



Indonesian Sarcasm Detection Using Convolutional Neural Network

Kiefer Stefano Ranti, Abba Suganda Girsang

Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University
Jakarta, Indonesia 11480 Email: kiefer.ranti@binus.ac.id, agirsang@binus.edu

Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University
Jakarta, Indonesia 11480 Email : agirsang@binus.edu

ABSTRACT

In sentiment analysis, sarcasm is considered to be one of the most difficult problems to solve. Sarcasm or verbal irony is a linguistic phenomenon where the wording of a speaker signifies the opposite of the literal meaning of what they are saying. On social media, Indonesians tend to use sarcastic remarks to criticize something and this behavior can pose a problem when a sentiment analysis is performed where the analysis yields false-positive results. To solve this potential problem, the detection of sarcasm is necessary. A convolutional neural network model is used in this paper to handle the automated identification of sarcasm in Indonesian. Although it is mainly used in computer vision, this model has shown promising results on specific kinds of sentiment analysis and classification problems compared to other neural network models. Using this model can increase detection accuracy compared to discrete models. There have been many studies conducted on sarcasm detection, although the majority is in English. Other languages such as Indonesian are slow in improvement compared to English. This fact can be seen by the lack of publications on Indonesian sarcasm detection studies. This study is conducted to contribute to other Indonesian sarcasm detection studies.

Key words: convolutional neural network, indonesia, sarcasm, text

I INTRODUCTION

Sentiment analysis – also known as opinion mining is a method of evaluating the emotional tones behind a statement to gain an understanding of the behaviors, feelings, and emotions expressed in it. These statements can be a form of dialogues, texts, or conversations. The problem that occurs within these studies is that sometimes, the presence of a sarcastic statement can be falsely detected as a positive sentiment while it is a negative sentiment and vice versa. Sarcasm or verbal irony is a linguistic phenomenon where a speaker's utterance means the opposite of the literal meaning of what they are saying [1]. It is not that easy to detect sarcasm in a sentence, even for humans, as it can heavily depend on the context [2].

Detecting sarcasm is a tricky problem that still exists in recent sentiment analysis studies. The fact that it is widely employed in a lot of conversations also makes it a common thing to encounter in a dataset filled with people's conversations. In this case, it may give some negative impact on the performance of a sentiment analysis process, so it is necessary to have a way to differentiate these kinds of statements [3].

Sarcasm detection is very useful for any sentiment analysis studies, as it reduces the number of false-positive and false-negative, thus increasing the performance of a sentiment analysis model. Many studies have been conducted to improve the sarcasm detection model. One problem is that most of these studies are based on English. Other languages like Indonesian are still falling behind in improvement compared to English. This fact can be seen by the lack of publications on Indonesian sarcasm detection studies [4].

Most sarcasm detection studies use tweets as the dataset. While using tweets as dataset is very suitable as it includes people's conversation, it is still constrained to 280 characters. According to [5], [6], different lengths of text may have different impacts on the sarcasm detection result. For this reason, the dataset of this study will be collected from Reddit, where discussions, conversations, and sarcastic remarks are abundant and less constrained.

The majority of the studies, especially the ones in Indonesian, use machine learning methods such as support vector machine, naive bayes, or k-nearest neighbors [4], [7], [8]. These methods usually take a lot of time on hand engineering features of the data to further increase the accuracy. This feature extraction and engineering step can be simplified by using a neural network method to automate the process. The convolutional neural network is selected over the recurrent neural network in this case, as it performs well on sentiment analysis tasks [9].

This study will conduct a sarcasm detection on a dataset collected from Reddit by implementing the proposed convolutional neural network model. Other methods used in other studies such as k-nearest neighbors and naive bayes are also implemented to compare with the proposed model.

II LITERATURE REVIEW

Various methods had been used to detect sarcasm, one of them is Lexical analysis by extracting the opinion adjective and verbs in the sentence. Lexical and pragmatic features are used in [8], [10], [11], and [12] to distinguish sarcastic, positive and negative utterances. In [10] specifically, 2 lexical features were used, unigrams and dictionary-based; and 3 pragmatic factors were used, positive emoticon, negative emoticon, and ToUser feature that shows if a tweet is a reply to another tweet. The features are then classified by using sequential minimal optimization and logistic regression. The method proposed in [11] successfully identifies sarcasm, determines the polarity of the tweet, and predict some political results out of the political tweets. [8] uses machine learning classification methods such as Naïve

Bayes, Maximum Entropy, and Support Vector Machine to detect sarcasm on Indonesian tweets with additional features like negativity and interjection resulting in increased accuracy. [13] developed a pre-trained sentiment, emotion, and personality models for identifying sarcastic texts using a convolutional neural network and found that the model is very effective for sarcasm detection. While [3] proposed a context augmented neural network model for twitter sarcasm detection. The result of their studies shows that the proposed neural network model is very effective for sarcasm detection and has improved results compared to the state-of-the-art discrete model.

