Sentiment Analysis Of State Officials News On Online Media Based On Public Opinion Using Naive Bayes Classifier Algorithm And Particle Swarm Optimization

Ali Idrus
AMIK BSI Purwokerto, Management of Informatic Program
Purwokerto, Indonesia
ali.alu@bsi.ac.id

Herlambang Brawijaya
STMIK Nusa Mandiri, Informatic Technic Program
Jakarta, Indonesia
Herlambang.braw@gmail.com

Maruloh
STMIK Nusa Mandiri, Information System Program
Jakarta, Indonesia
maruloh.mru@gmail.com

***Abstract* - Online media in general, ie all types of media formats that can only be accessed via the Internet contains text, photos, video, and sound. With the understanding of online media in general then email, mailing list (mailing list), website, blog, WhatsApp, and social media into the category of online media. the advantages of online media are compared to the conventional media due to its very flexible because it can be accessed in various places connected to the Internet network, the information provided is up to date and also anyone can comment on related articles. Information is spreading very quickly and is accompanied by freedom of expression resulting in different types of opinions, whether negative or positive. In this study, the authors apply the text mining process using Naive Bayes Classifier algorithm and Particle Swarm Optimization to classify sentiments automatically. The data used are 200 comments online news portal readers about the sentiment to the attitude and actions of state officials in carrying out their duties. The data are classified manually and divided into each of 100 data for positive and negative sentiments. Then 200 data is used for testing. The results of this study resulted in a system that can classify sentiments automatically with accuracy ranging from 76.00%. With a fairly high degree of accuracy, Naïve Bayes model supported by Particle Swarm Optimization can provide solutions in classifying public opinion to the news of state officials to be more accurate and optimal.**

Keywords—online media, sentiment analysis, text mining

I. INTRODUCTION

At this time information is very easy to obtain especially information or news about public figures such as state officials, readers hungry for information about the latest news about attitudes, actions and policies submitted or issued by state officials such as presidents, ministers, members of Parliament and leaders this area is so vital that their role in determining the policy direction of this nation. The development of information media in addition to conventional media (newspapers and magazines), Media Online in general, ie all types of media formats that can only be accessed via the Internet contains text, photos, video, and sound. With the understanding of online media in general then email, mailing list (mailing list), website, blog, WhatsApp, and social media into the category of online media. the advantages of online media are compared to the conventional media due to its very flexible because it can be accessed in various places connected to the Internet network, the information provided is up to date and also anyone can comment on related articles. Information is spreading very quickly and is accompanied by freedom of expression resulting in different types of opinions, whether negative or positive. The negative opinion here implies that words or opinions that can cause hostility, humiliation, debate, and disputes in cyberspace. While a positive opinion is a word

or opinion that is positive, supportive and does not cause hostility, humiliation, debate, and disputes in cyberspace.

There is a tendency that a person who is overly sympathetic to a certain official will comment positively on the character and someone else who is less fond of the character will comment negatively. The Government, in this case, protects the public at the same time anticipates the existence of opinions that are despicable individuals or certain groups publish some laws and circulars related to the content of information that is negative or unlawful among them is a Circular issued by the police chief about the prohibition of hate speech (Hate Speech) numbered SE / 06 / X / 2015 which contains Shape, Aspects and Media Hate Speech In the 2nd number of letters (f) mentioned, hate speech may be a criminal act set forth in the Criminal Code other criminal law outside the Criminal Code, in the form of:

1. Humiliation
2. Defamation
3. Unpleasant behavior
4. Provocation
5. Instigate
6. Hoax

and all of the above acts have a purpose or may impact discrimination, violence, disappearance, and / or social conflict.

In the letter (h) mentioned, hate speech as mentioned above can be done through various media, among others:

1. In the oration of campaign activities.
2. Banners.
3. Social media networks.
4. Public opinion subsmission (demonstration)
5. Religious Lectures.
6. Print or electronic media.
7. Pamphlets.

