2017 International Conference on Sustainable Information Engineering and Technology (SIET)

Implementation Of Naive Bayes Classifier Algorithm On Social Media (Twitter) To The Teaching Of Indonesian Hate Speech

Naufal Riza Fatahillah
Program Studi Sistem Informasi
STMIK AKAKOM
Yogyakarta, Indonesia
rayzalzero@gmail.com

Pulut Suryati
Program Studi Sistem Informasi
STMIK AKAKOM
Yogyakarta, Indonesia
lut_surya@akakom.ac.id

Cosmas Haryawan
Program Studi Sistem Informasi
STMIK AKAKOM
Yogyakarta, Indonesia
cosmas@akakom.ac.id

***Abstract* - Twitter is a social media that is widely used as a sharing medium on the internet. There are tweets containing sentences shared by the user, thus they can be read by the other users. A lot of information can be obtained from Twitter. Twitter users can connect with the other Twitter users in an international scale.Technology that is growing as today can be used for various things, especially regarding the information distributed in social media, specifically Twitter. One of the problems derived from social media is that Twitter tweets containing speechesin the form of both positive and negative utterances. From the problems above, a research is required to classify tweets that contain positive and negative speeches orutterances using naive bayes classifier method. The results of this study are implemented into a system that can classify tweets on Twitter. The system is built using js Node technology and Naive Bayes classifier as the calculation method of classification. Based on the tests performed, the best accuracy generated by the systems using the Naive Bayes Classifier is 93%.**

Keywords: classify, Naive Bayes Classifier, Twitter, speech

I. INTRODUCTION

One of the social media used to communicate among users is Twitter. It is estimated that the number of Twitter users registered in 2016 has reached 317 million users[1]. The number of users which continues to increase makes Twitter to be a medium of information that is widely used by Internet users to share information. Twitter itself does not have filtering system regarding whether the written tweets are positive or negative. A tweet can contain a negative utterance called hate speech. Hate speech itself can be in the form of derogatory, defamatory, libelous, unpleasant acts or utterances, provocation, sedition and the spreading of false news. It all can be called as hate speech. It can lead to non-discrimination, violence, homicide, and social conflicts. Due to the above case, the researchers propose to create a system that can filter out negative or positive tweets so that users can distinguish between the two. The system will thus automatically distinguish positive tweets and negative tweets. The list of positive and negative utterances was quoted from Liu et al[2], Tala[3], wahid and azhari[4].

There are several methods within a classification method. In this case, the development of classification applications is using naive bayes classifier method. Tweets to be used are in Indonesian. The system classifies based on words instead of semantics.

A similar study was conducted by Kurniawan [5] analyzing twitter social networking data for street congestion conditions mapping in Yogyakarta province by the method of text mining. This condition monitoring system classifies tweet contents using several machine learning methods including Naive Bayes, Support Vector Machine, and Decision Tree. Rakhman [6] classifies Tweet Spam and Valid using Chi-Square Feature Selection and naive bayes classifier algorithm on tweets in Bahasa Indonesia. Based on the tests performed, the best accuracy generated by the systems using Naive bayes classifier Multinomial model combined with Chi-Square feature selection is 95%.

The naive bayes classifier algorithm is also used for SMS Spams Filtering in Bahasa Indonesia[7], on the application of multi-agent systems for the extraction of information on blood requirements on twitter [5] and SMS-based community complaint classification [8].

A. Twitter

Twitter is a social media that can connect all the people in the world and allows users to communicate with each other in a 140-character short message called tweet.Twitter itself allows users to follow the latest news regarding the topics that they are interested in. According to the official website, dev.twiiter.com Twitter platform provides access to twitter data through the API. Each API presents several aspects of Twitter and allows developers to build and expand their applications in a new way based on their own creativity. Twitter developer continues to expand API so that the Twitter API can undergo changes. [9]

B. Naive Bayes Classifier

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Studies comparing classification algorithms have found a simple Bayesian classifier known as the *Naive Bayesian classifier* to

be comparable in performance with decision tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. *Naive* Bayesian classifiers assume that the effect of an attribute of value on a given class is independent of the values of the other attributes [10][11].

