Roteiro apresentação

Slide 1

Sumário dos itens que serão apresentados

Slide 2

Visão geral do projeto e objetivos (texto descrito no material enviado)

Slide 3

Fluxograma do desenvolvimento do projeto

* Preparação dos dados
  + Compreensão e preparação dos dados
  + Verificação dos dados, valores nulos, tipos de dados, valores ausentes
  + Manipulação dos dados para as etapas seguintes, eliminando variáveis que não serão utilizadas e, caso necessário, realizando a transformação e/ou criação de novas variáveis
* Análise exploratória dos dados
  + Visualização dos dados para compreender a distribuição das variáveis
  + Identificar características relevantes para manipulação das variáveis e desenvolvimento dos modelos (feature selection)
* Modelos de machine learning
  + Preparação dos dados para executar modelos de machine learning: categorizar variáveis, padronizar variáveis numéricas
  + Separação dos dados em dados de treino e de teste
  + Seleção de atributos (feature selection) para selecionar as variáveis que melhor fornecem informações para construção dos modelos e predição
  + Executar diversos modelos para avaliar qual melhor apresenta resultados para a previsão da classificação
  + Selecionar os 3 melhores modelos
  + Realizar avaliação dos 3 melhores modelos para realizar análise comparativa entre eles
  + Realizar hyperparameter tuning nos modelos selecionados para encontrar melhores valores
  + Analisar a predição dos melhores modelos com os melhores parâmetros
  + Salvar os modelos e artefatos para construção do deploy dos modelos de machine learning
* Deploy
  + Importar os dados, modelos e artefatos para simular o deploy dos melhores modelos de machine learning
  + Aplicar a modelagem e preparação dos dados no formato ideal que os modelos foram treinados
  + Selecionar dados aleatórios
  + Aplicar os dados nos modelos e realizar a análise da previsão dos modelos

Slide 4

Graficos de barras com a distribuição das variáveis numéricas e categórica

Slide 5

Gráfico da matriz de correlação

Slide 6

Estrutura da divisão de treino e teste e a Feature selection: análise das variáveis que apresentaram melhores informações para os modelos

Slide 6

Fluxograma da modelagem dos modelos de machine learning. Os modelos inicialmente selecionados, a realização do treinamento e a escolha dos 3 modelos

Slide 7

Analise comparando os 3 modelos: acurácia e curva ROC

Slide 8

Escolha dos melhores parâmetros e salvar os modelos

Slide 9

Deploy do modelo e possíveis estratégia para a previsão das classes

Slide 10