

# Data Mining

---

Guilherme Guy de Andrade - 16/0123186

Leonardo dos Santos S. Barreiros - 15/0135521

# O que é Data Mining?

- Processo de análise de grande quantidade de dados
- Busca por padrões
- Descoberta de conhecimento
- Entendimento sobre relações que se manifestam dentro dos dados

# O que é Data Mining?

Descende fundamentalmente de 3 linhagens:

- Estatística
- Inteligência Artificial
- Machine learning

# Para que usar Data Mining?

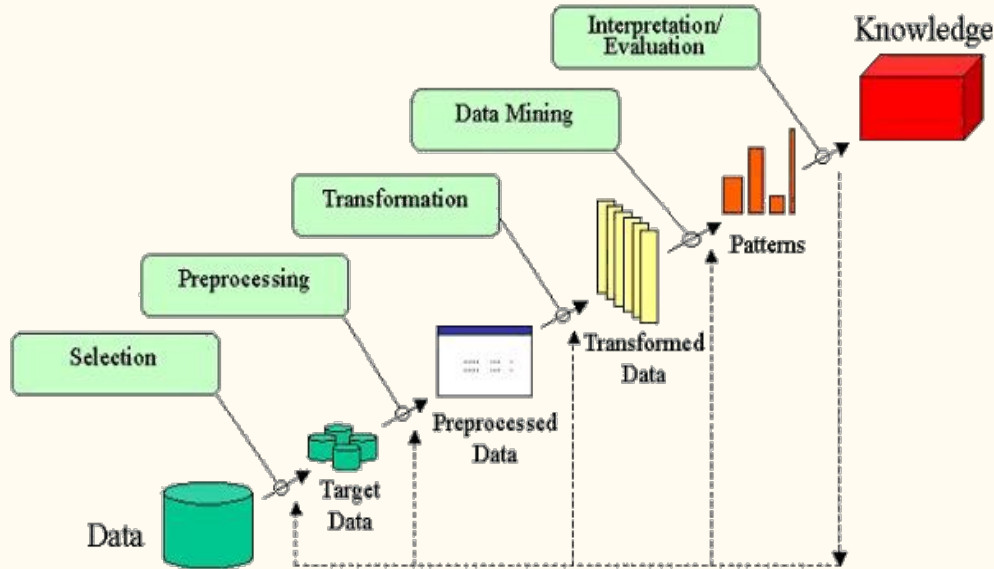
- Conhecer o comportamento de certos dados para tomar decisões de negócio
- Validar uma decisão de negócio com base em dados existentes
- Extrapolar dados de uma base de dados

# Como aplicar Data Mining?

- Geralmente é usado uma metodologia
- Se aplicam algoritmos e técnicas no processo
- A análise de dados depende de conjunto de dados

# Processo de Data Mining

## Knowledge Discovery in Databases (KDD)

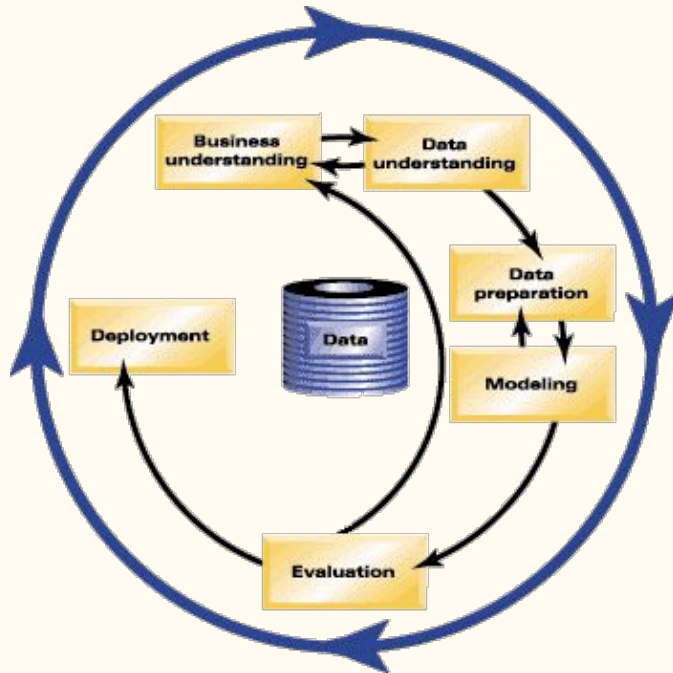


- Processo generalizado para descoberta de conhecimento

Fonte: [http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\\_kdd.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html)

# Processo de Data Mining

## Cross-Industry Standard Process for Data Mining (CRISP-DM)



- Modelo de data mining criado com base em necessidades do mercado
- Cíclico
- Apoiado em BI