In addition to the above circular criminal offenses in the case of communication in the online media shall be regulated in the 2008 ITE Law Article 27 paragraph (3) ie any person intentionally and without right to distribute and / or transmit and / or make accessible electronic information and / or electronic documents which is either insulting or defamatory. Any person who is proven intentionally to disseminate defamatory electronic information as meant in Article 27 paragraph (3) of the ITE Law shall be charged under Article 45 Paragraph (1) of the ITE Law, maximum imprisonment of 6 years and/or a maximum fine of 1 Billion Rupiah. Therefore required a system that can filter the words that should not be posted.

In recent years, internet users have grown very rapidly. Many forums, blogs, social networks, e-commerce websites, and news reports serve as a form of expressing opinions, which can be used to understand public and consumer opinions on social events, politics, corporate strategies, product preferences, and monitoring reputations[1], because the amount of data contained on the Internet, without being processed to be used deeper then appear Opinion Mining which is a research branch of Text Mining. The focus of Opinion Mining's research is to conduct an opinion analysis from a text document[2].

Based on the above explanation, the analysis is called sentiment analysis, which can generally be defined as a computational study of people's opinions, sentiments and emotions through entities and attributes possessed expressed in text form[3]. The sentiment analysis will classify the polarity of the text present in the sentence or document to know the opinions expressed in the sentence or document whether it is positive, negative or neutral[4].

The analysis of sentiments analyzed a lot of sentiment analysis on textual content on facebook and twitter[5], social media analysis[6], sentiments analysis of text documents[7], sentiment analysis of news content[8], while the author will conduct research on sentiments analysis public opinion news artist. Machine Learning that introduces text classifications such as Naive Bayes, K-NN, SVM and Rocchio Classification[9]. Naive Bayes. Experimental as well as evaluation show that SVM, KNN, and NB are traditional classification texts. Experiments and evaluations indicate valid cloaking texts[10]. For this reason, this study uses the Naïve Bayes Classifier method for text classification.

II. RELATED WORK

The following is a summary of the related research that the researchers used as a guide for this study:

TABLE I
Summary Of the related research

Tittle	Researcher	Classifier	Accuracy	Result
An Advanced Multi Class Instance Selection Based Support Vector Machine For Text Classification	Ramesh	AMCIS SVM	67.1224% (Glass), 78.9042% (Diabetees), 94.7869% (Lonosphere)	The researcher result that AMCISSVM have optimal work and high accuracy
The Role Of Text Pre-Processing In Sentiment Analysis	Haddi	SVM, (TF-IDF, FF, FP)	18.88% (TF-IDF), 76.33% (FF), 82.7% (FP)[11]	The researcher result that data testing using SVM + (TF-IDF, FF, FP) have high accuracy
Opinion Mining Of Movie Review Using Hybrid Method Of Support Vector Machine And Particle Swarm Optimization	Basari, et al	Hybrid SVM PSO	77%	Researcher using PSO for optimization with 10 Fold-Cross Validation
More Than Words: Social Networks Text Mining For Consumer Branf Sentiment	Mostafa	Sentiment Analysis	SP (N 46%, P 54%), SC (N 29, P 34)[12]	The researcher using random sample 3516 tweets for evaluating

				consumer sentiment of branded product using qualitative and quantitative method
A text mining framework for advancing sustainability indicators	Rivera et al	SU, K-NN, SVM, NB	K-NN (85%) SVM (53%) NB (68%)[13]	researchers generate test data with sustainability indicators by analyzing the structured digital news articles with text mining method
Sentiment Analysis Of Artist News using support vector machine and particle swarm optimization	Norma	SVM PSO	?	researchers conducted a review of public opinion reviews of artist news on pso-based svm

III. LITERATURE REVIEW

A. Sentiment Analysis

Sentiment Analysis or opinion mining refers to a wide field of natural language processing, linguistic computing and text mining aimed at analyzing the opinions, sentiments, evaluations, attitudes, judgments and emotions of a person whether the speaker or author pertains to a topic, service product, organization, individual, or certain activities[14].

