

# Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis

**Mohammad Nur Habibi<sup>1</sup>**

Department of Informatics, Widyatama University Bandung, Indonesia  
Email: nur.habibi@widyatama.ac.id

**Sunjana<sup>2</sup>**

Department of Informatics, Widyatama University Bandung, Indonesia  
Email: sunjana@widyatama.ac.id

Received: 13 September 2019; Accepted: 19 October 2019; Published: 08 November 2019

**Abstract**—The development of the Internet in Indonesia is quite rapid, it is marked by the increasing use of social networks, especially Twitter. Not only to share status or stories, Twitter has become become a means of promotion and campaign for elections. The Twitter data can be used to find out the political polarization in Indonesia that is needed in the 2019 presidential election. The method used in this research is sentiment analysis using naïve bayes classifier and social network analysis using the calculation of network attribute values and centrality values. 8.814 Twitter data was collected using data crawling method. The data are divided into three subsets consisting of jokowi's sentiments, prabowo's sentiments, and pilpres's sentiments. Final result of the sentiment analysis is classified sentiments into positives, negatives, and neutral. The average value of the classification results was 91.27% positive sentiment, 7.56% negatives sentiment, and 1.17% neutral sentiment. This classification yielded the average accuracy of 69.2% for jokowi's sentiments and 100% for prabowo sentiments. The classification accuracy calculation uses ROCs method. Final results of the social network analysis based on the calculation of network attributes yielded 277 nodes, 7.950 edges, 57,401 average degree, 56.44 average weighted degree, network diameter is 4, 1.853 average path length, 0.201 density, and 5 of number communities. Centrality values generates the 5 most influential actors in social network interactions are jokowi's of first rank, 2<sup>nd</sup> SBYudhoyono's, 3<sup>rd</sup> detikcom, 4<sup>th</sup> yuniardi, 5<sup>th</sup> mohmahfudmd.

**Index Terms**—Centrality values, classification, naïve bayes classifier, network attributes values, sentiment analysis, social network analysis

## I. INTRODUCTION

Less than a year before the Indonesia presidential election, however it appears that political polarization has strengthened in all strata of society. The community began

to split into two groups that faced each other. This can be seen in the action of attacking each other between the two groups either directly, such as in the Car Free Day area of Jakarta some time ago or indirectly such as on social networks. The rigors of abuse and attack on social networks do not rule out the possibility that this event could be the seeds of polarization that lead to harder physical friction. Like the polarization that occurred in the 2017 DKI Jakarta regional elections it is very likely to occur in the 2019 presidential election, even more likely to occur with greater resonance. There is great concern that if this political polarization does not stop growing it will damage trust in the government in the wider community. Regarding how to map community sentiment there are many ways that can be done, such as mapping community sentiment through social networks. The development of social networking users has made the availability of relevant data sources to analyze various community sentiments more accessible. Indirectly this causes sentiment analysis to be a very interesting research topic in web data mining. Sentiment analysis on this research is the process of classifying people's sentiments into three classes, namely positive sentiment class, negative sentiment class, and neutral sentiment class using the naïve bayes classifier method.

Social network analysis itself is a knowledge that studies the relationship between one entity unit with other entity units with the help of graph theory [1]. Methods and techniques of social network analysis were chosen because this method can provide a picture or visualization to the smallest relationship that occurs only one individual with other individuals in the network [2]. Using social network analysis, this research was conducted with the aim of evaluating by modeling information dissemination that occurs on the Twitter social network. Network visualization uses the undirected network method, then identifies influential actors who have a high number of interactions in the network by calculating the concept of the value of the centrality node (degree, betweenness, closeness, and eigenvector centrality) [2]. With this

mapping, the issues of political polarization that occur in social networks, especially Twitter can be known patterns and roots of the problem, as well as intellectual actors who hide behind it.

## II. RELATED WORK

Sentiment analysis Research on the classification of sentiments has been carried out by Bo Pang. In his journal, Bo Pang classified sentiments on film reviews using various machine learning techniques. The machine learning techniques used are Naïve Bayes, Maximum Entropy, and Support Vector Machines (SVM). In that study also used several approaches to perform feature extraction, namely: unigram, unigram, bigram, unigram, Part of Speech (POS), adjective, and position unigram. The results of experiments conducted in this study found that SVM became the best method when combined with unigram with an accuracy of 82.9% [17].

