Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

MAC0459/MAC5865 - Tópicos em Ciência e Engenharia de Dados

Aula 00

Sejam bem-vindas, sejam bem-vindos!

Entre no link https://app.sli.do/event/dbuxvuoz ou e faça suas perguntas da aula.

R. Hirata Jr.

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Objetivos de hoje

- Ao final da aula de hoje você deve saber responder:
  - Por que o Hirata está ministrando esta disciplina
  - Por que você está aqui
  - A diferença entre:
    - o método do cientista
    - o método do engenheiro
    - o método do cientista de dados
  - Os papéis, ou trabalhos, em Ciência de Dados

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

MAC0459/MAC5865 - Syllabus

- **Objetivos:**
  - Ao final da disciplina o estudante deverá saber os fundamentos e as técnicas para manipulação, representação, análise, modelagem e validação de grandes conjuntos de dados.
- **Rational:**
  - A facilidade de coleta de dados alcançada pela civilização atual nos impõe um grande desafio: armazenar, representar, analisar e modelar o que usualmente é conhecido como “avalanche de dados”. A formação dos graduandos e pós-graduandos em Ciência da Computação não pode dispensar o conhecimento dos referidos fundamentos e técnicas proporcionados por esta disciplina.

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

MAC0459/MAC5865 - Syllabus

- **Conteúdo:**
  - Importância da área e de suas aplicações. Processo de descoberta do conhecimento (KDD) em conjuntos de dados. Tratamento, representação e qualificação de grande volumes de dados. Armazém de dados e modelos multidimensionais. Indexação e recuperação de grande volumes de dados. Grafos em Bancos de Dados. Análise exploratória de dados (análise de agrupamentos e associações de dados). Modelagem de conhecimento (classificadores, regras de classificação, exemplos). Desenvolvimento e uso de software para KDD. Exercícios com utilização de dados simulados e reais.

- **Critério de aprovação:**
  - Provas escritas, exercícios programados e presença

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

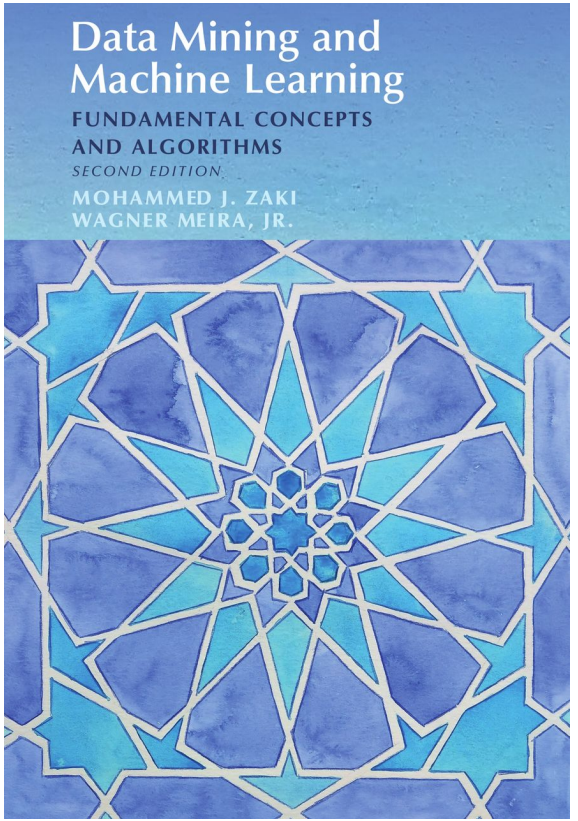
Referência principal

- The Data Science Design Manual
  - http://www.data-manual.com/
  - What is Data Science?

no e-disciplinas

Outras referências

- Data Mining and Machine Learning
- https://dataminingbook.info/
- Palestra do prof. Meira na disciplina de MAC5865
  - http://iptv.usp.br/portal/video.action?idItem=22508



INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
IME 50
UNIVERSIDADE DE SÃO PAULO

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Por que o Hirata está aqui?

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Quem é o Hirata

- Graduação
  - IC - Topologia diferencial, modelagem matemática, computadores e educação
- Pós-graduação
- Posdoc/professor de faculdade
- Pesquisador/professor
- Administrador

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
IME 50
UNIVERSIDADE DE SÃO PAULO

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Por que você está aqui?

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Quem é você?

