Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

Sentiment Analysis of Digitalization of Small and Medium Enterprise on Social Media X Using SVM and KNN Methods

Muhammad Dzakiyuddin Haidar, Kemas Muslim Lhaksmana*

School of Computing, Telkom University, Bandung, Indonesia
Email: [1]haidarzaki@student.telkomuniveristy.ac.id, [2,*]kemasmuslim@telkomuniversity.ac.id
Correspondence Author Email: kemasmuslim@telkomuniversity.ac.id
Submitted: **14/01/2025**; Accepted: **28/03/2025**; Published**: 28/03/2025**

Abstract−The rapid digitalization of Small and Medium Enterprises (SMEs) has led to significant shifts in business operations, especially in their adaptation to digital platforms. Public perception towards this digital transformation is crucial to understand, as it reflects the success and acceptance of these efforts. This research conducts sentiment analysis on social media platform X to classify public opinions regarding the digitalization of SMEs. The analysis employs two machine learning algorithms, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), using Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction. The study compares the performance of both models under baseline and hyperparameter-tuned conditions. The results show that the SVM model consistently outperforms KNN in terms of accuracy, precision, recall, and F1-score. The highest accuracy achieved by the SVM model is 81.97% after hyperparameter tuning with a sigmoid kernel. Meanwhile, the best KNN model records an accuracy of 81.31% using Manhattan distance with 11 neighbors. This study demonstrates that SVM provides better stability and performance in sentiment classification related to SME digitalization. The findings are expected to help policymakers better understand public sentiment and formulate more effective strategies for supporting SME digital transformation.

Keywords: Sentiment Analysis; Small and Medium Enterprises (SMEs); Digitalization; Support Vector Machine (SVM); K-Nearest Neighbor (KNN); Social Media Analysis; Hyperparameter Tuning

1. INTRODUCTION

Digitalization has become the main driver of transformation in various economic sectors, including Small and Medium Enterprises (SMEs), which play a strategic role in the national economy. SMEs contribute significantly to Indonesia's Gross Domestic Product (GDP) and employment. According to various studies, the contribution of SMEs to the national GDP reaches around 60%, while employment by SMEs is close to 97% of the total workforce [1]. The government continues to encourage the digitalization of SMEs to improve competitiveness and expand market reach. Digitalization allows small businesses to utilize digital platforms to market their products, improve efficiency, and reach a wider market [2]. However, despite this huge potential, there are still significant challenges in the digital adoption process, such as low digital literacy, limited technological infrastructure, and resistance to work culture change[3][2]. These challenges require intensive assistance and greater government support to help SMEs optimally utilize digital technology.

Social media, particularly *X*, has become one of the most widely used platforms for expressing public opinion. As a microblogging platform, Twitter presents very dynamic data and reflects the various views of the public on certain issues, including the digitalization of SMI. [4]. Sentiment analysis is an effective method for identifying public opinion expressed in text, especially on social media such as Twitter. It can classify opinions into positive, negative, or neutral sentiments, allowing researchers to understand trends and public perceptions of digitization policies by the government and industry players [5][6]. Studies show that sentiment analysis techniques can be used to monitor real-time community responses to social and economic issues [7]. This approach is very useful in assessing policy effectiveness and increasing public engagement in the process of digitizing SMEs.

In the context of SMI digitization, sentiment analysis can provide important insights into how people are responding to the changes taking place. Data from Twitter provides an overview of support or resistance to the digital transformation undertaken by SMEs. The sentiment analysis method used involves Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms, which have proven effective in classifying sentiments into positive and negative. [8]. The feature extraction technique used is Term Frequency-Inverse Document Frequency (TF-IDF), which allows processing data with large dimensions more accurately [9]. SVM was chosen due to its good performance in handling data with many features and has been applied in various studies for sentiment analysis on social media [10]. Meanwhile, KNN is used as a comparison algorithm in evaluating the performance of the model by measuring the classification accuracy on diverse data.

Several previous studies have explored this topic. Research conducted [11] highlights that various sentiment analysis techniques can be employed to categorize restaurant reviews as either positive or negative. The methods used include Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and a combination of SVM with Principal Component Analysis (PCA). The findings revealed that the SVM model achieved an accuracy of approximately 96%, outperforming other classification models. Another research [12] compared five machine learning algorithms, including SVM and KNN, in Twitter sentiment analysis related to the 2023 global recession. The results showed that the SVM algorithm achieved the highest accuracy in sentiment classification compared to the other algorithms. Research [13] evaluates the performance of SVM, KNN, and Naive Bayes algorithms in sentiment classification. The study found that SVM provides higher accuracy than KNN and Naive Bayes in sentiment analysis. Research [14]

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

analyzes consumer responses to Micro, Small, and Medium Enterprises (MSMEs) products that conduct business through social media. Sentiment analysis is done by collecting customer comments on MSME Instagram accounts and processing them using the SVM method. The obtained results are SVM accuracy of 87.09% before optimization and 89.28% after optimization, while the random forest achieves accuracy of 86.42% before optimization and 87.09% after optimization. These results indicate that the SVM model performs better than the Random Forest model in sentiment classification. Another research [15] evaluated the classifier performance using the terms recall, accuracy and precision, as well as the F1 score. Support vector machine obtained the highest accuracy (92%), followed by KNN and Naive Bayes, with 88% and 85% accuracy, respectively.

