Khazanah Journal of Religion and Technology
Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
https://journal.uinsgd.ac.id/kjrt
Online ISSN: 2987-6060
Volume 2 Issue 1 (2024)
DOI: https://doi.org/10.15575/kjrt.v2i1.227

Optimizing Qur'an Recitation Monitoring with Random Forest Algorithm

Ivan Wijayana
Department of Informatics
UIN Sunan Gunung Djati Indonesia
Bandung, Indonesia
ivanwijayana@gmail.com

Muhammad Ardiansyah
Department of Informatics
UIN Sunan Gunung Djati Indonesia
Bandung, Indonesia
ardi1035@gmail.com

Abstract— Every Muslim should improve their worship, including reciting the Al-Qur'an. Reciting the Al-Qur'an is a deep spiritual practice that provides spiritual benefits and blessings daily. To maximize the benefits of reciting Al-Qur'an during this holy month, implementing Machine Learning can make a significant contribution. This research explores the application of Machine Learning using the Random Forest algorithm to improve the practice of reciting the Al-Qur'an. By collecting data through surveys using questionnaires, this research identifies important factors that influence an individual's success in completing reading the Al-Qur'an. The research results show that the Random Forest algorithm can be used to predict the number of individuals who have the potential to complete reciting Al-Qur'an with an accuracy value of 80%.

Keywords- Classification, Machine Learning, Random Forest, Reciting Al-Qur'an.

I. INTRODUCTION

Reciting and understanding the Al-Qur'an is a very important practice for Muslims [1], [2], [3]. In an effort to improve the practice of reciting the Al-Qur'an, it is necessary to collect relevant data to understand the factors that contribute to individual success in completing reading the Al-Qur'an (khatam). Currently, understanding of the influence of factors such as reading fluency, daily reading frequency, and the habit of reading after prayer on a person's possibility of completing reading the Qur'an (*khatam*) needs to be improved. Therefore, this research aims to fill this knowledge gap and provide a more comprehensive understanding of the factors that play a role in achieving the objectives of reciting Al-Qur'an.

This research has great importance because by utilizing Machine Learning techniques, especially the Random Forest algorithm, we can predict the number of individuals who will successfully complete reading the Al-Qur'an based on the factors collected. This will provide better insight and understanding of the factors that influence the success of reciting the Al-Qur'an. The results of these predictions can be used to develop more effective strategies in increasing individual participation and success in reciting the Al-Qur'an.

Machine learning technology, as the process of making computers intelligent [4], [5], can be utilized optimally to help predict an individual's readiness to complete their Al-Qur'an reading. Therefore, this research focuses on developing a Machine Learning model using the Random Forest algorithm to predict the number of people who will successfully complete the khatam of the Al-Qur'an. Random Forest is a development of the decision tree algorithm by building many tree models so as to reduce the possibility of overfitting [6], [7]. The Random Forest algorithm is proven to have good performance in classification or prediction tasks.

The dataset collected will include information such as name, reading fluency, daily reading frequency, total *Juz* read per day, whether read after prayer or not and the number of khatam that the individual has achieved. The model developed will take these data as input and provide predictions about the number of individuals who have the potential to complete the khatam of the Qur'an based on the observed factors. Prediction results from the developed model can provide insight information that can help design more effective strategies in increasing individual participation and success in reciting the Al-Qur'an. Apart from that, this research also has the potential to encourage the application of technology in a religious context which can provide benefits and positive impacts on Muslims in carrying out the recitation of the Al-Qur'an.

II. RELATED WORKS

Some previous research related to this research includes:

1. Application of machine learning for voice recognition techniques to identify errors in reading the Al-Qur'an. The methodology they use focuses on analyzing voice features such as intonation, clarity and rhythm. This approach is similar to our research in terms of analyzing voice recordings for evaluating Al-Qur'an tadarus, however, we also consider other factors such as reading fluency and daily reading frequency [8].
2. Research on analyzing the recitation of the Al-Qur'an using Natural Language Processing (NLP) techniques. They utilized language processing methods to analyze the Qur'anic text and identify recitation errors. Although their methodology is different from our research which focuses on sound analysis, both aim to improve the quality of Al-Qur'an recital using the latest technology [9], [10], [11].
3. Research on the use of Virtual Reality (VR) in learning the Al-Qur'an. They developed an immersive virtual environment to help users read and understand the Al-Qur'an. Their research differs from ours in terms of the approach used, but both seek to improve the learning experience in the context of reciting Al-Qur'an [12].
4. Development of Machine Learning-based mobile applications to assist individuals in studying the Al-Qur'an. They used a classification algorithm to distinguish between correct and incorrect letters of the Qur'an in reading. Their methodology is comparable to our research in terms of using Machine Learning techniques, but we focused on predicting the number of people who khatam based on factors such as daily reading frequency and reading habits after prayers [13], [14].
5. Development of Augmented Reality (AR) technology in Al-Qur'an learning. They developed an AR application that allows users to view the written Al-Qur'an expanded with tajweed explanations and translations. Even though their approach is different from this research which uses Machine Learning, both are trying to increase understanding and experience in reciting the Al-Qur'an with the help of technology [15].

