OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 6, Issue. 1, April 2021. pp. 1-12
doi:10.34818/indojc.2021.6.1.408

Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation

Raginda Firdaus #1, Ibnu Asror #2, Anisa Herdiani #3

 # Informatics Faculty, Telkom University
Jl. Telekomunikasi Jl. Terusan Buah Batu No.01, Sukapura, Dayeuhkolot, Bandung, Jawa Barat, Indonesia

[1]firdausraginda@student.telkomuniversity.ac.id
[2]iasror@telkomuniversity.ac.id
[3] anisaherdiani@telkomuniversity.ac.id

Abstract

Student feedback for lecturer plays an important role, it used to improve the quality of lecturer in teaching. In general, student feedback consists of two types, quantitative feedback and qualitative feedback. For quantitative feedback, it can easily analyze using statistical calculations, because it contains closed questions with multiple choices. But for qualitative feedback, it is difficult to analyze, because it contains open questions with essay answers. Lecturers can analyze manually, but it takes extensive times and the results can be very subjective. To overcome this problem, sentiment analysis is applied to analyze textual data automatically in order to improve teaching evaluations. This research uses student feedback as dataset, and lexicon approach with InSet Lexicon. InSet Lexicon has been best tested for Indonesian language sentiment analysis, however, there is not much paper about Indonesian language sentiment analysis with lexicon approach that uses InSet Lexicon. Therefore, this research can give insight on how good the impact of InSet Lexicon towards Indonesian language sentiment analysis. The evaluation in this research performed in 2 levels, document level and sentence level. For document level, the accuracy is 90.9%, as for sentence level obtained better results after lexicon modification.

Keywords: sentiment analysis, student feedback, lexicon, inset lexicon

Abstrak

Dalam dunia pendidikan, data evaluasi dosen oleh mahasiswa (EDOM) memegang peranan penting, yaitu untuk meningkatkan kualitas mengajar dosen. Pada umumnya, data EDOM terdiri atas dua jenis, yaitu data kuantitatif dan kualitatif. Data kuantitatif mudah untuk dianalisis menggunakan perhitungan statistika, karena berisi pertanyaan tertutup dengan jawaban pilihan ganda. Sedangkan data kualitatif sulit dianalisis, karena berisi pertanyaan terbuka dengan jawaban esai. Dosen bisa saja melakukan analisis manual terhadap data kualitatif, namun akan memakan waktu yang lama dan hasilnya bisa sangat subjektif. Untuk mengatasi masalah ini, analisis sentimen diterapkan untuk menganalisis data tekstual secara otomatis agar meningkatkan evaluasi mengajar dosen. Penelitian ini menggunakan data EDOM sebagai dataset, dan pendekatan *lexicon* dengan InSet *Lexicon*. InSet *Lexicon* telah teruji paling baik untuk analisis sentimen bahasa Indonesia, namun, tidak banyak penelitian tentang analisis sentimen bahasa Indonesia dengan pendekatan *lexicon* yang menggunakan InSet *Lexicon*. Dari penelitian ini, bisa diketahui seberapa baik dampak InSet *Lexicon* terhadap analisis sentimen bahasa Indonesia. Evaluasi pada penelitian ini dilakukan pada 2 tingkat, tingkat dokumen dan tingkat kalimat. Untuk tingkat dokumen, didapatkan akurasi sebesar 90.9%, sedangkan untuk tingkat kalimat didapat hasil yang lebih baik setelah dilakukan modifikasi pada *lexicon*.