Although these studies have made significant contributions to the field of sentiment analysis, the focus has not been specifically on the digitization of SMEs. Most of the previous studies mostly discuss sentiment related to public policy, product reviews, or other social issues. In addition, the comparison between SVM and KNN algorithms in the context of digitalization of SMEs on social media has not been widely explored. Therefore, this study aims to fill the gap by conducting sentiment analysis on the digitalization of SMEs on Twitter social media, using SVM and KNN methods, and utilizing Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction..

The purpose of this research is to explore the public sentiment towards the digitization of SMEs expressed through Twitter. This research aims to compare the performance of SVM and KNN algorithms in classifying positive and negative sentiments contained in Twitter data. It is expected that the results of this study can provide deeper insights into public perceptions of SMI digitalization, as well as provide recommendations regarding the most effective sentiment analysis methods for similar contexts in the future. The results of this study are also expected to help policy makers in designing a more targeted and responsive SMI digitalization strategy to the needs of the community.

2. RESEARCH METHODOLOGY

2.1 Research Stages

In this study, the development of a sentiment analysis system for digitalization of small and medium industries from twitter datasets using the SVM and KNN methods is divided into several stages, the stages have been illustrated using a flowchart diagram in Figure 1.

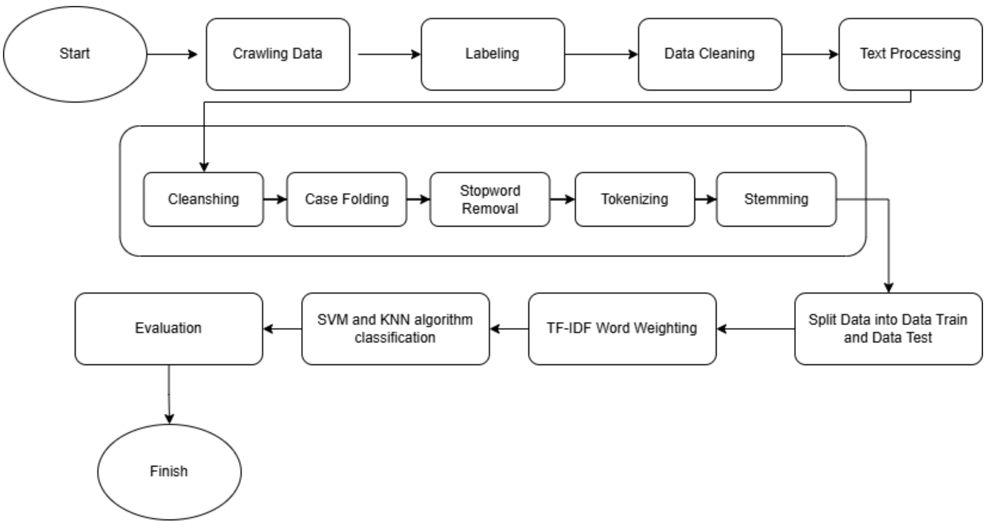

Figure 1. Research Process

The system model starts with data crawling, then sentiment data labeling (positive, negative, neutral), and data cleaning. Next, text processing is performed which includes cleansing, case folding, stopword removal, tokenizing, and stemming. After that, the data is divided into train data and test data, then weighted using TF-IDF. The processed data is classified using SVM and KNN algorithms.

2.2 Crawling Data

The data for this study was obtained from Platform X, which contains text data. The data collected covers a certain period of time, namely before, during the second term of the 7th President Joko Widodo in office. The data collection process was carried out by crawling techniques using Tweet Harvest software version 2.6.8 developed by Helmi Satria. The process of crawling data from Twitter is done by utilizing the Twitter API[16]. To determine whether the scraped news dataset is relevant to the topic of the digitalization of small and medium enterprises (SMEs). Data collected was 1,525 tweets, a WordCloud analysis is conducted. This process highlights the most frequently occurring words in the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

dataset, providing an overall reflection of its content. The resulting WordCloud image offers insights into the dataset's dominant themes and keywords in Figure 2.

