

Processamento e análise para Big Data

DATA SCIENCE & ANALYTICS
SENAI CIMATEC

Robespierre Pita

Bacharel em Sistemas de Informação - UNIFACS

Especialista em Redes de Computadores e Telecom - UNIFACS

Mestre em Ciência da Computação - UFBA (integração de dados em Big Data)

Doutorando em Ciência da Computação - UFBA (machine learning)

Cientista de Dados no CIDACS.



UNIVERSIDADE FEDERAL DA BAHIA

DISSERTAÇÃO DE MESTRADO

**CORRELAÇÃO PROBABILÍSTICA IMPLEMENTADA EM
SPARK PARA BIG DATA EM SAÚDE**

ROBESPIERRE DANTAS DA ROCHA PITA

Mestrado Multiinstitucional em Ciência da Computação

Salvador
05 de março de 2015



Federal University of Bahia
Institute of Mathematics and Statistics

Graduate Program in Computer Science

**CLUSTERING CATEGORICAL DATA USING
THE FREQUENCY FACTOR**

Robespierre Dantas da Rocha Pita

TESE DE DOUTORADO



UNIVERSIDADE FEDERAL DA BAHIA

DISSERTAÇÃO DE MESTRADO

**CORRELAÇÃO PROBABILÍSTICA IMPLEMENTADA EM
SPARK PARA BIG DATA EM SAÚDE**

ROBESPIERRE DANTAS DA ROCHA PITA

Mestrado Multiinstitucional em Ciência da Computação

Salvador
05 de março de 2015



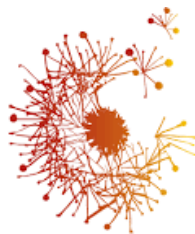
Federal University of Bahia
Institute of Mathematics and Statistics

Graduate Program in Computer Science

**CLUSTERING CATEGORICAL DATA USING
THE FREQUENCY FACTOR**

Robespierre Dantas da Rocha Pita

TESE DE DOUTORADO



cidacs
Centro de Integração de Dados
e Conhecimentos para Saúde
FIOCRUZ

E você?

- Qual a relação entre ciência de dados e Big Data?
- O que é aprendizado de máquina?
- O uso de dados já melhorou alguma atividade sua?
- Qual sua formação?
- Já trabalha com dados grandes? Onde? Como? Que análise fez?
- Já trabalha com aprendizado de máquina? Onde? Como?
- Quais as expectativas com a disciplina?



