Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

2024 Presidential Election Sentiment Analysis in News Media Using Support Vector Machine

Bayu Muhammad Iqbal*, Kemas Muslim Lhaksmana, Erwin Budi Setiawan

School of Computing, Telkom University, Bandung, Indonesia
Email: [1*]bayuiqballl@student.telkomuniversity.ac.id, [2]kemasmuslim@telkomuniversity.ac.id, [3]erwinbudisetiawan@telkomuniversity.ac.id
Correspondence Author Email: bayuiqballl@student.telkomuniversity.ac.id
Submitted: **07/01/2023;** Accepted: **27/02/2023;** Published: **28/02/2023**

Abstract−The 2024 presidential election is an event for all Indonesian people to determine their best leader. The presidential and vice presidential candidates are also competing to give their best efforts so that they can be elected as President and Vice President. The news media also provide news related to the 2024 presidential election with various titles that can interest their readers. Not infrequently the titles raised contain words that have sentiments, both positive and negative. In order to facilitate the analysis of the sentiments of these news titles, it is necessary to build a system that can detect the sentiments of these titles. In this study, we built a sentiment analysis system using the Support Vector Machine (SVM) method on news headline data obtained from online news media to detect whether news headlines contain positive or negative sentiments. For feature exctraction we compare the effect of FastText word embedding with TF-IDF for feature extraction. In the SVM method, several experiments were carried out on the attributes of C, kernel, gamma, and the ratio of the test data. The results obtained are a FastText can outperform TF-IDF for feature extraction. Also, combination of Kernel, C, and gamma values that give the best accuracy score of rbf, 1, and auto respectively at a test data ratio of 90:10, with an accuracy score of 99%.

Keywords: Sentiment Analysis; Presidential Election 2024; Online News; News Headlines; Support Vector Machine

1. INTRODUCTION

The presidential election in Indonesia is a once-every-five-year event that is attended by all Indonesians who already have rights under the provisions of the laws and regulations. General elections have been held several times throughout Indonesia's history. However, the first general elections in which the people of Indonesia directly participated were held during the reform era, specifically in 2004 [1]. Many issues have begun to spread on social media to the news media, both online and in print, during the presidential election preparation period, which is approximately 1 to 2 years before the presidential election. News is published in the news media by creating headlines that entice readers to read them. However, there is no filter in place to determine which news titles contain positive or negative sentiments [2]. With so many news stories, manually determining the sentiment in the headlines is difficult [3].

The news media is a medium that provides information about current events, both online and in printed form. Online news media is news information that is accessed online through a website [3]. The media plays an important role in an election event by presenting relevant news that can impact the electability of the presidential and vice-presidential candidate pairs. The headline on online news media is designed to be appealing to entice readers to access the news page and read it.

Sentiment analysis is a type of text that involves extracting information from a text in order to identify the sentiments contained in it [4]. Sentiment analysis is a branch of Natural Language Processing (NLP) that determines whether a text contains positive or negative sentiments [5]. Text data can be processed using sentiment analysis to determine the sentiment contained in a news title. Because of the volume of news about the 2024 presidential election, it is critical to conduct sentiment analysis on news titles.

During the previous presidential election cycle, research on sentiment analysis related to Indonesian presidential elections was conducted. The Support Vector Machine (SVM) approach was used in study [6] to do sentiment analysis on tweets on Twitter related to the 2019 Presidential Election. According to the results, the SVM method achieved an accuracy score of 91.5%. Additionally, Digna et al. in [7] developed a sentiment analysis system for the 2019 Presidential Candidates on Twitter using the Support Vector Machine (SVM) method. To cope with data that is not linearly separated, they incorporated the kernel function to SVM. In accordance with the results, the SVM model with kernel functions can give an accuracy score of 86%. Research [8] by Aditya et al., built a sentiment analysis system about the 2019 Presidential Election in online news media using a Support Vector Machine. The results obtained are that the SVM method can provide an accuracy score of 93.35%.

Based on the related studies above, it can be concluded that the SVM method can provide good performance on sentiment analysis systems on Indonesian language text data, in this case on the topic of Presidential Elections in Indonesia in online news media. Therefore, in this study we built a sentiment analysis system on the topic of the 2024 Presidential Election in online news media using the Support Vector Machine (SVM) method.

