❒     5409

# Twitter Sentiment Analysis on 2013 Curriculum Using Ensemble Features and K-Nearest Neighbor

**M. Rizzo Irfan, M. Ali Fauzi, Tibyani, Nurul Dyah Mentari**
Faculty of Computer Science, Brawijaya University, Indonesia

| Article Info | ABSTRACT |
|---|---|

2013 curriculum is a new curriculum in the Indonesian education system which has been enacted by the government to replace KTSP curriculum. The implementation of this curriculum in the last few years has sparked various opinions among students, teachers, and public in general, especially on social media twitter. In this study, a sentimental analysis on 2013 curriculum is conducted. Ensemble of several feature sets were used including textual features, twitter specific features, lexicon-based features, Parts of Speech (POS) features, and Bag of Words (BOW) features for the sentiment classification using K-Nearest Neighbor method. The experiment result showed that the the ensemble features have the best performance of sentiment classification compared to only using individual features. The best accuracy using ensemble features is 96% when k=5 is used.

*Corresponding Author:*

M. Ali Fauzi,
Faculty of Computer Science,
Brawijaya University,
Jl. Veteran, Malang, Indonesia.
Email: moch.ali.fauzi@ub.ac.id

## 1. INTRODUCTION

According to a survey conducted by IDC (International Data Corporation), a market research agency in the United States, in 2013 to 2020 the number of digital information will continue to grow corresponding the factor of 10, from 4 trillion gigabytes to 44 trillion gigabytes. This is commensurate with the growing number of users of social media nowadays since they want to be able to exchange information more quickly. However, not all information displayed always has a good opinion value. There are multiple opinions that can be either positive or negative to a particular topic that is being discussed.

One of the most widely circulated information today is the opinion of 2013 curriculum by Indonesian Ministry of Education and Culture. The 2013 curriculum is a new curriculum to succeed the old 2006 curriculum (often referred as KTSP) in the Indonesian education system [1-2]. The application of this new curriculum reaps a variety of opinions from public. There are some significant differences between this new curriculum and the old one such as students are required to be active, teachers only submit materials and students must find out for themselves, there are some lessons that are eliminated, require scouts and other things that increasingly provoked various opinions about the topic especially among twitter users.

Twitter is one of the largest and most dynamic social media contributors based on user-generated content. It is very popular among Indonesian people. In Twitter, users can post status or a message that is called as a tweet that is not more than 140 characters. It is estimated that there are about 400 million tweets posted by 200 million users daily [3].

In this study, sentiment analysis system is built to know the positive or negative opinion that developed in the society about 2013 curriculum through twitter media. Ensemble of several features will be used for classifying the polarity of tweets. One of the previous work conducted by [4] was using several statistical and semantic features including textual features, twitter specific features, lexicon based features, Parts of Speech (POS) features, and Bag of Words (BOW) features alone only give 73.8% accuracy. Meanwhile, the ensemble of features can improve the accuracy to become 87.7%. The use of this ensemble feature also give better accuracy than other features like unigram + bigram, propagation label, sentiment topic feature, sentistrength, meta level features, and semantria (online system).

In this study, we will explore the use of K-Nearest Neighbor (KNN) for the classification task. K-Nearest Neighbor (K-NN) is an algorithm that classifies objects based on learning data that resembles the closest resemblance to the object [5-6]. In a previous study conducted by [7], K-NN yielded the highest accuracy value when compared with Naive Bayes and Term Graph. The average accuracy result is 98.95% for K-NN method, 62.66% for Naive Bayes and 98.72% for Term Graph. Therefore, K-NN would be more suitable to use for this task.

## 2.      RESEARCH METHOD

This section describes the steps in the sentiment analysis system. The main workflow of the system can be seen in Figure 1. As shown in Figure 1, the first step conducted in this system is taking a tweet that entered by the user and then standardization of words is conducted. This standardization is the purpose of this standardization is to convert non-standard words into standard and to correct spelling errors. The next step is features extraction. Some feature used in this work including including textual features, twitter specific features, lexicon based features, Parts of Speech (POS) features, and Bag of Words (BOW) features.

