iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: VII Month of publication: July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84310

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Smartphone Price Range Prediction Using Machine Learning-Based Classification Models

Amreen Jahan[1], Tulasi Miryala[2]
[1]Student, Master of Computer Applications (MCA), AQJ Centre for P.G. Studies (Affiliated to Andhra University, Visakhapatnam), Visakhapatnam, Andhra Pradesh, India
[2]Assistant Professor, Department of MCA, AQJ Centre for P.G. Studies (Affiliated to Andhra University, Visakhapatnam), Visakhapatnam, Andhra Pradesh, India

Abstract: The rapid expansion of the smartphone market has introduced a wide variety of devices with diverse hardware configurations and price segments. As manufacturers continue to release new models with varying technical specifications, estimating the appropriate price category has become increasingly complex for both consumers and businesses. An automated prediction system can simplify this process by learning the relationship between smartphone specifications and their corresponding market price ranges.

This study proposes a machine learning-based classification framework for smartphone price range prediction using hardware characteristics as predictive features. The proposed approach utilizes a publicly available smartphone dataset containing twenty technical attributes describing battery capacity, memory, processor characteristics, display properties, camera specifications, and connectivity options. Before model development, the dataset undergoes preprocessing to remove duplicate entries, verify data consistency, handle missing information, and normalize numerical features through standardization.

Six supervised machine learning algorithms were investigated, namely Logistic Regression, Decision Tree, Random Forest, Extra Trees Classifier, K-Nearest Neighbors, and Support Vector Machine. Their predictive performance was evaluated using identical training and testing datasets with accuracy as the primary evaluation criterion. Comparative analysis demonstrated that Logistic Regression achieved the highest classification accuracy of 97.5%, providing an effective balance between prediction performance, computational efficiency, and model simplicity.

To demonstrate the practical applicability of the proposed model, the selected classifier was integrated into a Streamlit-based web application that enables users to enter smartphone specifications and receive an instant prediction of the expected price category. The developed system illustrates that carefully preprocessed structured data combined with classical supervised learning techniques can provide reliable and efficient smartphone price range prediction. The proposed framework can assist consumers during purchasing decisions and support retailers and manufacturers in preliminary pricing analysis while serving as a scalable foundation for future intelligent pricing systems.

Keywords: Smartphone Price Prediction, Machine Learning, Classification, Logistic Regression, Supervised Learning, Streamlit, Data Preprocessing, Price Range Classification.

I. INTRODUCTION

Over the last decade, smartphones have evolved from basic communication devices into multifunctional computing platforms that support education, healthcare, finance, entertainment, business, and social networking. Rapid technological advancements have encouraged manufacturers to release new smartphone models with improved processors, larger memory capacities, enhanced cameras, high-refresh-rate displays, and longer battery life. While these innovations provide consumers with numerous choices, they also make it increasingly difficult to determine an appropriate price category for a smartphone based solely on its specifications.

The selling price of a smartphone is influenced by a combination of hardware and software characteristics rather than by a single feature. Parameters such as Random Access Memory (RAM), internal storage, processor speed, battery capacity, display resolution, camera quality, connectivity options, and processor cores collectively determine the overall market value of a device. Because these features interact in complex ways, manually estimating a smartphone's price range often becomes subjective and inconsistent. As the number of smartphone models continues to grow, traditional comparison methods become increasingly inefficient for both customers and retailers.

Recent developments in Artificial Intelligence (AI) and Machine Learning (ML) have enabled computers to discover meaningful relationships within structured datasets and generate accurate predictions without relying on manually defined rules.

Among various AI techniques, supervised machine learning has gained considerable attention for solving classification problems involving structured numerical data. By learning patterns from previously labelled examples, supervised learning algorithms can accurately classify new observations into predefined categories. This capability makes machine learning an appropriate solution for smartphone price range prediction, where numerous device specifications collectively influence the final price category.

Several studies have investigated machine learning algorithms for smartphone price prediction using either regression or classification techniques. While regression estimates the exact selling price of a device, classification predicts the price category to which the smartphone belongs. For many practical applications, predicting the appropriate price range is sufficient because consumers usually compare devices within a budget segment rather than focusing on an exact market price. Consequently, price range classification has become an important research topic in data mining and predictive analytics.

