

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS
EDUCACIONAIS EM REDE – MESTRADO PROFISSIONAL

Jáder Adiel Schmitt

**IDENTIFICAÇÃO DE ALUNOS COM TENDÊNCIA À EVASÃO NOS
CURSOS DE GRADUAÇÃO A DISTÂNCIA POR MEIO DE
MINERAÇÃO DE DADOS EDUCACIONAIS**

Santa Maria, RS
2018

Jáder Adiel Schmitt

**IDENTIFICAÇÃO DE ALUNOS COM TENDÊNCIA À EVASÃO NOS CURSOS DE
GRADUAÇÃO A DISTÂNCIA POR MEIO DE MINERAÇÃO DE DADOS
EDUCACIONAIS**

Dissertação apresentada ao Curso de Mestrado Profissional do Programa de Pós-Graduação em Tecnologias Educacionais em Rede, na Área de Concentração Tecnologias Educacionais em Rede para Inovação e Democratização da Educação, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Tecnologias Educacionais em Rede**.

Orientadora: Profa. Dra. Giliane Bernardi

Coorientador: Dr. Gustavo Zanini Kantorski

Santa Maria, RS
2018

Schmitt, Jäder Adiel
IDENTIFICAÇÃO DE ALUNOS COM TENDÊNCIA A EVASÃO NOS
CURSOS DE GRADUAÇÃO A DISTÂNCIA POR MEIO DE MINERAÇÃO DE
DADOS EDUCACIONAIS / Jäder Adiel Schmitt.- 2018.
173 p.; 30 cm

Orientadora: Giliane Bernardi
Coorientador: Gustavo Zanini Kantorski
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Educação, Programa de Pós-Graduação em
Tecnologias Educacionais em Rede, RS, 2018

1. Mineração de Dados Educacionais 2. Previsão de
Evasão na EaD 3. Gestão Educacional na EaD I. Bernardi,
Giliane II. Kantorski, Gustavo Zanini III. Título.

sistema de geração automática de ficha catalográfica da unsm. dados fornecidos pelo
autor(a). sob supervisão da direção da divisão de processos técnicos da biblioteca
central. bibliotecária responsável paula schoenfeldt ratta cxa 10/1728.

Jáder Adiel Schmitt

**IDENTIFICAÇÃO DE ALUNOS COM TENDÊNCIA A EVASÃO NOS
CURSOS DE GRADUAÇÃO A DISTÂNCIA POR MEIO DE
MINERAÇÃO DE DADOS EDUCACIONAIS**

Dissertação apresentada ao Curso de Mestrado Profissional do Programa de Pós-Graduação em Tecnologias Educacionais em Rede, na Área de Concentração Tecnologias Educacionais em Rede para Inovação e Democratização da Educação, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Tecnologias Educacionais em Rede**.

Aprovado em 12 de junho de 2018:



Giliane Bernardi, Profa. Dra. (UFSM)
(Presidente/Orientadora)



Gustavo Zanini Kantorski, Dr. (UFSM)
(Coorientador)



Solange de Lurdes Pertile, Profa. Dra. (UFSM)



Gilse Antoninha Morgental Falkembach, Profa. Dra. (Faculdades QI)

Santa Maria, RS
2018

Dedico esta dissertação:

Aos meus pais, Aerton Schmitt e Eliane Schmitt que conduziram e incentivaram a minha educação e crescimento profissional.

À minha esposa, Cibele Viero Schmitt, pelo apoio incondicional e constante incentivo.

Ao meu filho, Fernando Viero Schmitt, que mesmo recém-chegado a este mundo permitiu a conclusão deste trabalho.

AGRADECIMENTOS

Primeiramente agradeço a Deus por ter me concedido a oportunidade de chegar até aqui para a concretização de mais esse sonho.

Em segundo lugar, gostaria de agradecer aos meus pais, Aerton e Eliane, que sempre primaram pelos meus estudos. Sem o incentivo deles com certeza não teria findado esta etapa da minha vida profissional.

À minha esposa, Cibele Viero Schmitt, por estar ao meu lado o tempo todo. Nos momentos mais difíceis sempre me fez acreditar que concluiria esta jornada.

À minha orientadora, Profa. Dra. Giliane Bernardi, pela paciência, oportunidade, incentivo, sugestões e orientações durante esta caminhada.

Ao meu coorientador, Dr. Gustavo Zanini Kantorski, pela constante disposição em ajudar e por todos os ensinamentos oferecidos.

Ao Núcleo de Tecnologias Educacionais da UFSM, em especial ao Alexandre Schlottgen, pela concessão dos dados e disposição em prestar quaisquer esclarecimentos em relação às informações sempre que necessário.

À banca examinadora, Profa. Dra. Solange Pertile e Profa. Dra. Gilse Morgental Falkembach, por todas as orientações e apontamentos.

Aos coordenadores dos cursos EaD da UFSM analisados, Profa. Dra. Vanessa Ribas Fialho e Prof. Dr. Marcus Vinicius Liessem Fontana (Letras – Espanhol), Profa. Dra. Andréa Forgiarini Cechin (Pedagogia) e Prof. José Luiz Padilha Damilano (Educação Especial), por concordarem com uso dos dados para o desenvolvimento da pesquisa.

E, por fim, a todos que, direta ou indiretamente, se fizeram presentes e contribuíram para a conclusão deste trabalho.

Aprendizagem é um produto da interação.

(Tanya Elias)

RESUMO

IDENTIFICAÇÃO DE ALUNOS COM TENDÊNCIA À EVASÃO NOS CURSOS DE GRADUAÇÃO A DISTÂNCIA POR MEIO DE MINERAÇÃO DE DADOS EDUCACIONAIS

AUTOR: Jáder AdiéI Schmitt
ORIENTADORA: Giliane Bernardi
COORIENTADOR: Gustavo Zanini Kantorski

A evasão é um dos grandes desafios constantes no contexto educacional em todas as modalidades. Por outro lado, o crescente aumento na quantidade de dados gerados por sistemas como Ambientes Virtuais de Ensino e Aprendizagem (AVEA) vem trazendo destaque para a área de Mineração de Dados Educacionais (MDE). Esse fato pode ser constatado por meio de diversos trabalhos que vêm sendo desenvolvidos nesta área com o propósito de prever alunos com tendência à evasão. No entanto, a grande maioria se detém nos aspectos técnicos pertinentes à mineração. Nessa perspectiva, o trabalho visa propiciar aos gestores educacionais dados estratégicos por meio de aplicação de MDE, para que eles possam avaliar, refletir e gerar ações para mitigar o processo de evasão. Para tanto, foi realizada uma pesquisa bibliográfica com a intenção de compreender em detalhes temas relacionados à EaD, evasão, bem como suas causas. Logo em seguida foi realizada uma Revisão Sistemática de Literatura (RSL) visando conhecer as abordagens tecnológicas aplicadas na previsão de evasão. Na sequência, foram abordados os aspectos relacionados a descoberta de conhecimento em base de dados, trabalhos correlatos e aspectos metodológicos. O desenvolvimento foi realizado por meio de dois experimentos abrangendo três cursos de graduação em que foram empregados dados de interações dos alunos no AVEA e dados do sistema de gestão acadêmico da instituição. Como resultados foram obtidos bons índices de acertos, permitindo assim, a conclusão de que a abordagem proposta é factível para detecção de alunos com tendência à evasão em cursos de graduação EaD. Por fim, para a disponibilização dos dados oriundos da mineração foi desenvolvido um aplicativo para dispositivos móveis. Na avaliação do aplicativo, junto aos coordenadores dos cursos avaliados, constatou-se que os dados descobertos são de suma importância para gestão, pois facilita o acompanhamento dos alunos com tendência à evasão, para realização de intervenções junto a eles a fim de evitar a sua desistência. A dissertação de mestrado apresentada está inserida na linha de pesquisa Desenvolvimento de Softwares Educacionais, do Programa de Pós-Graduação em Tecnologias Educacionais em Rede, e gerou como produtos o próprio texto aqui apresentado, bem como a estratégia de MDE criada e o aplicativo desenvolvido.

Palavras-chave: Mineração de Dados Educacionais, Previsão de Evasão na EaD, Gestão Educacional na EaD.

ABSTRACT

IDENTIFICATION OF STUDENTS WITH EVASION TREND IN THE DISTANCE GRADUATION COURSES BY MEANS OF EDUCATIONAL DATA MINING

AUTHOR: Jáder Adiel Schmitt
ADVISOR: Giliane Bernardi
CO-ADVISOR: Gustavo Zanini Kantorski

The evasion is one of the great constant challenges in the educational context in all the modalities. On the other hand, the increase in the amount of data generated by systems such as Virtual Environment for Teaching Learning (AVEA in Portuguese) has been highlighting the area of Educational Data Mining (EDM). This fact can be verified through several studies that have been developed in this area with the purpose of predicting students with tendency to evasion. However, the vast majority detain on the technical aspects pertaining to mining. In this perspective, the work aims provide to educational manager's strategic data, through the application of EDM, so that they can evaluate, reflect and generate actions to mitigate the evasion process. In order to do so, a bibliographical research was carried out with the intention of comprehending in detail themes related to distance education, evasion as well as its causes. Soon afterwards a Systematic Literature Review (SLR) was carried out aiming to know the technological approaches applied in the prediction of evasion. In the sequence, the aspects related to the discovery of knowledge in databases, related works and methodological aspects were approached. The development was carried out through two experiments covering three undergraduate courses in which data were used of student interactions in AVEA and data from the institution's academic management system. As results, good indexes of correct answers were obtained, thus allowing the conclusion that the proposed approach is feasible for the detection of students with a tendency to evasion in undergraduate courses. Lastly, an application for mobile devices was developed for the availability of mining data. In the assessment of the application, along with the coordinators of the evaluated courses, it was found that the data discovered are of paramount importance for management because it facilitates the follow-up of students with a tendency to evasion to carry out interventions with them in order to avoid their dropping out. The dissertation presented is part of the research line Development of Educational Softwares, of the Post-Graduate Program in Educational Technologies in Network, and generated as products the text presented here, as well as the EDM strategy created and the application developed.

Keywords: Educational Data Mining, Prediction of Evasion in Distance Education, Educational Management in Distance Education.

LISTA DE ILUSTRAÇÕES

Figura 1 — Etapas da pesquisa.....	33
Figura 2 — Gerações da EaD	38
Figura 3 — Modelo de parceria da UAB envolvendo instâncias do MEC, universidades e entes do governo	42
Figura 4 — Modelo de funcionamento da UAB	43
Figura 5 — Interface MOODLE – UAB – UFSM	52
Figura 6 — Processo de descoberta de conhecimento em bases de dados.....	69
Figura 7 — Interface gráfica do WEKA	78
Figura 8 — <i>Semi-Supervised Prediction Tool</i>	81
Figura 9 — Interface da tela de lista de predição de evasão	82
Figura 10 — Metodologia CRISP-DM adaptada para evasão	90
Figura 11 — Esquema de integração das fontes de dados utilizadas	98
Figura 12 — Árvore de decisão - Primeiro experimento	125
Figura 13 — Árvore de decisão - Segundo experimento	126
Figura 14 — Arquitetura do sistema	131
Figura 15 — Diagrama de casos de uso	132
Figura 16 — Fluxo de utilização	133
Figura 17 — Menu principal	134
Figura 18 — Dados do curso.....	135
Figura 19 — Seleção de dados	136
Figura 20 — Lista de polos	137
Figura 21 — Menu acesso detalhes dos polos.....	137
Figura 22 — Lista de alunos do polo	138
Figura 23 — Dados do polo e do aluno.....	139

LISTA DE GRÁFICOS

Gráfico 1 — Instituições que informaram conhecer os motivos da evasão dos seus alunos, em percentual, por categoria administrativa	45
Gráfico 2 — Número de trabalhos aceitos e não aceitos na seleção inicial e pós-leitura	58
Gráfico 3 — Quantidade de publicações por ano e critério de inclusão	59
Gráfico 4 — Quantidade de artigos por fonte de dados	60
Gráfico 5 — Dados utilizados nos trabalhos analisados	61
Gráfico 6 — Algoritmos mais utilizados nas abordagens analisadas.....	64
Gráfico 7 — Evasão anual média – Primeiro experimento	96
Gráfico 8 — Média de utilização dos recursos do AVEA	107
Gráfico 9 — Evasão geral – Segundo experimento	109
Gráfico 10 — Evasão anual média – Segundo experimento	109
Gráfico 11 — Quantidade média dos dados utilizados na MD – Letras – Espanhol	119
Gráfico 12 — Quantidade média dos dados utilizados na MD – Pedagogia	120
Gráfico 13 — Quantidade média dos dados utilizados na MD – Educação Especial	120
Gráfico 14 — Quantidade média dos dados utilizados na MD – Todos	121

LISTA DE QUADROS

Quadro 1 — Mapeamento entre etapas e os objetivos da pesquisa.....	34
Quadro 2 — Causas endógenas da evasão na EaD	47
Quadro 3 — Causas exógenas da evasão na EaD	48
Quadro 4 — Questões de pesquisa	55
Quadro 5 — Questões específicas de MD.....	56
Quadro 6 — Expressões de busca	57
Quadro 7 — Critérios de inclusão	57
Quadro 8 — Critérios de exclusão	57
Quadro 9 — Grupos de interesses na MDE	79
Quadro 10 — Análise dos trabalhos relacionados (Parte 1).....	85
Quadro 11 — Análise dos trabalhos relacionados (Parte 2).....	86
Quadro 12 — Atributos armazenados nas planilhas disponibilizadas do AVEA MOODLE ..	98
Quadro 13 — Unificação atributos AVEA MOODLE	101
Quadro 14 — Atributos finais selecionados para MD – Primeiro experimento	102
Quadro 15 — Atributos finais selecionados para MD – Segundo experimento	111
Quadro 16 — Requisitos funcionais.....	130
Quadro 17 — Requisitos não-funcionais.....	130
Quadro 18 — Respostas das questões de avaliação	142
Quadro 19 — Comparativo entre as características dos trabalhos correlatos	145

LISTA DE TABELAS

Tabela 1 — Métricas de estimativa de desempenho dos algoritmos utilizados – Primeiro experimento	103
Tabela 2 — Algoritmos utilizados <i>versus</i> evasões previstas <i>versus</i> acertos – Primeiro experimento	104
Tabela 3 — Métricas de avaliação da previsão – Primeiro experimento	105
Tabela 4 — Conjuntos de dados.....	112
Tabela 5 — Métricas de estimativa de desempenho dos algoritmos utilizados – Segundo experimento	113
Tabela 6 — Relação de alunos aptos à matrícula <i>versus</i> evadidos – Segundo experimento .	114
Tabela 7 — Algoritmos utilizados <i>versus</i> evasões previstas <i>versus</i> acertos – Segundo experimento	115
Tabela 8 — Métricas de avaliação da previsão – Segundo experimento	117

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
ARS	Análise de Redes Sociais
AVEA	Ambiente Virtual de Ensino e Aprendizagem
CAPES	Coordenação de Aperfeiçoamento de Nível Superior
CPD	Centro de Processamento de Dados
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
CSV	<i>Comma-Separated Values</i>
DM	<i>Data Mining</i>
EaD	Educação a Distância
EDM	<i>Educational Data Mining</i>
ER	Entidade-Relacionamento
GPL	<i>General Public License</i>
HTML	<i>HyperText Markup Language</i>
IES	Instituição de Ensino Superior
IGC	Índice Geral de Cursos
KDD	<i>Knowledge Discovery in Databases</i>
LDB	Lei de Diretrizes e Bases da Educação Nacional
LMS	<i>Learning Management System</i>
MD	Mineração de Dados
MDE	Mineração de Dados Educacionais
MEC	Ministério da Educação
MOODLE	<i>Modular Object-Oriented Dynamic Learning</i>
NTE	Núcleo de Tecnologia Educacional
ONG	Organização Não Governamental
PDI	Plano de Desenvolvimento Institucional
PNAP	Programa Nacional de Formação em Administração Pública
PROGRAD	Pró-reitoria de Graduação
PRONATEC	Programa Nacional de Acesso ao Ensino Técnico e Emprego
RBIE	Revista Brasileira de Informática na Educação
RENOTE	Revista de Novas Tecnologias na Educação
RSL	Revisão Sistemática de Literatura
RUP	<i>Rational Unified Process</i>
SBIE	Simpósio Brasileiro de Informática na Educação
SQL	<i>Structured Query Language</i>
SVM	<i>Support Vector Machine</i>
TIC	Tecnologia da Informação e Comunicação
UAB	Universidade Aberta do Brasil
UFPEL	Universidade Federal de Pelotas
UFRJ	Universidade Federal do Rio de Janeiro
UFSM	Universidade Federal de Santa Maria
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
WMDE	<i>Workshop de Mineração de Dados Educacionais</i>
XP	<i>EXtreme Programming</i>

SUMÁRIO

1	INTRODUÇÃO	29
1.1	PROBLEMA DE PESQUISA	31
1.2	OBJETIVOS	31
1.3	METODOLOGIA DE PESQUISA	32
1.3.1	Etapas da pesquisa	33
1.4	JUSTIFICATIVA	35
1.5	PRODUTOS DE PESQUISA	35
1.6	ORGANIZAÇÃO DO TRABALHO	36
2	A EVASÃO NA EAD	37
2.1	EDUCAÇÃO A DISTÂNCIA (EAD)	37
2.1.1	Perspectiva histórica	37
2.1.2	Definição e características da EaD	38
2.1.3	Universidade Aberta do Brasil (UAB)	41
2.2	EVASÃO NA EAD	44
2.3	AMBIENTE VIRTUAL DE ENSINO E APRENDIZAGEM (AVEA)	50
2.3.1	MOODLE	51
3	ABORDAGENS TECNOLÓGICAS COMO ESTRATÉGIA PARA DETECÇÃO E PREVENÇÃO DA EVASÃO NA EAD	55
3.1	METODOLOGIA	55
3.1.1	Planejamento	55
3.1.2	Execução	58
3.2	RESULTADOS OBTIDOS E DISCUSSÕES	59
3.2.1	Resposta às questões de pesquisa	60
3.2.1.1	Q1 – Quais dados foram utilizados para análise?	60
3.2.1.2	Q2 – Quais os resultados obtidos por meio da abordagem empregada?	62
3.2.1.3	Q3 – Que métodos de avaliação foram empregados?	63
3.2.1.4	Q4 – Nas abordagens que empregam MD, quais são os algoritmos mais utilizados?	64
3.2.1.5	Q5 – Quais ferramentas foram utilizadas na MD?	65
3.3	CONSIDERAÇÕES	65
4	DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS	67
4.1	OBJETIVOS DO PROCESSO DE KDD	68
4.2	PROCESSO DE KDD	69
4.2.1	Base de dados	70
4.2.2	Preparação ou pré-processamento de dados	70
4.2.2.1	Limpeza dos dados	70
4.2.2.2	Integração	70
4.2.2.3	Seleção	71
4.2.2.4	Transformação	71
4.2.2.4.1	Padronização	71
4.2.2.4.2	Normalização	71
4.2.3	Mineração de Dados (MD)	72
4.2.3.1	Associação	72
4.2.3.2	Classificação	73
4.2.3.3	Regressão	73
4.2.3.4	Agrupamento	73
4.2.3.5	Sumarização	74

4.2.3.6	Técnica para análise dos dados da pesquisa.....	74
4.2.3.6.1	Avaliação de algoritmos de classificação	74
4.2.4	Avaliação ou validação do conhecimento.....	76
4.3	FERRAMENTAS PARA KDD	77
4.4	MINERAÇÃO DE DADOS EDUCACIONAIS (MDE).....	78
5	TRABALHOS CORRELATOS.....	81
5.1	DESCRIÇÃO DOS TRABALHOS	81
5.2	ANÁLISE DOS TRABALHOS CORRELATOS	84
6	ASPECTOS METODOLÓGICOS.....	89
6.1	A UNIVERSIDADE FEDERAL DE SANTA MARIA (UFSM)	89
6.2	METODOLOGIA PARA KDD.....	90
6.3	TECNOLOGIAS UTILIZADAS	92
6.4	METODOLOGIA DE DESENVOLVIMENTO DA FERRAMENTA	92
6.5	METODOLOGIA DE AVALIAÇÃO	93
7	DESENVOLVIMENTO	95
7.1	CONTEXTO	95
7.2	PRIMEIRO EXPERIMENTO	96
7.2.1	Entendimento da evasão.....	96
7.2.2	Entendimento dos dados de alunos	97
7.2.3	Preparação dos dados	99
7.2.4	Modelagem.....	103
7.2.5	Avaliação.....	104
7.2.6	Disponibilização	108
7.3	SEGUNDO EXPERIMENTO	108
7.3.1	Entendimento da evasão.....	108
7.3.2	Entendimento dos dados de alunos	110
7.3.3	Preparação dos dados	110
7.3.4	Modelagem.....	112
7.3.5	Avaliação.....	114
7.3.6	Disponibilização	122
7.4	ANÁLISE DOS EXPERIMENTOS	122
8	<i>IPERMANENCY: UM APLICATIVO PARA DISPONIBILIZAÇÃO DOS DADOS DA MINERAÇÃO</i>	129
8.1	ETAPAS DE DESENVOLVIMENTO.....	129
8.1.1	Exploração	129
8.1.2	Inicialização	131
8.1.3	Produção	133
8.1.3.1	Menu principal	134
8.1.3.2	Dados do curso.....	134
8.1.3.3	Seleção de dados	135
8.1.3.4	Lista de polos	136
8.1.3.5	Lista de alunos do polo	138
8.1.3.6	Dados do polo e do aluno.....	138
8.1.4	Estabilização e testes	139
8.2	AVALIAÇÃO.....	140
8.2.1	Levantamento dos objetivos.....	140
8.2.2	Definição das questões	140
8.2.3	Escolha dos métodos	141
8.2.4	Identificação das questões práticas	141
8.2.5	Análise das questões éticas	141

8.2.6	Avaliação, interpretação e apresentação dos resultados.....	141
9	COMPARAÇÃO COM OS TRABALHOS CORRELATOS	145
10	CONCLUSÃO.....	147
10.1	TRABALHOS FUTUROS	149
11	ARTIGOS PUBLICADOS	151
	REFERÊNCIAS	153
	APENDICE A – TRABALHOS SELECIONADOS PÓS-LEITURA	161
	ANEXO A – AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS ACADÊMICOS DO SISTEMA DE INFORMAÇÃO INSTITUCIONAL	167
	ANEXO B – AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS DE INTERAÇÕES DO MOODLE – LETRAS ESPANHOL	161
	ANEXO C – AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS DE INTERAÇÕES DO MOODLE – PEDAGOGIA.....	171
	ANEXO D - AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS DE INTERAÇÕES DO MOODLE – EDUCAÇÃO ESPECIAL	173

1 INTRODUÇÃO

Significativas transformações sociais estão ocorrendo por meio da democratização global no emprego das Tecnologias de Informação e Comunicação (TIC) em rede. A popularização da Internet e o desenvolvimento tecnológico permitiu o aprimoramento de muitas áreas diante das possibilidades que a informação passou a oferecer.

Esta situação não é diferente em relação ao contexto educacional. O desenvolvimento tecnológico, aliado ao apoio das políticas públicas, faz com que a Educação a Distância (EaD) esteja se tornando cada vez mais consolidada no Brasil, atraindo assim, um número maior de adeptos. Em uma perspectiva em que o cotidiano das pessoas está cada vez mais atribulado e nem sempre os ambientes educacionais encontram-se geograficamente presentes, a EaD torna-se a melhor opção para quem não tem condições de se deslocar para cumprir as exigências de um curso presencial.

O Censo da EaD 2016, realizado no ano de 2015 pela Associação Brasileira de Educação a Distância (2016), demonstra o avanço da EaD através do número de alunos registrados em cursos desta modalidade no país. Em 2015 foram contabilizados 5.048.912 alunos, o que representa um aumento de 30,5% em relação a 2014. Além disso, os resultados demonstram que, no mesmo período, enquanto o ensino presencial teve um crescimento de 2,3% nas matrículas, a EaD teve uma expansão de 3,9%. Todavia, dados do censo também revelam um alto índice de evasão, sendo que a maioria das instituições pesquisadas apresenta taxas de evasão entre 26% e 50%. Diante das elevadas taxas registradas, a evasão vem tornando-se um dos principais temas de preocupação no âmbito da EaD no Brasil.

O cenário da evasão está presente em todas as modalidades de ensino, seja ela presencial, semipresencial ou a distância (BITTENCOURT e MERCADO, 2014). Devido a relevância e a complexidade do assunto, diversos trabalhos relacionados ao tema têm sido desenvolvidos, sendo que, em sua ampla maioria, aplicados às Instituições de Ensino Superior (IES).

A maior parte desses trabalhos buscam compreender os fatores que causam a evasão (SILVEIRA, 2012; BIZARRIA, SILVA e CARNEIRO, 2014; ABBAD, ZERBINI e SOUZA, 2010; MARTINS, GEBRAN e TERÇARIOL, 2014). No contexto da EaD, entre os vários fatores levantados, destacam-se a insatisfação com o tutor, dificuldade de acesso à complexidade das atividades, dificuldade de assimilação da cultura inerente à falha na elaboração do curso, expectativas erradas por parte dos alunos, tecnologia inadequada, falta de

habilidade para usar a tecnologia corretamente, tempo de realizar os estudos, nível de estudo anterior à matrícula, restrições de mercado de trabalho, entre outros.

Outros trabalhos dos autores Abu-Oda e El-Halees (2015); Azoumana (2014); Jadrić, Garača e Čukušić (2010); Kantorski *et al.* (2016); Orea, Vargas e Alonso (2005) abordam a evasão no sentido de prever a sua ocorrência no ensino superior presencial. Geralmente, esses trabalhos utilizam técnicas de descoberta de conhecimento em base de dados, quer seja através do inglês *Knowledge Discovery in Databases* (KDD) ou Mineração de Dados (MD), como também do inglês *Data Mining* (DM), para realizar a previsão de resultados.

Estudos semelhantes também estão sendo dirigidos aos cursos de graduação a distância, como pode ser visto nos trabalhos de Detoni, Araújo e Cechinel (2015); Ferreira e Elia (2013); Heidrich *et al.* (2007); Queiroga *et al.* (2016); Reino *et al.* (2015); Rigo e Cazella (2014). No entanto, boa parte destes trabalhos se restringem aos aspectos relacionados aos modelos de predição obtidos a partir do processo de MD. Eles não fornecem, assim, uma ferramenta para que os dados oriundos desses modelos possam ser analisados por gestores para auxiliar nas tomadas de decisões. Dentre algumas informações aplicadas na MD dos trabalhos listados destacam-se a frequência de uso do Ambiente Virtual de Ensino e Aprendizagem (AVEA) por parte dos alunos e professores, notas, disciplinas que mais influenciaram na evasão do curso de graduação, dentre outras.

Por outro lado, o crescente aumento na quantidade de dados gerados diariamente pelos sistemas informacionais vem trazendo maior destaque para a área de MD. A MD consiste na exploração de grandes quantidades de dados com o objetivo de detectar padrões que permitem a extração de novos conhecimentos (SILBERSCHATZ, KORTH e SUDARSHAN, 2006). Como exemplo, a MD pode ser aplicada em uma instituição financeira para prever o comportamento de clientes que desejam obter um novo empréstimo. Com base em dados históricos de seus antigos clientes é possível prever se um cliente novo será um bom pagador ou não. Dessa forma, em uma avaliação de crédito, é possível proceder com a concessão de créditos somente a clientes que são enquadrados como bons pagadores.

A aplicação de MD para exploração de dados em ambientes educacionais vem sendo designada na literatura como Mineração de Dados Educacionais (MDE), do inglês *Educational Data Mining* (EDM). A MDE busca adaptar métodos e algoritmos convencionais de MD com o objetivo de propiciar uma melhor compreensão em relação aos dados do contexto educacional (COSTA *et al.*, 2013).

O emprego das técnicas de MDE em dados gerados pelos AVEA, que consiste nos ambientes onde ocorrem grande parte das interações entre os discentes e os tutores/docentes na

EaD, pode possibilitar a descoberta de conhecimentos relacionados aos alunos com tendência à evasão para que ações possam ser desempenhadas a fim de evitar a sua ocorrência.

1.1 PROBLEMA DE PESQUISA

O êxito na formação do estudante é o propósito principal de qualquer programa de formação. Deste modo, os gestores educacionais necessitam de informações qualificadas que identifiquem tanto as causas de sucesso quanto as do insucesso estudantil.

Dentre as principais causas de insucesso está a evasão que, de acordo com Silva Filho *et al.* (2007), é um dos grandes desafios a ser superado no campo da educação, pois cada vez mais são aplicados recursos escassos na atração e captação de alunos e os resultados têm sido influenciados negativamente diante das elevadas taxas de evasão. O abandono do aluno representa prejuízo para si próprio ao não se diplomar, para o professor, que não atinge o seu objetivo como educador, para a instituição, pelo não atendimento da sua missão, para a sociedade, pelas perdas sociais/econômicas e para a família, pelo sonho não realizado.

No entanto, apesar do impacto negativo que a evasão provoca para todos os envolvidos, as suas causas não são amplamente conhecidas pelas IES, conforme aponta o Censo da EaD realizado pela Associação Brasileira de Educação a Distância (2017). Neste contexto, o censo também demonstra um cenário ainda mais agravante em relação as IES públicas onde apenas 46% afirmam ter conhecimento sobre as causas do abandono.

Diante destas considerações, a presente pesquisa busca responder a seguinte questão norteadora: de que modo as informações oriundas do processo de MDE podem apoiar os gestores na criação de estratégias para mitigar o processo de evasão de alunos nos cursos de graduação a distância em uma IES pública?

1.2 OBJETIVOS

O objetivo geral deste trabalho visa propiciar aos gestores educacionais dados estratégicos para que eles possam avaliar, refletir e gerar ações para mitigar o processo de evasão em cursos de graduação a distância.

Para isso foram delineados os seguintes objetivos específicos:

- a) Revisar o estado da arte relacionado à compreensão das causas da evasão educacional;

- b) Revisar o estado da arte sobre o emprego de técnicas de MD para prever a evasão educacional;
- c) Identificar as informações disponíveis no AVEA que podem influenciar no processo de previsão da evasão;
- d) Realizar a previsão de evasão de alunos de graduação da modalidade EaD, empregando técnicas de MDE, considerando os dados presentes no AVEA utilizado por uma IES pública;
- e) Desenvolver uma ferramenta para disponibilizar os resultados oriundos da previsão; e,
- f) Avaliar a ferramenta, bem como os dados disponibilizados de forma a verificar a sua contribuição para os gestores no que tange a análise e tomadas de decisões relacionadas à evasão.

1.3 METODOLOGIA DE PESQUISA

Geralmente, uma pesquisa científica pode ser classificada quanto à natureza, à abordagem do problema, aos objetivos e aos procedimentos técnicos (PRODANOV e FREITAS, 2013). Nessa concepção, esta pesquisa pode ser classificada, quanto a sua natureza, como aplicada – levando em consideração que busca verificar como os dados provenientes da aplicação de técnicas de MD podem auxiliar a gestão educacional na mitigação do processo da evasão.

Quanto à forma de abordagem do problema, a pesquisa pode ser classificada como quantitativa – por utilizar dados provenientes da MD para deduzir características dos alunos que tendem a evadir - e qualitativa – já que se objetiva à compreensão em detalhes das causas relacionadas à evasão de alunos da modalidade EaD de uma IES pública como, também, da aplicação de técnicas de MD para identificação da evasão.

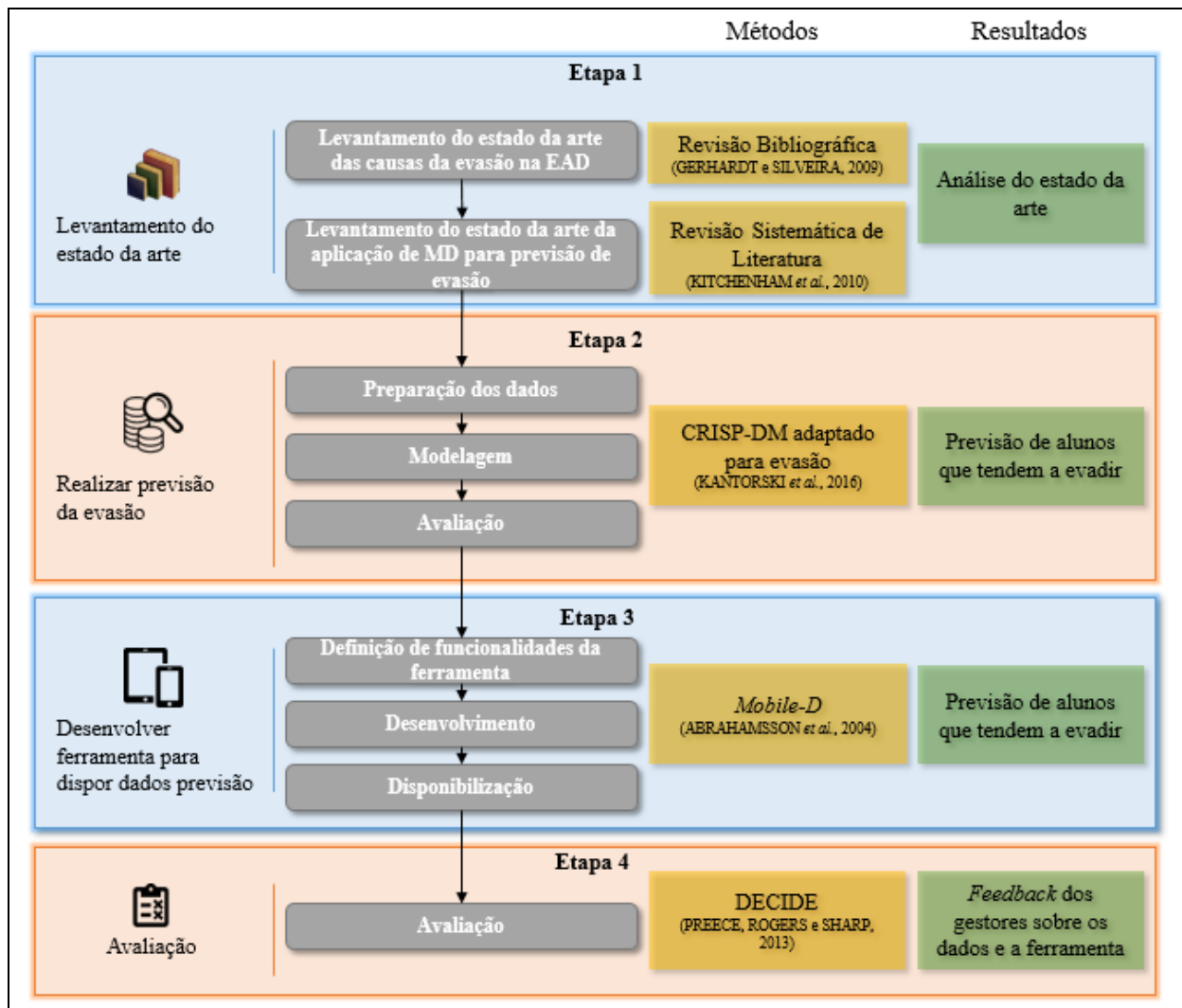
Com relação aos seus objetivos, a pesquisa pode ser classificada como exploratória e descritiva, pois pretende buscar maior familiaridade com o problema de pesquisa por meio do levantamento bibliográfico e, também, tem como objetivo a descrição dos resultados que foram alcançados. Da perspectiva dos procedimentos técnicos, este estudo pode ser classificado como bibliográfico e experimental.

De acordo com as classificações descritas, na próxima seção serão apresentadas as etapas da pesquisa a serem realizadas para atingir o objetivo estabelecido.

1.3.1 Etapas da pesquisa

As etapas realizadas para a condução desta pesquisa são apresentadas na Figura 1.

Figura 1 — Etapas da pesquisa



Fonte: Autor.

A primeira etapa, levantamento do estado da arte, visa conhecer as abordagens existentes que buscam a identificação das causas da evasão na EaD na educação superior, bem como aquelas que utilizam técnicas de MD para sua identificação. Para isso, no que tange as causas da evasão foi efetuada uma pesquisa bibliográfica. Já para o levantamento das técnicas de MD foi realizada uma Revisão Sistemática de Literatura (RSL) utilizando a metodologia definida por Kitchenham *et al.* (2010). O conhecimento das causas da evasão permite estabelecer quais dados disponíveis no AVEA possuem maior nível de relevância para posterior

utilização no processo de MD. Já o conhecimento das técnicas possibilita analisar aquelas que tiveram maior sucesso no processo de previsão que possam também ser utilizadas nesta pesquisa.

Na segunda etapa, realizar previsão da evasão, foram empregadas as técnicas de MD para identificar os alunos com tendência a abandonar um curso. Essa etapa foi realizada seguindo a metodologia proposta por Kantorski *et al.* (2016) que consiste de uma metodologia específica para MD (CRISP-DM) adaptada para a detecção da evasão. Dentre as atividades previstas na metodologia proposta estão: entendimento dos dados da evasão e dos alunos, atendidas na primeira etapa; preparação dos dados, que compreende os ajustes necessários nos dados para otimizar o processo de MD; modelagem, consiste na MD propriamente dita; e, avaliação, que busca analisar se o processo da etapa anterior apresenta resultados viáveis.

A terceira etapa, desenvolver ferramenta, teve como objetivo construir uma ferramenta para dispor o acesso aos dados procedentes da MD aos gestores de uma IES. Esta etapa foi executada seguindo o método *Mobile-D*, proposto por Abrahamsson *et al.* (2004) para o desenvolvimento de aplicativo para dispositivos móveis. As atividades executadas nesta etapa foram: definição das funcionalidades a serem disponibilizadas pela ferramenta, desenvolvimento e disponibilização. A escolha dos dispositivos móveis como plataforma para construção da ferramenta justifica-se pelo crescimento de sua utilização para os mais diversos fins. Além disso, pela agilidade que ela proporciona ao gestor no acesso das informações.

Por fim, a quarta etapa, avaliar ferramenta, teve como objetivo avaliar, junto aos coordenadores de cursos, aspectos relacionados à utilização da ferramenta, bem como os dados provenientes da MD que foram disponibilizados. Esta fase foi executada utilizando o *framework* DECIDE, proposto por Preece, Rogers e Sharp (2013), que orienta o planejamento, a execução e a análise da avaliação.

O Quadro 1 apresenta a relação entre o mapeamento de cada etapa da pesquisa e o atendimento aos objetivos específicos definidos para este trabalho.

Quadro 1 — Mapeamento entre etapas e os objetivos da pesquisa

Etapa da pesquisa	Objetivo específico
Etapa 1 - Levantamento do estado da arte	a, b
Etapa 2 - Realizar previsão da evasão	c, d
Etapa 3 - Desenvolver ferramenta para dispor dados da evasão	e
Etapa 4 - Avaliar ferramenta	f

1.4 JUSTIFICATIVA

A evasão estudantil na EaD pode ser definida como um fenômeno educacional complexo que está presente em todas as instituições educacionais e em todos os níveis de ensino (BITTENCOURT e MERCADO, 2014). São diversas causas que levam as instituições, públicas ou privadas, a dirigirem suas atenções ao assunto. De acordo com Silva Filho *et al.* (2007), os problemas da evasão na EaD são: para o setor público, os recursos investidos sem o devido retorno; para o setor privado, importante perda de receita; para ambos os setores, fonte de ociosidade de professores, funcionários, equipamentos e, em algumas situações, espaço físico.

Desse modo, verifica-se que é importante a realização de pesquisas sobre o tema, visto que existe a necessidade de uma melhor compreensão a respeito do conceito de evasão, suas variáveis e fatores motivadores. Alia-se a isso a necessidade de proporcionar aos gestores das IES um melhor entendimento sobre o tema, amparados pelos sistemas de informação.

Os avanços tecnológicos nos AVEA possibilitam a obtenção de novos conhecimentos através de processos de KDD. O KDD pode ser visto como o processo de descoberta de padrões e tendências através da análise de grande volume de informações, tendo como principal etapa o processo de MD, consistindo na execução prática de análise e algoritmos que produz uma relação particular de padrões a partir de dados (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).

Diante do acelerado crescimento da EaD no país, a relevância acadêmica da presente pesquisa justifica-se na medida em que se propõe a prever a evasão de alunos de graduação nesta modalidade de ensino através do emprego de técnicas de MD. Além disso, visa disponibilizar uma ferramenta que demonstre as informações provenientes dessa previsão com o objetivo de controlar a evasão, buscando facilitar o entendimento e análise da questão por parte dos envolvidos na gestão acadêmica.

A análise da evasão pode fornecer subsídios para a realização de ações estratégicas que norteiem a direção da educação superior e potencializem a permanência e a conclusão dos cursos.

1.5 PRODUTOS DE PESQUISA

De acordo com a exigência de um mestrado profissional o resultado da pesquisa deve consistir na elaboração de um produto. Assim sendo, os produtos resultantes deste trabalho são:

(a) a dissertação, consistindo de um produto teórico; (b) a abordagem de MDE, também consistindo de um produto teórico; e, (c) um aplicativo para dispositivos móveis para permitir a análise dos resultados provenientes da mineração.

1.6 ORGANIZAÇÃO DO TRABALHO

Esta pesquisa apresenta no capítulo dois os conceitos relacionados à EaD. Dentre esses conceitos são abordados a EaD propriamente dita; a Universidade Aberta do Brasil (UAB), que consiste em um programa do governo federal elaborado com o objetivo de fomentar a EaD e que regem os cursos abordados nesta pesquisa; os AVEA, que são os sistemas computacionais que permitem a interação entre os atores educacionais; por fim, os conceitos relacionados à evasão e as suas causas.

No terceiro capítulo é apresentada uma RSL, destacando-se as abordagens tecnológicas existentes para a previsão de evasão de alunos. No capítulo subsequente, o quarto, são abordados os conceitos de KDD e MDE. Na sequência, no capítulo cinco, são apresentados os trabalhos correlatos ao desenvolvido nesta pesquisa. Os aspectos metodológicos do trabalho são descritos no sexto capítulo. O desenvolvimento do processo de mineração é apresentado no capítulo sete, sendo que no capítulo oito é exposta a ferramenta desenvolvida para disponibilizar os resultados obtidos. O nono capítulo traz uma comparação entre a presente pesquisa e os trabalhos correlatos apresentados no capítulo cinco. Por fim, no capítulo dez são abordadas as conclusões, contribuições e trabalhos futuros provenientes desta dissertação.

2 A EVASÃO NA EAD

Este capítulo aborda os conceitos que serviram de base para a realização desta pesquisa. Inicialmente será exposta a definição da EaD bem como suas principais características. Sucedendo esta definição, será abordada a UAB, que é uma política pública brasileira que fomentou a EaD no ensino superior. Na sequência, serão discutidos aspectos da evasão nesta modalidade de ensino juntamente com algumas pesquisas encontradas na literatura que abordam as suas causas. Finalmente, caracterizam-se os AVEA, que são os ambientes computacionais que permitem a interação entre os atores da EaD.

2.1 EDUCAÇÃO A DISTÂNCIA (EAD)

A evolução da EaD está diretamente relacionada ao aluno e às tecnologias presentes para sua implementação. Diante disso é apresentado, no subcapítulo que segue, a EaD em sua perspectiva histórica, visando compreender as transformações ocorridas desde sua concepção até o modelo como ela é conhecida nos dias de hoje. Na sequência, são abordadas sua definição e principais características.

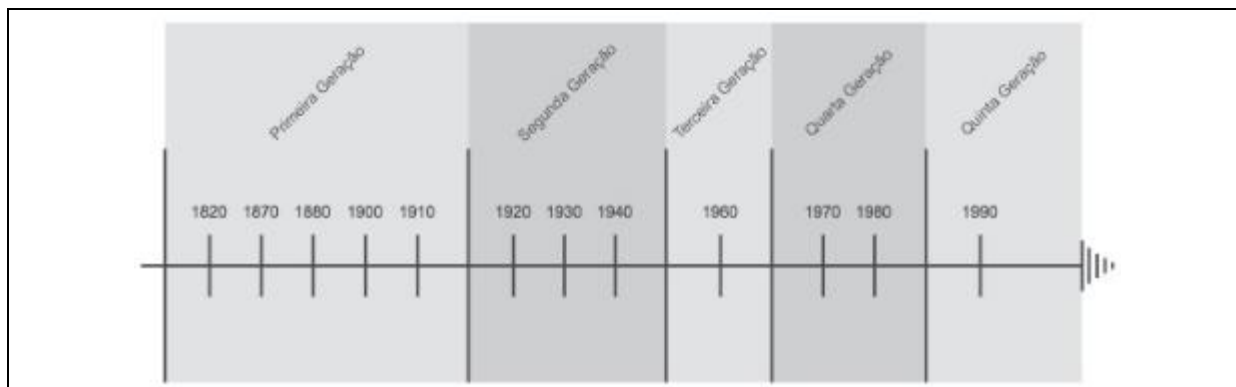
2.1.1 Perspectiva histórica

A EaD surgiu a partir da segunda metade do século XIX, visando oportunizar a qualificação de mão de obra para atender as necessidades que emergiram por meio da industrialização, da mecanização e divisão dos processos de trabalho.

