

## ***DATA MINING* APLICADO AO ENSINO A DISTÂNCIA**

Jeovane Aparecido de Oliveira<sup>1</sup>

Ary dos Santos Rocha Junior<sup>2</sup>

**RESUMO:** Atualmente, novos desafios aparecem no contexto de ensino a distância, exigindo cada vez mais dos responsáveis por cursos de tal cunho. Em contrapartida, diversas ferramentas vêm auxiliar no processo de planejamento, estruturação e definição das diretrizes de cursos a distância. O objetivo deste artigo é apresentar alguns dos benefícios que uma destas ferramentas, *Data Mining*, pode proporcionar aos gestores e aos promotores de cursos a distância. Obviamente que este não pretende esgotar todas as possibilidades aqui. Para uma melhor compreensão das técnicas usadas aqui, fez-se necessária a criação de uma base de dados de simulação, por meio da qual dados aleatórios foram gerados e posteriormente armazenados. Será apresentada, neste artigo, a metodologia utilizada para atingir tais benefícios, além de uma breve descrição das tecnologias apresentadas neste. Será apresentada ainda uma breve descrição da evolução do EaD (Ensino a Distância) e também análise e discussão dos resultados obtidos com a realização deste trabalho.

**PALAVRAS CHAVE:** *Data Mining*. Mineração de Conhecimento. Ensino a Distância.

**ABSTRACT:** In present days new challenges appear in the distance education context, demanding each time more of the responsible by these courses. On the other hand diverse tools come to assist in the process of planning and definition of the lines of direction of courses in the distance. The objective of this article is to present some of the benefits that one of these tools, *Data Mining*, can to provide to the managers and promoters of courses. Obviously that this article does not intend to deplete here all the possibilities. For one better understanding of the used techniques here, became necessary the creation of a simulation database , created from random data, generated and later stored. It will be presented in this article, the used methodology to reach such benefits, beyond one brief description of the technologies presented in this. One brief description of the evolution of the EaD will be presented still (Education in the Distance) and also analysis and quarrel of the results gotten

---

1 Aluno do curso de Sistemas de Informação da Faculdade de Ciências Administrativas (FACIA) do Centro Universitário de Patos de Minas (UNIPAM) e bolsista do VI PIBIC.

2 Professor do Centro universitário de Patos de Minas (UNIPAM) e orientador da pesquisa.

with the accomplishment of this work.

**KEYWORDS** : Data Mining, Knowledge Mining, Distance Education

## 1 INTRODUÇÃO

“As atividades que ocupam o lugar central das organizações não são mais aquelas que visam produzir ou distribuir objetos, mas aquelas que produzem e distribuem informação e conhecimento”. [1]

A primeira definição dos fatores de produção como terra, capital e trabalho foi feita por Jean Baptist Say (1767-1832). Hoje, mais de duzentos anos depois, vive-se em uma sociedade onde esses fatores não são mais importantes do que o “Conhecimento”, de acordo com Peter Drucker. Diante disso, pessoas e organizações estão cada vez mais em busca dessa nova necessidade. Com essa crescente demanda, o EaD (Ensino a Distância) assume uma função fundamental nessa nova sociedade.

EaD supre uma necessidade da sociedade atual: aprender a cada dia o mais rápido possível. Várias empresas já o adotam como meio de treinamento e implantação de novos produtos ou de novos processos, e essas empresas têm tido um retorno muito bom, uma vez que essa tecnologia tem alta flexibilidade e economiza custos com hospedagem, viagens, tempo de seus funcionários, entre outros.

O EaD é usado há muito tempo, desde a época dos cursos por correspondências e uma preocupação que sempre esteve presente foi a sua eficiência, pois fazer com que as pessoas acreditem no EaD sempre foi um desafio aos seus promotores.

Existem hoje várias ferramentas para melhorar a eficiência destes cursos, entre as quais se destaca *Data Mining*. Esta possibilita a extração de conhecimento sobre os dados operacionais, conhecimento que auxiliará na administração e no controle de qualidade destes cursos. Entre as funcionalidades que *Data Mining* propicia podem-se citar: revelar padrões de relação entre alunos e materiais, padrões de relação entre alunos e professores, interação entre aluno e turma, entre outros. Esses padrões possibilitam uma análise mais completa e precisa do ambiente educacional; gera-se, então, uma base de conhecimento que possibilitará o planejamento de diretrizes mais precisas em relação a diversos fatores como, por exemplo, qual curso oferecer a determinada faixa etária, ou que maneira dispor o conteúdo dos cursos a determinado perfil de estudante.

