

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL DO
CONHECIMENTO

RANILSON OSCAR ARAÚJO PAIVA

UMA FERRAMENTA PARA RECOMENDAÇÃO PEDAGÓGICA BASEADA EM
MINERAÇÃO DE DADOS EDUCACIONAIS

Maceió

2013

RANILSON OSCAR ARAÚJO PAIVA

Uma Ferramenta para Recomendação Pedagógica Baseada em Mineração de Dados Educacionais

Dissertação submetida à Coordenação do Mestrado em Modelagem Computacional do Conhecimento da Universidade Federal de Alagoas como parte dos requisitos necessários para obtenção do grau de Mestre em Modelagem Computacional do Conhecimento.

Orientador: Prof. Dr. Ig Ibert Bittencourt
Santana Pinto

Coorientador: Prof. Dr. Alan Pedro da Silva

Maceió

2013

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

BIBLIOTECÁRIA RESPONSÁVEL: HELENA CRISTINA PIMENTEL DO VALE

P149u Paiva, Ranilson Oscar Araújo.
Uma ferramenta para recomendação pedagógica baseada em mineração
de dados educacionais / Ranilson Oscar Araújo Paiva. - 2013.
117 f. : il.

Orientador: Ig Ibert Bittencourt Santana Pinto.

Coorientador: Alan Pedro da Silva.

Dissertação (mestrado em Modelagem Computacional de Conhecimento) –
Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2013.

Bibliografia: f. 100-103.

Apêndices: 104-117.

1. Cursos baseados na WEB. 2. Mineração de dados educacionais. 3. Processo de
recomendação pedagógica. I. Título.

CDU: 004.5: 37



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
Avenida Lourival Melo Mota, Km 14, Bloco 09, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364



Membros da Comissão Julgadora da Dissertação de Mestrado de Ranilson Oscar Araújo Paiva, intitulada: “Uma Ferramenta de Recomendação Pedagógica Baseada em Mineração de Dados Educacionais”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas em 03 de junho de 2013, às 14h00min, na sala de aula do Mestrado em Modelagem Computacional de Conhecimento.

COMISSÃO JULGADORA

Prof. Dr. Alan Pedro da Silva
UFAL – Instituto de Computação
Orientador

Prof. Dr. Ig Ibert Bittencourt Santana Pinto
UFAL – Instituto de Computação
Orientador

Prof. Dr. Aydano Pamponet Machado
UFAL – Instituto de Computação
Examinador

Prof. Dr. Seiji Isotani
USP – Instituto de Ciências Matemáticas e de Computação
Examinador

Maceió, junho de 2013.

Dedico essa dissertação aos meus pais **Keginaldo e Dilma**, pois vocês me deram a base e a essência que me permitiram iniciar, e concluir, essa jornada. À minha esposa **Andréia**, por ter suportado ao meu lado todas as dificuldades desse período tão árduo, por ser o meu norte e ter me dado amor, paz e carinho sempre que eu precisei. À minha filha **Sofia**, minha pequena "Sabedoria", por completar a nossa vida de uma forma que não é possível descrever, de uma forma que só é possível sentir! ... Sem vocês, eu não teria conseguido.

AGRADECIMENTOS

À **Deus**, pelas graças e benção. Pelos dias que vivi, e por todos os outros que viverei. Pelas lições que me fez aprender e pelos pedidos atendidos. É verdade que não o compreendo, mas sinto sua presença, todos os momentos, e isso me alegra e me conforta.

Aos meus orientadores, **Alan Pedro** e **Ig Ibert** por não se acomodarem na mediocridade, e não permitir que aqueles sob sua tutoria padeçam desse mal. Reproduzo, aqui, suas palavras: "Enxergo longe no ombro de gigantes". Muito obrigado!

À minha esposa **Andréia** e minha filha **Sofia**, minhas meninas, meus amores. Motivo de eu me alegrar ao chegar em casa, mesmo completamente esgotado. Minha esposa que é o ser humano de mais linda composição e conteúdo que já pude conhecer e, por isso, quis você para mim! E minha filha, minha pequena "Sabedoria". A única coisa que, realmente, fiz. Gerada da minha matéria e, como sua mãe mesmo sabe, você é como eu queria que fosse! É, em grande parte, por vocês que eu vivo a superar os desafios que surgem na minha vida! Que seja longa, saudável e feliz a nossa trajetória... o nosso caminho!

À minha mãe **Dilma**, que viveu uma vida de sacrifícios. Não por si, mas por toda a família! Por ter primado por minha educação e busca por conhecimento. Eu me enxergo em você... Eu sou você. De uma forma que conseguíamos nos entender sem que nenhuma palavra precisasse ser dita. Sinto a todo instante, que você está aqui perto de mim.

Ao meu pai **Keginaldo**, que criou a mim e ao meu irmão com base em exemplos. Não só falou, mas mostrou como se é que se vive. Suas palavras, acompanhadas de atitudes, demandam obediência! Obrigado por me criar com rigor, mas com carinho... muito carinho!

Ao meu irmão **Kinho**, tão parecido comigo, mas tão diferente ao mesmo tempo. Modulou o meu comportamento, e eu o dele. Parabéns por você ter a coragem de ser o que você realmente é. Sua autenticidade e sinceridade são, não os únicos, mas seus maiores brilhos.

À minha família: avós (**Maria das Neves** e **Maria Anunciada**), avôs (**Oscar** e **Ranilson**), tios (**Orlando**, **Rossini**, **Israel**, **Alberto**, **Francisco**, **Paulo**, **Oscar**, **Antônio** e **Dilton**), tias (**Itália** e **Flávia**), primos, primas. Sou apaixonado por vocês e, graças a Deus tenho pessoas de ótimo caráter em minha família. Vocês são um pilar da minha vida! Saibam que estou aqui por vocês, também.

À minha sogra **Rossana**, que cuida de minha esposa e filha na minha ausência, e está sempre preocupada com nosso bem-estar, muitas vezes esquecendo-se de si.

Aos meus amigos, que nunca hesitaram em responder meus chamados, peço desculpas pela ausência e agradeço o apoio.

Aos membros do NEES, MeuTutor, UFAL Línguas, amigos e grandes companheiros nesse processo. Com menção especial a **Daniel Borges, Diego Dermeval, Thyago Tenório, Endhe Elias, Gabriel Moretti, Olavo Holanda e Armando Barbosa**, os quais eu perturbei em demasia, mas estavam sempre dispostos a me ajudar com meus inúmeros "gaps" de conhecimento!

Espero ter agradecido a todos, mas caso não tenha mencionado alguém, me perdoe. Contudo, tenha certeza que minha gratidão é eterna!

A TODOS vocês, MUITO Obrigado!

RESUMO

A presente dissertação trata da criação de uma ferramenta para a recomendação pedagógica cujo objetivo é prover aos professores de cursos baseados na web, recomendações pedagógicas personalizadas geradas com base nos resultados da Mineração dos Dados Educacionais de seus alunos. Para orientar essa criação propomos o Processo de Recomendação Pedagógica, o qual conta com o trabalho conjunto e coordenado da Inteligência Humana (especialistas nos domínios envolvidos) e da Inteligência Artificial (ferramentas computacionais). O processo é constituído de quatro etapas que ocorrem de forma cíclica e sequencial, iniciando com “Detectar Práticas”, onde detectamos se existem ações afetando o processo de ensino e aprendizagem. Na etapa seguinte, “Descobrir Padrões”, utilizamos as técnicas de Mineração de Dados Educacionais, por meio de Cenários de Mineração predefinidos, para encontrar padrões de interesse pedagógico acerca das práticas detectadas. Na próxima etapa, “Recomendar”, são oferecidas recomendações apropriadas a atual situação pedagógica do aluno. Finalmente a etapa “Monitorar e Avaliar”, onde acompanhamos e analisamos se os alunos foram afetados positivamente pelas recomendações e se estas foram relevantes. A ferramenta de recomendação proposta foi utilizada em um estudo de caso, com dados reais provenientes de um curso de língua Espanhola com 200 alunos que produziram mais de **700 megabytes** de informações dispostas em, aproximadamente, **1220000 triplas**. Como resultados, fomos capazes de detectar práticas e os padrões associados a elas, que foram utilizados na criação de recomendações, avaliadas (relevância) por especialistas no domínio educacional/pedagógico, e disponibilizadas para que os usuários finais (professores) as ofereçam a seus alunos.

Palavras-chave: Cursos baseados na Web. Mineração de dados Educacionais. Processo de recomendação pedagógica. Recomendações pedagógicas. Sistema de recomendação.

ABSTRACT

This work is about the creation of a tool for pedagogical recommendation which objective is to provide teachers, from web-based courses, personalized pedagogical recommendations generated based on the mining results of their students' educational data. In order to guide this creation, we propose the Pedagogical Recommendation Process that counts on the coordinated work and cooperation of the Human Intelligence (domain specialists) and the Artificial Intelligence (computational tools). The process is constituted of four steps that occur in a sequential and cyclic way, starting with "Detect Practices", where we detect if there are actions affecting the teaching and learning process. Is the next step, "Discover Patterns", we use educational data mining techniques, based on predefined mining scenarios, to find patterns with pedagogical significance for the practices detected. In the following step, "Recommend", it is where appropriate recommendations are offered, given the students' current pedagogical situation. Finally, the "Monitor and Evaluate" step, where it is analyzed whether the students were positively affected by the recommendations and if they were relevant. The proposed tool was used in a case study with real data provided by a Spanish language course with 200 students enrolled, who produced more than 700 megabytes of information contained in, approximately, 1220000 triples. As results we were able to detected practices and the patterns associated to them, which were used to create recommendations, evaluated (relevance) by specialists in the educational/pedagogical domain and made available for the final users (teachers) to suggest them to their students.

Keywords: Educational Data Mining. Recommender Systems. Pedagogical Recommendations. Pedagogical Recommendation Process. Web-Based Courses.

LISTA DE ILUSTRAÇÕES

Figura 1 - Número de alunos por metas do sistema.	19
Figura 2 - Quantidade de exibições do primeiro e ultimo vídeos para cada semana do curso.	19
Figura 3 - Representação do processo CRISP-DM.	29
Figura 4 - Arquitetura do sistema colaborativo de recomendação proposto no artigo.	37
Figura 5 - Arquitetura do EDUCA.	39
Figura 6 - Arquitetura da ferramenta de mineração e do engenho de recomendação propostos.	41
Figura 7 - Arquitetura PROTUS 2.0.	43
Figura 8 - Arquitetura do I-HELP.	44
Figura 9 - O Processo de Recomendação Pedagógica na interseção das áreas de Inteligência Artificial, Psicologia Cognitiva e as Tecnologias Educacionais.	49
Figura 10 - A interação entre os dados e os atores do Processo de Recomendação Pedagógica.	50
Figura 11 - As quatro etapas do Processo de Recomendação Pedagógica.	50
Figura 12 - Gráficos apresentando prática de aprendizagem.	52
Figura 13 - Representação de alguns recursos educacionais disponíveis no ambiente UFAL Línguas. 1) Exercícios, 2) Preencher os Dados Pessoais em Espanhol, 3) Aulas em Vídeo, 4) Chat e 5) Fórum.	55
Figura 14 - Diagrama de casos de uso.	69
Figura 15 - Arquitetura de alto nível da Ferramenta de Recomendação.	70
Figura 16 - Diagrama de Componentes.	72
Figura 17 - Tela de apresentação do Assistente de Gestão de Aprendizagem.	76
Figura 18 - Os cenários de mineração disponíveis.	77
Figura 19 - Selecionando uma das recomendações pedagógicas propostas.	77
Figura 20 - Monitoramento e avaliação do impacto das recomendações no desempenho dos alunos.	79
Figura 21 - Diagrama de sequência com as etapas da mineração dos dados educacionais.	81
Figura 22 - Diagrama de sequência com as etapas da recomendação.	82
Figura 23 - Gráfico de qualidade das interações dos alunos no ambiente de aprendizagem. ..	84
Figura 24 - Detecção das Práticas.	85
Figura 25: Boxplots dos dados acesso e pontuação nos testes.	87

Figura 26 - Gráfico de qualidade das interações dos alunos no ambiente de aprendizagem. .	87
Figura 27 - Representação textual da árvore de decisão inferida pelo algoritmo C4.5.	88
Figura 28 - Árvore de decisão com pontos de interesse circulados.	89
Figura 29 - Tela de apresentação do Assistente de Gestão de Aprendizagem.	91
Figura 30 - Os cenários de mineração disponíveis.	92
Figura 31 - Selecionando uma das recomendações pedagógicas propostas.	93
Figura 32 - Monitoramento e avaliação do impacto das recomendações no desempenho dos alunos.	93
Figura 33 - Código R anotado. Análise estatística dos dados interacionais para o cenário 1.	105
Figura 34 - Código R anotado. Análise estatística dos dados interacionais para o cenário 1.	106
Continuação.	106
Figura 35 - Implementação de um cenário de mineração, que constitui da definição de suas configurações que serão utilizadas pelo componente de mineração (parte 1).	110
Figura 36 - Implementação de um cenário de mineração. (parte 2).	111
Figura 37 - Método de construção do arquivo ARFF, parte do componente de mineração.	112
Figura 38 - Método responsável pela transformação de dados. Parte do pré-processamento dos dados.	113
Figura 39 - Métodos de tratamento de instâncias ausentes e valores não numéricos.	114
Figura 40 - Método responsável por verificar a existência de um cenário.	114
Figura 41 - Método de mineração dos dados.	115
Figura 42 - Conector de cenários. Artefato através do qual o componente de mineração requer as configurações e dados providos pelos cenários (parte 1).	116
Figura 43 - Conector de cenários. (parte 2).	117
Figura 44 - Conector de cenários. (parte 3).	117

LISTA DE TABELAS

Tabela 1 - Resultado da enquete Onde você aplicou técnicas analíticas/mineração de dados em 2012.	26
Tabela 2 - Os 10 algoritmos de mineração mais utilizados.	30
Tabela 3 - Capacidades e limitações dos algoritmos escolhidos.	32
Tabela 4 - Tabela comparativa dos trabalhos relacionados.	46
Tabela 5 - Descrição dos cenários de mineração criados.	58
Tabela 6 - Respostas às perguntas criadas para compreender os cenários de mineração.	60
Tabela 7 - Especificação dos cenários criados nessa dissertação.	64
Tabela 8 - Formato de apresentação, para o aluno, das recomendações pedagógicas.	75
Tabela 9 - Dados do cenário Qualidade da Interação dos Alunos no Ambiente de Aprendizagem.	86
Tabela 10 - Parâmetros utilizados na mineração do cenário.	90
Tabela 11 - Resultados do questionário respondido pelos professores a respeito das recomendações.	109

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Motivação e Contextualização do Trabalho	15
1.2	Problemática	18
1.3	Objetivos	22
1.4	Contribuições do Trabalho	23
1.5	Organização da Dissertação	23
2	FUNDAMENTAÇÃO TEÓRICA	24
2.1	Introdução à Mineração de Dados	24
2.2	Mineração de Dados Educacionais	26
2.3	O Processo de Mineração dos Dados	28
2.4	Os Algoritmos de Mineração de Dados	29
2.5	Sistemas de Recomendação	33
3	TRABALHOS RELACIONADOS	36
3.1	Collaborative Recommender System Using Distributed Rule Mining for Improving Web-based Adaptative Courses	36
3.2	EDUCA	38
3.3	Applying Web Usage Mining for Personalizing Hyperlinks in Web-based Adaptative Educational Systems	40
3.4	PROTUS 2.0	41
3.5	I-HELP	43
3.6	Tabela Comparativa	45
4	A PROPOSTA	47
4.1	O Processo de Recomendação Pedagógica	47
4.1.1	Detectar Práticas	50
4.1.2	Descobrir Padrões	52
4.1.3	Recomendar	54
4.1.4	Monitorar e Avaliar	56
4.2	Cenários de Mineração de Dados	57
4.3	A Ferramenta para Recomendação Pedagógica	65
4.3.1	Casos de Uso	66

4.3.2	Arquitetura.....	69
4.3.3	Arquitetura de Implementação	71
	O Gerenciador da Mineração (<i>MINING MGT</i>)	73
	Gerenciador dos Cenários (<i>SCENARIOS MGT</i>).....	73
	O Conector de Cenários (<i>SCENARIOS CONNECTOR</i>)	73
	O Conector de API (<i>API CONNECTOR</i>).....	74
	O Micro Kernel (<i>MICROKERNEL</i>)	74
4.3.4	Execução dos Módulos de Mineração e Recomendação.....	79
5	ESTUDO DE CASO.....	83
5.1	Detectar Práticas	83
5.2	Descobrir Padrões	85
5.3	Recomendar	90
5.4	Monitorar e Avaliar	91
5.5	A Interface com o Usuário.....	91
5.6	Discussão	94
6	CONCLUSÃO	97
6.1	Os Objetivos Específicos da Dissertação.....	97
6.2	Limitações	99
6.3	Perspectivas Futuras	99
	REFERÊNCIAS	100
	APÊNDICE A	104
	A Ferramenta de Mineração de Dados	104
	O Workbench Weka®	104
	APÊNDICE B	105
	Análise Estatística dos Dados (Script R).....	105
	APÊNDICE C	107
	Questionário de Recomendações	107
	O Questionário	107
	Os Resultados	109
	APÊNDICE D	110
	Implementação da Ferramenta.....	110

1 INTRODUÇÃO

O presente trabalho trata sobre a proposição, criação e avaliação de uma Ferramenta para a Recomendação Pedagógica, que está situada na linha de pesquisa de Descoberta de Conhecimento do Mestrado em Modelagem Computacional de Conhecimento, do Instituto de Computação da Universidade Federal de Alagoas. Trata-se de um tema inerentemente interdisciplinar, envolvendo áreas do conhecimento como Informática na Educação, Inteligência Artificial e Engenharia de Software. **O resultado dessa dissertação visa auxiliar professores e tutores de cursos mediados por computador, a prover ajuda a seus alunos na forma de Recomendações Pedagógicas Personalizadas.** Desta forma a personalização leva em consideração as características pessoais dos alunos. As informações que permitirão criar tais recomendações são obtidas por meio da Mineração dos Dados Educacionais oriundos das interações desses alunos dentro do ambiente de aprendizagem, onde a ferramenta será utilizada.

1.1 Motivação e Contextualização do Trabalho

O tema desta dissertação está relacionado com os índices que medem a qualidade do ensino e da educação no Brasil nos âmbitos nacional e internacional. Isso nos levou a uma busca pelos desafios, planejamentos, objetivos e metas governamentais dentro do contexto da computação na educação, e de que forma poderíamos contribuir com base nessas orientações.

O primeiro índice observado foi o **PISA**¹ (Programa para a Avaliação Internacional dos Alunos), que é uma iniciativa internacional organizada pela **OCDE** (Organização para Cooperação e Desenvolvimento Econômico), que realiza uma avaliação comparando alunos que estão concluindo a escolaridade básica obrigatória em diversos países (com idade aproximada de 15 anos). Seu objetivo é avaliar o quão bem as escolas, de cada país participante, estão preparando esses jovens para exercer a cidadania, provendo resultados que auxiliem os governos na definição e refinamento de políticas educativas. Isso é alcançado por meio da criação de indicadores que sirvam como base para o planejamento e implementação de políticas de melhoria da qualidade do ensino básico (PISA, 2009).

O resultado de 2009 exhibe o Brasil com 412 pontos, ocupando o **58º lugar** (de um total de 75 países avaliados) (WALKER, 2011). Tal pontuação está **abaixo da média da**

1 Sigla em Inglês para: Programme for International Students Assessment.

OCDE, que foi de 493 pontos, e deixa o Brasil atrás de outros países da América do Sul, como Chile, (44º colocado), Uruguai (50º colocado) e Colômbia (57º colocado). Frisa-se ainda que o Brasil ficou apenas à frente da Índia, dentre os países que constituem o BRIC² (Brasil [58º], Rússia [43º], Índia [73º e 74º]³ e China [1º, 4º e 28º]⁴).

Em seguida observamos o **IDEB** (Índice de Desenvolvimento da Educação Básica) que mede a qualidade de todas as escolas (ensino fundamental e ensino médio) das redes de ensino do país (pública, estadual, municipal e privada). O IDEB é um índice calculado através do desempenho dos alunos em avaliações e nas taxas de aprovação das escolas, nas disciplinas de Matemática e Português. As medições são feitas a cada dois anos e em 2011, ano da última medição, as médias dos resultados foram: 5.0 para os anos iniciais do ensino fundamental, 4.1 para os anos finais do ensino fundamental e 3.7 para o ensino médio. Notas que se encontram **abaixo da meta 6.0** que, segundo o Ministério da Educação, quando superada, indicará que o Brasil possui educação com qualidade compatível com países desenvolvidos (MEC; IDEB, 2012).

Tais indicadores mostram a difícil situação em que se encontra a educação no nosso país. Com essas informações apresentadas e buscando avaliar as demandas para a educação no Brasil, chegamos ao **Plano Nacional da Educação**, documento onde são definidos os objetivos e orientações para a educação nacional nos 10 anos seguintes à sua aprovação (PNE, 2010). O atual plano contém 10 diretrizes: (1) Erradicação do analfabetismo, (2) Universalização do atendimento escolar, (3) Superação das desigualdades educacionais, (4) Melhoria da qualidade do ensino, (5) Formação para o trabalho, (6) Promoção da sustentabilidade socioambiental, (7) Promoção humanística, científica e tecnológica do país, (8) Estabelecimento de meta de aplicação de recursos públicos em educação como proporção do produto interno bruto, (9) Valorização dos profissionais de educação e, por fim, (10) Difusão dos princípios da equidade, do respeito à diversidade e à gestão democrática da educação.

Tais diretrizes devem servir como norte para a criação das metas (objetivos) educacionais a serem atingidas no período de vigência do plano e, também, para o planejamento das estratégias que devem ser utilizadas para alcançar essas metas (PNE, 2010).

2 Acrônimo cunhado por Jim O'Neil, economista da Goldman Sachs, e que representa os países: Brasil, Rússia, Índia e China, os quais, segundo o economista, estão passando por um estágio semelhante em relação ao desenvolvimento econômico (CNN MONEY, 2009).

3 A Índia foi um dos 10 países adicionais que constituíram o PISA 2009+ (WALKER, 2011).

4 As avaliações realizadas na China foram realizadas em Xangai, Hong King e Macao, com cada uma retornando um resultado independente dos demais. (PISA, 2009).

Ao observarmos as ações descritas no PNE, percebemos a relação que existe entre tais ações e a aplicação de tecnologias da informação e comunicação como mecanismos para suas realizações, como acompanhamento pedagógico e individualizado, monitoramento de acesso, universalização do acesso, dentre outras.

Além dos índices e documentos supracitados, destaca-se também o documento intitulado “Grandes Desafios da Computação 2006 – 2016” (de Leon F. de Carvalho et al., 2006), criado pela **Sociedade Brasileira de Computação**, com o objetivo de motivar e guiar as pesquisas em computação realizadas no país. Nele foram definidos os cinco grandes desafios científicos da computação para o período mencionado. Um desses desafios refere-se ao “**Acesso Participativo e Universal do Cidadão Brasileiro ao Conhecimento**”, que visa conceber sistemas, ferramentas, modelos, métodos, procedimentos e teorias que sejam capazes de vencer, competentemente, as barreiras que impedem o acesso do cidadão brasileiro ao conhecimento.

A importância dessas questões foi reforçada em Julho de 2012, quando a Sociedade Brasileira de Computação promoveu, na cidade de Curitiba – Paraná, o XXXII Congresso da Sociedade Brasileira de Computação, com o título: Computação e Inovação: ampliando fronteiras para solução de desafios no Brasil. O congresso sediou o **I Workshop de Desafios da Computação Aplicada à Educação (DesafIE!)**, evento criado com o objetivo de debater temas relacionados a desafios em educação sob a perspectiva computacional, tendo como alguns de seus principais temas:

- A adaptação ao aprendiz como sensibilidade ao contexto;
- O registro de longo prazo dos perfis de aprendizes e tutores;
- A mineração de dados educacionais para viabilizar a busca por padrões de aprendizagem;
- A personalização da aprendizagem como meio de melhorar a qualidade do ensino e o desempenho dos alunos.

Em consonância com as metas propostas no PNE, os Grandes Desafios da Computação identificados pela SBC e os temas abordados no DesafIE!, observa-se o crescimento do acesso à educação. Esse crescimento traz consigo novas necessidades que levam a uma mudança no paradigma de aprendizagem. Percebemos uma tendência à descentralização do processo de ensino e aprendizado, antes vinculado às aulas presenciais. Isso vem ocorrendo através de um processo mediado pela tecnologia, que busca permitir o

aprendizado a partir de **qualquer lugar, a qualquer momento e para qualquer pessoa, AAAL**⁵ (BITTENCOURT, 2008; 2009).

Esse paradigma tem como uma de suas características, o aumento da oferta de cursos online, mediados por plataformas/ambientes cuja finalidade é auxiliar professores e alunos na gestão do ensino e da aprendizagem. Ambientes como Moodle⁶, Teleduc⁷, Blackboard⁸, Sakai⁹ e, mais recentemente, os MOOCs¹⁰ como EDX¹¹, Coursera¹² e Udacity¹³, são exemplos desses tipos de ambiente.

Tanto os ambientes tradicionais (Moodle, Teleduc, Blackboard e Sakai) como os MOOCs (EDX, Coursera e Udacity) oferecem cursos online. A grande diferença entre eles é que os últimos focam na capacidade de prover interatividade em larga escala, através dos materiais de curso como vídeos, exercícios, fóruns, peer assessment¹⁴, dentre outros.

1.2 Problemática

A quantidade e diversidade de dados decorrente das interações dos usuários nos ambientes de aprendizagem, as formas de analisar, processar e encontrar informações úteis nesses dados, e como apresentar esses resultados aos professores, de maneira simples e que os isente de conhecimento técnico sobre os processos envolvidos, são problemas relevantes que identificamos em ambientes online de aprendizagem (DUKE, 2013).

Para que seja possível obter informações relevantes das interações dos usuários que utilizam esses ambientes, e com elas tomar decisões pedagógicas pertinentes, é preciso que o professor seja capaz de analisar os dados armazenados pelo sistema. No entanto a quantidade e diversidade desses dados torna impossível que isso seja realizado manualmente, e em tempo hábil a oferecer a ajuda desejada.

Tomemos como exemplos o curso, ofertado através do ambiente *Coursera*¹⁵, intitulado *Bioelectricity: A Quantitative Approach*¹⁶. Embora nenhum pré-requisito tenha sido

5 Sigla em Inglês para: Anywhere, Anytime, Anyone Learning.

6 Moodle, disponível em: <https://moodle.org/>

7 Teleduc, disponível em: <http://www.teleduc.org.br/>

8 Blackboard, disponível em: <http://www.blackboard.com/>

9 Sakai, disponível em: <http://www.sakaiproject.org/>

10 Sigla em Inglês para: Massive Open Online Courses – Cursos Online Massivos Abertos.

11 EDX, disponível em: <https://www.edx.org/>

12 Coursera, disponível em: <https://www.coursera.org/>

13 Udacity, disponível em: <https://www.udacity.com/>

14 Avaliação por pares, onde a responsabilidade de avaliar um aluno é dada a outro aluno do mesmo curso e de igual nível/série.

15 www.coursera.org

16 Bioeletricidade: Uma Abordagem Quantitativa.

exigido, o curso tinha como foco alunos de ensino superior, uma vez que se tratava de uma disciplina ofertada pela Universidade de Duke.

Figura 1 - Número de alunos por metas do sistema¹⁷.

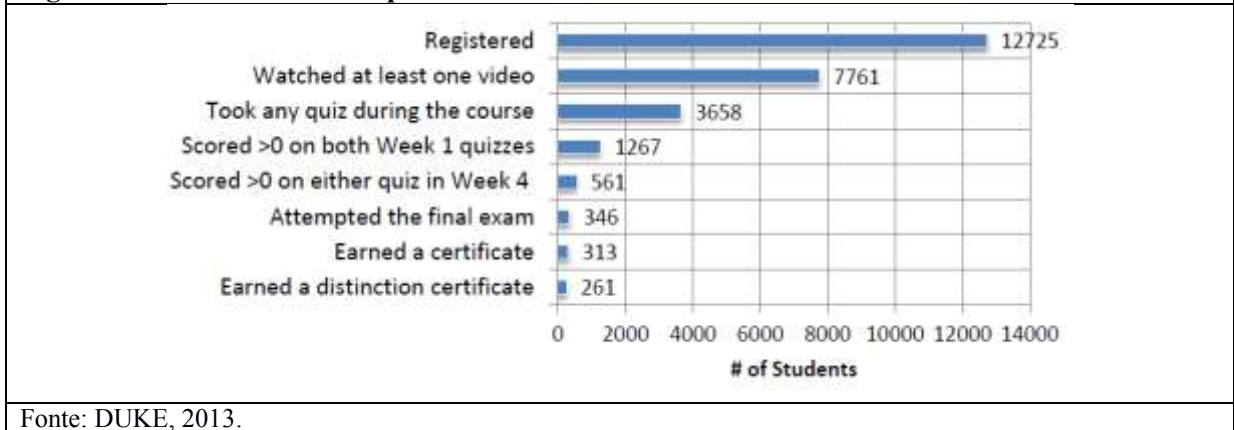
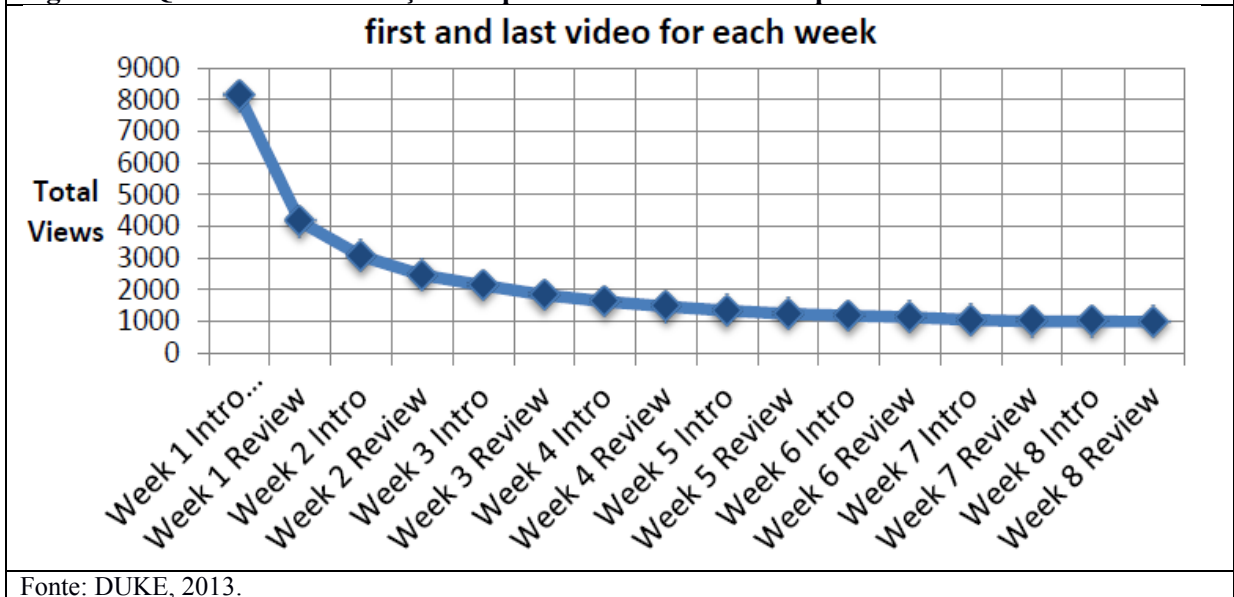


Figura 2 - Quantidade de exibições do primeiro e último vídeos para cada semana do curso¹⁸.



Após o final do mencionado curso, seus responsáveis publicaram um relatório exibindo detalhes sobre o planejamento, execução e conclusão do mesmo (DUKE, 2013). Com base nesse relatório e observando as figuras 1 e 2, alguns dados nos chamaram a atenção, são eles: **12725** alunos se registrados no curso, **156000** exibições de vídeo foram realizadas, sendo o vídeo introdutório do curso responsável por **8000** dessas exibições. Mais de **800** alunos distintos realizaram alguma postagem no fórum, com mais de **550** mensagens

17 Tradução das variáveis do gráfico (Eixo Y): alunos registrados, que assistiram pelo menos a um vídeo, realizaram pelo menos uma prova durante o curso, pontuação superior a zero em ambas as provas da primeira semana, pontuação superior a zero em qualquer das provas da quarta semana, tentou realizar o exame final, recebeu o certificado, recebeu o certificado com distinção. (Eixo X): número de alunos

18 Tradução das variáveis do gráfico (Eixo Y): total de visualizações. (Eixo X): introdução semana X e revisão semana x.

já na primeira semana. **3658** alunos realizaram pelo menos uma das provas, sendo que **3200** deles realizaram a prova da primeira semana do curso. Aproximadamente **700** alunos conseguiram nota máxima nas avaliações da primeira semana.

O que podemos observar é que apesar da centena de milhares de visualizações das vídeo-aulas, apenas 3658 alunos (i.e. 28.29% dos alunos) fizeram apenas uma das atividades avaliativas, conforme pode ser visto na figura 1.

Frisa-se ainda que apenas 6.28% dos alunos realizou alguma postagem nos fóruns. Entretanto, estes dados só foram apresentados ao final do curso, e durante o curso, torna-se complexo e bastante custoso ao professor analisar os dados interacionais de 12725 alunos e compreender o desempenho de sua turma. Com base nesses números, e nas situações mencionadas, surge a nossa primeira questão.

Q1: Como processar esses dados para extrair deles informações relevantes?

Ao final do processamento desses dados, provenientes de contexto educacional, espera-se obter informações novas e úteis, que possam ser usadas pelos professores e/ou tutores responsáveis pelo curso, para aprimorar o aprendizado dos alunos. Como exemplo, este processamento de dados poderia comparar se existe alguma relação entre as postagens nos fóruns e a nota alcançada pelo aluno.

Entretanto, segundo (GARCIA, 2009), obter os resultados desse tipo de processamento é uma tarefa bastante árdua para os usuários. Este fato se dá devido à quantidade e diversidade desses dados que os professores precisam analisar para obter informações úteis. Porém, é fundamental que haja este tipo de análise de dados, pois do outro lado temos os alunos, cada um com diferentes necessidades e níveis educacionais distintos, que precisam, com frequência, de ajuda para progredir em matérias ou cursos.

Levando isto em consideração, é essencial que haja algum nível de automatização de modo a processar tais dados interacionais. Não obstante, simplesmente apresentar os resultados do processamento desses dados para o professor não representa a efetivação de uma ação pedagógica deste professor de modo a melhorar a experiência de aprendizagem dos usuários. Tais fatos dão origem à nossa segunda questão.

Q2: O que fazer com o resultado do processamento dos dados educacionais?

Observando a situação dos professores, vemos que sua rotina diária de trabalho é dificultada devido à sobrecarga de atividades e ao acúmulo de funções em decorrência de reformas educacionais (OLIVEIRA, 2004). Grande parte do tempo de um professor é utilizada na gestão do aprendizado de suas turmas e de seus alunos. Isso é feito coletando dados que o auxiliam na avaliação e na tomada de decisões pedagógicas, e ocorre tanto em modelos tradicionais de ensino (professor e alunos em uma sala de aula física), como em modelos virtuais (baseados em vídeo aulas e baseados em aulas na web). Porém, mesmo com a automatização e apresentação dos dados processados, ainda é uma tarefa complexa para o professor decidir qual a melhor (mais apropriada) ação pedagógica a tomar.

Retomemos o exemplo do sobre o curso de Bioeletricidade no *Coursera*, onde 12412 alunos não conseguiram finalizar o curso. Mesmo que o ambiente de aprendizagem possua alguma ferramenta para identificar e apresentar os problemas de cada um dos alunos, o professor não conseguirá, em tempo hábil, analisar essas informações e tomar uma decisão que atenda as necessidades dessa quantidade de alunos.

