Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

# Sentiment Analysis of Flood Disaster Management in Jakarta on Twitter Using Support Vector Machines

**M Amiruddin Saddam[1)*], Erno Kurniawan Dewantara [2)], Achmad Solichin[3)]**
[1,2,3)] Faculty of Information Technology Budiluhur University, Indonesia
[1)]2111600827@student.budiluhur.ac.id, [2)] 2111600413@student.budiluhur.ac.id,
[3)]achmad.solichin@budiluhur.ac.id

**Submitted**: Dec 25, 2022 | **Accepted**: Jan 2, 2023 | **Published**: Jan 6, 2023

**Abstract:** Floods can cause negative impacts in various aspects, starting from economic, social, to health aspects. Even quoted from the site gis.bnpb.go.id, during 2022, there have been 1031 cases of flood disasters in Indonesia. Meanwhile in Jakarta, in 2022 there have been 14 cases of floods that caused hundreds of people to lose their homes. Several approaches can be taken to determine public opinion about flooding, one of which is text mining with an analysis of community sentiment. Sentiment analysis aims to determine public opinion regarding flood management in the capital city of Jakarta based on positive, neutral, and negative categories. To get the public sentiment, researchers carried out several stages, including the preprocessing stage. After obtaining public sentiment regarding flood management through the preprocessing stage, then classification is carried out based on public opinion so that it can be used by related parties for evaluation material in flood handling. In this study, the classification method used is the SVM method which is one of the supervised learning methods in machine learning. After classification, the next stage is the testing process using the K-Fold Cross Validation method. From the various sentiments obtained from Twitter data, it can be concluded that there are around 414 positive sentiments and 2464 negative sentiments related to flood handling in DKI Jakarta, while the results obtained from the test results show that the accuracy reaches 88.6%, the precision reaches 88.6% and recall reached 89.4%.

**Keywords:** Classification; Flood; Sentiment Analysis; SVM; Text Mining; Twitter

## INTRODUCTION

Floods are a problem that often occurs in various regions, both in Indonesia and abroad. Floods can be caused by various aspects, ranging from increased rainfall, clogged drains, excessive groundwater exploitation, and so on. Floods can cause negative impacts in various aspects, starting from economic, social, to health aspects. Even quoted from the site gis.bnpb.go.id, during 2022, there have been 1031 cases of flood disasters in Indonesia. Meanwhile in Jakarta, in 2022 there have been 14 cases of floods that caused hundreds of people to lose their homes.

Through technological developments, many people express their opinions and expressions through technology such as social media, one of the social media that is widely used by the public in expressing opinions and expressions is Twitter social media, where on Twitter people can write their opinions in posts so that many people can see them. (Setiawan et al., 2021).

Twitter is a social networking service that helps its users send and read text-based messages of up to 140 characters. In early 2013, Twitter users were sending more than 500 million tweets per day. The high popularity of Twitter causes this service to be utilized for various purposes in various aspects. In addition, Twitter also allows access to parts of the service through APIs to allow people to build software that integrates with Twitter, such as solutions that help companies respond to customer feedback on Twitter. Twitter data differs from data shared by most other social platforms in that it reflects information that users choose to share publicly (Duei Putri et al., 2022).

The contents of the tweet can be used to express a person's feelings towards the presidential and vice-presidential candidate pairs, for example, "I am very happy with his leadership." This is a subjective assessment or opinion (Atsqalani et al., 2022). In this tweet, the opinion used for analysis is regarding flood management in

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

DKI Jakarta. There are several approaches that can be taken to find out public opinion about flooding, one of which is text mining with community sentiment analysis. Sentiment analysis is a computational study of recognizing and expressing opinions, sentiments, evaluations, attitudes, emotions, subjectivity, judgments, or views contained in a text. (Eldha Oktaviana & Arum Sari, 2022). Sentiment analysis can also be included in the field of NLP (natural language programming), which is a data processing process with the aim of identifying and understanding the issues mentioned by the company and their own emotions on these matters. Sentiment analysis, also known as "opinion mining," is a scientific field that analyzes the opinions, feelings, evaluations, ethics, and sentiments of society towards products, services, organizations, and individuals. To perform sentiment analysis, data containing opinions on an issue is required. There are many ways to get this data, one of which is through social networks. Unlike before, public opinion can only be known through polls, group opinions, newspapers, radio, and TV (Romli et al., 2021).

Sentiment analysis aims to determine public opinion regarding flood management in the capital city of Jakarta based on positive, neutral, and negative categories. Following the collection of public sentiment regarding flood management, classification is carried out based on that sentiment so that related parties can use it for evaluation material in flood handling.

