



ANA CAROLINA LE SENECHAL

**ANÁLISE E PRÉ-PROCESSAMENTO DE
DADOS UTILIZANDO TÉCNICAS DE
MINERAÇÃO DE DADOS EDUCACIONAIS
PARA O MOODLE**

LAVRAS – MG

2013

ANA CAROLINA LE SENECHAL

**ANÁLISE E PRÉ-PROCESSAMENTO DE DADOS UTILIZANDO
TÉCNICAS DE MINERAÇÃO DE DADOS EDUCACIONAIS PARA O
MOODLE**

Monografia apresentada ao Colegiado do Curso de
Ciência da Computação, para a obtenção do título
de Bacharel em Ciência da Computação.

Orientador

Prof. Ms. Eric Fernandes de Mello Araújo

LAVRAS – MG

2013

ANA CAROLINA LE SENECHAL

**ANÁLISE E PRÉ-PROCESSAMENTO DE DADOS UTILIZANDO
TÉCNICAS DE MINERAÇÃO DE DADOS EDUCACIONAIS PARA O
MOODLE**

Monografia apresentada ao Colegiado do Curso de
Ciência da Computação, para a obtenção do título
de Bacharel em Ciência da Computação.

APROVADA em 26 de agosto de 2013.

Prof. Dra. Ana Paula Piovesan Melchiori

DCC/UFLA

Prof. Ms. Lívia Naiara de Andrade

DCC/UFLA

Prof. Dra. Marluce Rodrigues Pereira

DCC/UFLA



Prof. Ms. Eric Fernandes de Mello Araújo
(Orientador)

LAVRAS – MG

2013

Dedico esta vitória minha mãe, a meu “papito”, a minha avó Amnerys, a minha avó Rita, à meu avô Antonio, que acompanhando de perto ou de longe, sabendo ou não tendo ideia do que eu faço, sempre me apoiaram e não me deixaram desistir! Amo vocês!

AGRADECIMENTOS

Agradeço à Deus, pela força quando o desespero me dominava, pela força para aguentar os percalços durante este período. Agradeço à Deus, também, por ter colocado pessoas tão maravilhosas em minha vida, que com seu amor incondicional não deixaram a minha peteca cair. Esta vitória é sua, mãe, pelo imenso amor, carinho, paciência, dedicação e esforços sem fim para que eu concluísse a graduação. Por pior que fossem os problemas, sempre me acalmava e dizia: Calma, vai dar tudo certo, confie em você. Esta vitória é nossa! Amo você, mãe! À meu pai, que olha por mim sempre. Porque fiz esse curso? Porque, desde criança, ele fez despertar um encantamento por computadores que viraria minha profissão e tudo que amo fazer! Amo você, “papito”! Saudades eternas! À minha avó Rita, minha avó Amnerys, meu avô Antonio e meu tio Airton, por todo o apoio e carinho, sempre dispostos a me ajudar! Obrigada, amo vocês! À minhas amigas-irmãs, Isabela e Jéssica, pelo IMENSO apoio e paciência, sei que quase coloquei vocês doidas, obrigada por tudo sempre! Amo vocês! À meu namorado, Moizes, pelo amor, força, paciência e por fazer com que eu acreditasse que era capaz, eu consegui, nós conseguimos! Te amo, obrigada por tudo! À meu orientador, Eric, pela imensa paciência, obrigada! À todos, amigos, familiares, professores do DCC/UFLA, Marina e demais companheiras de república, a todos que acompanharam de perto esta importante etapa! Essa vitória é nossa!

“Quando eu deixei de olhar tão ansiosamente para o que me faltava e passei a olhar com gentileza para o que eu tinha, descobri que, de verdade, há muito mais a agradecer do que a pedir. Tanto, que às vezes, quando lembro, eu me comovo. Pelo que há, mas também por conseguir ver.” (Ana Jácomo)

RESUMO

Neste trabalho é realizado o pré-processamento de dados em um ambiente virtual de aprendizagem que permita identificar perfis de aprendizagem em um curso *online*. O ambiente virtual em questão é o ambiente *Moodle*, o qual possui recursos e atividades que permitem a criação e aplicação de cursos *online*. É mostrado o planejamento, criação e execução de um curso preparatório para a Olimpíada Brasileira de Informática. As características do curso são mostradas, em que o mesmo possui formato de vídeo-aulas e simulados referentes ao conteúdo visualizado. O curso preparatório foi disponibilizado *online* e aplicado em duas escolas municipais de Lavras, em que o ambiente *Moodle* armazenou as informações referentes às atividades dos usuários no curso. Será mostrada a utilização de técnicas de mineração de dados educacionais para analisar as informações armazenadas no banco de dados do *Moodle*. Foi realizada uma análise minuciosa das informações armazenadas na base de dados, representando as etapas percorridas e o processo de refinamento até a obtenção de uma tabela única, o que representa o resultado do processo de pré-processamento de dados. Tal tabela é chamada de tabela de sumarização e contém o resumo das informações referentes às atividades realizadas durante o curso. Os dados resultantes, então, representam todo o processo de análise e refinamento das informações, de forma que tais dados possam ser utilizados para avaliar o aprendizado e identificar os perfis de aprendizagem presentes em um curso *online* preparatório para a Olimpíada Brasileira de Informática.

Palavras-Chave: ambiente virtual de aprendizagem; *Moodle*; mineração de dados educacionais; pré-processamento de dados; Olimpíada Brasileira de Informática

ABSTRACT

In this work is carried out the preprocessing of data in a virtual learning environment which identifies learning profiles in an online course. The virtual environment in question is the Moodle environment, which has resources and activities that enable the creation and implementation of online courses. Shown is the planning, creation and implementation of a preparatory course for the Brazilian Informatics Olympiad. The characteristics of current are shown in which has the same format and simulated video lessons regarding viewed content. The preparatory course is available online and applied in two schools of Lavras in the Moodle environment stored information related to users activities on this course. It will show the use of data mining techniques to analyze educational information stored in the Moodle database. We performed a thorough analysis of the information stored in the database, representing the steps taken and the refinement process to obtain a single table, which is the result of the pre-processing of data. Such a table is called table summarization and contains summary information regarding the activities carried out during the course. The resulting data will then represent the entire process of analysis and refinement of information, so that such data can be used to assess learning and identifying the learning profiles present in an online course in preparation for the Olympiad of Informatics.

Keywords: virtual learning environment; *Moodle*; educational data mining; preprocessing of data; Olympiad of Informatics

SUMÁRIO

1	Introdução	17
1.1	Motivação e Justificativa	19
1.2	Objetivos	20
1.3	Organização do trabalho	21
2	Referencial Teórico	22
2.1	Ambientes Virtuais de Aprendizagem	22
2.2	Perfis de Aprendizagem	24
2.3	Formas de avaliação <i>online</i>	29
2.4	Mineração de Dados	30
2.5	Mineração de Dados Educacionais	32
2.6	Estratégia de mineração no Moodle	34
2.6.1	Redes neurais artificiais	37
2.6.2	Rede neural de Kohonen	39
3	Metodologia	42
3.1	Tipo da pesquisa	42
3.2	Procedimentos Metodológicos	43
3.3	Estudo, instalação e configuração do <i>Moodle</i>	45
3.4	Olimpíada Brasileira de Informática	48
3.5	Aprovação do projeto na COEP	49

3.6	Montagem do curso preparatório para a OBI	50
3.7	Coleta de Dados	52
3.8	Ferramentas utilizadas	53
3.8.1	Gismo	54
3.8.2	Weka	55
3.9	Pré-processamento	56
4	Resultados	64
4.1	Resultados - Ferramenta GISMO	64
4.2	Resultados - Ferramenta WEKA	68
4.3	Resultados - Questionário Índice de Estilos de Aprendizagem (ILS)	80
4.4	O Pré-processamento de dados sob a perspectiva dos estilos de aprendizagem	87
4.4.1	Resultados - Estilos de Aprendizagem Sequencial/Global	90
5	Conclusão	93
A	Apêndice	99
A.1	Apêndice 1	99
A.2	Apêndice 2	102
A.3	Apêndice 3	105
A.4	Apêndice 4	108
A.5	Apêndice 5	110

A.6 Apêndice 6	114
--------------------------	-----

LISTA DE FIGURAS

2.1	Representa a escala do índice ILS. As 3 (três) cores representam os intervalos 1-3, 5-7 e 9-11.	28
2.2	Figura adaptada de (HAN; KAMBER, 2000). Representa as etapas necessárias para realizar o processo de KDD.	31
2.3	Figura retirada de (ROMERO; VENTURA; GARCIA, 2008). Apresenta um exemplo de <i>log</i> mostrado no <i>Moodle</i>	35
2.4	Figura retirada de (PIAZZA <i>et al.</i> , 2011). Apresenta o resultado da aplicação de uma técnica de <i>clusterização</i>	38
2.5	Figura retirada de (SEN; UÇAR; DELEN, 2012). Representa uma RNA do tipo Feedforward.	39
2.6	Figura retirada de (CARDON; MÜLLER; NAVAU, 1994). Apresenta a topologia de uma rede neural de Kohonen.	41
3.1	A figura mostra a estrutura modular do curso. Apresenta a visão geral de como são os módulos do curso.	51
3.2	A figura mostra a estrutura de um dos módulos. Apresenta o módulo do vídeo e o questionário com os exercícios.	52
3.3	A figura mostra o módulo que contém o vídeo. O módulo possui um player que permite assistir ao vídeo no ambiente.	52
3.4	A figura mostra um dos questionários realizados pelos alunos no curso.	53
3.5	Figura retirada de (ROMERO; VENTURA; GARCIA, 2008) representando os acessos a um curso no <i>Moodle</i>	54

3.6	Exemplo de arquivo ARFF. Apresenta as informações da seção de cabeçalho e da seção de dados.	56
4.1	Escola X - Figura gerada pelo GISMO, representa as notas obtidas nos exercícios.	65
4.2	Escola X - Figura gerada pelo GISMO, representa os acessos de cada aluno ao curso.	66
4.3	Escola X - Figura gerada pelo GISMO, representa os acessos gerais dos alunos ao curso.	67
4.4	Escola Y - Figura gerada pelo GISMO, representa as notas obtidas nos exercícios.	67
4.5	Escola Y - Figura gerada pelo GISMO, representa os acessos de cada aluno ao curso.	68
4.6	Escola Y - Figura gerada pelo GISMO, representa os acessos gerais dos alunos ao curso.	68
4.7	Figura retirada do apêndice A.1. Apresenta os atributos da tabela de sumarização.	69
4.8	Figura retirada do apêndice A.1. Mostra o número de alunos inscritos no <i>Moodle</i> (variável <i>Instances</i>) e o número de atributos para a análise (variável <i>Attributes</i>)	70
4.9	Figura retirada do apêndice A.1. Mostra os agrupamentos obtidos pela RNA de Kohonen.	70
4.10	Figura retirada do apêndice A.1. Mostra que os alunos do agrupamento 0 realizaram, em média, até o segundo módulo.	71

4.11	Figura retirada do apêndice A.1. Mostra as porcentagens médias de questionários finalizados pelos alunos dos agrupamentos 1 e 3.	72
4.12	Figura retirada do apêndice A.1. Mostra a porcentagem média obtida pelos alunos do agrupamento 2.	72
4.13	Figura retirada do apêndice A.2. Mostra o número de alunos que terminaram o curso (variável <i>Instances</i>) e o número de atributos para a análise (variável <i>Attibutes</i>)	73
4.14	Figura retirada do apêndice A.2. Mostra os agrupamentos obtidos pela RNA de Kohonen.	73
4.15	Figura retirada do apêndice A.2. Mostra a porcentagem média de questionários finalizados pelos alunos do agrupamento 2.	73
4.16	Figura retirada do apêndice A.2. Mostra as notas médias obtidas pelos alunos do agrupamento 1.	74
4.17	Figura retirada do apêndice A.2. Mostra as notas mínimas e máximas obtidas pelos alunos do agrupamento 2.	75
4.18	Figura retirada do apêndice A.2. Mostra as notas mínimas e máximas obtidas pelos alunos do agrupamento 3.	76
4.19	Figura retirada do apêndice A.2. Apresenta o percentual de questionários finalizados pelo alunos do agrupamento 3.	77
4.20	Figura retirada do apêndice A.2. Mostra as notas mínimas e máximas obtidas pelos alunos do agrupamento 3.	78
4.21	Figura retirada do apêndice A.3. Apresenta os agrupamentos obtidos pelo algoritmo da rede neural de Kohonen.	90

4.22	Figura retirada do apêndice A.3. Contém os atributos que foram considerados para a análise.	91
4.23	Figura retirada do apêndice A.3. Apresenta o número de instâncias e atributos considerados na análise.	91
4.24	Figura retirada do apêndice A.3. Apresenta os valores de desvios dos progressos alcançados nos módulos.	92

LISTA DE TABELAS

3.1	Grupos de tabelas do banco de dados do <i>Moodle</i>	45
3.2	Tabelas importantes para os propósitos do estudo	57
3.3	Campos das tabelas do banco de dados do <i>Moodle</i> que foram considerados	59
3.4	Consultas realizadas para obter a tabela de sumarização (tabela 3.5)	61
3.5	Tabela de sumarização.	62
4.1	Tabela que apresenta, respectivamente, o número de alunos inscritos e o número de alunos que terminaram o curso.	65
4.2	Tabela que apresenta o período de duração do curso nas 2 (duas) escolas parceiras.	65
4.3	Perfis de aprendizagem detectados com a aplicação do questionário ILS	80
4.4	Estilos de aprendizagem Ativo/Reflexivo encontrados com a aplicação do questionário ILS	83
4.5	Estilos de aprendizagem Sensorial/Intuitivo encontrados com a aplicação do questionário ILS	84
4.6	Estilos de aprendizagem Visual/Verbal encontrados com a aplicação do questionário ILS	85
4.7	Estilos de aprendizagem Sequencial/Global encontrados com a aplicação do questionário ILS	86
4.8	Nova tabela de sumarização.	88

1 INTRODUÇÃO

A Mineração de Dados Educacionais (do inglês *Educational Data Mining*, EDM), é uma área recente de pesquisa que tem como objetivo principal o desenvolvimento de métodos para explorar dados originados em ambientes educacionais. Apesar de recente, a área vem se estabelecendo como uma linha de pesquisa eficaz e consolidada, que possui grande potencial para melhorar a qualidade de ensino. Baker (2010)

A EDM permite utilizar informações vindas de ambientes educacionais para melhorar o processo de aprendizagem e, com a popularização da Internet, diversas estratégias educacionais têm sido aplicadas com propósito de melhorar a forma como os alunos aprendem. Uma alternativa aos ambientes educacionais presenciais, nas salas de aula, é um ambiente em que o processo de aprendizagem é *online*, utilizando a Internet e os recursos que pode oferecer. O aprendizado *online*, chamado de “*e-learning*”, veio para colaborar com o processo de aprendizagem antes feito apenas de maneira presencial.

Este trabalho é parte de um projeto que busca melhorar o processo de ensino em ambientes virtuais de aprendizagem. O mapeamento dos perfis de aprendizagem encontrados em um curso *online* deve fornecer subsídio para que o próprio ambiente virtual adapte as atividades de acordo com os perfis de aprendizagem dos alunos. Existem alunos de graduação em Sistemas de Informação e Física envolvidos em dar continuidade às atividades desenvolvidas neste trabalho.

O ambiente educacional utilizado neste trabalho é o Sistema de Gerenciamento de Aprendizagem *Moodle*, que pode ser utilizado tanto para complementar algum curso presencial, como também pode ocorrer totalmente *online*. O ambiente *Moodle* é uma plataforma de gerenciamento virtual de aprendizagem para criação de salas de aula virtuais utilizando-se de recursos comumente encontrados

no aprendizado presencial. A plataforma permite trabalhar com inúmeros tipos de atividades, desde provas e questionários, até páginas de construção colaborativa, entre outros. Pode-se também criar salas de bate-papo, em que pode ser monitorada, por exemplo, a interação entre os alunos.

O *Moodle* foi utilizado para criação de um curso preparatório para a Olimpíada Brasileira de Informática (OBI) ¹. A Sociedade Brasileira de Computação (SBC) ² é a organizadora da olimpíada e o Instituto de Computação da UNICAMP é o responsável pela organização do evento. Esta olimpíada está nos mesmos moldes de outras olimpíadas, como Matemática, Física e Astronomia. A OBI apresenta exercícios que estimulam o pensamento lógico e, em níveis mais avançados, trabalha com o conhecimento em programação. A OBI atende alunos de ensino fundamental, de sexto ao nono ano (ou equivalente), na modalidade Iniciação (nível 1 ou nível 2) e alunos de ensino médio na modalidade programação (nível júnior, nível 1 ou nível 2). O curso proposto neste trabalho dedica-se a modalidade Iniciação da prova (ambos os níveis) atendendo alunos do ensino fundamental de algumas escolas públicas municipais de Lavras.

O sistema *Moodle* armazena, em sua base de dados, informações referentes ao uso do ambiente. Especificamente, informações que referem-se as atividades realizadas pelos usuários (alunos das escolas parceiras) no curso. O curso contém vídeo aulas em que, basicamente, o aluno visualiza a vídeo aula e faz exercícios para fixação do conteúdo visto, com total liberdade para interagir com o ambiente. Todas estas informações de interação do usuário são armazenadas no banco de dados do *Moodle*. Estas informações, por si só, não podem ser utilizadas “cruas” e precisam, então, de análise e refinamento para que apenas informações relevantes sejam selecionadas.

¹<https://olimpiada.ic.unicamp.br/>, acessado em 20 de janeiro de 2013

²<http://www.sbc.org.br/>, acessado em 20 de janeiro de 2013

A mineração de dados educacionais, em suas técnicas e estratégias, representa uma maneira de analisar os dados obtidos do ambiente *Moodle* de forma a encontrar padrões nos dados de utilização. Tais padrões devem ser analisados para determinar o que representam no contexto educacional em questão.

1.1 Motivação e Justificativa

Segundo Hernandez *et al.* (2011), com o advento das tecnologias de informação e comunicação (do inglês *Information and Communication Technology*, ICT) e especialmente da Internet, um grande número de instituições estão integrando essas tecnologias a seus sistemas de aprendizagem para que os mesmos as suportem e assim melhorem o processo de aprendizagem.

O *e-learning* surgiu para completar o processo de aprendizado com intenção de melhorá-lo, auxiliando na comunicação e compreensão de informações vindas de professores para alunos e vice-versa. O aumento da popularidade do *e-learning* levou também a um aumento na identificação de fatores que levam ao sucesso da implementação das ferramentas interativas das ICT, e para atingi-lo, segundo Limayem e Cheung (2008), é necessária sua aceitação por parte dos alunos e o uso contínuo dessas tecnologias.

Uma das tecnologias ao alcance de educadores e pesquisadores são os Ambientes Virtuais de Aprendizagem (AVAs), os quais tornaram-se muito populares e adotados como forma de complementar o processo de aprendizagem. Um AVA é um ambiente virtual que permite que professores disponibilizem materiais adicionais, aproximando professores e alunos através de atividades e materiais interativos (COSTA; FRANCO, 2005).

O processo de aprendizagem *online*, além de completar as aulas tradicionais, é também um processo de aprendizagem feito exclusivamente pela Internet. O

e-learning utiliza as tecnologias disponíveis na Internet para criar um ambiente educacional. A obtenção de informações que referem-se às ações dos usuários no ambiente virtual, além de facilitar a relação professor/aluno, torna-se um importante repositório para melhorar esse e outros processos de aprendizado. A utilização de técnicas de mineração de dados educacionais permite avaliar inúmeros aspectos em um processo de aprendizagem *online*. Por exemplo, pode-se determinar qual a interação dos alunos com o ambiente, qual a interação dos alunos entre si, como está ocorrendo o progresso do aluno no curso, entre outros.

1.2 Objetivos

Este trabalho tem por objetivo realizar o pré-processamento de dados que permita identificar os perfis de aprendizagem presentes em um curso preparatório para a Olimpíada Brasileira de Informática. A partir da análise dos dados referentes a utilização do ambiente e de suas ferramentas por parte dos alunos, pretende-se detectar maneiras de melhorar o processo de aprendizagem em um curso *online*.

Especificamente, tem-se por objetivo identificar perfis de aprendizagem que podem ser encontradas em um curso preparatório para a Olimpíada Brasileira de Informática e perceber quais recursos favorecem de forma eficiente o aprendizado para esse tipo específico de evento. Foram utilizados elementos visuais (vídeos, imagens, etc.) e elementos textuais que facilitem a interação dos alunos com o ambiente. Todos os recursos e ferramentas utilizados visam mapear os perfis de aprendizagem dos alunos para que, em trabalhos futuros, o próprio ambiente adapte as atividades de acordo com os perfis encontrados.

1.3 Organização do trabalho

O capítulo 1 apresenta a mineração de dados educacionais como forma de avaliação de aprendizagem no ambiente *Moodle* e são apresentados, também, os benefícios que podem ser alcançados com a realização deste trabalho. O estado da arte deste trabalho é explicado no capítulo 2, contendo o embasamento teórico para sua realização. Este capítulo mostra o contexto e a popularização dos ambientes virtuais de aprendizagem, qual o tipo de avaliação será utilizada no ambiente *Moodle*, a mineração de dados com suas etapas e a contextualização da mineração de dados educacionais, representando a base para a metodologia deste trabalho.

O processo de realização deste trabalho é descrito no capítulo 3, com a apresentação do tipo de pesquisa realizada a partir das atividades que este trabalho envolve. É mostrado, também, o contexto da prova da OBI, o planejamento/execução do curso, as ferramentas utilizadas para atender aos objetivos do trabalho, as etapas envolvidas no pré-processamento dos dados e o questionário Índices de Estilos de Aprendizagem, utilizado para identificar os perfis de aprendizagem no curso *online*.

No capítulo 4 são mostrados os resultados obtidos com a ferramenta de visualização GISMO. É mostrado, também, o resultado da mineração de dados educacionais realizada no *software* WEKA, referente ao pré-processamento de dados dos alunos que participaram e também dos alunos que terminaram o curso. Neste capítulo, ainda, é mostrado o resultado da mineração de dados educacionais realizada no *software* WEKA, referente ao pré-processamento de dados que identifica os estilos de aprendizagem dos alunos que terminaram o curso.

2 REFERENCIAL TEÓRICO

Neste capítulo, na seção 2.1 são apresentados conceitos sobre Ambientes Virtuais de Aprendizagem, bem como o ambiente *Moodle*. A seção 2.2 apresenta a classificação de Estilos de Aprendizagem que é utilizada neste trabalho. As formas de avaliação *online* são apresentadas na seção 2.3. A seção 2.4 apresenta os conceitos e definições de mineração de dados para melhor compreensão de sua importância neste trabalho. A mineração de dados educacionais aplicada ao contexto em questão é explicada na seção 2.5 e o detalhamento das abordagens utilizadas está na seção 2.6.

