

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**ANÁLISE VISUAL PARA
MONITORAMENTO DE ALUNOS
DE CURSOS À DISTÂNCIA**

AUGUSTO WEIAND

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof^a. Dr^a Isabel Harb Manssour

**Porto Alegre
2016**

Dados Internacionais de Catalogação na Publicação (CIP)

W415a Weiland, Augusto

Análise visual para monitoramento de alunos de cursos à distância / Augusto Weiland. – 2016.

114 p.

Dissertação (Mestrado) – Faculdade de Informática,
Pontifícia Universidade Católica do Rio Grande do Sul.
Orientador: Profª. Drª. Isabel Harb Manssour

1. Educação a Distância. 2. Ambientes Virtuais
Compartilhados. 3. Mineração de Dados (Computação).
4. Informática. I. Manssour, Isabel Harb. II. Título.

CDD 23 ed. 005.72

Salete Maria Sartori CRB 10/1363
Setor de Tratamento da Informação da BC-PUCRS



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Análise Visual para Monitoramento de Alunos de Cursos a Distância" apresentada por Augusto Weiland como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, aprovada em 1º de março de 2016 pela Comissão Examinadora:

Isabel Manssour

Profa. Dra; Isabel Harb Manssour -
Orientadora

PPGCC/PUCRS

Milene Selbach

Profa. Dra. Milene Selbach Silveira -

PPGCC/PUCRS

Carla Maria Dal Sasso Freitas

Profa. Dra. Carla Maria Dal Sasso Freitas -

UFRGS

Joyce Munarski Pernigotti

Profa. Dra. Joyce Munarski Pernigotti -

FACOS

Homologada em 19/05/2016, conforme Ata No. 010 pela Comissão Coordenadora.

Luiz Gustavo Leão Fernandes

Prof. Dr. Luiz Gustavo Leão Fernandes
Coordenador.

DEDICATÓRIA

Dedico este trabalho a minha mãe, esposa e mestres.

“Não é o que você faz, mas quanto amor você dedica no que faz que realmente importa.”
(Agnes Gonxha Bojaxhiu (Madre Teresa de Calcutá))

AGRADECIMENTOS

Agradeço primeiramente a minha mãe, Silvana, que me auxiliou sempre que foi necessário, sendo em diversas ocasiões um pai e um ombro amigo, com quem muito aprendi e sempre sentirei saudades. Assim como agradeço ao meu indispensável Mestre, Gilberto, que me ensinou muito, e jamais esquecerei.

Não posso deixar de agradecer de forma especial a minha esposa Fernanda, pelas noites em claro, pela paciência e pelas diversas vezes que me levantou e me deu forças quando elas já não me habitavam.

Aos Mestres, em especial à Prof^a. Isabel, que me recebeu na PUCRS, que orientou-me com muita dedicação, paciência e atenção na elaboração desta pesquisa, e na oportunidade de participar de uma equipe de pesquisa fantástica no Centro de Pesquisa em Computação Aplicada, onde aprendi muito e tive a oportunidade de trabalhar com grandes pessoas.

Agradeço ao Convênio PUCRS/HP pela bolsa de estudos bem como o aprendizado durante o projeto junto aos colegas que me auxiliaram nesta caminhada.

Ao Programa de Pós-Graduação em Ciência da Computação e a todos os professores dos quais pude conviver durante esse período, que me ensinaram muito mais do que apenas aquilo que o ofício requer, me ensinaram a pensar.

A todos aqueles que de alguma forma contribuíram ou torceram pela concretização desta pesquisa e desta trajetória.

ANÁLISE VISUAL PARA MONITORAMENTO DE ALUNOS DE CURSOS À DISTÂNCIA

RESUMO

Com o avanço da tecnologia, a educação a distância tem sido muito discutida nos últimos anos, especialmente com o surgimento de diversos tipos de Ambientes Virtuais de Aprendizagem (AVA's). Estes ambientes, quando utilizados em cursos de educação a distância, normalmente geram uma grande quantidade de dados devido ao elevado número de alunos e as diversas tarefas que envolvem as suas interações. Técnicas de mineração de dados auxiliam na descoberta de conhecimentos implícitos que possibilitem dar suporte à tomada de decisão. Porém, eventualmente surgem dificuldades no entendimento dos resultados obtidos pela mineração, devido ao volume de dados analisado. Neste caso, o uso de técnicas de visualização e interação auxiliam nesta tarefa. Este trabalho tem como objetivo apresentar o desenvolvimento de uma abordagem de análise visual, que utiliza algoritmos de mineração de dados e técnicas de visualização para auxiliar no acompanhamento de alunos de cursos a distância nas instituições que utilizam ambientes virtuais de aprendizado. Estes alunos são classificados considerando o seu desempenho, possibilitando a investigação e predição de possíveis aprovações, reprovações ou evasões. É possível analisar tanto o comportamento geral dos alunos de uma disciplina selecionada, como seus comportamentos individuais. Comparações de desempenho de um aluno entre diferentes disciplinas, e das interações realizadas em um conjunto de disciplinas também são permitidas. Testes iniciais demonstraram que foi possível efetuar previsões de maneira satisfatória, assim como, possibilitar aos usuários visualizações e interações para interpretar estas informações advindas dos algoritmos de mineração.

Palavras-Chave: Ambientes Virtuais de Aprendizagem, Análise Visual, Descoberta de Conhecimento, Educação a Distância.

VISUAL ANALYSIS FOR MONITORING STUDENTS IN DISTANCE COURSES

ABSTRACT

With the technology advancement, distance education has been very discussed in recent years, especially with the emergence of several kinds of Virtual Learning Environments (VLE's). These environments used in distance education courses, usually generate a lot of data due to the high number of students and the various tasks which involve their interactions. Thus, arises the need to search efficient and intelligent ways to find relevant information. Data mining techniques help in the discovery of implicit knowledge that can support decision making. However, eventually appear difficulties in understanding the obtained results of the mining due to the analyzed volume. In these cases, the use of visualization and interaction techniques assists in this task. The main goal of this work is to present the development of a visual analysis approach that uses data mining algorithms and visualization techniques to help monitoring students of distance learning courses in the institutions that use virtual learning environments. These students are classified considering their performance, providing ways to investigate and predict possible approvals, disapprovals and evasions. The visualizations aim to improve the understanding of the generated data by the mining algorithms, providing different ways of interaction. It is possible to analyze both the general behavior of students in a selected course, as their individual behaviors. Performance comparisons of a student between different courses, and from interactions performed in a set of courses are also allowed. Initial tests demonstrated that it was possible to make predictions in a satisfactory way, as well as enable visualizations and interactions to the users for interpreting the information resulting from mining algorithms.

Keywords: Virtual Learning Environment, Visual Analysis, Knowledge Discovery, Distance Education.

LISTA DE FIGURAS

Figura 2.1 – Moodle.	26
Figura 2.2 – TelEduc.	26
Figura 2.3 – Solar.	27
Figura 2.4 – BlackBoard.	28
Figura 2.5 – Exemplo de Table Lens que apresenta uma análise da população com idade maior de 65 anos em mais de 200 regiões da Europa.	30
Figura 2.6 – Exemplo de Coordenadas Paralelas com dados pluviométricos relacionados em cada coluna.	31
Figura 2.7 – Rede de padrões de co-edição da Wikipédia demonstrada com a técnica de Grafo.	32
Figura 2.8 – Exemplo de mapa representando dados variados sobre a agricultura dos Estados Unidos.	33
Figura 2.9 – Exemplo de Árvore de Decisão sobre a aprovação de um aluno utilizando a nota e a frequência como critérios.	40
Figura 2.10 – Processos de Descoberta de Conhecimento.	42
Figura 3.1 – Exemplo de interface que apresenta uma análise de trocas de mensagens e teste de atividade (representadas nos eixos x e y). As cores representam os agrupamentos, os círculos o centro do grupo, e cada estrela um usuário.	46
Figura 3.2 – Interface da aplicação AAT para realização do cadastro de um padrão.	47
Figura 3.3 – Matriz cognitiva do desempenho dos estudantes baseado em um <i>quiz</i> relacionado com conceitos do domínio estudado.	48
Figura 3.4 – Visualização de tópicos de discussão de um fórum.	48
Figura 3.5 – Apresentação dos acessos dos estudantes ao curso.	49
Figura 3.6 – Primeiro componente da abordagem desenvolvida por Romero et al., que possibilita a seleção de um conjunto de dados e atributos dos alunos, advindos do AVA Moodle, para posterior geração de gráficos estatísticos.	50
Figura 3.7 – Segundo componente da abordagem desenvolvida por Romero et al., que permite ao usuário selecionar um arquivo e atributos, para aplicação de um algoritmo de mineração. Os resultados podem ser visualizados ou baixados em formato texto ou PDF.	51
Figura 4.1 – Abordagem proposta e apresentada na primeira etapa de entrevistas, que apresenta a interação com o mouse sobre as barras de total de acessos.	56

Figura 4.2 – Ambientes já utilizados pelos entrevistados.	58
Figura 4.3 – Ambientes em utilização pelos entrevistados.	58
Figura 4.4 – Demonstrativo dos elementos gráficos disponibilizados no sistema de monitoramento utilizado.	59
Figura 4.5 – Opinião dos entrevistados para a avaliação do aluno.	59
Figura 5.1 – Modelagem UML das tabelas internas para armazenamento dos dados.	65
Figura 5.2 – Modelo UML das tabelas e campos utilizados do AVA Moodle.	67
Figura 5.3 – Sequência de ações na análise de uma nova disciplina.	68
Figura 5.4 – Visualização dos acessos dos alunos e do resultado da mineração.	70
Figura 5.5 – <i>Table lens</i> com uma linha selecionada (a). Visualização de bolhas com a interação de mouse sobre uma bolha (b).	71
Figura 5.6 – Visualização de bolhas com os círculos agrupados de acordo com a classificação do aluno.	72
Figura 5.7 – Seleção de aluno, recursos e disciplina.	72
Figura 5.8 – Seleção de disciplina.	72
Figura 5.9 – Seleção de aluno e recursos.	73
Figura 5.10 – Mouse sobre o gráfico de colunas empilhadas (a, b). Seleção de período (c).	74
Figura 5.11 – Visualizações da mineração dos fóruns do AVA utilizando TF-IDF. O tamanho da bolha representa o índice de similaridade, cada coluna um fórum e cada linha um aluno <i>a</i> , assim como em <i>b</i> , porém o índice é demonstrado com variações de cor. Já em <i>c</i> , cada anel representa um fórum e cada coluna um aluno, utilizando também variações de cor para a similaridade.	75
Figura 5.12 – Interações nas visualizações de mineração dos fóruns.	76
Figura 5.13 – Comparação das classificações e acessos do aluno nas disciplinas.	76
Figura 5.14 – Interação nas comparações de classificação do aluno.	77
Figura 5.15 – Interações nas barras de comparação de recursos do aluno.	77
Figura 5.16 – Comparação entre disciplinas.	78
Figura 5.17 – Interações nas barras de comparação de classificações dos alunos nas disciplinas.	79
Figura 5.18 – Interações nas barras de comparação de médias de acesso aos recursos nas disciplinas.	80
Figura C.1 – Tela Inicial.	107
Figura C.2 – Linha clicada.	108

Figura C.3 – Ordenação por coluna.	108
Figura C.4 – Filtro por recurso.	109
Figura C.5 – Filtro por aluno.	109
Figura C.6 – Mouse sobre as barras.	110
Figura D.1 – Tela da primeira semana.	111
Figura D.2 – Tela da segunda semana.	112
Figura D.3 – Tela da terceira semana.	113
Figura D.4 – Tela da quarta semana.	114

SUMÁRIO

1	INTRODUÇÃO	21
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	EDUCAÇÃO A DISTÂNCIA	23
2.1.1	AMBIENTES VIRTUAIS DE APRENDIZAGEM	24
2.1.2	CURSOS ON-LINE ABERTOS E MASSIVOS	28
2.2	VISUALIZAÇÃO DE DADOS MULTIVARIADOS	29
2.2.1	<i>TABLE LENS</i>	30
2.2.2	COORDENADAS PARALELAS	31
2.2.3	GRAFOS	32
2.2.4	MAPAS	33
2.3	ANÁLISE VISUAL	33
2.4	INTERAÇÃO E ANÁLISE VISUAL	35
2.5	MINERAÇÃO DE DADOS	37
2.5.1	ALGORITMOS DE MINERAÇÃO	39
2.6	DESCOBERTA DE CONHECIMENTO	41
3	TRABALHOS RELACIONADOS	45
3.1	ANÁLISE DOS TRABALHOS	52
4	LEVANTAMENTO DE REQUISITOS	55
4.1	DESCRIÇÃO DO PROCESSO	55
4.2	PERFIL	56
4.2.1	EXPERIÊNCIA	57
4.3	OPINIÃO SOBRE AS FERRAMENTAS EXISTENTES	58
4.4	AVALIAÇÃO DA PROPOSTA	60
5	ABORDAGEM DE ANÁLISE VISUAL PARA ACOMPANHAMENTO DISCENTE	63
5.1	ENTRADA DE DADOS	64
5.2	IMPLEMENTAÇÃO	68
5.2.1	ALGORITMOS DE MINERAÇÃO	69
5.2.2	VISUALIZAÇÕES DE RECURSOS E ACESSOS	69
5.2.3	FORMAS DE INTERAÇÃO	72

5.2.4	VISUALIZAÇÕES DOS FÓRUNS E COMPARAÇÕES	73
6	AVALIAÇÃO DE RESULTADOS	81
6.1	METODOLOGIA DE AVALIAÇÃO	81
6.2	DISCIPLINA E PERFIL DO TUTOR	82
6.3	AVALIAÇÃO SEMANAL DA ABORDAGEM	83
6.4	ANÁLISE FINAL	85
6.5	CONSIDERAÇÕES DA ANÁLISE DA ABORDAGEM	86
7	CONCLUSÕES E TRABALHOS FUTUROS	87
	REFERÊNCIAS	91
	APÊNDICE A – Termo de consentimento livre e esclarecido	101
	APÊNDICE B – Questionário de levantamento de requisitos	103
	APÊNDICE C – Telas apresentadas no levantamento de requisitos	107
	APÊNDICE D – Telas capturadas durante a avaliação da disciplina	111

1. INTRODUÇÃO

A popularização da internet e o avanço da tecnologia trouxeram consigo o aperfeiçoamento da comunicação à distância, a disseminação de informações, conhecimentos e o aumento da globalização. O surgimento, por exemplo, de diversas redes sociais, ambientes virtuais de aprendizagem e, até mesmo, de dispositivos vestíveis, levou à geração de grandes volumes de dados de diversos tipos, os quais, na maioria das vezes, são armazenados de forma não estruturada, tornando complexa a tarefa de analisá-los [76].

Da mesma forma, acontece com o crescimento exponencial dos cursos a distância. E, devido ao grande número de alunos que participam destes cursos, principalmente considerando os cursos on-line abertos e massivos (MOOC's), aumentou também o volume de dados gerado pelos ambientes de aprendizagem que eles utilizam. Porém, apesar do crescimento, estes cursos enfrentam um grande problema: a evasão dos alunos [7, 19, 42, 58, 101, 113].

Neste contexto, surge a demanda de fornecer mecanismos para a análise dos dados gerados pelos ambientes de aprendizagem para acompanhar o desempenho dos alunos, pois o seu grande volume dificulta uma análise manual [62]. Além disso, na medida em que os cursos à distância se popularizam e aumenta a competitividade entre eles, tais análises ganham ainda mais importância, pois possibilitam que as organizações consigam desenvolver um trabalho mais adequado ao perfil de cada aluno, e também acompanhar o seu desenvolvimento no curso.

Considerando estes fatores, alguns trabalhos que implementam algoritmos de mineração de dados para auxiliar nesta tarefa, foram desenvolvidos nos últimos anos [2, 19, 42, 69, 86]. Estes trabalhos apresentam algoritmos de mineração de dados relevantes, porém, as visualizações, quando existem, apenas mostram os resultados em tabelas, sem utilizar outros elementos gráficos e interativos que poderiam auxiliar na análise dos dados.

Uma análise sobre estes ambientes, e se possuem ou não abordagens de visualização, foi feita e descrita no capítulo 3. Observou-se, então, que o fato de que os *datasets* de alguns destes ambientes foram considerados pequenos por alguns pesquisadores pode ser um dos fatores que acarretam nas poucas abordagens de visualização desenvolvidas. Entretanto, algumas instituições especializadas em Educação a Distância (EAD) desenvolvem novas modalidades de ensino, como os MOOC's. Estes ambientes, assim como os ambientes virtuais de aprendizagem (AVA's) de grandes instituições, tendem a fornecer *datasets* maiores para estas análises, que poderão se beneficiar do desenvolvimento de visualizações.

Considerando este contexto, foi elaborada a seguinte questão de pesquisa: "De que forma diferentes técnicas de visualização e interação, integradas com técnicas de mi-

neração de dados, podem auxiliar professores e administradores no processo de acompanhamento e predição de desempenho de alunos em ambientes virtuais de aprendizagem?"

Baseando-se nesta questão de pesquisa, este trabalho apresenta uma abordagem para o acompanhamento de alunos de cursos à distância, integrando diferentes formas de análise visual e algoritmos de mineração aos dados gerados pelos ambientes de aprendizagem utilizados. Assim, com o apoio de técnicas de visualização, procura possibilitar aos professores, tutores e administradores o acompanhamento de seus cursos, fornecendo uma visão geral dos mesmos, com o foco nos alunos e nos recursos utilizados no ambiente. Estas informações são apresentadas juntamente com os recursos disponibilizados na disciplina a ser analisada pela abordagem, permitindo também fazer uma análise individual dos alunos e entre disciplinas. Desta maneira, pretende-se auxiliar na diminuição de possíveis evasões e reprovações.

O restante deste trabalho está organizado da seguinte forma: primeiro é apresentada a fundamentação teórica desta proposta. No capítulo 3 são apresentados os trabalhos relacionados a EAD e Análise Visual. Na sequência, no capítulo 4, são descritas as entrevistas realizadas com docentes para o levantamento dos requisitos da abordagem desenvolvida. Logo após, no capítulo 5, a abordagem desenvolvida é descrita. Os testes efetuados em um ambiente de produção são apresentados no capítulo 6. Por fim, no último capítulo são colocadas algumas conclusões sobre o trabalho realizado, incluindo possibilidades para sua extensão.

2. FUNDAMENTAÇÃO TEÓRICA

Com o amplo uso de sistemas informatizados em diversas áreas, como comércio, saúde e educação, cresce a cada instante o volume de dados gerados e armazenados em bases de dados digitais. Conseqüentemente, esforços são empregados para analisá-los, uma vez que é possível identificar padrões novos e relevantes a partir destes dados. Neste contexto, e para o entendimento deste trabalho, é relevante compreender os seguintes tópicos, que são descritos neste capítulo: Educação a Distância, Visualização, Análise Visual, Interação, Mineração de Dados e Descoberta de Conhecimento.

2.1 Educação a Distância

A Educação a Distância aparece cada vez mais no contexto da sociedade contemporânea como uma modalidade de educação mais adequada as novas demandas educacionais, decorrentes das mudanças pessoais e profissionais vividas na sociedade atualmente [9].

Com isso, percebe-se que há uma igual ascensão de sistemas que se propõe a desenvolver Ambientes Virtuais de Aprendizagem (AVA), de modo a agregar valor às estratégias de ensino-aprendizagem [67, 76]. Dentre os mais comumente utilizados [20, 28, 81], pode-se citar: Moodle, TelEduc, Solar e BlackBoard, os quais estão descritos na seção 2.1.1.

Juntamente com este alto crescimento surge também outra modalidade de EAD, os Cursos On-line Abertos e Massivos (MOOC's, ou *Massive Open On-line Course*). Estes cursos tem a intenção de possibilitar a modalidade de aprendizado a distância em um ambiente diferenciado, com mais participantes que em cursos EAD tradicionais, além de, em sua maioria, um menor valor monetário [112].

Com todo este rol de ambientes surgindo, identifica-se a necessidade de desenvolvimento e utilização de sistemas que auxiliam os professores e administradores a perceberem de uma melhor forma o estado de sua organização. Assim, tem sido implementados sistemas de visualização integrados a suas bases, para possibilitar a compreensão e a tomada de decisões de forma mais rápida e direta.

Todos estes mecanismos são utilizados também para evitar um dos grandes problemas do ensino a distância, a evasão. De acordo com a Associação Brasileira de Educação a Distância (ABED) [31], a falta de tempo para estudar ou participar do curso, a falta de adaptação à metodologia e o aumento de tarefas no trabalho, são apontados como a principal causa para a evasão dos alunos, o que chega a representar 25% dos casos, segundo a entidade.

Segundo Tinto [101], uma das definições mais coesas sobre a evasão, a qual foi utilizada neste trabalho, descreve-a como o abandono do aluno a instituição, permanecendo infrequente e sem as devidas avaliações, em qualquer momento antes da conclusão do curso. O que ocorre também no âmbito de EAD, quando o aluno deixa de frequentar o ambiente virtual utilizado e, não participa ativamente das propostas solicitadas, é considerado evadido, mesmo a modalidade a distância possuindo flexibilidade de tempo e local de estudo [58].

Assim como Tinto e Kampff et. al. [58, 101], demais autores como [7, 19, 42, 113] desenvolveram através dos anos, métodos e abordagens para, além de tentar identificar as causas da evasão, possibilitar prever estes casos de maneira a evita-los, ou reduzi-los. Alguns destes modelos são descritos mais adiante, no capítulo 3.

2.1.1 Ambientes Virtuais de Aprendizagem

Os Ambientes Virtuais de Aprendizagem tornam-se espaços cada vez mais acessados e discutidos dentro da academia. Andrade e Vicari [3] destacam que a Web se tornou um espaço de apoio ao desenvolvimento de cursos à distância, suportados por uma diversidade de ambientes informatizados on-line direcionados às atividades educacionais. Tendo início em cursos semipresenciais [76], os AVA's se desenvolvem cada vez mais rapidamente, acompanhando a evolução das tecnologias de informação e comunicação (TIC's).

Os AVA's auxiliam a diferenciar o processo metodológico de ensino e aprendizagem, no qual o professor exerce o papel de mediador durante o processo de aprendizagem, com o aluno no centro do processo [88]. Desta maneira, o professor deixa o papel de transmissor de conhecimentos, e assume um perfil de moderador, que prepara o espaço de diálogo e interação entre os alunos. Assim, o aluno passa a ser construtor e socializador do seu próprio conhecimento [56].

Torres e Fialho [102] veem no uso das tecnologias de informação e comunicação em ambientes EAD, uma metodologia para aperfeiçoar o aprendizado colaborativo, trazendo consigo, através do passar dos anos, um agregado de formatos distintos à estes ambientes.

Para Teles [98], o ensino baseado na Web possui um grande potencial para modelos pedagógicos colaborativos, tendo em vista as seguintes características:

- Comunicação de grupo a grupo: Permitindo que o estudante se comunique diretamente com outros colegas através do ambiente virtual;
- Independência temporal: Observando que o acesso poderá ser realizado a qualquer momento, possibilitando tempo necessário ao aluno para suas reflexões e análises de temas postados em discussões no ambiente;

- Independência espacial: O acesso ao ambiente on-line pode ser realizado de qualquer lugar;
- Interação entre os participantes: Comunicação mediada por computadores que requer do estudante a organização de ideias através da escrita.

De acordo com Schlemmer [90], o ponto mais importante a ser observado na utilização destes ambientes é o critério didático-pedagógico com o qual ele foi construído. Para ela, o desenvolvimento de um produto para a educação é permeado por uma concepção epistemológica, ou seja, por uma crença de como se dá a aquisição do conhecimento, relacionado com o aprendizado do sujeito. E de fato, os AVA's tendem a reduzir a distância transacional na educação on-line [66] no momento em que se constituem de ambientes coletivos que favorecem a interação entre os sujeitos participantes [8], reduzindo de certa forma, esta distância, através das diversas formas de interação disponíveis nestes ambientes.

Já no âmbito acadêmico, alguns AVA's se destacam por sua utilização em diversas instituições nacionais e internacionais como já descrito anteriormente, a saber: Moodle, TelEduc, Solar e BlackBoard. Estes são descritos brevemente a seguir.

O Moodle (*Modular Object-Oriented Dynamic Learning Environment*) foi desenvolvido pelo australiano Martin Dougiamas, que possui formação em ciência da computação e em educação por volta dos anos 1990 [64, 81]. Desde o princípio ele acreditava que um sistema de gerenciamento de cursos deveria ser elaborado por alguém com formação na área da educação e não da engenharia. De acordo com Franco [40], Dougiamas desenvolveu o Moodle inspirado na epistemologia sócio-construcionista, baseando-se na ideia de que “as pessoas aprendem melhor quando estão engajadas em um processo social de construção de conhecimento, construindo artefatos para os outros” [22]. Também por este motivo ele é disponibilizado de forma *Open Source* a comunidade.

Segundo Garcia e Lacleita [41] o Moodle apresenta uma estrutura modular, juntamente com uma ampla comunidade de desenvolvedores. Além disso, traz uma grande quantidade de documentação, escalabilidade, facilidade de uso, interoperabilidade, estabilidade e segurança. A figura 2.1 ilustra a interface de um curso, na qual apresentam-se alguns dos seus recursos, tais como fóruns, links, atividades, questionários, calendário e tarefas.

Aliando todos estes fatores o Moodle destaca-se como um dos principais AVA's em uso até o presente, como é possível observar nas estatísticas disponíveis na própria página do ambiente [72]. Sua comunidade também se destaca, possuindo mais de 130.000 usuários registrados em 160 países e 75 idiomas, entre estes, diversas entidades, professores autônomos e até grandes universidades e escolas [40]. Também observa-se este uso na literatura, como em Leite [64], Ribeiro et al. [81] e Franco [40].



Figura 2.1: Moodle.

Fonte: Interface de um curso no AVA Moodle [72].

Segundo Ribeiro e Mendonça [81] o TelEduc é uma plataforma *Open Source* e um ambiente de suporte à EAD (figura 2.2). O seu desenvolvimento é realizado de acordo com as necessidades, tanto tecnológicas como metodológicas, por desenvolvedores do Núcleo de Informática Aplicada à Educação (NIED) e do Instituto de Computação (IC) da Universidade Estadual de Campinas (UNICAMP) [96, 84, 83]. Foi desenvolvido inicialmente em 1997 com a dissertação de mestrado de Alessandra de Dutra e Cerceau, intitulada Formação à Distância de Recursos Humanos para Informática Educativa [17].



Figura 2.2: TelEduc.

Fonte: Interface de um curso no AVA TelEduc [97].

Cada curso que utiliza o ambiente TelEduc pode utilizar um subconjunto de ferramentas disponíveis, sendo a escolha de quais ferramentas que serão utilizadas atrelada a metodologia adotada por cada professor. Dentre estas ferramentas disponíveis, destacam-se as seguintes: agenda, atividades, leituras, perguntas frequentes, mural, correio, grupos e acessos [84, 83].

O AVA Solar (*Sistema On-line de Aprendizado*[94]) foi desenvolvido em 2001 através do Instituto UFC Virtual, da Universidade Federal do Ceará, o qual é voltado para a pesquisa e desenvolvimento de tecnologias e recursos para serem aplicados a EAD.



Figura 2.3: Solar.

Fonte: Interface de um curso no AVA Solar [95].

A interface do Solar, apresentada na figura 2.3, propõe uma aplicação distribuída, que pode ter seus componentes de software e base de dados em diferentes computadores. No painel da esquerda, os itens disponibilizados pelo professor neste curso podem ser acessados, e ao clicar em algum destes itens, a página central é carregada com este recurso. Como pode-se observar, alguns dos principais recursos disponibilizados por este AVA são: fórum, chat, webconferências, mensagens, links e agenda.

O BlackBoard [12] foi criado em 1997, em colaboração com membros do corpo docente da Universidade Cornell, com a intenção de fornecer a educadores e professores de diversas áreas um ambiente em que fosse possível montar um curso virtual. Ele possui uma interface para criação de cursos que combina ferramentas educacionais e administrativas e, é divulgado como de fácil manipulação, flexível e prático. Desta forma, não requer conhecimento especializado em linguagens de programação ou de qualquer outra habilidade avançada na área de computação. Com isso, os professores/instrutores e/ou designers podem adicionar componentes a suas aulas, abrigar um curso completo na web e publicar materiais suplementares para seus cursos presenciais de forma simples [23]. A figura 2.4 ilustra um exemplo de interface de curso. Como nos AVA's citados anteriormente, este tam-

bém utiliza o layout de duas colunas, permanecendo a esquerda os recursos utilizados no curso em questão, e a direita o item selecionado.

