

RAMAKRISHNA MISSION VIVEKANANDA EDUCATIONAL AND RESEARCH INSTITUTE

MACHINE LEARNING

FINAL PROJECT REPORT

A Classification Based approach for predicting Smartphone Price Categories

Submitted By

SAYAN DAS
B2430035

RAIHAN UDDIN
B2430070

Submitted To

BR. BHASWARACHAITANYA
(**TAMAL MAHARAJ**)

November 25, 2024

I	Introduction	3
I.1	Background	3
I.2	Motivation	4
I.3	Objectives	4
II	Literature Review	6
III	Dataset Description	9
III.1	Source	9
III.2	Features	9
III.3	Target Variable	10
IV	Data Preprocessing	11
V	Methodology	12
V.1	Algorithms Used	12
V.2	Justification	12
VI	Implementation	13
VI.1	Tools and Libraries	13
VI.2	Parameters	13
VI.3	Training Process	13

VII Results	14
VII.1 Evaluation Metrics	14
VII.2 Comparison between different Models	14
VIII Discussion	15
VIII.1 Analysis of Results	15
VIII.2 Anomalies	15
VIII.3 Limitations	15
IX Conclusion	16
IX.1 Summary	16
IX.2 Reflection on Objectives	16
IX.3 Future Work	16
References	18

I.1 Background

Now a days, mobile phones are more than just a means of communication. The global smartphone market is characterized by rapid technological innovation, intense competition, and increasingly sophisticated consumer expectations. With the proliferation of smartphones, the market dynamics have become increasingly complex, driven by continuous technological advancements, changing consumer preferences, and competitive pricing strategies.

The smartphone industry represents a highly dynamic technological ecosystem where manufacturers constantly strive to differentiate their products through innovative features, design, and pricing. Each season, hundreds of new smartphones are launched, each targeting different market segments and consumer needs. This rapid evolution creates significant challenges for both manufacturers and consumers in understanding and predicting smartphone pricing.

From a manufacturer's perspective, optimal pricing is crucial for maintaining market competitiveness and profitability. Pricing strategies must balance multiple factors including technological features, production costs, market positioning, and consumer purchasing power. Inaccurate pricing can lead to significant market share losses or reduced profit margins.

For consumers, purchasing a smartphone requires understanding of the complex landscape of technical specifications, brand reputation, and market trends. The ability to predict or understand the factors influencing smartphone

prices can help consumers make more informed purchasing decisions and assess the value proposition of different devices.

The emergence of machine learning techniques offers promising approaches to address these pricing challenges. By leveraging historical data and advanced predictive modeling, it becomes possible to develop more sophisticated and accurate methods of smartphone price categorization and prediction.

I.2 Motivation

The motivation for this research stems from the increasingly complex and dynamic nature of the smartphone market. Several critical challenges drive the need for an advanced smartphone price categorization approach:

1. **Economic Significance for Manufacturers :** Inaccurate pricing can result in significant financial losses or missed market opportunities.
2. **Tranceparent Pricing for Consumers :** Many consumers face challenges in understanding the intrinsic value of mobile phones. A data-driven approach to price categorization can provide transparent insights into the factors that genuinely influence mobile phone pricing.

I.3 Objectives

The primary objectives of this project are:

1. **Develop a Robust Classification Model :**
 - Create a machine learning model capable of accurately categorizing smartphones into distinct price segments
 - Achieve high predictive accuracy using multiple classification algorithms
 - Identify and leverage the most significant features influencing smartphone pricing
2. **Feature Analysis and Selection :**
 - Conduct comprehensive analysis of smartphone features
 - Determine the most influential factors in price categorization
 - Develop a systematic approach to feature selection and importance ranking
3. **Comparative Algorithm Performance :**
 - Implement and evaluate multiple machine learning algorithms
 - Compare the performance of different classification techniques
 - Identify the most effective algorithm for smartphone price category prediction
4. **Practical Applicability :**
 - Demonstrate the practical utility of the developed model for both manufacturers and consumers

5. Methodological Contribution :

- Develop a systematic approach to smartphone price categorization
- Contribute to the existing body of knowledge in machine learning applications in market analysis
- Establish a replicable methodology for similar predictive modeling challenges

SECTION II

LITERATURE REVIEW

In this paper [2] the authors investigate price class prediction for smartphones using machine learning, focusing on determining whether a mobile phone is economical or expensive. Data was collected from GSMArena, including features like display size, weight, RAM, and battery capacity. The dataset was preprocessed and categorized into four price classes: very economical, economical, expensive, and very expensive. Feature selection methods like InfoGain and WrapperAttributeEval were employed to reduce dataset dimensionality, optimizing computational efficiency. The study tested classifiers such as Decision Tree (J48) and Naïve Bayes, achieving a maximum accuracy of 78% with Decision Tree when combined with WrapperAttributeEval. Challenges included converting a regression problem into classification, which introduced errors, and the limited dataset size impacting accuracy. The paper concludes with recommendations for improved feature selection techniques, larger datasets, and extending the model for other product categories.

