Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 2 (2023) 254 - 262 ISSN Media Electronic: 2580-0760

A Hybrid Method on Emotion Detection for Indonesian Tweets of COVID-19

Adi Surya Suwardi Ansyah[1], Arya Putra Kurniawan[2], Asiyah Nur Kholifah[3], Diana Purwitasari[4*]
[1,2,3,4]Informatics Engineering, Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember
[1]adisur.sa@gmail.com, [2]aryapkurniawan15@gmail.com, [3]asiyahnurk@gmail.com, [4]diana@if.its.ac.id

Abstract

As a result of the COVID-19 pandemic, there have been restrictions on activities outside the home which has caused people to interact more and express their emotions through social media platforms, one of which is Twitter. Previous studies on emotion classification used only one feature extraction, namely the lexicon based or word embedding. Feature extraction using the emotion lexicon has the advantage of recognizing emotional words in a sentence while feature extraction using word embedding has the advantage of recognizing the semantic meaning. Therefore, the main contribution to this research is to use two lexicon feature extraction and word embedding to classify emotions. The classification technique used in this research is the Ensemble Voting Classifier by selecting the two best classifiers to try on both types of feature extraction. The experimental results for both types of feature extraction are the same, indicating that the best classifiers are Random Forest and SVM. Models using both types of feature extraction show increased accuracy compared to using only one feature extraction. The results of this emotional analysis can be used to determine the public's reaction to an event, product, or public policy.

Keywords: emotion detection; ensemble; hybrid; indonesian tweets; lexicon

1. Introduction

Due to the COVID-19 pandemic, people tend to express their emotions on Twitter, such as happiness, love, fear, anger, and sadness. The results of the emotional analysis of the text can be used to determine the public's response to an event, product, or public policy. Therefore, research on emotion detection is growing and has become an important research area used to understand human emotions. Emotion is recognizable from various input types, such as speech, facial expressions, and text. It can be detected in text in song lyrics [1], blogs [2], social media [3–5], etc.

To process data in text, it is necessary to convert it into a numerical form. Several existing models are conceived for projecting documents in vector space. These models are based on the frequency of words appearing in sentences (i.e., TF-IDF), semantic relationship (i.e., GloVe), or contextually (i.e., BERT). However, these models rely on the distributional hypothesis, which states that words in the same context have similar meanings. This conjecture assumes closer or similar word representations in vector space for semantically equivalent word pairs. Sad and happy emotions will have exact representations, even though they have opposite meanings. Thus, the use of the representation of the word embedding model allows for failure to identify emotions. Therefore, a model representation enriched with emotion and sentiment is needed. The enhanced methods can be concentrated into four categories: keyword-based, lexicon-based, learning-based, and hybrid [6].

Keyword-based detects existing emotions by matching text words with emotional keywords representing specific categories of emotion. For example, in a study [7], proverbs, keywords, short forms of expression, and emoticons were matched to each feeling they represented. Aphorisms such as "A Day of sorrow is longer than a month of joy" will be marked as a sad emotion in this study. There is also a list of tuples, such as ":-E" for the angry feeling and the abbreviation "ih8u" for "I hate you" and for the emotion of hurt. The method used in that study tends to be rarely used because there still needs to be an equivalent evaluation measure.

Lexicon-based is the most widely used method in emotion detection. This is because understanding lexical meaning can explain phenomena facing problems related to inflexibility and lack of predictive ability. A lexicon is a group of words labeled with emotional categories or dimensions. Single-word

Accepted: 01-01-2023 | Received in revised: 05-02-2023 | Published: 15-03-2023

weights can be found when searching for an item and finding its score from the lexicon. The emotional score is calculated from the total weight of each constituent word in the text. WordNet-Affect and Affective Norms for English Words (ANEW) is an example of the lexicon of emotions used in English. But this method has language limitations when used to detect emotions in other languages, mainly Indonesian.

The number of Indonesian language lexicons that can be accessed could be higher. All the words need to be adequately represented. The lexicon also has limitations when used in emotion domains such as social media because the slang word conversion process used in social media must be carried out into the standard language as contained in the lexicon [8].

Meanwhile, learning–based methods often used for supervised emotion detection are naïve Bayes algorithms, decision trees, and support vector machines [9]. For the supervised method, data with emotions are used for training and testing with a supervised classifier. Unsupervised machine learning methods can also detect emotion that has not been labeled emotion as an example of emotion detection from the YouTube platform comment column in [10]. To calculate the semantics between the words of a sentence with emotional labels, Pointwise Mutual Information (PMI) parameter measurements are used. Measures were calculated based on the co-occurrence between the terms to classify and the representative words sourced from the corpus or dataset. However, this machine learning method is not appropriate for large-scale training datasets because it takes a long time to learn.

The last one, detecting emotions from text, is the hybrid method, a combination of different approaches [11]. The processes carried out in previous studies have proven that the hybrid method produces the best results because it combines several ways that complement one another's weaknesses.

Several studies have used text mining and word embedding [1], [12], [13]. The traditional word embedding model still needs to be more accurate when applied directly to analyzing sentiments and emotions because the main problem of the learning word insertion algorithm is that it can only model the word context without involving feeling or emotional information in the text [14].

In the study [15], they have proposed using lexicon and word embedding in semantic analysis. After carrying out various experimental scenarios, the accuracy results are above 80% for the multiple datasets trained in these scenarios. Then, [16] conducted research by comparing three models: a lexicon-based feature model, a word embedding–based model, and a hybrid model comprising a combination of lexicon and word embedding. The results of this study, the hybrid model, are the best carried out with f-score results above 80%. However, the domain of this research is stress classification. No research uses this method to detect emotions from a text, mainly in Indonesian tweets.

Therefore, this study proposes the contribution of a new hybrid approach to detecting emotions. This research uses the lexicon with polarized words, emotional context, and the word embedding approach. Word representation in the dataset is also considered using Word2Vec and FastText. Meanwhile, to extract the emotional context, SenticNet and NRC Emotion Lexicon (EmoLex) are used, which are two Indonesian emotion lexicons that are publicly available.

The lexicon expansion is also carried out on the lexicon which gives the best results in the classification model. The lexicon is expanded by adding words from the dataset that have been manually labeled. The results of these feature extractions will each form a classification model from the six predetermined classifiers. Then the best model of word representation and emotion feature extraction will be ensembled with a voting classifier.

Furthermore, these models will be assembled using the soft voting rules classifier to form a new model. This method is implemented into an emotional dataset derived from tweets in the Indonesian language during the first three months of the COVID-19 outbreak in the world. The aim is to determine how well the hybrid method detects the emotions of the Indonesian people before the pandemic was officially announced in Indonesia in the dataset.

Finally, the paper is structured as follows: Section 2 presents the research methodology undertaken in this study and the implementation of the datasets used. The experimental results will be presented in Section 3, which is intended to evaluate the emotion detection model from three scenarios: (1) word insertion model, (2) lexicon-based model along with corpus-based lexicon extension, and (3) hybrid model that assembles a classification model. Best of word embedding and lexicon. The paper ends with a conclusion in Section 4, where an outline of this research is summarized, and possible future work that could be undertaken for further study.

2. Research Methods

Figure 1 illustrates our contribution framework for emotion analysis on the Twitter post dataset. The proposed framework can become a solution in determining a responsive action based on public opinion during an emergency disaster such as the COVID-19 pandemic. This paper focuses on research using TF-IDF with the lexicons, Word2Vec, and FastText when building the model. It compares the efficacy result of the proposed framework with the standard classification method.

DOI: https://doi.org/10.29207/resti.v7i2.4816

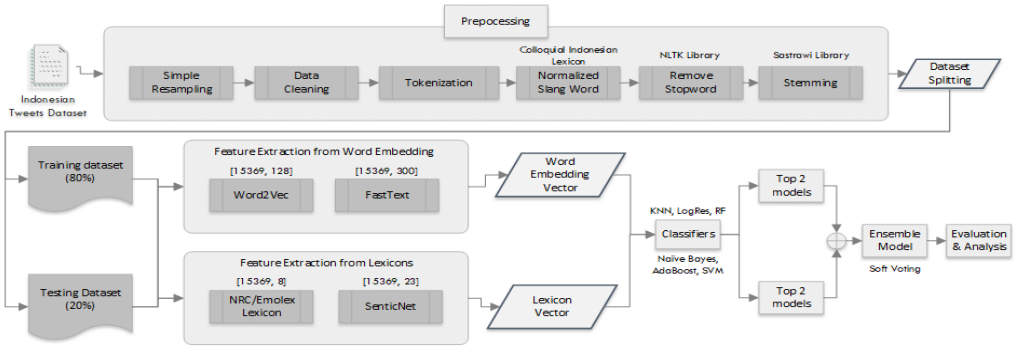

Figure 1. The Hybrid Method Using Word Embedding and Lexicon for Emotion Detection

2.1 Dataset

The data set in this research area [3] was crawled using the Twitter API and manually labeled by experts with emotions represented from actual tweets. This research will use a database of Indonesian Twitter posts before COVID-19 was declared a pandemic in Indonesia.

Previously, the study used the dataset to analyze the Indonesian Twitter user's behavior changes caused by the COVID-19 pandemic. The dataset was collected between 13 December 2019 and 13 March 2020. The method for collecting this dataset was by crawling the area around Setia Budi in the city of Jakarta using Twitter API combined with the Twint Library method to acquire many tweets in a short time. Tweets result from crawling using this criterion and method, averaging around 100 words per tweet.

There are 23,647 data on the dataset obtained by crawling on Twitter during the selected timeline. The data is labeled with one of the emotions expressed in the tweets: happiness, sadness, fear, love, and anger. The ratio of the label for each tweet is 9,939 tweets are labeled as happy; 3,965 tweets are ed as sadness; 3,354 tweets are led as fearful; 3,306 tweets are tagged as love; and finally, 3,083 are labeled as anger. Table 1 displays a sample of tweets from the dataset.

2.2 Preprocessing

The next step is preprocessing the dataset used in this study. Imbalanced classes are handled. In this scenario, on the Before COVID-19 Pandemic dataset, there is a severe imbalance in the happy category. Therefore, the data were reduced to equalize the data in other classes. After balancing, the happy class data were reduced to 4,039 records.

Cleaning is achieved by removing links within tweets, usernames, excess spaces, hashtags, RT words, punctuation, and numbers and changing the text to lowercase. This process is essential for this research because it will produce cleaner and better data, which can affect the model's performance [14].

Next, the text data are preprocessing processes such as tokenization, slang word normalization, and stemming. Normalizing slang words in Indonesian texts use the Colloquial Indonesian Lexicon, which contains 3,592 slang dictionaries [17].

This colloquial lexicon is used to normalize the Indonesian slang word "Bahasa alay," which was built from Instagram comments and manually annotated. For example, the abbreviation 'k' in chat or social media comments can mean two formal words in Indonesian: okay and 'kak.' This colloquial lexicon translates not only abbreviations but also assimilation (e.g., '*koq*' means '*kok*' in formal), vocal modification (e.g., '*sampe*' means '*sampai*'), naturalization (e.g., "*hepi*" from happy), clipping (e.g., '*liat*' from '*lihat*'), metathesis (e.g., '*sabi*' means '*bisa*'), and reversal (e.g., '*ucul*' from '*lucu*').

Furthermore, the process of removing stopwords using the NLTK library and stemming words with the Sastrawi library. The result is 15,369 clean words. The information is ready to be split into 80:20 for training and testing.

Table 1 shows a sample of words from the corpus or dataset used, along with the tweets translated into English and the cleaning results. In this table, tweets that have gone through all the cleaning processes are marked in italic format and colored red for tweets in the Indonesian language. At the same time, the sentences in blue are selected words that have been translated into English.

2.3 Data Training Model

Three models were formulated for solving the problem in this research, namely, the Word Embedding Model, Lexicon Model, and Ensemble Model. These were a combination of the previous models that will become the focus of the study. In building the word embedding model, Word2Vec and FastText were used.

DOI: https://doi.org/10.29207/resti.v7i2.4816

Table 1. Sample Tweets and After Cleaning

Emotion Expressed	Amount	Sample Tweet and Cleaning Tweet	Emotion Detected from NRC Lexicon
Happy	Before: 9939 After: 4039	Dulu *principal kerja* di *kantor* lama, *jam kerja* bener2 *kerja*, gak *main hp*, gak *sebat dll* . *Istirahat* mrk *beneran istirahat*, *makan*, *rokok*, *kopi*. *Jam pulang ya beberes* trus balik *penginapan*. Ada masalah atau *trouble dibicarain pas makan malem cari solusinya* bua t *besok* In the past, the *principal worked* in the old *office*, the *working hours* were really *workin g*, I didn't *use my cellphone*, I didn't *smoke*, etc. Their *rest* is *really resting*, *eating*, *smo king*, and *coffee*. It's *time* to *go home*, *get ready* then go back to the *inn*. There is a prob lem or *trouble discussed over dinner looking for* a *solution* for *tomorrow*.	Joy, Fear
Sadness	3965	*semalem iseng liat yutubnya* @ntsana_ aku ~~nangis~~ *menangis* sampe ~~gabisa~~ *nggak bisa* ~~nahan~~ *menahan* ~~lg,sakit~~ *lgsakit* rasanya :((((tp setelah itu aku *belajar* bahwa *Tuhan* sedang mempersiapkan yg ~~terbaik,mungkin~~ *terbaikmungkin* dia bukan *jodoh ku hehe*. *last night just for fun watching* @ntsana_'s *YouTube*, I *cried* until I *couldn't hold* it *anymore*, it *hurts* :((((but after that I *learned* that *God* is preparing *the best*, *maybe* he's not *my soul mate hehe*.	Anger, Fear, Sadness, Joy
Fear	3354	@martintambunan Kalau saya *jelasin* nanti *dibilang hoax*, *mending* tanya *langsung* ke pak @msaid_didu *seluk beluk tol* yg tak *layak dibangun* tapi *dibeli mahal* oleh *pemerintah era jokowi*.. *Btw*, *kalo transaksi jual beli* itu ada *feenya* nggak *ya*?ئ™ش™ش™ن™ˆ https://t.co/GrcqJNfs8l @martintambunan If I *explain* later that it will be *called* a *hoax*, it's *better* to ask Mr. @msaid_didu *directly* about the *details* of the *toll road* that was not *feasible to build* but was *bought expensively* by the *government* during the *Jokowi era*. *Btw*, *is there a fee* for *buying* and *selling transactions*? ˆ https://t.co/GrcqJNfs8l	Joy
Love	3306	@theMiniNino Aku sbg *big fan*, *diistora* udah *kayak kambing* ~~Conge~~ *conge* Di ~~Istora~~ *istora* pada ~~ngejar2~~ *ngejar atlitnya*, ~~klo~~ *kalo* aku cm bs ~~ngeliatin~~ *melihati* mereka نکً~~.~~ Krn ~~Caro~~ *caro* ~~lgsg~~ *langsung cabut* dr ~~Istora~~ *istora* نکً~~.~~ Sampe pernah dong nanya *satpam* "Pak, ~~Carolina Marin~~ *carolina marin* msh ada di dalam gak?"نکً~~.~~ @theMiniNino I'm a *big fan*, *at istora* I'm already *like* a *goat* ~~Conge~~ *conge* At ~~Istora~~ *istora chasing athletes*, I just could *see* them نکً Because ~~Caro~~ *caro immediately left* ~~Istora~~ *istora* نکً Until I asked the *security guard* "Sir, Is ~~Carolina Marin~~ *carolina marin* still inside or not?" نکً	Joy
Anger	3083	*Lo* bisa kok ~~nyalahin~~ *menyalahkan pemimpin lo* soal *banjir* tapi tetep: 1. ~~Bisa~~ *bisa buang sampang* pada *tempatnya*, 2. ~~Bisa~~ *bisa belajar* soal *sejarah* ~~Jakarta~~ *jaka a*, 3. ~~Bisa~~ *bisa belajar hukum fisika tentang air*, 4. ~~Bisa~~ *bisa paham kalo banjir kali ini emang yang terburuk dalam satu dekade terakhir*, 5. ~~Bisa~~ *bisa gak*!!!??? *You* can *blame your leaders* for *floods* but still: 1. Can *throw garbage* in *its place*, 2. Can *learn about* the *history* of *Jakarta*, 3. Can *learn* the *laws of physics about water*, 4. Can *understand that this flood is the worst* in *a last decade*, 5. *Can you*!!!???	Anger, Fear, Joy, Sadness

Then the Lexicon Model was formulated using two existing Indonesian language emotion lexicons, SenticNet, and NRC Lexicon (EmoLex). The best-performing word embedding and lexicon will be combined to build the ensemble model by comparing which scenario model has the best result. It is possible to compare the best from the models that have been made.

2.4 Word Embedding Model

In Natural Language Processing (NLP), since computers cannot process words like humans, they need to convert natural language into a digital language composed of numbers. To process the data into a vector that can be processed by computer, it is necessary to rely on a text representation library. In this research, Generate Similar (Gensim) library was used for this process. Gensim library's algorithm discovers the semantic structure of a document by analyzing the co-occurrence of patterns within a body of text in the training documents. From this library, two popular algorithms were used for text representation. These algorithms are Word2Vec and FastText.

Word2Vec maps sentences composed of words that the computer thinks are unrelated to one another into a higher dimensional matrix and replaces the semantic relations between these words with the mathematical references in the matrix. The computer can understand the sentences in natural language through mathematics and achieve the effect of making the comments in similar contexts have similar vectors [12]. This algorithm prepared the pre-trained model for the actual scenario model training. In the pre-trained model, Wikipedia's article corpus was fed into the Gensim library's Word2Vec algorithm. The results of this pre-training model vector were used in this study to model the actual Word2Vec algorithm on the COVID-19 dataset and obtained a vector of size (15369, 128).

FastText functions by exploiting subword information and considers the internal structure of words instead of learning word representations. FastText divides words into n-grams rather than using individual words and learns vectors for subparts of words, which are so-called characters of n-grams [18]. The FastText scenario model used pre-trained word vectors for 157 languages. This model is trained on Common Crawl and

DOI: https://doi.org/10.29207/resti.v7i2.4816

Wikipedia. These models were trained using CBOW architecture with position weights, in dimension 300, with character n-grams of length 5, a window of size 5, and 10 negatives. The trained model is loaded into the Gensim library's FastText algorithm, and a vector of size (15369, 300) is obtained.

Table 22. SenticNet Emotion and Word Count

Emotion	Number of Words	Description
Acceptance	847	Emotion related to the consent of something offered.
Anger	1,355	Intense feelings of annoyance, displeasure, or hostility
Annoyance	1,329	The feeling of being annoyed or a nuisance
Anxiety	2,568	The feeling of worry, nervousness, or unease
Bliss	2,135	State of a perfect happiness
Calmness	688	The feeling of being free from agitation or strong emotion
Contentment	6,247	A state of being fulfilled
Delight	1,210	The feeling of a great pleasure
Disgust	702	A feeling of revulsion or strong disapproval aroused by something unpleasant or offensive
Dislike	1,180	The feeling of distaste or hostility
Eagerness	2,133	Enthusiasm to do or to have something
Enthusiasm	12,380	Intense and eager enjoyment, interest, or approval
Fear	388	Afraid of (someone or something) as likely to be dangerous, painful, or threatening
Grief	28,351	Deep sorrow, especially that caused by someone's death
Joy	3,422	A feeling of great pleasure and happiness
Loathing	891	Feeling of intense dislike or disgust; hatred
Melancholy	3,682	A sad feeling
Pleasantness	624	Being enjoyable, attractive, friendly, or easy to like
Rage	5,665	Violent, uncontrollable anger.
Responsiveness	2,928	Reacting quickly and positively
Sadness	3,819	Condition or quality of being sad
Serenity	8,184	Being calm, peaceful, and untroubled
Ecstasy	30,590	The overwhelming feeling of great happiness or joyful excitement

2.5 Lexicon Model

Emotion Lexicons are registers of words and their expressed emotions (determined by annotating manually or automatically from large corpora). Textual emotion detection is the computational study of the natural language spoken in the text to identify its association with emotions such as anger, fear, joy, and sadness [8].

In the Lexicon model, tweets in the dataset were checked by comparing them with emotional words in the lexicon and extracting the feature using the Term Frequency (TF) algorithm. This feature extraction will result in several ranges between 0 to 1 for each tweet data. The closer it is to 1, the tweet checked will become relevant to the emotion expressed by the word on the lexicon. The lexicon vector is used for building the emotion lexicon model in this research framework. This research used two publicly available lexicons, SenticNet and NRC Emotion Lexicon (EmoLex).

SenticNet uses dimensionality reduction to infer the polarity of common-sense concepts and hence provides a public resource for mining opinions from natural language text at a semantic level rather than only at a syntactic level [19]. Table 2 displays the spread of each emotion complimented with the description and its word count. It is one of the publicly available complete lexicons, with a total of 121,318 words recorded in the SenticNet corpus.

The NRC Emotion Lexicon (EmoLex) includes entries for nearly 14,000 English terms. Each record contains ten binary scores (0 or 1), signaling no association or association with eight primary emotions and documenting positive and negative sentiments for each record. EmoLex was selected because the lexicon began in 2010 and has been used in numerous research projects [20], [21].

Rather than using the EmoLex as it is, an attempt was made to improve its performance by increasing its vocabulary. Initially, EmoLex consisted of 14,154 words expressing various emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. After checking these words, several exact words were discovered, which reduced the word count to 9,764. The Thesaurus Indonesian language dictionary helped increase the vocabulary [22]. In this corpus, the first column contains a parent word. The next column contains an array of synonyms based on the parent words in the previous column. Once the Thesaurus Indonesian model is loaded, the EmoLex corpus is checked for synonyms. When the checks find a parent word with a synonym, the synonym array is added to the EmoLex corpus and translated into an individual row.

These synonyms are labeled similarly to the emotion label of the parent word. After checking for the duplicates, 28,445 words were counted, which were further reduced to 8,015 words after neutral words (a word with 0 labels on every emotion) were eliminated. The combined synonyms from Thesaurus Indonesian and EmoLex resulted in a new corpus that has 17,779 words with the following ratio: 2,222 words are labeled as anger; 1,492 words are labeled as anticipation; 1,851 words are labeled as disgust; 2,610 words are labeled as fear; 1,246 words are labeled as joy; 2,228 words are labeled as sadness; 810 words are labeled as a surprise, and 2,693 words are labeled as trust. The increased vocabulary increases the model’s ability to tell apart the

DOI: https://doi.org/10.29207/resti.v7i2.4816

emotion expressed for each tweet. The Flow of the extension lexicon into the new Lexicon is illustrated in Figure 2.

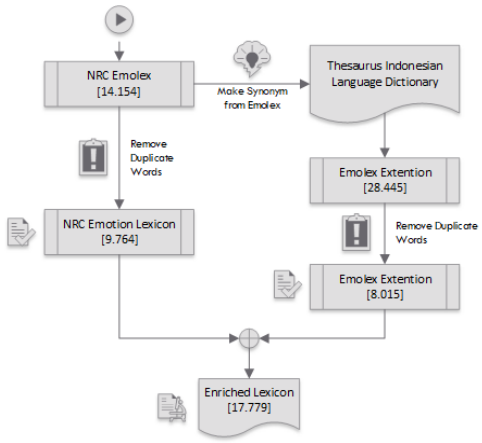

Figure 2. EmoLex Extension Flow

2.6 Hybrid or Ensemble Method

In this step, an instance of the word embedding, and lexicon model are combined into one hybrid model, then fed into a machine learning voting classifier. Word embedding and the Lexicon library in this model are selected from the comparison of the testing result of the previous two models.

2.7 Classifiers

The proposed model is compared with six machine learning methods: Naïve Bayes, K-Nearest Neighbour, Support Vector Machine, Logistic Regression, Random Forest, and AdaBoost. The models are trained with several different classifier methods to compare our models' performance.

The best model from word embedding is the Random Forest model from FastText. At the same time, the best model from the lexicon is the Random Forest model from EmoLex. These models are used to train for the ensemble model. These models are ensembled and voted on based on soft voting or majority rules classifier.

3. Results and Discussions

This section will discuss the research results and the following evaluation. Based on the hybrid model, the testing was performed in various scenarios. These scenarios include word embeddings, lexicon, and an ensemble model. The dataset for this testing is divided with a ratio of 80:20. The training data utilize 80% of the dataset, and testing data use the remaining 20%.

3.1 Word Embedding Model (*WEM*)

Based on the results of word representation using FastText representation and several previously mentioned classifiers, the following results are recorded as shown in Table 3.

From these results, it can be inferred that Naïve Bayes, with 34.3% accuracy, is the worst performance model of the text representation scenario. This result is quite surprising because although Naïve Bayes is an old method, it has proven to be quite effective while handling a large amount of data, and the robustness of this method can still be ensured to be more competitive when compared with other modern methods [12].

On the other hand, Random Forest, with an accuracy of 59.6%, has become the best-performed method in this test scenario. This result is unsurprising because Random Forest is one of the most stable and top methods. The result of machine learning models was recorded using Word2Vec as the word embedding in Table 3.

The results show that Naïve Bayes is the worst-performing model, with 34.9%. The average accuracy using this model is around 40%. Random Forest has served as the best classifier method again out of the five classifiers. This result shows that out of the tested word embedding library, FastText has been proven as a better choice in this dataset.

Table 33. Evaluation Performance of WEM

WEM	Classifier	Precision	Recall	F-score	Accuracy
FastText	KNN	41.3%	41.3%	40.9%	41.4 %
	LogRes	39.1%	39.5%	39.2%	39.5 %
	Random Forest	60.1%	59.6%	59.6%	59.6 %
	Naïve Bayes	35.2%	34.4%	33.2%	34.3 %
	AdaBoost	33.9%	34.6%	34%	34.6 %
	SVM	40.4%	40.8%	40.4%	40.8 %
Word2Vec	KNN	42.6%	42%	41.6%	42.6 %
	LogRes	37.7%	37.9%	37.7%	38.0 %
	Random Forest	59.6%	58.9%	59.0%	58.8 %
	Naïve Bayes	36.5%	34.1%	33.1%	34.2 %
	AdaBoost	34.9%	35.1%	34.9%	35.2 %
	SVM	37.1%	37.5%	36.9%	37.6 %

The explanation which we can give is that because FastText are inherently better at figuring the relationship of rare word rather than Word2Vec. While the Word2Vec feeds individual word into the Neural Network, FastText break words into several n-grams (sub-words). This results in rare words can be properly represented because the n-grams of these words can also appear in different words, therefore associating the words into the same category.

3.2 Lexicon Model

In this scenario, two lexicon models were tested: SenticNet and EmoLex. The same classifier setting was used as in the word embedding model. The results of the SenticNet model are shown in Table 4.

The result shows that the worst-performing model is Logistic Regression with 19.5% accuracy. And with the

DOI: https://doi.org/10.29207/resti.v7i2.4816

consistency of the Random Forest, it performed with the highest accuracy in this scenario with 41.3% accuracy. We used the same parameter for the EmoLex model. From the results, we can see that Naïve Bayes is the worst performed model with 26.3%. The average accuracy using this model is around 30% which is lower than the overall Word Embedding model. The best-performing model is Random Forest, with an accuracy of 50.7%, which is lower than the highest-performed model in the Word Embedding model. This result shows that EmoLex will be chosen for the ensemble model out of the tested Lexicon, as it performs better than the Word2Vec Model.

Table 4. Result using Lexicon

Lexicon	Classifier	Prec.	Recall	F-score	Acc.
Sentic-Net	KNN	30.4%	28.1%	27.1%	27.9%
	LogRes	3.9%	2%	6.5%	19.5%
	Random Forest	43.8%	41.5%	40.9%	41.3%
	Naïve Bayes	28.3%	24.6%	19.7%	23.8%
	AdaBoost	29.1%	28.3%	26.9%	27.8%
	SVM	29.4%	27.9%	25.9%	27.3%
Emo-Lex	KNN	31.0%	30.0%	29.9%	30.0 %
	LogRes	38.1%	38.3%	38.1%	38.4 %
	Random Forest	52.1%	50.8%	50.9%	50.7 %
	Naïve Bayes	21.4%	25.8%	20.6%	26.3 %
	AdaBoost	33.5%	34.0%	33.5%	34.0 %
	SVM	37.3%	37.8%	37.3%	37.9 %

The better performance of the EmoLex rather than SenticeNet could be attributed from the complexity of the Lexicon and inherent nature of the dataset used in this research. Although as a lexicon, SenticNet has larger database and much more complete word with 23 emotion labels rather than EmoLex with 8 emotion labels. In practice the complexity of the SenticNet backfired in this research because of confusion in training the model. Whereas EmoLex are positioned as a better tailored lexicon for our dataset which has much more similar and simple emotion labelling.

3.3 Ensemble Model

This is based on the preliminary results of the performance evaluation with several individual scenarios. FastText was used as a word embedding library for weighting words, and then added EmoLex as the lexicon semantic approach in our ensemble model. The Random Classifier method is also used for this scenario because of our dataset's best performance. The results of this approach are shown in Table 5 and compared with the best-performing model for each design.

Table 5. Result of The Hybrid Model

Scenario	Precision	Recall	F-score	Accuracy
FastText + Random Forest	60.1%	59.6%	59.6%	59.6 %
Word2Vec + Random Forest	59.6%	58.9%	59.0%	58.8 %
SenticNet + Random Forest	43.8%	41.5%	40.9%	41.3%
EmoLex + SenticNet	52.1%	50.8%	50.9%	50.7 %
FastText + EmoLex + Random Forest	65.9%	65.0%	65.0%	65.0 %

These results show that the accuracy of the ensemble classifying for Word Embedding and Lexicon, which is 65.0%, is the highest compared to the other scenarios. Compared with the second-highest model, FastText with Random Forest, it has an improvement of 5%, while the worst-performing model is the SenticNet Lexicon model. The improvement of the model performance can be attributed by the combined advantage of each text classification method which ensembled as one scenario.

3.4 Discussion

This study proposes an emotion detection method using two different types of feature extraction. The first feature extraction uses the word embedding technique, which has the advantage of semantic meaning. In this study, the word embedding techniques tested were FastText and Word2Vec. From the test results, FastText shows higher accuracy results than Word2Vec.

The highest score was obtained using the Random Forest classifier, with FastText having an accuracy of 59.6% and Word2Vec at 58%. FastText has higher accuracy because it doesn't handle the whole word to represent the word like Word2Vec. FastText represents each syllable in n-gram form, so FastText can better manage words that never appear in the vocabulary from models that have been made.

(a) (b)

Figure 3. Misclassification of Word Cloud from (a) Word Embedding and (b) Lexicon

The second feature extraction uses a knowledge base including a set of words labeled emotions in Indonesian, commonly called a lexicon. This study conducted

DOI: https://doi.org/10.29207/resti.v7i2.4816

experiments using a lexicon from EmoLex and SenticNet, which were tested on several classifiers. A synonym search was carried out using the Thesaurus Indonesian language dictionary to enrich the words in the lexicon.

The experiment results show that EmoLex is better at classifying with an accuracy of 52% using the Random Forest classifier. Using the Random Forest classifier, the most outstanding accuracy was from SenticNet at 41%. SenticNet has a low accuracy value because SenticNet's emotional word corpus does not detect many words from the dataset.

The results of the experiments on both types of feature extraction are the same, showing that the best classifiers are Random Forest and SVM. Both methods apply the ensemble voting classifier technique using different features. The hybrid approach successfully uses features from the lexicon and word embedding in applying the ensemble voting classifier technique. Using other components can fill one another's deficiencies in each element. Then the voting classifier votes to choose the best model. It can provide higher accuracy for emotion detection.

We concluded from the multiple outcomes that Naïve Bayes is incompatible with our model as it performed poorly. The hypothesis is that the low result of the Naïve Bayes classifier is because of this method's characteristics of conditional independence assumption, namely, the assumption that features are independent of one another when conditioned upon class labels. It was found that this assumption is rarely accurate. Features often depend on one another in non-trivial amounts, meaning multiple parts often contain similar signals. However, the Naive Bayes classifier's conditional independence assumption results in its treatment of features as distinct signals, each of which should independently contribute additional confidence to the classifier's prediction.

This phenomenon amplifies the contribution of signals to the ultimate classification confidence. The Naïve Bayes classifier nonetheless produces competitive classification accuracy because the extent to which the independence assumption favors different class labels roughly evens out on average. Although, in this research scenario, no class labels benefit this phenomenon.

Other algorithm that noticeably performed poorly in the research application is Logistic Regression. This model is one of the most interpretable classifiers used in this research. We use it to get a feeling for the most important features and the direction of the dependence. As the result, Logistic Regression performed poorly especially in SenticNet scenario. We reckon other reason is that Logistic Regression are mainly used to predict a binary outcome and therefore incompatible for our use cases which has several class labels outcome.

Several test instances also shown that Random Forest's algorithm is highly compatible with our model. The combination of the multiple Decision Trees helped Random Forest to conceive some biased classifiers. Each decision tree captures a different class label since a random subset of the instances is in this method's interest.

At the maximal randomness, Random Forest organizes nodes from a random subset of the features. In this way, feature-based randomness is also used. After creating n number of trees in this randomly, more cluttered decision boundaries than simple lines were obtained. (n decision trees use an n voting scheme to decide about unique instances).

The model performance is further analyzed by delving into the misclassified words. These misclassified words are projected into word clouds for each Word Embedding and Lexicon, as shown in Figure 3. The top five misclassified words from these two scenarios are the same. They are '*ya*' (yes), '*gue*' (i, me), '*kalo*' (if), '*banget*' (very), '*sih*' (anyway). All these words are hard to be classified based on the emotions on the dataset label from word embedding and emotion lexicon. The misclassification happened because these words are neutral, meaning they are not words that strongly express feeling and could only express a substantial emotion value when paired with certain words that describe an emotion.

4. Conclusion

In this research, the classification of emotions on Twitter using lexicon-based feature extraction and word embedding is proven to improve accuracy. The classification used in applying the ensemble voting classifier technique uses the 2 best methods when testing scenarios for each feature extraction. The best classification method using lexicon feature extraction or word embedding shows that SVM and Random forest are classifiers with the best accuracy. Using both classifiers in the ensemble voting classifier technique produces a model with an accuracy of 65.8%, the highest compared to other scenarios that have been tried. The ensemble method proved effective because there was an increase of 6.2% compared to the second best scenario models, FastText and Random Forest, which reached 59.6%.

From this research, the classification of emotions from Indonesian language texts can be further developed because several sentences still need to fit into the available emotion class. So, a neutral label is required to represent the tweet. Then, in applying feature extraction using the lexicon, many emotional words still need to be detected by the lexicon. This causes the information obtained in a text to be incomplete. Therefore, in future research, a classification of emotions can be developed by adding a neutral label

DOI: https://doi.org/10.29207/resti.v7i2.4816

and enriching the emotion lexicon with other methods to detect more expressive words.

Reference

[1] J. Abdillah, I. Asror, and Y. F. A. Wibowo, "Emotion Classification of Song Lyrics using Bidirectional LSTM Method with GloVe Word Representation Weighting," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 4, pp. 723–729, Aug. 2020, doi: 10.29207/resti.v4i4.2156.

[2] F. Keshtkar and D. Inkpen, "A Hierarchical Approach to Mood Classification in Blogs," *Nat. Lang. Eng.*, vol. 18, no. 1, pp. 61–81, Jan. 2012, doi: 10.1017/S1351324911000118.

[3] D. Purwitasari, A. Apriantoni, and A. B. Raharjo, "Identifikasi Pengaruh Pandemi Covid-19 terhadap Perilaku Pengguna Twitter dengan Pendekatan Social Network Analysis," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 8, no. 6, p. 1309, Nov. 2021, doi: 10.25126/jtiik.2021865213.

