

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION



journal homepage: www.joiv.org/index.php/joiv

Mining Opinions on a Prominent Health Insurance Provider from Social Media Microblog: Affective Model and Contextual Analysis Approach

Ihda Rasyada^a, Aliridho Barakbah^{a,*}, Elizabeth Anggraeni Amalo^a

^a Department of Information and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Jl. Raya ITS, Surabaya, 60111, Indonesia Corresponding author: ^{*}ridho@pens.ac.id

Abstract—Social media plays a significant role in enhancing communication among organizations, communities, and individuals. Besides being a mode of communication, the data generated from these interactions can also be leveraged to assess the performance of an institution or organization. People may evaluate public companies based on the opinions of their users. However, user-supplied information is brief and written in natural language. In addition to being brief, the process of sending messages or engaging in other social media interactions contains a great deal of context information. This multiplicity of context can be utilized to conduct a more indepth analysis of user opinion. This study presents a new approach to opinion mining for social media microblogging data by applying an affective model and contextual analyses. The affective model is applied for sentiment analysis to measure the degree of each adjective from user opinion by evaluating adjectives according to their varying levels of pleasure and arousal. The contextual analysis in this paper is modeled based on topic, user, adjective, and personal characteristics. The contextual analysis has four main features: (1) Temporal keyword sentiment context, (2) Temporal user sentiment context, (3) User impression context, and (4) Temporal user character context. Our affective model outperformed 75.6% the accuracy and 74.98% of F1-score, rather than SVM. In the experiment, the contextual analysis performed graph visualization of output results for each query feature for future development. Feature one to four successfully processes the query to produce a visualization graph.

Keywords— Affective model sentiment analysis; health context analysis; health opinion mining.

Manuscript received 28 Apr. 2023; revised 16 May 2023; accepted 6 Jun. 2023. Date of publication 30 Jun. 2023. International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

One way to categorize social media is by the technology used; microblogging is an instance of this. Microblogging is one kind of social media with a fast way of exchanging information [1]. The discussion should bridge the results and the literature review in the introduction. Twitter is a popular microblogging platform utilized by internet users in Indonesia. According to a Hootsuite survey (We are Social), there are 160 million social media users, of which 52% are Twitter users [2]. Twitter has a high rate of information exchange, allowing users to send and receive information on a topic more quickly. The enormous amount of information that can be gleaned from this medium enables an organization or institution to analyze user opinions regarding a product or service [3].

The case study for this research is Health Social Insurance, whose membership comprises 83.89% of the total population

of Indonesia [4]. Members of Health Social Insurance send compliments, complaints, and criticisms about its service through social media platforms such as Twitter. The Health Social Insurance can use this information to evaluate their service. As data increases, manual and one-by-one analysis of social media data will require more time and effort. This will lengthen the enhancement or decision-making process. Therefore, Sentiment Analysis can ease the job of manual analyzers. Sentiment Analysis is a technique that can classify comments as positive, negative, or neutral. While sentiment analysis does ease the process of analyzing commentary data, developing this Sentiment Analysis has several difficulties, such as language, sentence structure, and information delivery.

When communicating their opinions, one part of speech is used: adjectives that describe nouns; in practice, they can also describe a topic. The emotional value of adjectives used in sentences may vary depending on various factors, such as: 1) Some adjectives that have similar meanings may differ in their pleasure or displeasure values, such as "tidak suka" (dislike) and "benci" (hate), or "suka" (like) and "cinta" (love), which have different emotional intensities when used in a sentence [5]. 2) Adjectives can be followed by comparative words like more, less, and very. If the adjective is followed by a comparison word, the word's emotional depth value in the sentence will change. 3) The word "no" can be used to negate adjectives such as "not right," "not wrong," and "not good."

Even though the message is brief, the process of sending messages or other interactions on social media contains a lot of context information [6]. This multiplicity of context can be used to deepen analysis or augment sentiment analysis with additional features. For instance, a user's opinion on a topic will remain consistent over time [7], and an opinion may consist of one or more topics.

A. Sentiment Analysis

There are two approaches to analyzing sentiment: machine learning and lexicon-or rule-based analysis. Lexicon-based or rule-based methods are widely used in research because they can be applied to a variety of subjects. Combining several English lexical [8], [9] or creating one's own dictionaries for other languages based on sentence structure and word type [10]–[12] can be governed by rules. The sentiment analysis using rules-based derived from previous research does not account for the fact that different types of adjectives have distinct levels of pleasure and arousal. Russell's study on adjectives indicates the distinction between pleasure and arousal between adjectives [5]. The Russell Circumplex Model can be employed in sentiment analysis methods by utilizing the degree of adjectives. Several studies have explored the use of the Russell Circumplex Model in sentiment analysis. For instance, Erion and Morisio [13] conducted sentiment analysis on English song lyrics using the Circumplex Model. Naskar et al. [14] researched various witter topics using Russel's Circumplex Model and SentiWordnet; their technique is known as ANEW. All of these studies utilize English language data. However, Khodijah et al. [15] utilized a technique to automatically determine the level of pleasure and arousal in the Indonesian language. However, sentiment analysis has not been utilized thus far in the research.

B. Context Analysis

The methods of Sentiment Analysis utilized in the aforementioned studies [1], [8], [9], [16]–[19] were primarily based on textual content. Microblogging texts can comprise multiple topics with varying degrees of emotional weight, rendering microblogging messages context-specific. As explained by Kumar and Garg [16], the context in sentiment analysis refers to any supplementary source of evidence that can augment or reverse the polarity of the content. This study will discuss contexts based on topics (content), users, and adjectives. The four types of contexts will be applied based on the time of the user's query.

