MOBILE PRICE PREDICTION

Final Project Technical Report

Submitted by (GROUP #3)
MAMATHA TIPPANI
KOTESWAR ENAMADNI
TEJA REDDY SOMA
MAMATHA PAPALETI

Submitted for DSCI-6001 (FALL 2023)
DR. ARDIANA SULA

CONTENTS

1. INTRODUCTION

Mobile price prediction using python is an application that leverages the power of data to estimate the price of a Mobile phone based on various attributes. This can be immensely helpful for both consumers and businesses alike. For consumers, it aids them in making informed purchasing decisions by assessing whether a given mobile phone's price is justified by its specifications. Business entities, particularly those in the Mobile phone industry, can make use of it to guide pricing strategies and position themselves in the market.

2. Overview

When predicting the price of a mobile phone, several key features should be taken into consideration. The brand of the phone often significantly impacts its price, as does the amount of memory (RAM) and internal storage capacity. The size and resolution of the screen are also important factors, with larger, higher resolution screens typically found on more expensive models. The quality of both the front and rear cameras can affect price, as can the capacity of the battery. The type and speed of the phone’s processor, the operating system and its version, and connectivity features such as 4G, 5G, Wi-Fi, and Bluetooth versions are also relevant. Additional features, such as water resistance, fingerprint scanners, and face recognition, may also contribute to the cost. It’s important to note that the significance of these features may vary depending on the market segment being targeted. A feature important analysis can provide further insight into which features are most influential in driving price predictions.

3. BACKGROUND THEORY

Mobile price prediction is a valuable tool in the rapidly evolving smartphone market. With a multitude of brands, models, and features available, determining a fair price for a mobile phone can be a complex task. Accurate price prediction models can provide significant benefits to various stakeholders:

1. **Consumers**: An accurate price prediction model can help consumers make informed purchasing decisions. By understanding the factors that contribute to a phone's price, consumers can ensure they are getting good value for their money.
2. **Retailers and Resellers**: For businesses that sell mobile phones, price prediction models can aid in inventory management and pricing strategies. By predicting the price of new models or how the price of existing models may change over time, businesses can optimize their purchasing and pricing decisions to maximize profit.
3. **Manufacturers**: Mobile phone manufacturers can use price prediction models to inform their product development process. Understanding which features consumers value most can help manufacturers design phones that meet consumer demand and command higher prices in the market.
4. **Second-hand Market**: In the second-hand market, price prediction models can help sellers price their used phones appropriately and help buyers identify potential bargains.

In summary, mobile price prediction not only contributes to market transparency but also enables optimization of buying and selling strategies, ultimately benefiting both consumers and businesses.

4. METHODOLOGY

4.1. Data collection

The dataset used for the research was taken from kaggle. This dataset contains 1000 instances with 19 features. The target variable was Mobile price range (0 for Low, 1 for middle range, 2 for budget and 3 for expensive).

In total there are 1000 rows x 19 columns, which are:

- Battery
- Clock Speed
- Dual sim
- Four G
- Internal memory
- RAM
- No. of Cores
- Touch screen
- WiFi

4.2. CRISP DM

The approach to our mobile price prediction project using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is outlined here.

1. **Business Understanding**: This initial phase focuses on understanding the project objectives and requirements from a business perspective. For a mobile price prediction project, the objective is to build a model that accurately predicts the price of a mobile phone based on its features.
2. **Data Understanding**: In this phase, we gathered data relevant to the project. This involved collecting data on mobile phone prices and their features, such as brand, screen size, camera quality, battery life, etc. We worked on improving and understanding the quality of this data.

3. **Data Preparation**: This phase involves cleaning the data and transforming it into a format suitable for modeling. This involved handling missing values, dealing with outliers, encoding categorical variables, normalizing numerical variables, etc.
4. **Modeling**: In this phase, various modeling techniques are selected and applied. For a price prediction task, regression models like linear regression, decision tree regression, random forest regression, or even deep learning models could be used. We trained these models on the prepared dataset.
5. **Evaluation**: After building the models, we evaluated their performance. This involved using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) to quantify the difference between the predicted and actual mobile phone prices.
6. **Deployment**: Once a satisfactory model has been built and evaluated, we deployed it for use in a production environment. This could involve integrating the model into a mobile app or website where users can input a phone's features and receive a predicted price.

4.3. Machine Learning Algorithms for Classification

Machine learning algorithms enable us to analyze large datasets, identify patterns, and make predictions about potential mobile prices. They also help us detect potential price trends and early signs of market shifts.