Bayes' theorem is named after Thomas Bayes, a nonconformist English clergyman who did early work in probability and decision theory during the 18th century.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Let ***X*** be data tuple. In Bayesian terms, ***X*** is considered "evidence." As usual, it is described by measurements made on a set of *n* attributes. Let *H* be some hypothesis such as that the data tuple ***X*** belongs to a specified class *C*. For classification problems, we want to determine *P(H|X)*, the probability that the hypothesis *H* holds given the "evidence" or observed data tuple ***X***. In other words, we are looking for the probability that tuple ***X*** belongs to class *C*, given that we know the attribute description of ***X***.

II. METHODS

A. Analysis and System Design

The data or material required for this research is tweet data from Twitter or can be called as tweet generated from Twitter social media. Tweet data is a sentence with a length of 140 characters on each tweet. The tweet is pinned by hashtag search. The data collection technique is by means of Twitter API provided by Twitter.

The collected data is the official primary data obtained directly from Twitter using Twitter API. The data training collection is completed randomly using Twitter API based on the hashtag, whereas for classification data as in collecting data from the Twitter API is based on the hashtag with a time limit of 1 week.

The functional requirements analysis describes the service capabilities of a system. It consists of data requirements which include inputs, processes, and outputs. The explanation is as follows:

1. The input of this application is the hashtag search or keyword. From the keyword, it will generate data to be handled on the classification process.
2. The process of this application is a training process which is a stage of calculation and calcification process of tweet data
3. The output result of this application is in the form of a graph containing negative, positive and neutral tweet percentages.

B. Preprocessing

Preprocessing is implemented to avoid incomplete data, data interruptions, and inconsistent data[13]. The stages of preprocessing text in this research include:

1. Removing URL (http://www.situs.com) and email (nama@situs.com). They are deleted at this stage.
2. Replacing emoticons contained in the tweets with words which reflect the emoticons.
3. Deleting Special Twitter Characters This process is done by deleting special Twitter characters such as hashtag, username (@username), and special characters (eg RT, which indicates that the user retweet something).
4. Removing Symbols. This step is done to remove symbols and punctuations in the tweet.
5. Removing Stopwords. Stopwords are words that do not affect the process of classification

C. Training Proses

The training process starts from reading the data from the database then stopword process is conducted. Next, declare i=0 followed by iteration if variable i is smaller than the length of data to determine how much positive data is stored in docsByClass, and all of the words will be stored on the sumAllDocs variable. The checkwordIndoc process is a process of grouping positive words and negative words. The checkwordInDoc procedure is the process of checking whether the sentence on data training is positive or negative.

The tweet's classification probability calculation process is using naive bayes classifier method. The classification process calculates three prior calculation processes, likelihood, and probability. Prior calculation process is the calculation of occurrence level of a label on the previous training data. Likelihood calculation process is the calculation of the possibility of a word appears on a particular label. The process of probability calculation is the calculation measuring the occurrence of a label on a sentence.

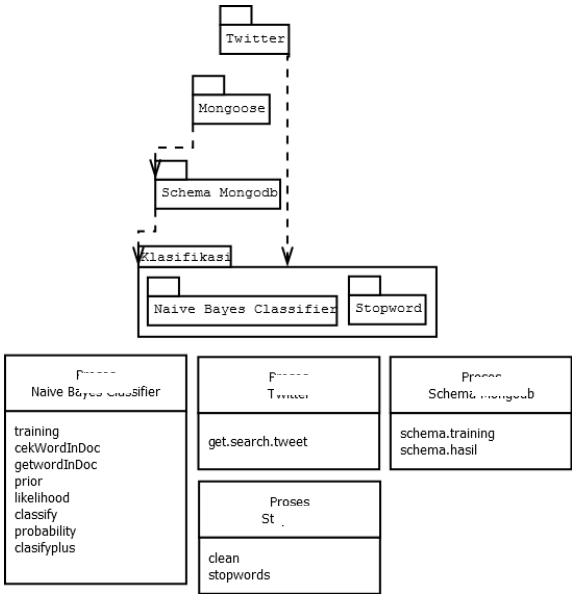

Fig. 1. Package Diagram

The classification process requires some additional packages to be able to classify data grouped into a package. Shown in Fig. 1. is a package diagram of the classification

process. Twitter Package is used to search tweet data. Package mongodb functions as a database for storing training data and classification data. Package classifier is used to classify the tweets.