O objetivo deste trabalho é usar *Data Mining* no contexto de EaD, revelando, assim, padrões nas atividades executadas pelos alunos como, por exemplo, baixa de

materiais, acesso a determinado conteúdo, frequência de acesso ao ambiente do curso, relação entre desistências, perfis de aluno e cursos, entre outros, que ajudarão no controle e no planejamento de Cursos a Distância. Com isso, podem-se criar ferramentas para controle e para análises mais precisas e, conseqüentemente, aumentar a qualidade dos cursos.

## **2 DESCRIÇÃO DA METODOLOGIA**

O desenvolvimento deste trabalho resume-se à criação da base de dados de simulação, geração aleatória dos dados, aplicação das técnicas de *Data Mining* nesses dados, apresentação e discussão dos resultados obtidos. No entanto, é importante também destacar algumas das características das técnicas utilizadas neste trabalho, assim como uma breve descrição da evolução do EaD. A seguir um breve histórico da evolução do EaD.

### **2.1 EVOLUÇÃO DO EAD**

A primeira geração de ensino a distância foi a dos cursos por correspondências ou televisionados, em que a interação do aluno era pequena. Na segunda, essas tecnologias ficaram mais interativas, deixando, assim, o ambiente mais parecido com o de uma sala de aula. Na terceira, com correio eletrônico, conferência computadorizada, o ambiente ficou mais interativo. E acredita-se existir hoje uma quarta geração, na qual os meios de comunicação se tornaram muito eficientes, rápidos e confiáveis, fazendo com que as preocupações mudem e estejam além da comunicação e da interatividade do ambiente de aprendizado [3].

Uma das causas da evolução do EaD é o avanço da tecnologia que o rodeia e o faz acontecer, como, por exemplo, o surgimento do *E-Learning* e, principalmente, a evolução dos computadores. Há quinze ou vinte anos, as perspectivas para um micro não incluíam usá-lo para realização de um curso ou treinamento. Hoje, contudo, os PC's estão em todos os lugares e cada vez mais computadores são usados para esse fim, facilitando a Aprendizagem a Distância. Outro fator marcante é a evolução dos meios de comunicação. Graças à Internet, temos hoje à disposição diversas funcionalidades, tais como, *email*, mensagens instantâneas, *chats* e muito mais.

### **2.2 DATA MINING**

*Data Mining* refere-se à extração ou à mineração de conhecimento em grandes

bases de dados. Existem outras denominações, quais sejam: Mineração de Conhecimento em Bases de Dados (*Knowledge Mining from Data*), ou somente Mineração de Conhecimento (*Knowledge Mining*). Muitas pessoas tendem a confundir *Data Mining* com KDD (*Knowledge Discovery in Databases*), que consiste em um processo de fases consecutivas das quais *Data Mining* faz parte, ou seja, *Data Mining* é uma fase do KDD. Outras definições, como Extração de Conhecimento (*Knowledge extraction*), Análise de Padrões em Dados (*Data Pattern Analysis*), Arqueologia de Dados (*Data Archaeology*), Dragagem de Dados (*Data Dredging*), também podem ser encontradas.

A primeira fase do KDD é a preparação dos dados. Essa fase consiste basicamente na limpeza dos dados, ou seja, na eliminação de dados desnecessários ao processo de *Data Mining*. O próximo passo é a integração dos dados, já que múltiplas fontes de dados podem ser combinadas. A seleção dos Dados é a próxima tarefa. Nessa etapa, os dados relevantes à análise são extraídos do banco. Na próxima fase, esses dados são transformados dentro dos parâmetros da mineração, geralmente por funções de agrupamento. A próxima fase, *Data Mining*, é quando serão aplicados métodos e técnicas inteligentes para extração de padrões nos dados. Esses padrões serão analisados na próxima fase, A Avaliação dos Padrões, para, assim, chegar-se ao objetivo: o Conhecimento [7].

