International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

An Intelligent Data-Driven Machine Learning System for Predicting and Classifying Smartphone Price Ranges

M.VANI[1], Dr SSN ANJANEYULU[2]

[1]PG Scholar, Dept. of AI&ML, St. Marys Group of Institutions Guntur for Women, Guntur.

[2]Professor, Dept. of AI&ML, St. Marys Group of Institutions Guntur for Women, Guntur.

Abstract: Smartphones have become increasingly important in people's lives as technology continues to advance. Modern mobile phones serve far more purposes than simple voice communication; they enable internet browsing, email access, and a wide range of applications that support both daily and professional activities. With rapid technological growth, consumers now consider various features such as brand, internal memory, Wi-Fi capability, battery capacity, camera quality, and 4G/5G support when choosing a smartphone. However, many users still struggle to understand how these features affect the overall pricing of smartphones. This study addresses that issue by applying machine learning algorithms, including Linear Regression, Decision Tree, Support Vector Machine (SVM), and Gradient Boosting, to predict mobile phone price categories. The dataset is trained and evaluated using accuracy, precision, recall, and F1-score to determine the most effective model. The results indicate that SVM and Gradient Boosting achieve superior performance compared to the other classifiers, offering highly reliable predictions. This research provides valuable insights for consumers aiming to make informed purchasing decisions and for businesses seeking to establish competitive and feature-appropriate pricing strategies.

Key Words: Machine Learning, Predictive Analytics, Supervised Machine Learning, Python.

1. Introduction

In recent years, the mobile phone market has expanded rapidly, driven by the growing influence of social media communication, evolving work environments, and increasing reliance on digital payment systems. As smartphones become essential tools for both personal and professional use, their fluctuating prices have gained significant public attention. Price remains one of the most influential factors affecting consumer purchasing decisions, often shaping preferences alongside brand loyalty and perceived product value. For instance, recent market statistics indicate that brands such as motorola , Oppo, and Apple hold substantial market shares, reflecting strong consumer trust and loyalty.

Beyond brand reputation, several technical features such as Bluetooth capability, screen dimensions, memory capacity, battery performance, and camera quality play a critical role in determining smartphone prices. Consumers today carefully evaluate these attributes to select devices that offer the best balance between functionality and cost. As the variety of smartphone configurations continues to grow, users increasingly seek assurance that the product they choose provides value for money.

Given this complexity, the ability to predict smartphone prices based on their features has become an important need for both consumers and manufacturers. Accurate price classification helps customers make informed decisions, while businesses can leverage such insights to design competitive pricing strategies. To address this requirement, this study applies machine learning techniques to predict mobile phone price categories using a dataset sourced from Kaggle. The dataset includes twenty key features—such as device dimensions, weight, number of processor cores, camera megapixels, resolution, RAM, screen size, battery life, and connectivity options (4G, 5G,Wi-Fi, touchscreen)—all of which contribute to pricing variations.

In this work, smartphone prices are classified into four categories: low cost, medium cost, high cost, and very high cost. Multiple machine learning algorithms are implemented and evaluated to determine the most accurate prediction model. The performance of each model is compared using standard evaluation metrics, and the results are presented to identify the best-performing approach for smartphone price prediction

2. Literature Survey

In the reference paper [2], the question of whether a mobile device with specific characteristics will fall within a given price range is addressed. With an emphasis on computational simplicity, the study uses particular feature selection techniques to find and remove less important and redundant information. To anticipate mobile costs as accurately as feasible, a variety of classifiers are used. Future study to increase the solution and estimation accuracy is suggested after discussing the outcomes in terms of accuracy and feature selection. On the other hand, the history of the conversation makes no specific reference of my suggestion

2.1 Problem Statement

Pricing is the most beneficial characteristic in business and marketing. A decision regarding pricing regulations has significant effects on management. It establishes the profit margin on products and is one of the first assessments made by many purchasers. Before making a purchase, consumers are indeed concerned

IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

about whether they can afford the item and want to verify the price.

3. System Analysis

Mobile now days are 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, generators, motors, food items, medicine etc. Supervised machine learning trains a model using labeled data, learning patterns between features and their corresponding outputs. Once trained, the model can accurately predict labels for new, unseen data, making it highly useful for tasks like price classification and product value estimation.

