Scientific Journal of Informatics
Vol. 10, No. 2, May 2023
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Automatic Scoring Using Term Frequency Inverse Document Frequency Document Frequency and Cosine Similarity

Winda Yulita[1], Meida Cahyo Untoro[2*], Mugi Praseptiawan[3], Ilham Firman Ashari[4], Aidil Afriansyah[5], Ahmad Naim Bin Che Pee[6]
[1,2,3,4,5]Department of Informatics Engineering, Institut Teknologi Sumatera, Indonesia
[6]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia

Abstract.
Purpose: In the learning process, most of the tests to assess learning achievement have been carried out by providing questions in the form of short answers or essay questions. The variety of answers given by students makes a teacher have to focus on reading them. This scoring process is difficult to guarantee quality if done manually. In addition, each class is taught by a different teacher, which can lead to unequal grades obtained by students due to the influence of differences in teacher experience. Therefore the purpose of this study is to develop an assessment of the answers. Automated short answer scoring is designed to automatically grade and evaluate students' answers based on a series of trained answer documents.
Methods: This is 'how' you did it. Let readers know exactly what you did to reach your results. For example, did you undertake interviews? Did you carry out an experiment in the lab? What tools, methods, protocols or datasets did you use The method used is TF-IDF-DF and Similarity and scoring computation. The word weight used is the term Frequency-Inverse Documents Frequency -Document Frequency (TF-IDF-DF) method. The data used is 5 questions with each question answered by 30 students, while the students' answers are assessed by teachers/experts to determine the real score. The study was evaluated by Mean Absolute Error (MAE).
Result: The evaluation results obtained Mean Absolute Error (MAE) with a resulting value of 0.123.
Novelty: The word weighting method used is the Term Frequency Inverse Document Frequency DocumentFrequency (TF-IDF-DF) which is an improvement over the Term Frequency Inverse Document Frequency (TF-IDF) method. This method is a method of weighting words that will be applied before calculating the similarity of sentences between teachers and students.

Keywords: TF-IDF-DF, Cosine similarity, Automatic answer scoring system
Received January 2023 / **Revised** February 2023 / **Accepted** March 2023

INTRODUCTION

In the learning process, most of the tests to assess learning achievement have been carried out by giving questions in the form of short answers or question essays. The variety of answers given by students makes a teacher focus on reading them. This scoring process is difficult to guarantee quality if it is done manually. Moreover, each class is taught by a different teacher can cause inequality in the value obtained by students because of the influence of differences in teacher experience [1], [2]. Therefore, an automated answer assessment research was developed.

Automatic short answer scoring is an important branch of education intelligent (intelligent education). Automatic short answer assessment is designed to assess and evaluate student answers automatically based on a collection of trained answer documents [3], [4]. This system can help teachers check student understanding and reduce the influence of teachers in giving subjective values. At the same time, automatic appraisal can reduce labor costs and material resources. Therefore, the automated short answer assessment task has significant convenience and great commercial value [5], [6].

Automated assessment has been widely researched and developed in various languages, such as Arabic, English and Arabic English, Indonesian language and others. Automatic answer scoring in Arabic was

*Corresponding author.
Email addresses: winda.yulita@if.itera.ac.id (Yulita), cahyo.untoro@if.itera.ac.id (Untoro), mugi.praseptiawan@if.itera.ac.id (Praseptiawan), firman.ashari@if.itera.ac.id (Ashari), aidil.afriansyah@if.itera.ac.id (Afriansyah), naim@utem.edu.my (Pee)
DOI: 10.15294/sji.v10i2.42209

performed by Abdeljaber [7] with use Longest Common Subsequence (LCS) algorithm and used the WordNet language Arabic. In English developed by Chang with use machine-generated concept map method [8]. Development of automatic short answer scoring in Indonesian language [9] was done with using similarity measurement (cosine similarity) to detect the similarity of students' answers with the teacher's answers. The study did not use the word weight measurement stage before measuring the similarity of answers.

One of the methods commonly used to measure the weight of a word is the Term Frequency-Inverse Document Frequency (TF-IDF). The TF-IDF algorithm is a statistical measure used to evaluate how important a word is in one word sentence on a text/document [10]. The TF-IDF method is a commonly known technique and is used as a weighting technique and its performance is even comparable to new techniques [11]. The TF-IDF method is widely applied in various fields including information retrieval [12]–[15]. The TF-IDF method is applied to information retrieval research to find news texts that match the given keywords. This study succeeded in displaying articles that were relevant or appropriate to the system created [16], [17].

Although the TF-IDF is familiar method but in other research found that TF-IDF method has deficiency in word weighting. The downside is that there is exists presumption that said a lot spread in other documents do not important, so considered no there is. Even though it's a frequent word that appears in another sentence can so is an important word. Consequently, high weight value obtained on the word that has frequency tall in document, while words that are scattered in other documents have calculation small weight. Because of that, this TF-IDF method developed more carry on for get weight representative of the extracted words with consider spread of words in other documents [18], [19].

