

A Framework of Educational Feedback System with Statistical Method and Sentiment Analysis

Sofhian Fazrin Nasrulloh
Universitas Gadjah Mada
Yogyakarta, Indonesia
sofhian.fn@mail.ugm.ac.id

Adhistya Erna Permanasari
Universitas Gadjah Mada
Yogyakarta, Indonesia
adhisty@ugm.ac.id

Sri Suning Kusumawardani
Universitas Gadjah Mada
Yogyakarta, Indonesia
suning@ugm.ac.id

Abstract—Sentiment analysis has long and widely used in attempt to know public sentiment towards an entity, including in education, specifically in evaluation of learning. The key to success of sentiment analysis is use of right method and data are valid. The Essay for analytical data is sometimes made carelessly but method used to eliminate this data only using stop word removal and no tendency to recognize pattern of user behavior. In addition, sentiment analysis results also require a comparison with other evaluation system. This study makes framework educational feedback system that can generate qualitative data from essay and quantitative data from performance assessment questionnaires. Data obtained from lecturers' evaluation by students in Educational Higher School of Muhammadiyah Kuningan. Qualitative data processed with sentiment analysis using Support Vector Machine (SVM) algorithm, but first performed similarity removal in preprocessing stage to remove essay that made carelessly. Then quantitative data processed with statistical method. Each output is a score that can be correlated to measure relationship between them. As a result, sentiment analysis with SVM able to produce 91% sentiment accuracy and correlation between performance score with sentiment score is 0.73 which means have a high relationship.

Keywords—*educational feedback system, sentiment analysis, support vector machine, similarity removal*

I. INTRODUCTION

Sentiment analysis used to know public sentiment towards an entity, such as market sentiment towards a product, public sentiment towards a public figure, audience sentiment towards a film [1]. The results of analysis commonly in positive sentiment orientation, neutral or negative that can be converted into numbers or quantitative scale. Accuracy of analysis in addition depending on method, also depend on how its valid data available. Data were analyzed are unstructured data such as essay text, comments, opinions, criticisms and recommendations. The essay was obtained through an application, such as social media, comment system or review system.

Especially in evaluation system, a valid essay obtained when users create essay honestly and objectively. Problems arise when users in give essay, has tendency to make essay carelessly and copy paste essay for some objects. As in review application, when review given was same as other review. This could be due to users tendency who carelessly in give review or users saturation when an application forcibly ask a lot of reviews. In addition, to produce sentiment analysis that has high accuracy, essay system can

be combined with other evaluation system, such as system rating, emoticons or questionnaires.

Same problems exist in educational feedback system. A system in education to evaluate performance of lecturers with students asked to fill out some questionnaires accompanied with essay of commentary, criticism or suggestion. Of course, student will evaluate several course lecturers. As mentioned above, problems arise when students fill out essay carelessly, indicated by tendency of same essay towards all lecturer, so data that given invalid and can affect sentiment analysis accuracy. Educational feedback system which combines Likert scale questionnaires and essay can be tried to generate a holistic evaluation.

The rest of this paper is organized as follow: Section II describes educational feedback system. Section III describes the notion of sentiment analysis. Section IV presents related works. Section V describes proposed research method. Section VI presents the obtained results and following by discussion. Finally, Section VII concludes this work.

II. EDUCATIONAL FEEDBACK SYSTEM

Evaluation should be applied after applying concept of PDCA (Plan, Do, Check, Act) in work management. In formal education, both at primary, secondary or higher education, evaluation is needed to assess quality of education process in various aspects, evaluation of teachers is the one. In higher education, evaluation of lecturers done from various aspects of assessment, lecturer teaching evaluation by students is the one, with evaluation made Internet based so evaluation process will be effective and efficient. Evaluation of lecturers by students called as feedback, because there are input from students to lecturers after one semester. Feedback system is a very useful evaluation instruments for evaluating quality of teaching so made according to needs of things what you want to be extracted from the students. Instruments can be essay and questionnaires, that made by Quality Assurance Department.

Performance assessment questionnaires at least made based on competencies that should be owned by a lecturer in teaching, that is pedagogical competence, professional competence, personal competence and social competence. This competences refers to the Law of Republic of Indonesia Number 14 2005 about Teachers and Lecturers [2]. The results of evaluation of lecturers by students will be used as one indicator of other policies by Human Resources (HR) in college, like lecturer performance index, assessment of credit, increase functional positions and evaluation of employment.