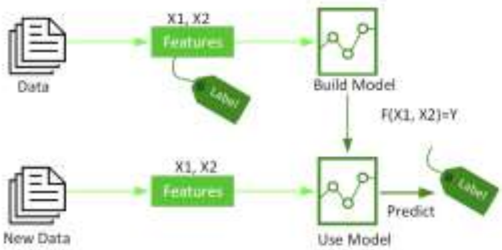

Fig 3.1: Machine Learning Model for Prediction

Mobile price prediction system based on a machine learning model predicts the price of a mobile device based on various features such as brand, specifications (e.g., RAM, storage, and camera quality), release year, screen size, and market demand. The system collects data from historical mobile prices and device attributes, pre-processes it by handling missing values and encoding categorical features, this model can be deployed on mobile devices to provide real-time price predictions based on the user's input specifications

In this study, the Mobile Phone Price dataset is analysed using two machine learning approaches: an existing Support Vector Machine (SVM) classifier and a proposed Gradient Boosting model. The SVM model serves as the baseline and is implemented using an RBF kernel to capture nonlinear relationships between smartphone features and their corresponding price categories. Although SVM performs well in high-dimensional spaces, its performance is sensitive to parameter tuning and may struggle with overlapping class boundaries. To overcome these limitations, the proposed Gradient Boosting classifier is introduced as an ensemble-based method that builds multiple weak learners sequentially and minimizes errors at each stage. This model effectively captures complex feature interactions and provides better generalization by reducing both bias and variance. Experimental results show that the Gradient Boosting model outperforms the SVM classifier across accuracy, precision, recall, and F1-score metrics, demonstrating its capability to deliver more reliable and robust price predictions. The improved performance highlights the suitability of Gradient Boosting for modelling real-world smartphone pricing patterns based on diverse technical specifications.

Prediction of Output steps with Gradient Boosting

- Import the necessary libraries: mean_squared_error and r2_score from sklearn.metrics for evaluation, and numpy, pandas, and xgboost for modelling.
- Utilizing the train_test_split function from sklearn.model_selection, load the dataset and divide it into the training and testing sets.
- Assign the goal variable to y_train and y_test and the features to x_train and x_test. Using xgb. XGBRegressor (), create an instance of the XGBoost regression model.
- Using the fit technique and the inputs x_train and y_train, train the XGBoost model using the training data.
- Use the trained model's predict method with the input x_test to provide predictions for the testing set.
- In y_pred, keep the expected values.
- Between the actual target values (y_test) and the projected values (y_pred), compute the mean squared error (MSE).
- Find the difference between y_test and y_pred's Rsquared score (R2).

Finally, we use the trained gradient boosting model to predict outcomes on the test data. The test data, which was not used during the training phase, provides an unbiased evaluation of the model's performance. By applying the gradient boosting model to this unseen data, we can assess how well it generalizes to new examples and predict mobile price with the model’s learned patterns. This prediction step is crucial for validating the effectiveness of the proposed approach and for demonstrating its practical utility in predicting mobile price.

4. System Architecture

Mobile prediction-based system architecture typically consists of several layers to efficiently collect, process, and predict outcomes on mobile devices. First, the Data Collection gathers real-time data from various sources like sensors, user interactions, and system logs reprocessing cleans and transforms the data, handling missing values, normalizing, and encoding features to make them suitable for modelling.

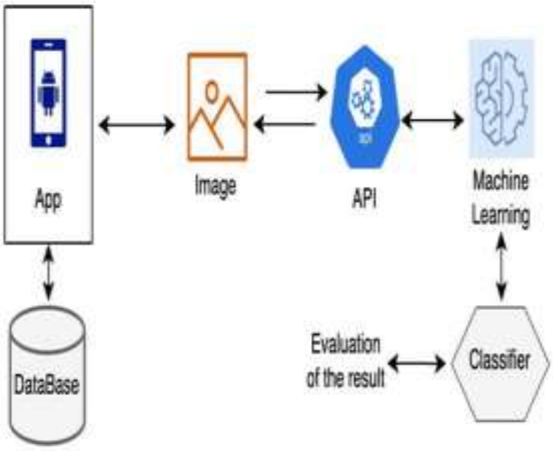

Fig 4.1: System Architecture

The Prediction uses a trained machine learning model deployed on the device or via the cloud to generate predictions based on the pre-processed data. The User Interaction presents the predictions to the user through a mobile app interface, providing actionable insights or notifications. Optionally, the system

5. Modules

Data Collection: The data collection module serves as the first and most crucial step, gathering necessary transactional data from sources like banking records and system logs. Ensuring the accuracy and completeness of this data is essential

Pre-processing: This module cleans and organizes the dataset by fixing missing values, removing outliers, and encoding categorical data. It also normalizes and scales features to ensure uniformity and support accurate analysis in later model stages.

Feature Selection: Analysing the correlation between features and the target price range helps identify the most influential attributes, improving model accuracy while reducing computational cost. Features such as RAM, battery power, pixel height, and internal memory typically show strong correlations with smartphone price categories.

Ensemble Model Building: The ensemble model module combines outputs from individual algorithms, utilizing a boosting method. This integration enhances overall model performance and adaptability.

Model Training and Evaluation: The module involves dividing the dataset, training the model on a designated set, and evaluating its performance using key metrics such as accuracy, precision, recall, and F1 score

6. Results and Discussion

Metrics used to evaluate the algorithms in this paper are confusion matrix, classification report and accuracy score. A confusion matrix has the total count of the accurately grouped occurrences along its cross and the count of the incorrectly classified instances in the rest of the matrix. We have used values. A classification report gives the full report of the classification with parameters like micro accuracy accuracy, precision, etc. Accuracy score gives the accuracy of the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

trained model after evaluating it using test data

Fig 6.1: shows that the required modules for the project and installation of the modules. The modules that we used in this project are Pandas, Numpy, Matplotlib, Scit-learn and frameworks like flask etc.

