



## DATA MINING: ESTUDO DO FRAMEWORK PENTAHO UTILIZANDO BASES DE DADOS DO IBGE E PNUD

Orientador: FORCHESATTO, André Luiz

Pesquisador: SILVEIRA, Fábio da

Curso: Ciência da Computação

Área de Conhecimento: ACET

Quantas informações há nos computadores? Há algum tempo as pessoas preenchem formulários e efetuam cadastros em diversos tipos de estabelecimentos, como lojas, *sites*, hospitais, entre outros. Estes dados foram e estão sendo armazenados por estas organizações, dos mais variados tipos, criando verdadeiros oceanos de dados. Entretanto, estes dados ficavam guardados e, muitas vezes, não eram mais utilizados, até que surgiu a seguinte questão: será que não existem informações úteis escondidas neste universo de *bits*? Nesse contexto, surge a Mineração de Dados ou Data Mining. Com estudos e pesquisas, foi verificada a relevância do Data Mining para o sucesso das empresas, e este se tornou o suporte para um ramo conhecido como *Business Intelligence* (BI), o qual tenta sustentar e orientar a tomada de decisão das organizações modernas. Contudo, para tornar o Data Mining viável e se efetuar um BI de qualidade, é necessário um sistema computacional capaz de realizar todas as tarefas requeridas por estes. Dessa forma, surgem *frameworks*, como o Pentaho. O estudo apresenta a aplicação do Data Mining como técnica de *Business Intelligence* (BI), utilizando o *framework* Pentaho em bases de dados volumosas, como as do IBGE e do Programa das Nações Unidas para o Desenvolvimento (PNUD), ressaltando, passo a passo, o processo desde a instalação até a geração de relatórios, além da execução de uma exploração de dados. O resultado apresenta a exploração de informações úteis, tomando como base de demonstração os dados de renda per capita individual, nos municípios do território nacional, e o Índice de Desenvolvimento Humano (IDH).

Palavras-chave: Data Mining. *Business intelligence*. Pentaho. IBGE. PNUD.

andre.forchesatto@unoesc.edu.br

silveira-fabio@hotmail.com