2. RESEARCH METHODOLOGY

2.1 Data Collection

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

To get news title data, scrapping is done on news media websites. The news media used for the scrapping stage is the CNN Indonesia news website with the search keyword "Pemilihan Presiden Indonesia 2024". After the scrapping stage was carried out, 2360 news title data were obtained about the 2024 Presidential Election from 2020 to 2022.

2.2 Labelling

In the labeling stage, we do this manually by labeling 0 and 1, where 0 represents a news title with negative sentiment, and 1 represents a news title with positive sentiment. Sentiment labeling in news headlines is done by paying attention to the use of the words. For example, the headline 'Fahri Hamzah Sindir Koalisi Sebelum Pemilu: Bisa Dibilang Tiket Palsu. (Fahri Hamzah Satire the Coalition Before the Election: You could say it's a fake ticket.).' is labeled as a headline with negative sentiment. Example of the dataset in this study is shown in Table 1.

Tabel 1. Example Dataset

Text	Label
Dinamika Politik 2022: Poros Perubahan Anies hingga Dugaan KPU Curang	0
Mahfud di Gereja Yogyakarta: Malam Natal 2022 di RI Aman dan Lancar	0
Survei Charta: Elektabilitas Ganjar Jauh Tinggalkan Anies dan Prabowo	1
NasDem Ingatkan KPU Cegah Menteri Ingin Nyapres Kampanye Terselubung	0
Jokowi Angkat Suara soal 'Kekuatan Besar' Intervensi Pemilu	0

2.3 Preprocessing

The preprocessing stage is carried out to prepare the data to get good quality before being processed in the following stages. The preprocessing stage carried out is Text Cleaning, Tokenization, Stopword Removal, Stemming. The text cleaning stage is carried out to change each letter in the text data to lowercase form, removing punctuation marks and other special characters. The tokenization stage is carried out to change sentences in text data into word tokens. The stopword removal stage is carried out to remove or delete words that appear frequently but do not represent the intent of the sentence. And the stemming stage is carried out to change each word to its basic form. An example of input and output at the preprocessing stage is shown in Table 2.

Tabel 2. Example Dataset

Stage	Input	Output
Text Cleaning	Jokowi: Nanti Ada Capres Tak Dapat Kendaraan Politik, Tuduh Presiden	jokowi nanti ada capres tak dapat kendaraan politik tuduh presiden
Tokenization	jokowi nanti ada capres tak dapat kendaraan politik tuduh presiden	['jokowi', 'nanti', 'ada', 'capres', 'tak', 'dapat', 'kendaraan', 'politik', 'tuduh', 'presiden']
Stopword Removal	['jokowi', 'nanti', 'ada', 'capres', 'tak', 'dapat', 'kendaraan', 'politik', 'tuduh', 'presiden']	['jokowi', 'capres', 'kendaraan', 'politik', 'tuduh', 'presiden']
Stemming	['jokowi', 'capres', 'kendaraan', 'politik', 'tuduh', 'presiden']	['jokowi', 'capres', 'kendara', 'politik', 'tuduh', 'presiden']

2.4 TF-DF Feature Extraction

Feature extraction is performed to convert text data into numbers, so that each word can have a weight that can be understood and processed by computers [9]. In TF-IDF, text is represented in the form of a vector space model so that it can be processed. In the vector space model, each vector has a value equal to the length of the vocabulary formed from the words contained in the dataset. [9]. TF-IDF is one of the most frequently used methods for feature extraction. TF-IDF is the combination of TF (Term Frequency) and IDF (Inverse Document Frequency). The frequency of occurrence of a word is represented by TF, while the number of documents containing a word is represented by IDF. IDF also denotes the importance of a word in a document [10]. TF-IDF is defined as the formula in equation (1).

$$\mathrm{w_{t,d}} = \mathrm{tf_{t,d}} \times \log\left(\frac{\mathrm{D}}{\mathrm{df_t}}\right) \quad (1)$$

Where $\mathrm{w_{t,d}}$ is the TF-IDF weight of term t for document d, $\mathrm{tf_{t,d}}$ is the number of occurrences of term t in document d, D is the number of total documents, and $\mathrm{df_t}$ is the number of documents containing the term t.