The types of context analysis studied include topic-based [1], [17]–[19], temporal-based [19], and user-based [1], [19]. Fangzhao [1] and Liu [17] conducted research utilizing context-based topics. Fangzhao [1] connects topics to thuser's and their friends' context. The topics were extracted from the Twitter hashtag (#). User context means that users typically

share the same viewpoints on particular topics over time. Whereas the friend context indicates that both accounts share the same viewpoint on a particular subject. To determine the sentiment value of a Twitter message, all contexts are modeled in a graph. Liu [17] obtained topics from website product labels. The method is known as the BaseLine Algorithm; text will be divided by a defined topic, with each word playing a significant role in the sentiment value of the text; for instance, "sound of the speaker is good in every way! But the battery runs out quickly. "Positive sentiment is associated with the speaker topic, while negative sentiment is associated with the battery topic. A topic can also be described by adjectives with contradictory values, such as "high performance" (positive) and "high maintenance" (negative). It implies that adjectives rely on the context in which the user is operating. Both Fangzhao's and Liu's works can enhance the precision of sentiment analysis. However, Fangzhao only uses topics based on hashtags; while using Liu's research to determine the sentence's focus, messages without hashtags can still be searched based on their text.

Abudalfa and Ahmed [18] used a microblogging message as a labeling token to identify all named entities about a person, organization, etc. Vanzo [19] conducted a study to increase the precision of sentiment analysis by developing a context-sensitive method. The first step is to create a social graph describing the relationship between a tweet and its reply; this will normalize user conversation. The following step involves classifying the model. The SVMhmm learning algorithm classifies tweets in the preceding social graph sequences, whereas the SVM method classifies tweets individually. Using the ColMustard method, the study also investigates the user's emotional state. With the contextual information combination described above, the presentation accuracy of tweets with multiple topics and contextual conversations is increased by 20%.

The research by Kumar and Garg [16] describes the benefits and drawbacks of utilizing context analysis to enhance the quality of sentiment analysis. The difference in the number of studies between these context types is quite striking, with most studies utilizing content-based contexts. Nevertheless, some studies include multiple contexts. As described by Kumar and Garg [16], describing the contextual text is crucial in sentiment analysis when identifying the context of the situation, specific topics, and the environment.

II. MATERIAL AND METHOD

This research proposes analyzing sentiment using the Affective Model Method based on the Russell Circumplex Model and applying context analysis based on topics (content), users, adjectives, and personal characteristics. Topics context is called temporal keyword sentiment context, which generates a temporal graph of sentiment values in the context of keywords. User context is called temporal user sentiment context, which generates a temporal graph of sentiment values in the context of the user against time. Adjective context is called user impression context, which generates a graph of the word impression value on a user. Character context is called temporal user character context, which generates the character analysis of a user.

Fig. 1 illustrates the general system design. The process commences with data collection via the Twitter crawler,

followed by preprocessing, text mining, and sentiment analysis utilizing the Affective Model. Upon the completion of the analysis, the subsequent step involves generating a context space for the four-function or feature context analysis process.



Fig. 1 System Design

A. Crawler

The Twitter Crawler is a program that uses the Twitter Developer API to retrieve information from Twitter. API (Application Programming Interface) serves as a bridge between the server and the application. API Twitter Developer is therefore used to retrieve data from the Twitter server. This API will retrieve data from users who mention Health Insurance in their Twitter accounts. The Twitter API searches for the most recent tweet published within the past week.

This crawler returns JSON data containing detailed attributes. Table 1 displays the primary attributes that will be used in the system. This process yields a total of 30532 tweets, but only 1064 tweets are utilized in the experiment. This data will also be evaluated for accuracy in terms of sentiment and context analysis.

TABLE I
DATA ATTRIBUTES

No.	Attribute Name	Function
1	id	Id data
2	created at	Data creation time
3	full text	User opinion (text)
4	user.screen name	User name

B. Pre-processing

Twitter text data typically contains numerous non-word components and varies in format. This preprocessing eliminates unnecessary components and standardizes the format. Delete Symbol is the initial phase of preprocessing. In the text, unnecessary symbols will be removed. The words that come after hashtags (#) and mentions (@) symbols will be eliminated along with the symbols themselves. For example, the phrases @AccountName and #tolakkenaikanpremi will be deleted. Some punctuation is not eliminated because it influences the process of calculating the degree of emotion. The dots, commas, question marks, and exclamation points are not to be removed. These symbols are text delimiters, indicating that the data contains multiple sentences or clauses. These characters will be removed:

The Number Eraser is the second phase of the preprocessing procedure. Sometimes numerous connotations are attached to numbers in sentences, but numbers are not used in this study. In addition to standalone numbers, this procedure will eliminate numbers that are consistently positioned before and after words. The final phase is case transformation. This procedure aims to convert all text to lowercase so that the resulting query has a uniform format.

C. Text Mining

Jalal [20] stated that Text Mining involves extracting patterns, useful information, and knowledge from text data resources. A single comment can comprise multiple sentences or a single sentence can contain multiple clauses. Additionally, each sentence or clause in a given text data may express different sentiments. Data is separated by dots, commas, question marks, or exclamation points to carry out text mining.

