# Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter

Dinar Ajeng Kristiyanti STMIK Nusa Mandiri Jakarta Jakarta, Indonesia dinar@nusaman diri.ac.id Akhmad Hairul Umam Universitas Tanri Abeng Jakarta, Indonesia ahmad.umam@t au.ac.id Mochamad Wahyudi STMIK Nusa Mandiri Jakarta Jakarta, Indonesia wahyudi@nusa mandiri.ac.id Ruhul Amin STMIK Nusa Mandiri Jakarta Jakarta, Indonesia ruhul.ran@nusa mandiri.ac.id Linda Marlinda STMIK Nusa Mandiri Jakarta Jakarta, Indonesia linda.ldm@nusa mandiri.ac.id

Abstract- Opinion is a statement conveyed by a person or group of people in addressing a problem by providing predictions or expectations about the event. No guarantee that an opinion automatically will be true because it is not reinforced by the facts, it is subjective, and there is a different opinion about an event. Everyone has different views and same rights to express opinions or give opinions toward particular event. Public opinion is view of someone for certain problem comes out due to prior conversation with another person who may have an effect on the opinion given. Public opinion comes from a discussion process in addressing the problem then lead to a conclusion as a joint decision and a shared opinion. One of the media to convey public opinion is through social media like twitter. Public opinion about the election of West Java governor candidate period 2018-2023 on twitter was increasingly widespread. There are several sentiments emerged for four candidates elected on twitter accounts such as Ridwan Kamil-Uu Ruzhanul Ulum, Tubagus Hasanuddin-Anton Charliyan, Sudrajat-Ahmad Syaikhu, and Deddy Mizwar-Dedi Mulyadi. Therefore, it is necessary to classify the sentiments to the existing opinion so that it can be predicted in advance which of the governor candidate pair of West Java who has more positive and predictable sentiments will be elected as governor period 2018-2023. The data used by the researchers is tweet in Indonesian Language with keywords Rindu, Hasanah, Asvik, 2DM with datasets number is 800 tweets. There are many classification techniques commonly used for opinion sentiment analysis. This study compares two classification techniques namely Support Vector Machine Algorithm (SVM) and Naïve Bayes Classifier (NBC). The results show that the Algorithm of Naïve Bayes Classifier (NBC) has a higher accuracy level of Support Vector Machine (SVM), up to 94% for Deddy Mizwar-Dedi Mulvadi.

Keywords: West Java Governor Candidate; Opinion Public; Sentiment Analysis; Twitter.

# I. INTRODUCTION

Soon after the registration stages and stipulation of governor candidate pair of West Java 2018, a lot people both

in the real world and cyberspace are discussing the names of candidates. The election of West Java Governor will be held in June 2018 and the crowd has begun to be felt. The General Election Commission (KPU) of West Java officially declared four candidates of Governor and Vice Governor for West Java 2018-2023. This is in line with the results of the plenary meeting about the announcement of candidates for governor and vice governor election of West Java 2018 in Setia Permana Hall, West Java KPU, Garut Street, Bandung, Monday, February 12, 2018 [1]. [2] According to West Java KPU four candidate pairs for Governor of West Java 2018- 2023 are H. Moch. Ridwan Kamil, ST, M.UD (Major of Bandung 2013-2018) and H. Uu Ruzhanul Ulum, SE (Tasikmalaya Regent 2011-2021), Major General (Ret. H. Tubagus Hasanuddin, SE, MM (Vice Chairman of House Commission I 2014-2018) and Inspector General of Police (Ret. Drs. H. Anton Charliyan Amanah, MPKN (West Java Police 2016-2017), Sudrajat (Retired Army 2005) and Ahmad Syaikhu (Vice Mayor of Bekasi 2013-2018), and Deddy Mizwar (Deputy Governor of West Java Period 2013-2018) and Dedi Mulyadi (Purwakarta Regent 2008-2018). The schedule of election for Governor and Vice Governor of West Java will take place on 28 June 2018 [2]. Everyone is free to argue or give opinion about the candidate of West Java Governor 2018 so that raises a lot opinion not only positive or neutral but also negative opinion.

