

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA FLUMINENSE**

**PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À
ENGENHARIA E GESTÃO**

Deivison Lamonica Barreto

**Análise da Evasão em Cursos de Ensino Superior do Instituto
Federal Fluminense utilizando Mineração de Dados**

Campos dos Goytacazes/RJ

2019

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA FLUMINENSE**

**PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À
ENGENHARIA E GESTÃO**

DEIVISON LAMONICA BARRETO

**Análise da Evasão em Cursos de Ensino Superior do Instituto
Federal Fluminense utilizando Mineração de Dados**

Aline Pires Vieira de Vasconcelos

(Orientadora)

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Instituto Federal de Educação, Ciência e Tecnologia Fluminense, no Curso de Mestrado Profissional em Sistemas Aplicados à Engenharia e Gestão (MPSAEG), como parte dos requisitos necessários à obtenção do título de Mestre em Sistemas Aplicados à Engenharia e Gestão.

Campos dos Goytacazes/RJ

2019

Biblioteca Anton Dakitsch
CIP - Catalogação na Publicação

B273a Barreto, Deivison Lamônica
Análise da Evasão em Cursos de Ensino Superior do Instituto Federal Fluminense utilizando Mineração de Dados / Deivison Lamônica Barreto - 2019.
84 f.: il. color.

Orientadora: Aline Pires Vieira de Vasconcelos

Dissertação (mestrado) -- Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campus Campos Centro, Curso de Mestrado Profissional em Sistemas Aplicados à Engenharia e Gestão, Campos dos Goytacazes, RJ, 2019.
Referências: f. 79 a 84.

1. mineração de dados. 2. , evasão. 3. ensino superior. I. Pires Vieira de Vasconcelos, Aline, orient. II. Título.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA FLUMINENSE

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS APLICADOS À
ENGENHARIA E GESTÃO

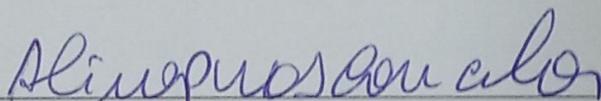
Deivison Lamonica Barreto

**Análise da Evasão em Cursos de Ensino Superior do Instituto Federal Fluminense
utilizando Mineração de Dados**

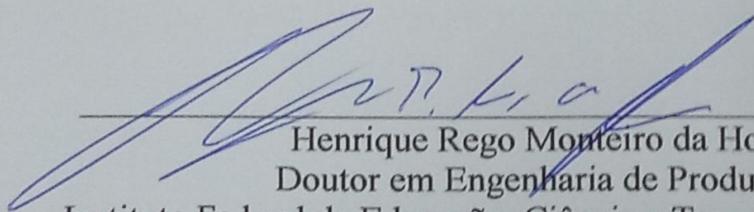
Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Instituto Federal de Educação, Ciência e Tecnologia Fluminense, no Curso de Mestrado Profissional em Sistemas Aplicados à Engenharia e Gestão (MPSAEG), como parte dos requisitos necessários à obtenção do título de Mestre em Sistemas Aplicados à Engenharia e Gestão.

Aprovado em 19 de dezembro de 2019.

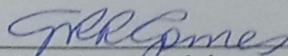
Banca Examinadora:



Aline Pires Vieira de Vasconcelos
Doutora em Engenharia de Sistemas e Computação
Instituto Federal de Educação, Ciência e Tecnologia Fluminense
(Orientadora)



Henrique Rego Monteiro da Hora
Doutor em Engenharia de Produção
Instituto Federal de Educação, Ciência e Tecnologia Fluminense



Geórgia Regina Rodrigues Gomes
Doutora em Informática
Universidade Federal Fluminense

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me ajudar a enfrentar as dificuldades ao longo do caminho. Foram três participações consecutivas no processo seletivo do Mestrado, fazendo do ingresso um desafio. Agora tenho a oportunidade de chegar a esse momento tão esperado.

À minha noiva Rafaela, por sua compreensão, paciência e bom humor, sempre afirmando que iniciei o Mestrado após o início do nosso relacionamento, fazendo dela a responsável pelo meu ingresso!

Aos meus pais, Sebastião e Eva, por todas as oportunidades que me deram ao longo da vida, o que vai muito além do tempo decorrido entre o início e a conclusão do Mestrado.

Ao meu irmão Gleidison pela colaboração em diversos momentos que precisei.

À minha orientadora Aline Pires, pelo direcionamento, ajudando-me tantas vezes no decorrer deste trabalho, permitindo-me concluí-lo com sucesso.

A todos os amigos do Polo de Inovação do Instituto Federal Fluminense, pelo apoio durante esse período, o que foi fundamental para a chegada desse momento.

RESUMO

A evasão pode ser definida como a saída de um estudante de um determinado curso, sem que o mesmo tenha sido concluído. Este problema afeta tanto instituições de ensino públicas como privadas, além de afetar também ao aluno que evadiu. Nesse sentido, o presente trabalho busca identificar padrões nos dados de alunos evadidos de cursos superiores por meio da aplicação de técnicas de mineração de dados, utilizando como estudo de caso o *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Para tanto, foram realizadas as seguintes etapas: definição dos objetivos da mineração de dados; seleção de dados do sistema acadêmico; extração de dados desse sistema; pré-processamento, quando foram realizadas todas as tarefas relativas à limpeza dos dados e demais procedimentos necessários para prepará-los para a etapa de mineração; mineração de dados, quando os dados foram submetidos a algoritmos a fim de encontrar padrões e, por fim, interpretação de resultados, quando foram analisadas as informações descobertas por meio da etapa de mineração de dados. De acordo com os resultados, a forma de ingresso de matrícula apareceu como raiz em seis de um total de dez cursos, demonstrando a relevância desse atributo na análise da evasão no contexto estudado. Entre os cursos com maiores taxas de evasão podem ser destacados o curso de Bacharelado em Sistemas de Informação, com 63,6% de evasão e o curso de Tecnologia em Sistemas de Telecomunicações, com 90%. Em relação a esses cursos, os alunos que ingressaram pelas modalidades Vestibular – Cota e Sisu – Cota aparecem entre os perfis que mais evadiram, estando presentes nos dois cursos mencionados. Nesse sentido, as informações encontradas permitem diagnosticar características dos perfis dos alunos evadidos de acordo com atributos disponíveis no sistema acadêmico, como por exemplo, a forma de ingresso, turno, sexo, entre outros. Assim, essas informações podem ser utilizadas pelos gestores da instituição de ensino nos processos de tomada de decisão e na elaboração de políticas públicas visando aumentar taxas de permanência dos alunos.

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

ABSTRACT

Dropout can be defined as a student leaving a course without having completed it. This problem affects both public and private educational institutions, as well as the student who has evaded. In this sense, the present work seeks to identify patterns in the data of students evaded from higher education through the application of data mining techniques, using as a case study the Campos Centro campus of Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Therefore, the following steps were performed: definition of data mining objectives; data selection from the academic system; data extraction from this system; data pre-processing, when all tasks related to data cleaning and other procedures necessary to prepare the data for the data mining phase were performed; data mining, when data were submitted to algorithms to find patterns and, finally, interpretation of results, when the information discovered through the data mining step was analyzed. According to the results, the form of enrollment appeared as root in six out of ten courses, demonstrating the relevance of this attribute in the evasion analysis in the studied context. Among the courses with the highest dropout rates, it can be highlighted the Bachelor Degree in Information Systems course, with 63.6% of dropout rates and the Technology Degree in Telecommunications Systems course, with 90%. In relation to these courses, students who entered Vestibular Quota and Sisu Quota modalities appear among the most evaded profiles, being present in the two mentioned courses. In this sense, the information found allows us to diagnose the characteristics of the students' profiles that have been evaded according to attributes available in the academic system, such as the form of admission, shift, gender, among others. Thus, this information can be used by educational institution managers in decision-making processes and in the elaboration of public policies aiming at increasing students' permanence rates.

Keywords: data mining, evasion, higher education.

LISTAS

Lista de Figuras

Figura 1: Etapas do processo de KDD.....	27
Figura 2: Exemplo de árvore de decisão.	30
Figura 3: Etapas da metodologia.	32
Figura 4: Tarefas realizadas para desenvolvimento do método proposto.	33
Figura 5: Diagrama de Venn representando a quantidade de publicações encontradas na base <i>Scopus</i>	44
Figura 6: Quantidade de publicações por ano.	44
Figura 7: Artigos mais citados.....	45
Figura 8: Quantidade de publicações por autor.	46
Figura 9: Fontes que mais publicaram sobre o tema.	46
Figura 10: Países que mais publicaram sobre o tema.....	47
Figura 11: Demonstrativo de alunos concluintes e evadidos por curso no período 2017-2 a 2019-1.....	54
Figura 12: Árvore de decisão para o curso de Bacharelado em Arquitetura e Urbanismo	56
Figura 13: Árvore de decisão para o curso de Bacharelado em Engenharia de Controle e Automação Fonte: Elaborado pelo autor.	57
Figura 14: Árvore de decisão para o curso de Bacharelado em Sistemas de Informação Fonte: Elaborado pelo autor.....	58
Figura 15: Árvore de decisão para o curso de Licenciatura em Ciências da Natureza. Fonte: Elaborado pelo autor.....	59
Figura 16: Árvore de decisão para o curso de Licenciatura em Geografia Fonte: Elaborado pelo autor.	60
Figura 17: Árvore de decisão para o curso de Licenciatura em Letras Fonte: Elaborado pelo autor.....	62
Figura 18: Árvore de decisão para o curso de Licenciatura em Matemática Fonte: Elaborado pelo autor.	63
Figura 19: Árvore de decisão para o curso de Tecnologia em Design Gráfico Fonte: Elaborado pelo autor.	65
Figura 20: Árvore de decisão para o curso de Tecnologia em Manutenção Industrial. Fonte: Elaborado pelo autor.....	66

Figura 21: Árvore de decisão para o curso de Tecnologia em Sistemas de Telecomunicações.
Fonte: Elaborado pelo autor. 67

Lista de Quadros

Quadro 1: Número de alunos com matrículas desvinculadas em relação ao total de alunos matriculados	17
Quadro 2: Metas do Plano Nacional de Educação (PNE)	20
Quadro 3: Quantidade de registros por semestre	34
Quadro 4: Relação de atributos e seus respectivos domínios	39
Quadro 5: Conceitos e tesouros utilizados.....	43
Quadro 6: String de busca realizada na base Scopus utilizando os conceitos ABC.....	43
Quadro 7: Comparação dos trabalhos relacionados	52
Quadro 8: Perfis de evasão encontrados nos trabalhos relacionados	53
Quadro 9: Principais regras de associação encontradas	69
Quadro 10: Panorama geral de evasão nos cursos investigados.....	71
Quadro 11: Comparação de trabalhos relacionados com a presente pesquisa	73
Quadro 12: Perfis de evasão encontrados nos trabalhos relacionados e na presente pesquisa.	74

Lista de Siglas e Abreviações

ACM	Association for Computing Machinery
BPMN	Business Process Model and Notation
CEFET	Centro Federal de Educação Tecnológica
CHAID	Chi-squared Automatic Interaction Detector
CNE	Conselho Nacional de Educação
CTE	Career and Technical Education
EDM	Educational Data Mining
EIT	Escola Industrial e Técnica
ENADE	Exame Nacional de Desempenho dos Estudantes
ENEM	Exame Nacional do Ensino Médio
EPT	Educação Profissional e Tecnológica
ETF	Escola Técnica Federal
ETFC	Escola Técnica Federal de Campos
IBM	International Business Machines Corporation
IFCE	Instituto Federal de Educação, Ciência e Tecnologia do Ceará
IFF	Instituto Federal de Educação, Ciência e Tecnologia Fluminense
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
KDD	Knowledge Discovery in Databases
MEC	Ministério da Educação
NPI	Negros, Pardos e Índios
PCD	Pessoa com Deficiência
PNAES	Programa Nacional de Assistência Estudantil
PNE	Plano Nacional de Educação
PROUNI	Programa Universidade para Todos
SISU	Sistema de Seleção Unificada
SOFD	Success Or Failure Determiner
SUAP	Sistema Unificado de Administração Pública
SVM	Support Vector Machine
TVET	Technical Vocational Education and Training
UFRJ	Universidade Federal do Rio de Janeiro
VET	Vocational Education and Training

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 Contextualização.....	12
1.2 Problema	15
1.3 Motivação	17
1.4 Objetivos.....	17
1.4.1 Objetivo Geral.....	18
1.4.2 Objetivos Específicos.....	18
1.5 Organização do trabalho	18
2 REFERENCIAL TEÓRICO.....	19
2.1 Ensino Superior no Brasil	19
2.2 Evasão no Ensino Superior	21
2.3 Institutos Federais de Educação, Ciência e Tecnologia.....	23
2.3.1 Instituto Federal de Educação, Ciência e Tecnologia Fluminense.....	25
2.4 Descoberta de conhecimento em bases de dados.....	26
2.4.1 Pré-processamento	27
2.4.2 Processamento (mineração de dados)	29
2.4.3 Pós-processamento (interpretação e análise de resultados).....	31
3 METODOLOGIA.....	32
3.1 Ferramentas utilizadas	32
3.2 Classificação da pesquisa.....	32
3.3 Procedimentos técnicos.....	33
3.3.1 Obtenção e seleção de dados	34
3.3.2 Pré-processamento	34
3.3.3 Mineração de dados e interpretação e análise dos resultados	40
4 ESTUDO BIBLIOMÉTRICO	42
4.1 Utilização de palavras-chaves.....	42
4.2 Resultados encontrados.....	44
4.3 Trabalhos Relacionados	47
4.3.1 Trabalhos de instituições localizadas na América.....	47
4.3.2 Trabalhos de instituições localizadas na Europa.....	49
4.3.3 Trabalhos de instituições localizadas na África	50

4.3.4 Trabalhos de instituições localizadas na Ásia	50
5 RESULTADOS E DISCUSSÃO	54
5.1 Árvores de Decisão	55
5.1.1 Cursos de Bacharelado	55
5.1.2 Cursos de Licenciatura	59
5.1.3 Cursos de Tecnologia	64
5.2 Regras de Associação	68
5.3 Panorama geral de evasão nos cursos investigados	71
5.4 Comparações com Trabalhos Relacionados	71
6 CONCLUSÕES	75
6.1 Contribuições	75
6.1.1 Realização da bibliometria da área de estudo	75
6.1.2 Definição e análise de fatores possivelmente relacionados à evasão	76
6.1.3 Identificação de perfis de alunos evadidos por curso	76
6.2 Trabalhos futuros	77
REFERÊNCIAS BIBLIOGRÁFICAS	79

1 INTRODUÇÃO

1.1 Contextualização

O Brasil é um país que apesar de possuir muitas riquezas, é marcado por uma gama de problemas relacionados à desigualdade social. Essa desigualdade acaba sendo refletida em diferentes áreas, como por exemplo, segurança, saúde e educação – esta última diretamente relacionada ao presente trabalho.

De maneira geral, existe um consenso na sociedade que a qualidade do ensino ofertada nos cursos de ensino fundamental e médio reflete no desempenho dos estudantes quando estes realizam processos seletivos, como vestibular e ENEM. Alunos que receberam uma boa base educacional tendem a estar mais preparados dos não a receberam. Além disso, Manfredi (2002) destaca que a maior parte da população associa o nível de escolaridade com melhores empregos e obtenção de melhores posições na carreira profissional. De acordo com Cretton (2016), após o meado da década de 1990, houve um aumento significativo no número de matrículas em cursos de nível superior, o que levou à criação de novas instituições de ensino para atenderem à crescente demanda.

Nesse sentido, Neves e Martins (2016) destacam que a expansão e a qualificação do ensino superior é algo necessário para o desenvolvimento econômico e social no Brasil, de maneira duradoura e consistente. Além disso, os autores destacam que a taxa de matrícula em cursos de nível superior no Brasil é uma das mais baixas entre países de nível de desenvolvimento equivalente e que o desempenho geral do sistema é algo passível de diversos questionamentos.

Tendo em vista esse cenário, Lima e Zago (2018) destacam que o Estado vem investindo em diversas políticas públicas com o objetivo de ampliar o acesso ao ensino superior, como por exemplo, o Programa Universidade para Todos (Prouni), criado em 2005, o Sistema de Seleção Unificada (Sisu), criado em 2010, a Lei de Cotas, aprovada em 2012, entre outros.

Entretanto, apesar dos esforços que vem sendo feitos pelo Estado para ampliação do acesso à educação superior, cada vez mais tem se mostrado necessário o desenvolvimento de políticas que deem aos estudantes o suporte necessário para que estes se mantenham nos cursos de graduação, visto que a evasão escolar ainda representa um problema grave de expressividade no país. Definida pelo MEC (BRASIL, 1996) como a saída definitiva do

estudante do curso de origem sem concluí-lo, Martins *et al.* (2017) apontam que no caso das universidades públicas brasileiras, por exemplo, a evasão representa um desperdício significativo de recursos públicos, extremamente necessários em diversos setores da sociedade.

Mashiloane e Mchunu (2013) destacam que muitas instituições de ensino aplicam testes de seleção com o objetivo de garantir apenas o ingresso de estudantes com maior probabilidade de obterem sucesso. Porém, apesar de todos os esforços realizados pelas instituições, muitos dos alunos selecionados passam a ser considerados como estudantes em risco de evasão no decorrer dos cursos em que estão matriculados. Nesse sentido, diversos fatores podem influenciar a evasão, como por exemplo, questões pessoais, familiares, formação anterior, desempenho no curso em que o aluno está matriculado, entre outras.

Lopes (2006) aponta que muitas ações têm sido adotadas para conquistar novos estudantes, mas que pouco esforço tem sido feito para reter ou aumentar o nível de satisfação dos alunos atuais, fazendo com que a permanência destes seja, cada vez mais, uma preocupação compartilhada entre instituições. Desse modo, as taxas de evasão crescem na mesma medida em que surgem novas ofertas de cursos e instituições.

Nesse sentido, Demetriou e Schmitz-Sciborski (2011) destacam que as principais razões apontadas na literatura a respeito da evasão são: o despreparo acadêmico, uma vez que muitos estudantes ingressam no ensino superior sem terem a base educacional necessária para este nível; a falta de interação do estudante com sua vida acadêmica e instituição, que pode ser apoiada através de programas de enriquecimento curricular e de atividades extracurriculares; a falta de integração social com colegas, professores e funcionários da instituição; a necessidade de trabalhar enquanto realiza o curso; a dependência financeira para custear os estudos, como por exemplo, a realização de empréstimos; características sociodemográficas, como nível de educação dos pais, gênero, etnia, renda familiar e distância entre a instituição de ensino e residência do aluno, dentre outros.

De acordo com Lima e Zago (2018), anualmente são registradas mais de dois milhões de novas matrículas na educação superior brasileira, porém a quantidade de alunos concluintes por ano possui uma média bastante inferior, em torno de um milhão. Tendo em vista a média anual de alunos que não concluem os cursos nos quais estão matriculados, percebe-se que a evasão não é um caso isolado, se repetindo diversas vezes, nas mais variadas instituições. A repetição desse problema indica que podem existir padrões de comportamento que motivem a evasão. Considerando a quantidade de informações existentes referentes aos alunos, esses padrões podem estar escondidos em meio a inúmeros dados. Descobri-los

manualmente seria uma tarefa inviável e extremamente onerosa, tornando necessária a sua automatização. Isso pode ser realizado através da utilização de técnicas de mineração de dados (do inglês, *data mining*).

Devasia, Vinushree e Hegde (2016) definem mineração de dados como a área que analisa grandes repositórios de dados com o objetivo de extrair informações necessárias ou úteis. Além disso, os autores destacam que computadores são capazes de processar diferentes tipos de dados, como números, textos e imagens, de modo que a análise é realizada com base em padrões, associações e relações entre os dados, de modo a obter conhecimento útil. Desse modo, Fayyad *et al.* (1996) destacam a diferença entre a mineração de dados e o processo de Descoberta de Conhecimento em Bases de Dados (do Inglês *Knowledge Discovery in Databases* – KDD). De acordo com os autores, KDD se refere ao processo global de descoberta de conhecimento útil presente no meio de vários dados, ao passo que mineração de dados refere-se a uma etapa desse processo, onde é realizada a aplicação de algoritmos específicos para descoberta desse conhecimento.

De acordo com Romero e Ventura (2010), existe uma área de pesquisa interdisciplinar que lida com o desenvolvimento de métodos para explorar dados originados em um contexto educacional. Tal área denomina-se mineração de dados educacionais (*Educational Data Mining* – EDM).

Miller *et al.* (2015) destacam que a mineração de dados educacionais preocupa-se com o desenvolvimento de métodos para exploração dos tipos exclusivos de dados oriundos de ambientes educacionais, utilizando esses métodos para que seja obtido um melhor entendimento a respeito do comportamento dos alunos, a fim de otimizar o processo de ensino aprendizagem. De acordo com os autores, a mineração de dados educacionais abrange diversos paradigmas de pesquisa e ferramentas de análise, como, por exemplo, estatística educacional quantitativa, psicometria, modelagem computacional, mineração de dados e aprendizado de máquina.

Desse modo, Gualhano (2019) destaca que a mineração de dados educacionais trabalha com diferentes grupos que, sob diferentes perspectivas, analisam informações educacionais, conforme seus objetivos. A partir disso, o conhecimento descoberto por meio de algoritmos de mineração de dados educacionais pode ser utilizado para auxiliar os professores na compreensão dos processos de aprendizado de seus alunos e apoiar a reflexão sobre os métodos de ensino. Além disso, também apoia reflexões do próprio aluno no que tange a sua situação escolar.

Nesse contexto, observa-se que a evasão pode estar relacionada tanto a fatores pessoais como a fatores institucionais. Por esta razão, é necessário conhecê-los, para que as medidas de combate à evasão a serem adotadas sejam pensadas e discutidas de acordo com cada necessidade identificada. Por exemplo, casos em que a evasão ocorre devido a fatores pessoais podem ser minimizados através de políticas públicas ou com acompanhamento de determinados profissionais, como psicólogos. Por outro lado, os casos em que a evasão ocorre por problemas relacionados ao aprendizado dos alunos podem indicar a necessidade de revisão dos métodos de ensino utilizados. Desse modo, os problemas são identificados, permitindo a discussão de soluções adequadas para diferentes casos de evasão.

1.2 Problema

A evasão, independente do nível de escolaridade, representa um problema multiforme. Tal problema pode ser enxergado sob diferentes pontos de vistas, como, por exemplo, da instituição de ensino, independente desta ser pública ou privada, do mercado de trabalho e também do aluno.

Em relação às instituições de ensino privadas, do ponto de vista financeiro, o aluno pode ser considerado um cliente. A instituição presta um serviço ao aluno, que paga por esse serviço. A instituição depende desse recurso para manter seu funcionamento, que envolve custos como pagamento de funcionários, energia elétrica, entre outros. Se um aluno abandona um determinado curso, isso representa um cliente a menos, ou seja, menos recursos para a instituição. Altas taxas de evasão podem, por exemplo, inviabilizar cursos. Nesse sentido, Nunes (2005) destaca que o aluno matriculado (cliente adquirido) compra um produto representado por uma vaga em determinado curso, algo que somente estará concretizado por meio da conclusão deste. Caso esse aluno desista do curso (cliente perdido), ele deixa de pagar pelo produto, de modo que a universidade poderá repor a venda apenas por meio das modalidades de reingresso (cliente reconquistado) ou transferência (cliente adquirido, proveniente de outra instituição).

Por outro lado, o orçamento de muitas instituições públicas, como Universidades e Institutos Federais, é baseado no número de matrículas, não dependendo de mensalidades pagas pelos alunos. Assim, semelhantemente às instituições privadas, menos alunos representarão menos recursos. Além disso, existe todo o investimento que é feito para manter o aluno na instituição, como por exemplo, os processos de seleção. Desse modo, Cunha, Nascimento e Durso (2016) apontam que a evasão em instituições públicas representa uma

significativa perda para a sociedade, uma vez que gera desperdício de recursos que poderiam ser utilizados em outras atividades dentro da própria instituição. Quando um candidato aprovado no vestibular deixa de realizar sua matrícula, esse problema pode ser resolvido com as reclassificações. Entretanto, dependendo do momento em que um aluno abandona um curso, a vaga deixada por ele já não pode mais ser aproveitada.

Do ponto de vista do mercado de trabalho, Linke, Nogueira e Linke (2017) destacam que a evasão impacta na falta de mão de obra qualificada, atingindo diretamente a economia de uma sociedade. Em outras palavras, esse problema pode afetar o desenvolvimento de uma determinada região, ocasionando um problema econômico. Tal problema pode, por exemplo, tornar necessária a contratação de profissionais de outras cidades e regiões, trazendo reflexos para a sociedade que ali reside.

Do ponto de vista do aluno, o fato de não concluir um curso pode causar frustração e até desmotivação. Martins (2007) destaca que para as famílias e o acadêmico, a evasão representa um sonho não realizado, um ciclo que não se fechou, representando também desperdício de tempo e/ou de dinheiro. A autora destaca ainda que em uma sociedade cada vez mais dependente de recursos intelectuais, essa questão se torna uma grande preocupação. Nesse sentido, o conhecimento e a educação permitem que as pessoas consigam melhores condições de vida e de trabalho. Assim, o aluno evadido tende a sofrer com impactos em sua renda familiar, padrão de vida, entre outros, o que futuramente pode gerar prejuízos. Desse modo, são observados diversos motivos que apontam a necessidade de conhecer melhor as razões que levam os alunos a evadirem. Assim, podem ser identificadas maneiras para que esse problema seja trabalhado, a fim de reduzir as taxas de evasão, aumentando índices de permanência e êxito.