Researchers conducted several stages, including preprocessing, to gather public sentiment. Following the collection of public sentiment regarding flood management during the preprocessing stage, classification is performed based on public sentiment so that it can be used by related parties for evaluation material in flood handling. At the classification stage, the authors use the SVM method, which is a supervised learning classification method in machine learning. K-Fold Cross Validation will be used to evaluate the model.

## LITERATURE REVIEW

Gata & Bayhaqy conducted a sentiment analysis related to Islamophobia. Opinions and public opinion expressed in large numbers on Twitter can, at the very least, provide a global analysis of anti-Muslim sentiment toward Muslims around the world on the day that news of the church attack Islamization of the Church of Christ in New Zealand was reported. Researchers use Nave Bayes and SVM algorithms in combination with SMOTE to balance VADER labeling data to improve algorithm test results (Gata & Bayhaqy, 2020).

In the research conducted by Syahputra, et al, after comparing the SVM and Nave Bayes algorithms using a cross-validation evaluation model, Syahputra et al. discovered that the algorithm has a higher accuracy value than Nave Bayes. The accuracy rate of the Naïve Bayes algorithm is 85%, while the SVM algorithm achieves higher accuracy with an accuracy rate of 86%. Even though the accuracy level is determined by testing the 8020-decomposition technique, the Nave Bayes algorithm has an 80% higher accuracy rate than the SVM algorithm, which has a 79% accuracy rate (Syahputra et al., 2022).

Research conducted by Hussein, et al, said that analysis of the COVID-19 cluster can be of great help to policymakers and governments in monitoring and planning appropriate plans. Sentiment analysis is a valuable way to find out what people think and feel about events or other aspects. In this study, a cluster analysis of the COVID-19 outbreak was presented using the K-means algorithm. Twitter data related to COVID-19 is explored and presented using the TF-IDF technique as a weighting plot. The data were clustered using the K-means algorithm and Euclidean distance measurements. In addition, this study revealed sentiment toward COVID-19 by applying sentiment analysis to each cluster using the vocabulary-based TextBlob engine. Using data visualization tools such as t-SNE and Word Cloud, the results of the analysis are illustrated. According to the experimental results, there are 9 relative clusters obtained with various subjects, with the highest score of 83.25% positive reports and 16.75% negative reports (Hussein et al., 2021).

In the research conducted by Nofiyanti, et al, sentiment analysis for disaster management, in this case referring to the Regional Disaster Management Authority (BPBD), is based on public opinion with positive, neutral, or negative social media categories on Twitter, using the Python library using the NLP method. The results of this sentiment analysis can be used by disaster managers to evaluate the management of disasters so they can get better in the future. According to the study, there are 23.53% positive tweets, 57.35% neutral tweets, and 19.12% negative tweets. According to the results, most Indonesians have a neutral view of disaster management. (Nofiyanti & Oki Nur Haryanto, 2021) .

Sentiment analysis was conducted by Khalid et al. to understand public opinion about the COVID-19 vaccination on the social network Twitter. The introduced system proposes to develop a model architecture based on a two-way deep short- and long-term memory (LSTM) neural network to analyze tweet data in positive, neutral, and negative forms. Therefore, the overall accuracy of the model developed based on the validation data is 74.92%. The results obtained from the psychological analysis system on tweet data collected from the COVID-19 vaccine show that neutral sentiment accounts for 69.5% of tweets, negative sentiment has a rate of tweets below 20.75 percent, and positive sentiment accounts for a high percentage. low rate of tweets reached 9.67% (Khalid, Talal, Faraj, & Yassin, 2022).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sentiment analysis performed by Kartika et al. was able to predict catastrophic events in predetermined classes. The technique used to preprocess the data affects how accurate the performance is. In addition, a Support Vector Machine (SVM) algorithm was run to classify the data into three (3) categories: witness, non-witness, and unknown. The total amount of data processed for the three class labels is 3000. The SVM algorithm has two methods: one-to-one (OVO) for labels with two layers and one-to-one (OVA) for labels with more than two layers. Multilayer SVM with the OVA method was used in this study. In comparison to previous studies and methods, the OVA method with RBF multiplier has a classification accuracy value of 87.03% (Kartika Delimayanti et al., 2021).

In a study conducted by Fridom et al., positive sentiment prevailed in the sentiment analysis of obesity-related tweets by text mining over negative sentiment and neutral sentiment. create. The accuracy value with the Naïve Bayes Classifier algorithm is in the "Excellent Classifier" category. That means the Naive Bayes Classifier algorithm successfully predicts the sentiment category in this search. (Fridom Mailo & Lazuardi, 2019).

