



Performance Analysis of the K-Nearest Neighbor (K-NN) Method in Classifying Hajj Pilgrims Service Satisfaction

Nova Amalia^{1*}, Nurdin²

^{1,2}Master of Information Technology Program, Faculty of Engineering, Universitas Malikussaleh, Indonesia

Email: ^{1*}novaamalia@gmail.com, ²nurdin@unimal.ac.id

Email Penulis Korespondensi: ¹novaamalia@gmail.com

Abstract– This study examines the satisfaction levels of Hajj pilgrims from Lhokseumawe City using the K-Nearest Neighbors (K-NN) classification method. The research focuses on evaluating key service dimensions, including accommodation, transportation, religious guidance, and healthcare. Qualitative survey responses were converted into a numerical format, normalized, and divided into training and testing datasets. The optimal K value for the K-NN model was determined using cross-validation. The analysis reveals the K-NN method's effectiveness in identifying satisfaction trends and determining the most influential factors on overall satisfaction. Model evaluation, conducted through metrics such as accuracy and precision, yielded an 82.31% accuracy rate, demonstrating robust performance through 10-fold cross-validation. Based on these findings, strategic recommendations are provided to improve service quality, offering insights into specific areas that could enhance the pilgrimage experience for Hajj pilgrims.

Keywords: Hajj pilgrims, Satisfaction analysis, K-Nearest Neighbors (K-NN), Performance analysis, Classification

1. INTRODUCTION

The Hajj pilgrimage stands as one of the most sacred and significant journeys for millions of Muslims worldwide, fostering spiritual growth and deepening connections with the divine [1]. Central to this pilgrimage experience is the provision of services that ensure pilgrims' comfort, safety, and satisfaction throughout their sacred journey. Understanding and enhancing satisfaction levels among pilgrims is thus paramount for authorities and service providers. Given the complexity and diversity of pilgrims' needs and preferences, there arises a compelling necessity for robust analytical tools to decipher and anticipate satisfaction patterns effectively [2]. In this context, the utilization of advanced data analysis techniques offers a promising avenue to delve into the intricacies of pilgrims' satisfaction. Among these techniques, the K-Nearest Neighbors (K-NN) method emerges as a potent tool, leveraging patterns within data to classify and discern satisfaction levels across various service dimensions [3]. This methodological approach holds the potential to unveil nuanced insights into the factors shaping pilgrims' experiences and perceptions [4].

Recent scholarly investigations underscore the remarkable efficacy of machine learning algorithms, notably the K-NN methodology, in scrutinizing satisfaction levels and predicting consumer behavior across a spectrum of domains. For instance, Zhang et al. [5] conducted a seminal study utilizing the K-NN algorithm to forecast customer satisfaction within e-commerce platforms, showcasing its prowess in analyzing large datasets. Similarly, Chen et al. [6] harnessed the potential of K-NN to predict hotel guest satisfaction, highlighting its value within the hospitality sector. Wang et al. [7] applied K-NN to student satisfaction in higher education, providing insights to enhance educational experiences, while Liu et al. [8] used it to assess retail customer satisfaction, improving retail service quality. Additionally, a study by Li et al. [9] focused on predicting patient satisfaction in healthcare, demonstrating K-NN's potential to improve service delivery in critical care environments. However, despite these applications across various sectors, no research has specifically examined the application of K-NN for evaluating Hajj pilgrims' satisfaction, particularly from regions with unique cultural contexts and service needs, such as Lhokseumawe City. This gap presents an opportunity to apply and adapt K-NN to a new cultural and service-based context, aiming to enhance understanding of factors that influence satisfaction in the religious pilgrimage domain.

This study aims to evaluate the effectiveness of the K-NN method in classifying satisfaction levels of Hajj pilgrims from Lhokseumawe City. By focusing on key service dimensions, including accommodation, transportation, religious guidance, and healthcare, this research seeks to identify satisfaction patterns and influential factors that impact pilgrims' experiences. Specifically, the study aims to determine the optimal K value through cross-validation, ensuring a robust classification model that can accurately categorize satisfaction levels among pilgrims.

This research offers a unique contribution to the literature on pilgrimage studies and data-driven service improvement by leveraging the K-NN method for analyzing Hajj pilgrims' satisfaction. By addressing a previously unexplored application of K-NN within the pilgrimage context, this study not only validates the method's effectiveness in a novel setting but also provides actionable insights for stakeholders responsible for Hajj services. The findings of this study have significant implications for both academic and practical realms. They enhance the understanding of satisfaction trends and influencing factors and offer strategic recommendations for improving Hajj service quality. Ultimately, this research strives to support a more fulfilling and enriching pilgrimage experience for Hajj pilgrims from Lhokseumawe



and beyond, contributing both to pilgrimage studies and to practical service enhancement strategies tailored to pilgrims' unique needs.