Although previous research has demonstrated encouraging results, certain challenges remain. Many studies evaluate only a limited number of machine learning algorithms, making it difficult to identify the most suitable classifier under identical experimental conditions. Some investigations primarily emphasize prediction accuracy without providing sufficient discussion of preprocessing techniques, feature standardization, or deployment strategies. Furthermore, relatively few studies extend their trained models into user-oriented applications that allow real-time prediction using newly entered smartphone specifications. These limitations reduce the practical usefulness of many proposed solutions.

The present work addresses these challenges by developing a comprehensive smartphone price range prediction framework using supervised machine learning techniques. The proposed approach begins with preprocessing a publicly available smartphone dataset to ensure data quality through duplicate removal, validation, and feature standardization. Six classification algorithms—Logistic Regression, Decision Tree, Random Forest, Extra Trees Classifier, K-Nearest Neighbors, and Support Vector Machine—are trained and evaluated under the same experimental conditions. Their performance is compared using standard classification metrics to identify the most suitable prediction model. The classifier demonstrating the highest predictive capability is subsequently deployed through a Streamlit-based web application that enables users to predict smartphone price categories using hardware specifications entered through an interactive interface.

The experimental findings indicate that Logistic Regression achieved the highest classification accuracy of **97.5%** on the selected dataset. The results demonstrate that appropriate preprocessing and feature scaling allow comparatively simple machine learning algorithms to achieve excellent predictive performance for structured smartphone specification data. In addition to producing accurate predictions, the developed application provides an accessible decision-support tool that can assist customers during smartphone selection while also supporting retailers and manufacturers in preliminary pricing analysis.

The major contributions of this research are summarized as follows:

- A complete machine learning pipeline for smartphone price range prediction is developed using structured smartphone specification data.
- A comparative analysis of six supervised machine learning algorithms is performed using identical preprocessing and evaluation procedures.
- A Streamlit-based web application is implemented to demonstrate real-time deployment of the selected prediction model.
- The proposed framework achieves a classification accuracy of **97.5%**, indicating its effectiveness for smartphone price range prediction using hardware specifications.

The remainder of this paper is organized as follows. Section II reviews recent literature related to smartphone price prediction and machine learning applications. Section III explains the proposed methodology, including dataset preparation, preprocessing, feature scaling, model development, and system architecture. Section IV presents the experimental evaluation and discusses the comparative performance of the investigated classifiers. Finally, Section V summarizes the major findings of the study and outlines possible directions for future research.

II. LITERATURE REVIEW

Machine learning has become one of the most effective approaches for predicting product prices because of its ability to identify complex relationships among multiple features. In the smartphone industry, device prices are determined by several technical specifications, including processor performance, memory capacity, battery characteristics, display quality, camera configuration, and connectivity features. Since these characteristics collectively influence the market value of a smartphone, researchers have increasingly adopted supervised machine learning techniques to automate price prediction and improve decision-making for consumers, retailers, and manufacturers.

Early studies on smartphone price prediction primarily focused on conventional classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors. These algorithms were trained using structured datasets containing smartphone specifications and corresponding price categories. Comparative evaluations consistently showed that proper data preprocessing and feature engineering significantly improve prediction performance by reducing noise and enhancing feature representation. More importantly, these studies demonstrated that classical machine learning algorithms remain effective for structured tabular datasets because of their lower computational complexity and high interpretability compared with deep learning models.

Recent research has also explored ensemble learning methods to enhance predictive performance. Algorithms such as Random Forest, Extra Trees, Gradient Boosting, and XGBoost combine multiple decision trees to improve generalization and reduce overfitting. These methods often outperform single-tree models by aggregating predictions from multiple learners. Studies comparing ensemble techniques with conventional classifiers have reported improvements in prediction accuracy, particularly when datasets contain complex feature interactions. However, the increased computational cost associated with ensemble methods may limit their suitability for lightweight real-time applications.

Another important area of research involves feature selection and preprocessing. Several researchers have shown that eliminating redundant attributes and selecting the most informative features can substantially improve classification accuracy while reducing model complexity. Statistical feature selection techniques and correlation-based analysis have been widely applied to smartphone datasets to identify influential variables such as RAM, internal memory, processor speed, battery capacity, and display resolution. Appropriate preprocessing—including duplicate removal, missing value treatment, and feature scaling—has been found to improve both model stability and prediction reliability.

