e-ISSN: 3089-5081 | p-ISSN: 1979-8911
VoL. 14 No. 1 June 2025: 39-43
DOI: 10.15575/istek.v14i1.2154

Price Prediction of Second-Hand Iphones Using Random Forest Regression Based on Unit Conditions

Denta Pratama Anggayana[1], Ichsan Taufik[2], Yana Aditia Gerhana[3]
[1,2,3] Department of Informatic Engineering, UIN Sunan Gunung Djati Bandung, Indonesia

Article Info

Keywords:

Price prediction
Second-hand Iphone
Random Forest Regression
CRISP-DM
Machine Learning

ABSTRACT

This study presents the development of a price prediction model for second-hand Iphones based on unit conditions using the Random Forest Regression algorithm, implemented in a web-based application. A dataset of 542 records was collected from Facebook Marketplace and iPhone trading groups, with variables including Iphone type, storage capacity, warranty status, Face ID, and Truetone. The research employed the CRISP-DM methodology through the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The model was tested using data splits of 80%–20%, 70%–30%, and 60%–40%, resulting in MAE values of 8.32%–8.42% and RMSE values of 10.64%–10.88%, indicating good and consistent accuracy. The developed system can automatically provide price recommendations based on unit conditions, assisting both sellers and buyers in determining fair market prices.

Corresponding Author:

Denta Pratama Anggayana
Informatic Engineering Department, Faculty of Science & Technology, UIN Sunan Gunung Djati Bandung Jl. A. H. Nasution No. 105, Cibiru, Bandung, Indonesia. 40614
Email: 1187050020@student.uinsgd.ac.id

1. INTRODUCTION

Over the past decades, technological advancements have developed rapidly, with communication devices continuously introducing new innovations each year, such as fingerprint sensors, facial recognition, and other advanced features. Currently, one of the most popular communication devices in the global market is the iPhone, which has a large and growing user base. In Indonesia, several official distributors such as Ibox, Digimap, and Blibli operate as authorized Apple resellers, including iPhone sales. However, numerous non-official distributors also market iPhones under international warranty schemes. Many of these devices are not registered with customs due to high associated fees. Officially registered iPhones have their IMEI numbers recorded in the Ministry of Industry's database, whereas unregistered ones do not appear in the system.

Iphone prices in the Indonesian market vary according to several factors, including warranty type and device condition. Apart from warranty, the physical and functional condition of the iPhone significantly influences its resale value. For both sellers and buyers, the ability to analyze and predict iPhone prices offers a competitive advantage in determining an optimal transaction price. Price prediction can be defined as the systematic process of estimating the most likely value of a product in the future.

Machine learning techniques, particularly decision tree–based algorithms such as Random Forest Regression, provide accurate prediction capabilities. Random Forest Regression is an ensemble learning method that constructs multiple decision trees and combines their outputs to generate more stable and reliable predictions. This algorithm is highly effective in handling datasets with multiple variables, while simultaneously reducing the risk of overfitting. Consequently, it is well-suited for modeling complex relationships between unit conditions and the resale prices of second-hand iPhones.

Several studies have demonstrated the effectiveness of Random Forest in predictive modeling. For example, Septiyanah and Athalina reported that Random Forest Regression achieved an accuracy of 0.83, outperforming Multiple Linear Regression, which achieved an accuracy of 0.68 [1]. Similarly, Kriswantara and Sadikin concluded that, after evaluating five different models for car price prediction, Random Forest produced the lowest error rates, with MAE = 1.006 and RMSE = 1.452 [2]. Furthermore, Pratiwi and Qoiriah found that Random Forest with 100 trees, evaluated using 10-fold cross- validation, achieved the best results, yielding an accuracy of 83.750% and a MAPE of 16.249% [3]. Ricky Verdiyanto, Dwi Hartanti, Eko Purwanto Based on the analysis from the development stage of the Random Forest Regression model to its integration into the application, it was able to provide prediction results with a very small mean absolute percentage error of 2.85%, indicating that the model can predict well [4]. Green Arther Sandag conducted a study which showed that, based on the problems encountered, the Random Forest algorithm outperformed other algorithms in identifying weaknesses in the Apple's AppStore dataset. The results achieved were an accuracy of 86.23%, recall of 84.80%, precision of 84.45%, and RMSE of 0.316 for independent testing. Furthermore, the 10-fold cross-validation results showed an accuracy of 86.08%, recall of 84.56%, precision of 84.38%, and RMSE of 0.316 in predicting ratings on Apple's AppStore [5]. Vito Hafizh Cahaya Putra, Muhammad Al- Husaini, Ari Purno Wahyu, and Agung Rachmat Raharja conducted a study in which the prediction of cherry tomato plant height using Random Forest Regression was performed based on the latest sensor data from ThingSpeak, with the evaluation resulting in a Mean Squared Error (MSE) of 0.8294 and an R^2 score of 0.8939 [6]. Rahmat Hidayat, Haris Tri Saputra, Mirdatul Husnah, Nabila, M. Bintang Hidayatullah, Muhammad Naufal Nazhmi, Jauzaa Azra, and Astri Rana concluded in their study that the model demonstrated excellent performance. The performance evaluation produced an Out-of-Bag (OOB) value of 99.99%, a Mean Squared Error (MSE) of 2.4899, an R-squared value close to 1 (0.99996), and a Mean Absolute Error (MAE) of only 0.9305. These results indicate that the model has very high accuracy, low error, and optimal generalization capability for new data [7]. Rizqatasyaa Achmad Zahra's research showed that the prediction accuracy of rental prices using Random Forest Regression, after applying Hyperparameter Tuning, was evaluated using RMSE, R^2, and MAPE on the testing data. The results yielded an RMSE of 971,099, an R^2 value of 0.92 indicating that the independent variables influenced the dependent variable (rental price), and a MAPE of 19.32%, with an overall accuracy of 80.68% [8]. Nazmia Kurniawati, Aisyah Novfitri, and Yuli Kurnia Ningsih, from the prediction results using the Random Forest Regression algorithm with seven estimators, obtained channel gain threshold values of 0.1 and 0.3 [9]. Ardianto, Agus Budi Raharjo, and Diana Purwitasari concluded that Random Forest Regression (RFR) achieved the best performance with identical scores of R^2 = 0.9679 and RMSE = 0.0438. From the comparison of the highest R^2 values, it was shown that Gradient Boosted Trees (GBT) and RFR outperformed the others, with RFR performing 4.91% better than Decision Tree Regression (DTR) and 5.32% better than Linear Regression (LR). Selecting the appropriate scenario for each algorithm was proven to improve forecasting accuracy, with an increase of 2.78% for GBT, 2.90% for RFR, 8.28% for DTR, and 5.79% for LR, derived from the performance difference between the lowest and highest results for each algorithm [10].