In a study conducted by Rozi, I. F., Pramono, S. H. & Dahlan, E. A. (2012) an opinion mining system was developed to analyze public opinion in tertiary institutions. In the subjectivity and target detection subprocesses, Part of Speech (POS) Tagging uses the Hidden Markov Model (HMM). In the results of the POS Tagging process, rules are applied to find out whether a document is an opinion or not, and to know which part of the sentence is the object of the opinion being targeted. Documents that are recognized as opinions are then classified into negative and positive opinions (subprocess opinion orientation) using the Naïve Bayes Classifier (NBC). Based on these tests obtained precision and recall values for the subjectivity document subprocesses are 0.99 and 0.88, for the target detection subprocesses are 0.92 and 0.93, and for opinion orientation subprocesses are 0.95 and 0.94 [16].

Research on classification was also carried out by Rodiyansyah, S. F. and Winarko Edi (2013) by conducting data mining techniques that are used to visualize traffic congestion in a city. In this research the method used is Naïve Bayes by combining prior knowledge with new knowledge. From the test results, the application shows that the smallest accuracy value of 78% is generated in testing with a sample of 100 and produces a high accuracy value of 91.60% in testing with a sample of 13106. The test results with Rapid Miner 5.1 tools obtained the smallest accuracy value of 72% with a sample of 100 and the highest accuracy value of 93.58% with a sample of 13106 for the Naive Bayesian Classification method, while for the Support Vector Machine method obtained the smallest accuracy value of 92% with a sample of 100 and the highest accuracy value of 99.11% with a sample of 13106 [18].

Other studies on the classification of film review sentiments were conducted by Chandani, V., Wahono, RS, & Purwanto (2015) by comparing Machine Learning classification methods such as Naïve Bayes (NB), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) and feature selection such as Information Gain, Chi Square, Forward Selection and

Backward Elimination. The comparative results of the SVM method get the best results with an accuracy of 81.10% and AUC 0.904. The results of the comparative feature selection Information Gain get the best results with an average accuracy of 84.57% and an average of AUC 0.899. The results of the integration of the best classification method and the best feature selection method resulted in an accuracy of 81.50% and AUC 0.929. These results have increased when compared to the results of experiments using SVM without feature selection. The results of testing the feature selection method is that Information Gain gets the best results for use on the NB, SVM and ANN methods [19].

### A. Sentiment Analysis

Sentiment analysis, also known as opinion mining is the method used for enabling computers to recognize and classifying opinions from big unstructured texts datasets with machine language and computer programming. Its main purpose to determine the context and emotion of online text data. Sentiment analysis, concept and techniques first introduced by Liu, B. He defined sentiments has quintuplet aspect, which are: target object, a feature of the object, the sentiment value of the opinion of the opinion holder, the polarity of the opinion and opinion, opinion holder and the time when the opinion is expressed [3]. The basic problem of opinion mining is opinion extraction. It is required to know the linguistic terms and get the idea from the text classification of subjective and objective terms identified by syntactic features. Another main focus is on subjectivity detection. Subjectivity is used to express the context or specific domain of online reviews [4].

Nowadays, sentiment analysis has become a large and growing field. In this sentiment analysis approach we used supervised methods, because its nature to be generally more accurate than unsupervised approaches. But On the other hand, this method requires labeled training data and the goal is to classify an online reviews as referring to one or more of the aspects. The research to Classification sentiment has been done by another researcher with various approaches and algorithm. Most of the algorithms for sentiment analysis are based on a classifier trained using a collection of annotated text data. Before training, data is preprocessed so as to extract the main features. Some classification methods have been proposed: Naïve Bayes Classifier, Support Vector Machine, K Nearest Neighbour, etc. However, because its varied result in various research, it is still not clear which of these classification strategies are the more appropriate to perform sentiment analysis system [5]. In this study, we design a sentiment analysis system to help gain information and knowledge regarding Indonesian online customer's opinion towards Distro's products and services using Naïve Bayes Classifier (NBC). NBC is a probabilistic learning algorithm that derives from Bayesian decision theory. NBC would combine previous knowledge with new knowledge. This classification algorithms are simple has performance similar to other approaches. In NBC, the probability of a message being in

class  $c$ ,  $P(cj)$ , is computed as shown in these equation formula:

$$P(cj | wi) = \frac{P(cj) \times P(wi | cj)}{P(wi)} \quad (1)$$

where  $P(cj | wi)$  is Posterior is an opportunity category  $j$  when there is the appearance of the word  $i$ .  $P(cj)$  Prior is an opportunity for the appearance of category  $j$ .  $P(wi | cj)$  conditional probability for word  $i$  to be classified for the  $j$  category.  $P(wi)$  is an opportunity for the appearance of a word.

