Automated Essay Scoring Using Natural Language Processing And Text Mining Method

1st Gunawansyah
Department of Informatics Engineering,
Faculty of Engineering
Universitas Sangga Buana
Bandung, Indonesia
gsyahbdg@gmail.com

2nd Riska Rahayu
Department of Informatics Engineering,
Faculty of Engineering
Universitas Sangga Buana
Bandung, Indonesia
rriska851@gmail.com

3rd Nurwathi
Department of Industrial Engineering,
Faculty of Engineering
Universitas Sangga Buana
Bandung, Indonesia
nurwathi68@yahoo.co.id

4th Bambang Sugiarto
Department of Informatics Engineering,
Faculty of Engineering
Universitas Sangga Buana
Bandung, Indonesia
bambang.sugiarto@usbypkp.ac.id

Computer vision Research Group,
Research Center of Informatics
Indonesian Institute of Sciences (LIPI)
Bandung, Indonesia
bamb044@lipi.go.id

5th Gunawan
Department of Informatics Engineering,
Faculty of Engineering
Universitas Sangga Buana
Bandung, Indonesia
gunawan@usbypkp.ac.id.

Abstract— **The use of technology really helps to maximized the effectiveness and efficiency of work expecially in the education field. Elearning is the concept of education that has begun to be widely implemented at this covid-19 pandemic to avoid the spread of transmission through social distancing. One of elearning types is essay but for large participants, it need much effort for evaluate by human rater. The inconsistency of assessment by the rater due to fatigue can also affect the quality of the assessment. Developing a system that can learn and understand on its own without having to be repeatedly programmed by humans used machine learning and computational linguistics to study the interaction between computers and human natural language used natural language processing proposed in this research. Natural language processing and text mining methods are able to provide a good assessment which is influenced by several processes, namely tokenization, stopword, stemming and support with the number of keywords, and the synonym of more complex keywords. The automated essay scoring system is proven to provide consistent and objective assessments and is able to approach human raters assessments.**

Keywords*: E-learning, Natural Language Processing, Text Mining, Automated Essay Scoring, Preprocessing.*

I. INTRODUCTION

The development of information and communication technology has a significant influence in various fields, especially education. The development of various technology based on educational services such as elearning, ebooks and so on, has begun to be used by many educational institutions in the learning process.

The use of technology in the world of education is very helpful, especially the effectiveness and efficiency of work can be maximized. One of the concepts that are widely used at this time is elearning. Elearning is an educational concept that implements information technology in the learning process and one of the processes in elearning is an online exam. Examination is carried out online, starting from answering exam questions and then giving test scores. This certainly provides an advantage for teachers and students because the implementation of essay exams is more effective and efficient. Another advantage is that the computer-assisted evaluation system can provide a faster and more accurate assessment than conventional scoring systems. With this system, the implementation of tests carried out with a large number of students can be done better and faster [1].

There are several online exam applications used in elearning, with the types of questions provided in the form of multiple choices including Multiple Choice Single Answer (MCSA) and Multiple Choice Multiple Answer (MCMA), true or false (True False) , match (Matching). The problem is that they still cannot provide a direct assessment of the answer questions for descriptions or essays. This causes the quality of the assessment decreases and sometimes the assessment is no longer objective. So that there is a subjective assessment that uses the teacher's reasoning and perspective only. In the end, the variation of assessment in the class is different, which is perceived by students as a source of injustice. So, it is necessary to create an automated essay scoring system that can provide objective, fast and accurate assessments. There are several previous studies in making automated essay scoring using the Latent Semantic Analysis method [1], then other studies using naïve bayes method and neuro fuzzy [2].

978-1-7281-7598-0/20/$31.00 ©2020 IEEE

Similarity level is a challenge in the automated essay scoring system. In this research, natural language processing and text mining methods are proposed to compare the level of similarity between this method and the assessment of human raters. Then an automated essay scoring application was created using natural language processing and web based on text mining methods in the online exam process to increase the effectiveness and efficiency of the essay exam to be more objective and accurate.

