

Relatório - Lista #11 de IA

Aluno: Breno Pires Santos

Matricula: 808238

1. Pré-processamento de Dados

Comecei analisando a base pra identificar valores ausentes, entender os tipos de variáveis e a distribuição dos dados. Transformei variáveis categóricas usando One-Hot Encoding e Label Encoding. Para lidar com os valores nulos, usei estratégias como média e moda, dependendo do caso. Também criei algumas variáveis novas, como "familiares a bordo" e o "título" (tipo Sr., Sra., etc.) que extraí do nome dos passageiros. As variáveis numéricas foram normalizadas com o StandardScaler.

2. Modelagem com Algoritmos de Classificação

Testei dois algoritmos de classificação: Random Forest e uma Rede Neural. Avaliei ambos com precisão, recall e F1-score. O modelo de Random Forest teve o melhor desempenho, mostrando uma boa capacidade de generalização. A rede neural teve resultados razoáveis, mas apresentou um pouco de overfitting em alguns testes.

3. Modelagem com Algoritmos de Agrupamento

Utilizei o algoritmo K-Means para tentar encontrar grupos de passageiros com perfis parecidos. Reduzi a dimensionalidade com PCA pra poder visualizar melhor os clusters, e consegui identificar padrões, como um grupo composto majoritariamente por mulheres da 1ª classe que sobreviveram. Essa etapa ajudou bastante a interpretar os dados de forma mais visual.

4. Extração de Regras de Associação

Apliquei o algoritmo Apriori para extrair regras de associação e encontrei padrões bem interessantes. Por exemplo: mulheres da 1ª classe tinham mais chances de sobreviver, passageiros que embarcaram com familiares também apresentaram maior taxa de sobrevivência, e homens sozinhos da 3ª classe tinham as menores chances. Essas regras ajudaram a confirmar o que os modelos anteriores já apontavam.

5. Conclusão

Os modelos supervisionados mostraram bons resultados na previsão de sobrevivência, principalmente o Random Forest. A clusterização revelou grupos bem distintos, e as regras de associação reforçaram padrões importantes.