Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

Sentiment Analysis On Twitter Using The Target-Dependent Approach And The Support Vector Machine (SVM) Method

Tansa Trisna Astono Putri[1], Mhd. Dominique Mendoza[2], Muhammad Fadhiel Alie[3]

[1,2] Universitas Negeri Medan, Jl. Williem Iskandar Ps.V Medan Estate, Sumatera Utara, Indonesia
[3] Universitas Indo Global Mandiri, Jl. Jenderal Sudirman No. 629 KM. 4 Palembang 30129

E-mail: tansatrisna@unimed.ac.id, aenaen@unimed.ac.id

ARTICLE INFO

Article history:
Received: 26 February 2020
Revised: 01 Maret 2020
Accepted: o1 May 2020

***Keywords**:*
sentiment analysis, target-dependent, support vector machine

ABSTRACT

Opinions or sentiments given by the public on Twitter social media have attracted attention for research. In this study, what will be done in connection with sentiment analysis on Twitter is to use a target-dependent approach. Where given a data twit or opinion relating to 3 major cities in Indonesia, namely Jakarta, Bandung and Medan, and will be classified as opinions that are positive, negative or neutral. Due to the social media used is Twitter, which can only write opinions in 140 characters, it is possible to find data on tweets or ambiguous opinions. Therefore, this study focuses on sentiment analysis with a target-dependent approach using the SVM method for classification. The target-dependent approach itself consists of 2 stages, namely determining matters relating to the topic of discussion and collecting data on tweets related to the topic. Broadly speaking, this research was carried out in 4 stages, namely preprocessing, classification with a target-dependent approach and SVM method and finally the determination of opinion classification)

1. Introduction

The internet that is developing now, makes it easier for the public to know more about the development of the environment. Existing social media now also supports the community to stay updated. Not a few people who use social media to shed all their hearts and minds about something that is experienced and thought, especially social media Twitter. Twitter is a micro blogging site operated by Twitter, Inc. Called micro blogging because this site allows users to send and read messages like blogs in general. The message is called tweets, which are 140 characters of written text displayed on the user's profile page.

Opinions given by the public in Twitter social media also vary. There are positive and negative. Research related to opinion analysis on social media is opinion mining or sentiment analysis. Sentiment analysis is used to analyze sentiments created by people through reviews from the internet. Sentiment analysis can influence a topic discussed in the community. In general, sentiment analysis helps to gather positive and negative aspects of a particular topic [9]. In the end, the final opinion value can be used as a parameter for the community on the topic.

This research focuses on Twitter sentiment analysis which applies a target-dependent classification approach in the process of grouping data and determining sentiment classification using the SVM method. The use of the target-dependent classification approach in this study was conducted to address the shortcomings of the classification of sentiments that normally use a target-independent approach where to classify sentiments, target-independent focus only on the topic as a keyword to select data to be managed without regard to the context of the opinion. Meanwhile, if we explore deeper, there are a number of opinions that are spread within the community that do not go directly to the topic but instead use things related to the topic to express opinions. In sentiment analysis, indirect opinion such as this must be entered into the data to be managed for better classification. The case studies taken in this study are 3 major cities in Indonesia, namely Bandung, Medan and Jakarta, because recently major cities in

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

Indonesia have become a hot topic of discussion in the community as feedback on the city. Public opinion of the city, can be used as a review for the government to better control the course of government or other communities to find out the condition of other cities.

This study aims to analyze the results of the implementation and analysis of Indonesian-language sentiments on Twitter for cities using a target-dependent approach for grouping data and SVM methods for sentiment classification and also to find out the level of positive and negative opinions on the cities of Bandung, Medan and Jakarta.

2. Literature Review

2.1. Twitter

Twitter is a microblogging service that has received a lot of attention nowadays. Unlike other social media like Facebook or MySpace where the relationship between people who follow and people who follow cannot respond to each other, Twitter allows users to follow or be followed by others. One user can follow another user and the user that follows does not need to re-follow the user who follows him. Being a follower on Twitter means that the user receives all messages called tweets from the user he follows. The terms used in Twitter include RT to retweet, '@' to identify a username and '#' represents a hashtag or topic. Twitter itself has a writing limit, where in each tweet can only consist of 140 characters. The retweet mechanism supports Twitter users to spread information of their choice. [2]

2.2. Sentiment Analysis

Sentiment analysis, also known as opinion mining is a field of science that analyzes opinions, sentiments, evaluations, judgments, attitudes and emotions of people towards several objects such as products, services, organizations, individuals, problems, events, topics and others [4]. This is a very big problem area. There are also many other names from this analysis, such as sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, mining review, and others [12][13]. However, all of them are now in a large container called sentiment analysis or opinion mining. In the industrial world, sentiment analysis is more often used but in the academic world sentiment analysis and opinion mining are often implemented. Both basically represent the same field of science.

