PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 20, No. 2 September 2024 | DOI: 10.33480/pilar.v20i2.5736

COMPARISON OF MACHINE LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS OF DIGITAL IDENTITY APPLICATION USERS

Muhammad Naufal Maulana Abrari[1*]; Ferian Fauzi Abdulloh[2]

Informatika[1, 2]
Universitas Amikom, Yogyakarta, Indonesia[1, 2]
https://www.amikom.ac.id/[1, 2]
naufal.abrari@students.amikom.ac.id[1*], ferian@amikom.ac.id[2]
(*) Corresponding Author

Abstract— *In the rapidly evolving digital era, the Population Identity Application (IKD) plays a crucial role in streamlining civil administration processes in Indonesia, allowing easier and faster access to population services. This study aims to explore the application of machine learning algorithms in analyzing user responses to the IKD application. Three popular algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Naïve Bayes were chosen to classify sentiment from 1301 user reviews on the Google Play Store into positive and negative categories. After performing data preprocessing such as tokenization and stemming, hyperparameter optimization was conducted using GridSearchCV to enhance classification accuracy. The research results indicate that the SVM algorithm, optimized with hyperparameters, including the use of the rbf kernel and a parameter value of C = 1, achieved the highest accuracy of 85.60%, making it the most effective method for sentiment classification of the IKD application. These findings provide valuable insights for the government and developers in refining the features and performance of IKD, contributing to the efficiency and security of digital administration in Indonesia. Furthermore, this study opens opportunities for further development that is more responsive to user needs and expectations in the future.*

Keywords: *digital identity application, k-nearest neighbors, naïve bayes, sentiment analysis, support vector machine.*

Abstrak— *Dalam era digital yang berkembang pesat, Aplikasi Identitas Kependudukan (IKD) memainkan peran penting dalam menyederhanakan proses administrasi sipil di Indonesia, memungkinkan akses yang lebih mudah dan cepat terhadap layanan kependudukan. Studi ini bertujuan untuk mengeksplorasi penerapan algoritma pembelajaran mesin dalam menganalisis respons pengguna terhadap aplikasi IKD. Tiga algoritma populer, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), dan Naïve Bayes, dipilih untuk mengklasifikasikan sentimen dari 1301 ulasan pengguna di Google Play Store menjadi kategori positif dan negatif. Setelah melakukan prapemrosesan data seperti tokenisasi dan stemming, optimasi hiperparameter dilakukan menggunakan GridSearchCV untuk meningkatkan akurasi klasifikasi. Hasil penelitian menunjukkan bahwa algoritma SVM dengan optimasi hyperparameter, termasuk penggunaan kernel rbf dan nilai parameter C = 1, mencapai akurasi tertinggi sebesar 85,60%, menjadikannya metode paling efektif untuk klasifikasi sentimen pada aplikasi IKD. Temuan ini memberikan wawasan berharga bagi pemerintah dan pengembang dalam menyempurnakan fitur dan kinerja IKD, serta berkontribusi pada efisiensi dan keamanan administrasi digital di Indonesia. Selain itu, studi ini membuka peluang untuk pengembangan lebih lanjut yang lebih responsif terhadap kebutuhan dan ekspektasi pengguna di masa depan.*

Kata Kunci: *identitas kependidikan digital, k-nearest neighbors, naïve bayes, analisis sentimen, support vector machine.*

INTRODUCTION

The Digital Identity Application (IKD) represents an innovative approach in civil administration, utilizing digital technology to streamline the access and verification of population data. This application is a digital version of the e-KTP, introduced as a mobile application in the form of photos or QR codes (Dukcapiladmin, 2023). The government aims for the IKD application to eliminate the need for printed documents or physical KTP cards for administrative purposes, thus saving on the costs associated with electronic KTP blanks (Purnamasari, 2023). This application is connected and integrated with various services, including healthcare, education, social services, banking, taxes, and payment gateways, facilitating easy use by the public in today's technological landscape (Purnamasari, 2023).

