International Journal of Engineering & Technology, 7 (4.44) (2018) 110-116

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Geo-Sentiment Analysis as a Location-Based Opinion Analysis System on Public Opinion Data about Governor Candidates

Imam Fahrur Rozi[1]*, Dika Rizky Yunianto[2], Mustika Mentari[3], Awan Setiawan[4], Rudy Ariyanto[5], Indrazno Siradjuddin[6]

[1,2,3,4,5,6] Information Department, State Polytechnics of Malang, Malang, Indonesia
**Corresponding author E-mail: imam.rozi@polinema.com*

Abstract

Ahead of governor elections, there were a lot of news and opinions related to the candidates through social media. The candidates could map the positive public opinions as their political supports that need to be strengthened, and the negative opinions that need for correction. To map those opinions, it is necessary for an opinion classification system from textual opinions. It became the focus of this research. The system was designed to work on textual opinions in Bahasa since the proposed case study was the opinion of East Java governor candidates mainly written in Bahasa. Classification method that was used to classify the opinions in this system, is Naive Bayes Classifier (NBC). The opinions would be classified into 2 classes, negative and positive opinion. The classified opinions then grouped by region. It would make users easier to map the opinion in each region. The visualization became more user-friendly since the count of classified opinion displayed as a pie chart on a geographical mode or a map. After testing on the classification results, the accuracy value that we got was 78%. It indicated that NBC could perform very well as a simple text classification method with a good result.

Keywords: *Sentiment Analysis, Regional Election, Geosentiment, Naïve Bayes Classifier*

1. Introduction

Considering the opinions is usually important to support the decision making process. For example, the customers of online shop usually read the reviews/opinions of a product that they will buy. If there are many good opinions on it, then the customers will be encouraged to buy it. Otherwise, they will be discouraged. And for product maker, by analyzing reviews of their product, they will be able to make product evaluation and improvement. In other case, politicians usually use social media networking to campaign and increase their popularity ahead of the election. It shows that nowadays, there are numerous opinion data on the internet. It could be a valuable asset to those who can mine the knowledge inside.

In the democratic political system in Indonesia that allows people to directly elect their candidate, social media and website take an important place as a campaign media for elections. By using social media, people can easily access the latest political news and public opinion that may not be published in national newspapers or television. People are able to easily post their opinions through the website as news comments and social media as well. So that, it is important for candidate or politician to be aware with opinion posted on the news website and social media. Twitter is one of the most popular social media for Indonesian. People are interested to use it because it is easy to access, it has no limitation for followers, and it limits a post contains up to 140 characters. Indonesia occupies the third position as a country with the most Twitter users. According to Twitter CEO Dick Costlo, in the middle of 2015, it reached 50 million [1]. It shows that social media, especially Twitter becomes another space where many Indonesian people gather and share their opinions. Some of the political participations of people in Twitter could be following candidates/politician, giving comments, posting opinions and following political parties account [2]. People become more interested in posting their political opinion through Twitter [3]. In anoother side, Twitter as social media, becomes prospective campaign media that could be used by political parties/politicians/candidates to reach people and voters instantly [4]. By knowing the trending topic and its public opinion that vastly spread in social media, candidates or politician can make some adjustments on their campaign strategy as well as do opposition research [5]. So that, any further analysis will be relevant to be carried out. The problem will arise when the analysis is performed on the text by text, opinion by opinion and tweet by tweet manually. In this case, a computing system is needed to analyze the collections of textual opinion or sentiment, that is known as sentiment analysis [6].

With massive users, Twitter will be able to generate big volumes of opinion texts in the form of tweets which is available for the sentiment analysis [7]. Some previous researches have focused on sentiment analysis applied Twitter data [6][8][9][10][11]. They have implemented a machine learning method to analyze the sentiment, and some of them combine with another method. One of machine learning method that is popular in sentiment analysis as well as text classification because of its simplicity and its good result is Naïve Bayes Classifier (NBC).