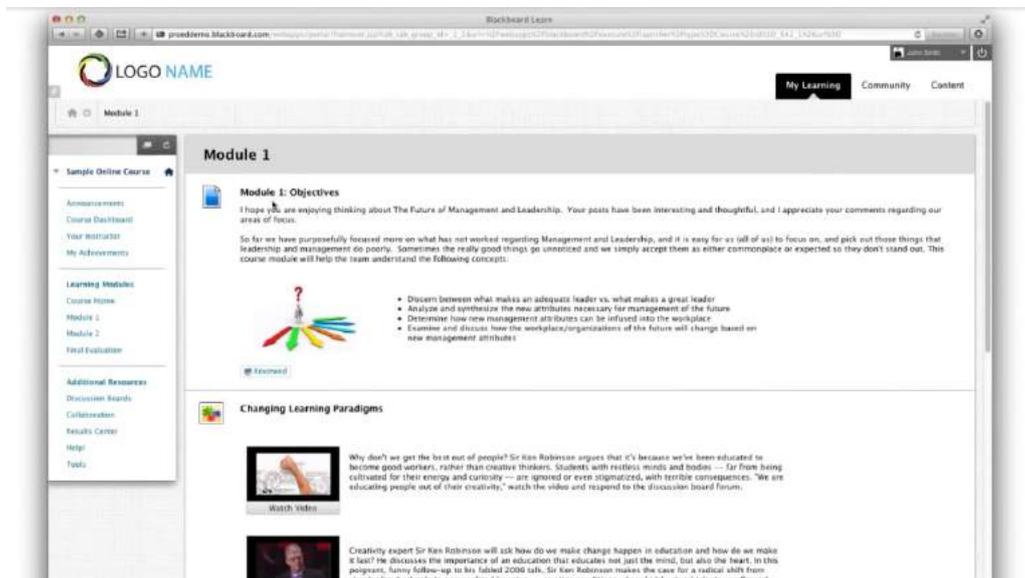


Figura 2.4: BlackBoard.

Fonte: Interface de um curso no AVA BlackBoard [12].

O BlackBoard possui diversas funções para comunicação e conteúdo, sendo as principais: anúncios, que os professores podem postar para os alunos lerem; possibilidade de postar artigos, vídeos; chat; discussões; e-mail; calendário; avaliações e biblioteca de mídia.

2.1.2 Cursos On-line Abertos e Massivos

Os MOOC's surgiram como uma modalidade de EAD que se propõe à permitir o acesso gratuito a cursos de curta duração em universidades renomadas internacionalmente [29]. Com isto, tem se difundido mundialmente, tanto nas instituições quanto no desenvolvimento de novas abordagens de ensino.

Segundo Yuan et al. [112] os MOOC's tem recebido uma atenção especial nos últimos tempos com a promessa de fornecer cursos de nível universitário de baixo, ou até mesmo sem, custo, com o potencial de transformar de uma forma abrangente os métodos tradicionais de ensino. O surgimento de plataformas de aprendizagem abertas, como o EDX [36], e novas *Start-Up's* como Coursera [27] e Udacity [104], também têm sido parte integrante desta nova modalidade.

Os MOOC's geralmente oferecem cursos de extensão, se diferenciando das modalidades tradicionais, como os tecnólogos e graduações em EAD, os quais não são clas-

sificados como MOOC's, tendo em vista além da formação, a abrangência, documentos e protocolos exigidos.

Apesar de serem cursos abertos, é necessário que o aluno se inscreva em uma plataforma, comumente sem exigência de documentação, sem seleção e sem datas de entrada fixas [29]. Com isto, estes ambientes atingem um número considerável de alunos no mundo todo. Este modelo de curso apresenta o fato das universidades estarem interessadas não somente em dispor seus cursos de forma regional, mas também de fornecerem uma visibilidade mundial. [29].

2.2 Visualização de Dados Multivariados

Diversas representações gráficas têm sido utilizadas desde os primórdios da humanidade para demonstrar significados distintos. Segundo Card et al. [16], com a evolução da ciência, estas representações passam a ter significados mais específicos, atribuindo o propósito da visualização à sua percepção (*insigth*), e não pela simples representação de figuras. Desta forma, sob este panorama, procuram-se desenvolver as representações visuais que mais se apropriam as características inerentes a cada aplicação, inclusive sob suas próprias normas, como, por exemplo, gráficos matemáticos e mapas cartográficos [60].

Paralelamente a isto, ocorre o crescente aumento do volume de dados armazenado e disponível para análise, e também diferentes abordagens que facilitam a visualização destas informações. Assim, novas formas de representação são desenvolvidas para auxiliar a busca por conhecimentos contidos nestes dados, não abstando-se a análise de números em tabelas. Desta maneira, a implementação de um formato gráfico, baseado não somente nas representações visuais, mas também em mecanismos de interação se torna cada vez mais presente nas visualizações [14, 99].

Mas os avanços das representações gráficas também perpassam por um processo de transformação, os dados armazenados e passíveis de visualizações começaram a não ter somente um ou dois tipos estruturais definidos, como números e textos, por exemplo [78]. No decorrer dos anos, diversos novos formatos foram surgindo e se integrando as bases de dados, tais como: links, documentos, coordenadas geográficas e objetos. E, assim, novas necessidades de representar visualmente estes dados multivariados, de forma a possibilitar seu entendimento, compondo visualizações com várias dimensões [59]. Diversas técnicas de visualização são desenvolvidas, de maneira a auxiliar nesta tarefa, como por exemplo: *table lens*, coordenadas paralelas, grafos e mapas [46, 78]. Estas técnicas são descritas nas subseções a seguir, pois foram identificadas como as mais relacionadas com este trabalho.

2.2.1 Table Lens

Table lens é uma técnica de visualização de informações utilizada para explorar dinamicamente grandes quantidades de dados tabulares. Foi desenvolvida por Rao et. al. [77] em meados de 1994, para simplificar a interação com grandes tabelas de dados. Advinda do modelo de planilha, esta técnica mostra os valores dos dados de cada registro em colunas e linhas. Cada linha da *table lens* representa um único registro e as colunas representam um indicador específico [77].

A figura 2.5 apresenta um exemplo no qual é possível observar informações de mais de 200 regiões da Europa classificadas em "População com idade 65+". Além disso, pode-se verificar que foi aplicado *zoom* em três linhas para analisar os números representados em cada coluna. Portanto, a grande vantagem desta abordagem é ter as habilidades combinadas para classificar um grande número de registros e aplicar filtros em áreas de interesse nos dados, possibilitando também a sua visualização numérica exata. Deste modo, esta abordagem permite a interação com grandes estruturas de informação, distorcendo dinamicamente o layout espacial da estrutura de acordo com diferentes níveis [73].

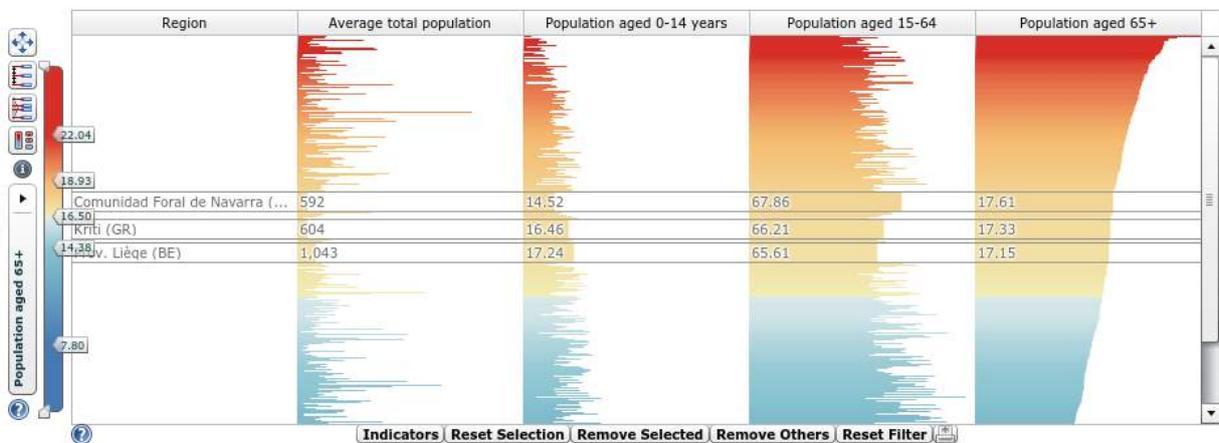


Figura 2.5: Exemplo de Table Lens que apresenta uma análise da população com idade maior de 65 anos em mais de 200 regiões da Europa.

Fonte: NCOMVA [73].

A relação entre o *zoom* e contexto, apoia a busca de padrões no quadro geral e possibilita a investigação de forma dinâmica de detalhes interessantes, sem perder o contexto. Contudo, esta abordagem não deixa de oferecer a característica mais importante de uma tabela, a regularidade de seu conteúdo com informações ao longo de linhas e colunas. A *table lens* pode ser usada para explorar padrões em tabelas e investigar várias estatísticas usando um conjunto de dados multivariado [77].

2.2.2 Coordenadas Paralelas

Esta técnica foi inicialmente apresentada por Inselberg e Dimsdale [52] como uma técnica de geometria computacional e posteriormente contextualizada em visualização de informação por Wegman [107].

A técnica de coordenadas paralelas destaca-se pela sua perspectiva multidimensional retratada em sua representação visual. Onde, cada dimensão é mapeada para uma coordenada bidimensional, usando eixos equidistantes e paralelos a um dos principais eixos (x ou y). Cada eixo representa uma dimensão do conjunto de dados, sobre o qual é mapeado linearmente, do menor ao maior valor, o intervalo de dados correspondente [14].

Cada item do conjunto de dados é exibido como uma linha longitudinal que intercepta cada eixo no ponto correspondente ao valor do atributo associado àquela grandeza. Um exemplo desta visualização que descreve uma base de dados pluviométrica é demonstrado na figura 2.6.

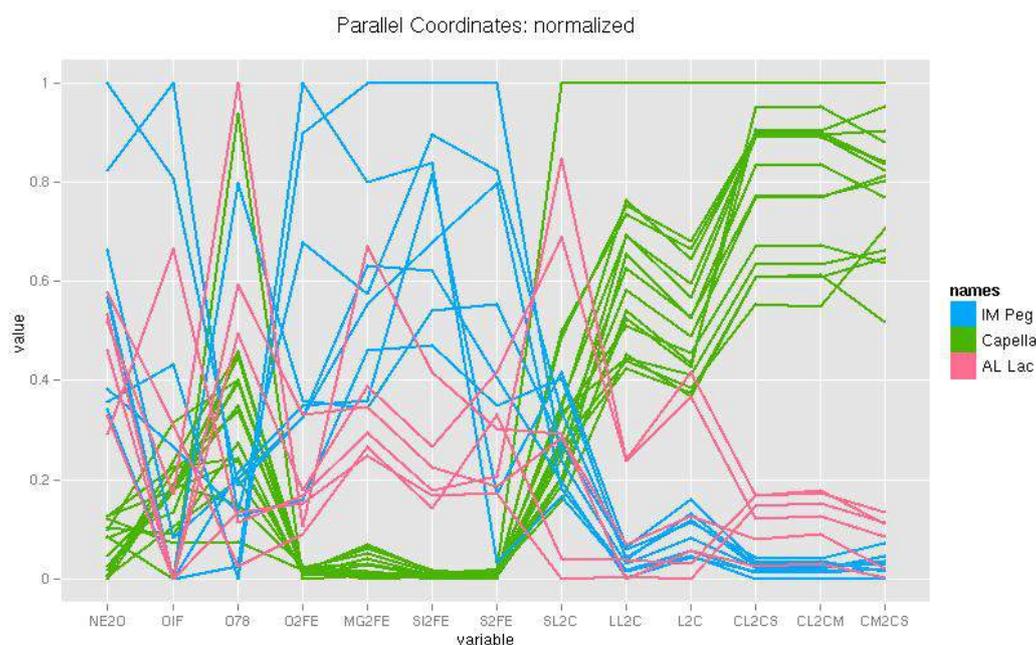


Figura 2.6: Exemplo de Coordenadas Paralelas com dados pluviométricos relacionados em cada coluna.

Fonte: Branco [14].

Pode-se observar que esta técnica gera uma representação plana, transformando representações de dados multivariados em padrões bidimensionais [52, 108], e possibilitando identificar características como distribuição ou aglomeração de dados em grandezas específicas.

Por outro lado, um dos principais problemas apresentados nesta técnica é a sobreposição de linhas quando o conjunto de dados e as grandezas desejadas são muitas. Keim

e Kriegel [61] afirmam que podem ser representados até 1.000 itens em uma visualização sem sobreposição acentuada de linhas.

2.2.3 Grafos

Grafos são estruturas utilizadas para modelar relações entre os objetos, eles são compostos por vértices e arestas que os conectam. Um grafo pode ser não-dirigido, o que significa que não há nenhuma distinção entre os dois vértices associados a cada aresta, ou as suas extremidades podem ser dirigidas a partir de um vértice ao outro [11, 47, 51].

Na figura 2.7, é demonstrado um exemplo de visualização em grafo de uma rede de padrões de co-edição da Wikipédia. Ele demonstra nos vértices, os idiomas, interligados por arestas direcionadas representando o sentido das edições. O seu tamanho representa o número de artigos correspondentes ao idioma em questão. Já as cores, representam a detecção, pelo algoritmo utilizado, de uma estrutura de comunidade semelhante.

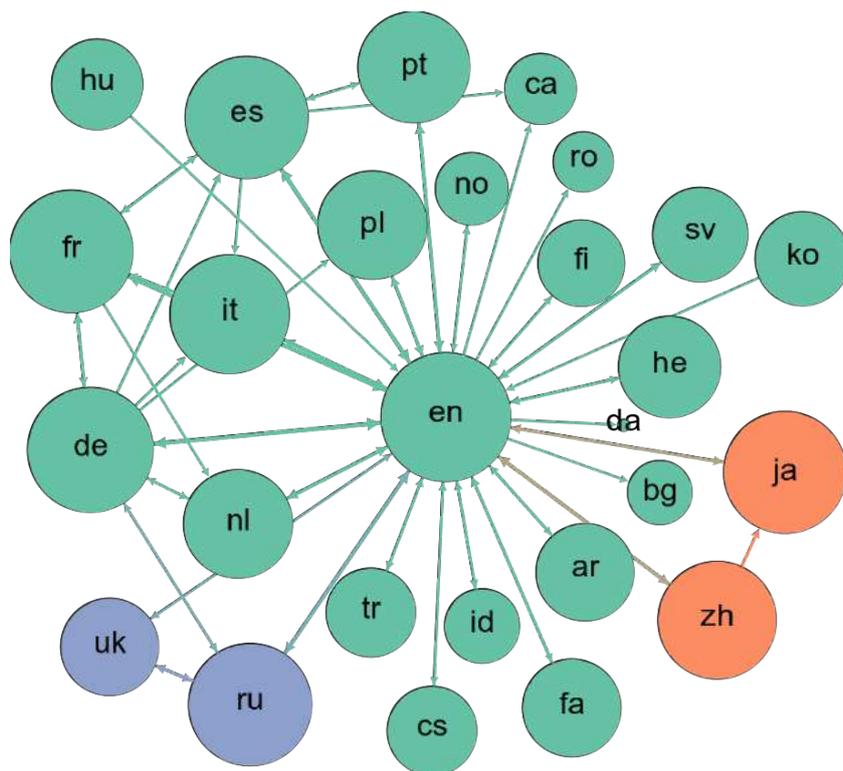


Figura 2.7: Rede de padrões de co-edição da Wikipédia demonstrada com a técnica de Grafo.

Fonte: Hale [47].

Esta técnica, segundo Herman et. al. [51] pode ser aplicada a conjuntos de dados, nos quais existe uma relação intrínseca entre os dados a serem visualizados. Desta

maneira, sua utilização se torna um tanto restrita a bancos de dados que contenham tal relação.

2.2.4 Mapas

Mapas tem sido utilizados desde os primórdios da visualização para demonstrar localizações específicas de objetos no espaço onde se vive. Os mapas são utilizados em diversos sistemas, desde analógicos até os digitais. Sua aplicação é muito extensa por se tratar de uma visualização simples, requerendo basicamente dados cartográficos, para sua localização, e mapeando o que se deseja visualizar neste mapa, bem como a forma desta visualização [18]. Porém, também podem ser utilizados para demonstrar dados multivariados, como representado na figura 2.8.

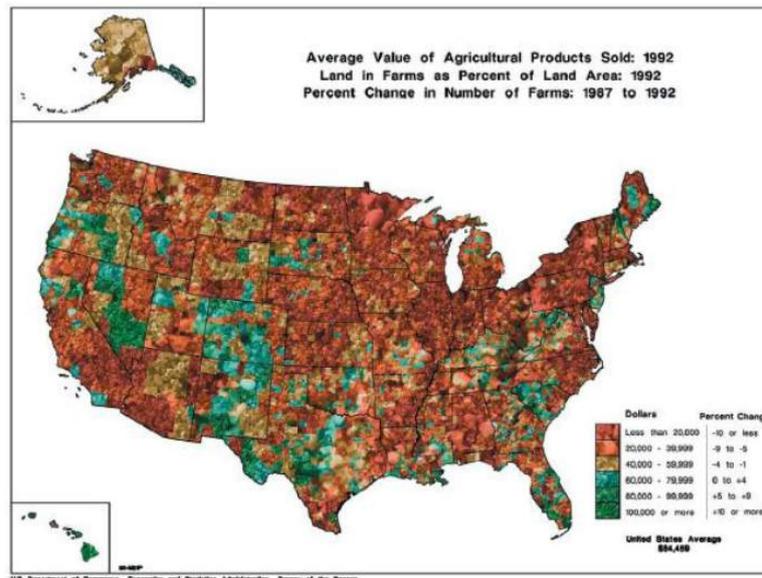


Figura 2.8: Exemplo de mapa representando dados variados sobre a agricultura dos Estados Unidos.

Fonte: Interrante [53].

Este mapa dos Estados Unidos, demonstra em com texturas sob cada localização, três variáveis de dados diferentes sobre a agricultura do país [53].

2.3 Análise Visual

Apesar de línguas faladas e escritas servirem a um propósito de transmitir informações entre as pessoas, o sistema visual humano é inigualável, funcionando tanto como um

processador de dados como quanto um reconhecedor de padrões. Desta maneira, surge a análise visual, juntamente com as técnicas gráficas clássicas, como um meio diversificado de transmitir informações entre as pessoas [5].

Nos dias atuais, é possível observar representações gráficas em diversas áreas do conhecimento, tais como na mecânica, arquitetura, área médica, mapas e área administrativa de empresas. Juntamente com a evolução da utilização da visualização, os computadores se tornaram pontos centrais para o desenvolvimento e utilização das visualizações. Em conjunto com o advento da internet, as visualizações passaram a ser em tempo real e vieram a ser utilizadas em larga escala em todo o mundo [63].

Apesar disto, não é fácil definir, em um sentido geral, análise visual, devido ao fato de sua natureza ser multidisciplinar. Esta é uma área que envolve múltiplos processos e uma grande variedade de áreas de aplicação [60].

Uma definição genérica descrita por Wong e Thomas [111], descreve análise visual como a formação de metáforas visuais abstratas, que ao serem combinadas com as técnicas de interação permitem a detecção e descoberta de dados esperados ou não, dentro de bases de dados dinâmicas e com grandes volumes de informação.

Já Wong e Thomas [111] a descrevem como: "a ciência do raciocínio analítico facilitada pelas interfaces interativas homem-máquina". Contudo, com base na prática atual, pode-se utilizar uma definição mais específica de Keim et al., na qual baseia-se este trabalho: "análise visual combina técnicas de análise automatizadas com visualizações interativas para uma compreensão eficaz, raciocínio e tomada de decisão com base em grandes e complexos conjuntos de dados"[60].

Com o fato desta definição estar cada vez mais presente no contexto dos profissionais da área de análise visual, a habilidade de transformar dados em novos conhecimentos é cada vez mais exigida destes profissionais. Competências para examinar dados dos mais variados e em diversas quantidades se apresentam como fatores de interesse para as empresas [6, 80]. Isto inclui desde pequenas bases de dados até *Big Data*, sem esquecer que estas bases contemplam os mais diversos formatos, como fluxos de informações de redes sociais e o cruzamento de fontes de dados variadas. Com destes dados se deve conseguir tomar decisões em um curto prazo de tempo.

Para o gerenciamento destes diversos tipos de informação a análise do ser humano é indispensável, combinando a flexibilidade, a criatividade e o profundo conhecimento do especialista do domínio em questão, com a enorme capacidade de armazenamento e o poder computacional dos computadores. Pode-se citar como sendo uma das principais vantagens da Análise Visual, o fato de que os usuários destes sistemas envolvem as suas capacidades cognitivas e perceptivas somente no processo analítico, permitindo-lhes aplicar o poder computacional para aperfeiçoá-lo.

Neste âmbito, de modo a conceitualizar a Análise Visual, alguns autores [4, 30, 59, 92] descrevem dois grupos que englobam-na: o processo de descoberta de conhecimento em bases de dados e o de visualização dos dados. Estes, combinam métodos de análise automática com a interação humana, de modo a adquirir conhecimento nos dados.

O processo de descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases - KDD*), descrito por Fayyad e Usama [38], enfatiza que o conhecimento é o produto final de um processo de descoberta guiado por dados [92]. No modelo proposto por Fayyad, Piatetsky-Shapiro e Smyth [38, 39], muito difundido na literatura, KDD é o processo global de identificar nos dados um padrão, uma estrutura válida, nova, potencialmente útil e interpretável [92, 30].

Já o processo de visualização de dados fortalece a descoberta do conhecimento com a incorporação de técnicas de visualização, como um instrumento para estimular a participação mais ativa do usuário na exploração e análise de dados. Com isso é oferecido a ele um suporte para monitoramento, avaliação e controle dos processos, elevando o grau de confiança nos resultados [4, 59].

Com estes dois processos interagindo, torna-se possível a integração das habilidades únicas da mente humana na percepção de padrões, exceções, tendências e relacionamentos. Desta forma a complementação mútua entre os recursos de visualização e procedimentos analíticos, desenvolvidos nos processos apresentados, descreve o processo de Análise Visual como um todo [92].

2.4 Interação e Análise Visual

Interações são necessárias para apoiar processos como comparar e classificar dados, extrair e recombina-los, criar/testar hipóteses e anotar dados. Desta maneira, autores como Thomas e Cook [99] reforçam o desenvolvimento desta ciência, cada vez mais necessária para o aperfeiçoamento dos processos desenvolvidos.

A "Ciência da Interação", deve possuir o foco no design da interação dos dados, amarrada no entendimento das diferentes formas de interação e seus benefícios, fornecendo prioridades iguais as da representação visual, complementando-as [99].

Sob este ponto de vista, percebe-se cada vez mais a imersão desta ciência em nosso cotidiano, dado que diversos dispositivos que utilizamos apresentam várias formas interações, como por exemplo: telas *touchscreen*; dispositivos multi-toques; interações de *zoom/panning*, seleções personalizadas, etc. Tais abordagens permitem um aumento no grau de interatividade possibilitando novas descobertas e usos das tecnologias.

Juntamente com estas formas de interação inseridas nos processos, apresenta-se a necessidade de pesquisar e de desenvolver abordagens diferenciadas para as diversas áreas do conhecimento.

De acordo com Thomas e Cook [99], a área de análise visual necessita de uma reformulação no que tange a apresentação das informações, devido a inevitáveis dúvidas que surgem conforme a análise mais detalhada dos dados é feita, acarretando na utilização de abordagens de interação para auxiliar neste processo. Para isto, considera-se que a análise visual é mais que um diálogo entre o analista e os dados dispostos a ele na visualização, a qual deve ser capaz de fornecer diferentes perspectivas e modelagens nos dados de forma interativa.

Assim como Thomas e Cook [99], diversos autores, entre eles, Heer e Shneiderman [50] e Keim e Kohlhammer [60], citam uma frase de Shneiderman [93], que descreve a ideia principal da análise visual, com o foco da interação: "*Overview first, zoom and filter, details on demand*". Ou seja, no desenvolvimento de uma nova ferramenta de análise visual, primeiramente deve-se ter uma a visão geral, após aplicar *zoom* e filtro, deixando por último e sobre demanda, detalhes mais específicos. Mas, assim como Thomas e Cook [99] questionam: o que vem a seguir, quais são as próximas etapas?

É possível iniciar uma reflexão sobre um novo conceito mais abrangente do que o de Shneiderman [93], com o auxílio de Keim et al.[63]. Efetuando uma releitura desta frase da seguinte forma: "Análise primeiro, mostre os dados importantes, permita *zoom* e filtro, analise se necessário, e exiba os detalhes sob demanda"[63].

Este conceito mais abrangente, possibilita explorar mais recursos sobre os métodos atuais de interação tornando, assim, mais atrativas as visualizações. Também se permite a descoberta de novos conhecimentos baseados nos dados já existentes, porém filtrados. Da mesma forma, age de forma mais completa sob as bases de dados atuais sem a perda de informações importantes [63].

Com estes conceitos, Card et al. [16] apresentam três sistemáticas de uso para a visualização, as quais permitem a interação para visualização da informação: modificar os dados (filtragem), modificar mapeamentos visuais e modificar visualizações. Adicionando, também um quarto uso de interação, o discurso de informação humana, originado por Thomas e Cook [99]. Estes usos são descritos de acordo com uma adaptação deste mesmo autor, descrita a seguir:

- Modificação para transformações dos dados, ou filtragem: Diversas técnicas se aplicam a este conceito, como manipulação direta e consultas dinâmicas;
- Modificação para mapeamentos visuais: Visualizações de fluxo de dados, assim como tabelas pivotantes, são dois exemplos de técnicas que permitem a interação do usuário para modificar os mapeamentos entre os dados e suas representações visuais;

- Modificação para visualização das informações, ou navegação: Técnicas de *zoom* e *panning* são exemplos de modificação de visualização das informações, e podem ocorrer de forma simples, em modelos 2D, como destaques ou seleções dos objetos de interesse, assim como mais avançadas em modelos 3D.

Pode-se concluir, com vistas a estes conceitos e usos que interações são necessárias para apoiar processos como comparar e classificar dados, extrair e recombina-los, criar e testar hipóteses, e anotar dados. Desta maneira autores como os citados nesta seção reforçam o desenvolvimento desta ciência, cada vez mais necessária para o aperfeiçoamento dos processos desenvolvidos.

2.5 Mineração de Dados

A mineração de dados é constantemente definida na literatura como a análise de grandes volumes de dados em busca de padrões consistentes [26, 48, 57, 109]. Para isso são utilizados algoritmos de aprendizado de máquina, como busca e análise, que acabam gerando novos conjuntos de dados processados.

Apesar das diversas descrições encontradas na literatura [48, 109, 114] sobre a definição da mineração de dados, é possível sintetiza-las na seguinte frase de Weiss e Indurkha [109]: "a mineração de dados consiste na busca de informações valiosas em grandes Bancos de Dados. É uma tarefa de busca e auxílio entre homens e sistemas. Os homens desenvolvem os sistemas, descrevem problemas e definem seus objetivos. Os computadores analisam dados e padrões que conciliem com as metas estabelecidas pelos homens."

Assim, relacionando este conceito a alguns exemplos, pode-se descrever como uma de suas aplicações, a utilização em empresas que necessitam analisar grandes volumes de informações armazenadas em suas bases de dados, sejam estes dados advindos de websites, aplicativos de celular ou de outras fontes. O objetivo é obter informações relevantes, como, por exemplo, análise de perfis de compra e satisfação de clientes relacionados aos sistemas ou produtos [26]. Outro exemplo consiste no uso de algoritmos genéticos de análise de dados, em sistemas hospitalares, para definir relações importantes para possíveis problemas relacionados a pacientes.

Estas informações muitas vezes estão distribuídas em vários locais diferentes, além de serem vários os formatos e tipos de dados. Por este motivo, a análise e sistematização destes dados de modo a criar padrões comuns são muitas vezes tão complexos que se tornam impossíveis, visto que a capacidade humana de interpretar tais dados é limitada [57]. É neste segmento que os algoritmos e os softwares especializados em mineração de dados se encontram.

Desta maneira, a mineração de dados tem sido considerada e classificada como a intersecção de diversas áreas do conhecimento, como a Estatística, a Inteligência Artificial e Banco de Dados. É possível identificar conteúdos de cada uma destas áreas dentro do processo principal de mineração e suas inter-relações [26, 57].