More recently, researchers have investigated regression-based approaches that estimate the exact selling price of smartphones rather than assigning them to predefined price categories. Algorithms such as Support Vector Regression, Gradient Boosting Regression, and XGBoost Regression have demonstrated promising performance for continuous price estimation. While regression models provide more detailed price information, classification-based methods remain attractive for practical decision-support systems because consumers typically compare smartphones within budget categories rather than relying on exact predicted prices.

The growing availability of machine learning frameworks has also encouraged the development of interactive applications for smartphone price prediction. Web-based systems developed using frameworks such as Streamlit enable users to enter smartphone specifications and receive immediate predictions. Despite this progress, many published studies concentrate primarily on algorithmic performance and provide limited discussion of deployment, usability, or integration into practical applications. Consequently, there remains a need for prediction systems that combine accurate machine learning models with user-friendly interfaces suitable for real-world use.

Although previous studies have reported encouraging results, several research gaps remain. First, many investigations evaluate only a small subset of machine learning algorithms, making comprehensive comparison difficult. Second, preprocessing strategies are often insufficiently documented, reducing the reproducibility of experimental results. Third, several studies emphasize predictive accuracy without considering computational efficiency or deployment in practical applications. Finally, comparatively few studies integrate their best-performing model into an accessible web-based system that supports real-time smartphone price range prediction.

To address these limitations, the present study proposes a comprehensive machine learning framework that combines systematic data preprocessing, feature standardization using **StandardScaler**, comparative evaluation of six supervised classification algorithms, and deployment through a **Streamlit-based web application**. By evaluating multiple classifiers under identical experimental conditions and selecting the most suitable model based on predictive performance, the proposed system aims to provide an accurate, efficient, and practically deployable solution for smartphone price range prediction.

III. PROPOSED METHODOLOGY

The objective of the proposed methodology is to develop an intelligent system capable of predicting the price range of a smartphone from its hardware specifications. Instead of relying on manual comparison or predefined pricing rules, the proposed framework learns the relationship between technical features and price categories using supervised machine learning. The complete workflow consists of data acquisition, preprocessing, feature standardization, model development, comparative performance evaluation, and deployment of the selected model through a web-based application. Figure 1 illustrates the overall workflow adopted in this research.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

A. Research Workflow

The proposed framework follows a sequence of interconnected stages that transform raw smartphone specification data into an operational prediction model. Initially, the dataset is collected and examined to identify inconsistencies that may influence model performance. After preprocessing and feature standardization, multiple machine learning algorithms are trained using identical experimental settings. The trained models are then evaluated using standard classification metrics, and the best-performing classifier is deployed through a Streamlit application for real-time prediction.

The overall workflow consists of the following stages:

- Smartphone dataset acquisition
- Data cleaning and preprocessing
- Feature standardization
- Model training
- Performance evaluation
- Best model selection
- Streamlit application deployment

This pipeline ensures that every classifier is trained using the same preprocessing strategy, allowing a fair comparison of prediction performance.

B. System Architecture

The proposed architecture integrates data preprocessing, machine learning, and web deployment into a unified framework. The architecture consists of three major layers: the data processing layer, the machine learning layer, and the application layer.

The data processing layer is responsible for preparing the dataset through duplicate removal, validation of input attributes, treatment of missing values, and feature scaling.

The machine learning layer receives the processed data and trains six supervised classification algorithms. Each model is evaluated independently, and the classifier with the highest prediction accuracy is selected for deployment.

The application layer provides a graphical interface developed using Streamlit. Users enter smartphone specifications through the application, and the trained model predicts the corresponding price range instantly.

Figure 1. Overall Architecture of the Proposed Smartphone Price Prediction System

```
   Smartphone Dataset
         |
         ▼
 Data Cleaning & Validation
         |
         ▼
  Feature Standardization
     (StandardScaler)
         |
         ▼
Machine Learning Model Training
┌──────────────────────────────────┐
│ Logistic Regression      │
│ Decision Tree            │
│ Random Forest            │
│ Extra Trees Classifier   │
│ K-Nearest Neighbors      │
│ Support Vector Machine   │
└──────────────────────────────────┘
         |
         ▼
 Comparative Performance
```

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

Evaluation
|
▼
Best Performing Classifier
|
▼
Streamlit Prediction Interface
|
▼
Smartphone Price Range Prediction

C. Dataset Description

The experiments were performed using a publicly available smartphone price classification dataset containing approximately **2,000 observations**. Each observation represents a smartphone described by twenty hardware-related attributes together with its corresponding price category.