The detailed features can be seen in Table 1. For POS fatures, we utilize kateglo API to get POS tag of each words. We also use data from previous research for lexicon of positive and negative words, emoticons, data dictionary word amplifier or intensifier word by [8]. We also use dictionary of non-standard or slang language by [9]. Special for the BOW features extraction, preprocessing generaly should be conducted first before the extraction begin [10]. This preprocessing step including tokenization, filtering, and stemming. In the tokenization process, each documents is splitted into smaller units called token [11]. In this step, all letters are converted into lowercase and some characters like punctuation, numbers, and HTML tags are also removed [12-13]. In filtering, uninformative words are removed based on the existing stoplist by by Tala [14]. The last process in preprocessing is stemming or restoring every words to its root [15-16]. In this case, we use Sastrawi Stemmer.

The last stage is sentiment classification using K-Nearest Neighbor. This stage output is test data category wheter they are positive or negative. For the term weighting method, we use TF.IDF since it is a very poplar method and generally gives very good performance on classification task [17]. The neighbor proximity calculation in this study is using cosine similarity instead of Euclidian distance. Based on the previous works [18-19], cosine similarity gives performs very well on NLP task.
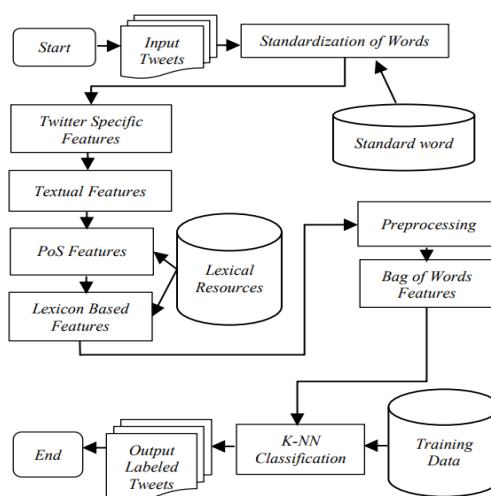


Figure 1. Sentiment Analysis System using Ensemble Features

Table 1. The Ensemble Features

| Type | ID | Feature Description |
|---|---|---|
| Twitter Specific | F1 | Whether the tweet contains a #hashtag or not. |
| | F2 | Whether the tweet is a retweet or not. |
| | F3 | Whether the tweet contains a user name or not. |
| | F4 | Whether the tweet contains a URL or not. |
| Textual Features | F5 | TweetLength: Number of words in the tweet. |
| | F6 | AvgWordLength: Average character length of words. |
| | F7 | Number of question marks in the tweet. |
| | F8 | Number of exclamation marks in the tweet. |
| | F9 | Number of quotes in the tweet. |
| | F10 | Number of words start with the uppercase letter in tweet. |
| | F11 | Whether the tweet contains a positive emoticon or not. |
| | F12 | Whether the tweet contains a negative emoticon or not. |
| Parts of Speech (PoS) Features | F13 | Number of noun PoS in the tweet. |
| | F14 | Number of adjective PoS in the tweet. |
| | F15 | Number of verb PoS in the tweet. |
| | F16 | Number of adverb PoS in the tweet. |
| | F17 | Number of interjection PoS in the tweet |
| | F18 | Percentage of noun PoS in the tweet. |
| | F19 | Percentage of adjective PoS in the tweet. |
| | F20 | Percentage of verb PoS in the tweet. |
| | F21 | Percentage of adverb PoS in the tweet. |
| | F22 | Percentage of interjection PoS in the tweet. |
| Lexicon Based Features | F23 | Number of positive words in the tweet. |
| | F24 | Number of negative words in the tweet. |
| | F25 | Number of positive words with adjective PoS. |
| | F26 | Number of negative words with adjective PoS. |
| | F27 | Number of positive words with verb PoS. |
| | F28 | Number of negative words with verb PoS. |
| | F29 | Number of positive words with adverb PoS. |
| | F30 | Number of negative words with adverb PoS. |
| | F31 | Percentage of positive words with adjective PoS. |
| | F32 | Percentage of negative words with adjective PoS. |
| | F33 | Percentage of positive words with verb PoS. |
| | F34 | Percentage of negative words with verb PoS. |
| | F35 | Percentage of positive words with adverb PoS. |
| | F36 | Percentage of negative words with adverb PoS. |
| | F37 | Number of intensifier words in the tweet. |
| BOW Features | F38 | Term1 |
| | … | … |
| | F38+n | Termn |

