

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO PÚBLICA**

**SERGIO RODRIGUES SILVEIRA**

**EVASÃO DE ALUNOS DE GRADUAÇÃO: ESTUDO COM BANCOS DE DADOS  
NOS CENTROS DA UFES EM ALEGRE**

ALEGRE

2019

SERGIO RODRIGUES SILVEIRA

**EVASÃO DE ALUNOS DE GRADUAÇÃO: ESTUDO COM BANCOS DE DADOS  
NOS CENTROS DA UFES EM ALEGRE**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão Pública do Centro de Ciências Jurídicas e Econômicas da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Título de Mestre em Gestão Pública.

Orientador: Prof. Dr. Clovis Eduardo Nunes Hegedus.

ALEGRE

2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de  
Bibliotecas - SIBI/UFES e elaborada pelo autor

---

S587e Silveira, Sergio Rodrigues, 1973-  
Evasão de alunos de graduação : estudo com banco de dados  
nos centros da UFES em Alegre / Sergio Rodrigues Silveira. -  
2019.  
75 f. : il.

Orientador: Clovis Eduardo Nunes Hegedus.  
Dissertação (Mestrado Profissional em Gestão Pública) -  
Universidade Federal do Espírito Santo, Centro de Ciências  
Jurídicas e Econômicas.

1. Evasão universitária. 2. Mineração de dados (Computação).  
3. Classificação. I. Hegedus, Clovis Eduardo Nunes. II.  
Universidade Federal do Espírito Santo. Centro de Ciências  
Jurídicas e Econômicas. III. Título.

CDU: 35

---

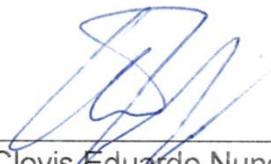
SERGIO RODRIGUES SILVEIRA

**EVASÃO DE ALUNOS DE GRADUAÇÃO: ESTUDO COM BANCOS DE DADOS  
NOS CENTROS DA UFES EM ALEGRE**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão Pública do Centro de Ciências Jurídicas e Econômicas da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Título de Mestre em Gestão Pública.

Aprovado em 17 de junho de 2019.

**COMISSÃO EXAMINADORA**



---

Prof. Dr. Clovis Eduardo Nunes Hegedus  
Orientador



---

Prof. Dr. Wendel Sandro de Paula Andrade  
Membro Interno



---

Prof. Dr. Antonio Almeida de Barros Junior  
Membro Externo – UFES

Dedicado a todos que buscam produzir e popularizar o conhecimento.

## **AGRADECIMENTOS**

Ao Professor Clovis Eduardo Nunes Hegedus, meu orientador, pela devoção à docência e pelo companheirismo expressado durante toda esta etapa de minha formação acadêmica.

Ao Professor Bruno Vilela Oliveira, pela valiosa ajuda prestada em minha busca pela compreensão de diversos conceitos em mineração de dados.

Aos Professores Antonio Almeida de Barros Junior e Wendel Sandro de Paula Andrade, que aceitaram participar das bancas de qualificação e defesa desta dissertação, dedicando tempo e atenção que resultaram em importantes contribuições para o presente trabalho.

Aos professores do mestrado e aos companheiros de turma, pelas experiências compartilhadas durante as noites de curso.

Aos meus familiares e amigos, pelas palavras de incentivo e por compreenderem os momentos em que precisei estar ausente.

“Um oráculo declarara que a Esfinge se destruiria a si própria no dia em que seus enigmas fossem decifrados.” (SÓFOCLES, 1990, p. 98)

## RESUMO

A evasão é um fenômeno que representa um problema para as instituições de ensino, sejam públicas ou particulares, ocasionando perda de recursos de todos os envolvidos no processo de ensino. A Mineração de Dados Educacionais é uma área de pesquisa recente que tem se demonstrado promissora no desenvolvimento de métodos para a exploração de dados produzidos em ambientes educacionais, ajudando na elaboração de indicadores para mitigação dos fatores que impedem a permanência de alunos. O objetivo principal da presente pesquisa foi obter informações que proporcionem a aquisição de conhecimentos relacionados à evasão estudantil na Universidade Federal do Espírito Santo e que possibilitem a tomada de decisões que favoreçam a permanência de seus alunos. A pesquisa realizou uma abordagem quantitativa, utilizando dados secundários para obter padrões e modelos que permitiram realizar análises preditivas com suas variáveis. Quanto a seus fins, tratou-se de uma pesquisa descritiva e metodológica. Como fonte de dados, foram utilizadas tabelas contendo informações sociais e acadêmicas de alunos matriculados nos anos de 2007 a 2018 nos dois centros da UFES, localizados no município de Alegre. O pré-processamento, a transformação e a mineração dos dados foram realizados com auxílio do programa *RapidMiner Studio*, versão 9.2.001, plataforma WIN64. Com os resultados obtidos, concluiu-se que é possível ampliar o conhecimento e a capacidade de tomada de decisões da gestão universitária da UFES para reduzir os índices de evasão na instituição. Como produto técnico resultante da pesquisa, elaborou-se um conjunto de instruções para que a gestão pública da UFES possa utilizar o mesmo processo aplicado neste trabalho para realizar previsões com alunos futuros.

Palavras-chave: Evasão. Permanência. Mineração de Dados. Classificação.

## **ABSTRACT**

Evasion is a phenomenon that represents a problem for educational institutions, whether public or private, causing loss of resources for all those involved in the teaching process. Educational Data Mining is a recent area of research that has shown promise in the development of methods for the exploration of data produced in educational environments, helping in the elaboration of indicators to mitigate the factors that prevent the permanence of students. The main objective of the present research was to obtain information that provides the acquisition of knowledge related to student evasion at the Federal University of Espírito Santo and that make it possible to make decisions that favor the permanence of its students. The research carried out a quantitative approach, using secondary data to obtain patterns and models that allowed to perform predictive analyzes with its variables. As for its purposes, it was a descriptive and methodological research. As data source, tables containing social and academic information of students enrolled in the years 2007 to 2018 were used in the two UFES centers, located in the city of Alegre. Data preprocessing, transformation and mining were performed using the RapidMiner Studio program, version 9.2.001, WIN64 platform. With the obtained results, it was concluded that it is possible to increase the knowledge and the decision-making capacity of the UFES university management to reduce the rates of evasion in the institution. As a technical product resulting from the research, a set of instructions was elaborated so that the UFES public management can use the same process applied in this work to make predictions with future students.

**Keywords:** Evasion. Permanence. Data Mining. Classification.

## LISTA DE FIGURAS

Figura 1 – Visão geral das etapas do processo de descoberta de conhecimento em base de dados.....	29
Figura 2 – Modelo esquemático do método Hold Out .....	33
Figura 3 – Modelo esquemático de validação cruzada.....	34
Figura 4 – Processo utilizado na mineração dos dados obtidos.....	42
Figura 5 – Operadores de classificação e de desempenho utilizados.....	43
Figura 6 – Interface de predições do programa <i>RapidMiner</i> .....	51

## LISTA DE QUADROS

Quadro 1 – Modelo esquemático de uma matriz de confusão .....	34
Quadro 2 – Rótulos qualitativos do índice Kappa .....	37
Quadro 3 – Relação de atributos das tabelas obtidas na Secretaria Única de Graduação – Setorial Sul .....	44
Quadro 4 – Características dos atributos encontrados .....	45
Quadro 5 – Atributos utilizados na construção do modelo de predição.....	47
Quadro 6 – Valores de atributos com maior probabilidade para evasão voluntária.....	49
Quadro 7 – Matriz de confusão resultante da validação cruzada.....	53

## LISTA DE SIGLAS

AVA	Ambiente Virtual de Aprendizagem
CAUFES	Centro Agropecuário da Universidade Federal do Espírito Santo
CCAIE	Centro de Ciências Agrárias e Engenharias
CCAUFES	Centro de Ciências Agrárias da Universidade Federal do Espírito Santo
CCENS	Centro de Ciências Exatas, Naturais e da Saúde
CRA	Coeficiente de Rendimento do Aluno
CRN	Coeficiente de Rendimento Normalizado
ENEM	Exame Nacional do Ensino Médio
ESAES	Escola Superior de Agronomia do Espírito Santo
MDE	Mineração de Dados Educacionais
MEC	Ministério da Educação e Cultura
PNAES	Programa Nacional de Assistência Estudantil
PROAECI	Pró-Reitoria de Assuntos Estudantis e Cidadania
PROAES-UFES	Programa de Assistência Estudantil da Universidade Federal do Espírito Santo
RU	Restaurante Universitário
SIE	Sistema de Informações para o Ensino
SISU/MEC	Sistema de Seleção Unificada do Ministério da Educação
UFES	Universidade Federal do Espírito Santo
UFPB	Universidade Federal da Paraíba
UFRJ	Universidade Federal do Rio de Janeiro
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	13
1.1	CONTEXTO E PROBLEMA DE PESQUISA .....	14
1.2	OBJETIVO GERAL.....	14
<b>1.2.1</b>	<b>Objetivos específicos</b> .....	15
1.3	DELIMITAÇÃO DA PESQUISA .....	15
1.4	JUSTIFICATIVA.....	16
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	17
2.1	HISTÓRICO DO ENSINO SUPERIOR PÚBLICO NO BRASIL .....	17
2.2	A INSTITUIÇÃO E OS CENTROS PESQUISADOS.....	18
2.3	EVASÃO NO ENSINO SUPERIOR E SEUS FATORES .....	20
<b>2.3.1</b>	<b>A UFES e seu contexto de evasão</b> .....	22
2.4	RESERVA DE VAGAS E ASSISTÊNCIA ESTUDANTIL NO BRASIL .....	23
<b>2.4.1</b>	<b>O programa de assistência estudantil da UFES</b> .....	24
2.5	MINERAÇÃO DE DADOS E SUAS APLICAÇÕES EDUCACIONAIS .....	26
<b>2.5.1</b>	<b>Tipos de dados utilizados</b> .....	28
<b>2.5.2</b>	<b>O processo de descoberta de conhecimento com base de dados</b> .....	29
<b>2.5.3</b>	<b>Regra <i>Naive</i> de Bayes</b> .....	30
<b>2.5.4</b>	<b>Ganho de informação de atributos</b> .....	31
<b>2.5.5</b>	<b>Avaliação de modelos de aprendizado de máquina</b> .....	32
2.5.5.1	Método <i>Hold Out</i> .....	33
2.5.5.2	Validação cruzada .....	33
2.5.5.3	Matriz de confusão .....	34
2.5.5.4	Coeficiente Kappa .....	35
<b>2.5.6</b>	<b>Mineração de dados educacionais</b> .....	37
2.5.6.1	Ferramentas utilizadas .....	38
2.5.6.2	Trabalhos relacionados .....	38
<b>3</b>	<b>MÉTODOS E PROCEDIMENTOS</b> .....	40
3.1	COLETA DOS DADOS .....	40
3.2	TRATAMENTO DOS DADOS.....	41
3.3	MINERAÇÃO DOS DADOS.....	42
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	44
4.1	CARACTERÍSTICAS DOS DADOS COLETADOS.....	44

4.2	PREPARAÇÃO DOS DADOS ENCONTRADOS.....	47
4.3	CARACTERÍSTICAS DO ALUNO QUE EVADE VOLUNTARIAMENTE .....	49
4.4	CLASSIFICAÇÃO DE ALUNOS EM RISCO POTENCIAL DE EVASÃO .....	51
4.5	AVALIAÇÃO DO MÉTODO EMPREGADO .....	53
<b>5</b>	<b>CONCLUSÃO</b> .....	<b>54</b>
	<b>REFERÊNCIAS</b> .....	<b>56</b>
	APÊNDICE A – PRODUTO TÉCNICO RESULTANTE DA DISSERTAÇÃO..	63

## 1 INTRODUÇÃO

A evasão é um problema que tem causado preocupação para as instituições de ensino em geral, sejam elas públicas ou particulares. A saída dos alunos traz consigo sérias consequências, uma vez que representa uma perda de recursos não só de organizações governamentais e privadas, mas também de alunos e seus familiares, que investem tempo e parte de suas provisões no esforço de construir seu futuro.

A evasão é medida como a diferença entre a quantidade de alunos que ingressam num determinado curso, em determinado ano, e a quantidade desses mesmos alunos que conseguem concluir seus respectivos cursos. Essa medida inclui todos os alunos que abandonaram seus cursos independentemente dos motivos ou causas, excluindo-se os falecidos (ROSA, 1994).

As pesquisas com o propósito de investigar a evasão de alunos em instituições públicas ou privadas costumam apontar uma taxa de evasão em torno de 20% a 50%, sendo que as maiores incidências ocorrem no primeiro ano de curso, e os três primeiros meses são os que apresentam o maior número de alunos evadidos (BARDAGI; HUTZ, 2009).

A adoção de estratégias de combate à evasão estudantil ajuda a reduzir esses índices entre alunos pertencentes a grupos socioeconomicamente vulneráveis. Silveira (2012) defende que as ações de assistência estudantil são fundamentais para a permanência no ensino superior dos grupos estudantis que se enquadram na renda *per capita* familiar de até um salário mínimo e meio, possibilitando a conclusão da graduação desses alunos com qualidade e igualdade de oportunidades.

Uma recente área de pesquisa que tem se demonstrado promissora é a Mineração de Dados Educacionais (MDE), do inglês *Educational Data Mining*, pois permite o desenvolvimento de métodos para a exploração de dados que são produzidos em ambientes educacionais. Nesta abordagem, Baker, Carvalho e Isotani (2011) afirmam que, atualmente, a MDE vem se estabelecendo como uma forte linha de pesquisa com potencial para produzir melhorias na qualidade de ensino. Cambuzzi, Cazella e Rigo (2012) ressaltam que, perante a diversidade de dados envolvidos no contexto da evasão escolar, a exploração de algoritmos, técnicas e mecanismos de

MDE são cruciais para a obtenção de resultados abrangentes e úteis para a mitigação dos fatores que impedem a permanência dos alunos.

### 1.1 CONTEXTO E PROBLEMA DE PESQUISA

A evasão estudantil no ensino superior é um fenômeno que vem crescendo no Brasil nos últimos anos. De acordo com Sales Junior (2013), entre as diversas causas estão os programas de expansão que promoveram a ampliação do número de instituições de ensino superior, de cursos oferecidos e de novas vagas. Conseqüentemente, ocorreu também o aumento do número de estudantes dos cursos de graduação que abandonam seus estudos antes de obter a diplomação.

Nos últimos anos, a Universidade Federal do Espírito Santo (UFES) também passou por um crescimento em evasões. Em sua pesquisa contendo dados estatísticos de evasão na UFES, o autor supracitado observou um expressivo aumento no índice de evasões ocorrido a partir de 2010, passando de 4,2% em 2009 para 7,7% em 2010. Também destacou que apenas no primeiro semestre de 2012 já se registravam mais evasões que em todo o ano de 2011.

Perante o exposto, o presente estudo pretendeu responder o seguinte questionamento: é possível para a gestão universitária da UFES ampliar seu conhecimento e sua capacidade de tomada de decisões através da descoberta de informações contidas em seus bancos de dados de evasão estudantil?

### 1.2 OBJETIVO GERAL

Na expectativa de responder a questão que envolve o problema, o presente trabalho teve como principal objetivo: obter informações que proporcionem a aquisição de conhecimentos relacionados à evasão estudantil na UFES e possibilitem a tomada de decisões que favoreçam a permanência de seus alunos.

O estudo adotou a hipótese de que a MDE permite identificar fatores potenciais que conduzem à evasão nos centros da UFES em Alegre, permitindo classificar, antecipadamente, alunos com possibilidade de evadirem-se.

### 1.2.1 Objetivos específicos

Para alcançar o objetivo almejado, a pesquisa se norteou pelos seguintes objetivos específicos:

- a) obter informações sociais e acadêmicas de alunos dos dois centros de ciências da UFES em Alegre, através de acesso a banco de dados digitais relacionados à evasão de estudantes de graduação;
- b) realizar o tratamento e mineração dos dados obtidos;
- c) identificar características dos alunos que evadem voluntariamente nos centros pesquisados;
- d) possibilitar a identificação prévia de alunos em risco potencial de evasão.

### 1.3 DELIMITAÇÃO DA PESQUISA

Atualmente, a UFES dispõe de 103 cursos de graduação presencial, oferecendo 5.004 vagas para novos alunos a cada ano. Na pós-graduação, a universidade possui 47 cursos de mestrado acadêmico, 9 de mestrado profissional e 26 cursos de doutorado oferecidos. O número total de matrículas na graduação é de 19.997 alunos; na pós-graduação, conta com 3.174 matriculados. Os cursos são ministrados em quatro *campi* universitários distribuídos no estado do Espírito Santo: Goiabeiras e Maruípe, na capital Vitória; e nos municípios de Alegre, localizado no interior sul, e São Mateus, no norte do estado (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2013a).

Devido à proximidade e maior facilidade de acesso aos dados, optou-se pela aplicação desta pesquisa nos 17 cursos de graduação presencial ofertados pelo Centro de Ciências Agrárias e Engenharias (CCAEE) e pelo Centro de Ciências Exatas, Naturais e da Saúde (CCENS), ambos localizados no município de Alegre.

#### 1.4 JUSTIFICATIVA

Os crescentes índices de evasão estudantil na UFES representam um desafio que requer a ação de seus gestores visando combatê-la constantemente. As diretrizes adotadas através do Programa de Assistência Estudantil da UFES (PROAES-UFES) são iniciativas que viabilizam a permanência de alunos em situação de vulnerabilidade socioeconômica. Ações como essas podem ser potencializadas mediante a aquisição de um maior conhecimento da realidade de seus alunos.

A MDE é uma área de estudos capaz de produzir mecanismos que facilitam a obtenção de conhecimento e a tomada de decisão, aproveitando os bancos de dados produzidos em instituições de ensino. Sua utilização enriquece o conjunto de possibilidades para a melhoria do desempenho escolar.

Deste modo, a pesquisa justifica-se por oferecer à gestão pública da UFES a possibilidade de ampliar seu conhecimento relacionado à evasão estudantil, ajudando-a a maximizar as iniciativas já implantadas e a encontrar novas estratégias para propiciar a permanência nos cursos a um maior número de alunos.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo traz um breve histórico do ensino superior no Brasil, as iniciativas de auxílio estudantil do Governo Federal no País e os principais aspectos da assistência estudantil na UFES. Em seguida, relata sobre a Mineração de Dados e sua importância na área da educação.