It has been used by Sandi and Edi (2013) to classify tweet of traffic congestion in Bandung with the highest accuracy reaching 93.58% [12]. Ahmad and Azhari (2014) also conducted a research on classifying tweets that contain public sentiments on certain public figures using NBC and term frequency feature with accuracy 79.91%. In the different research, by using NBC and TF-IDF feature reached 79.68% [13]. The identical result has been obtained by the research that combined NBC with HMM POS tagger with the best precision and recall are 0.95 and 0.94 [14]. To accommodate that sometimes users post their tweet with emoji which is the simplest way to express emotion and is so popular to use, a research added a weighting of emoji [15]. Some of tweets are satire opinion, in which the sentiments contained in non-textual are more dominating than the sentiments contained in textual (sentences), which leads to ambiguity of opinion [16]. With the emoticons weighted, it is expected that the sentiment results will really have a clear boundary between positive and negative sentiments [17] so that the accuracy is increased. Some of other previous researches addressed different method on sentiment analysis, such as Hybrid Topic Based Sentiment Analysis (HTBSA) [18] and combination of network science and sentiment analysis [19], towards the political elections. In different case, novel approach based on the combination of word-based n-grams and character-based q-grams addressed as sentiment analysis method with the best accuracy 65,53% [20]. Since a good result of NBC, this research will develop the sentiment analysis system to analyze public opinion on regional (East Java) head candidates by using NBC as well, with TF-IDF feature. This research focuses on the opinion data in Bahasa. The East Java head election was chosen as a case study in this research because the beginning time of this research is close to the election time.

The sentiment analysis system developed in this research will classify the tweets that have been crawled from Twitter, into positive and negative class. People can use the system to analyze the trend of a regional head candidate, whether the positive is more dominant or vice versa. As well as people, the candidate or politician can analyze the trend of their own opinion and trend other candidate or competitor opinion. Candidates or politician may need a more specific feature, likes visualizing the opinion recapitulation in each region/area. By knowing the distribution of opinion in each area, candidates can map out the support and campaign strategies. Which area that already gives a good opinion, so that it needs a strengthening opinion, and which area that contributes many negative sentiments, so that it needs to be countered and needs a reconciliation. And of course, by visualizing the sentiment distribution in a geographical mode, it will make it more user-friendly.

2. Research Method

2.1. Geosentiment Analysis

Geosentiment is a combination of geoinformatics and sentiment analysis. Sentiment analysis is related to computational research to analyze the opinions, sentiments, and emotions that expressed textually. Sentiment analysis aims to extract attributes from textual comments (opinions, sentiments, and emotions). One of the main result of sentiment analysis is positive or negative orientation of each opinion [21]. By using sentiment analysis system, all collected opinions could be determined its orientation automatically. Then, the opinion is grouped by the location where it comes and is visualized in geographical view or map. The result of geosentiment analysis will help to those who need to analyze the opinion based on the location where it is issued.

2.2. Regional Heads Election

UU No 32 of 2004 brings the revision for UU No 22 of 1999. In UU No 22 of 1999, DPRD that is the representative of people, does the election for regional heads. And it has been revised by UU No 32 of 2004 that tells that regional heads will be directly elected by people. DPRD will no longer elect the regional head. It gives the chance for people to be actively involved in a very strategic decision-making process of regional governance through this direct regional head election [22]. And for regional head candidates, it will change the way of campaign. Now, they need to reach people directly to promote their strategical vision and mission. They need to acquire the people opinion massively, towards giving the constructive feedback on it.

2.3. Naïve Bayes Classifier (NBC)

One of the popular classifier method used in text classification is NBC. NBC is a classification method based on the Bayes theorem. The main characteristic of the NBC is a very strong (naïf) assumption of the independence of each condition/event. It is a supervised learning method. It learns from the collection of training dataset. The learning will produce a probabilistic model that will be used to classify a testing data. In the training dataset, all data is classified into *k* classes/categories, so that the class will be $C_j = \{C_1, C_2, C_3, \dots, C_k\}$ and the prior probability of each class will be $p(C_j)$ where *j=1,2,3,...,k*.

Document that will be classified, is represented by $d_i = (w_1, \dots, w_j, \dots, w_m)$ with w_j is collection of features/words in a document. The document will be classified into C_j class/category.