Lately social media become a tremendous trend. The role of social media is very influential for the development of the current global situation. According to Forrester Research, 75% of Internet surfers used Social Media in the second quarter of 2008 by joining social networks, reading blogs, or contributing reviews to shopping sites, this represents a significant rise from 56% in 2007 [3]. Its function currently no longer as a medium or means of friendship, looking for friends, fans page, self-existence, but has switched the

function as a medium of promotion for merchandise products, services, even today more than that promote political parties or campaigns of Legislative Candidates, Regents, Governor up to the President. Just like Facebook, Blackberry Messager (BBM), Whatsapp, Line, Youtube, Instagram, Google+, Tumblr, Linkedin and even Twitter are widely used ranging from ordinary people to among officials. Social media, especially Twitter is now a communication tool that is very popular among Internet users [4]. The developer of Chirp Twitter 2010 at the official conference, the company delivered statistics about Twitter's site and users. The statistics say that in April 2010, Twitter had 106 million accounts and 180 million unique visitors every month. The number of Twitter users stated continues to increase 300,000 users every day [5]. Social media especially Twitter is now one of the places of effective and efficient promotion or campaign [4].

Sentiments or opinion mining analysis is a computational study of people's opinions, behaviors and emotions to the entity [6]. The big effects and benefits of sentiment analysis led to research and application-based sentiment analysis growing rapidly [4]. Even in America there are about 20-30 companies that focus on sentiment analysis services [7]. Classification techniques commonly used for opinion sentiment analysis include Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) [8] Research related to the theory of sentiment analysis has been done mainly related to the general election among them are Sentimental Causal Rule Discovery From Twitter [8]. Sentiment Analysis Of Turkish Political News [9]. Sentiment Analysis in Czech Social Media Using Supervised Machine Learning [10]. Predicting Political Elections with Social Networks (The Case of Twitter in the 2012 U.S. Presidential Election) [11]. Analysis Candidates of Indonesian President 2014 with Five Class Attribute [12]. An Islamic Party in Urban Local Politics: The PKS Candidacy in 2012 for Jakarta Gubernatorial Election [13]. Ridwan Kamil for Mayor A study of a Political Figure on Twitter [14]. Citizens' Political Information Behaviors During Elections On Twitter In South Korea: Information Worlds Of Opinion Leaders [15]. Estimating Popularity by Sentiment and Polarization Classification on Social Media [16]. The Role of Religion and Ethnicity in Jakarta's 2012 Gubernatorial Election [17]. Sentiment Analysis of Smartphone Product Review Using Support Vector Machine Algorithm-Based Particle Swarm Optimization [6]. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization [18]. Feature Selection Based On Genetic Algorithm, Particle Swarm Optimization And Principal Component Analysis For Opinion Mining Cosmetic Product Review [19]. Sentiment Classification Of Online Reviews To Travel Destinations By Supervised Machine Learning Approaches [20]. Implementation of Sentiment Analysis Using Naive Bayes Algorithm Against the Election of Governor DKI Jakarta On Social Media Twitter [21]. Indonesian Presidential Election: Will Social Media Forecasts Prove Right? Naïve Bayes Classifier And Vector Machines

Support For Sentiment Analysis [23]. Sentiment Analysis of Governor Candidate of Jakarta Capital City 2017 On Twitter [4]. Text Mining on Social Media Twitter Case Study: The Quiet Period of DKI Governor Election 2017 Round 2 [24]. Twitter Sentiment Analysis Using the Naïve Bayes Method (Case Study of Governor Election for DKI Jakarta 2017) [25]. Sentiment Analysis About Opinion of DKI Governor Election 2017 On Indonesian Language Document Using Näive Bayes and Emoji Weighting [26].

With the increasing need of information organization and knowledge discovery from text data, many supervised learning algorithms have been used for text classification document. Among these methods, Naïve Bayes and Support Vector Machines (SVM) are always in the comparison list [27]. According to Thorsten Joachims in [27], SVM a discriminative classifier is considered the best text classification method to date. According to Tom Michael Mitchell in [27], Naïve Bayes a generative classifier is considered a simple but effective classification algorithm. Both have advantages and disadvantages in classification techniques.

In this study, Support Vector Machines (SVM) and Naïve Bayes algorithms will be applied as a method of classification techniques that researchers use to predict which of the candidate pairs who have more positive and predictable sentiments will be elected as Governor of West Java period 2018-2023. The researchers will compare the two algorithms to prove which algorithm of both has the best accuracy value in sentiment analysis.