Por esta razão há a necessidade de outro suporte computacional para auxiliar o professor neste processo de tomada de decisão pedagógica. Isto pode ser feito utilizando técnicas da mineração de dados com o objetivo de melhorar/aprimorar os processos de ensino e da aprendizagem, com base nos dados provenientes dos ambientes educacionais (ROMERO, 2011) e propondo recomendações pedagógicas sobre como o professor pode atuar diante dos cenários existentes.

Uma vez que as recomendações para ajudar os alunos partem dos professores, devemos considerar uma maneira de realizar o processamento dos dados educacionais (Q1) e tratar os resultados para serem apresentados aos professores (Q2), de forma a não demandar deles qualquer conhecimento sobre os processos envolvidos, ou conhecimento sobre as técnicas de computação utilizadas. Gerando nosso terceiro questionamento.

Q3: Como tornar o processo transparente para os professores?

Para atender a essa demanda precisaremos de um sistema capaz de auxiliar professores a utilizar as informações obtidas do processamento dos dados educacionais, de forma a oferecer suporte à tomada e propagação de decisões pedagógicas, sem com isso requerer conhecimento técnico sobre mineração de dados, ou sobre os detalhes dos processos envolvidos na recomendação pedagógica.

Identificamos que, segundo (RICCI, 2011), os **sistemas de recomendação** são capazes de realizar essas tarefas e se tornaram formas valiosas, populares e poderosas de aborda-los. Sendo assim, esses foram o sistema de escolha para responder à nossa terceira questão de pesquisa.

Esses problemas, e as questões levantadas a partir deles, guiaram a nossa busca por uma forma de auxiliar o processo de aprendizagem dentro do contexto da computação na educação.

Apesar de existirem trabalhos relacionados que, por um lado, aplicaram técnicas de mineração de dados para a detecção de padrões educacionais e, por outro lado, propõem a utilização de sistemas de recomendação para auxiliar os professores em processos de tomada de decisão, a literatura da área ainda carece de trabalhos que tratam desse tipo de sistema para a educação online, aplicando técnicas para a recomendação pedagógica e de mineração de dados baseadas em cenários.

Na seção seguinte definiremos, de forma mais detalhada, os objetivos dessa dissertação.

1.3 Objetivos

O objetivo geral desta dissertação é a criação de uma ferramenta para a recomendação pedagógica que auxilie professores e tutores, de cursos mediados por computador, a prover ajuda a seus alunos na forma de recomendações pedagógicas personalizadas. Essas recomendações serão obtidas minerando-se os dados educacionais oriundos das interações dos alunos no ambiente educacional onde o sistema será utilizado, com a posterior aplicação de questionários onde especialistas no domínio pedagógico, participam diretamente do processo informando qual decisão pedagógica tomar para os cenários criados e explorados.

Nossos objetivos específicos são:

- Fazer o levantamento do estado da arte sobre mineração de dados educacionais e sistemas de recomendação;
- Implementar componentes que reutilizam algoritmos de mineração de dados educacionais.
- Especificar cenários de mineração de dados educacionais, os quais serão encapsulados em componentes distintos.

- Realizar recomendações personalizadas, com base na situação pedagógica dos alunos.
- Construir uma ferramenta de recomendação a qual possibilite que professores e tutores utilizem a mineração de dados educacionais, com escopo predeterminado, e os permita fornecer recomendação pedagógica, sem que para isso necessitem de conhecimento sobre essas técnicas que foram empregadas no processo.
- Avaliar a ferramenta através de um estudo de caso real.

1.4 Contribuições do Trabalho

O presente trabalho contribui com a comunidade de Informática na Educação, i) com um processo de recomendação pedagógica baseada em mineração de dados educacionais. Além disso, para tornar este processo transparente para os usuários finais (professores e tutores), ii) um sistema de recomendação foi desenvolvido, possibilitando professores e/ou tutores a gerar recomendações úteis e relevantes para os alunos, na forma de recomendações pedagógicas personalizadas.

1.5 Organização da Dissertação

Essa dissertação encontra-se organizada da seguinte maneira: no **Capítulo 2** apresentamos a fundamentação teórica que aborda o conhecimento utilizado nessa dissertação, seguido do **Capítulo 3** que contém os trabalhos relacionados. A proposta da dissertação, que trata do processo de recomendação pedagógica e do sistema de recomendação, encontra-se no **Capítulo 4**. No **Capítulo 5** exibimos um estudo de caso, mostrando a aplicação do sistema proposto em um cenário real, e uma discussão acerca dos resultados alcançados. Finalmente, o **Capítulo 6** apresenta as conclusões, seguido das referências e apêndices deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo abordaremos os principais tópicos acerca dos fundamentos teóricos utilizados nessa dissertação, e necessários para uma apropriada compreensão da mesma. Nesse sentido, distribuimos as seções da forma como segue: na **seção 2.1**, será apresentada uma introdução à mineração de dados, seguida de mineração no contexto da educação, conhecido como mineração de dados educacionais, introduzida na **seção 2.2**. O processo de mineração dos dados CRISP-DM, será tratado na **seção 2.3** e algoritmos para mineração destes dados, na **seção 2.4**. Logo após, abordaremos os conceitos acerca dos sistemas de recomendação, na **seção 2.5**.

2.1 Introdução à Mineração de Dados

Com o barateamento das tecnologias para armazenamento e transmitir dados, a informatização de diversos processos, substituindo o papel por documentos digitais, e com a constante necessidade de se armazenar, manipular e apresentar esses dados e informações, estamos cada vez mais consumindo, guardando e propagando dados das diversas áreas do conhecimento. Aquilo que há alguns anos seria eliminado por falta de espaço físico em discos rígidos, ou pelo alto custo de meios digitais de armazenamento¹⁹, hoje é armazenado para uso futuro (DEOGUN, 1998). É possível, ainda, que esses dados nos acompanhem até onde estivermos graças à *World Wide Web* e aos dispositivos que utilizam memória flash. Por isso, quando há a necessidade de extrairmos informações desses dados, precisamos encontrar formas de **analisar, classificar, sumarizar, descobrir e caracterizar tendências**, bem como **detectar anomalias** nesses dados, de forma **automática e mais precisa possível**. Uma das áreas da computação que está em consonância com tais objetivos e necessidades é a área de **mineração de dados**.

As pesquisas em mineração de dados preocupam-se com essas questões criando e aprimorando técnicas destinadas a aborda-las, ou seja, **obter de maneira automática e precisa, informações relevantes de grande quantidade de dados**. Isso é feito por meio da busca automatizada ou, pelo menos, aprimorada com o uso de sistemas computacionais e da descoberta de padrões com grande significância, nesses dados (dados de treinamento),

¹⁹ Seja em relação ao custo de tempo para realizar a transferência dos dados para esses meios de armazenamento, ou ao custo de aquisição dos mesmos ou, ainda, ao custo de se arquivar apropriadamente.

permitindo que se façam descobertas não triviais em um conjunto novo de dados (dados de teste) (WITTEN, 2011).

A mineração de dados surgiu por volta de 1990 (ZAIANE, 1999) e evoluiu bastante em teoria e aplicações. De uma forma geral, seu objetivo é extrair de grandes conjuntos de dados, padrões que levem à descoberta de informações relevantes para um determinado contexto. Essa descoberta é feita pela criação de aplicação baseadas em algoritmos para essa finalidade (por exemplo: **K-means** para identificar agrupamentos de dados com características semelhantes), idealmente, devem manter um determinado limite de especificação de maneira que possam ser aproveitadas para analisar novos conjuntos de dados, mantendo uma boa capacidade e qualidade de avaliação (i.e. uma função com capacidade de generalização).

Para cada tipo de padrão que se deseja encontrar em um repositório de dados, pode haver uma, ou mais, tarefas de mineração de dados as quais são classificadas como: **descritivas** e **preditivas**. Segundo (HAN, 2011) as tarefas descritivas caracterizam as propriedades gerais dos dados no repositório. Já as preditivas realizam inferências sobre esses dados, com o objetivo de fazer previsões.

O conjunto de funcionalidades de mineração de dados, listadas a seguir, é o resultado da junção do conteúdo sobre o tema, da forma com é abordado em (WITTEN, 2011), (WU, 2007) e (HAN, 2011):

- **Caracterização e Discriminação:** sumariza conceitos e classes de forma concisa, mas precisa.
- **Mineração de Padrões Frequentes, Associações e Correlações:** detecta itens que ocorram com frequência em um conjunto de dados, bem como verificar se existem associações e correlações entre itens.
- **Classificação e Previsão:** gera modelos capazes de distinguir classes e conceitos.
- **Análise de Agrupamentos:** agrupa itens semelhantes cuja classe não é conhecida.
- **Análise de *Outliers*:** analisa os dados considerados "anormais" e detectar padrões por trás do surgimento desses dados.
- **Análise de Evolução:** analisa a forma como os dados evoluem, em busca de padrões relevantes nesse processo.

As tarefas de mineração são executadas pela implementação das técnicas de mineração, podendo existir diversas maneiras de se implementar tarefas de uma mesma funcionalidade.

2.2 Mineração de Dados Educacionais

Desde meados de 2005 tem-se notado o aumento do interesse e da aplicação da mineração de dados na educação (ROMERO, 2007). Abaixo, na tabela 1, listamos 10 áreas onde técnicas de mineração de dados são aplicadas. Essas áreas correspondem às mais votadas na enquete: **Onde você aplicou Técnicas Analíticas/Mineração de Dados em 2012?**^{20 21}.

Tabela 1 - Resultado da enquete “Onde você aplicou técnicas analíticas/mineração de dados em 2012?”

Posição	Área de Aplicação	Percentual de Votos
1º	Relacionamento com Clientes	28.6%
2º	Saúde	16.3%
3º	Varejo	14.8%
4º	Operações Bancárias	14.3%
5º	Educação	14.3%
6º	Propaganda	13.3%
7º	Deteção de Fraudes	12.8%
8º	Mídias Sociais / Redes Sociais	12.2%
9º	Ciências	11.7%
10º	Finanças	10.2%

Fonte: <http://www.kdnuggets.com/polls/2012/where-applied-analytics-data-mining.html>.

Notamos o destaque no uso da mineração de dados na educação, que é o contexto de interesse nessa dissertação. Notamos que seu percentual de votos está próximo ao de áreas onde a aplicação de técnicas de mineração de dados já é tradicional, como saúde, varejo e operações bancárias.

Isto se deve, em parte, pelo recente aumento da oferta de cursos a distância mediados por ambientes com suporte computacional, onde pesquisadores vêm adaptando métodos provenientes da mineração de dados (BAKER, 2010) e os aplicando à grande quantidade de

²⁰ Em Inglês: Where did you apply Analytics/Data Mining in 2012?

²¹ Realizada pelo site KDNuggets entre 26 de Novembro e 11 de Dezembro de 2012, (KDD NUGGETS, 2012).

dados oriundos desses ambientes, dando origem à uma nova área de pesquisa, a **Mineração de Dados Educacionais - EDM**²².

A EDM utiliza conceitos e técnicas da mineração de dados tradicional com o objetivo de **aprimorar os processos de ensino e da aprendizagem**, bem como a **gestão desses processos**, com base em **dados de meios educacionais**. Dados desse contexto podem ser provenientes de sistemas tutores inteligentes, sistemas educacionais clássicos baseados em computação, dados administrativos da própria escola, testes padronizados, dentre outros (ROMERO, 2011).

De acordo com (ROMERO, 2011), a EDM vem sendo utilizada para a obtenção de diversos objetivos educacionais. Essas tarefas podem ser agrupadas nas seguintes categorias:

- **Comunicação com *stakeholders***²³: visa prover auxílio a administradores de cursos e professores para avaliar as atividades realizadas pelos alunos, bem como a participação dos mesmos no curso. Mineração de processos, geração de relatórios, visualização de dados e a análise estatística de dados, são as técnicas mais utilizadas para esse grupo de aplicações.
- **Realizar melhorias e manutenções em cursos**: tem como objetivo ajudar gestores de curso e educadores sobre quais estratégias utilizar para obter melhorias. Associação, agrupamento e classificação, são as técnicas mais usuais nesse grupo.
- **Gerar recomendações**: objetiva recomendar conteúdo apropriado para o momento educacional vivenciado pelo estudante. Segundo (BRITO, 2012) as recomendações devem atender às necessidades dos alunos e levar em consideração seu nível de conhecimento. Associação, sequenciação, agrupamento e classificação, são as técnicas mais utilizadas nesse grupo.
- **Prever resultados de atividades/provas ou de avaliações de aprendizado**: busca antever o resultado de testes e de outras avaliações educacionais, com base na análise das atividades realizadas pelos estudantes. Mais uma vez, as técnicas de associação, agrupamento e classificação, são as mais utilizadas.
- **Criar modelos de alunos**: o objetivo é estudar determinadas características dos alunos. Para esse tipo de aplicação, há uma demanda maior e as técnicas mais

22 Em Inglês: Educational Data Mining - EDM

23 Stakeholders, nesse contexto, se refere a pessoas diretamente interessadas no ambiente educacional e no desempenho dos alunos.

utilizadas são: Análises estatísticas, redes Bayesianas, modelos psicométricos e aprendizado por reforço.

- **Análise da estrutura do domínio:** busca avaliar a estrutura do domínio (por exemplo: do ambiente de aprendizado), analisando seu desempenho, ou seja, como o domínio realiza uma determinada tarefa (por exemplo: quão eficiente é a detecção de desistentes do ambiente avaliado?). As técnicas mais utilizadas incluem regras de associação, métodos de agrupamento e algoritmos de busca.

2.3 O Processo de Mineração dos Dados

O processo de mineração de dados pode ser dividido em três partes: pré-processamento dos dados, mineração dos dados e pós-processamento dos dados. Esse processo é, muitas vezes, referenciado como Descoberta de Conhecimento a partir de Dados²⁴ (HAN, 2011)²⁵.

Por ser um processo com considerável complexidade, existem algumas iniciativas para padronizar os passos a serem seguidos para realizar a mineração de dados, com estratégias orientadas à solução de problemas. Com essa finalidade, criou-se o **CRISP-DM**²⁶, uma solução não proprietária, e livremente disponível para organizar o ciclo de vida de aplicação de mineração de dados em um projeto.

CRISP-DM é um processo constituído de seis fases, em uma sequência adaptativa, onde as saídas de uma fase são associadas à fase seguinte (NISBET, 2009). Listamos abaixo as fases que constituem esse processo, e na figura 3 demonstramos uma imagem desse ciclo e a relação entre suas fases.

1. **Fase de compreensão do negócio:** nessa fase faz-se um estudo dos objetivos e demandas do projeto, do ponto de vista do ambiente no qual ele está inserido. O intuito é definir um problema de mineração de dados e traçar, em linhas gerais, um plano de ação.
2. **Fase de compreensão dos dados:** nessa fase busca-se realizar uma primeira análise dos dados, familiarizando-se com algumas de suas qualidades e, ao mesmo tempo, formulando possíveis hipóteses para extrair deles informações relevantes.

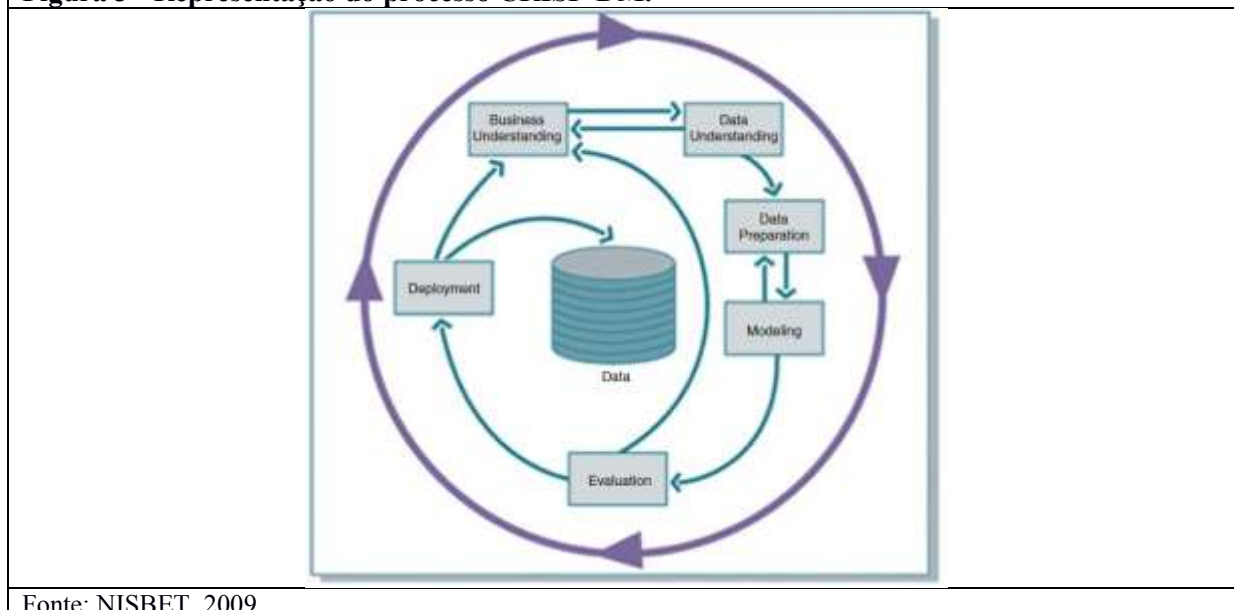
24 Em Inglês: Knowledge Discovery from Data - KDD

25 Nessa dissertação utilizaremos o termo mineração de dados nesse sentido amplo, ou seja, como sinônimo do Processo de Mineração de Dados.

26 Em Inglês: Cross-Industry Standard Process for Data Mining.

3. **Fase de preparação dos dados:** essa fase tem por objetivo tratar os dados, através de tarefas de pré-processamento, de forma que o resultado seja o *conjunto final de dados*, o qual alimentará a ferramenta de modelagem.
4. **Fase de modelagem:** nessa fase selecionam-se as técnicas de modelagem mais apropriadas para o problema e para os dados analisados nas fases anteriores.
5. **Fase de avaliação:** a meta dessa fase é analisar se o modelo criado é capaz de atingir os objetivos pretendidos. Busca-se determinar se algum ponto importante para o problema de mineração não foi considerado de forma apropriada.
6. **Fase de implantação:** nessa fase organiza-se o conhecimento obtido para que o mesmo seja disponibilizado de uma forma que o usuário final possa utilizá-lo/consultá-lo.

Figura 3 - Representação do processo CRISP-DM.



Fonte: NISBET, 2009.

2.4. Os Algoritmos de Mineração de Dados

Nessa seção abordaremos os algoritmos que utilizaremos para o processo de mineração dos dados, avaliando suas capacidades, mas também suas limitações.

2.4.1. OS 10 ALGORITMOS DE MINERAÇÃO DE DADOS MAIS CONHECIDOS

Tomamos como base o artigo *Top 10 Algorithms in Data Mining*²⁷ (WU, 2007), onde 10 algoritmos de mineração de dados são apresentados, considerando-se suas indicações de

²⁷ Os 10 Melhores Algoritmos de Mineração de Dados.

uso, capacidades e limitações. Os autores do referido artigo, em um primeiro passo, convidaram pesquisadores, vencedores dos prêmios *ACM KDD Innovation Award*²⁸ e *IEEE ICDM Research Contributions Award*²⁹ de 2006, a listar até 10 dos "mais conhecidos" algoritmos de mineração de dados, citando: **(a)** o nome do algoritmo, **(b)** uma breve justificativa a respeito da escolha e **(c)** publicações representativas a respeito do algoritmo.

Como segundo passo, os autores escolheram, dentre os algoritmos citados pelos pesquisadores no passo acima, aqueles que continham mais de 50 citações no Google Scholar, chegando a uma lista de 18 algoritmos candidatos.

Por fim os pesquisadores foram solicitados a votar em 10 algoritmos da lista de candidatos, resultando nos algoritmos presentes no artigo supracitado. Na tabela 2 descrevemos, de forma breve, os 10 algoritmos.

Tabela 2 - Os 10 algoritmos de mineração mais utilizados

Continua.

Algoritmo	Descrição
C4.5	C4.5 é um indutor de árvores de decisão, que utiliza fórmulas baseada na teoria da informação para avaliar o quão bom é um teste. Para um conjunto de casos, escolhe-se o teste do qual se obtém o maior ganho de informação para a construção da árvore.
K-Means	O algoritmo K-Means define grupos para as instâncias de dados, baseado no número de instâncias centrais (centróides, definido pelo usuário) e pela distância euclidiana das instâncias em relação às instâncias centrais.

28 Disponível em: http://www.sigkdd.org/awards_innovation.php

29 Disponível em: <http://www.cs.uvm.edu/~icdm/>

Tabela 2 - Os 10 algoritmos de mineração mais utilizados.

Continuação

Algoritmo	Descrição
Support Vector Machines	As support vector machines têm por objetivo inferir a melhor função de classificação que gera um hiperplano capaz de separar as diferentes classes de dados. O algoritmo garante que a função é de fato a melhor, maximizando a margem que separa as classes.
Apriori	Através da metodologia "gerar e testar", o algoritmo Apriori cria conjuntos de itens mais longos, a partir de conjuntos mais curtos, mas já conhecidos, que sofrem incrementos sucessivos. Cada novo conjunto tem sua frequência testada e comparada com um valor de suporte mínimo.
Expectation-Maximization	O algoritmo EM, assim como o K-Means, define grupos de instâncias de dados, entretanto a escolha do número de centróides é feita pelo próprio algoritmo através de um processo de cálculo da "probabilidade de agrupamento" utilizada para reestimar os parâmetros de busca dos centróides. Esse processo é repetido até a convergência.
Page Rank	O Page Rank é um algoritmo para classificar, no sentido de atribuir uma ordem de importância, buscas por páginas da web. O algoritmo considera os hiperlinks de uma página X apontando para uma página Y como um voto (de X em Y). Quanto mais votos uma página web possui, mais alta é a sua classificação. Entretanto ele, também, considera um fator "prestígio" de cada voto, ou seja, votos de páginas bem classificadas têm "peso" superior ao de outras numa classificação inferior.
AdaBoost	AdaBoost utiliza o método de Aprendizado em Conjunto criando, inicialmente, um modelo básico onde todos os exemplos de teste recebem o mesmo peso. Em seguida outros modelos são criados e os pesos dos exemplos vão sendo modificados. Após um determinado número de repetições do processo, seleciona-se o melhor dentre os modelos gerados.
K-Nearest Neighbor	Dados um conjunto de treinamento e um conjunto de teste, o algoritmo KNN procura grupos de dados de treinamento que estão "mais próximos" ao objeto sendo avaliado. A classe dada é aquela com maior predominância.

Tabela 2 - Os 10 algoritmos de mineração mais utilizados.

Conclusão

Algoritmo	Descrição
Naive Bayes	O objetivo do algoritmo Naive Bayes é inferir um classificador com base em dados de treinamento. Isso é feito atribuindo uma faixa de escores a cada uma das classes conhecidas. A classificação ocorre avaliando-se a probabilidade de um objeto de teste pertencer a uma classe dado o número de objetos dessa classe (probabilidade prévia), multiplicada pelo grau de similaridade do objeto com a classe avaliada.
Classification and Regression Tree	O algoritmo CART é um indutor de árvores de decisão que utiliza um método de poda mais robusto que o C4.5. O CART gera as árvores podando as sub-árvores que levam ao menor incremento de erro nos dados de treinamento, considerando o tamanho dessa sub-árvore.

Fonte: WU, 2007.

Para atingir os objetivos dessa dissertação definimos que não serão utilizados os seguintes algoritmos dessa lista: support vector machines, page rank, adaboost, k-nearest neighbor e naive bayes.

Para os algoritmos restantes acreditamos ser sensato conhecer não apenas sua descrição de funcionamento, mas, também, suas principais capacidades e limitações as quais se encontram sumarizadas, de acordo com (WU, 2007; WITTEN, 2011; LAROSE, 2005), na tabela a seguir (tabela 3).

Tabela 3 - Capacidades e limitações dos algoritmos escolhidos.

Continua.

Algoritmo	Capacidades	Limitações
C4.5	<ul style="list-style-type: none"> • Gerar classificadores na forma de árvores de decisão • Gerar classificadores na forma de conjunto de regras 	<ul style="list-style-type: none"> • Tempo elevado para processamento. • Grande uso de memória • Bastante sensível a <i>outliers</i>.
K-Means	<ul style="list-style-type: none"> • Particionar um conjunto de dados em agrupamentos de acordo com as características dos dados 	<ul style="list-style-type: none"> • Custo de ajuste para se definir a quantidade ótima de clusters • O desempenho é sensível à seleção de inicialização

Tabela 3 - Capacidades e limitações dos algoritmos escolhidos.

		Conclusão.
Apriori	<ul style="list-style-type: none"> • Descobrir itens frequentes em conjuntos • Gerar regras de associação 	<ul style="list-style-type: none"> • Não há garantia de se encontrar todos os itens frequentes • Deve-se escolher entre precisão ou eficiência
Expectation-Maximization	<ul style="list-style-type: none"> • Encontrar o melhor valor para o agrupamento de dados 	<ul style="list-style-type: none"> • Valores ausentes (em quantidade moderada) e alta dimensionalidade dos dados diminuem a velocidade de convergência.
Classification and Regression Tree	<ul style="list-style-type: none"> • Processar atributos contínuos e nominais • Balancear classes • Tratar valores ausentes • Aprendizado sensível ao custo • Estimar árvores de probabilidade 	<ul style="list-style-type: none"> • Maior custo computacional que os demais indutores de árvores de decisão

Fonte: WU, 2007.

2.5 Sistemas de Recomendação

Sistemas de recomendação podem ser definidos como:

"Ferramentas de software e técnicas cujo objetivo é prover sugestões de itens a serem utilizados por um usuário. As sugestões providas focam em ajudar os usuários em diversos processos de tomada de decisão" (RICCI, 2011).

Tradicionalmente as pessoas solicitam recomendações sobre os mais diversos assuntos, de diversas áreas, usando várias estratégias para consegui-las: através de conversas informais, consultas com pessoas de confiança, consultas com especialistas, pesquisas na internet, através de métodos oriundos da teoria da decisão (modo bastante formal), seguindo os próprios "instintos" ou, simplesmente, fazendo o que a maioria das pessoas está fazendo (JANNACH, 2011).

O processo de busca por bons conselhos e recomendações é um processo bastante custoso, em termos financeiros e de tempo, e nem sempre é possível medir sua qualidade. Por esse motivo, de acordo com (RICCI, 2011 e JANNACH, 2011), no início dos anos 90, surgiram os primeiros sistemas cuja natureza era a de agregar as opiniões de milhões de pessoas online com o objetivo de ajudá-los a encontrar e recomendar conteúdo útil e

interessante. Tais sistemas encontraram aceitação e evoluíram em técnica e passaram a contemplar diversas áreas.

Nessa dissertação utilizaremos os as técnicas de recomendação para a construção de uma ferramenta de apoio na decisão para professores e tutores de cursos mediados por ambiente de aprendizagem online.

Contudo, para que se consigam recomendações de qualidade é preciso escolher a técnica apropriada dada o contexto, e a finalidade, de utilização do sistema. Atualmente os sistemas de recomendação utilizam, essencialmente, as seguintes técnicas (RICCI, 2011):

1. **Recomendações baseadas em filtros colaborativos:** os sistemas de recomendação desse tipo levam em consideração as similaridades nos interesses dos usuários, ou seja, se usuários apresentaram o mesmo interesse no passado, é possível que isso se repita no futuro. Por exemplo, esses sistemas recomendam itens³⁰ que um usuário, com interesses semelhantes, interagiu ou recomendou.
2. **Recomendações baseadas em conteúdo:** sistemas de recomendação desse tipo utilizam as informações sobre os itens de recomendação (descrições, categorias, características, etc.), para propor novas recomendações, podendo organizar os itens recomendados de acordo com a importância de cada uma dessas informações. Por exemplo, alunos com problema no assunto "**Los Pronombres**" em um curso online de Espanhol, receberão recomendações de vídeos, textos, exercícios, dentre outros recursos educacionais (RE), sobre esse mesmo assunto. Pode-se ainda considerar outras características como dificuldade de um recurso específico (por exemplo, uma questão de múltipla escolha).
3. **Recomendações baseadas em conhecimento:** sistemas de recomendação desse tipo não apenas consideram a necessidade dos usuários finais em relação a características de um determinado objeto de aprendizagem, mas levam em consideração o modelo do usuário e dos RE, consultando e elicitando de tal usuário a importância relativa de todos os atributos desses recursos. É esse o tipo de sistema de recomendação no qual nos basearemos, para a criação da nossa ferramenta, uma vez que desejamos oferecer recomendações de acordo com a situação pedagógica do aluno. Por exemplo, um aluno pode estar apresentando um desempenho ruim por diversos motivos. O sistema

³⁰ Itens são os objetos da recomendação. Para sistemas comerciais podem ser produtos, mas no contexto dessa dissertação são recomendações pedagógicas (constituídas por recursos educacionais) que objetivam auxiliar os alunos a superar uma dificuldade ou melhorar sua experiência de aprendizagem.

deve poder obter conhecimento sobre esse contexto e recomendar de acordo com as prioridades do aluno.

4. **Recomendações baseadas em abordagens híbridas:** é possível, entretanto, que apenas uma dessas abordagens não seja suficiente para oferecer recomendações de forma satisfatória. Se esse for o caso, é possível criar sistemas de natureza híbrida que, como o próprio nome indica, fazem uso de mais de uma das técnicas apresentadas anteriormente. Como exemplo poderíamos ter um sistema que recomenda RE com base em suas características (conteúdo) e, também, considerando avaliações de outros alunos ou por similaridade de preferências/dificuldades (filtro colaborativo).

Existem outras técnicas associadas a outros temas abordados na área de sistemas de recomendação, que serão apenas mencionadas, brevemente, a seguir. São elas: **explicar as recomendações**, que visa detalhar os motivos de uma determinada recomendação ter sido feita (algo que mostra a relevância da recomendação para os usuários finais). **Avaliar recomendações**, que aplica métricas para analisar o nível de sucesso das recomendações, em um determinado contexto, dadas as reações dos usuários. Há, ainda, técnicas para **visualização, segurança e recomendações em áreas multidisciplinares**.

3 TRABALHOS RELACIONADOS

Neste capítulo trazemos alguns trabalhos que consideramos relacionados às ideias apresentadas nessa dissertação. Pesquisamos por aqueles que abordassem sistemas/ferramentas de recomendação, juntamente com o emprego de técnicas capazes de promover auxílio a professores e alunos. O objetivo foi encontrar ferramentas com a capacidade de descobrir conhecimento útil a partir de grandes quantidades de dados provenientes das interações dos usuários de ambientes educacionais baseados na web. Fomos capazes de encontrar trabalhos de grande significância, em cada uma das áreas já mencionadas, para realizar uma análise comparativa com o que apresentamos nessa dissertação.

Os trabalhos são apresentados por ordem decrescente de relação com a proposta dessa dissertação e, no final desse capítulo, criamos uma tabela (tabela 4) para sumarizar as contribuições de cada trabalho e compará-los com o sistema proposto nesta dissertação.

3.1 Collaborative Recommender System Using Distributed Rule Mining for Improving Web-based Adaptive Courses

O trabalho, realizado por pesquisadores da *University of Córdoba*³¹, descreve um sistema que auxilia os professores, de cursos baseados na web, a realizar melhorias nestes, através da mineração de regras que são utilizadas para a geração de recomendações. O sistema, ainda, promove a colaboração, pontuando e compartilhando entre outros professores e especialistas no domínio educacional, as recomendações obtidas (GARCIA, 2009).

O sistema proposto pode ser definido como um “recomendador” colaborativo que permite que professores, e especialistas no domínio educacional, troquem entre si informações sobre padrões e experiências sobre a forma como seus alunos aprendem. O sistema personaliza as recomendações geradas, com base nas características das turmas de cada professor. Tal capacidade é decorrente do uso de **técnicas de mineração de dados**, com o objetivo de **adaptar o curso ao perfil de seus usuários**.

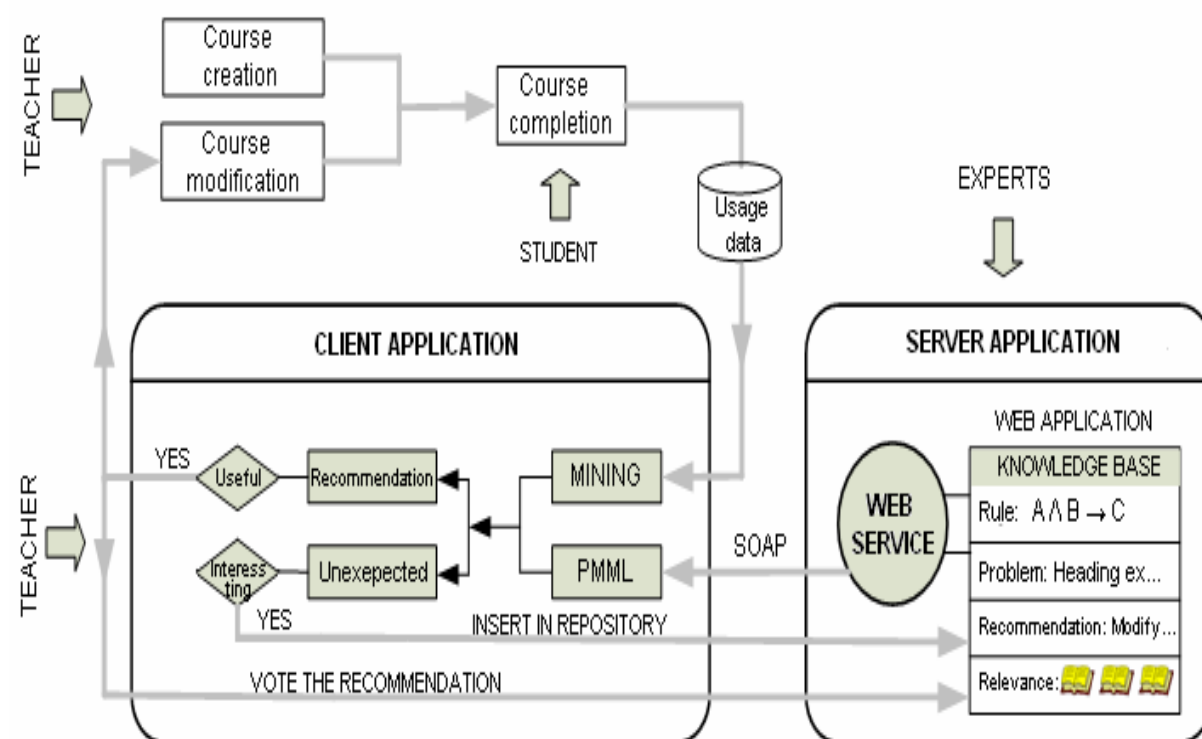
O sistema é baseado em uma arquitetura cliente-servidor, conforme pode ser observado na figura 4. Os dados de interesse são aqueles resultantes das interações dos alunos dentro do sistema³², os quais serão minerados utilizando-se um algoritmo de mineração de

31 Universidade de Córdoba, disponível em: <http://www.uco.es/internacional/>

32 Usage Data: dados (logs) de uso do sistema pelos usuários.

regras de associação. Os resultados serão apresentados aos professores na forma de **recomendações do tipo REGRA-PROBLEMA**, cujo propósito é melhorar o curso.

Figura 4 - Arquitetura do sistema colaborativo de recomendação proposto no artigo.



Fonte: GARCIA, 2009

Essas recomendações podem ser compartilhadas com outros professores, e especialistas do domínio, que as classificam com a finalidade de avaliar sua relevância.

O **módulo servidor** é destinado, apenas, para especialistas no domínio que através do qual, gerenciam a base de conhecimento (constituída de regras, problemas, recomendações geradas e suas respectivas relevâncias), podendo, também, avaliar (através de voto) as contribuições de outros especialistas.