2.1 Ambientes Virtuais de Aprendizagem

So *et al.* (2012) explicam que a geração atual, Genes-N, também chamada de Geração Net por Tapscott (2009), pensa e aprende de maneira diferente das gerações anteriores. Por crescer rodeada de tecnologias digitais, procura e processa informação rapidamente. A Geração Net utiliza o mais recente em tecnologia, esperando que tudo funcione corretamente e trabalhe rápido. Eles ficam entediados se não são desafiados da forma correta, mas, quando desafiados, superam de forma criativa e inovadora. A maneira que a Geração Net aprende é: fazendo. São os alunos que devem ser alcançados (TAPSCOTT, 2009).

Os AVAs são ambientes virtuais onde o processo de aprendizagem se dá, ou (1) complementando as aulas presenciais, ou (2) exclusivamente *online*, potencializando o uso das ICT, aproveitando ao máximo seus recursos e funcionalidades. No que tange a este trabalho, a abordagem (2) é utilizada.

De acordo com Moura *et al.* (2010), a escola se materializa através dos ambientes virtuais de aprendizagem (AVAs), estabelecendo uma nova cultura e lingua-

gem de comunicação e informação. Com isso, são criadas novas possibilidades de processos de ensino e aprendizagem. Os AVAs através de sua proposta de inovar o processo ensino aprendizagem, conseguem atingir, então, os jovens da Geração Net.

Devido à progressiva expansão da Internet e das condições gerais de sua acessibilidade, Gomes (2005) explica que um novo domínio de utilização das ICT na educação se afirma, e que passaram a ser consideradas uma “*extensão virtual da sala de aula presencial*”. O aprendizado *online* e, por consequência, tais tecnologias, não somente complementam o processo de aprendizado mas também são empregadas em um contexto exclusivamente virtual. Visto isso, a utilização das ICT permite criar ambientes de educação a distância, ou seja, aplicações das ICT no contexto da educação.

Muitos dos ambientes virtuais de aprendizagem criados utilizam o sistema de gerenciamento de aprendizagem (*learning management system*, do inglês LMS) *Moodle*, também conhecido como sistema de gerenciamento de cursos (*Course Management System*, do inglês CMS).

Moodle ou *Modular Object - Oriented Dynamic Learning Environment* é um software de código aberto que segue os termos da Licença Pública Geral (General Public License, do inglês GPL). De acordo com Free Software Foundation (2013) esta licença garante a liberdade de compartilhar e alterar todas as versões de um programa, garantindo que o software continue de código aberto porém sem modificar a licença original e *copyrights*.

O *Moodle* segue os princípios do construcionismo social. Esta abordagem baseia-se na interação entre o ambiente e envolvidos com o mesmo, em que o conhecimento é construído a partir do aprendizado colaborativo.

Por seguir os princípios acima, a plataforma prioriza aprendizagem, em que seus recursos e/ou funcionalidades a favorecem. Sendo assim, de acordo com

Bechara e Haguenaer (2010), a estrutura do *Moodle* é dividida em recursos e atividades. Recursos correspondem a conteúdos estáticos, como páginas de texto, páginas *web* e também arquivos para *download*. Já as atividades, podem ser classificadas como interativas ou sociais.

Intuitivamente, atividades sociais envolvem interação entre os envolvidos, ou como Bechara e Haguenaer (2010) explicam, favorecem a construção coletiva do conhecimento. Como exemplos de atividades sociais, tem-se: salas de bate-papo (do inglês *chats*), fóruns de discussão e wikis (página criada que pode ser alterada por vários usuários).

Nas atividades interativas há comunicação com um monitor ou tutor e/ou com outros envolvidos no ambiente virtual. Como exemplos de atividades interativas, tem-se tarefas (que devem ser feitas e enviadas ao ambiente para avaliação), lições e questionários. Os conteúdos disponibilizados podem ser divididos em tópicos ou seguir uma estrutura semanal. Em resumo, os recursos e atividade disponíveis no ambiente permitem acompanhar o desenvolvimento da aprendizagem (BECHARA; HAGUENAUER, 2010).

2.2 Perfis de Aprendizagem

A plataforma *Moodle*, ao ser projetada com enfoque no construcionismo social, prioriza recursos e atividades que estimulem a aprendizagem. Desta forma, permite analisar como está sendo o processo de aprendizagem dos envolvidos no ambiente virtual. De forma mais específica, pode-se classificar os perfis de aprendizagem, determinando quais alunos possuem perfis que são favoráveis à aprendizagem *online* e ir além, em que verificam-se ações para melhorar tal processo.

Para identificar um perfil de aprendizagem, deve-se partir de alguma classificação pré-existente. (FELDER; SILVERMAN, 1988) propuseram um modelo de

ensino-aprendizagem baseado na experiência de Linda Silverman em psicologia educacional e na experiência de Richard M. Felder em engenharia.

De acordo com Felder e Silverman (1988) existe incompatibilidade entre estilos de aprendizagem e de ensino, em que alunos aprendem de várias maneiras (lendo ou ouvindo, refletindo e agido, pensando logicamente ou intuitivamente, memorizando e visualizando, entre outras formas). Devido à tais incompatibilidades, tem-se como resultado alunos desinteressados, evasão de determinado curso, comportamento inadequado, etc.

Felder e Silverman (1988) explicam que aprendizado é um processo que envolve dois passos: recepção e processamento da informação. No passo de recepção, informações externas e internas estão disponíveis ao aluno e basicamente o que é feito é selecionar o que se mostra interessante e ignora-se o resto. Na fase seguinte, em que ocorre o processamento, há a memorização da informação.

Felder (2002) também desenvolveu o Índice de Estilos de Aprendizagem (Index of Learning Styles, do inglês ILS) atribuindo preferências de aprendizado em quatro escalas baseado no modelo de estilo de aprendizagem feito pelo próprio autor e por Linda Silverman no ano de 1988. Felder (2002) expõe que houve necessidade de modificação de alguns parâmetros, tais como: Exclusão da dimensão dedutiva/indutiva e alteração na dimensão visual/auditiva para dimensão visual/verbal.

As quatro escalas ou dimensões do ILS exposto acima correspondem às maneiras como as informações são recebidas/processadas e estão divididas em (FELDER; SILVERMAN, 1988):

1. **Sensorial ou Intuitiva:** Pessoas sensoriais e intuitivas aprendem observando, através dos sentidos, especulação e imaginação. Sensoriais gostam de fatos, dados e experimentos e gostam de resolver problemas através de

métodos padrão e não gostam de surpresas. Já intuitivos preferem princípios e teorias, gostam de inovação e não gostam de repetição. Sensoriais são cuidadosos porém podem ser lentos, uma vez que resolvem os problemas seguindo algum método, já intuitivos são mais rápidos, porém menos cuidadosos. Apesar das diferentes características, ambos os tipos são individuais. É importante notar também que sensoriais estão menos confortáveis com símbolos do que intuitivos, pois palavras são traduzidas em símbolos, de forma a serem representadas naturalmente para os intuitivos.

2. **Visual ou Verbal:** O modo como as pessoas recebem informação pode ser dividido em três categorias: visual (figuras, diagramas, símbolos), auditiva (sons, palavras) e cinestésica (tato, toque e cheiro). A melhor forma que os indivíduos visuais aprendem é o que vêem: figuras, diagramas, gráficos de fluxo, linhas do tempo, filmes, demonstrações. Se algo é apenas dito, rapidamente esquecem. Já indivíduos verbais, preferem escrita e explicações faladas. A questão chave é que muitos alunos são visuais enquanto professores são verbais, apresentando informações visuais de forma auditiva (como palestras) ou utilizando a representação visual de uma informação auditiva como por exemplo através do uso de transparências, estabelecendo uma das incompatibilidades encontradas entre os estilos ensino-aprendizagem.
3. **Ativa ou Refletiva:** A maneira como a informação é transformada em conhecimento apresenta duas formas: experimentação ativa e observação reflexiva. A experimentação ativa envolve, intuitivamente, fazer algo no mundo externo a partir de uma informação, como por exemplo, estabelecer discussões ou explanação ou testar a informação de alguma maneira. A partir disso, a observação reflexiva consiste, então, em examinar e manipular a informação introspectivamente. Assim, um aluno reflexivo tem mais facilidade com observação reflexiva do que experimentação ativa, visto que lida melhor com informações introspectivas. Consequentemente, alunos ativos

lidam melhor com experimentação ativa, aprendem melhor em situações que estimulem sua principal característica, de forma que em um ambiente de ensino essa característica deva ser trabalhada. Da mesma forma, atividades de aprendizagem para alunos reflexivos devem, então, trabalhar sua característica passiva.

4. **Sequencial ou Global:** As formas de apresentação de determinado conteúdo educacional segue uma ordem, em que temos dois estilos de aprendizagem: aprender sequencialmente e aprender globalmente. Alunos que aprendem sequencialmente dominam o conteúdo na ordem que é apresentado, enquanto que os alunos globais, capazes de ficar perdidos por dias, semanas, incapazes de resolver o problema mais simples e demonstrar qualquer entendimento. De repente, adquirem compreensão do conteúdo, deixando alunos sequenciais perplexos.

O Índice de Estilos de Aprendizagem é um instrumento que permite determinar como uma pessoa recebe e processa as informações. O índice apresenta 44 (quarenta e quatro) questões que tratam os diferentes estilos de aprendizagem. O questionário pode ser encontrado em Felder e Soloman (2013). O questionário também pode ser visualizado no apêndice A.6.

A partir das respostas assinaladas obtém-se a indicação dos estilos de aprendizagem que uma pessoa apresenta. Cada estilo é dividido em duas dimensões e obtém-se 4 (quatro) valores como resultado. Tais valores pertencem a uma escala, a qual indica a proximidade com uma das 4 (quatro) dimensões possíveis. Esta escala permite encontrar um estilo de aprendizagem de acordo com uma pontuação obtida. Se a pontuação estiver no intervalo de 1 a 3, o estilo de aprendizado é balanceado entre as duas dimensões. Se a pontuação está no intervalo de 5 a 7, a pessoa possui estilo de aprendizagem próximo a uma das dimensões e, caso

a pontuação esteja no intervalo de 9 a 11, há uma forte preferência por uma das dimensões.

A figura 2.2 apresenta a escala do índice ILS. Os 3 (três) intervalos estão representados nas 3 (três) cores que estão representadas. Pode-se observar que, quanto maior o valor obtido, maior é a proximidade de determinado estilo de aprendizagem.

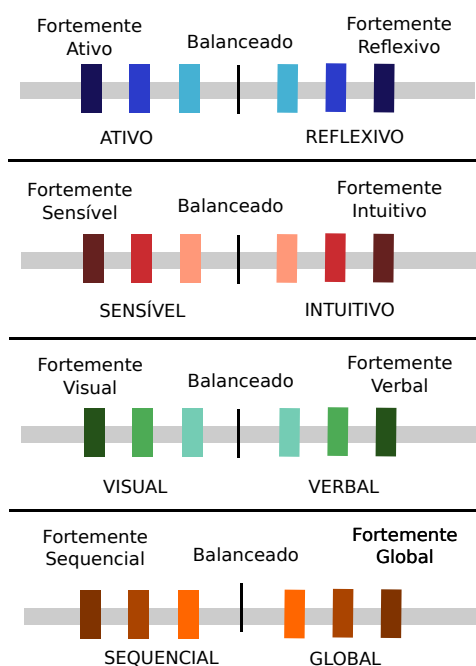


Figura 2.1: Representa a escala do índice ILS. As 3 (três) cores representam os intervalos 1-3, 5-7 e 9-11.

Em resumo, pode-se planejar as atividades de determinado ambiente educacional para que a maioria dos perfis de aprendizagem sejam alcançados. De acordo com estes perfis encontrados, pode-se aprimorar as atividades de forma a tratá-los eficientemente.

2.3 Formas de avaliação *online*

As diretrizes que surgiram acopladas à popularização do “*e-learning*” permitem avaliar o curso *online* e/ou avaliar o processo de aprendizagem no ambiente virtual. Para avaliar a aprendizagem em educação *online*, (GOMES, 2009) explica que uma função de avaliação deve ser escolhida. Essa função pode ser diagnóstica, formativa ou somativa.

A avaliação diagnóstica é aplicada no início do processo de aprendizagem para identificar aptidões e conhecimentos que colaborem com o processo em questão. Neste tipo de avaliação pode-se utilizar, por exemplo, questionários que permitam identificar as expectativas dos alunos em relação a um curso *online*. Na avaliação formativa, o foco é no processo de aprendizagem, em que é realizado o acompanhamento do aluno. Este tipo de avaliação é realizada durante a ocorrência do curso, visando obter “*feedback*” referente ao aprendizado dos alunos. Já a avaliação somativa tem como foco principal os resultados obtidos e constitui uma forma de avaliar, ao final do curso *online*, como foi o aprendizado dos alunos.

Segundo Gomes (2009), os “registros automáticos” (arquivos de *log*) armazenados no ambiente permitem analisar os dados de utilização como forma de detectar o interesse ou abandono das atividades propostas. Forner e Trevisol (2010) explicam, ainda, que os tipos de avaliação representam conceitos distintos em termos de ensino e aprendizagem. Neste trabalho, a avaliação somativa é realizada ao final do curso, quando será feita a análise dos dados armazenados. As abordagens e métodos utilizados são explicados no capítulo 3.

Hoffmann (1998) aborda os aspectos construtivistas da aprendizagem, expondo que, pelo fato do processo de avaliação ser complexo, não se pode adotar uma estratégia determinista de que tal aluno aprendeu mais que outro, pois cada

aluno tem seu tempo e estilo de aprendizagem, sendo um importante aspecto que deve ser levado em consideração como, se não o principal, essencial.

2.4 Mineração de Dados

A mineração de dados (do inglês *data mining*, DM) se tornou popular frente a necessidade de se transformar grandes quantidades de dados em informação e conhecimento, podendo ser aplicada em todas as áreas que possuem essa demanda. Os dados devem passar por um processo de tratamento para que representem informações úteis para a identificação de padrões e associações para a mineração de dados.

A mineração de dados refere-se a extração ou mineração de conhecimento a partir de grandes conjuntos de dados. De acordo com Han e Kamber (2000), o termo “data mining” é frequentemente utilizado como Descoberta de conhecimento em dados (do inglês *Knowledge Discovery from Data*, KDD).

Han e Kamber (2000) definem os passos para KDD, os quais são mostrados na figura 2.2.

A descrição detalhada de cada um dos passos de KDD pode ser visualizada como segue:

1. Limpeza dos dados: para que os dados possam ser utilizados na mineração e representem informações “utilizáveis”, os mesmos devem estar “limpos”, não apresentando inconsistências e nem serem incompletos.
2. Integração dos dados: quando se representa uma informação em um banco de dados deve-se garantir que a referência àquela informação seja a mesma em todos os repositórios em que está. O mesmo vale para dados que sejam arquivos, planilhas, entre outros. Deve-se garantir que os dados estejam

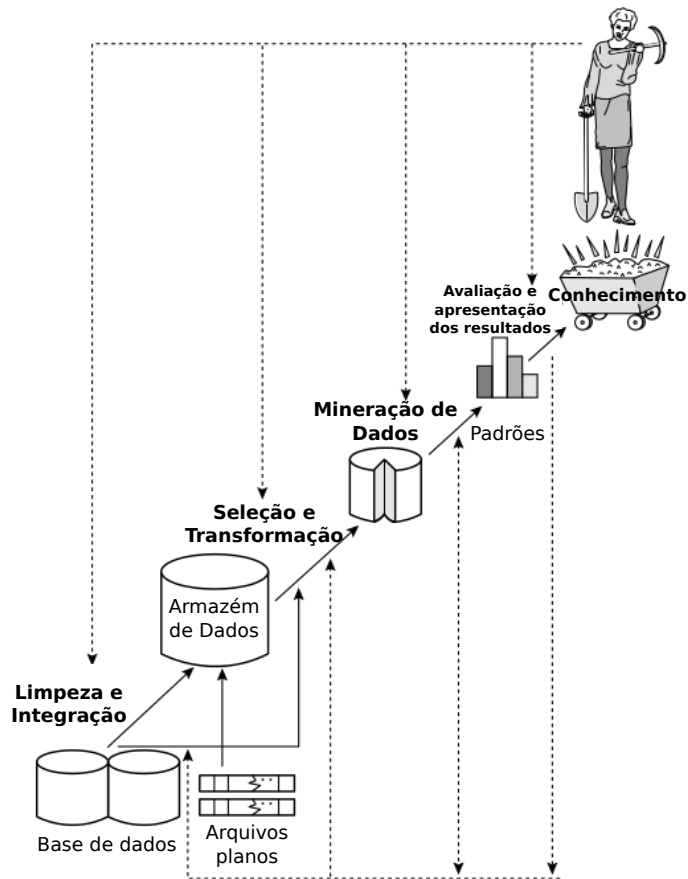


Figura 2.2: Figura adaptada de (HAN; KAMBER, 2000). Representa as etapas necessárias para realizar o processo de KDD.

representados de maneira única. Portanto, após verificar se os dados estão completos e são consistentes, deve-se garantir que estejam representados de maneira uniforme.

3. Seleção dos dados: após os passos 1 e 2, nem todos os dados que estão disponíveis são úteis para a identificação de padrões. Assim, apenas os dados relevantes para o contexto do problema são selecionados.
4. Sumarização dos dados: Como o processo de *data mining* é focado no comportamento geral dos dados, rotinas de sumarização, agregação e genera-

lização são importantes para melhorar a performance de identificação de padrões.

5. Mineração dos dados: As tarefas de *data mining* são realizadas, as quais incluem manipular os dados para que se tornem entrada do algoritmo de mineração de dados e à devida execução do algoritmo.
6. Descoberta de padrões: São selecionados os padrões relevantes para o problema em questão, avaliando sua importância contextualizada.
7. Visualização do conhecimento: Consiste em visualizar os resultados do processo de *data mining* e aplicar técnicas de representação de conhecimento apresentando o conhecimento minerado ao usuário.

2.5 Mineração de Dados Educacionais

A área Mineração de Dados Educacionais (*educational data mining*, do inglês EDM) é uma área de pesquisa considerada recente e com grande potencial para melhorar a qualidade do ensino, impactando diretamente no processo de aprendizagem (BAKER, 2010).

A Mineração de Dados Educacionais trata do desenvolvimento de métodos e técnicas para explorar dados originados em ambientes educacionais, encontrando padrões entre esses dados com o propósito de contribuir para o entendimento de como é o processo de aprendizagem e, por consequência, realizar ações que possam melhorar as práticas e materiais utilizados.

Baker, Isotani e Carvalho (2011) explicam que a EDM, além de determinar o que afeta a aprendizagem, fornece melhor compreensão dos sistemas educacionais existentes e o que pode ser feito para melhorá-los, proporcionando uma visão mais ampla e abrangente dos processos de aprendizagem.

Os métodos de EDM fornecem diversas formas para minerar dados educacionais, porém, a hierarquia dos dados deve ser considerada. É importante notar também a distinção entre dados e informação. Dados (fluxo de entrada para o processo de mineração de dados) são diferentes da saída do processo de mineração de dados (informações), em que a saída é definida de acordo com os padrões encontrados. Ou seja, é realizada a contextualização destes padrões para o problema no qual a EDM está inserida.

O direcionamento do fluxo atual de pesquisa na área de EDM, segundo Calders e Pechenizkiy (2012) está focado principalmente na mineração em sistemas tutores inteligentes e em análise dos *logs* dos LMS, os sistemas de gerenciamento de aprendizagem. Baker, Isotani e Carvalho (2011) apresentam a taxonomia das técnicas que podem ser utilizadas em mineração de dados educacionais. As técnicas correspondem a:

- Predição
 - Classificação
 - Regressão
 - Estimação de Densidade
- Agrupamento
- Mineração de Relações
 - Mineração de Regras de Associação
 - Mineração de Correlações
 - Mineração de Padrões Sequenciais
 - Mineração de Causas
- Destilação dos dados para facilitar decisões humanas

- Descobrimto com modelos

Calders e Pechenizkiy (2012) destacam, ainda, que a EDM ajuda a resolver problemas relacionados a diferentes fases do processo de aprendizado, que pode ser tanto formal (provas, testes) ou informal (*games*), quanto intencional (tutoria) ou inesperada (utilizando mídias sociais). Neste trabalho, o pré-processamento de dados é realizado com base nos arquivos de log e na base de dados do ambiente *Moodle*.

2.6 Estratégia de mineração no Moodle

A Mineração de Dados Educacionais, como o próprio nome sugere, trata de dados obtidos a partir de ambientes educacionais. Partindo disso, pode ser realizado o mapeamento das demandas de mineração de dados educacionais em problemas de *data mining*. O enfoque deste trabalho está na abordagem de Aprendizado de Máquina denominada agrupamento (clusterização).

Os sistemas de aprendizagem armazenam em arquivos de *log* as interações do usuário com o ambiente *online*. A quantidade de dados nestes arquivos é muito extensa, fato pelo qual um processo de análise manual demandaria tempo e esforço consideráveis. A figura 2.3 mostra um exemplo de *log* gerado pelo *Moodle*. Na figura, pode-se observar algumas informações que o *Moodle* mostra, tais como: data do acesso, o endereço IP que o usuário acessou o ambiente, nome do usuário, a ação realizada no ambiente e a informação referente a ação.

A proposta da área EDM é analisar os dados em ambientes educacionais para melhoria do processo de aprendizagem, tornando-se, então, uma estratégia promissora para atender aos objetivos.

Course Fullname 101 You are logged in as Admin User (Logout)

moodle » CF101 » Logs » All participants, All days

Course Fullname 101: All participants, All days (Server's local time)

Course Fullname 101 | All participants | All days | All activities | Show these logs

Displaying 78 records

Time	IP Address	Full name	Action	Information
Sun 11 February 2007, 10:00 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Sun 11 February 2007, 09:57 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz report	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz report	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz preview	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz view	5
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz preview	4
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	quiz view	4
Mon 15 January 2007, 01:44 PM	127.0.0.1	Admin User	course view	Course Fullname 101
Mon 15 January 2007, 01:43 PM	127.0.0.1	Admin User	quiz preview	3
Mon 15 January 2007, 01:43 PM	127.0.0.1	Admin User	quiz view	3

Figura 2.3: Figura retirada de (ROMERO; VENTURA; GARCIA, 2008). Apresenta um exemplo de *log* mostrado no *Moodle*.

Romero e Ventura (2010) explicam que para o sucesso da EDM é necessário que os dados apresentem a mesma estrutura, para que constituam bases de dados que possam ser disponibilizadas e utilizadas para várias pesquisas, com enfoques diferentes.

