Type of Contribution:
Research Paper
Review Paper
Case Study

ENERGY: JURNAL ILMIAH ILMU-ILMU TEKNIK
Vol. 15, No. 1 (2025) pp 67-76
DOI: 10.51747/energy.v15i1.1516

E-ISSN: 2962-2565

This article contributes to:

Sentiment Analysis on NU Online Using KNN Method Important Findings of the Article

Mas Nurul Hamidah[1]*, Fardanto Setyatama[1], R Dimas Adityo[1], Rahmawati Febrifyaning Tias[1]

[1] Department of Engineering, Bhayangkara University Surabaya, 60235, **Indonesia**

*nurul@ubhara.ac.id

Abstract

The large population of Indonesia results in the rapid growth of the mobile economy very quickly, Business of Apps in January 2025 states that more than 200 million Indonesians spend an average of 5.7 hours per day in using cell phones or applications contained on cell phones. One of the applications on PlayStore is NU Online with more than 1 million downloads and there are 27 thousand user reviews with an average rating of 4.8. NU Online App was created by one of the largest organizations in Indonesia, Nahdlatul Ulama, the purpose of making the application is to provide convenience to users, especially among pesantren and NU members. Data Mining is the process of finding patterns, trends, and useful information from large amounts of data. One of the algorithms in data mining is KNN. This research focuses on sentiment analysis on the NU Online application to find out that the review is worth POSITIVE, and NEGATIVE using the KNN method, resulting in 72% showing positive sentiment, 28% showing negative sentiment, this shows the performance and performance of the NU Online love application is very good with the support of the highest accuracy evaluation value at K10 which is 0.83.

Keywords: NU Online, Data Mining, KNN

Article Info
Submitted: 2025-04-02
Revised: 2025-05-19
Accepted: 2025-05-28

Publisher
Universitas Panca Marga

1. Introduction

The rapid development of digital technology is very influential in several fields, one of which is the relationship between an organization's followers and the organization itself. Mobile applications are one of the important platforms in conveying information to followers of the organization, because they provide communication service[1]. In Indonesia NU (Nahdlatul Ulama) is among the largest organizations in Indonesia and even in the world, also utilizing digital platforms through the NU Online application. The goal is as a forum for communication and

information and even services needed by millions of followers and outside community[2].

User engagement with the NU Online App generates a large amount of data on reviews, much of this data contains opinions on features provided by the NU Online app, therefore understanding user sentiment towards the NU Online app is crucial for developers and managers to evaluate, identify areas of improvement, and improve the quality and relevance of the app in the future.

NLP (Natural Language Processing) techniques focus on identifying and extracting subjective opinions from text[3]. This technique is used in processing sentiment analysis in NU Online application. One of the popular and effective classification algorithms in classification tasks, including sentiment analysis, is K-Nearest Neighbors (KNN)[4]. KNN is a simple yet powerful non-parametric algorithm that works by classifying new data based on the majority class of its K nearest neighbors in the feature space[5].

Sentiment analysis aims to analyze the sentiment between users towards NU Online application based on user reviews. The K-Nearest Neighbors (KNN) method was chosen as a classification technique to identify the sentiment polarity (positive, negative) contained in user reviews[6]. It is hoped that the results of this study can provide deep insight into how users perceive their experience using the NU Online application, as well as provide constructive recommendations for future application development. This research contributes to the understanding of the utilization of sentiment analysis in the context of religious organization applications and explores the effectiveness of the KNN method in classifying sentiment on Indonesian language data.

2. Methods

This research methodology consists of 4 stages: 1) Data processing 2) Design 3) Implementation 4) Evaluation. **Figure 1** shows the stages in this research methodology.

2.1 Stage 1: data processing

This stage starts from collecting review data taken from google palystrore on the NU Online application, as many as 100 reviews are taken for data processing, then cleaning irrelevant data for sentiment analysis will be carried out, after which a tokenization technique is carried out, namely breaking the review text into smaller units, called tokens[7]. Furthermore, stemming techniques to remove affixes (prefixes, suffixes, infixes) so that words with the same root can be treated as one entity [8] and stop word techniques, namely the process of removing common words that often appear in text but are considered to have less significant meaning in sentiment analysis[9].