O **módulo cliente** é destinado ao professor e permite que ele realize as melhorias no curso (em andamento), através da aplicação das recomendações³³. Esse módulo gera, e compara essas recomendações com as regras criadas e armazenadas na base de conhecimento (parte do módulo servidor), classificando-as como **esperadas**, quando estão de acordo com as regras criadas, ou **inesperadas**, em caso contrário. Os votos alimentam a base de conhecimento e as recomendações que foram classificadas como **inesperadas**, mas aplicadas

³³ A aplicação de uma recomendação caracteriza, automaticamente, um voto a favor de sua utilidade.

pelos professores, ou seja, foram classificadas como **úteis** por eles, são marcadas para avaliação dos especialistas.

Apesar de o sistema apresentado ser focado em ajudar o professor na melhoria do curso, algo que também promove melhorias para os alunos, acreditamos que esse sistema tem capacidade limitada, pois não oferece uma forma de abordar a descoberta e solução das dificuldades pedagógicas dos alunos, o que poderia aprimorar de forma mais direta, a experiência de aprendizagem destes.

3.2 EDUCA

EDUCA é uma Ferramenta de Autoria, cuja finalidade é a criação de material de aprendizado adaptativo, que é disponibilizado aos alunos através de dispositivos móveis (CABADA, 2011). A ferramenta foi desenvolvida através de um esforço colaborativo entre o *Instituto Tecnológico de Culiacán*³⁴, e o *Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)*³⁵, ambos no México.

EDUCA utiliza Mapas Auto Organizáveis³⁶, que consiste em uma técnica de classificação não supervisionada que utiliza Redes Neurais, para identificar o estilo de aprendizagem de cada aluno e, dessa forma, adaptar o conteúdo de acordo com essa informação. A ferramenta, ainda, faz uso de tecnologias web 2.0 para recomendar esses “recursos de aprendizagem”, os quais são minerados da web.

O sistema é constituído, basicamente, de cinco componentes/módulos principais:

- Ferramenta de Autoria (*Authoring Tool*).
- Repositório de Recursos de Aprendizagem (*Learning Resources Repository*).
- Repositório dos Cursos (*Course Repository*).
- Engenho de Entrega Inteligente (*Intelligent Delivery Engine*).
- Engenho de Recomendação (*Recommendation Engine*).

Através da Ferramenta de Autoria, um tutor/instrutor cria o material, que será armazenado tanto no Repositório de Cursos, como no Repositório de Recursos de Aprendizagem. Tal material será posteriormente mantido e atualizado pela comunidade de usuários/aprendizes, para cada um dos cursos. Os cursos são capazes de realizar o

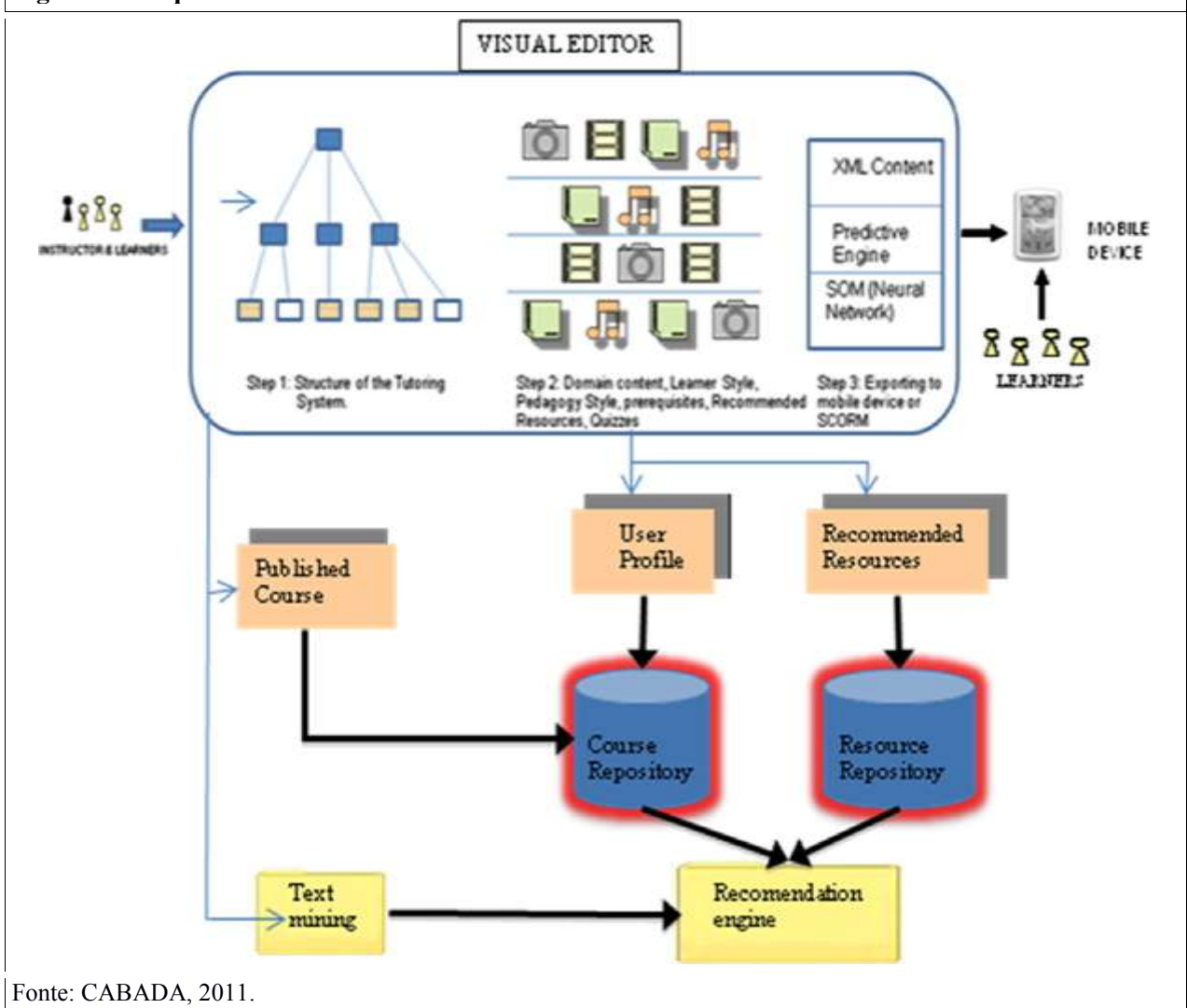
³⁴ Instituto Tecnológico de Culiacán, disponível em: <http://itculiacan.edu.mx/>

³⁵ Instituto Nacional de Astrofísica, Óptica e Eletrônica, disponível em: <http://www.inaoe.mx/>

³⁶ Em Inglês: *Self-Organizing Maps (SOMs)*.

reconhecimento dinâmico das características de aprendizado dos usuários e encaminhar para eles o conteúdo de aprendizagem mais apropriado para seu perfil, e o faz por meio do Engenho de Recomendação. Novos recursos de aprendizagem são minerados da web pelo uso de técnicas de mineração de texto, e serão utilizados em futuras recomendações. Por fim, esse conteúdo é disponibilizado para os dispositivos móveis dos alunos/aprendizes, utilizando o Engenho de Entrega Inteligente. A arquitetura do sistema pode ser visualizada na figura 5.

Figura 5 - Arquitetura do EDUCA.



Fonte: CABADA, 2011.

O EDUCA apresenta a capacidade de realizar recomendações personalizadas com base nas características de aprendizagem dos alunos, ou seja, define para quais alunos um material, inserido no sistema por um professor, é apropriado. Entretanto não se identifica uma forma de contextualizar essas recomendações com base em situações pedagógicas de interesse nem a participação de professores e tutores na geração das recomendações. Por exemplo, não há a possibilidade de recomendar conteúdo para **alunos com dificuldade** em determinado assunto.

3.3 Applying Web Usage Mining for Personalizing Hyperlinks in Web-based Adaptative Educational Systems

A intenção dos autores, pesquisadores da *University of Córdoba*³⁷ e da *Eindhoven University of Technology*³⁸, foi o de criar uma ferramenta para a Mineração de Dados da Web, possível de ser utilizada por professores, auxiliando-os a executar um processo de **Mineração de Uso da Web**³⁹ com a finalidade de recomendar links. Tal ferramenta foi integrada a um sistema de Hipermídia Adaptativa Baseado na Web, de nome **AHA!** Essa integração permitiu que as recomendações resultantes do processo, pudessem ser diretamente aplicadas aos cursos (ROMERO, 2009).

Em termos de funcionalidade, a ferramenta foi concebida para que pudesse ser utilizada de dois modos, conforme exibido na figura 6 e descrito a seguir:

1. **Modo Básico:** nesse modo utilizam-se as informações dos alunos que estão presentes nos logs do sistema web, com um conjunto reduzido dos recursos disponíveis na ferramenta, por exemplo, apenas um algoritmo de mineração é utilizado. É uma forma mais simples e rápida de prover as recomendações, contudo novos alunos são prejudicados, pois não se tem dados suficientes a respeito da interação dos mesmos.
2. **Modo Avançado:** nesse modo utilizam-se informações adicionais dos alunos, além de diversos recursos de mineração de dados e permite que as recomendações sejam personalizadas.

A ferramenta implementada é composta de duas aplicações:

1. **A Ferramenta de Mineração:** responsável por realizar o processo completo mineração dos arquivos de log. Tais arquivos são armazenados no formato XML e durante a mineração, esses arquivos e dados de perfis de alunos, passam por um processo de preparação, onde os logs são agrupados em um único arquivo. Tal arquivo é **minerado em busca de padrões sequenciais**, ou seja, comportamentos que se repetem em diferentes interações, os quais são enviados para o engenho de recomendação. Há ainda o agrupamento dos aprendizes de acordo com seus perfis. O resultado desse agrupamento é, também, enviado para a geração de recomendações. É importante mencionar que essa fase ocorre de modo *offline*.

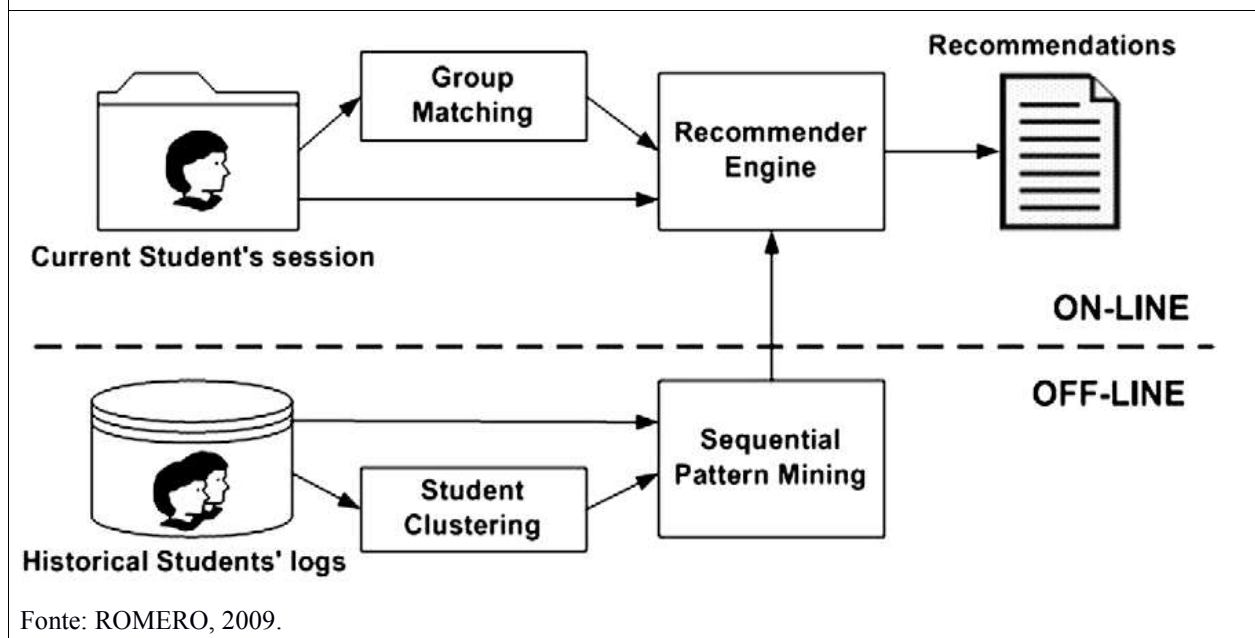
37 Universidade de Córdoba, disponível em: <http://www.uco.es/internacional/>

38 Universidade de Tecnologia de Eindhoven, disponível em: <http://www.tue.nl/en/>

39 Mineração de Uso da Web: também conhecida como mineração de “logs” de acesso, é o processo de extrair padrões interessantes de logs de acesso da web (ROMERO, 2009).

2. **O Engenho de Recomendação:** a recomendação, por outro lado, utiliza o resultado do processo de mineração. Esses dados alimentarão o engenho que, por sua vez irá verificar quais recomendações (lista de links) são **apropriadas considerando as características do aprendiz**. Esse processo ocorre de forma online, e é iniciado logo que o aprendiz estabelece uma sessão com o sistema AHA!

Figura 6 - Arquitetura da ferramenta de mineração e do engenho de recomendação propostos.



A ferramenta permite que usuários sem conhecimento especializado em mineração de dados, façam uso da técnica de **mineração de logs de uso da web**, armazenados pelo sistema, no caso o AHA! para a **recomendação de links**. Contudo essa capacidade de recomendar, para a ferramenta em questão, está atrelada a uma única técnica de mineração dos dados e a uma única fonte de dados, o que pode limitar a capacidade de recomendação da ferramenta.

3.4 PROTUS 2.0

PROTUS 2.0 é um Sistema Tutor para o aprendizado dos princípios básicos da linguagem de programação Java. Foi desenvolvido por pesquisadores da *Higher School of Professional Business Studies*⁴⁰ e pela *Faculty of Sciences*⁴¹ (*Department of Mathematics and Informatics*), ambas na Sérvia. Segundo descrição dos autores o sistema é capaz de promover

⁴⁰ Escola Superior de Estudos em Negócios Profissionais, disponível em: <http://www.vps.ns.ac.rs/en/home.1.61.html>

⁴¹ Faculdade de Ciências (Departamento de Matemática e Informática), disponível em: <http://www.pmf.uns.ac.rs/en>

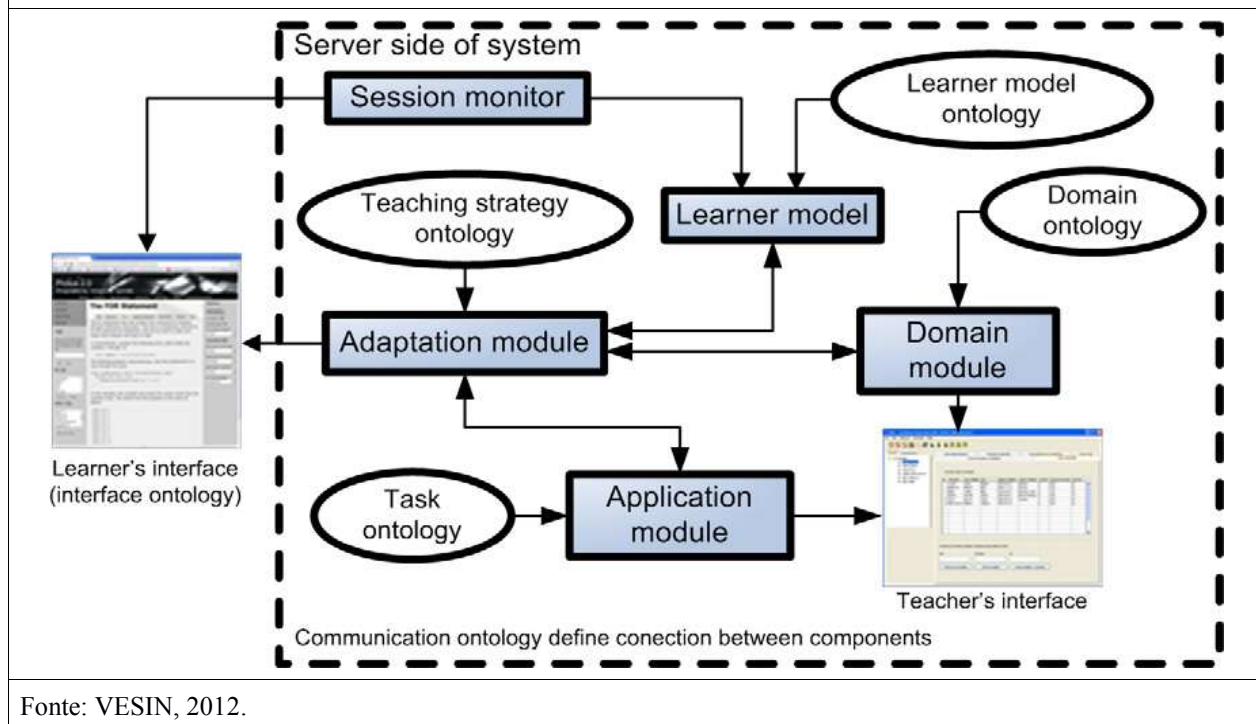
a personalização do curso, através da identificação do estilo de aprendizado dos usuários e, a partir dessa informação, recomendar conteúdo (VESIN, 2012). O sistema encontra-se na segunda versão, a qual faz uso de ontologia e regras de adaptação para representar o conhecimento, bem como engenhos de inferência para realizar raciocínio sobre a ontologia.

Conforme pode ser observado na arquitetura do sistema (figura 7), PROTUS 2.0 é constituído, do lado do servidor, de três módulos centrais: o Módulo da Aplicação, o Módulo de Adaptação e o Módulo do Domínio, estando cada um desses módulos conectados a uma ontologia específica, que são a Ontologia de Tarefas, a Ontologia de Estratégias de Ensino, a Ontologia de Modelo do Aprendiz e a Ontologia do Domínio, respectivamente. Além disso o sistema conta com um Modelo do Aprendiz e um monitorador das ações realizadas dentro do escopo de uma sessão de interações (o Componente de Monitoramento da Sessão).

O sistema monitora as ações durante uma sessão de acesso do aprendiz, por meio do Componente de Monitoramento da Sessão, assim como detecta e corrige erros cometidos por eles. Ao final de cada sessão essas informações são associadas àquelas já coletadas, e o Modelo do Aprendiz é atualizado e usado em uma nova sessão pelo Módulo de Adaptação, para promover a personalização do sistema. Desse processo participam o Módulo da Aplicação, que é onde serão aplicadas as personalizações e o Módulo do Domínio, responsável por descrever como o material de aprendizagem (tutoriais e testes) está estruturado.

O PROTUS 2.0 apresenta a capacidade de adaptar a interface do aprendiz de acordo com as características de aprendizagem dos mesmos, considerando o domínio abordado. Trata-se de uma característica desejável. Entretanto o sistema não oferece ao professor uma forma de identificar questões pedagógicas através da análise dos dados educacionais, nem permite a participação dos especialistas no domínio educacional para contribuir com recomendações para os cenários pedagógicos que o sistema é capaz de identificar.

Figura 7 - Arquitetura PROTUS 2.0



Fonte: VESIN, 2012.

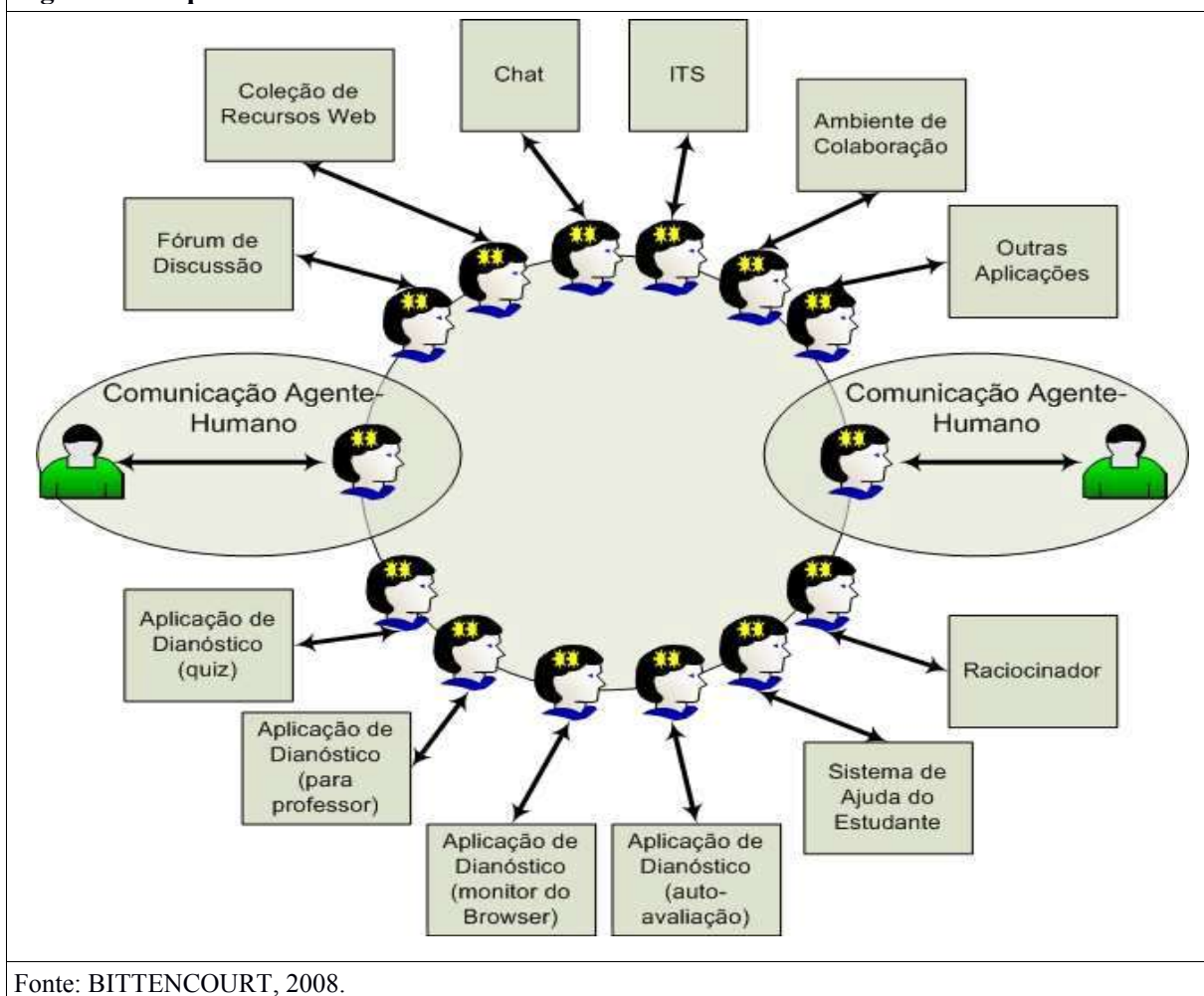
3.5 I-HELP

I-HELP, desenvolvido no *Laboratory for the Advanced Researches in Intelligent Educational Systems*⁴², pertencente à *Saskatchewan University*⁴³, é um sistema que visa auxiliar os aprendizes na resolução de questões, através da **localização de recursos** (online e humanos), de acordo com as especificações da necessidade do aprendiz. O sistema assemelha-se, em sua interface com o usuário, com um ambiente de compartilhamento de mensagens em áreas de discussão públicas e privadas. Contudo sua funcionalidade ocorre devido a um complexo sistema envolvendo vários agentes, conforme é exibido na figura 8, que se comunicam entre si, modelos do aprendiz que são propagados entre esses agentes e mecanismos de inferência capazes de localizar “ajudantes” apropriados para o modelo do aprendiz (GREER, 2001).

42 Laboratório de Pesquisas Avançadas em Sistemas Educacionais Inteligentes, disponível em: http://www.cs.usask.ca/research/research_groups/aries/projects.htm

43 Universidade de Saskatchewan, disponível em: <http://www.usask.ca/>

Figura 8 - Arquitetura do I-HELP.



Fonte: BITTENCOURT, 2008.

O sistema é constituído de dois componentes: o **I-HELP Pub**, um ambiente público de discussão, e o **I-HELP 1-on-1**, para discussões privadas.

Através do ambiente **I-HELP Pub** os aprendizes podem, de forma assíncrona, postar perguntas, comentários e respostas nos fóruns. As postagens feitas são compartilhadas com os demais aprendizes

O **I-HELP 1-on-1** permite que ocorram interações privadas entre um aprendiz, aquele no papel de quem solicita a ajuda, e outro aprendiz (ou *expert*), que tem o papel de prover a ajuda, e funciona segundo a sequência a seguir:

1. Um aprendiz entra em contato com seu agente pessoal e faz uma requisição de ajuda.
2. Esse agente negocia com os agentes de outros aprendizes, em busca de potenciais “ajudantes”.
3. Os mais bem qualificados são notificados sobre uma solicitação de ajuda em aguardo.
4. O primeiro que aceitar a solicitação iniciará uma interação um a um (1-on-1) com o

aprendiz em busca de ajuda, e as solicitações para os demais “ajudantes” são canceladas.

5. Após a conclusão da interação, cada envolvido recebe um formulário de avaliação do parceiro de interação, para auxiliar na modelagem dos alunos.

O I-HELP depende da busca por "ajudantes", ou material online, apropriados para a solução das questões dos alunos. Contudo não há nenhuma menção a uma forma de verificar a relevância do material online, ou do quão apropriado é o ajudante para o problema específico do aluno, nem se a busca por esses recursos considera as características e situação pedagógica dos alunos. Acreditamos que esses fatos colocam em risco a validade da ferramenta.

3.6 Tabela Comparativa

Nessa seção exibimos a tabela criada com o propósito de sumarizar as características dos sistemas e ferramentas apresentados nos trabalhos relacionados, bem como compará-los com ferramenta proposta nessa dissertação (tabela 4). Nessa sumarização consideramos características como a uso da mineração de dados orientada a cenários, a construção dos sistemas seguindo, explicitamente, um processo próprio, o uso da mineração dos dados educacionais, a proposição de recomendações de natureza pedagógica, a proposição de recomendações personalizadas que consideram as características dos alunos, e, por fim, se há a possibilidade de os professores participarem/contribuírem para o processo de recomendação pedagógica.

Tabela 4 - Tabela comparativa dos trabalhos relacionados.

	Mineração Orientada a Cenários	Utiliza um Processo Pedagógico Próprio	Mineração de Dados Educacionais	Recomendações Pedagógicas	Recomendações Personalizadas	Participação dos Professores no Processo
<u>Ferramenta Proposta</u>	SIM	SIM	SIM	SIM	SIM	SIM
Collaborative Recommender	NÃO	NÃO	SIM	SIM	SIM	SIM
EDUCA	NÃO	NÃO	SIM	SIM	SIM	SIM
Web Usage Mining on AHA!	NÃO	NÃO	SIM	SIM	SIM	NÃO
PROTUS 2.0	NÃO	NÃO	SIM	NÃO	SIM	NÃO
I-Help	NÃO	NÃO	NÃO	NÃO	SIM	NÃO

Fonte: criada pelo autor.

4 A PROPOSTA

A proposta apresentada nessa dissertação é a criação de uma **Ferramenta para Recomendação Pedagógica**, que disponibilize aos professores de cursos baseados na web, **Recomendações Pedagógicas Personalizadas** geradas com o auxílio da **Mineração dos Dados** oriundos das interações de seus alunos (i.e.: dados educacionais) com o ambiente de aprendizagem. O processo de mineração é orientado, conforme seus objetivos pedagógicos, pelos **Cenários de Mineração**.

A proposta visa atender às questões de pesquisa (Seção 1.2). O uso da Mineração de Dados Educacionais, orientada através dos cenários de mineração, tem por objetivo **descobrir informações e padrões úteis** (ROMERO, 2010), para o contexto educacional, a partir dos registros das interações dos alunos dentro do ambiente educacional, o que atende à primeira questão de pesquisa.

Sabemos que, simplesmente, apresentar os resultados da mineração aos professores pode sobrecarregá-los com informações que deverão ser filtradas e interpretadas antes de seu uso. Uma abordagem mais apropriada seria definir um **propósito mais prático de utilização dessas informações** (GARCIA, 2009), justificando assim o uso das recomendações pedagógicas e respondendo à segunda questão de pesquisa.

Entretanto essas soluções não podem demandar dos professores conhecimento sobre técnicas diferentes daquelas próprias de seu ofício. Tal motivo justifica o uso da Ferramenta de Recomendação e o uso de Cenários de Mineração, respondendo à terceira questão de pesquisa.

Nas subseções desse capítulo explicaremos com mais detalhes o conteúdo da proposta, partindo do Processo de Recomendação Pedagógica. Logo após, abordaremos o conceito de Cenários de Mineração, que tem por objetivo definir o que são os Cenários e como se dá a mineração dos dados seguindo essa técnica. Em seguida o apresentaremos os requisitos da Ferramenta de Recomendação, incluindo a arquitetura do software, casos de uso e detalhes sobre cada componente. Por fim exibiremos as interfaces da Ferramenta de Recomendação e os diagramas de sequência que representam seu funcionamento.

4.1 O Processo de Recomendação Pedagógica

Uma das principais tarefas dos professores é identificar as necessidades pedagógicas de seus alunos (OLIVEIRA, 2009), oferecendo ajuda personalizada e direcionamento, que são normalmente apresentados como recomendações de diversos tipos para superar uma

determinada situação. Isso constitui a base das **recomendações pedagógicas**, que nessa dissertação são definidas como *"ações de natureza preventiva, ou reativa, associadas a situações pedagógicas definidas. Tais ações podem ser realizadas utilizando os recursos nativos do ambiente de aprendizado, ou através de recursos externos, e têm o objetivo de solucionar os problemas (pedagógicos) identificados, melhorando a experiência de aprendizado dos alunos"*.

A recomendação pedagógica é uma atividade que consome bastante tempo, e que envolve não apenas o conhecimento das possíveis soluções para um problema, mas quais soluções estão de acordo com as capacidades de cada aluno, com base em seu perfil pedagógico. Consideremos, por exemplo, os professores de cursos baseados na web, foco dessa dissertação: esses professores são responsáveis por diversas turmas, cada uma contendo vários alunos, cada um deles com suas necessidades, características e limitações (DUKE, 2013). A partir dessa situação, percebemos a grande quantidade de dados que se deve coletar, pré-processar, avaliar buscando informações úteis (dentro do contexto pedagógico), planejar e agir, realizando todo esse processo de forma acurada e no tempo adequado para prover o auxílio que os alunos precisam.

Os professores realizam todo esse processo, normalmente, **sem qualquer tipo de suporte/apoio** tendo ainda que, muitas vezes, checar os resultados, realizar ajustes e repetir todo o processo até que cessem os problemas ou que o curso chegue ao seu fim.

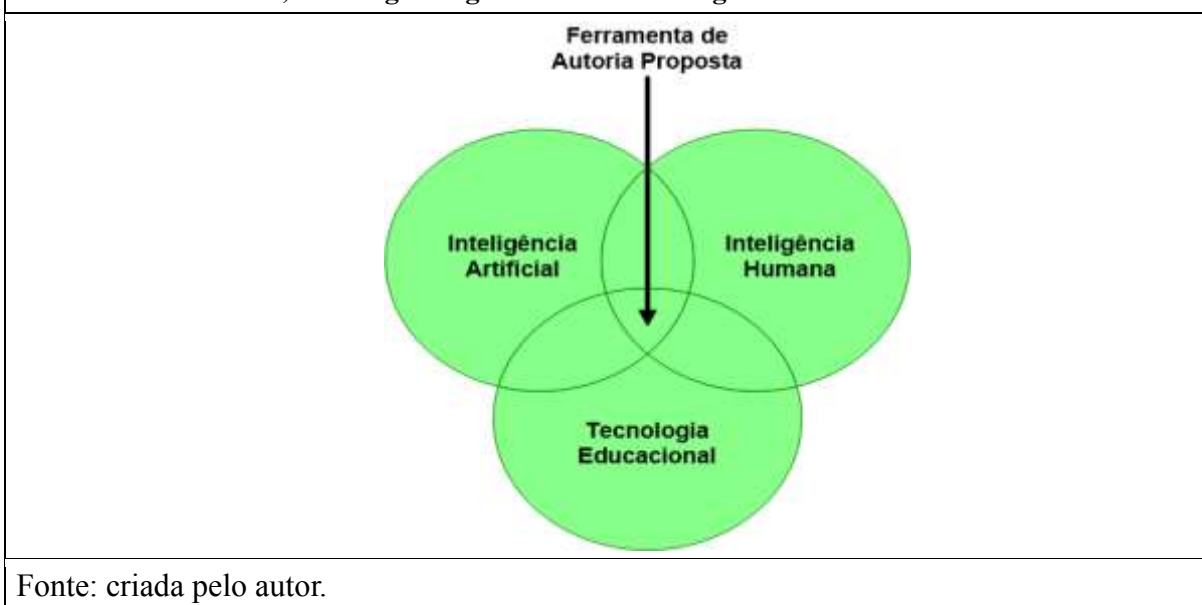
Para que sejamos capazes de prover essas recomendações, é essencial **conhecer o ambiente de aprendizagem**, suas capacidades e limitações. É necessário, também, ser capaz de **coletar o máximo de informações possíveis** sobre o desempenho dos alunos e sobre a forma que eles se comportam (i.e.: interagem) dentro desse ambiente. Torna-se necessário **o monitoramento e avaliação das recomendações**, e seu impacto sobre o desempenho dos alunos, para determinar se as mesmas estão sendo efetivas ou não, uma vez que recomendar algo inapropriado pode acarretar danos mais graves, ao aprendizado, que não realizar recomendação alguma.

Com base no que foi exposto anteriormente, identificamos na intersecção de três grandes subáreas da pesquisa em **Computação e Educação**, (vide Figura 9) um processo que denominamos de **Processo de Recomendação Pedagógica**, que trataremos com mais detalhes adiante. Essas "três grandes subáreas" estão listadas abaixo:

1. **Tecnologias Educacionais:** representada pelos ambientes online de aprendizagem, ferramentas síncronas e assíncronas.

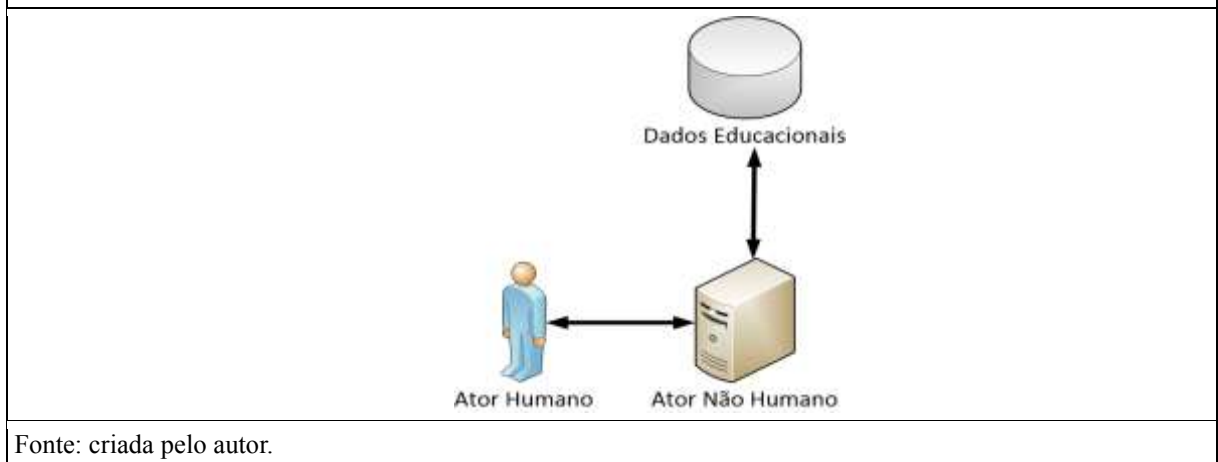
2. **Inteligência Artificial:** representada nessa dissertação por técnicas de mineração de dados, mineração de dados educacionais e aprendizado de máquina.
3. **Psicologia Cognitiva**⁴⁴: por exemplo, o modelo cognitivo dos alunos, que é uma representação, no âmbito computacional, de suas características, não apenas suas habilidades e conhecimentos em determinado assunto, mas podendo abordar aspectos afetivos, motivacionais, meta-cognitivos e comportamentais, (BAKER, 2010).