Numerous sarcasm detection studies have already been performed in languages other than English such as Dutch, Japanese, and Czech [14]. Most of the studies, however, are conducted in English. There are only a handful of studies in Indonesian that focuses on detecting sarcasm. [8] proposed two additional features to detect sarcasm after a sentiment analysis is conducted: negativity information and number of interjection words. These additional features helped in improving the accuracy of their model by about 4%. [4] tackled the sarcasm detection technique on Indonesia sentences by combining punctuation and interjection feature extraction methods with two different weighting methods and two classification algorithms. They used a pre-processing and stemming algorithm that is designed for the Indonesian language. [7] focuses on using emoticon as a feature and effectively improve their model. [15] created a system that can be used to detect sarcasm for any topic on twitter, using Support Vector Machine and a corpus of slang words. The features used are N-gram, POS-Tag, Pragmatic, Punctuation, and a combination of all the features.

III PROPOSED METHOD

In this study, the sarcasm detection will be done using the deep learning approach, and the deep learning model is a convolutional neural network. The data collection step is carried out by retrieving posts from social media. The retrieved data will then undergo the pre-processing steps; URL removal, punctuation removal, number removal, lowercasing, abbreviation expansion, normalization, stopword removal, stemming, and tokenization. The pre-processing result will be used as an input for the embedding layer, where it will be transformed into word embeddings, and the embeddings will be passed through a convolutional neural network. After the classification results came out, there will be an evaluation of the results to determine the accuracy.

To create the dataset used in this study, any posted comments in the subreddit “r/indonesia” dated from April 1st, 2020 to May 12th, 2020 are requested using the API. The only problem is the number of a non-sarcastic comment far exceeded the sarcastic ones. Using the API, we specifically search for comments containing “/s” from January 1st, 2020 to May 12th, 2020 to increase the number of sarcastic comments to make the dataset more balanced.

The pre-processing step is carried out on the data gathered so that the comments can be transformed into vectors before being used as an input to the convolutional neural network. All the process in the pre-processing steps is visualized in Figure 3.2.

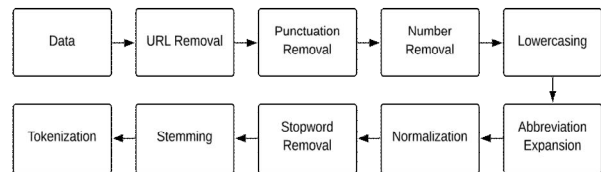


Figure 1: Pre-processing steps

In some state-of-the-art studies, the performance of convolutional neural network in classification is proven to be better than other classification models such as naive bayes . For this reason, convolutional neural network is the classification method that is chosen to classify the dataset in this study. Methods used in other Indonesian sarcasm detection studies will also be implemented as a comparison with the proposed method. These methods are picked based on their popularity and results on other studies. Three of the Indonesian sarcasm detection studies [4], [7], [8] used the naive bayes algorithm in their experiment. The classification method proposed by [4] using k-nearest neighbors and tf-idf is one of the best results obtained from these three studies. In total, there are 3 methods that are going to be used to detect sarcasm in this study: k-nearest neighbors, naive bayes, and convolutional neural network.

For evaluation, the performance of the model will be measured using a confusion matrix to measure accuracy, precision, recall, and F1-score.

IV EXPERIMENT RESULTS

Dataset

The data used in this study are comments from the subreddit /r/indonesia, the comments taken are any comments posted from April 1st, 2020 to May 12th, 2020, but since the number of a non-sarcastic comment far exceeds the sarcastic ones, we manually searched for comments containing “/s” and include them in the dataset. Reddit users assumed a common sarcasm annotation method by adding a “/s” mark at the end of any sarcastic statements they posted. This means almost every post, either submissions or comments that contain the marker “/s” is, proclaimed by the poster, a sarcastic statement.

The collected data consists of 11378 rows of comments, 5654 are labeled sarcastic and 5724 are labeled not sarcastic. All the collected data are pre-processed using the steps on proposed method. Table I shows an example of a comment before and after the pre-processing step.

