RESEARCH ARTICLE OPEN ACCESS

Sarcasm Detection on Indonesian Politics Tweet Using Multi Labeling Method and Support Vector Machine

Fariz Ardiansyah[1], Aghistina Kartikadewi[2]

1(Master of Computer Science, Faculty of Information TechnologyBudi Luhur University, Jakarta, Indonesia
Email: farizardiansyah60@gmail.com)

2(Master of Computer Science, Faculty of Information TechnologyBudi Luhur University, Jakarta, Indonesia
Email: aghistinakd@gmail.com)

Abstract:

Detecting a sentence of sarcasm is considered as one of the difficult problems in sentiment analysis. In observing Twitter's social media in Indonesia, for some topics such as politics, people tend to criticize something using sarcasm. In previous studies, many researchers used the corpus of SentiWordNet or commonly called the Lexical method which has many limitations both from the number of words and from the language used that is only English and must be translated first into the desired language. In this study the author made Sarcasm Detection not with the help of a corpus or data dictionary but with the help of manual labeling using the Multi Labeling method. For the modeling method the author uses the TF-IDF method to convert words or sentences into vectors and Support Vector Machines. The whole modeling process uses the python programming language while for the display uses the PHP programming language. The training data set or training data for the author's model is obtained from the Twitter API crawling results with the help of a crawler engine made with the java programming language. Then for testing the author use the Confusion Matrix to determine the level of accuracy, sensitivity, specificity and precision of the proposed model. The final result of this study is for the SVM algorithm get an accuracy value of 86% while Naive Bayes as a comparison method gets an accuracy value of 78%. This proves that using machine learning that uses training data is proven to be able to overcome the limitations of the Lexical method because the training data can be added as much as the writer wants if the model is judged to be poor and the author does not need to do two classifications and there is no need to change the language process in the dictionary as in the Lexical method.

Keywords — **Sarcasm Detection, Multi Labelling, Support Vector Machine.**

I. INTRODUCTION

According to the results of a survey conducted by We Are Social in 2018, active users of social networks in Indonesia reached around 130 million users and 27% or around 35 million users were active users on the Twitter social network [1]. This year in 2018 is the year where political turmoil is happening because the next year 2019 Presidential Election will be held. One way to find out public opinion about the 2019 Presidential Election is by conducting Analysis Sentiment. But in conducting Sentiment Analysis when viewed from the culture of users of social networks in Indonesia in writing sentences of opinion tweets towards pairs of presidential candidates more often towards the opposite of the true meaning of the sentence that is intended to insinuate someone or compare something like one presidential candidate with another presidential candidate or commonly called sarcasm tweet. As a result, the results of the Sentiment Analysis become less precise or not in accordance with the actual meaning of the tweet sentence. Therefore we need sentence sarcasm detection or commonly called Sarcasm Detection to help the sentiment analysis results so that the results are more precise and in accordance with the actual meaning of the sentence.According to the KamusBesar Bahasa Indonesia (KBBI) sarcasm, which means the use of words intended to offend others in the form of ridicule or rude ridicule [5].

In a previous study entitled Indonesian Social Media Sentiment Analysis with Sarcasm Detection, the Support Vector Machine produced better accuracy compared to other methods in Sentiment Analysis and Sarcasm Detection. In this study, researchers used the Lexical method or relied on a corpus or word dictionary to be a feature and divide the classification process into 2 (two) processes. First, classifying each tweet into 3 (three) sentiment

classes, namely positive, negative and neutral. Second, classifying sentences of sarcasm based on the results of positive sentiment only. The resulting accuracy is only up to 78% for Sentiment Analysis and 54% for Sarcasm Detection [2]. From these studies the author found a problem that is if relying on the Lexical method or the use of a corpus or word dictionary used as a feature does not help much in detecting sarcasm sentences because of the limited number of words in the corpus.

Then in a previous study entitled Contextualized Sarcasm Detection on Twitter, researchers said that hashtags or hash marks like #sarcasm did not mean that the tweet sentences were sarcasm sentences because it had to be seen in advance the purpose of the tweet writer embed the hashtag [3]. From this research, the writer found a problem, namely the help of hashtag like #sarcasm does not mean that the sentence is a sentence of sarcasm because it must be seen in advance whether the sentence is intended to insinuate someone or there is no specific purpose in embedding the fence sign.