II. RESEARCH CONCEPT

Elearning is an educational system or concept that implements information technology in the process of teaching and learning activities. During the Covid-19 pandemic, this system is the main choice to avoid the spread of transmission through social distancing. One of the assessments carried out to evaluate the success of learning is with doing an essay assignment. For a large number of students, it takes a lot of time to assess the results of the evaluation. The inconsistency of assessment by the rater due to fatigue can also affect the quality of the assessment, therefore it is necessary to create a system that can provide a consistent assessment.

One approach is machine learning which is a part of artificial intelligence (AI) application. AI focuses on developing a system that can learn and understand on its own without having to be repeatedly programmed by humans. On the other hand, natural language processing (NLP), which is a field of AI too, especially computational linguistics to study the interaction between computers and human natural language is also very important. Natural language processing seeks to solve the problem of understanding human natural language with all its grammatical and semantic rules and transforming the language into a formal representation that can be processed and understood by computers [3].

Synonym is one of the challenge in the assessment process in order to provide more objective results and match the assessment by human rates. Therefore, the process of pattern extraction (information and useful knowledge) from a large number of unstructured data sources such as word documents, PDFs, text quotes etc. in Text Mining is very necessary [3]

A. Essay exam

In general, essay tests or descriptions are questions that require students to answer in the form of describing, explaining, discussing, comparing, giving reasons, and other similar forms in accordance with the demands of the question using their own words and language. So the test requires the ability of students to generalize their ideas through written language.

Before processing sentences into preprocessing text, sentences that will be processed are taken from the database, the goal is to take similarity as a point to get the score for each keyword. After the data is collected, preprocessing begins with the tokenization process, then continues with the stopwords step from tokenized results and stemming step to produce the root word as a reference for the keyword. The preprocessing text process will produce an array of keywords that have been processed in an intersect array, where each keyword will be classified according to the similarity of the word. After the word array is generated, question calculation is carried out and each calculation result is combined for all the questions so that the final calculation is produced to display the overall student score.

The automated essay scoring process architecture in this research can be seen in the picture below.

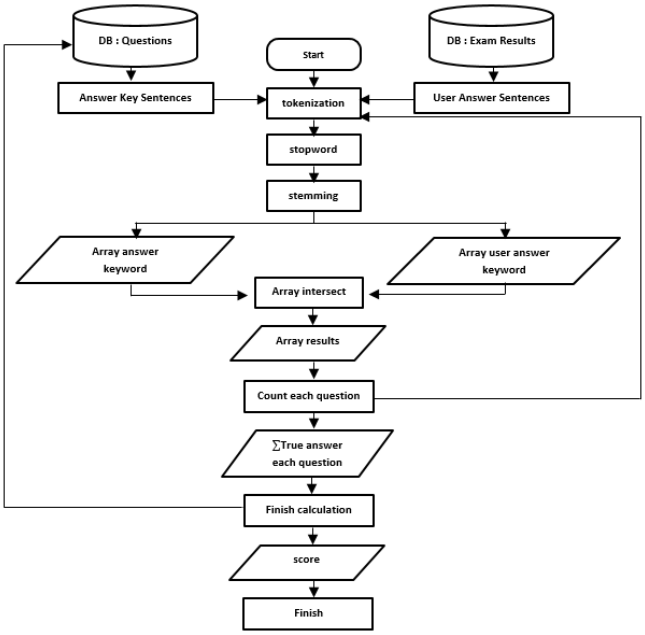

Fig 1. Automated Essay Scoring architecture

B. Preprocessing Text

Preprocessing text is a stage from the initial process of the text to prepare the text into data that will be processed further. A text cannot be processed directly by the search algorithm, therefore text preprocessing is needed to convert text into numeric data [4]. At this stage there are 3 processes including [5]:

1. Tokenization

At this stage, the input string is cut based on each word composing it. Tokenization is the process of breaking a set of characters or sentences in a text into word units, how to distinguish certain characters that can be treated as word separators or not.

In this research, the process of tokenizing question sentences and answers is taken from the database, then processed using an explode code where sentence strings are broken down into word by word units, then decomposed into a string array.

2. Stopword

The stopword process is defined as a collection of words that are not related (irrelevant) to the main subject in question, even though these words often appear in the data being used.

The words in question are usually types of conjunctions, affixes, and so on. The stopword process of any irrelevant or unrelated words is removed. The stopword data provided in this system database is 1075 data.