Assim, apesar da grande importância de programas de assistência estudantil para a permanência e êxito dos estudantes, é necessário que as instituições de ensino se atentem à qualidade do ensino oferecido. É de grande relevância ações que motivem a participação dos estudantes em atividades de ensino, pesquisa e extensão, no sentido de garantir condições de permanência e permitir que os estudantes obtenham sucesso no decorrer de sua trajetória de formação acadêmica e profissional (IFCE, 2017).

Conforme apontado por Saccaro *et al.* (2019), os estudos sobre evasão no ensino superior brasileiro são escassos quando comparados à literatura a respeito do tema nos Estados Unidos e Europa. Desse modo, Evangelista *et al.* (2019) destacam que tais estudos ainda necessitam de maiores esforços para que seja obtida uma compreensão do fenômeno, permitindo a identificação e mitigação de fatores que desencadeiam o abandono da graduação.

1.3 Motivação

A evasão é um grave problema, que atinge tanto instituições públicas como privadas. Como forma de visualizar essa problemática a nível nacional, o Quadro 1 apresenta a proporção de alunos que se desvincularam dos cursos nos quais estavam matriculados e o número total de matrículas no ensino superior brasileiro, durante os anos de 2013 a 2017.

Quadro 1: Número de alunos com matrículas desvinculadas em relação ao total de alunos matriculados

Ano	Matriculados	Matrículas Desvinculadas	Percentual de Matrículas Desvinculadas em Relação ao Número de Matriculados
2013	9.904.051	1.465.292	15%
2014	7.828.013	1.625.135	21%
2015	8.027.297	1.735.546	22%
2016	8.048.701	2.029.687	25%
2017	8.286.663	1.900.586	23%

Fonte: Elaborado pelo autor, a partir dos dados do Censo da Educação Superior (INEP – 2013 a 2017).

Nesse sentido, em relação ao período mencionado, pode ser observado que a evasão em cursos de nível superior no Brasil variou entre 15% e 25%, sendo que seu ápice ocorreu em 2016. Como forma de enfrentar esse problema, é necessário que estudos sejam realizados a nível local, uma vez que cada instituição possui particularidades que precisam ser conhecidas para que a questão seja trabalhada de maneira efetiva. Assim, por meio do presente trabalho, busca-se estudar a evasão nos cursos superiores do *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense, localizado no município de Campos dos Goytacazes, no estado do Rio de Janeiro. O referido município recebe estudantes de várias cidades vizinhas, fazendo com que seja considerado um polo universitário. Do mesmo modo, a instituição de ensino em que a presente pesquisa foi aplicada possui alunos residentes em várias cidades. Ações que visem reduzir as taxas de evasão na referida instituição são necessárias, impactando de forma positiva o desenvolvimento regional.

1.4 Objetivos

Nesta seção, são apresentados o objetivo geral e os objetivos específicos do presente trabalho.

1.4.1 Objetivo Geral

O presente trabalho visa identificar características de alunos evadidos em diferentes cursos de nível superior, utilizando como estudo de caso o *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense.

1.4.2 Objetivos Específicos

Além do objetivo geral, podem ser apontados como objetivos específicos da presente pesquisa:

- a) Realizar a bibliometria da área de estudo;
- b) Definir e analisar fatores que podem estar relacionados à evasão;
- c) Identificar, para cada curso investigado, perfis de alunos evadidos (evasão de curso).

1.5 Organização do trabalho

O presente trabalho possui mais cinco capítulos, além desta Introdução.

O Capítulo 2 apresenta o referencial teórico, destacando conceitos importantes relacionados ao tema da presente pesquisa.

O Capítulo 3 apresenta a metodologia utilizada no desenvolvimento da presente pesquisa.

O Capítulo 4 apresenta o estudo bibliométrico e os trabalhos relacionados, destacando artigos que possuem grande relevância para o tema do presente trabalho.

O Capítulo 5 faz uma apresentação dos resultados encontrados e apresenta as discussões dos mesmos.

O Capítulo 6 relata as conclusões obtidas através da etapa de interpretação dos resultados alcançados.

2 REFERENCIAL TEÓRICO

2.1 Ensino Superior no Brasil

Gilioli (2016) destaca que o Brasil possui uma tradição universitária que em termos históricos pode ser considerada recente. De acordo com o autor, as universidades brasileiras passaram a se configurar de fato a partir da década de 1930 e foi a partir desse momento que houve uma ampliação na quantidade de instituições de ensino superior, ocorrendo em 1968 o marco da Reforma Universitária. O autor aponta ainda que essa reforma foi responsável por dar feições contemporâneas ao sistema universitário brasileiro. Como destaque, é citado o crescimento da rede privada e a Carta Magna de 1988, que elevou a autonomia universitária a princípio constitucional. Outras mudanças ocorreram somente na primeira década do século XXI, principalmente relacionadas à expansão da rede federal pública de instituições de ensino superior e à introdução de novos meios de seleção.

Em 2014 surgiu o Plano Nacional de Educação (PNE), que é uma lei responsável por estabelecer diretrizes, metas e estratégias relacionadas à educação no Brasil em um horizonte de dez anos. Essas metas e estratégias abrangem todos os níveis, modalidades e etapas educacionais, desde a Educação Infantil até a Pós Graduação. O PNE está relacionado à Constituição Federal de 1988, no sentido de que o Brasil possui a obrigação de planejar o futuro do ensino, objetivando oferecer educação com qualidade para toda a população brasileira. Em consonância com esse objetivo, os municípios, estados e Distrito Federal devem aprovar planos compreendendo suas realidades, porém orientados ao PNE (PNE, 2019).

O PNE possui vinte metas relacionadas a diferentes níveis de ensino. O Quadro 2 apresenta os dados relacionados à meta 12 – Educação Superior. Visto que o referido plano se encontra em execução, são apresentados os dados disponibilizados para essa meta, que correspondem à porcentagem já alcançada em três períodos: 2014 (ano de início do PNE), 2019 (ano de realização da presente pesquisa) e 2024 (horizonte de dez anos definido pelo PNE).

Quadro 2: Metas do Plano Nacional de Educação (PNE)

Meta	Situação em 2014	Situação em 2019	Meta para 2024
Porcentagem de matrículas na Educação Superior em relação à população de 18 a 24 anos	34,2%	34,6%	50%
Porcentagem de matrículas da população de 18 a 24 anos na Educação Superior	17,7%	18,1%	33%
Porcentagem de matrículas novas na Rede Pública em relação ao total de matrículas novas na Educação Superior	5,5%	7,7%	40%

Fonte: Observatório do Plano Nacional de Educação (PNE) (adaptado).

Lima e Zago (2018) destacam que para que essas metas sejam alcançadas, diversas políticas públicas de ampliação de acesso e permanência no ensino superior têm sido foco de investimento pelo Estado. Como exemplos, são citadas o Programa Universidade para Todos (Prouni), criado em 2005, o Programa Nacional de Assistência Estudantil (PNAES), criado em 2010, o Sistema de Seleção Unificada (Sisu), criado em 2010, entre outras. Os autores destacam ainda que essas políticas permitiram o ingresso de muitos estudantes com carência socioeconômica, por meio de notas obtidas no Exame Nacional do Ensino Médio (ENEM). Apesar dessas políticas terem sido criadas antes do PNE, pode-se observar que as mesmas vão de encontro com as metas traçadas, sendo meios utilizados com o objetivo de que elas sejam alcançadas.

Nesse sentido, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) realiza anualmente o Censo da Educação Superior, um instrumento de pesquisa que reúne informações sobre as instituições de ensino superior, como por exemplo, seus cursos (de graduação presencial ou a distância), vagas oferecidas, inscrições, matrículas, ingressantes, concluintes, entre outras informações (BRASIL, 2019). De acordo com os dados do Censo da Educação Superior, o número de estudantes matriculados nesse nível de ensino era de 5.302.373 em 2007 e 8.290.911 em 2017, o que representa um aumento de 56,4% para o referido período, com uma média de crescimento anual de 4,6%. Em relação às instituições de ensino superior, no ano de 2017, 87,9% eram privadas, enquanto apenas 12,1% eram públicas.

Sobre esse aspecto, Silva e Santos (2017) destacam que no Brasil o acesso ao ensino superior por meio da oferta de vagas em instituições públicas de ensino ainda é algo limitado, apesar do acesso à educação ser um direito constitucional. Isso acontece porque as instituições possuem limites de infraestrutura, o que as impedem de atender grande parte da demanda existente.

2.2 Evasão no Ensino Superior

Hoffmann *et al.* (2019) apontam que com o processo de ampliação de vagas no ensino superior brasileiro, ingressar no mesmo se tornou algo relativamente mais fácil. Além disso, destacam que a realização de um curso superior representa, em muitos casos, o atendimento às expectativas pessoais e profissionais. Apesar da grande expectativa por parte dos estudantes, os autores destacam que o abandono e o trancamento de matrícula nos primeiros anos do ensino superior é algo comum de acontecer. Outra questão apontada pelos autores é que as Instituições de Ensino Superior potencializam o desenvolvimento econômico e social, onde se busca formar não somente profissionais com qualidade técnica desejada, mas também em quantidade suficiente para atender às demandas do mercado de trabalho. Nesse sentido, é importante conhecer as causas que levam os alunos a evadirem, a fim de que essas taxas possam ser reduzidas por meio de ações de permanência e êxito.

Desse modo, Silva e Santos (2017) definem evasão como a saída definitiva do estudante do seu curso, sendo um problema de nível internacional que gera problemas tanto para instituições públicas como para instituições privadas. Para o MEC (BRASIL, 1996), a evasão é definida como a saída definitiva do estudante do curso de origem sem concluí-lo. Entretanto, essa definição contém limitações, uma vez que um mesmo estudante pode evadir de um curso e se tornar concluinte em outro. Com o objetivo de esclarecer limitações como esta, o MEC (BRASIL, 1996) propõe as seguintes formas de evasão:

- a) evasão de curso, quando o estudante se desliga do curso superior em situações como abandono e transferência;
- b) evasão da instituição, quando o estudante se desliga da instituição em que está matriculado;
- c) evasão do sistema, quando o estudante abandona o ensino superior de forma definitiva ou temporária.

Essas formas de evasão são classificadas por Gilioli (2016) em três modalidades: microevasão (evasão de curso), mesoevasão (evasão da instituição) e macroevasão (evasão da educação superior). Além dessas, Lima e Zago (2018) apontam a nanoevasão, que ocorre quando o estudante pode permanecer no mesmo curso e instituição, porém, transfere-se para outro campus, turno, modalidade de ensino (presencial ou ensino a distância) ou formação (bacharelado ou licenciatura).

De acordo com Prestes e Fialho (2018), a evasão causa prejuízos que atingem indivíduos, organizações e sociedade, de modo que há mais de cinquenta anos as políticas

públicas educacionais de diferentes países passaram a incluir a diminuição da evasão escolar em seus propósitos e metas. Hoffmann *et al.* (2019) destacam que o aluno evadido representa um desperdício direto, pois este utilizou recursos providos pela instituição de ensino (como materiais, professores, infraestrutura, entre outros) sem que o objetivo de concluir o curso tenha sido alcançado.

Silva Filho *et al.* (2007) destacam que a evasão no ensino superior é um problema de nível internacional que afeta o resultado dos sistemas educacionais, refletindo desperdícios sociais, acadêmicos e econômicos. O autor destaca que no setor público, os recursos públicos investidos não têm o devido retorno, enquanto no setor privado, representa perda de receitas.

Em relação ao Brasil, Prestes e Fialho (2018) destacam que a evasão acompanha a história educacional. Isso faz dela um dos principais pontos de Programas e Planos Educacionais, visando a implantação de estratégias de monitoramento e de prevenção e buscando meios de estimular a permanência e êxito dos alunos matriculados. Nesse sentido, Hoffmann *et al.* (2019) apontam a necessidade que as instituições de ensino superior possuem em conhecer seus indicadores de evasão. Análises desses indicadores permitem identificar causas ou segmentos de alunos que possuem taxas de evasão elevadas. Desse modo, é possível elaborar políticas, atividades e programas tendo como objetivo a permanência e êxito dos estudantes.

Como uma medida essencial de reduzir a evasão, Gilioli (2016) aponta a necessidade do planejamento estratégico no que diz respeito à oferta de cursos para uma determinada região, o que deve considerar se existe demanda local de trabalho suficiente. O autor destaca ainda que a identificação e o comprometimento do aluno com o curso escolhido são fatores decisivos em relação à evasão. Desse modo, se o curso não for a primeira opção do aluno, existe maior possibilidade de que a evasão ocorra.

Outro ponto de grande importância destacado pelo autor é que, apesar do governo federal promover programas de permanência e êxito para instituições de ensino superior, o sucesso dos mesmos depende de cada instituição, conforme sua autonomia. Em outras palavras, isso significa que iniciativas de nível nacional não são suficientes se não estiverem em consonância com as realidades locais onde cada instituição se encontra. O autor também destaca a necessidade de acompanhamento dos estudantes, para identificação de dificuldades, sejam elas acadêmicas, operacionais ou socioeconômicas. Além disso, destaca que ao tratar os desiguais de maneira igual, desigualdades previamente existentes acabam sendo reforçadas.

Castro e Teixeira (2013) destacam que a evasão pode estar relacionada tanto a fatores pessoais como a fatores institucionais. Nesse sentido, os autores defendem a ideia de que

mesmo que a universidade adote ações que visem amenizar a evasão, algumas questões podem extrapolar a sua capacidade de ação. Como exemplo, são citadas questões relacionadas à vocação dos estudantes, algo que é prévio ao ingresso na instituição. Assim, os autores apontam a necessidade de realização de intervenções que alcancem crianças e adolescentes na escola, tratando de questões como a importância de buscar informações sobre as profissões, além de promover reflexões sobre as características pessoais. Por outro lado, em relação à evasão ligada a fatores institucionais, é apontado que estes podem ser trabalhados pela instituição, como por exemplo, pontos relacionados à qualificação de professores e questões curriculares.

Oliveira Júnior (2015) destaca que uma possível detecção antecipada de quais alunos tendem a evadir pode fornecer informações que permitem tomadas de decisões pelos gestores educacionais visando reverter esta situação. De acordo com o autor, pesquisadores da área de Informática em Educação têm utilizado técnicas computacionais de mineração de dados com o objetivo de realizar essa detecção antecipada.

2.3 Institutos Federais de Educação, Ciência e Tecnologia

Os Institutos Federais de Educação, Ciência e Tecnologia foram criados em 2008 por meio da Lei nº. 11.892, durante o governo do então presidente Luiz Inácio Lula da Silva. Esta seção apresenta um pouco do contexto histórico dessas instituições, que foi iniciado em 1909, com a criação das Escolas de Aprendizes Artífices pelo então presidente Nilo Peçanha (BRASIL, 2011).

Em 1909 foi assinado pelo presidente Nilo Peçanha o decreto 7.566, que criou dezenove Escolas de Aprendizes Artífices. O objetivo dessas escolas era fornecer ensino profissional primário e gratuito para classes menos favorecidas. Em 1937 foi promulgada por Getúlio Vargas a Constituição que fez com que o ensino técnico passasse a ser contemplado como um elemento estratégico para o desenvolvimento da economia. Tal ato transformou as Escolas de Aprendizes Artífices em Liceus Industriais. Os Liceus trabalhavam em sintonia com o crescimento da indústria, de modo que para manter esse crescimento era necessário formar mão de obra qualificada. Em 1942, foi realizada uma reforma no sistema educacional brasileiro pelo então ministro da Educação e Saúde, Gustavo Capanema. Através dessa reforma, o ensino profissional e técnico foi equiparado ao nível médio, transformando os Liceus Industriais em Escolas Industriais e Técnicas (EITs). Tais escolas prevaleceram até

1959, quando as EITs foram transformadas em Escolas Técnicas Federais (ETFs), ganhando autonomia pedagógica e administrativa (BRASIL, 2011).

Nesse sentido, em 1961, foi fixado por lei as Diretrizes e Bases da Educação Nacional, equiparando o ensino profissional ao ensino acadêmico e fazendo com que a partir desse momento, o ensino profissional e técnico passasse a ser considerado essencial para a expansão da economia e não como um tipo de formação destinada apenas a indivíduos carentes, como vinha sendo considerada até então. Com a aceleração do crescimento econômico, houve forte expansão da oferta de ensino técnico e profissional, de modo que em 1978, surgiram os três primeiros Centros Federais de Educação Tecnológica (CEFETs), com o objetivo de formar engenheiros de operação e tecnólogos (BRASIL, 2011).

Em 1994, a Lei Federal nº 8.948, de 8 de dezembro, estabeleceu a transformação gradativa das Escolas Técnicas Federais (ETFs) em Centros Federais de Educação Tecnológica (CEFETs), mediante decreto específico para cada instituição. Assim, os CEFETs absorveram as atividades das ETFs e das Escolas Agrotécnicas Federais, se transformando na unidade padrão da Rede Federal de Ensino Profissional, Científico e Tecnológico. Em 2008, por meio da Lei nº. 11.892, o sistema foi reorganizado com a criação dos Institutos Federais de Educação, Ciência e Tecnologia, absorvendo os CEFETs e as Escolas Técnicas remanescentes (BRASIL, 2011).

De acordo com o MEC (2019a), os Institutos Federais são instituições, pluricurriculares e multicampi, possuindo em sua estrutura Reitoria, *campus*, *campus* avançado, Polos de Inovação e Polos de Educação a Distância. São especializados na oferta de Educação Profissional e Tecnológica (EPT) em todos os seus níveis e formas de articulação com os demais níveis e modalidades da Educação Nacional. Ofertam diferentes tipos de cursos de EPT, além de licenciaturas, bacharelados e pós-graduação *stricto sensu*. Foram instituídos no momento de constituição da Rede Federal e possuem como obrigatoriedade legal garantir um mínimo de 50% de suas vagas para a oferta de cursos técnicos de nível médio, prioritariamente na forma integrada.

Além disso, tais instituições devem garantir o mínimo de 20% de suas vagas para atender a oferta de cursos de licenciatura, bem como programas especiais de formação pedagógica, visando a formação de professores para a educação básica, especialmente nas áreas de Ciências e Matemática, e para a educação profissional. Também atua no desenvolvimento de soluções técnicas e tecnológicas através de pesquisas aplicadas e as ações de extensão junto à comunidade, visando o avanço econômico e social local e regional.

Atualmente, existem 38 Institutos Federais com mais de 661 unidades estão vinculadas a eles (MEC, 2019a).

É importante, porém, destacar que existe uma diferença entre a realidade brasileira e a realidade de outros países. De acordo com o Conselho Nacional de Educação (CNE), a Educação e Formação Profissional é um termo abrangente que se aproxima de denominações em inglês como *Vocational Education and Training* (VET), *Technical Vocational Education and Training* (TVET) ou ainda *Career and Technical Education* (CTE). Assim, o estudo realizado pelo CNE esclarece que no Brasil existe uma diferença conceitual entre formação vocacional e formação profissional, de modo que a formação vocacional é mais voltada para o comportamento do aluno, enquanto a formação profissional abrange um aspecto mais técnico (BRASIL, 2015).

Entretanto, o estudo aponta ainda que em diversos países essa diferença não é tão notada, de modo que os programas VET são mais direcionados para o ensino básico ou educação obrigatória como uma oportunidade de fornecer aos adolescentes conhecimentos práticos para ingressarem no mercado de trabalho. Em alguns países, esse processo tem início no ensino fundamental, enquanto que em outros está integrado ao ensino médio, sendo o termo VET também aplicado para os cursos técnicos pós-secundários. Já os termos TVET e CTE são mais utilizados no ensino superior, estando relacionados aos estudos mais aprofundados das tecnologias e ciências afins.

2.3.1 Instituto Federal de Educação, Ciência e Tecnologia Fluminense

A criação das Escolas de Aprendizes e Artífices em 1909 consistiu, a princípio, na implantação de escolas nas capitais dos Estados, por estas serem cidades com maior capacidade de absorção de mão de obra e destino de pessoas que buscavam novas alternativas de empregabilidade nos espaços urbanos. Porém, no Estado do Rio de Janeiro, a escola não foi instalada na capital e sim na cidade de Campos dos Goytacazes. Entrou em funcionamento em janeiro de 1910, sendo a nona a ser criada no Brasil, com cinco cursos: alfaiataria, marcenaria, tornearia, sapataria e eletricidade (IFF, 2015a).

Em 1945, a Escola de Aprendizes Artífices de Campos passou a ser denominada Escola Técnica Federal de Campos (ETFC) e, como as demais, atrela-se às políticas de desenvolvimento, com interesse voltado para o crescimento e consolidação da indústria (IFF, 2018).

Na década de 60, foram criados os cursos de edificações, eletrotécnica e mecânica de máquinas na Escola Técnica Federal de Campos. Posteriormente, na década de 70, a instituição passou a oferecer também o curso de química, voltado para a indústria açucareira, uma indústria muito importante para a economia da cidade na época. Em 1974, a ETFC passou a oferecer somente cursos técnicos em seu currículo oficial, pondo fim às antigas oficinas. No mesmo ano, a Petrobras anuncia a descoberta de campos de petróleo no litoral norte do estado do Rio de Janeiro. Desse modo, a ETFC passou a ser a principal formadora de mão de obra para as empresas que operavam na bacia de Campos. Na década de 90, as Escolas Técnicas Federais são transformadas em Centros Federais de Educação Tecnológica, de modo que a Escola Técnica Federal de Campos se transformou no Centro Federal de Educação Tecnológica de Campos (IFF, 2015a).

Em 1999, o CEFET foi autorizado a oferecer cursos em nível de terceiro grau. Em outubro de 2004, sob decretos assinados pelo então Presidente Luís Inácio Lula da Silva, o CEFET Campos passou ao *status* de Centro Universitário. Assim, além do ensino médio e técnico, passou a oferecer cursos superiores de Automação, Manutenção Industrial, Indústria do Petróleo e Gás, Desenvolvimento de Software, Design Gráfico, Geografia, Matemática, Arquitetura e Ciências da Natureza nas modalidades: Química, Física e Biologia, além de três pós-graduações *lato sensu*: Educação Ambiental, Produção e Sistemas e Literatura, Memória Cultural Sociedade e um curso de mestrado em Engenharia de Meio Ambiente (IFF, 2015b).

Devido ao seu movimento de territorialização, atualmente o Instituto Federal de Educação, Ciência e Tecnologia Fluminense encontra-se em vários municípios, possuindo treze *campi*, um Polo de Inovação, um Centro de Referência em Tecnologia, Informação e Comunicação na Educação e a Reitoria. Tais municípios são: Bom Jesus do Itabapoana, Itaperuna, Cambuci e Santo Antônio de Pádua na região Noroeste Fluminense; Campos dos Goytacazes, São João da Barra, Quissamã e Macaé na região Norte Fluminense; na região das Baixadas Litorâneas, o de Cabo Frio; e os municípios de Itaboraí e Maricá na região Metropolitana (IFF, 2015b).

2.4 Descoberta de conhecimento em bases de dados

De acordo com Chen *et al.* (2012), o desenvolvimento da internet durante a década de 1970 e a ampliação da *World Wide Web* (WWW) em larga escala durante a década de 1990 impulsionaram de forma exponencial a criação e a coleta de dados de negócio. Desse modo, Assis (2017) aponta que devido a esse aumento no volume de dados armazenados, surgiu a

necessidade de técnicas que fossem capazes de extrair informações potencialmente úteis em meio à massa de dados. Oliveira Júnior (2015) destaca que essas técnicas compõem o assunto do campo do KDD. Além deste, outro conceito muito importante relacionado à extração de conhecimento em meio a diversas informações é a mineração de dados (do inglês *Data Mining*).

Oliveira Júnior (2015) destaca que em muitas vezes, o termo “Mineração de Dados” é utilizado como sinônimo de KDD, o que pode ser considerado um contexto amplo, tendo em vista as diferenças entre esses dois conceitos. Nesse sentido, Fayyad *et al.* (1996) apontam que KDD se refere ao processo global de descoberta de conhecimento útil em meio a diversos dados, enquanto mineração de dados se refere a uma etapa específica desse processo. Desse modo, os autores definem o KDD como um processo não trivial que busca identificar padrões válidos, potencialmente úteis e compreensíveis em meio à massa de dados. Já a mineração de dados é definida pelos autores como a aplicação de algoritmos específicos, que possuem como objetivo encontrar padrões em meio aos dados.

As etapas que compõem o processo de KDD podem ser agrupadas da seguinte forma: pré-processamento, envolvendo as etapas de seleção, limpeza, pré-processamento e transformação de dados; processamento, envolvendo a etapa de mineração de dados; e pós-processamento, envolvendo a etapa de interpretação e análise dos resultados, representadas na Figura 1.



Figura 1: Etapas do processo de KDD.

Fonte: Elaborado pelo autor, adaptado de Fayyad *et al.* (1996).

2.4.1 Pré-processamento

Nesta seção são apresentadas as etapas de seleção, limpeza, pré-processamento e transformação de dados.

