

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO

FABRÍCIA DAMANDO SANTOS

**DESCOBERTA DO DESÂNIMO DE ALUNOS EM AMBIENTES
VIRTUAIS DE ENSINO E APRENDIZAGEM: UM MODELO A PARTIR DA
MINERAÇÃO DE DADOS EDUCACIONAIS**

PORTO ALEGRE

2016

Fabrcia Damando Santos

**DESCOBERTA DO DESÂNIMO DE ALUNOS EM AMBIENTES
VIRTUAIS DE ENSINO E APRENDIZAGEM: UM MODELO A PARTIR DA
MINERAÇÃO DE DADOS EDUCACIONAIS**

Tese apresentada ao Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS), como requisito parcial para a obtenção do título de Doutor em Informática na Educação.

Orientadora: Prof. Dra. Magda Bercht

Coorientador: Prof. Dr. Leandro Krug Wives

Linha de Pesquisa: Paradigmas para a Pesquisa sobre o Ensino Científico e Tecnológico.

Porto Alegre

2016

CIP - Catalogação na Publicação

Damando Santos, Fabrícia
Descoberta do Desânimo de Alunos em Ambientes
Virtuais de Ensino e Aprendizagem: Um Modelo a
partir da Mineração de Dados Educacionais / Fabrícia
Damando Santos. -- 2016.
175 f.

Orientadora: Magda Bercht.
Coorientador: Leandro Krug Wives.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Centro de Estudos Interdisciplinares
em Novas Tecnologias na Educação, Programa de Pós-
Graduação em Informática na Educação, Porto Alegre, BR-
RS, 2016.

1. Computação afetiva . 2. Mineração de dados
educacionais. I. Bercht, Magda, orient. II. Wives,
Leandro Krug, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os
dados fornecidos pelo(a) autor(a).



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO

**ATA SOBRE A DEFESA DE TESE DE DOUTORADO
FABRÍCIA DAMANDO SANTOS**

Às quatorze horas do dia seis de abril de dois mil e dezesseis, na sala 329 do PPGIE/CINTED, nesta Universidade, reuniu-se a Comissão de Avaliação, composta pelos Professores Doutores: Eliseo Berni Reategui, Evandro Alves e Patrícia Augustin Jaques Maillard para a análise da defesa de Tese intitulada **"DESCOBERTA DO DESÂNIMO DE ALUNOS EM AMBIENTES VIRTUAIS DE ENSINO E APRENDIZAGEM: um modelo a partir da Mineração de Dados Educacionais"**, da doutoranda do Programa de Pós-Graduação em Informática na Educação Fabrícia Damando Santos, sob a orientação da Prof^a. Dr^a. Magda Bercht e coorientação do Prof. Dr. Leandro Krug Wives.

A Banca, reunida, após a apresentação e arguição, emite o parecer abaixo assinalado.

Considera a Tese aprovada

(x) sem alterações;

() e recomenda que sejam efetuadas as reformulações e atendidas as sugestões contidas nos pareceres individuais dos membros da Banca;

Considera a Tese reprovada.

Considerações adicionais (a critério da Banca):

A banca destaca a relevância do tema bem como a
contribuição da pesquisa na área de Informática na Educação.
Sugere também que sejam considerados os comentários
dos pareceres individuais.

Porto Alegre, 06 de abril de 2016.

Prof^a. Dr^a. Magda Bercht
Presidente e Orientadora

Prof. Dr. Leandro Krug Wives
Coorientador

Prof. Dr. Eliseo Berni Reategui
PPGIE/UFRGS

Prof. Dr. Evandro Alves
FACED/UFRGS

Prof^a. Dr^a. Patrícia Augustin-Jaques Maillard
UNISINOS

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do CINTED: Prof. José Valdeni de Lima

Coordenador do PPGIE: Prof. Eliseo Berni Reategui

*Para meu filho Marcelo, luz da minha
vida e minha maior alegria, por quem meu
coração bate feliz!*

*Para Humberto, meu grande amor, por
ser um grande parceiro na realização dos meus
sonhos!*

AGRADECIMENTOS

Chegou o momento pelo qual esperei desde o início do doutorado e, finalmente, deixarei minhas emoções aflorarem para o momento de agradecimento. Então vamos fechar esse ciclo!

A construção de uma pesquisa tão longa como esta tese iniciou-se com um movimento apaixonado pelo saber e amor ao ensino, trilhou vários afetos e, para concluí-la, foi preciso muita perseverança e apoio contínuo de algumas pessoas ao meu redor, tornando importantíssimo agradecer aos que fizeram parte dessa caminhada.

Agradeço de alma e coração aos meus dois amores, meu filho Marcelo e meu marido Humberto, por fazerem pulsar alegria em minha vida e serem o meu tudo. Obrigada ao meu marido Humberto, meu grande motivador. Desculpem-me pelas diversas ausências, em que muito do meu tempo foi dedicado a esta pesquisa, mas tenham certeza de que vocês foram os alicerces para a construção deste projeto. Aos dois, minha eterna gratidão e meu amor mais puro e sincero.

Um agradecimento especial para a pessoa que me ajudou a vir nessa vida e segurou na minha mão, mostrando que sou capaz de fazer e ser o que eu quiser, obrigada, mamãe.

Também teve aquela pessoa doce, com quem eu tenho um dos maiores laços que me foram permitidos nessa vida, minha linda irmã Alessandra, obrigada por compreender minha dedicação ao estudo.

Obrigada à minha avó Laíde, que sempre tem um olhar afetuoso para mim. Agradeço aos meus tios, tias, primos e primas e, também, ao meu lindo afilhado Wallace que, ao longo desse tempo, se transformou em um belo adolescente. Agradeço à minha madrinha Neide, que sempre opinou sabiamente nas minhas decisões, participando dessa minha caminhada, mesmo que distante.

Obrigada ao meu sogrão Getúlio, por ser essa pessoa alegre com quem eu posso contar em todas as horas, além de proporcionar inúmeras distrações, ajudando a amenizar a tensão de alguns dias. Duda, minha sogra, obrigada por apoiar na língua estrangeira.

Agradeço de coração àqueles que abraçaram a minha ideia, me ajudaram a conduzir esta tese, presentes do início ao fim. Meus mentores, orientadora Magda Bercht e coorientador Leandro Wives, obrigada por todos os ensinamentos, incentivos, parcerias e acolhida.

Agradeço aos professores que participaram da banca do projeto de tese, cada um com suas contribuições e apontamentos possibilitou que eu seguisse o melhor caminho para a conclusão desta pesquisa. Obrigada à doce e talentosa professora Patrícia Jaques, ao grande professor Palazzo e ao queridão Eliseo Reategui.

Também preciso agradecer a alguns professores que fizeram diferença em minha vida acadêmica e profissional, ao meu querido professor Weber Martins, que desde a época da UFG me apresentou e despertou em mim o interesse sobre ambientes de aprendizagem. Ao meu orientador de mestrado Leonardo Guedes, por ter sido um grande mestre e motivador.

Aproveito para agradecer aos docentes do Programa de Pós-Graduação em Informática na Educação, pelo pioneirismo e por tornarem o PGIE em um lugar de excelência no ensino e em pesquisas. Voltarei!

Quero agradecer em especial ao prof. Silvio Cazella, que, com todo seu humor e sabedoria, fez da disciplina de MDE um momento de aprendizado único. À Profa. Liane Tarouco, que, mesmo sem saber, me apresentou ao PGIE, um lugar no qual sempre quero estar. Obrigada ao professor Eliseo B. Reategui, que com sua sapiência e segurança me fez acreditar e apaixonar pela mineração de dados.

A todos meus amigos e, principalmente, aos companheiros de doutorado. Cristiane, Renata, Cícero, Aninha, Fernanda e Maristela, obrigada por todas as trocas, compreensões e companheirismo nos momentos de desespero. Vocês são amigos com que o PGIE me presenteou! Um eterno agradecimento à minha colega e agora grande amiga Maria Ivanice, por ter disponibilizado o cenário para a realização dos meus experimentos.

Agradeço à UERGS, instituição em que trabalho, por ter proporcionado a imersão nos últimos dois anos desta tese, os quais foram fundamentais para o desenvolvimento da pesquisa.

Agradeço aos meus alunos, por buscarem dentro de uma instituição de ensino a realização de seus sonhos e por me permitirem participar de suas vidas acadêmicas. Vocês sempre serão lembrados. Um agradecimento especial ao meu bolsista Max, por ajudar a codificar as minhas ideias.

Agradeço, finalmente, ao meu pai (*in memoriam*), por sempre falar que eu seria doutora, talvez não na área em que ele imaginava, mas tenho certeza de que estaria feliz em me ver assim.

*Por falar em sonho.
Nós somos sonhos encarnados.
A vida inteira é uma tentativa desesperada
para fazer com que nossos sonhos se tornem
realidades. Não basta ter sonho, é preciso
transformar a coisa em uma realidade.
Se não tivermos sonhos, fantasias e utopias,
jamais poderemos construir um jardim.*

RUBEM ALVES

RESUMO

A presente pesquisa aborda uma investigação interdisciplinar (Educação e Computação) sobre estudos que estabeleceram como foco a influência da afetividade na educação e sobre como reconhecer o desânimo do aluno em interação em um ambiente virtual de ensino e aprendizagem (AVEA) utilizando mineração de dados educacionais (MDE). A afetividade pode influenciar na aprendizagem do aluno, principalmente com relação aos aspectos negativos, frustrações, sensações de solidão, desânimo, fazendo com que o aluno possa, inclusive, desistir de um curso, tornando-se uma problemática no ensino. Identificar esses aspectos em cursos à distância torna-se desafiador para o professor devido à distância temporal e assincronicidade desse meio. Nos cursos à distância, essa possibilidade pode ser permitida através das análises dos dados das interações do aluno no ambiente, porém, o volume de dados existentes torna-se muito grande para ser analisado pelo professor, fazendo com que seja mais difícil realizar essa identificação. Na busca por identificar o estado de ânimo desanimado, esta tese apresenta um Modelo de Predição do Desânimo baseado em comportamento observável e autorrelato armazenados em AVEA, utilizando regras de associação. Para desenvolver o Modelo de Predição do aluno, as variáveis comportamentais indicadoras do desânimo foram evidenciadas na pesquisa, bem como a utilização dos fundamentos e instrumento de Scherer para identificação dos estados afetivos, mais precisamente do estado de ânimo desanimado, que duram por longos períodos, possibilitando sua identificação após determinados fatos terem ocorrido no processo de aprendizagem, o que possibilitou ter uma metodologia de acompanhamento do aluno. As regras de associação foram descobertas devido ao potencial da MDE, que, além de propiciar a inferência e predição, pode ser usada para fornecer apoio tanto ao professor, no processo de ensino e acompanhamento do aluno, quanto ao aluno, no processo de aprendizagem. Nesse contexto, a pesquisa é aplicada ao processo de ensino e aprendizagem utilizando como procedimento técnico experimentos para coleta de dados. Foram feitos experimentos com aplicação de técnicas computacionais para apoio à inferência e geração do modelo de predição. Em cada experimento onde se aplicou a MDE, as melhores regras foram escolhidas com base nas medidas de interesse e presença do estado de ânimo desanimado. A partir dessas melhores regras, uma validação foi realizada em um novo experimento propondo o Modelo de Predição do aluno desanimado em interação no AVEA Moodle. Além de apresentar o Modelo de Predição do Aluno Desanimado, este modelo foi implementado e integrado como ferramenta computacional à plataforma Moodle. A pesquisa justifica-se na medida em que apresenta inovação tecnológica para investigar a influência da afetividade na aprendizagem dentro do contexto da Educação a Distância (EAD) e aplica técnicas computacionais desenvolvendo um Modelo de Predição do Aluno Desanimado, que fornece para o professor uma visão geral do modelo e melhor acompanhamento de seus alunos, através de *dashboard*, contribuindo na sua prática docente. Logo, a tese apresenta como destaque inovador um produto de pesquisa com utilidade na prática docente no ensino superior, principalmente em cursos EAD, para o reconhecimento de aspectos relacionados à afetividade no contexto educacional. Através da ferramenta computacional, um melhor acompanhamento de alunos desanimados em interação em AVEA pode ser feito pelo professor, permitindo a este fomentar uma metodologia de acompanhamento desses alunos, a fim de minimizar futuras evasões, bem como desistências em cursos e disciplinas, beneficiando a comunidade acadêmica.

Palavras-chave: **Computação Afetiva. Afetividade em Ambientes Virtuais de Ensino e Aprendizagem. Desânimo. Mineração de Dados Educacionais.**

ABSTRACT

This research addresses an interdisciplinary research (Education and Computer) on studies that established focus on the influence of affectivity in education and how to recognize the dismay of student interaction in a virtual teaching and learning environment (VTLE) using educational data mining (EDM). Affection can influence student learning, particularly with respect to the negative, frustration, feelings of loneliness, discouragement, causing the student can even give up a course, becoming a problematic teaching. Identify these aspects in distance courses becomes challenging for the teacher due to the temporal distance and asynchronicity that medium. In distance learning courses, this possibility may be permitted by the data analysis of student interactions in the environment, however, the amount of data becomes too large to be analyzed by the teacher, making it more difficult to carry out such identification. In seeking to identify the state of despondent mood, this thesis presents a prediction model of the observable behavior-based Discouragement and self-report stored in VTLE using association rules. To develop the prediction model student, the indicator behavioral variables of discouragement were evident in the research, and the use of the grounds and Scherer tool to identify the affective states, specifically the state of despondent mood that last for long periods, enabling identification after certain events have occurred in the learning process, making it possible to have a follow-up methodology of the student. Association rules were discovered due to the potential of the EAW, which, besides providing the inference and prediction, can be used to provide support to both the teacher in the teaching and monitoring of the student as the student in the learning process. In this context, the research is applied to the teaching and learning process using as a technical procedure experiments to collect data. experiments were made with application of computational techniques to support the inference and generation of the prediction model. In each experiment where we applied the MED, the best rules were chosen based on measures of interest and presence in the state of despondent mood. From these best rules, a validation was performed on a new experiment proposing the Prediction Model discouraged student interaction in VTLE Moodle. In addition to presenting the Prediction Model of Student Discouraged, this model was implemented and integrated as a computational tool to the Moodle platform. The research is justified in that it presents technological innovation to investigate the influence of affect on learning within the education context Distance Learning and applies computational techniques developing a prediction model Discouraged Student, which provides for the teacher a view general model and better monitoring of their students through dashboard, contributing to their teaching practice. Therefore, the thesis shows how innovative highlight a research product to use in teaching practice in higher education, especially in distance education courses, for the recognition of aspects related to affectivity in the educational context. Through computational tool for better monitoring of disheartened students interacting in VTLE it can be done by the teacher, allowing him to promote a follow-up methodology of these students in order to minimize future evasions and dropouts courses and disciplines, benefiting the community academic.

Keywords: Affective Computing. Affectivity in Virtual Environments for Teaching and Learning. Discouragement. Educational Data Mining.

RESUMEN

Esta investigación se ocupa de una investigación interdisciplinaria (Educación e Informática) en los que se estableció el enfoque sobre la influencia de la afectividad en la educación y cómo reconocer la consternación de la interacción del estudiante en un entorno virtual de enseñanza aprendizaje (AVEA) utilizando la minería de datos educativa (MDE). El afecto puede influir en el aprendizaje del estudiante, en particular con respecto a la negativa, frustración, sentimientos de soledad, desánimo, haciendo que el estudiante puede incluso renunciar a un curso, convirtiéndose en una enseñanza problemática. Identificar estos aspectos en los cursos a distancia se convierte en un reto para el maestro debido a la distancia temporal y asincronía ese medio. En los cursos de enseñanza a distancia, esta posibilidad puede ser permitido por el análisis de los datos de las interacciones de los estudiantes en el ambiente, sin embargo, la cantidad de datos es demasiado grande para ser analizados por el profesor, lo que hace más difícil llevar a cabo dicha identificación. Al tratar de identificar el estado de ánimo deprimido, esta tesis presenta un modelo de predicción del desaliento observables basada en el comportamiento y auto-informe almacenado en AVEA las reglas de asociación. Para desarrollar el estudiante modelo de predicción, las variables de comportamiento del indicador de desaliento eran evidentes en la investigación, y el uso de los terrenos y Scherer herramienta para identificar los estados afectivos, específicamente el estado de ánimo deprimido que duran por largos períodos de tiempo, que permite la identificación después de ciertos acontecimientos se han producido en el proceso de aprendizaje, por lo que es posible tener una metodología de seguimiento del estudiante. reglas de asociación fueron descubiertos debido al potencial de la orden de detención europea, que, además de proporcionar la inferencia y la predicción, se puede utilizar para proporcionar apoyo tanto a la maestra en la enseñanza y el seguimiento del alumno como estudiante en el proceso de aprendizaje. En este contexto, la investigación se aplica al proceso de enseñanza y aprendizaje mediante experimentos como un procedimiento técnico para recopilar datos. experimentos se hicieron con la aplicación de técnicas computacionales para apoyar la inferencia y la generación del modelo de predicción. En cada experimento en el que se aplicó el MDE, las mejores reglas fueron elegidos en base a medidas de interés y presencia en el estado de ánimo deprimido. A partir de estas mejores reglas, una validación se realizó en un nuevo experimento que propone la interacción de los estudiantes desalentado modelo de predicción de AVEA Moodle. Además de presentar el Modelo de Predicción del Estudiante Desalentado, este modelo fue implementado e integrado como una herramienta computacional para la plataforma Moodle. La investigación se justifica porque presenta la innovación tecnológica para investigar la influencia del efecto sobre el aprendizaje en el contexto de la educación a distancia (EAD) y aplica técnicas computacionales en desarrollo un modelo de predicción de Estudiantes Desalentado, que prevé el profesor una vista modelo general y un mejor seguimiento de sus estudiantes a través de tablero de instrumentos, contribuyendo a su práctica docente. Por lo tanto, la tesis muestra cómo destacado innovador de un producto de investigación a utilizar en la práctica docente en la enseñanza superior, sobre todo en los cursos de educación a distancia, para el reconocimiento de los aspectos relacionados con la afectividad en el contexto educativo. A través de la herramienta computacional para un mejor seguimiento de los estudiantes desanimados que interactúan en AVEA se puede hacer por el profesor, lo que le permite promover una metodología de seguimiento de estos estudiantes con el fin de reducir al mínimo las evasivas y abandonos futuros cursos y disciplinas, en beneficio de la comunidad académica.

Palabras clave: **Computación afectiva. Afectividad en los entornos virtuales de enseñanza y aprendizaje. El desánimo. La minería de datos para la Educación.**

LISTA DE FIGURAS

Figura 1: Visão geral do projeto	26
Figura 2: Mapa conceitual	30
Figura 3: Geneva Emotion Wheel	33
Figura 4: Roda de Estados de Ânimo 2.0 / GEW 2.0.....	36
Figura 5: Processo de descoberta de conhecimento em banco de dados.....	44
Figura 6: Técnicas de MDE - Taxonomia	50
Figura 7: Etapas para MDE	60
Figura 8: Tela inicial do Weka	62
Figura 9: Apresentação do Weka.....	62
Figura 10: Exemplo de arquivo ARFF	63
Figura 11: Passos para exportar e usar um arquivo do Moodle no Weka	64
Figura 12: Tela do <i>RapidMiner</i>	65
Figura 13: Tela Knime.....	66
Figura 14: Tela do SPSS.....	67
Figura 15: Processos fundamentais do processo geral de pesquisa.....	70
Figura 16: Procedimento Metodológico	72
Figura 17: Visão geral das etapas de elaboração do projeto.....	73
Figura 18: Histograma da distribuição de frequência.....	79
Figura 19: Histograma dos estados de ânimo	80
Figura 20: Preferência de ferramentas	82
Figura 21: Solicitação de ajuda ao professor.....	83
Figura 22: Com relação às atividades individuais	84
Figura 23: Com relação às atividades em grupo.....	84
Figura 24: Conteúdo da disciplina.....	85
Figura 25: Exemplo de tela com pergunta.....	89
Figura 26: Componentes e seus eventos no log.....	90
Figura 27: Frequência dos componentes de <i>log</i>	91
Figura 28: Componentes de logs utilizados na mineração	92
Figura 29: Exemplo de arquivo ARFF	94
Figura 30: Componentes para mineração	94
Figura 31: Seleção de atributos	97

Figura 32: Mineração componente tarefa.....	98
Figura 33: Mineração e métodos usados na Classificação J48.....	105
Figura 34: Atributos usados no modelo de Classificação.....	106
Figura 35: Resultado do treinamento	106
Figura 36: Resultado da mineração J48 – base de teste	107
Figura 37: Método <i>Cross-Validation</i>	108
Figura 38: Árvore de Decisão.....	109
Figura 39: Seleção de atributos J48.....	110
Figura 40: Resultado do teste	111
Figura 41: Mineração J48 – <i>Cross-Validation</i> 9 atributos	112
Figura 42: Árvore de Decisão.....	113
Figura 43: Mineração fórum - Apriori.....	115
Figura 44: Atributos selecionados	119
Figura 45: Mineração componente arquivo – Apriori.....	120
Figura 46: Visão geral das técnicas e métodos minerados	123
Figura 47: Validação a partir das regras da mineração	125
Figura 48: Resultado da validação componente fórum – <i>Holdout</i>	129
Figura 49: Resultado da validação método <i>Cross-Validation</i>	130
Figura 50: Integração PHP/Java	135
Figura 51: Tela inicial do projeto	136
Figura 52: Arquivos ARFF no Moodle	137
Figura 53: Resultado da mineração do algoritmo Apriori – componente tarefa etapa 2.....	137
Figura 54: Definição dos parâmetros para o algoritmo <i>Cross-Validation</i>	138
Figura 55: Resultado da mineração <i>Cross-Validation</i> 9 atributos.....	138
Figura 56: Seleção do arquivo de teste para o método <i>Holdout</i>	139
Figura 57: Mineração <i>Holdout</i> 9 atributos	140
Figura 58: Modelo geral de integração Weka/Moodle e apresentação <i>Dashboard</i>	141
Figura 59: Ferramenta de visualização das regras selecionadas.....	142
Figura 60: Regra 1 do componente tarefa	143
Figura 61: Regra 1 componente arquivo	144

LISTA DE QUADROS

Quadro 1: Fenômenos afetivos relacionados ao estado de ânimo desanimado – REA 2.0.....	37
Quadro 2: Exemplos de trabalhos desenvolvidos utilizando MDE.....	58
Quadro 3: Termos afetivos usados na REA 2.0	76
Quadro 4: Modelo de <i>log</i> resultante do Moodle.....	90
Quadro 5: Exemplo dos itens do <i>log</i> componente tarefa.....	95
Quadro 6: Exemplo dos itens do <i>log</i> componente fórum	103
Quadro 7: Atributos para mineração	110
Quadro 8: Características do componente arquivo.....	117
Quadro 9: Variáveis usadas para minerar os arquivos	119
Quadro 10: Regras do Modelo de Predição.....	134

LISTA DE TABELAS

Tabela 1: Medidas estatísticas das questões REA 2.0	77
Tabela 2: Frequência estatística.....	78
Tabela 3: Tabela cruzada para atividades individuais	80
Tabela 4: Dificuldade em realizar atividades individuais	81
Tabela 5: Atributos usadas em cada etapa.....	99
Tabela 6: Resultado das melhores regras geradas para o estado de desânimo.....	100
Tabela 7: Precisão dos modelos.....	114
Tabela 8: Resultado das Regras de Associação selecionadas.....	115
Tabela 9: Regras obtidas para o componente arquivo.....	121
Tabela 10: Taxas de acerto da validação das regras.....	126
Tabela 11: Taxas de acerto da validação das regras adaptadas	127
Tabela 12: Comparação da mineração e validação	130
Tabela 13: Taxas de acerto da validação das regras do experimento C	131
Tabela 14: Medida descritiva componente tarefa.....	132
Tabela 15: Taxas de acerto da validação das regras componente arquivo	133

LISTA DE ABREVIATURAS E SIGLAS

API	Interface de Programação de Aplicativos
ARFF	<i>Attribute Relation File Format</i>
AVEA	Ambiente Virtual de Ensino e Aprendizagem
BROMP	<i>The Baker Rodrigo Ocumpaugh Monitoring Protocol</i>
CA	Computação Afetiva
CNPQ	Conselho Nacional de Desenvolvimento Científico e Tecnológico
EAD	Ensino a Distância
ERP	<i>Enterprise Resource Planning</i>
GEW	<i>Geneva Emotion Wheel</i>
GLP	<i>General Public License</i>
HART	<i>The Human Affect Recording Tool</i>
IES	Instituição de Ensino Superior
KDD	<i>Knowledge Discovery in Database</i>
LA	<i>Learning Analytics</i>
MD	Mineração de Dados
MDE	Mineração de Dados Educacionais
MOOC	Massive Open Online Course
NAE	Núcleo de Assessoria Estatística
NZ	Nova Zelândia
OA	Objetos de Aprendizagem
PHP	<i>Hipertext Preprocessor</i>
PPGIE	Programa de Pós-Graduação em Informática na Educação
PSLC	<i>Pittsburgh Science of Learning Center</i>
RA	Regras de Associação
REA	Roda de Estados de Ânimo
SEAD	Secretaria de Educação a Distância
SPSS	<i>Statistical Package for the Social Sciences</i>
SQL	<i>Structured Query Language</i>
UAB	Universidade Aberta do Brasil
UFRGS	Universidade Federal do Rio Grande do Sul

SUMÁRIO

1	INTRODUÇÃO	19
1.1	CONTEXTUALIZAÇÃO	19
1.2	MOTIVAÇÃO E CARACTERIZAÇÃO DO PROBLEMA	22
1.3	A QUESTÃO DE PESQUISA	24
1.4	OBJETIVOS	24
1.5	CONTRIBUIÇÕES DA PESQUISA	25
1.6	PLANEJAMENTO DA PESQUISA / ETAPAS DA PESQUISA	25
1.7	VISÃO GERAL DOS CAPÍTULOS	26
2	DIMENSÃO AFETIVA	28
2.1	AFETIVIDADE: EMOÇÕES E SENTIMENTOS	28
2.2	ESTADO DE ÂNIMO	32
2.3	ESTADO DE ÂNIMO DESANIMADO	35
2.4	TRABALHOS ASSOCIADOS	38
2.5	CONSIDERAÇÕES SOBRE O CAPÍTULO	41
3	TÉCNICAS COMPUTACIONAIS	43
3.1	MINERAÇÃO DE DADOS.....	43
3.2	MINERAÇÃO DE DADOS EDUCACIONAIS.....	45
3.2.1	Áreas da Mineração de Dados Educacionais	48
3.2.2	Técnicas Aplicadas na MDE	49
3.2.3	Trabalhos Associados	55
3.2.4	Processo de Mineração de Dados Educacionais na Pesquisa	59
3.3	FERRAMENTAS PARA MINERAÇÃO DE DADOS EDUCACIONAIS.....	61
3.3.1	Weka	61
3.3.2	RapidMiner	65
3.3.3	KNIME	66
3.3.4	SPSS	67
3.4	CONSIDERAÇÕES SOBRE O CAPÍTULO	68
4	PROCEDIMENTO METODOLÓGICO	70
4.1	CLASSIFICAÇÃO DA PESQUISA.....	71

4.2	ETAPAS DA ELABORAÇÃO DO PROJETO E DEFINIÇÃO DA AMOSTRA....	72
4.3	DESCRIÇÃO DOS EXPERIMENTOS	75
4.3.1	Procedimento para o experimento A	75
4.3.2	Procedimentos para a mineração de dados (experimento B)	86
4.3.3	Procedimento para a validação da mineração e apresentação do modelo (experimento C)	87
5	APLICAÇÃO E ANÁLISE DA MINERAÇÃO DE DADOS	88
5.1	COMPONENTE TAREFA	95
5.1.1	Mineração de dados – Componente Tarefa	97
5.2	COMPONENTE FÓRUM	103
5.2.1	Mineração de dados – Componente Fórum	104
5.2.2	Classificação J48.....	105
5.2.3	Algoritmo Apriori.....	114
5.2.4	Considerações finais sobre a mineração do componente fórum	116
5.3	COMPONENTE ARQUIVO	117
5.3.1	Mineração de dados – Componente Arquivo.....	120
5.4	CONSIDERAÇÕES SOBRE O CAPÍTULO	122
6	EM BUSCA DO MODELO DE PREDIÇÃO DO ALUNO DESANIMADO... 125	
6.1	VALIDAÇÃO DA MINERAÇÃO DE DADOS – COMPONENTE TAREFA	126
6.2	VALIDAÇÃO DA MINERAÇÃO DE DADOS – COMPONENTE FÓRUM.....	128
6.2.1	Classificação J48.....	128
6.2.2	Algoritmo Apriori.....	131
6.3	VALIDAÇÃO DA MINERAÇÃO DE DADOS – COMPONENTE ARQUIVO ..	132
6.4	MODELO DE PREDIÇÃO DO ALUNO DESANIMADO	133
6.5	VISUALIZAÇÃO DO MODELO PROPOSTO	134
6.5.1	Integração Weka/Moodle.....	135
6.5.2	Ferramenta de visualização	141
6.6	CONSIDERAÇÕES SOBRE O CAPÍTULO	144
7	CONCLUSÕES	146
7.1	LIMITAÇÕES E TRABALHOS FUTUROS	148
	REFERÊNCIAS	151

APÊNDICE A – TERMO DE CONSENTIMENTO.....	160
APÊNDICE B – QUESTIONÁRIO	161
APÊNDICE C – MINERAÇÃO COMPONENTE TAREFA – ETAPA 1.....	164
APÊNDICE D – MINERAÇÃO COMPONENTE TAREFA – ETAPA 2.....	165
APÊNDICE E – MINERAÇÃO COMPONENTE TAREFA – ETAPA 3.....	166
APÊNDICE F – MINERAÇÃO COMPONENTE ARQUIVO – ETAPA 1	167
APÊNDICE G – MINERAÇÃO COMPONENTE ARQUIVO – ETAPA 2.....	168
APÊNDICE H – MINERAÇÃO COMPONENTE ARQUIVO – ETAPA 2.....	169
APÊNDICE I – MINERAÇÃO COMPONENTE FÓRUM – APRIORI.....	170
APÊNDICE J – MINERAÇÃO COMPONENTE FÓRUM – VALIDAÇÃO.....	171
ANEXO A – GRUPOS DE PESQUISAS CNPQ – COMPUTAÇÃO AFETIVA	172

1 INTRODUÇÃO

A pesquisa desenvolvida nesta tese é um dos requisitos parciais para a obtenção do título de Doutor através do Programa de Pós Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS). Esta pesquisa foi elaborada com enfoque principal na linha de pesquisa *Paradigmas para a Pesquisa sobre o Ensino Científico e Tecnológico*.

Esta tese apresenta uma investigação na busca da identificação do aluno que apresenta desânimo quando está em interação em um ambiente virtual de ensino e aprendizagem (AVEA). Aborda a construção de um Modelo de Predição do desânimo com o apoio da mineração de dados educacionais (MDE). Para essa finalidade, foi necessário tecer um olhar interdisciplinar, integrando conhecimentos relacionados às áreas da Ciência da Computação e Computação Afetiva aplicadas na Educação. Dessa forma, a pesquisa tem alicerces nas teorias e trabalhos de Scherer (2000, 2005), Sacharin, Schlegel e Scherer (2012), Picard (2000), Sabourin e Lester (2014), Pontarolo (2008), Longhi (2011), Bercht (2001), Iepsen (2013) e Damásio (1996; 2000; 2004; 2011).

Também foram usadas técnicas computacionais aplicadas aos dados advindos das interações dos alunos no ambiente utilizando mineração de dados educacionais. Para este fim, o aporte teórico fundamentou-se em Baker (2014), Ocumpaugh *et al.* (2015), Baker, Isotani e Carvalho (2011), Rigo *et al.* (2014), Gottardo, Noronha e Kaestner (2014), Tan, Steinbach e Kumar (2009), Romero e Ventura (2006), Costa *et al.* (2012), bem como mineração de dados tradicional, descrita por Feldman e Sanger (2007).

1.1 CONTEXTUALIZAÇÃO

À medida que o Ensino a Distância (EAD) se dissemina, crescem as possibilidades das pessoas realizarem cursos, ofertados em instituições de ensino, que antes só estavam disponíveis de forma presencial. Conforme Moore e Kearsley (2008), os cursos ofertados via correspondência para conclusão de curso superior foram usados pela primeira vez em 1883 pelo *Chautauqua Correspondence College*. Porém, com o uso e popularização da internet nos cursos à distância, tornou-se possível a utilização de AVEA, objetos de aprendizagem (OA),

simuladores, ambientes imersivos, dentre outros, como suporte ao EAD. Através dos AVEA, criaram-se espaços para organizar cursos e disciplinas, administrar conteúdos e monitorar alunos nas modalidades presenciais, semipresenciais e a distância. As potencialidades dos ambientes virtuais vão além de repositórios de materiais e gerenciadores de conteúdos e/ou de participantes, pois permitem a interação dos participantes através das ferramentas integradas ao ambiente (BEHAR *et al.*, 2013 p. 58).

Por meio das interações entre os participantes (tutores, professores e alunos), ocorrem possibilidades de trocas mediadas através de um sistema de comunicação (fórum, *chat*, *e-mail*, entre outros). A partir dessas interações, muitos laços se criam, se fortalecem, causando sentimento de pertencimento ao curso e, principalmente, minimizando o sentimento de distância física entre os sujeitos envolvidos (BASTOS, 2012; BEHAR *et al.*, 2013, p. 47).

Por outro lado, há alunos que sentem dificuldade de interação ou falta dela, sentindo-se isolados, desanimados e desmotivados. Mas estas sensações devem ser evitadas, pois um AVEA deve ser um local em que professores e alunos trabalham juntos, emergindo a colaboração e a interação, caso contrário, pode levar à desistência, à desmotivação e até à evasão do aluno (SOUSA, 2006; BASTOS, 2012).

Observa-se que, através dessas interações, os aspectos relacionados à afetividade do aluno manifestam-se tanto positivamente quanto negativamente e, por meio das ações que o aluno realiza em um ambiente, torna-se possível observar seus dados comportamentais e dados fisiológicos, que podem ser usados como indicadores para identificar os seus estados afetivos (BERCHT, 2001). Nesse mesmo sentido, Picard e Daily (2008) afirmam que dados de um usuário podem ser adquiridos também pelo desempenho em tarefas. Estes dados se constituem na denominação de *observações comportamentais* ou *variáveis comportamentais*.

Fundamenta-se, assim, a possibilidade de obter variáveis comportamentais a partir de um AVEA, desde a manifestação explícita da emoção, por exemplo, via pistas textuais, até o seu comportamento observável, que pode ocorrer a partir das ações do aluno no ambiente, passando por análise dos *logs*¹ e fóruns, até o uso e análise de sinais por meio de sensores fisiológicos (PICARD, 2000; BERCHT, 2001; LONGHI, 2011; MAJADAS *et al.*, 2013).

Sob esse enfoque, pesquisas associadas à afetividade no processo de ensino e aprendizagem, suportadas pela Computação Afetiva (CA) e pela Educação, vêm sendo desenvolvidas, devido à possibilidade de se inferir a afetividade e usar tais dados como apoio

¹ Log é o registro de dados das atividades/eventos que ficam armazenados no computador ou AVEA.

ao professor, visando à melhor forma de atender ao aluno. Alguns exemplos dessas pesquisas são listados a seguir:

- a) Desenvolvimento de agentes pedagógicos com dimensões afetivas (BERCHT, 2001);
- b) Desenvolvimento de modelo afetivo para a inferência dos estados de ânimo dos alunos (LONGHI, 2011);
- c) Identificação da frustração do aluno ao realizar um algoritmo e posterior direcionamento com apoio específico (IEPSEN, 2013);
- d) Análise do afeto e engajamento em jogo baseado em ambiente de aprendizagem (SABOURIN; LESTER, 2014);
- e) Modelagem de aspectos afetivos do aluno em jogo educacional colaborativo (PONTAROLO, 2008);
- f) Desenvolvimento de um agente pedagógico animado para interagir afetivamente com o aluno (JAQUES, 2004), entre outros.

Mesmo assim, ao considerar o atual estágio da tecnologia, as pesquisas ainda estão incipientes quando se trata da abordagem e consideração da dimensão afetiva em sistemas computacionais (LONGHI; BEHAR; BERCHT, 2009; BAKER, 2014). Fato observado no contexto nacional após realizar-se uma busca pelo termo *computação afetiva* no grupo de pesquisas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), no qual, atualmente (janeiro de 2016), há dez (10) grupos trabalhando com linhas de pesquisas relacionadas à CA (Anexo A). Ainda assim, a maioria dos grupos não se dedica de forma enfática à CA, explorando somente linhas de pesquisa que abordam o tema, o que reforça e evidencia uma área carente de pesquisas e projetos.

Apesar de os estudos serem incipientes na área da CA associada à Educação, eles são transformadores, vistos os resultados alcançados pelas pesquisas supracitadas, apresentando-se como motivadoras e engajadoras na perspectiva de criação de melhorias no processo educacional, ora apoiando o professor, ora informatizando e dotando computadores com novas possibilidades de interação com alunos e professores. Da mesma forma, a MDE se mostra nova, com potencial crescente, na medida em que novas inferências e predições podem ser realizadas, antecipando fatos que podem vir a contribuir com a melhoria do ensino e aprendizagem (BAKER, 2014).

1.2 MOTIVAÇÃO E CARACTERIZAÇÃO DO PROBLEMA

De forma geral, quando a aula é presencial, questões relacionadas à afetividade no contexto da aprendizagem podem até ser percebidas pelo professor, pois algumas vezes ele pode identificar se a turma ou o aluno está motivado ou desanimado com um conteúdo ou uma atividade. Quando o professor detecta essa situação, como o caso do desânimo, ele pode adotar rapidamente novas ações ou práticas pedagógicas a fim de apoiar o aluno, a turma ou um grupo específico de alunos.

Silva e Figueiredo (2012) corroboram essa concepção ao expor que no ensino presencial o professor consegue “sentir” a turma, diagnosticar problemas, dificuldades e, por isso, pode alterar a sua didática e metodologia. Porém, notar essa situação em uma disciplina ou curso EAD exige muito mais do professor, devido ao grande volume de dados e interações existentes nestes ambientes, além da distância temporal e da assincronicidade.

Devido a isso, percebe-se a necessidade de se reconhecerem aspectos relacionados à afetividade e à possibilidade de se usar dados observáveis para fornecer indicativos sobre o comportamento afetivo do aluno. Por outro lado, pesquisas apontam a necessidade e as formas de oferecer apoio ao professor para identificar alunos propensos ao desânimo e à evasão, o que pode ocorrer ora instrumentalizando professores para repensar ou adotar novas práticas pedagógicas, ora exibindo ações pedagógicas de apoio ao estudante (IEPSEN, 2013; LONGHI, 2011; KAMPFF, 2009; PICARD, 2010; BAKER, 2014).

Uma das técnicas que pode ser utilizada para prover ferramentas de apoio ao professor se dá através da MDE. Para Costa *et al.* (2012), há diversas tarefas em MDE para analisar dados gerados nas interações dos estudantes no AVEA, além de uma necessidade significativa e urgente de prover ambientes computacionais apropriados para a MDE e oferecer ao professor facilidade de uso desses dados nesses ambientes.

Frente a esse cenário, investigou-se, nesta tese, o comportamento observável do aluno que está no estado de ânimo desanimado, pois, para a área educacional, o desânimo surge quando o aluno não consegue acompanhar um conteúdo ou quando não consegue cumprir os prazos para a entrega das atividades, o que causa um obstáculo na sua aprendizagem e até mesmo o abandono da disciplina ou curso, tornando-se uma problemática no processo de aprendizagem (LONGHI, 2011).

De forma ainda mais preocupante, o desânimo pode ser um problema avassalador no processo de aprendizagem do aluno, conforme apontado por Seligman e Csikszentmihalyi (2000) e Figueiredo (2011) quando asseguram que indivíduos que estão em situações de desânimo se veem confrontados com situações as quais não conseguem ultrapassar, conduzindo, no caso da educação, o aluno a um estado de apatia e passividade representado por afetos negativos. Associado a isso, quando as emoções negativas persistem, conforme justificado por Figueiredo (2011), tornam-se incompatíveis com os objetivos escolares, levando o aluno à desistência.

Esse fato também é observado por Picard (2000) quando afirma que ações negativas possuem intensa influência sobre a memória, podendo afetar o aprendizado do aluno, justificando, novamente, o infortúnio que esse tipo de afeto pode acarretar.

Apesar de haver sistemas informatizados que se propõem a tratar os aspectos afetivos, a sua maioria, atualmente, aborda como foco principal o reconhecimento de emoções básicas. Poucas aplicações foram desenvolvidas sob o enfoque dos estados de ânimo (animado/desanimado e/ou satisfeito/insatisfeito) do aluno no processo de aprendizagem (LONGHI; BEHAR; BERCHT, 2009).

Essas constatações asseguram, novamente, a importância e o enfoque da presente pesquisa com relação aos estados de ânimo, principalmente com relação ao estado de ânimo desanimado, por ser um estado que alavanca a falta de participação do aluno em atividades de aprendizagem e, até mesmo, podendo levá-lo ao abandono, o que leva a presente pesquisa a explorar essa lacuna.

Portanto, a pesquisa é impulsionada pela possibilidade da identificação das variáveis comportamentais indicadoras do estado de ânimo desanimado dos alunos em interação em um AVEA, aplicando técnicas computacionais, principalmente a MDE e técnicas estatísticas (histogramas, correlações, geração de quartis), com intuito em identificar quais alunos estão propensos ao desânimo, no sentido de apoiar e instrumentalizar o professor na medida em que identifica esses alunos.

1.3 A QUESTÃO DE PESQUISA

Em consonância com o exposto, esta tese propõe como questão norteadora:

Como identificar o estado de ânimo desanimado a partir de indicadores das variáveis comportamentais dos alunos em um ambiente virtual de ensino e aprendizagem?

Ainda a partir da questão norteadora, foram investigadas as seguintes questões específicas:

Como identificar os alunos desanimados por meio de suas interações?

Quais mecanismos computacionais podem ser utilizados para reconhecer o estado de ânimo desanimado de um aluno em um ambiente virtual?

1.4 OBJETIVOS

Diante da questão de pesquisa apresentada, abordam-se os objetivos descritos abaixo, tendo como objetivo principal:

Investigar quais as variáveis comportamentais que representam o desânimo e que podem ser descobertas pela interação dos alunos em um AVEA para compor um Modelo de Predição do aluno desanimado.

Para tal, os seguintes objetivos específicos foram desenvolvidos a fim de se alcançar o objetivo geral:

- a) Investigar quais são as variáveis comportamentais que indicam o estado de ânimo desanimado;
- b) Investigar que tecnologias computacionais podem ser adotadas para apoiar a descoberta do desânimo no aluno;
- c) Desenvolver um Modelo de Predição do estado de ânimo desanimado para o aluno em interação no AVEA.

1.5 CONTRIBUIÇÕES DA PESQUISA

As principais contribuições da presente tese, de forma inovadora, à academia e à sociedade são:

- a) Obtenção de um mapeamento das variáveis observáveis indicadoras do desânimo;
- b) Desenvolvimento de um modelo que identifique o comportamento de um aluno desanimado na modalidade EAD, contribuindo para o reconhecimento de aspectos relacionados à afetividade no contexto educacional;
- c) Construção do Modelo de Predição para contribuir na identificação dos alunos desanimados em um AVEA, a fim de apresentar ao professor uma visão geral do comportamento do aluno desanimado, para que o professor possa atuar da melhor forma no apoio e acompanhamento do processo de aprendizagem desses alunos, buscando minimizar os efeitos que o desânimo pode causar na aprendizagem, bem como na diminuição de possíveis evasões e/ou abandonos de cursos;
- d) Apresentação dos dados minerados via ferramenta visual, integrada ao Moodle em forma de *dashboard*, possibilitando uma interface mais amigável para que o professor identifique e acompanhe o aluno desanimado.

1.6 PLANEJAMENTO DA PESQUISA / ETAPAS DA PESQUISA

Esta pesquisa possui *natureza aplicada ao processo de ensino e aprendizagem*, com ênfase na Educação Superior, incluída nas modalidades presencial, semipresencial ou à distância, e que utiliza um sistema computacional que permita a análise das interações do aluno com o ambiente. É classificada, com relação aos seus objetivos, como *explicativa*, preocupando-se em identificar fatores que contribuem para ocorrência de determinados fenômenos. Com relação aos procedimentos técnicos, é classificada como *experimento*, investigando dados após a sua ocorrência, como o caso do aluno desanimado. Os detalhes e justificativas da metodologia encontram-se no Capítulo 4.

A fim de alcançar a questão de pesquisa e questões secundárias, a partir do levantamento bibliográfico, foi realizado estudo piloto preliminar no semestre 2013/1. Utilizando o instrumento questionário, o estudo piloto abordou a percepção da afetividade em um AVEA a partir das interações de alunos que estavam em atividade de grupo, analisando, por exemplo, como o aluno considerava o grupo com relação à interação entre os participantes, com relação à satisfação, à colaboração do grupo e à participação do grupo nas atividades. Mais detalhes encontram-se na seção 4.2.

A partir desse estudo piloto, foram desenvolvidos experimentos (A, B, e C), para os semestres de 2014/1 e 2014/2, a fim de mapear as variáveis comportamentais dos alunos em interação no AVEA, aplicar de técnicas computacionais (mineração de dados educacionais e análise estatística) e apresentar um modelo, baseado em regras mineradas, para a inferência do desânimo de alunos.

Uma visão geral da proposta para a pesquisa, enfatizando tanto o mapeamento quanto a captura e apresentação dos dados, é exposta na Figura 1.

Figura 1: Visão geral do projeto



Fonte: Elaborado pela autora.

1.7 VISÃO GERAL DOS CAPÍTULOS

A presente tese encontra-se organizada em sete capítulos, contemplando esta Introdução para a construção do objeto de pesquisa.

O Capítulo 2 aborda de forma mais detalhada os aspectos relacionados à dimensão afetiva, discorrendo sobre a afetividade e suas características, fundamentadas por Damásio (1996; 2000; 2004; 2011). Com relação aos estados de ânimo, buscou respaldo em Scherer (2005) e Longhi (2011).

O Capítulo 3 dedica-se a tratar de conceitos, técnicas e ferramentas computacionais, como mineração de dados e mineração de dados educacionais, tendo como fundamentação as pesquisas de Feldman e Sanger (2007); Baker *et al.* (2014); Baker; Yacef (2009); Costa *et al.* (2009); Romero; Ventura (2006; 2008), entre outros.

O Capítulo 4 discorre sobre os aspectos metodológicos que norteiam a pesquisa. Descreve a natureza da pesquisa, sua classificação quanto aos objetivos, procedimentos técnicos e tipo de abordagem do problema. Além da definição da população e amostra, também são apresentados os experimentos desenvolvidos.

O Capítulo 5 detalha o processo realizado de mineração de dados educacionais, apresentando a análise dos experimentos desenvolvidos, as minerações, técnicas e métodos aplicados. Foram feitas análises e comparações entre as melhores minerações, a fim de se apresentar um conjunto de variáveis comportamentais e regras mineradas para compor o Modelo de Predição do aluno desanimado. Esse capítulo apoia a explanação da primeira contribuição da tese.

No Capítulo 6, apresentam-se os procedimentos de validação das regras em um novo conjunto de dados, permitindo, assim, testá-las e validá-las, atingindo um grupo de regras para a criação de um Modelo de Predição do desânimo. A partir da validação, foi desenvolvido o Modelo de Predição e foi implementada uma integração entre o minerador de dados e o AVEA Moodle que permitiu que o modelo de regras estivesse disponível dentro do ambiente. Associada a esse modelo, foi desenvolvida uma ferramenta mais visual de apresentação do Modelo de Predição, na forma de *dashboards*. Esse capítulo atende a segunda e terceira contribuições desta tese.

Finalizando a tese, o Capítulo 7 expõe as considerações finais deste trabalho, com reflexões sobre todo o processo, além de detalhar limitações e apresentar perspectivas futuras para a continuidade dessa pesquisa.

2 DIMENSÃO AFETIVA

Este capítulo apresenta a revisão bibliográfica sobre aspectos relacionados à afetividade e apresenta abordagens que fundamentam os conceitos, teorias e técnicas adotadas referentes à dimensão afetiva.

2.1 AFETIVIDADE: EMOÇÕES E SENTIMENTOS

A definição de emoção apresentada nesta pesquisa é pautada nas concepções de Scherer (2000; 2005) e Damásio (1996; 2000; 2004; 2011). Scherer (2005) inicia o seu trabalho “*What are the emotions? And how can they be measured?*” justamente com esta provocação, citando que já foram apresentadas várias definições sobre emoções, tanto por cientistas como por leigos, cuja resposta não apresenta um consenso.

Para Scherer (2005), a emoção é um episódio inter-relacionado, sincronizado em resposta à avaliação de um determinado evento, possuindo duração e intensidade das reações corporais e ações comportamentais. A duração da emoção deve ser relativamente curta, devido à grande mobilização de ações e respostas corporais, já a intensidade de padrões de resposta a uma emoção está associada à adaptação comportamental (Scherer, 2005).

Além disso, para Scherer (2005), a emoção ainda envolve o sistema nervoso central, o sistema neuroendócrino, o sistema nervoso somático e o sistema nervoso autônomo, podendo atuar isoladamente ou em conjunto sobre o componente emocional. O componente emocional pode ser considerado o componente cognitivo, neurobiológico, motivacional, de expressões faciais e do sentimento subjetivo.

No contexto de Damásio (2011, p. 142), a definição de emoção é embasada na neurobiologia:

Emoções são programas de ações complexos e em grande medida automatizados, engendrados pela evolução. As ações são complementadas por um programa cognitivo que inclui certas ideias e modos de cognição, mas o mundo das emoções é, sobretudo, feito de ações executadas no nosso corpo, desde expressões faciais e posturas até mudanças nas vísceras e meio interno.