In the research conducted by Turmudi, it is said that the main problem of sentiment analysis is that the size of the classification space is quite large compared to the feature space, which often occurs in a text with tens of thousands of features, and the data has too much noise in it. The model used for this study is Nave Bayes. Nave Bayes can be combined with feature extraction. In the test, the feature extraction of the Count Vectorizer and TFIDF Vectorizer is compared using the cross-validation technique to improve the Nave Bayes classifier. The measurement of value is done by comparing trials without validation and using assertions. Accuracy can be measured using confusion, precision, and recall matrices. According to the research findings, TF-IDF Vectorizer function extraction outperforms Count Vectorizer with the highest accuracy of 85.98%, and the extraction function with cross-validation outperforms none. use cross-validation with the highest accuracy. value of 97.67%. So, the best-used feature extraction test is the TF-IDF vector, and by using the cross-validation technique, it can improve the performance of the Nave Bayes model in Twitter sentiment analysis in Indonesian to achieve a differential accuracy of 11.69% (Turmudi & Syarief Yasah, 2020).

Sitepu et al., in their research, apply the support vector machine algorithm to classify Shopee user review data. To solve this problem, the research is conducted in several steps, which are: After preprocessing the text from the dataset, feature extraction is performed, followed by word weighting using the TF-IDF method, and finally, the SVM algorithm is used to evaluate the model. In the research results, it was found that the word expressing the most positive opinion of Shopee customers was "good," with a total of 4684 words. while the word with the most negative reviews is "seller," with 68 words. From the five sentiment analysis models tested, the mean of the confusion matrix obtained is precision = 1, recall = 0.97, and f1-score = 0.98. From this study, it can be concluded that the SVM algorithm has been proven to apply to performing sentiment analysis on user reviews of Shopee products with an average accuracy rate of 97.3% (Sitepu, Munthe, & Harahap, 2022).

In research conducted by Muktafin, et al, Perceived analysis of public service at a hospital in Yogyakarta, Indonesia, using the K-nearest neighbor (KNN) algorithm with inverse frequency document frequency weighting (TF-IDF) and classified into two categories "satisfied" and "dissatisfied". Then find out which attributes are rated as "satisfied" and "dissatisfied." An NLP approach is needed to improve the conversational language so that the system is easier to understand. The results of this study received an accuracy of 74.00%, a precision of 76.00%, and a recovery of 73.08%. less satisfied with the waiting time (Muktafin & Kusrini, 2021).

## METHOD

The research method used is to use quantitative methods. The sample used in this study is Twitter data with the keyword "Jakarta Flood." The stages carried out in this study are depicted in Fig. 1.

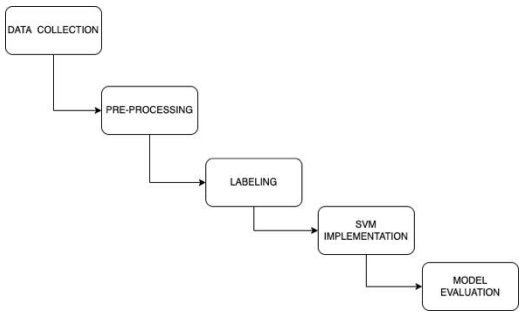


Fig 1 Step of Research Work

### Data Collection

Twitter data crawling is the process of extracting user data and tweets originating from Twitter based on specific keywords. Data extraction is accomplished through the use of Application Programming Integration (API)

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

(Alhakiem & Setiawan, 2022). The crawling is done using the SNScrape library provided in the Python programming language. Only tweets in Indonesian were collected over three months, from 1 September 2022 to 29 November 2022, totaling 28,778 tweets with the keyword "Jakarta Flood."

**Pre-Processing**
Preprocessing is a series of steps performed to perform data cleaning, data integration, data transformation, and data reduction. This stage is important because if you analyze data that has not been processed carefully, it can provide misleading information. If there is a lot of irrelevant information or a lot of data noise, the data quality will decrease. To produce high-quality data requires understanding and applying the appropriate preprocessing methods used for each case study (Fatihah Rahmadayana & Yuliant Sibaroni, 2021). The preprocessing stages carried out are data cleansing, case folding, tokenizing, stopword removal, normalization, and stemming. The preprocessing stage is carried out in Python using the provided NLTK library.

**Data Cleansing**
The first step in preprocessing is data cleansing. The data obtained from Twitter is selected during the data cleansing process based on its relevance to the topic chosen, namely flood management in DKI Jakarta.

**Case Folding**
This stage is the process of converting text that is irregular in the use of letters in writing commentary text, resulting in the inconsistent text of the comments. This case-folding process functions to change the letters in the commentary text that have been cleaned into standard form, that is, all lowercase letters (Kurniawan et al., 2019).