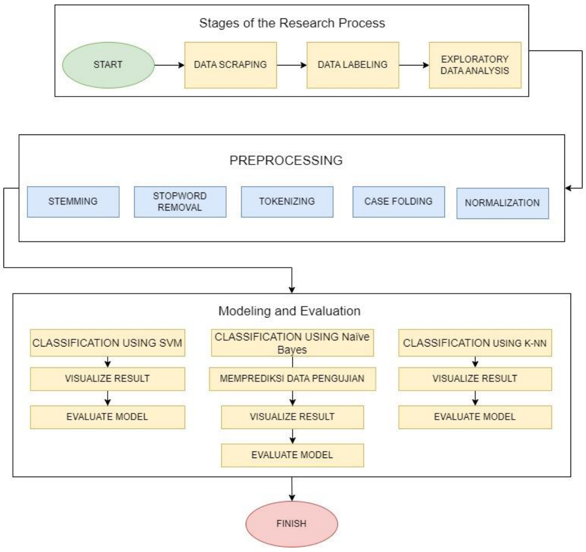
While this innovation aims to improve the efficiency and quality of population data management, there is public debate regarding the implementation and security of the IKD (Jayasinga & Triono, 2023). Some people, particularly in remote areas, are concerned about limited internet access and a lack of information about the application (Maimori, Eliwatis, & Syafriwaldi, 2022). Concerns have also been raised about the security of digital identity data and difficulties in accessing the IKD application. However, previous research has not specifically analyzed public sentiment towards the IKD application on the Play Store (Hidayat et al., 2024), highlighting a gap in understanding public perceptions more deeply.

To support this digitalization effort, it is crucial to gather information on public evaluations of IKD through sentiment analysis (Zhang, Zhong, Wei, & Zhang, 2022). Previous research that serves as a reference for this study compared the performance of Support Vector Machine (SVM) and Naïve Bayes algorithms in sentiment analysis of regional elections (PILKADA) based on Twitter data. This research aimed to analyze public sentiment towards PILKADA by leveraging social media data, highlighting the strengths and weaknesses of each algorithm. The study found that NBC achieved an accuracy of 81.7%, recall of 81.7%, and precision of 80%, while SVM achieved an accuracy of 80.7%, recall of 80.7%, and precision of 84%. These results indicate that NBC outperforms SVM in terms of accuracy and recall, while SVM excels in precision (Romaito, Anam, Rahmaddeni, Ulfah, & Noviciate, 2021). These findings provided the basis for choosing the classification methods used in this research, aiming to achieve a comprehensive understanding of public sentiment towards the IKD application based on user comments from the Google Play Store.

Based on the background described, this research is conducted to determine the sentiment generated from discussions about the Digital Identity Application (IKD) from user comments on the Play Store, by comparing the methods K-Nearest Neighbors (K-NN), Naive Bayes, SVM, and Neural Network (Wahyuningsih & Hendry, 2023). This study aims to provide a deeper understanding of public views on the IKD application and offer valuable insights to the government for developing a more efficient and secure multifunctional identity system. In this study, data is collected based on user comments who downloaded the Digital Identity Application on the Play Store (KEMENDAGRI, 2022). The samples taken will be classified into two categories: positive and negative sentiment (Safitri, Umaidah, & Maulana, 2023).

MATERIALS AND METHODS

This research employs three methods: Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Naïve Bayes, to determine which is most effective in analyzing user sentiment regarding the IKD application from reviews by users who have downloaded it on the Play Store. The main goal of this study is to delve into user complaints about the IKD application to assist the government in further improving the app. The research methodology is outlined in Figure 1.



Source: (Research Results, 2024)
Figure 1. Research methodology

Data Scrapping

The data used in this study consists of user reviews collected through data scraping techniques (Maulana, Voutama, & Ridwan, 2023) on the Google Play Store. The researcher utilized the `google-play-scraper` library (JoMingyu, 2024) during this phase. A total of 1301 reviews were randomly selected for the analysis.

Data Labeling

The data obtained from scraping is then labeled as either positive or negative. Labeling is conducted collaboratively by several labelers to ensure maximum accuracy (Doloksaribu & Yusran Timur Samuel, 2022). The researchers did not rely on the score, as some reviews were found to be irrelevant to the given ratings. Positive reviews

Vol. 20, No. 2 September 2024 | DOI: 10.33480/pilar.v20i2.5736

typically contain suggestions, praise, and user satisfaction, while negative reviews express complaints, protests, and negative feelings such as disappointment and anger towards the IKD application.

Exploratory Data Analysis (EDA)

The purpose of Exploratory Data Analysis (EDA) is to gain initial insights into the existing data set before proceeding with more detailed research processes (Merdiansah, Siska, & Ali Ridha, 2024). EDA enables the extraction of various preliminary insights about the analyzed data, such as data quantity, completeness, and the number of positive and negative label.

Data Preprocessing

Data preprocessing is a crucial step before developing text mining models (Hermawan, Jowensen, Junaedi, & Edy, 2023). Researchers briefly analysis of the data characteristics using several preprocessing steps, including normalization, case folding, tokenizing, stopword removal, and stemming.

