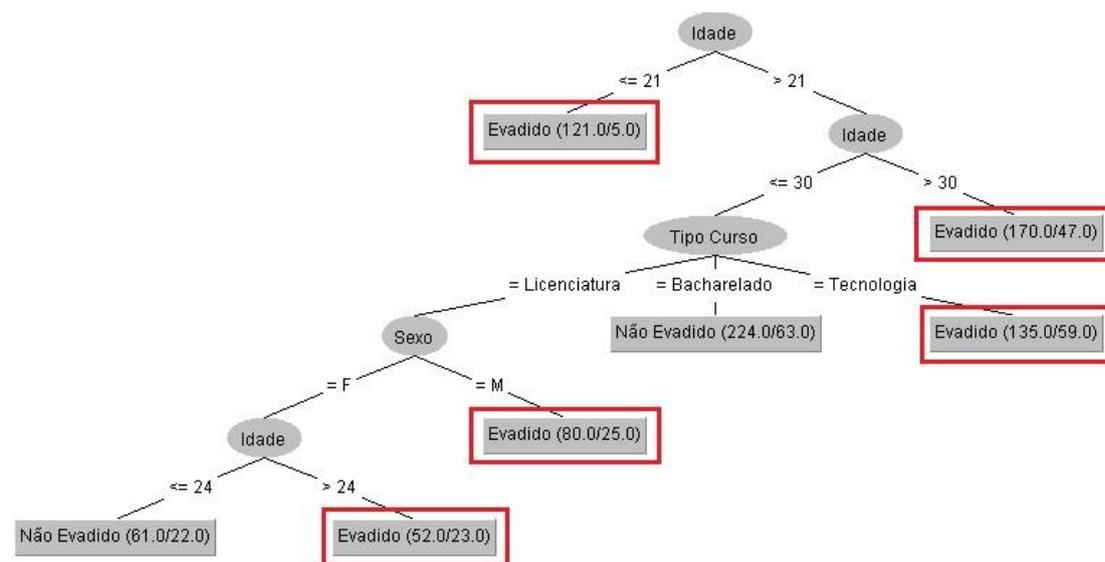


Figura 2: Árvore de decisão utilizando o algoritmo J48 para o semestre 2017.1.
Fonte: Elaboração Própria.

Pode-se observar na Figura 2 que para o semestre 2017.1 foram identificadas duas faixas de idade com maior evasão: alunos com idade menor ou igual a 21 anos e alunos com idade maior que

30 anos. Para os alunos na faixa etária entre 21 e 30 anos, a evasão mostrou maior relação com tipo de curso e o sexo do estudante. Em relação ao tipo de curso, os cursos de Tecnologia possuem mais alunos evadidos. Quando se trata do atributo sexo, observa-se maior relação com os cursos de Licenciatura, sendo que a taxa de evasão nesses cursos é mais alta entre os alunos do sexo masculino. Em relação ao sexo feminino, também nos cursos de Licenciatura, observa-se maior evasão para alunas com idade superior a 24 anos.

A Tabela 2 apresenta os resultados obtidos pelo algoritmo J48 para o primeiro semestre de 2017:

Tabela 1 - Resultados do algoritmo J48 para o semestre 2017.1

	Evadido	Não Evadido
Evadido	389	95
Não Evadido	159	200

Instâncias Classificadas Corretamente: 589 (69,87%)
Instâncias Classificadas Incorretamente: 254 (30,13%)
Precisão (Evadido): 71%
Precisão (Não Evadido): 68%

Fonte: Elaboração Própria.

