Mining Student's Reviews to Obtain Their Perception toward College Department Performance

1st Yuliska
Informatics Engineering
Politeknik Caltex Riau
Pekanbaru, Indonesia
yuliska@pcr.ac.id

2nd Dini Hidayatul Qudsi
Information System
Politeknik Caltex Riau
Pekanbaru, Indonesia
dinihq@pcr.ac.id

3rd Lia Anggraini
Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
23520037@std.stei.itb.ac.id

4th Juanda Hakim Lubis
Informatics Engineering
Universitas Medan Area
Medan, Indonesia
juandahakim@gmail.com

Abstract—Student's perception toward department performance can be crucial, therefore, it can be used to evaluate the department outcome and take an immediate action to improve its management. This study applies sentiment analysis and topic modeling to the student's reviews of college department at Politeknik Caltex Riau in order to mine student's perception for seven college departments performance. Sentiment analysis with Support Vector Machine (SVM) is employed to obtain student's sentiment. There are 3 types of sentiments to be analyzed; positive, negative and neutral. Topic modeling with Latent Dirichlet Allocation (LDA) is also carried out to get some important keywords in the student's reviews. Our experiments show that Positive is the most prominent sentiment in the student's reviews while LDA reveals some important topics toward preferences.

Keywords—Student's Review, Student's Perception, Department Performance, Sentiment Analysis, Topic Modeling

I. INTRODUCTION

As technology develops, more and more users use the internet as a source of information. The role of the internet as a source of information makes users use the internet to share their personal opinions in the form of reviews. The review can be in the form of a user's personal opinion about the quality of a movie, products, hotel services and services of an institution. Many websites provide a questionnaire feature so that users can express their opinions. The results of the review will then be used for better decision making [1].

In this study, we utilize student's reviews on the performance of seven college departments at Politeknik Caltex Riau (PCR) to mine student's perception toward those departments. The student's reviews on the performance of each department can reach 1500 reviews; with the number of words can be up to 150 of each review. Therefore, mining these many and long reviews becomes essential because it allows the department to improve their performance based on the mining result.

The mining process is performed using sentiment analysis and topic modeling. We carry out Sentiment analysis with Support Vector Machine (SVM) [2] to get 3 types of student's sentiments in the review, namely, positive, negative and neutral. The Latent Dirichlet Allocation (LDA) [3] is then used to reveals some important keywords hidden on the student's review.

II. LITERATURE REVIEW

A. Sentiment Analysis

Sentiment analysis is a process that analyzes and detects the sentiment of a text input that has positive, negative or neutral sentiments [4]. However, until now, the sentiments that can be detected have become more diverse and detailed and are not limited to only positive and negative, but also happiness, sadness, anger, fear, disgust and surprised [5].

Sentiment analysis can be used to monitor the performance of a product or institution's services. By applying sentiment analysis, product developers and service owners can easily find out whether a product or service is received positively by customers or vice versa.

B. Support Vector Machine (SVM)

Support Vector Machine is one of the powerful machine learning techniques for data and text classification. It is also a supervised method, so we need labeled dataset to train a text classifier which based on Support Vector Machine.

Suppose we have some data points which each belongs to one of 2 classes, and the goal is to predict which class those points are belong. Support Vector Machine works by building some hyperplanes between those points that might classify the data points. An optimal hyperplane is the one with the largest margin or separation between the two classes.
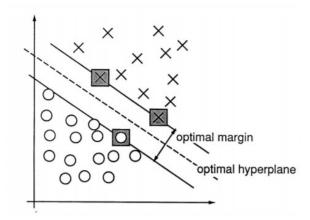
Fig. 1. An Optimal Hypeplane [2]

C. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation is a method which is largely used for topic modeling. It is a technique used to find some topics or hidden keywords on a collection of text document. Suppose we have following documents which contains some words:

Document1: $word_1$, $word_2$, $word_3$, …, $word_n$
Document2: $word_1$, $word_2$, $word_3$, …, $word_n$
Document3: $word_1$, $word_2$, $word_3$, …, $word_n$

The words on each document are ordered by the frequency of its occurrence. With Latent Dirichlet Allocation, we can figure out the words which belong to different topics, as can be seen in Table I below:

TABLE I. WORDS WITH PROBABILITY SCORE ON DIFFERENT TOPICS

Topic	$Word_1$	$Word_2$	$Word_3$	$Word_n$
Topic1	0.23	0.001	0.019	0.007
Topic2	0.015	0.49	0.02	0.011
Topic3	0.010	0.40	0.021	0.001

III. EXPERIMENT

In this section, we describe some stages we conducted during the experiment, the datasets, and our experimental settings.