Automatic short answer assessment is an important branch of education intelligent (intelligent education). Automatic short answer assessment is designed to automatically assess and evaluate student answers based on a set of trained answer documents [3], [20]. This system can help teachers check students' understanding of learning and reduce the influence of teachers in giving subjective scores. Figure 1 plots research conducted with gradually that is studies literature, design system, test system and analysis and discussion Literature review containing explanation related to text preprocessing data, Term Frequency- Inverse Document Frequency-Document Frequency (TF-IDF-DF), cosine similarity such as under this:

Text Preprocessing

Text preprocessing is step for prepare data for processed on processing next. A collection of sentences contained in the document will be prepared and processed so that be formed word group [21]. The process consists from segmentation of sentences, case folding, tokenizing, filtering with make stopwords, stemming

1. sentence segmentation
   Sentence segmentation is the first step of the text preprocessing process. In this process, the news text consists of paragraphs that are broken down into several sentences. Separation of each sentence based on punctuation marks, such as periods (.), exclamation points (!) and question marks (?) [22].
2. case folding
   News paragraphs that have been cut into sentences will carry out the case folding process. Case folding is the process of converting all text to lowercase characters and discarding all characters other than az. If there are punctuation marks, numeric numbers and symbols are omitted [23].
3. tokenizing
   Tokenizing is a process to change the form of sentences into single words. Cutting sentence based on the delimiter that composes it, namely a space (" ") [24].
4. filtering
   In the filtering process, stopword disposal. Stopwords are words that have no meaning or words that are less meaningful and often appear in a collection of words [25].
5. Stemming
   The next process is stemming, namely: return a word to a form the root (root word) with certain rules, so that every word has the same representation [26]. Stemming in study this use algorithm Nazief & Adriani [18].

Basic Concepts of Term Frequency-Inverse Document Frequency-Document Frequency (TF-IDF-DF)

Frequency (TF-IDF-DF) is modified from TF-IDF method because TF-IDF method has deficiency in word weighting. The downside is that exists presumption that word spread in other documents is not important, so considered no there is. Even though it's a frequent word appearing in another sentence, so is an important word. Consequently, the value of high weight obtained on the word that has frequency tall in document, while words that are scattered in other documents have calculation small weight. Because of that, this TF-IDF method developed more carry on for getting weight representative of the extracted words with consider spread of words in the document another. Document Frequency (DF) which contains the i word effect on the topic whole document, so score in word weighting will multiplied with the i word DF [14]. Formula word weighting of TF-IDF modifications are:

$$wi,j = (tfi,j \; x \; log \; (\frac{N}{\text{dfi}})) \; x \; dfi \; x \; dfi \quad (1)$$

Where is wi,j is weight from the i word from the document j. tfi,j (term frequency) is number of words from word i in the document j. $log \; (\frac{N}{\text{dfi}})$ is Inverse Document Frequency (IDF) formula, N is amount whole document or sentence. dfi is Document Frequency is many sentences containing the word i in gathering document.

Measure the similarity between sentences with cosine similarity

Relevance/similarity of teacher answers with student answers. The value of relevance is obtained by measuring the similarity between 2 vectors, namely the teacher's answer vector and the student's answer vector. The greater the value of relevance, the teacher's answers and answers the student will more relevant/similar.

According to Patidar et al. [27], size similarity (similarity measure) is the distance between various data points. Similarity measure is also used to measure the similarity between sets based on the intersection of two sets. Similarity measure is also known as a function that calculates the degree of similarity between a pair of text objects. In short, similarity reflects amount of strong connection between two data [28].

One of size most common similarities used is cosine similarity. Cosine similarity is basis of calculation to get the value relevance between queries with documents and the relevance between documents. Cosine similarity is the cosine of the angle θ between vectors. Cosine similarity has the formula as below:

$$\boldsymbol{Sim \; (S_1, S_2) = \frac{\sum_i t_{1i} t_{2i}}{\sqrt{\sum_i t_{1i}^2} \; x \; \sqrt{\sum_i t_{2i}^2}}} \quad (2)$$

Information:
$\boldsymbol{S_1}$ = candidate word weight vector
$\boldsymbol{S_2}$ = word weight vector other than candidate
Where $\boldsymbol{t_i}$is the word weight of the word $\boldsymbol{w_i}$.

METHODS

Study this using several steps, starting from preprocessing, calculating TF-IDF value and calculating score similarity among answer students with expert answers.