In order to improve the accuracy of the results, the writer will use the Support Vector Machine method added with the Multi Labeling method and feature extraction with the TF-IDF method, the results of which will be used as features. Multi Labeling is a classification that produces multiple outputs or multiple outputs where multiple labels can exist on multiple instances or one instance can have two or more labels to solve a classification problem. Multi Labeling is a generalization of Multi Class, which is a problem in labeling a single in categorizing instances that can be categorized as having more than two classes [6]. Term Frequency - Inverse Document Frequency (TF-IDF) is a statistical-based weighting technique that is often applied in the process of extracting information. TF-IDF is used by giving weight to each sentence in a document. Sentence weight is obtained from the sum of the term weights in a sentence, where terms can be words, phrases or other syntetic types [7]. Support Vector Machines (SVM) are a class of linear algorithms that can be used for classification, regression, and other applications. In the case of the classification of the two simplest classes, SVM has a hyperplane that separates two data classes [8].

The author tests not only the level of accuracy of the modeling results, but also will see the sensitivity, specificity and precision of the model so that it can be known whether the results are better than previous studies or not. To be able to do so, the author divide the feature data into a training set and test set, the results of which will later be made into a Confusion Matrix so that the accuracy, sensitivity, specificity and precision of the modeling can be calculated and seen the success rate.Confusion Matrix is a useful tool for analyzing how well a classification model can correctly predict classification results on a large number of classes [9].

II. RESEARCH METHODS

The method in this research is descriptive which is to express data based on actual and comparative conditions that is comparing one method with another method to find out which method is more suitable or better for this research.

A. Collecting Data Method

Data collection methods used in this study the author divided into two, namely first literature studies such as reading journals, books, internet and other sources relating to the problem analyzed by researchers. Second, the tweet data crawler from the Twitter API is then used as data for the train set and test set.

B. SamplingSelection Method

The sampling method in this research is Systematic Random Sampling, which is random and systematic sampling using intervals in selecting research samples. The criteria for the selection of sampling in this study is the 5th tweet of thousands of results provided by the Twitter API that has been made into an excel file. So every tweet in the excel file that is ranked 5, 10, 15, 20 and so on that is what is used as a sample for this research.

B. Research Steps

In this research, there are 10 research steps that author use. These steps are explain in the figure below.

