

Pós-Graduação a distância

Ciência de Dados e Big Data

Disciplinas:

- 1) Ciência de Dados e Big Data em Negócios
- 2) Linguagens de Programação para Ciência de Dados
- 3) Banco de dados Não Relacionais
- 4) Estatística Geral - Teoria e aplicações
- 5) Técnicas Estatísticas de Predição: Teoria e Aplicações
- 6) Soluções para Processamento Paralelo e Distribuído de Dados
- 7) Recuperação da informação na Web e em Redes Sociais
- 8) Tecnologias para o Ecossistema de Big Data
- 9) Integração e Processamento de fluxo Contínuo de Dados
- 10) Machine Learning
- 11) Data Discovery, OLAP e Visualização de Dados
- 12) Arquitetura e Qualidade de Dados

Ementas:

Disciplina 01 – CIÊNCIA DE DADOS E BIG DATA EM NEGÓCIOS

Importância da informação no negócio. Necessidades em decisões de negócio. Conceitos de Big Data. Big Data em relação a outras disciplinas. Ciência dos dados. Ciclo de vida do processo de ciência de dados. Papéis dos envolvidos em projetos de Ciência de dados e Big Data. Computação em nuvens. Arquitetura de Big Data. Modelos de entrega e distribuição de serviços de Big Data. Principais plataformas de *Cloud Computing* para Big Data.

Disciplina 02 – LINGUAGENS DE PROGRAMAÇÃO PARA CIÊNCIA DE DADOS

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

Disciplina 03 – BANCO DE DADOS NÃO RELACIONAIS

Bancos de Dados NoSQL: definição; motivação; modelo de Transações. Modelos Nosql. Propriedades Modelo Relacional x Propriedades Modelos Nosql. Principais SGBD's. Soluções para Big Data.

Disciplina 04 – 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 05 – TÉCNICAS ESTATÍSTICAS DE PREDIÇÃO: TEORIA E APLICAÇÕES

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

Disciplina 06 – SOLUÇÕES PARA 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. Hadoop. Spark. Outros ambientes de processamento. Aplicações.

Disciplina 07 – RECUPERAÇÃO DA INFORMAÇÃO NA WEB E EM REDES SOCIAIS

Conceito de inteligência. Conceito de inteligência coletiva. Conceito de crowdsourcing. Ferramentas de análise, monitorização e benchmark. Web mining. 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. Text Mining.

Disciplina 08 – TECNOLOGIAS PARA O ECOSISTMA DE BIG DATA

Frameworks sobre Hadoop: PIG, Hive, Impala e outras soluções. Introdução linguagem SQL. Programação de aplicações. Conexão de clientes.

Disciplina 09 – INTEGRAÇÃO E PROCESSAMENTO DE FLUXO CONTÍNUO DE DADOS

Conceitos. 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. Estrutura de dados ETL. Projeto e desenvolvimento de aplicação ETL. Plano de Teste. Operação. Principais ferramentas e tecnologias.

Disciplina 10 – MACHINE LEARNING

Metodologia para descoberta de conhecimento em banco de dados. Exploração do espaço problema e espaço solução. Técnicas de aprendizado supervisionado e não-supervisionado. Regras de associação, agrupamento (clustering) e classificação. Rede neural, Agrupamento com K Means. Classificador Naïve Bayesian. Árvore de decisão. Outros algoritmos.

Disciplina 11 – DATA DISCOVERY, OLAP E VISUALIZAÇÃO DE DADOS

Fundamentos e requisitos de aplicações de suporte a decisão. Projeto, construção e tecnologias de aplicações OLAP. Dashboards. Data Storytelling. Fundamentos da descoberta de dados. Projetos em design de informação. Métodos e técnicas de visualização de dados. Self Service BI. Geoanálises.

Disciplina 12 – ARQUITETURA E QUALIDADE DE DADOS

Arquitetura de dados. Arquitetura de Dados no DMBOK e Open Group Togaf. A relação do papéis: AD, DBA, Arquiteto Corporativo e Arquiteto de Dados. Conceitos e motivações para governança de dados. Maturidade em governança de dados. Conceitos de qualidade de dados. Atividades e técnicas para qualidade de dados. Avaliação da qualidade de dados. Metadados. Master Data Management.