Na perspectiva da área de inteligência artificial, Witten e Frank [110] retratam-a como a “extração de informação implícita, previamente desconhecida, e potencialmente útil a partir de dados”. Para a estatística, de acordo com Han et al. [49], mineração de dados é “a análise de (frequentemente grande) conjuntos de dados para encontrar relacionamentos inesperados e para resumir os dados em novas formas que sejam compreensíveis e úteis para o seu proprietário”.

Os bancos de dados estão presentes em boa parte do processo de mineração, pois armazenam e gerenciam os dados essenciais. Desta maneira são eles que “alimentam” o sistema, fornecendo os materiais, ou as amostras, necessárias à pesquisa e ao desenvolvimento de toda a mineração.

Assim como os bancos de dados fornecem os materiais brutos a serem analisados, a inteligência artificial faz a maior parte do trabalho. Utilizando diversas técnicas, todas as amostras e materiais vindas das bases são processados e refinados para implementar o processo de mineração, chegando ao processo final esperado pelo usuário. Para tal, a inteligência artificial lança mão de vários algoritmos computacionais dentro dos quais pode-se citar, por exemplo, os de classificação [48], de agrupamento [25, 35] e de regressão [48].

Como pode-se observar nos itens a seguir, ocorre uma divergência entre os autores sobre a definição e os processos que envolvem a mineração de dados e sua utilização. Percebe-se que são tratadas muito mais por sua área de atuação do que pelo próprio tema:

- Adriaans e Zantinge [1] descrevem que a mineração de dados é realizada por descoberta de conhecimento e predição;
- Berry e Linoff [10], definem por classificação, estimação, predição, afinidade em grupos, agrupamentos (*clustering*) e descrição;
- Thuraisingham [100] define por classificação, detecção de sequência, análise de dependência de dados e análise de desvio;
- Elmasri [37], define por previsão, identificação, classificação e otimização;
- Han e Kamber [48] define por descrição e predição.

Considerando as opiniões dos autores citados, neste trabalho se assume que o processo de mineração de dados e a descoberta de conhecimento perpassam as etapas de classificação, agrupamento e predição.

Porém, definir bem quais são os dados disponíveis e a quais dados se quer chegar, é essencial e ponto de partida para o processo de mineração, assim como definir seu objetivo e se há a possibilidade de obtenção dos dados.

É possível sintetizar que, durante a identificação dos problemas, é relevante: definir as metas do processo; estabelecer os critérios de desempenho; estabelecer a relação entre simplicidade e precisão do conhecimento extraído; e definir se um conhecimento resultante deve ser compreensível a seres humanos ou se um modelo do tipo caixa-preta é suficiente. Esses pontos interferem nas escolhas posteriores e nas diversas etapas do processo de descoberta de conhecimento [79].

Para estas etapas, a utilização de algoritmos que auxiliam na descoberta do conhecimento muitas vezes é necessária. Dentre os mais utilizados podem-se citar os seguintes: Classificação [48], Predição [48], Árvores de Decisão [26, 48], Algoritmo J4.8 [25], Classificação Bayesiana [48], Apriori [25, 89], Agrupamento [25, 35] e Genéticos [26, 48].

2.5.1 Algoritmos de Mineração

Para uma melhor compreensão deste trabalho, a seguir são descritos alguns dos algoritmos de mineração de dados utilizados e/ou testados para o desenvolvimento desta pesquisa.

- K-Means: método de quantização vetorial, originalmente de processamento de sinal, que é popular para a análise de grupos em mineração de dados. O agrupamento por K-Means particiona 'n' registros, em 'k' grupos, sendo que cada registro pertence ao grupo mais próximo da média [35, 48].
- DBSCAN: O algoritmo "aglomeração espacial baseada em densidade de aplicações com ruído" (*Density-based spatial clustering of applications with noise*, DBSCAN) consiste em uma abordagem de agrupamento de dados baseada em densidade porque ele encontra um número de grupos a partir da distribuição da densidade estimada dos pontos correspondentes [48].

Basicamente, requer apenas dois parâmetros: a distância máxima entre duas amostras para serem considerados na mesma vizinhança, e o número mínimo de pontos necessários para formar uma região densa. Deste modo, se inicia com um ponto de partida arbitrário que não tenha sido visitado, os pontos vizinhos da distância máxima deste são recuperados e, se ele contém um número suficiente de pontos, um grupo é iniciado, caso contrário, o ponto é rotulado como ruído. Este fato porém, não o desconsidera, de modo que ele poderá ser novamente visitado se encontrado em outro grupo de tamanho suficiente.

- **Árvore de Decisão:** consiste em um algoritmo que analogamente se assemelha a um desenho de estrutura de uma árvore, no qual cada nó interno, não folha, denota o teste de um atributo, cada galho representa um lado do teste, e, por sua vez, cada nó terminal (ou nó folha), identifica uma classe, assim como ilustrado no exemplo da figura 2.9 [48].

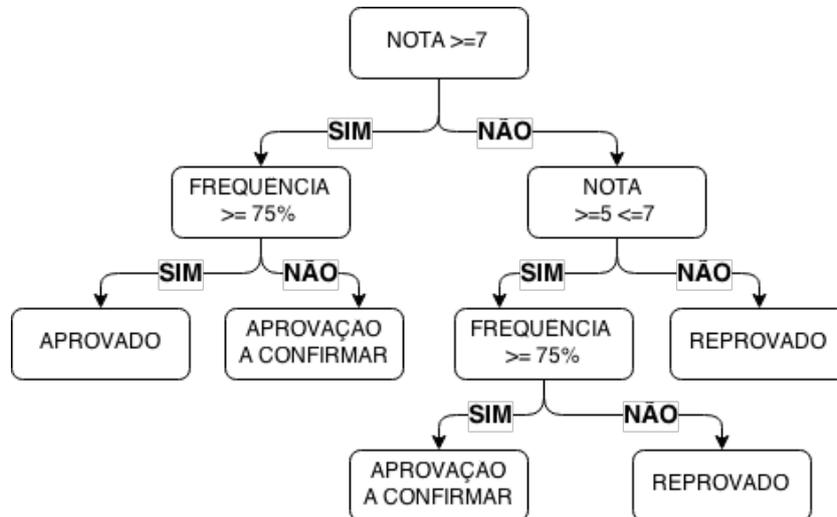


Figura 2.9: Exemplo de Árvore de Decisão sobre a aprovação de um aluno utilizando a nota e a frequência como critérios.

Fonte: Autoria própria, 2016.

As árvores de decisão são denotadas pela literatura [48, 26] como as mais utilizadas. Sua implementação e validação é de certa maneira simples e também fácil de ser entendida, quando demonstrada em visualizações, como na figura 2.9. Além disso, em geral apresentam um bom nível de acurácia. São muito utilizadas em diversas áreas do conhecimento, como medicina, manufatura, produção e finanças.

Sua utilização necessita ser acompanhada de um treinamento, no qual um *dataset* pré-classificado deve ser informado como entrada ao algoritmo, de modo a treiná-lo sobre os padrões que encontrará para poder classifica-los.

- **Classificação Bayesiana:** funciona como um classificador estatístico, baseado nas regras de Bayes. Prevê a classificação de acordo com a probabilidade da tupla pertencer a uma classificação em particular [103]. Diversos estudos efetuados demonstram que classificadores Bayesianos como Naïve Bayes, tem performance comparável as árvores de decisão e classificadores desenvolvidos com redes neurais, sem deixar de lado sua alta acurácia e velocidade, mesmo lidando com grandes bases de dados [103, 48].

Naïve Bayes consiste em um modelo probabilístico simples com base na regra de Bayes com seleção de recursos independentes. Assim, corresponde a uma maneira

de olhar para as probabilidades condicionais que permite que se alterne em torno da condição de uma forma conveniente. A condicional é um evento X que provavelmente irá ocorrer, dada a evidência Y . Isso é normalmente escrito por $P(X | Y)$.

A regra de Bayes permite determinar essa probabilidade, quando tudo o que se tem é a probabilidade de o resultado oposto e dos dois componentes individualmente: $P(X | Y) = P(X) P(Y | X) / P(Y)$. Essa correção pode ser muito útil quando se está tentando estimar a probabilidade de algo baseado em exemplos de sua ocorrência [103].

- TF-IDF: O algoritmo de frequência de termos (*Term Frequency–Inverse Document Frequency*, TF-IDF), gera uma eficaz representação de documentos na forma vetorial, expressando o seguinte raciocínio: quanto maior a frequência de um termo em um documento, mais representativo ele é para o conteúdo, e quanto mais documentos contiverem um termo, menos discriminante ele é para o conteúdo [87]. Assim, atribuindo-se um peso $w_{t,d}$ para um termo t em um documento d , obtém-se as equações a seguir (SALTON e BUCKLEY [87]).

$$w_{t,d} = tf_{t,d} * idf_t \quad (2.1)$$

$$idf_t = \log \frac{N}{df_t} \quad (2.2)$$

Onde, $tf_{t,d}$ é a quantidade de vezes que o termo t aparece no documento d , sendo N a quantidade total de documentos que compõem o conjunto de dados e df_t é a quantidade de documentos que contém o termo t .

Desta forma, o algoritmo TF-IDF gera uma medida estatística utilizada para avaliar o quanto uma palavra é importante para um documento em relação a uma coleção (corpus). Essa importância aumenta proporcionalmente com o número de vezes em que a palavra aparece no documento, e diminui de acordo com a frequência da palavra na coleção. Da mesma maneira, com todas as palavras de um determinado documento, pode-se ter a similaridade entre este e demais a serem testados entre si [82].

2.6 Descoberta de Conhecimento

O processo de descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases - KDD*), foi estabelecido no primeiro workshop da área em 1989 por Fayyad e Usama [38], no qual se enfatizou que o conhecimento (*knowledge*) é o produto final de um processo de descoberta guiado por dados [92]. No modelo proposto por Fayyad, Piatetsky-Shapiro e Smyth [38, 39], referenciado em diversos artigos, KDD é o processo

global de identificar nos dados um padrão, uma estrutura válida, nova, potencialmente útil e interpretável [92, 30].

Desta forma, na figura 2.10, é apresentado um exemplo difundido na literatura [92, 48, 59], sobre os processos envolvidos para o desenvolvimento da descoberta de conhecimento, da mesma forma como a descrição sobre cada processo:

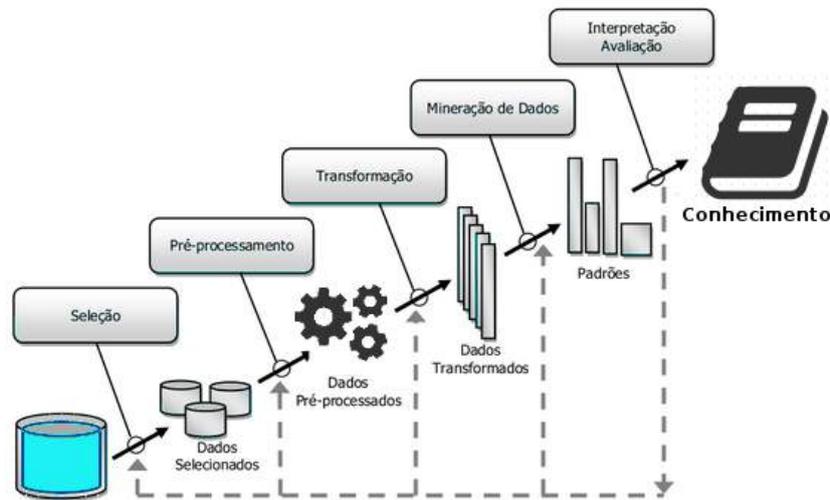


Figura 2.10: Processos de Descoberta de Conhecimento.

Fonte: adaptada de Fayyad et al. [39]

1. Seleção: Consiste na seleção dos atributos de interesse do usuário.
2. Pré-Processamento: Após a seleção dos dados, prepara-os de forma a tornar valores padronizados, verificar, corrigir, adicionar e/ou retirar inconsistências, assim como, demais processos que sejam necessários. Esta etapa é vital em KDD, pois dados faltantes podem resultar em problemas de classificação. Para isto podem ser utilizados algoritmos para prever valores faltantes ou simplesmente remover este registro.
3. Transformação: Com os dados pré-processados torna-se necessário modelar a sua estrutura de modo a possibilitar a adequação de acordo com o especificado pelos algoritmos de mineração. Desta forma, nesta etapa os dados são transformados para os formatos adequados, como CSV (*Comma-Separated Values*), ARFF (*Attribute-Relation File Format*), em outros.
4. Mineração de Dados: Na etapa de Mineração de Dados são aplicados algoritmos de inteligência artificial e aprendizado de máquina para extrair padrões de dados ou gerar novas regras que descrevem o comportamento da base de dados.
5. Interpretação e Avaliação: Esta etapa consiste em avaliar o conhecimento extraído da base de dados, identificando padrões e interpretando-os, auxiliando os tomadores de

decisão. O objetivo é filtrar as informações que serão apresentadas a estes usuários, de forma que se os resultados não forem satisfatórios, é então executada novamente a etapa de Mineração de Dados.

É importante destacar que a etapa de avaliação tem seu processo auxiliado pelos usuários que tem o conhecimento do domínio da aplicação, para que os mesmos visualizem os dados obtidos pelos algoritmos e verifiquem a sua veracidade, pois é possível que os algoritmos de mineração tenham resultado em informações que não condizem com a realidade analisada.

Já para o processo posterior à descoberta de conhecimento, encontra-se a visualização, à qual fortalece a descoberta incorporando diversas abordagens da análise visual, e, como isto, torna-se um instrumento para estimular a participação mais ativa do usuário na exploração e análise dos dados. Assim, oferecendo um suporte para monitoramento, avaliação e controle dos processos, elevando o grau de confiança nos resultados [4, 59].

Com estas abordagens juntas, torna-se possível a integração das habilidades da mente humana na percepção de padrões, exceções, tendências e relacionamentos. Desta forma a complementação mútua entre os recursos de visualização e procedimentos analíticos, desenvolvidos nos processos descritos, demonstra o processo de descoberta de conhecimento e análise visual como um todo [92].

3. TRABALHOS RELACIONADOS

Da mesma forma como no ambiente empresarial, a utilização de aplicações de análise visual se mostram úteis aos ambientes de EAD, possibilitando a análise dos dados gerados pelos ambientes de aprendizagem para acompanhar o desempenho dos alunos, pois grandes volumes de dados dificultam análises manuais [62]. Além disso, na medida em que os cursos à distância se popularizam, e com o aumento da competitividade, estas análises ganham ainda mais importância, possibilitando que as organizações consigam desenvolver um trabalho mais adequado ao perfil de cada aluno, bem como acompanhar o seu desenvolvimento no curso [76].

Com isso, técnicas de análise visual desenvolvem um papel importante para as instituições de ensino, pois podem auxiliar nas tomadas de decisão, norteando-as sob a situação de seus cursos e, em muitos casos, prevenindo eventuais evasões e reprovações.

Sendo assim, um meio de auxiliar a área de educação a distância consiste na utilização de técnicas de análise visual em conjunto com algoritmos de mineração de dados [7]. A partir dos dados das interações realizadas pelos alunos nos ambientes de aprendizagem, é possível identificar diferentes perfis de alunos, por exemplo, quais possuem uma tendência à evasão, reprovação ou aprovação. Este tipo de predição é possível a partir da aplicação de algoritmos de mineração para identificar padrões de comportamento [58]. A visualização destas informações pode auxiliar no acompanhamento de um grande número de alunos e na análise dos cursos, possibilitando identificar o valor agregado pelos recursos utilizados e a tomada de decisões.

Através do estudo da literatura, pode-se perceber que, alguns autores utilizam métodos de classificação semelhantes para obter seus resultados, sendo que eles também buscam a generalização de suas abordagens. Amershi e Conati [2], por exemplo, utilizam estes métodos e também de clusterização, tal como K-Means, para efetuar suas predições. Outros autores, como Bravo e Ortigosa [15], os quais descrevem um estudo sobre a qualidade dos sistemas que são oferecidos a estudantes, também utilizam estes mesmos algoritmos de mineração de dados.

A figura 3.1 ilustra o resultado da utilização de uma abordagem de mineração de dados, em uma pesquisa realizada por Chiritoiu et al. [19]. Juntamente com técnicas de clusterização, foi desenvolvida uma visualização de modelo de mapa mental que permite algumas interações. Para a construção desta visualização, foram utilizados dados de atividades desenvolvidas pelos alunos em um curso regular de graduação. Deste modo, na figura 3.1, cada estrela representa um usuário, a sua cor (vermelho, verde ou amarelo) representa um agrupamento e o círculo da mesma cor representa o centro deste agrupamento. As cores de fundo ilustram os agrupamentos mais específicos. Os eixos x e y representam os dados selecionados pelo usuário, que neste exemplo são as atividades de

troca de mensagens e teste de atividade, respectivamente. Os valores que aparecem no eixo X são resultantes de uma fórmula que utiliza o número de mensagens e a média de caracteres de cada mensagem, e no eixo Y os valores também são resultantes de um cálculo composto pelo número de tarefas e a média dos resultados de cada tarefa.

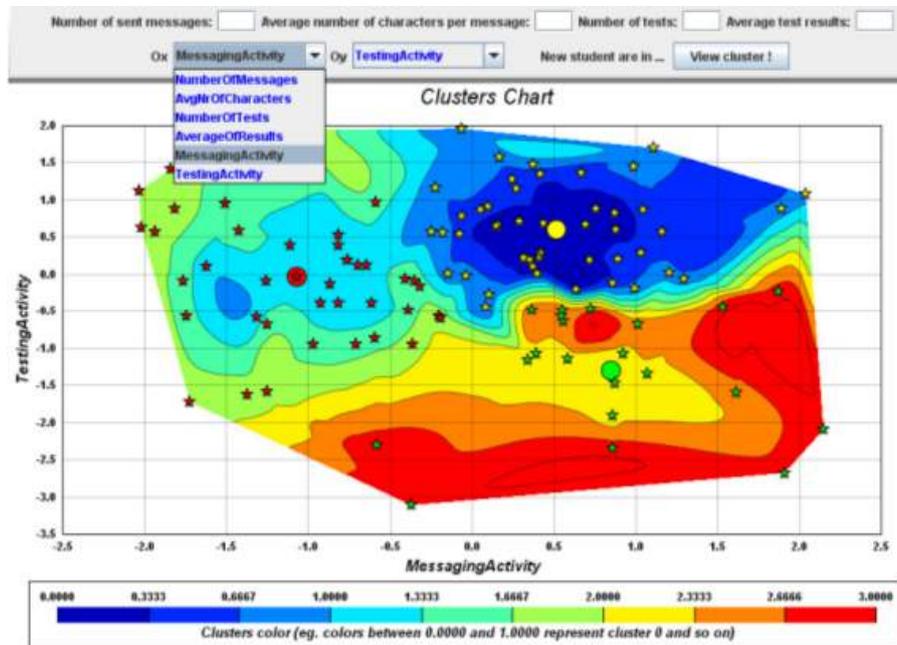


Figura 3.1: Exemplo de interface que apresenta uma análise de trocas de mensagens e teste de atividade (representadas nos eixos x e y). As cores representam os agrupamentos, os círculos o centro do grupo, e cada estrela um usuário.

Fonte: Chirioiu et al. [19].

Aliado à utilização destas abordagens de mineração, autores como Garcia-Saiz e Pantaleon [42] e Graf et al. [44] descrevem o desenvolvimento de um conjunto de visualizações integradas a ambientes virtuais de aprendizagem. Estes autores usaram sistemas de mineração com técnicas de visualização para simplificar a busca por informações relevantes. A figura 3.2, elaborada por Graf et al. [44], apresenta uma das interfaces utilizadas em sua abordagem, que tem por objetivo auxiliar designers de aprendizagem no desenvolvimento de conteúdo para cursos EAD.

Seguindo esta mesma linha, Mazza e Dimitrova [69, 68] desenvolveram uma abordagem com a utilização de algoritmos de classificação e técnicas de visualização. Através dela, os autores procuram reduzir a curva de aprendizagem de seus usuários e propiciar um melhor entendimento das informações apresentadas. Exemplos das visualizações fornecidas são apresentados nas figuras 3.3, 3.4 e 3.5, sendo que, neste último exemplo, após um novo estudo [70], os autores apresentaram a sua integração na forma de um *plugin* para o AVA Moodle.

Da mesma forma como os autores descritos até aqui, outras abordagens estudadas buscaram a utilização de mais de um algoritmo de mineração de forma conjunta para

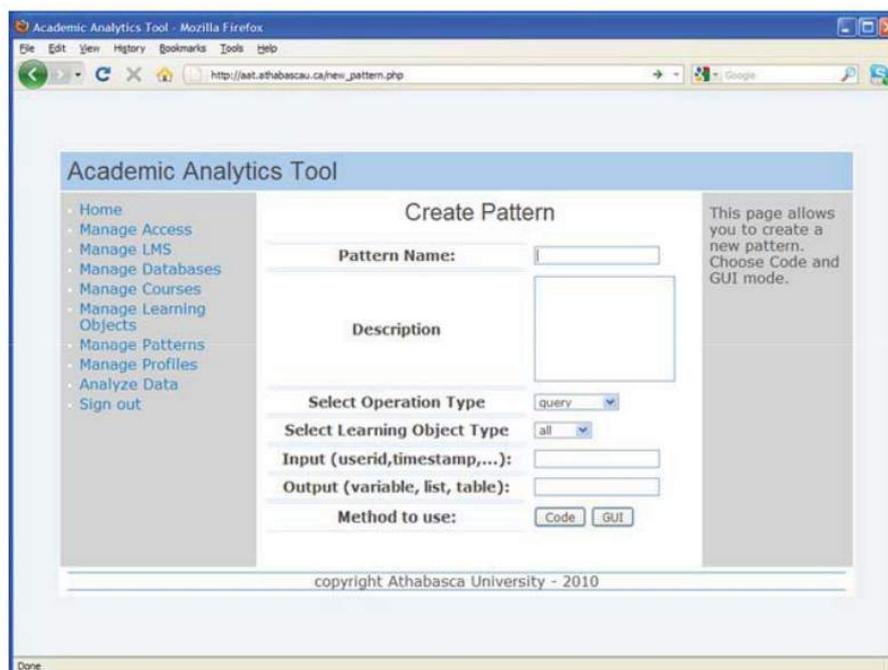


Figura 3.2: Interface da aplicação AAT para realização do cadastro de um padrão.

Fonte: Graf et al. [44].

auxiliar na tarefa de descoberta de conhecimento. Baruque et al. [7] e Romero et al. [86], por exemplo, utilizam redes neurais em conjunto com técnicas de clusterização para efetuar a mineração dos dados. Suas pesquisas demonstraram que a utilização destas técnicas de mineração em conjunto, resultaram em ganhos de performance, e possibilitaram um estudo mais completo da interação destes algoritmos.

Duas imagens que representam a abordagem desenvolvida por Romero et al. [86] são demonstradas nas figuras 3.6 e 3.7. Elas ilustram os dois principais componentes de interação para a mineração de dados. O primeiro componente (figura 3.6), demonstra visualizações possíveis com os dados disponibilizados no ambiente, onde, o usuário pode selecionar um ou vários alunos, o tipo estatística (notas, *logs*, envios, fóruns, etc), o período de avaliação, e, após a captação dos dados pelo algoritmo, o tipo de gráfico a ser visualizado. Já o segundo componente (figura 3.7), auxilia no processo de mineração de maneira mais detalhada, pois possibilita o usuário selecionar o tipo de algoritmo a ser aplicado (agrupamento, classificação, regressão, etc), os atributos utilizados para ele, e então visualizar o resultado em tela, com a possibilidade de efetuar o *download* do mesmo em modo texto ou PDF.

Da mesma forma que os demais, os trabalhos de Bovo et al. [13] e Pedraza-Perez et al. [75] também utilizam clusterização, porém, buscam a técnica de regressão como mais um recurso de mineração em suas pesquisas. Além disso, diferem-se no que tange a utilização de ferramentas gráficas para auxílio de seus usuários. O trabalho de Bovo et al. não inclui técnicas de visualização e necessita que o usuário seja conhecedor da área de

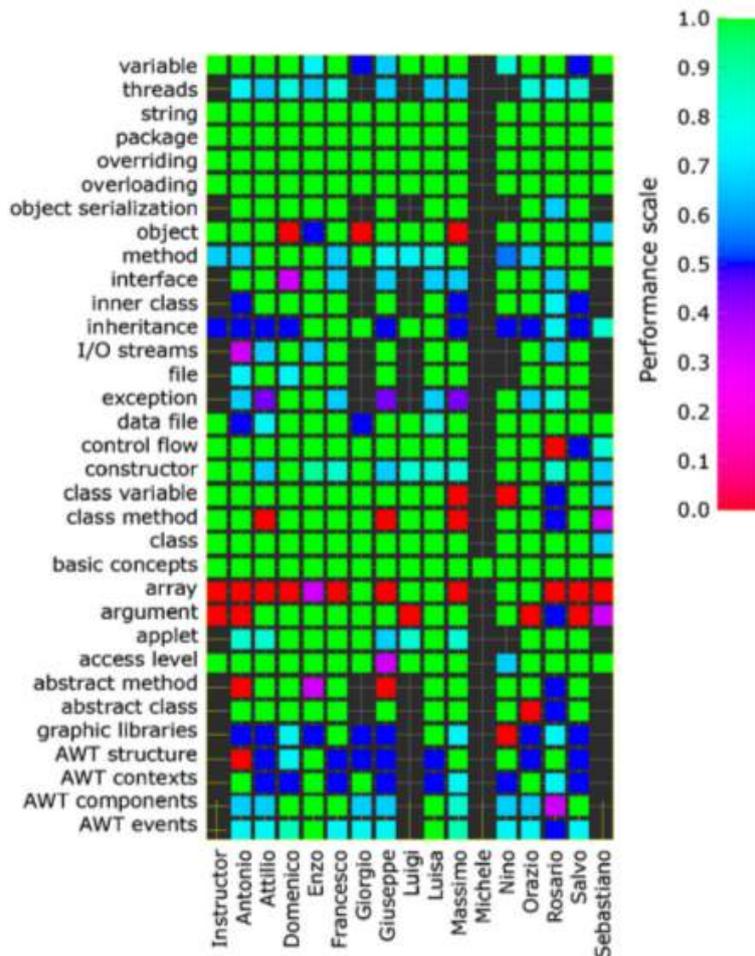


Figura 3.3: Matriz cognitiva do desempenho dos estudantes baseado em um *quiz* relacionado com conceitos do domínio estudado.

Fonte: Mazza e Dimitrova [69].

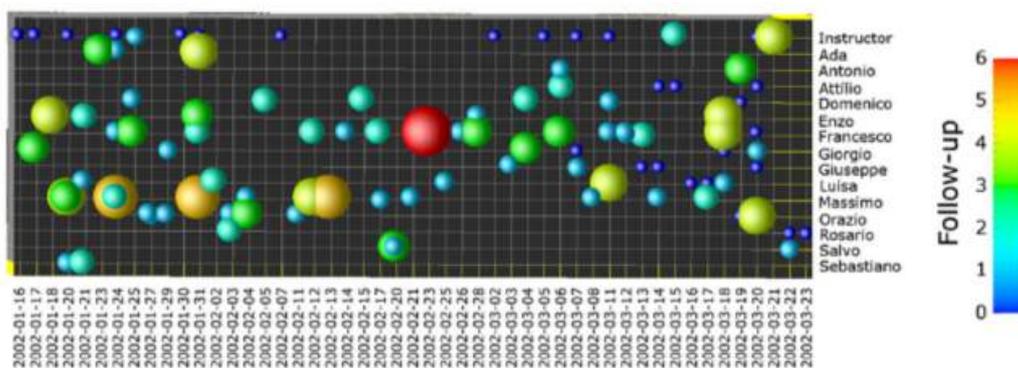


Figura 3.4: Visualização de tópicos de discussão de um fórum.