Além das transformações sociais, que exigem maior qualificação por parte das pessoas, a EaD está diretamente atrelada aos avanços tecnológicos. Tais avanços, para Damásio (2008), criaram um processo de cinco gerações de EaD sucessivas, conforme exposto na Figura 2.

Com a popularização dos serviços postais, entre os séculos XIX e XX, surgiu a primeira geração da EaD. Esta geração relacionou-se diretamente com a disponibilização de cursos via correios que permitiu o estudo “independente” e “autônomo” dos indivíduos (CORREIA, 2016).

Figura 2 — Gerações da EaD



Fonte: (MACHADO e MORAES, 2015, p. 10).

Já no início do século XX emergiu a segunda geração da EaD mediante as transmissões via rádio e televisão. Tais dispositivos proporcionaram aos docentes a elaboração de cursos com o uso de som e imagem nos processos (MACHADO e MORAES, 2015).

A partir do surgimento de tecnologias de comunicação, na década de 1960, iniciou-se a terceira geração. Nesta geração a transmissão via satélite permitiu o acesso às informações de forma mais dinâmica, mobilizando a interação e interatividade (DAMÁSIO, 2008).

Relacionada ao acesso à tecnologia e a globalização se estabelecia a quarta geração. Nesta geração, os cursos regulares de EaD passaram a empregar ferramentas que possibilitaram uma comunicação mais ativa e dinâmica, permitindo maior interatividade como teleconferências, fax, páginas de WEB, grupos *on-line* e correios eletrônicos (CORREIA, 2016).

Por fim, surge a geração corrente, a quinta geração, com a popularização da Internet em que a interatividade remodelou os métodos de ensino e aprendizagem com rapidez. Nesta fase emergiram as plataformas de ensino com repositório de conteúdo, estratégias, materiais e ferramentas de ensino diversificando e inovando os AVEA (CORREIA, 2016).

2.1.2 Definição e características da EaD

No Brasil, apesar do ensino via correspondência existir a mais de um século, foi a partir da promulgação da Lei de Diretrizes e Bases da Educação Nacional (LDB) – 9.394, de 20 de dezembro de 1996, que esta modalidade de ensino se tornou oficializada. O Decreto nº 5.622 da Presidência da República, que regulamenta o art. 80 da LDB, destaca em seu primeiro artigo a definição de EaD:

A Educação a Distância é a modalidade educacional na qual a mediação didático-pedagógica nos processos de ensino e aprendizagem ocorre com a utilização de meios e tecnologias de informação e comunicação, com estudantes e professores desenvolvendo atividades educativas em lugares ou tempos diversos (BRASIL, 2005, não paginado).

É possível encontrar na literatura diversas definições para EaD, das quais podem ser extraídos alguns elementos fundamentais para sua caracterização e distinção em relação à educação tradicional. De acordo com Walter (2006), esses elementos são:

- a) Distanciamento físico entre professor e aluno;
- b) Processo de aprendizagem administrado pelo próprio aluno e não pelo professor, que nessa modalidade desempenha um papel de auxiliar e facilitador;
- c) Relação entre professor e aluno é realizada por meio de mídias tecnológicas diversas (televisão, Internet, intranet, rádio, CD-ROM, material impresso são exemplo desses recursos); e,
- d) Dinamicidade de conteúdos a serem apreendidos (o aluno, de acordo com suas necessidades, determina a sua aprendizagem).

No que tange aos objetivos da EaD, de maneira geral, Aretio (1994) elenca:

- a) Democratização do acesso à educação: possibilita o alcance de um grande número de pessoas nos seus locais de trabalho ou meios culturais;
- b) Estímulo permanente à educação: proporciona, por meio de cursos, a atualização profissional ou acadêmica;
- c) Possibilidade de uma aprendizagem autônoma e vinculada à experiência: os alunos são sujeitos da sua própria aprendizagem e definem o currículo a ser seguido de acordo com suas necessidades profissionais e aptidões;
- d) Promoção de um ensino inovador e de qualidade: a EaD propõe novos papéis aos atores envolvidos no processo educacional o que necessita de um maior planejamento nos materiais e no processo instrucional, além de utilização de recursos multimídia, o que pressupõe uma maior qualidade no ensino; e,
- e) Redução de custos com educação: possibilita a redução de custos de locomoção dos alunos já que os encontros presenciais ocorrem de acordo com as suas necessidades.

A EaD propicia a muitas pessoas acesso à educação e conhecimento, democratizando o ensino com qualidade e por lugares onde as instituições de ensino não conseguem chegar. Esse

processo só foi viável com a melhoria da qualidade da Internet e incorporação das TIC na educação que, de acordo com Alves (2011), promoveram um grande avanço dessa modalidade de ensino nos últimos anos, permitindo a implementação de projetos educacionais mais diversos e para as mais complexas situações. Na literatura, os cursos EaD que são ofertados por estes aparatos tecnológicos conectados à Internet estão ganhando novas definições. No caso dos cursos oferecidos por intermédio do computador eles são definidos como *e-Learning* que vem de “*electronic learning*” (do inglês, aprendizagem eletrônica). Recentemente, com a disseminação no uso dos dispositivos móveis, como celulares e *tablets*, para fins educacionais, despontou o surgimento de um novo termo conhecido como *m-Learning* que vem de “*mobile learning*” (do inglês, aprendizagem móvel).

É possível constatar o lugar de destaque que o modelo da EaD vem ganhando no Brasil a partir dos dados relacionados ao número de alunos que efetuaram a matrícula nessa modalidade de ensino no ano de 2015. De acordo com a Associação Brasileira de Educação a Distância (2016), neste ano houve um aumento de 30,5% das matrículas na EaD, totalizando 5.048.912 alunos.

O Censo da EaD ainda demonstra que, no país, no que se refere à oferta de cursos na EaD, o nível com maior número de vagas são os cursos de especialização e graduação. Conforme o Decreto nº 5.622 que regulamenta a EaD no Brasil, os níveis de ensino que podem ser ofertados na modalidade são: educação básica - jovens e adultos, especial e profissional; e, superior - sequenciais, graduação, especialização, mestrado e doutorado.

Apesar de a EaD permitir que o estudante não precise comparecer à Instituição para estudar e assistir as suas aulas, bem como para suprir as suas dúvidas, existem alguns momentos em que a sua presença é obrigatória. Conforme o Decreto nº 5.622, de 19 de dezembro de 2005, esses momentos ocorrem nas seguintes ocasiões: avaliação final dos estudantes, estágios (quando previstos), defesa de trabalho de conclusão de curso e atividades em laboratórios.

Diante da obrigatoriedade destes momentos presenciais é necessário que as instituições que ofertam este tipo de curso disponham de um polo de apoio para a realização destas atividades. As suas estruturas devem preencher todos os requisitos de uma instituição presencial com: acervo da biblioteca atualizado, laboratório de informática com acesso à banda larga e laboratórios de ensino. Além disso, no que tange a duração dos cursos ofertados na modalidade EaD, eles devem possuir a mesma carga horária de curso correspondente ofertado na modalidade presencial.

Conforme a Associação Brasileira de Educação a Distância (2016), apesar da EaD possibilitar o alcance à educação para um maior número de pessoas, dentre os principais

problemas enfrentados pela modalidade no país estão: (a) a evasão, caracterizada pela desistência do estudante do curso; (b) a resistência dos docentes em relação à modalidade, ou seja, a falta de motivação dos professores em atuar na EaD, muitas vezes, devido à falta de domínio das TIC ou, até mesmo, por não acreditar que a EaD propicie o mesmo nível de aprendizagem ofertado no ensino presencial; e (c) a dificuldade de adaptação dos estudantes a essa modalidade, o que ocorre, na maioria das vezes, devido ao pensamento equivocado de que a dedicação a um curso a distância pode ser menor do que a empregada no ensino presencial pela não realização de encontros presenciais. Neste cenário, vale destacar que o foco da presente pesquisa visa proporcionar novos aparatos para enfrentar um dos seus principais problemas, a evasão.

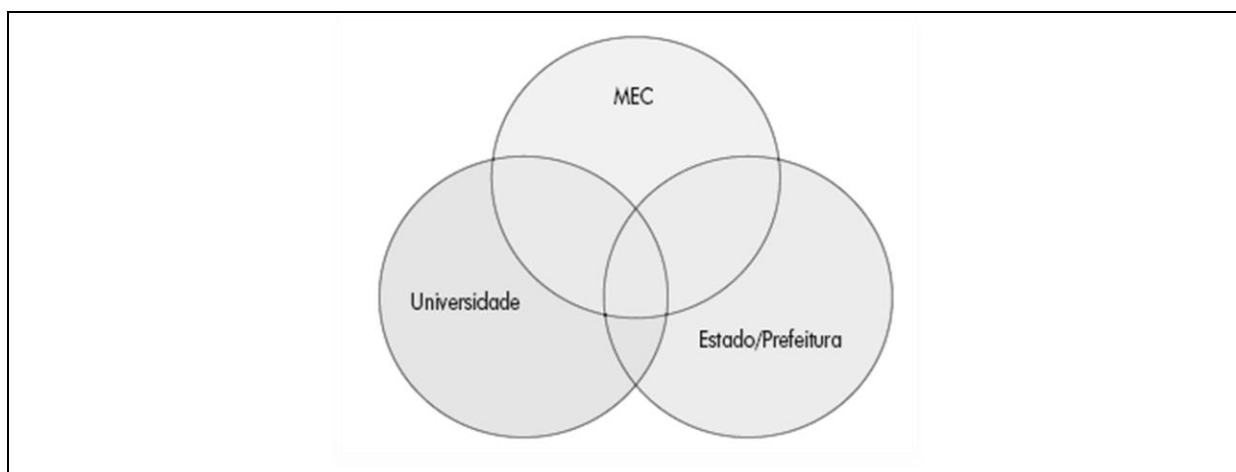
O Governo Federal, através do MEC, lançou diversos programas de EaD, em parceria com outros entes da federação, tendo em vista ampliar e democratizar a oferta de ensino. Dentre eles estão: (a) a UAB, na oferta de cursos de graduação; (b) o Programa Nacional de Acesso ao Ensino Técnico e Emprego (PRONATEC), na oferta de educação profissional e tecnológica; (c) o ProInfantil, para capacitação de profissionais que atuam na sala de aula da educação infantil; (d) a TV Escola, canal da educação destinados a todos interessados em aprender; e, (e) Programa Nacional de Formação em Administração Pública (PNAP), que visa ofertar cursos na área de administração pública.

Este estudo aborda o problema da evasão em cursos de graduação na EaD em uma IES pública. Desse modo, para facilitar o entendimento de algumas diretrizes que operacionalizam estes cursos serão expostos, na próxima seção, maiores detalhes referente à UAB.

2.1.3 Universidade Aberta do Brasil (UAB)

A UAB é um sistema do Governo Federal criado por meio do Decreto nº 5800, de 9 de junho de 2006, para a expansão da formação em nível superior (BRASIL, 2006). Ela é composta por universidades públicas brasileiras, governos estaduais e municipais, conforme pode ser visto na Figura 3.

Figura 3 — Modelo de parceria da UAB envolvendo instâncias do MEC, universidades e entes do governo



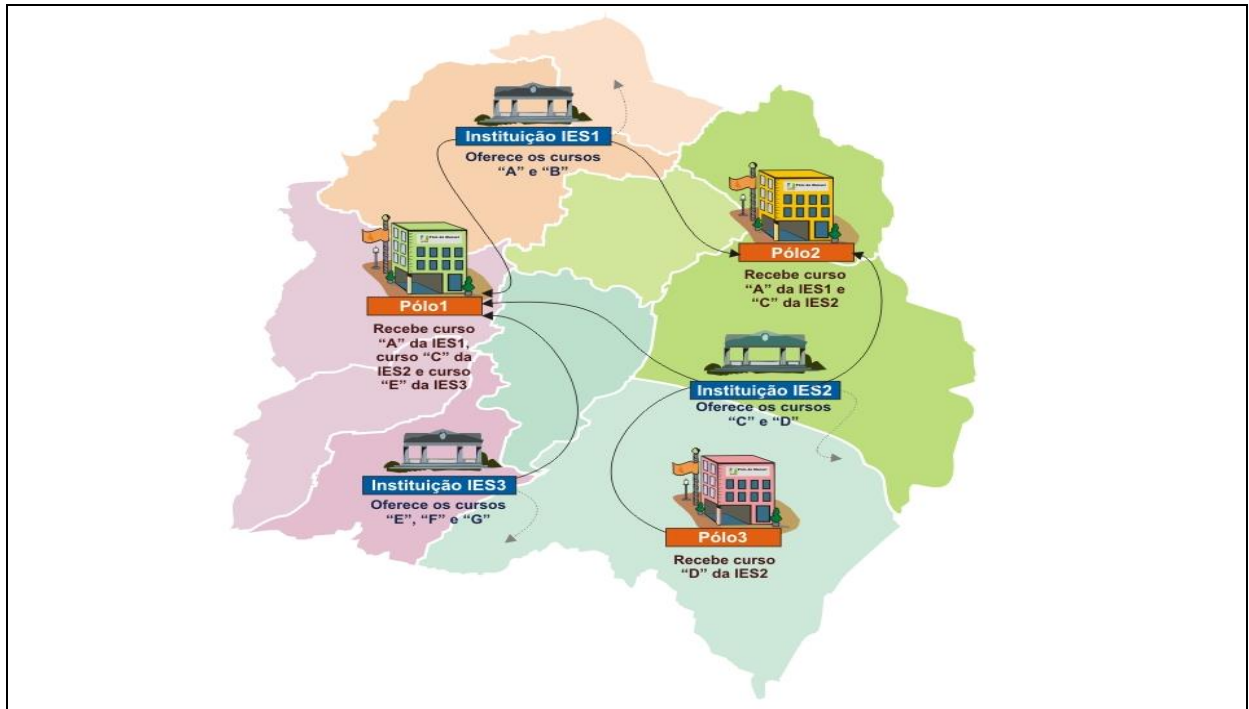
Fonte: (LITTO e FORMIGA, 2012, p. 282).

Ao invés de recomendar a criação de uma nova IES específica para a oferta de cursos EaD, o sistema UAB atua como um organizador entre os três níveis governamentais e as IES públicas existentes com o objetivo de atender as necessidades locais por educação superior (LITTO e FORMIGA, 2012). Conforme BRASIL (2006), a UAB tem como objetivo a expansão e interiorização de cursos de educação superior pública no país, a ampliação do seu acesso e o estímulo à criação de centros de formação permanente por meio de polos de apoio presencial.

A UAB é formada por uma ampla articulação entre as IES, os Estados/Municípios e a Coordenação de Aperfeiçoamento de Nível Superior (CAPES). Nesta articulação, a oferta dos cursos fica sob responsabilidade das IES públicas já estabelecidas em todo o país e com o conhecimento e tradição na oferta de cursos presenciais. Já a manutenção dos polos presenciais fica a cargo dos Estados/Municípios, sendo que à CAPES compete a regulação e o financiamento do sistema UAB (LITTO e FORMIGA, 2012).

O modelo de articulação define quais IES devem ser responsáveis pela oferta de cursos em determinadas regiões por meio de polos de apoio presencial (BRASIL, 2006). Dessa forma, a UAB garante o apoio em determinadas ações a fim de assegurar o bom funcionamento dos cursos. Na Figura 4 é sintetizada a articulação das IES e os polos de ensino.

Figura 4 — Modelo de funcionamento da UAB



Fonte: (COMISSÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2017).

O modelo apresentado permite bastante liberdade para exploração das potencialidades da parceria. Nela é possível observar que um mesmo polo de apoio presencial pode receber um ou mais cursos de uma ou mais IES como pode ser visto nos polos 1 e 2.

No entanto, a flexibilidade proporcionada por este modelo pode representar um grande desafio no âmbito da gestão educacional. É possível demonstrar este aspecto, considerando a situação típica de um curso que é ofertado para cinco ou seis polos, com, usualmente, 30 alunos por polo, onde existirão sempre uma média de 150/180 alunos atendidos por um único professor, 5 a 6 tutores, e um único gestor¹. Neste caso, monitorar e acompanhar todos estes alunos torna-se uma tarefa complexa, o que pode levar a ocorrência da evasão em um número maior do que o apresentado em um mesmo curso na modalidade presencial, em que existem turmas menores.

Diante dos conhecimentos relacionados à EaD e o programa governamental que estabelece as suas diretrizes, vistos até então, na próxima seção serão apresentados os aspectos relacionados à evasão na EaD, bem como as causas que corroboram com a sua ocorrência.

¹ Os números apresentados são meramente ilustrativos já que a definição de quantos alunos devem ser alocados para cada tutor encontra-se no projeto ou plano de cada curso.

2.2 EVASÃO NA EAD

O termo evasão vem do latim *evasio* e teve a sua origem em meados do século XV significando fuga, saída, fora, abandono, fracasso, insucesso (HOFFMANN, 2016). As nomenclaturas relacionadas ao termo variam de acordo com o autor estudado, no entanto, no âmbito educacional, todas elas levam a um caminho em comum que é a não conclusão de um curso, treinamento, qualificação, especialização ou qualquer modalidade educacional que oriente o aluno à obtenção de conhecimento.

Considerando o ensino superior, trata-se de um dos grandes problemas que cercam este contexto, por se tratar de um fenômeno complexo (SCALI, 2009), heterogêneo e macrossocial. Assim, define-se evasão no ensino superior como o movimento do aluno abandonar a IES e nunca receber o diploma (TINTO, 1975), ou mesmo, a interrupção no ciclo dos estudos, em qualquer nível de ensino (MOROSINI *et al.*, 2012). São considerados evadidos inclusive os estudantes que se matriculam e desistem antes mesmo de iniciar o curso (COMMARELLA, 2009).

Lobo (2012), por meio da Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras, classifica a evasão no ensino superior em três espécies: (a) evasão de curso, quando o estudante se desliga do curso de origem sem concluí-lo; (b) evasão de instituição, quando ele abandona a IES na qual está matriculado; e (c) evasão de sistema, quando o aluno se ausenta de forma permanente ou temporária da academia.

Já na EaD, a evasão pode ser classificada em quatro tipos que, conforme Martínez, García e Montoro (2003) são: (a) *dropout* (evasão), quando o aluno abandona o curso durante sua realização e nunca mais retorna; (b) *stopout* (trancamento), quando ocorre a interrupção temporária do curso, (c) *atteainer* (evasão do curso), que acontece quando o acadêmico sai do curso antes do seu término, no entanto ele obtém o conhecimento; por fim, (d) *nonstarter* (não iniciado), quando o estudante sequer inicia o curso.

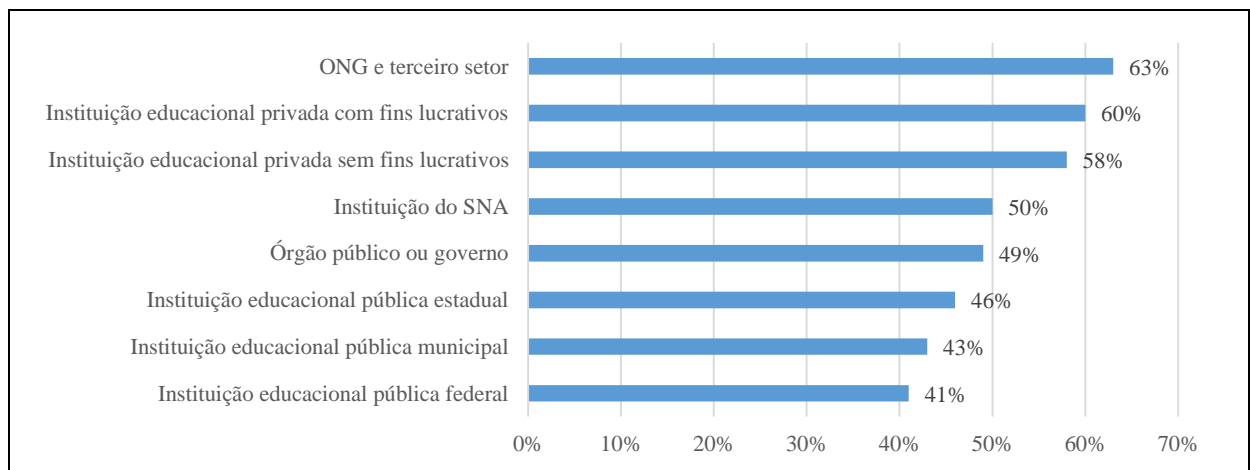
Na modalidade a distância existe uma maior atenção em relação ao tema evasão em comparação à presencial, pois devido a flexibilidade que ela proporciona ao aluno na realização de suas tarefas - tanto em seu lar quanto no seu próprio trabalho - e, ainda, por permitir que o aluno defina os seus momentos de estudos, geralmente existem mais estímulos concorrenciais (filhos, esposa, barulhos da vizinhança, dentre outros). Além disso, depende também diretamente de algumas aptidões do estudante como capacidade de organização e de concentração para os estudos (BITTENCOURT e MERCADO, 2014).

Outro fator é a não existência da figura do professor como centralizador do conhecimento e que todos devem estar de acordo com suas definições. O aluno, neste caso, passa a fazer parte do processo de aprendizagem intervindo, participando e contribuindo para a construção do conhecimento.

Essas exigências da EaD necessitam que os estudantes assumam um comportamento oposto, quando se tornam os atores principais do seu processo de aprendizagem e de sua formação. Tais características exigem adaptabilidade e persistência e nem todos conseguem atingir esse objetivo.

Conforme o Censo da EaD realizado pela Associação Brasileira de Educação a Distância (2017), a evasão é um problema constante em todas as instituições educacionais e em todos os níveis de ensino de forma geral, ocupando o topo das atenções das instituições participantes nos últimos levantamentos realizados. Ainda, de acordo com as informações levantadas pelo censo, chama a atenção a quantidade de instituições que afirmaram conhecer os motivos que levaram à evasão do aluno. Conforme pode ser visualizado no Gráfico 1, 63% das ONGs, 60% das instituições privadas com fins lucrativos e 58% das instituições privadas sem fins lucrativos afirmam conhecê-las. As IES públicas são as que menos conhecem estes motivos (entre 41% e 46%).

Gráfico 1 — Instituições que informaram conhecer os motivos da evasão dos seus alunos, em percentual, por categoria administrativa



Fonte: Censo da EaD (ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA, 2017).

Silva Filho *et al.* (2007) afirmam que a evasão deve ser compreendida sob dois aspectos semelhantes, mas não iguais:

- a) Evasão anual média: avalia o percentual de alunos matriculados em um curso que, no ano seguinte, não tenha se matriculado. Por exemplo, se uma IES tivesse 100 alunos matriculados em um certo curso que deveriam se rematricular no ano seguinte, mas apenas 70 o fizessem, a evasão anual média no curso seria de 30%; e,
- b) Evasão total: verifica a quantidade de alunos que tendo iniciado um determinado curso não obteve o diploma. Por exemplo, se 100 discentes iniciaram um curso em um determinado período e 65 se formaram a evasão nesse curso é de 35%.

O cálculo do percentual da evasão referente ao ano n é dado por:

$$E(n) = 1 - \frac{[M(n) - I(n)]}{[M(n-1) - C(n-1)]}$$

No qual E é a evasão, M o número de alunos matriculados, C o número de concluintes, I é o número de ingressantes, n é o ano em estudo e $(n - 1)$ é o ano anterior.

A complexidade do problema da evasão exige um esforço na busca de suas causas. Dessa forma, diversas pesquisas buscam identificar as causas da evasão, a exemplo de Tresman (2002), Shannon e Bylsma (2006), Bittencourt e Mercado (2014), Silveira (2012), Bizarria, Silva e Carneiro (2014), Abbad, Zerbini e Souza (2010), Coelho (2002), Martins, Gebran e Terçariol (2014), dentre outros.

No trabalho realizado por Tresman (2002) na *Open University*, na Inglaterra, onde foram questionados, no período de um ano, meio milhão de prováveis estudantes de cursos EaD, os principais argumentos que os levariam a desistir de um curso a distância seriam: custo do curso, incerteza em relação à disponibilidade de tempo para dedicar ao curso, mudanças na vida pessoal e dificuldade de escolher um curso entre diversas possibilidades.

Já no estudo realizado por Shannon e Bylsma (2006), os autores apontam diversos fatores relacionados com o estudante, com a família e à comunidade, que podem influenciar a evasão bem como fatores relacionados à IES: questão econômica, realização acadêmica pobre, cursos com poucos encontros presenciais, número de disciplinas com reprovações, doença na família, baixa autoestima, falta de um currículo relevante, estratégias instrucionais passivas, uso inadequado da tecnologia e desrespeito aos estilos de aprendizagem dos estudantes.

Bittencourt e Mercado (2014) buscaram as causas da evasão em turmas do curso de Administração na Universidade Federal de Alagoas. O resultado da pesquisa evidenciou que as principais causas da evasão estão atreladas a problemas endógenos à instituição, ou seja, fatores

institucionais e didático-pedagógicos. Os fatores endógenos estão associados, segundo os autores, a causas relacionadas ao aluno quando está presente na instituição, tais como: atitude comportamental, motivos institucionais e requisitos didático-pedagógicos do curso, conforme pode ser visto no Quadro 2.

Quadro 2 — Causas endógenas da evasão na EaD

Atitude comportamental
Didática dos professores
Orientação da coordenação do curso
Motivação e incentivo por parte do tutor
Insatisfação com o tutor
Contato com os professores
Motivos institucionais
Ausência de tutores nos polos
Acesso às bibliotecas
Estrutura dos polos de ensino
Laboratório de informática nos polos de ensino
Interatividade no AVEA
Meios de comunicação oferecidos para contato
Tecnologia inadequada utilizada
Requisitos didático-pedagógicos
Carga horária curricular do curso
Relação do currículo com o mercado
Critérios de avaliação do aluno
Associação entre a teoria e a prática
Relação entre conteúdo das disciplinas
Encontros presenciais
Complexidade das atividades
Contato entre colegas de curso
Reprovação em mais de duas disciplinas no semestre
Prazos de entrega das atividades
Avaliação dos exercícios
Avaliações das provas
Material didático oferecido
Qualidade do curso
Falha de elaboração do curso

Fonte: (BITTENCOURT e MERCADO, 2014, p. 475).

Além das causas endógenas os autores ainda citam as causas exógenas. Essas causas, na visão deles, se referem àquelas nas quais a instituição não tem possibilidade de intervir junto ao aluno e, dessa forma, não pode evitá-las. Porém, é possível minimizá-la através do levantamento do perfil do aluno buscando identificá-las. No Quadro 3 é possível visualizá-las.

Quadro 3 — Causas exógenas da evasão na EaD

Sócio-político-econômicos
Apoio da instituição que trabalha
Valorização do diploma no mercado
Tempo para estudar
Carga horária semanal de trabalho
Deslocamento até o polo de ensino
Dificuldade de acesso à Internet
Entendimento das matérias
Adequação do conteúdo com o trabalho
Vocação pessoal
Aptidão para a profissão
Possuir outro curso superior
Adaptação ao sistema universitário
Mudança de interesse pessoal ou profissional
Estar cursando paralelamente outro curso
Desconhecimento prévio a respeito do curso
Características individuais
Problemas de saúde
Atendimento do curso às expectativas prévias
Dificuldade de assimilação da cultura de EaD
Falta de habilidade para usar as TIC
Conjunturais
Problemas financeiros
Influência familiar
Mudança de residência ou cidade
Mudança de estado civil
Responsabilidade econômica no sustento da família

Fonte: (BITTENCOURT e MERCADO, 2014, p. 475).

Silveira (2012) realizou um estudo de caso nos cursos EaD ofertados no Polo UAB de Franca para verificar as causas da evasão. Neste estudo, as principais causas foram agrupadas

em cinco categorias: pessoal e profissional, relação interpessoal, aspectos referentes ao curso e tutoria. Dentre as causas, três estão relacionadas à instituição onde a pesquisa foi feita, o que permite uma intervenção pelos gestores para amenizar os índices de evasão.

Bizarria, Silva e Carneiro (2014) focam o estudo no papel do tutor em uma IES pública no estado do Ceará. Por meio da análise de discursos, os pesquisadores constataram que o tutor que tem uma abordagem pedagógica possui uma possibilidade maior de constituir vínculos junto aos alunos e que estes vínculos aliados à interação são fatores determinantes para a redução da evasão.

Abbad, Zerbini e Souza (2010) estudaram a evasão em cursos oferecidos no Brasil onde concluíram que a não utilização de recursos do AVEA empregados pelos cursos, tais como mural de notícias, *chats* e mensagens eletrônicas estão diretamente relacionados com a ocorrência da evasão.

Coelho (2002) aponta como principais fatores da ocorrência da evasão na EaD: a falta da relação frente a frente entre professor e estudante, já que se acredita que nesta forma de relacionamento ocorre maior interação e respostas afetivas entre os envolvidos; pouco conhecimento em relação à utilização das tecnologias envolvidas, desse modo a falta de habilidade cria empecilhos para os alunos acompanharem as atividades propostas pelos cursos; dificuldade na exposição de ideia em uma comunicação escrita por parte dos acadêmicos; por fim, o sentimento de exclusão do aluno no processo educacional devido à falta de interação de pessoas em um mesmo espaço físico.

Além das causas apontadas por Coelho (2002), existem outras que se relacionam ao estudante, como a família, a comunidade e, também, a IES. Segundo Martins, Gebran e Terçariol (2014), esses fatores são: questão socioeconômica, realização acadêmica pobre, cursos com poucos encontros presenciais, diversas disciplinas com reprovações, doença na família, baixa autoestima, falta de um currículo relevante, uso inadequado das tecnologias e desrespeito aos estilos de aprendizagem.

As chances de mitigar o processo de evasão são superiores quando se tem um conhecimento maior sobre os alunos e as causas do abandono. No entanto, diante do alto número de estudantes que, geralmente, estão alocados nos cursos EaD sob responsabilidade de poucos gestores, torna-se inviável, por parte da gestão, compreender detalhadamente todos os aspectos sobre cada estudante e relacioná-los com os motivos da evasão estudados. Nesse sentido, o AVEA, utilizado na interação entre os participantes do processo educativo, pode ser um aliado importante da gestão acadêmica na obtenção de indícios que podem antecipar e, ao mesmo tempo, prever uma possível desistência dos alunos.

Sendo assim, na próxima seção será exposto o conceito de AVEA e os principais componentes que integram estas ferramentas.

2.3 AMBIENTE VIRTUAL DE ENSINO E APRENDIZAGEM (AVEA)

Conforme Bastos, Alberti e Mazzardo (2005), com o surgimento da Internet, a EaD tomou um novo impulso. O que antes era feito por correspondência, rádio ou TV passa a ser mediado pelo computador conectado, possibilitando, inclusive, uma comunicação síncrona. Seguindo a mesma linha de raciocínio, os autores afirmam que, nesse novo contexto, mudam os espaços e tempos de ensinar e aprender, bem como muda a forma de organizar-se o trabalho escolar. O espaço é outro, e os AVEA surgem como possibilidade para que esse processo se articule.

Vieira, Almeida e Alonso (2003) definem AVEA como:

[...] sistemas computacionais geralmente acessados via Internet, destinado ao suporte de atividades mediadas pelas TICs e por um professor-orientador. Permitem integrar múltiplas mídias e recursos, apresentar informações de maneira organizada, desenvolver interações entre pessoas e objetos de conhecimento, elaborar e socializar produções (VIEIRA, ALMEIDA e ALONSO, 2003, p. 118-119).

Envolvendo recursos e procedimentos metodológicos inovadores, os AVEA funcionam como mediadores do processo de ensino e aprendizagem, fomentando a busca pela informação por parte dos alunos e uma postura disciplinar frente às atividades propostas. Bastos, Alberti e Mazzardo (2005) compartilham essa ideia, afirmando que:

Os Ambientes Virtuais de Ensino e Aprendizagem representam novas oportunidades de ensino-aprendizagem, pois comportam um grande número de informações, disponibilidade e acesso, independente de horários preestabelecidos e distâncias geográficas, possibilitam interação através de comunicação síncrona e assíncrona entre os participantes e trabalho colaborativo (BASTOS, ALBERTI e MAZZARDO, 2005, p. 1).

Estes ambientes podem ser empregados em diversos contextos educacionais como: atividades presenciais, permitindo aumentar o número de interações para além da sala de aula; em atividades semipresenciais, nos encontros presenciais e nas atividades a distância (BRASIL, 2005).

Atualmente, existe uma gama de AVEA disponíveis para o suporte a processos de ensino e aprendizagem nesses contextos. Dentre os exemplos de ambientes mais utilizados no

âmbito nacional, pode-se citar: MOODLE², Teleduc³, Sakai⁴ e E-Proinfo⁵. No entanto, em nível mundial, o MOODLE é o mais utilizado por ser um *software* livre e gratuito, contando com uma comunidade diversificada de usuários com propósito de aperfeiçoamento e inovação do desenvolvimento de uma ferramenta educacional.

Os AVEA também são reconhecidos como sistema de gestão do ensino e aprendizagem, do inglês *Learning Management System* (LMS), pois compõe diversas funcionalidades que proporcionam aos educadores a criação de cursos *on-line* de alta qualidade e com diversos tipos de recursos.

A IES onde será aplicada a pesquisa utiliza o AVEA MOODLE. Dessa forma, na subseção que segue, serão descritas características relacionadas ao ambiente que podem proporcionar a descoberta de conhecimento relacionado à evasão dos alunos.

2.3.1 MOODLE

O *Modular Object-Oriented Dynamic Learning* (MOODLE) é um *software* livre, de apoio à aprendizagem, executado em um ambiente virtual, que foi desenvolvido nos anos 1990, por Martin Dougiamas, baseado nas abordagens de aprendizado do construtivismo e do construcionismo social, possibilitando a criação e administração de cursos focado no trabalho colaborativo em um ambiente de simples e intuitiva utilização (BEHAR, 2013).

Entre os principais aspectos do MOODLE, pode-se destacar a tradução do ambiente para mais de 50 idiomas e o seu sistema modular que permite a inclusão de diversos recursos e atividades durante a oferta de um curso. Essa flexibilidade permite a adequação do sistema a diferentes contextos e níveis de trabalho conforme a necessidade de cada instituição. Tais características fizeram com que o MOODLE se tornasse rapidamente um dos AVEA mais utilizados pelas principais instituições de ensino de vários países do mundo. A Figura 5 demonstra um exemplo da interface do MOODLE utilizada pela Universidade Federal de Santa Maria (UFSM).

² Disponível em: <https://moodle.org>.

³ Disponível em: <http://www.teleduc.org.br>.

⁴ Disponível em: <http://sakaiproject.org>.

⁵ Disponível em: <http://e-proinfo.mec.gov.br/e-proinfo>.

Figura 5 — Interface MOODLE – UAB – UFSM

The screenshot shows the Moodle login page for UAB at UFSM. The interface is clean and professional, with a green and white color scheme. The Moodle logo is prominently displayed on the left, alongside the UFSM logo. The login form is located on the right side of the page, featuring a 'Selecionar ambiente:' section with tabs for 'UAB/EAD', 'PRESENCIAL', 'CAPACITAÇÃO', and 'UAB/EAD Histórico'. Below this, there are input fields for 'Usuário' and 'Senha', a large green 'ACESSAR' button, and a checkbox for 'Lembrar identificação de usuário'. A link for 'Dúvidas Frequentes' is also present. The footer includes the Moodle logo, the UFSM logo, and contact information: 'Contato: suporte@moodleufsm@gmail.com | Telefone de contato'. A time display in the bottom right corner shows 22:01.

Fonte: https://ead08.proj.ufsm.br/moodle2_UAB/login/.

O MOODLE possui um conjunto de ferramentas que podem ser selecionadas pelo professor, de acordo com seus objetivos pedagógicos, para aumentar a eficácia de um curso *on-line*. De acordo com Mattar (2012), as principais funcionalidades disponíveis no módulo básico do MOODLE são:

- a) *Chat*: também conhecido como sala de bate-papo, consiste de uma ferramenta síncrona que possibilita ao docente e ao aluno a realização de atividades ou discussões em tempo real;
- b) *Questionário*: funcionalidade que permite a realização de pesquisas e exercícios relacionados a um determinado conteúdo;
- c) *Fórum*: é um recurso comum do AVEA que possibilita o estabelecimento de debates entre professores e alunos por meio da publicação de comentários;
- d) *Glossário*: viabiliza a definição de um dicionário com os termos que os alunos poderão ter dúvidas no decorrer dos seus estudos; e,
- e) *Wiki*: é uma ferramenta que oferece a possibilidade da construção colaborativa de textos promovendo, assim, a interação entre o grupo sendo que qualquer participante pode efetuar modificações no conteúdo.

Silva (2010) apresenta outras ferramentas que podem ser inseridas no MOODLE como: *Podcast* (áudio), *Team Assignment Type* (trabalho em equipe), *Web Conference* (conferências

com imagens e som), Vídeo-Tutor (conjugação de páginas HTML com vídeo), *Mobile Learning Engine* (dispositivos móveis) e *Mobile QUIZ* (questionários em forma de testes para serem respondidos por dispositivos móveis).

A realização de um curso na plataforma MOODLE ocorre através da disponibilização de um espaço, também concebido como sala de aula virtual, para cada disciplina. Nesta sala de aula virtual é possível a interação com os professores, tutores e com os próprios colegas por meio da utilização dos diversos recursos educacionais disponíveis.

Além de recursos educacionais, o MOODLE fornece algumas ferramentas de gestão sendo, dentre as mais conhecidas, a ferramenta relatórios. Por meio desta ferramenta são disponibilizados relatórios que permitem aos gestores a visualização dos acessos dos alunos no AVEA, da realização de tarefas, de participação em fóruns dentre outras informações necessárias para o acompanhamento do estudante. Para disponibilizar tais informações, todas as interações realizadas pelos alunos no AVEA são armazenadas na base de dados do MOODLE. Tais dados dispõem de um grande potencial para a área de MD para descoberta de conhecimento que estão implícitos, porém, podem ser estratégicos para a gestão acadêmica.

No próximo capítulo, serão abordadas as tecnologias que estão sendo empregadas para detecção e prevenção de evasão na EaD.

3 ABORDAGENS TECNOLÓGICAS COMO ESTRATÉGIA PARA DETECÇÃO E PREVENÇÃO DA EVASÃO NA EAD

Diversas tecnologias vêm sendo empregadas para auxiliar na identificação, de forma a prever e buscar, evitar ou minimizar as elevadas taxas de evasão. Como forma de conhecer quais estratégias são estas e como têm sido empregadas, foi realizada uma RSL. Esta RSL buscou evidenciar também se a descoberta de conhecimento por meio de MD, foco desta pesquisa, tem sido utilizada e de que forma. Sendo assim, na subseção seguinte será apresentada a metodologia empregada nesta RSL bem como a sua execução.

3.1 METODOLOGIA

A RSL desenvolvida é baseada nas fases elaboradas por Kitchenham *et al.* (2010) que são: Planejamento, Execução e Resultados. Cada uma dessas fases será abordada nas subseções a seguir.

3.1.1 Planejamento

Na etapa de Planejamento é sugerida a definição de algumas questões de pesquisa que orientarão a seleção e leitura dos artigos na busca de respostas. De acordo com o objetivo da revisão, a questão principal que foi estabelecida é: “ **Que abordagens tecnológicas têm sido empregadas na predição da evasão de alunos de curso superior?** ”. Com base nesta questão, algumas questões de pesquisa mais específicas foram elaboradas conforme demonstradas no Quadro 4.

Quadro 4 — Questões de pesquisa

Questão	Descrição
Q1	Quais dados foram utilizados para análise?
Q2	Quais os resultados obtidos por meio da abordagem empregada?
Q3	Que métodos de avaliação foram empregados?

Fonte: Autor.

Ainda, por mais que esta RSL investigue abordagens em geral, o objetivo principal é a busca por trabalhos que contemplem as técnicas de MD como estratégia para prever evasão de estudantes em nível superior, considerando que este é o foco deste projeto. Assim, foram definidas algumas questões a serem respondidas, no caso dos artigos encontrados que envolvam MD. Estas questões podem ser visualizadas no Quadro 5.

Quadro 5 — Questões específicas de MD

Questão	Descrição
Q4	Nas abordagens que empregam MD, quais são os algoritmos mais utilizados?
Q5	Quais ferramentas foram utilizadas na MD?

Fonte: Autor.

A Q1 serve para compreender os dados educacionais que estão sendo utilizados para o viés da evasão. Esses dados podem servir como base para novas abordagens a serem criadas. Já os resultados e as avaliações, obtidos através das respostas da Q2 e Q3, permitem avaliar se os trabalhos realizados foram aplicados e se a predição foi bem-sucedida. Já a Q4, aplicada especificamente nas abordagens que empregam MD, busca verificar quais são os algoritmos de MD que vêm sendo empregados nos trabalhos. Por fim, a Q5, também particular à mineração, visa detectar quais ferramentas estão sendo utilizadas na execução da mineração.

Para responder essas questões, foram considerados artigos disponibilizados em meios eletrônicos, escritos em inglês e português, publicados entre os anos de 2006 e 2017. A escolha do ano de 2006 é justificada pelo surgimento da UAB, o que possibilitou uma expressiva expansão da EaD no Brasil. Como bases de pesquisa foram escolhidas: *IEEE Xplore Digital Library*, *ACM Digital Library*, *Springer*, Simpósio Brasileiro de Informática na Educação (SBIE), *Workshop de Mineração de Dados Educacionais (WMDE)*, *Revista Brasileira de Informática na Educação (RBIE)* e *Revista de Novas Tecnologias na Educação (RENTE)*. Além das fontes citadas, também foram analisadas as referências utilizadas nos artigos em busca de novos estudos.

Para realizar a pesquisa nas fontes que disponibilizam um motor de busca foram utilizadas as expressões disponíveis no Quadro 6. Essa expressão foi gerada a partir da combinação dos termos das palavras-chave e seus sinônimos, usando operadores lógicos *OR* e *AND* e foi traduzida para os dois idiomas consultados. Já para as bases que não disponibilizam

um motor de busca foi necessário realizar a busca manualmente através da leitura do título de cada artigo.

Quadro 6 — Expressões de busca

Idioma	Expressão
Português	(abordagens OR estratégias OR práticas OR metodologias OR métodos OR técnicas OR modelo) AND (previsão OR predição OR preditivo OR diagnóstico OR identificação) AND (evasão OR permanência OR conclusão OR abandono OR “não concluinte”) AND (ensino OR educação) AND superior
Inglês	(<i>approaches OR strategies OR practices OR methodologies OR methods OR techniques OR model</i>) AND (<i>predict OR prevision OR predictive OR diagnostic OR identify</i>) AND (<i>dropout OR permanency OR conclusion OR abandon OR quittance OR achievement OR dropout OR dismiss</i>) AND (“ <i>data mining</i> ” OR “ <i>machine learning</i> ” OR <i>KDD</i> OR “ <i>educational data mining</i> ” OR <i>EDM</i> OR <i>MDE</i>) AND (“ <i>higher education</i> ” OR <i>degree</i>)

Fonte: Autor.

Para seleção dos estudos foram delimitados os seguintes critérios de inclusão que podem ser vistos no Quadro 7. Já os critérios de exclusão são detalhados no Quadro 8.

Quadro 7 — Critérios de inclusão

Critério	Descrição
I1	Abordagem utilizada em cursos presenciais.
I2	Abordagem utilizada em cursos a distância.

Fonte: Autor.

Quadro 8 — Critérios de exclusão

Critério	Descrição
E1	Artigo redigido antes de 2006.
E2	Não emprega abordagens tecnológicas relacionadas à evasão.
E3	Não emprega a abordagem em cursos de graduação ou pós-graduação.
E4	Estudos duplicados e redundantes (nos estudos que envolvem os mesmos autores relacionados à mesma solução, apenas o mais recente será considerado).
E5	Resumo.

Fonte: Autor.

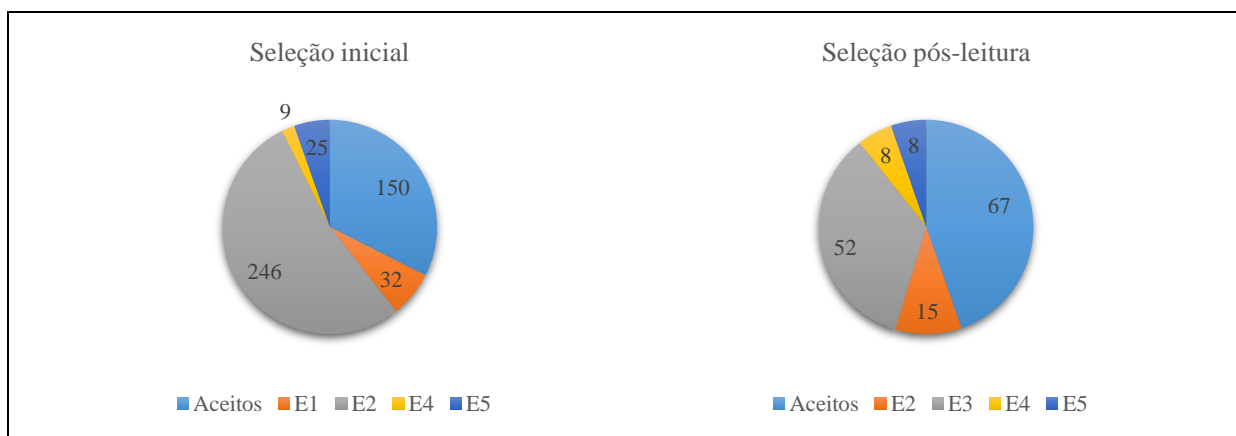
Na seção seguinte serão detalhados os aspectos da execução da RSL.

