EMITTER International Journal of Engineering Technology
Vol. 6, No. 2, December 2018
ISSN: 2443-1168

Rule-based Sentiment Degree Measurement of Opinion Mining of Community Participatory in the Government of Surabaya

Berlian Juliartha Martin Putra, Afrida Helen, Ali Ridho Barakbah

Politeknik Elektronika Negeri Surabaya
Jln. Raya ITS, Sukolilo, Surabaya, 60111
Email : berlianjuliartha@pasca.student.pens.ac.id, helen@pens.ac.id, ridho@pens.ac.id

Abstract

Diskominfo Surabaya, as a government agency, received much community participatory for improvement of governmental services, with increasing number of 698, 2717, 4176 and 4298 participatory data respectively in 2011, 2012, 2013 and 2014. It is challenging for Diskominfo Surabaya to set a target by giving the response back within 24 hours. Due to task complexity to address the degree of participatory and to categorize the group of participatory, they faced difficulty to fulfill the target. In this research, we present a new system for measuring the sentiment degree of community participatory. We provide 5 functions in our system, which are: (1) Data Collection, (2) Data Preprocessing, (3) Text Mining, (4) Sentiment Analysis and (5) Validation. We propose our rule-based technique for the sentiment analysis of opinion mining with detection of 8 important parts, which are (1) Verb, (2) Adjective, (3) Preposition, (4) Noun, (5) Adverb, (6) Symbol, (7) Phrase, and (8) Complimentary. For applicability of our proposed system, we made a series of experiment with 408 data of community participatory in Twitter for Diskominfo Surabaya and compared with other sentiment classification algorithms which are SVM and Naive Bayes Classifier. Our system performed 81.33% rate of accuracy and outperformed to other comparing algorithms.

Keywords: Sentiment Analysis, Sentiment Classification, Twitter Opinion Mining, Diskominfo Surabaya.

1. INTRODUCTION

The main function of government serves their society. But, the government services still find much weakness. Those are known from many complaints that were posted on social media and print media[1]. Diskominfo Surabaya reports the complaint data that increase every year. There were

[1]https://dinkominfo.surabaya.go.id/old/download2.php?id=165 was accessed in 22 December 2017 at 12.29 wib.

698 complaints in 2011, 2.717 complaints in 2012, 4.176 complaints in 2013 and 4.298 complaints in 2014[2]. Those become suggestions for fixing the government performance so the Surabaya government builds a media center to handle those complaints.

Twitter becomes the second highest of a medium which receives complaints from the societies of Surabaya[1]. The twitter text consists of non-sentiment and sentiment. Non-sentiment doesn't relate to the government of Surabaya while sentiment consists of negative, neutral and positive sentiment. The negative sentiment is the complaint that must be responded immediately. Media Center has a rule, it must submit its response to the complaint of societies maximum 1x24 hours[2] but the twitter of Media Center doesn't respond like standard procedure. It needs attention to several factors in order to media center run as expected. Those are the openness to technological progress, quality of infrastructure, network, infrastructure and human resources[2]. So one of the media center challenges is to accelerate the response to public complaints by continuing to update the technology[2].

Sentiment analysis eases to know the sentiment of public opinion. It divides opinion into its sentiment. Sentiment analysis can be applied to any textual form of opinions such as blogs, reviews and microblogs[1]. Microblogs are the small text message such as twitter, facebook etc so this can be applied to the media center twitter. Researchers use sentiment analysis to analyze public opinion. This helps to determine what public say about products or service[2]. Many advantages of Sentiment analysis are product review[3], government feedback analysis[1][4][5][6][7], restaurants review[8], students' comments analysis[9], political review[10][11], movie review[12][13], depression detection and monitoring[14], airline service analysis[15] etc. This research focuses on government feedback analysis. Many previous researches used the government of Surabaya's twitter text[4][5][6]. In this research, we propose sentiment analysis using rule-based method approach to analysis the government of Surabaya's twitter text.

2. RELATED WORKS

Sentiment analysis for government feedback has been studied[1][4][5][6][7]. Sentiment analysis divides into the classification-based and rule-based method. The classification-based method uses machine learning methods[5][6][7]. While rule-based method uses sentiment lexicon and rule which build manually by the human[1][4][16][17][18].

Naiknaware, Kawathekar and Deshmukh proposed Indian government schemes classification[1]. The tweet was accessed over time with different topics. It used sentiment lexicons and scoring function. The scoring function calculated the score of each tweet. If the number of negative words was

[2]http://www.surabaya.go.id/pemerintahan/3892-media-center,-wadah-pengaduan publik-yang-diapresiasi-secara-internasional was accessed in 17 January 2017 at 23.06 wib.
[3]http://sentiwordnet.isti.cnr.it/

greater than the number of positive words, the score would be negative. The score would be neutral if the number of positive and negative words were same. And if the number of positive words was greater than the number of negative words, the score would be positive.