A. Normalization: This process involves removing disruptive elements from unstructured text to make it easier to analyze sentiment (Rahmawati & Sukmasetya, 2022). At this stage, researchers corrected non-standard words into standard forms in the user reviews (Safitri et al., 2023). Examples of normalized words can be seen in Table 1.

Table 1. Normalization of Words

No	Non-Standard Words	Standard Words
1	gk,g,gak,ngak,gx,tdk,ga,ngga	Tidak
2	tp,tetapi,tapi	Namun
3	ok,oke,okee,okk	Baik
4	sdh,udh,udah	Sudah
5	apk	Aplikasi
6	eror	Error
7	bs,bsa	Bisa
8	dpt,dpat,dapet	Dapat
9	yg	Yang
10	krn,krna	Karena
11	smtr,smntara	Sementara

Source: (Research Results, 2024)

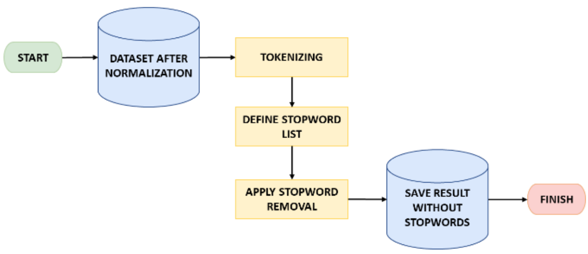
B. Case Folding: This step involves converting all text in user reviews to lowercase (Ditami, Ripanti, & Sujaini, 2022). The goal here is to ensure that the text processed by the model is in a uniform format. The steps in the case folding process can be seen in Figure 2.

Source: (Research Results, 2024)
Figure 2. Case Folding Process

C. Tokenizing: This stage breaks down text into individual words (Balit, M. N. B., Utomo, 2024). The purpose of this step is to present text in a format that is easier for the model to understand, allowing the algorithm to more easily recognize patterns in the text. The steps in the tokenization process are shown in Figure 3.

Source: (Research Results, 2024)
Figure 3. Tokenizing Process

D. Stopword Removal: In software engineering, stopword removal significantly enhances tool performance compared to a general stop list, as shown in previous studies (Fan, Arora, & Treude, 2023). Examples of commonly appearing words that do not provide information include "and," "or," "is," and others; these words are known as stopwords. The stopword removal stage aims to eliminate these words from the text so that the model can focus more on meaningful and relevant words. By reducing this noise, the model can analyze the text more efficiently. The process is depicted in Figure 4.



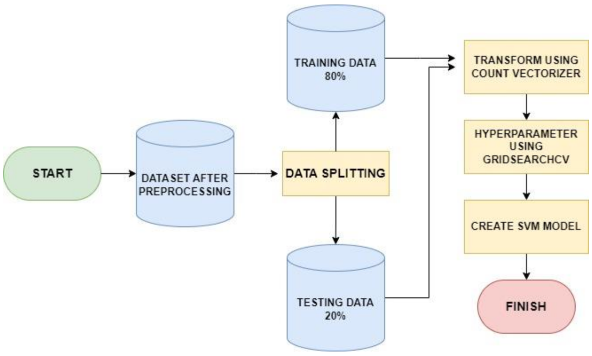
Source: (Research Results, 2024)
Figure 4. Stopword Removal Process

E. Stemming: Stemming is the process of transforming inflected words into their base form (Safryda Putri & Ridwan, 2023). This stage aims to reduce the number of unique words that need to be processed by the model, thereby improving its performance. The steps in the stemming process are illustrated in Figure 5.

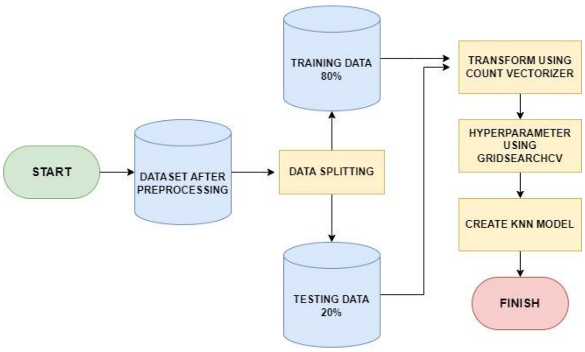
Source: (Research Results, 2024)
Figure 5. Stemming Process

