Indonesian sentiment summarization for lecturer learning evaluation by using textrank algorithm

I G M Darmawiguna¹, G A Pradnyana² and I B Jyotisananda³

¹,²,³Informatics Department, Faculty of Engineering & Vocational Universitas Pendidikan Ganesha

Email: mahendra.darmawiguna@undiksha.ac.id

Abstract. Learning evaluation is a very important process to assess lecturer performances. Evaluation is carried out at the end of the semester by students. Student can give comments, critics, suggestions related to the subject taught by the lecturers. To analyse student comments, the lecturer can read the entire comments one by one. The main focus in this research was to develop the system that will be applied in summarizing student comments on the implementation of course learning. The algorithm that was used to perform text summarization was the TextRank algorithm. TextRank is a graph-based ranking algorithm (a graph with a ranking model) for processing text from natural or human language documents. We used 100 data sentiments by students to lecturers who teach one course in one semester. It was directly taken from SIAK (academic information system) by using web services. Based on the evaluation of the summaries by the expert, it was found that the textRank algorithm is quite representative in performing summarization for sentiments with the accuracy of 82%.

1. Introduction

The evaluation process is very important for the running of an activity or program. Usually the purpose of conducting an evaluation is to find out whether the activities or programs that have been carried out have been conveyed to participants well, or are in accordance with the targets/objectives of the program, or not at all. In education, the evaluation that is often done is the evaluation of the implementation of learning. In tertiary institutions, the learning evaluation process is carried out at the end of the semester. This process is carried out to assess lecturer performance. It can be seen that the effect of lecturer teaching on students. Good teaching will help students achieve good learning. The quality of teaching and academic standards certainly need to be evaluated and improved. Good teaching is a challenge for every lecturer. Basically, lecturer evaluation is used to identify the contribution of the lecturer in achieving the objectives of the study program and assess the need for the lecturer for guidance and training in a particular teaching field.

Universitas Pendidikan Ganesha (Undiksha) provides student facilities to provide evaluation of lecturers' teaching in the form of closed and open questionnaires facilitated in the Undiksha Academic Information System. The closed questionnaire is a questionnaire containing questions or statements where alternative answers have been included in the questionnaire so that students only choose from these alternative answers, while the open questionnaire is a questionnaire that contains a list of questions that give students opportunity to write their opinions about the questions given in the questionnaire. The results of the closed questionnaire are easier to be analysed with appropriate statistical techniques, but on the other hand, open questionnaires, namely commentary analysis, are quite difficult to do. Some researchers who examine learning evaluation only focus on quantitative data, namely closed
questionnaire analysis, whereas qualitative learning evaluation data such as comments from the student's point of view on the learning process and outcomes will provide feedback for the lecturer on the improvement of the learning system. Therefore, analysis of student comments is very important to be analysed in order to increase the effectiveness of teaching lecturers.

“Comments” according to the Indonesian dictionary are responses, criticisms, suggestions which are given by someone for another related to a specific purpose. Comments can be facts or opinions about a problem. Facts are objective expressions of entity, while opinions are subjective expressions that describe people's sentiments, judgments, or feelings towards the entity. To analyse student’s comments, the lecturer can read the entire commentary one by one from each student, but there are several obstacles. The first obstacle is that reading a lot of comments will take time, so for lecturers who have little time to reflect, reading tens of hundreds of comments for one course is quite time-consuming work. The second obstacle is the presence of comments that may not be relevant, which will cause the lecturer to read them in vain. Therefore, we need a solution in this problem, the solution that the researcher offers is to summarize student comments and look for the most relevant comments and describe the comments as a whole.

