

Identificação Padrões na Experiência de Produtos Digitais – Parte II

Clustering e Técnicas de Classificação

Com base nos mesmos dados do projeto anterior, mantendo todas as correções já efetuadas (out-of-range, missing values, outliers, etc), agora o seu grupo deve aprofundar a análise utilizando algumas técnicas de ML não supervisionadas e supervisionadas.

Considere as seguintes tarefas:

- (i) Aplicar os algoritmos de clusters hierárquicos (considere: **single**, **complete** e **average linkage**, pelo menos) e também um algoritmo de cluster não hierárquico (**k-means**, por exemplo).
 - a. Apresentar e comparar os clusters obtidos (**dendrogramas**, **número de clusters resultantes**, etc)
 - b. Interpretar os clusters criados (**características médias**, **perfis**, **variáveis mais influentes**, etc)
 - c. Avaliar e justificar a técnica de cluster mais adequada para o seu data set
 - d. Criar uma nova coluna chamada “classe” no seu data set, onde cada observação receberá a classe correspondente ao cluster escolhido em (c)
- (ii) Considerando a nova coluna “classe” gerada em (i) como a sua output variable, aplicar os seguintes algoritmos de classificação: **Árvore de Decisão**, **Random Forest**, **Bagging** e **Boosting** (ex: **AdaBoost**, **Gradient Boosting** ou **XGBoost**). Para cada modelo:
 - a. Apresentar as métricas de desempenho (matriz de confusão, accuracy, precision, recall, etc)
 - b. Interpretar a importância das variáveis, quando aplicável, e discutir os resultados (justificar o melhor modelo)
- (iii) Considerando a nova coluna “classe” gerada em (i) como a sua output variable, aplicar um algoritmo de Support Vector Machine (SVM) para realizar a classificação.
 - a. Avaliar diferentes kernels (linear, RBF, polinomial, etc) quando relevante
 - b. Apresentar métricas e discutir os resultados
- (iv) Considerando a nova coluna “classe” gerada em (i) como a sua output variable, aplicar um modelo de redes neurais artificiais (ANN) para realizar a classificação – utilize uma framework para isto
 - a. Especificar a arquitetura utilizada (número de camadas, neurónios, funções de ativação, epochs de treino, etc)
 - b. Apresentar métricas e discutir os resultados

Critérios de avaliação:

- Clareza e organização do código

- Profundidade da análise no relatório, o que inclui a interpretação dos resultados e a sua relevância
- Apresentação oral

Entregáveis:

- Relatório (máx 15 páginas, excluindo referências e apêndice)
- Código:
 - Cluster:
 - numerogrupo_i_projeto2.ipynb e requirements_i.txt
 - Decision Trees:
 - numerogrupo_ii_projeto2.ipynb e requirements_ii.txt,
 - SVM
 - numerogrupo_iii_projeto2.ipynb e requirements_iii.txt,
 - NN
 - numerogrupo_iv_projeto2.ipynb e requirements_iv.txt
- Apresentação (20 min)

Deadline: 10/01/2026