Fonte: Mazza e Dimitrova [69].

mineração e programação para que possa utilizar e analisar os resultados obtidos. Já o trabalho de Pedraza-Perez et al. contém uma implementação conjunta de um assistente para

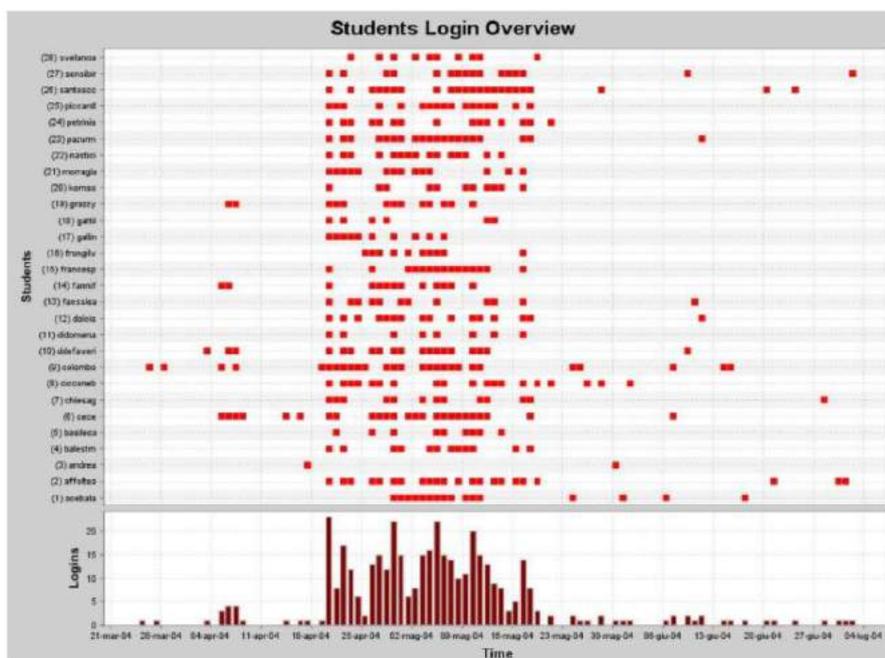


Figura 3.5: Apresentação dos acessos dos estudantes ao curso.

Fonte: Mazza e Milani [70].

o processo de mineração de dados e descoberta de conhecimento, buscando possibilitar a transparência e o entendimento das etapas e das operações realizadas.

A classificação através de *clustering* também é abordada por outros autores [65, 85] que utilizaram dados de fóruns de um curso universitário para prever as notas finais do mesmo. Mas, estes autores optaram por disponibilizar seus resultados em formato de tabela e não desenvolveram gráficos.

Além dos trabalhos já citados, durante as pesquisas efetuadas na literatura, foram encontradas abordagens mais específicas, que buscam aplicar seus métodos e técnicas em AVA's desenvolvidos pelas próprias instituições, ou em outras plataformas de ensino comerciais. Dentre estes, pode-se perceber o uso de abordagens de mineração distintas das descritas até o momento.

A Utilização de técnicas de predição baseadas em árvore de decisão, descrita por Cocea e Weibelzahl [21], Golding e Donaldson [43] e Dekker et al. [34], é apresentada como de grande importância para a melhoria na busca da descoberta de conhecimento em ambientes educação a distância. Segundo os autores, a sua utilização como preditora de alguns índices, tais como, evasão e reprovação, a torna deveras relevante para este contexto. Assim como, a utilização de diversos campos do banco de dados no processo de mineração, dado o fato que, foram unidos diferentes ambientes e *datasets* em um mesmo local para este processo.

Já DeBoer et al. [33], descrevem a utilização de dados coletados de diferentes plataformas, de modo a constituir um conjunto maior para a etapa de mineração, na qual

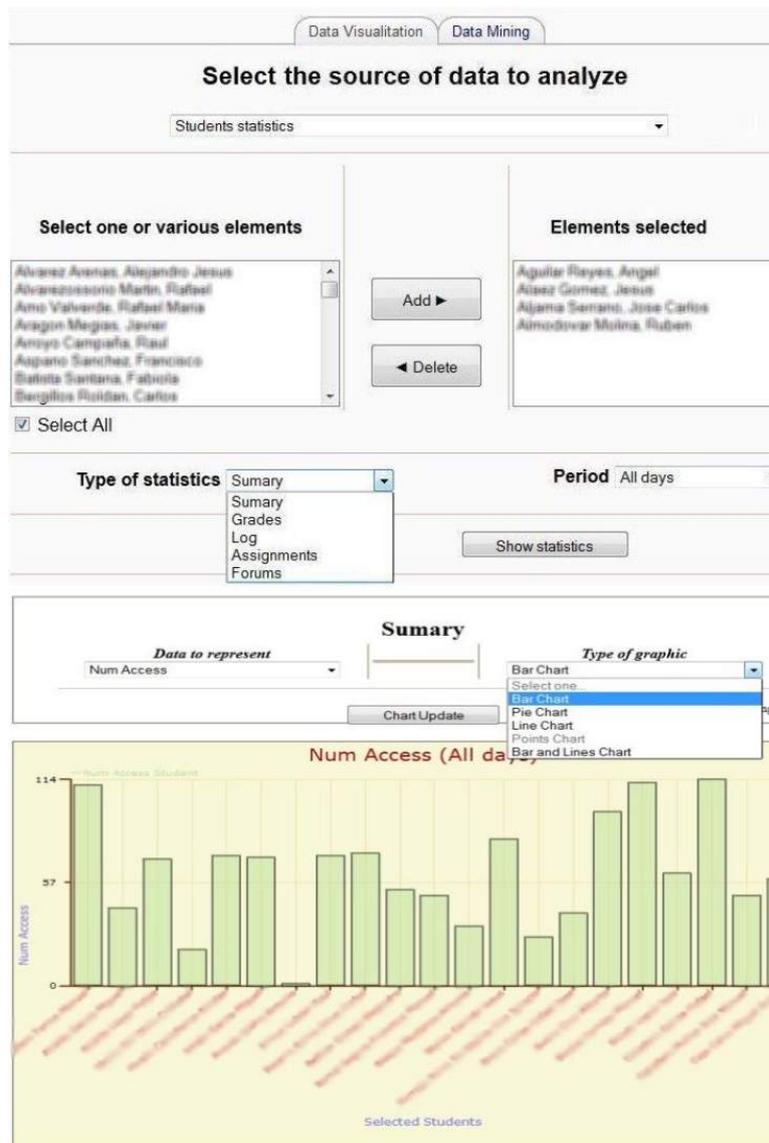


Figura 3.6: Primeiro componente da abordagem desenvolvida por Romero et al., que possibilita a seleção de um conjunto de dados e atributos dos alunos, advindos do AVA Moodle, para posterior geração de gráficos estatísticos.

Fonte: Romero et al. [86].

o método de regressão foi aplicado. Zhong-mei et al. [113], também apresentam a utilização de regressão para a tarefa de mineração, porém, combinada com algoritmos de redes neurais, possibilitando, segundo os autores, uma melhor exploração do conjunto de dados utilizado.

Assim como as abordagens descritas anteriormente apresentam estudos sobre mineração em AVA's de modo a prever notas e evasões, as demais pesquisas oferecem resultados para a análise dos alunos, observando os textos publicados nos fóruns destes ambientes. Também são encontrados na literatura alguns métodos para efetuar estas análises. Ribeiro [82] descreve a utilização do algoritmo de medida estatística TF-IDF (descrito

The screenshot displays a web-based interface for data mining. At the top, there are tabs for 'Data Visualisation' and 'Data Mining'. Below this, a section titled 'Select a exported data file' shows a file path 'C:\Users\aaaa\Desktop' with 'Examinar...' and 'Send' buttons. The next section, 'Select a Data Mining Task', has a dropdown menu with 'Grouping / Clustering' selected. Below this, there are two columns: 'Select one or more attributes' and 'Selected attributes'. The 'Selected attributes' column contains: 'Acceso a los Recursos', 'Num Actividades Realizadas', 'Num Cuestionarios Realizados', 'Num de Accesos < 30 seg', and 'Num de Accesos'. The 'Clustering algorithm parameters' section shows 'Number of clusters' set to 3. Below this, a list of clusters is shown with their respective centroids and sizes. The 'CENTROIDS' section lists three centroid coordinates. At the bottom, a 'Save results' section has a dropdown menu with 'PDF' selected and an 'Export' button.

Figura 3.7: Segundo componente da abordagem desenvolvida por Romero et al., que permite ao usuário selecionar um arquivo e atributos, para aplicação de um algoritmo de mineração. Os resultados podem ser visualizados ou baixados em formato texto ou PDF.

Fonte: Romero et al. [86].

no subcapítulo 2.5.1) para recomendar documentos aos alunos, baseando-se em *hashtags* utilizadas por eles nos fóruns das disciplinas estudadas.

Com o amplo uso dos fóruns muitas mensagens são trocadas nos AVA's, e isso pode acarretar em problemas no monitoramento pelos professores e tutores. Pensando nisto, Júnior et al. [32] mostram em seus estudos a utilização do algoritmo TF-IDF com o propósito de monitoramento destas mensagens de forma de classificá-las em positivas e negativas, permitindo um filtro e assim, uma maior atenção as possíveis mensagens negativas.

3.1 Análise dos Trabalhos

A partir do estudo dos trabalhos relacionados e considerando o objetivo desta pesquisa, foi feita uma análise dos algoritmos e técnicas que as abordagens descritas anteriormente contemplam, bem como foi possível observar algumas questões que não são contempladas.

Sob este ponto de vista, as abordagens foram estudadas para elucidar o entendimento dos métodos propostos e suas aplicações. Algumas delas são utilizadas com o AVA Moodle como fonte de dados, outras são aplicadas a ambientes proprietários, não descritos, e outras ainda são mais genéricas, procurando permitir seu re-uso em outros ambientes. Considerando as técnicas de visualização fornecidas e os algoritmos de mineração de dados utilizados, nota-se que a visualização é pouco empregada entre as abordagens analisadas. Isto deve-se ao fato de que os dados utilizados possivelmente não continham um grande volume que demandasse o uso de representações visuais mais elaboradas para suas abordagens, ou este não era seu foco.

Também percebeu-se a não definição do uso específico de um algoritmo de mineração de dados para os ambientes educacionais, resultando na incerteza da melhor abordagem a ser utilizada. Contudo, alguns trabalhos relatam que a utilização de mais de um algoritmo resulta em aumento significativo de acurácia.

Desta forma, a Tabela 3.1 apresenta sucintamente os algoritmos abordados pelos autores, possibilitando uma visualização mais abrangente com relação a aplicação dos algoritmos de mineração. É possível observar, nesta tabela, que os três algoritmos de mineração mais utilizados são: clusterização, classificação e *k-means*.

Tabela 3.1: Algoritmos de mineração utilizados nos ambientes educacionais

	Clusterização	Classificação	C.4.5	Regressão	Redes Neurais	Árvore de Decisão	K-Means	Associação
Amershi e Conati [2]	X	X					X	
Baruque et al. [7]	X				X			X
Bovo et al. [13]	X			X			X	
Bravo e Ortigosa [15]			X					
Chiritoiu et al. [19]	X						X	
Cocea e Weibelzahl [21]						X		
DeBoer et al. [33]				X				
Dekker et al. [34]						X		
García-Saiz e Pantaleon [42]	X						X	
Graf et al. [44]	X	X						X
López et al. [65]	X							
Mazza e Dimitrova [69, 68]	X	X						
Pedraza-Perez et al. [75]	X	X	X	X		X		X
Romero et al. [85]	X	X	X				X	X
Romero et al. [86]		X			X	X		
Zhong-mei et al. [113]				X	X			

Contudo, observa-se que o desenvolvimento destas abordagens é restrito ao aperfeiçoamento da acurácia dos algoritmos, algo relevante. Porém, não consideram com a

mesma importância o ambiente desenvolvido no que tange a sua interface e forma de interação, conforme pode-se observar na Tabela 3.2.

Tabela 3.2: Lista de trabalhos relacionados considerando as visualizações disponibilizadas e a possibilidade de interação

	Sem Gráficos	Interatividade	Linha	Barra	Mapa de Calor
Amershi e Conati [2]		X	X		x
Baruque et al. [7]	X				
Bovo et al. [13]	X				
Bravo e Ortigosa [15]	X				
Chirioiu et al. [19]		X			X
Cocea e Weibelzahl [21]	X				
DeBoer et al. [33]	X				
Dekker et al. [34]	X				
Garcia-Saiz e Pantaleon [42]	X				
Graf et al. [44]		X			
Johnson e Barnes [55, 54]		X			
Lopez et at. [65]	X				
Mazza e Dimitrova [69, 68]		X	X	X	
Pedraza-Perez et al. [75]	X				
Romero et al. [85]	X				
Romero et al. [86]	X				
Zhong-mei et al. [113]				X	

Considerando as Tabelas 3.1 e 3.2 fica claro que diversas abordagens utilizaram os algoritmos de agrupamento, mas poucas destas abordagens possuíam recursos visuais. Entretanto, pode-se considerar que, caso elas utilizassem recursos visuais e interativos, disponibilizariam um ganho para seus usuários, que poderiam, por exemplo, descobrir novos conhecimentos com os mesmos dados através da manipulação destes recursos.

A partir destas análises, surgem as seguintes hipóteses:

- É possível facilitar a descoberta de conhecimento por parte do usuário através do uso de algoritmos de mineração associados a ferramentas de visualização;
- A análise visual de diferentes interações dos recursos disponibilizados ao usuário no AVA, pode facilitar a predição do desempenho do aluno;
- A utilização de uma abordagem de acompanhamento e predição, pode contribuir no resgate de alunos e no planejamento de novas disciplinas.

Estas hipóteses serviram como questões norteadoras para o desenvolvimento da abordagem de visualização descrita neste trabalho, a qual é apoiada por um processo de

mineração de dados com algoritmos de agrupamento. Também, procurou-se analisar o ponto de vista do usuário, de modo a identificar as questões que poderiam ser relevantes a ele e a construção de visualizações interativas, de modo a facilitar seu uso. Através de entrevistas com professores e administradores de uma Instituição de Ensino Superior (IES), foi realizado um levantamento de requisitos, descrito no capítulo 4, para auxiliar neste fim. Logo após, foi realizado o desenvolvimento da abordagem proposta, juntamente com as questões levantadas pelos usuários entrevistados, descrita no capítulo 5.

4. LEVANTAMENTO DE REQUISITOS

Durante o desenvolvimento deste trabalho foram consideradas as opiniões de diversos professores, tutores e administradores que atuam na área de educação a distância há pelo menos um ano e que trabalham em diversas universidades e faculdades, particulares e federais. Para isto, houveram duas etapas de levantamento de opiniões. A primeira, teve como objetivo analisar tanto a eficácia da abordagem de visualização que seria desenvolvida, como as funcionalidades desta proposta e as necessidades dos futuros usuários. A segunda foi aplicada com o intuito de avaliar a abordagem já implementada, testando-a durante o andamento de uma disciplina, conforme descrito no capítulo 6. Este capítulo apresenta a descrição do processo aplicado na primeira etapa, quando foi feito um levantamento de requisitos, juntamente com suas respectivas análises. Também é apresentada uma análise do perfil dos entrevistados (especialistas de domínio), suas experiências, opiniões sobre as ferramentas que já utilizaram ou utilizam e sobre a abordagem proposta.

4.1 Descrição do Processo

O levantamento de requisitos foi realizado através da aplicação de um questionário, seguido de uma entrevista. Para o questionário foram elaboradas 17 perguntas, sendo 9 de única escolha, 5 de múltipla escolha e 3 dissertativas, para coleta de opinião. Todos os participantes assinaram o termo de consentimento livre e esclarecido e após responderam ao questionário. Ambos documentos foram anexados nos apêndices A e B, respectivamente.

Para realizar o questionário, foram elaboradas 17 perguntas, sendo 9 de única escolha, 5 de múltipla escolha e 3 dissertativas, pelas quais o entrevistado poderia manifestar sua opinião. Todos os participantes assinaram o termo de consentimento livre e esclarecido e após responderam ao questionário. Ambos documentos foram anexados nos apêndices A e B, respectivamente.

De modo a facilitar o preenchimento do questionário, as questões foram agrupadas por assuntos da seguinte maneira: solicitação de informações relativas ao perfil acadêmico do entrevistado, suas experiências com sistemas de mineração e com visualização de dados educacionais em ambientes virtuais de aprendizado.

Depois que os participantes respondiam o questionário, eram apresentadas 7 imagens que ilustravam a abordagem de visualização proposta, cada uma delas relacionada às diferentes possibilidades de visualização e interação. A figura 4.1 contém uma destas imagens, que ilustra uma interação com o mouse sobre as barras de total de acessos. A

partir disso as entrevistas foram realizadas utilizando a metodologia semi-estruturada [105], na qual, como roteiro, utilizaram-se as perguntas relacionadas a seguir:

- A visualização no seu ponto de vista é adequada?
- Esta abordagem possibilita realizar interpretações que o auxiliem a avaliar a situação da disciplina e de cada aluno, auxiliando a tomada de decisões?
- Esta abordagem seria útil no seu dia-a-dia?
- Qual(is) elementos, poderia(am) ser aperfeiçoado(s)?



Figura 4.1: Abordagem proposta e apresentada na primeira etapa de entrevistas, que apresenta a interação com o mouse sobre as barras de total de acessos.

Fonte: Autoria Própria, 2016.

4.2 Perfil

As entrevistas foram realizadas com 20 professores, com experiência no ensino de EAD, na cidade de Osório. Entre estes professores, 8 homens e 12 mulheres, metade

possuía de 31 a 40 anos de idade, e a outra metade tinha idade acima ou abaixo deste intervalo, como mostra a tabela 4.1. Dentre eles, todos trabalham ou já trabalharam em diferentes universidades e faculdades públicas e privadas no estado Rio Grande do Sul.

Tabela 4.1: Dados dos entrevistados.

Id	Idade	Sexo	Área de formação	Titulação	Experiência em EAD (anos)
P1	Até 30 anos	Feminino	Administração/Contabilidade/Economia	Especialista	1
P2	De 31 a 40 anos	Feminino	Administração/Contabilidade/Economia	Especialista	2
P3	De 31 a 40 anos	Feminino	Administração/Contabilidade/Economia	Especialista	2
P4	Até 30 anos	Feminino	Administração/Contabilidade/Economia	Especialista	1
P5	De 31 a 40 anos	Feminino	Administração/Contabilidade/Economia	Especialista	3
P6	De 31 a 40 anos	Feminino	Administração/Contabilidade/Economia	Mestre	3
P7	De 31 a 40 anos	Masculino	Administração/Contabilidade/Economia	Mestre	1
P8	De 31 a 40 anos	Masculino	Administração/Contabilidade/Economia	Mestre	10
P9	De 31 a 40 anos	Masculino	Administração/Contabilidade/Economia	Mestre	4
P10	De 41 a 50 anos	Feminino	Educação	Especialista	6
P11	De 31 a 40 anos	Feminino	Educação	Especialista	8
P12	De 41 a 50 anos	Feminino	Educação	Mestre	5
P13	Até 30 anos	Masculino	Educação	Mestre	5
P14	Até 30 anos	Masculino	Educação	Mestre	5
P15	Mais de 50 anos	Feminino	Educação	Mestre	11
P16	Até 30 anos	Feminino	Matemática	Especialista	1
P17	Mais de 50 anos	Feminino	Psicologia	Doutor	15
P18	De 31 a 40 anos	Masculino	Psicologia	Mestre	5
P19	Mais de 50 anos	Masculino	Teologia	Doutor	5
P20	De 31 a 40 anos	Masculino	Teologia	Doutor	4

Analisando a tabela 4.1, identifica-se entre os entrevistados cinco áreas do conhecimento nas quais eles possuem formação principal: Educação, Matemática, Administração/Contabilidade/Economia, Teologia e Psicologia. Além disso, todos os entrevistados eram especialistas, mestres ou doutores nestas áreas.

4.2.1 Experiência

Entre o público entrevistado, encontram-se pesquisadores, professores, tutores, coordenadores de curso e administradores do centro de EAD de uma faculdade. Portanto, foram realizados questionamentos sobre o seu tempo de atuação ministrando disciplinas na modalidade EAD, os ambientes já utilizados e os que utilizam atualmente.

Considerando o tempo de experiência na modalidade EAD, foram obtidos diferentes valores, variando de 1 até 15 anos de experiência na área (Tabela 4.1). Com as informações relacionadas aos ambientes já utilizados e em utilização pelos entrevistados, geraram-se os gráficos apresentados nas figuras 4.2 e 4.3.

Qual(is) destes ambientes já utilizou?

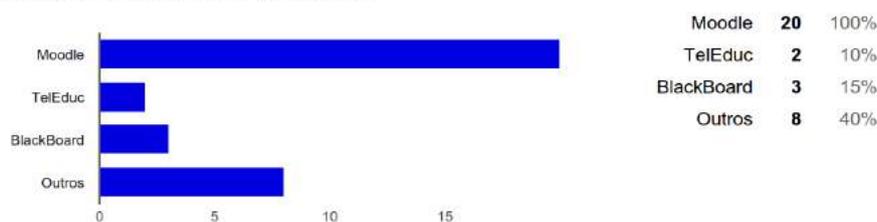


Figura 4.2: Ambientes já utilizados pelos entrevistados.

Fonte: Autoria Própria, 2016.

Qual(is) destes ambientes utiliza hoje, nos cursos que ministra?



Figura 4.3: Ambientes em utilização pelos entrevistados.

Fonte: Autoria Própria, 2016.

Pode-se observar que, de acordo com os questionamentos feitos, e considerando os gráficos das figuras 4.2 e 4.3, o ambiente Moodle é predominantemente o mais utilizado e em uso entre os entrevistados. Foi possível observar que todos já utilizaram ou o utilizam atualmente.

4.3 Opinião sobre as Ferramentas Existentes

De modo a elucidar a experiência dos entrevistados com relação a utilização de ferramentas de acompanhamento e predição de desempenho de alunos, foi solicitado que informassem a utilização ou não, de ferramentas deste tipo em sua atuação. No caso de resposta afirmativa, também foram solicitadas informações adicionais.

Considerando o uso de sistemas de auxílio de monitoramento de atividades dos alunos, obteve-se um número igual de respostas positivas e negativas, dez. Os entrevistados que responderam positivamente à pergunta, foram questionados sobre a disponibilidade ou não de gráficos nestes sistemas, e se haviam, como eram estes gráficos. Destes, 70% responderam que eram disponibilizados gráficos, conforme ilustra a figura 4.4.

Além disso, também foram solicitados os dados demonstrados nestes sistemas de auxílio de monitoramento e seus gráficos. Dentre as respostas informadas, destacam-se as seguintes: número de acessos; atendimentos na plataforma de gestão de relacionamento

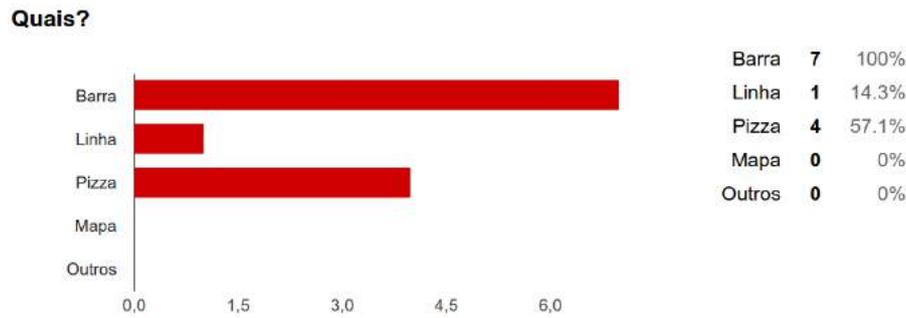


Figura 4.4: Demonstrativo dos elementos gráficos disponibilizados no sistema de monitoramento utilizado.

Fonte: Autoria Própria, 2016.

com o cliente; acompanhamento de correção de provas presenciais; notas dos alunos; e atividades no curso.

Quando os entrevistados foram questionados em relação ao uso de um sistema de predição/mineração de dados, apenas um informou que utilizava. Da mesma forma como nos sistemas de auxílio de monitoramento, o entrevistado informou que era disponibilizado somente um gráfico de barras que possibilita analisar possíveis aprovações e/ou reprovações de alunos no AVA utilizado pela instituição, e ele não soube informar o método de análise utilizado.

Os entrevistados também listaram, de acordo com a sua opinião, alguns recursos considerados importantes para a avaliação do aluno. Desta forma, foi gerado o gráfico apresentado na figura 4.5, o qual representa as respostas e transparece a diferenciação de elementos considerados importantes para o conjunto da avaliação de um aluno, na opinião dos entrevistados.



Figura 4.5: Opinião dos entrevistados para a avaliação do aluno.

Fonte: Autoria Própria, 2016.

Para entender melhor as funcionalidades que um sistema deveria ter para auxiliar o trabalho do professor, solicitou-se que os entrevistados refletissem sobre uma comparação entre duas possíveis abordagens de visualização: uma contendo um elemento gráfico

para cada um dos recursos usados para a avaliação do aluno, e outra contendo um elemento gráfico que representasse toda a participação do aluno, representando a média dos recursos usados. Neste questionamento, obteve-se 70% das respostas afirmando que seria melhor a abordagem com o gráfico da média.

Para finalizar, solicitou-se aos entrevistados que dissertassem sobre sua posição em relação a um sistema de auxílio de avaliação e predição de desempenho de alunos. A partir das respostas fornecidas, destacam-se a seguir algumas:

- P1: "Um sistema que pudesse indicar um nível de rendimento do aluno, diversas medidas poderiam ser tomadas antes que ele tivesse um desempenho baixo ou deixasse o curso."
- P8: "Ferramentas de predição seriam úteis para analisarmos o comportamento dos alunos ao longo do curso e realizarmos ações voltadas a recuperar alunos."
- P9: "Acho oportuna a criação de um sistema em que o professor/tutor possa acompanhar todos os recursos acessados e utilizados pelos estudantes. Um sistema gráfico permitiria uma leitura dinâmica e rápida, dado o número de alunos atendidos, de modo a visualizarmos e monitorarmos o desenvolvimento deles."
- P13: "Qualquer forma de informação que indique a situação de aprendizagem do aluno, particularmente quando estiver em situação 'de risco' é útil."
- P17: "Para melhor visualização das atividades, que demonstram o interesse maior da participação do aluno. Por exemplo, eu gostaria de poder utilizar um gráfico específico, para acompanhar o acesso e a participação do aluno no fórum e ao mesmo tempo ter disponível um gráfico, que contemple todas as outras ferramentas, que o aluno participou."
- P20: "Gostaria muito, pois, quando se atende um número grande de alunos as ferramentas gráficas nos ajudam a identificar os casos que necessitam de maior intervenção e/ou que necessitam maior brevidade no atendimento."

4.4 Avaliação da Proposta

Após a etapa do questionário, descrita na seção 4.3, foram mostradas aos entrevistados sete imagens diferentes, disponíveis no Apêndice C, que ilustravam a abordagem de visualização proposta e suas interações. Nesta parte da entrevista, buscou-se a opinião sobre a abordagem apresentada, utilizando a metodologia semi-estruturada [105], na qual, como roteiro, utilizaram-se perguntas relativas a clareza da abordagem, facilidade de uso

e das informações dispostas (Apêndice B). De forma a simplificar este processo, o áudio destas entrevistas foi gravado, com a autorização dos entrevistados.

A maioria dos entrevistados relatou que a abordagem apresentada teria utilidade significativa em seus cursos e disciplinas. Conforme pode-se observar em alguns trechos do áudio transcritos a seguir, a facilidade, simplicidade e a reunião de diversas informações em um único local, são alguns dos pontos considerados mais importantes da proposta:

- P10: "facilitaria muito o nosso trabalho, a gente acaba na correria acompanhando dados semelhantes, mas não de forma gráfica e automática e por aluno, facilitaria muito.";
- P11: "foi o que eu fiz hoje, um aluno me questionou sobre a nota de interatividade em uma determinada semana, eu tive que pegar o calendário, acessar diversos locais e reunir as informações. Com esta ferramenta eu teria 'perdido' bem menos tempo."

Da mesma forma, alguns entrevistados demonstraram um grande interesse na proposta de auxiliar no combate a evasão e avaliação dos recursos utilizados na disciplina:

- P2: "quanto mais ferramentas para evitar a evasão de alunos melhor";
- P19: "possibilita tanto para o professor quanto para o coordenador do curso, uma dimensão rápida do que é necessário. Possibilitando identificar evasões, dificuldades dos alunos. Dentro do que é proposto, a abordagem é fantástica";
- P17: "a abordagem contribui no aspecto da avaliação, possibilitando alcançar uma informação mais tangível acerca disto. Possibilitando desenvolver um parecer descritivo do aluno ou da disciplina. Possibilita analisar o curso, para colocar ou retirar recursos, por exemplo, um chat pouco acessado. Possibilitaria a demonstração deste relatório inclusive para autoridades como o MEC, para demonstrar dados da disciplina/curso";
- P7: "uma das nossas grandes dificuldades gerenciais são as atividades nos cursos, para evitar a perda de alunos. Algumas coisas que visualizo aqui eu poderia analisar os dados através da visualização de algumas informações e mentalmente deduzir isto, por isso o sistema me auxilia e agiliza o processo".