The purpose of the sentiment analysis is to determine the behavior or opinion of an author by considering a particular topic. Behavior may indicate the reason, opinion or judgment, condition of tendency[15]. Sentiment analysis can also express emotional feelings of sadness, joy, or anger.

According to Moraes[16], the steps commonly found in the text classification of sentiment analysis are:

1. Define the domain dataset

Collection of datasets that surround a domain, such as a movie review dataset, product review dataset, and so on.

2. Pre-processing

Initial processing stages are generally performed with Tokenization, Stopwords removal, and Stemming processes.

3. Transformation

The process of representation of numbers calculated from textual data. Binary representations are commonly used and only count the presence or absence of a word in a document. How many times a word appears in a document is also used as a weighting scheme of textual data. Commonly used processes are TF-IDF, Binary transformation, and Frequency transformation.

4. Feature Selection

Feature selection can make the classifier more efficient by reducing the amount of data to be analyzed by identifying relevant features which will then be processed. The usual feature selection method is Expert. Knowledge, Minimum Frequency, Information gain, Chi-Square, and so forth.

5. Classification

The classification process generally uses classifiers such as Naïve Bayes, Support Vector Machine, and so forth.

6. Interpretation / Evaluation

The evaluation stage usually calculates the accuracy, recall, precision, and F-1.

B. Text Mining

Text mining or text analytics is a term that describes a technology capable of analyzing semi-structured and unstructured text data, this is what distinguishes it from data mining where data mining processes data that are structured. Basically, text mining is an interdisciplinary field that refers to information retrieval, data mining, machine learning, statistics, and linguistic computing[17]

Text mining generally includes the categorization of information or text, grouping text, extraction entities or concepts, developing and formulating a general taxonomy. Text mining deals with structured information or textual extraction of meaningful information and knowledge of large amounts of text[18].

Text mining is a mining performed by a computer to get something new, something unknown or rediscover implicitly implicit information, derived from information extracted automatically from different text data sources[19]. Text mining is a technique used to handle classification, clustering, information extraction and information retrieval[20].

Text mining can analyze documents, group documents based on the words contained therein, and determine the similarities between documents to find out how they relate to other variables.

From the five expert opinions above, it can be concluded that text mining is structured information that is used to analyze or classify documents or text from a large number of documents or text.

C. Opinion Review of state officials

Reviews or reviews contained on the internet are numerous but not processed into useful information. Public sentiment can serve as an indicator to see if the news is qualified or not. Social media is a medium that is often used to pour sentiment or public opinion about the news.

Detikcom is a web portal that contains news and articles online in Indonesia. detikcom is one of the most popular news sites in Indonesia. Different from other Indonesian language news sites, AFP only has online editions and hangs revenue from the advertising field. Even so, detikcom is the leader in terms of new news (Breaking News). Since August 3, 2011, detikcom became part of PT Trans Corporation, one of CT Corp's subsidiaries.

D. Naiive Bayes Classifier Algorithm

The naive bayes classifier algorithm is an algorithm used to find the highest probability value for classifying test data in the most appropriate category[19]. In this study which becomes test data is a document reader news article on news

portal detik.com. There are two stages in the classification of documents. The first stage is the training of documents that are known to the category. While the second stage is the process of classification of documents that have not been known category. In the naïve bayes classifier algorithm, each document is represented by the attribute pair "x1, x2, x3, ... xn" where x1 is the first word, x2 is the second word and so on. While V is a set of comment categories. At the time of classification the algorithm will look for the highest probability of all categories of documents tested (VMAP), where the equation is as follows:

$$V_{MAP} = \underset{V_j \in V}{\arg\max} \frac{P(x_1, x_2, x_3, \ldots x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \ldots x_n)}$$

For P (x1, x2, x3, ... xn) the value is constant for all categories (Vj) so that the equation can be written as follows:

$$V_{MAP} = \underset{V_j \in V}{\arg\max} \prod_{i=1}^{n} P(x_i | V_j) P(V_j)$$

Information:
Vj = Category comments j = 1, 2, 3, ... n. Where in this study j1 = negative comment comment categories, j2 = category of positive sentiment comments
P (xi | Vj) = Probability xi in category Vj
P (Vj) = Probability of Vj