Lailiyah, Sumpeno and Purnama proposed sentiment analysis of public complaints. The data got from twitter and media center website of Surabaya government[4]. It used Indonesian sentiment lexicon and Sentiwordnet. Indonesian sentiment lexicon gave the best result. Faradhillah, Kusumawardani and Hafidz classified the Government's twitter data using Naive Bayes and Support Vector Machine(SVM) method in 2016[5]. SVM method gave the best accuracy. Another research, Nomleni, Hariadi and Purnama classified the Surabaya government's twitter using SVM method[6].

Laksana and Purwarianti also classified the tweets of Bandung's Government. In this research, They employed Naive Bayes, Decision Tree and SVM algorithm with Label Power Set and Binary Relevance for multi-label classification method. SVM method and classification multi-label technique Label Power Set, 1-gram feature and presence of complaint words gave the best result[7].

Rengga, Achmad and M. Udin proposed gender-based temporal sentiment analysis in Indonesian on Culinary Places in Surabaya City. It used temporal sentiment analysis based opinion mining[16]. The first process was text mining which gave important words. The second process was opinion analysis which based on table design and data aggregation. This was a temporal-based study so the result data classified based on time. The third process was opinion mining which used word degrees, stoplist and rule. This research got 67.32% accuracy.

Our previous research was Temporal Sentiment Analysis for Opinion Mining of ASEAN Free Trade Area on Social Media[17]. We proposed sentiment analysis about AFTA on the social media using a new approach temporal sentiment analysis. This process started from taking data. Data was taken from media social about AFTA. The next process was Text Mining. It was used for getting keywords which were used for searching trend. The next process was opinion mining from comment data for resulting comment value. Opinion mining used the word-degree and rule-based method. Word-degree created manually, it based on Sentiwordnet. The opinion mining result synchronized with news keyword and saved into a database. The last step was user interaction into a system. The user keyword became the search key of the analysis result. From this result gave the accurate of 60%.

We also proposed a new approach temporal sentiment analysis of the cellular phone operator service in our previous research[18]. Data got from twitter then the next process was text mining which gave keywords. Sentiment analysis used word-degree and rule-based method. Twitter comment projected based on time so the user can know about the sentiment of cellular phone operator service in certain time. This research got 50% accuracy.

Sentiment analysis for the government of Surabaya's twitter text has been done[4][5][6]. They used classification-based methods[5][6] and rule-based method[4] but it didn't consider type and location of words in a sentence[4]. In this research, the rule-based method was used on Twitter of Surabaya government too but it used our previous work method[17][18] which considers the types of word and location of words in a sentence by using the rule-based method and word-degree with some rule modification.

3. ORIGINALITY

In this research, we proposed a new approach analyzing the sentiment of the Surabaya Government's twitter text using the rule-based method. This method used word-degree to know the value of words and the rule-based method as the rule to calculate the sentiment value of a sentence. Word-degree consisted the list of words on Indonesian which was created manually based on Sentiwordnet. Word-degree values were -1, 0 and 1. A sentence has many words so the meaning of a sentence is supported by the composition of the words whereas every word has a word type. Therefore, a sentence has a different form with other sentences. According to those reasons, it created the rule based on the forms of the sentence. Each form has a rule that different each other. The rule created after researching some Indonesian sentences because the twitter data was Indonesian. The rule considered of type and location of words in a sentence which was created based on the structure of Indonesian sentences.

4. SYSTEM DESIGN

The problem of machine learning methods is typically not accurate for a simple task separating positive and negative sentiment on social media[2]. So this research used the rule-based method.

Figure 1 is the system design of our proposed research. We provide 5 functions in our system, which are: (1) Data Collection, (2) Data Preprocessing, (3) Text Mining, (4) Sentiment Analysis and (5) Validation. Each process is described in section 4.1 – 4.4.

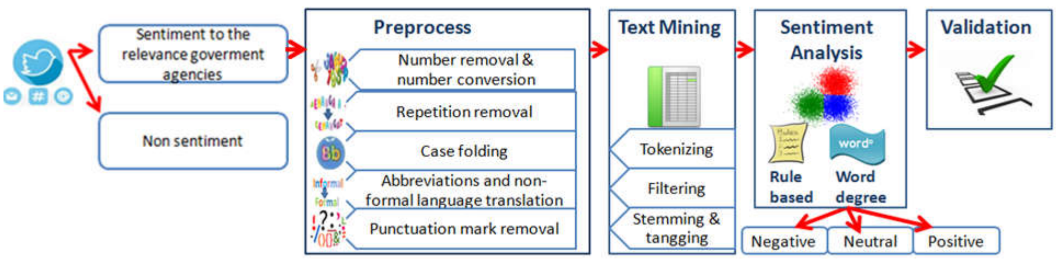

Figure 1. The System Design of our proposed research

4.1 Data Collection

The first step, the system got data. Data was gotten by the twitter4j library. Data was labeled by researchers manually like their class because it was used for validating the result of classification.