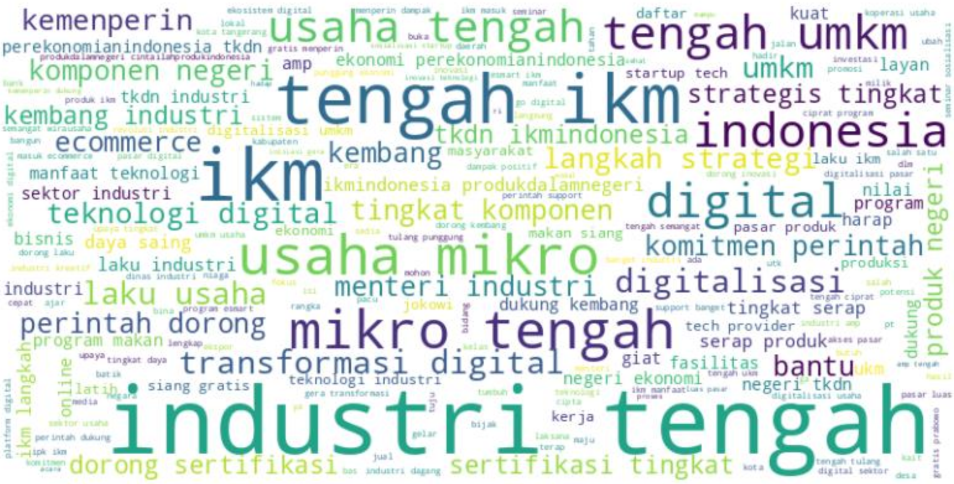

Figure 2. Frequently Occurring Words in the X Dataset

2.3 Labeling

Labeling is the process of assigning positive or negative labels to those in the dataset. This process classifies data into two categories based on the sentiment or polarity conveyed. In sentiment analysis, data such as user reviews or comments on social media are generally classified as “positive” or “negative” according to the tone or meaning of the text[17]. There were 876 tweets with positive sentiment and 649 tweets with negative sentiment.

2.4 Data Cleaning

Data cleaning is a crucial step in processing raw data to ensure it is prepared for analysis. This process involves several actions aimed at eliminating errors, reducing noise, resolving inconsistencies, and removing irrelevant or unnecessary information, thereby enhancing the accuracy and reliability of the analysis outcomes [18]. Table 1 shows the data that has gone through the labeling and cleaning process.

Table 1.Dataset That has been Cleaned and Labeled

Full Text	Sentiment
Indibiz meluncurkan 4 rangkaian solusi digital terbaru menjadikan total 11 island untuk mendukung digitalisasi usaha kecil dan menengah di Indonesia. Dengan solusi ini Indibiz mempermudah pelaku bisnis serta meningkatkan efektivitas dan efisiensi. #Indibiz #indibizjtd https://t.co/QAaxarscR7	1
@Indibiz_jtdiy Indibiz emang juara dalam mendukung digitalisasi usaha kecil menengah ya. Semoga semakin banyak yang terbantu!	1
Unpopular opinion: IKM mestinya mandiri secara finansial. Dalam artian menerima Rp0 dari PPKM. Sistem administrasi selama ini amburadul dan meski sudah digitalisasi tahun ini kok rasanya tetep rawan miskom (gap generasi ortu dan mhs tinggi). Jadi lebih baik hapus saja	0

2.4 Text Processing

Data processing is the process of processing and organizing the data that has been collected so that it can be used effectively in the analysis. The data obtained is usually unstructured and contains various irrelevant elements that need to be filtered first [19]. In the text processing stage, there are several techniques used, including cleansing, changing letters into a consistent format (case folding), removing common words that have no significant meaning (stopword removal), breaking text into words or tokens (tokenizing), and returning words to their basic form (stemming) [19].

2.4.1 Cleansing

Cleansing involves the removal of irrelevant or distracting elements from the data, including punctuation marks, emoticons, and noise. This step is essential to ensure that the data is clean and structured, as the presence of unnecessary elements can interfere with the accuracy and consistency of the data processing. By eliminating these disruptions, the cleansing process helps improve the quality of the data, making it more suitable for reliable analysis and reducing the risk of errors in subsequent stages [19].

2.4.2 Case Folding

Case folding is the process of converting letters in a text into a uniform format, namely by converting uppercase (capital) letters into lowercase letters or vice versa [19]. This process aims to equalize the form of words that have the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

same meaning but differ in lettering, such as "Data" and "data," so that they are considered the same entity during analysis.

