Emerging Information Science and Technology
Vol. 4, No. 1, (2023), pp. 1-7

Sentiment Analysis of Public Responses on Indonesia Government Using Naïve Bayes and Support Vector Machine

Haris Setyawan*, Laila Ma'rifatul Azizah, Alvira Yusnia Pradani

Universitas Muhammadiyah Yogyakarta, Jln.Brawijaya, Tamantirto, Kasihan, Bantul, Yogyakarta 55183, Indonesia
**Corresponding author: haris.setyawan@umy.ac.id*

Abstract

Many people are interested in knowing how the public views President Joko Widodo's administration. Text Mining analysis can be one way to collect and analyze text data about Joko Widodo's administration and extract relevant information from the data. Data was obtained by collecting tweet data about Joko Widodo's government in 2022 on Twitter using Netlyitic. Then the Text Mining analysis of Joko Widodo's government was carried out using the Navie Bayes (NVB) classification and Support Vector Machine (SVM). This classification can be used to predict sentiment or public views of the government based on the tweets collected. Based on a case study of the classification results of President Joko Widodo using Naive Bayesian classification, we obtained a precision value of 79%, a recall value of 91% and a precision value of 82%. And by using SVM, we get 85% precision, 95% recall, and 83% precision. Due to the high accuracy, recall, and precision, it can be said that SVM classification is more accurate than NVB.

Keywords: *Naïve Bayes, Support Vector Machine, Text Mining, analysis, classification*

1. Introduction

Text Mining and Sentiment Analysis is a field of research that has a lot of interest, it is because of the increasing number of social media users Twitter, Facebook, Instagram and other social media, so. a lot of data has been obtained that can be used as research material [1], [2]. One of the benefits that can be obtained from this data is being able to find out the opinions or sentiments of social network users regarding a topic that is being discussed on social media [3], [4]. Because from the results of sentiment analysis can be obtained new knowledge that can be used to make decisions [5].

Joko Widodo's government is one of the governments that has become a public concern in Indonesia. Many people are interested in knowing how Joko Widodo's government is performing and how the public views his government. Text mining analysis can be one way to collect and analyze text data about Joko Widodo's administration and extract relevant information from the data.

So the author conducts Text Mining Analysis on the data that the author has scraped on twitter using Netlytic with the keywords #Joko Widodo and #pemerintahanjokowi taken between August-October 2022, so that the data can be used to identify trends and patterns in public views on government policies and actions. This can be done by collecting and analyzing news from social media posts and other sources of information about the government using Text Mining techniques[6] to identify the sentiment of the public towards president Joko Widodo's administration. Text Mining can also identify key issues of public concern, such as economic development, corruption, or social welfare.

And the results of tetxt mining can be concluded that the public's views are positive or negative towards the issues obtained.

Text Mining analysis of Joko Widodo's government using Navie Bayes classification [7], [8] and Support Vector Machine (SVM) [9]which previously collected data from twitter using Netlyitic regarding Jokowi's government in 2022 with tweet data from Twitter, this classification can be used to predict sentiment or public views on the government based on tweets collected.

.