1.1. Previous Researches

In this study, the main focus is opinion summarization which will be applied in summarizing student comments on the implementation of course learning. Opinion summarization is a technique in opinion mining to summarize opinions. Several previous studies related to the title of this study. [1] This study explores the features of a person's opinion, captures whether the opinion includes positive or negative opinions and then summarizes the results. In this article, it does not explain what it means to summarize the results, whether to summarize the opinion or just to list positive and negative opinions. Based on the results of testing and analysis conducted in this study, it can be concluded that the accuracy obtained is 93.44% with a precision of 97.66% and a recall of 92.88%. The next research is [2] where this study uses a single document. TextRank is a graphical ranking algorithm for processing text. TextRank generates sentence extractions as summaries. One of the advantages of this algorithm is that there is no need for training using training data on the algorithm used. The formulation is carried out at the following stages: pre-processing, calculating the value of overlapping content similarities, calculating the Text Rank value for each sentence, and making a graphic. The result is an informative summary text. Automatic Summarization was tested by Q & A Evaluation which was given to several respondents. Tests show that this algorithm is able to provide summaries with informative content up to 82.48% for summary text 50% and informative content 93.76% for summary text summarized 75%. In this research, what is summarized is a news document, not an opinion or comment, which is quite different because of the sentiment in it. The next research was [3]. In this research, Text Rank is used to summarize multi documents. From the results of the tests performed, the results obtained accuracy of 74%, recall of 73%, precision of 74%, and f-measure of 73%. In this research, what is summarized is news, not opinions or comments. The next research was [4]. This research was conducted to analyse opinions on products using the random forest classification algorithm. There are two stages before entering summary generation, first is product feature extraction which is done using the association mining method to get frequent itemset with two-words selection schemes, namely noun filtering and noun phrase filtering. The second stage is the classification process of extracted product features for positive and negative orientations using a supervised learning approach with a random forest algorithm. The results are not optimal, especially in feature extraction using association mining.

Based on the related studies above, the researcher wishes to develop “Opinion Summarization for Lecturer Learning Evaluation with Text Rank Algorithm”. The contribution that can be made is a combination of sentiment analysis and summarizing opinions / comments. This is done because comments have a different feature than documents or news, namely there are sentiments in them. Therefore, what will be done is to conduct a sentiment analysis to determine positive and negative opinions, then summarize the opinions using the text rank algorithm. In this paper, it will describe part of the research, which is comment/opinion summarization. With the implementation of this opinion summarization, it is hoped that it will be able to accelerate the lecturers in reading comments / criticisms and suggestions from students regarding the learning process of the courses taught by the lecturer.
2. Methodology
This research was part of sentiment analysis research for lectures’ learning evaluation. The stages of the implementation of the development of Indonesian Sentiment Summarization for Lecturer Learning Evaluation by Using TextRank Algorithm can be visualized by flowchart in the figure 1.

![Flowchart of Indonesian sentiment summarization by using TextRank algorithm.](image)

From the flowchart of Indonesian sentiment summarization by using TextRank algorithm that can be seen in figure 1, it can be explained as follows:

1. Sentiment data were retrieved from SIAK (Academic System) by using web services. The parameter requests were identification number, majors, academic years, and the sentiments.

2. Do the pre-processing language:
   a. Clean Texts, this was done because based on the results of sentiment data taken from SIAK, there were several HTML and PHP tags so they need to be cleaned. In addition, this step was taken to clean emojis, because there are many comments that are not in accordance with EYD, one of which was by filling in emojis in comments.
   b. Tokenization, this process was carried out to divide text which can be comments from each user into sentences.
   c. Case Folding, this process was carried out to convert the entire text in the document into a standard form where all words in the comment data which were made into a lower case.
   d. Cleansing & Filtering, this process removes stopwords and corrects some shortened words which were not in the format of EYD.
   e. Stemming, this process was carried out to convert affixed words into basic words in Indonesian using the Sastrawi algorithm.

3. Perform the summarization process using the Textrank algorithm.
   a. Creating a similarity matrix.
   b. Perform calculations to determine ranking
   c. Normalize scores to create scores
d. Sort by score

e. Determine the number of sentences you want, then display the sentences.

