# Comparing the Performance of Data Mining Algorithms in Predicting Sentiments on Twitter

Rusydi Umar[1], Sunardi[2], Muhammad Nur Ardhiansyah[3*]
[1,3]Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta
[2]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta
[1]rusydi@mti.uad.ac.id, [2]sunardi@mti.uad.ac.id, [3]muhammad1808048033@webmail.uad.ac.id

*Abstract*

*On the social networking site Twitter, users can post tweets, videos, and images. It can, however, also be disruptive and difficult. In order to categorize material and improve searchability, hashtags are crucial. This study focuses on examining the opinions of Twitter users who participate in trending topics. The algorithms K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are employed for sentiment analysis. The dataset comprises of tweet information on popular subjects that was collected using the Twitter API and saved in Excel format. SVM and K-NN are used for data preparation, weighting, and sentiment analysis. With 105 data points, the study provides insights into user sentiment. SVM identified 99% of positive and 1% of negative replies with accuracy of 80%. KNN successfully identified 90% of positive and 10% of negative responses, with an accuracy rate of 71.4%. According to the results, SVM performs better when analyzing the sentiment of hashtag users on Twitter.*

*Keywords: sentiment analysis; twitter; SVM; K-NN*

## 1. Introduction

Technology is advancing so quickly, as shown by the ease with which daily operations may now be carried out. The Internet plays a role in facilitating human effort in information gathering; any information, including that pertaining to business, politics, and entertainment, can be accessible and known via the Internet. The growth of creative industries in Indonesia shows that the people's economy is becoming more [1]. Being a YouTuber and a celebrity are both creative careers that call for innovation and originality in order to provide engaging, current, and top-notch content that viewers find appealing. Due to escalating competition and the requirement to keep up with trends in order to avoid falling behind, creativity is required.

Twitter which is connected to the internet can do the sharing [2], [3]. Twiter has grown in the last decade, and a great deal of information is provided by tweets and instant messages [4]. Twitter application users have now reached 310 million and in the world reached 2.31 trillion [5], with the possibility of increasing every year. Some many users making twitter a means for gathering text-based data and increasing tweets that are inappropriate for public consumption. The growth of social media has made it so easy to obtain information

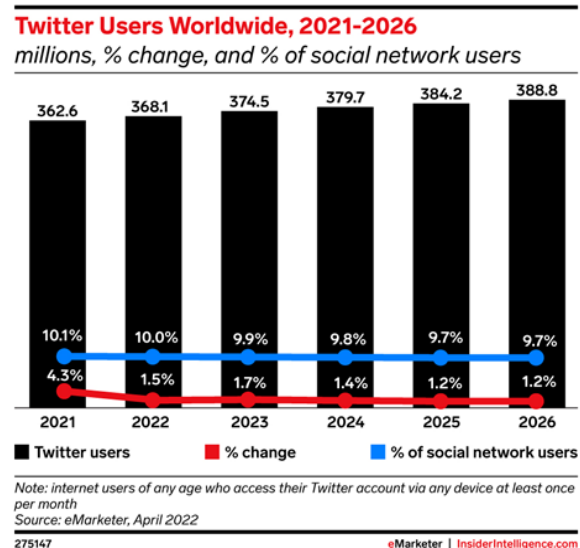[6] - [7]. Among them is information about trends that are developing in society.


Figure 1. User's Twitter prediction

Figure 1 is data sourced from eMarketer on the website insideintelligence.com shows the growth of Twitter users from 2021 to 2026, twitter usage is increasing every year.

There is the term hashtag on Twitter which functions to group content and make it easier to search, the most hashtags will become trending topics. Social media is not only for making friends or sharing but can be used to carry out trade promos or buying and selling to political party promos [8]. The text data that develops on Twitter is generally the result of presenting something that is currently developing to the community and is a reflection of the current trend in society [9].

Sentiment analysis is a field of science that analyzes the opinions, sentiments, evaluations, judgments, attitudes and emotions of the public towards entities such as products, services, organizations, individuals, problems, events, topics, and their attributes [10]. Sentiment analysis can be in the form of opinions to express or reveal sentiment values. Contextual text mining extracts subjective information from data sources. The benefits of doing Twitter hashtag sentiment analysis can be used to assist content creators in getting ideas to become content, so that data obtained from Twitter that is developing can be retrieved and processed to obtain information.

