Journal of Emerging Information Systems and Business Intelligence
ISSN: 2774-3993
Journal homepage: https://ejournal.unesa.ac.id/index.php/JEISBI/

Web-Based Umrah Departure Quota Prediction Application Using a Machine Learning Approach (Case Study: PT. Rahmatan Berkah Wisata)

Muhammad Rifqi Ardani[1], Ardhini Warih Utami[2]

[1,2]*Universitas Negeri Surabaya, Surabaya, Indonesia*

muhammadrifqi.22014@mhs.unesa.ac.id, *ardhiniwarih@unesa.ac.id*

ABSTRACT

PT. Rahmatan Berkah Wisata faces uncertainties regarding Umrah departure quotas due to capacity determination that is still conducted manually. This study aims to design a web-based quota prediction application utilizing Machine Learning with the Random Forest Regression algorithm, as well as to evaluate the accuracy and functionality of the system. System development employs the Rapid Application Development (RAD) method, encompassing the stages of requirement planning, user design, construction, and cutover. The research data consists of historical pilgrim data from the 2022–2026 period, which includes the year, package name, package type, trip duration, number of pilgrims, and departure season.The results show that the Random Forest Regression model achieved a Mean Absolute Error (MAE) of 3.265, a Mean Squared Error (MSE) of 15.186, and an R-squared (R2) value of 0.707 (70.7%), demonstrating its capability to provide sufficiently accurate predictions. The model was successfully implemented into a web-based application featuring Umrah data management, prediction and retraining scheduling, departure prediction, prediction reporting, and model retraining. Furthermore, functional testing using Black Box Testing across three user roles (President Director, Operational Manager, and Pilgrim Staff) was successfully executed across all test scenarios. Based on the findings, this application can effectively serve as a decision support tool for determining Umrah departure quotas.

Keyword: Random Forest Regression, Quota Prediction, Umrah, RAD, Black Box Testing.

Article Info:
Article history:
Received May 23, 2026
Revised June 10, 2026
Accepted June 15, 2026

Corresponding Author
Muhammad Rifqi Ardani
Universitas Negeri Surabaya, Surabaya, Indonesia
muhammadrifqi.22014@mhs.unesa.ac.id

1. INTRODUCTION

The umrah pilgrimage holds profound spiritual significance for Muslims and can be performed throughout the year, unlike the hajj pilgrimage [1]. In recent years, umrah has evolved into a growing lifestyle trend and religious travel industry, increasing competition among travel agencies, including PT Rahmatan Berkah Wisata [2]. To remain competitive, the company must improve its operational management and service quality.

Based on interviews with the Chief Commissioner of PT Rahmatan Berkah Wisata, the company experienced quota management issues 2–3 times in the past year, including unfilled quotas and ticket demand exceeding available capacity. These problems occurred because quota determination relied on experience rather than data analysis. Meanwhile, registrant numbers

fluctuate and are influenced by factors such as package type, trip duration, departure season, and year.

The uncertainty of future demand can be addressed through predictive modeling using machine learning techniques [3]. Previous studies have demonstrated the effectiveness of Random Forest Regression (RFR) in forecasting tasks. The algorithm achieved an accuracy of 96.35% in predicting hajj registration numbers [4], outperformed Linear Regression with an R^2 value of 0.99 compared to 0.64 [5], and maintained strong performance even when trained on relatively small datasets [6].

To overcome quota management challenges, this study proposes the development of a web-based umrah departure quota prediction application using the Random Forest Regression algorithm. The model utilizes historical departure data and supporting variables, including package name, package type, trip duration, departure season, and departure year. Model performance will be evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) metrics [7], [8], while system functionality will be validated through Black Box Testing [9].

Although Random Forest Regression has been successfully applied in various prediction studies, its implementation for umrah departure quota prediction has not been specifically explored. Therefore, the novelty of this research lies in applying the algorithm within the religious travel sector and integrating it into a web-based application to support more accurate and data-driven quota planning at PT Rahmatan Berkah Wisata.

