# Mineração de Dados - tp 3
# KNN

Thales Filizola Costa
*thalesfc@dcc.ufmg.br*

15 de janeiro de 2013

---

## 1 Introdução

A seção 2 retrata a base de dados utilizada neste trabalho, tendo como foco a análise de stemming e de retirada de stopwords e relacionando os mesmos com os efeitos na acurácia e na performance do algoritmo.

## 2 Base de Dados

A base de dados utilizada consiste em *reviews* de usuários para filmes. Para cada *review* são fornecidos o comentário e a classificação do comentário, que pode ser positiva (1) ou negativa (0).

### 2.1 Vocabulário

Nessa subseção queremos determinar se vamos utilizar ou não *stemming* e se vamos retirar ou não *stopwords* para os próximos experimentos.

A tabela **??** apresenta as diferentes configurações que foram executadas. Nesses experimentos, o número de vizinhos utilizados foi dez (`k=10`) e a distância entre os pontos foi calculada utilizando a distância euclideana (`d=euclidean`).

## Referências

[1] Michael Hahsler. A Comparison of Commonly Used Interest Measures for Association Rules. `http://michael.hahsler.net/research/association_rules/measures.html`, 2011. [Online; accessed 13-October-2012].

[2] Mohammed Zaki and Wagner Meira Jr. *Fundamentals of Data Mining Algorithms*. Cambridge University Press, 2010.