The initial step in text mining is Tokenizing. Tokenizing is the process of separating the white-space-separate text into tokens. Words, numbers, punctuation, and emoticons can be considered as tokens. In this case, however, a token consists only of textual words.

The second phase is referred to as the Filtering phase, which involves eliminating insignificant words, commonly referred to as stopwords. For example, in this study, the author will generate specific stopwords, for example, "bahwa," "apabila," and "bagaimana," because some existing stopwords in other libraries contain multiple words that will be used in the Sentiment Analysis process. The final stage of text mining is stemming, which involves reducing a word to its basic form by removing affixes and converting it into its root form.

D. Sentiment Analysis

According to Liu [21], sentiment analysis involves examining people's viewpoints, emotions, judgments, recognition, and feelings towards entities like products, services, individuals, organizations, issues, and events. The Affective Model, used in this study for analyzing sentiment, is a method that focuses on the value differences between adjectives in a text. This adjective's values are based on research conducted by Russell [5], and Rasyada et al [22]. The three main components of the Affective Model are Affective space, polarity of sentiment, and degree of affection. These three components are used to calculate the sentiment degree value from text data.



Fig. 2 Coordinates for 28 words by direct circular scaling [5]

1) Affective Space: The Affective Space is a matrix containing Indonesian's "pleasure-arousing" adjective values. This matrix shows that adjectives have different values of "pleasure-arousal," such as "happy" and "delighted." In Fig. 2, the polar coordinate depicts 28 adjectives that have been mapped by Russell, including happy, delighted, excited, astonished, aroused, tense, alarmed, angry, afraid, annoyed, distressed, frustrated, miserable, sad, gloomy, depressed, bored, droopy, tired, sleepy, calm, relaxed, satisfied, at ease, content, serene, glad, and pleased. Horizontal axes can be interpreted as proposed dimensions of pleasure and displeasure. Both pleasure and displeasure are contradictory human emotions. The closer an affective word is to the east, the greater its positive pleasure value, and the closer it is to the west, the greater its negative displeasure value. The vertical axis can be interpreted as each adjective's proposed level of arousal. The greater the affective word's position to the north, the greater the passion. The greater a word's position to the south, the less passionate it is.

Each of the 28 adjectives should have a degree value, but 13 of them are missing from Russell's list. 13 words do not specify the degree value: angry, afraid, annoyed, distressed, frustrated, gloomy, depressed, bored, satisfied, at ease, content, glad, and relaxed [5]. The angles of these 13 words can be determined using a protractor. Once the degree values for all adjectives have been determined, the next step is to measure each adjective's radius (r). Degree value (α) and radius are used to calculate adjectives x and y values.

By utilizing the trigonometric ratio cosine and equation (1), the value of x can be determined.

$$\cos \alpha = \frac{x}{r} \tag{1}$$

By employing the trigonometric ratio sine and equation (2), the value of y can be ascertained.

$$\sin \alpha = \frac{y}{r} \tag{2}$$

The values of x represent the pleasure-displeasure of the adjectives. The values of y represent the degree of arousal of the adjectives. These two dimensions can be utilized to determine the degree of sentiment for each adjective. In view of their low correlation, these two affective dimensions have been regarded as separate or orthogonal [23]. To determine

the sentiment value of each word according to Russell's quadrant, equations (3) and (4) are applied. If adjectives are in quadrants two or three, then equation (4) is applied, as the term has a negative sentiment. If adjectives are in quadrants 1 or 4, equation (3) is applied.

$$d = x - y - 5 \tag{3}$$

$$d = x + y + 5 \tag{4}$$

The sentiment values will be normalized as they fall within the range of -1 to 1. The obtained sentiment values are normalized using minmax in equation (5), which has a range of -1 to 1.

$x' = \frac{(x - x_{min})}{(x_{max} - x_{min})} * (new_{max} - new_{min}) + new_{min}$					
TABLE II					
A FEE CTUVE SDACE					

	А	FFECTIVE SI	ACE		
Value	Delighted	Annoyed	Satisfied	Miserable	Alarmed
gelisah	0	0	0	0	-0.9
gembira	0.91	0	0	0	0
malang	0	0	0	-0.8	0
jengkel	0	-0.95	0	0	0
puas	0	0	0.47	0	0
konsisten	0	0	0.47	0	0

The sentiment values of Russell's adjectives are the benchmark for the sentiment values of Indonesia's adjectives [24]. Indonesia's adjectives have sentiment value because they are mapped to Russell adjectives. Affective Space is a matrix containing Indonesian adjectives with varying degrees of sentiment and different pleasure-displeasure values and arousal levels. Table 2 displays an excerpt of Affective Space. Given the value of alarmed, "gelisah" has a value of -0.9. Based on the word delighted, "gembira" is worth 0.91. "Malang" has a value of -0.8 based on miserable. Based on the value of annoyed, "jengkel" has a value of -0.95. And based on the value of satisfied, the value of "puas" is 0.47.

2) Polarity of Sentiment: The polarity of a sentence can be either positive or negative. The negation word indicates negative polarity. The term negation word refers to the word for denial. Negation words that can be found in Indonesian are are "bukan," "tidak," "jangan," and "belum." Observations indicate that many social media users employ nonstandard word usage. Therefore, we have added non-standard negation words such as "ga," "nggak," "gak," and "bkn" to the system.

3) Degree of Affection: The degree of affection is a value that decreases or increases the degree of sentiment in a sentence. Table 3 displays the words that indicate the degree of affection in a sentence, as identified by observing and examining the Indonesian dictionary [24]. For words that indicate enhancement of sentiment, a score of +0.5 is given. For words that indicate the reduction of sentiment, a score of -0.5 is given.