3.1.2 Execução

Nessa fase ocorreu a busca de artigos nas fontes de dados, onde foram encontradas 462 publicações utilizando a expressão de busca e busca manual. Sucessivamente, houve a seleção preliminar dos artigos através da leitura do título, palavras-chave e resumo de cada publicação. Desses artigos, 32 foram excluídos por terem sido publicados em anos anteriores ao definido no trabalho (E1), 246 por não empregar abordagens tecnológicas relacionadas à evasão (E2), 9 por serem duplicados (E4) e 25 por não serem artigos completos (E5).

Os 150 trabalhos restantes foram lidos na íntegra e, após a leitura, foram reavaliados para verificar se ainda estavam enquadrados nos critérios de inclusão definidos na seleção preliminar. Com a leitura completa 8 trabalhos foram excluídos por estarem duplicados (E4), 52 por não empregar abordagem tecnológica relacionada à evasão (E3), 15 por não empregar uma abordagem em cursos de graduação ou pós-graduação (E2) e 8 por não se tratarem de artigos completos (E5). Portanto, 67 trabalhos foram aceitos e utilizados como base para a extração dos dados. No Apêndice A pode ser visualizada a lista completa desses trabalhos. O Gráfico 2 apresenta os dados da seleção inicial e pós-leitura dos trabalhos relacionados.

Gráfico 2 — Número de trabalhos aceitos e não aceitos na seleção inicial e pós-leitura



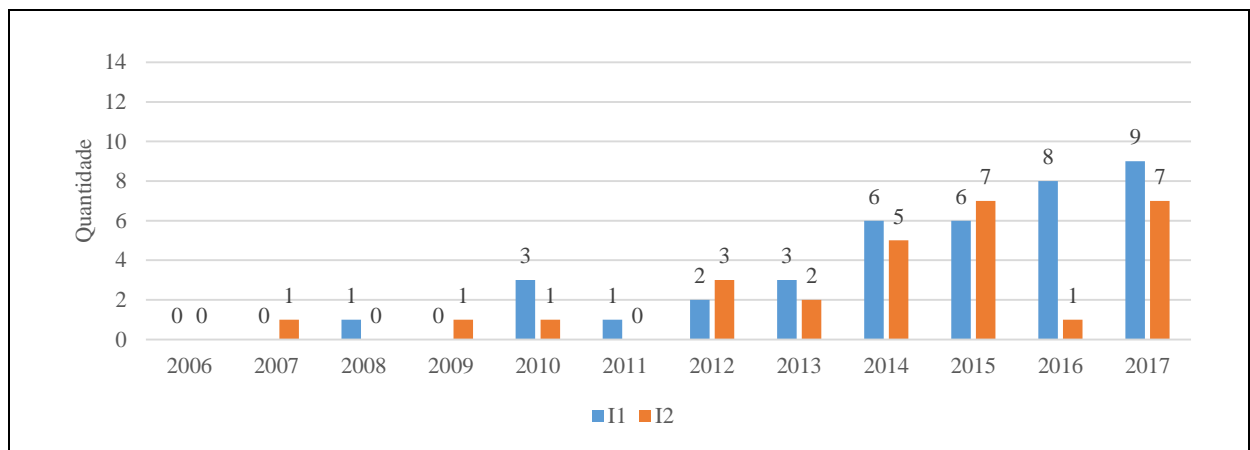
Fonte: Autor.

Cabe salientar, ainda, que a maior parte dos trabalhos que se enquadraram no terceiro critério de exclusão (E3), após a seleção inicial, não possuem uma abordagem tecnológica relacionada à evasão, mas estudam quais são os fatores determinantes para a sua causa o que reforça ainda mais a importância do tema.

3.2 RESULTADOS OBTIDOS E DISCUSSÕES

Com o intuito de analisar a evolução histórica dos trabalhos desenvolvidos no contexto da predição de evasão, no Gráfico 3 é possível visualizar a distribuição dos artigos analisados conforme o ano em que foram publicados e o critério de inclusão em que se enquadraram.

Gráfico 3 — Quantidade de publicações por ano e critério de inclusão



Fonte: Autor.

A partir do Gráfico 3 é possível constatar que entre os anos de 2007 e 2011 foram publicados uma média aproximada de dois artigos por ano, sendo totalizado nesse período um total de 8 artigos. Além disso, no período, nota-se que a maior parte dos trabalhos foram relacionados à predição da evasão em cursos presenciais (I1) totalizando 5 artigos.

Depois do ano de 2012 observa-se um crescimento considerável no número de publicações, sendo que o ápice foi atingido no ano de 2017 somando um total de 16 artigos, o que representa 23,88% dos trabalhos abrangidos por essa RSL. Nesse período houve a produção de 59 trabalhos. Ainda, no que tange aos critérios de inclusão, os artigos relacionados à previsão em cursos presenciais (I1) totalizaram 39 trabalhos, já os trabalhos de previsão na EaD (I2) atingiram 28 publicações.

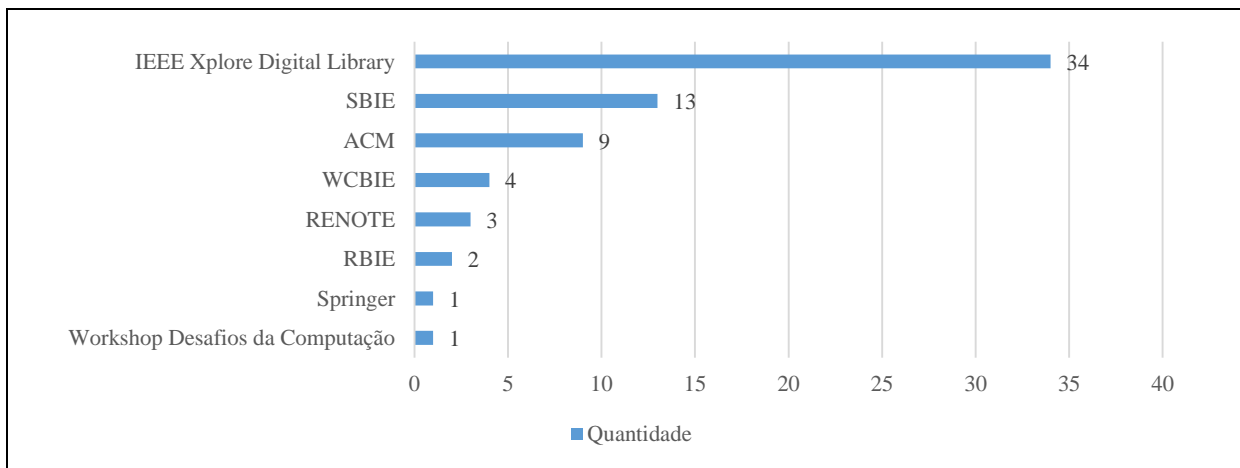
O crescimento apontado nos últimos anos pode ser influenciado por alguns fatores como: avanço das TIC, expansão dos cursos a distância, aumento no número de alunos ingressando no ensino superior, dentre outros.

Percebe-se, em relação aos critérios de inclusão estabelecidos, uma superioridade dos trabalhos publicados na modalidade presencial em relação aos trabalhos na EaD, o que pode ser explicado por essa modalidade de ensino possuir um número maior de alunos em relação à EaD.

Além disso, outro fator importante é o de que as políticas públicas desenvolvidas para sua expansão, tais como UAB criada em 2006, são recentes e, consecutivamente, somente nos últimos anos estão tendo o seu objetivo alcançado. Apesar da diferença constante entre os participantes destas modalidades de ensino, ela não se aplica aos estudos existentes entre elas o que pode ser explicado pelas possibilidades que os dados disponíveis nos AVEA concernem no sentido de prever características dos alunos que, como resultado, permitem a prever a evasão.

Em relação a fonte dos dados é possível verificar, através do Gráfico 4, que mais da metade dos trabalhos estão distribuídas em duas fontes: a primeira, que é internacional, a *IEEE Xplore Digital Library* onde foram encontrados 34 trabalhos; e, a segunda, nacional, o SBIE onde foram localizados 13.

Gráfico 4 — Quantidade de artigos por fonte de dados



Fonte: Autor.

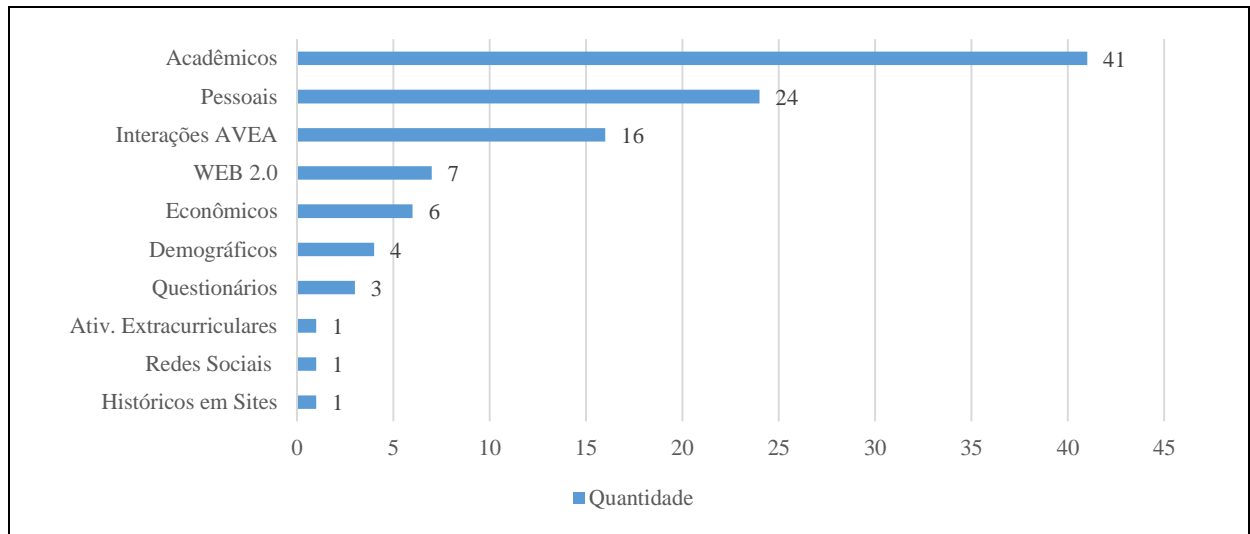
3.2.1 Resposta às questões de pesquisa

Com o objetivo de apoiar a questão de pesquisa principal da RSL serão respondidas algumas perguntas mais específicas com base nos dados extraídos dos artigos.

3.2.1.1 Q1 – Quais dados foram utilizados para análise?

O Gráfico 5 permite visualizar as informações e o número de trabalhos em que foram utilizadas.

Gráfico 5 — Dados utilizados nos trabalhos analisados



Fonte: Autor.

Buscando facilitar a compreensão das informações que foram empregadas, os atributos utilizados nos trabalhos foram organizados em 10 grupos: dados demográficos, acadêmicos, pessoais, econômicos, de interações em um AVEA, disponíveis em recursos da WEB 2.0, de históricos de utilização de *sites*, de redes sociais, provenientes de questionários e de atividades extracurriculares.

Com base no Gráfico 5 percebe-se que os dados acadêmicos são os mais empregados com um total de 41 trabalhos. Alguns atributos desse grupo que merecem ser destacados são: notas, tempo no curso, número de disciplinas aprovadas/reprovadas, trancamentos e média de utilização da biblioteca.

Logo em seguida aparecem os dados pessoais empregados em 24 abordagens. Dentre os mais utilizados são: gênero, idade e estado civil. Junto com os dados pessoais, aparecem os dados que podem ser obtidos através da interação do aluno no AVEA. Destacam-se dentre esses dados: número de acessos ao AVEA, número de interações nos fóruns de discussões, notas das atividades e utilização de mensagens.

Alguns trabalhos também utilizaram recursos provenientes da WEB 2.0 (*fóruns* e *wikis*) para obter dados. Em algumas dessas abordagens foram utilizadas uma outra modalidade de mineração para auxiliar na previsão de alunos. Essa outra técnica é a mineração de informações textuais também conhecida como *Text Mining*. No que se refere a dados econômicos, utilizados em 6 trabalhos, vale citar os dados de renda familiar, se o aluno possui renda própria, se necessita trabalhar para poder se manter na faculdade e etc.

E, por fim, algumas abordagens mais específicas empregaram dados demográficos (4), questionários (3), redes sociais (1), histórico de acessos a *sites* (1) e atividades extracurriculares (1). Exceto o histórico de acesso a *sites*, as demais foram utilizadas em conjunto com outros dados a fim de ampliar o leque de informações de entrada no processo de mineração que, consecutivamente, tendem a produzir resultados mais precisos.

3.2.1.2 Q2 – *Quais os resultados obtidos por meio da abordagem empregada?*

No que tange aos resultados, a maioria dos trabalhos avaliou se os dados definidos para mineração apresentaram resultados satisfatórios, podendo assim, serem empregados para previsão da evasão. Nesses trabalhos, o conjunto de dados foi preparado para ser minerado. Depois disso, um conjunto de algoritmos de mineração foi utilizado em que cada um gerou um modelo. No fim, avaliando a eficiência dos modelos, os autores concluíram se os dados poderiam contribuir para previsão.

Através dos modelos gerados, os autores contribuíram com alguns fatores relevantes para a ocorrência da evasão, como: alunos com um maior número de participações em fóruns de discussões tendem a ter notas mais altas; em relação aos alunos da EaD, verificou-se que muitos alunos desistem no primeiro semestre do curso pela baixa motivação em estudar por conta própria; notas iniciais dos alunos podem identificar as notas finais; disciplinas de cursos que são responsáveis pelo maior número de reprovações; alunos que realizam o empréstimo de maior número de livros na biblioteca tendem a permanecer no curso; e, alunos com maior número de interações no AVEA tendem a concluir o curso.

Os trabalhos de Kostopoulos, Sotiris e Panagiotis (2015) e Rigo e Cazella (2014) apresentaram um viés diferenciado através do desenvolvimento de uma ferramenta que utiliza os modelos gerados. Nessa ferramenta era permitido que usuários pudessem visualizar/executar o processamento. Na primeira abordagem, o conjunto de dados de um aluno era solicitado ao usuário para que, a partir dele, uma análise fosse feita e, assim, disponibilizado um percentual referente à probabilidade que aquele aluno possuía de evadir ou não. No segundo, era disponibilizada uma ferramenta WEB com a listagem dos alunos de uma turma em que era exibido um alerta naqueles com tendência a evadir. Nessa ferramenta, era possível que o docente/tutor informasse quais foram as ações pedagógicas realizadas a fim de evitar a desistência. Esses dados eram utilizados, posteriormente, pelos demais docentes, para que ações que apresentaram bons resultados pudessem ser utilizadas em outros casos.

No trabalho de Palazuelos, García-Saiz e Zorrilla (2013) uma nova ferramenta também foi desenvolvida buscando facilitar o uso da ferramenta WEKA. Nessa ferramenta foram abstraídas as características mais técnicas dos usuários, possibilitando assim, que o processo de mineração fosse utilizado por pessoas que não possuem um grande conhecimento em banco de dados. Com o emprego de outras técnicas apareceram os trabalhos de Zafra, Cristóbal e Sebatían (2009); Zheng, Zeyu, Changjun (2013) e Delgado *et al.* (2014). No primeiro, foram utilizados algoritmos de programação genética para prever o desempenho final dos alunos. No segundo, o desempenho foi avaliado utilizando modelos de redes neurais. Por fim, no terceiro, a avaliação de desempenho foi aplicada adotando Máquinas de Vetores de Suporte (*Support Vector Machine - SVM*).

3.2.1.3 Q3 – *Que métodos de avaliação foram empregados?*

Conforme a resposta da questão anterior muitos dos trabalhos analisados na RSL se restringe à proposição de um modelo conceitual para MD. Dessa forma, é avaliado, através de um conjunto de dados, se a sua utilização apresenta ou não uma quantidade mínima de “acertos” que possibilitem a sua adoção na previsão da evasão. Sendo assim, os métodos utilizados para a avaliação desses artigos se restringem nas métricas que analisam o modelo como: Acurácia, Precisão, Revocação e Medida-F.

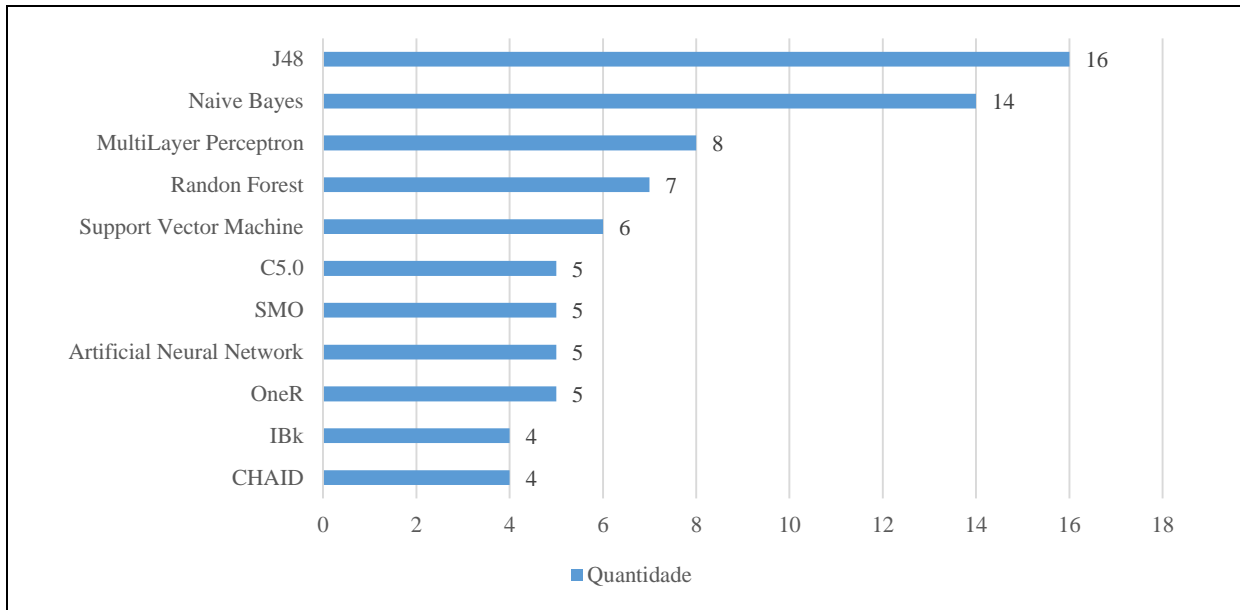
Alguns trabalhos como Rigo e Cazella (2014); Sales, Leite e Joye (2012); Martins, Lopes, Raabe (2011); Al-Shargabi e Nusari (2010); Zafra, Cristóbal e Sebatían (2009) e Kantorski *et al.* (2016) apresentaram um estudo de caso em que a abordagem proposta foi validada. Dos trabalhos citados, vale destacar o artigo de Rigo e Cazella (2014), que apresenta o desenvolvimento de uma nova ferramenta para disponibilização dos resultados e a abordagem foi testada em duas turmas, sendo que, em uma delas, se utilizou a ferramenta e se efetuou intervenções junto aos alunos que nela foram apontados como possíveis desistentes. De acordo com os autores, essa turma apresentou redução na evasão em comparação com a outra turma que não houve intervenções.

No trabalho Kantorski *et al.* (2016), antes do final do semestre letivo, foi disponibilizada uma lista com possíveis alunos com tendência a evadir. Ao encerrar o período de matrículas para o semestre seguinte foi verificado quais dos alunos apontados na lista não tinham realizado a matrícula, ou seja, evadiram. Validações semelhantes foram aplicadas nos trabalhos Zafra, Cristóbal e Sebatían (2009); Wong e Simon (2016) e Sales, Leite e Joye (2012).

3.2.1.4 Q4 – Nas abordagens que empregam MD, quais são os algoritmos mais utilizados?

Nos trabalhos avaliados foram utilizados um total de 39 algoritmos de MD. Os mais utilizados podem ser visualizados no Gráfico 6.

Gráfico 6 — Algoritmos mais utilizados nas abordagens analisadas



Fonte: Autor.

Para fins de visualização, no Gráfico 6 foram demonstrados os algoritmos que foram utilizados em pelo menos quatro trabalhos. Dentre os mais utilizados destacam-se, em primeiro lugar, o algoritmo de árvore de classificação J48 sendo empregado em 16 abordagens. Logo em seguida, aparece o algoritmo de classificação *Naive Bayes*, empregado em 14 trabalhos. E, por fim, o *MultiLayer Perceptron* utilizado em 8.

Dentre os algoritmos menos citados e não indicados no gráfico apresentado estão: *AdaBoost*, *AdTree*, *Apriori Inverse*, *BayesNet*, *BFTree*, *C&R Tree*, *CART*, *Decision Table*, *Hellinger Distance Decision Tree*, *IB1*, *ID3*, *J48-Bin*, *JRip*, *K-Means*, *Logistic Regression*, *Multiple Instance Learning*, *Neareast Neighbor*, *Neuro Fuzzy*, *PART*, *Prism*, *Quest*, *Radial Basis Function Network*, *Simple Cart* e *SimpleLogistic*.

3.2.1.5 Q5 – Quais ferramentas foram utilizadas na MD?

Dentre os trabalhos que fizeram referência à utilização de uma ferramenta para mineração (37), em 22 deles foi citada a utilização da ferramenta WEKA. Logo em seguida, aparece a ferramenta SPSS que foi utilizada em 3 artigos. Dentre outras ferramentas utilizadas aparecem: *KeyGraph* (análise de textos), *PASW Modeler* (IBM), *Principal Component Analysis* (PCA), *Microsoft SQL Server 2005 Analysis Services* (SSAS), *Clementine*, *SVMLight*, *RapidMiner*, *SPSS* e *R*.

3.3 CONSIDERAÇÕES

Na RSL apresentada verificou-se a predominância de dados de interação dos AVEA, bem como os dados acadêmicos dos estudantes para predição da evasão. Além disso, em relação aos resultados, constata-se que grande parte dos trabalhos apresentou um modelo capaz de inferir a possibilidade da desistência do aluno de um curso. Devido a esse fato, as avaliações restringiram-se nas métricas dos algoritmos. Poucos trabalhos utilizaram os resultados dos algoritmos e validaram se na realidade a previsão se concretizou.

Como o foco da pesquisa tem como objetivo a utilização de MD para previsão da evasão de alunos na EaD, no próximo capítulo serão apresentados os detalhes desta área computacional e dos processos envolvidos.

4 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

Os contínuos avanços das tecnologias da informação têm viabilizado o armazenamento de grandes e múltiplas bases de dados. Tecnologias como a Internet, redes sociais, AVEA, dispositivos móveis, sensores de coletas de dados, sistemas de comunicações e sistemas de informação em geral são alguns exemplos que têm tornado possíveis a criação e o crescimento de inúmeras bases de dados de natureza administrativa, comercial, educacional, governamental e social (GOLDSCHMIDT, BEZERRA e PASSOS, 2015).

Dados armazenados nestas bases podem fornecer padrões ou tendências oportunas que, se descobertos, possuem potencial, por exemplo, para auxiliar na compreensão de resultados de um experimento científico, ajudar médicos a interpretar os efeitos de um tratamento, otimizar processos de uma corporação, detectar a evasão de alunos, dentre outros.

A análise de grandes quantidades de dados não é viável para o homem sem o auxílio de ferramentas computacionais apropriadas. Dessa forma, é essencial a disponibilização destes recursos para amparar o indivíduo na tarefa de analisar, interpretar e relacionar esses dados, para que estratégias de ação possam ser elaboradas em cada domínio de aplicação (GOLDSCHMIDT, BEZERRA e PASSOS, 2015). Neste cenário surgiu a área de Descoberta de Conhecimento em Base de Dados (mais conhecida do inglês, *Knowledge Discovery in Databases* – KDD) para responder a estas necessidades.

KDD é definido por Fayyad, Piatetsky-Shapiro e Smyth (1996) como “o processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. Ainda de acordo com os autores, sua utilização visa a melhoria do entendimento de um problema ou o fornecimento de subsídios para a tomada de decisões.

Para um melhor entendimento do conceito apresentado, Prass (2004) apresenta a análise de alguns dos termos utilizados pelos autores. (a) não trivial: evidencia a existência de um processo de busca ou inferência. (b) previamente desconhecidos: sugere que a informação deve ser nova tanto para o sistema como para o usuário. Por fim, (c) potencialmente úteis, indica que a informação deve fornecer algum benefício, ou seja, deverá proporcionar ao interessado algum ganho.

Dentre as áreas em que o KDD pode ser aplicado estão:

- a) Retenção de clientes: detecção de perfis de clientes para determinados produtos (PETERMANN, 2006; FERREIRA, 2005);

- b) Bancos: busca de padrões para apoiar o gerenciamento do relacionamento (SOUSA e FIGUEIREDO, 2014);
- c) Cartão de crédito: identificar segmentos de mercado como também padrões de rotatividade (KISAHLEITNER, 2008);
- d) *Telemarketing*: facilidade de acesso aos dados de clientes (JALAL, HOSSEINI e KARLSSON, 2016; COUSSEMENT e VAN DEN POEL, 2008);
- e) Eleitoral: indicação de perfil de candidatos a vereador (CAMARGO *et al.*, 2016; TAVARES, BOZZA e KONO, 2007);
- f) Segurança: revelação de atividades terroristas e criminais (MCCUE, 2007; CHEN *et al.*, 2008);
- g) Recursos humanos: designação de capacidades em currículos (CABRAL e SIEBRA, 2006) e investigação de acidentes de trabalhos (BARTOLOMEU, 2002);
- h) Tomada de decisão: fornecimento de indicadores de probabilidades (ADEODATO *et al.*, 2005);
- i) Comércio: descoberta de combinações simultâneas em transações de vendas (NUNES, 2009);
- j) *Marketing*: análise de comportamento do consumidor (KIPPES, 2010); e,
- k) Evasão escolar: descoberta de alunos com tendência à evasão (KANTORSKI *et al.*, 2016; QUEIROGA *et al.*, 2016).

Na próxima seção serão abordadas as categorias de objetivos a serem alcançados na execução do processo de KDD.

4.1 OBJETIVOS DO PROCESSO DE KDD

Normalmente, o processo de KDD inicia-se pela análise de um conjunto de dados. Depois disso, devem ser delineados os objetivos a serem alcançados durante o processo. Han, Kamber e Pei (2012) classificam os objetivos a serem alcançados em duas categorias:

- a) Descritiva: visa a análise dos dados para o detalhamento de suas características e atributos principais; e,
- b) Preditiva: tem como objetivo analisar os dados para a construção de modelos para tentar realizar inferências sobre os mesmos de maneira que possibilite o sistema a fazer previsões de comportamento para um novo conjunto de dados. Os modelos originados destas técnicas caracterizam aspectos próprios dos dados. Sendo desse

modo, é necessário um conjunto de exemplos (registros) e características (atributos) que definam de forma correta os grupos ou classes diferentes.

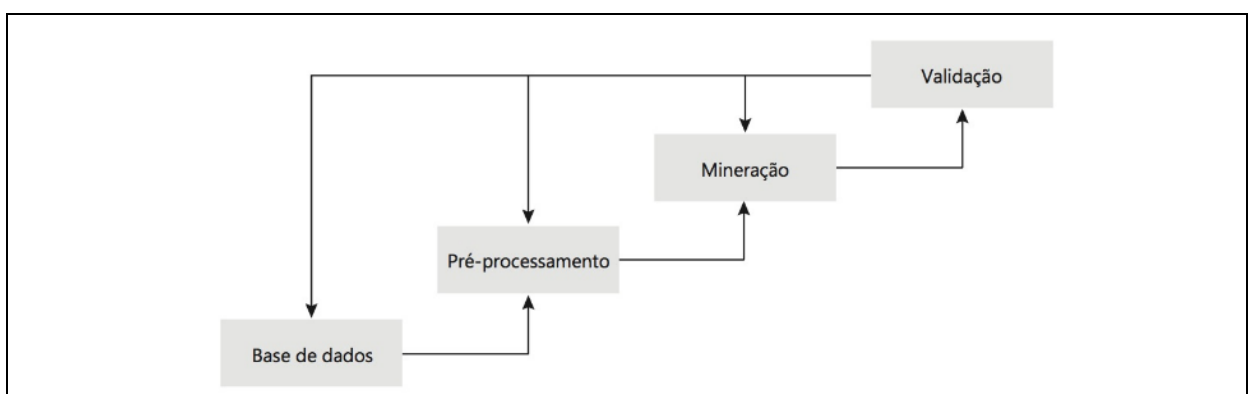
Considerando as categorias descritas, esta pesquisa se enquadra como preditiva. Isso porque os dados históricos do AVEA MOODLE serão utilizados para a construção de modelos que façam a predição em relação à permanência/desistência de alunos que estejam matriculados em cursos EaD.

Logo após a exposição das categorias de objetivos a serem alcançados na execução do processo de KDD, na subseção a seguir as etapas do processo serão descritas.

4.2 PROCESSO DE KDD

O KDD é representado como um processo composto por várias etapas. Para facilitar a sua compreensão ele é sintetizado em quatro elementos principais observando-se, no primeiro, os dados obtidos por meio de uma base de dados. No segundo elemento, os dados são modificados na fase de pré-processamento. Logo em seguida, são aplicadas técnicas de MD. Por fim, o resultado da etapa anterior é transformado para gerar a visualização da informação. A sintetização dessas etapas pode ser vista na Figura 6.

Figura 6 — Processo de descoberta de conhecimento em bases de dados



Fonte: (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).

Para um melhor entendimento, os itens que compõem o processo serão caracterizados nas próximas subseções.

4.2.1 Base de dados

Uma base de dados consiste de uma coleção sistematizada de dados que possibilita a sua recuperação de maneira eficiente. Neste caso, os dados se referem ao menor nível de abstração a partir do qual a informação e, posteriormente, os conhecimentos podem ser obtidos (CASTRO e FERRARI, 2016). As bases de dados empregadas no processo de KDD podem ser obtidas a partir de diversas formas de estruturas e armazenamento de dados: bancos de dados relacionais, planilhas eletrônicas, *data warehouse*, arquivos de log, *data stream*, dados da WEB, entre outros (MANHÃES, 2015). Independente da origem, os dados devem ser transformados, na etapa de pré-processamento, para um formato adequado no qual é possível a aplicação das técnicas de MD.

4.2.2 Preparação ou pré-processamento de dados

A preparação ou pré-processamento de dados consiste na fase que antecede à mineração. Esta fase tem como objetivo a organização dos dados para uma análise eficaz e eficiente pelos algoritmos de MD (CASTRO e FERRARI, 2016). As funções básicas de pré-processamento dos dados serão descritas a seguir.

4.2.2.1 Limpeza dos dados

A limpeza consiste de qualquer manipulação realizada sobre os dados obtidos, de forma a assegurar a qualidade da informação. Um exemplo de manipulação seria o estabelecimento de um intervalo de valores permitidos para um determinado atributo. No âmbito da pesquisa, um intervalo de valores pode ser estabelecido para o atributo de quantidade de interações do aluno no AVEA, onde o seu valor precisa ser maior ou igual a zero. Se o atributo do registro possuir um valor que não se enquadre no critério estabelecido, ele pode até mesmo ser desconsiderado.

4.2.2.2 Integração

A integração, também conhecida como enriquecimento dos dados, prevê a união de dados oriundos de fontes distintas com o objetivo de agregar mais informações aos registros existentes, fornecendo assim, mais subsídios para MD.

4.2.2.3 Seleção

A seleção, também denominada como redução de dados, compreende a definição dos dados que são relevantes para análise. De acordo com Goldschmidt e Passos (2005), a seleção pode ser executada sob dois aspectos diferentes: a seleção de atributos ou a seleção de registros que devem ser submetidos à MD. Como exemplo da seleção de atributos, o nome do aluno é irrelevante em uma aplicação de KDD cujo intuito seja prever o comportamento de alunos com tendência à evasão. Já a seleção de registros ocorre nos casos em que não é possível utilizar todo o conjunto de dados disponíveis como, por exemplo, a ausência de valores de atributos em alguns registros.

4.2.2.4 Transformação

A transformação, também conhecida como codificação dos dados, trata da conversão dos dados em formatos apropriados para a mineração. A transformação pode ser Numérica – Categórica, que transforma valores reais em categorias ou intervalos; ou Categórica – Numérica, que representa numericamente valores de atributos categóricos. A seguir serão abordados os dois tipos de transformações empregados na pesquisa, a padronização e a normalização.

4.2.2.4.1 Padronização

A padronização é uma espécie de transformação que objetiva resolver as diferenças de unidades e escala dos dados tais como: capitalização, caracteres especiais, padronização de formatos e conversão de unidades.

4.2.2.4.2 Normalização

A normalização visa a conversão dos dados para um formato mais apropriado para aplicação em alguns tipos de algoritmos de MD. Ela é frequentemente utilizada na distribuição de todos os atributos de uma base de dados sob o mesmo intervalo de valores, por exemplo no intervalo entre 0 e 1 (CASTRO e FERRARI, 2016).

Na aplicação da pesquisa os dados também são convertidos para este formato a fim de verificar se estes não apresentam melhor performance em relação ao formato sem transformação. Desse modo, a partir de então, o formato de dados em que a normalização foi empregada será referenciado como dados normalizados. Já, os dados em que esta técnica não será aplicada, será denominado de dados não normalizados.

Na próxima subseção será abordada a etapa subsequente delineada no processo de KDD, a mineração de dados.

4.2.3 Mineração de Dados (MD)

Mineração de dados é a principal etapa do processo de KDD tanto que pode ser encontrada na literatura autores que definem MD e KDD como sinônimos. Ela compreende a aplicação de técnicas e algoritmos sobre os dados em busca de conhecimento útil e implícito (GOLDSCHMIDT, BEZERRA e PASSOS, 2015). A definição da técnica a ser utilizada depende, na maioria das vezes, do tipo de tarefa de KDD a ser realizada.

Nas subseções subsequentes serão detalhadas as técnicas de MD que podem ser aplicadas nas tarefas de descoberta de conhecimentos.

4.2.3.1 Associação

A associação, também conhecida como descoberta de associação, define quais fatos tendem a ocorrer simultaneamente e com grande frequência em um conjunto de dados. Um exemplo clássico da aplicação desta técnica provém da área de *marketing* em que, durante a aplicação do processo de KDD em uma base de dados de uma rede de supermercados, foi descoberto que boa parte dos compradores de fralda também comprava cerveja na véspera de finais de semana com jogos transmitidos na televisão. Por meio de uma análise mais detalhada, foi possível constatar que os compradores eram homens e ao comprarem fraldas para seus filhos, compravam também cerveja para o consumo enquanto assistiam aos jogos e cuidavam de seus filhos durante o final de semana. Diante desse novo conhecimento, a rede de mercados aproximou os dois produtos a fim de incrementar as suas vendas. São exemplos de algoritmos que implementam esta técnica: *Apriori*, GSP, DHP, entre outros (ZAKI, 2000).

4.2.3.2 Classificação

A classificação consiste na montagem de um modelo, tendo como entrada um conjunto de dados classificados que possam ser utilizados para predizer qual seria a classificação de novos registros, ou seja, registros cujas classificações ainda não são conhecidas. Este tipo de técnica é conhecido como predição e pode ser de dois tipos: classificação ou estimação. Como exemplo, uma empresa de cartão de crédito quer saber se concede ou não crédito para um cliente, o que é feito por meio da classificação. Além disso, se conceder o crédito, pode querer saber qual o valor de limite de crédito que pode ser disponibilizado, o que é feito através da estimação. Redes Neurais, Algoritmos Genéticos e Lógica Nebulosa são exemplos de algoritmos que podem ser aplicados na classificação (CASTRO e FERRARI, 2016).

4.2.3.3 Regressão

A regressão, de acordo com Goldschmidt, Bezerra e Passos (2015), compreende a busca por uma função linear que mapeie os registros de uma base de dados em um intervalo de valores reais, semelhante à técnica de classificação. Como exemplo de aplicações de regressão, pode-se citar: estimativa da probabilidade de um paciente sobreviver, dado o resultado de um conjunto de exames; predição de risco de um determinado investimento, dentre outros. Dentre as ferramentas que implementam a tarefa de regressão estão: Estatística e Redes Neurais.

4.2.3.4 Agrupamento

O agrupamento, também conhecido como clusterização, visa agrupar os registros de uma base de dados em subconjuntos, ou *clusters*, de tal forma que cada grupo compartilhe atributos em comuns que os diferenciem dos demais. Como exemplo, uma empresa de seguro pode aplicar o processo de clusterização em sua base de dados para detectar o perfil dos segurados que realizam comunicação de sinistro com alta frequência. Para a realização desta técnica podem ser utilizados os algoritmos: *k-Means*, *k-Modes*, *k-Prototypes*, Mapas de Kohonen, dentre outros (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).

4.2.3.5 Sumarização

A sumarização visa identificar e indicar aspectos iguais entre registros do conjunto de dados (WEISS e INDURKHYA, 1998). Como exemplo de aplicação desta técnica pode ser considerado o caso de um conjunto de dados sobre clientes que assinam um determinado tipo de revista. Neste caso a sumarização pode ser aplicada para buscar os aspectos comuns a todos os clientes. Esses aspectos podem ser empregados pela equipe de *marketing* da revista para direcionar a oferta de novas assinaturas para potenciais clientes. Alguns exemplos de algoritmos que implementam a técnica de sumarização são: Lógica Indutiva e Algoritmos Genéticos.

4.2.3.6 Técnica para análise dos dados da pesquisa

Analisando as técnicas de MD abordadas, percebe-se que cada uma delas, com suas especificidades, pode apresentar informações úteis no que diz respeito tanto à previsão de alunos com tendência à evasão (classificação e regressão) quanto na obtenção de dados que permitem a identificação do perfil de alunos que se enquadrem nesta situação (associação, agrupamento e sumarização).

Desse modo, tendo em vista o objetivo proposto na pesquisa, a técnica a ser empregada é a de classificação. Isso em razão de que os modelos gerados para a predição permitem o levantamento de informações necessárias para a obtenção do perfil do aluno com tendência à evasão. Além disso, as predições geradas permitem que ações pedagógicas possam ser realizadas junto aos alunos previstos.

Sendo assim, na subseção que segue serão abordados aspectos relacionados a avaliação dos algoritmos que fazem o uso desta técnica. O conhecimento das avaliações é necessário já que permitirão determinar se os resultados de um algoritmo são os desejados para atendimento do objetivo estabelecido.

4.2.3.6.1 Avaliação de algoritmos de classificação

Usualmente, a avaliação dos algoritmos de classificação é realizada em dois momentos distintos. No primeiro é analisada a estimativa de desempenho com base nos dados previamente classificados (treinamento), visando avaliar se os modelos gerados pelos algoritmos são factíveis para a classificação. Logo em seguida, no segundo momento, é avaliado o desempenho da predição dos classificadores.

Para o primeiro momento, a estimativa, comumente é empregado o método de avaliação cruzada, do inglês *cross-validation*. A forma mais conhecida de validação cruzada é a chamada validação cruzada em k -pastas, do inglês *k-fold cross-validation*, que consiste na divisão da base de dados em k subconjuntos, sendo $k-1$ conjuntos para treinamento dos algoritmos e o conjunto restante para teste. Esse processo é realizado para todos os k subconjuntos e a média do desempenho para os conjuntos analisados é utilizado como indicador de qualidade de desempenho (CASTRO e FERRARI, 2016). A validação cruzada adotando $k=10$ é a mais utilizada para esta finalidade sendo também utilizada neste trabalho. Os resultados desta avaliação são disponibilizados por meio das mesmas métricas abordadas na análise de desempenho.

No que condiz a avaliação de desempenho da predição comumente são empregadas as métricas de Acurácia, Precisão, Revocação e Medida-F. Para facilitar a compreensão dessas métricas suas fórmulas serão demonstradas no campo de aplicação da pesquisa, sendo elas:

- a) Acurácia (*accuracy*): mais conhecida como a taxa geral de sucesso do classificador.

Diante de um determinado conjunto de teste é obtido o percentual dos exemplos que foram corretamente classificados. Ou seja, demonstra quanto o modelo foi preciso no acerto dos dados do conjunto de testes. O cálculo de acurácia é dado por:

$$\text{Acurácia} = \frac{\text{Número de Classificações Corretas}}{\text{Número de Instâncias}}$$

Em que o Número de Classificações Corretas é a quantidade total de classificações previstas corretamente, tanto relacionadas aos alunos que foram classificados como desistentes quanto aos permanentes. Já o Número de Instâncias representa o número de alunos avaliados;

- b) Precisão (*precision*): determina a taxa de acertos do algoritmo em relação a previsão realizada. O cálculo é dado por:

$$\text{Precisão} = \frac{\text{Nro. Alunos Previstos} \cap \text{Nro. Alunos Evadidos}}{\text{Nro. Alunos Previstos}}$$

Em que o Número de Alunos Previstos representa a quantidade de alunos que foi classificado, no âmbito da pesquisa, como “Evadido” já que o objetivo é encontrar os alunos com tendência à evasão. Já o Número de Alunos Evadidos compreende a

quantidade de alunos que de fato desistiu. Em suma, a métrica de precisão é calculada pelo número de alunos previstos e evadidos dividido pelo número de alunos previstos;

- c) Revocação (*recall*): mede a integridade do algoritmo em relação aos alunos evadidos. O cálculo é dado por:

$$\text{Revocação} = \frac{\text{Nro. Alunos Previstos} \cap \text{Nro. Alunos Evadidos}}{\text{Nro. Alunos Evadidos}}$$

Em que o Número de Alunos Previstos representa a quantidade de alunos que foi classificado, no âmbito da pesquisa, como “Evadido”. Já o Número de Alunos Evadidos compreende a quantidade de alunos que de fato desistiu. Em suma, a métrica de revocação é calculada pelo número de alunos previstos e evadidos pelo número de alunos que evadiram; e,

- d) Medida-F (*f-Measure*): métrica de avaliação que consiste na média harmônica de Precisão e Revocação. O cálculo é dado por:

$$\text{Medida} - F_{\beta} = (1 + \beta^2) * \left(\frac{\text{Precisão} * \text{Revocação}}{(\beta^2 * \text{Precisão}) + \text{Revocação}} \right)$$

Existem outras fórmulas que podem ser utilizadas para cálculo da Medida-F. No entanto, a fórmula demonstrada foi adotada por possibilitar que pesos diferentes possam ser utilizados para Precisão e Revocação. Com a utilização de $\beta = 2$, a maior ênfase é dada para Revocação e $\beta = 0,5$ para Precisão. Dessa forma, nesta pesquisa será utilizada a abordagem que emprega maior peso para a Revocação já que, além do acerto da previsão, também é importante considerar o maior número de alunos que possam se enquadrar como evadido.

4.2.4 Avaliação ou validação do conhecimento

A etapa de avaliação ou validação do conhecimento, também referenciada na literatura como pós-processamento, compreende as etapas que visam facilitar a interpretação e a avaliação dos dados pelos especialistas do domínio da aplicação no que se refere à utilidade do conhecimento descoberto (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).

De acordo com Goldschmidt e Passos (2005), as etapas deste processo são:

- a) Elaboração e organização do conhecimento obtido, podendo incluir a simplificação de gráficos, diagramas, ou relatórios demonstrativos; e,
- b) Conversão da forma de representação do conhecimento obtido.

Na próxima seção serão abordadas as ferramentas que auxiliam no processo de descoberta de conhecimento.

