## A Capstone Project Report on

MOBILE PRICE PREDICTION USING MACHINE LEARNING

Submitted by

VENIS C

## ABSTRACT

To predict “If the mobile with given features will be Economical or Expensive” is the main motive of this project. Real Dataset is collected by using Web scraping. Web scraping is the process of collecting and parsing raw data from the Web. Different feature selection algorithms are used to identify and remove less important and redundant features and have minimum computational complexity. Different Regressors are used to achieve as higher accuracy as possible. Results are compared in terms of highest accuracy achieved and minimum features selected. Conclusion is made on the base of best feature selection algorithm and best Regressor for the given dataset. This work can be used in any type of marketing and business to find optimal product(with minimum cost and maximum features). Future work is suggested to extend this research and find more sophisticated solution to the given problem and more accurate tool for price estimation.

Dataset Link: <https://github.com/venis-c/CAPSTONE-PROJECT/blob/main/mobile%20data.csv>

Source Code: https://github.com/venis-c/CAPSTONE-PROJECT

## ACKNOWLEDGEMENTS

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of my capstone project. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

Further, I have fortunate to have Mr. PRASAD as my mentor. He has readily shared his immense knowledge in data science and guides me in a manner that the outcome resulted in enhancing my data skills.

I certify that the work done by me for conceptualizing and completing this project is original and authentic.

Date: July 10, 2022 Name: VENIS C

## CERTIFICATE OF COMPLETION

I hereby certify that the project titled “ Mobile Price Prediction Using Machine Learning ” was undertaken and completed the project (10th July, 2022).

Mentor : Mr. Prasad

Date : 10th July,2022

Place : Karur

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER**  **NO.** | **TITLE** | **PAGE NO.** |
| **1**  **2**  **2.1**  **2.2**  **2.3**  **2.4**  **2.5**  **2.6**  **2.7**  **3**  **4**  **4.1**  **4.2**    **4.3**  **5**  **5.1**  **5.2**  **5.3**  **5.4**  **6**  **7** | **ABSTRACT**  **ACKNOWLEDGEMENT**  **INTRODUCTION**  **DATA COLLECTION AND PREPARATION**  **DATA OVERVIEW**  **DATA PREPROCESSING**  **HANDLING MISSING VALUES**  **DATA TYPE CHANGING**  **EXPLORATORY DATA ANALYSIS**  **FEATURE OBSERVATION**  **FEATURE SELECTION**  **TEST VALIDATION SPLIT**  **FITTING MODEL TO DATA**  **DECISION TREE REGRESSOR**  **RANDOM FOREST REGRESSOR**    **MODEL PERFORMANCE**  **STREAMLIT CONNECTIVITY**  **STREAMLIT**  **STREAMLIT WEB APPLICATION**  **MOBILE PRICE PREDICTOR**  **OUTPUT**  **CONCLUSION**  **REFERENCE** | **2**  **3**  **9**  **11**  **11**  **13**  **14**  **15**  **16**  **17**  **18**  **19**  **20**  **20**  **21**  **23**  **24**  **24**  **25**  **27**  **28**  **29**  **30** |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE**  **NO.** | **TITLE** | **PAGE NO.** |
| **1.1**  **2.1**  **2.2**  **2.3**  **2.4**  **2.5**  **2.6**  **2.7**  **2.8**  **2.9**  **3.1**  **4.1**  **4.2**  **4.3**  **4.4**    **4.5**  **4.6**  **5.1**  **5.2**  **5.3**  **5.4**  **5.5**  **5.6** | **Mobiles**  **Mobile Details Dataset**  **Missing Values Analysis Missing Values Handling**  **data types**  **data types after conversion**  **information**  **description**  **correlation heatmap**  **feature selection**  **train validation split**  **decision tree**  **decision tree regressor model**  **random forest**  **random forest regressor model**    **random forest accuracy**  **decision tree accuracy**  **pickle files**  **import pickle files**  **streamlit code**  **run streamlit code**  **web application**  **prediction** | **9**  **11**  **13**  **14**  **15**  **15**  **16**  **16**  **17**  **18**  **19**  **20**  **21**  **22**  **22**  **23**  **23**  **24**  **25**  **26**  **26**  **27**  **28** |

## CHAPTER 1 INTRODUCTION

Price is the most effective attribute of marketing and business. The very first question of costumer is about the price of items. All the costumers are first worried and thinks “If he would be able to purchase something with given specifications or not”. So to estimate price at home is the basic purpose of the work. This paper is only the first step toward the above mentioned destination.