	TABLE III	
	DEGREE OF AFFECTION	
Word		Value

sangat

Sungui	10.5
paling	+0.5
sekali	+0.5
amat	+0.5
terlalu	+0.5
agak	-0.5
kurang	-0.5
sedikit	-0.5
minim	-0.5

To calculate the degree of the sentiment of a tweet is based on the three components of the Affective Model: (1) finding adjectives in each sentence or clause; (2) determining whether the adjectives have words in the element of polarity of sentiment; and (3) determining whether the adjectives have words in the degree of affection. For example, "saya sangat gelisah menunggu pembayaran iuran yang belum diproses". Suppose the word "gelisah" means to have -0.9, then there is no negation word, but it has a degree of affection word. The word "sangat" appears in the sentence, so the value of $(0.9+0.5)^*-1$ is -1.4. The sentiment, therefore, is -1.4.



Fig. 3 Character Analysis

E. Character Analysis

Character Analysis is a method for determining a user's personality based on the Big Five Personality Traits. They are Openness, Conscientiousness, Experience, Extraversions, Agreeableness, and Neuroticism [25]. This big five personality has been a model for categorizing personalities in social media data[25]–[27]. Fig. 3 shows the process of character analysis. The process starts from data preprocessing in the previous stage. The output of data preprocessing is a token from the user's tweet. The token will be extracted using the NRC Emotion Lexicon, which includes eight fundamental emotions from Plutchik's Emotion Wheel. The outcome of feature extraction will be correlated with Big Five Personality Metadata to determine the personality value range.

1) NRC Emotion Lexicon: Based on Plutchik's theory and the sentiment value of words, The NRC Emotion Lexicon was developed by Mohammad and Turney [28] and comprises a compilation of English words and their corresponding associations with eight fundamental emotions: fear, surprise, sadness, disgust, anger, anticipation, and joy. There are numerous translations of the NRC Emotion Lexicon's terms, including Indonesian. This study will utilize the NRC Emotion Lexicon with Indonesian words associated with Plutchik's eight basic emotions to extract emotion characteristics from a user's tweet. Maharani [29] conducted research to determine the correlation between personality characteristics and behavior on social media by integrating the big five and the NRC lexicon.

2) Emotion Features Extraction: One of the results of data preprocessing is tokens. Tokens from this user's tweet will be matched with NRC Emotion Lexicon to produce the result of emotion feature extraction. This attraction process is carried out temporally based on a set of user tweets. The process of feature attraction begins with the extraction of emotions for each token from tweets and concludes with the aggregation of emotions from all tweets using binary normalization.

3) Correlation Measurement: This method uses metadata, which is the correlation of the eight basic emotions of the Plutchik Wheel of Emotions with the Big Five Personality [28]. The results of the emotion feature extraction in the previous process were used to calculate the correlation measurement with Big Five Personality Metadata [30] using the Inner Product method.

F. Context Analysis

In this study, context in sentiment analysis refers to supplementary source of evidence that can enhance or reverse the polarity of the content. Context Analysis in this study describes the context of keywords used by users, the context of users who tend to have the same opinion over a period, and the context of impression words (adjectives).



Fig. 4 Temporal Keyword Sentiment context

1) Temporal Keyword Sentiment Context: This feature generates a temporal graph of sentiment values in the context of keywords. The graph displays the average sentiment value for each keyword over the specified time period. This graph allows users to view the opinions of others regarding an object. User-supplied keywords and a time period are utilized in the process. The steps for this function are depicted in Fig. 4. Each unit of time (t) contains tweet data from users, and each time can contain multiple tweets. Each of these tweets contains a word (w) attribute derived from the text mining result and a sentiment value derived from the sentiment analysis result. The value of each w is the sentiment value of the original tweet. Fig. 4 shows that in t_1 , there are two tweets; tweet₁ has $[w_1, w_2, w_3]$, and the value of these words is s_1 . tweet₂ has $[w_1, w_3]$ and the value of these words is s_2 . To determine the sentiment value of the keyword context at a time (t_1) , the average value of the keyword $[w_1, w_2, w_3]$ in tweet₁ and tweet₂ is used. This average value is calculated every time (t_n) based on the time period inputted by the user. The value of each w at each time (t) is combined into a single matrix. From this matrix, sentiment value graphs based on

user-entered keywords can be generated. This graph displays the sentiment associated with the keywords entered by the user.



Fig. 5 Temporal User Sentiment Context

2) Temporal User Sentiment Context: This feature generates a temporal graph of sentiment values in the context of the user. The result shown in the graph is the average sentiment value for every tweet that a user sends over a specified time period. This feature aims to show user sentiment towards the subject service over a certain period of time. The process uses user input such as username account and time period from the user. The steps for this function are depicted in Fig. 5.

The system retrieves tweet data from a user and divides it by time (t). Each unit of time (t) may contain multiple tweets or none at all. Each of these tweets has impressions or adjective values (I) and sentiment values from the sentiment analysis process. The value of each I is the sentiment value from Affective Space in Table 2. Fig. 5 shows that t_1 has two tweets: (1) tweet₁ containing $[I_1, I_2]$ and the sentiment value is -1, and (2) *tweet*₂ contains $[I_3]$ and the sentiment value is -1. The sentiment value of a user at a given time (t_1) is taken from the average value of sentiment and impression word value in tweet₁ and tweet₂. This average value is calculated each time (t_n) based on the user-specified time period. The result of the value of each w at each time (t) is combined into one matrix. From this matrix, a sentiment value based on a user's sentiment can be made. This graph can show whether a user's sentiment is the same or not on the subject's service.