Calders e Pechenizkiy (2012) resumizam o processo de EDM em *e-learning* nos passos como a seguir:

- Coleta de dados;
- Pré-processamento dos dados;
- Aplicar *data mining*;
- Interpretação e avaliação dos resultados.

A coleta dos dados corresponde exclusivamente ao armazenamento das interações e uso do ambiente *Moodle* em arquivos de *log* (base de dados do *Moodle*).

Durante o pré-processamento, os dados são “limpos”, sumarizados e generalizados, para que os padrões desejados possam ser encontrados. Em seguida, os dados já sumarizados, apenas com informações relevantes, são transformados no formato apropriado que a entrada do algoritmo de mineração de dados requer. Esse processo é necessário para a limpeza dos dados, eliminando informações que não representem padrões úteis.

Em seguida, tem-se a fase de visualização dos resultados obtidos aplicados ao contexto da EDM. Uma das ferramentas de visualização utilizadas para visualização é o *GISMO*. Segundo Romero, Ventura e Garcia (2008), “*GISMO* fornece diferentes tipos de representações gráficas, relatórios, análise de itens, entre outros”. A partir de sua integração ao ambiente *Moodle*, esta ferramenta enriquece os resultados deste estudo quando apresenta visualizações que permitem melhor compreensão de como está o progresso dos alunos no curso. A fase de interpretação dos resultados refere-se ao contexto que cada agrupamento encontrado representa no ambiente educacional em questão.

As fases de interpretação e visualização dos resultados obtidos aplicados ao contexto da EDM permitem que as informações interpretadas e visualizadas sirvam de auxílio para tomar decisões referentes ao *Moodle* que melhorem o processo de aprendizagem no ambiente *online*.

Uma das técnicas de mineração de dados que é bastante utilizada é a *clusterização*. Antonenko, Toy e Niederhauser (2012) expõem que a *clusterização* é tipicamente utilizada como ferramenta de análise exploratória. Neste tipo de análise, os dados que tenham algum grau de associação são agrupados, de tal forma que a associação é máxima se pertencem ao mesmo grupo e mínima, em caso contrário.

Agrupamento ou *clusterização* é uma técnica de aprendizado de máquina não supervisionado. Em uma abordagem não supervisionada, os dados não possuem

rótulos pré-definidos e o objetivo do algoritmo é justamente determinar grupos ou *clusters* baseados na localidade dos dados e sua conectividade em um espaço n -dimensional (ROMERO; VENTURA; GARCIA, 2008). Primeiramente, a (dis)similaridade entre os dados é calculada para a variável de interesse, prosseguindo até que os dados não mudem de *cluster*. Ao final do processo, verifica-se a importância do *cluster* no contexto do problema.

A *clusterização* faz uma categorização dos dados similares em grupos homogêneos ou clusters, em que, de acordo com Antonenko, Toy e Niederhauser (2012), os algoritmos de *clusterização* podem ser hierárquicos e não hierárquicos.

Antonenko, Toy e Niederhauser (2012) explicam que algoritmos de *clusterização* hierárquicos são utilizados quando não há predefinição do valor K (não se sabe quantos dados vizinhos serão considerados) e também quando a quantidade de dados é grande. Os grupos ou *clusters* podem ser utilizados, por exemplo, para encontrar grupos de alunos com perfis de aprendizagem semelhantes. Estes agrupamentos, no contexto da EDM, representam conjuntos com muitos dados. O método de *clusterização* que será utilizado é a rede neural de Kohonen, sendo também conhecida como mapas auto-organizáveis de Kohonen, a qual será explicada na seção 2.6.2. Um exemplo de *clusterização* é apresentado na figura 2.4. Na figura, pode-se observar que existem 3 (três) agrupamentos e cada um está representado com uma cor (um agrupamento utiliza a cor verde, um agrupamento utiliza a cor azul e um agrupamento está representado com a cor vermelha).

2.6.1 Redes neurais artificiais

Redes neurais artificiais (RNAs) são inspiradas no funcionamento dos neurônios biológicos. São modelos matemáticos que simulam o funcionamento do cérebro humano, possuindo neurônios artificiais (que imitam o comportamento dos neurô-

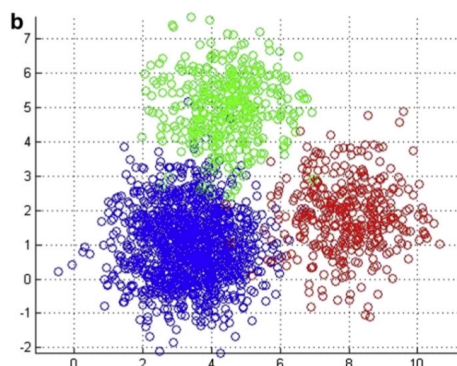


Figura 2.4: Figura retirada de (PIAZZA *et al.*, 2011). Apresenta o resultado da aplicação de uma técnica de *clusterização*.

nios humanos), bem como suas ligações (sinapses) e componentes (axônios e dendritos).

Normalmente, as RNAs são utilizadas para problemas supervisionados. Nas redes neurais *feedforward* de múltiplas camadas, um padrão é apresentado a camada de entrada da rede neural. Nas camadas intermediárias ocorre todo o processamento, e o resultado dos cálculos é repassado para as camadas de saída. A saída desejada é comparada com a saída obtida, em que o erro é calculado. O erro é retropropagado das camadas de saída até a camada de entrada, ajustando a rede para que possa se aproximar da saída desejada. O processo ocorre até satisfazer um critério de parada, que pode ser, por exemplo, atingir o menor erro desejado. Um exemplo de uma Rede Neural Feedforward pode ser visualizado na figura 2.5. Na figura pode-se observar que os neurônios de cor verde são os neurônios da camada de entrada (*input layer*). Os valores são apresentados à camada de entrada e são propagados pela rede até a camada de saída (*output layer*), representada pelos neurônios de cor vermelha. A saída obtida é comparada com a saída desejada e, assim, o erro é calculado. Este erro é retropropagado pela rede de forma que os pesos dos neurônios da rede são ajustados. Este processo ocorre até que algum critério de parada seja satisfeito. Este critério pode ser, por exemplo, atingir o menor

erro desejado, indicando, assim, que a rede aprendeu a reconhecer os padrões que lhe são apresentados.

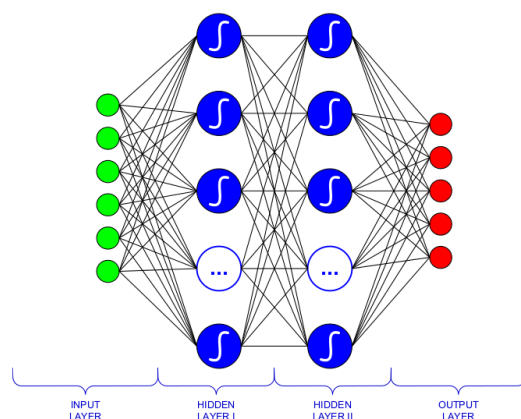


Figura 2.5: Figura retirada de (SEN; UÇAR; DELEN, 2012). Representa uma RNA do tipo Feed-forward.

2.6.2 Rede neural de Kohonen

A Rede neural de Kohonen é uma técnica de aprendizado não supervisionado que difere das redes neurais “*feedforward*” em aspectos referentes a treinamento e reconhecimento de padrões (HEATON, 2008). Também chamada de mapa de auto organização, a RNA de Kohonen difere das Redes Neurais “*feedforward backpropagation*” uma vez que é treinada em modo não supervisionado. Conforme explicado na seção 2.6, os dados apresentados não possuem classificação, de forma que a RNA de Kohonen encontra padrões nos dados e forma agrupamentos.

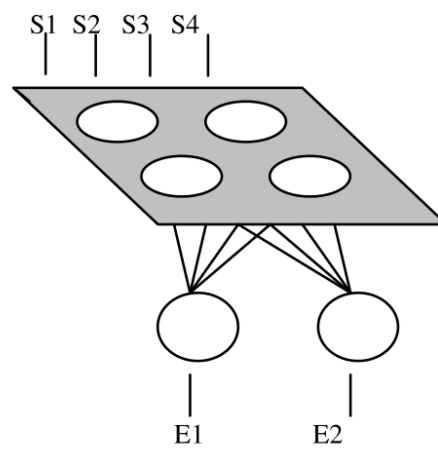
Heaton (2008) explica que o mapa auto organizável de Kohonen contém uma camada de neurônios de entrada, uma camada de neurônios de saída e não existem camadas ocultas.

O funcionamento da RNA de Kohonen é como segue: quando um padrão de entrada é apresentado a rede, utilizam-se os neurônios de entrada. Tal padrão é

constituído de números reais que devem estar normalizados, com valores entre -1 e 1. Um neurônio produz um valor, que pode ser verdadeiro ou falso. Um neurônio produz um valor verdadeiro e é escolhido como neurônio de saída. Este tipo de rede neural possui processo de treinamento competitivo, em que os neurônios são escolhidos a medida que seus valores são considerados satisfatórios. Esta decisão de qual será o neurônio vencedor envolve algumas considerações, tais como calcular a saída do neurônio e a comparação desta saída com as saídas produzidas pelos outros neurônios. Heaton (2008) ainda explica que, “para cada conjunto de treinamento um neurônio irá vencer. Este neurônio vencedor terá que ajustar seu peso, de forma que reaja mais fortemente da próxima vez.” O autor ainda destaca que “uma vez que os neurônios vencem para diferentes padrões, sua habilidade de reconhecer uma padrão em particular será aumentada.”

Da mesma forma que em outras redes neurais, uma época consiste na apresentação dos dados de entrada para a rede, em que os pesos são ajustados em cada época. O treinamento ocorre pela repetição das épocas até que o erro seja considerado aceitável e, então, o processo é encerrado. Heaton (2008) salienta que quando o erro é alterado por uma quantidade marginal, ou seja, a alteração é muito pequena, o processo de treinamento é abortado.

Na figura 2.6, retirada de Cardon, MÜLLER e Navaux (1994), pode-se observar a estrutura da rede neural de Kohonen. Neste exemplo, existem duas entradas (E_1 e E_2) e quatro saídas (S_1, S_2, S_3, S_4). Cardon, MÜLLER e Navaux (1994) explica que o número de entradas é definido de acordo com a quantidade de agrupamentos desejados e o conjunto de neurônios de saída é composto de tal forma que cada agrupamento (representando um padrão) deve estar associado a um neurônio de saída.



E_i - entrada i
S_i - saída i

Figura 2.6: Figura retirada de (CARDON; MÜLLER; NAVAUX, 1994). Apresenta a topologia de uma rede neural de Kohonen.

3 METODOLOGIA

As atividades envolvidas na execução deste trabalho definem o tipo de pesquisa que foi realizado. Tais atividades e o tipo de pesquisa associada, são mostrados na seção 3.1. Este capítulo apresenta também todo o processo de criação do curso preparatório para a OBI. A configuração do ambiente *Moodle* está detalhada na seção 3.3. As atividades de planejamento e execução do curso *online* envolvem estudar como é o evento da Olimpíada Brasileira de Informática. As informações referentes ao evento podem ser visualizadas na seção 3.4. A submissão do projeto na COEP, o planejamento do curso e a montagem das vídeo aulas são apresentados, respectivamente, na seção 3.5 e 3.6. A seção 3.7 mostra como foi realizada a coleta de dados no ambiente *Moodle*.

3.1 Tipo da pesquisa

Este projeto trata-se de uma pesquisa tecnológica, analisando o pré-processamento de dados de um curso *online* a partir de um algoritmo de mineração de dados, a fim de melhorar o processo de aprendizagem. Em relação a seus objetivos, trata-se tanto de pesquisa exploratória quanto de pesquisa descritiva. Este projeto é exploratório pois estão disponíveis estudos sobre a utilização de algoritmos de mineração de dados para ambientes educacionais. Este projeto é, ainda, descritivo, uma vez que analisará as características de utilização do ambiente *online*. Este projeto é também diagnóstico, uma vez que utiliza um instrumento que identifica as preferências de estilos de aprendizagem nos alunos que terminaram o curso proposto. Os procedimentos utilizados fazem parte da pesquisa experimental, utilizando a ferramenta *Moodle* e minerando os dados oriundos de sua utilização através da utilização de algoritmos de mineração de dados educacionais. Este projeto consiste

de uma pesquisa em campo, onde os dados armazenados retratam a real utilização do ambiente.

3.2 Procedimentos Metodológicos

Para realizar as atividades do projeto, foi necessário entender o funcionamento do ambiente *Moodle*, incluindo aspectos de instalação e configuração, além de explorar e entender o banco de dados do ambiente. Antes de aplicar as técnicas de mineração de dados foi realizado o pré-processamento dos dados com intenção de torná-los representativos. As atividades envolvidas na realização deste trabalho são descritas nos passos a seguir. São elas:

1. Estudo, instalação e configuração do *Moodle*: Estudo sobre o ambiente *Moodle* para entender quais requisitos são necessários a sua instalação e configuração. Os detalhes destas atividades podem ser encontrados na seção 3.3.
2. Estudo sobre o evento Olimpíada Brasileira de Informática: Foi realizado um estudo sobre o evento da OBI para conhecer o funcionamento tanto do evento quanto dos conteúdos abordados pela prova. Os detalhes do evento, bem como os detalhes referentes a público alvo desta olimpíada podem ser encontrados na seção 3.4.
3. Coleta de dados: O processo descrito nas seções 3.3, 3.5 e 3.6 constituiu o planejamento, preparação e execução do curso nas escolas parceiras. Os dados referentes ao uso do ambiente foram armazenados na base de dados do *Moodle* e esta foi “populada” a medida que os alunos utilizavam o curso. Os detalhes do processo de coleta de dados são apresentados na seção 3.7.

4. Pré-processamento de dados: O estudo realizado na etapa 1 contribuiu para a realização do pré-processamento. O estudo sobre o banco de dados da plataforma *Moodle* possibilitou refinar quais informações da base de dados seriam realmente necessárias a este trabalho. Os detalhes do processo de refinamento até a obtenção de uma tabela de sumarização são apresentados na seção 3.9.
5. Aplicação das técnicas de mineração de dados: A etapa 3 forneceu como resultado uma tabela de sumarização, que contém os dados quantitativos referentes aos progressos obtidos pelos alunos no curso *online* preparatório para a OBI. As ferramentas que auxiliaram para a obtenção dos resultados foram a ferramenta de visualização GISMO e o *software* WEKA (ver seção 3.8). Os dados da tabela de sumarização foram, então, convertidos para o formato de arquivo ARFF, requerido pelo WEKA. Os detalhes referentes a aplicação das técnicas de mineração de dados são apresentados na seção 4.2.
6. Análise dos perfis de aprendizagem: Para analisar quais perfis de aprendizagem são encontrados em um curso preparatório para a OBI, primeiramente, deve-se utilizar algum instrumento que consiga detectar os estilos de aprendizagem preferenciais para os alunos participantes do curso. Neste trabalho, o instrumento utilizado é o Questionário Índice de Estilos de Aprendizagem (ILS) (ver seção 2.2). Este questionário foi preenchido por todos os 43 alunos das 2 (duas) escolas parceiras e os estilos de aprendizagem encontrados a partir das respostas assinaladas são apresentados na seção 4.3. Os resultados referentes a análise dos perfis de aprendizagem podem ser encontrados na seção 4.4.1.

3.3 Estudo, instalação e configuração do *Moodle*

Para o funcionamento da plataforma *Moodle* é necessário instalar o conjunto Apache + MySQL + PHP (LAMP) e, inicialmente, foi instalado e configurado em servidor local. O próximo passo consistiu no *download* e instalação da plataforma ¹ no servidor local mencionado e em sua configuração (criação dos cursos, configurações padrão, etc.). O nome de cada curso seguiu o padrão “Curso Olimpíada Brasileira de Informática - Nome da Escola”, com a criação de dois cursos, um para cada escola participante.

A plataforma *Moodle* possui um banco de dados extenso e com muitas tabelas, o que demanda estudo para conhecer as tabelas e a estrutura da base de dados.

Banco de dados do *Moodle*

A plataforma *Moodle* armazena as informações em um banco de dados que contém aproximadamente 200 (duzentas) tabelas. Este número pode sofrer alterações caso novas funcionalidades sejam acrescentadas ao ambiente. As tabelas do banco de dados do *Moodle* são divididas em grupos e tal divisão pode ser observada na tabela 3.1. O nome de cada tabela do *Moodle* sugere a qual grupo ela está ligada e seguem o padrão: mdl_(nome_do_grupo).

Tabela 3.1: Grupos de tabelas do banco de dados do *Moodle*

Grupo	Descrição
Grupo e agrupamento	Grupos e agrupamentos são conjuntos de grupos relacionados em uso ou em lógica funcional. Seu uso é opcional.
Sistema de <i>logs</i>	<i>Logs</i> de utilização do ambiente <i>Moodle</i> .

Continua na próxima página.

¹<http://download.moodle.org/>, acessado em 20 de junho de 2013

Tabela 3.1 – *Grupos de tabelas do banco de dados do Moodle*

Grupo	Descrição
Sistema de blocos	Blocos são itens que podem ser adicionados na esquerda, direita ou no centro de qualquer página no <i>Moodle</i> .
Eventos	Um evento é quando algo considerado “interessante” acontece no <i>Moodle</i> e é emitido um alerta ao sistema.
<i>Backup</i> e restauração	Armazenam informações referentes a cópia de um curso, restauração de um curso e sobre a cópia de todo o site.
Estatísticas	Tabelas responsáveis por guardar as estatísticas sobre a atividade no ambiente.
<i>Tags</i>	Tabelas que contêm os rótulos das tabelas.
Notas	Tabelas que armazenam informações sobre os itens que valem nota.
Livro de Notas (<i>gradebook</i>)	Tabelas que representam a localização central em que as notas dos alunos são armazenadas e mostradas.
Banco de Questões e Mecanismos de Questões	As tabelas do banco de questões armazenam as definições de questões, organizadas em categorias. As tabelas de mecanismos de questões são responsáveis por interpretar as questões e processar as respostas dos alunos (submissão ou submissões do usuário).

Continua na próxima página.

Tabela 3.1 – *Grupos de tabelas do banco de dados do Moodle*

Grupo	Descrição
Sistema de mensagens	Tabelas que possuem a definição de quais mensagens de saída estão habilitadas, quais saídas podem ou não podem e quais poderiam ser utilizadas para quais mensagens, configurações de usuário que definem quais mensagens o usuário irá receber
Rede do <i>Moodle</i>	Tabelas que armazenam informações relacionadas a rede a qual o <i>Moodle</i> está conectada.
<i>Caching</i>	Tabelas que contém o armazenamento de dados em cache.
Miscelânea	Tabelas que contém informações sobre fuso horário e as sessões no <i>Moodle</i> .
Módulos de atividade	Tabelas que guardam informações sobre os módulos de atividade, tais como: provas, bate-papo, questionários, atividades com prazo, <i>feedback</i> , fórum, glossário, SCORM, WIKI e workshop.
Plugins de autenticação	Tabelas que possuem informações referentes a autenticação no <i>Moodle</i> .
Blocos	Tabelas que armazenam informações sobre os blocos existentes no <i>Moodle</i> .
Plugins de ambiente	Tabelas que armazenam informações referentes a situação e inscrição dos usuários.
Tipos de Questões	Tabelas que armazenam informações extras sobre as questões.

A estrutura do banco de dados é definida, editada e atualizada utilizando o sistema XMLDB ². Este sistema compõe a camada abstrata do banco de dados do *Moodle* que permite interação com o ambiente e o acesso à sua base de dados.

3.4 Olimpíada Brasileira de Informática

A Olimpíada Brasileira de Informática (OBI) é uma competição organizada pela Sociedade Brasileira de Computação (SBC) nos moldes das outras olimpíadas científicas brasileiras, como Matemática, Física e Astronomia, em que o Instituto de Computação da UNICAMP é o responsável pela organização do evento.

O objetivo desta olimpíada é despertar nos alunos o interesse pela Ciência da Computação, através de uma atividade que envolve desafio, engenhosidade e uma saudável dose de competição. A OBI possui duas modalidades: modalidade iniciação e modalidade programação.

A modalidade iniciação possui, por sua vez, dois níveis: nível 1 e nível 2. O nível 1 é para alunos até a sexta série do ensino fundamental ou equivalente e o nível 2 é para alunos até a oitava série do ensino fundamental ou equivalente.

A modalidade programação possui três níveis: nível júnior, nível 1 e nível 2. Esta modalidade destina-se a alunos do ensino fundamental que já tenham algum conhecimento em programação e a alunos do ensino médio. Os alunos do ensino fundamental podem participar caso já tenham algum conhecimento básico em programação e encaixam-se no nível júnior. Os alunos do ensino médio, dependendo de seu ano escolar, participarão ou do nível 1 ou do nível 2. O nível 1 desta modalidade abrange alunos até o segundo ano do ensino médio e o nível 2 abrange os

²http://docs.moodle.org/dev/XMLDB_Documentation, acessado em 20 de junho de 2013

alunos até o terceiro ano do ensino médio. A OBI também permite a participação de alunos que tenham cursado o ensino médio até dezembro do ano anterior.

A modalidade iniciação é feita utilizando lápis e papel, em que os alunos resolvem problemas de lógica, sem o uso do computador. A modalidade programação para alunos do ensino fundamental e para os competidores do nível 1 exige um conhecimento básico em programação, como comando *if*), estruturas de repetição (*while, for*) e estruturas de dados (vetores). Para os alunos participantes do nível 2, as questões exigem algum conhecimento em programação, em que as mesmas possuem vários níveis de dificuldade. Existem tarefas mais fáceis, que exigem um conhecimento básico em programação, e existem tarefas mais complexas, que exigem conhecimento mais avançado.

Os alunos competem individualmente e cada aluno pode participar de apenas uma modalidade. Todos os alunos recebem certificado de participação independente de sua colocação. Os melhores classificados são convidados a realizar cursos de programação no Instituto de Computação da UNICAMP.

3.5 Aprovação do projeto na COEP

Foi firmada uma parceria com três escolas municipais de Lavras, em que os alunos interessados realizariam as atividades nos laboratórios de informática de suas escolas. Este projeto, então, necessitou de aprovação do Comitê de Ética e Pesquisa em Seres Humanos (COEP) da UFLA³. Todos os projetos que necessitem de aprovação do comitê devem ser submetidos à Plataforma Brasil⁴. O número CAAE do projeto é 07514413.8.0000.5148 (identificação do projeto na Plataforma Brasil). Os documentos anexados a este trabalho são: Termo de Consentimento de Uso de Imagem, em que os alunos que gravaram as vídeo-aulas autorizam que sua

³http://www.prp.ufla.br/site/?page_id=83, acessado em 26 de agosto de 2013

⁴<http://www.saude.gov.br/plataformabrasil>, acessado em 18 de agosto de 2013

imagem seja disponibilizada no curso online (este documento pode ser visualizado no apêndice A.4) e o Termo de Consentimento Livre e Esclarecido (TCLE), que o responsável pelo menor o autoriza a participar do curso (este documento pode ser visualizado no apêndice A.5). O projeto foi aprovado, sem restrições, e foi dado início ao planejamento e execução do curso.

