PROJECT REPORT

**Smart Phone Price Prediction**

*Submitted towards the partial fulfillment of the criteria for award of PGA by Imarticus*

*Submitted By: Sagar Gupta*

*Course and Batch: PGA 32*



# Abstract

The aim of this research is to develop a model to predict the price of a mobile when the specifications of a mobile are given and to find the ML algorithm that predicts the price most accurately. The usage of archival data to accurately forecast forthcoming instances is the essence of Predictive Analytics. One of the ways Predictive Analytics can be performed is by using Machine Learning.

**Keywords**

1. Price
2. Prediction
3. Variable
4. Specifications
5. Visualization

# Acknowledgements

I am using this opportunity to express my gratitude to everyone who supported us throughout the course of this project. I 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 am fortunate to have Purushottam Sharma as our mentor. He has readily shared his immense knowledge in data analytics and guided us in a manner that the outcome resulted in enhancing our data skills.

I wish to thank, all the faculties, as this project utilized knowledge gained from every course that formed the PGA program.

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

Date: August 24, 2023 Sagar Gupta

Place: Delhi

# Certificate of Completion

I hereby certify that the project titled “Smart Phone Price Prediction” was undertaken and completed under my supervision by Sagar Gupta from the batch of PGA (32)

Mentor: Purushottam Sharma

Date: August 24, 2023

Place – Delhi

**Table of Contents**

[Abstract 2](#_bookmark0)

[Acknowledgements 3](#_bookmark1)

[Certificate of Completion 4](#_bookmark2)

[CHAPTER 1: INTRODUCTION](#_bookmark3) 7

* 1. [Title & Objective of the study](#_bookmark4) 7
  2. [Need of the Study 7](#_bookmark5)
  3. [Data under study 8](#_bookmark6)
  4. [Data Sources 9](#_bookmark8)
  5. [Tools & Techniques 10](#_bookmark9)

[CHAPTER 2: DATA PREPARATION AND UNDERSTANDING 11](#_bookmark10)

* 1. [Phase I – Data Extraction and Cleaning](#_bookmark11) 11
  2. [Phase II - Feature Engineering](#_bookmark12) 12
  3. [Data Dictionary](#_bookmark13) 12
  4. [Exploratory Data Analysis](#_bookmark14) 13

[CHAPTER 3: FITTING MODELS TO DATA](#_bookmark15) 14

* 1. [Linear Regression](#_bookmark16) 14
  2. [Decision Tree Regressor](#_bookmark17) 14
  3. Random Forest Regressor 14
  4. KNN 15
  5. XGBoost Regressor 15
  6. [INFRASTRUCTURE CHALLENGES](#_bookmark18) 15

[CHAPTER 4: KEY FINDINGS](#_bookmark19) 16

[Figure 4.1 Accuracy Table](#_bookmark20) 16

[CHAPTER 5: RECOMMENDATIONS AND CONCLUSION](#_bookmark21) 17

[CHAPTER 6: REFERENCES](#_bookmark22) 18

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Figure 2.1** | **Data** | **Page 12** |
| **Figure 4.1** | **MAE Table** | **Page 16** |

# CHAPTER 1: INTRODUCTION

## Title & Objective of the study

Title of the project is Smart Phone Price Prediction.

Mobile now a days is one of the most selling and purchasing device. Mobile phones come in all sorts of prices, features and specifications. Every day new mobiles with new version and more features are launched. Hundreds and thousands of mobile are sold and purchased on daily basis. Setting an optimal price before the release of a smartphone is imperative for any company. Deciding on the correct price of a smartphone is very important for its market success. A new smart phone that has to be launched, must have the correct price so that consumers find it appropriate to buy that smart phone.

Objective : Predictive Analytics - Build a ML model to predict the price of a smartphone for a company.

A tool that gives the estimated price of a smartphone after weighing in the features it provides can come in handy and can help the company in making an informed decision while setting its market price.

Such a tool can also be used by a consumer to get an estimated price based on the features they are looking for in the smartphone.

## Need of the Study

* + - **Strategic decision-making:** By understanding the factors that influence the price of a smartphone. It enables company to focus on opportunities with appropriate prices of mobile phones, leading to improved business outcomes.
    - **Resource management:** By accurately assessing the market success or demand of a smartphone, the company can allocate their resources, including human resources, time, and finances, more efficiently. It helps in optimizing resources.
    - **Sales Growth planning:** The study provides insights into the influence of specifications on the price of a smartphone as well as the influence of price on the demand and market success of the smartphone. This enables better planning and forecasting, facilitating sustainable and controlled growth.
    - **Continuous improvement:** Through data analysis and predictive modeling, the company can gather insights from the data of mobile specifications and learn how it influences their prices. It enables continuous improvement by identifying patterns, best practices, and areas for enhancement, leading to better price predictions over time.