### B. Social Network Analysis

In the social sciences, social network analysis has become a powerful agreement tool in addition to statistics, network concepts that have been approved, approved and applied in the tradition of research in the social sciences, ranging from anthropology, sociology, business administration and history. SNA focuses on the structure of human groups (groups), communities, organizations, societies, or systems that exist in the world (world systems). The basis of visualization of social networks is placed by researchers who refer to themselves as sociometric. J.L. Moreno composed a social science called sociometry, which is an interpersonal relationship, those who communicate society are not an aggregate of individuals and their characteristics, but are a structure of interpersonal relationships. Therefore, individuals must use basic social units. Society (social atom) consists of individuals and social, economic, or cultural relations and discussion into groups and ultimately consists of interrelated groups described in a sociogram (structure of relationships between groups).

There are several concepts in the approach of social network analysis, in addition to describing the patterns formed by the relationship between nodes or actors, SNA is more often used to determine the central node in a network, by calculating some centrality values including those commonly calculated are:

#### 1. Degree Centrality

Degree centrality is used to count the number of interactions a node has. To calculate the degree centrality of this node, you can use the following formula:

$$CD(ni) = d(ni) \quad (2)$$

Where  $d(ni)$  = the number of interactions that this node has with the other network nodes.

#### 2. Betweenness Centrality

Betweenness centrality calculates how often a node is passed by another node to go to a particular node in the network. This value serves to determine the role of the actor who is a bridge connecting interaction in the network. To calculate the degree centrality of a node, you can use the following formula:

$$CB(ni) = \sum \frac{gjk(ni)}{gjk} \quad (3)$$

Where  $gjk(ni)$  = number of shortest paths from node  $j$  to node  $k$  that passes through node  $i$ .

$gjk$  = number of shortest paths between 2 nodes in the network

#### 3. Closeness Centrality

Closeness centrality calculates the average distance between a node and all other nodes in the network or in other words measures the proximity of a node to another node. In a network with  $g$  nodes, the closeness of the centrality of this node is as follows:

$$CC(ni) = \left[ \frac{N-1}{\sum d(ni, nj)} \right] \quad (4)$$

$N$  = number of nodes in the network

$d(ni, nj)$  = the number of shortest paths connecting nodes  $ni$  and  $nj$

#### 4. Eigenvector centrality

Eigenvector centrality performs measurements that give higher weight to nodes connected to other nodes that also have high centrality values. To calculate the centrality eigenvector value of a node can be done using the following formula:

$$\begin{aligned} Ci(\beta) &= \sum (\alpha + \beta cj) Aji \\ C(\beta) &= \alpha(I - \beta\alpha) - 1A1 \end{aligned} \quad (5)$$

$\alpha$  = normalized constant (vector scale).

$\beta$  = symbolizes how much a node has a centrality weight in a node which also has a high centrality value [1]

### III. METHODOLOGY

The methodology used in this study is experimental research. The basis for making this research experimental research is a form of Indonesian political polarization in the 2019 presidential election, research focused in depth on public sentiment through Twitter accounts namely @jokowi and @prabowo which are used as a comparative analysis.

This model is divided into two submodels as describe on fig. 1 down below. First submodel is sentiment analysis process. This submodel doing the classification process on input dataset. The second process is social network analysis. This submodel doing the social network visualization and find the influencer around social network.

#### A. Data Crawling using Twitter API

Develop a Twitter API for crawling the tweets. The Twitter API directly communicate with the source. The Access keys and tokens are established that helps in communication over Twitter Server. The data is

