p-ISSN: 2252-3006
e-ISSN: 2685-2411

Performance Analysis of the K-Nearest Neighbors (K-NN) for Sentiment Analysis of Online Loan Application X

I Ketut Agus Leo Gunadarma[a1], Gusti Made Arya Sasmita[a2], I Nyoman Prayana Trisna[a3]
[a]Information Technology Department, Faculty of Engineering,
Udayana University Bukit Jimbaran, Bali, Indonesia, telp. (0361) 701806
e-mail: [1]leogunadarma16@gmail.com, [2]aryasasmita@unud.ac.id, [3]prayana.trisna@unud.ac.id

Abstrak

Ekonomi digital di Indonesia berkembang pesat, termasuk dalam layanan pinjaman online. Aplikasi X adalah salah satu aplikasi pinjaman online yang populer, memberikan kemudahan bagi pengguna dalam mengajukan pinjaman secara online. Penelitian ini menggunakan analisis sentimen terhadap ulasan pengguna aplikasi X untuk memahami preferensi dan kebutuhan mereka. Metode K-Nearest Neighbours (K-NN) diterapkan sebagai algoritma utama untuk klasifikasi sentimen. Data yang dikumpulkan melalui proses scraping ulasan pengguna mengalami serangkaian tahap prapemrosesan, seperti tokenisasi, penghapusan stop word, dan stemming yang ditujukan untuk meningkatkan kualitas data. Model K-NN diuji dalam berbagai skenario untuk menemukan hasil yang terbaik. Skenario terbaik menunjukkan bahwa akurasi tertinggi diperoleh oleh model K-NN ketika proses stop word tidak digunakan pada tahap prapemrosesan data dimana dengan perbandingan akurasi tanpa menggunakan proses stop word sebesar 92.9% dibandingkan dengan menggunakan stopword sebesar 89.9%.

Kata kunci: *Analisis Sentimen, K-Nearest Neighbours, Pinjaman Online, Stop Word, Ulasan*

Abstract

The digital economy in Indonesia is growing rapidly, including in online lending services. Application X is one of the popular online lending applications, offering users convenience in applying for loans online. This research employs sentiment analysis on user reviews of Application X to understand their preferences and needs. The K-Nearest Neighbours (K-NN) method is applied as the primary algorithm for sentiment classification. Data collected through user review scraping undergoes a series of preprocessing stages, such as tokenization, stop word removal, and stemming, aimed at improving data quality. The K-NN model is tested in various scenarios to achieve the best results. The best scenario reveals that the highest accuracy is achieved by the K-NN model when the stop word removal process is not applied during the data preprocessing stage where the accuracy without using the stop word process was 92.9%, compared to 89.9% when using stop words.

Keywords: *Sentiment Analysis, K-Nearest Neighbours, Online Lending, Stop Word, Reviews*

1. Introduction

The development of internet technology in Indonesia has shown a significant trend in recent years. The number of internet users in Indonesia reached around 25 million in 2020. Indonesia ranked 73rd out of 190 countries in ease of doing business, according to the World Bank (2019). This ranking highlights Indonesia's unique appeal to investors and business actors. The influx of foreign capital into portfolio instruments reflects investor confidence in the stability and condition of the Indonesian economy, as well as the prospects and projections of Indonesia's economic policies[1]. The digital business industry in Indonesia has been established for quite some time, but its rapid expansion only started in 2014. According to Euromonitor, e-commerce sales in Indonesia reached USD 1.1 billion that year. Additionally, data from the Central Statistics

Agency indicated a 17% growth in the country's digital business sector in recent years. This figure reflects a significant increase and serves as an indicator that Indonesia is undergoing a transformation towards a rapidly developing digital economy [1]. However, Indonesia is still in the early stages of digital economy development compared to developed countries. Various efforts to enhance and expand digital infrastructure have had a clear positive impact, especially in the business sector, particularly in e-commerce and ride-hailing services [1]. One of the sectors impacted by the development of the digital economy is online lending services.