2.4.1.1 Seleção de dados

De modo geral, uma base com dados brutos é o ponto de partida do processo de KDD. Tais bases possuem inúmeras informações, que precisam ser trabalhadas para que o conhecimento possa ser extraído das mesmas. Desse modo, Han, Kamber e Pei (2012) definem a etapa de seleção como o momento em que dados relevantes para a tarefa de análise são recuperados da base. Isso permite reduções significativas de grandes massas de dados, extraindo informações que sejam realmente importantes de acordo com o objetivo de um determinado estudo.

2.4.1.2 Limpeza

Dentre os processos constantes no KDD, a limpeza é o que demanda mais tempo para ser executado, devido à sua natureza, que envolve considerável trabalho manual. Nesse sentido, Han, Kamber e Pei (2012) destacam que dados reais tendem a ser incompletos, inconsistentes e com algum tipo de ruído. Isso acontece porque, por exemplo, uma mesma informação pode ser escrita de formas diferentes por diferentes usuários. Assim, os autores apontam que a limpeza de dados busca preencher valores que estejam ausentes, identificar discrepâncias a fim de reduzir ruídos e corrigir inconsistências presentes nos dados. Tais procedimentos são de suma importância, uma vez que esses problemas tornam os dados imprecisos, comprometendo o sucesso das etapas posteriores, como por exemplo, as etapas de mineração de dados e interpretação de resultados. Em outras palavras, o processo de limpeza exerce grande influência sobre os resultados, de modo que uma boa execução do mesmo está relacionada ao cumprimento (ou não) do objetivo de um determinado estudo.

2.4.1.3 Transformação de dados

Além da seleção e da limpeza, existe a transformação, outra etapa necessária a fim de preparar os dados antes que neles sejam aplicados algoritmos de mineração. Moscoso-Zea, Vizcaino e Mora (2017) definem a transformação de dados como a etapa na qual são criados conjuntos de dados com as variáveis necessárias, visando a redução da complexidade da análise. Nesta etapa, os dados sofrem modificações a fim de permitir que os resultados da aplicação dos algoritmos sejam interpretáveis, além de buscar melhorias no desempenho desses algoritmos. Um exemplo seria transformar valores numéricos em faixas de valores,

reduzindo significativamente o volume dessas informações e contribuindo com os objetivos dessa etapa.

2.4.2 Processamento (mineração de dados)

Bases de dados podem conter informações valiosas, porém como tais bases geralmente possuem um volume muito grande de informações, analisá-las manualmente seria uma tarefa árdua e ineficiente. Assim, grandes volumes de dados podem possuir padrões que dificilmente seriam identificados em uma análise manual e mesmo que fossem percebidos, o desgaste para encontrá-los seria muito grande. Por essa razão, a tarefa de busca por padrões em dados precisa ser realizada de forma automatizada, por meio da mineração de dados.

Fayyad *et al.* (1996) apontam que a mineração de dados corresponde a uma etapa do processo de KDD, onde ocorre a aplicação de algoritmos que buscam encontrar padrões em meio à massa de dados. Essa etapa equivale à fase do processamento. Nesse sentido, Lorenzetti e Telöcken (2016) destacam que a mineração de dados é composta por diversas técnicas (algoritmos) criadas para facilitar a descoberta de algo novo e que possa contribuir no conhecimento a partir de milhares de informações. Desse modo, os algoritmos são classificados de acordo com a tarefa que eles executam. Alguns exemplos de tarefas são a classificação e a associação.

2.4.2.1 Classificação

De acordo com Ahmed e Elaraby (2014), as técnicas de classificação empregam um conjunto de exemplos pré-classificados (aprendizado supervisionado) para desenvolver um modelo e a partir deste, classificar os registros de uma determinada base de dados. Assim, a classificação envolve as etapas de aprendizado e classificação. Na aprendizagem, os dados utilizados para treinar o modelo são analisados pelo algoritmo. Já na etapa de classificação, os dados da fase de teste são utilizados para estimar a precisão das regras de classificação. Um exemplo de técnica de classificação são as Árvores de Decisão.

2.4.2.1.1 Árvores de decisão

Daí e Ji (2014) apontam que as árvores de decisão são uma técnica de classificação em mineração de dados onde cada árvore é direcionada a partir de um nó raiz. Esse nó não possui

nenhum meio de entrada, o que o difere dos demais, conhecidos como nós de decisão. Os autores destacam que no estágio de treinamento, cada nó interno divide o espaço da instância em duas ou mais partes com o objetivo de otimizar o desempenho do classificador. Depois disso, todo caminho do nó raiz ao nó folha forma uma regra de decisão para determinar a qual classe uma nova instância pertence.

Nesse sentido, Nandeshwar (2009) definem árvore de decisão como uma coleção de nós, ramos e folhas. De acordo com os autores, as árvores de decisão trabalham com a abordagem da divisão e conquista, onde cada nó é dividido, usando determinados critérios, até que os dados sejam classificados para atender a uma condição de parada.

Quinlan (1987) destaca ainda que uma árvore de decisão é uma estrutura recursiva simples, que visa expressar um processo de classificação sequencial para um determinado caso, descrito por um conjunto de atributos. Desse modo, cada folha da árvore indica uma classe e cada nó representa um teste em um ou mais dos atributos, havendo uma árvore de decisão para cada resultado possível do teste. Para classificar um caso, o processo é iniciado na raiz da árvore. Se for uma folha, o caso é atribuído para a classe indicada; se for um teste, o resultado para este caso é determinado e o processo continua com a árvore apropriada para esse resultado. Tais procedimentos são realizados, por exemplo, pelo algoritmo J48.

A Figura 2 apresenta um exemplo de árvore de decisão, destacando os elementos mencionados (nó raiz, ramos, nós de decisão e folhas).

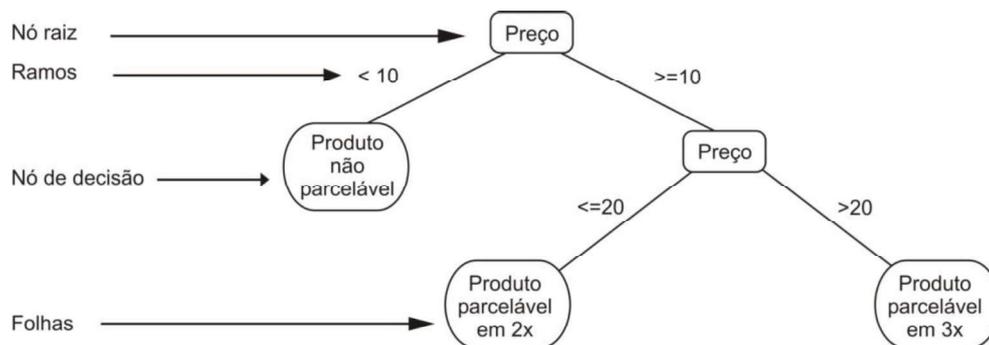


Figura 2: Exemplo de árvore de decisão.
Fonte: Elaborado pelo autor.

No exemplo ilustrado, caso o preço de um determinado produto seja inferior a R\$ 10,00 o mesmo não poderá ser parcelado. Porém, um produto que custe entre R\$ 10,00 e 20,00 pode ser parcelado em duas vezes, enquanto um produto com valor superior a R\$ 20,00 pode ser parcelado em três vezes.

2.4.2.3 Associação

De acordo com Raval (2012), na associação, um padrão é descoberto com base em um relacionamento de um item específico com outros itens na mesma transação. Desse modo, a associação é utilizada, por exemplo, para identificar quais produtos clientes frequentemente compram juntos. Um exemplo de algoritmo de associação é o *Apriori*.

De acordo com Herpich *et al.* (2016), o algoritmo *Apriori* é utilizado para extrair regras de associações entre os itens das relações de conjuntos de dados. Tais regras realizam a análise da base de dados, buscando encontrar relacionamentos entre uma variável e outra, representando-as da seguinte maneira: X (antecedente) \Rightarrow (implica em) Y (consequente). Por exemplo, {café, pão} \Rightarrow {manteiga} é uma regra de associação, onde os clientes que compram juntamente os produtos café e pão possuem a tendência de também comprar manteiga.

De Amo (2004) aponta que esse conhecimento pode ser útil em diversas tarefas, como planejamento de catálogo de produtos, folhetos de promoções, campanhas de publicidade, melhor organização da localização de produtos em prateleiras, entre outros. Além de ser útil para lojas físicas, esse conhecimento é frequentemente utilizado por lojas virtuais, fazendo com que esta técnica esteja presente nas sugestões oferecidas por sites de compras, aplicativos de músicas, entre outros. Tais sugestões se baseiam no conteúdo que um usuário viu ou buscou, visando apresentar a ele conteúdo que seja de seu interesse.

2.4.3 Pós-processamento (interpretação e análise de resultados)

A última etapa do processo de KDD corresponde à interpretação e análise dos resultados. De acordo com Fayyad *et al.* (1996), nessa etapa os padrões encontrados são analisados e interpretados, a fim de que sejam determinados quais deles podem ser considerados conhecimento novo e útil.

Desse modo, essa análise faz com que o conhecimento descoberto por algoritmos de mineração de dados se torne conhecido por *stakeholders*, auxiliando o processo de tomada de decisão, podendo indicar, por exemplo, elaboração de novas políticas, modificação de políticas existentes, entre outros.

3 METODOLOGIA

O presente trabalho foi realizado em cinco etapas, a saber: Pesquisa Bibliométrica, Extração de Dados do Sistema Acadêmico, Pré-Processamento, Mineração de Dados, Interpretação e Análise dos Resultados, conforme apresentado na Figura 3.



Figura 3: Etapas da metodologia.
Fonte: Elaborado pelo autor.

3.1 Ferramentas utilizadas

Para a realização das etapas do presente trabalho, houve a necessidade de utilização de duas ferramentas. A primeira ferramenta utilizada foi *LibreOffice Calc*, um aplicativo de planilhas utilizado para calcular, analisar e gerenciar dados (LIBREOFFICE, 2011). Tal ferramenta foi utilizada para realizar as etapas de pré-processamento e transformação de dados exportados por meio do Sistema Acadêmico, preparando-os para a etapa de mineração de dados.

Já a etapa de mineração de dados foi realizada utilizando o *Orange*, uma ferramenta de código aberto para mineração de dados, aprendizado de máquina e visualização de dados, desenvolvido na Faculdade de Computação e Ciência da Informação da Universidade de Liubliana, na Eslovênia. A ferramenta funciona por meio de scripts Python e programação visual, com fluxos interativos para análise de dados (DEMSAR *et al.*, 2013). De acordo com Antunes *et al.* (2018), esses fluxos são montados no ambiente de programação visual do sistema, por meio de blocos de construção chamados *widgets*. Os *widgets* são adicionados e interligados por meio da interface gráfica, permitindo, assim, a criação dos fluxos. Dessa forma, a ferramenta foi selecionada para o presente trabalho, pois, além da programação visual facilitar o processo de mineração de dados, as várias formas de visualização dos dados disponibilizadas pela ferramenta auxiliam na análise dos mesmos.

3.2 Classificação da pesquisa

De acordo com Silva e Menezes (2005) e Gil (2008), a pesquisa pode ser classificada em relação à natureza, forma de abordagem do problema, objetivos e procedimentos técnicos.

Em relação à natureza, o presente trabalho pode ser classificado como pesquisa aplicada, pois seu foco consiste na aplicação de conhecimentos existentes em uma situação concreta. Em relação à forma de abordagem do problema, pode ser classificado como pesquisa quantitativa, pois envolve a utilização de técnicas estatísticas. Exemplos de utilização dessas técnicas podem ser observados na análise de desempenho dos algoritmos de mineração de dados, como porcentagem de acertos e de erros. Em relação aos objetivos, o presente trabalho pode ser classificado como pesquisa descritiva, pois visa descrever as características de uma determinada população. No caso da presente pesquisa, trata-se de uma população de alunos evadidos. Em relação aos procedimentos técnicos, pode ser classificada como pesquisa experimental, pois após determinação do objeto de estudo, é possível selecionar variáveis que podem exercer alguma influência sobre ele e realizar observações a respeito do quanto essa influência pode ou não ocorrer.

3.3 Procedimentos técnicos

Neste trabalho, é utilizado o processo de KDD que, conforme já mencionado, é composto por cinco etapas: obtenção e seleção de dados, pré-processamento dos dados, transformação de dados, mineração de dados (*data mining*) e interpretação e análise dos resultados. A Figura 4 apresenta as tarefas realizadas durante o desenvolvimento do método proposto através da notação BPMN (*Business Process Model and Notation*).

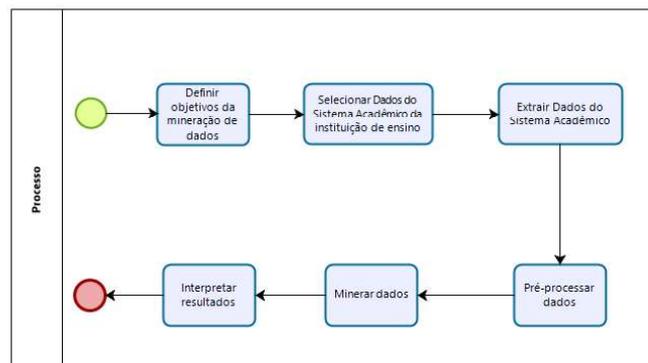


Figura 4: Tarefas realizadas para desenvolvimento do método proposto.

Fonte: Elaborado pelo autor

A seguir são apresentados os procedimentos realizados durante essas etapas para o Estudo de Caso onde esse método foi aplicado, a fim de ser verificada a sua aplicabilidade. Tal método foi realizado nos cursos de nível superior oferecidos pelo *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense. O referido *campus* situa-

se no município de Campos dos Goytacazes, no norte do estado do Rio de Janeiro, sendo o maior *campus* do Instituto, com aproximadamente 7.000 alunos em cursos que vão desde a educação básica até a pós-graduação.

3.3.1 Obtenção e seleção de dados

Esta é a etapa inicial do processo de KDD, onde os dados são obtidos e analisados, visando identificar quais deles serão selecionados. Essa seleção pode variar de acordo com o objetivo da mineração de dados a ser realizada. Por essa razão, os objetivos são definidos e a partir desse ponto, os dados são selecionados e extraídos do Sistema Acadêmico.

Nesse sentido, para realização da presente pesquisa, foram exportadas quatro bases de dados por meio do sistema acadêmico utilizado na instituição de ensino onde este estudo foi aplicado. Cada uma dessas bases era referente a um semestre: o segundo semestre de 2017, o primeiro e o segundo semestre de 2018 e o primeiro semestre de 2019. A justificativa para escolha desse período se dá porque o mesmo corresponde ao período da gestão atual, quando também ocorreu a criação de políticas afirmativas realizadas pela Diretoria de Gestão Acadêmica, incluindo ações para permanência e êxito dos estudantes.

Inicialmente, cada base possuía 132 atributos referentes a diversas informações dos alunos. Algumas dessas informações eram pessoais, como estado civil e data de nascimento. Outras eram relacionadas ao desempenho dos estudantes, como por exemplo, o Coeficiente de Rendimento. O Quadro 3 apresenta a quantidade de registros presentes em cada uma das bases trabalhadas.

Quadro 3: Quantidade de registros por semestre

Semestre	Quantidade de registros
2017-2	3.260
2018-1	3.430
2018-2	3.556
2019-1	3.700

Fonte: Elaborado pelo autor

3.3.2 Pré-processamento

Após a obtenção e seleção dos dados, foi realizado o pré-processamento e a transformação dos mesmos, uma vez que eles não se encontravam adequados para que a etapa de mineração fosse iniciada. Os dados foram analisados com o objetivo de identificar quais inconsistências os mesmos possuíam, permitindo assim, identificar quais procedimentos

seriam necessários para corrigi-los. Dessa forma, eles foram organizados acrescentando o atributo “Semestre” em cada uma das bases (fazendo com que, por exemplo, todos os registros da base 2017-2 passassem a ter o atributo “Semestre” preenchido com o valor “2017.2”). Além disso, também foi realizada a unificação das quatro bases em uma única, com o objetivo de facilitar a realização dos procedimentos referentes às etapas de limpeza e transformação de dados de todos os semestres trabalhados.

Caso essa unificação não fosse feita, toda vez que fosse necessário realizar um procedimento de ajuste em alguma das bases, o mesmo teria de ser replicado nas outras três. Dessa forma, a unificação das bases foi realizada com o objetivo de otimizar esses procedimentos. Após a criação do atributo “Semestre” e a unificação, passou-se a ter uma base com 13.946 registros e 133 atributos.

3.3.2.1 Limpeza de atributos não relevantes

Foi realizada a limpeza de atributos que estavam vazios em todos os registros, uma vez que por não conterem nenhuma informação, não possuíam nenhuma relevância para a etapa de mineração de dados. Foram excluídos doze atributos nessa situação, como por exemplo, atributos referentes ao número de registro de registro de diploma, número da folha em que o diploma foi registrado, entre outros. Após a limpeza desses atributos, a base passou a ter 121 atributos.

O procedimento seguinte foi realizar a limpeza dos atributos que possuíam o mesmo valor em todos os registros. Uma vez que um atributo possui essa característica, ele também se torna irrelevante para a etapa de mineração de dados, visto que todos os registros possuem a mesma informação. Como exemplos, podem ser citados o atributo “Desc_Instituicao”, uma vez que seu valor sempre era “Campus Campos Centro” e o atributo “Desc_Regime_Internato”, que sempre possuía a informação “Não Internos”. Após a limpeza dos atributos nessa condição, a base passou a ter 112 atributos.

Dando continuidade à etapa de pré-processamento dos dados, foi observada a existência de diferentes atributos relacionados a uma mesma informação. Por exemplo, foi observado que existiam alguns atributos que armazenavam códigos e outros que armazenavam as descrições correspondentes a esses códigos. Nesse sentido, foram identificadas quais eram as informações a que esses atributos faziam referência e ao invés de manter todos eles, passou-se a manter apenas um, aquele que fosse mais adequado. Por exemplo, o atributo “Cod_Forma_Ingresso_Matricula” foi excluído, mas o atributo

“Desc_Forma_Ingresso_Matricula” foi mantido. Após a realização desses procedimentos de limpeza, a base passou a ter 85 atributos.

Posteriormente, foi realizada uma análise dos atributos restantes, buscando verificar se estes possuíam relação com a evasão. Foi observado que muitos deles não possuíam nenhuma relação com o objetivo da presente pesquisa, como por exemplo, Diploma_Pasta, Ano_Reservista, Regiao_Militar, Possui_Foto, Secao_Ele, entre outros. Por estes serem atributos irrelevantes em relação à evasão, os mesmos foram excluídos, fazendo com que a base passasse a ter 37 atributos.

3.3.2.2 Limpeza de registros

Para descobrir os perfis de alunos evadidos, optou-se por limpar os registros que representavam situações em andamento, ou seja, situações que ainda não tiveram um desfecho (uma exceção é a situação “Falecido”). Nesse sentido, foram excluídos os registros dos alunos com os seguintes valores no campo Desc_Sit_matricula: “Aguardando ENADE”, “Falecido”, “Trancado”, “Matriculado”, “Transferido Externo” e “Transferido Interno”. Por outro lado, foram mantidos os registros em que atributo “Desc_Sit_matricula” representava desfechos de sucesso ou insucesso. Os valores mantidos foram: “Egresso”, “Concluído”, “Cancelado” e “Evasão”. Cabe destacar que os desfechos “Egresso”, “Concluído” representam casos de sucesso, enquanto os desfechos “Cancelado” e “Evasão” representam casos de insucesso.

Após a limpeza dos registros com situações em andamento, a base passou a ter 1.767 registros.

3.3.2.2.1 Limpeza de registros repetidos – casos de insucesso

Após a realização dos procedimentos de limpeza de registros, alguns registros permaneceram com o atributo “Situação” vazio. Esses registros foram excluídos, pois se tratavam de alunos com o atributo “Desc_Sit_matricula” correspondente à evasão em semestres em que esta não ocorreu, mas por esta ser a situação atual da matrícula. Se esses registros fossem mantidos, um mesmo aluno que evadiu seria contado em diferentes semestres.

Após a realização do procedimento de limpeza dos registros que se encontravam nessa condição, a base passou a ter a quantidade de 1.252 registros.

3.3.2.2.2 Limpeza de registros repetidos – casos de sucesso

Foi observado que nos registros cujo valor do atributo “Situação” era “Concluinte” também existiam alunos que se repetiam. Isso aconteceu porque o atributo “Desc_Sit_matricula” possuía a situação atual da matrícula dos alunos (nesse caso, “Egresso” e “Concluído”), repetindo a situação da matrícula em diferentes semestres. Para corrigir esse problema, foram selecionados todos os concluintes e após isso, foram excluídos os registros em que o atributo “Desc_Sit_matricula_Periodo” correspondia aos valores “Aprov. c/Dependência”, “Matriculado” e “Reprovado”, sendo mantido apenas o valor “Aprovado”, que representa o último semestre, quando o aluno de fato concluiu o curso.

Após a realização do procedimento de limpeza dos registros que se encontravam nessa condição, a base passou a ter 1.022 registros.

3.3.2.2.3 Limpeza de registros referentes a cursos sem alunos concluintes

Foi observado que os cursos de Licenciatura em Teatro, Licenciatura em Educação Física, Bacharelado em Engenharia Elétrica e Bacharelado em Engenharia de Computação não possuíam alunos concluintes, tendo apenas alunos evadidos. Essa situação ocorreu porque nesses cursos ainda não houve formação da primeira turma. Esse tipo de situação impede que algoritmo utilizado consiga classificar os dados, uma vez que existe apenas uma classe. Por essa razão, os registros referentes aos alunos desse curso foram excluídos. Após esse procedimento de limpeza, a base de dados passou a ter 932 registros.

3.3.2.3 Realização de ajustes

A análise dos dados apontou a necessidade de realização de alguns ajustes a fim de adequá-los para posteriormente serem minerados. Um exemplo de ajuste identificado foi a necessidade de realizar a substituição das vírgulas presentes nos valores dos atributos “Coeficiente_Rendimento” e “Percentual_Frequencia” por ponto final. Isso foi necessário para padronizar os números de acordo com o idioma Inglês, idioma padrão da ferramenta *Orange*, utilizada no presente trabalho. Dessa forma, os valores desses dois atributos puderam ser reconhecidos como números pela ferramenta.

3.3.2.4 Transformação de dados

Em alguns atributos foi realizada a tarefa de transformação, como por exemplo, a redução da quantidade de categorias presentes nos valores de um determinado atributo, de modo a melhorar o desempenho do algoritmo de classificação e facilitar a interpretação dos resultados. Por exemplo, o atributo “Desc_Cor” possuía os seguintes possíveis valores: “Amarela”, “Branca”, “Indígena”, “Não dispõe da informação”, “Não quis declarar cor/raça”, “Parda” e “Preta”. Esses valores foram reduzidos para: “Não dispõe da informação”, “NPI” (equivalente a negros, pardos e índios) e “Outros”, ou seja, os valores foram agrupados em dois casos: alunos que entraram por meio de cota devido à cor declarada (NPI) e alunos que não ingressaram por cota nesse quesito.

O mesmo procedimento também foi realizado no atributo “Desc_Forma_Ingresso_Matricula”, que entre outros possíveis valores, possuía diferentes tipos de cotas para as opções “Vestibular” e “Sisu”. Ao invés de manter todos os tipos de cotas, elas foram reduzidas para “Vestibular – Cota” e “Sisu – Cota”. Do mesmo modo, o atributo “Desc_Estado_Civil” também foi transformado, agrupando os valores “Divorciado” e “Separado” em uma mesma categoria. Por fim, valores vazios de diversos atributos foram substituídos por “Não Declarado”.

3.3.2.5 Criação de novos atributos

Além dos atributos já existentes, alguns novos atributos foram criados a partir destes. Estes novos atributos foram criados por dois motivos: pela motivação de investigar se eles teriam relação com a evasão e como uma forma de tratar atributos com muitos valores distintos, pois isso faria com que as árvores de decisão ficassem muito grandes, de modo que dificultaria a interpretação das mesmas. Por exemplo, foi criado o atributo “Idade”, para que na etapa de mineração de dados ele fosse utilizado no lugar do atributo “Dt_Nascimento”.

Além disso, com o objetivo de facilitar a interpretação de resultados, foi observada a necessidade de criar o atributo “Tipo_curso”, podendo ter os seguintes valores: “Bacharelado”, “Licenciatura” e “Tecnologia”. Uma vez que esse atributo foi criado, a informação constante no mesmo foi retirada da descrição do curso, que também foi abreviada, com o objetivo de facilitar a interpretação de resultados.

Após a inclusão dessas informações, a base passou a ter 42 atributos.

3.3.2.6 Inclusão do atributo classificador

Na etapa de mineração de dados, para realizar a tarefa de classificação, é necessário que exista um atributo que contenha a classe à qual cada registro pertence. Por essa razão, foi criado o atributo “Situação”, que pode armazenar apenas um dos seguintes valores: “Evadido” ou “Concluente”. Essas são as classes utilizadas na etapa de mineração de dados.

A classe “Evadido” é referente aos casos de insucesso, ou seja, alunos com situação de “Cancelado” ou “Evasão”. Por outro lado, a classe Concluente foi utilizada para tratar os casos de sucesso (alunos egressos e concluintes).

Após a inclusão do atributo “Situação”, que é o atributo classificador, a base passou a ter 43 atributos, conforme demonstrado no Quadro 4.