The importance of knowing the price of a product before buying it cannot be overstated. It allows consumers to make informed purchasing decisions and ensures they get value for their money. For businesses, it aids in developing effective pricing strategies and staying competitive in the market.

The most used machine learning algorithms for predictions are K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Support Vector Machines (SVM).

4.3.1. K-Nearest Neighbors Classifier

KNN is a non-parametric algorithm that identifies a group of data points that are like the data point being analyzed. It uses a distance measure such as Euclidean to calculate the distance between each data point and the query point. KNN can be used to classify data points into different classes of price ranges.

4.3.2. Random Forest Classifier

Random Forest is an ensemble learning algorithm that uses multiple decision trees to make a prediction.

It can be used for both classification and regression. It harnesses the power of multiple decision trees to make decisions. Because it's a forest of randomly generated decision trees. Each node in the decision tree operates on a random subset of features to compute the output. Random Forest then combines the outputs of the individual decision trees to produce the final output.

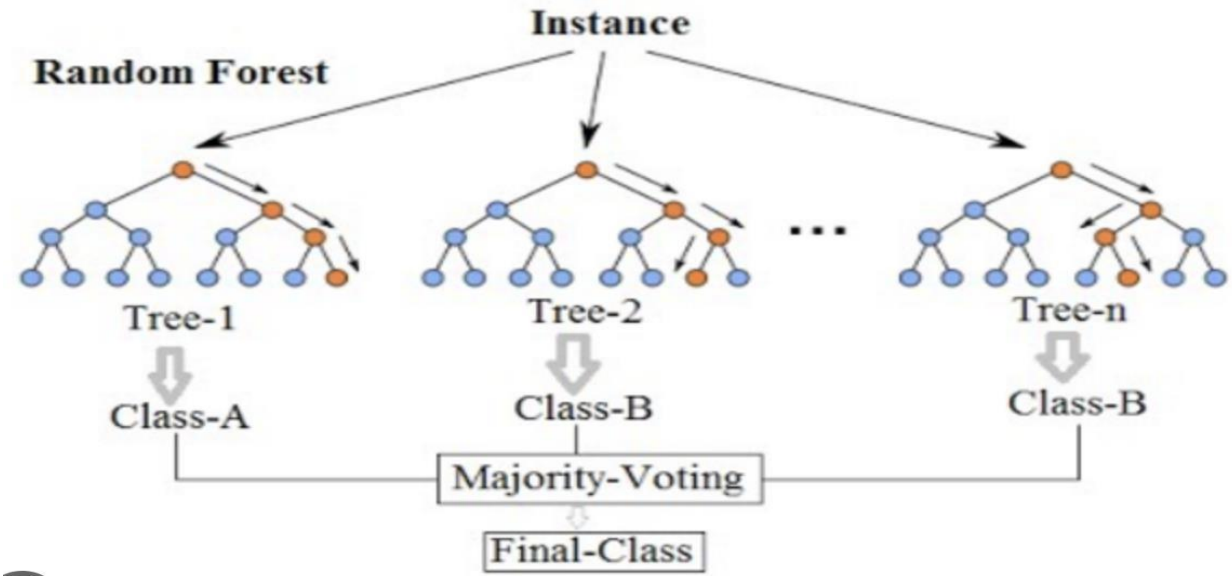

Fig. 4.2.2: Random Forest

4.3.3. Logistic Regression

Logistic regression is a supervised learning algorithm that uses a linear combination of independent variables to estimate the probability of a certain outcome. It can be used to identify factors that are associated with the development of liver disease.

This classification algorithm is used to predict the probability of categorical dependent variables. In logistic regression, the dependent variable is a binary variable with data encoded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.

$f(\mathbf{x})$: $p(\mathbf{x}) = 1 / (1 + \exp(-f(\mathbf{x}))$

4.3.4. Support Vector Machines Classifier (SVC)

SVC is a nonparametric clustering algorithm that makes no assumptions about the number or shape of clusters in the data. Support Vector Classifier is a supervised machine learning algorithm typically used for classification tasks. SVC works by mapping data points into a high-dimensional space and finding the best hyperplane that splits the data into two classes. It uses a kernel function to map data points into a higher dimensional space.