## 3. RESULTS AND ANALYSIS

The dataset used in this study is obtained from twitter. A total of 200 tweets containing the keyword 'Kurikulum2013' were taken. Of the 200 data, 100 data are positive tweets and the other 100 are negative ones. The category of the tweets is annotated manually by an expert. Datasets then be divided into training data and test data. A total of 150 tweet data were used as training data (75 positive categorical data and 75 negative categorical data) and 50 as test data (25 positive categorical data and 25 negative categorical data).

In this study, several experiments are conducted and the results are analyzed. The first experiment is to determine the effect of k value of K-NN to the accuracy of sentiment analysis system. The next experiment to explore the use of the BOW features, the ensemble features without BOW (textual features, twitter specific features, lexicon-based features, and POS features), and the combination of them all.

### 3.1. K Value Experiment Result anad Analysis

The first experiment is to analyze the effect of k value of K-NN to the accuracy of sentiment analysis system and determine which the k value of K-NN that has the best accuracy value. In this experiment, the features used are the complete ensemble features. The experiment is conducted using several values of k started from 3 to 31. The experiment result displayed in Figure 2.
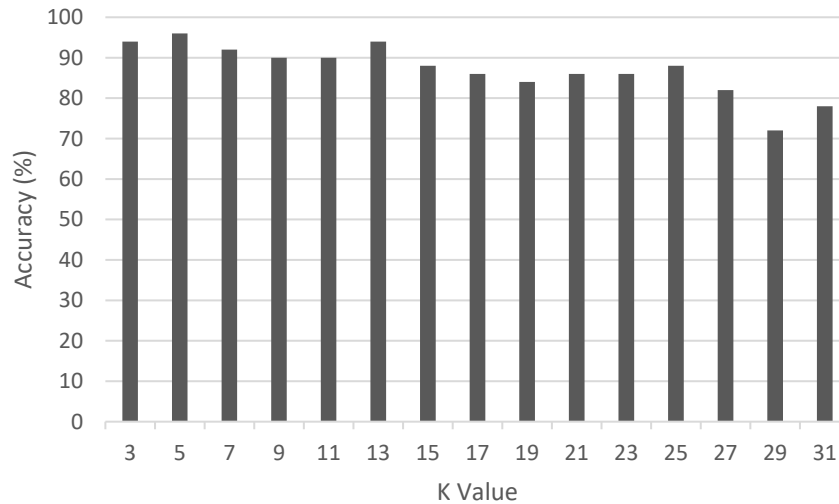
Figure 2. K Value Experiment Result

The result showed that when the value of k was too small, for example the value k=3, the classification accuracy could not reach the maximum point because there are some relevant data that are not involved in the category voting by K-NN. However, when the value of k was too big, for example when the k value was more than 13, the accuracy decreased slowly because there are many irrelevant data that had been involved in the category voting. The best accuracy value is obtained when k=5 with 96% accuracy. Therefore, this best value of k would be used for the next experiment.

