

O que é Mineração de Dados?

Carlos Diego Rodrigues

28 de setembro de 2021

Universidade Federal do Ceará

Introdução

- A mineração de dados é uma tecnologia que vem amadurecendo nos últimos 40 anos e recentemente ganhou muita evidência por aplicações de sucesso.
- Para alguns ela ainda é um mistério.
- Nosso primeiro objetivo neste curso é desvendar este mistério.
- Nosso segundo objetivo neste curso é entender o que está por trás das cortinas, os princípios matemáticos, estatísticos e computacionais dos principais algoritmos de Aprendizado de Máquina.
- Nosso terceiro objetivo neste curso é desenvolver a capacidade de propor soluções em Mineração de Dados para problemas reais.

Nuvem de palavras



Mineração de dados é o futuro!

- Até 2005 os seres humanos criaram 130 Exabytes de dados...
- O que é um exabyte?
 - Um caractere... 1 byte
 - Uma página de um livro... 1 kilobyte
 - Um livro de 1024 páginas... 1 megabyte
 - 1024 livros... 1 gigabyte
 - Se você transformar toda a Amazônia em papel... 1 terabyte
 - 1024 terabytes = 1 exabyte
- Até 2010 os seres humanos criaram 1200 exabytes!
- Até 2020 os seres humanos criaram 40.000 exabytes!

Primeira parte: um curso prático!

- A primeira parte do nosso curso será dedicada a desvendar na prática a Mineração de Dados.
 - Conceitos
 - Algoritmos simples e populares.
 - Avaliação e validação.
 - Aplicação prática
- Software especializado: Weka!

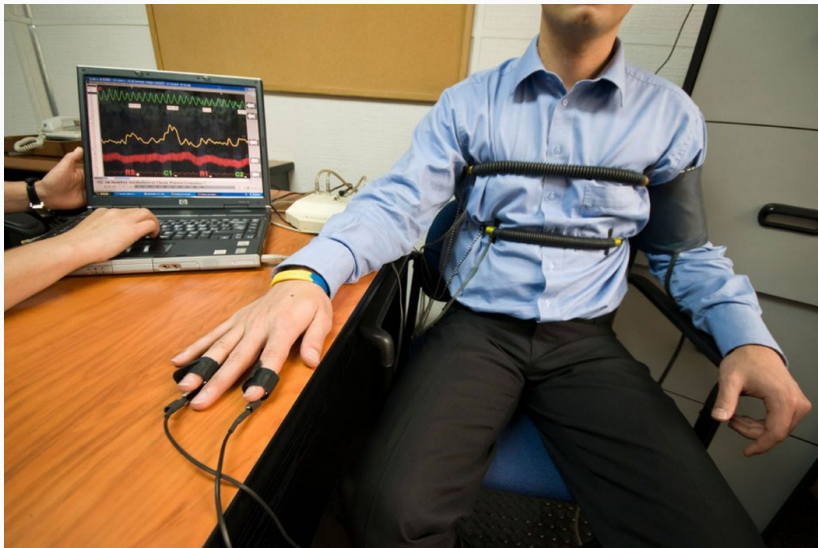
O que é Mineração de Dados?

- Contexto: vivemos em um mundo que transborda dados!
Nossos bancos de dados estão cheios!
- Mineração de Dados é o conjunto de ciências e tecnologias capazes de transformar esse mar de dados em informações capazes de fazer previsões úteis.
- Exemplos:
 - carrinho de compras no supermercado;
 - prevenção e repressão de crimes (detecção de fraudes, detecção de mentiras);
 - redes sociais;
 - fertilidade humana.
- **Mineração de Dados vs. Aprendizado de Máquina.**

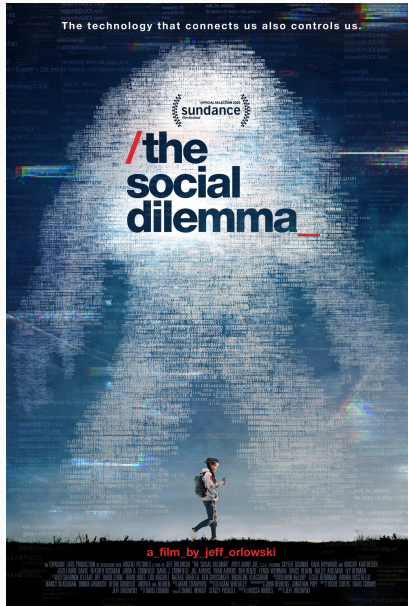
Farmácias e o seu CPF



Prevenção/repressão de crimes



Redes sociais





- O que é Weka?
- Waikato Environment for Knowledge Analysis
- É um mesa de trabalho para Mineração de Dados.
- Diversos algoritmos de Aprendizado de Máquina implementados:
 - 100+ Algoritmos de classificação
 - 75+ Algoritmos de pré-processamento de dados
 - 25+ Algoritmos de seleção
 - 20+ Algoritmos de agrupamento/associação
- Ferramentas de avaliação, visualização, etc.

O que vamos aprender?

- Carregar uma base de dados.
- Preprocessamento através de filtros.
- Exploração e visualização interativa.
- Aplicar algoritmos de classificação.
- Interpretar os resultados dos algoritmos.
- Entender os métodos de avaliação e suas implicações.
- Entender as representações dos modelos.
- Entender como funcionam alguns algoritmos de Aprendizado de Máquina.
- Entender as principais dificuldades na aplicação da Mineração de Dados.