

Sentiment Analysis of Public Opinion on Hajj Pilgrimage Travel Costs Using the K-Nearest Neighbors (KNN) Method

Zeinor Rahman^a, Haryono^b, Rachmad Hidayat^c, Mohammad Iqbal Bachtiar^d, Rizki Anantama^e

^aDepartment of Informatics, Faculty of Science and Technology, University of Bahaudin Mudhary Madura, Sumenep, Indonesia

ABSTRACT

In the context of Islam, humans have the duty to educate themselves, purify their souls, and control their desires. Worship performed with sincerity and pure faith is considered a blessing. One example of worship in Islam is performing the Hajj pilgrimage, which is one of the five pillars of Islam, along with the declaration of faith (shahada), prayer (salat), almsgiving (zakat), and fasting (sawm). Hajj requires self-control, dedication, and sacrifice, including financial resources (Noor, 2018). However, the cost of the Hajj pilgrimage, proposed by the Ministry of Religious Affairs (Kemenag), is set to increase by almost 100% from the previous year, reaching IDR 69 million per person in 2023, eliciting various responses from the public. Some responses are constructive and positive, while others oppose the increase. Sentiment analysis is the method used to analyze public reactions to this change. Data for sentiment analysis was gathered from Twitter, with tweets processed using the Tf-Idf and Tf-Rf word weighting methods. The K-Nearest Neighbors (KNN) algorithm was then employed to assess the accuracy and effectiveness of these methods. The study findings revealed that the Tf-Idf word weighting method outperformed Tf-Rf in categorizing public sentiment regarding the cost of the Hajj pilgrimage. Tf-Idf achieved 84% accuracy, 81% precision, and 79% recall, while Tf-Rf achieved 79% accuracy, 75% precision, and 77% recall. Tf-Idf is considered superior in this sentiment analysis, providing precise predictions while effectively capturing important sentiments. Therefore, this method is deemed a reliable choice for accurate sentiment analysis.

Keywords: Sentiment Analysis, K-Nearest Neighbors (KNN), Tf-Idf, Tf-Rf.

Article History

Received 15 April 25

Received in revised form 12 May 25

Accepted 20 June 25

1. Introduction

In Islam, individuals are tasked with educating their external selves, purifying their souls, and liberating themselves from worldly desires. Those who achieve this are those who worship sincerely and possess pure faith aligned with Allah's will. Islam encompasses various forms of worship, including the Hajj pilgrimage. Hajj, the fifth pillar of Islam, is obligatory for those who can afford it, coming after the declaration of faith (shahada), prayer (salat), almsgiving (zakat), and fasting (sawm). Hajj is considered a noble act of worship as it demands self-discipline, physical effort, enthusiasm, and financial sacrifice (Noor, 2018). As reported by Bareksa.com, the Ministry of Religious Affairs (Kemenag) proposed that the cost of the Hajj pilgrimage, known as the Hajj fare (ONK), would nearly double in 2023. The proposed increase would raise the fare from IDR 39.8 million in 2022 to IDR 69 million per person in 2023. This significant increase in Hajj costs has elicited various responses from the public, both positive and negative. One platform where individuals freely express their opinions is social media, particularly Twitter (Nano, 2020).

This study employs the K-Nearest Neighbors (KNN) algorithm to classify the sentiment of tweets on Twitter. KNN is a classification model that relies on the similarity or nearest distance of objects within the training data (Siregar et al., 2019). Initially, Twitter text data is processed and then analyzed using KNN. Before applying the KNN algorithm, the text data from Twitter must be extracted and weighted using feature extraction and word weighting methods such as Tf-Idf and Tf-Rf. Tf-Idf is a statistical measure used to determine the significance of a word in a document or a collection of words, while Tf-Rf is a newer method that assesses the relevance of a document based on the frequency of term occurrences in related categories, treating each sentence as a single document (Assidyk et al., 2020a). The frequency of word occurrences in a document reflects the importance and commonality of that word within the document (Amrizal, 2018).