Opinion is the center of all human activity because they are the key that causes our attitude and treatment in society. Whenever we want to make a decision, we want to know other people's opinions on the topics we discuss. In real life, businesses and organizations always want to know public and consumer opinions about their products and services. Consumers also want to know the opinions of previous users of a product before buying it and other people's opinions about political candidates before making decisions in political elections. In the past, when someone needed an opinion, he had to ask a friend or family [6]. When an organization or business body needs public and consumer opinion, they must make a survey, gather opinions and discuss groups. To find out public opinion and consumers, more effort is needed to support marketing, public relations and public promotion.

With the growth of social media such as discussion forums, blogs, micro-blogs, Twitter, comments and posts on social networking sites on the internet, organizations and individuals are already using this feature as a support for decision makers [11]. At this time, if someone wants to buy a product, there is no need to ask the opinion of friends or family because there are many user reviews and discussions relating to the product on the internet.

The most important indicator in opinion is words called opinion words. These are examples of words that are usually used to indicate positive or negative opinions. For example, "bagus", "mengagumkan" dan "mengesankan" are words of positive opinion while "buruk", "payah" dan "mengerikan" are words of opinion that indicate negative opinions. Besides that, there are some phrases or terms that might mean positive or negative opinions for example "film ini bagus sekali sampai aku tidak mau melihatnya lagi". Words and phrases in opinion are very meaningful in sentiment analysis. [2] [4] [5]

2.3. Target-Dependent

It is not uncommon for people to express their opinions about a particular target or topic by commenting not only on the target but on matters relating to the target or topic. For example, someone

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

might express his opinion about a company through comments on products or technology produced by the company. To show opinions about a product, someone might comment about the features or functionality of the product. It is intended that the reader can clearly link opinions about the main topic through comments or opinions on matters relating to the topic. As shown below, the author shows positive opinions about "Microsoft" by making positive opinions directly about "teknologi Microsoft".

"Saya merasa bahagia dengan teknologi Microsoft, khususnya Silverlight"

Comments that show positive or negative opinion by using things related to this topic can be categorized as positive or negative opinions about the main topic. Therefore, for the classification of Twitter sentiments using a target-dependent approach, the first thing to do is to identify all things that might be related to the main topic. [3]

This approach states that opinions containing nouns related to targets are also the result of analysis of user sentiments towards targets even if the target or topic is not mentioned in the opinion.

1. The first thing to do in this approach is to identify nouns related to the target or topic discussed in the study. It is normal for users to express their opinions on targets through nouns related to targets with or without writing down target keywords in their opinions. For example "Bandung. Aku cinta kamu."The word "kamu" in that sentence refers to "Bandung". With this we add opinions that contain the word target as the target dataset.
2. Determine nouns or noun phrases that have a strong connection to the target. Here we use Pointwise Mutual Information (PMI) to measure the connectedness.

$$PMI(w,t) = \log \frac{p(w,t)}{p(w)\,p(t)} \quad (1)$$

Where p (w, t), p (w) and p (t) are the possibilities of the occurrence of w and t interconnected. w appears and t then arises because of w.

3. The last thing to do in this approach is to identify the number of occurrences of the related nouns in the opinions given by the user. Determine what nouns are related to what. For example, at this stage we can see that "kuliner" has a connection to " kota kuliner Indonesia ya Bandung". "Durian" relates tothe "Medan" target of " kangen Medan dan duriannya " and "macet" relate to "Jakarta" from " kenapa Jakarta selalu macet?".