Figura 9 - O Processo de Recomendação Pedagógica na interseção das áreas de Inteligência Artificial, Psicologia Cognitiva e as Tecnologias Educacionais.



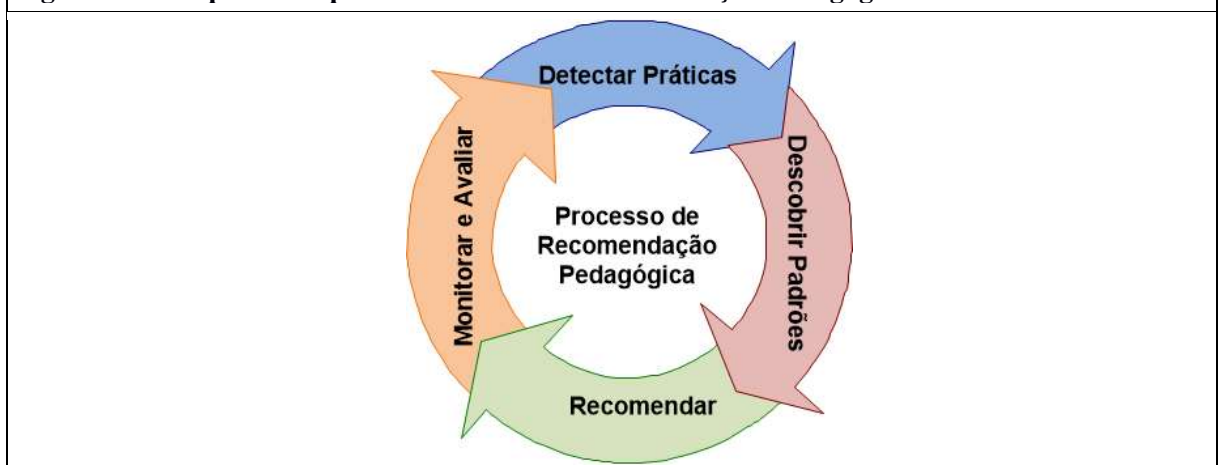
É muito importante mencionar que a integração das subáreas supracitadas, para o **Processo de Recomendação Pedagógica**, ocorre de forma **semiautomática**, ou seja, a **Inteligência Humana** e a **Inteligência Artificial** deverão trabalhar juntas e de forma coordenada, com vistas a proporcionar o aprimoramento no desempenho dos alunos. Por um lado, a inteligência artificial (**Ator Não-Humano**) é encarregada da coleta, processamento e análise dos dados, devendo apresentar os resultados aos **Atores Humanos**, de forma que sejam facilmente compreendidos. Por outro lado, a inteligência humana, através dos **Atores Humanos**, deve avaliar os resultados da análise automática e a partir delas tomar decisões no contexto pedagógico. Essas interações são representadas na figura 10.

⁴⁴ Cognição: processo de conhecer. O termo é usado pelos psicólogos para atividades mentais, como usar a linguagem, pensar, raciocinar, resolver problemas, conceituar, lembrar, imaginar e aprender matérias complexas (DAVIDOFF, 1983).

Figura 10 - A interação entre os dados e os atores do Processo de Recomendação Pedagógica.

É importante, também, frisar que esse processo é de natureza **iterativa e incremental**, constituído por quatro grandes etapas (figura 11), listadas a seguir, e serão detalhadas mais adiante nessa seção.

1. **Detectar Práticas** (positivas e negativas) no processo de aprendizagem (O Que?).
2. **Descobrir padrões** responsáveis pelas práticas detectadas (Por quê?).
3. **Recomendar** conteúdo pedagógico com base nos padrões descobertos, com o propósito de solucionar problemas detectados (Como?).
4. **Monitorar e Avaliar** a efetividade/eficiência das recomendações (Quais os Resultados?).

Figura 11 - As quatro etapas do Processo de Recomendação Pedagógica.

4.1.1 Detectar Práticas

Detectar práticas é a primeira etapa do processo de recomendação pedagógica e antes de descrevermos seus detalhes, gostaríamos de destacar algumas definições:

- Práticas é o termo utilizado para as possíveis formas de interagir por meio do ambiente de aprendizagem. Essas ações podem ocorrer, porém não se limitando, entre:
 - Alunos e professores/tutores.
 - Alunos e outros alunos.
 - Alunos e o ambiente de aprendizagem.
 - Professores e o ambiente de aprendizagem.
- Aquelas práticas que promovem, ou facilitam, o aprendizado e o desenvolvimento acadêmico, serão denominadas de Práticas Positivas.
- De forma contrária, aquelas que prejudicam o aprendizado, serão chamadas de Práticas Negativas.

Mencionado isso, podemos afirmar que o objetivo dessa etapa é o de identificar se as interações dos alunos, dentro do ambiente de aprendizagem, estão resultando em progresso para a experiência de aprendizagem, individual ou em grupo, ou se os resultados observados indicam prejuízo ao processo de aprendizado dos alunos e turmas.

Não é, portanto, objetivo dessa etapa identificar os detalhes e padrões dessas práticas. Expressando de outra forma podemos dizer que a detecção de práticas se ocupa da identificação de ocorrências e resultados acima ou abaixo de um valor, ou faixa de valores, esperados. São exemplos de práticas a identificação de resultados de provas abaixo da média, a detecção de alunos com número de faltas, para um determinado período, superior à média ou a um valor de referência, etc.

Como exemplo sobre a detecção de práticas pedagógicas, apresentamos os gráficos da figura 12. O gráfico apresenta uma situação hipotética onde um mesmo professor possui três turmas diferentes, porém aplica o mesmo plano de aula (conteúdos e métodos de avaliação).

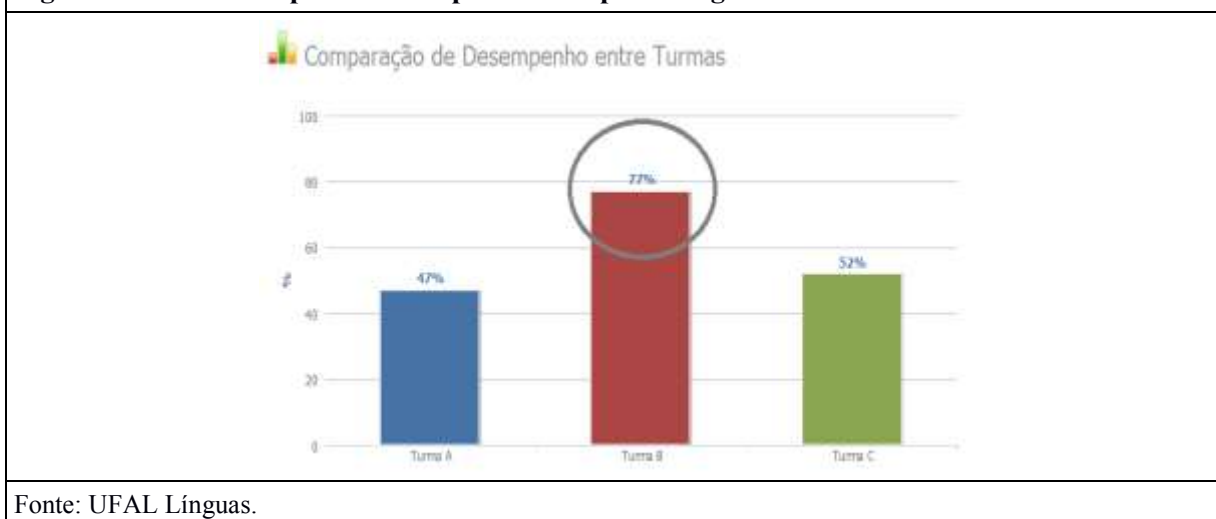
Podemos avaliar a situação exposta nesse gráfico de duas formas:

1. Uma, ou várias, práticas positivas estão ocorrendo na turma B e isso a coloca em vantagem em relação às outras turmas.
2. As turmas A e C, devido a alguma **prática negativa** sendo realizada, estão sofrendo danos em seus desempenhos.

A partir da constatação da ocorrência de um tipo de prática, devemos formular **hipóteses** as quais tentarão identificar/explicar os motivos de essas práticas estarem acontecendo. A confirmação, ou não, dessas hipóteses ocorrerá por meio da utilização das técnicas e análise dos resultados que serão feitas na próxima etapa do processo, as quais são realizadas pelos Atores Humanos (Inteligência Humana), ou seja, os especialistas no domínio pedagógico e os especialistas no domínio tecnológico. Hipóteses válidas resultarão nos

Cenários de Mineração que têm como um de seus objetivos, **orientar** a mineração dos dados educacionais dentro do Processo de Recomendação Pedagógica, abstraindo do professor os conceitos sobre as técnicas utilizadas (Mineração de Dados Educacionais, por exemplo).

Figura 12 - Gráficos apresentando prática de aprendizagem.



4.1.2 Descobrir Padrões

O foco dessa etapa é **descobrir os padrões** relacionados às práticas que foram **detectadas** na primeira etapa (vide **Subseção 4.1.1**), ou seja, descobrir o motivo pelo qual cada prática está ocorrendo. Como mencionamos na subseção anterior, nessa etapa utilizaremos as técnicas de mineração e análise estatística dos dados com o intuito de confirmar, ou rejeitar, as hipóteses que foram criadas.

A primeira coisa a se fazer, então, é conhecer o ambiente de aprendizagem. Dentre suas capacidades, a de coletar dados sobre o desempenho dos alunos e sobre a forma como eles se comportam dentro desse ambiente (i.e.: dados interacionais) é essencial para que essa etapa do processo seja bem-sucedida. De acordo com (HAN, 2011), um dos mais importantes requisitos da mineração de dados é a “**qualidade dos dados**”, que deve ser entendida como dados corretos, úteis e em uma quantidade apropriada (i.e.: quantidade de instâncias) para a finalidade pretendida.

Se conhecermos as capacidades e limitações quanto à coleta de dados por parte do ambiente de aprendizagem, poderemos analisar se dados apropriados, para se realizar a descoberta dos padrões que justificam as práticas, estão disponíveis e quais deles podem ser úteis para cada descoberta.

Em seguida será necessário definir se esses dados precisam de algum tipo de tratamento antes de serem utilizados no processo de mineração (i.e.: pré-processamento dos dados). Nesse ponto será preciso utilizar técnicas da estatística descritiva para se conhecer os dados escolhidos (por exemplo, podemos analisar uma variável em busca de valores muito diferente dos demais - *outliers*).

De posse dos dados escolhidos e tratados, precisamos definir qual tarefa de mineração é apropriada para a análise das hipóteses (por exemplo, obter quais as características comuns dos alunos com baixo rendimento, obter quais as características comuns dos alunos que desistiram do curso, etc.). Feita essa escolha, e conhecendo os algoritmos de mineração de dados, saberemos qual deles fornece os resultados desejados.

Por fim, em alguns casos é necessário um tratamento final que possibilite a leitura dos resultados ou para desfazer alguma transformação de dados realizada no pré-processamento. Esse passo é chamado de pós-processamento dos dados ou resultados.

Com isso teremos o necessário para decidir se desejamos aceitar ou rejeitar uma hipótese. Sendo aceita, todos esses passos anteriores serão utilizados para a criação de um **Cenário de Mineração** que representam uma forma de definir e limitar rotinas/situações específicas de mineração de dados. Eles devem conter todas as especificações necessárias para orientar as tarefas de mineração. Gostaríamos de fornecer um exemplo bastante simples desse conceito. Considerando, ainda, a situação da figura 12, os especialistas no domínio pedagógico acreditam que o motivo de a turma B ter obtido destaque nas avaliações é que os alunos dessa turma interagem mais através do fórum. Após análise dos dados (por exemplo: quantidade de posts por semana, quantidade de resposta por post, quantidade de visitas ao fórum, etc.), e da escolha da tarefa de mineração (por exemplo, classificação), confirmou-se a hipótese e, a partir de agora, temos um cenário de mineração.

Sumarizando o que ocorre nessa etapa temos que foi feita a escolha dos dados adequados, que serão pré-processados, minerados (de acordo com os algoritmos escolhidos e seus parâmetros) e, por fim, pós-processados de forma que possam ser utilizados na etapa posterior. Essa etapa está de acordo com as fases 2 a 4 do processo CRISP-DM (mais detalhes sobre o processo CRISP-DM encontram-se na **Seção 2.3**), e requer a participação conjunta de especialistas no domínio pedagógico e do domínio tecnológico, para que sejam definidos os atributos do processo de mineração dos dados.

4.1.3 Recomendar

Essa etapa é responsável por disponibilizar aos professores opções de recomendações pedagógicas para que estas sejam encaminhadas aos usuários finais (i.e.: os alunos). Essas recomendações são criadas por especialistas no domínio pedagógico com base na análise dos padrões descobertos na etapa anterior. Elas são realizadas com foco na resolução das situações pedagógicas trazidas à luz por meio da descoberta dos padrões, cuja mineração é feita com base nos cenários definidos (**Seção 4.2**).

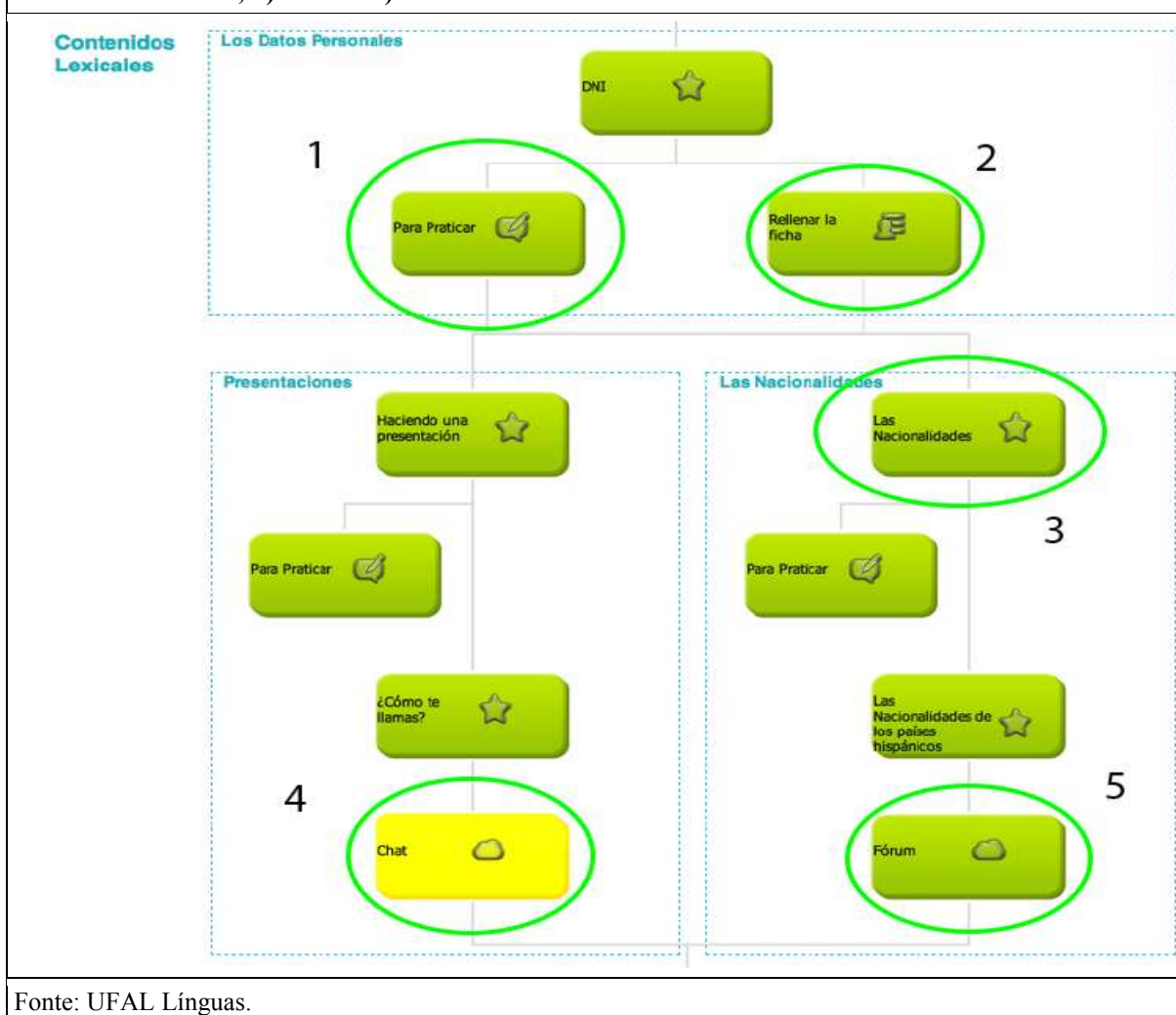
Contudo essas recomendações só estarão disponíveis para os alunos, após a escolha daquela considerada apropriada, tarefa realizada pelos usuários finais, ou seja, os professores responsáveis pelos cursos.

Para sermos capazes de recomendar algo através de um ambiente de aprendizagem, precisamos conhecer quais **recursos educacionais** estão disponíveis (vide figura 13).

Esses recursos localizam-se na interface dos usuários entre si (i.e.: professores, tutores e alunos), e dos usuários com o ambiente de aprendizagem, tendo dentre as suas finalidades **apresentar o conteúdo** (aulas em diversos formatos de mídia), **avaliar a aquisição de conhecimento** (exercícios e testes), **coletar e apresentar resultados e dados educacionais** (gráficos e relatórios), **prover ajuda e apoio** (tutoriais, material extra, links externos), **comunicar, engajar e agregar seus usuários** (chat e fórum).

Diante dessas considerações, temos que as **Recomendações Pedagógicas** consistem de entidades **externas** ou **próprias** do ambiente de aprendizagem, cujo objetivo é **oferecer orientações** para que se possa superar **dificuldades pedagógicas já detectadas** (i.e.: latentes), ou **previstas**. Uma recomendação pedagógica para ser classificada como útil, deve ser capaz de aprimorar a experiência de aprendizagem dos alunos aos quais ela foi direcionada.

Figura 13 - Representação de alguns recursos educacionais disponíveis no ambiente UFAL Línguas. 1) Exercícios, 2) Preencher os Dados Pessoais em Espanhol, 3) Aulas em Vídeo, 4) Chat e 5) Fórum.



Fonte: UFAL Línguas.

Entretanto diferentes alunos podem ter necessidades distintas e, certamente, características e capacidades diferentes. Percebemos, então, que a personalização dessas recomendações é um atributo capaz de maximizar suas chances de sucesso, estando de acordo com objetivos como o "Acompanhamento Pedagógico Personalizado" (PNE, 2010) e a "Personalização da Aprendizagem" (DESAFIE!, 2012), mencionados no **Capítulo 1**. Assim, para se oferecer recomendações personalizadas, precisamos considerar a situação pedagógica dos alunos.

Por fim, listamos abaixo algumas orientações para se criar recomendações pedagógicas personalizadas e úteis. São elas:

1. **Ser criada em tempo adequado:** isso significa que a recomendação deve ser gerada enquanto ainda houver **tempo** para que o aluno possa segui-las.

2. **Ser relevante para o contexto do problema:** nesse caso precisamos garantir a correta detecção da "prática pedagógica" e descobrir com precisão os seus padrões, caso contrário ofereceremos "uma recomendação certa, para a situação pedagógica errada", ou o contrário.
3. **Considerar as características dos alunos:** tal orientação é primordial se pretendemos promover a **personalização** da recomendação e, como mencionado, maximizar suas chances de sucesso. Se não a seguirmos, corremos o risco de recomendar algo que pode estar além das condições de o aluno realizar, ou que seja muito simples e não ofereça ajuda ou aprimoramento significativos.
4. **Promover melhoria no processo de aprendizagem:** as recomendações devem ser atraentes para os alunos e, caso eles as realizem, devem sentir e demonstrar que foram impactados positivamente, superando os problemas pedagógicos que enfrentavam e para os quais as recomendações foram realizadas.

4.1.4 Monitorar e Avaliar

Essa etapa tem por objetivo avaliar se os alunos estão progredindo positivamente em decorrência das recomendações recebidas e, também, avaliar o quão relevante estão sendo essas recomendações para as situações pedagógicas para as quais elas foram propostas. Essa tarefa é realizada da seguinte forma:

1. **Monitorar o progresso dos alunos:** medindo o quão bem os alunos estão **progredindo**, após terem sido expostos a uma determinada recomendação. Para isso é necessário saber se o aluno aceitou a recomendação, conhecer como ele estava progredindo antes de receber a recomendação e comparar com seu os desempenhos atual e anterior à recomendação.
2. **Avaliar a relevância das recomendações:** através da criação de uma pontuação que permita classificar as recomendações quanto à sua relevância. Tal pontuação deve considerar o número de vezes que a recomendação foi sugerida, o número de vezes que foi aceita e realizada, quantas vezes foi rejeitada, quantas vezes promoveu resultados pedagógicos positivos e, possivelmente, uma avaliação manual feita por especialistas no domínio pedagógico.

É através desta etapa que novos dados serão coletados com vistas a avaliar a própria ferramenta por meio da análise da qualidade das recomendações, bem como prover novos dados de entrada para a detecção de outras práticas pedagógicas.

4.2 Cenários de Mineração de Dados

Conforme citado no processo CRISP-DM (NISBET, 2009), e mencionado por outros autores (LAROSE, 2006; HAN, 2011; NORTH, 2012), para que um processo de mineração seja bem-sucedido, é necessário compreender o domínio onde a mineração de dados será aplicada, compreender os dados e definir claramente que resultado se pretende obter a partir dessa mineração.

Para conseguirmos seguir essas orientações, criamos componentes para minerar dados educacionais em busca dos padrões de uma determinada situação pedagógica. Esses componentes, denominados **Cenários de Mineração de Dados**, contêm as especificações que serão utilizadas pelo processo de mineração de acordo com os objetivos pretendidos.

Para a criação dos cenários exibidos nessa dissertação, consultamos o que alguns autores consideram problemas importantes e recorrentes em ambientes de aprendizagem. Além disso, foi avaliada a possibilidade desses problemas serem tratados por meio do uso das TICs, e se os mesmos poderiam ser encontrados no ambiente educacional de onde os dados, para o estudo de caso demonstrado nessa dissertação, foram extraídos.

De acordo com (MORAN, 2006) a avaliação dos alunos deve ser feita de forma continuada, o que significa que é necessário que aconteça durante todo o curso, por meio de atividades que possibilitem a avaliação qualitativa da aprendizagem. Essas atividades são classificadas pelo autor em três categorias: i) atividades relacionadas ao conteúdo; ii) pesquisas sobre temas próximos à vida e interesses do aluno; e iii) avaliação da qualidade da participação do aluno no ambiente virtual (por exemplo: as tradicionais provas, a participação dos alunos, comentários, críticas e atitudes em relação aos assuntos abordados, aos professores e colegas). Essa preocupação nos motivou para a criação do nosso primeiro cenário, denominado: **Avaliar a qualidade das interações dos alunos no ambiente de aprendizagem**, que analisa como os alunos interagem com o ambiente de aprendizagem, indicando em que pontos eles podem melhorar.

Nosso segundo cenário baseia-se na possibilidade de oferecer aos professores uma forma de identificar quais alunos estão sob o risco de falha em uma prova, ou outro tipo de avaliação para que, segundo (HUANG, 2012), seja possível realizar intervenções pedagógicas que visem recuperar e sensibilizar o aluno para a melhoria de seu desempenho, evitando a reprovação. Tal cenário foi denominado **Prever o resultado de provas e avaliações**.

O último cenário criado para essa dissertação tem o objetivo de identificar comportamentos de risco dentro do ambiente de aprendizagem, como por exemplo aqueles

que antecedem a desistência de um aluno. De acordo com (KANELLOPOULOS, 2008), o número de desistentes nos cursos a distância é "definitivamente" superior ao de cursos convencionais, por isso há a necessidade de se obter antecipadamente, de forma confiável, esse conhecimento a partir dos dados educacionais armazenados pelo ambiente de aprendizagem (BAYER, 2012). Com esse objetivo foi criado o cenário: **Agrupar alunos por nível de participação, comprometimento e engajamento com o curso.**

Os três cenários, mencionados anteriormente, foram criados para uso na ferramenta de recomendação proposta nesse trabalho, e são apresentados na tabela 5.

Tabela 5 - Descrição dos cenários de mineração criados.		
Código	Cenário	Descrição
C1	Avaliar a qualidade das interações dos alunos no ambiente de aprendizagem	Esse cenário avalia se a forma como determinado aluno interage com o sistema favorece seu aprendizado resultando em bom desempenho na disciplina, ou se o que acontece é o contrário.
C2	Prever o resultado de provas e avaliações	Esse cenário avalia o desempenho, com o progredir do curso, de determinado aluno nos exercícios e atividades, bem como a qualidade de suas interações, com o propósito de antecipar o resultado em provas e avaliações.
C3	Agrupar alunos por tipo de participação, comprometimento e engajamento com o curso	Esse cenário compara a forma como os alunos se comportam dentro do sistema e os agrupa de acordo com as características de participação dos mesmos dentro do ambiente de aprendizagem.

A criação dos cenários é, inicialmente, orientada por três perguntas as quais apresentam uma visão geral sobre o cenário. Para chegarmos às respostas (vide tabela 6), focamos no objetivo que deve ser alcançado para cada cenário e nas características do ambiente de aprendizagem utilizado no estudo de caso.

Para o cenário C1, cujo objetivo é avaliar se a forma de os alunos interagirem com o ambiente de aprendizagem aprimora seu desempenho, identificamos os dados desejáveis com base na afirmação de (MORAN, 2006) de que é preciso avaliar a qualidade da participação do aluno no ambiente virtual, observando suas provas, participação, comentários, críticas e atitudes em relação aos assuntos abordados, aos professores e colegas.

Para o cenário C2, cujo objetivo é prever o resultado de provas e avaliações, para determinar que dados utilizar, observamos o que afirma (PARK, 2009), citando seu framework teórico para adultos desistente no aprendizado online (PARK, 2007), onde existem

fatores que influenciam a decisão de os alunos desistirem de um curso ou disciplina, que são: i) Características pessoais do aluno, como idade, sexo, nível educacional, situação de emprego, dentre outras; ii) Fatores externos, como problemas com horários, questões familiares, problemas financeiros, e outros; e iii) Fatores internos, que podem surgir com o decorrer do curso (ambiente de aprendizagem), como a capacidade de promover/facilitar a integração social, questões acadêmicas (o nível dos instrutores, o nível das atividades, a metodologia utilizada, dentre outras), questões tecnológicas/técnicas e de usabilidade, e questões motivacionais. Segundo (MANHÃES, 2011), grandes dificuldades enfrentadas no início do curso (disciplinas, atividades e provas muito difíceis) são fatores decisivos para a perda da motivação com conseqüente desistência.

Para o cenário C3, cujo objetivo é agregar os alunos por suas características de participação e engajamento no curso, com vistas a oferecer melhores recomendações de recursos educacionais (BAKER, 2009), nos baseamos em (NISBET, 2004) que descreve um modelo proposto, e fundamentado com pesquisas, por (SALMON, 2000) onde são estabelecidos 5 estágio de engajamento online, que são: i) acesso e motivação, representados por acesso ao sistema e a fóruns com postagens; ii) socialização online, quando os alunos compartilham um pouco de suas experiências; iii) troca de informações; iv) construção do conhecimento; e v) desenvolvimento do próprio aprendizado. Dessa forma listamos na tabela 6 os dados desejáveis para observar esses estágios.

Para a criação da decisão sobre quais as possíveis ações/tarefas poderíamos realizar com os dados, consultamos (ROMERO, 2011) em busca de informações a respeito das técnicas mais comumente utilizadas e aceitas, para os objetivos desejados.

Conforme já foi mencionado, nessa etapa há uma cooperação dos Atores Humanos que compartilham seus conhecimentos (domínios pedagógico/educacional e tecnológico). As perguntas são as seguintes:

1. Qual o objetivo desse cenário?
2. Quais os dados necessários para atingir esse objetivo?
3. O que deve ser feito com esses dados para se chegar ao objetivo pretendido?

Tabela 6 - Respostas às perguntas criadas para compreender os cenários de mineração.

Código	Objetivo	Dados	Ação com os Dados
C1	Avaliar o quão produtivas, pedagogicamente, são as interações e o comportamento de cada estudante no ambiente de aprendizagem.	Por Curso <ul style="list-style-type: none"> • Acessos • Exercícios • Testes • Interações via fórum • Interações via chat • Desempenho 	<ul style="list-style-type: none"> • Classificar • Gerar árvore de decisão
C2	Prever o resultado de avaliações, com base no desempenho do estudante nos exercícios e na qualidade das interações.	Por Unidade <ul style="list-style-type: none"> • Acessos • Número de exercícios • Nível dos exercícios • Exercícios respondidos • Exercícios respondidos por nível • Exercícios corretos • Exercícios corretos por nível • Resultados de testes anteriores • Resultado na análise da qualidade das interações 	<ul style="list-style-type: none"> • Regras de associação • Detectar grupos de risco • Classificar
C3	Agrupar alunos por característica de participação, comprometimento e engajamento, para facilitar a recomendação de recursos de aprendizagem.	Por Semana <ul style="list-style-type: none"> • Acessos • Relação entre exercícios respondidos / total de exercícios disponíveis • Características dos recursos de aprendizagem que interagiu • Frequência de interação com tipo de recurso de aprendizagem • Interações via fórum • Interações via chat 	<ul style="list-style-type: none"> • Regras de associação • Agrupamento • Árvores de decisão • Classificar

Em seguida, definimos as especificações de mineração dos cenários (tabela 7). Nessa parte é necessário conhecimento sobre as capacidades e limitações dos algoritmos que se pretende utilizar, informações sobre a implementação destes (atributos necessários e API) e sua relação com os dados educacionais obtidos. De forma resumida devemos obter as seguintes informações:

1. Os dados de interesse.
2. A forma que esses dados serão pré-processados.
3. A maneira como ocorrerá o processo de mineração dos dados.

4. O modo como os resultados serão pós-processados.

Para os cenários criados, e considerando a capacidade de armazenamento de dados do ambiente educacionais do estudo de caso (EDUCA_®), chegamos aos dados relacionados na primeira coluna da tabela 7 após diversas iterações de experimentos.

A coluna que trata o pré-processamento dos dados está intimamente relacionada com as três técnicas de mineração de dados escolhidas [Árvore de Decisão (J48), Árvores de Classificação e Regressão (SimpleCart), Agrupamento (Simple K-Means)]. Para essa análise consultamos as descrições das técnicas e algoritmos presentes em (LAROSE, 2005; WU, 2007; WITTEN, 2011; HAN, 2011).

Algumas formas de tratamento dos dados se repetem para todos os cenários como, por exemplo, a remoção de valores superiores, e inferiores, a três vezes a variação interquartil (IQR⁴⁵), os quais podem ser classificados como *outliers* cuja presença afeta os resultados obtidos. Os registros dos desistentes são removidos, pois para os cenários criados esses valores influenciam os resultados. Removemos, também, o campo ID uma vez que são valores numéricos e geram alterações nos resultados para os algoritmos utilizados.

É recomendado, por questão de desempenho (WU, 2007; WITTEN, 2011) que valores não numéricos e categóricos sejam transformados em valores numéricos.

O uso de técnicas de imputação⁴⁶, para os dados ausentes, pode ser realizado quando há a suspeita de falha no registro do valor, ou quando utilizamos K-Means que é um algoritmo sensível (WU, 2007) ao valor zero. Em caso contrário podemos indicar com zero os valores ausentes, indicando o não registro dos mesmos.

O último tratamento dessa coluna refere-se ao tipo de arquivo utilizado. Devido ao fato de utilizarmos as implementações dos algoritmos contidas na API WEKA®, torna-se essencial convertermos os dados, já pré-processados, para o formato ARFF⁴⁷. Para tanto a própria API (Java®) oferece métodos que auxiliam na criação desse tipo de arquivo.

A coluna que trata sobre a mineração propriamente, contém os parâmetros que podem ser definidos pelo usuário para cada um dos algoritmos escolhidos (BOUCKAERT, 2012).

A validação cruzada em 10 dobras (10 fold cross-validation) consiste em dividir os dados em 10 conjuntos de dados com tamanho $n/10$, treinar o algoritmo utilizando nove

45 Interquartile Range, em Inglês.

46 Imputação é um processo que prove um valor substituto para aqueles que estão ausentes com base nos demais valores existentes, mas relacionados (BOSLAUGH, 2012).

47 Acrônimo para Attribute-Relation File Format (Formato de arquivo atributo-relação, em português), é o formato de arquivo processado, nativamente, pelo Weka e os métodos contidos em sua API Java (WITTEN, 2011).

desses conjuntos e realizar os testes no conjunto de dados restante. Esse processo é repetido 10 vezes e, no fim, calcula-se a média das acurácias obtidas em cada repetição.

O algoritmo J48 é executado com os seguintes parâmetros:

- **binarySplits(false):** dispensa a divisão de atributos nominais.
- **confidenceFactor(0.5):** determina o fator de confiança no algoritmo que definirá quanta poda deve ocorrer (quanto menos o fator de confiança, maior a poda. 0.5 é o maior valor possível).
- **minNumObj(2):** informa o número mínimo de observações por folha da árvore gerada.
- **reducedErrorPruning(false):** dispensa a simplificação da árvore de decisão através de podas. Esse parâmetro visa minimizar a quantidade de valores classificados de forma errada.
- **subtreeRaising(true):** expande as subárvores, sempre que possível. Esse parâmetro visa criar árvores maiores, contemplando maior número de possibilidades.
- **unpruned(false):** se selecionado (true) esse parâmetro inibiria podas à árvore gerada.
- **useLaplace(false):** se selecionado (true) esse parâmetro suavizaria o número de folhas, considerando funções de classificação mais gerais.

O algoritmo SimpleCart, para o cenário criado, é executado com os seguintes parâmetros:

- **debug(false):** informações adicionais sobre a execução do processo de mineração são exibidas no console de saída de dados.
- **heuristic(true):** adiciona heurística à busca para problemas com múltiplas classes.
- **minNumObj(3):** informa o número mínimo de observações por folha da árvore gerada.
- **numFoldsPruning(5):** define o número de dobras na validação cruzada realizada internamente.
- **seed(1):** número aleatório de seed a ser utilizado.
- **sizePer(1.0):** percentual das instâncias (1.0 = 100%) que são utilizadas no conjunto de treinamento.
- **useOneSE(false):** não utiliza a regra 1SE para as decisões de poda da árvore.
- **usePrune(true):** minimiza o custo de complexidade para a poda.

O algoritmo SimpleKMeans para o cenário criado, é executado com os seguintes parâmetros:

- **displayStdDevs(false):** não exibe o desvio padrão de atributos numéricos e não contaos atributos nominais.
- **distanceFunction(Euclidean):** calcula a distância entre os dados por meio da função Euclidiana.
- **dontReplaceMissingValues(false):** se selecionado (true), substituiria os valores ausentes pela média ou moda global. Contudo já realizamos o tratamento dos valores ausentes no pré-processamento dos dados.
- **maxIterations(500):** define o número máximo de iterações, evitando que o algoritmo nunca pare o processamento por não convergir.
- **numClusters(3):** define o número de agrupamentos.
- **preserveInstanceOrder(true):** manter a ordem original das instâncias de dados.
- **seed(10):** número aleatório de seed a ser utilizado.

