JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2104-2117

SENTIMENT ANALYSIS ABOUT THE 2024 PRESIDENTIAL ELECTION USING CNN METHOD

Alif Faidhil Ahmad*[1)], Kemas Muslim L.[2)]
1. Telkom University, Indonesia
2. Telkom University, Indonesia

Article Info
Keywords: Sentiment Analysis, Convolutional Neural Network (CNN), Support Vector Machine (SVM), Media *Online*, Pemilu 2024, Akurasi *Method, Algoritma,* Opini, Afiliasi.

Article history:
Received 16 Agustus 2024
Revised 21 September 2024
Accepted 1 Oktober 2024
Available online 1 September 2025

DOI :
https://doi.org/10.29100/jipi.v10i3.6457

* Corresponding author.
Alif Faidhil Ahmad
E-mail address:
alifahmad@student.telkomuniversity.ac.id

ABSTRACT
The upcoming 2024 Indonesian General Election (Pemilu 2024) will be interesting news for online media users. With so much news about the election, online media has become one of the most effective media used to guide public opinion. Apart from that, public opinion is that the coverage in online media for each candidate is not balanced or not because a media is considered to have an affiliation with a particular candidate. To prove this opinion, sentiment analysis will be carried out on several online media in order to prove whether people's opinions are correct or not. Although previous research has used various platforms and achieved various levels of accuracy using the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) methods with various features, this analysis will be developed using the Convolutional Neural Network (CNN) method to obtain higher accuracy and will be compared with the Support Vector Machine (SVM) method from the media platforms Detik.com, CNN Indonesia and CNBC Indonesia. The final results prove that the use of the Convolutional Neural Network (CNN) method shows an average combined performance of 65% (Cancidate 1 = 61%, Candidate 2 = 69%, Candidate 3 = 65%) smaller than the performance of the Support Vector Machine (SVM) method. with a combined average of 74% (Candidate 1 = 73%, Candidate Candidate 2 = 77%, Candidate Candidate 3 = 72%). This study provides insights into optimizing sentiment classification techniques for Online Media platforms, emphasizing the importance of leveraging semantic and contextual information in sentiment analysis tasks.

I. INTRODUCTION

Sentiment analysis is a technique of computationally recognizing and classifying the viewpoints represented in a text, to determine the writer's opinion whether positive or negative. Currently, the development of online media in the form of news is experiencing very rapid growth. Ahead of the 2024 General Election (Pemilu) for Presidential Candidates (Capres) and Vice Presidential Candidates (Cawapres), many parties are using social media as a channel to campaign and increase popularity [4]. Some online news media that are very actively used are detikcom, cnnindonesiacom, and cnbcindonesia.

Convolutional Neural Networks, or CNN as it is more often known, are one of the two deep learning algorithms. They typically use two different algorithms and use the multi-layer perceptron (MLP) principle, which is intended to be used for data augmentation, among other things[15]. This is used for capturing data such as images and text. CNN is a type of deep neural network algorithm that has a depth layer in its network and many other features that are applied to citra data. CNN is typically used to identify objects or individuals. The line operation in the CNN method is a complex process that involves eight dimensions in the kernel's collection form. CNN has a layer with the ability to do feature extraction and classification[16]. Concept SVM may be easily explained as an attempt to find an ideal hyperplane that is bounded as a pemisah between two input kelas. Support Vector Machines (SVM) are a safe learning method used for classification. The SVM classification model operates by selecting a sample from the class or label with a maximum allowable margin. For this reason, support vector machines (SVM) are often used in safe learning scenarios and are particularly useful for classifying regression vectors or pendukung vectors. The SVM algorithm searches for the best hyperplane by searching the kelas rank. In SVM, a hyperplane is a feature that can be used to balance one class against another. Functions one and two that are used in classifying the members of the class are called garis, functions three and four are called bidang, and functions that are typically used to classify the members of the class that are more tinggi than the other three are called hyper selection.

This research analyzes online media sentiment (Detikcom, CNN Indonesia, CNBC Indonesia) regarding the 2024 Indonesian Election, focusing on presidential and vice presidential candidates, using CNN and SVM methods.

There have been several previous studies related to sentiment analysis that focused on the 2019 Indonesian Presidential Election on Twitter social media using the Naïve Bayes Classifier method. Of the 240 data, 134 negative sentiments and 106 positive sentiments. This study shows that the results of data classification using the this method can provide an accuracy of 73%, negative class precision of 78%, resulting in a class classification of 22% and positive precision of 66%, resulting in a class classification of 34% [2].For the current research, we will use the CNN and SVM methods which will measure the accuracy of the two methods after going through the preprocessing stage [3]. The CNN method is a method that can extract features from global information and can perform convolution operations from previous layers so that data can be extracted [4]. SVM is a method in supervised learning that can be used for classification and regression.

Thus, in the previous study, Naïve Bayes Classifier was used quite frequently to classify text identification so that a good analysis could be stated. On the other hand, CNN can be defined as having the ability to identify an object, an adegan, and text from a few sources so that research on whether or not to use CNN's motor to classify the data accurately and SVM as a comparison because they are similar to each other. Research Summary: Several scholars have recently examined this sentimen analysis, such as "Sentimen Analysis of the 2024 Presidential Proposal Using Support Vector Machine" by Azzawagama Firdaus, Anton Yudhana, and Imam Riadi of Ahmad Darhlan University. According to Ziz Perdana, Arief Hermawan, and Donny Avianto of Universitas Teknologi Yogyakarta, "Sentiment analysis of the election postponement issue on Twitter using the Naive Bayes classifier".

There is an opinion in the community that online media coverage of each candidate is unbalanced or unreasonable because a media outlet is considered to have an affiliation with a particular candidate. With this research, we can prove whether the public's opinion is correct or not. The domain of natural language processing Language study and computational linguistics in the field of artificial intelligence have influenced the development of Natural Language Processing (NLP), a subfield of computer science and engineering. Using natural language to facilitate human-machine or other-device interaction is the primary objective of natural language processing (NLP). NLP focuses on natural language processing, namely the language generally used by humans to communicate with each other. In order for a computer to understand human or user intent, the language received by the computer needs to be processed and understood first.

Sentiment analysis is a process of extracting opinions in text form. Information in text form is often found on the internet, both in blogs, social media, and sites that contain reviews. With the help of sentiment analysis, initially unstructured information can be transformed into more structured data. In general, sentiment analysis is divided into three main levels [4].

1. Document Level

Sentiment analysis at the document level assumes that each document conveys an opinion about a single entity. Documents are analyzed to determine whether they have a positive, negative, or neutral orientation. Below are a few examples of data that show how to classify the aforementioned parameter.