Para Damásio (2000), as emoções originam sentimentos, porém, nem todo sentimento é provocado por uma emoção. Nesses casos, seguem como exemplo o sentimento de honra, sentimento de justiça, sentimento de dever cumprido, dentre outros. Damásio (2000, p. 64) apresenta a diferença entre esses dois termos, citando que o termo emoção é usado para designar um conjunto de reações publicamente observáveis, ocorrendo no rosto, na voz, ou em comportamentos específicos. Já o sentimento é considerado a experiência mental privada de uma emoção, como as imagens mentais que são propriedades privadas do organismo em cujo cérebro ocorrem (DAMÁSIO, 2004).

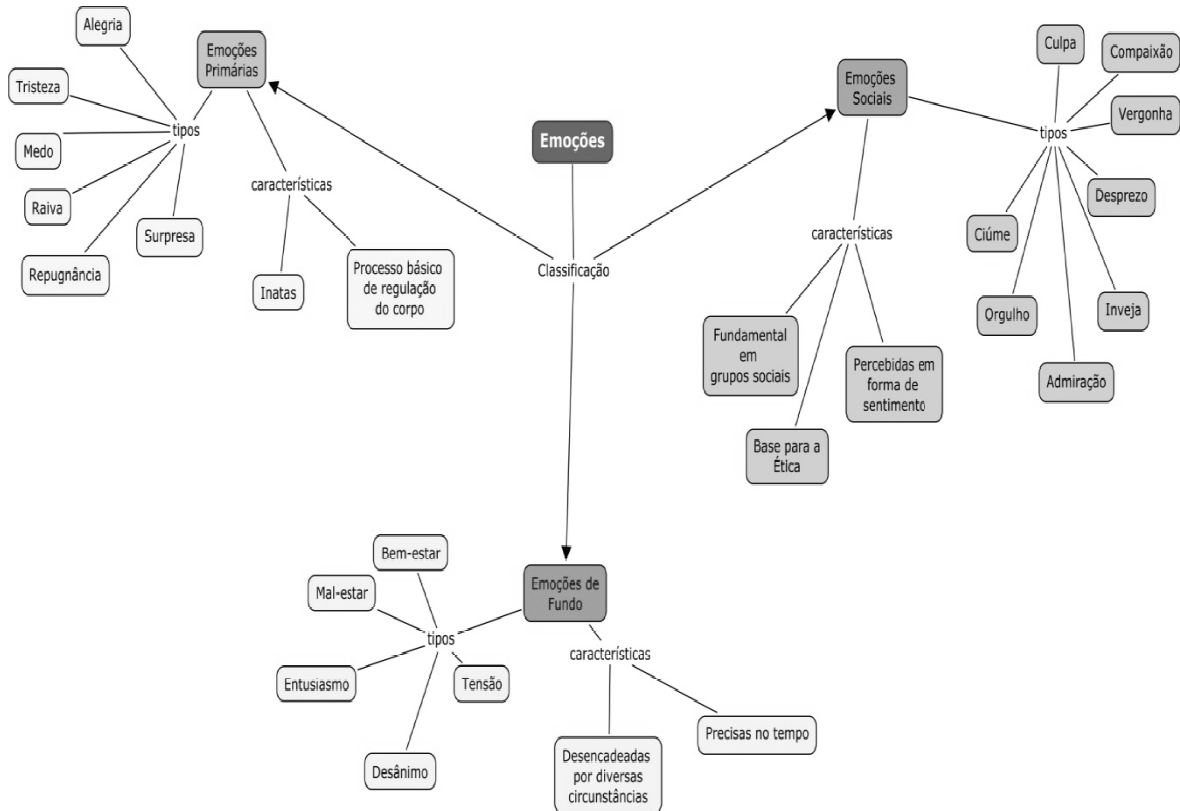
As emoções foram classificadas por Damásio (2004, p. 52-53; 2011, p. 161) em três categorias: emoções primárias, emoções de fundo e emoções sociais, observadas a seguir:

- a) As emoções primárias ou universais: consideradas inatas, de alta intensidade e fáceis de identificar nas mais variadas culturas e situações. Constituem o processo básico de reação, quando certos eventos são detectados, desencadeiam a ativação de um estado do corpo. Incluem: medo, raiva, nojo, surpresa, tristeza e felicidade. Scherer (2005) reforça que essas emoções facilitam a nossa adaptação ao meio em que vivemos;
- b) As emoções de fundo: podem ser desencadeadas por diversas circunstâncias e são consideradas próximas dos estados de humor², mas diferem-se por serem mais precisas no tempo. Incluem bem-estar/mal-estar, calma/tensão, entusiasmo/desânimo;
- c) As emoções sociais ou secundárias: incluem simpatia, compaixão, embaraço, vergonha, culpa, orgulho, ciúme, inveja, gratidão, admiração, espanto e desprezo. Possuem papel fundamental em grupos sociais, incorporam princípios morais e formam a base para a ética, além de serem percebidas em forma de sentimentos.

A Figura 2, a seguir, apresenta um mapa conceitual sobre a categorização das emoções.

² A noção de estados de humor e/ou estados de ânimo foi traduzida a partir do termo *states of mood*, fundamentado sob a ótica de Scherer.

Figura 2: Mapa conceitual



Fonte: Elaborado pela autora a partir de Damásio (2004; 2011).

Salienta-se que as emoções primárias manifestam-se espontaneamente e possuem alta intensidade, enquanto as emoções secundárias podem ser percebidas por mais tempo. Porém, conforme apresentado, as emoções de fundo são as que duram por mais tempo e influenciam na tomada de decisão.

Scherer (2005) também justifica a necessidade de classificar as emoções, não só para mapear os termos semanticamente e cientificamente, mas, principalmente, devido à importância de reconhecer os componentes do processo emocional.

Para Scherer, o processo emocional é o que pode provocar determinada emoção, o que uma pessoa faz de acordo com o evento disparador da emoção, bem como as suas consequências, padrões de respostas, impacto comportamental, intensidade e duração da experiência associada à emoção. De acordo com a emoção, esse processo emocional pode ser alterado e se, por exemplo, a emoção for o medo, a resposta pode vir em forma de grito, choro

intenso, podendo durar segundos ou perdurar mais tempo, instigando, por exemplo, uma tomada decisão que faz a pessoa a correr.

Nesse sentido, ao apresentar os termos emoções e sentimentos, torna-se necessário versar, também, sobre a *afetividade*. De modo geral, o termo *afetividade* engloba os termos emoções, sentimentos, paixão, personalidade, motivação, estados de ânimo, entre outros (BERCHT, 2001). A *afetividade* diz respeito à habilidade de experimentar sentimentos e a eles reagir, podendo incluir um conjunto de *fenômenos afetivos* que possuem diferentes origens, intensidade, duração, impacto comportamental e reações corporais (BERCHT, 2001; LONGHI, 2011, p.32; SCHERER, 2005).

No contexto desta pesquisa, que se baseia em Damásio (2004, 2011) e Klaus Scherer³ (2005), os termos *fenômeno afetivo* e *estado afetivo* são considerados como um conjunto de condições ou disposições psicológicas e biológicas ligadas à afetividade, influenciando a forma como são encaradas as situações e acontecimentos.

Bercht (2001, p.73) considera *estados afetivos* como um elemento da afetividade o qual se conceitua por “todo o domínio das emoções propriamente ditas, dos sentimentos das emoções, das experiências sensíveis e da capacidade de se poder entrar em contato com sensações”, além de serem afetados por desejos, experiências e aprendizado.

De acordo com Scherer (2000), os *estados afetivos* incluem a emoção, estados de ânimo (*mood*), posturas interpessoais, atitudes e traços de personalidade e são descritos de acordo com a intensidade, duração, sincronização a um evento, foco no evento, mudança de estado associada a uma situação, mudança de acordo com a natureza do estado, grau em que um estado afeta o comportamento.

Fundamenta-se, dessa forma, que os estados de ânimo são considerados estados afetivos e, na presente pesquisa, serão utilizados a partir dos pressupostos de Scherer (2000, 2005), enquanto são usados por meio de manifestações do estado de ânimo via autorrelato e comportamento observável de variáveis do ambiente.

³ Professor do Departamento de Psicologia da *University of Genève* e Pesquisador do *Swiss Center for Affective Sciences*.

2.2 ESTADO DE ÂNIMO

Scherer tem desenvolvido pesquisas abordando as emoções e, em especial, os estados de ânimo (*states of mood*⁴). Os estados de ânimo fazem parte dos fenômenos afetivos, são tidos como de baixa intensidade, perduram por mais tempo, se enquadram nas emoções de segundo plano ou emoções de fundo, conforme a categorização apresentada na seção anterior. Scherer (2000, 2005) detalha o estado de ânimo como um estado afetivo difuso, destacado como uma mudança de sentimentos subjetivos, de baixa intensidade, duração longa e sem causa aparente, podendo afetar o comportamento de uma pessoa.

Para Picard (2000), Damásio (2000) e Longhi (2011), os estados de ânimo são os que mais influenciam na tomada de decisão e na aprendizagem, uma vez que ocorrem mais comumente nas atividades de aprendizagem dos alunos.

Nesse sentido, justifica-se e fortalece a motivação para a investigação dos estados de ânimo nesta tese e o desenvolvimento de um Modelo de Predição do aluno desanimado quando em interação em um AVEA.

Conforme apontado por Scherer (2005), não há métodos objetivos de medição do estado afetivo experimentado por uma pessoa durante determinado evento e não há meios para obter esta inferência a não ser perguntando ao indivíduo. Nesses casos, podem ser oferecidas listas categorizadas de termos afetivos a fim de se obter informação qualitativa sobre o estado afetivo experimentado. A partir disso, Scherer (2005) sugeriu uma lista com emoções categorizadas em termos afetivos para ser empregada em pesquisas que usam respostas livres ou autorrelatos para emoções sentidas pelos participantes em forma de representação gráfica.

A proposta de Scherer (2005) culminou na representação, para medição do estado afetivo, intitulada GEW (*Geneva Emotion Wheel*). A GEW possui representação gráfica, é de fácil utilização e permite ao usuário identificar a emoção vivenciada de uma forma simples e visual. A GEW apresenta os termos afetivos que estão agrupados segundo a sua valência (positiva/prazerosa ou negativa/desagradável ou construtora/desconstrutora) e o grau de controle ou potencial de *coping*⁵ (alto/baixo), auxiliando na identificação do estado de ânimo.

⁴ *States of mood* foi um termo de introduzido por Aristóteles e traduzido na presente tese como “estados de ânimo”, a partir da perspectiva de Scherer (2000).

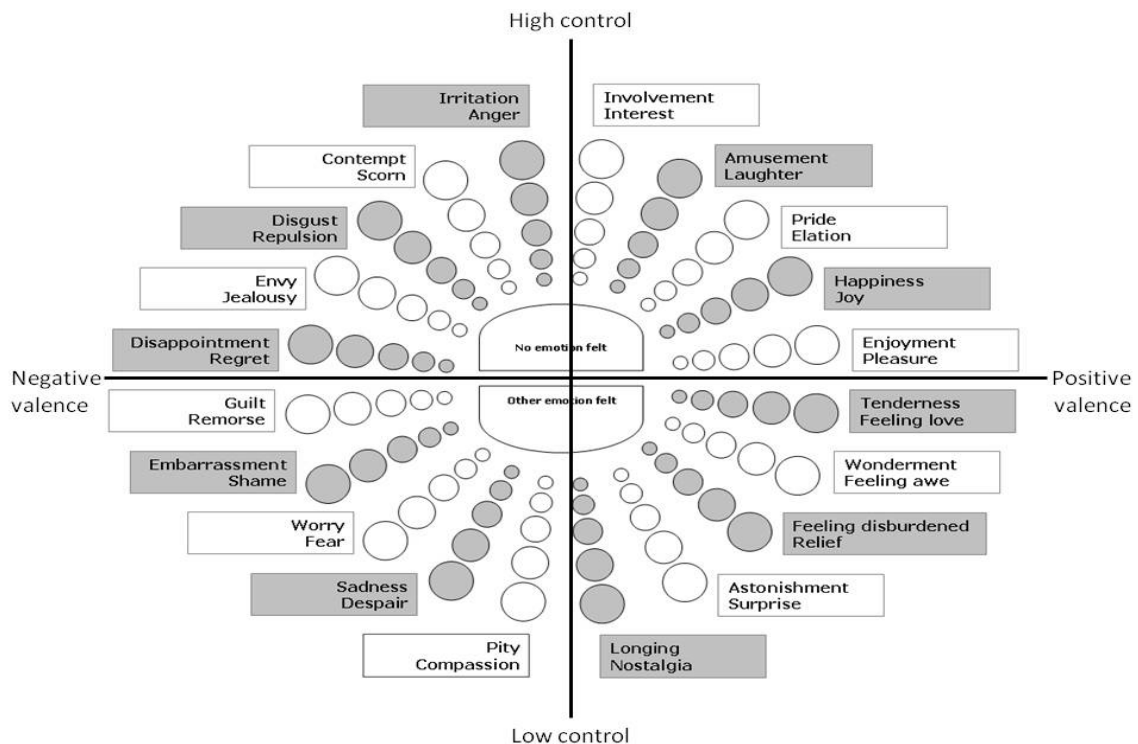
⁵ Potencial de *Coping* diz respeito à capacidade que um indivíduo tem de lidar com as situações, capacidade de superação e enfrentamento.

A GEW pode ser aplicada para respostas livres ou autorrelatos, cuja vantagem é a representação de termos afetivos de forma discreta, possuindo um número limitado de famílias afetivas e uma escala correspondente à intensidade dos sentimentos (SACHARIN; SCHLEGEL; SCHERER, 2012).

Ressalta-se, ainda, que, de acordo com Scherer (2000, 2005), a possibilidade de identificar o estado de ânimo via autorrelato, após a ocorrência de determinados eventos que o dispararam, só é possível porque os estados de ânimo duram mais tempo do que as emoções. Ainda nesse sentido, Scherer (2000, 2005) chegou a um conjunto de termos afetivos mapeados na GEW, o qual envolveu longos períodos de estudos e experimentos que permitem utilizar a GEW como instrumento de apoio na identificação dos estados de ânimo, pois, quando solicitado ao aluno, há a possibilidade de identificar o sentimento vivenciado através de um dos termos associados às famílias afetivas dispostas na GEW.

Na construção de Scherer (2005), uma família afetiva refere-se aos termos aplicados para a representação de uma emoção na GEW e está representada por duplas de termos nas 20 famílias afetivas disponibilizadas. A Figura 3 apresenta a GEW 2.0.

Figura 3: Geneva Emotion Wheel



Fonte: Sacharin, Schlegel e Scherer (2012).

O levantamento do autorrelato diz respeito a um questionamento respondido pelo participante sobre um passado próximo, fazendo com que o indivíduo relate, por meio da GEW, qual emoção ou termo afetivo ele sentiu durante determinado momento de aprendizagem, bem como a sua intensidade. Para Bower e Forgas (2000, p. 104), o registro de determinado evento na memória está associado ao armazenamento de determinada emoção vivenciada naquele momento, o que explica o fato de, ao lembrar-se de um evento passado, uma pessoa voltar a sentir um estado afetivo equivalente ao de quando realmente o evento ocorreu.

Nesse sentido, além de os estados de ânimo terem duração temporal relevante e influenciarem os processos de decisão e aprendizagem, eles podem ser utilizados via autorrelato, por serem evocados quando determinadas memórias forem lembradas, fazendo-se presente no momento do autorrelato.

Atualmente, a GEW encontra-se na sua segunda versão (GEW 2.0), a qual dispõe de 40 termos afetivos dispostos em pares e 20 famílias afetivas divididas em quatro quadrantes (SACHARIN; SCHLEGEL; SCHERER, 2012). Além de selecionar uma família afetiva, é possível escolher a intensidade sentida, o que é realizado de acordo com os círculos que aumentam de tamanho à medida que se distanciam do centro da roda, permitindo mensurar a intensidade do termo afetivo em cinco níveis. Dessa forma, quanto maior o círculo, mais intensamente foi sentido determinado termo afetivo, permitindo a avaliação da intensidade do estado de ânimo vivenciado.

Os termos “nenhuma” e “outra” estão dispostos na GEW caso o participante não tenha experienciado ou não tenha reconhecido nenhum dos termos das respectivas famílias afetivas para o sentimento vivido. Nesses casos, o participante pode optar por escolher a opção “outra” quando teve algum sentimento cujo termo não estiver presente ou, até mesmo, se não teve nenhum sentimento associado, pode recorrer à opção “nenhuma”.

Algumas aplicações da GEW já foram realizadas e dentre elas destacam-se as seguintes pesquisas:

- a) O trabalho desenvolvido por Caicedo e Beuzekom (2006) faz uma comparação/avaliação entre várias ferramentas para medir emoções relacionadas a produtos de consumo. Dentre as ferramentas utilizadas, a melhor forma de avaliação foi atribuída à GEW;
- b) No contexto educacional, uma das primeiras aplicações da GEW 1.0 foi o trabalho desenvolvido por Longhi (2011), o qual apresenta uma adaptação da GEW para a

REA (Roda de Estados de Ânimo) e a utiliza para compor a dimensão afetiva de alunos em um AVEA, juntamente com traços de personalidade e mineração textual.

No presente trabalho, o enfoque está na identificação das variáveis comportamentais relacionadas com o estado de ânimo desanimado, cuja inferência foi realizada por meio de comportamento observável do aluno, via variáveis comportamentais obtidas das interações do AVEA Moodle e manifestação sobre o estado de ânimo via autorrelato, vindos dos questionários utilizados com a GEW 2.0.

2.3 ESTADO DE ÂNIMO DESANIMADO

A palavra desânimo, etimologicamente, significa falta de sopro de vida, já psiquicamente, relaciona-se ao consumo excessivo de reservas emocionais e motivacionais que, quando consumidas, faz surgir o sentimento de que não há nada mais a ser feito, culminando em esgotamento (JBEILI, 2004). O autor descreve que quem vive o desânimo não aceita desafios, esquivava-se da ansiedade provocada pelas tentativas, tenta evitar repetições de fracasso, o que torna a pessoa apática e indiferente, fazendo com que evite tomar decisões.

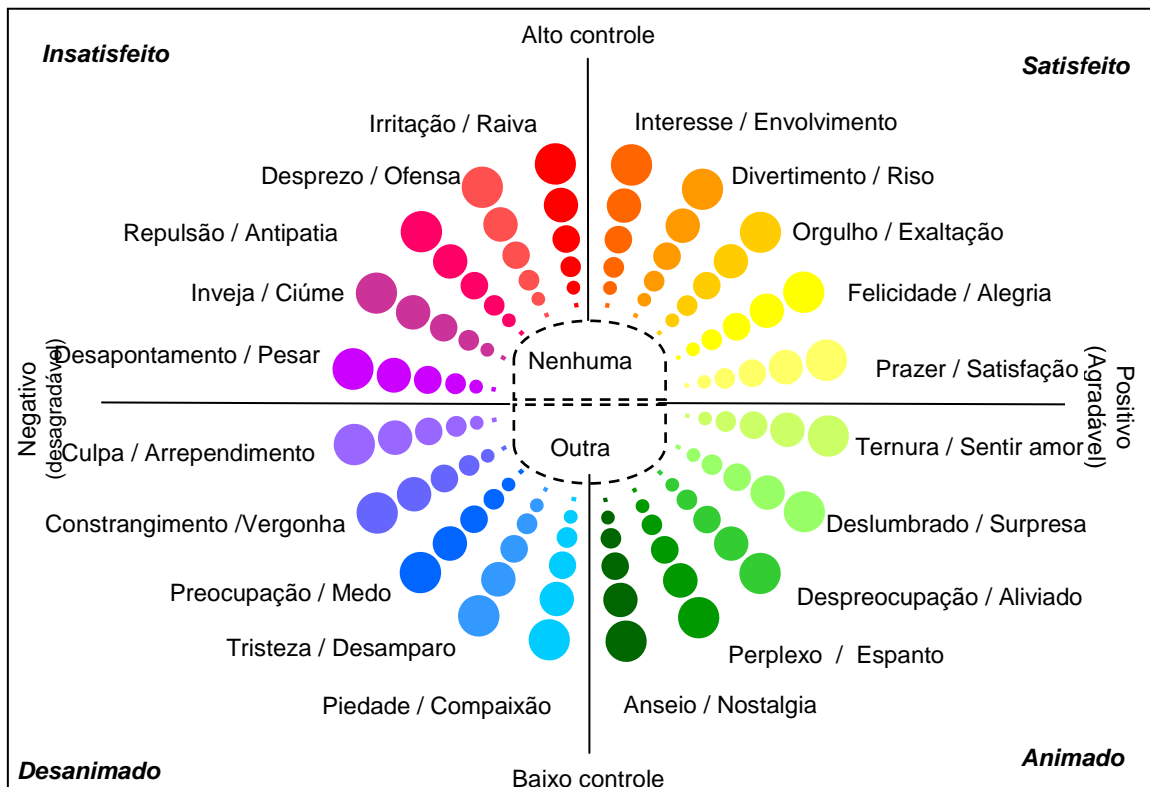
Scherer (2005) considera o desânimo como um estado de ânimo de baixo controle e valência negativa e é representado em um dos quadrantes da GEW 2.0. O uso da GEW 2.0 foi adaptado⁶ para o português, utilizando o acrônimo REA (Roda de Estados de ânimo) 2.0.

A presente roda foi adaptada a partir de Sacharin, Schlegel e Scherer (2012) e Longhi (2011) e é dividida em quatro quadrantes (satisfeito, insatisfeito, animado e desanimado), que representam famílias afetivas classificadas em: logrativas (satisfação), antagônicas (insatisfação), resignativas (desânimo) e condutivas (animado).

A REA 2.0 pode ser observada por meio da Figura 4.

⁶ A adaptação da GEW 2.0 para o português foi realizada por Renata Bonotto, especialista em Linguística Aplicada e Dimensão Afetiva nos Processos Educacionais.

Figura 4: Roda de Estados de Ânimo 2.0 / GEW 2.0



Fonte: Adaptado pela autora a partir de Sacharin; Schlegel; Scherer (2012, p. 3) e Longhi (2011, p. 169).

Na REA 2.0, há um conjunto de famílias afetivas apresentadas por meio de termos afetivos mais representativos para cada uma, os quais estão dispostos ao participante. Além das 20 famílias afetivas, há a opção “nenhuma”, caso o participante não se identifique com nenhum dos termos apresentados. Caso o participante tenha sentido outro afeto que não esteja representado, o mesmo pode selecionar a opção “outra” e relatar o afeto sentido.

Para cada família afetiva representada na REA 2.0, há um rol de termos que são utilizados segundo sua similaridade de significados, apresentando a mesma equivalência para as famílias afetivas e afetos que são apresentados na roda.

No caso particular do estado de ânimo desanimado, esse detalhamento é apresentado no Quadro 1.

Quadro 1: Fenômenos afetivos relacionados ao estado de ânimo desanimado – REA 2.0

ESTADO DE ÂNIMO	FAMÍLIA AFETIVA	TERMOS AFETIVOS
Desanimado	Culpa / Arrependimento	Culpado, arrependido, irresponsável, faltoso, pesaroso,
	Constrangimento/Vergonha	Envergonhado, acanhado, humilhado, constrangido, dificuldade, desapontamento
	Preocupação/Medo	Tenso, amedrontado, ansioso, apavorado, assustado, receoso, temeroso, intimidado, aflito, preocupado
	Tristeza / Desamparo	Entristecido, abatido, choroso, magoado, melancólico, sofrido, deprimido, aflito, fracassado, frustrado, sentir-se incompetente, pesaroso, mal humorado, incomodado, abatido, arrependido, desesperança
	Piedade/Compaixão	Com pena, piedoso, compaixão, consentimento, dó, lástima, compreensão

Fonte: Adaptado a partir de Scherer (2000) e Longhi (2011).

No âmbito da educação, estar em um dos estados de ânimo apontados por Scherer (2005) pode influenciar positivamente ou negativamente na aprendizagem e estar no estado de ânimo desanimado sugere que o aluno demonstra ou reprime, de alguma forma, as emoções de tristeza, medo, vergonha ou culpa, bem como os possíveis termos afetivos que as representam (LONGHI, 2011). Para a autora, no estado de ânimo desanimado, há uma tendência de conduzir o aluno à desistência ou submissão, porém, podem ocorrer implicações desejáveis e reversíveis para o aluno, como ter um tempo para repensar suas atitudes ou readaptar-se às novas condições, sendo um estado de ânimo passível de enfrentamento e mudanças, designado por emoções resignativas.

Na pesquisa de Brand; Reimer; Opwis (2007), os autores dividem indivíduos em dois grupos, os que estão no estado de ânimo positivo e os que estão no estado de ânimo negativo. Para os autores, os indivíduos que estão no estado de ânimo negativo estão com mais dificuldade em apresentar uma solução a um problema, levam mais tempo para realizar determinada tarefa e repetem muitas vezes uma atividade até encontrar a sua solução. Dessa forma, o estado de ânimo negativo repercute negativamente no sucesso da aprendizagem.

Picard (2000), ao abordar as emoções e a aprendizagem, relata que a aprendizagem pode ser abandonada por sentimentos negativos, ou seja, pode levar o aluno à desistência.

É nesse sentido que se acredita que, ao inferir o desânimo do aluno em interação no ambiente, há possibilidades de se evitar a desistência ou abandono do aluno na disciplina/curso, e o principal, a possibilidade de que o aluno saia desse estado, com o apoio do professor, e retome o seu aprendizado.

2.4 TRABALHOS ASSOCIADOS

Há mais de uma década, a Computação Afetiva (CA) foi apresentada pela pesquisadora Picard (2000), pois esta estava preocupada em fazer com que computadores pudessem reconhecer emoções, expressar emoções e responder adequadamente às emoções de quem interagisse por intermédio deles, devido ao fato das emoções serem essenciais no processo de tomada de decisão, na percepção, influenciando o aprendizado, além da memória e processos cognitivos.

Uma das áreas de aplicação da CA é a Educação, visto que pode auxiliar na solução de desafios da aprendizagem ao utilizar métodos e técnicas que permitam desenvolver programas que reconheçam emoções, podendo subsidiar o professor, prover assistência personalizada ao aluno, prever possível evasão, inferir frustração do aluno em um AVEA, identificar o tédio e sua influência e propensão ao abandono, dentre outras (PICARD, 2000; JAQUES; NUNES, 2012; KAMPPFF, 2009; BAKER, *et al.* 2011; IEPSEN, 2013, RIGO *et al.*, 2014).

Nesse sentido, Jaques e Nunes (2012) apontam que trazer características psicológicas humanas para o ambiente virtual, basicamente relacionado aos afetos negativos, proporciona à EAD a diminuição do *gap* existente entre o ensino presencial e o virtual, principalmente no que diz respeito à CA. Aspectos relacionados com a afetividade do aluno estão presentes em sala de aula presencial ou à distância. Porém, no ambiente à distância, muitas vezes, a forma da expressão desses estados afetivos encontra-se presente no ambiente educacional no qual estão inseridos. Conforme apontado por Kort, Reilly e Picard (2001), em ambientes informatizados, os pesquisadores começam a dar aos computadores habilidades de reconhecer expressões afetivas e, mais cedo ou mais tarde, haverá capacidade de reconhecer comportamentos humanos que levam a inferências sobre o estado afetivo do aluno.

A influência da afetividade na aprendizagem pode ser observada pelos professores em todos os níveis escolares, podendo, inclusive, desencadear o abandono do curso ou disciplina por causa de sentimentos negativos, como por exemplo a confusão, frustração e ansiedade (PICARD, 2001). Nesse mesmo sentido, Piaget (2014) afirmava que não há mecanismos cognitivos sem afetos, pois os afetos sempre intervieram na inteligência, confirmando a importância envolvida entre a inteligência, aprendizagem e afetos.

Ainda corroborando com o exposto, Lago (2011), em suas pesquisas sobre a influência da afetividade na aprendizagem, aponta que, ao abordar o pensamento como algo autônomo, independente dos motivos, dos interesses, das inclinações e das emoções do indivíduo,

apresenta-se uma postura danosa para a educação. Nesse mesmo sentido, a autora afirma que “Estados e fatores emocionais como raiva, ansiedade, desmotivação e falta de autoestima podem limitar, se não impossibilitar, o que é compreendido e aprendido” (LAGO, 2011 p. 51).

Pesquisas desenvolvidas na área educacional por Mastrella-de-Andrade (2011) e Mastrella-de-Andrade e Norton (2011) revelam a influência da afetividade no processo de aprendizagem, bem como a evidência do papel da motivação na aprendizagem, ao abordar a ansiedade no processo de aprendizagem. No trabalho de Figueiredo (2011), os fatores afetivos apresentam-se como fundamentais no processo de escrita e correção de erros na aprendizagem de línguas estrangeiras.

Além desses, destaca-se a pesquisa de Sabourin e Lester (2014), a qual aponta que o envolvimento e a motivação dos alunos têm se mostrado fundamental na melhoria da aprendizagem em AVEA. Os autores enfatizam que estados afetivos positivos possuem uma correlação positiva na aprendizagem e os estados afetivos negativos possuem o efeito oposto. Sabourin e Lester aplicam a CA ao desenvolver ambientes baseados em jogos para promover maior engajamento do aluno, em consequente, afetos positivos foram evidenciados e influenciaram positivamente a aprendizagem dos alunos.

Majadas, Santos e Boticario (2013) estão em processo de desenvolvimento de uma pesquisa que se concentra em identificar e modelar estados afetivos para detectar a emoção a fim de fornecer *feedback* personalizado ao aluno. Os autores convergem esforços na tentativa de utilizar um maior número de dispositivos (sensores fisiológicos, reconhecimento de expressões faciais, traços de personalidade, entre outros) que reconheçam emoções, a fim de contemplar a evidência da emoção em todas as situações em que possam ocorrer no ser humano. É proposto o uso de dados fisiológicos para obter a frequência cardíaca e a temperatura, aplicação do questionário de inferência de traços de personalidade SAM⁷ (*Self-Assessment Manikin*) e mineração de textos para identificar palavras com conotação afetiva. Inicialmente, foi feito um experimento em que se utilizou a mineração de texto para identificar palavras com conotação afetiva, o questionário SAM, reconhecimento de expressão facial vinda do *Kinect*⁸ e o tempo de pressionar e soltar teclas do mouse. Com a junção de todos os dados, os mesmos foram minerados através dos algoritmos *Nayve Bayes* e C4.5. Os resultados sugeriram que fossem usadas formas menos invasivas para a detecção da emoção e que dados vindos de sensores precisam ser remodelados a fim de se obterem

⁷ SAM (*Self-Assessment Manikin*) – questionário usado para identificar satisfação, motivação e sentimento de domínio quando um indivíduo usa uma aplicação computacional.
<<http://lifes.dc.ufscar.br/competicaoihc2012/sam.html>>

⁸ Dispositivo capaz de reconhecer movimentos usado comumente em videogame Xbox.

melhores resultados. Como essa pesquisa está em andamento, futuramente, os autores pretendem desenvolver um modelo de usuário baseado em padrões para personalização ao se considerar o estado afetivo do aluno.

Na pesquisa de Ocumpaugh, Baker e Rodrigo (2012), é apresentado o BROMP (*The Baker Rodrigo Ocumpaugh Monitoring Protocol*) como um protocolo/método de observação de estados afetivos e comportamento observável do aluno em interação em um ambiente educacional, a fim de registrar o comportamento do estudante diretamente nessa ferramenta, na maioria das vezes sincronizada com dados de *log* do aluno. Esse protocolo é preenchido por um observador, podendo ser ou não o professor, cuja finalidade é identificar situações em que o aluno necessite de intervenção, como, por exemplo, quando apresentar frustração, confusão, tédio, entre outros. Atualmente, está em sua segunda versão, cujo protocolo é capaz de inferir aspectos relacionados à afetividade, quando o aluno encontra-se em confusão, tédio, frustração e concentração, assim como já foi integrado a várias aplicações educacionais como *Aplusix*, *Cognitive Tutor Algebra*, *Chemistry Virtual Laboratory*, entre outros (OCUMPAUGH; BAKER; RODRIGO, 2012, 2015).

A pesquisa de Bastos (2012) aborda a presença social⁹ e permite observar o grau de envolvimento, interação do aluno no AVEA, o acompanhamento dos estados afetivos dos alunos em um AVEA, principalmente nos eventos comunicativos de fóruns e chats que possibilitam a expressão da afetividade em texto, a fim de fortalecer o sentimento que o indivíduo percebe de pertencimento com relação ao seu grupo.

Observa-se, dessa forma, que há pesquisas que abordam aplicações para a inferência do estado afetivo, como o caso do trabalho desenvolvido por Iepsen (2013). Nesse trabalho, o autor buscou a inferência da frustração no contexto da disciplina de algoritmos, dadas as altas taxas de evasão, vindo a contribuir no processo de aprendizagem, pois, ao identificar os alunos em estado afetivo frustrado, se consegue minimizar a desistência na disciplina (IEPSEN, 2013).

No trabalho de Kapoor, Burleson e Picard (2007), a frustração também foi inferida quando o aluno utilizava um Sistema Tutor Inteligente (STI) e, dessa forma, poderia prover sistemas mais inteligentes ao fornecer respostas apropriadas para quem estivesse neste estado afetivo.

De forma geral, as pesquisas supracitadas evidenciam a importância de considerar aspectos relacionados à afetividade no contexto educacional, como o aluno se sente e como a

⁹ Presença Social é definida como a manifestação verbal e percepção da afetividade e interatividade dos sujeitos com relação ao AVEA (BASTOS, 2012).

afetividade pode vir a afetar a sua aprendizagem, através do reconhecimento de estados afetivos, os quais podem influenciar mais ou menos, positivamente ou negativamente a aprendizagem do aluno, além de buscarem formas de identificar e reconhecer emoções no contexto educacional.

Como pode ser observado com as pesquisas apresentadas, há várias aplicações da CA no âmbito educacional, as quais buscam o desenvolvimento de sistemas que auxiliem na inferência das emoções nas mais variadas situações de aprendizagem, e que vêm ao encontro do que foi desenvolvido na presente pesquisa, mostrando-se como uma área em expansão, principalmente em contexto nacional.

Porém, o Modelo de Predição do aluno desanimado proposto nesta tese constitui uma ampliação de reconhecimento de vários termos afetivos (Quadro 1) característicos do estado de ânimo desanimado, não só o caso da frustração, conforme apresentado em trabalhos supracitados. Esta pesquisa também exige a necessidade de apoio de psicólogos para a aplicação de questionários, como o caso do SAM, sugerindo que o uso da REA 2.0 e os dados comportamentais dos alunos no AVEA sejam satisfatórios para a predição do desânimo.

Quando Sherer (2000, 2005) propôs a GEW 2.0 como forma de aplicação em autorrelatos, fundamentou-se em pesquisas as quais garantem que, mesmo um evento tendo ocorrido há tempos, a memória resgatada e os termos apresentados na GEW 2.0 são representativos dos estados de ânimos vivenciados no passado.

Salienta-se, que o Modelo de Predição proposto foi implementado computacionalmente, integrado ao AVEA Moodle, e pode ser utilizado por qualquer professor que utilize a ferramenta. Também há a possibilidade de se visualizar os alunos desanimados por meio de uma ferramenta visual, utilizando *dashboards*, o que propicia uma interpretação visual dos resultados, sugerindo maior facilidade de análise do modelo.

2.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

No presente capítulo, foi exposto um conjunto de definições sobre emoções e suas classificações, bem como sentimentos e estado afetivo, delimitando os conceitos e apresentando uma visão geral sobre eles. Versou sobre a computação afetiva, bem como apresentou os trabalhos relacionados da área e a influência da afetividade na aprendizagem.

Também foi apresentado o modelo de Scherer (2005) em relação aos estados de ânimo, principal linha teórica da pesquisa, e a nova versão da GEW 2.0, a qual foi adaptada para REA 2.0. Foi realizado um levantamento dos termos afetivos aplicados às famílias afetivas representativas do estado de ânimo desanimado e apresentado em formato tabular (Quadro 1).

A partir do aperfeiçoamento da REA 2.0, foi desenvolvido um questionário para apoio na verificação de variáveis relacionadas ao desânimo (Apêndice B) e aplicado aos experimentos descritos no capítulo 4.

Como o enfoque é o desânimo, percebe-se que ser, estar ou sentir-se desanimado no processo de aprendizagem é algo que acomete os estudantes, podendo levá-los ao processo de desistência de uma atividade ou até mesmo de um curso. Este fato pode ser derivado de várias causas – vergonha por não ter entregado atividades no prazo estipulado, culpa ou vergonha por não ter conseguido acompanhar um conteúdo, entre outras. Porém, uma questão importante nestes casos é que esse processo pode ser revertido se ocorrerem interferências por parte do professor ou de forma automática, por um sistema inteligente, tornando-se fator motivador para o desenvolvimento da tese.

3 TÉCNICAS COMPUTACIONAIS

Neste capítulo, é exposto um panorama geral sobre mineração de dados e, mais detalhadamente, sobre mineração de dados educacionais, incluindo conceitos, ferramentas e técnicas, apresentando-se como técnica computacional adotada para apoio à inferência do desânimo a partir dos dados das interações dos alunos no AVEA.

3.1 MINERAÇÃO DE DADOS

Com os avanços tecnológicos, a queda do preço do *hardware* e o aumento da capacidade de armazenamento, muitas instituições, desde as mais simples até as mais complexas, começaram a armazenar suas informações em meio digital. A *EMC Corporation*¹⁰, empresa americana conhecida por desenvolver soluções para armazenamento dados, divulgou, em abril de 2014, um levantamento¹¹ em que o universo digital cresce cerca de 40% ao ano e a previsão é de que alcance 44 trilhões de *gigabytes* em 2020.

Observa-se, assim, um aumento evidente do volume de dados produzido. Porém, esse volume de dados precisa fazer algum sentido, gerar alguma informação que seja útil a fim de servir de apoio em tomadas de decisões.

De acordo com Tan, Steinbach e Kumar (2009), para extrair informações dessas bases de dados, podem-se aplicar técnicas tradicionais de consultas em bases de dados relacionados à área de recuperação de dados, porém, estas se tornam limitantes quando há necessidade de se apresentar resultados mais elaborados, como busca de padrões, detecção de conjunto de dados, descoberta de conhecimento, entre outros.

Frente a isso, a proposta de realizar a descoberta de conhecimentos nos bancos de dados pode ser efetivada por meio da mineração de dados (MD), um processo de descoberta automática de informações úteis em grandes depósitos de dados (KDD – *Knowledge Discovery in Database*) com o intuito de descobrir padrões úteis e recentes que poderiam permanecer ignorados (TAN; STEINBACH; KUMAR, 2009).

¹⁰ *EMC Corporation* é uma empresa que trabalha com tecnologia de armazenamento de dados. Site: <http://www.emc.com/en-us/index-2016.htm>

¹¹ Disponível em: <http://brazil.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

Para realizar a descoberta de conhecimento em bases de dados é preciso passar pelo processo de converter dados brutos em informações úteis, identificando novas relações entre dados que podem produzir novos conhecimentos e gerar descobertas originais (TAN; STEINBACH; KUMAR, 2009; BAKER; ISOTANI; CAVALHO, 2011). Dessa forma, há várias etapas para a descoberta de conhecimento, que vão desde o pré-processamento até o pós-processamento. A Figura 5 apresenta o processo de descoberta de conhecimento em bases de dados utilizando MD.

Figura 5: Processo de descoberta de conhecimento em banco de dados



Fonte: Adaptado de Tan, Steinback e Kumar (2009).

O pré-processamento é considerado uma das fases mais trabalhosas da mineração de dados, cujo propósito é transformar dados de estrada, considerados brutos, em um formato próprio para a análise. Algumas vezes, é necessário fazer a fusão de dados de várias fontes, realizar a limpeza dos dados e eliminação de ruído e selecionar registros.

No pós-processamento, devem ser apresentados resultados válidos e úteis, que muitas vezes são incorporados a um sistema de apoio a tomada de decisões (TAN; STEINBACH; KUMAR, 2009).

Porém, para descobrir o conhecimento necessário, é preciso ter metas claras e bem definidas de qual conhecimento se deseja obter (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Também é preciso identificar qual tipo de tarefa que a MD busca explorar, a tarefa de previsão ou tarefa descritiva. Resumidamente, a tarefa de previsão tem como objetivo prever valores de uma variável a partir dos valores de outras variáveis e a tarefa descritiva tem o objetivo de derivar padrões que resumam os relacionamentos nos dados (TAN; STEINBACH; KUMAR, 2009).

Na presente tese, foram empregadas as duas modelagens, a modelagem de previsão, que é uma tarefa de previsão aplicando a Classificação, e também a tarefa descritiva, ao empregar análise de associação para descoberta de padrões.

As seções subsequentes descrevem questões relacionadas à mineração de dados educacionais como suporte à descoberta de informações para apoio na área educacional. Também são apresentadas técnicas, tipos de conhecimentos descobertos e ferramentas para mineração de dados educacionais.

3.2 MINERAÇÃO DE DADOS EDUCACIONAIS

A MDE é uma área de pesquisa cujo principal foco é o desenvolvimento de métodos para explorar conjunto de dados coletados em ambientes educacionais (BAKER; ISOTANI; CARVALHO, 2011). Neste enfoque, a MDE desenvolve ou adapta métodos e algoritmos existentes na mineração de dados, aplicando-os na área educacional, por exemplo, para entender dados produzidos por alunos, professores e suas interações ou para verificar a relação de uma abordagem pedagógica e o aprendizado do aluno (BAKER, ISOTANI e CARVALHO, 2011; COSTA *et al.*, 2012; WEBBER, ZAT e LIMA, 2013).

Porém, o desenvolvimento da mineração de dados na área educacional se deu, de forma geral, há pouco tempo. Baker (2014) admite que os primeiros artigos e *workshops* sobre MDE começaram há cerca de apenas uma década e que o primeiro *Journal (Journal of Educational Data Mining)* produzido especificamente na área teve sua primeira publicação em 2009. Uma das razões para esse início tardio é o fato de que a disponibilidade dos dados em formato utilizável na área educacional se deu mais tarde, pois muitos dados eram armazenados em formato impresso, como em fichas, e, quando armazenados eletronicamente, não eram capazes de gerar dados possíveis de minerar, por não guardarem informação semântica necessária, como os dados de cliques do mouse (BAKER, 2014).

Mesmo com a pouca disponibilidade de dados, aqueles que se encontram armazenados em ambientes educacionais estão sendo usados. Na área educacional, há dados que podem ser oriundos de ambientes educacionais ou de apoio ao ensino, que podem armazenar informações referentes aos alunos, professores, disciplinas, cursos, notas ou todo tipo de informação acadêmica de uma instituição. Compreende os ambientes educacionais, desde *softwares* aplicados e utilizados para a gestão educacional como o *Enterprise Resource Planning* (ERP), GVColege¹², até os AVEA como o Moodle¹³, Rooda¹⁴, TopClass¹⁵,

¹² GVColege é um ERP desenvolvido pela GVDasa, cuja especialidade é desenvolver sistemas de gestão educacional.

Claroline¹⁶, entre outros. Também é possível obter dados das interações ocorridas, por exemplo, em simuladores, objetos de aprendizagem, jogos educacionais, fóruns e *chats*.

Mesmo que alguns ambientes educacionais ofereçam relatórios aos professores, tutores e gestores, é difícil extrair informações somente através destes relatórios, como por exemplo, reconhecer padrões de comportamento do aluno no ambiente (GAUDIOSO; TALAVERA, 2006). Apesar de os relatórios possuírem informações relevantes, podem ser considerados superficiais diante da quantidade de dados que estão armazenados e, muitas vezes, pode tornar-se árduo para o professor extrair informação útil através da junção dos relatórios disponíveis (ROMERO; VENTURA; GARCIA, 2008).

De acordo com Romero, Ventura e Garcia (2008), devido à vasta quantidade de dados gerados diariamente nos AVEA, torna-se cansativo para professores e tutores gerenciá-los manualmente, tornando iminente a necessidade de fornecer ferramentas para assisti-los nesta tarefa. Dessa forma, a questão de como explorar adequadamente esses dados a fim de extrair informações e realizar uma análise apropriada é fomentada pela área da mineração de dados aplicados na educação, na busca por respostas a perguntas específicas envolvidas no processo de aprendizagem (RIGO *et al.*, 2014).

Porém, como muitos desses dados estão armazenados localmente, cada um em sua instituição, estas bases de dados são privadas aos pesquisadores locais. Para Baker, Isotani e Carvalho (2011), quando os dados estão disponibilizados publicamente, propicia-se que mais pesquisadores tenham acesso aos mesmos, proporcionando resultados mais precisos para a MDE, além de fomentar o desenvolvimento de pesquisas colaborativas. Exemplos de bases públicas começam a ser difundidos, como o caso do *Cognitive Tutor*¹⁷, que é um sistema tutor inteligente utilizado nos Estados Unidos por mais de 500 mil alunos em aproximadamente 2000 escolas (BAKER; ISOTANI; CARVALHO, 2011).

Outro exemplo de base pública é o repositório aberto de dados e aplicações *Web* conhecido por *Pittsburgh Science of Learning Center (PSLC) DataShop*¹⁸ que possibilita o uso de suas bases para aplicações de MDE. O *DataShop* é especializado em dados sobre as interações entre alunos e professores, incluindo dados de cursos *on-line*, dados de sistemas

¹³ Moodle é um AVEA open source que possui licença GLP, podendo ser usado livremente. Download em: <https://download.Moodle.org/>.

¹⁴ Rooda é um AVEA desenvolvido e mantido pela UFRGS/NUTED para as atividades de graduação e pós-graduação. Mais informações no site: <https://ead.ufrgs.br/rooda/>.

¹⁵ TopClass AVEA disponível em: <https://www.wbtsystems.com/products/topclass-learning-management-system/>.

¹⁶ Claroline é um AVEA open source distribuído sob a licença GLP. Disponível em: <http://www.claroline.net/>.

¹⁷ *Cognitive Tutor* é um sistema tutor inteligente que produz dados de boa qualidade para MD.

¹⁸ PSLC *DataShop* disponível em: <https://pslclatashop.web.cmu.edu/index.jsp>.

tutores inteligentes, dados de ambientes virtuais disponíveis em vários formatos, incluindo arquivos de *log*, o que propicia uma colaboração mais efetiva com a comunidade de *data mining* (KOEDINGER *et al.*, 2011).

No âmbito nacional, os dados encontram-se organizados nas próprias instituições e, por enquanto, não foi viabilizada uma base que consolidasse dados de instituições públicas. Seria necessário desenvolver ferramentas, *scripts*, *plugins* que facilitassem a extração de dados pelas Instituições de Ensino Superior (IES) para a consolidação de várias bases.

Nota-se que a expansão nacional da EAD é evidente em muitas das IES, com destaque especial à UAB¹⁹ (Universidade Aberta do Brasil). Como os cursos da UAB são oferecidos por diversas IES e cada uma usa o próprio AVEA, não há uma base única de dados. A UAB oferta 997 cursos à distância nas modalidades aperfeiçoamento, bacharelado, especialização, extensão, formação pedagógica, licenciatura, sequencial e tecnólogo em 103 IES (FERREIRA, 2013).

Dessa forma, observa-se que há um potencial para se desenvolverem bases públicas no Brasil, principalmente pela UAB, por ser um sistema integrado por universidades públicas que, além de fornecer acesso universal ao ensino, poderia oferecer acesso público aos dados de suas bases, a fim de beneficiar e desenvolver uma cultura orientada por dados, através da aplicação de MDE.

Apesar dessas adversidades, não há empecilhos em desenvolver pesquisas com MDE e explorar dados gerados de instituições, a fim de contribuir para teorias e práticas na educação. Em consonância com o exposto, Baker (2014) assegura que, no campo da educação, tanto a mineração de dados quanto o *learnig analytics*²⁰ possuem um potencial transformador.

Corroborando com o exposto, Rigo *et al.* (2004) justificam o uso da mineração de dados aplicada ao contexto educacional pela possibilidade de gerar padrões de comportamento que representem aspectos importantes para apoiar determinadas práticas pedagógicas.

Sendo assim, Baker (2014) reconhece e defende que MDE tem sido bem sucedida em várias aplicações relevantes para a aprendizagem do aluno em ambientes *on-line* e que, a cada dia, modelos melhores são desenvolvidos, testados e validados para se tornarem mais generalizáveis. Muitas das pesquisas realizadas, que abordam aplicações de MDE, são

¹⁹ Universidade Aberta do Brasil é um sistema integrado de universidades que foi instituída pelo Decreto nº 5.800, de 8 de Junho de 2006, tendo por finalidade expandir e interiorizar a oferta de cursos e programas de educação superior a distância

²⁰ *Learning Analytics* é a medição, coleta, análise e comunicação dos dados dos alunos e seus contextos, a fim de se compreender e otimizar a aprendizagem em AVEA (LIÑÁN; PÉREZ, 2015).

divulgadas através de *Journals*²¹, conferências²², *workshops*²³. Porém, o autor afirma que é preciso criar parcerias entre pesquisadores de MDE e comunidade de sistemas inteligentes.

Nesse sentido, apesar da MDE ser considerada uma área nova, progressos relevantes foram feitos nos últimos anos, conforme citado por Baker (2014). Para o autor, o desenvolvimento de modelos que podem inferir sobre quando o aluno desenvolve a aprendizagem e aplicações de métodos de mineração de dados para modelar o afeto e emoções são pertinentes. Observa-se que a questão da modelagem do aluno relacionada com a afetividade é o tema central pesquisado na presente tese, vindo a corroborar com o exposto por Baker (2014).

Portanto, para utilizar a MDE, é preciso definir a área de pesquisa utilizada bem como as técnicas aplicadas. As seções 3.2.1 e 3.2.2 descrevem conceitos sobre esses dois enfoques e apresentam o que foi adotado na pesquisa.