4. Evaluate the result of summarization by expert. We asked lecturer in Indonesian Language to determine whether the result of the summarization had been suitable or not.

2.1. Opinion Summarization
Sentiment Analysis is a broad area that includes opinion mining, sentiment classification and opinion summarization. Opinion Summarization is the process of automatically summarizing many opinions related to the same topic [5]. In sentiment analysis, Opinion Summarization involves many pre-processing steps such as tokenization, part of speech, stemming; makes it different from traditional summaries. This is one of NLP's most valued and powerful technologies. In social media, it is all about finding the most relevant posts with opinions for a given topic. This will allow understanding of hidden events and sentiment on various incidents [6]. Sentiment summarization is also different from summarization of factual data, because sentences that are seen as instructive from a factual point of view may not contain any sentiment at all, making them useless from a sentiment perspective [7].

Opinion Summarization can be easily integrated into real-life applications, which saves users time and effort. For example, via Twitter opinion, politicians can review their public image and companies can check their customer feedback. It also plays an important role in Social Media semantic analysis and Social Media Analysis.

There are two main approaches to generating textual summaries: 1) Extractive Summarization where the summary consists entirely of content extracted from the input and 2) Abstractive Summarization where the summary contains some content that is not in the source, for example like a paraphrased material [8]. In this research, it uses extractive summarization with textrank algorithm combined with cosine similarity algorithm.

2.2. TextRank Algorithm
Graph-based ranking algorithms are essentially a way of deciding the importance vertex within a graph, based on information recursively drawn from graph. The basic idea of implementation is “recommendation” or “voting”. If there is link from one vertex to another, it means that it is casting vote for other vertex. Higher number of votes means higher importance of vertex. The score is determined based on the votes that are cast for it [9].

Formally, let G = (V,E) be a directed graph with the set vertices V and set of edges E, where E is the subset of V x V. For a given vertex Vi, let In(Vi) be the set of vertices that point to it (predecessors), and let Out(Vi) be the set of vertices that vertex Vi points to (successors). The score of a vertex Vi is defined as follows [9]:

\[ S(V_i) = (1 - d) + d * \sum_{j \in \text{In}(V_i)} \frac{1}{|\text{Out}(V_j)|} S(V_j) \] (1)

Where d is a damping factor that can be set between 0 and 1, which is usually set by 0.85, which has the role of integrating into the model the probability of moving from a given vertex to another vertex in graph [10].

In model graphs that proposed by [9] are built from natural language text, and may include multiple or partial links between units (vertices) that are extracted from text. The weight (wij) added to corresponding edge that connects the two vertices. So, the formula can be defined as follows:

\[ WS(V_i) = (1 - d) + d * \sum_{j \in \text{In}(V_i)} \frac{w_{ij}}{\sum_{k \in \text{Out}(V_j)} w_{jk}} WS(V_j) \] (2)

To enable the application of graph-based ranking algorithms in natural language texts, we have to build a graph that represents the text and interconnects words or other text entities with meaningful
relations, like words, collocations, sentences, etc. The application of graph-based ranking algorithms to natural language texts consists of the following steps:

1. Identify text units that best define the task and add them in vertices in graph.
2. Identify relations that connect such text units, and use that to draw edges between vertices in the graph.
3. Iterate the graph-based ranking algorithm until convergence
4. Sort vertices based on their final score. [9]

Furthermore, the natural processing tasks which involves the ranking of text units consist of two tasks. The first is keyword extractions tasks. This task consists of the selection of key phrases representative for a given text and sentence. The second is extraction tasks. This task consists of the identification of most important sentences in the text, which can be used to do extractive summaries [9].