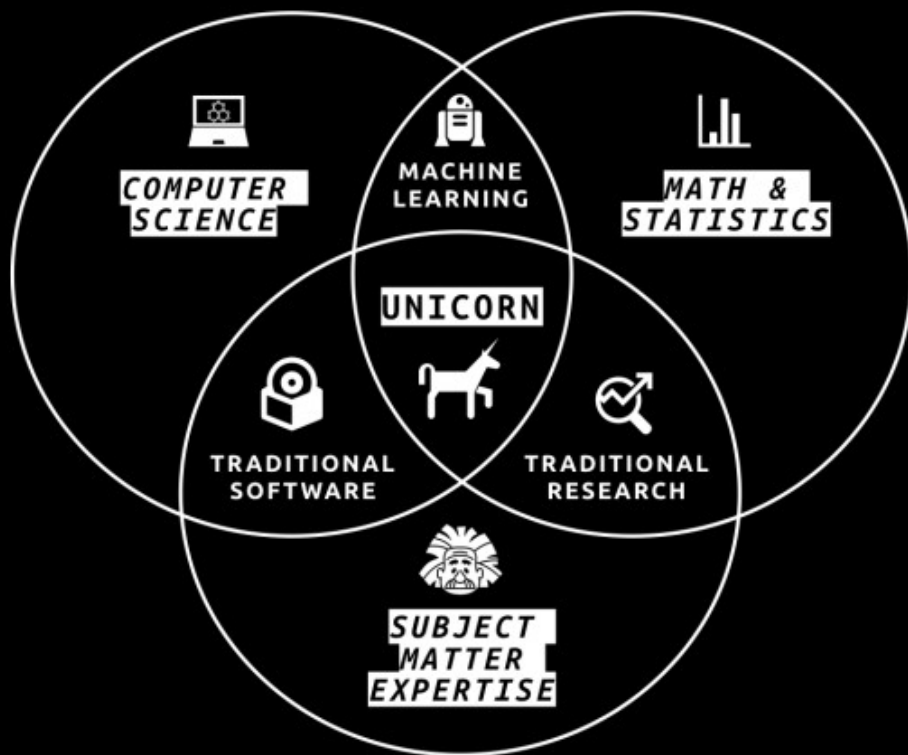
DATA

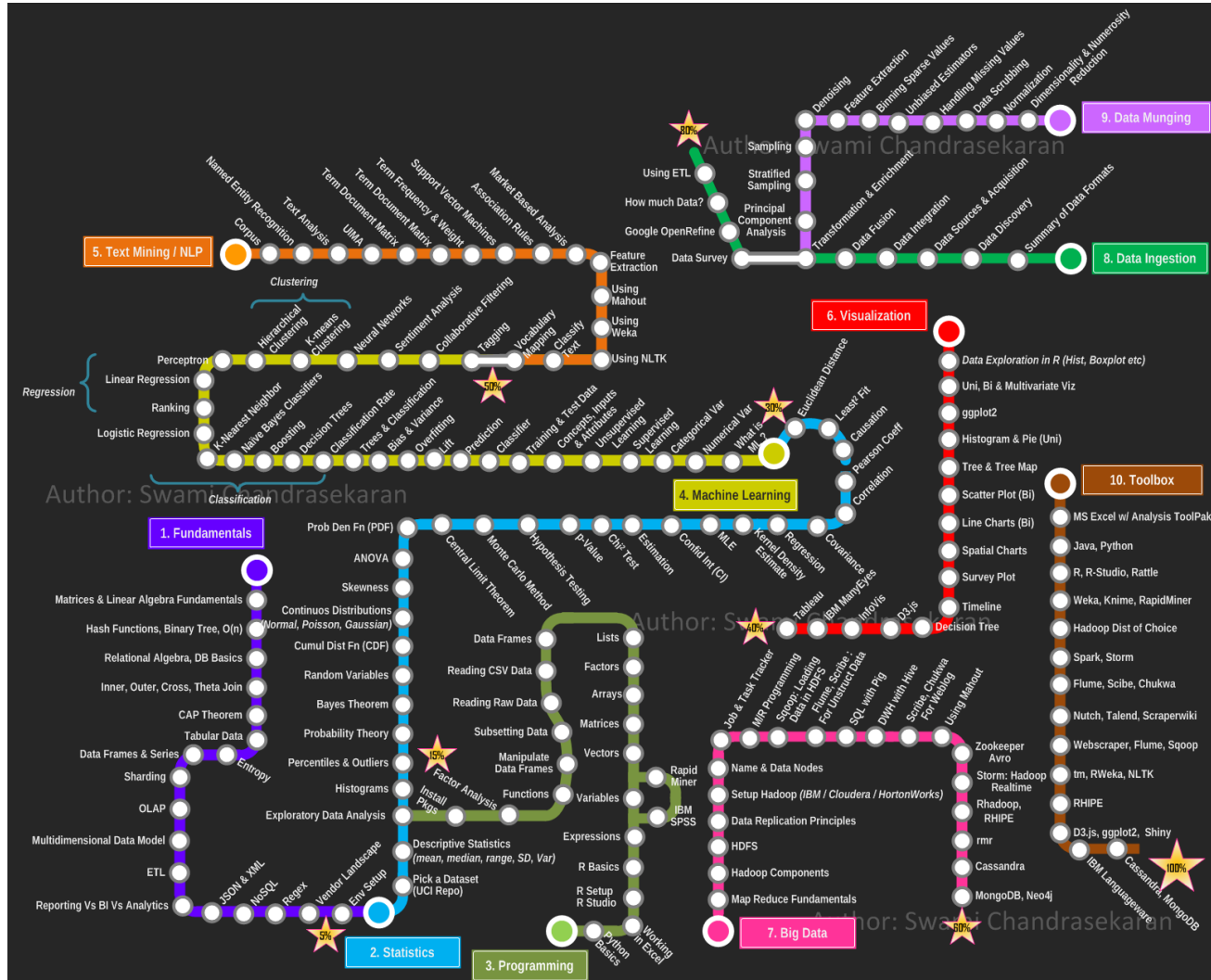
Engineer

r

Scientist

t





Competências

- ✓ Python
- ✓ Reconhecimento de padrões
- ✓ Aprendizado de máquina
- Mineração de dados

- ✓ Ciência de dados
- ✓ Programação de computadores
- ✓ Programação pareada
- ✓ R (linguagem de programação)

Requisitos

- Experiência na manipulação de Big Data;
- Experiência com banco de dados para modelagem preditiva;
- Habilidade em Programação nas linguagens: SQL, Python, PLSQL, R, bancos de dados não relacionais;
- Vivência com algoritmos para aprendizado de máquina (tanto supervisionado quanto não-supervisionado)
- Experiência em Deep Learning e demais técnicas de inteligência artificial/computacional;
- Sólidos conhecimentos em estatística avançada, modelagem espacial e temporal e estatística Bayesiana.
- Amplos conhecimentos em Inferência Estatísticas, distribuições de probabilidade, testes de hipótese, análise de variância, modelos de regressão linear e não linear
- Conhecimentos consolidados de modelos de otimização, a partir de algoritmos evolutivos; e modelos de advanced analytics com foco na predição de eventos;
- Graduação concluída em Ciência da Computação, Tecnologia da Informação, Matemática, Estatística, Engenharia, ou áreas correlatas.
- Disponibilidade para atuar em São Luiz, MA.

Competências

- ✓ Inglês
- ✓ Python
- ✓ Computação em nuvem
- ✓ Ferramentas de colaboração