For P (Vj) and P (xi | Vj) is calculated during training where the equation is as follows:

$$P(V_j) = \frac{|docs\ j|}{|contoh|}$$

$$P(x_i | V_j) = \frac{n_k + 1}{n + |kosakata|}$$

Information :
| docs j | = number of documents per category j
| contoh | = number of documents from all categories
nk = number of times the occurrence of each word
n = number of times word occurrence of each category
| kosakata | = number of all words from all categories

IV. METHOD

A. Research Design

Basically, research is an organized investigation, conducted to present information and solve problems. The research method used by the authors use experimental research methods. The research method that the author uses several stages as follows:

1. Data Collection

The data used to conduct the experiment is collected through the website detik.com then data public opinion news officials are selected and collected in the notepad to be processed in testing the data.

2. Preliminary Data Processing

Selects the method to be used when testing the data. The preferred method, based on previous research. The author uses the Naive Bayes Classifier Algorithm Method.

3. Proposed Method

The method proposed by the authors added optimization in order to increase the value of accuracy. The optimization used is Particle Swarm Optimization (PSO).

4. Experiments and Testing Methods

Experiments conducted by researchers, using the framework RapidMiner 6.5.1 to process the data so as to produce accurate accuracy value and for testing methods, the author makes the application using PHP and HTML programming language.

5. Evaluation and Validation of Evaluation Results

Evaluation serves to determine the accuracy of the proposed algorithm model. Validation is used to see the comparison of accuracy results from the model used with pre-existing results. The validation technique used is Cross-Validation. The accuracy of the algorithm will be measured using Confusion Matrix and the calculation results will be shown in the form of ROC Curve (Receiver Operating Characteristic).

B. Preliminary Data Processing

Text Mining is a process that aims to find information or the latest trends that were not previously revealed, by processing and analyzing large amounts of data. In analyzing some or all unstructured text, text mining tries to associate one part of the text with another based on certain rules. Unprocessed text usually has high dimensional characteristics, there is noise on the data and there is a bad text structure. For that, in the processing of initial data, text mining must go through several stages called preprocessing. The stages are:

1. Tokenization

The process of taking every word in the text and changing the letters in the document into lowercase. Only letters received, while special characters or punctuation will be omitted. So the result of the tokenization process is the words that are the compiler of sentences or strings inserted without any punctuation.

2. Transform Cases

Converts whole letters into lowercase or all capital.

V. EXPERIMENT RESULT

A. Research Results

Training data used at the time of testing data taken from newsmedia.co.id, kapanlagi.com, and tribunnews.com. Testing

the data, done by using artist news review (300 data training, consisting of 150 negative reviews and 150 positive reviews) then do testing and training dataset so get accuracy and AUC. The following will be explained in more detail about the results of research obtained.

Here are the stages of doing data processing are:

1. Data Collection

Review artist news, each grouped by way of being saved into a folder that is the positive folder and negative folder, then each document is given extension .txt so it can be opened with Notepad application.

2. Initial Data Processing (Preprocessing)

Here are the steps in preprocessing:

a. Tokenization

In this tokenization process, all the words in each document are collected and punctuated and removed if there are symbols, special characters or anything other than letters.

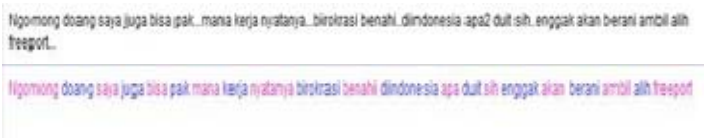

Fig. 1. Text Comparing before and after tokenization process

b. Transform Cases

In this process of transformation, all letters are changed into all lowercase or all capital letters.

Fig. 2. Text Comparing before and after transform cases process

B. Result Evaluation Analysis And Validation Model

Validation is the process to evaluate the predictive accuracy of the model. Validation is used to obtain predictions using existing models and then compare those results with known results, representing the most important step in the process of building a model[21].