2.4.3 Stopword Removal

Stopword removal is the process of removing words that are considered not to make an important contribution to data analysis, such as conjunctions or common words that often appear in the text. This step aims to filter out words that are irrelevant and do not affect the main meaning of the sentence, leaving only words that are more meaningful and significant [19]. By removing stopwords, the data mining process becomes more efficient as it focuses only on words that have an effect on the analysis results, thus increasing the accuracy in extracting relevant information from the data set.

2.4.4 Stemming

Stemming is the process of removing affixes on a word to return it to its base form or root word [19]. This stage aims to simplify words that have a variety of forms due to the use of prefixes, suffixes, or inserts, so that all word forms derived from the same root word can be considered as one entity during analysis [19]. For example, the word "*menengah*," will be processed into the root word "*tengah*." This stemming process is important to improve the accuracy of text analysis, especially in grouping words with similar meanings that differ in writing form.

2.4.5 Tokenization

Tokenization is the process of breaking a sentence or text into smaller units, called words or tokens [19]. This stage aims to separate each word in a text so that it can be analyzed separately [19]. The tokenization process is very important in text analysis because it allows algorithms to recognize each word individually and identify patterns or meanings in the data. Data that has gone through the data cleaning stage will be displayed in Table 2

Table 2. Data That Has Gone Through The Text Processing Stage

Full text	Cleansing	Case Folding	Stopword Removal	Stemming	Tokenization
Indibiz meluncurkan 4 rangkaian solusi digital terbaru menjadikan total 11 island untuk mendukung digitalisasi usaha kecil dan menengah di Indonesia. Dengan solusi ini Indibiz mempermudah pelaku bisnis serta meningkatkan efektivitas dan efisiensi. #Indibiz #indibizjtd https://t.co/QAaxarscR7	indibiz meluncurkan 4 rangkaian solusi digital terbaru menjadikan total 11 island untuk mendukung digitalisasi usaha kecil dan menengah di indonesia dengan solusi ini indibiz mempermudah pelaku bisnis serta meningkatkan efektivitas dan efisiensi indibiz indibizjtd	indibiz meluncurkan 4 rangkaian solusi digital terbaru menjadikan total 11 island untuk mendukung digitalisasi usaha kecil dan menengah di indonesia dengan solusi ini indibiz mempermudah pelaku bisnis serta meningkatkan efektivitas dan efisiensi indibiz indibizjtd	indibiz meluncurkan 4 rangkaian solusi digital terbaru menjadikan total 11 island mendukung digitalisasi usaha menengah indonesia solusi indibiz mempermudah pelaku bisnis meningkatkan efektivitas efisiensi indibiz indibizjtd	indibiz luncur 4 rangkai solusi digital baru jadi total 11 island dukung digitalisasi usaha tengah indonesia solusi indibiz mudah laku bisnis tingkat efektivitas efisiensi	['indibiz', 'luncur', '4', 'rangkai', 'solusi', 'digital', 'baru', 'jadi', 'total', '11', 'island', 'dukung', 'digitalisasi', 'usaha', 'tengah', 'indonesia', 'solusi', 'indibiz', 'mudah', 'laku', 'bisnis', 'tingkat', 'efektivitas', 'efisiensi'
@Indibiz_jtdiy Indibiz emang juara dalam mendukung digitalisasi usaha kecil menengah ya. Semoga semakin banyak yang terbantu!	indibizjtdiy indibiz emang juara dalam mendukung digitalisasi usaha kecil menengah ya semoga semakin banyak yang terbantu	indibizjtdiy indibiz emang juara dalam mendukung digitalisasi usaha kecil menengah ya semoga semakin banyak yang terbantu	indibizjtdiy indibiz emang juara mendukung digitalisasi usaha menengah ya semoga terbantu	indibizjtdiy indibiz emang juara dukung digitalisasi usaha tengah ya moga bantu	'indibiz', 'emang', 'juara', 'dukung', 'digitalisasi', 'usaha', 'tengah', 'ya', 'moga', 'bantu'

2.5 Split Data Into Data Train and Data Test

The division of data into train and test data is a step in the data processing process where the available data is divided into two main parts. Training data is used to train the model, while test data is used to test the performance of the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

trained model [20]. The benefit of this sharing is to ensure that the developed model can be objectively evaluated using data that was never seen before during the training process, thus providing a more accurate picture of the model's performance in the real world [20].