Figure 1.1: Mobiles

Mobile now a days is one of the most selling and purchasing device. Every day new mobiles with new version and more features are launched. Hundreds and thousands of mobile are sold and purchased on daily basis. So here the mobile price\_class prediction is a case study for the given type of problem i.e finding optimal product. The same work can be done to estimate real price of all products like cars, bikes etc

Many features are very important to be considered to estimate price of mobile. For example Processor of the mobile. Battery timing is also very important in todays busy schedule of human being. Size and thickness of the mobile are also important decision factors. Internal memory, Camera pixels, and video quality must be under consideration. Internet browsing is also one of the most important constraints in this technological era of 21st century. And so is the list of many features based upon those, mobile price is decided. So we will use many of above mentioned features to classify whether the mobile would be very\_economical, economical, expensive or very\_ expensive

## CHAPTER 2

**DATA COLLECTION AND DATA PREPARATION**

In this dataset I wasn’t downloading from Kaggle or any other data collecting websites. I just make or create the dataset using one of the web scrapping tools. So a little bit of overview we understand about the data and its features. Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites.Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser.

# ****2.1 DATA OVERVIEW****

# 

Figure 2.1: Mobile Details Dataset

* **Brand me** — This is first feature of our dataset. Its Denotes name of the mobile phones and Brands.
* **Ratings** — This Feature Denotes Number of the ratings gave by the consumers for each mobile.
* **RAM** — It’s have RAM size of the phone.
* **ROM** — It’s have ROM **(Internal Memory)** size of the phone.
* **Mobile\_Size** — It’s represents how many inches of the particular mobile phone have. Here all the values are gave in **inches**
* **Primary\_Cam**— It’s Denotes Number of the pixels of the primary camera **(Back Camera)** for each mobile.
* **Selfi\_Cam** — It’s Denotes Number of the pixels of the Selfi camera **(Front Camera)** for each mobile.
* **Battery\_Power** — It’s Denotes amount of the battery power in each mobiles in **mAh**.
* **Price** — It’s a Dependent Feature of the dataset. It’s just denoting prices of the each mobile.

# ****2.2 DATA PREPROCESSING****

Data preprocessing is an important step in the data mining process. The phrase “garbage in, garbage out” is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc.

In this project you might be performing lot of preprocessing steps. Because in this dataset is not downloaded from Kaggle or any other data source website. This data retrieve from E-Commerce website. But after I was get the dataset I was make a dataset for model prediction. So you need not to and data preprocessing steps except handling the missing values.

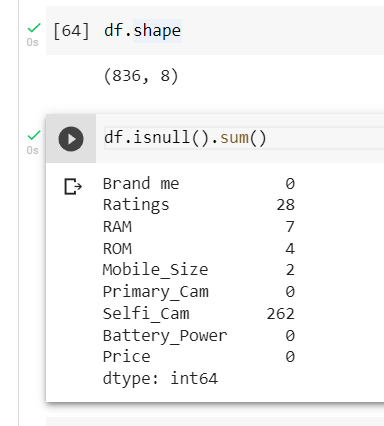


Figure 2.2 Missing Values Analysis

# ****2.3 HANDLING THE MISSING VALUES****

# 

# Figure 2.3 Missing Values Handling

We need not to have the **Brand me**feature for prediction because it just a mobile name. So we should drop the first column.

After handling all the null or missing values will be look like Fig 2.3.

# ****2.4 DATA TYPES CHANGING****

# 

# Figure 2.4 data types

# Here some of the data types are floating point values. We need to change the into integer values except Rating feature. After changing the data types dataset and data types will be look like Fig 2.3.