To classify document d_i into C_j, is done by calculating probability value of all documents (posterior probability). Equation (1) shows the formula to calculate posterior probability of class C_j if given the document d_i.

$$p(C_j|d_i) = \frac{p(d_i|C_j)p(C_j)}{p(d_i)} \quad (1)$$

NBC based text classification could be done by maximizing the result of equation (1). Since the value of $p(d_i)$ will be the same or constant for any class or category, then it could be ignored. And the posterior probability equation could be wrote as equation (2).

$$\max_{C_j \in C} p(C_j|d_i) = \max_{C_j \in C} p(d_i|C_j)p(C_j \quad (2)$$

Based on the Bayesian hypothesis which states that every feature or word $w_1, \dots, w_j, \dots, w_m$ in the document d_i is independent, the total probability distribution (also known as likelihood) is the product of the probability distribution of each feature or word, as shown in the equation (3).

$$p(d_i|C_j) = p(w_1, \dots, w_j, \dots, w_m|C_j) = \prod_{i=1}^{m} p(w_i|C_j) \quad (3)$$

By substituting equation (3) to equation (2), then equation (2) will become equation (4).

$$\max_{C_j \in C} p(C_j|d_i) = \max_{C_j \in C} p(C_j) \prod^{m} p(w_i|C_j) \quad (4)$$

Equation (4) is known as classifying formula. The value of $p(C_j)$ could be calculated by using equation (5), while equation (6) will be used to calculate the value of $p(w_i|C_j)$.

$$p(C_j) = \frac{N_{cj}}{N} \quad (5)$$

$$p(w_i|C_j) = \frac{N_{cw}+1}{N_c+V} \quad (6)$$

N_{cj} is the number of documents in the training data (dataset) classified to class C_j, while N is the total number of documents in dataset. N_{cw} is the number of words w_i in the dataset or training data classified in class C_j, N_c is the number of all unique words in the dataset classified in class C_j, and V is the total number of unique words in the dataset. And to avoid 0 in numerator in equation (6), then it is added 1 called Laplace Smoothing [21].

2.4. System Design

Steps to analyze the sentimen is shown in Figure 1. The sentiment analysis system globally consists of subprocess data collection, preprocessing, classification and visualization of the results of the analysis. Data collection is conducted by crawling text commentar or opinion from social media, especially Twitter. Tweets that discuss or talk about regional head election, are collected in a certain time. Beside the tweet, the location where the tweet posted, is also collected. The location represent the area where the Twitter user posts the opinion. Collected data will be used as training data and testing data. Training data will be used in learning process to build a probabilistic model to classify the testing data.

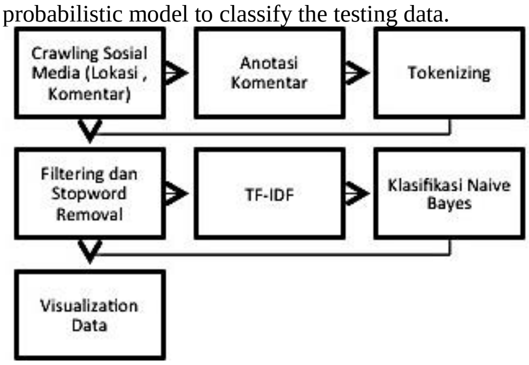

Fig. 1: System Design Chart

All collected data will be given a label/anotation that represents in which class it being classified. Labelling or anotating the tweet is done manually, and not automatically. After the comment annotation is completed, the next step is tokenizing. Tokenizing is the process of breaking sentences into words. After tokenizing, collections of words will be produced. And it will need to be filtered, so that the un-informative word or irrelevant word will be removed. It is aimed to reduce the dimmension of word that will be processed further. The filtering is done by eliminating the symbols continued with stopword removal. After getting the clean word, then the weighted term is calculated by using TF-IDF.

Then the wighted terms will be used in the learning process and classifying process of NBC. The data that have been acquired from Twitter is devided into two, they are training data and testing data. NBC will learn from training data and based on the resulting probabilistic model, NBC will classify the testing data into two class (positive and negative). Detailed NBC method has been explained at section 2. The results of the classification of tweet data then will be visualized in a map. Mapping is adjusted according to the location of the author's tweet. This location is from several regions in Indonesia, especially in Java. From this visualization, the users can easily view in which region the positive/negative sentiments come. And further, they can use this information to make some adjusment in their campaign.

2.4. Technology Stack

This research applies some technologies to work together towards analyzing the sentiment data. Tweets are crawled from Twitter through Twitter API. And some comments of online news website connected to Facebook are crawled by using Facebook API to get the location of user. Both are providing method to grab the location data. The crawled data then formatted as .csv. Python, with its massive library for data science analytic likes numphy, pandas, and sklearn, takes an important place on manipulating and analyzing data. Before analysis, the data needs to be preprocessed by removing stopword and irrelevant symbol by using sastrawi library in Python. After data analysis, the number of positive and negative sentiment in each region will be resulted. And finally, the result is visualized as pie chart on a map using PowerBI. PowerBI applies map that is provided by Bing using geospatial data from tanahair.indonesia.go.id.