Previous research has been carried out and there are similarities in objects such as Zuhdi et al [11] utilizing the existing API on Twitter to obtain data from the two 2019 President Nomination, resulting in that the community has 300 positive opinions, 68 negative, 132 neutral for candidate pair 1 and 312 positive, negative 91, neutral 97 for candidate pair 2. Ernawati & Wati [12] conducted research on the Application of the K-Nearest Neighbor Algorithm in Sentiment Analysis of Travel Agent Reviews. Data retrieval used travel agent review data from the site https://www.trustpilot.com/categories/travel_holidays, resulting in a k-NN classifier of 100 positive and negative reviews where the resulting accuracy value reached 87% with an AUC value of 0.916.

Ipmawati et al [13] do a Comparison of Text Mining Classification Techniques on Sentiment Analysis, combining the three methods of Naive Bayes, K-NN, and SVM aims to determine good performance in terms of accuracy. SVM comparison results are superior to Naïve Bayes and k-NN. Sari and Wibowo [14] conduct a sentiment analysis of JD.id users on Twitter social media. The method used is NBC with TF-IDF weighting based on emoticon conversion. The results of TF-IDF weighting and emoticon conversion have an accuracy of 98%.

Vatambeti et al [15] conducted research on Twitter sentiment analysis on online food services based on elephant herd optimization with a hybrid deep-learning technique. Aiming to help app-based food delivery businesses, the research transforms social broadcasting data into production data for competitive decision-makers about social broadcasting. This study compared

Zomato, Uber Eats, and Swiggy. Combining the CNN and Bi-directional Long Short-Term Memory (Bi-LSTM) models shows that Zomato gets the most positive feedback (29%), followed by Swiggy (26%), and Uber Eats (25%).

Umar et al [16] perform a comparison of the K-NN and SVM methods for image matching. The stage begins with data processing, feature extraction and finally the classification stage. A total of 10 images were used consisting of 9 manipulated images and the remaining original images. The test uses the WEKA application by applying the 10-fold cross validation. The result of the comparison is K-NN works better than SVM for image matching. Abdar et al [17] Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. The purpose of this study is comparison of different data mining algorithm on prediction of heart diseases compared data mining techniques to predict the risk of heart diseases. After feature analysis, models by five algorithms including C5.0, Neural Network, SVM, K-NN, and Logistic Regression, developed and validated. C5.0 Decision tree has been able to build a model with greatest accuracy 93.02%, KNN, SVM, Neural Network have been 88.37%, 86.05% and 80.23% respectively. Results of decision tree can be simply interpretable and applicable; their rules can be understood easily by different clinical practitioner.

This study compared the two methods, which are SVM, a technique for predicting classification that is quite popular. The concept of SVM classification is to find the best hyperplane [16], which functions as a separator for two data classes using selected data points [19]. The K-NN method classifies objects based on learning data that has the closest distance to the object as well as guided data with a limited sample size [20]. This method aims to classify newly entered objects based on training sample data. In this study, an analysis of hashtags on Twitter was carried out.

## 2. Research Methods

The research object to be carried out is the use of hashtags which are trending topics on Twitter. In conducting this research, tools and materials are needed to assist the research process as shown in Table 1.

Table 1. Tools and Materials

| No | Tools | Description | Information |
|---|---|---|---|
| 1 | Laptop | Compaq Presario CQ40 | Hardware |
| 2 | OS | Windows 7 Ultimate 64 Bit | Software |
| 3 | HD Drive | Hard Disk Capacity 230 GB | Hardware |
| 4 | Web Browser | Google Chrome, Mozilla Firefox, Internet Explorer | Software |
| 5 | Web Server | Receive requests from web browsers using XAMPP | Software |
| 6 | Database | Database Management System (DBMS) PHPMyAdmin | Software |
| 7 | Text Editor | VS Code, Sublime Text, Atom | Software |

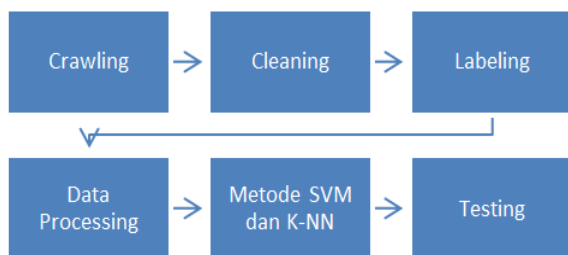The stages in the research can be seen in Figure 2.



