

Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text

Aprilia Nurkasanah¹, Mardhiya Hayaty^{2*}

^{1,2} Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

¹ aprilia.nurkasanah@students.amikom.ac.id, ² mardhiya_hayati@amikom.ac.id

*corresponding author

Accepted 06 April 2022

Approved 24 May 2022

Abstract— Text Mining is a part of Neural Language Processing (NLP), also known as text analytics. Text mining includes sentiment analysis and emotion analysis which are often used to analyse social media, news, or other media in written form. The emotional breakdown is a level of sentiment analysis that categorises text into negative, neutral, and positive sentiments. Emotion is organised into several classes. This study categorized emotion into anger, fear, happiness, love, and sadness. This study proposed feature extraction using Lexicon and TF-IDF on the emotion recognition dataset of Indonesian texts. InSet Lexicon Dictionary is used as the corpus in performing the feature extraction. Therefore, InSet Lexicon was chosen as the dictionary to perform feature extraction in this study. The results show that InSet Lexicon has poor performance in feature extraction by showing an accuracy of 30%, while TF-IDF is 62%.

Index Terms— Emotion Recognition Text, Feature Extraction, InSet Lexicon, Lexicon, Random Forest.

I. INTRODUCTION

Communication between individuals is generally done verbally and non-verbally. Verbal communication begins and is obtained through words, sentences, or paragraphs. We often do not catch the emotion implied in the written words, sentences, and sections in verbal communication. As time goes by, the way of communication between individuals has changed. When technology is developing, the exchange of information flows spreads and develops rapidly. Social media is an oral communication tool that is often accessed and has become a trend. In 2021, Indonesia will have active social media users, as much as 61.8% of the existing population[1]. The number increased by 6.3% from the last year. Not only as a trend, but social media has also become a necessity. It can be seen from the statistical data of the social media users that it continuously grows every year. Various video, voice, and text data are added every minute. The flow of data information carried by social media is enormous. Thus, it is necessary to have an accurate processing model so that the information on social media can be processed correctly. Social media has

become a part of life in modern society. Through social media, people often express themselves regarding existing changes. Not only do they give objective opinions on an event, but social media users also often express their subjective views. Social media is considered a critical communication medium [2]. Not only opinions or criticisms about an event, but social media users also often criticize products, services, and even government decisions currently hotly discussed. Social media can be used as a medium for monitoring public opinion. This social media user shares their emotions in posts that many people from various platforms watch. Emotion influences humans to express hundreds of words. Therefore, it is considered a crucial thing in communication.

Natural Language Processing (NLP) in computational linguistics is a modern study of linguistics using computer science tools. NLP is also known as a subfield of Artificial Intelligence (AI). [3] NLP extract meaningful information from various texts, including text from social media. Social media has big data containing essential informations. These data give valuable information about a particular phenomenon, government policies, or reviews of using specific products. [4] NLP allows researchers to easily extract beneficial insights in textual datasets while avoiding burdensome computational work. These responses, comments, and reviews are then assessed based on sentiment class, commonly known as sentiment analysis.

Starting from the field of sentiment analysis, which analyzes opinions or messages into negative, positive, and neutral sentiment, emotion recognition is a sub-field of more detailed sentiment analysis that recognizes happy, sad, and angry emotions. Emotional interaction has been a part of a psychological phenomenon in daily life. It can be easily found in everyday interaction and encourages the existence of emotion recognition. Emotion recognition becomes the basis of successful human interaction, communication, and decision-making. The development of big data

influences it to become a significant issue both academically and industrially. [5] Text data produced in emotional communication is often used to understand the human's emotional state. A person's emotional state is often associated with feelings influenced by the individual's interaction with their environment [6]. According to psychology, basic emotions are classified into six classes: joy, sadness, fear, disgust, surprise, and anger [7].

Emotion recognition is currently done through facial expressions and utterances and written text. In applying human-computer interaction, recognizing human emotion helps improve the impression in executing the system. Emotion recognition in text form can be analyzed by several approaches: keyword-based, rule-based, classical learning-based, deep learning-based, and hybrid systems [8]. The methods usually used for emotion classification are Naïve Bayes, J48, K-Nearest Neighbor (KNN), and Support Vector Machine-Sequential Minimal Optimization (SVM-SMO) [9]. [10] compared automated text classification methods and showed that Naïve Bayes and Random Forest provide high accuracy. Naïve Bayes is often used in classifying emotions based on conditional probabilities to organize data in predefined classes [11].

However, this research has not achieved maximum accuracy because there are sentences that do not represent the actual emotions; for example, a sentence belongs to the fear class but is indicated to be in the happy class [11]. Machine learning is a method for allowing machines to learn from empirical data that experts have confirmed. Although the machine learning approach is used in many studies, our research takes a new approach to the problem by using the lexical approach to overcome data processing. Lexicon is also evaluated as a feature extraction tool in this study. As a result, this study utilizes a keyword-based approach to the Lexicon method. [12] Lexicon proved capable of providing high accuracy in sentiment analysis. This method is expected to perform feature extraction in emotion recognition in Indonesian texts more accurately.

