

Introdução à Ciência dos Dados

Aula 02

Gabriel Pacheco

Bacharel em Sistemas de Informação (UFMG)

Mestrando em Ciência da Computação (UFMG)

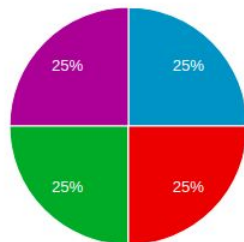
Trabalha em Digicade desde 2016



Resultado do formulário - Aula 01

Qual dos conteúdos abaixo você gostaria de ver na próxima aula?

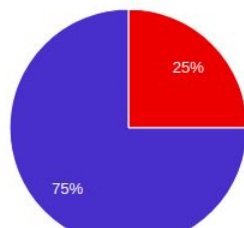
8 respostas



- Limpeza e análise dos dados
- Treino, teste e validações
- Agrupamento
- Classificação
- Associação
- Técnicas de regressão avançadas
- Maldição da Dimensionalidade

Qual dos exemplos práticos você gostaria de ver na próxima aula?

8 respostas



- Utilizando regressão com dados do mercado financeiro
- Regras de associação com dados sobre arborização urbana
- Agrupamento com dados sobre arborização urbana



Resultado do formulário - Aula 01

Outros tópicos

- Aulas 2x por semana
- “Exemplificar a inteligência artificial utilizada no nosso dia a dia, por exemplo o que pode ser IOT num futuro próximo...”