- “Prabowo 'Controlled' TikTok Residents Allegedly Because of Emotion, Not Rationality.” (Negative)

- “See in more detail the Health Vision and Mission of 3 Candidates for the 2024 Election.” (Neutral)

- “Prabowo-Gibran Wins National Recapitulation Calculations Abroad.” (Positive)

2. Sentence Level

Sentiment analysis at the sentence level breaks down a document into sentences. Each sentence is considered as a single entity that is analyzed separately. At this level, a distinction is made between objective sentences that convey factual information and subjective sentences that express opinions. Below are a few examples of data that show how to classify the aforementioned parameter.

- “The Vice Presidential Debate Suddenly Disappeared? KPU Opens Vote”. (Factual & Subjective)

- “Otto Hasibuan repeats election lawsuit: There is no legal basis.” (Factual & Subjective)

- “Cak Imin on Crowded Campuses Criticizes Jokowi: I'm Disappointed by the Palace's Response.” (Factual & Subjective)

3. Aspect Level

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2104-2117

Sentiment analysis at the aspect level focuses on extracting product aspects and then analyzing feedback or evaluating opinions related to the extracted aspects. Below are a few examples of data that show how to classify the aforementioned parameter.

- "Veracity Survey Results Predict Two Rounds of Presidential Election, Prabowo-Gibran Wins: Jokowi believes the inauguration must be hastened." (Evaluation Opinion)

- "Reason for Village Head Revoking AMIN Campaign Location Permit: New Field Fixed, residents are taking back the promise of the field." (Evaluation Opinion)

- "Airlangga: Prabowo-Gibran's votes in various regions are above 60 percent, Surya Palo denies asking for a repeat vote in these areas." (Evaluation Opinion)

The corelation between aspects of each level of analysis in this research uses all these aspects to identify data from online media so as to achieve accurate results because each data is separated from each level starting from the title of the opinion which states the positive and negative of the sentence. factual and subjective opinions as well as feedback analysis or evaluation of opinions from these data.

II. METHODS

This chapter describes the implementation processes strategy into practice to build on previous research that have been discussed. The dataset is derived through data scraping from Indonesian reviews of the article on the Media Online (CNN Indonesia, Detik.com, CNBC Indonesia). After deriving the dataset, preprocessing data is executed, then data labeling is executed that specifically method Convolutional Neural Network (CNN) is a structure in the realm of deep learning that involves a number of representation layers. With this deep structure, CNN has the ability to automatically extract representative characteristics from data through nonlinear transformation and nonlinear function approximation and The Support Vector Machine (SVM) method is included in the machine learning supervised learning category, which requires sample data for the process. The SVM algorithm was developed by Boser, Guyon, and Vapnik in 1992, combining concepts from previous computational theory. SVM has the ability to transform training data into higher dimensions using non-linear patterns.

The Convolutional Neural Network (CNN) method will explain how this CNN configuration architecture can run with the categories of convolution layers, pooling layers and fully connected layers.

- Convolution Layer

  Applying convolution operations to the input data is done at a crucial step in the CNN structure called the convolution layer. Convolution, as a mathematical concept, is the process of integrating two sets of data to create a third function by combining two functions. The input and filter interact to produce the output of the convolution layer, where the kernel contributes to the formation of a feature map.

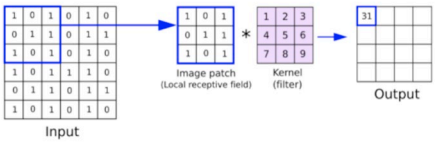

Fig. 1. Convolution Layer

- Pooling Layer

  By using filters and step sizes (shifts) that have been calculated on the feature map, the procedure known as pooling or pooling seeks to reduce the dimensions of the matrix [9]. A number of pooling strategies exist, such as average pooling, which determines the average value while accounting for the input's fluctuating size, and maximum pooling, which chooses the highest value. A quick summary of maximal pooling and its similarities to average pooling are provided below.

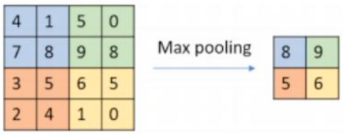

Fig. 2. Pooling Layer

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2104-2117

- Fully Connected Layer

  The goal of fully linked layers, which are frequently utilized in MLPs, is to alter the dimensionality of the data to enable linear classification. The neurons in the convolutional layer are sent to the fully connected layer once they have produced one-dimensional data. Spatial information in the data is permanently lost during this operation. This completely connected layer is limited to use at the network's termination.

The Support Vector Machine (SVM) method will explain how this SVM kernel & parameters can run. Support Vector Machine (SVM) is the model utilized in this study to classify sentiment on the prepared dataset. The way SVM operates is by identifying the best hyperplane to divide data into different classes.

- Kernel

  This SVM model makes use of the Radial Basis Function (RBF) kernel. The RBF kernel was selected due to its general good performance in classification issues and its ability to handle data that is not linearly separable. The RBF kernel is an excellent option for complicated datasets since it takes into account all dimensions when calculating the distance between two locations in feature space.

- Parameter C

  The C parameter in an SVM controls how much the model tolerates misclassification. In this study, the C value was set at 1.0. This value was chosen to maintain a balance between finding the optimal separation margin and preventing overfitting. With this value of C, the model penalizes classification errors but still provides a certain tolerance for errors in the training data, allowing wider margins on the hyperplane.

TABLE I.
OVERVIEW OF RESEARCH METHODOLOGY STAGES

Research Stage	Description
1. Dataset	The dataset used comes from crawling results on 3 popular online media such as Detik.com, CNN Indonesia and CNBC Indonesia with a total of 3200+ data.
2. Data Preprocessing	Cleaning and preparing the text for analysis with steps such as attribute selection, Cleaning, normalization, stopwords, Tokenize and Stemming data.
3. Data Labeling	Labeling the data using star score labeling.
4. Convolutional Neural Network (CNN) and The Support Vector Machine (SVM) method	Modeling and visualization of data that displays sentiment results (neutral, positive and negative), displays sentiment distribution graphs, displays words that often appear in the data and displays positive, negative and neutral words using CNN and SVM methods
5. Evaluation	Evaluating the model in this comparison of results, a comparison of the precise accuracy results obtained by the Convolutional Neural Network (CNN) method and the accuracy obtained by the Support Vector Machine (SVM) was carried out to achieve consistent results.

Table I summarizes the key stages of the research methodology. The detailed processes that include these steps can be seen in Figure 1.