The K-Nearest Neighbors (KNN) algorithm utilized in this study focuses on classifying tweets from Twitter based on predefined categories. The final step involves assigning a class to each sentence and assessing the model's accuracy. KNN aims to categorize objects into predefined classes using sample data (Pratama et al., 2021). This research distinguishes itself by comparing the word weighting methods Tf-Idf and Tf-Rf. Previous

* Corresponding author.

E-mail address: zenio.spd@gmail.com.

studies employing KNN with Tf-Idf achieved a high accuracy of 96.61% (Putri Fitrianti et al., 2018), with an overall accuracy of 89.4% and an error rate of 10.6% (Alrajak et al., 2020). The objective of this study is to offer accurate insights into public sentiment on Twitter regarding the cost of the Hajj pilgrimage, aiming to provide valuable information to the public.

2. Research Methodology

2.1 Sentiment Analysis

Sentiment analysis is a technique used to automatically extract, process, and comprehend textual data in order to discern the sentiment expressed in opinions (Sari, F. V., & Wibowo, 2019). The aim is to gather user opinions from a platform. Situated within the field of Natural Language Processing (NLP), sentiment analysis involves developing systems to identify and extract text sentiments. Textual information is widely available on the internet today, such as on social media, blogs, forums, and review sites.

Hajj is an act of worship that requires sincerity and strong determination. It involves the resolve to leave one's hometown and beloved family and the commitment to abandon all bad behaviors. Hajj is obligatory for every Muslim, with the condition "for those who are able." This includes the ability to perform Hajj both physically and financially (Istianah, 2017). Preparing oneself as a guest of Allah is very important. Hajj is a form of worship that does not distinguish between rank and social status.

2.2 Cosine Similarity

The Cosine Similarity method is employed to gauge the likeness between two documents or texts. In this study, cosine similarity is utilized to compute the frequency of term occurrences across referenced pages indexed. Each document or text is treated as a vector, where in text matching, vectors A and B denote the term-frequency vectors of a document (Samuel et al., 2018). The formula for calculating cosine similarity is as follows:

$$\cos(\theta_{ij}) = \frac{\sum_k (d_{ik} d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \quad (2.1)$$

Description:

dik = weight of *k* towards document *i*

djk = weight of *k* towards document *j*

If the similarity score is higher, then the level of similarity between the test document and the training document is greater

2.3 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a supervised learning classification method used to classify objects based on training data and their associated class attributes. The algorithm operates by classifying a test data object based on the classes of the training data points that are closest in distance to it. Each training data point has a known class attribute, and the model predicts the class for test data points that lack class attributes (Septian et al., 2019).

2.4 Weighting

To assign weights to words in a sentence or document, the Tf-Idf (Term Frequency-Inverse Document Frequency) and Tf-Rf (Term Frequency-Relevance Frequency) methods are employed. Tf-Idf converts textual data into numerical data by assigning weights to each word or feature. It is a statistical measure used to evaluate the importance of a word in a document. Tf-Idf integrates two primary concepts in its weighting calculation: the frequency of a word's occurrence in a specific document and the inverse frequency of documents containing that word. The

frequency of a word in a document indicates its significance within that document. Consequently, the weight of the relationship between a word and a document is higher when the word's frequency is high within the document and lower across the collection of documents containing that word (Nurjannah & Fitri Astuti, 2013). The Tf-Idf equation is expressed as follows:

$$tf = \begin{cases} 1 + \log(f_{t,d}), & f_{t,d} > 0 \\ 0, & f_{t,d} = 0 \end{cases} \quad (2.2)$$

$$idf_t = \log\left(\frac{D}{d_{ft}}\right) + 1 \quad (2.3)$$

$$W_{d,t} = tf_{d,t} \times idf_{d,t} \quad (2.4)$$

Description:

tf = Number of times a term appears in a document.

ft = Frequency of term (t) in document.

D = Total number of documents.

dft = Total number of documents containing term t.

Idf = Inverse Document Frequency.

d = Document index d.

t = Term index t from the keyword.

W = Weight of a document d against term t.