2.5 TF-IDF Word Weighting

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical technique used to measure how significant a word is within a document. Unlike one-hot encoding, this method assigns a specific value to each word, calculated by multiplying the term frequency with the inverse document frequency [21]. The function of TF-IDF is to identify features as well as the frequency of occurrence of each feature in the document, so it can be used in text analysis and data clustering [21]. TF-IDF formula is explained in formula (1)

$$TF - IDF(t,d) = TF(t,d) \times IDF(t,d) \quad (1)$$

$TF(t,d)$, is the term frequency of term t in the document d, as calculated in formula (2)

$$TF(t,d) = \frac{f_{t,d}}{\sum_{tf \in d} f_{tf,d}} \quad (2)$$

$IDF(t,d)$, is the inverse document of term t, as calculated in formula (3)

$$IDF(t,d) = \log(\frac{N}{|\{d \in D : t \in D\}|}) \quad (3)$$

Here, $f_{t,d}$ represent the frequency of term t in document d, N represent the total of number of document in the corpus, and $|\{d \in D: t \in D\}|$ indicates the count of document that include in term t.

2.4 SVM and KNN Algorithm Classification

The algorithms that will be used to classify sentiment are SVM and KNN. KNN is used as a comparison of SVM results.

a. SVM
Support Vector Machine (SVM) is a method used to make predictions in various cases, both classification and regression. SVM basically functions as a linear classifier. However, this method has been further developed to address non-linear problems by utilizing the concept of kernels that map data to a high-dimensional space. In this space, SVM attempts to find a hyperplane that maximizes the margin or separation distance between data groups [22].
The kernel function in SVM is used to map the initial dimension (lower dimension) of the data set to a new dimension (relatively higher dimension). This allows SVMs to work on non-linear data by finding a hyperplane that separates the classes of data in a higher feature space [22].

b. KNN
K-Nearest Neighbor (KNN) is one of the classification methods used in data mining to classify new data based on pre-existing data. The KNN process begins by calculating the distance between new data and existing data using the Euclidean distance formula. After the distance is calculated, the next step is to sort the distance values from smallest to largest. Next, the algorithm will select a number of K closest data to determine the classification. The resulting category or class is the class that appears most often among the K closest data.

2.4 Evaluation

In this research, the system's performance was evaluated by measuring accuracy and precision. Accuracy reflects the proportion of correctly classified instances. The accuracy score is calculated based on True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values derived from the confusion matrix. The confusion matrix is a tool used in machine learning to evaluate and visualize a model's performance by displaying the distribution of predictions. It provides essential data to calculate accuracy-related metrics and identify errors in the model's predictions. [23].

Precision is a metric used to assess how accurate the positive predictions made by the model are in classifying the data. Recall is a metric used to measure the extent to which a model is able to detect all corresponding or relevant positive data in a dataset. F1-score is a metric used to assess the balance between the precision and recall values of a classification model, thus providing an overall picture of the model's performance in handling positive data. Accuracy, precision, recall, and f1-score are defined in the following equations

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - Score = \frac{2(precison \times recall)}{precision+recall} \quad (7)$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

3. RESULT AND DISCUSSION

The results of this study are divided into 3 scenarios. The first scenario is the baseline of the SVM and KNN models. The second scenario uses hyperparameter tuning for SVM and KNN models. The third scenario compares the SVM and KNN models.

3.1 Baseline

In the baseline scenario, SVM and KNN models are used to classify sentiment into two categories, namely Negative and Positive. The training data is processed by utilizing the TF-IDF Vectorizer which is limited to 5000 features, then the SVM and KNN models are trained using the data that has gone through the vectorization process. Figure 3 shows the accuracy, precision, recall, and F1-Score of the SVM baseline model.

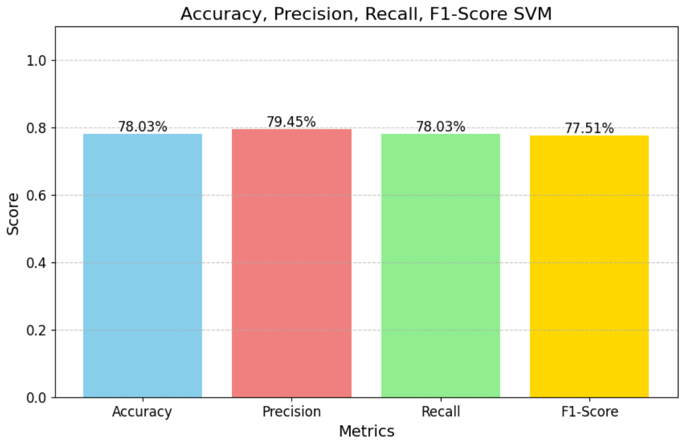

Figure 3. Classification Report SVM Model

Based on Figure 3, the SVM model achieved an accuracy of 78.03%, with a precision of 79.45% and a recall of 78.03%, indicating the model's ability to make fairly accurate predictions. The F1-Score of 77.51% reflects the balance between precision and recall, indicating a stable performance in avoiding prediction errors. Although these results are quite good for sentiment classification, there is still room for improvement, especially in reducing classification errors.