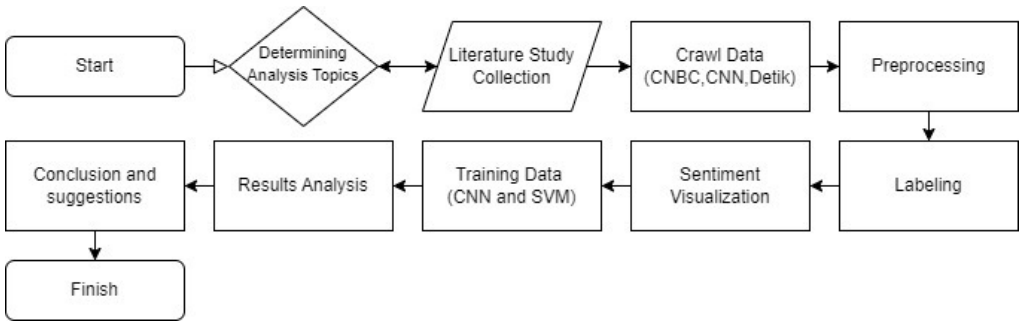

Fig. 2. System Flow of the Processes

A. *Dataset*

Crawling data taken from 3 online media (Detik.com, CNN Indonesia and CNBC Indonesia) with the keywords

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2104-2117

"2024 election and the name of each candidate pair". Crawling data is taken in text form which is the title of the news. The data taken was 3,215 from 3 online media after going through the data processing stage, amounting to 1,594 data/Online Media containing news on 3 election candidate pairs discussed in 3 online media. Some examples of datasets that are the result of data scraping can be seen in Table I.

TABLE II
RESEARCH DATASET

Media Online		Content	Candidate
Detik.com	-	Abu Bakar Ba'asyir supports Anies-Cak Imin in the 2024 presidential selection. (Abu Bakar Ba'asyir mendukung Anies-Cak Imin di Pilpres 2024)	1
	-	Indicator Survey: Prabowo-Gibran 48.55%, AMIN 24.17%, Ganjar-Mahfud 21.60% (Indicator Survey: Prabowo-Gibran 48.55%, AMIN 24.17%, Ganjar-Mahfud 21.60%)	2
	-	The moment when Ganjar-Mahfud were given the Wisanggeni puppet and the Semar puppet. (The moment when Ganjar-Mahfud were given the Wisanggeni puppet and the Semar puppet.)	3
CNN Indonesia	-	Anies asks supporters to guard the polling stations and guard the vote count (Anies asks supporters to guard the polling stations and guard the vote count)	1
	-	VIDEO: Prabowo Accepts NasDem Joining the Coalition. (VIDEO: Prabowo Accepts NasDem Joining the Coalition.)	2
	-	The Ganjar-Mahfud Legal Team Still Can't Believe They Lost in Bali and Central Java. (Tim Hukum Ganjar-Mahfud Masih Tak Percaya Kalah di Bali dan Jawa Tengah)	3
CNBC Indonesia	-	Anies wants to challenge the 2024 election results at the Constitutional Court, Prabowo prepares ammunition. (Anies Mau Tantang Hasil Pemilu 2024 ke Mahkamah Konstitusi, Prabowo Siapkan Amunisi)	1
	-	28 Provinces Completed, Prabowo Wins in West Java to Aceh! (28 Provinsi Rampung, Prabowo Menang di Jabar hingga Aceh!)	2
	-	Prabowo Wins Quick Count, Ganjar: So Many Election Violations! (Prabowo Menang Quick Count, Ganjar: Banyak Pelanggaran Pemilu!)	3

B. Data Preprocessing

Data preprocessing has a significant role in text data, the aim is to standardize the format and clean the data from some noise for effective feature extraction. Several processes used in preprocessing this data include datasets, normalization, stopwords, tokenization and data stemming. in Figure 2.

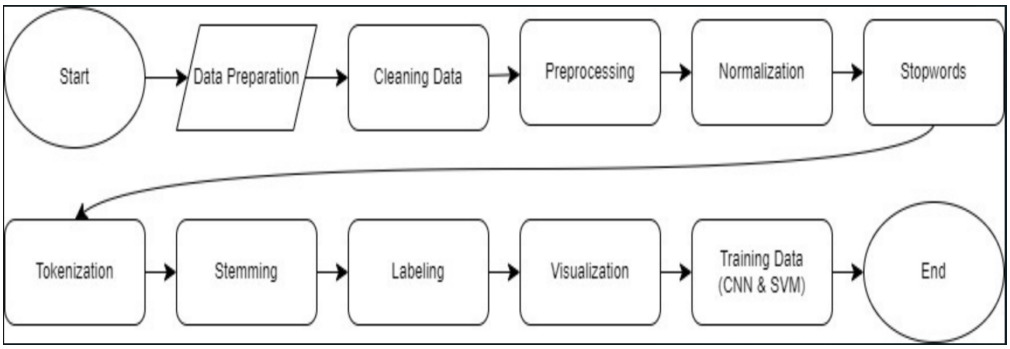

Fig. 3. Preprocessing Process

C. Data Labeling

The subsequent step is data labeling, which comes after data preprocessing. Data labeling is Sentiment class labeling of the review data is required because the machine learning classification procedure is still a part of the

supervised classification process. Generally speaking, labeling Three classes can be formed from the sentiment analysis categorization. There are three types of sentiments: neutral, negative, and positive.

It is necessary to label sentiment classes for review data. In general, labeling The classification carried out by sentiment analysis can be divided into three classes sentiment, namely positive sentiment, neutral sentiment and negative sentiment. Based on the ratings given by users, you can get a general idea of the user's assessment of online media. Therefore, data containing Hate Speech Text will be entered into the "Negative" category, while data that does not contain Hate Speech Text based on the TextBlob Library will be entered into the "Positive" category outside the *TextBlob* or *BlankText* Library filter will be entered into the "Neutral" category.

TABLE III
DATA LABELING ONLINE MEDIA RESULT

CANDIDATE	Detik.com			CNN Indonesia			CNBC Indonesia		
	1	2	3	1	2	3	1	2	3
Positive	151	172	149	156	94	108	158	173	162
Negative	219	199	218	203	278	261	148	139	136
Neutral	23	14	23	41	26	31	41	32	33

D. Convolutional Neural Network (CNN)

On Convolutional Neural Network is one of the deep learning algorithms usually used to classify image, can be used for identify faces, past signs cross, tumor, and various aspects of the data visual. CNN consists of three main layers that are connected to each other for processing entered data. Extraction layer features consist of convolution layers and pooling layer. These two layers used to process and learn features on dataentered. Next layer is a fully connected layerused to perform classification on input data [6]. Formation CNN algorithm model that has a split 80:20 data amounts to 3200+ data. Then 500 test data will be entered data where it will be carried out classification by providing hyperplane results to separate space from two classes, which is simple This method is a grouping directionally (linear classifier) and then developed to become capable to work with non-linear problems [2].