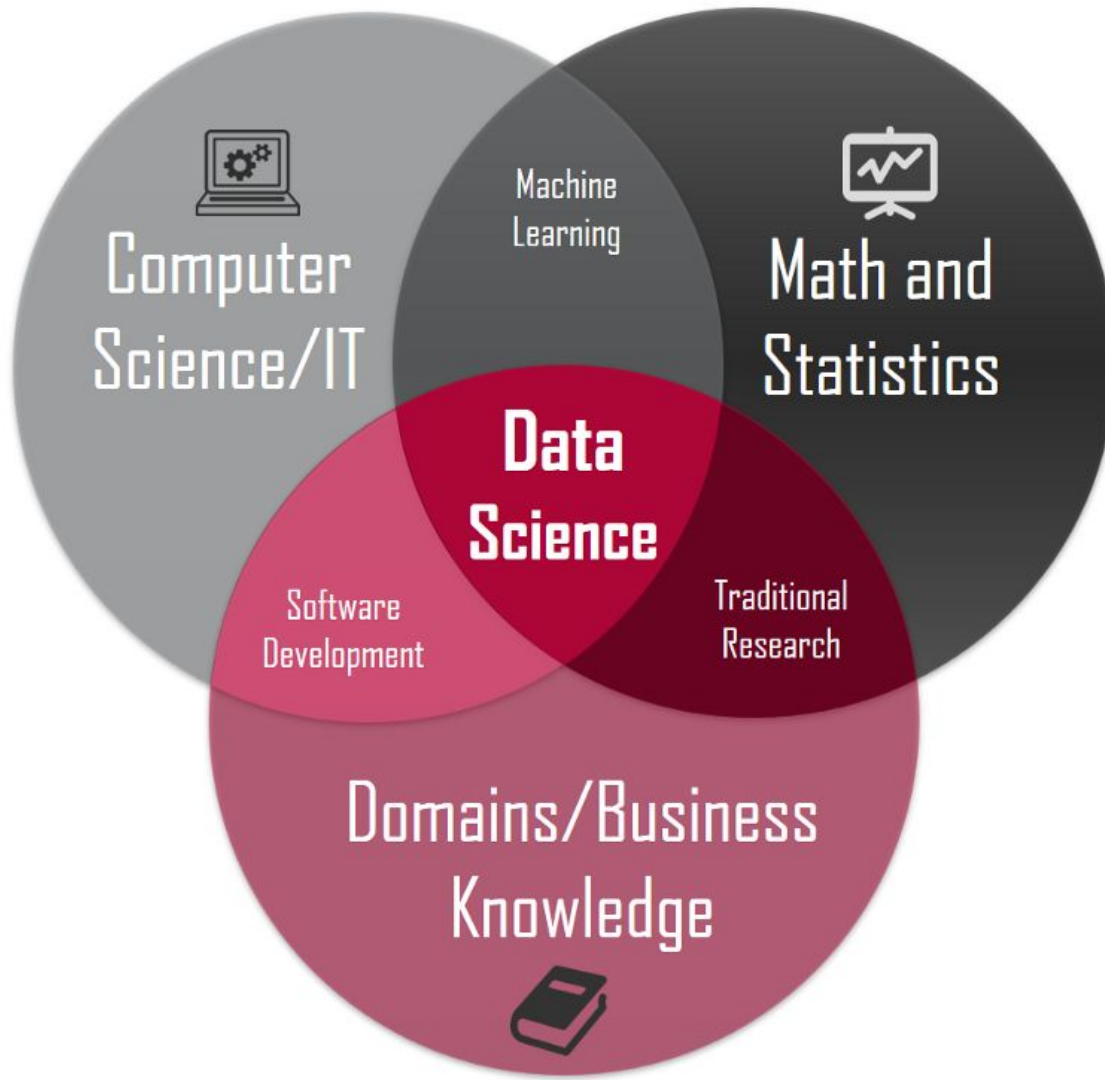
Curiosidade

- <https://quickdraw.withgoogle.com/>
- <https://www.tecmundo.com.br/software/151406-novo-profissional-inteligencia-artificial.htm>



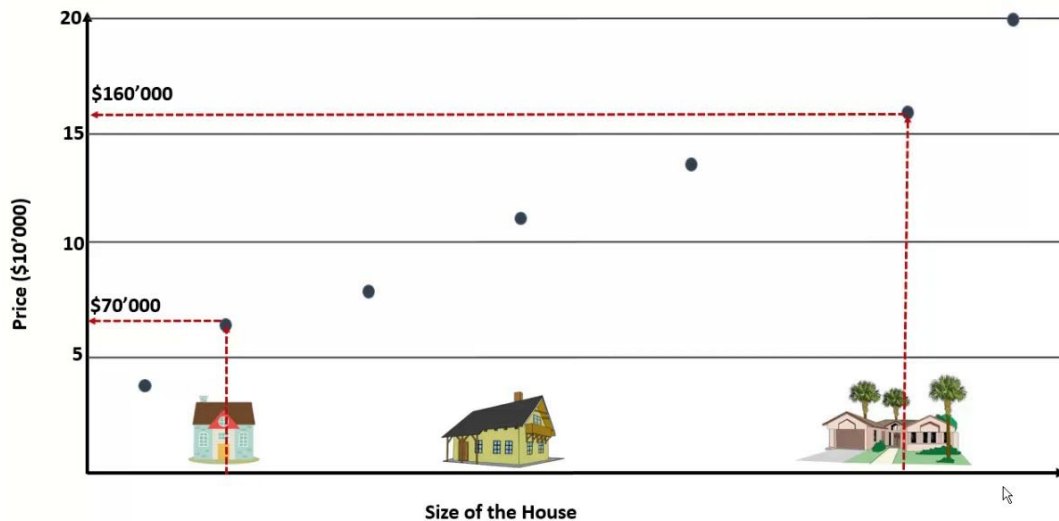
Cronograma

- Revisão
- Técnicas de regressão avançadas
- Exemplo prático
- Ferramentas



Regressão

- Predição de preços
- Predição de demanda*
- Predição de idade





Técnicas de regressão avançadas

- Árvores de decisão (decision trees)
- Florestas aleatórias (random forest)



Árvores de decisão

- Fácil de criar
- Fácil de utilizar
- Fácil de interpretar

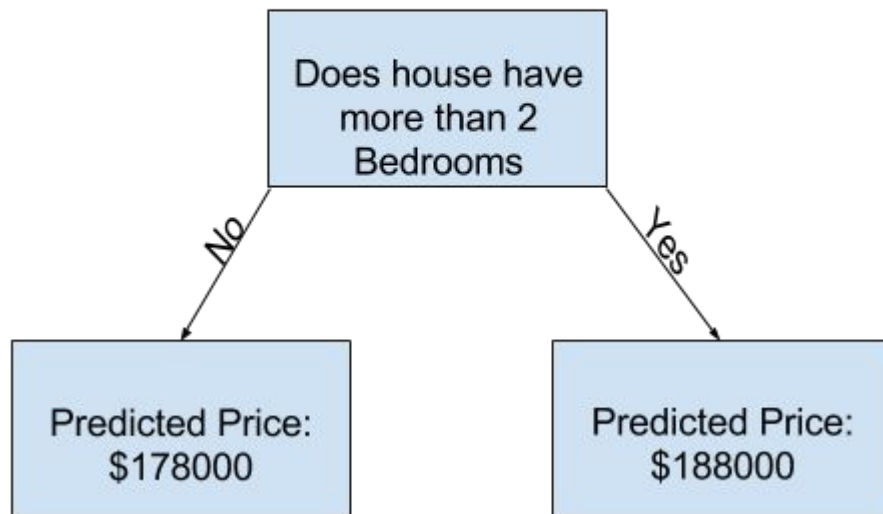
Porém...