Employs the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which consists of six phases—business understanding, data understanding, data preparation, modeling, evaluation, and deployment—allowing the predictive model to be built in alignment with real business needs. Accordingly, this research focuses on predicting second-hand iPhone prices based on unit conditions using Random Forest Regression within the CRISP-DM framework and implementing the model into a web-based application.

Building upon these previous studies, the present research introduces an enhanced system that incorporates additional predictive features. Specifically, the model considers iPhone condition attributes such as warranty type, storage capacity, Face ID functionality, and True Tone feature, thereby enabling users to obtain more accurate and optimal predictions of second-hand iPhone prices.

2. METHODOLOGY

This study adopts the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology, Figure 1 is the flow of CRISP-DM.

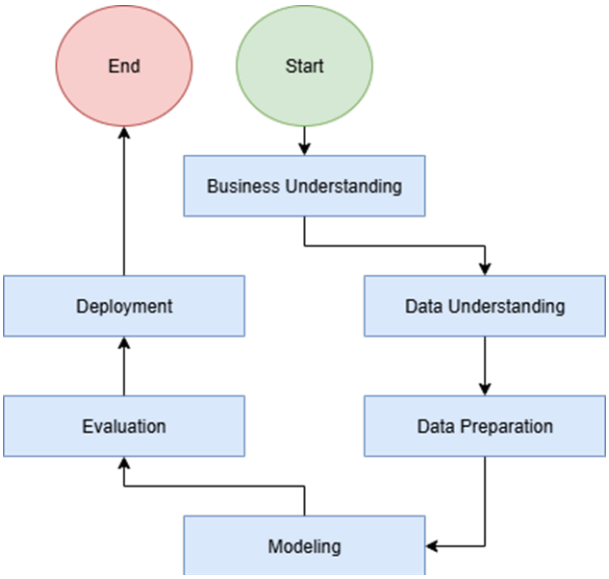

Figure 1 CRISP-DM

which includes the following steps:

1. Business Understanding – identifying the problem of inconsistent second-hand iPhone prices.
2. Data Understanding – collecting 542 data entries from Facebook Marketplace and iPhone trading groups, considering variables such as iPhone type, storage capacity, warranty (official/inter), Face ID, and TrueTone.
3. Data Preparation – manually converting raw data into structured Excel format, followed by mapping categorical variables into numerical form using Google Colab.
4. Modeling – applying Random Forest Regression with parameters n_estimators=100 and random_state=42. Data was split into 80:20, 70:30, and 60:40 for training and testing.
5. Evaluation – measuring model performance using MAE, MSE, and RMSE metrics.
6. Deployment – implementing the final model into a web-based system using the Flask framework.

3. RESULT AND DISCUSSION

The evaluation results, in this stage the prepared testing data is tested using MAE, MSE, and RMSE, producing the model's performance.

3.1. The Result Evaluation 80% Training and 20% Testing

Table 1 presents the evaluation results from 80% training data and 20% testing data,

Table 1 Result Evaluation 80% Training and 20% Testing

Train-Test Split	MAE	Percentage MAE	RMSE	Percentage RMSE
80% : 20%	260.615	8.32%	335.911	10.73%

The model evaluation results from 433 training data and 109 testing data show an MAE of 260,615 and an RMSE of 335,911. The MAE of 8.32% indicates the average prediction error from the actual iPhone price, while the RMSE of 10.73% shows that the model's predictions deviate on average from the actual price.