4.3.5. Summary

As a team of data science experts, we utilize machine learning algorithms to analyze, clean, and predict mobile phone prices. By employing these algorithms, we can assist both consumers and companies in more accurately determining the potential price range of mobile phones based on their features. This not only aids in making

informed purchasing decisions but also supports manufacturers and retailers in developing effective pricing strategies. Furthermore, the insights gained from these predictive models can guide the design and development of new mobile phones, ensuring they meet market demand and provide value for money.

5. EXPLORATORY DATA ANALYSIS

5.1. Distribution of RAM and price range

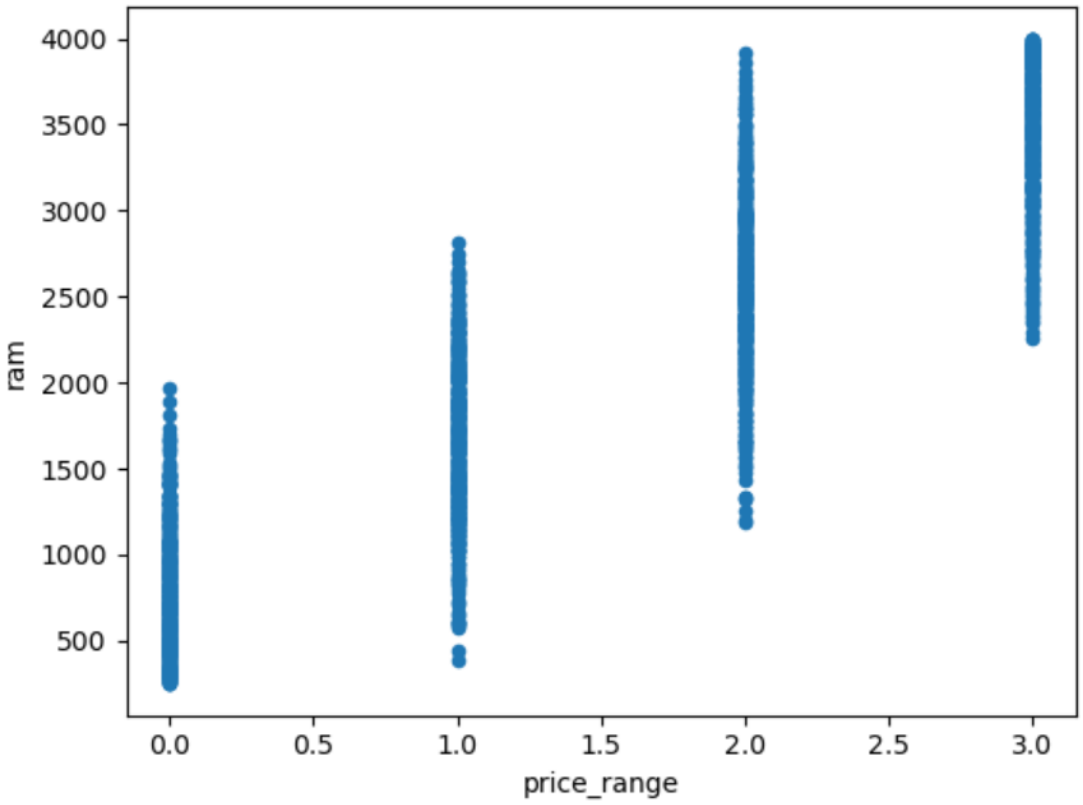

Fig. 5.1: Distribution of RAM among patients

5.2. Distribution of Battery Power

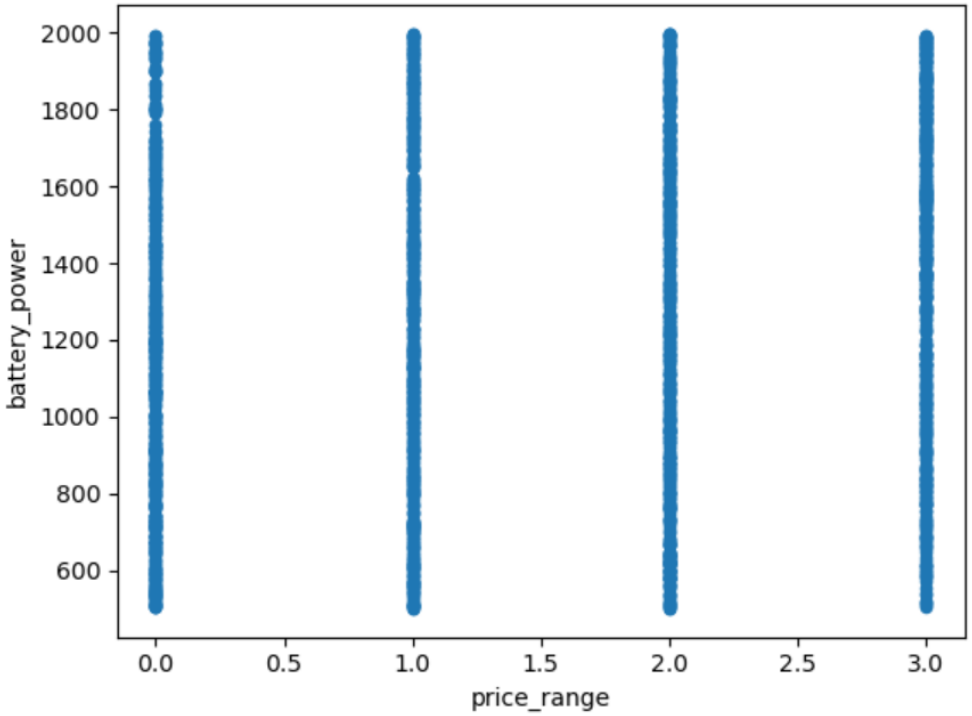

Fig. 5.2: Distribution of Battery Power

5.3. No of cores

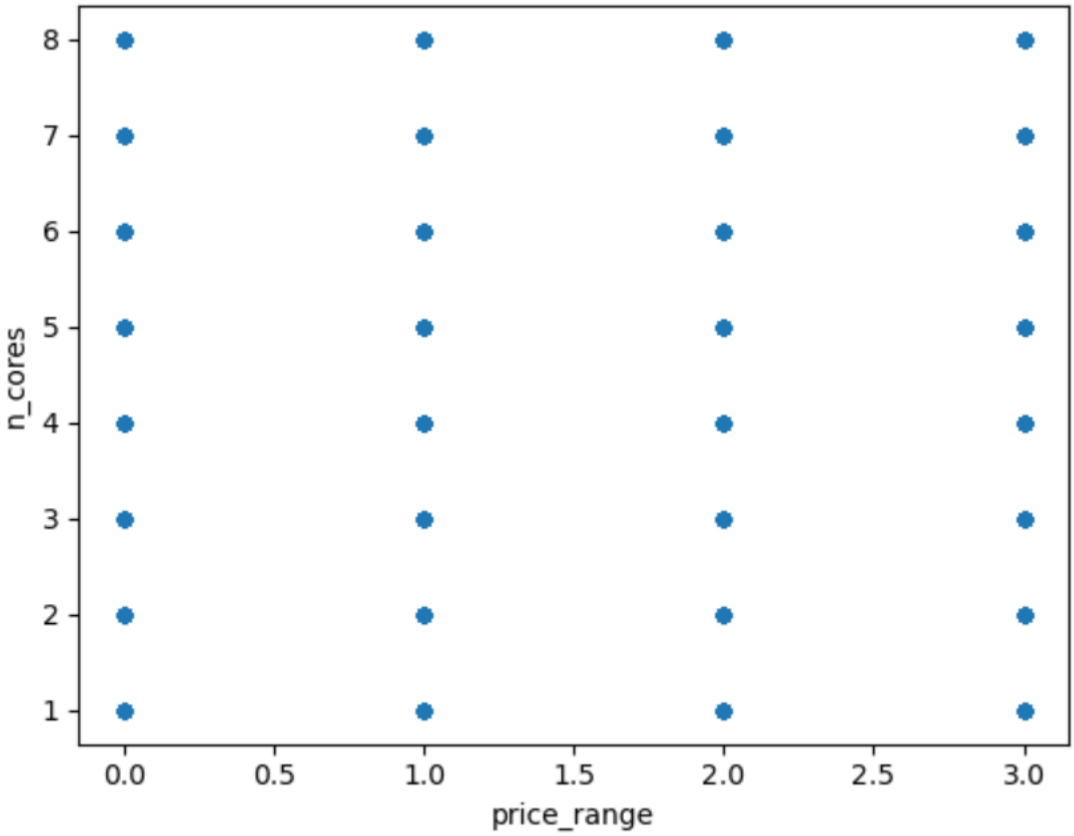

Fig. 5.3: No of cores

5.4. Correlation heatmap

The heatmap shown below appears to have some correlated parameters.