E. Support Vector Machine (SVM)

Support vector machine is one of the machine learning algorithms gives better results than other machine learning algorithms in perform classification, regression, and predictions. In carrying out SVM classification works by finding the hyperplane Best, this is helped by settings c parameters that work forcontrols hyperplane search [2]. After labeling, use lexicon based on train data, the data will be included in the formation SVM algorithm model that has a split 80:20 data amounts to 1500 data. Then 500 test data will be entered data where it will be carried out classification by providing hyperplane results to separate space from two classes.

F. Evaluation

One step in the system flow is evaluation, which tests the predicted data in order to evaluate the CNN & SVM sentiment classification model's performance and identify the optimal algorithm. Accuracy and F1-score are two metrics we used to assess the effectiveness of the Naïve Bayes sentiment classification model, which are obtained from the confusion matrix. Additionally, we used 10-fold cross-validation to guarantee the consistency and dependability of our model's performance. The dataset was split into ten equal pieces for this operation. To ensure that every part had an equal opportunity to be the test set, the model was then trained on nine parts and tested on the remaining part. This process was performed ten times. After that, performance measures were determined by averaging.

Confusion Matrix is an evaluation method used to evaluate object classification performance. Table is a simple illustration that presents the results of the classification to make understanding easier regarding the confusion matrix in evaluating or testing classification models in the table above there are several variables that have a role in carrying out the evaluation process of a model classification. Each variable's explanation is as follows: the True Negative (TN) variable represents data that the system successfully classifies as having an incorrect or negative value, and the True Positive (TP) variable represents data that the system successfully classifies as having a correct or positive value. Subsequently, data with a legitimate value of true or positive but an unsuccessful classification by the system is classed as False Positive (FP) variables [10]. Finally, data with an unsuccessful classification by the system is classified as False Negative (FN). Table V contains the structure of the confusion matrix.

TABLE IV.
CONFUSION MATRIX

Predicted	Actual	
	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

To evaluate the performance of the CNN & SVM sentiment classification model, several metrics will be used including accuracy, precision, recall, and F1-score. The following are the formulas for the metrics that will be used.

1) *Accuracy*

Accuracy measures the ratio of the number of correct predictions to the overall data. The formula for calculating Accuracy is found in formula (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

2) *Precision*

Precision measures the ratio of correct positive predictions to all positive detection results. The formula for calculating Precision is in formula (3).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

3) *Recall*

Recall measures the ratio of correctly predicted positive data. The formula for calculating Recall is found in formula (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

4) *F1-Score*

F1-Score is the harmonic mean of precision and recall. The formula for calculating F1-Score is found in formula (5).

$$F1\ Score = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right) \quad (5)$$

III. Result and Discussion

A. Result

In this chapter, testing has been conducted for feature extraction, feature combination, CNN & SVM model training for 2024 of presidential election sentiment classification on Media Online (Detik.com, CNN Indonesia & CNBC Indonesia) with CNN & SVM approaches. The class frequencies for the data labeling have balanced data. Then, feature extraction and feature combination for model training are performed.

The three model scenarios of the CNN & SVM model were executed aims to compare and evaluate the effectiveness of different feature extraction methods in sentiment analysis and Additionally, we used 10-fold cross-validation to guarantee the consistency and dependability of our model's performance. The dataset was split into ten equal pieces for this operation. To ensure that every part had an equal opportunity to be the test set, the model was then trained on nine parts and tested on the remaining part. This process was performed ten times. After that, performance measures were determined by averaging.

1. Analysis Result method Convolutional Neural Network (CNN)

- Analysis Sentiment by Media Online (Detik.com)

  On data classification modeling Candidate 1, Candidate 2 and Candidate 3 by using CNN Algorithm, once done division of the dataset into train data and test data with a ratio of 80:20 for the performance of classification results with using CNN method on Media Online: Detik.com is interpreted inward following table. Resulting accuracy for data classification for candidate par with using the CNN algorithm for Candidate 1 average accuracy of 57% with a F-Measure of 41%. Accuracy results of 57% incl into the fair classifier category, for Candidate 2 average of 58% with a F-Measure of 58%. Accuracy results of 58% incl into the fair classifier category, for Candidate 3 average of 62% with a F-Measure of 52%. Accuracy results of 62% incl into the good classifier category.

- Analysis Sentiment by Media Online (CNN Indonesia)

  On data classification modeling Candidate 1, Candidate 2 and Candidate 3 by using CNN Algorithm, once done division of the dataset into train data and test data with a ratio of 80:20 for the performance of classification results with using CNN method on Media Online CNN Indonesia is interpreted inward following table. Resulting accuracy for data classification for candidate par with using the CNN algorithm for Candidate 1 average of 71% with a F-Measure of 66%. Accuracy results of 71% incl into the good classifier category, for Candidate 2" average of 78% with a F-Measure of 68%. Accuracy results of 78% incl into the good classifier category, for Candidate 3 average of 69% with a F-Measure of 56%. Accuracy results of 69% incl into the good classifier category.

- Analysis Sentiment by Media Online (CNBC Indonesia)

  On data classification modeling Candidate 1, Candidate 2 and Candidate 3 by using CNN Algorithm, once done division of the dataset into train data and test data with a ratio of 80:20 for the performance of classification results with using CNN method on Media Online CNBC Indonesia is interpreted inward following table. Resulting accuracy for data classification for candidate par with using the CNN algorithm for Candidate 1 average of 56% with a F-Measure of 52%. Accuracy results of 56% incl into the fair classifier category, for Candidate 2 average of 72% with a F-Measure of 43%. Accuracy results of 72% incl into the good classifier category, for Candidate 3 average of 66% with a F-Measure of 61%. Accuracy results of 66% incl into the good classifier category.

2. Analysis Result method Support Vector Machine (SVM)

- Analysis Sentiment by Media Online (Detik.com)

  On data classification modeling Candidate 1, Candidate 2 and Candidate 3 by using CNN Algorithm, once done division of the dataset into train data and test data with a ratio of 80:20 for the performance of classification results with using CNN method on Media Online : Detik.com is interpreted inward following table. Resulting accuracy for data classification for candidate par with using the CNN algorithm for Candidate 1 average of 68% with a F-Measure of 65%. Accuracy results of 68% incl into the good classifier category, for Candidate 2 average of 77% with a F-Measure of 75%. Accuracy results of 77% incl into the good classifier category, for Candidate 3 average of 72% with a F-Measure of 67%. Accuracy results of 72% incl into the good classifier category.