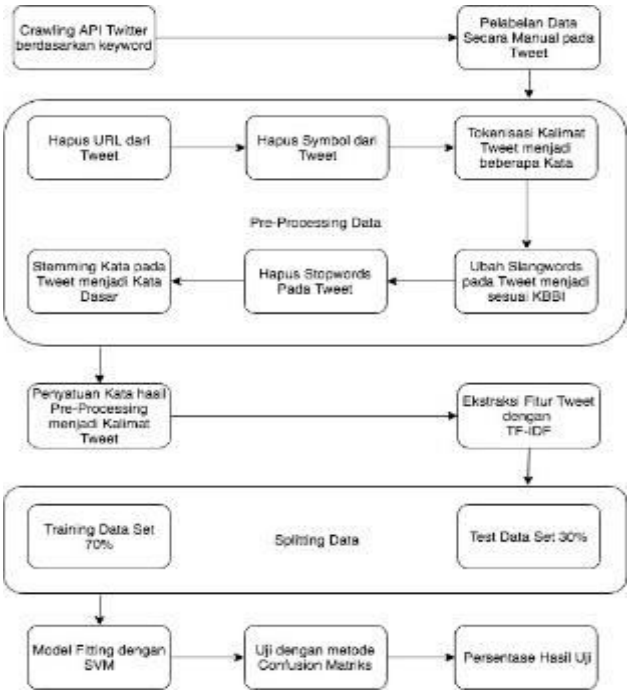

Fig. 1 Research Steps

1) In collecting data the author crawls into the Twitter API and receives data in the form of a collection of tweets. The programming language used to make crawlers is the Java programming language. After the data is received successfully, the data is moved into an Excel spreadsheet and placed in the tweet column.
2) There are two methods for making Sarcasm Detection that will be compared in this study, namely by using Multi Labeling with Support Vector Machines and with Naive Bayes. The programming language used to make Sarcasm Detection is the Python programming language. Crawling excel data is given 5 new columns, namely positive, negative, neutral, sarcasm, and non-sarcasm columns.
3) The next step for each method is to preprocessing data. This step is divided into 6 sections:
   a. Remove URL.
   b. Remove symbol and special character.
   c. Tokenization tweet sentence into words.
   d. Change slangwordsbased on KamusBesar Bahasa Indonesia (KBBI).
   e. Remove stopwords.
   f. Change affixed words become root words or stemming words.
4) Furthermore, after the collection of tweets on Excel successfully cleared of noise, reuniting the words was made into a tweet sentence and placed in a new Excel file.
5) Then feature extraction is performed using TF-IDF to convert tweet sentences to vectors.
6) Then the splitting data is carried out by dividing the data into training sets and test sets. In this study the author divided the data 80% for the training set and 20% for the test set in each method.
7) Then do the fitting or training on the model using Support Vector Machine and Naive Bayes. The data used in this training step is the training data set.
8) Next after the model has been trained, testing is done using the Confusion Matrix. The data used are training data sets and test sets.
9) The final result of this study is the percentage of test results in the form of a percentage of accuracy, sensitivity, specificity and precision of each method.

III. RESULTS AND DISCUSSION

Classification of tweets that the author made through several stages from starting to collect tweet data, tweet data processing, and several other stages. The following are the detailed stages of the classification that has been made.

A. Collecting Data

In collecting data, the author uses the help of a crawler engine that I made with the help of the Twitter API and java programming language. In crawling data, tweets are taken based on tweets containing usernames related to the accounts of several politicians, namely accounts from @jokowi, @prabowo, @rockygerung, @fahrihamzah, @fadlizon, and @bonihargens and hashtag #2019gantipresiden and #2019tetapjokowi. From the results of crawling obtained 1000 sample data tweet data which is then stored in an Excel file format.

B. Labeling Data

Labeling of crawling tweet data is done manually and using one of the Multi Labelinglabel methods, Label Powerset. Data is categorized into 5 (five) categories, namely positive, negative, neutral, non-sarcasm, and sarcasm. In categorizing the data, the writer gives the number 0 which means it does not fit into that category and the number 1 which means it belongs to that category.

TABLE 1Labeling Data Example

Tweet	Pos	Neg	Net	Non	Sar
PRESTASI GABENER Anies Di MusimHujaniniBisaMembuat Kolam RenangDi Tengah RumahWarga dg BanjirKirimanygbarokah. Banjirini BUKAN MUSIBAH JadiTakPerluDiantisipasi, Cukup dg Doa. FOKUS, Ayo TEBAK Benda-bendaapakahyglagimengambangitu? #Sarkas https://t.co/jX26iPpspe	1	0	0	0	1

The example at table above “PRESTASI GABENER Anies Di MusimHujaniniBisaMembuat Kolam RenangDi Tengah RumahWarga dg BanjirKirimanygbarokah. Banjirini BUKAN MUSIBAH JadiTakPerluDiantisipasi, Cukup dg Doa. FOKUS, Ayo TEBAK Benda-bendaapakahyglagimengambangitu? #Sarkas https://t.co/jX26iPpspe” is a tweet that has apositive sentiment category because it is expressed with good words like “bukanmusibah”, “prestasi”, and “barokah”. Although the tweet is categorized as positive sentiment, in reality the tweet is also included in the sentence of sarcasm because it means wanting to make an allusion to the governor's performance in dealing with floods. From the method of labeling above, each tweet will have a value of each whose form is a binary number, if you see the example above has a binary label 10001.

C. Pre-Processing Data

Data received from the Twitter API crawling produces tweets that are usually said to be dirty tweets because they still contain a lot of words or symbols that can become noise at the time of classification later. To reduce noise on tweets, data processing is usually done as pre-processing. In pre-processing, there are several stages so that the noise in a tweet can be minimized or reduced.