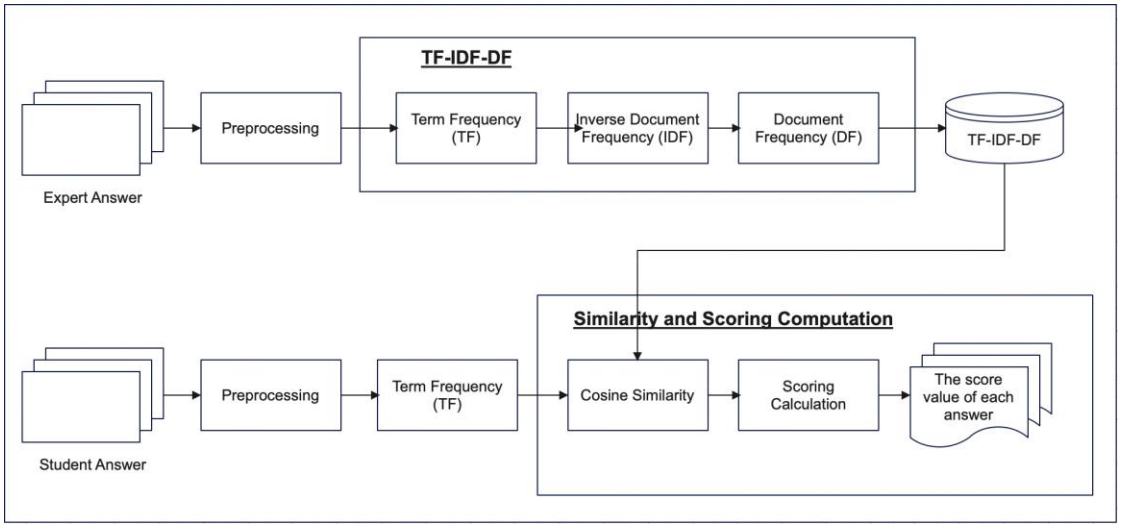

Figure 1. Research diagram

Based on Figure 1, it is explained that the expert's answer will be preprocessed, calculating the value of term frequency (TF), calculating the value of inverse document frequency (IDF) and calculating the value of document frequency so that the value of TF-IDF-DF is obtained. The TF-IDF-DF value will be used in calculating the similarity of expert answers with student answers. Students' answers still go through the preprocessing stage and calculate the value of the term frequency (TF). After calculating the cosine similarity value is found, the value of the system will be changed by multiplying the actual value of the answer. Furthermore, the value obtained by students will be stored.

Preprocessing

Destination from text preprocessing stage that is change article news be ready words processed for calculation word weight. Some of the processes of text preprocessing, namely segmentation sentences, case folding, tokenizing, filtering, and stemming. Following this is one _ example entered document _ in the process of text preprocessing accompanied by stages of the text preprocessing process:

Question: What is Kevin Spacey's attitude in civil court?

Answer: Kevin Spacey testifies in his defense in civil trial. Kevin Spacey has taken the stand as the first witness in his own defense in the sexual misconduct trial against him, brought by actor

Segmentation sentence

Is solving paragraph become sentence. Solving was conducted with separate based on sign read dot (.), sign question (?) and sign exclaimed (!). The result of the segmentation process sentence is shown in Table 1.

Table 1. Sentence segmentation results

No	Sentence
0	Kevin Spacey testifies in his defense in civil trial
1	Kevin Spacey has taken the stand as the first witness in his own defense in the sexual misconduct trial against him, brought by actor Anthony Rapp.
2	In a response to the first question from his attorney, Chase Scolnick, Spacey said Rapp's allegations are not true.

Case folding

Paragraph news that has been cut Becomes sentence will run the case folding process. Case folding is change all text becomes character with letter small and throw away all character other than az. Besides that, sign read, number numerics and symbols are also omitted. Table 2 is the results of the case folding process.

Table 2. Case folding results

No	Sentence
0	kevin spacey testifies in his defense in civil trial
1	kevin spacey has taken the stand as the first witness in his own defense in the sexual misconduct trial against him brought by actor anthony rapp
2	in a response to the first question from his attorney chase scolnick spacey said rapps allegations are not true

Word Tokenizing

It is a cutting process of sentences into words. Cutting sentence based on the delimiter that composes it, namely space (" ").

Table 3. Case folding results

No	Word	No	Word
1	kevin	28	trial
2	spacey	29	against
3	testifies	30	him
4	in	31	brought
5	his	32	by
6	defense	33	actor
7	in	34	anthony
8	civil	35	rapp
9	trial	36	in
10	kevin	37	a
11	spacey	38	response
12	has	39	to
13	taken	40	the
14	the	41	first
15	stand	42	question
16	as	43	from
17	the	44	his
18	first	45	attorney
19	witness	46	chase
20	in	47	scolnick
21	his	48	spacey
22	own	49	said
23	defense	50	rapps
24	in	51	allegations
25	the	52	are
26	sexual	53	not
27	misconduct	54	true

Filtering

In this filtering stage to disposal stopword. Stopword are words that are not have meaning or less words mean and often appear in word group. How to get rid of words that are not important with check the dictionary stopword. If the words are the same with stopword then the word will thrown away or deleted.