Kata Kunci: analisis sentimen, evaluasi dosen oleh mahasiswa, *lexicon*, inset *lexicon*

Received on January, 2021. Accepted on Mei, 2021

I. Introduction

SENTIMENT analysis is a set of methods and techniques about detecting and extracting subjective information, such as opinions and attitudes, from language [1]. Sentiment analysis will determine whether someone has a positive, neutral, or negative sentiment towards something [1]. Research on sentiment analysis has many domains, such as film reviews, product reviews, and hotel reviews. One application of sentiment analysis in education domain is sentiment analysis of student feedback. Student feedback consists of two types, quantitative and qualitative data [2]. In general, quantitative data contains closed questions with multiple choice answers, while qualitative data contains open questions with essay answers [2]. We can get a lot of insights from qualitative data, including suggestion from students to lecturers, student perceptions of the character of lecturers, student perceptions of lecturers' understanding of the material taught, and student perceptions of lecturers' presentation skills [3]. It is important to analyze qualitative data to improve the effectiveness of lecturer evaluations [2]. But to interpret qualitative data is difficult to do, unlike quantitative data in numerical rating scale that is easily analyzed by statistical techniques [2]. Lecturers can analyze manually students' feedback by giving a minus sign (-) to negative qualitative data, and plus (+) to positive ones [2]. After that, the lecturer can count the number of positive and negative data to find out which is more, but this is a laborious task, and the results can be very subjective [2]. Therefore, the system that can automatically interpret qualitative data from the students' feedbacks is needed [2]. Here, sentiment analysis is the solution, because it performs an effective automated qualitative data analysis method for enhancing lecturer evaluation [2].

There are two approaches to conduct sentiment analysis, machine learning (ML) approach and Lexicon approach [4], [5]. Although ML has the advantage of being able to identify implied sentiments, it requires training data with quite a large number [4], [6], [7]. For example, in the case of film review sentiment analysis, this case has a considerable amount of data, because there are many movie review sites with comments and ratings on internet, so that it fits using the ML approach [4]. But in this research, we use Telkom University student feedback data, which not spread much on the internet, kept by internal campus, and has not been manually labeled. This condition makes the data does not have a sufficient amount such as movie review data. When there is not enough data used, sentiment analysis with existing lexicon is needed [4]. There is a lot of research about Indonesian language sentiment analysis with ML approach, commonly uses Naïve Bayes or Support Vector Machine [8]–[12]. Although ML approach applied in many research, our research tries to see problem with different perspective, by overcome processing the limited data with the lexicon approach. This research also evaluates sentiment analysis in two levels, sentence level and document level. Furthermore, instead of building own lexicon, we use existing lexicon named InSet Lexicon. InSet Lexicon has been best tested for Indonesian language sentiment analysis compared to translated SentiWordNet, translated Liu Lexicon, translated AFINN Lexicon, and Vania Lexicon [13]. Even though InSet Lexicon has been best tested, there is not much paper about Indonesian language sentiment analysis with lexicon approach that uses InSet Lexicon. Therefore, this research can give insight on how good the impact of InSet Lexicon towards Indonesian language sentiment analysis.

II. Literature Review

Sentiment analysis of student feedback with lexicon has been done before. The lexicon-based approach itself is unsupervised learning with the advantage of fast computation, since it does not require training on its data [7]. Lexicon-based approach is divided into three steps, word-level calculation, sentence-level calculation, and document-level calculation [7]. One of lexicon-based implementation is research conducted by Quratulain Rajput, Sajjad Haider, and Sayeed Ghani [14]. They conducted a sentiment analysis by calculating the sentiment score [14]. But before calculating sentiment score, preprocessing is carried out first, the stages include tokenization, stemming, case conversion, punctuation removal, and stop word removal [14]. After preprocessing done, then they do polarity tagging, calculating the word frequency, determining the word attitude, calculating the overall attitude, and finally calculating the sentiment score [14].

1) *Polarity Tagging:* This stage analyze sentiment words in the sentence, then determines the polarity of the sentiment words [14].

2) *Word Frequency:* This stage calculates the number of occurrences of sentiment words [14].

3) *Word Attitude:* This stage determines the word attitude in each sentiment word, the +1 value if the sentiment words have positive polarity, and -1 if negative [14].