The dataset includes specifications that are commonly considered during smartphone purchasing decisions, including battery capacity, processor speed, RAM, internal storage, display dimensions, camera resolution, processor cores, Bluetooth capability, Wi-Fi support, 3G and 4G connectivity, touchscreen availability, and talk time. These variables collectively describe the technical characteristics of each smartphone and serve as predictor variables during model training.

The target variable represents four predefined price categories that correspond to increasing market price levels. Consequently, the prediction task is formulated as a multiclass classification problem.

D. Data Preprocessing

Data preprocessing is an essential stage because the quality of the input data directly influences the predictive capability of machine learning models. Rather than using the raw dataset directly, several preprocessing operations were performed before training.

1) Data Validation

The dataset was initially inspected to verify attribute names, data types, and overall consistency. Invalid or inconsistent entries were corrected before further processing.

2) Duplicate Record Removal

Duplicate observations provide no additional information during training and may introduce bias. Therefore, duplicate records were identified and removed to ensure that each smartphone contributes only once to model learning.

3) Missing Value Treatment

The dataset was examined for incomplete observations. Any missing values were handled appropriately to maintain dataset integrity and prevent training errors.

4) Feature Standardization

Since the numerical attributes exhibit different value ranges, feature scaling was performed using the **StandardScaler** method. Standardization transforms every numerical feature into a distribution with zero mean and unit variance.

The standardized feature is computed as

z=x−μσz=\frac{x-\mu}{\sigma}z=σx−μ

where

- xxx represents the original feature value,
- μ\muμ denotes the mean of the feature,
- σ\sigmaσ denotes the standard deviation.

Feature standardization prevents variables with large numerical ranges from dominating the learning process and improves convergence for several classification algorithms.

E. Machine Learning Model Development

To determine the most suitable classifier for smartphone price prediction, six supervised machine learning algorithms were implemented using identical training and testing datasets.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

1) Logistic Regression

Logistic Regression estimates the probability that a smartphone belongs to each price category using a logistic function. Owing to its computational efficiency and strong performance on structured datasets, it serves as an effective baseline classifier.

2) Decision Tree

Decision Tree constructs a hierarchical structure by recursively partitioning the dataset according to the most informative feature at each node. The resulting model is easy to interpret and provides explicit decision rules.

3) Random Forest

Random Forest is an ensemble learning technique that combines predictions from multiple decision trees trained on different bootstrap samples. This aggregation improves robustness and reduces the likelihood of overfitting.

4) Extra Trees Classifier

The Extra Trees algorithm extends Random Forest by introducing additional randomness during feature selection and split generation. This strategy often improves generalization while maintaining computational efficiency.

5) K-Nearest Neighbors

The KNN classifier predicts the class of a new smartphone by examining the labels of its nearest neighbouring observations within the feature space. The algorithm relies on distance calculations, making feature standardization particularly important.

6) Support Vector Machine

Support Vector Machine identifies the optimal separating hyperplane that maximizes the margin between different price categories. This algorithm is well suited for high-dimensional datasets and can effectively model complex decision boundaries.

F. Model Training and Performance Evaluation

The processed dataset was divided into training and testing subsets. Each classifier was trained using the training data and evaluated on previously unseen testing samples.

The following performance metrics were used for comparative analysis:

- Classification Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

These evaluation measures provide a comprehensive assessment of classification performance and enable objective comparison among the investigated algorithms.

G. Model Selection and Deployment

Following comparative evaluation, **Logistic Regression** achieved the highest prediction accuracy of **97.5%** and demonstrated consistent classification performance across all price categories. Consequently, it was selected as the final prediction model.