3. Results and Discussions

Location-based geo-sentiment analysis on public opinion data about the candidate of regional heads consists of several stages. These start from the data crawling/acquisition, classification and visualization in a map.

3.1. Data

This research uses 398 tweets data from Twitter. It was crawled from Twitter and comments of online news websites (kompas.com and detik.com) that are connected to Facebook account form January 2018 to July 2018. Some of keywords used in the crawling were *pilkada, pilkada jatim, gubernur jatim, wagub jatim, jawa timur*. The whole crawled data were partially used for training data and for testing data. For example, 60% or 239 tweets were used as training data, and the remaining 40% or 159 tweets data were used as testing data. Table 1 shows some tweets taken from several different locations/regions. It is containing 5 data from several different locations which have the same topic being discussed, it is about the candidate of regional heads.

Table 1: Example of the crawling data

No.	Location	Text
1	Batu	Semakin hari Gus Ipul - Puti semakin manyak menuai dukungan. Bahkan ketum Partai Gerindra Prabowo Subianto dijad-walkan ikut mengisi kampanye pasangan tersebut. Dengan begitu bisa dikatakan kemenangan sudah didepan mata. Positif Thinking #LawanPolitikHitam
2	Jember	Ibu Khofifah lebih hebat dari calon yg lain dan tidak pernah dipernah dipecat Gusdur
3	Kediri	Biasa aja kelezz. Mau titisan siapa aja BOLEH JADI PEMIMPIN ASAL JAN-GAN MALING KORUPSI. KESIAN RAKYAT DI PELOSOK.....KERJA KERJA KERJA. JANGAN NGERJAiN RAKYAT KORUPSI APBN....
4	Malang	Meskipun terkesan mendadak, tapi mbak Puti sudah membuktikan bahwa beliau layak untuk menjadi pemimpin Jawa Ti-mur. Karena semakin hari semakin banyak yang memberikan dukungannya untuk Gus Ipul dan mbak Puti.
5	Surabaya	Gus Ipul memang sdh ckp jadi wakilnya pak De. Dan wktnya sdh habis, hrs ming-gir

Each data that has been collected from Twitter, then it would be given the label or annotation (positive or negative) manually. Comments that have normal annotations are not included in this datashet. The label shows the orientation of the opinion in a tweet. Table 2 shows some examples of tweets that is already annotated or labelled. Tweet number 1,2, and 4 have positive annotations because they show a good opinion and they contain of many words with positive meaning. While the tweet number 3 and 5 all contain negative annotations. Tweet number 3 has the negative connotation since it contains a negative word *korupsi*. And tweet number 4 also has a negative connotation since it contains a negative word *minggir*. That negative words have a dominant meaning in the sentences, so that they make the sentences have a negative orientation.

Table 2: Examples of Annotation / Data Training

No.	Location	Text	Label
1	Batu	Semakin hari Gus Ipul - Puti semakin manyak menuai dukungan. Bahkan ketum Partai Gerindra Prabowo Subianto dijadwalkan ikut mengisi kampanye pasangan tersebut. Dengan begitu bisa dikatakan kemenangan sudah didepan mata. Positif Thinking #LawanPolitikHitam	Positif
2	Jember	Ibu Khofifah lebih hebat dari calon yg lain dan tidak pernah dipernah dipecat Gusdur	Positif
3	Kediri	Biasa aja kelezz. Mau titisan siapa aja BOLEH JADI PEMIMPIN ASAL JANGAN MALING KORUPSI. KESIAN RAKYAT DI PELOSOK.....KERJA KERJA KERJA. JAN-GAN NGERJAiN RAKYAT KORUPSI APBN....	Negatif
4	Malang	Meskipun terkesan men-dadak, tapi mbak Puti sudah membuktikan bahwa beliau layak untuk menjadi pemimpin Jawa Timur. Karena semakin hari semakin banyak yang mem-berikan dukungannya untuk Gus Ipul dan mbak Puti.	Positif
5	Surabaya	Gus Ipul memang sdh ckp jadi wakilnya pak De. Dan wktnya sdh habis, hrs minggir	Negatif

Table 3 contains examples of testing data. They do not have any lebel yet. The label will be given automatically by the system after processing NBC. Label that has been resulted by NBC then will be compared with the label that is given manually by human perception (ground truth). Human will know the orientation or label of each tweet will be positive or negative, by analyzing words that construct the sentence and correlation with other wiords. The comparation will bring the accuracy value that shows how accurate the classification result by NBC in this case.

