



MINERAÇÃO DE DADOS APLICADA NA DETECÇÃO DE METODOLOGIAS DE ESTUDO EM ESTUDANTES UNIVERSITÁRIOS

SCHNEIDER, Felipe Pires¹; CHICON, Patricia Mariotto Mozzaquatro²; RODRIGUES, Marcia Cristina Gomes³

Palavras- Chave: Mineração de Dados. Classificação. Perfil Universitário

INTRODUÇÃO

Durante muitos anos, e, com atividades relacionadas a escolas, o autor Nero (1977), teve ensejo de sentir de perto o problema do aproveitamento escolar. Neste contexto surge o problema de pesquisa: Como detectar o método de estudo universitário integrando fatores como o Perfil de hábitos relacionado aos seguintes tópicos: apontamento, método, provas, distribuição, concentração e compreensão? Com o intuito de solucionar a questão em si, serão utilizadas técnicas computacionais, a Mineração de dados, a qual irá detectar o perfil de estudo dos alunos universitários.

Nas seções a seguir serão descritos o embasamento teórico, modelagem, implementação e testes da aplicação.

MINERAÇÃO DE DADOS

Mineração de dados (MD) integrA uma das principais etapas na descoberta de conhecimento em banco de dados, denominada Knowledge Discovery Database (KDD), transformando dados brutos em informação.

Conforme Han:

Mineração de dados constitui o processo de descoberta de conhecimento interessante com a utilização de métodos e técnicas que permitem analisar grandes conjuntos de dados para a extração de informação previamente desconhecida, válida e que gera ações úteis, de grande ajuda para a tomada de decisão estratégica” (HAN, 2001).

A mineração de dados desdobra-se em técnicas específicas responsáveis pelo tratamento dos dados minerados, sendo elas: Classificação, Segmentação, Associação, Padrões Sequenciais e a Regressão. A pesquisa em desenvolvimento irá utilizar a técnica de classificação, descrita a seguir.

¹ Acadêmico do Curso de Ciência da Computação. E-mail: feschneider@unicruz.edu.br

² Professora do Curso de Ciência da Computação. E-mail: patriciamozzaquatro@gmail.com.

³ Funcionária do NAE - Universidade de Cruz Alta. E-mail: mrodrigues@unicruz.edu.br



A tarefa de classificação consiste em construir um modelo que possa ser aplicado em um conjunto de dados objetivando categorizá-los em classes. Um dado é analisado e classificado em uma classe definida, onde os algoritmos procuram a descoberta de funções e mapeiam registros em classes pré-definidas (HARRISON,1998) A técnica de classificação busca classificar itens similares conforme classes pré-definidas, identificando a qual classe cada registro pertence. Esta técnica integra métodos, neste pesquisa será abordado o método J48.

O método J48 tem a finalidade de gerar uma árvore de decisão baseada em um conjunto de dados de treinamento, sendo este modelo usado para classificar as instâncias no conjunto de teste (WITTEN et al., 2005). Para a montagem da árvore, o algoritmo J48 utiliza a abordagem de dividir-para-conquistar, onde um problema complexo é decomposto em subproblemas mais simples, aplicando recursivamente a mesma estratégia a cada subproblema, dividindo o espaço definido pelos atributos em subespaços, associando-se a eles uma classe (QUINLAN, 1993)

METODOLOGIA

No que se refere a abordagem, pode-se classificar a pesquisa como quantitativa, pois conforme o autor WAINER (2007, p. 05) A pesquisa quantitativa é baseada na medida (normalmente numérica) de poucas variáveis objetivas, na ênfase em comparação de resultados e no uso intensivo de técnicas estatísticas. As seguintes etapas são descritas: Etapa 1 – Estudo teórico: descoberta de conhecimento em base de dados, compreender diferentes técnicas e métodos, a relação das suas usabilidades no processo de Knowledge Discovery Database (KDD), pesquisar sobre a mineração de dados, técnica de classificação, funcionamento da árvore de decisão, descrever o método J48. Etapa 2 – Modelagem: construir o diagrama de casos de uso, classe, sequencia e modelagem ER. Etapa 3: implementação da aplicação na linguagem PHP e banco de dados MySql. Etapa 4: Validar a aplicação (Teste PHD) utilizando o teste caixa preta, ou funcional, aplicado no final, avaliando somente se os requisitos foram total ou parcialmente satisfeitos pelo software.