4) *Overall Attitude:* This stage calculates the overall attitude with the following equation [14].

$$\textit{Overall attitude} = \textit{word attitude} * \textit{word frequency} \tag{1}$$

5) *Sentiment Score:* This stage calculates the sentiment score by summing the overall attitude that has been calculated before [14].

$$Sentiment\ Score = \sum_{i=1}^{n} overallAttitude(i) \tag{2}$$

If the sentiment score of a sentence > 0 then the sentence has a positive sentiment, if the sentiment score < 0 then it has negative sentiment. Whereas if the sentiment score = 0, then the sentence has a neutral sentiment. Words extraction involves a lot of challenges, therefore Quratulain Rajput, Sajjad Haider, and Sayeed Ghani performed sentiment dictionary modification [14]. The words *miss*, *lecture*, *fine*, and *fun* are assigned negative, however in context of students' feedback they should not be considered as negative [14]. The word *miss* and *lecture* should be neutral because refers to a teacher and class session [14]. While the word *fine* and *fun* commonly used in a positive sense so they should be positive instead of negative [14]. This research obtained an accuracy of 91.2% and the best performance was for positive sentiments because it has the highest recall and precision rates [14]. Commonly, sentiment analysis performed in sentence level, but it can also be done in document level [15]. For document level sentiment analysis, what is done is counting all opinions in the document [15]. As for the sentence level, identify whether each sentence expresses an opinion or not, if so, then determine the polarity of the opinion [15].

Another example of student feedback sentiment analysis with lexicon is research conducted by Khin Zezawar Aung and Nyein Nyein Myo [5]. In that research, Khin Zezawar and Nyein Nyein developed 745 words in sentiment word databases that were labeled with range -3 to +3 by education experts [5]. To make lexicon compatible with education domain, they analyze dataset and find words that are close to education such as 'explain', 'interactive', 'ask', 'understand', 'practiced', 'knowledgeable', 'available', 'slow', 'fast', 'talented', 'complex', 'complicated', 'sleepy', 'overcome', 'concern', then add these words to their lexicon [5]. Sentiment analysis also can be done using existing lexicon, for example using InSet Lexicon that was made by Fajri Koto and Gemala Y. Rahmaningtyas [13].

InSet Lexicon is an Indonesian lexicon containing 3,609 positive words and 6,609 negative words with value in range -5 to +5 [13]. This lexicon does not only contain the standard words but also non-standard words, because InSet Lexicon is built using Twitter data [13]. Twitter was chosen because it represents the most frequently used social media in Indonesia [13]. Of approximately 10,000 tweets obtained, preprocessing is then performed. There are 4 stages in preprocessing, removing repetitive ads, case conversion, removing Twitter URLs and entities like @account, and stop word removal [13]. In selecting word candidates, InSet Lexicon uses n-gram with n = {1, 2, 3} [13]. After that, the value is given manually with range of -5 (very negative) to +5 (very positive) by 2 native Indonesian speakers [13]. Research shows InSet Lexicon provides better results for Indonesian language sentiment analysis than the translated SentiWordNet, translated Liu Lex, translated AFINN, and Vania Lexicon [13].

III. Research Method

The workflow of this research can be seen in Figure 1.

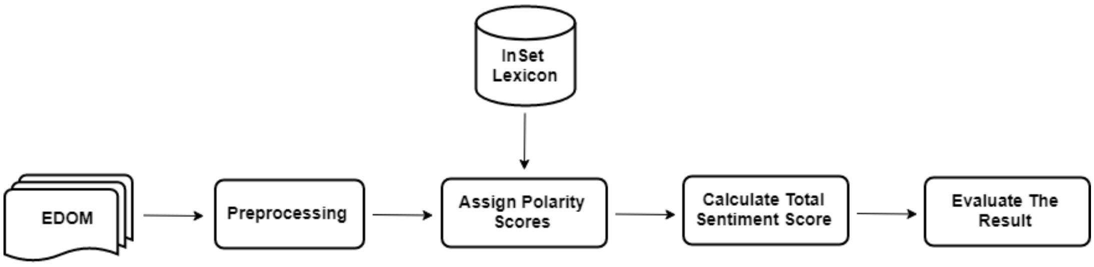

Fig. 1. Research Flowchart

A. Dataset

This research uses Telkom University student feedback data for Artificial Intelligence (AI) lecture in 2018/2019. There were 11 lecturers that taught AI in that period. Each lecturer teaches a different number of classes and students. Total student feedback that has been collected is 442 data. This student feedback has 3 labels: positive, negative, and neutral. Each label has a different amount of data, it can be seen in Table I.