# II. LITERATURE REVIEW

# A. Sentiment Analysis

Sentiment analysis refers to the general method to extract polarity and subjectivity from semantic orientation which refers to the strength of words and polarity text or phrases [28]. According James Spencer and Gulden Uchyigit in [29] Polarity refers to the most basic form, which is if a text or sentence is positive or negative. Sentiment analysis aims to extract opinions towards a topic generally from textual data sources [8]. The sentiment can be found in the comments or tweet to provide useful indicators for many different purposes [30].

# B. Opinion Mining

Collecting opinions of people about products and about social and political events and problems through the Web is becoming increasingly popular every day. The opinions of users are helpful for the public and for stakeholders when making certain decisions [31].

Opinion mining refers to the broad area of natural language processing, text mining, computational linguistics, which involves the computational study of sentiments, opinions and emotions expressed in text [32]. Millions of users share opinions on different aspects of life everyday [33]. Several messages express opinions about events, products, and services, political views or even their author's emotional state and mood [34]. According to Dave et al. in [35], the ideal opinion-mining tool would process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good).

# C. Twitter

Twitter, the most popular micro blogging platform, for the task of sentiment analysis [33]. Twitter is a popular micro blogging service where users create status messages (called "tweets") [7]. Twitter contains an enormous number of text posts and it grows every day. The collected corpus can be arbitrarily large. Twitter's audience varies from regular users to celebrities, company representatives, politicians, and even country presidents. Therefore, it is possible to collect text posts of users from different social and interests groups. Users represent twitter's audience from many countries. Although users from U.S. are prevailing, it is possible to collect data in different languages [33].

# D. Twitter Sentiment Analysis

Twitter has limited for a small number of words which are designed for the quick transmission of information or exchange of opinion [29]. Sentiment analysis is a natural language processing techniques to quantify an expressed opinion or sentiment within a selection of tweets [32]. Also, [36] stated that a sentiment can be categorized into two groups, which is negative and positive words.

#### E. Algorithm Support Vector Machine

SVM is a supervised learning method that analyze the data and recognize patterns that are used for classification [37]. SVM has the advantage of being able to identify separated hyperplane that maximizes the margin between the two different classes [38]. SVM is the best choice to get good results, because SVM can efficiently and effectively process the collection [36]. According Kai-Zhu Huang et al. in [19] With task-oriented, strong, tractable nature of computing, SVM has achieved great success and is considered a state ofthe art classifier today. Support Vector Machine is to detect the sentiments of tweets [39].

However Support Vector Machine has shortcomings on parameter election issues or suitable features [18]. Selection of features at once setup parameters in SVM significantly influence the results of the classification accuracy [40].

### F. Naïve Bayes Classifier

This method is often used in text classification because of its simplicity and speed. Basically, Naïve Bayes makes the assumption that features (words) are generated independently of word position [9]. The Naïve Bayes algorithm are the most commonly employed classification techniques [30]. The Naïve Bayes algorithm is very simple and efficient [41]. In addition, the Naïve Bayes algorithm is a popular machine learning technique for text classification, and performs well on many domains [20]. However, Naïve Bayes is apparently flawed, which is very sensitive in feature selection [41]. The number of features that are too much in the classification process not only increases the calculation time but also decreases the accuracy [42].

# G. Evaluation and Validation Technique

Validation and Evaluation Technique for classification technique of sentiment analysis are as a tool to facilitate researcher in measuring result of research. There are many methods used to validate a model based on existing data, such as the holdout, random sub-sampling, cross-validation, stratified sampling, bootstrap and others [19]. Performance evaluation is done to test the classification results by measuring the truth-value of the system. The parameter used to measure truth-value is accuracy. Accuracy is the percentage of documents successfully classified by the system. All parameters for obtaining accuracy are obtained from Confusion Matrix according Christopher D Manning et al. in [43] in the following table:

Table 1. Confusion Matrix			
	Relevant	Non Relevant	
Retrivied	True Possitive (TP)	False Possitive (FP)	
Non Retrivied	False Negative (FN)	True Negative (TN)	

K-fold cross-validation is a validation technique with initial data randomly split into k sections mutually exclusive or "fold" [44]. ROC curves will be used to measure the AUC (Area Under the Curve). ROC curve divides a positive result in the y-axis and a negative result in the x-axis [45]. According Vecerlis Carlo in [19] Graph curve ROC (Receiver Operating Characteristic) is used to evaluate the accuracy classifier and to compare the different classification models. So the larger the area under the curve, the better the prediction results.