### 2.1 HISTÓRICO DO ENSINO SUPERIOR PÚBLICO NO BRASIL

O ensino superior no Brasil teve seu início após o período colonial com a chegada da família real portuguesa no País em 1808, quando foram fundadas as primeiras escolas com a finalidade de atingir esse grau de ensino. Contudo, seu desenvolvimento ocorreu lentamente. Nesse período, as escolas de ensino superior funcionavam como faculdades isoladas, seguindo um modelo de formação de profissionais liberais, cujo diploma visava assegurar a ocupação de postos privilegiados em um mercado de trabalho restrito, garantindo prestígio social a seus formandos. Sobre esse aspecto, Martins (2002) afirma que esse caráter não universitário do ensino não representava um demérito para a formação superior da época, pois os cursos eram de longa duração, e o nível dos docentes devia se equiparar ao da Universidade de Coimbra.

Após a instauração da República em 1889, a ideologia positivista existente entre os oficiais que a proclamaram contribuiu para influenciar o atraso na criação de universidades no Brasil. Seus primeiros políticos eram estritamente adeptos à criação de cursos laicos de orientação ou técnica profissionalizante e consideravam as universidades como instituições ultrapassadas e anacrônicas do Velho Continente, não adaptadas à realidade do Novo Mundo (OLIVEN, 2002).

Em razão desse pensamento, somente em 1920 ocorreria a criação oficial da primeira universidade brasileira, como resultado do Decreto nº 14.343, que institucionalizou a Universidade do Rio de Janeiro. Conforme relata Oliven (2002), seu início não representou uma alternativa diferente do sistema de ensino superior realizado até então. A instituição mantinha o caráter profissionalizante de seus cursos e continuava mais voltada ao ensino do que à pesquisa.

Os anos seguintes, nessa década, foram marcados por uma série de debates que culminaram na Reforma Campos de 1931, que se consubstanciou no chamado Estatuto das Universidades Brasileiras. Esse documento previa a possibilidade de incluir, no conjunto de escolas que formariam as universidades, uma Faculdade de Educação, Ciências e Letras destinada a formar professores, especialmente para o ensino normal e secundário. Contudo, Mendonça (2000) afirma que o modelo universitário não se impôs nesse período, e as faculdades seguiram expandindo-se como instituições isoladas, formando professores para a escola secundária, muitas vezes precariamente.

Após a Segunda Guerra Mundial, com o fim do Estado Novo no Brasil, o País passa por um processo de redemocratização que se denota no caráter liberal encontrado na Constituição de 1946. Inicia-se, também, a consciência de diversos setores da sociedade sobre a necessidade de uma reformulação na universidade brasileira. Através de intensa participação do movimento estudantil, foram discutidas questões relevantes com o intuito de desfazer o caráter arcaico e elitista das universidades no País. Entretanto, conforme salienta Fávero (2006), nos finais de 1967 aumenta a preocupação do Governo com uma possível subversão estudantil. Com isso, são criadas medidas visando fortalecer o princípio da autoridade e disciplina nas instituições de ensino superior.

Segundo Mazzilli (2011), somente na década de 1980 a universidade brasileira daria os primeiros sinais no caminho de se tornar uma instituição voltada para o ensino, pesquisa e extensão em busca de soluções para os problemas sociais.

## 2.2 A INSTITUIÇÃO E OS CENTROS PESQUISADOS

A UFES foi criada em maio de 1954, após o Governador Jones dos Santos Neves sancionar a Lei nº 806, aprovada pela Assembleia Legislativa, que passou a reunir faculdades e escolas de ensino superior existentes no estado. Depois de a universidade passar seus primeiros quatro anos com dificuldades para se consolidar, o Presidente da República Juscelino Kubitschek sancionou a lei que tornava a UFES uma instituição federal de ensino. Conforme é relatado em seu livro que traz o histórico narrando os 60 anos da instituição, a Universidade Federal do Espírito

Santo (2014) relembra que a federalização partiu de uma negociação entre Governo do Estado e a Assembleia Legislativa com a finalidade de salvar a instituição, que passava por graves dificuldades para se consolidar devido à estagnação de suas instâncias deliberativas, da falta de reconhecimento pelo Ministério da Educação, e com a falta de soluções para a instalação de seu campus universitário.

Em 18 de março de 1971, a Universidade incorporou a antiga Escola Superior de Agronomia do Espírito Santo (ESAES). Sua primeira turma, composta por 23 alunos do curso de Agronomia, recebeu a colação de grau em dezembro de 1974. No ano seguinte, o curso de Agronomia é reconhecido pelo Ministério da Educação e Cultura (MEC) através do Decreto nº 75.235, de 16 de janeiro de 1975, assinado pelo Presidente da República Ernesto Geisel. Em 1976, é transformada no Centro Agropecuário da UFES (CAUFES), com sede no município de Alegre (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2013b).

Em 2001, com a criação dos cursos de Engenharia Florestal, Medicina Veterinária e Zootecnia, o Centro recebe nova estruturação jurídico-institucional e passa a denominar-se Centro de Ciências Agrárias da UFES (CCAUFES). Em 2005 e 2009, novos cursos foram criados através do Projeto de Expansão das Instituições Federais de Ensino e do Projeto de Apoio a Planos de Reestruturação das Universidades Federais (REUNI), são eles: Química; Física; Biologia; Licenciatura em Matemática e Ciência da Computação; Sistemas de Informação; Engenharia Química e Bacharelado em Farmácia (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2013b).

O CCAUFES foi desmembrado através da Resolução nº 44/2015 do Conselho Universitário, passando a formar dois centros: CCAE e CCENS (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2015).

Atualmente, as formas de ingresso nos cursos presenciais da UFES incluem: vestibular, processo seletivo que é realizado uma vez por ano, destinado a quem já concluiu ou está concluindo o ensino médio; remoção, para alunos matriculados na instituição que desejam mudar de turno; reopção de curso, para os que desejam mudar de curso dentro da própria universidade; transferência, destinada a alunos de outras instituições de ensino superior que desejam se transferir para a UFES; novo curso, onde o interessado que já possua curso superior poderá concluir outro curso

mediante processo seletivo que não o vestibular; aluno especial, em que são permitidas matrículas em disciplinas isoladas por quem possua diploma de curso superior ou esteja matriculado em outra instituição de ensino superior (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2013d).

Em 2016, a UFES passou a aderir ao Sistema de Seleção Unificada do Ministério da Educação (SISU/MEC), através da Resolução nº 13/2016, como forma de ingresso em seus cursos de graduação (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2016). De acordo com o portal do Ministério da Educação (2019), o SISU/MEC é o sistema informatizado utilizado pelo ministério através do qual as instituições públicas de ensino superior oferecem vagas a candidatos participantes do Exame Nacional do Ensino Médio (ENEM).

### 2.3 EVASÃO NO ENSINO SUPERIOR E SEUS FATORES

A evasão escolar, descrita como a saída de estudantes antes da conclusão dos cursos em que estavam matriculados, tem sido objeto de estudos devido à importância da compreensão de seus fatores para a adoção de medidas que amenizem os prejuízos causados por sua ocorrência. Sobre esse aspecto, Rosa (1994) salienta que a consequência imediata da evasão é a oneração do custo por aluno graduado, produto final das instituições de ensino. O autor destaca os prejuízos causados ao aluno, que investiu seu tempo e recursos para fazer um vestibular; e também o desperdício com estrutura e custos fixos elevados da instituição de ensino, resultando em prejuízo para a nação, que é privada de pessoas habilitadas.

Em relatório apresentado pela Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (1996), observou-se a necessidade de tornar clara a definição do conceito de evasão. No estudo, o termo foi considerado através de três dimensões concretas: evasão do curso, evasão da instituição e evasão do sistema de ensino superior. Evasão do curso ocorre quando o aluno desliga-se do curso em situações diversas, como: abandono, desistência, transferência/reopção e exclusão por norma institucional; evasão da instituição consiste no desligamento, por parte do aluno, da instituição em que estava matriculado; e evasão do sistema aborda as

situações onde o aluno abandona o ensino superior de forma definitiva ou temporária.

Ao estudar as causas da evasão em alunos do Instituto de Física da Universidade Federal do Rio de Janeiro (UFRJ), Barroso e Falcão (2014) classificaram os motivos que levam à evasão em três grandes grupos: evasão econômica, relacionada à impossibilidade de manutenção do vínculo com a instituição de ensino por questões socioeconômicas; evasão vocacional, que decorre da percepção de uma escolha inadequada aos interesses do estudante; e evasão institucional, descrita como o abandono por inadequação ou fracasso na escolha e na permanência dentro do instituto.

Em outro estudo elaborado por Pereira (2003), os fatores que influenciam na decisão do aluno por abandonar a instituição foram divididos em dois grupos: fatores internos às instituições e fatores inerentes ao estudante. O primeiro engloba aspectos como infraestrutura deficitária, acervo desatualizado, métodos de avaliação e deficiência didático-pedagógica dos professores; o outro, aspectos ligados a dificuldades financeiras do aluno, escolha equivocada do curso, ausência de base para acompanhar o curso escolhido e o fato de o curso não ser a sua primeira opção.

De acordo com Manhães *et al.* (2011), os índices de evasão variam entre as universidades e entre os cursos. Com isso, cada instituição de ensino realiza seus estudos sobre as causas de evasão e produz as medidas preventivas necessárias com aspectos fortemente ligados ao contexto de sua realidade. Nessa perspectiva, os mesmos autores afirmam que identificar esses fatores que influenciam na evasão e atribuir-lhes uma ordem de importância é uma tarefa complexa que está diretamente ligada à análise do conjunto de alunos. Conforme descrevem, é preciso considerar que:

Existem várias condições que propiciam evasão. Por exemplo, no contexto institucional deve ser considerado que cada universidade possui características que atraem maior ou menor número de alunos em função de sua localização geográfica, público alvo, adequação dos cursos ao contexto sócio-econômico da região. No contexto do curso de graduação algumas condições são a atualidade do currículo do curso, sua adequação à formação para o mercado de trabalho e tempo de duração. (MANHÃES *et al.*, 2011, p. 151)

Nessa ótica, observa-se a importância da adequação dos cursos ao contexto geográfico da região onde a instituição funciona, como também devem estar em sintonia com os anseios de seu público alvo e com as exigências do mercado de trabalho. A instituição deve ter o objetivo de atender as necessidades de seus alunos, baseando-se nas condições socioeconômicas em que estão inseridos.

### 2.3.1 A UFES e seu contexto de evasão

Nos últimos anos, a UFES tem passado por um crescimento no número de evasões, conforme pode ser observado na Tabela 1. De acordo com Sales Junior (2013), a universidade passou por um expressivo aumento em evasões a partir de 2010, principalmente através de desistência e desligamento por abandono, formas estas em que o aluno evade voluntariamente.

**Tabela 1 – Número de evasões ocorridas entre 2007/1 e 2012/1 na UFES**

Forma de Evasão	Ano						Total 2007 a 2012/1
	2007	2008	2009	2010	2011	2012/1	
Desistência	297	248	239	469	531	359	2.143
Desligamento por abandono	444	101	10	627	468	708	2.358
Desligamento por mandato judicial	0	0	2	1	4	1	8
Desligamento: três reprovações em uma disciplina	9	27	311	91	101	117	656
Desligamento: curso extinto	0	0	0	0	0	6	6
Desligamento: descumpriu plano de estudos	4	2	3	32	2	4	47
Falecimento	0	3	7	0	0	0	10
Jubilado	0	0	0	0	3	0	3
Matrícula desativada	0	0	0	0	1	1	2
Não informado	1	2	4	14	7	0	28
Reopção de curso	127	78	78	70	47	41	441
Sanção disciplinar	1	0	0	0	0	0	1
Transferência	0	0	2	0	0	0	2
Transferência interna	22	6	4	27	3	0	62
Transferido	29	15	26	10	22	17	119
<b>Total</b>	<b>934</b>	<b>482</b>	<b>686</b>	<b>1.341</b>	<b>1.189</b>	<b>1.254</b>	<b>5.886</b>

Fonte: Sales Junior (2013).

Nota: O total de evasões no ano de 2012 foi de 3.792 alunos, de acordo com Azevedo (2017).

O autor também destaca que apenas no primeiro semestre de 2012 registraram-se mais evasões do que em todo o ano letivo de 2011.

Esse pico de evasão em 2012 também foi observado por Azevedo (2017), cuja pesquisa totalizou 3.792 alunos evadidos durante todo aquele ano na instituição. Em 2013, os números voltaram ao mesmo patamar de 2010 e 2011. Nos resultados de seu estudo, demonstrou-se que o fenômeno da evasão de estudantes dos cursos de graduação é decorrente de diversos fatores, sendo o principal deles a dificuldade de conciliar os estudos com o sustento financeiro.

#### 2.4 RESERVA DE VAGAS E ASSISTÊNCIA ESTUDANTIL NO BRASIL

Em 12 de abril de 2012, o Governo Federal instituiu a Lei nº 12.711, que estabelece diretrizes a respeito do ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio. A lei prevê que as instituições federais de ensino reservem 50% de suas vagas para estudantes que tenham concluído integralmente o ensino médio em escolas da rede pública, e destine, entre essas vagas, 50% para estudantes com renda *per capita* familiar igual ou inferior a um salário mínimo e meio. É garantida também a participação nas vagas reservadas de estudantes autodeclarados pretos, pardos e indígenas em proporção à representação destes na população da unidade federativa em que a instituição de ensino esteja estabelecida (BRASIL, 2012).

Os estudantes portadores de deficiência também passaram a ser incluídos nas reservas de vagas, juntamente com estudantes pretos, pardos e indígenas, a partir de 2016 com a Lei nº 13.409/2016, que alterou a norma de 2012 (BRASIL, 2016).

Outra iniciativa adotada pelo Governo Federal para garantir acesso ao ensino superior a grupos sociais desfavorecidos é através da assistência estudantil. Constituída de um conjunto de ações, programas e projetos de inclusão social, as políticas de assistência estudantil possuem o propósito de dar aos estudantes de classes sociais desfavorecidas e com vulnerabilidade socioeconômica as condições necessárias à permanência nas instituições de ensino superior, buscando reduzir a retenção e a evasão escolar. Borges (2015) afirma que as primeiras políticas de assistência estudantil começaram a surgir a partir da década de 1930, mas que

somente a Constituição Federal de 1988 veio a tratar a permanência na escola através do princípio da igualdade.

Atualmente, o Programa Nacional de Assistência Estudantil (PNAES) estabelece as diretrizes nacionais voltadas a ampliar a permanência dos jovens estudantes na educação superior pública federal. Entre seus objetivos destacam-se:

- I – democratizar as condições de permanência dos jovens na educação superior pública federal;
- II – minimizar os efeitos das desigualdades sociais e regionais na permanência e conclusão da educação superior;
- III – reduzir as taxas de retenção e evasão; e
- IV – contribuir para a promoção da inclusão social pela educação. (BRASIL, 2010, p. 5)

Dessa forma, as ações do PNAES dão destaque não só à importância da inclusão social pela educação, mas também à minimização das desigualdades sociais que impedem a permanência dos jovens na educação superior pública federal. Entre as propostas do programa está a redução da evasão no ensino público superior, que é um desafio a ser superado na educação brasileira.

De acordo com Bardagi e Hutz (2009), as pesquisas costumam indicar uma taxa de evasão, em instituições públicas ou privadas, em torno de 20% a 50%. As maiores incidências ocorrem no primeiro ano de curso, onde os três primeiros meses de estudo tendem a apresentar o maior número de alunos evadidos.

Descrita como o abandono do curso pelo aluno, sem a conclusão de seus estudos, a evasão “representa uma perda social, de recursos e de tempo de todos os envolvidos no processo de ensino, pois perdeu aluno, seus professores, a instituição de ensino, o sistema de educação e toda a sociedade” (LOBO, 2012, p. 1).

#### **2.4.1 O programa de assistência estudantil da UFES**

Através de sua Pró-Reitoria de Assuntos Estudantis e Cidadania (PROAECI), a Universidade desenvolve o Programa de Assistência Estudantil da UFES (PROAES-UFES), com recursos provenientes do PNAES. Entre seus objetivos, o programa visa:

- I. contribuir para o acesso aos direitos essenciais de alimentação, moradia e transporte;

- II. promover ações de caráter psicossocial;
- III. proporcionar condições de acesso e permanência na perspectiva da inclusão social e democratização do ensino; e
- IV. analisar, planejar e promover ações que visem à redução dos índices de evasão e retenção universitária, quando motivadas por fatores socioeconômicos. (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2013c, p. 1)

Desta forma, o programa estabelece suas diretrizes para a implementação de ações que contribuam para a inclusão social e democratização do ensino na universidade, dando enfoque à redução da evasão e da retenção universitária, quando estas são motivadas por fatores socioeconômicos.

De acordo com a Universidade Federal do Espírito Santo (2017a), atualmente o PROAES-UFES oferece aos alunos os seguintes auxílios: auxílio alimentação; auxílio material de consumo; auxílio moradia; auxílio transporte; acesso ao estudo de língua estrangeira; empréstimo estendido de livros; auxílio educação infantil; auxílio ao estudante em mobilidade internacional; e auxílio cidadania cultural. O auxílio alimentação traduz-se em desconto oferecido no valor do tíquete do Restaurante Universitário (RU). O desconto pode ser de 50% ou 100%, de acordo com critérios estabelecidos em instrução normativa, e pode ser estendido aos filhos de estudantes cadastrados no programa, com idade até seis anos e onze meses. O auxílio material de consumo oferece recursos financeiros para ajudar a custear parte dos gastos do aluno com material didático. Conforme critérios estabelecidos em instrução normativa, o auxílio moradia oferecido pelo programa destina-se a contribuir com parte das despesas de moradia dos alunos cadastrados que migraram de sua cidade de origem com objetivo de acesso aos *campi* e adjacências da UFES. O auxílio transporte contribui com parte dos custos de locomoção do aluno até a universidade. O acesso ao estudo de língua estrangeira oferece bolsas de estudo em língua estrangeira aos estudantes cadastrados, distribuídas conforme condições delimitadas em edital próprio. O empréstimo estendido de livros amplia o tempo de permanência com livro de bibliotecas da UFES, em conformidade com resolução vigente. O auxílio educação infantil consiste em ajudar a custear parte das despesas com creche e pré-escola de estudantes que possuam filho ou menor de idade sob sua guarda ou tutela, com idade até 5 anos, 11 meses e 29 dias, recebendo prioridade aqueles com idade inferior a 2 anos, 11 meses e 29 dias. O auxílio ao estudante em mobilidade internacional é concedido a alunos que ingressaram na universidade através do Convênio PEC-G que não recebem auxílio

financeiro pela Bolsa Promissas. Por último, o auxílio cidadania cultural oferece auxílio financeiro ao estudante cadastrado no PROAES-UFES que desenvolva projetos culturais e atividades voltadas a ações afirmativas nos *campi* da UFES.