**Tokenizing**
The next stage is tokenizing. In this process, characters, URLs, emoticons, hashtags, and other characters are removed. This process is an important process in sentiment analysis to minimize errors in writing. After removing characters, tokenizing will split the sentence into several words.

**Stopword Removal**
The fourth stage is stopword removal. In this process, the tokenized words are filtered. Words that have no meaning will be discarded or deleted. This will be very helpful in the process of sentiment analysis.

**Normalization**
The next stage is the word normalization process. Words that have been case-folded, tokenized, and stopword removed will be normalized again in order to equate words with the same meaning but are different words or correct slang to become standard words.

**Stemming**
The last stage is stemming. The process of converting word forms into root words by finding the root words of each filtered word for each affixed word to be converted into a base word, the purpose of stemming is to maximize and optimize the text processing process (Putri Nirwandani & Cahya Wihandika, 2021). In the stemming stage of this study, using the Indonesian literary library in python.

**Labeling**
A machine learning algorithm (ML) is a two-step process, training, and testing dataset. Thus, we need to label some of the tweets with real opinions to be used in the training stage (Alabid & Katheeth, 2021). The labeling process is carried out to determine the category of sentiment that has been obtained, whether the sentiment is positive or negative.

**SVM Implementation**
After the labeling process, the next stage is the implementation of the support vector machine model. Support Vector Machine (SVM) has proven to be useful in machine learning, especially for classifying multiclass data. There are two approaches for multilayer SVMs. First, process all the live data into a single optimized formula. The second part describes the multilayer in a binary SVM sequence. The idea behind binary SVM is to build a multiclass classifier from binary with a one-to-one technique (Alita et al., 2019).

**Model Evaluation**
The next stage is to test the model using the K-Fold Cross Validation technique. Model performance evaluation is performed based on error metrics to get model accuracy. To evaluate the performance of the model, the K-fold

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Cross-Validation method was used. The number of tricks used for evaluation and training is 10 data. The data is then trained to get an exact value. In this assessment, the confusion matrix was used not only to obtain accuracy, but also the precision, recall, and f-score.

## RESULT

Before carrying out the analysis process, it is necessary to carry out several stages starting from data crawling, and pre-processing to testing. In the previous discussion, it was explained that there were 6 pre-processing stages carried out in this study, the first pre-processing stage was data cleansing. At the data cleaning stage, the researcher examined the data based on relevance and also checked the similarity of the data for each dataset as shown in table 1. Checking based on relevance was seen from the linkage of the data with the topic discussed, namely flood management in Jakarta. Meanwhile, data that has similarities will be discarded and only the initial data will be taken.

Table 1 Data Cleansing Result

| Dataset Period | Total Data | Similarity | Relevant Count | Not Relevant Count |
|---|---|---|---|---|
| 01-29 September 2022 | 2054 | 55 | 347 | 1652 |
| 10-30 Oktober 2022 | 9838 | 523 | 1918 | 7397 |
| 01-29 November 2022 | 5114 | 217 | 613 | 4284 |

After the data cleansing process is carried out, the datasets are then merged or merged for further case folding, tokenizing, stopword, normalization, and stemming processes as shown in table 2.

Table 2 Pre-Processing Results

| Tweets | Case Folding | Tokenizing | Stopword Removal | Normalization | Stemming |
|---|---|---|---|---|---|
| Padahal faktanya penanganan banjir Jakarta era Anies, normalisasi mandek hingga sumur resapan tidak efektif, yang terjadi tetap saja banjir ketika musim hujan, yang terdampak di hampir semua kawasan di Jakarta. SPBU Sadtember | padahal faktanya penanganan banjir jakarta era anies, normalisasi mandek hingga sumur resapan tidak efektif, yang terjadi tetap saja banjir ketika musim hujan, yang terdampak di hampir semua kawasan di jakarta. spbu sadtember | padahal, faktanya, penanganan,banjir,jakarta,era,anies,normalisasi,mandek,hingga,sumur,resapan,tidak,efektif,yang,terjadi,tetap,saja,banjir,ketika,musim,hujan,yang,terdampak,di,hampir,semua,kawasan,di,jakarta,spbu,sadtember | faktanya, penanganan,banjir,jakarta,era,anies,normalisasi,mandek,sumur,resapan,efektif,banjir,musim,hujan,terdampak,kawasan,jakarta,spbu,sadtember | faktanya, penanganan,banjir,jakarta,era,anies,normalisasi, mandek,sumur,resapan,efektif,banjir,musim,hujan,terdampak,kawasan,jakarta, spbu,sadtember | fakta, tangan,banjir,jakarta,era,anies,normalisasi, mandek,sumur,resap,efektif, banjir,musim, hujan,dampak,kawasan,jakarta,spbu,sadtember |