#### III. RESEARCH METHOD

#### A. Research Framework

This study starts from the problem in text classification on sentiment analysis using a classifier Support Vector Machine (SVM) and Naïve Bayes. From this research want to know which of the two algorithms that have the best performance in sentiment analysis. The data used in this research is sentiment analysis based on public opinion toward Governor candidate period 2018-2023 on twitter. data obtained from www.twitter.com from each twitter based on twitter account of candidate pair of West Java Governor Period 2018-2023 consisting of 100 positive and negative opinion. Preprocessing is performed with tokenization, Generate N-Gram and Stemming. While the classifier used is Support Vector Machine (SVM) and Naïve Bayes. 10 Fold Cross Validation testing will be done, accuracy algorithm will be measured using the Confusion Matrix and the processed data in the form of ROC curves. RapidMiner Version 5.3 is used as a tool in

measuring the accuracy of the experimental data is done in research.

B. Research Methodology

The research method that researchers do is experimental research method with the following stages:

1. Data Collection

Researchers used data taken from the site www.twitter.com based on twitter account of the pair of West Java Governor Candidate Period 2018-2023 as @ridwankamil and @uuruzhan by keyword or jargon #rindu, @tbhasanuddin and @antoncharlian with jargon #hasanah, @ mayjensudrajat and @syaikhu ahmad with jargon #asyik, @deddy mizwar and @dedimulyadi71 with jargon #2DM which each consist of 100 positive opinion and 100 negative opinion of tweets on each candidate with Indonesian text.

2. Preliminary Data Processing

This dataset in the preprocessing stage must go through 2 processes, namely:

a. Tokenization

Collecting all the words that appear and remove any punctuation or symbols that are not letters.

b. Generate N-grams

Combining adjectives that often appear to indicate sentiments such as the word "very" and the word "good". The word "good" is already showing the sentiment of positive opinion. The word "very" does not mean it stands alone. But if the two words are combined to be "very good", it will greatly reinforce the positive opinion. Researchers use a combination of three words, called 2-grams (Bigrams).

3. Method Used

The method that researchers compare is the Naïve Bayes algorithm and Support Vector Machine. Where the two methods are already very popular used in research in the field of sentiment analysis, text classification, or opinion mining.

4. Experiments and Testing Methods

For experimental research data, researchers used Rapid Miner Version 5.3 to process the data.

5. Evaluation and Validation of Results Validation is done by using 10 fold cross validation. While the measurement accuracy is measured by confusion matrix and ROC curve to measure the value of AUC.

### IV. RESULTS & DICUSSION

A. Results

The dataset in this study is a separate set of texts in the form of documents collected from Twitter with Crawling methods from social media of Twitter. The data taken only tweet in Indonesian, which is tweet with keywords **RINDU** for West Java Governor Candidate Period 2018-2023 Ridwan Kamil-Uu Ruzhanul Ulum, **HASANAH** for West Java Governor Candidate Period 2018-2023 Tubagus HasannudinAnton Charliyan Amanah, **ASYIK** for Candidate of West Java Governor Period 2018-2023 Sudajat-Ahmad Syaikhu and **2DM** for West Java Governor Candidate Period 2018-2023 Deddy Mizwar-Dedi Mulyadi. Data is randomly picked from either a regular user or an online medium on Twitter.

The training data used in this text classification consists of 200 tweets from each of the four candidates for West Java Governor Period 2018-2023 with Indonesian subtitles. Before being classified, the data must go through several stages of the process in order to be classified in the next process.



The above data will go into a classification process that is useful for defining a sentence as a member of a positive class or negative class based on the probability calculation values of the larger Bayes and SVM formulas. If the sentence probability result for a positive class is greater than the negative class, then the sentence belongs to the positive class. If the probability for the positive class is less than the negative class, then the sentence belongs to the negative class.

In this research, model testing is done by using technique of 10 cross validation. This process divides the data randomly into 10 sections. The testing process begins with the formation of models with data in the first section. The model will be tested in the remaining 9 sections of data. Then the accuracy process is calculated by looking at how much data has been correctly classified. Model test results will be discussed through Confusion Matrix to show how well the model is formed. Here is the Confusion Matrix for every West Java Governor Candidate Pair Period 2018-2023 with Support Vector Machine and Naïve Bayes algorithm.