Figure 2. Flow of Research Stages

Crawling or data collection uses the API provided by Twitter. In this study retrieved hashtag data that was trending topic on Twitter. Retrieval of data by utilizing the API key by creating an account with the Twitter API developer, as shown in Figure 3.
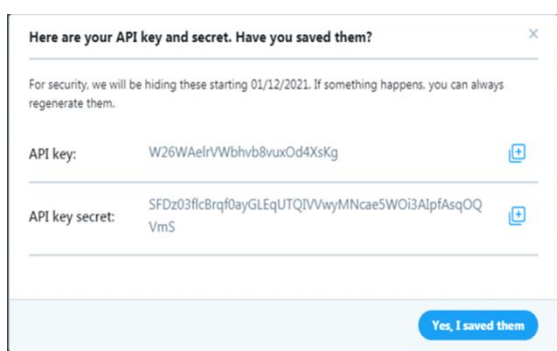


Figure 3. API Key Twitter

Cleaning is the process of removing empty row data. This process is carried out so that there are no empty rows that are processed thereby saving time when processed by the system.

Labeling is the labeling of data that has been cleaned. At this stage the data is labeled positive and negative with the decision-making method considering the positive connotations of words for positive labels and negative connotations for negative labels.

Data Processing is the stage to get the desired data from datasets that have been processed before [21]. The data processing stage is very important because to find out the sentiment value of a Twitter user's tweet, the steps taken start with tokenizing to break down sentences into words and remove the delimiters that make them up [22][23]. Then the stopword removal process is carried out by removing vocabulary that is not a unique word by removing pre, last, conjunctions and so on. After that, stemming is done to decompose affixed word forms into basic words and non-standard words into standard words.

The SVM method is to carry out sentiment analysis using the Support Vector Machine method for the training and testing process [24] and is used to facilitate classification tasks with minimized computational effort whether the data includes positive or negative sentiments [25]. SVM has a goal to build the best hyperplane that can divide a given sample set into two distinct classes [26].

The K-NN method after the training data has been transformed into a document vector is used to classify polarization data [27], one of the well-known clustering techniques for grouping large data sets [28]. The training data is in the form of numbers with labels to classify new objects based on the sample

Testing In this process is done to determine the value of accuracy using the Cross Validation method [29]. K-fold cross validation is used to estimate prediction errors in evaluating [30]. In this study will do 5-fold cross validation.

## 3. Results and Discussions

This study retrieves data by crawling using the Twitter API, retrieving tweet data based on hashtags that are currently trending topics. Data crawling can be done continuously by users when using web-based applications that are connected to the internet by means of updates.

From the data that has been obtained, it can be seen in Table 2. Then data selection will be carried out, namely by selecting the column that will be used for training. The column used is the "Tweet" column.

Table 2. Crawling Results

| No | Date | Hashtag | Tweet |
|---|---|---|---|
| 1 | 29 Dec 2022 | #BeliPulsaDiGrab | Kemarin pagi minta diantar pagi2 liputan video diluar Kota yang bawa kru dan alat byk beud. Sangat membantu Thanks #BeliPulsaDiGrab |
| 2 | 29 Dec 2022 | #BeliPulsaDiGrab | Walau deket karna sy mager jd pesen grab aja wkwk, pengemudinya juga pake masker n nyaranin syaa pake helm sendiri, sipp lha #BeliPulsaDiGrab |
| 3 | 29 Dec 2022 | #BeliPulsaDiGrab | beli makan siang buat calon mertua #Grabfood |
| 4 | 29 Dec 2022 | #BeliPulsaDiGrab | Layanan Grab yang terakhir dipakai adalah Grab Bike,jadi pastinya naik Grab Bike itu praktis banget. Tetap memperhatikan protokol kesehatan, pakai masker dan bawa helm pribadi #BeliPulsaDiGrab |
| 5 | 29 Dec 2022 | #BeliPulsaDiGrab | Grabike kin aja gaesss, anjaayyy #BeliPulsaDiGrab pic.twitter.com/pok28TpzNK |
| 6 | 29 Dec 2022 | #BeliPulsaDiGrab | GrabFood aja dah untuk menu favorit andalan Nasi Telor Pontianak #BeliPulsaDiGrab pic.twitter.com/EI3rJxU6Ys |
| 7 | 29 Dec 2022 | #BeliPulsaDiGrab | Nih min, gue pake dalam sehari udah dua kali, btw yang pertama dapat promo dipotong 25% makasih looo |