3.6 Montagem do curso preparatório para a OBI

Todo o conteúdo ministrado no curso está baseado em Martins (2011). O livro apresenta os capítulos com muitos exemplos e resoluções de exercício, além de simulados que abrangem todo o conteúdo.

Inicialmente, foi feito o planejamento de execução do curso, com definição da estrutura e a sequência de apresentação dos conteúdos. Após estas definições, as vídeo aulas foram gravadas por dois alunos de graduação dos cursos de Sistemas de Informação e Ciência da Computação. As vídeo aulas foram editadas e divididas em várias partes devido ao fato do ambiente *Moodle* não aceitar arquivos de tamanho grande.

A estruturação do curso é como segue: um módulo de conteúdo por semana, formado por uma vídeo aula e exercícios de fixação. As vídeo aulas são pacotes *Scorm*, configurados para possuir um *player* que fosse exibido dentro do ambiente virtual. Para montagem dos questionários foi utilizado o recurso *Quiz* do ambiente *Moodle*, em que cada questionário contém questões de anos anteriores da OBI, reforçando o conteúdo apresentado no vídeo. O curso possui 6 (seis) módulos no total. Após a montagem no servidor local, foi realizada a transferência do curso para o seguinte endereço: <http://pesquisa.dcc.ufla.br/olimpiada>.

A figura 3.1 mostra a estrutura modular do curso de maneira geral. Pode-se observar que os módulos seguem a mesma estrutura.

Bem vindos(as) ao curso!!
Olá! Bem vindos ao curso Preparatório para a Olimpíada Brasileira de Informática.

Este curso tem vídeo aulas e exercícios preparatórios muito interessantes, confirmam :)

O curso terá um novo conteúdo a cada semana, por isso, aproveitem bem o que é ensinado nos vídeos e façam os exercícios, ok?

Nesta primeira semana veremos um vídeo que conta o que é a Olimpíada, como ela acontece, quem pode participar, premiações e etc.

Após ver o vídeo, tem um tópico chamado Exercícios Introdutórios, é recomendado que façam :)

Espero que gostem!

Introdução

Aqui vocês verão um vídeo de Introdução a OBI, explicando como funciona e tudo direitinho!
Aproveitem!

Exercicios

Este questionário contém questões bem tranquilas de fazer, pra vocês se ambientarem com o tipo de questão da OBI. Boa sorte =)

Acesso restrito: 'Não disponível até que a atividade **Introdução** esteja marcada como concluída.'

+ Adicionar uma atividade ou recurso

Bem vindos(as) a semana 2 do curso!
Olá!
Nesta semana vocês verão os seguintes assuntos:

1. Estrutura de uma Questão
2. Tipos de Questões e Regras
3. Relação entre Regras e Tipos de Questões
4. Exemplos de Questões
5. Ordenação
6. Agrupamento
7. Cálculo
8. Método Geral de Resolução
9. Estatísticas das Últimas Competições

Veja os conteúdos na sequência: Parte 1, Parte 2 e Parte 3.

Figura 3.1: A figura mostra a estrutura modular do curso. Apresenta a visão geral de como são os módulos do curso.

A figura 3.2 mostra a estrutura de um módulo do curso, a qual repete-se para todos os outros módulos. É importante notar que cada módulo apresenta um vídeo e alguns exercícios e, somente após o vídeo ser visualizado, e marcado como tal, os exercícios podem ser feitos. Outra característica importante é que o aluno só avança no curso após ter realizado as atividades anteriores.

Na figura 3.3 pode-se visualizar o módulo no ambiente *Moodle* e o *player* para o vídeo.

A figura 3.4 mostra um exemplo dos questionários que foram realizados pelos alunos. Todos seguem a mesma estrutura.

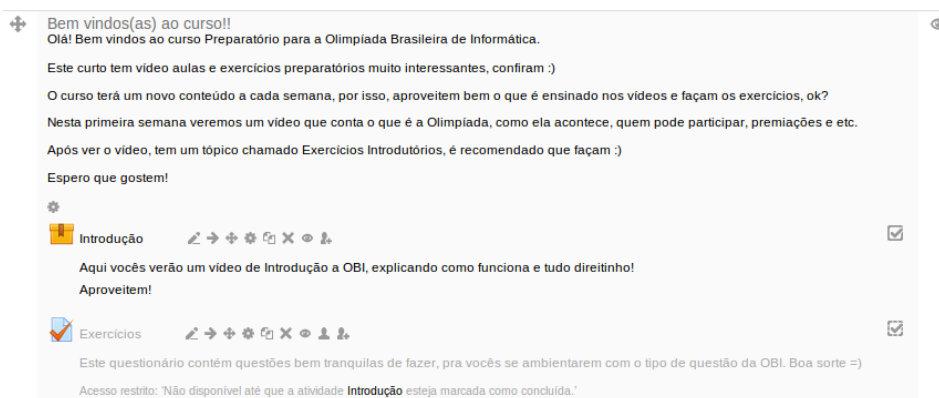


Figura 3.2: A figura mostra a estrutura de um dos módulos. Apresenta o módulo do vídeo e o questionário com os exercícios.



Figura 3.3: A figura mostra o módulo que contém o vídeo. O módulo possui um player que permite assistir ao vídeo no ambiente.


3.7 Coleta de Dados

O *Moodle* registra em sua base de dados todas as informações referentes ao uso do ambiente. A medida que os usuários o utilizam, as informações referentes as ações do usuário são registradas na base de dados. A estrutura do curso em formato modular e exercícios de fixação fazem com que os dados referentes a visualização dos módulos e realização dos exercícios sejam armazenados nas tabelas correspondentes (grupos de tabelas que armazenam informações sobre os módulos ativos e grupos de tabelas que armazenam informações referentes aos questionários).

Questão 3
Incompleto
Vale 1,0 ponto(s).

▼ Marcar questão
🔗 Editar questão

São dados N azulejos de dimensões $10\text{cm} \times 10\text{cm}$. Com eles, você deve montar um conjunto de quadrados (com espessura de um azulejo) de modo a utilizar TODOS os azulejos dados. Inicialmente você deve montar o maior quadrado possível com os azulejos dados; então, com os azulejos que sobraram, você deve montar o maior quadrado possível, e assim sucessivamente. Por exemplo, se forem dados 31 azulejos, o conjunto montado terá quatro quadrados, conforme ilustra a figura abaixo:



Conjunto com quatro quadrados, montado a partir de 31 azulejos

Qual o número de quadrados do conjunto montado se foram dados 148 azulejos?

Escolha uma:

- a. 5
- b. 3
- c. 6
- d. 4
- e. 2

[Verificar](#)

Figura 3.4: A figura mostra um dos questionários realizados pelos alunos no curso.

Sendo assim, durante o período de ocorrência do curso, os dados foram armazenados no banco de dados para, então, ser realizado o pré-processamento. O processo de pré-processamento dos dados tem a intenção de manipular as informações com intenção de torná-las dados úteis. Estes dados passaram, então, por um processo de mineração de dados com objetivo de analisar como foi o acompanhamento dos alunos no curso.

3.8 Ferramentas utilizadas

Foram selecionadas duas ferramentas para auxiliar neste trabalho: GISMO, um *plugin* instalado no *Moodle* que apresenta gráficos referentes a vários dados de utilização do ambiente e o *software* WEKA, que apresenta uma série de algoritmos de mineração de dados que resolvem os mais variados problemas de *data mining*.

3.8.1 Gismo

A ferramenta GISMO⁵ foi utilizada para a visualização dos dados do ambiente *Moodle*. Romero, Ventura e Garcia (2008) explicam que o *software* GISMO pode

⁵<http://gismo.sourceforge.net/>

ser instalado no sistema *Moodle* para fornecer representações gráficas que monitorem a interatividade dos alunos bem como rastrear informações que extraiam dados de uso do Moodle. Os exemplos de visualização incluem gráficos relatando, por exemplo, o acesso dos alunos ao curso, gráficos relatando os recursos mais acessados, entre outros. A figura 3.5 mostra os acessos dos alunos a um curso, em que as barras verticais azuis representam a quantidade de acessos.

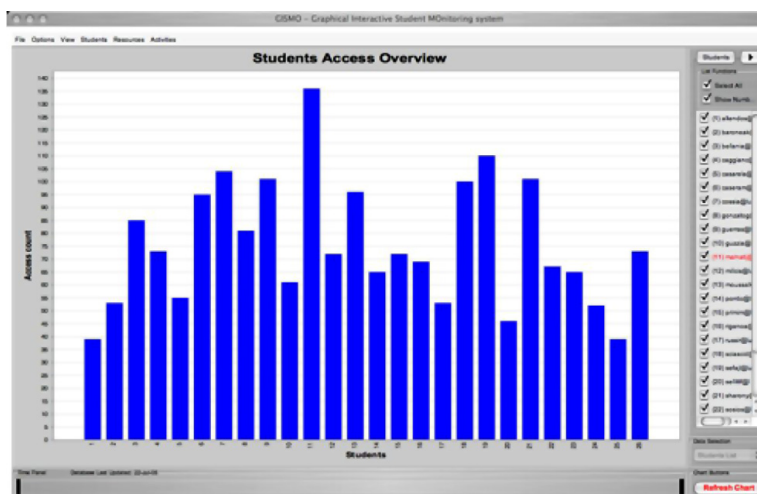


Figura 3.5: Figura retirada de (ROMERO; VENTURA; GARCIA, 2008) representando os acessos a um curso no *Moodle*

A ferramenta GISMO foi instalada no servidor do *Moodle* e habilitada em todos os cursos aplicados. A partir da utilização do ambiente virtual, a ferramenta gerou gráficos relacionados às notas obtidas e à frequência de acessos ao curso.

3.8.2 Weka

O *software* ambiente Waikato para análise de conhecimento (*Waikato Environment for Knowledge Analysis*, do inglês WEKA)⁶ foi desenvolvido pelo grupo de aprendizado de máquina da Universidade de Waikato. De acordo com Hall *et al.* (2009), o WEKA tem por objetivo fornecer uma ampla coleção de algoritmos de

⁶<http://www.cs.waikato.ac.nz/ml/weka/>

aprendizado de máquina que resolvam problemas de *data mining*. Os algoritmos podem ser aplicados diretamente sobre um conjunto de dados ou o programa pode ser executado a partir de um código Java. O software WEKA contém ferramentas para dados pré-processados com suporte a tarefas de classificação, regressão, agrupamento, regras de associação e visualização. O suporte a técnicas de agrupamento abrange técnicas de aprendizado supervisionado e não supervisionado, incluindo a rede neural de Kohonen (seção 2.6.2).

O formato de arquivos ARFF (Formato de arquivo atributo-relação)⁷ é utilizado pelo *software* e corresponde a um arquivo de texto ASCII que descreve um lista de instâncias que compartilham um conjunto de atributos. Estes arquivos possuem duas seções: a seção de cabeçalho (**@attribute**), que contém informações referentes ao conjunto de atributos a ser considerado e a seção de dados (**@data**), que contém os dados da base de dados, separados por vírgula.

A figura 3.6⁸ mostra a relação *Training*. A seção de cabeçalho contém atributos do tipo *numeric* e do tipo *quality*. Como exemplo, tem-se, respectivamente, “@attribute 'fixed acidity' numeric” e “@attribute quality good,bad”.

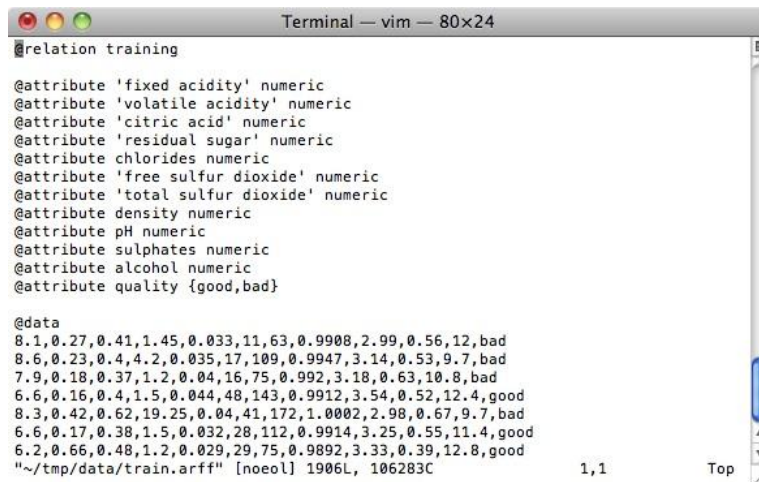
3.9 Pré-processamento

Para executar o pré-processamento dos dados, utilizou-se a linguagem de programação Java, a IDE NetBeans 7.0.1 e o sistema operacional Ubuntu 12.04. O pré-processamento foi dividido em algumas etapas, como segue:

- Etapa 1: estudo para selecionar as tabelas importantes do banco de dados do *Moodle*;
- Etapa 2: mapeamento em planilha das tabelas selecionadas na etapa 1;

⁷<http://weka.wikispaces.com/ARFF+%28book+version%29>, acessado em 02 de julho de 2013

⁸<http://users.eecs.northwestern.edu/~ahu340/eecs349-ps1/>, acessado em 02 de julho de 2013



```

relation training

@attribute 'fixed acidity' numeric
@attribute 'volatile acidity' numeric
@attribute 'citric acid' numeric
@attribute 'residual sugar' numeric
@attribute chlorides numeric
@attribute 'free sulfur dioxide' numeric
@attribute 'total sulfur dioxide' numeric
@attribute density numeric
@attribute pH numeric
@attribute sulphates numeric
@attribute alcohol numeric
@attribute quality {good,bad}

@data
8.1,0.27,0.41,1.45,0.033,11,63,0.9908,2.99,0.56,12,bad
8.6,0.23,0.4,4.2,0.035,17,109,0.9947,3.14,0.53,9.7,bad
7.9,0.18,0.37,1.2,0.04,16,75,0.992,3.18,0.63,10.8,bad
6.6,0.16,0.4,1.5,0.044,48,143,0.9912,3.54,0.52,12.4,good
8.3,0.42,0.62,19.25,0.04,41,172,1.0002,2.98,0.67,9.7,bad
6.6,0.17,0.38,1.5,0.032,28,112,0.9914,3.25,0.55,11.4,good
6.2,0.66,0.48,1.2,0.029,29,75,0.9892,3.33,0.39,12.8,good
"~/tmp/data/train.arff" [noeol] 1906L, 106283C
  
```

Figura 3.6: Exemplo de arquivo ARFF. Apresenta as informações da seção de cabeçalho e da seção de dados.

- Etapa 3: criação das tabelas selecionadas no banco de dados local;
- Etapa 4: análise dos dados quantitativos e criação da tabela de sumarização.

Etapa 1 - estudo para selecionar as tabelas importantes do banco de dados do Moodle

Foi realizado um estudo sobre quais seriam as tabelas interessantes para os propósitos deste estudo. A tabela 3.2 apresenta a listagem das tabelas selecionadas e quais informações elas armazenam. As tabelas selecionadas contém informações sobre as atividades dos usuários nos módulos juntamente com informações sobre a completude de um módulo (se o aluno marcou ou não a atividade como realizada). Apresentam, também, informações sobre as notas obtidas pelo usuário e o número de tentativas de submissão, permitindo saber, após análise, se o usuário está acompanhando o curso ou não. A tabela **mdl_log** apresenta informações referentes ao uso do ambiente. Tais informações permitem avaliar como cada usuário interage com o ambiente.

Etapa 2 - Mapeamento em planilha das tabelas selecionadas na etapa 1

Tabela	Tipo de informação armazenada
mdl_course_modules	Informações sobre quais módulos estão nos cursos
mdl_course_modules_completions	Informações sobre completude de atividades
mdl_quiz_attempts	Informações sobre as tentativas de submissão dos questionários
mdl_quiz_grades	Informações sobre as notas dos usuários nos questionários
mdl_log	Informações sobre as ações do usuário no ambiente virtual

Tabela 3.2: Tabelas importantes para os propósitos do estudo

Para compreensão de quais dados seriam necessários foi realizado um mapeamento em planilha, com especificação de quais campos de cada tabela seriam considerados. A tabela 3.3 apresenta os resultados deste mapeamento. As tabelas selecionadas possuem outros campos além dos listados, porém foram considerados apenas os campos necessários aos objetivos do estudo. Na tabela **mdl_course_modules** os campos *id*, *course* e *module* representam, respectivamente, o identificador da entrada na tabela, o número único que identificará o curso e o identificador do módulo que está naquele curso. Um módulo pode ser: um blog, um calendário, bate-papo, um curso, um fórum, login, questionários, módulo *Scorm* e um usuário. Neste projeto fez-se uso, principalmente, do módulo questionário.

Os campos considerados na tabela **mdl_course_modules_completions** são *id*, *userid*, *coursemoduleid* e *completionstate*. Os campos representam, respectivamente, o identificador da entrada na tabela, o identificador do usuário, o identificador exclusivo do módulo e o estado de “completude” do módulo (este campo assume valor 1 (um) se o módulo foi executado e 0 (zero), em caso contrário). É possível verificar se um módulo está completo devido a utilização do recurso do Moodle chamado *Completude de Atividades*, que permite marcar uma atividade

como concluída manualmente ou quando determinadas condições forem satisfeitas.

A tabela **mdl_quiz_attempts** apresenta informações sobre as tentativas e/ou submissões realizadas pelos usuários e a tabela **mdl_quiz_grades** contém informações sobre as notas obtidas pelos usuários.

Os campos *state*, *timestart* e *timefinish* refletem, respectivamente o estado de um questionário por parte de um usuário (o estado pode assumir os valores abandonado, em progresso ou finalizado), quando o questionário foi iniciado e quando o questionário foi finalizado.

A tabela **mdl_log** reflete as ações realizadas pelo usuário no ambiente virtual e é armazenada no campo *action*. Os valores considerados foram: *close attempt*, *continue attempt*, *login*, *pre-view*, *review*, *view*, *view all*, *view forum* e *view summary*.

Etapa 3 - Criação das tabelas selecionadas no banco de dados local

Nesta etapa foi iniciada a codificação utilizando a linguagem de programação Java e o driver de banco de dados *JDBC*. A conexão com o banco de dados do servidor *pesquisa.dcc.ufla.br/olimpiada* foi realizada para que o servidor local tivesse uma cópia das tabelas necessárias. Tal procedimento foi realizado para garantir dados sempre atualizados sem alterações na base de dados original.

Etapa 4 - Análise dos dados quantitativos e criação da tabela de sumarização

Após a seleção das tabelas importantes, do mapeamento dos campos úteis e a devida criação destes dados localmente, foi realizada, nesta etapa, a criação da tabela de sumarização. Esta tabela apresenta os dados quantitativos obtidos através de consultas ao banco de dados. As consultas realizadas podem ser visualizadas na tabela 3.4.

A consulta para obter o atributo *numberOfModulesMade* verifica, para cada usuá-

Tabela	Campos necessários
mdl_course_modules	<i>mdl_course_modules.id</i> <i>mdl_course_modules.course</i> <i>mdl_course_modules.module</i>
mdl_course_modules_completions	<i>mdl_course_modules_completions.id</i> <i>mdl_course_modules_completions.userid</i> <i>mdl_course_modules_completions.coursemoduleid</i> <i>mdl_course_modules_completions.completionstate</i>
mdl_quiz_attempts	<i>mdl_quiz_attempts.userid</i> <i>mdl_quiz_attempts.state</i> <i>mdl_quiz_attempts.timestart</i> <i>mdl_quiz_attempts.timefinish</i> <i>mdl_quiz_attempts.sumgrades</i>
mdl_quiz_grades	<i>mdl_quiz_grades.userid</i> <i>mdl_quiz_grades.grade</i>
mdl_log	<i>mdl_log.userid</i> <i>mdl_log.action</i>

Tabela 3.3: Campos das tabelas do banco de dados do *Moodle* que foram considerados

rio, quantos módulos estão com estado de completude igual a 1 (*mdl_course_modules_completion.completionstate=1*).

A consulta para obter o atributo *percentageOfAbandonment* realiza, para cada usuário, a contagem (*COUNT(state)*) de quantos questionários iniciados por ele (*mdl_quiz_attempts.userid = i*) foram abandonados (*mdl_quiz_attempts.state='abandoned'*). A consulta para obter o atributo *percentageOfCompleted* realiza, para cada usuário, a contagem (*COUNT(state)*) de quantos questionários iniciados por ele (*mdl_quiz_attempts.userid = i*) foram finalizados (*mdl_quiz_attempts.state='finished'*).

A consulta para obter o atributo *averageTimeAnswering* realizada, para cada usuário (*mdl_quiz_attempts.userid = i*) a soma dos tempos gastos para responder

aos exercícios (SUM(timefinish-timestart)) e considera apenas os exercícios que foram finalizados pelo usuário (mdl_quiz_attempts.state='finished'). O resultado da consulta representa apenas a soma e, para resultar na média de tempos gastos, o resultado dado pela consulta foi dividido pelo total de exercícios finalizados pelo usuário.

Os atributos de nota (*gradeQuiz1*, *gradeQuiz2*, *gradeQuiz3*, *gradeQuiz4*, *gradeQuiz5* e *gradeQuiz6*) representam as notas obtidas pelos usuários nos 6 (seis) exercícios que o curso apresenta. As consultas obtêm, para cada usuário (... where userid = i) as notas de acordo com a ordem dos questionários (SELECT grade ... order by quiz ASC, em que a restrição ASC ordena as notas obtidas do primeiro ao sexto módulo). Tais informações foram obtidas a partir da tabela mdl_quiz_grades.

Os 10 (dez) atributos *actionAttempt*, *actionCloseAttempt*, *actionContinueAttempt*, *actionLogin*, *actionPreView*, *actionReview*, *actionView*, *actionViewAll*, *actionViewForum* e *actionViewSummary* representam quantas ações de cada de tipo o usuário executou no ambiente. Para obtê-los, a consulta SQL considerou apenas valores da tabela mdl_log e, para cada usuário (mdl_log.userid = i), foram contabilizadas (COUNT(action)) quantas ações de cada tipo (mdl_log.action = name_action) o usuário realizou.

