TELKOMNIKA, Vol.15, No.3, September 2017, pp. 1301~1309
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v15i3.4711

The Addition Symptoms Parameter on Sentiment Analysis to Measure Public Health Concerns

Yohanssen Pratama*[1], Puspoko Ponco Ratno[2]
[1]Faculty of Informatics and Electrical Engineering, Institut Teknologi Del,
Jl. Sisingamangaraja, Sitoluama, Toba Samosir, Sumatera Utara, Indonesia, Tel: +62 632 331234
[2]School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia
*Corresponding author, e-mail: yohanssen.pratama@del.ac.id

Abstract

Information about public health has a very important role not only for health practitioners, but also for goverment. The importance of health information can also affect the emotional changes that occur in the community, especially if there is news about the spread of infectious disease (epidemic) in particular area at the time, such as case of outbreaks Ebola disease or Mers in specific area. Based on data obtained from Semiocast, Indonesia is the country with fifth largest number of Twitter users in the world, where every topic that lively discussed will also influence a global trending topic. This paper will discuss the measurement of public health concern (Degree of Concern) level by using sentiment analysis classification on the twitter status. Sentiment data of the tweets were analyzed and given some value by using a scoring method. The scoring method equation (Kumar A. et al., 2012) will be tested with new additional parameters, ie symptoms parameters. The value of any twitter user sentiment is determined based on adjectives, verbs, and adverbs that contained in the sentence. The method that we used to find the semantic value of adjectives is corpus-based method. While for finding the semantic value of the verb and adverb we used a dictionary-based method.

Keywords: *Twitter, epidemic, sentiment classification, scoring method*

1. Introduction

Information has a very important role in enhancing our knowledge and perspective about the environment and the outside world. As well as information about health, is needed by the community, so the individual's health condition can be maintained. Poor access to health facilites in Indonesia make people seemed not to care about health information. On the other hand, most of the media in Indonesia rarely present information about health, whereas the number of the disease and the death rate could be reduced if people get adequate health information. Lack of health information will make people more susceptible to the dangers of the disease.

The importance of health information can also affect the emotional changes that occur in society, especially if there is news about the spread of infectious diseases (epidemic) in a particular area at a time. Based on the research of Zhu, et al., [1] stated that at the moment there is information about outbreaks of infectious diseases, the majority of people that being interviewed (96.4%) showed changes in negative emotions such as panic (54.8%), anxiety (34.0%), and fear (7.6%). Real example of the emotional changes that caused by the disease actually happened when there is ebola outbreak in particular country, the economic level become decreasing because affected by the disease. From these examples it can be concluded that information on health have a crucial role in the life of society, and so we need a reliable system to obtain accurate and appropiate information about health.

To obtain accurate information we need a robust monitoring system. Monitoring information on public health has a very important role not only for health practitioners such as doctors, health specialist, nurses, and other health practitioners, but also for the goverment, particularly the health department. To do this surveillance, the supervisor collecting data systematically and continously, followed by analysis and interpretation of data which related to health event. The data results can be used to gain a better understanding of the health status of a population in order to plan, implement, describe, and evaluate public health programs to

Received April 3, 2017; Revised June 28, 2017; Accepted July 13, 2017

control and prevent the spread of disease. Therefore, to be useful, data must be accurate, timely, and available in a form that is easy to understand.

Some systems for monitoring health have been proposed in an innovative way to obtain data on health. The data could be use as an indicator of the spread disease activity, start by counting the number of searches for particular drugs, doctor, and the hospital. For example Google Flu Trends which understand that human interaction in virtual world are a valuable source for sensing health trends. Google Flu Trends utilizes aggregated web search queries pertaining to influenza to build a comprehensive model that can estimate nationwide as well as state-level influenza-like illness (ILI) cases [1]. In this paper we will try to present a similar system but by utilizing the data from Twitter.