Online lending is a financial service that allows users to access loans without having to visit physical financial institutions like banks. This service can be accessed through platform-based applications, either mobile apps or websites. As the number of internet users and digital economy penetration in Indonesia increases, online lending services have become a popular choice for people in need of quick funds. Online lending offers the convenience of a faster loan application process with relatively easier requirements compared to conventional loans.

One of the online lending platforms that has become increasingly popular among the Indonesian public is the Online Lending Application X. This application offers various attractive features to its users, such as a loan limit of up to 50 million rupiahs, quick disbursement, and requirements that are easy to meet for many people. These advantages have made the Online Lending Application X the primary choice for those in need of emergency funds.

The Online Lending Application X is also registered and supervised by Otoritas Jasa Keuangan (OJK) the Financial Services Authority in accordance with the regulations set out in POJK Number 77/POJK.01/2016 on Peer-to-Peer Lending Services. The supervision by OJK aims to ensure that the services provided by the online lending platform are safe and comply with applicable legal provisions, thereby preventing misuse that could harm consumers. OJK’s stringent oversight makes users feel safer and more protected when accessing this service. As the popularity and use of online lending applications like the Online Lending Application X increase, it is important for service providers to continue understanding the needs and preferences of users. One effective way to achieve this is by using sentiment analysis.

Sentiment analysis is a computational process to understand and interpret opinions, feelings, and attitudes contained in documents such as comments, reviews, or user feedback, where Natural Language Processing (NLP) is used to analyze text and identify emotions or opinions [2]. Online lending applications can use sentiment analysis to gain insights into user experiences, satisfaction, and expectations regarding the services provided. This information is crucial for decision-making in service improvement and the development of new features that better meet market needs.

Sentiment analysis works by classifying a text based on the emotions it attempts to display. Generally, the classifications in sentiment analysis are categorized into positive and negative. Sentiment analysis looks at opinions at different levels such as document level, sentence level, and aspect level, as opposed to understanding text through documents, paragraphs, sentences, clauses, or phrases [3]. The results of sentiment analysis are then used to help businesses monitor brand and product sentiment from the input provided by consumers and understand what consumers want [3]. It is important to note that when conducting effective sentiment analysis, selecting the right algorithm is crucial for obtaining accurate results. One algorithm that can be used in sentiment analysis is K-Nearest Neighbors (K-NN). The technique used in K-NN is to classify data with objects that have the closest neighboring values [4]. Essentially, this algorithm works by searching for the "nearest neighbors" of the data to be analyzed and then classifying the data based on the majority label of its nearest neighbors.

K-NN belongs to the instance-based learning group. This algorithm is also one of the lazy learning techniques. K-NN works by finding the k closest objects in the training data to a new or test object. After collecting the KNN, the majority of the KNN is then taken to predict the test sample. The proximity of the neighbors is typically calculated using the Euclidean distance [5].

The KNN method has several advantages, such as simple training, fast processing, easy to understand, and being effective when the training data size is large [5]. This makes K-NN highly suitable for sentiment analysis on online lending platforms, where the volume of data analyzed can be quite large. A study by Furqan et al. (2022) demonstrated that K-NN, combined with TF-IDF, performed well in analyzing sentiment related to COVID-19 policies, achieving an accuracy of 94.5%. This research uses the K-Nearest Neighbor method to classify opinions from 1,000 tweets collected from the social media platform Twitter. Through various data preprocessing techniques, the study successfully identified 811 positive sentiments and 189 negative

 p-ISSN: 2252-3006
e-ISSN: 2685-2411

sentiments[6]. However, despite its many advantages, there are some challenges that need to be addressed, such as handling imbalanced data that may affect classification results, as well as dealing with words that may influence the classification outcome.