TABLE I
NUMBER OF POSITIVE, NEGATIVE, AND NEUTRAL DATA

	Positive	Negative	Neutral	Total Data per Lecturer
Lecturer 1	80	13	8	101
Lecturer 2	57	3	3	63
Lecturer 3	26	4	0	30
Lecturer 4	55	17	10	82
Lecturer 5	22	3	0	25
Lecturer 6	24	2	1	27
Lecturer 7	29	3	0	32
Lecturer 8	7	0	2	9
Lecturer 9	13	5	0	18
Lecturer 10	16	7	3	26
Lecturer 11	15	13	1	29
Total Data Per Label	344	70	28	442

There are 2 levels of student feedback in this research:

1) *Sentence Level:* Each comment per lecturer. There are 442 sentence level student feedbacks.
2) *Document Level:* All comments per lecturer. Since there are 11 lecturers, so this research has 11 document level student feedbacks.

Both student feedback levels are labeled positive / negative / neutral manually by 3 native Indonesian speakers. For sentence level, native Indonesian speakers label each comment per a lecturer. As for document level, native Indonesian speakers first read all comments per lecturer, then summarize what the label is.

B. Preprocessing

Before calculating sentiment score, preprocessing is required first. Preprocessing includes tokenization, case conversion, stemming, punctuation removal, and stop word removal.

1) *Tokenization:* This process breaks sentence into a list of words [14], [16].
2) *Case Conversion:* Convert to lowercase letters [14], [16].
3) *Stemming:* Every word is changed into their root word [14]. In this research, stemming uses *Sastrawi*[1].
4) *Punctuation Removal:* Eliminates punctuation [14].
5) *Stop Word Removal:* Eliminates words that do not contribute to sentiment analysis, such as prepositions, help verbs, and so forth [14], [16].

Examples of student feedback can be seen in Table II.

TABLE II
EXAMPLE OF STUDENT FEEDBACK

Sentence
Ngajarnya enak, tegas, mudah di mengerti (The teaching is good, firm, easy to understand)

Result of preprocessing can be seen in Table III:

TABLE III
RESULT OF PREPROCESSING

Preprocessing
'ngajarnya', 'enak', 'tegas', 'mudah', 'erti' (‘teaching’, ‘good’, ‘firm’, ‘understand’)

C. Assign Polarity Scores

This stage analyze sentiment words in student feedback, then determines the polarity scores of the sentiment words based on the lexicon [14]. The results of assigning the polarity scores can be seen in Table IV.

[1] https://github.com/har07/PySastrawi

TABLE IV
RESULT OF ASSIGN POLARITY SCORES

Assign Polarity Scores				
Sentiment words	=	‘enak’ (good)	‘mudah’ (easy)	‘erti’ (understand)
Polarity score	=	5 + (-4)	4 + (-1)	2 + (-2)

From Table IV it is known that the words ‘enak’ (good), ‘mudah’ (easy), and ‘erti’ (understand) are sentiment words contained in student feedback. Each of these three words belongs to positive and negative classes at the same time, so to get the total polarity scores, each value of the positive and negative classes must be added together. In its implementation, unique conditions were found, including:

1) *Word with hyphen ‘-‘:* Example = ‘bertanya-tanya’ (wondering).

2) *Word with n-gram, n = 2:* Example = ‘kecil hati’ (discouraged).

3) *Word with n-gram, n = 3:* Example = ‘buang air besar’ (defecate).

These words require different treatment from other words, because these words cannot be separated and then conduct stemming. For example, the word ‘kecil hati’ (discouraged), without separating words and without stemming has a value of -4 in lexicon. But if it is separated and stemmed, it has a different value, kecil (small) = -3, hati (heart) = -2 + 2, becomes -3.

The word ‘kecil hati’ (discouraged) which should have a value of -4, when separated and stemmed will have a value of -3. Therefore, the 3 unique conditions above must be treated differently by assigning polarity scores without separating words and without stemming.