2. METHODS

This study is a Research and Development (R&D) project aimed at designing and developing a web-based umrah departure quota prediction application by applying the Random Forest Regression algorithm as a Machine Learning approach. In its development process, this study utilizes the Rapid Application Development (RAD) method, as it significantly accelerates system creation through an iterative and flexible approach tailored to user requirements [10].

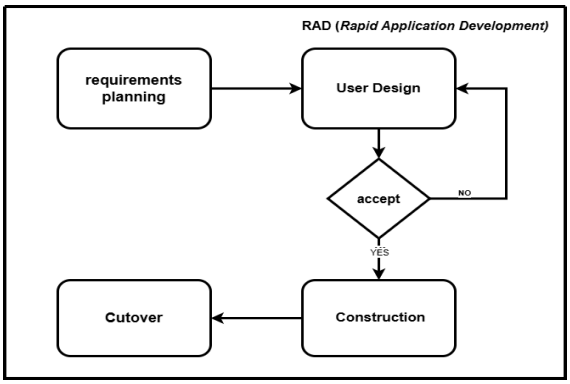

Figure 1. RAD Research flow

The research phases are structured based on the RAD approach in Figure 1. which includes: requirements planning, interactive user design with the company, construction, and final testing and evaluation (cutover) [10].

2.1 Data Collection

The data collection in this study was obtained from interviews and observations conducted directly at the company. Information gathering was carried out through semi-structured interviews, an interactive technique that combines a predefined question guide with field communication flexibility [11]. Through interview and observation, the core research

problems faced by the company in quota determination were successfully identified, while simultaneously gathering primary historical records of umrah pilgrim departures for the 2022–2026 period. These identification results are then analyzed in-depth as part of the Rapid Application Development (RAD) method.

2.2 Unified Modeling Language (UML)

For crafting the software blueprint this research will use Unified Modeling Language (UML) as a visual modeling language to specify the structure and behavior of the object-oriented system [12]. The developed UML diagrams include a use case diagram to model user interactions, an activity diagram to illustrate system activity flows, a sequence diagram to display the message sequences between objects, and a class diagram to explain the static structure of the system [12].

2.3 Random Forest Regression (RFR) algorithm.

Subsequently, the machine learning model training is executed using the Random Forest Regression (RFR) algorithm. This algorithm minimizes the risk of overfitting by constructing multiple independent decision trees through parameter tuning, including ntree (number of trees), mtry (number of random variables), and node size (tree depth) [13], [14].The splitting criteria for the tree branches are determined based on the lowest Mean Squared Error (MSE_n) value [7], formulated as:

$$\mathrm{MSE_n} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{y}_n)^2 \quad (1)$$

Description :

MSEn: MSE value for the nth tree.

N : Number of samples for the nth tree.

Yi : Value of the i-th sample for the nth tree.

$\bar{y}$n : Average value of the n-th tree samples.

The final predictive output for this regression case is obtained by averaging the predictions from all individual decision trees [7], calculated using the following equation:

$$\hat{Y}_i = \frac{1}{Ntree} \sum_{n=1}^{Ntree} \hat{Y}_n \quad (2)$$

Description :

$\hat{Y}$i : Final prediction result.

Ntree : Total number of trees in the Random Forest.

$\hat{Y}$n : Prediction result of the nth tree.

2.4 Predicted Model Evaluation

Next step involves evaluating the predicted model's performance using three quantitative metrics [7]. First, the MSE metric measures the average squared difference between the predicted and actual values [7], [8], expressed as:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (X_i - F_i)^2 \quad (3)$$

Description:

Xi : Actual data for period i.

Fi : Forecast or predicted value for period i.

n : Number of sample data.

Second, the Mean Absolute Error (MAE) metric calculates the average absolute error of the absolute differences in the data [7], derived from the equation:

$$MAE = \frac{1}{n} \sum_{j=1}^{n} | y_j - \hat{y}_j | \quad (4)$$

Description:
n : Number of sample data.
yj : Actual value for sample j.
ŷj : Predicted value for sample j.