Os principais componentes de um Ambiente de *Data Mining* são os seguintes :

- Database, Data Warehouse ou qualquer outra forma de armazenamento de dados: pode ser um ou mais repositórios, ou seja, podem ser combinadas várias fontes de dados. A limpeza e a integração dos dados acontece aqui;
- Servidor de Banco de dados ou *Data Warehouse*: responsável por retornar as informações desejadas;
- Base de Conhecimento do usuário: é usada para direcionar as buscas e os padrões buscados. Usada como base para definir o que se espera de todo o processo;
- Máquina de *Data Mining*: com ferramentas como Redes Neurais, Algoritmos Genéticos, realizam-se as tarefas de agrupamento, caracterização, associação, análise dos *clusters* e desvios;
- Módulo de avaliação de padrões: nessa fase, os padrões encontrados pela fase de *Data Mining* são analisados; opcionalmente pode vir junto à fase de *Data Mining* ;
- Interface gráfica com o usuário final: é usada para realizar a comunicação entre usuários e o sistema de *Data Mining*, permitindo que o usuário interaja com a busca [7];

A figura abaixo representa um ambiente de *Data Mining*.

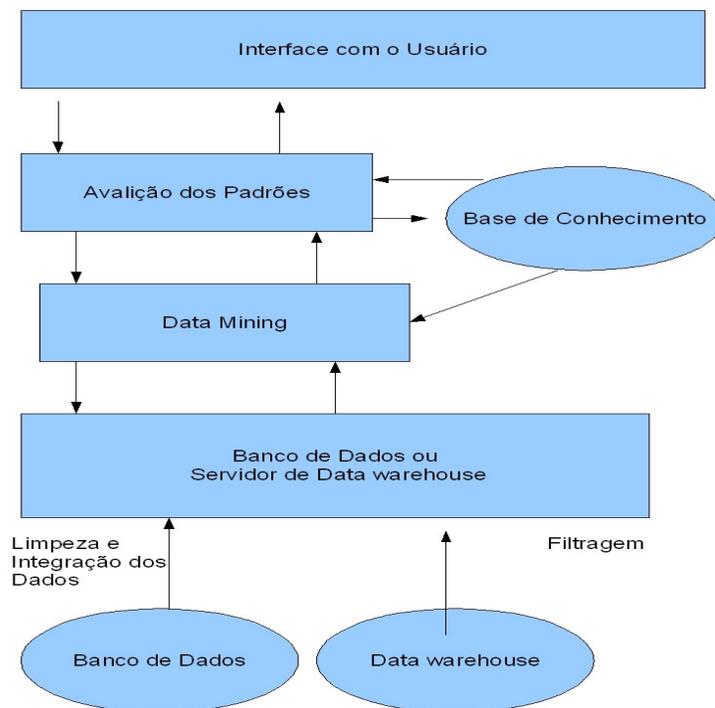


Figura 1. Arquitetura típica de um Sistema de *Data Mining*.

O conhecimento adquirido neste processo, especificamente no ambiente de EaD, possibilita um controle maior da qualidade dos cursos, um direcionamento das atividades oferecidas e ou a maneira de como são oferecidas, propiciando, assim, um maior rendimento. Com essa técnica, é possível descobrir possíveis associações entre perfis de alunos e cursos oferecidos. Um exemplo: alunos com faixa etária entre 20 e 30 anos estão mais propensos a cursos de Tecnologias do que a cursos voltados para a área Humanas. Assim, pode focar-se em alunos "idéias" para os cursos oferecidos. As informações descobertas aqui, ou o conhecimento minerado aqui, não seriam obtidos com simples consultas.

É importante ressaltar que, neste trabalho, fala-se de *Data Mining* aplicado a Banco de Dados Relacionais, não abordando Mineração de Dados Complexos, Espaciais, Multimídia, Textuais, ou qualquer outro tipo de Armazenagem de Dados.

### 2.3 CRIAÇÃO DA BASE DE DADOS PARA SIMULAÇÃO E GERAÇÃO ALEATÓRIA DOS DADOS

Nesta etapa do trabalho, foi criada a base de dados para a realização dos

processos necessários ao desenvolvimento deste trabalho.

