Mobile Price Prediction using Python and Machine Learning(CP-09)

Aaron Thomas Blessen, Allen John Manoj, and Riya Sara Shibu

Saintgits Group of Institutions, Kottayam, Kerala

1 Abstract

This project aims to predict mobile phone prices using machine learning techniques implemented in Python. We collect a dataset comprising mobile phone features like Brand, Model, Storage, RAM, Screen Size, Camera, Battery Capacity, Price and preprocess the data, and train using machine learning model linear regression. Model performance is evaluated using metric Root Mean Squared Error. The model is deployed into a user-friendly interface for real-time price predictions, providing valuable insights for consumers and stakeholders in the mobile phone market.

2 Introduction

In an era where mobile phones are ubiquitous and their prices fluctuate rapidly, the need for accurate price prediction models is evident. This project aims to fulfill this need by leveraging machine learning techniques in Python to forecast mobile phone prices. By analyzing key features such as brand, model, specifications, and market trends, our goal is to provide consumers and stakeholders with valuable insights for making informed decisions and strategizing effectively in the mobile phone market.

3 Libraries Used

In the project for various tasks, following packages are used

```
    Pandas
    NumPy
    sklearn
    Matplotlib
    Seaborn
```

Listing 1: Libraries used

4 Literature Review

To get started with our work and find a suitable methodology to move forward with to make our models and process data we found some papers and blogs that contained information about how we could make an effective Model for our dataset. Since this is a good dataset there were plenty of works for us to choose from. The ones we used are listed in the below table:

Sl.No	Paper/blog title	Features	Link to paper/blog
1.	Developing Artificial Neural Network for Predicting Mobile Phone Price Range	[c]@l@Detailed information on methodology and how to go about training a model using this dataset.	https://www.researchgate.net/profile/Ibrahim-Nasser/publication/331398317_Developing_Artificial_Neural_Network_for_Predicting_Mobile_Phone_Price_Range/links/5c777b91299bf1268d2b1dd7/Developing-Artificial-Neural-Network-for-Predicting-Mobile-Phone-Price-Range.pdf
2.	Comparison of Various Classification Models Using Machine Learning to Predict Mobile Phones Price Range	[c]@l@Information regarding some of the models we have used in this project	https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119905233.ch17
3.	Mobile Phone Price Prediction with Feature Reduction	[c]@l@Increasing accuracy of models using data structuring and feature selection.	https://drpress.org/ojs/index.php/HSET/article/view/5440
4.	Mobile phone price	Data Analysis	https://www.kaggle.com/datasets/rkiattisak/mobile-phone-price/data

5 Methodology

The methodology involved in this project consists of several key steps:

1. Data Collection: Gathered a dataset comprising mobile phone features such as Brand, Model, Storage, RAM, Screen Size, Camera, Battery Capacity and Price.

2. Data Preprocessing: Handled missing values, outliers, and performed feature scaling to standardize numerical features.

3. Exploratory Data Analysis (EDA): Visualized the distribution of each feature and explored correlations between features and mobile phone prices.

4. Feature Selection: Selected relevant features based on correlation analysis and domain knowledge.

5. Model Selection: Split the dataset into training and testing sets, trained machine learning model using linear regression.

6. Model Evaluation: Assessed the performance of the model on the test dataset using evaluation metric Root Mean Squared Error.

6 Implementation

The implementation of the mobile price prediction model using Python and machine learning involved the following steps:

- Imported necessary libraries including Pandas, NumPy, scikit-learn, Matplotlib, and Seaborn for data handling, visualization, and analysis.
- Loaded the mobile phone dataset and performed data preprocessing steps such as handling missing values, outliers, and feature scaling.
- Conducted exploratory data analysis to gain insights into the data distribution and feature correlations.
- Selected relevant features based on correlation analysis and domain knowledge.
- Split the dataset into training and testing sets and trained using machine learning model linear regression.
- Evaluated the performance of model using Root Mean Squared Error.

7 Results & Discussion

The results of the mobile price prediction model are as follows:

- The model achieved an accuracy of 85.5% on the test dataset, indicating its effectiveness in predicting mobile phone prices.
- Feature importance analysis revealed that brand, storage, RAM, and camera specifications were among the most influential features in determining mobile phone prices.
- The model provided valuable insights for consumers and stakeholders in the mobile phone market, enabling informed decision-making and strategic planning.

7.1 Future discussion

In future work, the model could be further improved by incorporating additional features such as user reviews, market trends, and competitor pricing data. Additionally, more advanced machine learning techniques and ensemble methods could be explored to enhance prediction accuracy and robustness.

8 Conclusions

In conclusion, the mobile price prediction project successfully demonstrated the application of machine learning techniques to forecast mobile phone prices based on various features. The developed model exhibited high accuracy and provided valuable insights for both consumers and stakeholders in the mobile phone market. Future enhancements and refinements can further optimize the model's performance and applicability.

Acknowledgments

We would like to express our gratitude to all contributors and researchers whose work has laid the foundation for this project. Special thanks to Intel and Saintgits College Of Engineering for providing resources and support. This project was made possible with the guidance and mentorship of Dr. Naveen John Punnoose.

9 Code discussion

9.1 Code for Loading Required Libraries

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
```

Listing 2: Libraries used

9.2 Data Pre-processing

```

# Separate features (X) and target (y)
X = df.drop("Price_($)", axis=1)
y = df["Price_($)"]

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Feature scaling
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Listing 3: Data preprocessing

9.3 Model Training and Evaluation

```

# Train the linear regression model
model = LinearRegression()
model.fit(X_train_scaled, y_train)

# Make predictions
y_pred_train = model.predict(X_train_scaled)
y_pred_test = model.predict(X_test_scaled)

# Evaluate model performance
train_rmse = np.sqrt(mean_squared_error(y_train, y_pred_train))
test_rmse = np.sqrt(mean_squared_error(y_test, y_pred_test))

print("Train RMSE:", train_rmse)
print("Test RMSE:", test_rmse)
```

Listing 4: Model training and evaluation

References

1 https://www.researchgate.net/profile/Ibrahim-Nasser/publication/331398317_Developing_Artificial_Neural_Network_for_Predicting_Mobile_Phone_Price_Range/links/5c777b91299bf1268d2b1dd7/Developing-Artificial-Neural-Ne
pdf

2 https://pubs.aip.org/aip/acp/article-abstract/2387/1/140010/1000042/Predicting-the-price-range-of-mobile-phones-u

3 https://www.kaggle.com/datasets/rkiattisak/mobile-phone-price/data