# PLANO DE ENSINO

|  |
| --- |
| **DISCIPLINA:**  Machine Learning |

|  |
| --- |
| **CURSO:**  Curso de Especialização em Ciência de Dados e Big Data |

|  |
| --- |
| **PROF (A):**  Hugo Bastos de Paula |

|  |
| --- |
| **CARGA HORÁRIA TOTAL (sala de aula + trabalho orientado): 24 h-a** |

|  |
| --- |
| **TRABALHO ORIENTADO (descrição das atividades extracurriculares):**  Será desenvolvido um trabalho orientado durante o curso que demandará 8 horas. O trabalho orientado consistirá de um exercício de aplicação de uma técnica de aprendizado de máquina e será acompanhado pelo professor. |

|  |
| --- |
| **EMENTA (Projeto Pedagógico):**  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. |

|  |
| --- |
| **UNIDADES DE ENSINO (Conteúdo Programático):**    **UNIDADES DE ENSINO: HORAS-AULA**  **Unidade 1: Identificação de Padrões**  **04 h/a**  1.1 O processo de Ciência de Dados.  1.2 Tarefas de aprendizado de máquina.  1.3 Preparação, seleção, melhoria e enriquecimento de dados.  1.4 Tipos de dados  1.5 Estilos de aprendizado.  **Unidade 2: Aprendizado supervisionado**  **12 h/a**  2.1 Conceitos de classificação e predição.  2.2 Indução por árvore de decisão.  2.3 Naïve Bayes.  2.4 Redes Neurais Artificiais  2.5 Aplicação e avaliação de modelos.  2.6 Algoritmos de previsão: regressão linear.  2.7 Boosting: algoritmo Adaboost.  **Unidade 3: Aprendizado não supervisionado 08 h/a**  3.1 Estruturas de dados.  3.2 Medidas de similaridade e de distância.  3.3 Algoritmos de agrupamento. |

|  |
| --- |
| **OBJETIVOS/MÉTODOS DIDÁTICOS:**  Ao final da disciplina, espera-se que o aluno seja capaz de:   * Descrever convenientemente um problema, definir seu espaço de soluções e utilizar estratégias de aprendizado de máquina adequadas para resolvê-lo. * Diferenciar quando aplicar aprendizado supervisionado e não supervisionado na solução de problemas. * Avaliar a performance dos modelos aprendidos. * Estruturar, definir o conjunto de treinamento e de validação para construção e avaliação de modelos de aprendizado de máquina.   A disciplina está estruturada em dois momentos. No primeiro momento, ela possui uma estrutura mais conceitual, onde são apresentados os requisitos mínimos necessários para se compreender o processo de ciência de dados e *analytics*, além de compreender questões básicas sobre tipos de dados, exploração e preparação de dados, e o impactos de ruídos e inconsistências de dados na indução de modelos de aprendizado. No segundo momento, desenvolvido em laboratório, diversos algoritmos de aprendizado serão apresentados, com suas aplicabilidades, capacidades e limitações. Esses algoritmos serão agrupados por tarefas de aprendizado e estilo de aprendizado, conforme apresentado na seção inicial da disciplina. |

|  |
| --- |
| **DISTRIBUIÇÃO DE PONTOS/CRITÉRIOS DE AVALIAÇÃO:**  01 Exercícios/avaliações em sala 40 pontos  02 Trabalho Orientado 40 pontos  03 Avaliação presencial individual 20 pontos |

|  |
| --- |
| **REFERÊNCIAS BIBLIOGRÁFICAS (indicação de livros, artigos e periódicos de acordo com o assunto):**  FACELI, Katti et al. *Inteligência artificial: uma abordagem de aprendizado de máquina*. Rio de Janeiro, RJ: LTC, 2011. xvi, 378 p. ISBN 9788521618805.  MUELLER, Andreas; GUIDO, Sarah. *Introduction to Machine Learning with Python*. O’Reilly. 2016. ISBN: 978-1491917213  TAN, Pang-Ning, STEINBACH, Michael, KUMAR, Vipin. *Introdução ao Data Mining – Mineração de dados*. Ciência Moderna, 2012. ISBN 978-8573937619. |