The related research mentioned above shows various approaches to improve Al-Qur'an recitation using the latest technology such as Machine Learning, Natural Language Processing, Augmented Reality and Virtual Reality. Although the methodologies used are different, all of this research has the same goal, namely improving the quality of Al-Qur'an recitation and providing a better learning experience for individuals.

Like previous studies, this research adopts a machine learning approach using the Random Forest algorithm to predict the number of people who have the potential to finish reading the Al-Qur'an (*khatam*). This research methodology is different from previous related research which focused more on sound analysis, daily reading frequency, and other factors that influence the success of reciting the Al-Qur'an.

In developing our methodology, we took inspiration from approaches that have been applied in related research, such as the use of technology in Al-Qur'an learning. However, we also incorporated unique aspects in our research, namely the prediction of the number of people who died based on the factors collected.

III. RESEARCH METHODS

This research includes several stages, namely data collection, data preprocessing, modeling, and evaluation from the research presented in Figure 1. Data collection was carried out by conducting a survey through distributing questionnaires to respondents randomly. The questions in the questionnaire relate to the factors you want to research, such as reading fluency, daily reading frequency, total *Juz* read per day, whether you read after prayer or not, and the number of khatam you have achieved. There were 25 correspondents who filled out the survey.

Fig 1. Research Activities

After obtaining relevant data, data analysis was carried out in three main stages, namely data processing, modeling and evaluation. In the data preparation stage, an examination of the dataset used is carried out. This includes selecting relevant data, cleaning data to deal with empty and duplicate values, and data construction. Apart from that, this stage also involves labeling the data by converting the data type from string to integer.

Next, at the modeling and model evaluation stage, data visualization and correlation checking between each variable are carried out [16]. To identify an increase in Al-Qur'an Tadarus in the month of Ramadan, the Random Forest algorithm was used. The Random Forest (RF) method is a development of the Classification and Regression Tree (CART) method, which involves the use of bootstrap aggregating (bagging) and random feature selection [17], [18]. This method can increase accuracy by generating child nodes for each node (the nodes above it) and selecting features randomly. The classification results of each tree are accumulated, and the classification results that appear most frequently are selected [18]. This method consists of a root node, an internal node and a leaf node [19], [20].

The root node is the node located at the top of the Decision tree [21]. An internal node is a node that has two or more branches, with only one input. Meanwhile, the leaf node is the last node that only has one input and no output. The decision tree starts by calculating the entropy value as an indicator of the level of attribute diversity, and the information gain value is used for attribute selection [22]. To calculate the entropy value, the formula used is the same as in equation 1.

The following is the formula for Random Forest:

$$Entropy\ (Y) = -\sum_i p(c|Y) log^2 p(c|Y) \qquad (1)$$

Where Y is a set of cases, while P(c|Y) is the proportion of Y values in class c.

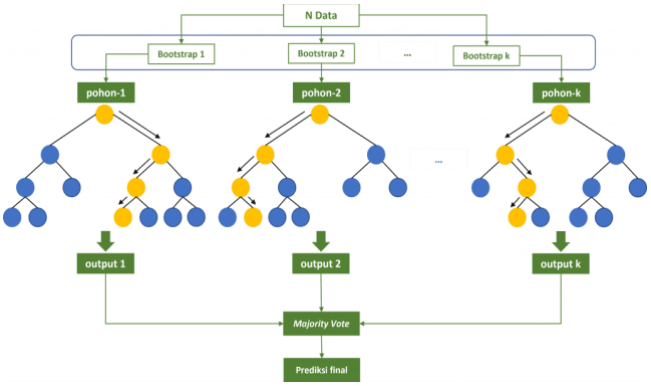

Fig 2. Example of Random Forest Model

Figure 2 is the structure of the Random Forest algorithm. Random Forest can be applied to regression and classification modeling. In the random forest model for regression, predictions are calculated based on the average value (averaging) of the output of each decision tree . Meanwhile, for classification models, predictions are determined using the majority vote [21], [23]. In this research, we use a classification model to provide predictions of completion in reading the Al-Qur'an.