Classification Using the Three Algorithms

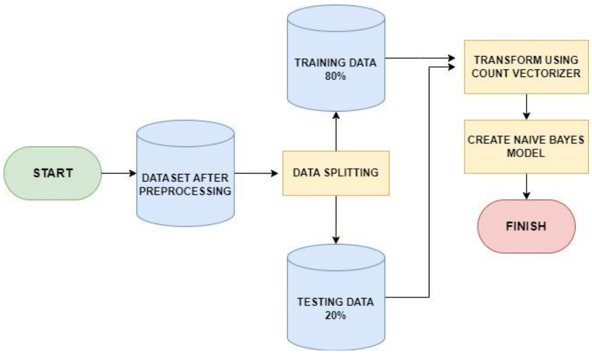
This study employs three algorithms: SVM, K-NN, and Naïve Bayes, to classify user reviews of the IKD application. The data is split into training (80%) and testing (20%) sets and then transformed into numerical representations using Count Vectorizer to learn the vocabulary from the training data and create features based on word frequency (Turki & Roy, 2022). The three models are then optimized using GridSearchCV to find the best parameters (Muslim et al., 2020). The results show that SVM performs best in classifying reviews with a score of 85.6%, significantly outperforming the other two methods. At this stage, the classification process is shown in three different images. Figure 6 illustrates the classification process using the SVM (Support Vector Machine) method. Figure 7 shows the classification process using the K-NN (K-Nearest Neighbors) method. Finally, Figure 8 shows the classification process using the Naïve Bayes method.



Source: (Research Results, 2024)
Figure 6. Classification Process Using the SVM Method



Source: (Research Results, 2024)
Figure 7. Classification Process Using the K-NN Method



Source: (Research Results, 2024)
Figure 8. Classification Process Using the Naïve Bayes Method

Visualization of Results

Word clouds are an effective visual tool for displaying frequently appearing words in a text or review. By using word clouds, researchers can quickly visualize the frequency of words and gain insight into potential patterns in the reviews or texts being analyzed (Supriyanto & Rosalin, 2023).

Evaluation of Results

This study uses a confusion matrix to calculate accuracy by comparing the correct or incorrect predictions of the classification methods against actual data or prediction targets (Cahyono & Dewi Setiyawati, 2023). The calculation formulas are as follows :

$$\text{Accuracy} = \frac{TP+TN}{\text{Total}} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

The confusion matrix for calculating accuracy, precision, and recall is explained in these formulas (Normawati & Prayogi, 2021),(Krstinić, Braović, Šerić, & Božić-Štulić, 2020).

Definitions:

A. TP (True Positive) = The number of data points from the positive class correctly classified as positive.
B. TN (True Negative) = The number of data points from the negative class correctly classified as negative.
C. FP (False Positive) = The number of data points from the negative class incorrectly classified as positive.
D. FN (False Negative) = The number of data points from the positive class incorrectly classified as negative.

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 20, No. 2 September 2024 | DOI: 10.33480/pilar.v20i2.5736

RESULTS AND DISCUSSION

This study uses user reviews of the Digital Identity Application (IKD) available on the Google Play Store. The data, consisting of user comments, were randomly collected on May 29, 2024. Examples of the data scraped can be seen in Table 2.

Table 2. Results of Data Scraping

Username	Score	Date	Content
Ferbi Fadha Utama	1	5/14/2024 8:38:00 AM	Tolong aplikasi nya diperbaiki, tidak bisa masuk LOADING TERUS, padahal koneksi internet stabil, loading selalu gagal & aplikasi masih mudah di Hack (diretas)
Kristo Kepler	1	5/4/2024 2:51:00 AM	Aplikasi sampah banget nih, penanggung jawabnya hoflok, aplikasinya gak guna, respon lama, selalu lost koneksi, hadeuhh, pembodohan digital ini namanya.
Reza Purba	4	6/24/2023 6:20:00 AM	Sungguh aplikasi yg bisa digunakan untuk pemula buat KTP gk sulit dan cepat untuk di pelajari

Source : (Research Results, 2024)

After collecting a total of 1300 data points, the next step is data labeling. The labeling process aims to classify the data into two categories: positive and negative. An example of the data labeling process is shown in Table 3.