2. Method

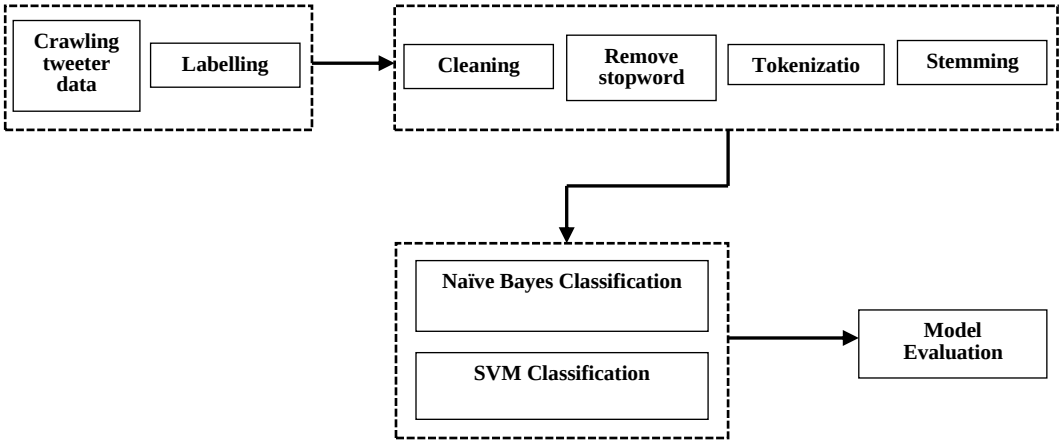
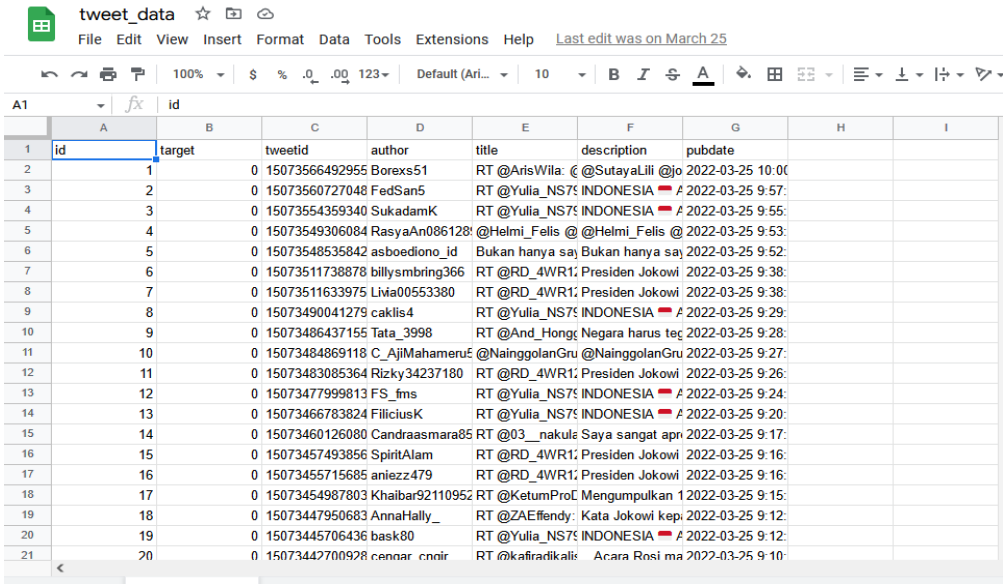


Figure 1. Text Mining and Analysis

2.1. Crawling Twitter data

The data used in this study is Twitter user tweet post data on Twitter.com. The data collected is tweet text data taken using crawling techniques using the Netlytic web[10]. At this stage, 4532 data were obtained that could be processed.


tweet_data

id	target	tweetid	author	title	description	pubdate
1	0	15073566492955	Borexs51	RT @ArisWila: @	@SutayaLili @jo	2022-03-25 10:0
2	0	15073560727048	FedSan5	RT @Yulia_NS7	INDONESIA 🇮🇩 A	2022-03-25 9:57:
3	0	15073554359340	SukadamK	RT @Yulia_NS7	INDONESIA 🇮🇩 A	2022-03-25 9:55:
4	0	15073549306084	RasyaAn086128	@Helmi_Felis @	@Helmi_Felis @	2022-03-25 9:53:
5	0	15073548535842	asboediono_id	Bukan hanya say	Bukan hanya say	2022-03-25 9:52:
6	0	15073511738878	billysmbring366	RT @RD_4WR1	Presiden Jokowi	2022-03-25 9:38:
7	0	15073511633975	Livia00553380	RT @RD_4WR1	Presiden Jokowi	2022-03-25 9:38:
8	0	15073490041279	caklis4	RT @Yulia_NS7	INDONESIA 🇮🇩 A	2022-03-25 9:29:
9	0	15073486437155	Tata_3998	RT @And_Hongg	Negara harus teg	2022-03-25 9:28:
10	0	15073484869118	C_AjiMahameru5	@NainggolanGru	@NainggolanGru	2022-03-25 9:27:
11	0	15073483085364	Rizky34237180	RT @RD_4WR1	Presiden Jokowi	2022-03-25 9:26:
12	0	15073477999813	FS_fms	RT @Yulia_NS7	INDONESIA 🇮🇩 A	2022-03-25 9:24:
13	0	15073466783824	FiliciusK	RT @Yulia_NS7	INDONESIA 🇮🇩 A	2022-03-25 9:20:
14	0	15073460126080	Candraasmara85	RT @03__nakula	Saya sangat apr	2022-03-25 9:17:
15	0	15073457493856	SpiritAlam	RT @RD_4WR1	Presiden Jokowi	2022-03-25 9:16:
16	0	15073455715685	aniezz479	RT @RD_4WR1	Presiden Jokowi	2022-03-25 9:16:
17	0	15073454987803	Khaibar92110952	RT @KetumProD	Mengumpulkan 1	2022-03-25 9:15:
18	0	15073447950683	AnnaHally_	RT @ZAEffendy:	Kata Jokowi kepa	2022-03-25 9:12:
19	0	15073445706436	bask80	RT @Yulia_NS7	INDONESIA 🇮🇩 A	2022-03-25 9:12:
20	0	15073442700928	cenqar_cngir	RT @kafiradikalis	Acara Rosi ma	2022-03-25 9:10:

Figure 2. Data Samples

2.2. Labelling

Then perform the data labeling process to determine the classification of opinions or views from the results of the tweets that have been crawled earlier. In this labeling process, it is divided into 3 classes, namely positive, negative, and neutral.