- ✓ Capacidade analítica

- ✓ Análise de dados
- ✓ Empreendedorismo
- ✓ Programação de computadores
- ✓ R (linguagem de programação)
- ✓ Microsoft Office

Competências

- ✓ Estatística
- ✓ SQL
- ✓ Aprendizado de máquina
- ✓ Análises
- ✓ Negócios

- ✓ Inteligência
- ✓ SAS
- ✓ Liderança
- ✓ Marketing
- ✓ R (linguagem de programação)

Data science
Business Intelligence
Big Data

CRESCEMOS E GANHAMOS RELEVÂNCIA

Venda Site Share Site Crescimento



magalu

Unicórnios brasileiros
Centro de dados como estratégia

fklearn



Unicórnios brasileiros
Centro de dados como estratégia



Unicórnios brasileiros
Centro de dados como estratégia

A word cloud of Portuguese terms related to digital security and fraud prevention. The terms are arranged in a cluster, with varying font sizes and colors (dark green and light green). The most prominent terms are 'Aprovação de Crédito' and 'Segmentação de Clientes'. Other visible terms include 'Marketing um-para-um', 'Detecção de Fraude', 'Lançamento de produtos', 'Análise de Risco', 'Reconhecimento facial', 'Prova de vida', 'Reconhecimento de assinaturas', 'Análise comportamental', and 'Melhoria de processos'.

Aprovação de
Crédito
Segmentação de Clientes
Marketing um-para-um
Detecção de Fraude
Lançamento de produtos
Análise de Risco
Reconhecimento facial
Prova de vida
Reconhecimento de assinaturas
Análise comportamental
Melhoria de processos

Expectativas...