But some challenges found and also confirmed by J. T. Richardson study [3]. In a paper that entitled "Instruments for obtaining student feedback: a review of the literature", a literature review discusses issues in order to gain feedback from students. There are several issues to be discussed, such as why to explore feedback, the instruments used why should be formal, what should be object of evaluation and what are criteria, whether one instrument is suitable for all students and how serious and carelessly feedback is given by students.

The results of literature review concluded that student feedback is very important to assess quality and to support effort of quality improvement. Qualitative instruments are all dictated by educational institutions. Feedback should be focused on student's perceptions of lecturers' teaching affairs or program quality. Feedback should be done as soon as possible after program is implemented. Questionnaires can be made with a wide range of applications, with curriculum innovations likely to lead to a radical overhaul of instrument, moreover, assessment of each program should take differences in context and number of students. The average student who responds to feedback seriously only 60%, with student that respond usually they are that successfully completing the program, but student failing to complete may not respond. Many teachers and students agree that feedback is useful, but its usefulness is not really maximized by teachers and institutions because one of problems is processing and interpreting feedback results. Especially qualitative data.

The problem of feedback utilization is not maximal because the processing problem has discussed in previous research [4] [5], when qualitative data is not utilized. From the above study, feedback system should focus on the teaching aspect with the student's point of view of lecturer, done after completion of course, instrument must be completely contextual and, most importantly, feedback result should be interpreted for all questionnaires, whether quantitative or qualitative, and really utilized.

III. SENTIMENT ANALYSIS

Sentiment, opinion mining or text mining analysis has been done on many data, such as data from social media, web or blog, product reviews and essay of evaluation results. The goal is diverse, as is the orientation. But from such diverse cases, at least similarities can be drawn in terms of methods. The methods used in sentiment analysis are Supervised Machine Learning and Lexicon-based. Supervised Machine Learning requires an example of training data that is used as a reference model for analyzing new data, characterizing probability or frequency of word occurrence. While Lexicon-based required identification of grammar and dictionary of language, with the main characteristic is identification on words or sentences semantic.

There have been many studies that apply, combine [6] [7] [8] and compare two methods [9]. Even in same method, comparisons are made to algorithm used [10]. The results of these studies have not found an agreement on which method is better in giving accuracy, because the results of research show the contradiction of each other. Instead of arguing that one method is better than another, or one algorithm is better

than another, the suitability of the method with the case is more advanced.

Methods and algorithms selected according to case that face. The Lexicon-based method is suitable for less data but has a high sensitivity of meaning. In this study, Supervised Machine Learning method selected because the data used quite a lot but there are carelessly essay, so it can be given examples as a model. Researchers chose Support Vector Machine (SVM) algorithm because SVM is lightweight and easy to use and has been widely used. SVM is able to accommodate a model with concept of Support Vector.

IV. RELATED WORKS

There are some research's that is relevant even underlie this study and from previous research's found deficiencies that will be equipped and used as a recommendation. First, research in Singapore Management University [4] that made a conceptual framework for analyzing student feedback. All this time, only quantitative data that processed for evaluating lecturer although there are qualitative data available too. This framework try to processing that qualitative data by sentiment analysis and try to correlating the result with quantitative data. But this study has done only for sentiment analysis and correlating the result postponed until future work. The result shown that framework can get precision 80.1%, recall 86,4% and F-Score 83,5%.

Research in 2016 [11] has made design of teacher's performance evaluation system that filled out by students, in form of feedback from students and rating. Using lexicon based sentiment analysis, the results is in a score of sentiment. In this design, tendency of students that make essay carelessly is not considered and rating system is not used as a comparison tool for sentiment analysis.

Another study in 2015 [12], simple statistics were used to analyze the results of evaluation of students towards teachers. Instrument of evaluation on Likert scale questionnaires given consists of four categories. The results of this study can be applied in analyzing Likert scale questionnaires.

In 2016, a study [13] made design combination of Machine Learning, Lexicon-based, and Ontology in analyzing text evaluation of teachers. Combination of these methods is expected to generate a positive sentiment, neutral and negative then tested for validation and accuracy of results. Design of this research can be consideration in choice what method will be used for sentiment analysis.

Other relevant research is research by Hamzah in 2014 [5]. Research using a Naive Bayes classifier (NBC) to analyzing sentiment on data that obtained from advice of students after completed course in one semester in IST AKPRIND Yogyakarta. Accuracy of this study reached 85.95%. In preprocessing stage, this study did not clean up advice that made carelessly by student. In addition, no other evaluation systems are used besides just essay.