TextRank does very simple things. It finds how similar each sentence is to all other sentences in the text. The most important sentence is the one that is most similar to all the others, with this in mind the similarity function should be oriented to the semantic of the sentence, cosine similarity based on a bag of words approach can work well for TextRank. [11]

To apply TextRank, first the graph associated with the text has to be built, where the graph vertices are representative for the units to be ranked. For sentence task extraction, the goal is to rank entire sentences, therefore, a vertex is each sentence in the text. Formally, given two sentences $S_i$ and $S_j$, with sentence being represented by the set of $N_i$ words that appear in the sentence: $S_i = \{w_1^i, w_2^i, w_3^i, ..., w_{N_i}^i\}$, the similarity of $S_i$ dan $S_j$ is defined as:

$$Similarity(S_i, S_j) = \frac{|\{w_k| w_k \in S_i \& w_k \in S_j\}|}{\log(|S_i|) + \log(|S_j|)}$$

After the ranking algorithm is run on the graph, sentences are sorted in reversed order of their score, and the top ranked sentences are selected for inclusion.

3. Results and Discussion

3.1. Result

The implementation of Indonesian sentiment summarization for lecturer learning evaluation was used TextRank Algorithm supported by python programming language for data processing and Laravel for the front end. Python script was wrapped in a Python-powered REST API with Flask which that can be called from the Laravel app. The web-based front-end interface was made simpler because it was integrated with the existing Undiksha Academic Information System. The GUI of this implementation can be seen in figure 2.
Figure 2. Interface of indonesian sentiment summarization for lecturer learning evaluation by using textrank algorithm.

The results of summarized sentiment in one semester per department were displayed in the system. The number of sentences displayed in the summary was adjusted to the wishes of the lecturer or system user.

In this system, we retrieved three comments data from a lecturer with positive sentiment that is shown in Table 1.

Table 1. Sentiment data.

| No. | Comments |
|-----|----------|
| 1.  | Bapak sudah sangat baik dalam pemberian materi dari mata kuliah PBO ini sehingga mahasiswa dapat cepat mengerti dari materi yang bapak berikan kepada kami. |
| 2.  | Kesam saya mengikuti pelajaran bapak saya mudah mengerti apa yang disampaikan. Bapak memberikan materi sebelum mencoba ke praktek yang sangat memudahkan untuk saya mengerti. Cara bapak mengajar sangat teliti. |
| 3.  | Bapak sudah sangat bagus dalam hal mengajar dan saran saya adalah bapak harus mempertahankannya |

First step is to clean the text and case folding.

Table 2. Cleaned and case folded data.

| No. | Comments |
|-----|----------|
| 1.  | bapak sudah sangat baik dalam pemberian materi dari mata kuliah PBO ini sehingga mahasiswa dapat cepat mengerti dari materi yang bapak berikan kepada kami. |
| 2.  | kesam saya mengikuti pelajaran bapak saya mudah mengerti apa yang disampaikan. bapak memberikan materi sebelum mencoba ke praktek yang sangat memudahkan untuk saya mengerti. cara bapak mengajar sangat teliti. |
| 3.  | bapak sudah sangat bagus dalam hal mengajar dan saran saya adalah bapak harus mempertahankannya |

The second step is tokenization. It will tokenize to sentences and then words.
Table 3. Tokenized data.

| No. | Comments                                                                 |
|-----|---------------------------------------------------------------------------|
| S1. | bapak, sudah, sangat, baik, dalam, pemberian, materi, dari, mata, kuliah, PBO, ini, sehingga, mahasiswa, dapat, cepat, mengerti, dari, materi, yang, bapak, berikan, kepada, kami |
| S2. | kesan, saya, mengikuti, pelajaran, bapak, saya, mudah, mengerti, apa, yang, disampaikan |
| S3. | bapak, memberikan, materi, sebelum, mencoba, ke, praktek, yang, sangat, memudahkan, untuk, saya, mengerti |
| S4. | cara, bapak, mengajar, sangat, teliti |
| S5. | bapak, sudah, sangat, bagus, dalam, hal, mengajar, dan, saran, saya, adalah, bapak, harus, mempertahankannya, |

The third step is filtering and stopword removing.