III. Result and Discussion

The system architecture built consists of three components: client, web server, Twitter API, and mongodb server. On client will access web server using browser with certain URL. There is an app.js in the web server that serves as the parent server, naive bayes classifier module is used to perform classification calculations. On the mongodb side, it uses a schema from mongoose to perform data request. Fig. 2 is the outlook as a user page who will do the classification. There are several classifier results shown on the main page as well as the form to classify.

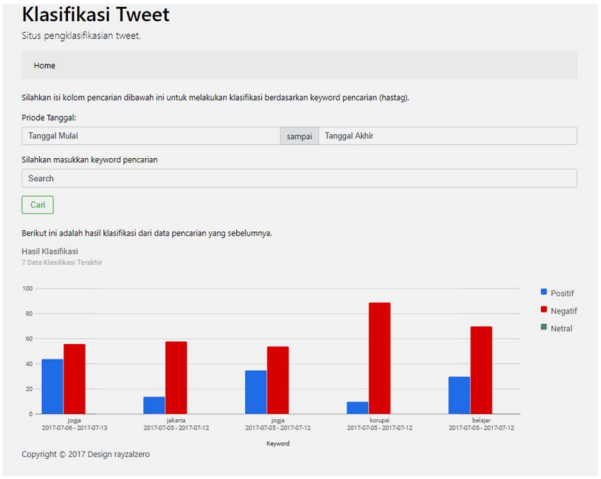

Fig.2 the view as a user page

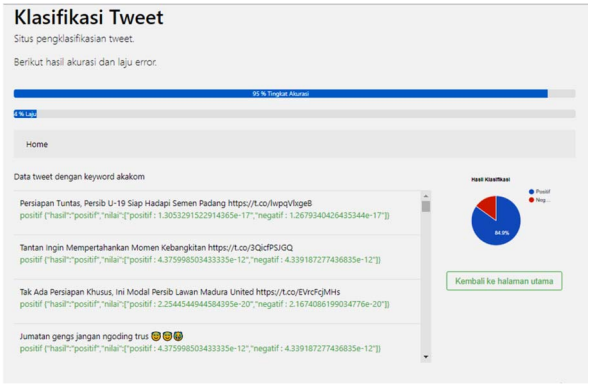

Fig.3. Tweet Search Results and Classification Display

The tweet search and classification process page are completed after the user enters a hashtag and a particular date of tweeting at a specified time of 1 week. Then a tweet search is conducted based on the hashtag. There is some information gained, ie:

1. the system's level of accuracy at the top,
2. the classification results in the form of text and tweet,
3. the result of classification in the form of graphs.

The search results and classification page is shown in Fig.3.

Display of admin pages to collect training data are shown in Fig. 4. There is a search field to search for new training data. The training data is displayed below in the search field.

Fig. 4. Display of admin pages

The data training search page using Twitter data obtained after filling out the search form is shown in Fig. 5. After the data is grouped, the data is then stored in a database.