The latest research [14] that referenced is a study evaluating performance of teachers by fill out an improved method by combining data mining and evaluation indicators that amounted to 12 indicators. The evaluation indicators are categorized poor, good and excellent by students. This research does not use sentiment analysis, but data mining

with tabular data sets. Through Support Vector Machine (SVM) algorithm, can be predicted based on input of teacher performance towards 12 indicators. The results of this study can be applied in analyzing performance assessment questionnaires.

From above previous research's, it can be concluded that there has not been previous research that eliminating carelessly essay at preprocessing stage for sentiment analysis. Whereas with essay that made carelessly, data becomes invalid. So this research will find a solution to eliminate that data.

Furthermore, there is no holistic implementation in educational feedback research which in analysis of the results really combines qualitative approach that is the result of sentiment analysis on essay with quantitative approach that is the result of other assessment system, in this case performance questionnaire. So this study will combine two approaches then performed correlation analysis, to know relationship between two approaches. Because if a quantitative approach produces a high value, ideally a qualitative approach produces a high value too, or otherwise. The ecoefficiency of correlation will also show how success sentiment analysis is.

V. METHOD AND FRAMEWORK SYSTEM

Based on previous research and to solve new problems that found, a framework of educational feedback system as in Fig. 1 has been proposed. Fig. 2 shows detail process in sentiment analysis.

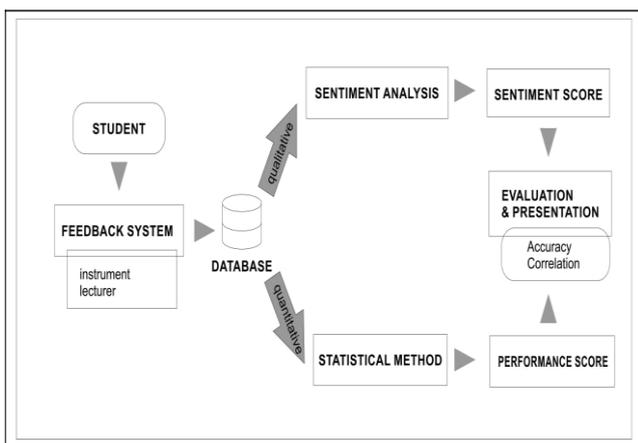


Fig. 1. Framework system

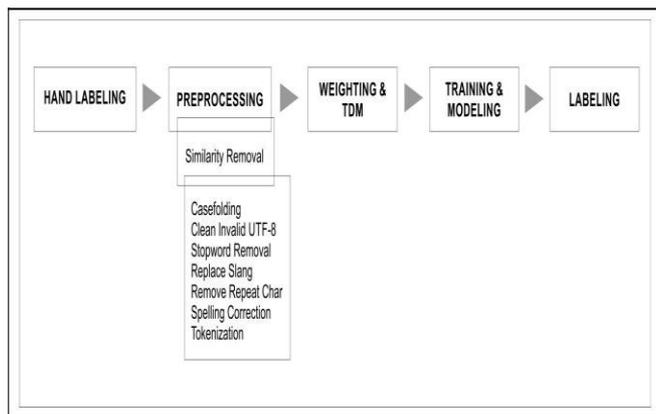


Fig. 2. Sentiment Analysis Process

From Figure 2, in every stage, can described as follows:

A. Feedback System

This is beginning stages of how the data is collected through a evaluation system, filled out by students as an input, through performance assessment instrument with a Likert scale and essay towards lecturers. Web-based evaluation system created to make the evaluation process effective and efficient.

B. Database

The results from student input stored in database, complete with student attributes. Data in database can be categorized into two types of data, quantitative and qualitative. Quantitative data is data from performance assessment questionnaire Likert scale. While qualitative data is essay from students.

C. Statistical Method

This phase converting any performance assessment questionnaires results Likert scale into quantitative data. Each selection is weighted score of 1-4. The better performance, the higher score.

The results of performance assessment questionnaires that has been converted into a score scale of 1-4, then summed by lecturer and divided by number of items questionnaire so this process results are in mean value. Due to number in the thousands, the mean value of these have a high validation.

D. Performance Score

Mean score lecturer is converted to scale of 100 to produce performance score. Qualitative data is further processed in pre-processing stage for sentiment analysis.