## 2.5 MINERAÇÃO DE DADOS E SUAS APLICAÇÕES EDUCACIONAIS

A Mineração de Dados, *Data Mining* em inglês, teve seu início nos anos 80, quando os profissionais de empresas e organizações começaram a se concentrar no aumento do volume de informações não utilizadas que estocavam dentro das empresas. Naquele período, esse ramo da Computação ocupava-se em extrair informações de grandes bancos de dados da maneira mais automatizada possível. Segundo Amo (2004), atualmente a Mineração de Dados consiste basicamente na análise dos dados após serem extraídos e busca, por exemplo, levantar necessidades reais e hipotéticas de clientes de uma empresa para realizar campanhas de *marketing*. Desta forma, a autora destaca que:

[...] uma empresa de cartões de crédito, por exemplo, tem uma mina de ouro de informações: ela sabe os hábitos de compra de cada um dos seus seis milhões de clientes. O que costuma consumir, qual o seu padrão de gastos, grau de endividamento, etc. Para a empresa essas informações são extremamente úteis no estabelecimento do limite de crédito para cada cliente, e além disso, contém dados comportamentais de compra de altíssimo valor. (AMO, 2004, p. 2)

Nesse contexto, a autora exemplifica como a Mineração de Dados pode ajudar na extração de conhecimento a partir do conteúdo registrado em bancos de dados mantidos pelas empresas.

Existem também outros aspectos em que a Mineração de Dados tem se mostrado necessária a uma boa gestão empresarial, tais como: a importância dos volumes de dados para serem tratados apenas com técnicas clássicas de análise; o usuário final não precisa ser um especialista em estatística; a possibilidade de acesso aos dados aumenta com a intensificação de seu tráfego, como navegação na internet, catálogos *online* etc (AMO, 2004).

A Mineração de Dados é definida como um processo que permite a descoberta de padrões em um conjunto de dados. Esse procedimento deve ser automático ou semiautomático e os padrões descobertos devem ser significativos, levando a

alguma vantagem, geralmente de proveito econômico. Trata-se de um tópico que envolve o desenvolvimento de habilidades práticas que possibilitem encontrar e descrever padrões estruturais em dados para explicá-los e fazer previsões a partir deles (FRANK; WITTEN, 2005).

Sobre as tecnologias de mineração de dados, Côtres, Porcaro e Lifschitz (2002) afirmam que elas podem ser aplicadas a uma grande variedade de contextos em que a tomada de decisões no ramo de negócios se faz presente, tais como: *marketing*, finanças, manufatura, saúde, área biomédica etc. Nas finanças, por exemplo, suas aplicações incluem a análise para concessão de crédito a clientes, a segmentação de contas a receber, análise de investimentos como ações, bônus e fundos mútuos, para avaliar opções financeiras e ajudar na detecção de fraudes.

Na Mineração de Dados, a escolha das tarefas e algoritmos utilizados é feita a partir dos objetivos do estudo, visando obter respostas para o problema em questão. Essa abordagem é feita por Galvão e Marin (2009), que afirmam ser possível agrupar todas as tarefas possíveis de um algoritmo de extração de padrões em atividades preditivas e descritivas. Entre as principais tarefas de mineração de dados, apontam:

- a) classificação, que consiste em encontrar uma função que mapeie um conjunto de registros em um conjunto de variáveis, chamadas de classes;
- b) regressão, que funciona de modo similar à classificação, porém as variáveis a serem preditas possuem valores numéricos e busca-se funções lineares ou não lineares entre os conjuntos de variáveis;
- c) clusterização, que é usada para separar registros de um banco de dados em subconjuntos ou *clusters*, também chamados de agrupamentos, onde busca-se encontrar maior similaridade dentro de cada grupo do que entre eles;
- d) associação, que busca identificar e descrever associações entre variáveis no mesmo item ou entre itens diferentes;
- e) sumarização, usada para identificar e indicar características comuns em um conjunto de dados, é normalmente empregada nos agrupamentos obtidos pela clusterização.

Amaral (2016) ressalta que a classificação é a tarefa mais comum empregada no processo de aprendizado de máquina. Nesta tarefa existe um atributo especial denominado classe, e o objetivo da classificação é utilizar todos os outros atributos que compõem a relação para predizer a classe.

Amo (2004) descreve a tarefa de classificação como um processo que visa encontrar modelos ou funções baseados na análise prévia de um conjunto de dados que contenha objetos corretamente classificados, permitindo descrever ou distinguir classes ou conceitos. Deste modo, o modelo encontrado poderá ser utilizado para descobrir a classe de objetos que ainda não foram classificados. Ilustrando esse ponto, menciona que um gerente de supermercado interessado em saber quais tipos de características de seus clientes os classificam como “bom comprador” ou “mau comprador” pode utilizar um modelo de classificação que inclua a seguinte regra: “Clientes da faixa econômica B, com idade entre 50 e 60 são maus compradores” (AMO, 2004, p. 4).

### **2.5.1 Tipos de dados utilizados**

Segundo Camilo e Silva (2009), os dados empregados na Mineração de Dados podem ser caracterizados em dois tipos: quantitativos e qualitativos. Os quantitativos são representados pelos valores numéricos; os qualitativos, também chamados de categóricos, podem ser ordinais e nominais. Giolo (2017) aponta que os ordinais apresentam categorias ordenadas, por exemplo: o grau de pureza da água (baixo, médio ou alto); e as que não exibem ordenação são nominais, por exemplo: a preferência de local para passar as férias (praia, montanha ou fazenda).

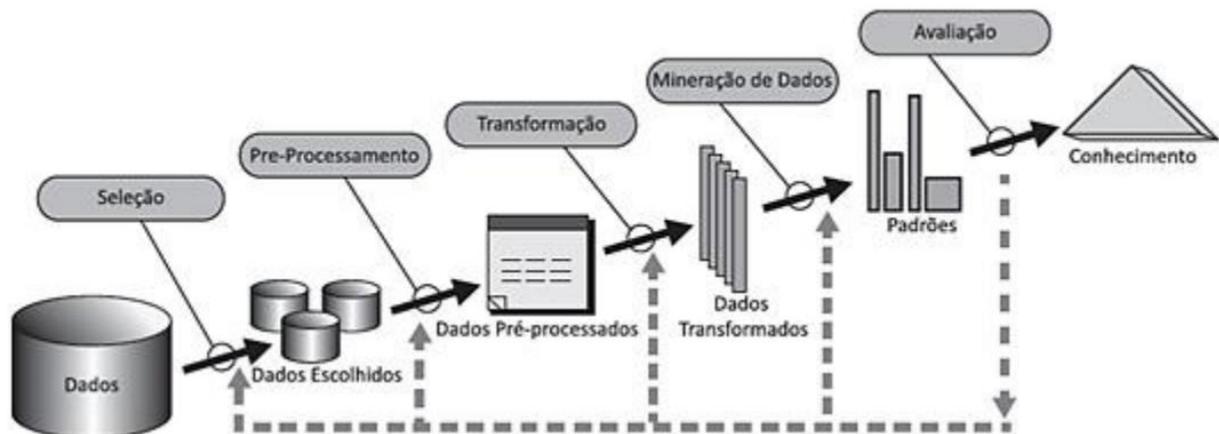
Para Amaral (2016), os tipos de dados utilizados na Mineração de Dados dividem-se em dois grupos: contínuos e nominais. Os contínuos incluem valores numéricos como os do conjunto dos reais. Os nominais podem incluir uma descrição, um nome ou uma categoria e, por isso, podem ser chamados de dados categóricos. Este autor também define como dados discretos aqueles que são finitos, normalmente valores inteiros.

Os algoritmos de classificação conseguem resultados mais efetivos quando os dados utilizados são categorizados, ou seja, segmentados em categorias (ALVES, CECHINEL; QUEIROGA, 2018).

### 2.5.2 O processo de descoberta de conhecimento com base de dados

De acordo com Fayyad, Piatetsky-Shapiro e Smith (1996), o processo de descoberta de conhecimento em base de dados é composto por cinco etapas: seleção, pré-processamento, transformação, mineração de dados e interpretação/avaliação (Figura 1). Segundo os autores, a mineração de dados é a etapa que mais tem recebido contribuições da literatura, mas todas elas possuem importância para o sucesso da descoberta de conhecimento em base de dados na prática.

**Figura 1 – Visão geral das etapas do processo de descoberta de conhecimento em base de dados**



Fonte: Castillo (2017).

Castanheira (2008) descreve as cinco etapas do processo de descoberta de conhecimento em base de dados da seguinte maneira: seleção, como o agrupamento organizado dos dados que se pretende conhecer; o pré-processamento refere-se à limpeza dos dados para adequá-los aos algoritmos em que serão empregados; a transformação consiste em modificar os dados para armazená-los adequadamente em arquivos que serão lidos pelos algoritmos; após esta fase, inicia-se especificamente o processo de mineração de dados, onde será feita a escolha dos algoritmos a serem utilizados; após isso, será gerado um arquivo de descobertas na forma de relatório ou gráfico, que serão interpretados e permitirão as conclusões que fornecem o conhecimento da base de dados estudada.

### 2.5.3 Regra *Naive* de Bayes

O algoritmo de classificação *Naive Bayes* é uma técnica estatística baseada no Teorema de Thomas Bayes, que afirma ser possível obter a probabilidade de certo evento ocorrer, dada a probabilidade de outro evento que já ocorreu (CAMILO; SILVA, 2009). Esse processo é utilizado quando se desconhece as relações de dependência entre os dados de entrada usados por um classificador (BRAGA; LACERDA, 2004).

Segundo Pardo e Nunes (2002), o classificador *Naive Bayes* trabalha com a suposição “ingênua” de que todos os atributos ( $a_1...a_n$ ) da instância que se pretende classificar são independentes. O classificador produz resultados satisfatórios mesmo com essa suposição de independência sendo falsa na maioria dos casos. A fórmula final utilizada pelo algoritmo é apresentada pelos autores conforme se segue:

$$\arg \max P(\text{classe}|a_1 \dots a_n) = \arg \max \prod_i P(a_i|\text{classe}) \times P(\text{classe}) \quad (1)$$

Em que:

*classe* – informação que se pretende prever;

$a_1$  – valor conhecido do primeiro atributo;

$a_n$  – valor conhecido do último atributo;

$n$  – número total de atributos.

Na fórmula acima,  $P(\text{classe}|a_1...a_n)$  representa a probabilidade de ocorrer um valor específico, representado pela palavra “classe”, para a informação que se pretende prever em um evento onde já se conhecem os valores de seus demais atributos, representados por “ $a_1...a_n$ ”.  $P(a_i|\text{classe})$  e  $P(\text{classe})$  aplicam-se sobre o conjunto de dados a serem estudados. O primeiro equivale à proporção encontrada através do número de casos com valor  $a_i$  sobre o total de casos da classe em questão; o segundo é a proporção obtida com a quantidade de casos pertencente à mesma classe sobre o número total de casos. A fórmula é aplicada sobre todas as possíveis classes que se pretende prever, devendo ser escolhida aquela que apresente maior probabilidade (PARDO; NUNES, 2002).

Segundo Andrade *et al.* (2016), o algoritmo de classificação *Naive Bayes* possui ampla utilização em diversos trabalhos na área de MDE, voltada para a identificação precoce de estudantes propensos à evasão.

### 2.5.4 Ganho de informação de atributos

De acordo com Almeida *et al.* (2003), é possível realizar o cálculo completo do ganho de informação obtido através de um determinado atributo utilizando os métodos *Gain*, *Split* e *Gain Ratio*. Segundo os autores, esses métodos derivam da entropia, grandeza utilizada para medir a desordem tanto de objetos físicos quanto de informações, e são expressos pelas fórmulas 2, 3, 4 e 5 a seguir.

$$E(A = v_j) = -\sum_{i=1}^n p(i) \times \log_2(p(i)) \quad (2)$$

Em que:

$E$  – valor de entropia do conjunto de informações;  
 $A$  – nome do atributo;  
 $v_j$  – valor do atributo;  
 $n$  – número de classes diferentes ( $c_1, c_2, \dots, c_n$ );  
 $i$  – número da classe;  
 $p(i)$  – probabilidade de pertencer à classe  $c_n$ .

Onde  $E$  representa o valor da entropia na qual o atributo  $A$  possui valor  $v_j$ ;  $n$  indica o número de classes diferentes ( $c_1, c_2, \dots, c_n$ ); e  $p(i)$  informa a probabilidade de um registro pertencer à classe  $c_n$ . Após calcular todos os valores de entropia possíveis para um determinado atributo, utiliza-se a fórmula seguinte para relacioná-los e obter o ganho de informação do atributo.

$$Gain(D, T) = Info(D) - \sum_{i=1}^k \frac{|D_i|}{|D|} \times Info(D_i) \quad (3)$$

Em que:

$D$  – número total de exemplos da base;  
 $T$  – nome do atributo;  
 $k$  – número total de subconjuntos do atributo  $T$ ;  
 $i$  – número do subconjunto ( $D_1, D_2, D_3, \dots, D_n$ );  
 $D_i$  – quantidade de exemplos do subconjunto;  
 $Info(D_i)$  – entropia do conjunto  $D_i$ ;  
 $Info(D)$  – entropia de toda a base de dados.

Na fórmula acima,  $Info(D)$  constitui o valor de entropia da base como um todo;  $k$  refere-se à quantidade de subconjuntos do atributo  $T$ ;  $Info(D_i)$  reproduz a entropia de cada subconjunto ( $D_1, D_2, D_3, \dots, D_n$ );  $D_i$  representa a quantidade de exemplos de cada subconjunto e  $D$  a quantidade de exemplos de toda a base.

Os autores supracitados ressaltam que o método *Gain*, apesar de apresentar bons resultados, favorece atributos com muitos valores. Como forma de contornar esse problema, apontam o cálculo do método *Split Info*.

$$Split(D, T) = - \sum_{i=1}^k \frac{|D_i|}{D} \times \log_2 \left( \frac{|D_i|}{D} \right) \quad (4)$$

Em que:

$D$  – número total de exemplos da base;

$T$  – nome do atributo;

$k$  – número total de subconjuntos do atributo  $T$ ;

$i$  – número do subconjunto ( $D_1, D_2, D_3, \dots, D_n$ );

$D_i$  – quantidade de exemplos do subconjunto.

No cálculo do *Split*,  $k$  traz a quantidade de subconjuntos do atributo  $T$ ;  $D_i$  reproduz a quantidade de exemplos de cada subconjunto do atributo  $T$ ; e  $D$  refere-se à quantidade de exemplos de toda a base. Com isso, obtêm-se os elementos necessários para o cálculo completo do ganho de informação através da fórmula:

$$GainRatio(D, T) = \frac{Gain(D, T)}{Split(D, T)} \quad (5)$$

Em que:

$D$  – número total de exemplos da base;

$T$  – nome do atributo.

Aqui,  $D$  refere-se também à quantidade de exemplos de toda a base e  $T$  é o atributo que se pretende calcular o ganho de informação.

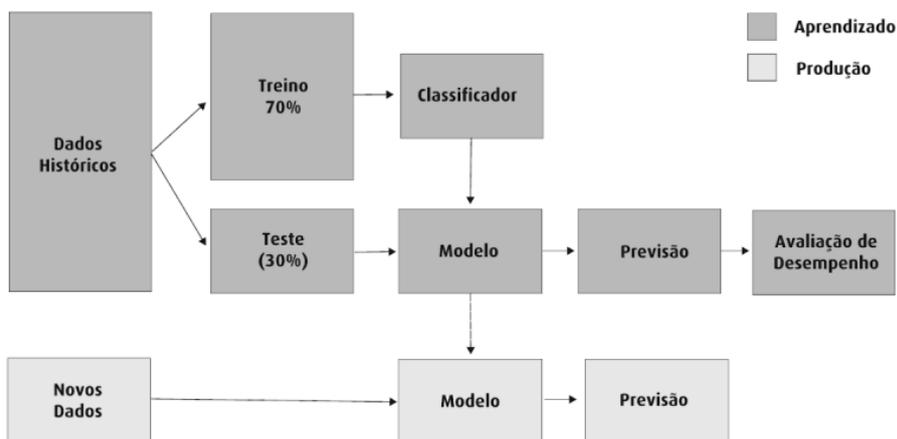
### 2.5.5 Avaliação de modelos de aprendizado de máquina

Quando se dispõe de dados históricos, aqueles cuja classificação já se conhece, não basta aplicá-los a um algoritmo e utilizar o modelo gerado. É preciso medir o quanto o modelo construído é capaz de prever com eficiência os dados que ainda não conheceu. Uma forma confiável de avaliar um modelo é aplicá-lo a um conjunto de dados que ele ainda não conheça. Pode-se conseguir isso dividindo parte dos dados históricos para gerar o modelo e outra parte para testá-lo (AMARAL, 2016).

### 2.5.5.1 Método *Hold Out*

O método *Hold Out* é uma das maneiras mais comuns de separar os dados para construir o modelo e para testá-lo. Normalmente, submetem-se 70% dos dados para que o algoritmo de classificação produza o modelo, que em seguida é aplicado aos 30% restantes dos dados para que seja feita a previsão (Figura 2). Por se tratarem de dados históricos já classificados, é possível comparar a classe real com a classe prevista (AMARAL, 2016).

**Figura 2 – Modelo esquemático do método *Hold Out***



Fonte: Amaral (2016).