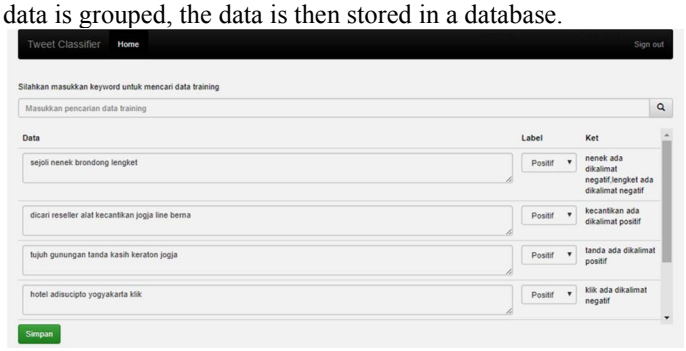

Fig. 5. Admin Data Search Training Page

The classification calculation uses naive bayes classifier method based on the formula (1). Shown in Table 1 is an example of training data that will become a pattern for classification calculations.

TABLE 1 EXAMPLE OF DATA TRAINING

No.	Data	Label
1	Budi bohong kepada guru.	Negatif
2.	Belajar berbuat baik di sekolah.	Positif
3.	Yani anak yang jahat dan sering bohong.	Negatif
4.	Selalu jujur dalam berkata baik dalam perbuatan.	Positif
5.	Mengapa selalu saja bohong.	Negatif
6	Ibu berkata jujur itu adalah baik.	?

The stopword process is completed before the training process to eliminate some meaningless words. The stopword process will generate data as shown in TABLE 2.

TABEL 2. EXAMPLE OF THE STOPWORD PROCESS

No.	Data	Label
1.	Budi bohong guru.	Negatif
2.	Belajar berbuat baik sekolah.	Positif
3.	Yani jahat sering bohong.	Negatif
4.	Selalu jujur berkata baik perbuatan.	Positif
5.	Mengapa selalu bohong.	Negatif
6	Ibu berkata jujur baik.	?

Training data will produce a calculation pattern used in the classification calculation process. From TABLE 2, we can get the prior value and calculation of naive bayes classifier.

$$P(positif) = \frac{2}{5}$$
$$P(negatif) = \frac{3}{5}$$

Calculation of probability value:

P(ibu|positif) = (0+1)/(9+19)=0.35
P(berkata|positif) = (0+1)/(9+19)=0.35
P(jujur|positif) = (1+1)/(9+19)=0.71
P(baik|positif) = (1+1)/(9+19)=0.71
P(ibu|negatif) = (0+1)/(10+19)=0.34
P(berkata|negatif) = (0+1)/(10+19)=0.34
P(jujur|negatif) = (0+1)/(10+19)=0.34
P(baik|negatif) = (0+1)/(10+19)=0.34

The probability that classification**:**

P(positif|d6) = 2/5 * 0.35 * 0.35 * 0.71 * 0.35 ≈ 0.247
P(negatif|d6) = 3/5 * 0.34 * 0.34 * 0.34 * 0.34 ≈ 0.008

Based on the probability value, it can be concluded that sentence number 6 is positive with the value 0.247 higher than 0.008 at the negative label value. A system test is performed on all modules of the program to determine whether the program is made in accordance with its functional needs or not. The process of training data testing is obtained manually from Twitter. The first test process is done without using the preprocessing stage out of 100 training data. 60 data are predicted accurately. Obtained classification testing accuracy using naive bayes classifier reaches 60% of accuracy with an error prediction of 40%.

```
data,label
Kita harus bersyukur hati kita masih ditautkan dengan masjid
dengan menjalankan itikaf hari terakhir,positif
You too Have a bless ramadhan and eid in advance and selamat
membaca,positif
Jika kita menghendaki Ramadhan berkah berarti kita berharap agar
Ramadhan bisa menambah melanggengkan kebaikan bagi kita.,positif
karakter di awal ramadhan pun sudah tiada apatah lagi di akhir .
Betulkan saya kalau salah,negatif
warga Yaman maut dalam serangan udara,negatif
lancar Waqaf Saham Larkin Sentral,positif
Nutrisari Jeruk Peras Ramadhan Kios,negatif
RADIO KESAYANGAN KAMU            ,positif
Dalam kesempatan silaturahmi di bulan Suci ramadhan,positif
Berkah Ramadhan Alasantri Kolaborasi dan Tema MUDIK Selamat
kepada pemenang,positif
Malam ganjil 23 ramadhan,negatif
Sedih itu Ramadhan udah mau habis tapi belum khatam,negatif
```

Fig. 6. Sample result of data testing from twitter

The second testing process is completed to assess the training data using the preprocessing stage as to clean up tweets from words that are not meaningful and prepare tweets for the classification process. The second testing process was done using 200 training data. 186 data are predicted correctly. Fig. 6. The sample result of data testing is from twitter. The second training data test results are better in accuracy than the first one, with 93% of accuracy out of 200 training data.