2.5 Model Evaluation

This stage evaluates the accuracy of methods using confusion matrices, accuracy tables, and precision tables for each model. The research model is evaluated using the K-fold cross-validation technique, where the data used for testing is divided into several (k) parts, and classification tests are conducted multiple times based on these parts (Wijaya, 2019). Typically, k is set to 10 (k=10), meaning the data is split into 10 parts and tested 10 times (Ratnawati, 2018). In K-fold cross-validation, data is divided into training and testing sets.

K-fold cross-validation provides accuracy, precision, and recall values. Precision measures the accuracy of predicted positive cases that are actually positive in the actual data. Recall measures the proportion of actual positive cases that are correctly predicted as positive. Accuracy, on the other hand, is the percentage of total data classified correctly overall (Andriani, 2013).

Table 2. 1 Table Cofusion Matrix

Prediction value	The actual value	
	P	N
P	true positif	false positif
N	false negative	true negative

3. Results and Discussion

3.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

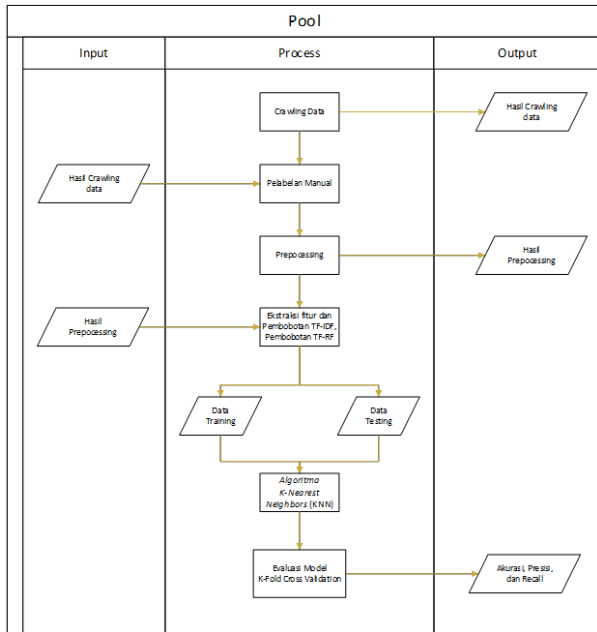
This research will employ the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which encompasses several stages: Business Understanding, Data Understanding, Data Collection, Data Preparation, Modeling, and Evaluation (Fazar et al., 2020).

3.2 System Design

In Figure 3.1, the system design flow illustrates the process flow of the system. It begins with data crawling from Twitter. Next, manual labeling of the crawled data based on the opinions of 3 volunteers (Indonesian Language Teachers) will be conducted. The data then proceeds to the preprocessing stage, resulting in preprocessed data output. The preprocessed data will undergo weighting and feature extraction processes to obtain weights for each word that will serve as features. Next, the data is

split into training and testing sets for classification. The classification stage utilizes the K-Nearest Neighbors (KNN) model, with the researcher opting for K=10. This choice of K was based on a review of literature, selecting the value that produced optimal results for the KNN classification model. The next step involves testing the model using K-fold cross-validation to display accuracy, precision, and recall rates.

Figure 3.1 System Process Diagram



3.3 Testing Scenario

In this study, the testing scenario assesses the accuracy of a custom-built program (developed without using a library) and compares its performance with a program that utilizes a library.

4. Interface Implementation

4.1 Data Understanding

In the data understanding phase, the researcher aims to understand the data to be used. Data collection is performed through crawling. The data to be crawled consists of public sentiment on Twitter regarding the government's policy on the Hajj pilgrimage cost increase. The data was collected on February 7, 2022, resulting in 157 tweet data. Subsequently, the researcher manually labels the data with the assistance of Indonesian language teachers.