Na ilustração, a avaliação de desempenho é realizada na fase final do aprendizado, após a previsão feita sobre os dados históricos. Uma vez medido o desempenho e considerado satisfatório, pode-se então utilizar o modelo na fase de produção de previsão em novos dados (AMARAL, 2016).

### 2.5.5.2 Validação cruzada

A validação cruzada trata-se, tipicamente, de um processo de aprendizagem supervisionada na mineração de dados em que, após serem pré-processados e formatados, os dados são divididos em dois subconjuntos: base de treinamento e base de testes. Inicialmente, é aplicado um algoritmo de indução de conhecimento à base de treinamento, extraíndo o conhecimento que, numa segunda etapa, será aplicado ao fragmento do conjunto de dados chamado base de testes. Esta base, que também possui rotulação prévia, permitirá medir a taxa de acerto do

conhecimento extraído na base de treinamento ao comparar o resultado obtido com o rótulo encontrado na base de testes (KAESTNER *et al.*, 2009).

As instâncias dos dados de treinamento são divididas aleatoriamente em  $k$  partições mutuamente exclusivas, onde a quantidade de partições geradas determina a quantidade de treinamentos e testes realizados, de modo que cada partição é usada  $k-1$  vezes para treinamento e uma vez para teste (Figura 3). Ao final, o desempenho é obtido a partir da média aritmética das avaliações (SOUSA, 2017).

**Figura 3 – Modelo esquemático de validação cruzada**

Interação 1	Conjunto de Treinamento	Conjunto de Treinamento	Conjunto de Treinamento	Conjunto de Teste
Interação 2	Conjunto de Treinamento	Conjunto de Treinamento	Conjunto de Teste	Conjunto de Treinamento
Interação 3	Conjunto de Treinamento	Conjunto de Teste	Conjunto de Treinamento	Conjunto de Treinamento
Interação 4	Conjunto de Teste	Conjunto de Treinamento	Conjunto de Treinamento	Conjunto de Treinamento

Fonte: Sousa (2017).

Na Figura 3, as quatro interações geraram doze conjuntos como base de treinamento e quatro como base de testes.

### 2.5.5.3 Matriz de confusão

Amplamente utilizada em análise de estatística de concordância, a matriz de erros ou matriz de confusão (Quadro 1) é obtida após a aplicação do modelo extraído no conjunto de treinamento sobre o conjunto de testes selecionado (NONATO; OLIVEIRA, 2013).

**Quadro 1 – Modelo esquemático de uma matriz de confusão**

		Predito	
		Classe A	Classe B
Verdadeiro	Classe A	VP	FN
	Classe B	FP	VN

Fonte: Nonato e Oliveira (2013).

No quadro acima, “VP” representa a quantidade de valores denominados como verdadeiros positivos, ou seja, aqueles em que o algoritmo de classificação predisse como Classe A, considerada como positiva, e coincidiu com a classificação verdadeira preexistente no conjunto de testes. De forma semelhante, “VN” corresponde à quantidade de acertos para a Classe B, considerada como negativa. As predições falsas estão representadas por “FP” e “FN”.

A partir da matriz de confusão, é possível calcular as taxas de precisão, revocação, acurácia e especificidade da classificação realizada pelo algoritmo. Segundo Ciferri *et al.* (2009), a precisão é a taxa com que todos os valores apresentados como positivos são realmente positivos, não incluindo os valores negativos; a taxa de revocação é aquela que indica o quanto do total de informação relevante foi extraída; a acurácia é a mais utilizada para avaliação de problemas de classificação de aprendizado de máquina; e especificidade reproduz a taxa na qual uma instância negativa foi classificada como negativa. As fórmulas para cálculo das referidas taxas são apresentadas pelos autores conforme mostrado pelas fórmulas 6, 7, 8 e 9 a seguir:

$$Precisão = \frac{VP}{VP+FP} \quad (6)$$

$$Revocação = \frac{VP}{VP+FN} \quad (7)$$

$$Acurácia = \frac{VP+VN}{VP+VN+FP+FN} \quad (8)$$

$$Especificidade = \frac{VN}{VN+FP} \quad (9)$$

Em que:

VP – verdadeiro positivo;  
 VN – verdadeiro negativo;  
 FP – falso positivo;  
 FN – falso negativo.

#### 2.5.5.4 Coeficiente Kappa

O coeficiente Kappa é usado para medir a concordância entre classificações realizadas por dois anotadores/juízes, podendo ser eles humanos ou não. Essa definição é apresentada por Ciferri *et al.* (2009), que citam os algoritmos de

classificação como exemplos de anotadores não humanos. Segundo os autores, o coeficiente define um limiar de concordância entre os anotadores em tarefas de classificação ou de extração de informação, ajudando a avaliar o desempenho do sistema, contribuindo para avaliar se o conjunto de treinamento é um material válido e ajudando a descartar exemplos que não sejam fáceis para os anotadores concordarem.

A fórmula para cálculo do coeficiente *Kappa*, expressa pela letra *k*, foi apresentada por Cohen (1960), conforme equação 10, e exprime a proporção de acordos encontrados entre juízes após desconsiderar seus acordos esperados pelo acaso.

$$k = \frac{p_o - p_c}{1 - p_c} \quad (10)$$

Em que:

$p_o$  – proporção de concordância;

$p_c$  – proporção de concordância ao acaso.

Na fórmula,  $p_o$  representa a proporção em que os juízes concordam; e  $p_c$ , a proporção esperada em que eles concordariam por mero acaso. A proporção  $p_o$  é obtida dividindo-se o número de casos em que os juízes concordam pelo total de casos existentes. Já a proporção  $p_c$  obtém-se realizando o somatório dos valores alcançados após multiplicar as proporções encontradas por ambos juízes em cada valor da classe.

De acordo com Fonseca, Silva e Silva (2007), embora não exista um valor objetivo específico do índice *Kappa* a partir do qual se deva considerá-lo adequado, na literatura é possível encontrar algumas sugestões que normalmente orientam essa decisão.

Koch e Landis (1977) atribuíram rótulos a faixas do índice *Kappa* para descrever a força relativa da concordância associada (Quadro 2).

**Quadro 2 – Rótulos qualitativos do índice Kappa**

<b>Estatística Kappa</b>	<b>Força da Concordância</b>
< 0,00	Pobre
0,00 a 0,20	Leve
0,21 a 0,40	Justa
0,41 a 0,60	Moderada
0,61 a 0,80	Substancial
0,81 a 1,00	Quase Perfeito

Fonte: Adaptado de Koch e Landis (1977).

Através dos rótulos, é possível obter uma alternativa de orientação para descrever a força da concordância encontrada em classificações feitas por dois anotadores/juízes.

### **2.5.6 Mineração de dados educacionais**

Cambruzzi, Cazzela e Rigo (2012) destacam que a difusão do uso de sistemas informatizados em escolas e universidades produz a cada dia um crescente volume de informações geradas e armazenadas em banco de dados. Isso tem proporcionado inúmeras oportunidades para a MDE, que trata da aplicação de técnicas de mineração junto ao novo conjunto de dados produzidos nos diversos contextos educacionais. Nesse entendimento, observam que:

Esta é uma área de pesquisa em expansão, tendo como principais enfoques os trabalhos relacionados com predição, agrupamento, mineração de relações, descoberta com modelos e tratamento de dados para apoio à decisão. Em todos estes casos, em maior ou menor grau, podem ser vislumbradas aplicações ligadas às questões discutidas no âmbito dos modelos teóricos sobre evasão escolar e ligadas aos processos de aquisições de modelos gerais e de antecipação de diagnósticos. (CAMBRUZZI; CAZELLA; RIGO, 2012, p. 173)

Nessa perspectiva, os referidos autores ressaltam a importância da utilização de recursos de mineração de dados, uma vez que possibilitam a aquisição de modelos que permitem realizar diagnósticos voltados à prevenção de fatores ligados à evasão escolar.

### 2.5.6.1 Ferramentas utilizadas

Baker *et al.* (2013) apontam algumas das principais ferramentas empregadas na área de descoberta de conhecimento no contexto educacional: DBMiner, Clementine, IBM Intelligent Miner, WEKA e RapidMiner. As duas últimas sendo muito utilizadas na literatura segundo os autores.

*Waikato Environment for Knowledge Analysis* (WEKA) é um programa de código aberto desenvolvido na Universidade de Waikato na Nova Zelândia. Entre suas funcionalidades, possui uma variedade de algoritmos de aprendizagem, incluindo ferramentas de pré-processamento. Oferece suporte a todo processo de mineração, como: preparação dos dados de entrada, avaliação estatística de aprendizagem, visualização dos dados de entrada e dos resultados (BAKER *et al.*, 2013).

O *RapidMiner* é um software para a realização de análises preditivas e aprendizagem de máquina. A execução de comandos sem a necessidade de desenvolvimento de código, apenas utilizando a interface gráfica, é uma de suas principais características. Através da inclusão de elementos gráficos, que representam uma operação em questão, e da união de seus elementos, o usuário constrói um fluxo de execução intuitivo (KOMATI; REZENDE; RESENDO, 2017).

### 2.5.6.2 Trabalhos relacionados

Nesta seção, são apresentados alguns trabalhos com diferentes metodologias, que tiveram o propósito de ampliar o conhecimento sobre evasão estudantil através da utilização de softwares de mineração de dados.

No trabalho de Andrade *et al.* (2016), buscou-se alcançar o diagnóstico precoce de estudantes propensos à evasão estudantil. Os autores utilizaram registros socioeconômicos e acadêmicos de 241 estudantes concluintes e evadidos do curso de Ciência da Computação da Universidade Federal da Paraíba (UFPB) para avaliar atributos de classificação e a capacidade preditiva dos métodos empregados pelo software de mineração de dados WEKA. Na ocasião, constatou-se que as informações acadêmicas apresentaram maior impacto que as socioeconômicas na situação final dos alunos pesquisados. Sobre a predição de evasão dos estudantes,

o método *Naive Bayes* utilizado pelo programa apresentou uma acurácia de 85,48%, considerada viável e sensata pelos autores.

Num estudo de Ferreira *et al.* (2014), foi proposto um sistema de alertas em Ambiente Virtual de Aprendizagem (AVA), empregado no ensino a distância, a partir de dados gerados por meio de processo de mineração de dados com a utilização do programa *RapidMiner*. Os alertas ofereciam suporte à atuação dos professores no acompanhamento do processo de aprendizagem dos alunos, notificando-os sobre grupos de alunos que compartilham de necessidades específicas. Com o resultado dos experimentos realizados com dados de 1780 alunos, os autores concluíram que o sistema de alertas proposto pode contribuir para o aumento nos índices de aprovação e com a redução dos índices de evasão na educação a distância.

Na Universidade de Santa Cruz do Sul (UNISC), outro trabalho utilizando a ferramenta de mineração de dados WEKA foi conduzido por Cornelius Junior (2015). O autor utilizou técnicas de mineração de dados buscando encontrar razões para a evasão no ensino superior, visando contribuir com a gestão universitária no planejamento de ações efetivas para reduzir a saída antecipada de alunos. O trabalho permitiu análises detalhadas do padrão de estudantes com perfil de evasão.

### 3 MÉTODOS E PROCEDIMENTOS

A pesquisa proposta neste trabalho realizou uma abordagem quantitativa, uma vez que se baseou “em dados mensuráveis das variáveis, procurando verificar e explicar sua existência, relação ou influência sobre outra variável” (FONSECA, 2009, p. 35). Foram utilizados dados secundários que tiveram a incidência de suas variáveis medidas a fim de possibilitar uma análise preditiva de eventos futuros. Quanto a seus fins, tratou-se de uma pesquisa descritiva e metodológica.

A população alvo, ou objeto da pesquisa, foi composta por estudantes de graduação do CCAE e CCENS que realizaram matrícula no período compreendido entre os anos de 2007 a 2018 e que passaram por processo de evasão nos anos de 2016 a 2018.

O pré-processamento e a transformação dos dados foram realizados com o auxílio dos programas Microsoft Excel 2010 (MICROSOFT CORPORATION, 2010) e *RapidMiner Studio*<sup>1</sup> (RAPIDMINER GMBH, 2019). Este último, em sua versão 9.2.001, plataforma WIN64, foi também utilizado para realizar a mineração dos dados.

#### 3.1 COLETA DOS DADOS

Como fonte de dados, o trabalho utilizou bancos de dados que foram solicitados à Secretaria Única de Graduação – Setorial Sul, unidade subordinada ao Departamento de Suporte à Gestão, localizada no município de Alegre.

Os bancos de dados fornecidos estavam dispostos em duas tabelas contendo dados digitais em formato xls extraídos do programa informatizado Sistema de Informações para o Ensino (SIE). Uma delas com informações sociais e acadêmicas de 1340 alunos dos cursos de graduação que haviam passado por processo de evasão nos referidos centros nos anos de 2016 a 2018, e outra contendo matrículas de 6.827 alunos no período de 2007 a 2018.

---

<sup>1</sup> Programa obtido no sítio <http://rapidminer.com>.

### 3.2 TRATAMENTO DOS DADOS

Os dados fornecidos foram reunidos em duas tabelas com finalidades distintas. A primeira reuniu atributos de ambas as tabelas fornecidas com informações relacionadas a alunos que passaram por processo de evasão no período de 2016 a 2018 e foi utilizada pelo algoritmo classificador para a construção do modelo de predição. A segunda, contendo as mesmas colunas de atributos da primeira, agrupou informações de todos os alunos encontrados que ainda não passaram por processo de evasão e foi utilizada pelo modelo de predição para classificá-los de acordo com sua probabilidade de evadir voluntariamente.

O anonimato dos alunos pesquisados foi preservado. Com isso, foram descartadas as colunas de atributos referentes a valores que continham suas informações pessoais ou que possibilitassem contatá-los.

O número de matrícula dos estudantes foi utilizado como elemento de ligação entre as tabelas de dados fornecidas, possibilitando a junção de colunas com atributos de alunos comuns entre elas. Nas ocasiões em que a visualização de tais números se fez necessária para auxiliar a demonstração dos resultados, seus dois últimos dígitos foram ocultados para impossibilitar a identificação do aluno.

Durante a preparação dos dados, foram também descartados os atributos:

- a) cujos valores não eram aplicáveis a futuros alunos;
- b) que reproduziam informações existentes em outra coluna de atributos.

Também foram descartados os registros de alunos cujas formas de evasão apresentaram os valores “Erro de cadastro” e “Falecimento”.

Durante o pré-processamento dos dados, os valores de atributo relacionados à forma de evasão foram reduzidos em duas categorias. Com isso, as formas de evasão “desistência” e “desligamento por abandono” foram substituídas pelo valor “Evasão voluntária”; e as demais formas encontradas foram denominadas como “Outras evasões”. Também foram discretizados os valores de atributos do tipo “real”, sendo convertidos para “inteiro”.

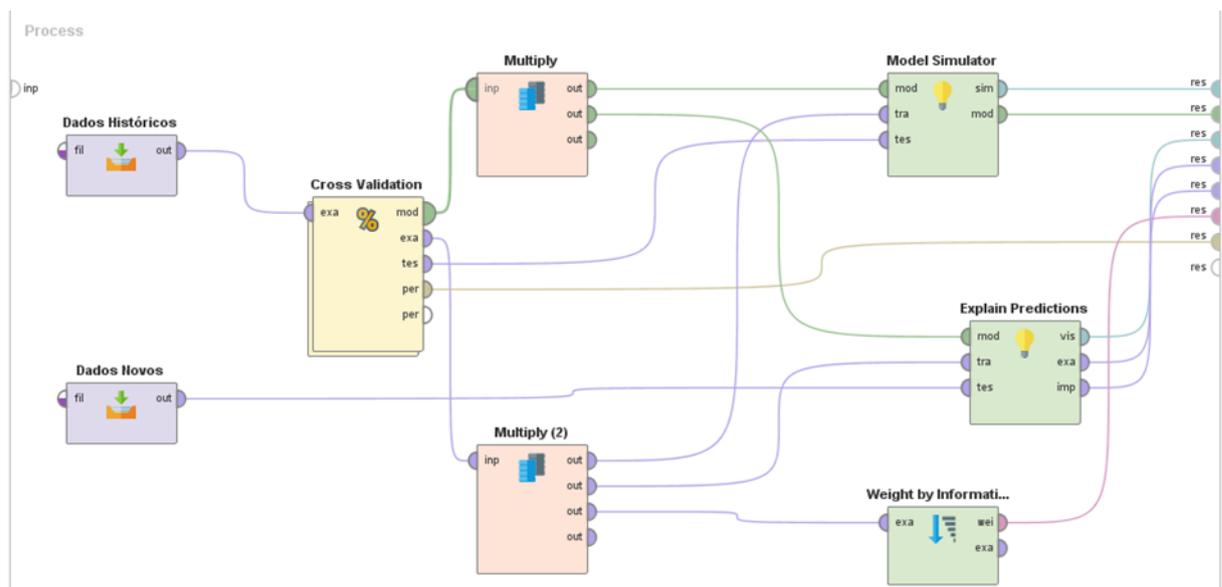
Após passarem pelos processos de tratamento descritos nesta seção, ambas as tabelas tiveram os valores de todos os seus atributos convertidos para o tipo “nominal”.

### 3.3 MINERAÇÃO DOS DADOS

Utilizou-se o programa *RapidMiner Studio* para encontrar os fatores de maior importância para a evasão voluntária, bem como para a construção do modelo de predição empregado, através do algoritmo de classificação *Naive Bayes*, que faz parte do conjunto de operadores encontrados neste programa. Este algoritmo foi escolhido devido a sua ampla utilização em trabalhos relacionados à identificação prévia de estudantes propensos à evasão (ANDRADE *et al.*, 2016).

A tabela contendo dados dos alunos evadidos foi utilizada pelo algoritmo de classificação para a construção do modelo de predição, que foi aplicado na tabela com dados dos alunos que ainda não passaram por processo de evasão para obter-se a relação de estudantes com risco potencial de evasão. O processo utilizado no programa está disposto na Figura 4.