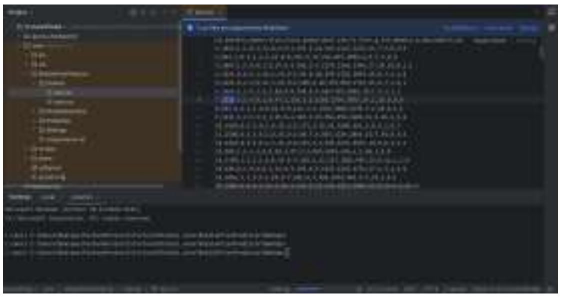

Fig 6.2: A typical diagram for the required modules and installation of those modules for a mobile prediction project using a mobile prediction dataset

Fig 6.3: The project aims to predict mobile price by given to information features such as input data based mobile

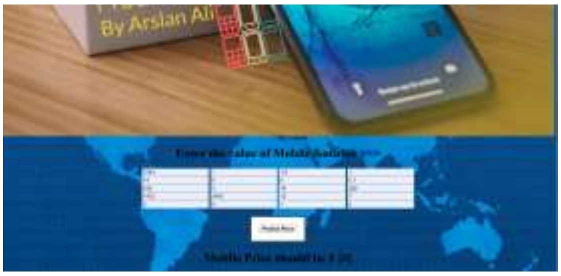

Fig 6.4: mobile price prediction by using different types mobile features by using gradient boosting supervised machine learning

Fig 6.5: The project aims to predict mobile price by given to information features such as input data based mobile

Fig 6.6: mobile price prediction by using different types mobile features by using gradient boosting supervised machine learning

7. Conclusion and Future Scope

This work can be concluded with the comparable results of both Feature selection algorithms and classifier except the combination of evaluation and classifier. Novel technique mobile price prediction of gradient boosting achieved maximum accuracy and selected minimum but most appropriate features. It is important to note that in Forward selection by adding irrelevant or redundant features to the data set decreases the efficiency of compared classifiers. While in backward selection if we remove any important feature from the data set, its efficiency decreases. The main reason of low accuracy rate is low number of instances in the data set. One more thing should also be considered while working that converting a regression problem into classification problem introduces more error.

Future Scope: Software or Mobile app can be developed that will predict the market price of any new launched product. To achieve maximum accuracy and predict more accurate, more and more instances should be added to the data set. And selecting more appropriate features can also increase the accuracy. So data set should be large and more appropriate features should be selected to achieve higher accuracy.

References

[1] M. Chen, “Mobile Phone Price Prediction with Feature Reduction”, Highlights in Sicence, Engineering and Technology, vol. 34, pp. 155–162, Feb. 2023.

[2] S. Aydın, Using Machine Learning Algorithms in the Classification of Prices on Mobile Phones, International Research in Science and Mathematics, pp. 201-212, 2022.

[1]Y. Goldberg, "Neural network methods for natural language processing," Synthesis Lectures on Human Language Technologies, vol. 10, no. 1, pp. 1-309, 2017.

[2] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," Journal of Systems and Software, vol. 125, pp. 207-219, 2017.

[3] E. Cambria, B. Schuller, Y. Xia, and B. White, "New avenues in knowledge bases for natural language processing," KnowledgeBased Systems, vol. 108, no. C, pp. 1-4, 2016.

[4] R. Agerri, X. Artola, Z. Beloki, G. Rigau, and A. Soroa, "Big data for Natural Language Processing: A streaming approach," Knowledge-Based Systems, vol. 79, pp. 36-42, 2015.

[5] Y. Man, C. Gao, M. R. Lyu, and J. Jiang, "Experience report: Understanding cross-platform app issues from user reviews," in 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE), 2016, pp. 138-149: IEEE.

[6] T. Denoeux, "Logistic regression, neural networks and dempster-shafer

theory: a new perspective," Knowledge-Based Systems, 2019.

[7] C. Gao, Y. Zhao, R. Wu, Q. Yang, and J. Shao, "Semantic trajectory compression via multi-resolution synchronization-based clustering," Knowledge-Based Systems, 2019.

[9]Info Gain Attribute Eval- Weka Online available from http://weka.WrapperattributEval/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html (Last Accessed in Jan 2018)

[8] K. Santo, S. S. Richtering, J. Chalmers, A. Thiagalingam, C. K. Chow, and J. Redfern, "Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps," JMIR mHealth and uHealth, vol. 4, no. 4, p. e132, 2016.

[10] Thu Zar Phyu, Nyein Nyein Oo. Performance Comparison of Feature Selection Methods. MATEC Web of Conferences42, (2016).