3.2.1 Áreas da Mineração de Dados Educacionais

A utilização de MDE pode ser feita em grandes volumes de dados e para diversas finalidades educacionais, sendo assim, Baker e Yacef (2009) propuseram a divisão da MDE em quatro áreas, cada qual com uma finalidade de pesquisa e aplicações diferentes. As quatro áreas propostas por Baker e Yacef (2009) são apresentadas a seguir:

- a) A primeira área é destinada ao *modelo do aluno*, que apresenta informações sobre características do aluno, conhecimento atual, motivação, cognição e atitudes. Os autores também relatam o uso da MDE para identificar quando alunos estão frustrados ou entediados, assim como a modelagem do estudante para inferir por predição de insucesso escolar;
- b) A segunda área diz respeito à *descoberta na estrutura do conhecimento*;
- c) A terceira área tem como foco a aplicação de métodos de MDE para *suporte pedagógico* em ambientes de aprendizagem colaborativa, cujo intuito é descobrir

²¹ *Journal of Education Data Mining* (<http://www.educationaldatamining.org/JEDM/index.php/JEDM>).

²² *International Education Data Mining Society* (<http://educationaldatamining.org/>).

²³ *Workshop WMDAVA 2014* (<http://lucianosilvasp.lwsite.com.br/>).

quais tipos de suporte pedagógico são mais eficazes para diferentes grupos de alunos ou em diferentes situações;

- d) A quarta área é destinada a evidenciar e estender as *teorias educacionais* para aprofundar os fenômenos que afetam a aprendizagem, desenvolvendo melhores sistemas de aprendizagem.

A fim de elucidar, a presente pesquisa se enquadra na primeira área, na busca pelo modelo do aluno desanimado. Para Pontarolo (2008, p. 22), a modelagem do aluno diz respeito ao “processo pelo qual são definidas as categorias de informações que representam o aluno no sistema e o tipo de abstração computacional utilizado para organizar e obter essas informações durante o uso do sistema pelo aluno”.

3.2.2 Técnicas Aplicadas na MDE

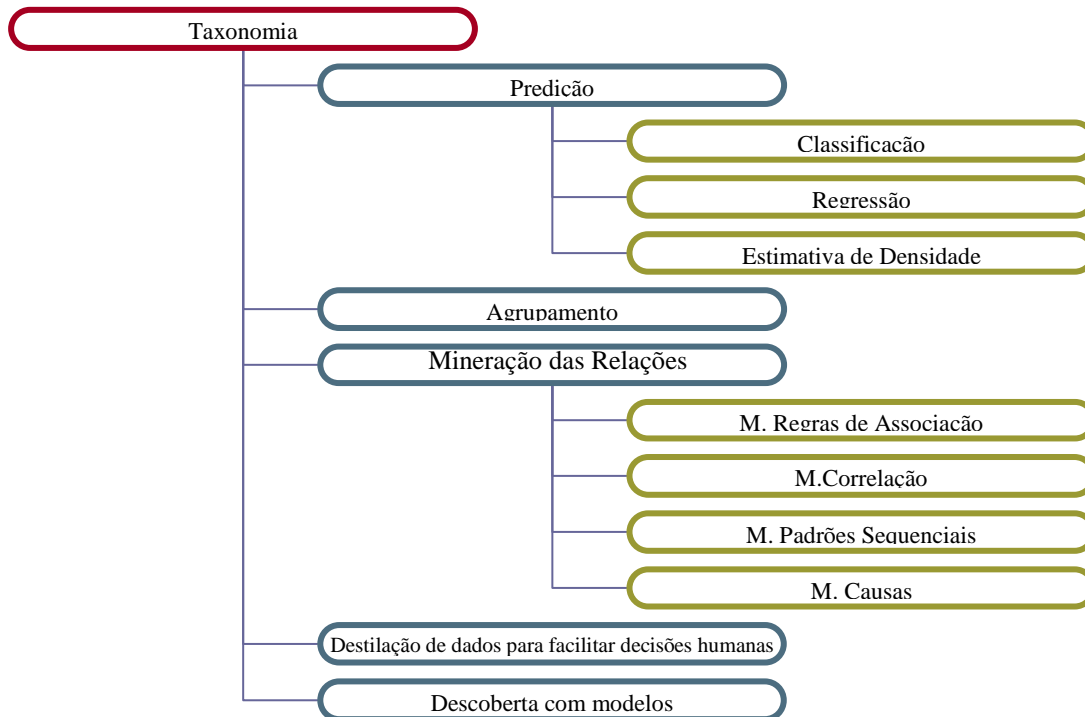
As técnicas empregadas na MDE são oriundas da MD e sua aplicação e escolha dependem do tipo de conhecimento que se deseja descobrir, bem como dos dados disponíveis para realizar a análise. Baker (2010) apresentou uma classificação das técnicas empregadas na MDE em formato de taxonomia, a qual também foi exposta por Baker, Isotani e Carvalho (2011).

A taxonomia é dividida em cinco (5) categorias: a) Predição; b) Agrupamento; c) Mineração das relações; d) Destilação de Dados para Facilitar Decisões Humanas, e e) Descobertas de Modelos. A taxonomia é apresentada a seguir, por meio da Figura 6. Nessa proposta, as três primeiras categorias estão presentes tanto em MD como em MDE.

Ressalta-se que são detalhadas as categorias Predição, Agrupamento e Mineração das Relações. Esta tese dedicou-se ao estudo da Predição (Classificação) e Mineração de Relações (Regras de Associação) por serem compatíveis com o objetivo da presente pesquisa no que diz respeito ao Modelo de Predição do aluno.

A figura 6 apresenta as taxonomias para MDE.

Figura 6: Técnicas de MDE - Taxonomia



Fonte: Adaptado de Baker e Yacef (2009).

a) Predição

A Predição é uma técnica usada para descobrir um valor futuro de uma determinada variável em função de valores de outras. A primeira pode ser chamada de variável preditiva ou dependente e a segunda de independente ou preditora. As variáveis preditoras são as variáveis cuja resposta é o que se quer prever. Um exemplo seria a partir de uma variável x quer-se prever uma variável y , sendo x o nome de uma pessoa e y o seu sexo, considerando que se conhece o nome da pessoa pode-se prever o seu sexo. Neste caso, x é uma variável preditiva e y é a variável preditora (BAKER; YACEF, 2009).

Para Baker, Isotani e Carvalho (2011), os métodos de Predição são usados para estudar quais aspectos de um modelo são importantes para Predição, geralmente aplicados em pesquisas que tentam prever os benefícios educacionais para um conjunto de estudantes. O objetivo é verificar o quanto um aluno aprende sem considerar as diversas variáveis que possam influenciar a aprendizagem, como as variáveis de comportamento. Além disso, esse tipo de técnica pode auxiliar no desenvolvimento de atividades, pois consegue estimar os

benefícios educacionais antes mesmos de eles terem sido aplicados aos alunos (BAKER; ISOTANI; CARVALHO, 2011).

Uma das formas de fazer a Predição é através de tarefas e métodos de Classificação, Regressão e Estimação de Densidade. A Classificação utiliza uma variável categórica e inclui os Métodos de Árvore de Decisão, Regressão Logística, Máquina de Vetores de Suporte (Baker, 2010). Baker (2010) descreve que na regressão a variável prevista é contínua e utiliza métodos de regressão linear, redes neurais, entre outros. Já para o caso da tarefa de estimação de densidade, a variável é prevista em função da densidade de probabilidade e pode utilizar funções gaussianas.

A Classificação visa identificar a qual classe pertence um determinado registro²⁴, ou seja, este método associa um item de dado a uma ou várias classes pré-definidas. Pode ser utilizado, por exemplo, para identificar, em uma escola, qual turma seria mais indicada para um determinado aluno (CAMILO; SILVA, 2009).

A tarefa de Classificação consiste em encontrar um modelo que descreve e distingue classe de dados e conceitos, associa um item de dado a uma ou várias classes pré-definidas. Na Classificação, a Predição é feita usando um atributo classificador. Os modelos podem ser representados: Árvore de Decisão, máquina de vetores de suporte, regras de Classificação, funções matemáticas, redes neurais, entre outros (BAKER; ISOTANI; CARVALHO, 2011; COSTA *et al.*, 2012).

A Classificação pode ser realizada a partir do algoritmo J48, usando dois métodos: o método *Holdout* e o método *Cross-Validation*. Para usar um Modelo Classificador de Predição, é preciso trabalhar com duas bases de dados, sendo uma base de treinamento e uma base de teste. Tan, Steinbach e Kumar (2009) explicam que o modelo do conjunto de treinamento é usado para construir um modelo de Classificação, que subsequentemente é aplicado ao conjunto de testes, e a avaliação de seu desempenho é baseada na contagem de registros de testes previstos correta e incorretamente. Faz parte desse modelo o método *Holdout*.

No método *Cross-Validation* ou validação cruzada, um conjunto de dados é dividido aleatoriamente em K partes, em que uma das partes é escolhida para teste e as demais são usadas para treinamento. Esse procedimento é repetido K -vezes, de modo que cada partição seja usada para teste uma única vez, ou seja, torna-se interativo na medida em que a parte

²⁴ Registro – um conjunto de dados relacionados pertencentes a uma tabela no banco de dados relacional. Pode representar, por exemplo, a informação sobre um aluno (número de matrícula, nome, endereço, turma, curso, entre outras).

selecionada para testes é trocada até que todas tenham sido escolhidas (TAN; STEINBACH; KUMAR, 2009). Por padrão, nesta pesquisa, será utilizado *Cross-Validation* com $K= 10$ folds.

Na presente tese, a proporção de dados usados na mineração seguiu o sugerido por Tan, Steinbach e Kumar (2009, p. 221), em que dois terços dos dados foram reservados para treinamento e um terço para teste. Dessa forma, a precisão do modelo é avaliada com base na precisão do modelo do conjunto de teste ao aplicar o método *Holdout*.

Uma das técnicas de Classificação simples e bastante usada é a Árvore de Decisão. Árvore de decisão é um exemplo de modelo estatístico que usa Classificação e Predição de dados, além de ser um dos algoritmos mais usados para Predição. Dentre os algoritmos de Árvore de Decisão, optou-se pelo uso do J48, por ser um dos algoritmos que gera melhor resultado, além de se mostrar adequado para procedimentos envolvendo variáveis contínuas e discretas, conforme mencionado por Librelotto e Mozzaquatro (2013).

Como há diferentes objetivos em se utilizar a Classificação, os autores Chen *et al.* (2000) aplicaram a Classificação, basicamente através de Árvore de Decisão, para descobrir grupos de alunos com características e reações semelhantes a uma determinada estratégia pedagógica. Outro exemplo foi o trabalho proposto por Cocea e Weibelzahl (2006), o qual realizou a Classificação para identificar alunos com baixa motivação e encontrou ações que foram utilizadas para reduzir taxas de abandono.

Nesta pesquisa, para o Modelo de Predição desenvolvido, foi necessário fazer uma Classificação de alunos desanimados, a fim de obter um modelo para essa Classificação e Predição. Foi aplicado na Classificação o algoritmo J48 juntamente com os Métodos *Holdout* e o *Cross-Validation*, a fim de verificar qual dos métodos encontra melhores resultados.

Técnicas preditivas abordam o aprendizado supervisionado. A aprendizagem supervisionada é muito aplicada em técnicas preditivas, pois são definidos previamente os dados de uma classe específica, ou seja, é preciso definir o atributo classe ou objetivo e os atributos preditivos (DAMASCENO, 2010; REATEGUI; CAZELLA, 2005).

A mineração realizada encontra-se no capítulo 5, que detalha minuciosamente as técnicas e algoritmos aplicados, assim como os resultados obtidos em cada mineração.

b) Agrupamento

Agrupamento ou *Clustering* é uma técnica utilizada para descrever dados e seus relacionamentos. O objetivo é encontrar dados semelhantes entre si. Quanto maior a

homogeneidade dentro de um grupo e maior a diferença entre grupos, melhor será o agrupamento (TAN; STEINBACH; KUMAR, 2009).

O principal objetivo do agrupamento é fazer a divisão dos dados em grupos, ficando estes dados reunidos por semelhança, ou seja, busca associar um item de dados com um ou vários agrupamentos determinados pelos dados, utilizando medida de similaridades (COSTA *et al.*, 2012). Um dos algoritmos mais utilizados na Classificação é o *K-means*.

Algumas aplicações na área educacional que aplicam agrupamento são utilizadas, por exemplo, para agrupar alunos para fins de atividades colaborativas, considerando características dos estudantes, nível de conhecimento, entre outros dados (MORENO; OVALLE; VICARI, 2012).

c) Mineração de Relações (Regras de Associação)

Tem como objetivo descobrir possíveis relacionamentos entre variáveis escondidos em grandes conjuntos de dados (TAN; STEINBACH; KUMAR, 2009). A Mineração de Relações ainda abrange a Mineração de Correlações, tendo como exemplo a correlação linear positiva ou negativa entre variáveis, considerando a existência de uma correlação positiva quando, por exemplo, um aluno manifesta comportamentos que não são relacionados à atividade que o professor solicitou com a sua nota (BAKER; ISOTANI; CARVALHO, 2011).

Além das categorias supracitadas para Mineração de Relações, Baker (2010) cita mais duas: a Mineração de Sequências e a Mineração de Causas (análise de padrões de covariância, inferindo qual evento foi a causa do outro).

Dentre as categorias que compreendem a Mineração de Relações, um enfoque especial é dado para as Regras de Associação, detalhadas a seguir:

i. Regras de Associação

As Regras de Associação (RA) representam combinações de itens que ocorrem com uma determinada frequência em bases de dados e, diante disso, se tornam úteis para descobrir relacionamentos interessantes escondidos nessas bases de dados (AGRAWAL; IMIELINSKI; SWAMI, 1993; TAN; STEINBACH; KUMAR, 2009; GONÇALVES, 2005). As RA são usadas para descobrir padrões que descrevem características associadas entre os dados, e esses padrões descobertos são representados em forma de regras de implicação, obtendo padrões interessantes (TAN; STEINBACH; KUMAR, 2009).

Uma RA é tida como uma expressão de implicação no formato $X \rightarrow Y$, sendo que X e Y são conjuntos distintos de dados. Para a implicação $X \rightarrow Y$, X é o lado esquerdo da regra ou o antecedente e Y é o lado direito da regra, ou seja, o conseqüente, podendo ser interpretada como uma expressão lógica do tipo “Se X, então Y” (NOMELINI *et al.*, 2010) . De forma simplista, uma regra pode ser expressa por uma implicação do tipo $X \rightarrow Y$, onde X e Y representam um conjunto de itens.

Uma das formas de se obter as RA é através do algoritmo Apriori. O algoritmo Apriori, descrito por Tan, Steinbach e Kumar (2009), parte do princípio de que se um conjunto de itens é frequente, então todos os seus subconjuntos também deverão ser frequentes, ou seja, se $\{x, y, z\}$ é um conjunto de itens frequentes, qualquer transação que contenha $\{x, y, z\}$ deve conter seus subconjuntos $\{x, y\}$, $\{x, z\}$, $\{y, z\}$, $\{z\}$, $\{x\}$, $\{y\}$.

Como as RA são aplicadas para explorar um conjunto de dados com o intuito de encontrar regras que representem o padrão de comportamento, no presente caso, utilizou-se a mineração de Regras de Associação (RA) com a aplicação do algoritmo Apriori, para identificar o padrão de comportamento do aluno desanimado, apresentado no capítulo 5. O algoritmo Apriori foi elencado para ser utilizado na presente tese, partindo do pressuposto de que é capaz de gerar RA e, a partir dessas regras, apoiar a descoberta do desânimo.

As técnicas de associação usam o tipo de aprendizado de máquina não supervisionado. Na aprendizagem não supervisionada, utilizam-se instâncias sem a determinação do atributo classe ou atributo objetivo, usando análise exploratória dos dados e ficando a cargo do algoritmo utilizado a definição de quais atributos e valores serão usados para o desenvolvimento das RA (REATEGUI; CAZELLA, 2005).

ii. Qualidade e Medidas de Interesse de Regras de Associação

Geralmente, mede-se a qualidade ou força da Regra de Associação a partir de duas métricas²⁵, uma é o suporte e a outra é a confiança. O suporte de uma regra é usado para determinar a frequência na qual uma regra é aplicada em um determinado conjunto de dados, e a confiança para determinar a frequência na qual os itens Y aparecem nas transações que contenham X (SILVA; CRUZ; SILVA, 2014). Além disso, uma regra de baixo suporte pode ser eliminada por não apresentar interesse, sendo usado valor baixo para eliminar regras. Por outro lado, a confiança mede a confiabilidade da inferência de uma regra, quanto maior o seu

²⁵ Métricas são regras que regulam uma medida, no caso, regras que serviram para medir a qualidade de uma regra, facilitando a escolha de melhores regras.

valor, mais confiável é a regra. O Capítulo 5 delinea o uso e aplicação da confiança e suporte nas RA encontradas.

Para Agrawal, Imielisnki e Swami (1993), não se deve confundir o suporte com confiança, visto que a confiança é uma medida de força de uma regra e o suporte corresponde a uma significância estatística, e se o suporte não for suficientemente grande, significa que a regra não é digna de ser considerada ou não ser uma das preferidas, sendo usada mais tarde.

Além do suporte e confiança, foram usadas as medidas de interesse objetivas, ou seja, medidas que empregam índices estatísticos para avaliar a força de uma regra como o *lift* (avaliar dependência) e *conviction* (avalia implicação). Para Gonçalves (2005), ao considerar uma regra $A \rightarrow B$, a medida *lift* indica o quanto mais frequente se torna B quando A ocorre, e quanto maior o valor encontrado, melhor é a regra. Já o *conviction*, diferentemente do *lift*, avalia a implicação da regra, considerando o suporte do antecedente como do consequente. Quando o *lift* ($A \rightarrow B$) < 1 , A e B são negativamente dependentes, e se o valor da implicação da regra ($A \rightarrow B$) > 1 , então, a regra é considerada positivamente dependente, e caso o valor de *lift* ($A \rightarrow B$) = 1, então, A e B são independentes (GONÇALVES, 2005). Para o autor, os valores usuais de *conviction* a serem selecionados para as melhores regras devem estar entre 1,01 e 5, pois, os valores acima de 5 geralmente apresentam regras óbvias.

Ambas as métricas foram usadas, de acordo com a literatura, a fim de apoiar na seleção de melhores regras no contexto desta pesquisa, onde se selecionou as regras que possuíam maiores escores para as duas métricas. As Regras de Associação obtidas (capítulo 5) no processo de mineração foram validadas por especialistas²⁶ da área de MDE e CA, e as melhores regras foram apontadas para serem validadas e comporem o Modelo de Predição.

3.2.3 Trabalhos Associados

Algumas pesquisas que aplicam MDE são apresentadas nesta seção com intuito de mostrar as potencialidades da MD aplicada na educação, bem como apresentar um referencial teórico sobre importantes pesquisas desenvolvidas.

A MDE pode ser usada para projetar tecnologias/sistemas inteligentes e fornecer melhores informações aos alunos e professores (BAKER, 2014). Para o autor, aplicação da

²⁶ Dois especialistas nas áreas de mineração de dados e um especialista em computação afetiva apoiaram no entendimento de variáveis e interpretação das regras obtidas no capítulo 5.

MDE propicia, por exemplo, a descoberta de como os alunos aprendem, podendo apoiar na predição e entendimento de comportamento para a aprendizagem.

O trabalho desenvolvido por Iepsen (2013) utilizou a MDE para detectar padrões de comportamento associados com o estado afetivo de frustração no aluno. O autor utilizou a ferramenta de mineração *WizRule*²⁷ para a geração das Regras de Associação associadas ao estado afetivo de frustração do aluno ao realizar exercícios de Algoritmos. Foi possível evidenciar que algumas variáveis relacionadas à ocorrência repetida de erros no código-fonte e à demora na conclusão do exercício de algoritmo estavam associadas ao estado afetivo de frustração do aluno. Após a descoberta das regras, o autor propôs uma ferramenta de recomendação de um novo exercício com nível menor de complexidade em relação àquele que o aluno não conseguiu resolver.

O trabalho de Majadas, Santos e Boticario (2013), conforme também citado na seção 2.4, investigou como detectar afetividade do aluno a partir de várias fontes de dados. Uma delas é através da produção textual dos alunos em interação em um ambiente educacional. Os autores utilizaram a mineração de texto a fim de encontrar palavras com conotação afetiva, onde termos positivos e negativos foram identificados. A técnica de MDE aplicada foi a de redes *bayesianas*, porém, os autores apontam a necessidade de redefinir certos dados para obter melhores resultados nos novos experimentos.

A pesquisa desenvolvida por Inventado *et al.* (2013) propõe um modelo para ajudar o aluno a monitorar os estados afetivos (tédio, confusão, encantado, engajado e neutro) durante a realização de atividades. A proposta dos autores é de que o estudante seja informado sobre o seu estado afetivo quando estiver exposto a uma sobrecarga cognitiva, para que possa mudar o seu comportamento. Tal sistema ainda poderá sugerir atividades para melhorar o aprendizado, quando determinados estados afetivos são ou não são identificados. Os autores utilizaram a ferramenta de mineração *RapidMiner* para a aplicação da MD. Foram feitas comparações entre os resultados alcançados pelos algoritmos J48, *Naive Bayes*, *Jrip*, *SVM*, *Rep-Tree*, *Bayesian Network*. Melhor resultado foi apresentado pela *Naive Bayes*.

A pesquisa de Rigo *et al.* (2014) foi desenvolvida em uma aplicação a qual possibilita a descoberta de conhecimento em bases educacionais, através de MDE juntamente com *Learning Analytics* (LA), fornecendo aos gestores e professores informações sobre o processo de ensino-aprendizagem, cujo objetivo principal era analisar e identificar situações que levam à evasão do aluno. Foram utilizados dados das interações dos alunos no AVEA para a

²⁷ O *WizRule* é uma ferramenta de mineração desenvolvida pela *WizSoft* e *TextMinig*. Versão demo: www.wizsoft.com.

predição da evasão baseada na quantidade e datas de acessos às ferramentas do ambiente. Os autores aplicaram redes neurais como técnica de MD para predição. O algoritmo manipulado foi o *backpropagation*.

No trabalho de Gottardo, Noronha e Kaestner (2014), foi feito um estudo de aplicação da MDE para a inferência do desempenho do estudante em cursos a distância, sendo útil na medida em que os professores acompanham os estudantes e definem estratégias pedagógicas que minimizem reprovações. Nos experimentos, foram utilizados o método de classificação aplicando os algoritmos *RandomForest* e *MultilayerPerceptron* diretamente na ferramenta de mineração Weka. A proposta dos autores encontra-se em desenvolvimento.

A pesquisa de Ocumpaugh, Baker e Rodrigo (2012; 2015), apesar de ter sido apresentada na seção 2.2, merece destaque nesta seção com relação à MDE. Através do BROMP (*The Baker Rodrigo Ocumpaugh Monitoring Protocol*) foi desenvolvido um modelo padrão, em forma de protocolo, para que professores ou outro observador pudessem observar estados afetivos e comportamento observável do aluno em interação em um ambiente educacional, e anotassem via registro no protocolo BROMP. A principal finalidade do BROMP é identificar situações em que o aluno necessite de intervenção quando apresentar frustração, confusão, tédio, entre outras.

Atualmente, o BROMP está em sua segunda versão (BROMP 2.0) e é capaz de inferir aspectos relacionados à afetividade quando o aluno encontra-se em confusão, tédio, frustração e concentração e já foi integrado a várias aplicações educacionais (OCUMPAUGH; BAKER; RODRIGO, 2012, OCUMPAUGH *et al.* 2015).

Os dados do protocolo associados aos dados de *log* dos alunos, quando minerados, podem inferir questões relacionadas à afetividade. A partir das observações anotadas no protocolo, uma sincronização com os *logs* dos dados do AVEA ou outro tipo software educacional que esteja sendo usado pelos alunos é realizada. De posse dessas informações, a MDE é aplicada e é feita uma correlação dos padrões de interação com as observações de campo vindas do protocolo. A partir dessa junção de dados, os autores apresentam uma inferência sobre os estados afetivos detectados.

Para usar o protocolo, é preciso ter a certificação BROMP, em que é necessário receber o treinamento e alcançar a avaliação da medida estatística *Kappa* para o valor de 0,6. Para facilitar a observação e seguir o protocolo, foi desenvolvida uma ferramenta em Android chamada de HART (*The Human Affect Recording Tool*), apresentada na pesquisa de Ocumpaugh *et al.* (2015), que facilita a coleta e integração dos dados do protocolo com os de *log* do ambiente usado pelo aluno.

A pesquisa de Shute *et al.* (2015) usa MDE para modelar o conhecimento do aluno incluindo seus estados afetivos e a sua influência na aprendizagem em um jogo educacional. O jogo utilizado é o *Physics Playground* (Shute e Ventura, 2013). Os autores (Shute *et al.*, 2015) buscaram estabelecer a forma como os estados afetivos (frustração, confusão e concentração) influenciaram a aprendizagem no jogo e analisar como esses fatores interagem para influenciar uns aos outros e promover a aprendizagem. A observação dos estados afetivos foi feita utilizando-se o protocolo BROMP 2.0.

Após analisar as pesquisas supracitadas, percebe-se que há forte engajamento da comunidade acadêmica no desenvolvimento de pesquisas que aplicam MDE. A pesquisa desenvolvida nesta tese teve como foco o estado de ânimo desanimado e utilizou métodos e técnicas da MDE para apoio na obtenção do Modelo de Predição do aluno desanimado.

O Quadro 2 faz uma apresentação resumida dos trabalhos supracitados, juntamente com o desenvolvido nesta pesquisa.

Quadro 2: Exemplos de trabalhos desenvolvidos utilizando MDE

AUTOR	APLICAÇÃO	TÉCNICAS UTILIZADAS MDE	FERRAMENTA DESENVOLVIDA	FERRAMENTA UTILIZADA
Damando (2016)	Reconhecimento do estado de ânimo desanimado do aluno em interação no AVEA Moodle	Classificação (J48 e <i>Cross-Validation</i>). Regras de Associação (Apriori)	Modelo de Predição do Aluno Desanimado juntamente com a Visualização do Modelo em forma Gráfica	Weka/SPSS
Rigo <i>et al.</i> (2014)	Evasão	Redes Neurais com o algoritmo <i>backpropagation</i>	Ferramenta que apresenta os alunos com predição da evasão.	Não cita
Gottardo, Noronha e Kaestner (2014)	Estimativa de desempenho de estudantes	Classificação através dos algoritmos <i>RandomForest</i> e <i>MultilayerPerceptron</i> .	Não foi desenvolvida ferramenta	Weka
Ocuppaugh, Baker e Rodrigo (2012; 2015)	Protocolo para observar de estados afetivos (frustração, confusão, tédio) e comportamento observável do aluno em interação em um ambiente educacional.	Não consta	BROMP- <i>The Baker Rodrigo Ocuppaugh Monitoring Protocol</i> Ferramenta Android HART (<i>The Human Affect Recording Tool</i>)	Ferramenta própria.
Shute <i>et al.</i> (2015)	Identificação dos estados afetivos: frustração, confusão e concentração para o aluno em interação no jogo <i>Physics Playground</i> .	Não consta	Utiliza o BROMP 2.0	Não cita

AUTOR	APLICAÇÃO	TÉCNICAS UTILIZADAS MDE	FERRAMENTA DESENVOLVIDA	FERRAMENTA UTILIZADA
Iepsen (2013)	Estado afetivo frustrado	Regras de associação - algoritmo Apriori	Recomendação de exercícios de menor complexidade para o aluno frustrado.	<i>WizRule</i>
Majadas, Santos e Boticario (2013)	Reconhecer emoções em produções textuais	Mineração de Texto - Árvore de decisão, Redes Bayesianas	Não tem	Não cita
Inventado et al. (2013)	Reconhecer estados afetivos (tédio, confusão, encantado, engajado e neutro)	Variáveis comportamentais - Redes Bayesianas, J48, <i>JRip</i> , <i>Rep-Tree</i>	Proposta de modelo	<i>RapidMiner</i>

Fonte: Elaborado pela autora

Ao analisar as pesquisas supracitadas, algumas delas englobaram reconhecimento de questões relacionadas com os afetos como as pesquisas de Iepsen (2013), Ocumpaugh, Baker e Rodrigo (2012; 2015) e Majadas, Santos e Boticario (2013). Algumas incluem a frustração, que é considerado um termo afetivo que faz parte do estado de ânimo desanimado. Dessa forma, o estado de ânimo desanimado, o qual é abordado na nesta tese, engloba muitos termos que estão presentes isoladamente em algumas pesquisas, tornando-o mais completo quando se refere aos termos afetivos evidenciados.

Já com relação às técnicas de MDE, estas são específicas a cada pesquisa e vão ao encontro do objetivo proposto em cada uma, porém, observa-se que tarefas preditivas foram desenvolvidas na maioria das pesquisas.

Na maior parte das pesquisas, modelos de aluno são propostos e muitos dos dados são oriundos dos AVEA ou softwares educativos os quais ocorrem interação dos alunos.

No que diz respeito à presente pesquisa, foi desenvolvido o Modelo de Predição do aluno desanimado, juntamente com a integração da ferramenta de mineração Weka no AVEA Moodle, e foi implementada uma ferramenta de visualização do Modelo de Predição, via *dashboard*, demonstrando um diferencial com relação aos demais.

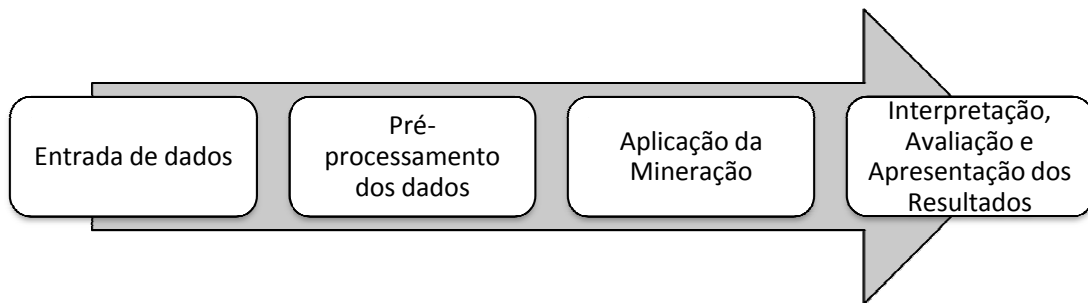
3.2.4 Processo de Mineração de Dados Educacionais na Pesquisa

Ao associar áreas e técnicas ao processo da mineração, é possível ter uma visão macro da abordagem da MDE na presente pesquisa. Inicialmente, foi feita uma análise estatística

para o questionário aplicado (Apêndice B) e para os *logs* dos experimentos A, B, e C (detalhados na seção 4.3). Essa análise incluiu a geração de frequências estatísticas, a geração de histogramas da distribuição estatística, a tabulação cruzada, a fim de verificar associações entre as questões e dados de *log* e obter melhor entendimento das variáveis levantadas. Após essas definições, o processo de MDE pode ser executado.

O procedimento adotado na presente pesquisa para a MD é apresentado através da Figura 7 e, logo após, são apresentados os detalhes previstos para cada etapa do processo de mineração.

Figura 7: Etapas para MDE



Fonte: Elaborado pela autora

Detalhamento das etapas do processo de mineração:

- a) Entrada de dados: a busca dos dados foi feita através da base de dados do AVEA Moodle.
- b) Pré-processamento dos dados: o AVEA Moodle é composto por várias tabelas relacionadas em seu banco de dados, cujos dados necessários para realizar a mineração. Os dados foram “limpados” e “transformados” no formato apropriado para ser minerado, ou seja, precisaram ser transformados para o formato a ser usado no minerador Weka.
- c) Aplicação da mineração de dados: aplicou-se algoritmos de mineração de dados a fim de descobrir o conteúdo de conhecimento de interesse. Utilizou-se o minerador Weka (mais detalhes na seção 3.3.1) para realização da Classificação e Regras de Associação, através do algoritmo J48 (método *Holdout* e *Cross-Validation*) e algoritmo Apriori.
- d) Interpretar, avaliar e apresentar os resultados (pós-processamento): após a mineração, o resultado alcançado foi apresentado aos especialistas de MD e CA,

conforme mencionado anteriormente, a fim de verificar o resultado da mineração e validar regras obtidas.

3.3 FERRAMENTAS PARA MINERAÇÃO DE DADOS EDUCACIONAIS

Esta seção se dedica à apresentação de algumas ferramentas usadas para MD e aplicadas em MDE com enfoque no Weka, pois foi a ferramenta selecionada para realizar a mineração na presente tese. Além do Weka, são apresentadas as ferramentas *RapidMiner*, *KNIME* e *SPSS*.

3.3.1 Weka

Uma das ferramentas de mineração de dados bastante utilizada é o Weka (*Waikato Environment for Knowledge Analysis*) que possui um conjunto de algoritmos de aprendizado de máquina e ferramentas de processamento de dados para prover a mineração de dados. Possui uma interface que permite ao usuário poder comparar diferentes métodos e identificar qual é o mais apropriado para o problema no qual precisa resolver (WITTEN; FRANK, 2005).

O Weka foi desenvolvido pela Universidade de Waikato, na Nova Zelândia (NZ), particularmente, pelo *Machine Learning Group*, um grupo que busca aplicar métodos de aprendizado de máquina a fim de fazer com que computadores possam analisar automaticamente um grande volume de dados, e decidir quais informações são mais relevantes, apoiando em previsões e precisão a tomada de decisões. O objetivo do grupo é desenvolver técnicas de aprendizado de máquina, soluções práticas e importantes para a NZ, além de contribuir com o quadro teórico da área (Weka, 2014).

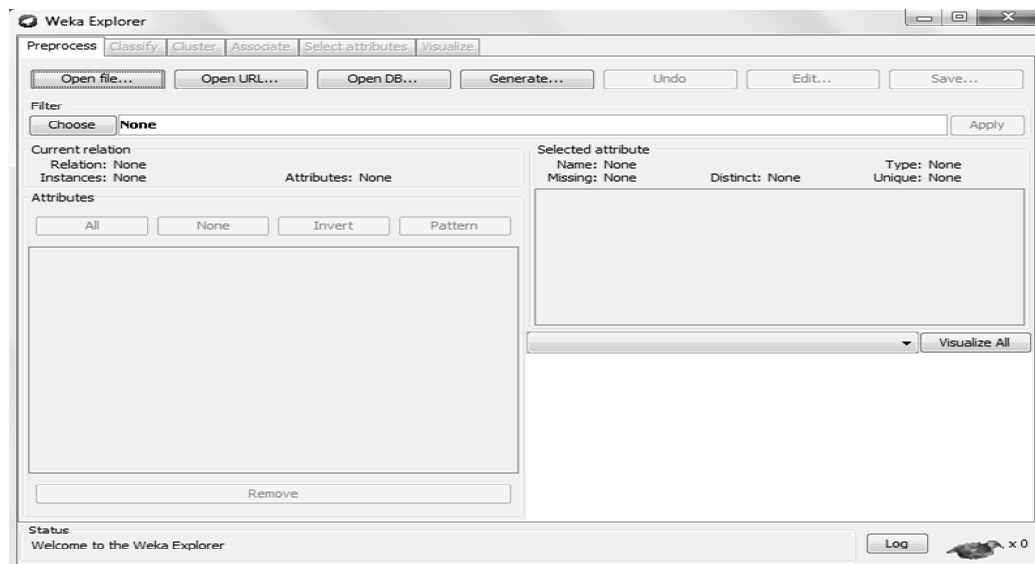
A tela inicial do Weka é exibida na Figura 8.

Figura 8: Tela inicial do Weka



Com relação aos detalhes técnicos, o Weka é desenvolvido em Java, multiplataforma, e é distribuído pela licença GNU (*General Public Licence*), podendo ser obtido por *download* diretamente do site da Universidade de Waikato²⁸. O Weka possui muitos algoritmos de aprendizado que podem ser facilmente aplicados a um *dataset*,²⁹ além de ferramentas para pré-processamento, regressão, classificação, agrupamento, regras de associação, seleção de atributos e visualização. A Figura 9 exibe a tela principal do Weka.

Figura 9: Apresentação do Weka



A ferramenta trabalha com o tipo de arquivo no formato ARFF (*Attribute Relation File Format*), que pode ser lido de um arquivo ou gerado por uma consulta ao banco de dados.

²⁸ Para download do Weka: <http://www.cs.waikato.ac.nz/ml/Weka/downloading.html>.

²⁹ *Dataset* são conjuntos de dados que podem estar armazenados em banco de dados.

Arquivos ARFF possuem duas seções. A primeira diz respeito ao *Header*, ou seja, descreve informações no cabeçalho, que inclui o nome da relação e a lista de atributos com seus respectivos dados. A segunda seção diz respeito aos dados propriamente ditos (Weka, 2014). Arquivos do tipo ARFF aceitam os seguintes tipos de dados: nominal (lista de possíveis valores) entre colchetes e separados por vírgula, numérico (real ou inteiro), string (muito usado para mineração de texto) e data (entre aspas o *timestamp* “yyyy-MM-dd HH:mm:ss”). Geralmente, no cabeçalho, há detalhes sobre o arquivo (em comentários), seguido por @ e nome da relação e @ e os atributos a serem utilizados, neste caso, com o respectivo tipo de dado. Para a segunda parte dos dados, o seu início se dá com @data, seguido por todas as instâncias ou registros e seus atributos. Cabe ressaltar que os atributos devem ser separados por vírgulas e cada linha deve corresponder a um registro, além do uso de % para comentários.

Ao importar um arquivo ARFF para o Weka, torna-se possível escolher uma técnica e o algoritmo desejado para realizar a mineração. Um exemplo de arquivo ARFF pode ser analisado como demonstrado na Figura 10.

Figura 10: Exemplo de arquivo ARFF

```

%Primeira parte - Cabeçalho
% Exemplo de um arquivo ARFF
%
@relation Teste
type numeric
@attribute Resultado {aprovado, reprovado, recuperação}
@attribute Media numeric
@attribute InteracaoForum {sim, não}
@attribute Chat {sim, não}

% Segunda parte – Dados
@data

% instancias – 5 para exemplo

1111, aprovado, 9, sim, sim
2222, recuperação, 6, sim, não
3333, recuperação, 7, não, sim
4444, aprovado, 8, sim, sim
5555, reprovado, 6, não, não

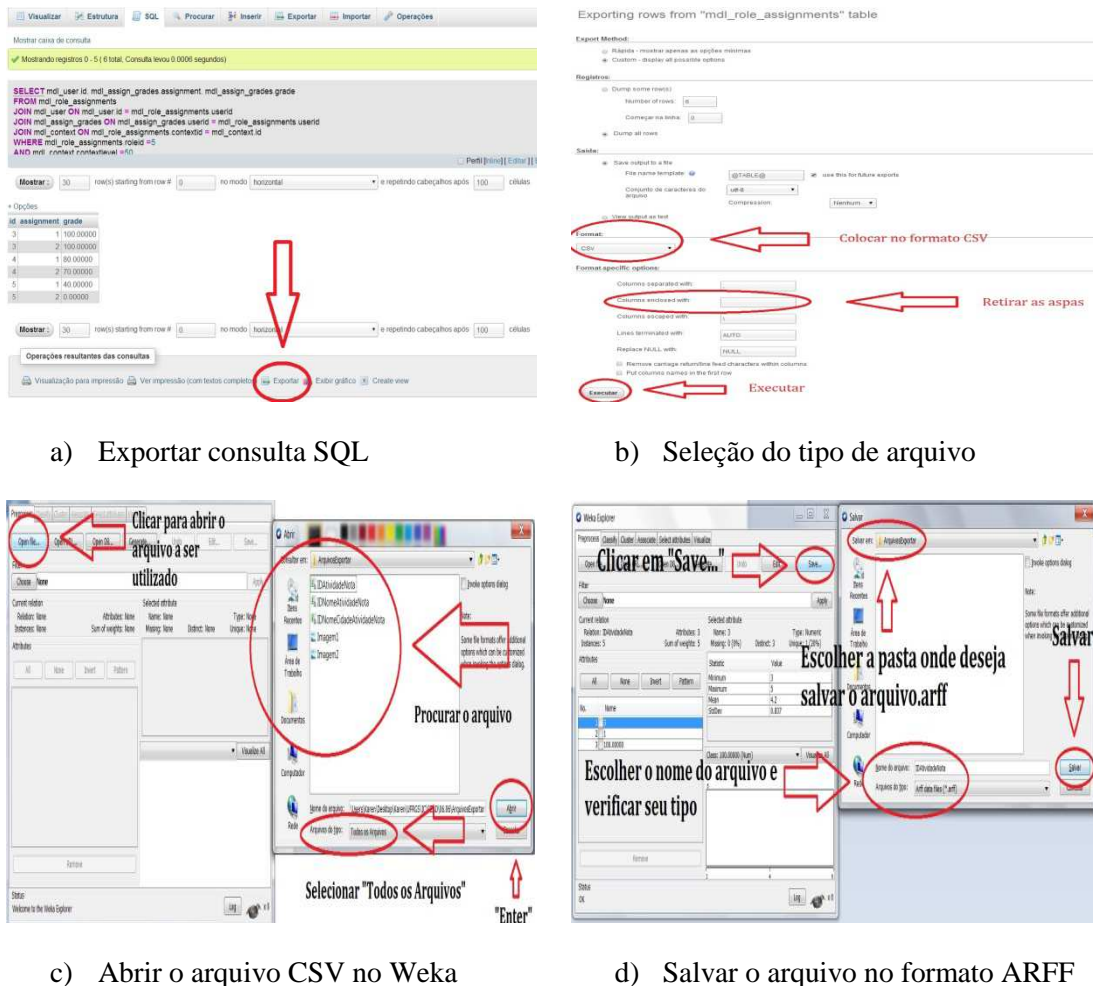
```

Fonte: Elaborado pela autora

Se os dados estão no Moodle, é possível realizar a exportação dos dados necessários para um formato mais apropriado para gerar o ARFF. Para isso, geralmente, é preciso realizar uma consulta SQL (*Structured Query Language*) no Banco de Dados (*MySQL*) e exportar os dados no formato CSV, retira-se as aspas e, em seguida, basta selecionar o botão de executar. Dentro do Weka, é possível abrir o arquivo do formato CSV exportado pelo Moodle e posteriormente salvá-lo no formato ARFF. Desta forma, é possível transformar os dados do Moodle para serem usados diretamente no Weka.

A Figura 11 exibe um passo-a-passo para esta execução.

Figura 11: Passos para exportar e usar um arquivo do Moodle no Weka



Fonte: Elaborado pela autora

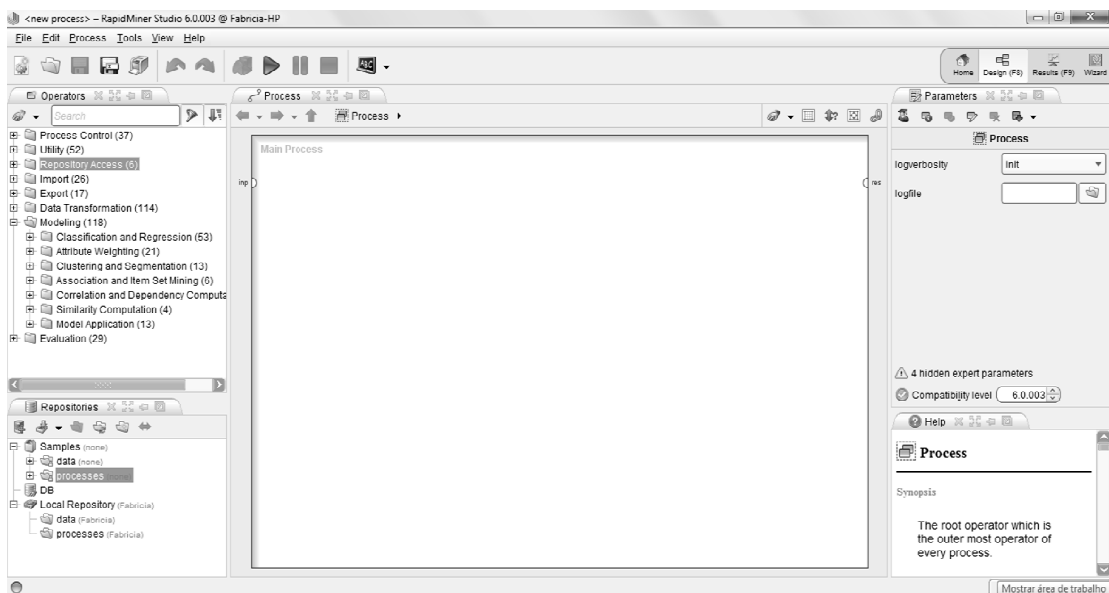
O processo de mineração realizado a partir dos dados do arquivo ARFF está descrito no capítulo 5.

Além da mineração, uma integração do Weka ao AVEA Moodle foi desenvolvida e implementada com o propósito de possibilitar a mineração dentro do próprio ambiente. Dessa forma, permite ao professor utilizar a mineração com os dados integrados ao ambiente e possibilita obter os resultados mais rapidamente do que se eles não estiverem integrados. Essa integração oferece ao professor, além do apoio computacional para a realização da mineração, a facilidade de uso para realizar efetivamente a mineração. Essa integração é apresentada no Capítulo 6.

3.3.2 RapidMiner

O *RapidMiner*³⁰ é um minerador de código aberto com capacidade de realizar análise preditiva. Esta ferramenta, desenvolvida em Java e na sua versão 7, permite aplicação de várias técnicas e algoritmos para a mineração dos dados e mineração de texto. A Figura 12 apresenta a tela do *RapidMiner*.

Figura 12: Tela do *RapidMiner*



Algumas pesquisas na área educacional foram desenvolvidas utilizando o *RapidMiner*, por exemplo, o trabalho de Wilges *et al.* (2010) que aplicou árvore de decisão para determinar

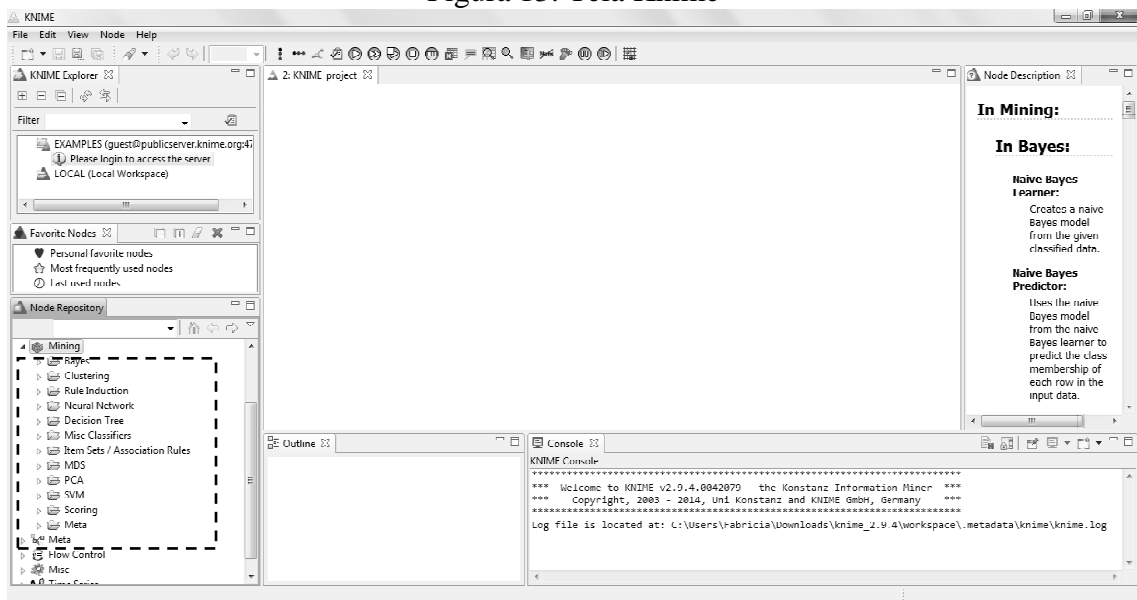
³⁰ Disponível para download em: <https://my.rapidminer.com/nexus/account/index.html#downloads>

o perfil de aprendizagem dos estudantes. A pesquisa de Leonardi, Catapan e Bastos (2013) aplicou o *RapidMiner* para realizar a análise de perfil de estudantes de um curso *Lato Sensu* na modalidade EAD, em que aplicou o algoritmo de classificação ID3 para possibilitar uma visualização da formação e experiência em EAD dos candidatos selecionados para o curso, a fim de propor uma abordagem pedagógica do curso congruente com o perfil selecionado.

3.3.3 KNIME

Outra opção para a MD é propiciada pelo *Knime Platform Analytics*³¹, que é uma plataforma utilizada para mineração de dados, possuindo código fonte aberto, licença GLP (*General Public License*) e facilmente extensível e direcionada ao desenvolvimento de API (Interface de Programação de Aplicativos). O Knime é uma plataforma de análise de dados que possibilita realizar a mineração em um grande volume de dados a fim de analisar tendências e prever resultados (KNIME, 2014). A Figura 13 apresenta a tela do Knime, com destaque para a mineração.

Figura 13: Tela Knime



Encarado como uma poderosa ferramenta para análise e visualização de dados, fornece um ambiente para intuitivo e fácil de usar. Além da possibilidade da mineração de dados por

³¹ Disponível para download em: <http://www.knime.org/knime>

meio de arquivos, o KNIME pode chamar e trabalhar com *WebService* para manipular e minerar os dados (SILIPO; MAZANETZ, 2012).

O trabalho desenvolvido por Wiswedel, Kotter e Silipo (2013) utilizou o KNIME para realizar a mineração de textos dos dados coletados a partir de um fórum na *Web*, além de monitorar suas estatísticas. Dessa forma, foi possível observar os temas discutidos nos fóruns, apresentados pela mineração de texto.

