# **Documentação do Projeto: Análise e Predição de Qualidade das Bananas**

## **1. Objetivo do Projeto**

O projeto tem como objetivo analisar e realizar predições sobre dados relacionados à qualidade das bananas. Através de gráficos e modelos preditivos, o sistema permite que o usuário faça upload de dados, visualize as informações, receba predições sobre a qualidade das bananas e re-treine os modelos com novos dados.

## **2. Ferramentas Utilizadas**

* **Flask**: Framework utilizado para a criação da aplicação web.
* **Pandas**: Biblioteca utilizada para manipulação e análise de dados em formato tabular (CSV).
* **Matplotlib e Seaborn**: Bibliotecas para geração de gráficos e visualizações.
* **Scikit-learn**: Biblioteca usada para construção dos modelos de predição (Random Forest).
* **Joblib**: Usada para salvar e carregar os modelos treinados.

## **3. Funcionalidades do Sistema**

### **3.1 Upload de Dados**

O sistema começa com o **upload de um arquivo CSV**, que deve conter dados sobre as bananas. O arquivo pode ser enviado pela interface web da aplicação. O processo é simples:

1. O usuário clica no botão "Escolher arquivo" e seleciona o arquivo CSV.
2. O sistema lê o arquivo, valida se é um CSV válido e o carrega para o processamento.
3. Após o upload, o sistema realiza a análise dos dados e os exibe ao usuário.

**O que acontece no backend:**

* O arquivo CSV é carregado usando o **Pandas**, que converte os dados em um formato que o sistema pode manipular.
* Após a leitura, os gráficos e modelos de predição são gerados com base nesses dados.

### **3.2 Tabela de Dados**

Após o upload do arquivo CSV, o sistema gera uma tabela com as **10 primeiras linhas dos dados**, mostrando as variáveis e suas características. A tabela é exibida para que o usuário possa verificar os dados que foram carregados.

* **Como a tabela é gerada**: O Pandas é utilizado para pegar as primeiras 10 linhas dos dados e gerar uma tabela HTML.
* **O que é mostrado na tabela**: As 10 primeiras colunas (se existirem) dos dados, com as informações organizadas.

### **Exemplo de como a tabela pode aparecer:**

| **Variedade** | **Teor de Açúcar (Brix)** | **Firmeza (kgf)** | **Comprimento (cm)** | **...** |
| --- | --- | --- | --- | --- |
| Banana A | 18.5 | 3.4 | 12.5 | ... |
| Banana B | 20.1 | 3.0 | 11.8 | ... |

### **3.3 Gráficos de Análise**

Após o upload dos dados e a exibição da tabela, o sistema gera **gráficos de análise**. Esses gráficos ajudam a entender visualmente os dados, destacando padrões e correlações importantes. Os gráficos apresentados incluem:

1. **Gráfico de Barras - Distribuição por Variedade**:
   * **Objetivo**: Exibe a quantidade de bananas por variedade.
   * **Interpretar**: Ajuda a identificar quais variedades são mais predominantes no dataset.
2. **Gráfico de Pizza - Proporção de Categorias de Maturação**:
   * **Objetivo**: Mostra a distribuição percentual das categorias de maturação (como 'Maduras', 'Verdes', etc.).
   * **Interpretar**: Facilita a visualização da proporção de cada categoria de maturação.
3. **Gráfico de Dispersão - Pontuação de Qualidade vs. Teor de Açúcar**:
   * **Objetivo**: Exibe a relação entre a pontuação de qualidade das bananas e o teor de açúcar.
   * **Interpretar**: Identifica possíveis correlações entre a qualidade e o teor de açúcar.
4. **Boxplot - Distribuição da Firmeza por Categoria de Qualidade**:
   * **Objetivo**: Mostra como a firmeza das bananas varia entre diferentes categorias de qualidade.
   * **Interpretar**: Ajuda a entender a distribuição da firmeza entre as categorias.
5. **Heatmap - Média do Índice de Maturação por Categoria de Qualidade**:
   * **Objetivo**: Visualiza a média do índice de maturação das bananas para cada categoria de qualidade.
   * **Interpretar**: Fornece uma visão geral da maturação das bananas em relação à sua qualidade.

### **3.4 Predições**

Após visualizar os dados e os gráficos, o sistema permite que o usuário **faça predições** sobre a pontuação de qualidade das bananas ou sobre a categoria de qualidade (como 'Madura' ou 'Verde'). O sistema possui dois tipos de predição:

1. **Predição para Pontuação de Qualidade**:
   * **O que o usuário faz**: Preenche os valores das variáveis (exemplo: teor de açúcar, firmeza e comprimento da banana) e clica no botão "Predict".
   * **Como funciona**: O modelo de regressão (Random Forest Regressor) é utilizado para prever a pontuação de qualidade com base nos dados preenchidos.
2. **Predição para Categoria de Qualidade**:
   * **O que o usuário faz**: Preenche as variáveis (como teor de açúcar e índice de maturação) e clica no botão "Predict".
   * **Como funciona**: O modelo de classificação (Random Forest Classifier) é utilizado para prever a categoria da banana (por exemplo, "Madura", "Verde").

### **Exemplo de Previsão:**

* Para o input:
  + Teor de Açúcar: 18.5
  + Firmeza: 3.0
  + Comprimento: 12.5
* O sistema pode prever:
  + **Pontuação de Qualidade**: 85
  + **Categoria de Qualidade**: "Madura"

### **3.5 Reupload e Re-treinamento de Modelos**

O sistema também permite que o usuário **envie um novo arquivo CSV** para re-treinar os modelos com dados mais recentes.

* **O que acontece**: O usuário faz o upload de um novo arquivo CSV e clica em "Re-treinar". O sistema irá:
  + Processar o novo arquivo.
  + Gerar os gráficos atualizados.
  + Re-treinar os modelos de predição com os novos dados.
  + Atualizar a tabela de visualização com os novos dados.
* **Requisitos**: O arquivo deve ser um CSV válido com a mesma estrutura do anterior (com as mesmas colunas e tipos de dados).

## **4. Fluxo de Trabalho do Sistema**

### **Passo a Passo:**

1. **Upload de Dados**: O usuário envia um arquivo CSV.
2. **Exibição de Dados**: O sistema exibe as 10 primeiras linhas do arquivo e os gráficos de análise.
3. **Realizar Predições**: O usuário pode preencher os valores das variáveis e fazer predições sobre a qualidade das bananas.
4. **Re-treinamento de Modelos**: O usuário pode re-enviar um arquivo CSV para re-treinar os modelos e gerar novos gráficos e predições.

## **5. Conclusão**

Este projeto foi desenvolvido para facilitar a análise e predição de dados relacionados à qualidade das bananas. Usando ferramentas como Flask, Pandas, Scikit-learn, Matplotlib e Seaborn, ele oferece uma maneira simples e visual de explorar os dados e tomar decisões baseadas em modelos preditivos.