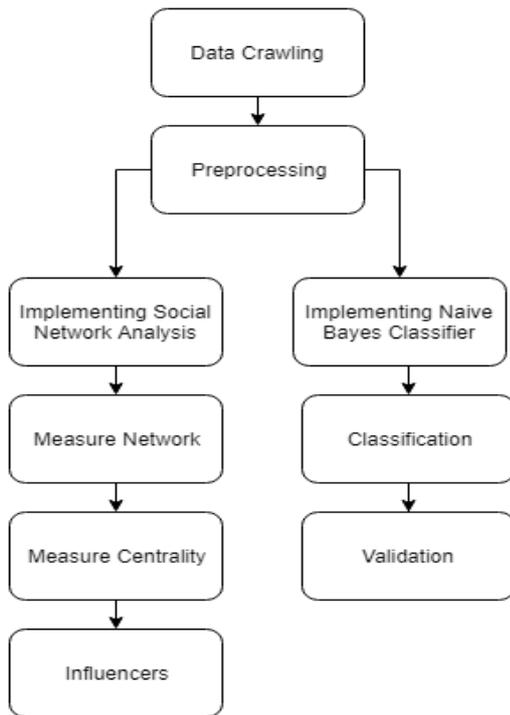


Fig. 1 Research Process

Indonesian political polarization tweet. This data was obtained from April 1 2019 to April 16 2019, exactly a day before the Indonesia president elections held on April 17 2019. The process of crawling data was carried out using Rstudio and R programming language.

**B. Preprocessing for Sentiment Analysis**

The data from Twitter contains various type of contents such as url, emoticons, spaces, hashtag, etc. which should be removed before next process. This process can highly increase the accuracy of result.

**1. Document Filtering**

This process aims to eliminate words or symbols that are not needed to reduce noise during the classification process, so that the list of words stored by the machine becomes more effective and efficient for the next process. The words and symbols that were omitted include:

- a. URL (*http://*)
- b. RT (RT)
- c. Mention (*@username*)
- d. Hashtag (*#*)

**2. Case Folding**

Text in tweets tends to have a variety of writing, one of which is writing uppercase and lowercase letters. The solution to this problem is that the text will be changed in lowercase letters. Another purpose of the case folding process is to equalize the context of a word, because the machine will assume two different words based on the use of uppercase or lowercase letters even though the shape is the same.

**3. Remove Duplicate**

This process uses the remove duplicate operator to handle duplicate sentences. The remove duplicate process is performed on the text attribute. The following are the results of the Tweet after going through the process of removing duplicates. The delete duplicate operator will eliminate duplicate sentences and save the last Tweets saved according to the order of the lines on the system.

**4. Tokenization**

This process will break up a set of characters in a text into units of words. This is done by distinguishing certain characters that can be treated as separators. Examples of separating characters such as whitespace, enter, tabulation, and spaces are considered word separators.

Table 1. Illustration of preprocessing

Process	Text
Data Example	RT @adianna70fans: Adian Napitupulu Praised Jokowi about the Land Title Deed Distribution Program, because it was real work #realwork <"ahref = http://twitter.com"> Twitter for Android
URL Remove	RT @adianna70fans: Adian Napitupulu Praised Jokowi about the Land Title Deed Distribution Program, because it was real work #realwork
RT Remove	@adianna70fans: Adian Napitupulu Praised Jokowi about the Land Title Deed Distribution Program, because it was real work #realwork
Mention Remove	Adian Napitupulu Praised Jokowi about the Land Title Deed Distribution Program, because it was real work #realwork
Hashtag Remove	Adian Napitupulu Praised Jokowi about the Land Title Deed Distribution Program, because it was real work
Case Folding	adian napitupulu praised jokowi about the land title deed distribution program, because it was real work
Tokenization	adian napitupulu praised jokowi about the land title deed distribution program, because it was real work

Before Remove Duplicate

Status	Date	Text
Removed	5 April 2019 06:02:33	adian napitupulu praised jokowi about the land title deed distribution program, because it was real work
	5 April 2019 06:08:54	adian napitupulu praised jokowi about the land title deed distribution program, because it was real work

After Remove Duplicate

Date	Text
5 April 2019 06:08:54	adian napitupulu praised jokowi about the land title deed distribution program, because it was real work

C. Preprocessing for Social Network Analysis

In this process we use twecoll for fetching data from Twitter. Twecoll is a Twitter command-line tool written in Python. It can be used to retrieve data from Twitter and purge likes (its only data-altering feature). It is based on a sub-command principle meaning calls to twecoll are based on a keyword which instructs twecoll what to do. Running twecoll requires Python 2.7 and the argparse library. Next step after got all the necessary data is combine it. We do that by using the edgelist command. Per default it creates a .gml file and tries to create a visualization with python package igraph. The igraph library is optional and is used to generate a clustered graph of the network.