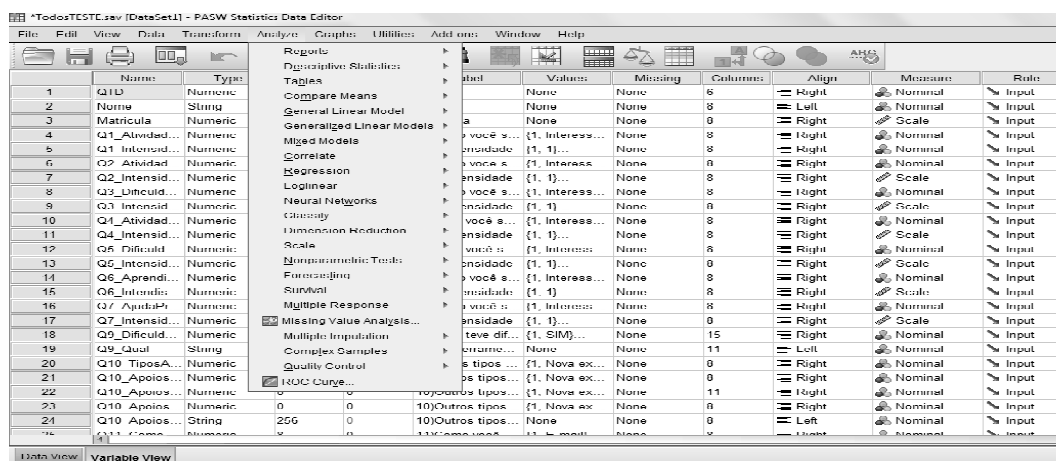
3.3.4 SPSS

O SPSS³², cujo acrônimo é *Statistical Package for the Social Sciences*, é um *software* utilizado para análises estatísticas que inclui a possibilidade de realizar mineração de dados e mineração de texto.

Para utilizá-lo, é preciso ter um arquivo preparado para ser importado pelo SPSS, podendo ter o formato de dados, por exemplo, do Excel (*xls*), Access³³, entre outros. É necessário declarar cada variável utilizada, informando o nome da variável, tipo de dados (número, *string*), tamanho da variável, entre outras informações.

A Figura 14 apresenta a tela inicial do SPSS.

Figura 14: Tela do SPSS



³² Download do SPSS em: <http://www-01.ibm.com/software/analytics/spss/products/statistics/downloads.html>.

³³ Access é um sistema gerenciador de banco de dados desenvolvido pela Microsoft.

Após ter todos os dados, é possível iniciar o seu uso através dos vários menus e submenus, realizando análise estatística, geração de gráficos, regressão linear, transformação de variáveis, geração de quartis³⁴ e demais opções.

Um manual do SPSS foi desenvolvido pela UFRGS e é disponibilizado através do *link*: <http://www.ufrgs.br/gpat/wp-content/uploads/2012/05/Manual-de-Spss-pt.pdf>.

Ressalta-se que a presente pesquisa contou, inicialmente, com a assessoria do Núcleo de Assessoria Estatística (NAE)³⁵. Foram agendadas duas reuniões para apoio no uso do SPSS. A versão utilizada do SPSS foi a disponibilizada pela UFRGS aos alunos – *PAWS Statistics 18* – através do *download* diretamente no site da IES.

O SPSS foi aplicado para fazer a tabulação dos dados dos questionários (apêndice B), também para tabular os dados dos *logs* do Moodle nos devidos experimentos e serviu de apoio para a geração de histogramas de frequências, geração de quartis usados nas minerações e apoio estatístico detalhado nos capítulos 4 e 5, apoiando nas tarefas de pré-processamento e transformação dos dados.

3.4 CONSIDERAÇÕES SOBRE O CAPÍTULO

O capítulo apresentou os principais conceitos, técnicas, métricas e ferramentas utilizadas na mineração de dados educacionais e apontou as principais técnicas computacionais abordadas na presente pesquisa, como a Classificação e Regras de Associação, juntamente com os algoritmos J48 e Apriori.

Além dos principais conceitos, foi possível identificar, através das áreas da mineração de dados apresentadas, que a pesquisa se enquadra à primeira área, modelagem do aluno. Além disso, foi possível apresentar e compreender a aplicação da MDE em várias pesquisas apresentadas, onde muitas delas buscam identificar aspectos relacionados com a afetividade, mostrando o potencial da MDE para a área educacional.

Buscou, através das ferramentas disponíveis para MDE, identificar e detalhar as principais ferramentas e apontar a ferramenta escolhida, optando pelo uso do Weka por ser uma ferramenta utilizada em muitas pesquisas na área de MDE, devido à sua facilidade de uso

³⁴ Para Espinheira *et al.* (2015), separatrizes são medidas usadas para separar dados e o quartil é usado para dividir dados de uma série em quatro partes iguais, dividindo a população em grupos.

³⁵ NAE é um órgão do departamento de estatística da UFRGS com a finalidade de ofertar assessoria estatística em pesquisas acadêmicas dentro da Universidade. Disponível em: <http://www.mat.ufrgs.br/~nae/onaie.htm>.

e, também, pela possibilidade de a exportação dos dados do Moodle para Weka se dar de forma amigável. Além disso, o Weka possui código livre, o que facilitou quando foi necessário integrá-lo ao Moodle.

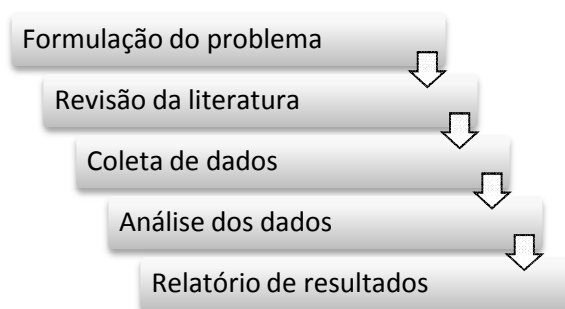
Também foi empregado o SPSS como apoio à análise estatística e auxílio na utilização das variáveis na mineração, principalmente as que tiveram que ser transformadas, como o caso da geração de quartis. Todos os valores de quartis estão discriminados no capítulo 5.

4 PROCEDIMENTO METODOLÓGICO

Este capítulo destina-se à apresentação do procedimento metodológico adotado nesta tese. Foram utilizados para a fundamentação da metodologia de pesquisa os autores Gil (2002), Yin (2010), Flick (2013) e Silva e Menezes (2005).

Para Sampieri, Collado e Lucio (2013, p. 30), uma pesquisa abrange um “conjunto de processos sistemáticos, críticos e empíricos aplicados ao estudo de um fenômeno” os quais são apresentados, através da Figura 15, como processos fundamentais do processo de pesquisa.

Figura 15: Processos fundamentais do processo geral de pesquisa



Fonte: Adaptado de Sampieri, Collado e Lucio (2013, p. 30).

A primeira fase, de formulação do problema, foi desenvolvida no capítulo 1 e abordou a exploração, descrição e entendimento do problema. Na fase de revisão de literatura, foram apresentadas bases teóricas e conceituais para a pesquisa no que tange aos aspectos relacionados à afetividade e sobre a fundamentação da MDE, os quais foram descritos nos capítulos 2 e 3, respectivamente.

A coleta de dados, apresentada no presente capítulo, descreve o experimento piloto e os experimentos utilizados na coleta de dados. A análise de dados diz respeito à descrição sobre os resultados e a análise propriamente dita dos dados coletados, os quais incluíram os procedimentos de MD e a validação da mineração.

4.1 CLASSIFICAÇÃO DA PESQUISA

A *pesquisa* proposta é de *natureza aplicada* ao processo de ensino e aprendizagem, fundamentada no conceitos de Silva e Menezes (2005), enquanto afirmam que a pesquisa aplicada objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas. A classificação da pesquisa também é fundamentada nas concepções de Gil (2002), ao descrever que toda pesquisa possui *classificação* baseada de acordo com seus *objetivos* gerais. Dessa forma, a pesquisa é classificada como *explicativa*, a qual se preocupa em identificar os fatores que determinam ou contribuem para a ocorrência de fenômenos. Para Gil (2002), as pesquisas explicativas podem ser classificadas como experimentais ou *ex-post-facto*.

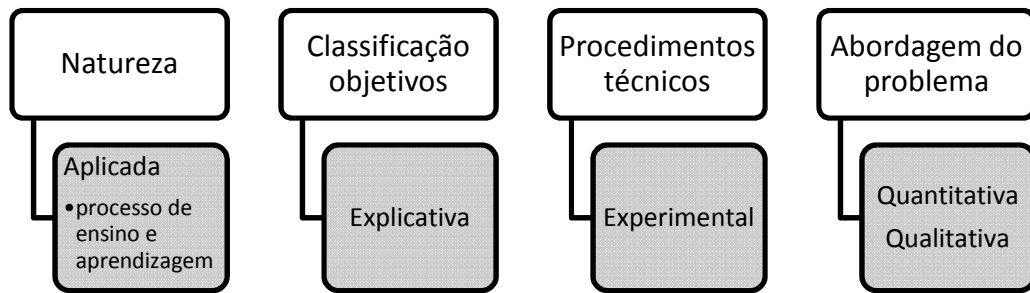
Sendo assim, com relação aos *procedimentos técnicos*, optou-se em desenvolver pesquisa experimental sob a ótica de Gil (2002), ao expor que pesquisas experimentais determinam um objeto de estudo, selecionam-se variáveis que podem influenciá-lo além da definição das formas de controle e observação dos efeitos das variáveis sob o objeto de estudo. O autor aponta a necessidade de realizar testes com o instrumento de coleta, a fim de medir o que é proposto na pesquisa, os quais foram realizados e estão presentes no capítulo 6, onde se aborda a validação.

Nesse sentido, observa-se que o objeto de estudo foi definido através da questão de pesquisa, e as variáveis que podem influenciá-lo foram identificadas pelas variáveis observáveis do aluno em processo de desânimo e são descritas no capítulo 5. As variáveis foram descobertas e validadas/testadas por meio da mineração de dados, apresentadas nos capítulos 5 e 6.

A abordagem do problema se deu de forma quanti-quali. Qualitativa por ter realizado uma interpretação dos dados empíricos, os quais foram possíveis de interpretação a partir da MDE realizada. Também se adotou uma abordagem quantitativa, que se deu ao utilizar procedimentos estatísticos da análise dos dados dos *logs* dos alunos, sua tabulação e aplicação no SPSS.

Diante do exposto, a metodologia guiou-se através dos dados procedimentos apresentados pela Figura 16.

Figura 16: Procedimento Metodológico



Fonte: Elaborado pela autora

4.2 ETAPAS DA ELABORAÇÃO DO PROJETO E DEFINIÇÃO DA AMOSTRA

Para a elaboração da pesquisa, foi realizada inicialmente uma pesquisa bibliográfica para apropriação dos conceitos fundamentais, dos temas centrais relacionados à afetividade bem como suas relações com a aprendizagem em AVEA. Além dos termos relacionados à afetividade, foi necessário buscar aporte na análise estatística e mineração de dados educacionais, a fim de dar sustentação à técnica computacional utilizada. Através da presente apropriação, foi possível escolher termos e técnicas para o direcionamento da pesquisa, os quais incluem:

- Estudo aprofundado sobre estados de ânimo e, mais especificamente, no estado de ânimo “desanimado”;
- Definição do uso da MDE e técnicas estatísticas para predizer os alunos propensos ao desânimo;
- Desenvolvimento do estudo piloto, onde foi possível apontar o mapeamento de variáveis utilizadas, conforme apresentado na Figura 17.

Portanto, para a elaboração desta pesquisa, foi necessário realizar a *coleta de dados* a partir de um estudo piloto e da realização de três experimentos (A, B e C).

Para a coleta de dados, foi utilizado o instrumento questionário, análise de documentos do AVEA Moodle (*logs*, dados de tabelas do banco de dados), tabulação e preparação de dados, mineração dos dados levantados, obtenção de regras de associação através da MDE.

Após essa etapa, foi realizada a análise dos dados, geração do modelo de predição do desânimo e apresentação do modelo em forma gráfica via *dashboards*.

A etapa de avaliação e análise dos dados diz respeito à análise e interpretação dos dados. De acordo com a abordagem do problema, foi realizada a análise qualitativa e quantitativa dos dados, via questionário e análise dos dados e interação dos alunos no AVEA Moodle. As etapas realizadas para o desenvolvimento da pesquisa podem ser analisadas resumidamente através da Figura 17.

Figura 17: Visão geral das etapas de elaboração do projeto



Fonte: Elaborado pela autora

A população alvo da pesquisa é específica e pertencente à unidade de estudo ECO 3010 – Análise de Demonstrações Contábeis, disciplina ofertada na modalidade semipresencial, a qual iniciou um estudo piloto na turma de 2013/1, para validação de questionário e efetivou-se nos semestres de 2014/1 e 2014/2 através dos experimentos A, B e C.

Conforme mencionado na seção 1.6, o estudo realizado em 2013/1 permitiu analisar as interações dos alunos nos fóruns de grupos e verificar a influência do estado de ânimo nas atividades de grupo à distância, através das interações e uso do Moodle. Foi desenvolvida uma coleta de dados através do instrumento questionário, composto por questões mistas, que está disponível pelo link: <<https://docs.google.com/forms/d/16yDHzdwoHLmR1YQBy8b5ogjVLfBR7P5tp99Iai5816I/viewform>>.

O questionário foi dividido em duas partes, sendo que a primeira parte continha enfoque em questões relacionadas ao grupo de trabalho, o qual pertencia um aluno, e a segunda parte está relacionada com a disposição afetiva do aluno frente a sua aprendizagem no AVEA Moodle, além de dificuldade de uso do ambiente.

Foi aplicada a REA (detalhada no capítulo 2) e a mineração de textos ao fórum de grupos, utilizando a ferramenta SOBEK³⁶ (Klemann, Reategui, Lorenzatti, 2009) para analisar o conteúdo das postagens dos alunos e de seus grupos. Através das análises, foi possível identificar que o estado de ânimo influenciava na cooperação do aluno no grupo, na medida em que apresentava maior satisfação e interesse pelas atividades desenvolvidas, impulsionando-o a enfrentar os desafios de aprendizagem.

Este estudo serviu de apoio para identificar melhorias implementadas no instrumento de coleta de dados, chegando ao questionário que se encontra no Apêndice B, e iniciar a pesquisa efetivamente com os experimentos A, B e C.

Após esse estudo preliminar, dois projetos foram desenvolvidos e submetidos à Secretaria de Educação a Distância (SEAD), na UFRGS, através dos Editais 19 (2014) e 21 (2015), os quais foram contemplados, sendo que o último foi finalizado em fevereiro de 2016. No primeiro, foi implementado o questionário do Apêndice B, onde REA 2.0 foi implementada computacionalmente, em linguagem de programação PHP (*Hipertext Preprocessor*) e banco de dados MySQL. O objetivo do segundo edital foi de apoiar na

³⁶ SOBEK é um minerador de textos capaz de identificar conceitos relevantes em um texto. Disponível para download em: <http://sobek.ufrgs.br/download-pt.html>.

tabulação dos dados, análise estatística e a mineração. Além disso, também foi possível propor um Modelo de Predição, o desenvolvimento da integração do Weka ao Moodle e desenvolvimento de uma ferramenta de visualização por meio de *dashboard* (capítulo 6).

O período de execução do primeiro projeto foi de abril de 2014 a fevereiro de 2015 e serviu de apoio aos experimentos para a disciplina ECO 3010. O segundo projeto iniciou-se em abril de 2015 e encerrou-se em fevereiro de 2016. Ambos foram contemplados com uma bolsa de graduação.

A seção 4.3 apresenta uma visão geral dos experimentos realizados, o detalhamento encontra-se nos capítulos 5 e 6, respectivamente.

4.3 DESCRIÇÃO DOS EXPERIMENTOS

As subseções seguintes descrevem o processo realizado para a aplicação dos experimentos A, B e C.

4.3.1 Procedimento para o experimento A

Nesse experimento, realizado em 2014/1, a turma ECO03010 – Análise das Demonstrações Contábeis possuía 58 alunos inscritos, porém, a amostra do questionário contou com 41 alunos participantes. O questionário foi aplicado durante uma aula presencial no dia 24/05/2014, turno da manhã. No início, se teve o cuidado de explicar os objetivos da pesquisa, bem como foi apresentado aos alunos o Termo de Consentimento (Apêndice A) e o Questionário a ser respondido (Apêndice B). Sendo assim, participaram somente os alunos que consentiram em fornecer os seus dados de interação do AVEA Moodle bem como os dados das respostas ao questionário.

Posteriormente à aplicação do questionário, foi realizada a tabulação dos dados via Excel e SPSS, cuja intenção, nesse primeiro momento, era de tentar verificar indicadores observáveis dos alunos propensos ao desânimo, identificando as variáveis prováveis de serem exploradas no AVEA.

De forma resumida, foram realizados os seguintes passos:

- a) Aplicação do questionário para análise da disposição afetiva e validação das variáveis indicadoras do desânimo;
- b) Tabulação e tratamento de dados no excel , mapeamento dos resultados tabulados através do SPSS, utilizando análise estatística, tabulações cruzadas, correlações, geração de quartis, dentre outros.

A finalidade foi buscar, através do montante de variáveis constantes nos questionários respondidos, quais seriam as variáveis prováveis a serem utilizadas na mineração. A seção 4.3.1.1 apresenta os resultados alcançados com a aplicação do experimento A.

4.3.1.1. Análise descritiva dos dados do experimento A

Os primeiros dados tabulados do experimento A se referem às questões que utilizaram a REA 2.0, cujas perguntas encontram-se no Apêndice B, onde o enunciado das sete (07) questões se iniciava por “*Como você se sente ...*”. Nessas questões, o aluno faz um autorrelato selecionando um par de emoções e a sua intensidade, a qual pode variar em cinco níveis.

O questionário contou com 15 perguntas, sendo que o primeiro grupo de questões (7 questões), diz respeito à disposição afetiva do aluno frente a situações de aprendizagem. As demais questões elencaram informações sobre conteúdo, ferramentas, formas de ajuda, entre outras, sendo utilizados outros modelos de questões.

Nas sete primeiras questões, foi aplicada a REA 2.0. Como todos os resultados estão mapeados em valores numéricos e, a fim de facilitar a visualização e entendimento da REA 2.0, usou-se os valores numéricos ao lado dos pares de termos, conforme apresentado no Quadro 3.

Quadro 3: Termos afetivos usados na REA 2.0

SATISFEITO		ANIMADO		DESANIMADO		INSATISFEITO	
1	Interesse	6	Ternura	11	Piedade	16	Desapontamento
	Envolvimento		Sentir amor		Compaixão		Pesar
2	Divertimento / Riso	7	Deslumbrado	12	Tristeza	17	Inveja
			Surpresa		Desamparo		Ciúme

SATISFEITO		ANIMADO		DESANIMADO		INSATISFEITO	
3	Orgulhoso Exaltação	8	Despreocupação Alívio	13	Preocupação Medo	18	Repulsão Antipatia
4	Felicidade Alegria	9	Perplexo Espanto	14	Constrangimento Vergonha	19	Desprezo Ofensa
5	Prazer/ Satisfação	10	Anseio Nostalgia	15	Culpa Arrependimento	20	Irritação Raiva
21 – Nenhuma		22 – Outra		23 – Não respondeu			

Fonte: Elaborado pela autora

➤ Medidas estatísticas (média, mediana, desvio padrão)

Ao aplicar medidas estatísticas para as questões 1 a 7, o total de respostas válidas foi em torno de 39, pois alguns participantes não responderam uma ou outra questão. Ao analisar os resultados obtidos, através da Tabela 1, a média³⁷ encontrada para as questões Q1, Q2, Q3 e Q5 sugere que as respostas representaram o grupo de termos utilizado para o desânimo (termos de 11 a 15). Da mesma forma, a mediana³⁸ sugere o mesmo resultado ao analisar as questões Q1, Q2, Q3, Q4 e Q5.

Tabela 1: Medidas estatísticas das questões REA 2.0

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Válidos	39	39	40	38	39	38	40
Média	15,08	15,18	14,50	9,89	12,90	5,24	6,88
Mediana	15,00	15,00	14,50	13,00	13,00	3,50	5,00
Moda	16	14 ^a	20	1	13	1	1
D. Padrão	2,679	2,270	5,204	7,942	5,251	5,185	6,653
Variância	7,178	5,151	27,077	63,070	27,568	26,888	44,266
Mínimo	7	12	1	1	1	1	1
Máximo	21	23	20	22	23	16	22

^a A menor moda é mostrada por existir mais de uma.

Fonte: Elaborado pela autora

➤ Frequências estatísticas

³⁷ A média é entendida, neste contexto, como o valor considerado o ponto de equilíbrio para as respostas.

³⁸ A mediana pode ser considerada uma média dos valores médios.

As frequências estatísticas foram levantadas para cada questão, e são apresentadas na Tabela 2. Ao analisar a questão Q1, cerca de 44% das respostas se referem a algum dos termos representativos do desânimo. A mesma situação acontece, respectivamente, para as questões Q2 com 68%, Q3 com 37%, Q4 com 24%, Q5 com 44%, Q6 com 10%, Q7 com 24% dos termos escolhidos representarem o desânimo. Dessa forma, verifica-se que as variáveis relacionadas às questões Q1, Q2, Q3 e Q5, possuem um potencial para a representação do desânimo.

Tabela 2: Frequência estatística

Termos	Q1 F	Q2 F	Q3 F	Q4 F	Q5 F	Q6 F	Q7 F
Interesse / Envolvimento	0	0	2	11	2	12	18
Divertimento / Riso	0	0	0	3	1	6	1
Orgulhoso / Exaltação	0	0	0	0	0	1	0
Felicidade / Alegria	0	0	1	1	0	5	0
Prazer/ Satisfação	0	0	0	1	0	3	2
Ternura / Sentir amor	0	0	0	0	0	0	0
Deslumbrado / Surpreso	1	0	2	1	4	2	3
Despreocupação / Aliviado	1	0	0	1	2	1	3
Perplexo /Espanto	0	0	1	0	0	0	0
Anseio / Nostalgia	0	0	0	0	0	0	0
Piedade / Compaixão	0	0	0	0	0	0	0
Tristeza / Desamparo	1	1	4	0	4	1	1
Preocupação / Medo	5	5	7	7	12	3	1
Constrangimento / Vergonha	5	11	3	3	2	0	8
Culpa / Arrependimento	7	11	1	0	0	0	0
Desapontamento / Pesar	15,0	7	6	1	5	4	0
Inveja / Ciúme	0	0	0	0	0	0	0
Repulsão / Antipatia	0	0	0	0	0	0	0
Desprezo / Ofensa	0	0	1	0	0	0	0
Irritação / Raiva	3	2	12	5	6	0	1
Nenhuma	1	1	0	2	0	0	1
Outra	0	0	0	2	0	0	1
Não Respondeu	2	2	1	3	2	3	1

Fonte: Elaborado pela autora

Também foi feita uma análise com relação às questões Q1 e Q2, sobre o envio de atividades individuais e em grupo, onde a primeira questão: “*como o aluno se sente quando envia atividades com atraso*” e segunda questão: “*como o aluno se sente quando envia atividades em grupo com atraso*”. O resultado mostrou que, os termos com maior frequência estavam presentes justamente no quadrante do desânimo, quando questionado sobre qual tipo de emoção o aluno evidencia na ocorrência das atividades individuais ou em grupo atrasadas. Para a primeira questão, 18 alunos manifestaram sentir alguma das emoções relacionadas ao

estado de ânimo desanimado e, para a segunda questão, a frequência foi de 28 alunos. Sendo assim, há indícios que sugerem que essas questões possam apoiar a evidência do desânimo, baseado nas frequências de estado de ânimo das respostas obtidas.

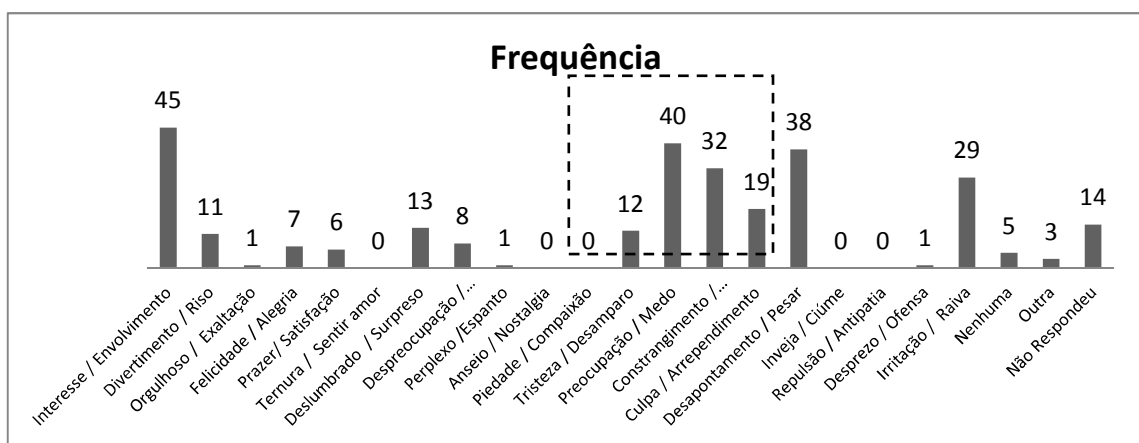
➤ **Histograma da distribuição das frequências**

Através da distribuição de frequência, foi possível organizar os dados em categorias de termos afetivos. Como são 7 questões, 41 participantes e 20 termos afetivos, obteve-se 287 respostas. A distribuição de frequência resultante pode ser observada através da Figura 19.

Ao analisar a Q6 – *como o aluno se sente com relação ao seu aprendizado*, com 12 ocorrências, apresenta-se como um indicador satisfatório para o aprendizado.

Os escores de frequência encontram-se ilustrados através do histograma apresentado na Figura 18.

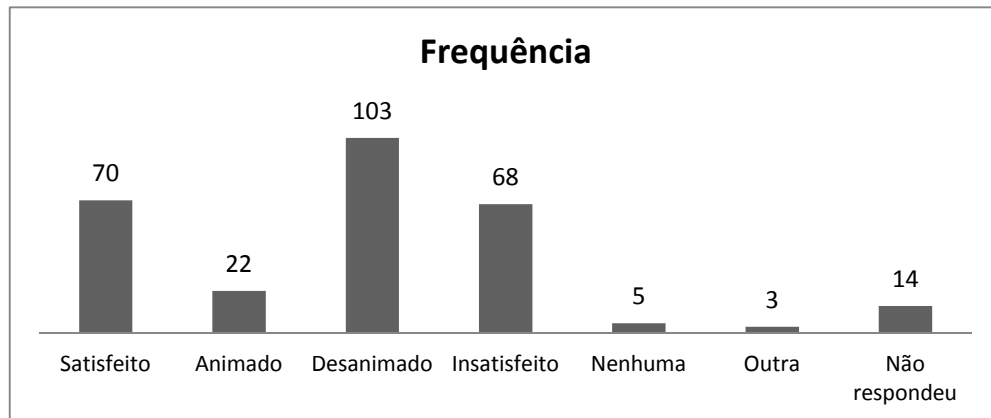
Figura 18: Histograma da distribuição de frequência



Fonte: Elaborado pela autora

Através do histograma pôde-se observar que os termos relacionados ao desânimo tiveram maiores frequências. Fato que é confirmado pelo histograma apresentado na Figura 19, a qual apresenta os escores da distribuição de frequência de acordo com os estados de ânimo (animado, desanimado, satisfeito e insatisfeito). Nesse histograma, os termos representativos das famílias afetivas foram agrupados de acordo com os termos satisfeito, animado, insatisfeito, além do agrupamento de nenhuma, outra e não respondeu. Esse agrupamento possibilita observar claramente a frequência obtida para o estado de ânimo desanimado.

Figura 19: Histograma dos estados de ânimo



Fonte: Elaborado pela autora

➤ **Tabulação cruzada**

Uma tabulação cruzada foi realizada para apoiar a identificação de relações entre variáveis. A Tabela 3 apresenta a tabulação cruzada entre a questão da REA 2.0 – *como você se sente quando envia atividades individuais em atraso* – e a questão que usa a escala *Likert* – *com relação às atividades individuais, se o aluno as envia com atraso*. É possível representar frequências estatísticas distribuídas de acordo com a escala usada e, nesse sentido, os alunos que entregam atividades em atraso informam termos afetivos que representam o desânimo, apresentando correlações entre as variáveis.

Tabela 3: Tabela cruzada para atividades individuais

1) Como você se sente quando envia atividades individuais em atraso?	13 a - Atividades individuais - Você as envia com atraso?				Total
	Algumas Vezes	Nunca	Quase Sempre	Raramente	
Deslumbrado/Surpresa	0	0	0	1	1
Despreocupação/ Aliviado	0	0	0	1	1
Tristeza / Desamparo	0	0	0	1	1
Preocupação / Medo	0	2	0	3	5
Constrangimento/Vergonha	1	2	0	2	5
Culpa / Arrependimento	3	1	1	2	7
Despontamento / Pesar	8	3	0	4	15
Irritação / Raiva	0	0	0	3	3
Nenhuma	0	1	0	0	1
Não respondeu	1	1	0	0	2
Total	13	10	1	17	41

Fonte: Elaborado pela autora

Observa-se, pela análise da Tabela 3, que a maioria dos alunos que entregam atividades em atraso também sentiu dificuldade em realizá-las e, em ambos os casos, apresentaram emoções relacionadas ao desânimo. Sugere-se que essas variáveis comportamentais sejam utilizadas no mapeamento das observáveis comportamentais indicadoras do desânimo.

Da mesma forma, foi questionado ao aluno se ele *sentia dificuldade em realizar as atividades individuais* (escala Likert) e realizou-se a tabulação cruzada desta questão com a questão: *como o aluno se sentia ao enviar atividades em atraso* (REA 2.0/GEW 2.0). O resultado pode ser analisado através da Tabela 4 .

Tabela 4: Dificuldade em realizar atividades individuais

Como você se sente quando envia atividades individuais em atraso?	13b - Atividades individuais - Você sente dificuldade em realizá-las?				Total
	Algumas Vezes	Nunca	Quase Sempre	Raramente	
Deslumbrado/Surpresa	1	0	0	0	1
Despreocupação/ Aliviado	1	0	0	0	1
Tristeza / Desamparo	0	0	0	1	1
Preocupação / Medo	4	0	0	1	5
Constrangimento/Vergonha	5	0	0	0	5
Culpa / Arrependimento	3	1	1	2	7
Despontamento / Pesar	8	0	2	5	15
Irritação / Raiva	2	0	0	1	3
Nenhuma	0	0	1	0	1
Não respondeu	0	0	1	1	2
Total	24	1	5	11	41

Fonte: Elaborado pela autora

Através da análise dos resultados, sugere-se que essas variáveis possam ser utilizadas para o mapeamento das observáveis indicadoras do desânimo, haja vista a frequência e o inter-relacionamento das mesmas.

Também foram realizados questionamentos sobre como o aluno se sentia ao enviar atividades em grupo em atraso, frequência de envios atrasados e se sentia dificuldades em realizar tarefas em grupo. Resultados da tabulação foram similares aos apresentados nas Tabelas 3 e 4 utilizadas para casos individuais e que também foram evidenciadas para o caso de atividades em grupo, o que motiva o uso dessas variáveis para o mapeamento das observáveis como indicadoras ao desânimo.

Os demais dados analisados se referem às questões de 9 a 15. As questões analisam uso e preferências de ferramentas no contexto da aprendizagem, perguntas relacionadas à solicitação de ajuda ao professor e apoios que podem ser ofertados, e por fim, analisa atividades individuais e em grupo e acompanhamento de conteúdos.

Portanto, sugere-se que as questões 13a, 13b, 14a, 14b, 15a e 15b possam ser usadas para a mineração e verificação de possibilidade de apoio na evidência do desânimo.

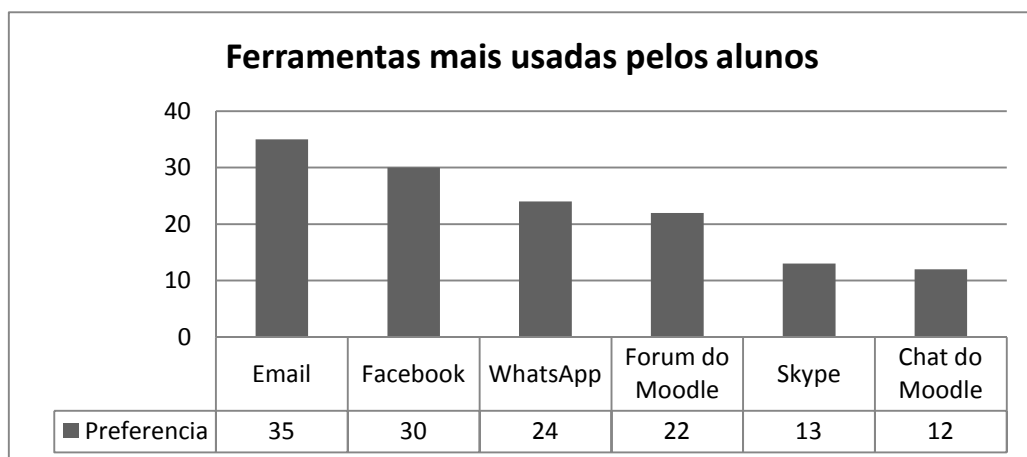
➤ Ferramentas

Com relação à dificuldade de usar alguma ferramenta do Moodle, aproximadamente 80% relataram não ter dificuldades com ferramentas, alguns relatos mostraram dificuldade com a ferramenta bate-papo, bem como no uso de questionários.

Outro questionamento diz respeito a quais ferramentas os alunos preferem usar para trocar informações sobre trabalhos da disciplina, tendo o *email*, *facebook*³⁹ e *whatsapp*⁴⁰ como os de maior preferência, sugerindo que os alunos estão usando ferramentas vinculadas aos celulares, devido à facilidade de comunicação a qualquer momento.

A Figura 20 apresenta as ferramentas ordenadas por preferência de uso por parte do aluno.

Figura 20: Preferência de ferramentas



Fonte: Elaborado pela autora

³⁹ Facebook – Rede social.

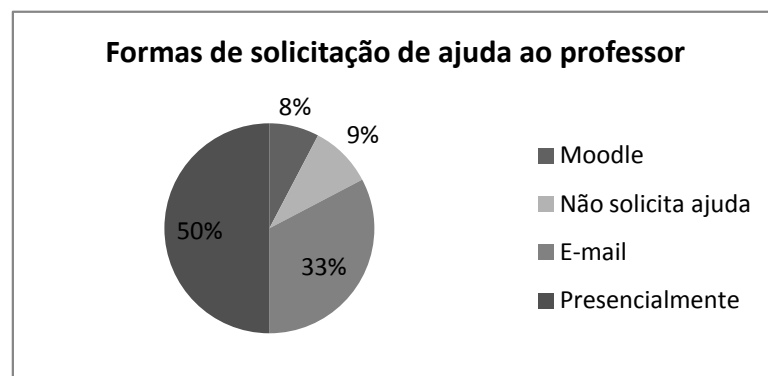
⁴⁰ WhatsApp – Ferramenta utilizada para comunicação assíncrona através de aparelhos celulares.

Mesmo o AVEA Moodle possuindo ferramentas de comunicação, a preferência do uso de ferramentas se deu fora do ambiente.

➤ Professor

No que diz respeito às formas de solicitação de ajuda ao professor, o que se destaca é que, como a disciplina é realizada de forma semipresencial, porém com alguns encontros presenciais e utiliza o AVEA Moodle, somente 8% dos alunos solicitaram ajuda ao professor por este meio e, no caso de usar alguma ferramenta, eles recorrem aos e-mails. Esses resultados podem ser observados pela Figura 21, sugerindo que muitas das solicitações são deixadas para os momentos presenciais.

Figura 21: Solicitação de ajuda ao professor



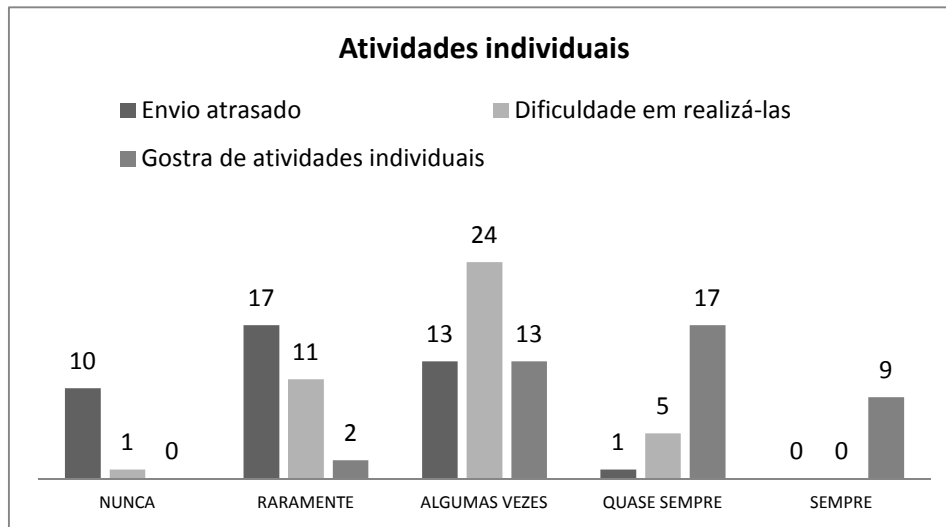
Fonte: Elaborado pela autora

➤ Atividades individuais e em grupo

As questões 13 e 14 abordaram, de forma geral, a forma com que as atividades (individuais ou em grupo) são desenvolvidas, questionando sobre: atraso no envio, dificuldade em realizá-las e se o aluno gosta de realizá-las.

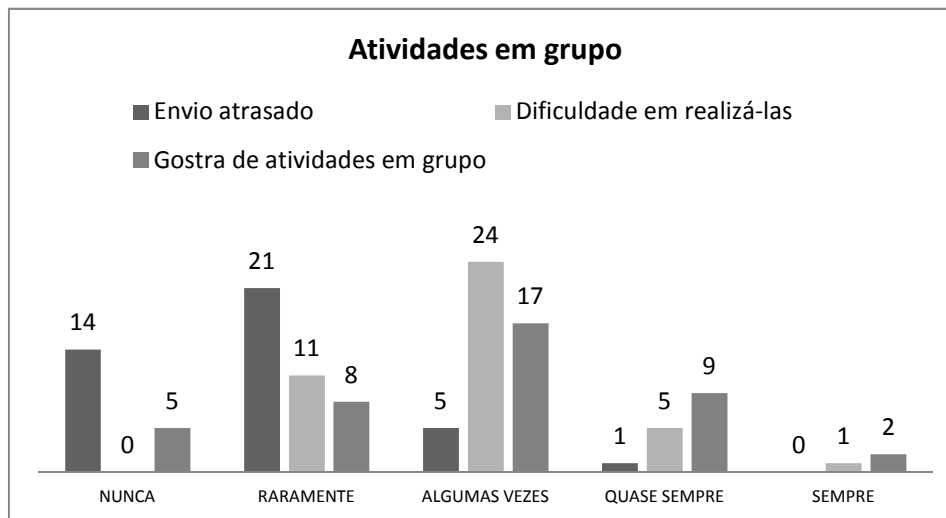
Ambas foram desenvolvidas utilizando a escala *Likert* (nunca, raramente, algumas vezes, quase sempre e sempre) e são observadas pelas Figuras 22 e 23, respectivamente.

Figura 22: Com relação às atividades individuais



Fonte: Elaborado pela autora

Figura 23: Com relação às atividades em grupo



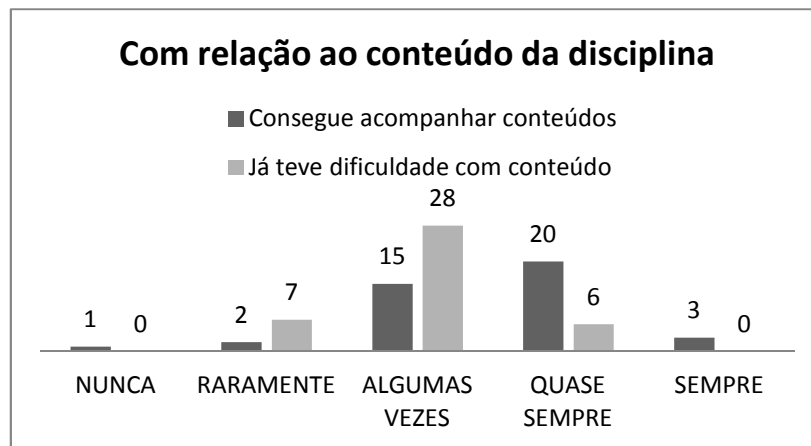
Fonte: Elaborado pela autora

As respostas relatadas sugerem que os alunos sentem algum tipo de dificuldade de realizar as atividades, tanto individualmente quanto em grupo, porém, nem sempre as envia com atraso. Isso pode ser justificado pelo fato de a maioria das atividades possuir data para ser entregue, apesar de a professora da disciplina negociar esta possibilidade.

➤ **Conteúdo**

Foi feita uma análise através de questões relativas ao conteúdo, as quais abordavam se o aluno conseguia acompanhar o conteúdo e se já teve dificuldade com o mesmo.

Figura 24: Conteúdo da disciplina



Fonte: Elaborado pela autora

As frequências encontradas podem ser observadas através da Figura 24. Verificou-se que os alunos apresentaram, algumas vezes, dificuldades com o conteúdo.

➤ **Considerações sobre o estudo piloto**

Ressalta-se que, nesse estudo, as variáveis foram levantadas a partir da manifestação da emoção do comportamento desanimado, oriundos somente da análise e tabulação do questionário aplicado.

A partir da manifestação e tabulação dos dados, foi possível verificar as seguintes variáveis como possíveis indicadoras do desânimo:

- Quantidade de vezes em que enviou atividades individuais em atraso;
- Quantos dias de atraso por atividade individual;
- Quantidade de vezes em que enviou atividades em grupo com atraso;
- Quantos dias de atraso por atividade em grupo;
- Quantidade de tentativas para enviar material;

- Quantidade e tipos de solicitações de ajuda ao professor;
- Qual ferramenta mais solicita ajuda do professor;
- Autorrelato: Dificuldade de utilização de ferramentas do AVEA Moodle;
- Autorrelato: Dificuldade na realização das atividades individuais e em grupo;
- Autorrelato: Dificuldade de acompanhar conteúdos.

As variáveis acima foram sugeridas, num primeiro momento, para a utilização na mineração do experimento B, descrito no capítulo 5. Porém, as variáveis efetivamente elencadas para a mineração são apresentadas em cada técnica e método utilizado na MDE como pode ser observado no capítulo supracitado.

4.3.2 Procedimentos para a mineração de dados (experimento B)

O experimento B realizou a aplicação do questionário, teve como foco a aplicação da mineração de dados, análise dos resultados minerados e escolha dos melhores resultados alcançados.

Para conduzir o experimento, foram realizadas as seguintes etapas:

- Aplicação do questionário para inferência da disposição afetiva (*online*);
- Análise do *log* do Moodle;
- Seleção e preparação dos dados de *log* e do questionário;
- Seleção dos algoritmos usados pelo MDE Weka;
- Seleção de variáveis comportamentais para a mineração;
- Análise dos resultados minerados;
- Seleção dos melhores resultados da mineração;
- Apresentação das regras/padrões de comportamento que levam ao desânimo.

Excluindo as duas atividades iniciais, as demais atividades foram destinadas no processo de MDE.

Todo o detalhe do experimento B encontra-se no capítulo 5.

4.3.3 Procedimento para a validação da mineração e apresentação do modelo (experimento C)

O procedimento adotado no experimento C foi destinado à validação das regras obtidas das minerações dos componentes-tarefa, fórum e arquivo em um novo conjunto de dados. O objetivo foi verificar quais regras poderiam ser generalizadas ou, pelo menos, regras que fossem identificadas a partir de escores mínimos adotados.

A validação foi realizada para um grupo de alunos que já havia respondido ao questionário, mas que não havia tido os seus dados minerados. De posse das regras encontradas, as mesmas foram testadas nesse novo conjunto de dados, realizando-se buscas dentro no banco de dados e armazenando seus resultados.

Também foi feita uma nova mineração, a partir das variáveis comportamentais de cada componente, para essa turma. Os dados foram comparados e as melhores regras sugeridas foram validadas pelas duas estratégias, a da mineração e a da pesquisa no BD.

A partir das regras validadas, foi proposto um modelo de predição do aluno desanimado. Foi feita uma integração do Weka ao Moodle para as técnicas e algoritmos utilizados no experimento.

Além disso, foi desenvolvido um modelo de apresentação das regras obtidas de forma mais visual, obtendo uma ferramenta de visualização do Modelo de Predição em forma de *dashboards*.

Sugere-se que a partir dos *dashboards* seja possível o professor ter uma visão geral das regras e, assim, poder melhor acompanhar alunos propensos ao desânimo.

Todo o processo desenvolvido nesse experimento é relatado no capítulo 6.

5 APLICAÇÃO E ANÁLISE DA MINERAÇÃO DE DADOS

Este capítulo se dedica à descrição das técnicas e métodos utilizados para a realização da MDE para o experimento B, bem como a análise dos resultados minerados.

O experimento B ocorreu para o componente curricular/disciplina ECO03010 – Análise das Demonstrações Contábeis, no semestre 2014/2. Essa disciplina foi composta por duas turmas (turma “a” e turma “b”) e é ofertada pela mesma docente do experimento A.

Inicialmente, os participantes responderam ao questionário da pesquisa, apresentado no apêndice B, porém, realizado de forma on-line e dividido em duas etapas. A primeira etapa apresenta a REA 2.0 implementada⁴¹, onde era possível ao aluno responder a sete (07) questões referentes à sua disposição afetiva. Somente os alunos matriculados em 2014/2 na disciplina ECO3010 é que poderiam responder ao questionário.

Após a validação do usuário, feita através de *login* e senha, uma tela de consentimento era apresentada e, caso o aluno consentisse em participar, ele era redirecionado para as questões a serem respondidas utilizando a REA 2.0. Caso o aluno não aceitasse participar, uma mensagem de agradecimento era apresentada e encerrava a sua participação. Sendo assim, a amostra para o experimento B foi composta por 44 alunos que consentiram em participar e permitiram acesso aos seus dados de *log* do Moodle.

Ao aceitarem participar, os alunos eram redirecionados para a tela a qual apresentava explicações sobre o questionário e realizava elucidações sobre como deveriam proceder para respondê-lo. Após esse passo, as questões eram apresentadas uma a uma. A Figura 25 apresenta a primeira pergunta “*O que você sente quando envia atividades individuais com atraso?*”. Com isso, o aluno devia selecionar uma das opções apresentadas. Ressalta-se que a pergunta realizada envolve “*o que você sente...*” para todas as questões relacionadas à REA 2.0.

Quando o *mouse* é posicionado sobre qualquer uma das famílias afetivas, vários termos afetivos associados são disponíveis e apresentados ao aluno. Esse procedimento visa facilitar o entendimento do aluno com relação ao termo exposto. Assim que uma das opções fosse selecionada, o aluno era redirecionado para a próxima pergunta, sendo esse processo repetido até a sétima pergunta, quando finalizava esse primeiro questionário.

⁴¹ A REA 2.0 foi implementada computacionalmente utilizando linguagem de programação PHP e banco de dados MySQL. A REA 2.0 encontra-se disponível em: <http://www.computacaoafetiva.com/>.

Figura 25: Exemplo de tela com pergunta



Fonte: Elaborado pela autora

A segunda etapa do questionário, complementar à primeira, contou com as demais questões presentes no apêndice B e foi desenvolvida usando o *GoogleDocs*. Ambos endereços eletrônicos foram disponibilizados para os alunos dentro do Moodle.

Os dados dos questionários foram tabulados de forma similar ao que foi realizado no experimento A. O próximo passo foi ter acesso aos *logs* das interações dos alunos no Moodle. Os *logs* dos alunos foram gerados pela ferramenta de administração do próprio Moodle (disponível para professor) e exportados para a planilha eletrônica Excel, e manipulados no SPSS para apoio estatístico e apoio na geração de histogramas de frequências, geração de quartis usados nas minerações.

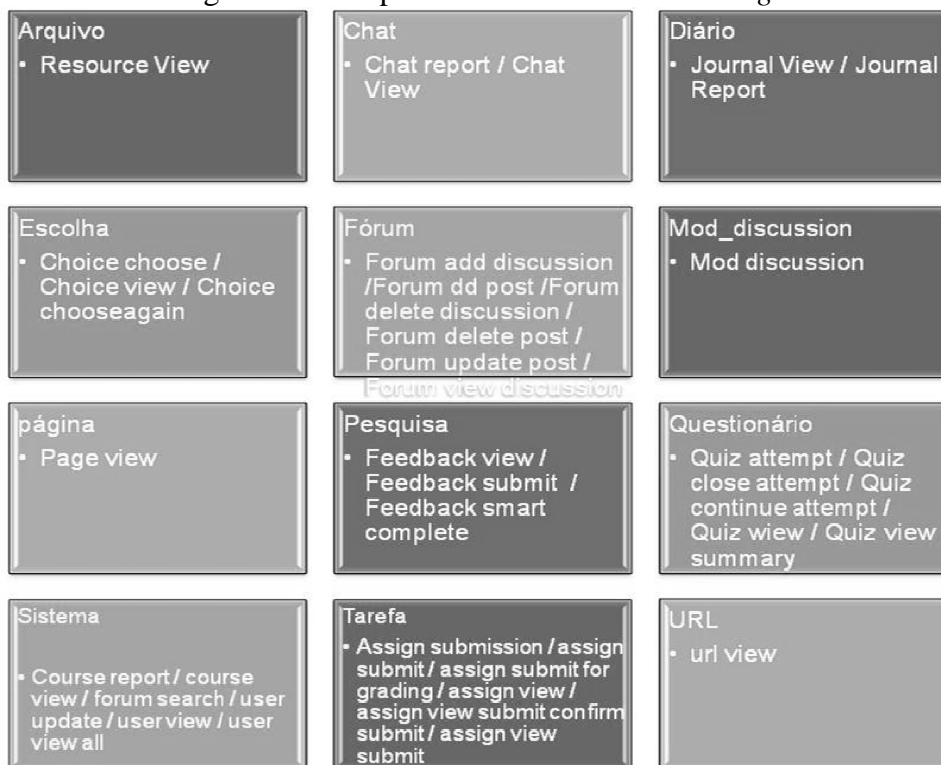
O *log* do Moodle é composto pelos atributos: *id_usuario*, usuário afetado, contexto do evento, componente, nome do evento e descrição. Como foram manipulados *logs* de duas turmas para o experimento B, foi acrescido o atributo “turma”. O quadro 4 apresenta o modelo resultante do *log* exportado pelo Moodle.

