Project Report : Mobile Phone Price Prediction System

This project focuses on building a machine learning system that can predict the **price range of mobile phones** based on their specifications. The target variable includes four classes: **0 (Low Cost), 1 (Medium Cost), 2 (High Cost), and 3 (Very High Cost)**. The aim was to analyze the given dataset, train an effective model, and build an interactive interface that demonstrates how the model performs in real-world usage.

1. Problem Understanding

Mobile phones vary widely in hardware capabilities, and these specifications strongly influence their market price. The dataset provided includes features such as:

- Battery power
- RAM
- Camera specifications
- Pixel resolution
- Storage capacity
- Screen dimensions
- Connectivity options like 3G, 4G, Bluetooth, WiFi, etc.

The objective was to use these features to **classify phones into price categories** rather than predicting an exact numerical price.

2. Data Exploration & Preprocessing

I began by loading the dataset and performing exploratory checks such as:

- Detecting missing values
- Understanding feature distributions
- Observing correlations among variables

Since the dataset was clean and well-structured, minimal preprocessing was required. I separated the **features (X)** and **target variable (y)**, then applied a **StandardScaler** to normalize the numerical values, ensuring that all features contributed proportionally to the model.

3. Model Selection & Training

I experimented with a few algorithms conceptually, and ultimately used a **Random Forest Classifier**, which is well known for:

- Handling non-linear relationships
- Performing well with mixed feature types
- Being resistant to overfitting
- Providing interpretability through feature importance

The data was split into 80% training and 20% testing sets. After training, the model achieved a strong accuracy of **approximately 89%**.

4. Evaluation & Insights

The evaluation metrics (precision, recall, and F1-score) showed balanced performance across all four price categories.
Key insights included:

- **RAM** was the single most influential feature.
- **Pixel resolution** (height and width) significantly affected price range.
- Battery capacity, internal memory, and processor details also played important roles.

The confusion matrix confirmed that the model rarely misclassified phones into completely wrong categories (e.g., Low → Very High), proving that the predictions were reliable.

5. Building the Application

To demonstrate the model, I created a **Streamlit web application** (app.py) where users can enter mobile specifications through a clean and simple interface.
Key features of the app:

- Yes/No choices for features like Bluetooth, 3G, 4G, WiFi, Dual SIM, etc.
- Sliders and numeric inputs for RAM, screen size, camera specs, etc.
- A "Predict" button that outputs the phone's predicted price range.

This interactive app provides a practical example of how machine learning can be integrated into a real-world decision-support system.

6. Conclusion

This project helped me understand the complete machine learning lifecycle:

- Dataset understanding
- Preprocessing and scaling
- Model training and analysis
- Saving and deploying a model
- Building a working application around it

The final system is accurate, easy to use, and demonstrates how data-driven insights can assist businesses or consumers in evaluating mobile devices. It also showcases my ability to take a project from concept to deployment-ready implementation.