4.3 FERRAMENTAS PARA KDD

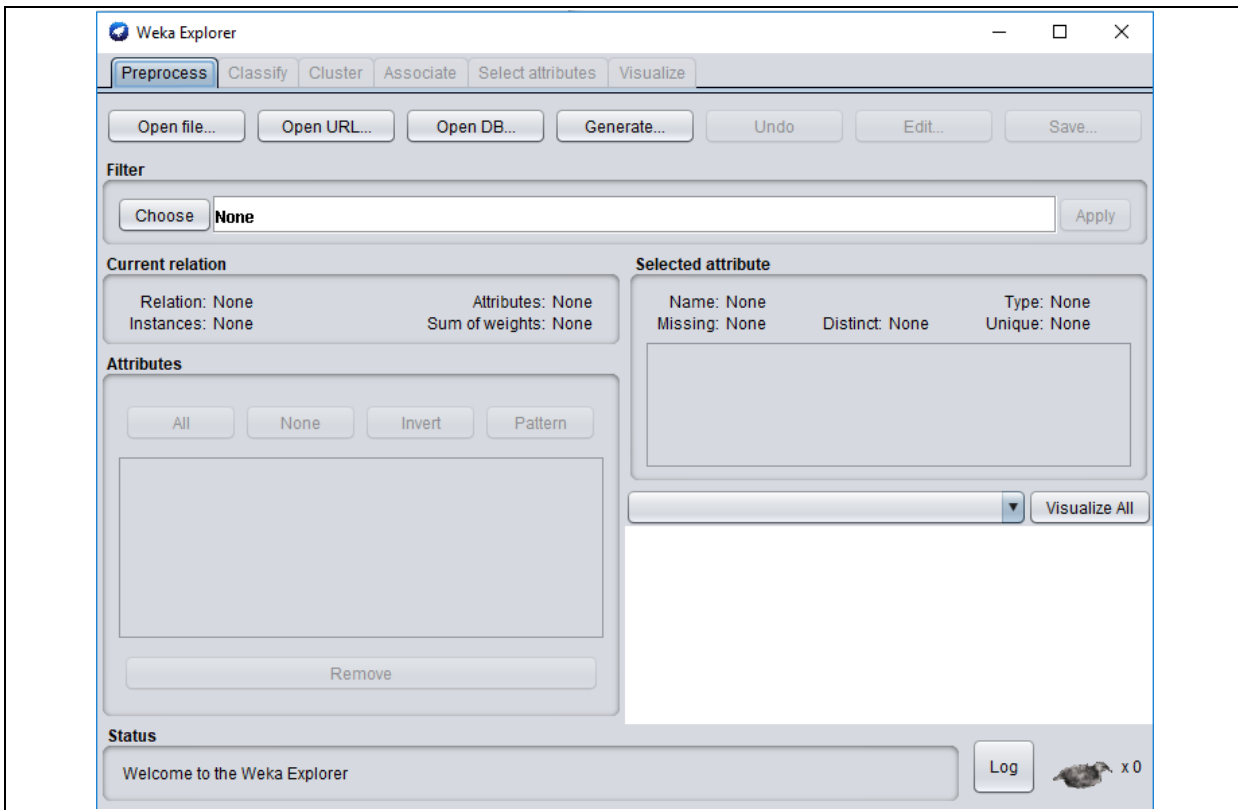
Atualmente, existe uma gama de ferramentas que implementam ambientes integrados para proporcionar a execução das etapas operacionais do KDD. *RapidMiner*, *RProgramming*, *Orange*, *Knime*, *NLTK*, *Oracle Data Mining*, *SAS*, *Cognos* e *WEKA* são algumas dessas ferramentas.

Neste trabalho optou-se pela utilização da ferramenta WEKA⁶ (*Waikato Environment for Knowledge Analysis*), atualmente disponível na versão 3.9.0, que é mantida por universitários da Universidade de Waikato, na Nova Zelândia. A escolha desse *software* se justifica por algumas características, tais como facilidade de utilização, obtenção direta da página do desenvolvedor sem custos, número considerável de algoritmos disponíveis para a utilização, junto com a possibilidade de alteração dos parâmetros de execução dos mesmos e a possibilidade de fácil comparação entre os algoritmos verificados.

A WEKA teve sua primeira versão criada no ano de 1999 com o principal intuito de ser aplicado para análise de dados agrícolas (HALL *et al.*, 2009). É uma ferramenta desenvolvida utilizando a linguagem de programação JAVA e de código aberto sob a *General Public License* (GPL), podendo ser utilizada tanto por meio de sua interface gráfica quanto em uma aplicação customizada através da importação de sua *Application Programming Interface* (API). Na Figura 7 é possível visualizar a interface gráfica de exploração de dados do WEKA.

⁶ Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>.

Figura 7 — Interface gráfica do WEKA



Fonte: <https://www.cs.waikato.ac.nz/ml/weka>.

Algumas de suas principais funcionalidades são: o pré-processamento de dados, a visualização e análise preditiva, além de técnicas de modelagem, *listening*, agrupamento, associação, regressão e classificação.

4.4 MINERAÇÃO DE DADOS EDUCACIONAIS (MDE)

A Mineração de Dados Educacionais (do inglês *Educational Data Mining – EDM*), é conceituada por Costa *et al.* (2013) como a aplicação, o desenvolvimento e a adaptação dos métodos de KDD para o contexto educacional.

Os primeiros trabalhos desenvolvidos na área concentravam-se, basicamente, na construção de modelos para predição do desempenho dos estudantes com base em dados históricos anteriores.

Nos últimos anos, com o fortalecimento da EaD, novas possibilidades de pesquisa começaram a surgir. Neste contexto, Goldschmidt, Bezerra e Passos (2015) afirmam que outros grupos de participantes do contexto educacional, além de estudantes e docentes, passaram a ter

interesse no campo, traçando, assim, novas abordagens para a MDE. Os grupos e alguns exemplos de interesses são exemplificados pelos autores no Quadro 9.

Quadro 9 — Grupos de interesses na MDE

Grupo	Exemplos de interesses
Estudantes Aprendizes	<ul style="list-style-type: none"> • Personalização de ambientes; • Identificação de estratégias de aprendizado; • Recomendação de recursos de apoio; • Sugestão de tarefas de reforço; e, • Identificação e tratamento de lacunas de aprendizado.
Instrutores Professores Tutores	<ul style="list-style-type: none"> • Análise do aprendizado e do comportamento dos estudantes; • Identificar estudantes que requerem apoio diferenciado; • Agrupamento de aprendizes; • Caracterização de estratégias de ensino bem-sucedidas; • Aprimoramento na customização de cursos e materiais didáticos; e, • Predição de desempenho discente.
Pesquisadores Desenvolvedores de Cursos e Materiais Didáticos	<ul style="list-style-type: none"> • Avaliação de material didático; • Avaliação de cursos; • Aprimoramento do aprendizado discente; • Estudo comparativo de técnicas de mineração; e, • Desenvolvimento de ferramentas de mineração.
Organizações Patrocinadores	<ul style="list-style-type: none"> • Recomendação de cursos para grupos de estudantes; • Aprimoramento do processo decisório nos investimentos em educação; e, • Apoio no processo seletivo de admissão de estudantes.
Gestores Diretores Administradores	<ul style="list-style-type: none"> • Apoio no investimento em recursos educacionais; • Uso eficiente de recursos educacionais; • Avaliação das ofertas de cursos e necessidades de melhorias; • Avaliação de instrutores e currículos; e, • Apoio na configuração de <i>websites</i> e serviços.

Fonte: (GOLDSCHMIDT, BEZERRA e PASSOS, 2015).

No próximo capítulo, Trabalhos Correlatos, serão apresentados os trabalhos relacionados com a presente pesquisa.

5 TRABALHOS CORRELATOS

A evasão em cursos de ensino superior, sejam a distância ou presencial, é um dos temas mais explorados na MDE tanto com a criação de modelos preditivos quanto no desenvolvimento de ferramentas que apoiem os gestores na detecção de estudantes em risco. Cada vez mais trabalhos buscam melhorar o processo de aprendizagem com a aplicação de alguma tecnologia educacional. Dessa forma, este capítulo visa descrever algumas das pesquisas que estão sendo realizadas neste cenário.

5.1 DESCRIÇÃO DOS TRABALHOS

No trabalho apresentado por Kostopoulos, Sotiris e Panagiotis (2015), em uma turma do curso de informática da *Hellenic Open University*, foi desenvolvida a ferramenta *Semi-Supervised Prediction*, em que um conjunto de dados pessoais do aluno como sexo, idade, estado civil, quantidade de filhos, tempo de trabalho e conhecimentos em informática era solicitado pelos docentes para que, a partir deles, uma análise fosse realizada por meio da ferramenta KEEL com a utilização dos algoritmos de MD C4.5 e *Naive Bayes*. Após a análise, era disponibilizado um percentual referente à probabilidade de um determinado aluno evadir ou não do curso. Os resultados alcançados na pesquisa tiveram uma acurácia entre 60% e 84% sendo considerados satisfatórios para o reconhecimento dos estudantes que possuem um risco elevado de abandonar o curso. Na Figura 8 é possível visualizar a interface da ferramenta desenvolvida pelos autores.

Figura 8 — *Semi-Supervised Prediction Tool*

The screenshot shows a web browser window with the URL `file:///C:/Users/GeorgeK/Desktop/paper1-Tool.html`. The page title is "SEMI-SUPERVISED PREDICTION TOOL for students dropout". The form contains the following fields:

Gender: male	OCS1 : absent
Age: <32	TEST1: 5-7
Domestic: single	OCS2 : present
Children: 0	TEST2: 5-7
Work: part-time	
ComputKnowledge: yes	
ComputerJob: no-related	

Below the form are two buttons: "Predict" and "Clear". At the bottom of the page, it displays "Probability of Dropout = 18.75 %".

Fonte: (KOSTOPOULOS, SOTIRIS e PANAGIOTIS, 2015).

Já no estudo realizado por Rigo e Cazella (2014), em cursos de graduação na modalidade a distância, foram utilizados dados relacionados à interação dos discentes no AVEA como a quantidade de participações em fóruns e *chats*, o tempo de acesso, a quantidade de acessos total e resultados de questionários e avaliações. Por meio de uma ferramenta WEB, desenvolvida pelos autores, era disponibilizada para os docentes a listagem dos alunos com tendência a evadir. Na ferramenta apresentada, era possível que o docente/tutor informasse quais foram as ações pedagógicas realizadas a fim de evitar a desistência. Tais dados eram utilizados, posteriormente, pelos demais docentes, para que as ações que apresentaram bons resultados pudessem ser reutilizadas em outros casos. A pesquisa foi validada em duas turmas sendo que, em uma delas, utilizou-se a ferramenta e efetuadas intervenções junto aos alunos que nela foram apontados como possíveis desistentes. De acordo com os autores, essa turma apresentou redução média de 11% na evasão em comparação com a outra turma em que não houve intervenção. Na Figura 9 é possível visualizar a interface da tela de listagem dos alunos da previsão da evasão.

Figura 9 — Interface da tela de lista de predição de evasão



The screenshot shows a web interface titled "Acesso e Predição de Evasão". It includes a search bar, buttons for "Registrar Ação" and "Voltar", and a "LISTA DE ESTUDANTES" section with filters for "Ativos" and "Inativos". Below the filters are three icons: "Ações encerradas", "Ações com alerta", and "Ações com alerta para você". The main part of the interface is a table with the following columns: #, Ações, Nome, 1º Acesso, Último Acesso, Nota GA, Módulo 3, Módulo 4, Módulo 5, and Módulo 6. The table contains 11 rows of student data. Rows 4, 5, 7, 8, 9, and 10 have a red background, indicating a predicted dropout risk. Row 10 also has a yellow icon in the "Ações" column.

#	Ações	Nome	1º Acesso	Último Acesso	Nota GA	Módulo 3	Módulo 4	Módulo 5	Módulo 6
1			Módulo 1	9 horas 24 min.	10,00	0,01	0,01	0,01	0,01
2			Módulo 1	11 horas 44 min.	9,86	1,36	1,36	1,43	1,51
3			Módulo 1	20 horas 54 min.	10,00	0,01	0,01	0,01	0,01
4			Módulo 1	18 horas 3 min.	0,00	99,89	100,00	100,00	100,00
5			Módulo 1	22 horas 32 min.	2,66	88,82	88,96	72,55	75,88
6			Módulo 1	18 horas 9 min.	9,94	0,53	0,53	0,57	0,59
7			Módulo 1	1 dia 14 horas	0,82	85,54	86,07	91,91	95,72
8			Módulo 2	18 horas 20 min.	0,00	97,80	98,41	100,00	100,00
9			Módulo 1	1 dia	0,00	99,98	100,00	100,00	100,00
10			Nunca	Nunca	0,00	99,98	100,00	100,00	100,00
11			Módulo 1	45 min.	0,00	0,54	0,59	0,15	0,55

Fonte: (RIGO e CAZELLA, 2014).

Kantorski *et al.* (2016) propõem uma abordagem com o objetivo de prever a evasão no curso de graduação em Administração, na modalidade presencial, aplicada sobre os dados de 791 estudantes da UFSM. Foram utilizados para o processo de MD dados pessoais, acadêmicos e sobre a participação em programas de auxílio estudantil. Antes do final de cada semestre letivo, foi disponibilizada uma lista com possíveis alunos com tendência a evadir. Ao

encerrar o período de matrículas para o semestre seguinte foi verificado quais alunos previstos não tinham realizado a matrícula, ou seja, evadiram. Para MD foram aplicados os algoritmos CART, J48, J48-Bin e *Naive Bayes*, em que foi obtida uma precisão de até 73% no número de acertos.

A pesquisa de Manhães (2015) apresenta uma proposta de arquitetura baseada em MDE para identificar precocemente alunos com risco de evasão nos cursos de graduação em Engenharia Civil, Mecânica e da Produção, na modalidade presencial, da Universidade Federal do Rio de Janeiro (UFRJ). Foram empregados dados relacionados ao desempenho dos alunos em duas disciplinas do primeiro semestre do curso em que foram utilizados os algoritmos *SimpleLogistic*, *SimpleCart*, J48, *Random Forest*, OneR, JRip e *DecisionTable* na ferramenta WEKA. Dentre os resultados foi alcançada uma acurácia média entre 75% e 80%.

Detoni, Araújo e Cechinel (2015), em uma abordagem semelhante, buscaram detectar os alunos por meio da utilização da contagem de interações no AVEA MOODLE em disciplinas na EaD. O experimento foi realizado na Universidade Federal de Pelotas (UFPEL), nos cursos de Licenciatura em Educação do Campo e Licenciatura em Pedagogia. A partir das interações foram calculadas as médias de utilização do AVEA por parte dos alunos, tutores e professores e, posteriormente, foram aplicados os algoritmos de Redes Bayesianas, Redes Neurais, J48 e *Random Forest*, onde foram obtidos resultados de até 67% de acurácia na predição do desempenho do aluno. Uma extensão dessa pesquisa pode ser vista também em Queiroga *et al.* (2016) em que a abordagem foi aplicada na predição de evasão de alunos de dois cursos técnicos a distância. Nesta foram utilizadas somente a simples contagem de interações de alunos no AVEA durante as primeiras semanas do curso. Apenas com esses dados foi obtida uma acurácia de 98,67% na metade do curso.

Burgos *et al.* (2017) empregaram a MD para a criação de modelos de predição para detectar o risco da evasão dos alunos. Foram utilizados dados de 104 alunos de diversos cursos de curta duração a distância. Até a quarta semana de curso foi obtido 100% de acurácia nos acertos. De acordo com os autores, a aplicação da técnica possibilitou a redução aproximada de 14% no abandono escolar nos cursos onde foram aplicados.

Já no trabalho desenvolvido por Bayer *et al.* (2012) foram empregadas informações dos alunos a partir do sistema da Universidade de Masaryk – República Tcheca. Além das informações acadêmicas do aluno, como notas e faltas, o sistema forneceu informações do tipo: amizade explícita, troca de e-mail, coautoria de publicação, dentre outros. Por meio da utilização de dados de desempenho nos algoritmos de MD, em conjunto com dados das redes

sociais, foi obtida uma acurácia entre 89,89% e 93,67% no final do curso de Bacharelado em Informática Aplicada.

Por fim, o trabalho de Cássia, Brito e Medeiros (2016) buscou verificar o potencial do uso de técnicas de Análise de Redes Sociais (ARS) entre alunos de um curso EaD que realizam interações em fóruns de discussões no AVEA MOODLE. Por meio da aplicação das métricas de centralidades de grau, intermediação e centralidade, foi possível detectar a existência de correlações entre as métricas e os desempenhos finais dos alunos em uma disciplina de um curso técnico de uma instituição federal de ensino.

5.2 ANÁLISE DOS TRABALHOS CORRELATOS

Ao analisar os trabalhos apresentados, é possível visualizar que existem diversas pesquisas na área, mas, no entanto, elas se diferenciam em relação à modalidade de ensino em que são aplicadas, dados utilizados na previsão, ferramentas de apoio ao processo de KDD e resultados demonstrados. Nos Quadros 10 e 11 podem ser vistas a sintetização destas diferenças.

Quadro 10 — Análise dos trabalhos relacionados (Parte 1)

Autor	Dados analisados	Técnicas/Algoritmos utilizados	Ferramentas de KDD	Tipo de ensino	Aspectos positivos	Aspectos negativos
Kostopoulos, Sotiris e Panagiotis (2015)	Pessoais dos alunos.	C4.5 e <i>Naive Bayes</i>	KEEL	EaD	Desenvolvimento de uma ferramenta que permite a avaliação da estimativa de evasão por uma pessoa que não é especialista na área de MD.	A ferramenta desenvolvida é limitada a disponibilização de um percentual relacionada a probabilidade da evasão. Não demonstra os fatos que determinaram a estimativa.
Rigo e Cazella (2014)	Interação dos discentes no AVEA.	<i>MultiLayer Perceptron</i>	Não especificada	EaD	Disponibilização de uma ferramenta que permite o usuário visualizar a lista de alunos com tendência a evadir. Além disso, permite a descrição de ações pedagógicas realizadas a fim de evitar o abandono.	Não permite a análise dos fatores causadores da evasão para apoiar a tomada de decisões.
Kantorski <i>et al.</i> (2016)	Pessoais, acadêmicos e assistência estudantil.	J48, J48-Bin, CART, <i>Naive Bayes</i>	WEKA	Presencial	Alta taxa de acertos.	Não permite a análise dos fatores causadores da evasão por parte dos gestores.
Manhães (2015)	Dados de desempenho dos estudantes em duas disciplinas.	<i>Simple Logistic</i> , <i>SimpleCart</i> , J48, <i>Random Forest</i> , OneR, JRip e <i>DecisionTable</i>	WEKA	Presencial	Proposta de uma nova arquitetura para previsão da evasão e alto índice de acertos.	Não permite a análise dos fatores causadores da evasão por parte dos gestores.

Fonte: Autor.

Quadro 11 — Análise dos trabalhos relacionados (Parte 2)

Autor	Dados analisados	Técnicas/Algoritmos utilizados	Ferramentas de KDD	Tipo de ensino	Aspectos positivos	Aspectos negativos
Detoni, Araújo e Cechinel (2015)	Contagem de interações nas disciplinas no AVEA.	Redes Bayesianas, Redes Neurais, J48 e <i>Random Forest</i>	WEKA	EaD	Elevadas taxas de acerto na previsão.	Direcionado aos aspectos técnicos de MD sem preocupações referente a gestão acadêmica.
Burgos <i>et al.</i> (2017)	Contagem de interações no AVEA.	FEED Forward, <i>Neural Network</i> (FFNN), <i>Support Vector Machine</i> , ARTMAP	Não especificada.	EaD	Alto percentual de acertos na previsão da evasão de alunos.	Restrição na análise dos modelos avaliados.
Bayer <i>et al.</i> (2012)	Informações acadêmicas e sociais dos alunos.	J48, IB1, PART, <i>Support Vector Machine</i> , <i>Naive Bayes</i>	WEKA	Presencial	Alto percentual de acertos na previsão da evasão de alunos.	Restrição na análise dos modelos avaliados.
Cássia, Brito e Medeiros (2016)	Interações em fóruns no AVEA.	Análise de Redes Sociais	Não especificada	EaD	Viabilidade da abordagem de análise de redes sociais para previsão de evasão de alunos.	Restrição na análise da viabilidade da utilização da abordagem.

Fonte: Autor.

No que tange à modalidade de ensino, a EaD aparece em cinco trabalhos e a presencial em três. Percebe-se, também, que este fato impacta diretamente nos dados a serem utilizados para análise, já que a maioria das pesquisas na EaD emprega dados relacionados à interação realizada pelos alunos no AVEA, e as abordagens no ensino presencial tomam como base os dados disponíveis nos sistemas de gerenciamento acadêmico, como dados pessoais e acadêmicos.

Em relação às ferramentas de apoio ao processo de descoberta de conhecimento, dos trabalhos que citaram sua utilização, apenas um não empregou a ferramenta WEKA. Os fatores que justificam a sua ampla adoção podem ser vistos na seção 4.3 deste trabalho.

No que compete aos resultados, verifica-se que a maioria dos trabalhos busca demonstrar que os modelos gerados pelos algoritmos/técnicas de MD apresentam boas taxas de acertos e podem ser utilizados para previsão da evasão de alunos. Entretanto, alguns trabalhos como Kostopoulos, Sotiris e Panagiotis (2015) e Rigo e Cazella (2014), além da demonstração dos modelos, buscaram prover soluções para permitir que docentes/gestores façam uso dos resultados do processo de KDD. No primeiro trabalho é calculado um percentual da probabilidade de um aluno evadir com base em um conjunto de dados fornecidos pelo usuário. E, no segundo, os alunos com tendência à evasão são relacionados por meio de uma ferramenta que permite, também, o registro de informações relacionadas a ações pedagógicas efetivadas para evitar o abandono.

A partir dos trabalhos analisados, fica aparente uma lacuna na área de predição de evasão, já que a ampla maioria das pesquisas se detém nos aspectos mais técnicos relacionados à MD e, os trabalhos que disponibilizam alguns dos seus resultados, não permite uma análise, por parte da gestão acadêmica, dos fatores que tiveram impacto na sua geração. Desse modo, este trabalho visa propiciar uma ferramenta de apoio aos gestores que demonstre os prováveis alunos a evadirem em conjunto com os fatores que possam ter sido determinantes para enquadrá-los nesta situação.

Sendo assim, após a conclusão da análise, serão abordados no próximo capítulo os aspectos metodológicos utilizados para alcançar o objetivo proposto na pesquisa.

6 ASPECTOS METODOLÓGICOS

Primeiramente, por meio de uma pesquisa bibliográfica, procurou-se um maior entendimento em relação aos temas EaD, UAB, evasão na EaD, bem como suas causas. Na sequência foi apresentada uma RSL com o objetivo de conhecer as abordagens tecnológicas que estão sendo empregadas na previsão de evasão de alunos. Logo em seguida foram abordados os aspectos relacionados ao processo de descoberta de conhecimento em bases de dados (KDD).

Depois disso, para representar a relevância e contextualizar este trabalho, buscou-se por trabalhos correlatos que da mesma forma propuseram a utilização de técnicas de MD para previsão da evasão de alunos. Dessa maneira, foi possível também verificar as principais características de cada trabalho, permitindo, no final, diferenciá-las em relação à abordagem proposta na presente pesquisa. A UFSM, IES objeto desta pesquisa será brevemente descrita na próxima seção.

6.1 A UNIVERSIDADE FEDERAL DE SANTA MARIA (UFSM)

A UFSM foi a pioneira na interiorização do ensino superior no Brasil, sendo a primeira universidade federal criada fora do eixo das capitais brasileiras (PDI/UFSM/2016-2026, p.14). Esse fato contribuiu para o Rio Grande do Sul tornar-se o primeiro estado da federação a contar com duas universidades federais.

A instituição oferece ensino nas modalidades presenciais e a distância e possui cursos, programas e projetos nas mais diversas áreas do conhecimento humano. Possui em sua estrutura: restaurantes universitários, biblioteca central e setoriais, laboratórios de ensino e de pesquisa, hospital veterinário universitário, dentre outros. O corpo discente é constituído por mais de 30000 estudantes em todas as modalidades de ensino. O quadro de pessoal conta com 4769 servidores, incluindo docentes e técnicos administrativos em educação (PDI/UFSM/2016-2026, p. 17).

De acordo com o Plano de Desenvolvimento Institucional (PDI) 2016-2026 da instituição, a UFSM é a 16ª melhor universidade do Brasil de acordo com a pontuação do Índice Geral de Cursos (IGC) no ano de 2014.

A instituição atualmente oferta cursos de graduação e pós-graduação na modalidade EaD. Dentre os cursos de graduação estão: Administração Pública, Educação Especial, Formação de Professores para Educação Profissional, Licenciatura em Física, Licenciatura em

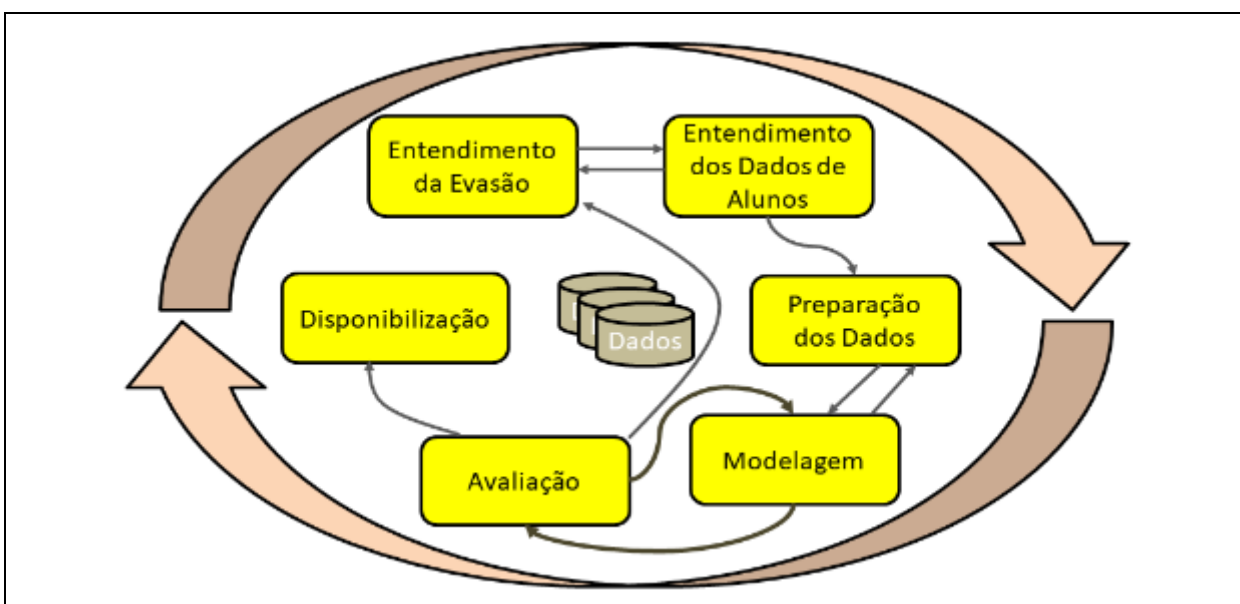
Geografia, Letras – Espanhol e Literaturas, Letras – Português e Literaturas, Licenciatura em Sociologia, Pedagogia, Tecnólogo em Agricultura Familiar e Sustentabilidade, Licenciatura em Computação, Licenciatura em Educação do Campo e Ciências da Religião. Na seção a seguir será abordada a metodologia adotada para guiar o processo de KDD.

6.2 METODOLOGIA PARA KDD

Do ponto de vista do processo de KDD a metodologia empregada é baseada no modelo proposto por Kantorski *et al.* (2016). O modelo mencionado foi adotado nesta pesquisa pelo objetivo comum compartilhado por ambas que é a descoberta de alunos com tendência à evasão. No entanto, o trabalho proposto por Kantorski *et al.* (2016) aplica-se no âmbito de cursos de graduação ofertados na modalidade de ensino presencial.

A metodologia em questão consiste de uma adaptação da metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) (CHAPMAN *et al.*, 2000) que define e padroniza as fases e atividades do processo de KDD, sendo aplicada especificamente para este fim. Dentre as principais características, relacionadas a sua utilização, destaca-se o fato de ser uma metodologia genérica que não se limita na utilização de uma ferramenta ou tecnologia específica e possui relações com KDD. A Figura 10 ilustra as fases estabelecidas para a metodologia.

Figura 10 — Metodologia CRISP-DM adaptada para evasão



De acordo com Kantorski *et al.* (2016), as etapas da metodologia compreendem:

- a) Entendimento da evasão: visa identificar os objetivos sob o ponto de vista do KDD. Fazem parte destes objetivos o tipo de tarefa a ser executada como, por exemplo, previsão da evasão e os critérios de avaliação dos modelos utilizados;
- b) Entendimento dos dados de alunos: relaciona-se com a coleta preliminar dos dados e na sua descrição e exploração por meio da análise de suas propriedades e qualidades. Entende-se por exploração a análise dos dados juntamente com o emprego de técnicas de visualização. Já a descrição compreende a avaliação de propriedades, tais como faixas de valores, número de atributos, significado de cada atributo e sua importância para atingir o objetivo desejado;
- c) Preparação dos dados: nesta fase um conjunto de dados é gerado de acordo com os modelos de descoberta de conhecimento que serão utilizados. Para obtenção deste conjunto é necessária a realização de uma sequência de passos tais como: seleção de dados e das instâncias – realizada manualmente ou por meio de algoritmos; limpeza – visa a melhoria na qualidade dos dados como remoção de dados com erros e padronização; construção dos dados – consiste na normalização, na transformação de valores simbólicos para numéricos ou vice-versa e na discretização dos atributos; por fim, a integração – objetiva a combinação de múltiplas fontes de dados e a sua formatação sem alterar o significado;
- d) Modelagem: é a etapa de descoberta de conhecimento propriamente dita. Nesta etapa são definidos os modelos que serão utilizados para a solução do problema;
- e) Avaliação: compreende a interpretação e a avaliação, propriamente dita, dos resultados em relação aos objetivos de descoberta de conhecimento definidos na primeira etapa. Entre a presente etapa e a anterior há um ciclo que permite a avaliação e comparação entre os diversos modelos de aprendizagem; e,
- f) Disponibilização: dispõe os resultados obtidos por meio da comparação entre os múltiplos modelos gerados.

A aplicação da metodologia descrita será realizada em duas etapas. Na primeira, descrita na seção 7.2, a proposta foi aplicada em apenas um curso de graduação EaD da UFSM para avaliar os resultados preliminares e possíveis melhorias que pudessem ser realizadas nos dados para lapidar os resultados da mineração. Já na segunda etapa, a mineração foi aplicada em mais

dois cursos de graduação, totalizando três, de forma a obter um modelo genérico que possa ser aplicado nos demais cursos ofertados, bem como em outras IES.

Na próxima subseção são detalhadas as tecnologias utilizadas para apoiar o desenvolvimento do trabalho.

6.3 TECNOLOGIAS UTILIZADAS

Para implementação da ferramenta proposta foi utilizada a plataforma de desenvolvimento para aplicativos móveis Android⁷. Esta plataforma foi escolhida, visando facilitar o acesso aos dados por parte dos gestores por meio dos seus próprios dispositivos móveis. Além disso, por ela ser gratuita e incluir um dos sistemas operacionais mais utilizados em dispositivos móveis no mundo.

Já com o objetivo de apoiar o processo de KDD, foi utilizada a ferramenta WEKA descrita na seção 4.3.

6.4 METODOLOGIA DE DESENVOLVIMENTO DA FERRAMENTA

Para apoiar o desenvolvimento da ferramenta, foi escolhida a metodologia *Mobile-D* proposta por Abrahamsson *et al.* (2004). A escolha justifica-se por ser uma metodologia específica para o desenvolvimento de aplicativos para dispositivos móveis. Além disso, é baseada nas melhores práticas das abordagens *EXtreme Programming* (XP), *Crystal Methodologies* e *Rational Unified Process* (RUP) consolidadas no desenvolvimento de *softwares* corporativos para as mais diversas finalidades.

De acordo com Abrahamsson *et al.* (2004), o uso da metodologia é recomendado para equipes pequenas, de no máximo 10 pessoas, e com prazo de entrega do produto não ultrapassando as 10 semanas. O *Mobile-D* está dividido em cinco fases em que cada uma delas possui uma série de etapas e práticas associadas. As fases da metodologia são descritas a seguir:

- a) Exploração: compreende a geração de um plano pela equipe de desenvolvimento contendo todas as características do projeto. Esta fase é realizada em três etapas: identificar as partes interessadas, definir o escopo e estabelecer o projeto;
- b) Inicialização: nesta fase devem ser compreendidos os aspectos relacionados ao produto em desenvolvimento e a preparação dos recursos necessários para as

⁷ Disponível em: <https://www.android.com>.

atividades de produção como recursos físicos, tecnológicos e de comunicação. Isto é realizado por meio das etapas: configuração do projeto, planejamento inicial e dia de avaliação;

- c) Produção: envolve essencialmente as atividades de implementação. No término desta fase maior parte da aplicação deve estar pronta. Esta fase é contemplada pelas etapas dias de planejamento, trabalho e lançamento. Na etapa de dias de planejamento, entre outras tarefas são especificados os testes de aceitação que serão executados na etapa dias de lançamento;
- d) Estabilização: fase utilizada para finalização do produto; e,
- e) Testes e reparos: como o próprio nome indica, fase que compreende os testes e eventuais reparos no produto.

Na seção seguinte, será discutida a metodologia empregada para avaliação do aplicativo e dos dados disponibilizados.

6.5 METODOLOGIA DE AVALIAÇÃO

Nesta pesquisa, a metodologia de avaliação a ser utilizada no âmbito do KDD já faz parte do seu processo, tomando como base as métricas de avaliação dos resultados obtidos pelos modelos gerados durante a MD. As métricas e suas respectivas definições podem ser vistas na seção 4.2.3.6.1.

Em relação à análise da ferramenta proposta foi utilizado o *framework* DECIDE desenvolvido por Preece, Rogers e Sharp (2013). Esse *framework* orienta o planejamento, a execução e a análise do ponto de vista de usabilidade e aspectos de interação, bem como os objetivos/funcionalidades junto aos especialistas da área, no caso, os coordenadores de curso. As etapas estabelecidas pelo *framework* podem ser vistas a seguir:

- a) Determinar os objetivos da avaliação: nesta etapa devem ser estabelecidos os objetivos a serem atingidos com a avaliação, bem como os possíveis interessados pelos seus resultados. As demais etapas serão delineadas por estes objetivos;
- b) Explorar as questões a serem respondidas: fase onde são elaboradas as questões a serem respondidas para se obter as informações desejadas na avaliação;
- c) Escolher os métodos de avaliação: após a definição dos objetivos e questões, nesta etapa seguinte devem ser definidos os métodos mais adequados para obtê-los;

- d) Identificar as questões práticas da avaliação: nesta etapa devem ser identificadas e administradas as questões práticas relacionadas a avaliação como, por exemplo, o recrutamento de participantes, a preparação e o uso de equipamentos, prazos, dentre outros;
- e) Decidir como lidar com questões éticas: quando a avaliação envolver a presença de usuários devem ser tomados todos os cuidados éticos necessários. Os participantes de nenhuma forma podem ser desrespeitados ou prejudicados em decorrência da avaliação; e,
- f) Avaliar, interpretar e apresentar os dados: nesta fase os resultados obtidos são analisados e posteriormente divulgados para os interessados.

No próximo capítulo será apresentado o processo de desenvolvimento do trabalho.

7 DESENVOLVIMENTO

Neste capítulo será apresentado o contexto de aplicação da pesquisa, bem como os experimentos realizados para seu desenvolvimento.

7.1 CONTEXTO

Para o desenvolvimento da pesquisa foram utilizados dados dos cursos de Letras – Espanhol, Pedagogia e Educação Especial ofertados na modalidade EaD pela UAB - UFSM. A escolha fundamenta-se pelo tempo de existência dos cursos, sendo que eles foram os precursores na modalidade EaD na instituição e, consecutivamente, fornecem um maior conjunto de dados para serem analisados pelos algoritmos de MD.

Os cursos analisados são ofertados em diversos polos de ensino espalhados pelo interior do estado do Rio Grande do Sul, e possuem as disciplinas de sua estrutura curricular distribuídas em oito semestres letivos. Dessa forma, no âmbito da pesquisa, será enquadrado como evadido o aluno que não se matriculou em pelo menos uma disciplina durante o decorrer de um semestre, ou seja, não manteve o vínculo com a instituição durante o período.

Para fins de análise dos dados o período de abrangência considerado foi o semestral. Isso porque algumas disciplinas têm seus conteúdos abordados semanalmente, já outras, quinzenalmente. Desse modo, a análise semestral dos dados abrange todas as atividades realizadas pelo aluno durante a realização de uma disciplina. Outro fator que motivou essa definição relaciona-se com a descoberta da ocorrência da evasão, que somente é possível no início de cada semestre, em decorrência das rematrículas.

O desenvolvimento deste projeto foi dividido em dois experimentos. O primeiro foi realizado junto ao projeto de qualificação da pesquisa abrangendo os dados do curso de Letras – Espanhol. Este experimento dispôs da previsão da evasão para o segundo semestre de 2017. Logo após a qualificação, foi realizado o segundo experimento, abrangendo também os cursos de Pedagogia e Educação Especial com o objetivo de validar a primeira etapa. Para o segundo experimento a previsão foi efetuada para o primeiro semestre de 2018. Desse modo, nas seções 7.2 e 7.3 serão detalhados os experimentos realizados.

7.2 PRIMEIRO EXPERIMENTO

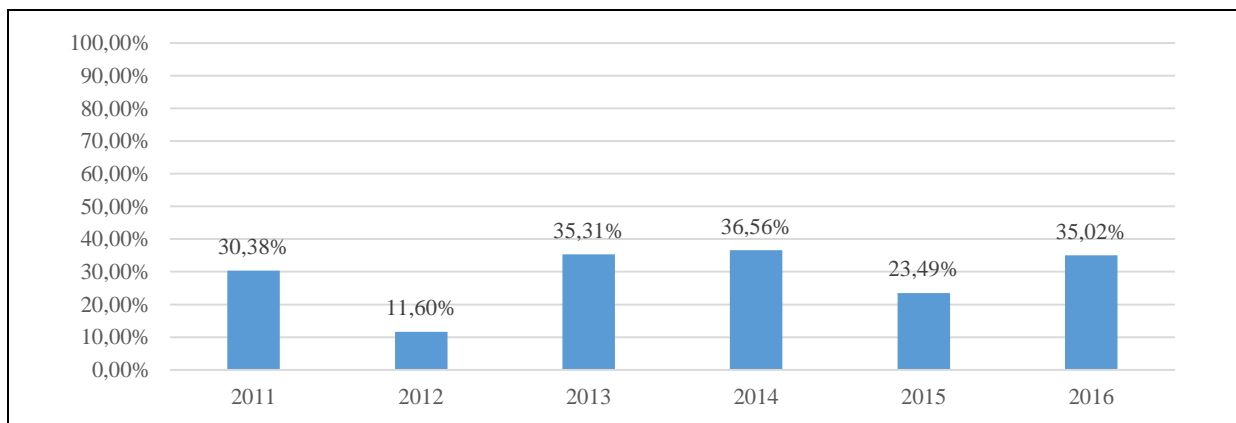
Nesta seção serão abordadas as etapas da metodologia de KDD seguidas na elaboração do primeiro experimento.

7.2.1 Entendimento da evasão

Conforme detalhado na seção 6.2, a aplicação da metodologia de KDD inicia-se por meio da compreensão dos dados da evasão. A partir deles, devem ser estipulados o tipo de tarefa a ser executada e os critérios para avaliação dos modelos utilizados.

Para analisar a evasão do curso selecionado no primeiro experimento, Letras – Espanhol, foram calculadas as taxas de evasão total e evasão anual média detalhadas na seção 2.2 considerando os dados dos alunos registrados no sistema acadêmico da instituição. Sendo assim, em relação a taxa de evasão total, foi encontrado o índice de 50,91%. Já a média em relação a taxa anual de evasão, considerando os dados de 2011 a 2016, foi de 28,73%. No Gráfico 7 é possível visualizar as médias anuais durante o período compreendido no cálculo.

Gráfico 7 — Evasão anual média – Primeiro experimento



Fonte: Autor.

O Gráfico 7 contém os dados do período compreendido entre os anos de 2011 e 2016. Isso porque o ano de 2011 é o primeiro período dos dados considerados para análise, alunos ingressantes após o segundo semestre de 2010, em que é possível calcular o índice. E, o ano de 2017 foi desconsiderado por estar em andamento no momento da elaboração do experimento.

Os dados apresentados revelam um cenário preocupante. Com base no valor de evasão geral do curso, percebe-se que mais da metade dos alunos que ingressaram no curso não obtiveram sua conclusão. Além disso, em entrevista com os coordenadores, foi levantado, além dos fatores obtidos na revisão de literatura das causas da evasão, um outro fator que pode alavancar ainda mais os índices futuros. Este fator está relacionado com mudanças curriculares que removeram a obrigatoriedade da disciplina de Espanhol da base curricular do ensino médio. Tal fator gera uma demanda menor pelos profissionais da área que, consecutivamente, acarreta na sua desistência pela menor oferta de trabalho. Diante dos fatos relacionados, ratifica-se a importância do presente trabalho seja pela possibilidade de diminuir os índices apresentados, como também por permitir a identificação dos possíveis desistentes para que ações possam ser tomadas a fim de mantê-los no curso.

Desse modo, com base nos dados apresentados, a tarefa a ser executada consiste na previsão da evasão de alunos do curso de graduação em Letras – Espanhol a distância, utilizando dados de interações do AVEA MOODLE com o intuito de diminuir os índices elencados. Como resultado paralelo é possível delinear os fatores mais influentes para a previsão.

7.2.2 Entendimento dos dados de alunos

Esta etapa visa a análise dos dados dos alunos disponíveis para o processo de MD para uma melhor compreensão dos mesmos e, consecutivamente, com base na revisão de literatura, possibilitar a projeção das informações disponíveis que possam influenciar na desistência do discente no curso.

Os dados utilizados na pesquisa estão armazenados em duas fontes de dados distintas. Na primeira, estão as informações relacionadas às interações dos alunos no AVEA MOODLE. Estas informações consistem de logs brutos gerados pelo sistema durante cada ação realizada pelo estudante. Tais dados foram concedidos pelo Núcleo de Tecnologia Educacional (NTE) da UFSM no formato de planilhas eletrônicas (CSV) sendo que a autorização para sua utilização foi obtida junto aos coordenadores do curso analisado, conforme pode ser visto no Anexo B. Os atributos armazenados nas planilhas disponibilizadas podem ser vistos no Quadro 12.

Quadro 12 — Atributos armazenados nas planilhas disponibilizadas do AVEA MOODLE

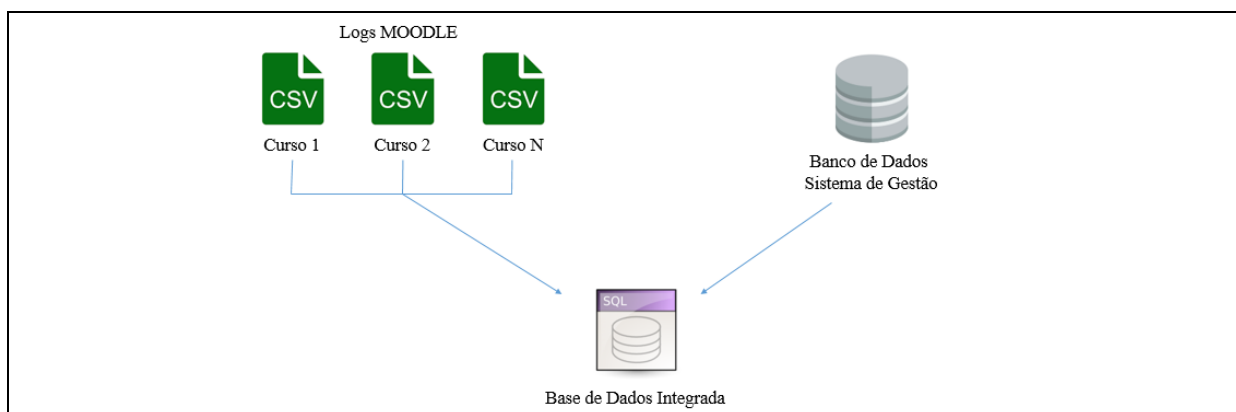
Atributo	Descrição
matricula	Matrícula do aluno que efetuou a interação.
nome_curso	Nome do curso em que a disciplina foi cursada.
semestre_letivo	Semestre letivo de realização da disciplina.
cod_disciplina	Código do curso e da disciplina em que a interação foi realizada.
nome_disciplina	Nome da disciplina em que a interação foi realizada.
componente_log	Recurso/Componente do AVEA utilizado pelo aluno.
acao_log	Ação realizada no recurso/componente do AVEA pelo aluno.
data_ano_mes_acesso	Mês e ano em que a ação foi executada.
count	Quantidade de interações realizadas com base nos atributos cod_disciplina, componente_log, acao_log e data_ano_mes_acesso.

Fonte: Base de dados AVEA MOODLE.

A segunda fonte de dados armazena as informações acadêmicas dos alunos, tais como ano e semestre de ingresso, que não foram disponibilizadas na primeira fonte. Estas informações encontram-se armazenadas no banco de dados relacional do sistema de gestão acadêmico utilizado na instituição, e foram disponibilizadas pelo Centro de Processamento de Dados (CPD) e pela Pró-reitoria de Graduação (PROGRAD) da UFSM, com quem foi assinado um termo de autorização de uso dos dados, conforme pode ser visto no Anexo A.