1. Naïve Bayes Classsifier

The value of cycles training in this study was determined by testing C, epsilon. The following is the result of experiments that have been made for the determination of the value of training cycles.

TABLE II
NBC Cycles Training Value Determination Experiment

C	Epsilon	SVM	
		Accuracy	AUC
0.0	0.0	73.33%	0.774
0.1	0.1	71.00%	0.765
0.2	0.2	67.67%	0.739
0.3	0.3	71.33%	0.764
0.4	0.4	70.33%	0.764
0.5	0.5	73.33%	0.770
0.6	0.6	70.00%	0.768
0.7	0.7	71.33%	0.766
0.8	0.8	72.33%	0.762

The test results showed that the application of Naives Bayes Classifier method in Table IV.3 with C = 0.0 and Epsilon E = 0.0 produced Accuracy = 73.33% and AUC = 0.774.

The result of the model test is to classify the review of public opinion of negative artist news and public opinion review of positive artist news using Naive Bayes Classifier algorithm on the RapidMiner framework with the following model design:

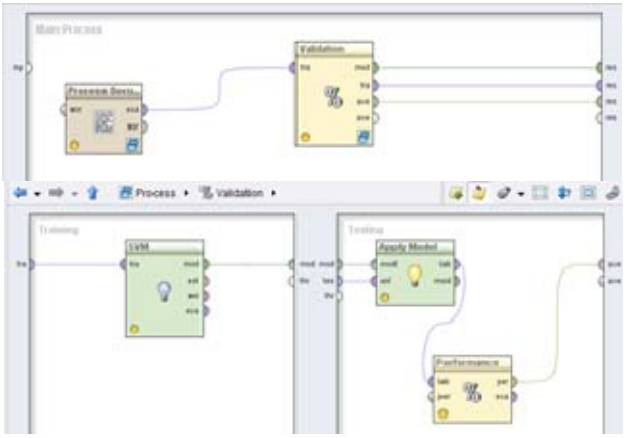

Fig. 3. Naiive Bayes Classifier Validation Model Design

a. Confusion Matrix

Providing decisions gained in training and testing, confusion matrix provides an assessment of classification performance based on true or false objects. Confusion matrix contains actual information (actual) and predicted (predicted) on the classification system

TABLE III
Confusion Matrix Naives Bayes Classifier

Accuracy: 67% +/- 12.29% (micro: 67.00%)			
	Negative true	Positive true	Class precision
Negative pred.	65	31	67.71%
Positive pred.	35	69	66.35%
Class recall	65.00%	69.00%	

$$\text{Acc (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{91+129}{91+21+59+129} = \frac{220}{300} = 0.73$$

b. ROC Curve

The ROC (Receiver Operating Characteristic) curve is another way to evaluate the accuracy of the classification visually. A ROC graph is a two-dimensional plot with a false positive proportion on the X-axis and is positively correct on the Y-axis. The calculation results on the ROC curve, depicting the ROC curve for the Naives Bayes Classifier algorithm. It can be concluded that one point on the ROC curve is better than the other if the direction of the transverse line from the lower left to the top is in the graph. ROC curve Naives Bayes Classifier with the value of AUC (Area Under Curve) of 0.774 where the diagnosis of the results fair classification. The following can be seen ROC curve of Naives Bayes Classifier in Figure 4.