# 

# Figure 2.5 data type after conversion

# ****2.5 EXPLORATORY DATA ANALYSIS****

# In statistics, exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

# 

# Figure 2.6 information

# 

# Figure 2.7 description

# ****2.6 FEATURE OBSERVATION****

# First Understanding the correlation of features between target and other features.

# 

# Figure 2.8 correlation heatmap

# ****2.7 FEATURE SELECTION****

# **Feature Selection** is the process where you automatically or manually **select** those **features** which contribute most to your prediction **variable** or output in which you are interested in. Having irrelevant **features** in your data can decrease the accuracy of the models and Make your model learn based on irrelevant **features**.

# 

# Figure 2.9 Featue selection

## CHAPTER 3

**TRAIN VALIDATION SPLIT**

The dataset was divided into two parts for traing and testing purposes.here we are taking 25% of the dataset for testing and the balance 75% of the dataset for training the model.

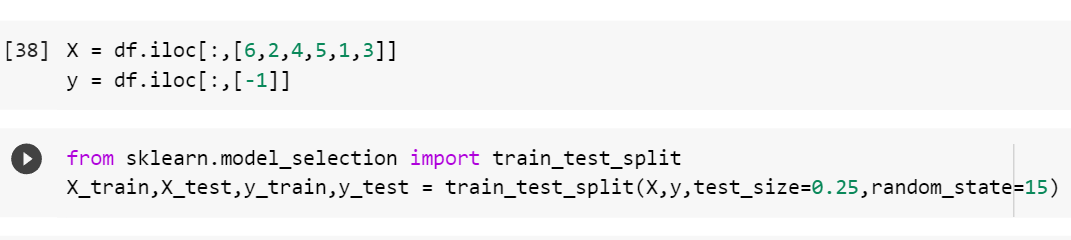


Figure 3.1: Test Validation Split

## CHAPTER 4

**FITTING MODELS TO DATA**

# 4.1 DECISION TREE ****REGRESSOR****

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

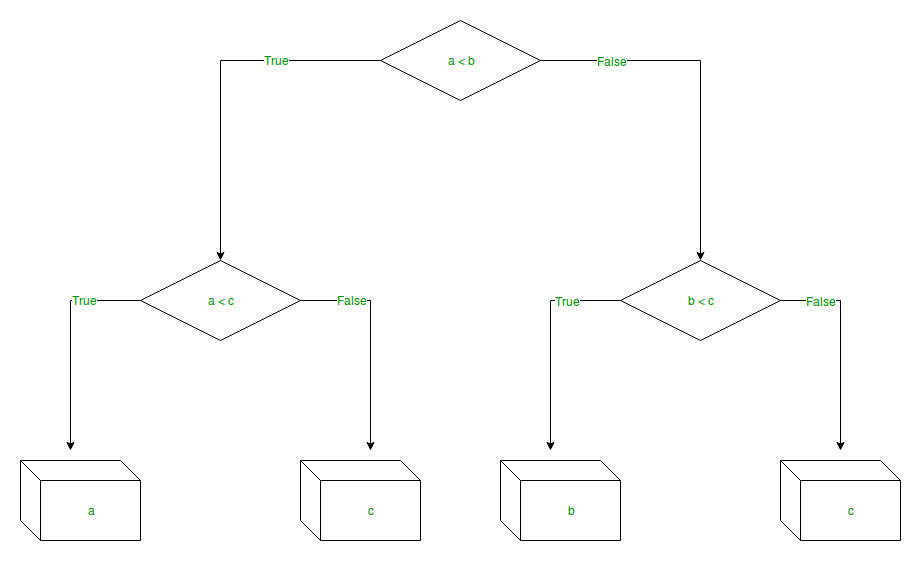


Figure 4.1: Decision Tree

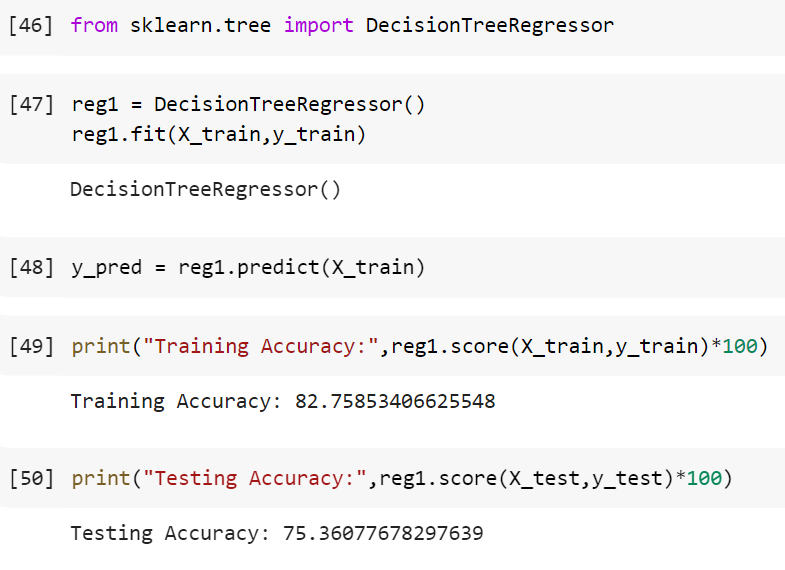


Figure 4.2 Decision Tree Regressor model