Quadro 4: Relação de atributos e seus respectivos domínios

Atributo	Domínio
Ano_Conclusao_1_grau	Número natural
ANO_CONCLUSAO_2_GRAU	Número natural
ano_letivo_ini	Número natural
clPeriodo_Let_ini	Número formatado (Ex.: 2017/2)
COD_NACIONALIDADE	Texto
Coeficiente_Rendimento	Número decimal
Desc_Area_Procedencia_Escola_Origem	[Rural Urbana]
Desc_Cor	[NPI Outros]
tipo_curso	[Bacharelado Licenciatura Tecnologia]
desc_curso	[Arq. e Urb. Ciênc. da Nat. D. Gráfico Eng. Cont. Aut. Geografia Letras Manut. Ind. Matemática Sist. de Inf. Sist. de Telecom.]
Desc_Estado_Civil	[Solteiro Casado Divorciado Outro Separado União Estável Viúvo]
Desc_Estado_Civil_Pais	[Solteiro Casado Divorciado Outro Separado União Estável Viúvo]
Desc_Forma_Ingresso_Matricula	[Portadores de Diploma Sisu Ampla Concorrência Sisu Cota Sisu PCD Transferência Externa Transferência Interna Vestibular Ampla Concorrência Vestibular Cota]
Desc_renda_familiar	[Até 1 salário 1 a 2 salários 2 a 3 salários 3 a 5 salários 5 a 10 salários 10 a 20 salários Mais de 20 salários Não declarado]
Desc_Renda_Per_Capita	[Menos de 1 salário mínimo 1 salário mínimo 2 salários mínimos 3 salários mínimos 4 salários mínimos De 5 a 6 salários mínimos De 7 a 10 salários mínimos Não Declarado]
Desc_Renda_Per_Capita_SIG	[RFP <= 0,5 SM 0,5 SM < RFP <= 1 SM 1 SM < RFP <= 1,5 SM 1,5 SM < RFP <= 2,5 SM 2,5 SM < RFP <= 3 SM RFP > 3 SM Não Declarado]
Desc_Sit_matricula	[Cancelado Concluído Egresso Evasão]
Desc_Sit_matricula_Periodo	[Aprovado Reprovado Egresso Evasão]

Atributo	Domínio
Desc_Tip_Forma_Ingresso_Matricula_Periodo	[Reabertura de matrícula Reintegração Renovação por Aprovação Renovação por Reprovação Seleção]
Desc_Tipo_Escola_Origem	[Privada Pública Municipal Pública Estadual Pública Federal]
Desc_Turno	[Integral Matutino Vespertino Noturno]
GRAU_PARENTESCO_RESPONSAVEL	[Avó Tia]
MAE_FALECIDA	[0 1]
N_FILHOS	[0 1 2 3 4]
necessidade_auditiva	[0 1]
necessidade_fisica	[0 1]
necessidade_mental	[0 1]
necessidade_multipla	[0 1]
necessidade_visual	[0 1]
outras_necessidades	[0 1]
PAI_FALECIDO	[0 1]
Percentual_Frequencia	Número decimal
Periodo	[0 1 2 3 4 5 6 7 8 9 10]
Periodo_Let	[1 2]
Periodo_letivo_ini	[1 2]
Sexo	[F M]
TIPO_RESPONSAVEL	[A M O P]
Dif_con_ens_medio_ini_grad	Número natural
ja_possuia_graduacao	[Sim Não]
ja_possuia_pos_graduacao	[Sim Não]
Idade	Número natural
Semestre	Número formatado (Ex.: 2017/2)
Situação	[Concluinte Evadido]

Fonte: Elaborado pelo autor.

3.3.2.7 Divisão da base unificada

Com o objetivo de descobrir se existiam diferenças nos perfis de alunos evadidos de cursos distintos, a base unificada foi dividida em dez bases menores, sendo cada uma delas referente a um determinado curso.

3.3.3 Mineração de dados e interpretação e análise dos resultados

Para as bases geradas, foram criados modelos onde foram utilizados o algoritmo C4.5 (*Decision Tree*) e o algoritmo *CN2 Rule Induction* na ferramenta *Orange*. A partir desses modelos, foi realizada a etapa de mineração de dados, através da qual foi possível identificar padrões relacionados à evasão em cada curso estudado. Para validá-los, foi utilizada a técnica de validação cruzada (*cross-validation*), que consiste na divisão do conjunto de dados em

subconjuntos, nos quais um é utilizado para treino e os demais para validação. No presente trabalho, o conjunto original de dados foi subdividido em dez subconjuntos.

No Capítulo 4 é apresentado o estudo bibliométrico realizado, enquanto no Capítulo 5 são apresentadas as árvores de decisão geradas e as regras de associação obtidas, bem como a interpretação e análise das mesmas.

4 ESTUDO BIBLIOMÉTRICO

Para a realização do presente trabalho, foi realizado um estudo bibliométrico utilizando a base *Scopus*. De acordo com Cretton (2016), esta é uma base de grande importância para o levantamento do conhecimento científico, sendo uma das bibliotecas de publicações mais volumosas disponíveis em âmbito acadêmico.

A busca realizada encontrou setenta publicações entre os anos de 2005 e 2019 que tratam da utilização de técnicas de mineração de dados na descoberta de perfis de alunos evadidos em cursos superiores. Dos resultados encontrados, sessenta e oito possuíam relação com o tema do presente trabalho, sendo que cinquenta e oito abordavam a questão da evasão em cursos de graduação presenciais, enquanto dez estavam relacionados a cursos à distância. Nesse sentido, foram considerados somente os trabalhos relacionados aos cursos presenciais, uma vez que a presente pesquisa também buscava analisar a evasão em cursos presenciais de graduação. Esta seção apresenta, além das palavras chaves utilizadas, dados estatísticos sobre os países, revistas e autores que possuem maior quantidade de publicações na área, assim como os artigos mais citados.

4.1 Utilização de palavras-chaves

Para a realização desse estudo bibliométrico, foram realizadas buscas na base *Scopus* utilizando três conceitos e seus respectivos tesouros. Os conceitos utilizados foram: mineração de dados (conceito A), evasão (conceito B) e ensino superior (conceito C). O Quadro 5 apresenta esses conceitos e os tesouros pesquisados.

Quadro 5: Conceitos e tesouros utilizados

Conceitos →	Conceito A: Mineração de Dados	Conceito B: Evasão	Conceito C: Ensino Superior
Tesouros →	<i>data mining</i> <i>machine learning</i>	<i>drop out</i> <i>drop-out</i> <i>dropout</i> <i>evasion</i> <i>quitter</i> <i>truancy</i>	<i>higher education</i> <i>higher learning</i> <i>tertiary education</i> <i>further education</i> <i>tertiary level</i> <i>university education</i> <i>university degree</i> <i>undergraduate program</i> <i>undergraduate degree</i> <i>undergraduate education</i> <i>degree course</i> <i>tertiary degree</i> <i>graduation course</i>

Fonte: Elaborado pelo autor.

Após a definição dos conceitos e tesouros, foram realizadas buscas na base *Scopus*, contemplando cada conceito separadamente. Em seguida, foram feitas buscas combinando os conceitos, de modo que fossem consideradas todas as formas possíveis de combinações entre eles. Em todas as buscas realizadas, foram utilizados como parâmetros de pesquisa o título, o resumo e as palavras-chaves das publicações.

A busca envolvendo os três conceitos trabalhados é apresentada através do Quadro 6.

Quadro 6: String de busca realizada na base Scopus utilizando os conceitos ABC

```
( TITLE-ABS-KEY ( "data mining" OR "machine learning" ) AND TITLE-ABS-KEY ( "drop out" OR "drop-out" OR dropout OR evasion OR quitter OR truancy ) AND TITLE-ABS-KEY ( "higher education" OR "higher learning" OR "tertiary education" OR "further education" OR "tertiary level" OR "university education" OR "university degree" OR "undergraduate program" OR "undergraduate degree" OR "undergraduate education" OR "degree course" OR "tertiary degree" OR "graduation course" ) ) AND PUBYEAR < 2020
```

Fonte: Elaborado pelo autor.

Com o objetivo de apresentar uma visão geral sobre a quantidade de publicações por tema, assim como a quantidade de publicações envolvendo mais de um tema em questão, foi elaborado o Diagrama de Venn presente na Figura 5.

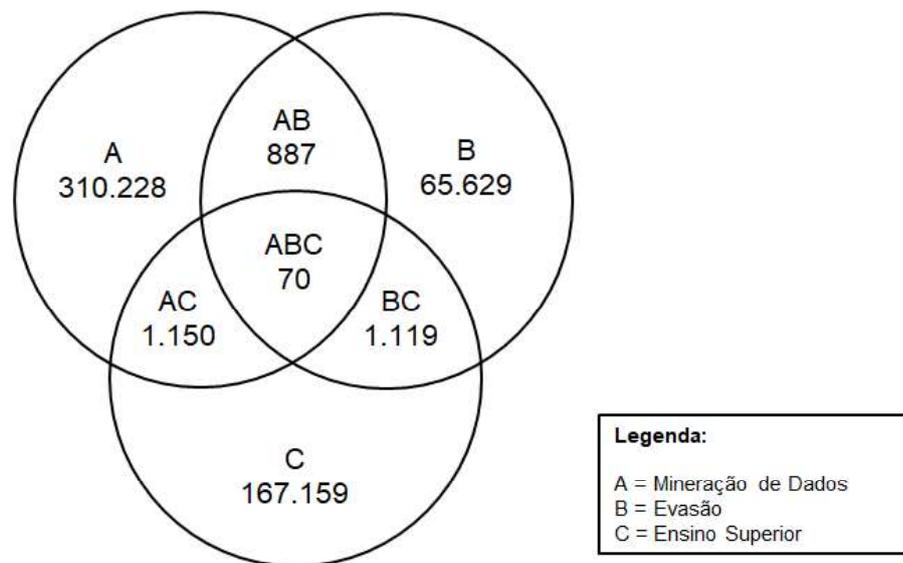


Figura 5: Diagrama de Venn representando a quantidade de publicações encontradas na base Scopus

Fonte: Elaborado pelo autor.

4.2 Resultados encontrados

Nesta seção, os resultados encontrados são apresentados em forma de gráficos, permitindo uma melhor visualização e análise das informações obtidas.

Na Figura 6, é apresentada a quantidade de publicações por ano, envolvendo a interseção dos três temas pesquisados, ou seja, mineração de dados, evasão e ensino superior. Analisando o gráfico, pode ser observado que 2005 foi o primeiro ano encontrado em que houve publicação envolvendo os temas relacionados. Pode-se observar também que, em relação a todos os anos, 2018 e 2019 foram os que possuíram o maior número de publicações, ultrapassando vinte em cada um desses anos.

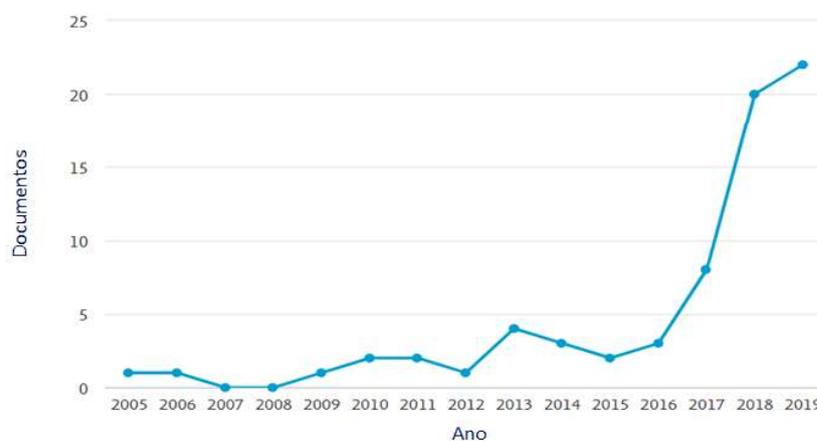


Figura 6: Quantidade de publicações por ano.

Fonte: Elaborado pelo autor.

A Figura 7 apresenta os artigos mais citados em relação ao tema pesquisado. Ayesha *et al.* (2010) aprezem em primeiro lugar, com quarenta e oito citações. O segundo lugar desse ranking é ocupado por Devasia, Vinushree e Hegde (2016), com trinta e nove citações. Na terceira posição, tem-se Hershkovitz e Nachmias (2011) e Hussain *et al.* (2018) com dezenove citações. Em termos proporcionais, pode ser observada que a maior diferença ocorre entre as três primeiras posições, sendo reduzida significativamente entre os demais autores.

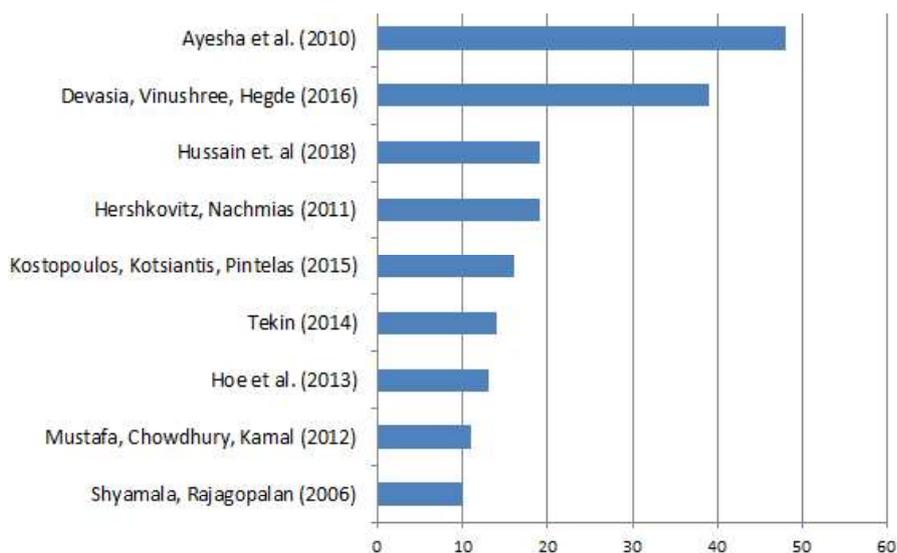


Figura 7: Artigos mais citados.
Fonte: Elaborado pelo autor.

A Figura 8 apresenta a quantidade de publicações por autor. Uma análise da mesma permite verificar que os seis primeiros colocados possuem duas publicações cada. Comparando esta com Figura 6 (publicações por ano), pode-se observar que o tema em questão representa uma área em crescimento, tendo sido mais explorada durante os anos de 2018 e 2019, porém com um número pequeno de publicações por autor. Um fato que corrobora com essa afirmação se dá ao comparar a Figura 7 (artigos mais citados) com a Figura 8 (quantidade de publicações por autor). Na Figura 7, pode ser observado que Ayesha *et al.* (2010) é a publicação mais citada, entretanto, a Figura 8 mostra que há apenas uma publicação de seus autores.

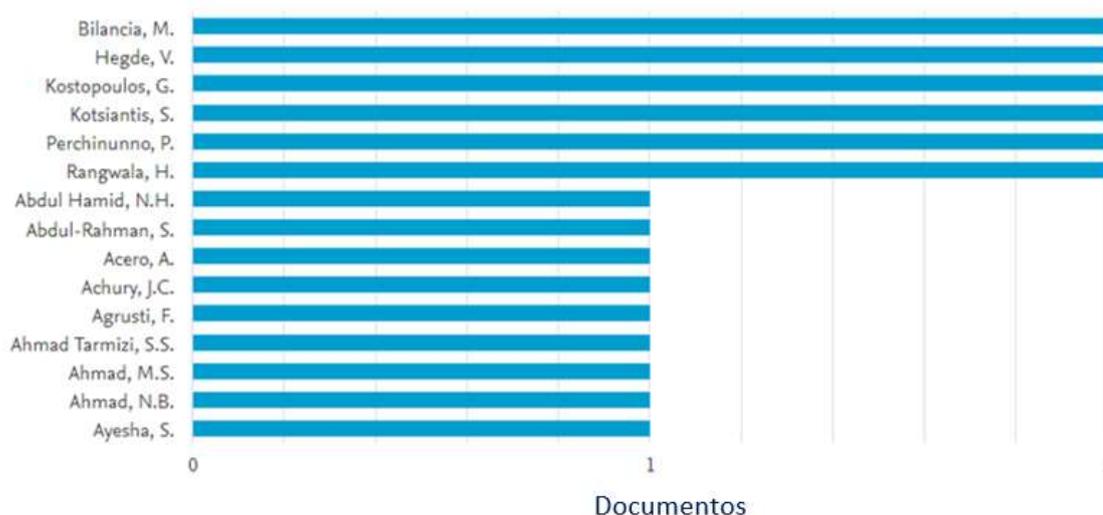


Figura 8: Quantidade de publicações por autor.
Fonte: Elaborado pelo autor.

A Figura 9 apresenta as fontes que mais publicaram sobre o tema. Analisando-o, pode ser observado que o *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lectures Notes in Bioinformatics* é a fonte com mais publicações sobre o tema. Foram cinco publicações, sendo uma em 2011, uma em 2013 e três em 2018. Em segundo lugar, tem-se o *ACM Internation Conference Proceeding Series*, com quatro publicações, sendo uma em 2015 e duas em 2018 e uma em 2019. O *Ceur Workshop Proceedings* ocupa o terceiro lugar, com três publicações: uma em 2017, uma em 2018 e uma em 2019, juntamente com o *Communications in Computer And Information Science*, que possui três publicações em 2019. Em quarto lugar tem-se o *Internation Journal Of Innovative Technology And Exploring Engineering*, com duas publicações em 2019.

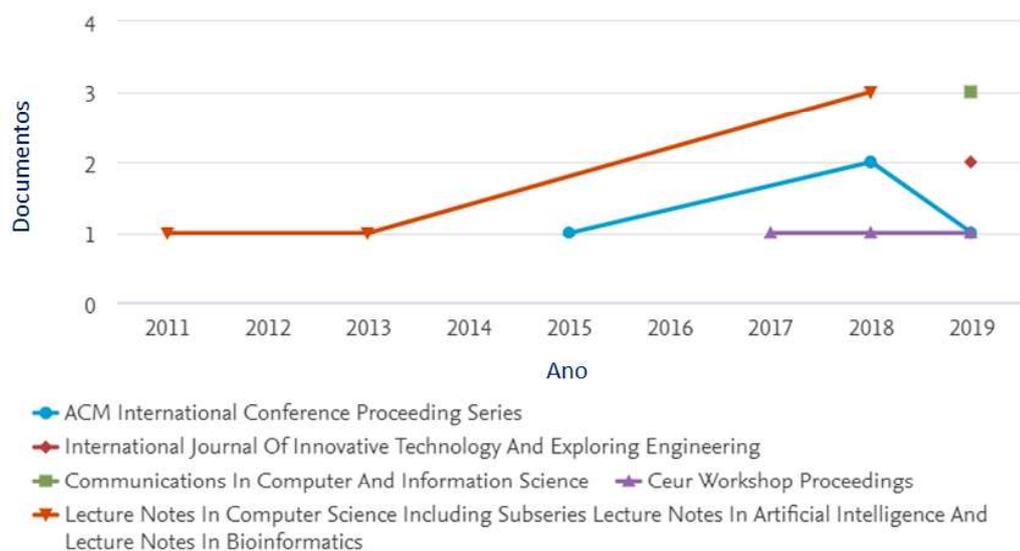


Figura 9: Fontes que mais publicaram sobre o tema.
Fonte: Elaborado pelo autor.

A Figura 10 apresenta os países que mais publicaram sobre o tema em questão. A Índia lidera o ranking, com nove artigos. O Brasil aparece em segundo lugar, empatado com os Estados Unidos, com seis publicações.

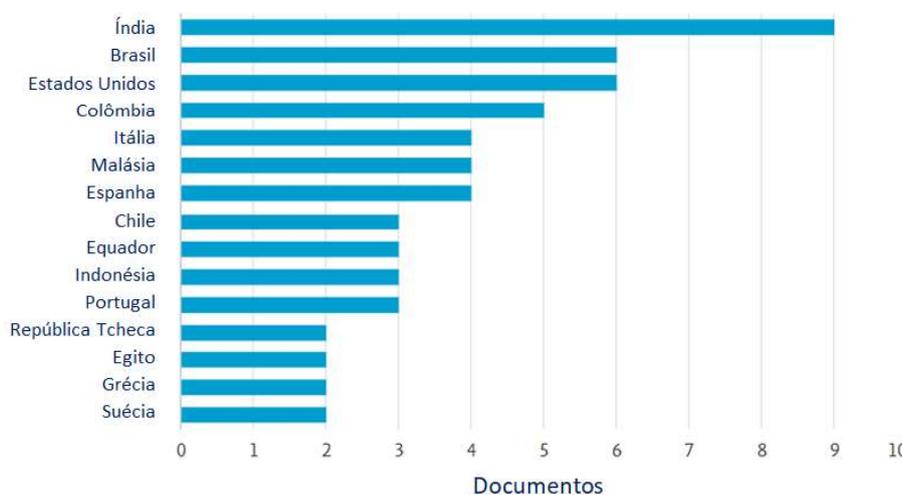


Figura 10: Países que mais publicaram sobre o tema.
Fonte: Elaborado pelo autor.

4.3 Trabalhos Relacionados

Esta seção apresenta trabalhos relacionados ao tema dessa pesquisa, ressaltando suas contribuições e estabelecendo um paralelo. Para tal, foram selecionados os trabalhos que, de acordo com a pesquisa realizada, possuíam forte relação com o presente trabalho.

4.3.1 Trabalhos de instituições localizadas na América

Manhães, Cruz e Zimbrão (2014) apontam que todos os anos, o sistema educacional brasileiro oferece um número crescente de vagas nas universidades públicas, motivado pela necessidade de preparar grandes quantidades de trabalhadores para se engajar na economia emergente do país. Além disso, os autores destacam que uma universidade pública federal brasileira é totalmente financiada pelo governo federal e que ser admitido em uma universidade pública é uma expectativa da maior parte dos jovens. No entanto, muitos alunos ingressam nessas instituições e não completam o curso no qual se matricularam, representando um elevado custo, sem retorno para o governo, sociedade e às instituições de ensino.

Nesse sentido, os autores analisaram dados de todos os estudantes da Universidade Federal do Rio de Janeiro, com matrículas realizadas nos anos de 2003 e 2004. Foram analisados dados de doze semestres, a partir do período de realização das matrículas, de modo

que o período investigado foi até o segundo semestre de 2010. Os autores estabeleceram três categorias de alunos (evadidos, matriculados e concluintes), a fim de investigar o desempenho desses alunos e as taxas de evasão. Foram analisados dados de cerca de 14.000 alunos, em 155 cursos de graduação, sendo possível encontrar uma precisão de mais de 80% em diferentes algoritmos de classificação (J48, SimpleCart, Support Vector Machine (SVM), Naïve Bayes e redes neurais (Multilayer Perceptron - MP).

Apesar de não ter obtido a melhor acurácia entre os algoritmos testados, o algoritmo Naïve Bayes foi utilizado para dar suporte à análise quantitativa. Isso se deu pelo fato de ser considerado de fácil interpretação por seres humanos e de fácil adaptação para o processo de visualização de dados. Assim, o trabalho permite a identificação dos alunos com maior probabilidade de desistirem, permitindo também que a universidade utilize não somente análises estatísticas para enfrentar o problema das altas taxas de evasão.

Moscoso-Zea, Vizcaino e Mora (2017) destacam que a aplicação de técnicas e algoritmos de mineração de dados permite que instituições de ensino compreendam de uma melhor maneira a forma que os professores ensinam e a forma que os alunos aprendem. De acordo com os autores, ela também permite uma melhor compreensão das atividades relacionadas aos processos organizacionais, de modo a aperfeiçoar o processo de tomada de decisão.

Nesse sentido, os autores buscam identificar quais os métodos e algoritmos de mineração de dados educacionais mais adequados para analisar as taxas de graduação e evasão de alunos. Para tal, foram analisados dados de alunos do curso de Ciência da Computação de uma determinada instituição privada, envolvendo alunos que ingressaram na universidade no primeiro semestre e alunos que ingressaram validando disciplinas. Foram utilizados cinco algoritmos: Naïve Bayes, Stacking, One R, J48 e Random Tree. De acordo com os autores, os algoritmos J48 e Random Tree apresentaram melhor precisão para classificar os estudantes em casos de sucesso e de evasão, permitindo que novas estratégias possam ser implementadas, a fim de melhorar os indicadores acadêmicos.

Balaniuk et al (2011) utilizaram técnicas de mineração de dados para detectar alunos em alto risco de evasão em uma instituição de ensino superior privada de Brasília. Foram analisados registros históricos de 11.495 alunos para construir um modelo de previsão que apontava, para cada estudante que concluiu o segundo semestre, a probabilidade deste se formar ou evadir. Os dados analisados combinavam as informações socioeconômicas e acadêmicas de cada aluno, como por exemplo, faixa etária, sexo, renda familiar, média geral de notas e frequência número de disciplinas reprovadas, entre outros. Foram utilizados 3.058

registros de alunos matriculados em 2005 e 2006 para treinar e testar o modelo, ao passo que 8.437 registros de alunos que ingressaram em 2010 foram utilizados para gerar as previsões. De acordo com o modelo construído, 3.250 dos 8.437 alunos matriculados iriam evadir, correspondendo a 38,5% dos alunos. Por se tratar de uma identificação precoce, os autores apontam que a universidade pode planejar estratégias para combater os fatores de riscos, sejam estes individuais ou que exijam mudanças institucionais.