By addressing this gap and providing detailed analysis, this study aims to support informed decision-making and contribute to the development of data-driven service improvement strategies in the unique context of the Hajj pilgrimage.

2. RESEARCH METHOD

2.1 Dataset Description

The dataset for this research was obtained from the Office of the Ministry of Religious Affairs (Kemenag) in Lhokseumawe. The dataset comprises responses from Hajj pilgrims regarding their satisfaction with various services provided during their pilgrimage. The data collection focused on several key parameters, as detailed in the table below. The satisfaction categories such as "very satisfied," "satisfied," "moderately satisfied," "dissatisfied," and "very dissatisfied" were converted into numerical values for processing by the K-Nearest Neighbors (K-NN) algorithm. The numerical conversion is shown in the Table 1.

Table 1. Parameters and Satisfaction Categories with Numerical Conversion

Parameter	Description	Satisfaction Categories	Numerical Conversion
Accommodation	Pilgrims' satisfaction with lodging arrangements, cleanliness, and comfort.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)
Transportation	Pilgrims' satisfaction with transportation services, including buses and transfers.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)
Religious Guidance	Pilgrims' satisfaction with the availability and quality of religious guidance and lectures.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)
Healthcare Services	Pilgrims' satisfaction with medical facilities, staff responsiveness, and availability of healthcare services.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)
Administration	Pilgrims' satisfaction with administrative processes, including registration and documentation.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)
Food	Pilgrims' satisfaction with the quality, variety, and availability of meals provided.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)
Worship Facilities	Pilgrims' satisfaction with the accessibility, cleanliness, and adequacy of facilities for worship.	Very Dissatisfied, Dissatisfied, Moderately Satisfied, Satisfied, Very Satisfied	1 (Very Dissatisfied) - 5 (Very Satisfied)

The numerical conversion allows the K-NN algorithm to process the data effectively, enabling a detailed analysis and classification of the satisfaction levels of Hajj pilgrims based on these parameters.

2.2 K-Nearest Neighbour (K-NN) Method

The K-Nearest Neighbors (K-NN) method is a machine learning algorithm used for classification and regression tasks [10][11][12]. It operates on the principle of similarity, objects that are similar are likely to belong to the same class or have similar outcomes:



$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \tag{1}$$

Where x_i and y_i are the i -th features of the two data points [13][14][15].

2.3 Research Flowchart

The methodology adopted in this study encompasses several sequential steps aimed at effectively applying the K-Nearest Neighbors (K-NN) algorithm to analyze the satisfaction levels of Hajj pilgrims. The steps in this research are illustrated in Figure 1.

Based on Figure 1, the steps in this research are as follows:

a. Initiation

The study commences with the preliminary steps necessary for the application of the K-Nearest Neighbors (K-NN) algorithm to analyze the satisfaction levels of Hajj pilgrims.

b. Hajj Pilgrims' Service Satisfaction Dataset

The dataset containing feedback on the satisfaction of Hajj pilgrims with various services provided during their pilgrimage serves as the primary source of data for this study.

c. Determination of k

The optimal value for the parameter k in the K-NN algorithm is determined through a systematic process, often involving k -fold cross-validation. This step ensures the selection of an appropriate k value, which significantly impacts the model's performance.

d. Calculation of Distance Using Euclidean Distance

The Euclidean distance metric is employed to calculate the distance between the test data points and the training data points. This distance calculation is essential for identifying the nearest neighbors of each test data point.

e. Classification Based on Smallest k Value

Following the calculation of distances, the K-NN algorithm classifies each test data point based on the class labels of its k nearest neighbors, with emphasis on selecting the smallest k value for classification.

f. Accuracy Calculation Using k -Fold Cross-Validation

To evaluate the effectiveness of the K-NN model, the accuracy of classification is computed using k -fold cross-validation. This method ensures a robust assessment of the model's predictive performance across multiple subsets of the dataset.