We utilizing data from Twitter because based on data which obtained from Semiocast (www.semiocast.com), Indonesia is the fifth country in the world with the largest number of Twitter users, where it affects the global trending topic. This fact proves that Indonesia has abundant information and potential information sources that could be used in decision making [2]. We could use twitter services to collect potential information about health and then make some decisions from it. In Indonesia, netizens (internet users) tend to use social media such as Twitter to criticize, express opinions, and share about the conditions or events that occur in the vicinity. This is not the first time research that conducted by using data from Twitter. There are some previous research show that reports from twitter can be used to predict 2009a (H1N1) swine flu pandemic. More recently, Collier, et al., [3] analyze the status (tweets) on Twitter and show that there is a strong enough correlation between the results of studies with laboratory data report from WHO / NREVSS for A (H1N1) in United States in 2009-2010. Lampos and Cristianini [4] and Culotta [5] conduct the research by using data from Twitter to track the spread of influenza disease.

Some research has also been conducted to obtain information about public health based on public sentiment for illnesses. Xiang Ji, et al., [6] through they research tried to determine the magnitude of public concern about the spread of disease that occur in the vicinity, by classifying public sentiment into positive and negative class, using the Multinomial Naïve Bayes method. In they research, Xiang Ji, et al. [6], separate between the tweet that contains news about a disease with a tweet that contains the user tweets that affected by disease (personal tweet). Then followed by applying sentiment analysis on the results of personal tweet to separate between the tweet that has a value of negative sentiment and positive sentiment. To do this sentiment classification they conduct experiments using two methods, there are Clue Based which combined with F-Measure and machine learning method that uses Naïve Bayes and SVM. The results of the classification that has the best accuracy will be used as the parameter value in the calculation degree of concern. From these research it was concluded that the Multinomial Naïve Bayes method has the best results [6].

In connection with all previous research, our research will design and implement a method for calculating the level of concern Twitter users (degree of concern) based on the Twitter sentiment score status that expressed by the user at any given time. The idea for the method that we use is by adding symptoms parameter into scoring method which proposed by Kumar, et al., [6]. The first step that conducted in this method is filtering the data which taken from the Twitter by the disease name that we want to search. Then we do the pre-processing to eliminate non-essential attribute. Results of pre-processing will experienced the sentiment classification process by using the scoring method [7]. The scoring method classify the data based on the words contained in the tweet status, where the sentiment classification results will be analyzed to measure the levels of concern (Degree of Concern) people in the area [6]. Finally information degree of concern for some diseases in particular area that acquired from twitter could be used as warning indicators. This indicator can be used by health authorities to control the health of residents.

2. Research Method

In this section will be explained about the sentiment analysis research design that used in the measurement degree of concern a community to the health. There are 3 stages (input, process, and output) that must be followed until we get the degree of concern as an output. The twitter text become an input for the process stage and in the process stage there are preprocessing step and comparison between scoring method with and without addition symptoms parameter. Figure 1 below presented the research step and design of this paper:

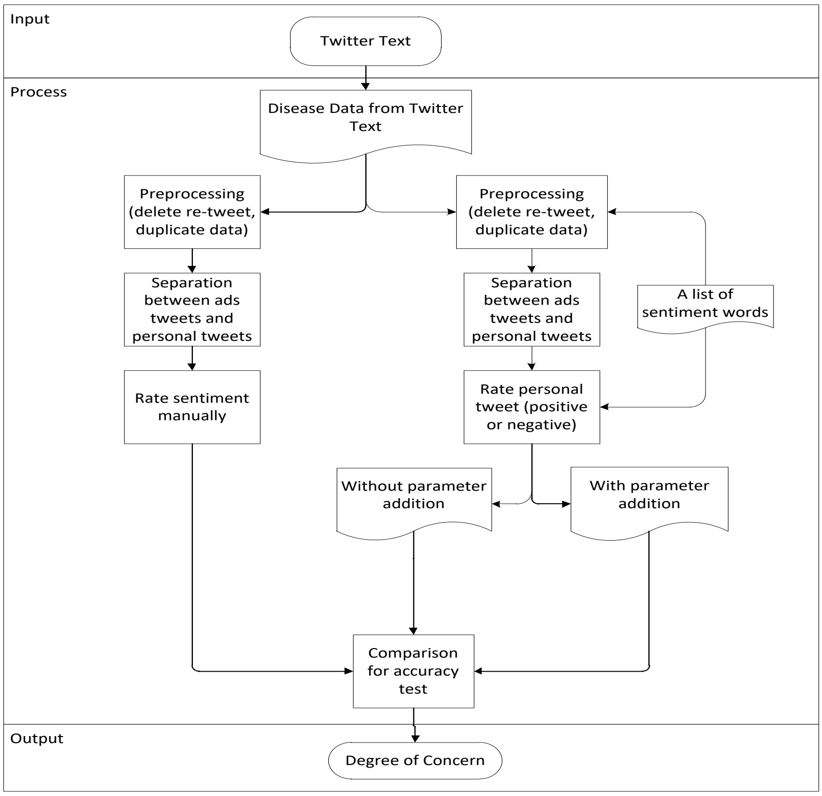

Figure 1. Effects of selecting different switching under dynamic condition

a) Input

The data input or commonly reffered as dataset are comes from tweet timeline of twitter users who speak Indonesian. In this case the data that collected from twitter is influenza disease data.

b) Process

The data that has been collected in the database will undergo a preprocessing process. In this process performed removal of retweet data, deletion of duplicate data, also separation between personal tweets with a tweets that contains advertising. After that the sentiment rating is conducted by sampling some of the data. The data sample will be given the value in 3 ways: manually, by sentiment rating based on Akhsi. et al. method, and by addtion symptom parameter into Akhsi. et al. method. Results of the assessment from the last two methods will be compared with manual methods to get the best accuracy values and to see wether the influence of the symptom parameter give some impact into the level of accuracy of the public sentiment.

c) Output

The output of the application is the value of public concern (degree of concern) which obtained from the best comparison between akhsi K., et al methods that use parameters to those not using parameters.

In every step (Figure 1) there were many process and parts that involved, below is the explanation for:

2.1. Dataset

The dataset used in this study were taken from twitter status. Data retrieval is based on the name of disease that wants to be analyzed. We choose the name that has similarities with influenza such as cough, colds, and flu. In addition, data for diarrheal diseases was also taken which will be used as a test of the method that generated in this study. Determination of data collection for diseases that have similarities with influenza illness is based on the amount of data on influenza disease in Twitter and the ammount of supporting data such as influenza disease sypmtoms data, complaints, etc.

The taken data came from the someone timeline who has such words: ‘cough’, ‘cold’, ‘flu’, ‘influenza’ for the dataset illness that similar to influenza. On the other hand the word ‘diarrhea’ was taken for diarrheal disease dataset. All taken data was searched by using Twitter API. Results of the search will provide feedback in the form of a json data structure. The existence of this json data structure provide the flexibility for researchers to change the data format accordance the need. Data was taken from 1 August 2014 until 31 October 2014 with 213.900 total amount of data for influenza illness and 12.450 data for diarrheal diseases.

2.2. Pre-processing

The dataset will be going through the preprocessing step and this processing is intended to perform cutting at every tweet sentence and choose the words that correspond to the raw words in Indonesian. The words that contained in the tweet will be formalized accordance to Indonesian language rules which commonly referred to ‘ejaan yang disempurnakan’ (EYD). This is due to the number of words that are not follow the standard rules (EYD) being used in the tweet, as well as the use of abbreviations, punctuation, and emoticons. To formalize the terms that not follow EYD we do formalization for some terms like “stratejik”→”strategi” (strategy), “spekulas”→”spekulasi” (speculation), “melamah”→”melemah” (declining) [8]. The preprocessing stage could be se in Figure 2.

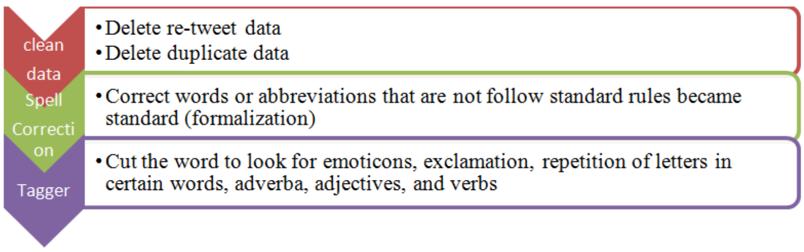

Figure 2. Pre-processing Stages on Research

2.3. Sentiment Analysis

The results of the data preprocessing will produce parts of words that will be used to determine the sentiment value of every tweets from the twitter user. Adjectives and adverbs value from the sentence will be combined and calculated together where the value is:

$$AG = nAdj \times nAdv \tag{1}$$

AG = Adjectives Group (The combined value between adjectives and adverbs)
nAdj = Adjectives Value
nAdv = Adverbs Value