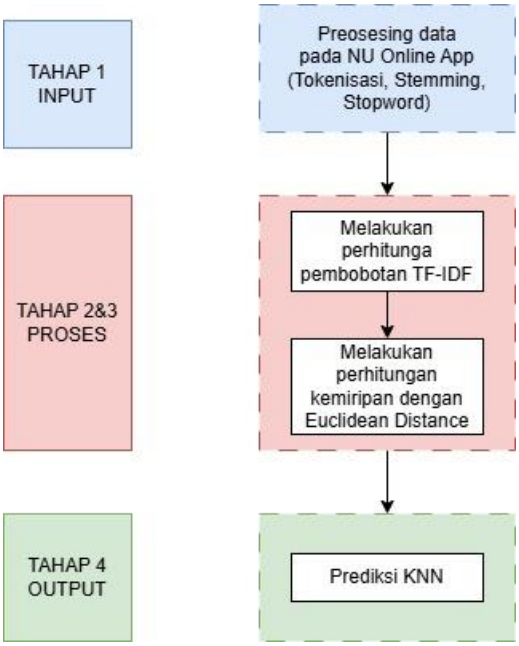


Figure 1. Stages of Research Methodology

Table 1. Review list or feedback list Classification

Document	Review	Class
D1	"NU Online *sangat* membantu. Informasinya lengkap ,akurat dan cepat."	Positif
D2	"Aplikasi ini *memang* bagus. Desainnya menarik dan mudah digunakan."	Positif
D3	"Aplikasi ini *terkadang* lambat dan sering error."	Negatif
D4	"Saya kecewa, desainnya *kurang* modern dan navigasinya membingungkan."	Negatif
D5	"Tampilan NU Online cukup baik, tapi beritanya kadang kurang cepat."	?

2.1 Stage 2: Design

The algorithm used at this stage is the KNN algorithm to perform sentiment analysis on the review data [10] of the NU Online application, starting with manual labeling to determine the polarity of the sentiment, the labels used are positive, and negative [11], after labeling, weighting will be carried out using the Term Frequency-Inverse Document Frequency (TF-IDF) technique[12] . TF-IDF gives weight to each word in the document based on its frequency in the document and the inverse of the word's document frequency in the entire document collection. It aims to highlight words that are important in distinguishing sentiment between reviews [13]. The K value (number of nearest neighbors) is an important parameter in the KNN algorithm[14].

Tabel 2. Preposesing data

Document	Review
D1	◎ Tokenisasi: ['nu', 'online', 'sangat', 'bantu', 'info', 'lengkap', 'akurat', , 'cepat'] ◎ Penghapusan Stopword: ['nu', 'online', 'sangat', 'bantu', 'info', 'lengkap', 'akurat', 'cepat'] ◎ Stemming: ['nu', 'online', 'sangat', 'bantu', 'info', 'lengkap', 'akurat', 'cepat']
D2	◎ Tokenisasi: ['aplikasi', 'memang', 'bagus', 'desain', 'tarik', 'mudah', 'guna'] ◎ Penghapusan Stopword: ['aplikasi', 'memang', 'bagus', 'desain', 'tarik', 'mudah', 'guna'] ◎Stemming: ['aplikasi', 'bagus', 'desain', 'tarik', 'mudah', 'guna']
D3	◎Tokenisasi: ['Aplikasi', 'ini', 'terkadang', 'lambat', 'dan', 'sering', 'error', '.'] ◎Penghapusan Stopword (Opsional: ['aplikasi', 'terkadang', lambat, 'sering', 'eror',] ◎Stemming: ['aplikasi, 'kadang', 'lambat, 'sering', 'eror',]
D4	◎ Tokenisasi: ['Saya', 'kecewa', ',', 'desainnya', 'kurang', 'modern', 'dan', 'navigasinya', 'membingungkan', '.'] ◎ Penghapusan Stopword: ['Saya', 'kecewa', 'desainnya', 'kurang', 'modern', 'navigasinya', 'membingungkan'] ◎ Stemming: ['Saya', 'kecewa', 'desain', 'kurang', 'modern', 'navigasi', 'bingung']
D5	◎ Tokenisasi: ['Tampilan', 'NU', 'Online', 'cukup', 'bagus', ',', 'tapi', 'beritanya', 'kadang', 'kurang', 'cepat', '.'] ◎ Penghapusan Stopword: ['Tampilan', 'NU', 'Online', 'cukup', 'bagus', 'beritanya', 'kadang', 'kurang', 'cepat'] ◎ Stemming: ['Tampil', 'NU', 'Online', 'cukup', 'bagus, 'berita', 'kadang', 'kurang', 'cepat']