Table 4. Filtering results

No	Word	No	Word
1	kevin	18	against
2	spacey	19	brought
3	testifies	20	actor
4	defense	21	anthony
5	civil	22	rapp
6	trial	23	response
7	kevin	24	first
8	spacey	25	question
9	taken	26	attorney
10	stand	27	chase
11	first	28	scolnick
12	witness	29	spacey
13	own	30	said
14	defense	31	rapps
15	sexual	39	allegations
16	misconduct	40	true
17	trial		

Stemming

Stemming, that is return a word to form the root (root word), so that every word has the same representation. In method, this only handles affix (affix) prefix (prefix) and suffix (suffix) only. This thing caused by infrequent occur case addition affix infix (insert) in Indonesian. The results of stemming are shown in Table 5.

Table 5. Stemming results

No	Word	No	Word
1	kevin	18	against
2	spacey	19	brought
3	testifi	20	actor
4	defens	21	anthoni
5	civil	22	rapp
6	trial	23	respons
7	kevin	24	first
8	spacey	25	question
9	taken	26	attorney
10	stand	27	chase
11	first	28	scolnick
12	wit	29	spacey
13	own	30	said
14	defens	31	rapp
15	sexual	39	alleg
16	misconduct	40	true
17	trial		

TF-IDF-DF
After the text preprocessing process, the next stage is counting word weight with the TF-IDF algorithm using generated words from the stemming process.

Cosine Similarity
If the word weight has obtained, next look for the cosine similarity value generated by the system. Table 6 represents score similarity Among teacher and student answers generated by the system on questions 1 to with question 5 (Q1-Q5).

Table 6. The similarity value of system

No	Q1	Q2	Q3	Q4	Q5
1	0,67	0,56	0,74	0,68	0,56
2	0,73	0,78	0,75	0,58	0,67
3	0,34	0,65	0,93	0,56	0,54
4	0,52	0,77	0,71	0,85	0,76
5	0,36	0,68	0,38	0,81	0,98
6	0,56	0,74	0,58	0,59	0,45
7	0,78	0,56	0,59	0,63	0,56
8	0,89	0,56	0,93	0,77	0,56
9	0,45	0,78	0,84	0,85	0,39
10	0,52	0,94	0,58	0,74	0,78
11	0,53	0,85	0,83	0,83	0,56
12	0,69	0,86	0,85	0,59	0,78
13	0,88	0,88	0,83	0,66	0,56
14	0,83	0,93	0,77	0,59	0,87
15	0,75	0,75	0,71	0,84	0,67
16	0,56	0,78	0,59	0,72	0,53
17	0,72	0,77	0,63	0,82	0,69
18	0,93	0,73	0,68	0,56	0,78
19	0,47	0,76	0,92	0,56	0,89
20	0,23	0,64	0,47	0,67	0,63
21	0,56	0,73	0,57	0,79	0,78
22	0,45	0,82	0,65	0,58	0,73
23	0,69	0,57	0,56	0,37	0,56
24	0,83	0,59	0,47	0,58	0,78
25	0,87	0,73	0,78	0,67	0,67
26	0,79	0,67	0,59	0,68	0,67
27	0,84	0,94	0,61	0,49	0,52
28	0,59	0,45	0,77	0,79	0,65
29	0,79	0,83	0,74	0,56	0,43
30	0,88	0,82	0,69	0,73	0,89

The similarity value generated by the system ranges from between 0 to 1. The similarity value is close to number 1 means similarity answer among students and teachers are increasingly similar, vice versa score close to 0 represents answer students and teachers are increasingly no similar [28], [29]. The value obtained by the system is multiplied by with original score for every number, i.e. 20 points. So that the value

generated by the system is between 1 and with 20 for every the number. Amount student as many as 30 students.

Table 7. Conversion result in student 1

Document	Similarity value system	Expert value
ßQ1	0,67 x 20 = 13,4	15
Q2	0,56 x 20 = 11,2	15
Q3	0,74 x 20 = 14,8	15
Q4	0,68 x 20 = 13,6	15
D5	0,56 x 20 = 11,2	10

Testing

Testing is done by looking at the student scores generated by the system with the values given by the teacher manually. At this stage, an assessment of the accuracy of the value generated by the system will be carried out. To evaluate the performance of the system that has been created, the Mean Absolute Error (MAE) calculation between the results of the system assessment and the results of the manual assessment will be used [30]. The equation used to calculate the correlation value is as follows:

$$MAE = \frac{1}{n}\sum_{i-1}^{n} |x_i - j_i| \quad (3)$$

RESULT AND DISCUSSION

Data

Data used is questions given by the history teacher with amount question that is 5 questions in short essay. Engaged students for answer question is students of SMK class XI, totaling 30 students because that, every question will answered by 30 students.

Research Results

Figure 2 to with Figure 5 shows comparison score the similarity produced by the system with value obtained from expert to every questions (Q1-Q5).