Para facilitar a compreensão e manipulação das informações, as fontes de dados descritas foram integradas em um mesmo banco de dados conforme pode ser visualizado na Figura 11.

Figura 11 — Esquema de integração das fontes de dados utilizadas



Fonte: Autor.

Este processo precisou ser realizado manualmente já que os dados de cada curso foram disponibilizados em diversas planilhas. Por meio desta unificação, foi possível a extração dos dados mediante a utilização da linguagem de consulta SQL (*Structured Query Language*).

Diante dos atributos das tabelas de logs do MOODLE, demonstrados no Quadro 12, merecem destaque os atributos MATRICULA, ACAO_LOG e COMPONENTE_LOG. Através do primeiro é possível relacionar os dados de logs do MOODLE com os dados do sistema de gestão acadêmico da instituição. O segundo permite determinar a ação realizada pelo aluno no ambiente como, por exemplo, o valor “assign_submit” especifica a ação de submissão de uma tarefa pelo aluno. E, o terceiro, qual é o módulo/recurso do MOODLE em que esta ação foi efetuada, por exemplo o valor “forum” determina que a ação foi efetuada durante a utilização da ferramenta Fórum. Nos dados disponibilizados existem 109 diferentes ações realizadas em 48 módulos gerados por 571 alunos, perfazendo um total de 6.571.140 interações. Além disso, os dados extraídos estão compreendidos entre os anos de 2010 e 2017.

Os dados acadêmicos encontrados no sistema de gestão, como ano e semestre de ingresso do aluno, ano e semestre de evasão/conclusão do curso, são utilizados por permitirem o cálculo da quantidade de interações realizadas pelo aluno de acordo com o número de semestres cursados. Este cálculo é importante, sendo que a utilização da quantidade total de interações realizadas pelo aluno no decorrer do curso pode influenciar no resultado da mineração, pois os alunos que estão em semestres mais avançados tendem a acumular um maior número de interações no AVEA.

Devido ao desenvolvimento de novas versões do MOODLE que, no decorrer do tempo, foram adotadas pela instituição, o conjunto de dados dos logs de monitoração disponíveis são provenientes do segundo semestre do ano de 2010. Desse modo, para não distorcer os dados a serem avaliados pela mineração, os alunos que ingressaram em períodos anteriores foram descartados da análise.

7.2.3 Preparação dos dados

O foco desta etapa se deu no conjunto de tabelas de logs brutos do MOODLE. Primeiramente, foram analisados os dados dos atributos “nome_disciplina” para a realização do levantamento dos recursos do AVEA que são mais utilizados no curso. Na análise verificou-se que algumas disciplinas tinham seus nomes registrados de diversas formas que, apesar de se referirem ao mesmo nome, se distinguem no conjunto de dados. Essas ocorrências foram ocasionadas devido as mudanças na sua descrição no decorrer do curso. Nestes casos, o nome

que foi adotado para remover essas inconsistências é o que está atualmente definido para a disciplina na base de dados do sistema de gestão acadêmica.

No próximo passo foram integrados os dados do sistema acadêmico com os dados do MOODLE por meio do atributo de matrícula dos alunos comum em ambas as fontes de dados. Essa integração teve como objetivo verificar se os dados das interações registrados nos logs correspondiam aos dados dos alunos dos cursos analisados. Ou seja, buscou-se verificar se não havia dados de alunos que estavam realizando a disciplina no curso em questão, mas pertenciam a um outro curso. Como resultado da abordagem foram encontrados 32 registros de 2 alunos que estavam nesta situação que foram, consecutivamente, desconsiderados do processo de mineração.

As informações do atributo “data_ano_mes_acesso” foram transformados por meio da separação dos seus valores em dois novos atributos: “ano_acesso” e “mês_acesso”. Essa transformação visa possibilitar que a contagem das interações dos alunos seja relativa ao número de semestres cursados por eles no curso.

Ainda, foi necessário unificar os dados do atributo “COMPONENTE_LOG” e “ACAO_LOG”. Isso se deu pela mudança de nomenclatura que ocorreu nos nomes dos recursos do AVEA durante as alterações de versões do MOODLE. Do mesmo modo que a informação do nome da disciplina, o valor adotado para a nomenclatura foi aquela utilizada na versão mais recente do AVEA. No Quadro 13 são relacionados os ajustes efetuados nestes itens sendo a coluna “Recurso” representando o recurso do AVEA considerado. Já as colunas “COMPONENT_LOG (MOODLE)” e “ACAO_LOG (MOODLE)” representam as informações constantes nos logs do MOODLE. Por fim, as colunas “COMPONENT_LOG (MD)” e “ACAO_LOG (MD)” descrevem os respectivos dados transformados para utilização no processo de MD.

Quadro 13 — Unificação atributos AVEA MOODLE

Recurso	COMPONENT_LOG (MOODLE)	COMPONENT_LOG (MD)	ACAO_LOG (MOODLE)	ACAO_LOG (MD)
Fórum	Forum	Forum	Add discussion	Forum_add
			Add post	Forum_add
			View forum	Forum_view
			View forums	Forum_view
	Mod_forum	Forum	Created	Forum_add
			Viewed	Forum_view
Tarefas	Assign	Assign	Upload	Assign_submit
			View	Assign_view
	Mod_assign	Assign	Submitted	Assign_submit
			Viewed	Assign_view
Materiais	Resource	Resource	View	Resource_view
	Mod_resource	Resource	Viewed	Resource_view
Links	Url	Url	View	Url_view
	Mod_url	Url	Viewed	Url_view
Pasta	Folder	Folder	View	Folder_view
	Mod_folder	Folder	Viewed	Folder_view
Wiki	Wiki	Wiki	View	Wiki_view
	Mod_wiki	Wiki	Viewed	Wiki_view
Quiz	Quis	Quiz	View	Quiz_view
	Mod_Quiz	Quiz	Viewed	Quiz_view

Fonte: Autor.

Dentre os componentes selecionados, com o intuito de melhor caracterizar o problema abordado, os dados foram classificados em duas classes: “Aluno Regular” e “Evadido”. A primeira classe representa os alunos cuja permanência no curso é prevista. E, na classe “Evadido”, aqueles com tendência a evadirem. Não se optou pela criação de uma terceira classe que representasse os alunos formados, já que os alunos que se encontram nos semestres finais e são enquadrados como “Aluno Regular” consecutivamente estarão enquadrados nesta situação.

No Quadro 14 é possível visualizar os atributos finais selecionados para o processo de MD.

Quadro 14 — Atributos finais selecionados para MD – Primeiro experimento

Recurso	Atributo	Descrição
Fórum	forum_add	Número de postagens em fóruns.
	forum_view	Número de visualizações nos fóruns.
Tarefas	assign_submit	Número de tarefas submetidas.
	assign_view	Número de visualizações de tarefas.
Materiais	resource_view	Número de visualizações de materiais disponibilizados.
<i>Links</i>	url_view	Número de visualizações de <i>links</i> disponibilizados.
Pasta	folder_view	Número de visualizações a diretórios de materiais.
<i>Wiki</i>	wiki_view	Número de interações no <i>wiki</i> .
Questionário	quiz_view	Número de interações nos questionários.

Fonte: Autor.

Após a definição dos atributos para MD as informações dos alunos foram separadas em dois conjuntos. O primeiro deles, chamado de conjunto de treinamento, contém dados de todos os alunos que já possuem a sua situação definida junto ao curso, ou seja, são formados ou evadidos. Dessa forma, estes dados servirão de base para “treinar” os algoritmos selecionados, gerando assim, o modelo preditivo de dados. Para o curso de Letras – Espanhol, no segundo semestre de 2017 os dados de treinamento possuem informações de 351 alunos, sendo 98 deles formados - classe “Aluno Regular” - e 253 evadidos - classe “Evadido”. E o segundo, chamado de conjunto de testes, é composto por informações de alunos que não possuem sua situação definida no curso, ou seja, ainda se encontram matriculados. Estes alunos serão “testados” (avaliados) para se conhecer a sua situação futura com base nos dados de treinamento. O conjunto de testes no curso é composto por 220 alunos que permaneciam matriculados no fim do primeiro semestre de 2017. É importante ressaltar que os alunos que estão no conjunto de testes não estão presentes no conjunto de treinamento.

Alguns algoritmos de MD são sensíveis em relação ao formato de dados numéricos. Por esta razão a geração dos modelos foi realizada em dois formatos: dados não normalizados e normalizados. Os dados não normalizados contém os atributos obtidos diretamente da base de dados. Já os dados normalizados têm os atributos numéricos modificados para uma mesma faixa de valores, neste caso compreendidas entre 0 e 1. A normalização, do mesmo modo que o processo de MD, é realizada através da ferramenta WEKA.

Por fim, os dados foram manipulados para serem utilizados na ferramenta WEKA empregada na etapa a seguir.

7.2.4 Modelagem

Após a preparação dos dados é promovida a etapa de modelagem por meio da implementação dos algoritmos. A seleção dos algoritmos iniciou-se na RSL descrita no capítulo 3. No Gráfico 6 é possível visualizar os algoritmos mais utilizados nos trabalhos abordados pela RSL. Dessa forma, considerando os algoritmos mencionados e sua disponibilidade na ferramenta WEKA, foram selecionados para realização desta etapa os algoritmos: J48, *Naive Bayes*, *MultiLayer Perceptron*, *Random Forest*, OneR, SMO e IBk.

Para a estimativa de desempenho foi utilizado o método de validação cruzada, assumindo o valor de 10 pastas. Na Tabela 1 é possível visualizar os resultados dos testes realizados por meio das métricas de classificação Precisão, Revocação e Medida-F⁸.

Tabela 1 — Métricas de estimativa de desempenho dos algoritmos utilizados – Primeiro experimento

Algoritmo	Não Normalizado			Normalizado		
	Precisão	Revocação	Medida-F	Precisão	Revocação	Medida-F
J48	95,20%	95,20%	95,20%	95,20%	95,20%	95,20%
<i>Naive Bayes</i>	91,70%	91,50%	91,54%	91,70%	91,50%	91,54%
<i>MultiLayer Perceptron</i>	95,30%	95,20%	95,22%	95,30%	95,20%	95,22%
<i>Random Forest</i>	96,60%	96,60%	96,60%	96,90%	96,90%	96,90%
OneR	95,50%	95,40%	95,42%	95,50%	95,40%	95,42%
SMO	96,90%	96,90%	96,90%	96,90%	96,90%	96,90%
IBk	95,30%	95,20%	95,22%	95,30%	95,20%	95,22%

Fonte: Autor.

Por meio da Tabela 1 é possível perceber que os testes efetuados na base de treinamento apresentam resultados semelhantes entre suas métricas de avaliação, bem como não apresentam diferenças significativas entre as métricas dos dados normalizados e não normalizados. Além disso, os resultados foram considerados satisfatórios para todos os algoritmos selecionados em que houve uma Precisão média de 95,21% para os dados não normalizados e 95,26% para os

⁸ Para maiores informações sobre as métricas de avaliação consulte a seção 4.2.3.6.1.

dados normalizados. Desse modo, conforme destacado na Tabela 1, o maior percentual de acertos entre os dados não normalizados foi atingido pelo algoritmo SMO com 96,90% e, entre os dados normalizados, foi atingido pelos algoritmos *Random Forest* e SMO com o mesmo valor.

Após a geração dos modelos foi procedida com a previsão por meio da aplicação do conjunto de testes. A aplicação deste procedimento será descrita na etapa de avaliação.

7.2.5 Avaliação

Após a modelagem inicia-se a etapa de avaliação. Na etapa de avaliação o conjunto de teste é submetido aos modelos gerados e, logo em seguida, são procedidas a análise dos resultados previstos pelos algoritmos após o processamento de matrícula ser realizado pela Instituição.

Para o conjunto de testes foram considerados todos os alunos aptos a realizarem matrícula no segundo semestre de 2017. O número de alunos que se enquadraram neste critério totalizou 220. Após o prazo de realização de matrículas foram identificadas 40 desistências no curso de Letras – Espanhol. A relação dos algoritmos utilizados, evasões previstas e acertos (Revocação) pode ser visualizada na Tabela 2.

Tabela 2 — Algoritmos utilizados *versus* evasões previstas *versus* acertos – Primeiro experimento

Algoritmo	Não Normalizado		Normalizado	
	Alunos previstos	Revocação	Alunos previstos	Revocação
J48	177	37 (92,50%)	30	15 (37,50%)
<i>Naive Bayes</i>	114	32 (80,00%)	92	26 (65,00%)
<i>MultiLayer Perceptron</i>	142	34 (85,00%)	20	7 (17,50%)
<i>Random Forest</i>	175	38 (95,00%)	162	35 (87,50%)
OneR	177	37 (92,50%)	30	15 (37,50%)
SMO	177	38 (95,00%)	9	6 (15,00%)
IBk	172	37 (92,50%)	104	23 (57,50%)
Todos ¹	191	39 (97,50%)	170	35 (87,50%)

Fonte: Autor.

¹ Considera o universo dos diferentes alunos previstos em todos os algoritmos analisados.

A Tabela 2 detalha os dados da quantidade de alunos apontados pelos algoritmos como possíveis desistentes, na coluna Alunos Previstos e, na coluna Revocação, a quantidade de acertos da previsão em relação aos alunos que não se matricularam no segundo semestre de 2017, ou seja, de fato evadiram no período mencionado. Da mesma forma da modelagem, as avaliações também foram realizadas com dados não normalizados e normalizados. Além disso, para permitir uma análise global das previsões, foi relacionado na tabela como um algoritmo o elemento “Todos”, contendo a união dos alunos previstos em todos os algoritmos.

O maior índice foi encontrado nos dados não normalizados na união das previsões (Todos) com 97,50%. Já para os dados normalizados, o maior percentual foi obtido pela união das previsões (Todos) e o algoritmo *Random Forest* com 87,50%. No entanto, considerando os demais resultados, constata-se um elevado percentual de acertos nas previsões realizadas nos dados não normalizados. O algoritmo com o menor índice de acertos neste formato, o *Naive Bayes*, apresentou uma taxa de 80,00%. Em relação aos dados normalizados tanto o número de alunos previstos quanto de acertos é inferior. Mesmo assim, algoritmos como *Naive Bayes*, *Random Forest*, IBk e a união (Todos) apresentaram boas taxas de acertos.

Para avaliação da qualidade dos resultados das previsões foram utilizadas as métricas de Acurácia, Precisão e Medida-F. A Tabela 3 apresenta os valores alcançados para as métricas mencionadas em cada um dos algoritmos para os dados não normalizados e normalizados.

Tabela 3 — Métricas de avaliação da previsão – Primeiro experimento

Algoritmo	Não Normalizado			Normalizado		
	Acurácia	Precisão	Medida-F	Acurácia	Precisão	Medida-F
J48	35,00%	20,90%	54,89%	81,89%	50,00%	39,47%
<i>Naive Bayes</i>	59,09%	28,07%	58,39%	63,64%	28,26%	51,59%
<i>MultiLayer Perceptron</i>	48,18%	23,94%	56,29%	79,09%	35,00%	19,44%
<i>Random Forest</i>	36,81%	21,71%	56,71%	40,00%	21,60%	54,34%
OneR	35,00%	20,90%	54,89%	81,82%	50,00%	39,47%
SMO	35,91%	21,47%	56,38%	83,18%	66,67%	17,75%
IBk	37,27%	21,51%	55,72%	55,45%	22,11%	43,56%
Todos ¹	68,18%	20,42%	55,56%	95,90%	20,59%	53,03%

Fonte: Autor.

¹ Considera o universo de alunos previstos em todos os algoritmos analisados.

Na Tabela 3 são destacados os maiores índices encontrados para cada métrica. A métrica de Acurácia tem o índice mais elevado na união das previsões dos algoritmos (Todos),

alcançando 95,90% no formato de dados normalizados. Também para os dados normalizados a Precisão atingiu 66,67% no algoritmo SMO. Por fim, nos dados não normalizados, o algoritmo *Naive Bayes* atingiu o maior índice de Medida-F totalizando 58,39%.

Em relação à taxa de acertos dos algoritmos, a Acurácia, os dados normalizados apresentaram melhores resultados. Isto pode ser visto pela Acurácia média dos algoritmos analisados em que os dados normalizados atingiram a 72,62% e os dados não normalizados alcançaram 44,43%. A inferioridade dos resultados apresentados nos dados não normalizados justifica-se pela grande parte dos alunos contidos no conjunto de testes terem sido indicados como “Evadido”, conforme pode ser visto na Tabela 2.

No que diz respeito à taxa de acertos dos algoritmos em relação à previsão realizada, a Precisão, os melhores índices também se encontram nos dados normalizados. A precisão média entre os algoritmos com dados normalizados chegou a 36,78% e para os dados não normalizados atingiu 22,36%. Da mesma forma que a Acurácia, os resultados dos dados não normalizados foram inferiores já que a quantidade de alunos previstos como “Evadido” foram superiores em relação aos dados normalizados, conforme pode ser visto na Tabela 2.

Já a Medida-F, que combina as métricas de Precisão e Revocação, alcançou melhor resultado para os dados não normalizados em que a média alcançada pelos algoritmos chegou a 56,10%. Quanto aos dados normalizados a média obtida foi de 39,83%. Um dos objetivos do trabalho consiste na previsão de alunos que tendem a evadir em um determinado semestre, mas, no entanto, ainda que ele não venha a evadir no semestre avaliado, ele pode ser um futuro desistente. Por esta razão, por mais que seja importante que o algoritmo seja preciso nas suas previsões também é interessante ter a indicação de alunos que possam vir a se enquadrar nesta situação. Desse modo, neste trabalho foi utilizado uma fórmula de cálculo da Medida-F que prioriza a métrica de Revocação, ou seja, dá ênfase para a quantidade de acertos em prol da precisão. Sendo assim, apesar da Precisão ter apresentado melhor resultado para os dados normalizados, ainda assim a Medida-F apresentou melhores taxas para os dados não normalizados já que a Revocação neste formato atingiu melhores índices.

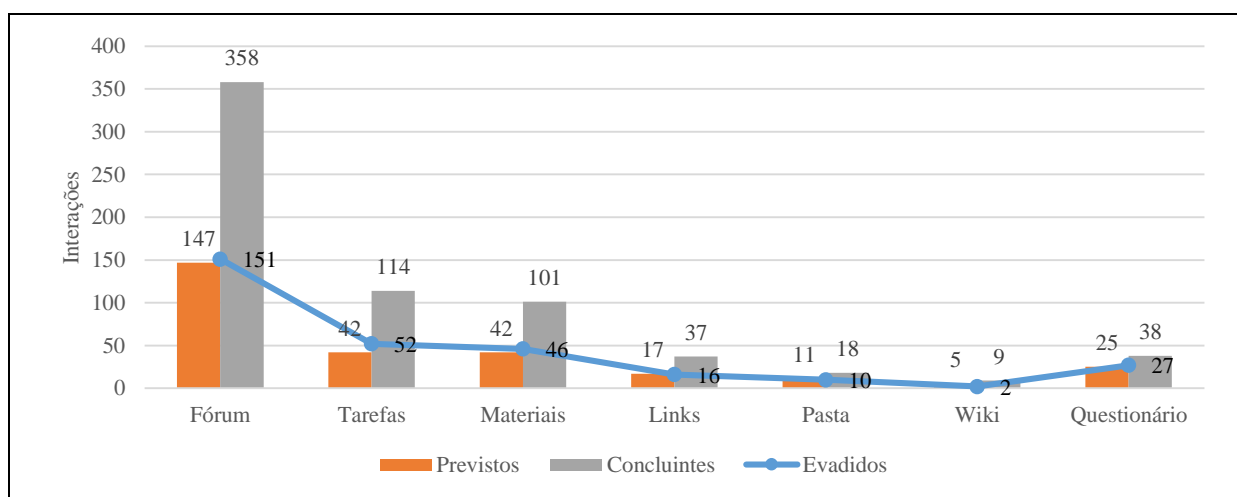
Nesse contexto, com base nos índices de Medida-F, é possível afirmar que os algoritmos utilizados, em especial com os dados não normalizados, são apropriados para a previsão de evasão de alunos de graduação na modalidade EaD. No entanto, alguns algoritmos utilizando o formato normalizado, como *Naive Bayes* e *Random Forest*, também apresentaram resultados satisfatórios.

Outro fator que corrobora com a viabilidade da previsão de evasão pela MD é a relação de alunos evadidos que foram indicados pelos algoritmos. Considerando os alunos previstos

nas simulações, todos os 40 alunos que evadiram no segundo semestre de 2017 foram apontados em pelo menos uma delas.

Por fim, no Gráfico 8 é possível visualizar a média de utilização dos recursos do AVEA que foram empregados no processo de MD. Os dados estão distribuídos em três séries: alunos concluintes, abrangendo os alunos que tiveram sua situação final no processo de mineração definida como “Aluno Regular”; previstos, contendo os alunos que foram indicados no final do processo como “Evadido”; e, evadidos, sendo composto pelos alunos que vieram a evadir de fato após o período de matrícula.

Gráfico 8 — Média de utilização dos recursos do AVEA



Fonte: Autor.

Para o cálculo da média dos alunos previstos e concluintes foram utilizadas as previsões oriundas da simulação que apresentou o maior índice de Medida-F, isto é, o algoritmo *Naive Bayes* com dados não normalizados.

O Gráfico 8 apresentado permite uma análise por dois ângulos diferentes. No primeiro é possível constatar a proximidade dos valores de todos os atributos considerados entre os alunos que foram previstos e aqueles que evadiram de fato. Sendo este mais um fator que ratifica a viabilidade da utilização das técnicas de MD para previsão de evasão, pois permite a identificação de todos os alunos que se enquadrem num perfil de aluno evadido. E, no segundo viés, permite identificar uma grande lacuna na utilização do AVEA entre os alunos previstos como concluintes e os previstos como evadidos. Isso permite concluir que os alunos que fazem mais uso do ambiente virtual tendem a concluir o curso, indo ao encontro com as constatações obtidas no trabalho de Abbad, Zerbini e Souza (2010) relacionadas na seção 2.2.

7.2.6 Disponibilização

Após a avaliação da viabilidade de utilização dos algoritmos para previsão de evasão, os dados relacionados devem ser disponibilizados de forma a proporcionar a análise para a tomada de decisões. No entanto, no primeiro experimento, o objetivo principal era a demonstração da aplicabilidade da MDE para o fim estabelecido. Dessa forma, a etapa de disponibilização deste experimento foi efetuada após, em conjunto com o segundo, por meio do desenvolvimento de um aplicativo para dispositivo móvel a ser detalhado no capítulo 8.

7.3 SEGUNDO EXPERIMENTO

Após a qualificação do projeto foi efetuado o segundo experimento. Nele foram considerados os dados dos alunos dos cursos de Letras – Espanhol, avaliados anteriormente, Pedagogia e Educação Especial. Além de abranger um maior número de cursos, a previsão foi realizada para o primeiro semestre de 2018.

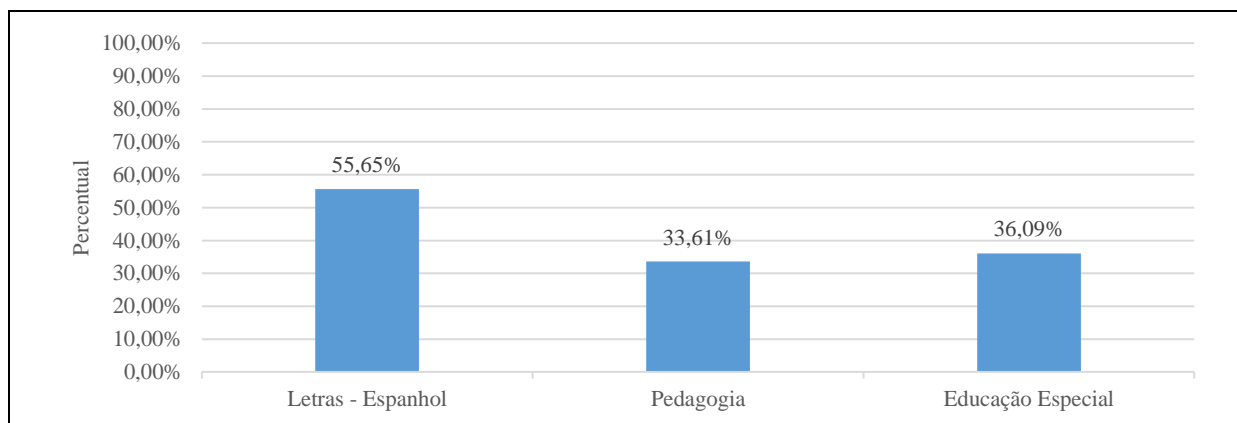
Nas subseções que seguem serão detalhadas as etapas da metodologia de KDD adotada para a realização desta etapa.

7.3.1 Entendimento da evasão

Para analisar a evasão dos cursos selecionados foram calculadas as taxas de evasão total e evasão anual média para os cursos abrangidos. No Gráfico 9 é possível visualizar os valores correspondentes à evasão geral de cada curso.

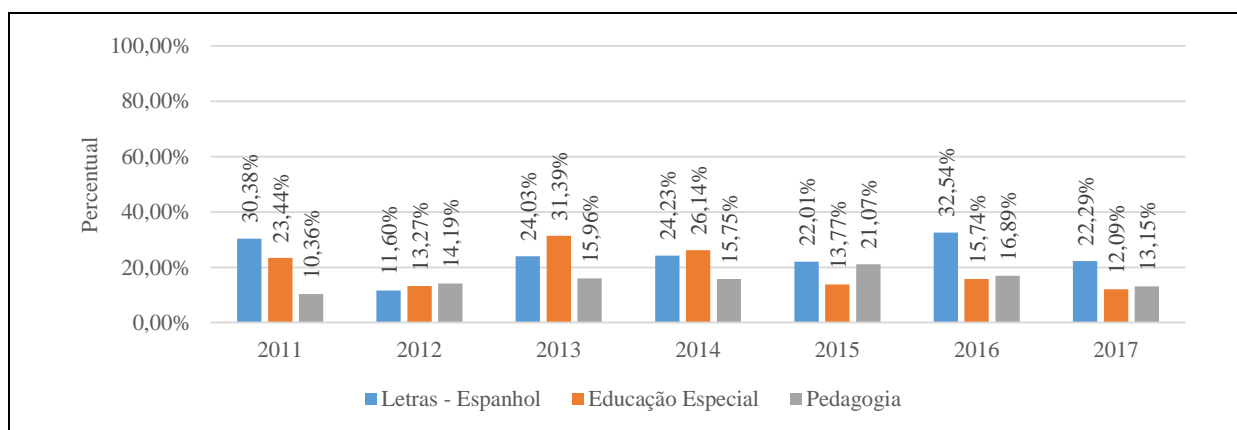
Como base no Gráfico 9 percebe-se que o curso de Letras – Espanhol lidera o índice de evasão total com 55,65%. Em segundo lugar aparece o curso de Educação Especial com 36,09%. Por fim, em terceiro lugar, surge o curso de Pedagogia com 33,61%. No Gráfico 10 são demonstrados os índices de evasão anual média para os cursos selecionados entre os anos de 2011 e 2017.

Gráfico 9 — Evasão geral – Segundo experimento



Fonte: Autor.

Gráfico 10 — Evasão anual média – Segundo experimento



Fonte: Autor.

O período selecionado para demonstração do gráfico inicia-se em 2011, pois o ano de 2010 foi o período inicial considerado no processo de MD, não sendo possível calcular o seu índice. Por fim, termina no ano de 2017, já que 2018 estava em andamento no momento da elaboração do experimento.

A partir das médias dos valores da evasão anual é possível detectar a mesma tendência apresentada no indicador de evasão geral. O curso de Letras – Espanhol segue na liderança, apresentando uma média anual de 23,87%. Em segundo lugar, segue o curso de Educação Especial com 19,40%. E, no fim, aparece o curso de pedagogia com 15,34%.

No curso de Letras – Espanhol, por exemplo, com Evasão Geral de 55,65%, mais da metade dos alunos que ingressaram no curso não obtiveram sua conclusão. Nos demais cursos,

apesar das taxas inferiores, o cenário também é preocupante já que a cada dez alunos que ingressam no curso de Educação Especial, quatro tendem a não finalizá-lo e, na Pedagogia, esse índice se mantém em três.

Consolidando os dados dos três cursos selecionados, a média de evasão geral alcança 41,78%, ou seja, em números gerais, a cada dez alunos ingressantes quatro não obtém a sua conclusão o que justifica, assim, a relevância do trabalho proposto.

7.3.2 Entendimento dos dados de alunos

A organização dos dados para a realização deste experimento segue a mesma exposta no anterior, conforme detalhado na seção 7.2.2. No entanto, por considerar um volume maior de informações e visando a melhoria dos resultados da primeira etapa, cabem destacar algumas modificações.

O primeiro diferencial neste experimento está na quantidade e diversidade dos dados disponíveis para mineração. Nos dados disponibilizados existem 128 diferentes ações realizadas em 60 módulos gerados por 3989 alunos, perfazendo um total de 30.561.288 interações. Além disso, os dados extraídos estão compreendidos entre os anos de 2010 e 2017.

Outro ponto a considerar são os dados utilizados no segundo experimento. Além dos dados dos logs do AVEA MOODLE e acadêmicos, já analisados no primeiro, foram incluídas, também, as informações relacionadas ao número de disciplinas aprovadas e reprovadas de cada estudante. De acordo com o exposto nos trabalhos de Shannon e Bylsma (2006), Martins, Gebran e Terçariol (2014) e Bittencourt e Mercado (2014)⁹ estes são potenciais fatores causadores da evasão.

Além disso, as respectivas autorizações para uso dos dados dos cursos de Educação Especial e Pedagogia encontram-se nos Anexos C e D.

7.3.3 Preparação dos dados

O foco inicial da preparação dos dados foi na remoção das inconsistências e padronização dos dados, conforme relatado no primeiro experimento (seção 7.2.3). Do novo conjunto de dados analisados foram encontrados 704 registros de 18 alunos que pertenciam a outros cursos, dessa forma, eles foram desconsiderados do processo de mineração.

⁹ Maiores detalhes desses trabalhos podem ser visualizados na seção 2.2.

Ainda na preparação, foram analisados atributos disponíveis no Sistema de Informação institucional que permitissem melhorar os resultados apresentados no primeiro experimento. No final desta análise foram encontrados novos atributos relacionados à quantidade de cursos e disciplinas realizados pelos alunos. O Quadro 15 demonstra como ficou a seleção final dos atributos para MD.

Quadro 15 — Atributos finais selecionados para MD – Segundo experimento

Recurso/Informação	Atributo	Descrição
Fórum	forum_add	Número de postagens em fóruns.
	forum_view	Número de visualizações nos fóruns.
Tarefas	assign_submit	Número de tarefas submetidas.
	assign_view	Número de visualizações de tarefas.
Materiais	resource_view	Número de visualizações de materiais disponibilizados.
<i>Links</i>	url_view	Número de visualizações de <i>links</i> disponibilizados.
Pasta	folder_view	Número de visualizações a diretórios de materiais.
<i>Wiki</i>	wiki_view	Número de interações no <i>wiki</i> .
Questionário	quiz_view	Número de interações nos questionários.
Cursos externos	num_cursos_externos	Número de cursos de graduação realizado em outras instituições.
Cursos concluídos	num_cursos_concluidos	Número de cursos de graduação concluído na instituição.
Cursos em andamento	num_cursos_andamento	Número de cursos de graduação ou pós-graduação em realização na instituição.
Cursos evadidos	num_cursos_evadidos	Número de cursos evadidos na instituição.
Disc. aprovadas 1º sem.	num_aprovadas_1	Número de disciplinas aprovadas no primeiro semestre do curso.
Disc. reprovadas 1º sem.	num_reprovadas_1	Número de disciplinas reprovadas no primeiro semestre do curso.
Disc. aprovadas semestre corrente	num_aprovadas_ultimo	Número de disciplinas aprovadas no semestre corrente.
Disc. reprovadas semestre corrente	num_reprovadas_ultimo	Número de disciplinas reprovadas no semestre corrente.

Além dos atributos provenientes do primeiro experimento foram consideradas as informações relacionadas à realização de outros cursos de graduação tanto na própria instituição quanto em outra. Conforme as causas da evasão expostas por Bittencourt e Mercado (2014), estes atributos foram considerados por acreditar-se que o fato de o aluno já ter concluído ou evadido de um curso pode prever o comportamento dele no curso em que está sendo analisado pelos algoritmos de MD.

Por fim, também foi considerado o número de disciplinas que o aluno aprovou/reprovou no primeiro e no último semestre cursado por ele. Estas quantidades, de acordo com Shannon e Bylisma (2006), Bittencourt e Mercado (2014) e Martins, Gebran e Terçariol (2014), podem indicar o comportamento do aluno durante o decorrer do curso.

Após a definição dos atributos para MD, as informações dos alunos foram separadas nos conjuntos de treinamento e teste para cada curso, de acordo com o exposto na Tabela 4.

Tabela 4 — Conjuntos de dados

Curso	Treinamento	Teste
Letras – Espanhol	421	150
Pedagogia	838	517
Educação Especial	284	165
Total	1543	832

Fonte: Autor.

Como pode ser visto na Tabela 4, o conjunto de treinamento do curso de Letras – Espanhol é composto por 421 alunos, sendo 128 deles formados – classe “Aluno Regular” e 293 evadidos – classe “Evadido”. Já o curso de Pedagogia é composto por 838 alunos, sendo 378 deles formados e 460 evadidos. Por fim, o conjunto de treinamento do curso de Educação Especial é composto de 284 alunos, sendo 118 formados e 166 evadidos.

7.3.4 Modelagem

Os algoritmos selecionados no experimento foram os mesmos adotados no seu precursor. O diferencial na modelagem, como nas demais etapas deste experimento, é o número de cursos considerados, bem como o período de abrangência dos dados.

Seguindo os mesmos critérios empregados no experimento anterior, a estimativa de desempenho foi realizada considerando dados não normalizados e normalizados. Além disso,

foi utilizado o método de avaliação cruzada assumindo o valor de 10 pastas. Na Tabela 5 é possível visualizar os resultados das métricas de classificação aplicados na base de treinamento dos cursos.

Tabela 5 — Métricas de estimativa de desempenho dos algoritmos utilizados – Segundo experimento

Curso	Algoritmo	Não Normalizado			Normalizado		
		Precisão	Revocação	Medida-F	Precisão	Revocação	Medida-F
Letras – Espanhol	J48	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%
	<i>Naive Bayes</i>	98,80%	98,80%	98,80%	98,80%	98,80%	98,80%
	<i>MultiLayer Perceptron</i>	99,10%	99,00%	99,02%	99,10%	99,00%	99,02%
	<i>Random Forest</i>	99,30%	99,30%	99,30%	99,10%	99,00%	99,02%
	OneR	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%
	SMO	99,10%	99,00%	99,02%	99,10%	99,00%	99,02%
	IBk	98,20%	98,10%	98,12%	98,20%	98,10%	98,12%
Pedagogia	J48	99,60%	99,60%	99,60%	99,60%	99,60%	99,60%
	<i>Naive Bayes</i>	98,70%	98,70%	98,70%	98,70%	98,70%	98,70%
	<i>MultiLayer Perceptron</i>	99,50%	99,50%	99,50%	99,50%	99,50%	99,50%
	<i>Random Forest</i>	99,60%	99,60%	99,60%	100,00%	100,00%	100,00%
	OneR	99,60%	99,60%	99,60%	99,60%	99,60%	99,60%
	SMO	99,50%	99,50%	99,50%	99,50%	99,50%	99,50%
	IBk	98,60%	98,60%	98,60%	98,60%	98,60%	98,60%
Educação Especial	J48	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%
	<i>Naive Bayes</i>	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%
	<i>MultiLayer Perceptron</i>	99,00%	98,90%	98,92%	99,00%	98,90%	98,92%
	<i>Random Forest</i>	100,00%	100,00%	100,00%	99,70%	99,60%	99,62%
	OneR	99,30%	99,30%	99,30%	99,70%	99,60%	99,62%
	SMO	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%
	IBk	97,90%	97,90%	97,90%	97,90%	97,90%	97,90%
Todos ¹	J48	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%
	<i>Naive Bayes</i>	97,80%	97,80%	97,80%	97,80%	97,80%	97,80%
	<i>MultiLayer Perceptron</i>	98,80%	98,80%	98,80%	99,00%	99,00%	99,00%
	<i>Random Forest</i>	99,50%	99,50%	99,50%	99,50%	99,50%	99,50%
	OneR	98,50%	98,40%	98,42%	98,50%	98,40%	98,42%
	SMO	99,10%	99,10%	99,10%	99,10%	99,10%	99,10%
	IBk	98,50%	98,50%	98,50%	98,50%	98,50%	98,50%

Fonte: Autor.

¹ Engloba no mesmo conjunto de dados todos os cursos analisados.

Para realizar uma análise mais ampla foi definido um curso chamado “Todos”, contendo no conjunto de treinamento os alunos de todos os cursos analisados. Por meio da análise da Tabela 5 é possível observar que os testes efetuados nas bases de treinamento apresentam resultados semelhantes entre suas métricas de avaliação sem variações significativas entre os algoritmos, formato de dados e cursos analisados. Além disso, os resultados foram considerados satisfatórios para todas as simulações já que nenhuma métrica apresentou resultado abaixo de 97,00%.

Os resultados apresentados nesta etapa demonstram que os algoritmos analisados estão aptos a serem utilizados na fase seguinte, a avaliação. Dessa forma, a aplicação deste procedimento será descrita na subseção a seguir.

7.3.5 Avaliação

Após a modelagem inicia-se a etapa de avaliação. Nos dados a serem avaliados, o conjunto de teste, são considerados todos os alunos aptos a realizarem matrícula no primeiro semestre de 2018. Na Tabela 6 é possível visualizar a relação desses alunos juntamente com o número de evasões ocorridas para cada curso.

Tabela 6 — Relação de alunos aptos à matrícula *versus* evadidos – Segundo experimento

Curso	Aptos a matrícula	Evadidos
Letras – Espanhol	150	22
Pedagogia	517	16
Educação Especial	165	18
Total	832	56

Fonte: Autor.

A relação dos cursos avaliados, algoritmos utilizados, evasões previstas e acertos (Revocação) podem ser visualizadas na Tabela 7.

Tabela 7 — Algoritmos utilizados *versus* evasões previstas *versus* acertos – Segundo experimento

Curso	Algoritmo	Não Normalizado		Normalizado	
		Alunos previstos	Revocação	Alunos previstos	Revocação
Letras – Espanhol	J48	149	22 (100,00%)	147	21 (95,45%)
	<i>Naive Bayes</i>	129	22 (100,00%)	78	15 (68,18%)
	<i>MultiLayer Perceptron</i>	132	21 (95,45%)	1	1 (4,54%)
	<i>Random Forest</i>	137	22 (100,00%)	132	22 (100,00%)
	OneR	140	22 (100,00%)	7	4 (18,18%)
	SMO	126	22 (100,00%)	40	15 (68,18%)
	IBk	128	22 (100,00%)	69	22 (100,00%)
	Todos ¹	149	22 (100,00%)	150	22 (100,00%)
Pedagogia	J48	486	13 (81,25%)	41	6 (37,50%)
	<i>Naive Bayes</i>	454	15 (93,75%)	264	15 (93,75%)
	<i>MultiLayer Perceptron</i>	490	15 (93,75%)	44	4 (25,00%)
	<i>Random Forest</i>	496	15 (93,75%)	483	13 (81,25%)
	OneR	501	15 (93,75%)	41	6 (37,50%)
	SMO	479	14 (87,50%)	31	3 (18,75%)
	IBk	460	14 (87,50%)	160	15 (93,75%)
	Todos ¹	504	15 (93,75%)	483	16 (100,00%)
Educação Especial	J48	163	18 (100,00%)	2	1 (5,56%)
	<i>Naive Bayes</i>	164	18 (100,00%)	38	12 (66,67%)
	<i>MultiLayer Perceptron</i>	133	17 (94,44%)	3	1 (5,56%)
	<i>Random Forest</i>	162	18 (100,00%)	163	18 (100,00%)
	OneR	163	18 (100,00%)	2	1 (5,56%)
	SMO	160	18 (100,00%)	12	10 (55,56%)
	IBk	151	18 (100,00%)	37	14 (77,78%)
	Todos ¹	164	18 (100,00%)	163	18 (100,00%)
Todos ²	J48	818	55 (98,21%)	804	50 (89,29%)
	<i>Naive Bayes</i>	563	53 (94,64%)	314	38 (67,86%)
	<i>MultiLayer Perceptron</i>	680	52 (92,86%)	22	4 (7,14%)
	<i>Random Forest</i>	803	55 (98,21%)	794	55 (98,21%)
	OneR	783	54 (96,42%)	50	11 (19,64%)
	SMO	648	52 (92,86%)	68	9 (16,07%)
	IBk	708	52 (92,86%)	239	53 (94,64%)
	Todos ¹	820	56 (100,00%)	820	56 (100,00%)

Fonte: Autor.

¹ Considera o universo de alunos distintos previstos em todos os algoritmos analisados.

² Engloba no mesmo conjunto de dados todos os cursos analisados.

A Tabela 7 detalha os dados da quantidade de alunos apontados pelos algoritmos como possíveis desistentes, na coluna Alunos Previstos, e na coluna Revocação a quantidade de acertos da previsão em relação aos alunos que não efetuaram a matrícula no primeiro semestre de 2018. Além disso, as informações constantes nela estão divididas em duas partes. Na primeira é representado o conjunto de dados utilizados para o processo de mineração sendo sua descrição definida na coluna Curso. Já a segunda divisão é relacionada ao algoritmo de MD utilizado descrito na coluna Algoritmo.

De forma a ampliar a análise foram realizadas simulações com a utilização dos dados exclusivos dos cursos – representados pelo nome de cada curso – como também uma única simulação contendo os conjuntos de dados de todos os cursos – representado na tabela pela nomenclatura “Todos”. Além disso, para cada conjunto citado foi analisado o resultado da união das previsões – representado em cada conjunto de dados pela denominação “Todos”.

Considerando o percentual de acertos é possível verificar que a simulação nos dados não normalizados apresentaram resultados superiores em comparação com os dados normalizados. Nos dados não normalizados, o menor percentual de acertos obtidos foi de 81,25%. Já em relação aos dados normalizados poucos algoritmos obtiveram acertos superiores a este valor, mas, no entanto, é possível encontrar a ocorrência de algumas simulações com dados normalizados que obtiveram até 100,00% de acertos. Além disso, realizando uma comparação entre as simulações realizadas com dados de um curso e as efetuadas com dados de todos os cursos em uma mesma simulação, não apresentam diferenças significativas entre os seus resultados tanto de uma simulação quanto de outra.

Para avaliação da qualidade dos resultados foram utilizadas as métricas de Acurácia, Precisão e Medida-F. A Tabela 8 apresenta os valores alcançados para as métricas mencionadas em cada um dos algoritmos para os dados não normalizados e normalizados.

Tabela 8 — Métricas de avaliação da previsão – Segundo experimento

Curso	Algoritmo	Não Normalizado			Normalizado		
		Acurácia	Precisão	Medida-F	Acurácia	Precisão	Medida-F
Letras-Espanhol	J48	15,33%	14,76%	46,40%	15,33%	14,28%	44,67%
	<i>Naive Bayes</i>	28,67%	17,05%	50,68%	53,33%	19,23%	45,18%
	<i>MultiLayer Perceptron</i>	25,33%	15,91%	47,73%	86,00%	100,00%	5,61%
	<i>Random Forest</i>	23,33%	16,06%	48,89%	26,67%	16,67%	50,00%
	OneR	21,33%	15,71%	48,24%	86,00%	57,14%	21,05%
	SMO	30,67%	17,46%	51,40%	78,67%	37,50%	58,59%
	IBk	29,33%	17,19%	50,93%	68,67%	31,88%	70,06%
	Todos ¹	34,67%	14,76%	46,40%	100,00%	14,67%	46,22%
Pedagogia	J48	7,93%	2,67%	11,80%	91,27%	14,63%	28,57%
	<i>Naive Bayes</i>	14,89%	3,30%	14,46%	51,54%	5,68%	22,86%
	<i>MultiLayer Perceptron</i>	7,93%	3,06%	13,53%	89,94%	9,09%	18,52%
	<i>Random Forest</i>	6,77%	3,02%	13,38%	8,51%	2,69%	11,88%
	OneR	5,80%	2,99%	13,26%	91,27%	14,63%	28,57%
	SMO	9,67%	2,92%	12,88%	92,07%	9,67%	15,78%
	IBk	13,35%	3,04%	13,34%	71,76%	9,37%	33,47%
	Todos ¹	21,28%	2,97%	13,18%	97,68%	3,31%	14,61%
Educação Especial	J48	12,12%	11,04%	38,29%	89,09%	50,00%	6,76%
	<i>Naive Bayes</i>	11,51%	10,98%	38,14%	80,61%	31,58%	54,55%
	<i>MultiLayer Perceptron</i>	29,09%	12,78%	41,46%	88,48%	33,33%	6,67%
	<i>Random Forest</i>	12,72%	11,11%	38,46%	12,12%	11,04%	38,29%
	OneR	12,12%	11,04%	38,29%	89,09%	50,00%	6,77%
	SMO	13,94%	11,25%	38,79%	93,94%	83,33%	59,53%
	IBk	19,39%	11,92%	40,36%	83,64%	37,84%	64,22%
	Todos ¹	35,15%	10,97%	38,12%	98,18%	9,81%	35,27%
Todos ²	J48	8,17%	6,72%	26,38%	8,65%	6,22%	24,32%
	<i>Naive Bayes</i>	38,34%	9,41%	33,66%	64,66%	12,10%	35,31%
	<i>MultiLayer Perceptron</i>	24,04%	7,65%	28,77%	91,59%	18,18%	8,13%
	<i>Random Forest</i>	9,37%	6,85%	26,78%	11,05%	6,93%	27,02%
	OneR	12,14%	6,90%	26,82%	89,90%	22,00%	20,07%
	SMO	27,88%	8,02%	29,80%	87,26%	13,23%	15,41%
	IBk	20,67%	7,34%	27,88%	77,28%	22,18%	57,24%
	Todos ¹	46,15%	6,83%	26,82%	99,16%	6,83%	26,82%

Fonte: Autor.