**Figura 4 – Processo utilizado na mineração dos dados obtidos**

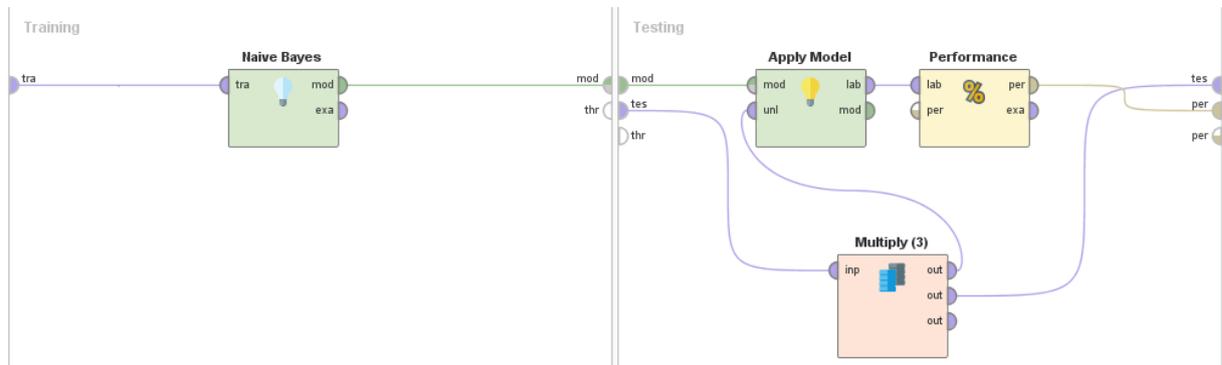


Fonte: Dados da pesquisa.

Nota: Imagem extraída do programa *RapidMiner Studio*.

No processo mostrado, “Dados Históricos” corresponde ao operador “*Read Excel*” vinculado aos dados da tabela de alunos evadidos; e “Dados Novos” refere-se ao mesmo tipo de operador, porém vinculado aos dados da tabela dos alunos que ainda estavam cursando nos dois centros pesquisados. O operador “*Cross Validation*”, aplicado para a realização da validação cruzada, abriga outros operadores que foram utilizados para a criação e para o teste de desempenho do modelo de predição utilizado na classificação (Figura 5).

**Figura 5 – Operadores de classificação e de desempenho utilizados**



Fonte: Dados da pesquisa.

Nota: Imagem extraída do programa *RapidMiner Studio*.

Em todo o processo, os operadores utilizados mantiveram sua configuração padrão; exceto o operador “*Performance (Binominal Classification)*”, que teve o parâmetro “*kappa*” marcado para que seu coeficiente fosse exibido entre os resultados.

A acurácia e a força de concordância do modelo foram verificadas através de matriz de confusão obtida pelo método de Validação Cruzada e através do coeficiente *Kappa* encontrado.

## 4 RESULTADOS E DISCUSSÃO

Esta seção apresenta a estrutura de metadados das bases obtidas nos dois centros acadêmicos da UFES em Alegre, contendo informações de seus estudantes de graduação presencial, bem como os resultados do tratamento e mineração desses dados com enfoque na identificação de características do aluno que evade voluntariamente e na identificação prévia de alunos com potencial risco de evasão.

### 4.1 CARACTERÍSTICAS DOS DADOS COLETADOS

Os bancos de dados foram fornecidos pela Secretaria Única de Graduação – Setorial Sul, em 11 de abril de 2019, através de duas planilhas em formato xls extraídas do programa informatizado Sistema de Informações para o Ensino (SIE). A primeira com registros de 1340 alunos evadidos nos anos de 2016 a 2018, e a segunda com 6.827 registros de alunos ingressantes de 2007 até o segundo semestre de 2018. O Quadro 3 apresenta os atributos contidos em cada planilha.

**Quadro 3 – Relação de atributos das tabelas obtidas na Secretaria Única de Graduação – Setorial Sul**

<b>Código SIE</b>	<b>Conteúdo</b>	<b>Atributos</b>
11.02.06.03.04	Alunos evadidos de 2016 a 2018	ANO_EVASAO; ANO_INGRESSO; COD_CURSO; COTISTA; CPF; DT_SAIDA; FORMA_EVASAO; MATR_ALUNO; NOME_ALUNO; NOME_CURSO; NUM_VERSAO; PERIODO_EVASAO; PERIODO_INGRESSO; SEXO
11.02.06.03.13	Ingressantes de 2007 a 2018	ANO_EVASAO; ANO_INGRESSO; BAIRRO; CEP; CH_TOTAL_CURSO; COD_CURSO; COMPLEMENTO; COTISTA; CPF; CRA; CRN; DEFICIENCIA; DESCR_MAIL; DT_NASCIMENTO; ESTADO; ESTADO_CIVIL; ETNIA; FORMA_EVASAO; FORMA_INGRESSO; MATR_ALUNO; MUNICIPIO; NATURALIDADE; NOME_ALUNO; NOME_CURSO; NUM_MAX_PERIODOS; NUM_PERIODOS; NUMERO; PAIS; PERIODO_EVASAO; PERIODO_INGRESSO; RG; RG_ESTADO; RG_ORGAO_EMISSOR; RUA; SEXO; TIPO_COTA; TIPO_LOGRADOURO; TURNO; TURNO_ALUNO_ITEM; UF; UF_NATURALIDADE

Fonte: Dados da pesquisa.

Os nomes dos arquivos fornecidos continham a numeração de código das rotinas do programa SIE utilizadas para gerar as tabelas.

Observou-se que a planilha de alunos ingressantes continha todos os atributos disponíveis na planilha de alunos evadidos fornecida, com exceção dos atributos “DT\_SAIDA”, onde é informada a data em que ocorreu a evasão do aluno, e “NUM\_VERSAO”, que informa o ano de versão do projeto pedagógico do curso. O Quadro 4 descreve as características de todos os atributos encontrados.

**Quadro 4 – Características dos atributos encontrados**

(Continua)

<b>Atributo</b>	<b>Descrição</b>	<b>Tipo</b>
ANO_EVASAO	Ano de evasão	Contínuo
ANO_INGRESSO	Ano de ingresso	Contínuo
BAIRRO	Bairro de residência	Nominal
CEP	CEP de residência	Nominal
CH_TOTAL_CURSO	Carga horária total do curso	Contínuo
COD_CURSO	Código do curso	Contínuo
COMPLEMENTO	Complemento de residência	Nominal
COTISTA	Se está incluído em cotas	Nominal
CPF	Número de CPF	Nominal
CRA	Coefficiente de rendimento do aluno	Contínuo
CRN	Coefficiente de rendimento normalizado	Nominal
DEFICIENCIA	Tipo de deficiência	Nominal
DESCR_MAIL	Endereço eletrônico do aluno	Nominal
DT_NASCIMENTO	Data de nascimento	Nominal
DT_SAIDA	Data de evasão	Nominal
ESTADO	Estado de residência	Nominal
ESTADO_CIVIL	Estado civil do aluno	Nominal
ETNIA	Etnia do aluno	Nominal
FORMA_EVASAO	Forma de evasão	Nominal
FORMA_INGRESSO	Forma de ingresso	Nominal
MATR_ALUNO	Número de matrícula do aluno	Nominal
MUNICIPIO	Município de residência	Nominal
NATURALIDADE	Naturalidade	Nominal
NOME_ALUNO	Nome do aluno	Nominal
NOME_CURSO	Nome do curso	Nominal
NUM_MAX_PERIODOS	Número máximo de períodos do curso	Contínuo
NUM_PERIODOS	Número de períodos	Contínuo
NUM_VERSAO	Ano de versão do projeto pedagógico do curso	Contínuo

Quadro 4 – Características dos atributos encontrados

(Conclusão)

Atributo	Descrição	Tipo
NUMERO	Número de residência do aluno	Nominal
PAIS	País de origem	Nominal
PERIODO_EVASAO	Período da evasão	Nominal
PERIODO_INGRESSO	Período de ingresso	Nominal
RG	Número de identidade do aluno	Nominal
RG_ESTADO	Estado emissor da identidade	Nominal
RG_ORGAO_EMISSOR	Órgão emissor da identidade	Nominal
RUA	Nome do logradouro de residência do aluno	Nominal
SEXO	Sexo do aluno	Nominal
TIPO_COTA	Tipo de cota em que está inserido	Nominal
TIPO_LOGRADOURO	Tipo de logradouro	Nominal
TURNO	Turno de estudos	Nominal
TURNO_ALUNO_ITEM	Código do Turno de estudos	Contínuo
UF	Sigla do Estado de residência	Nominal
UF_NATURALIDADE	Sigla do Estado de naturalidade	Nominal

Fonte: Dados da pesquisa.

Os valores do atributo “CRN” estavam formatados como nominal, embora representassem dados contínuos do conjunto dos números reais, tais como os encontrados no atributo “CRA”. Esses atributos estão relacionados ao Coeficiente de Rendimento Normalizado (CRN) e ao Coeficiente de Rendimento do Aluno (CRA).

O CRA reproduz o valor médio do aproveitamento registrado no histórico do aluno em todas as disciplinas em que foi matriculado nos diversos semestres (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2017b).

A Universidade Federal do Espírito Santo (2008), através de sua Resolução nº 58/2008, estabeleceu que o CRN deverá ser utilizado como critério de seleção dos alunos quando a demanda por uma disciplina for superior ao número de vagas oferecidas, e seu cálculo é obtido pela fórmula seguinte:

$$CRN = \frac{5,000 + 1,250(CRA - CRM)}{DESV} \quad (11)$$

Em que:

*CRA* – coeficiente de rendimento do aluno;

*CRM* – coeficiente de rendimento médio do curso;

*DESV* – desvio padrão dos coeficientes de rendimento do curso.

Na fórmula, CRM representa a média dos coeficientes de rendimento dos alunos do curso em que o aluno possui vínculo; e DESV, o desvio padrão dos coeficientes de rendimento dos alunos do curso.

#### 4.2 PREPARAÇÃO DOS DADOS ENCONTRADOS

Após a preparação dos dados e o descarte previsto para alguns atributos, obteve-se uma tabela, cujas características estão dispostas no Quadro 5, contendo dados de 1.337 estudantes que passaram por processo de evasão. Destes, encontrou-se 421 alunos cuja saída antes da conclusão de seus cursos ocorreu de forma voluntária, totalizando 31,49% do grupo de alunos evadidos utilizados na mineração dos dados.

**Quadro 5 – Atributos utilizados na construção do modelo de predição**

(Continua)

<b>Atributo</b>	<b>Tipo</b>	<b>Valores encontrados</b>
CH_TOTAL_CURSO	Nominal	2915; 3020; 3095; 3125; 3200; 3215; 3285; 3440; 3675; 3855; 3930; 4005; 4050; 4190; 4215; 4275; 4370; 4410; 4635
CRA	Nominal	0 a 17
CRN	Nominal	0 a 17
DEFICIENCIA	Nominal	Auditiva/Surdez; Autismo; Física; Sem deficiência; Visual
ESTADO_CIVIL	Nominal	Casado(a); Divorciado(a); Não informado; Outros; Separado(a) Judicialmente; Solteiro(a); Uniao Estavel
ETNIA	Nominal	Amarela; Branca; Não Declarada; Não informado; Parda; Preta
FORMA_EVASAO	Nominal	Evasão voluntária; Outras evasões
FORMA_INGRESSO	Nominal	Novo Curso; Outros; Reopção de Curso; SISU; Transferência Facultativa; Vestibular
MATR_ALUNO	Nominal	20071000XX a 20182072XX
MUNICIPIO	Nominal	134 municípios
NATURALIDADE	Nominal	201 municípios

Quadro 5 – Atributos utilizados na construção do modelo de predição

(Conclusão)

Atributo	Tipo	Valores encontrados
NOME_CURSO	Nominal	Agronomia - Alegre; Ciência da Computação - Bacharelado - Alegre; Ciências Biológicas - Bacharelado - Alegre; Ciências Biológicas - Licenciatura - Alegre; Engenharia de Alimentos - Alegre; Engenharia Florestal - Alegre; Engenharia Industrial Madeireira - Alegre; Engenharia Química - Bacharelado - Alegre; Farmácia - Bacharelado - Alegre; Física - Licenciatura - Alegre; Geologia - Alegre; Matemática - Licenciatura - Alegre; Medicina Veterinária - Alegre; Nutrição - Alegre; Química - Licenciatura - Alegre; Sistemas de Informação - Bacharelado - Alegre; Zootecnia - Alegre
NUM_MAX_PERIODOS	Nominal	13; 14; 15; 16
NUM_PERIODOS	Nominal	8; 9; 10
PAIS	Nominal	Brasil; Outros
PERIODO_INGRESSO	Nominal	1º Semestre; 2º Semestre
SEXO	Nominal	F; M
TIPO_COTA	Nominal	Baixa Renda e Não PPI; Baixa Renda e PPI; Cotista; NÃO COTISTA; Renda Normal e Não PPI; Renda Normal e PPI; Reserva
TURNO	Nominal	Diurno; Integral; Matutino; Não informado; Noturno; Vespertino
UF	Nominal	BA; DF; ES; GO; MG; PE; PR; RJ; SP
UF_NATURALIDADE	Nominal	Alagoas; Amazonas; Bahia; Ceará; Distrito Federal; Espírito Santo; Goiás; Maranhão; Mato Grosso; Mato Grosso do Sul; Minas Gerais; Pará; Parana; Pernambuco; Piauí; Rio de Janeiro; Rio Grande do Norte; Rondônia; São Paulo

Fonte: Dados da pesquisa.

As informações contidas nessa tabela passaram pelo algoritmo de classificação para a construção do modelo de predição.

Outra tabela com os mesmos atributos foi preparada, contendo dados de 2.553 alunos extraídos da lista de ingressantes que ainda não haviam passado por processo de evasão. Nesta tabela aplicou-se o modelo de predição resultante da mineração dos dados para classificar os estudantes quanto à possibilidade de evadirem-se de forma voluntária.

### 4.3 CARACTERÍSTICAS DO ALUNO QUE EVADE VOLUNTARIAMENTE

Após o emprego do algoritmo de classificação *Naive-Bayes* (RAPIDMINER GMBH, 2019), foram obtidos os valores dos atributos onde a probabilidade de resultar em evasão voluntária alcançou maiores índices perante a probabilidade de ocorrerem outras evasões (Quadro 6).

**Quadro 6 – Valores de atributos com maior probabilidade para evasão voluntária**

Atributo	Valor	Probabilidade Outras Evasões	Probabilidade Evasão Voluntária
CH_TOTAL_CURSO	3215	0,001	0,017
CRA	3	0,011	0,124
CRN	1	0,003	0,076
DEFICIENCIA	Visual	0,003	0,010
ESTADO_CIVIL	Uniao Estavel	0,003	0,007
	Divorciado(a)	0,003	0,007
ETNIA	Não Declarada	0,013	0,024
	Preta	0,062	0,093
FORMA_INGRESSO	SISU	0,072	0,511
MUNICIPIO	MUQUI	0,001	0,019
NATURALIDADE	VOLTA REDONDA	0,001	0,009
NOME_CURSO	Matemática - Licenciatura - Alegre	0,016	0,081
NUM_MAX_PERIODOS	13	0,029	0,105
NUM_PERIODOS	8	0,068	0,159
PAIS	Brasil	0,999	1,000
PERIODO_INGRESSO	2º Semestre	0,401	0,667
SEXO	M	0,474	0,542
TIPO_COTA	Renda Normal e PPI	0,031	0,119
TURNO	Noturno	0,210	0,496
UF	SP	0,004	0,012
UF_NATURALIDADE	Ceará	0,001	0,007

Fonte: Dados da pesquisa.

Os valores de coeficiente de rendimento com maior probabilidade de evasão observados, CRA e CRN em torno 3 e 1, respectivamente, são compatíveis com os resultados encontrados por Sales Junior (2013), que pesquisou dados de alunos de

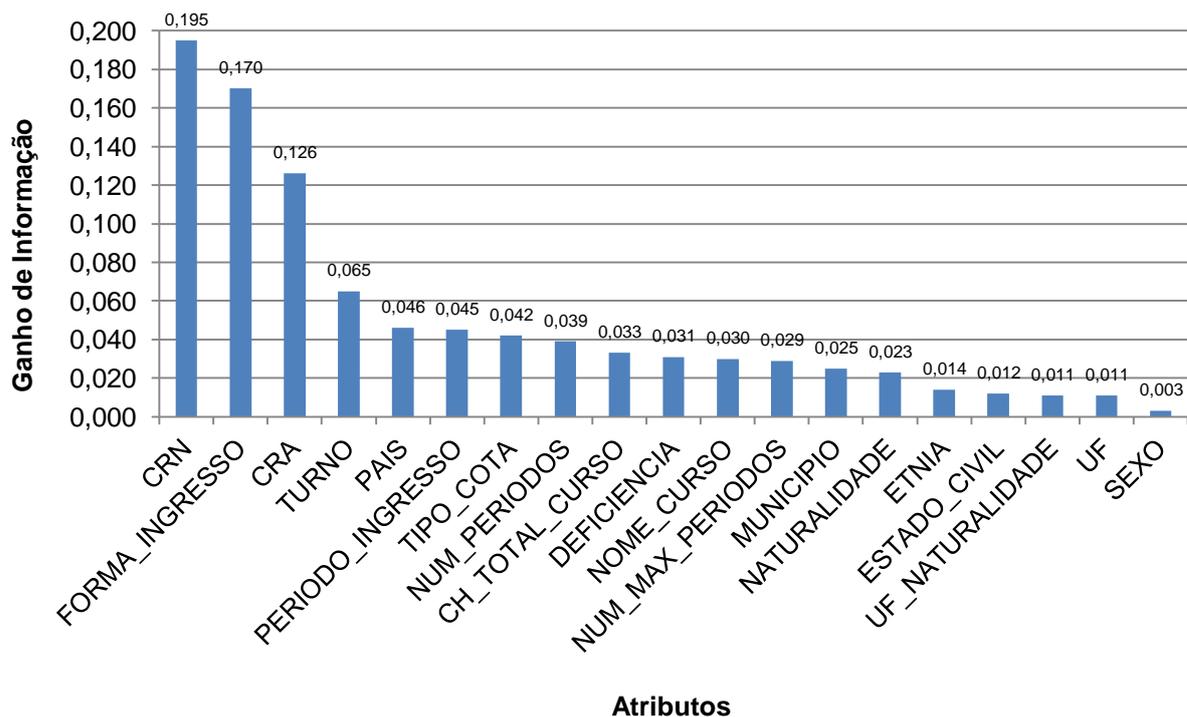
todos os cursos de graduação da UFES e observou o fato de nenhum dos estudantes formados possuírem coeficiente de rendimento inferior a 4,1.