The twitter text problems are misspell, abbreviation and non-formal word. The Societies of Surabaya also use local language beside the Indonesian. This is an example of twitter texts which was written in Indonesian, local language, and non-formal word.

ibu nya bonek mana? kpn diskusi kyok ngene? @SapawargaSby

The tweets of Surabaya Government consist of sentiment and non-sentiment. The sentiment is the public opinion about their government. The non-sentiment examples are news and advertisement. The sentiment is divided into 3 classes, there are negative, neutral, and positive.

This is an example of the negative sentiment :

@wahju_wibowo:@SapawargaSby mohon diperhatikan kualitas sepanjang jl kayoon, banyak bekas pekerjaan utilitas asal tambal, tdk mulus dan rapi.

This is an example of the neutral sentiment :

Hari ini Saya melakukan kunjungan kerja ke Surabaya, didampingin Ibu Walikota Tri Risma Harini. @RismaHarini

This is an example of the positive sentiment :

Surabaya di bw kepemimpinan Ibu Risma akan selalu jadi cermin keindahan kota dr seluruh kota yg ada di indonesia, Surabaya ttp Berjaya...

4.2 Preprocessing

In Twitter, We often find the non-formal words which need preprocessing step for correcting the word. Those examples are misspell, abbreviation and non-formal word.

Preprocessing consists of number removing and number converting, repetition removing, case folding, abbreviation and non-formal converting, punctuation mark removing (except dot, comma, question mark and exclamation mark).

Number removal removes all number[5] which don't have the function on the text. The example is *2 jam*, the system will delete *2* and will use *jam* to become a keyword. Number conversion converts numbers in a word into proper character[19]. The example is *se7*, it will change into *setuju*. Repetition removal removes the repeated character in a word[19]. The example is *setuuuujuuuu*, it will change into *setuju*. Case folding replaces all character into lowercase[5]. The example is *Saya setuju dengan Pemerintah Surabaya*, it will change into *saya setuju dengan pemerintah surabaya*. Abbreviation and non-formal conversion[5] correct abbreviation and nonformal words. The example is *sy ogah melok aturan*, it will change into *saya tidak mau ikut aturan*. Punctuation mark removal removes all punctuation mark[5] except dot, comma, question mark and exclamation mark because they use as separated words for two or more sentences.

4.3 Text Mining

Text mining process gives the result as keywords. These Keywords are used in the sentiment analysis. Tokenizing, filtering, stemming, tagging and analyzing are the text mining processes[17][18][20]. This research skips analyzing process because it wants to determine opinion into its sentiment.

Tokenizing is a process to break character sequences into tokens which are used in the next processing[21][22]. Filtering is a process to remove unimportant words or so-called stoplist[20]. It removes all words which include on the stoplist. Stoplist is a list that contains words and unrepresentative words, those cannot represent a document. Unrepresentative can be conjunction, auxiliaries etc[20]. Stemming is a process to delete affix, suffix, prefix etc. Tagging is a process to change the result of stemming into its root word.

4.4 Sentiment Analysis

Sentiment Analysis uses classification-based and rule-based method[23]. The rule of the classification-based method is built by machine learning algorithm which are SVM[23], Naive Bayes, Decision Tree etc while rule-based method uses sentiment lexicon and rule. It builds rule manually by the human.

The sentiment lexicon of this research was created manually by researchers based on Sentiwordnet. Sentiwordnet was built for supporting sentiment analysis. It has negative and positive score[3]. Each word has a score between 0 and 1[3]. A word score can be converted from -1 to 1 [16][17][18] or -3 to +3[9].

There are many variations of rule-based technique. Because the meaning of each word must be known and the rule of words in a sentence depends on the language used[24]. Even though any standard rule in language, but it cannot be applied completely[24]. The rule can define by SUM value form each word value and then compare the result but some researchers consider types of the word[9][16][17][18]. It can divide into adjective[9][16][17][18], verb[9][16][17][18], adverb[9], intensifier[9], preposition[16][17][18], symbol[16][17][18], modal operators[25] and modifier[25]. The sentiment value of sentence also consider the word location on the sentence[9][16][17][18] and it uses the certain rule-based.

This research used the list of sentiment lexicon which its word values only used -1, 0 and 1 then we call it as word-degree. It also used the rule-based method that considers type and location of the word in an Indonesian sentence.

4.4.1 Word-Degree

Word-Degree is a list of Indonesian words which have value. Each word has the different value. The list of word-degree was created manually based on Sentiwordnet which can be accessed in http://sentiwordnet.isti.cnr.it/

[3]http://sentiwordnet.isti.cnr.it/ was accessed in 17 December 2017 at 23.06 wib.

because Indonesian sentiwordnet is difficult to find[16][17][18]. Figure 2 shows how to convert sentiwordnet score to word-degree score. If the positive score of a word smaller than the negative score of a word so the word-degree value is -1. If the positive score of a word same as the negative score of a word so the word-degree value is 0. If the positive score of a word greater than the negative score of a word so the word-degree value is 1.