2.4 FastText Word Embedding

FastText is the result of a Word embedding project worked on by the Facebook research team. FastText is a fast and lightweight method for creating word representation and text classification. FastText is a library for learning word representations and performing text classification tasks. It works by using a combination of traditional word embeddings and character-level representations, which allows it to handle out-of-vocabulary words effectively.

FastText starts by breaking words into subwords or character n-grams, and then learning vector representations (embeddings) for these subwords. These embeddings are combined to form the representation of the whole word. The embeddings of the subwords capture information about the word's morphological structure, which is important for languages with many rare or out-of-vocabulary words [11]. Once the embeddings have been learned, FastText can be used for various text classification tasks, such as sentiment analysis or language identification. The text is represented as the average of the embeddings of the words it contains. A classifier, such as a linear classifier or a deep neural network, is then trained on this text representation to predict the class label.

FastText's architecture is designed to be computationally efficient and handle large datasets, making it a popular choice for text classification tasks in real-world applications. Additionally, FastText has been shown to outperform traditional word embeddings in many text classification tasks, especially when the training data is limited [12].

2.5 Support Vector Machine

SVM is one of the most powerful classification algorithms. SVM is also an algorithm that is suitable to be applied to structured and unstructured such as text and images [10]. SVM is a supervised algorithm that divides data into a number of classes by calculating the hyperplane's maximum margin. SVM is a learning technique that seeks for the optimal hyperplane in the input space using the Structural Risk Minimization (SRM) concept [13]. As shown in Figure 1, the best hyperplane of the two data is the hyperplane with the greatest margin. The margin is the space between each class's closest pattern and the hyperplane. Support vectors are used to describe the closest pattern. The sample data point from each class that is most similar to the other classes is referred as a support vector. The algorithm computes the margin to identify the highest margin after receiving the support vector [13].

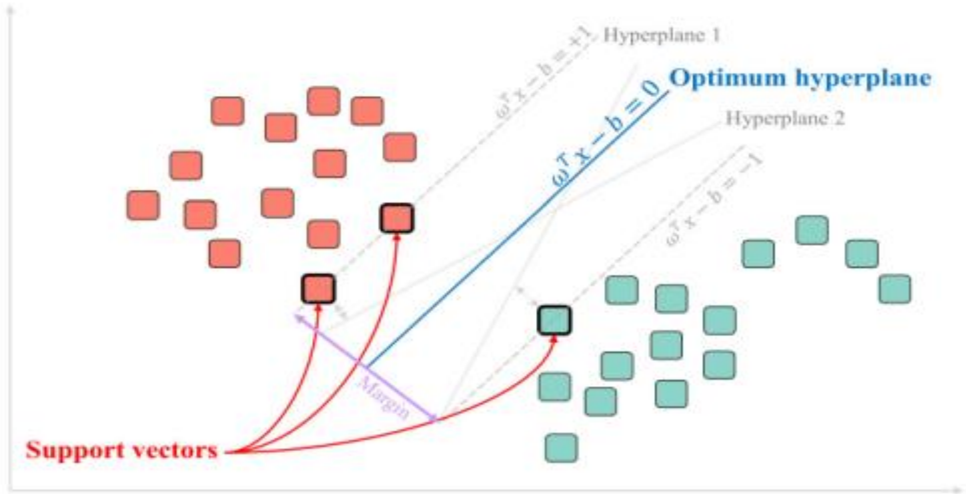

Figure 1. Optimum Hyperplane of SVM [14]

The SVM algorithm works with the kernel namely linear and RBF. With a linear kernel sample data will be separated by a hyperplane linearly. Data with high dimensions are projected in large dimensions so that sample data can be separated linearly [14]. The RBF kernel is a nonlinear kernel on SVM that works well on data with high complexity [14].