A. Dataset

We used Quality Assurance (QA) dataset of the department performance at Politeknik Caltex Riau (PCR) in the second semester of the Academic Year of 2018/2019. It is generated from the PCR Planning, Development and Quality Assurance Department, called BP3M. Elaborately, the data contained of reviews from students on the performance of seven college departments at PCR. The seven departments consist of Academic and Student Administration (BAAK), Information System and Technology (BSTI), Student department, Cooperation, Finance, Library, and Infrastructure department. The number of reviews in each department is from 1470 to 1500, but only 1200 reviews will be used, while the remaining 300 reviews are eliminated based on the number of words contained in it. The number of words in the reviews ranges from 2 words to 150 words. Table II shows an example of reviews in one of the departments.

TABLE II. AN EXAMPLE OF REVIEWS OF DEPARTMENTS

Number	Reviews
1	Pelayanan Sangat baik dan diberikan arahan dengan baik dan ramah ☺☺
2	Terapkan pembayaran ansuran agar mahasiswa yang tidak mampu dapat kemudahan
3	Jangan sampai antrian panjang menanti
4	Pelayanan yang d berikan sudah baik, penjaga nya juga ramah, dalam menindaklanjuti pembludakan pembayaran serentak yang dilakukan mahasiswa harus d perbaiki.
5	Saran dari saya supaya tempat pembayaran SPP di tambah karna menurut saya kalau 1 bank masi kurang cukup, sebab admin bank akan kerepotan mengurus mahasiswa yang mengurus SPP tersebut
6	Bagian keuangan sudah memberikan pelayanan dan tanggapan yang memuaskan

B. Automatic Labelling

To be able to use the data explained in sub section A to train SVM model, we employed an automatic labeling with a tool called *SentiStrength* [6]. *SentiStrength* gives label to the review based on the sentiment score. For instance, if the total scores of words which have positive sentiment greater than the score of words with negative and neutral sentiment, then the review is labeled as positive review. This scenario is also applied to negative and neutral label.

In this study, we take 200 reviews from each department for training purpose. Therefore, there are 1400 reviews used for training. Subsequently, we perform automatic labeling to give sentiment label on each review.

C. Text Preprocessing

This study uses *NLTK (Natural Language ToolKit)* as a library for word processing. The preprocessing stages applied in this study are as follows:

1) Case Folding: The process of converting the uppercase to lowercase.

TABLE III. CASE FOLDING

Review	*Case Folding*
Pelayanan Sangat baik dan diberikan arahan dengan baik dan ramah ☺☺	pelayanan sangat baik dan diberikan arahan dengan baik dan ramah ☺☺

2) Stop word Removal: The process of eliminating the unnecessary words.

TABLE IV. STOPWORD REMOVAL

Review	*Stopword Removal*
pelayanan sangat baik dan diberikan arahan dengan baik dan ramah ☺☺	pelayanan sangat baik diberikan arahan baik ramah ☺☺

3) Emoticon Removal: The process of removing the emoticons in the text.

TABLE VI. EMOTICON REMOVAL

Review	Emoticon Removal
pelayanan sangat baik dan diberikan arahan dengan baik dan ramah ☺☺	pelayanan sangat baik diberikan arahan baik ramah

4) Tokenization: The process of tokenizing sentence into a collection of words.

TABLE VII. TOKENIZATION

Review	Tokenization
pelayanan sangat baik diberikan arahan baik ramah	["pelayanan", "sangat", "baik", "diberikan", "arahan", "baik", "ramah"]

D. Experimental Setting

We use a library called *Scikit-Learn* to train SVM model, with the kernel is set to 'linier', regularization parameter is set to 1.0, and the remaining parameters are set to default. SVM model is trained using the data explained in sub section B, and the model will be employed to predict the sentiment of 1000 reviews on each department. In the next stage, to perform Latent Dirichlet Allocation, we still utilize *Scikit-Learn* Library with all parameters is set to default. For word representation, we use TF-IDF representation. In addition, because data training is imbalance, we apply oversampling to balance it.

IV. RESULT & DISCUSSION

This section reports on our experimental results and discusses the reason behind those results.

A. Sentiment Analysis Result

As we explained in section III, we used 1400 labeled review for training purpose. During training, the SVM model obtained the accuracy of 96%. However, we have to mention that although the accuracy of the model is good, we found that the model do not perform effectively in predicting negative sentiment. We then use this model to predict the 1000 reviews on each department. The results are shown below:
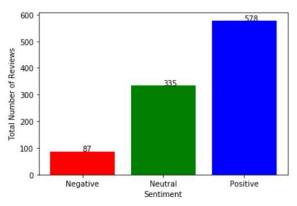

Fig. 2. Sentiment Analysis of BAAK

Figure 1 shows the review sentiments for Academic and Student Administration (BAAK) department. It shows that reviews with positive sentiment are the most prominent. There are 578 positive reviews, 335 neutral reviews, and 87 negative reviews.
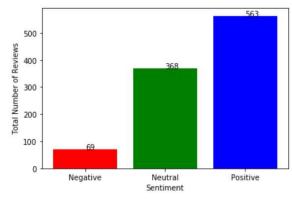

Fig. 3. Sentiment Analysis of BSTI

The review sentiments for Information System and Technology (BSTI) department can be seen in Figure 2. It shows that positive sentiments are also the most notable review. In this department, there are 563 positive reviews, 368 neutral reviews, and 69 negative reviews. It can be seen that Information System and Technology (BSTI) department has less positive and negative reviews than Academic and Student Administration (BAAK) department.
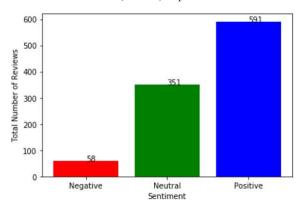

Fig. 4. Sentiment Analysis of Student Department

Moreover, the review sentiments of Student Department have the same trend with Academic and Student Administration (BAAK) and Information System and Technology (BSTI) departments which the positive sentiment is the most significant, and the negative reviews have the least total number. This department has 591 positive reviews, 351 neutral reviews, and 58 negative reviews. It can be observed that this department has more positive reviews and less negative reviews than the previous two departments.

After that, we analyze the sentiment analysis result of Finance department. As can be seen in Figure 4, the Finance Department also has the same trend with previous departments. In this department, reviews with positive sentiment have the highest number and reviews with negative sentiment come with the least number. There 601 positive reviews, 347 neutral reviews and 52 negative reviews.
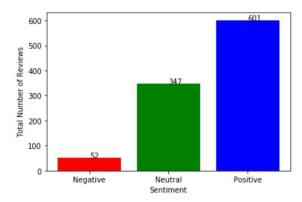
Fig. 5. Sentiment Analysis of Finance Department

Figure 5 shows the review sentiments for Cooperation Department. Expectedly, this department has the same trend with the sentiment analysis of previous departments. In this department, the positive sentiments also become the most striking review with the total number of 600. Additionally, with 41 negative reviews, this department has the least number of negative sentiments than the previous departments. The neutral sentiment comes with the number of 359 reviews.
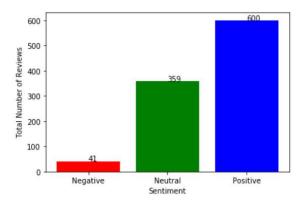

Fig. 6. Sentiment Analysis of Cooperation Department

Figure 6 and 7 display the review sentiments for Infrastructure Department and Library, respectively. These departments have the same trend that the positive reviews have the highest number. It is followed by neutral reviews in the second high number and negative reviews with the least number. As can be seen in Figure 6, there are 90 negative reviews in Infrastructure department, which is the highest among 7 departments.
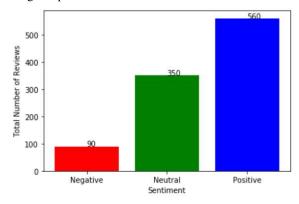

Fig. 7. Sentiment Analysis of Infrastructure Department

Furthermore, we analyze the reviews which have positive sentiments in 7 departments, and we found that all departments have similar reviews for positive sentiments. For positive sentiment, "good service" is the common topic.
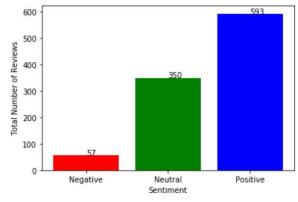

Fig. 8. Sentiment Analysis of Library

B. Topic Modelling Result

We perform topic modeling on the positive reviews with Latent Dirichlet Allocation (LDA) with the number of topics is set to 10 topics. Figure 8 shows the information about topic word weight or words with probability score in 10 different topics.

	akademik	alhamdulillah	baak	bagus	belajar	berani
Topic0	0.100007	21.400820	0.100022	0.100086	0.100017	0.100009
Topic1	0.100019	0.100011	0.100020	0.100024	0.100016	0.100010
Topic2	0.110206	0.100021	0.147426	506.581892	0.100012	0.100010
Topic3	0.100009	0.100009	0.100018	0.100020	0.100012	0.100010
Topic4	0.100026	0.100008	62.656136	0.100035	0.100013	0.100008
Topic5	0.100013	0.100008	0.101251	0.100024	0.100015	0.100008
Topic6	0.100008	0.100009	0.100083	0.100025	0.100019	0.100008
Topic7	25.028356	0.100010	35.942779	0.100014	0.100033	0.100008
Topic8	0.100012	0.100009	0.100016	0.100010	0.100017	0.100008
Topic9	0.100015	0.100009	0.100013	0.100142	14.953149	11.360504

Fig. 9. Topic Word Weight

The top 10 topic keywords in 10 different topics can be observed in Table VII.