### 3.2. Ensemble Features Experiment Result anad Analysis

This experiment aim to analyze the use of ensemble features. In this experiment, we compared the use of the the BOW features, the ensemble features without BOW (textual features, twitter specific features, lexicon based features, and POS features), and the combination of them all. The experiment result displayed in Figure 3.
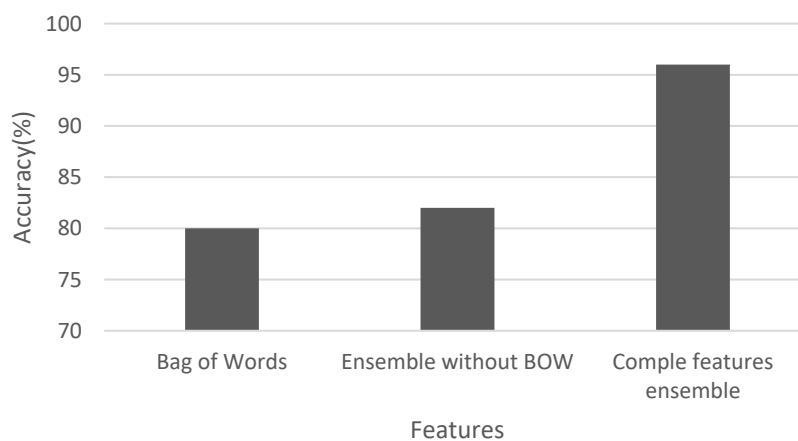


Figure 3. Ensemble Features Experiment Result

It is clear to see from Figure 3 that the most inferior performance is obtained when only BOW features were used with accuracy value of 80%. It happened because there are some short tweets which only has a very few words that can lead to sparsity and ambiguity. Consequently, most words contained in the test

tweet data never appeared in the training data. This shows that the use of this feature is highly dependent on word statistics contained in the training data.

The ensemble features without BOW (textual features, twitter specific features, lexicon-based features, and POS features) had slightly better performance than only involving BOW features. The accuracy value was 82%. This feature is very dependent on the dictionary or lexicon used. Words on test data tweets that have not been well-recognized or not contained within the lexicon will affect the feature's value so that it impacts the classification result.

The complete combination of all features sets perform the best accuracy by 96%. There is an improvement compared to the previous features. By combining all of the feature sets, it can cover the weakness of each features sets and get the best out of them.

## 4. CONCLUSION

In this study, we built sentiment analysis of 2013 curriculum using K-NN and ensemble features. Various test scenarios have been conducted to specify the effect of k value and the effect of feature combination on sentiment classification accuracy. The value of k is very prominent in the accuracy of the K-NN method, the best value k obtained when k was 5 with the accuracy of 96%. The k value that is too small causes the accuracy obtained has not reached the maximum point otherwise the k value too much will cause the accuracy to decrease.

Apart from the k values, feature combinations also have significant significant influence in improving the accuracy. Combining BOW features and other features including textual features, twitter-specific features, POS features, and lexicon-based features can improve the accuracy compared to only using independent features. Incorporating this feature can cover the weaknesses of each feature sets and and get the best out of them. The best accuracy gained by combining all features sets reaches 96% accuracy value.