IV. RESULT AND DISCUSSION

A. *Data Pre-processing Result*

	nama	pernyataan	kelancaran_baca	frekuensi_bacaharian	totaljuz_perhari	baca_setelahsolat	total_khatam
0	Ivan Wijayana	Ya	Lancar	5	1 Juz	Ya	1
1	Dani Kusuma	Ya	Sedang	3	< 1 Juz	Ya	1
2	Karina Ramaditha	Ya	Sangat Lancar	7	> 1 Juz	Ya	1
3	Tika Permatasari	Ya	Sangat Lancar	10	> 1 Juz	Ya	2
4	Abdillah Nuriman	Ya	Lambat	2	< 1 Juz	Ya	0
5	Alfin	Ya	Lancar	5	1 Juz	Ya	1
6	Mail Nurrohman	Ya	Sedang	1	< 1 Juz	Tidak	0
7	Indra Gumilar	Ya	Lancar	4	1 Juz	Ya	1
8	M Rizki P.S.	Ya	Sedang	3	< 1 Juz	Ya	0
9	Alan Nur Ramadhan	Ya	Sangat Lancar	7	> 1 Juz	Ya	2
10	Hidayah Nur Sa'adah	Ya	Sangat Lancar	8	> 1 Juz	Ya	3
11	Nurul Habibah Azzahra	Ya	Lancar	6	1 Juz	Ya	1
12	Fikri Aufa	Ya	Lambat	2	< 1 Juz	Tidak	0
13	Syerika Putri	Ya	Lancar	5	1 Juz	Ya	1
14	Aziz Muhlis	Ya	Lancar	6	> 1 Juz	Ya	1

Fig 3. Example of Random Forest Model

Before entering the modeling process, the data is prepared properly to produce a good model. Figure 3 is an example of the dataset that will be used in this research. This dataset has several attributes including name, statement, reading fluency, daily reading frequency, total *Juz* per day, reading after prayer, and also total khatam. Figure 4 is an implementation of a bar plot using the matplotlib library in Python. The bar graph will provide a visualization of the number of juz or people who have completed the tadarus of the Al-Qur'an. The well of the stem will represent each individual's name, while the height of the stem will represent the number of juz or khatam people.

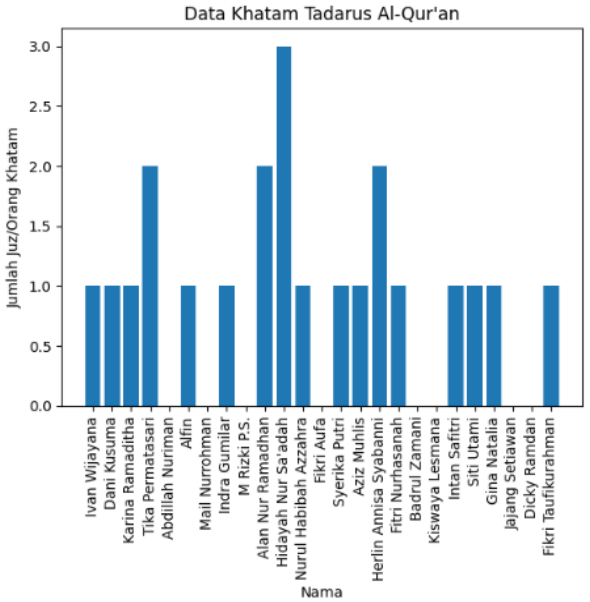

Fig 4. Amount of *Juz* per *khatam's person*

Data preparation for model training is done by separating features and labels, and also calculating the number of features. Figure 5 is a visualization of how much data is available for each feature in the dataset. The height of the bars on the graph indicates the amount of data present for each feature. By looking at this plot, an understanding of the data distribution between the features that will be used in the modeling process can be obtained.