Figure 2. The Score of Sentiwordnet Converted into Word-Degree

The society opinion has a sentiment value, It got from the value of each word which was added and also considered a rule to calculate the sentiment value. The rule is explained in the next point.

4.4.2 Rule-Based

The rule-based depends on the structure of Indonesian sentences. The set of words value depends on this rule. The rule was created after researching some Indonesian sentences. It analyzed the location of words and the type of words in a sentence[16][17][18]. The word types divide into adjective, verb, adverb, noun, preposition, symbol, phrase and complimentary. The rule definitions are described below.

Figure 3 shows rule-based method architecture. The first process is word types detection. The second process is the form detection because the meaning of a sentence depends on the words and the arrangement of its words. This research built rules based on the forms which were analyzed. The last process is to calculate the value of words and the set of word that have rules with SUM operator. The total value is used to define the sentiment. If the total value less than 0 so the sentiment is negative, if the total value same as 0 so the sentiment is neutral and if the total value greater than 0 so the sentiment is positive. The rules were described below.

4.4.2.1 Single Adjective

If a sentence contains adjective without following verb and preposition, so the sentence value is the adjective value[16][17][18].

warga marah

The value of *marah* is -1. This value gets from the list of word-degree that creates before so the sentence value is -1(negative).

4.4.2.2 Single Verb

If a sentence contains verb without adjective and preposition, so the sentence value is the verb value[16][17][18]. Example :

warga setuju

The value of *setuju* is 1. The value gets from the list of word-degree that creates before so the sentence value is 1(positive).

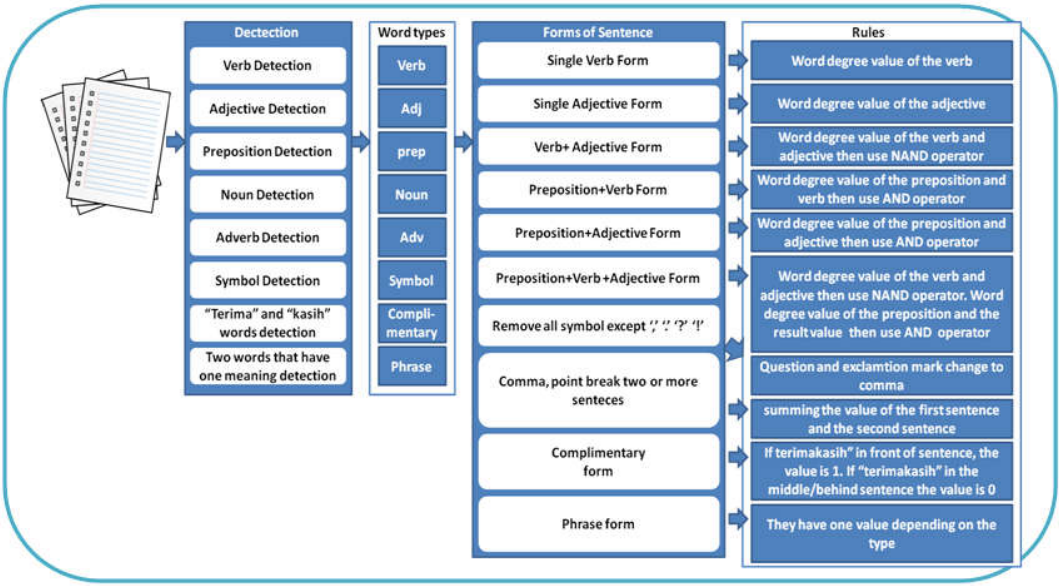

Figure 3. The Architecture of Rule-Based

4.4.2.3 Preposition + Adjective

If there is an adjective after a preposition, so the sentence value is calculated by AND operator[16][17][18]. Example:

warga tidak sabar

The value of *tidak* is -1 and *sabar* is 1 so the AND operator value gives -1 (negative).

4.4.2.4 Preposition + Verb

If there is a verb after a preposition so the sentence value is calculated by AND operator[16][17][18]. Example:

warga tidak setuju

The value of *tidak* is -1 and *setuju* is 1 so the AND operator value is -1 (negative).

4.4.2.5 Verb + Adjective

The Opposite of NAND can represent the sentiment value which contains verb and adjective. Table 1. is the rule of NAND definition[16][17][18]. Example :

Dia memahami dengan baik

The basic word of *memahami* is *paham*. The value of *paham* is 1 and *baik* is 1. Based on the table above, the sentiment value result is 1 (positive).

Table 1. Verb + Adjective rule

Verb	Adjektif	Value
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1

4.4.2.6 Preposition + Verb + Adjective

Sometime preposition, verb and adjective are seen together in a sentence. First, the system counts verb and adjective using NAND operator. Then it gives a result. Second, the system counts preposition and the result value before using AND operator[16][17][18]. Example :

Warga tidak memahami dengan baik

memahami is a verb and the value is 1, *baik* is an adjective and the value is 1. Based on the NAND operator, the result is 1 then *tidak* is a preposition. Its value is -1 so the result of -1 AND 1 is -1 (negative).