Table 2. Model *Confusion Matrix* RINDU For Support Vector Machine Method

Accuracy: 74.50%, +/- 8.50% (Micro: 74.50%)			
	True Positive	True Negative	Class Precission
Predictions Positive	85	36	70.25%
Predictions Negative	15	64	81.01%
Class Recall	85.00%	64.00%	
Table 3. Model Confusion Matrix RINDU			

For Naïve Bayes Method

Accuracy: 89.00%, +/- 6.24% (Micro: 89.00%)			
	True Positive	True Negative	Class Precission
Predictions Positive	79	1	98.75%
Predictions Negative	21	99	82.50%
Class Recall	79.00%	99.00%	

Table 4. Mode	l Confusion	Matrix HASA	NAH
Accuracy: 63.50%, +/- 8.	38% (Micro:	63.50%)	
	True Positive	True Negative	Class Precission
Predictions Positive	30	3	90.91%
Predictions Negative	70	97	58.08%
Class Recall	30.00%	97.00%	
Table 5. Model Confusion Matrix HASANAH			
For Naïve Bayes Method			
Accuracy: 84.50%, +/- 5.22% (Micro: 84.50%)			
	True	True	Class
	Positive	Negative	Precission
Predictions Positive	92	23	80.00%
Predictions Negative	8	77	90.59%
Class Recall	92.00%	77.00%	
Table 6. Mo	del Confusio	n Matrix ASY	ΊK
For Suppo	ort Vector M	achine Method	ł
Accuracy: 72.00%, +/- 7.	14% (Micro:	72.00%)	
	True	True	Class
	Positive	Negative	Precission
Predictions Positive	96	52	64.86%
Predictions Negative	4	48	92.31%
Class Recall	96.00%	48.00%	
Table 7. Mo For	del <i>Confusio</i> Naïve Bayes	<i>n Matrix</i> ASY Method	ΊK
Accuracy: 87.00%, +/- 7.	48% (Micro:	87.00%)	
	True	True	Class
	Positive	Negative	Precission
Predictions Positive	90	16	84.91%
Predictions Negative	10	84	89.36%
Class Recall	90.00%	84.00%	
Table 8. M	odel Confusi	on Matrix 2D	M
For Suppo	ort Vector M	achine Method	1
Accuracy: 75.50.00%, +/	- 6.24% (Mici	ro: 89.00%)	
	True	True	Class
	Positive	Negative	Precission
Predictions Positive	62	11	84.93%
Predictions Negative	38	89	70.08%
Class Recall	62.00%	89.00%	
Table 9. M	odel Confusi	on Matrix 2D	M
FOr Accuracy: 94.00% ±/ 4	naive Dayes	• 04 00%	
Accuracy : 94.00 70, +/- 4	.20 70 (WHKF0	. 34.00 /0J	Class
		1 1 1 1 1	I SAME

**Predictions** Negative 10 90.74% **Class Recall** 90.00% 98.00% The researchers illustrate the comparison of classification results of the Support Vector Machine (SVM) with Naïve Bayes in Table 10 and Figure 2 in the Graph of Classification

Negative

98

Precission

97 83%

Positive

90

**Predictions Positive** 

Results Comparison.

Table 10. Comparison of Classification Results

Support Vector Machine (SVM)			
West Java Governor Candidate Period 2018-2023	Accuracy (%)	TP Rate	
RINDU	74.50	85	
HASANAH	63.50	30	
ASYIK	72.00	96	
2DM	75.50	62	
Naïve Bayes			
Naïve B	ayes		
Naïve B West Java Governor Candidate Period 2018-2023	ayes Accuracy (%)	TP Rate	
Naïve B West Java Governor Candidate Period 2018-2023 RINDU	ayes Accuracy (%) 89.00	TP Rate 79	
Naïve B West Java Governor Candidate Period 2018-2023 RINDU HASANAH	ayes Accuracy (%) 89.00 84.50	<b>TP Rate</b> 79 92	
Naïve B West Java Governor Candidate Period 2018-2023 RINDU HASANAH ASYIK	Accuracy (%) 89.00 84.50 87.00	<b>TP Rate</b> 79 92 90	



Fig. 2. Comparison of Classification Results West Java Governor Candidate

#### В. Discussion

The Naïve Bayes and Support Vector Machine algorithms are popular in classifying texts, especially for the field of sentimental analysis. Both have good performance and high accuracy results. However, each study uses different data, so the result of accuracy is vary. In this research Indonesian textual opinion data is used where generally the data used in English text. This is likely to be one of the causes of the accuracy of the Support Vector Machine algorithm is not optimal. In addition, the use of 2-grams generates more tested data than ever before. It can be seen that the Naïve Bayes algorithm excels with the highest accuracy of 94%. While the Support Vector Machine algorithm only produces the highest accuracy of 75.50%.

