

Plano de Ensino

■ Código e nome da disciplina

DGT0134 TÓPICOS DE BIG DATA EM PYTHON

2 Carga horária semestral 🕾

3 Carga horária semanal ∑

4 Perfil docente 🤬

O docente deve ser graduado em áreas afins de Tecnologia da Informação e possuir Pós-Graduação Lato Sensu (especialização), embora seja desejável a Pós-Graduação Stricto Sensu (Mestrado e/ou Doutorado) na área do curso ou áreas afins.

É desejável que o docente possua experiência de desenvolvimento com linguagem Python e Big Data, além de conhecimentos teóricos e práticos, habilidades de comunicação em ambiente acadêmico, capacidade de interação e fluência digital para utilizar as ferramentas necessárias ao desenvolvimento do processo de ensino-aprendizagem (SGC, SAVA, BdQ e SIA). Importante, também, o conhecimento do Projeto Pedagógico dos Cursos que a disciplina faz parte na Matriz Curricular.

É necessário que o docente domine as metodologias ativas inerentes à educação por competências e ferramentas digitais que tornam a sala de aula mais interativa. A articulação entre teoria e prática deve ser o eixo direcionador das estratégias em sala de aula. Além disso, é imprescindível que o docente estimule o autoconhecimento e autoaprendizagem entre seus alunos.

5 Ementa 📳

Princípios de Big Data. Hadoop e Armazenamento de Dados. Princípios de Desenvolvimento com PySpark. Análise de Dados em Python com Pandas. Big Data Analytics.

6 Objetivos

- Criticar os processos de trabalho e obter interpretações rápidas e valiosas sobre tendências de mercado, baseando-se nos princípios de Big Data, para planejar o processamento de uma quantidade enorme de informações fazendo isso com alto desempenho e disponibilidade.
- Esquematizar o armazenamento e processamento distribuído de grandes conjuntos de dados, utilizando a plataforma de código aberto Hadoop, para criar clusters de computadores com hardware commodity obtendo armazenamento, processamento, acesso, governança, segurança e operações de dados.

- Construir aplicações de computação paralela, tendo como ferramenta o PySpark, para criar aplicações com pipeline de aprendizado de máquina realizando previsões nos dados.
- Analisar grandes volumes de dados, utilizando a biblioteca Pandas, para propor a utilização de técnicas avançadas de análise de dados.
- Esquematizar modelo de analises de dados estruturados e semiestruturados, baseando-se em técnicas de descoberta do conhecimento, para construir algoritmos de inteligência artificial.

7 Procedimentos de ensino-aprendizagem

Aulas interativas em ambiente virtual de aprendizagem, didaticamente planejadas para o desenvolvimento de competências, tornando o processo de aprendizado mais significativo para os alunos. Na sala de aula virtual, a metodologia de ensino contempla diversas estratégias capazes de alcançar os objetivos da disciplina.

Os temas das aulas são discutidos e apresentados em diversos formatos como leitura de textos, vídeos, hipertextos, links orientados para pesquisa, estudos de caso, podcasts, atividades animadas de aplicação do conhecimento, simuladores virtuais, quiz interativo, simulados, biblioteca virtual e Explore + para que o aluno possa explorar conteúdos complementares e aprofundar seu conhecimento sobre as temáticas propostas.