Com relação ao aspecto visual, a abordagem proposta procurou oferecer gráficos simples, interativos e intuitivos, elementos que, conforme o trecho das entrevistas transcrito a seguir, foram contemplados pela proposta:

- P12: "visualmente agradável, possibilitando um controle maior sobre a disciplina. Nós não temos atualmente como saber as falhas dos alunos, assim nós poderemos ver problemas específicos dos alunos, sem a necessidade de entrar em contato com ele. Tem toda a vida do aluno 'aí'. E não é necessário abrir três, quatro programas para saber tudo isso.";

- P16: "as cores e as formas ajudam muito."

A utilização da abordagem *table lens* (seção 2.2.1), de acordo com os entrevistados, possibilitou uma visão ampla de todos os alunos na mesma visualização sem perder seus detalhes. Também a mineração dos dados apresentados aparece como um ponto importante para a composição completa da proposta. Estes elementos os entrevistados descreveram como fundamentais para o entendimento da disciplina como um todo, assim como para a análise específica dos recursos. Como percebe-se através destes trechos transcritos:

- P5: "Possibilidade de visualizar o todo dos alunos em um único local, aperfeiçoando o entendimento de uma maneira melhor.";
- P14: "um recurso para ser utilizado como norte para a avaliação.";
- P3: "a proposta me parece bem completa, com a mineração eu consigo ter um alcance de quantidade. Possibilita analisar recursos com baixa adesão.";
- P20: "O sistema apresenta uma radiografia de todo o curso/disciplina.";
- P17: "achei interessante a visualização da mineração dos dados, juntamente com o clique para a expansão das barras. Também a visualização dos acessos, que possibilita uma visualização do aluno e geral, com todos. Possibilita uma atenção maior ao aluno, com maior detalhe, mas sem perder o contexto do curso.";

De acordo com as entrevistas e os questionários aplicados, foram elencados os seguintes requisitos:

- Possibilitar a visualização de todos os alunos e suas classificações;
- Permitir a visualização das informações de um único aluno;
- Apresentar os acessos aos recursos da disciplina;
- Permitir a seleção de um período para análise de acessos, tanto de um aluno quanto de todos;
- Possibilitar a comparação do desempenho de um aluno entre as disciplinas que ele cursa;
- Permitir a comparação entre alunos;
- Possibilitar a comparação entre disciplinas;

5. ABORDAGEM DE ANÁLISE VISUAL PARA ACOMPANHAMENTO DISCENTE

Como descrito no capítulo 3, a mineração de dados educacionais para a análise de evasão, baixa produtividade e demais indicadores tem sido abordada por diversos autores [7, 19, 42, 101, 113]. Com isto, pesquisas sobre algoritmos necessários para o desenvolvimento de tais abordagens tem aumentado consideravelmente, dos quais muitos já são utilizados na área, sendo que alguns destes foram descritos na seção 2.5.1 [58].

Assim, esta área tem se desenvolvido de forma a auxiliar administradores, professores e alunos no acompanhamento de suas atividades. Percebe-se também, que com o crescimento elevado da cultura EAD, diversos ambientes virtuais de aprendizagem são disponibilizados, gratuita ou comercialmente, para as instituições de ensino (seção 2.1.1). Com isso, aliado às diversas metodologias de ensino utilizadas nestes sistemas, as quais geralmente possuem estruturas distintas, a complexidade para os processos de mineração também aumenta, sendo necessário gastos com implementação e/ou adaptação das plataformas de mineração de dados a estes ambientes [48].

Observa-se também que muitos destes ambientes não possuem representações visuais associadas com as técnicas de mineração de dados. Este fato pode ser comprovado através da literatura estudada, pois, como mostra a tabela 3.2 apresentada no capítulo 3, muitos dos trabalhos nesta área não apresentam visualizações e interações para auxiliar na mineração de dados. A maioria dos trabalhos estudados disponibilizam somente as análises resultantes do processo de mineração, o que pode acarretar em dificuldades na utilização por pessoas que não possuam conhecimentos estatísticos. Da mesma maneira, a utilização destes trabalhos sem visualização, principalmente em turmas com muitos alunos, dificulta a análise dos resultados, devido aos grandes volumes de informações resultantes.

Considerando este contexto e as hipóteses apresentadas na seção 3.1, surge a seguinte questão de pesquisa: "De que forma diferentes técnicas de visualização e interação, integradas com técnicas de mineração de dados, podem auxiliar professores e administradores no processo de acompanhamento e predição de desempenho de alunos em ambientes virtuais de aprendizagem?". Além desta questão de pesquisa, outros questionamentos também podem ser feitos:

- As interações dos alunos diferem em relação aos recursos disponibilizados em uma disciplina?
- Existe um número mínimo de alunos que deve ser considerado para a utilização de visualização associada à mineração de AVA's?
- Como representar graficamente resultados de algoritmos de mineração de dados de diferentes recursos neste contexto?

Para responder a estes questionamentos, o presente trabalho desenvolveu uma abordagem de visualização de dados de disciplinas EAD processados por algoritmos de mineração. Através desta abordagem, buscou-se integrar formas de análise visual e algoritmos de mineração aos dados gerados pelos ambientes de aprendizagem utilizados. O uso de técnicas de visualização visa possibilitar aos professores, tutores e administradores o acompanhamento de suas disciplinas, fornecendo uma visão geral dos mesmos, com o foco nos alunos e nos recursos utilizados no ambiente.

De modo a tornar transparente alguns dos conceitos utilizados neste trabalho, e considerando as pesquisas efetuadas na literatura, definiu-se que o aluno considerado aprovado, reprovado ou evadido, é aquele que demonstra um padrão de acessos e interações compatíveis ou não com o treinamento efetuado. Esta avaliação é feita com base no histórico de acessos deste aluno no banco de dados da instituição em questão.

Assim, para este treinamento, foram utilizadas as metodologias e técnicas descritas na literatura [58, 113, 7, 19, 42], as quais resultaram em regras de classificação de cada aluno para um destes perfis. A tabela 5.1, ilustra as regras obtidas.

Tabela 5.1: Regras de classificação obtidas após o treinamento.

Classificação	Interações	Tempo médio de acesso por interação
Evadido	< 30	< 1 min
Reprovado	entre 30 e 40	entre 1 e 40 min
Aprovado	> 40	> 40 min

Considerando este contexto, para o desenvolvimento desta abordagem foram utilizadas técnicas de mineração de dados para identificar três perfis de alunos nas disciplinas a serem analisados: com tendência à aprovação, reprovação ou evasão. Para auxiliar nesta análise, foram elaboradas duas visualizações principais: uma para os dados minerados e classificados, e outra sobre as interações realizadas pelos alunos. Além destas, também foram implementadas visualizações para demonstrar os dados da mineração sobre os fóruns da disciplina, visualizações para comparação de todas as disciplinas de um mesmo aluno, e comparação entre diversas disciplinas. Estas visualizações, bem como os algoritmos e o conjunto de dados de teste utilizados são descritos nas seções a seguir.

5.1 Entrada de Dados

Diferentes cursos de educação a distância usam diferentes ambientes virtuais de aprendizagem. Com intuito de possibilitar que a abordagem proposta seja utilizada com dados de diferentes ambientes, e após um estudo sobre os dados disponibilizados pelos

AVA's Moodle, TelEduc e BlackBoard, desenvolveu-se um conjunto de tabelas padrão para uso interno (figura 5.1), que, após o processo de mineração, são populadas com estes dados já minerados. Então, as informações a serem utilizadas na abordagem de visualização já encontram-se prontas, não sendo mais necessário utilizar algoritmos de mineração.



Figura 5.1: Modelagem UML das tabelas internas para armazenamento dos dados.

Fonte: Autoria Própria, 2016.

Pode-se observar na figura 5.1 nove tabelas, as quais armazenam as informações já mineradas de maneira a auxiliar na posterior recuperação dos dados. Estas tabelas são descritas a seguir:

- *cursos*: armazena informações relativas aos cursos, como nome e identificador;
- *user_stats*: contém informações inerentes a cálculos efetuados com os dados do banco, como o tempo de acesso ao AVA e número de interações nele realizadas;
- *forum_mineds*: armazena as informações da mineração dos fóruns, tais como: resultado do cálculo de similaridade entre o solicitado no fórum e as mensagens de cada aluno e o número de mensagens postadas no fórum;

- *clusters*: contém as classificações efetuadas pelo algoritmo de mineração sobre cada aluno. Além disto, armazena o número de acessos efetuado e a nota final, caso esteja disponível;
- *alunos*: armazena as informações dos alunos, como nome, sobrenome, identificador único, etc.;
- *calendarios*: funciona como um calendário, armazenando somente as datas que possuem interações nas disciplinas analisadas pela abordagem, para facilitar a recuperação de informações e o armazenamento de maneira mais otimizada;
- *recursos*: armazena algumas informações dos recursos utilizados na disciplina analisada, como por exemplo fórum, wiki e quiz;
- *nota*: armazena as notas dos alunos para cada recurso utilizado, caso tenham sido disponibilizadas;
- *acos*: possui o registro de todas as atividades realizadas pelos alunos e armazenadas no banco de dados do AVA.

Assim, dados de diferentes ambientes podem ser importados e convertidos para esta estrutura. Como os três AVA's destacados anteriormente possuem tais informações armazenadas na sua base, basta implementar um mecanismo para importação destes dados e esta abordagem pode ser utilizada. Dados de outros AVA's também podem ser usados, desde que possuam os campos definidos como obrigatórios, que são: registros de acessos efetuados; dados sobre os cursos/disciplinas, tais como acessos aos recursos, visualizações e comentários em fóruns. Os dados opcionais para inclusão são as notas finais e parciais (que podem ser usadas para alterar o treinamento padrão).

De acordo com as pesquisas realizadas na literatura [44, 69, 70, 13, 7, 65, 86] observou-se que o Moodle é um dos AVA's mais utilizado em diversas universidades. Deste modo, inicialmente, desenvolveu-se um mecanismo de importação da base de dados do AVA Moodle.

Este mecanismo de importação faz uma conexão direta ao banco de dados, pois os dados de registro de uma disciplina EAD com atividade normal, tendem a gerar arquivos muito grandes, tornando a importação por arquivos ineficiente. Portanto, estabeleceu-se que esta importação se daria através de um acesso a base de dados na qual o Moodle está hospedado. Também, de modo a possibilitar a utilização desta abordagem em outros AVA's, são necessárias apenas alterações na classe *MoodleAdapter* para que as consultas ao banco de dados (*sql*), se adequem a estrutura de tabelas deste outro ambiente. Desta maneira, é possível realizar as devidas consultas, importando para as tabelas padrão desta abordagem somente as informações necessárias e já filtradas.

Com isso, neste caso, para este mecanismo de importação funcionar, foram utilizadas as tabelas ilustradas na figura 5.2 e descritas a seguir. Estas tabelas fazem parte do ambiente Moodle e os nomes dos campos ilustrados na figura foram alterados e alguns omitidos para melhor compreensão. A tabela *mdl_log* é responsável por armazenar todos os registros de atividades efetuadas no AVA; *mdl_course* por sua vez, armazena dados gerais de cadastro das disciplinas; *mdl_grade_grades*, armazena dados das notas relativas ao aluno e a um item descrito na tabela *mdl_grade_items*; por fim, a tabela *mdl_grade_items* armazena os dados de cadastro da nota em questão, como módulo a qual pertence, disciplina e nota máxima, por exemplo.

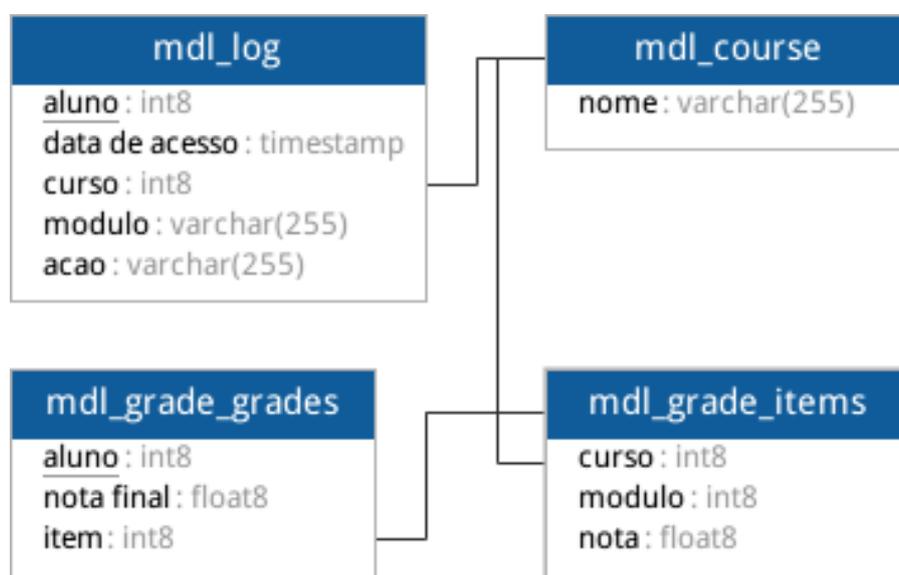


Figura 5.2: Modelo UML das tabelas e campos utilizados do AVA Moodle.

Fonte: Autoria Própria, 2016.

Para o armazenamento destes dados foi estabelecido o uso do banco de dados relacional MySQL, no qual foram armazenadas informações extraídas da base de dados importada. Como fundamento para a criação desta base de dados, foram usados os conceitos apresentados na seção 2.6, assim como, técnicas de *Business Intelligence* [91], as quais buscam auxiliar na posterior recuperação e no gerenciamento de grandes volumes de dados.

Portanto, de maneira a efetuar testes e validações, este mecanismo de importação foi testado com dados do Moodle que foram disponibilizados pela Faculdade Cenecista de Osório (FACOS). Estes dados correspondem a diversos cursos de EAD oferecidos nos últimos três anos pela instituição nesta modalidade de ensino. Estes cursos possuem diferentes recursos de interação, tais como fóruns, wikis, quiz e url externos, e cada um contém dados de 100 a 600 alunos regularmente matriculados. É importante ressaltar que estes dados foram entregues com autorização da direção e de forma anônima, somente para a

realização deste trabalho. Para a associação das notas com as interações realizadas no ambiente estão sendo usados identificadores internos do próprio Moodle.

5.2 Implementação

A implementação desta abordagem foi efetuada com linguagens que possibilitam a sua utilização no ambiente web. Para isto, foi escolhido o *framework Ruby & Rails* com a linguagem de programação *Ruby*. Além disso, foi utilizado o banco de dados relacional MySQL e funções da biblioteca *Javascript D3js* para o desenvolvimento da parte gráfica.

De modo a efetuar as minerações e classificações já descritas anteriormente, foram utilizadas bibliotecas para o *framework Ruby & Rails*. Para a classificação utilizando k-means foi usada a biblioteca *k_means_pp* [106], naïve bayes a biblioteca *stuff-classifier* [74], árvore de decisão a biblioteca *decisiontree* [45] e para tf-idf a biblioteca *tf-idf-similarity* [71].

Com o conjunto de atividades utilizadas na disciplina avaliada, foram aplicados algoritmos de mineração que classificaram o aluno de acordo com sua interação, utilizando para isto, um treinamento efetuado inicialmente com os dados da IES. Para este treinamento, foram utilizadas 10 disciplinas já finalizadas do mesmo curso onde foi validada a abordagem. Estas disciplinas foram escolhidas aleatoriamente, os únicos critérios de seleção foram, já terem sido finalizadas e possuírem os dados de acesso e notas dos alunos. Assim, os algoritmos foram treinados para identificar os padrões de classificação e, após isto, pode-se verificar se os alunos foram classificados em um dos grupos a seguir de maneira correta: com tendência à aprovação, com tendência à reprovação ou com tendência à evasão.

Assim, utilizando as bibliotecas e os critérios mencionadas anteriormente, quando solicitada uma nova análise de disciplina é executada a sequência de ações ilustrada na figura 5.3.

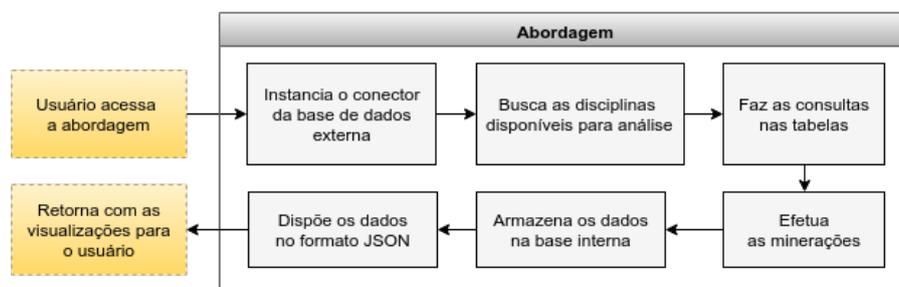


Figura 5.3: Sequência de ações na análise de uma nova disciplina.

Fonte: Autoria Própria, 2016.

5.2.1 Algoritmos de Mineração

Para a mineração de dados foram testados três algoritmos: k-means, naïve bayes e árvore de decisão. Porém, para a utilização destes dois últimos, é necessário fazer um pré-treinamento, que possibilita efetuar as predições e classificações. Para isto, foi elaborado um conjunto de dados anotado com registros sobre as interações dos alunos e suas classificações. Este conjunto de dados foi construído utilizando 100 alunos de 10 disciplinas aleatórias, os quais, possuíam notas e status definido como aprovado ou reprovado. Porém, de modo a classificá-los como evadidos, foram utilizados os conceitos da literatura estudada e, as informações dos professores destas disciplinas. Com isso, foi construído o perfil de acesso de um aluno considerado como provável evasão, que é aquele cuja o número de acessos em relação aos outros alunos é mais baixo.

Após a etapa de treinamento, foram aplicados novamente os três algoritmos nestas mesmas 10 disciplinas. E, conforme demonstrado na tabela 5.2, os algoritmos k-means e naïve bayes apresentaram resultados insatisfatórios, sendo que árvore de decisão teve uma melhor acurácia, sendo, portanto, o escolhido.

Tabela 5.2: Comparação dos algoritmos testados

	Correto %	Correto	Errado %	Errado
Árvore de Decisão	76,00%	23	23,00%	7
K-means	30,00%	9	70,00%	21
Naïve Bayes	23,00%	7	76,00%	23

Os valores apresentados na tabela 5.2 são resultados de cálculos que utilizam as notas e os acessos dos alunos, advindos de 10 disciplinas do banco de dados fornecido pela instituição já descrita. Desta forma, a validação da acurácia foi obtida verificando os dados acadêmicos finais e a nota final do aluno, se ele obteve nota maior do que 7, considerou-se aprovado, entre 5 e 7, reprovado, e caso tenha sido abaixo de 5, considerou-se os acessos do aluno. Caso ele possua um número de acessos baixo em relação aos outros alunos, é considerado evadido. Estes valores foram originados do conjunto de treinamento e da base dados utilizada, e envolveu um total de 10 disciplinas já finalizadas para a validação.

5.2.2 Visualizações de Recursos e Acessos

Depois que é feita a entrada e a mineração dos dados, estes são processados e passados para o módulo de visualização, que gera a saída apresentada na figura 5.4. Nesta

figura é possível verificar que existem duas visualizações distintas: uma com o resultado da aplicação do algoritmo de mineração (figura 5.4a); e outra com informações sobre os acessos realizados pelos estudantes nos diferentes recursos do ambiente (figura 5.4b).

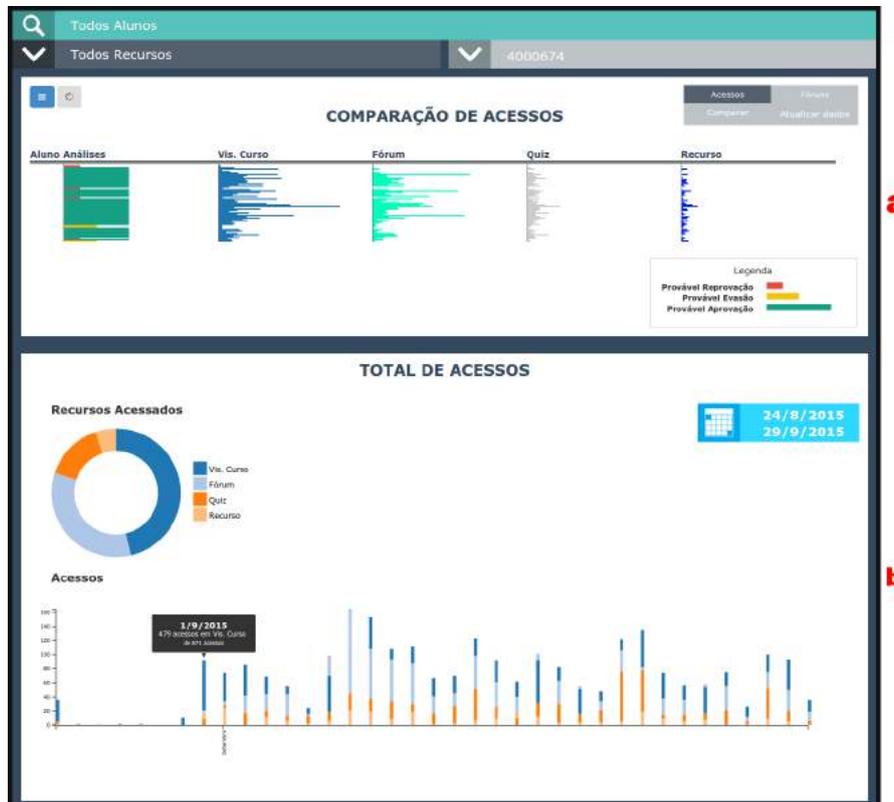


Figura 5.4: Visualização dos acessos dos alunos e do resultado da mineração.

Fonte: Autoria Própria, 2016.

Para viabilizar a exibição de informações de vários alunos, ocupando o menor espaço possível, e evitando a necessidade de rolar a tela, foram implementadas duas visualizações distintas: uma baseada na técnica de *table lens* [77] (descrita na seção 2.2.1) e a outra baseada em um gráfico de bolha. Estas visualizações possibilitam uma rápida análise e comparação de todos os alunos e recursos sem ocupar muito espaço em tela.

Na visualização baseada em *table lens*, cada linha representa as informações de um único aluno e cada coluna representa um recurso da disciplina em questão. Ao clicar em uma linha, ela se expande permitindo visualizar mais detalhes sobre o aluno selecionado, como mostra a figura 5.5a.

As cores na coluna *Análises* (figura 5.4a) representam o perfil identificado para o aluno: amarelo para o aluno com tendência à evasão, vermelho para o aluno com tendência à reprovação e verde para o aluno com tendência a aprovação. A identificação do perfil do aluno é resultado do algoritmo de mineração, tomando como base as interações do aluno nos diferentes recursos. As demais colunas possuem uma cor única e representam, neste exemplo, o total de acessos de cada aluno em cada recurso.

Na visualização sobre os acessos (figura 5.4b) são apresentados todos os acessos efetuados pelos alunos na disciplina ao longo do tempo, classificados pelo recurso acessado. O gráfico de colunas permite que seja selecionado um intervalo de tempo, o que ocasiona um recálculo da proporção no gráfico de pizza.

Já na visualização baseada em gráfico de bolhas, cada círculo representa um aluno, e seu tamanho corresponde a soma dos seus acessos nos recursos da disciplina. As cores, ilustram sua classificação, da mesma forma como na visualização *table lens*. Ao posicionar o mouse sobre um dos círculos, é possível visualizar mais dados, conforme ilustra a figura 5.5b. De modo a auxiliar na visualização destes dados, também pode ser efetuado o agrupamento por classificação, selecionando o valor no campo "Agrupar Por", conforme demonstrado na figura 5.6.

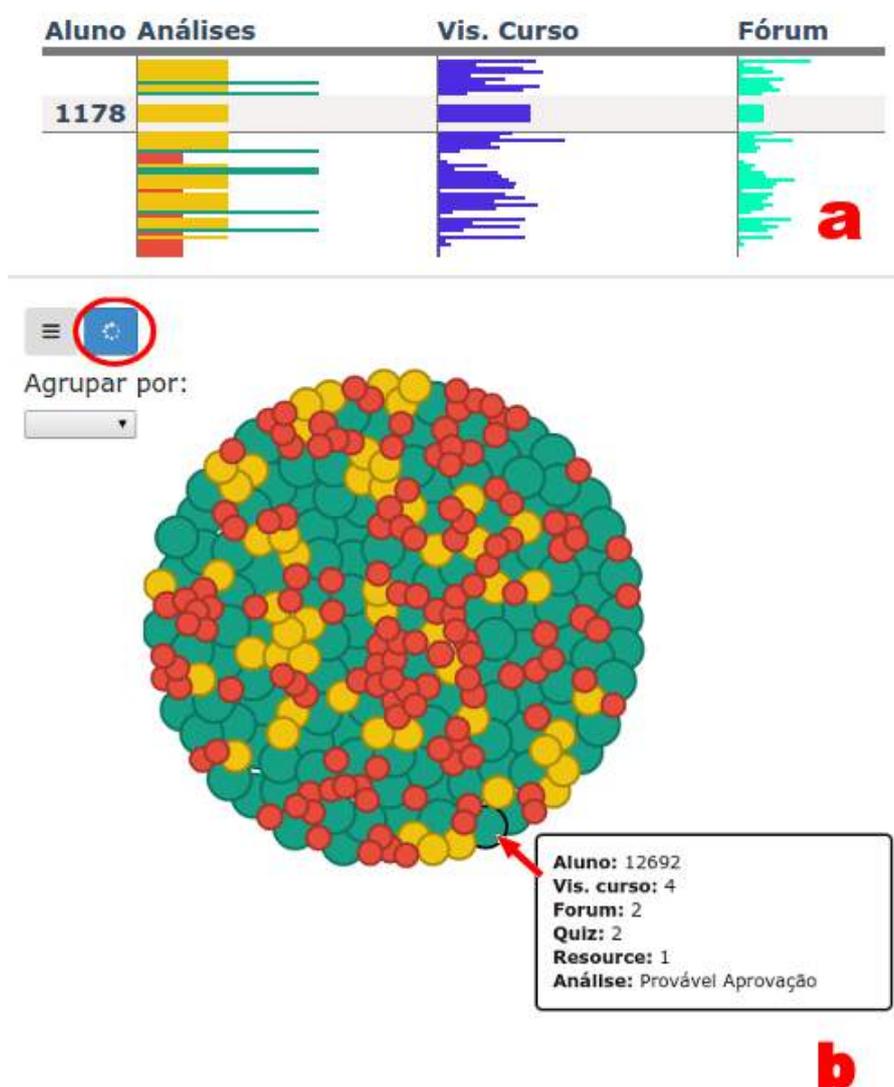


Figura 5.5: *Table lens* com uma linha selecionada (a). Visualização de bolhas com a interação de mouse sobre uma bolha (b).

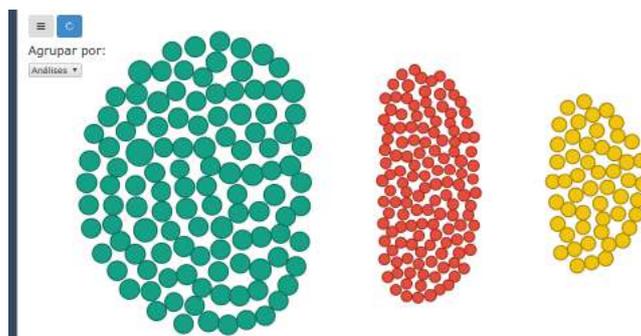


Figura 5.6: Visualização de bolhas com os círculos agrupados de acordo com a classificação do aluno.

Fonte: Aatoria Própria, 2016.

5.2.3 Formas de Interação

Por se tratar de uma abordagem que busca uma visualização simples e intuitiva para o usuário, foram desenvolvidas formas de interação que permitem a manipulação dos dados. Algumas destas são disponibilizadas através dos menus apresentados na figura 5.7. Se no local de *Todos Alunos* for digitado o nome de um aluno, é feita uma busca, e os dados do aluno são destacados automaticamente, como ilustra a figura 5.5a. Esta ação também afeta os gráficos de colunas e de pizza, de maneira a representar apenas os dados do aluno selecionado. Da mesma forma, deixando o campo em branco, os gráficos voltam ao seu estado inicial, apresentando informações da turma como um todo.

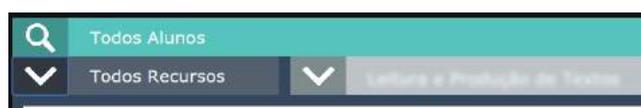


Figura 5.7: Seleção de aluno, recursos e disciplina.

Fonte: Aatoria Própria, 2016.



Figura 5.8: Seleção de disciplina.

Fonte: Aatoria Própria, 2016.

