

---

# Aplicação de Técnicas de Mineração de Dados no Processo de Aprendizagem na Educação a Distância

Maxwel Macedo Dias<sup>1</sup>, Luiz Alberto da Silva Filho<sup>2</sup>, Adriano Del Pino Lino<sup>1</sup>,  
Eloi Luiz Favero<sup>2</sup>, Edson Marcos Leal Soares Ramos<sup>3</sup>

<sup>1</sup> Faculdade de Informática – Universidade Federal do Pará (UFPA)

<sup>2</sup> Programa de Pós Graduação em Ciência da Computação (PPGCC-UFPA)

<sup>3</sup> Programa de Pós Graduação em Matemática e Estatística (PPGME-UFPA)

{maxwelmdias, lasf.bel}@gmail.com, {adrianod, favero1, edson}@ufpa.br

**Abstract.** *The use of e-learning environment has grown too much in lately, however, it is one of the greatest difficulties is in accompany the learning of students. In that context, this paper has the objective to analyze data e-learning environment through of the application of Data Mining techniques to find relevant information about the student's profile with the use of this technology.*

**Resumo.** *A utilização do Ensino a Distância a partir Internet tem crescido bastante nos últimos anos, no entanto, uma das maiores dificuldades está em acompanhar o aprendizado dos estudantes. Nesse contexto, este artigo tem o objetivo de analisar os dados de um ambiente de ensino-aprendizagem a partir da aplicação de técnicas de Mineração de Dados para descobrir informações relevantes sobre o perfil do aluno com relação à utilização dessa tecnologia.*

## 1. Introdução

A utilização da Internet tem crescido demasiadamente nos últimos anos, proporcionando a disseminação dos mais variados tipos de informações e serviços, como por exemplo, comércio eletrônico, sites de relacionamentos, bibliotecas digitais, educação a distância, dentre outros. A acessibilidade e a facilidade do uso de ferramentas para manipular os recursos da *web* tem tornado esta tecnologia uma escolha para educação a distância. [Machado e Becker, 2002].

Segundo Souza (2007), a Educação a Distância (EaD) é uma forma de organização de ensino-aprendizagem na qual alunos estudam, quer em grupo, quer individualmente em seus lares, locais de trabalho ou outros lugares com materiais auto-instrutivos distribuídos por meios de comunicação, possibilitando a comunicação com docentes, monitores ou outros alunos.

Os educadores deste novo processo de aprendizagem utilizam estes ambientes para disponibilizar informações *online*, porém possuem pouco suporte para avaliar e discriminar os diferentes comportamentos dos alunos sobre o ambiente de ensino virtual e a forma de execução das atividades *online* propostas durante a realização dos cursos [Zaiane e Luo, 2001].

---

---

A Mineração de Dados, definida como a descoberta de conhecimento não-trivial e útil em grandes bases de dados [Goldschmidt e Passos, 2005], apresenta variedade de aplicabilidade de suas técnicas e tarefas a partir da base de dados dos ambientes de EaD. Tais dados são provenientes, geralmente, do histórico dos acessos aos recursos do sistema pelos alunos, suas avaliações, comunicação (*chat* e *e-mail*) entre alunos e entre alunos e professor, tempo utilizando o sistema, dentre outros.

Neste sentido, o objetivo deste trabalho é analisar os dados obtidos a partir um ambiente de ensino-aprendizagem, aplicando técnicas de Mineração de Dados para descobrir informações relevantes sobre o perfil do aluno com relação à utilização dessa tecnologia.

Além desta Seção introdutória, este artigo está organizado como segue: na Seção 2 é apresentado uma visão geral dos ambientes de ensino-aprendizagem, sendo enfatizado o ambiente LabSQL; na Seção 3 é abordada a definição de Mineração de Dados, sendo enfatizadas as técnicas de Árvore de Decisão e Redes Bayesianas; na Seção 4 são mostrados os resultados da aplicação das técnicas de Mineração de Dados; os trabalhos relacionados são vistos na Seção 5 e, finalmente, na Seção 6 são realizadas as conclusões.

