

Plano de Aula

Disciplina/módulo: Machine Learning aplicado a Bioinformática

Professor: Gilderlanio Santana de Araújo

Período: 16/04/2024 a 04/06/2024

Horários: Terças e Quintas – 14:00 às 17:00 horas

1. Conteúdo programático

Introdução a aprendizagem de máquina (machine learning) e bioinformática. Coleta de dados. Pré-processamento de dados (limpeza, normalização, imputação e visualização). Engenharia de recursos/características. Aprendizado Supervisionado. Aprendizado não-supervisionado. Redução de dimensionalidade. Aplicações em genômica, transcriptômica e proteômica.

2. Objetivos

Capacitar os alunos a compreenderem os princípios fundamentais de aprendizagem de máquina e sua aplicação em ciências “ômicas”, fornecendo-lhes habilidades práticas para analisar e interpretar dados biológicos. Ao final do curso, os alunos devem ser capazes de selecionar e implementar adequadamente algoritmos de aprendizagem de máquina e interpretar os resultados obtidos, bem como comunicar suas descobertas de forma clara e eficaz.

3. Avaliação

Etapa 1) - Coleta de dados e análise exploratória de dados (15%):

- Coletar dados biológicos relevantes de fontes confiáveis.
- Realizar uma análise exploratória inicial dos dados, identificando padrões, distribuições e possíveis relações entre as variáveis.

Etapa 2) - Pré-processamento (15%):

- Realizar o pré-processamento dos dados, incluindo limpeza, tratamento de valores ausentes, normalização e padronização, conforme necessário para preparar os dados para a modelagem.

Etapa 3) - Projeto final e apresentação de seminário (70%):

- Desenvolver um projeto final onde aplicarão técnicas de aprendizagem de máquina para resolver um problema específico em bioinformática relacionado ao seu tema de pesquisa (40%).
- Documentar todo o processo em um repositório GitHub pessoal, seguindo o roteiro disponibilizado (10%).
- Apresentar seus resultados em um seminário no final da disciplina (15 minutos de apresentação e 5 minutos de arguição) (20%).

4. Metodologia

A aula será conduzida de forma dialogada e expositiva, incorporando apresentações de slides e recursos visuais em vídeos, além da execução de tutoriais de bancos de dados de forma prática para fixação do conteúdo.

5. Recursos didáticos

O conteúdo será ministrado com a utilização de recursos instrucionais (quadro branco e pincel) e audiovisuais (apresentação de slides por data show).

6. Pré-requisitos

Programação em Python ou R. Bioestatística. Linux Shell.

7. Bibliografia

- James, Gareth, et al. An Introduction to statistical learning. With Application in Python. Springer nature. 2023
- Marsland, Stephen. *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC, 2011.
- Learning, M. *Tom Mitchell*; McGraw Hill: New York, NY, USA, 1997