Para concluir a descrição técnica dos cenários criados, trataremos sobre a última coluna da tabela, pós-processamento dos dados, que basicamente tem por objetivo (GIBERT, 2008) tratar os dados de forma que seja possível:

- Entender os resultados.
- Comunicar os resultados para uma etapa seguinte do processo, ou encaminhá-los diretamente àqueles responsáveis pela tomada de decisão.
- Identificar informações relevantes com rapidez e de maneira simples.
- Oferecer a melhor forma de apresentar os resultados para o usuário final.

Para os cenários criados são desfeitas as transformações que foram realizadas nos dados originais (por exemplo, fizemos a conversão do gênero dos alunos de masculino e feminino, para 0 e 1, e não foi do nosso interesse apresentar/armazenar esses dados em valores numéricos, o que nos levou a reverter a transformação). Recolocamos, também, os IDs dos registros.

Com base no exemplo anterior, percebemos que é necessário manter um registro de todas as transformações realizadas nos dados para tornar possível desfazê-las.

Cada cenário é representado por um componente, conforme pode ser observado, mais adiante, no diagrama de componentes (vide figura 16), os quais são gerenciados pelo Controlador de Cenários, estrutura que funciona como uma *facade*^{48 49}, redirecionando para o

48 Em Português, Fachada.

49 Uma classe cuja funcionalidade é oferecer uso comum para as classes de um pacote ou de um subsistema, tornando-o fácil de usar (METSKER, 2004).

cenário apropriado as requisições feitas pelo componente de Mineração. Essa abordagem visa simplificar a inclusão, modificação ou remoção de cenários.

Tabela 7 - Especificação dos cenários criados nessa dissertação.

	Dados de Interesse	Pré-Processamento	Mineração	Pós-Processamento
C1	<p>Agrupar dados por curso</p> <ul style="list-style-type: none"> • Número de acessos • Número de exercícios disponíveis • Número de exercícios respondidos • Número de exercícios respondidos por nível • Número de testes disponíveis • Número de testes realizados • Interações via fórum • Interações via chat • Média final dos testes • Status final de aprovação 	<ul style="list-style-type: none"> • Remover registros com outliers abaixo de $Q1 - 3 * IQR$ e acima de $Q3 + 3 * IQR$ • Remover registro de desistentes • Remover ID • Tratar valores categóricos • Tratar valores não numéricos • Preencher valores ausentes com 0 • Converter resultado para arquivo ARFF 	<ul style="list-style-type: none"> • API: Weka • Algoritmo: J48 • Validação Cruzada: 10 fold • Parâmetros (<ul style="list-style-type: none"> • binarySplits = false • confidenceFactor = 0.5 • minNumObj = 3 • reducedErrorPruning = true • subtreeRaising = true • unpruned = false • useLaplace = false) 	<ul style="list-style-type: none"> • Adicionar respectivos IDs • Converter valores categóricos • Gerar visualização da árvore de decisão
C2	<p>Agrupar os dados por unidade</p> <ul style="list-style-type: none"> • Acessos • Número de exercícios • Nível dos exercícios • Exercícios respondidos • Nível Exercícios respondidos • Exercícios corretos • Nível Exercícios corretos • Número testes anteriores • Resultado testes anteriores • Resultado C1 anteriores 	<ul style="list-style-type: none"> • Remover registros com outliers abaixo de $Q1 - 3 * IQR$ e acima de $Q3 + 3 * IQR$ • Remover registro de desistentes • Remover ID • Tratar valores categóricos • Tratar valores não numéricos • Preencher valores ausentes com 0 • Converter resultado para arquivo ARFF 	<ul style="list-style-type: none"> • API: Weka • Algoritmo: SimpleCart • Validação Cruzada: 10 fold • Parâmetros (<ul style="list-style-type: none"> • debug = false • heuristic = true • minNumObj = 3 • numFoldsPruning = 5 • seed = 1 • sizePer = 1.0 • useOneSE = false • usePrune = true) 	<ul style="list-style-type: none"> • Adicionar respectivos IDs • Converter valores categóricos
C3	<p>Agrupar dados por semana</p> <ul style="list-style-type: none"> • Acessos • Relação entre total de exercícios /exercícios respondidos • Interações via fórum • Interações via chat 	<ul style="list-style-type: none"> • Remover registros com outliers abaixo de $Q1 - 3 * IQR$ e acima de $Q3 + 3 * IQR$ • Remover registro de desistentes • Remover ID • Tratar valores categóricos • Tratar valores não numéricos • Tratar os valores ausentes (imputação) • Converter resultado para arquivo ARFF 	<ul style="list-style-type: none"> • API: Weka • Algoritmo: SimpleKMeans • Parâmetros (<ul style="list-style-type: none"> • displayStdDevs = false • distanceFunction = Euclidean • dontReplaceMissingValues = false • maxIterations = 500 • numClusters = 3 • PreserveInstancesOrder = true • seed = 10) 	<ul style="list-style-type: none"> • Adicionar respectivos IDs • Converter valores categóricos

4.3 A Ferramenta para Recomendação Pedagógica

Ferramentas para Recomendação, como o nome sugere, são ferramentas de software e cujo objetivo é prover sugestões/recomendações de itens a serem utilizados por um usuário, com o intuito de ajuda-los em diversos processos de tomada de decisão (RICCI, 2011). No contexto dessa dissertação, essas ferramentas realizam sua função utilizando os dados decorrentes das interações dos alunos no ambiente de aprendizagem estudado (EDUCA_), e por meio das estratégias comuns a esse tipo de ferramentas, por exemplo: a criação de modelos dos estudantes, obtenção das características dos recursos educacionais, identificando relações de interesses entre os alunos, etc. Maiores detalhes sobre sistemas/ferramentas de recomendação encontram-se na **Seção 2.5**.

A Ferramenta de Recomendação proposta nessa dissertação utiliza o Processo de Recomendação Pedagógica (detalhado na **Seção 4.1**), para oferecer aos professores a capacidade de minerar os dados educacionais gerados pelos alunos em decorrência de suas interações. Essa capacidade tem por meta buscar padrões de interesse pedagógico (detectar práticas) nesses dados e utilizar as informações descobertas para oferecer ajuda personalizada, na forma de recomendações pedagógicas, com o objetivo de melhorar/aprimorar a experiência de aprendizagem dos alunos.

A ferramenta proposta está dividida em três módulos, que são **a Interface com o Usuário**, através da qual os professores podem interagir com os resultados (recomendações pedagógicas, obtidas com base no cenário pedagógico que foi minerado) obtidos pelo **Módulo de Mineração dos Dados Educacionais**, e encaminhá-los para os respectivos alunos, ou turmas, bem como, avaliar e monitorar seu impacto no desempenho acadêmico destes, que é feito pelo **Módulo de Recomendação Pedagógica**.

É através do módulo de mineração dos dados educacionais que o processo de mineração é realizado. Esse módulo orienta suas tarefas com base nos **Cenários de Mineração**, detalhados na **Seção 4.2**. De acordo com os objetivos do professor, um cenário apropriado é escolhido (diferentes cenários, prestam-se a diferentes objetivos de mineração de dados para uma determinada situação pedagógica), e os padrões de interesse pedagógico descobertos são direcionados para o módulo de recomendação pedagógica, onde serão usados para propor recomendações personalizadas. Tais recomendações serão encaminhadas aos alunos, após escolha da mais apropriada e aprovação por parte do professor responsável pelo curso. Sendo aprovadas, as recomendações receberão uma pontuação positiva (**avaliação da recomendação**), em caso contrário serão desconsideradas (i.e.: não serão enviadas aos

alunos), sendo pontuadas negativamente. O desempenho dos alunos expostos às recomendações será monitorado para que se possa avaliar se a recomendação promoveu **melhoria na experiência de aprendizagem**, ou seja, foi efetiva em promover auxílio à situação pedagógica para a qual foi proposta.

A contribuição dessa ferramenta de recomendação é a de realizar esse processo sem que os professores necessitem de conhecimento sobre as técnicas envolvidas, bastando **conhecer os objetivos pedagógicos de todos os cenários de mineração** disponíveis.

A descrição da ferramenta se dá pela apresentação dos Casos de Uso (**Subseção 4.3.1**), da Arquitetura de Alto Nível de Abstração (**Subseção 4.3.2**) e da Arquitetura de Implementação, através do diagrama de classe e diagramas de componentes, exibidos na **Subseção 4.2.3**.

4.3.1 Casos de Uso

Nessa subseção exibimos o diagrama de casos de uso (figura 14) onde estão especificadas as atividades para os atores que fazem parte do sistema.

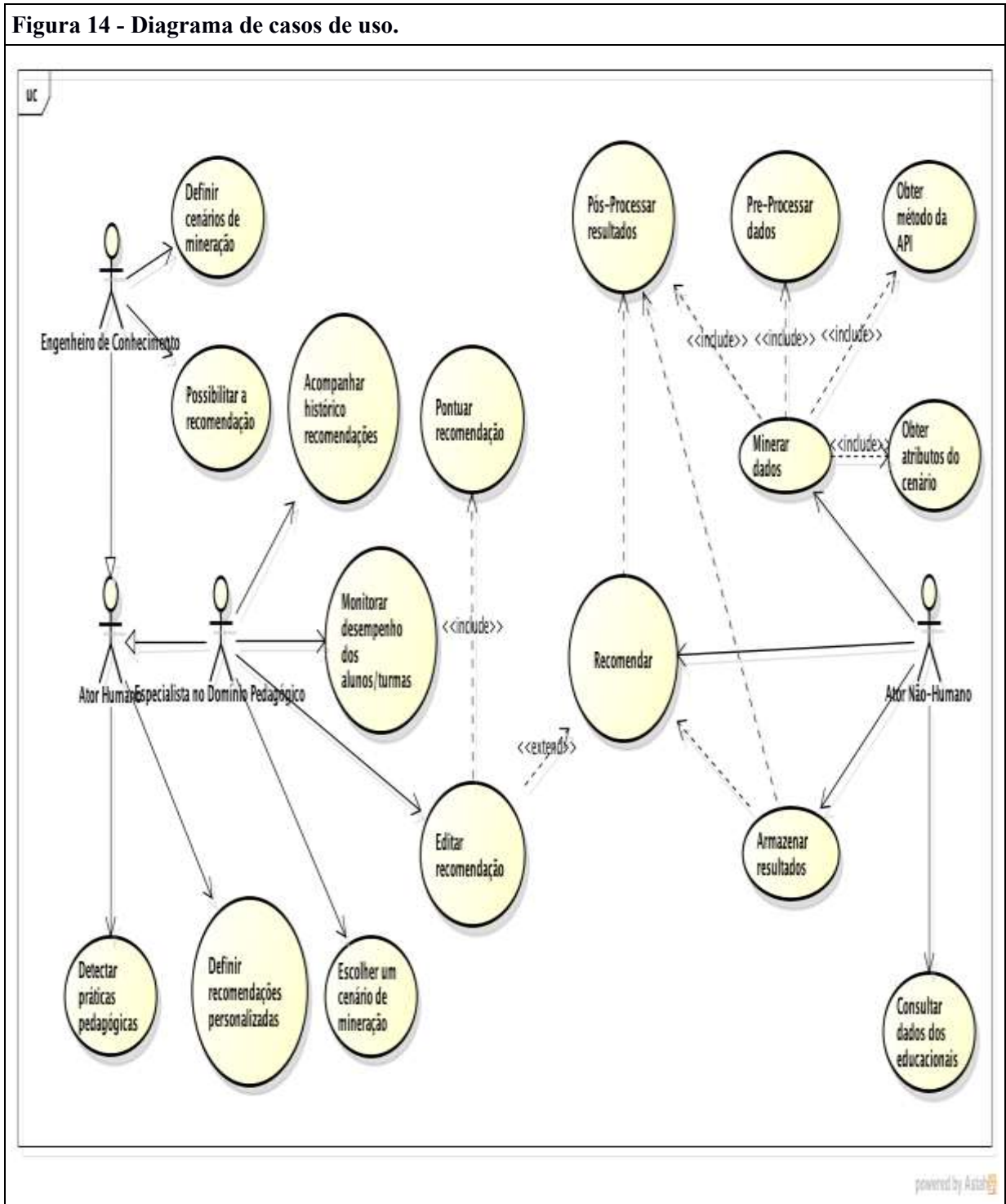
Os casos de uso descrevem as unidades funcionais providas pela ferramenta. Por meio deles representamos as responsabilidades do ator não humano, bem como as funcionalidades disponíveis para os atores humanos, em suas especializações (especialistas no domínio pedagógico/educacional, tecnológico e usuários finais). Listamos a seguir os requisitos funcionais e não funcionais da ferramenta:

- Requisitos Funcionais
 - Atores Humanos
 - **Detectar práticas pedagógicas:** responsabilidade dos atores humanos que consiste na detecção de situações pedagógicas fora dos limites esperados (exemplo: número elevado de faltas em uma turma).
 - **Definir cenários de mineração:** responsabilidade dos atores humanos que consiste no uso de técnicas de mineração de dados educacionais para descobrir os padrões responsáveis pelas práticas encapsular estratégias de mineração de dados, de forma a criar um processo de mineração guiado sem a ajuda dos especialistas.
 - **Escolher cenário de mineração:** funcionalidade disponível para que os usuários finais utilizem técnicas de mineração de dados para um contexto pedagógico definido.

- **Definir recomendações [pedagógicas] personalizadas:** responsabilidade dos atores humanos que consiste em, descobertos os padrões para as práticas identificadas, criar as recomendações que irão abordar essas práticas.
 - **Possibilitar a recomendação:** responsabilidade dos atores humanos que consiste em utilizar regras de associação para combinar as recomendações com as respectivas situações pedagógicas para as quais elas podem ser apropriadas.
 - **Editar recomendação:** funcionalidade disponível para que os usuários finais possam selecionar a recomendação mais apropriada, considerando a situação pedagógica que foi identificada e seu conhecimento sobre a turma ou os alunos.
 - **Monitorar desempenho dos alunos/turmas:** funcionalidade disponível para que os usuários finais possam acompanhar os resultados dos alunos após terem sido expostos a recomendações.
 - **Acompanhar histórico [das] recomendações:** funcionalidade disponível para que os usuários finais possam acompanhar, cronologicamente, quais recomendações já foram sugeridas.
 - **Pontuar recomendação:** funcionalidade disponível na interface com o usuário que, implicitamente, pontua as recomendações escolhidas pelos usuários finais.
- Ator Não-Humano
 - **Obter atributos do cenário [de mineração escolhido]:** a ferramenta obtém os atributos definidos para o cenário escolhido pelo usuário.
 - **Obter método da API:** a ferramenta obtém, das APIs disponíveis, os métodos que serão utilizados no processo de mineração dos dados e encaminhamento das recomendações.
 - **Consultar dados educacionais:** a ferramenta consulta os dados necessários para mineração, segundo especificações do cenário escolhido.
 - **Pré-processar dados [educacionais]:** a ferramenta trata os dados educacionais consultados de acordo com os requisitos do cenário escolhido pelo usuário.

- **Minerar dados [educacionais]:** a ferramenta minera os dados educacionais consultados, fazendo uso dos métodos obtidos das APIs disponíveis, de acordo com os requisitos do cenário escolhido pelo usuário.
 - **Pós-processar resultados:** a ferramenta trata o resultado da mineração dos dados educacionais de acordo com os requisitos do cenário escolhido pelo usuário.
 - **Recomendar:** a ferramenta busca, no repositório de dados, as recomendações disponíveis e avalia quais delas são apropriadas para os resultados obtidos.
 - **Armazenar resultados:** a ferramenta registra os detalhes de sua execução em log para consultas posteriores (exemplo: histórico de recomendações).
- Requisitos Não-Funcionais
 - **Arquitetura baseada em componentes**
 - **Implementação orientada a objetos**
 - **Escalabilidade**
 - **Manutenabilidade**
 - **Confiabilidade dos resultados da mineração**
 - **Facilidade de uso por parte dos usuários finais**
 - **Simplicidade de uso por parte dos usuários finais**

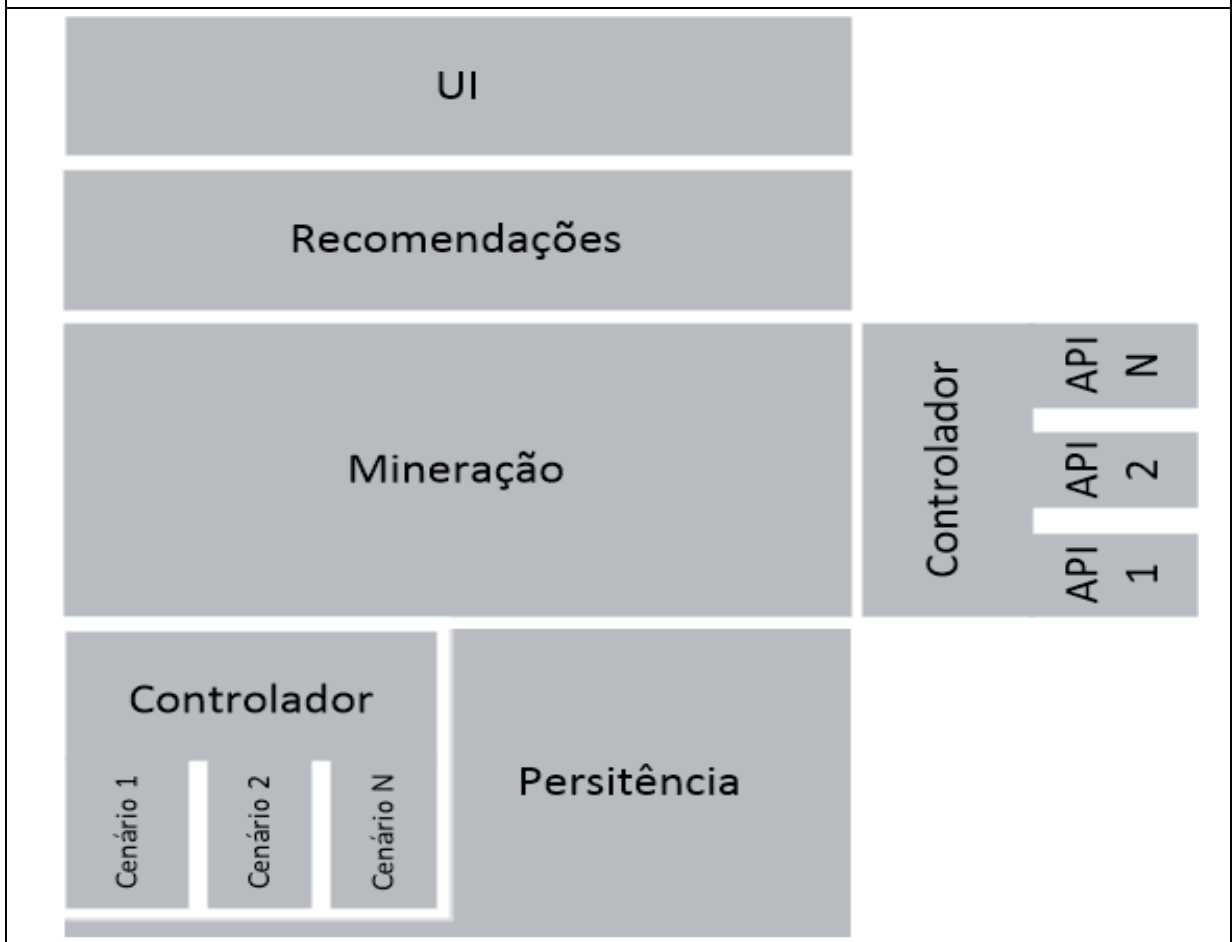
Figura 14 - Diagrama de casos de uso.



4.3.2 Arquitetura

Na figura 15 representamos a arquitetura de alto nível de abstração da ferramenta proposta. Nela exibimos os três módulos: **a interface com o usuário (UI)**, **módulo de mineração dos dados educacionais (Mineração)** e o **módulo de recomendação pedagógica (Recomendações)**, mencionados na **Seção 4.3** e que juntos constituem a ferramenta em questão. Detalharemos em seguida cada um desses módulos.

Figura 15 - Arquitetura de alto nível da Ferramenta de Recomendação.



1. **UI:** a Interface com o Usuário⁵⁰, no caso da ferramenta proposta é a **Interface com o Usuário**, ou seja, é o módulo através do qual os professores interagem com as funcionalidades disponíveis na ferramenta, permitindo-os ter acesso aos resultados da mineração dos dados interacionais dos alunos (orientada por um determinado cenário), recomendar conteúdo pedagógico aos alunos em situação que demande atenção, acompanhar o impacto dessas recomendações e acessar os registros de suas ações realizadas por meio dessa interface. Mais detalhes sobre a UI serão apresentados na **Subseção 4.3.3**.
2. **Recomendações:** módulo cuja responsabilidade é propor as recomendações que foram criadas, e adicionadas à ferramenta, pelos especialistas no domínio pedagógico, com base nos padrões pedagógicos descobertos. Esse módulo, também, é responsável por armazenar registros das recomendações realizadas para proporcionar o monitoramento e a avaliação da relevância de cada recomendação.

⁵⁰Em Inglês: User Interface, UI

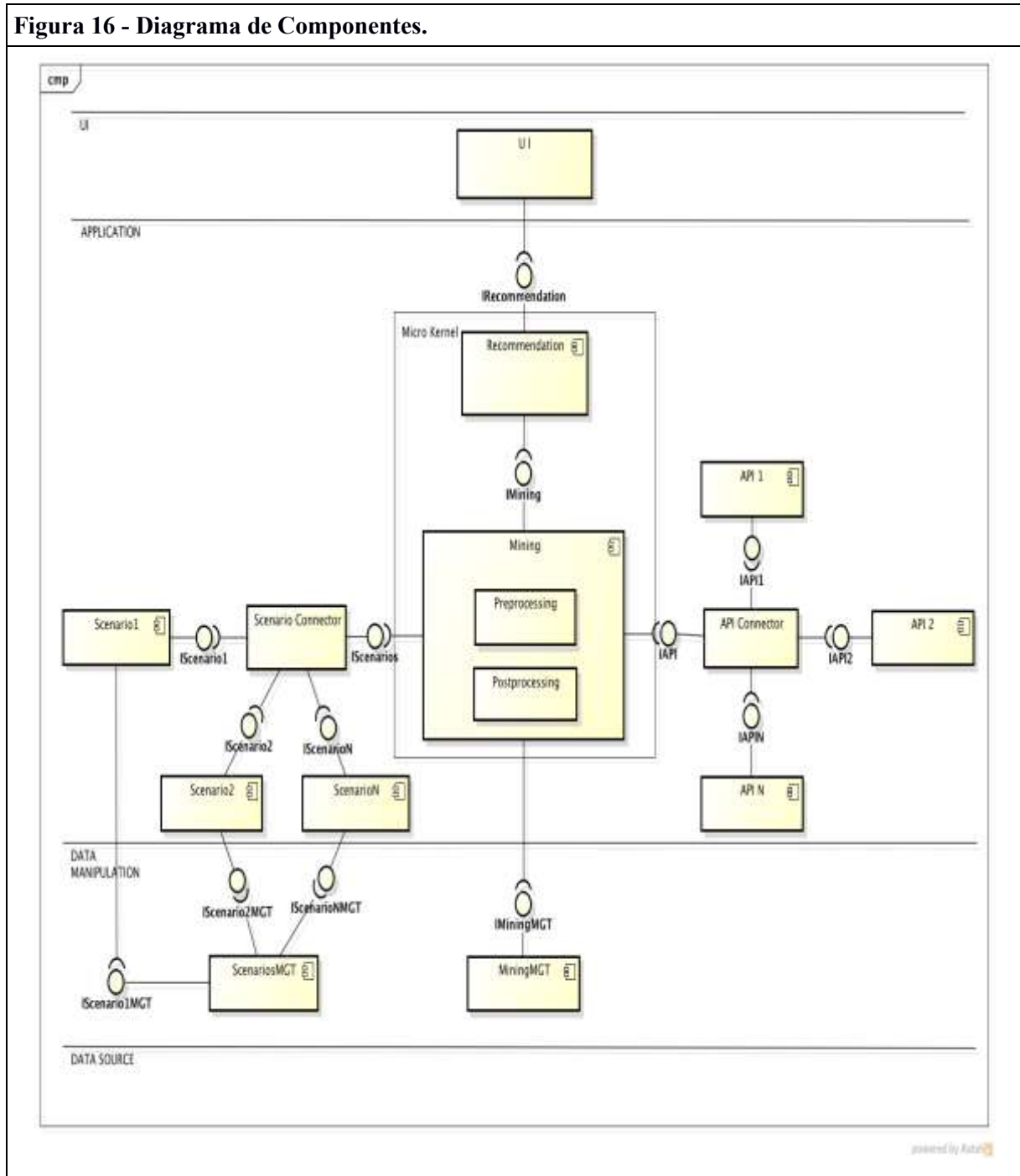
3. **Mineração:** módulo responsável pelo processo de mineração dos dados educacionais, o que inclui pré-processamento dos dados, a mineração dos dados educacionais (guiada por um determinado cenário de mineração escolhido) e o pós-processado dos resultados, tornando-os apreciáveis por atores humanos ou em formato apropriado para uso pelo módulo de recomendações. As ações desse módulo são orientadas pelos cenários de mineração, utiliza os dados educacionais obtidos do ambiente de aprendizagem e faz uso de API⁵¹ que, dentre outras funcionalidades, implementam os algoritmos de mineração necessários.
4. **Controlador de API:** funciona seguindo o padrão de projetos *Facade*, que equivale a uma classe cuja funcionalidade é oferecer uso comum para as classes de um pacote ou de um subsistema, tornando-o fácil de usar (METSKER, 2004). Nesse caso o pacote/subsistema corresponde a uma API.
5. **API:** as API são um meio (i.e.: interface) de comunicação entre componentes de software os quais contêm implementações de terceiros para funcionalidades úteis ao sistema. Por exemplo, a API dos algoritmos de mineração de dados do WEKA® (detalhado no **Apêndice B**).
6. **Controlador de Cenários:** assim como os demais controladores, funciona como um *Facade*, oferecendo uso comum dos Cenários de Mineração implementados, pelo módulo de mineração.
7. **Cenários:** contém a implementação para cada **Cenário de Mineração de Dados**, conforme detalhado na **Seção 4.2**.
8. **Persistência:** contém os métodos para gerenciamento dos dados em suas respectivas fontes/repositórios.

4.3.3 Arquitetura de Implementação

A partir do momento que definimos a arquitetura de alto nível da ferramenta proposta, passamos a definir uma arquitetura de implementação, descrita no diagrama de componentes na figura 16.

⁵¹ Sigla em Inglês para: Application Programming Interface (Interface de Programação de Aplicação), que é um meio (i.e.: interface) de comunicação entre componentes de software. Definição obtida de (WIKIPEDIA.COM) em 10/04/2013, acessível em: https://en.wikipedia.org/wiki/Application_programming_interface.

Figura 16 - Diagrama de Componentes.



O desenvolvimento da ferramenta foi feito em Java®⁵², uma linguagem de programação que utiliza os conceitos de Orientação a Objetos.

A capacidade de reuso de código, escalabilidade e a independência de algumas partes que constituiriam o sistema foram definidas como requisitos do projeto (vide subseção 4.3.1), por esse motivo a ferramenta foi desenvolvida seguindo os princípios do **Processo de**

⁵² Disponível em: <http://www.java.com/en/>

Desenvolvimento Baseado em Componentes que, de acordo com (SOMMERVILLE, 2011), é o processo de definir, implementar e integrar, em sistemas, componentes independentes ou fracamente acoplados. Segundo o mesmo autor a importância de tal abordagem aumentou bastante devido à sua capacidade de promover o reuso, o que está em consonância com o nosso requisito e justifica a sua utilização.

Os **Componentes** são abstrações de alto nível os quais são definidos por suas interfaces (SOMMERVILLER, 2011) e para alcançar esse requisito utilizamos o modelo **COSMOS**⁵³, o qual define as especificações e implementação de componentes, conectores, configurações arquiteturais e componentes compostos (GAYARD, 2008). Descreveremos a seguir cada um dos componentes que constitui a ferramenta. O código anotado dos principais componentes encontram-se no **Apêndice D**.

O GERENCIADOR DA MINERAÇÃO (*MINING MGT*)

Os componentes do tipo *gerenciador* têm como função controlar as operações entre um determinado componente e a fonte de dados. Nesse caso o componente é o **Componente de Mineração** e seu gerenciador torna possível consultar resultados de minerações anteriores, ou que novos resultados sejam armazenados na fonte de dados.

GERENCIADOR DOS CENÁRIOS (*SCENARIOS MGT*)

Esse componente controla as operações com os dados entre os cenários e a fonte de dados, permitindo consultar os dados definidos na especificação do mesmo.

O CONECTOR DE CENÁRIOS (*SCENARIOS CONNECTOR*)

Os componentes do tipo *conector* funcionam como um *Facade*, oferecendo uso comum para os componentes aos quais se comunica. No caso da ferramenta proposta, o conector de cenários torna possível o uso dos **Cenários** pelo **Componente de Mineração dos Dados Educacionais** que, como já mencionado anteriormente nessa dissertação (vide **Seção 4.2**), são responsáveis por orientar o processo de mineração dos dados.

⁵³ Em Inglês: COmponent System MOdel for Software architectures (Modelo de Sistema de Componentes para Arquiteturas de Software).

Outro objetivo do conector de cenários é o de possibilitar que a adição de novos cenários de mineração, bem como a remoção e atualizações dos cenários já existentes, seja feita de uma forma mais simples.

O CONECTOR DE API (*API CONNECTOR*)

Possui o mesmo propósito geral do conector de cenários, permitindo que o componente de mineração faça uso comum das *API* vinculadas a esse conector e torna possível a adição de novas *API* e a remoção/atualização daquelas já existentes.

O MICRO KERNEL (*MICROKERNEL*)

A arquitetura de *microkernel* oferece mínimo possível de software para que seja possível prover os mecanismos básicos necessários para implementar uma determinada aplicação (MINIX.ORG).

O *microkernel* é constituído dos componentes essenciais ao sistema, que são o **Componente de Mineração dos Dados Educacionais** e o **Componente de Recomendação**, sobre os quais trataremos abaixo.

O COMPONENTE DE MINERAÇÃO DOS DADOS (*MINING*)

Este componente é responsável por pré-processar os dados educacionais, descobrir os padrões de interesse pedagógico e, por fim, pós-processar os dados para serem utilizados pelo componente de recomendação, como consta no **Processo de Recomendação Pedagógica**, apresentado na **Seção 4.1**. Ele é guiado pelas especificações definidas em cada **Cenário de Mineração** (mais detalhes sobre as especificações dos cenários, vide Tabelas 5, 6 e 7), que descrevem os cenários criados e especificam como cada tarefa deve ser executada e, ainda, quais *API* devem ser utilizadas.

O processamento de todos os cenários disponíveis é o que denominamos de **Ciclo de Mineração dos Cenários** (i.e.: processamento da fila de cenários de mineração). Essa atividade consome um tempo considerável e faria com que os usuários tivessem que aguardar um período considerável por um resultado. Por essa razão decidimos que o Processo de Mineração ocorrerá de forma *off-line* (i.e.: iniciado e processado automaticamente, independente de solicitação dos usuários), e os ciclos ocorrerão de forma iterativa, o que significa que o final de um ciclo agenda a permissão para o início do próximo ciclo.

O COMPONENTE DE RECOMENDAÇÃO (*RECOMMENDATION*)

Esse componente tem por objetivo buscar relação entre os resultados provenientes da mineração dos cenários, com regras no formato **IF ___ THEN ___**, e os dados interacionais dos alunos. Combinações positivas identificam alunos em determinada situação pedagógica e os torna elegíveis a receber uma recomendação pedagógica.

Para que essas regras sejam recomendadas é preciso que já existam interações com o ambiente de aprendizagem e que sejam em número suficiente para que delas possam ser detectadas práticas, com posterior descoberta dos padrões responsáveis por sua ocorrência.

Para essa finalidade usamos dados decorrentes do curso de Espanhol lecionado através do **UFAL Línguas**. Os resultados originaram árvores de decisão e regras de associação que foram transformadas em um questionário contendo detalhes sobre as práticas detectadas associadas a possíveis recomendações. Tal questionário foi avaliado por diversos professores que pontuaram as recomendações apresentadas de acordo com o quão apropriado elas eram (i.e.: relevância dentro do contexto de uma situação pedagógica). Com base nessa avaliação as regras de associação foram criadas e adicionadas ao sistema (esse estudo de caso encontra-se descrito, com maiores detalhes, no **Capítulo 5**).

As recomendações, conforme mencionado na **Subseção 4.1.3**, utilizam recursos educacionais presentes no ambiente e que foram propostos como recomendação no questionário (utilizado no estudo de caso e que se encontra no **Apêndice C**).

As recomendações são apresentadas para o aluno no formato exposto tabela 8 e são acompanhadas por um link para o **OA**, cujos objetivos são: i) facilitar o acesso ao objeto de aprendizagem e ii) monitorar a realização da recomendação por parte do aluno.

Tabela 8 - Formato de apresentação, para o aluno, das recomendações pedagógicas.

Ação	Recurso Educacional	Subconjunto OA	Localização
Ler	Livro A	Páginas B, C e D	Acessível em <u>Link E</u>
Responder	Exercício F	Questões G e J	Acessível em <u>Link K</u>

A INTERFACE DO USUÁRIO (*UI*)

A Interface com o Usuário permite que professores tenham acesso às recomendações geradas para cada cenário que foi minerado. No estudo de caso apresentamos algumas telas da Interface do Usuário, as quais são referidas nessa subseção.

É através dessa interface que o **Processo de Recomendação Pedagógica** é apresentado aos professores (vide figura 17) descrevendo, de forma geral, como os resultados são obtidos através desse processo.

Figura 17 - Tela de apresentação do Assistente de Gestão de Aprendizagem.

Olá, prof. Alan Pedro! Seja bem-vindo ao Assistente de Gestão de Aprendizagem

O Processo de Recomendação Pedagógica é capaz de descobrir informações úteis, através da análise dos dados interacionais dos alunos, contando com a sinergia da Inteligência Humana e da Inteligência Artificial, tendo como a finalidade de prover apoio à tomada de decisões pedagógicas, para promover o aprimoramento da experiência de aprendizado para alunos e turmas.