## **2. Ambientes de Ensino-Aprendizagem**

O ambiente de aprendizagem ou Ambiente de Educação a Distância é um sistema que fornece suporte a qualquer tipo de atividade realizada pelo aluno, isto é, um conjunto de ferramentas que são utilizadas em diferentes situações do processo de aprendizagem. Estes ambientes têm o objetivo de apoiar classes de usuários por meio da Internet, sendo útil para usuários que não residem perto de instituições de ensino, ou não dispõem de horários regulares para estudar [Martins e Campestrini, 2004]. Na Subseção 2.1 é descrito o ambiente LabSQL, estudo de caso deste trabalho.

### **2.1. LabSQL**

O LabSQL é um ambiente interativo para auxiliar os alunos no aprendizado da linguagem SQL e pode ser utilizado como ferramenta de apoio ao mediador para realizar automaticamente as avaliações nas atividades de laboratório [Lino *et al.*, 2007].

No ambiente de aprendizagem de SQL, o aprendiz visualiza o texto didático acompanhado de exemplos executáveis. Juntamente com o conteúdo são apresentadas listas de exercícios para que o aprendiz treine suas habilidades. Existem três tipos de exercícios: objetivos de múltipla escolha (ou V/F); não objetivos descritivos e exercícios de programação.

No momento em que o aprendiz interage com o sistema, enviando sua consulta SQL, o sistema executa e avalia a complexidade desta consulta em relação à consulta do mediador. Dessa forma, o aprendiz pode receber o *feedback* automático, contendo: o resultado da consulta, permitindo avaliar se a resposta está correta ou não; a avaliação automática da resposta do aprendiz, levando em consideração o resultado da execução e o grau de complexidade comparado com a resposta do mediador; o número de tentativas; e a avaliação global da prova ou exercício.

Além do *feedback* para o aprendiz, é gerado um relatório detalhado para o mediador, contendo as informações de cada aprendiz e da turma em geral; permite

---

---

visualizar a avaliação de cada questão resolvida por aprendizes e identificar os aprendizes com dificuldade de concluir os exercícios. Por exemplo, o ambiente mostra os alunos que já tentaram mais de 10 vezes. A partir dessa interface, o mediador pode enviar comentários associados às questões de cada aprendiz.

No relatório de acompanhamento de avaliação, o mediador tem uma visão geral do andamento da turma em relação às avaliações cadastradas (listas de exercícios e provas). Este relatório objetiva visualizar um *ranking* dos aprendizes por turma; facilitar o planejamento do tempo necessário para os aprendizes concluírem os exercícios e identificar grupos de aprendizes mais (ou menos) adiantados para propor exercícios em grupos.

A partir dos dados armazenados no banco de dados do ambiente LabSQL, aplicaram-se as técnicas de Mineração de Dados, as quais serão abordadas na Seção 3 para realização da análise do perfil dos alunos com relação à utilização deste ambiente de ensino-aprendizagem.

### **3. Mineração de Dados**

A Mineração de Dados consiste de um conjunto de técnicas reunidas da Estatística e da Inteligência Artificial com o objetivo de descobrir conhecimento novo, útil, relevante e não-trivial que porventura esteja escondido em uma grande massa de dados [Goldschmidt e Passos, 2005]. Descobrir padrões e tendências escondidos em grandes massas de dados não é processo trivial. Em Mineração de Dados esse processo envolve o uso de diversas tarefas. Dentre as existentes, pode-se citar: Classificação, Clusterização (Agrupamento), Estimativas (Regressão), Associação, dentre outras.

No contexto do processo de aprendizagem na EaD, este trabalho propõe obter conhecimento relevante para o entendimento do perfil dos alunos na utilização do ambiente LabSQL. A tarefa de Mineração de Dados mais adequada e viável para responder perguntas de interesse neste segmento é a tarefa de classificação, pois esta tarefa pode ser considerada preditiva, permitindo, dessa forma, prever o desempenho obtido pelos alunos. Na Seção 3.1, são apresentados os principais conceitos e técnicas empregados na tarefa de classificação.