Third, the *R*-squared (R^2) metric is employed to determine the proportion of variance in the dependent variable that can be explained by the independent variables in the regression model [7], [8], calculated using the formula:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\Sigma(Yactual - Ypredict)2}{\Sigma(Yactual - \bar{y})\,2} \quad (5)$$

Description:
Y*actual* : Test data.
Y*predict* : Test data, predicted results.
Ȳ : Average of test data.
SSE : Sum of squares of differences between actual and predicted values.
SST : Sum of squares of differences between actual and average values.
SSR : Sum of squares of differences between predicted and average values.

The R^2 score is also utilized to measure the strength of the prediction accuracy generated by the model [8]. The R^2 score ranges from 0 (meaning no correlation) to 1 (indicating a very strong correlation with a positive direction) [15]. To determine the level of relationship strength or accuracy based on the coefficient, the test results will be consulted and interpreted with reference to Table 1. [15].

Table 1. Correlation Coefficient Interpretation and Relationship Strength

Coefficient Interval	Relationship Level
0.800 - 1.000	Very Strong
0.600 - 0.799	Strong
0.400 - 0.599	Moderately Strong
0.200 - 0.399	Low
0.000 - 0.199	Very Low

2.5 Database MySQL

The application database is using MySQL, an open-source Relational Database Management System (RDBMS) [16]. MySQL serves to store, organize, and process data structurally and efficiently using SQL (Structured Query Language) commands [16].

2.6 Flask

The application is developed using the Flask framework, a lightweight and efficient Python-based micro web framework designed to accelerate web development [17]. Flask serves

as the core connector that integrates the predictive model with the user interface, while systematically managing the website components including HTML, CSS, and JavaScript [17].

2.7 Black Box Testing

For system functionality testing using the Black Box Testing method to ensure all features operate according to the company's requirements without examining the internal code structure [18]. This testing is implemented through the Equivalence Partitioning technique, which divides test data into several groups with similar characteristics to validate the input, process, and output of each application form [18].

3. RESULTS AND DISCUSSION

This section presents the system development results through four phases: problem and requirements analysis, system architecture design, application development including data processing, model training and evaluation, and web interface implementation. The section concludes with the results of application functionality testing.

3.1 Problem and System Requirements Analysis

Based on interviews with the Chief Commissioner of PT. Rahmatan Berkah Wisata, the determination of umrah pilgrim departure quotas relies on conventional estimation (considering package name, package types, trip duration, and peak seasons like Ramadan or Maulid). This approach caused issues 2–3 times over the past year, resulting in either unfulfilled quotas or excessive ticket demands, which led to financial losses due to reactive handling.

To address this problem, a web-based prediction system utilizing the Random Forest Regression algorithm was developed. The system supports historical data management, model training, prediction generation, accuracy evaluation (MSE, MAE, and R^2), automated prediction scheduling, manual retraining, and secure user authentication. Additionally, the system is designed to provide accurate predictions, efficient processing, easy model updates, an intuitive user interface, and secure cross-device accessibility.

3.2 Use Case Diagram

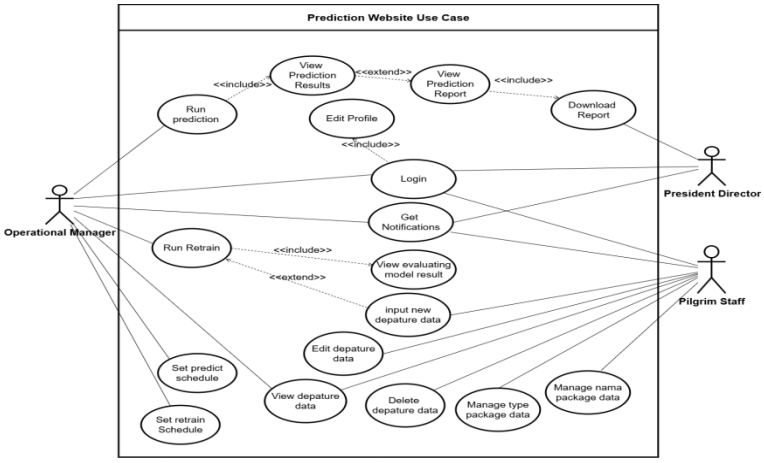

Figure 2. Use Case Predict Application

Based on Figure 2. The system restricts functionalities into three distinct roles:

1. Pilgrim Staff: Manages (Create, Read, Update, Delete) departure data, package data, and departure types.

2. Operational Manager: Views data, executes model training/retraining, configures automated schedules, and reviews prediction and evaluation outcomes.
3. President Director: Accesses, views, and downloads the final umrah pilgrim departure quota prediction reports.