E. Sentiment Analysis: Hand Labeling

Previously for training data, has created label manually by Indonesian language expert (hand labelling), in this case Indonesian language lecturers. Hand labelling done by reviewing one by one essay that decided what sentiment will give.

F. Sentiment Analysis: Preprocessing

At this stage, processed data is qualitative data, in the form essay from student. Essay data before analyzed, first performed processing raw data. In this stage, there is a similarity removal process that will resolve problem of students who evaluating carelessly. Carelessly is filling an evaluation but same for all lecturers. The method used in this process is grouping for calculate frequency of occurrence similar essay made by students toward other lecturers. If frequency of occurrence only one, it means that essay is unique and can be used. But if frequency of occurrence more than one, then essay should be removed. Role of student attribute is very important because as identifying who the maker of essay. If essay same but from different students, then data is not deleted. Case folding will change case of word into lower case. Clean invalid UTF-8 will remove other character that is not UTF-8. Stop word removal process will eliminate common words that do not need and do not have sentiment. Tokenization process will transform unstructured text data to be semi-structured in form word of matrix.

G. Sentiment Analysis: Weighting and TDM

After unstructured text data transform into semi-structured data inf form word of matrix, next process is how to give that matrix a weight. One word or token, have different significance than another word. Common word that have many times occurrence, is meaningless than have less occurrence. Weighting used method of TF-IDF (Term Frequency-Invers Document Frequency) that calculate weight by occlumency frequency.

H. Sentiment Analysis: Training and Modeling

Training data that has been given label manually (hand labelling) then trained by SVM algorithm. The training process generates an SVM model that stored in a file. The model contains type and kernel types, classes and vectors. The SVM model can be loaded for labelling purposes.

I. Sentiment Analysis: Labeling

Labelling is process give an essay a value or in this case, sentiment orientation. Training data that has been labelled analysed by algorithm of Support Vector Machine (SVM) to be made a model that consist a pattern. SVM used because based on research [10], it has slightly better accuracy than other algorithms such a Naive Bayes Classifier (NBC). The model was used to labelling essay automatically. The output of this process is orientation whether essay have positive or negative sentiment.

J. Sentiment Score

The essay that already labelled positive or negative sentiment grouped by lecturer so resulting total number of sentiment, number of positive sentiment and number negative sentiment per lecturer. This sentiment number will produce a sentiment score in scale of 100. It can be processed by number of positive sentiment divided by total number of sentiment. Until this stage, obtained each sentiment score and performance score. Both of these scores can be referred as variables, that can be continued to evaluation process and presentation stage.

K. Evaluation and Presentation

To assess accuracy of sentiment analysis using SVM models, samples taken from results of labelling with SVM models then compared with hand labelling. Number of right labelling using SVM models, whether positive or negative sentiment, divided by number of samples to produce accuracy score. In addition, to measure whether there is a relationship between essay questionnaires processed with sentiment analysis with performance assessment questionnaires processed with statistical methods, processed correlation analysis between two variables. Correlation coefficient would indicate that both instruments are feasible to be used in the education feedback system.

VI. RESULT AND DISCUSSION

Research result and discussion will divide into every process, described belows:

A. Feedback System

Stage of feedback system has been implemented in Educational Higher School of Muhammadiyah Kuningan, a college in Kuningan, West Java, Indonesia. By creating a web-based lecturer evaluation system with PHP

programming language and MySQL database. Instruments consist of 21 performance assessment questionnaires and essay column for opinions, criticism and suggestions. Evaluation is done by students every semester after completion of final examination and before students can access grades because to avoid unobjective. Interface systems such as shown in Fig. 3.