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

| | | | | | |
|---|---|---|---|---|---|
| @BiLLRaY2019 @Pencerah___ Banjir di jakarta Diperparah dgn buruknya drainase pdhl sumur serapan anggaran sdh dibikin tp unfaedah ,sumur dibuat utk serap anggaran ke kantong wan ambruk dkk & tgupp..yg patut diacungi jempol DKI ketika dipimpin oleh Anies adalah korupsinya .klo ini sih juara nasional.. | @billray2019 @pencerah___ banjir di jakarta diperparah dgn buruknya drainase pdhl sumur serapan anggaran sdh dibikin tp unfaedah ,sumur dibuat utk serap anggaran ke kantong wan ambruk dkk & tgupp..yg patut diacungi jempol dki ketika dipimpin oleh anies adalah korupsinya .klo ini sih juara nasional.. | banjir, di, jakarta, diperparah, dgn,buruknya, drainase,pdhl,sumur,serapan, anggaran,sdh, dibikin,tp,unfaedah,sumur,dibuat,utk,serap, anggaran,ke,kantong,wan,ambruk,dkk,amp,tguppyg,patut,diacungi,jempol,dki,ketika, dipimpin,oleh, anies,adalah,korupsinya,klo,ini,sih,juara,nasional | banjir, jakarta, diperparah, dgn, buruknya,drainase,pdhl,sumur, serapan,anggaran,sdh,dibikin,tp,unfaedah,sumur,utk,serap,anggaran,kantong,wan,ambruk, dkk,amp,tguppyg,patut,diacungi,jempol,dki,dipimpin,anies,korupsinya,klo,sih,juara,nasional | banjir, jakarta, diperparah, dengan, buruknya,drainase,padahal,sumur,serapan,anggaran,sudah,dibikin,tapi,unfaedah,sumur,untuk,serap,anggaran,kantong,wan,ambruk,dan kawan-kawan,amp,tguppyg,patut,diacungi,jempol,dki,dipimpin,anies, korupsinya,kalau,sih,juara,nasional | banjir, jakarta, parah, dengan, buruk,drainase,padahal,sumur,serap,anggar,sudah,bikin,tapi,unfaedah,sumur,untuk,serap,anggar,kantong,wan,ambruk,dan kawan,amp,tguppyg,patut,acung,jempol,dki,pimpin,anies,korupsi,kalau,sih,juara,nasional |
| Bendungan Ciawi dan Sukamahi akan jadi Dry Dam pertama di Indonesia, dalam upaya untuk mengurangi kerentanan banjir kawasan Metropolitan Jakarta. | bendungan ciawi dan sukamahi akan jadi dry dam pertama di indonesia, dalam upaya untuk mengurangi kerentanan banjir kawasan metropolitan jakarta. | bendungan, ciawi, dan, sukamahi, akan,jadi,dry,dam,pertama,di, indonesia,dalam,upaya,untuk,mengurangi,kerentanan,banjir,kawasan,metropolitan,jakarta | bendungan, ciawi, sukamahi, dry,dam,indonesia,upaya,mengurangi,kerentanan,banjir,kawasan,metropolitan,jakarta | bendungan, ciawi, sukamahi, kering,bendungan,indonesia,upaya,mengurangi,kerentanan,banjir,kawasan,metropolitan,jakarta | bendung, ciawi, sukamahi, kering,bendung,indonesia,upaya,kurang,rentan,banjir,kawasan,metropolitan,jakarta |
| Banjir Jakarta Dinilai Cepat Surut Berkat Anies: Dia Sangat Saintifik, Layak Jadi Presiden | banjir jakarta dinilai cepat surut berkat anies: dia sangat saintifik, layak jadi presiden | banjir, jakarta, dinilai, cepat,surut,berkat,anies,dia,sangat,saintifik, layak,jadi,presiden | banjir, jakarta, dinilai,cepat,surut,berkat,anies,saintifik,layak, presiden | banjir, jakarta, dinilai,cepat,surut,berkat,anies,saintifik,layak,presiden | banjir, jakarta,nilai,cepat,surut,berkat,anies,saintifik,layak,presiden |
| @Siregar_najeges Banjir Jakarta lebih terkendali dan waktu surut lebih cepat, terimakasih dedikasi @aniesbaswedan & pemerintah DKI atas kinerjanya 👍 | @siregar_najeges banjir jakarta lebih terkendali dan waktu surut lebih cepat, terimakasih dedikasi @aniesbaswedan & pemerintah dki atas kinerjanya 👍 | banjir, jakarta, lebih,terkendali,dan,waktu,surut,lebih,cepat, terimakasih,dedikasi,amp,pemerintah,dki,atas,kinerjanya | banjir, jakarta, terkendali, surut,cepat,terimakasih,dedikasi,amp,pemerintah,dki,kinerjanya | banjir, jakarta, terkendali, surut,cepat,terimakasih,dedikasi,amp,pemerintah,dki,kinerjanya | banjir, jakarta, kendali,surut, cepat,terimakasih,dedikasi, amp,perintah, dki,kerja |