	A	B	C	D	E	F
1	Dokumen	Orang1	Orang2	Orang3	sentimen	label
2	@spotipay Kenaikan harga haji ini benar-benar membuat saya khawatir. Bagaimana generasi mendatang?	negatif	negatif	negatif	0	
3	@HambaAllah_411 Harga naik haji makin luar biasa, rasanya seperti hanya yang kaya yang bisa melaksan...	negatif	negatif	negatif	0	
4	@merfessyBagaimana kita bisa berharap untuk melaksanakan haji dengan harga yang semakin tinggi? negatif	negatif	negatif	negatif	0	
5	@spotipay Halo kak @spotipay, Berdasarkan SE kemenhub No. 96 Tahun 2021 dan SE Setgas No. 22 Tahun...	positif	positif	positif	1	
6	@banguntamine Halo kak @banguntamine. Saya sangat mendukung keinginan untuk menjalani haji, tetapi...	positif	positif	positif	1	
7	@tiket Kak beda antigen sama test pcr apa kak apa harus wajib dua2 nya untuk berangkat naik haji?	positif	positif	positif	1	
8	Saya berharap agar pemerintah dapat mengatasi kenaikan harga haji ini. Semakin banyak orang yang bisa...	positif	positif	positif	1	
9	@nanderahin Kenaikan harga haji membuat banyak orang terjebak dalam utang dan beban keuangan yang negatif	negatif	negatif	negatif	0	
10	Sungguh disayangkan melihat harga haji yang semakin tinggi. Banyak orang berhenti bernimpi untuk mel...	negatif	negatif	negatif	0	
11	@noid7 merujuk SE 97/2021 Kemenhub, anak usia	positif	positif	positif	1	
12	@arnybanker Saya bercita-cita untuk naik haji, tetapi dengan harga yang semakin melonjak, rasanya sepe...	negatif	negatif	negatif	0	
13	@HeuraBessa anak usia di bawah 12 tahun (termasuk balita, batita dan bayi) sudah boleh naik Ka anar k...	positif	positif	positif	1	
14	@chocochandra Harga haji naik lagi? Semakin sulit untuk mewujudkan impian ini. Mungkin perlu lebih ba...	positif	positif	positif	1	
15	@iniave Kami informasikan saat ini PCR Test/Rapid Test Antigen wajib dilakukan di rumah sakit/klini...	positif	positif	positif	1	
16	@panca66 @msaid_didu Cuma di Indon perjalanan domestik wajib test PCR. Cuma di Indon orang gak ber...	positif	negatif	negatif	0	
17	@dianchrisa Sebagai umat Muslim, kita semua ingin menjalani haji. Tetapi kenaikan harga yang terus m...	positif	positif	positif	1	
18	@mimipiduluaa Harga haji kembali naik, dan ini sangat mempengaruhi semangat umat Islam untuk mela...	positif	positif	positif	1	
19	@pekyuv @andhydt @solihsolihun Betul utk gejala batuli slim ke dkr umum wajib pcr di tggp hasil test k...	positif	positif	positif	1	
20	dok adek aku mulai dr kemarin udh wajib ptn 100%, eh ternyata pas ud nyampe rumah diumumin di grup a...	positif	positif	positif	1	
21	@sdpdohm123 bulan depan kak, di KUA aku sm calon hrs ada surat keterangan sehat dr puskesmas ma...	positif	positif	positif	1	
22	@akusperimadu Halo kak @akusperimadu. Kenaikan harga haji terus meningkat, sementara penghasil...	negatif	negatif	negatif	0	
23	@Mlusiadikim88 Ketika harga haji terus naik, banyak umat Islam yang merasa putus asa. Semoga pemer...	negatif	negatif	negatif	0	
24	@faizalm07 Semoga ada kebijakan yang dapat membantu meringankan beban finansial @haji2023 & sentri...	negatif	negatif	negatif	0	
25	@Indrininggal Saya sangat ingin naik haji, tetapi harga yang terus naik membuatnya seperti mimpi yang t...	positif	positif	positif	1	
26	@henrvr Harga haji yang semakin mahal membuat saya merasa putus asa. Bagaimana bisa kita melakukan...	negatif	negatif	negatif	0	

