Sarcasm Detection on Indonesian Twitter Feeds

Dwi A. P. Rahayu
Informatics Department
University of Muhammadiyah Malang
Malang, Indonesia
dwi.ap.rahayu@umm.ac.id

Soveatin Kuntur
Graduate of Informatics Department
University of Muhammadiyah Malang
Malang, Indonesia
kuntursoveatin@gmail.com

Nur Hayatin
Informatics Department
University of Muhammadiyah Malang
Malang, Indonesia
noorhayatin@umm.ac.id

***Abstract*. In social media, some people use positive words to express negative opinion on a topic which is known as sarcasm. The existence of sarcasm becomes special because it is hard to be detected using simple sentiment analysis technique. Research on sarcasm detection in Indonesia is still very limited. Therefore, this research proposes a technique in detecting sarcasm in Indonesian Twitter feeds particularly on several critical issues such as politics, public figure and tourism. Our proposed technique uses two feature extraction methods namely interjection and punctuation. These methods are later used in two different weighting and classification algorithms. The empirical results demonstrate that combination of feature extraction methods, tf-idf, k-Nearest Neighbor yields the best performance in detecting sarcasm.**

Keywords—social media, negative opinion, sentiment analysis, sarcasm detection, feature extraction

I. INTRODUCTION

In this modern day, online data grows significantly every minute. Twitter is one of social media which produces millions of data every day. Indonesia ranked as 5th biggest country of Twitter users[1]. Thus, Indonesian tweets data is abundant and worth to be analyzed. Twitter limits a message to have a maximum of 280 characters which leads users either be concise or be creative in writing their messages. Most of Indonesian Twitter's users are active and expressive, they can creatively express their tweet on trending topics in that limited number of characters[1]. As part of their creativity, some of them often use sarcasm, i.e. positive words to express negative opinion, in their Twitter message.

Sarcasm or irony has been extensively explored in linguistic and psychology field. Nevertheless, in natural language processing field, detecting sarcasm within a sentence or message is still considered as a big challenge because the lexical features extracted from the sentence do not give enough information to detect sarcasm[2]. The existence of sarcasm can also drop the performance of sentiment analysis techniques[2]. While sarcasm detection is an emerging research field in English natural language processing. There are only very limited researches which focus on sarcasm detection in Indonesian. To the best of our knowledge, there is only one research on Indonesian sarcasm detection using full machine learning algorthm[2]. Therefore, this research aims to fill this gap by proposing a technique in Indonesian sarcasm detection.

This research investigates and detects sarcasm used in Indonesian Twitter feeds on several trending topics in 2018 such as politics, public figure and tourism. Our sarcasm detection technique uses combination of feature extraction method, weighting method and classification algoritm. The writers first use the combination interjection and punctuation, Bag of Words and Naïve Bayes to detect sarcasm. The combination of interjection and punctuation, tf-idf and k-Nearest Neighbor are employed. We compare these two combinations to get the best technique in detecting sarcasm.

We discuss current techniques used in sarcasm detection in section two, followed by details of our techniques in section three. We then present our experiment data, settings and results in section four, followed by conclusion and future work in section five.

II. RELATED WORK

Sentiment analysis is a technique to identify people's opinion, emotion towards any situation and attitude. Sentiment analysis is used to determine whether people's opinion or emotion is positive, neutral or negative based on words used in their sentences. Researchers use machine learning to further investigate sarcasm on text data collected from various sources[3][4][5][6].

Some of feature extraction methods used in sarcasm detection on English sentences are punctuation and interjection. Early work on sarcasm detection on Twitter data using punctuation and interjection successfully gained a f-measure score of 0.813[3]. In another work which detects sarcasm in Facebook comment posts, combination of interjection and punctuation with syntactic feature increased the f-measure score into 0.852[4].

Despite many researches have been conducted to detect sarcasm in English, there is only one of a kind on Indonesian. The only machine learning based sarcasm detection on Indonesian social media messages is proposed by Edwin Lunando and Ayu Purwarianti[2]. They used unigram, the number of interjection words, negativity and question word as feature extraction method, then use these features in classifiers such as Naïve bayes, Maximum Enthropy and Support Vector to detect sarcasm. The accuracy of their proposed technique was still very low. This low accuracy was caused by many sarcasm texts in their dataset have no global topic. They also recognized terms using translated SentiWordnet. They translated English SentiWordNet terms into Indonesian using Google Translate which may lead to undetected terms as Indonesian words used in social media are very rich[2].

