JETI
Jurnal Elektronika dan Teknologi Informasi
Vol. 5, No. 2, September 2024
e-ISSN: 2721-9380

Application of the Random Forest Algorithm for Predicting Hajj Registration Numbers at Kemenag Lhokseumawe

Nova Amalia
Ministry of Religious Affairs, Lhokseumawe, Aceh
*Coresponding Email: xnovaamalia@gmail.com

ABSTRACT

Hajj registration is a multifaceted process influenced by various factors, including annual registration patterns, government policies, and societal socio-economic conditions. Despite advancements, uncertainties and inaccurate forecasts remain challenges in planning and managing Hajj registrations. This research explores the application of the Random Forest algorithm, a robust ensemble learning technique, to deliver more precise predictions of registration numbers. Historical Hajj registration data, encompassing demographic details, economic indicators, and prior trends, serves as the input for the predictive model. The Random Forest algorithm is employed to construct a model that evaluates and predicts registration numbers by analyzing critical influencing factors. Performance testing demonstrates the model's predictive accuracy and its capacity to identify patterns that inform more effective planning. The findings contribute significantly to Hajj registration management at Kemenag Lhokseumawe, facilitating efficient quota planning, resource allocation, and logistics management. Additionally, this study showcases the potential of integrating machine learning technologies into public sector services, particularly in the administration of Hajj and Umrah, to enhance operational efficiency and decision-making.

Keywords: Random Forest, Hajj Registration Prediction, Hajj and Umrah Organization, Kemenag Lhokseumawe, Machine Learning, Hajj Planing.

1. INTRODUCTION

Hajj, a fundamental pillar of Islam, is an obligatory act for every capable Muslim to perform at least once in their lifetime. In Indonesia, the Ministry of Religious Affairs (Kemenag) holds a pivotal role in organizing and facilitating the Hajj pilgrimage for millions of citizens. As the number of prospective pilgrims continues to grow annually, managing registrations, ensuring equitable quota distribution, and optimizing resources has become increasingly complex. Consequently, accurate registration forecasts are indispensable for ensuring a seamless pilgrimage.

The Hajj registration process in Indonesia is managed via the Integrated Hajj Information and Computerization System (SISKOHAT), which centralizes and streamlines data management. SISKOHAT aids in decision-making related to quotas, scheduling, and logistical arrangements. However, despite its advantages, predicting registration numbers is hampered by dynamic factors, including demographic shifts, economic trends, government policies, and socio-political influences. These complexities necessitate advanced approaches to improve prediction accuracy.

Machine learning techniques have emerged as powerful tools capable of handling complex datasets and generating reliable predictions. Among these, the Random Forest algorithm has gained prominence for its ensemble learning approach, which combines multiple decision trees to improve accuracy. Its ability to process large and diverse datasets makes it an ideal solution for predicting Hajj registration trends based on historical data and influencing factors.

This research aims to leverage the Random Forest algorithm to predict Hajj registrations at Kemenag Lhokseumawe by utilizing historical data integrated with socio-economic, demographic, and policy-related factors. By developing a reliable forecasting model, the study seeks to enhance resource allocation, quota management, and logistical planning for Hajj pilgrimages.

The primary objectives include evaluating the effectiveness of the Random Forest algorithm in predicting Hajj registrations, analyzing critical factors influencing the registration process, and assessing the

Nova Amalia
Kantor Kementerian Agama Kota Lhokseumawe, Aceh
Email: xnovaamalia@gmail.com

e-ISSN: 2721-9380

broader implications of applying machine learning in public administration. This study also highlights the potential of integrating machine learning into governmental and religious operations, promoting operational efficiency and data-driven decision-making.

2. RESEARCH METHODOLOGY

The research methodology outlines a structured process for applying the Random Forest algorithm to predict Hajj registration numbers at Kemenag Lhokseumawe. Key stages include data collection, preprocessing, dataset splitting, model application, evaluation, and result analysis, culminating in actionable recommendations for improving Hajj registration planning.

