**PLANO DE ENSINO**

**CURSO:** Ciência da Computação

**SÉRIE:** 6º semestre

**DISCIPLINA:** Ciência de Dados – Oferecida em EaD

**CARGA HORÁRIA SEMANAL:** 1,5 horas-aula

**CARGA HORÁRIA SEMESTRAL:** 30 horas-aula

**I – EMENTA**

Apresentar ao aluno uma visão geral sobre os principais conceitos, técnicas e ferramentas referentes a ciência de dados, e dos diversos paradigmas e algoritmos existentes de aprendizado de máquina e inferência indutiva para a aquisição automática de conhecimento a partir de conjuntos de dados.

**II – OBJETIVOS GERAIS**

Fornecer os principais conceitos, técnicas e ferramentas referentes a ciência de dados e big data, e assim prover teoria básica para que os alunos possam aplicar as novas técnicas e ferramentas estudadas em problemas reais frente a grande quantidade de dados gerados por diferentes fontes.

**III – OBJETIVOS ESPECÍFICOS**

Entender os processos de descoberta de conhecimento em bases de dados, mineração e preparação dos dados, preparação dos dados e pré-processamento, modelagem dos dados, planejamento e a análise dos resultados.

**IV – Competências**

Compreender os principais conceitos, técnicas e ferramentas referentes a ciência de dados, para utilização nos processos de mineração e preparação dos dados, preparação dos dados e pré-processamento, modelagem dos dados, planejamento e a análise dos resultados.

**V – CONTEÚDO PROGRAMÁTICO**

1. **Introdução e Visão Geral sobre Ciência de Dados.**
2. Big Data e Ciência de dados – além do hype
3. Atual cenário Data-driven
4. Problemas e Desafios
5. Soluções baseadas em dados
6. Conjuntos de habilidades do professional cientista de dados
7. Perspectivas de um projeto de ciência de dados
8. **Descoberta de Conhecimento em Bases de Dados**
9. Visão geral sobre KDD *(do inglês, Knowledge Discovery in Databases)*
10. Representação e extração de conhecimento
11. Fontes de dados

1. **Introdução e Visão Geral sobre Aprendizado de Máquina.**
2. O que é Machine Learning (ML)?
3. Modelagem
4. *Pipeline* da Aprendizagem do Modelo
5. *Overfiting e Underfitting*
6. Balanço entre Viés e Variância em modelos de ML

1. **Viés indutivo.**
2. Sistema de Aprendizado
3. Tipos de aprendizagem
4. Espaço de Hipóteses
5. Viés de busca – ajuste aos dados

1. **Aprendizado descritivo e preditivo.**
2. Aprendizado supervisionado:
3. Classificação
4. Regressão
5. Aprendizado não-supervisionado:
6. Agrupamento
7. Associação
8. Sumarização

1. **Mineração de Dados**
2. Visão geral
3. Modelos de ML:
4. Árvore de Decisão
5. Naive-Bayes
6. K-Vizinhos mais próximos (KNN)
7. K-Médias
8. **Preparação de Dados**
9. Principais fontes de dados
10. Coleta, limpeza e organização das informações
11. Métodos de Raspagem
12. Tabulação

1. **Pré-processamento de Dados**
2. Seleção de atributos
3. Engenharia de características
4. Normalização dos dados
5. Dados ausentes

1. **Modelos Preditivos**
2. Regressão linear simples
3. Ajuste com Mínimos Quadrados
4. Gradiente Descendente
5. Regressão Linear múltipla
6. Regressão Logística

 10. **Planejamento de experimentos**

1. Split de dados – treino, teste e validação
2. Validação cruzada
3. *Benchmarking*

1. **Análise de resultados experimentais**
2. Métricas
3. Classificação
4. Regressão
5. Seleção de modelos

1. **Aplicações avançadas de ML**
2. Visão computacional
3. Processamento de linguagem natural
4. Reconhecimento de fala
5. APIs de Inteligência Artificial

**VI - ESTRATÉGIAS DE TRABALHO**

A disciplina é ministrada por meio de tele aulas, metodologias ativas e diversificadas apoiadas no Plano de Ensino. O desenvolvimento dos conceitos e conteúdos ocorre com apoio de propostas de leituras das Unidades Instrucionais disponíveis, de livros e artigos científicos básicos e complementares, exercícios, discussões em fórum/chats, sugestões de filmes, vídeos e demais recursos audiovisuais. Com o objetivo de aprofundar e enriquecer o domínio dos conhecimentos e incentivar a pesquisa, o docente pode propor trabalhos individuais ou em grupo, palestras, atividades complementares e práticas em diferentes cenários, que permitam aos alunos assimilarem os conhecimentos essenciais para sua formação. Há acesso a monitores no caso do aluno necessitar de suporte para seus estudos.

**VII – AVALIAÇÃO**

O processo de avaliação ocorre no laboratório de informática, através de agendamento para a realização das provas, e consiste de duas notas bimestrais. No primeiro bimestre o ambiente AVA disponibilizará uma lista de exercícios online cuja correção gerará a nota NP1. No segundo bimestre o ambiente AVA gerará uma prova impressa cuja correção (através de cartão de respostas) gerará a nota NP2. As provas Substitutiva e Exame também serão impressas e geradas pelo ambiente AVA.

**VIII – BIBLIOGRAFIA**

**Básica**

FACELI, KATTI; LORENA, ANA CAROLINA; GAMA, JOÃO ; DE CARVALHO, A. C. P. L. F. Inteligência Artificial : uma abordagem de aprendizado de máquina. 2. ed. Rio de Janeiro: LTC, 2022.

<https://integrada.minhabiblioteca.com.br/books/9788521637509>. Acesso: 10 nov. 2022

GRUS, JOEL. Data Science do Zero: Noções fundamentais com Python. 2. ed. Rio de Janeiro: Alta Books, 2021.

<https://integrada.minhabiblioteca.com.br/#/books/9788550816463/>. Acesso em: 10 nov. 2022.

MITCHELL, T. M. (1997). Machine Learning. McGraw-Hill.

**Complementar**

ALPAYDIN, E. (2004). Introduction to Machine Learning. MIT Press.

FLACH, P. (2012). Machine Learning:The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press.

PROVOST, F.; FAWCETT, T. Data Science for Business: What you need to know about data mining and data-analytic thinking by O'Reilly Media, 2013.

HAN, J.; KAMBER, M.; PEI, J. Data Mining: Concepts and Techniques, Third Edition (The

MORGAN KAUFMANN Series in Data Management Systems). Morgan Kaufmann, 2011.

WITTEN, I.; FRANK, E. Third Edition (The Morgan Kaufmann Series in Data Management Systems). 2011.

TAN, P.-N.; STEINBACH, M.; KUMAR, T. Introduction to Data Mining. Addison Wesley, 2005.