This study aims to evaluate the performance of the K-NN algorithm in performing sentiment analysis on user reviews of the Online Lending Application X. The main focus of this research is to compare the effectiveness of the K-NN algorithm in classifying user sentiment and to identify parameters that can influence its performance, such as accuracy, variations in sentiment count, and the use of stop words. The results of this study are expected to contribute significantly to the development of marketing strategies and management of online lending applications by providing a better understanding of user preferences and satisfaction.

2. Research Method / Proposed Method

This study analyzes user sentiment toward the X online loan application using the K-Nearest Neighbors (K-NN) algorithm as the primary classification method. Data used in this study includes 10,000 user reviews collected from the Google Play Store. The research process is illustrated in the accompanying flow diagram as shown in Figure 1, which outlines each stage from data collection to visualization of results.

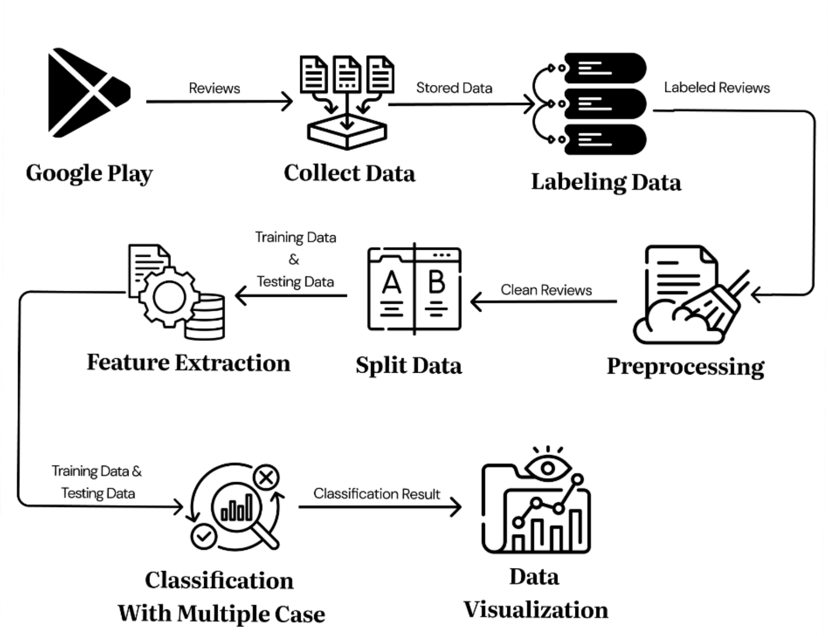

Figure 1. Research Overview

The first stage, Data Collection, involves gathering user reviews from the Google Play platform. These reviews are stored and subsequently processed in the Labeling phase, where each review is manually labeled as either "positive" or "negative" to ensure accurate sentiment classification. This labeling provides a solid foundation for the training and testing data used in the subsequent stages.

Once labeled, the reviews move to the Preprocessing stage, where they are cleaned to remove unnecessary characters, standardize case formatting, and tokenize words. Stopwords are also removed to focus on words with meaningful sentiment value, while stemming reduces words to their base forms, enhancing consistency across the dataset. The result is a set of "clean reviews" ready for further processing.

After preprocessing, the data enters the Split Data stage, where it is divided into training and testing sets to allow for model evaluation. Two split scenarios are used: one with an 80-20 split and another with a 90-10 split, to assess how training set size affects model accuracy.

In the case of the balanced dataset scenario, which consists of 10,000 reviews that have passed the preprocessing stage, the 80-20 split will divide the dataset into 8,000 training reviews and 2,000 testing reviews, while the 90-10 split will divide it into 9,000 training reviews and 1,000 testing reviews. On the other hand, for the imbalanced dataset scenario, which consists of 6,000 reviews with 3,000 positive reviews and 3,000 negative reviews, the 80-20 split will divide the

 p-ISSN: 2252-3006
e-ISSN: 2685-2411

dataset into 4,800 training reviews and 1,200 testing reviews, while the 90-10 split will divide it into 5,400 training reviews and 600 testing reviews.

