

Exercícios de Modelos de Classificação, Regressão e Associação

Questão 5 (Intermediário) - Diagnóstico de Doenças Cardíacas

Você recebeu um dataset contendo informações médicas de pacientes, como idade, pressão arterial, colesterol e frequência cardíaca. Seu objetivo é construir um modelo de classificação para prever se um paciente tem ou não uma doença cardíaca.

Tarefas:

- Carregue um dataset de doenças cardíacas (exemplo: Heart Disease UCI do Kaggle).
- Realize um pré-processamento adequado, tratando valores ausentes e normalizando as variáveis.
- Utilize pelo menos dois modelos de classificação (exemplo: Random Forest e SVM) e compare os resultados.
- Avalie os modelos utilizando precisão, recall e curva ROC-AUC.

Pergunta: Qual modelo apresentou melhor desempenho? Quais variáveis mais impactaram na previsão?

Questão 6 (Avançado) - Previsão do Valor de Imóveis

Uma imobiliária deseja prever o valor de imóveis com base em características como localização, número de quartos, tamanho do terreno, entre outros.

Tarefas:

- Utilize um dataset de preços de imóveis (exemplo: California Housing do Scikit-Learn).
- Aplique técnicas de feature engineering para melhorar o desempenho do modelo.
- Teste diferentes algoritmos de regressão, como Regressão Linear, XGBoost e Redes Neurais Artificiais (ANNs).
- Avalie os modelos com métricas como RMSE e R^2 .

Pergunta: Qual modelo teve menor erro de previsão? Como otimizar ainda mais o desempenho?

Questão 7 (Intermediário) - Recomendação de Produtos em um Supermercado

Você trabalha para um supermercado e deseja implementar um sistema de recomendação baseado no histórico de compras dos clientes.

Tarefas:

- Utilize um dataset de transações de supermercado contendo itens comprados em conjunto.
- Aplique o algoritmo Apriori para descobrir padrões de associação entre produtos.
- Defina regras de associação com base em suporte, confiança e lift.
- Interprete os resultados e gere recomendações de produtos.

Pergunta: Quais foram as regras de associação mais relevantes? Como elas podem ser aplicadas

para aumentar as vendas?

Questão 8 (Avançado) - Recomendação de Filmes com Filtragem Colaborativa

Uma plataforma de streaming deseja sugerir filmes para os usuários com base nas avaliações de outros usuários.

Tarefas:

- Utilize um dataset de avaliações de filmes (exemplo: MovieLens).
- Implemente um modelo de filtragem colaborativa baseado em usuários e itens.
- Compare a filtragem colaborativa com abordagens baseadas em aprendizado profundo (exemplo: Autoencoders).
- Avalie o desempenho com métricas como RMSE e MAE.

Pergunta: Qual abordagem foi mais eficiente na recomendação de filmes? Como melhorar o sistema de recomendação?

Questão 9 (Avançado) - Classificação de Imagens de Raio-X com CNNs

Você recebeu um conjunto de imagens de raio-X e precisa desenvolver um modelo de Deep Learning para classificar doenças pulmonares.

Tarefas:

- Utilize um dataset de imagens médicas (exemplo: Chest X-ray dataset).
- Construa uma Rede Neural Convolucional (CNN) para classificar imagens saudáveis e com pneumonia.
- Aplique técnicas de data augmentation para melhorar a generalização.
- Avalie o modelo com métricas como precisão e matriz de confusão.

Pergunta: Quais desafios foram encontrados no treinamento? Como melhorar o desempenho do modelo?

Questão 10 (Intermediário) - Previsão de Vendas Mensais em Rede de Varejo

Contexto:

Uma rede de varejo deseja prever as vendas mensais de suas lojas com base em variáveis como gastos com publicidade, número de funcionários, presença de promoções e sazonalidade.

Tarefas:

- Carregar um dataset de vendas históricas (exemplo: *Rossmann Store Sales* do Kaggle).
- Realizar feature engineering (ex: codificar promoções como variáveis binárias, extrair mês/trimestre).
- Aplicar modelos de regressão como Regressão Linear, Árvore de Decisão e XGBoost.
- Avaliar os modelos usando RMSE, MAE e R^2 .

Pergunta:

Qual modelo apresentou a menor margem de erro? Quais variáveis tiveram maior impacto nas previsões?