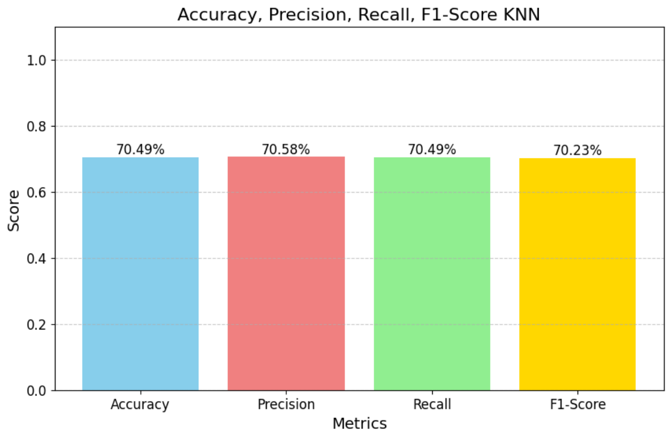

Figure 4. Classification Report KNN Model

Figure 4 shows the evaluation of the KNN model with TF-IDF, resulting in an accuracy of 70.49%, which means 70.49% of the model's predictions are correct. Precision and recall were recorded at 70.58% and 70.49% respectively, indicating the model's ability to recognize positive data well. The F1-Score of 70.23% reflects the balance between precision and recall.

This evaluation shows that the KNN model performs adequately in sentiment classification, although its accuracy is still lower than that of the SVM model under the same conditions. The KNN model requires additional optimization to improve the prediction accuracy and reduce the error rate.

3.2 Hyperparameter Tuning

In the Hyperparameter Tuning scenario, GridSearchCV is used to find the best combination of parameters for the SVM model. The tuned parameters include C as regularization, gamma as kernel coefficient, and various kernel types such as 'rbf', 'poly', 'sigmoid', and 'linear'. This process uses 5-fold cross-validation and evaluates the model based on

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

accuracy. The result is that the best kernel to use is the sigmoid kernel, with gamma 0.1 and C 10. Figure 5 shows the evaluation of the SVM model that has gone through the hyperparameter tuning process.

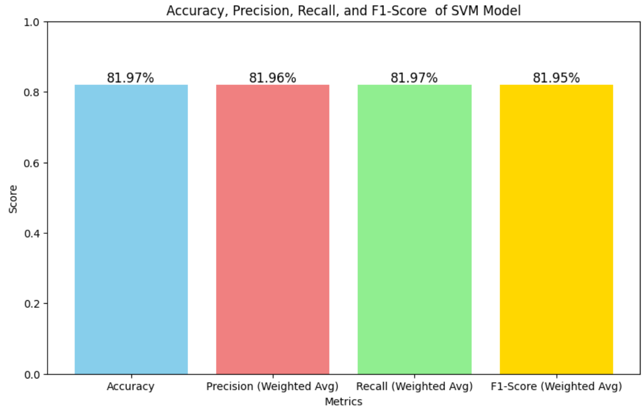

Figure 5. Classification Report of SVM Model with Hyperparameter Tuning

Support Vector Machine (SVM) model with sigmoid kernel, gamma 0.1, and C 10 performed well in sentiment classification. The model produced an accuracy of 81.97%, with a precision weighted average of 81.96%, and a recall of 81.97%, indicating accurate prediction and good recognition of positive samples. In addition, the model recorded a weighted average F1-score of 81.95%, reflecting a balance between precision and recall. Overall, the SVM model showed stable performance, although further experiments with different parameters can be conducted to improve performance.

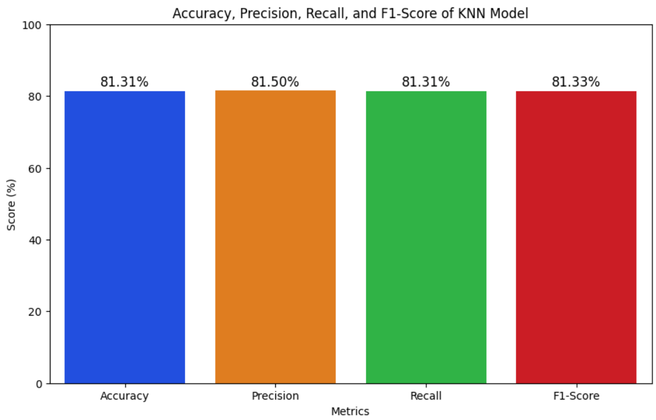

Figure 6. Classification Report of KNN Model with Hyperparameter Tuning

Figure 6 shows the KNN classification results that have been optimized with feature selection using SelectKBest, and hyperparameter tuning using GridSearchCV. Feature selection aims to reduce the dimensionality of the data by selecting features that are most relevant to the target label. GridSearchCV performs 5-fold cross-validation to evaluate each parameter combination and select the combination with the best accuracy.