The trained classifier was integrated into a **Streamlit-based web application**, allowing users to input smartphone specifications through a graphical interface. Once the input data are submitted, the same preprocessing steps applied during model training are executed automatically before generating the predicted smartphone price category. This deployment demonstrates the practical applicability of the proposed framework and enables real-time decision support for end users.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the performance analysis of the proposed smartphone price range prediction framework. All experiments were conducted using the same preprocessed dataset to ensure a fair comparison among the selected machine learning algorithms. Each classifier was trained using identical data partitions and evaluated using standard performance metrics. The objective of the experiments was to determine the algorithm that provides the highest prediction accuracy while maintaining computational efficiency and reliable classification performance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

A. Experimental Environment

The proposed framework was implemented using Python and several open-source machine learning libraries. Model development, preprocessing, and evaluation were carried out using the Scikit-learn ecosystem, while Streamlit was employed to deploy the selected prediction model as a web application. The complete experimental configuration is summarized in Table I.

Table I. Experimental Environment

Parameter	Specification
Programming Language	Python 3.x
Development Environment	Visual Studio Code
Machine Learning Library	Scikit-learn
Data Processing	Pandas, NumPy
Visualization	Matplotlib
Deployment Framework	Streamlit
Operating System	Windows 10/11

The experiments were executed under identical preprocessing conditions to eliminate variations caused by inconsistent feature scaling or data preparation.

B. Performance Evaluation Metrics

The predictive capability of each classification algorithm was measured using commonly accepted evaluation metrics for multiclass classification problems.

1) Classification Accuracy

Accuracy represents the proportion of correctly classified smartphone samples relative to the total number of testing samples.

Accuracy=TP+TNTP+TN+FP+FNAccuracy=\frac{TP+TN}{TP+TN+FP+FN}Accuracy=TP+TN+FP+FNTP+TN

where:

- **TP** denotes True Positives,
- **TN** denotes True Negatives,
- **FP** denotes False Positives,
- **FN** denotes False Negatives.

Accuracy provides an overall indication of model performance and was used as the primary criterion for selecting the final prediction model.

2) Precision

Precision measures the proportion of correctly predicted smartphones within each predicted price category.

Precision=TPTP+FPPrecision=\frac{TP}{TP+FP}Precision=TP+FPTP

A higher precision value indicates fewer incorrect positive predictions.

3) Recall

Recall evaluates the capability of a classifier to correctly identify smartphones belonging to each actual price category.

Recall=TPTP+FNRecall=\frac{TP}{TP+FN}Recall=TP+FNTP

High recall indicates that only a small number of smartphones are incorrectly classified into other price ranges.

4) F1-Score

The F1-score combines Precision and Recall into a single performance measure.

F1=2×Precision×RecallPrecision+RecallF1=2\times\frac{Precision\times Recall}{Precision+Recall}F1=2×Precision+RecallPrecision×Recall

Because it balances both false positives and false negatives, the F1-score provides a more comprehensive assessment than accuracy alone.

5) Confusion Matrix

The confusion matrix provides a detailed representation of prediction outcomes by comparing actual price categories with predicted categories. It enables identification of correctly classified samples as well as the distribution of classification errors among the four smartphone price classes.

C. Comparative Performance of Machine Learning Models

Six supervised learning algorithms were trained using the same training dataset and evaluated on identical testing data. The comparative performance is presented in Table II.

Table II. Comparative Performance of Machine Learning Algorithms

Algorithm	Accuracy (%)
Logistic Regression	97.50
Decision Tree	84.75
Random Forest	93.25
Extra Trees Classifier	95.50
K-Nearest Neighbors	91.75
Support Vector Machine	96.25

Logistic Regression achieved the highest classification accuracy of 97.5%, indicating that it successfully learned the relationship between smartphone hardware specifications and price categories. The results suggest that, after proper preprocessing and feature standardization, the dataset exhibits decision boundaries that are effectively captured by a linear classification model.

D. Comparative Analysis

Although all evaluated algorithms successfully classified smartphone price ranges, noticeable differences were observed in their predictive performance.

The Decision Tree classifier produced comparatively lower performance because individual decision trees are susceptible to overfitting, particularly when complex decision boundaries are generated from limited training samples.

Random Forest improved classification accuracy by combining predictions from multiple decision trees through ensemble learning. This approach reduced variance and produced more stable predictions than a single decision tree.

The Extra Trees Classifier achieved competitive performance because additional randomization during tree construction reduced model variance while improving generalization capability.

