Mobile Price Prediction

Introduction

Mobile Price Prediction project uses machine learning to estimate phone prices based on features. By analyzing a comprehensive dataset, the model learns correlations between RAM, storage, camera quality, battery capacity, etc. Its objective is to develop an accurate predictive tool for mobile price range estimation.

Literature Survey

1. "Mobile Price Class prediction using Machine Learning Techniques"

Authors: Muhammad Asim and Zafar Khan

The study encompasses key steps: data collection, dimensionality reduction, utilizing decision tree and Naive Bayes algorithms, and conducting a comparative study on accuracy based on selected features and algorithms employed.

2. "Prediction of Mobile Phone Price Class using Supervised Machine Learning Techniques"

Authors: Varun Kiran and Dr. Jebakumar R.

Investigates the use of supervised machine learning classification algorithms, including Decision Tree, Linear Discriminant Analysis, Naive Bayes, K-Nearest Neighbors (KNN), and Random Forest Classification, for predicting the price class of mobile phones.

Data Collection and Exploratory Data Analysis

Data Collection

Total of 13 features of mobiles are collected from 91mobiles.com .

Brand, RAM (GB), Rear Camera (MP), Front Camera (MP), Battery (mAh), Display Size Resolution (inches), Resolution (pixels x pixels), Screen To Body Ratio (% percentage), ROM (GB), Fingerprint sensor, Processor (number of cores + GHz)

Intentionally there were supposed to be 14 features, but Unfortunately, only 12 of them were fetched as expected.

Initial dataset dimension was 1021 x 12

Data Preprocessing

Data has been cleaned from noisy and irrelevant entries, transformed into numeric data types, Duplicate records are removed , Features which have no impact on price are removed. After the preprocessing stage our final dataset dimension are 305 x 10

Feature	DataType	MIN	MAX	MEAN	StdDIV	CORRELATION-COEFFICIENT TO PRICE
Brand	int	1.00	18.00	12.32	12.32	0.001768
RAM	int	8.00	16.00	6.21	2.56	0.662607
Rear Camera	int	8.00	232.00	60.15	31.95	0.423225
Front Camera	float	5.00	25.00	12.06	4.53	0.175510
Battery	float	3100.00	6000.00	4951.35	505.86	-0.355946
Display Size	float	5.80	7.60	6.59	0.20	0.372888
Resolution	int	2320.00	4656.00	3254.85	603.83	0.537929
Screen To Body Ratio	float	78.70	98.470	84.77	2.57	0.614781
RAM	int	32.00	512.00	119.08	71.91	0.729783
Processor	float	7.20	20.56	16.19	2.11	0.495853

Metric and Model Selection

Model Selection

1. Regression - It assumes a linear relationship between the independent variables (features) and the dependent variable (price in this case). It allows us to understand the impact of each feature on the predicted price and provides a baseline for comparison.

2. Random Forest - Random Forest is an ensemble learning technique that combines decision trees to predict complex relationships. It handles missing values, outliers, and nonlinearity well, making it suitable for our dataset. It also provides feature rankings, identifying influential factors in mobile phone price.

Metric Selection

1. Mean Absolute Error (MAE) - It measures the proportion of predictable variance in the dependent variable (price) explained by independent variables (features). It evaluates the model's fit to the data, where higher values indicate a better fit.

2. Root Mean Squared Error (RMSE) - It calculates the square root of the average of squared differences between predicted and actual prices. It emphasizes larger errors and provides insight into error magnitude

3. R-squared (R^2) Score - It measures the proportion of predictable variance in the dependent variable (price) explained by independent variables (features). It evaluates the model's fit to the data, where higher values indicate a better fit.

Model	Data	MAE	RMSE	R^2
Linear Regression	**Train**	1.200513e+08	10956.793795	0.717488
Linear Regression	**Test**	1.532262e+08	12378.456330	0.687668
Random Forest	**Train**	1.020359e+07	3194.305402	0.975988
Random Forest	**Test**	6.912799e+07	8314.324157	0.859091

Model Evaluation and Future Works

Model Evaluation

The Random Forest model outperformed Linear Regression in both datasets, as indicated by a higher R-squared (R^2) score. This is because the Random Forest model excels in capturing complex relationships and interactions among features. Mobile phone prices are influenced by various factors that may not have a clear linear relationship with price, such as brand, front camera resolution, display size, and battery life. Linear Regression assumes linearity, which may not hold true in this case. Random Forest, with its ensemble of

decision trees, effectively handles non-linear relationships, captures intricate patterns, and interactions among features, resulting in improved predictions of mobile phone prices.

Code

Language and dependencies

The project used python programming language. The project's dependencies and requirements are documented in the "**requirements.txt**" file.

The code files

1. Data Pre-processing-1.ipynb - Conversion of data into numeric datatypes and Removing duplicates

2. Data Pre-processing-2.ipynb - Dropping non useful features and Removing outliers

3. Exploratory-Data-Analysis.ipynb - understand the relationships between variables, identify patterns.

4.Model-Building-&-Prediction-Linear-Regression.ipynb - dedicated to building, training, and evaluating the Linear Regression model

5.Model-Building-&-Prediction-Random-Forest.ipynb - dedicated to building, training, and evaluating the Random Forest model.

The Dataset csv files

1.Mobile-DataSet.csv

Dataset created by Data Pre-processing-1.ipynb

2.Mobile-DataSet-Pre-Processed.csv

Dataset created by Data Pre-processing-2.ipynb

The final DataSet

Rowdata folder : multiple CSV files that serve unprocessed rowdata