

Bancos de Dados Relacionais e Não - Relacionais

Modelo Relacional. SQL. Bancos de Dados NoSQL: definição; motivação; modelo de Transações. Modelos NoSQL. Propriedades Modelo Relacional x Propriedades Modelos NoSQL. Principais SGBD's.

Cultura e Práticas Dataops e Mlops

A cultura DevOps. Integração contínua e entrega contínua. Estratégias de deploy. Projeto de pipeline para build e deployment. Automação de testes. Infrastructure as Code (IaC). Ferramentas e infraestrutura do ambiente integrado DevOps: Containers, Docker, Kubernetes e OpenShift.

Estatística Geral: Teoria e Aplicações

Estatística descritiva. Probabilidade e distribuições de probabilidade. Inferência: estimação pontual e intervalar e testes de hipóteses. Utilização de software para análises estatísticas e análise de casos aplicados à gestão.

Integração e Processamento de Fluxo Contínuo de Dados

Conceitos. Estrutura de dados ETL. Data Lake. Data Streaming e dados em tempo real. Conceitos de eventos e sua topologia. Identificação e processamento de eventos complexos. Sistemas de gestão de fluxo de dados. Projeto e desenvolvimento de aplicação ETL. Plano de Teste. Principais ferramentas e tecnologias.

Introdução à Linguagem Python

Introdução à computação. Tipos de dados. Estruturas de controle: condicional e repetição. Estruturas de dados: listas, tuplas, conjuntos e dicionários. Funções.

Introdução à Linguagem R

Introdução ao R. Estruturas de Dados. Estruturas de Controle. Funções. Visualização de Dados em R. Manipulação de Dados.

Machine Learning

Processo de aprendizagem de máquina. Tipos de dados. Técnicas e algoritmos de aprendizado supervisionado e não-supervisionado. Combinação de modelos. Métricas e avaliação de modelos.

Modelagem e Preparação de Dados para Machine Learning

Modelagem conceitual de domínio. Montagem do conjunto de dados. Tipos de dados. Melhoramento e enriquecimento. Preparação: eliminação de dados irrelevantes, granulação e agregação, consistência, poluição, análise de domínios de atributos, integridade, concordância, duplicação e redundância. Análise e detecção de outliers. Dados ausentes: mecanismos e recuperação. Discretização de variáveis. Fusão de variáveis. Medidas de similaridade e de-similaridade. Feature Engineering. Teste em dados.

Modelos Estatísticos

Modelos Preditivos e tipos de análise. Abordagens para análise preditiva. Séries temporais. Regressão Linear simples e múltipla. Regressão logística.

Processamento de Linguagem Natural

Algoritmos e técnicas de processamento em linguagem natural. Expressões regulares. Medidas de similaridade textual. Parsing, tokenização, lematização, stemming. Marcação textual. Reconhecimento de entidades nomeadas. Extração de



Especialização em Ciência de Dados e Big Data

OFERTA 7, TURMA 1, MODALIDADE Assíncrono

EMENTÁRIO DAS DISCIPLINAS

informação. Arquitetura de aplicação para processamento de Linguagem Natural. Análise de sentimento.

Processamento Paralelo e Distribuído de Dados

Princípios de processamento e de volumes de dados massivos. Conceitos básicos de sistemas distribuídos. Modelo de Computação MapReduce: definição e motivação. Ambientes de processamento: Hadoop, Spark. Frameworks sobre Hadoop: Spark, Hive e Impala. Aplicações com ML Lib.

Programação para Ciência de Dados

Fundamentos da Linguagem Python. Vetorização. Bibliotecas para manipulação dados. Bibliotecas para visualização de dados.

Recuperação da Informação na Web e Redes Sociais

Ferramentas de análise, monitorização e benchmark. Algoritmos e soluções para problemas de busca e extração de informação da Web. Algoritmos e soluções para a análise de redes sociais online e em sites de conteúdo. Web crawling.

Redes Neurais e Deep Learning

Introdução a redes neurais artificiais. Deep Learning. Técnicas, métodos e arquiteturas para redes neurais. Redes convolucionais modernas. Aplicações de redes convolucionais e deep learning.

Visualização de Dados

Fundamentos de visualização de Dados. Uso de cores. Tipos de Dados. Técnicas de visualização para dados categóricos, séries temporais, distribuição, correlação, dados multivariados, dados geoespaciais, texto, grafos/redes, dados em alta dimensão. Dashboards. Data storytelling.