

Resumo do Projeto

Objetivo: Identificar uma oportunidade de projeto a partir da base fornecida e construir uma solução de ponta a ponta em Ciência de Dados — desde o entendimento do problema até a entrega de um modelo validado e interpretado.

Roteiro Completo do Projeto

1. Entendimento do Negócio e Definição do Problema

- **Análise dos stakeholders:** Qual a motivação por trás do problema?
- **Hipótese inicial:** A base se refere a problemas de tireoide? Classificação de pacientes com risco?

2. Exploração Inicial dos Dados

- Carregamento da base com `pandas`
- Verificação de tipos, valores nulos e estrutura
- Dicionário de dados (se não fornecido, inferido)

3. Análise Exploratória (EDA)

- Estatísticas descritivas
- Visualização das variáveis (categorias, distribuições)
- Correlações e insights para feature engineering

4. Preparação dos Dados

- Tratamento de valores nulos
- Conversão de tipos
- Codificação de variáveis categóricas
- Normalização/padronização se necessário

5. Modelagem

- Definição do problema: **Classificação**
- Separação treino/teste
- Aplicação de modelos:
 - Regressão Logística
 - Random Forest
 - XGBoost
- Avaliação com métricas (Accuracy, Precision, Recall, F1-score, ROC AUC)

6. Ajuste e Validação

- Tunagem de hiperparâmetros com GridSearchCV ou RandomSearch
- Validação cruzada (cross-validation)
- Análise de importância das variáveis

7. Conclusão e Apresentação dos Resultados

- Interpretação dos resultados
- Recomendações para os stakeholders
- Exportação do modelo (opcional: `joblib` ou `pickle`)

8. Documentação

- Comentários em cada célula
- Título, subtítulos e explicações claras
- Gráficos interpretáveis
- Sumário executivo ao final