As for the adverbs value which combined with verbs are calculated by the equation:

$$VG = nVerb \times nAdv \tag{2}$$

VG = Verbs Group (The combined value between adverbs and verbs)
NVerb = Verbs Value
NAdv = Adverbs Value

Should the results form the data preprocessing has no adverb, then the adverb would have a 0.5 value. The total of sentiment value from each tweet can be calculated using following Equation [7]:

$$S(T)=\frac{\left(1+\left(Pc+\log\left(Ns\right)+\log\left(Nx\right)\right)/3\right)}{\left|OI\left(R\right)\right|}*\sum_{i=1}^{|OI(R)|}S(AGi)+S(VGi)+Nei*S(Ei) \quad (3)$$

S(T) = Sentiment Value from the tweet
Pc = the number of capital letters in the tweet word
Ns = the number of recurring characters
Nx = the number of exclamation points that contained in the tweet
S(AGi) = Adjectives group value
S(VGi) = Versbs group value
S(Ei) = the emoticons value that contained in the tweet
Nei = the number of emoticons in the tweet

The result value from these calculations will experience rounding, wherein if the sentiment value is smaller than -1 then the value will be rounded to -1 and if the value is greater than 1, the value is rounded to 1.

The value for verbs and adverbs are given based on Kumar A. and T.M. Sebastian paper [7], where the words is translated and adapted by looking for its synonyms in Indonesian. Value for adverbs and verbs are shown in Table 1.

Table 1. List of Verbs and Adverbs in Indonesian

Verbs (In Indonesian)	English	Value	Adverbs (In Indonesian)	English	Value
Cinta	Love	1	sempurna, tuntas, sembuh, sehat	Perfect, done, cure, healthy	1
Kagum	Admire	0,9	paling, amat	Most	0,9
Suka	Like	0,8	Sangat		0,8
Nikmat	Favor	0,7	Selalu	Always	0,7
Senyum	Smile	0,6	Sekali	Once	0,6
terkesan	Impress	0,5	Banyak	Many	0,4
Tertarik	Attract	0,4	Cukup	Enough	0,3
Senang	Happy	0,3	Lebih	More	0,2
Santai	Enjoy	0,2	Beberapa	Some	-0,2
bosan, tolak	Bored, ignore	-0,2	Agak	Rather	-0,3
Jijik	Disgust	-0,3	Sedikit	A little	-0,4
Derita	Suffer	-0,4	Kurang	Less	-0,6
tidak suka	Don’t like	-0,6	Jarang	Seldom	-0,8
sebal	resentful	-0,9	Tidak	No	-0,9
benci	Hate	-1	tidak pernah	Never	-1

Emoticon values also drawn from Kumar A. paper [7] by making additions and adjustments to the emoticons that have the same meaning. Emoticons values are shown in Table 2.

Table 2. List of Emoticons and its meaning in Indonesian

Emoticon	Meaning (In Indonesian)	English	Value
:D	Tertawa lebar	Laughing out loud	1
BD	Tertawa lebar dengan kaca mata	Laughing with eyeglass	1
XD,:)),;),=D>	Tertawa	Laughing	1
\m/	Hai 5	Hi 5	1
:),=),:-), :3	Tersenyum	Smile	0,5
:*	Ciuman	Kiss	0,5
:|	Muka datar	Poker face	0
:\	Belum memutuskan	Uncertain	0
:(	Sedih	Sad	-0,5
</3	Patah Hati	Broken hearted	-0,5
B(	Sedih dengan kaca mata	Sad with eyeglass	-0,5
:’(	Menangis	Cry	-1
X-(	Marah	Angry	-1

The adjectives words were taken from www.wiktionary.com (Wiktionary, 2014). Each word has a different value, where the values are given manually between -0.5, 0, and 0.5. For a 0.5 value was given to a group of words adjectives that have a positive meaning. A value of 0 is given to words that are not commonly used or give a neutral meaning to the sentence. Adjectives with value -0.5 were given to a group of words that have a negative meaning. Here are some example of adjectives words and their values (Figure 3),