#### **3.1. Tarefa de Classificação**

Classificação consiste em examinar as características de um objeto ou situação e atribuir a ele uma classe pré-definida, ou seja, esta tarefa tem como objetivo a construção de modelos que permitam agrupamento de dados em classes [Goldschmidt e Passos, 2005]. Esta tarefa é considerada preditiva, pois uma vez que as classes são definidas, ela pode prever automaticamente a classe de um novo dado. Os modelos de classificação possuem atributos de dois tipos: os preditivos e os objetivos. Geralmente, um atributo objetivo corresponde a uma variável categórica que representa as classes previamente definidas. Os atributos preditivos são os utilizados pela técnica para inferir a que classe um novo objeto pertence.

Existem diferentes técnicas para a realização desta tarefa, como Redes Neurais, Algoritmos Genéticos, Métodos Bayesianos, Árvores de Decisão, entre outras. Dentre estas, utilizou-se Árvore de Decisão e Redes Bayesianas. Tais técnicas foram utilizadas, pois, a partir das Redes Bayesianas, é possível contabilizar as relações de dependência entre as ações envolvidas no processo de aprendizagem e o desempenho obtido pelos

---

---

alunos. As principais vantagens da utilização das Árvores de Decisão são que elas fazem decisões levando em consideração os atributos que são mais representativos, além de serem compreensíveis para a maioria das pessoas. Ao escolher e apresentar as regras em ordem de importância, as árvores de decisão permitem aos usuários observarem quais fatores mais influenciam os seus trabalhos. Nas Subseções 3.1.1 e 3.1.2, tais técnicas serão detalhadas.

### **3.1.1. Árvore de Decisão**

A Árvore de Decisão consiste de uma hierarquia de nós internos e externos que são conectados por ramos. Uma das principais características de uma Árvore de Decisão é o seu tipo de representação: uma estrutura hierárquica que traduz uma árvore invertida a qual se desenvolve da raiz para as folhas. A estrutura hierárquica traduz uma progressão da análise de dados no sentido de desempenhar uma tarefa de previsão/classificação.

Uma árvore de decisão é formada por um conjunto de regras de classificação. Cada caminho da raiz até uma folha representa uma destas regras. Cada percurso da árvore de decisão, desde um nó raiz até um nó folha, é convertido em uma regra, onde a classe do nó folha corresponde à classe prevista pelo consequente (parte “Então” da regra) e as condições ao longo do caminho correspondem às condições do antecedente (parte “Se” da regra).

### **3.1.2. Redes Bayesianas**

A noção fundamental da Estatística Bayesiana é a Probabilidade Condicional  $P(H|E)$  no qual  $H$  é a hipótese e  $E$  é a evidência. Para computar a probabilidade de uma hipótese  $H$ , é necessário levar em consideração o valor da evidência  $E$ . Quando não existir evidências, tem-se a probabilidade incondicional  $P(H)$  [Russell e Norvig, 2004].

A formulação do teorema de Bayes envolve estas probabilidades. A Equação (1) apresenta o teorema formulado por Thomas Bayes.

$$P(H | E) = \frac{P(E | H)P(H)}{P(E)}. \quad (1)$$

As Redes Bayesianas são grafos acíclicos direcionados, mostrando as relações de causalidade entre os atributos. Nestes grafos, as elipses representam os atributos e as ligações representam os relacionamentos de influência entre os atributos. A partir dos cálculos estatísticos, cada atributo terá uma tabela de valores de probabilidades para que suas possíveis ações sejam realizadas. Dessa forma, na utilização de uma ferramenta de análise de Redes Bayesianas é possível definir hipóteses sobre um determinado atributo, tendo respostas sobre as influências dele de acordo com as ligações existentes entre os outros atributos.