Table 3. Example of Data Labeling Results

Content	Sentiment
Tolong aplikasi nya diperbaiki, tidak bisa masuk LOADING TERUS, padahal koneksi internet stabil, loading selalu gagal & aplikasi masih mudah di Hack (diretas)	Negative
Aplikasi sampah banget nih, penanggungjawabnya hoflok, aplikasinya gak guna, respon lama, selalu lost koneksi, hadeuhh, pembodohan digital ini namanya.	Negative
Sungguh aplikasi yg bisa digunakan untuk pemula buat KTP gk sulit dan cepat untuk di pelajari	Positive

Source: (Research Results, 2024)

Following the labeling process, an exploratory data analysis is conducted to perform an initial examination of the dataset. The analysis revealed no missing or null values within the dataset. Out of the 1300 data points categorized, 650 were labeled as positive sentiment, while the remaining 650 were labeled as negative sentiment.

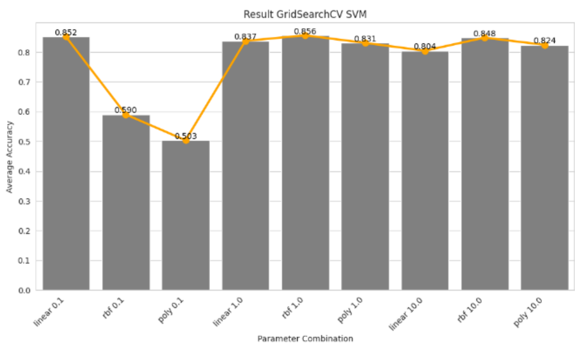
The next phase involves data preprocessing, which is a critical step in text mining. An example of data preprocessing can be seen in Table 4.

Table 4. Results of Data Preprocessing

Stage	Preprocessing Result
Initial Data	Aplikasinya memudahkan kita buat yang sering mobile, jadi ga ribet keluarin KTP dari dompet. Dan urusan lainnya. Terus aman juga buat share, krn yg di share bukan berbentuk foto KTP.
Normalization	Aplikasinya memudahkan kita buat yang sering mobile jadi tidak ribet keluarin KTP dari dompet Dan urusan lainnya Terus aman juga buat share karena yang di share bukan berbentuk foto KTP
Case Folding	aplikasinya memudahkan kita buat yang sering mobile jadi tidak ribet keluarin ktp dari dompet dan urusan lainnya terus aman juga buat share karena yang di share bukan berbentuk foto ktp
Tokenizing	['aplikasinya', 'memudahkan', 'kita', 'buat', 'yang', 'sering', 'mobile', 'jadi', 'tidak', 'ribet', 'keluarin', 'ktp', 'dari', 'dompet', 'dan', 'urusan', 'lainnya', 'terus', 'aman', 'juga', 'buat', 'share', 'karena', 'yang', 'di', 'share', 'bukan', 'berbentuk', 'foto', 'ktp']
Stopword Removal	['aplikasinya', 'memudahkan', 'mobile', 'ribet', 'keluarin', 'ktp', 'dompet', 'urusan', 'aman', 'share', 'share', 'berbentuk', 'foto', 'ktp']
Stemming	['aplikasi', 'mudah', 'mobile', 'ribet', 'keluarin', 'ktp', 'dompet', 'urus', 'aman', 'share', 'share', 'bentuk', 'foto', 'ktp']

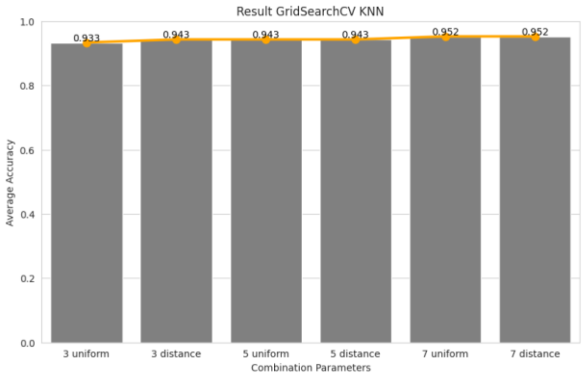
Source : (Research Results, 2024)

The next step involves hyperparameter optimization using GridSearchCV to find the most optimal parameter combinations for the SVM method. The parameter C is tested with values including 0.01, 0.05, 0.25, 0.5, 0.75, 1, and 10. Meanwhile, the kernel types tested include linear, RBF, and poly. This process uses GridSearchCV to obtain the best parameter combinations, as shown in Figure 9.