## REFERENCES

[1]   Poerwati, Loeloek Endah, and Sofan Amri. "Panduan Memahami Kurikulum 2013." *Jakarta: Prestasi Pustaka* 2013.
[2]   Ima Nurdiana, Yulia. "Comparative Study of Implementation of 2013 Curriculum in Class X Between State High School 1 Taman Sidoarjo and Madrasah Aliyah Sidoarjo" (in Bahasa), PhD diss., UIN Sunan Ampel, 2015.
[3]   Da Silva, Nadia FF, Eduardo R. Hruschka, and Estevam R. Hruschka. "Tweet sentiment analysis with classifier ensembles", *Decision Support Systems* 66: pp. 170-179; 2014
[4]   Siddiqua, Umme Aymun, Tanveer Ahsan, and Abu Nowshed Chy. "Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog." In *Computer and Information Technology (ICCIT), 2016 19th International Conference on*, IEEE, pp. 304-309; 2016.
[5]   Hardiyanto, Erik, and Faisal Rahutomo. "Preliminary Classification Study Indonesian Wikipedia Articles Using the K-Nearest Neighbor Method (in Bahasa)" *Sentrino,* 2(1): 158-165; 2016.
[6]   Suharno, Claudio Fresta, M. Ali Fauzi, and Rizal Setya Perdana. "Indonesian Text Classification On Online Combustion Documents Using K-Nearest Neighbors And Chi-Square Methods." *Systemic: Information System and Informatics Journal,* Vol 3 no. 1: 25-32; 2017.
[7]   Bijalwan, Vishwanath, Vinay Kumar, Pinki Kumari, and Jordan Pascual. "KNN based machine learning approach for text and document mining." *International Journal of Database Theory and Application*, Vol 7, no. 1: 61-70; 2014.
[8]   Wahid, Devid Haryalesmana, and S. N. Azhari. "Summarizing Extractive Sentiments on Twitter Using Hybrid TF-IDF and Cosine Similarity (in Bahasa)", *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 10, no. 2: 207-218.
[9]   Antinasari, Prananda, Rizal Setya Perdana, and M. Ali Fauzi. "Sentiment Analysis of Film Opinion on Indonesian Language Twitter Documents Using Naive Bayes with Non-Standard Word Repair (in Bahasa)" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN* 2548: 964X.
[10]  Feldman, Ronen, and James Sanger. "The text mining handbook: advanced approaches in analyzing unstructured data". Cambridge university press, 2007.
[11]  Fauzi, M. Ali, Agus Zainal Arifin, and Sonny Christiano Gosaria. "Indonesian News Classification Using Naïve Bayes and Two-Phase Feature Selection Model", *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),* Vol 8, no. 3: 610-615; 2017.
[12]  Fauzi, M. Ali, Agus Zainal Arifin, and Anny Yuniarti. "Arabic Book Retrieval using Class and Book Index Based Term Weighting." *International Journal of Electrical and Computer Engineering (IJECE)*, Vol 7, no. 6: 3705-3710; 2017.
[13]  Fauzi, M. Ali, Agus Arifin, and Anny Yuniarti. "Term Weighting Based on Book Index and Class for Ranking of Arabic Documents (in Bahasa)", *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi* 5, no. 2 (2013).
[14]  Tala, Fadillah Z. "A study of stemming effects on information retrieval in Bahasa Indonesia." *Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands* (2003).
[15]  Fauzi, M. Ali, and Anny Yuniarti. "Ensemble Method for Indonesian Twitter Hate Speech Detection." *Indonesian Journal of Electrical Engineering and Computer Science* 11, no. 1 (2018).

[16] Fauzi, M. Ali, Ro'I. Fahreza Nur Firmansyah, and Tri Afirianto. "Improving Sentiment Analysis of Short Informal Indonesian Product Reviews using Synonym Based Feature Expansion." *TELKOMNIKA Telecommunication, Computing, Electronics and Control.* Vol 16, no. 3; 1345-1350; 2018.

[17] Christopher, D. Manning, Raghavan Prabhakar, and S. C. H. Ü. T. Z. E. Hinrich. "Introduction to information retrieval", *An Introduction to Information Retrieval*, 151: 177; 2008.

[18] Pramukantoro, Eko Sakti, and M. Ali Fauzi. "Comparative analysis of string similarity and corpus-based similarity for automatic essay scoring system on e-learning gamification." In *Advanced Computer Science and Information Systems (ICACSIS), 2016 International Conference on*, pp. 149-155. IEEE, 2016.

[19] Fauzi, M. Ali, Djoko Cahyo Utomo, Eko Sakti Pramukantoro, and Budi Darma Setiawan. "Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Gamification Based E-Learning." In *International Conference on Advances in Image Processing (ICAIP), 2017 International Conference on*, pp. 151-155. ACM, 2017.