

Lista #11

Curso: Ciência da Computação

Disciplina: Inteligência Artificial

Prof^a. Cristiane Neri Nobre

Data de entrega: 01/06

Valor: 5 pontos

Questão 01

A base de dados do Titanic contém informações sobre os passageiros do famoso navio, como sexo, idade, classe da passagem, se embarcaram sozinhos, entre outros atributos. A tarefa principal é prever a sobrevivência dos passageiros. Esta base está disponível em diversos repositórios, como o Kaggle (<https://www.kaggle.com/c/titanic/data>).

Utilizando a base de dados do Titanic, crie um pipeline completo de Inteligência Artificial em Python com as seguintes etapas:

Objetivos de aprendizagem desta atividade

- Implementar pipelines completos de processamento e modelagem com dados reais.
- Avaliar criticamente diferentes abordagens de IA (supervisionada, não supervisionada e associação).
- Interpretar resultados e comunicar descobertas de forma clara.

1. Pré-processamento de Dados

Realize as seguintes operações:

- Carregamento da base de dados.
- Análise exploratória inicial (ex. estatísticas descritivas, valores ausentes, etc.).
- Preenchimento de valores nulos e transformação de variáveis categóricas (por exemplo, com One-Hot Encoding ou Label Encoding).
- Criação de variáveis derivadas que possam ser úteis para a modelagem (por exemplo: “família a bordo”, “título da pessoa” etc.).

- Normalização ou padronização das variáveis numéricas, quando necessário.

2. Modelagem com Algoritmos de Classificação

Utilize ao menos **dois algoritmos de classificação supervisionada** para prever a sobrevivência dos passageiros. Sugestões:

- **Random Forest**
- **Árvore de decisão**
- **Naive Bayes**
- **Rede Neural**
- **etc**

Avalie os modelos com métricas apropriadas precisão, recall e F1-Score.

3. Modelagem com Algoritmos de Agrupamento

Utilize ao menos **um algoritmo de clusterização** para identificar agrupamentos de passageiros com perfis semelhantes. Sugestões:

- **K-Means**
- **DBSCAN**
- **SOM**

Visualize os clusters utilizando PCA ou t-SNE e interprete os principais padrões encontrados (ex: cluster com maioria de mulheres sobreviventes etc.).

4. Extração de Regras de Associação

Aplique **algoritmos de regras de associação** para identificar padrões interessantes no comportamento dos passageiros, como por exemplo:

- A combinação de “sexo = feminino” e “classe = 1ª” implica em alta probabilidade de sobrevivência?

Use o algoritmo **Apriori** (ou outro similar) com os parâmetros adequados e interprete **ao menos 3 regras** extraídas com suporte, confiança e lift.

5. Entrega

- O notebook deverá conter explicações curtas sobre cada etapa.
- Visualizações e comentários são bem-vindos para justificar as decisões tomadas.
- Inclua a conclusão final comparando os modelos e os insights obtidos.
- Um pequeno relatório contendo todas as explicações deve ser entregue. Ou seja, apenas o notebook não é suficiente.