Fig. 6 User Impression Context

3) User Impression Context: This feature generates a graph of the value of a user's word impression. The graph represents the average sentiment value for each impression word over a specified time period. The procedure utilizes user-supplied impressions or adjective word inputs such as "kecewa," "marah," or "ramah" and time period. The steps for this feature are shown in Fig. 6.

All users in the database are logged in for each tweet, as well as the sentiment value attributes and impression words used. Fig. 6 shows Ul has two tweets: (1) *tweet*₁ has sentiment value s_1 and impression words $[I_1, I_2, I_3]$, and (2)

*tweet*² has sentiment value s_2 and impression words $[I_1, I_3]$. The value of $[I_1, I_2, I_3]$ in *tweet*¹ is s_1 , and the value of $[I_1, I_3]$ in *tweet*² is s_2 . To get the sentiment value of impression words on a user (U1), which is taken from the average value of the impression word $[I_1, I_2, I_3]$ in *tweet*¹ and *tweet*². This average value is calculated for every user in the database (U_n) . The result of the value of each *I* for each user is combined into one matrix. From this matrix, sentiment value based on impression words input by the user can be derived.

4) Temporal User Character Context: This feature visually represents a user's character over time based on their tweets. The value displayed in the graph is the result of their character analysis. The procedure employs user input as a username account with a daily time interval. This feature employs the method for character analysis described in the previous section.

III. RESULT AND DISCUSSION

The experimental research methods were used to develop the system method in this study. The system begins by identifying and cleansing the data utilized for data preprocessing. The second step is sentiment analysis using the Affective Model. The third step is the context detection process for storing keywords, impressions, and sentiment value in the database. The final step is context analysis, which has four features. The system is developed using the programming language Python. The use of libraries such as pandas, NumPy, JSON, re, and matplotlib simplifies the coding process. The experiment begins with data preprocessing, followed by sentiment analysis and context analysis.

A. Sentiment Analysis

The outcome of the sentiment analysis method is the degree of sentiment of a comment using Affective Space, which contains 144 Indonesian adjectives. This sentiment's performance outcomes are accuracy, precision, recall, and F1score. These four-performance metrics are derived from the confusion matrix, a table containing four unique predicted and actual value combinations[31]. In the confusion matrix, four terms represent the outcome of the classification process: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Accuracy is the proportion of accurate predictions to total data. Equation (6) is the accuracy formula. Precision is the ratio of correct positive predictions to total positive predictions. Equation (7) is the precision formula. Recall is the proportion of accurate predictions compared to the total number of accurate data. Recall is represented in equation (8). Due to the fact that the performance test has three measurements, precision, recall, and F1-score helps to represent the precision and recall measurements. The F1-score compares the averages of precision and recall. Equation (9) is the F1-score formula.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$
(6)

$$Precision = \frac{TP}{TP+FP}$$
(7)

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - score = \frac{2*recall*precision}{recall+precision}$$
(9)

TABLE IV PERFORMANCE ANALYSIS OF AFFECTIVE MODEL

Performance Analysis	Value
Accuracy	75.6%
Precision	80.22%
Recall	75.6%
F1-Score	74.98%

Affective model sentiment analysis experiments utilize 250 tweet data samples. Table 4 displays the results of the study's sentiment analysis. The Affective Model method has an accuracy of 75.6%. The precision value is 80.22%, recall value is 75.6% and F1-Score value is 74.98%. The results of this experiment are not satisfactory because the value of performance test is not too high. Recall that is lower than precision indicates that the system has low accuracy in predicting positive data labels that match the actual label.

TABLE V Confusion matrix				
A streal		Predict	ed	
Actual	-1	0	1	
-1	120	2	40	
0	7	3	10	
1	1	1	66	

The data result has three types of labels: positive, negative, and neutral. To calculate precision, recall, and F1-Score, each label is extracted from the confusion matrix. Table 5 displays the outcome of this experiment's confusion matrix. This confusion matrix can be used to calculate the TP, TN, FP, and FN values for each label. For instance, the value of TP in label -1 is the number of correct predictions predicted in label -1. TN value is the number of predictions that correctly predicted a class other than -1. FP value is the number of incorrect predictions of the value -1. FN represents the number of incorrect predicted outcomes, while the other classes predicted -1.

TABLE VI Value of TP. TN. FP. FN

VALUE OF IP, IN, FP, FN						
Label	ТР	TN	FP	FN		
-1	120	(3+10+1+66) = 80	(2+40) = 42	(7+1) = 8		
0	3	(120+1+40+66) =227	(7+10) = 17	(2+1) = 3		
1	66	(120+2+7+3) = 132	(1+1) = 2	(40+10) = 30		
TABLE VII PERFORMANCE ANALYSIS FOR FACH LAREL USING AFFECTIVE MODEL						

Label	Precision	Recall	F1-Score	
-1	74.07%	93.75%	82.82%	
0	15%	50%	23.07%	
1	97.05%	68.75%	80.4%	

TP, TN, FP, and FN values for each label are shown in Table 6. Precision, recall, and F1-score values for each label can be calculated using Eq.(7), Eq.(8), and Eq.(9). As shown in Table 7, the recall value for labels 0 and 1 is lower, indicating that they are less likely to be correctly predicted. In contrast, the precision and recall values for the label -1 class are both high, indicating that the system has successfully identified label -1 with a high predictive value. The fact that the recall value of label -1 is greater than its precision

indicates that label -1 makes few false negative predictions. The number of false negatives in the -1 label class, which is a little bit, indicates that only a small number of 0 and 1 label classes are incorrectly predicted to be -1.