Neste conjunto de menus (figura 5.7) também é possível selecionar um ou mais recursos (figura 5.9) para que sejam analisados somente os selecionados, ou deixar que todos eles sejam exibidos. Além disso, ainda é possível efetuar a troca de disciplina (figura 5.8).

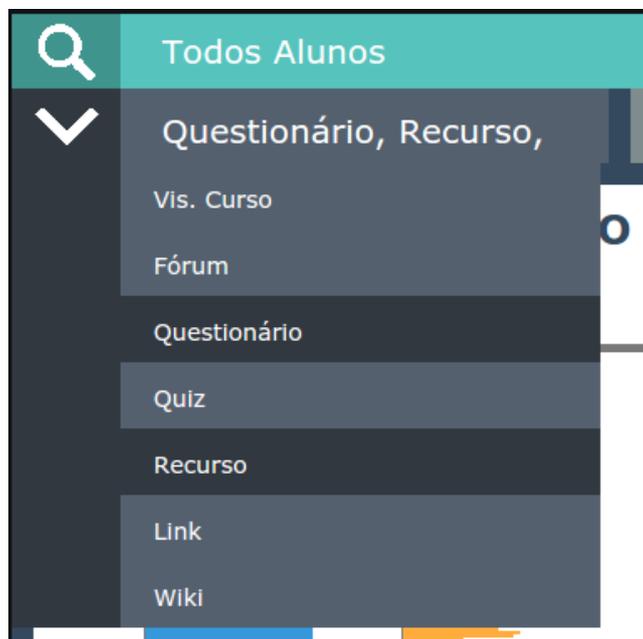


Figura 5.9: Seleção de aluno e recursos.

Fonte: Autoria Própria, 2016.

Considerando as colunas na figura 5.4a, é possível clicar sobre uma delas para efetuar a reordenação levando em conta os dados da coluna clicada, do maior para o menor e vice-versa.

Outra forma de interação implementada possibilita que o usuário selecione com o mouse um período específico do gráfico de colunas, como pode ser visualizado na figura 5.10c. Desta forma, o gráfico de rosca representará os totais de acessos dos recursos dentro do período marcado. Além disso, quando o mouse está sobre uma barra (figura 5.10a e figura 5.4b), informações mais detalhadas sobre a barra selecionada são apresentadas, tais como número de acessos totais em um dia e também o número de acessos relativos ao recurso. Isto é possível pelo fato do gráfico ser do tipo empilhado. Sendo assim, caso o usuário posicione o mouse em outra parte da barra, definida por outra cor (figura 5.10b), verá os acessos relativos ao recurso que ela representa.

5.2.4 Visualizações dos Fóruns e Comparações

Visando fornecer outras formas de auxílio aos professores e tutores, após a implementação e testes com especialista de domínio das visualizações apresentadas em 5.2.2

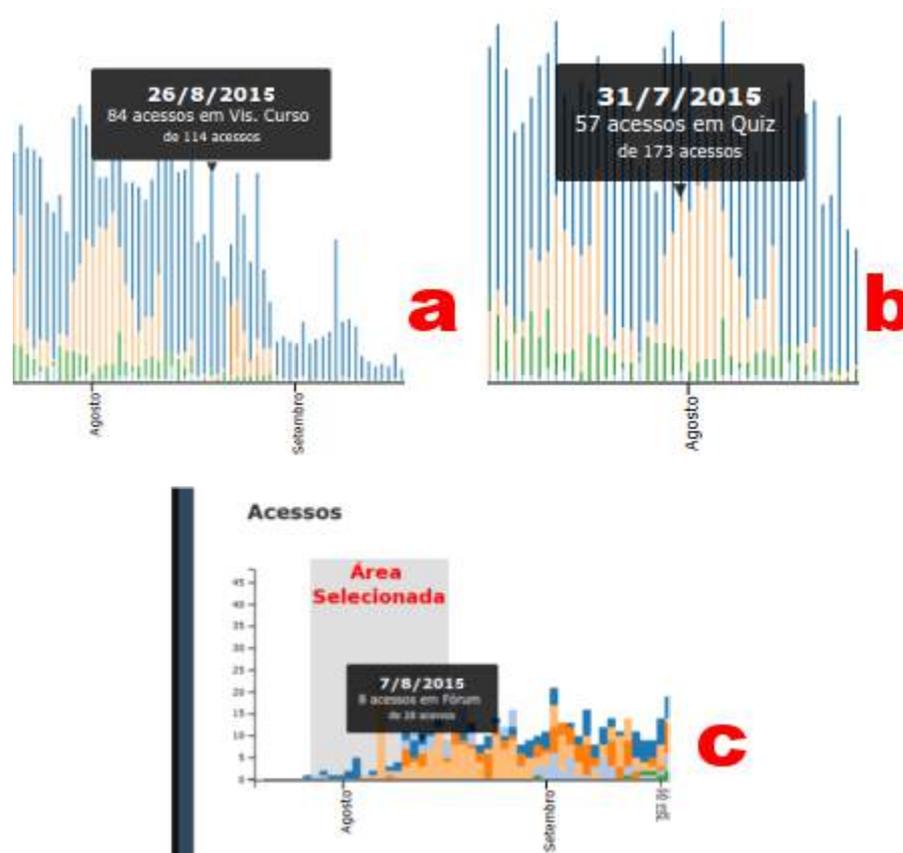


Figura 5.10: Mouse sobre o gráfico de colunas empilhadas (a, b). Seleção de período (c).

Fonte: Autoria Própria, 2016.

utilizando uma base de dados real, foi constatado também seria interessante analisar o recurso de fórum dos AVA's. Como a avaliação deste recurso possibilita um entendimento mais completo das informações sobre os alunos, desenvolveu-se outra visualização, com foco exclusivamente nos fóruns.

Três visualizações distintas com os mesmos dados foram implementadas, de modo a auxiliar na percepção e melhor exploração dos mesmos. A figura 5.11 exemplifica estas visualizações. Os dados apresentados são obtidos através da aplicação do algoritmo de mineração TF-IDF (descrito na seção 2.5.1), sendo que como documento principal foi utilizado o texto fornecido pelo professor na abertura do fórum, e, para a comparação, foram concatenadas todas as mensagens postadas por cada aluno neste mesmo fórum em um único documento. Assim, foi possível calcular o índice de similaridade entre estes documentos, para, então, representar estes valores nestas visualizações.

Na figura 5.11a, cada coluna representa um fórum, e cada linha um aluno. O tamanho dos círculos é relativo ao índice de similaridade calculado pelo algoritmo, que varia de 0 (nada similar) a 1 (totalmente similar). Na figura 5.11b, esta diferença de similaridade é representada com a escala de cores, mantendo o mesmo conceito de colunas e linhas. Diferentemente, a figura 5.11c, representa estes dados em uma visualização de rosca com

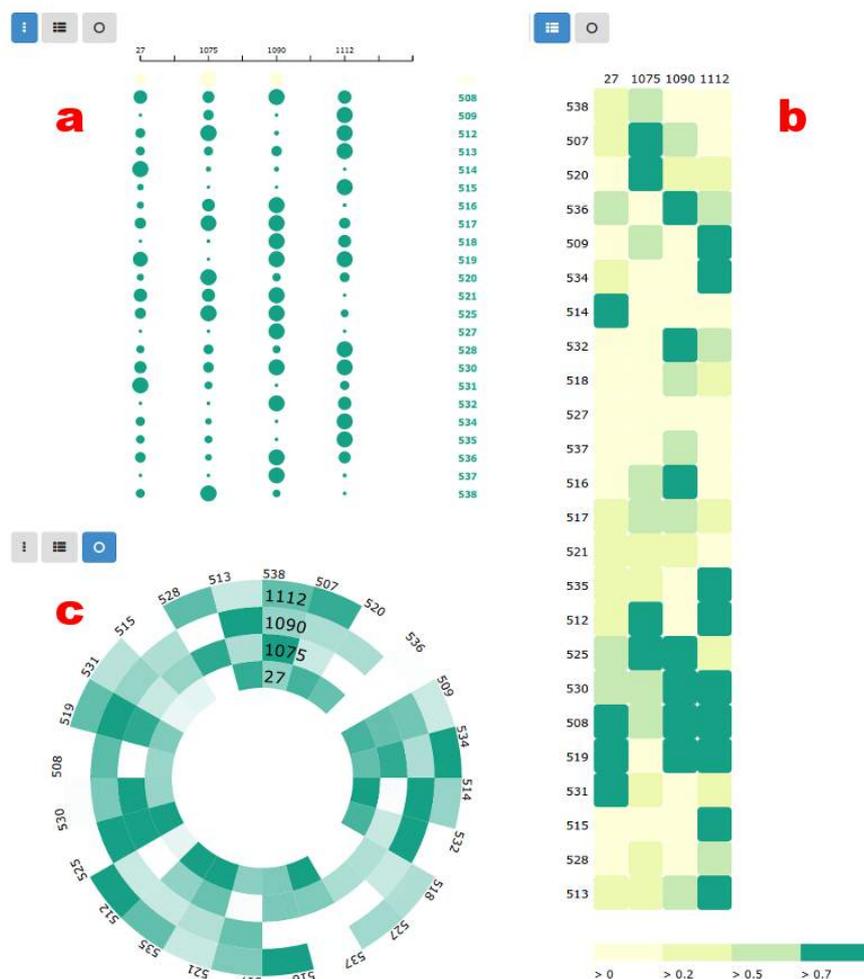


Figura 5.11: Visualizações da mineração dos fóruns do AVA utilizando TF-IDF. O tamanho da bolha representa o índice de similaridade, cada coluna um fórum e cada linha um aluno *a*, assim como em *b*, porém o índice é demonstrado com variações de cor. Já em *c*, cada anel representa um fórum e cada coluna um aluno, utilizando também variações de cor para a similaridade.

Fonte: Autoria Própria, 2016.

vários anéis. Cada anel desta visualização representa um fórum e cada coluna o aluno. Assim como na visualização anterior, para a representação da similaridade, é utilizada a escala de cores, da mais fraca (nada similar) para a mais forte (totalmente similar).

Estas visualizações possuem recurso de interação com o mouse, o qual, quando posicionado em cima de uma das classificações, indica o índice de similaridade calculado em valores numéricos, como demonstrado na figura 5.12, sendo: *a* e *b* duas interações possíveis na mesma visualização (figura 5.11a); *c* a interação da visualização de radar (figura 5.11c); *d* a interação da visualização de mapa de calor (figura 5.11b).

Outras duas visualizações também foram desenvolvidas com o objetivo de possibilitar uma análise individual dos alunos e entre disciplinas. Na figura 5.13 pode-se observar

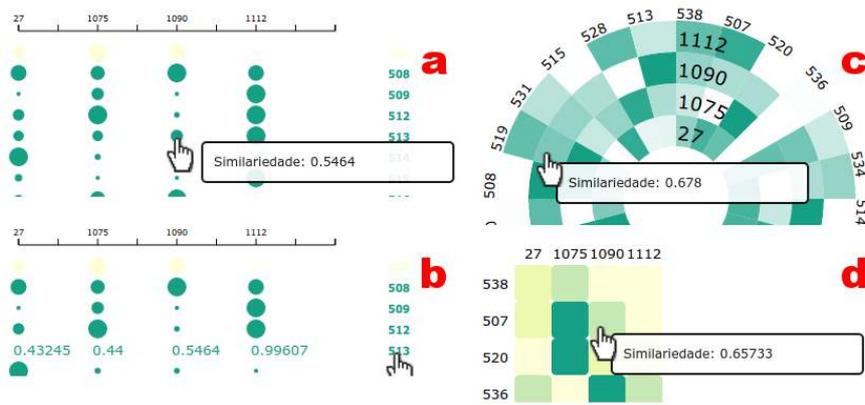


Figura 5.12: Interações nas visualizações de mineração dos fóruns.

Fonte: Autoria Própria, 2016.

a comparação da análise efetuada em cada disciplina deste aluno. As identificações por ícones coloridos demonstram as previsões obtidas. O *check* verde, significa que foi feita a previsão de "Provável Aprovação" para o aluno; a bandeira amarela corresponde a "Provável Evasão"; e a exclamação vermelha, "Provável Reprovação". Juntamente com estas classificações são demonstrados o número de interações, e o tempo médio de cada interação em minutos. Ainda, nesta abordagem, pode-se verificar os acessos efetuados pelo aluno em diferentes disciplinas, agrupados por tipo de recurso no gráfico de barras horizontal. Ambas as visualizações permitem ao usuário, visualizar informações numéricas relativas ao ícone/barra posicionando o mouse sobre ele.

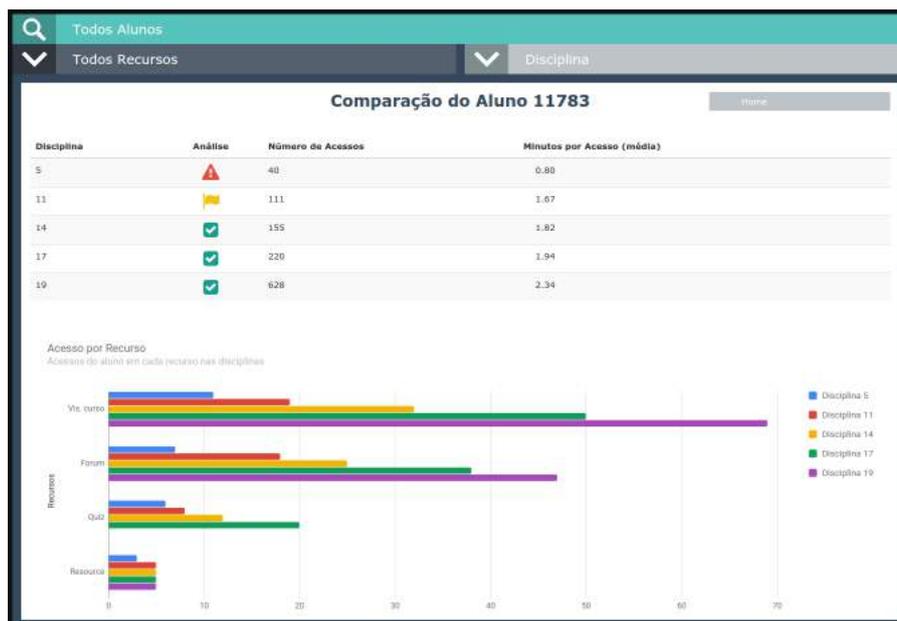


Figura 5.13: Comparação das classificações e acessos do aluno nas disciplinas.

Fonte: Autoria Própria, 2016.

Estas visualizações permitem ao docente uma análise mais completa de cada aluno ao possibilitar a visualização do conjunto de disciplinas cursadas por ele. Assim, identifica-se mais facilmente o seu perfil para poder auxiliá-lo, se necessário, de forma mais pontual. Estas visualizações também são interativas para facilitar a leitura dos dados. A figura 5.14 ilustra o evento do mouse sobre um dos ícones de classificação do aluno em uma disciplina, o qual apresenta seu valor de modo textual.

Disciplina	Análise	Número de Acessos
5		40
11		111
14	 Provável Evasão	
17		220

Figura 5.14: Interação nas comparações de classificação do aluno.

Fonte: Autoria Própria, 2016.

Da mesma forma, a figura 5.15a demonstra o evento de mouse sobre uma das barras, apresentando uma janela *popup* com o seu valor numérico. A figura 5.15b, ilustra o evento de *click* do mouse sobre uma disciplina na legenda, fazendo com que as barras que não pertencem a disciplina selecionada sejam esmaecidas.

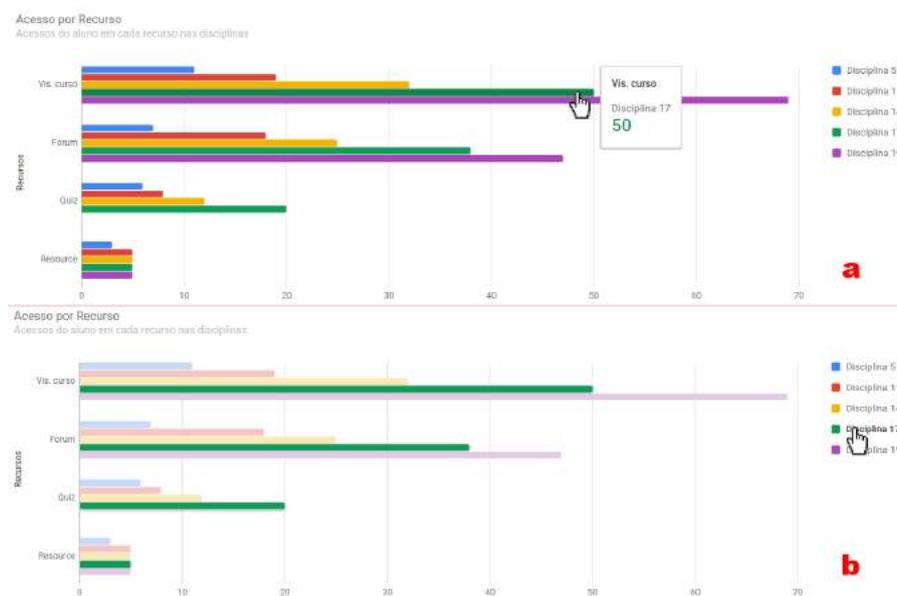


Figura 5.15: Interações nas barras de comparação de recursos do aluno.

Fonte: Autoria Própria, 2016.

A figura 5.16 ilustra a abordagem de comparação de disciplinas. Nesta figura, o gráfico de barras vertical apresenta as classificações efetuadas dos alunos, agrupadas por tipo. O gráfico de barras horizontal apresenta a média de acessos em cada recurso das

disciplinas, agrupado por recurso. Os demais marcadores correspondem ao número total de alunos e ao número total de interações em cada disciplina.

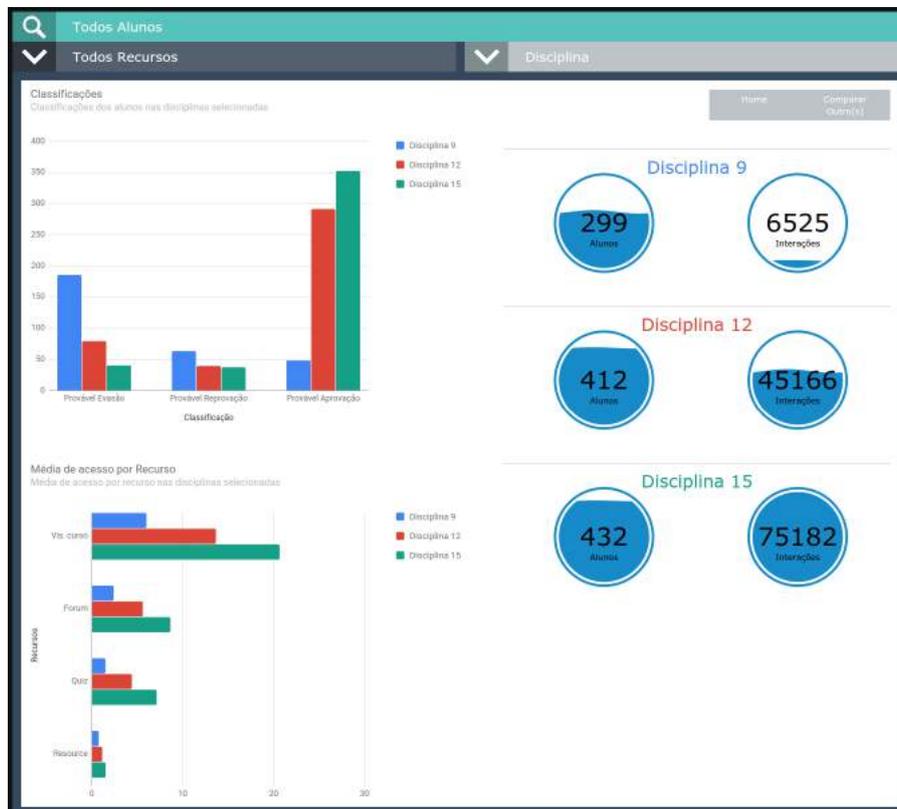


Figura 5.16: Comparação entre disciplinas.

Fonte: Autoria Própria, 2016.

Para manter os padrões de interação já descritos anteriormente para os demais recursos desenvolvidos neste trabalho, a comparação de disciplinas também apresenta a interação demonstrada na figura 5.15, ilustrada nas figuras 5.17 e 5.18.

Esta visualização demonstra como os recursos das disciplinas foram utilizados, e as predições obtidas em cada uma. Com estas informações, é possível refletir sobre algumas questões, tais como: O recurso mais acessado em todas as disciplinas, foi um fator considerável para reduzir as evasões e/ou reprovações? O número de interações também foi responsável por isto? O aluno apresenta o mesmo desempenho nas diferentes disciplinas cursadas?

Em virtude da abordagem de mineração de textos e das demais visualizações descritas nesta seção terem sido estudadas e implementadas após os testes com o especialista de domínio, elas não puderam ser testadas em um ambiente real, com docentes, mas sim, em um ambiente simulado. Por isso, a análise destas visualizações não será descrita no capítulo 6. De acordo com os testes realizados em ambiente simulado e com a literatura pesquisada (capítulos 2.5.1 e 3), nas quais são utilizadas abordagens semelhantes, o nível

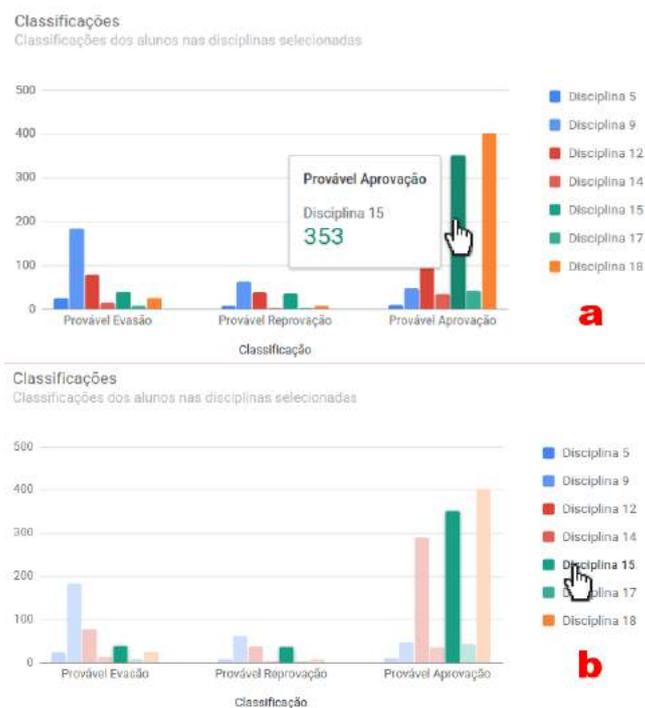


Figura 5.17: Interações nas barras de comparação de classificações dos alunos nas disciplinas.

Fonte: Autoria Própria, 2016.

de acurácia é satisfatório para a mineração nos fóruns. Para estes testes foram utilizadas 10 disciplinas aleatórias do mesmo banco de dados da IES já descrito na seção 5.1.

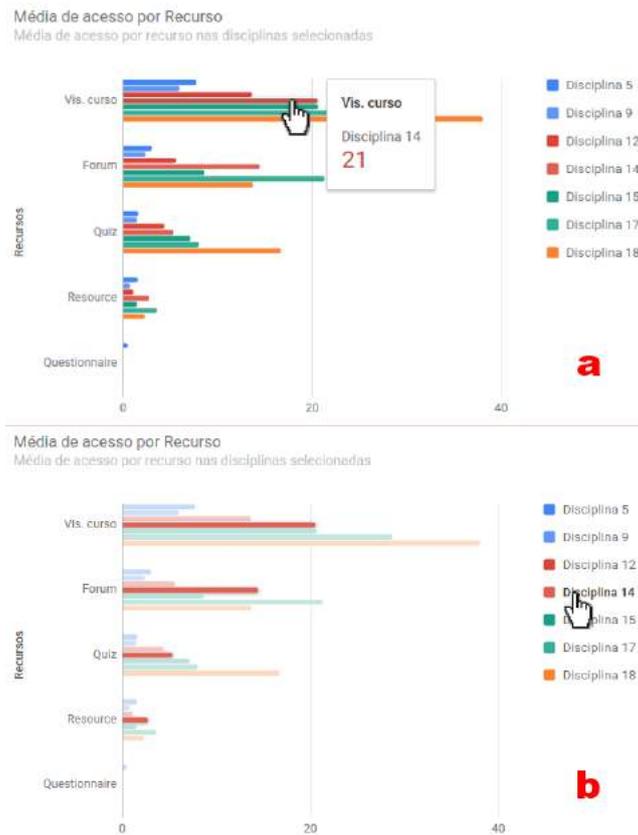


Figura 5.18: Interações nas barras de comparação de médias de acesso aos recursos nas disciplinas.

Fonte: Autoria Própria, 2016.

6. AVALIAÇÃO DE RESULTADOS

Este capítulo apresenta a descrição do processo de avaliação da abordagem, a qual foi efetuada durante o andamento de uma disciplina de uma instituição de ensino que utiliza o AVA Moodle. Os dados para a análise e avaliação da abordagem foram disponibilizados de maneira anônima. As visualizações e interações analisadas são as apresentadas nas seções 5.2.2 e 5.2.3. Desta forma, são descritas nas seções a seguir: a metodologia de avaliação da abordagem; o perfil do tutor/professor responsável pela disciplina e que usou e avaliou a abordagem desenvolvida; o perfil dos alunos da disciplina; as análises efetuadas após a avaliação semanal; e a avaliação final da abordagem desenvolvida e utilizada.

6.1 Metodologia de Avaliação

Para validação do algoritmo de mineração escolhido e das técnicas de visualização implementadas, foi definida uma metodologia de avaliação. Para isto, inicialmente, foi feita uma alteração da interface para permitir a aplicação de mais de um algoritmo de mineração. O objetivo foi de possibilitar a análise e a comparação dos resultados.

Após, foi escolhida uma disciplina de um curso EAD, a qual teve duração de 1 mês, para efetuar avaliações semanais. Assim, foi possível analisar as ações dos alunos após as interações com os tutores, e a previsão dos resultados inferidos por cada algoritmo de mineração, visando avaliar a sua precisão.

Para estas avaliações semanais, e considerando as hipóteses do trabalho, foi elaborado um questionário semi-estruturado que busca responder as seguintes questões que foram utilizadas nas entrevistas:

- O processo de análise dos dados é simples, facilitando a compreensão e a tomada de decisões?
- As visualizações implementadas auxiliam nestas tarefas?
- A análise de diferentes interações do usuário no AVA auxilia na predição do desempenho dos alunos? Como?
- O uso de diferentes algoritmos geraram resultados e análises diferentes? Qual o mais adequado?

Para algumas perguntas deste questionário foi utilizada a escala Likert como escala de respostas, de modo a padronizá-las. Outras perguntas exigiam respostas descritivas para dar uma maior liberdade para o usuário colocar suas impressões e sugestões.

Além disso, as questões para a avaliação da abordagem foram divididas em quatro grupos: perfil do usuário; perfil da disciplina analisada; uso da abordagem no decorrer da disciplina; uso da abordagem de maneira geral.

Com relação ao perfil do entrevistado, foram solicitadas informações acadêmicas e de experiência docente, para permitir análises posteriores a utilização. Da mesma forma, foram solicitados dados sobre a disciplina, tais como área do conhecimento, número de alunos e período. Finalizando os demais grupos, elaborou-se perguntas relativas a análise semanal da abordagem, de modo a acompanhar a evolução dos alunos na disciplina. Assim, pode-se verificar a identificação dos alunos de risco, e também, se os mesmos evoluem após o auxílio do tutor/professor. Para finalizar o questionário, foram elaboradas perguntas direcionadas a avaliação da abordagem como ferramenta de auxílio, para verificar o quão alinhada com as hipóteses deste projeto a abordagem se demonstrou.

Para cada avaliação semanal, além da aplicação do questionário, foram capturadas as imagens das visualizações geradas para cada algoritmo de mineração que estava sendo testado. Estas telas e as avaliações realizadas foram analisadas e comparadas, assim como as sugestões foram consideradas para indicar possíveis aperfeiçoamentos futuros. Optou-se por, inicialmente, apenas identificar as alterações que poderiam ser realizadas deixando a implementação para o final do período de avaliação, de modo a não prejudicar o andamento dos testes e da geração da sequência de análise dos dados.

6.2 Disciplina e Perfil do Tutor

Para avaliar a abordagem, solicitou-se à instituição acesso a uma disciplina em andamento. Esta disciplina faz parte da grade curricular de dois cursos da área de gestão desta IES, que são oferecidos integralmente na modalidade EAD. A disciplina teve 486 alunos regularmente inscritos e a duração de 4 semanas.