In this research, we investigate sarcasm detection technique on Indonesian sentences by combining punctuation and interjection feature extraction methods with two different weighting methods and two classification algorithms. Our technique does not involve any translation process to avoid similar problems faced by previous researchers happened[2]. Instead, we use pre-processing and stemming algorithm which are designed for Indonesian. Details of our technique will be further discussed in next section.

978-1-5386-8402-3/18/$31.00 ©2018 IEEE

III. Proposed Technique

Opinions posted in social media can be categorized as positive, negative, and neutral. A positive sentiment can be further classified as actual positive or sarcasm as shown in Fig. 1[2]. Therefore, positive tweets have to be extracted from crawled tweets prior to sarcasm detection process.

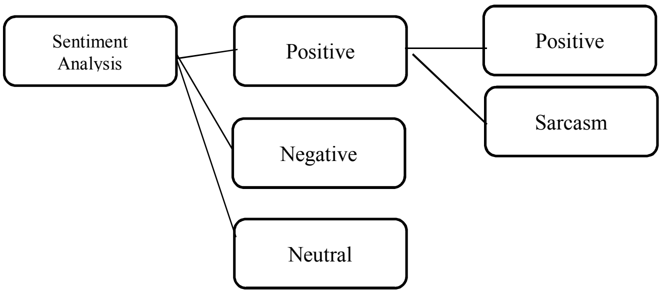

Fig. 1. Levelled method in sentiment analysis [2]

As displayed in Fig. 2, our technique includes two phases. The first phase of our technique preprocesses the tweets and categorizes the sentiment of tweets. This preprocessing technique is very important to extract meaningful words from sentences and discard common words and symbols[7]. In preprocessing, we firstly use a case folding method to make all sentences have a uniform case. We then use a filtering method to remove URLs, mentions, and hashtags within the tweets. We also leverage stemming algorithm on Indonesian, Sastrawi Stemmer, which can be accessed on github[8]. We use this stemmer to remove suffices and prefixes from words within tweets.

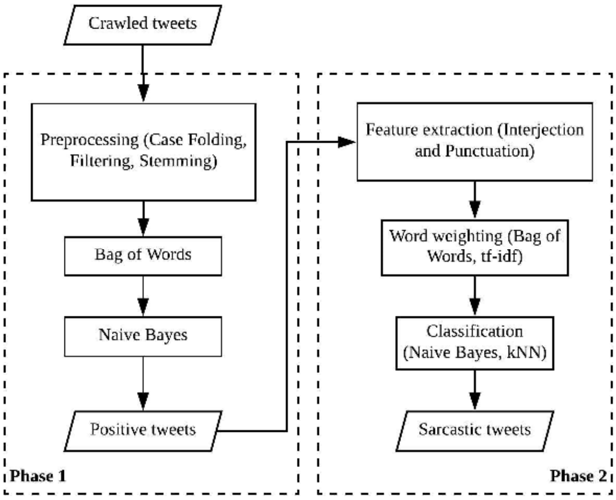
Crawled tweets
Preprocessing (Case Folding, Filtering, Stemming)
Bag of Words
Naive Bayes
Positive tweets
Phase 1
Feature extraction (Interjection and Punctuation)
Word weighting (Bag of Words, tf-idf)
Classification (Naive Bayes, kNN)
Sarcastic tweets
Phase 2

Fig. 2. Sarcasm detection phases

Once all tweets are preprocessed, we use Bag of Words to count the frequency of words and run Naïve Bayes classifier to categories the tweets as positive, negative and neutral tweets. The positive tweets are then further investigated in the second phase of our technique which detects the sarcasm.

The second phase of our technique combines two feature extraction methods, a word weighting method and a classification algorithm to extract sarcasm tweets from positive tweets. The feature extraction methods extract any words which indicated sarcasm and any sign that showed emotional suppression. The word weighting method rank preprocessed words based on its importance/frequency within the documents. The classification algorithm categories tweets based on similarity between testing data and training data which contained feature that have implemented.