¹ Considera o universo de alunos previstos em todos os algoritmos analisados.

² Engloba no mesmo conjunto de dados todos os cursos analisados.

Em relação aos resultados obtidos no curso de Letras – Espanhol verifica-se que os melhores índices foram apresentados no formato de dados normalizados. Para a Acurácia a média geral dos algoritmos chegou a 26,08% nos dados não normalizados e, para os dados normalizados, totalizou 64,33% sendo, para esta métrica, o melhor resultado alcançado na união dos algoritmos (Todos) nos dados normalizados com 100,00%. A média da Precisão para os dados não normalizados foi de 16,11% e, para os dados normalizados, 36,42% em que o melhor resultado foi alcançado pelo algoritmo *MultiLayer Perceptron* com 100% nos dados normalizados. Por fim, a média geral da Medida-F teve resultado superior para os dados não normalizados, atingindo 48,83% em que os dados normalizados chegaram a 42,67%. Em contrapartida, o algoritmo que atingiu melhor resultado para esta métrica foi o IBk com 70,06% nos dados normalizados.

Analisando os resultados obtidos em cada métrica é possível verificar que o mesmo comportamento se aplica na mineração dos dados do curso de Pedagogia, podendo-se observar que os melhores resultados foram apresentados no formato normalizado. A Acurácia média obtida para os dados não normalizados atingiu 10,95% e para os dados normalizados chegou a 74,25% sendo que a união de todos os algoritmos (Todos) alcançou o melhor resultado com 97,68% para os dados normalizados. A Precisão média para os dados não normalizados foi de 3,00% já os dados normalizados alcançaram 8,63% sendo os algoritmos J48 e OneR neste formato que atingiram o melhor resultado com 14,63%. Por fim, a Medida-F alcançou a média de 13,23% para os dados não normalizados e 21,78% para os dados normalizados, atingindo o algoritmo IBk neste formato o melhor resultado com 33,47%.

Quanto à aplicação da mineração nos dados do curso de Educação Especial os dados normalizados mantiveram resultados superiores. A Acurácia média das simulações sobre os dados não normalizados foi de 18,25% e para os dados normalizados atingiu 79,39% em que a união dos algoritmos (Todos) atingiu o melhor índice com 98,18%. Da mesma forma acontece com a métrica de Precisão em que para os dados não normalizados a média atingida foi de 11,39% e nos dados normalizados chegou a 38,37% sendo que o algoritmo SMO atingiu o melhor resultado com 83,33% para os dados normalizados. E a métrica de Medida-F no formato não normalizado teve média de 38,99% já para os dados normalizados a média é inferior, ou seja, sendo 33,85%, no entanto, o algoritmo que apresentou melhor resultado foi o IBk com 64,22% com dados normalizados.

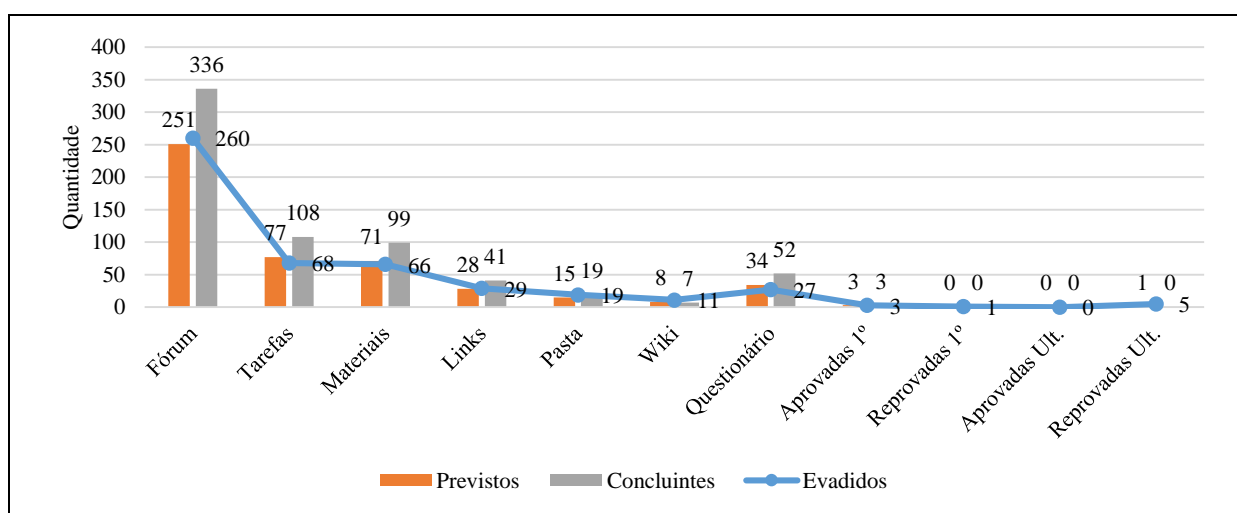
A fim de analisar a aplicação das técnicas de MD no conjunto de dados de todos os cursos analisados, foi criada uma simulação denominada de “Todos”, na Tabela 8. Nesta simulação as médias obtidas pela métrica de Acurácia para os dados não normalizados e

normalizados foram respectivamente 23,34% e 66,19% sendo o melhor resultado obtido pela união dos algoritmos (Todos) com 99,16% para os dados normalizados. Já a Precisão teve como médias 7,46% e 13,46% para os dados não normalizados e normalizados, respectivamente. O algoritmo com melhor resultado foi o IBk com 22,18% para os dados normalizados. Por fim, a Medida-F apresentou média de 28,36% para os dados não normalizados e 26,79% para os dados normalizados. Para esta métrica o algoritmo com melhor resultado foi o IBk nos dados normalizados com 57,24%.

Analisando os resultados obtidos por meio das métricas de avaliação dos algoritmos é possível afirmar que a utilização das técnicas de MD são factíveis para detectar a evasão de alunos de cursos de graduação EaD. Os resultados foram satisfatórios tanto com a utilização do conjunto de dados restritos a um determinado curso, quanto com o emprego de dados de diversos cursos em um mesmo conjunto. Além disso, o formato de dados que apresentou melhores taxas para a maior parte das simulações foram os dados normalizados. Essa constatação corrobora com a importância de realizar a simulação para ambos os formatos e confirma a sensibilidade de alguns algoritmos sobre este aspecto.

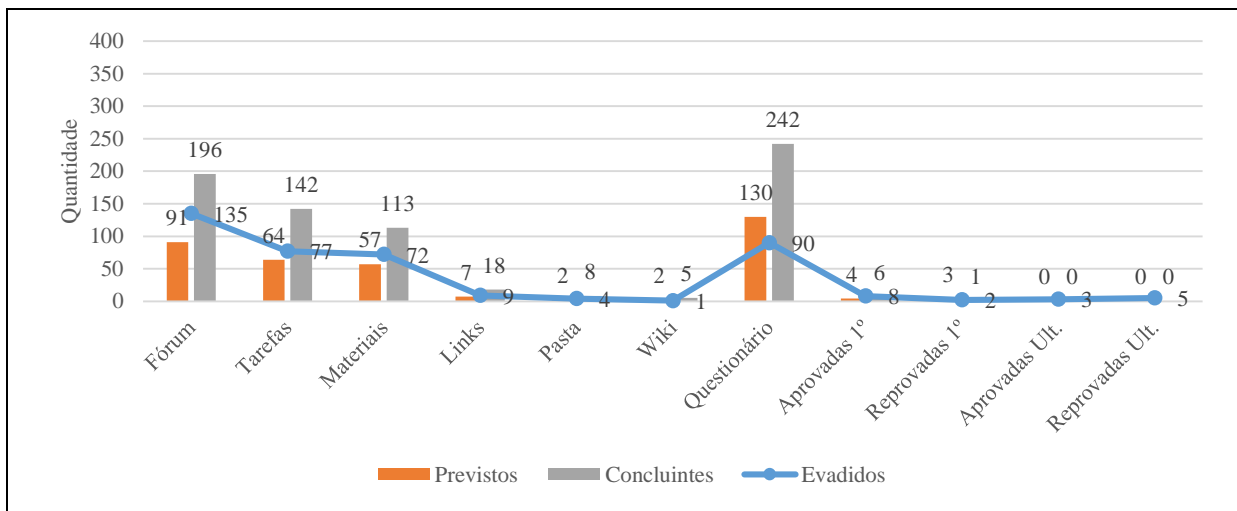
Nos Gráficos 11, 12, 13 e 14 é possível visualizar as quantidades médias dos dados utilizados no processo de MD para os conjuntos de dados analisados.

Gráfico 11 — Quantidade média dos dados utilizados na MD – Letras – Espanhol



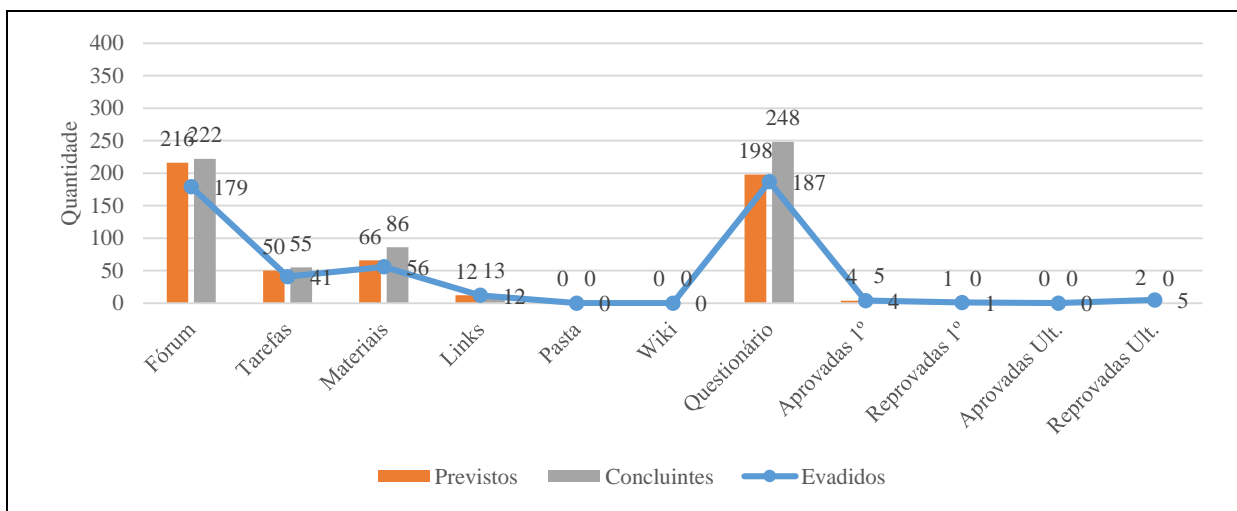
Fonte: Autor.

Gráfico 12 — Quantidade média dos dados utilizados na MD – Pedagogia



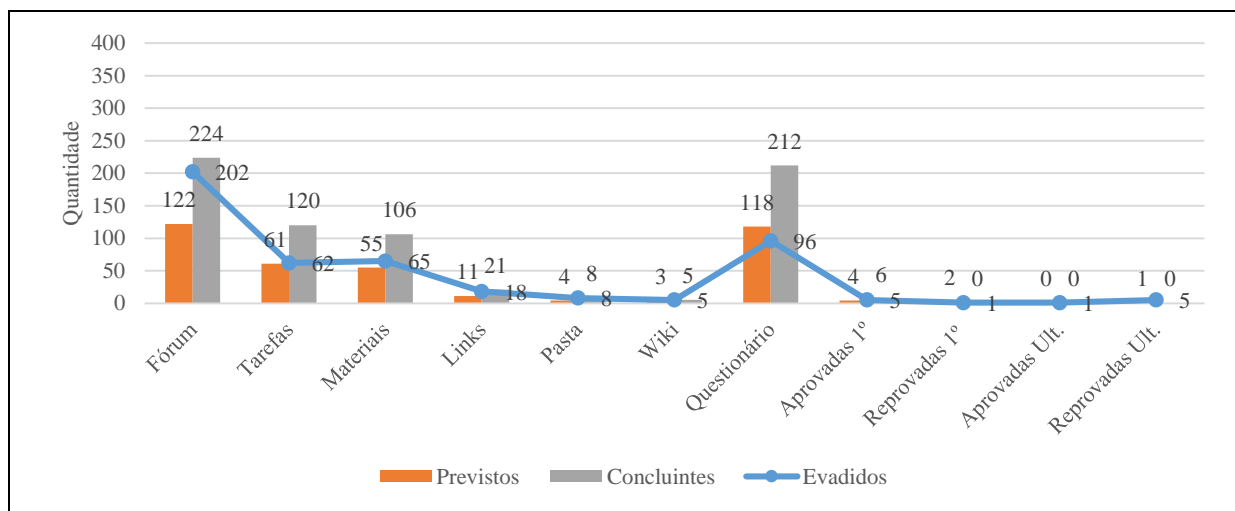
Fonte: Autor.

Gráfico 13 — Quantidade média dos dados utilizados na MD – Educação Especial



Fonte: Autor.

Gráfico 14 — Quantidade média dos dados utilizados na MD – Todos



Fonte: Autor.

Os dados estão distribuídos em três séries: alunos concluintes, abrangendo os alunos que tiveram sua situação final no processo de mineração definida como “Aluno Regular”; previstos, contendo os alunos que foram indicados no final do processo como “Evadido”; e, evadidos, sendo composto pelos alunos que vieram a evadir no final do período analisado.

Os atributos relacionados à realização de outros cursos não foram expostos nos gráficos pelo fato de que seus valores médios estão compreendidos no intervalo entre zero e um.

Para o cálculo da média dos alunos previstos e concluintes foram utilizados os alunos indicados nas simulações que apresentaram o maior índice de Medida-F. Neste experimento, para todos os conjuntos analisados, o algoritmo IBk com dados normalizados obteve o melhor desempenho.

No Gráfico 11, que dispõe as quantidades médias do curso de Letras – Espanhol, é possível verificar a proximidade dos valores entre os alunos previstos a evadir e aqueles que evadiram de fato. O que permite constatar que mesmo a indicação da previsão não estando correta, ou seja, não veio a ocorrer, o aluno previsto se enquadra no comportamento dos alunos que já evadiram, desse modo pode se tornar futuro desistente.

Nos Gráficos 12, 13 e 14 o comportamento se mantém em relação à semelhança entre os valores encontrados para os alunos previstos e evadidos. O que fica mais evidente, em uma análise conjunta, é a utilização de alguns recursos do AVEA em determinados cursos por parte dos alunos. A partir dos gráficos é possível constatar que os recursos utilizados pelos cursos de Pedagogia e Educação Especial se restringem basicamente nos Fóruns, Tarefas, Materiais e Questionários. Já o curso de Letras – Espanhol apresenta uma leve superioridade em relação ao

uso dos recursos de *Wiki* e *Links*. No entanto, considerando os resultados obtidos em ambos os experimentos, verifica-se que essa variação não influenciou na previsão.

7.3.6 Disponibilização

Após a avaliação da viabilidade de utilização dos algoritmos para previsão de evasão, os dados relacionados devem ser disponibilizados de forma a proporcionar a análise para a tomada de decisões. Desse modo, na presente pesquisa, a disponibilização destas informações foi promovida através do desenvolvimento de um aplicativo móvel. A escolha da plataforma foi motivada por proporcionar um fácil acesso às informações disponibilizadas. O desenvolvimento do aplicativo será detalhado no capítulo 8.

7.4 ANÁLISE DOS EXPERIMENTOS

O primeiro experimento foi elaborado exclusivamente com os dados provenientes de logs do AVEA MOODLE, sendo suas simulações efetuadas com dois formatos de dados, não normalizados e normalizados. Foram aplicados sete algoritmos diferentes, abrangendo dados sobre as interações dos alunos do curso de Letras – Espanhol. Todas as simulações realizadas com dados não normalizados apresentaram bons resultados podendo-se observar que o melhor deles foi obtido pelo algoritmo *Naive Bayes* com 58,39% de Medida-F. No entanto, algoritmos como o próprio *Naive Bayes*, *Random Forest* e *IBk* obtiveram respectivamente 51,59%, 54,34% e 43,56% de Medida-F para o formato normalizado, atingindo também bons índices.

Diante dos bons resultados alcançados no primeiro experimento foi possível determinar que a utilização das técnicas de MD para prever a evasão de alunos de cursos de graduação EaD era factível. Desse modo, foi dado prosseguimento no desenvolvimento da pesquisa com a elaboração do segundo experimento.

O segundo experimento contemplou também, além dos dados de interações do AVEA, alguns dados quantitativos relacionados ao número de disciplinas aprovadas/reprovadas pelos alunos. Estes dados foram incluídos com o intuito de melhorar as previsões dos algoritmos. Ademais, foram analisados outros cursos, mais especificamente, Letras – Espanhol, Pedagogia e Educação Especial, sendo empregados os mesmos algoritmos e formatos de dados do experimento anterior. Nesse experimento, por envolver um número maior de cursos e, assim, ser possível uma análise mais ampla, as simulações tiveram quatro conjuntos de dados distintos

em que três continham apenas os dados de cada curso e o último compreendia, na mesma simulação, os dados de todos os cursos.

Na mesma trajetória do primeiro experimento, o segundo também alcançou resultados satisfatórios para todos os conjuntos de dados. No entanto, diferentemente do primeiro, o formato normalizado apresentou melhores índices, obtendo-se o melhor deles pelo algoritmo IBk para todos os conjuntos de dados. Além disso, foi possível observar que tanto as simulações realizadas para os dados exclusivos dos cursos quanto para o conjunto abrangendo a união dos dados apresentaram bons resultados, permitindo a escolha de qualquer uma das abordagens para previsão da evasão.

Diante do exposto é viável afirmar que o fator chave de sucesso dos experimentos está diretamente relacionado com a qualidade dos dados utilizados no processo de mineração. Esta circunstância demonstra a importância das fases de entendimento e preparação dos dados existentes na metodologia de KDD utilizada.

Observando os algoritmos utilizados, o *Naive Bayes*, *Random Forest* e IBk apresentaram os melhores resultados e mantiveram uma regularidade no desempenho em praticamente todas as simulações realizadas. No entanto, com base na métrica de Medida-F, é possível verificar que o algoritmo IBk obteve destaque entre eles apresentando a melhor performance no segundo experimento, mas, também, resultados interessantes no primeiro.

No âmbito dos cursos, o curso de Letras - Espanhol foi o único constante nos dois experimentos. Neles é possível verificar que as quantidades totais de alunos disponíveis não tiveram alterações, totalizando 571. A diferença ocorrida entre eles foi em relação aos dados do conjunto de treinamento em que foram incluídos os alunos que tiveram sua situação definida – evadiram/formaram-se – no decorrer do desenvolvimento. Com esses novos alunos no conjunto de treinamento foi possível perceber uma melhora nos resultados apresentados para o curso no segundo experimento. Por outro ângulo, é possível fazer uma análise com relação ao número de alunos previstos no primeiro experimento que não evadiram de fato no período – segundo semestre de 2017, mas, no entanto, vieram a evadir no período abrangido pelo segundo experimento – primeiro semestre de 2018. Como exemplo, o algoritmo IBk com dados normalizados previu a evasão de 104 alunos no primeiro período dos quais 23 vieram a evadir no mesmo. Dos 81 alunos previstos que não evadiram no período, 13 deles vieram a evadir no período posterior. Esta situação reforça o argumento de que se os alunos previstos não venham a evadir no período analisado podem vir a se tornar futuros desistentes.

O curso de Pedagogia continha o maior conjunto de dados entre os cursos analisados. Apesar disso, analisando as métricas de Medida-F, foi o que obteve os piores resultados entre

as simulações. Isso, provavelmente, se deve ao fato de que boa parte dos alunos contidos no conjunto de testes possuíam características semelhantes aos alunos evadidos, mas, no entanto, não evadiram no período analisado. Apesar da métrica de Medida-F não ter alcançado índices elevados, quase a totalidade dos alunos que vieram a evadir estavam presentes entre os alunos previstos pelos algoritmos.

Já o curso de Educação Especial continha 449 alunos no seu conjunto de dados. Apesar da baixa quantidade de alunos, as simulações realizadas para o curso apresentaram bons índices.

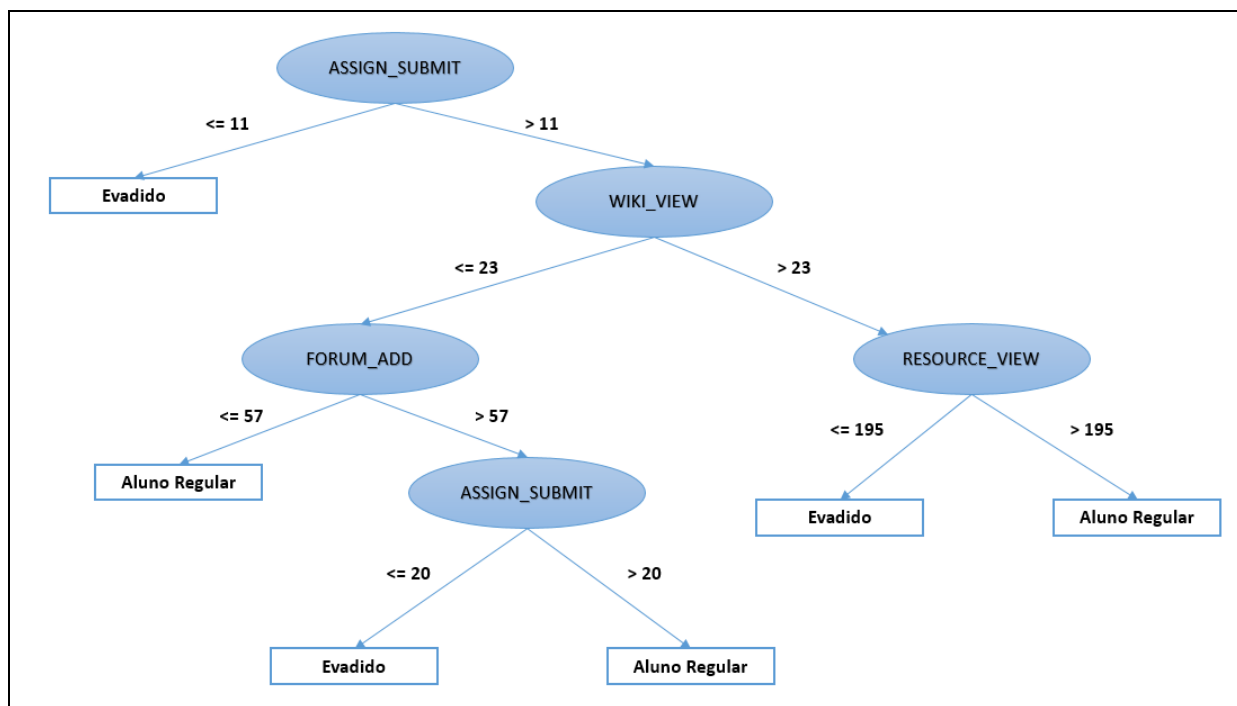
A melhor performance obtida pelos algoritmos no segundo experimento pode estar relacionada a dois fatores. O primeiro está diretamente relacionado com o conjunto de dados de treinamento, que com o passar do tempo, tende a reunir uma maior a quantidade de dados e, assim, proporcionar a montagem de modelos mais apurados pelos algoritmos. E o segundo fator associa-se com a abrangência de mais informações no conjunto de dados, o que pode ser visto com a inclusão dos atributos de disciplinas aprovadas e reprovadas juntamente com as quantidades de interações dos alunos no AVEA.

Por fim, no que tange ao levantamento do perfil dos alunos com tendência à evasão, por meio dos modelos gerados pelos algoritmos cabem, inicialmente, algumas considerações. Conforme abordado no capítulo 4, para efetuar uma previsão, cada algoritmo gera um modelo com base nos padrões de dados obtidos no conjunto de treinamento. Esses modelos são gerados de acordo com o tipo de cada algoritmo, podendo ser baseados em árvores de decisões, distância, conexionistas, probabilístico, dentre outros.

Porém, a obtenção do conhecimento/padrões dos modelos gerados, principalmente dos algoritmos que empregam abordagens matemáticas, não são factíveis, o que dificulta a caracterização do perfil de um aluno desistente. Este é o caso do algoritmo que obteve os melhores resultados nas simulações realizadas, o IBk.

Dentre os algoritmos utilizados, o que proporciona a visualização do modelo gerado pela ferramenta WEKA é o J48. Apesar de sua performance não estar entre as melhores, ele nos permite deduzir algumas características interessantes. Na Figura 12 é possível visualizar o modelo gerado no primeiro experimento, considerando apenas os dados do AVEA.

Figura 12 — Árvore de decisão - Primeiro experimento



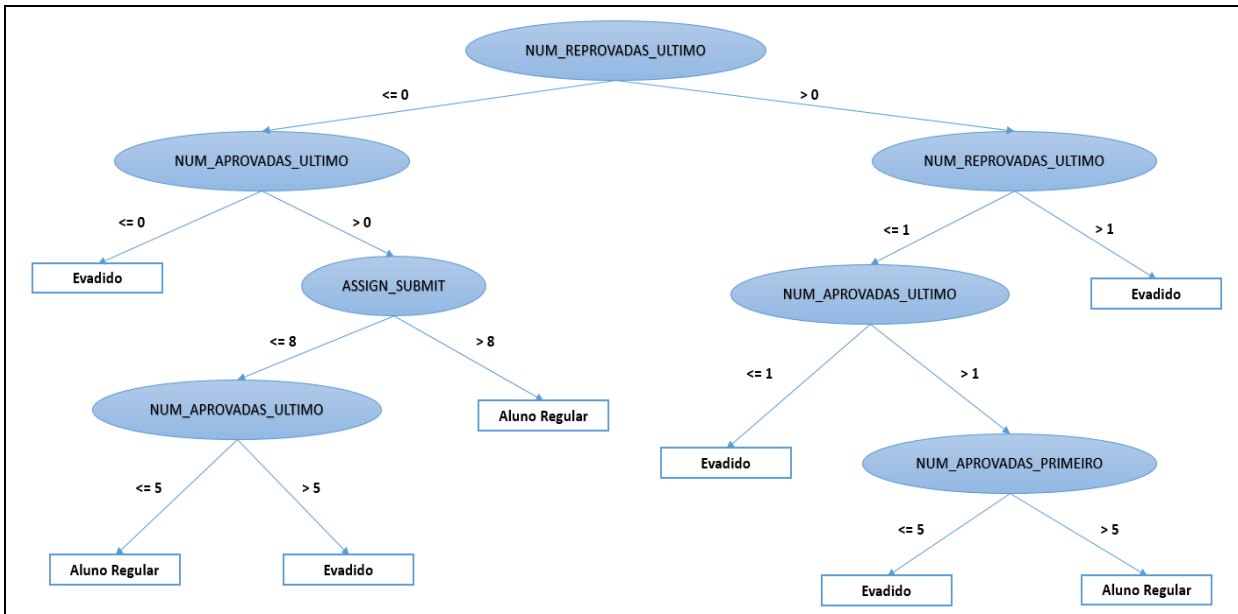
Fonte: Autor.

Conforme pode ser visto no nó da árvore demonstrado na Figura 12, o fator elementar que determina a evasão do aluno no modelo gerado é o número de envio de tarefas (ASSIGN_SUBMIT) no decorrer de um semestre. O aluno que submete 11 ou menos tarefas tende a evadir. Para os alunos que enviam regularmente as tarefas é analisado o uso do recurso *Wiki* (WIKI_VIEW). Caso o recurso seja pouco utilizado, ou seja, menor ou igual a 23 vezes, a evasão é definida com base na utilização do fórum de discussões. Caso contrário, define-se de acordo com os acessos aos materiais disponibilizados no ambiente (RESOURCE_VIEW). Esse ponto permite deduzir duas situações: (a) para as disciplinas que utilizam o recurso de *Wiki* os alunos tendem a utilizar os materiais disponibilizados no ambiente para colaborar com a construção de conteúdos no recurso; e, (b) para as disciplinas que não empregam o *Wiki*, o que determina a situação do aluno, além da submissão das tarefas, é a participação nos fóruns de discussões. Em suma, a utilização dos recursos do AVEA tais como Tarefas, *Wiki*, Fóruns e Recursos podem ser indicativos da situação final do aluno no curso. Este fato vai ao encontro dos resultados obtidos na RSL abordada no capítulo 3 e, também, com as causas da evasão levantadas por Abbad, Zerbini e Souza (2010)¹⁰.

¹⁰ Maiores detalhamentos das causas levantadas pelos autores podem ser vistos na seção 2.2.

Na Figura 13 é possível visualizar a árvore de decisão gerada para o segundo experimento, considerando as informações acadêmicas juntamente com as interações do AVEA.

Figura 13 — Árvore de decisão - Segundo experimento



Fonte: Autor.

Na árvore demonstrada na Figura 13 verifica-se a predominância dos dados de números de aprovações/reprovações nas disciplinas dos semestres iniciais e finais. O número de disciplinas reprovadas no último semestre encontra-se no nó da árvore sendo o elemento principal. O aluno com mais de uma disciplina reprovada no último semestre cursado tende a evadir. No entanto, o aluno com uma disciplina reprovada, mas sem nenhuma disciplina aprovada é, também, um provável desistente. Por fim, o aluno com mais de uma disciplina aprovada no último semestre e que não foi aprovado na maior parte das disciplinas do primeiro, tende a abandonar o curso. Ou seja, o número de disciplinas aprovadas no primeiro semestre realizado pelo aluno pode predizer o seu comportamento no final do curso.

No outro lado da árvore é possível verificar que por mais que o aluno não venha a reprovar em nenhuma disciplina no último semestre, caso não seja participativo no AVEA (no envio de tarefas – ASSIGN_SUBMIT), é um provável candidato a desistir.

Esse fator permite deduzir que a utilização de dados de interação do AVEA com dados do sistema de gestão acadêmico pode ser realizada em conjunto para propiciar uma melhoria nos resultados das predições.

Enfim, apesar dos modelos gerados pelo algoritmo J48 não terem alcançado os melhores resultados permite obter conhecimentos que não seriam facilmente detectados com a simples visualização dos dados brutos, tais como: a interação dos alunos nos recursos do MOODLE, em especial Tarefas, *Wikis*, Fóruns e Recursos podem ser indicativos no que se refere a evasão do aluno de um curso.

Os conhecimentos obtidos por meio da árvore de decisão do algoritmo J48 já foram relacionados em alguns dos trabalhos descritos na revisão de literatura das causas da evasão (seção 2.2). Esse fato ocorreu devido a limitação dos dados do AVEA MOODLE utilizados no processo de MD, os logs brutos de interações dos alunos. Com dados mais abrangentes do AVEA seria possível a obtenção de conhecimentos que talvez permitissem a conclusão de novas causas de evasão.

Cabe ressaltar aqui que os demais algoritmos podem ter obtido modelos que empregam os demais atributos utilizados para mineração, mas que não foram utilizados pelo J48.

No próximo capítulo será abordado o aplicativo desenvolvido para a disponibilização dos dados da mineração, o *IPermanency*.

8 IPERMANENCY: UM APLICATIVO PARA DISPONIBILIZAÇÃO DOS DADOS DA MINERAÇÃO

Para a disponibilização dos dados provenientes do desenvolvimento da pesquisa, abordados no capítulo anterior, foi desenvolvido um aplicativo móvel chamado *iPermanency*. O objetivo do aplicativo é disponibilizar as informações oriundas da MD para auxiliar os gestores nas tomadas de decisões. A funcionalidade principal consiste na listagem de alunos com tendência à evasão indicados pelos algoritmos. Além disso, serão viabilizadas funcionalidades para possibilitar a comparação dos dados dos alunos com evasão prevista e os alunos indicados como regulares. Essa comparação permite a análise das causas que possam ter influenciado na previsão.

Os dados demonstrados pelo aplicativo serão aqueles resultantes do algoritmo que atingiu o melhor índice de Medida-F nas simulações realizadas. No desenvolvimento da pesquisa, o algoritmo que se enquadrou nessa situação foi o IBk com dados no formato normalizado.

Na seção que segue são abordadas as etapas para o desenvolvimento da ferramenta.

8.1 ETAPAS DE DESENVOLVIMENTO

Para o desenvolvimento foi utilizada a metodologia de desenvolvimento de aplicativos para dispositivos móveis chamada *Mobile-D* (abordada na seção 6.4). Desse modo, nas subseções que seguem serão abordadas as suas etapas.

8.1.1 Exploração

A primeira fase da metodologia consiste na exploração dos requisitos a serem atendidos para o estabelecimento das características básicas do aplicativo. Como artefato de saída é proposta a confecção da seguinte documentação: requisitos iniciais, plano de projeto e *CheckList* do plano de projeto. Pelo aplicativo proposto não se tratar de um projeto comercial não serão tratados os aspectos relacionados à documentação de plano de projeto e *CheckList*. Sendo assim, os requisitos funcionais estabelecidos são demonstrados no Quadro 16.

Quadro 16 — Requisitos funcionais

Código	Requisito
RF1	Listar os índices de evasão geral e evasão média anual para cada curso.
RF2	Listar os índices de evasão média anual para os polos de ensino de um curso.
RF3	Listar os alunos com tendência à evasão de acordo com o seu polo de ensino.
RF4	Relacionar as informações consideradas para a MD entre os alunos evadidos e aqueles em que se estima a sua continuidade no curso.
RF5	Relacionar as informações consideradas para a MD dos alunos de um polo de ensino e os concluintes do curso.

Fonte: Autor.

Já os requisitos não-funcionais serão demonstrados no Quadro 17.

Quadro 17 — Requisitos não-funcionais

Código	Requisito
RNF1	Sincronizar as informações da MD de forma automática.

Fonte: Autor.

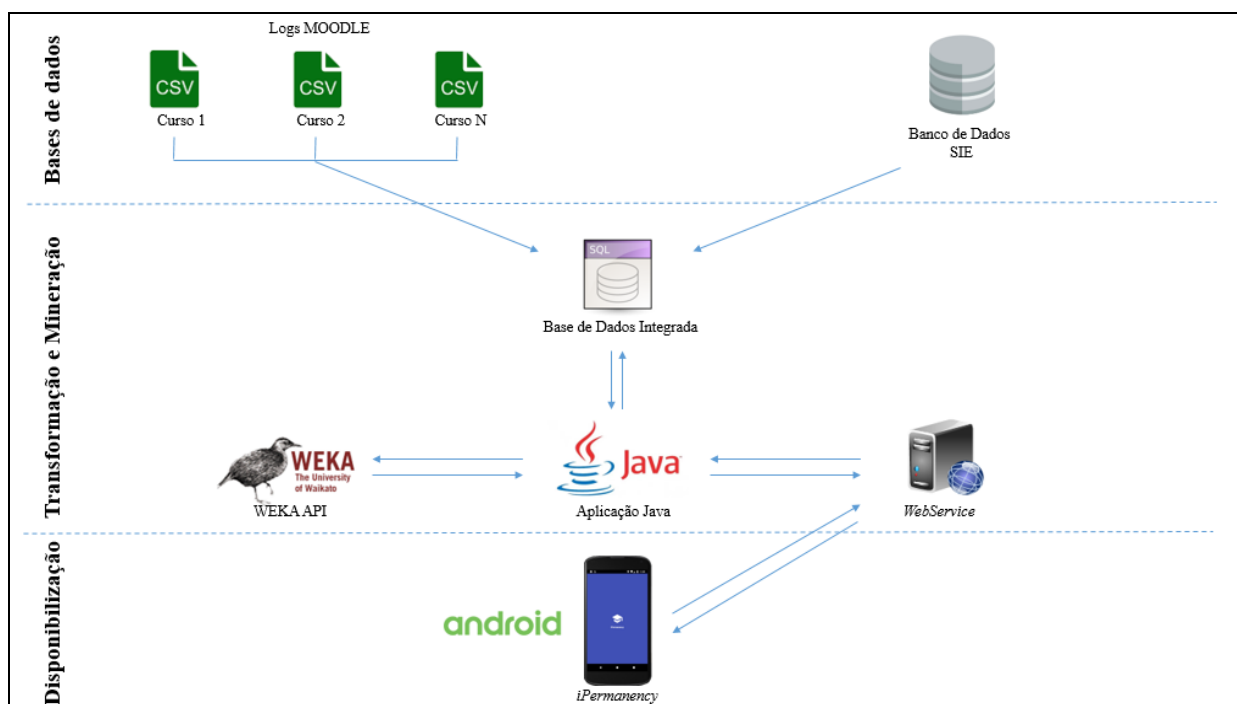
Os requisitos funcionais foram estabelecidos visando facilitar a compreensão dos dados resultantes da MD pelos gestores. Sendo assim, o primeiro requisito funcional (RF1) visa possibilitar uma análise quantitativa em relação à evasão no curso. Já a listagem de índices de evasão média anual por polo de ensino, RF2, possibilita uma visão da ocorrência de evasão por polo, permitindo assim, que os gestores direcionem seus esforços no que visa atenuar os índices apresentados. O RF3 surge da necessidade de se conhecer os alunos com tendência à evasão para que ações possam ser tomadas junto a eles a fim de evitá-las. Já os requisitos RF4 e RF5 permitem que o gestor visualize os dados analisados pelos algoritmos de MD que possam estar influenciando na indicação da evasão. Com esses dados, é possível realizar intervenções junto aos docentes/tutores no sentido de incentivar a utilização de determinados recursos do AVEA que, por exemplo, podem não estar sendo utilizados de forma adequada.

O requisito não-funcional, RNF1, foi definido para possibilitar a utilização do aplicativo sem a necessidade de possuir conexão à Internet cada vez que for preciso consultar os dados. Após o estabelecimento dos requisitos é possível seguir para a fase seguinte da metodologia, a Inicialização.

8.1.2 Inicialização

Na fase de inicialização são elaborados os artefatos que servem de base para as demais fases do desenvolvimento, tais como: modelo entidade-relacionamento (ER), diagrama de casos de uso e protótipo de interfaces. Para facilitar o entendimento na Figura 14 é apresentada a arquitetura do sistema.

Figura 14 — Arquitetura do sistema



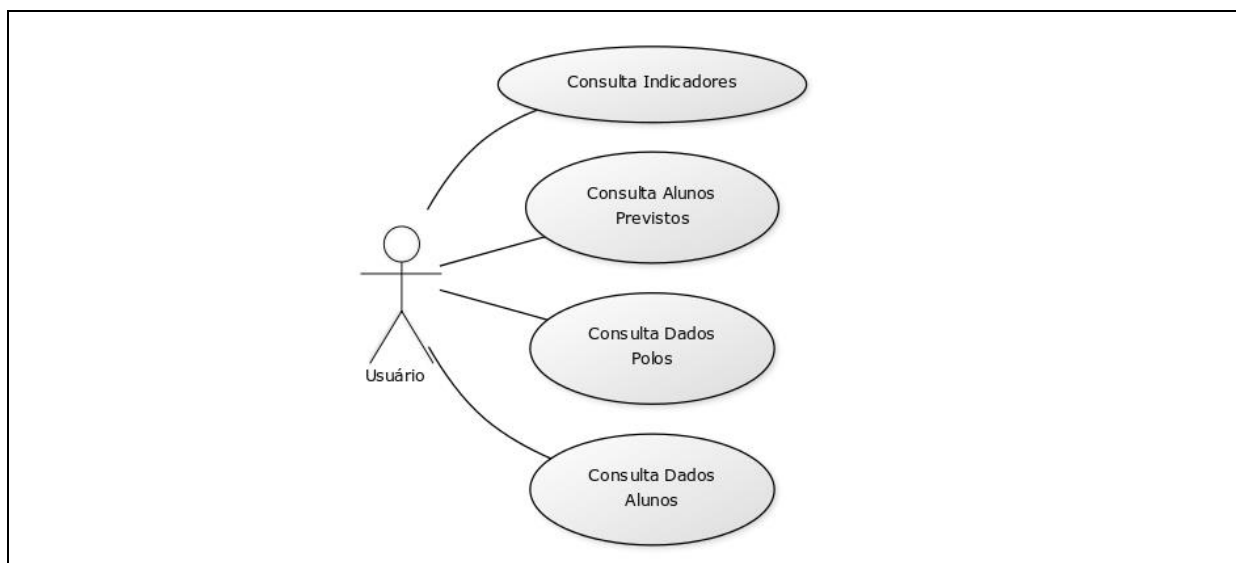
Fonte: Autor.

A arquitetura do sistema é dividida em três partes. A primeira contém as bases de dados que dispõem das informações que serão utilizadas no processo de MD. São elas: dados de interação do MOODLE e dados do sistema institucional, o SIE. Na segunda parte consta os componentes que realizam a transformação e MD, sendo eles: uma base de dados única, contendo tanto os dados do MOODLE quanto os dados do SIE; uma aplicação Java que utiliza a API do WEKA para efetuar o processo de MD e armazenar os dados resultantes do processo; e, um *Webservice* que disponibiliza os dados resultantes do processamento para serem acessados pelo aplicativo de disponibilização, o *iPermanency*. Por fim, a terceira parte corresponde à disponibilização que contém o aplicativo desenvolvido utilizando a metodologia.

A plataforma Android foi escolhida para o desenvolvimento do aplicativo por ser uma das mais utilizadas atualmente, o que facilita a distribuição e adoção pelos interessados.

Com o intuito de possibilitar a consulta dos dados no aplicativo sem necessidade de estar conectado na Internet foi desenvolvida uma base de dados simples, com apenas uma tabela, para armazenar os dados da mineração. Os dados a serem armazenados nessa tabela serão obtidos por meio de um *WebService* descrito na arquitetura do sistema.

Figura 15 — Diagrama de casos de uso



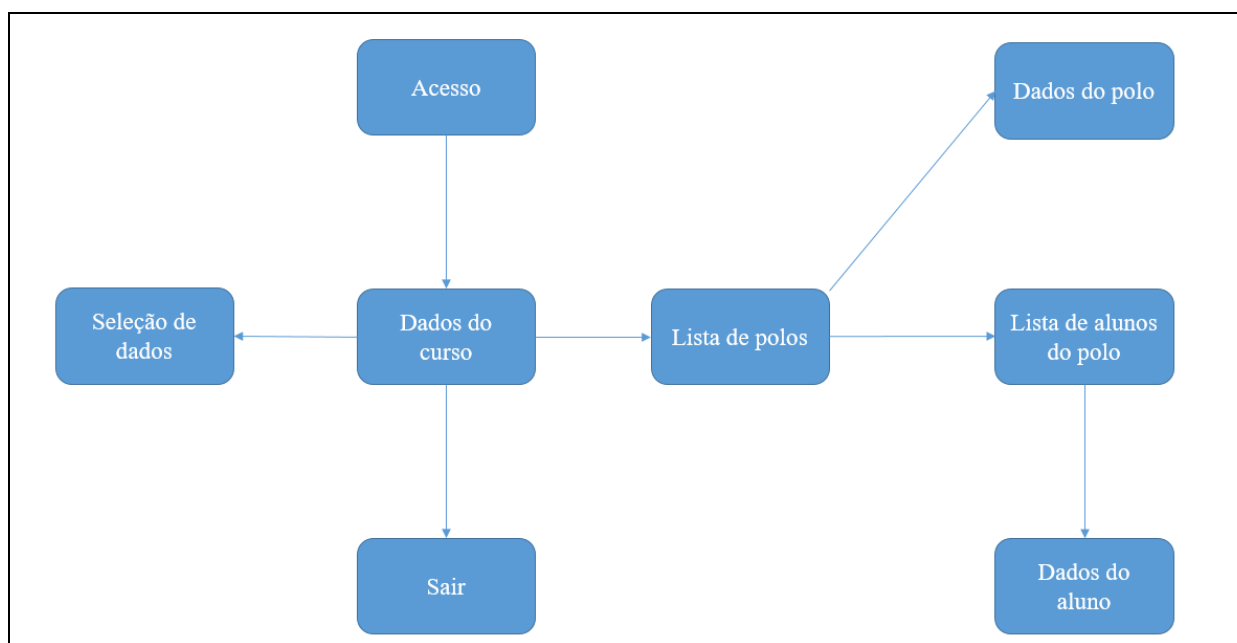
Fonte: Autor.