# ****4.2 RANDOM FOREST REGRESSOR****

# Every decision tree has high variance, but when we combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that particular sample data, and hence the output doesn’t depend on one decision tree but on multiple decision trees. In the case of a classification problem, the final output is taken by using the majority voting classifier. In the case of a regression problem, the final output is the mean of all the outputs. This part is called **Aggregation**.

# 

# Figure 4.3 Random Forest

# 

Figure 4.4 Random Forest Regressor model

# ****4.3 MODEL PERFORMANCE****

# By analysing the Decision Tree Regressor and Random Forest Regressor performance the Decision Tree algorithms are gave more prediction accuracy than Random Forest. So we can choose Decision Tree for predicitions.

# Random Forest Regressor Accuracy

# 

# Figure 4.5 Random Forest Accuracy

# Decision Tree Regressor Accuracy

# 

# Figure 4.6 Decision Tree Accuracy

## CHAPTER 5 STREAMLIT CONNECTIVITY

**5.1 STREAMLIT**

Streamlit is an open source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc. With Streamlit, no callbacks are needed since widgets are treated as variables. Data caching simplifies and speeds up computation pipelines. Streamlit watches for changes on updates of the linked Git repository and the application will be deployed automatically in the shared link.



Figure 5.1 pickle files

**5.2 STREAMLIT WEB APPLICATION**

The next step is to upload the pickle file to the pycharm. And the write a code to get the input from the user.other style like a background images, page title and Text boxes to get the input from the user. Convert the input into array and then fed the input to the model.

