PAPER • OPEN ACCESS

Opinion Analysis on Rohingya using Twitter Data

View the article online for updates and enhancements.

Related content

- Join the *Twitter* bandwagon

- Naive Bayes as opinion classifier to evaluate students satisfaction based on student sentiment in Twitter Social Media
  Fahmi Candra Permana, Yusep Rosmansyah and Atje Setiawan Abdullah

- Pre-processing Tasks in Indonesian Twitter Messages
  A F Hidayatullah and M R Ma'arif

Opinion Analysis on Rohingya using Twitter Data

N Rochmawati*, S C Wibawa

Department of Informatics, Universitas Negeri Surabaya, Indonesia

*naim.rochmawati@gmail.com

Abstract. Rohingya is an ethnicity in Myanmar. Recently there was a conflict in the area between the Rakhine population and the Myanmar army. Many opinions are pro and contra in addressing this issue. There is a critic, there is a support and there is a neutral. The purpose of this paper is to analyze the world public opinion about the case of Rohingya. The opinion data to be processed is taken from twitter. the reason for using twitter is because twitter has become one of the popular social media and includes the most frequently visited social media. Therefore, it would be a lot of data that can be taken from twitter to be processed in the process of sentiment analysis. The grouping of opinions will be divided into 3 parts of positive, negative and neutral. the method used in grouping is the naïve Bayes method.

1. Introduction
Rohingya is an ethnic Indo-Aryan from the Rakhine region (also known as Arakan, or Rohang in Rohingya) in Myanmar. Recently there was violence in Rohingya between the Rakhine population and the Myanmar army. Many opinions are pro and contra in addressing this issue. Some are criticized and some are supportive and some are neutral.

Twitter is one of the social media that allows users to send and read text-based messages. this message is called tweet. tweet is limited to 140 characters. twitter was first launched in 2006 (Daniel, Twitter, Edward, Twitter, & Earth, n.d.). people use microblogging like twitter to talk about their daily activities and share information(Finin & Tseng, 2007) .

The high popularity of Twitter can be used for various purposes in various aspects, for example for political campaign, learning tool, and as an emergency communication media. Twitter is also faced with various issues and controversies such as user security and privacy issues, lawsuits, and censorship. Twitter can also be used to know the opinions of the world community on news about Rohingya.

The purpose of this paper is to know the opinion analysis of people about Rohingya where the opinion will be divided into 3 groups: positive, negative and neutral. An opinion analysis will be taken from twitter. Grouping data will use naive Bayes.

2. .Literature

2.1. State of the art
Sentiment analysis is widely developed in various fields such as to analyze hotel rating (Elango & Narayanan, 2002; Hargreaves, 2015; Kasper & Vela, nd), politics (Spoelstra, nd), stock (Joshi, nd), health (Database, 2016) and so on.

2.2. Sentiment analysis
Sentiment analysis is one of the fastest growing research areas in computer science (Mäntylä, Graziotin, & Kuutila, 1937). Hence it is not surprising that the sentiment analysis is much in demand as a research material.

Sentiment analysis is a way to retrieve information from opinions or language. Therefore information sentiment analysis is mostly taken from social media because opinions and user information can be extracted from users tweets. Sentiment analysis is divided into two levels: document level (coarse grained sentiment analysis) and fined grained sentiment analysis (Fink, Chou, Kopecky, & Llorens, 2011).

2.3. Text mining
Text mining is the application of data mining concepts and techniques to find patterns in the text. the process of analyzing texts is intended to extract useful information for a particular purpose. Based on the irregularities of text data structures, the text mining process requires several initial steps which in essence is preparing for the text to be changed to be more structured. Text mining has many methods and techniques (Gaikwad, 2014).

The purpose of text mining is to get useful information from a set of documents. Thus, the data source used in text mining is a collection of text that has a unstructured or at least semi-structured format The specific tasks of text mining include categorization of text (text categorization) and text clustering text groupings).

2.4. Naïve bayes
Bayes's theorem is the theorem used in statistics to calculate the odds for a hypothesis. Bayes Optimal Classifier calculates the chances of a class from each of the existing attribute groups, and determines which classes are the most optimal.