## **4. Estudo de Caso**

O objetivo deste estudo de caso é analisar a utilização do ambiente de ensino-aprendizagem LabSQL a partir da aplicação das técnicas de Mineração de Dados denominadas Árvore de Decisão e Redes Bayesianas. Este trabalho é do tipo exploratório, onde se analisa os dados buscando relacionamentos novos e não previstos.

---



---

Os dados utilizados neste trabalho foram obtidos a partir do banco de dados do LabSQL correspondentes a sete turmas em um modelo de ensino- aprendizagem semi-presencial, contendo em média 30 alunos, durante dois semestres letivos da Universidade Federal do Pará. Onde quatro turmas eram de pós-graduação em especialização em Banco de Dados e três turmas de graduação: sendo duas pertencentes ao curso de Ciência da computação e uma de Sistemas de Informação.

#### **4.1. Pré-processamento dos Dados**

Antes de realizar a etapa de extração de padrões, realizou-se um tratamento nos dados, adequando o formato dos dados selecionados para o processo de extração de conhecimento. A construção de uma árvore de decisão utilizando atributos contínuos exigiria a criação de um ramo para cada valor distinto do atributo, tornando a árvore pouco generalista. Para evitar este problema, é recomendado submeter os atributos contínuos a um processo de discretização [Fayyad, 1993]. Assim, foi realizada a discretização de alguns atributos para a redução do número de valores contínuos em classes.

Além disso, criaram-se novos atributos a partir de outros, como a média do nível de dificuldade dos problemas resolvidos e a média de pontos dos problemas resolvidos. Criou-se, também, alguns atributos dicotômicos visando avaliar o desempenho dos alunos, por exemplo, para avaliar se o aluno está abaixo ou acima da média de pontos ou de acessos ou se o aluno usou ou não determinado recurso do ambiente LabSQL, como a agenda e o trabalho em grupo.

Foram analisados 272 registros de usuários do ambiente LabSQL, sendo trabalhados com 18 atributos: sexo do usuário (masculino ou feminino); código do curso (Ciência da Computação, Sistema de Informação ou Especialização em Banco de Dados); código do tipo de curso (graduação ou especialização); código da disciplina (3 valores); código da turma (7 valores); código do coordenador da turma (2 valores); o tempo que o usuário levou para se inscrever na turma após o início de inscrição (em dias); trabalhou em equipe (sim ou não); usou agenda de anotações do sistema (sim ou não); total de pontos obtidos pelo usuário na resolução dos exercícios (valor contínuo); total de problemas resolvidos (valor contínuo); média de pontos dos problemas resolvidos (valor contínuo); média do nível de dificuldades dos problemas resolvidos (valor contínuo); quantidade de acessos as páginas do ambiente LabSQL (valor contínuo); ficou acima da média de acessos de todas as turmas (sim ou não); ficou acima da média de acessos de sua turma (sim ou não); ficou acima da média de pontos de todas as turmas (sim ou não) e ficou acima da média de pontos da sua turma (sim ou não).

#### **4.2. Resultados - Redes Bayesianas**

Para a aplicação da técnica de Redes Bayesianas, foi utilizado o *software Bayesware Discoverer* [Bayesware, 2008] que proporciona a criação e a geração de redes bayesianas. Este *software* construiu as redes a partir dos atributos do banco de dados criado, exibindo as tabelas de probabilidade condicional ou incondicional de cada nó (atributo).

Dentre as ligações observadas na Figura 1, gerada após a execução do *software*, destaca-se que a demora para inscrição influencia diretamente na média da quantidade

---

de acessos, na utilização ou não da agenda e no total de pontos obtidos pelos usuários. Por outro lado, a demora para inscrição é influenciada pelo curso. Além disso, observou-se que a quantidade de acessos influenciou no total de problemas respondidos e que o sexo do usuário não influencia nenhum outro atributo.