4.4.2.7 All symbols on tweet is deleted except dot, comma, exlamation mark, and question mark

All characters or symbols will be deleted, except dot, comma, question mark and exclamation mark. Exclamation mark and question mark are changed into comma[16][17][18].

4.4.2.8 Dot and comma break the sentence

Each Comment usually consists of multiple sentences, it is separated by dot and comma[16][17][18]. Example :

warga marah, mengurus ktp sulit

The comment will be divided into 2 sentences. First *warga marah* and second *mengurus ktp sulit*. The value of the first sentence is -1 and the value of the second sentence is -1 so the sentiment value result is -1+-1=-2(negative).

4.4.2.9 Two words which have one meaning cannot be seperated

In this research, phrase is two words which have one meaning usually meet on Indonesian sentences. Those examples are *terima kasih, air mata, sepak bola* etc. Those have only one value.

4.4.2.10 The value of "terima kasih" or complimentary depend on its location

If *terima kasih* is in front of a sentence, the value is 1. But if If *terima kasih* is in the middle or behind of a sentence, the value is 0. Example :

terima kasih atas pelayanannya

This sentence means positive sentiment because *terima kasih* in front of a sentence so the value of *terima kasih* on this sentence is 1. It will help to give positive value. The second Example :

tolong pohon segera dipangkas. terima kasih

The second example maybe means negative or neutral sentiment because *terima kasih* is in the behind of a sentence so the value of *terima kasih* is 0.

4.4.2.11 Words are not on Indonesian Dictionary will be ignored

We usually meet non-formal, misspell and English language on the government of Surabaya's twitter text. Those can be fixed in the preprocessing step but if the result is still not accordance with words in Indonesian dictionary, it will be ignored and will be considered as unknown.

5. EXPERIMENT AND ANALYSIS

This point discusses system performance. We analyzed 3 sentences. We will see the system to finish processes and to give system performance. In this experiment, our approach also was compared by SVM and Naive Bayes.

5.1 Preprocessing and Text Mining Experiment

It was taken 3 comments that represented each sentiment. Table 2 is sentences before text preprocessing and text mining.

Table 2. The Example of Sentences

Comment	
1	@sbotv 4 Okt traffictlight depan PGS ini sudah tertutup pohon #pemkotsurabaya #dkp http://fb.me/2HLhdYBTf
2	@KaraMozavie:@Tri_Rismaharini pagi mba' .. assalamualaikum
3	@tweetspiring:Luwar biyasa @DivHumasPolri @PemkotSurabaya @Tri_Rismaharini https://t.co/X7ozOwXJXL

Table 3. was the text preprocessing and text mining result. It shows keywords. The Keywords represented a sentence.

Table 3. Text Mining Result

Keywords of Comment 1	Keywords of Comment 2	Keywords of Comment 3
oktober trafictlight pgs sudah tutup pohon pkotsurabaya dkp http fb . me hlhdybtf	rismaharini pagi kakak . assalamualaikum	luar biasa rismaharini htpp . co x7ozowxjxl

Table 3 knows that the numbers were removed, the punctuation marks were removed except point, the informal words were changed into formal words, the words were changed into lower case and split into tokens. The unimportant words were removed by the filtering process. The words also were changed into the root of words by stemming and tagging process. So the text mining process was done correctly.

5.2 Sentiment Analysis Experiment

The next step was sentiment analysis. The sentiment analysis calculations are shown in Table 4, Table 5 and Table 6.

Table 4. The Sentiment Analysis Result of first sentence

keyword comment 1	type	value	rule/not rule
oktober	noun	0	not rule
trafictlight	noun	0	not rule
depan			
pgs	uknown	0	not rule
ini			
sudah	preposition	1	rule
tutup	adj	-1	rule
pohon	noun	0	not rule
pkotsurabaya	noun	0	not rule
dkp	uknown	0	not rule
http	uknown	0	not rule
fb	uknown	0	not rule
.	symbol	0	not rule
me	uknown	0	not rule
hlhdybtf	uknown	0	not rule
Total		-1	negative

Table 5. The Sentiment Analysis Result of second sentence

keyword comment 2	type	value	rule/not rule
rismaharini	uknown	0	not rule
pagi	noun	0	not rule
kakak	noun	0	not rule
.	symbol	0	not rule
assalamualaikum	uknown	0	not rule
Total		0	neutral

A sentence has words. Their type and value of words were different each other. In Table 4, a preposition followed by an adjective, it used rule in point 4.4.2.3 so it used AND operator. The value of *sudah* was 1 and the value of *tutup* was -1. Using AND operator, the value result was -1 while the value of other words were zero so the sentiment value of this sentence was -1. In

Table 5, It was nothing rules occurrence so it can be used SUM to calculate all the value of words.