Figure 4.1 Manually labeled crawled data

4.2 Data Prepaion

In this stage, data crawled from Twitter will be manually labeled for sentiment, followed by preprocessing. Before preprocessing, there are 3946 words in the dataset consisting of 157 data points, which have been filtered to remove non-word characters. The outcomes are depicted in the figure provided below:

```

tweet_tokens
0 [kenaikan, harga, haji, ini, benarbenar, membu...
1 [harga, naik, haji, makin, luar, biasa, rasany...
2 [kita, bisa, berharap, untuk, melaksanakan, ha...
3 [halo, kak, berdasarkan, se, kemenhub, no, tah...
4 [halo, kak, saya, sangat, mendukung, keinginan...
..
152 [selamat, siang, kak, merujuk, se, kemenhub, a...
153 [kami, informasikan, saat, ini, pcr, testrapid...
154 [didu, cuma, di, indon, perjalanan, domestik, ...
155 [selamat, sore, kak, merujuk, se, kemenhub, an...
156 [selamat, sore, kak, merujuk, se, kemenhub, an...

```

[157 rows x 7 columns]

Jumlah 'Kata' sebelum dilakukan Preprocessing dan Drop Duplikat = 3946

Figure 4.2 Process of counting words before the step of dropping duplicates and preprocessing

4.3 Modelling

In this phase, the researcher partitions the data into training and testing sets in an 80:20 ratio for both TF-Idf and Tf-Rf word weighting methods. The data is then processed using the K-Nearest Neighbors (KNN) algorithm. Subsequently, testing is conducted using 10-fold cross-validation. The initial dataset, comprising 153 data points, is divided into training data for model training and testing data for evaluating the model's performance. This ensures that the developed model can accurately predict and generalize to new, unseen data. Through this testing, the effectiveness of the K-Nearest Neighbors (KNN) model employing Tf-Idf and Tf-Rf word weighting in classifying public sentiment towards the cost of the Hajj pilgrimage can be assessed.

Figure 4.3 K-fold Cross Validation

Uji	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih
Latih	Uji	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih
Latih	Latih	Uji	Latih	Latih	Latih	Latih	Latih	Latih	Latih
Latih	Latih	Latih	Uji	Latih	Latih	Latih	Latih	Latih	Latih
Latih	Latih	Latih	Latih	Uji	Latih	Latih	Latih	Latih	Latih
Latih	Latih	Latih	Latih	Latih	Uji	Latih	Latih	Latih	Latih
Latih	Latih	Latih	Latih	Latih	Latih	Uji	Latih	Latih	Latih
Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji	Latih	Latih
Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji	Latih
Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Latih	Uji

Visualization of the KNN-TfIdf method and the distribution of positive and negative data:

<AxesSubplot: xlabel='sentimen', ylabel='count'>

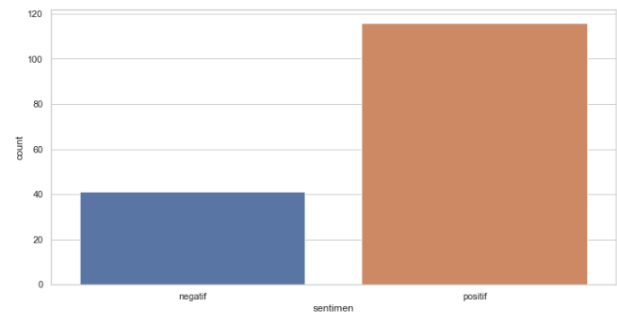


Figure 4.4 Visualization of positive and negative data distribution

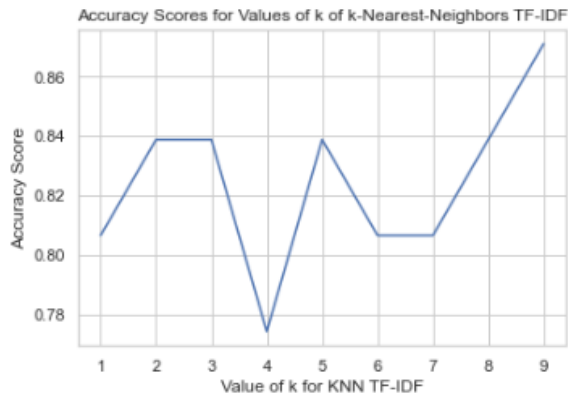


Figure 4.5 Tf-idf KNN Graph

Looking at the graph results above, it can be concluded that the optimal k value for the KNN method with Tf-Idf word weighting is k=9, achieving a KNN score of 87%.