É importante atentar-se para o fato de que os dados apresentados aqui são oriundos de simulações, podendo, portanto, apresentar um cenário diferente do cenário realístico. O objetivo principal deste trabalho não é apresentar este cenário realístico, mas, sim, apresentar as possibilidades que se terão ao implantar os mecanismos proporcionados pelas técnicas de *Data Mining*.

Após a criação deste ambiente de simulação, foram realizados os passos necessários ao processo de *Data Mining*. Para maiores detalhes sobre estes passos, consultar a seção anterior (cf. subseção 2.3). Nesta etapa, várias passagens sobre o banco de dados se fizeram necessárias, assim como diversos testes validando, assim, os resultados alcançados. Após esta etapa, têm-se os resultados. Serão apresentados a seguir alguns exemplos desses resultados.

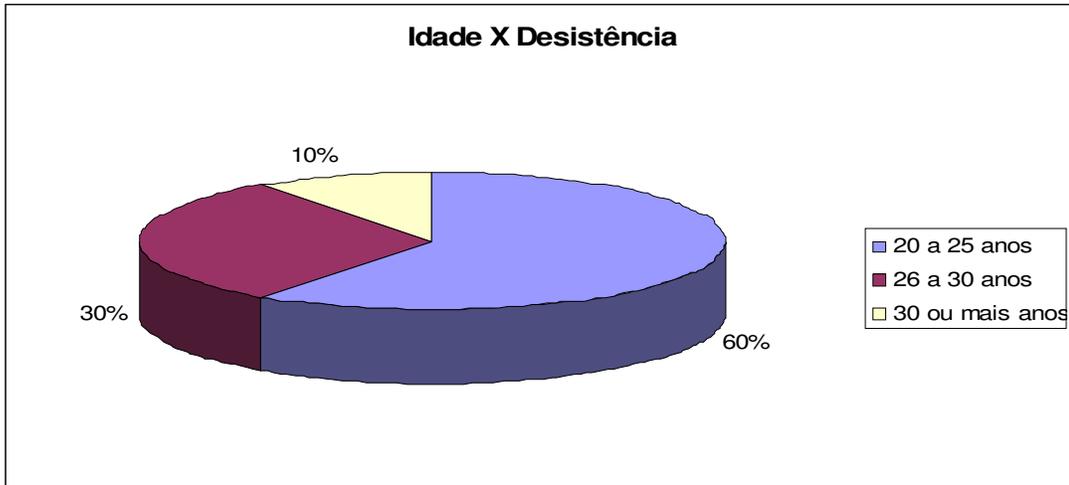
### **3 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS**

Em um processo de mineração de dados, com o objetivo de minerar regras de associação, têm-se, inicialmente, muitas dessas regras que, na maioria das vezes, não são interessantes ao processo, ou seja, não dizem nada de novo, ou não são significativas. A presença de uma pessoa que conhece as regras de negócio onde o processo de Data Mining está sendo aplicado se faz muito importante, uma vez que será esta quem validará essas regras de associação. O processo de *Data Mining* é um processo iterativo que necessita, algumas vezes, da interação do usuário. Mais uma vez, um técnico na área de conhecimento na qual estão sendo aplicadas as técnicas de Data Mining se faz necessário.

Após todo esse trabalho de geração das regras de associação e análise das mesmas, surgem algumas regras interessantes que podem contribuir para o processo de tomada de decisão. A seguir alguns exemplos destas:

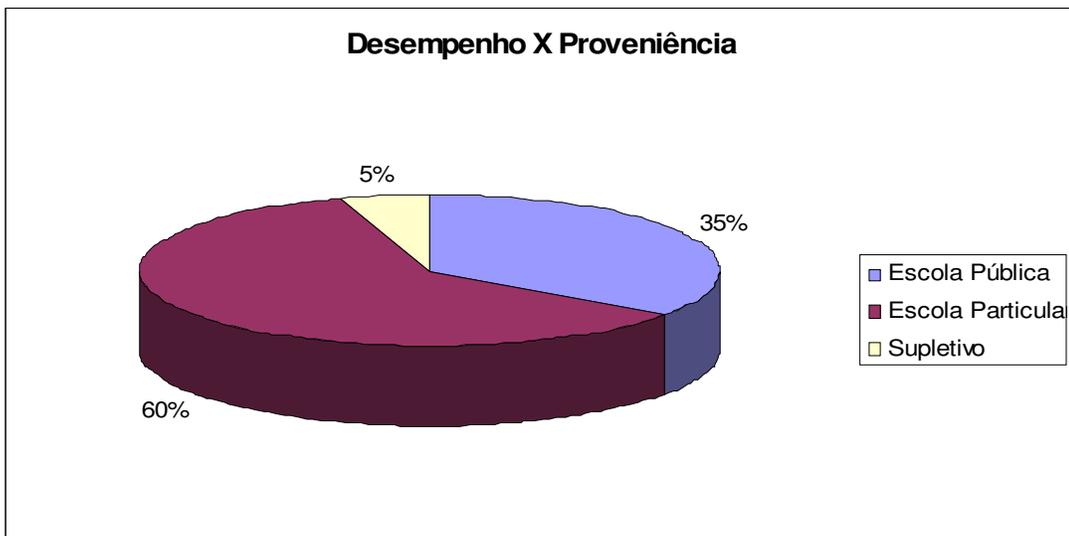
Em 70% das desistências, o aluno teve uma média de faltas maior do que a média da turma nas três primeiras semanas de aula.

Cruzando essa informação com a idade dos alunos, obteve-se o resultado de que 60% das pessoas com desistência possuem 30 ou mais anos de idade conforme gráfico a seguir:

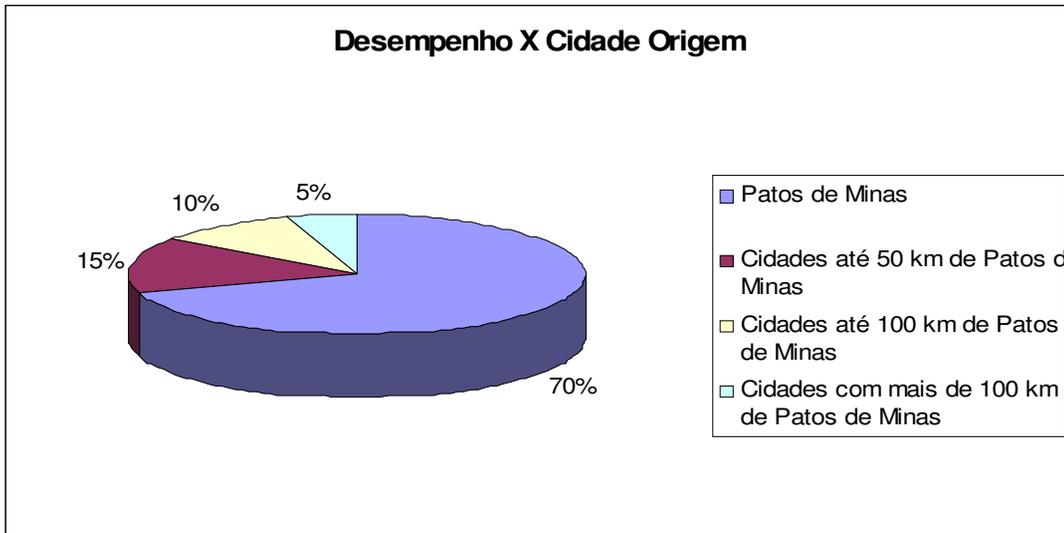


Em 60% das vezes que um aluno teve desempenho superior ou igual a 80% nas disciplinas, o aluno teve também uma participação maior do que a média nos fóruns do ambiente do curso.