2.4. Support Vector Machine (SVM)

The classification technique that has received attention now is the Support Vector Machine (SVM). This technique has a basis on statistical learning theory and has shown promising results in many research applications, from recognition of handwritten numbers to text categorization. SVM also works well with high-dimensional data and successfully avoids the problem of dimensional text data. Another unique thing from this approach is that SVM is able to represent decision boundaries using a subset of training data, which has come to be known as a support vector. [8]

One method that has recently received much attention as a state of the art in pattern recognition is Support Vector Machine (SVM) [1] [2] [7]. The Support Vector Machine (SVM) was developed by Boser, Guyon, Vapnik, and was first presented in 1992 at the Annual Workshop on Computational Learning Theory. The basic concept of SVM is actually a harmonious combination of computational theories that have existed decades before, such as hyperplane margins [9], kernels were introduced by Aronszajn in 1950, and likewise with other supporting concepts. However, until 1992, there had never been any efforts to assemble these components [3] [8]. Unlike the neural network strategy that tries to find hyperplane separators between classes, SVM tries to find the best hyperplane in the input space. The basic principle of SVM is a linear classifier, and subsequently developed to work on non-linear problems [10]. by incorporating the concept of kernel tricks in high-dimensional workspaces. This development has stimulated research interest in the field of pattern recognition for theoretical and potential application of SVM investigations. Today SVM has been successfully applied to real-world problems, and in general provides better solutions than conventional methods such as artificial neural networks. This paper introduces the basic concept of SVM, and discusses its application in bioinformatics, which is currently one of the fields that is developing quite rapidly.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

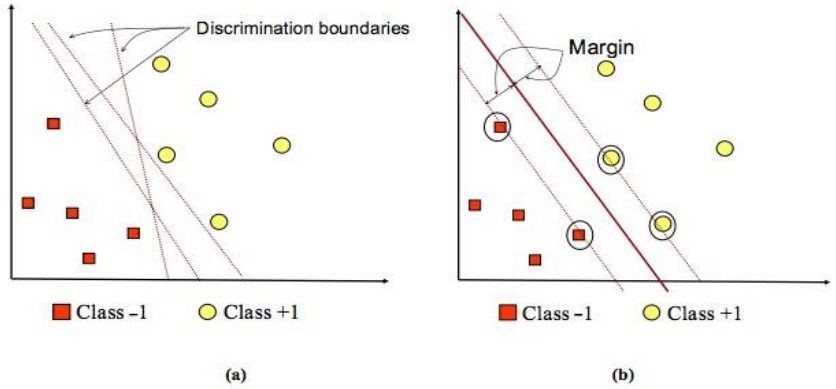

Figure. 1. Classifying in SVM

The concept of SVM can be explained simply as an attempt to find the best hyperplane2 that functions as a separator of two classes in the input space. Figure 1a shows several patterns that are members of two classes: +1 and –1. Patterns that are joined in class-1 are symbolized in red (box), while patterns in class +1, are symbolized in yellow (circle). The classification problem can be translated by trying to find a line (hyperplane) that separates the two groups. Various alternative discrimination boundaries are shown in Figure 1-a. The best hyperplane separator between the two classes can be found by measuring the margin of the hyperplane. and look for the maximum point. Margin is the distance between the hyperplane and the closest pattern of each class. The closest pattern is called a support vector. The solid line in figure 1-b shows the best hyperplane, which is located right in the middle of the two classes, while the red and yellow dots in the black circle are support vectors. The effort to find the location of a hyperplane is the core of the learning process in SVM.

3. Analysis and Discussion

In this final project, the system built is a system that ultimately produces target-dependent information on the cities of Bandung, Jakarta and Medan which can be used for the process of classifying sentiments and grouping data.

After the system has been built, a test of the data will be carried out to measure the correctness of the classification of sentiments and the grouping of data on the tested data.

The system test was conducted on public opinion data on the cities of Bandung, Jakarta and Medan on social media Twitter with the following scenarios:

1. Observation of the results of the implementation and analysis of Indonesian sentiments on Twitter to the city using the target-dependent approach for data classification and SVM method for sentiment classification.
2. Observe the effect of the target-dependent approach in classifying data on public opinion of the cities of Bandung, Medan and Jakarta produced by the system.
3. Analyzing the influence of the frequency of the bigram consisting of the cities of Bandung, Medan and Jakarta and words that are often described by the public against these three cities with the frequency of each word on the value of Pointwise Mutual Information (PMI) generated by the system.

The author tests the training data and test data on this system with scenarios using 300 twit training data for the training data and uses 150 tweets for the test data.

3.1. Analysis of Implementation Results and Sentiment Analysis

Analysis of the results of the implementation was carried out by observing the results of the calculation of Pointwise Mutual Information (PMI) for each group of data which was then used for the process of finding more data to compare the results of positive, negative and neutral opinions on the cities of Bandung, Jakarta and Medan.