Sánchez, Gutiérrez e Meneses (2005) utilizaram técnicas de mineração de dados (regras de associação e árvores de decisão) para estudar a evasão na *Universidad Católica del Norte*, no Chile. Para tanto, foram analisados dados de 4793 alunos que ingressaram na referida universidade entre os anos de 1995 e 2004. A evasão foi estudada nos cursos de Arquitetura, Engenharia de Computação, Direito, Jornalismo e Geologia, de modo que a mesma foi relacionada com variáveis de ingresso, como notas do ensino médio, nota no vestibular e colégio de origem.

4.3.2 Trabalhos de instituições localizadas na Europa

Bernardo *et al.* (2017) destacam que o desenvolvimento de um país é frequentemente impactado por taxas de empregabilidade e pela educação de nível superior, de modo que o crescimento das taxas de evasão representa uma desaceleração no avanço do desenvolvimento. Nesse sentido, os autores apontam que é necessário que as instituições de ensino estudem a fundo essa questão, a fim de elaborarem medidas adequadas para tratar o problema. Outro ponto que os autores destacam é que diversos estudos focam nas características de estudantes que evadem, desconsiderando diferentes tipos de abandono, como por exemplo, casos em que o estudante opta por seguir em outro curso.

Desse modo, o trabalho analisou registros de 1.311 alunos iniciantes em uma universidade no norte da Espanha. Os dados foram obtidos através de duas vertentes: através das informações dos alunos constantes no sistema utilizado na universidade e através da aplicação de questionários. As informações obtidas através do sistema consistiam em informações pessoais, sociodemográficas e acadêmicas. Já o questionário aplicado era utilizado para levantar informações sobre as decisões de abandono e quais as suas causas.

Assim, os autores identificaram cinco grupos: estudantes que persistem em seus cursos, estudantes que mudam de programa dentro da mesma universidade, estudantes que se transferem para outra universidade, estudantes que se matriculam em estudos que não são de nível superior e estudantes que desistem de estudar. Por fim, foram gerados modelos

preditivos que utilizam técnicas de mineração de dados para orientarem ações a serem adotadas como estratégias de intervenção, conforme a necessidade de cada grupo.

Em seu trabalho, Nagy e Molontay (2018) analisaram dados de 15.825 estudantes de graduação da Universidade de Tecnologia e Economia de Budapeste, matriculados entre 2010 e 2017. Foi realizada a previsão de desempenho futuro dos alunos, buscando identificar possíveis tendências de evasão. Para isso, os autores optaram por não utilizar dados referentes ao desempenho dos alunos no primeiro ano acadêmico, utilizando exclusivamente informações relativas ao ensino médio e dados pessoais, como sexo, idade, entre outros. Desse modo, os autores desenvolveram uma ferramenta web utilizando o algoritmo *Gradient Boosted Trees*, por meio da qual os alunos podem informar seus resultados do ensino médio e tal ferramenta, com base no conjunto de dados de treinamento, prevê se o aluno irá se formar ou não. Assim, o modelo pode ser usado para prever o status final dos novos alunos e para identificar possíveis situações de risco já no momento da inscrição dos estudantes, o que permite a recomendação de diferentes tipos de assistência, como por exemplo, aulas de reforço, entre outras.

4.3.3 Trabalhos de instituições localizadas na África

Mashiloane e Mchunu (2013) analisaram dados dos anos de 2009 a 2011 referentes aos de alunos do primeiro ano do curso de Ciência da Computação da Universidade de Witwatersrand, buscando realizar a predição de sucesso e falha dos alunos ao final do ano letivo. Foram utilizados três diferentes algoritmos: J48, Naive Bayes e Decision Table. Durante a fase de treino, o algoritmo J48 apresentou desempenho superior aos demais, sendo integrada à ferramenta *Success Or Failure Determiner (SOFD)*, desenvolvida pelos autores. Tal ferramenta permite que recursos sejam alocados onde eles são mais necessários, além de permitir a realização de intervenções de maneira precoce, a tempo de fornecer suporte e reverter a situação dos alunos em risco de evasão.

4.3.4 Trabalhos de instituições localizadas na Ásia

Hegde e Prageeth (2018) destacam que nos últimos anos, o número de alunos evadidos em diversas instituições de ensino tem crescido de maneira muito rápida, representando um sério problema para essas instituições. Os autores apontam ainda que o aluno, ao ingressar em um curso, possui vários sonhos e expectativas, que quando não satisfeitas, exercem um

impacto sobre o estudante, podendo fazer com o que o mesmo desista do curso que iniciou. Além disso, também é destacado que a decisão de evadir é influenciada tanto por fatores individuais como institucionais, de modo que a possibilidade de evasão, quando identificada em estágios iniciais de um curso, auxilia as instituições a manterem esses estudantes.

Nesse sentido, os autores propõem uma abordagem que busca combinar fatores como dados demográficos, desempenho acadêmico, problemas de saúde, local de residência, entre outros. Para tal, são aplicados questionários para alunos em diferentes estágios de uma determinada instituição, a fim de comparar o ponto de vista de novos alunos com os de alunos mais avançados. Essas informações são armazenadas em um banco de dados, pré-processadas e submetidas a técnicas de mineração de dados, com o objetivo de realizar predições sobre a possibilidade dos alunos evadirem.

Shyamala, Rajagopalan e Mohammed (2006) propõem um modelo que busca encontrar padrões para realizar predições a respeito do desempenho dos estudantes de uma universidade indiana. Tal modelo monitora o progresso dos estudantes através da obtenção de variáveis como a nota do semestre anterior, frequência, autoconfiança intelectual, tempo gasto em atividades em grupo, questões financeiras, nível de escolaridade dos pais, preferências acadêmicas, entre outras. Através de técnicas de mineração de dados, essas informações são utilizadas para identificar alunos que tendem a evadir. Desse modo, os autores buscam identificar o quanto a mineração de dados pode auxiliar na realização de intervenções com o objetivo de reduzir a evasão, bem como melhorar o processo de tomada de decisões na educação superior. Outra questão apontada pelos autores é que a influência exercida por grupos de amizade é um fator importante para prever o desempenho dos estudantes.

Devasia, Vinushree e Hegde (2016) analisaram dados de 700 alunos, da Amrita Vishwa Vidyapeetham, Mysuru, na Índia. Cada registro possuía 19 atributos e era referente ao curso de Ciência da Computação, de modo que foi analisado o período de 2013 a 2016. Até então o sistema presente na instituição permitia apenas o armazenamento e realização de consultas aos dados, sem nenhuma inteligência capaz de analisá-los. Assim, os autores desenvolveram um sistema web utilizando o algoritmo Naïve Bayes que, além de armazenar informações relacionadas à admissão, curso, notas, frequência, entre outros, utiliza o histórico acadêmico dos alunos para prever seus futuros desempenhos.

O Quadro 7 destaca características importantes de alguns dos trabalhos relacionados e faz uma comparação entre as mesmas.

Quadro 7: Comparação dos trabalhos relacionados

	Bernardo <i>et al.</i> (2017)	Manhães, Cruz e Zimbrão (2014)	Hegde e Prageeth (2018)	Shyamala, Rajagopalan e Mohammed (2006)	Moscoso- Zea, Vizcaino e Mora (2017)	Mashiloane e Mchunu (2013)
Quantidade de alunos	1.311	14.000	50	180	441	391
Período	2008 – 2012	2003 – 2010	Não informado	Não informado	2002 – 2015	2009 – 2011
Instituição	Universidade de Oviedo, Espanha	Universidade Federal do Rio de Janeiro (UFRJ), Brasil	Amrita School of Arts and Sciences, Índia	Dr. Ambedkar Government College, Índia	Não informado	University of The Witwatersrand, África do Sul
Ferramentas utilizadas	IBM Statistical Package for the Social Sciences	Weka	Weka	Clementine	Weka; Orange; Rapid Miner.	Weka
Algoritmos utilizados	<i>Exhaustive CHAID</i>	J48; SimpleCart; Support Vector Machine (SVM); Naïve Bayes (algoritmo selecionado); Multilayer Perceptron - MP.	Naïve Bayes	C5.0	Naïve Bayes; Stacking; One R; J48 (algoritmo selecionado); Random Tree.	J48 (algoritmo selecionado); Naïve Bayes; Decision Table.
Bases de dados consideradas	Base acadêmica; Questionários aplicados aos alunos.	Base acadêmica.	Base acadêmica; Questionários aplicados aos alunos.	Base acadêmica.	Base acadêmica.	Base acadêmica.

Fonte: Elaborado pelo autor.

O Quadro 8 apresenta os padrões relacionados à evasão encontrados nos trabalhos relacionados.

Quadro 8: Perfis de evasão encontrados nos trabalhos relacionados

	Bernardo <i>et al.</i> (2017)	Manhães, Cruz e Zimbrão (2014)	Hegde e Prageeth (2018)	Shyamala, Rajagopalan e Mohammed (2006)	Moscoso-Zea, Vizcaino e Mora (2017)	Mashiloane e Mchunu (2013)
Padrões encontrados de alunos evadidos	<ul style="list-style-type: none"> - Estavam na fase inicial do curso; - Possuíam idade igual ou superior a 20 anos no momento de ingresso no curso; - Declararam ter dedicado pouco tempo aos estudos durante a fase inicial do curso; - Moravam sozinhos, com parentes que não sejam os pais ou com parceiro(a). 	<ul style="list-style-type: none"> - Reduziram a quantidade de disciplinas em que estão inscritos a cada semestre; - Possuíram um número decrescente de disciplinas com notas de aprovação a cada semestre; - Possuíram, pelo menos, uma disciplina em que foram reprovados (incluindo também reprovação por falta) no primeiro semestre acadêmico; - Ao final do primeiro semestre, possuíam notas de aprovação inferiores às dos alunos concludentes; - Possuíam boas notas em poucas disciplinas. 	<ul style="list-style-type: none"> - Possuíram quatro reprovações em disciplinas; - Possuíram três reprovações por faltas; - Possuíam algum problema de saúde; - Não se adaptaram ao ambiente institucional. 	<p>Não foi apontado nenhum padrão de evasão. O modelo agrupou os estudantes de acordo com várias notas. Associou as notas do semestre anterior com notas do semestre em andamento para prever o resultado final. Apontou como fatores que influenciam a evasão: dificuldades acadêmicas, preferências acadêmicas, falta de ajuda financeira, frequência, entre outros.</p>	<ul style="list-style-type: none"> - Cursaram o ensino médio em escola pública; - Eram casados; - Possuíam bolsa de estudos no momento de ingresso na universidade, mas perderam a mesma no decorrer do curso. 	<p>Não foi apontado nenhum padrão de evasão. Foi desenvolvida a ferramenta <i>Success Or Failure Determiner</i> (SOFD), que realizou a predição da situação do aluno no final do ano acadêmico (sucesso ou insucesso) com base em sua nota no primeiro semestre.</p>

Fonte: Elaborado pelo autor.

5 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados e interpretados os resultados da etapa de mineração de dados do Estudo de Caso realizado na presente pesquisa. Tais resultados são apresentados por curso, organizados de acordo com o tipo do mesmo (Bacharelado, Licenciatura ou Tecnologia).

É importante destacar que de quatorze cursos superiores existentes na instituição onde esta pesquisa foi aplicada, quatro cursos não permitiram a obtenção de padrões. Tais cursos são: Licenciatura em Teatro, Licenciatura em Educação Física, Bacharelado em Engenharia Elétrica e Bacharelado em Engenharia de Computação. Essa impossibilidade se deu pelo fato de que esses cursos ainda não formaram a primeira turma, fazendo com que eles possuam alunos evadidos, mas não possuam alunos concluintes, as duas classes utilizadas na tarefa de classificação. Uma vez que alunos com situação como matriculado, por exemplo, foram excluídos (em virtude desta ser uma situação provisória e não final), as bases desses quatro cursos ficaram apenas com alunos evadidos. Havendo apenas uma classe, o algoritmo não consegue ser executado.

Nesse sentido, a fim de permitir uma visualização do cenário de evasão no período estudado (segundo semestre de 2017 até o primeiro semestre de 2019) é apresentado a Figura 11.

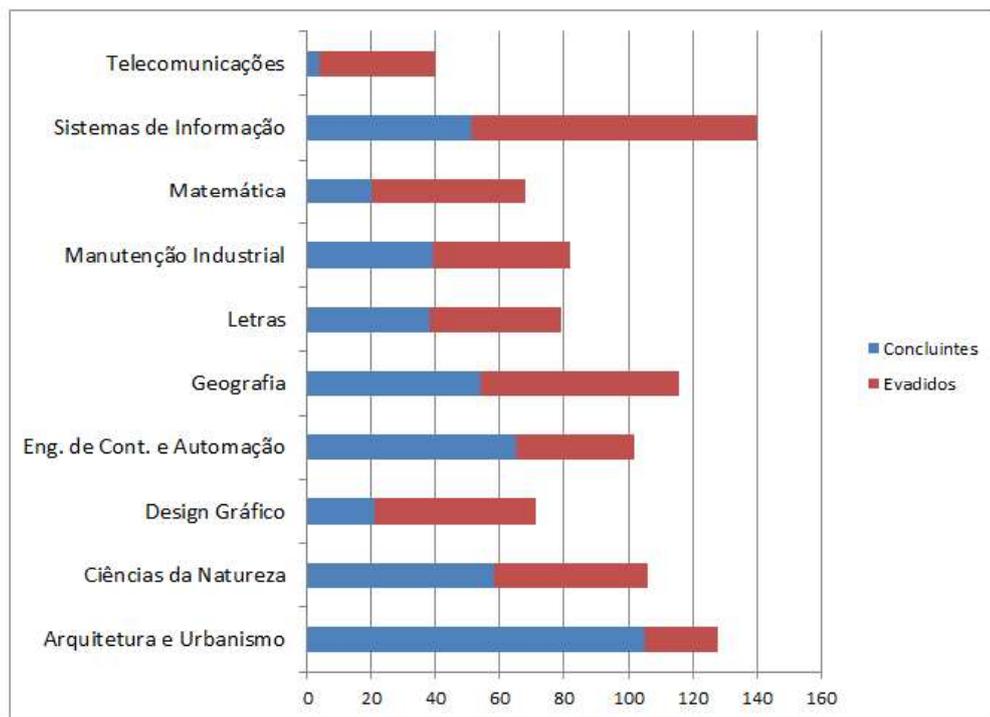


Figura 11: Demonstrativo de alunos concluintes e evadidos por curso no período 2017-2 a 2019-1.
 Fonte: Elaborado pelo autor.

Em relação à Figura 11, é importante destacar que sob o ponto de vista de conclusão e evasão, os cursos podem ser agrupados da seguinte maneira:

- a) Cursos com taxas de conclusão altamente superiores às taxas de evasão (Bacharelado em Arquitetura e Urbanismo);
- b) Cursos com taxas de conclusão consideravelmente superiores às taxas de evasão (Bacharelado em Engenharia de Controle e Automação);
- c) Cursos com taxas de conclusão um pouco superior à taxa de evasão (Licenciatura em Ciências da Natureza);
- d) Cursos com taxas de evasão um pouco superior à taxa de conclusão (Licenciatura em Geografia, Licenciatura em Letras e Tecnologia em Manutenção Industrial);
- e) Cursos com altas taxas de evasão (Tecnologia em Design Gráfico, Licenciatura em Matemática, Bacharelado em Sistemas de Informação e Tecnologia em Sistemas de Telecomunicações).

Esse agrupamento permite uma visão ampla sobre a evasão nos cursos estudados. Entretanto, para que políticas de combate à evasão, ou seja, de ações para permanência e êxito dos estudantes, possam ser adotadas, é necessário que haja maior detalhamento quanto ao perfil dos alunos evadidos em cada curso. As subseções a seguir trazem maiores detalhes a respeito dos perfis de alunos que evadiram nos dez cursos investigados.

5.1 Árvores de Decisão

5.1.1 Cursos de Bacharelado

A seguir são apresentadas as árvores de decisão geradas e as interpretações das mesmas para os cursos de Bacharelado. O número mínimo de objetos para a criação de um ramo na árvore foi de 10% do total de registros em cada curso.

5.1.1.1 Bacharelado em Arquitetura e Urbanismo

Em relação ao curso de Bacharelado em Arquitetura e Urbanismo é apresentada a árvore de decisão conforme a Figura 12.

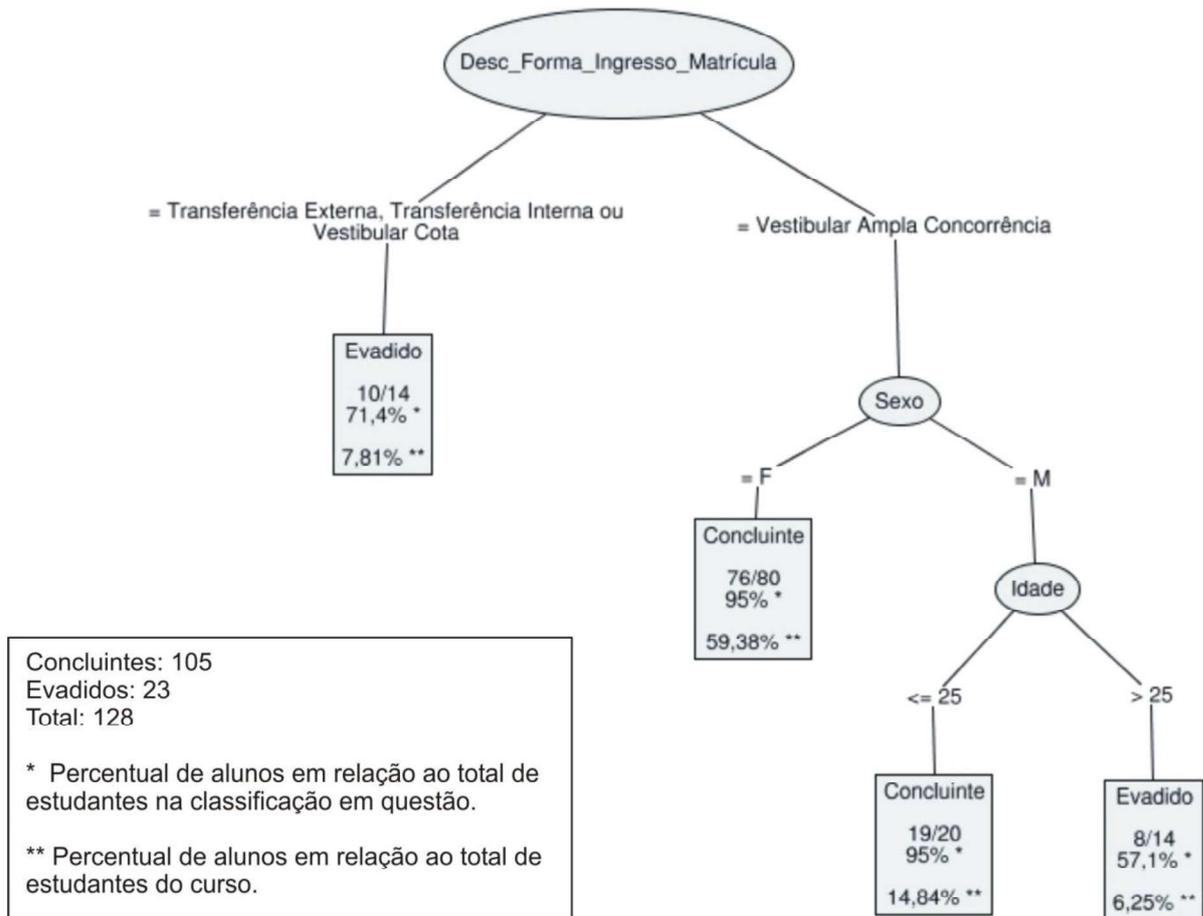


Figura 12: Árvore de deciso para o curso de Bacharelado em Arquitetura e Urbanismo

Fonte: Elaborado pelo autor.

A base do curso de Bacharelado em Arquitetura e Urbanismo possuía cento e vinte e oito registros, dos quais cento e cinco eram concluïntes, sendo esta a classe majoritria para esse curso. O percentual de concluïntes correspondeu a 82% do nmero total de registros na base. Para o referido curso, foram encontrados dois perfis de evaso: no primeiro, foi identificado que dez alunos, em um total de quatorze que ingressaram por meio de Transferncia Externa, Transferncia Interna ou Vestibular (Cota) evadiram, correspondendo a 71,4% de evaso para esse perfil e 7,81% em relao ao total de registros da base. J no segundo perfil de evaso encontrado, oito de quatorze alunos evadiram, correspondendo a 57,1% para esse perfil, que  composto por alunos com as seguintes característïcas: idade superior a 25 anos, do sexo masculino e que ingressaram pela modalidade de Vestibular (Ampla Concorrncia).

No caso do curso de Arquitetura e Urbanismo,  importante destacar que, apesar do resultado obtido, o nmero de alunos evadidos  significativamente inferior ao total de alunos, fazendo com que a evaso se torne um caso raro no contexto dessa anlise. Por essa razo,

não há formação de um padrão forte para o estudo da evasão nesse curso, de modo que as características encontradas podem estar muito aderentes aos poucos alunos que evadiram. Caso a quantidade de alunos evadidos fosse mais balanceada em relação à quantidade de concluintes, os padrões encontrados poderiam ser diferentes e mais fortes, pois abrangeriam um número maior de estudantes.

5.1.1.2 Bacharelado em Engenharia de Controle e Automação

Em relação ao curso de Bacharelado em Engenharia de Controle e Automação, é apresentada a árvore de decisão conforme a Figura 13.

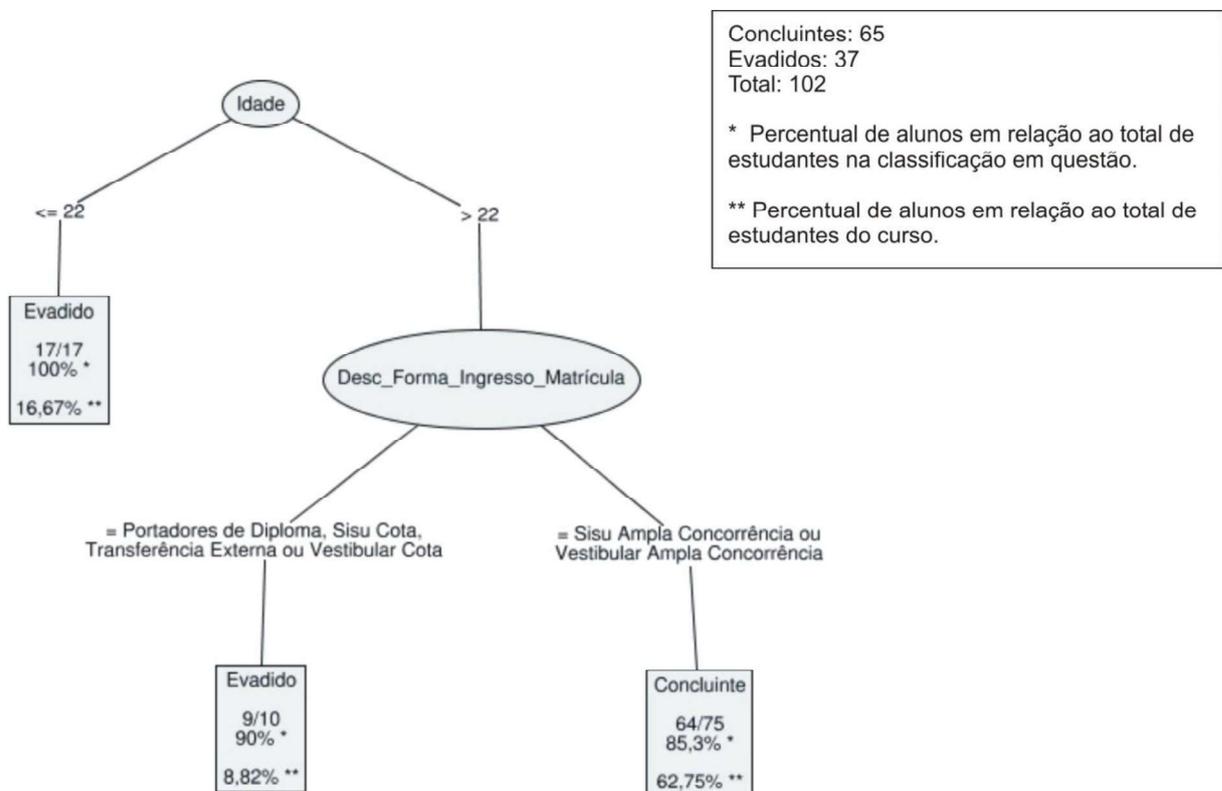


Figura 13: Árvore de decisão para o curso de Bacharelado em Engenharia de Controle e Automação

Fonte: Elaborado pelo autor.