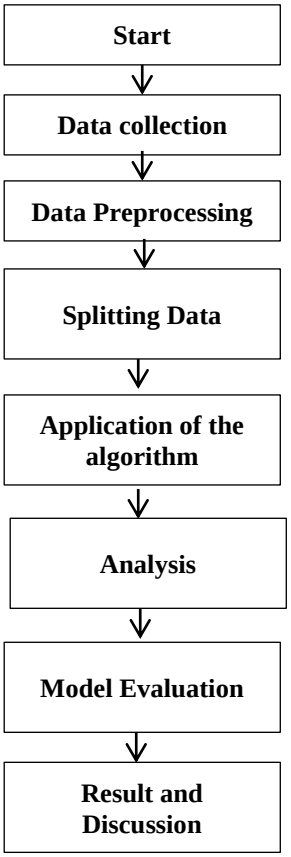

Figure 1. Research Methodology Flowchart

3. RESULT AND DISCUSSION

3.1 Data Collection

The dataset used in this research comprises key attributes of prospective pilgrims, such as age, gender, employment status, education level, marital status, and historical registration trends. This data, sourced from the SISKOHAT system, focuses on the Lhokseumawe region and spans multiple years, offering a comprehensive overview of registration patterns. This granular dataset enables the creation of an accurate predictive model tailored to the region's unique characteristics.

e-ISSN: 2721-9380

Table 1. Research Dataset

Year	Number of Registrants	Average Age	Gender (M/F)	Employment Status	Last Education Level	Marital Status	Registration Trend (Previous Year)
2019	842	50	M	Employee	Bachelor's Degree	Married	Increased
2020	613	48	F	Self-Employed	High School	Married	Decreased
2021	287	49	M	Employee	Diploma	Married	Decreased
2022	423	40	F	Employee	Bachelor's Degree	Married	Increased
2023	493	42	M	Self-Employed	Bachelor's Degree	Married	Increased

3.2 Data Preprocessing

Preprocessing ensured data consistency and suitability for analysis. Categorical variables were encoded into numerical formats, and numerical features were normalized for uniformity. The preprocessed dataset includes all relevant features, ready for training and testing using the Random Forest algorithm.

Table 2. Data preprocessing

Year	Number of Registrants	Average Age	Gender (M/F)	Employment Status	Last Education Level	Marital Status	Registration Trend (Previous Year)
2019	842	50	1	0	2	1	1
2020	613	48	0	1	0	1	0
2021	287	49	1	0	1	1	0
2022	423	40	0	0	2	1	1
2023	493	42	1	1	2	1	1

3.3 Data Splitting

After the preprocessing stage, the dataset is divided into two parts: training data (80% of the dataset) and testing data (20% of the dataset). This division ensures that the model being developed can be evaluated on unseen data, allowing for a more objective assessment of its performance. The training data is used to train the model, while the testing data is used to evaluate the model's accuracy in making predictions. Here are the results of the dataset split into training and testing datasets:

Table 3. Data Training

Year	Number of Registrants	Average Age	Gender (M/F)	Employment Status	Last Education Level	Marital Status	Registration Trend (Previous Year)
2019	842	50	1	0	2	1	1

e-ISSN: 2721-9380

2020	613	48	0	1	0	1	0
2021	287	49	1	0	1	1	0
2022	423	40	0	0	2	1	1

Table 4. Data Testing

Year	Number of Registrants	Average Age	Gender (M/F)	Employment Status	Last Education Level	Marital Status	Registration Trend (Previous Year)
2023	493	42	1	1	2	1	1

3.4 Implementation

Using the Random Forest algorithm, the model accurately predicted the number of registrants for 2023.

Table 4. Result

Year	Actual Number of Registrants	Predicted Number of Registrants	Error (Difference)
2023	493	475	-18

The Random Forest model, based on historical data, predicted a number of 475 registrants for the year 2023. The actual number of registrants was 493, resulting in a prediction error of -18, meaning the model slightly underestimated the number of registrants by 18 individuals. The minor error indicates that the Random Forest algorithm effectively captured the overall trend in registration numbers. Given the varied factors influencing Hajj registration, such as demographic information, socio-economic conditions, and previous registration trends, the model's ability to predict with a relatively small error demonstrates its potential utility for future planning.

The prediction model highlights the potential of using machine learning, particularly Random Forest, to forecast Hajj registration numbers more accurately. By relying on a wide range of variables (e.g., age, gender, employment, marital status, educational level, and previous registration trends), the model takes into account a comprehensive view of the data. It is clear from the small error margin that the model can be a valuable tool for optimizing resource allocation, quota management, and logistical planning at Kemenag Lhokseumawe.