In each kernel there are 2 parameters used, namely C and gamma. Parameter C works as a regularization parameter that trades off the correct classification results from the training data to maximization on the margins of the decision function. For larger values of C, a relatively small margin would be accepted if the decision function performs better at detecting and classifying all training points. A lower C decreases training accuracy but supports a greater margin and a simpler decision function [15]. With low values signifying "far" and high values signifying "close," the gamma parameter describes how far an individual training example's influence spreads. The inverse of the radius of influence of the samples picked by the model as support vectors can be thought of as the gamma parameters [15].

2.6 Evaluation

In this study, the accuracy score and F1-Score were used to evaluate system performance. System accuracy is obtained from the TP (True Positive) and FN (False Negative) values of the confusion matrix. The system accuracy score can be calculated using the formula in equation (2).

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

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

The F1-Score value can be calculated with the help of the Precision and Recall values. The F1-Score value is obtained from the average calculation of Precision and Recall. The higher the F1-Score value indicates the higher the Precision and Recall values. Precision is the ratio between TP and the total predicted positive data, and Recall is the ratio between TP and the actual total data. Precision, recall, and F1-Score values can be calculated using the formulas in equations (3), (4), and (5).

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

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

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

3. RESULTS AND DISCUSSION

3.1 FastText Word Embedding

FastText starts by breaking words into subwords or character n-grams. For example, the word "apple" could be broken down into the subwords "ap", "app", "ple", and "apple". The n-grams could be character 3-grams, 4-grams, 5-grams, etc. The choice of the n-gram length is a trade-off between capturing enough information about the word's morphological structure and the computational cost of learning the embeddings. FastText then learns vector representations (embeddings) for each subword using a neural network architecture. These embeddings are learned through a process called backpropagation, where the network's parameters are updated based on the error between the predicted and actual class labels.

In this study, to form sentence vectors in the dataset, each word vector that has been formed by FastText is then summed based on the words contained in a sentence. That way the sentence vector still has the same vector size. The size of the word vector is 100 and the sentence vector also has a vector size of 100. After the sentence vector is formed, then these vectors are used as features used at the classification stage by the Support Vector Machine (SVM).

3.2 Experimental Result

In this study, experiments were conducted to compare the performance of the SVM algorithm in performing sentiment classification in the feature extraction method with TF-IDF and FastText.. Also, experiments were carried out on kernel, C, and gamma parameters in the SVM algorithm. In addition, the experiment was also carried out at 3 test data ratios namely 70:30, 80:20 and 90:10. The kernel used is RBF and linear. The C values used are 0.1, 1, and 10. As well as the gamma values used are 0.1, 1, auto, and scale. The experimental results can be seen in Tables 3, 4 and 5.

Tabel 3. Result of 70:30 Test Data Ratio

Feature Extraction	Kernel	C	Gamma	Accuracy	Precision	Recall	F1-Score
TF-IDF	rbf	0.1	0.1	0.5692	0.5692	1.0	0.7255
	rbf	0.1	1	0.5932	0.5832	1.0	0.7367
	rbf	0.1	auto	0.5692	0.5692	1.0	0.7255
	rbf	0.1	scale	0.5932	0.5832	1.0	0.7367
	linear	0.1	0.1	0.8347	0.7771	0.995	0.8727
	linear	0.1	1	0.8347	0.7771	0.995	0.8727
	linear	0.1	auto	0.8347	0.7771	0.995	0.8727
	linear	0.1	scale	0.8347	0.7771	0.995	0.8727
	rbf	1	0.1	0.8828	0.8361	0.9876	0.9056
	rbf	1	1	0.9237	0.9068	0.9653	0.9351
	rbf	1	auto	0.5692	0.5692	1.0	0.7255
	rbf	1	scale	0.9237	0.9068	0.9653	0.9351
	linear	1	0.1	0.9209	0.9121	0.9529	0.932
	linear	1	1	0.9209	0.9121	0.9529	0.932
	linear	1	auto	0.9209	0.9121	0.9529	0.932
	linear	1	scale	0.9209	0.9121	0.9529	0.932
	rbf	10	0.1	0.9138	0.9091	0.9429	0.9257
	rbf	10	1	0.9251	0.9127	0.9603	0.9359
	rbf	10	auto	0.5692	0.5692	1.0	0.7255
	rbf	10	scale	0.9251	0.9127	0.9603	0.9359