RESULTADOS

Nesse tópico serão apresentadas algumas telas do sistema. Ao acessar o sistema o aluno deverá inserir o seu RA (Registro Acadêmico) e sua senha, que utiliza para acesso no portal neste momento seus dados são coletados. A Figura 1 descreve a tela com o



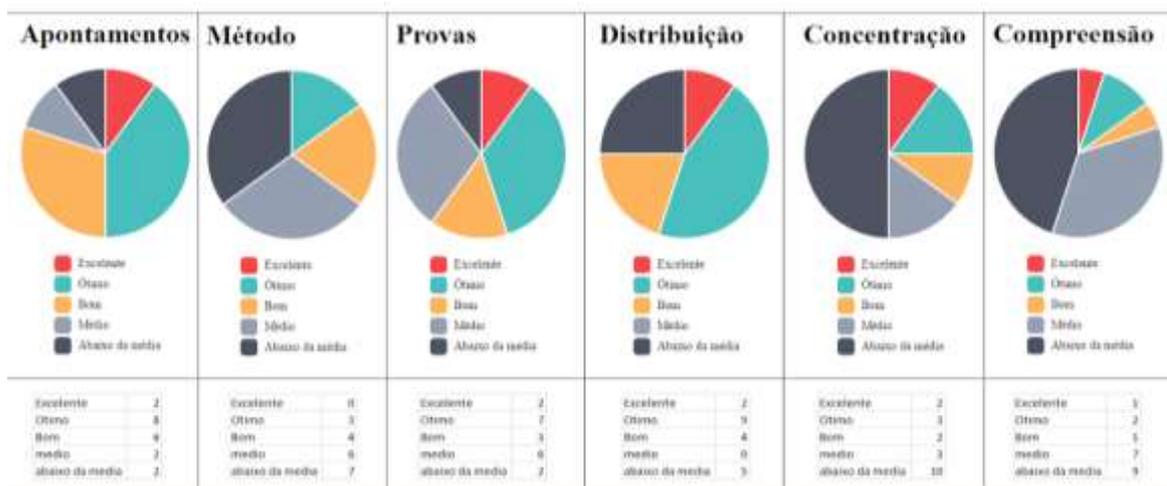
questionário. O aluno deverá responder uma série de perguntas, que compõem a Programação de Hábitos de Desempenho, elaborado por Carlos Del Nero (DEL NERO, 1977).

Figura 1. Tela Questionário



O sistema atualmente encontra-se em fase de testes, já foram aplicados os testes do tipo caixa branca, ou seja, estes testes foram realizados pelo programador. Após estão sendo aplicados os testes do tipo caixa preta, ou seja, testes realizados pela população em geral. O sistema foi aplicado na turma de Lógica para Computação integrante do curso de Ciência da Computação. Vinte alunos participaram da pesquisa, conforme a Figura 2.

Figura 2. Perfil da Turma



CONSIDERAÇÕES FINAIS

Foi desenvolvido um sistema computacional com a integração da técnica de classificação, mais especificamente o método J48, o qual realiza o diagnóstico de fatores como Apontamento, método, provas, distribuição, concentração e compreensão. A Mineração de dados transforma estes dados em conhecimento. A pesquisa encontra-se em fase de testes.



O sistema foi aplicado na turma de Lógica para Computação (primeiro semestre do Curso de Ciência da Computação).

O sistema encontra-se em fase de teste. Até o momento estão sendo aplicados testes do tipo caixa branca e caixa preta. O sistema será de fundamental importância para a Universidade de Cruz Alta. Como justificativa computacional, o presente trabalho busca aplicar na prática a técnica de classificação para a descoberta de conhecimento de forma automática. A justificativa social objetiva disponibilizar um sistema computacional para automatizar o processo de classificação de perfil universitário da Unicruz.

REFERÊNCIAS BIBLIOGRÁFICAS

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. California, 2. Ed. Morgan Kaufmann, 2001

HARRISON, T. H. **Intranet Data Warehouses**. Sao Paulo: Berkley Brasil, 1998.

NERO, Del Carlos, **PHD, Programa de hábitos e desempenho no estudo**. Editora Psico-Pedagogica Ltda. São Paulo, 1977.

QUINLAN, J. R.; **C4.5: programs for machine learning**. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

WAINER, J. **Métodos de pesquisa quantitativa e qualitativa para a ciência computação**. In T. KOWALTOWSKI e K. BREITMAN (Org.), Atualização em Informática da Sociedade Brasileira de Computação, pp. 221-262, 2007

WITTEN, I. H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2 ed. San Francisco: Morgan Kaufmann Publishers, 2005.