Cruzando esta informação com a proveniência dos alunos, visualiza-se que 60% da desistência é proveniente de alunos de escola pública, conforme se verifica no gráfico abaixo:

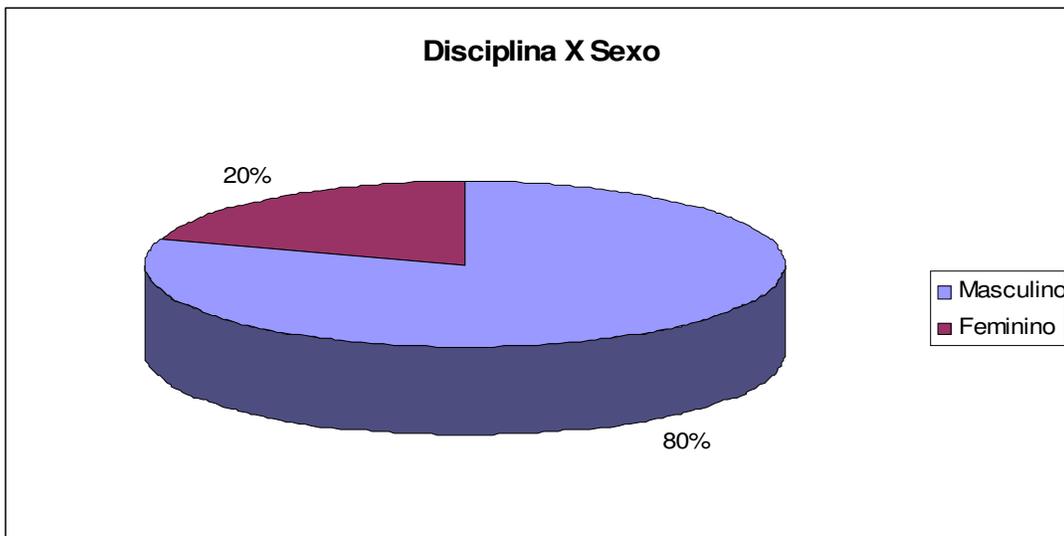


Cruzando esta informação com a cidade de origem do aluno, obteve-se que 70% das desistências são de alunos provenientes de Patos de Minas, de acordo com o gráfico a seguir:



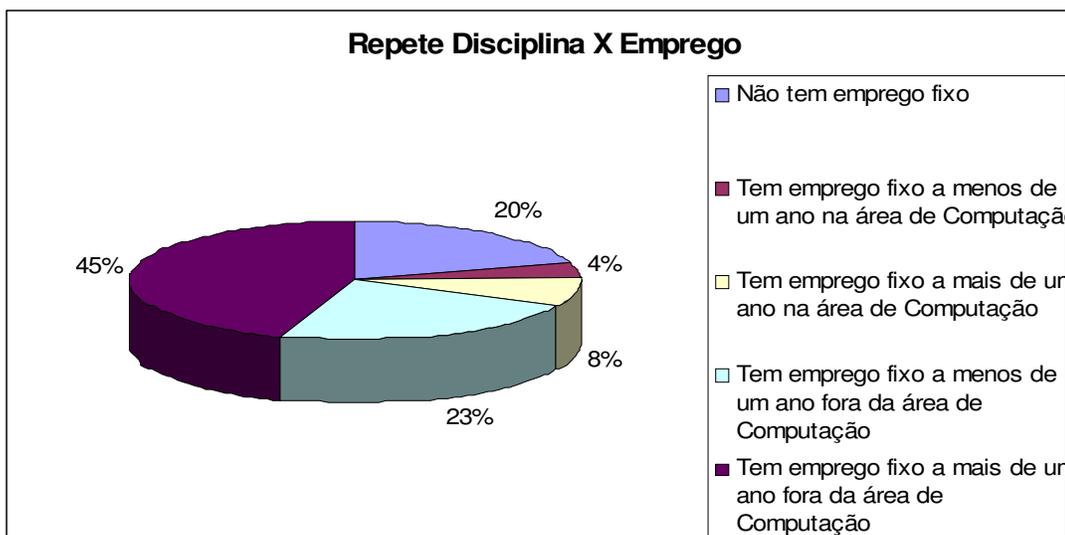
Em 65% das vezes que um aluno cursa a Disciplina Opcional "Comandos e técnicas Linux" também cursam a disciplina "Comandos e técnicas do SGDB PostgreSQL"

Cruzando essa informação com o sexo do aluno, obteve-se que 80% dos alunos são do sexo masculino, conforme gráfico a seguir:



Em 60% das vezes que o aluno repete a disciplina "Lógica de Programação" repete também a disciplina "Algoritmos e Estrutura de Dados em Java"

Cruzando essa informação com o emprego do aluno, obteve-se que 45% dos alunos que repetem disciplinas têm emprego fixo há mais de um ano fora da área de computação, conforme se verifica no gráfico a seguir:



### 3 CONCLUSÃO

Atualmente, percebe-se que a sociedade vivencia grandes e importantes modificações, principalmente no que tange à comunicação e ao conhecimento. É nesse cenário que a nova geração de EaD, a geração do Ensino Eletrônico, conquista cada vez mais espaço.