2.3. Preprocessing

Preprocessing is a process to clean data from words or tweets that are not needed and words that have no meaning. So this process needs to be done according to the data from the data retrieval process or crawling Twitter data. The steps of preprocessing have the following sequence:

2.3.1. Cleaning

Cleaning is the process of removing symbols, punctuation marks, capital letters and numbers that often appear in Twitter user tweets so that the data becomes ineffective and meaningless data. An example of the application of the cleaning process can be seen in the table below.

Table 1. Cleaning Process

Before cleaning	*After cleaning*
RT @RD_4WR1212: Presiden Jokowi Marah, Seandainya saja, pengadaan barang dan jasa pemerintahan secara konsisten membeli produk buatan industry	*presiden jokowi marah, seandainya saja, pengadaan barang dan jasa pemerintahan secara konsisten membeli produk buatan industry*
The Power of Emak-Emak.. Ditujukan kepada rezim pemerintahan acakadut & zhalim @jokowi yg didukung mayoritas poli_tikus busuk Senayan @mprgoid @DPR_RI yg hobi kasak-kusuk demi kepentingan perut sendiri. Gemar mbideg mbudeg thd kepentingan rakyat. Waras kalian, dab?! Hmm.. https://t.co/FKZjLtyWaN	*the power of emak-emak ditujukan kepada rezim pemerintahan acakadut yg didukung mayoritas poli tikus busuk senayan yg hobi kasak-kusuk demi kepentingan perut sendiri. gemar mbideg mbudeg thd kepentingan rakyat.*
@Helmi_Felis Cari masalah lu, mi? Mau ngajak makar ya? Penghianat! Lu aja yg keluar dari Indonesia. Lu gak ada gunanya di sini. Sampah! Lu berani macam2 sama pemerintahan resmi pak Jokowi? @CCICPolri tangkap nih si Helmi, mau makar!	*cari masalah lu mi mau ngajak makar ya penghianat lu aja yg keluar dari Indonesia lu gak ada gunanya di sini sampah lu berani macam2 sama pemerintahan resmi pak Jokowi*
@detikcom Pak Jokowi sedang mengkritik pemerintahan, semoga didengar presiden yah pak	*pak jokowi sedang mengkritik pemerintahan semoga didengar presiden yah pak*

2.3.2. Remove Stopword

Remove Stopword is the process of removing words that are less meaningful or words that have no meaning such as the words and, or, you, me. An example of the application process at the Remove Stopword stage can be seen in the following table.

Table 2. Remove Stopword Process

Before Remove Stopword	*After Remove Stopword*
presiden jokowi marah, seandainya saja, pengadaan barang dan jasa pemerintahan secara konsisten membeli produk buatan industry sendiri	*presiden jokowi marah seandainya saja pengadaan barang dan jasa pemerintahan secara konsisten membeli produk buatan industry sendiri*
Pak jokowi sedang mengkritik pemerintahan semoga didengar presiden yah pak	*pak jokowi sedang mengkritik pemerintahan semoga didengar presiden yah pak*

2.3.5. Tokenization

Tokenization is the process of breaking a sentence into parts called tokens. A token can be thought of as a single word, phrase, or meaningful element. Examples of the process at the tokenization stage can be seen in the following table.

Table 3. Tokenization Process

Before Tokenization	*After Tokenization*
presiden jokowi marah seandainya saja pengadaan barang dan jasa pemerintahan secara konsisten membeli produk buatan industry	*['presiden','jokowi','marah','seandainya', 'saja','pengadaan','barang','dan','jasa','pemeri ntahan','secara','konsisten','membeli','produk',' buatan','industri']*
pak jokowi sedang mengkritik pemerintahan semoga didengar presiden yah pak	*['pak','jokowi','sedang','mengkritik','pemerintah an', 'semoga', 'didengar', 'presiden' 'yah',' pak']*

2.3.4. Stemming

Stemming is the process of converting words into basic form by removing additional words in front or behind the word. Examples of the stemming application process can be seen in the following table.