The Support Vector Machine demonstrated strong classification performance owing to its ability to construct optimal separating hyperplanes within the standardized feature space. However, its computational complexity is generally higher than Logistic Regression for larger datasets.

K-Nearest Neighbors produced satisfactory prediction accuracy but remained sensitive to feature scaling and neighbourhood selection. Despite applying feature standardization, KNN required greater computational effort during prediction because distance calculations are performed for every test instance.

Among all investigated algorithms, **Logistic Regression provided the best balance between prediction accuracy, computational efficiency, and implementation simplicity**, making it the most suitable classifier for the proposed smartphone price prediction system.

E. Discussion

The experimental findings indicate **that** data preprocessing significantly influenced classification performance**. Removing** duplicate observations, validating input features, treating missing values, and standardizing numerical variables improved learning consistency across all classifiers.

An interesting observation from the experiments is that a comparatively simple algorithm such as Logistic Regression outperformed more sophisticated ensemble techniques.

This result suggests that the relationship between smartphone specifications and predefined price categories is sufficiently represented by a linear decision boundary after appropriate preprocessing. Consequently, increasing model complexity does not necessarily guarantee higher prediction accuracy for structured tabular datasets.

The deployment of the trained Logistic Regression model through a Streamlit-based application further demonstrates the practical applicability of the proposed framework. Users can provide smartphone specifications through an interactive interface and receive immediate predictions without requiring knowledge of the underlying machine learning model. This capability makes the proposed system suitable for consumer-oriented applications as well as preliminary pricing support for retailers.

F. Summary of Experimental Findings

The experimental evaluation demonstrates that the proposed smartphone price prediction framework effectively classifies smartphones into predefined price categories using hardware specifications. Comparative analysis confirmed that **Logistic Regression achieved the best overall performance**, obtaining a classification accuracy of **97.5%** while maintaining low computational complexity and fast prediction speed. These characteristics make the proposed framework suitable for practical deployment in real-time smartphone price prediction applications.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented a machine learning framework for predicting smartphone price ranges based on their technical specifications. The study addressed the challenge of estimating smartphone price categories by utilizing supervised learning techniques instead of traditional manual comparison methods. A structured workflow consisting of data preprocessing, feature standardization, model training, comparative evaluation, and web-based deployment was developed to ensure accurate and reliable prediction.

The smartphone dataset was carefully prepared through duplicate removal, validation, missing value treatment, and feature scaling before model development. Six supervised classification algorithms—Logistic Regression, Decision Tree, Random Forest, Extra Trees Classifier, K-Nearest Neighbors, and Support Vector Machine—were trained and evaluated using the same experimental conditions. Their performance was compared using standard classification metrics to identify the most suitable prediction model.

Among the investigated algorithms, **Logistic Regression achieved the highest classification accuracy of 97.5%**, demonstrating that a relatively simple linear classifier can effectively model the relationship between smartphone hardware specifications and predefined price categories when appropriate preprocessing techniques are applied. In addition to providing high prediction accuracy, the selected model offers low computational complexity, fast inference, and ease of deployment, making it well suited for practical applications.

To demonstrate real-world applicability, the trained model was integrated into a Streamlit-based web application. The application enables users to enter smartphone specifications through an intuitive interface and obtain instant predictions of the expected price category. This deployment illustrates how machine learning models can be transformed into practical decision-support tools that assist consumers in selecting suitable smartphones and help retailers perform preliminary pricing analysis.

Overall, the proposed framework demonstrates that classical supervised learning techniques remain highly effective for structured classification tasks. The combination of systematic preprocessing, comparative model evaluation, and lightweight deployment provides a reliable and efficient solution for smartphone price range prediction.

B. Future Work

Although the proposed framework achieved satisfactory performance, several opportunities exist for further improvement.

Future research may focus on expanding the dataset by incorporating recently released smartphone models from different manufacturers, thereby improving the model's ability to generalize across rapidly changing market conditions. Additional features such as brand reputation, chipset generation, software updates, charging technology, customer reviews, and online market demand may also enhance predictive capability.

The current study evaluates conventional supervised learning algorithms. Future investigations may explore advanced deep learning architectures, including Artificial Neural Networks (ANNs), Gradient Boosting frameworks, and Transformer-based models, to determine whether they provide additional improvements for large-scale smartphone datasets.