D. Calculate Total Sentiment Score

Sentiment score is a value that will determine the sentiment of student feedback. In this research, sentiment score is a summation of overall polarity scores in student feedback.

$$Sentiment\ Score = \sum_{i=1}^{n} overallPolarityScores(i) \tag{3}$$

If the sentiment score < 0 then the student feedback has a negative sentiment. And if the sentiment score > 0, then the student feedback has positive sentiment. Student feedback can also have neutral sentiment if the sentiment score = 0. The result of sentiment score can be seen in Table V.

TABLE V
RESULT OF SENTIMENT SCORE

Sentiment Score								
Sentiment words	=	‘enak’ (good)		‘mudah’(easy)		‘erti’ (understand)		
Sentiment score	=	1	+	3	+	0	=	4

From the above calculation, it can be seen that the level sentence student feedback has a positive sentiment with a sentiment score = 4. To get the document level sentiment score, it is necessary to add all sentence level sentiment scores in the document [15]. For example, in Table VI there are 5 sentence level student feedbacks.

TABLE VI
SENTENCE LEVEL STUDENT FEEDBACK IN ONE DOCUMENT

No	Lecturer	Student Feedback	Sentence Level Sentiment Score
1	Lecturer 1	Terkadang penjelasan beberapa materinya masih kurang jelas (Sometimes the explanation of some materials is still unclear)	-4
2	Lecturer 1	terimkasih sudah mau mengajar dengan sabar pak (Thank you for teaching patiently)	+7
3	Lecturer 1	Baik (Well)	+2
4	Lecturer 1	Mantap (Great)	+5
5	Lecturer 1	sudah baik, penyampaian materi mudah dipahami mahasiswa (Decent, the delivery of material easy for student to understand)	+5
Document Level Sentiment Score			**+15**

From Table VI, it is known that the sentiment score of sentence levels student feedbacks are -4, +7, +2, +5, and +5. While the document level student feedback is +15.

E. Lexicon Modification

After reviewing the InSet Lexicon and the student feedback, it was found that InSet Lexicon must be modified to overcome certain words. These conditions are:

1) When there is a word 'tidak ada' (nothing) and 'ga ada' (nothing) in student feedback: There are a lot of student feedback that contains the words 'tidak ada' (nothing) and 'ga ada' (nothing), which means that students do not have any comments for the lecturer. In this condition, the student feedback should have neutral sentiment, but InSet Lexicon does not contain word 'tidak ada' and 'ga ada', then what happens can be seen in Table VII and VIII.

TABLE VII
WORD 'TIDAK ADA'

'tidak ada'				
'tidak' (no)		'ada' (exist)		
-5	+	4+(-3)	=	-4

TABLE VIII
WORD 'GA ADA'

'ga ada'				
'ga' (no)		'ada' (exist)		
0	+	4+(-3)	=	1

Therefore, it is necessary to add the word ‘tidak ada’ (nothing) and ‘ga ada’ (nothing) to lexicon with value 0 (neutral).

2) When there is a word ‘masukkan’ (suggestion) and ‘masukan’ (suggestion) in student feedback: Word ‘masukkan’ (suggestion) and ‘masukan’ (suggestion) in student feedback means suggestions from students for lecturers. For example, student feedback in table IX.

TABLE IX
SENTENCE CONTAINS ‘MASUKAN’

Sentence
“Terimakasih atas ilmu yang sudah Ibu berikan. **Masukan** saya, sebaiknya dijelaskan lagi lebih detail tentang materi untuk tugas-tugasnya. Agar mahasiswa dapat lebih paham dalam pengerjaan tugas-tugasnya” (Thank you for the knowledge that you have given. My suggestion is, you should explain more detail about assignments you gave. So students can understand more in the execution of their tasks)

InSet Lexicon does not have the word ‘masukkan’ (suggestion) and ‘masukan’ (suggestion), so when it stemmed, the word becomes ‘masuk’ (enter) with value -3. Not only the value changed, but also the meaning of the word. Therefore, it is necessary to add the word ‘masukkan’ and ‘masukan’ to lexicon with value -1. The value -1 is obtained from the synonym word ‘masukkan’, which is ‘saran’ with value 2 + (-3) = -1 based on InSet Lexicon.