Tabela 3.4: Consultas realizadas para obter a tabela de sumarização (tabela 3.5)

Atributo	Consulta ao banco de dados
<i>numberOfModulesMade</i>	<pre>SELECT COUNT FROM mdl_course_modules, mdl_course_modules_completion where mdl_course_modules_completion.userid = i AND (mdl_course_modules.course = 2 OR mdl_course_modules.course = 4) AND mdl_course_modules.id = mdl_course_modules_completion.coursemoduleid AND mdl_course_modules_completion.completionstate=1</pre>
<i>percentageOfAbandonment</i>	<pre>SELECT COUNT(state) from mdl_quiz_attempts where mdl_quiz_attempts.userid = + i + and mdl_quiz_attempts.state = 'abandoned'</pre>
<i>percentageOfCompleted</i>	<pre>SELECT COUNT(state) from mdl_quiz_attempts where mdl_quiz_attempts.userid = + i + and mdl_quiz_attempts.state= 'finished'</pre>
<i>averageTimeAnswering</i>	<pre>SELECT SUM(timefinish-timestart) from mdl_quiz_attempts where mdl_quiz_attempts.userid = + i + and mdl_quiz_attempts.state='finished'</pre>
<i>gradeQuiz1, gradeQuiz2, grade-Quiz3, gradeQuiz4, gradeQuiz5 e gradeQuiz6</i>	<pre>SELECT grade from mdl_quiz_grades where userid = + i + order by quiz ASC; aqui foi obtida uma variável do tipo ResultSet que contém todas as notas</pre>

Continua na próxima página.

Tabela 3.4 – Consultas realizadas para obter a tabela de sumarização (tabela 3.5)

Grupo	Descrição
<i>actionAttempt, actionCloseAttempt, actionContinueAttempt, actionLogin, actionPreView, actionReview, actionView, actionViewAll, actionViewForum e actionViewSummary</i>	SELECT COUNT(action) from mdl_log where mdl_log.userid = + i + and mdl_log.action = <i>name_action</i> ; <i>name_action</i> foi substituída pelo nome de cada ação

A tabela 3.5 apresenta a tabela de sumarização criada. Esta tabela apresenta o resultado do processo de pré-processamento dos dados de uso do *Moodle*, em que seus campos são atributos quantitativos resultantes do processo percorrido durante as etapas 1, 2 e 3.

Tabela 3.5: Tabela de sumarização.

Tabelas utilizadas	Dados obtidos
mdl_course_modules, mdl_course_modules_completions	<i>numberOfModulesMade</i> - número de módulos feitos pelo usuário
mdl_quiz_attempts	<i>percentageOfAbandonment</i> - Porcentagem de abandono do usuário nos questionários <i>percentOfCompleted</i> - Porcentagem de questionários finalizados pelo usuário <i>averageTimeAnswering</i> - média de tempo que o usuário gastou respondendo aos questionários

Continua na próxima página.

Tabela 3.5 – Tabela de sumarização.

Tabelas utilizadas	Dados obtidos
mdl_quiz_grades	<i>gradeQuiz1, gradeQuiz2, gradeQuiz3, gradeQuiz4, gradeQuiz5, gradeQuiz6</i> - Notas do usuário nos questionários
mdl_log	<p><i>actionAttempt</i> - Quantas vezes o usuário fez submissões em questionários</p> <p><i>actionCloseAttempt</i> - Quantas vezes o usuário encerrou uma tentativa de submissão (nota)</p> <p><i>actionContinueAttempt</i> - Quantas vezes o usuário continuou uma tentativa anterior não finalizada</p> <p><i>actionLogin</i> - Quantas vezes o usuário fez <i>login</i> no Moodle não finalizada</p> <p><i>actionPreView</i> - Quantas vezes o usuário pré-visualizou algum recurso</p> <p><i>actionReview</i> - Quantas vezes o usuário repetiu a visualização de um recurso</p> <p><i>actionView</i> - Quantas vezes o usuário visualizou um recurso</p> <p><i>actionViewAll</i> - Quantas vezes o usuário visualizou o ambiente de forma geral</p> <p><i>actionViewForum</i> - Quantas vezes o usuário visualizou o fórum</p> <p><i>actionViewSummary</i> - Quantas vezes o usuário visualizou o sumário do curso</p>

4 RESULTADOS

4.1 Resultados - Ferramenta GISMO

Neste trabalho, foi realizado o pré-processamento dos dados obtidos no curso preparatório para a OBI. Houveram 5 (cinco) escolas inscritas, porém, apenas 2 (duas) escolas tiveram infraestrutura para participar do projeto. As principais dificuldades encontradas foram:

- Os laboratórios de informática das escolas não são frequentemente utilizados e não há manutenção nos computadores;
- Os computadores das escolas, muitas vezes, não apresentam componentes de som.
- A Internet disponível nas escolas possui velocidade lenta e a conexão é instável, com atraso na execução do curso;

O problema dos computadores não possuírem som foi solucionado através da visualização conjunta dos vídeos e com a resolução separada dos exercícios. Na outra escola, as atividades foram realizadas normalmente. Outra consideração importante é a falta de assiduidade e comprometimento dos alunos, que hora compareciam e hora não compareciam, o que atrapalhou o rendimento das turmas.

A tabela 4.1 mostra, para as 2 (duas) escolas parceiras, 1) o número de alunos que se inscreveram no curso e 2) o número de alunos que terminaram o curso.

A tabela 4.2 mostra o período inicial e final do curso em cada uma das escolas parceiras. Pode-se observar que o curso na escola X durou quase 3 meses. O curso na escola Y durou aproximadamente 1 mês e 15 dias.

Escola	Quantidade de alunos inscritos	Quantidade de alunos que terminaram o curso
X	29 alunos	19 alunos
Y	8 alunos	4 alunos

Tabela 4.1: Tabela que apresenta, respectivamente, o número de alunos inscritos e o número de alunos que terminaram o curso.

Escola	Início	Término
X	10 de maio	9 de agosto
Y	01 de junho	18 de julho

Tabela 4.2: Tabela que apresenta o período de duração do curso nas 2 (duas) escolas parceiras.

A ferramenta GISMO gerou gráficos que representam como foi a interação dos alunos com o curso. A figura 4.1 representa as notas obtidas pelos alunos da escola X. Nesta escola houveram 29 (trinta) alunos inscritos e, deste total, 19 terminaram o curso (aproximadamente 63%).

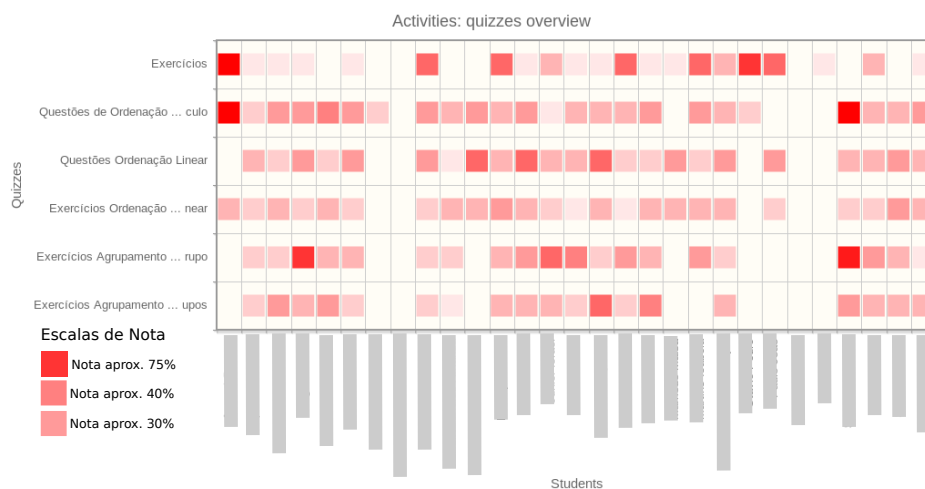


Figura 4.1: Escola X - Figura gerada pelo GISMO, representa as notas obtidas nos exercícios.

A figura 4.2 representa os acessos dos alunos da escola X ao curso. Nesta escola, os acessos não foram sempre estáveis devido a falta de assiduidade de alguns alunos.

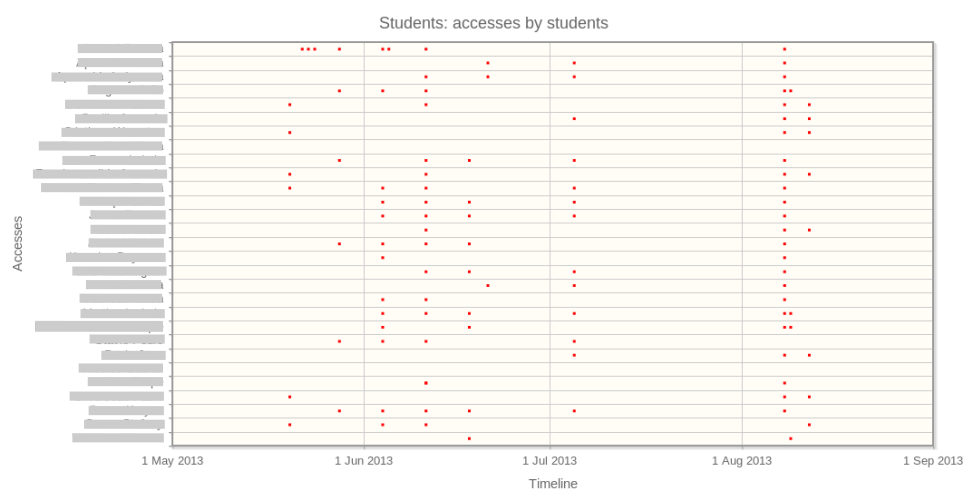


Figura 4.2: Escola X - Figura gerada pelo GISMO, representa os acessos de cada aluno ao curso.

A figura 4.3 representa a frequência de acessos ao ambiente *Moodle*. No período inicial do curso houveram aproximadamente 100 acessos. No mês de junho, houve certo desinteresse devido a proximidade com as férias e, com a volta das aulas, o interesse foi retomado. Neste período, houve grande quantidade de acessos (a maior quantidade corresponde a aproximadamente 900 acessos).

A figura 4.4 representa as notas obtidas pelos alunos da escola Y. Pode ser observado que, dos 8 alunos inscritos, 4 completaram o curso, o que representa 44%. Os alunos que estão sem nota ou com poucas notas do total possuíram pouca ou nenhuma frequência.

A figura 4.5 representa a frequência de acessos dos alunos da escola Y ao curso. Pode-se observar que os alunos assíduos tiveram acessos constantes ao curso.

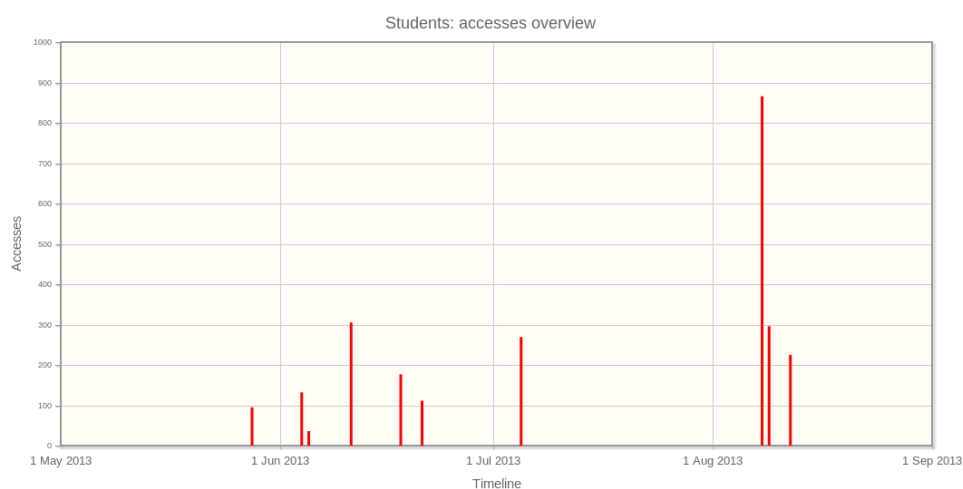


Figura 4.3: Escola X - Figura gerada pelo GISMO, representa os acessos gerais dos alunos ao curso.

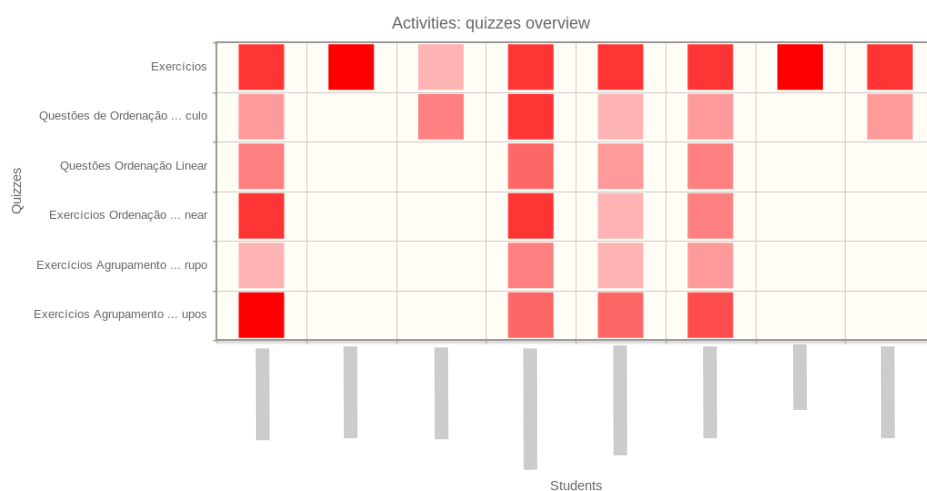


Figura 4.4: Escola Y - Figura gerada pelo GISMO, representa as notas obtidas nos exercícios.

A figura 4.6 representa a frequência de acessos ao ambiente nos meses da ocorrência do curso. Nesta escola, o curso foi iniciado no mês de junho. Nos meses de junho a julho os acessos se mantiveram estáveis, com auge de, aproximadamente, 250 acessos. O período de férias compreendeu parte do mês de julho e, por isso, os acessos voltaram a ocorrer no mês de agosto, período de volta às aulas.

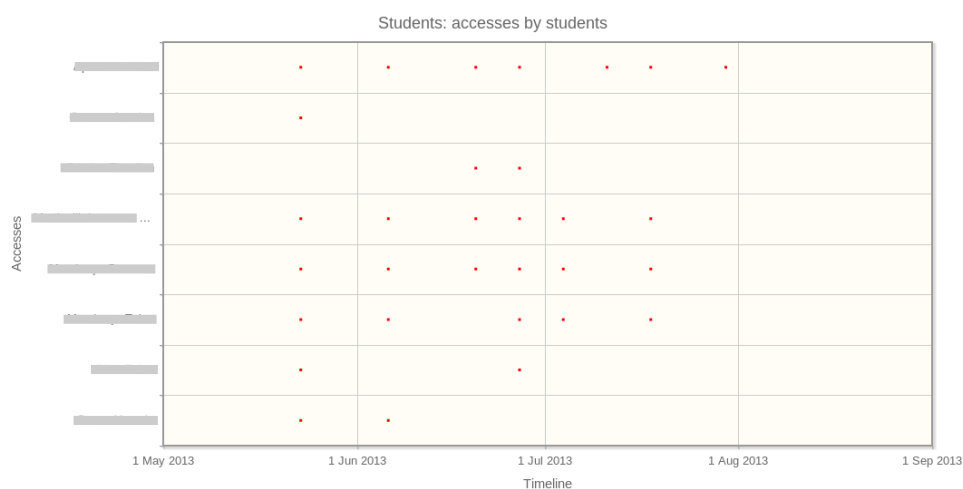


Figura 4.5: Escola Y - Figura gerada pelo GISMO, representa os acessos de cada aluno ao curso.

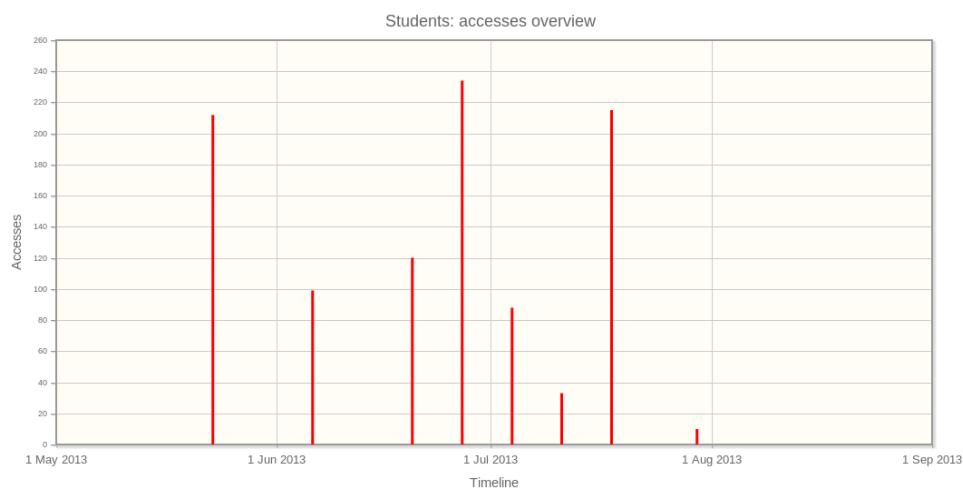


Figura 4.6: Escola Y - Figura gerada pelo GISMO, representa os acessos gerais dos alunos ao curso.

4.2 Resultados - Ferramenta WEKA

O resultado do pré-processamento dos dados do curso estão representados na tabela de sumarização. Os dados armazenados na tabela de sumarização foram convertidos para o formato requerido pelo *software* WEKA (arquivo ARFF - ver seção

3.8.2). Tais dados, já no formato apropriado, constituíram a entrada para o algoritmo da rede neural de Kohonen. O algoritmo de Kohonen foi executado com taxa de aprendizagem definida como 0.4 e os atributos considerados são os 20 atributos descritos na tabela de Sumarização (ver tabela 3.5) e que estão representados na figura 4.7.

```
Attributes: 20
            gradeQuiz1
            gradeQuiz2
            gradeQuiz3
            gradeQuiz4
            gradeQuiz5
            gradeQuiz6
            numberOfModulesMade
            percentageOfAbandonment
            percentageOfCompleted
            averageTimeAnswering
            actionAttempt
            actionCloseAttempt
            actionContinueAttempt
            actionLogin
            actionPreView
            actionReview
            actionView
            actionViewAll
            actionViewForum
            actionViewSummary
```

Figura 4.7: Figura retirada do apêndice A.1. Apresenta os atributos da tabela de sumarização.

Foram realizadas duas análises. Em ambas as análises, as variáveis que armazenam as notas (*gradeQuiz1*, *gradeQuiz2*, *gradeQuiz3*, *gradeQuiz4*, *gradeQuiz5* e *gradeQuiz6*) poderão assumir valor igual a -10, o que indica que houveram usuários que não realizaram as atividades correspondentes àquele módulo (módulo 1, 2, 3, 4, 5 ou 6).

Análise que considera todos os alunos que realizaram o curso

Nesta análise, pode-se observar o acompanhamento dos alunos perante o curso. Os dados obtidos podem ser visualizados no apêndice A.1. A figura 4.8 mostra que o número de alunos inscritos no ambiente é 43 (representado pela variável *Instances*).

```

=== Run information ===

Scheme:      weka.clusterers.Sel
Relation:    moodleData
Instances:   43
Attributes:  20

```

Figura 4.8: Figura retirada do apêndice A.1. Mostra o número de alunos inscritos no *Moodle* (variável *Instances*) e o número de atributos para a análise (variável *Attributes*)

Os agrupamentos obtidos são apresentados na figura 4.9. Pode-se observar que, do total de 43 instâncias, 35% delas (agrupamento 0) correspondem a alunos que se cadastraram e que realizaram a primeira, a segunda ou nenhuma das duas atividades, o que indica abandono do curso. A partir da observação das variáveis referentes as notas, verifica-se que estes alunos possuem, em média, notas até o segundo módulo. Os campos mínimo e máximo dos atributos de nota assumirem

Clustered Instances	
(Instâncias Agrupadas)	
Agrupamento 0	15 (35%)
Agrupamento 1	4 (9%)
Agrupamento 2	16 (37%)
Agrupamento 3	8 (19%)

Figura 4.9: Figura retirada do apêndice A.1. Mostra os agrupamentos obtidos pela RNA de Kohonen.

valor -10 indica que os usuários deste agrupamento não realizaram os 4 últimos módulos. A figura 4.10 mostra que os alunos do agrupamento 0 realizaram, em média, até o segundo módulo. Pode-se observar que os campos mínimo e máximo das variáveis de nota desse agrupamento, a partir do módulo 4, possuem valor igual a -10, o que indica abandono das atividades do curso.

Attribute	Cluster
	0
	(15)
=====	
gradeQuiz1	
value	33.6067
<u>min</u>	<u>-10</u>
<u>max</u>	<u>100</u>
mean	34
std. dev.	40.585
gradeQuiz2	
value	-1.496
<u>min</u>	<u>-10</u>
<u>max</u>	<u>40</u>
mean	-1.3333
std. dev.	16.4172
gradeQuiz3	
value	-10
<u>min</u>	<u>-10</u>
<u>max</u>	<u>-10</u>
mean	-10
std. dev.	0
gradeQuiz4	
value	-10
<u>min</u>	<u>-10</u>
<u>max</u>	<u>-10</u>
mean	-10
std. dev.	0
gradeQuiz5	
value	-10
<u>min</u>	<u>-10</u>
<u>max</u>	<u>-10</u>
mean	-10
std. dev.	0
gradeQuiz6	
value	-10
<u>min</u>	<u>-10</u>
<u>max</u>	<u>-10</u>

Figura 4.10: Figura retirada do apêndice A.1. Mostra que os alunos do agrupamento 0 realizaram, em média, até o segundo módulo.

O agrupamento 1 e o agrupamento 3 representam os alunos que realizaram, em média, até a sexta atividade (estas informações estão presentes nas variáveis referentes às notas). A diferenciação se faz quando observamos que o agrupamento 3 possui melhores notas quando comparado ao agrupamento 1. Pode-se observar que os alunos do agrupamento 3 completaram 91,56% dos questionários que iniciaram e o grupo 1, por sua vez, completou 74,05%. A porcentagem de “completude” considera quantos exercícios (questionários) o usuário começou e

quantos foram finalizados. A figura 4.11 mostra as porcentagens médias de questionários finalizados pelos alunos que pertencem aos agrupamentos 1 e 3.

percentageOfCompleted		Agrupamento 1		Agrupamento 3
value	0.6639	<u>0.7405</u>	0.885	<u>0.9156</u>
min	0	0.5	0.6154	0.7273
max	1	1	1	1
mean	0.6611	0.7417	0.8855	0.9156
std. dev.	0.4591	0.2115	0.1182	0.1022

Figura 4.11: Figura retirada do apêndice A.1. Mostra as porcentagens médias de questionários finalizados pelos alunos dos agrupamentos 1 e 3.

