كلية الهندسة
College of Engineering

جامعة عفت
EFFAT UNIVERSITY

Computer Science Department
Fall 24
Data Science

Students' Names	IDs	Section
Lujain Almarri	S20106753	1
Leen Sharab	S21107195	1
Mawaddah Alagha	S20106707	1

Project Proposal

I. Project Name:

Predicting Mobile Phone Prices Using Machine Learning

II. Description:

This project focuses on analyzing smartphone specifications and pricing data to identify key factors influencing smartphone prices and to develop predictive models for price estimation. The datasets include:

- **Mobile Specifications Dataset:** Contains detailed information about smartphone features such as RAM, processor speed, storage, and battery capacity.
- **Pricing Dataset:** Provides historical pricing data for various smartphone models across different brands.
- **Feature Engineering Dataset:** Includes engineered variables such as performance scores, price per RAM, and price per storage.
- **Market Insights Dataset:** Offers insights into consumer preferences and market trends, where available.

III. The Goal:

This project aims to contribute to the understanding of smartphone pricing dynamics by:

- Providing manufacturers with data-driven recommendations for product development and pricing.
- Empowering consumers with insights into price-to-performance trade-offs.
- Advancing research in the field of predictive modeling for consumer electronics.

- **Proposal:**

1. High-Level Statement of the Problem:

Computer Science Department
Fall 24
Data Science

In today's fast-evolving smartphone market, understanding the factors influencing mobile phone prices is crucial for manufacturers aiming to optimize pricing strategies and for consumers seeking value for money. Smartphones vary widely in their specifications, performance metrics, and branding, which collectively impact their market prices. However, the extent to which each feature contributes to price determination remains unclear.

Research Question:

What are the most important factors influencing mobile phone prices, and how do these factors interact to predict price using machine learning?

Existing Research:

While there has been extensive research on individual smartphone features such as camera quality, battery life, and processor speed, comprehensive studies integrating multiple features for price prediction remain limited. Advanced machine learning techniques provide an opportunity to analyze complex interactions between features, offering insights that traditional statistical methods may overlook. This project builds on existing studies by employing a holistic approach to modeling smartphone prices using diverse datasets.

2. Outcome Variable:

The outcome variable for this project is the *price in USD* of mobile phones, representing their market value.

Conceptual Relationship: Price serves as a measure of a smartphone's value, influenced by a combination of specifications, performance, and branding. Modeling this relationship helps identify the most impactful features driving pricing strategies.

Summary Statistics:

- Mean price: $359.88
- Standard deviation: $112.50
- Minimum price: $239.99
- Maximum price: $491.88
- Distribution: Prices are right-skewed, with the majority of devices priced below $600.

3. Predictor Variables:

To model the outcome variable, the following 10 predictors will be used:

1. **RAM (GB):** Affects multitasking performance.

2. **Processor Speed (GHz):** Reflects computational capabilities.
3. **Storage Capacity (GB):** Determines internal storage availability.
4. **Battery Capacity (mAh):** Indicates longevity between charges.
5. **Fast Charging (W):** Represents charging speed.
6. **Specs Score:** Combines multiple performance metrics into a single value.
7. **Display Size (inches):** Influences usability and user experience.
8. **Price per RAM (USD/GB):** Highlights cost-efficiency of memory.
9. **Performance Score:** An engineered feature combining processor speed, RAM, and specs score.
10. **Price per Storage (USD/GB):** Demonstrates cost-efficiency for storage.

Data Sources: The primary dataset for this project includes publicly available smartphone specifications and pricing data. Additional preprocessing and feature engineering steps will ensure the data's readiness for modeling.

4. Definition of "Success":

A successful project will achieve the following objectives:

- **Predictive Accuracy:** Develop a machine learning model with an R^2 score above 0.95 and a Mean Squared Error (MSE) below 2000 on the test dataset.
- **Insights Generated:** Identify the most significant factors influencing smartphone prices.
- **Practical Applications:** Provide actionable recommendations for manufacturers to refine pricing strategies.

Key Metrics for Success:

- **Model Performance:** High accuracy and low error rates in predicting smartphone prices.
- **Feature Importance:** Clear identification of key variables driving pricing decisions.

- **Interpretability:** Models should provide transparent and actionable insights.

5. Methodology and Tools

This project will follow a structured approach:

1. **Data Preprocessing:**
    - Address missing values and outliers.
    - Normalize numerical features for consistent scaling.
    - Engineer additional features to enhance predictive power.
2. **Exploratory Data Analysis (EDA):**
    - Visualize relationships between predictors and price.
    - Identify correlations and trends using scatter plots and heatmaps.
3. **Model Development:**
    - Train multiple machine learning models, including Random Forest, Gradient Boosting, and Linear Regression.
    - Select the best-performing model based on cross-validation metrics.
4. **Model Evaluation:**
    - Evaluate models using R^2, Mean Absolute Error (MAE), and Mean Squared Error (MSE).
    - Validate model performance on a holdout test dataset.
5. **Insights and Recommendations:**
    - Generate feature importance plots to rank predictors.
    - Provide actionable insights for manufacturers and consumers.

Tools and Libraries:

- Python (pandas, numpy, sklearn, matplotlib, seaborn)
- Machine learning frameworks (e.g., XGBoost, LightGBM)