

Trabalho P2

Pedido

- Relatório no formato de um jupyter notebook, de forma a aplicar um (ou mais) modelos de mineração de dados estudado(s) durante o curso.
- Os dados devem ser coletados o Kaggle (sugestão). Caso necessário, deve ser incluído também script para a extração de dados.
- O racional sobre a análise dos dados será considerado na avaliação do trabalho!
- Utilizem markdown para explicar o processo.

Organização do relatório

- Apresentação
- Dados
 - O que são?
 - Qual a fonte?
 - Link?
 - Contexto dos dados
- Objetivo
 - Definir uma questão que o modelo deve responder ("qual seria a função necessária para prever um determinado valor?", "como uma nova amostra se classificari em um determinado atributo?")
 - Apresentar uma ou mais hipóteses sobre os dados ("quanto maior a área do imóvel, maior o seu preço de venda")
- Análise de dados
 - Apresentar os dados
 - Qual o DataFrame
 - Tipos de dados
 - Resumos (médias, desvios-padrão, etc.)
 - Shape
 - Gráfico de distribuição de valores
- Limpeza dos dados
 - Eliminar colunas desnecessárias
 - Converter dados de texto
 - Eliminar dados inválidos
- Implementação do modelo
 - Dividir os dados em dados de treino e de teste

- Aplicar o modelo
 - Ajustar
- Testar o modelo
 - Aplicar o modelo ajustado sobre os dados de teste
 - Levantar métricas adequadas
- Conclusões
 - Apresentar gráficos
 - Apresentar tabelas
 - Observações finais sobre o modelo

Informações Gerais

- Prazo: 28/11/2021
- Sejam criativos!
- Submissão do trabalho pelo Teams, **em um arquivo .zip**, contendo:
 - Notebook (.ipynb)
 - Dados (.csv, .json, etc.)
- Se pegarem dados de outras fontes, apresentar o ETL