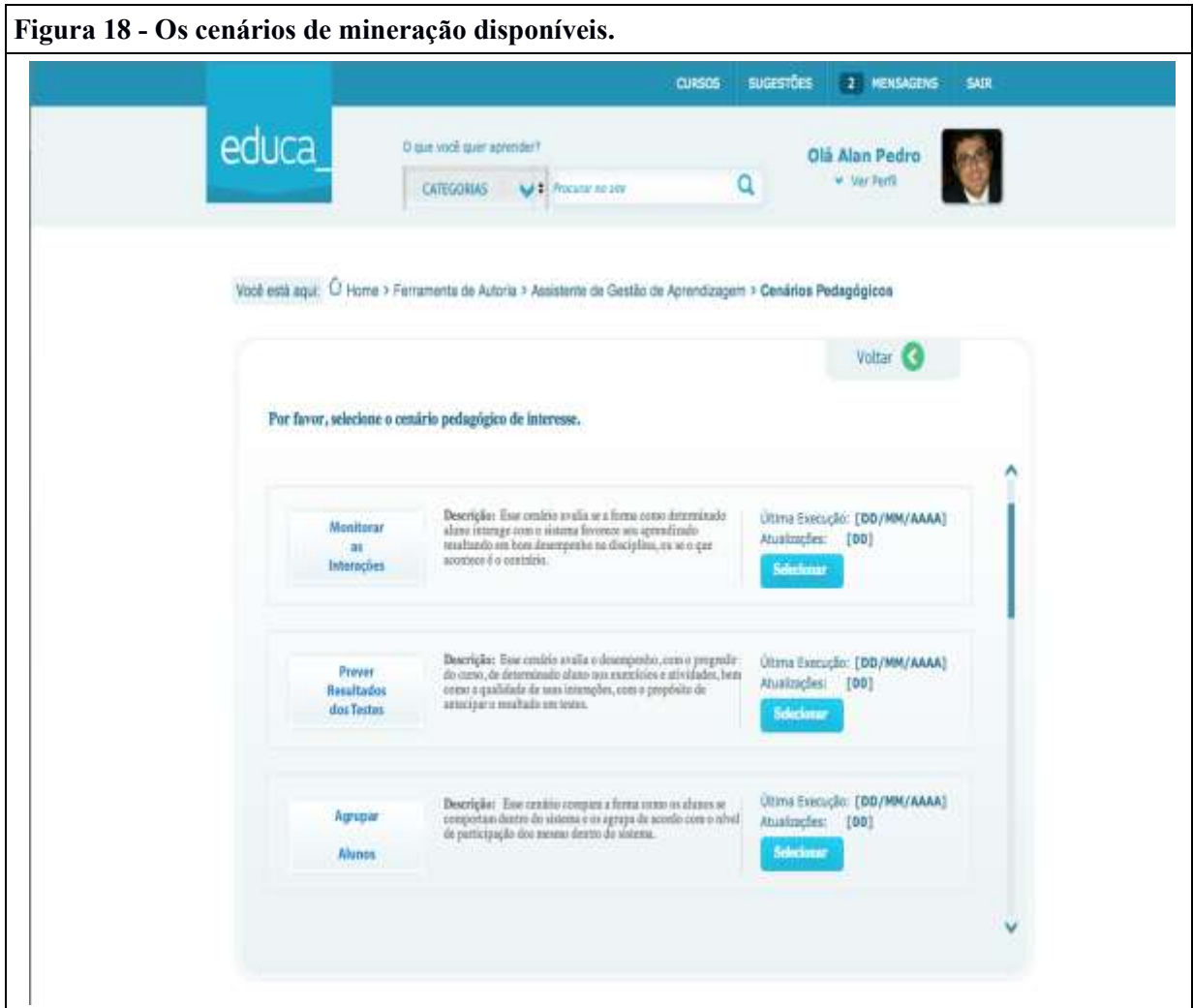
[Saber Mais...](#)

- 1 Detectar Práticas
- 2 Descrever Práticas
- 3 Recomendar
- 4 Monitorar e Avaliar

Iniciar

Na etapa seguinte são exibidos os **Cenários de Mineração** disponíveis (figura 18), contendo informações sobre o propósito pedagógico de cada um, bem como os dados sobre a atualização dos resultados da mineração *off-line* e estatísticas de uso de cada um deles.

Figura 18 - Os cenários de mineração disponíveis.



Escolhendo-se um dos cenários, os resultados mais atuais da mineração *off-line* são exibidos, consistindo dos dados do aluno, juntamente com **5 opções de recomendação** para cada aluno (figura 19), ordenadas pelo **valor de relevância** (tabela 10, localizada no **Apêndice C**), sendo a **5ª opção proposta randomicamente**, dentre todas as recomendações disponíveis, e oferecida ao professor. A decisão final de enviar, ou não, as recomendações sugeridas, é do professor, o que implicará em pontuação para avaliar a relevância das recomendações.

Figura 19 - Selecionando uma das recomendações pedagógicas propostas.

The screenshot shows the Educa system interface. At the top, there is a navigation bar with 'CURSOS', 'SUGESTÕES', '2 MENSAGENS', and 'SAIR'. Below this, the 'educa' logo is on the left, and a search bar with 'O que você quer aprender?' is in the center. On the right, the user's name 'Olá Alan Pedro' and a profile picture are displayed. The main content area shows a breadcrumb trail: 'Você está aqui: Home > Ferramenta de Autoria > Assistente de Gestão de Aprendizagem > Resultados (Cenário 1)'. Below the breadcrumb, there is a 'Voltar' button. The main instruction reads: 'Selecione a recomendação mais apropriada para cada aluno e clique em Iniciar. As recomendações escolhidas serão enviadas para os respectivos alunos.' Below this is a table with columns: '#', 'Aluno', 'Turma', 'Matrícula', and 'Recomendações'. The table contains three rows for 'Aluno A', 'Aluno B', and 'Aluno C', all in 'Espanhol A' class. The 'Recomendações' column has dropdown menus for each student. A dropdown menu is open for 'Aluno C', showing options: 'AtividadeExtraClas', 'Refazer/ProvasPenalidade(50)', 'Recomendar', 'ExerciciosExtra', 'ReverVideosUnidades[2,3]', 'ConteudoExterno(LNK, Pronombres)', 'RefazerExerciciosPenalidade(0)', and 'AtividadeNivel(Fácil)'. The 'ReverVideosUnidades[2,3]' option is selected. An 'Iniciar' button is at the bottom right of the table area.

#	Aluno	Turma	Matrícula	Recomendações
1	Aluno A	Espanhol A	000001	AtividadeExtraClas
2	Aluno B	Espanhol A	000002	Refazer/ProvasPenalidade(50)
3	Aluno C	Espanhol A	000003	ReverVideosUnidades[2,3]

Através da Interface com o Usuário é possível monitorar e avaliar os alunos que receberam uma recomendação. Nessa etapa avalia-se a diferença entre as médias do desempenho através das interações com os recursos normais de uma unidade ou período, e do desempenho obtidos nos recursos recomendado (ponderadas pelo tipo e dificuldade dos recursos educacionais considerados). O objetivo é avaliar o desempenho do aluno após a recomendação, conforme pode ser visualizado na figura 20.

Figura 20 - Monitoramento e avaliação do impacto das recomendações no desempenho dos alunos.

The screenshot displays the Educa system interface. At the top, there is a navigation bar with 'CURSOS', 'SUGESTÕES', '2 MENSAGENS', and 'SAIR'. Below this, the 'educa' logo is visible on the left, and a search bar with the text 'O que você quer aprender?' is in the center. On the right, the user's name 'Olá Alan Pedro' and a profile picture are shown. The main content area shows a breadcrumb trail: 'Você está aqui: Home > Ferramenta de Autoria > Assistente de Gestão de Aprendizagem > Histórico de Recomendações'. Below the breadcrumb, there is a 'Voltar' button with a left arrow. The 'Histórico de Recomendações' section contains a table with the following data:

#	Matrícula	Aluno	Turma	Última Atualização	Progresso	Recomendação
1	000001	Aluno A	Espanhol A	DDMMWAAAA	3%	Atividade Extra via Chat
2	000002	Aluno B	Espanhol A	DDMMWAAAA	5%	Realizar Provas (Penalidade = 50%)
3	000003	Aluno C	Espanhol A	DDMMWAAAA	1%	Rever Vídeos (Unidades = 2, 3)

At the bottom right of the table area, there is a blue 'Iniciar' button.

4.3.4 Execução dos Módulos de Mineração e Recomendação

Nessa seção descreveremos como ocorre a mineração dos dados educacionais com a posterior realização das recomendações pedagógica.

A primeira tarefa do Componente de Mineração é definir uma fila⁵⁴ para processamento dos Cenários de Mineração disponíveis, ou seja, em que ordem esses cenários serão processados. É necessário verificar se os referidos cenários estão disponíveis e ativos, antes de incluí-los à fila. Em seguida é iniciado o processamento dessa fila, que segue as especificações do cenário, apresentadas logo abaixo:

1. Obtenção dos dados dos respectivos repositórios.
2. Pré-Processar os dados.

⁵⁴ A ordem sequencial com que os cenários serão minerados.

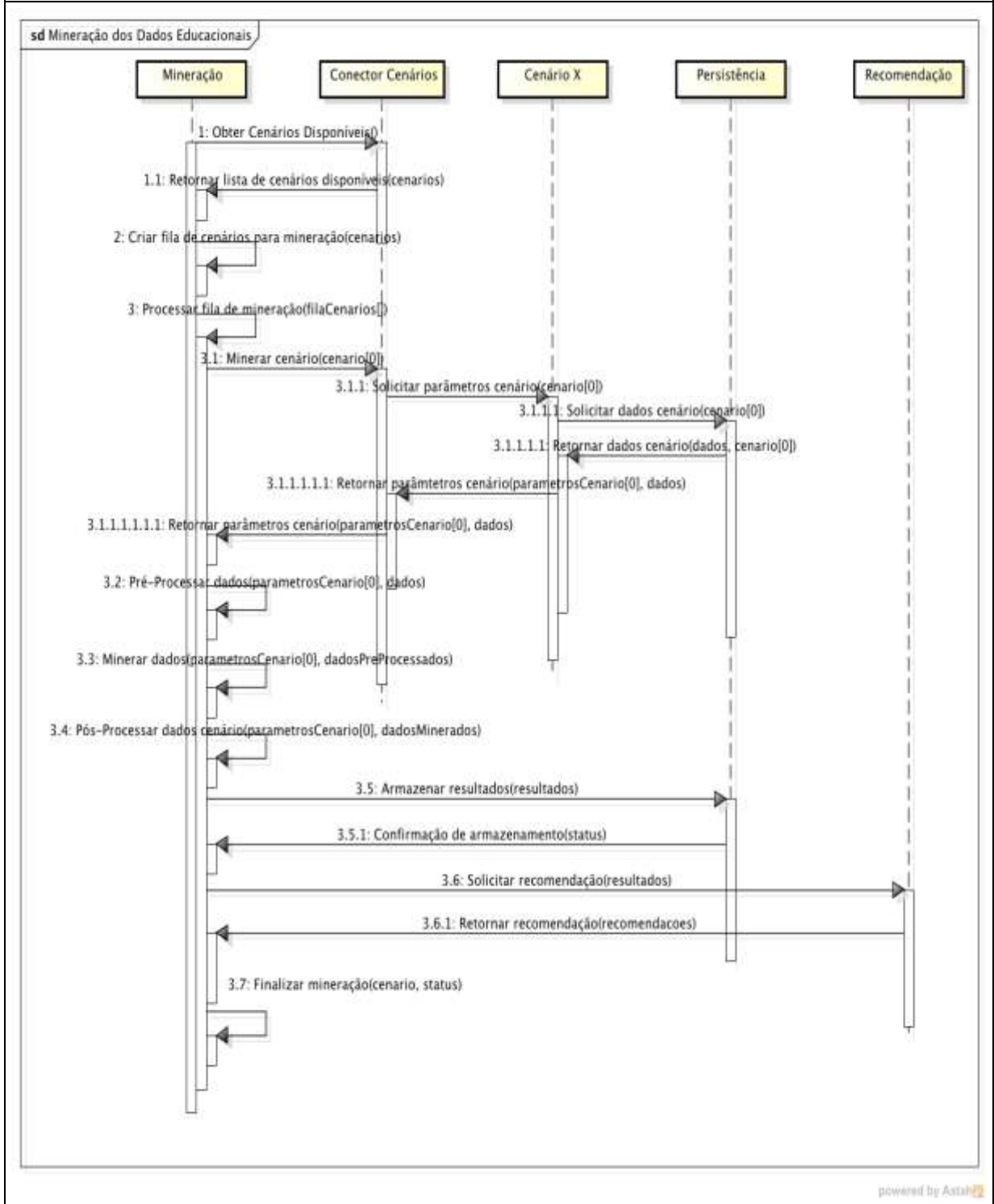
3. Criação do arquivo contendo os dados para a mineração (formato ARFF⁵⁵).
4. Solicitar, do controlados de API os algoritmos necessários para a mineração.
5. Minerar os dados utilizando o algoritmo solicitado e os parâmetros especificados no cenário.
6. Pós-Processar os resultados, tratando-os para que sejam utilizados pelo componente de recomendação.
7. O componente de mineração disponibiliza as recomendações através da Interface com o Usuário, atualizando os dados a cada novo resultado da mineração.

As etapas desse processo, e como elas se relacionam entre si, podem ser observadas no **Diagrama de Sequência** na figura 21.

Essas atividades correspondem ao processamento de um cenário. O ciclo de mineração dos cenários equivale à repetição de toda essa sequência de atividade para cada um dos cenários na fila e, como já mencionado anteriormente, isso ocorre de forma *off-line*, sendo o resultado enviado ao componente de recomendação após o final de cada processamento.

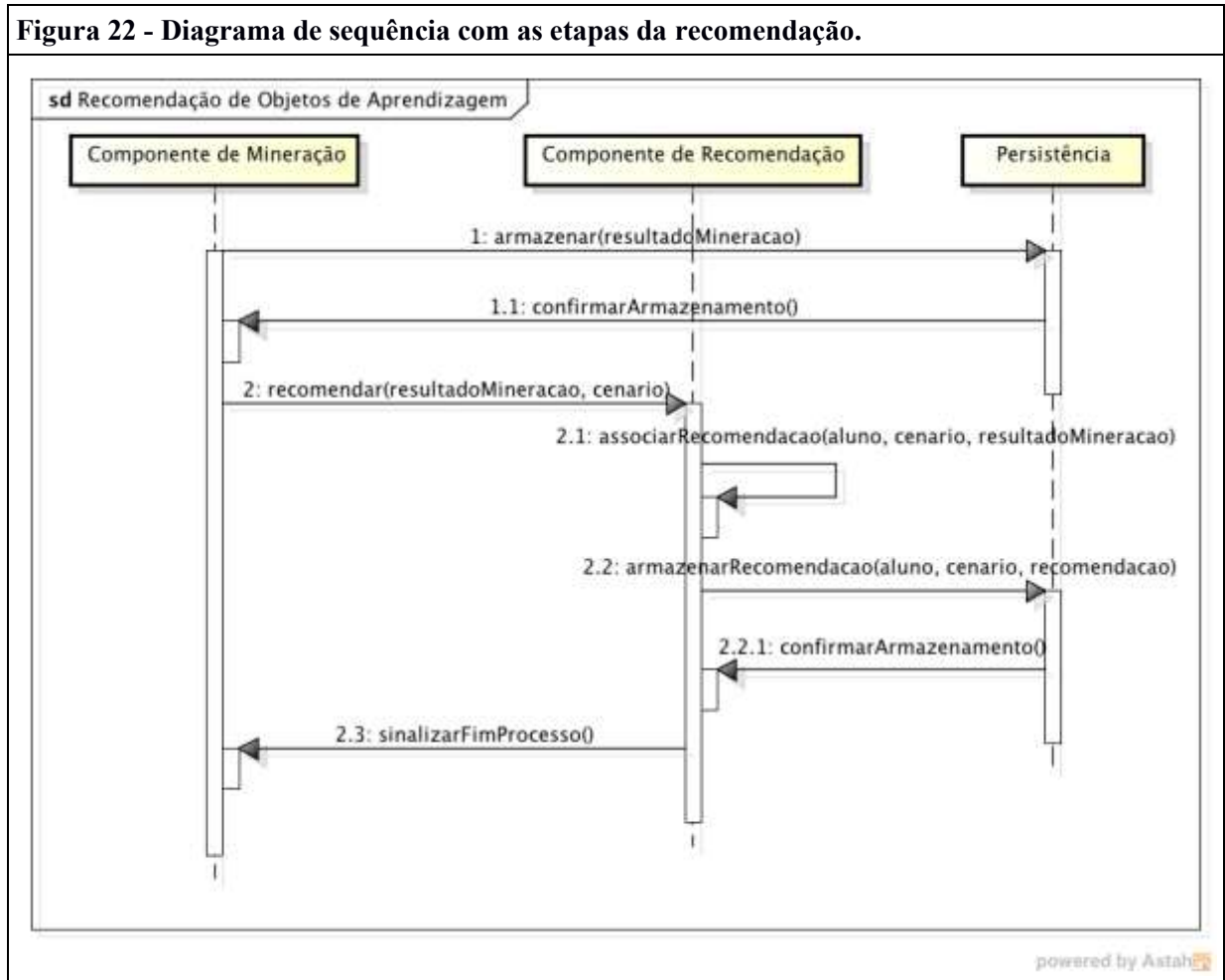
⁵⁵Formato nativo do WEKA®.

Figura 21 - Diagrama de sequência com as etapas da mineração dos dados educacionais.



Concluído o processo de mineração dos dados, os resultados são enviados ao Componente de Recomendação que se encarrega de **recomendar conteúdo, armazenar os dados da recomendação** para que possam ser acessados pela interface com o usuário. O diagrama da figura 22 ilustra esse processo.

Figura 22 - Diagrama de sequência com as etapas da recomendação.



5 ESTUDO DE CASO

Neste capítulo descreveremos a aplicação das técnicas apresentadas nessa dissertação. Mostraremos como o Processo de Recomendação Pedagógica e a ferramenta proposta, foram utilizados no contexto de um curso de língua Espanhola online, denominado **UFAL Línguas - Espanhol**. Tal curso teve um total de 2075 alunos inscritos, sendo que 200 foram matriculados. A representação e organização do conhecimento gerado pelo ambiente de aprendizagem foram feitas pelo uso de ontologias. Ao final do curso o repositório continha mais de **700 megabytes** de informações (cinco meses de curso), dispostas em, aproximadamente, **1220000 triplas**.

O curso teve caráter de extensão (Espanhol nível básico), e foi oferecido pela **Universidade Federal de Alagoas**, gratuitamente, com 100 vagas para alunos UFAL regularmente matriculados, e as demais para alunos concluintes do Ensino Médio (UFAL Línguas, 2012a; UFAL Línguas, 2012b), como demanda social.

O curso ocorreu na modalidade online, com duração de cinco meses (de outubro de 2012 a Fevereiro de 2013), foi constituído por 6 unidades. Sua parte pedagógica foi administrada por um professor e 8 tutores, os quais eram responsáveis pelo progresso do curso, bem como o acompanhamento dos alunos e correção das atividades e testes. Temos então uma proporção de 25 alunos, para cada tutor. Ao seu final os alunos concluintes receberam um certificado de **Espanhol Nível Básico**.

Descreveremos nas seções seguintes como aplicamos o **Processo de Recomendação Pedagógica** (PRP), bem como o modo de utilização da ferramenta proposta na dissertação.

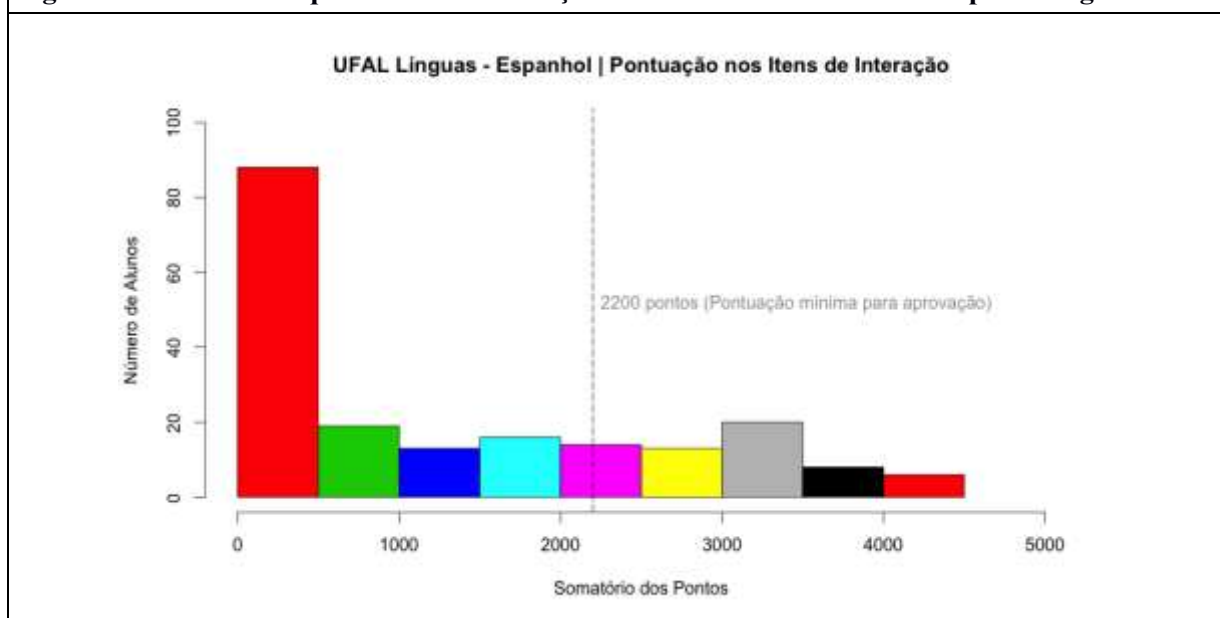
5.1 Detectar Práticas

Essa primeira etapa visou **Detectar Práticas Pedagógicas**, ou seja, situações do contexto pedagógico que estavam ocorrendo dentro do ambiente de ensino, e influenciaram o desempenho dos alunos e/ou das turmas.

Como mencionado anteriormente, o PRP (vide Subseção 4.1) utiliza a colaboração de atores humanos e artificiais. Para o caso em estudo, utilizamos o **Gráfico da Pontuação nos Itens de Interação** (figura 23), onde os resultados foram calculados da seguinte forma: atividades e exercícios respondidos corretamente receberam 10 pontos multiplicados pelo peso referente ao nível de dificuldade da atividade/exercício (fácil = 1, médio = 3 e difícil = 5). A isso se somou o resultado das provas (0 a 100 para cada prova realizada), acesso, a

quantidade de interações via chat e fórum (multiplicadas por 1, 20 e 10, respectivamente). O gráfico foi obtido por meio da análise dos dados extraídos do sistema (**ator não humano**), utilizando a ferramenta R⁵⁶. Através da visualização dos resultados foi possível notar uma grande concentração de resultados entre os valores 0 e 500 pontos, valores que caracterizam um baixíssimo aproveitamento no curso, insuficiente para a aprovação. Tal gráfico foi avaliado por especialistas no domínio pedagógico (**atores humanos**), em busca das mencionadas “Práticas Pedagógicas”.

Figura 23 - Gráfico de qualidade das interações dos alunos no ambiente de aprendizagem.



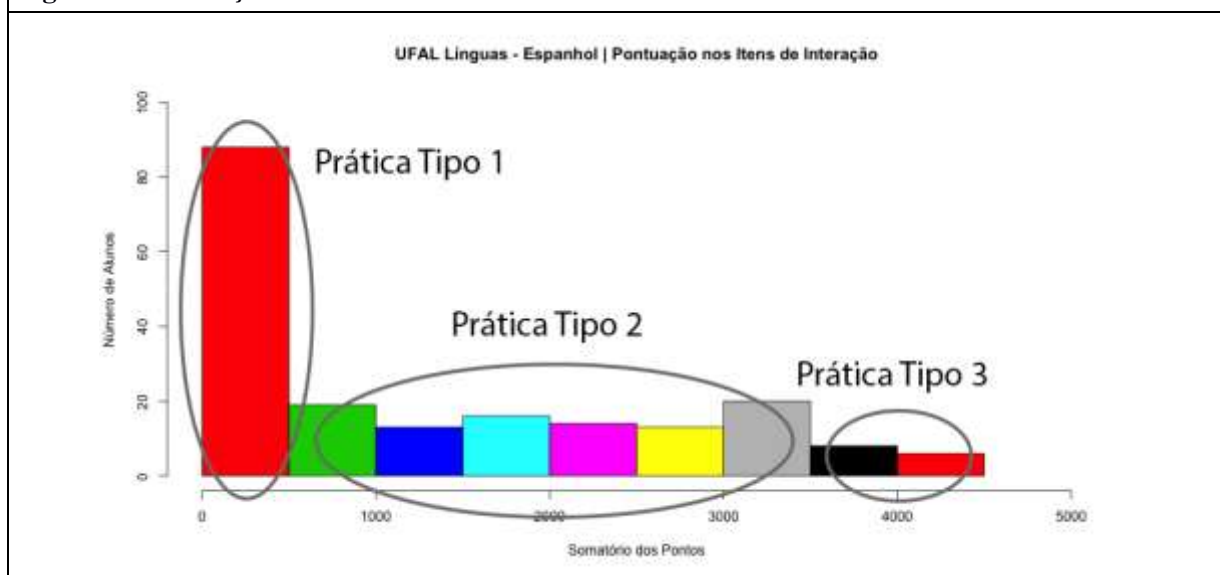
Na análise desse gráfico constatou-se a existência de três tipos distintos de práticas (figura 24) ocorrendo dentro do ambiente de aprendizagem:

1. Um tipo que leva a resultados bem abaixo daqueles que são esperados e que, infelizmente, culminam com desistências (provas realizadas com pontuação zero) e reprovação.
2. Um segundo tipo que leva a resultados medianos, os quais podem ser suficientes, ou não, para a aprovação no curso, mas nesse tipo residem os alunos com maior chance de serem "resgatados".
3. E um terceiro tipo, não tão comum, que leva os alunos a um nível de pontuação bastante elevado e, conseqüentemente, um aproveitamento bem melhor. Os

⁵⁶ Um sistema para análises estatísticas diversas, disponível para download em: <http://cran.r-project.org/>

detalhes desse tipo de prática podem ser identificados e estudados para que professores e tutores possam utilizá-los no auxílio a outros alunos.

Figura 24 - Detecção das Práticas.



A detecção dessas três práticas levantou aos seguintes questionamentos:

- O que **faz** um aluno **estar** em um desses tipos de prática?
- Que **dados** podemos coletar para **compreender** esses tipos de prática?
- Como **nomear/definir** os tipos de prática detectados?

A etapa seguinte tornará possível responder essas perguntas.

5.2 Descobrir Padrões

Nessa etapa buscamos os padrões relacionados às práticas que identificamos anteriormente, e utilizamos os resultados da detecção de práticas da etapa anterior. A primeira atitude tomada foi a de supor um motivo que justificasse as práticas detectadas (hipótese) e obter os dados que nos permitissem confirmar ou rejeitar essa suposição, considerando a capacidade de coleta do ambiente de aprendizagem.

Para que isso fosse realizado de forma organizada, foi necessário definirmos os **objetivos de mineração**, uma vez que vários fatores poderiam ser responsáveis por cada uma das práticas detectadas.

Partimos da premissa de que a qualidade das interações dos alunos dentro do ambiente de aprendizagem, ou seja, o quão bem esses alunos interagem com os recursos oferecidos pelo

ambiente, eram responsáveis pelas práticas. Diversas tentativas e experimentos com algoritmos de mineração de dados, e diferentes parâmetros, deram origem ao **Cenário 1: Qualidade das Interações dos Alunos** (tabelas 5, 6, 7 e Figura 30). O resultado final dessas experiências está relatado a seguir.

Inicialmente trataremos dos dados utilizados. Na tabela 9 temos uma listagem de algumas características dos dados utilizados para a mineração desse cenário, tais como: o nome da variável, se alguma dessas variáveis é subconjunto de outra (por exemplo, vemos na tabela que as atividades nível fácil, médio e difícil estão contidas na variável “atividades respondidas”), tipo de dado (numérico, *string*, booleano, categórico), e os valores possíveis (mínimo e máximo) que essa variável pode assumir. Essas características podem nos servir para planejar os passos de preparação, e de modelagem.

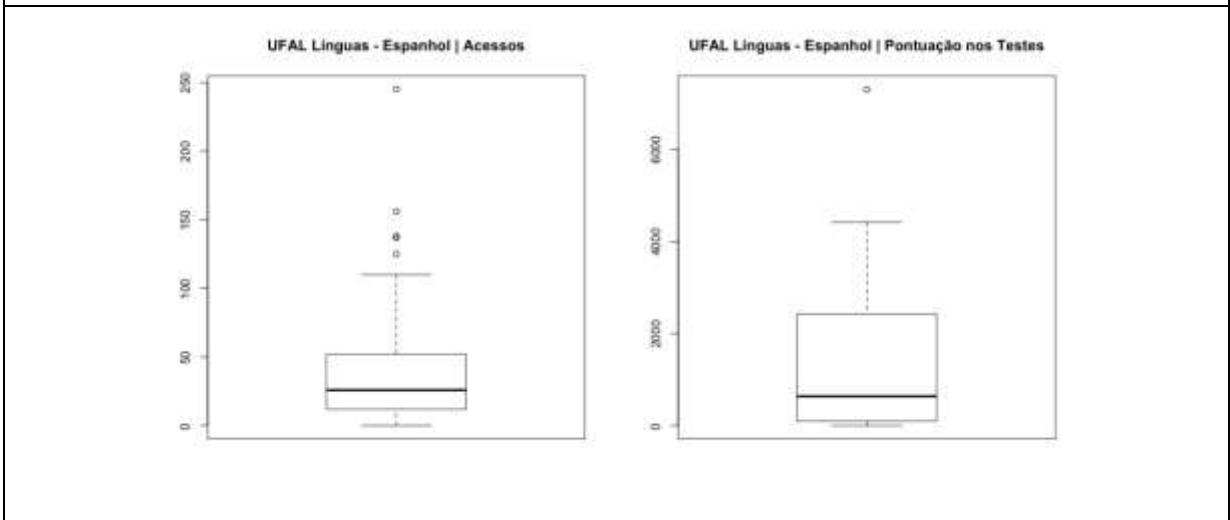
Tabela 9 - Dados do cenário Qualidade da Interação dos Alunos no Ambiente de Aprendizagem.

Nome da Variável	Relação	Tipo de Dado	Valores Possíveis
ID do Aluno		Texto	
Acessos ao Sistema		Numérico / Discreto	[0 ; 245]
Atividades Respondidas	A	Numérico / Discreto	[0 ; 13]
Atividades Nível Fácil	a	Numérico / Discreto	[0 ; 10]
Atividades Nível Médio	a	Numérico / Discreto	[0 ; 3]
Atividades Nível Difícil	a	Numérico / Discreto	[0 ; 0]
Respostas Corretas Nível Fácil	B	Numérico / Discreto	[0 ; 381]
Respostas Corretas Nível Médio	b	Numérico / Discreto	[0 ; 71]
Respostas Corretas Nível Difícil	b	Numérico / Discreto	[0 ; 23]
Interações via Fórum		Numérico / Discreto	[0 ; 7]
Interações via Chat		Numérico / Discreto	[0 ; 22]
Testes	C	Numérico / Discreto	[0 ; 57]
Resultado dos Testes	c	Numérico / Real	[0.0 ; 100.0]
Aprovação		Boolean	[true ; false]
Qualidade da Interação		Categórico	[baixa ; media ; alta]

No passo seguinte avaliamos a necessidade de tratar *outliers* e valores ausentes, normalizar, agrupar e/ou transformar os dados. Essa fase é denominada de pré-processamento dos dados e é importante, pois alguns tipos de dados ou não são aceitos, ou diminuem a eficiência dos algoritmos utilizados. Um passo importante, também, é que após a preparação os dados foram adicionados a um arquivo ARFF (formato aceito, nativamente, pela ferramenta de mineração WEKA). Foi necessário então avaliar vários aspectos estatísticos dos

dados (o *boxplot* apresentado na figura 25 é um exemplo dessa análise). Nele foi possível perceber a presença de alguns *outliers*, os quais foram removidos para evitar que sua influência distorcesse os resultados. O script R utilizado nessa análise está disponível (vide **Apêndice B**).

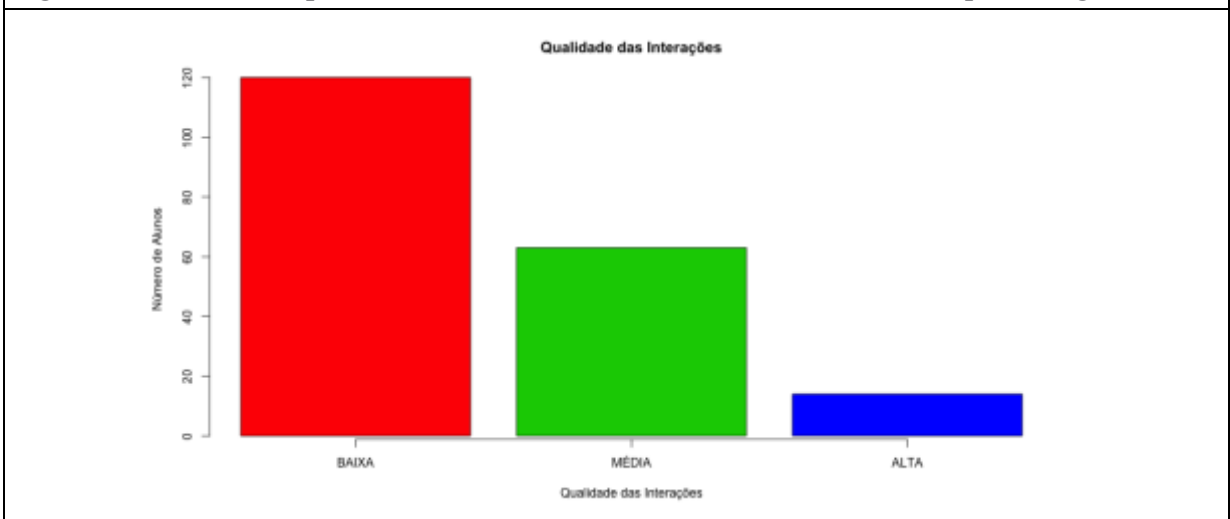
Figura 25: Boxplots dos dados acesso e pontuação nos testes.



O valor da classe "qualidade da interação" foi definida com base na variável "**Pontuação nos Itens de Interação**" e foi dividida três em faixas (figura 26):

- **Baixa:** de 0 a 1499 pontos (120 instâncias)
- **Média:** de 1500 a 3499 pontos (63 instâncias)
- **Alta:** de 3500 a 5000 pontos (14 instâncias)

Figura 26 - Gráfico de qualidade das interações dos alunos no ambiente de aprendizagem.



Os dados, originalmente em formato CSV, foram convertidos para ARFF que é o formato nativo do WEKA que foi utilizado no processo de mineração. Nosso objetivo foi gerar um "classificador", do tipo **árvore de decisão**, para que pudéssemos conhecer os critérios da classificação feita pelo WEKA. O classificador foi inferido utilizando o algoritmo C4.5, implementado no WEKA através do algoritmo J48 (utilizamos validação cruzada em 10 dobras, 10 fold cross-validation, que consiste em dividir os dados em 10 conjuntos de dados com tamanho $n/10$, treinar o algoritmo utilizando nove desses conjuntos e realizar os testes no conjunto de dados restante). Com isso fomos capazes de gerar um classificador com acurácia de 88.89%. A árvore resultante desse processo pode ser visualizada em forma textual na Figura 27 e em forma gráfica, na Figura 28.