- Analysis Sentiment by Media Online (CNN Indonesia)

  On data classification modeling Candidate 1, Candidate 2 and Candidate 3 by using CNN Algorithm, once done division of the dataset into train data and test data with a ratio of 80:20 for the performance of classification results with using CNN method on Media Online CNN Indonesia is interpreted inward following table. Resulting accuracy for data classification for candidate par with using the CNN algorithm for Candidate 1 average of 74% with a F-Measure of 70%. Accuracy results of 74% incl into the good classifier category, for Candidate 2 average of 81% with a F-Measure of 75%. Accuracy results of 81% incl into the good classifier category, for Candidate 3 average of 78% with a F-Measure of 76%. Accuracy results of 78% incl into the good classifier category.

- Analysis Sentiment by Media Online (CNBC Indonesia)

  On data classification modeling Candidate 1, Candidate 2 and Candidate 3 by using CNN Algorithm, once done division of the dataset into train data and test data with a ratio of 80:20 for the performance of classification results with using CNN method on Media Online CNBC Indonesia is interpreted inward following table. Resulting accuracy for data classification for candidate par with using the CNN algorithm

for Candidate 1 average of 77% with a F-Measure of 74%. Accuracy results of 77% incl into the good classifier category, for Candidate 2 average of 75% with a F-Measure of 73%. Accuracy results of 75% incl into the good classifier category, for Candidate 3 average of 66% with a F-Measure of 62%. Accuracy results of 66% incl into the good classifier category.

3. The model evaluation of three Candidate of CNN & SVM models can be seen in Table VII & Table IX:

TABLE V
CNN MODEL TRAINING RESULT

Mount	Accuracy		Precision		Recall		F1-Score	
	SVM	CNN	SVM	CNN	SVM	CNN	SVM	CNN
Candidate Number 1	73.0%	61,3%	75.0%	49.0%	73.0%	61.3%	69.6%	53.0%
Candidate Number 2	77.6%	69.3%	75.0%	62.3%	77.6%	69.3%	74.3%	64.3%
Candidate Number 3	72.0%	65.6%	67.6%	57.3%	72.0%	65.6%	68.3%	56.3%

4. Evaluation Algoritma Model

The classification process using machine learning is still included In supervised classifications, sentiment class labeling of review data is needed. In general, labeling The classification carried out in sentiment analysis can be divided into three classes Sentiments are positive sentiment, neutral sentiment and negative sentiment. Based on the ratings given by users, you can get a general idea of the user's assessment of one of Detik.com's online media. Therefore, data containing Hate Speech Text will be entered into the "Negative" category, while data that does not contain Hate Speech Text based on the TextBlob Library will be entered into the "Positive" category outside the TextBlob or BlankText Library filter will be entered into the "Neutral" category.

TABLE X
SENTIMENT DATA LABELING REDUCTION RESULTS

CLASS	CNN	SVM
Positive	47.6%	66.3%
Neutral	42.0%	77.8%
Negative	0.0%	10.2%

Table X provides information on the number of reduction results for positive sentiment labeling class positive for CNN, which is 47.6%, neutral for 42%, and negative for 0.0%, and for SVM, which is 66.3%, neutral for 77.8%, and negative for 1-.2%.

5. Visualisasi Wordcloud

Wordcloud is a system that able to visualize words by highlighting the frequency of that word used in a written text and as a tool to do it analysis to obtain information about things that need to be developed and corrected from online media Data applied to wordcloud is the data resulting from the classification of the algorithm CNN because it has better accuracy and f-measure than SVM.

Wordcloud Function by highlighting the words that appear most frequently in the news regarding presidential candidates. By using wordcloud, researchers can quickly identify key words or phrases that appear frequently in a particular context. For example, if negative words are predominant in the resulting wordcloud, this indicates that news about a particular candidate may tend to have a negative sentiment. For example, in this analysis, words like “corruption” or “scandal” may appear more frequently if they are mentioned frequently in the context of negative news. Additional Insights from Wordcloud visualizations provide additional insight into patterns in data that may not be readily apparent from numerical results alone. By visually displaying the frequency of words, wordcloud helps researchers to better understand how CNN models capture and classify context in text. For example, if a wordcloud shows that 70% of the most frequently occurring words relate to negative issues, this could help explain why the model may be classifying

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2104-2117

sentiment as negative more often. Wordcloud can also help in identifying specific words or phrases that may influence sentiment classification, providing a clearer picture of bias or trends in the data.

Overall, wordcloud is an effective tool for uncovering dominant themes in sentiment and providing deeper insight into how data is processed by CNN models. This visualization allows researchers to look further into the sentiment patterns present in the data and supports more comprehensive analysis.