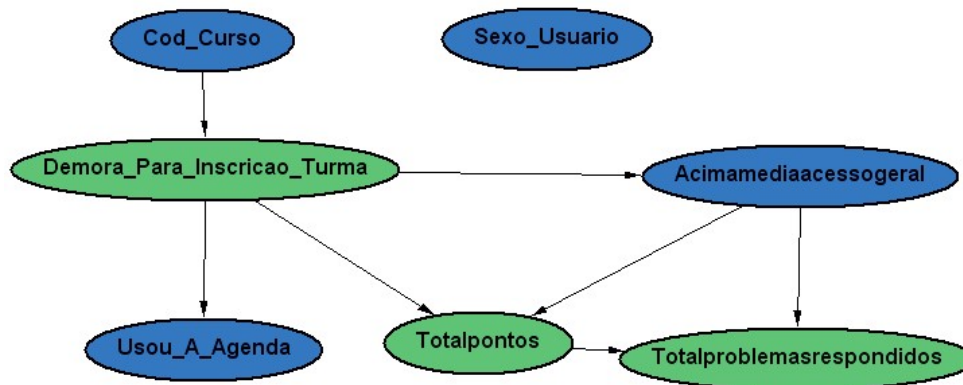


Figura 1. Rede Bayesiana para Análise da Demora para Inscrição.

Na Figura 2 é possível observar as tabelas de probabilidade dos nós da rede gerada. Em cada tabela, tem-se a distribuição probabilística dos seus possíveis valores. Após realizar algumas inferências, foi possível observar que ao colocar o atributo *demora\_para\_inscricao\_turma* em 100% para a menor demora (0 a 6 dias), inferiu ao atributo *acimaMediaAcessoGeral* um aumento de 0,388 (38,8% de probabilidades a priori) para 0,509 (50,9% de probabilidades a posteriori) em “S” (acima da media de acesso). Da mesma forma, inferiu ao atributo *AcimaMediaPontosTurma* um aumento de 0,398 (39,8% de probabilidades a priori) para 0,500 (50% de probabilidades a posteriori) em “S” (acima da media de pontos da turma).



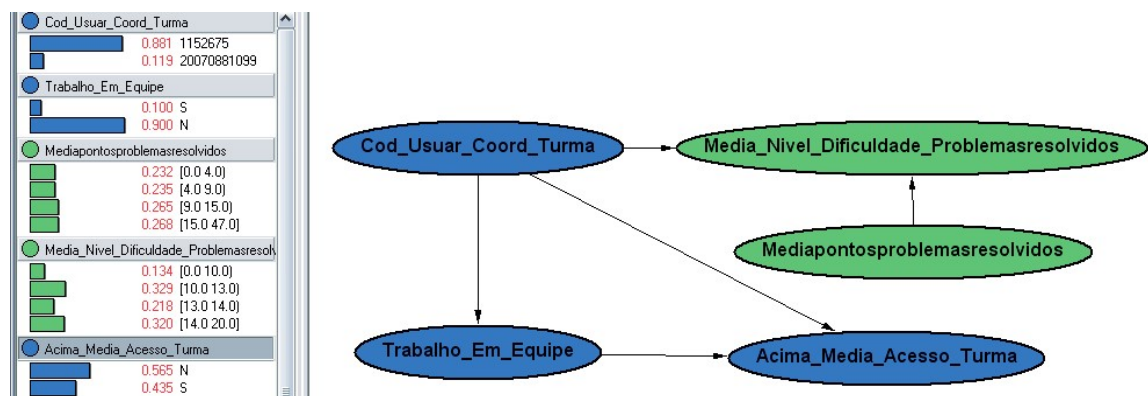
Figura 2. Tabelas de Probabilidade para Análise da Demora para Inscrição.

Portanto, os alunos que iniciam a disciplina mais cedo no ambiente, possuem um desempenho elevado em relação à quantidade de acessos e ao total de pontos na resolução dos exercícios. Desde o início, deve-se motivar a participação dos alunos na utilização do ambiente para desenvolver melhor seu aprendizado, e criar outras

possibilidades para aproximar os alunos que estão atrasados em relação ao restante da turma.