Table 4. Stopword removed data.

| No. | Komentar                                                                 |
|-----|---------------------------------------------------------------------------|
| S1. | pemberian, materi, mata, kuliah, pbo, mahasiswa, cepat, mengerti, materi |
| S2. | kesan, mengikuti, pelajaran, mudah, mengerti |
| S3. | materi, praktek, memudahkan, mengerti |
| S4. | mengajar, teliti |
| S5. | bagus, mengajar, saran, mempertahankannya, |

Then, create the similarity matrix from five sentences.

Table 5. Similarity matrix.

| Kalimat | S1  | S2  | S3  | S4  | S5  |
|---------|-----|-----|-----|-----|-----|
| S1      | 0   | 0.47140452 | 0.94280904 | 0.33333333 | 0.25819889 |
| S2      | 0.47140452 | 0 | 0.66666667 | 0.47140452 | 0.36514837 |
| S3      | 0.94280904 | 0.66666667 | 0 | 0.47140452 | 0.36514837 |
| S4      | 0.33333333 | 0.47140452 | 0.47140452 | 0 | 1.03279556 |
| S5      | 0.25819889 | 0.36514837 | 0.36514837 | 1.03279556 | 0 |

The last step is to calculate and rank the sentences with (2) formula.

Table 6. The result of ranking.

| Kalimat | S1  | S2  | S3  | S4  | S5  |
|---------|-----|-----|-----|-----|-----|
| Score   | 0.94992728 | 0.93313051 | 1.10644553 | 1.05496137 | 0.9555353 |

From table 6, it was found that the third sentence got the highest score and the fourth sentence got the second highest. If we take 2 sentences to be the summarized opinion, the sentence will be "Bapak memberikan materi sebelum mencoba ke praktek yang sangat memudahkan untuk saya mengerti. Cara bapak mengajar sangat teliti.”

To evaluate the results of the sentiment summarization process, we asked experts in the field of language, who was an Indonesian language lecturers to be able to check whether the results of the summarized sentiment were good or not. The role of the expert was to rank the sentences based on the meaning of sentences that describe opinions or opinions on lecturer learning. The results then were compared with the result of the algorithm. We provided 100 students sentiment data taken from comments ten lecturers where one hundred data are then broken down into 10. Ten sentiment data does not mean 10 sentences, because in one sentiment it could consist of more than one sentence. Each list of 10 data (S1-S2) is analysed to find three sentences that represent the ten data. The linguist was asked to score against 10 sentiment lists with a range of 1 to 5 (likert scores). Score 5 means that the ranking
done was correct according to her, that means the sentence chosen has represented all sentiments. Conversely, Score 1 revealed that the summary results do not represent all sentiments at all.

| Sentences | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 |
|-----------|----|----|----|----|----|----|----|----|----|-----|
| Score     | 4  | 4  | 5  | 4  | 3  | 4  | 5  | 4  | 4  | 4   |

From the results of the evaluation by linguists, it can be calculated that the accuracy of the summary results carried out by this system has reached 82%.

3.2. Discussion
Based on the implementation of Indonesian sentiment summarization by using TextRank Algorithm, it can be seen that TextRank succeed in identifying the most important sentences in the text based on information exclusively drawn from the text itself. This algorithm does not need training data from previous sentiments in order to learn to make good summary, because textrank is fully unsupervised learning which depend on the given text to derive an extractive summary. Because this algorithm performs a ranking, the results of the summary can be in the form of one sentence or even up to 100 sentences depending on the needs.

The biggest challenge of this summarization process is in the pre-processing of sentiments where the sentiments obtained were not always clean. It was found that there were many sentiments that the sentences do not match the spelling perfected in Indonesian. Some of the sentiments include emoticons that should be removed in the pre-processing steps. In addition, there were some sentiments that were only one or two words such as "already good", "good" which must be recognized as a sentence. However, based on the results of the evaluation, it can be seen that the summary results were good enough to represent all comments.

