

Campus: São José dos Campos		
Curso (s): Engenharia de Computação e Ciência da Computação		
Unidade Curricular (UC): Aprendizado de Máquina		
Unidade Curricular (UC): <i>Machine Learning</i>		
Código da UC:		
Docente Responsável: Márcio P. Basgalupp		Contato (e-mail): <i>basgalupp@unifesp.br</i>
Docente (s) Colaborador/a (es/as):		Contato (e-mail): [opcional]
Ano letivo: 2023	Termo:	Turma (s): Noturno
Nome do Grupo/Módulo/Eixo da UC (se houver):		Idioma predominante em que a UC será oferecida: <input checked="" type="checkbox"/> Português <input type="checkbox"/> English <input type="checkbox"/> Español <input type="checkbox"/> Français <input type="checkbox"/> Libras <input type="checkbox"/> Outro:
UC: <input type="checkbox"/> Fixa <input checked="" type="checkbox"/> Eletiva <input type="checkbox"/> Optativa	Oferecida como: <input checked="" type="checkbox"/> Disciplina <input type="checkbox"/> Módulo <input type="checkbox"/> Estágio <input type="checkbox"/> Outro:	Oferta da UC: <input type="checkbox"/> Semestral <input type="checkbox"/> Anual
Ambiente Virtual de Aprendizagem: <input type="checkbox"/> Moodle <input type="checkbox"/> Classroom <input type="checkbox"/> Outro: <input type="checkbox"/> Não se aplica		
Pré-Requisito (s) - Indicar Código e Nome (s) da (s) UC:		
Carga horária total (em horas): 72		
Carga horária teórica (em horas): 48	Carga horária prática (em horas): 24	Carga horária de extensão (em horas, se houver):
Se houver atividades de extensão, indicar código e nome do projeto ou programa vinculado na Pró-Reitoria de Extensão e Cultura (ProEC):		
Ementa: Análise e pré-processamento de dados. Aprendizado supervisionado. Aprendizado não-supervisionado. Avaliação de modelos preditivos e descritivos. Tópicos avançados em aprendizado de máquina e Mineração de Dados.		
Conteúdo programático: 1 Análise e Pré-Processamento de Dados 1.1 Análise de Dados 1.1.1 Caracterização de Dados 1.1.2 Exploração de Dados 1.1.3 Medidas de Distância 1.2 Pré-Processamento de Dados 1.2.1 Amostragem de Dados 1.2.2 Limpeza de Dados 1.2.3 Transformação de Dados 1.2.4 Redução de Dimensionalidade 2 Aprendizado Supervisionado 2.1 Paradigma Baseado em Distâncias 2.1.1 K-NN 2.1.2 Raciocínio Baseado em Casos 2.2 Paradigma Probabilístico 2.2.1 Aprendizado Bayesiano 2.2.2 Naïve Bayes		

- 2.2.3 Redes Bayesianas para Classificação
- 2.3 Paradigma Baseado em Procura
 - 2.3.1 Árvores de Decisão e Regressão
 - 2.3.2 Regras de Decisão
- 2.4 Paradigma Baseado em Otimização
 - 2.4.1 Regressão Linear
 - 2.4.2 Regressão Logística
 - 2.4.2 Redes Neurais Artificiais
 - 2.4.3 Máquinas de Vetores de Suporte
- 3 Aprendizado Não-Supervisionado
 - 3.1 Análise de Agrupamentos
 - 3.1.1 Agrupamento Hierárquico
 - 3.1.2 Agrupamento Particional
 - 3.2 Índices de Validação de Agrupamentos
 - 3.2.1 Índices Relativos
 - 3.2.2 Índices Internos
 - 3.2.3 Índices Externos
- 4 Tópicos Avançados de Aprendizado de Máquina
 - 4.1 Aprendizado em Fluxos Contínuos de Dados
 - 4.2 Meta-Aprendizado
 - 4.3 Classificação Hierárquica e/ou Multirrótulo
 - 4.4 Aprendizado baseado em Computação Natural
 - 4.4.1 Colônias de Formigas
 - 4.4.2 Enxames de Partículas
 - 4.4.3 Computação Evolutiva

Objetivos:

O cumprimento da disciplina busca dar ao aluno, ao final do semestre, condições de:

1. Compreender os passos de análise e pré-processamento de dados que se fazem necessários para posterior aplicação de algoritmos e técnicas de aprendizado de máquina.
2. Identificar quando uma determinada aplicação deve ser analisada sob a ótica do aprendizado supervisionado, semisupervisionado ou não-supervisionado.
3. Conhecer e ser capaz de implementar algoritmos clássicos dos principais tipos de paradigmas de aprendizado supervisionado (paradigmas probabilísticos, baseado em distâncias, baseado em procura e baseado em otimização).
4. Comparar e avaliar o desempenho de diferentes algoritmos quando aplicados a um determinado domínio de interesse.
5. Conhecer o estado da arte do aprendizado de máquina nos seus diversos paradigmas e conceituações.

Metodologia de ensino:

Exposição teórica, seguida de exercícios, seminários e trabalhos práticos com o uso de computador.

Avaliação:

Serão atribuídas notas aos trabalhos práticos e seminários, compondo duas notas: T1 e T2. Será também realizada uma prova escrita (P) em sala de aula. A nota final será calculada pela média simples de T1, T2 e P.

Bibliografia:

Básica:

1. FACELI, K., LORENA, A.C., GAMA, J., CARVALHO, A.C.P.L.F. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. Rio de Janeiro: LTC, 2011. 378 p.
2. TAN, P. N. STEINBACH, M. KUMAR, V. Introduction to Data Mining. Addison-Wesley, 2005. 769 p.
3. Mitchell, T. Machine Learning. McGraw-Hill, 1997. 432 p.

Complementar:

1. ALPAYDIN, E. Introduction to machine learning. 2nd edition, Cambridge (UK): The MIT Press, 2010. 537 p.
 2. BISHOP, C.M., Pattern Recognition and Machine Learning. Springer, 2006. 738 p.
 3. WITTEN, I.; FRANK, E. Data mining: practical machine learning tools and techniques with java implementations. San Francisco (CA): Morgan Kaufmann, 2000. 371 p.
 4. DUDA, R., HART, P., STORK, D. Pattern Classification. 2nd edition, Wiley, 2000. 680 p.
- MARSLAND, S. Machine Learning: An Algorithmic Perspective. Chapman & Hall/CRC, 2009. 406 p.