TABLE VIII. TOP 10 TOPIC KEYWORDS

Topic	Keywords
Topic 1	Mantap, semangat, ruangan, alhamdulillah, bersih, lumayan, oke, konsisten, layanannya, puas
Topic 2	Terimakasih, sarana, sukses, prasana, ditingkatkan, membantu, cepat, diharapkan, maju, keramahan
Topic 3	Kedepannya, bagus, semoga, layanan, memuaskan, puas, kemahasiswaan, pelayanan, baak, sistem
Topic 4	Terimakasih, pelayanannya, maju, kinerja, semoga, mohon, kedepannya, keuangan, sistem, tolong
Topic 5	Pelayanan, baak, sistem, tingkatkan, informasi, industri, senyum, mudah, dipertahankan, teknologi
Topic 6	Keuangan, saran, terbaik, tolong, fasilitas, layanan, suka, pelayanan, mahasiswa, semoga
Topic 7	Perpustakaan, kerjasama, rapi, kedepannya, buku, keren, sopan, melayani, industri, terimakasih

Topic 8	Mahasiswa, ramah, baak, melayani, proses, akademik, pengelola, mohon, pembelajaran, sesuai
Topic 9	Semoga, nyaman, kampus, berkembang, depannya, senang, penilaian, pcr, layanan, perpusatakaan
Topic 10	Pcr, kedepan, pertahankan, lanjutkan, kerja, kembangkan, belajar, ruang, berani, meningkatkan

The common keywords on the 10 topics are "puas", "bagus", "ramah", and "terimakasih". The first, the second and the last topics are mainly related to the service of infrastructure department. The third is notably focused on the service of Student Department. Furthermore, the fourth and the sixth topic are mainly related to the Service of Finance Department. The fifth and the eight topics consist of the keywords related to the service of Academic and Student Administration (BAAK) department. Lastly, the seventh and the ninth topics are discussed about the service of Library. As can be seen in Table VII, all keywords on each department show that the students were satisfied with the service of all departments. From the keywords, in addition, we can also conclude that the students hoped for the department to improve their performance and service in the future.

V. CONCLUSIONS & FUTURE WORK

Our experiment implemented sentiment analysis with Support Vector Machine and Topic Modeling with Latent Dirichlet Allocation to mine student's perception toward performance of college departments. Our findings are summarized as follow:

- Sentiment analysis can be utilized to discover the sentiment expressed by students through their reviews on the performance of college departments. Based on the sentiment analysis results, it can be concluded that the Finance and Cooperation departments have the highest number positive reviews and has the least number of negative reviews, respectively.

- Topic modeling can be used to identify some hidden keywords associated to some topics. With Topic modeling, we can find the characteristic of each department through the keywords displayed on each topic.

We perform sentiment analysis with Support Vector Machine (SVM). The model has great accuracy, but has poor performance in predicting negative sentiments. Thus, utilizing another machine or deep learning model might result in better and more accurate performance.

REFERENCES

[1] Raut, V. B., & Londhe, D. D. (2014). Opinion mining and summarization of hotel reviews. *Proceedings - 2014 6th International Conference on Computational Intelligence and Communication Networks, CICN 2014*, 556–559. https://doi.org/10.1109/CICN.2014.126

[2] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995). https://doi.org/10.1007/BF00994018

[3] Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3, 993--1022. doi: http://dx.doi.org/10.1162/jmlr.2003.3.4-5.993

[4] Gupta, P., Tiwari, R., & Robert, N. (2016). Sentiment analysis and text summarization of online reviews: A survey. *International Conference on Communication and Signal Processing, ICCSP 2016*, 241–245. https://doi.org/10.1109/ICCSP.2016.7754131

[5] Seyeditabari, A., Tabari, N., & Zadrozny, W. (2018). *Emotion Detection in Text: a Review*. http://arxiv.org/abs/1806.00674

[6] Wahid, D. H., & SN, A. (2016). Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *10*(2), 207. https://doi.org/10.22146/ijccs.16625