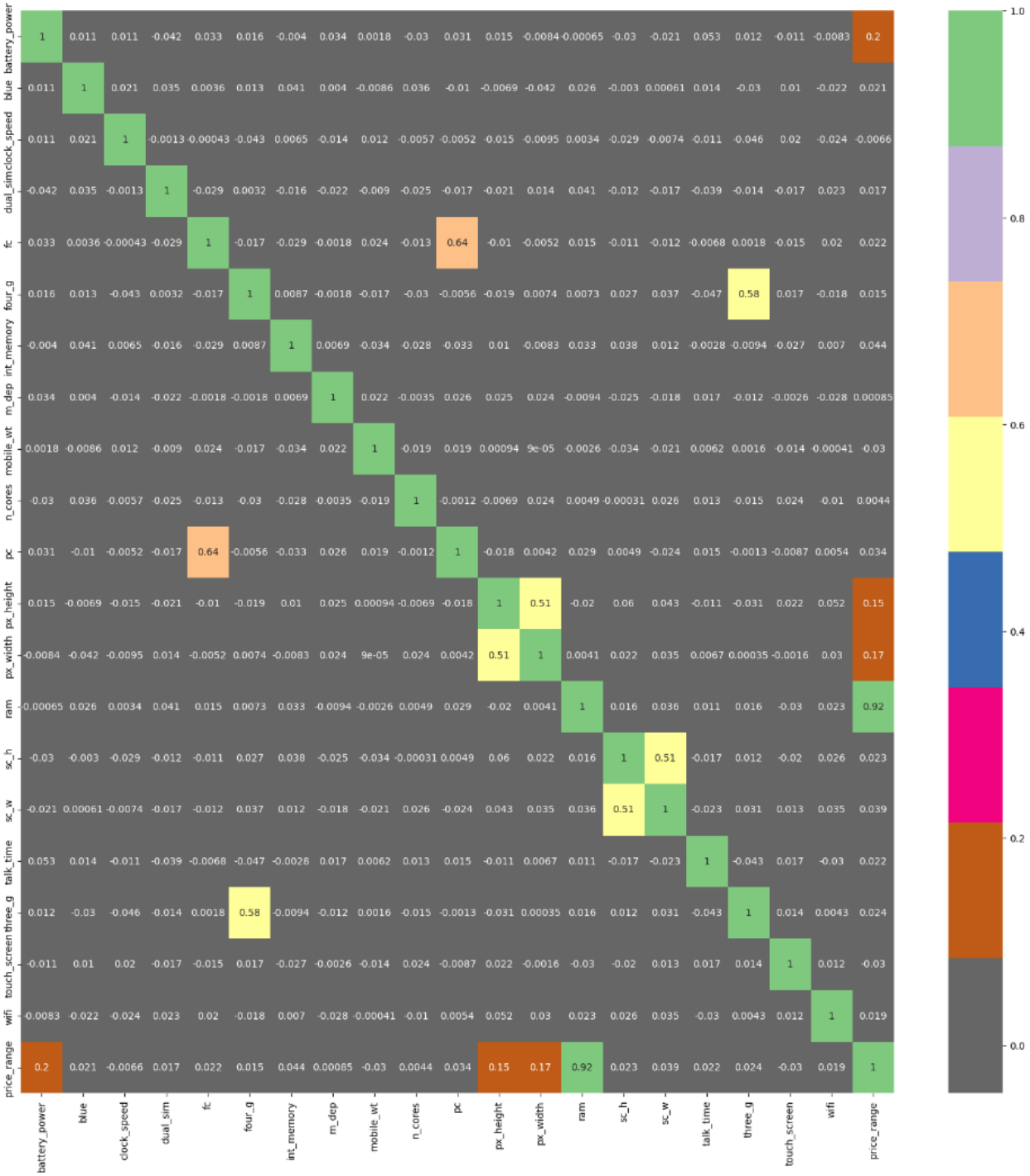

Fig. 5.4: Correlation between the features using a heatmap.

6. RESULTS

The Logistic Regression model has been found to be the most effective for live disease prediction. Its superior performance over other algorithms such as KNN, random forest, indicates that logistic Regression is the best choice for mobile price prediction. Its ability to efficiently handle large datasets, its robustness to noise, and its high accuracy make it an ideal choice for such applications. With Logistic regression, we can make more accurate predictions in real time, which can help in finding the best mobile in budget and this helps the companies in improving customer satisfaction and increasing productivity.