Total there are 4 words added to lexicon. For those four words, after the case conversion process on preprocessing, assign polarity scores then performed and not through the stemming process so it doesn't change its meaning.

IV. RESEARCH RESULT

A. Test Scenario

This research performed 2 scenarios for testing.

1) *Scenario 1:* Sentiment analysis with InSet Lexicon.

2) *Scenario 2:* Sentiment analysis with a modified InSet Lexicon.

Both of scenarios are made to compare which scenario is better. The comparison applied in 2 levels, document level and sentence level.

1) *Document Level:* Compare scenario 1 & scenario 2.

2) *Sentence Level:* Compare scenario 1 & scenario 2.

By that, the best scenario for document level and sentence level can be obtained. Since the number of student feedback with positive, negative, and neutral label are not balanced, so the calculations of accuracy, precision, recall, and f-measure are performed. The calculations can focus on one particular label, so information about what label is best recognized by the system can also be obtained [17].

B. Test Result

Based on test scenario, the aim of this testing is to know which scenario has better performance in document level and sentence level. The test results are shown as follows:

1) *Document Level:* The results of the comparison between scenario 1 and 2 on documents level can be seen in Table X and XI.

TABLE X
DOCUMENT LEVEL TEST SCENARIO 1

Accuracy	0.909	
Precision	**Positive**	0.875
	Negative	1
	Neutral	None
Recall	**Positive**	1
	Negative	0.75
	Neutral	None
F-Measure	**Positive**	0.933
	Negative	0.857
	Neutral	None
Sentiment Score	295	
Sentiment	Positive	

TABLE XI
DOCUMENT LEVEL TEST SCENARIO 2

Accuracy	0.909	
Precision	**Positive**	0.875
	Negative	1
	Neutral	None
Recall	**Positive**	1
	Negative	0.75
	Neutral	None
F-Measure	**Positive**	0.933
	Negative	0.857
	Neutral	None
Sentiment Score	417	
Sentiment	Positive	

Based on Table X and XI, it can be seen that scenario 1 and 2 have the same good performance for document level with both accuracy 0.909 or 90.9%.

2) *Sentence Level:* The results of the comparison between scenario 1 and 2 on sentence level can be seen in Table XII and XIII.

TABLE XII
SENTENCE LEVEL TEST SCENARIO 1

		Lecturer 1	Lecturer 2	Lecturer 3	Lecturer 4	Lecturer 5	Lecturer 6	Lecturer 7	Lecturer 8	Lecturer 9	Lecturer 10	Lecturer 11
Accuracy		0.644	0.651	0.667	0.585	0.560	0.741	0.656	0.556	0.778	0.731	0.586
Precision	**positive**	0.881	0.952	0.947	0.814	0.867	1.000	0.952	0.833	0.909	0.923	0.750
	negative	0.188	0.056	0.250	0.353	0.111	0.250	0.143	0.000	0.571	0.667	0.571
	neutral	0.000	0.000	0.000	0.200	0.000	0.000	0.000	none	none	0.250	0.000
Recall	**positive**	0.738	0.702	0.692	0.636	0.591	0.750	0.690	0.714	0.769	0.750	0.600
	negative	0.462	0.333	0.500	0.706	0.333	1.000	0.333	none	0.800	0.857	0.615
	neutral	0.000	0.000	none	0.100	none	0.000	none	0.000	none	0.333	0.000
F-Measure	**positive**	0.803	0.808	0.800	0.714	0.703	0.857	0.800	0.769	0.833	0.828	0.667
	negative	0.267	0.095	0.333	0.471	0.167	0.400	0.200	none	0.667	0.750	0.593
	neutral	none	none	none	0.133	none	none	none	none	none	0.286	none
Sentiment Score		102	96	53	17	38	15	43	7	-12	-49	-15
Sentiment		positive	positive	positive	positive	positive	positive	positive	positive	negative	negative	negative

