Emotion Classification on Indonesian Twitter Dataset

3 authors, including:

Mei Silviana Saputri
University of Indonesia
8 PUBLICATIONS   **46** CITATIONS

SEE PROFILE

Rahmad Mahendra
University of Indonesia
54 PUBLICATIONS   **203** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Recommender Systems for e-Commerce View project

Indonesian Morphological Analysis View project

Emotion Classification on Indonesian Twitter Dataset

Mei Silviana Saputri, Rahmad Mahendra, Mirna Adriani
Faculty of Computer Science
Universitas Indonesia
Depok, Indonesia
mei.silviana@ui.ac.id,{rahmad.mahendra,mirna}@cs.ui.ac.id

Abstract—The rapid growth of Twitter usage attracts many researchers to utilize Twitter data for several purposes, including emotion analysis. However, there is a resource limitation in standard dataset for emotion analysis task for under-resourced language, especially Indonesian. In this study, we build an Indonesian twitter dataset for emotion classification task which is publicly available. In addition, we conduct feature engineering to decide the best feature in emotion classification. The features used in this research are lexicon-based, Bag-of-Words, word embeddings, orthography and Part-Of-Speech (POS) tag features. We test those features in two datasets with different characteristics. F1-score is employed as an evaluation metric. The results of our experiments show that implementing the combination of all proposed features in our built dataset can achieve 69.73% of F1-Score, which outperforms the baseline model by 26.64%.

Keywords-natural language processing; emotion classification; indonesian tweet; feature engineering

I. INTRODUCTION

Social media has become a new trend for people to interact and communicate. Hence, the growth rate of social media users is increasing rapidly over the years. A social media which has the highest user growth is Twitter. The content of the Twitter post, which is called as tweet, has been widely used by researchers, government or industry to gain knowledge which helps them to solve everyday problems. Various actual human behaviors can be captured from tweets. One of the most popular tasks is emotion analysis.

Emotion is an ongoing state of mind, characterized by mental, physical, and behavioral symptoms [1]. People emotion can be identified directly through their facial expression and speech. Automatically detecting emotion is crucial because it can be implemented in various fields. In education, emotion analysis can be utilized for intelligent e-learning environment [2]. Moreover, emotion analysis can be used in the business for identifying customer complaint in email [3]. In nowadays world where the technology has grown rapidly, people also tend to express their emotion through text in a social medias post. In social media data such as Twitter, emotion detection can be beneficial in government to monitor public response regarding policy or political event. Moreover, emotion analysis from social media also can be utilized by companies to monitor public responses about services or product thus help them in deciding the target market. Identifying emotion in Twitter is also challenging because its short text with informal words and unstructured grammar can-

not be handled using normal text processing techniques. Because of its importance, several datasets are created as a benchmark to obtain state-of-the-art techniques for emotion analysis. Those standard datasets mostly used for English emotion task. However, the standard dataset for another language is limited.

Indonesian tweet is potential for emotion analysis study. According to Statista, an online statistics portal, Indonesia is marked as the third largest active Twitter users in the Asia Pacific from 2012 to 2018[1]. It can be inferred that conducting emotion analysis for Indonesian tweet would be beneficial for many purposes. However, there is not any public dataset for emotion analysis in Indonesia. Previous works in Indonesian emotion analysis [4],[5] not publish their dataset for the public. In addition, their datasets are limited in small data dan less variety. Therefore, we construct an Indonesian Twitter dataset for emotion classification task which has various characteristics and available for public.

In addition, we also propose feature engineering to discover the best features for Indonesian emotion classification. Those features include Bag-of-Words, word embeddings, lexicon-based, Part-Of-Speech (POS) tag, and orthographic features. For classifier, there are three methods used: Logistic Regression, Support Vector Machine, and Random Forest. F1-score is utilized as a metric to evaluate the best performance of feature and classifier.

To sum up, our main contributions are:

- We build a dataset for Indonesian emotion classification from Twitter data. This dataset consists of 4.403 tweets which divided into five classes of emotions (love, joy, anger, sadness, fear) and publicly available for research purpose[2].
- We propose feature engineering which recommends the best features to identify emotion in Indonesian tweet.

II. RELATED WORK

The earliest study in emotion mining in text was conducted by Alm et al. [6]. They identified emotion expresses in children fairy tales using Valence and Arousal model. The dataset built in their research has been widely used in emotion analysis study. On the other hand, the initial study

---

of emotion analysis on Twitter data was introduced by Mohammad [7]. They used n-gram and emotion lexicon-based features for detecting the emotion in English tweet based on Ekman's emotion model. Since then, the study of emotion analysis using tweet is increased, both using supervised and unsupervised methods.

Most emotion analysis studies utilize emotion lexicon for classification features. There are several emotion lexicons for English which have been widely used for emotion classification, such as NRC emotion lexicon and WordNet Affect (WNA) lexicon which construct based on Ekman's emotion class. However, there is only one emotion lexicon for Indonesian which was developed by Shaver [8] based on Shaver's emotion definition [9]. Therefore, the study of emotion analysis in Indonesia mostly uses n-gram based feature instead of lexicon-based. Early research on Indonesian emotion analysis on tweet data was conducted by Arifin et. al [5]. They use Non-negative Matrix Factorization, an extension of TF-IDF model, to classify emotion in tweets. TF-IDF based features also used by [10] to classify emotion in Indonesian tweet. On the other hand, The et al. [4] used more various features for detecting emotion in Indonesian tweet, including n-gram, linguistic, sentiment lexicon, and orthographic features. They used Shaver's emotion word list as query filters in data collection thus their dataset consists of explicit emotion only. However, all experiments in Indonesian emotion analysis are conducted using their own dataset because there is no standard dataset for Indonesian emotion classification which publicly available.

In recent years, word embeddings dominantly used as a feature for emotion classification. Word embedding features for English emotion detection has been implemented by Heriz et al. [11]. They compared the use of basic Bag-of-words (BOW) features and word embeddings (Word2Vec and Glove). The results of their experiment show that combining basic BOW features and word embeddings can improve the performance. Word embeddings for tweet emotion classification also used by Vora et al. [12]. Using Random Forest, their model can achieve 91% precision for four classes of emotion in English tweet. However, word embeddings have not been yet utilized for Indonesia emotion classification task.

III. Methodology

There are two main processes conducted in this study: dataset building and emotion classification.

A. Dataset Building

Our goal is to create a tweet emotion dataset for Indonesian language. This process consists of two steps: data collection and data annotation.

1) Data Collection: We collected tweet using Twitter Streaming API for about 2 weeks, starting from June 1, 2018 until June 14, 2018. Indonesian geolocation coordinate was used to filter streaming tweet. We did not utilize emotion words list as a query filter because we want to minimize the bias caused by the emotion words list. Emotion words list would be utilized as a feature for classification, while the majority of emotion words list is correspondence to the emotion class explicitly. Furthermore, we focus to define emotion in personal tweets. Therefore, we exclude the tweet from news portal and government office. Commercial promotion tweets also eliminated from our collection.

2) Data Annotation: We use Shaver's basic emotion [9] which later popularized by Parrot as Parrot's basic emotion [1] to define the emotion class of our dataset. This emotion theory consists of six classes: love, joy, surprise, anger, sadness, and fear. Shaver and Murdaya have defined the structure of the emotion lexicon in Indonesia based on this emotion theory but without surprise class [8]. Thus, there are five classes defined for Indonesian emotion lexicon: love, joy, anger, sadness, and fear. We use the definition of emotion classes in their work to create annotation guideline.

There are two annotators employed for data annotation. First, each annotator is asked to filter out the tweets which not contain any emotion. After that, annotators asked to annotate tweets with one or multi-label scheme. For one label annotation, there are five emotion labels have to be chosen: love, joy, anger, sadness, and fear. However, because there is a possibility of tweet contains more than one emotion, we introduce a multi-label scheme. The example of a tweet which categorized as multi-label is as follows:

> Kesel banget jauh jauh ke Kuta buat beli clay mask di Guardian, tapi pas nyampe sana stoknya kosong. Karna marah, trus malah ke em em jus beli es teler durian. Nambah siomay ama nasgor ikan asin. Perut kenyang hati senang.
> (*It's very annoying to go far to Kuta to buy clay mask in the Guardian, but the stock is empty. Out of anger, I went to em em juice to buy durian ice. Also, dumplings and salted fish friend rice. My stomach is full, and I am happy*)

This tweet is categorized as a multi-label emotion because it expresses two emotion classes, i.e expresses anger because of the emptiness of a product as well as happy in the end because of food. However, multi-label emotion tweets are not considered in this study. We evaluate the agreement of five classes of emotion using Cohen Kappa measurement.

B. Emotion Classification

The result of dataset building is used for emotion classification. The steps conducted in this phase are as follows:

1) Pre-processing: Twitter limits maximum characters in a tweet into 280 characters. However, there are no writing rules for people who want to tweet. Thus, the data collected is unstructured and need to be cleaned. The pre-processing conducted in this research consists of several steps, including:

- Data Normalization
  We change all characters into lowercase to increase

the possibility of same words with different folding counted as one. However, we also keep the original dataset in a separate file because we can use the information of orthography (such as capital characters and punctuations) for detecting emotion. Moreover, we implement some removals, such as username, hyperlink and stop words. Furthermore, we utilize a manually created Indonesian typography dictionary to normalize abbreviation and misspelling in tweets.

- Part-Of-Speech (POS) Tagging
  We tag the token in each tweet with Part-Of-Speech (POS) definition from Dinakaramani et. al. [13].
- Stemming
  For each token in each tweet, we apply stemming to get root word. This process aims to increase the probability of similar words with similar root word are counted as one.

2) Feature Extraction: Feature extraction is a process of transforming data from text into numerical features for machine learning process. In this study, several features are implemented, including:

- Basic Features
  - Indonesian Emotion Words List [8]
    There are 94 emotion words which are divided into five main classes, i.e love, joy, anger, sadness, and fear. We use this feature as a baseline feature.
  - Bag-of-Words (BoW)
    In a BoW model, the frequency of each unique word in the text is used as features for classification.
  - Word Embeddings
    Word embedding is a technique for mapping semantic meaning into a geometric space. For example, the word "happy" and "joyful" are expected to have similar vector representation. Hence, this model is expected can improve the performance of emotion classification.
    In this study, we utilize two models of word embeddings model: Word2Vec [14] and FastText [15]. In this experiment, we train 1.026.483 Indonesian tweets into Word2Vec and FastText representation with several variations of dimension. We randomly set the dimension into 100, 200, 300, 400, 500 and choose the dimension with the highest performance as features for classification.
- Lexicon
  - Vania's Sentiment Lexicon [16]
    This lexicon consists of the list of 415 positive words and 581 negative words. For each tweet, the number of positive and negative terms are counted and used as features for classification.
  - InSet Sentiment Lexicon [17]
    InSet lexicon is developed from Indonesian microblog. This lexicon consists of 3.609 positive words and 6.609 negative words list along with

its sentiment score. For each tweet, we sum up the sentiment score of positive and negative terms for classification feature.
  - Emoticon List
    Emotion class in a tweet can be detected by using emoticons, such as :(expresses sadness and :) expresses joy. We manually create this list of emoticons.
- Additional Features
  - Part-Of-Speech (POS) Tag
    Part-of-Speech (POS) tag indicates the class of a word in a sentence, for example, proper noun, verb, adjective, noun, and negation. In this research, we count the occurrence of adjective and negation words in a tweet as features. Emotion words usually indicated by adjective words. Furthermore, the use of negation, such as "not" is crucial because it can show the opposite emotion.
  - Orthographic
    Orthographic is a set of rules explains the using of capitalization, punctuation, hyphenation, and spelling in writing. In emotion analysis, especially in Indonesian text, orthographic is important because the using of capitalization, exclamation and question mark express a specific emotion. In this study, we count the number of capitalization and exclamation in every tweet as features. In addition, we count the number of characters in a tweet as well as total words in a tweet as features.
- Combined Features
  After knowing the performance of each individual features, we combine the features which have good performance. Feature fusion is conducted by combining individuals feature vectors into one combined vector.

3) Learning: In this phase, the labeled training features are used as input for the Machine Learning classifier. This learning process generates a model which used for testing unlabeled features. We use 10-fold K-Fold cross-validation to split training and testing data. There are three classifiers used in the experiment: Logistic Regression, Linear Support Vector Machine, and Random Forest. Logistic Regression and Support Vector Machine have been implemented in multiclass text classification on Indonesian tweet [4]. On the other hand, Random Forest is an ensemble method that using decision tree as the weak learner. Multiple learners are used in Random Forest in order to improve the generalizability of individual learning. Therefore, we include the Random Forest in the experiment to test the effectivity of using ensemble learner.

4) Evaluation: We use macro F1-score to measure the performance of classifiers and features used. F1-score is a weighted harmonic mean of precision and recall. For features evaluation, we measure the performances of the individual feature as well as the combination of different features. In addition, we present the evaluation of best features by showing the precision, recall, and F1-score in

each emotion class as well as its confusion matrix.

IV. Dataset Characteristics

Our dataset is built based on manual annotation by two annotators. There are 7.500 tweets that should be annotated by annotators. After annotation, the proportion of five basic emotion class, no-emotion, and multi-label emotion are 64%, 32%, and 4% respectively.

In this study, we consider to focus on five basic emotion classes. To measure the quality of annotation, we calculate the Kappa score of five basic emotion classes. The Kappa score of our annotation is 0.917 which considered being very good. The final dataset is taken from the dataset with the agreed label, which consists of 4.403 tweets. The distribution of our dataset is summarized in Figure. 1. Figure 1 shows that there is a balanced number of joy, anger, and sad class. On the hand, the number of love and fear tweet are limited.
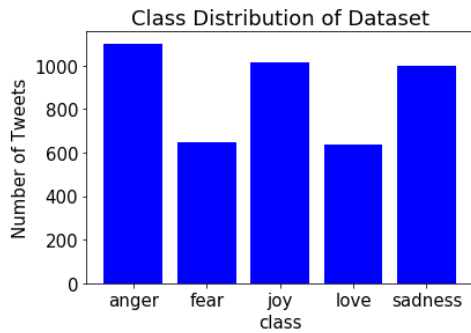

Figure 1. Class Distribution of Dataset

To show the variety of our data, we put on the example of tweets in anger class.
First example:

hari ini libur, rencananya mau nonton Jurassic World, tapi kayanya gajadi deh mengingat kondisi yg gak fit bgt ini **sebel**. Rusak rencana sebelanga.. **sebel** akutu
*(Today is holiday, I am going to watch Jurassic World, but maybe it should be canceled because I am extremely not fit. **annoying**. What a broke plan. I am **annoyed**.)*

Second example:

Ini aja membuktikan anda sudah TIDAK BENAR....!!! MASA NAPI KORUPTOR BISA PUNYA HP DI PENJARA ITU SDH MELANGGAR ATURAN.... DAN ANDA DG ENAKNYA MELANGGAR ATURAN...!! INI MENANDAKAN BAHWA ITULAH KARAKTER ANDA.
(It proves that you are NOT TRUE!!! HOW CAN THE CORUPTOR CONVICT HAVE A HAND PHONE IN THE PRISON THAT HAVE BEEN BREAKING THE RULES ... AND YOU ENJOY BREAK THE RULES..!! THIS INDICATES THAT'S YOUR CHARACTER)

The first example contains emotion word, i.e. annoying, hence anger emotion can be indicated explicitly. On the other hand, the second example does not contain any emotion words, but we can identify this tweet as anger because of capitalized characters and exclamation mark. This kind of implicit emotion can be captured in our dataset because we do not use emotion words list on the data collection process. This characteristic is different from another Indonesian tweet dataset which commonly contains explicit emotion only.

V. Experiment and Result

We implement our proposed features which have been described in Section III to our built dataset. In addition, we also applied our proposed features into Indonesian tweet dataset from J. E. The et. al. [4] for comparison. Their dataset consists of 942 tweet which has similar emotion classes but has different characteristics from ours. Their dataset has explicit emotion because it was build based on emotion words list. On the other hand, our dataset has more variety of data as mentioned in Section IV. We compare the contribution of different features in different Machine Learning classifier for both datasets. We implement several individual features as mentioned in Section III as well as the combination of those individual features.

The results of our experiment are summarized in TABLE I. We examine the use of different individual features and the combined features. The results show that the use of emotion words list as our baseline feature performs better on The's dataset which contains emotion words explicitly. This feature achieves 57.85% on F1-score when Logistic Regression applied. On the other hand, the highest F1-score for this baseline feature on our new built dataset is 43.09%. The use of emotion word list is not enough to capture the emotion expressed in our dataset due to the variety of our data.

Other individual features are Bag-of-Words and word embeddings. The use of Bag-of-Words can boost performance on both datasets. For word embeddings features, we compare the use of Word2Vec and FastText features. In general, FastText obtain better score both in two datasets although Word2Vec perform better on The's dataset when Logistic Regression applied. The great result obtained when we combine the emotion word list, Bag-of-Words, and FastText features.

For the lexicon-based feature, InSet sentiment lexicon, get the best F1-Score compared to Vania's lexicon and emoticon list. Vania's lexicon contains formal words while InSet lexicon is developed using Twitter data thus it more suitable for our task. However, there is a slight difference of F1-score obtained from emoticon lexicon feature in both datasets. Combining Vania's lexicon, InSet Lexicon and emotion list obtain slightly higher F1-score than the result of individual InSet. In addition, we examine the effect of combine emotion words list, Vania's Lexicon, InSet lexicon, and emoticon list for feature combination. The result shows that the combination of these features achieve

Features	The's Dataset [4]			New Dataset		
	LR	SVM	RF	LR	SVM	RF
Basic Features						
Emotion word's list (EW) [8]	57.85%	58.34%	56.04%	43.09%	43.12%	42.20%
Bag-of-Words (BOW)	69.53%	67.52%	67.31%	65.13%	57.88%	64.07%
Word2Vec (WV)	67.32%	53.96%	40.32%	61.83%	61.37%	53.03%
FastText (FT)	66.46%	65.42%	55.01%	62.49%	62.27%	55.18%
EW + BOW + WV	70.34%	64.77%	62.52%	68.25%	61.20%	59.60%
EW + BOW + FT (Basic)	73.72%	71.46%	64.53%	68.39%	61.58%	62.23%
Lexicon						
Vania's Sentimen (VSent) [16]	11.02%	11.59%	11.71%	8.15%	8.79%	9.59%
InSet Sentimen (ISent) [17]	22.34%	19.05%	28.17%	19.36%	14.78%	24.92%
Emoticon List (Emot)	11.56%	11.53%	13.05%	11.45%	11.52%	11.48%
VSent + ISent + Emot (Lex)	23.80%	19.16%	28.85%	22.48%	15.37%	25.87%
EW + Lex	62.91%	62.19%	55.78%	50.30%	37.08%	46.56%
Other Features						
Orthographic (Ort)	20.39%	16.19%	26.62%	21.16%	9.25%	25.31%
Pos Tag (POS)	16.87%	16.91%	17.37%	16.79%	17.14%	17.84%
Ort + POS	27.93%	16.93%	30.66%	22.98%	15.00%	25.36%
Features Combination						
Basic + Lex	74.60%	68.43%	65.94%	69.43%	63.06%	62.30%
Basic + Ort + POS	74.60%	69.90%	66.90%	69.22%	57.99%	62.72%
Basic + Lex + POS + Ort	**75.98%**	70.84%	66.,60%	**69.73%**	64.24%	63.06%

better performance compare to emotion word list only.

To boost the performance of emotion classification model, we also examine the use of POS tag and orthographic features. The results show that both features not perform well as individual feature, but shows better performances when combined.

For increasing the F1-score, we consider implementing several feature combination scenarios. We take the best feature for each feature group and combine those features. Based on the results in TABLE I, it can be inferred that the most significant features are formed based on the combination of Emotion Words List, Bag-of-Words and FastText. This combination achieve 73.72% of F1-Score in The's dataset and 68.39% in our new dataset. Adding lexicon and additional features (orthographic and POS tag) to the combination of basic features can increase the F1-Score. Both The's dataset and our new dataset achieve the highest F1-score when the combination of basic (emotion word list, Bag-of-Words, FastText), Lex (Vania's lexicon, InSet lexicon, emoticon list), orthography and POS tag features used in the Logistic Regression model. This combination achieves 75.98% of F1-Score on The's dataset and 69.73% of F1-Score on our new dataset.

Regarding the classifier model, Logistic Regression performs the best in almost scenarios, followed by Support Vector Machine and Random Forest. In general, our proposed feature combination can boost performance in both datasets. The implementation of our proposed features to The's dataset can achieve 75.98% F1-score which is better

TABLE II. EVALUATION OF EACH EMOTION CLASS ON OUR NEW DATASET

Class	Precision	Recall	F1-Score
love	64%	75%	69%
joy	81%	60%	69%
anger	61%	81%	70%
sadness	89%	72%	80%
fear	65%	53%	59%
avg/total	**70%**	**68%**	**68%**

compared to the result of The et. al. implementation [4] with the same dataset with 71.96% accuracy. On the other hand, the implementation of our combined features on our new dataset achieve 69.73%, which outperforms the baseline by 26,64%. Due to the variety and complexity of our new dataset, which consists of explicit and implicit emotion, the learning model cannot perform better than the implementation in The's dataset, which contains explicit emotion only.

We present the detail evaluation of each emotion class of our new built dataset in TABLE II. The best-combined features and Logistic Regression classifier are used in this evaluation. TABLE II shows that a balanced score of precision and recall is achieved by sadness class. Sadness class obtains the best evaluation in precision, i.e. 89%. It means that there is only 11% false positive for sadness label. Recall score for sadness class is also quite high, i.e 72%. On the other hand, joy class achieves high precision but low recall. There is 40% of joy class is predicted as false negative. In contrast, anger class obtains low precision but high recall. The lowest score of precision and recall is obtained from fear class. The limited number of samples in fear class impacts to its classification performance.
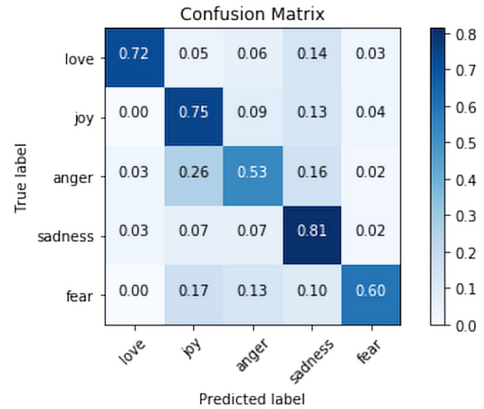

Figure 2. Confusion Matrix of Testing Label

In addition, we present the confusion matrix of testing classes in Figure 2 to understand the errors being made by our model. Based on Figure. 2, it can be understood that sadness class has the highest true positive value. Moreover, sadness class has a small proportion of false positive value and moderate proportion of false negative value. Therefore, sadness class achieves high precision but moderate recall. However, the significant error is found in the number of anger class which predicted as joy class.

In general, the performance of classification model on our new dataset can be considered as moderate and can be used as a benchmark model for detecting implicit emotion on Indonesian tweet.