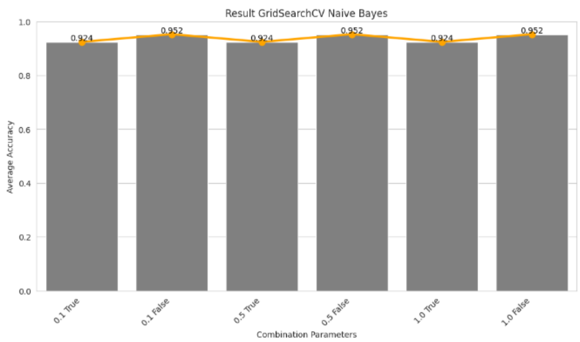
Source : (Research Results, 2024)
Figure 9. GridSearchCV Support Vector Machine

Next, for the K-NN method, hyperparameter optimization is also performed using GridSearchCV to find the best parameter combinations. The difference lies in the parameters used, which include the number of nearest neighbors (n_neighbors) and distance metrics (metric). The n_neighbors values tested include 3, 5, 7, 9, and 11, while the distance metrics tested include Euclidean, Manhattan, and Minkowski. The GridSearch for K-NN can be seen in Figure 10.



Source : (Research Results, 2024)
Figure 10. GridSearchCV K-Nearest Neighbors

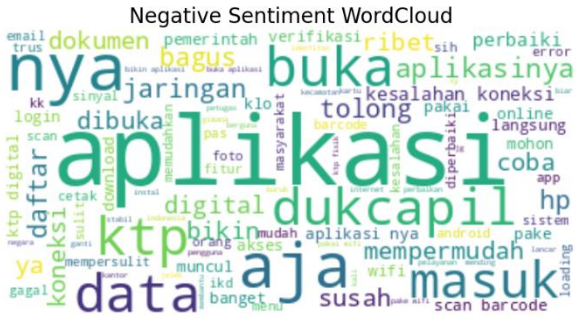
Lastly, hyperparameter optimization using GridSearchCV is conducted for the Naïve Bayes method. In this model, researchers tested several parameter combinations, including smoothing values (var_smoothing) and other settings relevant to the Naïve Bayes model. The results of the GridSearch can be seen in Figure 11.



Source : (Research Results, 2024)
Figure 11. GridSearchCV Naïve Bayes

Based on these results, the SVM method, when developed with parameter values of C at 1 and kernel RBF, achieved the best accuracy compared to the other methods, reaching 85.60%. This configuration allowed the SVM model to effectively classify text data into relevant categories. The model was then trained using the preprocessed training data, and its performance was evaluated on test data to ensure the desired accuracy.

The visualization of the negative word cloud provides a visual representation of words frequently appearing in the data labeled as negative, offering insights into the most commonly expressed complaints or shortcomings of the IKD application, as shown in Figure 12.



Source : (Research Results, 2024)
Figure 12. Negative Sentiment Word Cloud

According to the results, the SVM method, with parameter values of 'C' at 1 and RBF kernel, achieved the highest accuracy of 85.60%, compared to Naïve Bayes at 81.28% and K-Nearest Neighbor (KNN) at 74.10%. This configuration enabled the SVM model to effectively classify text data into relevant categories. The superior performance of SVM can be attributed to several factors: First, SVM is well-suited for text classification tasks involving high-dimensional data, as it finds the optimal hyperplane that maximizes the margin between different classes. This characteristic is particularly useful in this study, where text data is represented by a large number of features. Second, SVM's robustness to overfitting, achieved through the regularization parameter 'C', helps control the trade-off between a low error on training data and minimizing the norm of the weights. This prevents overfitting and ensures better generalization to unseen data. Third, the effective use of the RBF kernel in SVM allows it to handle non-linear relationships in the data by mapping input features into higher-dimensional space.

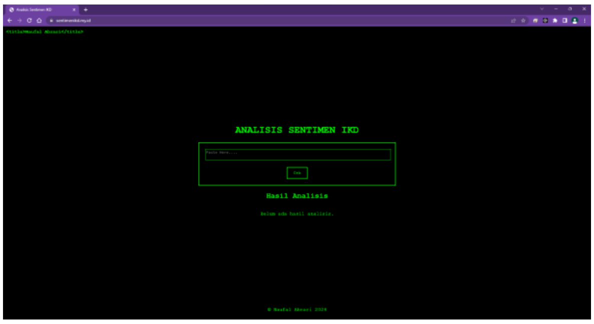
In contrast, the Naïve Bayes algorithm, which achieved an accuracy of 81.28%, performed moderately well but was not as effective as SVM. Naïve Bayes is a probabilistic classifier based on Bayes' theorem, which works well with large datasets and is computationally efficient. However, its assumption of feature independence can be a limitation in text classification tasks where features (words) are often correlated. Despite this, Naïve Bayes showed a precision of 0.78, recall of 0.86, and an F1-score of 0.82 for the negative class, and a precision of 0.85, recall of 0.77, and an F1-score of 0.80 for the positive class.