Overall, the study of mobile specifications and prices provides valuable information for effective decision-making, resource management, growth planning and competitive positioning in companies that deal with the sales of smartphones. It helps these companies optimize its efforts, increase the likelihood of sales and market success of smartphones, and achieve its growth targets.

## Data under study

The Mobile Price Specifications and Prices dataset sourced from the Kaggle data science communitywebsite(<https://www.kaggle.com/datasets/pratikgarai/mobile-phone-specifications-and-prices?resource=download>) that shows the price of mobiles was used to study and create a prediction model.

The dataset contains 21 attributes in total – 20 features and a target variable which is the price. The features include Name, Brand, Model, Touch Screen, Battery capacity, Resolution x, Resolution y, RAM, Screen Size, Internal Storage, Operating system, Front Camera, Rear Camera, Number of SIMs, Bluetooth, GPS, 3G, Wi-Fi, 4G/LTE. The target variable is the price.

The specific focus of this project is to analyze and predict the prices of smartphones. By understanding the factors that influence the prices of smartphones, the company can make data-driven decisions, manage resources effectively, and meet their growth targets. The project aims to provide insights into how variables such as Battery capacity, RAM, Screen Size, Internal Storage, Operating system, Front Camera, Rear Camera and other factors impact the price of smartphones.

## Data Sources

The Mobile Price Specifications and Prices dataset sourced from the Kaggle data science community website(<https://www.kaggle.com/datasets/pratikgarai/mobile-phone-specifications-and-prices?resource=download>) that shows the price of mobiles was used to train the prediction model. The dataset contains data scraped and preprocessed from [Gadgets360](https://gadgets360.com/)'s website.

## Tools & Techniques

**Tools :** Jupyter Notebook, Power BI.

**Techniques :** Python Libraries, Machine Learning, Data Visualization, Data Analysis.

# CHAPTER 2: DATA PREPARATION AND UNDERSTANDING

## Phase I – Data Extraction and Cleaning:

* + - **Missing Value Analysis and Treatment**
    - **Handling Outliers**
    - **Feature Extraction**

## Phase II - Feature Engineering

* + - **Encoding (Label Encoding)**
    - **EDA (Exploratory Data Analysis)**
    - **Feature Engineering**
    - **Feature Scaling**

## Data Dictionary:

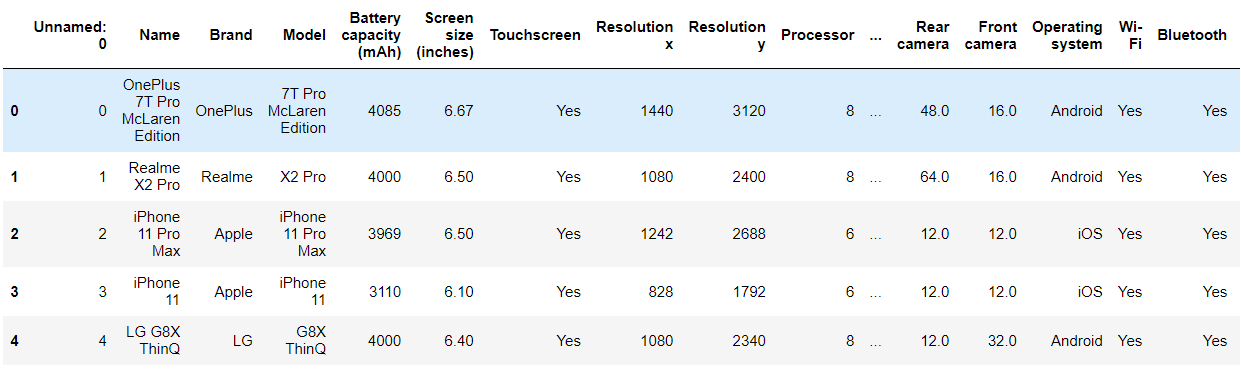


Figure 2.1 Data

1. Name– Object
2. Brand– Object
3. Model– Object
4. Battery capacity(mAh) – Int64
5. Screen size(inches) – Float64
6. Touchscreen – Object
7. Resolution x– Int64
8. Resolution y– Int64
9. Processor – Object
10. RAM(MB) – Int64
11. Internal storage(GB) – Float64
12. Rear Camera – Float64
13. Front Camera – Float64
14. Operating system – Object
15. Wi-Fi – Object
16. Bluetooth – Object
17. GPS – Object
18. Number of SIMs – Int64
19. 3G – Object
20. 4G/LTE – Object
21. Price – Int64