7. APPLICATION DEPLOYMENT

The demo application has been deployed using Flask on localhost:

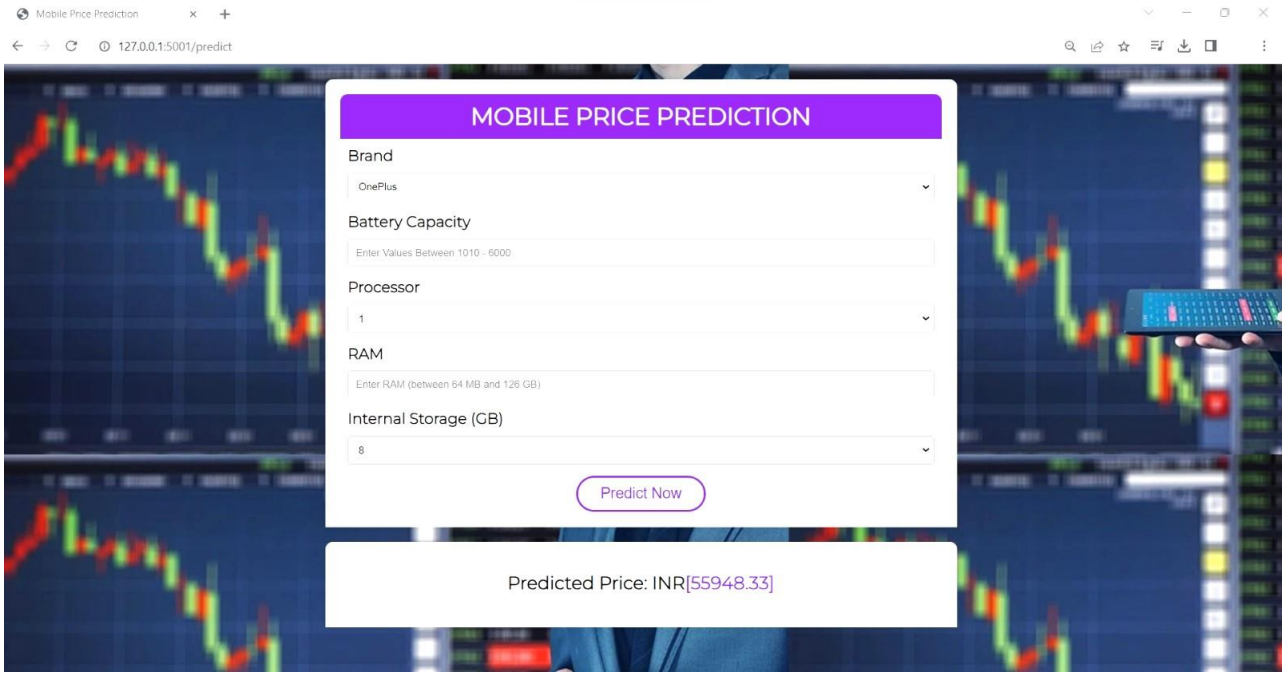

Fig.7.1. Application Demo on Microsoft Edge

The code is hosted on GitHub:

[koteswar-e/Data-Science-midterm_Project_1 (github.com)]

8. CONCLUSION

In conclusion, machine learning algorithms play a significant role in the field of price prediction. They are becoming increasingly important in analyzing large datasets, identifying patterns, and making predictions about potential mobile prices. These algorithms are helping to uncover new insights into the factors influencing mobile prices and potential price trends. As such, the use of machine learning algorithms in this field is likely to increase in the future and will continue to play a key role in improving our understanding of this complex market landscape.

9. FUTURE IMPROVEMENTS

Collaborating with industry professionals can provide valuable insights into the features that are most relevant for predicting mobile phone prices. Their on-site feedback can help refine the prediction algorithm, making it more accurate and reliable.

Collecting more data on mobile prices, as well as additional information on factors such as battery consumption and mobile application usage, can further enhance your model. Understanding how different applications and practices affect battery health could also be a significant factor in predicting mobile prices, as battery life is a key consideration for many consumers.

This approach not only allows for continuous improvement of the model but also ensures that it stays relevant as new trends emerge in the mobile phone industry.

10. REFERENCES

[1] https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[2] Machine Learning Algorithms for Prediction of Mobile Phone Prices | SpringerLink

[3] Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization | IEEE Conference Publication | IEEE Xplore

[4] Prediction of Phone Prices Using Machine Learning Techniques | SpringerLink