#### V. CONCLUSION

From the data processing that has been done, it is proven that Naïve Bayes algorithm is superior to the Support Vector Machine algorithm in classifying public opinion with Indonesian text in twitter to the candidate of Governor of West Java Period 2018-2023. The highest accuracy of the Naïve Bayes algorithm is 94% for the 2DM pair, while the Support Vector Machine algorithm produces only the highest accuracy of 75.50% for the 2DM pair as well. The difference in accuracy is quite far, at 18.50%. Based on the results of the above research, in addition to the superior Naïve Bayes algorithm, it can also be predicted to date based on opinion from the public on Twitter to the candidate of West Java Governor period 2018-2023 which has the most positive sentiment with the best accuracy is the couple 2DM Deddy Mizwar and Dedi Mulyadi.

#### REFERENCES

- [1] N. Nurulliah, "Empat Pasangan Calon Pilgub Jabar 2018 Resmi
- Ditetapkan," http://www.pikiran-rakyat.com, Bandung, Feb-2018.
- [2] P. J. B. KPU, "PENETAPAN PASANGAN CALON GUBERNUR DAN WAKIL GUBERNUR JAWA BARAT TAHUN 2018," 2018, p. 1.
- A. M. Kaplan and M. Haenlein, "Users of the world, unite! The [3]

challenges and opportunities of Social Media," *Bus. Horiz.*, vol. 53, no. 1, pp. 59–68, 2010.

- [4] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J. Maret*, vol. 1, no. 1, pp. 32–41,2016.
- [5] M. R. Yarrow, J. A. Clausen, and P. R. Robbins, "The Social Meaning of Mental Illness," *J. Soc. Issues*, vol. 11, no. 4, pp. 33–48, Oct. 2010.
- [6] M. Wahyudi and D. A. Kristiyanti, "Sentiment Analysis of Smartphone Product Review Using Support Vector Machine Algorithm-Based Particle Swarm Optimization.," J. Theor. Appl. Inf. Technol., vol. 91, no. 1, p. 189, 2016.
- [7] J. Spencer and G. Uchyigit, "Sentimentor: Sentiment analysis of twitter data," CEUR Workshop Proc., vol. 917, pp. 56–66, 2012.
- [8] R. Dehkharghani, H. Mercan, A. Javeed, and Y. Saygin, "Sentimental causal rule discovery from Twitter," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4950–4958, Aug. 2014.
- [9] M. Kaya, G. Fidan, and I. H. Toroslu, "Sentiment analysis of Turkish Political News," in *Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012*, 2012, pp. 174–180.
- [10] I. Habernal, "Sentiment Analysis in Czech Social Media Using Supervised Machine Learning," no. June, pp. 65–74, 2013.
- [11] S. Ceri, "Predicting Political Elections with Social Networks (The Case of Twitter in the 2012 U.S. Presidential Election)," Politecnico Di Milano, 2014.
- [12] G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, "Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute," *Int. J. Comput. Appl.*, vol. 136, no. 2, pp. 23–29, 2016.
- [13] S. Hidayat, "An Islamic Party in Urban Local Politics: The PKS Candidacy at the 2012 Jakarta Gubernatorial Election," J. Polit., vol. 2, pp. 5–40, 2016.
- [14] M. Iqbal, "Ridwan Kamil for Mayor A study of a Political Figure on Twitter," Stockholm University, 2016.
- [15] J. S. LEE, "Citizens' Political Information Behaviors During Elections On Twitter In South Korea: Information Worlds Of Opinion Leaders," The Florida State University, 2016.
- K. Charalampidou, "Estimating Popularity by Sentiment and Polarization Classification on Social Media," Delft University of Technology, 2012.
- [17] K. Miichi, "The Role of Religion and Ethnicity in Jakarta's 2012 Gubernatorial Election.," J. Curr. Southeast Asian Aff., vol. 33, no. 1, pp. 55–83, 2014.
- [18] 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," *Procedia Eng.*, vol. 53, pp. 453–462, Jan. 2013.
- [19] D. A. Kristiyanti and M. Wahyudi, "Feature Selection Based on Genetic Algorithm, Particle Swarm Optimization and Principal Component Analysis for Opinion Mining Cosmetic Product Review," in 2017 5th International Conference on Cyber and IT Service Management (CITSM), 2017.
- [20] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6527–6535, Apr. 2009.
- [21] A. S. Adinugroho, "Implementasi Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Pemilihan Gubernur DKI Jakarta Pada Media Sosial Twitter," UDINUS, 2016.
- [22] J. H. Yang, "Indonesian Presidential Election: Will Social Media Forecasts Prove Right?," *RSIS Comment.*, no. 120, pp. 1–3, 2014.
- [23] N. W. S. Saraswati, "Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis," *Semin. Nas. Sist. Inf. Indones.*, pp. 587–591, 2013.
- [24] A. F. Hadi, D. B. C. W, M. Hasan, and A. D. Penelitian, "Text Mining Pada Media Sosial Twitter Studi Kasus: Masa Tenang Pilkada Dki 2017 Putaran 2," 2017.
- [25] M. H. Rasyadi, "Analisis Sentimen Pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus Pemilihan Gubernur Dki Jakarta 2017)," Institut Pertanian Bogor, 2017.
- [26] A. R. T. Lestari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen

Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718– 1724, 2017.

- [27] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, Jun. 2011.
- [28] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.
- [29] A. Sarlan, C. Nadam, and S. Basri, "Twitter sentiment analysis," Conf. Proc. - 6th Int. Conf. Inf. Technol. Multimed. UNITEN Cultiv. Creat. Enabling Technol. Through Internet Things, ICIMU 2014, no. November 2014, pp. 212–216, 2015.
- [30] M. Annett and G. Kondrak, "A comparison of sentiment analysis techniques: Polarizing movie blogs," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5032 LNAI, no. Figure 1, pp. 25–35, 2008.
- [31] K. Khan, B. Baharudin, A. Khan, and A. Ullah, "Mining opinion components from unstructured reviews: A review," J. King Saud Univ. - Comput. Inf. Sci., May 2014.
- [32] T. Carpenter and T. Way, "Tracking Sentiment Analysis through Twitter," Proc. 2012 Int. Conf. Inf. Knowl. Eng., no. Figure 1, 2013.
- [33] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," *Ijarcce*, vol. 5, no. 12, pp. 320–322, 2016.
- [34] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and Combining Sentiment Analysis Methods," in *Proceedings of the first ACM conference on Online social networks*, 2014, pp. 27–38.
- [35] B. Pang and L. Lee, *Opinion Mining and Sentiment Analysis*, vol. 2, no. 1–2. Boston, USA: Foundations and Trends R? in Information Retrieval, 2008.
- [36] R. Prabowo and M. Thelwall, "Sentiment Analysis: A Combined Approach," J. Informetr., vol. 3, no. 2, pp. 143–157, 2009.
- [37] 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," *Procedia Eng.*, vol. 53, pp. 453–462, 2013.
- [38] J.-S. Chou, M.-Y. Cheng, Y.-W. Wu, and A.-D. Pham, "Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3955–3964, Jun. 2014.
- [39] S. Sharma, "Damage Detection Method Using Support Vector Machine and First Three Natural Frequencies for Shear Structures," WORCESTER POLYTECHNIC INSTITUTE In, 2008.
- [40] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5197–5204, May 2011.
- [41] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature selection for text classification with Naïve Bayes," *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 5432–5435, 2009.
- [42] A. K. Uysal and S. Gunal, "A novel probabilistic feature selection method for text classification," *Knowledge-Based Syst.*, vol. 36, pp. 226–235, 2012.
- [43] V. I. Santoso, G. Virginia, Y. Lukito, U. Kristen, and D. Wacana, "Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine," vol. 14, no. 1, pp. 79–83, 2017.
- [44] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*, vol. 278. Elsevier, 2011.
- [45] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Belanda: Morgan Kaufmann, 2016.