The stages in this process can be explained as follows:

1. Identification problem: Indonesia political polarization ahead of the 2019 president election.
2. Data extraction: extracting and collecting data uses data crawling.
3. Preprocessing: data preparation which eliminates noise in data.
4. Implementating: processing the network interactions, the interaction patterns are visualized using undirected graph.
5. Network Measure: calculating network attribute values that counts eight attributes such as total nodes, total edges, average degree, average weighted degree, network diameter, average path length, density, and number of communities.
6. Centrality Measure: calculating centrality value such as harmonic closeness, page rank, closeness centrality, and eigenvector centrality.
7. Rank: ranking the centrality values in the network to find influencers.

synonym and its polarity. The synonym represents the similar word meaning which will be having same polarity. The polarity represents the positivity and negativity of the word in the context of the sentence.

- Steps:
1. Data Crawling (data were crawled using API Twitter divided into 2 datasets).
  2. Positive and negative sentiments were kept in two files pos.txt and neg.txt as data training
  3. Preprocessing
  4. Implementing naïve bayes classifier for calculated the score of each sentiment
  5. The classifier was trained using the dataset just prepared
  6. Labelled sentiments were kept correctly in reference sets and the predicatively labelled version in data testing
  7. Metrics were calculated accordingly

The result is down below:

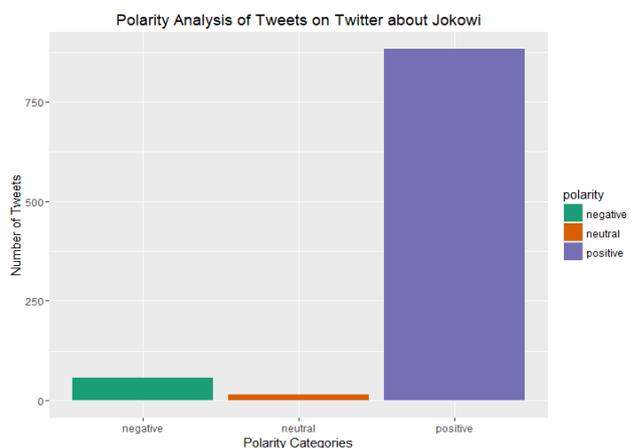


Fig. 2. Jokowi's Polarity

IV. IMPLEMENTATION

A. Implementing Naïve Bayes Classifier

To implement naïve bayes classifier we need a trained word dictionary. It consists of collection of word with its

The picture above shows the results of the classification sentiments towards Jokowi in jokowi dataset. The amount of positive sentiment acquisition tends to be lame with negative sentiment. Positive sentiment towards Jokowi amounted to 882 or 92.65% of the total sentiment, while for negative sentiment it amounted to 56 or 5.88%, and for neutral sentiment it amounted to 14 or 1.47% of the total sentiment implemented. Based on the above results it can be concluded that the majority sentiment for Jokowi is positive. The following is an example of sentiment from Jokowi's sentiment classification results.

Table 2. Jokowi's Sentiment Classification

Sentiment	Text
Positive	good luck for president joko widodo, whom we love
Negative	perpetrators who threaten beheaded jokowi admitted i was emotional
Neutral	president joko widodo jokowi launched the indonesian sharia economic master plan

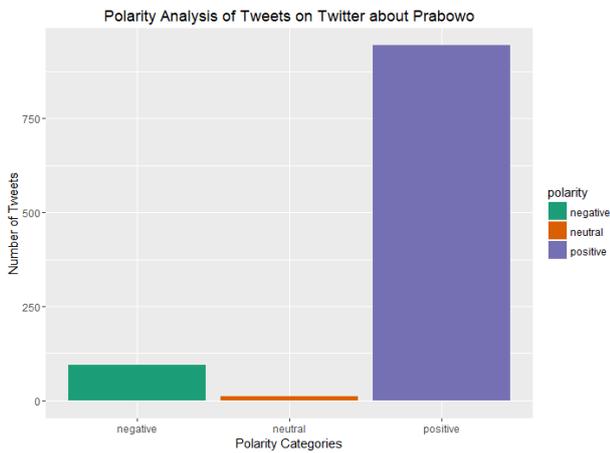


Fig. 3. Prabowo's Polarity

The picture above shows the results of the sentiment classification of Prabowo in prabowo dataset. The amount of positive sentiment obtained has almost the same value as Jokowi's sentiment. Positive sentiments towards Prabowo amounted to 943 or 89.99% of total sentiments, while negative sentiments amounted to 95 or 9.06%, and neutral sentiments amounted to 10 or 0.95% of the total implemented sentiments. Based on the above results it can be concluded that the majority sentiment for Prabowo is positive. The following are examples of sentiments from the results of Prabowo's sentiment classification.