TABLE VIII Performance analysis of SVM

Performance Analysis	Value
Accuracy	76.19%
Precision	64.86%
Recall	76.19%
F1-Score	69.38%

This experiment compares both the Affective Model and SVM methods. In SVM classification, 250 data are used in total. The data will be separated into data for training and data for testing. There are 63 test data and 187 training data sets used in the experiment. The accuracy of SMV is 76.19%, the precision value is 64.86 percent, recall value is 76.19 percent, and F1-Score is 69.38 percent, as shown in Table 8. When the Affective Model is compared with SVM for the same dataset, SVM has slightly higher accuracy and a lower F1-score. A larger F1-score value indicates the Affective Model has a good harmonic between precision and recall value.

B. Temporal Keyword Sentiment Context

The Temporal Keyword Sentiment Context system is deemed successful if it can provide a response based on a user's query and display a visualization graph. There is a set of queries with expected correct answers, i.e., the collection of queries is processed by the system, and the system's output is compared with the correct answers derived from the observations. An example of a user query for this feature experiment is provided in Table 9.

TABLE IX User query feature 1						
No.	Start Date	Start Time	End Date	End Time	Keyword	
1.	2020- 01-01	00.00	2020- 01-01	23.00	aplikasi, mobile, pembayaran	
3.	2020- 01-08	08.00	2020- 01-09	23.00	uang, rakyat, and iuaran	

The first query is for data between 00:00 and 23:00 on 2020-01-01 containing the keywords "aplikasi," "mobile," and "pembayaran." The second query is data from 2020-01-08 at 08.00 until 2020-01-09 at 23.00 containing keywords "uang," "rakyat," and "iuran."

This system's time interval is 60 minutes. With the initial query in Table 9, the system retrieves 265 records and a time range between t0 and t22. A text mining process executes the keyword query to generate words that match keywords in the matrix system. The first query keywords are "aplikasi," "mobile," and "pembayaran." This keyword seeks to determine the attitude towards the mobile application's payment system. Fig. 7 depicts the graph for the first query. During t1, the sentiment towards the keyword "mobile" is positive, as shown in Fig 7. (2020-01-01:01.00 to 2020-01-01:02.00). At t6, the sentiment value of the keyword "aplikasi" is also positive. At t3 and t15, sentiment is negative towards "pembayaran" keywords. From Fig. 7, we can conclude that the Health Insurance mobile application has a

positive sentiment while the payments have a negative sentiment, and the three keywords also do not intersect with each other.



Fig. 7 First Query's Graph of Temporal Keyword Sentiment Context



Fig. 8 Second Query's Graph of Temporal Keyword Sentiment Context

The second query in Table 9 returns 80 records and a time interval between t0 and t9. These are the keywords: "uang," "rakyat," and "iuran." The keywords seek to determine the sentiment of users who emphasize their opinions on the Health Insurance program. Fig. 8 shows a graph of the second query's visualization. From Fig. 8, we can conclude that the payment of Health Insurance has a negative sentiment, while "rakyat" has a positive sentiment and that the three keywords do not overlap.

	USER QUERY FEATURE 2				
No.	Start	Start	End	End	Username
	Date	Time	Date	Time	
1.	2020-02-	00.00	2020-	23.00	'user_l'
	26		02-26		_
2.	2020-01-	18.00	2020-	06.00	'user_s'
	05		01-06		_

TABLE X USER OUERY FEATURE 2

C. Temporal User Sentiment Context

The Temporal User Sentiment Context system will be successful if it can provide a response based on a user's query and display a visualization graph. There is a set of queries with expected correct answers, i.e., the system processes the collection of queries, and the system's output is compared with the correct answers derived from the observations. Table 10 contains examples of user queries for this feature. The first query is for data between 00 and 23:00 on 2020-02-26 for user 'user_1'. The second query is for data between 2020-01-05 18:00 and 2020-01-06 06:00 applicable to user 'user s'.



Fig. 10 Second Query's Graph of Temporal User Sentiment Context

The time interval for this feature is 60 minutes, the same as the first feature. From query number one in Table 10, the system obtains 3 data from user 'user_l' and a time range between t0 and t23. The result of this query is shown in Fig. 9. Three obtained data are at time t14. The graph in Fig. 9 concludes that on t14, user 'user_l' posted about the Health Insurance with an average negative sentiment value of -2.67. This means that within an hour, user 'user_l' tends to have a negative opinion of the services provided by The Health Insurance.

The second query in Table 10 obtained 11 data from user 'user_s' and a time range between t0 and t12. Fig. 9 depicts the resulting graph from this query's visualization. The user 'user_s' tweets with a sentiment value at t0 and t8, so the sentiment value at other times is 0. The graph in Fig. 9 reveals that at t0 and t8, the average sentiment value for the 'user_s' make status regarding Health Insurance is -0.906. At time t0, user 'user_s' updates his status twice with values of -0.906. At time t8, the user 'user_s' updates his status with the value of -0.906. This indicates that within the given time range, user 'user_s' tend to have a negative opinion of the Health Insurance.