In this second phase, we use two combinations. These two combinations are summarized in Table I.

TABLE I. Sarcasm Detection Technique Combination

Combination 1	Interjection + Punctuation + Bag of Words + Naïve Bayes classifier
Combination 2	Interjection + Punctuation + tf-idf + Cosine similarity + k-Nearest neighbor

For the first combination, we combine interjection and punctuation in feature extraction process, Bag of Words in word ranking process, and Naïve Bayes in classification process. Interjection extracts any words which indicated to be sarcasm such as “wow”, “anjir”, “anjay”, “njir”. Punctuation extracts any sign indicated emotional suppression such as “!!”, “?!”, “??”. Bag of words weighting method ranks extracted words(features) based on its frequency within the sentence. Naïve Bayes classifies the tweets by calculating the probability of each tweet as sarcasm tweets based on extracted ranked words (features).

We then utilize a different combination in comparison with the first one. This combination also uses interjection and punctuation for feature extraction process. However, instead of using Bag of Words and Naïve Bayes, tf-idf and k-Nearest Neighbor are used instead for word weighting and classification process. Note that tf-idf ranks the words based on its appearance in total documents, whereas k-Nearest Neighbor classifies the tweets by calculating cosine similarity of tweets’ features towards positive and sarcasm tweets in training data set.

IV. Experiment

A. Dataset and Software

In our experiment we use Indonesian tweets crawled from Twitter. We crawled 2315 tweets on various topics such as public figure, politics, places, and tourism. We crawled tweets which contains trending topics such as “apbd”, “apbn”, “#thepowerofsetnov”, “Jogja Baik Saja”, “Bu Dendy” and “lgbt”. These words are selected based on Indonesian Twitter feed trending topics in 2018.

We divide this data into 1389 training data and 926 testing data. The training data consist of 538 positive tweets, 213 negative tweets, 638 neutral tweets. The positive tweets include 217 sarcasm tweets. In order to develop the ground truth, each tweet is manually labeled as positive, sarcasm, neutral and negative by two linguistic teachers.

These tweets are then preproccessed using case folding, filtering and stemming as discussed in previous section. Table II shows an example of sarcastic tweet and its preprocessed result. The underlined words show the transformation from unprocessed into proceesed tweets.

We mainly use Phyton to conduct our study. The first phase of our technique uses Bag of Words method and Naïve Bayes classifier provided in free TextBlob Python library [9] The second phase of our technique is also implemented in Python. We developed our own code to implement the weighting methods and classifier algorithms.

TABLE II. TWEET PREPROCESSING

Preprocessing Step	Tweet
Initial	"@denradityaa: Gue kalo jadi anak bu dendy pas dimarahin dikasih tiket umroh sampe kiamat kali ya?"
Case folding	"@denradityaa: gue kalo jadi anak bu dendy pas dimarahin dikasih tiket umroh sampe kiamat kali ya?"
Filtering	gue kalo jadi anak bu dendy pas dimarahin dikasih tiket umroh sampe kiamat kali ya
Stemming	gue kalo jadi anak bu dendy pas marah kasih tiket umroh sampe kiamat kali ya

To measure the performance of our technique we use three parameters which are commonly used in information retrieval namely precision, recall and f-measures. Precision shows the fraction of correctly classified tweets out of all retrieved tweets with the reference of ground truth. Recall shows the fraction of correctly classified tweets out of all relevant tweets. f-measure shows harmonic means of precision and recall. These three parameters give better insight of learning performance on internet based phrases than simple accuracy since, commonly, there is a big data imbalance within crawled documents[10].

B. *Experiment Result*

The technique performances are measured in three different states. The first measurement is to analyze the performance of first phase of our technique.