Table 3: Examples of Data Testing

No.	Location	Text	Label
1	Mojokerto	Ha ha ha,setalah kemarin koar2 ingin menjaga NU agar tdk pecah akhirnya cari jwban lain krn jwban kemarin tdk sesuai keadaan yg sebenarnya sbb meski mbak Yenny gak maju NU sdh pecah dr dulu,mulai dr dibentuknya PKB sbg pecahan NU dibidang politik,trus trjadinya penolakan ula-ma2 NU atas petusan Fatwa kemarin soal pe-nistaan AGAMA yg mana ketua MUI sekaligus Rais Aam NU atau ketuam Umum NU berseberangan dgn ketua Harian yakni Said Aqil dan ulama2 NU lain mengindikasikan NU sdh pecah,penolakan Ustad Khalid basamalah yg merupakan cucu ketua NU sulawesi dan persekusi dan penolakan ustad Abdul Somad yg juga Anggota NU di pekanbaru Riau menunjukan perpecahan ditubuh NU sdh tdk dpt ditutupi lagi.	?
2	Pacitan	Mbak yenny sdh bertindak sangat bijak. Beliau orang yang mengerti menghargai persaudaraan dibandingkan kekuasaan. Beliau samgat memahami ibu khofifah adalah salah satu murid Gus Dur . Salut utk mbak Yenny.	?
3	Sumenep	pilih b.khofifah amanah dn sudah terbukti	?
4	Tulungagung	Meskipun partai pen-gusung utama pasangan bacalon Gubernur dan cawagub Jawa Timur, Saifullah Yusuf (Gus Ipul) dan Puti Guntur Soekarno, adalah PDI-P dan Partai Kebangkitan Bangsa (PKB). pasti nya akan banyak partai lain yang ikut mendukung keme-nangan Gus Ipul-Mba Puti	?

3.2 Naive Bayes Classification

Classification with NBC bring good results. It could be shown at table 4. The testing was conducted in several times with different composition of traning data and testing data. First testing is done by taking 80% of data as a training data and 20% as a testing data. Up to the number 5 trial conducted by taking 40% of the total data as a training data and the remaining 60% as a testing data. The highest results can be seen from the trial No. 3 which produces an accuracy of 78.3% and the lowest trial number 4 which produces a percentage of 64%.

The best accuracy was found when we employed the 80% training data and 20% testing data. This describes that the more training data, the better accuracy value will be. The more training data means providing more dataset where NBC will learn from. It will increase the possibility of classifier to be able to classify exactly as expected. From table 4, the accuracy value decreases as the number of training data decreases.

Table 4: Accuracy results of Naive Bayes

No.	Training Data Percentage (%)	Testing Data Percentage (%)	Accuracy(%)
1	80	20	78,3%
2	70	30	75,2%
3	60	40	66,2%
4	50	50	64,0%
5	40	60	64,2%

3.3 Data Visualization

After the classification process using the NBC, all data now have their own label. Then we count the number of tweets that have positive or negative label in each region. It will be visualized by using a pie chart on a map. The visualization will show trend of tweet or opinion in each region for a candidate of governor.

Fig. 2: Visualization of geo-sentiment of overall opinions

The overall sentiment of the governor candidates can be seen at Figure 2. It shows us the accumulation of positive and negative opinions of all candidates in each region. Instead of showing only opinions of one candidate, this visualization mode will show overall opinion of people about the governor election without considering who the candidate is. The pie chart with two color is using to show the number of positive and negative opinions. Pink color means negative opinions and blue is a positive opinions.