A avaliação foi feita com a tutora responsável pela disciplina, que possui o seguinte perfil: mestrado na área de concentração Administração/Contabilidade/Economia, tem até 30 anos de idade e experiência de 2 anos de atuação especificamente na modalidade EAD. Ela utiliza o ambiente Moodle e até o momento não usou nenhum sistema de previsão de desempenho, nem de monitoramento de atividades dos alunos, exceto as ferramentas disponibilizadas pelo próprio Moodle.

6.3 Avaliação Semanal da Abordagem

Como a disciplina teve duração de 4 semanas, foram feitas 4 avaliações semanais, além de uma avaliação final descrita na próxima seção. As perguntas utilizadas nestas avaliações semanais estão listadas na tabela 6.1.

Tabela 6.1: Perguntas da avaliação semanal.

Sigla	Pergunta
P1	Qual a semana da disciplina que está sendo avaliada neste momento?
P2	Com as visualizações disponibilizadas na abordagem, os alunos destacados como provável evasão e/ou reprovação parecem corretos.
P3	Com base no gráfico de acesso dos recursos no tempo, é possível planejar ações tentando reduzir o número de reprovações e/ou evasões.
P4	As informações demonstradas nas visualizações e as interações disponibilizadas, auxiliam na identificação e busca de alunos considerados evadidos e/ou reprovados.
P5	Qual informação neste momento que consideras mais importante a respeito do aluno?
P6	Considerando os resultados apresentados pelos algoritmos, existe algum que parece mais representativo? Qual? Por que?
P7	Você tem alguma sugestão para alteração da forma de apresentação dos resultados? Explique.
P8	Na sua opinião, na semana analisada neste momento, a abordagem poderia apresentar alguma informação que não está presente? Qual?

As respostas para as cinco primeiras perguntas que utilizam a escala Likert, estão apresentadas na tabela 6.2. Estas são identificadas pelas siglas P1, P2, P3, P4 E P5 e separadas por semana. As demais perguntas são comentadas após a tabela por serem questões com respostas descritivas.

Durante a avaliação da primeira semana a tutora ressaltou, de acordo com a pergunta P6, que a análise dos dados dos fóruns e dos questionários é algo que pode auxiliar o tutor a estimular o aluno a realizar tais atividades. Através da questão P7, ela sugeriu que o sistema demonstrasse o tempo que o aluno tem de curso, de forma a possibilitar análises de relação, entre tempo de curso por aluno, e suas interações/classificações já demonstradas na abordagem. Considerando a pergunta P8, ela disse achar necessário mais informações dos alunos como, e-mail, matrícula, curso.

Tabela 6.2: Respostas para a avaliação semanal da abordagem.

Semana (P1)	P2	P3	P4	P5
1	Concordo totalmente	Concordo totalmente	Concordo parcialmente	Possível evasão
2	Concordo totalmente	Concordo totalmente	Concordo totalmente	Possível evasão
3	Concordo totalmente	Concordo totalmente	Concordo totalmente	Possível evasão, Interesse do aluno pela disciplina e pelos recursos utilizados
4	Concordo totalmente	Concordo totalmente	Concordo totalmente	Possível evasão, Interesse do aluno pelos recursos utilizados

Segundo a tutora, a avaliação da segunda semana se mostrou eficiente na predição dos alunos com tendências a evasão e reprovação, auxiliando tanto na identificação destes alunos, como na verificação da efetividade das ações da tutora na semana anterior. Ela também ressaltou a visualização do acesso aos fóruns e quiz como mais representativo neste momento, principalmente para a análise dos períodos de acesso na disciplina. Respondendo a pergunta P8, foi sugerido pela tutora a apresentação do número de disciplinas que o aluno cursou e está cursando, para possibilitar mais análises.

Com o final da disciplina se aproximando, a utilização dos recursos do AVA tendem a aumentar antes do período de provas finais, segundo a tutora. Durante a avaliação na terceira semana, pode-se comprovar esta afirmação, pois as visualizações apresentaram aumentos significativos de acesso aos recursos, principalmente a visualização da disciplina e a utilização dos fóruns. Estes dados, segundo a tutora, são os mais representativos desta semana, conforme a pergunta P6. Foi sugerido também, a complementação de algumas informações, como, o número de tentativas do aluno no quiz, juntamente com suas respectivas notas. A tutora ressaltou que eles não possuem nenhuma visualização que possa fornecer dados de análise da interação dos alunos com os vídeos utilizados pela instituição, o que possibilitaria a descoberta de informações sobre eles, tais como as apresentadas na abordagem desenvolvida e em avaliação.

Na última semana da disciplina realizou-se uma avaliação mais profunda em relação aos alunos que tiveram as classificações alteradas durante a análise das demais semanas. Foi constatado juntamente com a tutora que pelo menos 10 alunos que apareciam com tendência a evasão ou reprovação nas visualizações, devido ao baixo acesso aos recursos, foram contatados e acabaram sendo aprovados. Esta constatação também foi levantada pela tutora na pergunta P6, na qual ela salienta que as informações de acesso ao longo do tempo demonstraram, além do número de acessos, o real interesse dos alunos pela disciplina, auxiliando tanto no "resgate" de alguns, como no melhor entendimento dos recursos dispostos na disciplina. Nas demais perguntas realizadas, foi levantada novamente a questão da apresentação de mais informações pessoais sobre os alunos para que o processo de análise e contato seja ainda mais breve. Estas informações incluem endereço de e-mail, número de matrícula e total de disciplinas cursadas.

6.4 Análise Final

Após a análise semanal descrita na seção 6.3, foi aplicado um último questionário com a finalidade de avaliar a proposta de mineração e visualização como um todo. Foram elaboradas quatro questões, sendo três de múltipla escolha, utilizando a escala Likert, e uma descritiva, na qual a tutora pode opinar livremente sobre sua visão da abordagem desenvolvida, aplicada e testada durante a disciplina. Da mesma forma como foram efetuadas as demais entrevistas, seguiu-se o roteiro semi-estruturado, utilizando as questões descritas na tabela 6.3, juntamente com uma conversa gravada para que fosse possível analisar o maior número de informações possíveis sobre a pesquisa.

Tabela 6.3: Questões da avaliação final.

Sigla	Pergunta
Q1	A utilização da abordagem é simples, no que diz respeito às informações e interações disponíveis?
Q2	A informação dos acessos por recurso auxilia na escolha dos recursos, trazendo informações que podem ser relevantes para a sua utilização ou não em futuras disciplinas.
Q3	As informações são claras e auxiliam o seu trabalho no que diz respeito a identificação do desempenho dos alunos, possibilitando evitar possíveis reprovações e/ou evasões.
Q4	Você tem alguma crítica, sugestão e/ou alteração sobre a abordagem ou esta pesquisa?

A tutora respondeu que concordava totalmente com as afirmações das três primeiras questões, e na questão dissertativa ela descreveu o seguinte: "achei fantástica a possibilidade de ver cada aluno com todas suas informações de acesso. A abordagem de visualização compila dados que usualmente acabamos não controlando e verificando, mas que são necessárias para identificar os alunos com risco de evasão ou mesmo de reprovação".

Desta maneira, a abordagem como um todo foi classificada pela tutora como inovadora, útil e de grande valia para o ambiente de EAD.

6.5 Considerações da Análise da Abordagem

Considerando as entrevistas e análises efetuadas e descritas nas seções 6.3 e 6.4, pode-se concluir que a abordagem apresentou indícios de que o objetivo foi alcançado da forma esperada.

Quanto as minerações e visualizações, observou-se que, dentre os algoritmos implementados foi possível identificar que a árvore de decisão apresentou os melhores resultados no processo. Uma análise dos seus resultados auxiliou a tutora a "resgatar" alunos que poderiam vir a reprovar e/ou evadir.

Quanto as visualizações implementadas, obteve-se uma boa aceitação, o que pode ser comprovado pelo relato da tutora e também dos demais entrevistados para o levantamento de requisitos descrito no capítulo 4. Algumas das questões levantadas pela tutora, tais como o e-mail e a matrícula dos estudantes, são informações pertinentes, porém, dados os quesitos e normas de ética estabelecidas pela pesquisa, não se pode-se informar estes dados, mas a abordagem pode ser facilmente adaptada para mostrar estas informações futuramente. Da mesma forma, outras informações também apontadas pela tutora, a qual achava interessante que constasse na visualização, não foram implementadas por falta de acesso aos dados disponíveis AVA Moodle da instituição. As imagens resultantes das quatro semanas de interação constam no apêndice D.

7. CONCLUSÕES E TRABALHOS FUTUROS

A EAD é, a cada dia mais utilizada pelas instituições de ensino. Uma ferramenta muito importante para apoio, tanto a cursos à distância como a cursos presenciais, são os AVA's, que possuem uma vasta gama de recursos que auxiliam professores e alunos, conforme já descrito no capítulo 2. Desta maneira, um grande volume de dados é gerado, principalmente em cursos que podem possuir centenas de alunos. Neste trabalho, buscou-se apresentar um estudo sobre a mineração de dados destes ambientes e diferentes formas de apresentar tanto os resultados da mineração, como os demais dados de acesso aos recursos disponibilizados, visando auxiliar na solução de problemas diretamente relacionados com a área, como, por exemplo, altas taxas de evasão e reprovação.

Sendo assim, como forma de auxiliar professores e tutores na tarefa de reduzir tais taxas e possibilitar análises com os dados gerados pelos ambientes, este trabalho teve como objetivo desenvolver uma abordagem para o acompanhamento de alunos de cursos à distância, integrando formas de análise visual e algoritmos de mineração a estes dados. Desta maneira, possibilita-se aos professores, tutores e administradores o acompanhamento de suas disciplinas, fornecendo uma visão geral dos mesmos, com o foco nos alunos e nos recursos utilizados no ambiente. Portanto, algumas contribuições deste trabalho são:

- Permitir a visualização dos recursos disponibilizados em uma disciplina e seus percentuais de uso pelos alunos;
- Possibilitar a análise individual dos alunos em uma única visualização interativa;
- Apresentar as previsões advindas dos algoritmos de mineração, através de técnicas de visualização.

Resultados parciais deste trabalho foram apresentados no *6th Workshop on Visual Analytics, Information Visualization and Scientific Visualization* [24].

As classificações geradas só foram possíveis com o auxílio dos algoritmos de mineração estudados, e descritos no capítulo 2, combinados com as visualizações apresentadas. Com isto, percebe-se que na validação da abordagem com os especialistas de domínio (capítulo 6), quando foi efetuado o teste em um ambiente real, a mesma foi descrita satisfatoriamente, respondendo as questões levantadas no decorrer desta pesquisa, a saber:

- As interações dos alunos diferem em relação aos recursos disponibilizados em uma disciplina?

Foi constatado, juntamente com a tutora da disciplina analisada que, as interações dos alunos se diferem nos recursos disponibilizados na disciplina, e esta constatação

possibilitou a tutora, explorar mais os recursos que não estavam sendo utilizados em sua totalidade.

- Existe um número mínimo de alunos que deve ser considerado para a utilização de visualização associada à mineração de AVA's?

A existência de um número mínimo de alunos para a utilização de visualizações associadas à mineração de dados em AVA's, não é uma questão muito difundida na literatura pesquisada, e apresentada no capítulo 2. Da mesma forma, na disciplina utilizada para teste deste trabalho, que possuía um número considerável de alunos, foi possível determinar um número mínimo de alunos para a utilização de visualização. Porém, constatou-se que, as visualizações auxiliaram no rápido entendimento e aplicação dos conhecimentos obtidos através da análise das informações dispostas nas visualizações, independente do número de alunos visualizados.

- Como representar graficamente resultados de algoritmos de mineração de dados de diferentes recursos neste contexto?

As representações gráficas escolhidas para os resultados dos algoritmos de mineração, conforme descrito no capítulo 5, mostraram segundo a avaliação efetuada com os especialistas de domínio, que a demonstração dos dados minerados pode ser feita utilizando visualizações simples, de modo a permitir comparações intuitivas que possibilitam, de acordo com a tutora da disciplina, o resgate de alguns alunos que possivelmente não teriam sido identificados sem o auxílio desta abordagem, pelo padrão irregular de acesso aos recursos utilizados que eles demonstraram.

Pode-se concluir que diferentes técnicas de visualização integradas com a mineração de dados podem auxiliar professores e administradores no acompanhamento discente, possibilitando prever as taxas de evasão e reprovação. Através das entrevistas e testes da abordagem desenvolvida foi possível responder a questão de pesquisa apresentada, pois estas demonstraram que a utilização de visualizações, mesmo que simples, auxiliam no entendimento dos dados descobertos pelos algoritmos de mineração, assim como, na análise visual dos recursos envolvidos nas disciplinas.

Considerando os resultados obtidos com este trabalho, o conhecimento adquirido e buscando possibilitar a sua continuidade, sugere-se que as seguintes pesquisas sejam desenvolvidas futuramente:

- Elaboração de testes com a abordagem utilizando outras fontes de dados, obtidas através de MOOC's e demais AVA's, que não os estudados;
- Acompanhamento dos fóruns com o algoritmo TF-IDF, para verificar o comportamento dos alunos no decorrer da disciplina considerando a adequação da proposta do fórum em relação às suas respostas;

- Comparação entre os dados de diferentes disciplinas buscando identificar informações relevantes para a elaboração do ambiente de aprendizagem de outras disciplinas do mesmo curso;
- Aperfeiçoamento dos algoritmos de mineração possibilitando uma melhor acurácia;
- Desenvolvimento de novas visualizações para apresentar mais informações em uma mesma página, como por exemplo, apresentando as classificações dos alunos nas demais disciplinas analisadas e o índice de similaridade nos fóruns da disciplina, em uma mesma visualização.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Adriaans, P.; Zantinge, D. "Data mining", *England: Addison Wesley*, vol. 1–1, 1996, pp. 1.
- [2] Amershi, S.; Conati, C. "Combining unsupervised and supervised classification to build user models for exploratory", *Journal of Educational Data Mining*, vol. 1–1, 2009, pp. 18–71.
- [3] Andrade, A. F.; Vicari, R. M. "Construindo um ambiente de aprendizagem a distância inspirado na concepção sociointeracionista de vygotsky", *Educação online: teorias, práticas, legislação e formação corporativa*, vol. 2 ed–1, 2003, pp. 255–272.
- [4] Andrienko, G.; Andrienko, N. "Java tools for exploratory cartographic visualization in the internet", *The 10th International Conference on Computer Graphics and Vision GRAPHICON'2000*, vol. 1–10, Aug/Sep 2000, pp. 18–25.
- [5] Angel, E.; Shreiner, D. "Interactive Computer Graphics: A Top down Approach with Shader-based OpenGL". Novo México, EUA: Pearson, 2014, 7 ed ed., 736p.
- [6] Barrosoq, A. C. O.; Gomes, E. B. P. "Tentando entender a gestão do conhecimento", *Revista de Administração Pública*, vol. 33–2, 1999, pp. 147–170.
- [7] Baruque, C. B.; Amaral, M. A.; Barcellos, A.; da Silva Freitas, J. C.; Longo, C. J. "Analysing users access logs in moodle to improve e learning". In: *Proceedings of the 2007 Euro American conference on Telematics and information systems, 2007*, pp. 72:1–72:4.
- [8] Behar, P. A.; Meirelles, S.; Mazzocato, S. B.; de Souza, L. B.; Siqueira, L. G. "Avaliação de ambientes virtuais de aprendizagem: O caso do rooda na ufrgs", *Revista Avances en Sistemas e Informática*, vol. 4–1, 2007, pp. 88.
- [9] Belloni, M. L. "Educação a distância". Campinas, SP, Brasil: Autores Associados, 2004, 1 st ed., 126p.
- [10] Berry, M. J.; Linoff, G. S. "Data mining techniques: for marketing, sales, and customer relationship management". New York, USA: John Wiley and Sons, 2011, 3 rd ed., 888p.
- [11] Biggs, N. L.; Lloyd, E.; Wilson, R. J. "Graph Theory, 1736-1936". New York, USA: Oxford University Press Inc, 1976, 4 ed., vol. 1, 223p.
- [12] Blackboard. "Blackboard". Capturado em: <http://blackboard.grupoa.com.br>, 2014-09-29.

- [13] Bovo, A.; Sanchez, S.; Héguy, O.; Duthen, Y. "Demonstration of a moodle student monitoring web application", *Proceedings of the 6th International Conference on Educational Data Mining*, vol. 6–1, 2013, pp. 390–391.
- [14] Branco, V. M. A. "Visualização como suporte à exploração de uma base de dados pluviométrica", Dissertação de Mestrado, Universidade de São Paulo - ICMC, São Carlos, SP, Brasil, 2003, 87p.
- [15] Bravo, J.; Ortigosa, A. "Detecting symptoms of low performance using production rules", *Proceedings of 2nd International Conference on Educational Data Mining*, vol. 2–1, 2009, pp. 31–40.
- [16] Card, S. K.; Mackinlay, J. D.; Shneiderman, B. "Readings in Information Visualization: Using Vision to Think". San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, 1st ed., 712p.
- [17] Cerceau, A. d. D. "Formação a distancia de recursos humanos para informatica educativa", Dissertação de Mestrado, Universidade Estadual de Campinas - UNICAMP, Campinas, SP, Brasil, 1998, 118p.
- [18] Chan, W. W. Y. "A survey on multivariate data visualization", *Department of Computer Science and Engineering. Hong Kong University of Science and Technology*, vol. 8–6, Mar 2006, pp. 1–29.
- [19] Chirioiu, M. S.; Mihăescu, M. C.; Burdescu, D. D. "Students activity visualization tool", *Proceedings of the 6th International Conference on Educational Data Mining*, vol. 6–1, 2013, pp. 392–393.
- [20] Coates, H.; James, R.; Baldwin, G. "A critical examination of the effects of learning management systems on university teaching and learning", *Tertiary education and management*, vol. 11–1, 2005, pp. 19–36.
- [21] Cocea, M.; Weibelzahl, S. "Can log files analysis estimate learners level of motivation?", *14th Workshop on Adaptivity and User Modeling in Interactive Systems*, vol. 14–1, 2011, pp. 32–35.
- [22] Cole, J.; Foster, H. "Using Moodle: Teaching with the popular open source course management system". Sebastopol, Califórnia: O'Reilly Media, Inc, 2007, 2 nd ed., 284p.
- [23] Collins, H.; Ferreira, A. "Relatos de experiência de ensino e aprendizagem de línguas na Internet". Campinas, SP, Brasil: Mercado de Letras, 2004, 1 st ed., *Coleção As faces da lingüística aplicada*, vol. 7, 336p.

- [24] Comba, J. L. D.; Nonao, L. G. (Editores), "Towards visual analysis techniques for monitoring students of distance education courses", vol. 4, Sociedade Brasileira de Computação, 2015.
- [25] Conti, F.; Charão, A. S. "Análise de prazos de entrega de atividades no moodle: um estudo de caso utilizando mineração de dados", *Revista Novas Tecnologias na Educação*, vol. 9–2, 2011, pp. 10.
- [26] Côrtes, S. d. C.; Porcaro, R. M.; Lifschitz, S. "Mineração de dados-funcionalidades, técnicas e abordagens". Rio de Janeiro, RJ, Brasil: Pontifícia Universidade Católica do Rio de Janeiro, 2002, 33p.
- [27] Coursera. "Coursera". Capturado em: <https://www.coursera.org/>, 2014-08-05.
- [28] Cruz, D. M. d. "A produção audiovisual na virtualização do ensino superior: subsídios para a formação docente", *Educação Temática Digital*, vol. 8–2, Jun 2007, pp. 23–44.
- [29] Dal Forno, J. P.; Knoll, G. F. "Os moocs no mundo: Um levantamento de cursos online abertos massivos", *Nuances: estudos sobre Educação*, vol. 24–3, Set/Dez 2013, pp. 178–194.
- [30] de Amo, S. "Técnicas de mineração de dados", *XXIV Congresso da Sociedade Brasileira de Computação*, vol. 1–1, Jul/Ago 2004, pp. 43.
- [31] de Educação a Distância, A. B. "Censo ead.br: Relatório analítico da aprendizagem a distância no brasil 2014", *Censoead.Br*, vol. 1–1, Dez 2014, pp. 155.
- [32] de Oliveira Júnior, R. L.; Esmín, A. A.; Coelho, T. A.; Araújo, D. L.; Alonso, L.; Giroto, R. "Uma ferramenta de monitoramento automático de mensagens de fóruns em ambientes virtuais de aprendizagem", *Anais do XXII SBIE - XVII WIE*, vol. 22–1, Nov 2011, pp. 60–69.
- [33] DeBoer, J.; Ho, A.; Stump, G. S.; Pritchard, D. E.; Seaton, D.; Breslow, L. "Bringing student backgrounds online: Mooc user demographics, site usage, and online learning", *Proceedings of the 6th International Conference on Educational Data Mining*, vol. 6–1, 2013, pp. 312–313.
- [34] Dekker, G. W.; Pechenizkiy, M.; Vleeshouwers, J. M. "Predicting students drop out: A case study", *Proceedings of 2nd International Conference on Educational Data Mining*, vol. 2–1, 2009, pp. 41–50.
- [35] Dilly, R. "Data mining for design and manufacturing". In: *Data mining for design and manufacturing*, 2002 nd ed., Springer (Editor), Norwell, MA, USA: Kluwer Academic Publishers, 2001, vol. 1, cap. Data mining: An introduction, pp. 544.

- [36] EDX. “Edx”. Capturado em: <https://www.edx.org/>, 2014-08-05.
- [37] Elmasri, R.; Navathe, S. B. “Fundamentals of database systems”. USA: Addison-Wesley Publishing Company, 2010, 6 th ed., 1200p.
- [38] Fayyad, U. “Mining databases: Towards algorithms for knowledge discovery”, *IEEE Data Engineering Bulletin*, vol. 21–1, Mar 1998, pp. 39–48.
- [39] Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.; et al.. “Knowledge discovery and data mining: Towards a unifying framework”. In: Proceedings of KDD, 1996, pp. 82–88.
- [40] Franco, C. P. “A plataforma moodle como alternativa para uma educação flexível”, *Revista EducaOnline*, vol. 4–1, Jan/Abr 2010, pp. 27–41.
- [41] Garcia, P. L.; Lacleta, M. L. S. E. “A revolução pedagógica: o meio moodle”. Capturado em: http://contenidos.universia.es/html_trad/traducirEspecial/params/especial/bc/seccion/6/titulo/REVOLUCIONPEDAGOGICA-ENTORNO-MOODLE.html, 2015-02-25.
- [42] García-Saiz, D.; Pantaleon, M. E. Z. “E-learning web miner: A data mining application to help instructors involved in virtual courses.” In: Proceedings of the 4th International Conference on Educational Data Mining, 2011, pp. 323–324.
- [43] Golding, P.; Donaldson, O. “Predicting academic performance”. In: Frontiers in Education Conference, 36th Annual, 2006, pp. 21–26.
- [44] Graf, S.; Ives, C.; Rahman, N.; Ferri, A. “Aat: a tool for accessing and analysing students’ behaviour data in learning systems”. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, 2011, pp. 174–179.
- [45] Grigorik, I. “Decision tree”. Capturado em: <https://github.com/igrigorik/decisiontree>, 2016-01-26.
- [46] Hair, J. F.; Black, W. C.; Babin, B. J.; Anderson, R. E.; Tatham, R. L. “Multivariate Data Analysis”. USA: Pearson Prentice Hall Upper Saddle River, NJ, 2010, 7th ed., 761p.
- [47] Hale, S. A. “Multilinguals and wikipedia editing”. In: Proceedings of the 2014 ACM Conference on Web Science, 2014, pp. 99–108.
- [48] Han, J.; Kamber, M. “Data Mining: Concepts and Techniques”. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011, 3 rd ed., 744p.
- [49] Hand, D. J.; Smyth, P.; Mannila, H. “Principles of Data Mining (Adaptive Computation and Machine Learning)”. Massachusetts, USA: A Bradford Book The MIT Press, 2001, 1 st ed., 584p.

- [50] Heer, J.; Shneiderman, B. "Interactive dynamics for visual analysis", *Queue*, vol. 10–2, Feb 2012, pp. 30–55.
- [51] Herman, I.; Melancon, G.; Marshall, M. "Graph visualization and navigation in information visualization: A survey", *Visualization and Computer Graphics, IEEE Transactions on*, vol. 6–1, Jan 2000, pp. 24–43.
- [52] Inselberg, A.; Dimsdale, B. "Parallel coordinates: A tool for visualizing multi-dimensional geometry". In: Proceedings of the 1st Conference on Visualization '90, 1990, pp. 361–378.
- [53] Interrante, V. "Harnessing natural textures for multivariate visualization", *Computer Graphics and Applications, IEEE*, vol. 20–6, Nov 2000, pp. 6–11.
- [54] Johnson, M.; Barnes, T. "Visualizing educational data from logic tutors". In: Intelligent Tutoring Systems, Vincent Aleven, Judy Kay, J. M. (Editor), 2010, pp. 233–235.
- [55] Johnson, M.; Eagle, M.; Joseph, L.; Barnes, T. "The edm vis tool". In: Proceedings of the 4th International Conference on Educational Data Mining, Pechenizkiy, M.; Calders, T.; Conati, C.; Ventura, S.; Romero, C.; Stamper, J. C. (Editores), 2011, pp. 349–350.
- [56] Jonassen, D. "O uso das novas tecnologias na educação a distância e a aprendizagem construtivista", *Em Aberto, Brasília, ano 16*, vol. 16–70, Abr/Jun 1996, pp. 70–88.
- [57] Kampff, A. J. C. "Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente", Tese de Doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil, 2009, 186p.
- [58] Kampff, A. J. C.; Ferreira, V. H.; Reategui, E.; Lima, J. V. "Identifying evasion and poor performance profiles for the generation of alerts in a distance learning context", *Revista Latinoamericana de Tecnología Educativa*, vol. 13–2, 2014, pp. 61–76.
- [59] Keim, D. A. "Information visualization and visual data mining", *IEEE Transactions on Visualization and Computer Graphics*, vol. 8–1, jan 2002, pp. 8.
- [60] Keim, D. A.; Kohlhammer, J.; Ellis, G.; Mansmann, F. "Mastering The Information Age-Solving Problems with Visual Analytics". Goslar, Alemanha: Eurographics Association, 2011, 1 st ed., vol. 7, 170p, proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11).
- [61] Keim, D. A.; Kriegel, H.-P. "Visualization techniques for mining large databases: A comparison", *Knowledge and Data Engineering, IEEE Transactions on*, vol. 8–6, Dec 1996, pp. 923–938.

- [62] Keim, D. A.; Mansmann, F.; Schneidewind, J.; Thomas, J.; Ziegler, H. “Visual analytics: Scope and challenges”. In: *Visual Data Mining*, Simoff, S. J.; Böhlen, M. H.; Mazeika, A. (Editores), Berlin, Heidelberg: Springer-Verlag, 2008, cap. Visual Analytics: Scope and Challenges, pp. 76–90.
- [63] Keim, D. A.; Mansmann, F.; Schneidewind, J.; Ziegler, H. “Challenges in visual data analysis”. In: *Information Visualization, 2006. IV 2006. Tenth International Conference on*, 2006, pp. 9–16.
- [64] Leite, M. T. M. “O ambiente virtual de aprendizagem moodle na prática docente: conteúdos pedagógicos”. Capturado em: <http://www.virtual.unifesp.br/cursos/oficinamoodle/textomoodlevirtual.pdf>, 2015-06-08.
- [65] Lopez, M. I.; Luna, J.; Romero, C.; Ventura, S. “Classification via clustering for predicting final marks based on student participation in forums.”, *5th International conference on educational data mining*, vol. 5–1, 2012, pp. 148–151.
- [66] Machado, L. D. “Concepções de espaço e tempo nas teorias de educação a distância”. In: *Congresso Internacional de Educação a Distância*, 2005, pp. 10.
- [67] Maciel, C. “Ambientes Virtuais de Aprendizagem”. Cuiabá, MT, Brasil: Universidade Federal do Mato Grosso, 2013, 1 st ed., 262p.
- [68] Mazza, R.; Dimitrova, V. “Coursevis: Externalising student information to facilitate instructors in distance learning”, *Proceedings of the 11th International conference in Artificial Intelligence in Education*, vol. 11–1, Jul 2003, pp. 279–286.
- [69] Mazza, R.; Dimitrova, V. “Coursevis: A graphical student monitoring tool for supporting instructors in web-based distance courses”, *International Journal of Human-Computer Studies*, vol. 65–2, Feb 2007, pp. 125–139.
- [70] Mazza, R.; Milani, C. “Gismo: a graphical interactive student monitoring tool for course management systems”, *International Conference on Technology Enhanced Learning*, vol. 4–1, Nov 2004, pp. 18–19.
- [71] McKinney, J. “Ruby vector space model (vsm) with tf*idf weights”. Capturado em: <https://github.com/jpmckinney/tf-idf-similarity>, 2016-01-26.
- [72] Moodle. “Moodle”. Capturado em: <http://moodle.org/>, 2014-12-23.
- [73] NCOMVA. “Table lens”. Capturado em: http://www.ncomva.se/guide/?chapter=Visualizations§ion=Table%20Lens#_General, 2014-12-16.
- [74] Nedelcu, A. “Stuff classifier”. Capturado em: <https://github.com/alexandru/stuff-classifier>, 2016-01-26.