In the Feature Extraction phase, the cleaned reviews are transformed into a numerical format using Term Frequency-Inverse Document Frequency (TF-IDF). This method assigns weight to words based on their importance within the dataset, allowing the K-NN algorithm to focus on words that carry significant sentiment information.

After feature extraction, the Classification with Multiple Cases step is performed using the K-NN algorithm. Zero conditions are applied to evaluate the model's performance. The first condition tests the model with stopwords included and excluded, to understand how the presence of common words impacts classification accuracy. The second condition involves testing the model on both balanced and imbalanced datasets, where the imbalanced dataset uses a total of 10,000 reviews, and the balanced dataset uses 6,000 reviews, to provide insight into how K-NN performs with different data distributions. To enhance model reliability, the optimal k value (the number of nearest neighbors) is determined by selecting the neighborhood size that yields the highest performance from the predefined sets of neighbors. This strategy helps mitigate potential overfitting and underfitting issues, optimizing the model's generalization capabilities.

The final step, Data Visualization, presents the results through bar charts showing the distribution of positive and negative sentiments, as well as word clouds highlighting the most frequently used words in each category. These visualizations provide a clear understanding of sentiment trends and user preferences within the dataset.

The sequence of stages in this methodology, as depicted in the diagram, reflects a logical and structured approach to obtaining accurate sentiment analysis results. By following this flow, the study aims to build a robust K-NN model that can provide valuable insights into user satisfaction and identify potential areas for service improvement in online loan applications.

3. Literature Study

Literature study is a reference that contains material related to the research taken. This section explores foundational concepts and previous studies relevant to sentiment analysis using the K-Nearest Neighbors (K-NN) algorithm. This review is structured to provide clarity on the techniques, evaluation metrics, and algorithmic decisions specific to sentiment analysis within the digital financial services sector.

3.1. Sentiment Analysis

Sentiment analysis or opinion mining is the process of understanding, extracting, and processing textual data automatically to obtain the sentiment information contained within an opinion. Sentiment analysis is conducted to observe the tendency of opinions or views on an issue or an object by an individual, whether it tends to be viewed negatively or positively[7]. The categories of positive and negative grouping can be simply determined based on the words used in the sentiment. For example, words like "good," "excellent," "satisfied," and similar terms fall into the positive category, while words like "disappointed," "bad," "sad," and similar terms fall into the negative category[6]. The benefits of sentiment analysis in a business context are highly significant. Developers can evaluate user feedback on their products and make informed decisions for further development quickly and efficiently [8].

3.2. Preprocessing

Preprocessing of text is one of the components in the Text Mining algorithm. It is the process of transforming documents into structured data according to its purpose, so that it can be further processed [9]. This structured data can then be further processed for analysis or additional information processing, the steps involved in text processing include case folding, cleaning, normalization, stop words removal, tokenization, and stemming. At this stage, the scraped data will be processed to standardize its form and format [10].

3.3. Term Frequency-Inverse Document Frequency (TF-IDF)

To achieve effective sentiment classification with K-NN, robust feature extraction is essential, with Term Frequency-Inverse Document Frequency (TF-IDF) being a favored method. TF-IDF is a process of calculating or extracting words into a numerical vector format, which is used to determine the weight of a word within a document or corpus[11]. The TF-IDF technique eliminates terms that are very common and extracts terms with high relevance from the corpus. The more frequently a term appears in a document, the higher its value will be [12].

3.4. K-Nearest Neighbors (K-NN)

The K-Nearest Neighbors (K-NN) algorithm is a popular choice for sentiment analysis, particularly for text data classification. In the data classification process, KNN works by classifying new data points based on their proximity or distance to existing data points in the training set. Each data point in the training set has a class label, and KNN identifies the K nearest neighbors of the test data point by measuring the distance using metrics such as Euclidean distance [13]. K-NN's simplicity and accuracy in proximity-based classification make it widely applicable in sentiment analysis [5]. For example, Rahayu et al. (2022) applied K-NN in their study on the Flip application, a financial service, achieving an accuracy of 76.86% in sentiment classification [14].