# Técnicas

- Classificação
- Estimativa
- Previsão
- Análise de afinidade
- Análise de Agrupamentos
- Protocolo de mineração de dados



# Algoritmos

- Redes neurais
- Árvore de Decisão
- Regressão Linear
- Regressão Lógica
- Clustering
- Série Temporal
- Algoritmo de Redução Dimensional

# Vantagens

- Tratamento e tomada de decisões em Big data → Volume
- Precisão de dados → Machine learning
- Identificação de padrão ou divergência → Comportamento
- Previsão de tendências → Informações que o dado trás
- Gerar informações que estão ocultas → Derivações
- Dados fáceis de serem interpretados

# Desvantagens

- Alto custo → relacionado a complexidade
- Exige uma base de dados relativamente volumosa
- Complexidade dos dados em relação a comportamento/tendência

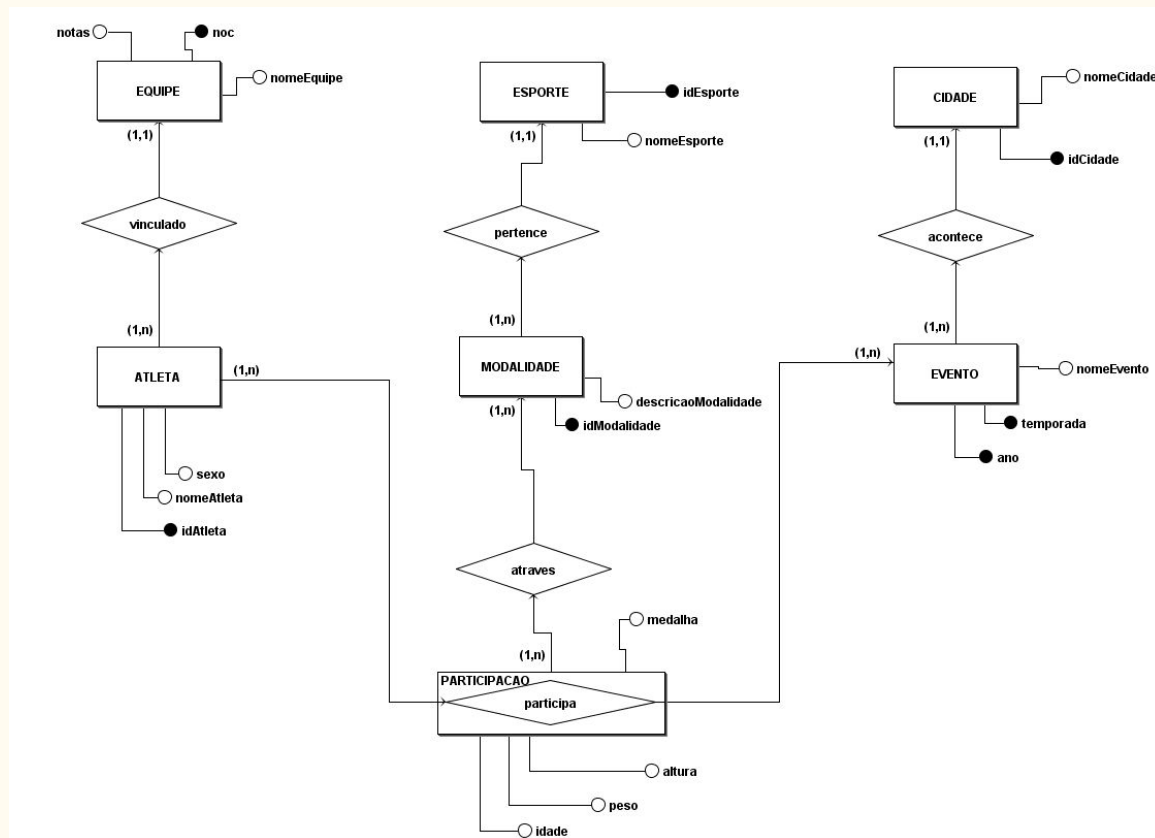
# Aplicação

---

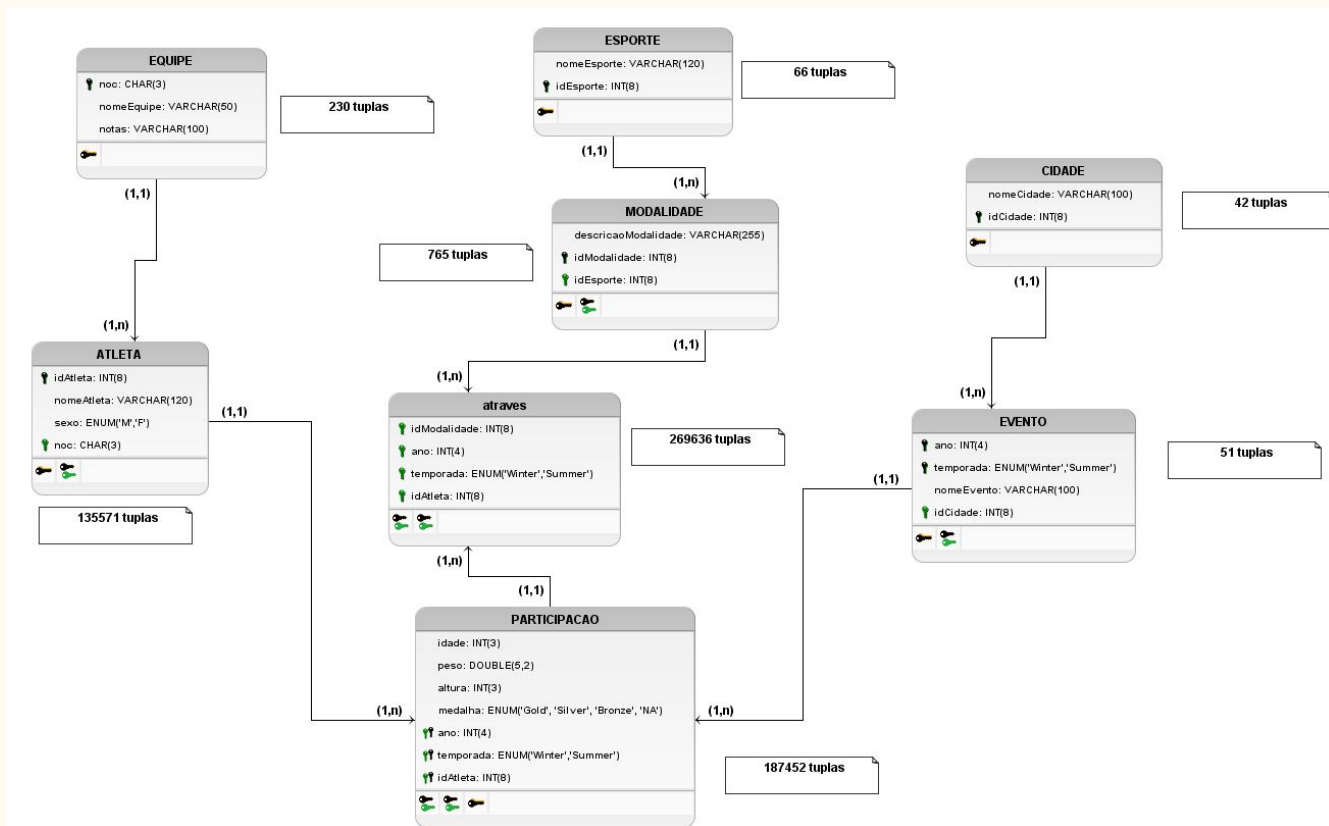
# 120 anos de história olímpica: Atletas e resultados

- Dataset com os dados históricos dos jogos Olímpicos modernos, incluindo todos os jogos de Atenas 1896 ao Rio 2016.