Table 3. List of Adjectives in Indonesian

Adjectives (In Indonesian)	Value
Bahagia (happy), ceria (cheerful), gemar (love to), gembira (happy), istimewa (special), jempolan (top), nikmat (enjoyable), sembuh (preserve), sehat (well), senang (joyful), suka (like), senyum (smile), etc.	0,5
Babil, awawarna, awamineral, azal, baki, bangkas, batangan, etc.	0
Agresif (aggressive), akut (acute), alergis (allergic), ambekan, aneh (weird), amburadul (broke), bajingan (bastard), ceroboh (careless), galak (impudent), ganas (fierce), ironis (ironic), jahat (evil), jahanam (cursed), jelek (ugly), etc.	-0,5

2.4. Degree of Concerns

The results of the sentiment value will be added together to be a reason in determining the value of the degree of concern for particular disease in public. The value of the degree of concern can be calculated using the following equation [5],

$$DOC\left[d,t\right]=\frac{NN^2}{PN} \tag{4}$$

DOC [d,t] = degree of concern value
NN = number of tweets that have a negative sentiment
PN = number of tweets that have a positive sentiment

By knowing the value of the degree of concern, we can detect areas that have the highest concentration of the disease [6].

3. Results and Analysis

Analysis conducted in 3 stages of test scenarios. Scenario that being used are:

1. Application testing

This phase was conducted to test whether the application that has been made in accordance with prior planning or not.

2. Sentiment classification testing

This test was conducted to test the accuracy of the classification program on sentiment. This sentiment classification testing will be made to the influenza dataset disease.

3. Testing of additional parameters on sentiment classification method

This test was conducted to examine the increase in accuracy for the classification program on sentiment. This sentiment classification testing will be conducted on influenza dataset disease. Here is the scenario that the test will be conducted along with the test results,

Table 4 below show the scenario test that will be conducted and the test results are shown in Table 5.

Table 4. List of Test Scenarios

No	Description
1	Programs can be collecting data (*crawling*) via Twitter API
2	Program for data collection can collect data in accordance with the search terms and eliminate the tweet that in the form of re-tweet
3	The program can perform the elimination of duplicate data (pre-processing)
4	The program can detect the presence of emoticons, words that written in capital letters, repetition of letters in a word, the number of exclamation points
5	The program can perform counting on adverbs group and verbs group
6	The program can perform sentiment classification
7	The program can perform accuracy testing for the sentiment classification that has been generated

Table 5. List of Test Scenarios Results

Test Case	Description	Testing procedures	Input	Expected output	Results	Conclusion
1 &2	*Crawling data*	Using the Twitter Search API to gather data from Twitter	Query to display the expected data	The obtained data should be relevant to the query	The collected data can be displayed in a browser	Accepted
3	Delete duplicate data	Choose duplicate data	Data from the database is result from the tweet data crawling	Dataset without any duplicate data	Datasets without duplicates	Accepted
4&5	Parameter detection and assessment	The tweet data were split for each sentence, this was intended to double check the parameters that have been determined	The tweet disease data that already clean, data parameters that have been determined	Detection results with its value (score)	Value / score of each parameter	Accepted
6&7	Sentiment classification process including its accuracy	Enter the tweet disease data then choose the equation that will be tested	Tweet disease data from the database	Sentiment classification against the disease data and followed by assessing the degree of accuracy	Classification results appear in the browser	Accepted

3.1. Validity of Test Results

The results of validity testing is consisting of test results from the test phase methods. The test method is being implemented by adding the sentiment parameters and then test its accuracy. The testing is using one kind of dataset, i.e. dataset for influenza disease. Here are the results of accuraccy testing that showed in Figure 3,