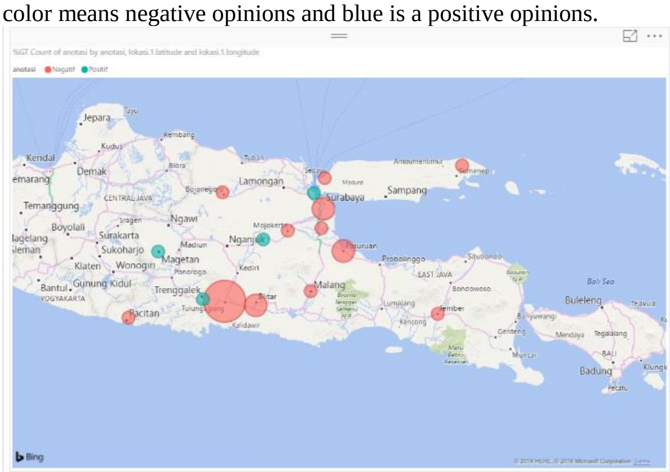

Fig. 3: Visualization of geo-sentiment of Khofifah-Emil Dardak opinions

Figure 3 shows the results of negative-positive geo sentiment for a governor candidate Khofifah and Emil Dardak as the deputy candidate. The size of the pie is in accordance with the population in the region. The more population in a region, the wider pie chart will be. From the pie chart provided at Figure 3, this candidate has negative opinions from Tulungagung, Pacitan, Blitar, Jember, Malang, Pasuruan and so on. While the other regions that give positive opinion for this candidate are Tulungagung, Magetan, Nganjuk and Surabaya. This could be a consideration for the candidate to make both strengthening opinion and to conduct a campaign intensively on those regions that has negative opinions dominantly.

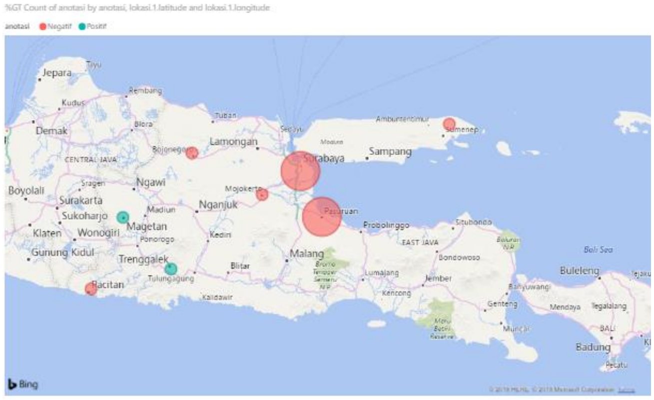

Fig. 4: Visualization of geo-sentiment of Khofifah opinions

In some cases, the trend of opinion of a governor candidate need to be analyzed itself, without including the deputy candidate. It is used to know the significance of opinion of governor candidate. Figure 4 shows this mode of visualization. Figure 4 describes the opinion mapping of Khifufah. Most of regions where there are opinions tell about her, gives negative opinions. It is at Surabaya, Pasuruan, Mojokerto, Bojonegoro, Sumenep and Pacitan. And other regions like Trenggalek and Magetan give positive opinions. The mapping of opinions of Emil Dardak is shown at Figure 5. Based on the crawled data, as the deputy candidate of Khofifah, Emil gets good opinion from Nganjuk and Surabaya. Even the number of this positive opinion is not significant, it may bring a strengthening and good effect for this candidates.

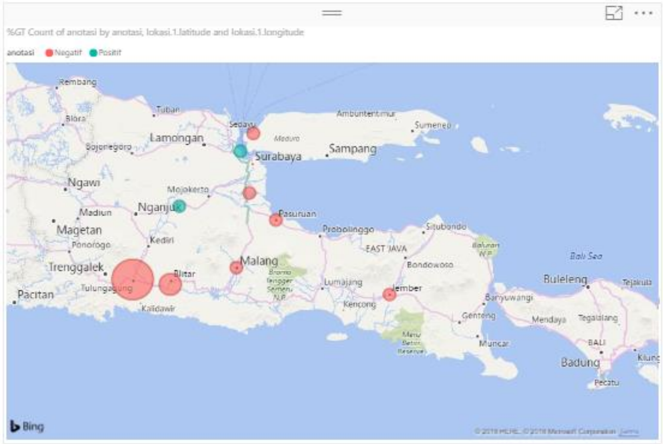

Fig. 5: Visualization of geo-sentiment of Emil opinions

The visualization of positive and negative geo sentiments for candidates Syaifulah-Puti, can be seen in Figure 6. Opinions for this candidate come from many regions compared with previous candidate (Khofifah-Emil). Based on Figure 2 there are 16 regions that gives opinion for Khofifah-Emil. While the opinions for Syaifullah–Puti come from 21 regions. It could be seen from Figure 6.