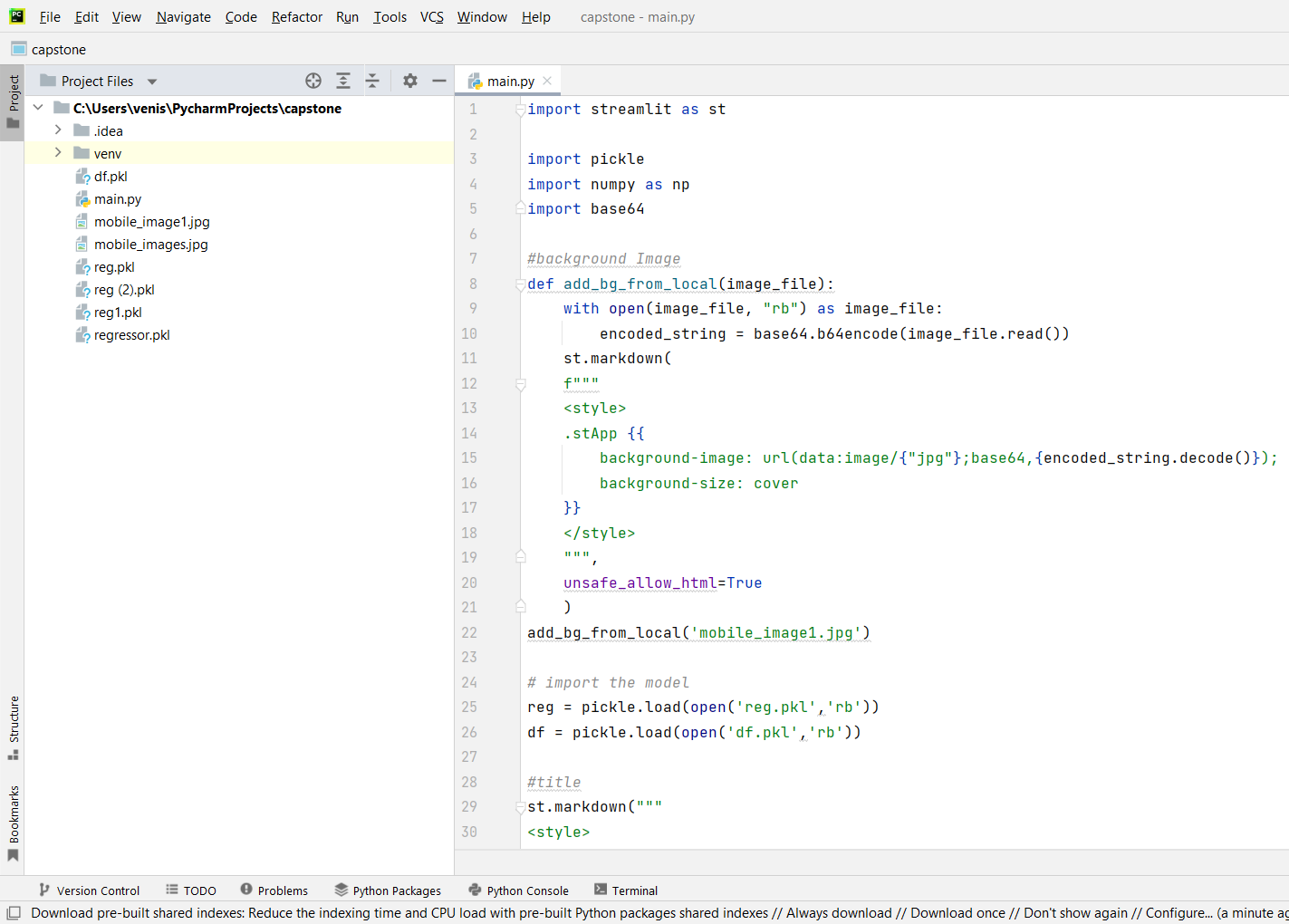


Figure 5.2 import pickle file

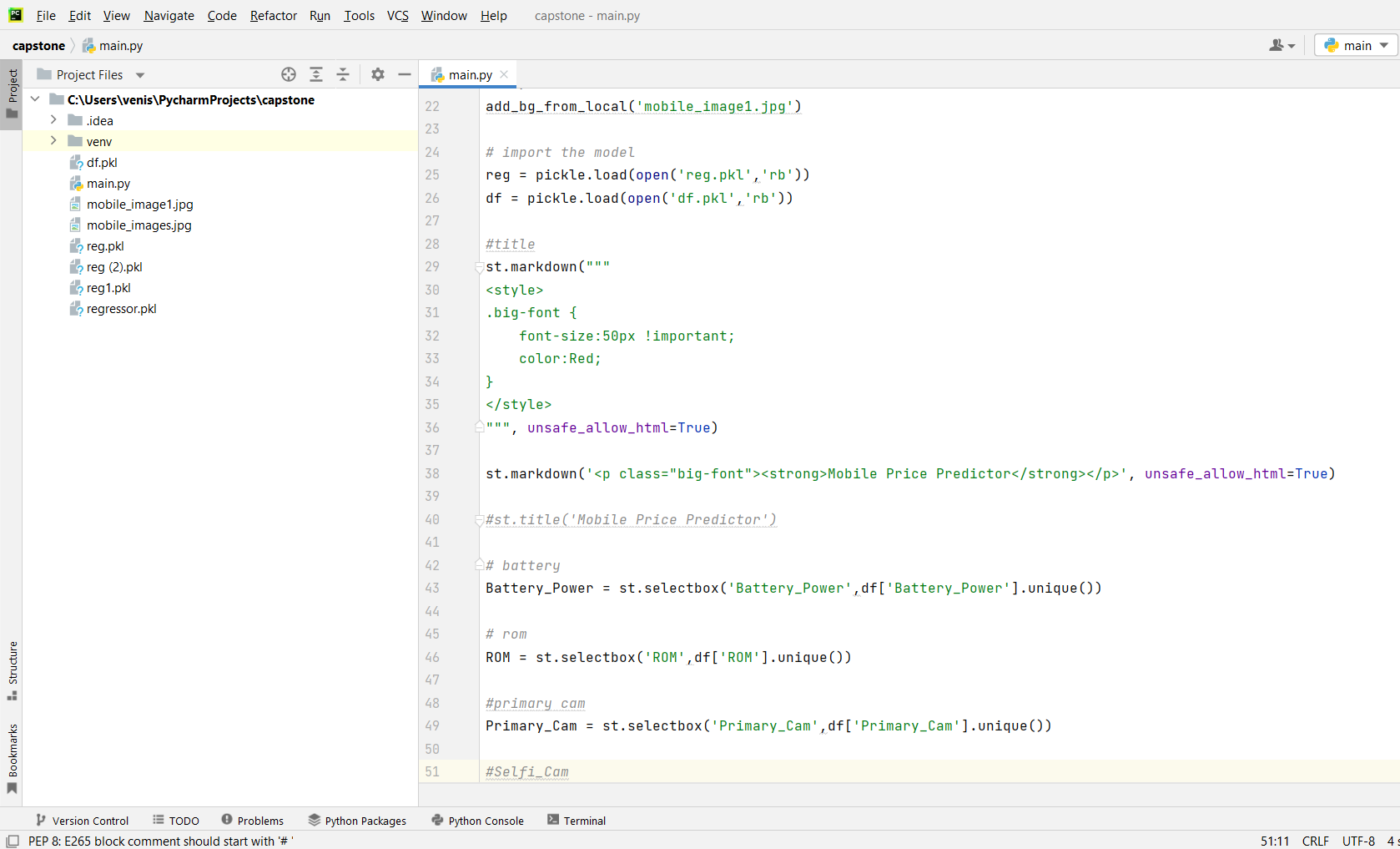


Figure 5.3 Streamlit code