D. User Impression Context

When the User Impression Context system provides a response that corresponds to the user's query and visualization graph, it is deemed successful. This feature's data consists of 177 users and 1064 tweets. This feature extraction process yields a matrix containing user and impression words used in tweets. The system processes the collection of queries, then compares the results with the correct answers derived from the observations. Examples of queries for this feature are in Table 11.

TABLE XI User query feature 3

No	Impression		
1.	buruk, parah		
2.	aman, cepat, nyaman		

The first two query words are "buruk" and "parah." The second query is "aman," "cepat," and "nyaman."



From query number one in Table 11, the system searched for the words impression, "buruk," and "parah." The visualization graph for this query is shown in Fig. 11. From this graph, user 'user_l' uses both words, "buruk ," and "parah." The overall sentiment value from the graph is negative. This indicates that the words "buruk" and "parah" are used to describe badly about Health Insurance service. Query number two in Table 11 searches for words impression "aman," "cepat," and "nyaman." Fig. 12 shows a graph for the second query. There is nobody who uses these three words together to describe the Health Insurance service, and most users use words like "aman," "cepat," and "nyaman" to describe positively the Health Insurance service.

E. Temporal User Character Context

Similar to User Impression Context, this function utilizes 177 users and 1064 tweets. The result of this feature extraction process is a matrix containing user and Big Five Personality that is calculated using Character Analysis. A user will have five personalities based on the Big Five Personalities, with different values for each personality at each time. The first query is user 'user_p'. The second query is user 'user s'.



2-29 2020-01-05 202 date Fig. 14 Second Query's Graph of User Impression Context

2020-01-06

User 'user_p' has tweeted on seven different days. A visualization graph of user 'user_p' is shown in Fig. 13. From Fig. 12, it can be seen that user 'user_p' has a greater value of neuroticism than the other personalities, and neuroticism values also have the highest value among other days on the graph. User 'user_s' in Fig. 14 also has the highest neuroticism value among other personalities in the graph of the result of the second query.

IV. CONCLUSIONS

This study describes a method for constructing sentiment and context analyses. This system consists of a Twitter data crawler, data preprocessing, text mining, Affective Modelbased sentiment analysis, and four-feature context analysis. Context analysis is comprised of the following components: (1) Temporal Keyword Sentiment Context, (2) Temporal User Sentiment Context, (3) User Impression Context, and (4) Temporal User Character Context.

The Affective Model Method for Sentiment Analysis has not been able to produce sufficient accuracy, but it has a higher precision value than the SVM method. One of the benefits of using the rule-based method is that it can evaluate an object based on differences in the adjectives used in an expression, and users can modify it by adding additional features during the process of analyzing an opinion, one example being the use of Context Analysis. The challenge of sentiment analysis on social media is posed by the large number of users who do not speak the Indonesian standard

1.5

2019-12-29

language and the use of conjunctions that influence the sentiment value of a sentence.

Each Context Analysis feature is evaluated with ten queries, each of which contains data in the database and therefore does not return a null result. The system has succeeded in displaying a graphical representation of each query based on the features that are being run and on the query results that have been culled. Therefore, the Context Analysis system can aid in providing a more comprehensive description of sentiment analysis in general. To make it easier to use, a webbased user interface is currently being developed to display the feature's visualization graph.

References

- F. Wu, Y. Huang, and Y. Song, "Structured microblog sentiment classification via social context regularization," *Neurocomputing*, vol. 175, pp. 599–609, Jan. 2016, doi: 10.1016/j.neucom.2015.10.101.
- S. Kemp, "Digital 2020: Indonesia," Indonesia, Feb. 18, 2020.
 [Online]. Available: https://datareportal.com/reports/digital-2020indonesia, Accessed on: Dec 01, 2021.
- [3] K. Catur Bagus Wicaksono, "Mengukur Efektivitas Social Media Bagi Perusahaan," *Binus Business Review*, vol. 4, no. 1, pp. 551–564, May 2013, doi: 10.21512/bbr.v4i1.1419.
- [4] M. I. Mahdi, Jakarta, Indonesia. [Online]. Available: https://dataindonesia.id, Accessed on: May 11 2022.
- J. A. Russell, "A circumplex model of affect," *J Pers Soc Psychol*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.
- [6] A. Kumar and G. Garg, "Systematic literature review on context-based sentiment analysis in social multimedia," *Multimed Tools Appl*, vol. 79, no. 21–22, pp. 15349–15380, Jun. 2020, doi: 10.1007/s11042-019-7346-5.
- [7] D. Naskar, S. R. Singh, D. Kumar, S. Nandi, and E. O. de la Rivaherrera, "Emotion Dynamics of Public Opinions on Twitter," *ACM Trans Inf Syst*, vol. 38, no. 2, pp. 1–24, Apr. 2020, doi: 10.1145/3379340.
- [8] A. T. Kabakus and R. Kara, "TwitterSentiDetector: a domainindependent Twitter sentiment analyser," *INFOR: Information Systems and Operational Research*, vol. 56, no. 2, pp. 137–162, Apr. 2018, doi: 10.1080/03155986.2017.1340797.
- [9] W. He, Z. Zhang, and V. Akula, "Comparing consumer-produced product reviews across multiple websites with sentiment classification," *Journal of Organizational Computing and Electronic Commerce*, vol. 28, no. 2, pp. 142–156, Apr. 2018, doi: 10.1080/10919392.2018.1444350.
- [10] M. T. F. Al Islami, A. R. Barakbah and A. Helen, "Ekstrak isu maskapai penerbangan pada media sosial dengan computational linguistics," Politeknik Elektronika Negeri Suarabaya, Surabaya, 2019.
- [11] M. T. F. Al Islami, A. R. Barakbah, and T. Harsono, "Social Media Engineering for Issues Feature Extraction using Categorization Knowledge Modelling and Rule-based Sentiment Analysis," *JOIV*: *International Journal on Informatics Visualization*, vol. 5, no. 1, Mar. 2021, doi: 10.30630/joiv.5.1.397.
- [12] B. J. M. Putra, A. Helen, and A. R. Barakbah, "Rule-based Sentiment Degree Measurement of Opinion Mining of Community Participatory in the Government of Surabaya," *EMITTER International Journal of Engineering Technology*, vol. 6, no. 2, pp. 200–216, Dec. 2018, doi: 10.24003/emitter.v6i2.275.
- [13] E. Çano and M. Morisio, "MoodyLyrics," in Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, New York, NY, USA: ACM, Mar. 2017, pp. 118– 124. doi: 10.1145/3059336.3059340.
- [14] D. Naskar, S. Mokaddem, M. Rebollo, and E. Onaindia, "Sentiment Analysis in Social Networks through Topic Modeling".