Hoje EaD possui como apoio tecnologias altamente eficientes, das quais destacam-se as seguintes: *Data Mining*, Redes Neurais, Algoritmos Genéticos, Internet. Cada uma delas contribui diferentemente para o seu desenvolvimento. Com isso, EaD se fortalece e conquista cada vez mais espaço na sociedade, conforme pudemos verificar ao longo deste artigo.

Essas novas tecnologias, além de possibilitarem uma interatividade maior, possibilitam um “raio X” completo do ambiente educacional, possibilitando, assim, controles antes desconhecidos, ou inalcançáveis. Permitem respostas rápidas e eficientes a problemas complexos, muitas das vezes se baseando em processos naturais, como, por exemplo, Algoritmos Genéticos e Redes Neurais Artificiais. Destacou-se, neste trabalho, *Data Mining*, apresentando de forma breve sua metodologia e principalmente os benefícios que este pode proporcionar ao ambiente de ensino a distância.

De uma forma mais geral, percebe-se que, atualmente, caminha-se a passos mais largos do que antes. Percebe-se que os clientes, nesse caso estudantes, estão mais exigentes, e que tecnologias se tornam obsoletas mais rapidamente. Assim, o conhecimento

transformou-se em um diferencial indispensável entre as pessoas e, diante desse cenário, percebe-se o quão impactante e importante essas tecnologias são.

## REFERÊNCIAS

HAN, Jiawei; KAMBER, Micheline. **Data Mining Concepts and Techniques**. 1. ed. San Diego: Academic Press, 2001.

OGLIARE, Paulo José. Curso de Data Mining. **Departamento de Informática e Estatística. - UFSC (Universidade Federal de Santa Catarina)**. Disponível em: <<http://www.inf.ufsc.br/~ogliari/cursodedatamining.html>> Acesso em: 15 mar. 2006.

[1] DRUCKER, P. F. **Sociedade pós-capitalista**. 7.ed. Rio de Janeiro: Campus, 1999. 208 p.

[2] JÁCOME, Taisa Fernandes. **Fisicare: Um E-Learning para a área da Física**. 2004 72 f. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação)-Centro Universitário de Palmas, 2004

[3] SALES, Mary Valda Souza. **UMA REFLEXÃO SOBRE A PRODUÇÃO DO MATERIAL DIDÁTICO PARA EAD**. Maio de 2005. 7 f. Descrição de Projeto em Andamento. Universidade do Estado da Bahia – UNEB.

[4] BECKER, Fernando. **Educação e construção do conhecimento**. Porto Alegre: Artmed Editora, 2001.

[5] PIAGET, Jean. **Biologia e conhecimento**. Porto: Rés Editora, 1978

[6] SLEVIN, James. **Internet e Sociedade**. 1. ed. Lisboa, Agosto 2002.

[7] HAN, Jiawei e KAMBER, Micheline. **Data Mining Concepts and Techniques**. San Diego, Academic Press, 2001.

[8] **Algoritmos Genéticos. Introdução**. Disponível em <<http://www.din.uem.br>> Acesso em: 26 de fevereiro de 2006.

[9] ROCHA JUNIOR, Ary dos Santos. **MINERAÇÃO DE PADRÕES SEQUENCIAIS GENERALIZADOS UTILIZANDO PROGRAMAÇÃO GENÉTICA**. Março de 2003. 111 f. Dissertação de Mestrado. Universidade Federal de Uberlândia. UFU.

[10] **Uma Introdução a Redes Neurais**. Disponível em <<http://www.din.uem.br>> Acesso em: 26 de fevereiro de 2006.