Karunya INSTITUTE OF TECHNOLOGY AND SCIENCES
(Declared as Deemed to be University under Sec.3 of the UGC Act, 1956)
MoE, UGC & AICTE Approved
NAAC A++ Accredited

An internship report submitted by

Team Falcon

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE AND ENGINEERING

under the supervision of

Dr. T. Jemima Jebaseeli
(Associate Professor)

DIVISION OF COMPUTER SCIENCE AND ENGINEERING
KARUNYA INSTITUTE OF TECHNOLOGY AND SCIENCES
(Declared as Deemed to be University under Sec-3 of the UGC Act, 1956)
Karunya Nagar, Coimbatore - 641 114. INDIA

DIVISION OF COMPUTER SCIENCE AND ENGINEERING

TITLE

MOBILE PRICE PREDICTION USING MACHINE LEARNING

TEAM NAME

FALCON

TEAM MEMBERS

N Benjamin Studd
Karthik M D

DATE OF SUBMISSION

15 JULY 2023

LITERATURE SURVEY

1.1 Introduction

The mobile phones market has undergone colossal changes from 2010. Prior to 2010. Blackberry OS and Microsoft OS were the ones having the biggest share in the market till 2005, but then they were overtook by the Symbian OS. Motorola, Samsung and Sony Ericsson held the most shares of the mobile phones market. But soon with the introduction of Android OS by Google changes the dynamic of mobile phones. Mobile phones started becoming smarter and powerful. Now it's the era of smartphones. Smartphones are not strange things anymore for people anymore. Smartphones are mobile phones with computer abilities and internet search. It has become a source of entertainment and a communication tool for the vast population. With technological innovations, the structure of smartphone market has undergone continuous evolution. Market structures according to Sutton's Model for tech products has evolved has a two stage game. The first stage brand invests in sunk costs that is research and development (R&D) and advertisement. In the second stage the brand competes on price based on the decisions in stage one.

In this project we are going to study about the second stage that is how the smartphone brands decide the prices of each phone.

1.2 Evolution of Mobile Industry

Mobile phone was officially launched on April 3, 1973, named Motorola Dyna Tac, which was invented by Martin Cooper. Motorola Dyna Tac had the same shape as today's mobile phones, but it was quite bulky weighing more than 1 kilogram and didn't gain much popularity. Since then mobile phones have developed constantly.

The era of smartphones began from 2007 when Apple showcased its first iPhone. At that time android was under development and was slowly growing and was on the way to become one of the most fascinating foundations. Android continues to grow and expand, lots of manufacturers supported Android. In today's era, smartphones is not only growing in popularity but also gives people a series of new possibilities in all fields like

entertainment anytime anywhere, information exchange, etc. Global smartphone audience count surpassed 1 billion in 2012.

1.3 Competition and Strategies

The process by which the mobile manufacturers chooses to charge customers for their services is one of the most important duties that a company must do. It includes the choices on benchmarks that are to be made while calculating prices, higher and lower price limits, and the price policies implemented by competitors in the market. Keeping all these in mind, the compilation of data about costs has a lot of significance. Telecommunication firms used to rely on cost based pricing strategies, demand based pricing strategies, and competition based pricing strategies while imposing their prices. However, the Indian mobile phone industry has implemented a very creative pricing techniques in light of the fact that cost, demand, and competition all need to be taken into consideration simultaneously while deciding the prices for mobile phones. So predicting the price of mobile phones is a crucial step for each brand for maximizing their profit while keeping the price range in reach of customers so that the particular phones becomes successful. Prediction of price range for a particular mobile phone can be done by applying various machine learning algorithms like Linear Regression, Logistic Regression, Decision Trees, etc.

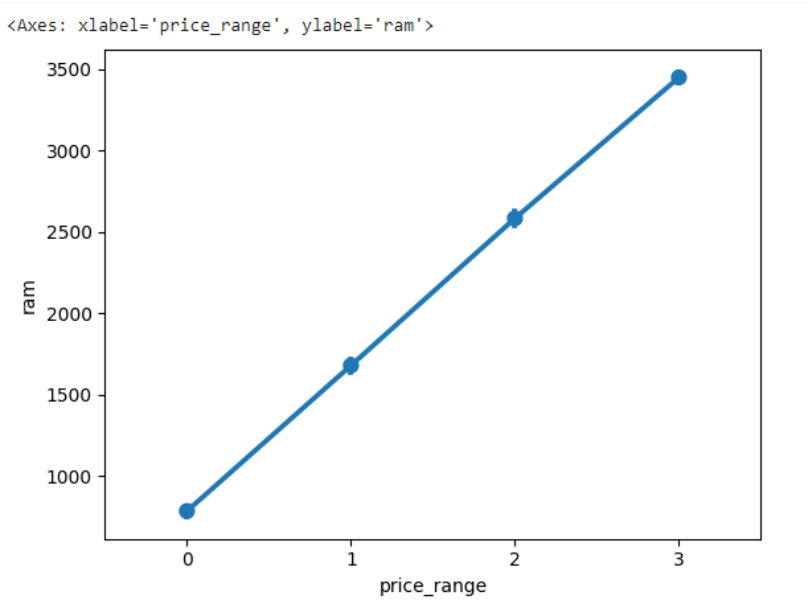
PROJECT EXPLANATION

2.1 Dataset Selection

In this project we have analyzed the factors that play an important role in the price of particular mobile phone by applying suitable machine learning algorithms. We have used a dataset from Kaggle to analyze and explore the factors which affect the price of mobile phones. This dataset contains the specifications of about two thousand mobile phones. The dataset contains over 21 features of a mobile phone.