Table 3. Prabowo's Sentiment Classification

Sentiment	Text
Positive	god willing, prabowo and sandiaga is just waiting for the official time to be announced
Negative	don't be afraid of scaring us with weapons given by the folks
Neutral	president prabowo subianto's will to all indonesians

**B. Implementing ROC Validation**

Testing the results of the sentiment classification is done with the Receiver Operating Characteristic (ROC) curve. This curve shows accuracy and compares the visual classification which shows the true positive level and the false positive rate. The ROC curve has a diagnostic value, which is:

- A. Accuracy is 0.90 - 1.00 = very good classification
- b. Accuracy is worth 0.80 to 0.90 = good classification
- c. The accuracy is 0.70 - 0.80 = fair classification
- d. Accuracy is 0.60 - 0.70 = bad classification
- e. Accuracy is 0.50 - 0.60 = failure

The results obtained from ROC processing for the results of the "Jokowi" sentiment analysis of 0.69 with a poor diagnosis classification level and for "Prabowo" sentiment of 1.00 with an excellent classification diagnosis level can be shown in the following below:

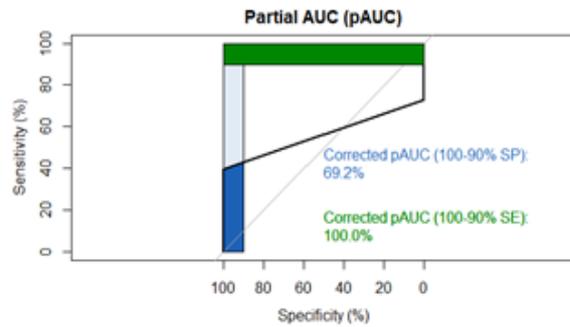


Fig.4. ROC Curve

**C. Implementing Social Network Analysis**

This process is carried out by visualizing the data interaction network of political polarization information dissemination that occurs in Indonesia on the social networking site Twitter by using the Gephi tool version 0.9.2. The data is visualized into a sociogram where the point in the image is called a node or node that represents an individual connected by a line called edge. Two connected nodes are represented by a line connecting the two. Network visualization using the Frucherman Reingold algorithm and Force Atlas. The following are the results of the process of implementing social network analysis.

Table 4. Network Attributes Value

Network Attributes	Value
Total Nodes	277
Total Edges	7.950
Average Degree	57.401
Average Weighted Degree	56.44
Network Diameter	4
Average Path Length	1.853
Density	0.201
Number of Communities	5

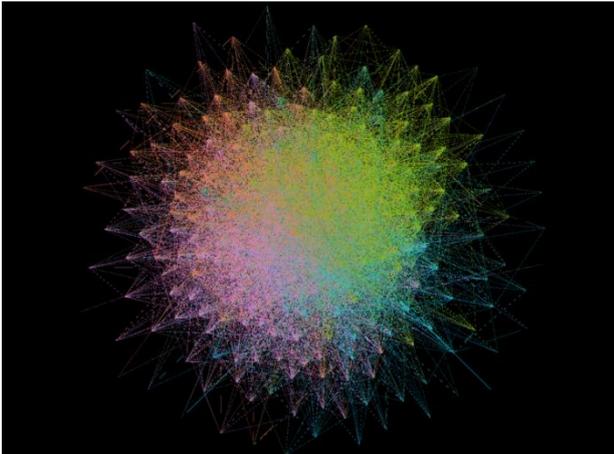


Fig.5 Fruchterman Reingold Implementation

The first visualization was done using the Fruchterman Reingold algorithm. The network formed is very difficult to see because the location of the nodes is very close together and difficult to identify with the naked eye. It can be concluded that these individuals are quite close to one another.