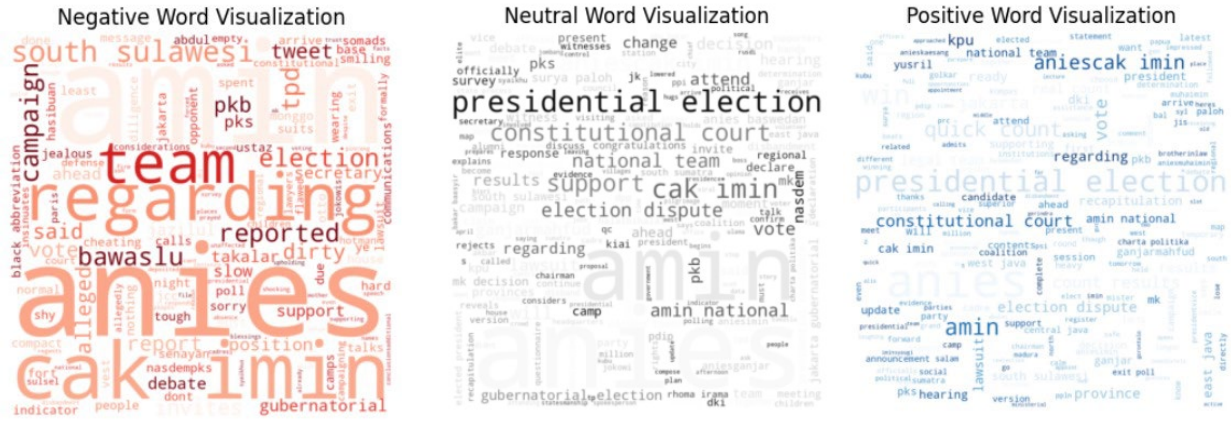

Fig. 3. Candidate 1 World Visualization

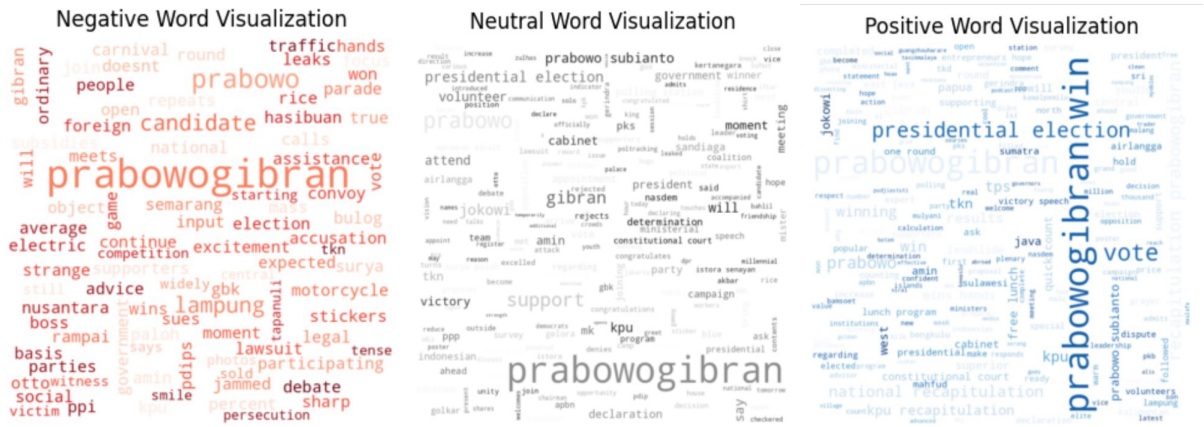

Fig. 4. Candidate 2 World Visualization

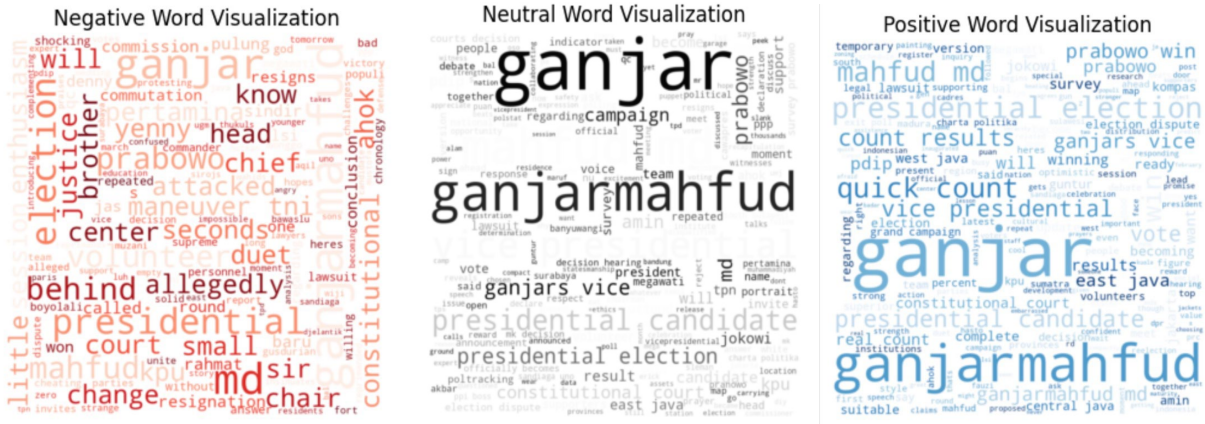

Fig. 5. Candidate 3 World Visualization

B. *Discussion*

Parameters in comparison Support Vector Machine algorithm and Convolutional Neural Networks. Mark Accuracy has a function in measuring the level of truth of the case is moderate identified. While the f-measure is

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2104-2117

is the result of a combination of values precision and recall values used to measure the performance of the system classification. Comparison results of accuracy and f-measure between the two algorithms on each is visualized on following diagram.performance indicators: Accuracy, Precision, Recall, and F1-score.

TABLE VI
CNN MODEL TRAINING RESULT

Mount	Accuracy		F-Measure	
	SVM	CNN	SVM	CNN
Detik.com	72%	59%	69%	50%
CNN Indonesia	78%	65%	74%	78%
CNBC Indonesia	77%	65%	74%	52%

Fig. 5. Accuracy & F-measure Comparison

From the diagram above, it can be concluded that the SVM algorithm has better performance in classifying sentiment in data from the three social media, namely detik.com, CNN Indonesia and CNBC Indonesia. Support Vector Mechine (SVM) is an algorithm that is considered effective in carrying out classification because it has layers that function to analyze data features. In modeling data that has been processed on both platforms. From the results above it can be seen from the evaluation of the CNN and SVM algorithms From this data it is known that the accuracy and f-measure values of SVM are higher than CNN and this is in accordance with the results of the comparison of modeling diagrams above. The parameters in the algorithm comparison have a high function in measuring the level of truth of the case being identified, while the results of each f-measure value will be compared for accuracy and represented through diagrams. If both algorithms know that the Support Vector Machine algorithm and Convolutional Neural Networks are included in the good classifier, then the algorithm for this research is already good at modeling, but if it does not reach a good classifier, new modeling can be done to achieve the right accuracy.

In this research, Convolutional Neural Network (CNN) and Support Vector Machine (SVM) are used as methods to perform sentiment analysis on data from three online media. Although both are powerful classification algorithms, the results obtained show significant differences in their performance.

Performance Comparison CNN tend to excel at handling very complex data, such as visual or text data that requires automatic feature extraction via convolutional networks. In the context of this research, CNN shows good performance in detecting hidden patterns in text, but the results show that SVM gives better performance in terms of overall accuracy. This may be due to the simpler nature of SVM in finding the best hyperplane to separate classes, so it is more effective on datasets that have clear margins between positive, negative and neutral classes.

Effectiveness of SVM may be more effective than CNN on this dataset due to its nature of working well on more structured datasets. SVM utilizes the maximal margin principle to separate data, which is very effective in text classification problems with a smaller number of features and clear margins between categories. Meanwhile, CNNs, although powerful, can experience difficulties when working with text data if the feature structure extracted from the text is not robust enough to provide the expected accuracy.

Another comparison based on the F-Measure or F1-Score is an important metric in this research because it provides a balance between precision and recall. F-Measure is calculated as the harmonic average of precision and recall, making it an ideal metric for assessing model performance in situations where both are important. where it can also be attributed that the accuracy results are based on Precision vs. Recall, Precision measures how many of the model's positive predictions are truly accurate, while recall measures how much of the overall positive data the model successfully detected. In sentiment analysis, precision is important to ensure that positive predictions truly reflect true sentiment. While recall is important to ensure that the model does not miss a lot of data that is actually positive. The Importance of F-Measure In this research, F-Measure is the main metric because it helps capture the trade-off between precision and recall. If precision is high but recall is low, it means the model is ignoring a lot of data that is actually positive, and conversely, if recall is high but precision is low, then the model is giving too many false positive predictions. By using F-Measure, this research can evaluate model performance more comprehensively, ensuring that the model is not only good in one aspect (such as precision or recall) but also provides an optimal balance between the two.