3.5. Evaluation Methods

The evaluation stage is the phase where the performance of the classification algorithm used in the research is assessed [15]. To measure the performance of the K-NN model, this study uses a 2x2 confusion matrix as the primary evaluation method. The confusion matrix provides a detailed view of the model's classification performance, comparing the predicted results with the actual outcomes and categorizing them into four main components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values enable the calculation of evaluation metrics such as accuracy, precision, recall, and F1-score, which offer a comprehensive assessment of the model's effectiveness in sentiment classification studies provide a solid foundation for using K-NN and TF-IDF in sentiment analysis within digital financial services. The current study applies K-NN exclusively to analyze sentiment in user reviews of application X, focusing on evaluating K-NN's classification effectiveness in this context [16].

4. Result and Discussion

The K-NN algorithm in this study will be tested using 10,000 user reviews from the X online loan application under different scenarios. To facilitate understanding, the scenarios will be grouped and presented in a table, as shown in Table 1.

Table 1. Analysis Testing Scenario

Testing Scenarios	**Analysis Using Stop Word Processing at the Preprocessing Stage.**	**Analysis Using Sentiment Data with a Balanced Ratio of Positive and Negative Sentiments (50:50)**
Normal analysis	✓	-
Balanced analysis	✓	✓
Normal analysis without using stop word	-	-
Balanced analysis without using stop word	-	✓

The Normal scenario has no modifications are made to the sentiment data, it is tested in its original distribution. The Balanced scenario indicates that the sentiment data is adjusted to have an equal proportion of positive and negative reviews. Each scenario will be tested using the K-NN algorithm with two data split ratios for training and testing: 80:20 and 90:10. This approach allows for a thorough assessment of the K-NN algorithm's performance across different data distributions and proportions.

4.1. Performance Comparison

The performance comparison is conducted by examining the accuracy levels produced by the K-NN model across various testing scenarios. The analysis focuses on understanding the effectiveness of K-NN in accurately classifying the data. The results of these tests are presented in detail in Table 2, providing a comprehensive overview of the model's performance in each scenario.

Table 2. Comparison Accuracy of K-NN model

Testing Scenarios	K-NN Model Accuracy
Normal Analysis with 80:20 ratio data	0.8980 (K=22)
Balanced Analysis with a data ratio of 80:20	0.8792 (K=20)
Normal Analysis Without Using Stop Word 80:20	0.9275 (K=25)
Balanced Analysis Without Using Stop Word with a data ratio of 80:20	0.8967 (K=15)
Normal Analysis with 90:10 ratio data comparison	0.8990 (K=18)
Balanced Analysis with 90:10 ratio data comparison	0.8917 (K=16)
Normal Analysis Without Using Stop Word with 90:10 ratio data comparison	0.9290 (K=19)
Balanced Analysis Without Using Stop Word with 90:10 ratio data comparison	0.9033 (K=18)

Table 2 presents a comparison of model performance, focusing on the K-NN model across various testing scenarios. Based on the analysis results, the K-NN model achieved the highest accuracy of 0.9290. This accuracy was reached in the normal analysis scenario without stop words, using a 90:10 data ratio and a K parameter set to 19. These results indicate that the K-NN model performs exceptionally well in classifying data. The highest accuracy achieved highlights its effectiveness compared to other methods in the study. The removal of stop words was found to play a crucial role in enhancing the model's accuracy. A more balanced dataset and optimal ratio further supported this analysis, providing deeper insight into the strengths of K-NN in specific testing scenarios.

4.2. Data Usage Comparison

The comparison of data usage will test the accuracy of the K-NN model by comparing its performance on balanced and imbalanced data. This testing process aims to observe the differences in results obtained when using two different types of data. Table 3 provides a clear overview of the comparison results from the tests conducted on both data conditions.