## Exploratory Data Analysis:

1. **Data collection and understanding:** It involves collecting the data and understanding its shape and format.
2. **Data cleaning and preprocessing:** This involves cleaning and preprocessing the data to handle missing values, outliers, and inconsistencies. This involve techniques such as treating outliers, normalization, or scaling.
3. **Data visualization:** Creating visual representations of data using plots, charts, histograms, scatter plots, box plots, etc. Visualization helps in identifying patterns, trends, outliers, and relationships between variables.
4. **Feature engineering:** Exploring potential transformations that might improve the performance of our models. This involve selecting relevant features, scaling features, or transforming variables to achieve better predictive power.

# CHAPTER 3: FITTING MODELS TO DATA

**Linear Regression**

**Decision Tree**

**Random Forest**

**K-NN**

**XGBoost**

* 1. **Linear Regression Model**

Linear Regression was applied on the Training data set. Below were the parameters with which the model was evaluated.

* + - Mean Absolute Error
    - Mean Absolute Percentage Error
    - Mean Squared Error
    - Root Mean Squared Error
    - R^2
  1. **Decision Tree Regressor Model**

Decision Tree Regressor was applied on the Training data set. Below were the parameters which were applied for Decision Tree.

* + - Max\_depth
    - Min\_samples\_split
    - Min\_samples\_leaf
    - GridSearchCV
  1. **Random Forest Regressor Model**

Random Forest Regressor was applied on the Training data set to validate if any further improvement of the model can be performed post the Decision Tree. Below were the parameters which were applied for Random Forest.

* + - N\_estimators
    - Max\_depth
    - Min\_samples\_split
    - Min\_samples\_leaf
    - Max\_features
    - Bootstrap
  1. **KNN Model**

K Nearest Neighbour was applied on the Training data set to validate if any further improvement of the model can be performed post the Random Forest Regressor. Below were the parameters which were applied for KNN.

* + - n\_neighbors
    - p
    - algorithm
  1. **XGBoost Regressor Model**

XGBoost Regressor was applied on the Training data set to and the accuracy was calculated to find out its performance. Below were the parameters which were applied for XGBoost Regressor.

* + - N\_estimators
    - Max\_depth
    - Max\_leaves
    - Learning\_rate
  1. **INFRASTRUCTURE CHALLENGES**
     + Handling Outliers: Implement appropriate techniques to address outliers in the dataset, such as IQR or Z-score method.
     + Feature Selection: Utilize feature selection methods (e.g., statistical tests, feature importance measures) to identify and drop columns that are deemed not important for achieving higher accuracy in the model. This step helps reduce dimensionality and focus on the most relevant features for predictive modeling.
     + Feature Scaling: Perform feature scaling by standardizing or normalizing data. This will help to bring the various features in the dataset on the scale and hence making predictions more accurate.

**CHAPTER 4: KEY FINDINGS**

Below table provides a snapshot of the various models which the business can choose from based on the pros and cons of each model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No** | **Model Name** | **MAE** | **Accuracy** | **Output** |
| **1** | Linear Regression | 5271 | - | Low Accuracy |
| **2** | Decision Tree Model | 4603 | 0.60 | Medium Accuracy |
| **3** | Random Forest Model | 4017 | 0.63 | High Accuracy |
| **4** | K-NN | 3867 | 0.56 | Low Accuracy |
| **5** | XGBoost | 4159 | 0.56 | Low Accuracy |

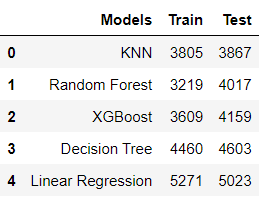


Figure 4.1 MAE Table

**CHAPTER 5: RECOMMENDATIONS AND CONCLUSION**

* For a regression problem, I applied various regression algorithms to the dataset to determine the most suitable model.
* After training the models using the training dataset, I evaluated their performance using the mean absolute error, mean absolute percentage error, mean squared error, root mean squared error, r^2 and accuracy score.
* After evaluating the performance of all the models, it was observed that the Random Forest Regressor algorithms consistently achieved highest accuracy compared to the other models followed by the Decision Tree Regressor.

**CHAPTER 6: REFERENCES**

* + - GeeksForGeeks
      * <https://www.geeksforgeeks.org/machine-learning/>
    - Kaggle
      * <https://www.kaggle.com/datasets/pratikgarai/mobile-phone-specifications-and-prices?resource=download>