Naïve Bayes Classifier is a classification method rooted in Bayes's theorem. The method of classification using the probability and statistical methods proposed by the British scientist Thomas Bayes, predicting future opportunities based on past experience, is known as Bayes's Theorem. The main feature of the Naïve Bayes Classifier is a very strong (naïve) assumption of the independence of each condition or event.

3. Methodology
The steps used to solve the problem are:

3.1. Collecting data
The first step is to collect data from twitter. Because that will be analyzed is the opinion of Rohing so that the data taken using the keyword "Rohingya". The data will be used as a sample of 3500 samples of data.

3.2. Cleaning data
The second step is cleaning the data (Paper, 2013). Before the data is analyzed, the data must be cleaned first because if not cleaned it will disturb the data analysis. What will be removed here are: retweet, @ people, punctuation, numbers, html links, unnecessary spaces, unnecessary symbols. In addition, the removal of punctuation and symbols other than alphabet and change it from uppercase to lowercase.

3.3. Perform sentiment analysis
In this stage the cleared data will be classified using the naïve Bayes algorithm. To classify emotions and positive or negative, use library sentiment.

4. Result
The data in this paper is processed using R in order to do wordcloud. In R there is a facility to create a word cloud. Word cloud (also called text cloud or tag cloud) is one method to display text data visually. These graphs are popular in text mining because they are easy to understand. By using the word cloud, the picture frequency of words can be displayed in an interesting yet informative form. The more often one word is used, the larger the word size is displayed in the word cloud.

The results of the testing that has been done is as follows:

1. Crawling data. Data
    Data taken from twitter as many as 3500 with keyword Rohingya. The result is shown in figure 1:

```
[[1]]
[1] "klinecj: Rohingya - the most persecuted people in the world - Thai PBS English News
https://t.co/jnJNkPu3zB via @twitterapi"

[[2]]
[1] "shuriemfraha: RT @cjwerleman: Thousands of Rohingya Muslims crosssed this river, fl
eeing Burma's death squads today. https://t.co/W1Z9TEeMVd"

[[3]]
[1] "JFCrisp: Very confusing for the public. Homepages of IOM, UNHCR and UNICEF all feat
ure separate appeals for Rohingya refugee… https://t.co/Lx08IRbAsk"

[[4]]
[1] "PanickerUma: RT @TrueIndology: Even Islamic Republic of Bangladesh arrests Rohingya
terrorists posing as refugees, but @ndtv finds \"no evidence of any R…"
```

Figure 1. Crawling data

2. Cleaning Data
    One example of the results when cleaned re-tweet then the results obtained can be seen in Figure2:

```
[1] "Rohingya - the most persecuted people in the world - Thai PBS English News https://t.co/jnJNkPu3zB "
```

Figure 2. Cleaning data

There is a difference between the data in Figure 1 before cleaned retweetnya and image 2 after cleaning retweets. Looks via and @ is gone. The purpose of cleaning this data is to remove non-essential data which can interfere with the process of opinion analysis.

3. Sentiment analysis
    After the process of analysis, the results obtained in Figure 3

```
       emotion  polarity
1      sadness  positive
2      unknown  negative
3      unknown  negative
4      unknown  positive
5      sadness  negative
6      unknown  positive
7      unknown   neutral
8      unknown   neutral
9      unknown  positive
10     unknown  positive
11     unknown  negative
12     unknown  positive
13     sadness  negative
14     unknown  positive
```
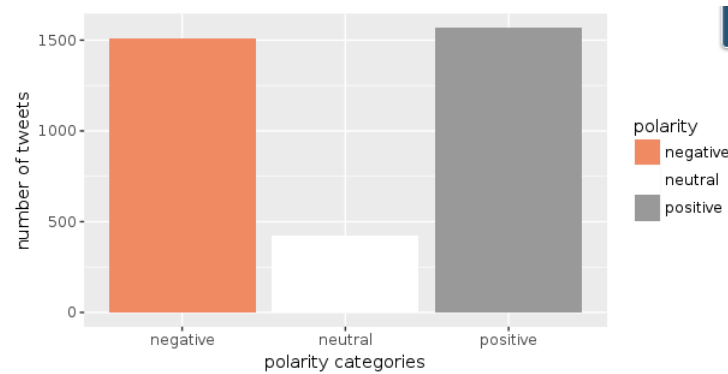
Figure 3. Result of sentiment analysis