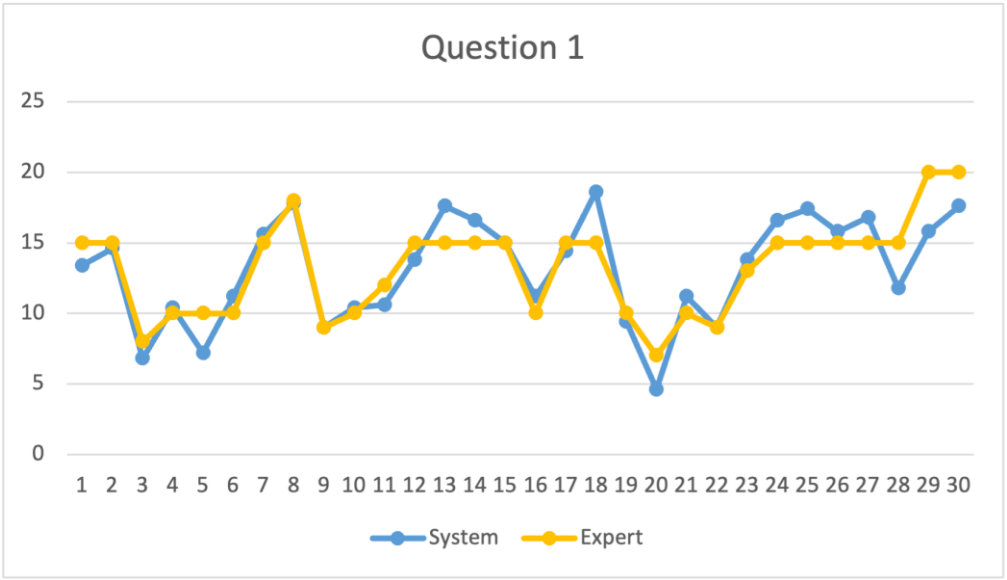

Figure 2. Comparison the value generated by the system and the expert in Question 1

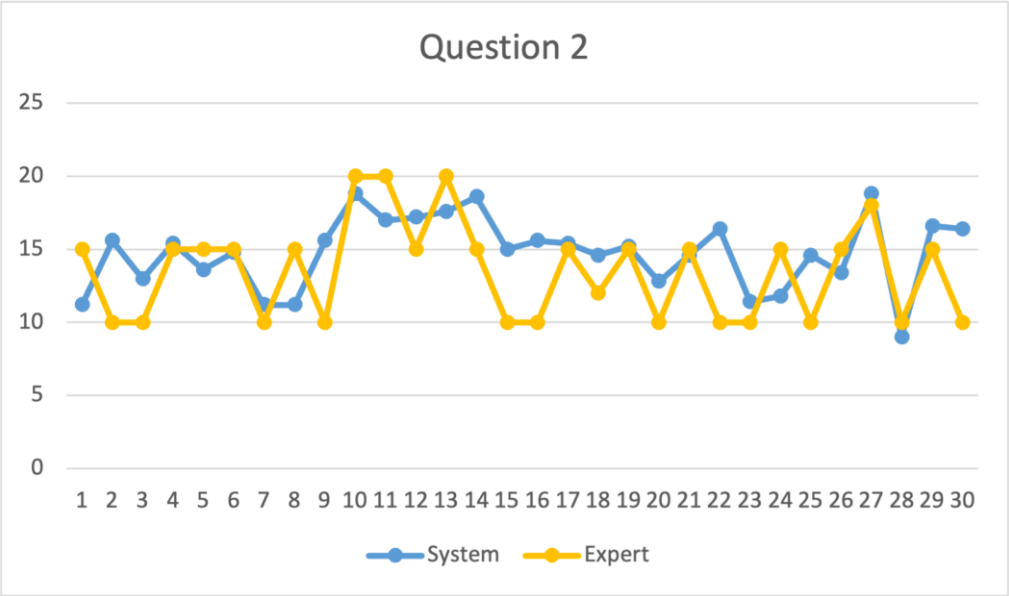

Figure 3. Comparison the value generated by the system and expert in Question 2

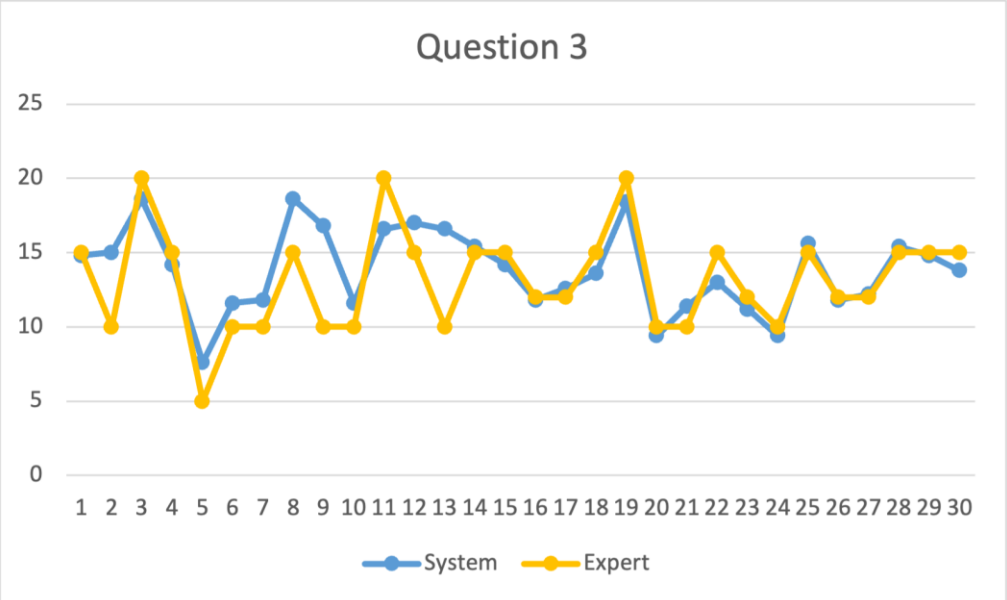

Figure 4. Comparison the value generated by the system and the expert in Question 3