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

Feature Extraction	Kernel	C	Gamma	Accuracy	Precision	Recall	F1-Score
	linear	10	0.1	0.9096	0.9104	0.933	0.9216
	linear	10	1	0.9096	0.9104	0.933	0.9216
	linear	10	auto	0.9096	0.9104	0.933	0.9216
	linear	10	scale	0.9096	0.9104	0.933	0.9216
FASTTEXT	rbf	0.1	0.1	0.5692	0.5692	1.0	0.7255
	rbf	0.1	1	0.5692	0.5692	1.0	0.7255
	rbf	0.1	auto	0.9661	0.9523	0.9901	0.9708
	rbf	0.1	scale	0.9718	0.9549	0.9975	0.9757
	linear	**0.1**	**0.1**	**0.9887**	**0.9877**	**0.9926**	**0.9901**
	linear	0.1	1	0.9887	0.9877	0.9926	0.9901
	linear	0.1	auto	0.9887	0.9877	0.9926	0.9901
	linear	0.1	scale	0.9887	0.9877	0.9926	0.9901
	rbf	1	0.1	0.5791	0.5749	1.0	0.7301
	rbf	1	1	0.5749	0.5724	1.0	0.7281
	rbf	1	auto	0.9774	0.9685	0.9926	0.9804
	rbf	1	scale	0.9802	0.9709	0.995	0.9828
	linear	1	0.1	0.9845	0.9828	0.9901	0.9864
	linear	1	1	0.9845	0.9828	0.9901	0.9864
	linear	1	auto	0.9845	0.9828	0.9901	0.9864
	linear	1	scale	0.9845	0.9828	0.9901	0.9864
	rbf	10	0.1	0.5819	0.5765	1.0	0.7314
	rbf	10	1	0.5749	0.5724	1.0	0.7281
	rbf	10	auto	0.9788	0.9732	0.9901	0.9815
	rbf	10	scale	0.9774	0.9708	0.9901	0.9803
	linear	10	0.1	0.9845	0.9828	0.9901	0.9864
	linear	10	1	0.9845	0.9828	0.9901	0.9864
	linear	10	auto	0.9845	0.9828	0.9901	0.9864
	linear	10	scale	0.9845	0.9828	0.9901	0.9864

Tabel 4. Result of 80:20 Test Data Ratio

Feature Extraction	Kernel	C	Gamma	Accuracy	Precision	Recall	F1-Score
TF-IDF	rbf	0.1	0.1	0.5466	0.5466	1.0	0.7068
	rbf	0.1	1	0.5869	0.5695	1.0	0.7257
	rbf	0.1	auto	0.5466	0.5466	1.0	0.7068
	rbf	0.1	scale	0.5869	0.5695	1.0	0.7257
	linear	0.1	0.1	0.8347	0.7727	0.9884	0.8673
	linear	0.1	1	0.8347	0.7727	0.9884	0.8673
	linear	0.1	auto	0.8347	0.7727	0.9884	0.8673
	linear	0.1	scale	0.8347	0.7727	0.9884	0.8673
	rbf	1	0.1	0.8708	0.8188	0.9806	0.8924
	rbf	1	1	0.9153	0.8865	0.969	0.9259
	rbf	1	auto	0.5466	0.5466	1.0	0.7068
	rbf	1	scale	0.9153	0.8865	0.969	0.9259
	linear	1	0.1	0.9153	0.9007	0.9496	0.9245
	linear	1	1	0.9153	0.9007	0.9496	0.9245
	linear	1	auto	0.9153	0.9007	0.9496	0.9245
	linear	1	scale	0.9153	0.9007	0.9496	0.9245
	rbf	10	0.1	0.9174	0.9011	0.9535	0.9266
	rbf	10	1	0.9216	0.9048	0.9574	0.9303
	rbf	10	auto	0.5466	0.5466	1.0	0.7068
	rbf	10	scale	0.9216	0.9048	0.9574	0.9303
	linear	10	0.1	0.9068	0.8934	0.9419	0.917
	linear	10	1	0.9068	0.8934	0.9419	0.917
	linear	10	auto	0.9068	0.8934	0.9419	0.917
	linear	10	scale	0.9068	0.8934	0.9419	0.917
FASTTEXT	rbf	0.1	0.1	0.5466	0.5466	1.0	0.7068
	rbf	0.1	1	0.5466	0.5466	1.0	0.7068
	rbf	0.1	auto	0.9576	0.9407	0.9845	0.9621
	rbf	0.1	scale	0.964	0.9446	0.9922	0.9679
	linear	**0.1**	**0.1**	**0.9809**	**0.9807**	**0.9845**	**0.9826**

