PROJECT: 1

Mobile Price Prediction

Name: Mani P

Email: maniprabu991@gmail.com

Program: Unified Mentor ML Internship Program

1. Problem Statement

The mobile phone market contains a wide range of devices with different hardware and software specifications, making pricing decisions complex. Accurately predicting the price range of a mobile phone based on its features helps manufacturers and sellers position products effectively and supports consumers in making informed choices. This project aims to classify mobile phones into different price ranges using machine learning.

2. Objective

- To build a machine learning model that predicts the mobile phone price range.
- To explore and analyze mobile phone features affecting pricing.
- To perform feature selection to improve prediction accuracy.
- To tune a Random Forest model for optimal performance.
- To save the trained model for future deployment.

3. Tech Stack Used

- **Programming Language:** Python
- **Libraries:** NumPy, Pandas, Matplotlib, Seaborn
- **Machine Learning:** Scikit-learn
- **Model:** Random Forest Classifier
- **Feature Selection:** SelectKBest, RFE, PCA
- **Model Saving:** Pickle
- **Platform:** Jupyter Notebook

4. Project Architecture / Workflow

1. Load and explore the dataset
2. Perform data preprocessing
3. Feature engineering
4. Feature selection
5. Split dataset into training and testing sets
6. Train Random Forest model
7. Hyperparameter tuning using GridSearchCV
8. Model evaluation
9. Save trained model

5. Implementation Details

The dataset contains 2000 mobile phone records with features such as RAM, battery power, screen dimensions, camera details, connectivity options, and processor specifications.
Feature engineering was applied by creating derived attributes like:

- Pixel Area
- Screen Area
- Total Camera Megapixels
- 3G/4G Support Indicator

Feature selection techniques such as SelectKBest, RFE, and PCA were used to reduce dimensionality and improve accuracy. The dataset was split into 80% training and 20% testing. Random Forest with GridSearchCV was used to fine-tune hyperparameters.

6. Output / Results

- **Top Features Selected:**
  ram, battery_power, pixel_area, int_memory, n_cores
- **Model Accuracy: 92.75%**

Classification Report Summary:

- High precision and recall across all price categories
- Balanced performance for low, medium, high, and very high price ranges

Confusion Matrix Highlights:

- Correct classification for majority of samples
- Minimal misclassification between adjacent price ranges

Feature importance visualization showed RAM and battery power as the most influential features.

7. Challenges Faced

- Handling multicollinearity between features
- Selecting optimal feature combinations
- Preventing overfitting during model training

8. Future Enhancements

- Deploy the model using a web application (Flask / Streamlit)
- Add real-time mobile data scraping
- Try advanced models like XGBoost or LightGBM
- Improve prediction explainability using SHAP

9. Conclusion

This project successfully implemented a Random Forest classifier to predict mobile phone price ranges with an accuracy of 92.75%. Feature selection significantly improved model performance, and the trained model is ready for deployment to predict pricing for new mobile devices.