Table 3. Comparison Accuracy of K-NN model

Testing Scenarios	Accuracy of using Balanced Data	Accuracy of using Unbalanced Data
K-NN Using Stop Word with a data ratio of 80:20	0.8792 (K=20)	0.8980 (K=22)
K-NN Without Using Stop Word with data comparison ratio 80:20	0.8967 (K=15)	0.9275 (K=25)
K-NN Using Stop Word with 90:10 ratio data comparison	0.8917 (K=16)	0.8990 (K=18)
K-NN Without Using Stop Word with data ratio 90:10	0.9033 (K=18)	0.9290 (K=19)

Table 3 presents a comparison of model performance with balanced and imbalanced data. The K-NN model achieved the highest accuracy of 0.9290 in a normal testing scenario without using stop words, with a data ratio of 90:10 and K=19. This indicates that the K-NN model is capable of making highly accurate predictions on imbalanced data. Although this result is the main highlight, it is important to note that the model's performance may vary depending on the data distribution and parameters used in other tests. A more detailed analysis of metrics such as the percentage of False Positives (FP), False Negatives (FN), and the F1-score for each scenario provides a more comprehensive view of K-NN's performance across various tests, as shown in Table 4.

Table 4. Percentage of FP, FN and f1-score of Naïve Bayes and K-NN Models

Types of Analysis	**FP K-NN**	**FN K-NN**	**F1- Score K-NN**
80:20 Normal Analysis	6.90%	3.30%	0.8473
90:10 Normal Analysis	6.60%	3.50%	0.8477
80:20 Balanced Analysis	9.50%	2.58%	0.8870
90:10 Balanced Analysis	8.33%	2.50%	0.8976
No Stop Word Analysis 80:20	4.05%	3.20%	0.8868
No Stop Word Analysis 90:10	4.00%	3.10%	0.8892
Balanced Analysis without Stop Word 80:20	7.25%	3.08%	0.8973
Balanced Analysis without Stop Word 90:10	6.50%	3.17%	0.9065

Table 4 presents the percentage of False Positive (FP) and False Negative (FN) errors in the K-NN model. In the tests, the K-NN model shows specific tendencies regarding the impact of data distribution and stop word removal on its performance. The model records an improvement in the F1-score when stop word removal is applied, indicating that this step enhances prediction quality, particularly in balanced data scenarios. In imbalanced data scenarios, the K-NN model faces challenges in predicting the less dominant class, which can affect the distribution of FN and FP. Review data with a higher proportion of negative categories compared to positive ones becomes a key factor in the distribution of these prediction errors.

4.2. Stop Word Performance Comparison

The analysis of stop word usage compares the accuracy of the K-NN model based on the application or removal of stop words during the preprocessing stage. The results of this accuracy comparison are presented in Table 5 to provide an overview of the impact of stop word handling on the model's performance.

Table 5. Comparison of Stop Word Usage

Testing Scenario	Accuracy using the Stop Word process	Accuracy not using the Stop Word process
K-NN Normal with data ratio 80:20	0.8980 (K=22)	0.9275 (K=25)
K-NN Balanced with data ratio 80:20	0.8792 (K=20)	0.8967 (K=15)
K-NN Normal with 90:10 ratio data comparison	0.8990 (K=18)	0.9290 (K=19)
K-NN Balanced with data ratio 90:10	0.8917 (K=16)	0.9033 (K=18)

Table 5 compares the model performance based on the use of balanced and imbalanced data. The K-NN model achieved the highest accuracy of 0.9290 in a normal analysis scenario without stop word removal, with a data ratio of 90:10 and K=19. This result demonstrates the advantage of K-NN under these conditions. Additionally, Table 4.6 reveals that the accuracy of the K-NN model improves when the stop word removal process is not applied. This improvement is related to the number of words in the sentiment data, which influences the model's ability to better understand the context of the data.