3.2. The Result Evaluation 70% Training and 30% Testing

Table 2 presents the evaluation results from 70% training data and 30% testing data,

Table 2 Result Evaluation 70% Training and 30% Testing

Train-Test Split	MAE	Percentage MAE	RMSE	Percentage RMSE
70% : 30%	265.892	8.42%	343.488	10.88%

The model evaluation results from 379 training data and 163 testing data show an MAE of 265,892 and an RMSE of 343,488. The MAE of 8.42% indicates the average prediction error from the actual Iphone price, while the RMSE of 10.88% shows that the model's predictions deviate on average from the actual price.

3.3. The Result Evaluation 60% Training and 40% Testing

Table 3 presents the evaluation results from 60% training data and 40% testing data,

Table 3 Result Evaluation 60% Training and 40% Testing

Train-Test Split	MAE	Percentage MAE	RMSE	Percentage RMSE
60% : 40%	263.582	8.35%	336.023	10.64%

The model evaluation results from 325 training data and 217 testing data show an MAE of 263,582 and an RMSE of 336.023. The MAE of 8.35% indicates the average prediction error from the actual Iphone price, while the RMSE of 10.64% shows that the model's predictions deviate on average from the actual price.

3.4 System Deployment

The trained model was deployed into a web-based application using Flask. The system allows users to input unit details (iPhone type, warranty, storage capacity, Face ID, and TrueTone). Based on these inputs, the system generates automatic price recommendations, thus assisting users in determining fair market values for second-hand Iphone

4. CONCLUSION

From the results and discussion of the study titled "Prediction of Second-Hand iPhone Prices Using Random Forest Regression Based on Unit Condition," it can be concluded that:

1. The Random Forest Regression algorithm was successfully applied in developing a system to predict second-hand iPhone prices based on unit condition, with input variables such as iPhone type, storage capacity, warranty status (official/third-party), presence of Face ID, and the Truetone feature.
2. The evaluation results (MAE 8.32%–8.42% and RMSE 10.64%–10.88%) indicate stable and accurate model performance.
   These results indicate that the model's performance is quite good, with a relatively low error percentage, and remains consistent even when the training and testing data are changed. This demonstrates that the Random Forest Regression algorithm has strong performance and can be relied upon to accurately predict second-hand iPhone prices.

REFERENCES

[1] S. Septiyanah and G. Athalina, "S. Septiyanah and G. Athalina, 'IMPLEMENTASI ALGORITMA RANDOM FOREST REGRESSION DALAM PREDIKSI HARGA LAPTOP,' Jurnal MNEMONIC, vol. 8, no. 1, pp. 70-73, Feb. 2025.," *J. Mnemon.*, vol. 8, no. 1, pp. 70–73, 2025, [Online]. Available: https://www.kaggle.com/datasets/mdriazhossain/gl

[2] B. Kriswantara and R. Sadikin, "Used Car Price Prediction with Random Forest Regressor Model," *J. Inf. Syst. Informatics Comput. Issue Period*, vol. 6, no. 1, pp. 40–49, 2022, doi: 10.52362/jisicom.v6i1.752.

[3] D. A. Pratiwi and A. Qoiriah, "Sistem Rekomendasi Wedding Organizer Menggunakan Metode Content-Based Filtering Dengan Algoritma Random Forest Regression," *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 231–239, 2021, doi: 10.26740/jinacs.v3n03.p231-239.

[4] R. Verdiyanto, D. Hartanti, and E. Purwanto, "Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest Regression," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 128–139, 2025, doi: 10.29408/jit.v8i1.28386.

[5] G. Arther Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest," *Cogito Smart J.* |, vol. 6, no. 2, pp. 167–178, 2020.

[6] V. H. C. Putra, M. Al-Husaini, A. P. Wahyu, and A. R. Raharja, "Perancangan Sistem Monitoring Cerdas Berbasis Internet of Things (IoT) dengan Algoritma Random Forest Regression untuk Deteksi Ketinggian pada Tanaman Tomat Cherry," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 10–25, 2024, doi: 10.57152/malcom.v5i1.1612.

[7] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *Simkom*, vol. 10, no. 1, pp. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.

[8] R. A. Zahra, "Penerapan Algoritma Random Forest Dengan Hyperparameter Tuning Untuk Memprediksi Harga Sewa Kost Di Kota Bandar Lampung," *Skripsi*, pp. 1–23, 2016.

[9] N. Kurniawati, A. Novfitri, and Y. K. NINGSIH, "Prediksi Channel Gain Threshold untuk Modulasi Adaptif V2V menggunakan Algoritma Random Forest Regression," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 10, no. 3, p. 544, 2022, doi: 10.26760/elkomika.v10i3.544.

[10] A. B. Ardianto, Raharjo and Purwitasari, "Random Forest Regression Untuk Prediksi Produksi Daya," vol. 7, no. 5, pp. 1058– 1075, 2022.