According to Salton, the IDF value is obtained with the following equation:

$$IDF = log \frac{D}{dfi} \quad (1)$$

Where D = Number of documents and dfi = number of term occurrences against D Calculation of TF-IDF weight (W) for each document against keywords with the formula:

$$W_{ij} = tf_{ij} log \frac{D}{dfi} \quad (2)$$

Table 3. Example of weighting in NU online application with TF-IDF

Term	tf						idf	Tf*idf				
	D1	D2	D3	D4	D5	df		D1	D2	D3	D4	D5
Nu	1				1	1	0.69	0.69				0.69
Online	1				1	2	0.39	0.39				0.39
Info	1					1	0.69	0.69				
Lengkap	1					1	0.69	0.69				
Akurat	1					1	0.69	0.69				
aplikasi		1	1			2	0.39		0.39	0.39		
Bagus		1			1	2	0.39		0.39			0.39
Desain		1				1	0.69		0.69			
Tarik		1				1	0.69		0.69			
Mudah		1				1	0.69		0.69			
Guna		1				1	0.69		0.69			
Kecewa				1		1	0.69				0.69	
Kurang				1		1	0.69				0.69	
navigasi				1		1	0.69				0.69	
Bingung				1		1	0.69				0.69	
Lambat			1			1	0.69			0.69		
eror			1			1	0.69			0.69		
Kurang				1	1	2	0.39				0.39	0.39
Cepat	1				1	2	0.39	0.39				0.39
cukup					1	1	0.69					0.69
berita					1	1	0.69					0.69

2.3 Stage 3: Implementation

The implementation used in this study uses the Python programming language with libraries such as Scikit-learn. KNN modeling begins with processing training data that has been labeled The training process involves storing a feature representation of the training data [15]. In the KNN process with this training data used to predict sentiment in test data[16] , KNN modeling needs to find the K closest neighbors in the training data based on feature distance using cosine distance [17] and predict sentiment labels based on the majority of labels from these K neighbors [18]. To calculate similarity using the cosine distance formula. As in the formula below:

$$cosine = \sum_k \frac{(d_{ik} d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \quad (3)$$

Where test (i,k) is all records and attributes in the testing data while train (j,k) is all records and attributes in the testing data. Table 4 example of calculating similarity using Cosine distance.

Table 4. similarity with Cosine distance with D5

0.4761	0	0	0
0.1521	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0.1521	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0.1521
0.1521	0	0	0
0	0	0	0
0	0	0	0

2.4 Stage 4: Evaluation

The evaluation stage is needed to find out how well the KNN method performs when applied to review data to determine sentiment analysis, The evaluation metric formula used is Accuracy[19]: The proportion of reviews that are correctly predicted sentiment by the model. The accuracy formula can be seen below:

$$Accuracy = \frac{Jumlah\ prediksi\ (positif, netral\ atau\ negatif)}{jumlah\ keseluruhan\ data} \tag{4}$$

3.Results and Discussion

From the calculation of KNN with a value of K = 3 on 100 review datasets, as done in table 3.2 for prepossessing data removing some conjunctions and making basic words [20] can be described as in figure 2 where the total percentage of positive sentiment is 72% and negative sentiment is 28%.

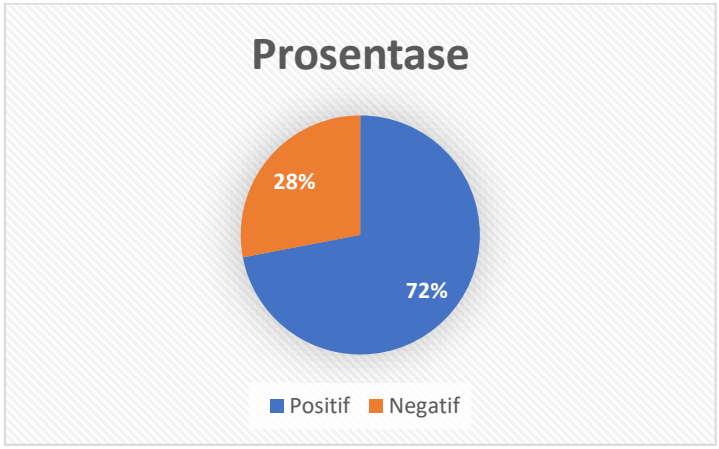
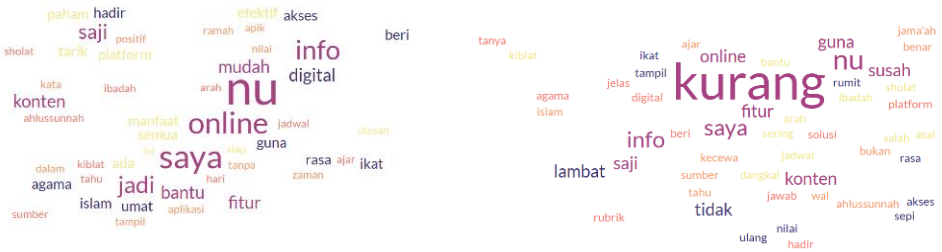