After the pre-processing stage, the next step is labeling. At the labeling stage, the data will be categorized into 2 categories, namely positive and negative. The data labeling process can be done manually or using a sentiment dictionary. The sentiment dictionary contains words as well as the weight for each word. The Python

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

programming language has a library that can do labeling automatically based on the polarity of each word. However, for now, the library can only be used for words containing English. Therefore, there is a need for a manual labeling process or using a dictionary as in table 3. If the weight of the sentence is less than 0 then the sentence is negative but if it is the other way around then it is positive.

Table 3 Data Labelling

| Tweets | Positive Word | Negatif Word | Label |
|---|---|---|---|
| fakta tangan banjir jakarta era anies normalisasi mandek sumur resap efektif banjir musim hujan dampak kawasan jakarta spbu sadtember | Tangan(1), resap(3),efektif(5) | Fakta (-3), banjir(-4),hujan(-2),dampak(-3) | Negative |
| banjir jakarta parah dengan buruk drainase padahal sumur serap anggar sudah bikin tapi unfaedah sumur untuk serap anggar kantong wan ambruk dan kawan amp tguppyg patut acung jempol dki pimpin anies korupsi kalau sih juara nasional | Anggar(3), kantong(1),jempol(4) | Banjir(-4), parah(-5),buruk(-5),ambruk(-5),patut(-5),korupsi(-4) | Negative |
| bendung ciawi sukamahi kering bendung indonesia upaya kurang rentan banjir kawasan metropolitan jakarta | Upaya(1) | bendung(-3), kering(-3),rentan(-4),banjir(-4) | Negative |
| banjir jakarta nilai cepat surut berkat anies saintifik layak presiden | Nilai (5), cepat(3),berkat(1),layak(3) | Banjir (-4) | Positive |
| banjir jakarta kendali surut cepat terimakasih dedikasi amp perintah dki kerja | cepat(3), terimakasih(5),kerja(2) | Banjir(-4), | Positive |

Meanwhile, to find out the number of positive and negative sentiments, the writer visualizes it in the form of a bar chart as shown in Fig 2

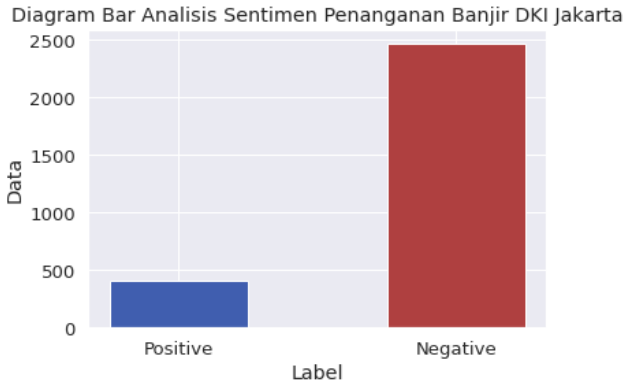


*Fig 2 Sentiment analysis bar diagram for DKI Jakarta flood handling*

Fig 2 shows that 414 of the 2878 opinion data collected contain positive sentiment. Meanwhile, 2464 other data points contain negative sentiment. After labeling the data, the next stage is the implementation of the SVM model or commonly known as the Support Vector Machine model.

Support Vector Machine (SVM) is a classification method using Machine Learning (supervised learning) which predicts the class of a model or pattern based on the results of the training process. Grading is done by finding the hyperplane or decision boundary that separates one class from another, which in this case plays a role in separating positive sentiment tweets (marked +1) from negative sentiment tweets (marked -1). SVM looks for hyperplane values using auxiliary vectors and margin values. In this study the input data which has a vector

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Issue 1, January 2023
DOI : https://doi.org/10.33395/sinkron.v8i1.12063
e-ISSN : 2541-2019
p-ISSN : 2541-044X

representation is obtained from the weighing process. By conducting training in SVM classification, it will then produce values or patterns that will be used in the SVM testing process to mark sentiments in tweets (Sulastomo et al., 2022).