Árvore de Decisão para o Semestre 2017.2

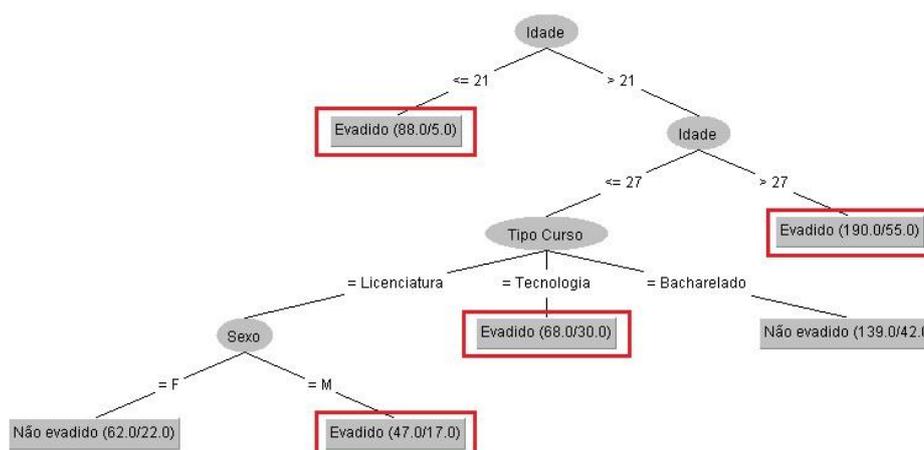


Figura 3: Árvore de decisão utilizando o algoritmo J48 para o semestre 2017.2.

Fonte: Elaboração Própria.

Na Figura 3 também são observadas duas faixas de idade com maior evasão: alunos com idade menor ou igual a 21 anos e alunos com idade superior a 27 anos. Cabe destacar que a faixa de idade até 21 anos se repete no semestre 2017.2, uma vez que ela também havia ocorrido no semestre anterior (2017.1). Para os alunos com idade entre 21 e 27 anos, foi identificada maior evasão para os atributos tipo de curso e sexo. Do mesmo modo que havia ocorrido no semestre 2017.1, os cursos de Tecnologia novamente foram os que possuíram mais alunos evadidos. Já para os cursos de Licenciatura, observa-se novamente que alunos do sexo masculino possuem maior evasão. A Tabela 3 apresenta os resultados obtidos pelo algoritmo J48 para o segundo semestre do ano de 2017:

Tabela 2 - Resultados do algoritmo J48 para o semestre 2017.2

	Evadido	Não Evadido
Evadido	276	74
Não Evadido	110	134

Instâncias Classificadas Corretamente: 410 (69,02%)
Instâncias Classificadas Incorretamente: 184 (30,98%)
Precisão (Evadido): 72%
Precisão (Não Evadido): 64%

Fonte: Elaboração Própria.

Árvore de Decisão para o Semestre 2018.1

A Figura 4 apresenta Árvore de decisão utilizando o algoritmo J48 para o semestre 2018.1:

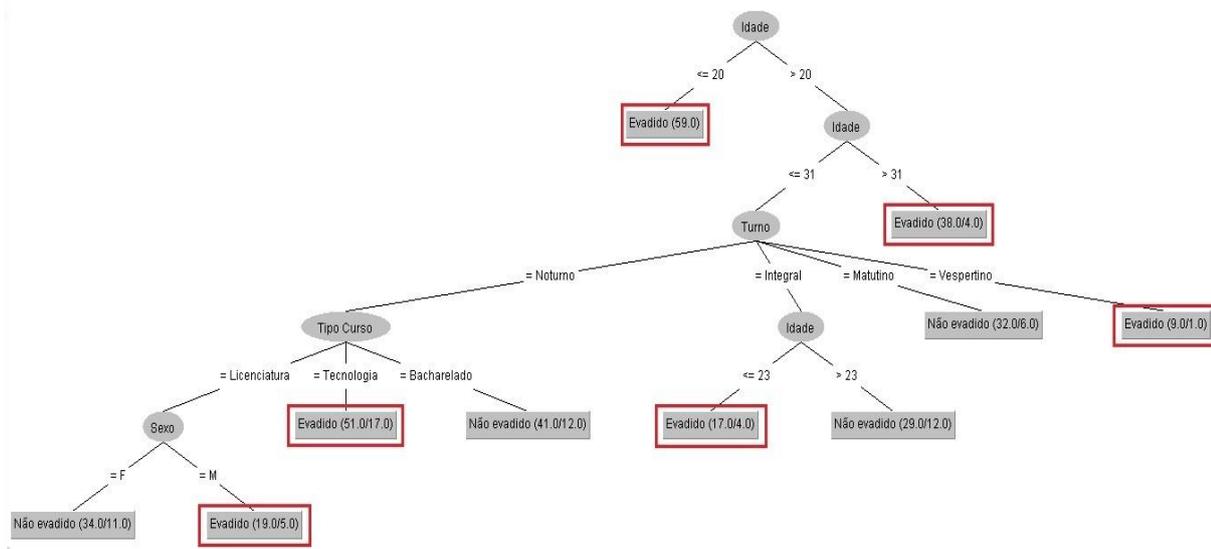


Figura 4: Árvore de decisão utilizando o algoritmo J48 para o semestre 2018.1.
Fonte: Elaboração Própria.

A árvore de decisão do semestre 2018.1 possui mais detalhes em relação aos atributos, como pode ser visto na Figura 4. Observa-se que mais uma vez apresentam-se duas faixas de idade diferentes com altas taxas de evasão: idade menor ou igual a 20 anos e idade maior que 31 anos. Para os alunos com idade entre 20 e 31 anos, a evasão estava relacionada principalmente aos atributos turno, tipo de curso e sexo. O atributo turno foi considerado relevante pelo algoritmo, destacando-se, em relação à evasão, os turnos noturno, integral e vespertino. O turno noturno possui relação com dois atributos: tipo de curso e sexo. Nesse turno, alunos do sexo masculino dos cursos de Licenciatura apresentaram maiores taxas de evasão, algo que também ocorreu nos semestres 2017.1 e 2017.2. Os cursos de Tecnologia, assim como nos outros semestres estudados, foram os que possuíram mais alunos evadidos. O turno integral ficou atrelado ao atributo idade, de modo que alunos desse turno, com idade menor ou igual a 23 anos são os mais evadidos. Em relação ao turno vespertino, o algoritmo apontou alta evasão, porém não relacionou esse turno a nenhum atributo específico.

A Tabela 4 apresenta os resultados obtidos pelo algoritmo J48 para o primeiro semestre de 2018:

Tabela 3 - Resultados do algoritmo J48 para o semestre 2018.1

	Evadido	Não Evadido
Evadido	150	53
Não Evadido	40	86

Instâncias Classificadas Corretamente: 236 (71,73%)
Instâncias Classificadas Incorretamente: 93 (28,26%)
Precisão (Evadido): 79%
Precisão (Não Evadido): 62%

Fonte: Elaboração Própria.

Identificação do comportamento de alunos evadidos

A utilização de árvores de decisão colabora com o cumprimento do objetivo do presente trabalho, que visa identificar o comportamento de alunos que por algum motivo evadiram. Desse modo, alunos matriculados, porém com comportamento semelhante aos de alunos evadidos podem ser identificados e trabalhados, a fim de evitar a evasão desses.

Em relação às árvores de decisão geradas, pode-se observar que o algoritmo J48 gera árvores através das quais o comportamento de alunos evadidos pode ser identificado. Apesar disso, essa é uma tarefa que nem sempre é fácil de ser realizada, pois dependendo do tamanho da árvore, existem muitos caminhos que podem ser percorridos, o que dificulta a identificação rápida de um aluno evadido, por exemplo.

Desse modo, visando facilitar a identificação do comportamento de alunos evadidos nas árvores geradas pelo algoritmo J48, foi elaborada a Tabela 5. Assim, ao realizar a busca do comportamento de um aluno evadido, essa consulta pode ser feita tanto através de análise das árvores de decisão, como através de consulta à referida tabela.

Tabela 4 - Comportamento de alunos evadidos utilizando o algoritmo J48

Semestre	Idade	Tipo de Curso	Sexo	Turno	Evadidos / Total	Porcent. de Evasão	
2017.1	<= 21				116 / 121	95,87%	
	> 30				123 / 170	72,35%	
	> 21 <= 30	Tecnologia			76 / 135	56,30%	
		Licenciatura	M		55 / 80	68,75%	
	> 24 <= 30	Licenciatura	F		29 / 52	55,77%	
2017.2	<= 21				83 / 88	94,32%	
	> 27				135 / 190	71%	
	> 21 <= 27	Tecnologia			38 / 68	56%	
		Licenciatura	M		30 / 47	63,83%	
2018.1	<= 20				59 / 59	100%	
	> 31				34 / 38	89,47%	
	> 20 <= 31				Vespertino	8 / 9	88,89%
		Tecnologia			Noturno	34 / 51	66,67%
		Licenciatura	M		Noturno	14 / 19	73,68%
	> 20 <= 23				Integral	13 / 17	76,47%

Fonte: Elaboração Própria.