The following table shows the results of Pointwise Mutual Information (PMI) generated by opinion data on the cities of Bandung, Medan and Jakarta.

Bigram frequency is the frequency of the two words said in the data, while frequency 1 shows the

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

number of repetitions of the first word in the data as well as frequency 2 shows the number of repetitions of the second word in the data.

3.2. Analysis of Target-Dependent and Non Target-Dependent Data Sentiment Classification Results

After the PMI value is generated, data collection and classification is based on the PMI value owned by the cities of Bandung, Jakarta and Medan. There are two types of data to be classified as sentiment values, namely data using the target-independent approach and data using the target-dependent approach. Where, target-independent data is data collected using Bandung, Jakarta or Medan keywords only and target-dependent data is data collected using keywords from PMI values generated by target-independent data.

1. Bandung Sentiment Classification Results

The following is table 1 of the results of the sentiment classification of the city of Bandung produced using the SVM method. Comparison of positive and negative opinions from the data using the target-independent approach and data using the target-dependent approach is carried out.

Table 1
Sentiment Classification Result Of Bandung

	Target-independent	*Target-dependent*
Positif	37	57
Netral	27	24
Negatif	36	19
Tingkat Precision	80,4%	79,5%
Tingkat Recall	75%	785

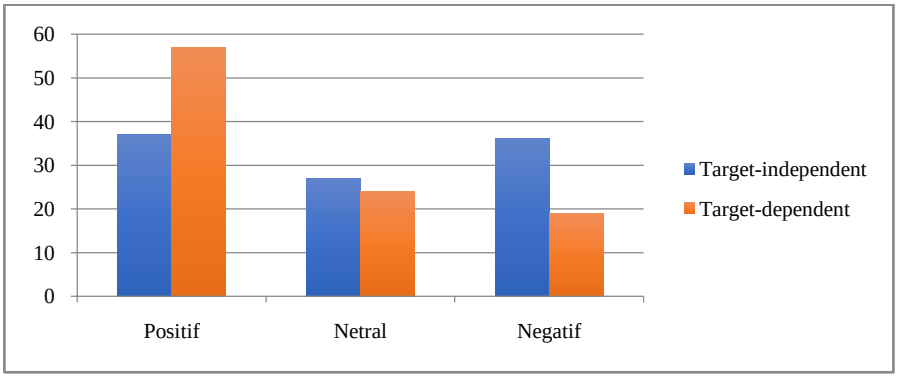

Figure. 2.Graph of Comparison of Positive, Negative and Neutral Sentiments between Target-Dependent and Target-Independent Data towards Bandung City

2. Medan Sentiment Classification Results

The following is table 2 of the results of the sentiment classification of Medan produced using the SVM method. Comparison of positive and negative opinions from the data using the target-independent approach and data using the target-dependent approach is carried out.

Table 2
Sentiment Classification Result of Medan

	Target-independent	*Target-dependent*
Positif	31	32
Netral	17	54
Negatif	52	14
Tingkat Precision	83,2%	77,6%
Tingkat Recall	75%	71%

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

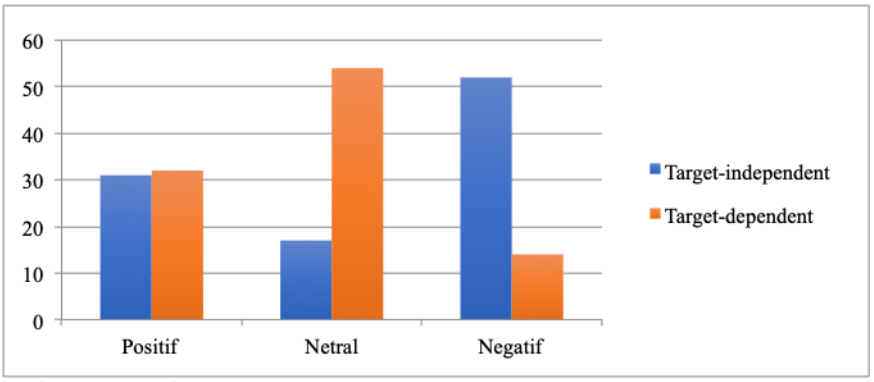

Figure. 3.Graph of Comparison of Positive, Negative and Neutral Sentiments between Target-Dependent and Target-Independent Data towards Medan City

3. Jakarta Sentiment Classification Results

Following is table 3 the results of the sentiment classification of the city of Jakarta produced using the SVM method. Comparison of positive and negative opinions from the data using the target-independent approach and data using the target-dependent approach is carried out.