TABLE XIII
SENTENCE LEVEL TEST SCENARIO 2

		Lecturer 1	Lecturer 2	Lecturer 3	Lecturer 4	Lecturer 5	Lecturer 6	Lecturer 7	Lecturer 8	Lecturer 9	Lecturer 10	Lecturer 11
Accuracy		0.663	0.714	0.667	0.659	0.600	0.815	0.688	0.556	0.833	0.808	0.621
Precision	**positive**	0.882	0.953	0.947	0.818	0.875	1.000	0.955	0.833	0.917	0.923	0.750
	negative	0.194	0.071	0.250	0.429	0.125	0.333	0.167	0.000	0.667	0.857	0.615
	neutral	0.500	0.500	0.000	0.600	0.000	0.500	0.000	none	none	0.500	0.250
Recall	**positive**	0.750	0.719	0.692	0.655	0.636	0.792	0.724	0.714	0.846	0.750	0.600
	negative	0.462	0.333	0.500	0.706	0.333	1.000	0.333	none	0.800	0.857	0.615
	neutral	0.125	1.000	none	0.600	none	1.000	none	0.000	none	1.000	1.000
F-Measure	**positive**	0.811	0.820	0.800	0.727	0.737	0.884	0.824	0.769	0.880	0.828	0.667
	negative	0.273	0.118	0.333	0.533	0.182	0.500	0.222	none	0.727	0.857	0.615
	neutral	0.200	0.667	none	0.600	none	0.667	none	none	none	0.667	0.400
Sentiment Score		112	118	53	51	43	32	48	12	-5	-39	-8
Sentiment		positive	positive	positive	positive	positive	positive	positive	positive	negative	negative	negative

Based on Table XII, and XIII, scenario 2 shows better performance than scenario 1. It can be seen from the accuracy and f-measure of scenario 2 > scenario 1. From Table XII and XIII can also be seen that system can better recognize positive label than the negative and neutral label, this is due to the f-measure in both scenarios for positive label almost always >= negative and neutral label.

In Table X, XI, XII, and XIII, there is a value of None. The None value is caused by division by zero in the precision / recall / f-measure calculation, so there is no insight that can be obtained from the results.

C. *Test Result Analysis*

The test result shows that for document level, scenario 2 and scenario 1 has the same accuracy, precision, recall, and f-measure. That means scenario 1 is as good as scenario 2 in document level. This is caused by lexicon modification that is not significantly impactful in document level, since there are only 4 words that modified. While in sentence level, scenario 2 is better than scenario 1. In sentence level, we can detect even to the small impact of lexicon modification. Suppose there is only 1 word that modified in this research, the impact is likely found in sentence level. However, lexicon modification needs to be made whenever applied to a specific case such as education, because InSet Lexicon has a general domain since it was built uses Twitter data [13].

The test result also shows that the system can better recognize student feedback with positive label than negative and neutral label. This is because student feedback with positive label has short sentences, therefore making it easier for the system to recognize it. Whereas negative label tends to have longer and more complex sentences. For neutral label, the system is more difficult to recognize, because the sentiment score must be exactly 0. For example, the length of student feedback with positive, negative, and neutral label can be seen in Table XIV.

TABLE XIV
STUDENT FEEDBACK WITH POSITIVE, NEGATIVE, AND NEUTRAL LABEL

No	Student Feedback	Label
1	Mantap (Great)	Positive
2	terimakasih pak (thank you sir)	Positive
3	baik (good)	Positive
4	semuanya sudah baik (everthing is good)	Positive
5	sangat puas (very satisfied)	Positive
6	Sudah baik (already well)	Positive
7	Sebaiknya dalam menjelaskan materi pelan dan sabar. jika mahasiswa tidak juga mengerti setelah dijelaskan, sebaiknya mengganti cara menjelaskannya dengan yang lebih mudah dipahami (It is better to explain material slowly and patiently. if students do not understand after being explained, it is better to change the way to explain it with a more easily understood one)	Negative
8	materi yang disampaikan susah untuk dipahami, kedepannya ditingkatkan lagi agar mahasiswanya mengerti tentang materi yang disampaikan (the material presented was difficult to understand, in the future, it should be improved so students can understand better)	Negative
9	setelah membahas sedikit lain langsung buat latihan untuk mahasiswa (after discussing a little more, directly conduct exercise for students)	Neutral

V. CONCLUSION

From the research and testing, it can be concluded that sentiment analysis of student feedback evaluation with InSet Lexicon was able to provide 90.9% accuracy in document level, while in the sentence level obtained better results after lexicon modification. It was also concluded that the system could recognize better student feedback with positive label than negative and neutral label. For the next research, it can be considered to make many modifications to the lexicon to fit with the research domain.