Table 6. The Sentiment Analysis Result of third sentence

keyword comment 3	type	value	rule/not rule
luar biasa	adj	1	rule
rismaharini	uknown	0	not rule
http	uknown	0	not rule
.	symbol	0	not rule
co	uknown	0	not rule
x7ozowxjxl	uknown	0	not rule
Total		1	positive

Table 6 was found phrase form, there is like rule point 4.4.2.9. Those value can be added to the other value of words which were not rule occurrence. The total value of the first example less than 0, so it is a negative sentiment. The total value of the second example same as 0, so it is a neutral sentiment and the total value of the last example sentence greater than 0 so it is a positive sentiment

Sometimes, the sentiment analysis gave the false prediction. The reason that affected false prediction will describe below.

- Question sentences used for saying negative sentiment so it was not used negative words. For example, *kapan GBT diperbagus?*
- A negative sentiment which used positive word. For example, *sudah dijelaskan kalau posisi saya diluar kota tetapi tidak didengarkan. Mantab ya pelayanannya.*
- Sometimes, Stemming changed the type of word. For example, *kegiatan* changed into *giat.*

5.3 System Performance

This was the result of the rule-based method using 408 data. Validation used LOO. It gave 81.33% accuracy. Table 8 shows the accuracy of the rule-based method then it was compared by the classification-based method.

The classification-based method used machine learning method. Basic feature types for classification-based are unigram, n-gram, punctuation, pattern features[26]. It commonly uses unigram or n-gram[24][27]. Unigram and complaint words feature got the best accuracy[7]. So, this research used unigram. Unigram assumes one word as one feature and it is unique. Then system built metadata aggregate. It is a big matric. The columns are features and the values are term frequency(TF) each word on twitter text. TF is the frequency value of a term t_i occurs on a document[28][29]. A term maybe appears more than one in a document. Formula 1 is TF formula.

$$tf = 1 + \log(n) \tag{1}$$

n : the value of times that term occurs in a document

The 0 value is the feature value which never appears in twitter text, the 1 value is the feature value which appears one time in twitter text and the value 1,... is the feature value which appears more than one in twitter text. This research used WEKA to handle classification-based method.

We compared the result of our research using SVM and Naive Bayes method because they is usually used for classify sentiment[5][6][7][12][30][31][32]. Using LOO validation, It gave 69.36% accuracy for SVM using SMO function and 71.1% accuracy for Naive Bayes.

Table 7 shows that rule-based method gave better accuracy than SVM and Naive Bayes method. Because rule-based method considers word-degree and location a word on Indonesian text behavior. While classification-based method uses features to classify data, even though one feature almost is used in others classes.

Table 7. Comparasion of rule-based, SVM and Naive Bayes method

	Rule-based	Classification-based	
		SVM	Naive Bayes
Negative accuration	84%	82.1%	79.5%
Neutral accuration	76.3%	62%	64.3%
Positive accuration	83.14%	70.4%	72.2%
Accuration	81.33%	69.36%	71.1%

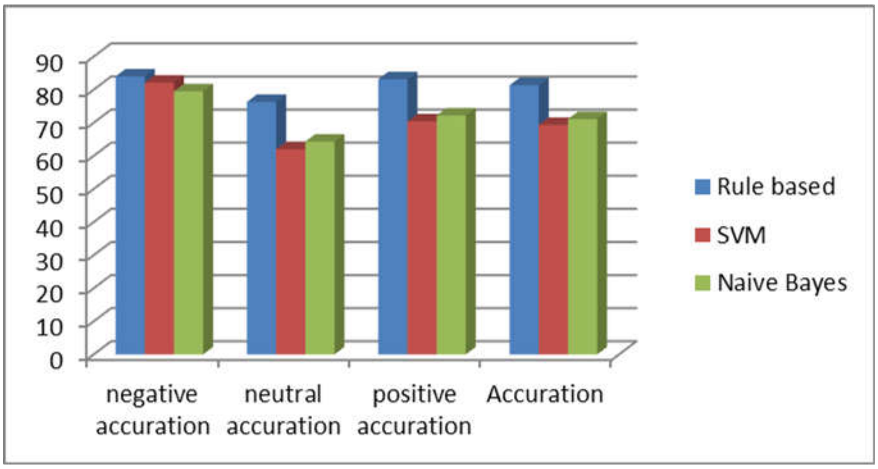

Figure 4. The Comparison Chart of rule-based, SVM and Naive Bayes method

Figure 4 shows Rule-based method gave the best accuracy. SVM was ranked second and the last was Naive Bayes.

6. CONCLUSION

The tweets of Surabaya's Government consist of sentiment and non-sentiment. Sentiments are used as feedback for the government of Surabaya. Sentiment divides into 3 classes. There are positive, neutral and negative.

