

IMD3003 - Aprendizado de Máquina Não-Supervisionado

2025.2

Projeto da Unidade III

Objetivo Geral

Criar uma apresentação e elaborar um relatório que descreve a aplicação de técnicas de aprendizado não supervisionado para explorar, organizar e interpretar uma base de dados por meio de clusterização e/ou redução da dimensionalidade, aplicados a um problema real.

Objetivos Específicos

- Selecionar um conjunto de dados adequado para técnicas não supervisionadas;
- Realizar pré-processamento, normalização e limpeza dos dados;
- Aplicar métodos de clusterização (ex.: K-means, DBSCAN, Agglomerative) e redução da dimensionalidade (ex.: PCA, t-SNE, UMAP);
- Comparar resultados, visualizar estruturas encontradas e interpretar os padrões identificados.

Descrição do Projeto

O trabalho será realizado em grupos de **até cinco pessoas**.

A equipe deve escolher um conjunto de dados relevante para sua área ou interesse. A base deve ser previamente analisada e preparada: remoção de inconsistências, normalização e seleção dos atributos necessários.

O grupo deve aplicar pelo menos dois algoritmos diferentes de aprendizagem não supervisionada, variando parâmetros e discutindo seus efeitos. Os resultados devem incluir visualizações, como projeções 2D/3D e gráficos que evidenciem padrões, agrupamentos ou

separações. A interpretação deve destacar estruturas reveladas pelos métodos, diferenças entre algoritmos, limitações detectadas e possíveis melhorias.

Ferramentas livres: Python, scikit-learn, PyTorch, seaborn, matplotlib, etc.

A apresentação ocorrerá ao final da unidade, junto com a entrega do relatório e do código-fonte.

Entrega do Projeto

- **Relatório técnico:** Documento conciso apresentando o dataset, o pré-processamento, os algoritmos aplicados, as projeções/visualizações e a interpretação dos padrões encontrados.
- **Apresentação oral:** Apresentação de até 10 minutos mostrando as visualizações, os métodos utilizados e os principais insights.
- **Código-fonte:** Notebook ou repositório contendo todo o código usado para pré-processamento, projeções, clusterização e gráficos.