# Ciências de dados

Análise Descritiva dos Dados

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Aplicação de técnicas estatísticas avançadas para modelagem dos dados:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Uso de ferramentas como regressão linear, classificação etc.:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Avaliação da adequação dos modelos estatísticos aos dados:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Implementação de modelos preditivos utilizando Python:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Avaliação da performance dos modelos preditivos:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Comparação entre diferentes abordagens de análise preditiva:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

# Aprendizado de máquinas

Coleta de Dados

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Limpeza e Pré-processamento dos Dados

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Construir um Modelo e Verificar a Matriz de Confusão

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Escolha de Algoritmos de ML

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Avaliação da Performance dos Modelos

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Otimização dos Hiperparâmetros

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Exemplo para Modelos de Regressão:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Validação Cruzada

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

### Documentação do Processo de Construção e Treinamento do Modelo

1. Definição do Problema

Objetivo:

Desenvolver um modelo preditivo para prever a demanda de peças de autopeças, visando melhorar a gestão de estoque e otimizar as operações de logística.

Tipo de Modelo:

Regressão (previsão de quantidade de demanda).

2. Coleta de Dados

Fontes de Dados:

Dados históricos de vendas obtidos do sistema de gerenciamento de vendas da empresa, incluindo vendas diárias, informações sobre produtos (tipo, categoria) e dados sazonais.

Período de Coleta:

Dados dos últimos 5 anos, de 2018 a 2023.

3. Pré-processamento de Dados

Limpeza de Dados:

Tratamento de valores ausentes, remoção de duplicatas e correção de erros de entrada.

Transformação de Dados:

Normalização dos dados numéricos e codificação one-hot para variáveis categóricas.

Divisão de Dados:

Divisão dos dados em 70% para treinamento e 30% para teste.

4. Seleção de Recursos

Recursos Utilizados:

Vendas diárias, informações de produtos e variáveis temporais como mês e dia da semana.

Técnicas de Seleção:

Análise de correlação e importância de recursos via modelos de árvores.

5. Escolha de Modelos

Modelos Avaliados:

* Regressão Linear: Baseline para comparação de desempenho.
* Random Forest Regressor: Devido à sua capacidade de lidar com relações não-lineares e importância de recursos integrada.

6. Otimização de Hiperparâmetros

Método Utilizado:

GridSearchCV para encontrar os melhores parâmetros do modelo Random Forest, como número de árvores e profundidade máxima.

Métricas de Avaliação:

RMSE (Root Mean Squared Error) e R² (Coeficiente de Determinação).

7. Treinamento do Modelo

Ambiente de Desenvolvimento:

* Python 3.8, Scikit-learn 0.24.

Processo de Treinamento:

Uso de validação cruzada com 5 folds para garantir a generalização do modelo.

8. Avaliação do Modelo

Resultados do Teste:

* Regressão Linear: RMSE = 250 unidades, R² = 0.60.
* Random Forest Regressor: RMSE = 180 unidades, R² = 0.75.