The K-Nearest Neighbor (KNN) algorithm showed the lowest accuracy at 74.10%. The lower performance of KNN can be explained by its sensitivity to irrelevant features and computational inefficiency. KNN considers all features equally when computing distances, which can be problematic in high-dimensional spaces where many features may be irrelevant or redundant. Additionally, KNN requires computing distances between the test instance and all training instances, making it computationally intensive, especially with large datasets. Furthermore, in high-dimensional spaces, the distance between data points becomes less meaningful, leading to poorer performance in classification tasks. For the negative class, KNN showed a precision of 0.70, recall of 0.84, and an F1-score of 0.76, while for the positive class, it had a precision of 0.80, recall of 0.64, and an F1-score of 0.71.

Thus, the SVM model provided more consistent results and higher accuracy in classifying text data into positive and negative categories. The effectiveness of SVM in handling high-dimensional text data and its robustness to overfitting made it the best-performing algorithm for sentiment analysis in this study.

Deployment

To implement the findings of this research, a sentiment analysis application for IKD has been deployed on a web platform. This deployment involves storing the trained model file and using Python as the deployment technique, allowing users to input review texts and receive sentiment analysis results in real time. This platform is expected to assist the government and application developers in better understanding user preferences and complaints, thereby improving the quality and performance of the IKD application. This deployment also supports the sustainability of innovation in digital population administration in Indonesia. The web deployment can be seen in Figure 13.



Source: (Research Results, 2024)
Figure 13. Display of the Cluster Model

CONCLUSION

Testing was conducted using three algorithms—Support Vector Machine (SVM), K-Nearest Neighbours (K-NN), and Naïve Bayes—for analysing the Digital Population Identity application. It was found that SVM had the best performance. With optimal parameters of C = 1 and an RBF kernel, SVM achieved an accuracy of 85.60%, outperforming both K-NN and Naïve Bayes. The SVM model also demonstrated superior results across evaluation metrics such as precision, recall, and F1-score. Based on these findings, SVM is identified as the best algorithm for sentiment analysis of the IKD application users in this study. These findings are expected to aid the government in developing the IKD application to better meet user needs and expectations.

However, there are still unresolved problems in this study. One limitation is the reliance on user reviews from the Google Play Store, which may not represent the entire user population. Future work could involve collecting data from multiple sources to get a more comprehensive understanding of user sentiment. Additionally, exploring other machine learning algorithms and deep learning approaches could further improve the accuracy and robustness of sentiment analysis. Future research could also investigate the integration of real-time sentiment analysis to provide immediate feedback to developers and policymakers. By addressing these limitations and exploring these potential areas, the development and implementation of the IKD application can be significantly enhanced.

REFERENCE

Balit, M. N. B., Utomo, F. S. (2024). Sentiment Analysis of pegipegi.com Review on Google Play Store with Naïve Bayes. *Sistemasi: Jurnal Sistem Informasi*, *13.3*, 1044–1053. https://doi.org/10.32520/stmsi.v13i3.3913

Cahyono, N., & Dewi Setiyawati. (2023). Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia. *Indonesian Journal of Computer Science*, *12*(1), 262–272. https://doi.org/10.33022/ijcs.v12i1.3154

Ditami, G. R., Ripanti, E. F., & Sujaini, H. (2022). Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *8*(3), 508. https://doi.org/10.26418/jp.v8i3.56478

Doloksaribu, H. P., & Yusran Timur Samuel. (2022). Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi. *Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, *16*(1), 1–11. https://doi.org/10.47111/jti.v16i1.3747

Dukcapiladmin. (2023). Identitas Kependudukan Digital (IKD) / Digital IDo Title. Retrieved July 1, 2024, from https://dukcapil.madiunkab.go.id/identitas-kependudukan-digital-digital-id/

Fan, Y., Arora, C., & Treude, C. (2023). Stop Words for Processing Software Engineering Documents: Do they Matter? *Proceedings - 2023 IEEE/ACM 2nd International Workshop on Natural Language-Based Software Engineering, NLBSE 2023*, 40–47. https://doi.org/10.1109/NLBSE59153.2023.00016

Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, *12*(1), 129–137. https://doi.org/10.23887/jstundiksha.v12i1.52358

Hidayat, R., Nur Rahman, R., Reifin Perdana, M., Teknik Informatika, P., Sains dan Teknologi, F., & Muhammadiyah Kalimantan Timur, U. (2024). Analisis Sentimen Aplikasi Identitas Kependudukan Digital (IKD) Menggunakan Metode Naïve Bayes. *Jurnal Sistem Informasi Dan Ilmu Komputer*, *2*(1), 129–140. Retrieved from https://doi.org/10.59581/jusiik-widyakarya.v2i1.2320