Figure 6 shows the evaluation of the optimized K-Nearest Neighbors (KNN) model with manhattan metric, n_neighbors 11, and uniform weight. The model achieved 81.31% accuracy, meaning 81.31% of the predictions on the test data were correct. The model also recorded a precision of 81.50%, indicating that 81.50% of the positive predictions were correct, as well as a recall of 81.31%, meaning the model recognized 81.31% of the actual samples. The F1-score of 81.33% reflects the balance between precision and recall. Hyperparameter tuning is proven to increase model accuracy for both SVM and KNN models. SVM model accuracy increased by 3.94% and KNN model accuracy increased by 10.36% when compared to the baseline.

3.3 Model Comparations

The third scenario compares the SVM and KNN models. Table 3 display the comparative result.

Table 3. Comparations of SVM and KNN Model

Model	Accuracy (%)	Precision (%)
Support Vector Machine (baseline)	78.03%	79.45%
K Neirest Neighbor (baseline)	70.49%	70.58%
Support Vector Machine (hyperparameter tuning)	81.97%	81.96%
K Neirest Neighbor (hyperparameter tuning)	81.31%	81.50%

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

In general, SVM that has been optimized through hyperparameter tuning shows superior performance compared to KNN, especially in terms of stability and accuracy in recognizing data. Although both models managed to achieve accuracy above 81% after optimization, SVM proved to be more effective in sentiment classification on the dataset used. Therefore, it can be concluded that hyperparameter tuning plays an important role in improving model performance, with SVM providing better results than KNN.

4. CONCLUSION

Based on the results and discussion, sentiment classification using SVM and KNN is proven to be able to classify sentiments regarding the digitalization of Small and Medium Enterpises. The baseline scenario and hyperparameter tuning on SVM and KNN models, it can be concluded that hyperparameter tuning significantly improves the performance of both models in sentiment classification. In the baseline scenario, SVM has an accuracy of 78.03%, higher than KNN which reaches 70.49%, showing that SVM is superior in the initial configuration. After tuning using GridSearchCV, the performance of both models improved. SVM with sigmoid kernel, gamma 0.1, and C 10 achieved 81.97% accuracy, while KNN with manhattan metric, n_neighbors 11, and weight uniform achieved 81.31% accuracy. In the evaluation using precision, recall, and F1-score, SVM showed a more stable performance with values in the range of 81.95% to 81.97%, while KNN recorded values in the range of 81.31% to 81.50%, confirming that SVM is more consistent than KNN. Overall, SVM with hyperparameter tuning performed better than KNN, especially in terms of stability and the ability to recognize data more accurately. However, both models performed well with accuracy above 81% after optimization. Thus, it can be concluded that the hyperparameter tuning process is essential to improve model performance, and in the context of the dataset used, the SVM model is superior in classifying sentiment than the KNN model. For future research, it is recommended to explore other machine learning algorithms such as Random Forest, Gradient Boosting (XGBoost, LightGBM), or Logistic Regression.

REFERENCES

[1] O. Juwita, A. Firdonsyah, M. Ali, A. P. Widodo, and R. R. Isnanto, "Studi Literatur Platform Digital Sebagai Sarana Pembangunan Ekosistem Dalam Mengembangkan UMKM," *INFORMAL: Informatics Journal*, vol. 7, no. 1, p. 59, Jun. 2022, doi: 10.19184/isj.v7i1.31547.

[2] P. Chairun, N. Fakultas, E. Dan Bisnis, and J. Manajemen, "PELUANG DAN TANTANGAN: KONSEP DIGITALISASI SMART CITY EKONOMI E-COMMERCE DI INDONESIA," 2019. [Online]. Available: https://economy.okezone.com/read/201

[3] V. Kurniawan, M. Faisal, R. Ansori, R. Yunus Pangaribuan, and I. Maritim Prasetya Mandiri, "PENGARUH UMKM (USAHA MIKRO KECIL MENEGAH) TERHADAP PENINGKATAN PEREKONOMIAN INDONESIA TAHUN 2024", doi: 10.8734/mnmae.v1i2.359.

[4] D. Rizqi Khusnul Khotimah, "Economic Growth and E-Commerce: Potensial for Digitizing SMSEs in East Java," *East Java Economic Journal*, vol. 5, no. 2, pp. 183–203, Sep. 2021, doi: 10.53572/ejavec.v5i2.69.