Na Figura 15 é possível visualizar o diagrama de casos de uso do aplicativo contendo as funcionalidades previstas para o sistema. A primeira funcionalidade consiste na consulta de indicadores relacionados à evasão. A segunda e principal funcionalidade refere-se à consulta de alunos previstos para evasão. Já a terceira e quarta funcionalidades estão relacionadas, respectivamente, à consulta dos dados dos polos de forma mais genérica e, também, dos alunos.

Por fim, na Figura 16, é possível visualizar o artefato de fluxo de utilização da aplicação. Por meio dele é possível verificar as funcionalidades projetadas juntamente com o fluxo para acessá-las. As funcionalidades de “Acesso” e “Sair” representam, respectivamente, as funcionalidades que permitem o acesso, bem como o encerramento do aplicativo. Depois delas aparece a funcionalidade de “Dados do curso” que representa a tela inicial da aplicação, contendo os dados dos indicadores de evasão geral e evasão média do curso. Além disso, por meio dos “Dados do curso” é possível visualizar as informações detalhadas por polo. Para isso, então, deve ser acessada a funcionalidade de “Lista de Polos”. Na listagem de polos é possível,

ainda, acessar outras duas funcionalidades: “Dados do polo” e “Lista de alunos do polo”. A primeira permite a visualização dos dados utilizados no processo de MD, onde estão relacionados os dados dos alunos concluintes do curso e dos alunos que estão matriculados no polo selecionado. Já a segunda, “Lista de alunos do polo” possibilita a visualização da listagem de alunos que foram previstos no processo de MD como prováveis evadidos. Por fim, para cada aluno é possível acessar os dados relacionados entre ele e os alunos concluintes.

Figura 16 — Fluxo de utilização



Fonte: Autor.

Após a definição dos itens necessários para a inicialização do desenvolvimento, na próxima seção será abordada a fase estabelecida na sequência da metodologia, a Produção.

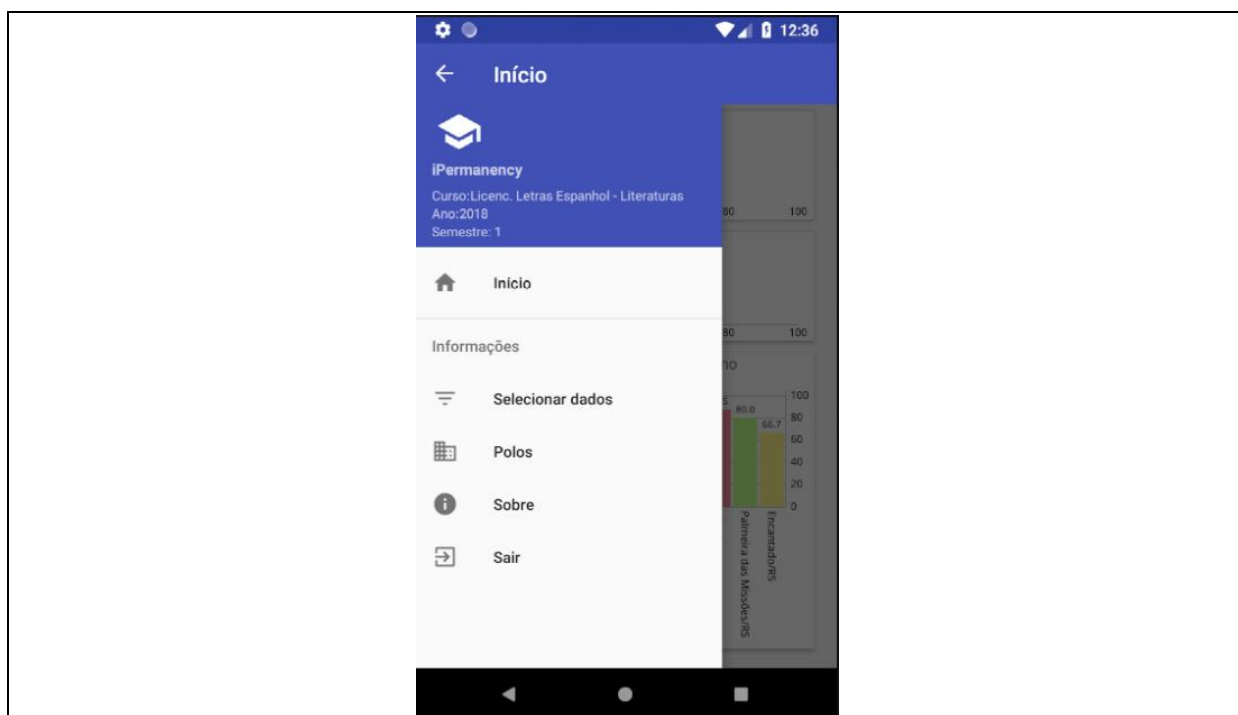
8.1.3 Produção

A produção consiste na fase em que se encontra a maior parte do esforço. Para cada funcionalidade a ser desenvolvida é criado um conjunto de documentos onde são estipulados os responsáveis por sua realização, prazos para entrega bem como quantidade de erros aceitáveis durante a realização dos testes. Logo após a documentação é inicializado, então, o desenvolvimento das funcionalidades do aplicativo. Nas subseções que seguem serão abordadas suas funcionalidades.

8.1.3.1 Menu principal

Na Figura 17 é possível visualizar o menu principal do aplicativo. Ele consiste de um menu lateral que pode ser acessado diretamente de qualquer funcionalidade por meio do botão de menu posicionado no canto superior do aplicativo.

Figura 17 — Menu principal



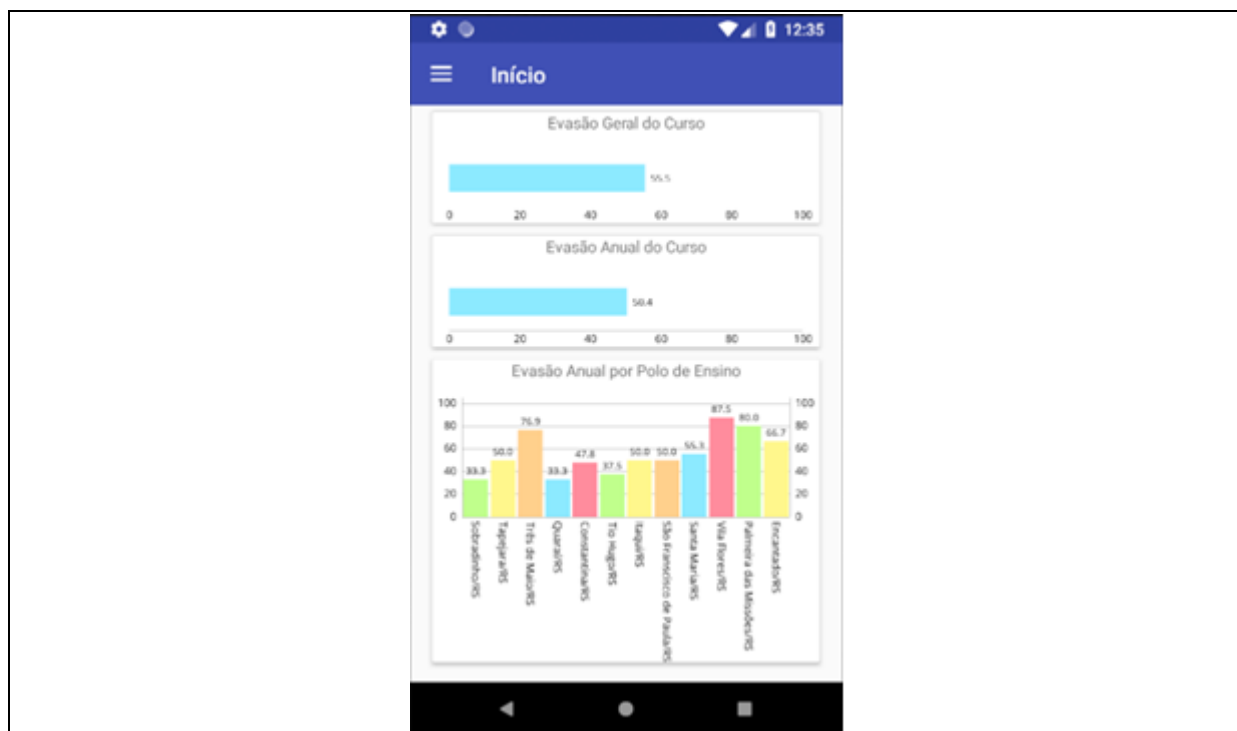
Fonte: Autor.

Na parte superior do menu é apresentado os dados do curso, ano e semestre selecionados na funcionalidade de seleção de dados. Logo abaixo são apresentadas as opções do menu, sendo elas: início, selecionar dados, polos, sobre e sair.

8.1.3.2 Dados do curso

Na Figura 18 é possível visualizar a tela de dados do curso. Nesta tela estão dispostos os indicadores de evasão geral, bem como os indicadores de evasão anual média do curso e dos polos de ensino.

Figura 18 — Dados do curso



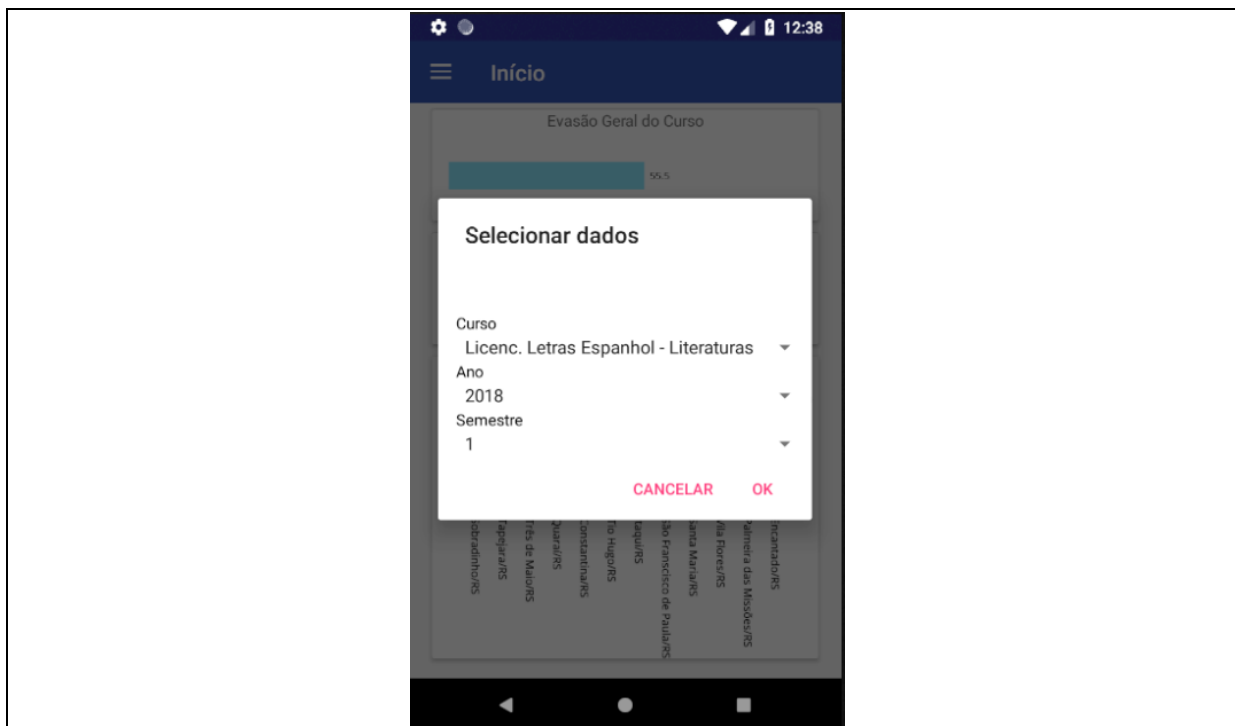
Fonte: Autor.

Esta funcionalidade consiste na tela principal do aplicativo, sendo exibida logo após a sua inicialização pelo usuário.

8.1.3.3 Seleção de dados

Na Figura 19 é possível visualizar a tela de seleção dos dados. Esta tela é composta pelos campos “Curso”, “Ano” e “Semestre” sendo nela que se define as informações que serão demonstradas nas demais funcionalidades do aplicativo.

Figura 19 — Seleção de dados

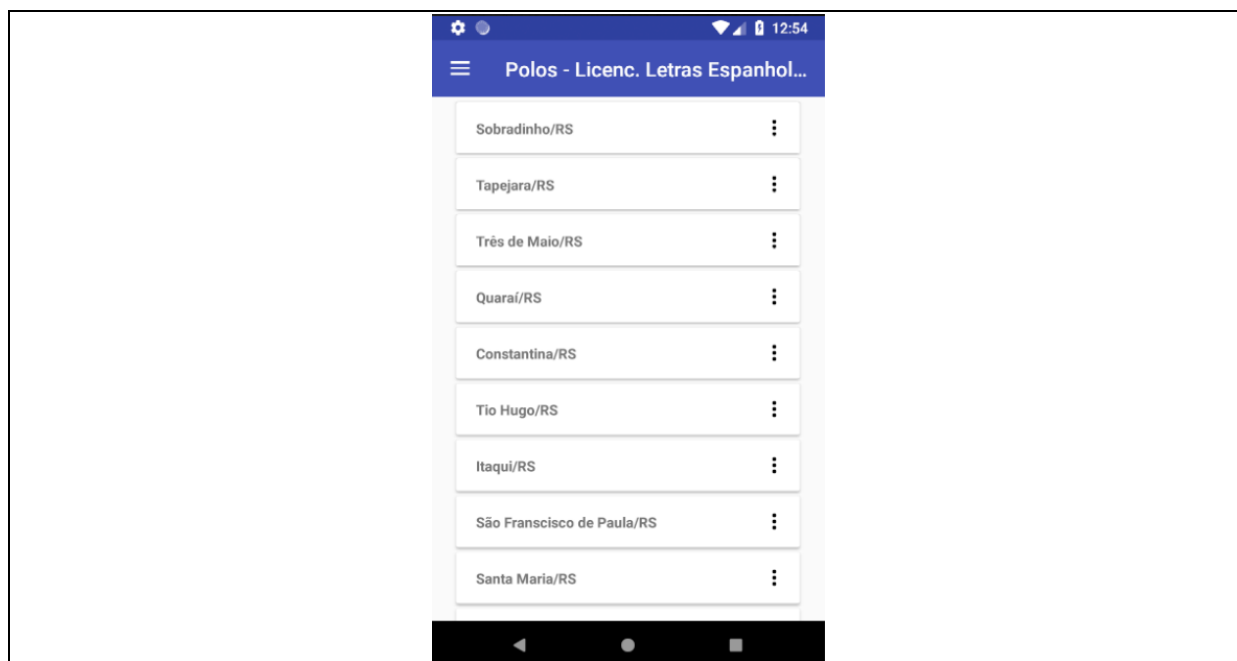


Fonte: Autor.

8.1.3.4 Lista de polos

Na Figura 20 é possível visualizar a lista de polos de acordo com o curso definido na funcionalidade de seleção.

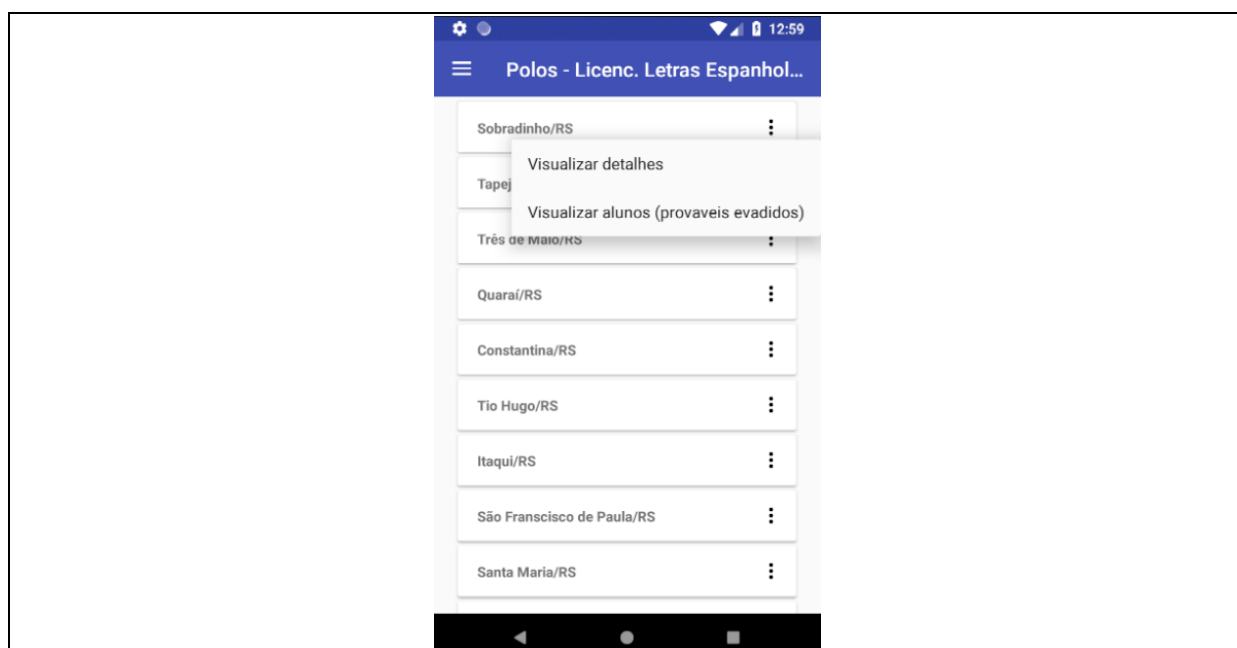
Figura 20 — Lista de polos



Fonte: Autor.

Para cada polo apresentado é possível acessar a lista de alunos com evasão prevista e dados do polo, conforme menu demonstrado na Figura 21.

Figura 21 — Menu acesso detalhes dos polos

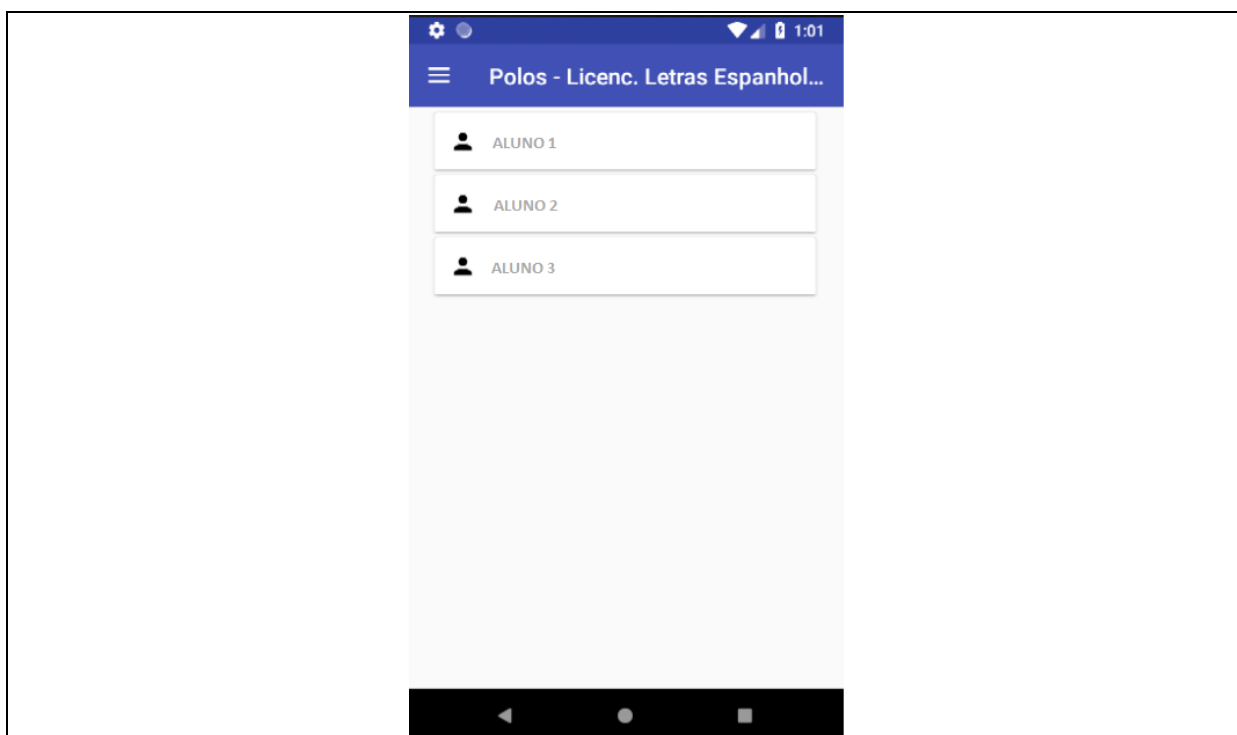


Fonte: Autor.

8.1.3.5 Lista de alunos do polo

Na Figura 22 é possível visualizar a lista de alunos com evasão prevista em um polo selecionado. Como pode ser visto por meio do mapa de navegação, esta funcionalidade é acessada pela listagem de polos.

Figura 22 — Lista de alunos do polo

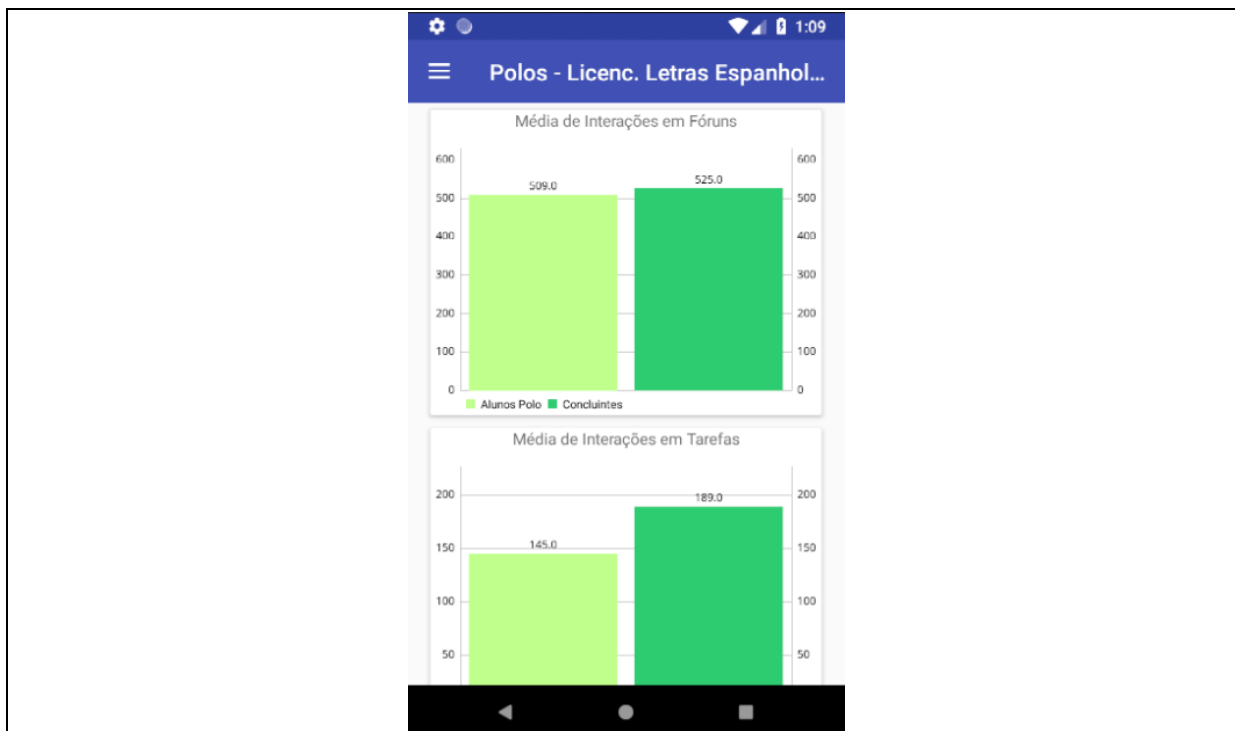


Fonte: Autor.

8.1.3.6 Dados do polo e do aluno

Na Figura 23 é possível visualizar a tela de visualização dos dados do polo considerados no processo de MD. Nela é possível comparar os dados dos alunos concluintes do curso com os alunos que estão alocados no polo de ensino selecionado.

Figura 23 — Dados do polo e do aluno



Fonte: Autor.

Conforme pode ser visto no mapa de navegação da aplicação na Figura 16, é prevista, ainda, uma funcionalidade de “Dados do aluno”. O *layout* desta funcionalidade é o mesmo apresentado na Figura 23. O diferencial entre eles encontra-se nos dados demonstrados. Na funcionalidade de “Dados do polo” são relacionados os dados dos alunos matriculados em um determinado polo e sua relação com os alunos concluintes do curso. Já a funcionalidade de “Dados do aluno”, ao invés de exibir os dados dos alunos do polo, demonstra os dados de um aluno em específico e sua relação com os alunos concluintes do curso.

8.1.4 Estabilização e testes

Após a implementação das funcionalidades, nesta fase foram realizados testes finais de integração com o intuito de assegurar a qualidade do desenvolvimento. Além disso, o aplicativo é verificado junto aos usuários quanto ao atendimento dos objetivos estabelecidos no início do projeto. Os aspectos relacionados com a avaliação serão abordados na seção a seguir.

8.2 AVALIAÇÃO

Após o desenvolvimento do aplicativo procedeu-se com a sua avaliação, junto aos coordenadores de curso. Para auxiliar neste aspecto foi utilizado o *framework* DECIDE proposto por Preece, Rogers e Sharp (2013) que estabelece um conjunto de etapas que orientam o planejamento e a realização da avaliação. Para facilitar a exposição algumas etapas definidas pelos autores tiveram sua nomenclatura adaptada, conforme descritas nas subseções a seguir.

8.2.1 Levantamento dos objetivos

A primeira etapa estabelecida no *framework* é o levantamento dos objetivos da avaliação. Sendo assim, no contexto da pesquisa, o objetivo principal da avaliação é verificar junto aos coordenadores de cursos o quanto os dados oriundos do processo de MD, disponibilizados pelo aplicativo, podem auxiliá-los nas atividades de gestão de permanência dos alunos.

8.2.2 Definição das questões

Após a definição dos objetivos, devem ser estabelecidas as questões a serem respondidas. Sendo assim, foram relacionadas as seguintes questões:

- a) As informações dos alunos com tendência à evasão no semestre são relevantes para a gestão?
- b) Dos alunos previstos, já havia conhecimento no seu enquadramento nesta situação?
Se sim, como?
- c) Em relação às informações utilizadas no processo de MD, elas auxiliam na identificação dos fatores que levaram ao apontamento do aluno na previsão?
- d) Se a disponibilização destes dados por meio do aplicativo fosse adotado no âmbito institucional, você iria utilizá-lo?
- e) Em relação ao aplicativo, achou fácil de usar?
- f) Teria alguma sugestão em relação às funcionalidades disponibilizadas?

8.2.3 Escolha dos métodos

A avaliação foi realizada individualmente junto a cada coordenador. Em um primeiro momento, foi demonstrado o aplicativo a eles. Após, será procedida com a entrevista abordando as questões definidas na etapa anterior.

8.2.4 Identificação das questões práticas

Em relação às questões práticas da avaliação, o avaliador deve possuir um dispositivo móvel na plataforma Android para demonstração do aplicativo para o participante. Isso é necessário, pois o aplicativo não foi concedido diretamente para os coordenadores, já que no escopo do desenvolvimento não estavam relacionadas questões de permissão de visualização de informações. Desse modo, o aplicativo não foi disponibilizado para os coordenadores para não permitir que dados de outros cursos que não sejam de sua responsabilidade possam ser acessados por ele.

8.2.5 Análise das questões éticas

Sobre as questões éticas, no convite ao coordenador foi esclarecido o intuito da avaliação que se relaciona com a opinião dele em relação aos resultados da pesquisa realizada. Nele, também foi esclarecido que, caso não deseje, sua identificação não será indicada na pesquisa. Além disso, como já detalhado nos aspectos práticos, o aplicativo não foi disponibilizado para os coordenadores, por manter a privacidade dos alunos que não pertencem ao curso pelo qual é responsável.

8.2.6 Avaliação, interpretação e apresentação dos resultados

A avaliação do aplicativo e dos dados obtidos no processo de mineração foi realizada junto aos coordenadores dos cursos analisados na pesquisa. Em um primeiro momento foram expostos os objetivos do trabalho e como eles foram abordados. Em seguida, foram levantadas informações junto a eles que pudessem esclarecer a forma como monitoravam/controlavam a evasão nos seus cursos. Por fim, o aplicativo foi demonstrado e realizadas as questões da avaliação, sendo que suas respostas se encontram sintetizadas no Quadro 18.

Quadro 18 — Respostas das questões de avaliação

Questão	Letras – Espanhol	Pedagogia	Educação Especial
A	Sim.	Sim.	Sim.
B	Sim, acompanhamento nos portais da instituição, e-mails tutores e AVEA.	Sim, acompanhamento nos portais da instituição e AVEA.	Sim, devido à proximidade dos tutores com os alunos.
C	Sim.	-	Sim.
D	Sim.	Sim.	Sim.
E	Sim, bem intuitivo.	Sim.	Sim.
F	Possibilidade de analisar os dados específicos por turmas	Disponibilidade do aplicativo para IOS.	-

Fonte: Autor.

Conforme demonstrado no Quadro 18, na questão “A”, relacionada com a importância das informações da evasão para a gestão, os coordenadores foram unânimes em afirmar que os dados disponibilizados são relevantes para assessorar nas suas atividades de gestão.

No que tange ao conhecimento do enquadramento dos alunos como prováveis desistentes, questão “B”, ambos reconheceram alguns alunos indicados no aplicativo. Além disso, afirmaram que dos alunos reconhecidos alguns haviam abandonado o curso no semestre avaliado e outros ainda estavam matriculados, porém somente para manter vínculo com o curso já que reprovam em grande parte das disciplinas do semestre.

A questão “B” também abordava a forma como os coordenadores obtinham o conhecimento em relação ao comportamento desses alunos. De acordo com suas respostas, a informação era obtida acompanhando o processo de matrícula dos alunos nos portais da instituição, por meio de relatórios de alunos matriculados – disponíveis no sistema acadêmico - e, pelo AVEA MOODLE. Boa parte desse conhecimento também é obtido pessoalmente pelos tutores e, posteriormente, repassado aos coordenadores.

Em relação aos dados comparativos, questão “C”, disponíveis nas funcionalidades de dados do polo e do aluno - demonstradas na seção 8.1.3.6, dois coordenadores afirmaram que eles auxiliam na descoberta dos fatores que levaram a indicação do aluno como provável evadido. Ainda, afirmaram que os dados comparativos por polo permitem a descoberta dos recursos do AVEA MOODLE que são pouco utilizados pelos alunos em um determinado polo de ensino. Essas informações permitem ao coordenador tomar ações junto aos professores/tutores no sentido de explorar melhor tais recursos.

Nas questões específicas do aplicativo, questões “D” e “E”, as respostas foram positivas. Os coordenadores afirmaram que utilizariam a ferramenta, no caso de ser promovida no âmbito institucional, e que acharam a mesma de fácil utilização.

Por fim, na questão “F”, em relação a novas funcionalidades, foram sugeridas a possibilidade de análise dos dados por turmas, a visualização dos dados históricos dos índices de evasão e a disponibilização do aplicativo para plataforma IOS.

Com base nas entrevistas foi possível perceber que a evasão é um tema que gera muitas preocupações para os gestores educacionais. No entanto, não existem aparatos projetados para a disponibilização de dados que possam indicar de forma objetiva os alunos com possibilidade de evadirem. Sendo assim, para obter estas informações os coordenadores analisam os dados disponíveis em diversos sistemas de informação, como: portais da instituição, relatórios do sistema acadêmico e relatórios do AVEA MOODLE. Além disso, é utilizado o conhecimento empírico dos tutores em relação aos alunos. Desse modo, considerando a quantidade de alunos matriculados nos cursos EaD, é inviável por parte da gestão analisar em detalhes os dados dos alunos e, ainda, buscar formas de impedir que a ocorrência da evasão se concretize. Nesse cenário, foi possível constatar que os dados obtidos pela MD são informações estratégicas para os gestores o que, consecutivamente, despertou o interesse de todos pela utilização do aplicativo.

Além da listagem dos alunos, foi destacada pelos coordenadores a importância dos dados relacionados aos indicadores de evasão por polo, disponível na tela principal do aplicativo. Segundo a opinião deles esse indicador permite definir, por exemplo, para quais polos de ensino serão ofertadas novas turmas de acordo com os índices apresentados.

Algumas causas de evasão também foram pontuadas pelos coordenadores, como: a evasão tende a ocorrer nos primeiros semestres do curso, os polos com participação ativa dos tutores apresentam menores índices de desistência, a exigência de um percentual mínimo de aproveitamento pelo aluno para manter vínculo com programas de estágio contribui para manter o aluno no curso e polos próximos à instituição sede tendem a apresentar maiores índices de evasão.

Concluindo a análise, cabe considerar um aspecto relacionado ao curso de Pedagogia. De acordo com a coordenadora é elevado o número de alunos que se matriculam com o intuito de apenas manter o vínculo com a instituição, sendo que boa parte desses alunos, no decorrer do tempo, abandonam o curso. Esse fato tem relação direta com os índices obtidos nas simulações dos algoritmos de MD para os dados do curso, levando em consideração que as

simulações com esses dados apresentaram índices inferiores em relação aos demais¹¹. Isso ocorreu por dois motivos sendo que, em primeiro lugar, as métricas são calculadas com base nos alunos efetivamente evadidos o que foi baixo em relação a quantidade de alunos do curso. E, em segundo lugar, os algoritmos indicaram uma grande quantidade de alunos como evadidos. Esse grande número de previsões, apesar de prejudicar nas métricas, provavelmente decorre dos alunos que apenas mantêm o vínculo já que, dessa forma, eles não fazem uso do AVEA e, também, reprovam nas disciplinas matriculadas o que, consecutivamente, os enquadra nos parâmetros de outros que já evadiram. Desse modo, apesar das métricas não terem apresentado os melhores índices no que tange aos aspectos de qualidade da mineração, os alunos indicados pelos algoritmos de alguma forma necessitam de um acompanhamento por parte dos coordenadores por estarem enquadrados dentro do perfil de um aluno evadido, conforme abordados nas seções 7.2 e 7.3.

¹¹ Maiores detalhes podem ser obtidos na seção 7.3.5.

9 COMPARAÇÃO COM OS TRABALHOS CORRELATOS

Diante do exposto é possível avaliar as diferenças existentes entre a presente proposta e os trabalhos correlatos apresentados no capítulo 5. Para isso, no Quadro 19 é demonstrada a sintetização das principais características de cada abordagem.

Quadro 19 — Comparativo entre as características dos trabalhos correlatos

Autor	Bases de dados envolvidas na MD	Período de abrangência dos dados	Meio de divulgação da previsão
Kostopoulos, Sotiris e Panagiotis (2015)	Sistema de gestão acadêmico.	Sob demanda.	Ferramenta WEB
Rigo e Cazella (2014)	AVEA	18 meses.	Ferramenta WEB
Kantorski <i>et al.</i> (2016)	Sistema de gestão acadêmico.	Semestral	Relatórios
Manhães (2015)	Sistema de gestão acadêmico.	Semestral	Restringe-se na avaliação dos algoritmos.
Detoni, Araújo, Cechinel (2014)	AVEA	Semanal	Restringe-se na avaliação dos algoritmos.
Burgos <i>et al.</i> (2017)	AVEA	Semanal	Restringe-se na avaliação dos algoritmos.
Bayer <i>et al.</i> (2012)	Sistema de gestão acadêmico.	Semestral	Restringe-se na avaliação dos algoritmos.
Cássia, Brito e Medeiros (2016)	AVEA	Semestral	Restringe-se na avaliação dos algoritmos.
Schmitt	AVEA e Sistema de gestão acadêmico.	Semestral	Aplicativo para dispositivos móveis contendo previsões da MD.

Fonte: Autor.

De acordo com o Quadro 19, a primeira diferença condiz com os conjuntos de dados envolvidos na mineração. Com exceção da presente pesquisa, as demais empregaram dados oriundos de uma única base de dados. Nos trabalhos que abordaram o ensino presencial foram utilizados dados do sistema de gestão acadêmico. Já as pesquisas na EaD empregaram os dados oriundos do AVEA. O presente trabalho foi o único da relação que integrou dados de mais de uma fonte para posterior utilização na MD.

Sobre a abrangência dos dados verifica-se a predominância do período semestral. Este fator é de suma importância para predição, pois abrange as ações efetuadas durante o decorrer de todo o período em que o aluno está cursando uma disciplina. Isso evita a ocorrência de anomalia nos resultados que podem ocorrer nas abordagens efetuadas em períodos menores, devido as especificidades na elaboração de cada disciplina.

No item relacionado à disponibilização de dados, verifica-se que a pesquisa é uma das poucas a propiciar os dados provenientes da mineração. Dentre elas, ainda, é a única que disponibiliza os dados por meio de um aplicativo para dispositivos móveis. Esse fato pode ser o grande diferencial do quesito pela facilidade na adoção do aplicativo por parte dos interessados.

Por fim, outro ponto que não foi relacionado, mas é de suma importância, refere-se à generalização da abordagem proposta. Como pode ser visto no segundo experimento, na seção 7.3, a abordagem proposta também apresenta bons resultados quando utilizadas, em um único conjunto de dados, informações de todos os cursos analisados. Esse fato permite concluir que a abordagem pode ser generalizada para os demais cursos da modalidade EaD da instituição como, também, para outras instituições de ensino superior que empregam o AVEA MOODLE.

10 CONCLUSÃO

A evasão é um dos temas de grande relevância para gestão educacional em todos os níveis de ensino. Sua ocorrência não traz perdas apenas para o aluno desistente, mas, também, para todos os envolvidos direta ou indiretamente no processo educacional.

Na EaD o tema ainda é mais preocupante devido a algumas características específicas da modalidade, como, por exemplo, a falta de acompanhamento pessoal do aluno por parte do professor, o que torna ainda mais complicada a identificação de alunos com tendência à evasão. Por outro lado, conforme pode ser visto na RSL, desenvolvida no capítulo 3, a utilização dos AVEA para interação entre os participantes nessa modalidade de ensino proporciona um grande potencial para o emprego de aparatos tecnológicos para auxiliar nesse sentido.

Desse modo, utilizando informações de interações nos recursos do AVEA MOODLE, a presente pesquisa empregou técnicas de MD para prover informações de alunos com tendência à evasão. Também, de forma a facilitar o acesso e entendimento desses dados, foi desenvolvido um aplicativo para dispositivos móveis com algumas funcionalidades que possibilitam ao gestor visualizar os resultados da MD e realizar a comparação dos dados de alunos e polos de ensino.

O índice de acertos dos alunos que vieram a evadir nos períodos analisados, de maneira geral, atingiu 94%. Além disso, os alunos que foram previstos, mas que não evadiram, se enquadravam no comportamento semelhante ao de um aluno que já desistiu, o que pode levar à conclusão de que esse aluno tende a ser um futuro desistente. Dessa forma, a posse desses dados pode propiciar uma grande contribuição para as tomadas de decisões por parte dos gestores como, também, possibilitar a intervenção junto a esses alunos de modo a evitar a sua desistência.

As contribuições do trabalho podem ser destacadas nos dois produtos práticos elaborados, a abordagem de MD e o aplicativo móvel desenvolvido.

Na abordagem de MD as contribuições foram as seguintes:

- a) A elaboração de uma RSL com o levantamento das principais características dos trabalhos realizados;
- b) A integração de duas bases de dados para o processo de mineração – base de dados do sistema de gestão e do AVEA;
- c) Alta precisão no acerto de alunos evadidos (94%);
- d) Indicação de alunos que podem se tornar futuros desistentes; e,

- e) Possibilidade de generalização do processo para outros cursos de EaD, tanto na instituição em que o trabalho foi desenvolvido quanto em outras que empregam o AVEA MOODLE.

Com relação ao aplicativo desenvolvido como parte desta pesquisa, as seguintes contribuições podem ser destacadas:

- a) Disponibilização da lista de alunos com tendência à evasão oriunda do processo de MD;
- b) Disponibilização de dados relacionados aos índices de evasão por curso e polo de ensino – evasão geral e evasão anual média;
- c) Disponibilização de dados comparativos de utilização dos recursos do AVEA, empregados na mineração, entre os alunos que foram previstos como evadidos e os alunos concluintes do curso; e,
- d) Disponibilização de dados comparativos de utilização dos recursos do AVEA, empregados na MD, entre os alunos de um determinado polo de ensino e os que concluíram o curso.

Após o delineamento das contribuições é possível responder à questão estabelecida na pesquisa: “de que modo as informações oriundas do processo de MDE podem apoiar os gestores na criação de estratégias para mitigar o processo de evasão de alunos nos cursos de graduação a distância em uma IES pública? ”.

A descoberta automatizada de alunos com tendência à evasão por si só representa uma grande ajuda para gestão. Isso porque, como pode ser visto na avaliação do aplicativo, a tarefa de descoberta desses alunos por parte dos coordenadores é extremamente trabalhosa, sendo necessária a consulta de dados em diversos sistemas, bem como a obtenção de informações empíricas por parte dos tutores dos polos de ensino. Além disso, cabe ressaltar que de nada adianta somente a obtenção do conhecimento, se o mesmo não chegar de forma clara até o gestor. Dessa forma, para suprir essa lacuna, foi concebido o aplicativo para disponibilização dos dados da MD.

Em suma, as informações oriundas do processo de mineração reduzem a necessidade do gestor de busca dos alunos que possuem tendência à evasão, permitindo assim, direcionar os seus esforços na busca de alternativas para evitar a sua ocorrência.

Os objetivos específicos relacionados à compreensão das causas da evasão (a), revisão sobre o emprego das técnicas de MD para previsão da evasão (b) e identificação das

informações disponíveis no AVEA que podem influenciar na previsão (c) foram atendidos nos capítulos 2 e 3. A realização da previsão dos alunos (d) foi atendida no capítulo 7. Por fim, o desenvolvimento (e) e a avaliação da ferramenta (f) foram atendidos no capítulo 8.

Da mesma forma, o objetivo geral do trabalho, que consistiu em propiciar aos gestores educacionais dados estratégicos para que eles possam avaliar, refletir e gerar ações para mitigar o processo de evasão em cursos de graduação a distância foi atendido. Inicialmente, por meio da obtenção do conhecimento dos alunos com tendência à evasão, utilizando as técnicas de MD e, por fim, com a disponibilização desses dados através de um aplicativo móvel.

Ainda, comparando os trabalhos correlatos com a presente abordagem, esta tem a vantagem de utilizar dados de duas fontes de dados distintas (base de dados do AVEA MOODLE e do sistema de gestão acadêmico) e a possibilidade de generalização para os demais cursos de graduação da modalidade EaD da instituição. Já a limitação da abordagem proposta está relacionada com os dados do AVEA utilizados no processo de MD (dados de logs brutos) que não permitiu uma análise mais abrangente das informações como, por exemplo, do número de tarefas entregues em atraso por um aluno.

Por fim, o presente trabalho permitiu a elaboração de artigos, sendo que um já foi publicado e outros encontram-se em processo de submissão.

10.1 TRABALHOS FUTUROS

Como trabalhos futuros sugere-se, em termos da abordagem de MD, a sua utilização nos demais cursos de graduação da instituição e, do mesmo modo, em outros níveis de ensino como, por exemplo, pós-graduação. Além disso, sugere-se o emprego de outras informações existentes no AVEA MOODLE que permitam aprimorar os resultados da mineração, tais como as interações realizadas por tutores e docentes, quantidade de tarefas propostas em uma disciplina durante um semestre, número de tarefas entregues em atraso e dados relacionados à localização do polo e da cidade de residência do aluno. Outras fontes de dados também são sugeridas como, por exemplo, redes sociais. Por fim, ainda em relação à abordagem de MD, recomenda-se a automatização da execução do processo no final de cada semestre letivo para disponibilizar, de antemão, o conhecimento dos alunos que tendem a evadir para os gestores.

Ainda no que concerne ao aplicativo desenvolvido, sugere-se algumas melhorias que permitam a sua adoção no âmbito institucional como, por exemplo, a implementação de permissões de acesso. Ademais, possibilitar a inclusão de dados por parte do gestor, tais como as ações pedagógicas adotadas com o intuito de evitar a ocorrência da evasão, se a ocorrência

da evasão ocorreu de fato ou não, quais os motivos levaram a ocorrência da evasão, dentre outros. O detalhamento desses dados pelo gestor possibilita a criação de uma base de conhecimentos que pode ser útil para futuras tomadas de decisões. Por fim, recomenda-se a disposição dos dados de indicadores no âmbito geral dos polos.