| No | Date | Hashtag | Tweet |
|---|---|---|---|
| 8 | 29 Dec 2022 | #BeliPulsaDiGrab | #BeliPulsaDiGrab pic.twitter.com/wPLyHbG3QA Sangat sangat terbantu banget sama @GrabID Terimakasih buat semua orang yang berada di balik suksesnya apk grab #BeliPulsaDiGrab pic.twitter.com/yZhcpKozgU |
| 9 | 29 Dec 2022 | #BeliPulsaDiGrab | #BeliPulsaDiGrab Aku td sore habis pake GrabExpress Sameday-Bike untuk ngesend hadiah ke temen kuu karna lg psbb enggak bisa ketemu2. Ngebantuuu bgt #BeliPulsaDiGrab pic.twitter.com/hkdH2zKaHG |
| 10 | 29 Dec 2022 | #BeliPulsaDiGrab | #BeliPulsaDiGrab Terakhir pake grabbike, ada diskon lagi hahaha pic.twitter.com/RrUNyiX6H5 |
| … | | | |
| 105 | 29 Dec 2022 | #BeliPulsaDiGrab | BeliPulsaDiGrab So pasti,yg pasti" aku ankny jajan makanan n sering bm lalu order makanan d grabfood krn sering promo n tmpt jajan ny enak" krn mnjlni kehidupan butuh tenaga apa lg yg overthingking mending makan lah biar kuat # pic.twitter.com/QxzXpp7wSK |

Table 2 is the trending hashtag data on October 29, 2022, namely 105 data with the hashtag "#BeliPulsaDiGrab". From the column that has been selected then the cleaning process is carried out by deleting empty rows. The results of the cleaning process can be seen in Table 3.

Tabel 3. Cleaning Result

| Tweet |
|---|
| Kemarin pagi minta diantar pagi2 liputan video diluar Kota yang bawa kru dan alat byk beud. Sangat membantu Thanks |
| Walau deket karna sy mager jd pesen grab aja wkwk, pengemudinya juga pake masker n nyaranin syaa pake helm sendiri, sipp lha |
| beli makan siang buat calon mertua |
| Layanan Grab yang terakhir dipakai adalah Grab Bike,jadi pastinya naik Grab Bike itu praktis banget. Tetap memperhatikan protokol kesehatan, pakai masker dan bawa helm pribadi |
| Grabike kin aja gaesss, anjaayyy |

After obtaining clean data, 5 data are taken from the initial data to be labeled positive (1) and negative (0) to be used as a dataset. The results of the dataset that has been labeled will then be stored in the system to be used as a reference for the system for the sentiment analysis process. The dataset can be seen in Table 4.

At this stage preprocessing is carried out so that text data in the form of sentences can be processed by the system.

Table 4. Sample Dataset

| Tweet | Label |
|---|---|
| Kemarin pagi minta diantar pagi2 liputan video diluar Kota yang bawa kru dan alat byk beud. Sangat membantu Thanks | 1 |
| Walau deket karna sy mager jd pesen grab aja wkwk, pengemudinya juga pake masker n nyaranin syaa pake helm sendiri, sipp lha | 1 |
| beli makan siang buat calon mertua | 1 |
| Layanan Grab yang terakhir dipakai adalah Grab Bike,jadi pastinya naik Grab Bike itu praktis banget. Tetap memperhatikan protokol kesehatan, pakai masker dan bawa helm pribadi | 1 |
| Grabike kin aja gaesss, anjaayyy | 1 |

The tokenizing process is breaking sentences into several words based on the words that compose them, as shown in Table 5.