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

Feature Extraction	Kernel	C	Gamma	Accuracy	Precision	Recall	F1-Score
	linear	0.1	1	0.9809	0.9807	0.9845	0.9826
	linear	0.1	auto	0.9809	0.9807	0.9845	0.9826
	linear	0.1	scale	0.9809	0.9807	0.9845	0.9826
	rbf	1	0.1	0.5551	0.5513	1.0	0.7107
	rbf	1	1	0.553	0.5501	1.0	0.7098
	rbf	1	auto	0.9831	0.9771	0.9922	0.9846
	rbf	1	scale	0.9809	0.9734	0.9922	0.9827
	linear	1	0.1	0.9852	0.9846	0.9884	0.9865
	linear	1	1	0.9852	0.9846	0.9884	0.9865
	linear	1	auto	0.9852	0.9846	0.9884	0.9865
	linear	1	scale	0.9852	0.9846	0.9884	0.9865
	rbf	10	0.1	0.5636	0.556	1.0	0.7147
	rbf	10	1	0.553	0.5501	1.0	0.7098
	rbf	10	auto	0.9788	0.9697	0.9922	0.9808
	rbf	10	scale	0.9809	0.9734	0.9922	0.9827
	linear	10	0.1	0.9852	0.9883	0.9845	0.9864
	linear	10	1	0.9852	0.9883	0.9845	0.9864
	linear	10	auto	0.9852	0.9883	0.9845	0.9864
	linear	10	scale	0.9852	0.9883	0.9845	0.9864

Tabel 5. Result of 90:10 Test Data Ratio

Feature Extraction	Kernel	C	Gamma	Accuracy	Precision	Recall	F1-Score
TF-IDF	rbf	0.1	0.1	0.5042	0.5042	1.0	0.6704
	rbf	0.1	1	0.5763	0.5434	1.0	0.7041
	rbf	0.1	auto	0.5042	0.5042	1.0	0.6704
	rbf	0.1	scale	0.572	0.5409	1.0	0.7021
	linear	0.1	0.1	0.8475	0.7748	0.9832	0.8667
	linear	0.1	1	0.8475	0.7748	0.9832	0.8667
	linear	0.1	auto	0.8475	0.7748	0.9832	0.8667
	linear	0.1	scale	0.8475	0.7748	0.9832	0.8667
	rbf	1	0.1	0.8983	0.8417	0.9832	0.907
	rbf	1	1	0.9237	0.8976	0.958	0.9268
	rbf	1	auto	0.5042	0.5042	1.0	0.6704
	rbf	1	scale	0.9237	0.8976	0.958	0.9268
	linear	1	0.1	0.9322	0.9328	0.9328	0.9328
	linear	1	1	0.9322	0.9328	0.9328	0.9328
	linear	1	auto	0.9322	0.9328	0.9328	0.9328
	linear	1	scale	0.9322	0.9328	0.9328	0.9328
	rbf	10	0.1	0.928	0.925	0.9328	0.9289
	rbf	10	1	0.9322	0.9187	0.9496	0.9339
	rbf	10	auto	0.5042	0.5042	1.0	0.6704
	rbf	10	scale	0.9322	0.9187	0.9496	0.9339
	linear	10	0.1	0.911	0.9083	0.916	0.9121
	linear	10	1	0.911	0.9083	0.916	0.9121
	linear	10	auto	0.911	0.9083	0.916	0.9121
	linear	10	scale	0.911	0.9083	0.916	0.9121
FASTTEXT	rbf	0.1	0.1	0.5042	0.5042	1.0	0.6704
	rbf	0.1	1	0.5042	0.5042	1.0	0.6704
	rbf	0.1	auto	0.9746	0.9593	0.9916	0.9752
	rbf	0.1	scale	0.9746	0.9593	0.9916	0.9752
	linear	0.1	0.1	0.9958	1.0	0.9916	0.9958
	linear	0.1	1	0.9958	1.0	0.9916	0.9958
	linear	0.1	auto	0.9958	1.0	0.9916	0.9958
	linear	0.1	scale	0.9958	1.0	0.9916	0.9958
	rbf	1	0.1	0.5127	0.5085	1.0	0.6742
	rbf	1	1	0.5127	0.5085	1.0	0.6742
	rbf	**1**	**auto**	**0.9958**	**0.9917**	**1.0**	**0.9958**
	rbf	1	scale	0.9915	0.9916	0.9916	0.9916
	linear	1	0.1	0.9958	0.9917	1.0	0.9958
	linear	1	1	0.9958	0.9917	1.0	0.9958