11 ARTIGOS PUBLICADOS

SCHMITT, J. A.; BERNARDI, G.; KANTORSKI, G. Z. Abordagens tecnológicas para mineração de dados sobre evasão em cursos a distância: Uma Revisão Sistemática de Literatura. In: CONGRESSO BRASILEIRO DE EDUCAÇÃO A DISTÂNCIA. 14. 2017, Rio Grande. **Anais...** Rio Grande: SEaD, 2017. Disponível em: <http://www.aunirede.org.br/anais/arquivos/ANAIS_ESUD2017.pdf>. Acesso em: 1 nov. 2017.

REFERÊNCIAS

ABBAD, G. S.; ZERBINI, T.; SOUZA, D. B. L. Panorama das pesquisas em educação a distância no Brasil. **Estudos de Psicologia**, Natal, v. 15, n. 3, p. 291-298, set./dez. 2010.

ABRAHAMSSON, P. et al. Mobile-D: An Agile Approach for Mobile Application Development. In: ACM SIGPLAN CONFERENCE ON OBJECT-ORIENTED PROGRAMMING SYSTEMS, LANGUAGES AND APPLICATIONS. 19. 2004, Vancouver. **Anais...** Vancouver: ACM, 2004. Disponível em: <<https://dl.acm.org/citation.cfm?doid=1028664.1028736>>. Acesso em: 14 nov. 2017.

ABU-ODA, G. S.; EL-HALEES, A. M. Data Mining in Higher Education: University Student Dropout Case Study. **International Journal of Data Mining & Knowledge Management Process**, Gaza, v. 5, n. 1, p. 15-27, jan. 2015.

ADEODATO P. J. L. et al. Uma aplicação de mineração de dados à previsão no-show no agendamento de serviços médicos. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. 25. 2005, São Leopoldo. **Anais...** São Leopoldo: SBC, 2005. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/enia/2005/039.pdf>>. Acesso em: 20 out. 2017.

AL-SHARGABI, A.; NUSARI, A. Discovering vital patterns from UST students data by applying data mining techniques. **Journal of Intelligent Computing**. v. 1, n. 2, p. 88-99, jun. 2010.

ALVES, L. Educação a distância: conceitos e história no Brasil e no mundo. Revista Brasileira de Aprendizagem Aberta e a Distância. v. 10, n. 1, p. 83-92, 2011.

ARETIO, L. G. **Educación a distancia hoy**. 1. ed. Madrid: UNED. 1994. 642 p.

ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA. **Censo EAD.BR 2016 - Relatório Analítico da Aprendizagem a Distância no Brasil**. São Paulo, 2016.

_____. **Censo EAD.BR 2017 - Relatório Analítico da Aprendizagem a Distância no Brasil**. São Paulo, 2017.

AZOUMANA, K. Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas. **Revista Pensamiento Americano**, Medellín, v. 6, p. 41-51, 2014.

BARTOLOMEU, T. A. **Modelo de investigação de acidentes do trabalho baseado na aplicação de tecnologias de extração de conhecimento**. 2002. 302 p. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2002.

BASTOS, F. P.; ALBERTI, T. F.; MAZZARDO, M. M. Ambientes Virtuais de Ensino-Aprendizagem: Os desafio dos novos espaços de ensinar e aprender e suas novas implicações no contexto escolar. **Revista Novas Tecnologias na Educação – RENOTE**. Porto Alegre, v. 3, n. 1, mai. 2005.

BAYER, J. et al. Predicting drop-out from social behaviour of students. In: INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING, 5. 2012, Chania. **Anais...** Chania: [s.n.], 2012. Disponível em: <http://educationaldatamining.org/EDM2012/uploads/procs/EDM_2012_proceedings.pdf>. Acesso em: 8 ago. 2016.

BEHAR, P. A (Org.). **Competências em Educação a Distância**. Porto Alegre: Penso, 2013. 311 p.

BITTENCOURT, I. M.; MERCADO, L. P. L. Evasão nos cursos na modalidade de educação a distância: estudo de caso do Curso Piloto de Administração da UFAL/UAB. **Revista Ensaio: Avaliação de Políticas Públicas em Educação**. Rio de Janeiro, v. 22, n. 83, p. 465-504, abr./jun. 2014.

BIZARRIA, F. P. A.; SILVA, M. A.; CARNEIRO, T. C. J. Evasão discente na EAD: percepções do papel do tutor em uma instituição de ensino superior. In: CONGRESSO BRASILEIRO DE ENSINO SUPERIOR A DISTÂNCIA – ESUD. 11., 2014, Florianópolis. **Anais...** Florianópolis: UFSC, 2014. Disponível em: <<http://esud2014.nute.ufsc.br/anais-esud2014/>>. Acesso em: 20 set. 2016.

BRASIL. Decreto Nº 5.622, 2005. **Regulamenta o art. 80 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional**. Brasília, DF, 19 dez. 2005.

_____. Decreto n. 5800, de 8 de Junho de 2006. **Dispõe sobre o Sistema Universidade Aberta do Brasil – UAB**. Brasília, DF, 8 jun. 2006.

BURGOS, C. et al. Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. **Computers & Electrical Engineering**. p. 1-16, mar. 2017.

CABRAL, L. S.; SIEBRA, S. A. **Identificação de competências em currículos usando ontologias: uma abordagem teórica**. [S.l.: s.n.], 2006.

CAMARGO A. et al. Mineração de dados eleitorais: descoberta de padrões de candidatos a vereador na região da campanha do Rio Grande do Sul. **Revista Brasileira de Computação Aplicada**. Passo Fundo, v. 8, p. 64-73, abr. 2016.

CÁSSIA, J.; BRITO, A.; MEDEIROS, F. Aplicação de métricas da Análise de Redes Sociais como apoio a avaliação das interações discentes em fóruns de discussão online. In: WORKSHOPS DO CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. 5. 2016, Uberlândia. **Anais...** Uberlândia: SBC, 2016. Disponível em: <<http://br-ie.org/pub/index.php/wcbie>> Acesso em: 28 ago. 2017.

CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. 1. ed. São Paulo: Saraiva, 2016. 376 p.

CHAPMAN, P. et al. **CRISP-DM 1.0 Step-by-step data mining guide**. [S.l.]: [s.n.], 2000. 97 p.

CHEN H. et al. **Terrorism informatics: Knowledge Management and Data Mining for Homeland Security**. [S.l.]: Springer, 2008. 558 p.

COELHO, M. L. **A formação continuada do docente universitário em cursos a distância via internet: um estado de caso**. Monografia, 2002. ABED, Belo Horizonte, 1 ago. 2002. Disponível em: <<http://www.abed.org.br/seminario2003/texto06.htm>>.

COMISSÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Página da Universidade Aberta do Brasil**. Disponível em: <<http://www.capes.gov.br/uab>>. Acesso em: 04 set. 2017.

COMMARELLA, R. F. **Educação superior a distância: evasão discente**. 2009. 146 p. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina, Florianópolis, 2009.

CORREIA, R. A. P. **Introdução à educação a distância**. São Paulo: Cengage Learning, 2016. 72 p.

COSTA, E. et al. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. In: Jornada de Atualização em Informática na Educação. 23. Rio de Janeiro, RJ: UFRJ. cap. 1. 2013. Disponível em: <<http://www.br-ie.org/pub/index.php/pie/article/view/2341>>. Acesso em: 23 nov. 2017

COUSSEMENT, K.; VAN DEN POEL, D. Integrating the voice of customers through call center emails into a decision support system for churn prediction. **Journal Information and Management**. Amsterdam, v. 45, n. 3, p. 164-174, abr. 2008.

DAMÁSIO, M. M. Educação Presencial x Educação a Distância: reflexões e considerações. TE em Revista, v. 2, p. 75-103, jan/dez. 2008.

DELGADO, M. F. et al. Learning analytics for the prediction of the educational objectives achievement. In: FRONTIERS IN EDUCATION CONFERENCE (FIE) PROCEEDINGS. 2014, Madrid. **Anais...** Madrid: IEEE, 2014. Disponível em: <<http://ieeexplore.ieee.org/document/7044402/>>. Acesso em: 28 ago. 2017.

DETONI, D; ARAÚJO, R.; CECHINEL, C. Modelling and Prediction of Distance Learning Students Failure by using the Count of Interactions. **Revista Brasileira de Informática na Educação**, v. 23, p. 1-11, 2015.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**. Providence, v. 17, n. 3, p. 37-54, jul. 1996.

FERREIRA, J. B. **Mineração de dados na retenção de clientes em telefonia celular**. 2005. 87 p. Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

FERREIRA, V. S.; ELIA, M. F. Uma modelagem conceitual para apoiar a identificação das causas da evasão escolar em EAD. In: WORKSHOP DE INFORMÁTICA NA ESCOLA - WIE. 19. 2013, São Paulo. **Anais...** São Paulo: UNICAMP, 2013. Disponível em <<http://www.br-ie.org/pub/index.php/wie/issue/view/80>>. Acesso em: 13 jul. 2017.

GERHARDT, T. E.; SILVEIRA, D. T. (Org.). **Métodos de Pesquisa**. 1. ed. Porto Alegre: Editora da UFRGS, 2009. 120p.

GOLDSCHMIDT, R.; BEZERRA, E.; PASSOS, E. **Data mining: conceitos, técnicas, algoritmos orientações e aplicações**. 2. ed. Rio de Janeiro: Elsevier, 2015. 296 p.

GOLDSCHMIDT, R.; PASSOS, E. **Data Mining: Um Guia Prático**. Rio de Janeiro: Campus, 2005. 262 p.

HALL, M et al. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**. New York, v. 11, n. 1, p. 10-18, jun. 2009.

HAN, J.; KAMBER, M. PEI, J. **Data Mining Concepts and Techniques**. 3. ed. Waltham: Elsevier, 2012.

HEIDRICH, L. et al. Diagnóstico do Comportamento dos Aprendizes na Educação a Distância com Base no Estilo de Aprendizagem. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. 16. 2007, Dourados. **Anais...** Dourados: 2007. Disponível em <<http://www.br-ie.org/pub/index.php/sbie/issue/view/27>>. Acesso em: 21 out. 2017.

HOFFMANN, I. L. **Metodologia para identificação de fatores estratégicos para acompanhamento sistemático da evasão em cursos de graduação**. 2016. 111 p. Dissertação (Mestrado em Engenharia da Produção) - Universidade Federal de Santa Maria, Santa Maria, 2016.

JADRIĆ, M.; GARAČA, Ž.; ČUKUŠIĆ, M. Student dropout analysis with application of data mining methods. In: INTERNATIONAL CONFERENCE ON TELECOMMUNICATIONS. 1. 2010, Zagreb. **Anais...** Zagreb: 2010. Disponível em: <<https://pdfs.semanticscholar.org/455e/0fff1275a936e062187d1ffdc80b89695151.pdf>>. Acesso em: 12 fev. 2017.

JALAL, M. E.; HOSSEINI, M.; KARLSSON, S. Forecasting incoming call volumes in call centers with recurrent Neural Networks. **Journal of Business Research**. [S.l.], v. 69, n. 11, p. 4811-4814, fev. 2016.

KANTORSKI, Z. G. et al. Uma Abordagem para Previsão de Evasão em Cursos de Graduação Presenciais. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE. 27. 2016, Uberlândia. **Anais...** Uberlândia: 2016. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/issue/view/155>>. Acesso em: 30 dez. 2016.

KITCHENHAM, B. et al. Systematic literature reviews in software engineering - A tertiary study. **Information and Software Technology**. [S.l.], v. 52, n. 1, p. 792-805, nov. 2010.

KIPPES, A. **Mineração de dados e análise do comportamento do consumidor: estudo de caso em um website de ensino a distância**. 2010. 117 p. Dissertação (Mestrado em Administração) - Universidade Federal de Lavras, Lavras, 2010.

- KISAHLEITNER, M. **Análise de técnicas de data mining na aquisição de clientes de cartão de crédito não correntistas**. 2008. 93 p. Dissertação (Mestrado em Administração) - Escola de Administração de Empresas de São Paulo, São Paulo, 2008.
- KOSTOPOULOS, G.; SOTIRIS, K.; PANAGIOTIS, P. Estimating student dropout in distance higher education using semi-supervised techniques. In: PANHELLENIC CONFERENCE ON INFORMATICS. 19. 2015, Athens. **Anais...** Athens: ACM, 2015. Disponível em: <<https://dl.acm.org/citation.cfm?id=2802013>>. Acesso em: 1 nov. 2017.
- LITTO, F. M.; FORMIGA, M. (Org.). **Educação a Distância: o estado da arte**. 2. ed. São Paulo: Pearson Education, v. 2, 2012. 456 p.
- LOBO, M. B. de C. M. Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. **ABMES Cadernos**. Brasília, set. / dez. 2012.
- MACHADO, D. P.; MORAES, M. G. S. **Educação a distância: fundamentos, tecnologias, estrutura e processo de ensino e aprendizagem**. 1. ed. São Paulo: Editora Érica. 2015. 120 p.
- MANHÃES, L. M. B. **Predição do desempenho acadêmico de graduandos utilizando mineração de dados educacionais**. 2015. 157 p. Tese (Doutorado em Engenharia de Sistemas e Computação) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Rio de Janeiro, 2015.
- MARTINS, C. Z.; GEBRAN, R. A.; TERÇARIOL, A. A. L. A evasão na perspectiva dos alunos: Uma análise no curso de administração a distância. In: SIMPÓSIO INTERNACIONAL DE EDUCAÇÃO A DISTÂNCIA – SIED. 2. 2014, São Carlos. **Anais...** São Carlos: SEaD, 2014. Disponível em: <<http://www.sied-enped2014.ead.ufscar.br/ojs/index.php/2014/pages/view/anais>>. Acesso em: 10 out. 2017.
- MARTINS, L. C.; LOPES, D. A.; RAABE, A. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE. 22. 2011, Aracaju. **Anais...** Aracaju: UFS, 2011. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/article/view/1585>>. Acesso em: 1 mar. 2017.
- MARTÍNEZ, M.; GARCÍA M. C.; MONTORO, J. M. **Dificuldades de aprendizagem**. 1. ed. Porto: Porto editora, 2003. 56 p.
- MATTAR, J. **Tutoria e interação em educação a distância**. São Paulo: Cengage Learning, 2012. 240 p.
- MCCUE, C. **Data Mining and Predictive Analysis: Intelligence Gathering and Crime**. 1. ed. Burlington: Elsevier, 2007. 368 p.
- MOROSINI, M. C. et al. A Evasão na Educação Superior no Brasil: uma análise da produção de conhecimento nos periódicos Qualis entre 2000-2011. In: CONFERENCIA LATINOAMERICANA SOBRE EL ABANDONO EN LÁ EDUCACIÓN SUPERIOR. 1. 2012, Managua. **Anais...** Managua: FADU, 2012. Disponível em: <<http://repositorio.pucrs.br/dspace/handle/10923/8762>>. Acesso em: 23 out. 2017.

NUNES, D. J. V. **Mineração de dados em comércio varejista para diminuição dos níveis de estoque**. 2009. 87 p. Dissertação (Mestrado em Engenharia da Produção) - Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, 2009.

OREA, S. V.; VARGAS, A. S.; ALONSO, M. G. Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. **Ene**. [S.l.], v. 779, p. 1-33, 2005.

PALAZUELOS, C.; GARCÍA-SAIZ, D.; ZORRILLA, M. Predicting Dropout-Prone Students in E-Learning Education System. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL COLLECTIVE INTELLIGENCE. 5. 2013, Craiova. **Anais...** Craiova: Springer Berlin Heidelberg, 2013. Disponível em: <http://software.ucv.ro/ICCCI2013/accepted_papers.html>. Acesso em: 20 out. 2017.

PETERMANN, R. J. **Modelo de mineração de dados para classificação de clientes em telecomunicações**. 2006. 164 p. Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2006.

PRASS, F. S. **Estudo comparativo entre algoritmos de análise de agrupamentos**. 2004. 71 p. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Santa Catarina, Florianópolis, 2004.

PREECE, J.; ROGERS, Y.; SHARP, H. **Interaction design: beyond human**. New York: John Wiley & Sons, 2013. 519 p.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013. 277 p.

QUEIROGA, E. M. et al. Generating models to predict at-risk students in technical e-learning courses. In: LATIN AMERICAN CONFERENCE ON LEARNING OBJECTS AND TECHNOLOGY – LACLO. 11. 2016, San Carlos. **Anais...** San Carlos: IEEE, 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7751770/>>. Acesso em: 1 nov. 2017.

REINO, L. R. A. C. et al. Análise das Causas da Evasão na Educação a Distância em uma Instituição Federal de Ensino Superior. In: Simpósio Brasileiro de Informática na Educação. 26. 2015, Maceió. **Anais...** Maceió: UFAL, 2015. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/issue/view/129>>. Acesso em: 13 out. 2017.

RIGO, S. J.; CAZELLA, S. C. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação – RBIE**. Dourados, v. 22, p. 132-146, 2014.

SALES, G. L.; LEITE, E. A. M.; JOYE, C. R. Gerenciamento da aprendizagem, Evasão em EAD online e possíveis soluções: Um Estudo de Caso no IFCE. **Revista Notas Tecnológicas em Educação – RENOTE**. Porto Alegre, v. 10, n. 3, dez. 2012.

SCALI, D. F. **Evasão nos Cursos Superiores de Tecnologia: a Percepção dos Estudantes sobre seus Determinantes**. 2009. 140 p. Dissertação (Mestrado em Educação) – Universidade Estadual de Campinas, Campinas, 2009.

SHANNON, G.; BYLSMA, P. Helping Students Finish School: Why Students Drop Out and How to Help Them Graduate. **Office of Superintendent of Public Instruction**, Olympia, 2006. 104 p.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. Sistema de Banco de Dados. 5. ed. Rio de Janeiro: Elsevier, 2006. 791 p.

SILVA, R. S. **Moodle para autores e tutores: educação a distância na web 2.0**. 1. ed. São Paulo: Novatec, 2010. 152 p.

SILVA FILHO, R. L. L. et al. A Evasão no Ensino Superior Brasileiro. **Cadernos de Pesquisa**. [S.i], v. 37, n. 132, p. 641-659, set./dez. 2007.

SILVEIRA, C. A. B. Educação a Distância e a Evasão: Estudo de Caso da Realidade no Polo UAB de Franca. In: SIMPÓSIO INTERNACIONAL DE EDUCAÇÃO A DISTÂNCIA. 2012, São Carlos. **Anais...** São Carlos: UFSCAR, 2012. Disponível em: <<http://sistemas3.sead.ufscar.br/ojs/index.php/sied/pages/view/Anais>> Acesso em: 31 jul 2017.

SOUSA, M. M.; FIGUEIREDO, R. S. Análise de crédito por meio de mineração de dados: Aplicação em cooperativa de crédito. **Revista de Gestão da Tecnologia e Sistemas de Informação**. São Paulo, v. 11, n. 2, p. 379-396, mai./ago. 2014.

TAVARES, C.; BOZZA, D.; KONO, F. Descoberta de Conhecimento Aplicado a Dados Eleitorais. **Revista Gestão e Conhecimento**. Curitiba, v. 5, n. 1, p. 54-94, jan./jun. 2007.

TINTO, V. Dropout from higher education: a theoretical synthesis of recent research. **Review of Educational Research**. p. 89-125. 1975.

TRESMAN, S. Towards a Strategy for Improved Studenty Retention in Programmers of Open, Distance Education: A Case Study from the Open University UK. **The International Review of Research in Open and Distributed Learning**. Athabasca, v. 3 n. 1, abr. 2002.

UNIVERSIDADE FEDERAL DE SANTA DE SANTA MARIA. Pró-Reitoria de Planejamento. **Plano de Desenvolvimento Institucional da UFSM/2016-2026**, Santa Maria, 2016.

VIEIRA, A. T. ALMEIDA, M. E. B.; ALONSO, M. (Org.). **Gestão educacional e tecnologia**. São Paulo: Avercamp, 2003.

WALTER, A. M. **Variáveis Predictoras de evasão em cursos a distância**. 2006. 177 p. Dissertação (Mestrado em Psicologia) – Universidade de Brasília Mestrado, Brasília, Distrito Federal, 2006.

WEISS, S.; INDURKHYA, N. **Predictive Data Mining: A Practical Guide**. 1. ed. São Francisco: Morgan Kaufmann Publishers, 1998. 228p.

WONG, G. K. W.; SIMON, L. Academic Performance Prediction Using Change Discovery from Online Discussion Forums. In: COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE – COMPSAC. 40. 2016, Atlanta. **Anais...** Atlanta: IEEE, 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7552094/>>. Acesso em: 23 jun. 2017.

ZAFRA, A.; CRISTÓBAL, R.; SEBATIÁN, V. Predicting academic achievement using multiple instance genetic programming. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS DESIGN AND APPLICATIONS. 9. 2009, Pisa. **Anais...** Pisa: IEEE, 2009. Disponível em: <<http://ieeexplore.ieee.org/document/5364212/>>. Acesso em: 29 out. 2017.

ZAKI, M. J. **Parallel and Distributed Data Mining: An Introduction**. Berlin: Springer-Verlag, 2000. 260 p.

ZHENG, J.; ZEYU, C.; CHANGJUN, Z. Applying NN-based data mining to learning performance assessment. In: CONFERENCE ANTHOLOGY. 2013, China. **Anais...** China: IEEE, 2013. Disponível em: <<http://ieeexplore.ieee.org/document/6784924/>>. Acesso em 15 ago. 2017.

APENDICE A – TRABALHOS SELECIONADOS PÓS-LEITURA

Ident.	Trabalho	Critério
A1	WONG, G. K. W.; SIMON L. Academic Performance Prediction Using Chance Discovery from Online Discussion Forums . In: COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE (COMPSAC), 2016 IEEE 40th Annual. v. 1, 2016.	I2
A2	RATNAPALA, I. P.; RAGEL, R. G.; DEEGALLA, S. Students be havioural analysis in an online learning environment using data mining . In: INTERNATIONAL CONFERENCE ON INFORMATION AND AUTOMATION FOR SUSTAINABILITY. v. 7, 2014.	I2
A3	ZHOU, Q.; YOUJIE, Z.; CHAO, M. Predicting students' performance of an offline course from their online behaviors . In: DIGITAL INFORMATION AND COMMUNICATION TECHNOLOGY AND ITS APPLICATIONS (DICTAP). v. 5, 2015.	I1
A4	PRADEEP, A.; SMIJA, D.; JUBILANT, J. K. Students dropout factor prediction using EDM techniques . In: SOFT-COMPUTING AND NETWORKS SECURITY (ICSNS). 2015.	I1
A5	NETO, F. A.; CASTRO, A. Elicited and mined rules for dropout prevention in online courses . In: FRONTIERS IN EDUCATION CONFERENCE (FIE). 2015.	I2
A6	BYDOVSKA, H.; LUBOMÍR, P. Predicting Student Performance in Higher Education . In: INTERNATIONAL WORKSHOP ON DATABASE AND EXPERT SYSTEMS APPLICATIONS. v. 24, 2013.	I2
A7	RIGO, S. J. et al. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. Revista Brasileira de Informática na Educação . v. 11, n. 2, 2014.	I2
A8	CECHINEL, C.; ARAUJO, R. M.; DETONI, D. Modelagem e Predição de Reprovação de Acadêmicos de Cursos de Educação a Distância a partir da Contagem de Interações. Revista Brasileira de Informática na Educação . v. 12, n. 3, 2015.	I2
A9	WILGES, B. et al. Sistemas multiagentes: mapeando a evasão na educação a distância. Revista Novas Tecnologias em Educação . v. 8, n.1, 2010.	I2
A10	SALES, G. L.; LEITE, E. A. M.; JOYE, C. R. Gerenciamento da Aprendizagem, Evasão em EaD Online e Possíveis Soluções: Um Estudo de Caso no IFCE. Revista Novas Tecnologias em Educação . v. 10, n. 3, 2012.	I2
A11	COSTA, S. S. et al. Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS. Revista Novas Tecnologias em Educação . v. 12, n. 2, 2012.	I2
A12	MANHÃES, L. M. B. et al. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados . In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 1, n. 1, 2011.	I1
A13	MARTINS, L. C.; LOPES, D. A.; RAABE, A. Um Assistente de Predição de Evasão aplicado a uma disciplina Introdutória do curso de Ciência da Computação . In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 23, n. 1, 2012.	I2
A14	BRITO, D. M. et al. Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina . In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 25, n. 1, 2014.	I1

A15	SILVA, F. et al. Um modelo preditivo para diagnóstico de evasão baseado nas interações de alunos em fóruns de discussão. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 26, n. 1, 2015.	I2
A16	SANTOS, R. N.; SIEBRA, C. A.; OLIVEIRA, E. D. S. Uma Abordagem Temporal para Identificação Precoce de Estudantes de Graduação a Distância com Risco de Evasão em um AVA utilizando Árvores de Decisão. In: WORKSHOPS DO CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 3, n. 1, 2014.	I2
A17	OLIVEIRA, J. et al. Análise da Correlação da Evasão de Cursos de Graduação com o Empréstimo de Livros em Biblioteca. In: WORKSHOPS DO CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 3, n. 1, 2014.	I1
A18	GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados. In: WORKSHOP DE DESAFIOS DA COMPUTAÇÃO APLICADA À EDUCAÇÃO. 2012.	I2
A19	HOE, A. C. K. et al. Analyzing students records to identify patterns of student's performance. In: INTERNATIONAL CONFERENCE ON RESEARCH AND INNOVATION IN INFORMATION SYSTEMS (ICRIIS). 2013.	I1
A20	NASIRI, M.; MINAEI, B. Predicting GPA and academic dismissal in LMS using educational data mining: A case mining. In: INTERNATIONAL CONFERENCE OF E-LEARNING AND E-TEACHING. 2012.	I1
A21	FUN, L. K.; RUSK, D.; SONG, F. Predicting student academic performance. In: COMPLEX, INTELLIGENT, AND SOFTWARE INTENSIVE SYSTEMS (CISIS). 2013.	I1
A22	ZHANG, Y. et al. Withdrawal prediction using the blackboard learning management system through SOM. In: SOFTWARE ENGINEERING AND DATA MINING (SEDM). 2010.	I1
A23	AL-SHARGABI, A. A.; NUSARI, A. N. Discovering vital patterns from UST students data by applying data mining techniques. In: COMPUTER AND AUTOMATION ENGINEERING (ICCAE). v. 2. 2010.	I1
A24	BUNYAMIN, N., MAT, U. B.; ARSHAD, P. M. Educational data mining for prediction and classification of engineering students achievement. In: INTERNATIONAL CONFERENCE ON ENGINEERING EDUCATION (ICEED). 2015.	I1
A25	ALHARBI, Z. et al. Using data mining techniques to predict students at risk of poor performance. In: SAI COMPUTING CONFERENCE. 2016.	I1
A26	AMRIEH, E. A.; THAIR, H.; IBRAHIM, A. Preprocessing and analyzing educational data set using X-API for improving student's performance. In: APPLIED ELECTRICAL ENGINEERING AND COMPUTING TECHNOLOGIES (AEECT). 2015.	I2
A27	DELGADO, M. F. et al. Learning analytics for the prediction of the educational objectives achievement. In: FRONTIERS IN EDUCATION CONFERENCE (FIE). 2014.	I1
A28	ZHENG, J.; ZEYU, C.; CHANGJUN, Z. Applying NN-based data mining to learning performance assessment. In: CONFERENCE ANTHOLOGY. 2013.	I2
A29	ZHANG, Z. Study and analysis of data mining technology in college courses students failed. In: INTELLIGENT COMPUTING AND INTEGRATED SYSTEMS (ICISS). 2010.	I1
A30	OGOR, E. N. Student academic performance monitoring and evaluation using data mining techniques. In: ELECTRONICS, ROBOTICS AND AUTOMOTIVE MECHANICS CONFERENCE (CERMA 2007). 2007.	I2

A31	SOROUR, S. E. et al. Predicting students' grades based on free style comments data by artificial neural network. In: FRONTIERS IN EDUCATION CONFERENCE (FIE) PROCEEDINGS. 2014.	I1
A32	ZAFRA, A.; CRISTÓBAL, R.; SEBATIÁN, V. Predicting academic achievement using multiple instance genetic programming. In: CONFERENCE ON INTELLIGENT SYSTEMS DESIGN AND APPLICATIONS. 2009.	I2
A33	PALAZUELOS, C.; GARCÍA-SAIZ, D.; ZORRILLA, M. Social network analysis and data mining: an application to the e-learning context. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL COLLECTIVE INTELLIGENCE. 2013.	I2
A34	DEWAN, M. A. A.; FUHUA, L.; DUNWEI, W. Predicting Dropout-Prone Students in E-Learning Education System. In: UBIQUITOUS INTELLIGENCE AND COMPUTING. 2015.	I2
A35	KOSTOPOULOS, G.; SOTIRIS, K.; PANAGIOTIS, P. Estimating student dropout in distance higher education using semi-supervised techniques. In: PANHELLENIC CONFERENCE ON INFORMATICS. 2015.	I2
A36	BARBOSA, M. et al. Towards automatic prediction of student performance in STEM undergraduate degree programs. In: SYMPOSIUM ON APPLIED COMPUTING. 2015.	I1
A37	AGUIAR, E. et al. Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. In: INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE. 2014.	I1
A38	SIMJANOSKA, M.; MARJAN, G.; BOGDANOVA, A. M. Intelligent modelling for predicting students' final grades. In: INFORMATION AND COMMUNICATION TECHNOLOGY, ELECTRONICS AND MICROELECTRONICS (MIPRO). 2014.	I2
A39	MUSTAFA, M. N.; LINKON C.; KAMAL, M. S. Students dropout prediction for intelligent system from tertiary level in developing country. In: INFORMATICS, ELECTRONICS & VISION (ICIEV). 2012.	I1
A40	COSTA, F. et al. Predição de sucesso de estudantes cotistas utilizando algoritmos de classificação. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 26, n. 1, 2015.	I1
A41	SILVA, R. et al. Mineração de dados educacionais na análise das interações dos alunos em um Ambiente Virtual de Aprendizagem. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 26, n. 1, 2015.	I2
A42	SANTANA, L. C.; MACIEL, A. M. A.; RODRIGUES, R. L. Avaliação do Perfil de Uso no Ambiente Moodle Utilizando Técnicas de Mineração de Dados. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 25, n. 1, 2014.	I2
A43	SANTOS, M. S. et al. Mining retention rules from student transcripts: A case study of the programs at a federal university. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 25, n. 1, 2014.	I1
A44	KANTORSKI, G. et al. Predição da Evasão em Cursos de Graduação em Instituições Públicas. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 27, n. 1, 2016.	I1
A45	PASCOAL, T. et al. Evasão de estudantes universitários: diagnóstico a partir de dados acadêmicos e socioeconômicos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 27, n. 1, 2016.	I1

A46	AMORIM, M. J. V.; BARONE, D.; MANSUR, A. U. Técnicas de aprendizado de máquina aplicadas na previsão de evasão acadêmica. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 1, n. 1, 2008.	I1
A47	RAMOS, J. L. C. et al. Um Modelo Preditivo da Evasão dos Alunos na EAD a partir dos Construtos da Teoria da Distância Transacional. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 28, n. 1, 2017.	I2
A48	CALIXTO, K. E.; SEGUNDO, C. V.N.; GUSMÃO, R. P. Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar. In: Anais do Simpósio Brasileiro de Informática na Educação. v. 28, n. 1, 2017.	I1
A49	PAZ, F. J.; CAZELLA, S. C. Identificando o perfil de evasão de alunos de graduação através da Mineração de Dados Educacionais: um estudo de caso de uma Universidade Comunitária. In: WORKSHOPS CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 6, n. 1, 2017.	I1
A50	FERNANDES, W. L. et al. Previsão de Desempenho de Estudantes usando o Algoritmo de Classificação Associativa. In: WORKSHOPS DO CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. v. 6, n. 1, 2017.	I2
A51	AMERI, S. et al. Survival Analysis based Framework for Early Prediction of Students Dropouts. In: PROCEEDINGS OF ACM INTERNATIONAL ON CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT. v. 25, n. 1, 2016.	I1
A52	FIGUEIRA, A. Mining Moodle Logs for Grade Prediction: A methodology walk-through. In: PROCEEDINGS OF INTERNATIONAL CONFERENCE ON TECHNOLOGICAL ECOSYSTEMS FOR ENHANCING MULTICULTURALITY. v. 5, n. 1, 2017.	I2
A53	KITANA, Y.; TAKEUCHI, K.; HIROKAWA, S. Predicting Learning Result of Learner in E-learning Course with Feature Selection Using SVM. In: PROCEEDINGS OF INTERNATIONAL CONFERENCE ON EDUCATION TECHNOLOGY AND COMPUTERS. v. 9, n. 1, 2017.	I2
A54	TANABE, Y. et al. Finding Key Integer Values in Many Features for Learner's Academic Performance Prediction. In: PROCEEDINGS OF INTERNATIONAL CONFERENCE ON EDUCATION TECHNOLOGY AND COMPUTERS. v. 9, n. 1, 2017.	I2
A55	DEVASIA, T.; VINUSHREE, T.; HEGDE, V. Prediction of Students Performance using Educational Data Mining. In: INTERNATIONAL CONFERENCE ON DATA MINING AND ADVANCED COMPUTING (SAPIENCE). 2016.	I1
A56	HASBUN, T.; ARAYA, A.; VILLALON, J. Extracurricular activities as dropout prediction factors in higher education using decision trees. In: INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES. v. 16, n. 1, 2016.	I1
A57	MAYA, N. E. et al. Data Mining: A Scholar Dropout Predictive Model. In: Mexican Humanitarian Technology Conference. 2017.	I1
A58	SOUNG, J.; WOO, Y.; PARK, S. Mining Course Trajectories of Successful and Failure Students: A Case Study. In: INTERNACIONAL CONFERENCE ON BIG KNOWLEDGE. 2017.	I1
A59	PEREIRA, R.; ZAMBRANO, J. Application of Decision Trees for Detection of Student Dropout Profiles. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS. v. 16, n. 1, 2017.	I1
A60	MARTINS, L. et al. Early prediction of college attrition using data mining. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS. v. 16, n. 1, 2017.	I1

A61	GUSTIAN, D.; HUNDAYANI, R. Combination of AHP Method With C4.5 in The Level Classification Level Out Students. In: INTERNATIONAL CONFERENCE ON COMPUTING, ENGINEERING, AND DESIGN (ICCED). 2017.	I2
A62	AHUJA, R.; KANKANE, Y. Predicting the Probability of Student's Degree Completion by Using Different Data Mining Techniques. In: INTERNATIONAL CONFERENCE ON IMAGE INFORMATION PROCESSING (ICIIP). v. 4, n. 1, 2017.	I1
A63	KOSTOPOULOS, G. et al. Early Dropout Prediction in Distance Higher Education Using Active Learning. In: INTERNATIONAL CONFERENCE ON INFORMATION, INTELLIGENCE, SYSTEMS & APPLICATIONS (IISA). v. 8, n. 1, 2017.	I2
A64	BADR, G. et al. Predicting Students' Performance in University Courses: A Case Study and Tool in KSU Mathematics Department. In: SYMPOSIUM ON DATA MINING APPLICATIONS (SDMA). 2016.	I1
A65	DIGIAMPIETRI, L.; NAKANO, F.; LAURETTO, M. Mineração de Dados para Identificação de Alunos com Alto Risco de Evasão: Um Estudo de Caso. Revista de Graduação USP. v. 1, n. 1, 2016.	I1
A66	CHAUDHARI, K. et al. Student Performance Prediction System using Data Mining Approach. In: INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING. v. 6, n. 3, 2017.	I1
A67	MEHBOOB, B.; LIAQAT, R.; ABBAS, N. Student Performance Prediction and Risk Analysis by Using Data Mining Approach. Journal of Intelligent Computing. v. 8, n. 2, 2017.	I1

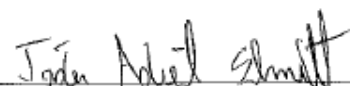
ANEXO A – AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS ACADÊMICOS DO SISTEMA DE INFORMAÇÃO INSTITUCIONAL

UNIVERSIDADE FEDERAL DE SANTA MARIA

À PROGRAD E CPD

SOLICITAÇÃO

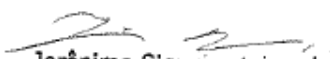
Eu, JADER ADIEL SCHMITT, aluno do Curso de Mestrado Profissional em Tecnologias Educacionais em Rede, da UFSM e Analista de TI do CPD/UFSM, venho solicitar ao CPD o acesso aos dados acadêmicos dos alunos de graduação da modalidade EAD que constam no SIE, os quais são necessários para elaboração do meu projeto de pesquisa e complementação da minha dissertação.


Jader Adiel Schmitt

De acordo. Fornecer os dados pessoais
como endereço, telefone não
será permitido.

De acordo,


Gustavo Chiapinotto da Silva
Diretor do CPD - UFSM


Jerônimo Siqueira Lybusch
Pró-Reitor Substituto/PROGRAD
Portaria n. 87.324/2018
SIAPE 3579368

ANEXO B – AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS DE INTERAÇÕES DO MOODLE – LETRAS ESPANHOL

Assunto Re: Permissão de Utilização dados Moodle
Remetente Espanhol UAB-UFSM EaD <espanholead@gmail.com>
Para Alexandre Schlöttgen <alexandre@nte.ufsm.br>, jaschmitt <jaschmitt@cpd.ufsm.br>, <bgiliane@gmail.com>
Data 2017-11-14 14:39



UNIVERSIDADE FEDERAL
DE SANTA MARIA

Alexandre,
autorizo o uso dos dados do curso de Espanhol e peço que o trabalho seja apresentado à comunidade, após a defesa.
Se for o caso, podemos, inclusive, agendar uma reunião com os professores do curso para a apresentação do trabalho.
Att.
Vanessa

Coordenação do Curso de Letras Espanhol a distância da UFSM

Coordenadora do Curso

Vanessa Ribas Fialho

Coordenador de Tutoria

Marcus Fontana

Coordenadora de Estágio

Daniele Bauermann Quinhones

Bolsista da Secretaria

Maria Clara Xavier

Em 14 de novembro de 2017 14:49, Alexandre Schlöttgen <alexandre@nte.ufsm.br> escreveu:

Olá Coordenações de Curso,

o servidor da UFSM - **Jáder Adiel Schmitt**, lotado no CPD, necessita de acesso aos dados do moodle UAB de alguns cursos para sua dissertação.

Não vemos nenhum problema em disponibilizar estes dados, que podem auxiliar futuramente o NTE e a UFSM a entender o comportamento dos estudantes e com isso tentar prever e diminuir a evasão dos alunos EaD. Mas precisamos da **autorização das coordenações de cursos**.

Esperamos retorno assim que possível, deste e-mail.

Atenciosamente,
Alexandre.



NÚCLEO
DE TECNOLOGIA
EDUCACIONAL
UFSM

.....
Alexandre Schlöttgen

Analista de Tecnologia da Informação

Núcleo de Tecnologia Educacional - NTE

Universidade Federal de Santa Maria - UFSM


ANEXO C – AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS DE INTERAÇÕES DO MOODLE – PEDAGOGIA

<p>Assunto Fwd: Retorno sobre disponibilização de dados para Mestrado</p> <p>Remetente Alexandre Schlöttgen <alexandre@nte.ufsm.br></p> <p>Para Jader Schmitt <jaschmitt@cpd.ufsm.br></p> <p>Data 2018-03-19 16:41</p>	 <p>UNIVERSIDADE FEDERAL DE SANTA MARIA</p>
--	---

Olá Jader,

abaixo a resposta da Coordenadora do curso de **Pedagogia EAD**, liberando acesso aos dados do curso de Pedagogia.

Att,
Alexandre.



.....

Alexandre Schlöttgen

Analista de Tecnologia da Informação

Núcleo de Tecnologia Educacional - NTE

Universidade Federal de Santa Maria - UFSM

----- Mensagem encaminhada -----

De: **Andréa Forgiarini Cechin** <afcechin@gmail.com>

Data: 24 de janeiro de 2018 12:34

Assunto: Re: Retorno sobre disponibilização de dados para Mestrado

Para: Alexandre Schlöttgen <alexandre@nte.ufsm.br>

Então tudo certo. Pode liberar os dados para o rapaz. Porque se precisasse verificar as disciplinas teríamos que solicitar autorização dos professores.

Um abraço

Andréa Forgiarini Cechin
Enviado de meu iPhone

Em 24 de jan de 2018, à(s) 09:50, Alexandre Schlöttgen <alexandre@nte.ufsm.br> escreveu:

Olá Professora Andréa,

o mestrando não ganhará nenhum acesso para se logar no sistema moodle, mas receberá **arquivos no formato de planilha eletrônica** com a matrícula do aluno e **quantidade de acessos** nos recursos e atividades do Moodle, conforme exemplo do arquivo em anexo.

Com isso, ele consegue ver **somente quantas vezes um aluno acessou** os fóruns de uma disciplina, quantas vezes visualizou uma página e assim por diante, por semestre, mês e disciplina.

Mas ele não saberá o conteúdo, não verá os textos das postagens, etc.

Att,
Alexandre.

Alexandre Schlöttgen

Analista de Tecnologia da Informação

Núcleo de Tecnologia Educacional - NTE

Universidade Federal de Santa Maria - UFSM

Em 23 de janeiro de 2018 14:56, Andréa Forgiarini Cechin <afcechin@gmail.com> escreveu:

Oi Alexandre, tudo bem?

Eu gostaria de saber a que dados ele quer acesso.

Andréa Forgiarini Cechin

Enviado de meu iPhone

Em 23 de jan de 2018, à(s) 13:33, Alexandre Schlöttgen <alexandre@nte.ufsm.br> escreveu:

Olá Coordenações de Curso,

em **Novembro de 2017 (14/11/2017)** encaminhei e-mail solicitando autorização para disponibilização de dados para o servidor da UFSM - **Jáder Adiel Schmitt**, lotado no CPD.

Vocês poderiam, por gentileza, responder este e-mail, indicando se permitem ou não a disponibilização dos dados.

Se a resposta for negativa, favor incluir justificativa.

Atenciosamente,
Alexandre.

"Ele necessita de acesso aos dados do moodle UAB de alguns cursos para sua dissertação.

*Não vemos nenhum problema em disponibilizar estes dados, que podem auxiliar futuramente o NTE e a UFSM a entender o comportamento dos estudantes e com isso tentar prever e diminuir a evasão dos alunos EaD. Mas precisamos da **autorização das coordenações de cursos.***

Esperamos retorno assim que possível, deste e-mail."

Alexandre Schlöttgen

Analista de Tecnologia da Informação
Núcleo de Tecnologia Educacional - NTE

Universidade Federal de Santa Maria - UFSM

<exemplo_para_coordenacao_pedagogia.ods>

ANEXO D - AUTORIZAÇÃO PARA UTILIZAÇÃO DOS DADOS DE INTERAÇÕES DO MOODLE – EDUCAÇÃO ESPECIAL

Autorização - Curso Educação Especial

José Luiz Padilha Damilano <jdamilano@hotmail.com>

Seg 19/03/2018, 16:12

Para: Jader Adiel <jaderadiel@hotmail.com>;

Cc: coordenacaoeedead@yahoo.com.br <coordenacaoeedead@yahoo.com.br>;

Olá Jader,

a Coordenação autoriza a utilização dos dados disponibilizados no moodle referentes ao Curso para a tua pesquisa. Entretanto, numa leitura rápida do teu material, referes Cursos: Letras; Pedagogia e Física para o teu estudo e não está a Educação Especial? Por....?

Saudações.

Prof. José Luiz Damilano

De: Jader Adiel <jaderadiel@hotmail.com>

Enviado: segunda-feira, 19 de março de 2018 09:42

Para: jdamilano@hotmail.com

Assunto: Autorização

Bom dia, Prof. José Luiz.

Meu nome é Jáder eu sou Analista de Tecnologia da Informação aqui da UFSM e aluno do Programa de Pós-Graduação em Tecnologias Educacionais em Rede orientando da professora Giliane Bernardi.

Venho através deste e-mail solicitar autorização para utilização dos dados do curso de Educação Especial (EAD) disponíveis no AVEA Moodle para o desenvolvimento da minha pesquisa de mestrado.

A minha pesquisa consiste na utilização de técnicas de Mineração de Dados Educacionais para detecção de possíveis alunos com tendência a evasão do curso com a utilização de dados que se referem as interações dos alunos dentro do AVEA Moodle.

Em conversa com o Alexandre do NTE ele me explicou que somente é possível a disponibilização dos dados do Moodle com a autorização do coordenador do curso.

Segue em anexo o meu projeto de pesquisa caso queira verificar o trabalho.

Desde já agradeço pela atenção.

Att.

Jáder Adiel Schmitt