3.3 Data Preprocessing

Table 2. Umrah depature data period 2022-2026

No	Tahun	Paket	Lama hari	Jumlah Jamaah	Jenis	Musim
1	2022	Paket 16 hari	16	35	Reguler	Low
2	2022	Paket 9 hari	9	22	Maulid	Peak
…	…	…	…	…	…	…
33	2026	Paket 12 hari	12	11	VIP	Low
34	2026	Paket 12 hari	12	10	VIP	Low

The data preprocessing phase prepares the historical umrah departure data in Table 2. for machine learning training. This process includes data loading and cleaning by removing missing values, transforming categorical variables (package name, package type, and season) into numerical values using Label Encoding, selecting the independent features (year, package name, duration, departure type, and season) and the target variable (total pilgrims), and finally splitting the dataset into 70% training data and 30% testing data using the train_test_split method for model evaluation.

3.4 Data Modeling and Evaluation

Following the preprocessing phase, the processed training dataset was used to train the predictive model using the RandomForestRegressor algorithm from the scikit-learn library. After the training process was completed, the model was evaluated using the remaining 30% of the dataset as testing data. To assess its performance, three evaluation metrics were calculated: Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). The evaluation results produced an MAE value of 3.265 and an MSE value of 15.186, indicating a relatively low prediction error. Furthermore, the R^2 value of 0.707 demonstrates that the model explains 70.7% of the variance in the dataset, reflecting strong predictive capability as presented in Table 1. However, these results may change over time as additional data are continuously incorporated into the umrah departure dataset.

3.5 User Interface and Implementation

This section discusses the visual and functional deployment of the application's main features, which encompass login, departure data, and prediction.