- Formulário Google:

https://docs.google.com/forms/d/e/1FAIpQLSfqCSJaIX0jEfNEVOJ3lsbmdlCMDlPUIva6do_r3qC4ZyPCDw/viewform?usp=pp_url

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Ciência, Engenharia e DS

Simple pipeline – Scientific Method

1. Pose a question
2. Formulate a hypothesis
3. Formulate an experiment
4. Observe (data collecting)
5. Analyse the results
6. Go back to step 2 if the hypothesis is not correct/supported
7. Report results

Simple pipeline – Engineering Method

1. Define a problem
2. Specify requirements
3. Brainstorm, evaluate, choose solution
4. Develop a prototype solution
5. Tests solution
6. Go back to step 3 if the results, or data, etc in case the solution does not meet requirements

Simple pipeline – Data Science Method

1. Pose question
2. Get the data
3. Explore the data
4. Model the data
5. Report results

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Data Science Jobs

Data Science Jobs

- Data analyst
  - SQL skills
  - Basic visualizations
  - Reporting dashboards
  - Analyse A/B/N tests

Data Science Jobs

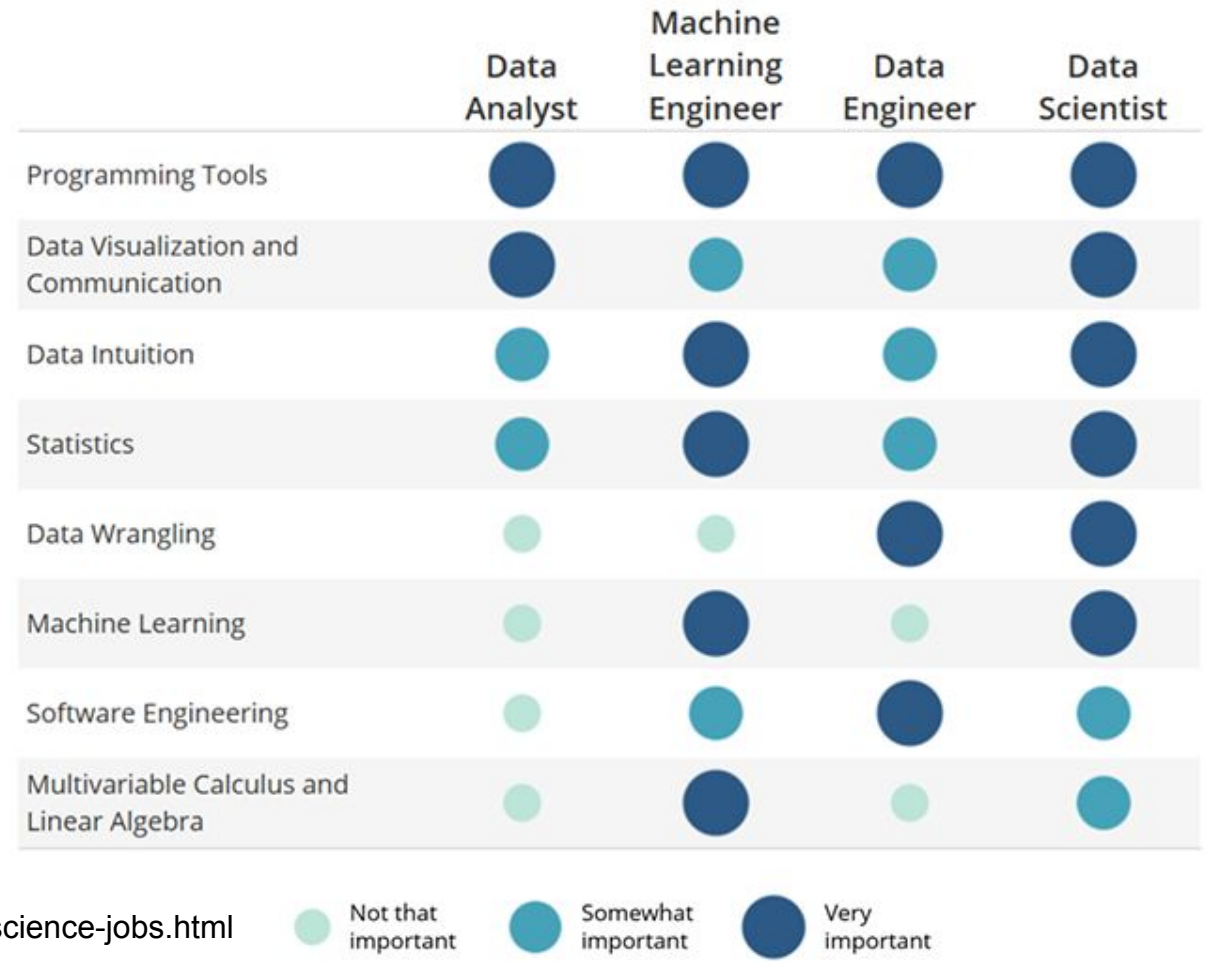
- Data engineer/scientist
  - Big data facilitator
  - Data intensive applications
  - Statistics and ML vs Software Engineering expertise

Data Science Jobs

- Machine learning engineer
  - Formal math/stat background
  - Data driven products
  - Massive amount of data
  - Data-based service

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Data Science Jobs



https://blog.udacity.com/2018/01/4-types-data-science-jobs.html

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
IME 50
UNIVERSIDADE DE SÃO PAULO

Escreva sua pergunta aqui: https://app.sli.do/event/dbuxvuoz

Obrigado!