Chandrashekhara et al. [3] 2019, focuses on predicting smartphone prices using machine learning techniques like Support Vector Regression (SVR), Backpropagation Neural Network (BNN), and Multiple Linear Regression (MLR). The dataset included features such as brand, RAM, memory, battery power, and display size, with 262 records spanning 2010–2018, sourced from e-commerce platforms. Data preprocessing steps included handling missing values, standardizing formats, and splitting the data into training (80%) and testing (20%) sets. Performance metrics such as R-squared and correlation values were used to evaluate models. SVR performed the best with an R-squared of 0.86 and correlation of 0.93. Graphical analyses demonstrated that SVR had the most accurate predictions, followed by BNN and MLR. The authors highlight the potential of SVR for broader applications in price prediction across retail industries, suggesting scalability through distributed systems like Hadoop.

Ercan & Şimşek [4] examines the classification of mobile phone prices into low, medium, high, and very high categories using a Kaggle dataset with 2000 entries and 20 features. Features included physical attributes (weight, dimensions), performance metrics (RAM, processor cores), and functionalities (3G, Wi-Fi). Four machine learning algorithms—Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—were applied and evaluated. SVM achieved the highest accuracy (96.16%), outperforming Logistic Regression (91%), Decision Tree (82%), and KNN (41%). Confusion matrices and performance metrics (accuracy, precision, recall, and F1-score) were used to validate results. The study emphasizes the superiority of SVM for this classification problem and suggests exploring additional algorithms and datasets for improved accuracy in future work.

In another study [5] the author focuses on predicting the price of second-hand electronic devices, particularly smartphones, using machine learning (ML) techniques. It addresses the growing trend of buying used electronics, driven by economic factors and technological advancements. The authors utilized web scraping to gather a dataset covering the last five years, containing various features influencing smartphone prices. They experimented with three ML algorithms—Random Forest, Linear Regression, and Multi-Layer Perceptron—to develop a predictive model. Performance evaluation was based on metrics like Absolute Percentage Difference (APD) and Root Mean Square Error (RMSE). Random Forest emerged as the best-performing model, demonstrating the lowest prediction error and superior generalization. The study highlights how accurate price prediction can aid both buyers and sellers, making the market more efficient. Data preprocessing and feature selection were critical steps, leading to a dataset with 24 variables. Techniques like 10-fold cross-validation were used to ensure robustness. The findings suggest that incorporating modern ML methods can significantly enhance price prediction accuracy compared to traditional approaches.

In this paper Akash Gupta, and Suhasini Kottur [6] explore mobile price prediction using machine learning, aiming to categorize mobile phones as economical or expensive based on their features. The study emphasizes the relevance of predictive modeling in aiding marketing and consumer decision-making. Key features considered include processor type, battery capacity, memory, and screen size. The authors employed algorithms like Linear Regression and K-Nearest Neighbors (KNN) to build predictive models. The seven-step machine learning process involved data gathering, preparation, model selection, training, evaluation, hyperparameter tuning, and prediction. The KNN model showed higher accuracy compared to Linear Regression, particularly in scenarios involving complex, non-linear relationships. The research demonstrated the importance of feature selection to reduce computational complexity and enhance prediction precision. The conclusion underscores the impact of ML-driven price prediction in optimizing marketing strategies and enhancing business decisions.