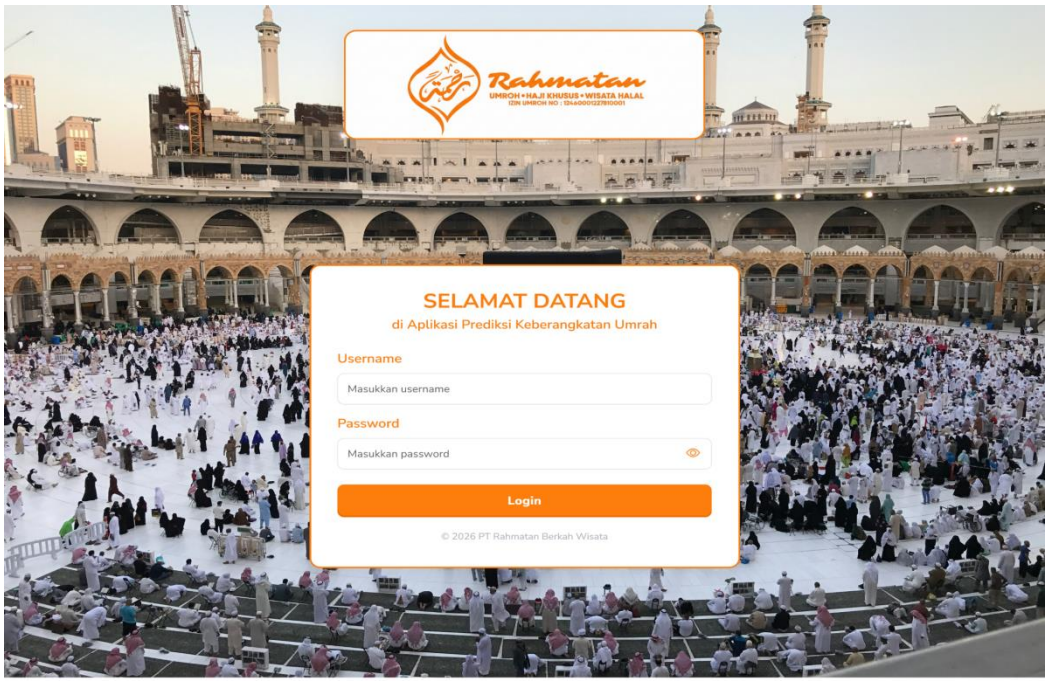

Figure 3. Login Page Application

The login page serves as the gateway to secure the application, restricting access to three authorized user roles (Operational Manager, President Director, and Pilgrim Staff) using credentials stored in the database. Users input their credentials into the login interface shown in Figure 3. Upon successful authentication, the system redirects them to their role-specific dashboard.

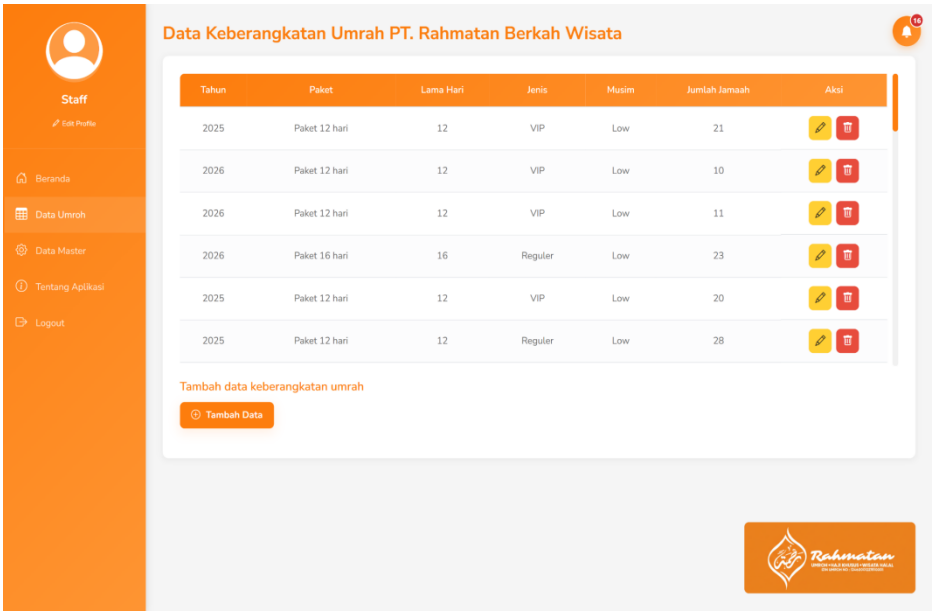

Figure 4. Umrah Departure Data Page

Figure 4. shows the Umrah Departure Data page, which can be accessed by Staff Jamaah. This page displays a table of historical umrah departure data, including the year, package, duration, type, season, and total number of pilgrims. Users can perform CRUD operations through the available action buttons. To add new data, users can click the “Tambah Data” button below the table, which will display the input form shown in the figure below.

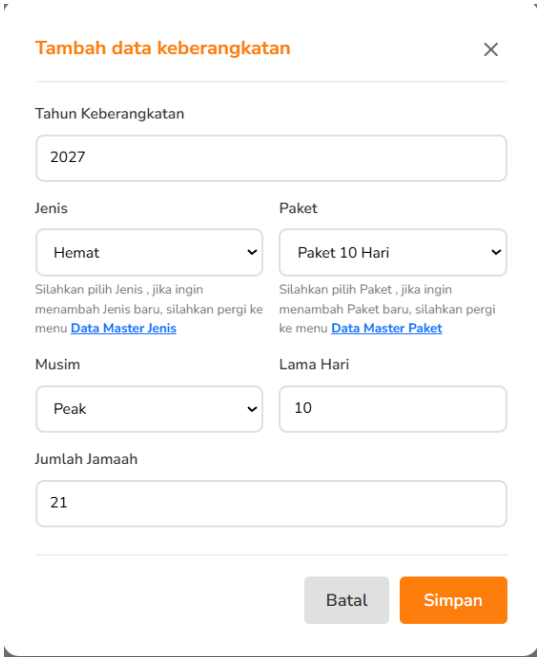

Figure 5. Add umrah departure data form

Figure 5. shows the Add Umrah Data form used to input new departure records into the system. The form includes fields for year, package, duration, type, season, and total number of pilgrims. If users need to add a new package or departure type that is not yet available, they can click the links provided in the package and type field descriptions, which will redirect them to the Master Package Data or Master Type Data pages for managing the required data.