II. RELATED WORK

Many studies on emotion mining have been done, including in Indonesian text. The [13] study included sentiment analysis and emotion recognition of readers. With the available approaches and features, texts from various sources such as news and social media posts can be used as the dataset in the study of emotion mining [14], [15].

The researchers of Affective Computing proposed many rule-based approaches for extracting the text emotion automatically. [16] combined Lexicon-based, Bag-of-Words, Words embedding, orthography, and Part-of-Speech (POS) tag, and got an F1-Score accuracy of 69.7%. The research classifies emotions

into 5 types: anger, fear, happiness, love, and sadness. Another model classifies emotions into six basic specifications of emotion: anger, disgust, fear, joy, sadness, and surprise [17]. Research [18] made a model for classifying emotions into five classes using an Indonesian language dataset and showed that the Maximum Entropy (ME) algorithm has better accuracy than SVM by 72%. Another feature extraction used in text extraction is TF-IDF [19]. TF-IDF works by extracting keywords based on the frequency of words that appear. In research [20], emotions were classified into 6 classes and showed that TF-IDF could provide high accuracy in text emotion analysis of the text. [21] showed that TF-IDF is a feature extraction based on a weighting method that can outperform feature extraction with N-Gram.

Research [22] combined the Random Forest Classifier with Unigram and SentiWordNet to produce the highest accuracy on sentiment analysis datasets in Malayam. [23] combined Lemmatization, TF-IDF, and Random Forest and showed a good result in speech emotion recognition models. The research used a Logistic Regression algorithm, Support Vector Machine, and Random Forest to calculate the accuracy of feature extraction. [24] used Random Forest on sentiment analysis with a deep learning approach. [25] The Random Forest classifier is often used in predictive modelling to reduce the number of variables required so that the burden of data collection is reduced and efficiency is increased. Research [26] shows that Random Forest worked based on various predictor variables without any assumption about the response variable. [27] Random Forest integrates trees, where a classification is made from several decision trees and produces an output based on the number of votes from all the tree outputs. [28] used domain-specific emotion lexicons (DSLs) and general-purpose emotion lexicons (GPELs) to study feature extraction emotional problems. [29] used the Lexicon approach to generate emotional weights or values.

Before a specified model recognizes the emotion in a text, the text must have a particular stage so the machine can understand the series of words or sentences. The stage is called pre-processing. [30] Pre-processing is a stage of converting unstructured data into structured data based on the needs. In extracting Lexicon features, research [31] carried out several pre-processing stages: case-folding, punctuation, conversation word, stopword removal, stemming, and tokenization. [32] Text pre-processing methods improve the predictive accuracy of the generated models for sentiment classification. Research [33] used NRC Affect Intensity Lexicon and SentiStrength techniques to extract and analyze the characteristics of Twitter's users towards sentiment and behaviors signaling the "suicide" sign. [34] showed that the Lexicon-based approach worked well for sentiment analysis. The Semantic Orientation Calculator (SOCAL) assigns a positive or negative label to a particular

text, capturing the text's opinion toward the main topic. [35] detected Lexicon emotion by relying on the relationship between words and emotions in WordNet. Research [36] showed that Lexicon feature extraction is more commonly found in handling medical sentiment than SentiWordNet (SWN) in the drug review dataset. [37] showed that Lexicon significantly surpassed the BoW feature in emotion classification. [38] used Lexicon to perform calculations by representing each dataset with a binary vector.

The use of Lexicon in the classification of emotions in English datasets has been widely applied. However, the use of Lexicon in emotion analysis in Indonesian text is still rarely used. Another study [39] of emotion analysis used the Term-Weighting Scheme in the approach. In this study, the InSet Lexicon dictionary [40], which contains a weighting of 3,609 positive words and 6,609 negative words in Indonesian, is used. The value weights in the InSet Lexicon are obtained from manual weighting by Indonesian language experts with weights between -5 to +5 as done by Affected Lexicon (AFFIN) [41]. InSet Lexicon is proved to be the best among translated SentiWordNet, translated Liu Lexicon, translated AFINN Lexicon, and Vania Lexicon in analyzing the sentiment for Indonesian. This dictionary is based on Twitter data, so it contains common words and non-standard words. It uses Lexicon to perform feature extraction from the emotion recognition dataset.

III. METHODOLOGY

A. Dataset

This dataset is taken from [16] applying the collection method by using the Twitter Streaming API for two weeks starting from June 1 2018, to June 14, 2018, and setting the geolocation filters in Indonesia. The datasets contains 4,403 Indonesian-language tweets with a 0.917 score of annotation, which has 5 (five) emotion classes, namely love, anger, sadness, joy, and fear. Fig 1. shows the balanced number of emotions in the emotions of anger, joy, and sadness. However, emotions with the categories of love and fear have a limited number of tweets.