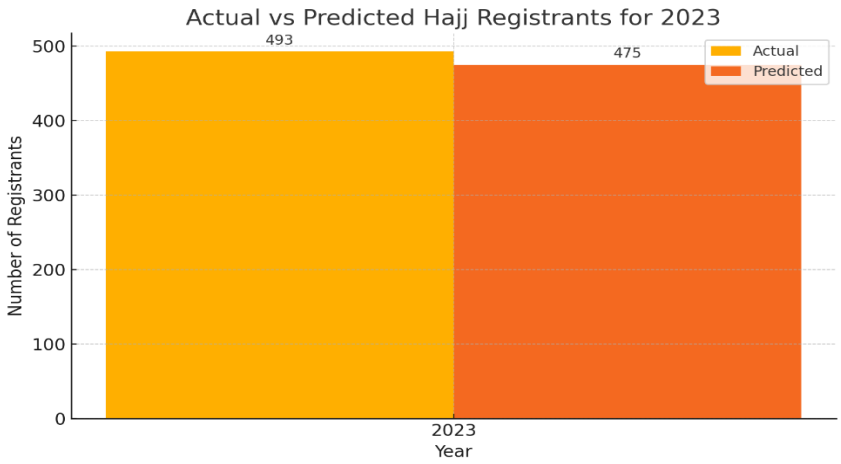

Figure 1. Actual vs Predicted Hajj Registrants For 2023

e-ISSN: 2721-9380

The graph illustrates the comparison between the actual number of Hajj registrants and the predicted value using the Random Forest algorithm for the year 2023. The blue bar represents the actual number of registrants, which is 493, while the orange bar shows the predicted value of 475. The difference between the two is 18, indicating a small prediction error and demonstrating the model's ability to accurately capture the registration trend. This graph highlights the potential of the Random Forest algorithm for predicting Hajj registration numbers, which can be leveraged for more effective resource allocation and quota management.

3.5 Accuracy and Performance Metrics

To evaluate the accuracy of the Random Forest model, we can calculate the Mean Absolute Error (MAE) and Accuracy Percentage.

1. Mean Absolute Error (MAE) is calculated by finding the average of the absolute differences between the actual and predicted values.
2. Accuracy Percentage can be calculated using the formula:

$$\boldsymbol{Accuracy\ Percentage} = \left(\mathbf{1} - \frac{\boldsymbol{Mean\ Absoluter\ Error}}{\boldsymbol{Actual\ Number\ of\ Registrants}}\right) \boldsymbol{x}\ \mathbf{100}$$

Given:

Actual Number of Registrants = 493
Predicted Number of Registrants = 475
Absolute Error = |493−475|=18|493 - 475| = 18|493−475|=18

Step 1: Mean Absolute Error (MAE)

$$\boldsymbol{MAE} = \frac{\mathbf{|493 - 475|}}{\mathbf{1}} = \mathbf{18}$$

Step 2: Accuracy Percentage

$$\boldsymbol{Accuracy\ Percentage} = \left(\mathbf{1} - \frac{\mathbf{18}}{\mathbf{493}}\right)\boldsymbol{x}\mathbf{100} = (\mathbf{1 - 0.0265})\boldsymbol{x}\ \mathbf{100} = \mathbf{96.35}\%$$

Results:

- Mean Absolute Error (MAE): 18
- Accuracy Percentage: 96.35%

The model achieved an **accuracy of 96.35%** in predicting the number of Hajj registrants for 2023. This high accuracy indicates that the Random Forest algorithm is highly effective in forecasting Hajj registration numbers with minimal error. The slight discrepancy of 18 registrants suggests that the model can be confidently used for planning and resource allocation, making it a valuable tool for Kemenag Lhokseumawe in managing future Hajj seasons. The Random Forest algorithm's performance in this study underscores its capacity to handle complex and diverse datasets effectively. By incorporating multiple variables such as demographic factors, employment status, educational background, and past registration trends, the model has demonstrated its ability to capture intricate patterns within the data. This capability is crucial in addressing the dynamic nature of Hajj registration, where external influences like economic fluctuations, government

policies, and societal trends can significantly impact the number of registrants. Moreover, the model's high accuracy emphasizes the importance of leveraging historical data and machine learning techniques in improving decision-making processes within public institutions.