TABLE III. SENTIMENT ANALYSIS RESULT

No	Parameter	Score
1	Recall Positive	0.81
2	Recall Negative	0.92
3	Recall Neutral	0.73
4	Precision Positive	0.60
5	Precision Negative	0.87
6	Precision Neutral	0.90
7	f-measure Positive	0.69
8	f-measure Negative	0.89
9	f-measure Neutral	0.81

Table III shows the result of phase 1 of our technique which categorizes crawled tweets into positive, negative, and neutral tweets. Despite high precision of negative and neutral class which is 0.87 and 0.90 respectively, the precision of positive class in this first phase is only 0.60 due to significant number of positive tweets categorized as neutral. The positive recall, negative recall, and neutral recall are 0.81, 0.92, and 0.73, respectively. Those precision and recall scores produces f-measure of positive, negative and neutral class as 0.69, 0.89 and 0.81 respectively.

These measurement parameter scores show that the combination of Bag of Words and Naïve Bayes is capable to classify each tweet class from the actual tweet class. However, the classification accuracy might be improved further if other word ranking methods is implemented.

We run an experiment on phase 2 of our technique on positive tweets resulted from previous phase. In phase 2, we use two different algorithm combinations as highlighted in Table I to identify tweets which contain sarcasm. Table III shows the combination 1 measurement parameter scores.

TABLE IV. SARCASM DETECTION RESULT – COMBINATION 1

No	Testing	Score
1	Recall Sarcasm	0.92
2	Recall Positive	0.65
3	Precision Sarcasm	0.34
4	Precision Positive	0.97
5	f-measure Sarcasm	0.50
6	f-measure Positive	0.78

As displayed in Table IV, precision of positive class is very high, 0.97. However, the precision of sarcasm class is very low, 0.34. The recall value of positive class and sarcasm class are 0.92 and 0.65 respectively.

These scores show that combination 1 is able to separate sarcasm class from actual sarcastic tweets, however combination 1 is only able to separate very few sarcastic tweets out of all positive tweets. On the other words, there are still many sarcastic tweets which are categorized as positive tweets.

The imbalance of precision and recall of both sarcasm and positive class leads to low f-measure scores. The f-measure score of sarcasm class is 0.50 and the f-measure of positive class is 0.78. These low f-measure scores imply that combination 1 is not able to accurately extract sarcastic tweets. Based on our analysis, Bag of Words might not be a good weighting method for sarcasm detection as it significantly affects the classification result.

The subsequent experiment of phase 2 uses combination 2 to detect sarcastic tweets out of positive tweets. The results of this experiment are shown in Table V. As shown in Table V, precision of positive class is 0.95. It is slightly lower than precision of positive class using combination 1. However, the recall of positive class is 0.82, which is significantly higher than combination 1. The precision and recall of sarcasm class are also better for combination 2. The precision and recall of sarcasm class are 0.74 and 0.92 respectively. These precision and recall scores leads to high f-measure scores. The f-measure score of positive class is 0.88 and the f-measure score of sarcasm class is 0.88. These higher f-measure scores suggest that combination 2 detects sarcastic tweets better than combination 1.

TABLE V. SARCASM DETECTION RESULT – COMBINATION 2

No	Testing	Score
1	Recall Sarcasm	0.92
2	Recall Positive	0.82
3	Precision Sarcasm	0.74
4	Precision Positive	0.95
5	f-measure Sarcasm	0.82
6	f-measure Positive	0.88

The comparison of combination 1 and combination 2 f-measure scores is re-highlighted in Fig 3 to give a better picture of each combination's performance. This comparison shows that the combination tf-idf and k-Nearest Neighbor gives significantly better prediction of sarcastic tweets than the combination of Bag of Words and Naïve Bayes. Combination 2 is 32% more accurate than combination 1 in detecting sarcasm within tweets. The tf-idf methods give smoother rank of words which leads to better features to be selected and used in classification. k-Nearest Neighbor which distinguishes sarcastic tweets based on similar tweets with smallest distance can accurately differentiate sarcastic tweets from positive tweets.