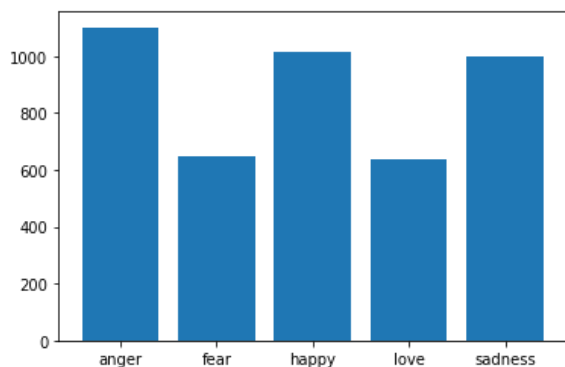


Fig. 1. Class Distribution on Dataset

B. Preprocessing

The pre-processing in this research is divided into six stages as follows:

- Case-folding: changing all letters to lowercase.
- Punctuation: deleting unnecessary symbols or characters.
- Convert Word: converting abbreviated words into standard language using the dictionary of abbreviations provided in the emotion dataset [16].
- Stopwords: removing words that are not needed in the used dataset.
- Stemming: finding the root word of a word that has an affix. The Indonesian stemming library, namely Sastrawi, is used in this process. In the Indonesian dataset, stemming is an important step. It is because Indonesian has many affixes. They are suffixes (addition at the end of a word), prefixes (addition at the beginning of a comment), and confixes (addition at the beginning and the end of a word).
- Tokenization: the last step of pre-processing is breaking sentences into text based on delimiters (spaces).

TABLE I. RESULT OF PREPROCESSING DATASET

Label	Text Clean	Text Processed
Anger	soal jalan jatiburpolisi tidak bisa gertak gubernur	['jalan', 'jatiburpolisi', 'gertak', 'gubernur']
Fear	pak ketua di temani sekretaris tapi selaku penanggung jawab	['ketua', 'teman', 'sekretaris', 'tanggung']
Happy	kepingin gudeg mbarek bu hj amad	['kepingin', 'gudeg', 'mbarek', 'bu', 'hj', 'amad']
Love	pasangan seringkali diibaratkan sebagai rumah karena rasa nyaman	['pasang', 'seringkali', 'rumah', 'nyaman']
Sadness	melihatmu sedih kebahagianku pudarseperti saat kala itu	['lihat', 'sedih', 'kebahagianku', 'pudarseperti']

Table 1 shows the example of the emotion recognition dataset of Indonesian text that passed the pre-processing phase.

C. Feature Extraction

Feature extraction is a part of feature engineering. Feature Engineering is divided into two, feature extraction and feature selection. After the pre-processing phase, the next step is feature extraction. Feature Extraction pulls out words from text data to be converted into features used by the classifier [42]. Feature Extraction helps get the best features by combining variables into components to reduce the

amount of data [19]. Feature extraction is the data transformation process in which raw data is changed into numerical features. These numerical features are processed by maintaining the information from the raw data set. Feature extraction is believed to give better results than using machine learning directly toward the raw data. It has a significant impact on the calculation results of the classifier. Through feature extraction, data is processed so that it can be read and used as input by the specified classifier. Fig. 2 shows the workflow of the Lexicon feature extraction used in this study.

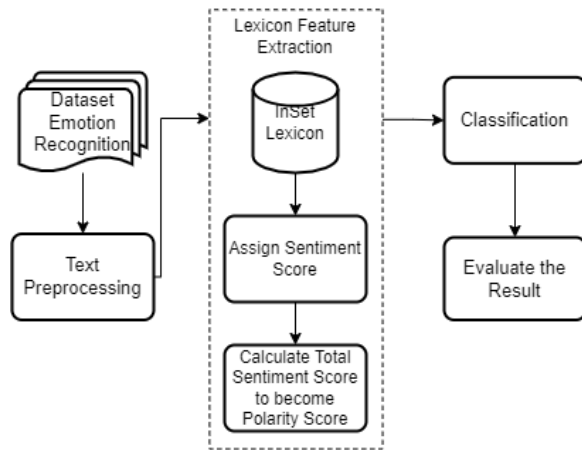


Fig. 2. Flowchart Diagram

The Lexicon-based approach relies on the emotions in the dictionary. This study uses the InSet Lexicon [37], an Indonesian sentiment dictionary accompanied by a weighting for each word to determine the polarity score. This dictionary is divided into two categories positive and negative. The following is the distribution of the Lexicon InSet dataset based on its type:

TABLE II. INSET LEXICON POSITIVE

Words	Weight
hai	3
merekam	2
ekstensif	3
paripurna	1
detail	2
pemik	3
belas	2
welas	4
kabung	1
rahayu	4

TABLE III. INSET LEXICON NEGATIVE

Words	Weight
putus tali gantung	-2

gelebah	-2
gobar hati	-2
tersentuh (perasaan)	-1
isak	-5
larat hati	-3
nelangsa	-3
remuk redam	-5
tidak segan	-2
gemar	-1

The dataset that passes the pre-processing stage is assigned a polarity score. This step analyzes sentiment words in the emotion recognition dataset of Indonesian texts and then determines the polarity score based on Lexicon [43].

TABLE IV. RESULT OF ASSIGN POLARITY SCORE

Assign Polarity Score				
Sentiment word	=	semangat	puasa	turun
Polarity Score	=	4+(-1)	0+(-2)	1+(-3)

From Table IV, it is known that the words 'semangat', 'puasa' and 'turun' are words that exist in the emotion recognition dataset of Indonesian texts. Among these three words, 'semangat' and 'turun' are categorized as positive and negative sentiments. At the same time, the word 'puasa' belongs to negative sentiments.