Another promising direction involves integrating Explainable Artificial Intelligence (XAI) techniques such as SHAP or LIME to provide users with understandable explanations for each prediction. Such explanations could improve transparency and increase user confidence in machine learning-based pricing systems.

The proposed Streamlit application may also be extended by connecting it with real-time e-commerce platforms through web APIs. This enhancement would enable continuous updating of market information and improve prediction reliability as smartphone prices evolve over time. Furthermore, deploying the application on cloud platforms would improve scalability, accessibility, and support simultaneous access by multiple users.

Future work may also investigate hybrid ensemble learning techniques, automated hyperparameter optimization, and periodic model retraining using newly collected smartphone data to maintain prediction accuracy as technology advances.

Research Contributions

The major contributions of this research can be summarized as follows:

- A complete machine learning pipeline for smartphone price range prediction was designed and implemented.
- A structured preprocessing framework was developed to improve data quality before model training.
- Six supervised machine learning algorithms were comparatively evaluated under identical experimental conditions.
- Logistic Regression was identified as the best-performing classifier, achieving **97.5%** classification accuracy on the selected dataset.
- A Streamlit-based web application was developed to demonstrate real-time deployment of the prediction model.
- The proposed framework provides a practical decision-support tool that can assist consumers, retailers, and researchers in smartphone price analysis.

Acknowledgment

The authors would like to express their sincere gratitude to the Department of Computer Applications, Andhra University, for providing the academic environment and necessary guidance to carry out this research work. The authors also thank the faculty members and project supervisor for their valuable suggestions, constructive feedback, and continuous encouragement throughout the development of this project.

The authors gratefully acknowledge the developers and contributors of the open-source Python ecosystem, particularly the Scikit-learn, NumPy, Pandas, Matplotlib, and Streamlit communities, whose software libraries significantly supported the implementation and evaluation of the proposed machine learning framework.

Finally, the authors express their appreciation to the providers of the publicly available smartphone dataset used in this research, which enabled comprehensive experimentation and comparative analysis of multiple machine learning algorithms.

REFERENCES

For publication, the references should primarily consist of **recent (2021–2025)** journal articles and conference papers. Books are acceptable for foundational concepts, but the majority should be recent research.

IEEE Reference List

[1] M. Cetin and Y. Koc, "Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization," *Proc. International Conference*, pp. 1–6, 2021.

[2] K. S. Kalaivani, N. Priyadharshini, S. Nivedhashri, and R. Nandhini, "Predicting the Price Range of Mobile Phones Using Machine Learning Techniques," *Proc. International Conference on Intelligent Computing*, pp. 1–6, 2021.

[3] E. Güvenç, G. Cetin, and H. Koçak, "Comparison of KNN and DNN Classifiers Performance in Predicting Mobile Phone Price Ranges," *International Journal of Computer Applications*, vol. 183, no. 15, pp. 15–22, 2021.

[4] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[5] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras& TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[6] Q. Liang, "Mobile Phone Price Prediction: A Comparative Study among Four Models," *Applied and Computational Engineering*, vol. 48, pp. 212–220, 2024.