Quanto à Etnia, o valor “Não Declarada” obteve a maior probabilidade de ocorrer “Evasão voluntária”, em relação a ocorrer “Outras evasões”. Contudo, quando declarada a etnia, aqueles pertencentes à denominação “Preta” apresentaram maior propensão para evadir voluntariamente, fato este compatível também com os resultados obtidos por Sales Junior (2013), que também encontrou dados sugerindo uma maior propensão desses estudantes à evasão. Em sua pesquisa, o autor verificou que 7,6% dos participantes se declararam pretos, porém o percentual de deste grupo entre os alunos evadidos era de 8,6%, e no de formados era de 6,5%.

O valor “Renda Normal e PPI”, encontrado no atributo “TIPO\_COTA”, está relacionado à reserva de vagas prevista pela Lei nº 12.711/2012 e refere-se a alunos autodeclarados pretos, pardos e indígenas, cuja renda familiar *per capita* é superior a um salário mínimo e meio.

No Gráfico 1, também foi possível identificar os atributos de maior influência na classificação dos alunos após o cálculo do ganho de informação realizado pelo programa *RapidMiner* (RAPIDMINER GMBH, 2019).

**Gráfico 1 – Ganho de informação dos atributos utilizados na pesquisa**



Fonte: Dados da pesquisa.

À esquerda, no Gráfico 1, estão presentes os atributos que tiveram maior influência durante a classificação dos alunos e, à direita, aqueles que apresentaram menor impacto durante essa tarefa. Segundo Andrade *et al.* (2016), ao analisar o ganho de informação dos atributos do conjunto de dados, é possível visualizar o quanto cada um deles contribui, individualmente, na classificação final.

A presença de coeficientes de rendimento entre os maiores valores de ganho de informação alinha com a pesquisa realizada pelos autores supracitados, que analisou informações sobre evasão de alunos de Computação na UFPB, onde foi constatado um maior impacto das informações acadêmicas dos estudantes em sua situação final ao serem utilizados os resultados de ganho de informação dos atributos pesquisados.

#### 4.4 CLASSIFICAÇÃO DE ALUNOS EM RISCO POTENCIAL DE EVASÃO

A mensuração do risco de evasão entre os estudantes ativos no quadro de alunos dos dois centros da UFES em Alegre foi outro resultado produzido pela mineração dos dados, conforme pode ser observado através da Figura 6.

**Figura 6 – Interface de predições do programa *RapidMiner***

Row No.	MATR_ALUNO	FORMA_EVASAO	prediction(FORMA_EVASAO)	SEXO	ESTADO_CIVIL	ETNIA	NOME_CURSO	FORMA_INGRESSO	PE
10	20101023	Sem evasão	Outras evasões	F	Solteiro(a)	Branca	Agronomia - Alegre	Reopção de Curso	2°
11	20151055	Sem evasão	Outras evasões	M	Solteiro(a)	Branca	Agronomia - Alegre	Vestibular	1°
12	20151055	Sem evasão	Outras evasões	M	Solteiro(a)	Branca	Agronomia - Alegre	Vestibular	1°
13	20151055	Sem evasão	Evasão voluntária	M	Outros	Parda	Agronomia - Alegre	Vestibular	1°
14	20182061	Sem evasão	Evasão voluntária	M	Solteiro(a)	Parda	Agronomia - Alegre	SISU	2°
15	20182062	Sem evasão	Evasão voluntária	M	Solteiro(a)	Parda	Agronomia - Alegre	SISU	2°
16	20171021	Sem evasão	Outras evasões	M	Solteiro(a)	Branca	Agronomia - Alegre	SISU	1°
17	20171021	Sem evasão	Outras evasões	F	Solteiro(a)	Branca	Agronomia - Alegre	SISU	1°
18	20171021	Sem evasão	Outras evasões	F	Solteiro(a)	Branca	Agronomia - Alegre	SISU	1°
19	20151057	Sem evasão	Evasão voluntária	F	Solteiro(a)	Branca	Agronomia - Alegre	Vestibular	1°
20	20181032	Sem evasão	Outras evasões	F	Solteiro(a)	Branca	Agronomia - Alegre	SISU	1°
21	20161024	Sem evasão	Evasão voluntária	M	Solteiro(a)	Parda	Agronomia - Alegre	Vestibular	1°
22	20141016	Sem evasão	Outras evasões	F	Solteiro(a)	Parda	Agronomia - Alegre	Vestibular	1°
23	20191096	Sem evasão	Evasão voluntária	F	Solteiro(a)	Branca	Agronomia - Alegre	SISU	1°

Fonte: Dados da pesquisa.

Nota: Imagem extraída do programa *RapidMiner Studio*.

Na Figura 6, é mostrada a interface do programa *RapidMiner* com os resultados após o modelo de predição gerado pela mineração dos dados de alunos evadidos entre os anos de 2016 e 2018 ser aplicado sobre os dados de alunos que ainda não passaram por processo de evasão. A coluna “*prediction(FORMA\_EVASAO)*” traz a classificação de cada aluno quanto a sua possibilidade de evadir voluntariamente ou através de outras formas.

Do total de 2.553 estudantes de graduação presencial, ativos até o momento da coleta dos dados, no CCAE e no CCENS, da UFES, o algoritmo classificou 1.560 deles como tendo a possibilidade de evadirem-se de forma voluntária, totalizando 61,10% das predições efetuadas (RAPIDMINER GMBH, 2019). Comparando-se com o percentual de 31,49% encontrado na tabela de alunos evadidos utilizada na construção do modelo de predição, o índice de evasões voluntárias previstas sugere que os dois centros pesquisados terão um aumento em torno de 29,61 pontos percentuais nesta forma de evasão nos próximos anos.

De maneira semelhante, observou-se um aumento no percentual de alunos com forma de ingresso pelo SISU/MEC nas mesmas tabelas supracitadas. Dos alunos que haviam passado por processo de evasão, 21,02% ingressaram por esse sistema de seleção unificada, enquanto os que ainda estavam em atividade totalizavam 65,18%.

O impacto do SISU/MEC em índices de evasão no ensino superior brasileiro tem sido objeto de estudos recentes. Gemaque e Souza (2016) destacam que, através deste sistema de seleção, o candidato tem a opção de escolher dois cursos para ingressar, sendo possível alterá-los durante o prazo das inscrições. Os candidatos são selecionados conforme a nota obtida no ENEM, dentro do número de vagas oferecidas em cada curso e por modalidade de concorrência. Segundo os autores, sua adesão pelas IFES tornou mais evidente a dinâmica de os alunos ingressarem na segunda opção, quando não classificados na primeira, visando prepararem-se para um novo ingresso no curso realmente desejado.

#### 4.5 AVALIAÇÃO DO MÉTODO EMPREGADO

Através do Quadro 7, é possível visualizar o total de erros e acertos de classificação realizados durante o teste de validação cruzada aplicado sobre os dados dos alunos que passaram por processo de evasão.

**Quadro 7 – Matriz de confusão resultante da validação cruzada**

		Predito	
		Outras Evasões	Evasão Voluntária
Verdadeiro	Outras Evasões	825	91
	Evasão Voluntária	87	334

Fonte: Dados da pesquisa.

Com os acertos obtidos, aqueles em que o valor predito coincide com o valor verdadeiro, a acurácia do algoritmo *Naive-Bayes* atingiu o percentual de 86,68% para o modelo de predição gerado a partir do grupo de dados de alunos evadidos utilizados na pesquisa (RAPIDMINER GMBH, 2019).

O valor alcançado com o coeficiente *Kappa* foi de 0,69 (RAPIDMINER GMBH, 2019). Esse índice alcançou uma força de concordância considerada substancial, por estar na faixa entre 0,61 a 0,80, conforme estabelecem os rótulos qualitativos apresentados por Koch e Landis (1977).

## 5 CONCLUSÃO

Os esforços empreendidos neste trabalho buscaram reunir informações que possibilitassem adquirir conhecimentos que viessem a favorecer a permanência de alunos integrantes do quadro de discentes da UFES.

Para isso, buscou-se contextualizar o histórico do ensino superior público no Brasil e na instituição pesquisada, seus fatores que influenciam a evasão de alunos bem como algumas iniciativas empregadas para a redução do problema. Também foram abordados conceitos e técnicas de mineração de dados aplicáveis para fins educacionais.

Após o tratamento e a mineração dos dados encontrados, as probabilidades obtidas pelo algoritmo de classificação *Naive-Bayes* e os atributos de maior influência obtidos pelo cálculo de ganho de informação permitiram identificar características dos alunos que evadem de forma voluntária nos centros pesquisados. Também foi possível obter, através da mineração dos dados, uma listagem identificando alunos com potencial risco de evasão.

Perante o exposto e diante da acurácia e da força de concordância alcançada com o método empregado, chegou-se à conclusão de que é possível ampliar o conhecimento da gestão universitária da UFES e sua capacidade de tomada de decisões, ao utilizar as informações contidas em seus bancos de dados com o intuito de reduzir os índices de evasão na instituição.

Tendo à sua disposição o conjunto de estudantes com maior probabilidade de abandonarem seus cursos, e conhecendo as características com maior influência em colocá-los nessa circunstância, a gestão universitária terá um acréscimo em sua capacidade de direcionar recursos para reduzir a saída desses alunos.

Para a realização de trabalhos futuros, sugere-se que pesquisas sejam conduzidas objetivando descobrir os fatores que motivariam a permanência de estudantes que ingressaram na instituição via SISU/MEC.

Os procedimentos empregados com os métodos deste trabalho permitiram a criação do produto técnico aplicado resultante da pesquisa, Apêndice A, que consistiu num grupo de instruções a serem seguidas por gestores da UFES que pretendam

reproduzir resultados semelhantes aos encontrados neste trabalho em bases de dados, atuais ou futuras, que contenham os mesmos atributos utilizados.

## REFERÊNCIAS

- ALMEIDA, L. M. *et al.* Uma ferramenta para extração de padrões. **Revista Eletrônica de Iniciação Científica**, v. 3, 2003. Disponível em: <http://www.cin.ufpe.br/~lma3/UmaFerramentaParaExtracaoDePadroes.pdf>. Acesso em: 18 abr. 2019.
- ALVES, R. D.; CECHINEL, C.; QUEIROGA, E. M. Predição do desempenho de matemática e suas tecnologias do ENEM utilizando técnicas de mineração de dados. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 7., 2018, Fortaleza. **Anais...** Fortaleza: Sociedade Brasileira de Computação, 2018. p. 469-479. Disponível em: <http://www.br-ie.org/pub/index.php/wcbie/article/view/8271/5948>. Acesso em: 28 maio 2019.
- AMARAL, F. **Aprenda mineração de dados: teoria e prática**. Rio de Janeiro: Alta Books, 2016. 240 p. ISBN 978-85-7608-988-9.
- AMO, S. Técnicas de mineração de dados. **Jornada de Atualização em Informática**, 2004. Disponível em: <http://files.sistemas2012.webnode.com.br/200000095-bf367bfb43/Tecnicas%20de%20Minera%C3%A7%C3%A3o%20de%20Dados.pdf>. Acesso em: 4 out. 2018.
- ANDRADE, L. P. *et al.* Evasão de estudantes universitários: diagnóstico a partir de dados acadêmicos e socioeconômicos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 27., 2016, Uberlândia. **Anais...** Uberlândia: Sociedade Brasileira de Computação, 2016. p. 926-935. Disponível em: <http://br-ie.org/pub/index.php/sbie/article/view/6778/4663>. Acesso em: 28 set. 2018.
- ASSOCIAÇÃO NACIONAL DOS DIRIGENTES DAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino**. Brasília: [s.n.], 1996. Disponível em: [http://www.andifes.org.br/wp-content/files\\_flutter/Diplomacao\\_Retencao\\_Evasao\\_Graduacao\\_em\\_IES\\_Publicas-1996.pdf](http://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf). Acesso em: 22 set. 2018. Relatório.
- AZEVEDO, S. L. A eficácia do programa de assistência estudantil (PROAES) na Universidade Federal do Espírito Santo para redução dos índices de evasão dos alunos de graduação. **Revista Guará**, v. 1, n. 6, 2017. Disponível em: <http://periodicos.ufes.br/guara/article/viewFile/15113/10699>. Acesso em: 9 out. 2018.
- BAKER, R. S. J. D. *et al.* Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1-29, 2013. Disponível em: <http://br-ie.org/pub/index.php/pie/article/view/2341>. Acesso em: 12 out. 2018.

BAKER, R. S. J.; CARVALHO, A. M. J. B.; ISOTANI, S. Mineração de dados educacionais: oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 2, p. 3, 2011. ISSN 1414-5685. Disponível em: <http://br-ie.org/pub/index.php/rbie/article/view/1301/1172>. Acesso em: 25 set. 2018.

BARDAGI, M. P.; HUTZ, C. S. Não havia outra saída: percepções de alunos evadidos sobre o abandono do curso superior. **PsicoUSF**, Itatiba, v. 14, n. 1, p. 95-105, abr. 2009. Disponível em: [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-82712009000100010&lng=pt&nrm=iso](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-82712009000100010&lng=pt&nrm=iso). Acesso em: 29 maio 2018.

BARROSO M. F.; FALCÃO, E. B. M. Evasão universitária: o caso do Instituto de Física da UFRJ. In: ENCONTRO NACIONAL DE PESQUISA EM ENSINO DE FÍSICA, 9., 2004, Jaboticatubas. **Anais...** Jaboticatubas: Sociedade Brasileira de Física, 2004. p. 1-14. Disponível em: <http://www.if.ufrgs.br/gra/agenda/co12-2.pdf>. Acesso em: 29 set. 2018.

BORGES, C. C. M. **A política de assistência estudantil a partir da execução do programa auxílio permanência no Instituto Federal de Alagoas - campus Palmeira dos Índios**: a permanência pelo auxílio. 2015. 126 f. Dissertação de Mestrado em Serviço Social - Universidade Federal da Paraíba, João Pessoa. Disponível em: <http://bdtd.biblioteca.ufpb.br/handle/tede/8395>. Acesso em: 26 maio 2018.

BRAGA, A. P.; LACERDA, W. S. Experimento de um Classificador de Padrões Baseado na Regra Naive de Bayes. **INFOCOMP**, v. 3, n. 1, p. 30-35, maio 2004. ISSN 1982-3363. Disponível em: <http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/60>. Acesso em: 19 out. 2018.

BRASIL. Decreto Lei nº 7.234, de 19 de julho de 2010. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 19 jul. 2010, Seção 1, p. 5. ISSN 1677-7042.

BRASIL. Decreto Lei nº 12.711, de 29 de agosto de 2012. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 30 ago. 2012, Seção 1, p. 1-2. ISSN 1677-7042.

BRASIL. Decreto Lei nº 13.409, de 28 de dezembro de 2016. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 29 dez. 2016, Seção 1, p. 3. ISSN 1677-7042.

CAMBRUZZI, W.; CAZELLA, S. C.; RIGO, S. J. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. **Workshop de Desafios da Computação Aplicada à Educação**, p. 168-177, 2012. Disponível em: <http://www.br-ie.org/pub/index.php/desafie/article/view/2787/2440>. Acesso em: 25 set. 2018.

CAMILO, C. O.; SILVA, J. C. Mineração de dados: conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás**, p. 1-28, ago. 2009. Disponível em: [https://rozero.webcindario.com/disciplinas/fbmg/dm/RT-INF\\_001-09.pdf](https://rozero.webcindario.com/disciplinas/fbmg/dm/RT-INF_001-09.pdf). Acesso em: 3 out. 2018.

CASTANHEIRA, L. G. **Aplicação de técnicas de mineração de dados em problemas de classificação de padrões**. 2008. 94 f. Dissertação de Mestrado - Universidade Federal de Minas Gerais, Belo Horizonte. Disponível em: [https://www.ppgee.ufmg.br/documentos/Defesas/777/Dissertacao\\_LucianaCastanheira.pdf](https://www.ppgee.ufmg.br/documentos/Defesas/777/Dissertacao_LucianaCastanheira.pdf). Acesso em: 3 out. 2018.

CASTILLO, F. L. **Técnicas para visualizar informação e extrair conhecimento de bases de informações**: estudo de caso da produção dos professores do PPGI-UFES. 2017. 107 f. Dissertação de Mestrado em Informática - Universidade Federal do Espírito Santo, Vitória. Disponível em: <http://repositorio.ufes.br/jspui/handle/10/9844>. Acesso em: 30 out. 2018.

CIFERRI, C. D. A. *et al.* Relatório técnico "métricas de avaliação". **SCA**, São Carlos, p. 15, set. 2009. Projeto "um ambiente para análise de dados da doença anemia falciforme". Disponível em: <http://conteudo.icmc.usp.br/pessoas/taspardo/TechReportUFSCar2009a-MatosEtAl.pdf>. Acesso em: 31 out. 2018.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37-46, abr. 1960. ISSN 1552-3888.

CORNELIUS JUNIOR, R. **Uso da mineração de dados na identificação de alunos com perfil de evasão do ensino superior**. 2015. 137 f. Dissertação - Universidade de Santa Cruz do Sul, Santa Cruz do Sul. Disponível em: <https://repositorio.unisc.br/jspui/handle/11624/535>. Acesso em: 02 out. 2018.

CÔRTEZ, S. C.; PORCARO, R. M.; LIFSCHITZ, S. Mineração de dados: funcionalidades, técnicas e abordagens. PUC-RioInf. **MCC10/02**, maio 2002. Disponível em: [ftp://139.82.16.194/pub/docs/techreports/02\\_10\\_cortes.pdf](ftp://139.82.16.194/pub/docs/techreports/02_10_cortes.pdf). Acesso em: 4 out. 2018.

FÁVERO, M. L. A. A universidade no Brasil: das origens à reforma universitária de 1968. **Educar em Revista**, Curitiba, n. 28, p. 17-36, dez. 2006. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0104-40602006000200003&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-40602006000200003&lng=en&nrm=iso). Acesso em: 15 maio 2018.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMITH, P. From data mining to knowledge discovery in databases. **AI magazine**, Providence, v. 17, n. 3, p. 37-54, 1996. ISSN 0738-4602-1996. Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>. Acesso em: 3 out. 2018.