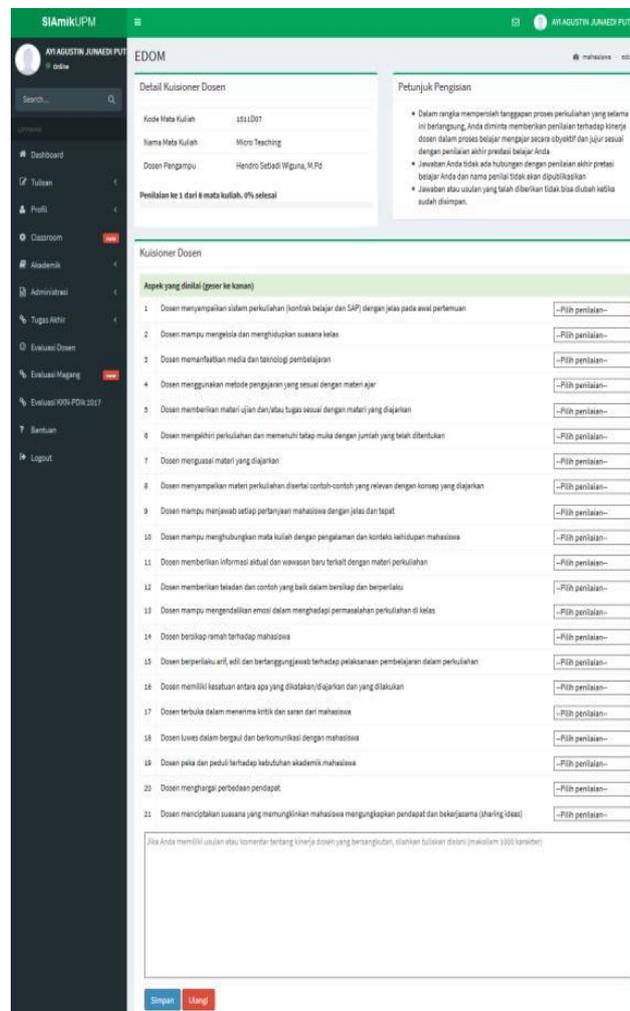


Fig. 3. Web based evaluation system

The evaluation instruments as discussed in the second part, consisting of 4 competence. The instruments are shown in Table I obtained from the Quality Assurance Agency (QAA) according to Internal Quality Assurance System (IQAS).

TABLE I. PERFORMANCE EVALUATION INSTRUMENT

No.	Competence	Number of Items
1	Pedagogic	6
2	Professional	5
3	Personality	5
4	Social	5

Each questionnaire given answers with a Likert scale. Higher the rating, better the score. Table II is a score each answer Likert scale

TABLE II. LIKERT SCALE

Score	Answer
1	Very good / very low / never
2	Bad / low
3	Good / high
4	Very good / very high

From results of implementation of evaluation systems for six semesters 2014/2015 even until 2017/2018 odd, as shown in Table III, quantitative data obtained from questionnaires as many as 584388 performance assessment data and qualitative data from as many as 11070 essay data. Data obtained from 1617 students toward 78 lecturers at Educational Higher School of Muhammadiyah Kuningan.

TABLE III. DATA ACQUISITION

Smt.	Quant.	Qual.	Lect.	Std.
2014-2	77154	2178	48	595
2015-1	112455	2419	51	978
2015-2	78057	1220	63	553
2016-1	108024	1744	66	755
2016-2	89943	1455	66	528
2017-1	118755	2054	69	690
Total	584388	11070	78	1.617

From these data, divided into two stages analysis, namely sentiment analysis for qualitative data and statistical methods for quantitative data.

B. Statistical Method and Performance Score

Quantitative data processed using a SQL query with the command "SELECT AVG (answer) FROM table GROUP BY lecturers". Output of this process is average performance score, then converted to a scale of 100 by equation (1). Table IV shows the results of mean score and performance score of each lecturer.

$$score = \frac{mean \times 4}{100} \tag{1}$$

TABLE IV. PERFORMANCE SCORE RESULT

Id.	Name of Lecturer	Mean	Score
2231001001	Kasdar Al Ade Saputra	3,19	80
2231009002	Dudung Abdu Salam	3,10	78
2231009003	Dodi Ahmad Haerudin	3,02	76
2231009004	Heti Triwahyuni	3,22	81
2231103005	Nanan Abdul Manan	3,14	79
...

Fig. 4. Shows that distribution of score normally have more located in middle area, in range of 76-80.

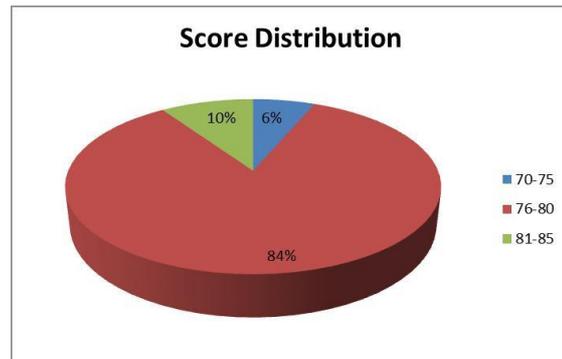


Fig. 4. Performance Score Distribution

C. Sentiment Analysis and Sentiment Score

After a statistical method analysis phase was completed, a sentiment analysis processed using RapidMiner software. The first pre-processing step is similarity removal, by SQL query "SELECT count (*) AS f, id FROM table HAVING f > 1". The frequency of essay that occurrences more than one then removed. Then stop word removal and skimming are done using a self-made language dictionary. Tokenization is done using RapidMiner. In labelling stage, Support Vector Machine (SVM) model created based on 500 essay that labelled manually by Indonesian expert. Fig. 5 shows SVM model creation process.