8 Temas de aprendizagem

- 1. PRINCÍPIOS DE BIG DATA
- 1.1 INTRODUÇÃO E APLICAÇÕES AO BIG DATA
- 1.2 CONCEITOS DE IOT E COMPUTAÇÃO DISTRIBUÍDA
- 1.3 PLATAFORMAS EM NUVEM PARA APLICAÇÕES DE BIGDATA
- 1.4 PROCESSAMENTO E STREAMING DE DADOS
- 2. HADOOP E ARMAZENAMENTO DE DADOS
- 2.1 INTRODUÇÃO E ARQUITETURA AO HADOOP
- 2.2 ECOSSISTEMA E SOLUÇÕES COM HADOOP
- 2.3 HDFS VS RDBMS
- 2.4 ENTENDO UM DATA LAKE
- 3. PRINCIPIOS DE DESENVOLVIMENTO DE SPARK COM PYTHON
- 3.1 INTRODUÇÃO AO SPARK
- 3.2 UTILIZANDO PYSPARK
- 3.3 OPERAÇÕES DE MAPREDUCE COM PYSPARK
- 3.4 TRANSFORMAÇÕES COM PYSPARK
- 4. ANÁLISE DE DADOS EM PYTHON COM PANDAS
- 4.1 COMPONENTES E SINTAXE DO PANDAS
- 4.2 PREPARAÇÃO DE DADOS COM PANDAS
- 4.3 MANIPULAÇÃO DE DADOS COM PANDAS
- 4.4 VISUALIZAÇÃO DE DADOS COM PANDAS
- 5. BIG DATA ANALYTICS (ATIVIDADE PRÁTICA SUPERVISIONADA)
- 5.1 COMPREENDENDO À DESCOBERTA DO CONHECIMENTO KDD
- 5.2 DESMISTIFICANDO INTELIGÊNCIA ARTIFICAL
- 5.3 APLICAÇÕES DE APRENDIZADO DE MÁQUINA COM TENSOR-FLOW

9 Procedimentos de avaliação

Os procedimentos de avaliação contemplarão as competências desenvolvidas durante a disciplina por meio de provas presenciais, denominadas AV e AVS, sendo a cada uma delas atribuído o grau de 0,0 (zero) a 10 (dez) no formato PNI - Prova Nacional Integrada.

Caso o aluno não atinja o resultado desejado na prova de AV, ele poderá recuperar sua nota na prova de AVS. Será composta por uma prova no formato PNI - Prova Nacional Integrada, com total de 10 pontos, e substituirá a nota da AV, caso seja maior.

Para aprovação na disciplina, o aluno deverá, ainda:

- atingir nota igual ou superior a 6 (seis) na prova de AV ou AVS;
- frequentar, no mínimo, 75% das aulas ministradas.

10 Bibliografia básica 📺

FACELI, Katti. Inteligência Artificial - Uma abordagem de aprendizado de máquina [BV:MB]. 2ª Ed. Rio de Janeiro: LTC, 2021.

Disponível em: https://integrada.minhabiblioteca.com.br/books/9788521637509/

GOLDSCHMIDT, Ronaldo. Data Mining - Conceitos, técnicas, algoritmos, orientações e aplicações [BV:MB]. Rio de Janeiro: Elsevier, 2015.

Disponível em: https://integrada.minhabiblioteca.com.br/books/9788595156395/

PEREIRA, Mariana Araújo. **Framework de Big Data [BV:MB]**. Porto Alegre: SAGAH, 2019. Disponível em: https://integrada.minhabiblioteca.com.br/books/9786556900803/

11 Bibliografia complementar 📦

BANIN, Sérgio Luiz. **Python 3 - Conceitos e Aplicações - Uma Abordagem Didática [BV:MB]**. São Paulo: Érica, 2018.

Disponível em: https://integrada.minhabiblioteca.com.br/books/9788536530253/

Batista, M.D. C. [BV:MB]. Data Mining. Porto Alegre: SAGAH, 2021.

Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786556900292/

MORAIS, Izabelly. **Introdução a Big Data e Internet das Coisas (IOT) [BV:MB]**. Porto Alegre: SAGAH, 2018.

Disponível em: https://integrada.minhabiblioteca.com.br/books/9788595027640/

MUELLER, John Paul. Aprendizado Profundo para Leigos [BV:MB]. Rio de Janeiro: Alta Books, 2020.

Disponível em: https://integrada.minhabiblioteca.com.br/books/9788550816982/

PERKOVIC, Ljubomir. Introdução à Computação Usando Python - Um Foco no Desenvolvimento

de Aplicações [BV:MB]. 1ª Ed. Rio de Janeiro: LTC, 2016. Disponível em: https://integrada.minhabiblioteca.com.br/books/9788521630937/