CONSIDERAÇÕES FINAIS

O objetivo do presente trabalho é realizar a identificação do comportamento de alunos evadidos de cursos superiores do *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Para esse fim, foram utilizadas informações desta instituição de ensino referentes ao primeiro e segundo semestres do ano de 2017 e do primeiro semestre do ano de 2018. A mineração de dados foi realizada utilizando o software livre Weka, com o algoritmo J48, que gerou as árvores de decisão.

Com a aplicação desse algoritmo, foi possível a identificação de comportamento encontrado em todos os semestres analisados, como por exemplo, a alta evasão entre jovens com idade até 21

anos. Outro ponto encontrado nos dois semestres do ano de 2017 foi o de que a evasão relacionada à renda familiar de até dois salários mínimos exerce maior impacto nos cursos de tecnologia. Além desses, alguns outros comportamentos foram encontrados, porém exclusivos de cada semestre analisado. Destaca-se, no primeiro semestre de 2017, a identificação de alto índice de evasão de pessoas do sexo masculino (com idade entre 21 e 30 anos) nos cursos de Licenciatura. Já no segundo semestre do mesmo ano, foi observado alto índice de evasão de alunos com mais de 27 anos no turno noturno. O primeiro semestre de 2018, por sua vez, apresentou alto índice de evasão de alunos com idade maior que 31 anos e também para alunos com idade entre 20 e 31 anos no turno noturno e em cursos de Tecnologia.

Pode-se observar que o comportamento de evasão que mais prevaleceu nos três semestres está ligado a três atributos: idade, curso e sexo. Em relação à idade, isto foi identificado para alunos que possuem até 21 anos ou acima de 27 anos. Para aqueles menores de 21 anos, pode-se considerar que são pessoas que acabaram de sair de algum curso profissionalizante e que possuem uma necessidade inicial de trabalhar, entrar no mercado de trabalho. Já para os alunos maiores de 27 anos, pode-se considerar que são pessoas que já estão no mercado de trabalho, em ascensão em suas empresas, priorizando o serviço. Com esta possibilidade de ascensão, buscam através de um nível superior melhoria na carreira. Por outro lado, esta é uma idade em que muitos estão iniciando uma família. A evasão pesa na escolha pelo serviço devido à dependência desses familiares e o próprio convívio familiar, pois muitos optam por realizar cursos noturnos pelo fato de trabalharem, e este convívio não é frequente.

Os cursos de Tecnologia, mesmo estando totalmente regulares com o MEC, ainda sofrem certo preconceito por parte das pessoas, pois em muitos concursos não há direcionamento para esse tipo de curso. Pode-se observar que os cursos tradicionais de Bacharelado são os que possuem menor evasão.

Sobre o atributo sexo, tem-se uma maior evasão para o masculino, atrelado aos cursos de Licenciatura. A formação de professor ainda possui um grande público feminino e como citado anteriormente, muitas pessoas optam por priorizarem o trabalho, principalmente para homens considerados como chefes de família, que em muitos casos, ganham mais que as mulheres.

Considerando estas informações, conclui-se que os alunos que possuem maiores chances de evadir são aqueles com idade até 21 anos, do curso de Tecnologia e do sexo masculino.

Assim, as características encontradas podem ser comparadas com características de alunos que estejam matriculados, de modo que facilite a identificação de estudantes que podem estar em risco de evasão. De posse dessas informações, a instituição poderá adotar medidas que busquem a redução da evasão dos seus alunos, algo que é importante não somente para ela, mas para toda a região onde está situada, bem como para outras regiões, como por exemplo, os locais de residência e trabalho onde esses alunos poderão atuar.