Finally, this paper [7] investigates the prediction of mobile phone prices using machine learning, with a

focus on market analysis and consumer behavior. The study outlines the challenges of determining accurate mobile prices due to diverse features and competitive market dynamics. A dataset containing information on screen size, memory, camera quality, and battery was used to train predictive models. Key algorithms employed included KNN, Forward Selection, and Backward Selection to manage data complexity and optimize feature selection. The data collection phase involved gathering specifications from various mobile phones to construct a comprehensive dataset. The authors emphasized the role of data visualization techniques, like the Elbow Method, to identify the optimal number of features. The study presented a detailed analysis of feature importance, highlighting key parameters influencing price prediction. Classification and testing phases assessed the model's accuracy, using preprocessed datasets for training and evaluation. Results were visualized using graphs that compared predicted prices against specifications like RAM and memory. The conclusion suggests that precise feature selection can improve predictive accuracy and aid in product launch decisions.

SECTION III

DATASET DESCRIPTION

III.1 Source

The dataset for this mobile price classification project was obtained from Kaggle, a prominent platform for data science and machine learning datasets.

Link - <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

The dataset is publicly available and contains 2000 mobile phone entries with 20 feature variables and a target variable representing price range.

III.2 Features

The dataset comprises 20 features that describe various characteristics of mobile phones. Each feature provides critical information about the smartphone's specifications. Following table provides a detailed breakdown of these features:

Feature Name	Description
battery_power	battery capacity in mAh
blue	has bluetooth or not
clock_speed	speed at which processor executes instructions
dual_sim	has dual sim support or not
fc	front Camera Megapixels
four_g	has 4G or not
int_memory	internal Memory capacity
m_dep	mobile Depth in cm
mobile_wt	weight of mobile phone
n_cores	number of cores in processor
pc	primary Camera Megapixels
px_height	pixel Resolution Height
px_width	pixel Resolution Width
ram	RAM in MB
touch_screen	has touch screen or not
wifi	has wifi or not
sc_h	screen Height in cm
sc_w	screen Width in cm
talk_time	longest time that a single battery charge will last over a call
three_g	has 3G or not

III.3 Target Variable

The target variable, `price_range`, represents the categorization of mobile phones into different price segments. It is a categorical variable with four distinct classes:

- 0 : Low-cost mobile phones
- 1 : Medium-low cost mobile phones
- 2 : Medium-high cost mobile phones
- 3 : High-end, premium mobile phones

SECTION IV

DATA PREPROCESSING

SECTION V _____

_____ METHODOLOGY

V.1 Algorithms Used

V.2 Justification

SECTION VI _____

IMPLEMENTATION

VI.1 Tools and Libraries

VI.2 Parameters

VI.3 Training Process

SECTION VII

RESULTS

VII.1 Evaluation Metrics

VII.2 Comparison between different Models

SECTION VIII _____

DISCUSSION

VIII.1 Analysis of Results

VIII.2 Anomalies

VIII.3 Limitations

SECTION IX _____

CONCLUSION

IX.1 Summary

IX.2 Reflection on Objectives

IX.3 Future Work

REFERENCES

- [1] Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 3rd ed., O'Reilly Media, 2023.
- [2] Asim, Muhammad, and Zafar Khan. "Mobile Price Class Prediction Using Machine Learning Techniques." *International Journal of Computer Applications*, March 2018. DOI: 10.5120/ijca2018916555.
- [3] Chandrashekhara, K. T., and M. Thungamani. "Smartphone Price Prediction in Retail Industry Using Machine Learning Techniques." In *Emerging Research in Electronics, Computer Science and Technology, Lecture Notes in Electrical Engineering 545*, edited by V. Sridhar et al., Springer Nature Singapore Pte Ltd., 2019. DOI: 10.1007/978-981-13-5802-9_34.
- [4] Ercan & Şimşek, "Mobile Phone Price Classification Using Machine Learning." *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 7, no. 4, 2023, pp. 458-462. DOI: 10.59287/ijanser.791.
- [5] Abbasi, Muhammad Hasnain, Abdul Sajid, Muhammad Arshad Awan, and Ayeb Amani. "Predicting The Price Of Used Electronic Devices Using Machine Learning Techniques." *International Journal of Computing and Related Technologies*, vol. 4, no. 1, 2024. Available at: <https://www.researchgate.net/publication/377526585>.
- [6] Gupta, Akash, and Suhasini Kottur. "Mobile Price Prediction by Its Features Using Predictive Model of Machine Learning." *UGC Care Journal*, August 2020. DOI: 10.13140/RG.2.2.20054.52800.

- [7] Kumuda S, Vishal Karur, and Karthick Balaje S. E. "Prediction of Mobile Model Price Using Machine Learning Techniques." *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 11, no. 1, October 2021. DOI: 10.35940/ijeat.A3219.1011121.