

# REFERÊNCIAS

# Cursos na plataforma

Stanford



Minning Massive Datasets



Machine Learning



Introdução a Ciência de Dados



Especialização Machine Learning

- (1) Fundamentos (2) Regressão (3) Classificação (4) Cluster (5) Sistemas de recomendação e redução de dimensão (6) Projeto (Capstone)



Especialização Data Mining:

- (1) Padrões, (2) Texto e Buscadores, (3) Cluster, Text Mining e Visualização de dados



Especialização Ciência de Dados:

- (1) Ferramentas, (2) R, (3) Extração e Limpeza, (4) Análise Exploratória, (5) Pesquisa Reprodutível, (6) Inferência Estatística, (7) Modelos Regressivos, (8) Aprendizagem Automática (Machine Learning), (9) Desenvolvendo Aplicação

Especialização Executiva em Data Science:

- (1) Introdução (2) Equipe (3) Gestão (4) Problemas reais (5) Projeto (Capstone)

# Minning Massive Datasets

Stanford



- MapReduce (modelo programação Hadoop)
- Uso de PageRank, grafos, medidas de similaridade
- Clustering, sistemas de recomendação
- Técnicas para grandes volumes de dados: redução de dimensão, data stream, LSH, SVM
- Boa parte do conteúdo em [mmds.org](http://mmds.org) (inclusive um livro)

# Machine Learning (Andrew Ng)

Stanford



- Forte embasamento matemático
- Uso de equações diferenciais e álgebra matricial (matemática aplicada) para método de mínimos quadrados
- Todos exercícios com uso de MatLab/Octave
- Transcrição das aplicações matemáticas em algoritmos

# Introdução a Ciência de Dados



Foco em programação

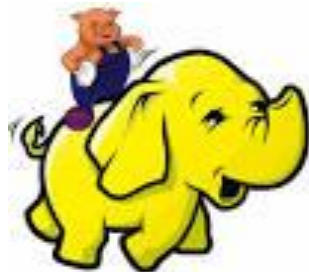
Fornece uma máquina virtual

Sentiment Analysis Python (twitter's crawler)

MapReduce: implementação em Python

Competição Kaggle

NoSQL: Pig (sobre Hadoop), utilizando AWS



# Especialização Data Mining



Embasamento matemático e estatístico

Ferramenta específica (Meta - C++) para Text Mining

## Aplicações

1. Ex.: fralda + cerveja
2. Ex.: google rank (progr.)
3. Ex.: vários tipos de cluster

1

Descobrendo padrões aplicáveis na mineração de dados

2

Recuperação de texto e motores de busca

3

Análise de Cluster em Data Mining

4

Mineração de Texto e Análise Estatística

5

Visualização de Dados

# Especialização Data Science



Todos os módulos com uso de R

1. Overview R, técnicas
2. Uso do R
3. Extração com R
4. Gráficos, descrição dados
5. R Markdown
6. Estatística clássica
7. Regressão linear e variantes
8. Básico de Machine Learning
9. Aplicações com RServer/Slides

- 1 As Ferramentas do Cientista de Dados
- 2 Linguagem R
- 3 Obtenção e Limpeza de Dados
- 4 Análise Exploratória de Dados
- 5 Pesquisa Reprodutível
- 6 Inferência Estatística
- 7 Modelos Regressivos
- 8 Aprendizagem Automática na Prática
- 9 Desenvolvimento de dados

# Plataformas Cloud

Máquinas Virtuais (Linux/Win),



Modelos de Machine Learning



R, Python, Scala, Spark integrados  Data Scientist Workbench



R Studio Cloud (rstudio.cloud)





# Cursos na plataforma

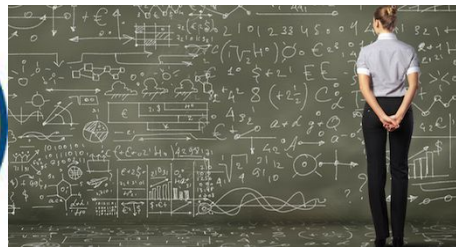


Aplicações

Cloud, Databases

Plataformas distribuídas

NoSQL, Fast Algorithms,  
Data Compression



Overview BigData

Spark, pySpark

spark MLlib

Introdução a Big Data com Apache Spark

# Bibliografia

- Estatística Aplicada:  
**Levine, Stepah, Krehbiel, Berenson**, ESTATÍSTICA TEORIA E APLICAÇÕES (com uso do Excel). 5.a edição, LTC.
- Aplicações em Business Intelligence e Big Data:  
**Dean, Jared**, BIG DATA, DATA MINING AND MACHINE LEARNING. Wiley.
- Data Mining:  
**Han, Kamber**, DATA MINING - Concepts and Techniques, 2.a edição, Morgan Kaufmann.
- Data Mining e Big Data:  
**Leskovec, Rajaraman, Ullman**, MINING OF MASSIVE DATASETS. 2014,  
<http://mmds.org>.

# Bibliografia

- **FÁVERO, L. P. L, et al.** *Análise de Dados - Modelagem Multivariada para Tomada de Decisão*, São Paulo, Campus/Elsevier, 2009.

Free books:

- **COGHLAN, Avril.** *A Little Book of R for Multivariate Analysis (free book)*. Disponível em: <http://bit.ly/1PTsN48>, 2014.
- **HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J.** *The Elements of Statistical Learning, 2nd Ed. (free book)*. Disponível em: <http://stanford.io/2OSfKqL>, 2008.
- **JAMES, Gareth, et al.** *An Introduction to Statistical Learning, with applications in R (free book)*, Disponível em: <http://bit.ly/1Grdeen>, 2014.