Before implementing the SVM model, the dataset obtained was divided into 2014 training data and 864 data testing data and then the feature extraction process. Feature extraction is used to explore potential information and represent words as feature vectors. This vector will be used as input for the classification method in the next stage. One of the techniques in feature extraction is using the IDF TF. Term Frequency or TF is calculated based on the number of occurrences of each word in each document, while Inverse Document Frequency or IDF is calculated based on the number of occurrences of each word in the entire document (Irmanda & Astriratma, 2020). The process in this SVM uses a linear kernel which will then train a sentiment model for data classification. The report of the classification can be seen in Fig 3 below.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Negatif | 0.90 | 0.98 | 0.94 | 738 |
| Positif | 0.80 | 0.37 | 0.51 | 126 |
| accuracy | | | 0.89 | 864 |
| macro avg | 0.85 | 0.68 | 0.72 | 864 |
| weighted avg | 0.89 | 0.89 | 0.88 | 864 |

*Fig 3 Sentiment classification report using SVM*

In Fig 3, it can be seen that the precision value of data containing negative sentiment is 90%, while the precision value of data containing positive sentiment is 80%. The recall value of data containing negative sentiment is 98%. While the recall value for data with positive sentiment is 37%, the f-score value for data with negative sentiment is 94%, while the f-score value for data with positive sentiment is 51%, and the value of support for data with negative sentiment is 738, while the value of support for data with positive sentiment is 126. The last step that needs to be done in sentiment analysis is the evaluation of the model used. In the evaluation process, the technique used is K-Fold Cross Validation. Cross-validation is a statistical method for evaluating and comparing learning algorithms by dividing the data into two segments, one segment is used to learn or train data and the other is used to validate the model. In cross-validation, the training and validation sets must be crossover successively so that each data has the opportunity to be validated (Ilahiyah & Nilogiri, 2018). In the evaluation stage, the data will be divided into 10 folds, which means 10 times the testing process will be carried out as shown in table 4 below.

Table 4 Evaluation of the K-Fold Cross Validation Model

| K-Fold | Accuracy | Precision | Recall | F-Score |
|---|---|---|---|---|
| K-2 | 0.875869 | 0.886514 | 0.894676 | 0.87789 |
| K-3 | 0.879342 | 0.886514 | 0.894676 | 0.87789 |
| K-4 | 0.883808 | 0.886514 | 0.894676 | 0.87789 |
| K-5 | 0.885303 | 0.886514 | 0.894676 | 0.87789 |
| K-6 | 0.885794 | 0.886514 | 0.894676 | 0.87789 |
| K-7 | 0.883816 | 0.886514 | 0.894676 | 0.87789 |
| K-8 | 0.883812 | 0.886514 | 0.894676 | 0.87789 |
| K-9 | 0.883306 | 0.886514 | 0.894676 | 0.87789 |
| K-10 | 0.886299 | 0.886514 | 0.894676 | 0.87789 |
| K-11 | 0.886299 | 0.886514 | 0.894676 | 0.87789 |

In table 4, it can be seen that the model evaluation was carried out 10 times, resulting in accuracy reaching 88.6%, precision reaching 88.6%, and recall reaching 89.4%.

## DISCUSSIONS

In this study, the results of sentiment analysis were carried out using the SVM method and tested using the K-Fold Cross Validation technique. The data used for testing is as much as 30% of the entire dataset used. Just

*name of corresponding author

like in the research conducted by Syahputra, et al, in this study, it was found that the accuracy obtained using SVM and K-Fold Cross Validation was quite high at 88.6%.

In addition, this study also uses the TF-IDF technique to carry out the process of extracting features of public sentiment. The amount of data that is still small makes the process of training and data testing still needs to be improved. And also, the need for comparisons between methods to be able to increase the level of accuracy of sentiment analysis.

## CONCLUSION

The research carried out can analyze public sentiment regarding flood management in the capital city of Jakarta, so it is hoped that the results of this sentiment analysis can be used as evaluation material for related parties. From the various sentiments obtained from Twitter data, it can be concluded that there are around 414 positive sentiments and 2464 negative sentiments related to flood handling in DKI Jakarta. Model evaluation was carried out 10 times with accuracy reaching 88.6%, precision reaching 88.6%, and recall reaching 89.4%. The existence of a preprocessing process using data cleansing, case folding, tokenizing, stopword, normalization, and stemming techniques makes the analysis carried out better. In addition, the SVM and TF-IDF methods for feature extraction also increase the accuracy of sentiment analysis.