Table 1: Before & After Pre-processing

	Comment
Before	[kopi aroma] (http://kopiaroma.id/) bukan sachet sih, tapi dikemas per 1/4kg, paling enak ditambihin carnation kental yang kaleng itu
After	'kopi', 'aroma', 'bukan', 'sachet', 'sih', 'tapi', 'kemas', 'per', 'kg', 'paling', 'enak', 'ditambihin', 'carnation', 'kental', 'kaleng'

For k-nearest neighbors and naive bayes, The dataset will be split into two parts, 90% training, and 10% test data. For the convolutional neural network, the dataset will be

split into three parts. 80% for training, 10% for validation, and 10% for testing.

B Results

Results obtained from each method will be presented in confusion matrices.

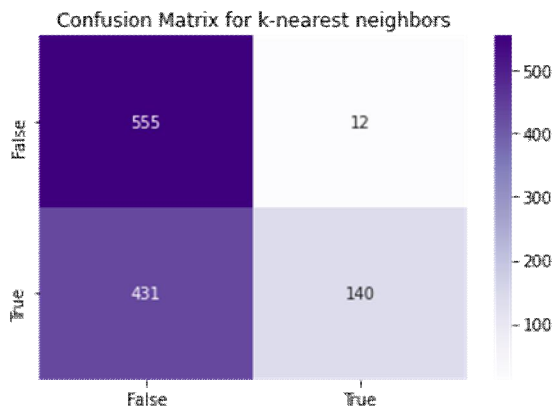


Figure 2: Confusion Matrix for K-nearest Neighbors

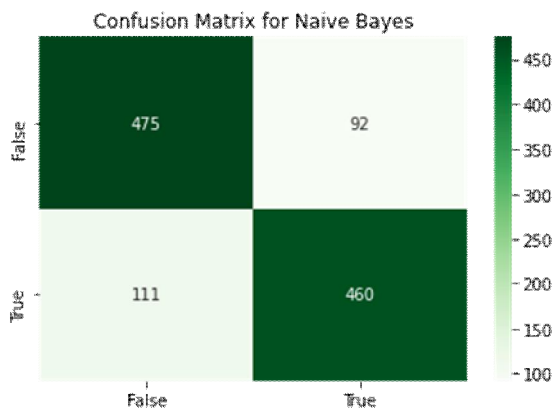


Figure 3: Confusion Matrix for Naive Bayes

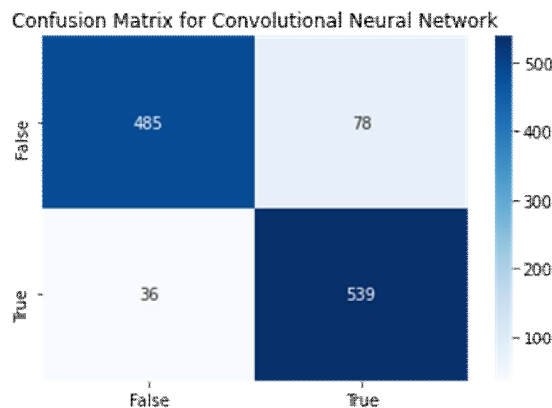


Figure 4: Confusion Matrix for Convolutional Neural Network

After the classification process, results from each method is presented and compared to each other in order to find the best performing model. Table shows the measurement results comparison.

Table 2: Results Comparison

Method	Accuracy	Precision	Recall	F1 Score
KNN	0.61	0.74	0.61	0.55
NB	0.82	0.82	0.82	0.82
CNN	0.90	0.88	0.93	0.90

Based on the comparison and evaluation in Table 4.19, the result shows that the proposed method using convolutional neural network yields the best result, that is 90% in F1 Score, followed by naive bayes for 82%, and lastly by k-nearest neighbors for 55%. Therefore it can be concluded that the proposed method using convolutional neural network is more accurate than k-nearest neighbors and naive bayes.

V CONCLUSIONS

This study shows that detecting sarcasm using the proposed convolutional neural network model can produce an F1 Score of 90% on 1138 comments on the test data. It is an 8% increase compared to the naive bayes model. This result proves that detecting sarcastic statements on text using a convolutional neural network can yield good results.

The challenge in Indonesian natural language processing studies is the lack of dataset. To further improve any Indonesian natural language processing studies, creating a bigger dataset is crucial in order to attract researchers. More datasets also mean more ways to compare and evaluate the results of a model. In terms of data pre-processing, since Indonesians have quite a unique way to communicate in social media (i.e., shortening words into different abbreviation such as “gue” into “gw” or “w”), the pre-processing step actually became more complicated compared to another language like English, for example. The pre-processed data in this study still contains some abbreviations that may hinder the performance of the model. There is a need to expand the Indonesian dictionary on words and abbreviations in the future, in order to make any classification model perform more accurately.

Future studies should also take into account the usage of emoji, stickers, or even images. A single emoji, sticker, or image can drastically change the meaning of a sentence. The challenge will be in interpreting the meaning of an emoji/sticker/image to a person since not everyone has the same comprehension of them.