Figure 2. Sentiment Percentage

The frequency of words in each sentiment class can be described as follows:



(a) (b)

Figure 3. Sentiment Word Cloud (a) Positive (b) Negative

In **Figure 3** in positive sentiment, the words that often appear are ('NU', 'online','I','content','info','serving'), while for negative sentiment, they are ('less', 'slow','difficult','content','feature','me'), and in this test, system evaluation is carried out using the accuracy formula, while the calculation results can be seen in Figure 4 below.

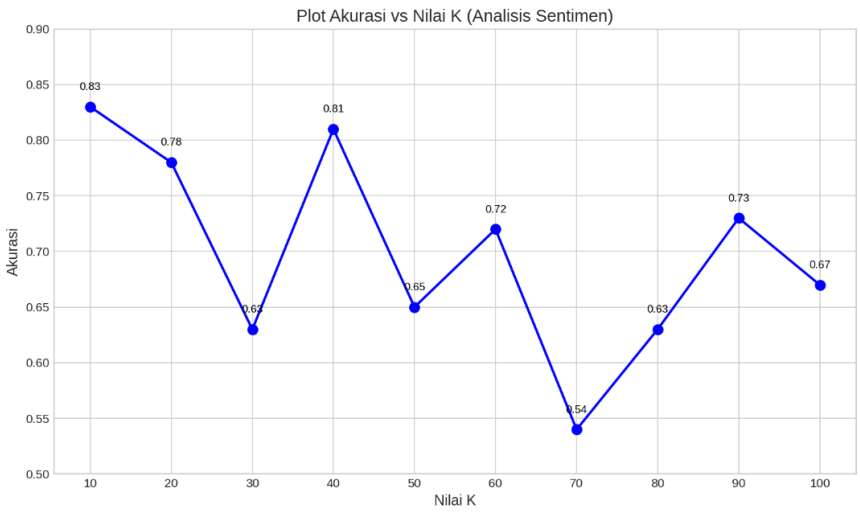


Figure 4. System evaluation with accuracy techniques

It can be seen in **Figure 3** that the best accuracy is at k = 10 which is worth 0.83 while the lowest accuracy is at K = 100 which is worth 0.67, with such results it can be concluded that the accuracy value is quite high with a value above 0.5 and close to the perfect value of 1.

The purpose of this research is to analyze sentiment in the NU online application using the KNN method, starting from data collection of 100 reviews which is carried out by means of scapping techniques on the google paly store which then performs data cleaning by removing unnecessary attributes, after which word processing is carried out on reviews using tokenization techniques, stemming and stop words which are then weighted using TF-IDF techniques, and classification of sentiment into positive and negative categories.

The test results using KNN and the value of k=100 resulted in the smallest accuracy value of 0.67 and the highest was 0.83 in k=10, these results indicate that KNN is quite reliable in capturing sentiment patterns in user reviews of the NU Online application, especially in distinguishing between positive and negative opinions.

An important finding of this study is that it produces higher positive sentiment, it can be seen that the words “help”, ‘easy’ and “use” are of high value, indicating that the NU Online application has met the expectations of its users in presenting digital Islamic content.

Overall, the main contribution of this research is to prove the effectiveness of the KNN method in sentiment classification in the context of digital da'wah applications and identify aspects that are viewed positively or negatively by users. This research can serve as a basis for further studies using other machine learning methods, such as SVM or Random Forest, to compare their performance and accuracy in similar context.