The sentiment score in the InSet Lexicon is the value used to determine the sentiment of the existing dataset. In this study, the polarity score was obtained from the sum of the sentiment scores in each row of the emotion recognition dataset of Indonesian texts.

$$Polarity\ Score = \sum_{i=1}^n Overall\ Sentiment\ Score(i) \quad (1)$$

In this step, after the polarity value is determined, the next step is labeling or categorizing the sentiment class of each data. The sentiments are classified according to the following rules:

$$\begin{aligned}
 &Polarity\ Score > 0 \rightarrow Positive \\
 &Polarity\ Score = 0 \rightarrow Neutral \\
 &Polarity\ Score < 0 \rightarrow Negative
 \end{aligned} \quad (2)$$

The sum of the polarity score can be seen in Table V.

TABLE V. RESULT OF POLARITY SCORE

Assign Polarity Score				
Sentiment word	=	semangat	puasa	turun
Polarity Score	=	3	-2	-2

From the calculation in Table V above, it can be seen that the emotion recognition data in Indonesian texts have negative sentiment or polarity with a score of = -1. This step will continue until the last row of data of the existing dataset. Fig. 3 displays the results of each data's polarity and polarity values.

label	text_clean	text_preprocessed	polarity_score	polarity
0	anger	soal jalan jatibaruipolis tidak bisa gerak g...	[jalari', 'jatibaruipolis', 'gerak', 'gubemu...	-3 negative
1	anger	sesama cewek lho kayaknya harusnya bisa lebih...	['cewek', 'lho', 'kayak', 'harusnya', 'bisa', 'lebih...	-34 negative
2	happy	kepingin gudeg mbarek bu hj amad foto dari goo...	['kepingin', 'gudeg', 'mbarek', 'bu', 'hj', 'a...	0 neutral
3	anger	jalan jatibaruibagian dari wilayah tn abangpen...	['jalari', 'jatibaruibagian', 'wilayah', 'tn', '...	0 neutral
4	happy	sharing pengalaman saja kemari jam batinai L...	['sharing', 'saja', 'kemari', 'jam', 'batainai...	2 positive

Fig. 3. Polarity Score

The existing polarity values are divided into negative, neutral, and positive sentiments. Fig 4 shows the sentiment percentage in the dataset after calculating the polarity score based on the InSet Lexicon.

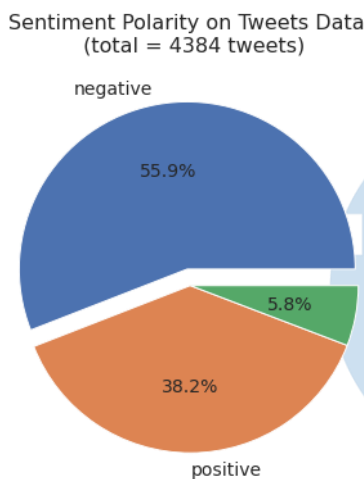


Fig. 4. Polarity Percentage

The following is the polarity value based on the emotion label with positive polarity in the dataset.

	text_clean	polarity_score	label
1	terimakasih papap terimakasih selalu ada terimakasih selalu siaga terimakasih selalu mencukupi terimakasih selalu memaklumi terimakasih selalu bersabar terimakasih selalu	36	happy
2	kegiatan ibadah sholat dzuhur uils koja penerima manfaat bersama petugas pendamping melakukan ibadah sholat dzuhur berjamaah indahny kebersamaan uri	31	happy
3	username bismillah happy bday kak wuatb gbu semoga selalu dilimpahi keberkahan rezeki kebahagiaan umur yang bermanfaat dan sukses dunia akhirat semoga citanya tercapai dan bisa membanggakan orangtua nusa negara dan agama amin btw kita sedom kak	30	happy
4	congratulation selamat ya adik cewek ke atas kelahiran putrinya yang ke saya sebagai kakak turut suka cita bersyukur senang sudah diberikan keselamatan allah swt dalam prosing persalinan putri ke nya semalam kakak berdoa semoga putri ke nya menjadi putri yang shalihah uri	30	happy
5	untukumu wanita yang selalu saya fikirkan bahagialah dengan bertambahnya usia mu sayang semoga kamu sehat semoga terjaga iman dan islammujaga bahagiamu kekasihku maka saya akan merasa tenang karena mencintaimu dalam diamsekalgi lagi selamat bertambah usia saya mencintaimu sampai akhir hdпку	30	love

Fig. 5. Positive Polarity

The following is the polarity value based on the emotion label with negative polarity in the dataset.