- [75] Pedraza-Perez, R.; Romero, C.; Ventura, S. "A java desktop tool for mining moodle data". In: *Proceedings of the 3rd Conference on Educational Data Mining*, 2010, pp. 319–320.
- [76] Pereira, A. T. C.; Schmitt, V.; Dias, M. "Virtual learning environments", *Virtual learning environments in different contexts*, vol. 1–1, 2007, pp. 5.
- [77] Rao, R.; Card, S. K. "The table lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1994, pp. 318–322.
- [78] Rencher, A. C. "Methods of Multivariate Analysis". Canada, USA: John Wiley and Sons, Inc., 2002, 2nd ed., 727p.
- [79] Rezende, S.; Pugliesi, J.; Melanda, E.; Pula, M. "Sistemas inteligentes: Fundamentos e aplicações". In: *Sistemas Inteligentes: Fundamentos e Aplicações*, 1 ed ed., LTDA, M. (Editor), Barueri, SP, Brasil: Manole, 2003, vol. 1, cap. Mineração de dados, pp. 307–336.
- [80] Rezende, Y. "Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual", *Ciência da Informação*, vol. 31–2, Mai/Ago 2002, pp. 120–128.
- [81] Ribeiro, E. N.; Mendonça, G. A. d. A.; Mendonça, A. F. "A importância dos ambientes virtuais de aprendizagem na busca de novos domínios na ead", *13º Congresso Internacional de Educação a Distância*, vol. 13–1, 2007, pp. 11.
- [82] Ribeiro, F. A. A. "Recomendação de documentos para os usuários do ava moodle a partir das hashtags postadas nos fóruns", *Dissertação de Mestrado*, Centro Tecnológico/Universidade Estadual do Maranhão, São Luis, MA, Brasil, 2013, 109p.
- [83] Rocha, H.; Silva, C.; Freire, F.; et al.. "Projeto teleduc: Pesquisa e desenvolvimento de tecnologia para educação a distância". In: *IX Congresso Internacional de Educação a Distância*, 2002, pp. 72.
- [84] Rocha, H. V.; Moraes, M. "Educação a distância: Fundamentos e práticas". Campinas, SP, Brasil: SP:UNICAMP/NIED, 2002, vol. 1, cap. O ambiente TelEduc para Educação à Distância baseada na Web: Princípios, Funcionalidades e Perspectivas de desenvolvimento, pp. 197–212.
- [85] Romero, C.; Castro, C.; Ventura, S. "A moodle block for selecting, visualizing and mining students usage data", *Proceedings of the 6th International Conference on Educational Data Mining*, vol. 6–1, 2013, pp. 400–401.

- [86] Romero, C.; Espejo, P. G.; Zafra, A.; Romero, J. R.; Ventura, S. “Web usage mining for predicting final marks of students that use moodle courses”, *Computer Applications in Engineering Education*, vol. 21–1, 2013, pp. 135–146.
- [87] Salton, G.; Buckley, C. “Term-weighting approaches in automatic text retrieval”, *Information processing and management*, vol. 24–5, 1988, pp. 513–523.
- [88] Santos, J. F. S. “Avaliação no ensino a distância”, *Revista Iberoamericana de Educación*, vol. 38–4, 2006, pp. 9.
- [89] Santos, R. “Weka na munheca: Um guia para uso do weka em scripts e integração com aplicações em java”, Relatório Técnico 1, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brasil, 2005, 20p.
- [90] Schlemmer, E. “Ambiente virtual de aprendizagem (ava): uma proposta para a sociedade em rede na cultura da aprendizagem”, *Aprendizagem em ambientes virtuais: compartilhando idéias e construindo cenários*, vol. 2 ed–1, 2005, pp. 135–159.
- [91] Sezões, C.; Oliveira, J.; Baptista, M. “Business intelligence”. São João do Estoril, Portugal: São João do Estoril, Portugal: Sociedade Portuguesa de Inovação, 2006, 1 st ed., 160p.
- [92] Shimabukuro, M. H. “Visualizações temporais em uma plataforma de software extensível e adaptável”, Tese de Doutorado, Universidade de São Paulo, São Paulo, SP, Brasil, 2004, 147p.
- [93] Shneiderman, B. “The eyes have it: A task by data type taxonomy for information visualizations”. In: Proceedings of the 1996 IEEE Symposium on Visual Languages, 1996, pp. 336–343.
- [94] Solar. “Solar”. Capturado em: <http://www2.virtual.ufc.br/portal2/index.php/ambiente/solar2>, 2014-12-23.
- [95] Solar. “Solar”. Capturado em: <http://www.solar.virtual.ufc.br/>, 2014-12-23.
- [96] TelEduc. “Teleduc”. Capturado em: <http://www.teleduc.org.br/>, 2014-12-23.
- [97] TelEduc. “Teleduc”. Capturado em: http://ggte.unicamp.br/~teleduc/pagina_inicial/autenticacao_cadastro.php, 2014-12-23.
- [98] Teles, L. “Educação a distância: o estado da arte”. São Paulo, SP, Brasil: Pearson Education do Brasil, 2009, 1 ed ed., cap. A aprendizagem por e-learning, pp. 72–80.

- [99] Thomas, J. J.; Cook, K. A. "Illuminating the path: The research and development agenda for visual analytics". USA: IEEE Computer Society Press, 2005, 1 st ed., 185p.
- [100] Thuraisingham, B. "Data Mining: Technologies, Techniques, Tools, and Trends". Boca Raton, FL, USA: CRC Press, Inc., 1998, 1 st ed., 288p.
- [101] Tinto, V. "Dropout from higher education: A theoretical synthesis of recent research". Columbia, SC, EUA: ERIC - Education Resources Information Center, 1975, 1 ed., 99p.
- [102] Torres, P. L.; Fialho, F. A. P. "Educação a distância: passado, presente e futuro", *Educação a distância: o estado da arte*, vol. 1–1, 2009, pp. 456–461.
- [103] Troussas, C.; Virvou, M.; Junshean Espinosa, K.; Llaguno, K.; Caro, J. "Sentiment analysis of facebook statuses using naive bayes classifier for language learning". In: Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on, 2013, pp. 6.
- [104] Udacity. "Udacity". Capturado em: <https://www.udacity.com>, 2014-08-05.
- [105] Valdete Boni, S. J. Q. "Aprendendo a entrevistar: como fazer entrevistas em ciências sociais", *Revista Eletrônica dos Pós-Graduandos em Sociologia Política da UFSC*, vol. 2–1, Jan 2005, pp. 68–80.
- [106] Vetesnik, O. "Kmeanspp". Capturado em: https://github.com/ollie/k_means_pp, 2016-01-26.
- [107] Wegman, E. J. "Hyperdimensional data analysis using parallel coordinates", *Journal of the American Statistical Association*, vol. 85–411, Sep 1990, pp. 664–675.
- [108] Wegman, E. J.; Luo, Q. "High dimensional clustering using parallel coordinates and the grand tour", Relatório Técnico, DTIC Document, Fairfax, VA, USA, 1996, 10p.
- [109] Weiss, S. M.; Indurkha, N. "Predictive Data Mining: A Practical Guide". San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, 1 st ed., 228p.
- [110] Witten, I. H.; Frank, E. "Data Mining: Pratical Machine learning tools and techniques". San Francisco, CA, USA: Morgan Kaufmann, 2005, 2 nd ed., 560p.
- [111] Wong, P. C.; Thomas, J. "Visual analytics", *IEEE Computer Graphics and Applications*, vol. 24–5, Sep/Oct 2004, pp. 20–21.
- [112] Yuan, L.; Powell, S.; CETIS, J. "Moocs and open education: Implications for higher education", *Centre for Educational Technology e interoperability standarts White Paper*, vol. 1–1, 2013, pp. 21.

- [113] Zhong-mei, S.; Qiong-fei, Q.; Lu-qi, F. “Educational data mining and analyzing of student learning outcome from the perspective of learning experience”, *Proceedings of the 7th International Conference on Educational Data Mining*, vol. 7–1, 2014, pp. 359–360.
- [114] Zhou, Z. H. “Book review: Three perspectives of data mining”, *Artif. Intell.*, vol. 143–1, Jan 2003, pp. 139–146, reviewer-Zhou, Zhi-Hua.

APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
Faculdade de Informática/PUCRS
Avenida Ipiranga, 6681 - Prédio 32 - 90619-900 - Porto Alegre - RS
Telefone: (51) 3320-3558

Termo de Consentimento Livre e Esclarecido

Obrigada pela sua disponibilidade em responder este questionário!

Nós somos pesquisadores e estudantes da PUCRS - Pontifícia Universidade Católica do Rio Grande do Sul, e o objetivo desta pesquisa é obter informações a respeito da utilização de sistemas de auxílio no monitoramento de alunos em Ambientes Virtuais de Aprendizado - AVA.

As questões não tem a intenção de avaliar o usuário. Não existem respostas certas ou erradas, e todas as questões disponibilizadas devem ser respondidas de acordo com a sua própria visão do resultado apresentado. Você estará colaborando com a pesquisa científica no Brasil e se estiver de acordo que suas respostas sejam utilizadas para fins de pesquisa, por favor, prossiga e responda às seguintes questões.

- Concordo plenamente.
 Abaixo registro condições adicionais para este teste.

Porto Alegre, ___ de _____ de 2015

Nome do participante: _____

Assinatura do participante: _____

Pesquisador responsável: _____ - Faculdade de Informática

APÊNDICE B – QUESTIONÁRIO DE LEVANTAMENTO DE REQUISITOS

Ambientes Virtuais de Aprendizado e Monitoramento de Alunos

Esse questionário foi elaborado com o objetivo de levantar informações a respeito da utilização de sistemas para auxiliar no monitoramento de alunos em Ambientes Virtuais de Aprendizado (AVA). Sua contribuição é de extrema importância. Todos os dados informados serão tratados sem a identificação dos respondentes, com o objetivo de garantir o anonimato dos mesmos. Por favor, dedique alguns minutos respondendo as questões a seguir, de acordo com a sua opinião a respeito das mesmas.

*Obrigatório

Qual a sua idade? *

marque somente uma opção

- Até 30 anos
- De 31 a 40 anos
- De 41 a 50 anos
- Mais de 50 anos

Qual a sua área de formação? *

- Tecnologia da Informação
- Educação
- Matemática
- Administração/Contabilidade/Economia
- Saúde
- Engenharia
- Outro:

Qual o seu sexo? *

marque somente uma opção

- Masculino
- Feminino

Qual a sua titulação mais alta? *

marque somente uma opção

- Graduado
- Especialista
- Mestre
- Doutor
- Pós-Doutor

Há quanto tempo atua/atuou ministrando disciplinas na modalidade EAD? *

em anos

Sistemas

Qual(is) destes ambientes já utilizou? *

marque tantos quantos já utilizastes

Moodle

TelEduc

BlackBoard

Outro:

Qual(is) destes ambientes utiliza hoje, nos cursos que ministra? *

marque tantos quantos utiliza

Moodle

TelEduc

BlackBoard

Outro:

Utiliza(ou) algum sistema de auxílio de monitoramento de atividades dos alunos?

Sim

Não

Sobre o auxílio de monitoramento

O sistema de auxílio de monitoramento utilizado disponibiliza gráficos? *

Sim

Não

Quais?

Barra

Linha

Pizza

Mapa

Outro:

Descreva quais informações são mostradas nestes gráficos.

Sistemas

Utiliza(ou) algum sistema de predição de desempenho dos alunos? *

- Sim
 Não

Sobre o sistema de predição

O sistema de predição de desempenho disponibiliza gráficos? *

- Sim
 Não

Quais?

- Barra
 Linha
 Pizza
 Mapa
 Outro:

Descreva quais informações são mostradas nestes gráficos.

Opinião

Na sua opinião, qual(is) dos recursos abaixo são importantes para a sua avaliação do aluno? *

- Notas
 Participação em Fóruns
 Participação em Chats
 Envio de Tarefas
 Frequencia de acesso ao AVA
 Tempo de navegação no AVA
 Quiz
 Wiki
 Outro:

Na sua opinião, qual(is) dos elementos abaixo o sistema deveria possuir para lhe auxiliar em seu trabalho? *

- Um elemento gráfico que represente a participação do aluno em cada um dos recursos apresentados na questão anterior
 Um elemento gráfico que represente toda a participação do aluno, unindo cada um dos recursos apresentados na questão anterior

Gostaria que fossem disponibilizadas ferramentas gráficas para avaliação e/ou predição do desempenho do aluno? Descreva.

Enviar

APÊNDICE C – TELAS APRESENTADAS NO LEVANTAMENTO DE REQUISITOS



Figura C.1: Tela Inicial.



Figura C.2: Linha clicada.



Figura C.3: Ordenação por coluna.



Figura C.4: Filtro por recurso.



Figura C.5: Filtro por aluno.



Figura C.6: Mouse sobre as barras.

APÊNDICE D – TELAS CAPTURADAS DURANTE A AVALIAÇÃO DA DISCIPLINA

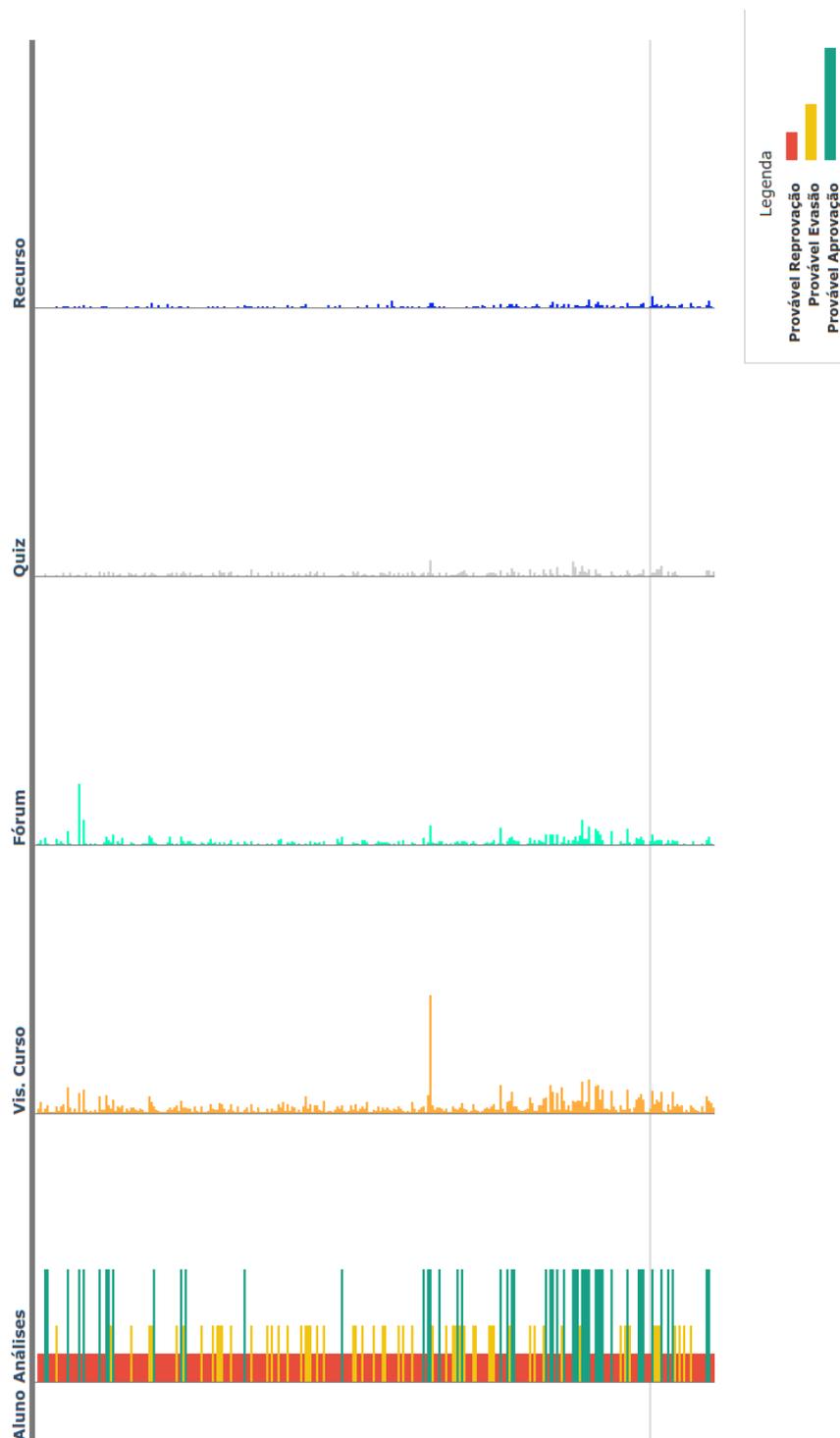


Figura D.1: Tela da primeira semana.

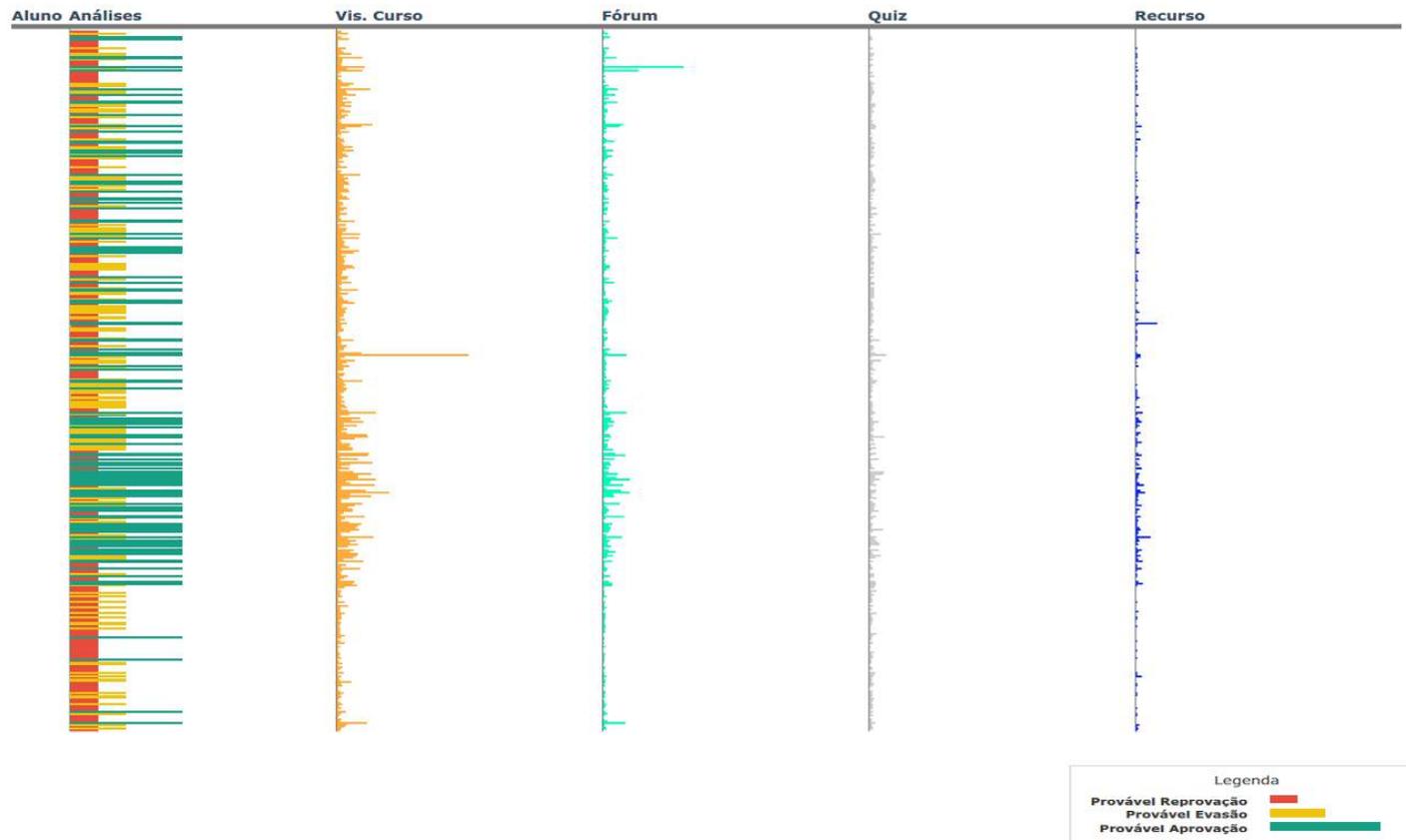


Figura D.2: Tela da segunda semana.

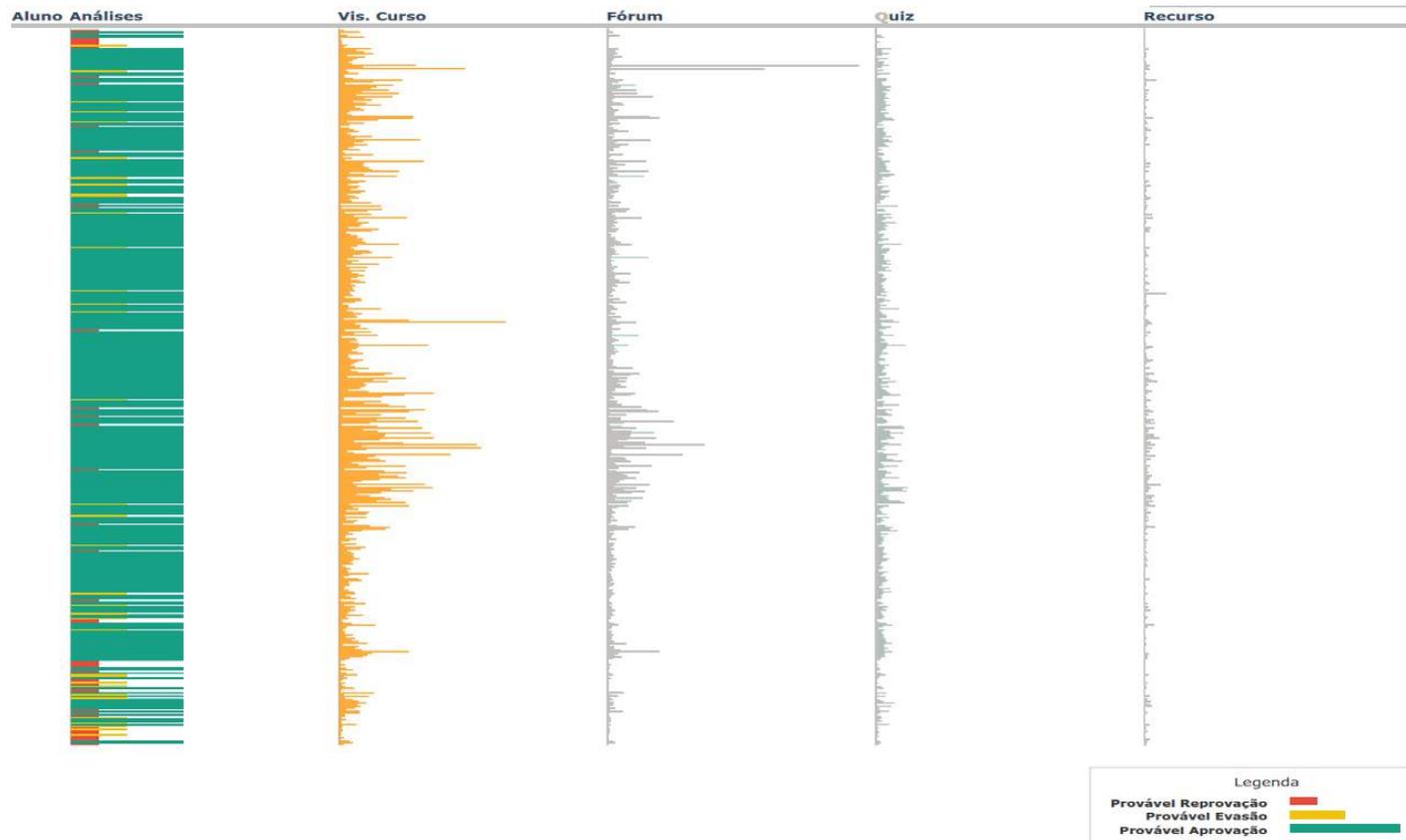


Figura D.3: Tela da terceira semana.

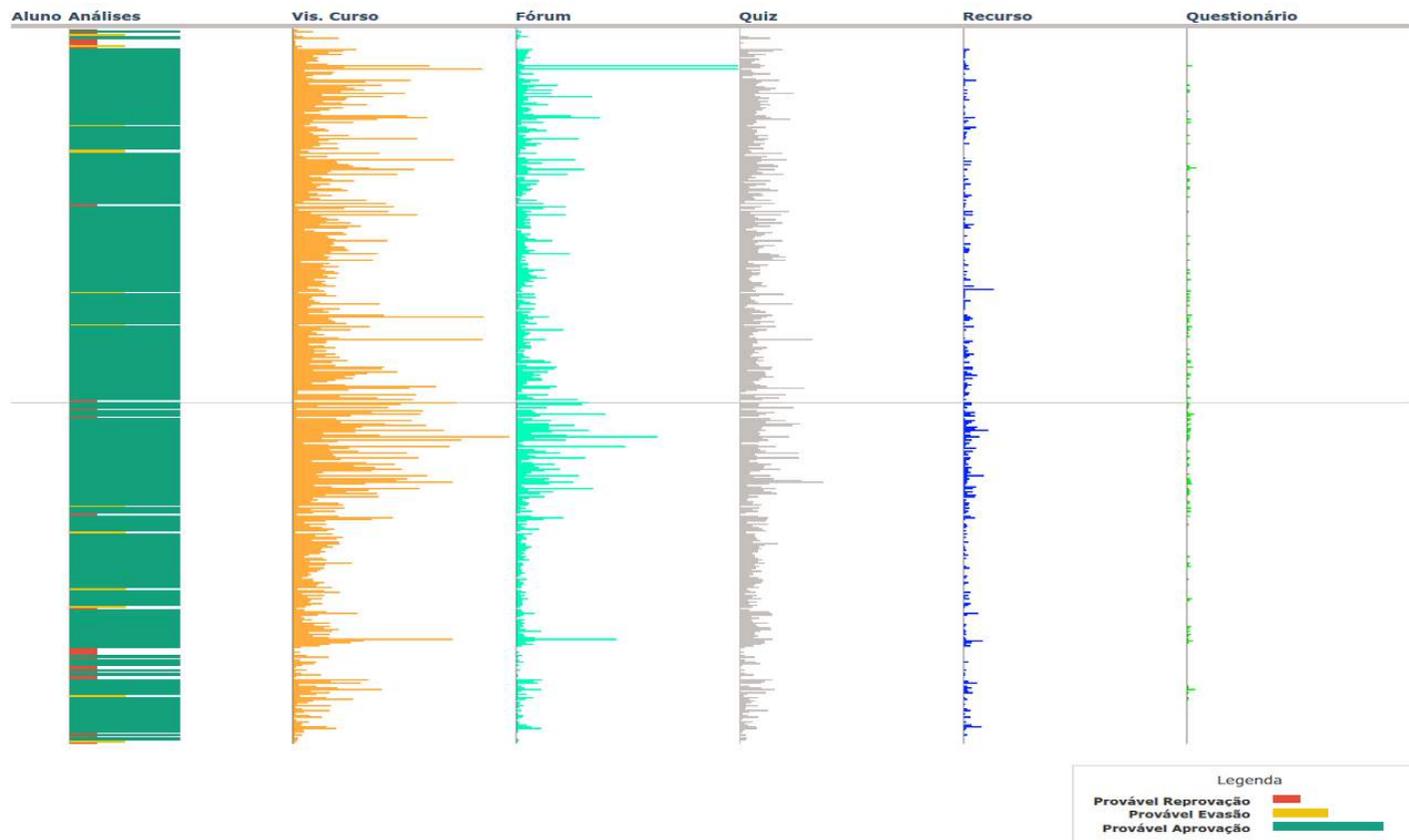


Figura D.4: Tela da quarta semana.