- The accuracy results of K-fold cross-validation for each fold in the 10-fold testing of Tf-Idf word weighting are shown in Table 4.1 below:

Table 4.1 Results of k-fold cross-validation for Tf-idf

Fold ke-	Accuracy	Precision	Recall
K1	0.812500	0.791667	0.754545
K2	0.812500	0.754545	0.791667
K3	0.812500	0.754545	0.791667
K4	0.866667	0.829545	0.829545
K5	0.933333	0.958333	0.875000
K6	0.800000	0.750000	0.704545
K7	0.666667	0.541667	0.534091
K8	0.933333	0.958333	0.875000
K9	1.000000	1.000000	1.000000
K10	0.800000	0.750000	0.784091
Average	84%	81%	79%

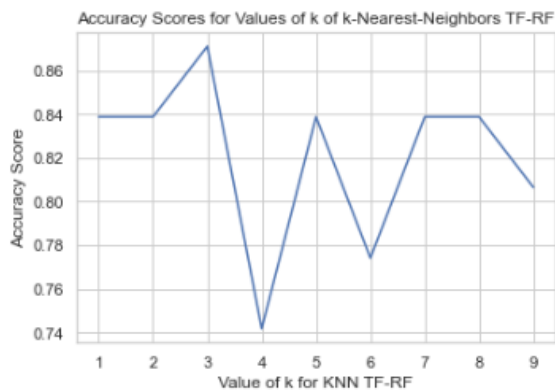


Figure 4.6 Grafik tf-rf KNN

Based on the graph results shown above, it can be concluded that the optimal k value for the KNN method with Tf-Rf word weighting is k=3, achieving a KNN score of 87%.

The accuracy results of K-fold cross-validation for each fold in the 10-fold

testing of Tf-Rf word weighting are shown in Table 4.2 below:

Table 4.2 Results of k-fold cross-validation for Tf-Rf

Fold ke-	Accuracy	Precision	Recall
K1	0.687500	0.650000	0.663636
K2	0.812500	0.754545	0.791667
K3	0.875000	0.833333	0.833333
K4	0.800000	0.750000	0.704545
K5	0.933333	0.958333	0.875000
K6	0.733333	0.659091	0.659091
K7	0.800000	0.750000	0.704545
K8	0.533333	0.517857	0.522727
K9	0.866667	0.833333	0.909091
K10	0.733333	0.694444	0.738636
Average	78%	74%	74%

4.4 Evaluate

In this stage, the researcher evaluates the classification method by measuring the performance accuracy, precision, and recall of each word weighting used in this study, namely the K-Nearest Neighbors (KNN) algorithm.

Table 4.3 Comparison of accuracy, precision, and recall results between Tf-Idf and Tf-Rf word weighting:

No	Weighting	Metode	Accuracy	Precision	Recall
1.	Tf-Idf	k-nearest neighbors (KNN).	84%	81%	79%
2.	Tf-Rf	k-nearest neighbors (KNN).	79%	75%	77%

Based on the test results, the best method was found to be Tf-Idf compared to Tf-Rf, achieving an accuracy of 84%. The Tf-Idf method is more effective in classifying public sentiment towards the cost of the Hajj pilgrimage compared to the Tf-Rf method in this testing context. With a balanced precision of 81% and recall of 79%, Tf-Idf achieves a good balance between the number of correct predictions (precision) and the ability to detect all true instances (recall). This indicates that the method can provide accurate predictions without sacrificing the ability to detect important sentiments.