The Result of Comparison

Twitter Content (in Indonesian)	Sentimen Manual	Sentimen Kumar	Sentimen +Parameter
Suara emasku nanti ganti bruwet ya mah ? :D "@UllyMartika: @herlina_laras haaa koe ojo batuk ta nak koe kudu sehat {}"	NETRAL	NETRAL	NETRAL
Aaa gegara si batuk semalem nih, jadi kesiangankan:(	Negatif (-)	Negatif (-)	Negatif (-)
Semangat pagi walau pilek melandaðŸ˜ƒðŸ˜†ðŸ˜ â™« Shake It Off by @taylorswift13 â€" https://t.co/MEttbLXVk5	Negatif (-)	NETRAL	Negatif (-)
Batuk terus sampe mampuuuuuus:&	Negatif (-)	Negatif (-)	Negatif (-)
Wahai batuk ku sayang... bilakah anda mau menjauhkan diri dari aq?!! Erghh	Negatif (-)	NETRAL	Negatif (-)
pilek berkepanjangan!! NO :(	Negatif (-)	Negatif (-)	Negatif (-)
Kesampaian juga dinyanyiin secara langsung, meskipun yg nyanyi pilek beratâ€¦ â™« Marry Your Daughter by BRKNRBTZ â€" https://t.co/LIwypfVXeK	Negatif (-)	NETRAL	Negatif (-)
@herlina_laras haaa koe ojo batuk ta nak koe kudu sehat {}	NETRAL	NETRAL	NETRAL
Pilek total "@CittaFitra: Menyiksa itu ketika makan dan bernafas lewat mulut :3"	NETRAL	NETRAL	NETRAL

Figure 3. Application display for accuracy testing

1. The test results that used a scoring method.

Sentiment data of the tweets were analyzed and given some value by using a scoring method [7]. The results of calculation by using the scoring method would be compared to manual counting. The test data that used was taken from the twitter sample dataset which associated with influenza illness. By using a scoring method [7], we get the results as can be seen in the Table 6.

Table 6. List of Test Scenarios Results

Methods	Neutral	Negative	Negative Right Value	Positive Right Value
Manual	548	452	452	548
Kumar A. et al.	701	299	255	504

From Table 6 can be calculated several values to get the F-Measure, which is used as a reference value for measuring the ability of the algorithm to classify correctly. Some of these values are as follows:

Accuracy Test = 56.4%
Precision = 75.9%
Recall = 56.41%
F-Measure = $2.\frac{precision.\,recall}{precision+recall}$ = 64.72%

This test result show that the F-measure value that formulated by Kumar A. et al., is equal to 0.6472, or it can be said that the algorithm has a validity value of 64.72% within the disease sentiment classification.

2. The test results that used a scoring method with the addition of symptoms parameters.

Data sentiment of the tweets were analyzed and given a value by using the scoring method with the addition of symptoms parameters for particular disease. The results would be compared to manual count results. The 1000 data that used was taken from the twitter sample dataset which associated with influenza disease. By using a scoring method with the addition of symptoms parameters, we get the results as can be seen in the Table 7.

Table 7. List of Test Scenarios Results

Methods	Neutral	Negative	Negative Right Value	Positive Right Value
Manual	548	452	452	548
Kumar A. et al.	565	435	401	514

From the Table 7, we get the F-Measure and another values are as follows:

Accuracy Test = 88.72%
Precision = 91.5%
Recall = 88.71%
F-Measure = $2.\frac{precision.\,recall}{precision+recall}$ = 90.09%

This test result show that the F-measure value that formulated by Kumar A, et al., with the addition of symptoms parameters is equal to 0.9009, or it can be said that the algorithm has a validity value of 90.09% within the disease sentiment classification. Table 8 show the accuracy comparison between the scoring method that with and without a symptoms parameter addition.

Table 8. Accuracy Comparison

Methods	F-Measure
Kumar A. et al. without symptoms parameter	64,72%
Kumar A. et al. with symptoms parameter	90,09%

3.2. Degree of Concern

Based on the results above, it is known that the largest F-Measure value that can be used as reference for the validity of the algorithms is equal to 0.9008 or 90.08%. This value was obtained after the addition of the symptoms parameters. This result will be used for searching the value of degree of concer n community against an influenza disease. Degree of concern value based on Equation (4) are:

$$DOC = \frac{NN^2}{PN} = 401/1000$$
$$= 0.401$$

The DOC value that obtained from the formula is 0.401, which means that there are approximately 40.01% of the population in the area that infected by influenza disease.