O agrupamento 2, por sua vez, agrupou os alunos que obtiveram notas em todos os questionários. Os alunos deste grupo finalizaram 88,5% dos questionários que iniciaram. A porcentagem de questionários iniciados e que foram finalizados pode ser observada na figura 4.12.

percentageOfCompleted			Agrupamento 2	
value	0.6639	0.7405	<u>0.885</u>	0.9156
min	0	0.5	0.6154	0.7273
max	1	1	1	1

Figura 4.12: Figura retirada do apêndice A.1. Mostra a porcentagem média obtida pelos alunos do agrupamento 2.

Análisa que considera apenas os alunos que terminaram o curso

Esta análise considera apenas os alunos que realizaram todos os módulos, em que procura-se entender os padrões encontrados entre os alunos que realizaram todas as atividades do curso. Buscou-se mostrar o rendimento alcançado pelos alunos que acompanharam o curso integralmente. Os dados obtidos com esta análise podem ser visualizados no apêndice A.2. O número de alunos que terminou o curso é 23 (representado na variável *Instances* - figura 4.13).


```

Scheme:      weka.clusterer:
Relation:    moodleData
Instances:   23
Attributes:  20

```

Figura 4.13: Figura retirada do apêndice A.2. Mostra o número de alunos que terminaram o curso (variável *Instances*) e o número de atributos para a análise (variável *Attributes*)

Os agrupamentos obtidos podem ser visualizados na figura 4.14.

Clustered Instances	
Agrupamento 0	4 (17%)
Agrupamento 1	10 (43%)
Agrupamento 2	5 (22%)
Agrupamento 3	4 (17%)

Figura 4.14: Figura retirada do apêndice A.2. Mostra os agrupamentos obtidos pela RNA de Kohonen.

Nesta análise, pode-se observar que os alunos do agrupamento 1 finalizaram 94,58% dos exercícios que iniciaram e, os alunos do agrupamento 2, possuem 96,63% dos seus questionários iniciados e finalizados. As porcentagens médias obtidas pelos alunos dos agrupamentos 1 e 2 são encontradas na figura 4.15.

percentageOfCompleted		Agrupamento 1	Agrupamento 2	
value	0.7401	<u>0.9458</u>	<u>0.9663</u>	0.9082
min	0.6154	0.8571	0.8333	0.8571
max	0.875	1	1	1
mean	0.7393	0.9464	0.9667	0.9087
std. dev.	0.1193	0.0694	0.0745	0.0627

Figura 4.15: Figura retirada do apêndice A.2. Mostra a porcentagem média de questionários finalizados pelos alunos do agrupamento 2.

Os alunos do agrupamento 1 não alcançaram boas notas, apesar da porcentagem média de aproximadamente 95% de questionários finalizados. Isto pode ser observado no campo *mean* (valor médio) em cada uma de suas variáveis referente a nota (conforme as indicações presentes na figura 4.16).

Attribute	Cluster	
	0 (4)	1 (10)
=====		
<u>gradeQuiz1</u>		
value	12.4994	2.4864
min	0	0
max	25	25
<u>mean</u>	12.5	<u>2.5</u>
std. dev.	14.4338	7.9057
<u>gradeQuiz2</u>		
value	17.4376	25.9404
min	0	10
max	30	40
<u>mean</u>	17.5	<u>26</u>
std. dev.	12.5831	8.4327
<u>gradeQuiz3</u>		
value	32.5632	21.9879
min	20	10
max	50	50
<u>mean</u>	32.5	<u>22</u>
std. dev.	12.5831	12.2927
<u>gradeQuiz4</u>		
value	16.6834	19.2769
min	13.3333	0
max	20	33.3333
<u>mean</u>	16.6666	<u>19.3333</u>
std. dev.	3.849	9.6609
<u>gradeQuiz5</u>		
value	37.2701	21.649
min	8.3333	0
max	83.3333	41.6667
<u>mean</u>	37.5	<u>21.6667</u>
std. dev.	36.3241	13.1468
<u>gradeQuiz6</u>		
value	30.1513	21.0168
min	20	10
max	60	40
<u>mean</u>	30	<u>21</u>
std. dev.	20	9.9443

Figura 4.16: Figura retirada do apêndice A.2. Mostra as notas médias obtidas pelos alunos do agrupamento 1.

Os alunos do agrupamento 2 possuem notas com valores mínimos (campo *min*) no intervalo de 15 a 30 e valores máximos (campo *max*) nos intervalos de 50 a 100. Estas informações podem ser visualizadas na figura 4.17 nos campos *min* e *max*. As médias alcançadas por este grupo tratam-se de valores em torno de

50% e, por isso, os desvios relacionados as notas obtidas são considerados estáveis (campo *mean*, também na figura 4.17).

Attribute	Agrupamento 2		
	Cluster 0 (4)	1 (10)	2 (5)
=====			
gradeQuiz1			
value	12.4994	2.4864	65.2
<u>min</u>	0	0	<u>25</u>
<u>max</u>	25	25	<u>75</u>
mean	12.5	2.5	65
std. dev.	14.4338	7.9057	22.3607
gradeQuiz2			
value	17.4376	25.9404	51.8402
<u>min</u>	0	10	<u>20</u>
<u>max</u>	30	40	<u>100</u>
mean	17.5	26	52
std. dev.	12.5831	8.4327	35.6371
gradeQuiz3			
value	32.5632	21.9879	36.1202
<u>min</u>	20	10	<u>20</u>
<u>max</u>	50	50	<u>50</u>
mean	32.5	22	36
std. dev.	12.5831	12.2927	11.4018
gradeQuiz4			
value	16.6834	19.2769	45.6682
<u>min</u>	13.3333	0	<u>13.3333</u>
<u>max</u>	20	33.3333	<u>73.3333</u>
mean	16.6666	19.3333	45.3333
std. dev.	3.849	9.6609	27.2437
gradeQuiz5			
value	37.2701	21.649	39.7497
<u>min</u>	8.3333	0	<u>16.6667</u>
<u>max</u>	83.3333	41.6667	<u>91.6667</u>
mean	37.5	21.6667	40
std. dev.	36.3241	13.1468	30.8446
gradeQuiz6			
value	30.1513	21.0168	64.2814
<u>min</u>	20	10	<u>30</u>
<u>max</u>	60	40	<u>100</u>
mean	30	21	64
std. dev.	20	9.9443	25.0998

Figura 4.17: Figura retirada do apêndice A.2. Mostra as notas mínimas e máximas obtidas pelos alunos do agrupamento 2.

Os alunos do agrupamento 3 obtiveram, todos, 50% de nota no primeiro módulo (campos *min* e *max* na variável *gradeQuiz1*. Nos módulos seguintes, as notas máximas alcançadas foram em torno de 30% (ver campo *max* dos outros módulos). Estas informações podem ser visualizadas na figura 4.18.

Attribute	Agrupamento 3			
	Cluster 0 (4)	1 (10)	2 (5)	3 (4)
gradeQuiz1				
value	12.4994	2.4864	65.2	50
<u>min</u>	0	0	25	<u>50</u>
<u>max</u>	25	25	75	<u>50</u>
mean	12.5	2.5	65	50
std. dev.	14.4338	7.9057	22.3607	0
gradeQuiz2				
value	17.4376	25.9404	51.8402	22.4873
<u>min</u>	0	10	20	<u>20</u>
<u>max</u>	30	40	100	<u>30</u>
mean	17.5	26	52	22.5
std. dev.	12.5831	8.4327	35.6371	5
gradeQuiz3				
value	32.5632	21.9879	36.1202	15.0246
<u>min</u>	20	10	20	<u>0</u>
<u>max</u>	50	50	50	<u>30</u>
mean	32.5	22	36	15
std. dev.	12.5831	12.2927	11.4018	12.9099
gradeQuiz4				
value	16.6834	19.2769	45.6682	14.9579
<u>min</u>	13.3333	0	13.3333	<u>0</u>
<u>max</u>	20	33.3333	73.3333	<u>33.3333</u>
mean	16.6666	19.3333	45.3333	15
std. dev.	3.849	9.6609	27.2437	14.7824
gradeQuiz5				
value	37.2701	21.649	39.7497	18.8653
<u>min</u>	8.3333	0	16.6667	<u>8.3333</u>
<u>max</u>	83.3333	41.6667	91.6667	<u>33.3333</u>
mean	37.5	21.6667	40	18.75
std. dev.	36.3241	13.1468	30.8446	12.5
gradeQuiz6				
value	30.1513	21.0168	64.2814	10.05
<u>min</u>	20	10	30	<u>0</u>
<u>max</u>	60	40	100	<u>20</u>
mean	30	21	64	10
std. dev.	20	9.9443	25.0998	8.165

Figura 4.18: Figura retirada do apêndice A.2. Mostra as notas mínimas e máximas obtidas pelos alunos do agrupamento 3.

É importante observar também, que o percentual de questionários finalizados pelos alunos do agrupamento 3 é em torno de 90% (figura 4.19) e, com exceção do primeiro módulo, possuem menores notas quando comparados ao agrupamento 1.

percentageOfCompleted				Agrupamento 3
value	0.7401	0.9458	0.9663	<u>0.9082</u>
min	0.6154	0.8571	0.8333	0.8571
max	0.875	1	1	1
mean	0.7393	0.9464	0.9667	0.9087
std. dev.	0.1193	0.0694	0.0745	0.0627

Figura 4.19: Figura retirada do apêndice A.2. Apresenta o percentual de questionários finalizados pelo alunos do agrupamento 3.

O agrupamento 0 representa os alunos que obtiveram notas em torno de 20 nos 4 primeiros módulos e aumento de aproximadamente 40% das notas nos módulos restantes (módulos 5 e 6). Tais dados podem ser observados através das notas mínima e máxima obtidas pelos alunos deste agrupamento (figura 4.20).

Attribute	Agrupamento 0 (4)
=====	
gradeQuiz1	
value	12.4994
<u>min</u>	<u>0</u>
<u>max</u>	<u>25</u>
mean	12.5
std. dev.	14.4338
gradeQuiz2	
value	17.4376
<u>min</u>	<u>0</u>
<u>max</u>	<u>30</u>
mean	17.5
std. dev.	12.5831
gradeQuiz3	
value	32.5632
<u>min</u>	<u>20</u>
<u>max</u>	<u>50</u>
mean	32.5
std. dev.	12.5831
gradeQuiz4	
value	16.6834
<u>min</u>	<u>13.3333</u>
<u>max</u>	<u>20</u>
mean	16.6666
std. dev.	3.849
gradeQuiz5	
value	37.2701
<u>min</u>	<u>8.3333</u>
<u>max</u>	<u>83.3333</u>
mean	37.5
std. dev.	36.3241
gradeQuiz6	
value	30.1513
<u>min</u>	<u>20</u>
<u>max</u>	<u>60</u>
mean	30
std. dev.	20

Figura 4.20: Figura retirada do apêndice A.2. Mostra as notas mínimas e máximas obtidas pelos alunos do agrupamento 3.

Apesar dos alunos do agrupamento 1 e 2 possuírem uma das maiores taxas de completude (94,58%), estes alunos não obtiveram boas notas. Já os alunos do agrupamento 2, que finalizaram 96,63% dos questionários que iniciaram, obtiveram notas mínimas baixas. Porém, as notas máximas obtiveram valores maiores (variando de 50 a 100). Então, em relação ao agrupamento 1, o agrupamento 2 obteve melhores notas. Já os alunos do agrupamento 3 possuíram nota de 50% no primeiro módulo e nos módulos seguintes as notas diminuíram (notas em torno de

30%) mas, ainda assim, possuíram melhores notas quando comparado ao grupo 2 e, conseqüentemente, em relação também ao grupo 1. Os alunos do agrupamento 0, por sua vez, obtiveram notas em torno de 20 nos 4 primeiros módulos com um pequeno aumento de 40% nos 2 (dois) módulos seguintes, o que indica que não melhoraram suas notas ao longo do curso.

4.3 Resultados - Questionário Índice de Estilos de Aprendizagem (ILS)

O questionário de Índices de Estilos de Aprendizagem foi aplicado para os alunos que terminaram o curso nas duas escolas parceiras. Os estilos de aprendizagem encontrados estão disponíveis na tabela 4.3.

Tabela 4.3: Perfis de aprendizagem detectados com a aplicação do questionário ILS

Aluno	Ativo/Reflexivo	Sensorial/Intuitivo	Visual/Verbal	Sequencial/Global
1	balanceado	balanceado	balanceado	balanceado
2	preferência moderada para Ativo	balanceado	balanceado	balanceado
3	preferência moderada para Ativo	balanceado	balanceado	moderada preferência para Global
4	preferência moderada para Reflexivo	balanceado	moderada preferência para Visual	balanceado
5	balanceado	balanceado	moderada preferência para Visual	moderada preferência para Sequencial
6	balanceado	moderada preferência para Sensorial	moderada preferência para Visual	balanceado
7	balanceado	balanceado	balanceado	balanceado

Continua na próxima página.

Tabela 4.3 – *Continua na próxima página.*

Aluno	Ativo/Reflexivo	Sensorial/Intuitivo	Visual/Verbal	Sequencial/Global
8	balanceado	moderada preferência para Intuitivo	moderada preferência para Visual	balanceado
9	balanceado	balanceado	balanceado	balanceado
10	moderada preferência para Ativo	balanceado	balanceado	balanceado
11	moderada preferência para Ativo	balanceado	balanceado	balanceado
12	moderada preferência para Ativo	balanceado	balanceado	moderada preferência para Sequencial
13	balanceado	balanceado	balanceado	balanceado
14	balanceado	forte preferência para Sensorial	moderada preferência para Visual	balanceado
15	balanceado	balanceado	moderada preferência para Visual	balanceado
16	balanceado	balanceado	balanceado	balanceado
17	forte preferência para Ativo	balanceado	balanceado	balanceado

Continua na próxima página.

Tabela 4.3 – *Continua na próxima página.*

Aluno	Ativo/Reflexivo	Sensorial/Intuitivo	Visual/Verbal	Sequencial/Global
18	balanceado	moderada preferência para Sensorial	moderada preferência para Visual	balanceado
19	balanceado	balanceado	moderada preferência para Verbal	balanceado
20	balanceado	forte preferência para Sensorial	moderada preferência para Verbal	balanceado
21	balanceado	balanceado	balanceado	balanceado
22	balanceado	balanceado	moderada preferência para Visual	moderada preferência para Global
23	forte preferência para Ativo	balanceado	balanceado	balanceado

Os dados obtidos na tabela 4.3 mostram os estilos individuais de aprendizagem para cada um dos 23 (vinte e três) alunos que terminaram o curso. Cada um dos 8 (oito) estilos de aprendizagem pode assumir um dos seguintes valores: estilo de aprendizagem balanceado entre as duas dimensões, preferência moderada para um dos estilos e forte preferência para um dos estilos.

Para o estilo de aprendizagem Ativo/Reflexivo, um aluno poderia ser classificado com preferência individual:

- Balanceada entre Ativo/Reflexivo
- Preferência moderada para
 - Ativo
 - Reflexivo
- Forte preferência para
 - Ativo
 - Reflexivo

A partir da tabela 4.3 foi criada a tabela 4.4, que contém o resumo dos estilos de aprendizagem encontrados para a dimensão Ativo/Reflexivo. Observa-se que 65,22% dos alunos possuem preferência balanceada entre as duas dimensões de aprendizagem. O restante, em sua maioria, apresenta preferência moderada para o estilo de aprendizagem Ativo (5 alunos).

Tabela 4.4: Estilos de aprendizagem Ativo/Reflexivo encontrados com a aplicação do questionário ILS

Porcentagem obtida	Estilo Ativo/Reflexivo
65,22% (15 alunos)	balanceado entre Ativo/Reflexivo
21,74% (5 alunos)	moderada preferência para Ativo
4,35% (1 aluno)	moderada preferência para Reflexivo
8,70% (2 alunos)	forte preferência para Ativo

A partir da tabela 4.3 foi criada a tabela 4.5, que contém o resumo dos estilos de aprendizagem encontrados para a dimensão Sensorial/Intuitivo.

- Balanceada entre Sensorial/Intuitivo
- Moderada preferência para
 - Sensorial
 - Intuitivo
- Forte preferência para
 - Sensorial
 - Intuitivo

Na tabela 4.5, observa-se que 78,26% dos alunos possuem preferência balanceada entre as duas dimensões de aprendizagem. Para estas dimensões, 2 (dois) alunos apresentavam forte preferência para Sensorial, 2 (dois) alunos apresentam moderada preferência para Sensorial e 1 (um) aluno apresenta moderada preferência para Intuitivo.

Tabela 4.5: Estilos de aprendizagem Sensorial/Intuitivo encontrados com a aplicação do questionário ILS

Porcentagem obtida	Estilo Sensorial/Intuitivo
78,26% (18 alunos)	balanceado entre Sensorial/Intuitivo
8,70% (2 alunos)	moderada preferência para Sensorial
4,35% (1 aluno)	moderada preferência para Intuitivo
8,70% (2 alunos)	forte preferência para Sensorial

Para as dimensões Visual/Verbal, um aluno poderia ser classificado com preferência individual:

- Balanceada entre Visual/Verbal
- Moderada preferência para
 - Visual
 - Verbal
- Forte preferência para
 - Visual
 - Verbal

A partir da tabela 4.3 foi criada a tabela 4.6, que contém o resumo dos estilos de aprendizagem encontrados para a dimensão Visual/Verbal. Na tabela 4.6, observa-se que aproximadamente metade dos alunos (56,25%) possuem preferência balanceada entre as duas dimensões de aprendizagem. O restante está dividido em: 8 (oito) alunos apresentavam moderada preferência para Visual e os outros 2 (dois) alunos apresentam moderada preferência para Verbal.

Tabela 4.6: Estilos de aprendizagem Visual/Verbal encontrados com a aplicação do questionário ILS

Porcentagem obtida	Estilo Visual/Verbal
56,52% (13 alunos)	balanceado entre Visual/Verbal
34,87% (8 alunos)	moderada preferência para Visual
8,70% (2 alunos)	moderada preferência para Verbal

Para as dimensões Sequencial e Global, um aluno poderia ser classificado com preferência individual:

- Balanceada entre Sequencial/Global
- Moderada preferência para
 - Sequencial
 - Global
- Forte preferência para
 - Sequencial
 - Global

A partir da tabela 4.3 foi criada a tabela 4.7, que contém o resumo dos estilos de aprendizagem encontrados para a dimensão Sequencial/Global. Na tabela 4.7, 19 dos 23 alunos (82,61%) possuem preferência balanceada entre as duas dimensões de aprendizagem. Os outros 4 (quatro) alunos estão divididos em 2 (dois) alunos com moderada preferência para Sequencial e 2 (dois) alunos com moderada preferência para Global.

Tabela 4.7: Estilos de aprendizagem Sequencial/Global encontrados com a aplicação do questionário ILS

Porcentagem obtida	Estilo Sequencial/Global
82,61% (19 alunos)	balanceado entre Sequencial/Global
8,70% (2 alunos)	moderada preferência para Sequencial
8,70% (2 alunos)	moderada preferência para Global

4.4 O Pré-processamento de dados sob a perspectiva dos estilos de aprendizagem

A estrutura do curso *online* preparatório para a OBI permite identificar apenas os estilos de aprendizagem Sequencial e Global, pois, a partir das atividades de vídeo-aula e questionários pode-se mensurar o progresso dos alunos durante o curso. Se os progressos são lineares há indicação de alunos com estilos de aprendizagem Sequenciais. Já se os progressos forem lineares e apresentarem um grande salto (alunos com notas lineares e de repente eles obtém notas muito boas), os alunos podem ser considerados com preferência para o estilo Global.

Para determinar os perfis de aprendizagem a partir dos resultados do pré-processamento, foram realizadas algumas mudanças na tabela de sumarização. Tais mudanças referem-se a manipulação dos campos de nota (*gradeQuiz1*, *gradeQuiz2*, *gradeQuiz3*, *gradeQuiz4*, *gradeQuiz5*, *gradeQuiz6*) para a obtenção dos progressos alcançados pelos alunos a medida que avançavam no curso.

A nova tabela de sumarização pode ser visualizada na tabela 4.8. Os campos relacionados a notas foram excluídos e foram acrescentados os campos que tratam dos progressos alcançados pelos alunos ao longo das atividades. A variável *progress1* trata do progresso obtido do módulo 1 para o módulo 2, a variável *progress2* trata do progresso obtido do módulo 2 para o módulo 3 e assim sucessivamente.

Tabela 4.8: Nova tabela de sumarização.

Tabelas utilizadas	Dados obtidos
mdl_course_modules, mdl_course_modules_completions	<i>numberOfModulesMade</i> - número de módulos feitos pelo usuário

Continua na próxima página.

Tabela 4.8 – Nova tabela de sumarização.

Tabelas utilizadas	Dados obtidos
mdl_quiz_attempts	<p><i>percentageOfAbandonment</i> - Porcentagem de abandono do usuário nos questionários</p> <p><i>percentOfCompleted</i> - Porcentagem de questionários finalizados pelo usuário</p> <p><i>averageTimeAnswering</i> - média de tempo que o usuário gastou respondendo aos questionários</p>
mdl_quiz_grades	<p><i>progress1,progress2,progress3,progress4,progress5</i> - Progressos obtidos pelos alunos de acordo com o avanço no curso.</p>

Continua na próxima página.

Tabela 4.8 – Nova tabela de sumarização.

Tabelas utilizadas	Dados obtidos
mdl_log	<p><i>actionAttempt</i> - Quantas vezes o usuário fez submissões em questionários</p> <p><i>actionCloseAttempt</i> - Quantas vezes o usuário encerrou uma tentativa de submissão (nota)</p> <p><i>actionContinueAttempt</i> - Quantas vezes o usuário continuou uma tentativa anterior não finalizada</p> <p><i>actionLogin</i> - Quantas vezes o usuário fez <i>login</i> no <i>Moodle</i> não finalizada</p> <p><i>actionPreView</i> - Quantas vezes o usuário pré-visualizou algum recurso</p> <p><i>actionReview</i> - Quantas vezes o usuário repetiu a visualização de um recurso</p> <p><i>actionView</i> - Quantas vezes o usuário visualizou um recurso</p> <p><i>actionViewAll</i> - Quantas vezes o usuário visualizou o ambiente de forma geral</p> <p><i>actionViewForum</i> - Quantas vezes o usuário visualizou o fórum</p> <p><i>actionViewSummary</i> - Quantas vezes o usuário visualizou o sumário do curso</p>

4.4.1 Resultados - Estilos de Aprendizagem Sequencial/Global

A tabela a seguir mostra os resultados obtidos através da aplicação do questionário ILS. Estes dados mostram que a grande maioria dos alunos (aproximadamente 83%) possuem estilo de aprendizagem balanceado e os 4 (quatro) outros alunos está igualmente dividido entre preferência moderada para Sequencial e Global.