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

Feature Extraction	Kernel	C	Gamma	Accuracy	Precision	Recall	F1-Score
	linear	1	auto	0.9958	0.9917	1.0	0.9958
	linear	1	scale	0.9958	0.9917	1.0	0.9958
	rbf	10	0.1	0.5297	0.5174	1.0	0.6819
	rbf	10	1	0.5127	0.5085	1.0	0.6742
	rbf	10	auto	0.9831	0.9752	0.9916	0.9833
	rbf	10	scale	0.9915	0.9835	1.0	0.9917
	linear	10	0.1	0.9915	0.9916	0.9916	0.9916
	linear	10	1	0.9915	0.9916	0.9916	0.9916
	linear	10	auto	0.9915	0.9916	0.9916	0.9916
	linear	10	scale	0.9915	0.9916	0.9916	0.9916

From the three tables above it can be seen that FastText can provide better performance than TF-IDF and so it can provide a better accuracy score and F1 score. From experiments on the test data ratio of 70:30 shown in Table 3, it can be seen that the combination of kernel, C, and gamma parameter values that can produce the best accuracy and F1-score values are linear, 0.1, and 0.1 with an accuracy score, respectively. of 0.9887 and F1-score of 0.9901. At the test data ratio of 80:20 in Table 4, it can be seen that the combination of kernel, C, and gamma values that can produce the best accuracy and F1-score values are linear, 0.1, and 0.1 respectively with an accuracy score of 0.9809 and F1- score of 0.9826. At the test data ratio of 90:10 in Table 5, it can be seen that the combination of kernel, C, and gamma values that produce the highest accuracy and F1-score values are respectively rbf, 1, and auto with an accuracy score of 0.9958 and F1-score of 0.9958.

Comparison of ratio values in test data can affect the SVM algorithm in the learning process on training data. The more the amount of training data, the better the SVM algorithm in learning on the training data [16]. Comparison of the performance of the SVM model on the ratio of the test data is shown in Table 6.

Tabel 6. Test Data Ratio

Ratio	Accuracy	F1-Score
70:30	0.9887	0.9901
80:20	0.9809	0.9826
90:10	**0.9958**	**0.9958**

In Table 6 it can be seen that at a test data ratio of 90:10 the SVM with FastText feature extraction can provide the highest accuracy value compared to the other 2 ratios. This means that a large amount of training data can make the SVM algorithm able to provide good performance in classifying text sentiments, in this case sentiment in news headlines about the 2024 Presidential Election. Also, with the use of FastText word embedding for feature extraction can improve the performance of SVM in the sentiment analysis task.