2.2 Exploratory Data Analysis

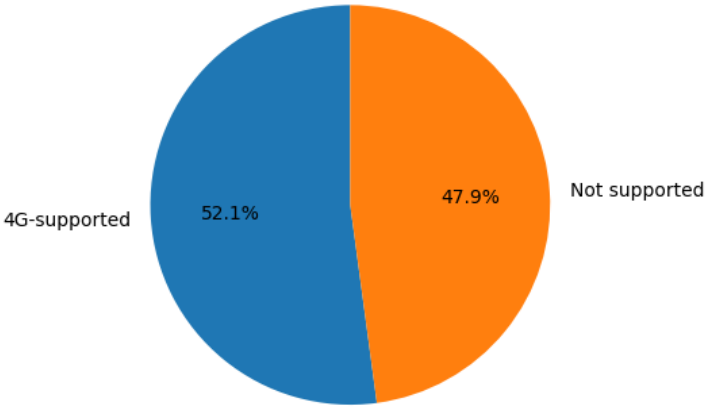
We then have did the relevant exploratory data analysis on our dataset in order to identify and rectify any errors if present in the dataset we have chosen as well as to understand the dataset by visualizing the patterns occurring in the dataset. The 'price_range' column is our target column and it is divided into 4 categories namely 0,1,2,3 which shows low cost to high cost price range. By testing for outliers/missing values and removing rows with missing values so that we can create efficient and much more accurate machine learning model. As missing values or outliers lead to noise in the dataset which decreases the real accuracy of machine learning algorithm. After that we have analyzed how some features are related to each other with the help of line plots, pie charts, boxplots, etc.



For instance, here we have analyzed how ‘RAM’ affects the price range of mobile phones. We can understand from the given lineplot that the price increases with the increase in RAM.

```
labels = ["4G-supported",'Not supported']
values=df['four_g'].value_counts().values

fig1, ax1 = plt.subplots()
ax1.pie(values, labels=labels, autopct='%1.1f%%',startangle=90)
plt.show()
```



Here we have analyzed how many of the mobile phones present in our dataset support 4-G cellular technology.

2.3 Implementation of Machine Learning Algorithms and Evaluation

For predicting mobile phone prices, we have used multiple machine learning algorithms in order to find out which model is the most suitable that is which model gives us the highest accuracy. We have applied the following machine learning algorithms:

- Linear Regression
- Logistic Regression
- KNN
- Random Forest
- Decision Tree

Dividing our dataset into training and validating with the help of ‘train_test_split’ present in sklearn library.

1. **Linear Regression :**

Linear regression is a supervised learning algorithm used for predicting a continuous numerical output variable based on one or more input features. It assumes a linear relationship between the input features and the target variable. The goal is to find the best-fit line that minimizes the sum of squared differences between the predicted and actual values. Therefore we have applied it on our dataset so that it will help in

In linear regression we got the R2 score of 0.91

```
[ ]  acc_L=lr.score(x_test,y_test)
     print("R2 Score for linear Regression: ",acc_L)

     R2 Score for linear Regression:  0.9140868303545098
```

2. **Logistic Regression :**

Logistic regression is a statistical algorithm used for binary or multi-class classification tasks. It models the relationship between a set of input variables and a binary or categorical output variable by estimating the probabilities using a logistic function. The algorithm learns the optimal parameters through maximum likelihood estimation, allowing it to make predictions and classify new instances based on the learned model. As the target variables of the dataset are discrete, we used logistic regression here.

The logistic regression model achieved an accuracy of 0.69 in predicting mobile device price ranges.

```
[ ]  #To find the accuracy of the model we will use accuracy_score

     acc=accuracy_score(y_test,y_pred)
     print("Accuracy of Logistic regression model is: ",round(acc,2))

     Accuracy of Logistic regression model is:  0.69
```

3. **K-Nearest Neighbors (KNN) :**

K-Nearest Neighbors is a very simple but effective algorithm used for classification tasks. If a data point is given, KNN determines its class or predicts its value based on the majority vote or average of the K nearest neighboring data points in its feature space. KNN is also non-parametric algorithm and does not

make assumptions about the underlying data distribution. It is also called a lazy learner algorithm as it learns only during the testing period.