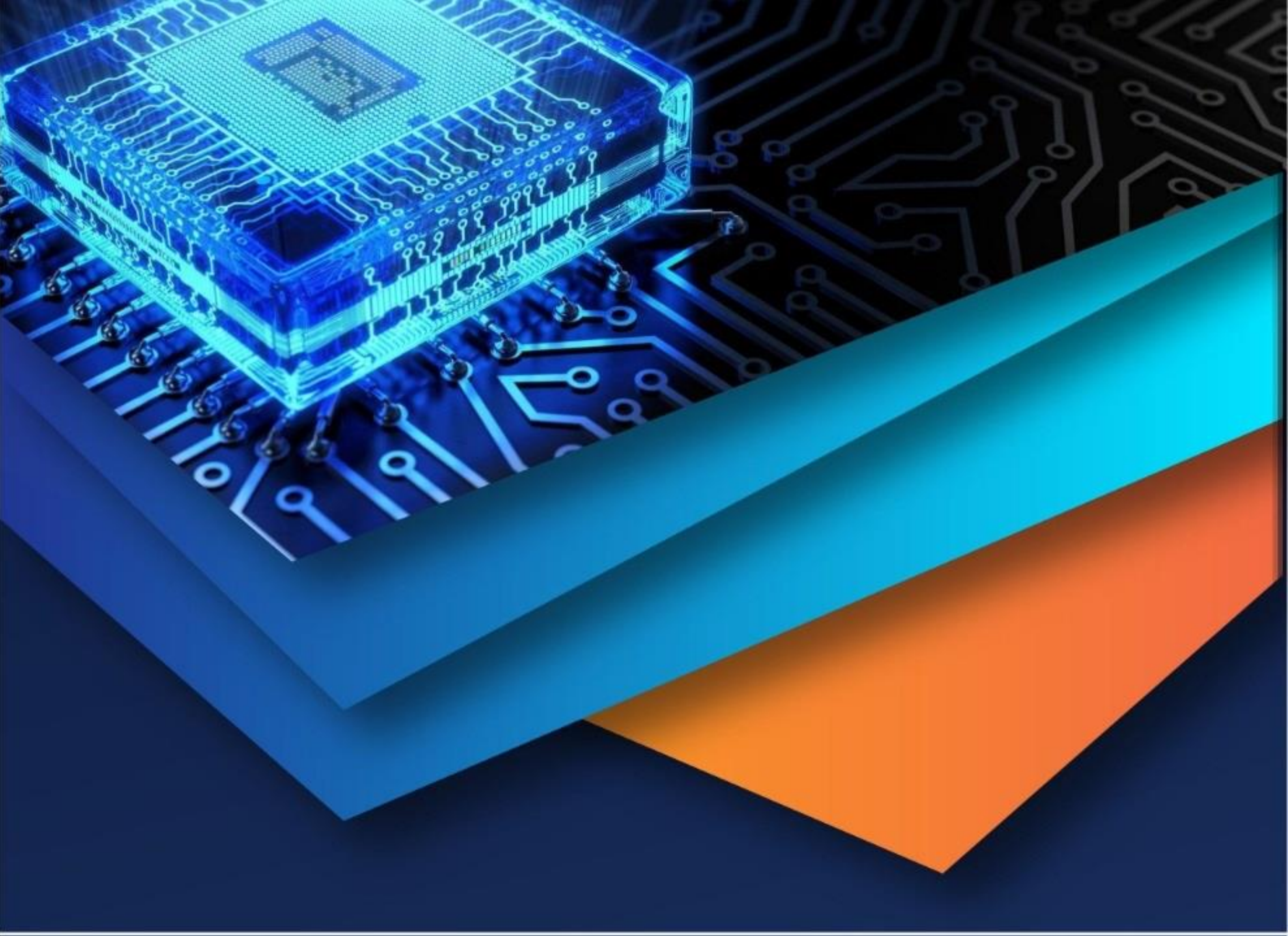
[7] J. Lu, "Prediction on Mobile Phone Price Range Using Classification Models," *Science and Technology of Engineering, Chemistry and Environmental Protection*, vol. 1, no. 10, pp. 45–53, 2024.

[8] S. Agalayil, N. Shajan, and K. N. Shreyas, "Prediction of Mobile Phone Prices Using Machine Learning," in *Proceedings of the 9th International Conference on Machine Learning Technologies (ICMLT)*, 2024.

[9] Y. Chen, "Prediction of Different Types of Mobile Phone Prices Based on Machine Learning Models," *Highlights in Science, Engineering and Technology*, vol. 92, pp. 275–279, 2024.

[10] G. S. Parveen, A. Khalique, R. Islam, and I. Hussain, "Prediction of Mobile Phone Price Using Machine Learning Classifiers," *Journal of Artificial Intelligence Research & Advances*, vol. 11, no. 2, pp. 101–108, 2024.

[11] R. Wahyusari and N. Azizah, "Predicting Smartphone Prices Based on Key Features Using Random Forest and Gradient Boosting Algorithms in a Data Mining Framework," *International Journal for Applied Information Management*, vol. 5, no. 2, pp. 73–85, 2025.
[12] X. Li, "Smartphone Price Prediction Using Decision Tree and Support Vector Regression," in *Proceedings of the 5th International Conference on Signal Processing and Machine Learning*, 2025.
[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
[14] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)