

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: **Computação Inteligente**

Título: **Mineração de Dados Educacionais utilizando Modelos de Regressão para Dados Abertos.**

Orientador – **Roberta Andrade de A. Fagundes** (roberta.fagundes@upe.br)

As universidades públicas, ao contrário do que ocorre no ensino fundamental, apresentam um nível de qualidade maior do que o encontrado nas universidades particulares. Diversos rankings nacionais e internacionais demonstram isso. Para manter tal nível de excelência e maximizar um dos maiores retornos que a instituição fornece à sociedade, que são os profissionais formados para a pesquisa científica e o mercado de trabalho, que a Universidade de Pernambuco, assim como as demais universidades públicas, precisam combater dois fenômenos que reduzem a taxa de conclusão dessas instituições: A evasão e a retenção. Diversas iniciativas já foram propostas, inclusive com a criação um programa do governo federal brasileiro para estimular as universidades, que estabeleceu metas para a diminuição desses dois problemas. A proposta desta pesquisa é criar uma abordagem apoiada na mineração de dados para a identificação de prováveis alunos retidos ou evadidos de acordo com suas informações sócio-econômico-culturais e de seu aproveitamento acadêmico no início do curso. Assim, a inteligência computacional [1] é uma técnica tem como objetivo fazer as máquinas resolverem problemas por meio da observação das pessoas ou apenas por meio do estudo de seus próprios métodos como por exemplo, regressão, redes neurais, árvores de decisão, algoritmos genéticos e etc. Com a explosão no volume de dados em gerando uma urgente necessidade de novas técnicas e ferramentas capazes de transformar, de forma inteligente e automática, esse imenso volume dado em informações significativas e em conhecimento para o negócio, gerando valor. Informações essas, muitas vezes implícitas sob uma montanha de dados, e não podem ser descobertas ou, no mínimo, facilmente identificadas utilizando-se sistemas convencionais de gerenciamento de banco de dados.

Em resposta a essa necessidade, surgiu o *Data Mining* (DM), também chamado de Mineração de Dados [2]. *Data Mining* é uma tecnologia que emergiu da intersecção de três áreas: estatística, inteligência artificial e aprendizado de máquina. Observa-se que o Data Mining é parte de um processo maior conhecido como KDD (*Knowledge Discovery in Databases*), em português, Os algoritmos utilizados no em DM são de diferentes áreas de conhecimento. Na estatística, essas técnicas podem ser aplicadas para descobrir estruturas ou associações em conjunto de dados, realizar previsões, etc. Dentre elas, destacam-se modelos de regressão [3], que são modelo matemáticos e tem como um dos objetivos prever o valor da variável dependente (Y) a partir das informações provenientes de um conjunto de variáveis independentes (X).

Neste contexto, este estudo propõe descrever as etapas para o desenvolvimento de modelos de previsão utilizando combinação de regressores para previsão. Para isso foram descritos os processos da técnica de mineração de dados, que extrai conhecimentos dados abertos na área de educação que serão utilizados para construção de um modelo de previsão. Para essa técnica, assume-se entendimento do negócio, dos dados e preparação, modelagem, avaliação e distribuição. Espera-se que a aplicação da técnica de mineração de dados permite a criação de um modelo de previsão utilizando combinação de regressores, a fim de promover avanços na área de educação. Durante o estudo um ambiente experimental para avaliação do modelo proposto será implementado, como também, serão utilizados dados (artificiais/reais) e o desempenho do modelo será medido pelo erro de predição estimado pelo método Monte Carlo no conjunto de teste utilizando a Linguagem R/Phyton.

Referências Bibliográficas

[1] NORVIG P. e RUSSELL S. **Inteligência Artificial**, 3ª Edição, 2013.

[2] GOLDSCHMIDT R., PASSOS E., BEZERRA E. **Mineração de Dados**, 2 ed., Editora Campus, 2015.

[3] MONTGOMERY D.C.; PECK, E.A. e VINING, G. G. **Introduction to Linear Regression Analysis**, Wisley-Interscience, 2006.