Pode-se observar ainda, que o atributo *Usou\_A\_Agenda* aumentou de 0,046 (4,6% de probabilidade a priori) para 0,480 (48% de probabilidade a posteriori) em “S” (usou a agenda). Portanto, os estudantes que iniciam mais cedo no ambiente exploraram mais as funcionalidades do sistema. Dessa forma, para auxiliar os estudantes atrasados, faz-se necessário criar uma espécie apoio na descoberta das funcionalidades do sistema.

Além disso, a Figura 3 apresenta uma rede bayesiana para análise do coordenador da turma. Nela, destaca-se a influência que o coordenador da turma tem em relação à quantidade de acesso dos usuários, o trabalho em equipe e o nível de dificuldade dos problemas resolvidos pelos usuários. Dessa forma, observa-se que determinados coordenadores de turma promovem uma maior utilização do ambiente em relação à quantidade de acessos dos alunos do que outros coordenadores. Além disso, a partir da rede bayesiana gerada é possível perceber diferenças de atitudes dos coordenadores em relação à utilização dos recursos do ambiente durante o processo de aprendizado, como o nível de dificuldade dos trabalhos oferecidos aos alunos e a opção de organizar trabalho em equipe.



**Figura 3. Tabelas de Probabilidade e Rede Bayesiana para Análise do Coordenador da Turma.**

A Subseção 4.3 apresenta os resultados da aplicação da técnica de Árvore de Decisão sobre a mesma base de dados utilizada na aplicação da Rede Bayesiana.

#### 4.3. Resultados: Árvore de Decisão

No processo inicial de análise dos dados, utilizou-se a técnica de Árvore de Decisão com o objetivo de formar uma árvore de decisão para a indução de regras. Para isso, utilizou-se a o sistema WEKA (*Waikato Environment for Knowledge Analysis*) [WEKA, 2008], versão 3.4, que executou a tarefa de classificação, utilizando o algoritmo J48 em validação cruzada (*cross validation*) para a construção da árvore de decisão. O método de validação cruzada *10-Fold-Cross-Validation* foi utilizado a fim testar a acurácia do modelo no mesmo conjunto de dados utilizado para construir o modelo. A acurácia é uma métrica que avalia os modelos de classificação a partir da porcentagem de predições corretas que o modelo executou sobre o total de predições realizadas. Ela é importante, pois permite avaliar um classificador para determinar o quanto ele será eficiente para prever dados futuros, ou seja, qual a sua capacidade de generalização. A média de acurácia dos modelos de classificação obtidos foi de 83,13%.

---

Foram geradas duas árvores de decisão, sendo uma para avaliar o aluno que está abaixo ou acima da média de pontos da sua turma e a outra para avaliar o aluno que está abaixo ou acima da média de acessos de sua turma. A partir das árvores de decisões geradas foram extraídas cerca de 50 regras, dentre as quais, foram selecionadas as mais relevantes. Um dos critérios que podem ser utilizados para medir a qualidade das regras geradas por um sistema de aprendizado é a precisão. A precisão é o grau de confiabilidade das regras, geralmente representada a partir da taxa de erro.

Para Berson e Smith (1997), o conceito de taxa de estimativa de erro pode ser obtido a partir da seguinte forma: se  $N$  exemplos são cobertos por determinado nó folha e  $E$  dentre estes  $N$  são classificados de forma incorreta, então a taxa de estimativa de erro dessa folha é  $E/N$ . O valor da classe (atributo-meta), que corresponde à parte “então” da regra, é apresentado após o símbolo de dois-pontos (:).

Por exemplo, para a regra extraída da árvore de decisão: “*cod\_Curso = ‘sistemas\_de\_informacao’ AND Media\_Nivel\_Dificuldade\_Problemasresolvidos > 1,2 : ‘acima’ (20/1)*”, tem-se dentro dos parênteses, respectivamente, que o peso das instâncias classificadas no nó folha *AcimaMediaPontosdaTurma* com valor “acima”, ou seja, acima da média de pontos da sua turma, é 20 e que o número de instâncias classificadas de forma incorreta para esse nó é 1. A partir da fórmula da estimativa de erro, a taxa de erro será 0,05 (1/20). Dessa forma, a precisão para esta regra será 0,95 (1 - 0,05), ou seja, aproximadamente 95% dos usuários do curso de Sistema de Informação que resolvem problemas com nível de dificuldade em média de 1,2 (numa escala de 1 a 3) estão acima da média de pontos da sua turma.