1) In Indonesia there is a term commonly used for a tweet in which there are strange characters or strange letters that are not commonly used in writing a tweet that is tweet “alayers”. These tweets often confuse people who read them. Therefore it is necessary to convert the tweet into normal tweet sentences.
2) In a tweet often accompanied by a URL. The URL is usually directed to an image, video or other site. URL is a noise because it is not needed in the classification of a text must therefore be removed.
3) Symbols and special characters are widely used in tweets and included as noise. Symbols that often appear include dots (.), Question marks (?), Exclamation points (!), Signs and (&), Dollar signs ($) and others.
4) Tokenisasi is a process of dividing or separating sentences into words.
5) In expressing an opinion or expression of feelings, the use of non-standard words or slangwords is a normal thing for the community to do. In cleaning tweets from nonstandard words the word should be noted because sometimes there are words that have the same meaning but the effect is not changed to make too many features later even though the word used has the same meaning or meaning as the word "afwan" with "maaf" has the same meaning and "ane" with “saya".
6) The collection of tweets where noise has been minimized will be used as a feature in this study. To be able to optimize the features obtained from a collection of tweets, words that often appear, or words that do not have meaning such as the conjunctions "yang", "itu", and others must be deleted. Words that often appear in a plain text are called stopwords.
7) At this stage the stemming is done or searching for basic words from the words in the tweet. Words that affect or have a prefix and a suffix will be changed with the basic word.

D. Feature Extraction

To recognize or classify objects / features in a text, first the extraction process must be carried out and then use the extracted feature to obtain classification. Feature extraction aims to look for areas of significant features in the text depending on their intrinsic characteristics. Extraction features tend to identify characteristics that can form a good representation of objects, so they can distinguish objects based on their respective categories. In this study the author used the help of the TF-IDF method to extract features.

E. Data Splitting

After performing the feature extraction step the next step is to divide the feature into 2 (two) types of datasets, namely training set and test set. The training set is the part of the dataset that we train to make predictions from the model. The test set is the part of the dataset that we test to see the accuracy or performance of the model. In this study the author use help to share data from the python library owned by sklearntrain_test_split.

In the train_test_split the definition process is divided into 4 (four) namely x_train as the data x we want to practice, x_test as the data x that we want to test, y_train is the dependent variable that is trained, and y_test is the dependent variable that is tested. We define these four variables together. In the train_test_split the data is filled with the arrays x parameter, y which shows the arrays used are x as an independent variable and y as the dependent variable. Next is the measurement for test data in this study the author uses 20% of data for the test set and 80% for the train set because usually as a general rule, the proportion of the test set is 20% and the train set is 80%. The author uses the help of the python library namely matplotlib to provide an overview of the condition of the feature data after splitting the data. The results are shown in figure below.

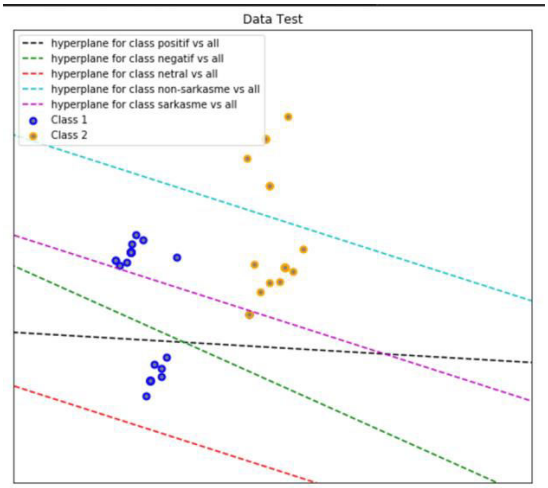

Fig. 2 Data Test Result

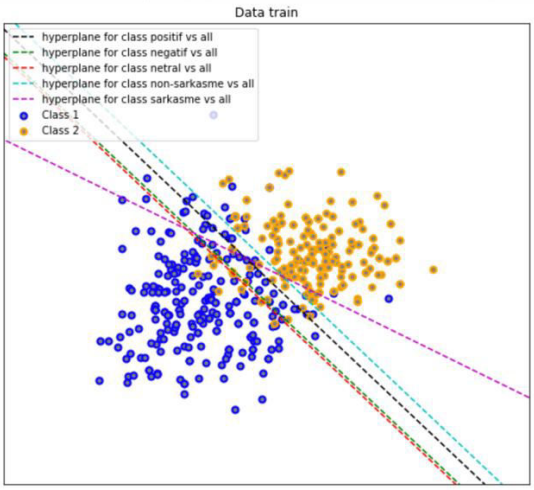

Fig. 3 Data Train Result

It can be seen that each data in the training data and test data are successfully separated by hyperplane lines, but there are still data with incorrect class separation. Therefore it is concluded that the data that I have are included in the type of soft margin.