In this research, a comparison was carried out between Convolutional Neural Network (CNN) and Support Vector Machine (SVM) to assess the effectiveness of both methods in classifying sentiment on data from online media. Both have different approaches to processing data and determining classification.

Fundamental Differences in Results: The results show that SVM is superior to CNN, with an average accuracy of 74% and an F1-Score of 70.7%. SVM is able to separate data more effectively due to its ability to find the optimal hyperplane, especially when the data has clear margins between positive, negative and neutral classes. The SVM algorithm is very suitable for more structured text data, where patterns in the text are more consistent and easy to recognize.

CNN Effectiveness: On the other hand, CNN showed lower results with an average accuracy of 65.4% and an F1-Score of 57.8%. CNNs are usually more effective in handling data that has a spatial structure, such as images, where the ability to extract complex features is required. However, in the context of text sentiment analysis, CNN may not be as effective as SVM, especially if the data is not varied enough or if the semantic context is difficult to extract. Although CNNs can capture contextual information from text, their performance may suffer if the features are irrelevant or difficult for the convolutional network to recognize.

From these results, it can be concluded that SVM is more effective in sentiment classification for this dataset due to its better ability to handle text data with clear patterns, whereas CNN, although powerful in other contexts, may be less optimal for this task with the available data.

In this research, Convolutional Neural Network (CNN) and Support Vector Machine (SVM) are used to analyze sentiment from online news data related to the 2024 Election. Both of these methods have their respective advantages and disadvantages which influence their effectiveness in the context of sentiment analysis.

Advantages of CNN: CNN has the advantage of extracting complex features from data that has a spatial structure, such as images or text with rich context. One of the main advantages of CNNs is their ability to automatically recognize patterns in data through convolution layers, making them very effective in capturing contextual and semantic information. In sentiment analysis, CNNs can help identify emotional nuances or language patterns that other methods may not capture. However, the effectiveness of CNNs is highly dependent on the quality and quantity of training data. If the dataset lacks variety or has irrelevant features, CNN performance can suffer, which is reflected in this study with an average accuracy of 65.4%.

Disadvantages of CNN: Although CNN has deep capabilities for capturing information, one of its disadvantages is the risk of overfitting, especially if the dataset used is not large enough or is not well structured. CNNs also require greater computing resources and longer time for training, especially on text data that does not have a clear spatial structure. In this case, the CNN provides an F1-Score of 57.8%, which shows that although it captures some contextual information, the final result is still below that of SVM.

Advantages of SVM: SVM is simpler and more effective in handling data with clear margins between classes. SVM has the ability to find optimal hyperplanes that clearly separate different classes, which makes it very suitable for text data that is structured and has consistent patterns. In this research, SVM shows superior performance with an accuracy of 74% and an F1-Score of 70.7%, indicating that SVM is better able to handle sentiment classification on more structured text data.

Disadvantages of SVM: Although SVM is effective in the classification of structured data, this method may be less capable of handling data that has many features or high complexity, such as text with complex emotional nuances or image data. SVM is also less flexible in capturing complex non-linear relationships between features in the data. In situations where the data has many interactive features, such as text containing sarcasm, SVM may not be as effective as CNN.

Overall, the choice between CNN and SVM largely depends on the nature of the data being analyzed. CNN excels at capturing contextual and semantic information from complex data, while SVM is more effective at classifying data that is clearly structured and has clear margins between classes. In the context of this research, SVM is proven to be superior because the data analyzed is better suited to simpler and more efficient approaches such as SVM. In this research, the results obtained show that Support Vector Machine (SVM) is superior in classifying sentiment in Twitter data related to the 2024 Election compared to Convolutional Neural Network (CNN) and Naive Bayes. To provide further context and emphasize the contribution of this study, the results can be compared with related studies that have been conducted previously.

In a previous study by Mahbubah and Zuliarso (2019), the Naive Bayes Classifier was used to analyze sentiment regarding the 2019 Indonesian Presidential Election on Twitter. The results of this study show that Naive Bayes achieves an accuracy of 73% with a precision for the negative class of 78% and for the positive class of 66%. These results show that Naive Bayes can classify sentiment with a fairly high level of accuracy, but there are limitations in handling complex nuances in the data, which can be seen from the relatively lower results for the positive class precision.[4]

Our research shows that SVM, with an accuracy of 74%, provides better results compared to CNN which only achieves an accuracy of 65.4%. SVM also excels at handling clear margins between positive and negative sentiment classes, allowing for more precise and consistent classification. These results are consistent with research by You

et al. (2017), which shows that SVM has strong capabilities in handling well-structured data and provides high accuracy in classification.[1]

On the other hand, research conducted by Sartini (2020) shows that CNN can achieve accuracy of up to 81.4% in sentiment analysis on Indonesian language Twitter data. However, in the context of our study, CNN showed lower performance, which could be caused by the complexity of poorly structured data or by limitations in fitting model parameters. This highlights the importance of choosing a method that suits the characteristics of the data being analyzed.[2]

By comparing the results of this research with previous studies, it can be seen that SVM is more suitable for sentiment classification in the context of the analyzed Twitter data. This research contributes by showing that although CNNs have great potential in various data analysis applications, SVMs are still a stronger choice in situations where the data is more structured and requires precise classification. In addition, this research enriches the literature by providing empirical evidence that supports the effectiveness of SVM in political sentiment analysis on social media.

IV. CONCLUSION

The sentiment analysis of online media for the 2024 election can be applied by separating the three candidates frequently discussed in the three online media platforms. This involves preprocessing, classification of online media using the Support Vector Machine (SVM) and Convolutional Neural Network (CNN) methods, word cloud visualization, and word association for sentiment information retrieval. This study analyzes the sentiment of online media (detikcom, CNN Indonesia, CNBC Indonesia) related to the 2024 Indonesian election, focusing on presidential and vice-presidential candidates, using CNN and SVM methods. Our findings indicate that the SVM model achieved the highest performance compared to the CNN method, with an SVM accuracy of 74.2%, precision of 72.5%, recall of 74.2%, and F1 score of 70.7%. This superior performance can capture a broader range of sentiment indicators, resulting in better sentiment identification.

The CNN model also performed in the fair classifier category with a CNN accuracy of 65.4%, precision of 56.2%, recall of 65.4%, and F1 score of 57.8%. This suggests that CNN is fairly effective in capturing contextual information from the text, contributing to a reasonably accurate sentiment classification, but its results are not on par with the SVM model. This indicates that the predefined sentiment dictionary may not capture contextual nuances as effectively as SVM or a combined approach. Although both algorithms are classified as good classifiers, SVM consistently demonstrates superior results in identifying and classifying sentiments from the analyzed data. Therefore, SVM is considered more effective in sentiment classification for the context of this study. The authors' research shows a significant improvement in model performance compared to some previous studies.