4. Conclusion

Based on the sentiment analysis of the words that appear in the positive sentiment, the application users feel helped by the existence of the NU online application, while in the negative sentiment the users sometimes feel that they lack additional features provided by the NU online application, while the evaluation results have an accuracy value of 0.83 in k=10, Entering more words significantly improves model performance, indicating that a richer vocabulary captures the nuances of sentiment better. The limitation in this research is the relatively small data collection, for further development with more varied and large data.

Authors’ Declaration

Authors’ contributions and responsibilities - The authors made substantial contributions to the conception and design of the study. The authors took

responsibility for data analysis, interpretation, and discussion of results. The authors read and approved the final manuscript.

Funding - No funding information from the authors.

Availability of data and materials - All data is available from the authors.

Competing interests - The authors declare no competing interest.

Additional information - No additional information from the authors.

References

[1] A. D. Adhi Putra, "Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.

[2] Suluh Gembyeng Ciptadi, "The Social Construction of Tolerance Discourse through Online Media: Study of NU Online," *KOMUNIKA J. Dakwah dan Komun.*, vol. 18, no. 1, pp. 113–123, 2024, doi: 10.24090/komunika.v18i1.8431.

[3] P. R. Nlp, B. A. I. Ndonesia, T. I. Tentang, P. Teknologi, D. A. N. A. Nlp, and M. Amien, "(nlp) b," no. 2007, pp. 1–7, 2022.

[4] A. Baita, Y. Pristyanto, N. Cahyono, P. Covid-, K. N. N. Akurasi, and K. Kunci, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) Dan K-Nearest Neighbor (KNN) Abstraksi Keywords :," vol. 4, no. 2, pp. 42–46, 2021.

[5] P. Astuti and N. Nuris, "Penerapan Algoritma KNN Pada Analisis Sentimen Review Aplikasi Peduli Lindungi," vol. 2, no. 2, pp. 137–142, 2022.

[6] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, pp. 302–307, 2023.

[7] V. No, K. Mustaqim, F. A. Maresti, and I. N. Dewi, "Edumatic : Jurnal Pendidikan Informatika Analisis Sentimen Ulasan Aplikasi PosPay untuk Meningkatkan Kepuasan Pengguna dengan Metode K-Nearest Neighbor (KNN)," vol. 8, no. 1, pp. 11–20, 2024, doi: 10.29408/edumatic.v8i1.24779.

[8] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," pp. 43–49.

[9] P. L. Jatika, J. Supriyanto, D. Alita, and A. R. Isnain, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring," vol. 4, pp. 74–80, 2023.

[10] H. Wicaksono, "Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily," vol. 10, no. 3, pp. 3591–3600, 2023.

[11] A. Sentimen, A. Gojek, M. Support, V. Machine, and D. A. N. K. Nearest, "UNNES Journal of Mathematics," vol. 10, no. 2, 2021.

[12] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksmana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer Dike : Jurnal Ilmu Multidisiplin," vol. 1, pp. 37–43,

2023.

[13] A. Uang, K. Baru, and M. Algoritma, "Analisis sentimen terhadap masyarakat adanya uang kertas baru menggunakan algoritma k-nearest neighbor(knn) 1," vol. 7, no. 2, pp. 105–114, 2023.

[14] R. Sari, "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN)," vol. 8, no. 1, pp. 10–17, 2020.

[15] V. No, S. Rahayu, Y. Mz, J. E. Bororing, and R. Hadiyat, "Edumatic : Jurnal Pendidikan Informatika Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP," vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.

[16] A. Y. Permana and M. Makmun, "Analisis Sentimen pada Teks Opini Penilaian Kinerja Dosen dengan Pendekatan Algoritma KNN," vol. 19, pp. 39–50, 2020.

[17] M. N. Fahriza and N. Riza, "Analisis Sentimen Pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer Gpt Menggunakan Metode Klasifikasi K-Nearest Neighbor (Knn) Sistematic Literature Review," vol. 7, no. 2, pp. 1351–1358, 2023.

[18] M. K. Neighbor, A. Asro, and H. Februariyanti, "Analisis Sentimen Pengguna Twitter terhadap Perpanjangan PPKM Menggunakan Diterima : Diterbitkan : Keyword :," vol. 10, no. 1, pp. 17–24, 2022.

[19] J. S. Komputer, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," vol. 5, no. September, pp. 897–910, 2021.

[20] V. No, A. S. Pamungkas, and N. Cahyono, "Edumatic : Jurnal Pendidikan Informatika Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor," vol. 8, no. 1, pp. 1–10, 2024, doi: 10.29408/edumatic.v8i1.24114.