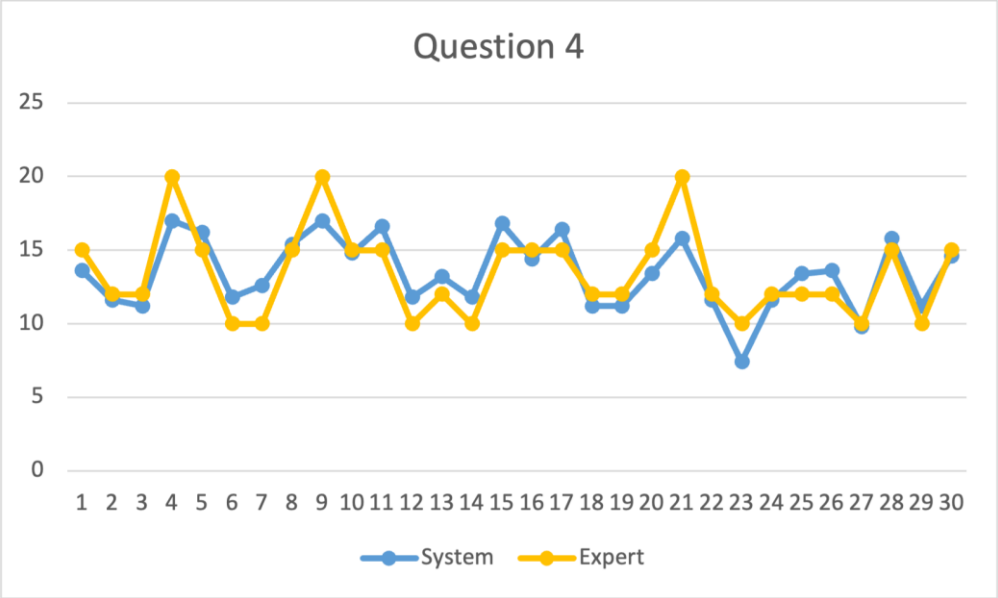

Figure 5. Comparison the value generated by the system and the expert in Question 4

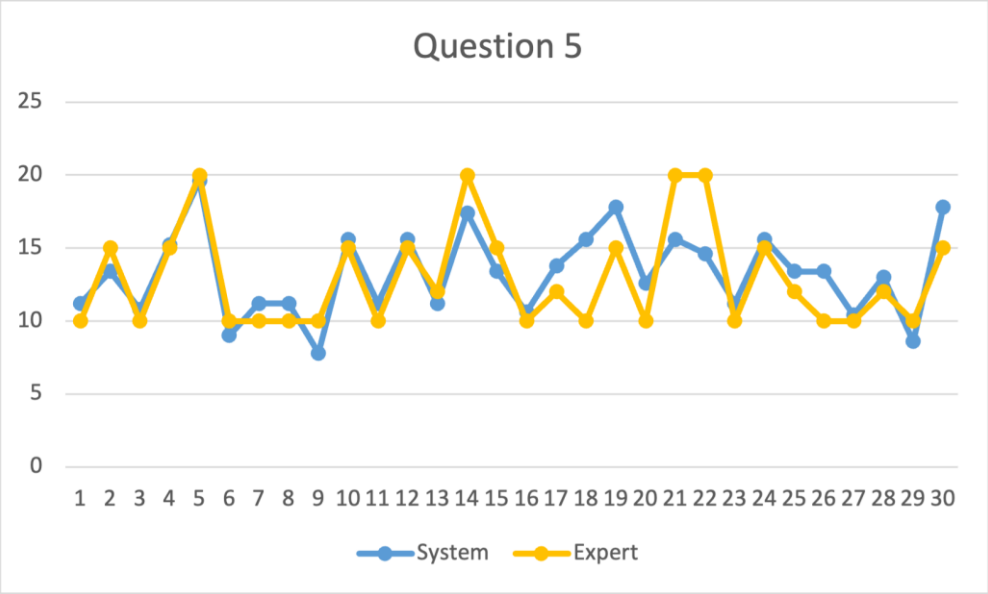

Figure 6. Comparison the value generated by the system and expert on Question 5

In Figure 2 to Figure 6 shows the similarity of the value given by the system and the value given by the expert for each question. Application the MAE (Mean Absolute Error) formula on the value the similarity produced by the system with the value that the expert produces the MAE value is 0.132. The difference between the system's assessment and the expert's assessment is because the student's answers contain different words from the expert's answers, but have the same meaning. The most significant difference in value between system and expert is shown in Figure 3.