4. Conclusion

In this study, we have been carried out analysis of public sentiment against influenza disease. The calculation of sentiment value refers to verbs, adjectives, and adverbs value that contained in the sentence. For the calculation of sentiment, we take into account the existing emoticon in a sentence, repeated word in sentence, adjectives and adverbs group values. So the sentiment values that obtained has a semantic meaning in a tweet sentence. The addition of symptoms parameter values to the equation that made by Kumar A. et al., further increasing the accuracy on sentiment classification. There is difference 25.37% between the scoring method that with and without addition of symptoms parameter. And then the sentiment classification results can be used as a parameter to search the degree of concern value community to a disease. For the future work, we need weighting techniques in verbs to get more accurate sentiment value, because in this case Indonesia language has a lot of verbs.

References

[1] Xichuan Z, Qin L, Zhenglin Z, Han Z, Hao T, Yujie F. Monitoring Epidemic Alert Levels by Analyzing Internet Search Volume. *IEEE Transaction on Biomedical Engineering*. 2012; 60(2): 446-452.
[2] Titin P, Iping, S, Ayu P. Determining Trust Scope Attributes Using Goodness of Fit Test: A Survey. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2015; 13(2): 654-660.
[3] Collier N, Nguyen S, Nguyen M. *Omg u got flu? Analysis of shared health messages for bio-surveillance*. Proceedings of the 4th Symposium on Semantic Mining in BioMedicine (SMBM'10). Cambridge. 2010: 18-26.
[4] Lampos V, Cristianini N. *Tracking the flu pandemic by monitoring the social web*. Proceedings of the 2nd International Workshop on Cognitive Information Processing (CIP). 2010: 411-416.
[5] Culotta A. *Towards detecting influenza epidemics by anlyzing twitter messages.* KDD Workshop on Social Media Analytics. 2010.
[6] Xiang J, Sooan A, James G. *Monitoring Public Health Concern Using Twitter Sentiment Classification*. IEEE International Conference on Healthcare Informatics.
[7] Kumar A, Sebastian T. Sentiment Analysis on Twitter. *IJCSI International Journal of Computer Science*. 9: 372.
[8] Irfan W, Taufik D, Wisnu A. Cluster Analysis for SME Risk Analysis Documents Bassed on Pillar K-Means. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2016; 14(2): 674-683.
[9] Balakrishnan G, Thiruchittampalam R, Nadarajah P, Pavalanathan P, Ashehan P. *Opinion Mining and Sentiment Analysis on a Twitter Data Stream*. The International Conference on Advances in ICT for Emerging Regions ICTer 2012. 2012: 182-188.
[10] Son D, Lucila O, Nigel C. *Enhancing Twitter Data Analysis with Simple Semantic Filtering: Example in Tracking Influenza-Like Illnesses*. IEEE Second Conference on Healthcare Informatics, Imaging and Systems Biology. 2012.
[11] Sivagowry, Durairaj, Persia. *An Empirical Study on Applying Data Mining Techniques for the Analysis and Prediction of Hearth Disease*. IEEE Second Conference on Healthcare Informatics, Imaging and Systems Biology. 2012.
[12] Xiuju F, Christina L, Harold S, Gary L, Terence H, Lee-Ching N. *Time-Series Infections Disease Data Analysis Using SVM and Genetic Algorithm*. IEEE Congress on Evolutionary Computation. 2007.
[13] Guo-Cheng L, Chao-Hui L, Yu-Yen L, Vincent S, Chu-Yu C, Miin-Luen D, Shyh-Chyi W, Ching-Nain C, Shyr-Yuan C. *Disease Risk Prediction by Mining Personalized Health Trend Patterns: A Case Study on Diabetes.* Conference on Technologies and Applications of Artifficial Intelligence. 2012.
[14] Hideo H, Liangliang W. *Prediction of Infectious Disease Spread using Twitter: A Case of Influenza.* Fifth International Symposium on Parallel Architectures, Algorithms and Programming. 2012.
[15] Nurul P. Analisis Sentimen Berdasarkan Isi Twitter Berbahasa Indonesia. Master Thesis. Bandung: Postgraduate ITB; 2012.