4.3. Visualization

The visualization with the highest accuracy in the analysis of normal data without stop words using the K-NN model, with a testing-to-training data ratio of 90:10, is presented as a bar chart comparing the total predicted results for positive and negative sentiments with the labeled sentiments. The bar chart can be seen in Figure 2.

p-ISSN: 2252-3006
e-ISSN: 2685-2411

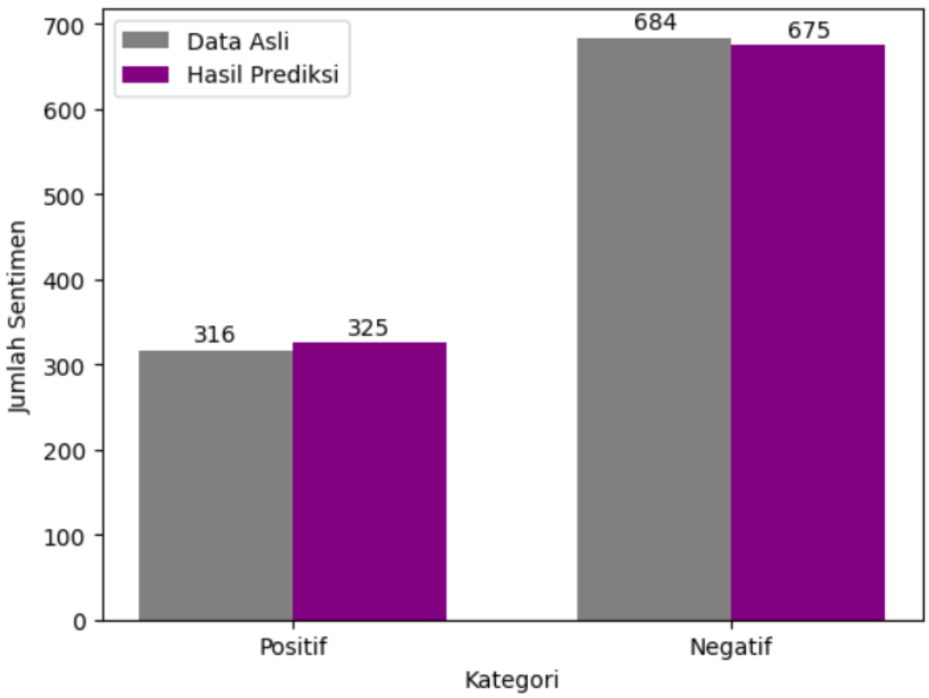

Figure 2. Bar Chart of K-NN Model Normal Analysis Scenario Without Using Stop Word With 90:10 Division

Figure 2 is a bar chart illustrating the total predicted results for positive and negative sentiments alongside the labeled sentiments from the analysis of normal data without stop words using the K-NN model with a testing-to-training data ratio of 90:10. The figure shows that the model predicted 325 positive reviews out of 316 and 675 negative reviews out of 684 based on a total of 1000 testing data. The next visualization is a word cloud for the positive category, which is shown in Figure 3.

Figure 3. Positive Word Cloud From K-NN Model Normal Analysis Scenario Without Using Stop Word With 90:10 Division

Figure 3 is a word cloud for the positive category, displaying a collection of words from the predicted positive sentiments in the analysis of normal data without stop words using the K-NN model with a testing-to-training data ratio of 90:10. The figure shows that the top three most frequent words are "dan", "sangat", and "cepat".The next visualization is a word cloud for the negative category, which is shown in Figure 4.76.

Figure 4. Negative Word Cloud From K-NN Model Normal Analysis Scenario Without Using Stop Word With 90:10 Division

Figure 4.76 is a word cloud for the negative category, displaying a collection of words from the predicted negative sentiments in the analysis of normal data without stop words using the K-NN model with a testing-to-training data ratio of 90:10. The figure shows that the top three most frequent words are "saya", "tidak", and "di".