O curso de Bacharelado em Engenharia de Controle e Automação possuía cento e dois registros em sua base, dos quais sessenta e cinco eram concluintes e trinta e sete eram evadidos, demonstrando que a classe “Concluinte” era majoritária. Foram obtidos dois perfis de evasão com essa árvore: o primeiro apontou que dezessete alunos, de um total de dezessete (100% dos registros nessa situação) com idade menor ou igual a vinte e dois anos evadiram. Em relação ao quantitativo total de alunos na base, esse valor corresponde a 16,67%. O segundo perfil encontrado apontou que nove alunos, em um total de dez, com idade maior que

vinte e dois anos e que ingressaram por meio das modalidades Portadores de Diploma, Sisu (Cota), Transferência Externa ou Vestibular (Cota) evadiram, o que representa uma taxa de evasão de 90% para os alunos com esse perfil. Já em relação ao total de alunos, esse valor corresponde a 8,82%

5.1.1.3 Bacharelado em Sistemas de Informação

Em relação ao curso de Bacharelado em Sistemas de Informação, é apresentada a árvore de decisão conforme a Figura 14:

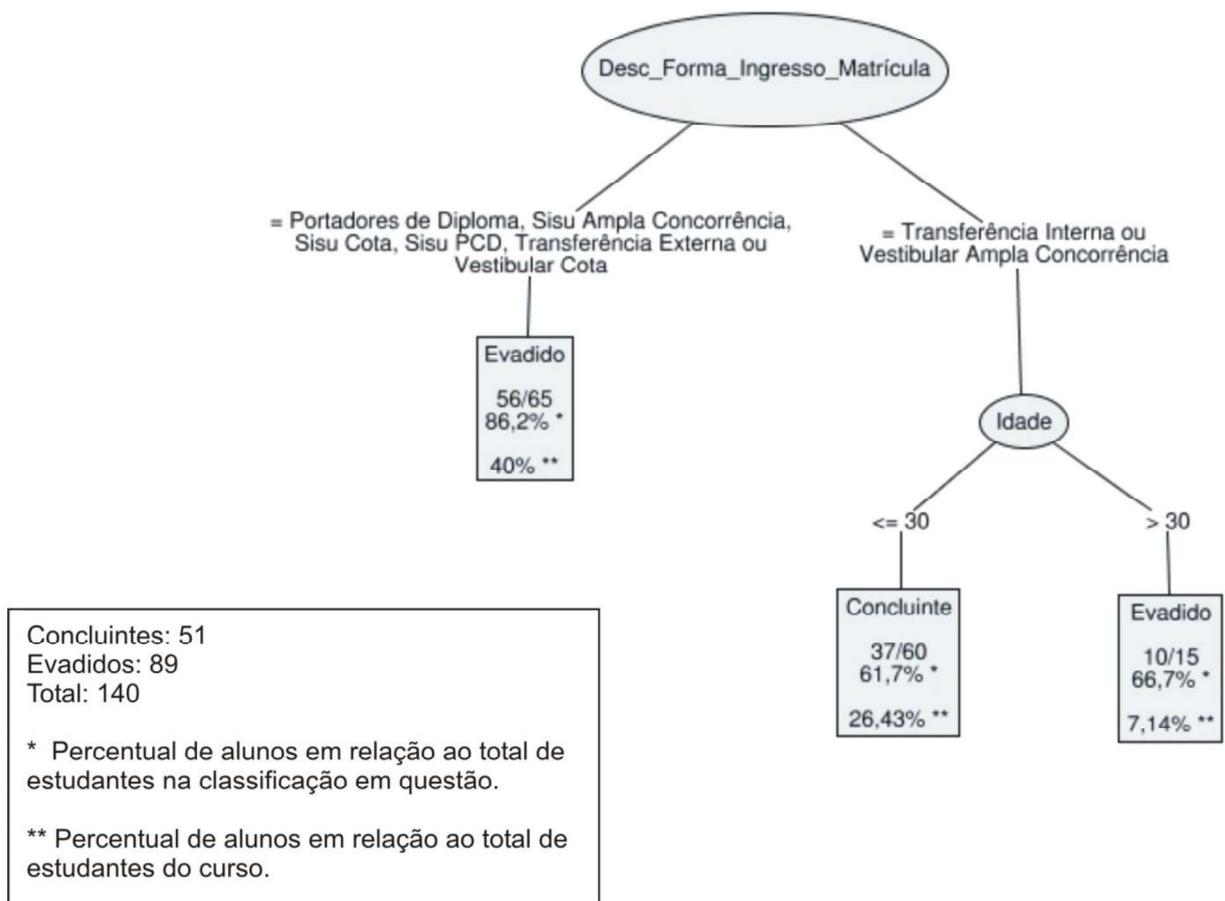


Figura 14: Árvore de decisão para o curso de Bacharelado em Sistemas de Informação
Fonte: Elaborado pelo autor.

O curso de Bacharelado em Sistemas de Informação possuía cento e quarenta registros em sua base. A árvore de decisão gerada demonstrou que a classe “Evadido” era majoritária, com oitenta e nove alunos, o que equivale a 63,6% do total de registros da respectiva base.

Foram obtidos dois perfis de evasão com essa árvore: o primeiro apontou que cinquenta e seis alunos, de um total de sessenta e cinco (86,2% dos registros nessa situação e 40% do total de registros) que ingressaram pelas modalidades Portadores de Diploma, Sisu

(Ampla Concorrência), Sisu (Cota), Sisu (PCD), Transferência Externa ou Vestibular (Cota) evadiram. O segundo perfil identificado apontou que dez estudantes, em um total de quinze, com idade acima de trinta anos e que ingressaram por meio das modalidades Transferência Interna ou Vestibular (Ampla Concorrência) evadiram, correspondendo a uma taxa de evasão de 66,7% para os estudantes com esse perfil, correspondendo a 7,14% em relação ao quantitativo total.

5.1.2 Cursos de Licenciatura

Nesta seção são apresentadas as árvores de decisão obtidas para os cursos de Licenciatura. O número mínimo de objetos para a criação de um ramo na árvore foi de 10% do total de registros em cada curso.

5.1.2.1 Licenciatura em Ciências da Natureza

Em relação ao curso de Licenciatura em Ciências da Natureza, é apresentada a árvore de decisão conforme a Figura 15.

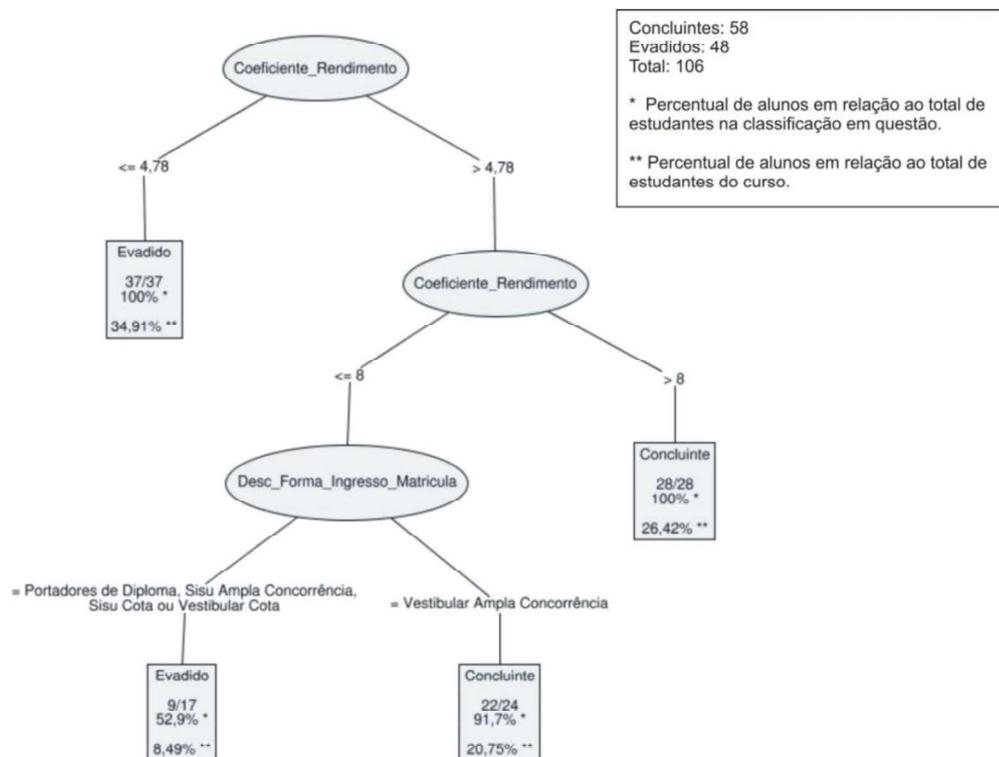


Figura 15: Árvore de decisão para o curso de Licenciatura em Ciências da Natureza.
Fonte: Elaborado pelo autor.

Em relação ao curso de Licenciatura em Ciências da Natureza, “Concluinte” foi a classe majoritária. A base desse curso possuía cento e seis alunos, dos quais cinquenta e oito concluíram, correspondendo a 54,7%. Nesse curso, foi observada uma grande importância do Coeficiente de Rendimento como fator determinante na evasão de alunos. Trinta e sete alunos, em um total de trinta e sete (ou seja, 100%) com Coeficiente de Rendimento menor ou igual a 4,78 evadiram. Esse valor corresponde a 34,91% do total de registros. Outro perfil de evasão também foi descoberto, porém com uma taxa bem menor de evasão: 52,9%. Nesse segundo caso, nove alunos, em um total de dezessete evadiram, o que corresponde a 8,49% do total de registros. Estes são alunos que ingressaram através das seguintes modalidades: Portadores de Diploma, Sisu (Ampla Concorrência), Sisu (Cota) ou Vestibular (Cota), possuindo Coeficiente de Rendimento entre 4,78 e 8.

5.1.2.2 Licenciatura em Geografia

Em relação ao curso de Licenciatura em Ciências da Natureza, é apresentada a árvore de decisão conforme a Figura 16.

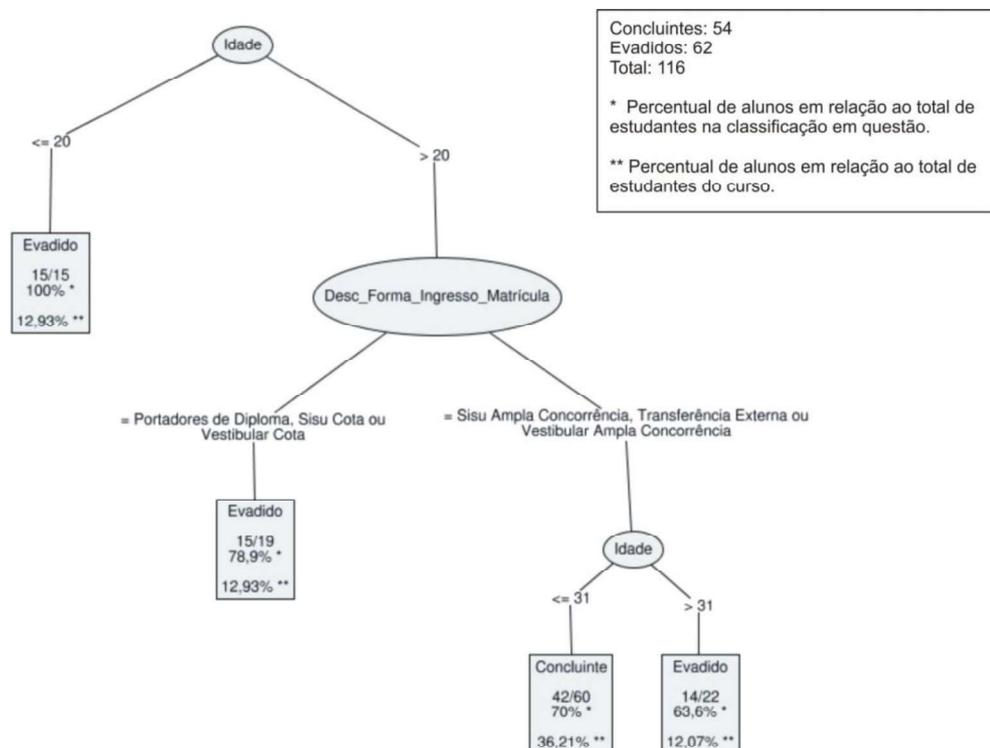


Figura 16: Árvore de decisão para o curso de Licenciatura em Geografia
Fonte: Elaborado pelo autor.

O curso de Licenciatura em Geografia apresentou a classe “Evadido” como majoritária. A base desse curso possuía um total de 116 registros, dos quais 62 eram evadidos, representando 53,4% do total de registros. Foram encontrados três perfis de evasão: no primeiro perfil se encontram alunos com idade menor ou igual a vinte anos. Quinze alunos, em um total de quinze que se enquadravam nessa situação evadiram, correspondendo a 100% dos alunos enquadrados nessa situação (12,93% do total de registros da base). No segundo perfil identificado, encontram-se alunos cuja idade é maior que vinte, possuindo como forma de ingresso as situações de Portadores de Diploma, Sisu (Cota) ou Vestibular (Cota). Nesse perfil, encontram-se quinze alunos em um total de dezenove, correspondendo a 78,9%. Já em relação total de registros da base, esse valor corresponde a 12,93%. No terceiro perfil encontrado estão os alunos cuja idade é maior que trinta e um anos, possuindo como forma de ingresso as situações de Sisu (Ampla Concorrência), Transferência Externa ou Vestibular (Ampla Concorrência). Nesse último perfil, quatorze de vinte e dois alunos evadiram, o que equivale a 63,6% dos alunos que estão nessa situação, correspondendo a 12,07% do total de alunos da base desse curso.

5.1.2.3 Licenciatura em Letras

Em relação ao curso de Licenciatura em Letras, é apresentada a árvore de decisão conforme a Figura 17.

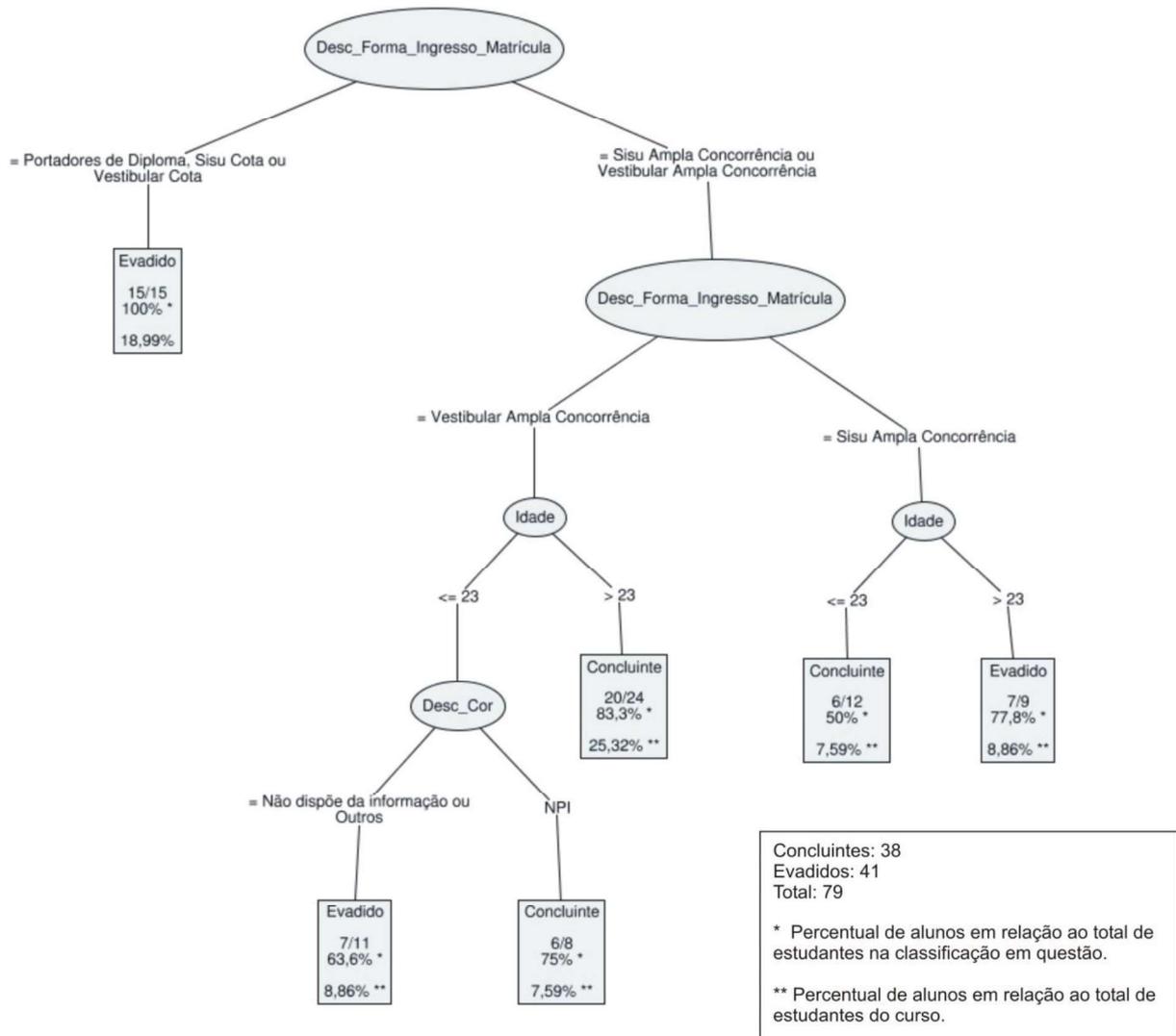


Figura 17: Árvore de decisão para o curso de Licenciatura em Letras
Fonte: Elaborado pelo autor.

O curso de Licenciatura em Letras possuía um total de setenta e nove registros, dos quais quarenta e um eram referentes a alunos evadidos, o que corresponde a 51,9% dos registros dessa base, fazendo desta a classe majoritária da mesma. Foram identificados três perfis de evasão: no primeiro, encontram-se alunos cuja forma de ingresso ocorreu por meio das modalidades de Portadores de Diploma, Sisu (Cota), ou Vestibular (Cota). Nessa situação havia quinze alunos em um total de quinze, correspondendo a 100%. Em relação ao total de registros da base, esse valor corresponde a 18,99%. No segundo perfil, encontram-se alunos cuja forma de ingresso ocorreu por meio de Vestibular (Ampla Concorrência), com idade menor ou igual a vinte e três anos e cuja cor é diferente de NPI (Negros, Pardos e Índios). Desses, sete alunos em um total de onze evadiram, representando 63,6%. Em relação ao total de registros, esse valor equivale a 8,86%. No terceiro perfil, encontram-se alunos cuja forma de ingresso ocorreu por meio de Sisu (Ampla Concorrência) e com idade maior que vinte e

três anos. Nessa situação, encontram-se sete de nove alunos, o que equivale a 77,8%. Em relação ao total de registros, o valor corresponde a 8,86%.

5.1.2.4 Licenciatura em Matemática

Em relação ao curso de Licenciatura em Matemática, é apresentada a árvore de decisão conforme a Figura 18.

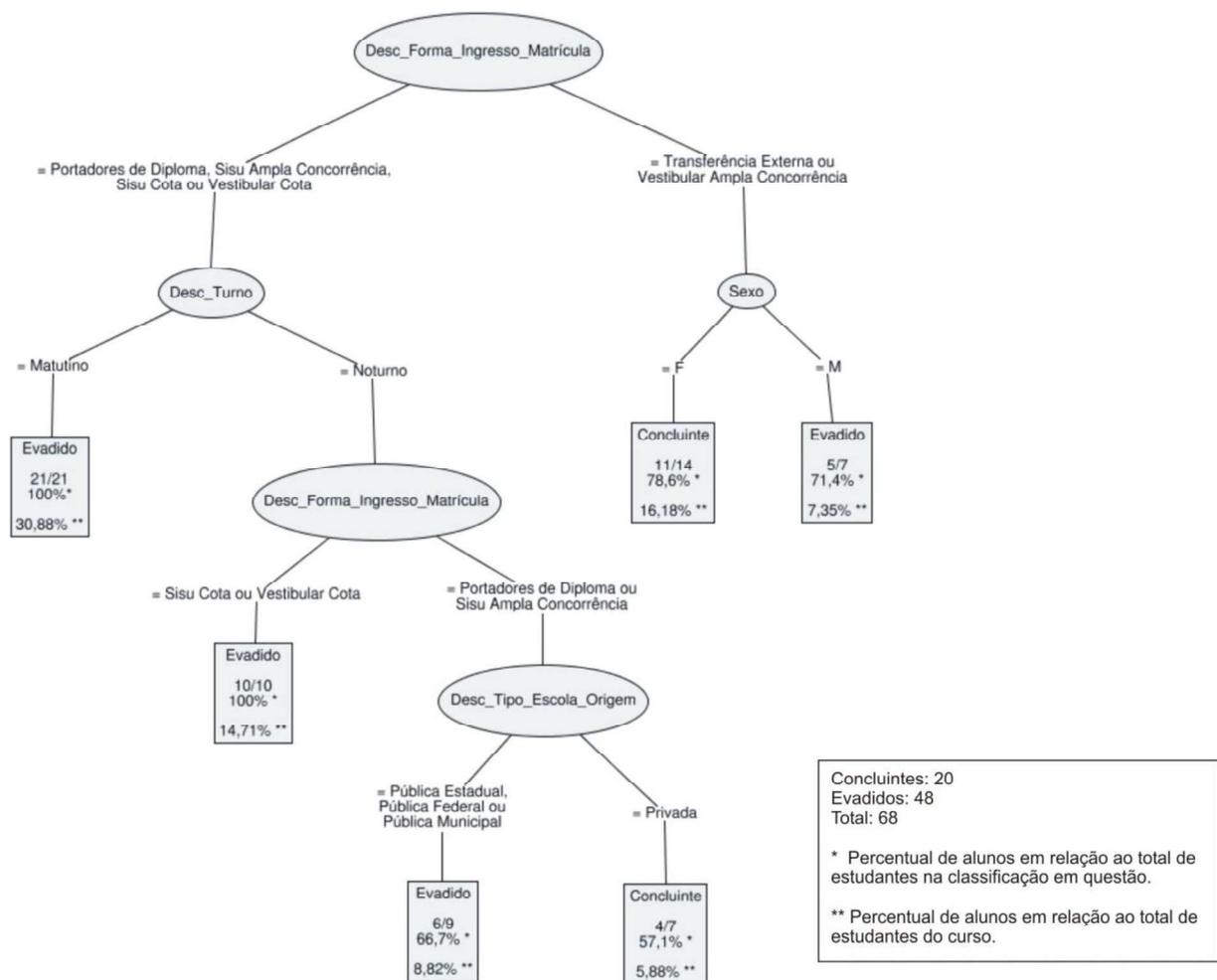


Figura 18: Árvore de decisão para o curso de Licenciatura em Matemática

Fonte: Elaborado pelo autor.

O curso de Licenciatura em Matemática teve a classe “Evadido” como majoritária. De um total de sessenta e oito registros, quarenta e oito eram referentes a alunos que evadiram. Esse número representa 70,6% do total de registros, demonstrando um alto índice de evasão para o referido curso. Foram descobertos quatro padrões de evasão:

a) Alunos que ingressaram através das modalidades Portadores de Diploma, Sisu (Ampla Concorrência), Sisu (Cota) ou Vestibular (Cota) do turno matutino: nesse perfil se

enquadraram vinte e um alunos, dos quais vinte e um evadiram, ou seja, 100% dos registros nessa situação. Em relação ao total de alunos, esse número equivale a 30,88%;

b) Alunos que ingressaram através das modalidades Sisu (Cota) ou Vestibular (Cota) do turno noturno: nesse perfil se enquadraram dez alunos, dos quais dez evadiram, representando 100% dos registros classificados nessa situação, equivalendo a 14,71% do total de alunos;

c) Alunos que ingressaram através das modalidades Portadores de Diploma ou Sisu (Ampla Concorrência) do turno noturno, que cursaram o Ensino Médio em escola pública: nesse perfil se enquadraram nove alunos, dos quais seis evadiram, o que equivale a 66,7% dos registros nessa situação (8,82% do total de alunos do curso);

d) Alunos do sexo masculino que ingressaram através das modalidades Transferência Externa ou Vestibular (Ampla Concorrência): nesse perfil se enquadraram sete alunos, dos quais cinco evadiram, equivalendo a 71,4% dos registros classificados nessa situação (7,35% do total de alunos do curso).

5.1.3 Cursos de Tecnologia

Nesta seção são apresentadas as árvores de decisão referentes aos cursos de Tecnologia. O número mínimo de objetos para a criação de um ramo na árvore foi de 10% do total de registros em cada curso.

5.1.3.1 Tecnologia em Design Gráfico

Em relação ao curso de Tecnologia em Design Gráfico, é apresentada a árvore de decisão conforme a Figura 19.