IV.CONCLUSION

Based on the lengthy process, starting from the planning to the implementation of this study, some conclusions are obtained, among others are:

1. The system can get tweet data based on hashtag tweet. Naive bayes classifier can be used to classify tweets based on labels. Classification testing using naive bayes classifier yields an accuracy of 93%.
2. The determination of data training can affect the test results because the pattern of data training will be used as a rule to determine the label on the classification data. The percentage of accuracy is also influenced by the determination of data training.

Subsequent development in order to obtain better results between the classification processes is completed in real time and the classification of tweets involves the meaning of words or phrases.

ACKNOWLEDGMENT

Through this article, the author would like to thank the friends of lecturers and students at STMIK AKAKOM Yogyakarta for their cooperation and support.

REFERENCES

[1] [1] https://www.statista.com/statistics/282087/number-of-monthly-active-Twitter-users/

[2] [2] Liu, Bing, Hu, Minqing, and Cheng, Junsheng,"Opinion Observer: Analyzing and Comparing Opinions on the Web.", Proceedings of the 14th International World Wide Web Conference (WWW-2005), May 10-14, Chiba, Japan, 2005.

[3] [3] F.Z. Tala, A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. M.S. thesis. M.Sc., Thesis, Master of Logic Project. Institute for Logic, Language and Computation, Universiteti van Amsterdam The Netherlands, 2003.

[4] [4] D.H.Wahid, S. N Azhari, Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity, IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 2016, pp. 207-218.

[5] [5] D. A. Kurniawan, Analisis Data Jejaring Sosial Twitter untuk Pemetaan Kondisi Kemacetan Jalan di Provinsi DIY dengan Metode Text Mining, Skripsi ,Universitas Gajah Mada Yogyakarta, 2016.

[6] [6] D. I. Rakhman, Klasifikasi Tweet Spam Dan Valid Menggunakan Seleksi Fitur Chi Square Dan Algoritma Naïve Bayes ClasifierPada Tweet Berbahasa Indonesia, Skripsi, Universitas Gajah Mada Yogyakarta, 2016

[7] [7] A. Muhantini, Collaborative Filtering SMS Spam Berbahasa Indonesia Menggunakan Algoritma Naive Bayes, Skripsi, Universitas Islam Negeri Yogyakarta, 2013.

[8] [8] A. Anggara, Penerapan Sistem Multi Agen Untuk Ekstraksi Informasi Kebutuhan Darah Pada Twitter, Skripsi, Universitas Gajah Mada Yogyakarta, 2013.

[9] [9] F. P. Azali, Klasifikasi Pengaduan Masyarakat Berbasis Sms Dengan Metode Naive Bayes Classifier. Skripsi Universitas Gajah Mada Yogyakarta, 2016.

[10] [10] https://Twitter.com

[11] [11] J. Han, M.Kamber, J.Pei, *Data Mining: Concepts and Techniques, 3rd Edition,* Morgan Kaufmann Publishers is an imprint of Elsevier,Waltham, MA 02451, USA, 2012, pp.350-351.

[12] [12] D.Xhemali. *Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages - Scientific Figure on Research Gate.* https://www.researchgate.net/41392270_fig1_Fig-1-System-Stages, 2009.

[13] [13] I. Hemalatha, P.G.Varma, A.Govardhan, Preprocessing the Informal Text for Efficient Sentiment Analysis, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol. 1, July – August 2012, ISSN 2278-6856.