Porcentagem obtida	Estilo Sequencial/Global
82,61% (19 alunos)	balanceado entre Sequencial/Global
8,70% (2 alunos)	moderada preferência para Sequencial
8,70% (2 alunos)	moderada preferência para Global

Os dados da nova tabela de sumarização (tabela 4.8) foram formatados para o tipo de arquivo requerido pelo WEKA. O algoritmo da rede neural de Kohonen foi executado sobre os dados com taxa de aprendizagem definida como 0.4. Os agrupamentos obtidos podem ser visualizados na figura 4.21, em que 91% dos alunos estão nos agrupamentos 0 e 3 e 9% dos alunos estão no agrupamento 2. O agrupamento 1 não possui instâncias.

	Clustered Instances (Instâncias agrupadas)
Agrupamento 0	12 (52%)
Agrupamento 2	2 (9%)
Agrupamento 3	9 (39%)

Figura 4.21: Figura retirada do apêndice A.3. Apresenta os agrupamentos obtidos pelo algoritmo da rede neural de Kohonen.

Os atributos considerados correspondem aos 19 atributos descritos na nova tabela de Sumarização (ver tabela 4.8) e estes atributos estão representados na figura 4.22.

```

Attributes: 19
            progress1
            progress2
            progress3
            progress4
            progress5
            numberOfModulesMade
            percentageOfAbandonment
            percentageOfCompleted
            averageTimeAnswering
            actionAttempt
            actionCloseAttempt
            actionContinueAttempt
            actionLogin
            actionPreView
            actionReview
            actionView
            actionViewAll
            actionViewForum
            actionViewSummary

```

Figura 4.22: Figura retirada do apêndice A.3. Contém os atributos que foram considerados para a análise.

A análise referente aos estilos de aprendizagem considera os dados de progresso de alunos que terminaram o curso, bem como baseia seus resultados nos estilos obtidos a partir da aplicação do questionário ILS. Na figura 4.23 pode-se observar que o número de alunos considerados é 23 (somente os que terminaram) e o número de atributos considerados é 19 (os mesmos atributos da nova tabela de sumarização).

```

Scheme:      weka.clusterers.Self0
Relation:    perfisDeAprendizagem
Instances:   23
Attributes: 19

```

Figura 4.23: Figura retirada do apêndice A.3. Apresenta o número de instâncias e atributos considerados na análise.

Pode-se observar que os alunos dos agrupamentos 0 (zero) e 3 (três) correspondem aos alunos que possuem os estilos de aprendizagem balanceados entre as dimensões Sequencial e Global. Tais alunos estão em agrupamentos diferentes

devido a sua atividade no ambiente virtual. Porém, seus progressos (variáveis *progress1*, *progress2*, *progress3*, *progress4*, *progress5* e *progress6*) possuem valores com desvio (campo *std. dev.*) em torno de 0.3, apresentando variações tanto em valores pequenos (indicando progressos lineares ou estilo Sequencial) quanto valores altos (progressos lineares que, de repente, assumem valores altos - indicando estilo Global). Estas informações podem ser obtidas na figura 4.24. O agrupamento 2 (dois) corresponde aos alunos que possuem moderada preferência para Global, visto os valores de seus desvios padrão (variável *std. dev.*).

	(0)		(2)	(3)
progress1				
value	0.0556	0.0213	0.5741	-0.1109
min	-0.3	NaN	0.4	-0.55
max	0.3	NaN	0.75	0.3
mean	0.0542	NaN	0.575	-0.1111
std. dev.	<u>0.2598</u>	NaN	<u>0.2475</u>	<u>0.3296</u>
progress2				
value	-0.0079	-0.0228	-0.5487	0.0347
min	-0.3	NaN	-0.8	-0.2
max	0.3	NaN	-0.3	0.1
mean	-0.0083	NaN	-0.55	0.0333
std. dev.	<u>0.1832</u>	NaN	<u>0.3536</u>	<u>0.1</u>
progress3				
value	-0.0408	-0.0184	0.0171	0.0031
min	-0.3	NaN	-0.0667	-0.2
max	0.2333	NaN	0.1	0.3333
mean	-0.0417	NaN	0.0167	0.0037
std. dev.	<u>0.177</u>	NaN	<u>0.1179</u>	<u>0.1695</u>
progress4				
value	0.0351	0.0255	0.4148	-0.0383
min	-0.3167	NaN	0.05	-0.5667
max	0.7	NaN	0.7833	0.4167
mean	0.0375	NaN	0.4167	-0.037
std. dev.	<u>0.2962</u>	NaN	<u>0.5185</u>	<u>0.2726</u>
progress5				
value	0.0094	0.0372	-0.2817	0.114
min	-0.6333	NaN	-0.6167	-0.3167
max	0.5167	NaN	0.05	0.8333
mean	0.0042	NaN	-0.2833	0.1111
std. dev.	<u>0.2999</u>	NaN	<u>0.4714</u>	<u>0.3631</u>

Figura 4.24: Figura retirada do apêndice A.3. Apresenta os valores de desvios dos progressos alcançados nos módulos.

5 CONCLUSÃO

O ambiente *Moodle* possui aproximadamente 250 tabelas em seu banco de dados e, para extrair determinada informação, deve ser feito um mapeamento das tabelas que ali estão. Assim, deve ocorrer um processo de refinamento para obter apenas informações relevantes aos propósitos do estudo.

Por ser este um projeto piloto, o pré-processamento de dados visou obter informações referentes ao acompanhamento do curso para, além de identificar focos de possíveis melhorias, encontrar ações que possam ser executadas sobre tais problemas encontrados. De posse dos resultados do pré-processamento, foi realizada a aplicação do questionário Índices de Estilos de Aprendizagem para detectar os perfis de aprendizagem presentes no curso proposto. Para confirmar as preferências de aprendizagem identificadas com o questionário, foram realizadas modificações na tabela de sumarização para que a mesma se adequasse aos estilos de aprendizagem encontrados. Em seguida, o algoritmo da rede neural de Kohonen foi executado no *software* WEKA e foi realizada a análise com base nos estilos de aprendizagem encontrados.

Os passos seguintes referem-se encontrar os outros estilos de aprendizagem presentes no curso proposto. Para tal, pode-se diversificar as atividades para que todas as dimensões sejam encontradas e, mais, pode-se também utilizar outros ambientes virtuais de aprendizagem para verificar quais os tipos de atividades que os mesmos possuem. Para a dimensão visual/verbal podem ser criadas atividades que se tenha como medir se o usuário entende melhor a informação se ele a visualiza (uma imagem interativa) ou o usuário entende melhor quando a informação é falada a ele. Para a dimensão sensorial/intuitiva podem ser criadas atividades que verifiquem qual o método de resolução que os alunos aplicam a ela. Se determinados alunos as resolvem utilizando-se de fatos e dados e empregam método

tradicional de resolução, então estes alunos podem ser sensoriais. Já se os alunos utilizam-se de princípios e teorias para resolução, então podem ser intuitivos. Para a dimensão Ativa/Reflexiva, podem ser criadas atividades que estimulem discussões sobre os tópicos do curso. Se determinado grupo de alunos lida melhor com estas discussões colaborativas, então podem ser ativos. Caso trabalhem a informação introspectivamente, então podem ser reflexivos.

REFERÊNCIAS BIBLIOGRÁFICAS

ANTONENKO, P. D.; TOY, S.; NIEDERHAUSER, D. S. Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, v. 60, n. 3, p. 383 – 398, 2012.

BAKER, R. S. J. d. Data mining for education. *International Encyclopedia of Education*, v. 7, n. 3, p. 112 – 118, 2010.

BAKER, R. S. J. d.; ISOTANI, S.; CARVALHO, A. de. Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v. 12, n. 2, p. 3 – 13, 2011.

BECHARA, J. J. B.; HAGUENAUER, C. J. Por uma aprendizagem adaptativa baseada na plataforma moodle. Laboratório de Pesquisa em Tecnologias da Informação e da Comunicação, Escola de Comunicação da Universidade Federal do Rio de Janeiro - LATEC/UFRJ, v. 4, n. 1, 2010. Disponível em: <<http://www.latec.ufrj.br/revistas/index%5B%5D=95>>.

CALDERS, T.; PECHENIZKIY, M. Introduction to the special section on educational data mining. *SIGKDD Explor. Newsl.*, ACM, New York, NY, USA, v. 13, n. 2, p. 3–6, maio 2012. ISSN 1931-0145. Disponível em: <<http://doi.acm.org/10.1145/2207243.2207245>>.

CARDON, A.; MÜLLER, D.; NAVAUX, P. *Introdução às Redes Neurais Artificiais*. 1994.

COSTA, L. A. C. da; FRANCO, S. R. K. Ambientes virtuais de aprendizagem e suas possibilidades construtivistas. 2005.

FELDER, R.; SOLOMAN, B. n. *Index of Learning Styles*. 2013. Acessado em 18 de agosto de 2013. Disponível em: <<http://www.ncsu.edu/felder/public/ILSpage.html>>.

FELDER, R. M. Author's preface to learning and teaching styles in engineering education [electronic version]. *Engineering Education*, v. 78, n. 7, p. 674–681, 2002. Disponível em: <<http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf>>.

FELDER, R. M.; SILVERMAN, L. Learning and teaching styles in engineering education. *Engineering Education*, v. 78, n. 7, p. 674–681, 1988.

FORNER, D. S. G.; TREVISOL, M. T. C. Significados e funções da avaliação da aprendizagem escolar do aluno do ensino fundamental: entendimento dos pais, professores e alunos. *III Colóquio Internacional de Educação e o I Seminário de Pesquisa sobre Indicadores de Qualidade do Ensino Fundamental (Projeto Unoesc / Programa Observatório da Educação – CAPES)*, 2010.

FREE SOFTWARE FOUNDATION. *GNU General Public License*. 2013. Acessado em 24 de março de 2013. Disponível em: <<http://www.gnu.org/licenses/gpl.html>>.

GOMES, M. J. E-learning: reflexões em torno do conceito. *Congresso Internacional sobre Tecnologias da Informação e Comunicação na Educação*, Centro de Competência da Universidade do Minho, n. 4, p. 229 – 236, 2005. Disponível em: <<http://hdl.handle.net/1822/2896>>.

GOMES, M. J. Problemáticas da avaliação em educação online. *Actas da Conferência Internacional de TIC na Educação : Challenges 2009*, Universidade do Minho, n. 6, p. 1675–1693, 2009. Disponível em: <<http://hdl.handle.net/1822-9420>>.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, nov 2009. ISSN 1931-0145. Disponível em: <<http://doi.acm.org/10.1145/1656274.1656278>>.

HAN, J.; KAMBER, M. *Data mining: concepts and techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.

HEATON, J. *Introduction to Neural Networks with Java*. 2nd. ed. Heaton Research, Inc., 2008. Disponível em: <<http://books.google.com.br/books?id=Swlcw7M4uD8C>>.

HERNANDEZ, B.; MONTANER, T.; SESE, F. J.; URQUIZU, P. The role of social motivations in e-learning: How do they affect usage and success of ict interactive tools? *Computers in Human Behavior*, v. 27, n. 6, p. 2224 – 2232, 2011.

HOFFMANN, J. *Avaliação: mito & desafio, uma perspectiva construtivista*. Mediação, 1998. Disponível em: <<http://books.google.com.br/books?id=TV2gOwAACAAJ>>.

LIMAYEM, M.; CHEUNG, C. M. K. Understanding information systems continuance: The case of internet-based learning technologies. *Information & Management*, v. 45, p. 227 – 232, 2008.

MARTINS, W. S. *Jogos de Lógica - Divirta-se e prepare-se para a: Olimpíada Brasileira de Informática*. Wellington Martins, 2011. Disponível em: <<http://books.google.com.br/books?id=GrYQywAACAAJ>>.

MOURA, D.; LÍVIO, L.; MAJEVSKI, R.; MARIA, R. *Tecnologias na EAD e o ambiente virtual de aprendizagem Moodle*. Planeta Educação, 2010. Disponível em: <<http://www.planetaeducacao.com.br/portal/artigo.asp?artigo=1866>>.

PIAZZA, A. D.; PIAZZA, M. C. D.; RAGUSA, A.; VITALE, G. Environmental data processing by clustering methods for energy forecast and planning.

Renewable Energy, v. 36, n. 3, p. 1063 – 1074, 2011. Disponível em:

<<http://www.sciencedirect.com/science/article/pii/S0960148110004234>>.

ROMERO, C.; VENTURA, S. Educational data mining: a review of the state of the art. *Trans. Sys. Man Cyber Part C*, IEEE Press, Piscataway, NJ, USA, v. 40, n. 6, p. 601–618, november 2010. ISSN 1094-6977.

ROMERO, C.; VENTURA, S.; GARCIA, H. Data mining in course management systems: Moodle case study and tutorial. *Computers e Education*, v. 51, n. 1, p.

368 – 384, 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0360131507000590>>.

SEN, B.; UÇAR, E.; DELEN, D. Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, v. 39, n. 10, p. 9468 – 9476, 2012. Disponível em:

<<http://www.sciencedirect.com/science/article/pii/S0957417412003752>>.

SO, H.-J.; CHOI, H.; LIM, W. Y.; XIONG, Y. Little experience with ict: Are they really the net generation student-teachers? *Computers & Education*, v. 59, n. 4,

p. 1234 – 1245, 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0360131512001297>>.

TAPSCOTT, D. *Grown up digital: how the net generation is changing the world*. New York: McGraw-Hill, 2009. 16 - 368 p. Disponível em:

<<http://nla.gov.au/nla.cat-vn4546527>>.

A APÊNDICE

A.1 Apêndice 1

O apêndice A.1 apresenta a saída fornecida pelo *software* WEKA, a qual refere-se ao pré-processamento dos dados de todos os alunos que participaram do curso *online*.

```
=== Run information ===
Scheme: weka.clusterers.SelfOrganizingMap -L 0.4 -O 200000 -C 100000 -H 2 -W 2
Relation: moodleData
Instances: 43
Attributes: 20
    gradeQuiz1
    gradeQuiz2
    gradeQuiz3
    gradeQuiz4
    gradeQuiz5
    gradeQuiz6
    numberOfModulesMade
    percentageOfAbandonment
    percentageOfCompleted
    averageTimeAnswering
    actionAttempt
    actionCloseAttempt
    actionContinueAttempt
    actionLogin
    actionPreView
    actionReview
    actionView
    actionViewAll
    actionViewForum
    actionViewSummary
Test mode: evaluate on training data

=== Clustering model (full training set) ===

Self Organized Map
=====
Number of clusters: 4

Attribute          Cluster
                   0          1          2          3
                   (15)        (4)        (16)        (8)
=====
gradeQuiz1
value              33.6067   18.8452   13.771    65.7624
min                -10       0         0         25
max                100      50        50        100
mean               34        18.75    14.0625   65.625
std. dev.          40.585   23.9357   20.3485   22.9031

gradeQuiz2
value              -1.496    27.5127   22.9716   52.162
min                -10      20         0         20
max                40       30         40        100
mean              -1.3333   27.5      23.125    52.5
std. dev.          16.4172   5         9.4648    34.5378

gradeQuiz3
value              -10       29.8577   22.0305   30.0485
min                -10     13.3333    0         10
max                -10      50         50         50
mean              -10      30         21.875    30
std. dev.          0        15.1535   13.7689   13.0931

gradeQuiz4
value              -10       8.1657    17.451    31.6823
min                -10     -10         0        -10
max                -10     33.3333   33.3333   73.3333
mean              -10      8.3333    17.5      31.25
std. dev.          0        21.8581   9.0677    29.9702
```

gradeQuiz5					
value	-10	-3.3672	25.8411	28.8393	
min	-10	-10	0	-10	
max	-10	16.6667	83.3333	91.6667	
mean	-10	-3.3333	26.0417	28.9583	
std. dev.	0	13.3334	20.8333	30.1838	
gradeQuiz6					
value	-10	-2.5381	21.5033	38.9914	
min	-10	-10	0	-10	
max	-10	20	60	100	
mean	-10	-2.5	21.25	38.75	
std. dev.	0	15	14.0831	40.1559	
numberOfModulesMade					
value	2.816	19.7411	27.0666	26.3782	
min	0	17	27	23	
max	16	27	28	27	
mean	2.8667	19.75	27.0625	26.375	
std. dev.	4.4218	4.8563	0.25	1.4079	
percentageOfAbandonment					
value	0	0	0	0	
min	0	0	0	0	
max	0	0	0	0	
mean	0	0	0	0	
std. dev.	0	0	0	0	
percentageOfCompleted					
value	0.6639	0.7405	0.885	0.9156	
min	0	0.5	0.6154	0.7273	
max	1	1	1	1	
mean	0.6611	0.7417	0.8855	0.9156	
std. dev.	0.4591	0.2115	0.1182	0.1022	
averageTimeAnswering					
value	39462.8067	559006.2995	8884.8002	179950.6342	
min	0	111	56	121	
max	603503	993080	86395	627258	
mean	40691	560784.25	9114	176112.875	
std. dev.	155697.9499	446473.5415	24005.6583	211452.1433	
actionAttempt					
value	1.5236	6.5151	7.9918	8.9862	
min	0	5	6	7	
max	5	9	13	12	
mean	1.5333	6.5	8	9	
std. dev.	1.5523	1.7321	2.0656	1.8516	
actionCloseAttempt					
value	1.2598	4.7538	6.8662	8.1118	
min	0	3	6	7	
max	5	6	8	10	
mean	1.2667	4.75	6.875	8.125	
std. dev.	1.3345	1.5	0.8062	1.2464	
actionContinueAttempt					
value	5.0084	15.7993	15.1836	18.4762	
min	0	9	7	15	
max	11	28	22	24	
mean	5	15.75	15.1875	18.5	
std. dev.	3.6645	8.4212	3.8161	3.0706	
actionLogin					
value	2.4305	4.9875	6.17	11.393	
min	0	4	2	6	
max	7	7	11	17	
mean	2.4667	5	6.1875	11.375	
std. dev.	2.1996	1.4142	2.5091	3.5431	

actionPreView				
value	1.414	1.5101	0.3183	13.5273
min	0	0	0	0
max	6	4	3	40
mean	1.4	1.5	0.3125	13.375
std. dev.	1.8048	1.7321	0.7932	14.9087
actionReview				
value	1.3251	5.0025	6.9316	9.4828
min	0	3	6	7
max	5	7	9	15
mean	1.3333	5	6.9375	9.5
std. dev.	1.3452	1.8257	0.9287	2.7255
actionView				
value	16.0829	50.216	76.7811	99.3417
min	0	39	59	53
max	55	66	106	147
mean	16.1333	50.25	76.9375	99.125
std. dev.	15.6883	11.3248	13.0152	26.9467
actionViewAll				
value	0.9325	0	0.0666	0
min	0	0	0	0
max	14	0	1	0
mean	0.9333	0	0.0625	0
std. dev.	3.6148	0	0.25	0
actionViewForum				
value	0.2039	0	0	0.7386
min	0	0	0	0
max	2	0	0	6
mean	0.2	0	0	0.75
std. dev.	0.5606	0	0	2.1213
actionViewSummary				
value	1.7379	6.7563	8.0425	10.5076
min	0	5	6	9
max	6	9	12	13
mean	1.7333	6.75	8.0625	10.5
std. dev.	1.7099	1.7078	1.652	1.1952

Time taken to build model (full training data) : 78.32 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	15 (35%)
1	4 (9%)
2	16 (37%)
3	8 (19%)

A.2 Apêndice 2

O apêndice A.2 apresenta a saída fornecida pelo *software* WEKA, a qual refere-se ao pré-processamento dos dados apenas dos alunos que finalizaram todas as atividades requeridas pelo curso.

```

=== Run information ===
Scheme:      weka.clusterers.SelfOrganizingMap -L 0.4 -O 200000 -C 100000 -H 2 -W 2
Relation:    moodleData
Instances:   23
Attributes:  20
              gradeQuiz1
              gradeQuiz2
              gradeQuiz3
              gradeQuiz4
              gradeQuiz5
              gradeQuiz6
              numberOfModulesMade
              percentageOfAbandonment
              percentageOfCompleted
              averageTimeAnswering
              actionAttempt
              actionCloseAttempt
              actionContinueAttempt
              actionLogin
              actionPreView
              actionReview
              actionView
              actionViewAll
              actionViewForum
              actionViewSummary
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

Self Organized Map
=====
Number of clusters: 4

Attribute          Cluster
                   0          1          2          3
                   (4)        (10)       (5)        (4)
=====
gradeQuiz1
value              12.4994   2.4864   65.2      50
min                0          0        25        50
max                25         25       75        50
mean               12.5      2.5      65        50
std. dev.         14.4338   7.9057   22.3607   0
gradeQuiz2
value              17.4376   25.9404  51.8402   22.4873
min                0          10       20        20
max                30         40       100       30
mean               17.5      26       52        22.5
std. dev.         12.5831   8.4327   35.6371   5
gradeQuiz3
value              32.5632   21.9879  36.1202   15.0246
min                20         10       20        0
max                50         50       50        30
mean               32.5      22       36        15
std. dev.         12.5831   12.2927  11.4018   12.9099
gradeQuiz4
value              16.6834   19.2769  45.6682   14.9579
min                13.3333   0        13.3333   0
max                20        33.3333  73.3333  33.3333
mean               16.6666  19.3333  45.3333  15
std. dev.         3.849     9.6609   27.2437  14.7824