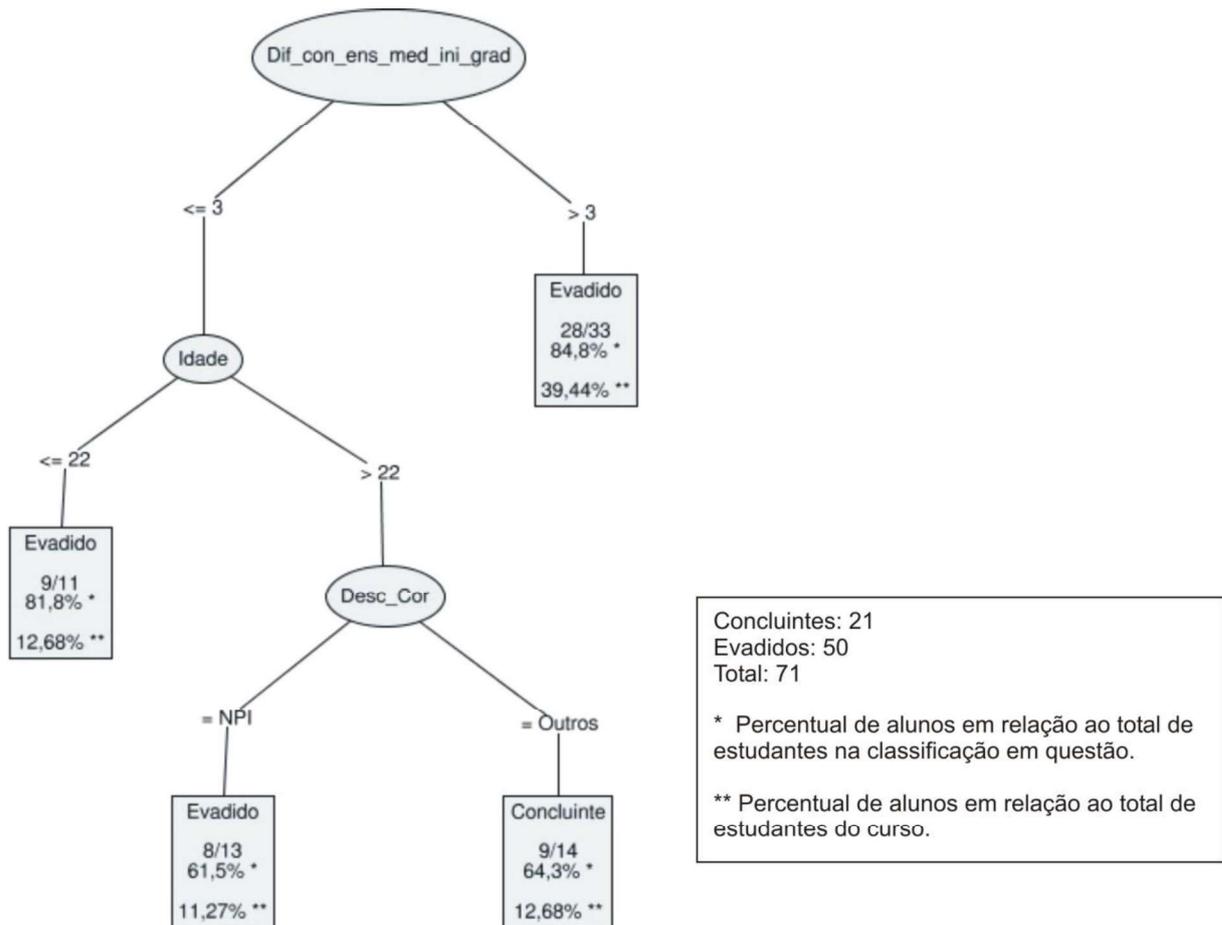


Figura 19: Árvore de decisão para o curso de Tecnologia em Design Gráfico
Fonte: Elaborado pelo autor.

No curso de Tecnologia em Design Gráfico, a classe “Evadido” também foi majoritária. De um total de setenta e um alunos, cinquenta evadiram, o que corresponde a 70,4% do total de registros. Foram descobertos três padrões de evasão para esse curso:

a) Alunos que levaram até três anos desde a conclusão do Ensino Médio até o início no curso superior, com idade menor ou igual a vinte e dois anos: nesse perfil se enquadraram onze alunos, sendo que destes, nove evadiram, o que equivale a 81,8% dos registros nessa situação. Em relação ao total de alunos do curso, o valor equivale a 12,68%;

b) Alunos que levaram até três anos desde a conclusão do Ensino Médio até o início no curso superior, com idade superior a vinte e dois anos, cuja a cor é NPI: nesse perfil foram identificados treze estudantes, dos quais oito evadiram, correspondendo a 61,5% dos registros classificados nessa situação (11,27% do total de alunos do curso);

c) Alunos que levaram mais de três anos desde a conclusão do Ensino Médio até o início no curso superior: nesse perfil foram identificados trinta e três alunos, dos quais vinte e oito

evadiram, equivalendo a 84,8% dos registros classificados nessa situação. Em relação ao total de alunos do curso, o valor corresponde a 39,44%.

5.1.3.2 Tecnologia em Manutenção Industrial

Em relação ao curso de Tecnologia em Manutenção Industrial, é apresentada a árvore de decisão conforme a Figura 20.

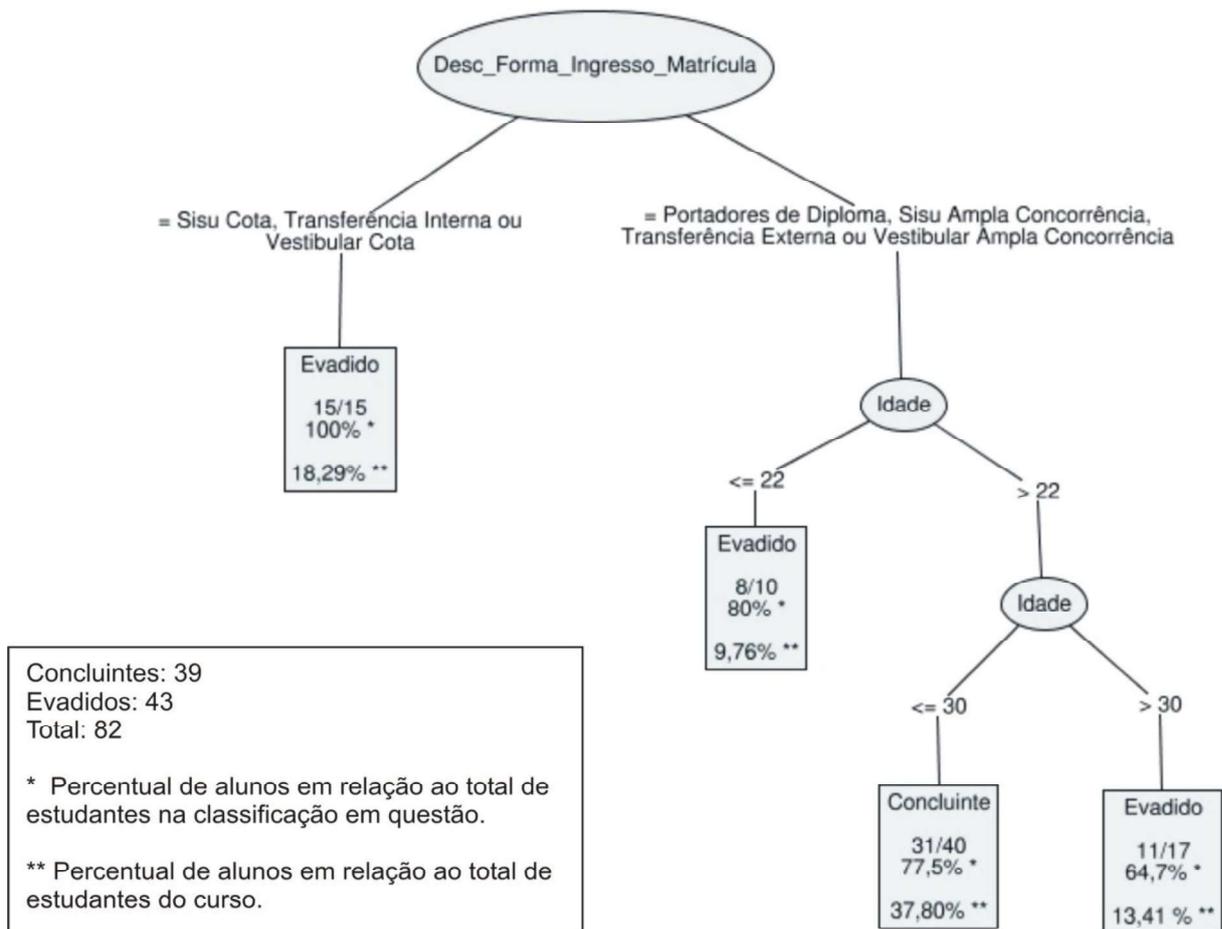


Figura 20: Árvore de decisão para o curso de Tecnologia em Manutenção Industrial.
Fonte: Elaborado pelo autor.

Em relação ao curso de Tecnologia em Manutenção Industrial, sua base possuía oitenta e dois registros, dos quais quarenta e três eram referentes a alunos evadidos, sendo esta a classe majoritária para o referido curso. Esse número corresponde a 52,4% do total de registros da base em questão. Para esse curso, foram descobertos três padrões de evasão:

a) Alunos que ingressaram por meio das modalidades Sisu (Cota), Transferência Interna ou Vestibular (Cota): nesse perfil foram identificados quinze alunos, dos quais quinze

evadiram, o que corresponde a 100% dos registros nessa situação. O valor corresponde a 18,29% do total de alunos do curso;

b) Alunos com idade menor ou igual a vinte e dois anos, que ingressaram por meio das modalidades Portadores de Diploma, Sisu (Ampla Concorrência), Transferência Externa ou Vestibular (Ampla Concorrência): nesse perfil foram identificados dez estudantes, dos quais oito evadiram, o que equivale a 80% dos registros classificados nessa situação (9,76% do total de alunos do curso);

c) Alunos que ingressaram por meio das modalidades Portadores de Diploma, Sisu (Ampla Concorrência), Transferência Externa ou Vestibular (Ampla Concorrência), com idade maior que trinta anos: nesse perfil foram identificados dezessete alunos, dos quais onze evadiram, correspondendo a 64,7% dos registros classificados nessa situação. Em relação ao total de alunos do curso, o valor corresponde a 13,41%.

5.1.3.3 Tecnologia em Sistemas de Telecomunicações

Em relação ao curso de Tecnologia em Sistemas de Telecomunicações, é apresentada a árvore de decisão conforme a Figura 21.

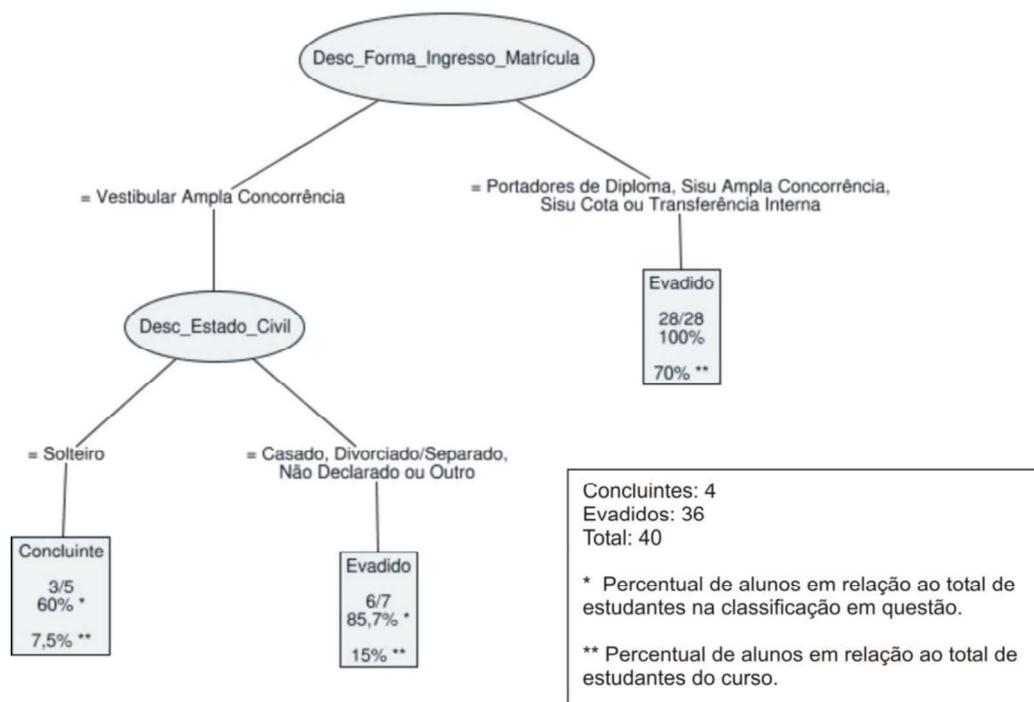


Figura 21: Árvore de decisão para o curso de Tecnologia em Sistemas de Telecomunicações.
Fonte: Elaborado pelo autor.

O curso de Tecnologia em Sistemas de Telecomunicações possuía uma base com quarenta registros, sendo que destes, trinta e seis eram referentes a alunos evadidos. Assim, esta é a classe majoritária para o referido curso, correspondendo a 90% do total de registros da base, o que representa um alto índice de evasão. Para esse curso, foram descobertos dois padrões de alunos evadidos:

a) Alunos que ingressaram por meio da modalidade Vestibular (Ampla Concorrência), cujo estado civil é diferente de solteiro: nesse perfil foram identificados sete alunos, dos quais seis evadiram, equivalendo a 85,7% dos registros nessa situação (15% do total de alunos do curso);

b) Alunos que ingressaram por meio das modalidades Portadores de Diploma, Sisu (Ampla Concorrência), Sisu (Cota) ou Transferência Interna: nesse perfil foram identificados vinte e oito alunos, dos quais vinte e oito evadiram, correspondendo a 100% dos registros nessa situação (70% em relação ao total de alunos do curso).

No caso do curso de Tecnologia em Sistemas de Telecomunicações, cabe destacar que, apesar do resultado encontrado, o número de alunos evadidos corresponde quase que à totalidade de estudantes, fazendo com que os alunos que concluem o curso se tornem um caso raro no contexto dessa análise. Por esse motivo, não há formação de um padrão forte para o estudo do sucesso nesse curso, de modo que as características encontradas podem estar muito aderentes aos poucos alunos que concluíram. Caso a quantidade de alunos concluintes fosse mais balanceada em relação à quantidade de alunos evadidos, os padrões encontrados poderiam ser diferentes e mais fortes, pois abrangeriam um número maior de estudantes. Isso poderia ocasionar mudanças na árvore de decisão desse curso, o que conseqüentemente também influenciaria os perfis encontrados dos alunos que evadiram.

5.2 Regras de Associação

Além das árvores de decisão, foi utilizado também o algoritmo *CN2 Rule Induction*, que obtém regras no formato “se [condição] então [classe]”. Esse algoritmo, porém, não conseguiu encontrar regras que fossem relevantes para os cursos de Bacharelado em Arquitetura e Urbanismo, Licenciatura em Ciências da Natureza, Licenciatura em Letras e Licenciatura em Matemática. Para os demais cursos, as principais regras encontradas estão relacionadas no Quadro 9. Nele, é apresentada também a distribuição de concluintes/evadidos e a probabilidade do aluno ser da classe “Evadido” quando cada regra ocorre.

Quadro 9: Principais regras de associação encontradas

Curso	Principais Regras	Distribuição [Concluintes, Evadidos]	Acurácia
Bacharelado em Arquitetura e Urbanismo	-----	-----	-----
Bacharelado em Engenharia de Controle e Automação	Idade \leq 23 E Desc_Forma_Ingresso_Matricula \neq Vestibular Ampla Concorrência	[0,15]	94%
	Dif_con_ens_med_ini_grad \geq 4	[0,10]	92%
Bacharelado em Sistemas de Informação	Desc_Forma_Ingresso_Matricula = Sisu Cota	[0,14]	94%
	Coefficiente de rendimento \leq 4,65 E Desc_Tipo_Escola_Origem = Privada	[0,22]	96%
	Coefficiente de rendimento \leq 4,65 E Desc_Forma_Ingresso_Matricula = Sisu Ampla Concorrência	[0,15]	94%
Licenciatura em Ciências da Natureza	-----	-----	-----
Licenciatura em Geografia	Estado Civil = Casado	[0,12]	93%
Licenciatura em Letras	-----	-----	-----
Licenciatura em Matemática	-----	-----	-----
Tecnologia em Design Gráfico	Dif_con_ens_med_ini_grad \geq 8	[0,15]	94%
	Desc_Estado_Civil_Pais = Divorciado	[1,7]	80%
	Dif_con_ens_med_ini_grad \geq 4 E Idade \leq 28 E N_Filhos = 0	[1,7]	80%
	Idade \geq 26	[1,7]	80%
Tecnologia em Manutenção Industrial	Desc_Forma_Ingresso_Matricula = Sisu Cota	[0,10]	92%

Curso	Principais Regras	Distribuição [Concluintes, Evadidos]	Acurácia
Tecnologia em Manutenção Industrial (continuação)	Desc_Forma_Ingresso_Matricula ≠ Vestibular Ampla Concorrência E Idade >= 30	[0,10]	92%
	Idade <= 23 E Desc_Tipo_Escola_Origem ≠ Pública Federal	[1,8]	82%
Tecnologia em Sistemas de Telecomunicações	Desc_Forma_Ingresso_Matricula = Sisu Cota	[0,5]	86%

Fonte: Elaborado pelo autor

Analisando o Quadro 9, pode-se observar, por exemplo, que em relação ao curso de Bacharelado em Engenharia de Controle e Automação, quando um aluno possui 23 anos de idade ou menos e ingressou no curso por qualquer modalidade diferente de Vestibular (Ampla Concorrência), este possui uma probabilidade de 94% de evadir. Na distribuição analisada pelo algoritmo para encontrar essa regra, em um grupo de quinze alunos com essas características, todos evadiram. No mesmo curso, alunos que levaram quatro anos ou mais entre a conclusão do ensino médio e o início no curso superior, possuem uma probabilidade de 92% de evadir. Nesse segundo caso, em um grupo de dez alunos que possuíam essas características, todos também evadiram.

Entre as principais regras, pode-se observar também que a forma de ingresso é a característica que mais se destacou, aparecendo seis vezes, das quais três correspondem à forma de ingresso Sisu (Cota). Em segundo lugar, a característica que mais aparece é a idade, totalizando cinco vezes, sendo que duas correspondem à idade menor ou igual a 23 anos. Em terceiro, está o tempo que o estudante leva entre a conclusão do ensino médio e o início do ensino superior. Ele aparece três vezes, das quais duas delas correspondem a um tempo maior ou igual a quatro anos. Essas são algumas das regras encontradas, recorrentes em diferentes cursos. Tais informações podem ser utilizadas como ponto de partida para que perfis de evasão sejam traçados, apontando direções para a elaboração de políticas de redução dessas taxas.

5.3 Panorama geral de evasão nos cursos investigados

Com o objetivo de apresentar um panorama geral dos cursos investigados, foi elaborado o Quadro 10.

Quadro 10: Panorama geral de evasão nos cursos investigados

Curso	Classe Majoritária	Porcentagem da Classe Majoritária
Bacharelado em Arquitetura e Urbanismo	Concluinte	82%
Bacharelado em Engenharia de Controle e Automação	Concluinte	63,7%
Bacharelado em Sistemas de Informação	Evadido	63,6%
Licenciatura em Ciências da Natureza	Concluinte	54,7%
Licenciatura em Geografia	Evadido	53,4%
Licenciatura em Letras	Evadido	51,9%
Licenciatura em Matemática	Evadido	70,6%
Tecnologia em Design Gráfico	Evadido	70,4%
Tecnologia em Manutenção Industrial	Evadido	52,4%
Tecnologia em Sistemas de Telecomunicações	Evadido	90%

Fonte: Elaborado pelo autor.

Comparando resultados obtidos por meio dos algoritmos utilizados, pode-se observar características que se confirmam como fatores altamente relacionados à evasão: alunos jovens (com idade até 23 anos) e alunos que ingressaram por meio de cota, tanto pelo Sisu como por Vestibular são os que mais evadem.

5.4 Comparações com Trabalhos Relacionados

Enquanto a presente pesquisa buscou identificar o perfil de alunos que evadiram de cursos superiores do *campus* Campos Centro do Instituto Federal Fluminense, Cordeiro (2017) pesquisou as causas da evasão escolar em cursos técnicos dos *campi* desta mesma instituição. Em seu trabalho, foram utilizados dados de dez *campi*, envolvendo cursos concomitantes e cursos subsequentes. De acordo com a autora, não há diferença no comportamento de alunos evadidos nessas duas modalidades, de modo que o mesmo é determinado por dois fatores: desempenho nas disciplinas dos primeiros módulos dos cursos (em sua maioria, disciplinas técnicas) e a ausência nas aulas.

No trabalho de Moscoso-Zea (2017), foram analisados dados de alunos do curso de Ciência da Computação de uma universidade privada, onde foram apontadas as principais características relacionadas ao sucesso ou à evasão por parte dos estudantes. Foram identificadas maiores taxas de sucesso no caso de alunos que nasceram e vivem na mesma

cidade onde está localizada a instituição de ensino. Também foram identificadas maiores taxas de sucesso no caso de alunos que ingressaram através da validação de cursos do que em relação aos alunos que ingressaram a partir do primeiro período. Já em relação aos casos de evasão, foram observadas maiores taxas para os casos de alunos que cursaram o ensino médio em escolas públicas, assim como para os estudantes que são casados. Outro caso de alunos considerados em situação de risco são aqueles que ingressaram no curso recebendo algum tipo de bolsa, mas que ao longo do tempo perderam esse benefício.

Manhães, Cruz e Zimbrão (2014) analisaram dados de cerca de 14.000 estudantes da Universidade Federal do Rio de Janeiro, a maior universidade pública federal brasileira. Foram analisados dados de doze semestres a partir do momento de matrícula desses alunos na universidade (matrículas realizadas nos anos de 2003 e 2004), de modo que o período analisado corresponde até o segundo semestre de 2010. Todos os dados utilizados foram obtidos através do Sistema de Gestão Acadêmica, de modo que não foram utilizadas informações consideradas não acadêmicas.

Algumas das características observadas pelos autores em relação aos alunos que abandonaram os estudos na UFRJ foram: reduzir inscrições em disciplinas a cada semestre até o abandono; possuir um número decrescente de disciplinas com notas de aprovação em cada semestre; possuir pelo menos uma disciplina com reprovação por nota ou por falta no primeiro semestre; possuir, ao final do primeiro semestre, notas de aprovação inferiores aos concluintes e por último, obter boas notas em poucas disciplinas.

O Quadro 11 apresenta a comparação de trabalhos relacionados com a presente pesquisa.

Quadro 11: Comparação de trabalhos relacionados com a presente pesquisa

	Bernardo <i>et al.</i> (2017)	Manhães, Cruz e Zimbrão (2014)	Hegde e Prageeth (2018)	Shyamala, Rajagopalan e Mohammed (2006)	Moscoso-Zea, Vizcaino e Mora (2017)	Mashiloane e Mchunu (2013)	Barreto (2019)
Quantidade de alunos	1.311	14.000	50	180	441	391	1.022
Período	2008 – 2012	2003 – 2010	Não informado	Não informado	2002 – 2015	2009 – 2011	2017 – 2019
Instituição	Universidade de Oviedo, Espanha	Universidade Federal do Rio de Janeiro (UFRJ), Brasil	Amrita School of Arts and Sciences, Índia	Dr. Ambedkar Government College, Índia	Não informado	University of The Witwatersrand, África do Sul	Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Brasil
Ferramentas utilizadas	IBM Statistical Package for the Social Sciences	Weka	Weka	Clementine	Weka; Orange; Rapid Miner	Weka	Orange
Algoritmos utilizados	<i>Exhaustive</i> CHAID	J48; SimpleCart; Support Vector Machine (SVM); Naïve Bayes (algoritmo selecionado); Multilayer Perceptron - MP.	Naïve Bayes	C5.0	Naïve Bayes; Stacking; One R; J48 (algoritmo selecionado); Random Tree.	J48 (algoritmo selecionado); Naïve Bayes; Decision Table.	C4.5 (Decision Tree)
Bases de dados consideradas	Base acadêmica; Questionários aplicados aos alunos.	Base acadêmica.	Base acadêmica; Questionários aplicados aos alunos.	Base acadêmica.	Base acadêmica.	Base acadêmica.	Base acadêmica

Fonte: Elaborado pelo autor.

O Quadro 12 apresenta os padrões relacionados à evasão encontrados nos trabalhos relacionados e também na presente pesquisa:

Quadro 12: Perfis de evasão encontrados nos trabalhos relacionados e na presente pesquisa

	Bernardo <i>et al.</i> (2017)	Manhães, Cruz e Zimbrão (2014)	Hegde e Prageeth (2018)	Shyamala, Rajagopalan e Mohammed (2006)	Moscoso-Zea, Vizcaino e Mora (2017)	Mashiloane e Mchunu (2013)	Barreto (2019)
Padrões encontrados de alunos evadidos	<ul style="list-style-type: none"> - Estavam na fase inicial do curso; - Possuíam idade igual ou superior a 20 anos no momento de ingresso no curso; - Declararam ter dedicado pouco tempo aos estudos durante a fase inicial do curso; - Moravam sozinhos, com parentes que não sejam os pais ou com parceiro(a). 	<ul style="list-style-type: none"> - Reduziram a quantidade de disciplinas em que estão inscritos a cada semestre; - Possuíram um número decrescente de disciplinas com notas de aprovação a cada semestre; - Possuíram, pelo menos, uma disciplina em que foram reprovados (incluindo também reprovação por falta) no primeiro semestre acadêmico; - Ao final do primeiro semestre, possuíam notas de aprovação inferiores às dos alunos concludentes; - Possuíam boas notas em poucas disciplinas. 	<ul style="list-style-type: none"> - Possuíram quatro reprovações em disciplinas; - Possuíram três reprovações por faltas; - Possuíam algum problema de saúde; - Não se adaptaram ao ambiente institucional. 	<p>Não foi apontado nenhum padrão de evasão. O modelo agrupou os estudantes de acordo com várias notas. Associou as notas do semestre anterior com notas do semestre em andamento para prever o resultado final. Apontou como fatores que influenciam a evasão: dificuldades acadêmicas, preferências acadêmicas, falta de ajuda financeira, frequência, entre outros.</p>	<ul style="list-style-type: none"> - Cursaram o ensino médio em escola pública; - Eram casados; - Possuíam bolsa de estudos no momento de ingresso na universidade, mas perderam a mesma no decorrer do curso. 	<p>Não foi apontado nenhum padrão de evasão. Foi desenvolvida a ferramenta <i>Success Or Failure Determiner</i> (SOFD), que realizou a predição da situação do aluno no final do ano acadêmico (sucesso ou insucesso) com base em sua nota no primeiro semestre.</p>	<ul style="list-style-type: none"> - Foram identificados perfis de evasão de acordo com cada curso; - De maneira geral, a evasão sofre maior influência da forma de ingresso e menor influência de características como sexo (maior evasão do sexo masculino) e cor (maior evasão de negros, pardos e índios); - A forma de ingresso com maior evasão é a modalidade “Sisu – Ampla Concorrência”.