4. Conclusion
In this paper, we have developed the Indonesian sentiment summarization for lecturer learning evaluation by using TextRank algorithm. Based on the results of the evaluation from the expert, the textrank algorithm can be used to summarize the opinions good with 82% of accuracy. In analysing sentiment, the biggest challenge was in the process of pre-processing sentiment, especially for sentiment sentences that are not in accordance with EYD. For further research, the sentiment data to be processed should be in the form of sentences, not just one or two words. Therefore, it is necessary to carry out a post tagging process to determine at least a subject and predicate in each sentence. In further research it can also be analysed, which sentiment is directed towards. Because not all sentiments lead to lecturers, but some discuss other things such as laboratory facilities, practicum schedules, and others. This will maximize the overall evaluation of learning.

References
[1] Suardani L G P, Divayana Y and Saputra K O 2019 Analisis Komentar Hasil Belajar Siswa Menggunakan Opinion Summarization Majalah Ilmiah Teknologi Elektro 61-68
[2] Eris V C M and Pragantha J 2017 Penerapan Algoritma TextRank untuk Automatic Summarization pada Dokumen Berbahasa Indonesia Jurnal Ilmu Tekni dan Komputer 71-78
[3] Putra R R and Evita K D 2019 Peringkasan Teks Otomatis Pada Multi Dokumen Menggunakan Textrank Digital library - Perpustakaan Pusat Unikom
[4] Aprianto A, Maharani W and Herdiani A 2016 Analisis Sentimen dan Peringkasan Opini pada Ulasan Produk Menggunakan Algoritma Random Forest (Bandung)
[5] Bhattacharjee S and Ray A K 2015 Sentiment analysis: approaches, applications and challenges IJIACS Int J Innov Adv Comput Sci 516-527
[6] Hole V and Takalikar M 2014 A survey on sentiment analysis and summarization for prediction Int J Eng Comput Sci (IEJCS) 9503-9506

[7] Moussa M E, Mohamed E H and Haggag M H A 2018 survey on opinion summarization techniques for social media Future Computing and Informatics Journal 3(1) 82-109

[8] Farzindar A 2014 Social network integration in document summarization. Innovative document summarization techniques revolutionizing knowledge understanding IGI-Global 139-162

[9] Mihalcea R and Tarau P 2004 TextRank: Bringing Order into Text Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing

[10] Brin S and Page L 1998 The anatomy of a large-scale hyper-textual Web search engine Computer Network and ISDN System 1-7

[11] Lee C 2018 Use TextRank to Extract Most Important Sentences in Article [Online]. Available: https://medium.com/the-artificial-impostor/use-textrank-to-extract-most-important-sentences-in-article-b8efc7e70b4.

[12] Condori R E L and Pardo T A S 2017 Opinion Summarization Methods: Comparing and Extending Extractive and Abstractive Approaches Expert System with Applications

[13] Tanwi S G, Kumar V, Jain Y S and B A 2019 Automatic Text Summarization using Text Rank International Research Journal of Engineering and Technology (IRJET) 2655-2657

[14] Pressman R S and Maxim B R 2015 Software Engineering: A Practitioner's Approach In McGraw-Hill Education

[15] Wahid D and Azhari S N 2016 Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity IJCCS (Indonesian Journal of Computing and Cybernetics Systems) 207-218

[16] Adriani M, Asian J, Nazief B, Tahaghoghi S M M and Williams H E 2007 Stemming Indonesian ACM Transactions on Asian Language Information Processing 1-33

[17] Nugroho K S 2019 Dasar Text Processing dengan Python [Online]. Available: https://medium.com/@ksnugroho/dasar-text-preprocessing-dengan-python-a4fa52608ffe.