VI. Conclusion

In this study, we have constructed a dataset for Indonesian emotion classification from Twitter which publicly availabe. The Kappa score for data annotation agreement is 0.917 which considered being very good. This dataset consists of five emotion classess (anger, fear, joy, love, sadness) which contains both explicit and implicit emotion.

In addition, we have proposed several features for emotion classification. Based on the experiment results, it can be inferred that the most significate features are formed based on the combination of emotion word list, Bag-of-Words and FastText. Adding sentiment and emoticon lexicon, orthography, as well as POS tag features to the basic combination features can boost the performance to our built dataset. The highest F1-score achieved by this feature combination is 69.73%, which outperform by 26.64% from the baseline.

For the future works, the problem of multi-label emotion classification which was excluded in the current study is interesting to be examined. Besides emotion classification, measuring the emotion degrees is also important to understand the intensity of emotion. In addition, we can consider building a larger dataset which can be constructed automatically using semi-supervised approach.

References

[1] W. G. Parrott, Ed., *Emotions in social psychology: Essential readings.* New York, NY, US: Psychology Press, 2001.

[2] T. Daouas and H. Lejmi, "Emotions recognition in an intelligent elearning environment," *Interactive Learning Environments*, vol. 0, no. 0, pp. 1–19, 2018.

[3] N. Gupta, M. Gilbert, and G. Di Fabbrizio, "Emotion detection in email customer care," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text.* Association for Computational Linguistics, 2010, pp. 10–16.

[4] J. E. The, A. F. Wicaksono, and M. Adriani, "A two-stage emotion detection on indonesian tweets," in *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct 2015, pp. 143–146.

[5] A. Zainal Arifin, Y. Arum Sari, E. Kamilah Ratnasari, and S. Mutrofin, "Emotion Detection of Tweets in Indonesian Language using Non-Negative Matrix Factorization," *I.J. Intelligent Systems and Applications Intelligent Systems and Applications*, vol. 09, no. 09, pp. 54–61, 2014.

[6] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, no. October, pp. 579–586, 2005.

[7] S. M. Mohammad, "#emotional tweets," in *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, ser. SemEval '12. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012, pp. 246–255.

[8] P. R. Shaver, U. Murdaya, and R. C. Fraley, "Structure of the Indonesian Emotion Lexicon," *Asian Journal of Social Psychology*, vol. 4, no. 3, pp. 201–224, 2001.

[9] P. Shaver, J. Schwartz, D. Kirson, and C. O'Connor, "Emotion Knowledge: Further Exploration of a Prototype Approach," *Journal of Personality and Social Psychology*, 1987.

[10] A. R. Atmadja and A. Purwarianti, "Comparison on the rule based method and statistical based method on emotion classification for Indonesian Twitter text," *2015 International Conference on Information Technology Systems and Innovation, ICITSI 2015 - Proceedings*, 2016.

[11] D. K. Jonathan Heriz, Michal Shmueli-Scchuer, "Emotion Detection from Text via Ensemble Classification Using Word Embeddings," in *he 2017 ACM SIGIR International Conference on the Theory of Information Retrieval*, 2017, pp. 1–6.

[12] P. Vora, M. Khara, and K. Kelkar, "Classification of Tweets based on Emotions using Word Embedding and Random Forest Classifiers," *International Journal of Computer Applications*, vol. 178, no. 3, pp. 1–7, 2017.

[13] A. Dinakaramani, F. Rashel, A. Luthfi, and R. Manurung, "Designing an indonesian part of speech tagset and manually tagged indonesian corpus," in *2014 International Conference on Asian Language Processing, IALP 2014, Kuching, Malaysia, October 20-22, 2014.* IEEE, 2014, pp. 66–69.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, vol. abs/1301.3781, 2013.

[15] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[16] C. Vania, M. Ibrahim, and M. Adriani, "Sentiment Lexicon Generation for an Under-Resourced Language," *International Journal of Computational Linguistics and Applications*, vol. 5, no. 1, pp. 59–72, 2014.

[17] F. Koto, "InSet Lexicon : Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 391–394.