Quadro 4: Modelo de *log* resultante do Moodle

ID-USER	USUÁRIO AFETADO	CONTEXTO DO EVENTO	COMPONENTE	NOME DO EVENTO	DESCRIÇÃO	TURMA
137538	User1	Fórum: Fórum dos Grupos Trabalho Setorial	Fórum	forum_view discussion	forum view discussion 107002	a
137538	User1	Curso: ECO03010 - Análise De Demonstrações Contábeis - Turma A (2014/2)	Sistema	course_view	course view 28136	a
137538	User1	Tarefa: Atividade de Revisão Avaliação P1 (Até 03/10)	Tarefa	assign_view	assign view Ver própria página de status de envio.	a
137538	User1	Arquivo: Aula – Alavancagens	Arquivo	resource_view	resource view 404214	a

Fonte: *Log* de dados Moodle

O *log* é composto por 12 componentes (arquivo, chat, diário, escolha, fórum, mod_discussion, página, pesquisa, questionário, sistema, tarefa, URL) além de possuir mais de 30 eventos possíveis de ocorrência, como exibido na Figura 26. Para cada evento, há aproximadamente 93 tipos de contextos possíveis, o que inclui, por exemplo, todas as atividades propostas, todos os arquivos de conteúdos disponibilizados, todos os questionários elaborados e *chats* desenvolvidos para estas turmas.

Figura 26: Componentes e seus eventos no *log*.

Fonte: Elaborado pela autora

O início do processo de descoberta de conhecimento foi aplicado ao presente trabalho, onde a *seleção dos dados* se deu a partir do AVEA Moodle, a partir dos dados selecionados dos *logs* dos alunos, além das questões dos questionários. Os *logs* continham o nome do usuário, contexto do evento, componente, nome do evento e descrição, totalizando 26.613 registros existentes para as duas turmas.

Porém, para minerar esses dados, foi preciso organizá-los. Foi necessário passar pela fase da MDE de *pré-processamento*, possibilitando a organização dos dados para posterior mineração. Dessa forma, foi criado um id (identificador) do usuário e excluído o seu nome, o contexto do evento que era uma *string* muito grande e de vários tipos. O componente passou a ser um atributo que continha apenas as opções: Arquivo, *Chat*, Diário, Escolha, Fórum, *mod_discussion*, Página, Pesquisa, Questionário, Sistema, Tarefa, URL. O nome do evento manteve em *string*, porém, sem espaços entre palavras e a turma definida como “a” ou “b”. O Quadro 4 apresenta um exemplo de *log* resultante.

Foi necessário excluir os *logs* da professora da disciplina e das respectivas tutoras existentes, além dos *logs* dos alunos que não responderam aos questionários ou que não aceitaram o termo de consentimento de acesso aos seus dados.

Ao término, os *logs* das turmas “a” e “b” somaram 26.353 registros considerados válidos. A frequência dos componentes de *logs* pode ser observada através da Figura 27.

Figura 27: Frequência dos componentes de *log*

Selected attribute		
Name: componente		Type: Nominal
Missing: 0 (0%)		Unique: 1 (0%)
Distinct: 12		
No.	Label	Count
1	Arquivo	3904
2	Chat	81
3	Diario	72
4	Escolha	595
5	Forum	2815
6	mod_discussion	1
7	Pagina	748
8	Pesquisa	62
9	Questionario	1084
10	Sistema	10737
11	Tarefa	5928
12	URL	326

Class: turma (Nom) Visualize All

Fonte: Elaborado pela autora

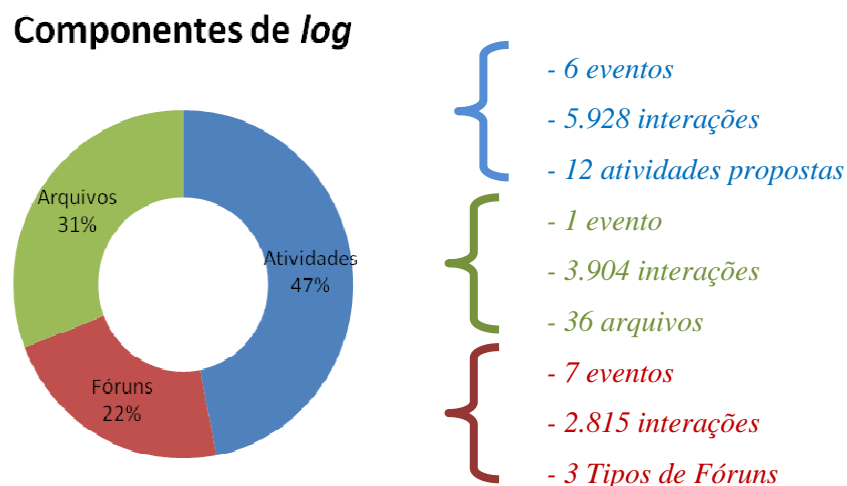
Os componentes de *log* de maior participação são evidenciados na Figura 28, destacando-se: os arquivos (incluem todos os documentos que o professor posta no ambiente,

bem como as visualizações dos alunos ao arquivo) com total de 3904 participações; o fórum (incluem todos os fóruns criados, bem como as postagens, visualizações, etc.) com um total de 2815 participações; sistema (contabiliza os acessos ao sistema) com 10737 acessos e, tarefa (incluem as visualizações, postagens de atividades, envio de atividades, confirmação de envio de tarefas) com um total de 5928 participações.

Diante dos componentes possíveis a serem explorados e minerados, optou-se pelos elementos atividades, fóruns e arquivos, pois representam os componentes de maior participação e interação dos alunos na disciplina, aliados a algumas variáveis sugeridas pelo experimento A para a evidência do desânimo, como as variáveis relacionadas ao envio de atividades individuais e em grupo. Ao analisar o *log* do moodle, estas variáveis estão presentes, por exemplo, no componente atividade, o qual identifica se uma atividade foi enviada ou não, no componente fórum e quando se tem atividade em grupo. Referente à variável dificuldade com conteúdo, esta pode ser explorada através do componente arquivo, quando o aluno não acessa o mesmo, quando o aluno não entrega atividade através componente atividade e nem participa de fórum de grupos.

Os componentes elencados para a mineração estão presentes na Figura 28. Juntamente a esses componentes, os dados dos questionários também foram utilizados para a mineração.

Figura 28: Componentes de *logs* utilizados na mineração



Fonte: Elaborado pela autora

Foram aplicados métodos de estatística descritiva utilizando o software SPSS, a fim de preparar os dados para a MDE. Foram realizadas a geração de distribuição de frequência e

histogramas de frequência a fim de organizar dados de mesmo tipo de classe para obter a quantidade de dados em cada classe, bem como foi feita a “discretização” dos dados para transformar atributos contínuos⁴² em atributos discretos⁴³. Também foram gerados quartis, os quais foram usados para definir, por exemplo, a quantidade de envio de atividades em grupo por aluno, adotando os rótulos “baixo”, “baixo médio”, “médio” e “alto”.

Sendo assim, após a separação dos componentes do *log*, conforme apresentado na Figura 28, utilizando o SPSS, também foi necessário separar componentes por aluno, obtendo a frequência de componentes por aluno. Esses valores estavam consolidados, ou seja, ao escolher o componente fórum, era apresentado o total de acessos, visualizações, exclusões, que poderiam ser manipulados. Como os valores desses componentes não estavam separados, foi preciso separá-los para obter informações mais precisas, por exemplo, se o aluno o acessou um determinado fórum, se visualizou o fórum, se postou algo, se respondeu àquele fórum. Dessa forma, foi necessário obter detalhes de cada componente para chegar a esse nível de informação.

Chegou-se a uma quantidade de 263 atributos a partir dos dados dos componentes de *log*. Ao considerar somente os componentes aplicados na mineração e os dados obtidos dos questionários, a quantidade de atributos foi reduzida para 195. Mesmo assim, como alguns atributos foram criados: conceito final na disciplina, turma, e quartis de envios de atividades, de quantidade de *posts* adicionados em fóruns de grupos, quartis de quantidade de visualizações de mensagens nos fóruns, quartis de quantidade de visualizações dos fóruns, quartis de quantidade de visualização e envio de arquivos, o total de atributos chegou a 202.

A próxima etapa do processo de MDE, a *transformação dos dados*, dedicou-se a transpor os dados para o tipo de arquivo a ser utilizado pelo minerador Weka, ou seja, os dados foram transformados em arquivos do tipo ARFF (necessários para o Weka), e adaptados para o algoritmo utilizado.

Para preparar os arquivos do tipo ARFF, foi necessário que eles estivessem no formato apropriado para cada tipo de algoritmo de mineração. Foi preciso agrupar o tipo de família afetiva selecionada pelo aluno, no caso das respostas usadas na REA 2.0, nos referidos quadrantes de cada família, o que proporcionou diminuir de 23 possibilidades de busca para 7 (satisfeito, insatisfeito, animado, desanimado, nenhuma, outra, não respondeu). Para o caso do conceito final, foram usadas as menções possíveis utilizadas pela professora da disciplina. Por

⁴² Atributos contínuos assumem qualquer valor em um determinado intervalo.

⁴³ Atributos discretos assumem valores finitos.

fim, alguns tipos de dados foram discretizados e gerados quartis para alguns atributos, conforme justificado acima. A Figura 29 apresenta um exemplo de arquivo ARFF.

Figura 29: Exemplo de arquivo ARFF

```

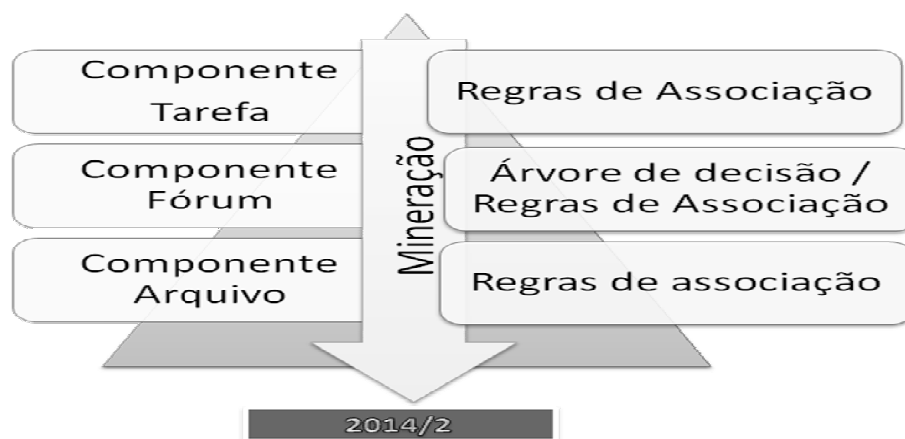
1 @relation J48_teste // para a tese
2
3 @attribute q2 {Satisfeito,Animado,Desanimado,Insatisfeito,Nenhuma,Outra,N_respondeu}
4 @attribute Grupo_Envio_Atraso {Nunca,Raramente,AlgumasVezes,QuaseSempre,Sempre}
5 @attribute Grupo_Sente_Dificuldade {Nunca,Raramente,AlgumasVezes,QuaseSempre,Sempre}
6 @attribute Disciplina_Consegue_AcompanharConteudos {Nunca,Raramente,AlgumasVezes,QuaseSempre,Sempre}
7 @attribute Disciplina_TeveDificuldadeConteudo {Nunca,Raramente,AlgumasVezes,QuaseSempre,Sempre}
8 @attribute ConceitoFinal {A,B,C,D,E,FF}
9 @attribute FGrup_viewDiscussion_Quartil {Baixo,BaixoMedio,Medio,Alta}
10 @attribute FGrup_viewForum_Quartil {Baixo,BaixoMedio,Medio,Alta}
11 @attribute FGrup_addPost_Quartil2 {Baixo,BaixoMedio,Medio,Alta}
12
13 @data
14
15 Insatisfeito,Nunca,AlgumasVezes,QuaseSempre,AlgumasVezes,B,BaixoMedio,Baixo,BaixoMedio
16 Desanimado,Nunca,AlgumasVezes,QuaseSempre,AlgumasVezes,B,Alta,Alta,Medio

```

Fonte: Elaborado pela autora

Somente após as etapas supracitadas, as etapas de *mineração* e de *interpretação* foram realizadas, e são descritas nas subseções seguintes, sendo realizadas individualmente para cada componente de *log* (tarefa, fórum e arquivos) juntamente com um algoritmo específico. A Figura 30 apresenta o modelo para mineração dos componentes, juntamente com os métodos de mineração para cada um.

Figura 30: Componentes para mineração



Fonte: Elaborado pela autora

A seguir, a descrição e a mineração de cada componente são apresentadas.

5.1 COMPONENTE TAREFA

O *componente tarefa* diz respeito a todas as atividades propostas na disciplina. Para essa turma, foram propostas 12 atividades as quais foram analisadas, tanto em termos de visualização, quanto de submissão, considerando que cada aluno poderia enviar e visualizar até 12 atividades.

Para isso, foi necessário compreender esse *log*, identificando seus eventos, componentes e contexto. Dessa maneira, o Quadro 5 apresenta os itens do *log* tarefa e, a partir desse *log* foi criado um arquivo SPSS, retratando eventos por contexto no componente tarefa.

Quadro 5: Exemplo dos itens do *log* componente tarefa

CONTEXTO: Tarefa Aula 20/10 Arquivo Word Setorial					
COMPONENTE: Tarefa					
EVENTO					
Assign_submission statement accepted	Assign_submit	Assign_submit for grading	Assign_view	Assign_view confirm submit assignment form	Assign_view submit assignment form

Fonte: Elaborado pela autora

Para cada componente tarefa, há 6 eventos possíveis: *assign_submission statement accepted*, *assign-submit*, *assign-submit for grading*, *assign_view*, *assign_view confirm submit assignment form*, *assign_view submit assignment form*. Dos 6 eventos apresentados, foram selecionados para mineração os eventos visualização (*assign_view*) e envio de atividades (*assign_submit*), pois possuem maior representatividade, tanto em quantidade de acesso quanto em função do conteúdo, além de resumirem praticamente os outros eventos e contemplarem o que é preciso para a mineração, quando o aluno visualiza e envia a atividade.

O quartil criado para o componente tarefa diz respeito ao atributo quantidade de envios de atividades, devido à necessidade em transformar um atributo contínuo em categórico. Sendo assim, os valores assumidos para os quartis gerados com relação à

quantidade de envios de atividades, foram: Baixo (≤ 6), Médio (> 6 e ≤ 7), Médio Alto (> 7 e ≤ 10) ou Alto (> 10).

Depois que todos os dados de *log* foram preparados, foi gerada uma única planilha no Excel com todas as respostas dos questionários (REA 2.0 e *docs*) e todos os *logs* relativos às atividades para cada aluno, com um total de 58 atributos. Além dos dados dos questionários e dos *logs*, havia os atributos responsáveis pelos quartis, conceito final na disciplina, número de matrícula do aluno e turma.

A partir dos 58 atributos, foi feita uma redução de atributos a partir da premissa de que o componente tarefa aborda as atividades individuais e em grupo, chegando ao conjunto de 14 atributos possíveis. A partir dessa redução, utilizou-se o algoritmo *CfsSubsetEval* ao conjunto de dados com intuito de reduzir e confirmar o uso desses atributos. Esse algoritmo faz parte do Weka.

O algoritmo *CfsSubsetEval* avalia a capacidade preditiva de cada atributo e o grau de redundância entre eles, utilizando o conjunto de atributos que são correlacionados com o atributo classe ou objetivo, mas com baixa intercorrelação (WITTEN; FRANK, 2005). Esse algoritmo é baseado em correlação, avaliando a capacidade preditiva de cada atributo no subconjunto, juntamente com o grau de redundância entre eles (NETTO, 2013).

Quando se opta por um algoritmo de seleção/avaliação de atributos no Weka, torna-se necessário escolher também um método de busca, que no caso da presente pesquisa foi o *BestFirst*. O método de busca *BestFirst* encontra um bom subconjunto de atributos (WITTEN; FRANK, 2005).

Sendo assim, foi aplicado o algoritmo *CfsSubsetEval* ao conjunto de dados, selecionado o método *BestFirst* e o modelo de seleção de atributos *full training set*. O conjunto de dados foi ajustado para os atributos classe ou objetivo q1, depois para o atributo q2 e q5, respectivamente, como pode ser observado na Figura 31. Do conjunto de atributos selecionados pelo algoritmo, todos foram usados para a mineração e, além deles, alguns atributos considerados importantes pelo especialista também foram incluídos.

Figura 31: Seleção de atributos

```

Attribute Subset Evaluator (supervised, Class (nominal)) { 1 q1}:
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 3,4,5,6 : 4
  Individual_Sente_Dificuldade
  Disciplina_TeveDificuldadeConteudo
  ConceitoFinal
  QuartilEnvioSIM

```

```

Attribute Subset Evaluator (supervised, Class (nominal)) { 1 q20}:
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 3,4,5,6 : 4
  Grupo_Sente_Dificuldade
  Disciplina_TeveDificuldadeConteudo
  ConceitoFinal
  QuartilEnvioSIM

```

```

Attribute Subset Evaluator (supervised, Class (nominal)) { 1 q5}:
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 2,3,4,5 : 4
  Disciplina_Consegue_AcompanharConteudos
  Disciplina_TeveDificuldadeConteudo
  ConceitoFinal
  QuartilEnvioSIM

```

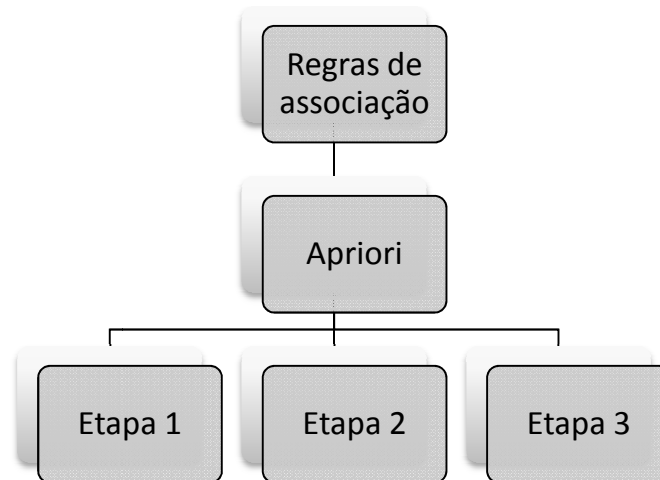
Fonte: Elaborado pela autora

Os atributos selecionados estão listados na Tabela 5 e, a partir desses atributos, foi criado o arquivo ARFF para ser submetido ao minerador Weka e dar início ao processo de mineração o qual é exposto na seção 5.1.1.

5.1.1 Mineração de dados – Componente Tarefa

Após realizar as etapas iniciais da *descoberta de conhecimento*, desde a seleção dos dados até a sua *transformação*, a mineração foi realizada. Para a etapa de mineração, utilizou-se a Tarefa de Regras de Associação através do algoritmo Apriori. A figura 32 apresenta um esquema adotado na mineração.

Figura 32: Mineração componente tarefa



Fonte: Elaborado pela autora

A mineração ocorreu em três etapas, com enfoque em separar em etapas para conquistar, buscando a implicação por questão (REA 2.0) e variáveis observáveis (*log* de atividades), a fim de encontrar um padrão de comportamento por contexto de questão e depois verificar se os padrões se repetiam, e considerá-los como globais.

Nesse experimento, o algoritmo Apriori foi configurado para gerar 10 regras para cada etapa, onde esta quantidade de regras mostrou um resultado mais significativo, sem redundâncias de regras. Além da definição da quantidade de regras, também foi necessário definir valores para o suporte e confiança. Os valores das métricas suporte e confiança teve sustentação teórica apresentada na seção 3.2.2, no tópico mineração de Regras de Associação.

Dessa forma, foram estipulados valores para o fator de suporte mínimo de 0.2 e suporte máximo 1.0 para as 44 instâncias, podendo ser ajustáveis em um minerador.

De acordo com exposto na seção 3.2.2., ao abordar sobre a mineração de Regras de Associação, valores de interesse como *confidence* (confiança), *lift* e *conviction* foram empregados. Conforme aponta Assunção (2012), as regras de interesse geradas que possuem maiores valores dos índices de interesse (*conviction*, *lift* e *confidence*) são consideradas as mais significativas.

Foram consideradas válidas somente as 10 regras mais significativas cujos valores adotados para as medidas de interesse foram o mínimo de 0,70 para a confiança e mínimo de 1.10 para *lift* e *convicção*. As regras cujos valores das medidas de interesse fossem menores que o estipulado não foram selecionadas.

O algoritmo Apriori foi aplicado em três etapas as quais apresentam os atributos utilizados (Tabela 5) e, posteriormente, a listagem das regras selecionadas (Tabela 6). Destaca-se que se encontram nos Apêndices C, D e E os resultados da mineração de cada etapa.

Tabela 5: Atributos usadas em cada etapa

R.	NOME ATRIBUTO	CATEGORIAS
ETAPA 1 – q1	Q1(O que você sente quando envia atividades individuais atrasadas? - aplicada usando a REA 2.0	Satisfeito, insatisfeito, animado, desanimado
	Individual_envio_atraso (Você envia atividades individuais com atraso?	Nunca, raramente, algumas vezes, quase sempre, sempre
	Individual_sente_dificuldade (Você sente dificuldades em realizar atividades individuais com atraso?	Nunca, raramente, algumas vezes, quase sempre, sempre
	Disciplina_dificuldadeConteudo (Você já teve dificuldade com algum conteúdo na disciplina?	Nunca, raramente, algumas vezes, quase sempre, sempre
	Conceito_Final	A, B, C, D, E, FF
	QuartilEnvioSIM (representa a quantidade de envios das atividades, podendo ter os valores	Baixo (≤ 6), Médio (>6 e ≤ 7), Médio Alto (>7 e ≤ 10) ou Alto (>10)
	NOME ATRIBUTO	CATEGORIAS
ETAPA 2 – q2	Q2 (O que você sente quando envia atividades em grupo atrasadas?)	Nunca, raramente, algumas vezes, quase sempre, sempre
	Grupo_Envio_Atraso (Você envia atividades em grupo com atraso? - respostas: nunca, raramente, algumas vezes, quase sempre e sempre);	Nunca, raramente, algumas vezes, quase sempre, sempre
	Grupo_Sente_Dificuldade (Você sente dificuldade em realizar atividades em grupo?	Nunca, raramente, algumas vezes, quase sempre, sempre
	Disciplina_dificuldade_Conteudo (idem etapa 1)	Nunca, raramente, algumas vezes, quase sempre, sempre
	Conceito_Final (idem etapa 1)	A, B, C, D, E, FF
	QuartilEnvioSIM (idem etapa 1).	Baixo (≤ 6), Médio (>6 e ≤ 7), Médio Alto (>7 e ≤ 10) ou Alto (>10)
	NOME ATRIBUTO	CATEGORIAS
ETAPA 3 – q5	Q5: (O que você sente quando tem dificuldade em algum conteúdo?)	Nunca, raramente, algumas vezes, quase sempre, sempre
	Disciplina_dificuldade_Conteudo (idem etapa 1)	Nunca, raramente, algumas vezes, quase sempre, sempre
	Conceito_Final (idem etapa 1)	A, B, C, D, E, FF
	QuartilEnvioSIM (idem etapa 1).	Baixo (≤ 6), Médio (>6 e ≤ 7), Médio Alto (>7 e ≤ 10) ou Alto (>10)

Fonte: Elaborado pela autora

A etapa de *interpretação* foi apoiada por especialistas de MDE já mencionados e usada para suporte de tomadas de decisão na disciplina em questão. A estratégia de dividir em etapas auxiliou a busca pelo objetivo do trabalho, evidência do desânimo. Para cada etapa, 10 regras foram geradas, porém, foram selecionadas somente as regras de medidas de interesse (*conviction lift e confidence*) mais significativas e que continham o atributo “desânimo”. As regras selecionadas para cada etapa foram consolidadas e são exibidas na Tabela 6.

Tabela 6: Resultado das melhores regras geradas para o estado de desânimo.

	N	R	Antecedente	Consequente	Conf	Lift	Conv
ETAPA 1	1	2	q1=Desanimado E Conceito_Final=B 12	Individual_Envio_Atraso=Algumas Vezes 10	0.83	1.47	1.73
	2	9	q1=Desanimado E Individual_Envio_Atraso=AlgumasVezes 14	Conceito_Final=B 10	0.71	1.37	1.34
	3	1 0	Individual_Envio_Atraso=AlgumasVezes E Conceito_Final=B 14	q1=Desanimado 10	0.71	1.37	1.34
ETAPA 2	4	1	q2=Desanimado Conceito_Final=B 12	Grupo_Sente_Dificuldade=Alguma sVezes 10	0.83	1.59	1.91
	5	5	Grupo_Envio_Atraso=Raramente E Disciplina_TeveDificuldadeConteudo=Alg umasVezes 12	q2=Desanimado 10	0.83	1.31	1.45
	6	9	q2=Desanimado E QuartilEnvioSIM=Baixo 10	Disciplina_TeveDificuldadeConteu do=AlgumasVezes 9	0.9	1.2	1.25
ETAPA 3	7	1	q5=Desanimado E QuartilEnvioSIM=MedioAlto 8	Conceito_Final=B 8	1	1.91	3.82
	8	2	q5=Desanimado E QuartilEnvioSIM=Baixo 8	Disciplina_Teve Dificuldade Conteudo=AlgumasVezes 8	1	1.33	2

*N: número da regra.

**R: número da regra encontrada no Weka.

Fonte: Elaborado pela autora

As regras anteriores puderam evidenciar algumas particularidades importantes e são destacadas a seguir.

A partir da **regra 2 da etapa 1**: “*Se o aluno se considera desanimado quando envia atividades individuais em atraso e obteve o conceito final B, então o aluno enviou algumas vezes atividades individuais em atraso*”.

Observa-se que:

- A regra apresenta uma confiança de 83%.
- O valor do *lift* indica que estar desanimado quando envia atividades individuais e possuir conceito = B, alavanca a ocorrência de envio de atividades atrasadas em 1,47, ou seja, para cada aluno que está desanimado quando envia atividades individuais atrasadas e possui conceito B, existe uma chance de 1,47 maior de enviar atividades atrasadas também.
- A medida de convicção indica que o sentido da leitura de regra é o apresentado, pois, sentir-se desanimado e ter um conceito B, acarreta em enviar atividades atrasadas e, não ao contrário, com valor de 1,73.
- Na análise sobre as instâncias selecionadas, foi possível observar que 40% dos alunos enviaram no máximo 7 atividades do curso. Os que se enquadraram nessa regra, foram considerados tendo entre baixo e médio envio de atividades, de acordo com os valores apresentados no quartil de envio de atividades, o que confirma o comportamento do aluno perante a regra encontrada.

Para a **regra 5 da etapa 2** – “*Se enviou raramente com atraso atividades em grupo e teve algumas vezes dificuldade com o conteúdo, então o aluno se sente desanimado com as atividades em grupo*”. Observa-se que:

- A regra apresentou uma confiança de e 83%, ou seja, 83% dos alunos que se sentem desanimados com as atividades em grupo também enviaram atividades atrasadas, mesmo que raramente, e apresentaram algumas vezes dificuldade com conteúdo.
- O *lift* indica que o antecedente alavanca o consequente em 1,31, ou seja, para cada aluno que envia atividades em grupo atrasadas e apresentam dificuldade com o conteúdo algumas vezes, existe uma chance 1,31 maior de ficar desanimado com atividades em grupo.
- A medida de *convicção* avalia a regra como uma verdadeira implicação, informa que o sentido da regra é o apresentado, e não o contrário, com um valor de 1,45.

- Ao analisar as instâncias selecionadas pela regra, foi possível identificar que as participações em fórum de grupos foram pequenas, sendo que somente 20% desses alunos fizeram postagens no fórum e 40% deles praticamente nem visualizaram os fóruns pelo grupo ao qual pertenciam. Desses alunos, somente 20% enviaram mais de 10 atividades, o que é considerado alto envio de atividades, de acordo com o quartil de envios.
- Sobre os 36 arquivos postados pela professora durante a disciplina, observa-se que cerca de 60% dos alunos chegam a visualizar entre 77% a 94% desses arquivos. De forma geral, os alunos participaram pouco dos fóruns de grupo e, mesmo visualizando os arquivos postados pela professora, poucos realmente enviaram praticamente todas as atividades propostas.

Já para a **regra 2 da etapa 3**, “Se o aluno se sente desanimado quando tem dificuldade com algum conteúdo e tem baixos envios de atividades, então ele apresentou algumas vezes dificuldades na disciplina”, observa-se que:

- A regra apresentou uma confiança de 100%.
- O valor do *lift* indica que ficar desanimado quando se tem dificuldade com conteúdo e, o envio de atividades for médio_alto, há uma chance de 1,33% de algumas vezes ter dificuldades com o conteúdo.
- A análise dos dados selecionados indica que 75% dos alunos apresentaram baixo acesso aos arquivos postados no Moodle, o que pode ter refletido na baixa quantidade de envios de atividades, já que poucos arquivos relacionados com o conteúdo foram vistos.
- Os alunos acessaram poucos arquivos, isso pode ter refletido na dificuldade com o conteúdo, já que os conteúdos estão disponíveis por meio desses arquivos não visualizados. Já com relação ao conceito, 75% obtiveram nota igual ou inferior a C, e houve uma desistência.

Sendo assim, conseguiu-se evidenciar, por meio das regras supracitadas, um padrão de comportamento. Sugere-se que estas regras possam ser utilizadas como apoio na evidência do estado de ânimo desanimado do aluno ao realizar atividades individuais ou em grupo. Os alunos que foram identificados pelas regras, através de consulta na base de dados, devem ser

monitorados e bem acompanhados, pois esse padrão de comportamento pode servir de subsídio para apoiar o professor na identificação desse aluno.

5.2 COMPONENTE FÓRUM

Ao analisar os *logs* com relação ao fórum, constatou-se que eles se dividiram em 3 segmentos (Fórum Avisos da Professora; Fórum Contatos com a Monitoria e Fórum dos Grupos de Trabalho), onde somaram 2.815 acessos válidos.

O *log* de fórum dos grupos, em destaque no Quadro 6, é composto por eventos como: *fórum_add_post*, *fórum_add_discussion*, *fórum_view_discussion*, *fórum_view_forum*, *fórum_delete_discussion*, *fórum_delete_post*, *fórum_update_post*. O *log* de monitor é composto pelos eventos *forum_add discussion*, *forum_add post*, *forum_view discussion* e *forum_view fórum*. E por último, o *log* do fórum de avisos da professora é composto pelos eventos *forum_add post*, *forum_view discussion*, *forum_view fórum* e *discussion_mark read*.

Quadro 6: Exemplo dos itens do *log* componente fórum

CONTEXTO: Fórum dos Grupos de Trabalho Setoriais						
COMPONENTE: Fórum						
EVENTO						
forum_add discuss ion	forum_add post	forum_delete discussion	forum_delete post	forum_updat e post	forum_view discussion	forum_view forum

Fonte: Elaborado pela autora

Como o fórum de grupo foi o que obteve maior participação dos alunos, além de ser o fórum em que os alunos poderiam trocar informações sobre seus trabalhos, os três eventos principais que compõem esse fórum foram selecionados para serem utilizados como atributos na mineração. Os eventos selecionados são: as visualizações dos fóruns criados (*forum_view fórum*), as visualizações das discussões de cada fórum (*forum_view discussion*) e as postagens feitas nos fóruns de grupo (*forum_add post*).

Como a quantidade de postagens nesses três eventos do fórum de grupos varia de aluno para aluno, foi necessário criar uma variável baseada em quartil para mensurar a

quantidade de acessos por aluno. A variável responsável pela quantidade de visualizações de discussão poderia ter os valores de acesso baixo, baixo_médio, médio e alto, sendo representada pela variável *FGrup_viewDiscussion_Quartil* {Baixo (≤ 5), Médio-Baixo (> 6 e ≤ 18), Médio (> 18 e ≤ 36), Alto (> 36)}. Da mesma forma, foram criadas variáveis para mensurar a visualização de fóruns e postagem nos fóruns, representadas respectivamente por: *FGrup_viewForum_Quartil* { Baixo(≤ 6), Médio-Baixo (> 6 e ≤ 13), Médio (> 13 e ≤ 27), Alto (> 27)} e *FGrup_addPost_Quartil* { Baixo (≤ 5), Médio-Baixo (> 6 e ≤ 18), Médio (> 18 e ≤ 36), Alto (> 36)}.

Além desses dados de *logs*, foram criados os atributos para totalizar a quantidade de acessos por cada aluno nos três segmentos de fóruns existentes, que foram: *total_forum_prof*, *total_forum_monitor* e *total_forum_grupos*.

Para compor a mineração, também foram utilizadas as respostas aos questionários aplicados, referentes às três questões iniciais 1: *O que você sente quando envia atividades individuais em atraso?*; 2: *O que você sente quando envia atividades em grupo com atraso?*; 3: *O que você sente quando tem dificuldades em realizar atividades individuais?*.

Com relação à segunda etapa do questionário, foram sugeridas as questões 13a, 13b, 14a, 14b, 15a e 15b para mineração. Essas questões utilizaram a escala *Likert* (nunca, raramente, algumas vezes, quase sempre e sempre) e abordavam a frequência com que o aluno enviava atividades individuais/grupo em atraso e se o aluno sentia dificuldades para realizá-las. A escolha das questões se deu pelas tabulações cruzadas encontradas entre elas através do SPSS, as quais se mostraram relevantes para o estudo, apontadas na seção 4.3.1.

Os atributos e métodos usados na mineração são apresentados nas seções subsequentes.

5.2.1 Mineração de dados – Componente Fórum

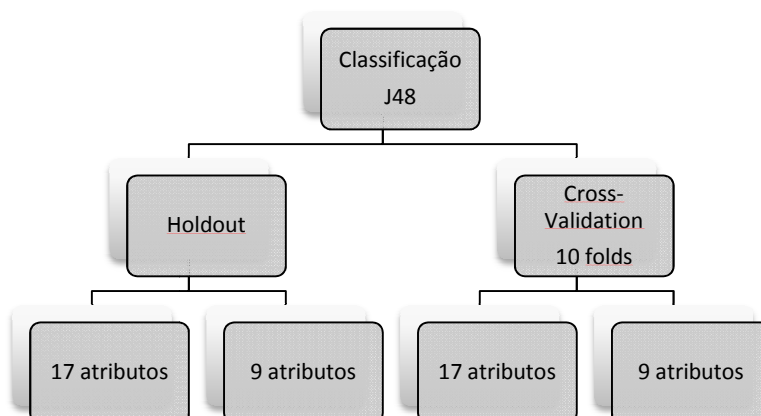
Após passar pelas etapas iniciais da descoberta de conhecimento, desde a seleção dos dados até a transformação desses dados, a mineração pode, então, ser realizada. Foram usados dois métodos de mineração na *etapa de mineração* do componente fórum, as quais incluíram os algoritmos J48 e Apriori.

5.2.2 Classificação J48

O primeiro método usado para a mineração foi Classificação, utilizando aprendizado supervisionado através do algoritmo de Árvore de Decisão J48. Os métodos *Holdout* e *Cross-Validation* foram aplicados a fim de verificar a acurácia dos dois e, ao término das minerações, uma comparação entre eles foi feita a fim de apresentar a melhor escolha para o modelo proposto.

Ambos foram aplicados usando as mesmas instâncias e quantidade de registros, porém, foram feitos dois exemplos de cada, adotando no primeiro experimento 17 atributos e no segundo 9 atributos, conforme exibido na Figura 33.

Figura 33: Mineração e métodos usados na Classificação J48



Fonte: Elaborado pela autora

- Método *Holdout* com 17 atributos

Para esse experimento, foi usado o método *Holdout*, em que os dados foram separados em dois conjuntos de dados, um para treinamento e outro para teste, conforme mencionado na seção 3.2.3. Dessa forma, os dados foram divididos em um conjunto de treinamento (29 instâncias) e um conjunto de testes (15 instâncias), ambos com 17 atributos.

A Figura 34 apresenta os atributos utilizados no treinamento para a geração do modelo preditivo, composta por 16 atributos preditivos e o atributo *q2*, usado como atributo objetivo ou atributo classe.

Figura 34: Atributos usados no modelo de Classificação

Instances: 29		Attributes: 17	
Attributes			
		All	None
		Invert	Pattern
No.	Name		
1	q2		
2	Grupo_Envio_Atraso		
3	Grupo_Sente_Dificuldade		
4	Disciplina_Consegue_AcompanharConteudos		
5	Disciplina_TeveDificuldadeConteudo		
6	FGrup_addDiscussion		
7	FGrup_addPost		
8	FGrup_deleteDiscussion		
9	FGrup_deletePost		
10	FGrup_updatePost		
11	FGrup_viewDiscussion		
12	FGrup_viewForum		
13	FGrup_Total		
14	ConceitoFinal		
15	FGrup_viewDiscussion_Quartil		
16	FGrup_viewForum_Quartil		
17	FGrup_addPost_Quartil2		

Fonte: Elaborado pela autora

Inicialmente, foi feita a mineração do conjunto de dados de treinamento e, posteriormente, aplicou-se o arquivo do conjunto de dados de teste para análise dos resultados minerados. A etapa de *interpretação* foi feita pelos mesmos especialistas (MD e CA) já citados e usados para suporte de tomadas de decisão. Sendo assim, com o modelo preditivo do conjunto de treinamento, foi possível classificar corretamente 72,41% dos dados, como observado pela Figura 35. Também foi evidenciada uma precisão de 0,72% para a Classificação do desânimo. Todas as 21 instâncias que apresentavam o desânimo foram classificadas corretamente.

Figura 35: Resultado do treinamento

```

Correctly Classified Instances      21      72.4138 %
Incorrectly Classified Instances    8      27.5862 %
Kappa statistic                    0.2241
Mean absolute error                0.1164
Root mean squared error            0.2412
Relative absolute error             75.4137 %
Root relative squared error        92.2624 %
Total Number of Instances         29

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      1      0.036   0.5       1      0.667     0.982    Satisfeito
      0      0       0       0       0       ?       Animado
      1      0.778   0.741    1      0.851     0.611    Desanimado
      0      0       0       0       0       0.549    Insatisfeito
      0      0       0       0       0       ?       Nenhuma
      0      0       0       0       0       ?       Outra
      0      0       0       0       0       0.536    N_respondeu
Weighted Avg.  0.724   0.538   0.528   0.724   0.61     0.606

=== Confusion Matrix ===

 a  b  c  d  e  f  g  <-- classified as
1  0  0  0  0  0  0 | a = Satisfeito
0  0  0  0  0  0  0 | b = Animado
0  0  20  0  0  0  0 | c = Desanimado
1  0  6  0  0  0  0 | d = Insatisfeito
0  0  0  0  0  0  0 | e = Nenhuma
0  0  0  0  0  0  0 | f = Outra

```

Fonte: Elaborado pela autora

Após o treinamento, foi aplicado o conjunto de teste. O resultado obtido para o conjunto de teste, representado pela Figura 36, mostra que, com esse conjunto, foi possível classificar corretamente 46,6% dos casos, e classificou corretamente praticamente todas as instâncias cujo desânimo estava presente para a questão 2, à qual somavam-se 8 instâncias com desânimo. Nesse caso, também se teve falsos/positivos classificados erroneamente.

Figura 36: Resultado da mineração J48 – base de teste

```

Correctly Classified Instances      7          46.6667 %
Incorrectly Classified Instances    8          53.3333 %
Kappa statistic                    0.0083
Mean absolute error                 0.1563
Root mean squared error             0.3117
Relative absolute error              91.4345 %
Root relative squared error         106.9559 %
Total Number of Instances          15

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0         0.133    0          0       0          ?        Satisfeito
      0         0         0          0       0          ?        Animado
      0.875     0.857    0.538     0.875   0.667     0.509    Desanimado
      0         0         0          0       0          0.528    Insatisfeito
      0         0         0          0       0          0.5       Nenhuma
      0         0         0          0       0          ?        Outra
      0         0         0          0       0          ?        N_respondeu
Weighted Avg.  0.467    0.457    0.287    0.467   0.356    0.516

=== Confusion Matrix ===

 a b c d e f g  <-- classified as
0 0 0 0 0 0 0 | a = Satisfeito
0 0 0 0 0 0 0 | b = Animado
1 0 7 0 0 0 0 | c = Desanimado
1 0 5 0 0 0 0 | d = Insatisfeito
0 0 1 0 0 0 0 | e = Nenhuma
0 0 0 0 0 0 0 | f = Outra

```

Fonte: Elaborado pela autora

De acordo com Tan, Steinbach e Kumar (2009), a precisão da classificação do modelo pode ser avaliada através da precisão do modelo induzido no conjunto de teste, e pode acontecer de o modelo induzido não ser tão bom como o modelo de treinamento. Diante disso, considera-se que realmente houve algumas limitações do método *Holdout*, portanto, o método *Cross-Validation* será aplicado, servindo de comparação e definição do melhor modelo.

- Método *Cross-Validation* com 17 atributos

O método *Cross-Validation* foi aplicado assumido 10 *Folds*, conforme é detalhado na seção 3.2.2 ao abordar o uso do *Cross-Validation*. Também foi considerado como atributo

objetivo a questão 2 (q2 – como você se sente ao enviar atividades em grupo atrasadas). Nesta mineração, foram utilizados os mesmos atributos da Figura 34.

Dessa forma, a *interpretação* do resultado minerado (Figura 37), apresentou um modelo preditivo com 52,27% de exatidão na Classificação dos dados. O modelo preditivo classificou corretamente 23 instâncias do total de 44. Com relação à acurácia⁴⁴ do modelo, a precisão (precision⁴⁵) da Classificação, ao considerar a classe desanimado, foi de 0.595 e a sensibilidade (recall⁴⁶) foi de 0,786.

Figura 37: Método *Cross-Validation*

```

Correctly Classified Instances      23          52.2727 %
Incorrectly Classified Instances    21          47.7273 %
Kappa statistic                    -0.1422
Mean absolute error                 0.1491
Root mean squared error             0.3203
Relative absolute error             92.9075 %
Root relative squared error         117.4181 %
Total Number of Instances          44

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0         0         0           0         0           0.453    Satisfeito
          0         0         0           0         0           ?        Animado
          0.786    0.938    0.595      0.786    0.677      0.508    Desanimado
          0.077    0.194    0.143      0.077    0.1         0.443    Insatisfeito
          0         0         0           0         0           0.326    Nenhuma
          0         0         0           0         0           ?        Outra
          0         0         0           0         0           0.326    N_respondeu
Weighted Avg.  0.523    0.654    0.421      0.523    0.46       0.479

=== Confusion Matrix ===

 a  b  c  d  e  f  g  <-- classified as
0  0  1  0  0  0  0  | a = Satisfeito
0  0  0  0  0  0  0  | b = Animado
0  0  22  6  0  0  0  | c = Desanimado
0  0  12  1  0  0  0  | d = Insatisfeito
0  0  1  0  0  0  0  | e = Nenhuma
0  0  0  0  0  0  0  | f = Outra

```

Fonte: Elaborado pela autora

Ao observar os resultados e a matriz de confusão⁴⁷ da Figura 37, nota-se que 23 instâncias foram classificadas corretamente e 21 incorretamente. Diante das 28 instâncias nas quais a classe desanimado constava, 22 foram classificadas corretamente e 6 delas incorretamente.

⁴⁴ Acurácia diz respeito ao percentual de amostras positivas/negativas classificadas corretamente.

⁴⁵ Precision é uma medida de desempenho que corresponde à porcentagem de amostras positivas preditas corretamente.

⁴⁶ Recall também é uma medida de desempenho que corresponde à porcentagem de amostras positivas classificadas corretamente.

⁴⁷ Matriz de confusão representa o número de registros da classe classificados corretamente e incorretamente.

Quanto à acurácia do modelo preditivo, tanto a *precision* (precisão) quanto o *recall* (abrangência) apresentaram percentuais satisfatórios para o caso do estado de ânimo desanimado, quando do envio de atividades em grupo atrasado.

A Árvore de Decisão gerada pode ser representada tanto em forma de grafo como em forma de regras, e a árvore gerada por esse modelo é apresentada na Figura 38 no formato de regras.

Figura 38: Árvore de Decisão

```
J48 pruned tree
-----
FGrup_updatePost <= 2
|   FGrup_viewForum_Quartil = Baixo: Desanimado (12.0/1.0)
|   FGrup_viewForum_Quartil = BaixoMedio
|   |   FGrup_addPost_Quartil2 = Baixo
|   |   |   Disciplina_Consegue_AcompanharConteudos = Nunca: Desanimado (0.0)
|   |   |   Disciplina_Consegue_AcompanharConteudos = Raramente: Desanimado (0.0)
|   |   |   Disciplina_Consegue_AcompanharConteudos = AlgumasVezes: Insatisfeito (2.0)
|   |   |   Disciplina_Consegue_AcompanharConteudos = QuaseSempre: Desanimado (4.0)
|   |   |   Disciplina_Consegue_AcompanharConteudos = Sempre: Desanimado (0.0)
|   |   |   FGrup_addPost_Quartil2 = BaixoMedio: Insatisfeito (1.0)
|   |   |   FGrup_addPost_Quartil2 = Medio: Desanimado (3.0)
|   |   |   FGrup_addPost_Quartil2 = Alto: N_respondeu (1.0)
|   |   FGrup_viewForum_Quartil = Medio
|   |   |   FGrup_Total <= 64: Insatisfeito (7.0/1.0)
|   |   |   FGrup_Total > 64: Desanimado (2.0)
|   |   FGrup_viewForum_Quartil = Alto
|   |   |   FGrup_viewForum <= 29: Insatisfeito (3.0/1.0)
|   |   |   FGrup_viewForum > 29: Desanimado (7.0)
FGrup_updatePost > 2: Satisfeito (2.0/1.0)
```

Fonte: Elaborado pela autora

Para esse modelo, a quantidade de atributos pode ter influenciado o resultado, visto que os 17 atributos geraram uma árvore de decisão relativamente grande. Foi realizada uma redução de atributos a fim de prover melhor interpretação.

Sendo assim, após a redução de atributos, a mineração foi realizada com o método *Holdout*.

- Método *Holdout* 9 atributos

Nesta aplicação, foi feita uma redução de atributos utilizando o algoritmo *CfsSubsetEval* e aplicado o método *Holdout*. O método de busca adotado foi *BestFirst* e o modelo de seleção de atributos *full training set*. Sendo assim, o conjunto de dados foi ajustado para o atributo classe q2 e utilizados os 17 atributos (Figura 34) no algoritmo de redução.

O resultado da aplicação do algoritmo é retratado pela Figura 39.

Figura 39: Seleção de atributos J48

```

Attribute Subset Evaluator (supervised, Class (nominal)): { 1 q2 } :
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 3,5,6,8,9 : 5
  Grupo_Sente_Dificuldade
  Disciplina_TeveDificuldadeConteudo
  ConceitoFinal
  FGrup_viewForum_Quartil
  FGrup_addPost_Quartil2

```

Fonte: Elaborado pela autora

Todos os atributos selecionados pelo algoritmo foram usados para a mineração e, além deles, atributos considerados importantes pelos especialistas também foram incluídos, como o caso dos atributos: grupo_envio_atraso, disciplina_consegue_acompanhar_conteúdo e Grup_viewDiscussion_Quartil, além do atributo objetivo.

Dessa forma, a finalidade desse experimento vem ao encontro da proposta reduzida de atributos, sendo que foram utilizados os atributos apresentados na Figura 39, bem como os atributos julgados importantes pelos especialistas da área, totalizando 9 atributos. O Quadro 7 apresenta os atributos utilizados nessa mineração.

Quadro 7: Atributos para mineração

NOME VARIÁVEL	CATEGORIAS
Q2 – O que você sente quanto envia atividades em grupo atrasadas?	Satisfeito, insatisfeito, animado, desanimado
Grupo_Envio_Atraso	Nunca, raramente, algumas vezes, quase sempre, sempre
Grupo_Sente_Dificuldade	Nunca, raramente, algumas vezes, quase sempre, sempre
Disciplina_Consegue_Acompanhar Conteúdos	Nunca, raramente, algumas vezes, quase sempre, sempre
Disciplina_TeveDificuldadeConteudo	Nunca, raramente, algumas vezes, quase sempre, sempre
Conceito Final	A, B, C, D, E, FF
Grup_viewDiscussion_Quartil	Baixo (<=5), Médio-Baixo (>6 e <=18), Médio (>18 e <=36), Alto (>36)
FGrup_viewForum_Quartil	Baixo(<=6), Médio-Baixo (>6 e <=13), Médio (>13 e <=27), Alto (>27)
FGrup_addPost_Quartil2	Baixo (<=5), Médio-Baixo (>6 e <=18), Médio (>18 e <=36), Alto (>36)

Fonte: Elaborado pela autora

Para realizar a mineração, o conjunto de dados foi dividido em dois, treinamento (29 instâncias) e um conjunto de testes (15 instâncias). A mineração feita, utilizando o algoritmo J48 e método *Holdout*, encontrou no treinamento uma exatidão de 68,96%, para a precisão da predição do aluno desanimado em torno de 0,69 e abrangência de 1.

Com relação ao teste, foram classificados corretamente cerca de 53,33% dos registros, como pode ser observado na Figura 40.

Figura 40: Resultado do teste

```

Correctly Classified Instances      8                53.3333 %
Incorrectly Classified Instances    7                46.6667 %
Kappa statistic                    0
Mean absolute error                 0.153
Root mean squared error             0.2946
Relative absolute error             89.5404 %
Root relative squared error         101.0947 %
Total Number of Instances          15

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0         0         0           0         0           ?        Satisfeito
      0         0         0           0         0           ?        Animado
      1         1         0.533       1         0.696       0.5       Desanimado
      0         0         0           0         0           0.5       Insatisfeito
      0         0         0           0         0           0.5       Nenhuma
      0         0         0           0         0           ?        Outra
      0         0         0           0         0           ?        N_respondeu
Weighted Avg.  0.533    0.533    0.284    0.533    0.371    0.5

=== Confusion Matrix ===

 a b c d e f g  <-- classified as
0 0 0 0 0 0 0 | a = Satisfeito
0 0 0 0 0 0 0 | b = Animado
0 0 8 0 0 0 0 | c = Desanimado
0 0 6 0 0 0 0 | d = Insatisfeito
0 0 1 0 0 0 0 | e = Nenhuma
0 0 0 0 0 0 0 | f = Outra

```

Fonte: Elaborado pela autora

Ao analisar a matriz de confusão desse modelo, observa-se que todos os 8 registros que apresentavam desânimo para a questão 2 foram classificados corretamente, sugerindo uma análise isolada para a Predição do desânimo com 100%. Os demais registros não foram classificados corretamente em suas devidas classes, gerando falsos positivos. De acordo com Witten e Frank (2005, p. 162), são considerados como falsos positivos quando o resultado previsto é incorretamente predito como positivo, mas, na realidade, não o é.