Table 4. Stemming Process

Before Stemming	*After Stemming*
['presiden','jokowi','marah','seandainya', 'saja','pengadaan','barang','dan','jasa','pemerintahan', 'secara','konsisten','membeli',' produk','buatan','industri']	*presiden jokowi marah seandainya saja pengadaan barang jasa pemerintahan secara konsisten membeli produk buatan industry*
['pak', 'jokowi', 'sedang', 'mengkritik', 'pemerintahan', 'semoga', 'didengar', 'presiden' 'yah',' pak']	*pak jokowi sedang mengkritik pemerintahan semoga didengar presiden pak*

The next step is to create a function that is useful for simplifying the classification of tweet data, this process is done by the feature extraction process. In this feature extraction process, two processes are carried out, namely the word vector generation

process. In this process, the system converts text into a vector representation and words into a weighting process.

Tf-idf *(term frequency - inverse document frequency)*. This word vector can be interpreted as an Indonesian word vector, which is the process of creating a matrix from existing sentences, each matrix row represents a row of documents, while each matrix column represents all the words contained in the tweet text. After all words are processed and converted into word vectors, the next step is to give weight to each word from each sentence or document using the Unigram and Tf-idf (*term frequency - inverse document frequency*) formulas. After the weighting process is complete, the dataset can be used to train the Naive BayesClassifier and Support Vector Machine calculations.

3. Result and Discussion

After testing the training data from the data, several classes are created from the test data, which are often called prediction classes. The actual prediction class comes from the previous test data which is hidden and then displayed and calculated the accuracy, precision, recall and f1 score results. As shown below, the accuracy table of the model estimation results using the Navie Bayesian Confusion Matrix method.

```
[ ]  model_Evaluate(nb)
```

	precision	recall	f1-score	support
0	0.42	0.63	0.51	38
2	0.92	0.83	0.87	189
accuracy			0.79	227
macro avg	0.67	0.73	0.69	227
weighted avg	0.83	0.79	0.81	227

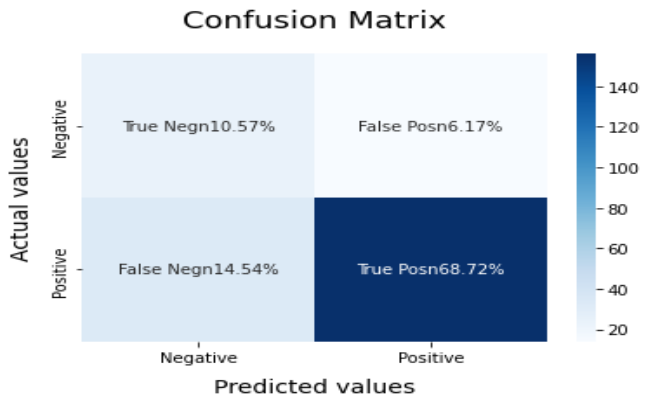


Figure 3. Matrix Result using the Naïve Bayes Method

The results of the classification using the Navie Bayes method can be seen from the results of the data above, the resulting level reaches 0.79 or can be called an accuracy level of 79%.

However, for the accuracy results using the SVM method, the results are much better than using the Navie Bayes method, which is 0.85 or can be called an 85% accuracy rate.

```
[ ]  model_Evaluate(SVM)
```

	precision	recall	f1-score	support
0	0.59	0.34	0.43	38
2	0.88	0.95	0.91	189
accuracy			0.85	227
macro avg	0.73	0.65	0.67	227
weighted avg	0.83	0.85	0.83	227