Jayasinga, I. P. A., & Triono, A. (2023). Digitalization of Population Administration to Facilitate Public Services in the Era of Regional Autonomy. *International Journal of Multicultural and Multireligious Understanding*, *10*(5), 484. https://doi.org/10.18415/ijmmu.v10i5.4725

JoMingyu. (2024). google-play-scraper 1.2.7. Retrieved July 1, 2024, from pypi.org website: https://pypi.org/project/google-play-scraper/

KEMENDAGRI, D. D. (2022). Identitas Kependudukan Digital. Retrieved July 1, 2024, from https://play.google.com/store/apps/details?id=gov.dukcapil.mobile_id&hl=

Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). *Multi-label Classifier Performance Evaluation with Confusion Matrix*. 01–14. https://doi.org/10.5121/csit.2020.100801

Maimori, R., Eliwatis, & Syafriwaldi. (2022). Utilization of Information Communication Technology on Education and Social Change of Village Community. *JURNAL AT-TAGHYIR*, *5*(1), 157–178. https://doi.org/10.24952/taghyir.v5i1.6011

Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC. *Jurnal Teknologi Terpadu*, *9*(1), 42–48. https://doi.org/10.54914/jtt.v9i1.609

Merdiansah, R., Siska, S., & Ali Ridha, A. (2024). Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, *7*(1), 221–228. https://doi.org/10.55338/jikomsi.v7i1.2895

Muslim, M. A., Sciences, N., Semarang, U. N., Info, A., Mining, T., Machine, S. V., & Search, G. (2020). Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy. *Journal of Soft Computing Exploration*, *1*(1), 8–15. https://doi.org/10.52465/joscex.v1i1.3

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *5*(2), 697–711. https://doi.org/10.30645/j-sakti.v5i2.369

Purnamasari, D. D. (2023). Pengguna Identitas Kependudukan Digital Ditarget Capai 50 Juta Jiwa di 2023Title. Retrieved July 1, 2024, from https://www.kompas.id/baca/polhuk/2023/05/03/pengguna-identitas-kependudukan-digital-ditarget-capai-50-juta-jiwa-di-2023

Rahmawati, C., & Sukmasetya, P. (2022). Sentimen Analisis Opini Masyarakat Terhadap Kebijakan Kominfo atas Pemblokiran Situs non-PSE pada Media Sosial Twitter. *JURIKOM (Jurnal Riset Komputer)*, *9*(5), 1393. https://doi.org/10.30865/jurikom.v9i5.4950

Romaito, E. S., Anam, M. K., Rahmaddeni, Ulfah, & Noviciate, A. (2021). *Dalam Analisa Sentimen Pilkada Pada Twitter*. 169–179. Retrieved from https://doi.org/10.22303/csrid.13.3.2021.169-179

Safitri, T., Umaidah, Y., & Maulana, I. (2023). Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine. *Journal of Applied Informatics and Computing*, *7*(1), 28–35. https://doi.org/10.30871/jaic.v7i1.5039

Safryda Putri, D., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine. *Jurnal Ilmiah Informatika*, *11*(01), 32–40. https://doi.org/10.33884/jif.v11i01.6611

Supriyanto, B. F. S., & Rosalin, S. (2023). Analisis Sentimen Program Merdeka Belajar dengan Text Analysis Wordcloud & Word Frequency. *Jurnal Minfo Polgan*, *12*(1), 25–32. https://doi.org/10.33395/jmp.v12i1.12312

Turki, T., & Roy, S. S. (2022). Novel Hate Speech Detection Using Word Cloud Visualization and Ensemble Learning Coupled with Count Vectorizer. *Applied Sciences (Switzerland)*, *12*(13). https://doi.org/10.3390/app12136611

Wahyuningsih, N., & Hendry, H. (2023). Perbandingan Metode Klasifikasi Dalam Analisis Sentimen Masyarakat Terhadap Identitas Kependudukan Digital (Ikd). *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *8*(4), 1218–1227. https://doi.org/10.29100/jipi.v8i4.4155

Zhang, S., Zhong, H., Wei, C., & Zhang, D. (2022). Research on Logistics Service Assessment for Smart City: A Users' Review Sentiment Analysis Approach. *Electronics (Switzerland)*, *11*(23), 1–18. https://doi.org/10.3390/electronics11234018