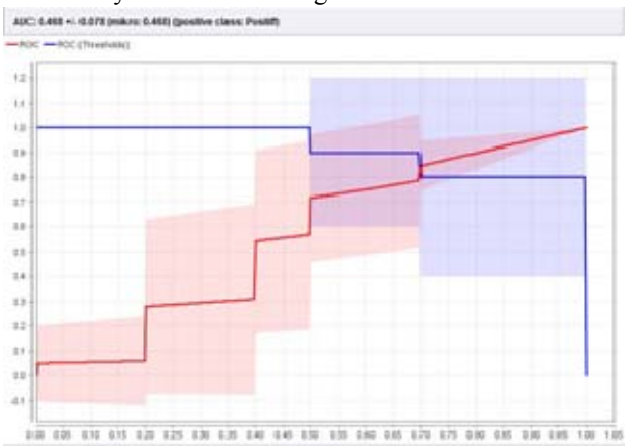

Fig. 4. ROC Curve Naïve Bayes Classifier

2. Naïve Bayes Classifier based on Particle Swarm Optimization

The value of cycles training in this study was determined by experimenting with C, epsilon and population size. The following is the result of experiments that have been made for the determination of the value of training cycles.

TABLE IV
NBC Cycles Training Determination Experiment Value Based on PSO

C	Epsilon	SVM		Population Size	C	Epsilon	SVM + PSO	
		Accuracy	AUC				Accuracy	AUC
0.0	0.0	73.33%	0.774	5	0.0	0.0	76.00%	0.794
0.1	0.1	71.00%	0.765	5	0.1	0.1	72.67%	0.783
0.2	0.2	67.67%	0.739	5	0.2	0.2	73.00%	0.771
0.3	0.3	71.33%	0.764	5	0.3	0.3	74.67%	0.781
0.4	0.4	70.33%	0.764	5	0.4	0.4	75.00%	0.792
0.5	0.5	73.33%	0.770	5	0.5	0.5	75.00%	0.794
0.6	0.6	70.00%	0.768	5	0.6	0.6	73.67%	0.773
0.7	0.7	71.33%	0.766	5	0.7	0.7	75.00%	0.796
0.8	0.8	72.33%	0.762	5	0.8	0.8	75.67%	0.798
0.9	0.9	71.33%	0.763	5	0.9	0.9	72.67%	0.788
0.0	1.0	50.00%	0.500	5	0.0	1.0	50.00%	0.500
1.0	1.0	50.00%	0.500	5	1.0	1.0	50.00%	0.500
1.0	0.0	70.00%	0.762	5	1.0	0.0	73.33%	0.790

The best results in PSO-based NBC experiments above are C = 0.0 and Epsilon E = 0.0 and population size = 5 produced accuracy = 76.00% and AUC = 0.794. This shows that using Particle Swarm Optimization optimization can improve the accuracy of the better.

The results of testing data training method of Naive Bayes Classifier based on Particle Swarm Optimization using Set Role that serves to determine the field in the class and then given optimization using Particle Swarm Optimization for higher yield accuracy. The measurement of the accuracy will be translated through the ROC Curves and Confusion Matrix below:

a. Confussion Matrix

TABLE V
Confusion Matrix Naives Bayes Classifier Based On Particle Swarm Optimizer

Accuracy: 76.00% +/- 6.63% (micro: 76.00%)			
	Negative true	Positive true	Class precision
Negative pred.	71	19	78.89%
Positive pred.	29	81	73.64%
Class recall	71.00%	81.00%	

$$\text{Acc (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{97+131}{97+19+53+131} = \frac{228}{300} = 0.76$$

b. ROC Curve

Fig. 5. ROC Curve Naïve Bayes Classifier Based On Particle Swarm Optimization

The resulting ROC curve based on the data tested in the figure above shows that there is an increase in accuracy using a Naive Bayes Classifier based on Particle Swarm Optimization of 76.00% and AUC of 0.794

C. Experiment Result

Based on the tests that have been conducted on the review of public opinion news artist by using the method of Naive Bayes Classifier, Particle Swarm Optimization based. The application of Particle Swarm Optimization (PSO) is proven to increase accuracy in the classification of public opinion reviews of artist news to identify between positive reviews and negative reviews. If you already have a text classification model in the review it will make it easier for readers to know

positive reviews and negative reviews. Based on the review data that has been processed through RapidMiner, then the result separated into words. These words, each of which has a weight so that it can be seen which words are associated with sentiments that often appear and have the highest weight. Thus it can be known that such reviews include positive restaurant reviews and negative restaurant reviews. In this study, the results of Naive Bayes Classier (NBC) method calculation has Accuracy of 73.33% and AUC of 0.774 while Naïve Bayes Classifier method based on Particle Swarm Optimization (PSO) resulted in Accuracy of 76.00% and AUC of 0.794. This shows that the use of Particle Swarm Optimization optimization can increase the accuracy value.