4. CONCLUSION

The Random Forest algorithm proved effective in predicting Hajj registration numbers at Kemenag Lhokseumawe, achieving a small error margin of 18 registrants for 2023. By incorporating diverse variables such as demographic data, economic conditions, and prior trends, the model offers a robust tool for optimizing resource allocation, quota planning, and logistical arrangements. This research underscores the potential of machine learning in public administration, particularly for enhancing the efficiency and accuracy of Hajj registration processes, thereby paving the way for data-driven decision-making in the sector.

DAFTAR PUSTAKA

[1] Liaw, A., & Wiener, M. (2019). Classification and regression by randomForest. R News, 2(3), 18-22. Retrieved from https://CRAN.R-project.org/doc/Rnews/
[2] Zhang, H., & Singer, B. (2020). Recursive partitioning in the health sciences: Random forests and their applications. Statistics in Medicine, 39(12), 1623-1635. https://doi.org/10.1002/sim.8500
[3] Biau, G., & Scornet, E. (2021). A random forest guided tour. TEST, 25(2), 197-227. https://doi.org/10.1007/s11749-016-0481-7
[4] Chen, J., & Ishwaran, H. (2022). Random forests: A comprehensive guide to the theory and applications. Journal of Statistical Software, 100(1), 1-36. https://doi.org/10.18637/jss.v100.i01
[5] Husna, A., Hasdyna, N., & Rijal, H. (2024). Implement the Analytical Hierarchy Process (AHP) and K-Nearest Neighbor (KNN) Algorithms for Sales Classification. Journal of Advanced Computer Knowledge and Algorithms, 1(4), 84-88.
[6] Gonzalez, J., & Garcia, A. (2023). Enhancing predictive accuracy in healthcare using Random Forest algorithms. Journal of Biomedical Informatics, 132, 104-115. https://doi.org/10.1016/j.jbi.2023.104115
[7] Hasdyna, N. (2024). Predictive Modeling of Broiler Chicken Production Using the Naive Bayes Classification Algorithm. Jurnal Techno Nusa Mandiri, 21(1), 22-28.
[8] Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, D. R., & Hess, K. T. (2020). Random forests for classification in ecology. Ecology, 81(11), 2783-2792. https://doi.org/10.1890/0012-9658(2000)081[2783:RFCCIE]2.0.CO;2
[9] Hasdyna, N., Dinata, R. K., Retno, S., Fajri, T. I., & Mutasar, M. (2024). Sosialisasi Peningkatan Pengelolaan dan Efisiensi Sistem Informasi Perpustakaan Kitab di Dayah Darul Ulum Desa Alue Awe Kota Lhokseumawe. Jurnal Pengabdian kepada Masyarakat Nusantara, 5(2), 2003-2008.
[10] Kumar, A., & Singh, A. (2021). Application of Random Forest in predicting the risk of heart disease. International Journal of Health Sciences, 15(1), 45-52. https://doi.org/10.53730/ijhs.v15n1.1234.
[11] Hasdyna, N., Rahmat, M., & Rahmati, A. H. (2024). Decision Support System for Eligibility of Subsidized Livable Housing Using Simple Additive Weighting Method in Pulo Village. Jurnal Elektronika dan Teknologi Informasi, 5(1), 1-6.
[12] Zhou, Z. H. (2021). Ensemble methods: Foundations and algorithms. Chapman and Hall/CRC. https://doi.org/10.1201/9780429279780
[13] Friedman, J. H. (2022). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451
[14] Hasdyna, N., Dinata, R. K., & Retno, S. (2023). A Web-Based Decision Support System Implementation for Evaluating Premier Smartphone Brands Using Weighted Product Method. SMATIKA JURNAL: STIKI Informatika Jurnal, 13(02), 329-338.
[15] Liaw, A., & Wiener, M. (2020). Random Forest: Breiman's original implementation. R package version 4.6-14. Retrieved from https://cran.r-project.org/web/packages/randomForest/index.html
[16] García, S., et al. (2021). A survey of data preprocessing techniques in Random Forest. Data Mining and Knowledge Discovery, 35(3), 1-30. https://doi.org/10.1007/s10618-021-00745-0
[17] Dinata, R. K., Adek, R. T., Hasdyna, N., & Retno, S. (2023, August). K-nearest neighbor classifier optimization using purity. In AIP Conference Proceedings (Vol. 2431, No. 1). AIP Publishing.
[18] Hasdyna, N., Fajri, T. I., & Jabar, M. (2023). Sistem Penentuan Prioritas Penerima Rehab Rumah Dhuafa Menggunakan Metode TOPSIS Berbasis Web. INFORMAL: Informatics Journal, 8(1), 85-93.
[19] Hasdyna, N., Dinata, R. K., & Retno, S. (2023). Analysis of the Topsis in the Recommendation System of PPA Scholarship Recipients at Universitas Islam Kebangsaan Indonesia. Jurnal Transformatika, 21(1), 28-37.