- [15] K. Hulliyah, H. T. Sukmana, N. S. A. Bakar, and A. R. Ismail, "Indonesian Affective Word Resources Construction in Valence and Arousal Dimension for Sentiment Analysis," in 2018 6th International Conference on Cyber and IT Service Management (CITSM), IEEE, Aug. 2018, pp. 1–5. doi: 10.1109/CITSM.2018.8674370.
- [16] A. Kumar and G. Garg, "The Multifaceted Concept of Context in Sentiment Analysis," 2020, pp. 413–421. doi: 10.1007/978-981-15-1451-7_44.
- [17] R. Liu and L. Jin, "Sentiment analysis method based on the focus sentence and context for Chinese comment text," in *Proceedings of* 2nd International Conference on Information Technology and Electronic Commerce, IEEE, Dec. 2014, pp. 267–270. doi: 10.1109/ICITEC.2014.7105616.
- [18] S. Abudalfa and M. Ahmed, "Open Domain Context-Based Targeted Sentiment Analysis System," in 2019 IEEE 7th Palestinian International Conference on Electrical and Computer Engineering (PICECE), IEEE, Mar. 2019, pp. 1–5. doi: 10.1109/PICECE.2019.8747187.
- [19] A. Vanzo, D. Croce, and R. Basili, "A context-based model for Sentiment Analysis in Twitter." [Online]. Available: http://expandedramblings.com/.
- [20] A. A. Jalal, "Text Mining: Design of Interactive Search Engine Based Regular Expressions of Online Automobile Advertisements," *International Journal of Engineering Pedagogy (iJEP)*, vol. 10, no. 3, p. 35, May 2020, doi: 10.3991/ijep.v10i3.12419.
- [21] B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.
- [22] I. Rasyada, Y. Setiowati, A. Barakbah, and M. T. Fiddin Al Islami, "Sentiment Analysis of BPJS Kesehatan's Services Based on Affective Models," in 2020 International Electronics Symposium (IES), IEEE, Sep. 2020, pp. 549–556. doi: 10.1109/IES50839.2020.9231940.
- [23] L. Feldman Barrett and J. A. Russell, "Independence and bipolarity in the structure of current affect.," *J Pers Soc Psychol*, vol. 74, no. 4, pp. 967–984, Apr. 1998, doi: 10.1037/0022-3514.74.4.967.
- [24] S. Wojowasito aand T. Wasito, Kamus Lengkap: Inggeris Indonesia; Indonesia - Inggeris dengan Ejaan Yang Disempurnakan. Jakarta: Hasta, 2007.
- [25] A. Hume, L. Cernuzzi, and J. L. Zarza, "Analysis of the Big-Five personality traits in the Chatbot 'UC-Paraguay," 2022. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx.
- [26] V. Balakrishnan, S. Khan, T. Fernandez, and H. R. Arabnia, "Cyberbullying detection on twitter using Big Five and Dark Triad features," *Pers Individ Dif*, vol. 141, pp. 252–257, Apr. 2019, doi: 10.1016/j.paid.2019.01.024.
- [27] K. Kircaburun, S. Alhabash, S. B. Tosuntaş, and M. D. Griffiths, "Uses and Gratifications of Problematic Social Media Use Among University Students: a Simultaneous Examination of the Big Five of Personality Traits, Social Media Platforms, and Social Media Use Motives," *Int J Ment Health Addict*, vol. 18, no. 3, pp. 525–547, Jun. 2020, doi: 10.1007/s11469-018-9940-6.
- [28] S. M. Mohammad and P. D. Turney, "Crowdsourcing A Word-Emotion Association Lexicon," *Comput Intell*, vol. 29, no. 3, pp. 436– 465, Aug. 2013, doi: 10.1111/j.1467-8640.2012.00460.x.
- [29] W. Maharani and V. Effendy, "Big five personality prediction based in Indonesian tweets using machine learning methods," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, pp. 1973–1981, Apr. 2022, doi: 10.11591/ijece.v12i2.pp1973-1981.
- [30] K. Faiqoh, A R. Brakbah and A. Helen, "Klasifikasi kepribadian user menggunakan multi label classification," Politeknik elektronika negeri Suarabaya, Surabaya, 2017.
- [31] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.