## REFERENCES

Alabid, N. N., & Katheeth, Z. D. (2021). Sentiment analysis of twitter posts related to the covid-19 vaccines. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(3), 1727–1734. Retrieved from https://doi.org/10.11591/ijeecs.v24.i3.pp1727-1734

Alhakiem, H. R., & Setiawan, E. B. (2022). Aspect-Based Sentiment Analysis on Twitter Using Logistic Regression with FastText Feature Expansion. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(5), 840–846. Retrieved from https://doi.org/10.29207/resti.v6i5.4429

Alita, D., Priyanta, S., & Rokhman, N. (2019). Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter. *Journal of Information Systems Engineering and Business Intelligence*, 5(2), 100. Retrieved from https://doi.org/10.20473/jisebi.5.2.100-109

Atsqalani, H., Hayatin, N., & Aditya, C. S. K. (2022). Sentiment Analysis from Indonesian Twitter Data Using Support Vector Machine and Query Expansion Ranking. *Jurnal Online Informatika*, 7(1), 116. Retrieved from https://doi.org/10.15575/join.v7i1.669

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1). Retrieved from https://doi.org/10.23960/jitet.v10i1.2262

Eldha Oktaviana, N., & Arum Sari, Y. (2022). Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features Dan Support Vector Machine. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, 9(2), 357–362. Retrieved from https://doi.org/10.25126/jtiik.202295625

Fatihah Rahmadayana, & Yuliant Sibaroni. (2021). Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 936–942. Retrieved from https://doi.org/10.29207/resti.v5i5.3457

Fridom Mailo, F., & Lazuardi, L. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health*, 4(1).

Gata, W., & Bayhaqy, A. (2020). Analysis sentiment about islamophobia when Christchurch attack on social media. *Telkomnika (Telecommunication Computing Electronics and Control)*, 18(4), 1819–1827. Retrieved from https://doi.org/10.12928/TELKOMNIKA.V18I4.14179

Hussein, A., Ahmad, F. K., & Kamaruddin, S. S. (2021). Cluster Analysis on Covid-19 Outbreak Sentiments from Twitter Data using K-means Algorithm. *Journal of System and Management Sciences*, 11(4), 167–189. Retrieved from https://doi.org/10.33168/JSMS.2021.0409

Irmanda, H. N., & Astriratma, R. (2020). Klasifikasi Jenis Pantun dengan Metode Support Vector Machines (SVM). *Jurnal RESTI*, 4(5), 915–922.

Kartika Delimayanti, M., Sari, R., Laya, M., Reza Faisal, M., & Pahrul, dan. (2021). Pemanfaatan Metode Multiclass-SVM pada Model Klasifikasi Pesan Bencana Banjir di Twitter. *Edu Komputika*, 8(1). Retrieved from http://journal.unnes.ac.id/sju/index.php/edukom

Khalid, E. T., Talal, E. B., Faraj, M. K., & Yassin, A. A. (2022). Sentiment analysis system for COVID-19 vaccinations using data of Twitter. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(2), 1156–1164. Retrieved from https://doi.org/10.11591/ijeecs.v26.i2.pp1156-1164

Kurniawan, S., Gata, W., Ayu Puspitawati, D., Nurmalasari, Tabrani, M., & Novel, K. (2019). Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI*, 3(2), 176–183.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Muktafin, E. H., & Kusrini, P. (2021). Sentiments analysis of customer satisfaction in public services using K-nearest neighbors' algorithm and natural language processing approach. *Telkomnika (Telecommunication Computing Electronics and Control)*, 19(1), 146–154. Retrieved from https://doi.org/10.12928/TELKOMNIKA.V19I1.17417

Nofiyanti, E., & Oki Nur Haryanto, E. M. (2021). Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia. *Jurnal Ilmiah SINUS*, 19(2), 17. Retrieved from https://doi.org/10.30646/sinus.v19i2.563

Putri Nirwandani, E., & Cahya Wihandika, R. (2021). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputere*, 5(3), 1039–1047. Retrieved from http://j-ptiik.ub.ac.id

Romli, I., Prameswari R, S., & Kamalia, A. Z. (2021). Sentiment Analysis about Large-Scale Social Restrictions in social media Twitter Using Algoritm K-Nearest Neighbor. *Jurnal Online Informatika*, 6(1), 96. Retrieved from https://doi.org/10.15575/join.v6i1.670

Setiawan, H., Utami, E., & Sudarmawan, S. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(1), 43–51. Retrieved from https://doi.org/10.31603/komtika.v5i1.5189

Sitepu, M. B., Munthe, I. R., & Harahap, S. Z. (2022). Implementation of Support Vector Machine Algorithm for Shopee Customer Sentiment Analysis. *Sinkron*, 7(2), 619–627. Retrieved from https://doi.org/10.33395/sinkron.v7i2.11408

Syahputra, R., Yanris, G. J., & Irmayani, D. (2022). SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter. *Sinkron*, 7(2), 671–678. Retrieved from https://doi.org/10.33395/sinkron.v7i2.11430

*name of corresponding author