The sentiment classification results and word association results provide information that can be retained from positive online media articles and aspects that need improvement from negative articles across the three online media platforms. The developed model has the highest prediction accuracy in the negative and positive classes. In future research, it is recommended to use a larger dataset with a balanced comparison for each label (positive, negative, and neutral), which would increase the accuracy in sentiment analysis classification. Additionally, for future research, it is suggested to use higher-level labeling to reduce noise and improve dataset quality by using a typo detection algorithm.

REFERENCES

[1] A. A. Firdaus, A. Yudhana and I. Riadi, "Analisis Sentimen Pada Proyeksi Pemilihan Presiden 2024 Menggunakan Metode Support Vector Machine," *https://tif.uad.ac.id/analisis-sentimen-pada-proyeksi-pemilihan-presiden-2024-menggunakan-metode-support-vector-machine/,* p. 1, 2024.

[2] A. Perdana, A. Hermawan and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier," *https://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/1412/0,* p. 1, 2022.

[3] M. M. Ismail and K. M. Lhaksamana, "Sentimen Analisis Pada Media Online Mengenai Pemilihan Presiden 2019 Dengan Menggunakan Metode Naive Bayes," *e-Proceeding of Engineering : Vol.6, No.2 Agustus 2019,* p. 9997, 2019.

[4] L. D. Mahbubah and E. Zuliarso, "ANALISIS SENTIMEN TWITTER PADA PILPRES 2019 MENGGUNAKAN ALGORITMA NAIVE BAYES," *16_ANALISA_SENTIMEN_TWITTER_PADA_PILPRES_2019_MENGGUNAKAN,* p. 198, 2019.

[5] N. A. Salsabila, "ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH GUSDUR MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE," *NUR ADINDA SALSABILA-FST,* p. 3, 2022.

[6] S. "ANALISIS SENTIMEN TWITTER BAHASA INDONESIA MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK," *5302416006 - Sartini Unnes,* pp. 3 - 12, 2020.

[7] A. Alwasi'a, "ANALISIS SENTIMEN PADA REVIEW APLIKASI BERITA ONLINE MENGGUNAKAN METODE MAXIMUM ENTROPY," *14611247 Annisa Alwasi'a,* p. 14, 2019.

[8] A. Alwasi'a, "ANALISIS SENTIMEN PADA REVIEW APLIKASI BERITA ONLINE MENGGUNAKAN METODE MAXIMUM ENTROPY," *14611247 Annisa Alwasi'a,* pp. 18 - 19, 2019.

[9] W. You, C. Shen, X. Guo, X. Jiang, J. Shi and Z. Zhu, "A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery," *Advances in Mechanical EngineeringVolume 9, Issue 6, June 2017, Advanced signature analysis for machine fault diagnosis and prognosis - Research Article,* pp. 463 - 473, 2017.

[10] R. FIRMANSYAH, "IMPLEMENTASI DEEP LEARNING MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI BUNGA," *https://repository.uinjkt.ac.id/dspace/bitstream/123456789/55347/1/REZKY%20FIRMANSYAH-FST.pdf,* pp. 19 - 28, 2021.

[11] S. H. Kusumahadi, H. Junaedi and S. Joan, "Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT), Vol.04, No.01, Januari 2019,* pp. 54-60, 2019.

[12] R. W. Pratiwi, D. D. Apriliani, D. I. Af'idah, S. F. Handayani, K. A. Gian and T. I. Saputra, "PENINGKATAN KETRAMPILAN DIGITAL ENTERPRENEUR BAGI SISWA SISWI SMK," *JMM (Jurnal Masyarakat Mandiri) http://journal.ummat.ac.id/index.php/jmm Vol. 6, No. 3, Juni 2022, Hal. 1746-1753,* pp. 40 - 46, 2022.

[13] S. F. Pane, A. Owen and C. Prianto, "Analisis Sentimen UU Omnibus Law pada Twitter menggunakan Metode Support Vector Machine," *InComTech: Jurnal Telekomunikasi dan Komputer vol.11, no.2, Agustus 2021 http://publikasi.mercubuana.ac.id/index.php/Incomtech,* p. 130, 2021.

[14] G. T. SANTOSO, "ANALISIS SENTIMENM PADA TWEET DENGAN TAGAR #BPJSRASARENTENIR MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)," *https://repository.uir.ac.id/16772/,* p. 11, 2021.

[15] A. K. A and M. M, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," *J. Inform. Univ. Pamulang, vol. 5, no. 4, p. 544, 2021, doi: 10.32493/informatika.v5i4.6760.,* 2021.

[16] M. R. L, U. S and S. D, "Studi Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network," *Semin. Nas. Teknol. Inf. dan Komun. STI&K, vol. 3, 2019..*

[17] M. D. Alizah, A. Nugroho, U. Rodiyah and W. Gota, "Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter," *IJSE – Indonesian Journal on Software Engineering, Vol.6, No.2, Desember 2020,* pp. 223 - 229, 2020.

[18] N. A. Salsabila, "ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH GUSDUR MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE," *ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP TOKOH GUSDUR MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE,* p. 11, 2022.

[19] W. Hidayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 5, Nomor 3, Juli 2021,* p. 1018, 2021.

[20] G. N. Arviana, "Sentiment Analysis, Teknik untuk Pahami Maksud di Balik Opini Pelanggan," 2021.

[21] Davies and P. Beynon, "Database Systems Third Edition," *New York: Palgrave Macmillan.,* 2004.

[22] I. Pramudiono, "Pengantar Data Mining:Menambang Permata Pengetahuan di Gunung Data," *Kuliah Umum IlmuKomputer.ComCopyright © 2003 IlmuKomputer.Com,* p. 2.

[23] T. Larose, "Discovering Knowledge in Data: an Introduction To Data Mining," *San Fransisco: John Wiley & Sons Inc,* p. xvi, 2014.

[24] A. Albadarneh and A. Ahmad2, "Automated Flower Species Detection and Recognition from Digital Images," *IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.4, April 2017,* p. 144, 2017.

[25] H. Almogdady, D. S. Manaseer and D. Hiary, "A Flower Recognition System Based On Image Processing And Neural Networks," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 7, ISSUE 11, NOVEMBER 2018,* pp. 166 - 173, 2018.

[26] M. -E. Nilsback and A. Zisserman, "A Visual Vocabulary for Flower Classification," *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing,* pp. 722-729, 2008.

[27] M. -E. Nilsback and A. Zisserman, "A Visual Vocabulary for Flower Classification," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 06) (Vol. 2).,* pp. 1447-1454, 2006.