- Método *Cross-Validation 9 atributos*

Nessa mineração, foi aplicado o método de teste *Cross-Validation* assumindo 10 *Folds* e os mesmos atributos preditivos do Quadro 7 como treinamento para a geração do modelo preditivo. O atributo q2 foi ajustado como atributo classe ou atributo objetivo.

Dessa forma, a *interpretação* do resultado minerado exibido pela Figura 41 apresentou um modelo preditivo com 59,09% de exatidão na classificação dos dados. O modelo preditivo classificou corretamente 26 instâncias do total das 44, e 18 instâncias foram classificadas incorretamente. Quanto à acurácia do modelo preditivo, foi obtido o valor de 0,649 para *precision* (precisão) e 0,857 para o valor de *recall* (abrangência).

Figura 41: Mineração J48 – *Cross-Validation* 9 atributos

```

Correctly Classified Instances      26          59.0909 %
Incorrectly Classified Instances    18          40.9091 %
Kappa statistic                    0.021
Mean absolute error                 0.1411
Root mean squared error             0.2934
Relative absolute error             87.9203 %
Root relative squared error         107.5498 %
Total Number of Instances          44

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0         0         0           0         0         0.314    Satisfeito
          0         0         0           0         0         ?        Animado
          0.857    0.813    0.649    0.857    0.738    0.54     Desanimado
          0.154    0.161    0.286    0.154    0.2       0.445    Insatisfeito
          0         0         0           0         0         0.267    Nenhuma
          0         0         0           0         0         ?        Outra
          0         0         0           0         0         0.337    N_respondeu
Weighted Avg.  0.591    0.565    0.497    0.591    0.529    0.496

=== Confusion Matrix ===

 a  b  c  d  e  f  g  <-- classified as
0  0  1  0  0  0  0  | a = Satisfeito
0  0  0  0  0  0  0  | b = Animado
0  0  24  4  0  0  0  | c = Desanimado
0  0  11  2  0  0  0  | d = Insatisfeito
0  0  0  1  0  0  0  | e = Nenhuma
0  0  0  0  0  0  0  | f = Outra
0  0  1  0  0  0  0  | g = N_respondeu

```

Fonte: Elaborado pela autora

Ao observar a matriz de confusão, das 44 instâncias do conjunto de treinamento, 26 instâncias foram classificadas corretamente pelo algoritmo J48.

Para este experimento, sugere-se que a árvore de decisão obtida (Figura 42) possa compor o Modelo de Predição do desânimo.

Figura 42: Árvore de Decisão

```

J48 pruned tree
-----

FGrup_viewForum_Quartil = Baixo: Desanimado (12.0/1.0)
FGrup_viewForum_Quartil = BaixoMedio
|   FGrup_addPost_Quartil2 = Baixo
|   |   Disciplina_Consegue_AcompanharConteudos = Nunca: Desanimado (0.0)
|   |   Disciplina_Consegue_AcompanharConteudos = Raramente: Desanimado (0.0)
|   |   Disciplina_Consegue_AcompanharConteudos = AlgumasVezes: Insatisfeito (2.0)
|   |   Disciplina_Consegue_AcompanharConteudos = QuaseSempre: Desanimado (4.0)
|   |   Disciplina_Consegue_AcompanharConteudos = Sempre: Desanimado (0.0)
|   FGrup_addPost_Quartil2 = BaixoMedio: Insatisfeito (1.0)
|   FGrup_addPost_Quartil2 = Medio: Desanimado (3.0)
|   FGrup_addPost_Quartil2 = Alto: N_respondeu (1.0)
FGrup_viewForum_Quartil = Medio: Insatisfeito (10.0/4.0)
FGrup_viewForum_Quartil = Alto: Desanimado (11.0/3.0)

```

Fonte: Elaborado pela autora

5.2.2.1. Considerações sobre a Classificação J48

Para Tan, Steinbach e Kumar (2009, p. 224), dependendo da quantidade de dados das amostras, a diferença observada da precisão entre dois classificadores pode não ser estatisticamente significativa, necessitando-se realizar alguns testes estatísticos com o fim de analisar as diferenças observadas. Para o autor, em um determinado conjunto de testes, um modelo de 30 registros obtém 85% de precisão e outro conjunto de testes com 5.000 registros obtém 75% de precisão, não sendo possível comparar os classificadores baseado somente no grau de precisão dos mesmos. Para os autores, quando há diferenças no conjunto de dados e no tamanho da amostra, a comparação através da precisão/exatidão não é suficiente.

Portanto, diante dos modelos apresentados na mineração, os quais foram aplicados sob o mesmo conjunto e quantidade de dados, a comparação utilizando o percentual de classificação correta entre os classificadores apresenta-se como suficiente, não necessitando de um método específico para comparação.

A partir de uma consolidação dos resultados, a Tabela 7 apresenta a precisão de cada modelo e apoiará a escolha do melhor modelo de classificação para os dados minerados através do algoritmo J48.

Tabela 7: Precisão dos modelos

ATRIBUTOS	MÉTODO	EXATIDÃO	PRECISÃO	ABRANGÊNCIA (RECALL)
17 atributos	<i>Holdout</i>	46,66%	0,538	0,875
	<i>Cross-Validation</i>	52,27%	0,595	0,786
9 atributos	<i>Holdout</i>	53,33%	0,533	1
	<i>Cross-Validation</i>	59,09%	0,649	0,857

Fonte: Elaborado pela autora

Buscando analisar os resultados dos modelos que obtiveram melhor escore, com relação à exatidão e árvore de decisão apropriada para apoio à predição do aluno desanimado, o método *Holdout* e *Cross-Validation* apresentaram melhores escores no caso da redução de atributos.

Dentre os modelos, o *Cross-Validation* com 9 atributos apresentou melhor exatidão, demonstrando ser o melhor Modelo de Predição encontrado. Dentre os dois, o *Cross-Validation* com 9 atributos apresentou melhor exatidão com 59,09%.

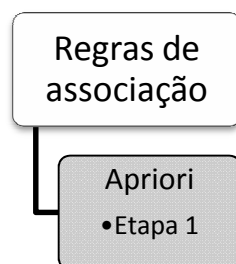
Porém, diante dos resultados obtidos pelo uso do algoritmo J48, novos experimentos foram realizados com o algoritmo Apriori, a fim de compor o modelo mais robusto, e são detalhados na próxima seção.

5.2.3 Algoritmo Apriori

Nesta seção, será descrita a mineração realizada com o objetivo de obter Regras de Associação utilizando o algoritmo Apriori. Foram mineradas informações do fórum, pois o fórum de grupos foi o que obteve maior participação dos alunos, o que propicia a investigação e mineração desses dados.

A Figura 43 apresenta a técnica e métodos abordados nesta seção.

Figura 43: Mineração fórum - Apriori



Fonte: Elaborado pela autora

Para realizar a mineração utilizando o Apriori, foram usados os mesmos atributos do modelo *Cross-Validation* do Quadro 13. O algoritmo Apriori manipula somente variáveis categóricas⁴⁸ ou qualitativas, e todos os nove atributos utilizados possuem estas características.

As regras foram mineradas usando o suporte mínimo de 0,25 para a geração de 20 regras para análise. Nesse exemplo, foi adotada a geração de 20 regras, pois essa quantidade de regras não apresentou redundâncias e foi possível obter regras representativas para o desânimo.

A partir das regras geradas, um processo de refinamento para a escolha das melhores regras foi utilizado, considerando-se regras que apresentassem o atributo objetivo q2 juntamente com as melhores medidas de interesse obtidas. As medidas de interesse utilizaram a confiança mínima de 0,70, bem como *lift* e convicção mínimos de 1.10.

A partir desses parâmetros, as melhores regras obtidas são listadas na Tabela 8 e o resultado da mineração encontra-se no apêndice I.

Tabela 8: Resultado das Regras de Associação selecionadas

N	R.	Antecedente	Consequente	Conf	Lift	Conv
1	3	Visualizações do fórum = Baixo 12	q2=Desanimado 11	0.92	1.44	2.18
2	4	Consegue acompanhar conteúdos da disciplina = Quase Sempre E Postagens no fórum = Baixo 12	q2=Desanimado 11	0.92	1.44	2.18
3	5	Postagens no fórum = Baixo 18	q2=Desanimado 15	0.83	1.31	1.64
4	7	Disciplina_TeveDificuldadeConteudo=AlgumasVezes E FGrup_addPost_Quartil2=Baixo 13	q2=Desanimado 11	0.85	1.33	1.58

⁴⁸ Variáveis categóricas são expressas por categorias ou palavras.

5	11	q2=Desanimado E Consegue acompanhar conteúdos da disciplina = Quase Sempre 21	Dificuldade com Conteúdo = Algumas Vezes 18	0.86	1.14	1.31
6	17	Consegue acompanhar conteúdos da disciplina = Quase Sempre E Dificuldade com Conteúdo = Algumas Vezes 25	q2=Desanimado 18	0.72	1.13	1.14

*N: número da regra.

**R: número da regra encontrada no Weka.

Fonte: Elaborado pela autora

Dentre as regras geradas, seis (06) delas foram selecionadas, pois atenderam às medidas de interesse e apresentaram o desânimo para a questão 2. Foram eliminadas 4 regras que, apesar de apresentarem o atributo objetivo, suas medidas de interesse estavam abaixo do estipulado, não sendo representativas para o modelo proposto.

As regras encontradas estão listadas na Tabela 8.

A **regra 3 (R=3)** pode ser interpretada da seguinte maneira: “*Se um aluno visualiza menos de 5 posts (valor do quartil baixo) no fórum de grupos, então ele tem uma propensão a ficar desanimado*”.

- Apresenta uma confiança de 92% das vezes.
- O valor do *lift* indica que o aluno fica desanimado quando envia atividades em grupo atrasadas (q2), é 1,44 vezes maior para os alunos que visualizam menos de 5 *posts* nos fóruns de grupo.
- A medida de *convicção* indica que o aluno estar desanimado (q2) sem que ocorra uma baixa visualização de *posts* de grupo é de 2,18 vezes menor do que o esperado.

A partir da análise acima, é possível interpretar as demais regras e, cabe salientar que as regras apresentadas estão ordenadas de acordo com as melhores medidas de interesse, sendo que as primeiras regras são consideradas as mais importantes.

5.2.4 Considerações finais sobre a mineração do componente fórum

No processo de composição do Modelo de Predição, que utiliza variáveis e regras do componente fórum, aplicou-se duas técnicas de mineração. As técnicas aplicadas foram a Classificação e a Associação.

Para a Classificação, foram adotados os métodos *Holdout* e *Cross-Validation*, sendo que o *Cross-Validation* com 9 atributos obteve melhor precisão da Classificação para o estado de ânimo desanimado.

Para a Associação, foram usadas Regras de Associação e aplicação do algoritmo Apriori. As melhores regras foram elencadas a partir das melhores métricas de medidas de interesse. As RA encontradas estão listadas na Tabela 8.

Sendo assim, sugere-se que regras de Classificação obtidas através do algoritmo J48 (*Cross-Validation*) e Regras de Associação obtidas através do algoritmo Apriori sejam usadas para compor o Modelo de Predição.

5.3 COMPONENTE ARQUIVO

O componente arquivo diz respeito aos arquivos postados pela professora para a disciplina e é a partir dele que se torna possível saber quais alunos acessaram os arquivos postados. Esta investigação pode vir a refletir se o aluno acessou o material, se realizou atividades referentes a esse material, se consultou o material quando teve dúvidas, entre outras.

O componente arquivo foi composto por 36 arquivos do tipo *pdf*. No Moodle, o componente arquivo é composto por um único evento, o *resource_view*, pois cabe ao aluno somente visualizar ou não um determinado arquivo. Inclusive, o aluno pode visualizar o mesmo arquivo várias vezes. O Quadro 8 apresenta as características presentes no *log* do componente arquivo.

Quadro 8: Características do componente arquivo

CONTEXTO: Ciclos operacionais
COMPONENTE: Arquivo
EVENTO
Resource_view

Fonte: Elaborado pela autora

O evento *resource-view* apresenta a quantidade de acessos por arquivo, mas não diz quantos arquivos dos 36 foram visualizados. Para obter esse tipo de dado, os dados do *log* do componente arquivo foram pré-processados e tabulados para cada arquivo de cada aluno. Foram criadas algumas variáveis, como a quantidade de arquivos visualizados, a quantidade de arquivos não visualizados e a quantidade de acessos aos arquivos, pois é possível que o aluno acesse, por exemplo, um arquivo cinco vezes, tornando possível obter essa informação.

Em média, os alunos visualizam 76% dos arquivos. Há alunos que visualizaram todos os 36 arquivos, porém, alguns alunos visualizaram somente 4 arquivos. A quantidade de acessos aos arquivos variou de 10 a 172. Dessa forma, observa-se que há uma variância muito grande de acessos aos arquivos da disciplina, o que torna necessário separar essas informações para serem mineradas.

A partir da preparação dos dados de *log*, foi gerada uma única planilha com todas as respostas dos questionários (REA 2.0 e docs), e todos os *logs* relativos às atividades para cada aluno, totalizando de 89 atributos. Além dos dados dos questionários e dos *logs*, havia os atributos responsáveis pelos quartis, conceito final na disciplina, número de matrícula do aluno e turma. Porém, como o componente tarefa incide sobre o conteúdo proposto durante a disciplina, sendo conduzido por materiais postados pela professora, esse tipo de dado é restrito individualmente ao aluno.

Dessa forma, deu-se foco para as questões da REA 2.0 que abordassem informações individuais, e o uso dos quartis apoiou o delineamento de uma melhor análise e redução da quantidade de atributos usados para os 36 arquivos. Para apoio na redução dos atributos, também foi aplicado o algoritmo *CfsSubsetEval* juntamente com o método de busca *BestFirst* e modelo de seleção de atributos *full traing set*.

A partir dessas configurações, o conjunto de dados foi ajustado, respectivamente, para os atributos classes q1, q3 e q5. Os resultados da seleção dos melhores atributos são apresentados na Figura 44.

Figura 44: Atributos selecionados

```

Attribute Subset Evaluator (supervised, Class (nominal): 1 q1):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 2,3,4,6,7,8 : 6
Individual_Envio_Atraso
Individual_Sente_Dificuldade
Disciplina_Consegue_AcompanharConteudos
ConceitoFinal
ARquivo_quartil
QuartilSim

```

```

Attribute Subset Evaluator (supervised, Class (nominal): 1 q3):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 2,3,4,6,8 : 5
Individual_Envio_Atraso
Individual_Sente_Dificuldade
Disciplina_Consegue_AcompanharConteudos
ConceitoFinal
QuartilSim

```

```

Attribute Subset Evaluator (supervised, Class (nominal): 1 q5):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 3,4,5,6,7,8 : 6
Individual_Sente_Dificuldade
Disciplina_Consegue_AcompanharConteudos
Disciplina_TeveDificuldadeConteudo
ConceitoFinal
ARquivo_quartil
QuartilSim

```

Fonte: Elaborado pela autora

Dessa forma, o Quadro 9 apresenta as variáveis/atributos utilizados para a mineração do componente arquivo.

Quadro 9: Variáveis usadas para minerar os arquivos

NOME VARIÁVEL	CATEGORIAS
Q1 – O que você sente quanto envia atividades individuais atrasadas	Satisfeito, insatisfeito, animado, desanimado
Q3 - Como você se sente quanto tem dificuldades em realizar atividades individuais?	Satisfeito, insatisfeito, animado, desanimado
Q5 - Como você se sente quando tem dificuldade em algum conteúdo?	Satisfeito, insatisfeito, animado, desanimado
Individual_Envio_Atraso	Nunca, raramente, algumas vezes, quase sempre, sempre

Individual_Sente_Dificuldade	Nunca, raramente, algumas vezes, quase sempre, sempre
Disciplina_Consegue_AcompanharConteudos	Nunca, raramente, algumas vezes, quase sempre, sempre
Disciplina_TeveDificuldadeConteudo	Nunca, raramente, algumas vezes, quase sempre, sempre
ConceitoFinal	A, B, C, D, E, FF
Quartil-visualizacoesSIM	Baixo (≤ 22), BaixoMedio (>22 e ≤ 29), Medio (>29 e ≤ 33), Alto (>33)
Quartil-Arquivos – resource_view	Baixo(≤ 54), BaixoMedio (>54 e ≤ 76), Medio (>76 e ≤ 111), Alto (>111)

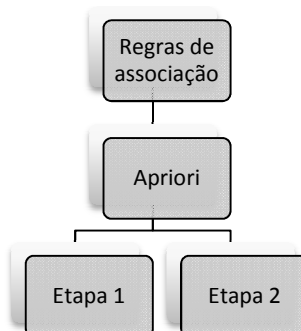
Fonte: Elaborado pela autora

A partir desses atributos, foi criado o arquivo ARFF para ser submetido ao minerador Weka e dar início ao processo de mineração.

5.3.1 Mineração de dados – Componente Arquivo

Nesta etapa de mineração, utilizou-se a Tarefa de Regras de Associação, assim como aplicados ao componente tarefa e ao componente fórum, adotando o algoritmo Apriori. A aplicação da mineração se deu em 2 etapas. A Figura 45 apresenta o esquema visual de técnicas e métodos utilizados.

Figura 45: Mineração componente arquivo – Apriori



Fonte: Elaborado pela autora

Esta mineração ocorreu em duas etapas. A etapa 1 minerou os 10 atributos apresentados no Quadro 9 e o resultado da mineração encontra-se nos apêndices F. A etapa 2 minerou 6 atributos (q1, q3, q5, conceito, quartil_aceessos, quartil_visualizações) e o resultado minerado encontra-se no apêndice G. Para realizar as minerações, foram adotados os valores de suporte mínimo de 0.2 , com a geração de 20 regras para as duas etapas.

Após a mineração, e de posse das regras geradas, o processo de refinamento e escolha das regras foi realizado considerando as melhores métricas de acordo com os valores das medidas de interesse *confiança*, *lift* e *conviction*. Foram consideradas as regras que nas questões q1, q2 e 13 apresentassem o desânimo e que tivessem confiança mínima de 0.70, *lift* e *convicção* mínimos de 1.10.

A partir desses parâmetros, as melhores regras foram elencadas e estão listadas na Tabela 9.

Tabela 9: Regras obtidas para o componente arquivo

	N.	R.	Antecedente	Consequente	Conf	Lift	Conv
ETAPA 1	1	9	q1=Desanimado Individual_Envio_Atraso= AlgumasVezes 14	q5=Desanimado 11	0.79	1.28	1.35
	2	13	q1=Desanimado q5=Desanimado 15	Individual_Envio _Atraso=Algumas Vezes 11	0.73	1.29	1.3
ETAPA 2	3	8	ARquivo_quartil=Medio 12	q5=Desanimado 10	0.83	1.36	1.55
	4	9	ConceitoFinal=B ARquivo_quartil=Medio 8	q5=Desanimado 7	0.88	1.43	1.55
	5	10	QuartilSim=Baixo 11	q3=Desanimado 8	0.73	1.45	1.38
	6	18	q5=Desanimado ARquivo_quartil=Medio 10	ConceitoFinal=B 7	0.70	1.34	1.19

*N: número da regra.

**R: número da regra encontrada no Weka.

Fonte: Elaborado pela autora

A partir das regras supracitadas, pode-se constatar as seguintes considerações com relação à **regra 9 da etapa 1**: “Se o aluno se sente desanimado ao enviar atividades

individuais e envia algumas vezes atividades atrasadas, então, esse aluno tem a propensão a sentir-se desanimado quanto tem dificuldade com algum conteúdo”. Observa-se que:

- A confiança da regra ocorrer é de 79% das vezes;
- O valor do *lift* indica que se o aluno sentir dificuldade com um conteúdo e tiver atividades individuais enviadas atrasadas, aumenta a chance em 1,28 de esse aluno sentir-se desanimado com o conteúdo;
- A medida convicção aponta que o sentido da leitura de regra é o apresentado, não ao contrário, com um valor de 1.35;
- Nos demais dados desses alunos, 36% que pertencem a esta regra entregaram menos da metade das atividades propostas, bem como visualizaram menos da metade dos arquivos propostos pela professora.

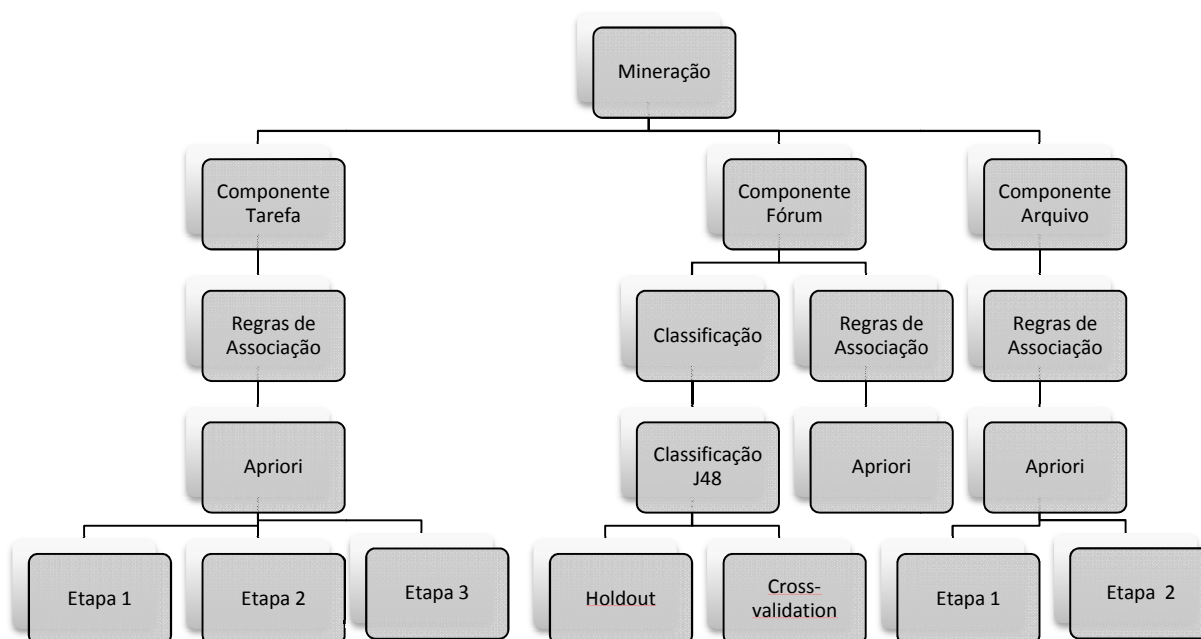
As demais regras podem ser interpretadas de forma equivalente à apresentada.

5.4 CONSIDERAÇÕES SOBRE O CAPÍTULO

O presente capítulo dedicou-se à apresentação detalhada do processo de mineração de cada componente, juntamente com os resultados obtidos de suas minerações e análises.

Através das minerações dos componentes tarefa, fórum e arquivo, foi possível obter melhores técnicas e métodos para cada um dos componentes, assim como selecionar as melhores regras. O esboço geral das minerações realizadas é exibido na Figura 46.

Figura 46: Visão geral das técnicas e métodos minerados



Fonte: Elaborado pela autora

Foram realizadas minerações para os componentes tarefa, fórum e arquivo a fim de identificar variáveis comportamentais e o reconhecimento dos padrões de comportamento do aluno desanimado que estão em interação no AVEA Moodle. A partir das minerações dos dados obtidos das interações dos alunos no AVEA, foram encontradas evidências de Regras de Associação e Regras de Classificação passíveis de identificar o aluno propenso ao desânimo.

A partir das análises estatísticas com o SPSS e das minerações realizadas usando o Weka e seus diversos algoritmos, pode-se chegar aos mecanismos computacionais utilizados para reconhecer o estado de ânimo desanimado, contemplando, todo este capítulo, a questão subsidiária da presente tese: “*Quais mecanismos computacionais podem ser utilizados para reconhecer o estado de ânimo desanimado de um aluno em um ambiente virtual?*”. Este capítulo também contemplou o objetivo específico “*investigar quais tecnologias computacionais podem ser adotadas para apoiar a descoberta do desânimo no aluno*”.

Apresentou todo o processo de como foi feito para obter e validar todo o mapeamento das variáveis observáveis indicadoras do desânimo, contemplando a primeira contribuição dessa tese.

As regras obtidas em cada componente precisam ser reavaliadas em um novo contexto para, então, serem consideradas possíveis de generalização. Dessa forma, uma validação das regras obtidas nesse capítulo em um novo contexto servirá para confirmá-las, permitindo,

assim, a confirmação das regras geradas pela mineração e uma possível generalização das regras para conjuntos de dados futuros.

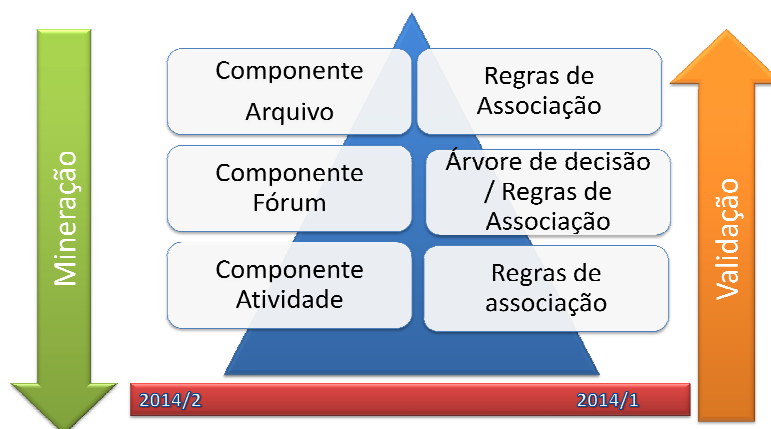
6 EM BUSCA DO MODELO DE PREDIÇÃO DO ALUNO DESANIMADO

Este capítulo é dedicado à apresentação da validação das melhores regras obtidas através das minerações realizadas no capítulo 5 e, a partir dessa validação, a indicação do Modelo de Predição do aluno desanimado.

Para a validação, foi realizado o experimento C, abordando um novo conjunto de dados os quais também dispunham dos componentes arquivo, fórum e atividade. O experimento C é composto pelas turmas de alunos da disciplina de Análise de Demonstrações Contábeis, ofertada pelo Departamento de Ciências Contábeis e Atuariais da UFRGS no ano de 2014/1, constituído por 41 participantes.

Esta validação foi realizada para perfis de estudantes semelhantes aos perfis de alunos realizados na mineração, ou seja, ocorreu com alunos da mesma disciplina, mesmo curso, mesmo professor, possuindo os mesmos materiais educacionais e atividades individuais e em grupo, porém como uma distância temporal de 6 meses. A Figura 47 apresenta um esboço da mineração e validação realizadas.

Figura 47: Validação a partir das regras da mineração



Fonte: Elaborado pela autora

Após a validação, as regras confirmadas com melhores escores foram consideradas para compor o modelo que possa prever o comportamento do aluno desanimado.

6.1 VALIDAÇÃO DA MINERAÇÃO DE DADOS – COMPONENTE TAREFA

A partir do experimento B, que resultou da mineração do componente tarefa, ao qual foi aplicado o algoritmo Apriori em três etapas, conforme descrito na seção 5.1, foram obtidas 8 RA (Tabela 6).

Para validar/testar as RA, inicialmente, os dados dos *logs* dos alunos tiveram que ser preparados e tabulados para serem analisados. Nesse experimento C, havia 11 atividades propostas pelo professor e, diante disso, foi necessário agrupar todas as 11 atividades realizadas no Moodle e analisar o *log* de cada uma delas utilizando os seguintes atributos: *assign-submit* e *assign_view*. Também foi preciso criar dois quartis, um para representar a quantidade de visualizações de atividades (baixo ≤ 7 , baixo médio > 7 e ≤ 9 , médio > 9 e ≤ 10 e alto > 10) e outro quartil para representar a quantidade de atividades enviadas pelos alunos (baixo ≤ 41 , baixo médio > 41 e ≤ 53 , médio > 53 e ≤ 72 e alto > 72), os quais receberam os valores: baixo, baixo médio, médio e alto.

Dessa forma, foram utilizados os mesmos atributos selecionados para a mineração do componente tarefa (Tabela 5), porém com a adequação dos quartis do conjunto de dados da validação. Cada RA apresentada na Tabela 6 foi testada nesse novo conjunto de dados, o que gerou uma consulta à base de dados do experimento C.

Nessa base de dados, analisaram-se os valores obtidos tanto para o antecedente quanto para o consequente de cada regra e os registros os quais apresentaram valores para o antecedente e para o consequente foram armazenados. A partir desses valores, foi calculada a taxa de acerto para cada regra testada (confiança de uma regra). As taxas de acerto das RA testadas no novo conjunto de dados podem ser observadas na Tabela 10.

Tabela 10: Taxas de acerto da validação das regras

N. Regra	Taxa de acerto / Confiança	N. Regra	Taxa de acerto / Confiança
1	Sem representatividade	5	100%
2	Sem representatividade	6	58%
3	Sem representatividade	7	70%
4	Sem representatividade	8	60%

Fonte: Elaborado pela autora

Nota-se que algumas regras na validação ficaram sem representatividade, pois não apresentaram nenhum registro com o padrão de comportamento da regra, porém, esse fato ocorreu somente quando o atributo Conceito final = B estava presente na regra. Esse fato foi analisado pelo especialista da área e notou-se que, nesse conjunto de dados da amostra, havia 1 único aluno com esse conceito.

Dessa forma, ao apresentar os dados para o especialista da área a fim de verificar a significância da regra, o mesmo sugeriu adaptar as regras que tinham o atributo “Conceito Final = B”, para um conceito que incluísse o valor B e C, chegando ao “Conceito Final \geq C e \leq B”. Sendo assim, após a adaptação das regras, as mesmas foram novamente aplicadas à amostra, apresentando melhoria para o uso de tais regras, pois representaram a realidade da turma ao considerar conceitos B e C. Dessa forma, as taxas de acerto foram novamente recalculadas e apresentadas na Tabela 11.

Tabela 11: Taxas de acerto da validação das regras adaptadas

N. Regra	Taxa de acerto	N. Regra	Taxa de acerto
1	59%	5	100%
2	83%	6	58%
3	100%	7	75%
4	62%	8	60%

Fonte: Elaborado pela autora

Após, foi possível identificar as melhores RA que puderam ser utilizadas nesse contexto. As regras sugeridas são as que possuem taxa de acerto com a confiança acima de 70%. O valor de 70% é o mesmo grau de confiança adotado no modelo de mineração do componente tarefa, exposto no capítulo 5.

A fim de confirmar as regras obtidas anteriormente, também foi feita a mineração nesse experimento utilizando os mesmos atributos e mesmas divisões de etapas. Todas as regras com 100% de taxa de acerto ocorreram também na mineração e tiveram uma taxa de confiança de 100%. As demais regras que foram utilizadas com taxa de acerto acima de 70% puderam ser evidenciadas com o mesmo grau de confiança da regra, sugerindo a possibilidade de generalização dessas regras para compor um único modelo.

Portanto, sugere-se que as regras de número 2, 3, 5 e 7 possam ser utilizadas para compor um modelo de apoio na descoberta do aluno desanimado, ou que se encontra em processo para o desânimo, por serem equivalentes às geradas pelo experimento anterior.

6.2 VALIDAÇÃO DA MINERAÇÃO DE DADOS – COMPONENTE FÓRUM

A validação do componente fórum foi feita com os melhores resultados obtidos em suas minerações, tanto para o algoritmo J48 usando os métodos *Holdout 9 atributos* e *Cross-Validation 9 atributos*, quanto para o Apriori, utilizando os mesmos atributos do Quadro 7. O processo para a validação está descrito na seção 6.2.1 e 6.2.2.

6.2.1 Classificação J48

- Método *Holdout* – 9 atributos

Na validação do experimento C, o conjunto de dados utilizados para o método *Holdout* foi dividido em dois, onde o treinamento contou com 26 instâncias e um conjunto de testes com 15.

A mineração utilizando o algoritmo J48, através do método *Holdout*, obteve no treinamento uma exatidão de 88,46%, com precisão para a predição do aluno desanimado em torno de 0.885 e abrangência igual a 1.

No resultado do teste, foi possível classificar corretamente 80% dos registros quando o atributo q2 foi considerado objetivo, obtendo uma precisão para a predição do aluno desanimado em torno de 0.8 e *recall* de 1, conforme apresentado na Figura 48.

Figura 48: Resultado da validação componente fórum – *Holdout*

```

Correctly Classified Instances      12          80  %
Incorrectly Classified Instances    3          20  %
Kappa statistic                     0
Mean absolute error                 0.0813
Root mean squared error             0.2265
Relative absolute error              70.0918  %
Root relative squared error         100.3518  %
Total Number of Instances          15

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	?	Satisfeito
	0	0	0	0	0	?	Animado
	1	1	0.8	1	0.889	0.5	Desanimado
	0	0	0	0	0	0.5	Insatisfeito
	0	0	0	0	0	0.5	Nenhuma
	0	0	0	0	0	?	Outra
	0	0	0	0	0	0.5	N_respondeu
Weighted Avg.	0.8	0.8	0.64	0.8	0.711	0.5	

```

=== Confusion Matrix ===
 a  b  c  d  e  f  g  <-- classified as
0  0  0  0  0  0  0  | a = Satisfeito
0  0  0  0  0  0  0  | b = Animado
0  0  12  0  0  0  0  | c = Desanimado
0  0  1  0  0  0  0  | d = Insatisfeito
0  0  1  0  0  0  0  | e = Nenhuma
0  0  0  0  0  0  0  | f = Outra
0  0  1  0  0  0  0  | g = N_respondeu

```

Fonte: Elaborado pela autora

Com relação à matriz de confusão, a classificação foi feita corretamente para os 12 alunos considerados desanimados ao enviar atividades em grupo atrasadas, fato similar ao resultado da mineração do experimento B, onde todas as instâncias que possuíam o desânimo foram classificadas corretamente, sugerindo um bom modelo preditivo.

Três instâncias foram classificadas erroneamente, gerando falsos positivos, tanto para o modelo de validação, quanto para o modelo preditivo da mineração inicial. Mesmo com alguns falsos/positivos, todas as instâncias em que efetivamente constava o estado de desânimo foram classificadas corretamente. Dessa forma, sugere-se que esse modelo possa ser usado para compor o modelo de predição para evidência considerando o desânimo.

- Método *Cross Validation* – 9 atributos

Para a realização da validação do método de teste *Cross-Validation*, os mesmos parâmetros e atributos foram utilizados, assumindo 10 *Folds*, e também foram utilizados os atributos do Quadro 7. Nesse experimento, todos os atributos são considerados atributos preditivos, exceto o atributo q2, que será ajustado como atributo classe ou atributo objetivo.

A interpretação do resultado minerado (Figura 49) aponta que o modelo preditivo classificou corretamente cerca de 85,36% das instâncias, o que corresponde a 35 instâncias classificadas corretamente do total das 41 instâncias. A acurácia do modelo preditivo foi de

0.854 e de abrangência 1 ao considerar o estado de ânimo desanimado. As 35 instâncias que contemplavam o aluno desanimado, quando enviava atividades em grupo atrasadas, foram classificadas corretamente. As demais instâncias caíram no caso dos falsos/positivos, classificando 6 instâncias que seriam de outros estados de ânimo.

Figura 49: Resultado da validação método *Cross-Validation*

```

Correctly Classified Instances      35          85.3659 %
Incorrectly Classified Instances    6          14.6341 %
Kappa statistic                    0
Mean absolute error                0.0765
Root mean squared error            0.1981
Relative absolute error            74.0303 %
Root relative squared error        97.761 %
Total Number of Instances         41

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0        0        0           0        0          ?       Satisfeito
      0        0        0           0        0          ?       Animado
      1        1        0.854       1        0.921     0.186   Desanimado
      0        0        0           0        0          0.083   Insatisfeito
      0        0        0           0        0          0.038   Nenhuma
      0        0        0           0        0          ?       Outra
      0        0        0           0        0          0.123   N_respondeu
Weighted Avg.   0.854   0.854   0.729   0.854   0.786   0.173

=== Confusion Matrix ===

 a  b  c  d  e  f  g  <-- classified as
 0  0  0  0  0  0  0 | a = Satisfeito
 0  0  0  0  0  0  0 | b = Animado
 0  0  35  0  0  0  0 | c = Desanimado
 0  0  2  0  0  0  0 | d = Insatisfeito
 0  0  1  0  0  0  0 | e = Nenhuma
 0  0  0  0  0  0  0 | f = Outra
 0  0  3  0  0  0  0 | g = N_respondeu

```

Fonte: Elaborado pela autora

A comparação entre os resultados obtidos da mineração e da validação pelo método *Holdout* e método *Cross-Validation* foi realizada e é apresentada pela Tabela 12.

Tabela 12: Comparação da mineração e validação

MÉTODO		% CLASSIFICAÇÃO
Mineração	<i>Holdout</i>	46,66%
	<i>Cross-Validation</i>	52,27%
Validação	<i>Holdout</i>	80%
	<i>Cross-Validation</i>	85,3%

Fonte: Elaborado pela autora

A partir dessa validação, sugere-se que o método *Cross-Validation* seja utilizado por ter alcançado melhor escore para a Classificação da Predição.

6.2.2 Algoritmo Apriori

A validação do segundo método utilizou Regras de Associação com a aplicação do algoritmo Apriori. Dessa forma, a validação seguiu o mesmo modelo da mineração realizada, e atributos do Quadro 7, bem como as regras obtidas pela Tabela 8.

Cada uma das regras obtidas pela Tabela 8 foi testada. Para cada regra pesquisada na base de dados do experimento C, os dados obtidos do antecedente e do consequente foram armazenados, o que possibilitou calcular as taxas de acerto/confiança para cada regra. As taxas de acerto da validação das RA testadas podem ser observadas na Tabela 22.

Tabela 13: Taxas de acerto da validação das regras do experimento C

Regra	Taxa de acerto / Confiança	Regra	Taxa de acerto / Confiança
1	100%	4	100%
2	100%	5	75%
3	100%	6	80%

Fonte: Elaborado pela autora

Nota-se que todas as RA testadas no experimento C (vide Tabela 13), apresentaram taxas de acerto acima de 70%, que é o valor mínimo adotado descrito na seção 5.2.3. Essa taxa de acerto corresponde ao valor da confiança encontrado quando a regra foi minerada na seção 5.2.3. e que também foi encontrada para a validação das regras expostas.

Associada à pesquisa na base de dados, também foi realizada a mineração dessa validação, gerando as mesmas RAs validadas. O resultado com todas as regras encontra-se no apêndice J.

Diante dos resultados das regras, sugere-se que ***todas as regras*** validadas e que estão listadas na Tabela 13 sejam usadas para compor o Modelo de Predição que evidencie o aluno desanimado quando realiza atividades em grupo, utilizando o componente fórum.

6.3 VALIDAÇÃO DA MINERAÇÃO DE DADOS – COMPONENTE ARQUIVO

A partir do experimento que resultou da mineração do componente arquivo, no qual foi aplicado o algoritmo Apriori em duas etapas, conforme descrito na seção 5.3, foram obtidas 6 RA (Tabela18). A partir dessas 6 regras, foi realizada a validação no experimento C.

Dessa forma, a validação foi dividida em duas etapas de acordo com o procedimento realizado no experimento B. Foi necessário tabular os dados e prepará-los para realizar as consultas das regras obtidas. Esse experimento teve um total de 49 arquivos postados pela professora para que os alunos pudessem acessá-los. A Tabela 14 apresenta uma medida descritiva sobre o componente arquivo.

Tabela 14: Medida descritiva componente tarefa

COMPONENTE TAREFA	VALORES
Média	31,41
Median	33,00
Moda	41
Std. Deviation	10,10
Mínimum	5
Maximum	47

Fonte: Elaborado pela autora

Foi necessário criar um atributo quartil para a quantidade de visualizações de arquivos feitas pelos alunos, onde os valores obtidos foram Baixo (≤ 25) acessos a arquivos, Baixo Médio (>25 e ≤ 33), Médio (>33 e ≤ 39), Alto (>40). Também foi criado um quartil para a quantidade de acessos aos arquivos, pois um mesmo arquivo poderia ser acessado várias vezes, variando entre 6 e 730 acessos para os 49 arquivos disponíveis. Também foi feita a alteração das regras que continham conceito final=B para conceito final $\geq C$ e $\leq B$, conforme já realizado com o componente tarefa.

A partir das regras obtidas e testadas nesse conjunto de dados, também foi calculada a taxa de acerto/confiança para cada regra. As taxas de acerto da validação das RA testadas no novo conjunto de dados podem ser observadas na Tabela 15.

Tabela 15: Taxas de acerto da validação das regras componente arquivo

Regra	Taxa de acerto / Confiança	Regra	Taxa de acerto / Confiança
1	100%	4	60%
2	33%	5	60%
3	70%	6	40%

Fonte: Elaborado pela autora

Das regras validadas no experimento C, somente as regras 1 e 3 obtiveram uma taxa de acerto ou a confiança acima de 70%, que foi o valor estipulado tanto na mineração quanto na validação das regras. Foi possível evidenciar, de forma geral, que a partir dos dados da primeira regra e analisando os demais atributos, os alunos que tiveram baixas visualizações de arquivos também apresentaram baixo envio de atividades e sentiram-se desanimados quando enviavam algumas vezes atividades individuais ou em grupo atrasadas.

Sendo assim, sugere-se que sejam usadas as **regras 1 e 3** (listadas na Tabela 9) para compor o modelo de apoio na evidência do aluno desanimado ou que se encontra em processo para o desânimo, quando analisado o componente arquivo.

6.4 MODELO DE PREDIÇÃO DO ALUNO DESANIMADO

Após a mineração e seleção das melhores regras elencadas no capítulo 5, o processo de validação foi realizado, testando cada uma das melhores regras identificadas para cada componente. Esse processo foi validado em um novo conjunto de dados, conforme descrito nas seções desse capítulo. As regras foram testadas e, quando necessário, foram adaptadas. Além disso, os dados foram minerados para confirmação e validação das regras obtidas.

Cada uma das regras encontradas foi validada, utilizando os mesmos valores mínimos das medidas de interesse, e chegou-se a um conjunto de regras que sugerem a identificação do aluno desanimado em interação no AVEA a partir de suas variáveis comportamentais.

Dessa forma, as regras listadas no Quadro 10 compõem um Modelo de Predição capaz de evidenciar o aluno propenso ao desânimo, ressaltando a sua importância no contexto educacional, na medida em que pode apoiar o professor a identificar esse aluno. O modelo sugerido atende ao objetivo específico “*desenvolver um Modelo de Predição do estado de ânimo desanimado para o aluno em interação no AVEA*”.

Quadro 10: Regras do Modelo de Predição

TAREFA
SE $q1=Desanimado$ E <i>Individual Envio Atraso=Algumas Vezes</i> ENTÃO <i>Conceito Final</i> $\geq C$ e $\leq B$
SE <i>Individual Envio Atraso = Algumas Vezes</i> E <i>Conceito Final</i> $\geq C$ e $\leq B$ ENTÃO $q1=Desanimado$
SE <i>Grupo Envio Atraso = Raramente</i> E <i>Disciplina Teve Dificuldade Conteúdo = Algumas Vezes</i> ENTÃO $q2=Desanimado$
SE $q5=Desanimado$ E <i>Quartil Envio = Medio Alto</i> ENTÃO <i>Conceito Final</i> $\geq C$ e $\leq B$
FÓRUM
SE <i>Visualizações do fórum = Baixo</i> ENTÃO $q2=Desanimado$
SE <i>Consegue acompanhar conteúdos da disciplina = Quase Sempre</i> E <i>Postagens no fórum = Baixo</i> ENTÃO $q2=Desanimado$
SE <i>Postagens no fórum = Baixo</i> ENTÃO $q2=Desanimado$
SE <i>Disciplina Teve Dificuldade Conteúdo = Algumas Vezes</i> E <i>Postagens no fórum=Baixo</i> ENTÃO $q2=Desanimado$
SE $q2=Desanimado$ E <i>Consegue acompanhar conteúdos da disciplina = Quase Sempre</i> ENTÃO <i>Dificuldade com Conteúdo = Algumas Vezes</i>
SE <i>Consegue acompanhar conteúdos da disciplina = Quase Sempre</i> E <i>Disciplina Teve Dificuldade Conteúdo = Algumas Vezes</i> ENTÃO $q2=Desanimado$
ARQUIVO
SE $q1=Desanimado$ E <i>Individual Envio Atraso = Algumas Vezes</i> ENTÃO $q5=Desanimado$
SE <i>Arquivo quartil = Médio</i> ENTÃO $q5=Desanimado$

Fonte: Elaborado pela autora

O modelo é composto por variáveis comportamentais descobertas como indicadoras do desânimo, bem como por Regras de Associação evidenciadas pela mineração e validadas. Através do Modelo de Predição obtido, que é composto por todas as regras do Quadro 10, cuja implementação foi realizada diretamente no AVEA Moodle, como em outros ambientes, torna-se factível na medida em que as variáveis comportamentais estejam presentes no mesmo.

Uma proposta de integração e apresentação do modelo proposto foi desenvolvida e é descrita na seção 6.5.

6.5 VISUALIZAÇÃO DO MODELO PROPOSTO

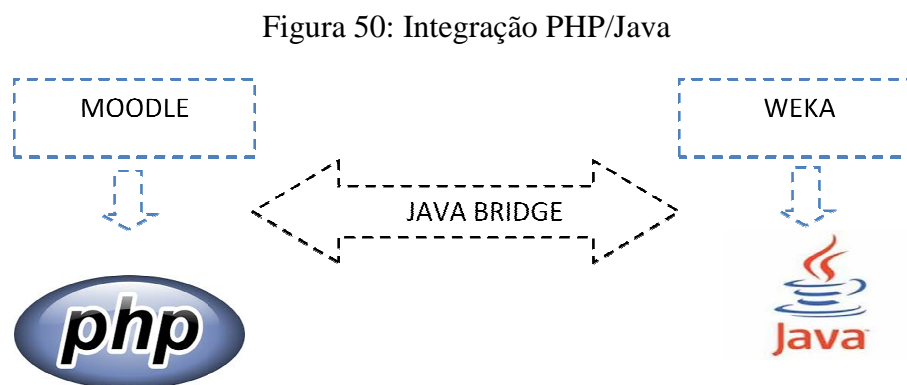
Com o intuito em apresentar o Modelo de Predição do aluno desanimado ao AVEA Moodle, foi realizada a integração do minerador Weka com o Moodle e, posteriormente, com

as regras dentro do Moodle, disponibilizou-se a apresentação delas através de uma ferramenta de visualização em formato de *dashboards*.

As seções seguintes apresentam detalhadamente todo o processo realizado para que fosse possível visualizar o resultado das minerações diretamente no Moodle, dado ora pela integração com o Weka, ora em formato de *Dashboard*.

6.5.1 Integração Weka/Moodle

Para consolidar a integração, inicialmente foi feita a instalação do ambiente Moodle bem como do ambiente Weka. Como o Moodle é desenvolvido utilizando a linguagem de programação PHP (*Hypertext Preprocessor*) e o Weka é desenvolvido em linguagem de programação Java, foi necessário utilizar o Java Bridge⁴⁹ para permitir conectar o PHP juntamente com a máquina virtual java. A Figura 50 apresenta um esquema da integração desenvolvida.

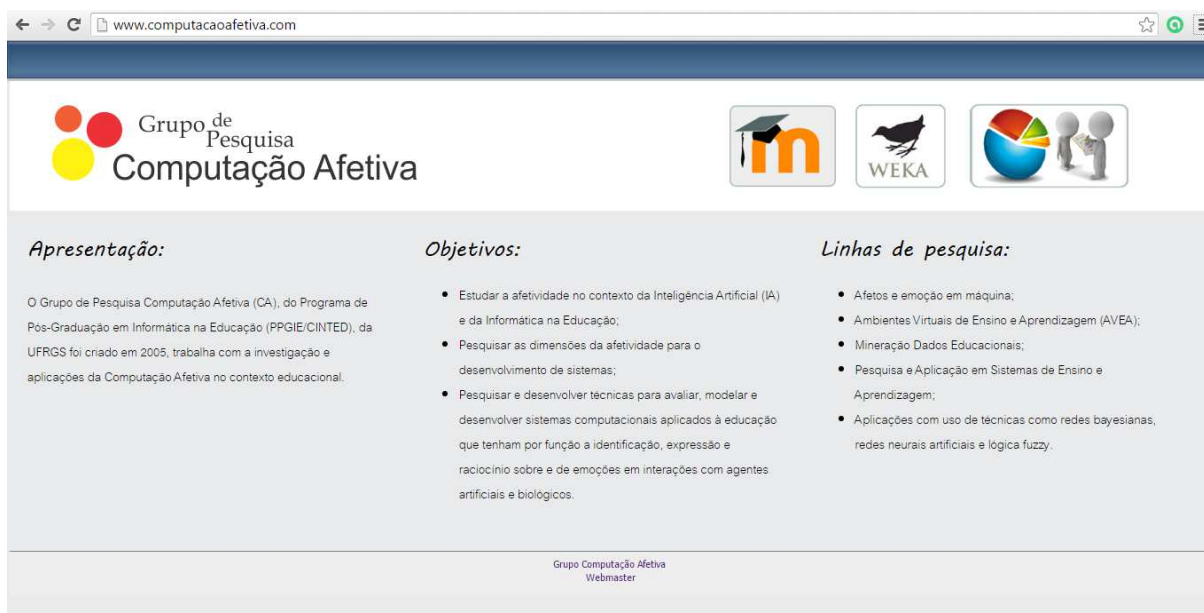


Fonte: Elaborado pela autora

Dessa forma, foi desenvolvido um site que está disponível através do endereço: <http://www.computacaoafetiva.com/> e está hospedado em um servidor que suporta as tecnologias necessárias para a integração e implementação. A Figura 51 apresenta a tela inicial do site, a qual aborda informações sobre o Grupo de Pesquisa Computação Afetiva, bem como botões para acessar o Moodle, Weka e *Dashboard*.

⁴⁹ Java Bridge é um protocolo que permite a implementação e integração de sistemas feitos em linguagens diferentes como o PHP e Java.

Figura 51: Tela inicial do projeto



Fonte: Elaborado pela autora

Estando no ambiente Moodle, é preciso entrar com *login* e senha e seguir normalmente no curso. Os perfis utilizados para usuários na ferramenta desenvolvida são: alunos, professores e o administrador do site. A opção do uso do Weka está disponível somente para professores e para o administrador, via botão implementado no canto superior esquerdo do ambiente.

Através do Weka, integrado ao Moodle, é possível minerar os algoritmos J48 (*Holdout* e *Cross Validation*) e Apriori. Para integrar os demais algoritmos do Weka, é preciso implementá-los através das classes disponíveis em uma página mantida pela Universidade de Waikato: <https://Weka.wikispaces.com/Use+Weka+in+your+Java+code>.

A possibilidade de minerações do Weka dentro do Moodle permite ao professor realizar a mineração dentro do próprio ambiente, onde o arquivo ARFF poderá ser gerado automaticamente pela ferramenta, utilizando as variáveis comportamentais prováveis usadas para o reconhecimento do desânimo.

A Figura 52 apresenta uma lista de arquivos ARFF prontos para serem minerados diretamente no Moodle. Além desses arquivos, também é possível ao usuário inserir um novo arquivo ARFF ao selecionar o botão “Carregar Ficheiro”.

Figura 52: Arquivos ARFF no Moodle

The screenshot shows the Moodle Weka interface. At the top, there's a navigation bar with 'Weka' and 'Dashboard' tabs. Below that, a breadcrumb trail reads 'Página principal > Weka > Mineração'. The main content area displays a table of files with columns: Designação, Tamanho, Modificado, Apriori, J48, N.Regras, and Ação. The 'Apriori' and 'N.Regras' columns are highlighted with red boxes. The 'N.Regras' column has a dropdown menu set to '10'. Below the table, there are buttons for 'Com os ficheiros escolhidos...', 'Selecionar todos', 'Desselecionar', and 'Carregar ficheiro'. At the bottom, there's a footer with 'Documentação Moodle para esta página' and 'Nome de utilizador: Fabricia Damando. (Sair)'.

Fonte: Elaborado pela autora

Após o arquivo ARFF estar no ambiente, o usuário escolhe o tipo de algoritmo com que deseja realizar a mineração e o processo pode iniciar. Caso seja definido o algoritmo Apriori (observar em destaque na Figura 53), é permitido ao usuário selecionar a quantidade de regras necessárias e iniciar a mineração. O resultado é apresentado automaticamente abaixo da lista de arquivos.

A Figura 53 expõe o resultado da mineração do algoritmo Apriori, podendo ser comparado com as regras obtidas pela etapa 2, do componente tarefa, expostos na Tabela 6.

Figura 53: Resultado da mineração do algoritmo Apriori – componente tarefa etapa 2

The screenshot shows the Moodle Weka interface displaying the results of the Apriori algorithm. The output is as follows:

```

Resultado
File: T1 - AtividadesApriori - NOMINAIS_atividadesAlunos2_q2.arff

Instances: 44
Attributes Numbers: 6
Attributes:
q2
Grupo_Envio_Atraso
Grupo_Sente_Dificuldade
Disciplina_TeveDificuldadeConteudo
ConceitoFinal
QuartilEnvioSIM

Apriori
=====

Minimum support: 0.2 (9 instances)
Minimum metric: 1.1
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13
Size of set of large itemsets L(2): 22
Size of set of large itemsets L(3): 6

Best rules found:

1. q2=Desanimado ConceitoFinal=B 12 ==> Grupo_Sente_Dificuldade=AlgumasVezes 10 conf:(0.83) lift:(1.59) lev:(0.08) [3] < conv:(1.91)>
2. QuartilEnvioSIM=Baixo 13 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVezes 12 conf:(0.92) lift:(1.23) lev:(0.05) [2] < conv:(1.63)>
3. QuartilEnvioSIM=MedioAlto 13 ==> ConceitoFinal=B 10 conf:(0.77) lift:(1.47) lev:(0.07) [3] < conv:(1.55)>
4. Disciplina_TeveDificuldadeConteudo=AlgumasVezes ConceitoFinal=B 16 ==> Grupo_Sente_Dificuldade=AlgumasVezes 12 conf:(0.75) lift:(1.43) lev:(0.08) [3] < conv:(1.53)>
5. Grupo_Envio_Atraso=Raramente Disciplina_TeveDificuldadeConteudo=AlgumasVezes 12 ==> q2=Desanimado 10 conf:(0.83) lift:(1.31) lev:(0.05) [2] < conv:(1.45)>
6. Grupo_Sente_Dificuldade=AlgumasVezes 23 ==> ConceitoFinal=B 16 conf:(0.7) lift:(1.33) lev:(0.09) [3] < conv:(1.37)>
7. ConceitoFinal=B 23 ==> Grupo_Sente_Dificuldade=AlgumasVezes 16 conf:(0.7) lift:(1.33) lev:(0.09) [3] < conv:(1.37)>
8. QuartilEnvioSIM=Baixo 13 ==> q2=Desanimado Disciplina_TeveDificuldadeConteudo=AlgumasVezes 9 conf:(0.69) lift:(1.38) lev:(0.06) [2] < conv:(1.3)>
9. q2=Desanimado QuartilEnvioSIM=Baixo 10 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVezes 9 conf:(0.9) lift:(1.2) lev:(0.03) [1] < conv:(1.25)>
10. q2=Insatisfeito 13 ==> ConceitoFinal=B 9 conf:(0.69) lift:(1.32) lev:(0.05) [2] < conv:(1.24)>

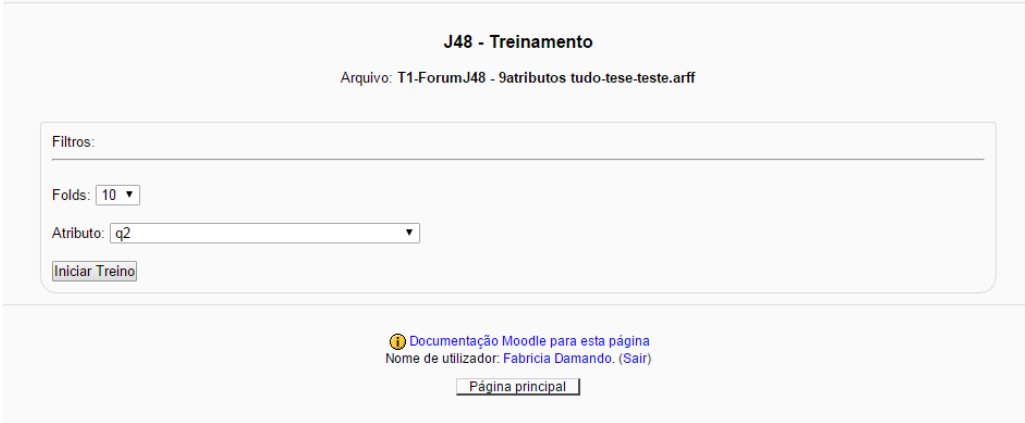
```

At the bottom, there's a footer with 'Documentação Moodle para esta página' and 'Nome de utilizador: Fabricia Damando. (Sair)'.

Fonte: Elaborado pela autora

Caso o algoritmo J48 tenha sido escolhido, automaticamente são habilitados dois links, um para realizar o método *Cross-Validation* e outro para o *Holdout*. No caso do *Cross-Validation*, o usuário é redirecionado para outra página onde é possível definir o atributo objetivo e a quantidade de “*folds*” necessários para a mineração (ver Figura 54).

Figura 54: Definição dos parâmetros para o algoritmo *Cross-Validation*



The screenshot shows a Moodle page titled "J48 - Treinamento". Below the title, it displays the file name "Arquivo: T1-Forum.J48 - 9atributos tudo-tese-teste.arff". There is a "Filtros:" section with a search bar. Below this, there are two dropdown menus: "Folds:" set to "10" and "Atributo:" set to "q2". A button labeled "Iniciar Treino" is positioned below the dropdowns. At the bottom of the page, there is a link for "Documentação Moodle para esta página" and the user information "Nome de utilizador: Fabricia Damando. (Sair)". A "Página principal" button is also present at the bottom.

Fonte: Elaborado pela autora

A Figura 55 apresenta o resultado da mineração *Cross-Validation* feita dentro do Moodle. O resultado obtido dessa mineração é o mesmo exibido na Figura 42, realizado somente na ferramenta Weka sem a integração com o Moodle. Dessa forma, é possível verificar que a implementação da integração foi realizada corretamente.

Figura 55: Resultado da mineração *Cross-Validation* 9 atributos

Resultado

File: T1-ForumJ48CrossValidation - 9atributos tudo-tese.arff

=== Run information ===

Relation: REAeForuns38AtributosTudoTESE-weka.filters.unsupervised.attribute.Remove-weka.filters.unsupervised.attribute.Remove-R1,3-weka.filters.unsupervised.attribute.Remove
Instances: 44
Attributes Numbers: 9
Attributes:
q2
Grupo_Envio_Atraso
Grupo_Sente_Dificuldade
Disciplina_Consegue_AcompanharConteudos
Disciplina_TeveDificuldadeConteudo
ConceitoFinal
FGrup_viewDiscussion_Quartil
FGrup_viewForum_Quartil
FGrup_addPost_Quartil2
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```

-----
FGrup_viewForum_Quartil = Baixo: Desanimado (12.0/1.0)
FGrup_viewForum_Quartil = BaixoMedio
|FGrup_addPost_Quartil2 = Baixo
||Disciplina_Consegue_AcompanharConteudos = Nunca: Desanimado (0.0)
||Disciplina_Consegue_AcompanharConteudos = Raramente: Desanimado (0.0)
||Disciplina_Consegue_AcompanharConteudos = AlgumasVezes: Insatisfeito (2.0)
||Disciplina_Consegue_AcompanharConteudos = QuaseSempre: Desanimado (4.0)
||Disciplina_Consegue_AcompanharConteudos = Sempre: Desanimado (0.0)
|FGrup_addPost_Quartil2 = BaixoMedio: Insatisfeito (1.0)
|FGrup_addPost_Quartil2 = Medio: Desanimado (3.0)
|FGrup_addPost_Quartil2 = Alto: N_respondeu (1.0)
FGrup_viewForum_Quartil = Medio: Insatisfeito (10.0/4.0)
FGrup_viewForum_Quartil = Alto: Desanimado (11.0/3.0)

```

Number of Leaves : 11

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	26	59.0909 %
Incorrectly Classified Instances	18	40.9091 %
Kappa statistic	0.021	
Mean absolute error	0.1411	
Root mean squared error	0.2634	
Relative absolute error	89.1028 %	
Root relative squared error	108.3878 %	
Total Number of Instances	44	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.314	Satisfeito
	0	0	0	0	0	?	Animado
	0.857	0.813	0.649	0.857	0.738	0.54	Desanimado
	0.154	0.181	0.288	0.154	0.2	0.445	Insatisfeito
	0	0	0	0	0	0.287	Nenhuma
	0	0	0	0	0	?	Outra
	0	0	0	0	0	0.337	N_respondeu
Weighted Avg.	0.591	0.585	0.467	0.591	0.529	0.468	

=== Confusion Matrix ===

a	b	c	d	e	f	g	<- classified as
0	0	1	0	0	0	0	a = Satisfeito
0	0	0	0	0	0	0	b = Animado
0	0	24	4	0	0	0	c = Desanimado
0	0	11	2	0	0	0	d = Insatisfeito
0	0	0	1	0	0	0	e = Nenhuma
0	0	0	0	0	0	0	f = Outra
0	0	1	0	0	0	0	g = N_respondeu

Fonte: Elaborado pela autora

Para o caso da opção pelo *Holdout*, o primeiro arquivo será usado para treinamento, enquanto em uma nova página (ver Figura 56) é possível selecionar um novo arquivo para teste, além do atributo objetivo.

Figura 56: Seleção do arquivo de teste para o método *Holdout*

J48 - Treinamento

Arquivo: T1-ForumJ48 - 9atributos tudo-tese-treinamento.arff

Filtros:

Arquivo de teste: Escolher arquivo Nenhum arquivo selecionado

(*Tamanho máximo 2Mb)

Atributo:

Fonte: Elaborado pela autora

Após a seleção dos parâmetros necessários, os resultados da mineração do método *Holdout*, tanto da mineração do treinamento quanto do teste, são exibidos.

A Figura 57 expõe o resultado da mineração do teste do método *Holdout* para o componente fórum, utilizando 9 atributos. O resultado alcançado é o mesmo apresentado na Figura 41, demonstrando que a integração apresentou o resultado desejável.

Figura 57: Mineração *Holdout* 9 atributos

Correctly Classified Instances	8	53.3333 %
Incorrectly Classified Instances	7	46.6667 %
Kappa statistic	0	
Mean absolute error	0.153	
Root mean squared error	0.2946	
Relative absolute error	82.5975 %	
Root relative squared error	102.1616 %	
Total Number of Instances	15	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0	?	Satisfeito
0	0	0	0	0	0	?	Animado
1	1	1	0.533	1	0.698	0.5	Desanimado
0	0	0	0	0	0	0.5	Insatisfeito
0	0	0	0	0	0	0.5	Nenhuma
0	0	0	0	0	0	?	Outra
0	0	0	0	0	0	?	N_respondeu
Weighted Avg.	0.533	0.533	0.284	0.533	0.371	0.5	

=== Confusion Matrix ===

a	b	c	d	e	f	g	<-- classified as
0	0	0	0	0	0	0	a = Satisfeito
0	0	0	0	0	0	0	b = Animado
0	0	8	0	0	0	0	c = Desanimado
0	0	0	6	0	0	0	d = Insatisfeito
0	0	1	0	0	0	0	e = Nenhuma
0	0	0	0	0	0	0	f = Outra
0	0	0	0	0	0	0	g = N_respondeu

Fonte: Elaborado pela autora

Portanto, observa-se que é possível realizar a mineração do Weka a partir do Moodle, o que pode vir a simplificar a tarefa do professor ao utilizar uma única ferramenta. Ressalta-se que, nos exemplos apresentados, o arquivo ARFF já havia sido preparado, porém, como o modelo já prevê todas as variáveis comportamentais necessárias, bem como toda a transformação de dados, gerar o arquivo ARFF automaticamente pelo Moodle é exequível e plausível de ser realizado.

Além da integração do Weka ao Moodle, as regras do Modelo de Predição foram implementadas via ferramenta de visualização, concebidas como *dashboards* e são descritas na seção 6.5.2.

6.5.2 Ferramenta de visualização

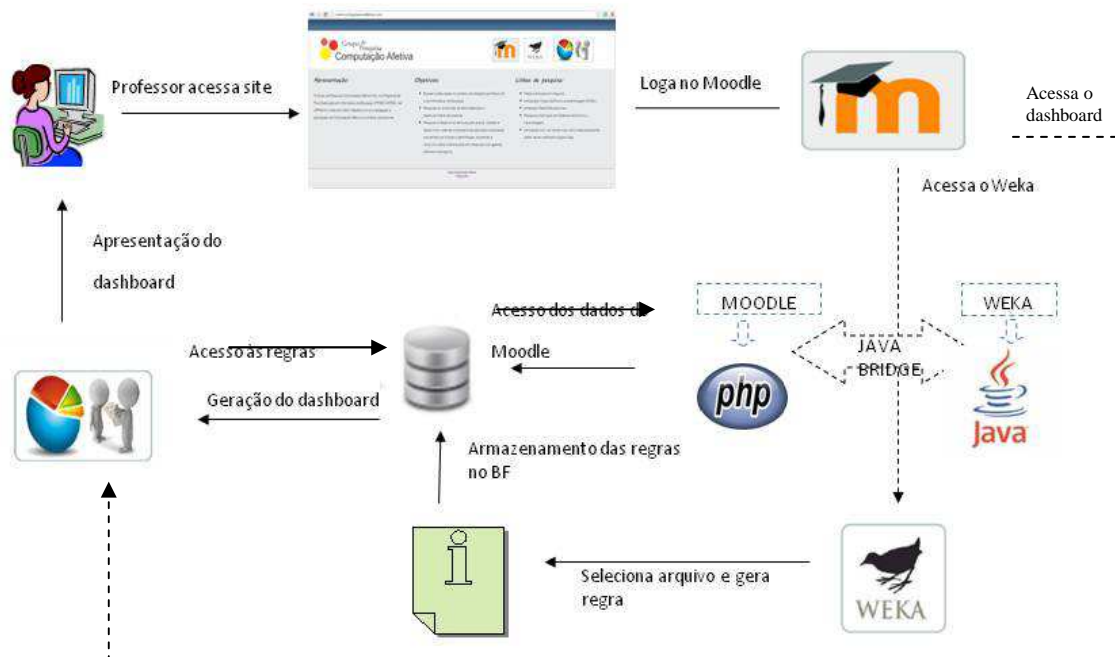
Uma apresentação visual do resultado foi desenvolvida e foi criada uma ferramenta de visualização das regras do Modelo de Predição. Essa ferramenta de visualização está expressa em forma de *dashboard*. Um *dashboard* é um painel de apresentações de resultados, usado em vários formatos, cuja apresentação permite visualização dos dados graficamente, propiciando uma visão geral e mais intuitiva sobre o conteúdo.

O Modelo de Predição foi implementado em forma de *dashboard* e integrado ao Moodle. O *dashboard* foi desenvolvido em PHP, HTML 5 (*Hypertext Markup Language*) e Java Script, utilizando as bibliotecas *JpGraph* e *RGraph* para geração de gráficos.

Como os dados para a geração dos *dashboards* são oriundos do Modelo de Predição, foram implementadas, inicialmente, duas regras de cada componente. Todas as regras estão armazenadas no banco de dados do Moodle a fim de poder gerar *dashboards*. As consultas foram realizadas diretamente no Moodle, o que permitiu a busca de todos os dados dos alunos os quais contemplavam determinada regra.

A Figura 58 apresenta o modelo geral de integração entre as ferramentas e apresentação do *dashboard* para o professor.

Figura 58: Modelo geral de integração Weka/Moodle e apresentação Dashboard

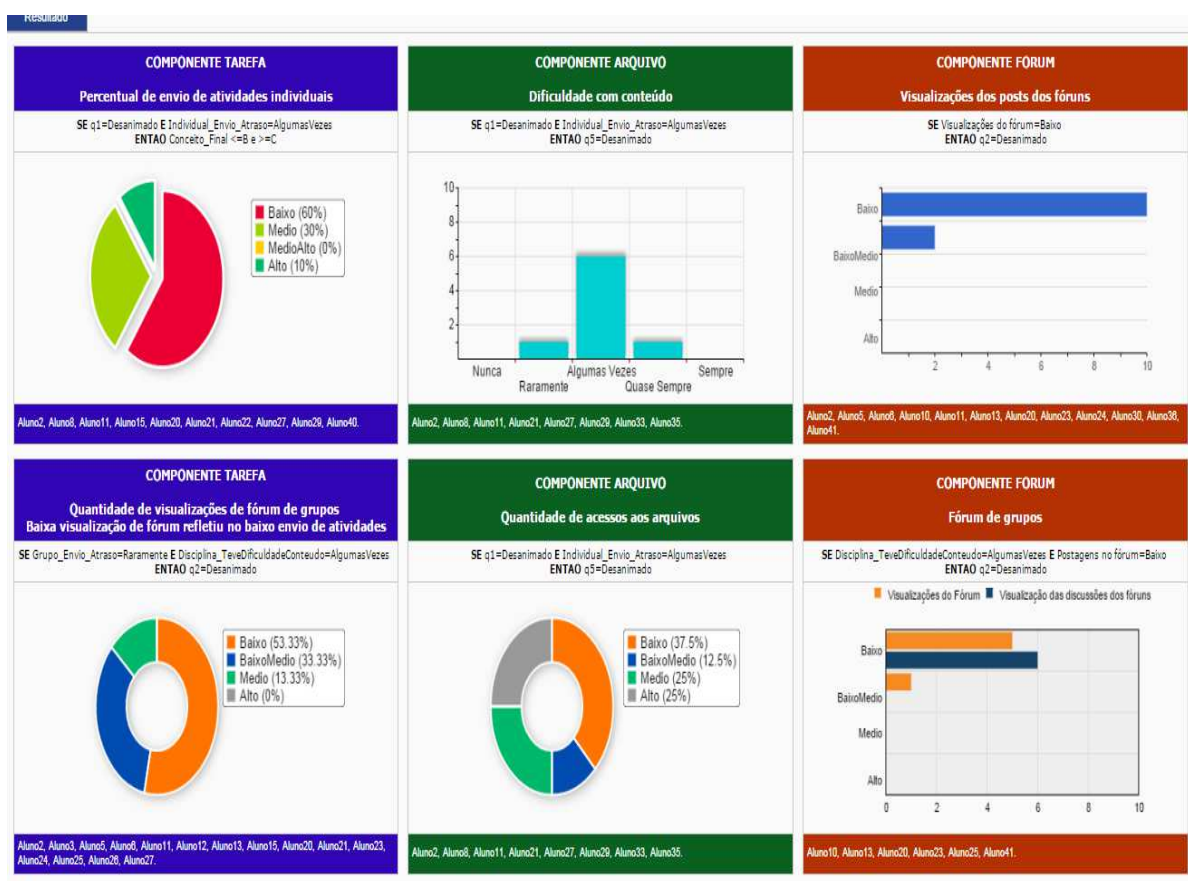


Fonte: Elaborado pela autora

Para o componente tarefa, foram implementadas as regras 1 e 3, para o componente fórum, foram implementadas as regras 1 e 4 e, por fim, as regras 1 e 3 do componente arquivo.

Todas as regras são apresentadas por meio da ferramenta de visualização em forma de *dashboards*, exibida na Figura 59, permitindo ter um painel com uma visão geral das regras, das variáveis comportamentais dos alunos no ambiente, bem como identificar alunos por regra. Os nomes dos alunos foram substituídos pela sequência aluno1, aluno2 e, assim, sucessivamente.

Figura 59: Ferramenta de visualização das regras selecionadas

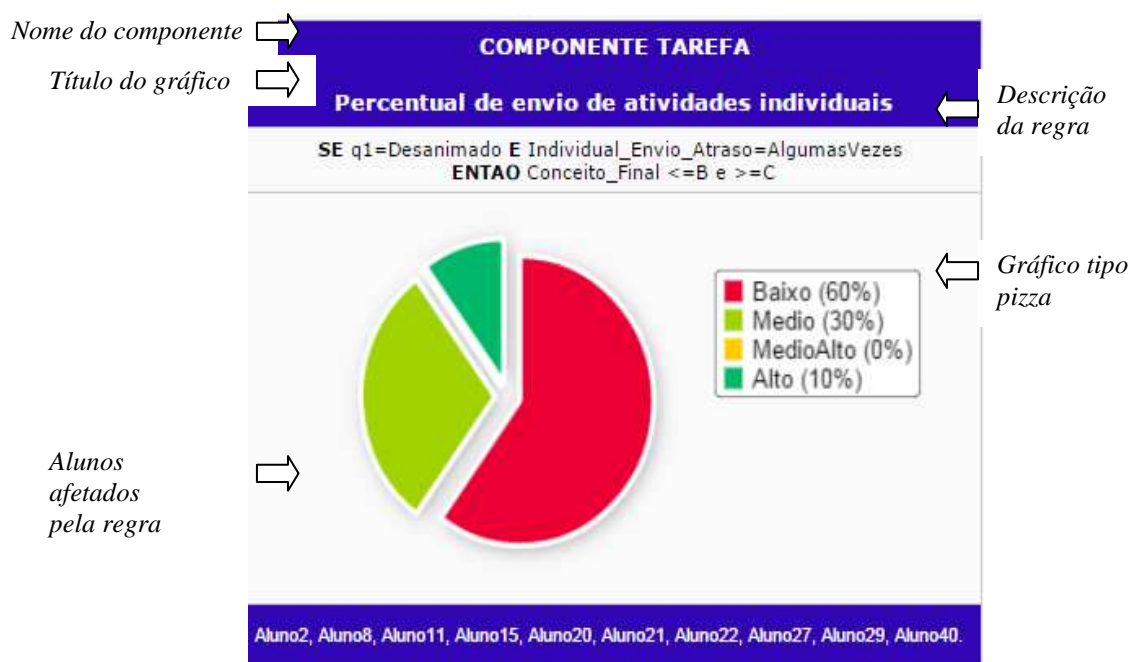


Fonte: Elaborado pela autora

Cada gráfico apresenta o nome do componente, um título para o gráfico e logo abaixo a descrição da regra. No centro, um gráfico é exibido e, abaixo dele, têm-se todos os alunos que foram identificados pela regra. Para utilizar o *dashboard*, não é preciso executar o Weka no Moodle todas as vezes. A partir do momento em que as regras já foram identificadas, o *dashboard* pode ser utilizado.

Uma análise individual do primeiro gráfico é exibida na Figura 60. Observa-se que apresenta o título “Percentual de envios de atividades individuais”, além da descrição da regra 1 e a identificação dos 10 alunos afetados pela regra. Da mesma forma, 60% destes alunos tiveram baixo envio de atividades, o que significa que enviaram menos de 7 atividades, do total de 11 atividades propostas na disciplina, indicando uma propensão ao desânimo quando o aluno envia atividades individuais na disciplina.

Figura 60: Regra 1 do componente tarefa



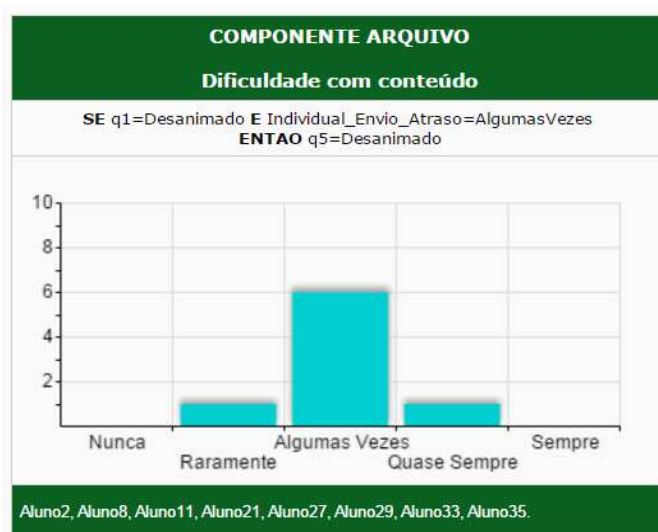
Fonte: Elaborado pela autora

O professor pode obter a regra pelo Weka, diretamente no Moodle, ou usar a ferramenta de visualização do Modelo de Predição.

Sugere-se que ao usar a ferramenta, o professor visualize os *dashboards* de modo a apoiá-lo na interpretação das regras. Também é possível ter uma visão geral das regras com o destaque dos alunos afetados em cada uma delas.

Analisando, em separado, um gráfico do componente arquivo (Figura 61), nota-se com relação à regra 1: “SE q1=Desanimado E Individual Envio Atraso = Algumas Vezes ENTÃO q5=Desanimado”, apresentou 8 alunos afetados pela regra. Além disso, foi possível analisar e retratar que estes alunos também relataram, em determinado momentos, dificuldade com o conteúdo refletindo no estado de ânimo desanimado.

Figura 61: Regra 1 componente arquivo



Fonte: Elaborado pela autora

Como a exibição do *dashboard* é feita com vários gráficos por tela, é possível ao professor comparar resultados de regras, analisar alunos que estão frequentemente em várias regras, apresentando várias situações de desânimo, constatando, por exemplo, que estes alunos precisam de maior atenção por parte do professor. Um exemplo observado diz respeito ao aluno 2, pois esse aluno apresenta situações de desânimo em 5 regras do total das 6 regras apresentadas, o que sugere que esse aluno precisa de uma atenção especial no curso.

Nota-se que a partir da identificação das variáveis que compõe o desânimo, e das regras que foram mineradas e validadas, foi possível desenvolver uma ferramenta dentro do próprio Moodle para apoiar o professor nessa inferência. A ferramenta visual propicia ao professor maior facilidade de interpretação das regras, fortalecendo a originalidade da tese.

6.6 CONSIDERAÇÕES SOBRE O CAPÍTULO

Esse capítulo se dedicou a realizar a validação da mineração dos componentes tarefa, fórum e arquivos em um novo conjunto de dados, na busca por descobrir um Modelo de Predição do aluno desanimado quando o aluno interagisse em um AVEA.

Foram validadas as melhores regras obtidas pela mineração descritas no capítulo 5. As regras que foram validadas e que puderam ser identificadas no novo conjunto de dados, via

consulta à base de dados e confirmação através de nova mineração, foram usadas para compor o modelo de Predição.

Destaca-se que através do Modelo de Predição proposto o objetivo específico “*Desenvolver um Modelo de Predição do estado de ânimo desanimado para o aluno em interação no AVEA*” pode ser atendido.

Porém, apresentar esse conjunto de regras ao professor exigiria do mesmo um esforço muito grande para identificar quais alunos seriam afetados por elas. Sendo assim, foi feita uma integração do Weka com o Moodle, somente das funcionalidades aplicadas para a mineração com as devidas variáveis comportamentais já definidas. Dessa forma, torna possível ao professor realizar a mineração dentro do próprio Moodle.

Além dessa integração, também foi implementada uma ferramenta de visualização (*dashboard*) com algumas regras.

A partir do *dashboard* representativo das regras do Modelo de Predição do desânimo, o professor tem uma visão macro sobre as regras e, ao mesmo tempo, ele pode obter informações específicas sobre cada uma, como a facilidade em identificar os alunos que “caíram” em determinada regra, além de poder comparar regras e identificar quais alunos apareceram com maior frequência.

Nesse sentido, sugere-se que o professor possa ter um instrumento de mineração de dados que o apoie na identificação do aluno propenso ao desânimo, tanto em forma de regras quanto em forma gráfica, para que o professor possa apoiar esse aluno de forma mais personalizada.

Verifica-se, nesse capítulo, que tanto a segunda quanto a terceira contribuições da tese puderam ser atendidas. A segunda contribuição “*Desenvolvimento de um modelo que identifique o comportamento de um aluno desanimado na modalidade EAD*”, foi contemplada através dos resultados obtidos pela mineração e a validação dos experimentos B e C.

A terceira contribuição “*Modelo de Predição para contribuir na identificação dos alunos desanimados em um AVEA a fim de apresentar, via dashboard, uma visão geral do comportamento do aluno desanimado, para que os professores possam atuar na melhor forma de apoio a esses alunos*”, foi contemplada, justificada e apresentada na seção 6.5.2., por meio da ferramenta de visualização.

7 CONCLUSÕES

Aspectos relacionados à afetividade influenciam a forma como os alunos aprendem, tanto positivamente quanto negativamente, tornando fonte de interesse e pesquisas na área da computação afetiva aplicada no contexto educacional. A influência que a afetividade exerce na aprendizagem foi citada em muitas das pesquisas descritas nessa tese.

Além de acompanhar o aprendizado do aluno sob o enfoque cognitivo, também é preciso identificar aspectos relacionados a sua afetividade, momentos onde o aluno se sente motivado, engajado, frustrado, desanimado, dentre outros, pois, sabe-se que os estados afetivos podem influenciar tanto positivamente quanto negativamente a aprendizagem (Scherer, 2005; Picard, 2000).

Geralmente, os AVEA oferecem ao professor ferramentas para acompanhar o aluno no ambiente, seja através de relatórios (frequência, controle de entregas de atividades, entre outros), ou por ferramentas de comunicação como os *chats*, fóruns, diário de bordo, entre outros, ou seja, ferramentas que apoiam o acompanhamento cognitivo do aluno.

Um aluno pode, por exemplo, sentir dificuldade com o conteúdo não assimilando o que é essencial para a realização de uma atividade e, em casos mais extremos, pode levar o aluno à desistência da disciplina por não conseguir acompanhar o conteúdo, por estar desanimado e sentir-se frustrado com essa situação.

Dessa maneira, esta pesquisa buscou reconhecer o aluno que esteja no estado de ânimo desanimado, pois esse estado é passível de mudança, pode ser alterado, não sendo estanque. Tornando momento oportuno para que o professor auxilie esse aluno.

Com esse intuito, a busca incessante durante toda essa pesquisa foi como identificar o aluno desanimado a partir do seu comportamento observável em um AVEA. Foi preciso identificar as variáveis observáveis que representassem o comportamento do aluno desanimado, que foi feito através da manifestação da emoção utilizando o instrumento questionário fundamentado em Scherer (2005), bem como análise dos *logs* dos alunos no Moodle utilizando MDE para apoio nessa inferência. A identificação das variáveis indicadoras do desânimo é considerada como uma das contribuições da tese.

A aplicação de técnicas de MDE teve dois objetivos, o primeiro e primordial, que foi obter regras que evidenciassem o comportamento do aluno desanimado, e o segundo objetivo, que foi de selecionar os atributos mais relevantes para a mineração, ou seja, reduzir a vasta quantidade de atributos possíveis para minerar.

Por meio da MDE, as variáveis comportamentais e as melhores regras puderam ser identificadas e selecionadas para compor o Modelo de Predição do aluno desanimado que esteja em interação em AVEA. Convém assinalar que os testes foram realizados com alunos em interação no AVEA Moodle, e que as variáveis comportamentais selecionadas estavam contidas nesse ambiente.

O procedimento metodológico foi conduzido por um estudo piloto e pelos experimentos A, B e C, permitindo utilizar mais de uma fonte de dados para coleta de dados e análise dos resultados. O estudo piloto auxiliou na aquisição do entendimento sobre questões relacionadas com a dimensão afetiva, na forma de desenvolvimento do instrumento de coleta de dados e, principalmente, apoiou a definição do escopo da pesquisa.

O experimento A conduziu o primeiro questionário desenvolvido a partir da REA e apoiou o mapeamento das variáveis observáveis. O experimento B, primordial para a pesquisa, conduziu efetivamente a mineração dos dados para a geração e seleção das variáveis comportamentais, e das melhores regras no contexto aplicado, além da implementação da REA 2.0 e tabulação dos dados dos *logs* do Moodle. O experimento C permitiu validar e testar regras sugeridas no experimento B, conduzindo a busca pelo Modelo de Predição. Uma ferramenta de visualização foi implementada a partir do Modelo de Predição, em forma de *dashboard*, visando facilitar a interpretação do modelo para apoiar o professor na melhor identificação e acompanhamento do aluno desanimado, o que contribuiu para a originalidade da tese.

De acordo com Costa *et al.* (2012), há diversas tarefas em MDE para analisar dados gerados nas interações dos estudantes no AVEA, além de uma necessidade significativa e urgente em prover ambientes computacionais apropriados para a MDE e oferecer ao professor facilidade de uso desses dados nesses ambientes.

Portanto, o uso da MDE juntamente com o desenvolvimento da ferramenta integrada ao Moodle, a qual propicia ao professor identificar e acompanhar o aluno com propensão ao desânimo, contribui para a melhoria do ensino e da aprendizagem do aluno. Além disso, com o *dashboard*, foi possível instrumentalizar o professor com uma ferramenta de acompanhamento do aluno, contribuindo com a comunidade acadêmica e científica.

O Modelo de Predição concebido possui variáveis comportamentais oriundas tanto dos *logs* do Moodle quanto de questões desenvolvidas utilizando a REA 2.0, a qual tem subsídio teórico nas concepções de Scherer (2005) e de Sacharin, Schlegel e Scherer (2012). A visualização do modelo desenvolvido e integrado ao Moodle sob a forma de *dashboard* teve como intuito apresentar uma comunhão da informática e afetividade aplicadas na educação,

além de propiciar uma melhor forma de apresentação e visualização dos alunos desanimados, vindo a fornecer ao professor um apoio computacional para essa identificação.

Dois projetos foram desenvolvidos nessa tese sob a orientação da profa. Dra. Magda Bercht e subsidiados pela SEAD/UFGRS e foram fundamentais para o desenvolvimento da pesquisa e divulgação de resultados alcançados.

- Edital UFRGS EAD 19/SEAD – Projeto Linha B - Pesquisas em educação à distância: Aplicação da mineração de dados educacionais em ambientes virtuais de ensino e aprendizagem: uma proposta para a descoberta do desânimo.
- Edital UFRGS EAD 21/SEAD - Projeto Linha B - Pesquisas em educação à distância: Descoberta do desânimo de alunos em ambientes virtuais de ensino e aprendizagem através do uso de mineração de dados educacionais.

Além dos projetos supracitados, artigos foram publicados em contexto nacional e internacional, repercutindo para a melhoria do ensino e aprendizagem.

Esta pesquisa recebeu destaque no Salão de Ensino da UFRGS no ano de 2015, na sessão de relato de experiência em atividades à distância, ao abordar a aplicação da mineração de dados educacionais em AVEA para a descoberta do desânimo.

Logo, a tese apresentou como destaque inovador um produto de pesquisa com utilidade na prática docente no ensino superior, principalmente em cursos EAD, para o reconhecimento de aspectos relacionados à afetividade no contexto educacional. Com o desenvolvimento de ferramentas que instrumentalizam professores para poderem melhorar o acompanhamento de alunos desanimados em interação em AVEA. Dessa forma, permite ao professor fomentar uma metodologia de acompanhamento dos alunos, a fim de minimizar futuras evasões, bem como desistências em cursos e disciplinas, beneficiando a comunidade acadêmica.

7.1 LIMITAÇÕES E TRABALHOS FUTUROS

Esta tese limitou-se, em seu escopo, a identificar o estado de ânimo desanimado do aluno, utilizando variáveis comportamentais observáveis através de suas interações no AVEA Moodle e, apresentou um Modelo de Predição desse aluno. Não contemplou os demais estados de ânimo definidos por *Scherer* e também não contemplou observações na prática

pedagógica do professor no acompanhamento do aluno desanimado. As limitações do escopo e da amostra da pesquisa abrem espaço para novas investigações, com vistas a propiciar um modelo mais completo possível.

A manifestação do estado de ânimo está associada às questões de autorrelato que utilizem a REA 2.0, por mais que se tenha desejado somente realizar essa etapa inicialmente e, posteriormente, que ela não fosse mais necessária. Percebeu-se, entretanto, que as questões da REA 2.0 são fundamentais na composição do modelo.

Ressalta-se que o modelo sugerido foi minerado e validado para uma amostra não probabilística, o que torna a amostra não generalizável, apesar dos resultados serem interessantes. Como o modelo revelou evidências para alunos desanimados, é preciso testá-lo em um novo conjunto de dados, uma nova amostra de preferência probabilística, na busca por um modelo generalizável.

Alguns detalhes a serem implementados futuramente podem aperfeiçoar o modelo e apoiar o professor na identificação e acompanhamento do aluno que esteja desanimado, como:

- a) Gerar automaticamente o arquivo ARFF para a mineração do Weka dentro do Moodle. Como as variáveis comportamentais necessárias para cada componente (tarefa, fórum e arquivo) já foram mapeadas, e estão presentes no ambiente, é preciso somente implementar o arquivo necessário para a mineração.
- b) Apresentar as demais regras em formato visual, por meio de *dashboards*. Como foram implementadas somente duas regras de cada componente, implementar as demais pode vir a subsidiar melhor o professor na identificação do aluno desanimado ou em processo de desânimo.
- c) Permitir ao professor selecionar qualquer um dos gráficos gerados e apresentados no *dashboard* e visualizá-lo mais detalhadamente. Pode ser apresentado quais alunos contemplam cada área do gráfico apresentado.
- d) Implementar computacionalmente somente as questões referentes à REA 2.0 e variáveis do questionário que estão presentes no modelo e são importantes para a manifestação do estado de ânimo. Dessa forma, o aluno responderá a menos questões, o que pode gerar uma amostra mais participativa.
- e) Permitir ao professor que dispare, quando necessário, o questionário implementado na opção “d”, e que faça um acompanhamento, ao longo da disciplina, dos alunos desanimados.

Sob outra perspectiva, além das possibilidades de melhorias na implementação, sugere-se que seja realizado uma pesquisa sob a ótica pedagógica, da aplicação do modelo proposto, com a finalidade em analisar melhor forma de apresentar subsídios para o professor.

Há possibilidade de usar a ferramenta além da sala de aula, podendo ser usada como ferramenta de apoio ao acompanhamento do aluno por pró-reitorias de graduação e comissões de graduação, servindo de apoio para a instituição na identificação do aluno desanimado e sua correlação com a evasão, por exemplo.

Nessa pesquisa, desenvolveu-se *dashboards* para apresentação visual do resultado do Modelo Proposto, também se torna necessário analisar qual é a melhor forma de apresentação desse modelo ao professor. Fundamentos de usabilidade e *user experience* podem ser aplicados.

Novas perspectivas com relação às técnicas computacionais podem ser agregadas através da mineração de séries temporais, com propósito da identificação e acompanhamento do aluno desanimado ao longo do tempo, em uma disciplina.

Sugere-se incorporar o modelo a outros AVEA, visto que a presente pesquisa utilizou o ambiente Moodle. Para que isso ocorra, talvez seja necessário identificar as variáveis correspondentes do novo ambiente com as variáveis comportamentais usadas diretamente do *log* do Moodle.

Por fim, uma abordagem mais desafiadora em consolidar, nacionalmente, uma cultura de análise de dados com o desenvolvimento de uma base de dados pública, inclusive que possa ser aplicada em cursos online abertos massivos (MOOC – *Massive Open Online Course*), onde tanto pesquisadores quanto estudantes, professores e instituições possam se beneficiar das aplicações de MDE.

REFERÊNCIAS

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. *Mining Association Rules Between Sets of Items in Large Databases*. In: ACM Sigmod International Conference on Management of Data, SIGMOD, p. 207-216, 1993.
- AMERSHI, S.; CONATI, C. *Combining unsupervised and supervised classification to build user models for exploratory learning environments*. Journal of Educational Data Mining, Vol. 1, Nº 1, 2009.
- ASSUNÇÃO, A. S. **Descoberta Direta e Eficiente de Regras de Associação Ótimas**. Dissertação de Mestrado. Universidade de São Paulo. Programa de Pós-Graduação em Ciências da Computação e Matemática Computacional, 2011.
- BAKER, R. S. J.; YACEF, K. *The State of Educational Data Mining in 2009: A Review and Future Visions*. JEDM – Journal of Education Data Mining . V.01, N.01, 2009.
- BAKER, R.S.J.d. *Data Mining for Education*. In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier, 2010.
- BAKER, R. S. J.; ISOTANI, S.; CARVALHO, A. M. J. B. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação. V.19, N.02, 2011.
- BAKER, R. S. J. *Educational Data Mining: An Advance for Intelligent Systems in Education*. *IEEE Intelligent Systems*, 29 (3), 78-82, 2014.
- BARADWAJ, B. K.; PAL, S. *Mining Education Data to Analyze Students Performance*. *International Journal of Advanced Computer Science and Applications*, Vol. 2, Nº. 6, 2011.
- BASTOS, H. P. P. **Presença Plus: modelo de identificação de presença social em ambientes virtuais de aprendizagem**. Tese de doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação, 2012.
- BEHAR, P. A. *et al.* **Competências: Conceito, elementos e recursos de suporte, mobilização e evolução**. Competências em Educação à distância. Porto Alegre: Penso, 2013.

BERCHT, M. Em **Direção a Agentes Pedagógicos com Dimensões Afetivas**. Tese de doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em computação, 2001.

BOWER, G. H.; FORGAS, J. P. *Mood and Social Memory. Handbook of Affect and Social Cognition. Psychology Press*, PP (95-120), Philadelphia, 2000.

BRAND, S.; REIMER, T.; OPWIS, K. *How to do learning in a negative mood? Effects of a negative mood on transfer and learning. Learning and Instruction* 17, 2007, pp. 1-16.

CAICEDO, D. G.; BEUZEKOM, M. V. *How do you feel? An assessment of existing tools for the measurement of emotions and their application in consumer product research*. Delf University of Technology, 2006.

CAMILO, C. O.; SILVA, J. C. **Mineração de Dados: conceitos, tarefas, métodos e ferramentas**. Relatório Técnico. Instituto de Informática – UFG, 2009.

CHEN, G.; LIU, C.; OU, K.; LIU, B. B. *Discovering Decision Knowledge from Web Log Portfolio for Managing Classroom Processes by Applying Decision Tree and Data Cube Technology*. Journal of Educational Computing Research, 23(3), pp.305–332, 2000.

COCEA, M.; WEBLZAHN, S. *Can Log Files Analysis Estimate Learners' Level of Motivation? Workshop on Adaptivity and User Modeling in Interactive Systems*. Hildesheim. pp.32-35, 2006.

COSTA, E., BAKER, R., AMORIN, L. MAGALHÃES, J., MARINHO, T. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. Jornada de Atualização em Informática na Educação – JAIE, 2012.

DAMÁSIO, A. **O erro de Descartes: emoção, razão e o cérebro humano**. São Paulo: Companhia das Letras, 1996.

_____. **O mistério da consciência: do corpo e das emoções ao conhecimento de si**. São Paulo: Companhia das Letras, 2000.

_____. **Em busca de Espinosa: prazer e dor na ciência dos sentimentos**. São Paulo: Companhia das Letras, 2004.

_____. **E o cérebro criou o homem.** São Paulo: Companhia das Letras, 2011.

DAMASCENO, M. **Introdução a mineração de dados usando Weka.** V Congresso Norte-Nordeste de Pesquisa e Inovação, 2010.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *Advances in knowledge discovery and data mining.* MIT PRESS. 1996.

FELDMAN, R.; SANGER, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* Cambridge University Press, 2007

FERREIRA, M. **I Fórum Nacional das Áreas do Sistema UAB - CAPES,** 2013.

FIGUEIREDO, M. F. H. **Locus de Controlo: Variável influente ou influenciada pelo desempenho em matemática? Estudo de caso com alunos do 3º ciclo.** Dissertação de mestrado. Universidade Aberta de Portugal, 2011.

FLICK, U. **Introdução à Metodologia de Pesquisa: um guia para iniciantes.** Porto Alegre: Penso, 2013.

GAUDIOSO. E.; TALAVERA, L. **Data mining to support tutoring in virtual learning communities: experiences and challenges.** In Data mining in e-learning. WIT Press, 2006.

GIL, A. C. **Como Elaborar Projetos de Pesquisa.** 4ª Edição. São Paulo: Atlas, 2002.

GONÇALVES, E. C. **Regras de Associação e suas medidas de interesse objetivas e subjetivas.** Infocomp Journal of Computer. V.04, N.01, 2005.

GOTTARDO, E.; NORONHA, R. V.; KAESTNER, C. A. A. **Estimativa de desempenho acadêmico de estudantes: análise da aplicação de técnicas de mineração de dados em cursos à distância.** Revista Brasileira de Informática na Educação. V.22, N. 01, 2014.

IEPSEN, E. F. **Ensino de Algoritmos: Detecção do Estado Afetivo de Frustração para Apoio no Processo de Aprendizagem.** Tese de Doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação, 2013.

- INVENTADO, P. S.; LEGASPI, R.; CABREDO, R.; NUMAO, MASAYUKI. *Modelling Affect in Student – driven Learning Scenarios*. International Educational Data Mining Society, 2013.
- JAQUES, P. A. **Um agente pedagógico animado para interagir afetivamente com o aluno**. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação, 2008.
- JAQUES, P. A.; NUNES, M. A. S. N. **Ambientes inteligentes de aprendizagem que inferem, expressam e possuem emoções e personalidade**. Jornada de Atualização em Informática na Educação. 23º Congresso Brasileiro de Informática na Educação , 2012.
- JBEILI, C. **Superando o desânimo antes que ele supere você**. São Paulo: Editora Nobel, 2004.
- KAMPFF, A. J. C. **Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente**. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação, 2009.
- KAPOOR, A.; BURLESON, W.; PICARD, R. W. *Automatic Prediction of Frustration*. International Journal of Human-Computer Studies, 2007, pp. 724-736.
- KLEMMANN, M.L; REATEGUI, E.; LORENZATTI, A. **O emprego da ferramenta SOBEK como apoio a produção textual**. XX Simpósio Brasileiro de Informática na Educação, 2009.
- KNIME. **Documentation**, 2014. Disponível em: <<http://tech.knime.org/documentation>>. Acessado em Maio, 2014.
- KOEDINGER, K.R.; BAKER, R.S.J.; CUNNINGHAM, K.; SKOGSHOLM, A.; LEBER, B.; STAMPER, J. A *Data Repository for the EDM community: The PSLC DataShop*. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press, 2011.
- KORT, B.; REILLY, R.; PICARD, R. W. *An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy – Bulding a Learning Companion*. IEEE International Conference on Advanced Learning Technologies - ICALT, 2001.