Table 3
Sentiment Classification Result of Medan

	Target-independent	*Target-dependent*
Positif	36	20
Netral	37	46
Negatif	27	34
Tingkat Precision	82,1%	88,9%
Tingkat Recall	81%	88%

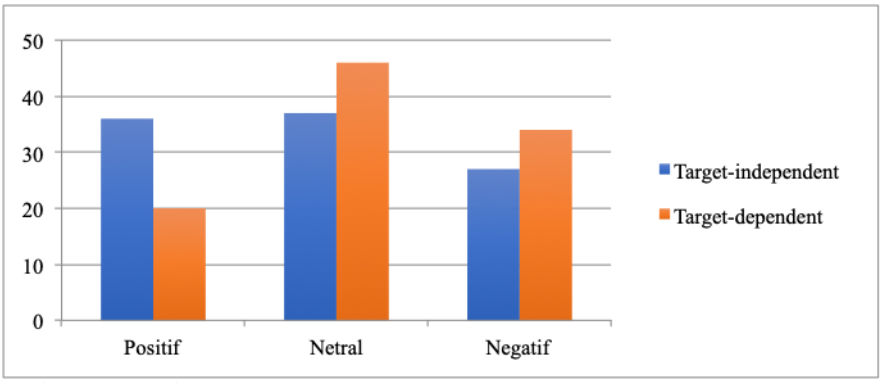

Figure. 4.Graph of Comparison of Positive, Negative and Neutral Sentiments between Target-Dependent and Target-Independent Data towards Jakarta City

4. Conclusion

Based on the analysis and testing carried out in the previous chapter, the conclusions that can be drawn are as follows:

1. The target-dependent approach to the classification of data was successfully implemented in the case of sentiment analysis of the cities of Bandung, Jakarta and Medan, after which sentiment classification was conducted using the SVM method.
2. The target-dependent approach affects the number of positive, negative and neutral opinions on the cities of Bandung, Jakarta and Medan. Where the classification of data is done by taking things that are closely related to the topics Bandung, Jakarta and Medan in accordance with the value of Pointwise Mutual Information generated from the system.
3. The value of the frequency of bigrams consisting of the cities of Bandung, Medan and Jakarta and words that are often described by the community of the three cities with the frequency of each word

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik
Volume 4 Number 1 May 2020, pp. 20-26 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

affect the value of Pointwise Mutual Information (PMI) which is then used to classify data according to the target approach -dependent.

5. References

[1] Yao, Yukai, et al. "K-SVM: An Effective SVM Algorithm Based on K-means Clustering." *JCP* 8.10: 2632-2639.2013.
[2] Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?." *Proceedings of the 19th international conference on World Wide Web*. 2010.
[3] Jiang, Long, et al. "Target-dependent twitter sentiment classification." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics. 2011.
[4] Putri, TTA., Sriadhi, S., Sari, RD., Rahmadani, R., Hutahaean, HD. "A Comparison of Classification Algorithms for Hate Speech Detection" *In Proceeding of ICIEVE 2019* (to be published).
[5] Agusta, Ledy. "Perbandingan algoritma stemming Porter dengan algoritma Nazief & Adriani untuk stemming dokumen teks bahasa indonesia." *Konferensi Nasional Sistem dan Informatika* 2009: 196-201. 2009.
[6] Wicaksono, A., Ellen Nio, and S. H. Myaeng. "Unsupervised Approach for Sentiment Analysis on Indonesian Movie Review." *Proceedings of the 6th Conference of Indonesian Students Association in Korea (CISAK 2013)*. Vol. 10. 2013.
[7] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1: 1-167. 2012.
[8] Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India. 2016.
[9] Vapnik, Vladimir. *The nature of statistical learning theory*. Springer science & business media, 2013.
[10] Cristianini, Nello, and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press. 2000.
[11] Weiss, Sholom M., et al. *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media. 2010.
[12] Triawati, Chandra. "Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia." *Institut Teknologi Telkom, Bandung*, 2009.
[13] Putri, Tansa Trisna Astono. "Analisis dan deteksi hate speech pada sosial twitter berbahasa indonesia." *Master's thesis, Faculty of Computer Science, Universitas Indonesia*, 2018.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.