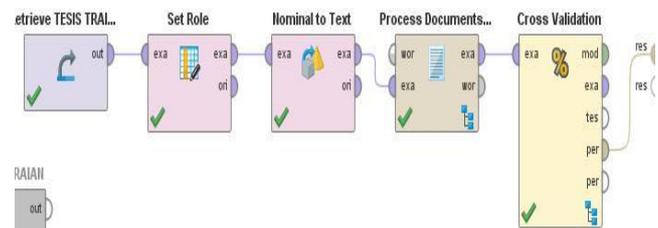


Fig 5. SVM modeling process with Rapidminer

Next step is to apply the SVM model to whole essay data to obtain sentiment label. Fig. 6 shows labelling process.

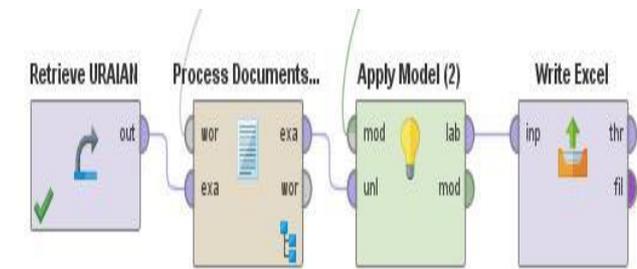


Fig. 6. Labeling process with SVM model

The Result is positive or negative sentiment for each essay in spreadsheet document. Every sentiment grouped by lecturer and by sentiment, then sentiment score calculated by equation (2).

$$sent. score = \frac{positive sentiment}{positive sent. + negative sent.} \times 100 \tag{2}$$

Table V shows the results of sentiment score.

TABLE V. SENTIMENT SCORE RESULT

Name of Lecturer	Positive Sentiment	Negative Sentiment	Score
Kasdar Al Ade Saputra	80	12	87
Dudung Abdu Salam	112	21	84
Dodi Ahmad Haerudin	93	18	84
Heti Triwahyuni	171	16	91
Nana Abdul Manan	73	10	88
...

Fig. 7. Shows that distribution of score normally have more located in middle area, in range of 51-60 and 61-70.

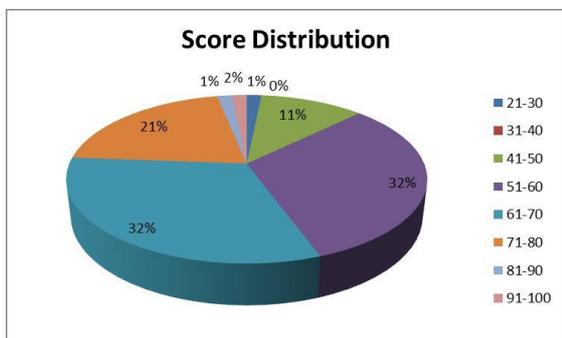


Fig. 7. Sentiment Score Distribution

In order to calculate accuracy of sentiment analysis, a total of 314 samples were taken randomly and results of the sentiment analysis compared with hand labelling. Table VI shows number of labelling results. Equation (3) is used to calculate accuracy.

TABLE VI. SENTIMENT ANALYSIS ACCURATION

SVM/Hand Labeling	Positive	Negative
Positive Label	204 (a)	27 (b)
Negative Label	0 (c)	83 (d)

$$accuracy = \frac{a + d}{a + b + c + d} \times 100 \quad (3)$$

Obtained accuracy of SVM model that created is 91%, which means that accuracy of labelling with sentiment analysis using SVM models is very high.

D. Evaluation and Correlation

Both of above stages resulting performance score and sentiments score, then analysed to measure correlation between two variables whether have a relationship or not. Using SPSS application with Pearson Correlation method 2-

tailed significance, as shown in Fig. 8, obtained correlation coefficient of 0.73, which means two variables have a high correlation according to Table VII. Performance score have high influence and high relationship with sentiment score.