3. Stemming

The stemming process is carried out after getting the results of the stopword process to remove affixes and produce the root word of a word. In the process, there are several rules for removing all affixes, both consisting of prefixes, infixes, suffixes and confixes (combinations of prefixes and suffixes) in derivative words.

III. IMPLEMENTATION

The initial stage of implementation is when the teacher create questions for the exam by entering questions and answer keys. When the teacher has finished creating the answer key, a process is carried out to display the keyword automatically.

When the program processes to generate keywords, the program will display several synonyms of the keyword. In this section the teacher can change synonym or add polysemy so that the word equation library data from the system will be increasingly complex. This will make the system smarter and broader in understanding human natural language.

Automated essay scoring in this research can be seen in the following figure:

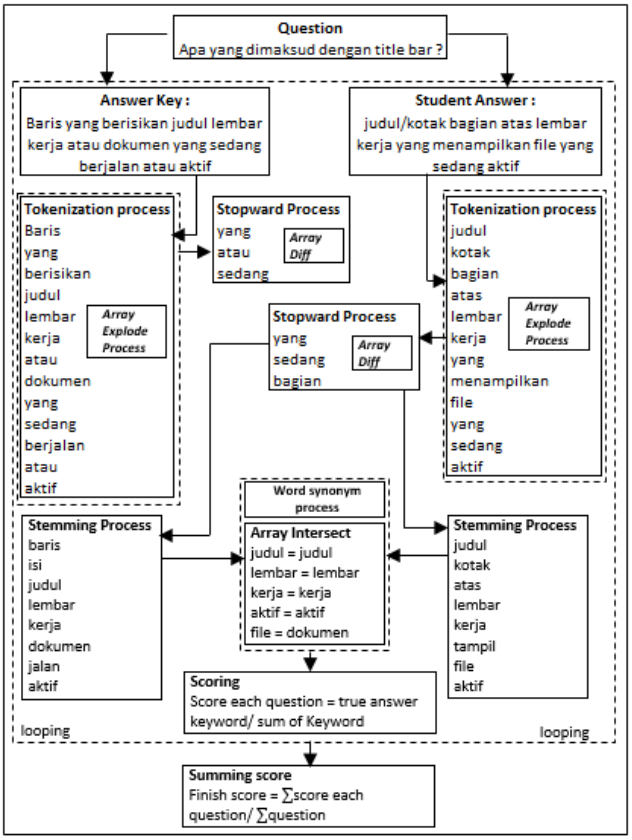

Fig.2 Automated Essay Scoring process

In preprocessing text, the data is taken from the data table of the teacher answer keys and the test results from each student. The process is carried out by breaking sentences with the tokenization method into a word list, then removing unnecessary words with the stopword method and the stemming process to produce basic words. After preprocessing text, the process of combining keywords and calculating student grades. In the combining keywords process, the synonym of keywords can be reach in synonym library from teacher input and previous answer from another question.

Online exams can be carried out by students after logging into their respective accounts and the exam time has started. Students enter the token that has been given by the teacher and start answering questions. After the exam time is over or the answers are sent by students, the answers will be stored and processed automatically by the system. The flow of online examinations by students can be seen in the following figure.

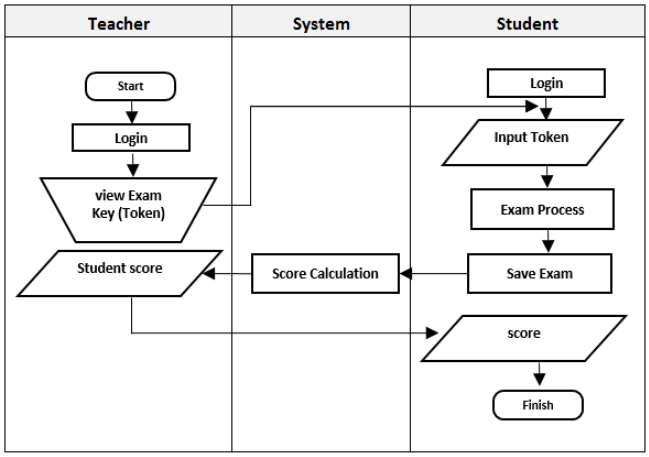

Fig 3. Online Exam process

Students perform automated essay scoring in realtime, after completion of the essay, test results will be calculated automatically by the system and display notification of student grades on the teacher's page.