4. CONCLUSION

In this research, a sentiment analysis system has been built to predict whether a news headline has positive or negative sentiment. The news title data was obtained from the CNN Indonesia news media as many as 2360 title data. Then do the labeling of positive and negative labels. In making the sentiment analysis model, experiments were carried out on the use of FastText in feature extraction by comparing it with TF-IDF.Also, experiments on SVM were carried out on the kernel, C, and gamma parameters as well as on the test data ratio. The use of FastText in feature extraction can outperform the TF-IDF method in SVM.. The combination of kernel, C, and gamma parameter values that can produce the best accuracy and F1-score values are rbf, 10, and 1 respectively at a test data ratio of 90:10 with an accuracy score of 93% and an F1-Score of 93%. From the experiment above it can be concluded that the SVM sentiment analysis model can classify sentiment in news headlines about the 2024 presidential election with good performance which can simplify the sentiment analysis process in news headlines. For future work, we recommend using a larger number of datasets, another feature extraction techniques, as well as using other machine learning algorithms or even using deep learning algorithms so as to improve performance in sentiment analysis systems.

REFERENCES

[1] I. Kurniawan and A. Susanto, “Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019,” Eksplora Informatika, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.
[2] A. F. Sabily, P. P. Adikara, and M. A. Fauzi, “Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer , vol. 3, pp. 4204–4209, May 2019.
[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, pp. 9997–10004, Aug. 2019.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 4, No. 2, Februari 2023, Page 397−404
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v4i2.3051

[4] R. A. A. Malik and Y. Sibaroni, "Multi-aspect Sentiment Analysis of Tiktok Application Usage Using FasText Feature Expansion and CNN Method," Journal of Computer System and Informatics (JoSYC), vol. 3, no. 4, pp. 277–285, Sep. 2022, doi: 10.47065/josyc.v3i4.2033.
[5] S. Kurniawan, W. Gata, D. A. Puspitawati, N. -, M. Tabrani, and K. Novel, "Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 3, no. 2, pp. 176–183, Aug. 2019, doi: 10.29207/resti.v3i2.935.
[6] D. T. Lukmana, S. Subanti, and Y. Susanti, "ANALISIS SENTIMEN TERHADAP CALON PRESIDEN 2019 DENGAN SUPPORT VECTOR MACHINE DI TWITTER," SEMINAR NASIONAL PENELITIAN PENDIDIKAN MATEMATIKA (SNP2M) , pp. 154–160, 2019.
[7] A. G. Soenarya and K. M. Lhaksmana, "Sentimen Analisis Politik Berita Media OnlineDalam Pemilihan Presiden 2019 Menggunakan Metode Support Vector Machine," e-Proceeding of Engineering, vol. 6, no. 2, pp. 10005–10012, Aug. 2019.
[8] O. Z. A. Tane and K. M. Lhaksmana, "Analisis Sentimen Pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode Svm (Support Vector Machine).," eProceedings of Engineering, vol. 6, no. 2, pp. 9716–9725, Aug. 2019.
[9] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika, vol. 8, no. 1, pp. 59–71, Mar. 2022, doi: 10.23917/khif.v8i1.15213.
[10] T. H. Jaya Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," Procedia Comput Sci, vol. 197, pp. 660–667, 2022, doi: 10.1016/j.procs.2021.12.187.
[11] I. Santos, N. Nedjah, and L. de Macedo Mourelle, "Sentiment analysis using convolutional neural network with fastText embeddings," in 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Nov. 2017, pp. 1–5. doi: 10.1109/LA-CCI.2017.8285683.
[12] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Trans Assoc Comput Linguist, vol. 5, pp. 135–146, Dec. 2017, doi: 10.1162/tacl_a_00051.
[13] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A Novel Active Learning Method Using SVM for Text Classification," International Journal of Automation and Computing, vol. 15, no. 3, pp. 290–298, Jun. 2018, doi: 10.1007/s11633-015-0912-z.
[14] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," IEEE J Sel Top Appl Earth Obs Remote Sens, vol. 13, pp. 6308–6325, 2020, doi: 10.1109/JSTARS.2020.3026724.
[15] "RBF SVM parameters." https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html#sphx-glr-auto-examples-svm-plot-rbf-parameters-py (accessed Jan. 31, 2023).
[16] A. A. Aldino, A. Saputra, A. Nurkholis, and S. Setiawansyah, "Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur," Building of Informatics, Technology and Science (BITS), vol. 3, no. 3, pp. 325–330, Dec. 2021, doi: 10.47065/bits.v3i3.1041.