	text_clean	polarity_score	label
1	wartawan di bunuh guru di bunuh aktivis di bunuh seseorang di bunuh belasan orang di bunuh puluhan orang di bunuh ratusan orang di bunuh ribuan orang di bunuh puluhan ribu orang di bunuh ratusan ribu orang di bunuh jutaan orang di bunuh	-52	anger
2	seandainya jokowi kalah sudah dipastikan hati bangsa bangkit teroris bangsa bangkit radikal bangsa bangkit semua politik balas dendam bangkit yang minoritas ditekan mending memisahkan diri dari nkri untuk itu harus tetappokowi	-46	anger
3	namanya orangtua ya memang wajib kan bayarin anaknya sekolah jadi tidak ada itu yang namanya buang duit semua yang sudah dilakuin sama kita baik benar apa tidak itu gabakal sia sia tidak usah takut gagal tidak usah mikir buruk justru yang sia sia itu dengan kalian ngejudge orglain kayak gini	-41	anger
4	jujur saja cukup yah kebobokan kegliaan saya harus sadar saya ingin segera keluar dari lingkaran setan hidup disini banyak saudara dengki n jahat sama saya tidak kuat saya n terutama sepupu sialan cin penghuni neraka si lina n lakinya n si lita juga najjilis kamu sialan	-40	anger
5	itu namanya fitnah mbak tidak ada bukti ko fitnah orang lain korupsi ingat mbk dunia ini berlaku sistem tabur tuai mgkin anda menupati keburukan anda dengan memfitnah orang lain saya tahu semua kok korupsi n pekerjaan terselubung anda	-39	anger

Fig. 6. Negative Polarity

Besides using Lexicon-based feature extraction, the researcher also uses TF-IDF to compare the accuracy. [44] TF-IDF Term Weighing, that is often used, is a result of an integration between Term Frequency and Inverse Document Frequency. This weighing assumes that infrequently appearing terms hold the highest importance. TF-IDF can be calculated using the formula as follows:

$$TF \cdot IDF(t_i, d_j) = TF(t_i, d_j) \times \log \frac{D}{d(t_i)} \quad (3)$$

Based on equation (3), Where is TF of the term t_i in document d_j , while $\frac{D}{d(t_i)}$ is IDF of term t_i .

D. Feature Extraction

The experiment in this study applies a programming language, namely Python 3.7.13, and Graphics Processing Units (GPU) based cloud service from Google called Google Collaboratory, also known as Collab. We utilise several libraries such as Pandas, Matplotlib, Sastrawi, and Sklearn. The investigation was started by importing the libraries used into the Google Collaboratory. Then, the pre-processing was done to process the dataset used in the study.

After the dataset was processed by adding the mass or value of each tweet, it was used as the input for the classifications of Machine Learning. This study uses 75% data training and 25% data testing. The researcher develops a supervised learning approach to analyse and detect emotion in a tweet text and classify them automatically. A Random Forest algorithm is applied to organise the data from the values generated by the extraction feature into emotional labels for each text. However, it is vital to highlight the focus of this research is the discussion of the performance of applying InSet-lexicon than the performance of the classification feature algorithm. There is no additional hyperparameter aiming to make a difference or improve the performance of the Random Forest algorithm in the classification.

Accuracy, precision, recall, and F-measure are applied to measure the performance with a confusion matrix. It can be seen in the following formula below:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

$$Precision = \frac{TP \text{ or } TN}{TP+FN \text{ or } TN+FP} \quad (5)$$

$$Recall = \frac{TP \text{ or } TN}{TP+FP \text{ or } TN+FN} \quad (6)$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

Explanation:

- TP (True Positive): Number of the sample that is correctly labelled positive.
- FP (False Positive): Number of the negative samples that are incorrectly labelled as positive.
- FN (False Negative): Number of the positive samples that are incorrectly labelled as negative.
- TN (True Negative): Number of the sample that is correctly labelled negative.

The accuracy is obtained by comparing the prediction ratio with the total of the existing data. Precision is a comparison between True Positive (TP) with the amount of data that is predicted to be positive. Meanwhile, Recall compares True Positive (TP) with the number of data that belongs to the positive class. In the F1 score, precision and recall are combined into a single metric. In short, the F1 score is a weighted average between precision and recall. The best value of the F1 score is indicated at one while the worst value is shown at 0.

IV. RESULT AND DISCUSSION

This research aims to test the Lexicon-based feature extraction in the form of a dictionary using the InSet Lexicon dictionary for the Indonesian language emotion recognition dataset. Feature extraction allows the simplification of the classification of the text data. The feature extraction process eliminates text dimensionality by removing unrelated features from the text data. InSet Lexicon itself provides high accuracy in analyzing sentiment datasets in Indonesian [12].

TABLE VI. EXPERIMENT RESULT ON RANDOM FOREST CLASSIFIER

Classifier	Accuracy
InSet Lexicon	30%
TF-IDF	62%

The tests carried out in this analysis used 1000 tweet data from Twitter with the keyword 'indihome' in Indonesian. This test uses 2 classes, namely positive and negative classes. The data collection is carried out using the Crawling technique that utilizes the tweepy library from the Python programming language. After that, labeling positive and negative sentiments on the dataset is done manually. For manual labeling of positive and negative sentiments, there should be the

help of linguists in determining the positive or negative sentiments of an opinion. The author only uses two classes of sentiment, namely positive and negative because the author wants a more conical conclusion between the two classes.