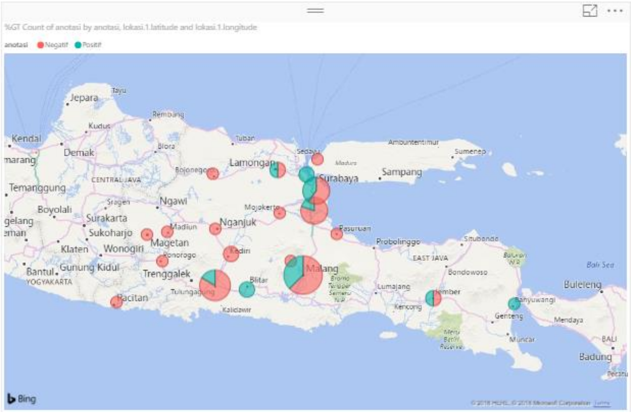

Fig. 6: Visualization of geo-sentiment of Syaifullah–Puti opinions

Figure 7 shows the opinions that come form Syaifullah, who is governor candidate. There are 15 regions that give opinion for him. 7 regions have completely negative, while the remaining regions give both positive and negative.

Fig. 7: Visualization of geo-sentiment of Syaifullah opinions

Syaifullah has a significance opinion for this candidate. While his partner, Puti does not. It could be seen from Figure 7. From 6 regions that give her opinions, 4 regions have positive opinions and the remains are negative opinions. It means that strengthening opinion for Puti is needed, especially at Malang, Tulungagung and Bojonegoro.

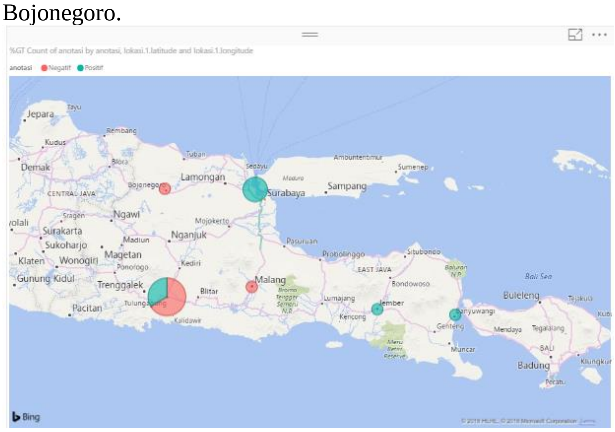

Fig. 8: Visualization of geo-sentiment of Puti opinions

Distribution of positive and negative opinion about the governor election in each region. Malang and Surabaya are the two regions with the biggest amount of opinion. It shows that people in these area is active to post their opinion through Twitter. So that, the campaign team of two candidates could optimize their campaign posting in Twitter, to reach people in these regions. Even there are many opinions come from these two regions, majority of the opinions is negative. It is important for campaign team to further analyze in what object the opinions tell about. Figure 9 depicts the distribution of opinion in each region.

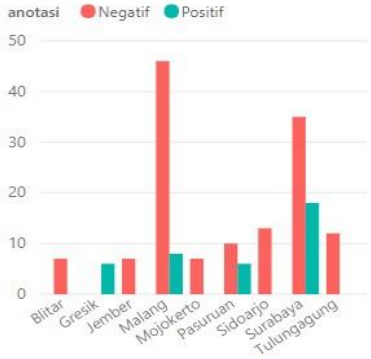

Fig. 9: Opinion distributian in each region

4. Conclusions and Suggestions

The system developed in this research, has successfully classified the textual opinion data crawled from social media. The classified opinions are then grouped by region where it came and finally displayed as pie chart on a map. NBC as the classifier could perform well on this research, as the resulted accuracy was 78,3%. The accuracy became better when there are more training data applied. It is suggested to develop further research with more varied data sources, so that it will supply more comprehensive data from each region. The more data crawled, then the chance to provide much more training data will be better as well.

Acknowledgement

The authors would like to thank to State Polytechnic of Malang who has supported this research.