In this research, we provide 5 functions in our system: (1) Data Collection, (2) Data Preprocessing, (3) Text Mining, (4) Sentiment Analysis and (5) Validation. In this experiment, we used 408 data of community participatory in Twitter of Diskominfo Surabaya. Data Preprocessing is used for normalizing text. Text mining is used for searching keywords which can represent a sentence. In this research, Sentiment analysis used the word-degree and rule-based method. Word-degree was the value of the word which created manually according to Sentiwordnet. The rule-based method is a set of rules to calculate the set of words value. The rule based on the structure of Indonesian sentences and created after researching some sentences. The first, system detected 8 type of words which are: (1) Verb, (2) Adjective, (3) Preposition, (4) Noun, (5) Adverb, (6) Symbol, (7) Phrase, and (8) Complimentary. The second, it detected the form of a sentence. Each form has a rule so the value of set words depend on this rule. The word that didn't identify as a rule, it will be identified as not-rule. The sentiment value get from calculating all value that we get from the word sets which are detected as a rule and all words which are not detected as not-rule by SUM operator. If the total value greater than 0 so the sentence is a positive sentiment, If the total value same as 0 so it is a neutral sentiment and If the total value less than 0 so it is a negative sentiment.

The rule-based method compared with SVM and Naive Bayes method. Using LOO validation, rule-based method gave 81.33% accuracy, SVM method using SMO function gave 69.36% and Naive Bayes method gave 71.1% so the rule-based method gave the best result.

The accuracy results of rule-based methods depend on the completeness of word-degree and the rules. Future research could focus on adding new rules and completing the list of word in the word-degree dictionary.

Acknowledgements

We want to thank for the government of Surabaya which has allowed us to use the crawling twitter data in our research.

REFERENCES

[1] Bharat R. Naiknaware, Seema Kawathekar, Sachin N. Deshmukh, **Sentiment Analysis of Indian Government Schemes Using Twitter Datasets**, *IOSR Journal of Computer Engineering*, pp. 70-78.

[2] Shreya Banker and Rupa Patel, **A Brief Review of Sentiment Analysis Method**, *International Journal of Information Sciences and Techniques*, Vol. 6, No.1/2, pp. 89-95, 2016.

[3] Alexandra Cernian, Valentin Sgarciu and Bogdan Martin, **Sentiment Analysis from Product Reviews Using Sentiwordnet as Lexical Resource**, *Electronics, Computers and Artificial Intelligence*, pp. WE-15, 2015.

[4] M. Lailiyah, S. Sumpeno and I.K.E. Purnama, **Sentiment Analysis of Public Complaints Using Lexical Resources Between Indonesian

Sentiment Lexicon and Sentiwordnet, *International Seminar on Intelligent Technology and Its Application*, pp. 307-312, Surabaya, 2017.
[5] Nuke Y.A. Faradhillah, Renny P. Kusumawardani and Irmasari Hafidz, **Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin**, *Seminar Nasional Sistem Informasi Indonesia*, pp. 15-24, Surabaya, 2016.
[6] Petrix Nomleni, Mochamad Hariadi and I Ketut Eddy Purnama, **Sentiment Analysis Based Big Data**, *Seminar Nasional ke-9 : Rekayasa Teknologi Industri dan Informasi*, pp. 142-149, 2015.
[7] Janice Laksana and Ayu Purwarianti, **Indonesian Twitter Text Authority Classification For Goverment in Bandung**, *International Conference of Andvanced Informatic: Concept, Theory and Application*, pp. 129-134, Bandung, 2014.
[8] Yu Mon Aye and Sint Sint Aung, **Sentiment Analysis for Reviews of Restaurants in Myanmar Text**, *International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel*, pp. 321-326, Kanazawa, 2017.
[9] Khin Zezawar Aung and Nyein Nyein Myo, **Sentiment Analysis of Students' Comment Using Lexicon Based Approach**, *International Conference on Computer and Information Science*, pp. 149-154, Wuhan, 2017.
[10] Md. Azza F. Yatim, Yulistiyan Wardhana, Ahmad Kamal, Anandra A. R. Soroinda, Febryan Rachim, M. Ismail Wonggo, **A Corpus-Based Lexicon Building in Indonesian Political Context Trough Indonesian Online News Media**, *International Conference on Advanced Computer Science and Information Systems*, pp. 347-352, Malang, 2016.
[11] Anass Bensrhir, **Big Data for Geo-political Analysis: Application on Barack Obama's Remark and Speeches**, *International Conference on Computer Systems Applications*, pp. 1-4, Casablanca, 2013.
[12] P. Kalaivani, Dr. K. L. Shunmuganathan, R.M.K. Engineering College, **Sentiment Classification of Movie Reviews By Supervised Machine Learning Approaches**, *Indian Jorunal of Computer Science and Engineering*, Vol.4, No.4, pp. 285-292, 2013.
[13] Rasika Wankhede and A.N. Thakare, **Design Approach for Accuracy in Movie Reviews Using Sentiment Analysis**, *International Conference on Electronics, Communication and Aerospace Technology*, Vol.1, pp. 6-11, Coimbatore, 2017.
[14] Chiara Zucco, Barbara Calabrese, Mario Cannataro, **Sentiment Analysis and Affective Computing for Depression Monitoring**, *International Conference on Bioinformatics and Biomedicine*, pp. 1988-1995, Kansas, 2017.
[15] Yun Wan and Dr. Qigang Gao, **An Ensemble Sentiment Classification System of Twitter Data for Airline Service Analysis**, *International Conference on Data Mining Workshop*, pp. 1318-1325, Atlantic, 2015.