Table 5. Tokenizing Process Samples

| Tweet |
|---|
| "Kemarin" "pagi" "minta" "diantar" "pagi2" "liputan" "video" "diluar" "Kota" "yang" "bawa" "kru" "dan" "alat" "byk" "beud" "Sangat" "membantu" "Thanks" |
| "Walau" "deket" "karna" "sy" "mager" "jd" "pesen" "grab" "aja" "wkwk" "pengemudinya" "juga" "pake" "masker" "nyaranin" "syaa" "pake" "helm" "sendiri" "sipp" "lha" |
| "beli" "makan" "siang" "buat" "calon" "mertua" |
| "Layanan" "Grab" "yang" "terakhir" "dipakai" "adalah" "Grab" "Bike" "jadi" "pastinya" "naik" "Grab" "Bike" "itu" "praktis" "banget" "Tetap" "memperhatikan" "protokol" "kesehatan" "pakai" "masker" "dan" "bawa" "helm" "pribadi" |
| "Grabike" "kin" "aja" "gaesss" "anjaayyy" |

The next stage is to carry out the stopword removal process by removing common and less important words such as the conjunctions 'and', 'or', 'anyway' and so on. For more details can be seen in Table 6.

Table 6. Stopword Removal Process

| Initial Sentence | Stopword Removal |
|---|---|
| "Kemarin" "pagi" "minta" "diantar" "pagi2" "liputan" "video" "diluar" "Kota" "yang" "bawa" "kru" "dan" "alat" "byk" "beud" "Sangat" "membantu" "Thanks" | "Kemarin" "pagi" "minta" "diantar" "pagi" "liputan" "video" "diluar" "Kota" "yang" "bawa" "kru" "alat" "byk" "beud" "Sangat" "membantu" "Thanks" |
| "Walau" "deket" "karna" "sy" "mager" "jd" "pesen" "grab" "aja" "wkwk" "pengemudinya" "juga" "pake" "masker" "nyaranin" "syaa" "pake" "helm" "sendiri" "sipp" "lha" | "Walau" "deket" "karna" "sy" "mager" "jd" "pesen" "grab" "aja" "wkwk" "pengemudinya" "juga" "pake" "masker" "nyaranin" "syaa" "pake" "helm" "sendiri" "sipp" "lha" |
| "beli" "makan" "siang" "buat" "calon" "mertua" | "beli" "makan" "siang" "buat" "calon" "mertua" |
| "Layanan" "Grab" "yang" "terakhir" "dipakai" "adalah" "Grab" "Bike" "jadi" "pastinya" "naik" "Grab" "Bike" "itu" "praktis" "banget" "Tetap" "memperhatikan" "protokol" "kesehatan" "pakai" "masker" "dan" "bawa" "helm" "pribadi" | "Layanan" "Grab" "yang" "terakhir" "dipakai" "adalah" "Grab" "Bike" "jadi" "pastinya" "naik" "Grab" "Bike" "itu" "praktis" "banget" "Tetap" "memperhatikan" "protokol" "kesehatan" "pakai" "masker" "bawa" "helm" "pribadi" |

| Initial Sentence | Stopword Removal |
|---|---|
| "Grabike" "kin" "aja" "gaesss" "anjaayyy" | "Grabike" "kin" "aja" "gaesss" "anjaayyy" |

The next step is to carry out the stemming process, namely changing all words into basic words and removing affixes. As can be seen in Table 7.

Table 7. Stemming Process

| Initial Sentence | After Stemming Process |
|---|---|
| "Kemarin" "pagi" "minta" "diantar" "pagi" "liputan" "video" "diluar" "Kota" "yang" "bawa" "kru" "alat" "byk" "beud" "Sangat" "membantu" "Thanks" | "Kemarin" "pagi" "minta" "antar" "pagi" "liputan" "video" "luar" "Kota" "yang" "bawa" "kru" "alat" "banyak" "banget" "Sangat" "membantu" "thank" |
| "Walau" "deket" "karna" "sy" "mager" "jd" "pesen" "grab" "aja" "wkwk" "pengemudinya" "juga" "pake" "masker" "nyaranin" "syaa" "pake" "helm" "sendiri" "sipp" "lha" | "Walau" "deket" "karna" "saya" "mager" "jadi" "pesen" "grab" "aja" "wkwk" "pengemudinya" "juga" "pake" "masker" "nyaranin" "saya" "pake" "helm" "sendiri" "sipp" "lha" |
| "beli" "makan" "siang" "buat" "calon" "mertua" | "beli" "makan" "siang" "buat" "calon" "mertua" |
| "Layanan" "Grab" "yang" "terakhir" "dipakai" "adalah" "Grab" "Bike" "jadi" "pastinya" "naik" "Grab" "Bike" "itu" "praktis" "banget" "Tetap" "memperhatikan" "protokol" "kesehatan" "pakai" "masker" "bawa" "helm" "pribadi" | "Layanan" "Grab" "yang" "terakhir" "pakai" "adalah" "Grab" "Bike" "jadi" "pasti" "naik" "Grab" "Bike" "itu" "praktis" "banget" "Tetap" "perhatikan" "protokol" "kesehatan" "pakai" "masker" "bawa" "helm" "pribadi" |
| "Grabike" "kin" "aja" "gaesss" "anjaayyy" | "Grabike" "kin" "aja" "gaesss" "anjaayyy" |