- Imprecisa, ou seja, não funciona bem para classificar novas amostras

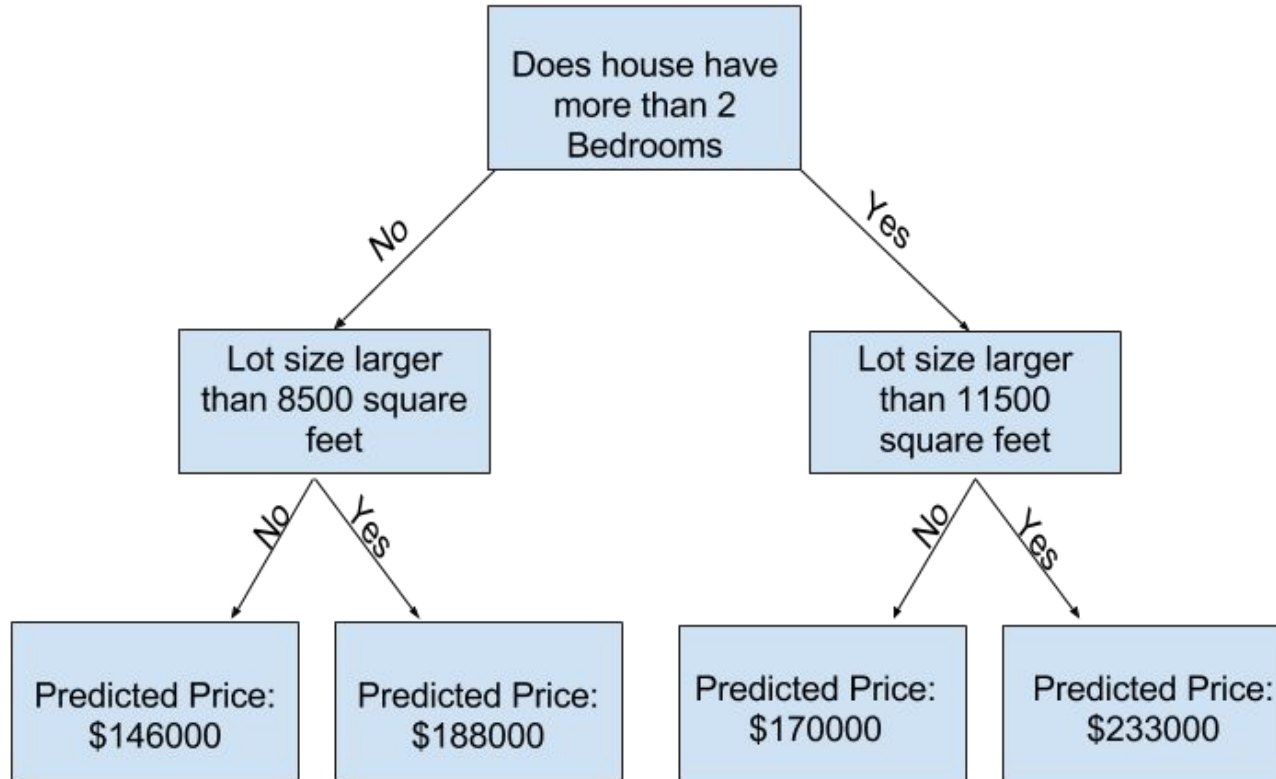


Árvores de decisão

Sample Decision Tree



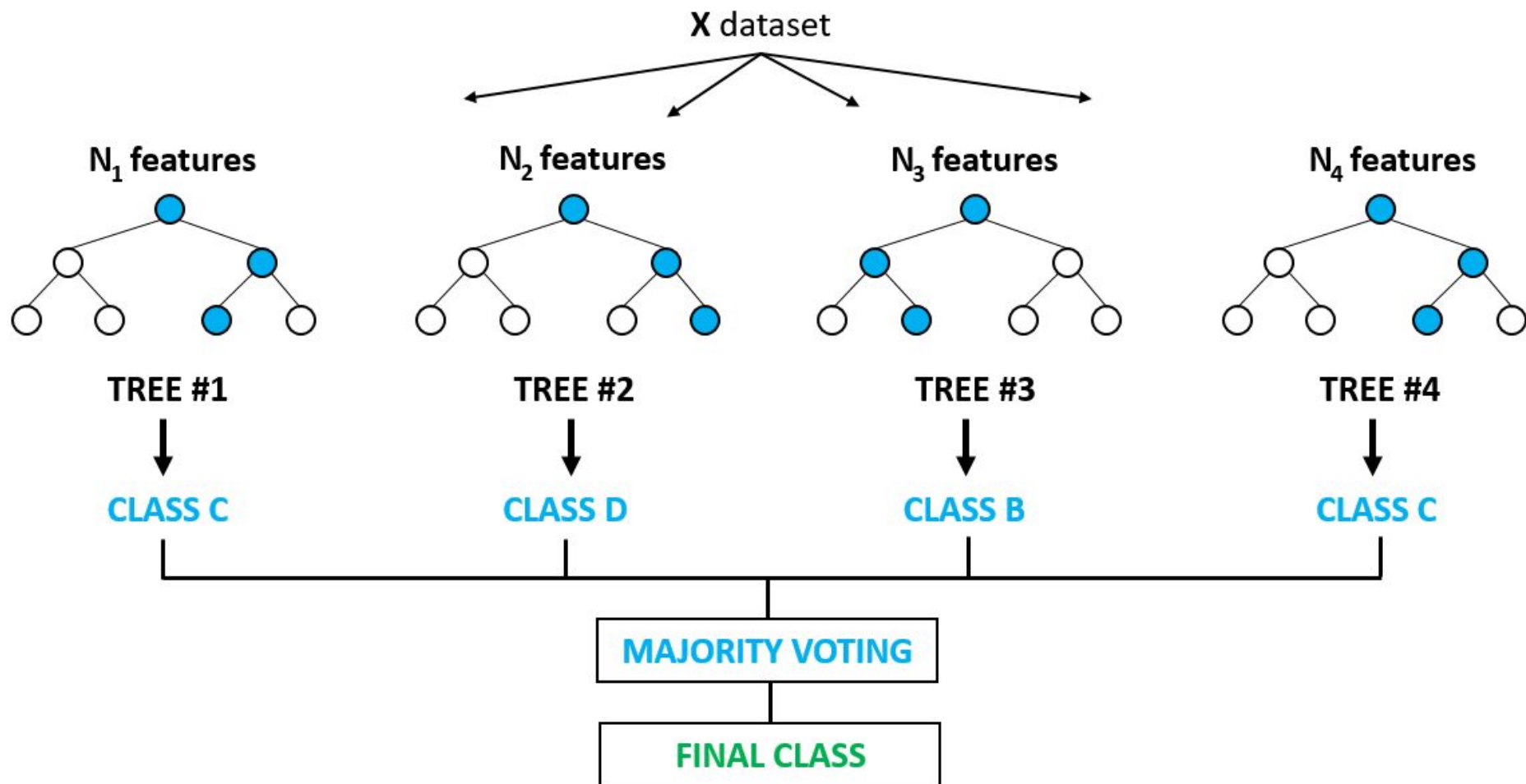
Árvores de decisão





Florestas aleatórias

- Técnica baseada em árvores de decisão
- Adiciona conceitos de aleatoriedade para aumentar a precisão





Exemplo prático: Mercado de Ações

- **Grafismo** x Fundamentalismo
- Base de dados
- Acesso aos dados
- Modelos...?
- Estratégia
- Backtesting (<https://www.quantopian.com/>)



Ferramentas e Links úteis

- Ferramentas e bibliotecas para programação
 - [Jupyter](#)
 - [Scikit-Learn](#)
 - [Pandas](#)
- Ferramentas “drag-and-drop”
 - [Knime](#)
- Base de dados e competições para treinar
 - [Kaggle](#)
- Backtesting
 - [Quantopian](#)
- [Smarttbot](#)



Tutoriais

- <https://www.kaggle.com/dinasinclair/housing-prices-from-basics-to-random-forest>
- <https://www.kaggle.com/dansbecker/using-categorical-data-with-one-hot-encoding>



Encerramento

- Materiais do curso: <https://github.com/gapacheco/ICD.git>
- Formulário de sugestões: <https://forms.gle/gvPWZGCjdgwra42i8>



Contato

- Email: gacampacheco@gmail.com
- LinkedIn: www.linkedin.com/in/gabriel-oc-pacheco