- LAGO, N. A. **Me, Myself and You: Autoestima e Aprendizagem de Línguas**. Afetividade e emoções no ensino/aprendizagem de línguas: múltiplos olhares. Campinas: Pontes Editores, 2011.
- LEONARDI, J.; CAPATAN, A. H.; BASTOS, R. C. **Análise do perfil dos estudantes de um curso Lato Sensu utilizando ferramenta RapidMiner**. Revista Eletrônica de Biblioteconomia e Ciência da Informação. V. 18, N. 37, p. 19-42, 2013.
- LIBRELOTTO, S. R.; MOZZAQUATRO, P. M. **Análise dos algoritmos de mineração J48 e Apriori aplicados na detecção de indicadores da qualidade de vida e saúde**. Revista Interdisciplinar de Ensino, Pesquisa e Extensão. V. 01, N. 01, 2013.
- LIÑÁN, L. C.; PÉREZ, A.A.J. *Educational Data Mining and Learning Analytics: differences, similarities, and time evolution*. Revista de Universidad y Sociedad del Conocimiento – RUSC. Universities and Knowledge Society Journal. Vol. 12, N.03, 2015.
- LONGHI, M. T. **Mapeamento de Aspectos Afetivos em um Ambiente Virtual de Aprendizagem**. Tese de doutorado. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação, 2011.
- LONGHI, M.; BEHAR, P. A.; BERCHT, M. **A Busca pela Dimensão Afetiva em Ambientes Virtuais de Aprendizagem**. Modelos Pedagógicos em Educação à distância. Porto Alegre: Artmed, 2009.
- MAJADAS, S. S.; SANTOS, O. C.; BOTICARIO, J. G.; CABESTRENO, R.; QUIRÓS, P.; SANIERO, M. *Gathering Emotional Data from Multiple Sources*. VI International Conference on Educational Data Mining, 2013.
- MAJADAS, S. S.; SANTOS, O. C.; BOTICARIO, J. G. *Affective State Detection in Educational Systems through Mining Mutimodal Data Sources*. International Educational Data Mining Society, 2013.
- MASTRELLA-DE-ANDRADE, M. R. **Falar, fazer, sentir, vir a ser: ansiedade e identidade no processo de aprendizagem de LE**. Afetividade e emoções no ensino/aprendizagem de línguas: múltiplos olhares. Campinas: Pontes Editores, 2011.

- MASTRELLA-DE-ANDRADE, M. R.; NORTON, B. **Querer é poder? Motivação, Identidade e Aprendizagem de Língua Estrangeira.** Afetividade e emoções no ensino/aprendizagem de línguas: múltiplos olhares. Campinas: Pontes Editores, 2011.
- MOORE, M. G.; KEARSLEY, G. **Educação a distância: uma visão integrada.** São Paulo: Cengage Learning, 2008.
- MORENO, J.; OVALLE, D.A.; VICARI, R. M. *A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics.* Journal Computers & Education. Vol.58, Issue 1, 2012, pp-560-569.
- NETTO, O. P. **Um filtro interativo de utilizando árvores de decisão.** Dissertação de Mestrado. Programa Interunidades de Bioinformática. Universidade de São Paulo, 2013.
- NOMELINI, J.; REZENDE, S. O.; YAMAMOTO, C. H.; BEZERRA, L. A. F. VOZZI, P. A. LÔBO, R. B. **Emprego de Regras de Associação para extração de padrões mercadológicos de touros Nelore com avaliação genética.** Revista Brasileira de Zootecnia. Vol.39 No.12, 2010.
- OCUMPAUGH, J.; BAKER, R. S.; RODRIGO, M. M. T.; SALVI, A.; VAN VELSEN, M.; AGHABABYAN, A.; MARTIN, T. HART: *The Human Affect Recording Tool.* ACM Special Interest Group on the Design of Communication (SIGDOC). New York, NY: ACM, 2015.
- OCUMPAUGH, J.; BAKER, R. S. J.; RODRIGO, M. T. *Baker Rodrigo Ocumpaugh Monitorin Protocol (BROMP) 2.0 Technical and Training Manual. Technical Report.* New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences, 2015.
- OCUMPAUGH, J.; BAKER, R. S. J.; RODRIGO, M. T. *Observation Method Protocol (BROMP) 1.0. Training Manual Version 1.0. Technical Report.* New York, NY: EdLab. Manila, Philippines: Ateneo Laboratory for the Learning Sciences, 2012.
- PIAGET, J. **Relações Entre a Afetividade e a Inteligência no Desenvolvimento Mental da Criança.** Organização e Tradução: Saltini e Cavenaghi. Rio de Janeiro. Wak Editora, 2014.
- PICARD, R. W. *Affective Computing.* The MIT Press, 2000.

- _____. *Affective Computing: from Laughter to IEEE*. *IEEE Transaction Education on Affective Computing*, Vol. 1, N.º 1, pp. -11-17. January – June, 2010.
- PICARD, R. W., DAILY, S. B. *Evaluating affective interactions: Alternatives to asking what users feel*. MIT Media Laboratory: Cambridge, MA. USA, 2008
- PONTAROLO, E. **Modelagem Probabilística de Aspectos Afetivos do Aluno em um Jogo Educacional Colaborativo**. Tese de doutorado. PPGIE. UFRGS, 2008.
- REATEGUI, E. B.; CAZELLA, S. C. **Sistemas de Recomendação**. XXV Congresso da Sociedade Brasileira de Computação. Unisinos, 2005.
- RIGO, S. J.; CAMBRUZZI, W.; BARBOSA, J. L. V.; 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*, V.22, N. 1, 2014.
- ROMERO, C. M.; VENTURA, S. **Mineração de Dados em E-Learning**. WIT Press Publishing, 2006.
- ROMERO, C. M.; VENTURA, S.; GARCÍA, E. *Data Mining in Course Management Systems: Moodle Case Study and Tutorial*. *Computers & Education* 51, pg. 368-384, Elsevier, 2008.
- SACHARIN, V.; SCHLEGEL, K.; SCHERER, K. R. *Geneva Emotion Wheel Rating Study*. Geneva, Switzerland: University of Geneva, Swiss Center for Affective Sciences, 2012.
- SABOURIN, J. L.; LESTER, J. C. *Affect and Engagement in Game-Based Learning Environments*. *IEEE Transactions on Affective Computing* . Vol 05, Nº 1, pp 45-56,doi: 10.1109/T-AFFC.2013.27, Março, 2014.
- SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, M. P. B. **Metodologia de Pesquisa**. 5ª Edição. Porto Alegre: Editora Penso, 2013.
- SCHERER, K. R. *Psychological Models of Emotion. The Neuropsychology of emotion*. Oxford University Press, 2000.
- SCHERER, K. R. *What are the emotions? And how can they measured?* In: *Social Science Information*, 2005.

- SELIGMAN, M. E. P.; CSIKSZENTMIHALYI, M. *Positive Psychology: An Introduction. American Psychologist*. Vol 55, pg. 5-14. Jan, 2000.
- SHUTE, V. J.; D'MELLO, S.; BAKER, R.; CHO, K.; BOSCH, N.; OCUMPAUGH, J.; VENTURA, M.; ALMEDA, V. *Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. Computers & Education* 86, 224-235, 2015.
- SHUTE, V. J.; VENTURA, M. *Measuring and supporting learning in games: Stealph assessment*. Cambridge, MA: MIT Press, 2013.
- SILIPO, R.; MAZANETZ, M. P. *The Knime cookbook*, 2012.
- SILVA, C. G.; FIGUEREDO, V. F. **Ambiente Virtual de Aprendizagem: comunicação, interação e afetividade na EAD**. Revista Aprendizagem em EAD. V01, 2012.
- SILVA, E. L.; MENEZES, E. M. **Metodologia de Pesquisa e Elaboração de Dissertação**. 4ª Edição. Universidade Federal de Santa Catarina, 2005.
- SILVA, M. L.; CRUZ, V. A.; SILVA, F. F. **A Dimensão Afetiva e sua Relevância no Processo de Ensino Aprendizagem: Uma abordagem Sociocognitiva**. REGET - Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental. V. 18, N. 3, 2014.
- SOUZA, S. **Confio em ambientes virtuais de aprendizagem**. Laboratório de Educação Digital, Universidade Jean Piaget de Cabo Verde, 2006.
- TAN, P-N.; STEINBACH, M.; KUMAR, V. **Introdução ao DataMining Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009.
- WEBBER, C. G.; ZAT, D.; LIMA, M. F. W. P. **Utilização de algoritmos de agrupamento na mineração de dados educacionais**. Revista Renote: Novas tecnologias na educação. V.11, Nº 1, 2013.
- WEKA. *Machine Learning Group at the University of Waikato*. The University of Waikato. Disponível em:<<http://www.cs.waikato.ac.nz/ml/Weka/documentation.html>>. Acessado em Maio de 2014.

WILGES, B.; MATEUS, G. P.; NASSAR, S. M.; BASTOS, R. C. **Avaliação da Aprendizagem por meio de lógica fuzzy validado por árvore de decisão ID3**. Renote – Novas Tecnologias na Educação. V. 08, N. 3, 2010.

WITTEN, I.; FRANK, E. *Data minig: practical machine learning tools and techniques*. 2ª Edição, Elsevier, 2005.

WISWEDEL, B.; KOTTER, T.; SILIPO, R. *Analysing the Web from Start to Finish: knowledge extraction from a web forum using Knime*. 2013. Disponível em:< http://www.knime.org/files/knime_web_knowledge_extraction.pdf> . Acessado em Maio de 2014.

YIN, R. K. **Estudo de Caso: planejamento e métodos**. 4ª Edição. Porto Alegre: Bookman, 2010.

APÊNDICE A – TERMO DE CONSENTIMENTO



Prezado(a) participante:

Meu nome é *Fabrcia Damando Santos* e estou realizando Doutorado no Programa de Pós-graduação em Informática na Educação (PPGIE), na Universidade Federal do Rio Grande do sul (UFRGS) sob Orientação da Professora Dr^a. Magda Bercht, atuando na linha de pesquisa “*Paradigmas para a Pesquisa sobre o Ensino Científico e Tecnológico*”, juntamente com o Grupo de Pesquisa em Computação Afetiva Aplicada a Educação (GCAE).

A presente pesquisa também faz parte de um projeto aprovado pela SEAD, onde se busca investigar como reconhecer o estado de ânimo dos alunos em ambientes virtuais de ensino e aprendizagem (AVEA), pois pode afetar a forma como desenvolverás teus estudos, tornando importante o seu estudo e formas de identificação.

Nesta etapa da pesquisa, precisamos coletar dados das suas interações no Moodle para a disciplina ECO03010 – Análise de demonstrações contábeis e, aplicaremos um questionário para apoio na identificação desta dimensão afetiva no AVEA Moodle. O questionário, apoiado com um mapa chamado de GEW (*Geneva Emotion Wheel*) e liderado por Klaus Scherer, será aplicado com o intuito em obter um autorrelato dos afetos sentidos por você em determinadas situações de aprendizagem.

A participação nesse trabalho é voluntária e em nenhum momento seu nome será citado nos resultados apresentados.

Gostaríamos de solicitar o seu auxílio na pesquisa respondendo ao questionário que segue.

Agradecemos a colaboração.

Profa. Dra. Magda Bercht – Coordenadora GCAE

Profa. MSc. Fabrcia Damando Santos – Pesquisadora

Eu _____, número do cartão UFRGS: _____ concordo em participar da pesquisa.

Assinatura do participante

Local e data

APÊNDICE B – QUESTIONÁRIO

Instruções

A GEW (*Geneva Emotion Wheel*) apresentada abaixo tem a finalidade de inferir a disposição afetiva sentida por você em situações de aprendizagem. Para cada pergunta, identifique um termo mais próximo possível do que você sentiu e a sua intensidade. Observe que há vários círculos, variando do menor para maior, onde o menor representa a menor intensidade e quanto maior, mais intenso o que você sentiu.

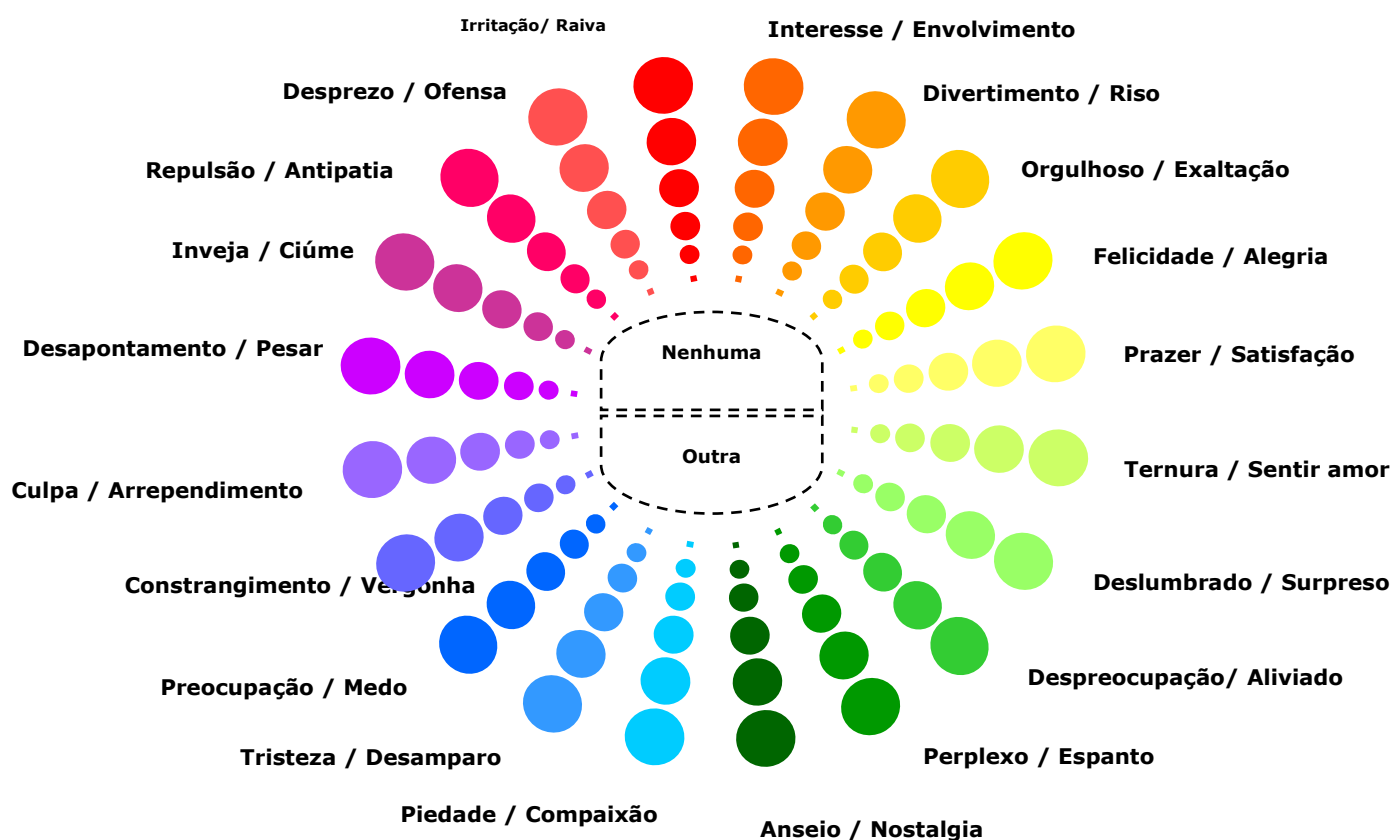


Se o que você sentiu não é apresentado na lista – marque opção: *Outra*, e escreva uma palavra que represente o que você sentiu. Caso não tenha sentido nada, marque no centro da roda a opção: *Nenhuma*.

QUESTIONÁRIO

Para cada questão abaixo, escolha apenas um estado e sua intensidade. Marque dentro do círculo o número da questão correspondente.

- 1) Como você se sente quando envia atividades individuais com atraso?
- 2) Como você se sente quando envia atividades em grupo com atraso?
- 3) Como você se sente quando tem dificuldades em realizar atividades individuais?
- 4) Como você se sente quando realiza atividades em grupo?
- 5) Como você se sente quando tem dificuldade em algum conteúdo?
- 6) Como você se sente com relação ao seu aprendizado?
- 7) Como você se sente ao solicitar ajuda do professor?



8) O que leva você a ficar desanimado com conteúdos ou atividades na disciplina?

9) Você já teve dificuldade em usar alguma ferramenta do Moodle?

() Não () Sim. Qual? _____

10) Quais os tipos de apoios você sugere que o professor faça quando você estiver com dificuldades em algum conteúdo ministrado?

- () Nova explicação () Monitoria
 () Sugestão de novo material para leitura () Outro. Qual? _____
 () Atendimento fora do horário de aula

11) Como você solicita ajuda ao professor, quando necessário?

() Email () Moodle () Presencialmente () Não solicita ajuda

12) Enumere em ordem de preferência quais ferramentas você e seu grupo mais usam para trocar informações sobre os trabalhos e atividades:

- | | |
|--|---|
| <input type="checkbox"/> Fórum do Moodle | <input type="checkbox"/> Skype |
| <input type="checkbox"/> Chat do Moodle | <input type="checkbox"/> WhatsApp |
| <input type="checkbox"/> Email | <input type="checkbox"/> Facebook |
| | <input type="checkbox"/> Outra. Qual? _____ |

13) Com relação às atividades individuais propostas pelo Moodle:

Você as envia com atraso? Nunca Raramente Algumas Vezes Quase Sempre Sempre

Você sente dificuldades em realizá-las? Nunca Raramente Algumas Vezes Quase Sempre Sempre

Você gosta de realizar atividades individualmente? Nunca Raramente Algumas Vezes Quase Sempre Sempre

14) Com relação às atividades em grupo propostas pelo Moodle:

Você as envia com atraso? Nunca Raramente Algumas Vezes Quase Sempre Sempre

Você sente dificuldades em realizá-las? Nunca Raramente Algumas Vezes Quase Sempre Sempre

Você gosta de realizar atividades em grupo? Nunca Raramente Algumas Vezes Quase Sempre Sempre

15) Com relação ao conteúdo da disciplina:

Você consegue acompanhar todos os conteúdos? Nunca Raramente Algumas Vezes Quase Sempre Sempre

Já teve dificuldade em algum conteúdo? Nunca Raramente Algumas Vezes Quase Sempre Sempre

Gratas pela contribuição,
 Fabrícia Damando Santos e Magda Bercht.

APÊNDICE C – MINERAÇÃO COMPONENTE TAREFA – ETAPA 1

```

Instances: 44
Attributes: 6
    q1
    Individual_Envio_Atraso
    Individual_Sente_Dificuldade
    Disciplina_TeveDificuldadeConteudo
    ConceitoFinal
    QuartilEnvioSIM
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.2 (9 instances)
Minimum metric <conviction>: 1.1
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 22

Size of set of large itemsets L(3): 10

Best rules found:

1. Individual_Envio_Atraso=AlgumasVeze QuartilEnvioSIM=Baixo 10 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 10   conf:(1) lift:(1.33) lev:(0.06) [2] < conv:(2.5)>
2. q1=Desanimado ConceitoFinal=B 12 ==> Individual_Envio_Atraso=AlgumasVeze 10   conf:(0.83) lift:(1.47) lev:(0.07) [3] < conv:(1.73)>
3. Disciplina_TeveDificuldadeConteudo=AlgumasVeze QuartilEnvioSIM=Baixo 12 ==> Individual_Envio_Atraso=AlgumasVeze 10   conf:(0.83) lift:(1.47) lev:(0.07) [3] < conv:(1.73)>
4. QuartilEnvioSIM=Baixo 13 ==> Individual_Envio_Atraso=AlgumasVeze Disciplina_TeveDificuldadeConteudo=AlgumasVeze 10   conf:(0.77) lift:(1.61) lev:(0.09) [3] < conv:(1.7)>
5. QuartilEnvioSIM=Baixo 13 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 12   conf:(0.92) lift:(1.23) lev:(0.05) [2] < conv:(1.63)>
6. q1=Insatisfeito 18 ==> Individual_Sente_Dificuldade=AlgumasVeze 14   conf:(0.78) lift:(1.37) lev:(0.09) [3] < conv:(1.55)>
7. QuartilEnvioSIM=MedioAlto 13 ==> ConceitoFinal=B 10   conf:(0.77) lift:(1.47) lev:(0.07) [3] < conv:(1.55)>
8. QuartilEnvioSIM=Baixo 13 ==> Individual_Envio_Atraso=AlgumasVeze 10   conf:(0.77) lift:(1.35) lev:(0.06) [2] < conv:(1.4)>
9. q1=Desanimado Individual_Envio_Atraso=AlgumasVeze 14 ==> ConceitoFinal=B 10   conf:(0.71) lift:(1.37) lev:(0.06) [2] < conv:(1.34)>
10. Individual_Envio_Atraso=AlgumasVeze ConceitoFinal=B 14 ==> q1=Desanimado 10   conf:(0.71) lift:(1.37) lev:(0.06) [2] < conv:(1.34)>

```

APÊNDICE D – MINERAÇÃO COMPONENTE TAREFA – ETAPA 2

```

Instances: 44
Attributes: 6
    q2
    Grupo_Envio_Atraso
    Grupo_Sente_Dificuldade
    Disciplina_TeveDificuldadeConteudo
    ConceitoFinal
    QuartilEnvioSIM
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.2 (9 instances)
Minimum metric <conviction>: 1.1
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 22

Size of set of large itemsets L(3): 6

Best rules found:

1. q2=Desanimado ConceitoFinal=B 12 ==> Grupo_Sente_Dificuldade=AlgumasVeze 10   conf:(0.83) lift:(1.59) lev:(0.08) [3] < conv:(1.91)>
2. QuartilEnvioSIM=Baixo 13 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 12   conf:(0.92) lift:(1.23) lev:(0.05) [2] < conv:(1.63)>
3. QuartilEnvioSIM=MedioAlto 13 ==> ConceitoFinal=B 10   conf:(0.77) lift:(1.47) lev:(0.07) [3] < conv:(1.55)>
4. Disciplina_TeveDificuldadeConteudo=AlgumasVeze ConceitoFinal=B 16 ==> Grupo_Sente_Dificuldade=AlgumasVeze 12   conf:(0.75) lift:(1.43) lev:(0.08) [3] < conv:(1.53)>
5. Grupo_Envio_Atraso=Raramente Disciplina_TeveDificuldadeConteudo=AlgumasVeze 12 ==> q2=Desanimado 10   conf:(0.83) lift:(1.31) lev:(0.05) [2] < conv:(1.45)>
6. Grupo_Sente_Dificuldade=AlgumasVeze 23 ==> ConceitoFinal=B 16   conf:(0.7) lift:(1.33) lev:(0.09) [3] < conv:(1.37)>
7. ConceitoFinal=B 23 ==> Grupo_Sente_Dificuldade=AlgumasVeze 16   conf:(0.7) lift:(1.33) lev:(0.09) [3] < conv:(1.37)>
8. QuartilEnvioSIM=Baixo 13 ==> q2=Desanimado Disciplina_TeveDificuldadeConteudo=AlgumasVeze 9   conf:(0.69) lift:(1.38) lev:(0.06) [2] < conv:(1.3)>
9. q2=Desanimado QuartilEnvioSIM=Baixo 10 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 9   conf:(0.9) lift:(1.2) lev:(0.03) [1] < conv:(1.25)>
10. q2=Insatisfeito 13 ==> ConceitoFinal=B 9   conf:(0.69) lift:(1.32) lev:(0.05) [2] < conv:(1.24)>

```

APÊNDICE E – MINERAÇÃO COMPONENTE TAREFA – ETAPA 3

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 6

Best rules found:

1. q5=Desanimado QuartilEnvioSIM=MedioAlto 8 ==> ConceitoFinal=B 8 conf:(1) lift:(1.91) lev:(0.09) [3] < conv:(3.82)>
2. q5=Desanimado QuartilEnvioSIM=Baixo 8 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 8 conf:(1) lift:(1.33) lev:(0.05) [2] < conv:(2)>
3. Disciplina_TeveDificuldadeConteudo=AlgumasVeze ConceitoFinal=C 10 ==> QuartilEnvioSIM=Baixo 7 conf:(0.7) lift:(2.37) lev:(0.09) [4] < conv:(1.76)>
4. ConceitoFinal=C QuartilEnvioSIM=Baixo 7 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 7 conf:(1) lift:(1.33) lev:(0.04) [1] < conv:(1.75)>
5. QuartilEnvioSIM=Baixo 13 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 12 conf:(0.92) lift:(1.23) lev:(0.05) [2] < conv:(1.63)>
6. QuartilEnvioSIM=MedioAlto 13 ==> ConceitoFinal=B 10 conf:(0.77) lift:(1.47) lev:(0.07) [3] < conv:(1.55)>
7. QuartilEnvioSIM=MedioAlto 13 ==> q5=Desanimado ConceitoFinal=B 8 conf:(0.62) lift:(1.93) lev:(0.09) [3] < conv:(1.48)>
8. ConceitoFinal=C 12 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze QuartilEnvioSIM=Baixo 7 conf:(0.58) lift:(2.14) lev:(0.08) [3] < conv:(1.45)>
9. Disciplina_TeveDificuldadeConteudo=AlgumasVeze QuartilEnvioSIM=Baixo 12 ==> ConceitoFinal=C 7 conf:(0.58) lift:(2.14) lev:(0.08) [3] < conv:(1.45)>
10. QuartilEnvioSIM=Baixo 13 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze ConceitoFinal=C 7 conf:(0.54) lift:(2.37) lev:(0.09) [4] < conv:(1.44)>

APÊNDICE F – MINERAÇÃO COMPONENTE ARQUIVO – ETAPA 1

Neste apêndice, podem ser constatadas as regras obtidas através da mineração do componente arquivo para 10 atributos.

1. QuartilSim=Alto 11 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 11 conf:(1) lift:(1.33) lev:(0.06) [2] < conv:(2.75)>
2. q5=Desanimado Individual_Sente_Dificuldade=Raramente 11 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 11 conf:(1) lift:(1.33) lev:(0.06) [2] < conv:(2.75)>
3. Individual_Sente_Dificuldade=Raramente 15 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 14 conf:(0.93) lift:(1.24) lev:(0.06) [2] < conv:(1.88)>
4. q5=Desanimado 27 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 24 conf:(0.89) lift:(1.19) lev:(0.09) [3] < conv:(1.69)>
5. q5=Desanimado Disciplina_TeveDificuldadeConteudo=AlgumasVeze 20 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 18 conf:(0.9) lift:(1.2) lev:(0.07) [3] < conv:(1.67)>
6. q1=Insatisfeito 18 ==> Individual_Sente_Dificuldade=AlgumasVeze 14 conf:(0.78) lift:(1.37) lev:(0.09) [3] < conv:(1.55)>
7. Individual_Sente_Dificuldade=Raramente Disciplina_TeveDificuldadeConteudo=AlgumasVeze 12 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 11 conf:(0.92) lift:(1.22) lev:(0.05) [2]
8. Individual_Sente_Dificuldade=Raramente 15 ==> q5=Desanimado Disciplina_Consegue_AcompanharConteudos=QuaseSempre 11 conf:(0.73) lift:(1.34) lev:(0.06) [2] < conv:(1.36)>
9. q1=Desanimado Individual_Envio_Atraso=AlgumasVeze 14 ==> q5=Desanimado 11 conf:(0.79) lift:(1.28) lev:(0.05) [2] < conv:(1.35)>
10. q3=Desanimado Disciplina_Consegue_AcompanharConteudos=QuaseSempre 14 ==> q5=Desanimado 11 conf:(0.79) lift:(1.28) lev:(0.05) [2] < conv:(1.35)>
11. Individual_Sente_Dificuldade=Raramente Disciplina_Consegue_AcompanharConteudos=QuaseSempre 14 ==> q5=Desanimado 11 conf:(0.79) lift:(1.28) lev:(0.05) [2] < conv:(1.35)>
12. q1=Desanimado Disciplina_Consegue_AcompanharConteudos=QuaseSempre 17 ==> q5=Desanimado 13 conf:(0.76) lift:(1.25) lev:(0.06) [2] < conv:(1.31)>
13. q1=Desanimado q5=Desanimado 15 ==> Individual_Envio_Atraso=AlgumasVeze 11 conf:(0.73) lift:(1.29) lev:(0.06) [2] < conv:(1.3)>
14. Individual_Sente_Dificuldade=Raramente 15 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre Disciplina_TeveDificuldadeConteudo=AlgumasVeze 11 conf:(0.73) lift:(1.29) lev:(0.06) [2]
15. Disciplina_Consegue_AcompanharConteudos=QuaseSempre 33 ==> q5=Desanimado 24 conf:(0.73) lift:(1.19) lev:(0.09) [3] < conv:(1.27)>
16. q5=Desanimado Individual_Envio_Atraso=AlgumasVeze 16 ==> q1=Desanimado 11 conf:(0.69) lift:(1.32) lev:(0.06) [2] < conv:(1.27)>
17. Individual_Envio_Atraso=AlgumasVeze 25 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 21 conf:(0.84) lift:(1.12) lev:(0.05) [2] < conv:(1.25)>
18. q1=Desanimado q5=Desanimado 15 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 13 conf:(0.87) lift:(1.16) lev:(0.04) [1] < conv:(1.25)>
19. Individual_Sente_Dificuldade=AlgumasVeze 25 ==> q1=Insatisfeito 14 conf:(0.56) lift:(1.37) lev:(0.09) [3] < conv:(1.23)>
20. Disciplina_Consegue_AcompanharConteudos=QuaseSempre Disciplina_TeveDificuldadeConteudo=AlgumasVeze 25 ==> q5=Desanimado 18 conf:(0.72) lift:(1.17) lev:(0.06) [2] < conv:(1.21)>

APÊNDICE G – MINERAÇÃO COMPONENTE ARQUIVO – ETAPA 2

Neste apêndice, podem ser constatadas as regras obtidas através da mineração do componente arquivo para 6 atributos.

Best rules found:

1. q3=Insatisfeito QuartilSim=Alto 7 ==> q1=Desanimado 7 conf:(1) lift:(1.91) lev:(0.08) [3] < conv:(3.34)>
2. ARquivo_quartil=Baixo 11 ==> QuartilSim=Baixo 9 conf:(0.82) lift:(3.27) lev:(0.14) [6] < conv:(2.75)>
3. QuartilSim=Baixo 11 ==> ARquivo_quartil=Baixo 9 conf:(0.82) lift:(3.27) lev:(0.14) [6] < conv:(2.75)>
4. q1=Desanimado q3=Insatisfeito 10 ==> QuartilSim=Alto 7 conf:(0.7) lift:(2.8) lev:(0.1) [4] < conv:(1.88)>
5. q1=Desanimado QuartilSim=Alto 9 ==> q3=Insatisfeito 7 conf:(0.78) lift:(2.01) lev:(0.08) [3] < conv:(1.84)>
6. QuartilSim=Alto 11 ==> q1=Desanimado 9 conf:(0.82) lift:(1.57) lev:(0.07) [3] < conv:(1.75)>
7. QuartilSim=Alto 11 ==> q1=Desanimado q3=Insatisfeito 7 conf:(0.64) lift:(2.8) lev:(0.1) [4] < conv:(1.7)>
8. ARquivo_quartil=Medio 12 ==> q5=Desanimado 10 conf:(0.83) lift:(1.36) lev:(0.06) [2] < conv:(1.55)>
9. ConceitoFinal=B ARquivo_quartil=Medio 8 ==> q5=Desanimado 7 conf:(0.88) lift:(1.43) lev:(0.05) [2] < conv:(1.55)>
10. QuartilSim=Baixo 11 ==> q3=Desanimado 8 conf:(0.73) lift:(1.45) lev:(0.06) [2] < conv:(1.38)>
11. QuartilSim=Medio 11 ==> q3=Desanimado 8 conf:(0.73) lift:(1.45) lev:(0.06) [2] < conv:(1.38)>
12. ARquivo_quartil=Medio 12 ==> q5=Desanimado ConceitoFinal=B 7 conf:(0.58) lift:(1.83) lev:(0.07) [3] < conv:(1.36)>
13. QuartilSim=Alto 11 ==> q3=Insatisfeito 7 conf:(0.64) lift:(1.65) lev:(0.06) [2] < conv:(1.35)>
14. ARquivo_quartil=Alto 11 ==> ConceitoFinal=B 8 conf:(0.73) lift:(1.39) lev:(0.05) [2] < conv:(1.31)>
15. q5=Desanimado ConceitoFinal=B 14 ==> ARquivo_quartil=Medio 7 conf:(0.5) lift:(1.83) lev:(0.07) [3] < conv:(1.27)>
16. q3=Insatisfeito 17 ==> q1=Desanimado QuartilSim=Alto 7 conf:(0.41) lift:(2.01) lev:(0.08) [3] < conv:(1.23)>
17. ConceitoFinal=C 12 ==> q3=Desanimado 8 conf:(0.67) lift:(1.33) lev:(0.05) [2] < conv:(1.2)>
18. q5=Desanimado ARquivo_quartil=Medio 10 ==> ConceitoFinal=B 7 conf:(0.7) lift:(1.34) lev:(0.04) [1] < conv:(1.19)>
19. q3=Insatisfeito 17 ==> ConceitoFinal=B 11 conf:(0.65) lift:(1.24) lev:(0.05) [2] < conv:(1.16)>
20. q3=Insatisfeito 17 ==> QuartilSim=Alto 7 conf:(0.41) lift:(1.65) lev:(0.06) [2] < conv:(1.16)>

APÊNDICE H – MINERAÇÃO COMPONENTE ARQUIVO – ETAPA 2

Mineração realizada para a validação das regras do componente arquivo para a turma do experimento B.

Best rules found:

1. Individual_Sente_Dificuldade=Raramente 11 ==> q3=Desanimado 10 conf:(0.91) lift:(1.77) lev:(0.11) [4] < conv:(2.68)>
2. q1=Desanimado Individual_Envio_Atraso=AlgumasVezes 12 ==> ConceitoFinal=C 10 conf:(0.83) lift:(1.71) lev:(0.1) [4] < conv:(2.05)>
3. QuartilACEssosARquivos=Medio 10 ==> q1=Desanimado 10 conf:(1) lift:(1.24) lev:(0.05) [1] < conv:(1.95)>
4. Individual_Envio_Atraso=AlgumasVezes ConceitoFinal=C 10 ==> q1=Desanimado 10 conf:(1) lift:(1.24) lev:(0.05) [1] < conv:(1.95)>
5. Individual_Envio_Atraso=AlgumasVezes 13 ==> q1=Desanimado ConceitoFinal=C 10 conf:(0.77) lift:(1.86) lev:(0.11) [4] < conv:(1.9)>
6. Individual_Envio_Atraso=AlgumasVezes 13 ==> ConceitoFinal=C 10 conf:(0.77) lift:(1.58) lev:(0.09) [3] < conv:(1.66)>
7. q3=Insatisfeito 15 ==> Individual_Sente_Dificuldade=AlgumasVezes 12 conf:(0.8) lift:(1.37) lev:(0.08) [3] < conv:(1.55)>
8. q1=Desanimado ConceitoFinal=C 17 ==> Individual_Envio_Atraso=AlgumasVezes 10 conf:(0.59) lift:(1.86) lev:(0.11) [4] < conv:(1.45)>
9. q3=Desanimado 21 ==> q5=Desanimado 15 conf:(0.71) lift:(1.27) lev:(0.08) [3] < conv:(1.32)>
10. q1=Desanimado Disciplina_teveDificuldadeConteudo=AlgumasVezes 21 ==> q5=Desanimado 15 conf:(0.71) lift:(1.27) lev:(0.08) [3] < conv:(1.32)>
11. q1=Desanimado q3=Desanimado 18 ==> q5=Desanimado 13 conf:(0.72) lift:(1.29) lev:(0.07) [2] < conv:(1.32)>
12. ConceitoFinal=C 20 ==> q1=Desanimado Individual_Envio_Atraso=AlgumasVezes 10 conf:(0.5) lift:(1.71) lev:(0.1) [4] < conv:(1.29)>
13. q3=Desanimado 21 ==> Individual_Sente_Dificuldade=Raramente 10 conf:(0.48) lift:(1.77) lev:(0.11) [4] < conv:(1.28)>
14. Individual_Envio_Atraso=AlgumasVezes 13 ==> q1=Desanimado 12 conf:(0.92) lift:(1.15) lev:(0.04) [1] < conv:(1.27)>
15. q5=Desanimado 23 ==> q3=Desanimado 15 conf:(0.65) lift:(1.27) lev:(0.08) [3] < conv:(1.25)>
16. q5=Desanimado 23 ==> q1=Desanimado Disciplina_teveDificuldadeConteudo=AlgumasVezes 15 conf:(0.65) lift:(1.27) lev:(0.08) [3] < conv:(1.25)>
17. ConceitoFinal=C 20 ==> Individual_Envio_Atraso=AlgumasVezes 10 conf:(0.5) lift:(1.58) lev:(0.09) [3] < conv:(1.24)>
18. q1=Desanimado ConceitoFinal=A 14 ==> q5=Desanimado 10 conf:(0.71) lift:(1.27) lev:(0.05) [2] < conv:(1.23)>
19. q3=Desanimado Disciplina_teveDificuldadeConteudo=AlgumasVezes 14 ==> q5=Desanimado 10 conf:(0.71) lift:(1.27) lev:(0.05) [2] < conv:(1.23)>
20. q1=Desanimado q5=Desanimado 20 ==> q3=Desanimado 13 conf:(0.65) lift:(1.27) lev:(0.07) [2] < conv:(1.22)>

APÊNDICE I – MINERAÇÃO COMPONENTE FÓRUM – APRIORI

A mineração realizada neste experimento diz respeito ao componente fórum. Foi realizada utilizando o algoritmo Apriori com nove atributos para as 44 instâncias.

```

Apriori
=====

Minimum support: 0.25 (11 instances)
Minimum metric <conviction>: 1.1
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 17
Size of set of large itemsets L(2): 21
Size of set of large itemsets L(3): 8

Best rules found:

1. FGrup_viewDiscussion_Quartil=Baixo 12 ==> FGrup_addPost_Quartil2=Baixo 11   conf:(0.92) lift:(2.24) lev:(0.14) [6] < conv:(3.55)>
2. Grupo_Sente_Dificuldade=Raramente 14 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre 14   conf:(1) lift:(1.33) lev:(0.08) [3] < conv:(3.5)>
3. FGrup_viewForum_Quartil=Baixo 12 ==> q2=Desanimado 11   conf:(0.92) lift:(1.44) lev:(0.08) [3] < conv:(2.18)>
4. Disciplina_Consegue_AcompanharConteudos=QuaseSempre FGrup_addPost_Quartil2=Baixo 12 ==> q2=Desanimado 11   conf:(0.92) lift:(1.44) lev:(0.08) [3] < conv:(2.18)>
5. FGrup_addPost_Quartil2=Baixo 18 ==> q2=Desanimado 15   conf:(0.83) lift:(1.31) lev:(0.08) [3] < conv:(1.64)>
6. FGrup_addPost_Quartil2=Baixo 18 ==> FGrup_viewDiscussion_Quartil=Baixo 11   conf:(0.61) lift:(2.24) lev:(0.14) [6] < conv:(1.64)>
7. Disciplina_TeveDificuldadeConteudo=AlgumasVeze FGrup_addPost_Quartil2=Baixo 13 ==> q2=Desanimado 11   conf:(0.85) lift:(1.33) lev:(0.06) [2] < conv:(1.58)>
8. Disciplina_TeveDificuldadeConteudo=AlgumasVeze ConceitoFinal=B 16 ==> Grupo_Sente_Dificuldade=AlgumasVeze 12   conf:(0.75) lift:(1.43) lev:(0.08) [3] < conv:(1.53)>
9. Grupo_Sente_Dificuldade=AlgumasVeze 23 ==> ConceitoFinal=B 16   conf:(0.7) lift:(1.33) lev:(0.09) [3] < conv:(1.37)>
10. ConceitoFinal=B 23 ==> Grupo_Sente_Dificuldade=AlgumasVeze 16   conf:(0.7) lift:(1.33) lev:(0.09) [3] < conv:(1.37)>
11. q2=Desanimado Disciplina_Consegue_AcompanharConteudos=QuaseSempre 21 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 18   conf:(0.86) lift:(1.14) lev:(0.05) [2] < conv:(1.31)>
12. Grupo_Sente_Dificuldade=AlgumasVeze Disciplina_Consegue_AcompanharConteudos=QuaseSempre 15 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 13   conf:(0.87) lift:(1.16) lev:(0.04) [1]
13. Grupo_Sente_Dificuldade=AlgumasVeze 23 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze ConceitoFinal=B 12   conf:(0.52) lift:(1.43) lev:(0.08) [3] < conv:(1.22)>
14. q2=Desanimado 28 ==> FGrup_addPost_Quartil2=Baixo 15   conf:(0.54) lift:(1.31) lev:(0.08) [3] < conv:(1.18)>
15. FGrup_addPost_Quartil2=Baixo 18 ==> q2=Desanimado Disciplina_Consegue_AcompanharConteudos=QuaseSempre 11   conf:(0.61) lift:(1.28) lev:(0.05) [2] < conv:(1.18)>
16. Grupo_Sente_Dificuldade=AlgumasVeze 23 ==> Disciplina_TeveDificuldadeConteudo=AlgumasVeze 19   conf:(0.83) lift:(1.1) lev:(0.04) [1] < conv:(1.15)>
17. Disciplina_Consegue_AcompanharConteudos=QuaseSempre Disciplina_TeveDificuldadeConteudo=AlgumasVeze 25 ==> q2=Desanimado 18   conf:(0.72) lift:(1.13) lev:(0.05) [2] < conv:(1.14)>
18. Grupo_Sente_Dificuldade=AlgumasVeze Disciplina_TeveDificuldadeConteudo=AlgumasVeze 19 ==> ConceitoFinal=B 12   conf:(0.63) lift:(1.21) lev:(0.05) [2] < conv:(1.13)>
19. q2=Desanimado 28 ==> FGrup_viewForum_Quartil=Baixo 11   conf:(0.39) lift:(1.44) lev:(0.08) [3] < conv:(1.13)>
20. q2=Desanimado 28 ==> Disciplina_Consegue_AcompanharConteudos=QuaseSempre FGrup_addPost_Quartil2=Baixo 11   conf:(0.39) lift:(1.44) lev:(0.08) [3] < conv:(1.13)>

```

APÊNDICE J – MINERAÇÃO COMPONENTE FÓRUM – VALIDAÇÃO

Neste apêndice, é apresentado o resultado da mineração feita para a validação do componente fórum ao utilizar o algoritmo Apriori para nove atributos.

```

Apriori
=====

Minimum support: 0.25 (10 instances)
Minimum metric <conviction>: 1.1
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 18

Size of set of large itemsets L(2): 37

Size of set of large itemsets L(3): 20

Best rules found:

1. Grupo_Envio_Atraso=Raramente 21 ==> q2=Desanimado 21   conf:(1) lift:(1.17) lev:(0.07) [3] < conv:(3.07)>
2. FG_ViewForum=Baixo 12 ==> FG_ViewDiscussion=Baixo 10   conf:(0.83) lift:(2.44) lev:(0.14) [5] < conv:(2.63)>
3. FG_ViewForum=Baixo 12 ==> q2=Desanimado FG_ViewDiscussion=Baixo 10   conf:(0.83) lift:(2.44) lev:(0.14) [5] < conv:(2.63)>
4. q2=Desanimado FG_ViewForum=Baixo 12 ==> FG_ViewDiscussion=Baixo 10   conf:(0.83) lift:(2.44) lev:(0.14) [5] < conv:(2.63)>
5. Grupo_Envio_Atraso=Raramente Disciplina_TeveDificuldadeConteudo=AlgumasVeze 15 ==> q2=Desanimado 15   conf:(1) lift:(1.17) lev:(0.05) [2] < conv:(2.2)>
6. q2=Desanimado Disciplina_Consegue_AcompanharConteudos=AlgumasVeze 13 ==> Grupo_Envio_Atraso=Raramente 11   conf:(0.85) lift:(1.65) lev:(0.11) [4] < conv:(2.11)>
7. FG_ViewDiscussion=Baixo 14 ==> q2=Desanimado 14   conf:(1) lift:(1.17) lev:(0.05) [2] < conv:(2.05)>
8. FG_ViewDiscussion=Baixo 14 ==> FG_ViewForum=Baixo 10   conf:(0.71) lift:(2.44) lev:(0.14) [5] < conv:(1.98)>
9. FG_ViewDiscussion=Baixo 14 ==> q2=Desanimado FG_ViewForum=Baixo 10   conf:(0.71) lift:(2.44) lev:(0.14) [5] < conv:(1.98)>
10. q2=Desanimado FG_ViewDiscussion=Baixo 14 ==> FG_ViewForum=Baixo 10   conf:(0.71) lift:(2.44) lev:(0.14) [5] < conv:(1.98)>
11. FG_ViewForum=Baixo 12 ==> q2=Desanimado 12   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.76)>
12. Disciplina_TeveDificuldadeConteudo=AlgumasVeze FG_addPost=BaixoMedio 12 ==> q2=Desanimado 12   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.76)>
13. FG_ViewForum=Baixo 12 ==> q2=Desanimado Disciplina_TeveDificuldadeConteudo=AlgumasVeze 10   conf:(0.83) lift:(1.49) lev:(0.08) [3] < conv:(1.76)>
14. Grupo_Envio_Atraso=Raramente Grupo_Sente_Dificuldade=AlgumasVeze 11 ==> q2=Desanimado 11   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.61)>
15. Grupo_Envio_Atraso=Raramente Disciplina_Consegue_AcompanharConteudos=AlgumasVeze 11 ==> q2=Desanimado 11   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.61)>
16. Disciplina_Consegue_AcompanharConteudos=AlgumasVeze 15 ==> Grupo_Envio_Atraso=Raramente 11   conf:(0.73) lift:(1.43) lev:(0.08) [3] < conv:(1.46)>
17. Disciplina_Consegue_AcompanharConteudos=AlgumasVeze 15 ==> q2=Desanimado Grupo_Envio_Atraso=Raramente 11   conf:(0.73) lift:(1.43) lev:(0.08) [3] < conv:(1.46)>
18. FG_ViewDiscussion=Medio 10 ==> q2=Desanimado 10   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.46)>
19. FG_ViewForum=Medio 10 ==> q2=Desanimado 10   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.46)>
20. Grupo_Envio_Atraso=Raramente ConceitoFinal=A 10 ==> q2=Desanimado 10   conf:(1) lift:(1.17) lev:(0.04) [1] < conv:(1.46)>

```

ANEXO A – GRUPOS DE PESQUISAS CNPQ – COMPUTAÇÃO AFETIVA

Qtd	Nome do grupo	Líder (es)	Instituição	Área	Objetivos	Contato
1	Grupo de Pesquisa Computação Afetiva	Magda Bercht Patrícia Alejandra Behar	UFRGS	Ciências Exatas e da Terra ; Ciência da Computação	- Estudar as emoções no contexto da Inteligência Artificial e da Informática na Educação; - pesquisar as dimensões da emoção numa perspectiva de utilidade para a construção de sistemas inteligentes; - pesquisar e desenvolver técnicas para avaliar e desenvolver sistemas computacionais aplicados a educação: sistemas computador aluno, sistemas na Web e de interação entre agentes quando um dos agentes é artificial; - desenvolver habilidades para a construção de protótipos que tenham por função a identificação, expressão e raciocínio sobre e de emoções em interações com agentes artificiais e biológicos	pgie@pgie.ufrgs.br
2	ACTDESIGN - Núcleo de Pesquisa em Arte, Processos Sensíveis de Comunicação e Tecnologias Afectivas	Maria Lucília Borges	Universidade Federal de Ouro Preto - UFOP	Linguística, Letras e Artes; Artes	Pretende trabalhar na intersecção e(ntr)e arte, ciência e tecnologia, com instrumentos teóricos e tecnológicos que atravessam o campo da filosofia, da biologia e da estética, desde os conceitos às (possibilidades de) experimentações poético-estéticas.	luciliaborges@gmail.com
3	Grupo de Computação Cognitiva Aplicada - COCA	Claudio Cesar de Sá Rafael Stubs Parpinelli	Universidade do Estado de Santa Catarina – UDESC	Ciências Exatas e da Terra; Ciência da Computação	Desenvolver estudos sobre temas que correlacionem três tópicos: Sistemas Computacionais, Comportamento Inteligente e Problemas Complexos.	claudio@colmeia.udesc.br
4	Grupo de Estudos e Pesquisa em Informática na Educação - GEPIED	Henrique Nou Schneider Maria Augusta Silveira Netto Nunes	Universidade Federal de Sergipe - UFS	Ciências Exatas e da Terra; Ciência da Computação	Desenvolver estudos, pesquisas e produtos digitais envolvendo a problemática da utilização da Informática na Educação visando o desenvolvimento econômico e social do Brasil	hns@terra.com.br

Qtd	Nome do grupo	Líder (es)	Instituição	Área	Objetivos	Contato
5	Grupo de Pesquisa em Inteligência e Imagens	Hendrik Teixeira Macedo Leonardo Nogueira Matos	Universidade Federal de Sergipe - UFS	Ciências Exatas e da Terra; Ciência da Computação	Não consta	hendrik@ufs.br
6	Inteligência Computacional e Pesquisa Operacional	Ádamo Lima de Santana	Universidade Federal do Pará – UFPA	Engenharias; Engenharia Elétrica	Investigar novos modelos, bem como aprimoramentos para sistemas e algoritmos clássicos, nas áreas de matemática, otimização e inteligência computacional; . Prover logística para o meio público e profissional, através do desenvolvimento de sistemas e metodologias para suporte à decisão, permitindo, particularmente, identificar padrões, realizar previsões e inferir cenários e comportamentos para análise e diagnóstico.	adamo@ufpa.br
7	Laboratório de Psicologia da Computação	Mauro Pinheiro Rodrigues Hugo Cristo Sant'Anna	Universidade Federal do Espírito Santo – UFES	Ciências Sociais Aplicadas; Desenho Industrial	Grupo criado para ampliar as pesquisas iniciadas pelo Núcleo de Interfaces Computacionais (2009-2011) nas áreas de Pensamento Computacional e Design de Interação.	ufes.labpc@gmail.com
8	NUTED-Núcleo de Tecnologia Digital aplicada à Educação	Patricia Alejandra Behar	Universidade Federal do Rio Grande do Sul – UFRGS	Ciências Humanas; Educação	O propósito deste grupo é desenvolver, usar, testar e validar ambientes virtuais de aprendizagem para a UFRGS e fora da instituição, que sirvam à Educação à distância.	pbehar@terra.com.br
9	Laboratório de Computação Aplicada à Educação e Tecnologia Social Avançada	Seiji Isotani Ellen Francine Barbosa	Universidade de São Paulo – USP	Ciências Exatas e da Terra; Ciência da Computação	Não informado	sisotani@icmc.usp.br
10	Núcleo de Mídias e Juventude	Jorge Fernando Silva de Araujo Marcia Martins de Oliveira		Ciências Humanas; Educação	Não informado	jorge.fernando@cp2.g12.br