The next step describes the appearance of the system that has been made along with the appearance of the application.
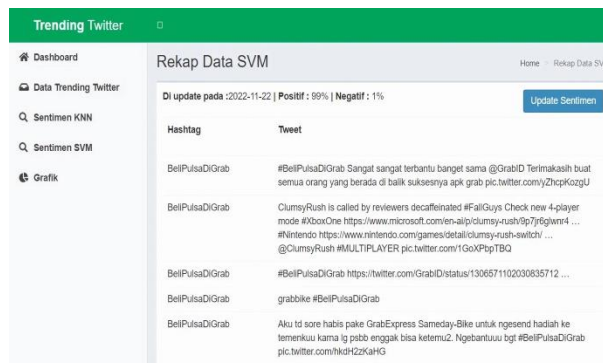


Figure 4. SVM Data Recap

Figure 4 displays the SVM data recap page. To perform a sentiment analysis, you can do it by selecting the Update sentiment button.

Program Code 1 SVM Data Recap

```
1    def prediksi_svm(tweet_svm, trend):
2    prepro=preprocessing(str(tweet_svm))
3    X= vectorizersvm.transform([prepro])
4    sentime = exsvm.predict(X)
5    senti = parseSentiment(sentime[0])
6    tanggal = date.today()
7    # print(prepro)
8    if tweet_svm == "" or tweet_svm == "nan":tweet_svm = "0"
9    elif senti == "":
10   return "0"
11   cursor = db.cursor()
12   sql = "INSERT INTO sentimen_svm (tanggal, hashtag, tweet, tweet_clean, sentimen) VALUES (%s, %s, %s, %s, %s)"val = (tanggal, trend, tweet_svm, prepro, senti)
13   cursor.execute(sql, val)
14   db.commit()
```

Program Code 1 is a display all of the trending topic data.

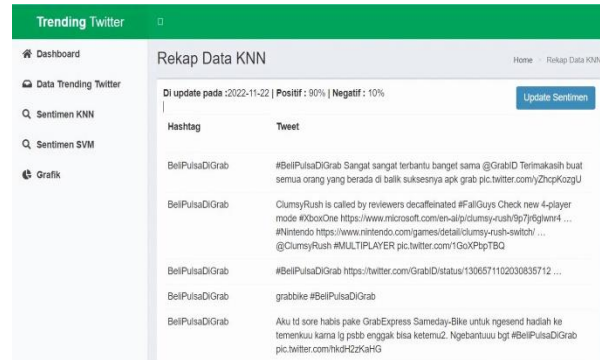The display of K-NN data recap on the application can be seen in Figure 5.



Figure 4. K-NN Data Recap

Figure 5 is a display of the K-NN data recap page.

Program Code 2 K-NN Data Recap

```
1    def prediksi_knn(tweet, trend):
2    preproces = preprocessing(str(tweet))
3    X = vectorizer.transform([preproces])
4    sentiment = exknn.predict(X)senti = parseSentiment(sentiment[0])
5    tanggal = date.today()
6    # print(X)
7    if tweet == "" or tweet == "nan": tweet = "0"
8    elif senti == "":
9    return "0"
10   cursor = db.cursor()
11   sql = "INSERT INTO sentimen_knn (tanggal, hashtag, tweet, tweet_clean, sentimen) VALUES (%s, %s, %s, %s, %s)"
12   val = (tanggal, trend, tweet, preproces, senti)
13   cursor.execute(sql, val)db.commit()
```

Program Code 3.2 is a display all of the trending topic data.