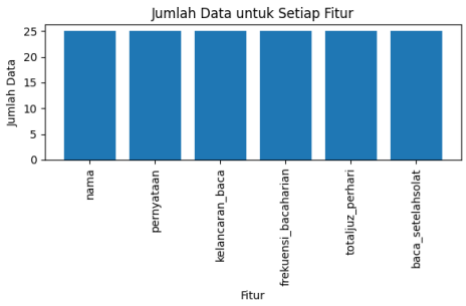

Fig 5. Example of Random Forest Model

B. *Modeling and Evaluation Result*

```
# Membangun model Random Forest
model = RandomForestClassifier(n_estimators=50, max_depth=3, random_state=42)

# Melatih model menggunakan data latih
model.fit(X_train, y_train)
```

```
RandomForestClassifier(max_depth=3, n_estimators=50, random_state=42)
```

```
# Melakukan prediksi pada data uji
y_pred = model.predict(X_test)

# Menghitung jumlah orang yang diprediksi khatam
jumlah_khatam = sum(prediksi == 1)  # Menghitung jumlah prediksi yang bernilai 1 (khatam)

# Menghitung akurasi
akurasi = model.score(X_test, y_test)
```

Fig 6. Example of Random Forest Model

Figure 6 shows the process of building a random forest model. This model will consist of 50 decision trees, and has a maximum depth of 3. Model training using training data and predictions on test data aims to measure the model's performance in predicting the number of people who have the potential to recite the Al-Qur'an. Figure 7 is a visualization that shows the number of people predicted to have read the Qur'an and the accuracy of the model in predicting. It can be seen from this visualization that the number of people who are predicted to have recited the Qur'an is four people, and the model has an accuracy of 80%.

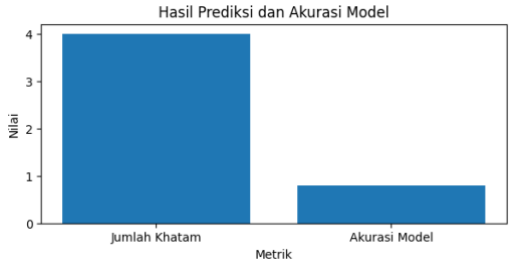

Fig 7. Example of Random Forest Model

C. Findings

The research results show that the use of the Random Forest algorithm can be used to predict the number of individuals who have the potential to complete Al-Qur'an tadarus (khatam) based on the factors that have been collected. In data collection, a survey was conducted using a questionnaire to collect information about reading fluency, daily reading frequency, total juz read per day, whether read after prayer or not, and the number of khatam that have been achieved.

After analyzing the data and training the Random Forest model, the prediction results showed that four people were predicted to successfully complete recite the Al-Qur'an based on the observed factors. The model also has an accuracy of 80%, which shows the model's ability to predict correctly.

Research findings based on the model produced in this study are that individuals who have a high level of reading fluency tend to be more successful in completing Al-Qur'an tadarus. Findings Then, the second hypothesis states that individuals who read the Al-Qur'an with high frequency every day have a greater chance of completing recite the Al-Qur'an.

V. CONCLUSION

Implementation of machine learning using the Random Forest algorithm has the potential to be an illustration of improving Al-Qur'an recitation. The research results show that factors such as reading fluency, daily reading frequency, and the habit of reading after prayer can influence the success of reciting the Al-Qur'an. Although this research has limitations, especially regarding the dataset used, these findings can contribute to increasing understanding and strategies to increase individual participation and success in reciting the Al-Qur'an. As for future research directions, there are still many things that can be learned. For example, expanding variables by considering other factors that can influence the success of reciting Al-Qur'an such as the social environment, family support, and learning methods used, and improving the model by considering the use of machine learning algorithms.

REFERENCES

[1] D. K. Wardani and M. S. N. Samsu, "Application of Tilawatil Qur'an Method to Improve the Ability Reading Al-Qur'an in Bahrul Ulum Islamic Boarding School," *SCHOOLAR: Social and Literature Study in Education*, vol. 2, no. 1, pp. 63–67, Jul. 2022, doi: 10.32764/schoolar.v2i1.1278.

[2] M. Alatiyyah, "Quran Reciter Identification: Techniques and Challenges," *J Theor Appl Inf Technol*, vol. 101, no. 16, 2023.

[3] Y. A. Gerhana *et al.*, "Computer speech recognition to text for recite Holy Quran," in *IOP Conference Series: Materials Science and Engineering*, 2018. doi: 10.1088/1757-899X/434/1/012044.

[4] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer," *Behav Ther*, vol. 51, no. 5, pp. 675–687, Sep. 2020, doi: 10.1016/j.beth.2020.05.002.

[5] M. O. Riedl, "Human-centered artificial intelligence and machine learning," *Hum Behav Emerg Technol*, vol. 1, no. 1, pp. 33–36, 2019, doi: https://doi.org/10.1002/hbe2.117.