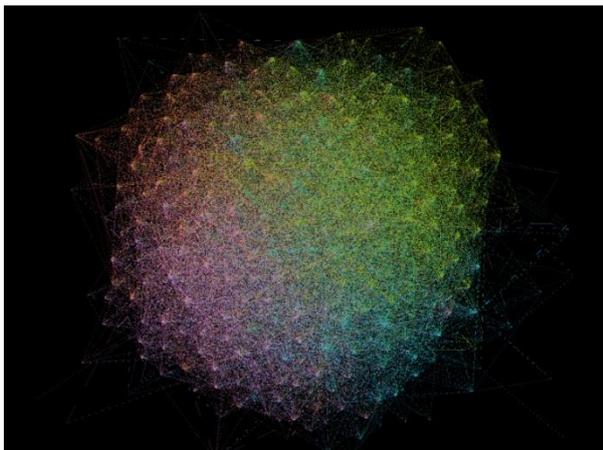


Fig.6 Force Atlas Implementation

The second visualization was done using the Force Atlas algorithm. The network that is formed although it is difficult to see it in plain sight, but the location between individuals tends to be more tenuous compared to using the Fruchterman Reingold algorithm, making it more possible to use the Force Atlas algorithm to implement complex data. Based on the results of the two algorithms above, the spread of network interactions is divided into five major communities. Here are the central actors in each community in the network the formula mentioned above, the results of the centrality calculation are as follows.

Table 5. Central Actor

Communities	Actor
Dark Green	SBYudhoyono
Light Green	Mohmahfudmd
Orange	Yjuniardi
Blue	Detikcom
Purple	Jokowi

Based on the calculation above, it is known that the Jokowi actor excels at all centrality values, so it can be said that the Jokowi actor is the most influential actor in the Indonesian political polarization interaction network on Twitter. The SBYudhoyono actor also has a stable calculation from the various centrality calculations above and it can be said that the SBYudhoyono actor was ranked second as an influential actor.

Table 6. Centrality Value

Actor	Betweenness Centrality	Degree Centrality	Closeness Centrality	Eigenvector Centrality
	Score	Score	Score	Score
SBYudhoyono	704.5	136	0.6618	0.9267
mohmahfudmd	432.4	120	0.6344	0.8340
yjuniardi	422.7	119	0.6359	0.8487
detikcom	712.9	132	0.6524	0.8577
jokowi	1087.7	156	0.6969	1.0

V. RESULT AND FUTURE WORKS

Data collection on Indonesian political polarization using data crawling, the programming language used is R language. Based on the results of sentiment classification of Indonesian political polarization, on the jokowi's dataset (92.65% positive sentiment, 5.88% negative sentiment, and 1.47% neutral sentiment) and prabowo's

dataset (89.99% positive sentiment, 9.06% negative sentiment, and 0.95% neutral sentiment). The accuracy results obtained from ROC processing for the jokowi's dataset is 0.69 with a poor diagnosis classification level and for prabowo's dataset is 1.00 with an excellent classification. One important thing that affects the outcome of accuracy is the amount of training data used, the more data the training used will be the higher the

accuracy. Mistakes in preprocessing can also affect overall accuracy. The results of the representation of relations between Indonesian political entities produced 5 large communities in the network. Interconnected actors numbered 277 and 7,950 relations. Based on the calculation of centrality value, the account of Jokowi is first rank as the account that most influences the course of interaction in the network. The second rank is occupied by the SBYudhoyono account and the detikcom account as third rank.

Suggestions for further research are as follows: It is necessary to use the server to retrieve data from the crawler automatically every time, this needs to be done because Twitter API limits the time of withdrawal of tweets. Choosing tweets that really contain community opinion according to the topic. Add a collection of words to the data dictionary so that the classification provided can be more accurate. Adding several stages in preprocessing, namely eliminating repetition of sentences and adding abbreviations and non-formal language that is not in accordance with the dictionary.