[5] E. Shintya Pratiwi, "ANALISIS SENTIMEN PADA DATA TWITTER KPK RI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," 2023. Accessed: Jan. 10, 2025. [Online]. Available: Teknologipintar.org

[6] A. Cahaya Puspitasari, "PENINGKATAN KINERJA ANALISIS SENTIMEN DENGAN TEKNIK FEATURE ENGINEERING PADA MACHINE LEARNING." Accessed: Jan. 10, 2025. [Online]. Available: Puspitasari

[7] A. Permata Informatika, "ANALISIS SENTIMEN MEDIA SOSIAL: MENGURAI OPINI PUBLIK DENGAN DATA," 2024. Accessed: Jan. 10, 2025. [Online]. Available: Teknologipintar.org

[8] N. Naw, "Twitter Sentiment Analysis Using Support Vector Machine and K-NN Classifiers," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 8, no. 10, Oct. 2018, doi: 10.29322/IJSRP.8.10.2018.p8252.

[9] Dhina Nur Fitriana and Yuliant Sibaroni, "Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 5, pp. 846–853, Oct. 2020, doi: 10.29207/resti.v4i5.2231.

[10] A. Deviyanto and M. D. R. Wahyudi, "PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 3, no. 1, p. 1, Dec. 2018, doi: 10.14421/jiska.2018.31-01.

[11] P. Verma, T. Bhardwaj, A. Bhatia, and M. Mursleen, "Sentiment Analysis 'Using SVM, KNN and SVM with PCA,'" in *Artificial Intelligence in Cyber Security: Theories and Applications*, vol. 240, Springer, Cham, 2023. doi: 10.1007/978-3-031-28581-3_5.

[12] V. A. Retnowati and V. Purwayoga, "Analisis Sentimen Twitter Tentang Isu Resesi 2023: Studi Komparatif Pendekatan Machine Learning Twitter Sentiment Analysis of Recession 2023: A Comparative Study of Machine Learning Approaches," vol. 11, p. 1, 2024, doi: 10.25124/jrsi.v11i01.612.

[13] M. D. H. Jasy, S. Al Hasan, M. I. K. Sagor, A. Noman, and J. M. Ji, "A Performance Evaluation of Sentiment Classification Applying SVM, KNN, and Naive Bayes," in *2021 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, IEEE, Dec. 2021, pp. 56–60. doi: 10.1109/CoNTESA52813.2021.9657115.

[14] E. K. Nararto, S. D. Budiwati, and S. K. Sari, "Opinion Classification on MSME Social Media Comments using Support Vector Machine and Random Forest Models," in *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, IEEE, Jul. 2024, pp. 165–170. doi: 10.1109/ICSINTESA62455.2024.10748163.

[15] M. D. H. Jasy, S. Al Hasan, M. I. K. Sagor, A. Noman, and J. M. Ji, "A Performance Evaluation of Sentiment Classification Applying SVM, KNN, and Naive Bayes," in *2021 International Conference on Computing, Networking,*

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2815−2823
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6723

Telecommunications & Engineering Sciences Applications (CoNTESA), IEEE, Dec. 2021, pp. 56–60. doi: 10.1109/CoNTESA52813.2021.9657115.
[16] D. Mustikasari, I. Widaningrum, R. Arifin, W. Henggal, and E. Putri, "Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents," 2021. [Online]. Available: http://tiny.cc/rootwords.
[17] B. Wijaya Rauf, "Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes," 2023. [Online]. Available: https://t.co/fSdh2dCADm
[18] N. Istiqomah and F. Novika, "Sentiment Analysis Penyedia layanan Asuransi dari Media Sosial Twitter," *Jurnal Tekno Kompak*, vol. 18, no. 1, p. 77, Feb. 2024, doi: 10.33365/jtk.v18i1.3465.
[19] R. Sanjaya, E. Tohidi, E. Wahyudi, and K. Kaslani, "ANALISIS SENTIMEN TERHADAP BERHENTINYA TIKTOKSHOP PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 507–514, Feb. 2024, doi: 10.36040/jati.v8i1.8443.
[20] F. Syahro and N. Fitriani, "PERBANDINGAN PERFORMA MODEL MACHINE LEARNING SUPPORT VECTOR MACHINE, NEURAL NETWORK, DAN K-NEAREST NEIGHBORS DALAM PREDIKSI HARGA SAHAM," *Jar's*, vol. 2, no. 1, p. 13, [Online]. Available: https://www.ejournalwiraraja.com/index.php/JARS
[21] J. Ravi and S. Kulkarni, "Text embedding techniques for efficient clustering of twitter data," *Evol Intell*, vol. 16, no. 5, pp. 1667–1677, Oct. 2023, doi: 10.1007/s12065-023-00825-3.
[22] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG," vol. 3, no. 4, pp. 811–820, 2014, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian