Mining Student Feedback to Improve The Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic

by Zaenal Abidin

Mining Student Feedback to Improve the Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic

Calandra Alencia Haryani
Fakultas Ilmu Komputer Universitas Indonesia[1]
Universitas Pelita Harapan[2]
Jakarta, Indonesia
calandra.alencia@ui.ac.id

Achmad Nizar Hidayanto
Fakultas Ilmu Komputer Universitas Indonesia
Jakarta, Indonesia
nizar@cs.ui.ac.id

Nur Fitriah Ayuning Budi
Fakultas Ilmu Komputer Universitas Indonesia
Jakarta, Indonesia
nurfitriah@cs.ui.ac.id

Zaenal Abidin
Fakultas Ilmu Komputer Universitas Negeri Semarang
Semarang, Indonesia
z.abidin@mail.unnes.ac.id

Theresia Wati
Fakultas Ilmu Komputer UPN "Veteran" Jakarta
Jakarta, Indonesia
theresia.atha@gmail.com

Abstract*— This research carried out the label aspect classification, sentiment analysis, and topic trends on the Open-Ended Question (OEQ) section for Student Feedback Questionnaire (SFQ). Multi-Class aspect label classification for SFQ will choose the best classification model by comparing the results of the evaluation of accuracy, precision, recall, and F1-score for each feature combination and comparison of four classification algorithms namely Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). The results of this research are Classification Techniques using a combination of features of TFIDF, Unigram, and Bigram with the SVM algorithm which is the best Multi-Class classification model for labeling SFQ aspects. In addition, the SentiStrenghtID algorithm used to get sentiments and also the LDA (Latent Dirichlet Allocation) used to get annual topic trends on each survey aspect label. The findings can help Higher Education to support decision making in taking proactive actions towards improvement for self-evaluation and quality.

Keywords— ***Classification, Education Data Mining, Higher Education, Multi Label Classification, Sentiment Analysis, Student Feedback, Survey, Trend Topic***

I. Introduction

Higher Education has an obligation to guarantee and provide quality education to students as one of the stakeholders that has a direct impact to the quality of a university. In Indonesia, Educational Development and Quality Assurance Institution are responsible to ensure the quality assurance of the university to meet the standard provided by the Indonesian Quality Assurance System of Educational Institution. One of their tasks is to carry out measurements and implementation of feedback every semester in the form of surveys as one part of direct quality control to the stakeholders. The results of the surveys can be used as a guideline for continuous improvement in the implementation of Higher Education quality assurance and management of Higher Education.

Conventional processing and measurement of feedback data are not enough to explore hidden information from surveys data [1]. Moreover, conventional processing and measurement are taken a long time to explore information from survey data [2]. The problems that occur are making it difficult for the Higher Education to get optimal results in extracting information and the inaccuracies in the expected improvements to the periodical survey. At present, Higher Education needs a new way to get a more in-depth analysis of survey results so that it can provide useful information for improving university quality, manage survey data more efficiently, and to get survey visualizations.

Education Data Mining (EDM) was used to process and analyze data to discover patterns and extracting information [3], from Higher Education in the form of Feedback Questionaire in the Close-Ended Question and Open-Ended Question (OEQ) section [4]. Student Feedback Questionaire (SFQ) is a feedback survey periodically to measure and obtain student's feedback on the lecturer. SFQ is used because it was the most impactful data information as students were the stakeholders who directly benefited from the quality of education [1]. In addition, student's feedback surveys are one of the tools that can be used to obtain direct evaluations from recipients of teaching and learning [5].

Educational Data Mining's Classification technique is used in this research because it is one of the most utilized Data Mining technique for managing OEQ [1][6][7][8][9][10]. OEQ is a feedback survey section that gives students the freedom to write comments in surveys. Analzing OEQ data is needed because it provides more tangible results experienced by survey recipients than the answers given at the CEQ [7]. OEQ also provides space for survey recipients to convey all the things contained in the recipient's thoughts and feelings without being limited to answering questions at the CEQ [7]. Furthermore, OEQ provides an opportunity for survey recipients to provide spontaneous responses that can provide clarity on answers that are still not obtained from the CEQ [11].

978-1-7281-5118-2/19/$31.00 ©2019 IEEE

Sorour, Goda, and Mine conduct research through Classification to predict student grades for each lesson based on comments form student feedback surveys using Latent Dirichlet Allocation (LDA) and Support Vector Machine (SVM) Algorithms [8]. Agaoglu conduct research through Classification to predict lecturer performance based on comments form student feedback surveys using the Decision Tree Algorithm, Support Vector Machine (SVM), Artifical Neural Network (ANN), and Discriminant Analysis [1]. Koufakou, Gosselin, and Guo conducted research through Classification and Association Rule Mining to get student's sentiments and views on the overall teaching using Naïve Bayes, K-Nearest Neighbor, and Frequent Item Set Algorithm [7]. In addition, Sivakumar and Reddy conducted research through classification and clustering in analyzing student feedback survey data using the Decision Tree, Naïve Bayes, Support Vector Machine (SVM), and K-Means Algorithm [10]. Based on previous studies, classification is used in this research to get the best model for multi-class aspect label, gain sentiments towards the labeling of survey aspects, and trends in survey topics. The purpose of this study is to utilize EDM in managing the feedback data for the OEQ section as an input for improving the quality of Higher Education by supporting decision-making and taking proactive actions towards appropriate improvement for Higher Education self-evaluation. This research is also expected to be a reference for further research regarding the use of EDM in managing feedback data survey.

II. REASERCH METHOD

This section will describe about data set, data processing stages, and evaluation method for this reasearch.

A. Data Set

TABLE 1. Recapitulation of Data Survey Collected

Survey	Period	*Sample Quantity*
SFQ	2017 – Odd Semester	13.440
SFQ	2017 – Even Semester	6.731
SFQ	2017 – Mid Semester	1.976
SFQ	2018 – Odd Semester	12.086

The data used in this paper is a feedback survey for students from University XYZ namely Student Feedback Questionnaire (SFQ) that contains a total of 34.233 data. University XYZ is one of the universities located in Tangerang-Indonesia. Table 1 shows the recapitulation of SFQ Data Collected. SFQ is conducted three times a year on odd semester, even semester, and short semester. The data used is feedback survey data to students obtained from the extraction of survey data through online feedback surveys in Excel. SFQ data is labeling into five aspects such as information provider, role model, facilitator, assessor, and others that show in Figure 1.

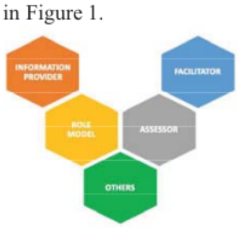

Fig. 1. SFQ Survey Aspect

B. Data Processing Stages

This study uses a quantitative approach in the data analysis phase using Classification Techniques to obtain label classification model aspect of the survey. In addition, this research also conducts sentiment analysis and obtains top tree trends topic in OEQ data survey using the stages of the Knowledge Data Discovery (KDD) process.

This research is processed using the Python programming language as it supports numerous methods and algorithms, along with Open Source libraries, and extension features that can be added based on needs, easy to understand, and easy to create Object functions Oriented [12][13]. In addition, Python is the most suitable programming language for this research because it supports various Data Mining methods and algorithms, especially Text Mining [14].

Data that has been collected will be processed with various stages: Data Cleaning, Data Integration, Data Selection, and Data Transformation [15][16][17][18].

- Data Cleaning is conducted to prepare the dataset before performing classification stages. The survey data that must be cleaned up in this study is that the data is not complete such as a lack of values or attributes on the survey (Incomplete Data). Cleansing data in this study is done by deleting each data that does not contain information or blank or (-) data.
- Data Integration is done by combining survey comments 1, comments 2, and comments 3 on odd semester 2017, 2017 short semester, 2017 even semester, and 2018 odd semester from multiple Microsoft Excel sheets.
- Data Selection is the identification and learning aspect labels in the survey. Aspect labels in the SFQ survey consisted of five labels, namely information providers, role models, facilitators, assessors, and others that show in Figure 1.
- Data transformation is done by labeling or annotating the OEQ survey data of SFQ which are used as data training for the learning process of the Data Mining classification model of this research. Labeling or annotation of OEQ survey data has been annotated by University XYZ survey team. Table 2 shows example of labeling comments in SFQ.

TABLE 2. Example of Comments SFQ

Aspect	Example of Comment
Information Provider	Additional Video Material regarding the subject, makes me understand the class more
Role Model	The kindness and patience from the teacher, clear when explaining the subject encourage me to attend the class
Facilitator	Interactive lecturer with the students, explain the material clearly, not boring.
Assessors	Too much assignment, please reduce it Sir. Don't be stingy giving marks.
Others	Unimporant subject, would never be use in the future. Replace it with the subject that has real correlation with our major

C. Evaluation

The results of data processing stages are ready to use data set, that has been divided into Train Datasets and Test Datasets at the Data Mining stage. After that, Train Dataset and Dataset Test will be processed using K-Means

Algorithm, Decision Tree, Naïve Bayes, SVM, and K-Nearest Neighbor. Figure 2 shows the methodology for analyzing the feedback survey data for the OEQ section used in this research which use Multi Class Classification.

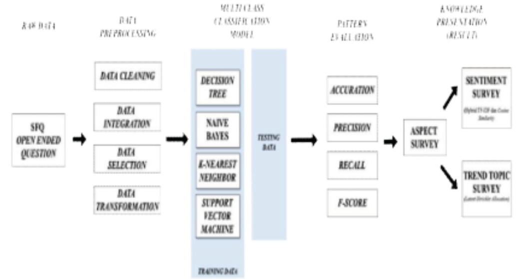

Fig. 2. Survey Methodology SFQ Open Ended Question

After the classification model has been carried out, then the model will be evaluated by measuring Accuration, Precision, Recall, and F1-Score, resulting in a comparison of the best models. The results of the comparison of the best models will be used to obtain maximum results in obtaining the survey aspects labeled SFQ. This study will also conduct an analysis sentiment on the survey and each of the survey aspects labeled SFQ in the OEQ section using the SentiStrengthId algorithm [19]. In addition, this study will also get top three trend survey topics each year on all aspects of the survey using the Sckit Learn Latent Dirichlet Allocation (LDA) library [20].

III. Result and Analysis

This section will describe about result and analysis for this reasearch such as multi class data exploration, model selection and comparison algorithms, sentiment analysis, and trend topic.

A. *Multi Class Data Exploration SFQ Aspects*

In this chapter an analysis of the multi class classification aspects of the OEQ SFQ survey. The multi class classification aspect of the OEQ survey SFQ consists of 5 aspects, namely information providers, role models, facilitators, assessors, and others. After data cleaning, the amount of clean data from the OEQ SFQ survey produced 60,282 data. After normalization at the data mining stage, the number of OEQ SFQ survey data becomes 58,754 data. This study, using the library langdetect to select only comments that use Indonesian. The data used are OEQ SFQ survey data using Indonesian as many as 44,510 data with 19,681 data at Comment 1, 15,615 data at Comment 2, 9,214 data at Comment 3.

The results of the exploration analysis of multi-class data on the OEQ aspect of the SFQ survey resulted in other aspects getting the first highest comment with 19,715 comments, then the information provider aspect had 17,447 comments, the facilitator aspect had 2,757 comments, and the assessment aspects and role models had the same number of comments namely 2,292 comments Figure 3 shows The results of exploration of SFQ aspect data.

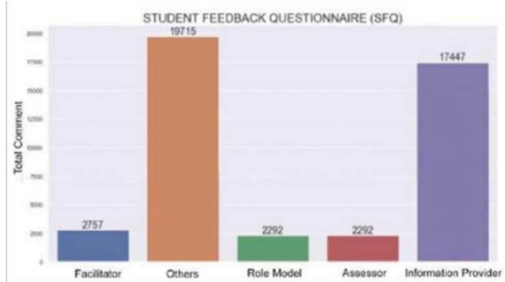

Fig. 3. The results of exploration of SFQ aspect data

The results of the multi class aspects of the OEQ survey SFQ exploration data analysis on Comment 1 resulted in the information provider aspect having the first highest comment with 10,511 comments, the other aspects having 6,478 comments, the facilitator aspect having 947 comments, the assessment aspect having 915 comments, and the role model having 830 comments Figure 4 is Eksploration Result Data Aspek for SFQ Comment 1.

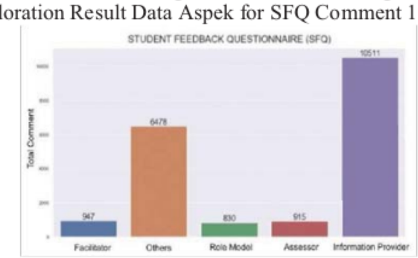

Fig. 4. Eksploration Result Data Aspek for SFQ Comment 1

The results of the multi class exploration analysis of the OEQ aspects of the SFQ survey on Comment 2 resulted in other aspects having the first highest comment with 8,857 comments, information provider aspects having 3,993 comments, facilitator aspects having 1,047 comments, role models having 1,017 comments, and assessors having 701 comments. Figure 5 is Eksploration Result Data Aspek for SFQ Comment 2.

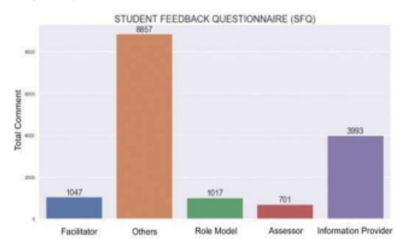

Fig. 5. Eksploration Result Data Aspek for SFQ Comment 2

The results of the multi class exploration analysis of the OEQ aspects of the SFQ survey on Comment 3 resulted in Other aspects having the highest comments in the first rank with 4,372 comments, Information provider aspect having 2,951 comments, Facilitator aspect having 765 comments, Assessment aspects having 682 comments, and Role Models having 444 comments. Figure 6 is Eksploration Result Data Aspek for SFQ Comment 3.

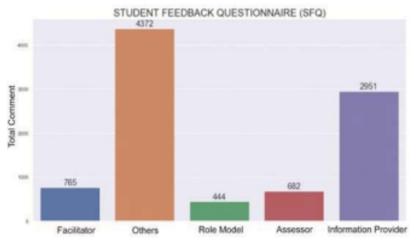

Fig. 6. Eksploration Result Data Aspek for SFQ Comment 3

B. *Model Selection and Comparison of SFQ Survey Algorithms*

This sub-chapter explained the selection of the best classification features and algorithms by looking at the evaluation results. The evaluation results of this assessment are a comparison of the values of accuration, precision, recall, and F1-score for each feature combination and comparison of four classification algorithms, namely Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). The following is a combination of each feature used as a model selection analysis and comparison of SFQ survey classification algorithms:

1. Combination of features A: Count Vectorizer and Unigram.
2. Combination of features B: Count Vectorizer and Bigram.
3. Combination of features C: Count Vectorizer, Unigram, and Bigram.
4. Combination of features D: TFIDF and Unigram.
5. Combination of features E: TFIDF and Bigram.
6. Combination of F features: TFIDF, Unigram, and Bigram.

TABLE 3. Performance Comparison Classification Model (%)

Feature	*Accuration*				*Precision*				*Recall*				*F1-Score*			
	DT	NB	KNN	SVM	DT	NB	KNN	SVM	DT	NB	KNN	SVM	DT	NB	KNN	SVM
A	68.9	71.7	69.4	72.7	68.6	70.6	68.3	71.8	68.9	71.7	69.4	72.7	68.7	70	68.6	72.1
B	62.5	64.1	60	63.7	62.8	62.4	63.8	64.2	62.5	64.1	60	63.7	61	61.6	56.4	62.1
C	69.2	71.5	69.3	72.2	69	70.5	68.5	71.5	69.2	71.4	69.3	72.2	69	69	68.3	71.8
D	68.6	70.9	67.3	73.3	67.9	70.1	66.6	72.4	68.6	70.9	67.3	73.3	68.2	67.9	66.6	72.5
E	61.6	63.9	**55.5**	64.4	62.2	66.4	**65.1**	65.3	61.6	63.4	**55.4**	64.3	60	60	**50**	62.4
F	69.2	**79.8**	69.3	**84.1**	68.9	**73.2**	68.5	**82.3**	69.2	**76.5**	69.3	**82.2**	69	**77.4**	68.3	**82.2**

Table 3 shows a comparison of the performance of the classification model of each feature in the SFQ survey. The first best model for SFQ label aspect survey classification algorithm based on table 3 are the use of combination TFIDF, Unigram, and Bigram features using the Support Vector Machine (SVM) algorithm with an accuracy of 84.1%, precision value of 82.31%, recall value of 82.2%, and F1-Score 82.2%. Furthermore, The second best model for SFQ label aspect survey classification algorithms based on table 3 is the use of a combination of features of TFIDF, Unigram, and Bigram using the Naïve Bayes algorithm (NB) with the value of accuracy 79.8%, precision value 73.2%, recall value 76.5%, and the value of F1-Score is 82.2%. Besides, table 3 shown that the lowest evaluation model value for SFQ label aspect survey classification is the combination of TFIDF and Bigram features using the K-Nearest Neighbor (KNN) algorithm with an accuracy of 55.5%, precision 65.1%, value 55.4% recall, and F1-Score 50%. Therefore, this research chose the SVM algorithm with features of TFIDF, Unigram, and Bigram as a label classification model for the SFQ survey label aspect.

C. *Sentiment Analysis SFQ Survey*

This sub-chapter explained about the analysis of SFQ survey sentiment classification with the SentiStrenghtID algorithm using Hybrid TF-IDF and Cosine Similarity. The data used is data from the Preprocessing stage. Figure 7 shows the results of the sentiment classification in the SFQ survey. The sentiment classification results showed that the comments given in the SFQ survey received as many positive sentiments as there were 24,315 comments compared to neutral sentiments of 16,199 comments and negative sentiments of 3,159 comments.

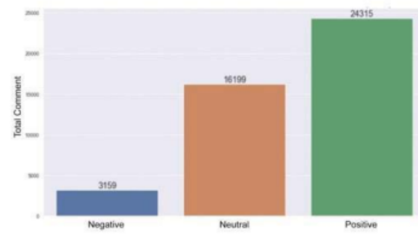

Fig. 7. SFQ Sentiment Analysis

In addition, the SFQ survey sentiment classification analysis in this study will also compare the results of the sentiment classification with the results of the SFQ aspect label classification. The comparison results show that on the label the aspects of Information Provider, Role Model, Assessor, and Others get more positive sentiments than neutral sentiments and negative sentiments, while Facilitator aspect labels get more neutral sentiments than positive sentiments and negative sentiments. Figure 8 shows the results of a label comparison of aspects of sentiment in the SFQ survey.

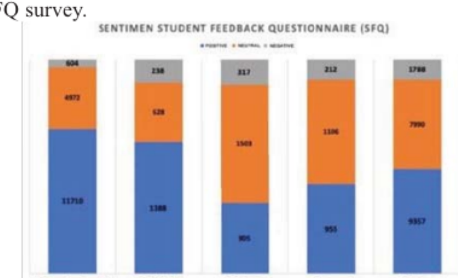

Fig. 8. SFQ's Aspect Label Sentiment Comparison Result

The SFQ survey sentiment evaluation was conducted to test the results of SFQ survey sentiment predictions using the SentiStrenghtID algorithm by calculating accuracy, precision, recall, and F1-score. The evaluation was carried out on 300 SFQ survey data that had been manually labeled. SFQ survey sentiment evaluation using the SentiStrenghtID algorithm produces an accuracy value of 75.7%, a precision value of 75.2%, a recall value of 76.7%, and an F1-Score value of 75.3%. Based on the evaluation results it can be concluded that the SentiStrenghtID algorithm can be used to identify SFQ survey sentiments. Figure 9 shows the results of visualization of the confusion matrix from SFQ survey sentiment evaluation using the SentiStrenghtID algorithm.

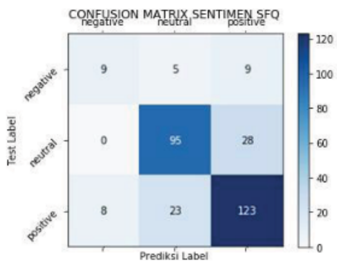

Fig. 9. SFQ's Confusion Matrix Sentiment

D. *Trend Topic SFQ Survey*

This sub-chapter explains the analysis of SFQ survey trends in 2017 and 2018 regarding label classification aspects and topic trend modeling using the Sckit Learn Latent Dirichlet Allocation (LDA) library in each aspect.

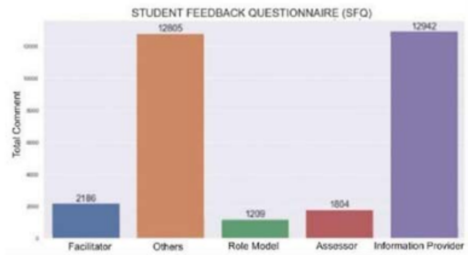

Fig. 10. Eksploration Result SFQ 2017

The results of the SFQ survey trend analysis in 2017 resulted in the highest aspect label classification, namely the information provider aspect of 12.942 comments and other aspects as many as 12.805 comments, while the lowest aspect label classification was an assessment aspect of 1,804 comments. Figure 10 show Eksploration Result SFQ 2017. The results of the SFQ survey trend analysis in 2018 resulted in the highest aspect label classification being 6.924 comments, while the lowest aspect label classification was the facilitator aspect of 559 comments. Figure 11 show Eksploration Result SFQ 2018.

Fig. 11. Eksploration Result SFQ 2018

In addition, this research conducted topic modeling of top three topics using the Scikit Learn LDA library to find trends in 2017 and 2018 SFQ survey topics on each survey aspect label. The results for trend SFQ topic model in 2017 are shown in following Table 4.

TABLE 4. Trend SFQ Topic Model in 2017

Rank	Aspect	Topic
1	Others	Topic 1: Kind
		Topic 2: Class
		Topic 3: Passionate
2	Information Provider	Topic 1: Material
		Topic 2: Teaching
		Topic 3: Explanation
3	Rule Model	Topic 1: Quick
		Topic 2: Class
		Topic 3: Understand
4	Facilitator	Topic 1: Group
		Topic 2: Room
		Topic 3: AC
5	Asessment	Topic 1: Practice
		Topic 2: Assignment
		Topic 3: Input

The results for trend SFQ topic model in 2018 are shown in following Table 5.

TABLE 5. Trend SFQ Topic Model in 2017

Rank	Aspect	Topic
1	Information Provider	Topic 1: Easy
		Topic 2: Explanation
		Topic 3: Teaching
2	Others	Topic 1: Kind
		Topic 2: Situation
		Topic 3: Sleepy
3	Facilitator	Topic 1: Wifi
		Topic 2: Discussion
		Topic 3: Class
4	Asessment	Topic 1: Example
		Topic 2: Practice
		Topic 3: Assignment
5	Rule Model	Topic 1: Fair
		Topic 2: Attention
		Topic 3: Class

IV. CONCLUSION AND FURTHER RESEARCH

This section will describe about conclusion and further research for this reasearch.

A. *Conclusion*

The conventional processing and measurement method to labeling the OEQ aspects of SFQ takes a long time. This has caused Higher Education difficulty in extracting survey results to classify surveys on predetermined aspects, obtain stakeholder sentiments, and analyze surveys appropriately quickly and optimally. The use of EDM by using classification techniques can help accelerate university to obtain the OEQ survey aspects and analysis of the SFQ surveys. The results of a fast and optimal label classification aspect of the survey, sentiment, and topic trends can be used as evaluation materials to develop strategies for improving university quality and services.

This study has conducted experiments on each feature combination and comparison of four classification algorithms such as Decision Tree, Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine (SVM) to obtain the Multi Class classification model for labeling SFQ aspect. The combination of TFIDF, Unigram, and Bigram features get the highest evaluation value such as accuration 84.1%, precision 82.31%, recall 82.2%, and F1-Score 82.2%. The SFQ survey is not recommended to use the TFIDF and bigram features with the K-Nearest Neighbor (KNN) algorithm because its only obtain an accuracy value of only 55.5%, the precision value of 65.1%, the recall value of 55.4%, and the F1-Score value of 50. Each survey has a different aspect label that can be used as a deeper input into the university. The SFQ survey produced more positive sentiments for each aspect label than neutral sentiments and negative sentiments. These results indicate that every label aspect of the survey conducted by the university has met the standards of meeting student satisfaction standards therefore the university can improve and maintain the quality and quality of each aspect label that gets a lot of positive sentiments. In addition, neutral sentiments can be analyzed more deeply to get student suggestions for each aspect label because neutral sentiments contain many suggestions that do not have positive sentiments or negative sentiments. The results of the SFQ survey sentiment evaluation using the SentiStrenghtID algorithm produce an accuracy value of 75.7%, a precision value of 75.2%, a recall value of 76.7%, and an F1-Score value of 75.3%. The evaluation results are good enough to make this algorithm used to get student sentiment towards the label aspects of student comments in the SFQ survey.

The Latent Dirichlet Allocation (LDA) algorithm is used to find hidden topics in a text to find out topics that are often commented on by students on surveys [6]. This study also uses the LDA Algorithm using the Sckit Learn Latent Dirichlet Allocation library to find three trending topics in each aspect of SFQ survey comments. Analysis using LDA is able to generate trends in the topic of student comments that represents important topics for students. The same topic trends can also represent topics that must be evaluated by universities every year because they are important for students to follow up on. The topic trend then could help universities to get a visualization of overall student comments. Therefore, the results of this topic trend can be used by the university as input in determining the right strategy in improving services and student satisfaction to improve the quality of the university.

B. *Further Reasearch*

This research has disadvantages due to limitations at the time of the study. The suggestion for further research are as follows:

1. Using EDM with techniques such as Association Rule Mining at CEQ to get a more in-depth analysis of CEQ results.
2. This research only uses a dictionary provided by the SentiStrenghtID algorithm to get student sentiment. In further research, it is recommended to use several algorithms to get better sentiment results.

ACKNOWLEDGEMENT

This study was supported by PIT9 Research Grant No NKB-0005/UN2.R3.1/HKP.05.00/2019 Universitas Indonesia

REFERENCES

[1] M. Agaoglu, "Predicting Instructor Performance Using Data (Sahu, Shrma, & Gondhalakar, 2017) Mining Techniques in Higher Education," *IEEE Access,* pp. 2379 - 2387, 2016.

[2] R. Saptarshi, "Big data in education," Gravity, the Great Lakes Magazine, vol. 20, pp. 8-10, 2013.

[3] N. Rajadhyax.and and R. Shirwaikar, "Data mining on educational domain," arXiv Preprint arXiv:1207.1535, pp. 1-6, 2012.

[4] C.´. Romero and S.´. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics,* pp. 601-618, 2010.

[5] A. Almahdi, S. Ahmida, A. A. Bakar, A. R. Hamdan and M. Alweshah, "Mining Survey Data on University Students to Determine Trends in the Selection of Majors," in *Science and Information Conference* , London,UK, 2015.

[6] A. M. Ahmed, A. Rizaner and A. H. Ulusoy, "Using data mining to predict instructor performance," in *12th International Conference on Application of Fuzzy Systems and Soft Computing*, Austria, 2016.

[7] A. Koufakou, J. Gosselin and D. Guo, "Using Data Mining to Extract Knowledge from Student Evaluation Comments in Undergraduate Courses," in *International Joint conference on Neural Networks*, Vancouver, BC, Canada, 2016.

[8] S. E. Sorour, K. Goda and T. Mine, "Estimation of Student Performance by Considering Consecutive Lessons," in *International Congress on Advanced Applied Informatics*, Okayama, Japan, 2015.

[9] H. Aldowah, H. Al-Samarraie and W. M. Fauzy, "Educational Data Mining and Learning Analytics for 21st Century Higher Education: A Review and Synthesis," *Telematics and Informatics,* pp. 13-49, 2019.

[10] M. Sivakumar and U. S. Reddy, "Aspect Based Sentiment Analysis of Students Opinion using Machine Learning Techniques," *International Conference on Inventive Computing and Informatics,* 2017.

[11] R. Popping, "Analyzing Open Ended Question by Means of Text Analysis Procedure," *BMS Bulletin of Sociological Methodology,* pp. 23-39, 2015.

[12] Phyton Software Foundation, "About Phyton," [Online].

[13] A. Nagpal and G. Gabrani, "Python for Data Analytics, Scientific and Technical Applications", 2019 Amity International Conference on Artificial Intelligence (AICAI).

[14] J. Mitrapanont, W. Sawangphol, T. Vithantirawat, S. Paengkaew, P. Suwannasing, A. Daramas, and Y. Chen. A Study on Using Python vs Weka on Dialysis Data Analysis". 2017 International Conference on Information Technology (INCIT).

[15] J. Han, M. Kamber and J. Pei, Data Mining, Elsevier Inc, 2012.

[16] H. Sahu, S. Shrma and S. Gondhalakar. "A Brief Overview on Data Mining Survey". 2011 International Journal of Computer Technology and Electronics Engineering (IJCTEE).

[17] Fernandez, D. B., & Lujan-Mora, S. (2017). Comparison of applications for educational data mining in Engineering Education. 2017 IEEE World Engineering Education Conference (EDUNINE).

[18] C. Márquez, C. Romero.and S.Ventura, "Predicción del fracaso escolar mediante técnicas de minería de datos," IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 7, nº 3, pp. 109117, 2012.

[19] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *Indonesian Journal of Computing and Cybernetics Systems,* pp. 207-218, 2016.

[20] Scikit-learn: Machine Learning in Python, Pedregosa *et al*., JMLR 12, pp. 2825-2830, 2011.

Mining Student Feedback to Improve The Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic

ORIGINALITY REPORT

8%	0%	0%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Telkom University Student Paper	8%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	Off		

Mining Student Feedback to Improve The Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6