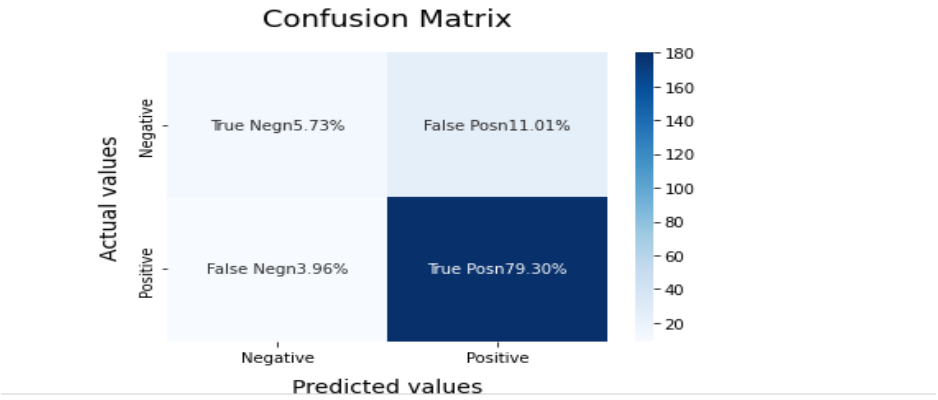


Figure 4. Matrix Result using the Support Vector Machine Method

4. Conclusion

Based on the case study of President Joko Widodo's classification results using Naive Bayesian classification, a precision value of 79%, a recall value of 91% and a precision value of 82% are obtained. And by using Support Vector Machine, a precision value of 85%, a recall of 95%, and a precision of 83% are obtained. Due to the high accuracy, recall, and precision, it can be said that Support Vector Machine (SVM) classification is more accurate than Naive Bayesian (NVB), which is true. Other measurements such as specificity, false positive rate, and area underestimation (AUC) also give correct values. It can be seen that the AUC value obtained is 0.85 or 85% which means that the value is very good or the classification is good with the SVM method compared to the results obtained by the Navie Bayes method.

References

[1] Maria Arista Ulfa, Budi Irmawati, and Ario Yudo Husodo, "Twitter Sentiment Analysis using Na¨ıve Bayes Classifier with Mutual Information Feature Selection," *J-COSINE*, vol. 2, no. 2, pp. 106–111, Desember 2018.

[2] Faisal Rahutomo,Pramana Yoga Saputra, Miftahul Agtamas Fidyawan, "IMPLEMENTATION OF TWITTER SENTIMENT ANALYSIS FOR MOVIE REVIEWS USING SUPPORT VECTOR MACHINE ALGORITHM," *Jurnal Informatika Polinema*, vol. 4, no. 2, pp. 93–100, Feb. 2018.(In Bahasa Indonesia)

[3] Yono Cahyono, Saprudin Saprudin, "Sentiment Analysis of Sundanese Tweets Using Naive Bayes Classifier with Feature Selection Chi Squared Statistic," *Jurnal Informatika Universitas Pamulang*, vol. 4, no. 3, pp. 87–94, Sep. 2019.(In Bahasa Indonesia)

[4] Mona Cindo, Dian Palupi Rini, Ermatita Ermatita, "SENTIMENT ANALYSIS ON TWITTER BY USING MAXIMUM ENTROPY AND SUPPORT VECTOR

MACHINE METHOD," *SINERGI*, vol. 24, no. 2, pp. 87–94, Jun. 2020, doi: http://doi.org/10.22441/sinergi.2020.2.002.

[5] Fazainsyah Azka Wicaksono, Ade Romadhony, Hasmawati, "Sentiment Analysis of University Social Media Using Support Vector Machine and Logistic Regression Methods," *Ind. Journal on Computing*, vol. 7, no. 2, pp. 15–24, Aug. 2022, doi: 10.34818/indojc.2022.7.2.638.

[6] Imamah, Husni, Eka Malasari Rachman, Ika Oktavia Suzanti, and Fifin Ayu and Mufarroha, "Text Mining and Support Vector Machine for Sentiment Analysis of Tourist Reviews in Bangkalan Regency," *Journal of Physics*, vol. 1477, no. 2, pp. 1–6, Feb. 2023, doi: 10.1088/1742-6596/1477/2/022023.

[7] I Gede Cahya Purnama YasaNgurah Agus Sanjaya ER, Luh Arida Ayu Rahning Putri, "Sentiment Analysis of Snack Review Using the Naïve Bayes Method," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 8, no. 3, pp. 333–338, Feb. 2020.

[8] Mohammad Abdul Manan1, , Sarwido, and , Gentur Wahyu Nyipto Wibowo, "APPLICATION OF NAIVE BAYES ALGORITHM FOR PREDICTING HEREGISTRATION OF NEW STUDENT CANDIDATES," *JURNAL TEKNIK INFORMATIKA*, vol. 1, no. 1, pp. 1–10, Feb. 2022.(In Bahasa Indonesia)

[9] Normah, "Naïve Bayes Algorithm For SentimentAnalysis Windows PhoneStore Application Reviews," *Sinkron, Journal Publications & Informatics Engineering Research*, vol. 3, no. 2, pp. 1–7, Apr. 2019, doi: https://doi.org/10.33395/sinkron.v3i2.242.

[10] Intania Eka Yanti, Oktariani Nurul Pratiwi and , Riska Yanu Fa'rifah, "ANALYSIS OF PUBLIC RESPONSES TO THE COVID-19 PANDEMIC ON TWITTER SOCIAL MEDIA USING THE SOCIAL NETWORK ANALYSIS METHOD," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 8988–8999, Oktober 2021.(In Bahasa Indonesia)