Hence we have used this algorithm to develop a model and we evaluated it based on accuracy score. For finding the optimal K value we have used GridSearch Method which helped increase the accuracy of our KNN model.

```
acc_knn=accuracy_score(y_test, y_pred_knn)
print("Accuracy Score of KNN Model: ",round(acc_knn,2))
```

```
Accuracy Score of KNN Model:  0.94
```

4. Random Forest :

Random forest is an ensemble learning algorithm that combines multiple decision trees to make predictions. Each tree is trained on a random subset of the data and features, reducing the risk of overfitting. The final prediction is determined by aggregating the predictions of individual trees. Random forest is effective in handling complex relationships and can provide accurate results for classification tasks.

By employing an ensemble of decision trees, random forest leveraged features such as battery power, memory, camera specifications, and connectivity options to accurately classify mobile devices into the respective price range categories. The algorithm's ability to handle complex relationships and reduce overfitting made it a suitable choice for achieving robust and accurate predictions in this mobile price prediction task.

This achieved an impressive accuracy score of 0.8583, indicating its strong performance in accurately predicting the price range of mobile devices.

```
acc_rf = accuracy_score(y_test, y_pred_rf)
print("Accuracy Score of Random Forest: ",acc_rf)
```

```
Accuracy Score of Random Forest:  0.8583333333333333
```

5. Decision Tree :

A decision tree is a supervised learning algorithm that uses a tree-like structure to make predictions. It partitions the data based on feature values and creates decision rules at each node to classify instances. It is capable of handling both categorical and numerical features, and its simplicity and interpretability

make it useful for understanding the decision-making process in classification tasks.

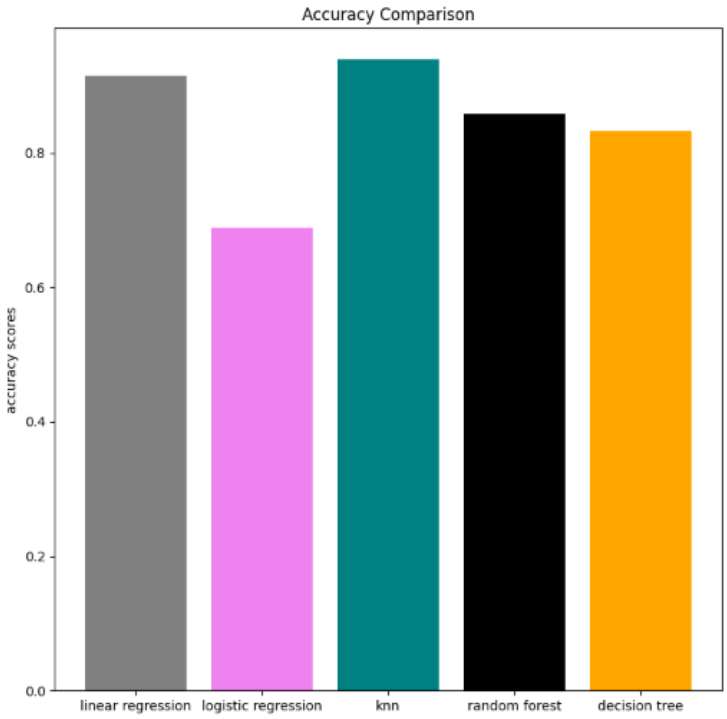
The decision tree algorithm divided the data based on features such as battery power, memory, camera specifications, and connectivity options to create decision rules for precisely classifying mobile devices into their respective price range categories. Decision trees are useful tools for understanding the variables influencing mobile device pricing because of their clarity and interpretability. The decision tree algorithm achieved an accuracy score of 0.83.

```
acc_dt = metrics.accuracy_score(y_test, y_pred_dt)
print('Accuracy of Decision Tree: ',round(acc_dt,2))
```

```
Accuracy of Decision Tree:  0.83
```

CONCLUSION

The KNN model showed the highest accuracy among the models evaluated, closely followed by the random forest model. Both models showed promise in accurately predicting mobile device price ranges for our dataset. The linear regression model performed well in explaining the variance, while logistic regression and decision tree models achieved moderate accuracy. Based on these findings, the KNN and random forest models are recommended for more accurate predictions of mobile device price ranges, offering valuable insights for pricing strategies and market analysis in the mobile industry.



FUTURE SCOPE

1. **Online Price Monitoring and Dynamic Updates**

One potential area for expansion is the development of a real-time price monitoring system that gathers ongoing data on mobile device prices. This system would be designed to continuously update the predictive models, ensuring they reflect the most current information. By incorporating real-time data and adapting to evolving market dynamics, the models can deliver more precise and timely predictions, enhancing the accuracy of price range estimations for mobile devices.

2. **Deployment as a Mobile App or Web Service**

Another potential avenue for development involves packaging the predictive models as either a mobile application or a web service. This would enable users to input mobile device features and obtain accurate price range predictions. Such a tool would prove invaluable to consumers, retailers, and online marketplaces, empowering them to make well-informed pricing and purchasing decisions.

Google Colab Link: **Click Here!**