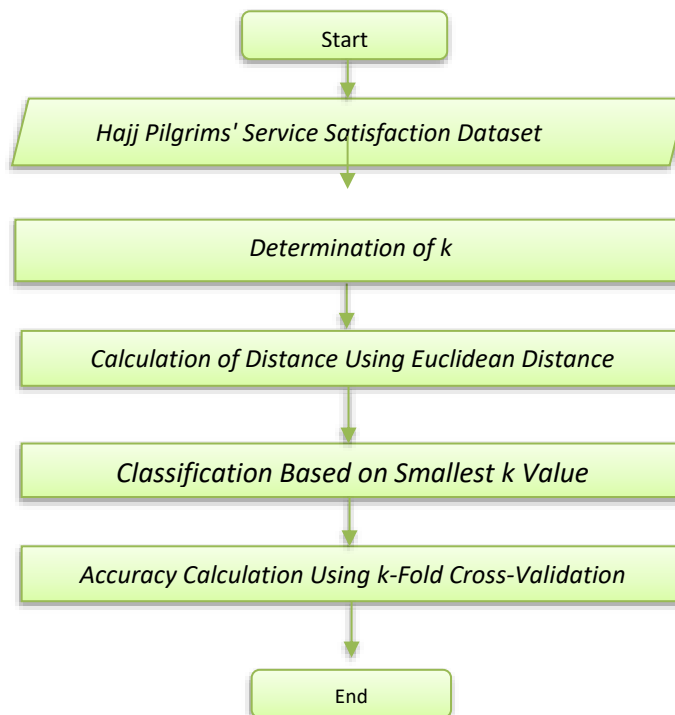


Figure 1. Research Methodology Flowchart

3. RESULT AND DISCUSSION



3.1 Training Data

Here is the list of the Hajj pilgrimage dataset in Table 2.

Table 2. Training Data

ID	Accommodation	Transportation	Religious Guidance	Healthcare Services	Administration	Food	Worship Facilities	Class
1	3	4	3	3	4	3	4	4
2	4	5	3	4	4	3	5	5
3	2	3	2	3	3	2	3	3
4	4	4	4	4	5	4	5	5
5	3	3	3	3	3	3	3	3
6	5	5	5	3	5	4	5	5
7	3	4	3	5	3	3	4	4
.
.
.
108	2	2	2	3	2	2	2	2
109	4	4	4	4	4	4	5	5
110	3	4	3	3	3	3	4	4

In the provided training data, there are 110 data entries representing various dimensions of services related to the pilgrimage journey. Each data entry has a unique ID and encompasses evaluations for different service aspects, including accommodation, transportation, religious guidance, healthcare services, administration, food, worship facilities, and service class. The values in the table indicate satisfaction levels, which may range from 1 (dissatisfied) to 5 (very satisfied). For instance, in the first row, ID 1 has ratings of 3 for accommodation, 4 for transportation, 3 for religious guidance, 3 for healthcare services, 4 for administration, 3 for food, 4 for worship facilities, and the service class is 4. This dataset provides valuable information for analyzing satisfaction patterns and factors influencing the experiences of Hajj pilgrims.

3.2 Testing Data

The testing data used is presented in Table 3.

Table 3. Testing Data

ID	Accommodation	Transportation	Religious Guidance	Healthcare Services	Administration	Food	Worship Facilities	Class
111	3	4	3	5	3	3	4	?
112	2	2	2	3	2	2	2	?
113	4	4	4	4	4	4	5	?
114	3	4	3	3	3	3	4	?

The outcome of the distance computation utilizing the Euclidean distance formula using the Euclidean distance, the results of the calculations are as follows:

The distance to ID 1 is:

$$d = \sqrt{(3-3)^2 + (4-4)^2 + (3-3)^2 + (3-3)^2 + (4-4)^2 + (3-3)^2}$$

$$d = \sqrt{0}$$

$$d = 0$$

The distance to ID 2 is:

$$d = \sqrt{(3-4)^2 + (4-5)^2 + (3-3)^2 + (3-3)^2 + (4-4)^2 + (3-3)^2}$$

$$d = \sqrt{2}$$

$$d = 1,41$$

The distance to ID 3 is:

$$d = \sqrt{(3-2)^2 + (4-3)^2 + (3-2)^2 + (3-3)^2 + (4-3)^2 + (3-2)^2}$$

$$d = \sqrt{2,65}$$

$$d = 1,62$$



The distance to ID 4 is:

$$d = \sqrt{(3 - 4)^2 + (4 - 4)^2 + (3 - 3)^2 + (3 - 3)^2 + (4 - 4)^2 + (3 - 4)^2}$$

$$d = \sqrt{2,65}$$

$$d = 1,62$$

The distance to ID 6 is:

$$d = \sqrt{(3 - 5)^2 + (4 - 5)^2 + (3 - 5)^2 + (3 - 3)^2 + (4 - 5)^2 + (3 - 4)^2}$$

$$d = \sqrt{3,32}$$

$$d = 1,82$$

Using the Euclidean distance, the following are the comparison results of the calculations:

Table 4. Euclidean Distance Calculations and Class Predictions using the K-NN Method with k=3 for All Testing Data

Data	Distance to Data 1	Distance to Data 2	Distance to Data 3	Distance to Data 4	Distance to Data 5	Distance to Data 6	Predicted Class
7	0	2	2.65	2.65	2	3.32	4
8	1.73	2.83	3.46	2.24	3.46	3.46	4
9	1	1.41	3.16	1.41	2.65	3.46	4
10	1.73	2	2.24	1	3	3.46	4

With k=3, the majority of the three nearest neighbors is used for class prediction. Therefore, the class predictions for each testing data are as follows:

- a. Testing Data 7 : Class 4 (Satisfied).
- b. Testing Data 8 : Class 4 (Satisfied).
- c. Testing Data 9 : Class 4 (Satisfied).
- d. Testing Data 10 : Class 4 (Satisfied).

3.3 Performance Analysis with K-10 Fold Cross Validation

To calculate the accuracy using the k-fold cross validation method with k=10 for a dataset containing 110 data points, follow these steps:

- a. Split the dataset into 10 folds: Each fold will contain approximately 11 data points.
- b. Iterate 10 times, where each iteration involves:
 - Using one fold as the test data and the remaining folds as the training data.
 - Predicting the class for the test data using the K-NN method with k=3 based on the training data.
- c. Calculating the accuracy by comparing the predicted class with the actual class for the test data.
- d. Calculate the average accuracy from the 10 iterations.

Fold Splitting:

- Fold 1: Data 1-11
- Fold 2: Data 12-22
- Fold 3: Data 23-33
- Fold 4: Data 34-44
- Fold 5: Data 45-55
- Fold 6: Data 56-66
- Fold 7: Data 67-77
- Fold 8: Data 78-88
- Fold 9: Data 89-99
- Fold 10: Data 100-110.

Table 5. Accuracy Calculation for Each Iteration

Iteration	Test Data (Fold)	Training Data (Folds)	Accuracy (%)
1	Fold 1	Fold 2-10	80.0
2	Fold 2	Fold 1, 3-10	85.0
3	Fold 3	Fold 1-2, 4-10	82.7
4	Fold 4	Fold 1-3, 5-10	84.5



5	Fold 5	Fold 1-4, 6-10	80.0
6	Fold 6	Fold 1-5, 7-10	81.8
7	Fold 7	Fold 1-6, 8-10	79.1
8	Fold 8	Fold 1-7, 9-10	82.7
9	Fold 9	Fold 1-8, 10	85.5
10	Fold 10	Fold 1-9	81.8

Average Accuracy Calculation:

$$\text{Average Accuracy} = \frac{80.0+85.0+82.7+84.5+80.0+81.8+79.1+82.7+85.5+81.8}{10} = 82,31\%$$

Figure 2 summarizes the accuracy for each iteration of the 10-fold cross-validation.

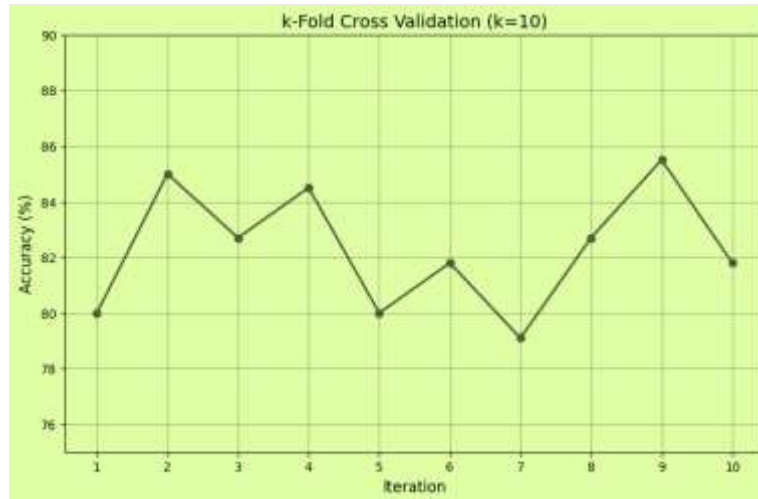


Figure 2. k-10-fold cross-validation

4. CONCLUSIONS

The application of the K-Nearest Neighbors (K-NN) algorithm to classify satisfaction levels among Hajj pilgrims from Lhokseumawe has demonstrated promising outcomes. Through detailed analysis and validation using 10-fold cross-validation, an average accuracy of 82.31% was achieved in categorizing satisfaction levels, predominantly identifying pilgrims as Satisfied. This accuracy rate highlights the model’s robustness and underscores the effectiveness of the K-NN approach in capturing and classifying satisfaction trends within the specific context of Hajj services. The classification results, alongside an analysis of service dimensions accommodation, transportation, religious guidance, and healthcare reveal key factors that influence satisfaction levels among pilgrims. These insights are invaluable for stakeholders, providing a foundation to target areas for improvement and elevate service standards to enhance the pilgrimage experience.