FERREIRA, V. H. *et al.* Identificação de perfis de evasão e mau desempenho para geração de alertas num contexto de educação à distância. **Revista Latinoamericana de Tecnologia Educativa**, Madri, v. 13, n. 2, p. 61-76, 2014. ISSN 1695-288X.

FONSECA, R. C. V. **Metodologia do trabalho científico**. Curitiba: IESDE Brasil S.A., 2009. 92 p. ISBN 978-85-387-2221-2.

FONSECA, R.; SILVA, P.; SILVA, R. Acordo inter-juízes: o caso do coeficiente kappa. **Instituto Superior de Psicologia Aplicada**, p. 81-90, 2007. ISSN 1645-7927. Disponível em: <http://hdl.handle.net/10400.12/1263>. Acesso em: 4 nov. 2018.

FRANK, E.; WITTEN, I. H. **Data mining**: practical machine learning tools and techniques. 2. ed. San Francisco: Elsevier Inc., 2005. ISBN 0-12-088407-0.

GALVÃO, N. D.; MARIN, H. F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, São Paulo, v. 22, n. 5, p. 686-690, out. 2009. ISSN 0103-2100.

GEMAQUE, L. S. B.; SOUZA, L. G. 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 & Multidisciplinaridade**, São Luís, v. 2, n. 1, p. 84-105, jan./jun. 2016. Disponível em: <http://www.periodicoseletronicos.ufma.br/index.php/ens-multidisciplinaridade/article/view/4872>. Acesso em: 15 maio 2019.

GIOLO, S. R. **Introdução à análise de dados categóricos com aplicações**. São Paulo: Blucher, 2017. 256 p. ISBN ISBN 978-85-212-1187-7.

KAESTNER, C. A. A. *et al.* Procedimentos de Validação Cruzada em Mineração de Dados para ambiente de Computação Paralela. **ERAD 2009**, Caxias do Sul, p. 233-236, mar. 2009. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/erad/2009/047.pdf>. Acesso em: 22 out. 2018.

KOCH, G. G.; LANDIS, J. R. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, jan. 1977. ISSN 0006-341X.

KOMATI, K. S.; REZENDE, J. M.; RESENDO, L. C. Text mining em documentos de patentes do USPTO: um estudo de caso usando o rapidminer, Serra, 2017. Disponível em: [http://ersi2017.nce.ufrj.br/wp-content/uploads/2017/05/177653\\_1.pdf](http://ersi2017.nce.ufrj.br/wp-content/uploads/2017/05/177653_1.pdf). Acesso em: 12 out. 2018.

LOBO, M. B. D. C. M. Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. **Instituto Lobo para Desenvolvimento da Educação, da Ciência e da Tecnologia**, 2012. Disponível em: [http://www.institutolobo.org.br/imagens/pdf/artigos/art\\_087.pdf](http://www.institutolobo.org.br/imagens/pdf/artigos/art_087.pdf). Acesso em: 28 maio 2018.

MANHÃES, L. M. B. *et al.* Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 22., 2011. Aracaju. **Anais...** Aracaju: [s.n], 2011. p. 150-159. Disponível em: <http://www.br-ie.org/pub/index.php/sbie/article/view/1585>. Acesso em: 28 maio 2018.

MARTINS, A. C. P. Ensino superior no Brasil: da descoberta aos dias atuais. **Acta Cirúrgica Brasileira**, São Paulo, v. 17, p. 4-6, 2002. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-86502002000900001&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-86502002000900001&lng=en&nrm=iso). Acesso em: 16 maio 2018.

MAZZILLI, S. Ensino, pesquisa e extensão: reconfiguração da universidade brasileira em tempos de redemocratização do Estado. **Revista Brasileira de Política e Administração da Educação - Periódico científico editado pela ANPAE**, [S.l.], v. 27, n. 2, dez. 2011. ISSN 2447-4193. Disponível em: <http://www.seer.ufrgs.br/index.php/rbpae/article/view/24770/14361>. Acesso em: 22 maio 2018.

MENDONÇA, A. W. P. C. A universidade no Brasil. **Revista Brasileira de Educação**, Rio de Janeiro, n. 14, p. 131-150, Agosto 2000. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-24782000000200008&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-24782000000200008&lng=en&nrm=iso). Acesso em: 25 maio 2018.

MICROSOFT CORPORATION. **Microsoft Excel 2010**. Versão 14.0.7166.5000. [S.l.]: Microsoft Corporation, 2010.

MINISTÉRIO DA EDUCAÇÃO. Sistema de Seleção Unificada, 2019. Disponível em: <http://www.sisu.mec.gov.br/inicial>. Acesso em: 28 maio 2019.

NONATO, R. T.; OLIVEIRA, S. R. M. Técnicas de mineração de dados para identificação de áreas com cana-de-açúcar em imagens Landsat 5. **Engenharia Agrícola**, Jaboticabal, v. 33, n. 6, p. 1268-1280, nov./dez. 2013. Disponível em: <http://www.scielo.br/pdf/eagri/v33n6/19.pdf>. Acesso em: 24 out. 2018.

OLIVEN, A. C. Histórico da Educação Superior no Brasil. **A educação superior no Brasil**, p. 31-42, 2002. Disponível em: [http://naipedigital.com/fid/images/docencia/moduloIII/Apostila\\_1\\_Alunos.pdf](http://naipedigital.com/fid/images/docencia/moduloIII/Apostila_1_Alunos.pdf). Acesso em: 16 maio 2018.

PARDO, T. A. S.; NUNES, M. D. G. V. Aprendizado Bayesiano Aplicado ao Processamento de Línguas Naturais. **NILC-ICMC-USP**, n. 180, 2002. Disponível em: [http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos\\_enviados/BIBLIOTECA\\_113\\_RT\\_180.pdf](http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_180.pdf). Acesso em: 22 out. 2018.

PEREIRA, F. C. B. **Determinantes da evasão de alunos e os custos ocultos para as instituições de ensino superior**: uma aplicação na Universidade do Extremo Sul Catarinense. 2003. 173 f. Tese de doutorado - Universidade Federal de Santa

Catarina, Florianópolis. Disponível em:  
<https://repositorio.ufsc.br/handle/123456789/86403>. Acesso em: 8 set. 2018.

RAPIDMINER GMBH. **RapidMiner Studio**. Versão 9.2.001. [S.l.]: RapidMiner GmbH, 2019.

ROSA, E. **Evasão no ensino superior**: causas e consequências. 1994. 239 f. Tese de doutorado - Escola de Administração de Empresas de São Paulo, São Paulo. Disponível em: <http://bibliotecadigital.fgv.br/dspace/handle/10438/4597>. Acesso em: 10 set. 2018.

SALES JUNIOR, J. S. **Uma análise estatística dos fatores de evasão e permanência de estudantes de graduação presencial da UFES**. 2013. 111 f. Dissertação de Mestrado - Universidade Federal do Espírito Santo, Vitória, 2013. Disponível em: <http://repositorio.ufes.br/handle/10/2483>. Acesso em: 26 set. 2018.

SILVEIRA, M. M. **A assistência estudantil no ensino superior**: uma análise sobre as políticas de permanência das universidades federais brasileiras. 2012. 137 f. Dissertação de mestrado em política social, Universidade Católica de Pelotas, Pelotas. Disponível em: [http://pos.ucpel.edu.br/ppgps/wp-content/uploads/sites/5/2018/03/MIRIAM.SILVEIRA\\_A-Assist%C3%Aancia-Estudantil-no-Ensino-Superior-%E2%80%93-uma-an%C3%A1lise-sobre-as-pol%C3%ADticas-de-perman%C3%Aancia-das-universidades-federais-brasileiras.pdf](http://pos.ucpel.edu.br/ppgps/wp-content/uploads/sites/5/2018/03/MIRIAM.SILVEIRA_A-Assist%C3%Aancia-Estudantil-no-Ensino-Superior-%E2%80%93-uma-an%C3%A1lise-sobre-as-pol%C3%ADticas-de-perman%C3%Aancia-das-universidades-federais-brasileiras.pdf). Acesso em: 26 set. 2018.

SÓFOCLES. **A trilogia tebana**: Édipo rei, Édipo em Colono, Antígona. Tradução de Mário da Gama Kury. Rio de Janeiro: Zahar, 1990. 262 p. ISBN 978-85-7110-081-7.

SOUSA, M. M. A. C. F. F. **Mineração de dados educacionais**: previsão de notas parciais utilizando classificação. 2017. 84 f. Dissertação de mestrado, Universidade Federal do Amazonas, Manaus. Disponível em:  
<https://tede.ufam.edu.br/handle/tede/6514>. Acesso em: 22 out. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Resolução nº 58/2008, de 17 de dezembro de 2008**, 2008. Disponível em:  
[https://aluno.ufes.br/Documentos/resolucao\\_058\\_2008.pdf](https://aluno.ufes.br/Documentos/resolucao_058_2008.pdf). Acesso em: 29 abr. 2019.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **A instituição**, 2013a. Disponível em: <http://portal.ufes.br/institui%C3%A7%C3%A3o>. Acesso em: 27 set. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **História**. Campus de Alegre, 2013b. Disponível em: <http://alegre.ufes.br/historia>. Acesso em: 9 out. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Sobre o Programa de Assistência Estudantil da Ufes**, 2013c. Disponível em: <http://proaeci.ufes.br/proaesufes>. Acesso em: 01 jun. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Como ingressar nos cursos presenciais de graduação**, 2013d. Disponível em: <http://www.ufes.br/como-ingressar-nos-cursos-presenciais-de-gradua%C3%A7%C3%A3o>. Acesso em: 28 maio 2019.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **UFES, 60 anos**. Vitória: EDUFES, 2014. 100 p. ISBN 978-85-7772-222-8. Disponível em: [https://issuu.com/ufes/docs/livro\\_60anos\\_final](https://issuu.com/ufes/docs/livro_60anos_final). Acesso em: 01 jun. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Resolução nº 44/2015, de 22 de dezembro de 2015**, 2015. Disponível em: [http://www.alegre.ufes.br/sites/alegre.ufes.br/files/resolucao\\_no.\\_44.2015\\_-\\_nova\\_estrutura\\_do\\_cca\\_2\\_1\\_1.pdf](http://www.alegre.ufes.br/sites/alegre.ufes.br/files/resolucao_no._44.2015_-_nova_estrutura_do_cca_2_1_1.pdf). Acesso em: 1 nov. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Resolução nº 13/2016, de 27 de abril de 2016**, 2016. Disponível em: <http://sisu.ufes.br/sites/sisu.ufes.br/files/field/anexo/Resolucao%20CEPE-UFES%20n%C2%BA%2013-2016.pdf>. Acesso em: 2019 abr. 28.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Portaria nº 1831, DE 25 de agosto de 2017**, 2017a. Disponível em: [http://www.ufes.br/sites/default/files/anexo/portaria\\_1831\\_proaes-ufes.pdf](http://www.ufes.br/sites/default/files/anexo/portaria_1831_proaes-ufes.pdf). Acesso em: 29 set. 2018.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Manual do Estudante**, 2017b. Disponível em: [http://www.prograd.ufes.br/sites/prograd.ufes.br/files/field/anexo/manual\\_do\\_estudanteweb.pdf](http://www.prograd.ufes.br/sites/prograd.ufes.br/files/field/anexo/manual_do_estudanteweb.pdf). Acesso em: 29 abr. 2019.

## APÊNDICE A – PRODUTO TÉCNICO RESULTANTE DA DISSERTAÇÃO



**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO – UFES  
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS – CCJE  
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO PÚBLICA**

**PRODUTO TÉCNICO APLICADO RESULTANTE DE DISSERTAÇÃO**

<b>Proponente:</b>	<b>Sergio Rodrigues Silveira</b>
<b>Matrícula:</b>	<b>2017230065</b>

<b>Título:</b>	<b>CLASSIFICAÇÃO DE ALUNOS DE GRADUAÇÃO QUANTO À FORMA DE EVASÃO ATRAVÉS DE BANCOS DE DADOS DOS CENTROS DA UFES EM ALEGRE</b>
----------------	---

<b>Instituição:</b>	<b>Universidade Federal do Espírito Santo</b>
<b>Alegre/ES - 2019</b>	

**PROGRAMA DE MESTRADO EM GESTÃO PÚBLICA – DADOS DO PROPONENTE**

<b>NOME:</b> Sergio Rodrigues Silveira		
<b>DATA DA TITULAÇÃO:</b>		
<b>NACIONALIDADE:</b> Brasileiro	<b>ESTADO CIVIL:</b> Casado	
<b>Endereço Residencial:</b> Rua Virgílio Rezende, 156 – Colina		
<b>CIDADE:</b> Alegre	<b>ESTADO:</b> ES	<b>CEP:</b> 29500-000
<b>INSTITUIÇÃO DE VÍNCULO:</b> Universidade Federal do Espírito Santo		
<b>DEPARTAMENTO:</b> Departamento de Suporte à Gestão – Setorial Sul		
<b>CARGO:</b> Assistente em Administração		
<b>EMAIL PESSOAL:</b> silveirasr@gmail.com		

## RESUMO

A evasão é um fenômeno que representa um problema para as instituições de ensino, sejam públicas ou particulares, ocasionando perda de recursos de todos os envolvidos no processo de ensino. A Mineração de Dados Educacionais é uma área de pesquisa recente que tem se demonstrado promissora no desenvolvimento de métodos para a exploração de dados produzidos em ambientes educacionais, ajudando na obtenção de resultados para mitigação dos fatores que impedem a permanência de alunos. O objetivo principal do presente produto técnico é a obtenção de informações que proporcionem a aquisição de conhecimentos relacionados à evasão estudantil na Universidade Federal do Espírito Santo e que possibilitem a tomada de decisões que favoreçam a permanência de seus alunos. O trabalho realizará uma abordagem quantitativa, utilizando dados secundários para obter padrões e modelos que permitam obter resultados preditivos com alunos que ainda não passaram por processo de evasão. Como fonte de dados, serão utilizadas tabelas contendo informações sociais e acadêmicas de alunos matriculados nos anos de 2007 a 2018 nos dois centros da UFES, localizados no município de Alegre. O pré-processamento, a transformação e a mineração dos dados serão realizados com auxílio dos programas Microsoft Excel e *RapidMiner Studio*.

Palavras-chave: Evasão. Permanência. Mineração de Dados. Classificação.

## 1 INTRODUÇÃO

A evasão é um problema que tem causado preocupação para as instituições de ensino em geral, sejam elas públicas ou particulares. A saída dos alunos traz consigo sérias consequências, uma vez que representa uma perda de recursos não só de organizações governamentais e privadas, mas também de alunos e seus familiares, que investem tempo e parte de suas provisões no esforço de construir seu futuro.

O problema da evasão é medido como a diferença entre a quantidade de alunos que ingressam num determinado curso, em determinado ano, e a quantidade desses mesmos alunos que conseguem concluir seus respectivos cursos. Essa medida inclui todos os alunos que abandonaram seus cursos independentemente dos motivos ou causas, excluindo-se os falecidos (ROSA, 1994).

As pesquisas com o propósito de investigar a evasão de alunos em instituições públicas ou privadas costumam apontar uma taxa de evasão em torno de 20% a 50%, sendo que as maiores incidências ocorrem no primeiro ano de curso, e os três primeiros meses são os que apresentam o maior número de alunos evadidos (BARDAGI; HUTZ, 2009).

A adoção de estratégias de combate à evasão estudantil ajuda a reduzir esses índices entre alunos pertencentes a grupos socioeconomicamente vulneráveis. Silveira (2012) defende que as ações de assistência estudantil são fundamentais para a permanência no ensino superior dos grupos estudantis que se enquadram na renda *per capita* familiar de até um salário mínimo e meio, possibilitando a conclusão da graduação desses alunos com qualidade e igualdade de oportunidades.

Uma recente área de pesquisa que tem se demonstrado promissora é a Mineração de Dados Educacionais (MDE), do inglês *Educational Data Mining*, pois permite o desenvolvimento de métodos para a exploração de dados que são produzidos em ambientes educacionais. Nesta abordagem, Baker, Carvalho e Isotani (2011) afirmam que atualmente a MDE vem se estabelecendo como uma forte linha de pesquisa com potencial para produzir melhorias na qualidade de ensino. Cambuzzi, Cazella e Rigo (2012) ressaltam que, perante a diversidade de dados envolvidos no contexto da evasão escolar, a exploração de algoritmos, técnicas e mecanismos de

MDE são cruciais para a obtenção de resultados abrangentes e úteis para a mitigação dos fatores que impedem a permanência dos alunos.

### 1.1 SITUAÇÃO PROBLEMA

A evasão estudantil no ensino superior é um fenômeno que vem crescendo no Brasil nos últimos anos. De acordo com Sales Junior (2013), entre as diversas causas estão os programas de expansão que promoveram a ampliação do número de instituições de ensino superior, de cursos oferecidos e de novas vagas. Conseqüentemente, ocorreu também o aumento do número de estudantes dos cursos de graduação que abandonam seus estudos antes de obter a diplomação.

Nos últimos anos, a Universidade Federal do Espírito Santo (UFES) também tem sido cenário de crescimento em evasões. Em sua pesquisa contendo dados estatísticos de evasão na UFES, o autor supracitado observou um expressivo aumento no índice de evasões ocorrido a partir de 2010, passando de 4,2% em 2009 para 7,7% em 2010. Também destacou que apenas no primeiro semestre de 2012 já se registravam mais evasões que em todo o ano de 2011.

### 1.2 OBJETIVOS

Perante o exposto, o presente produto técnico pretende oferecer um método que possibilite à gestão universitária obter informações que proporcionem a aquisição de conhecimentos relacionados à evasão estudantil na UFES e possibilitem a tomada de decisões que favoreçam a permanência de seus alunos.

### 1.3 JUSTIFICATIVA

Os crescentes índices de evasão estudantil na UFES representam um desafio que requer a ação de seus gestores visando combatê-la constantemente. As diretrizes adotadas através do Programa de Assistência Estudantil da UFES (PROAES-UFES) são iniciativas que viabilizam a permanência de alunos em situação de

vulnerabilidade socioeconômica. Essas ações podem ser potencializadas mediante a aquisição de um maior conhecimento da realidade de seus alunos.

A MDE é uma área de estudos capaz de produzir mecanismos que facilitam a obtenção de conhecimento e a tomada de decisão, aproveitando os bancos de dados produzidos em instituições de ensino. Sua utilização enriquece o conjunto de possibilidades para a melhoria do desempenho escolar.

