Project Brief: Predicting Mobile Price Range for Appclick

Dataset: https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification
File to Use: `train.csv` only

Background & Business Context

Appclick is a new and ambitious smartphone company aiming to disrupt the mobile device market by offering competitive, feature-rich smartphones at affordable prices. With plans to challenge industry leaders such as Apple, Samsung, and Xiaomi, Appclick's success depends on making data-driven decisions from the start.

One of the most critical decisions in launching a new device is setting its price. Pricing directly affects customer perception, market competitiveness, profit margins, and overall brand positioning. Inaccurate pricing can lead to revenue losses or missed market opportunities.

To address this challenge, Appclick's business team has collected structured data from various mobile phones already in the market. These data points include specifications such as battery capacity, memory size, processor speed, and connectivity options. The objective is to identify patterns that can help predict the **price category** a new mobile device belongs to, based on its features.

As a data science intern or apprentice at Appclick, your role is to develop a machine learning model that can classify mobile devices into price categories using this data. Your model will serve as a prototype for the pricing engine used by Appclick's product and marketing teams.

Objective

Build and evaluate a supervised machine learning model that classifies a smartphone into one of four **price categories** (0 to 3) based on its technical specifications. The goal is to deliver a system that is both interpretable and accurate, with a focus on **F1-score** as the primary evaluation metric.

This model will help Appclick:

- Understand how mobile specifications influence price category.
- Support the product design team in identifying feature trade-offs.
- Assist the marketing team in determining competitive pricing strategies.

Dataset Overview

Download the dataset from:
https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification

File to Use: `train.csv`

This dataset contains 21 columns:

- 20 feature variables (e.g., RAM, battery_power, internal_memory, 4G support, etc.)
- 1 target variable: `price_range`
  - 0: Low Cost
  - 1: Medium Cost
  - 2: High Cost
  - 3: Very High Cost

The dataset includes only numeric and binary values, which simplifies preprocessing and model development.

Deliverables

1. Exploratory Data Analysis (EDA)

- Perform descriptive statistics and visualizations to understand data distribution and relationships.
- Conduct univariate and bivariate analyses focused on identifying features most correlated with `price_range`.
- Identify and address any outliers or anomalies.
- Provide insights into which features may be most influential in determining price category.

2. Data Preprocessing

- Normalize or scale relevant features as needed.
- Ensure the dataset is clean and ready for modeling (no missing values or inconsistencies).
- Create a reproducible train-test split (e.g., 80/20 or 70/30) with a fixed random seed.

3. Model Development

- Train **at least three classification models** (e.g., Random Forest, Logistic Regression, SVM, XGBoost, etc.) and settle on the best performing one.

- Optimize for **F1-score** as the primary metric, particularly important for balanced class performance.
- Use appropriate validation techniques (e.g., k-fold cross-validation) to avoid overfitting.

4. Evaluation Metrics

Each model must be evaluated using:

- **F1-score**
- **Classification report** (precision, recall, F1-score for each class)
- **Confusion matrix**

Include visualizations where appropriate to support interpretation.

5. Model Interpretation

- Extract and interpret **feature importance** from at least one model.
- Explain how specific features (e.g., RAM, internal memory, 4G support) influence the predicted price range.
- Translate technical findings into actionable business insights for Appclick's design and marketing teams.

6. Final Submission Requirements

Each team or individual must submit the following:

- A **GitHub Gist** containing:
  o Clean, well-documented Python code (preferably in Jupyter Notebook format). Do the assignment on colab, click on save a copy as a github gist and submit the generated link.
  o Modular pipeline and clearly labeled sections for EDA, preprocessing, modeling, and evaluation (Extremely important to label each section).
- A **Medium article** that includes:
  o Summary of your approach and findings
  o Key insights and decisions made throughout the project
  o Visuals to support your explanation
  o Final model recommendation and business impact discussion

The Gist and article links are to be submitted via the submission tab on the portal.

Key Skills To Be Evaluated.

- Real-world supervised classification

- Data cleaning and transformation
- Model evaluation and performance tuning
- Communicating results to non-technical stakeholders
- Business-focused data storytelling

GOOD LUCK.