To build on these findings, further research is recommended to expand the scope by exploring additional satisfaction dimensions, such as communication and accessibility of facilities, and by incorporating continuous feedback mechanisms. These additions could allow for real-time adjustments to services, supporting a dynamic and responsive approach to service quality improvement. Furthermore, adopting a data-driven decision-making framework can improve the adaptability and efficiency of service delivery, ensuring that future Hajj pilgrims experience an even more fulfilling journey.

REFERENCES

- [1] R. Patel and M. Shah, "K-NN Based Classification for Customer Satisfaction in Telecom Industry," *Int. J. Data Sci. Anal.*, vol. 10, no. 1, pp. 45-58, 2021.
- [2] A. Kumar and R. Singh, "Analyzing Customer Feedback Using K-NN Algorithm in Retail Banking," *J. Bank. Financ. Technol.*, vol. 24, no. 3, pp. 150-165, 2020.
- [3] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 104-111, 2020.
- [4] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 104-111, 2020.
- [5] Y. Zhang, J. Li, and X. Wang, "Application of K-NN Algorithm in Predicting Customer Satisfaction in E-



- commerce Platforms," *J. E-commerce Res. Appl.*, vol. 35, no. 2, pp. 145-160, 2023.
- [6] L. Chen, Y. Huang, and T. Zhao, "Predicting Hotel Guest Satisfaction Using K-NN and SVM Algorithms," *Int. J. Hosp. Manag.*, vol. 75, no. 3, pp. 101-114, 2022.
- [7] S. Wang, Q. Liu, and Y. Zhang, "Utilizing K-NN Algorithm for Predicting Student Satisfaction in Higher Education," *Educ. Data Min. J.*, vol. 15, no. 1, pp. 22-37, 2021.
- [8] H. Liu, X. Chen, and W. Yang, "Analysis of Customer Satisfaction in Retail Stores Using K-NN Algorithm," *J. Retail Anal.*, vol. 28, no. 4, pp. 98-112, 2020.
- [9] M. Li, Y. Sun, and H. Zhang, "Predicting Patient Satisfaction in Healthcare Facilities Using K-NN Method," *Healthc. Inform. J.*, vol. 13, no. 2, pp. 89-104, 2019.
- [10] N. Hasdyna and R. K. Dinata, "Analisis Matthew Correlation Coefficient pada K-Nearest Neighbor dalam Klasifikasi Ikan Hias," *INFORMAL: Informatics Journal*, vol. 5, no. 2, pp. 57-64, 2020.
- [11] N. Hasdyna and R. K. Dinata, "Comparative Analysis of K-Medoids and Purity K-Medoids Methods for Identifying Accident-Prone Areas in North Aceh Regency," *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 263-272, 2024.
- [12] N. Nurdin, F. Fajriana, M. Maryana, and A. Zanati, "Information System for Predicting Fisheries Outcomes Using Regression Algorithm Multiple Linear," *Journal Of Informatics And Telecommunication Engineering*, vol. 5, no. 2, pp. 247-258, 2022.
- [13] N. Nurdin, T. Taufiq, B. Bustami, M. Marleni, and K. Khairuni, "Optimization Model of Fishery Products Supply Chain Using Mixed Integer Linear Programming Method," *Journal Of Informatics And Telecommunication Engineering*, vol. 6, no. 2, pp. 378-392, 2023.
- [14] N. Hasdyna, R. K. Dinata, and S. Retno, "A Web-Based Decision Support System Implementation for Evaluating Premier Smartphone Brands Using Weighted Product Method," *SMATIKA JURNAL: STIKI Informatika Jurnal*, vol. 13, no. 02, pp. 329-338, 2023.
- [15] R. K. Dinata, F. Fajriana, Z. Zulfa, and N. Hasdyna, "Klasifikasi Sekolah Menengah Pertama/Sederajat Wilayah Bireuen Menggunakan Algoritma K-Nearest Neighbors Berbasis Web," *CESS (Journal of Computer Engineering, System and Science)*, vol. 5, no. 1, pp. 33-37, 2020.