References

[1] CNN Indonesia. 2016. "Berita Teknologi Informasi : Twitter Rahasiakan Jumlah Pengguna di Indonesia". Diakses tanggal: 23 Maret 2018, Tersedia di: http://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitterrahasiakan-jumlah-pengguna-diindonesia
[2] Atmodjo. Juwono Tri, "Dinamika Partisipasi Politik Remaja melalui Media Sosial," *Jurnal Visi Komunikasi*, vol. 13, no. 04, November 2014.
[3] Bode. Leticia, Dalrymple. Kajsa E, "Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter," *Journal of Political Marketing*, vol. 15, issue 4, 2016.
[4] Abdillah. Leon Adretti, "Social Media as Political Party Campaign in Indonesia," *Jurnal ilmiah Matrik*, vol.16, no.1, April 2014.
[5] Howard. Phipil N, "Deep Democracy, Thin Citizenship: The Impact of Digital Media in Political Campaign Strategy," *The ANNALS of the American Academy of Political and Social Science Journal*, vol.597, issue 1, 2005.
[6] Sarlan. Aliza, Nadam. Chayanit, Basri. Shuib, "Twitter Sentiment Analysis," *International Conference on Information Technology and Multimedia (ICIMU)*, Outrajaya, Malaysia, 2014.
[7] P. Lai, "ExtractingStrongSentimentTrendfromTwitter". *Stanford University*, 2012.
[8] Sadida. Rizqon, et.al., "Perancangan Sistem Analisis Sentimen Masyarakat Pada Sosial Media Dan Portal Berita," *Seminar Nasional Teknologi Informasi dan Multimedia 2017*, 2017.
[9] Severyn. Aliaksei, Moschitti. Alessandro, "Twitter Sentiment Analysis with Deep Convolutional Neural Networks". *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015.
[10] Ghiassi. Manoochehr, Zimbra. David, "Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks," *Journal of Management Information Systems*, vol.33, issue 4, 2016.

[11] Perdana. R. S., Pinandito. A., "Combining Likes-Retweet Analysis and Naive Bayes Classifier Within Twitter for Sentiment Analysis," *International Conference On Communication And Computer Engineering (ICOCOE),* Penang, Malaysia: Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2017.

[12] Rodiyansyah. F. S., Winarko. E., "Klasifi-kasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayes," *Indonesian Journal of Computing and Cybernetics Systems (IJCCS),* 2013.

[13] Hidayatullah. A., F., Azhari, "Analisis Sentimen Dan Klasifikasi Katrgori Terhadap Tokoh Publik Pada Twitter," *Seminar Nasional Informatika,* Yogyakarta: UPN Veteran, 2014.

[14] Rozi. I. F., Pramono. S. H., Dahlan. E. A., "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *Jurnal EECCIS*, vol.6, no.1, 2012.

[15] Chin. D., Zappone. A., Zhao. J., "Analyzing Twitter Sentiment of the 2016 Presidential Candidates," 2016.

[16] Jayanthi, Sreeja, "Sentiment Analysis of Twitter Data containing Emoticons: A Survey," *International Journal of Innovations in Engineering and Technology (IJIET)*, VI(4), 147- 153, 2016.

[17] Dayalani. G. G., Patil. P. B., "Emoticon-based unsu-pervised sentiment classifier for polarity analysis in tweets," *International Journal of Engineering Research and General Sci-ence,* II(6), 438-445, 2014.

[18] Bansal, B., Srivastava. S., "On predicting elections with hybrid topic based sentiment analysis," *Procedia Computer Science 135, 346–353.* 2018.

[19] Kuše. E., Strembeck. M., "Politics, sentiments, and misinformation: An analysis of the Twitter," *Online Social Networks and Media 5, 37-50.* 2018

[20] Tellez. E. S., Miranda-Jim´enez. S., Graff. M., Moctezuma. D., Siordia. O. S., Villasenor. E. A., "A Case Study of Spanish Text Transformations for Twitter Sentiment Analysis," *Expert Systems With Applications,* doi: 10.1016/j.eswa.2017.03.071, 2017.

[21] Pang. Bo, Lee. Lillian, "Opinion Mining and Sentiment Analysis," *Now Publisher*, 2008.

[22] Nopyandri, "Pemilihan Kepala Daerah secara Langsung dalam Perspektif UUD 1945," *Inovatif Jurnal Ilmu Hukum*, vol. 6, no. 7, 2013.