[16] Rengga Asmara, Achmad Basuki and M. Udin Harun Al Rasyid, **Gender Based Temporal Sentiment Analysis in Indonesian on Culinary Places in Surabaya City**, *International Journal of Engineering and Technology Innovation*, Vol. 7, No. 4, 2017.

[17] Mustofa Kamal, Ali Ridho Barakbah and Nur Rosyid Mubtadai, **Temporal Sentiment Analysis for Opinion Mining of ASEAN Free Trade Area on Social Media**, *Knowledge Creation and Intelligent Computing*, pp. 207-212, Manado, 2016.

[18] Ni putu Nanda Aristiadewi, Tita Karlita, Nur Rosyid Mubtadai, Ali Ridho Barakbah, **Analisis Sentimen Opini Masyarakat terhadap Jasa Operator Telepon Seluler di Media Sosial Twitter**, *Jurnal Proyek Akhir PENS*, Vol. 1, 2016.

[19] Aqsath Rasyid Naradhipa, Ayu Purwarianti, **Sentiment classification for Indonesian message in social media**, *International Conference on Electrical Engineering and Informatics*, pp. 1-5, Bandung, 2011.

[20] Muh. Subhan, Amang Sudarsono and Ali Ridho Barakbah, **Classification of Radical Web Content in Indonesia using Web Content Mining and k-Nearest Neighbor Algorithm**, *EMITTER International Journal of Engineering Technology,* Vol. 5, No. 2, pp. 328-348, 2017.

[21] Twinkle Svadas, Jasmin Jha, **Document Cluster Mining on Text Document,** *International Journal of Computer Science and Mobile Computing,* Vol. 4, Issue. 6, pp. 778-782, 2015.

[22] Durga Bhavani Dasari, Dr. Venu Gopala Rao. K, **Text Categorization and Machine Learning Method: Current State of the Art**, *Global Journal of Computer Science and Technology***,** Vol. 12, Issue 11, pp. 37-46, 2012.

[23] Aldy Rialdy Atmadja and Ayu Purwarianti, **Comparison on the Rule based Method and Statistical based Method on Emotion Classification for Indonesian Twitter Text**, *International Conference on Information Technology Systems and Innovation*, pp. 1-6, 2015.

[24] Asniar and B.R. Aditya, **A Framework for Sentiment Analysis Implementation of Indonesian Language Tweet on Twitter**, *International Conference on Computing and Applied Informatics*, Vol. 801, No. 1, 2016.

[25] Alena Neviarouskaya, Helmut Prendinger, Mitsuru Ishizuka, Member and IEEE, **Sentiful: A Lexicon for Sentiment Analysis**, IEEE Transactions on Affective Computing, Vol. 2, No.1, pp. 22-36, 2011.

[26] Dmitry Davidov, Oren Tsur, Ari Rappoport, **Enhanced Sentiment Learning Using Twitter Hashtags and Smileys**, *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 241-249, Beijing, 2010.

[27] Payal B. Awachate, Prof. Vivek P. Kshirsagar, **Improved Twitter Sentiment Analysis Using N Gram Feature Selection and Combinations**, *International Journal of Advanced Research In Computer and Communication Engineering*, Vol. 5, Issue 9, pp. 154-157, 2016.

[28] Tian Xia, Yanmei Chai, **An Improvement to TF-IDF: Term Distribution based Term Weight Algorithm**, *Journal of Software*, Vol.6, No.3, pp. 413-420, 2011.
[29] Ying Liu, Han Tong Loh, Aixin Sun, **Imbalanced Text Classification: A Term Weighting Approach**, *Expert System with Application*, No.1, pp. 690-701, 2007.
[30] Elly Susilowati, Mira Kania Sabariah, Alfian Akbar Gozali, **Implementation Support Vector Machine Method for Traffic Jam Classification on Twitter**, *Jurnal Universitas Telkom*, Vol.2, No.1, pp. 1478-1484, 2015.
[31] Jayashri Khairnar and Mayura Kinikar, **Machine Learning Algorithm for Opinion Mining and Sentiment Classification**, *International Journal of Scientific and Research Publication,* Vol.3, Issue 6, pp. 1-6, 2013.
[32] Jumadi, Dian Sa'adillah Maylawati and Beki Subaeki, **Opinion Mining on Twitter Microblogging Using Support Vector Machine : Public Opinion abouy State Islamic University of Bandung**, *International Conference on Cyber and IT Service Management*, pp. 1-6, Bandung, 2016.