Além disso, observa-se que 100% dos usuários que demoram menos de 30 dias para se inscrever estão acima da média de pontos da sua turma (*demora\_para\_inscricao\_turma <= 30: ‘acima’ (20.0)*). Portanto, os alunos mais interessados na disciplina que buscam mais cedo iniciar a utilização da ferramenta, demonstram um desempenho maior em relação à pontuação dos exercícios. Outro exemplo de regra encontrado foi que 100% dos usuários que estão abaixo da média de pontos da sua turma e respondem menos de 145 problemas, estão abaixo da média de acesso da sua turma (*Acima\_media\_pontos\_turma = ‘abaixo’ AND total\_problemas\_respondidos <= 145: ‘abaixo’ (91.0)*). Portanto deve-se incentivar cada vez mais o acesso dos alunos no sistema.

Dessa forma, a partir da análise das regras de classificação encontradas é possível perceber padrões referentes ao processo de aprendizado relacionado ao comportamento dos alunos que podem ser aproveitados pelos gestores do domínio da aplicação.

## 5. Trabalhos Relacionados

Alguns trabalhos foram realizados, mostrando como técnicas de Mineração de Dados podem auxiliar na análise de um sistema de ensino a distância na Internet. Apresentando, dessa forma, novos mecanismos para analisar os atributos envolvidos na análise do desempenho dos aprendizes em um curso de ensino a distância.

Machado e Becker (2002) propõem um estudo de caso, para acompanhar uma aplicação voltada ao processo de ensino-aprendizagem a distância, onde se busca descobrir e analisar, através da mineração *Web*, as interações destes usuários com o

---



---

ambiente de ensino baseado na *Web*. Mais especificamente, buscou-se estabelecer um modelo de Mineração do Uso da *Web* pertinente para a descoberta de conhecimento aplicado a ambientes de ensino a distância a partir da análise de navegação dos usuários enquanto interagem neste ambiente, a fim de prover recursos de comparação entre os projetos atuais de um site educativo e seu uso real.

Lopes e Schiel (2004) propõem uma estratégia para o acompanhamento do aprendizado na educação a distância baseada nas práticas de acompanhamento do ensino presencial, acrescida da técnica de análise de dados. Estes fatores permitem verificar a aprendizagem de forma mais elaborada. Estes dados foram gerados com ferramentas de Mineração de Dados.

Wang e Meinel (2007) descobriram mudanças interessantes do aprendizado dos estudantes que usam os sistemas de aprendizado a distância via Internet. Seu trabalho foi implementado em um ambiente de ensino a distância denominado tele-TASK. Os resultados da mineração foram úteis para ajudar os professores a conhecer seus estudantes e ajustar o programa de ensino de maneira mais eficaz.

## 6. Conclusões

A avaliação e o entendimento do processo de ensino-aprendizado é um tópico de pesquisa importante devido ao crescente número de sistemas de EaD disponíveis como o LabSQL. Neste estudo, as técnicas de Mineração de Dados mostraram-se eficientes para analisar o comportamento dos estudantes e professores em um curso oferecido a partir de um ambiente de EaD, como observado nos resultados obtidos.

As Redes Bayesianas permitiram-se contabilizar relações de dependência entre as ações envolvidas no processo de aprendizagem e o desempenho obtido pelos alunos. E as árvores de decisão permitiram-se perceber padrões referentes ao processo de aprendizado relacionado ao comportamento dos alunos, levando em consideração as regras que são mais relevantes, além de serem compreensíveis para a maioria das pessoas.