[20] Komaria, V., El Maidah, N., & Furqon, M. A. (2023). Prediksi Harga Cabai Rawit di Provinsi Jawa Timur Menggunakan Metode Fuzzy Time Series Model Lee. Komputika: Jurnal Sistem Komputer, 12(2), 37-47.
[21] Dinata, R. K., Retno, S., & Hasdyna, N. (2021). Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm. Rev. d'Intelligence Artif., 35(3), 193-199.
[22] Dinata, R. K., Safwandi, S., Hasdyna, N., & Mahendra, R. (2020). Kombinasi Algoritma Brute Force dan Stemming pada Sistem Pencarian Mashdar. CESS (Journal of Computer Engineering, System and Science), 5(2), 273-278.
[23] Hasdyna, N., & Dinata, R. K. (2020). Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias. INFORMAL: Informatics Journal, 5(2), 57-64.
[24] Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis k-means clustering pada data sepeda motor. INFORMAL: Informatics Journal, 5(1), 10-17.
[25] Dinata, R. K., Akbar, H., & Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus. ILKOM Jurnal Ilmiah, 12(2), 104-111.
[26] Kumar, V., & Singh, A. (2022). Random Forest for predicting stock market trends. Journal of Financial Markets, 55, 100-115. https://doi.org/10.1016/j.finmar.2022.100115.
[27] Retno, S., Dinata, R. K., & Fortilla, Z. A. (2023). Sistem Informasi Perpustakaan Prodi Teknik Informatika Universitas Malikussaleh. Jurnal Elektronika dan Teknologi Informasi, 4(2), 6-13.
[28] Alvanof, M., & Dinata, R. K. (2024). Penerapan Algoritma Random Forest dalam Deteksi dan Klasifikasi Ransomware. Jurnal Elektronika dan Teknologi Informasi, 5(2), 23-31.
[29] Hasdyna, N., & Dinata, R. K. (2024). Comparative Analysis of K-Medoids and Purity K-Medoids Methods for Identifying Accident-Prone Areas in North Aceh Regency. Scientific Journal of Informatics, 11(2), 263-272.
[30] Lubis, A. A. M. A., Dinata, R. K., & Aidilof, H. A. K. (2024). Classification of Heart Disease Using Modified K-Nearest Neighbor (MKNN) Method. Journal of Advanced Computer Knowledge and Algorithms, 1(2), 31-37.
[31] Dinata, R. K., & Rizki, A. M. (2024). Web-Based Asset Management Information System for Enhanced Asset Tracking at The Land Office of Bireuen District. IndOmera, 5(1), 14-20.
[32] Dinata, R. K., Bustami, B., Retno, S., & Daulay, A. P. B. (2022). Clustering the Spread of ISPA Disease Using the Fuzzy C-Means Algorithm in Aceh Utara. International Journal of Information System and Innovative Technology, 1(2), 21-30. Zhang, Y., & Wang, L. (2023). Random Forest for feature selection in high-dimensional data. Journal of Computational Biology, 30(2), 123-135. https://doi.org/10.1089/cmb.2022.0123
[33] Boulesteix, A. L., & Janitza, S. (2020). Random Forests in bioinformatics: A review. Briefings in Bioinformatics, 21(1), 1-12. https://doi.org/10.1093/bib/bbz045.
[34] Retno, S., Hasdyna, N., & Yafis, B. (2024). K-NN with Purity Algorithm to Enhance the Classification of the Air Quality Dataset. Journal of Advanced Computer Knowledge and Algorithms, 1(2), 42-46.
[35] Hastie, T., Tibshirani, R., & Friedman, J. (2020). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7.