ACKNOWLEDGMENT

We would like to thank to all colleagues at Telkom University who have supported so that this research can be finished.

REFERENCES

[1] M. V. Mäntylä, D. Graziotin, and M. Kuutila, “The evolution of sentiment analysis—A review of research topics, venues, and top cited papers,” *Elsevier*, 2017.

[2] P. Kaewyong, A. Sukprasert, N. Salim, and F. A. Phang, “THE POSSIBILITY OF STUDENTS’ COMMENTS AUTOMATIC INTERPRET USING LEXICON BASED SENTIMENT ANALYSIS TO TEACHER EVALUATION,” *Proceeding 3rd Int. Conf. Artif. Intell. Comput. Sci.*, 2015.

[3] B. Jagtap and V. Dhotre, “SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment,” *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 3, no. 3, 2014.

[4] C. S. G. Khoo and S. B. Johnkhan, “Lexicon-based sentiment analysis : Comparative evaluation of six sentiment lexicons,” *J. Inf. Sci.*, 2017.

[5] K. Z. Aung and N. N. Myo, “Sentiment analysis of students’ comment using lexicon based approach,” *Proc. - 16th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2017*, 2017.

[6] M. E. Basiri and A. Kabiri, “Words Are Important: Improving Sentiment Analysis in the Persian Language by Lexicon Refining,” *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 17, no. 4, 2018.

[7] G. Isabelle, W. Maharani, and I. Asror, “Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes,” vol. 2, no. IcoIESE 2018, pp. 214–219, 2019.

[8] A. R. Naradhipa and A. Purwarianti, “Sentiment Classification for Indonesian Message in Social Media,” *Int. Conf. Cloud Comput. Soc. Netw.*, vol. 3, no. 2, pp. 29–33, 2012.

[9] N. A. Vidya, M. I. Fanany, and I. Budi, “Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers,” *Procedia Comput. Sci.*, vol. 72, pp. 519–526, 2015.

[10] Y. E. Soelistio and M. R. S. Surendra, “SIMPLE TEXT MINING FOR SENTIMENT ANALYSIS OF POLITICAL FIGURE USING NAÏVE BAYES CLASSIFIER METHOD,” *Proc. 7th ICTS*, pp. 99–104, 2013.

[11] C. Fiarni, H. Maharani, and R. Pratama, “Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique,” *2016 4th Int. Conf. Inf. Commun. Technol. ICoICT 2016*, vol. 4, no. c, 2016.

[12] E. Lunando and A. Purwarianti, “Indonesian social media sentiment analysis with sarcasm detection,” *2013 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2013*, pp. 195–198, 2013.

[13] F. Koto and G. Y. Rahmaningtyas, “Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs,” *2017 Int. Conf. Asian Lang. Process.*, 2017.

[14] Q. Rajput, S. Haider, and S. Ghani, “Lexicon−Based Sentiment Analysis of Teachers’ Evaluation,” *Hindawi*, 2016.

[15] S. Behdenna, F. Barigou, and G. Belalem, “Document Level Sentiment Analysis: A survey,” *EAI Endorsed Trans. Context. Syst. Appl.*, vol. 4, no. 13, 2018.

[16] I. Mentaruk, A. Herdiani, and D. Puspandari, “Analisis Sentimen Twitter Transportasi Online Berbasis Ontologi (Studi Kasus : Go-Jek),” *e-Proceeding Eng.*, vol. 6, no. 1, 2019.

[17] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation,” *Am. Assoc. Artif. Intell.*, 2006.