4.5 Testing Scenario

In this testing scenario, the parts of the program developed without using library assistance include the Tf-Idf word weighting stage. However, the cosine similarity, KNN algorithm, and K-fold cross-validation utilize library support. The testing strategy involves comparing the Tf-Idf process performed without libraries to the Tf-Idf process using libraries. This comparison is followed by evaluating the accuracy, precision, and recall of both approaches in the KNN and K-fold cross-validation processes. The results of this testing scenario are presented in Table 4.4 below:

Table 4.4 Results of the test scenario for the Tf-Idf program without using a library and using a library

Tf-Idf tanpa Library			Tf-Idf menggunakan Library		
Akurasi	Presisi	Recall	Accuracy	Presisi	Recall
84%	81%	79%	83%	82%	75%

Based on the results table provided, it is clear that the manually implemented Tf-Idf word weighting method achieved an accuracy of 84%, precision of 81%, and recall of 79%. In comparison, the Tf-Idf method using a library achieved an accuracy of 83%, precision of 82%, and recall of 75%. Comparing these results, it can be concluded that the manually

implemented Tf-Idf method yielded slightly higher accuracy (84% vs. 83%) and nearly the same precision (81% vs. 82%) compared to the Tf-Idf method using a library. However, the manual method showed better recall (79% vs. 75%).

This conclusion indicates that in this testing context, the manually implemented Tf-Idf word weighting method performs quite well, especially in terms of accuracy and recall. However, the decision to use manual methods or libraries may also depend on other factors such as time efficiency and implementation complexity.

Conclusion

In this study on sentiment analysis regarding the costs of the Hajj pilgrimage, the research findings indicate that the Tf-Idf word weighting method is more effective compared to the Tf-Rf method when used in conjunction with the K-Nearest Neighbors (KNN) algorithm. Based on the results provided, the Tf-Idf method demonstrated superior performance across all evaluation metrics, achieving an accuracy of 84%, precision of 81%, and recall of 79%. In contrast, the Tf-Rf method achieved an accuracy of 79%, precision of 75%, and recall of 77%. Therefore, it can be concluded that the Tf-Idf word weighting method is more effective in classifying public sentiment towards the costs of the Hajj pilgrimage. Additionally, Tf-Idf achieved a good balance between the number of correct predictions (precision) and the ability to detect all true instances (recall). This indicates that the method can provide accurate predictions without sacrificing the

REFERENCES

- [1] Alrajak, M. S., Emawati, I., & Nurlaili, I. (2020). Analisis Sentimen Terhadap Pelayanan PT PLN di Jakarta pada Twitter dengan Algoritma K-Nearest Neighbor (K-NN). *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 110–122.
- [2] Dwiki, A., Putra, A., Juanita, S., Studi, P., Informasi, S., Teknologi, F., Universitas, I., & Luhur, B. (2021). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN. 8(2), 636–646
- [3] Istiqlaliyyah, I., Sari, Y. A., & Fauzi, M. A. (2019). Pencarian Teks pada Terjemahan Ayat Al- Qur ' an dengan Menggunakan TF-RF dan Bray-Curtis Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(5), 4357–4363.
- [4] Joergensen Munthe, C. E., Astuti Hasibuan, N., & Hutabarat, H. (2022). Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace. *Media Online*, 2(3), 110–115
- [5] Kevin, V., Que, S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization). 9(2), 162–170
- [6] Laurensz, B., & Eko Sedyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- [7] Noor, M. (2018). Haji dan Umrah. *Jurnal Humaniora Teknologi*, 4(1), 38–42. <https://doi.org/10.34128/jht.v4i1.42>
- [8] Pratama, A. Y., Umaidah, Y., & Voutama, A. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(September), 897–910
- [9] Sari, D. I., Wati, Y. F., & Widiastuti. (2020). Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum Mrt Jakarta Menggunakan Naïve Bayes Classifier. *Jurnal Ilmiah Informatika Komputer*, 25(1), 64–75. <https://doi.org/10.35760/ik.2020.v25i1.2427>
- [10] Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. 43–49
- [11] Siregar, R. R. A., Siregar, Z. U., & Arianto, R. (2019). Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor. *Kilat*, 8(1), 81–92. <https://doi.org/10.33322/kilat.v8i1.421>
- [12] Wijaya, N. (2019). Penerapan Algoritma Klasifikasi Naive Bayes Untuk Data Status Huni Rumah Bantuan Dana Rehabilitasi Dan Rekonstruksi Pasca Bencana Erupsi Gunung Merapi 2010. *Seminar Nasional UNRIYO*, 1–10