Deste modo, a pesquisa justifica-se por oferecer à gestão pública da UFES a possibilidade de ampliar seu conhecimento relacionado à evasão estudantil, ajudando-a a maximizar as iniciativas já implantadas e a encontrar novas estratégias para propiciar a permanência nos cursos a um maior número de alunos.

#### 1.4 DESENVOLVIMENTO DO PRODUTO TÉCNICO

Nesta seção, apresenta-se o conjunto de procedimentos de mineração de dados adotados, visando adquirir de conhecimentos relacionados à evasão estudantil na UFES.

##### 1.4.1 Aquisição dos dados

Como fonte de dados para os procedimentos listados neste produto técnico, foram utilizados bancos de dados solicitados à Secretaria Única de Graduação – Setorial Sul, unidade da Universidade Federal do Espírito Santo subordinada ao Departamento de Suporte à Gestão, localizada no município de Alegre.

As tabelas fornecidas continham dados digitais em formato xls extraídos do programa informatizado Sistema de Informações para o Ensino – SIE, com informações sociais e acadêmicas de 1.340 alunos dos cursos de graduação, que passaram por processo de evasão nos referidos centros nos anos de 2016 a 2018, e outra contendo matrículas de 6.827 alunos no período de 2007 a 2018. Os códigos das rotinas do SIE para a emissão dos relatórios com os dados eram: 11.02.06.03.04, referente a alunos evadidos, e 11.02.06.03.13, referente a alunos ingressantes.

Os procedimentos empregados neste estudo podem ser aplicados a outros futuros conjuntos de alunos, desde que sejam utilizados os mesmos relatórios de dados do SIE contendo: alunos evadidos nos três anos mais recentes e tabela recente de ingressantes que inclua os alunos da tabela de evadidos.

Não deverão ser utilizados os dados de estudantes recém-matriculados que ainda não possuem coeficiente de rendimento indicando o desempenho de seus estudos.

#### 1.4.2 Preparação inicial dos dados

- Abra o arquivo contendo a tabela 11.02.06.03.04 de “Alunos Evadidos” utilizando o programa *Microsoft Excel 2010* (MICROSOFT CORPORATION, 2010);
- Na aba “Arquivo”, clique em “Salvar como”, nomeie o arquivo como “Alunos Evadidos.xlsx” e salve-o em local desejado;
- Volte para a aba “Página Inicial” e use o comando “Excluir linhas da planilha” após selecionar as linhas cujos registros contenham as formas de evasão “Erro de cadastro” e “Falecimento”;
- Use o comando “Excluir colunas da planilha” após selecionar todas as colunas, exceto “MATR\_ALUNO” e “FORMA\_EVASAO”;
- Salve e feche o arquivo “Alunos Evadidos.xlsx”;
- Abra a tabela 11.02.06.03.13 de “Ingressantes” com o programa *Microsoft Excel 2010*;
- Na aba “Arquivo”, clique em “Salvar como”, nomeie o arquivo como “Ingressantes.xlsx” e salve-o em local desejado;
- Selecione todos os valores de “CRN” clicando nas letras “AJ” identificadoras da coluna;
- Na aba “Dados”, clique em “Textos para coluna”, vá clicando em avançar e depois em concluir;

- Na aba “Página inicial”, use o comando “Excluir colunas da planilha” após selecionar as colunas “NOME\_ALUNO”, “DT\_NASCIMENTO”, “COD\_CURSO”, “ANO\_INGRESSO”, “TURNO\_ALUNO\_ITEM”, “ANO\_EVASAO”, “PERIODO\_EVASAO”, “DESCR\_MAIL”, “COTISTA”, “CEP”, “TIPO\_LOGRADOURO”, “RUA”, “NUMERO”, “COMPLEMENTO”, “BAIRRO”, “ESTADO”, “CPF”, “RG”, “RG\_ORGAO\_EMISSOR” e “RG\_ESTADO”;
- Na aba arquivo, clique em “Salvar”;
- Na aba “Página Inicial”, use o comando “Excluir linhas da planilha” após selecionar as linhas cujos registros não contenham a forma de evasão “Sem evasão”;
- Na aba “Arquivo”, clique em “Salvar como”, nomeie o arquivo como “Dados Novos.xlsx” e salve-o em local desejado;
- Feche o arquivo “Dados Novos.xlsx”.

#### 1.4.3 Preparação dos dados com o programa de mineração de dados

- Utilizando o programa *RapidMiner Studio*<sup>2</sup> (RAPIDMINER GMBH, 2019), vá ao menu “File” e clique em “New process”, escolhendo a opção “Turbo Prep”;
- Clique em “Load Data”, “Importar Data”, “My Computer”, selecione o arquivo “Ingressantes.xlsx” no local onde este foi gravado e clique em “Next”;
- Na janela “Select the cells to import”, clique em “Next”;
- Na janela “Format your columns”, clique ao lado do nome das colunas para selecionar a opção “Change Type” e mude os tipos de dados de todas as colunas para “polynomial”, exceto CRN e CRA, que deverão ser “Integer”;
- Na coluna “FORMA\_EVASAO”, selecione a opção “Exclude column” e depois clique em “Next”;

---

<sup>2</sup> Utilizou-se a versão 9.2.001, plataforma WIN64, obtida no sítio <http://rapidminer.com>.

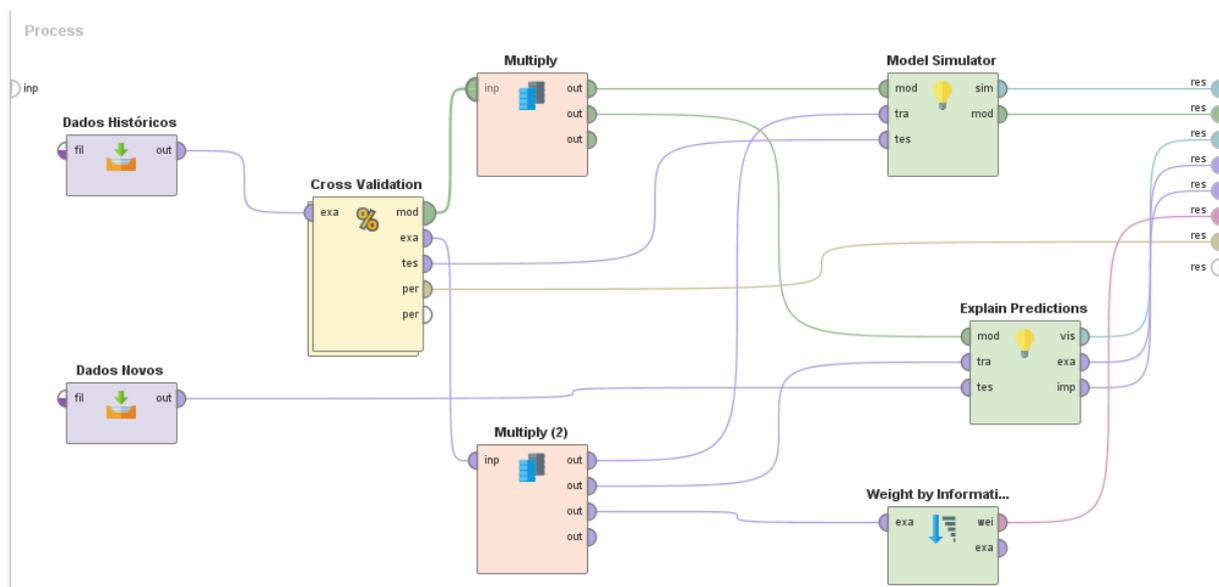
- Na janela “*Where to store the data?*”, selecione “*Local Repository/Data*” e clique em “*Finish*”;
- Clique em “*Load Data*”, “*Importar Data*”, “*My Computer*”, selecione o arquivo “*Alunos Evadidos.xlsx*” e clique em “*Next*”;
- Na janela “*Select the cells to import*”, clique em “*Next*”;
- Na janela “*Format your columns*”, clique ao lado do nome das colunas para selecionar a opção “*Change Type*” e mude os tipos de dados de todas as colunas para “*polynomial*” e depois clique em “*Next*”;
- Na janela “*Where to store the data?*”, selecione “*Local Repository/Data*” e clique em “*Finish*”;
- Clique na coluna “*FORMA\_EVASAO*” para selecioná-la, clique em “*Transform*” e utilize o comando “*Replace*” para renomear as formas de evasão “*Desistência*” e “*Desligamento por Abandono*” para “*Evasão Voluntária*”, e todas as demais formas para “*Outras evasões*”;
- Ao terminar as substituições, clique em “*Commit Transformation*”;
- Clique em “*Merge*” e, no campo “*Merge With*”, selecione “*Ingressantes*”;
- No campo “*Merge Type*”, selecione “*Inner Join*”;
- Em “*Add Join Keys*”, selecione “*MATR\_ALUNO*” nos dois campos;
- Clique em “*Update Preview*” e em “*Commit Merge*”;
- Ao lado da opção “*History*”, clique em “*...*” e selecione “*Export*”, “*Excel*”, “*Next*” e salve o arquivo como “*Dados Históricos*”;
- Clique em “*Load Data*”, “*Importar Data*”, “*My Computer*” e abra o arquivo “*Dados Novos.xlsx*”;
- Clique em “*Next*” e, na janela “*Format your columns*”, mude os tipos de dados de todas as colunas para “*polynomial*”, exceto CRN e CRA, que deverão ser “*Integer*”;

- Clique em “Next” e, na janela “Where to store the data?”, selecione “Local Repository/Data” e clique em “Finish”;
- Ao lado da opção “History”, clique em “...” e selecione “Export”, “Excel”, “Next” e salve o arquivo como “Dados Novos.xlsx”.

#### 1.4.4 Inclusão de processos para mineração dos dados

- Aperte F8 para ir até a tela “Design” e utilize a janela “Operators” para pesquisar o operador “Read Excel” e clique com o botão direito do mouse sobre este operador e peça para inseri-lo duas vezes;
- Após inseri-los, clique com o botão direito do mouse sobre os operadores e renomeie-os como “Dados Históricos” e “Dados Novos”;
- Pesquise e inclua também os operadores: “Cross Validation”, “Multiply” (duas vezes), “Model Simulator”, “Explain Predictions” e “Weight by Information Gain Ratio”;
- Conecte os operadores conforme o esquema disposto na Figura 1:

Figura 1 – Processo utilizado na mineração dos dados



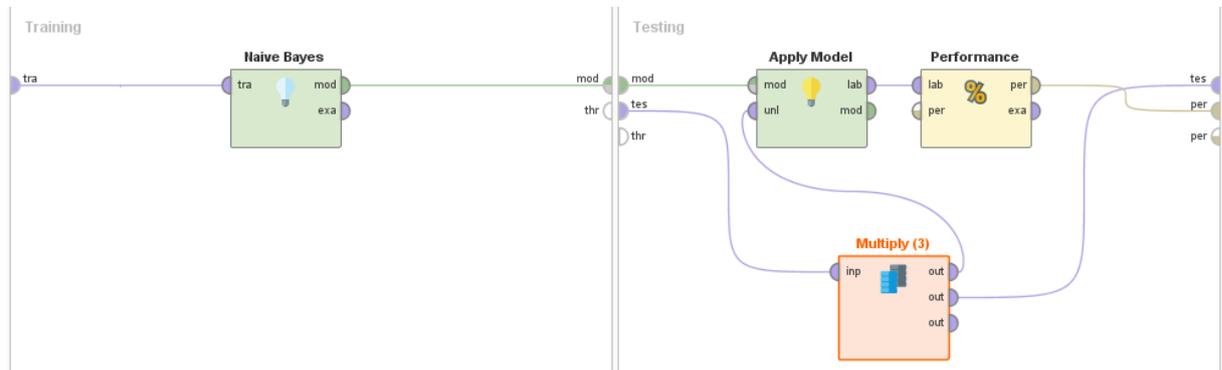
Fonte: Dados da pesquisa.

Nota: Imagem extraída do programa *RapidMiner Studio*.

- Dê clique-duplo no operador “Cross Validation”;

- Pesquise os operadores “*Naive Bayes*”, “*Multiply*”, “*Apply Model*”, “*Performance (Binominal Classification)*” e insira-os, arrastando e conectando-os conforme esquema disposto na Figura 2:

**Figura 2 – Operadores de classificação e de desempenho**



Fonte: Dados da pesquisa.

Nota: Imagem extraída do programa *RapidMiner Studio*.

- Clique no operador “*Performance (Binominal Classification)*” e marque a opção “*kappa*” na janela “*Parameters*”;
- Clique em “*Process*”;
- Clique no operador “*Dados Históricos*” e, na janela “*Parameters*”, clique em “*Import Configuration Wizard...*”;
- Selecione o arquivo “*Dados Históricos.xlsx*” e clique em “*Next*”;
- Na janela “*Select the cells to import*”, clique em “*Next*”;
- Na janela “*Format your columns*”, coloque todas as colunas com o tipo “*polynomial*”, exceto a coluna “*FORMA\_EVASAO*”, que deverá ser “*binominal*”;
- Clique ao lado do nome da coluna “*MATR\_ALUNO*” e, na opção “*Change Role*”, mude para “*id*”;
- Clique ao lado do nome da coluna “*FORMA\_EVASAO*” e, na opção “*Change Role*”, mude para “*label*”;
- Clique em “*Finish*”

- Clique no operador “Dados Novos” e, na janela “*Parameters*”, clique em “*Import Configuration Wizard...*”;
- Selecione o arquivo “Dados Novos.xlsx” e clique em “*Next*”;
- Na janela “*Select the cells to import*”, clique em “*Next*”;
- Na janela “*Format your columns*”, coloque todas as colunas com o tipo “*polynomial*”;
- Clique ao lado do nome da coluna “MATR\_ALUNO” e, na opção “*Change Role*”, mude para “*id*”;
- Clique ao lado do nome da coluna “FORMA\_EVASAO” e, na opção “*Change Role*”, mude para “*label*”;
- Clique em “*Finish*”.

#### 1.4.5 Mineração dos dados e visualização dos resultados

- No menu “*Process*”, clique na opção “*Run Process Locally*” ou aperte F11 no teclado.

Em “*Views/Results*” surgirão algumas abas com os resultados da mineração dos dados, tais como:

- “*Explain Predictions*”: onde é exibida a lista de alunos que ainda não passaram por processo de evasão, contendo a predição baseada na probabilidade de evadirem voluntariamente ou através de outras evasões;
- “*Model Simulator*”: onde é possível simular a probabilidade de evasão voluntária de acordo com a entrada de valores específicos dos atributos;
- “*Naive Bayes*”: onde são exibidas as probabilidades de cada valor dos atributos através de tabelas e gráficos;
- “*Performance*”: onde são exibidos o percentual de acurácia da classificação e o coeficiente Kappa;

- “*Weight by Information Gain Ratio*”: onde são exibidos os atributos que tiveram maior influência na classificação dos alunos, de acordo com o ganho de informação obtido por cada atributo;

Clicando na opção “*Open in Turbo Prep*”, que aparece em algumas das tabelas de resultados, será possível salvar o conteúdo exibido através do mesmo comando de exportar dados em formato Excel visto anteriormente.

## REFERÊNCIAS

- BAKER, R. S. J.; CARVALHO, A. M. J. B.; ISOTANI, S. Mineração de dados educacionais: oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 2, p. 3, 2011. ISSN 1414-5685. Disponível em: <http://br-ie.org/pub/index.php/rbie/article/view/1301/1172>. Acesso em: 25 set. 2018.
- BARDAGI, M. P.; HUTZ, C. S. Não havia outra saída: percepções de alunos evadidos sobre o abandono do curso superior. **PsicoUSF**, Itatiba, v. 14, n. 1, p. 95-105, abr. 2009. Disponível em: [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-82712009000100010&lng=pt&nrm=iso](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-82712009000100010&lng=pt&nrm=iso). Acesso em: 29 maio 2018.
- CAMBRUZZI, W.; CAZELLA, S. C.; RIGO, S. J. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. **Workshop de Desafios da Computação Aplicada à Educação**, p. 168-177, 2012. Disponível em: <http://www.br-ie.org/pub/index.php/desafie/article/view/2787/2440>. Acesso em: 25 set. 2018.
- MICROSOFT CORPORATION. **Microsoft Excel 2010**. Versão 14.0.7166.5000. [S.l.]: Microsoft Corporation, 2010.
- RAPIDMINER GMBH. **RapidMiner Studio**. Versão 9.2.001. [S.l.]: RapidMiner GmbH, 2019.
- ROSA, E. **Evasão no ensino superior**: causas e consequências. 1994. 239 f. Tese de doutorado - Escola de Administração de Empresas de São Paulo, São Paulo. Disponível em: <http://bibliotecadigital.fgv.br/dspace/handle/10438/4597>. Acesso em: 10 set. 2018.
- SALES JUNIOR, J. S. **Uma análise estatística dos fatores de evasão e permanência de estudantes de graduação presencial da UFES**. 2013. 111 f. Dissertação de Mestrado - Universidade Federal do Espírito Santo, Vitória, 2013. Disponível em: <http://repositorio.ufes.br/handle/10/2483>. Acesso em: 26 set. 2018.
- SILVEIRA, M. M. **A assistência estudantil no ensino superior**: uma análise sobre as políticas de permanência das universidades federais brasileiras. 2012. 137 f. Dissertação de mestrado em política social, Universidade Católica de Pelotas, Pelotas. Disponível em: [http://pos.ucpel.edu.br/ppgps/wp-content/uploads/sites/5/2018/03/MIRIAM.SILVEIRA\\_A-Assist%C3%Aancia-Estudantil-no-Ensino-Superior-%E2%80%93-uma-an%C3%A1lise-sobre-as-pol%C3%ADticas-de-perman%C3%Aancia-das-universidades-federais-brasileiras.pdf](http://pos.ucpel.edu.br/ppgps/wp-content/uploads/sites/5/2018/03/MIRIAM.SILVEIRA_A-Assist%C3%Aancia-Estudantil-no-Ensino-Superior-%E2%80%93-uma-an%C3%A1lise-sobre-as-pol%C3%ADticas-de-perman%C3%Aancia-das-universidades-federais-brasileiras.pdf). Acesso em: 26 set. 2018.