The cross validation testing process uses the Microsoft Excel application and the calculation results are not displayed on the system. The calculation is done with a value of k = 5 in the cross validation test.

Table 8. Iteration with k-cross validation

| Iteration | K-NN | | | | SVM | | | |
|---|---|---|---|---|---|---|---|---|
| | Tpo | Tne | Fpo | Fne | Tpo | Tne | Fpo | Fne |
| 1 | 8 | 2 | 3 | 8 | 9 | 1 | 2 | 9 |
| 2 | 7 | 3 | 3 | 8 | 8 | 2 | 3 | 8 |
| 3 | 7 | 3 | 5 | 6 | 8 | 2 | 3 | 8 |
| 4 | 9 | 2 | 3 | 7 | 9 | 1 | 2 | 9 |
| 5 | 7 | 2 | 4 | 8 | 8 | 2 | 3 | 8 |
| Subtotal | 38 | 12 | 18 | 37 | 42 | 8 | 13 | 42 |
| Total | 50 | | 55 | | 50 | | 55 | |

Table 8 is the result of a k-cross validation comparison iteration between K-NN and SVM. After that, the matrix results are obtained as shown in Table 9. The results of the k-fold cross validation matrix.

Table 9. The results of the k-fold cross validation matrix

| | | Predictions | | | | | |
|---|---|---|---|---|---|---|---|
| | | K-NN | | | SVM | | |
| | | P | N | Total | P | N | Total |
| Fact | Positive | 38 | 12 | 50 | 41 | 9 | 50 |
| | Negative | 18 | 37 | 55 | 15 | 40 | 55 |
| | Total | 56 | 49 | 105 | 56 | 49 | 105 |

So that the accuracy value of K-NN is obtained by Equation 1.

$$Accuracy = \frac{(TPP+TNN)}{(P+N)} \, x \, 100 \tag{1}$$

$$Accuracy = \frac{(38+37)}{(50+55)} x100$$

$$Accuracy = 71.4\%$$

While the SVM accuracy is obtained by Equation 2.

$$Accuracy = \frac{(TPP+TNN)}{(P+N)} \, x \, 100 \tag{2}$$

$$Accuracy = \frac{(42+42)}{(50+55)} x100$$

$$Accuracy = 80\%$$

## 4. Conclusion

Based on the research that has been carried out, it can be concluded that this research is a Twitter hashtag response analysis system to find out the sentiments of hashtag users on tweets that have been successfully developed so that it helps content creators improve the quality of their content and provide ideas continuously (realtime). Based on 105 research data, it was found that the K-NN method had an accuracy of 71.4% with positive responses of 90% and negative responses of 10%. While the SVM method has an accuracy value of 80% with positive responses of 99% and 1% negative. From these two results, it was found that the SVM method was better.

## References

[1] A.-K. Al-Khowarizmi, A. R. Lubis, M. Lubis, and R. F. Rahmat, "Information technology based smart farming model development in agriculture land," *IAES Int. J. Artif. Intell.*, vol. 11, no. 2, p. 564, 2022.

[2] R. A. K. N. Bintang, R. Umar, and U. Yudhana, "Perancangan perbandingan live forensics pada keamanan media sosial Instagram, Facebook dan Twitter di Windows 10," *Pros. SNST ke-9 Tahun 2018 Fak. Tek. Univ. Wahid Hasyim*, pp. 125–128, 2018.

[3] A. R. Royyan and E. B. Setiawan, "Feature Expansion Word2Vec for Sentiment Analysis of Public Policy in Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 78–84, 2022.

[4] S. Mann, J. Arora, M. Bhatia, R. Sharma, and R. Taragi, "Twitter sentiment analysis using enhanced bert," in *Intelligent Systems and Applications: Select Proceedings of ICISA 2022*, Springer, 2023, pp. 263–271.

[5] I. Zuhriyanto, A. Yudhana, and I. Riadi, "Perancangan Digital Forensik pada Aplikasi Twitter Menggunakan Metode Live Forensics," *Seminar Nasional Informatika 2008 (semnasIF 2008)*, vol. 2018, no. November. 2018.

[6] I. F. Rozi, E. N. Hamdana, M. Balya, and I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter ( Studi Kasus SAMSAT Kota Malang )," pp. 149–154, 2017.