Table VI compares the accuracy between InSet Lexicon and TF-IDF in the study. The TF-IDF feature extraction gets better accuracy results than Lexicon using the Lexicon InSet dictionary. The polarity value obtained by adding each word in the emotion recognition dataset based on the InSet Lexicon is not influenced by the emotion labels in the dataset.

However, the results of the Lexicon InSet in this study were higher than the study [16] that has the same dataset. The previous study provided an accuracy of 24.92% in the Random Forest classification. Meanwhile, the accuracy value in this study increased by 5.08% from the prior research. The increase in accuracy value is influenced by the pre-processing. In the previous research, the pre-processing stage that was carried out was data normalization, which includes changing the letters in the dataset into lowercase form, deleting usernames and hyperlinks, and doing stopwords. After performing data normalization, Part-of-Speech (POS) Tagging was completed and continued with stemming. Meanwhile, in this study, the pre-processing is case folding, punctuation (removing symbols, characters, links, and words that are not needed), converting words (changing abbreviated words into common words), stopwords, stemming, and tokenization.

TABLE VII. EVALUATION OF EACH EMOTION CLASS IN THE DATASET

Emotion Class	Feature Extraction	
	InSet	TF-IDF
Anger	38%	63%
Sadness	8%	49%
Happy	40%	60%
Fear	14%	76%
Love	23%	73%

This experiment used the F1-Score to determine the calculation results on the metric accuracy of each existing emotion class because the emotion recognition dataset of Indonesian texts had imbalanced data. The happy and anger emotion classes have the highest F1-Score value than other emotion classes in feature extraction using InSet Lexicon. The emotion class "Sadness" shows the lowest F1 score of the other emotion classes, 8%.

The emotion class "Sadness", with an 8% score, provides the lowest F1-Score value among the other emotion classes from Lexicon feature extraction. Meanwhile, this feature extraction also shows that the emotion classes "Anger" and "Fear" provide higher

accuracy than other classes. Fig 7 shows the confusion matrix of the Lexicon feature extraction in the Random Forest classifier.

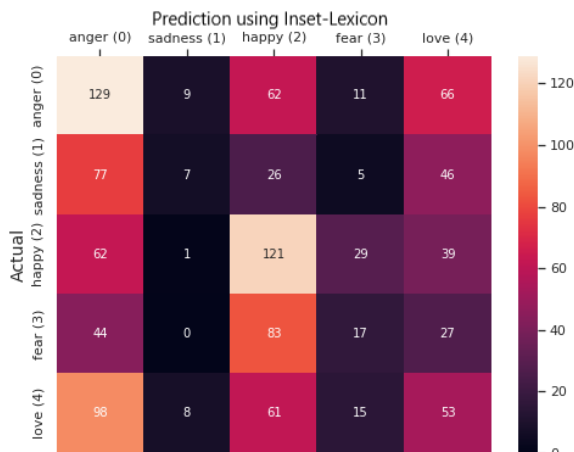


Fig. 7. Confusion Matrix of InSet Lexicon using Random Forest

The emotion class “Sadness” has a small proportion of false-positive values with moderate false negatives, which causes the “Sadness” class to have a small recall value even though it reaches a high precision value. Fig 8 shows the TF-IDF feature extraction confusion matrix in the Random Forest classifier.

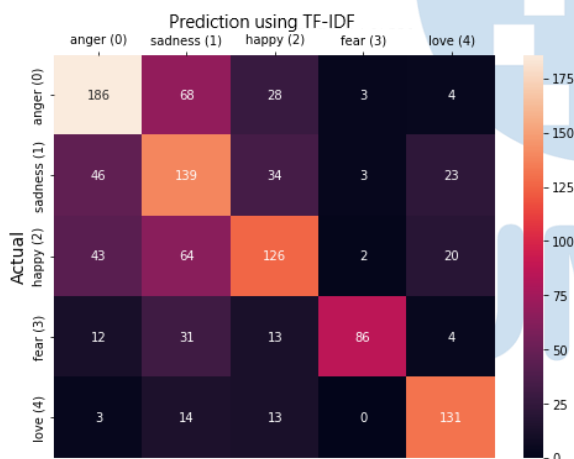


Fig. 8. Confusion Matrix of TF-IDF using Random Forest

Feature Extraction with TF-IDF provides better accuracy results on InSet Lexicon. Each emotion class can be predicted with a high F1-Score. The same with the Random Forest classification in InSet Lexicon feature extraction, the emotion class “Sadness”, with a 49% score, has the lowest F1-Score value among other emotion classes.

V. CONCLUSIONS

This study examines the Lexicon based feature extraction on the emotion recognition dataset of Indonesian texts in the form of a corpus or dictionary called InSet Lexicon. The results in this study show an accuracy of 30%. The accuracy of this result is higher

than the previous study that had the same dataset and classification. The difference in accuracy is influenced by the pre-processing stages carried out in both studies. However, it has less accuracy than feature extraction using TF-IDF which has an accuracy of 62% on the Random Forest classifier. In this study, each emotion class (Anger, Sadness, Happy, Fear, and Love) can be detected with the used classifier. The low accuracy value in the InSet lexicon is caused by the polarity value that is not influenced by the emotion labels in the dataset. Lexicon feature extraction using the Lexicon InSet, which usually provides high accuracy when used to analyze Indonesian sentiment, produces low accuracy in the emotion recognition dataset of Indonesian texts.