Fonte: Elaborado pelo autor

6 CONCLUSÕES

A evasão é um problema que atinge tanto instituições privadas como públicas, de modo que suas consequências atingem a sociedade de maneira geral. Nesse sentido, conhecer os motivos que levam os alunos a evadirem é algo de extrema importância para que o problema possa ser reduzido. Quanto antes for identificado os alunos que possuem alguma tendência para evasão, mais tempo as instituições possuem para tratar esse problema antes que ele se concretize. Uma forma de realizar essa identificação de maneira precoce é por meio de técnicas de mineração de dados.

6.1 Contribuições

O presente trabalho utilizou técnicas de mineração de dados para identificar possíveis alunos em risco de evasão. Para tal, a pesquisa foi aplicada nos cursos de nível superior oferecidos pelo *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense. Nesta seção é realizado um paralelo entre os objetivos específicos traçados por meio do presente trabalho e é feita uma análise se os mesmos foram ou não atingidos, demonstrando as contribuições da presente pesquisa. Os objetivos específicos do presente trabalho são: a) realizar a bibliometria da área de estudo; b) definir e analisar fatores que podem estar relacionados à evasão e c) identificar, para cada curso investigado, perfis de alunos evadidos (evasão de curso).

6.1.1 Realização da bibliometria da área de estudo

O primeiro objetivo específico da presente pesquisa foi realizar um estudo bibliométrico pautado em três conceitos: mineração de dados, evasão e ensino superior. Tal estudo foi realizado e por meio dele foram apresentados alguns dados estatísticos sobre os países, revistas e autores que possuem maior quantidade de publicações na área, assim como os artigos mais citados. Deste modo, o objetivo em questão foi alcançado, sendo também utilizado para embasar a seção de trabalhos relacionados.

6.1.2 Definição e análise de fatores possivelmente relacionados à evasão

Esse objetivo foi traçado com o propósito de avaliar, por meio de técnicas de mineração de dados, se determinadas ideias a respeito da evasão teriam fundamentos. No caso do presente trabalho, foi avaliado se a idade e se o tempo em que um estudante leva entre a conclusão do ensino médio e o início do curso superior possuiriam relação com a evasão. Para tal, esses atributos foram criados e calculados, uma vez que os mesmos não constavam nos dados originais exportados do sistema acadêmico.

Em relação ao atributo “idade”, foi observado que este se mostrou relevante em sete de um total de dez cursos. Destes sete, os alunos evadidos geralmente possuíam idade menor ou igual a 23 anos, o que demonstra que estudantes mais jovens possuem forte tendência à evasão, mesmo em diferentes cursos. Já em relação entre a diferença de tempo que um estudante leva entre a conclusão do ensino médio e o início da graduação, este foi observado em apenas um curso. Dessa forma, este objetivo específico foi cumprido, de modo que no contexto analisado, a idade possui forte relação com a evasão, mas a diferença de tempo entre a conclusão do ensino médio e o início no ensino superior não é relevante.

6.1.3 Identificação de perfis de alunos evadidos por curso

Por meio da presente pesquisa foi possível observar em quais cursos a evasão ocorre com mais frequência. Desse modo, os cursos de Bacharelado em Sistemas de Informação e Tecnologia em Sistemas de Telecomunicações foram identificados como os cursos com as taxas de evasão mais elevadas. Já o curso de Bacharelado em Arquitetura e Urbanismo possui uma taxa de evasão muito pequena.

Além de permitir observar em que proporção a evasão ocorre em cada curso, foi possível identificar quais são os padrões existentes em relação ao perfil dos alunos que evadiram, fazendo com que este objetivo também fosse alcançado. Cada curso possui suas particularidades em relação a esse perfil, mas, por exemplo, o ingresso por meio de Vestibular – Cota e Sisu – Cota, assim como a idade, geralmente inferior a 23 anos, foram as características mais comuns presentes em diferentes cursos.

Além disso, apesar das bases de dados terem ficado com 43 atributos após a realização de todos os procedimentos de limpeza, apenas nove atributos aparecerem nas árvores de decisão. São estes: forma de ingresso de matrícula, sexo, idade, coeficiente de rendimento, turno, diferença entre a conclusão do ensino médio e o início da graduação, cor, tipo de escola

de origem e estado civil. É importante destacar que a forma de ingresso de matrícula apareceu como raiz em seis de um total de dez cursos, demonstrando a relevância desse atributo na análise da evasão no contexto estudado.

Desse modo, as informações encontradas podem ser utilizadas como feedback aos professores e gestores, promovendo a conscientização dos problemas identificados e auxiliando ações de permanência e êxito a partir dos resultados apresentados. Tais informações também podem ser utilizadas no apoio ao desenvolvimento de atividades como reuniões de brainstorming, estudos de trabalhos sobre permanência e êxito, melhoria dos processos institucionais a partir da aplicação de ferramentas da qualidade, entre outros, a fim de auxiliar a elaboração e implantação de políticas com o objetivo de reduzirem as taxas de evasão. Além disso, o método aplicado não é restrito ao *campus* onde a pesquisa foi aplicada. O mesmo pode ser utilizado em outros *campi* do Instituto Federal de Educação, Ciência e Fluminense, outros Institutos Federais ou outras instituições de ensino, sejam elas públicas ou privadas.

6.2 Trabalhos futuros

Para a realização do presente trabalho, diversos dados de alunos foram analisados e submetidos a algoritmos de mineração de dados, com o objetivo de descobrir padrões relacionados à evasão nos cursos superiores do *campus* Campos Centro do Instituto Federal de Educação, Ciência e Tecnologia Fluminense.

Utilizando esses dados, foram identificadas características que são peculiares a cada curso, mas também chamou a atenção o fato de algumas características serem bastante comuns, se repetindo várias vezes, em diferentes cursos. Tais características estão relacionadas à idade e à forma de ingresso, uma vez que alunos jovens (geralmente com até 23 anos) e alunos que ingressaram por cota, tanto na modalidade Sisu como na modalidade Vestibular, compõem os grupos de estudantes que mais evadem no contexto em que essa pesquisa foi conduzida. Desse modo, o presente trabalho buscou apontar caminhos para a criação de políticas que visem reduzir as taxas de evasão, mas a criação propriamente dita dessas políticas, baseada na análise realizada é um extenso trabalho a ser desenvolvido.

Além disso, no presente trabalho foi estudada a evasão de curso, porém não necessariamente este tipo de evasão representa uma evasão da instituição, visto que o estudante pode migrar de curso e após esse ato se tornar um aluno concluinte. Desse modo,

outra possibilidade de trabalho futuro é a realização do estudo da evasão da instituição, a qual pode ser considerada ainda mais grave que a evasão de curso.

Também é necessário destacar que todos os dados utilizados na presente pesquisa foram obtidos por meio do sistema acadêmico da instituição. Esse sistema possui diversas informações relacionadas ao desempenho dos alunos e informações pessoais, como sexo, data de nascimento, entre outras. Apesar da relevância dessas informações, seria também importante descobrir se existe relação entre o suporte financeiro oferecido pela instituição aos alunos e as causas que motivam a evasão. Porém esse tipo de informação não consta no sistema acadêmico.

Desse modo, outra possibilidade de trabalho futuro é a realização da mineração de dados utilizando uma nova base, que unifique as informações constantes no sistema acadêmico, assim como as informações referentes às bolsas. Além de não constarem no referido sistema acadêmico, as informações relativas a essas bolsas não estão concentradas em um único local no *campus*. Por exemplo, existem bolsas administradas pelo setor de Assistência Estudantil, bolsas administradas pela Diretoria de Gestão Acadêmica, como a bolsa de desenvolvimento acadêmico e apoio tecnológico, bolsas administradas pelo setor de Educação Física, como a bolsa atleta e bolsas administradas pela Diretoria de Ensino, como a bolsa de monitoria. Além destas, existem também bolsa de cultura, bolsa de pesquisa e bolsa de extensão.

Parte dessas bolsas é administrada por meio de planilhas eletrônicas, mas outra parte é administrada manualmente. Além disso, bolsas mais recentes (a partir de 2018) começaram a ser administradas por meio do Sistema Unificado de Administração Pública (SUAP) devido a motivos de pagamento, porém bolsas de anos anteriores não constam no referido sistema. Dessa forma, unificar informações referentes aos pagamentos de bolsas e informações do sistema acadêmico consiste em um processo que exigirá diversos desdobramentos, fazendo dele uma possibilidade de trabalho futuro. Isso permitirá a descoberta de conhecimento útil em dados que hoje estão separados, organizando-os por uma nova perspectiva, o que poderá indicar novas direções para a elaboração de políticas que visem reduzir as taxas de evasão.

REFERÊNCIAS BIBLIOGRÁFICAS

AHMED, Abeer Badr El Din; ELARABY, Ibrahim Sayed. Data Mining: A prediction for Student's Performance Using Classification Method. **World Journal Of Computer Application And Technology**, Cairo, v. 2, n. 2, p.43-47, 2014.

AMO, Sandra de. **Técnicas de Mineração de Dados**. Uberlândia: Universidade Federal de Uberlândia, 2004. 43 p.

ANTUNES, Rodrigo Rodrigues et al. Análise de Integração de Mineradores de Dados com a Plataforma InterIMAGE – Qual a Melhor Solução? **Revista Brasileira de Cartografia**, Uberlândia, v. 70, n. 4, p.1470-1509, 15 dez. 2018. EDUFU - Editora da Universidade Federal de Uberlândia. <http://dx.doi.org/10.14393/rbcv70n4-46512>.

ASSIS, Lucas Rocha Soares de. **Perfil de Evasão no Ensino Superior Brasileiro: uma Abordagem de Mineração de Dados**. 2017. 153 f. Dissertação (Mestrado) - Curso de Mestrado Profissional em Computação Aplicada, Universidade de Brasília, Brasília, 2017.

AYESHA, Shaeela et al. Data mining model for higher education system. **European Journal of Science Research**, Faisalabad, v. 43, p.24-29, jun. 2010.

BALANIUK, Remis et al. Predicting Evasion Candidates in Higher Education Institutions. **Model And Data Engineering**, Brasília, p.143-151, set. 2011. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-24443-8_16.

BERNARDO, Ana et al. Freshmen Program Withdrawal: Types and Recommendations. **Frontiers In Psychology**, v. 8, 21 set. 2017. Frontiers Media SA. <http://dx.doi.org/10.3389/fpsyg.2017.01544>.

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

BRASIL. Conselho Nacional de Educação. **A qualidade social da educação brasileira nos referenciais de compromisso do plano e do sistema nacional de educação**. São Paulo: CNE/UNESCO, 2015. 209 p. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=41841-estudo-sobre-cenario-internacional-das-areas-educacional-empresarial-pdf&Itemid=30192>. Acesso em: 15 jan. 2020.

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

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Censo da Educação Superior**. Disponível em: <<http://portal.inep.gov.br/web/guest/censo-da-educacao-superior>>. Acesso em: 22 ago. 2019.

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

CASTRO, Alexandre Kurtz dos Santos Sisson de; TEIXEIRA, Marco Antônio Pereira. A evasão em um curso de psicologia: uma análise qualitativa. **Psicologia em Estudo**, Maringá, v. 18, n. 2, p.199-209, jun. 2013. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1413-73722013000200002>.

CHEN; CHIANG; STOREY. Business Intelligence and Analytics: From Big Data to Big Impact. **Mis Quarterly**, Mineápolis, v. 36, n. 4, p.1165-1188, 2012. JSTOR. <http://dx.doi.org/10.2307/41703503>.

CORDEIRO, Renata Gomes. **Identificação do comportamento dos estudantes evadidos de cursos técnicos utilizando técnicas de mineração de dados**. 2017. 91 f. Dissertação (Mestrado) - Curso de Mestrado em Sistemas Aplicados à Engenharia e Gestão, Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, 2017.

CRETTON, Nicollas Nogueira. **Mineração de dados aplicada na base do Enade com enfoque na criação de perfis dos estudantes que prestaram o exame utilizando o algoritmo J48**. 2016. 127 f. Dissertação (Mestrado) - Curso de Mestrado em Pesquisa Operacional e Inteligência Computacional, Universidade Cândido Mendes, Campos dos Goytacazes, 2016.

CUNHA, Jacqueline Veneroso Alves da; NASCIMENTO, Eduardo Mendes; DURSO, Samuel de Oliveira. Razões e influências para a evasão universitária: um estudo com estudantes ingressantes nos cursos de ciências contábeis de instituições públicas fede. **Advances In Scientific And Applied Accounting**, São Paulo, v. 9, n. 2, p.141-161, 31 ago. 2016. Associação Nacional de Programas de Pós-Graduação em Ciências Contábeis (ANPCONT). <http://dx.doi.org/10.14392/asaa.2016090202>.

DAI, Wei; JI, Wei. A MapReduce Implementation of C4.5 Decision Tree Algorithm. **International Journal Of Database Theory And Application**, Tasmânia, v. 7, n. 1, p.49-60, 28 fev. 2014. NADIA. <http://dx.doi.org/10.14257/ijdta.2014.7.1.05>.

DEMETRIOU, Cynthia; SCHMITZ-SCIBORSKI, Amy. **Integration, Motivation, Strengths and Optimism: Retention Theories Past, Present and Future**. Chapel Hill: The University Of North Carolina, 2011. 13 p.

DEMSAR, Janez et al. Orange: Data Mining Toolbox in Python. **Journal Of Machine Learning Research**, Liubliana, v. 14, p. 2349-2353, 2013.

DEVASIA, Tismy; P, Vinushree T; HEGDE, Vinayak. Prediction of students performance using Educational Data Mining. **2016 International Conference On Data Mining And Advanced Computing (SAPIENCE)**, Ernakulam, mar. 2016. IEEE. <http://dx.doi.org/10.1109/sapience.2016.7684167>.

EVANGELISTA, Bárbara Michelle; OLIVEIRA, Marcelo; FONSECA, Giovanni Campos. Produção científica em periódicos brasileiros sobre a evasão no ensino superior do país, 1998 a 2017. In: Congresso de Inovação e Metodologias no Ensino Superior, 2019, Belo Horizonte. **Anais...** . Belo Horizonte: Universidade Federal de Minas Gerais, 2019.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **Knowledge Discovery and Data Mining: Towards a Unifying Framework**. Portland: AAAI, 1996. 7 p.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008. 220 p.

GILIOLI, Renato de Souza Porto. **Evasão em instituições federais de ensino superior no Brasil: expansão da rede, SISU e desafios**. Brasília: Câmara dos Deputados, 2016.

GUALHANO, Mariana Abreu. **Análise do desempenho dos cursos de licenciatura da rede federal no cenário nacional por meio do conceito preliminar de curso utilizando mineração de dados**. 2019. 93 f. Dissertação (Mestrado) - Curso de Mestrado em Sistemas Aplicados à Engenharia e Gestão, Instituto Federal de Educação, Ciência e Tecnologia Fluminense, Campos dos Goytacazes, 2019.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. 3. ed. Waltham: Morgan Kaufmann, 2012. 740 p.

HEGDE, Vinayak; PRAGEETH, P. P.. Higher education student dropout prediction and analysis through educational data mining. **2018 2nd International Conference On Inventive Systems And Control (ICISC)**, Coimbatore, jan. 2018. IEEE. <http://dx.doi.org/10.1109/icisc.2018.8398887>.

HERPICH, Fabrício et al. Mineração de Dados Educacionais: uma análise sobre o Engajamento de Usuários em Mundos Virtuais. **Anais dos Workshops do V Congresso Brasileiro de Informática na Educação (CBIE 2016)**, Uberlândia, 10 nov. 2016. Sociedade Brasileira de Computação - SBC. <http://dx.doi.org/10.5753/cbie.webie.2016.910>.

HERSHKOVITZ, Arnon; NACHMIAS, Rafi. Online persistence in higher education web-supported courses. **The Internet And Higher Education**, Tel Aviv, v. 14, n. 2, p.98-106, mar. 2011. Elsevier BV. <http://dx.doi.org/10.1016/j.iheduc.2010.08.001>.

HOFFMANN, Ivan Londero; NUNES, Raul Ceretta; MULLER, Felipe Martins. As informações do Censo da Educação Superior na implementação da gestão do conhecimento organizacional sobre evasão. **Gestão & Produção**, São Paulo, v. 26, n. 2, 9 maio 2019. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/0104-530x-2852-19>.

HUSSAIN, Sadiq et al. Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. **Indonesian Journal of Electrical Engineering And Computer Science**, Yogyakarta, v. 9, n. 2, 1 fev. 2018. Institute of Advanced Engineering and Science. <http://dx.doi.org/10.11591/ijeecs.v9.i2.pp447-459>.

IFCE. **Plano Estratégico para Permanência e Êxito dos Estudantes do IFCE**. 2017. Disponível em: <<https://ifce.edu.br/proen/ensino/plano-de-permanencia-e-exito.pdf>>. Acesso em: 31 ago. 2019.

IFF. **Histórico**. 2015. Disponível em: <<http://portal1.iff.edu.br/conheca-o-iffuminense/historico>>. Acesso em: 19 out. 2018a.

IFF. **Apresentação**. 2015. Disponível em: <<http://portal1.iff.edu.br/conheca-o-iffuminense/apresentacao>>. Acesso em: 19 out. 2018b.

IFF. **Plano de Desenvolvimento Institucional**, 2018. Disponível em: <<http://portal1.iff.edu.br/desenvolvimento-institucional/arquivos/pdi-2018-2022-com-resolucao-menor.pdf>>. Acesso em: 3 ago. 2019

LORENZETT, Cassio dal Castel; TELOKEN, Alex. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão. In: Simpósio de Pesquisa e Desenvolvimento em Computação, 2016, Cruz Alta. **Anais...**. Cruz Alta: Universidade de Cruz Alta, 2016.

LIBREOFFICE. **Recursos do LibreOffice Calc - LibreOffice Help**. 2011. Disponível em: <https://help.libreoffice.org/Calc/Calc_Features/pt-BR>. Acesso em: 15 set. 2019.

LIMA, Franciele Santos de; ZAGO, Nadir. Desafios conceituais e tendências da evasão no ensino superior: a realidade de uma universidade comunitária. **Revista Internacional de Educação Superior**, Campinas, v. 4, n. 2, p.366-386, 2 abr. 2018. Universidade Estadual de Campinas. <http://dx.doi.org/10.20396/riesup.v4i2.8651587>.

LINKE, Elizandra Campos; NOGUEIRA, Bárbara Campos; LINKE, Elisangela Campos. A evasão escolar no ensino técnico profissionalizante. In: SEMINÁRIO INTERINSTITUCIONAL DE ENSINO, PESQUISA E EXTENSÃO, 22., 2017, Cruz Alta. **Anais...**. Cruz Alta: Universidade de Cruz Alta, 2017.

LOPES, Lila Reis. **O marketing nas instituições de educação superior privadas da Baha**: um estudo sobre o nível de conhecimento e potencialidade de uso do marketing e sobre as aspirações e necessidades dos estudantes candidatos. 2006. 192 f. Dissertação (Mestrado) - Curso de Mestrado em Administração, Universidade Federal da Bahia, Salvador, 2006.

MANFREDI, Sílvia Maria. **Educação profissional no Brasil**. São Paulo: Cortez Editora, 2002. 317 p.

MANHÃES, Laci; CRUZ, Sergio; SILVA, Geraldo. The Impact of High Dropout Rates in a Large Public Brazilian University - A Quantitative Approach Using Educational Data Mining. **Proceedings Of The 6th International Conference on Computer Supported Education**, Barcelona, p.124-129, 2014. SCITEPRESS - Science and Technology Publications. <http://dx.doi.org/10.5220/0004958601240129>.

MARTINS, Cleidis Beatriz Nogueira. **Evasão de alunos nos cursos de graduação em uma instituição de ensino superior**. 2007. 116 f. Dissertação (Mestrado) - Curso de Mestrado Profissional de Administração, Fundação Dr. Pedro Leopoldo, Pedro Leopoldo, 2007.

MARTINS, Luiz Carlos Barbosa et al. Early Prediction of College Attrition Using Data Mining. **16th IEEE International Conference on Machine Learning and Applications**, Cancún, dez. 2017. IEEE. <http://dx.doi.org/10.1109/icmla.2017.000-6>.

MASHILOANE, Lebogang; MCHUNU, Mike. Mining for Marks: A Comparison of Classification Algorithms when Predicting Academic Performance to Identify “Students at Risk”. **Mining Intelligence And Knowledge Exploration**, Joanesburgo, p.541-552, 2013. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-03844-5_54.

MILLER, L. Dee et al. A Comparison of Educational Statistics and Data Mining Approaches to Identify Characteristics That Impact Online Learning. **Journal Of Educational Data Mining**, Lincoln, v. 7, n. 3, p.117-150, 2015.

MOSCOSO-ZEA, Oswaldo; VIZCAINO, Mayra; LUJÁN-MORA, Sergio. **Evaluation of methods and algorithms of educational data mining**. Bogotá: Equinoctial Technological University, 2017. 11 p.

NAGY, Marcell; MOLONTAY, Roland. Predicting Dropout in Higher Education based on Secondary School Performance. In: IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT ENGINEERING SYSTEMS, 22, 2018, Las Palmas de Gran Canaria. **Proceedings...** . Las Palmas de Gran Canaria: 2018.

NANDESHWAR, Ashutosh; CHAUDHARI, Subodh. **Enrollment Prediction Models Using Data Mining**. 2009. 18 p.

NEVES, Clarissa Eckert Baeta; MARTINS, Carlos Benedito. **Ensino superior no Brasil: uma visão abrangente**. Brasília: Instituto de Pesquisa Econômica Aplicada, 2016. 30 p.

NUNES, Getúlio Tadeu. **Abordagem do marketing de relacionamento no ensino superior: um estudo exploratório**. 2005. 149 f. Dissertação (Mestrado) - Curso de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2005.

OLIVEIRA JÚNIOR, José Gonçalves de. **Identificação de padrões para a análise da evasão em cursos 2015 de graduação usando mineração de dados educacionais**. 2015. 86 f. Dissertação (Mestrado) - Curso de Mestre em Computação Aplicada, Universidade Tecnológica Federal do Paraná, Curitiba, 2015.

ORTIZ, Aurora Sanchez; GUTIERREZ, Marco Antonio; MENESES, Claudio. Using data mining to support the university decision process: a case in a chilean university. In: Americas Conference on Information Systems, 11., 2005, Omaha. **Proceedings...** . Omaha: Universidad Catolica del Norte, 2005. p. 2467 - 2479.

PNE. **Observatório do PNE**. 2019. Disponível em: <<http://www.observatoriodopne.org.br>>. Acesso em: 22 ago. 2019.

PRESTES, Emília Maria da Trindade; FIALHO, Maríllia Gabriella Duarte. Evasão na educação superior e gestão institucional: o caso da Universidade Federal da Paraíba. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 26, n. 100, p.869-889, jul. 2018. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0104-40362018002601104>.

QUINLAN, John Ross. **Generating Production Rules From Decision Trees**. Cambridge: Massachusetts Institute Of Technology, 1987. 4 p.

RAVAL, Kalyani. Data Mining Techniques. **International Journal Of Advanced Research In Computer Science And Software Engineering**, Bhavnagar, v. 2, n. 10, p.0-0, out. 2012.

ROMERO, Cristóbal; VENTURA, Sebastián. Educational Data Mining: A Review of the State of the Art. **IEEE Transactions On Systems, Man, And Cybernetics, Part C (Applications and Reviews)**, v. 40, n. 6, p.601-618, nov. 2010. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/tsmcc.2010.2053532>.

SACCARO, Alice; FRANÇA, Marco Túlio Aniceto; JACINTO, Paulo de Andrade. Fatores Associados à Evasão no Ensino Superior Brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de Ciência, Matemática e Computação e de Engenharia, Produção e Construção em instituições públicas e privadas. **Estudos Econômicos (São Paulo)**, São Paulo, v. 49, n. 2, p.337-373, abr. 2019. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/0101-41614925amp>.

SHYAMALA, K.; ., S.p. Rajagopalan. Data Mining Model for a Better Higher Educational System. **Information Technology Journal**, Faisalabad, v. 5, n. 3, p.560-564, 1 mar. 2006. Science Alert. <http://dx.doi.org/10.3923/itj.2006.560.564>.

SILVA FILHO, Roberto Leal Lobo e et al. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, São Paulo, v. 37, n. 132, p.641-659, dez. 2007. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-15742007000300007>.

SILVA, Adriano Maniçoba da; SANTOS, Beatriz Carolini Silva. Eficácia de políticas de acesso ao ensino superior privado na contenção da evasão. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, Sorocaba, v. 22, n. 3, p.741-757, dez. 2017. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1414-40772017000300009>.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 2005. 138 p.