- Entender o que é Big Data e quais as principais soluções para lidar com grandes volumes de dados hoje.
- Entender qual o papel de um cientista de dados e compreender os principais fundamentos desta profissão
- Entender a manipulação básica de dados para uso em data analytics.
- Entender fundamentos do aprendizado de máquina e validação dos modelos.
- Usar esse entendimento para liderar ou implementar projetos de Big Data Analytics.

-
- Se tornar expert:
 - nas ferramentas utilizadas na disciplina.
 - em aprendizado de máquina.
 - em soluções big data.

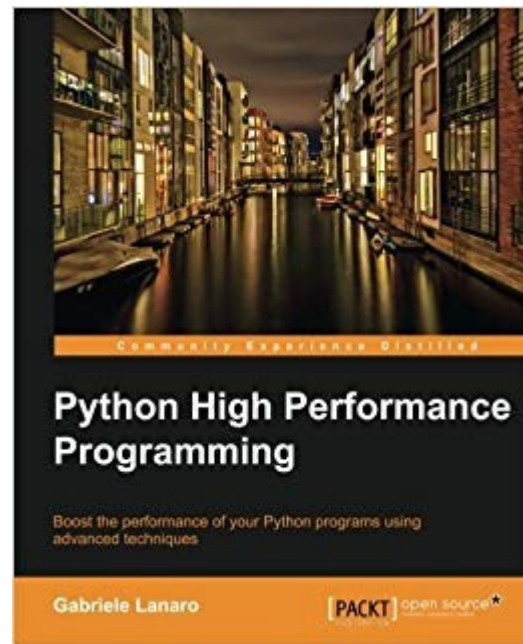
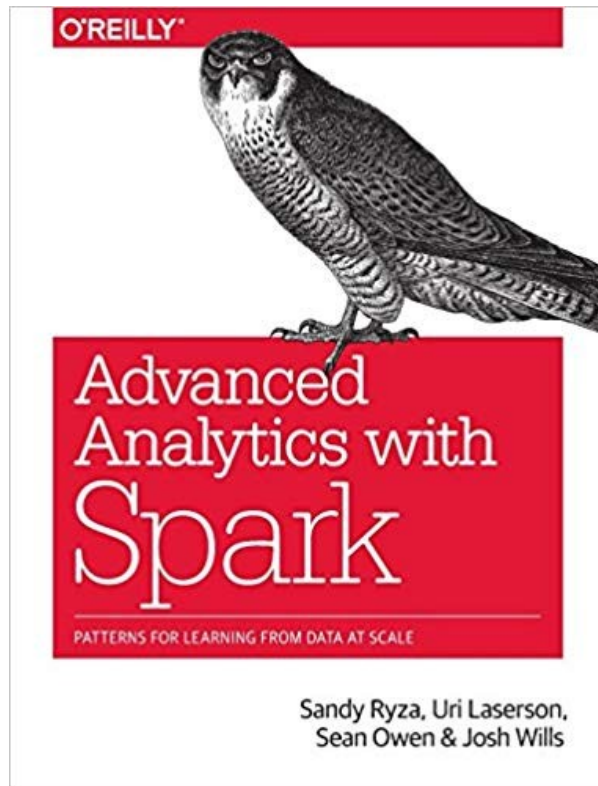
Processamento e análise para Big Data

Ementa: Conceitos, tecnologias. Mineração de dados. Machine learning. Tomada de decisões. Data Analytics. Ferramentas. Desafios na utilização de Big Data na indústria avançada: cenário, desafio, benefícios.

Conteúdo programático:

Módulo I - 22 a 24/07 [Fundamentos]	Módulo II - 24 e 27/07 [Integração e ML]	Módulo II [Indexação e Projeto]
Apresentação da disciplina	ML em Big Data	Indexação de bases
Fundamentos de computação paralela e distribuída	Integração de dados em Big Data	Orientação projeto
Ferramentas de Big Data	Avaliação prática – Predição em base aberta	Projeto – Integração de Big Data

Principais referências



Dúvidas?