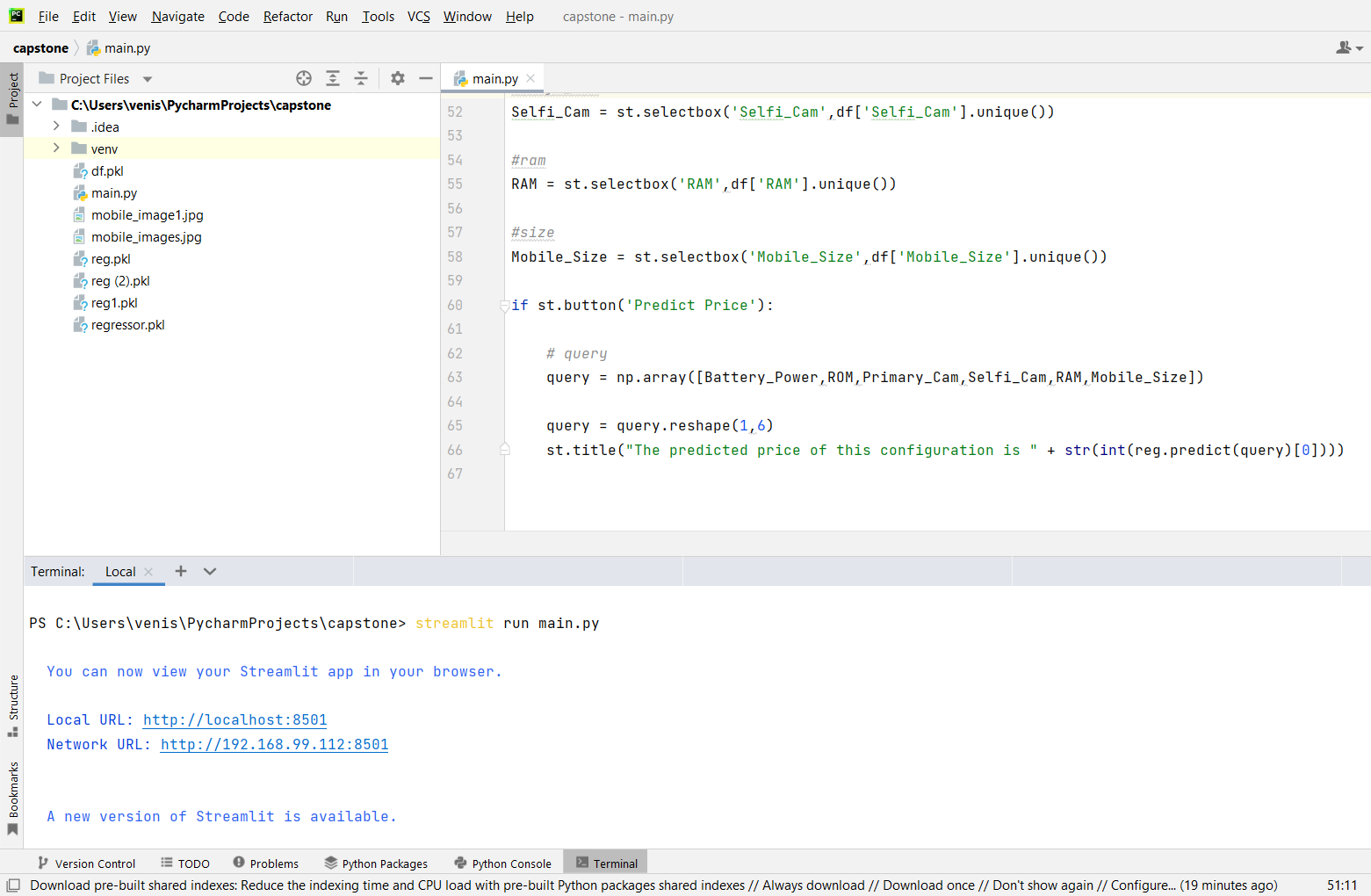


Figure 5.4 Run streamlit code

## 5.3 MOBILE PRICE PREDICTOR

This is how the mobile price predictor web application look.In this page the user need to Enter the details for which features of mobile to know the price of the mobile like a Battery\_power, ROM, Primary camera, Selfie camera, RAM, Mobile size.

