

**DEPARTAMENTO DE ENSINO
COORDENAÇÃO DO CURSO
PROGRAMA DE UNIDADE DIDÁTICA – PUD**

DISCIPLINA: CIÊNCIA DE DADOS	
Código: OPT02	
Carga Horária Total: 80h	CH Teórica: 40h CH Prática: 40h CH PPS: 0h CH Extensão: 0h
Número de Créditos: 4	
Pré-requisitos: -	
Semestre: Optativa	
Nível: Superior Tecnológico	
EMENTA	
Introdução a ciência dos dados: Business Intelligence, Business Analytics, Big data e ciência dos dados; Capacidades exigidas de um cientista de dados; Ecossistema de ciência de dados; Ética e privacidade no contexto de big data e ciência dos dados. Coleta e pré-processamentos de dados: Análise exploratória de dados: Conceitos de análise exploratória; Introdução à mineração de dados; Aprendizado supervisionado/não-supervisionado.	
OBJETIVO	
Objetivo geral Propiciar ao aluno o entendimento das tecnologias de ciência de dados para coleta, armazenamento, processamento, modelagem, visualização e análise de dados estruturados e não estruturados. Objetivos específicos - Entender os conceitos e práticas da ciência de dados; - Abordar a Gestão de Projetos em ciências de dados; - Conhecer e implementar técnicas de análise de dados avançadas; - Conhecer e implementar práticas de automação utilizando o desenvolvimento ágil na ciência de dados.	
PROGRAMA	
CIÊNCIA DE DADOS • Business Intelligence, Business Analytics, Big data e ciência dos dados; • Capacidades exigidas de um cientista de dados; • Ecossistema de ciência de dados; • Ética e privacidade no contexto de big data e ciência dos dados. COLETA E PRÉ-PROCESSAMENTO DE DADOS • Coleta de dados em tempo real (online) – data scraping (Web, APIs, tipos e formatos de dados); • Pré-processamento (limpeza, normalização, seleção de atributos e amostras); • Sistema de arquivos distribuído e MapReduce; • Apache Hadoop; • Apache Spark.	

ANÁLISE EXPLORATÓRIA DE DADOS

- Conceitos de análise exploratória;
- Estatísticas descritivas (revisão de conceitos e importação em Python);
- Visualização de dados (tipos de gráficos e dados, ferramentas para construção de gráficos estáticos, interativos e dinâmicos).

INTRODUÇÃO À MINERAÇÃO DE DADOS

- Aprendizado supervisionado/não-supervisionado;
- Aprendizado estatístico (regressão linear/logística);
- Classificação (árvores de decisão = naiveBayes, k-nearest);
- Agrupamentos (k-means, hierárquico);
- Mineração de padrões frequentes (regras de associação).

METODOLOGIA DE ENSINO

Aulas teóricas:

- As estratégias de aprendizado deverão priorizar o estudo por meio de aulas expositivas, apresentando os conteúdos necessários para o entendimento dos fundamentos de Ciência de Dados. Através de prática em laboratório, os conteúdos teóricos serão trabalhados, priorizando a contextualização desses em situações problema, e utilizando softwares de apoio aos modelos e análises em Ciência de Dados. Por meio de trabalhos em grupo, serão adotadas estratégias de aprendizagem colaborativa a fim de possibilitar troca de ideias e colaboração mútua, além da prática de soft skills como comunicação, proatividade e liderança. Aulas expositivas serão utilizadas com o intuito de introduzir assuntos e complementar processos de aprendizagem.

Aulas práticas:

- Ministradas em laboratório de informática, ou outro ambiente que facilite a consolidação dos conceitos fundamentais, por meio do uso e melhoramento de suas habilidades de trabalho ativo, onde a ênfase está na reflexão sobre o que se faz, provocando o encontro de significados no que for visto na aula teórica.
- Como recursos de apoio para o desenvolvimento de insights a partir de dados, tem-se a utilização de linguagens de programação mais específicas para esta área, tais como: Python, SQL e R. Algumas ferramentas/plataformas de apoio também são necessárias para o acesso, manipulação, armazenamento, análise e apresentação de dados. Com base no curso, uso da aprendizagem baseada em resolução de problemas com foco na interdisciplinaridade.

AValiação

A avaliação se dará de forma contínua, considerando aspectos qualitativos e quantitativos, em conformidade com as diretrizes estabelecidas no Regulamento da Organização Didática (ROD). Considerando seu caráter formativo, os instrumentos de avaliação deverão prever retorno aos estudantes sobre seus progressos e orientações para sanar dificuldades. Os instrumentos de avaliação serão diversificados e deverão não somente avaliar habilidade técnicas, mas também estimular o discente no desenvolvimento e aprimoramento de soft skills. Os instrumentos de avaliação a serem adotados serão realização de trabalhos práticos, individuais e em grupo, realização de seminários, relatórios de prática e visitas técnicas, dentre outros. A escolha dos instrumentos deve considerar o perfil da turma e ser um facilitador do processo de ensino e aprendizagem. Durante toda a continuidade do processo ensino-aprendizagem, sugere-se a promoção, em alta frequência, de avaliações formativas capazes de proporcionar ao docente um feedback imediato de como estão as interferências pedagógicas em sala de aula, e permitindo ao aluno uma reflexão sobre ele mesmo, exigindo autoconhecimento e controle sobre a sua responsabilidade, frente aos conteúdos já vistos em aula, privilegiando a preocupação com a satisfação pessoal do aluno e juntando informações importantes para mudanças na metodologia e intervenções decisivas na construção de conhecimento dos discentes, inclusive com subsídios para propostas de atividades de recuperação paralela na(s) reunião(es) de colegiado de curso, coordenadoria de curso e demais setores ligados ao ensino.

BIBLIOGRAFIA BÁSICA

- [1] ABEDIN, J.; DAS, K. **Data manipulation with R**. 2nd ed. EUA: Packt Publishing, 2015.
- [2] BENGFORT, B.; KIM, J. **Análítica de dados com Hadoop: uma introdução para cientistas de dados**. São Paulo: Novatec, 2016.
- [3] CHEN, D. **Pandas for everyone: Python data analysis**. São Paulo: Pearson Education, 2017.

BIBLIOGRAFIA COMPLEMENTAR

- [1] MACHADO, Felipe Nery Rodrigues. **Big Data o futuro dos dados e aplicações**. São Paulo: Saraiva, 2018. Ebooks.
- [2] MCKINNEY, W. **Python Para Análise de Dados: Tratamento de Dados com Pandas, NumPy e IPython**. EUA: O'Reilly Media, 2017.
- [3] BEELEY, C. **Web application development with R using Shiny**. EUA: Packt Publishing, 2013.
- [4] BROUCKE, S. VANDEN. **Practical web scraping for data science: best practices and examples with Python**. California: Apress, 2018
- [5] WICKHAM, H. **Ggplot2: elegant graphics for data analysis**. EUA: Springer International Publishing, 2016.

Coordenador do Curso

Setor Pedagógico