CONCLUSION

Study this has succeeded in develop system evaluation exam automatic for question essay use cosine similarity method. System was evaluated through the MAE (Mean Absolute Error) testing process with the error value is 0.132. Based on results research conducted using sample data answer student totaling 30 people with amount question as many as 5 questions. The value generated by the system is multiplied by

with score actually for every matter. In this study, every the similarity value of the system will be multiplied by 20 points, will continue evaluated with using MAE to see the resulting error rate.

REFERENCES

[1] H. Sun and Z. Shao, "Research on the Application of Students' Answered Record Analyze Model and Question Automatic Classify Based on K-Means Clustering Algorithm," in *2019 10th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, 2019, pp. 494–497, doi: 10.1109/ITME.2019.00116.

[2] A. Groza and R. Szabo, "Enacting textual entailment and ontologies for automated essay grading in chemical domain," in *2015 16th IEEE Int. Symp. Comput. Intell. Inform. (CINTI)*, 2015, pp. 221–226, doi: 10.1109/CINTI.2015.7382926.

[3] Z. Ke and V. Ng, "Automated Essay Scoring: A Survey of the State of the Art," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2019-Augus, pp. 6300–6308, 2019, doi: 10.24963/ijcai.2019/879.

[4] S. Drolia, S. Rupani, P. Agarwal, and A. Singh, "Automated Essay Rater using Natural Language Processing," *Int. J. Comput. Appl.*, vol. 163, no. 10, pp. 44–46, 2017, doi: 10.5120/ijca2017913766.

[5] G. Lv, W. Song, M. Cheng, and L. Liu, "Exploring the Effectiveness of Question for Neural Short Answer Scoring System," in *2021 IEEE 11th Int. Conf. Electron. Inf. Emerg. Commun. (ICEIEC)*, 2021, pp. 1–4, doi: 10.1109/ICEIEC51955.2021.9463814.

[6] N. Chamidah, M. M. Santoni, H. N. Irmanda, R. Astriratma, L. M. Tua, and T. Yuniati, "Word Expansion using Synonyms in Indonesian Short Essay Auto Scoring," in *2021 Int. Conf. Inform. Multimed. Cyber Inf. Syst. (ICIMCIS)*, 2021, pp. 296–300, doi: 10.1109/ICIMCIS53775.2021.9699374.

[7] T.-H. Chang, J.-L. Chen, H.-M. Chou, M.-H. Bai, F.-Y. Hsu, and Y.-C. Chen, "Automatic Scoring Method of Short-Answer Questions in the Context of Low-Resource Corpora," in *2021 Int. Conf. Asian Lang. Process. (IALP)*, 2021, pp. 25–29, doi: 10.1109/IALP54817.2021.9675160.

[8] U. Hasanah, T. Astuti, R. Wahyudi, Z. Rifai, and R. A. Pambudi, "An Experimental Study of Text Preprocessing Techniques for Automatic Short Answer Grading in Indonesian," in *2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. (ICITISEE)*, 2018, pp. 230–234, doi: 10.1109/ICITISEE.2018.8720957.

[9] A. Mishra and S. Vishwakarma, "Analysis of TF-IDF Model and its Variant for Document Retrieval," in *2015 Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, 2015, pp. 772–776, doi: 10.1109/CICN.2015.157.

[10] M. Ramya and J. A. Pinakas, "Different Type of Feature Selection for Text Classification," *Int. J. Comput. Trends Technol.*, vol. 10, no. 2, pp. 102–107, 2014, doi: 10.14445/22312803/IJCTT-V10P118.

[11] A. N. Khusna and I. Agustina, "Implementation of Information Retrieval Using Tf-Idf Weighting Method On Detik.Com's Website," in *2018 12th Int. Conf. Telecommun. Syst. Serv. Appl. (TSSA)*, 2018, pp. 1–4, doi: 10.1109/TSSA.2018.8708744.

[12] J. Jumanto, M. A. Muslim, Y. Dasril, and T. Mustaqim, "Accuracy of Malaysia Public Response to Economic Factors During the Covid-19 Pandemic Using Vader and Random Forest," *J. Inf. Syst. Explor. Res.*, vol. 1, no. 1, pp. 49–70, 2022, doi: 10.52465/joiser.v1i1.104.

[13] A. Falasari and M. A. Muslim, "Optimize Naïve Bayes Classifier Using Chi Square and Term Frequency Inverse Document Frequency For Amazon Review Sentiment Analysis," *J. Soft Comput. Explor.*, vol. 3, no. 1, pp. 31–36, 2022, doi: 10.52465/joscex.v3i1.68.