5. Conclusion

This study analyzes the performance of the K-NN algorithm in sentiment classification of reviews on the online loan application X, using 10,000 reviews collected from the Google Play Store between March and May 2023. The data was divided into 6,000 reviews for a balanced data testing scenario. The research process includes data collection through scraping, labeling positive and negative categories, preprocessing to standardize data formats, and constructing a K-NN model that was tested under various scenarios, such as with and without stop word removal. The results show that K-NN achieved the highest accuracy of 0.9290 in the normal analysis scenario without stop words, with a 90:10 data ratio and K=19. The model demonstrated stable performance on both balanced and unbalanced data, with a significant improvement in performance when stop words were excluded. This study emphasizes that K-NN has great potential to produce highly accurate sentiment classification, especially with optimization of parameters such as the value of K and the removal of stop words.

References

[1]. *Tanabrun -The Importance of Digital Economic Growth in Indonesia*. (n.d.).

[2]. Pratama, D., & Informatika, S. (n.d.). Mengintegrasikan Teknik Nlp Dan Machine Learning Untuk Meningkatkan Analisis Sentimen. In *Teknologipintar.org* (Vol. 3, Issue 11).

[3]. Rifqy, O. :, Turjaman, M., & Budi, I. (2022). *Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter)* (Vol. 30). Agustus.

[4]. Marita Febrianti, Y., & Wahyu Widodo, A. (2018). *Analisis Sentimen Pada Ulasan “Lazada” Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance* (Vol. 2, Issue 10). http://j-ptiik.ub.ac.id

[5]. Cahyanti, D., Rahmayani, A., & Ainy Husniar, S. (2020). *Indonesian Journal of Data and Science Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara*. *1*(2), 39–43.

[6]. Furqan, M., Mayang Sari, S., & Ilmu Komputer Fakultas Sains dan Teknologi, P. (2022). *Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia Sentiment Analysis using K-Nearest Neighbor towards the New Normal During the Covid-19 Period in Indonesia* (Vol. 21, Issue 1). www.tripadvisor.com

[7]. Rusydiana, A. S., & Marlina, L. (n.d.). Journal of Economics and Business Aseanomics Journal homepage Analisis Sentimen terkait Sertifikasi Halal. *Journal of Economics and Business Aseanomics*, *5*(1), 69–85. http://academicjournal.yarsi.ac.id/jeba

[8]. Prana, T., Sukma, W., & Pribadi, M. R. (n.d.). *Analisis Sentimen Review Pengguna Viu pada Play Store dengan Algoritma Random Forest*. *2*(1).

[9]. Ridwansyah, T. (2022). KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. *Media Online*, *2*(5), 178–185. https://djournals.com/klik

[10]. Darmawan, G., Alam, S., Imam Sulistyo, M., Studi Teknik Informatika, P., Tinggi Teknologi Wastukancana Purwakarta, S., & Artikel, R. (2023). *Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes Info Artikel Abstrak*. *2*(3), 100–108. https://doi.org/10.55123

[11]. Tri Putra, K., Amin Hariyadi, M., & Crysdian, C. (n.d.). *Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm*.

[12]. Riza Adrianti Supono, & Muhammad Azis Suprayogi. (2021). Perbandingan Metode TF-ABS dan TF-IDF Pada Klasifikasi Teks Helpdesk Menggunakan K-Nearest Neighbor. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(5), 911–918. https://doi.org/10.29207/resti.v5i5.3403

[13]. Yanti, S. N., Yuhandri, & Sumijan. (2024). Implementasi K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Generasi Sandwich. *Jurnal KomtekInfo*, 371–379. https://doi.org/10.35134/komtekinfo.v11i4.586

[14]. Rahayu, S., MZ, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, *6*(1), 98–106. https://doi.org/10.29408/edumatic.v6i1.5433

[15]. Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). *Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes*. *9*(2), 785–795.

[16]. Fikri, M. I., Sabrila, T. S., Azhar, Y., & Malang, U. M. (n.d.). *Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter*.