REFERÊNCIAS

ASSEISS, M. S. G. **Aplicação do processo de descoberta de conhecimento em banco de dados acadêmico utilizando as tarefas de agrupamento e classificação**. Dissertação de mestrado—Ilha Solteira: Universidade Estadual Paulista - UNESP, 2017.

ASSIS, L. R. S. **Perfil de Evasão no Ensino Superior Brasileiro: uma Abordagem de Mineração de Dados**. Dissertação de mestrado—Brasília: Universidade de Brasília, 2017.

BRASIL. Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. **Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas**, v. 1, n. 2, 1996.

BRASIL. **Surgimento das escolas técnicas**. Disponível em: <<http://www.brasil.gov.br/educacao/2011/10/surgimento-das-escolas-tecnicas/>>. Acesso em: 9 ago. 2019.

BRASIL. Ministério da Educação. **Instituições da Rede Federal**. Disponível em: <<http://portal.mec.gov.br/rede-federal-inicial/instituicoes>>. Acesso em: 19 jul. 2019a.

BRASIL. Ministério da Educação. **Educação Profissional e Tecnológica (EPT)**. Disponível em: <<http://portal.mec.gov.br/educacao-profissional-e-tecnologica-ept>>. Acesso em: 9 ago. 2019b.

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, v. 42, n. 3, p. 495–528, jun. 2008.

HEGDE, V.; PRAGEETH, P. P. **Higher education student dropout prediction and analysis through educational data mining**. 2018 2nd International Conference on Inventive Systems and Control (ICISC). **Anais...** In: 2018 2nd International Conference on Inventive Systems and Control (ICISC). Coimbatore: IEEE, jan. 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8398887/>>. Acesso em: 21 dez. 2018

HIPÓLITO, O. **O gargalo do Ensino Superior brasileiro — CartaCapital**, 27 abr. 2011. Disponível em: <<https://www.cartacapital.com.br/sociedade/o-gargalo-do-ensino-superior-brasileiro>>. Acesso em: 7 dez. 2018

LAM-ON, N.; BOONGOEN, T. **Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings.** [s.l.: s.n.]. v. 8, 2014.

LORENZETT, C. D. C.; TELOKEN, A. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão. **Simpósio de Pesquisa e Desenvolvimento em Computação**, v. 2, n. 1, 11 maio 2016.

MÁRQUEZ, C. et al. Early Dropout Prediction using Data Mining: A Case Study with High School Students. **Expert Systems**, v. 33, p. 107–124, 1 fev. 2016.

MEHTA, A. A.; BUCH, N. J. **Depth and breadth of educational data mining: Researchers' point of view.** 2016 10th International Conference on Intelligent Systems and Control (ISCO). **Anais...** In: 2016 10TH International Conference on Intelligent Systems and Control (ISCO). 7 jan. 2016.

SCHMITT, R. E. et al. **A evasão na Educação Superior no Brasil: uma análise da produção de conhecimento nos periódicos Qualis entre 2000-2011.** [s.l.] E.U.I.T. de Telecomunicación, 2012.

SILVA, A. M. DA; SANTOS, B. C. S. Eficácia de políticas de acesso ao ensino superior privado na contenção da evasão. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 22, n. 3, p. 741–757, dez. 2017.

SILVA FILHO, R. L. L. E et al. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, v. 37, n. 132, p. 641–659, dez. 2007.

SOUZA, C. T.; SILVA, C. DA; GESSINGER, R. M. Um estudo sobre evasão no ensino superior do Brasil nos últimos dez anos. **Congressos CLABES**, v. 0, n. 0, 9 out. 2017.

VIANA, G.; LIMA, J. F. DE. Capital humano e crescimento econômico. **Interações**, v. 11, n. 2, p. 137–148, 1 dez. 2010.

WEKA. Nova Zelândia: The University of Waikato, 2018.

WITTEN, I. H.; FRANK, E.; HALL, M. A. (EDS.). **Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)**. 3. ed. Boston: Morgan Kaufmann, 2011.