VI. CONCLUSION

Based on the explanations described in the previous chapter, this research produces accuracy in the form of Confusion Matrix and ROC Curves. The accuracy is generated on the Naive Bayes Classifier algorithm of 73.33% and AUC of 0.774, while the Naives Bayes Classifier and Particle Swarm Optimization with 76.00% accuracy and AUC of 0.794. Thus it can be concluded that the application of optimization can improve accuracy. Models in Naïve Bayes Classifier and Particle Swarm Optimization can provide solutions to the problem of classification review public opinion news state officials to be more accurate and optimal.

REFERENCES

[1] M. Rushdi Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, and L. A. Ureña-López, "Experiments with SVM to classify opinions in different domains," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14799–14804, 2011.

[2] I. F. Rozi, S. H. Pramono, and E. A. Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *Electr. Power, Electron. Commun. Control. Informatics Semin.*, vol. 6, no. 1, pp. 37–43, 2012.

[3] H. Liu, H. Q. Tian, C. Chen, and Y. F. Li, "An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization," *Int. J. Electr. Power Energy Syst.*, vol. 52, no. 1, pp. 161–173, 2013.

[4] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Inf. Retr. Boston.*, vol. 2, pp. 271–278, 2008.

[5] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Bus. Horiz.*, vol. 53, no. 1, pp. 59–68, 2010.

[6] I. Habernal, T. Ptáček, and J. Steinberger, "Reprint of 'supervised sentiment analysis in Czech social media,'" in *Information Processing and Management*, 2015, vol. 51, no. 4, pp. 532–546.

[7] V. Chandani, R. S. Wahono, and . Purwanto, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 55–59, 2015.

[8] B. Kurniawan, S. Effendi, and O. S. Sitompul, "Klasifikasi Konten Berita Dengan Metode Text Mining," *J. Dunia Teknol. Inf.*, vol. 1, no. 1, pp. 14–19, 2012.

[9] B. Ramesh and J. G. R. Sathiaseelan, "An Advanced Multi Class Instance Selection based Support Vector Machine for Text Classification," *Procedia Comput. Sci.*, vol. 57, pp. 1124–1130, 2015.

[10] Z. Yao and C. Zhi-Min, "An Optimized NBC Approach in Text Classification," *Phys. Procedia*, vol. 24, pp. 1910–1914, 2012.

[11] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," in *Procedia Computer Science*, 2013, vol. 17, pp. 26–32.

[12] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4241–4251, 2013.

[13] S. J. Rivera, B. S. Minsker, D. B. Work, and D. Roth, "A text mining framework for advancing sustainability indicators," *Environ. Model. Softw.*, vol. 62, pp. 128–138, 2014.

[14] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[15] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," in *Procedia Engineering*, 2013, vol. 53, pp. 453–462.

[16] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Systems with Applications*, vol. 40, no. 2. pp. 621–633, 2013.

[17] M. K. A. J. P. Jiawei Han, "Data Mining: Concepts and Techniques, Third Edition - Books24x7," *Morgan Kaufmann Publ.*, p. 745, 2012.

[18] H. Hashimi, A. Hafez, and H. Mathkour, "Selection criteria for text mining approaches," *Comput. Human Behav.*, vol. 51, pp. 729–733, 2015.

[19] R. Feldman and J. Sanger, "The text mining handbook: advanced approaches in analyzing unstructured data," *Imagine*, vol. 34, p. 410, 2007.

[20] M. W. Berry and J. Kogan, *Text Mining: Applications and Theory*. 2010.

[21] L. R. Angga Ginanjar Mabrur, "Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit," *J. Komput. dan Inform.*, vol. 1, no. 1, pp. 53–57, 2012.