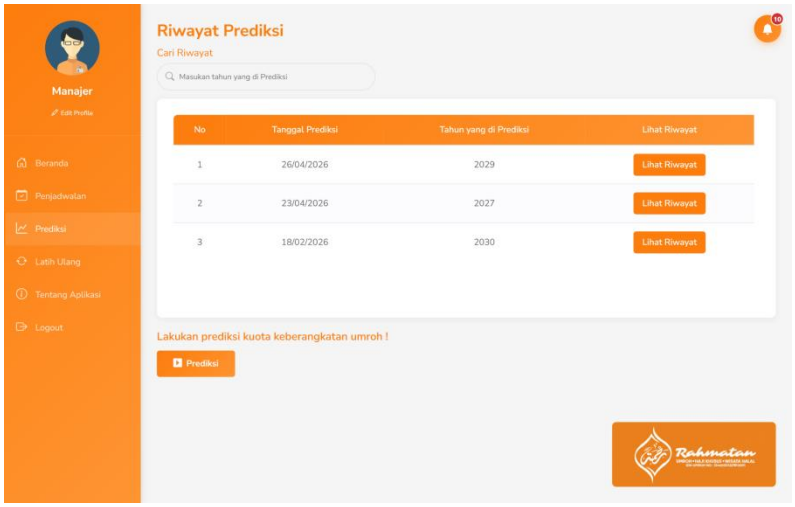

Figure 6. Predict Page

In Figure 6. shows the predict page that can be accessed by the Operations Manager which displays the prediction history table and the predict button, and if the user clicks on the predict button it will show the predict form like in Figure 7.

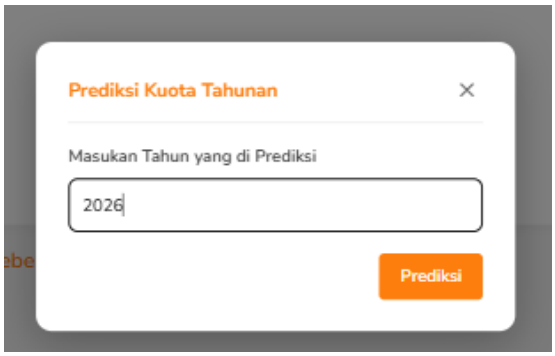

Figure 7. Modal of Predict Form

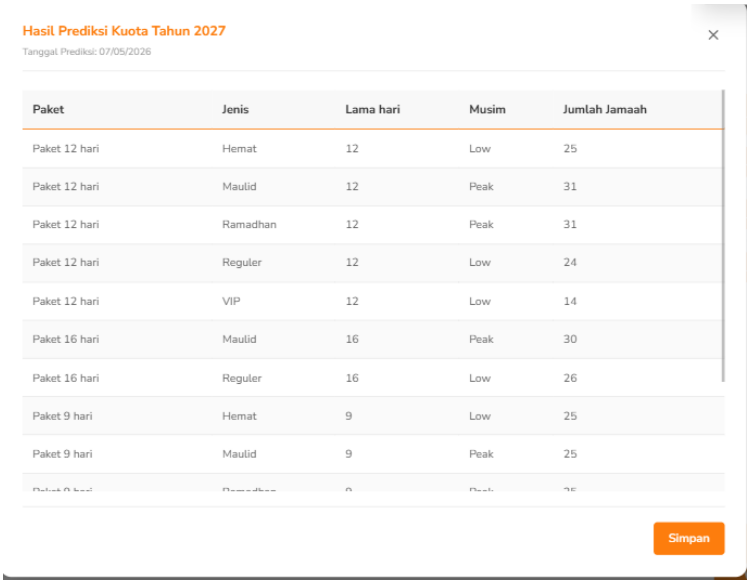

Figure 8. Modal of Predict Result Form

If the user already fills the predict form with the year of prediction, the predict results form will appear in the form of a table containing the package, type, duration, season and total pilgrims of umrah departure based on the year predicted by the user, as in Figure 8.