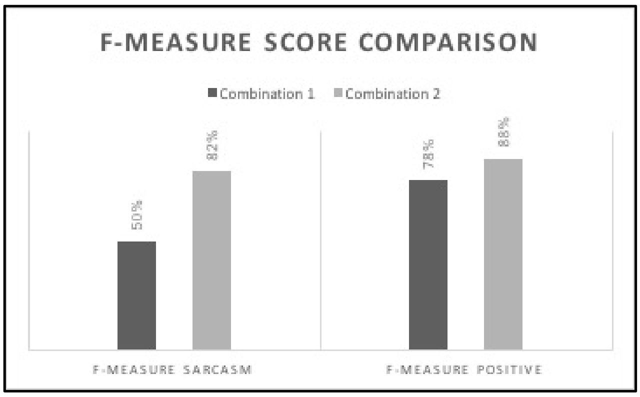

Fig. 3. f-measure of combination 1 and combination 2

Fig. 3 also shows that combination 2 has better f-measure score in detecting positive class than combination 1 even though the difference is not significant. These f-measure scores indicate that combination 2 performs better in both sarcasm and positive tweets detection.

The empirical results shown in this section concludes that the combination of interjection and punctuation, tf-idf and k-Nearest Neighbor is our recommended technique in sarcasm detection on Indonesian Twitter feeds.

V. CONCLUSION

Sarcasm is very special as it includes words which mean the opposite of what people really want to say. Sarcasm is widely used to mock someone or to be funny. Sarcasm existence within sentiment analysis field becomes important because its appearance influences sentiment analysis accuracy. Sarcasm detection on English messages has been widely researched, however there is very limited research on sarcasm detection in Indonesian. This research proposes a technique which extract positive tweets and further detects sarcastic tweets in Indonesian Twitter feeds.

The first phase of our technique uses Bag of Words and Naïve Bayes to separate positive tweets from crawled tweets of 2018 trending topics. The second phase of our technique analyses two combinations of feature extraction method, word weighting method and classification algorithms performance in detecting sarcastic tweets from previously classified positive tweets.

Empirical results show that combination of Bag of Words and Naïve Bayes that is used in first phase is able to extract positive tweets with f-measure score 0.69. Experiment results on second phase shows that combination of interjection, punctuation, tf-idf and k-Nearest Neighbor can accurately detect sarcastic tweets with f-measure score 0.82. The experiment results also show that this combination outweighs the combination of interjection, punctuation, Bag of Words and Naïve Bayes performance in detecting sarcasm. Thus, our technique is a promising technique to detect sarcasm in Indonesian sentences.

Since sarcasm detection research in Indonesia is currently very limited, there are many open research opportunities in this field. This work in proper can be extended further by combining different word weighting methods and classifiers, as well add more sophisticated feature extraction technique such that more sarcasm is detected.

REFERENCES

[1] I. Nurcahyani, "Tiga Karakter Pengguna Twitter di Indonesia," 2015. [Online]. Available: https://www.antaranews.com/berita/515549/tiga-karakter-pengguna-twitter-di-indonesia.

[2] E. Lunando and A. Purwarianti, "Indonesian social media sentiment analysis with sarcasm detection," *2013 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2013*, pp. 195–198, 2013.

[3] M. Bouazizi and T. Ohtsuki, "Sarcasm detection in twitter: »all your products are incredibly amazing!!!» - are they really?," *2015 IEEE Glob. Commun. Conf. GLOBECOM 2015*, pp. 1–6, 2015.

[4] M. S. M. Suhaimin, M. H. A. Hijazi, R. Alfred, and F. Coenen, "Natural language processing based features for sarcasm detection: An investigation using bilingual social media texts," *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, pp. 703–709, 2017.

[5] M. Hu and B. Liu, "Opinion extraction and summarization on the web," *Aaai*, pp. 1–4, 2006.

[6] E. Forslid, "Automatic irony- and sarcasm detection in Social media," 2015.

[7] A. Krouska, C. Troussas, and M. Virvou, "119. The effect of preprocessing techniques on Twitter Sentiment Analysis," *Information, Intell. Syst. Appl. (IISA), 2016 7th Int. Conf.*, pp. 1–5, 2016.

[8] "Sastrawi stemmer bahasa Indonesia." [Online]. Available: https://github.com/sastrawi/sastrawi/blob/master/README.en.md.

[9] S. Loria, "textblob Documentation," 2014.

[10] G. Hripcsak and A. S. Rothschild, "Agreement, the F-measure, and reliability in information retrieval," *J. Am. Med. Informatics Assoc.*, vol. 12, no. 3, pp. 296–298, 2005.