**Project Title:** Feature Extraction and Price Prediction for Mobile Phones

**Abstract**

This project focuses on building a predictive model to estimate the price of mobile phones based on their technical specifications. Through feature extraction, data preprocessing, exploratory analysis, and machine learning modeling, the system identifies key factors influencing price and predicts accurate values using regression models.

**1. Introduction**

Mobile phone pricing varies widely based on specifications such as memory, camera quality, and battery capacity. Predicting price from features helps retailers, manufacturers, and consumers understand market trends. This project aims to extract meaningful features and develop machine learning models to estimate mobile phone prices accurately.

**2. Problem Statement**

Given a dataset of mobile phone specifications, the objective is to:

- Extract relevant numerical features

- Clean and preprocess the data

- Train machine learning models

- Predict the mobile phone price

**3. Dataset Description**

The dataset contains columns such as:

- Memory

- RAM

- Rear Camera (MegaPixel)

- Front Camera (MegaPixel)

- Battery Capacity

- Price (target variable)

**4. Data Preprocessing**

The following steps were applied:

- Removal of special characters like "MP" from camera features

- Conversion of numeric strings with commas into integers

- Handling missing values using median imputation

- Ensuring correct data types using *pd.to_numeric()*

- Splitting dataset using a fixed random_state for reproducibility


## 5. Exploratory Data Analysis (EDA)

EDA revealed:

- Price strongly correlates with Memory, RAM, and Camera features

- Higher RAM and internal storage result in higher prices

- Camera megapixels also show a moderate positive relationship


## 6. Machine Learning Models Used

The following regression models were developed:

- Linear Regression

- Random Forest Regressor (with random_state=42)

- Gradient Boosting Regressor (with random_state=42)


## 7. Model Evaluation

The models were evaluated using:

- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)

- Mean Absolute Error (MAE)

- $R^2$ Score


Random Forest and Gradient Boosting outperformed Linear Regression due to their ability to capture non-linear feature interactions.


## 8. Results & Findings

- Gradient Boosting achieved the lowest RMSE

- Random Forest showed strong generalization and stability

- Memory, RAM, and camera specifications were the top predictors of price

## 9. Conclusion

The project successfully predicts mobile phone prices using machine learning techniques. Feature extraction and proper preprocessing significantly improved model performance. Tree-based models proved most effective in capturing complex relationships.