#### REFERENCES

- [1] Asri Maspupah, Asep ID Hadiana. (2018). *Pemetaan Bidang Keilmuan Organisasi Dengan Social Network Analysis*. Universitas Jenderal Ahmad Yani, Bandung.
- [2] Mahdi Shiddieqy, Dodie Tricahyono. (2017). Implementasi Social Network Analysis dalam Penyebaran Country Branding “Wonderful Indonesia”. Universitas Telkom, Bandung.
- [3] Liu B, “sentiment analysis and subjectivity, Hanbook of natural language processing”, vol.2, pp 627-666, 2010
- [4] S Padmajal and Prof. S. Sameen Fatima, “Opinion Mining and Sentiment Analysis -An assessment of People’s Belief: A Survey, International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol. 4 No. 1, February 2013
- [5] G. Fiarni, C., Maharani, H., & Pratama, R. (2016). Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique. 2016 4th International Conference on Information and Communication Technology (ICoICT)
- [6] Parveen, H., & Pandey, S. (2016). Sentiment analysis on Twitter Data-set using Naive Bayes algorithm. 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)
- [7] Yusainy Cleoputri, Anif Fatma Chawa, Siti Kholifah. (2017). Social Data Analytics sebagai Metode Alternatif dalam Riset Psikologi. Universitas Brawijaya, Malang.
- [8] Hu Jie, Junchi Zhang, Mengchi Liu. (2014). A Semantic Model for Academic Social Network Analysis. Hubei University, Wuhan.
- [9] Jaka Eka S, Erwin Budi S., Abdurahman Baizal (2018). Data Crawling pada Twitter. Telkom University, Bandung.
- [10] Liu Guojun, Zhang Ming, Fei Yan. (2010). Large – Scale Social Network Analysis based on MapReduce. Peking University, Beijing.
- [11] Akhtar Nadeem. (2014). Social Network Analysis Tools. Aligarh Muslim University. Aligarh
- [12] Cheliotis, D. G. (2010). Social Network Analysis (SNA). National University of Singapore, Singapore.
- [13] The Computer Advisor. Web site scraper the most effective tool for web data extraction, Retrieved November 28, 2018 from: <http://www.thecomputeradvisor.net/web-site-scraper-the-most-effective-tool-for-web-data-extraction/>
- [14] Vatrapu Ravi, Raghava Rao, Abid Hussain. (2016). Social Set Analysis: A Set Theoretical Approach to Big Data. Copenhagen Business School, Computational Social Science Laboratory, Frederiksberg, Denmark
- [15] Mujilawati Siti. (2016). Pre-Processing Text Mining pada Data Twitter. Universitas Islam Lamongan, Lamongan.
- [16] Rozi, Imam Fahrur; Pramono, Sholeh Hadi, Dahlan, Erfan Achmad. Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. Jurnal EECCIS, [S.l.], v. 6, n. 1, p. pp. 37-43, mar. 2013. ISSN 2460-8122. Available at: <<https://jurnaleeccis.ub.ac.id/index.php/eccis/article/view/164/142>>. Date accessed: 26 Maret 2019
- [17] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Sentiment Classification Using Machine Learning Techniques. EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 Pages 79-86.
- [18] Rodiyansyah, Sandi Fajar; Winarko, Edi. Klasifikasi Postinging Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), [S.l.], v. 6, n. 1, jan. 2013. ISSN 2460-7258. Available at: <<https://jurnal.ugm.ac.id/ijccs/article/view/2144>>. Date accessed: 26 Maret. 2019.
- [19] Vinita Chandani, Romi Satria Wahono, Purwanto. 2015. Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film. Universitas Dian Nuswantoro, Semarang.
- [20] Nur Azizah Vidya. 2015. Twitter Sentiment Analysis Terhadap Brand Reputation: Studi Kasus PT XL AXIATA Tbk. Universitas Indonesia, Jakarta.
- [21] Jurafsky, D S. (2000). Speech and Language Processing "An Introduction to Natural Language Processing, Cmputationak Linguistics, and Specch Reconition. Prentice-Hall, Inc. New Jersey.
- [22] M. Yusuf Nur Sumarno Putro. 2011 “Analisis Sentimen pada Dokumen berbahasa Indonesia dengan Pendekatan Support Vector Machine”. Masters, Binus.
- [23] Rustiana, deden. 2017. Analisis sentiment pasar otomotif mobil: tweet twitter Jurnal simetris, vol 8. No 1. April 2017
- [24] Novantirani A, Sabariah MK, Effendy V. 2004. Analisis Sentimen pada Twitter mengenai Penggunaan Transportasi Umum Darat dalam Kota dengan Metode Support Vector Machine. Universitas Telkom, Bandung.
- [25] Putranti, Noviah Dwi, Winarko, Edi. Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), [S.l.], v. 8, n. 1, p. 91-100, jan. 2014. ISSN 2460-7258. Available at: <<https://jurnal.ugm.ac.id/ijccs/article/view/3499>>. Date accessed: 26 Maret. 2019.

**Authors' Profiles**

**Mohammad Nur Habibi:** He has completed his Bachelor of Computer Science from Widyatama University on Faculty of Engineering. His areas interest are data mining, business intelligence, and social network analysis.



**Sunjana:** He is a lecturer in Informatics at the Faculty of Engineering of Widyatama University. In teaching, he has been focusing in artificial intelligence, numerical method, and machine learning. In research, his current interests include numerical analysis, machine learning, and data science.

**How to cite this paper:** Mohammad Nur Habibi, Sunjana, " Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis ", International Journal of Modern Education and Computer Science(IJMECS), Vol.11, No.11, pp. 22-30, 2019.DOI: 10.5815/ijmeecs.2019.11.04