Assim, a partir da combinação das tecnologias de Mineração de Dados com ambientes de EaD, é possível analisar e obter conclusões sobre as ações praticadas pelos usuários em relação ao seu desempenho e aprendizado, beneficiando os envolvidos com a utilização e evolução de ambientes de EaD.

Dessa forma, dentre outros resultados, nota-se que os usuários que se inscrevem mais tarde no ambiente demonstram deficiências na utilização dos recursos do ambiente LabSQL, uma vez que, proporcionalmente, obtiveram uma menor pontuação na resolução das atividades do ambiente LabSQL e conseqüentemente um desempenho inferior aos demais.

Como trabalhos futuros, pretende-se integrar as técnicas de Mineração de Dados avaliadas neste trabalho no ambiente LabSQL, a fim de obter os resultados da aplicação dessas técnicas de forma automática a partir do ambiente. Além disso, pretende-se construir um *data warehouse*, um repositório de dados, com os dados do ambiente LabSQL, a fim de ser utilizado para armazenar informações relativas aos atributos presentes no banco de dados que expressem o comportamento dos usuários, favorecendo a aplicação das técnicas de descoberta de conhecimento em base de dados.

---

---

## Referências

- Bayesware Limited. (2008), “Bayesware Discoverer”, Disponível em <http://www.bayesware.com>. Acessado em janeiro de 2008.
- Berson, A.; Smith, S.J. (1997), Data Warehousing, Data Mining and OLAP. USA. Mac-Graw-Hill.
- Fayyad, U. M. (1993), Multi-interval discretization of continuous valued attributes for classification learning. In: 13 th Int. Joint Conf., Alemanha.
- Goldschmidt, R. R. ; Passos, E. P. L. (2005), Data Mining: Um Guia Prático - Conceitos, Técnicas, Ferramentas, Orientações e Aplicações. 1. ed. Rio de Janeiro: Editora Campus. v. 1. 250 p.
- Lino, A. D. P.; Silva, A. do S.; Favero, E. L.; Brito, S. R.; Harb, M. P. A. A. (2007), Avaliação automática de consultas SQL em ambiente virtual de ensino-aprendizagem. 2ª Conferência Ibérica de Sistemas e Tecnologias de Informação, Porto.
- Lopes, C. C; Schiel, U. (2004), Uma Estratégia para Aplicar Mineração de Dados no Acompanhamento do Aprendizado na EaD. XIII Seminário de Computação. Blumenau/SC.
- Machado, L. S., Becker, K. (2002), O Uso da Mineração de Dados na Web Aplicado a um Ambiente de Ensino a Distância. In: I Workshop de Teses e Dissertações em Banco de Dados. XIX Simpósio Brasileiro de Banco de Dados. Gramado.
- Martins, J. G.; Campestrini, B. B. (2004), Ambiente virtual de aprendizagem favorecendo o processo ensino-aprendizagem em disciplinas na modalidade de educação a distância no ensino superior. Universidade do Vale do Itajaí, 2004. Disponível em: <<http://www.abed.org.br/congresso2004/por/htm/072-TC-C2.htm>>. Acessado em: março 2008.
- Russell, S. J., Norvig, P. (2004), “Inteligência Artificial”, 2ª Edição, Editora Elsevier, Rio de Janeiro - RJ.
- Souza, E. P. de. (2007), Avaliação Formativa em Educação a Distância via Web. 13º Congresso Internacional de Educação à Distância. Curitiba –PR. 2007.
- Wang, L; Meinel, C. (2007), Detecting the Changes of Web Students' Learning Interest. Proc. 6th IEEE/WIC/ACM WI, IEEE Press, pp. 816 - 819, Silicon Valley, USA.
- WEKA. (2008), Data Mining Software in Java. Disponível em <http://www.cs.waikato.ac.nz/ml/weka>. Acessado em março de 2008.
- Zaiane, O.; Luo, J. (2001), Towards Evaluating Learners' Behaviour in a Web-Based Distance Learning Environment. In: Proceedings of ICALT.
-