We will examine and improve the current results for further research to achieve better performance. The researcher also suggests future studies to investigate the effect of class imbalance on the dataset for each text weighting scheme. The imbalance plays a significant role in creating a presupposition toward selecting the majority class in the emotion recognition dataset of Indonesian texts.

REFERENCES

- [1] S. Kemp, “DIGITAL 2021: THE LATEST INSIGHTS INTO THE ‘STATE OF DIGITAL,’” *We Are Social*, 2021.
- [2] K. M. Carley, M. Malik, M. Kowalchuck, J. Pfeffer, and P. Landwehr, “Twitter Usage in Indonesia,” *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.2720332.
- [3] Y. Kang, Z. Cai, C. W. Tan, Q. Huang, and H. Liu, “Natural language processing (NLP) in management research: A literature review,” *J. Manag. Anal.*, vol. 7, no. 2, pp. 139–172, 2020, doi: 10.1080/23270012.2020.1756939.
- [4] K. Li, C. Zhou, X. (Robert) Luo, J. Benitez, and Q. Liao, “Impact of information timeliness and richness on public engagement on social media during COVID-19 pandemic: An empirical investigation based on NLP and machine learning,” *Decis. Support Syst.*, no. June 2021, p. 113752, 2022, doi: 10.1016/j.dss.2022.113752.
- [5] J. Deng and F. Ren, “A Survey of Textual Emotion Recognition and Its Challenges,” *IEEE Trans. Affect. Comput.*, vol. 3045, no. c, pp. 1–20, 2021, doi: 10.1109/TAFFC.2021.3053275.
- [6] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” *Eng. Reports*, vol. 2, no. 7, pp. 1–24, 2020, doi: 10.1002/eng2.12189.
- [7] T. M. Li, H. C. Chao, and J. Zhang, “Emotion classification based on brain wave: a survey,” *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0201-x.
- [8] N. Alswaidan and M. E. B. Menai, “A survey of state-of-the-art approaches for emotion recognition in text,” *Knowledge and Information Systems (2020) 62*, 2020.
- [9] Muljono, A. S. Winarsih, and C. Supriyanto, “Evaluation of classification methods for Indonesian text emotion detection,” *Proc. - 2016 Int. Semin. Appl. Technol. Inf. Commun. ISEMANTIC 2016*, pp. 130–133, 2016, doi: 10.1109/ISEMANTIC.2016.7873824.
- [10] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, “Comparing automated text classification methods,” *Int. J. Res. Mark.*, vol. 36, no. 1, pp. 20–38, 2019, doi: 10.1016/j.ijresmar.2018.09.009.
- [11] R. C. W. Fera Fanesya Indriati, “Deteksi Emosi pada Twitter Menggunakan Metode Naive Bayes dan Kombinasi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, p. 3,

