

Pós-Graduação a distância

Ciência de Dados e Big Data

Disciplinas:

- 1) Banco de Dados Relacionais e Não Relacionais
- 2) Linguagens de Programação para Ciência de Dados
- 3) Visualização de Dados
- 4) Data Discovery e Analytics
- 5) Integração e Processamento de Fluxo Contínuo de Dados
- 6) Processamento Paralelo e Distribuído de Dados
- 7) Estatística Geral - Teoria e aplicações
- 8) Recuperação da Informação na Web e Redes Sociais
- 9) Modelos Estatísticos
- 10) Machine Learning
- 11) Redes Neurais e Aprendizagem Profunda
- 12) Dataops e Implantação de Sistemas de Machine Learning

Ementas:

Disciplina 01 – BANCO 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.

Disciplina 02 – LINGUAGENS DE PROGRAMAÇÃO PARA CIÊNCIA DE DADOS

Aspectos básicos da programação em linguagem Python e R. Programação de aplicações.

Disciplina 03 – VISUALIZAÇÃO DE DADOS

Fundamentos de visualização de dados. Tipos de dados e técnicas de visualização: dados categóricos, séries temporais, dados multivariados, Geoanálises. Dashboards. Data Storytelling. Projetos em design de informação.

Disciplina 04 – DATA DISCOVERY & ANALYTICS

Fundamentos da descoberta de dados. Projeto e construção de aplicações OLAP. Ferramentas de Data Discovery e Self-Service Analytics.

Disciplina 05 – INTEGRAÇÃO E PROCESSAMENTO DE FLUXO CONTÍNUO DE DADOS

Conceitos. Estrutura de dados ETL. 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.

Disciplina 06 – 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: PIG, Hive, Impala e outras soluções. Aplicações.

Disciplina 07 – 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.

Disciplina 08 – 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.

Disciplina 9 – MODELOS ESTATÍSTICOS

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

Disciplina 10 – MACHINE LEARNING

Processo de aprendizagem de máquina. Tipos de dados. Técnicas e algoritmos de aprendizado supervisionado e não-supervisionado. Combinação de modelos: boosting e ensemble. Avaliação de modelos.

Disciplina 11 – REDES NEURAIIS E APRENDIZAGEM PROFUNDA

Taxonomia de redes neurais. Redes neurais recorrentes, convolutivas e redes de memória dinâmica. Aplicações e modelagem de problemas.

Disciplina 12 – DATAOPS E IMPLANTAÇÃO DE SISTEMAS DE MACHINE LEARNING

Provisionamento de ambientes de desenvolvimento, produção, integração e testes. Amazon SageMaker. Azure machine learning. Docker. API. Fluxo de retreinamento em produção.