The result in Fig. 3 if formed in the plot diagram will be like figure 4
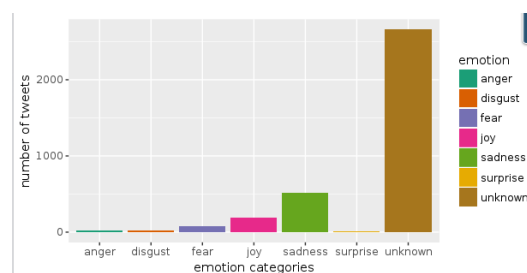
Figure 4. Diagram result of grouping emotion

Unknown results are from groupings that are not included in the group anger, disgust, fear, joy, sadness and surprise. It turns out that the largest group is unknown. The second highest group is sadness. It means that the people of the world turn out to be a lot of sympathizers and feel sorrow for what is experienced by the inhabitants of Rohingya. It is also interesting that the third highest group in the diagram shows some people who actually feel joy with the existence of Rohingya case.

The result of negative, positive and neutral opinion grouping can be seen in figure 5.

Figure 5. Diagram result of sentient analysis

From the result of figure 5, it can be concluded that between the positive and negative group has the same height. While the group with the lowest number is the neutral group.

The result of plotting wordcloud from this text analysis can be seen at figure 6.
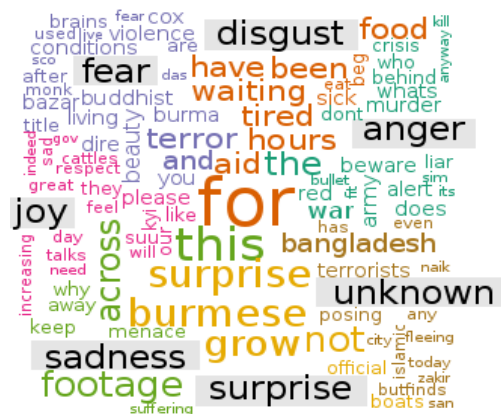
Figure 6. Word cloud

5. Conclusion

Based on the results of data analysis above can be concluded that between the positive vs negative groups balanced. While the neutral groups are fewer. The most number of tweets is unknown. This is due to the difficulty of processing text data so that many text that does not belong to the other 6 groups of emotions. This may be due to the incompleteness of the dictionary or it may be due to something else.

References

[1]   Daniel, B., Twitter, T., Edward, W., Twitter, A., & Earth, L. (n.d.). 7 Things You Should Know About Twitter.

[2]   Database, R. 2016. An effective model for store and retrieve big health data in cloud computing, *132*, 75–82. http://doi.org/10.1016/j.cmpb.2016.04.016

[3]   Elango, V., & Narayanan, G. 2002. Sentiment Analysis for Hotel Reviews.

[4]   Finin, T., & Tseng, B. 2007. Why We Twitter : Understanding Microblogging.

[5]   Fink, C. R., Chou, D. S., Kopecky, J. J., & Llorens, A. J. (2011). Coarse- and Fine-Grained Sentiment Analysis of Social Media Text, *30*(1), 22–30.

[6]   Gaikwad, S. V. 2014. Text Mining Methods and Techniques, *85*(17), 42–45.

[7]   Hargreaves, C. A. 2015. Analysis of Hotel Guest Satisfaction Ratings and Reviews : An Application in Singapore, *1*(4), 208–214.

[8]   Joshi, K. (n.d.). S TOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS.

[9]   Kasper, W., & Vela, M. (n.d.). Sentiment Analysis for Hotel Reviews, *231527*, 45–52.

[10] Mäntylä, M. V, Graziotin, D., & Kuutila, M. 1937. The Evolution of Sentiment Analysis - A Review of Research Topics , Venues , and Top Cited Papers.

[11] Paper, D. 2013. An introduction to data cleaning with R.

[12] Spoelstra,   J.   (n.d.).   Predicting   US   Primary   Elections   with   Twitter,   1–8.