- 2019.
- [12] Y. Fauziah, B. Yuwono, and A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review," *RSF Conf. Ser. Eng. Technol.*, vol. 1, no. 1, pp. 363–367, 2021, doi: 10.31098/cset.v1i1.397.
- [13] J. Bata, "# AkuGalau: Korpus Bahasa Indonesia untuk Deteksi Emosi dari Teks," *J. Elektro*, vol. 12, no. 2, pp. 103–110, 2019.
- [14] M. A. Azim and M. H. Bhuiyan, "Text to emotion extraction using supervised machine learning techniques," *Telkonnika (Telecommunication Comput. Electron. Control.)*, vol. 16, no. 3, pp. 1394–1401, 2018, doi: 10.12928/TELKOMNIKA.v16i3.8387.
- [15] D. William and D. Suhartono, "Text-based Depression Detection on Social Media Posts: A Systematic Literature Review," *Procedia Comput. Sci.*, vol. 179, no. 2019, pp. 582–589, 2021, doi: 10.1016/j.procs.2021.01.043.
- [16] M. S. Saputri, R. Mahendra, and M. Adriani, "Emotion Classification on Indonesian Twitter Dataset," *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, no. November, pp. 90–95, 2019, doi: 10.1109/IALP.2018.8629262.
- [17] P. Ekman, "An Argument for Basic Emotions," *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.
- [18] J. E. The, A. F. Wicaksono, and M. Adriani, "A two-stage emotion detection on Indonesian tweets," *ICACSIS 2015 - 2015 Int. Conf. Adv. Comput. Sci. Inf. Syst. Proc.*, pp. 143–146, 2016, doi: 10.1109/ICACSIS.2015.7415174.
- [19] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: a review," *Eurasip J. Wirel. Commun. Netw.*, vol. 2017, no. 1, pp. 1–12, 2017, doi: 10.1186/s13638-017-0993-1.
- [20] V. Sundaram, S. Ahmed, S. A. Muqtadeer, and R. Ravinder Reddy, "Emotion Analysis in Text using TF-IDF," *Proc. Conflu. 2021 11th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 292–297, 2021, doi: 10.1109/Confluence51648.2021.9377159.
- [21] M. Das, S. Kamalanathan, and P. Alphonse, "A Comparative Study on TF-IDF feature weighting method and its analysis using unstructured dataset," *CEUR Workshop Proc.*, vol. 2870, pp. 98–107, 2021.
- [22] S. S. and P. K.V., "Sentiment analysis of malayalam tweets using machine learning techniques," *ICT Express*, vol. 6, no. 4, pp. 300–305, 2020, doi: 10.1016/j.icte.2020.04.003.
- [23] N. Kholodna, V. Vysotska, and S. Albota, "A machine learning model for automatic emotion detection from speech," *CEUR Workshop Proc.*, vol. 2917, pp. 699–713, 2021.
- [24] A. Onan, "Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach," *Comput. Appl. Eng. Educ.*, vol. 29, no. 3, pp. 572–589, 2021, doi: 10.1002/cae.22253.
- [25] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Syst. Appl.*, vol. 134, pp. 93–101, 2019, doi: 10.1016/j.eswa.2019.05.028.
- [26] P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 3, pp. 1–15, 2019, doi: 10.1002/widm.1301.
- [27] H. Luzern, *Cyber Security Cyber Security*, no. March. 2017. doi: 10.1007/978-981-16-9229-1.
- [28] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, "Lexicon based feature extraction for emotion text classification," *Pattern Recognit. Lett.*, vol. 93, pp. 133–142, 2017, doi: 10.1016/j.patrec.2016.12.009.
- [29] J. Akram and A. Tahir, "Lexicon and Heuristics Based Approach for Identification of Emotion in Text," in *2018 International Conference on Frontiers of Information Technology (FIT)*, Dec. 2018, pp. 293–297. doi: 10.1109/FIT.2018.00058.
- [30] D. D. Palmer, "Text pre-processing," *Handb. Nat. Lang. Process. Second Ed.*, pp. 9–30, 2010, doi: 10.4018/978-1-5225-4990-1.ch006.
- [31] Y. Nooryuda Prasetya and D. Winarso, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," *J. Fasilkom*, vol. 11, no. 2, pp. 97–103, 2021.
- [32] V. V. Nhlabano and P. E. N. Lutu, "Impact of Text Pre-Processing on the Performance of Sentiment Analysis Models for Social Media Data," *2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. icABCD 2018*, pp. 1–6, 2018, doi: 10.1109/ICABCD.2018.8465135.
- [33] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, W. Alnumay, and A. P. Smith, "A lexicon-based approach to detecting suicide-related messages on Twitter," *Biomed. Signal Process. Control*, vol. 65, no. November 2020, 2021, doi: 10.1016/j.bspc.2020.102355.
- [34] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis DRAFT DRAFT DRAFT!," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.
- [35] F. Calefato, F. Lanubile, and N. Novielli, "EmoTxt: A toolkit for emotion recognition from text," *2017 7th Int. Conf. Affect. Comput. Intell. Interact. Work. Demos, ACIIW 2017*, vol. 2018-Janua, no. July, pp. 79–80, 2018, doi: 10.1109/ACIIW.2017.8272591.
- [36] S. Liu and I. Lee, "Extracting features with medical sentiment lexicon and position encoding for drug reviews," *Heal. Inf. Sci. Syst.*, vol. 7, no. 1, 2019, doi: 10.1007/s13755-019-0072-6.
- [37] A. Bandhakavi, N. Wiratunga, P. Deepak, and S. Massie, "Generating a word-emotion lexicon from #emotional tweets," *Proc. 3rd Jt. Conf. Lex. Comput. Semant. *SEM 2014*, pp. 12–21, 2014, doi: 10.3115/v1/s14-1002.
- [38] A. Gholipour Shahraiki and O. R. Zaiane, "Lexical and Learning-based Emotion Mining from Text," *Univ. Alberta, Canada*, pp. 24–55, 2017.
- [39] K. S. Nugroho, "Detecting Emotion in Indonesian Tweets : A Term-Weighting Scheme Study," vol. 8, no. 1, pp. 61–70, 2022, [Online]. Available: <https://www.e-journal.unair.ac.id/JISEBI/article/view/33017>
- [40] F. Koto and G. Y. Rahmangtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, 2018, vol. 2018-Janua, no. December, pp. 391–394. doi: 10.1109/IALP.2017.8300625.
- [41] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," in *CEUR Workshop Proceedings*, 2011, vol. 718, no. March, pp. 93–98.
- [42] R. N. Waykole and A. D. Thakare, "International Journal of Advance Engineering and Research A REVIEW OF FEATURE EXTRACTION METHODS FOR TEXT," pp. 351–354, 2018.
- [43] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-Based Sentiment Analysis of Teachers' Evaluation," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, pp. 1–12, 2016, doi: 10.1155/2016/2385429.
- [44] F. A. Bachtiar, W. Paulina, and A. N. Rusydi, "Text Mining for Aspect Based Sentiment Analysis on Customer Review : a Case Study in the Hotel Industry," *5th Int. Work. Innov. Inf. Commun. Sci. Technol.*, no. March, 2020.