- Apresenta dados dos atletas que competiram nos períodos em questão, informando nome, idade, peso, altura, time representante, qual olimpíada participou, modalidade, cidade, estação do ano, evento, medalha e comitê olímpico responsável.

# Diagrama Entidade Relacionamento



# Diagrama Lógico



# Referências

MORO, Sergio; LAUREANO, Raul; CORTEZ, Paulo. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In: Proceedings of European Simulation and Modelling Conference-ESM'2011. EUROSIS-ETI, 2011. p. 117-121.

CARVALHO, L. A. V.. Datamining A Mineração de Dados no Marketing , Medicina , Economia , Engenharia e Administração. 5. ed. Rio de Janeiro: CIENCIA MODERNA, 2015. 189p .

<https://docs.microsoft.com/en-us/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining>

<https://www.devmedia.com.br/conceitos-e-tecnicas-sobre-data-mining/19342>

<https://www.tiespecialistas.com.br/as-tecnicas-do-datamining/>

<https://www.dataversity.net/10-most-popular-data-mining-algorithms/>

<https://www.kdnuggets.com/2015/05/top-10-data-mining-algorithms-explained.html>

<https://portogente.com.br/portopedia/73758-datamining>

<https://transformacaodigital.com/o-que-e-data-mining/>



Fim

—