F. Classification using SVM and Confusion Matrix

After conducting the data splitting phase, the next step is to carry out the classification process. In this study the author used a classification with machine learning techniques namely supervised learning and used the Support Vector Machine model and the Naive Bayes model as a comparison method.Supervised Learning is one of the techniques in Machine Learning that directly learns from the operational data of a system, which makes it possible to predict the system [4]. Classification

using the Support Vector Machine aims to find the best hyperplane so that data can be divided according to each class. The author divides the class in this study into 5 (five) classes based on sentiment and type of tweet sentences namely positive, negative, neutral, sarcasm and non-sarcasm. The results obtained are that after testing with the Confusion Matrix the comparison results are obtained in the following table below.

TABLE 2 Comparation Result of Testing

Comparasion Result of Two Models		
Criteria	SVM	NB
Accuracy	**86%**	78%
Sensitivity	**86%**	78%
Spesifity	**87%**	75%
Presition	**87%**	82%

It can be seen that the SVM model gets the highest score in each of the testing criteria and is at odds between 5% to 12%. Therefore, it can be concluded that the SVM method can make predictions and recognize the characteristics of the sample better and more thoroughly than the NB method in this study.

IV. CONCLUSIONS

Based on the results of the research that has been discussed, the Sarcasm Detection research that the author made can be concluded as follows.

1. The use of the Multi Labeling method proved to be able to be applied into Sarcasm Detection and can be integrated with the Support Vector Machine algorithm and Naive Bayes.
2. From the results of this study, Support Vector Machine became the leading algorithm with an accuracy value of 86%, beating Naïve Bayes which received an accuracy value of 78%.
3. By using machine learning, a supervised learning technique that uses training data is proven to be able to overcome limitations on the lexical method or data dictionary because training data can be added as much as the author wants if the model is judged to be poor and does not need to do two classifications and there is no need for a process change language as in the lexical method.
4. It is proven that a tweet without a hashtag similar to #sarcasm can also be included in the category of sarcasm because this study the author use tweet data with a political topic and without the hashtag.

The suggestions for further research development are as follows.

1. Call a linguistic expert to help labelling the data so that objectivity in labeling becomes better.
2. The development of Sarcasm Detection was tried using techniques other than Support Vector Machine and Naive Bayes.

REFERENCES

1. *Kemp, S. (2018) Digital In 2018: World's Internet Users Pass The 4 Billion Mark. Available at: https://wearesocial.com/blog/2018/01/global-digital-report-2018 (Accessed: 29 December 2018).*
2. *Lunando, E. and Purwarianti, A. (2013) 'Indonesian Social Media Sentiment Analysis with Sarcasm Detection', in 2013 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2013, pp. 195–198.*
3. *Bamman, D. and Smith, N. A. (2015) 'Contextualized Sarcasm Detection on Twitter', in Proceedings of the Ninth International AAAI Conference on Web and Social Media, pp. 574–577.*
4. *Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018) Foundations of Machine Learning. 2nd edn. MIT Press.*
5. *Setiawan, E. (2012) Arti kata sarkasme - KamusBesar Bahasa Indonesia (KBBI) Online. Available at: https://kbbi.web.id/sarkasme (Accessed: 29 December 2018).*
6. *Khade, S. R. and Balwan, S. R. (2017) 'Study and Analysis of Multi-Label Classification Methods in Data Mining', International Journal of Computer Applications, 159(9), pp. 975–8887.*
7. *Wahid, D. H. and SN, A. (2016) 'PeringkasanSentimenEsktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity', IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 10(2), p. 207.*
8. *Sammut, C. and Webb, G. I. (2017) Encyclopedia of Machine Learning and Data Mining, Sammut& Webb. 2nd edn. Springer US.*
9. *Han, J., Kamber, M. and Pei, J. (2012) Data Mining – Concepts & Techniques. 3rd edn. Morgan Kaufmann Publishers.*