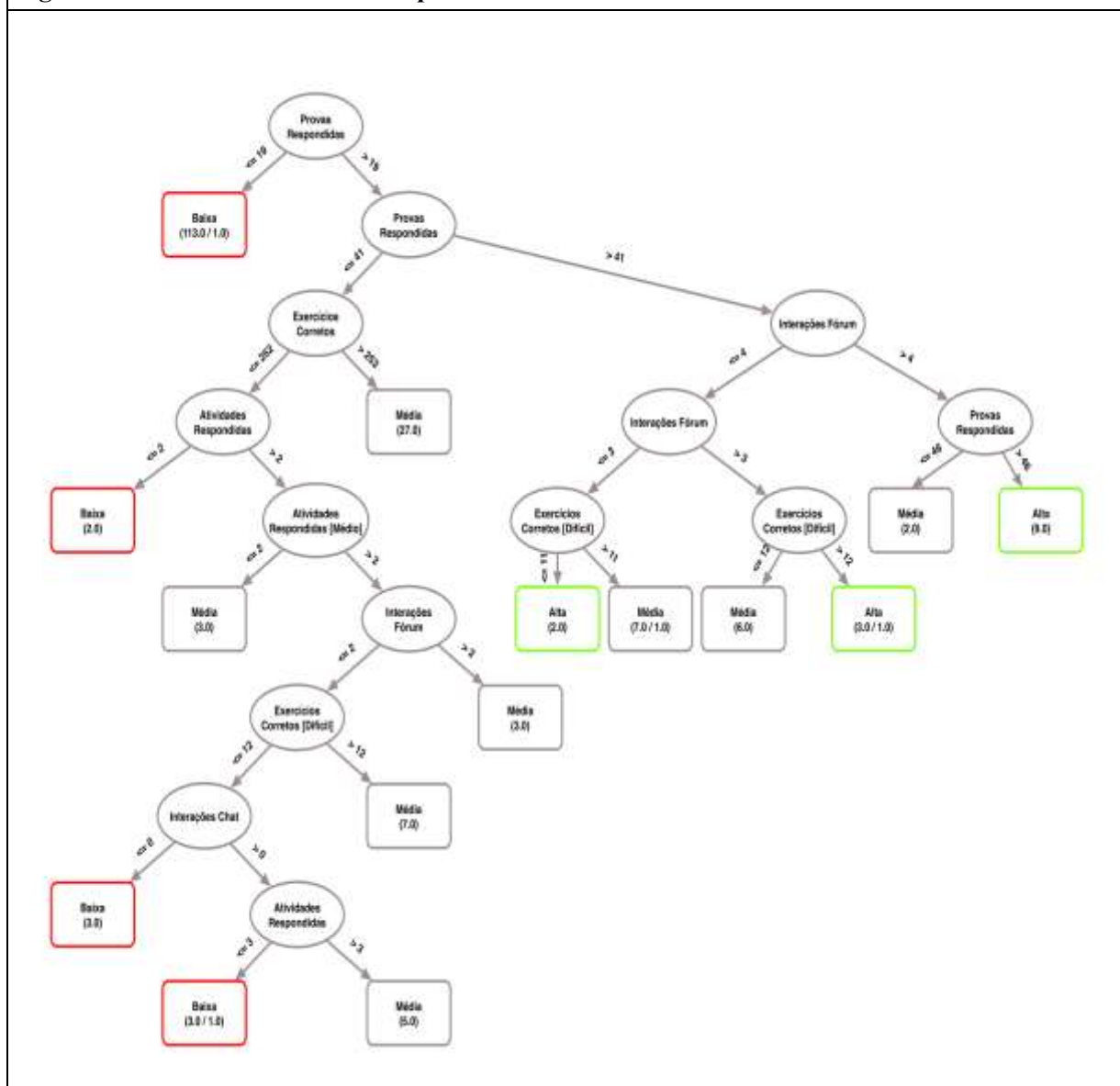
Figura 27 - Representação textual da árvore de decisão inferida pelo algoritmo C4.5.

```
J48 pruned tree
-----

ProvasRespondidas <= 19: BAIXA (113.0/1.0)
ProvasRespondidas > 19
|   ProvasRespondidas <= 41
|   |   ExerciciosCorretos <= 252
|   |   |   AtividadesRespondidas <= 2: BAIXA (2.0)
|   |   |   AtividadesRespondidas > 2
|   |   |   |   AtividadesRespondidasMedio <= 2: MEDIA (3.0)
|   |   |   |   AtividadesRespondidasMedio > 2
|   |   |   |   |   ExerciciosCorretosDificil <= 12
|   |   |   |   |   |   InteracoesChat <= 0: BAIXA (3.0)
|   |   |   |   |   |   InteracoesChat > 0
|   |   |   |   |   |   |   AtividadesRespondidas <= 3: BAIXA (3.0/1.0)
|   |   |   |   |   |   |   AtividadesRespondidas > 3: MEDIA (8.0)
|   |   |   |   |   |   |   ExerciciosCorretosDificil > 12: MEDIA (7.0)
|   |   |   |   |   ExerciciosCorretos > 252: MEDIA (27.0)
|   |   ProvasRespondidas > 41
|   |   |   InteracoesForum <= 4
|   |   |   |   InteracoesForum <= 3
|   |   |   |   |   ExerciciosCorretosDificil <= 11: ALTA (2.0)
|   |   |   |   |   ExerciciosCorretosDificil > 11: MEDIA (7.0/1.0)
|   |   |   |   InteracoesForum > 3
|   |   |   |   |   ExerciciosCorretosDificil <= 12: MEDIA (6.0)
|   |   |   |   |   ExerciciosCorretosDificil > 12: ALTA (3.0/1.0)
|   |   |   InteracoesForum > 4
|   |   |   |   ProvasRespondidas <= 46: MEDIA (2.0)
|   |   |   |   ProvasRespondidas > 46: ALTA (9.0)

Number of Leaves   :    14
Size of the tree   :    27
```


Figura 28 - Árvore de decisão com pontos de interesse circutados.



Notamos que ambos os formatos (textual e gráfico), não proporcionaram facilidade de interpretação para os professores e especialistas do domínio pedagógico, entretanto era essencial a participação dos mesmos no passo seguinte (dependente da análise dessa árvore de decisão). Como solução, resolvemos converter as questões pedagógicas identificadas, a partir dos resultados da mineração, em um questionário, que será abordado na descrição da próxima etapa.

Na tabela seguinte (tabela 10) listamos os parâmetros utilizados, conforme especificado na criação do cenário.

Tabela 10 - Parâmetros utilizados na mineração do cenário.

Variável	Valor	Descrição
Algoritmo	J48	Implementação do C4.5
Validação	10 fold cross-validation	
binarySplits (parâmetro)	<i>false</i>	Realiza a divisão de atributos nominais.
confidenceFactor (parâmetro)	0.5	Fator de confiança. Quanto menor, maior será a poda.
minNumObj (parâmetro)	2	Número mínimo de objetos por folha
reducedErrorPruning (parâmetro)	<i>false</i>	Maximiza a poda da árvore de forma de evite erros de classificação. Gera árvores mais simples
subtreeRaising (parâmetro)	<i>true</i>	Expande as sub-árvores, se possível.
unpruned (parâmetro)	<i>false</i>	Gera a árvore sem poda.
useLaplace (parâmetro)	<i>false</i>	Suaviza a quantidade de folhas, considerando funções mais gerais.
Acurácia (Resultado)	88,89%	

5.3 Recomendar

Na primeira iteração, devido ao problema de “cold start”^{57 58} foi necessário que fossem criadas, e adicionadas à ferramenta, algumas recomendações que já permitissem a aplicação do processo. Isso foi realizado através de um questionário (**Apêndice C**) para os especialistas do domínio pedagógico. O questionário foi gerado de uma análise da árvore de decisão (Figuras 27 e 28), obtida da etapa anterior. Nela identificamos conjuntos de interações que levam os alunos a terem a qualidade de suas interações classificadas como “baixa”, “média” ou “alta”. O questionário foi feito para professores e lista algumas recomendações para cada situação pedagógica identificada, solicitando que eles pontuem as recomendações de 1 a 5, sendo: **1)** recomendação totalmente inapropriada, **2)** recomendação pouco apropriada, **3)** recomendação razoável, **4)** recomendação apropriada e **5)** recomendação muito apropriada. Também foi permitido que os professores descrevessem outra recomendação, que eles julgassem adequada, mas que não constava no questionário.

⁵⁷ Início frio ou partida fria.

⁵⁸ É uma questão relacionada com Sistemas de Recomendação (SR), puramente, do tipo Filtro Colaborativo, onde itens novos não são considerados pelo SR por não possuírem “evidência de uso” adequada (IAQUINTA, 2011).

As recomendações apresentadas foram transformadas em **regras de associação do tipo IF ___ THEN ___**, e as notas atribuídas na avaliação dos professores foram utilizadas para gerar um **Valor de Relevância** para cada recomendação. As situações pedagógicas, as regras (i.e.: as recomendações) e seus respectivos valores de relevância foram adicionados ao sistema e usados para que a ferramenta já possuísse regras cadastradas, sendo capaz de propor recomendações.

A tabela 11 (localizada no **Apêndice C**) contém algumas das regras de associação para as recomendações e seus valores de relevância, para esse estudo de caso.

A Figura 31 exibe os resultados da mineração do cenário 1 (Qualidade das Interações dos Alunos), mostrando que alguns alunos precisavam de recomendações que proovessem melhoria na qualidade de suas interações. É possível notar, também, as recomendações disponíveis.

5.4 Monitorar e Avaliar

Infelizmente, por termos lidado com os dados finais do curso, não foi possível monitorar e avaliar o impacto das recomendações no desempenho dos alunos e o processo de monitoramento e avaliação das recomendações em si, ficou por conta dos resultados do questionário, apenas.

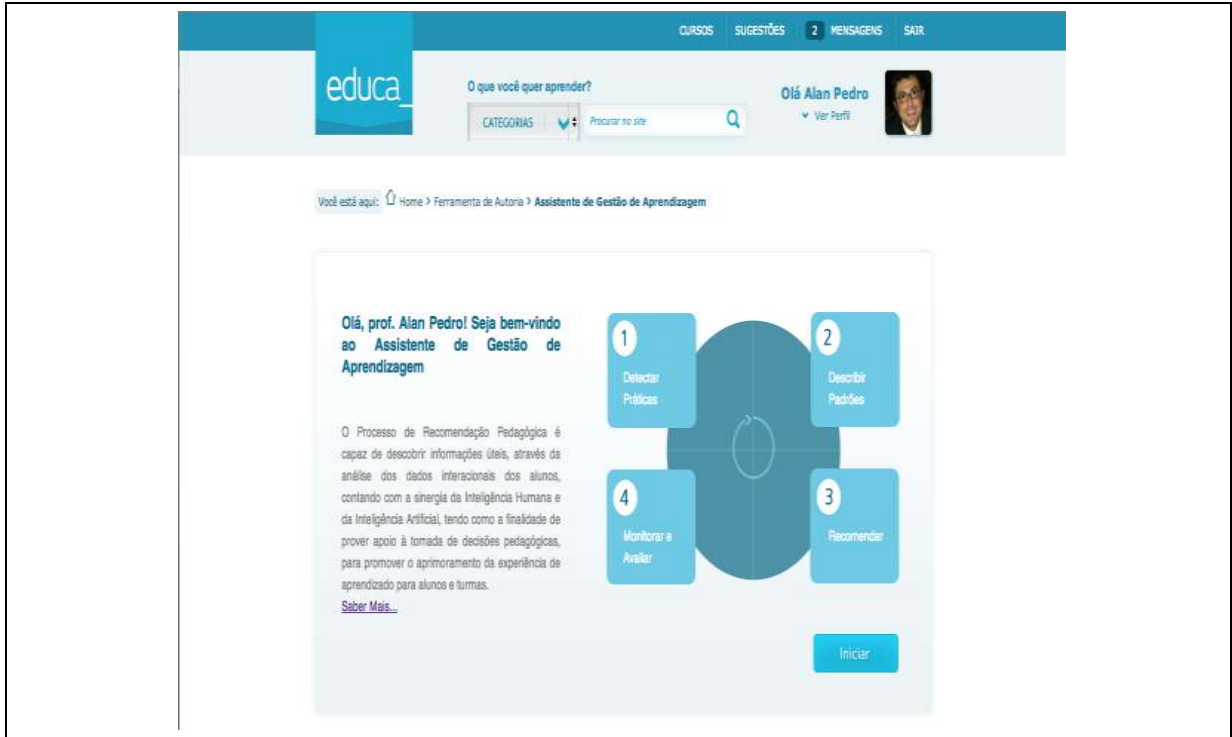
Uma simulação do que esperamos, parcialmente, obter dessa etapa encontra-se na Figura 32.

5.5 A Interface com o Usuário

Nessa seção exibimos a interface com o usuário final (i.e.: o professor), através do qual os professores podem propor recomendações pedagógicas, personalizadas de acordo com os cenários de mineração disponíveis.

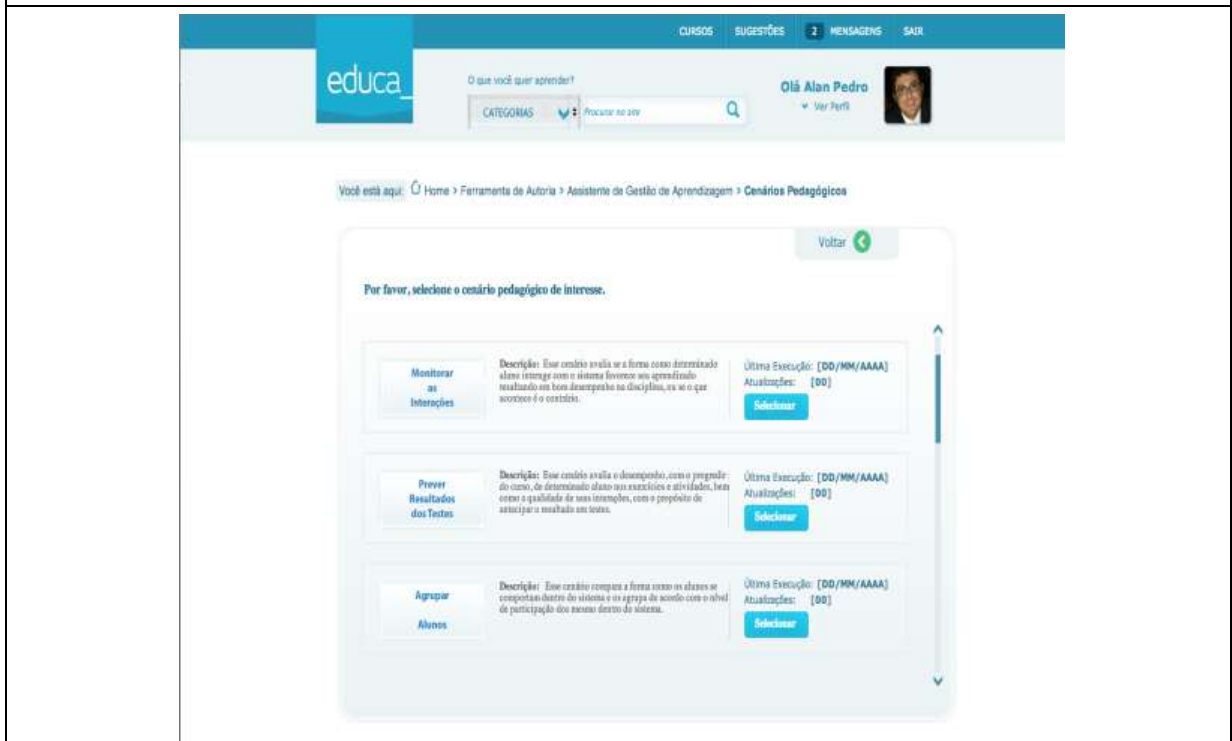
A Figura 29 exibe a tela de apresentação da Interface com o Usuário, onde o Processo de Recomendação Pedagógica é apresentado, podendo o professor obter mais detalhes se assim desejar.

Figura 29 - Tela de apresentação do Assistente de Gestão de Aprendizagem.



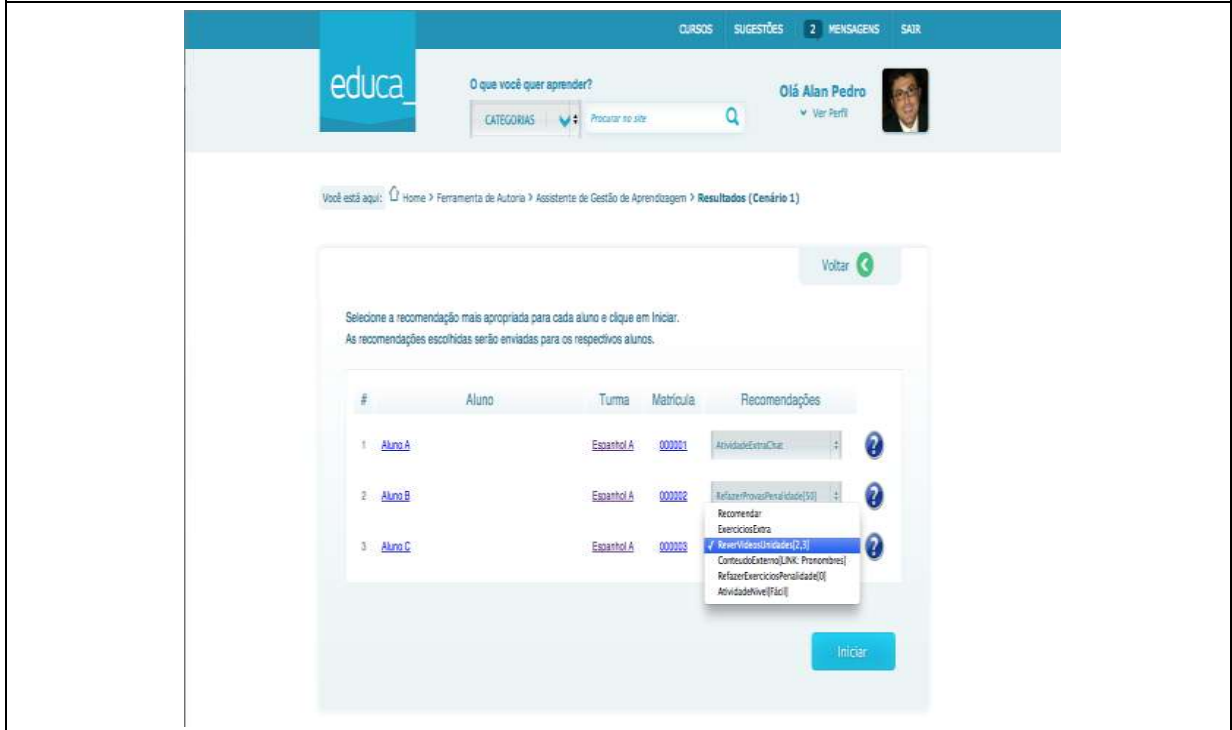
A Figura 30 exibe uma lista de cenários disponíveis para a mineração dos dados educacionais, com uma descrição do propósito pedagógico de cada um (informação de interesse para o professor) e dados sobre atualização da mineração *off-line*.

Figura 30 - Os cenários de mineração disponíveis.



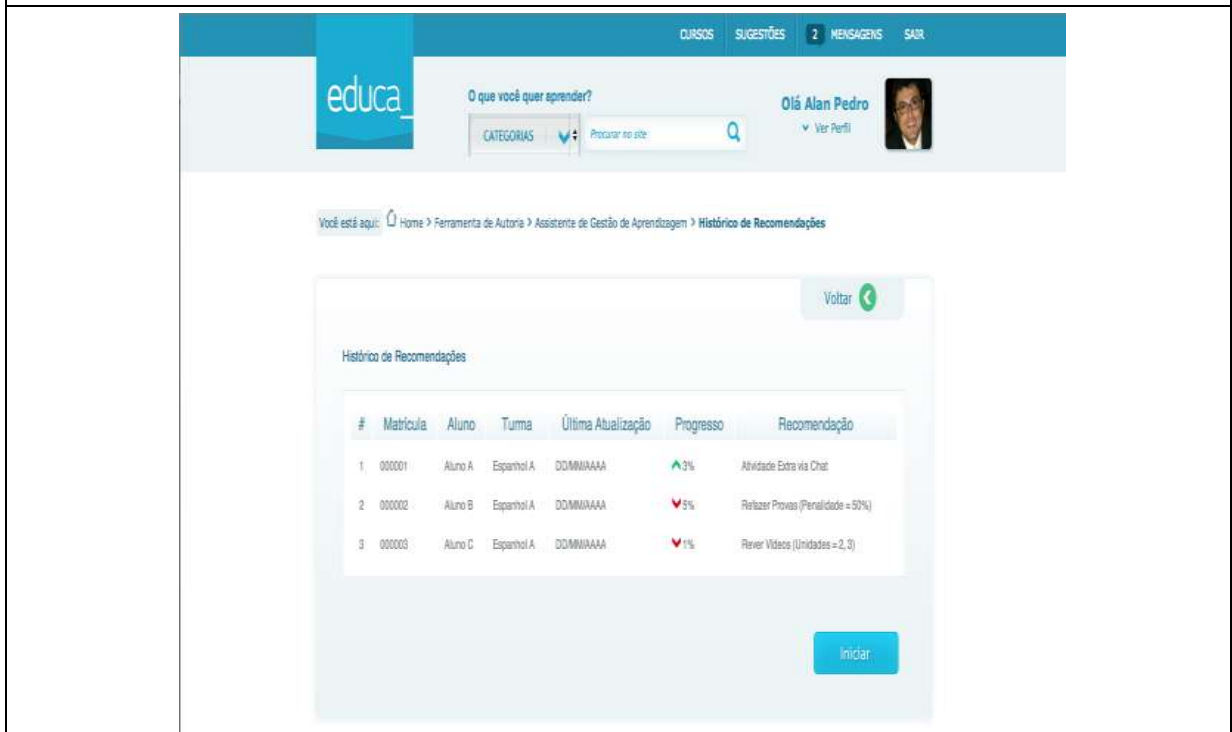
A Figura 31 mostra a escolha de uma recomendação apropriada para cada aluno.

Figura 31 - Selecionando uma das recomendações pedagógicas propostas.



A figura 32 apresenta uma simulação, da etapa de monitoramento e a avaliação do impacto das recomendações no desempenho dos alunos.

Figura 32 - Monitoramento e avaliação do impacto das recomendações no desempenho dos alunos.



5.6 Discussão

Nesse estudo de caso aplicamos o Processo de Recomendação Pedagógica (PRP), e a ferramenta, em um cenário real. O cenário em questão foi um curso de línguas, UFAL Línguas – Espanhol, oferecido na modalidade online, que ocorreu de 22 de Outubro de 2012 a 10 de Fevereiro de 2013, e teve 200 alunos inscritos.

Nesse estudo seguimos as etapas do PRP, utilizando os dados das interações dos alunos inscritos no UFAL Línguas e, na primeira etapa, conforme descrita no Capítulo 4 (Subseção 4.1.1), visamos detectar práticas (formas de interagir com vistas a obter um resultado pedagógico) positivas e negativas, dentre as interações dos alunos.

Assim como nas demais etapas, houve um esforço colaborativo entre atores humanos (*stakeholders* do sistema) e não humanos (o software utilizado para a análise estatística), o qual identificou no histograma de Pontuação de Itens de Interação⁵⁹ três agrupamentos de alunos. No primeiro agrupamento, percebemos que uma grande quantidade de alunos obteve resultado **inferior a 1500 pontos**, caracterizando **baixo desempenho** no curso. O agrupamento seguinte é mais diverso e abrange resultados **superiores a 1500 pontos até 3500 pontos**, os quais caracterizam desempenhos no curso que vão **de baixo a médio**. Por fim, uma pequena quantidade de alunos obteve resultados **superiores a 3500 pontos até, aproximadamente, 5000 pontos**, os quais foram classificados como tendo apresentado **alto desempenho** no curso.

Passamos, então, para a próxima etapa que consistiu em descobrir padrões que estivessem relacionados com essas práticas. Nessa etapa foi necessário o uso de Técnicas de Mineração de Dados (Tabela 5.2) e de formas de visualização dos resultados. A seleção do algoritmo baseou-se nas características, indicações e limitações (tabela 3) do conjunto de algoritmos considerados nessa dissertação (tabela 2), no caso em questão foi escolhido o J48, que nos gerou uma **Árvore de Decisão**. Entretanto, identificamos que a apresentação da árvore gerada (Figuras 27 e 28) não atingiu o objetivo pretendido, que era o de simplificar a análise dos resultados para os especialistas no domínio pedagógico.

⁵⁹ Os resultados foram calculados da seguinte forma: atividades e exercícios respondidos corretamente receberam 10 pontos multiplicado pelo peso referente ao nível de dificuldade (fácil = 1, médio = 3 e difícil = 5). A isso somou-se o resultado das provas (0 a 100 para cada prova) e a quantidade de interações via chat e fórum (multiplicadas por 20 e 10, respectivamente).

Para lidar com esse problema identificamos, nos resultados da mineração, cinco padrões de interesse pedagógicos e, para cada um deles, geramos, manualmente, algumas propostas de solução (recomendações pedagógicas), às quais foram encaminhadas para a avaliação dos professores em relação ao quão apropriadas elas eram para a situação pedagógica para a qual foram propostas. Isso foi feito através de um questionário respondido por, 17 professores cujas perguntas e o resultado encontram-se descritos no **Apêndice C**.

Os resultados do questionário foram utilizados para atribuir um grau de relevância para cada uma das recomendações criadas, eliminando, dessa forma o problema de "*cold start*" o qual geraria dificuldades, ou até impediria, que novas recomendações fossem propostas pelo sistema, por não haver evidência de uso (IAQUINTA, 2011), ou seja, não existirem indicações colaborativas (i.e.: feitas por outros usuários) de que a recomendação era relevante para a situação pedagógica para a qual ela foi proposta.

O problema de "*cold start*" também pode impactar negativamente a chance de adições posteriores de recomendações serem propostas. Antecipando essa situação, abordamos o problema adicionando uma recomendação extra para cada situação, feita de forma randômica.

Esse tratamento tornou possível que prosseguíssemos para o passo seguinte, onde as recomendações foram apresentadas aos professores por meio da Interface com o Usuário onde, para cada aluno, foram apresentadas algumas opções de recomendação pedagógica devendo o professor selecionar a que julgou mais adequada. Após a decisão do professor, as recomendações foram encaminhadas aos respectivos alunos, na forma de um link para a atividade, que ao ser clicado provê a capacidade de se monitorar o acesso à recomendação.

Nesse ponto reside uma limitação que foi o impediu de realizarmos o monitoramento e a avaliação do desempenho dos alunos após serem expostos à recomendação e, também, da própria avaliação onde utilizaremos técnicas de Aprendizado por Reforço. Retomaremos esse assunto na seção da conclusão que trata das limitações dessa dissertação.

Mesmo com essa limitação a ferramenta é capaz de responder as questões de pesquisa realizadas na Seção 1.2, que são: i) Como processar toda essa quantidade de dados educacionais de modo a extrair informações úteis? ii) O que fazer com o resultado obtido a partir do processamento dos dados educacionais? iii) Como tornar o processo transparente para os professores?

Concluimos respondendo que a Mineração de Dados Educacionais oferece as técnicas necessárias para esse processamento e **descoberta de informações relevantes**, dentro do

domínio educacional. Isso foi demonstrado na Seção 5.2 desse estudo de caso. Buscamos uma maneira de tornar útil o resultado desse processamento, em resposta à segunda questão de pesquisa, que foi o de **prover recomendações pedagógicas personalizadas**, como demonstrado na **Seção 5.3**. Por fim, toda a complexidade técnica dos processos envolvidos foi abstraída para os professores por intermédio da **Ferramenta para Recomendação Pedagógica**, proposta nessa dissertação, cuja interface é apresentada na **Seção 5.5**.

6 CONCLUSÃO

Nessa dissertação tivemos como objetivo a criação, e apresentação (através de um estudo de caso), de uma ferramenta capaz de auxiliar professores de cursos baseados na web, a extrair informações relevantes das interações desses alunos com o ambiente de aprendizagem, através da Mineração de Dados Educacionais. Esses dados foram utilizados para realizar recomendações pedagógicas que promovessem a melhoria/aprimoramento da experiência de aprendizado dos alunos.

Para atingir esse objetivo utilizamos o **Processo de Recomendação Pedagógica (PRP)**, o qual orienta a descoberta de informações úteis, através da análise dos dados interacionais dos alunos, contando com a sinergia da **Inteligência Humana** e da **Inteligência Artificial**, tendo como finalidade prover apoio à tomada de **decisões pedagógicas**, para promover o **aprimoramento da experiência de aprendizado** para alunos e turmas (PAIVA, 2012).

O desenvolvimento desse processo resultou na escrita de um artigo apresentado no **I Workshop de Desafios na Computação Aplicada à Educação (DesafIE!)**, parte do **XXXII Congresso da Sociedade Brasileira de Computação**, o qual ocorreu na cidade de Curitiba, em Julho de 2012. Tal apresentação nos possibilitou conhecer as algumas demandas da comunidade científica para essa área, o que culminou com a realização de ajustes no processo como, por exemplo, a participação dos especialistas no domínio pedagógico em todas as suas etapas.

O passo seguinte foi a utilização do PRP para, efetivamente, orientar a criação da ferramenta proposta.

Realizamos, por fim, um estudo de caso com dados de um ambiente de aprendizagem real, o UFAL Línguas - Espanhol, para avaliar a aplicabilidade da ferramenta. Através de tal estudo constatamos que o PRP é aplicável e traz benefícios ao processo de ensino e aprendizagem, uma vez que auxilia professores não apenas detectando práticas, mas, também, descobrindo os padrões que justificam tais práticas e oferecendo recomendações pedagógicas para as situações identificadas.

6.1 Os Objetivos Específicos da Dissertação

- Realizar o levantamento do estado da arte sobre Mineração de Dados Educacionais e Ferramentas para Recomendação Pedagógica;

- Mostramos na Seção 2.4 (tabelas 2 e 3) os 10 algoritmos mais utilizados pelos pesquisadores da área (WU, 2008), considerados como “estado da arte” na área de mineração de dados e aprendizado de máquina (WITTEN, 2011). Apresentamos, também, detalhes sobre a ferramenta e API contendo a implementação desses algoritmos, denominada WEKA.
- Implementar componentes que reutilizam algoritmos de mineração de dados educacionais.
 - Realizamos, no Capítulo 4, a descrição da ferramenta proposta e demonstramos no Apêndice D os principais detalhes de sua implementação, mostrando os componentes que constituem a ferramenta com alguns deles fazendo uso dos algoritmos disponibilizados pela API WEKA.
- Especificar cenários de mineração de dados educacionais, os quais serão encapsulados em componentes distintos.
 - A especificação desses cenários foi feita na Seção 4.2 e o uso de um desses cenários (Qualidade da Interação) foi demonstrado no estudo de caso, Seção 5.2. Detalhes sobre a implementação pode ser encontrada no Apêndice D.
- Realizar recomendações personalizadas, com base na situação pedagógica dos alunos.
 - Após a descoberta dos padrões, para as práticas pedagógicas detectadas, realizamos a criação de Recomendações Pedagógicas que foram avaliadas por professores (através de um questionário), com consequente medida de suas relevâncias para as práticas relatadas e que foram, posteriormente, transformadas em regras do tipo IF__THEN__. Esse processo está descrito na Seção 5.3.
- Construir uma Ferramenta para a Recomendação Pedagógica que possibilite que professores e tutores utilizem a Mineração de Dados Educacionais, bem como a Recomendação Pedagógica, sem que necessitem de conhecimento sobre essas técnicas.
 - Na Seção 5.5 apresentamos a Interface com o Usuário da ferramenta criada. É através dessa interface que os professores têm acesso ao resultado da mineração dos dados educacionais de seus cursos, para cada um dos cenários implementados. É, também, por meio dessa interface que eles podem identificar os alunos em determinada situação pedagógica e selecionar a recomendação mais apropriada para cada caso.
- Avaliar a ferramenta através de um estudo de caso real.

- Feita através da utilização dos dados educacionais do curso de Espanhol, UFAL Línguas. Processo descrito no Capítulo 5.

6.2 Limitações

O desenvolvimento da ferramenta foi limitado por algumas situações que ocorreram durante sua implementação.

Uma delas decorreu do uso de dados finais do curso. Isso nos impossibilitou de monitorar e avaliar os efeitos das recomendações no desempenho dos alunos, bem como avaliar a qualidade das próprias recomendações (através de técnicas de Aprendizado por Reforço).

Outra é a necessidade de interação entre os atores humanos e não humanos que, apesar de bastante desejável e útil, pode tornar lento o avanço da ferramenta.

Outra questão percebida é a de que a mineração dos registros de uso do sistema (logs) assim como os resultados de minerações, embora não implementados nessa versão, podem trazer diversos benefícios e serão planejados para uma versão futura.

6.3 Perspectivas Futuras

De uma forma geral, pretendemos evoluir a ferramenta de forma a promover melhorias em suas características de Recomendação e adicionar características de autoria à mesma.

Conforme mencionado nas limitações, queremos avaliar e medir (quantitativamente) a influência das recomendações no desempenho dos alunos e, além disso, fazer uso dos próprios dados gerados pela ferramenta como fonte para a mineração (retroalimentação).

Há planos de possibilitar que, sem o auxílio de especialistas no domínio tecnológico, os professores sejam capazes de criar seus próprios cenários de mineração, algo que envolverá atualização e controle dinâmicos dos cenários.

E por fim, consideramos bastante válida a completa implementação da ferramenta com base em ontologias, o que aprimoraria a capacidade da ferramenta de descrever conhecimento.

REFERÊNCIAS

- ABEL, F. et al. Recommendations in online discussion forums for e-learning systems. *IEEE Transactions on Learning Technologies*, v. 3, p. 165-176, 2010.
- AROYO, L. et al. EASE: Evolutional authoring support environment, 2004.
- AUDINO, D.F.; NASCIMENTO, R.d.S. *Revista Contemporânea de Educação*, v. 5, n. 10, julho/dezembro 2010.
- BAKER, R.S.J.d. Data mining for education. McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education*. Elsevier, 2010.
- BAKER, R.S.J.d. Data mining for student models. *Advances in Intelligent Tutoring Systems*. Springer, 2010. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-642-14363-2_16>. Acesso em: 03 Jul. 2013.
- BAKER, R.S.J.d.; ISOTANI, S.; CARVALHO, A.M.J.B.d. Mineração de dados educacionais: oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v. 19, n. 2, p. 3-13, 2011.
- BAYER, J. et al. Predicting drop-out from social behaviour of students. *Educational Data Mining Conference*, 2012.
- BITTENCOURT, I.I. et al. A computational model for developing semantic web-based educational systems, *Knowledge-Based Systems*, v. 22, p. 302-315, 2009.
- BITTENCOURT, I.I.; COSTA, E.d.B. Modelos e ferramentas para a construção de sistemas educacionais adaptativos e semânticos. *Revista Brasileira de Informação na Educação*, v. 19, p. 85-98, 2011.
- BOSLAUGH, S. *Statistics in a nutshell*. 2nd ed. O'Reilly, 2012.
- BOUCKAERT, R.R. et al. *WEKA manual for version 3.6.8*. University of Waikato, 2012.
- BRITO, P.H.S.; et al. A systematic approach for designing educational recommender systems. *Educational Recommender Systems and Technologies*, IGI Global, 2012.
- CABADA, R.Z.; ESTRADA, M.L.B.; GARCÍA, C.A.R. *EDUCA: A web 2.0 authoring tool for developing adaptative and intelligent tutoring systems using Kohonen network*. *Expert Systems with Applications*. Elsevier, 2011.
- CNN MONEY, Disponível em: <http://money.cnn.com/2009/06/17/news/economy/goldman_sachs_jim_oneill_interview.fortune/index.htm>, Acesso em: 05 Abr. 2013.
- DAVIDOFF, Linda L. *Introdução à psicologia*. McGraw-Hill do Brasil, p. 112-115, 1983.
- DEOGUN, J.S. et al. Data mining: trends in research and development. "Data mining: trends and issues-guest editors' introduction". *Journal of Amer. Soc. for Information Sci.*, v. 49(4), p. 397-402, Abril 1998.

DESAFIE!. I Workshop de desafios da computação aplicada à educação. Disponível em: <<http://www.imago.ufpr.br/csbc2012/desafie.php>>. Acesso em: 14 Mai. 2013.

DOMINGUES, A.K.; YACEF, K.; CURRAN, J.R. Data mining for individualized Hints in eLearning. Proceedings of the 3rd International Conference on Educational Data Mining. Junho, 2010.

DUKE. Bioelectricity: a quantitative approach. Duke University's First MOOC. Disponível em: <<http://dukespace.lib.duke.edu/dspace/handle/10161/6216>>. Acesso em: 27 Mar. 2013. Fevereiro, 2013.

FERREIRA, R. et al. A framework for building web mining applications in the world of blogs: a case study in product sentiment analysis. Expert Systems with Applications, v. 39, p. 4813-4834, 2012.

GAYARD, L.A.; RUBIRA, C.M.F.; GUERRA, P.A.d.C. COSMOS: a component system model for software architectures. Universidade Estadual de Campinas. Instituto de Computação, 2008.

GIBERT, K. et al. On the role of pre and post-processing in environmental data mining. International Congress on Environmental Modeling and Software, 2008.

HAN, J.; KAMBER, M. Data mining: concepts and techniques, ed. 03 - Elsevier/Morgan Kaufmann, 2011.

HUANG, S.; FANG, N. A work in progress: early prediction of students' academic performance in an introductory engineering course through different mathematical modeling techniques. Frontiers in Education, 2012.