[7] S. Aslan, S. Kızıloluk, and E. Sert, "TSA-CNN-AOA: Twitter sentiment analysis using CNN optimized via arithmetic optimization algorithm," *Neural Comput. Appl.*, pp. 1–18, 2023.

[8] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J. Maret*, vol. 2, p. 1, 2017.

[9] W. Setyobudi, A. Alwi, and I. P. Astuti, "Sentimen Analisis Twitter Terhadap Penyelenggaraan Gojek Traveloka Liga 1 Indonesia," *KOMPUTEK*, 2018, doi: 10.24269/jkt.v2i1.68.

[10] B. Liu, "Sentiment Analysis and Subjectivity in: Handbook of Natural Language Processing, Second Edition," *Handb. Nat. Lang. Process. Second Ed.*, 2010.

[11] A. M. Zuhdi, E. Utami, and S. Raharjo, "Analisis Ssentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN," *SSRN Electron. J.*, vol. 5, pp. 1–7, 2019, doi: 10.2139/ssrn.3368718.

[12] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," vol. VI, no. 1, 2018.

[13] J. Ipmawati, Kusrini, and E. Taufiq Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," *Indones. J. Netw. Secur.*, vol. 6, no. 1, pp. 28–36, 2017.

[14] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi," vol. 10, pp. 681–686, 2019.

[15] R. Vatambeti, S. V. Mantena, K. V. D. Kiran, M. Manohar, and C. Manjunath, "Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique," *Cluster Comput.*, pp. 1–17, 2023.

[16] R. Umar, I. Riadi, and D. A. Faroek, "Komparasi Image Matching Menggunakan Metode K-Nearest Neighbor ( KNN ) dan Metode Support Vector Machine ( SVM )," *J. Appl. Informatics Comput.*, vol. 4, no. 2, pp. 124–131, 2020.

[17] M. Abdar, S. R. N. Kalhori, T. Sutikno, I. M. I. Subroto, and G. Arji, "Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases.," *Int. J. Electr. Comput. Eng.*, vol. 5, no. 6, 2015.

[18] B. A. C. Martani and E. B. Setiawan, "Naïve Bayes-Support Vector Machine Combined BERT to Classified Big Five Personality on Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 6, pp. 1072–1078, 2022.

[19] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM parameter optimization using grid search and genetic algorithm to improve classification performance," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 4, pp. 1502–1509, 2016.

[20] A. Amkor and N. El Barbri, "Classification of potatoes according to their cultivated field by SVM and KNN approaches using an electronic nose," *Bull. Electr. Eng. Informatics*, vol. 12, no. 3, pp. 1471–1477, 2023.

[21] M. B. S. Rahmatullah, A. L. S. Hanani, A. M. Naim, Z. Sari, and Y. Azhar, "Detection of Credit Card Fraud with Machine Learning Methods and Resampling Techniques," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 6, pp. 923–929, 2022.

[22] N. Hafidz and D. Y. Liliana, "Klasifikasi Sentimen pada Twitter Terhadap WHO Terkait Covid-19 Menggunakan SVM, N-Gram, PSO," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 5, no. 2, pp. 213–219, 2021.

[23] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. M. Noordin, "Rice grain classification using multi-class support vector machine (SVM)," *IAES Int. J. Artif. Intell.*, vol. 8, no. 3, p. 215, 2019.

[24] L. Priyambodo *et al.*, "Klasifikasi Kematangan Tanaman Hidroponik Pakcoy Menggunakan Metode SVM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 153–160, 2022.

[25] J. M. Arockiam and A. C. Seraphim Pushpanathan, "MapReduce-iterative support vector machine classifier: novel fraud detection systems in healthcare insurance industry.," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 1, 2023.

[26] R. R. Septiawan, B. H. Prakoso, and I. Kurniawan, "DPP IV Inhibitors Activities Prediction as An Anti-Diabetic Agent using Particle Swarm Optimization-Support Vector Machine Method," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 6, pp. 974–980, 2022.

[27] F. Shamrat *et al.*, "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 1, pp. 463–470, 2021.

[28] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, "A dynamic K-means clustering for data mining," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 2, pp. 521–526, 2019.

[29] F. Rahmadayana and Y. Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 936–942, 2021.

[30] L. Mardiana, D. Kusnandar, and N. Satyahadewi, "Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 11, no. 1, pp. 97–102, 2022.