In addition to the main prediction and data management features, the application provides a Dashboard for data analysis, a Retraining feature to update the prediction model, and a Schedule feature for retraining and prediction reminders. The system also includes Master Data management for packages and departure types, downloadable Prediction Reports, and an About Application page containing user guides and tutorial videos.

3.6 Black Box Testing Result

The Black Box Testing results of the umrah departure quota prediction website were categorized into three user roles: President Director, Operations Manager, and Staff Jamaah. The testing involved actual users from PT. Rahmatan Berkah Wisata according to their respective access authorities. The results showed that all test cases were successfully executed and produced the expected outcomes, indicating that all website functions operated correctly, including data management, prediction, retraining, scheduling, reporting, and role-based access features. Therefore, the system successfully fulfilled all specified functional requirements.

CONCLUSION

This research successfully achieved all its objectives by developing a web-based Umrah departure quota prediction system using the Random Forest Regression algorithm as the Machine Learning model. The model demonstrated satisfactory performance with an MAE of 3.265, MSE of 15.186, and an R^2 value of 0.707, meaning the model was able to explain approximately 70.7% of the data patterns, making its predictions reasonably reliable based on historical data and the variables used. From a development standpoint, the system was built using the Rapid Application Development (RAD) method and successfully passed all Black Box Testing scenarios, confirming that every feature and function operates as intended. Academically, this study contributes to the application of Random Forest Regression in Umrah quota forecasting, while practically serving as a decision-support tool for PT. Rahmatan Berkah Wisata in planning departure pilgrims quotas. Future research may further improve prediction accuracy by incorporating additional relevant variables, expanding the historical dataset, and comparing Random Forest Regression with other machine learning algorithms to identify a more optimal approach.

REFERENCES

[1] A. R. Jawab and S. Salsabila, "PELAYANAN PRIMA DALAM MENINGKATKAN KEPUASAAN JAMAAH UMROH PADA PT. KHAZZANAH AL – ANSHARY TAHUN 2023," *Jurnal Manajemen Dakwah*, vol. 12, no. 1, pp. 141–167, Jun. 2024, doi: 10.15408/JMD.V12I1.39888.

[2] Zulkifli, Muhammad Arif, and Zulfadli Hamzah, "Analisis Pelayanan Ibadah Umroh Pada Perusahaan Jasa Penyelenggara Umroh Di Kota Pekanbaru," *Syarikat: Jurnal Rumpun Ekonomi Syariah*, vol. 5, no. 1, 2022, doi: 10.25299/syarikat.2022.vol5(1).9887.

[3] V. A. Lestari, A. Y. Ananta, and P. Basudewa, "Sistem Informasi Prediksi Persediaan Obat DI Apotek Naylun Farma Menggunakan Holt-Winters," *JIP (Jurnal Informatika Polinema)*, vol. 9, no. 2, pp. 229–236, Feb. 2023, doi: 10.33795/JIP.V9I2.1289.

[4] N. Amalia, "Application of the Random Forest Algorithm for Predicting Hajj Registration Numbers at Kemenag Lhokseumawe," *Jurnal Elektronika dan Teknologi Informasi*, vol. 5, no. 2, pp. 52–58, Sep. 2024, doi: 10.5201/JET.V5I2.487.

[5] Y. Christian, "Rental Price Prediction of Boarding Houses in Batam City Using Linear Regression and Random Forest Algorithms," 2023. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[6] R. Verdiyanto, D. Hartanti, and E. Purwanto, "Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest Regression," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 128–139, Jan. 2025, doi: 10.29408/JIT.V8I1.28386.

[7] E. S. Lestari and I. Astuti, "Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat," *Jurnal Ilmiah FIFO*, vol. 14, no. 2, pp. 131–146, Nov. 2022, doi: 10.22441/FIFO.2022.V14I2.003.

[8] D. Manurung, B. Zealtiel, and A. H. Lubis, "Prediksi Produksi Tanaman Padi di Indonesia dengan Menggunakan Algoritma Random Forest Regressor," *Journal of Computing and Informatics Research*, vol. 4, no. 3, p. 345, 2025, doi: 10.47065/comforch.v4i3.2125.

[9] J. Fahmi Idris *et al.*, "PENGUJIAN FUNGSIONAL DAN STRUKTURAL APLIKASI PENGAJUAN CUTI DENGAN METODE BLACK BOX DAN WHITE BOX," *Naratif : Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, vol. 7, no. 1, pp. 84–102, Jul. 2025, doi: 10.53580/NARATIF.V7I1.315.

[10] A. Putri, Z. Wulansari, and U. Mawaddah, "Prediksi Stok Barang Pada Toko Gudang Cangkring Menggunakan Metode Rapid Application Development (RAD)," *Jurnal*

Komputer, Informasi dan Teknologi, vol. 5, no. 2, pp. 1–17, Oct. 2025, doi: 10.53697/JKOMITEK.V5I2.3065.

[11] A. Rivaldi, "METODE PENGUMPULAN DATA MELALUI WAWANCARA," *Uinsu*, Jan. 2024, Accessed: Nov. 10, 2025. [Online]. Available: https://www.academia.edu/121718606/METODE_PENGUMPULAN_DATA_MELALUI_WAWANCARA

[12] S. W. Ramdany, S. Aulia Kaidar, B. Aguchino, C. Amelia, A. Putri, and R. Anggie, "Penerapan UML Class Diagram dalam Perancangan Sistem Informasi Perpustakaan Berbasis Web," *Journal of Industrial and Engineering System*, vol. 5, no. 1, pp. 30–41, Jul. 2024, doi: 10.31599/2E9AFP31.

[13] D. Aqsha, "PERBANDINGAN KINERJA ALGORITMA EXTREME GRADIENT BOOSTING DAN RANDOM FOREST UNTUK PREDIKSI HARGA RUMAH DI JABODETABEK," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 13, no. 1, Jan. 2025, doi: 10.24912/JIKSI.V13I1.32863.

[14] A. R. P. Masita, P. Dina, and R. M. Atok, "ANALISIS PERBANDINGAN METODE DECISION TREE REGRESSION DAN RANDOM FOREST REGRESSION PADA PREDIKSI KEPESERTAAN JAMINAN KESEHATAN NASIONAL," MATHunesa: Jurnal Ilmiah Matematika, vol. 13, no. 2, pp. 291–301, Aug. 2025, Accessed: Jan. 18, 2026. [Online]. Available: https://ejournal.unesa.ac.id/index.php/mathunesa/article/view/65473

[15] I. ,Gede Purnawinadi *et al.*, *Manajemen dan Analisis Data Penelitian Kuantitatif Kesehatan*. Yayasan Kita Menulis, 2023. Accessed: Jan. 14, 2026. [Online]. Available: https://www.researchgate.net/publication/374674743_Manajemen_dan_Analisis_Data_Penelitian_Kuantitatif_Kesehatan

[16] D. J. Evan and P. O. N. Saian, "IMPLEMENTASI PYTHON FRAMEWORK FLASK PADA MODUL TRANSFER OUT TOKO DI PT XYZ," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, pp. 1121–1131, Nov. 2023, doi: 10.29100/jipi.v8i4.4020.

[17] M. D. Firmansyah and Herman, "Analisa dan Perancangan Web E-Commerce Berbasis Website pada Toko Ida Shoes," *Journal of Information System and Technology*, vol. 02, no. 03, pp. 62–76, 2021.

[18] J. Shadiq, A. Safei, and R. W. R. Loly, "Pengujian Aplikasi Peminjaman Kendaraan Operasional Kantor Menggunakan BlackBox Testing," *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS : Journal of Information Management*, vol. 5, no. 2, p. 97, Jul. 2021, doi: 10.51211/imbi.v5i2.1561.