```

gradeQuiz5				
value	37.2701	21.649	39.7497	18.8653
min	8.3333	0	16.6667	8.3333
max	83.3333	41.6667	91.6667	33.3333
mean	37.5	21.6667	40	18.75
std. dev.	36.3241	13.1468	30.8446	12.5
gradeQuiz6				
value	30.1513	21.0168	64.2814	10.05
min	20	10	30	0
max	60	40	100	20
mean	30	21	64	10
std. dev.	20	9.9443	25.0998	8.165
numberOfModulesMade				
value	27	27.1035	27	27
min	27	27	27	27
max	27	28	27	27
mean	27	27.1	27	27
std. dev.	0	0.3162	0	0
percentageOfAbandonment				
value	0	0	0	0
min	0	0	0	0
max	0	0	0	0
mean	0	0	0	0
std. dev.	0	0	0	0
percentageOfCompleted				
value	0.7401	0.9458	0.9663	0.9082
min	0.6154	0.8571	0.8333	0.8571
max	0.875	1	1	1
mean	0.7393	0.9464	0.9667	0.9087
std. dev.	0.1193	0.0694	0.0745	0.0627
averageTimeAnswering				
value	1670.5081	96091.7788	157331.2571	34297.0648
min	56	119	34708	137
max	6522	993080	321815	86395
mean	1694.75	99563.6	156277.4	34156.5
std. dev.	3218.2705	313949.6819	106705.5055	41885.0161
actionAttempt				
value	10.7337	6.801	8.796	7.9999
min	8	6	7	7
max	13	8	12	9
mean	10.75	6.8	8.8	8
std. dev.	2.2174	0.7888	2.1679	1.1547
actionCloseAttempt				
value	7.7462	6.397	8.392	7.2462
min	7	6	7	6
max	8	7	10	8
mean	7.75	6.4	8.4	7.25
std. dev.	0.5	0.5164	1.5166	0.9574
actionContinueAttempt				
value	19.7337	13.2967	16.6	16.7436
min	17	7	15	13
max	22	18	18	24
mean	19.75	13.3	16.6	16.75
std. dev.	2.2174	3.1287	1.1402	4.9917
actionLogin				
value	3.9849	6.5374	11.0342	9.2185
min	2	2	6	5
max	6	9	17	14
mean	4	6.5	11	9.25
std. dev.	1.633	2.1731	4.4159	4.0311

actionPreView					
value	0	0.5035	20.9507	0.2538	
min	0	0	0	0	
max	0	3	40	1	
mean	0	0.5	20.8	0.25	
std. dev.	0	0.9718	14.2724	0.5	
actionReview					
value	7.9975	6.4935	9.574	7.2462	
min	7	6	7	6	
max	9	7	15	8	
mean	8	6.5	9.6	7.25	
std. dev.	0.8165	0.527	3.4351	0.9574	
actionView					
value	78.3193	76.8595	98.9318	80.2006	
min	59	65	53	70	
max	106	106	147	106	
mean	78.5	76.8	98.6	80.25	
std. dev.	19.7737	12.4258	34.4137	17.2119	
actionViewAll					
value	0	0.1035	0	0	
min	0	0	0	0	
max	0	1	0	0	
mean	0	0.1	0	0	
std. dev.	0	0.3162	0	0	
actionViewForum					
value	0	0	0	0	
min	0	0	0	0	
max	0	0	0	0	
mean	0	0	0	0	
std. dev.	0	0	0	0	
actionViewSummary					
value	9.2538	7.3811	10.8121	8.4975	
min	8	6	9	8	
max	11	12	13	10	
mean	9.25	7.4	10.8	8.5	
std. dev.	1.5	1.7764	1.4832	1	

Time taken to build model (full training data) : 42.83 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	4 (17%)
1	10 (43%)
2	5 (22%)
3	4 (17%)

A.3 Apêndice 3

O apêndice A.3 apresenta a saída fornecida pelo *software* WEKA referente aos estilos de aprendizagem Sequencial e Global encontrados no curso preparatório para a OBI.

```

=== Run information ===
Scheme:      weka.clusterers.SelfOrganizingMap -L 0.4 -O 200000 -C 100000 -H 2 -W 2
Relation:    perfisDeAprendizagem
Instances:   23
Attributes:  19
              progress1
              progress2
              progress3
              progress4
              progress5
              numberOfModulesMade
              percentageOfAbandonment
              percentageOfCompleted
              averageTimeAnswering
              actionAttempt
              actionCloseAttempt
              actionContinueAttempt
              actionLogin
              actionPreView
              actionReview
              actionView
              actionViewAll
              actionViewForum
              actionViewSummary
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

Self Organized Map
=====
Number of clusters: 4

Attribute          Cluster
                   0          1          2          3
                   (12)        (0)        (2)        (9)
=====
progress1
value              0.0556      0.0213      0.5741     -0.1109
min                -0.3        NaN         0.4        -0.55
max                0.3        NaN         0.75       0.3
mean              0.0542     NaN         0.575      -0.1111
std. dev.         0.2598     NaN         0.2475     0.3296

progress2
value             -0.0079     -0.0228     -0.5487     0.0347
min               -0.3        NaN         -0.8        -0.2
max               0.3        NaN         -0.3        0.1
mean             -0.0083     NaN         -0.55       0.0333
std. dev.        0.1832     NaN         0.3536     0.1

progress3
value            -0.0408     -0.0184     0.0171     0.0031
min              -0.3        NaN         -0.0667     -0.2
max              0.2333     NaN         0.1         0.3333
mean            -0.0417     NaN         0.0167     0.0037
std. dev.       0.177       NaN         0.1179     0.1695

progress4
value            0.0351     0.0255     0.4148     -0.0383
min             -0.3167     NaN         0.05        -0.5667
max              0.7        NaN         0.7833     0.4167
mean            0.0375     NaN         0.4167     -0.037
std. dev.       0.2962     NaN         0.5185     0.2726

progress5

```

value	0.0094	0.0372	-0.2817	0.114
min	-0.6333	NaN	-0.6167	-0.3167
max	0.5167	NaN	0.05	0.8333
mean	0.0042	NaN	-0.2833	0.1111
std. dev.	0.2999	NaN	0.4714	0.3631
numberOfModulesMade				
value	27	27.0469	27	27.1156
min	27	NaN	27	27
max	27	NaN	27	28
mean	27	NaN	27	27.1111
std. dev.	0	NaN	0	0.3333
percentageOfAbandonment				
value	0	0	0	0
min	0	NaN	0	0
max	0	NaN	0	0
mean	0	NaN	0	0
std. dev.	0	NaN	0	0
percentageOfCompleted				
value	0	0.4647	1	1
min	0	NaN	1	1
max	0	NaN	1	1
mean	0	NaN	1	1
std. dev.	0	NaN	0	0
averageTimeAnswering				
value	16253.7714	82463.7503	17477.8492	185126.9313
min	56	NaN	420	183
max	100866	NaN	34708	993080
mean	16286.5	NaN	17564	187762.4444
std. dev.	36304.2705	NaN	24245.2773	321648.0928
actionAttempt				
value	9.1501	8.0969	8.4925	6.6688
min	7	NaN	7	6
max	13	NaN	10	8
mean	9.1667	NaN	8.5	6.6667
std. dev.	2.1249	NaN	2.1213	0.7071
actionCloseAttempt				
value	7.4078	7.1628	8.4925	6.6688
min	6	NaN	7	6
max	10	NaN	10	8
mean	7.4167	NaN	8.5	6.6667
std. dev.	1.1645	NaN	2.1213	0.7071
actionContinueAttempt				
value	17.4519	15.7103	17	13.1389
min	13	NaN	17	7
max	24	NaN	17	18
mean	17.5	NaN	17	13.1111
std. dev.	3.451	NaN	0	3.1798
actionLogin				
value	6.2933	7.5412	4.995	9.682
min	2	NaN	4	5
max	14	NaN	6	17
mean	6.3333	NaN	5	9.6667
std. dev.	3.3394	NaN	1.4142	3.7749
actionPreView				
value	1.6732	4.9363	0	10.1162
min	0	NaN	0	0
max	19	NaN	0	40
mean	1.6667	NaN	0	10
std. dev.	5.4661	NaN	0	14.6969
actionReview				

value	7.5778	7.4214	10.9799	6.7766
min	6	NaN	7	6
max	11	NaN	15	8
mean	7.5833	NaN	11	6.7778
std. dev.	1.4434	NaN	5.6569	0.6667
actionView				
value	80.4967	82.9456	62.0452	89.3015
min	59	NaN	53	66
max	106	NaN	71	147
mean	80.8333	NaN	62	89.1111
std. dev.	16.7865	NaN	12.7279	25.8962
actionViewAll				
value	0	0.0469	0	0.1156
min	0	NaN	0	0
max	0	NaN	0	1
mean	0	NaN	0	0.1111
std. dev.	0	NaN	0	0.3333
actionViewForum				
value	0	0	0	0
min	0	NaN	0	0
max	0	NaN	0	0
mean	0	NaN	0	0
std. dev.	0	NaN	0	0
actionViewSummary				
value	8.4046	8.6028	11.5025	8.3544
min	6	NaN	11	6
max	11	NaN	12	13
mean	8.4167	NaN	11.5	8.3333
std. dev.	1.6765	NaN	0.7071	2.1794

Time taken to build model (full training data) : 39.17 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	12 (52%)
2	2 (9%)
3	9 (39%)

A.4 Apêndice 4

No apêndice A.4 pode ser visualizado o documento *Termo de Autorização de Uso de Imagem e Depoimentos*. Neste documento, os alunos de graduação que gravaram as vídeo-aulas autorizam a exibição de sua imagem aos alunos inscritos no curso.



TERMO DE AUTORIZAÇÃO DE USO DE IMAGEM E DEPOIMENTOS

Eu _____, CPF _____, RG _____, após conhecer e entender os objetivos, procedimentos metodológicos, riscos e benefícios da pesquisa, bem como de estar ciente da necessidade do uso de minha imagem e/ou depoimento, especificados no Termo de Consentimento Livre e Esclarecido (TCLE), AUTORIZO, através do presente termo, o pesquisador Eric Fernandes de Mello Araújo do projeto de pesquisa intitulado “Avaliação de aprendizagem de alunos em sistema à distância como instrumento de aprendizagem para a Olimpíada Brasileira de Informática (OBI)” a realizar os vídeos/imagens que se façam necessárias e/ou a colher meu depoimento sem quaisquer ônus financeiros a nenhuma das partes.

Ao mesmo tempo, libero a utilização destas imagens para fins científicos e de estudos (vídeos, slides e transparências), em favor dos pesquisadores da pesquisa, acima especificados, obedecendo ao que está previsto nas Leis que resguardam os direitos das crianças e adolescentes (Estatuto da Criança e do Adolescente – ECA, Lei Nº 8.069/1990), dos idosos (Estatuto do Idoso, Lei Nº 10.741/2003) e

das pessoas com deficiência (Decreto Nº 3.298/1999, alterado pelo Decreto Nº 5.296/2004).

Lavras, 22 de fevereiro de 2013

Pesquisador responsável pelo projeto

Sujeito da Pesquisa

A.5 Apêndice 5

No apêndice A.5 pode ser visualizado o documento *Termo de Consentimento Livre e Esclarecido - TCLE*. Este documento permite que os responsáveis pelos menores inscritos no curso conheçam a pesquisa, entendam os objetivos e autorizem a participação de seus filhos/sobrinhos/etc. no do curso preparatório para a OBI.

Projeto “Avaliação de aprendizagem de alunos em sistema à distância como instrumento de aprendizagem para a Olimpíada Brasileira de Informática (OBI)”

Pesquisador responsável (professor orientador): Eric Fernandes de Mello Araújo

Este documento que você está lendo é chamado de Termo de Consentimento Livre e Esclarecido (TCLE). Ele contém explicações sobre o estudo que seu filho (a)/sobrinho (a)/ etc. está sendo convidado a participar. Antes de decidir se autoriza a participação (de livre e espontânea vontade), você deverá ler e compreender todo o conteúdo. Ao final, caso decida autorizar a participação, você irá assinar este termo e receberá uma cópia do mesmo. Antes de assinar faça perguntas sobre tudo o que não tiver bem entendido. A equipe deste estudo responderá às suas perguntas a qualquer momento (antes, durante e após o estudo) através de uma página de FAQ.

Natureza e objetivos do estudo

O projeto de avaliação de aprendizado em treinamento para Olimpíada Brasileira de Informática (OBI) é um curso que será disponibilizado via plataforma Moodle para dar aulas virtuais, que irá preparar os alunos do ensino fundamental e médio para a prova da OBI. Durante o processo de aulas, serão coletados dados dos resultados obtidos, através de exercícios e ações dos participantes no ambiente virtual. Esses dados passarão por um processo de mineração de dados

em e-learning, para verificar o tipo de aprendizagem que se encaixa melhor com cada aluno.

Justificativa

A Olimpíada Brasileira de Informática, realizada anualmente desde 1999, é promovida pela Sociedade Brasileira de Computação (SBC) em conjunto com a Fundação Carlos Chagas (FCC). Ela tem o intuito de despertar nos mais jovens o interesse pela área da ciência da computação, seguindo os moldes das Olimpíadas de matemática, física e astronomia. A principal razão da pesquisa está na melhoria do método de ensino para alunos que utilizam a internet para aprendizado, visto que o número de cursos online disponíveis só tende a crescer com o aumento da disponibilidade de acesso à internet por parte da população. Procedimentos do estudo

Sua colaboração consiste em participar do curso online preparatório para a Olimpíada Brasileira de Informática. Os procedimentos se referem a assistir vídeo-aulas de conteúdo que serão disponibilizadas e realizar exercícios referentes ao conteúdo apresentado. Não haverá nenhuma outra forma de envolvimento ou comprometimento neste estudo. O curso para avaliação de aprendizado em treinamento para a OBI será disponibilizado via internet, e os alunos poderão fazer de suas residências, não implicando um local exato de pesquisa. São aproximadamente 200 alunos, na faixa etária de 12 a 15 anos, sem descrição de gênero que participarão do curso. Qualquer aluno nessa faixa-etária ou que esteja cursando o ensino fundamental (a partir da 6^a série) poderá participar.

Riscos e benefícios

A fonte de obtenção de material de estudo, são os resultados dos alunos, em aulas e exercícios, durante o curso, sendo assim, não existe nenhum risco aparente para a saúde do sujeito. O único risco eminente é a divulgação pública individualizada dos dados a serem usados na pesquisa, mas esse risco está controlado, pois

será mantida a confidencialidade dos dados sem divulgar nomes. O curso será disponibilizado em módulos em que se começa um novo módulo ao concluir todas as atividades anteriores. O aluno poderá fazer os módulos em seu tempo. Ao participar, seu filho(a)/sobrinho(a), etc., irá colaborar com o sucesso da pesquisa e ao mesmo tempo terá um enorme conhecimento a seu dispor que o(a) preparará para um evento tão interessante como a Olimpíada Brasileira de Informática.

Participação, recusa e direito de se retirar do estudo

O responsável pelo menor ou o próprio sujeito tem a liberdade de retirar seu consentimento a qualquer momento e deixar de participar do estudo, sem qualquer prejuízo ao atendimento a que está sendo ou será submetido. Conforme previsto pelas normas brasileiras de pesquisa com a participação de seres humanos não haverá nenhum tipo de compensação financeira pela participação neste estudo.

Confidencialidade

Os dados serão manuseados somente pelos pesquisadores e não será permitido o acesso a outras pessoas. O material com as informações do aluno (dados registrados no Moodle) ficarão guardados sob a responsabilidade do pesquisador responsável, com a garantia de manutenção do sigilo e confidencialidade. Os resultados deste trabalho serão apresentados (Pesquisas, Artigos, Trabalhos de Conclusão de Curso, etc.). Entretanto, serão mostrados apenas os resultados obtidos como um todo, sem revelar nome ou qualquer informação que esteja relacionada com a privacidade do mesmo.

Critérios para Suspender ou Encerrar a Pesquisa

Os resultados deste trabalho serão apresentados, não existindo, portanto, critério de encerramento. O curso continuará disponível na internet para que outros tenham acesso a esse conhecimento.

Consentimento

Eu _____, responsável pelo menor _____, certifico que, tendo lido as informações acima e suficientemente esclarecido (a) de todos os itens, estou plenamente de acordo com a realização do experimento. Assim, eu autorizo a execução do trabalho de pesquisa exposto acima.

Lavras, _____ de _____ de 2013.

NOME (legível) _____ RG _____

Pesquisador(a) responsável, (35) 3829-1535

Pesquisador(a) auxiliar, (35) 9169-7741

ATENÇÃO: A sua participação em qualquer tipo de pesquisa é voluntária. Em caso de dúvida quanto aos seus direitos escreva para o Comitê de Ética em Pesquisa em seres humanos da UFLA. Endereço – Campus Universitário da UFLA, Pró-reitoria de pesquisa, COEP, caixa postal 3037. Telefone: 3829-1127, falar com Andréa. No caso de qualquer emergência entrar em contato com o pesquisador responsável no Departamento de Ciência da Computação. Telefones de contato: 035 3829 1535.

A.6 Apêndice 6

Questionário Índices de Estilos de Aprendizagem

Instruções

Escolha apenas uma resposta para cada questão. Se as alternativas “a” e a “b” se aplicam a você, escolha a que melhor se aplica.

1. Eu entendo algo melhor depois que eu:
 - a) experimento.
 - b) penso sobre isso.

2. Eu poderia ser considerado(a):
 - a) realista.
 - b) inovador.

3. Quando eu penso sobre o que fiz ontem, sou mais propenso(a) a pensar sobre:
 - a) uma imagem.
 - b) palavras.

4. Eu tendo a:
 - a) compreender os detalhes de um assunto mas pode ser confuso entender sobre sua estrutura global.
 - b) compreender a estrutura como um todo mas pode ser confuso entender detalhes.

5. Quando estou aprendendo algo novo, isso me ajuda a:
 - a) falar sobre isso.
 - b) pensar sobre isso.

6. Se eu tenho um professor, eu prefiro que ele(a) dê aulas:
- a) que lide com fatos e situações da vida real.
 - b) que lide com ideias e teorias.
7. Eu prefiro obter novas informações em:
- a) figuras, diagramas, gráficos ou mapas.
 - b) escrevendo instruções ou informação verbal.
8. Umavez que eu entendo:
- a) todas as partes, eu entendo a coisa toda.
 - b) a coisa toda, eu vejo como as partes se encaixam.
9. Em um estudo de grupo trabalhando com assuntos difíceis, eu sou mais propenso(a) a:
- a) entrar e contribuir com ideias.
 - b) sentar e ouvir.
10. Eu acho que é mais fácil:
- a) para aprender fatos.
 - b) para aprender conceitos.
11. Em um livro com muitas figuras e gráficos, eu sou mais propenso(a) a:
- a) olhar as figuras e gráficos cuidadosamente.
 - b) focar no texto escrito.
12. Quando eu resolvo problemas de matemática:
- a) Eu costumo ter minha própria solução, um passo de cada vez.
 - b) Muitas vezes eu só vejo as soluções, mas depois tenho que me esforçar pra descobrir os passos para chegar nessas soluções.

13. Nas aulas eu:

- a) Eu geralmente comecei a conhecer muitos dos alunos.
- b) Eu raramente cheguei a conhecer muitos dos alunos.

14. Em uma leitura que envolve fatos reais, eu prefiro:

- a) algo que ensine novos fatos ou me diga como fazer alguma coisa.
- b) algo que me dê novas ideias para pensar.

15. Eu gosto de professores:

- a) que colocam muitos diagramas no quadro.
- b) que gastam muito tempo explicando.

16. Quando eu estou analisando uma história:

- a) eu penso nos incidentes e tento colocá-los juntos para solucionar o assunto.
- b) eu somente sei quais são os temas quando termino a leitura e então tenho que voltar e buscar os incidentes que o demonstram.

17. Quando eu começo uma lição de casa, eu sou mais propenso(a) a:

- a) começar a pensar na solução imediatamente.
- b) tentar entender completamente o problema primeiro.

18. Eu prefiro a noção de:

- a) certeza.
- b) teoria.

19. Eu lembro melhor:

- a) o que eu vejo.
- b) o que eu escuto.

20. Para mim, é mais importante que um instrutor:

- a) coloque o material em etapas sequenciais que sejam claras.
- b) me dê uma visão global e relacione com outros assuntos.

21. Eu prefiro estudar:

- a) em grupo de estudos.
- b) sozinho.

22. Eu sou mais propenso a ser considerado(a):

- a) cuidadoso sobre os detalhes do meu trabalho.
- b) criativo sobre como fazer o meu trabalho.

23. Quando eu tenho instruções sobre um novo lugar, eu prefiro:

- a) um mapa.
- b) instruções escritas.

24. Eu aprendo:

- a) em um ritmo bastante regular, se eu estudar bastante, eu “pego o jeito”.
- b) aos trancos e barrancos, estarei totalmente confuso(a) e, de repente, terei “cliques” e entenderei.

25. Eu prefiro primeiro:

- a) experimentar sobre as coisas.
- b) pensar sobre como vou fazer isso.

26. Quando eu leio por lazer, eu gosto de escritores que:

- a) dizem claramente o que querem dizer.
- b) dizem coisas criativas, de maneiras interessantes.

27. Quando eu vejo um diagrama ou um esboço na sala de aula, eu sou mais propenso(a) a lembrar:

- a) a imagem.
- b) o que o professor falou sobre aquilo.

28. Quando estou considerando um conjunto de informações, sou mais propenso(a) a:

- a) concentrar em detalhes e não ter visão da informação como um todo.
- b) tentar entender a informação como um todo antes de entrar em detalhes.

29. Eu lembro com mais facilidade:

- a) algo que eu tenha feito.
- b) algo que eu tenha pensado muito sobre.

30. Quando eu tenho que executar uma tarefa, eu prefiro:

- a) dominar uma maneira de fazê-la.
- b) chegar com novas maneira de fazê-la.

31. Quando algo é mostrado a mim, eu prefiro:

- a) ilustrações e gráficos.
- b) texto resumindo os resultados.

32. Quando escrevo um texto, eu sou mais propenso(a) a:

- a) trabalhar (pensar sobre ou escrever) o início do texto e progredir.
- b) trabalhar (pensar sobre ou escrever) diferentes partes do texto e então ordená-las.

33. Quando tenho que fazer um projeto em grupo, eu primeiro quero:

- a) ter uma reunião para debates, onde todos contribuem com ideias.
- b) ter um debate individual e então apresentar as ideias ao grupo e comparar com as ideias das outras p

34. Eu considero um elogio chamar alguém de:
- a) imaginativo.
 - b) imaginativo.
35. Quando eu conheço pessoas em uma festa, eu sou mais propenso(a) a lembrar:
- a) como eles são.
 - b) o que eles dizem sobre si mesmos.
36. Quando eu aprendo sobre um novo assunto, eu prefiro:
- a) ficar concentrado no assunto, aprendendo o máximo que eu puder.
 - b) tentar estabelecer conexões entre o assunto e assuntos relacionados.
37. Eu sou mais propenso(a) a ser considerado(a):
- a) extrovertido.
 - b) reservado.
38. Eu prefiro cursos que enfatizem:
- a) assunto concreto (fatos, dados).
 - b) assunto abstrato (conceitos, teoria).
39. Para diversão, eu preferiria mais:
- a) ver televisão.
 - b) ler um livro.
40. Alguns professores começam suas aulas com um roteiro do que vão ensinar. Esse roteiro:
- a) é algo pouco útil pra mim.
 - b) é algo muito útil pra mim.

41. A ideia de dever de casa em grupo, com uma nota inteira para o grupo:
- a) é interessante para mim.
 - b) não é interessante para mim.
42. Quando tenho que fazer contas grandes:
- a) eu tendo a repetir todos os passos e testar o resultado com cuidado.
 - b) eu acho cansativo verificar o resultado e me forço a fazer isso.
43. Eu tendo a desenhar lugares:
- a) de forma fácil e com fidelidade a lugares reais.
 - b) com dificuldade e sem muitos detalhes.
44. Quando estou resolvendo problemas em grupo, eu sou mais propenso(a) a:
- a) pensar nos passos para achar uma solução.
 - b) pensar nas possíveis consequências ou aplicações da solução em uma variedade de áreas.