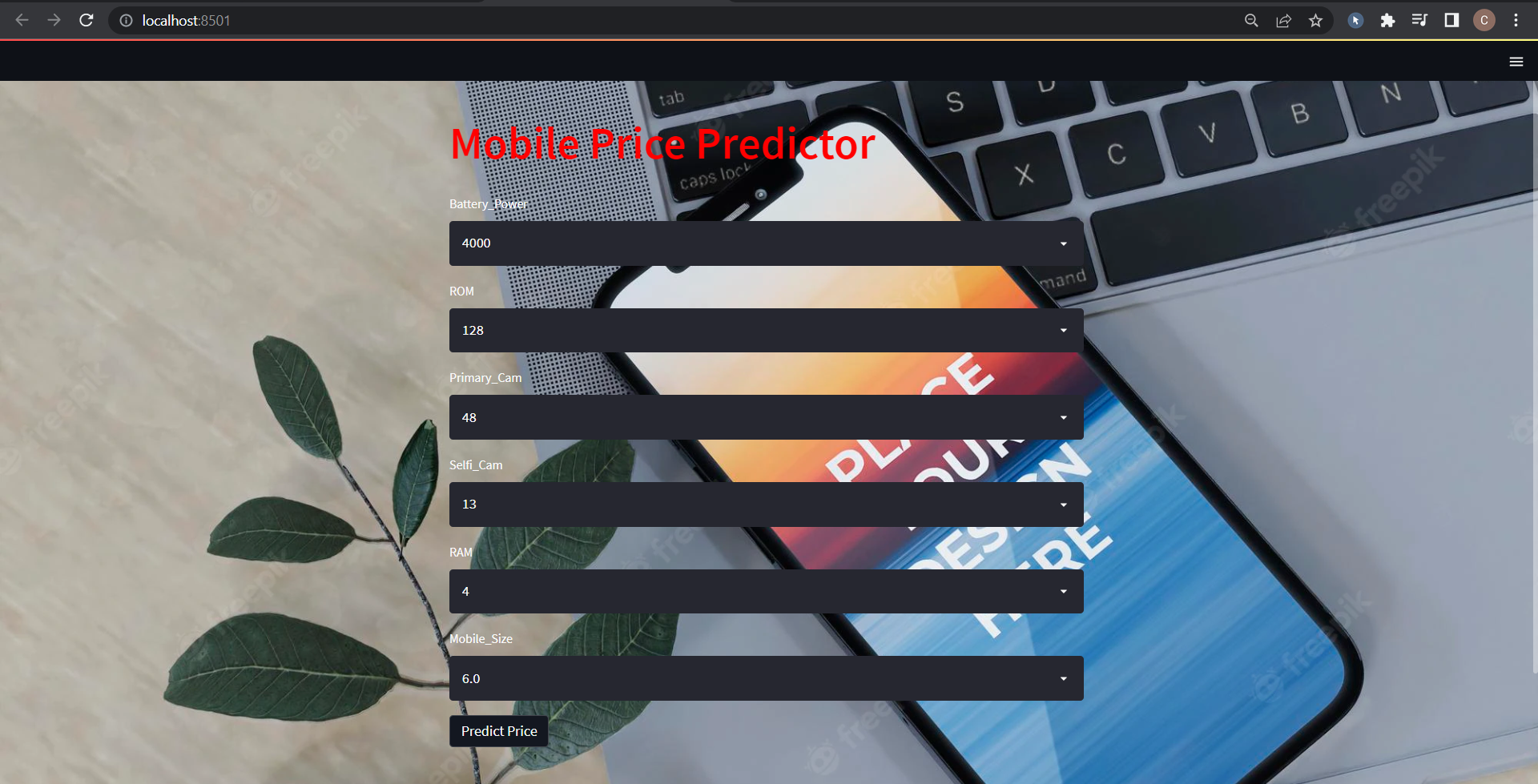
****

Figure 5.5 Web application

## 5.4 OUTPUT

## After entered the details of the mobile phone we need to click the “Predict Price” button. The input is feded to the model and the model’s output will be displayed in the web application.



Figure 5.6 Prediction

Network URL For the Application : <http://127.0.0.1:8501>

## CHAPTER 6 CONCLUSION

I have evaluated two kinds of machine learning algorithm for mobile price prediction based on the features of the mobile, Decision Tree and Random Forest. In my project, I have collect the data’s from the website using web scraping technique. I do some important processes like a handling missing values, data type conversion, find the correlation between the price and other variables and EDA process. After that I do the feature selection process to find the important features which is used to define the price of the mobile. And spliting the dataset into two parts 75% for traing and 25% for testing. Atlast fit the dataset to the model and Evaluate the performance of the model. Finally the Decision tree give the more accuracy while compare to the Random forest. So by using the Decision tree regressor model I was created the streamlit web application for the user experience. In future More sophisticated artificial intelligence techniques can be used to maximized the accuracy and predict the accurate price of the products.Software or Mobile app can be developed that will predict the market price of any new launched product.

## CHAPTER 7 REFERENCE

[1] Sameerchand Pudaruth . “Predicting the Price of Used Cars using Machine Learning Techniques”, International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 7 (2014), pp. 753-764

[2] Mariana Listiani , 2009. “Support Vector Regression Analysis for Price Prediction in a Car Leasing Application”. Master Thesis. Hamburg University of Technology.

[3] Kanwal Noor and Sadaqat Jan, “Vehicle Price Prediction System using Machine Learning Techniques” , International Journal of Computer Applications (0975 – 8887) Volume 167 – No.9, June 2017.s