		Performance_score	Sentiment_score
Performance_score	Pearson Correlation	1	,731**
	Sig. (2-tailed)		,000
	N	78	78
Sentiment_score	Pearson Correlation	,731**	1
	Sig. (2-tailed)	,000	
	N	78	78

** Correlation is significant at the 0.01 level (2-tailed).

Fig. 8. Correlation result with SPSS

TABLE VII. INTERPRETATION TABLE

Interval	Level of Relationship
0.00 – 0.199	Very Low
0.20 – 0.399	Low
0.40 – 0.59	Middle
0.60 – 0.79	High
0.80 – 1.00	Very High

VII. CONCLUSION

Educational feedback System is applied to evaluate quality of education, one of the aspects is teaching by teachers. The system built in addition to make evaluation process effective and efficient, as well as to produce an accurate analysis process. Framework designed in this study has been implemented in Educational Higher School of Muhammadiyah Kuningan in evaluating lecturers by students. Feedback system proved able to produce data that required for analysis materials, that is quantitative data as much as 584388 data and qualitative data as much as 11070 data. Sentiment analysis process able to eliminate carelessly essay and capable to produce high accuracy sentiment that is 91%. Both types of questionnaires and analysis process is capable to produce high correlation coefficient that is 0.73. For further research, sentiment analysis process can use other methods such as lexicon based method or Naive Bayes classifier to know whether the result of accuracy will be better than using SVM model.

ACKNOWLEDGMENT

This research supported by Educational Higher School of Muhammadiyah Kuningan, especially Quality Assurance Department and Information Technology Department. Thank you for giving evaluation instrument and evaluation data. May our research result can be used as long as it can be improved teaching quality.

REFERENCES

[1] B. Liu, *Sentiment analysis and opinion mining*. San Rafael: Morgan & Claypool, 2012.
 [2] Indonesia, "Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 Tentang Guru dan Dosen." 2005.

- [3] J. T. Richardson, "Instruments for obtaining student feedback: A review of the literature," *Assess. Eval. High. Educ.*, vol. 30. no. 4, pp. 387–415, 2005.
- [4] S. Gottipati, V. Shankararaman, and S. Gan, "A conceptual framework for analyzing students," in *2017 IEEE Frontiers in Education Conference (FIE)*, 2017, pp. 1–8.
- [5] A. Hamzah, "Sentiment analysis untuk memanfaatkan saran kuesioner dalam evaluasi pembelajaran dengan menggunakan Naive Bayes Classifier (NBC)," in *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, 2014, pp. 17–24.
- [6] O. Kharisman, "Analisis Sentimen pada Review Konsumen Maskapai Penerbangan Menggunakan Kombinasi Lexicon Berbasis Sentiwordnet dan Supervised Model," Universitas Gadjah Mada, Yogyakarta, 2017.
- [7] S. Suryadi, "Analisis Sentimen Review Hotel Menggunakan Algoritme Naive Bayes Classifier dan Pendekatan Lexicon Based," Universitas Gadjah Mada, Yogyakarta, 2017.
- [8] J.K. Wibisono, "Opinion Mining pada Twitter untuk Bahasa Indonesia dengan Metode Support Vector Machine dan Metode Berbasis Lexicon," Universitas Gadjah Mada, Yogyakarta, 2013.
- [9] V. Chandani and R. S. Wahono, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.
- [10] N. M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter," in *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016)*, 2016.
- [11] F. F. Balahadia, M. C. G. Fernando, and I. C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis," in *Region 10 Symposium (TENSYP)*, 2016 *IEEE*, 2016, pp. 95–98.
- [12] R. A. Kassim and N. Buniyamin, "Evaluating teaching quality using data from student online feedback system," in *Engineering Education (ICEED), 2015 IEEE 7th International Conference on*, 2015, pp. 64–68.
- [13] F. de Paula Santos, C. P. Lechugo, and I. F. Silveira-Mackenzie, "„Speak well“ or „complain“ about your teacher: A contribution of education data mining in the evaluation of teaching practices," in *Computers in Education (SHE), 2016 International Symposium on*, 2016, pp. 1–4.
- [14] L. Yu, "Application Research of SVM-based Mining Algorithm in Evaluation of College English Teaching," In *2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)* (pp. 73-76). *IEEE*. 2016, pp. 73–76.