[4] T. T. Sasidhar, P. B, and S. K. P, "Emotion Detection in Hinglish (Hindi+English) Code-Mixed Social Media Text," *Procedia Comput. Sci.*, vol. 171, pp. 1346–1352, 2020, doi: 10.1016/j.procs.2020.04.144.

[5] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare, G. Yogapriya, and P. Pandey, "An efficient way of text-based emotion analysis from social media using LRA-DNN," *Neurosci. Inform.*, vol. 2, no. 3, p. 100048, Sep. 2022, doi: 10.1016/j.neuri.2022.100048.

[6] H. Aka Uymaz and S. Kumova Metin, "Vector Based Sentiment and Emotion Analysis from Text: A survey," *Eng. Appl. Artif. Intell.*, vol. 113, p. 104922, Aug. 2022, doi: 10.1016/j.engappai.2022.104922.

[7] Rahman, Romana, Islam, Tajul, and Ahmed, Md. Humayan, "Detecting Emotion from Text and Emoticon," *Lond. J. Res. Comput. Sci. Technol.*, vol. 17, no. 3, pp. 9–13, 2017.

[8] A. Bandhakavi, N. Wiratunga, S. Massie, and D. Padmanabhan, "Lexicon Generation for Emotion Detection from Text," *IEEE Intell. Syst.*, vol. 32, no. 1, pp. 102–108, Jan. 2017, doi: 10.1109/MIS.2017.22.

[9] Md. Y. Kabir and S. Madria, "EMOCOV: Machine Learning for Emotion Detection, Analysis and Visualization Using COVID-19 Tweets," *Online Soc. Netw. Media*, vol. 23, p. 100135, May 2021, doi: 10.1016/j.osnem.2021.100135.

[10] D. yasmina, M. Hajar, and A. M. Hassan, "Using YouTube Comments for Text-based Emotion Recognition," *Procedia Comput. Sci.*, vol. 83, pp. 292–299, 2016, doi: 10.1016/j.procs.2016.04.128.

[11] A. R. Murthy and K. M. Anil Kumar, "A Review of Different Approaches for Detecting Emotion from Text," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1110, no. 1, p. 012009, Mar. 2021, doi: 10.1088/1757-899X/1110/1/012009.

[12] H. Tian and L. Wu, "Microblog Emotional Analysis Based on TF-IWF Weighted Word2vec Model," in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, Nov. 2018, pp. 893–896. doi: 10.1109/ICSESS.2018.8663837.

[13] Siti Khomsah, Rima Dias Ramadhani, and Sena Wijaya, "The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 6, no. 3, pp. 352–358, Jun. 2022, doi: 10.29207/resti.v6i3.3711.

[14] F. Incitti, F. Urli, and L. Snidaro, "Beyond Word Embeddings: A survey," *Inf. Fusion*, vol. 89, pp. 418–436, Jan. 2023, doi: 10.1016/j.inffus.2022.08.024.

[15] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. Ch. Chatzisavvas, "Sentiment Analysis Leveraging Emotions and Word Embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, Mar. 2017, doi: 10.1016/j.eswa.2016.10.043.

[16] S. Muñoz and C. A. Iglesias, "A Text Classification Approach to Detect Psychological Stress Combining a Lexicon-Based Feature Framework with Distributional Representations," *Inf. Process. Manag.*, vol. 59, no. 5, p. 103011, Sep. 2022, doi: 10.1016/j.ipm.2022.103011.

[17] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, and A. Jamal, "Colloquial Indonesian Lexicon," in *2018 International Conference on Asian Language Processing (IALP)*, Bandung, Indonesia, Nov. 2018, pp. 226–229. doi: 10.1109/IALP.2018.8629151.

[18] J. Serrano-Guerrero, M. Bani-Doumi, F. P. Romero, and J. A. Olivas, "Understanding What Patients Think about Hospitals: A Deep Learning Approach for Detecting Emotions in Patient Opinions," *Artif. Intell. Med.*, vol. 128, p. 102298, Jun. 2022, doi: 10.1016/j.artmed.2022.102298.

[19] E. Cambria, R. Speer, C. Havasi, and A. Hussain, "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," *AAAI Fall Symp. Commonsense Knowl.*, pp. 14–18, Feb. 2010.

[20] A. S. Aribowo and S. Khomsah, "Implementation of Text Mining for Emotion Detection Using the Lexicon Method (Case Study: Tweets About Covid-19)," *Telematika*, vol. 18, no. 1, p. 49, Mar. 2021, doi: 10.31315/telematika.v18i1.4341.

[21] B. Waspodo, Nuryasin, A. K. N. Bany, R. H. Kusumaningtyas, and E. Rustamaji, "Indonesia COVID-19 Online Media News Sentiment Analysis with Lexicon-based Approach and Emotion Detection," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, Yogyakarta, Indonesia, Sep. 2022, pp. 1–6. doi: 10.1109/CITSM56380.2022.9935884.

[22] J. Bata, "Leksikon untuk Deteksi Emosi dari Teks Bahasa Indonesia," in *Seminar Nasional Informatika 2015*, Yogyakarta, 2015, pp. 195–202. [Online]. Available: https://www.neliti.com/id/publications/174754/leksikon-untuk-deteksi-emosi-dari-teks-bahasa-indonesia

DOI: https://doi.org/10.29207/resti.v7i2.4816