[6] S. Polamuri, "How the random forest algorithm works in machine learning," *Retrieved December*, vol. 21, 2017.

[7] Y. L. Pavlov, *Random forests*. Walter de Gruyter GmbH & Co KG, 2019.

[8] A. M. Alagrami and M. M. Eljazzar, "Smartajweed Automatic Recognition of Arabic Quranic Recitation Rules," Dec. 2020.

[9] N. J. Ibrahim, M. Y. I. Idris, Z. Razak, and N. N. A. Rahman, "Automated tajweed checking rules engine for Quranic learning," *Multicultural Education and Technology Journal*, 2013, doi: 10.1108/METJ-03-2013-0012.

[10] A. Alfaries, M. Albahlal, M. Almazrua, and A. Almazrua, "A Rule-Based Annotation System to Extract Tajweed Rules from Quran," in *Proceedings - 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences, NOORIC 2013*, 2015. doi: 10.1109/NOORIC.2013.63.

[11] K. A. Mohammad, A. H. Kandil, A. M. EL-Bialy, and S. A. Fawzi, "Automated Detection and Consistency Analysis of Tajweed Recitation Rules in the Holy Quran," in *2022 4th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, IEEE, Oct. 2022, pp. 364–367. doi: 10.1109/NILES56402.2022.9942371.

[12] R. Jumansyah and S. Luckyardi, "Designing virtual reality game for learning Al-Quran," *Journal of Engineering, Science and Technology*, vol. 18, no. 1, pp. 227–234, 2023.

[13] M. M. Al Anazi and O. R. Shahin, "A Machine Learning Model for the Identification of the Holy Quran Reciter Utilizing K-Nearest Neighbor and Artificial Neural Networks," *Information Sciences Letters*, vol. 11, no. 4, pp. 1093–1102, Jul. 2022, doi: 10.18576/isl/110410.

[14] Nurul Husna Mat Isa, Noor Hazirah Abd Aziz, Mariyah Ishak, Wan Azani Mustafa, and Mohd Nizho Abd Rahman, "Quran Mobile Application: A Structured Review," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 34, no. 2, pp. 117–132, Dec. 2023, doi: 10.37934/araset.34.2.117132.

[15] A. P. Andriyandi, W. Darmalaksana, D. S. adillah Maylawati, F. S. Irwansyah, T. Mantoro, and M. A. Ramdhani, "Augmented reality using features accelerated segment test for learning tajweed," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 18, no. 1, pp. 208–216, 2020, doi: 10.12928/TELKOMNIKA.V18I1.14750.

[16] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst Appl*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/j.eswa.2019.05.028.

[17] T. S. Lestari and D. A. N. Sirodj, "Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest," *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 160–167, Feb. 2022, doi: 10.29313/jrs.v1i2.525.

[18] N. Widjiyati, "Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval," *Jurnal Janitra Informatika dan*

Sistem Informasi, vol. 1, no. 1, pp. 1–7, Apr. 2021, doi: 10.25008/janitra.v1i1.118.

[19] M. R. Raharjo and A. P. Windarto, "Penerapan Machine Learning dengan Konsep Data Mining Rough Set (Prediksi Tingkat Pemahaman Mahasiswa terhadap Matakuliah)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, p. 317, Jan. 2021, doi: 10.30865/mib.v5i1.2745.

[20] Nafisah Nurul Hakim, "Implementasi Machine Learning pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia," *Jurnal Sistem Cerdas*, vol. 3, no. 1, pp. 25–35, May 2020, doi: 10.37396/jsc.v3i1.58.

[21] M. B. Arya Darmawan, F. Dewanta, and S. Astuti, "Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot," *TELKA - Telekomunikasi Elektronika Komputasi dan Kontrol*, vol. 9, no. 1, pp. 52–61, May 2023, doi: 10.15575/telka.v9n1.52-61.

[22] A. Prasetio, M. H. Hasibuan, and P. Sitompul, "Simulasi Penerapan Metode Decision Tree (C4. 5) Pada Penentuan Status Gizi Balita," *Jurnal Nasional Komputasi Dan Teknologi Informasi*, vol. 4, no. 3, 2021.

[23] U. Indahyanti, N. L. Azizah, and H. Setiawan, "Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa," *Jurnal Sains dan Informatika*, vol. 8, no. 2, Dec. 2022, doi: 10.34128/jsi.v8i2.459.