In this research, three assessment scenarios were carried out to compare the final results of the assessment, namely manual assessment (human raters), the automated essay scoring system without adding word synonym (system 1) and scoring by adding word synonym (system 2). Synonym from a keyword can be added when creating new question or can be obtained from previous question that use the same keyword. So that if a word has many synonym, it will complement the word even more. The more often a keyword is used in various questions and completed with all synonyms, the application's capabilities will be increased and have competence like a human rater. All the keyword synonym will be stored in a synonym table.

The results of each assessment scenario in this research can be seen in the table below.

Table 1 Result comparison

No	Student name	Human Rate	System 1 Score	Sistem II Score
1	M.Fakhruzzaky A.	40	20.00	37.5
2	M Azis Maulana B	40	19.00	37.5
3	Kayla Ramadhani	50	24.00	47.22
4	Nazwa Aulia Khaerunisa	30	9.00	44.44
5	Sawerigading Anakahdin	50	24.00	47.22
6	Maharani Kirana Putri	65	29.00	51.38
7	Reza Putra Surya	50	19.00	48.61
8	M.Rahman Hidayat	60	24.00	51.38
9	Watik Ayu Ningtyas	45	29.00	43.05
10	Nisa Mutia Rahayu	50	29.00	43.05
11	Keisha Nathania	60	28.00	51.38
12	Anabela Nasya Syakirah	57	29.00	55.55
13	Neiva Nabila Waskita	50	34.00	47.22
14	Ataya Dyah Mega Saphira	50	34.00	47.05
15	Arimbi Nabila Putri Perdana Sabar	55	29.00	51.38
16	Falya Siti Fauzia	45	23.00	43.05
17	Shenya Mulyani S	65	44.00	51.38
18	Alyssa Shafira Dyannisa	54	23.00	43.05
19	Muhammad Rasya Azka Firdaus	35	23.00	47.22
20	Neiva Nabila	45	34.00	43.05
21	Ramadhan Fariz F	50	15.00	47.22
22	Azka Nasrullah	50	30.00	47.22
23	Nauval Zalendra Putra.R	50	23.00	47.22

Based on the comparison of the human raters' assessment results in a value of 30-65, the assessment system I produces a value of 9 – 44 and then the results of the assessment system II are 37.5 - 55.55. This proves that the keywords and the synonym of the keywords have a big influence in the automatic essay assessment when compared between teacher scores and system II scores which have a high level of similarity.

The results of the calculations above are influenced by the number of keywords and the word synonym of the keywords. From the test, the automated essay scoring system proves that the resulting value is objective and consistent, while in human raters, when re-evaluating, there is a difference in value with the values previously carried out.

IV.CONCLUSIONS & RECOMMENDATIONS

A. Conclusions

Based on this research, the natural language processing and text mining methods are able to provide a good assessment which is influenced by several processes, namely: tokenization, stopword, stemming, the number of keywords, and the synonym of more complex keywords. The automated essay scoring system is proven to provide consistent and objective assessments and it able to approach human raters' assessments. It can resulted consistent score for some scoring process in the different time, better than human rater.

B. Recommendations

More complex synonym library, automatic spelling and grammar correction for student when answer the question are challances to completed in this research because keyword is a important think in this reseacrh.

V. REFERENCES

[1] A. Agung, P. Ratna, B. Budiardjo, and D. Hartanto, “Simple : Sistem Penilai Esei Otomatis Untuk Menilai Ujian Dalam Bahasa Indonesia,” Makara,Teknologi, vol. 11, no. 1, pp. 5–11, 2007.

[2] M. Habibi, Y. F. A, and E. Rachmawati, “Automated Essay Scoring Menggunakan Naive Bayes dan Neuro Fuzzy,” pp. 1–6, 2013.

[5] “Teori Natural Language Processing,” pp. 1–14, 2012.

[4] “Preprocessing Text, ” 2017. [Online].Available:https://informatikalogi.com/text-preprocessing/. [Accessed: 18-Mar-2019].

[5] R. J. Mooney., Machine Learning Text Categorization. Austin: University of Texas at Austin, 2006.