IAQUINTA, L.; SEMERARO, G. Lightweight approach to the cold start problem in the video lecture recommendation. 2011.

JANNACH, D. et al. Recommender systems: an introduction. Cambridge University Press, 2011.

KANELLOPOULOS, D.; KOTSIANTIS, S. Towards an ontology-based system for intelligent prediction of student dropouts in distance education. International Journal of Management in Education. v. 2, n. 2, 2008.

KKD NUGGETS, Disponível em: <<http://www.kdnuggets.com/polls/2012/where-applied-analytics-data-mining.html>>. Acesso em: 22 Fev. 2013.

LAROSE, D.T. Data mining methods and models. Wiley, 2006.

LEFFA, V.J. Uma ferramenta de autoria para o professor: o que é e o que faz. Letras de Hoje, Porto Alegre, v. 41, n. 144, p. 189-214, 2006.

MANHAES, L.M.B. et al. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. Anais do XXII SBIE - XVII WIE, 2011.

MÁRQUEZ, V.; ROMERO C.; VENTURA S. Predicting school failure using data mining. In: proceedings of the international conference on educational data mining, 4th, Madrid, 2011.

MERCERON, A. Educational data mining: a case study. Proceeding of the conference on artificial intelligence in education. 12th, Amsterdam, 2005. Disponível em: <<http://dl.acm.org/citation.cfm?id=1562589>>

MINIX.ORG. The MINIX 3 Operating System. Disponível em: <<http://www.webcitation.org/64sAstWXM>>. Acesso em 20 Jan. 2012.

MORAN, J.M. O que aprendi sobre avaliação em cursos semi-presenciais. Avaliação da aprendizagem em educação online. Loyola, 2006. Disponível em: <<http://www.eca.usp.br/prof/moran/aprendi.html>>. Acesso em: 03 Jul. 2013.

NISBET, D. Measuring the quantity and quality of online discussion group interaction. Journal of eLiteracy, v. 1, p. 122-139, 2004.

NISBET, R.; ELDER, J.; MINER, G. Statistical analysis and data mining applications. Elsevier, 2009.

NORTH, M. Data mining for the masses. 2012

OLIVEIRA, D.A. A reestruturação do trabalho docente: precarização e flexibilização. Educ. Soc., Campinas, v. 25, n. 89, p. 1127-1144, Setembro/Dezembro 2004. Disponível em: <<http://www.cedes.unicamp.br>> Acesso em: 20 Dez. 2012.

PARK, J.; CHOI, H.J. Factors influencing adult learners' decision to drop out or persist in online learning. Educational technology & society, v. 12, p. 207-217, 2009.

PARK, J. Factors related to learner dropout in online learning, 2007. Disponível em: <<http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED504556>>. Acesso em: 03 Jul. 2013.

PNE. Plano nacional de educação – PNE. Disponível em: <http://portal.mec.gov.br/index.php?option=com_content&view=article&id=16478&Itemid=1107>. Acesso em: 24 Abr. 2012.

PNE(b). Projeto de lei 8035/2010. Câmara dos deputados. Disponível em: <<http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=490116>>. Acesso em: 12 Abr. 2012.

REIHANEH, R.K.; MANSOUREH, T.; ZAIANE, O.R. Analyzing participation of students in online courses using social network analysis techniques. In: proceedings of the international conference on educational data mining, 4th, Madrid, 2011.

REYNOLDS, K.; KOTSIANTIS, A.; EDWARDS, L. Using machine learning to detect cyberbullying. In: international conference on machine learning and applications and workshops, 10th, p. 241-244, 2011.

RICCI, F.; et al. Recommender systems handbook. Springer, 2011.

ROMERO, C.; VENTURA, S. Educational data mining: a survey from 1995 to 2005. Expert Systems with Applications, v. 33(1), p. 135-146, 2007.

ROMERO, C.; et al. Handbook of educational data mining. CRC Press, 2011.

RUSSELL, S.; NORVIG, P. Artificial intelligence: a modern approach. 3. ed. New Jersey: Pearson Education, 2010.

SALMON, G. E-Moderation - The key to teaching and learning online. Kogan Page, 2000.

SILVA, A.P.d.; COSTA, E.d.B.; BITTENCOURT, I.I. Uma linha de produto de software baseada na web semântica para sistemas tutores inteligentes. Revista brasileira de informação na educação, v. 20, p. 87, 2012.

SOMMERVILLE, I. Software Engineering. ed. 09. Addison-Wesley, 2011.

UFAL Línguas(a). Disponível em: <<http://www.ufal.edu.br/cied/informes/cied-torna-publico-o-edital-do-curso-201cel-mundo-hispanico-sin-fronteiras>>. Acesso em: 15 Jan. 2013.

UFAL Línguas(b). Disponível em:
<<http://www.ufal.edu.br/cied/documentos/editalsimplificadocursoextensaoespanhol.pdf>>.
Acesso em: 15 Jan. 2013.

VESIN, B. et al. Protus 2.0: Ontology-based semantic recommendation in programming tutoring system. Expert Systems with Applications. Elsevier, 2012.

W3C.ORG. Authoring tool definition. Disponível em: <<http://www.w3.org/TR/2012/WD-ATAG20-20120410/#def-Authoring-Tool>>. Acesso em: 03 Jan. 2013.

WALKER, M. PISA 2009 plus results: performance of 15-year-olds in reading, mathematics and science for 10 additional participants. Acer Press, 2011.

WITTEN, I.; FRANK, E.; HALL, M. Data mining: practical machine learning tools and techniques. ed. 3, Elsevier, 2011.

WU, X.; et. al. Top 10 Algorithms in data mining. Knowledge information systems, Springer, 2007.

ZAIANE, O.R. Principles of knowledge discovery in databases - Chapter 1: introduction to data mining. Disponível em:
<<http://webdocs.cs.ualberta.ca/~zaiane/courses/cmp690/slides/Chapter1/sld001.htm>>.
Acesso em: 29 Dez. 2012.

APÊNDICE A

A Ferramenta de Mineração de Dados

Trataremos aqui sobre a ferramenta escolhida para realizar as tarefas de mineração de dados e, também, quais implementações foram utilizadas para os algoritmos escolhidos.

O Workbench Weka®

Weka®⁶⁰ acrônimo que significa *Waikato Environment for Knowledge Analysis* (Ambiente Waikato para Análise do Conhecimento), foi desenvolvido pela **Universidade de Waikato**, na **Nova Zelândia**, com o objetivo de prover a rápida experimentação de métodos de análise em novos conjuntos de dados, de forma rápida e flexível.

O sistema foi escrito em Java® e é distribuído de acordo com os termos da licença GPL⁶¹. Weka® funciona nos principais sistemas operacionais: Linux, Windows e Macintosh, e é classificado como uma "bancada de trabalho"⁶² contendo uma coleção de ferramentas de pré-processamento de dados, e das mais atuais implementações de algoritmos de aprendizado de máquina. O ambiente oferece suporte completo ao processo de experimentação em mineração de dados, ou seja: suporte ao preparo dos dados de entrada (pré-processamento dos dados), avaliação estatística de esquemas de aprendizado, bem como métodos para o processamento dos resultados (pós-processamento) e integração com aplicações externas através de uma API Java® (WITTEN, 2011).

O suporte aprimorado ao processo de compreensão dos dados (pré-processamento dos dados), a integração de recursos estatísticos, a API Java® e a presença de um ambiente capaz de comparar o desempenho de diversos algoritmos de mineração de dados (MANHÃES, 2011), o ambiente *Experimenter*, são os principais motivos para a escolha do workbench Weka® em detrimento a outras ferramentas.

⁶⁰Disponível para download em: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>

⁶¹GPL, sigla em Inglês para: General Public License (Licença Pública Geral), que é uma licença de uso de software que garante ao usuário final o direito de estudar, compartilhar e modificar um software sob essa licença (GNU.ORG).

⁶²Em inglês: workbench.

APÊNDICE B

Análise Estatística dos Dados (Script R)

Aqui apresentamos o código anotado do script desenvolvido em **R** para análise estatística dos dados educacionais, utilizado no estudo de caso (descrito no **Capítulo 5**).

Na **linha 4**, realizamos a leitura, e armazenamento na variável `dados`, do arquivo contendo os dados coletados do ambiente UFAL Línguas.

Figura 33 - Código R anotado. Análise estatística dos dados interacionais para o cenário 1.

```

1 # Ferramenta de Autoria para a Recomendação Pedagógica Baseada em Mineração de Dados Educacionais
2 # Análise Estatística dos Dados Educacionais - UFAL Línguas Espanhol
3
4 dados = read.csv("UFALLinguas2.csv", header=TRUE, sep=";")
5
6 summary(dados)
7 summary(dados$Resultado)
8
9 png(filename="UFALLinguas_Histograma_Pontuacao_Testes.png", width=900, bg="white")
10
11 hist(
12   dados$Resultado,
13   col = c(2:10),
14   ylim = c(0,100),
15   xlim = c(0,5000),
16   main = "UFAL Línguas - Espanhol | Pontuação nos Itens de Interação",
17   xlab = "Somatório dos Pontos",
18   ylab = "Número de Alunos"
19 )
20
21 dev.off()
22
23 png(filename="UFALLinguas_Boxplot_Pontuacao_Itens_Interação_Acessos.png", width=900, bg="white")
24
25 par(mfrow=c(1,2))
26
27 boxplot(
28   dados$acessos,
29   main = "UFAL Línguas - Espanhol | Acessos"
30 )
31
32 boxplot(
33   dados$Resultado,
34   ylim = c(1, 7300),
35   main = "UFAL Línguas - Espanhol | Pontuação nos Itens de Interação"
36 )
37
38 dev.off()
39
40 png(filename="UFALLinguas_Barplot_Pontuacao_Itens_Interação.png", width=900, bg="white")
41
42 barplot(
43   dados$Resultado,
44   col = c(2:4),
45   axis.lty = 1,
46   main = "UFAL Línguas - Espanhol | Pontuação nos Itens de Interação",

```

Nas **linhas 5 e 6** imprimimos um sumário de todos os dados e da coluna `Resultado` (a pontuação nos itens de interação com o ambiente de aprendizagem), contendo informações sobre valores máximo e mínimo, média, mediana, 1º e 3º quartis.

Nas **linhas 9 a 21** solicitamos a criação de um histograma com os dados referentes à coluna `Resultados`, para análise da distribuição dos dados. O resultado foi salvo em uma imagem.

Figura 34 - Código R anotado. Análise estatística dos dados interacionais para o cenário 1.

Continuação.

```

47   xlab = "Somatório dos Pontos",
48   ylab = "Número de Alunos",
49   width = 5,
50   cex.names = 1
51 )
52
53 # Fechar arquivo
54 dev.off()
55
56 qld_baixa = subset(dados, Resultado < 1500, select = c(Resultado))
57 qld_media = subset(dados, Resultado < 3500 & Resultado >= 1500, select = c(Resultado))
58 qld_alta = subset(dados, Resultado >= 3500, select = c(Resultado))
59
60 qualidade = c(nrow(qld_baixa), nrow(qld_media), nrow(qld_alta))
61
62 hist(qld_baixa$Resultado)
63 hist(qld_media$Resultado)
64 hist(qld_alta$Resultado)
65
66 # Criar e abrir o arquivo imagem
67 png(filename="UFALLinguas_Boxplot.png", width=900, bg="white")
68
69 # Boxplot
70 boxplot(
71   qualidade,
72   main = "UFAL Línguas - Espanhol | Qualidade das Interações"
73 )
74
75 dev.off()
76
77 png(filename="UFALLinguas_Barplot.png", width=900, bg="white")
78
79 barplot(
80   qualidade,
81   names.arg = c("BAIXA", "MÉDIA", "ALTA"),
82   col = c(2:4),
83   axis.lty = 1,
84   main = "Qualidade das Interações",
85   xlab = "Qualidade das Interações",
86   ylab = "Número de Alunos",
87   width = 5,
88   cex.names = 1
89 )
90
91 dev.off()

```

Nas **linhas 23 a 38** criamos, lado a lado, um histograma e um *boxplot* para a coluna Acessos, para realizarmos uma avaliação da distribuição dos dados e busca por *outliers*. Essa análise foi repetida para as demais colunas (não representadas no script).

Nas **linhas 40 a 54** criamos um gráfico em barras contendo cada dado da coluna Resultado, com o resultado impresso em uma imagem.

Nas **linhas 56 a 64** segmentamos os dados de acordo com o nível de qualidade (baixo, médio e alto) e em seguida criamos um histograma para cada nível, com a finalidade de exibir a distribuição desses dados.

Nas demais **linhas, 67 a 91**, criamos um *boxplot* e um gráfico em barras para analisarmos a presença de *outliers* e os valores absolutos, por nível, respectivamente.

APÊNDICE C

Questionário de Recomendações

Apresentamos aqui o questionário que foi utilizado para obter a opinião dos professores sobre a relevância das primeiras recomendações, criadas para a Ferramenta para a Recomendação Pedagógica, durante o estudo de caso e, em seguida, exibimos a Tabela contendo os resultados.

O Questionário



Núcleo de Excelência em Tecnologias Sociais

Nome	_____		Nascimento	___/___/___
Sexo	<input type="checkbox"/> F <input type="checkbox"/> M	Nível Educacional	_____	
Anos de Experiência no Ensino	_____	Experiência com EAD ⁶³	<input type="checkbox"/> Sim <input type="checkbox"/> Não	
Disciplina	_____	Experiência como Tutor	<input type="checkbox"/> Sim <input type="checkbox"/> Não	

Caríssimos professores, gostaríamos de gentilmente solicitar sua ajuda na avaliação de algumas **recomendações pedagógicas** baseadas em situações detectadas em um curso online. Afirmamos que os dados pessoais fornecidos serão mantidos em sigilo, com sua utilização para fins estatísticos, apenas.

O curso em questão disponibiliza diversos exercícios e atividades para desenvolver o conhecimento do aluno, que deve realizar algumas tarefas que são avaliadas e pontuadas pela professora responsável pelo curso. Essas tarefas incluem **exercícios, atividades de revisão, interações via chat, postagens via fórum e as provas**.

Os dados das interações dos alunos, para esse curso, foram analisados resultando em 5 situações pedagógicas, para as quais foram criadas algumas recomendações (disponibilizamos um espaço livre para que possam adicionar recomendações diferentes das apresentadas). **Gostaríamos de contar com sua experiência para atribuir uma nota, de 1 a 5, para cada um dos quesitos de cada uma das 5 situações.**

O significado das notas é o seguinte: **1 = recomendação totalmente inapropriada, 2 = recomendação pouco apropriada, 3 = recomendação razoável, 4 = recomendação apropriada e 5 = recomendação muito apropriada.**

Situação 01

⁶³ Ensino a Distância.

De um total de 46 provas, aqueles alunos que responderam 19 ou menos, obtiveram baixo desempenho no curso.

A	Solicitar resolução das provas não respondidas, sem penalidades.	
B	Solicitar resolução das provas não respondidas, com penalidade de 50%.	
C	Refazer as provas já realizadas, sem penalidade.	
D	Refazer as provas já realizadas, com penalidade de 50%.	
E	Criar novas provas para alunos nessa situação.	
F		

Situação 02

De um total de 13 atividades, aqueles alunos que realizaram 2 ou menos, obtiveram baixo desempenho no curso.

A	Solicitar resolução das atividades não realizadas, sem penalidades.	
B	Solicitar resolução das atividades não realizadas, com penalidade de 50%.	
C	Solicitar resolução das atividades de um determinado nível (fácil, por exemplo).	
D	Refazer as atividades já realizadas.	
E	Assistir, novamente, todos os vídeos da unidade.	
F	Acessar um conteúdo de outro site, tratando sobre o assunto.	
G		

Situação 03

Alunos que obtiveram 50% ou menos, de aproveitamento nos exercícios, tiveram baixo desempenho no curso.

A	Refazer as atividades já realizadas.	
B	Assistir, novamente, todos os vídeos da unidade com piores desempenhos.	
C	Acessar um conteúdo de outro site, tratando sobre os assuntos abordados.	
D	Propor atividade extra.	
E	Refazer as atividades do nível fácil.	
F	Refazer as atividades do nível médio.	
G		

Situação 04

Alunos que não realizaram interações via chat, obtiveram desempenho baixo no curso.

A	Enviar solicitação de acesso ao chat.	
B	Solicitar resolução de atividade via chat.	
C	Interagir via chat com o aluno.	
D	Anunciar, antecipadamente, atividades via chat.	
E	Criar atividade extra para resolução via chat.	
F		

Situação 05	
A frequência de alguns alunos às aulas está diminuindo com o progredir do curso	
A	Solicitar que o aluno simplesmente acesse o sistema.
B	Alertar sobre as atividades/exercícios cujo prazo de conclusão está próximo.
C	Exibir o quanto (tempo, atividades e/ou provas) falta para a conclusão do curso.
D	Exibir o quanto o aluno já progrediu (tempo, atividades e/ou provas realizadas).
E	Bonificar o aluno com maior o prazo para realizar as atividades/exercícios.
F	

"Permito que **minhas avaliações sejam utilizadas para pontuar as recomendações aqui apresentadas**, e que os **meus dados pessoais sejam utilizados, exclusivamente, com propósito estatístico**, sem que minha identidade seja revelada, nem que meus dados pessoais sejam usados, ou compartilhados, para outra finalidade senão a que foi aqui apresentada".

Maceió, ____ de ____ de 20 ____

Assinatura do Participante

Os Resultados

Tabela 11 - Resultados do questionário respondido pelos professores a respeito das recomendações.

Regra de Associação	Recomendação	Relevância																
IF provasRespondidas <= 19	ProvasNaoRespondidasPenalidade[0]	5	1	4	1	1	5	1	1	1	2	3	3	2	4	1	1	2
	ProvasNaoRespondidasPenalidade[50]	1	3	1	4	2	1	1	5	2	4	2	2	2	1	3	4	2
	RefazerProvasPenalidade[0]	1	1	2	1	1	2	1	1	1	1	2	1	1	4	1	1	2
	RefazerProvasPenalidade[50]	1	4	1	4	2	2	1	1	1	1	2	2	5	1	4	1	2
	NovasProvas	3	5	2	2	4	3	5	1	1	3	3	5	4	5	4	5	4
IF provasRespondidas <= 19 AND exerciciosCorretos <= 252 AND atividadesRealizadas <= 2	AtividadesNaoRespondidasPenalidade[0]	5	3	1	1	1	5	1	1	3	2	3	3	2	4	1	1	2
	AtividadesNaoRespondidasPenalidade[50]	1	4	2	2	2	1	1	5	4	5	2	2	3	1	4	4	2
	AtividadeNivel[X]	1	1	1	2	1	3	1	4	1	2	4	1	1	4	2	3	1
	RefazerAtividadesPenalidade[0]	1	2	3	2	3	1	3	1	1	1	2	1	3	4	1	1	2
	ReverVideosUnidades[X]	3	5	5	3	4	5	5	3	3	2	4	3	4	4	2	1	4
	ConteudoExterno[X]	3	4	5	1	4	4	2	1	1	3	4	2	4	3	2	2	2
IF provasRespondidas <= 19 AND exerciciosCorretos <= 252	RefazerExerciciosPenalidade[0]	1	5	4	1	3	1	5	1	1	1	4	?	4	4	2	1	2
	ReverVideosUnidades[X]	3	5	3	3	4	5	2	4	3	4	3	?	4	1	3	1	4
	ConteudoExterno[X]	4	4	4	1	3	4	2	2	1	5	3	?	3	1	3	4	3
	ExerciciosExtra	5	4	4	4	4	5	4	4	3	4	4	?	5	4	4	5	4
	RefazerExerciciosNivel[facil]	2	2	2	2	2	1	1	5	2	2	3	?	2	3	2	2	1
	RefazerExerciciosNivel[medio]	2	4	2	2	2	1	1	5	2	2	3	?	3	5	2	2	1
IF interacoesForum <= 3 AND interacoesChat == 0	SolicitarAcessoChat	4	5	1	3	4	5	4	5	3	3	2	?	5	3	4	4	3
	AtividadeChat	4	5	1	3	4	4	2	5	4	4	3	?	5	3	1	4	3
	InteracaoChat	5	5	5	4	3	5	4	3	3	4	3	?	5	4	2	3	4
	AnunciarAtividadesChat	3	5	5	3	5	5	3	5	5	4	4	?	5	5	4	3	5
	AtividadeExtraChat	4	5	4	4	5	5	1	4	4	3	4	?	5	4	4	3	4

APÊNDICE D

Implementação da Ferramenta

Nessa parte dos apêndices exibiremos a implementação de alguns componentes, apresentando seu código-fonte acompanhado de comentários.

Figura 35 - Implementação de um cenário de mineração, que constitui da definição de suas configurações que serão utilizadas pelo componente de mineração (parte 1).

```

1 public class InteractionsQualityScenario {
2     private ManagerInteractionsQualityScenario manager;
3
4     public InteractionsQualityScenario(ManagerInteractionsQualityScenario manager) {
5         this.manager = manager;
6     }
7
8     //Configurações do pré-processamento
9     public Map<String, String> preprocessingConfig () {
10        Map<String, String> preprocessingConfigParameters = new HashMap<>();
11
12        preprocessingConfigParameters.put("api", "weka");
13        preprocessingConfigParameters.put("createFile", "arff");
14        preprocessingConfigParameters.put("normalize", null);
15        preprocessingConfigParameters.put("transformData", "NonNumericToZero");
16        preprocessingConfigParameters.put("treatEmpty", "fillWithZero");
17        preprocessingConfigParameters.put("treatNonNumeric", null);
18        preprocessingConfigParameters.put("treatOutliers", "3");
19
20        return preprocessingConfigParameters;
21
22    //Configurações do processo de mineração
23    public Map<String, String> miningConfig() {
24        Map<String, String> miningConfigParameters = new HashMap<>();
25
26        miningConfigParameters.put("algorithm", "j48");
27        miningConfigParameters.put("api", "weka");
28        miningConfigParameters.put("scenarioName", "InteractionsQuality");
29
30        return miningConfigParameters;
31    }
32
33    //Configurações do pós-processamento
34    public Map<String, String> postprocessingConfig() {
35        Map<String, String> postprocessingConfigParameters = new HashMap<>();
36
37        postprocessingConfigParameters.put("api", "weka");
38        postprocessingConfigParameters.put("denormalize", null);
39        postprocessingConfigParameters.put("transformData", null);
40        postprocessingConfigParameters.put("visualization", null);
41
42        return postprocessingConfigParameters;
43    }
44

```

Figura 36 - Implementação de um cenário de mineração. (parte 2).

```
45 //Configurações da recomendação
46 public Map<String, String> recommendingConfig() {
47     Map<String, String> recommendingConfigParameters = new HashMap<>();
48
49     recommendingConfigParameters.put("algorithm", "ibk");
50     recommendingConfigParameters.put("api", "weka");
51     recommendingConfigParameters.put("minimumRelevance", "weka");
52     recommendingConfigParameters.put("numberOfRecommendations", 5);
53     recommendingConfigParameters.put("numberOfRandomRecommendations", 1);
54
55     return recommendingConfigParameters;
56 }
57
58 }
```


Figura 37 - Método de construção do arquivo ARFF, parte do componente de mineração.

```

1 //Criação dos arquivos ARFF utilizando os dados oriundos do ambiente educacional
2 public void createArffFile(ArrayList educationalData, String scenarioName) throws ParseException, IOException, Exception {
3     FastVector attributes;
4     Instances dataSet;
5     double[] values = null;
6     attributes = new FastVector();
7
8     //Obter o número de linhas e colunas (primeiro elemento do ArrayList educationalData)
9     int numberColumns = Integer.parseInt(educationalData.get(0).toString());
10    int numberRows = educationalData.size() / numberColumns;
11
12    //Nomear os atributos do ARFF de acordo com o título das colunas do banco de dados
13    for(int i = 1; i <= numberColumns; i++) {
14        String attributeName = educationalData.get(i).toString();
15
16        attributes.addElement(new Attribute(attributeName));
17
18    }//End of for
19
20    //Criar e definir o atributo com propriedade de CLASS do ARFF
21    attVals = new FastVector();
22
23    for (int i = 0; i < 3; i++) {
24        attVals.addElement("val" + (i+1));
25
26    }//End of for
27
28    attributes.addElement(new Attribute("class", attVals));
29
30    //Adicionar instâncias ao arquivo ARFF (oriundos do banco de dados)
31    dataSet = new Instances(scenarioName.toString(), attributes, 0);
32
33    for(int i = 1; i < numberRows ; i++) {
34        values = new double[dataSet.numAttributes()];
35
36        int k = 0;
37
38        for(int j = 1; j <= numberColumns; j++) {
39            int index = (i * numberColumns) + j;
40
41            if(index < educationalData.size()) {
42
43                int value = Integer.parseInt(educationalData.get(index).toString());
44
45                values[k] = value;
46
47                k++;
48
49            }
50
51        }//End of for
52
53        dataSet.add(new Instance(1.0, values));
54
55    }//End of for
56
57    //Salvar o arquivo ARFF
58    ArffSaver arffSaverInstance = new ArffSaver();
59    arffSaverInstance.setInstances(dataSet);
60    arffSaverInstance.setFile(new File(scenarioName.toString() + ".arff"));
61    arffSaverInstance.writeBatch();
62 }
63

```


Figura 38 - Método responsável pela transformação de dados. Parte do pré-processamento dos dados.

```
64 //Transformar dados para outros tipos
65 public ArrayList transformData(String config, ArrayList educationalData) {
66     ArrayList result = new ArrayList();
67
68     //Transformar referências a sexo: de String para Integer (F = 0, M = 1)
69     for(int i = 0; i < educationalData.size(); i++) {
70         if(config.equals("Gender")) {
71             if(
72                 educationalData.get(i).toString().toLowerCase().equals("f") ||
73                 educationalData.get(i).toString().toLowerCase().equals("fem") ||
74                 educationalData.get(i).toString().toLowerCase().equals("feminino")
75             ) {
76
77                 educationalData.set(i, (Object) 0);
78
79             } else if(
80                 educationalData.get(i).toString().toLowerCase().equals("m") ||
81                 educationalData.get(i).toString().toLowerCase().equals("masc") ||
82                 educationalData.get(i).toString().toLowerCase().equals("masculino")
83             ) {
84                 educationalData.set(i, (Object) 1);
85
86             } //End of if
87
88         } //End of if
89
90     } //End of transformData
91
92     //Retornar novo ArrayList com as alterações realizadas
93     result = educationalData;
94
95     return result;
96
97 } //End of transformData
98
```

Figura 39 - Métodos de tratamento de instâncias ausentes e valores não numéricos.

```

99 //Tratar as instâncias vazias
100 public ArrayList treatEmpty(String config, ArrayList educationalData) {
101     ArrayList result = new ArrayList();
102
103     //Substituir valores vazias por 0
104     for(int i = 0; i < educationalData.size(); i++) {
105         if(config.equals("fillWithZero")) {
106             if(educationalData.get(i).toString().equals("")) {
107                 educationalData.set(i, (Object) 0);
108
109             }//End of if
110
111         }//End of if
112
113     }//End of for
114
115     //Retornar novo ArrayList com as alterações realizadas
116     result = educationalData;
117
118     return result;
119
120 }//End of treatEmpty
121
122 //Tratar variáveis não numéricas
123 public ArrayList treatNonNumeric(String config, ArrayList educationalData) {
124     ArrayList result = new ArrayList();
125
126     //Substituir caractere de marcação de valor inexistente por 0
127     for(int i = 0; i < educationalData.size(); i++) {
128         if(config.equals("NonNumericToZero")) {
129             if(educationalData.get(i).toString().equals("-")) {
130
131                 educationalData.set(i, (Object) 0);
132
133             }//End of if
134
135         }//End of if
136
137     }//End of for
138
139     //Retornar novo ArrayList com as alterações realizadas
140     result = educationalData;
141
142     return result;
143
144 }//End of treatNonNumeric
145

```

Figura 40 - Método responsável por verificar a existência de um cenário.

```

146 //Verificar a existência de um determinado cenário de mineração (Buscar pela classe)
147 public boolean checkActiveScenario(String className) {
148     boolean active = true;
149
150     try {
151         Class.forName(className);
152
153     } catch( ClassNotFoundException e ) {
154         active = false;
155
156     }//End of try...catch
157
158     return active;
159
160 }//End of checkActiveScenario method

```

Figura 41 - Método de mineração dos dados.

```

162 //Minerar os dados presentes em arquivos ARFF (treinamento, e teste se houver)
163 public void mining(
164     String trainingFileLocation,
165     String trainingFileName,
166     String testFileLocation,
167     String testFileName,
168     Boolean provideTestFile,
169     String algorithm,
170     String[] parameters
171 ) throws Exception {
172     //Criar arquivo de treinamento
173     BufferedReader breader = null;
174     breader = new BufferedReader(new FileReader(trainingFileLocation + trainingFileName));
175     Instances trainingSet = new Instances(breader);
176     trainingSet.setClassIndex(trainingSet.numAttributes() -1);
177     //Fechar buffered reader
178     breader.close();
179
180     //Minerar os dados utilizando o classificador J48
181     if(algorithm.toLowerCase().equals("j48")) {
182         J48 tree = new J48();
183         PART p = new PART();
184         DecisionTable dt = new DecisionTable();
185
186         //Remover o atributo com o ID das instâncias (pode influenciar nos resultados)
187         Remove remove = new Remove();
188         remove.setAttributeIndices("1");
189         remove.setInputFormat(trainingSet);
190         Instances instNew = Filter.useFilter(trainingSet, remove);
191
192         //Definir os parâmetros da mineração
193         tree.setBinarySplits(false);
194         tree.setConfidenceFactor((float) 0.90);
195         tree.setDebug(false);
196         tree.setMinNumObj(3);
197         tree.setNumFolds(3);
198         tree.setReducedErrorPruning(false);
199         tree.setSaveInstanceData(false);
200         tree.setSeed(1);
201         tree.setSubtreeRaising(false);
202         tree.setUnpruned(false);
203         tree.setUseLaplace(false);
204
205         //Construir o classificador
206         tree.buildClassifier(instNew);
207
208         //Criar arquivo de teste e realizar a classificação de suas instâncias.
209         if(provideTestFile) {
210             breader = new BufferedReader(new FileReader(testFileLocation + testFileName));
211             Instances testSet = new Instances(breader);
212             testSet.setClassIndex(testSet.numAttributes() -1);
213             Instances labeled = new Instances(testSet);
214
215             for(int i = 0; i < testSet.numInstances(); i++) {
216                 double classLabel = tree.classifyInstance(testSet.instance(i));
217                 labeled.instance(i).setClassValue(classLabel);
218             }
219
220             //End of for
221
222             BufferedWriter writer = new BufferedWriter(new FileWriter(testFileLocation + "labeled.arff"));
223             writer.write(labeled.toString());
224
225             writer.close();
226
227             //End of if
228
229             //Inserir resultado no BD
230             this.insertMiningResult(tree);
231
232             //End of if
233
234             //End of mining method

```

Figura 42 - Conector de cenários. Artefato através do qual o componente de mineração requer as configurações e dados providos pelos cenários (parte 1).

```

1 public class ScenariosConnector implements components.mining.spec.req.IScenarios {
2
3     private IManager dropoutsScenarioManager;
4     private IManager failingStudentsScenarioManager;
5     private IManager interactionsQualityScenarioManager;
6
7     public ScenariosConnector() {
8         this.dropoutsScenarioManager = ComponentFactoryDropoutsScenario.createInstance();
9         this.failingStudentsScenarioManager = ComponentFactoryFailingStudentsScenario.createInstance();
10        this.interactionsQualityScenarioManager = ComponentFactoryInteractionsQualityScenario.createInstance();
11    }
12
13    @Override
14    public Map<String, String> preprocessingConfig(String scenarioName) {
15
16
17
18
19
20        if (scenarioName.equals("InteractionQuality")) {
21
22            IInteractionsQualityScenario interactionsQualityScenario = (IInteractionsQualityScenario)
23            interactionsQualityScenarioManager.getProvidedInterface("IInteractionsQualityScenario");
24
25            return interactionsQualityScenario.preprocessingConfig();
26
27        } else if (scenarioName.equals("Dropouts")) {
28            IDropoutsScenario dropoutsScenario = (IDropoutsScenario) dropoutsScenarioManager.getProvidedInterface("IDropoutsScenario");
29
30            return dropoutsScenario.preprocessingConfig();
31
32        } else if (scenarioName.equals("FailingStudents")) {
33            IFailingStudentsScenario failingStudentsScenario = (IFailingStudentsScenario)
34            failingStudentsScenarioManager.getProvidedInterface("IFailingStudentsScenario");
35
36            return failingStudentsScenario.preprocessingConfig();
37
38        } else {
39
40            return null;
41
42        }
43    }

```

Figura 43 - Conector de cenários. (parte 2).

```

44  @Override
45  public Map<String, String> miningConfig(String scenarioName) {
46      if (scenarioName.equals("InteractionQuality")) {
47          IInteractionsQualityScenario interactionsQualityScenario = (IInteractionsQualityScenario)
interactionsQualityScenarioManager.getProvidedInterface("IInteractionsQualityScenario");
48
49          return interactionsQualityScenario.miningConfig();
50
51      } else if (scenarioName.equals("Dropouts")) {
52          IDropoutsScenario dropoutsScenario = (IDropoutsScenario) dropoutsScenarioManager.getProvidedInterface("IDropoutsScenario");
53
54          return dropoutsScenario.miningConfig();
55
56      } else if (scenarioName.equals("FailingStudents")) {
57          IFailingStudentsScenario failingStudentsScenario = (IFailingStudentsScenario)
failingStudentsScenarioManager.getProvidedInterface("IFailingStudentsScenario");
58
59          return failingStudentsScenario.miningConfig();
60
61      } else {
62
63          return null;
64
65      }
66  }
67
68  @Override
69  public Map<String, String> postprocessingConfig(String scenarioName) {
70      if (scenarioName.equals("InteractionQuality")) {
71          IInteractionsQualityScenario interactionsQualityScenario = (IInteractionsQualityScenario)
interactionsQualityScenarioManager.getProvidedInterface("IInteractionsQualityScenario");
72
73          return interactionsQualityScenario.postprocessingConfig();
74
75      } else if (scenarioName.equals("Dropouts")) {
76          IDropoutsScenario dropoutsScenario = (IDropoutsScenario) dropoutsScenarioManager.getProvidedInterface("IDropoutsScenario");
77
78          return dropoutsScenario.postprocessingConfig();
79
80      } else if (scenarioName.equals("FailingStudents")) {
81          IFailingStudentsScenario failingStudentsScenario = (IFailingStudentsScenario)
failingStudentsScenarioManager.getProvidedInterface("IFailingStudentsScenario");
82
83          return failingStudentsScenario.postprocessingConfig();
84
85      } else {
86
87          return null;
88
89      }
90  }
91
92  }

```

Figura 44 - Conector de cenários. (parte 3).

```

94  @Override
95  public Map<String, String> recommendingConfig(String scenarioName) {
96      if (scenarioName.equals("InteractionQuality")) {
97
98          IInteractionsQualityScenario interactionsQualityScenario = (IInteractionsQualityScenario)
interactionsQualityScenarioManager.getProvidedInterface("IInteractionsQualityScenario");
99
100         return interactionsQualityScenario.recommendingConfig();
101
102     } else if (scenarioName.equals("Dropouts")) {
103         IDropoutsScenario dropoutsScenario = (IDropoutsScenario) dropoutsScenarioManager.getProvidedInterface("IDropoutsScenario");
104
105         return dropoutsScenario.recommendingConfig();
106
107     } else if (scenarioName.equals("FailingStudents")) {
108         IFailingStudentsScenario failingStudentsScenario = (IFailingStudentsScenario)
failingStudentsScenarioManager.getProvidedInterface("IFailingStudentsScenario");
109
110         return failingStudentsScenario.recommendingConfig();
111
112     } else {
113
114         return null;
115
116     }
117
118  }
119
120  }

```