REFERENCES

- [1] Y. Alotaibi, ‘A New Analysis of Verbal Irony’, *Int. J. Appl. Linguist. Engl. Lit.*, vol. 6, no. 5, p. 154, Jul. 2017, doi: 10.7575/aiac.ijalel.v.6n.5p.154.
- [2] N. Parde and R. Nielsen, ‘Detecting Sarcasm is Extremely Easy ;-)', in *Proceedings of the Workshop on Computational Semantics beyond Events and Roles*, New Orleans, Louisiana, 2018, pp. 21–26, doi: 10.18653/v1/W18-1303.
- [3] Y. Ren, D. Ji, and H. Ren, ‘Context-augmented convolutional neural networks for twitter sarcasm

- detection', *Neurocomputing*, vol. 308, pp. 1–7, Sep. 2018, doi: 10.1016/j.neucom.2018.03.047.
- [4] D. A. P. Rahayu, S. Kuntur, and N. Hayatin, 'Sarcasm Detection on Indonesian Twitter Feeds', *EECSI*, vol. 5, no. 1, pp. 137–141, Nov. 2018, doi: 10.11591/eeesi.v5.1724.
- [5] S. K. Bharti, R. Pradhan, K. S. Babu, and S. K. Jena, 'Sarcasm Analysis on Twitter Data Using Machine Learning Approaches', in *Trends in Social Network Analysis*, R. Missaoui, T. Abdessalem, and M. Latapy, Eds. Cham: Springer International Publishing, 2017, pp. 51–76.
- [6] A. Joshi, V. Sharma, and P. Bhattacharyya, 'Harnessing Context Incongruity for Sarcasm Detection', in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Beijing, China, 2015, pp. 757–762, doi: 10.3115/v1/P15-2124.
- [7] D. Alita, S. Priyanta, and N. Rokhman, 'Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter', *JISEBI*, vol. 5, no. 2, p. 100, Oct. 2019, doi: 10.20473/jisebi.5.2.100-109.
- [8] E. Lunando and A. Purwarianti, 'Indonesian social media sentiment analysis with sarcasm detection', in *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Sanur Bali, Indonesia, Sep. 2013, pp. 195–198, doi: 10.1109/ICACSIS.2013.6761575.
- [9] W. Yin, K. Kann, M. Yu, and H. Schütze, 'Comparative Study of CNN and RNN for Natural Language Processing', *arXiv:1702.01923 [cs]*, Feb. 2017, Accessed: Aug. 24, 2020. [Online]. Available: <http://arxiv.org/abs/1702.01923>.
- [10] R. González-Ibáñez, S. Muresan, and N. Wacholder, 'Identifying Sarcasm in Twitter: A Closer Look', p. 6.
- [11] D. K. Tayal, S. Yadav, K. Gupta, B. Rajput, and K. Kumari, 'Polarity detection of sarcastic political tweets', in *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, Mar. 2014, pp. 625–628, doi: 10.1109/IndiaCom.2014.6828037.
- [12] S. Saha, Department of Computer Science, Central University of South Bihar, Patna – 800014, Bihar, India, J. Yadav, Department of Computer Science, Central University of South Bihar, Patna – 800014, Bihar, India, P. Ranjan, and Department of Computer Science, Central University of South Bihar, Patna – 800014, Bihar, India, 'Proposed Approach for Sarcasm Detection in Twitter', *Indian Journal of Science and Technology*, vol. 10, no. 25, pp. 1–8, Jun. 2017, doi: 10.17485/ijst/2017/v10i25/114443.
- [13] S. Poria, E. Cambria, D. Hazarika, and P. Vij, 'A Deeper Look into Sarcastic Tweets Using Deep Convolutional Neural Networks', *arXiv:1610.08815 [cs]*, Jul. 2017, Accessed: May 09, 2020. [Online]. Available: <http://arxiv.org/abs/1610.08815>.
- [14] T. Ptacek, I. Habernal, and J. Hong, 'Sarcasm Detection on Czech and English Twitter', p. 11.
- [15] N. A. Arifuddin, Indrabayu, and I. S. Areni, 'Comparison of Feature Extraction for Sarcasm on Twitter in Bahasa', in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia, Oct. 2019, pp. 1–5, doi: 10.1109/ICIC47613.2019.8985805.
- [16] A. Joshi, P. Bhattacharyya, and M. J. Carman, 'Automatic Sarcasm Detection: A Survey', *ACM Comput. Surv.*, vol. 50, no. 5, pp. 1–22, Nov. 2017, doi: 10.1145/3124420.