[14] W. B. Trihanto, R. Arifudin, and M. A. Muslim, "Information Retrieval System for Determining The Title of Journal Trends in Indonesian Language Using TF-IDF and Na?ve Bayes Classifier," *Sci. J. Inform.*, vol. 4, no. 2, pp. 179–190, 2017, doi: 10.15294/sji.v4i2.11876.

[15] S. Fransiska, Rianto, and A. I. Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Inform.*, vol. 7, no. 2, pp. 2407–7658, 2020.

[16] K. Taghipour and H. T. Ng, "A Neural Approach to Automated Essay Scoring," in *Proc. 2016 Conf. Empir. Methods Nat. Lang. Process.*, 2016, pp. 1882–1891, doi: 10.18653/v1/D16-1193.

[17] L. H. Pramono, A. S. Rohman, and D. H. Hindersah, "Modified Weighting Method in TF*IDF Algorithm for Extracting User Topic Based On Email and Social Media in Integrated Digital Assistant," in *2013 Jt. Int. Conf. Rural Inf. Commun. Technol. Electr.-Veh. Technol. (rICT ICeV-T)*, 2013, pp. 1–6, doi: 10.1109/rICT-ICeVT.2013.6741547.

[18] D. Palma and J. Atkinson, "Coherence-Based Automatic Essay Assessment," *IEEE Intell. Syst.*, vol. 33, no. 5, pp. 26–36, 2018, doi: 10.1109/MIS.2018.2877278.

[19] P. Lagakis and S. Demetriadis, "Automated essay scoring: A review of the field," in *2021 Int. Conf. Comput. Inf. Telecommun. Syst. (CITS)*, 2021, pp. 1–6, doi: 10.1109/CITS52676.2021.9618476.
[20] N. P. Ririanti and A. Purwinarko, "Implementation of Support Vector Machine Algorithm with Correlation-Based Feature Selection and Term Frequency Inverse Document Frequency for Sentiment Analysis Review Hotel," *Sci. J. Inform.*, vol. 8, no. 2, pp. 297–303, 2021, doi: 10.15294/sji.v8i2.29992.
[21] G. Liu, K. Y. Lee, and H. F. Jordan, "TDM and TWDM de Bruijn networks and ShuffleNets for optical communications," *IEEE Trans. Comput.*, vol. 46, no. 6, pp. 695–701, 1997, doi: 10.1109/12.600827.
[22] J. O. Contreras, S. Hilles, and Z. A. Bakar, "Essay Question Generator based on Bloom's Taxonomy for Assessing Automated Essay Scoring System," in *2021 2nd Int. Conf. Smart Comput. Electron. Enterp. (ICSCEE)*, 2021, pp. 55–62, doi: 10.1109/ICSCEE50312.2021.9498166.
[23] Gunawansyah, R. Rahayu, Nurwathi, B. Sugiarto, and Gunawan, "Automated Essay Scoring Using Natural Language Processing And Text Mining Method," in *2020 14th Int. Conf. Telecommun. Syst. Serv. Appl. (TSSA)*, 2020, pp. 1–4, doi: 10.1109/TSSA51342.2020.9310845.
[24] A. A. P. Ratna, L. Santiar, I. Ibrahim, P. D. Purnamasari, D. L. Luhurkinanti, and A. Larasati, "Latent Semantic Analysis and Winnowing Algorithm Based Automatic Japanese Short Essay Answer Grading System Comparative Performance," in *2019 IEEE 10th Int. Conf. Aware. Sci. Technol. (iCAST)*, 2019, pp. 1–7, doi: 10.1109/ICAwST.2019.8923226.
[25] P. Pathak, S. Raghav, S. Jain, and S. Jalal, "Essay Rating System Using Machine Learning," in *2021 5th Int. Conf. Inf. Syst. Comput. Netw. (ISCON)*, 2021, pp. 1–6, doi: 10.1109/ISCON52037.2021.9702504.
[26] S. Robertson, "Understanding Inverse Document Frequency: On Theoretical Arguments for IDF," *J. Doc.*, vol. 60, no. 5, pp. 503–520, Oct. 2004, doi: 10.1108/00220410410560582.
[27] A. Ayob, "Comparison Between Conventional and Digital Essay Writing Assessment System: Consumer Concept and User Friendly," *Res. World Econ.*, vol. 10, no. 2, p. 96, 2019, doi: 10.5430/rwe.v10n2p96.
[28] Y. Yang, L. Xia, and Q. Zhao, "An Automated Grader for Chinese Essay Combining Shallow and Deep Semantic Attributes," *IEEE Access*, vol. 7, pp. 176306–176316, 2019, doi: 10.1109/ACCESS.2019.2957582.
[29] A. K. Patidar, J. Agrawal, and N. Mishra, "Analysis of Different Similarity Measure Functions and Their Impacts on Shared Nearest Neighbor Clustering Approach," *Int. J. Comput. Appl.*, vol. 40, no. 16, pp. 1–5, 2012, doi: 10.5120/5061-7221.
[30] B. A. A. Nazief and M. Adriani, "Confix-Stripping : Approach to Stemming Algorithm for Bahasa Indonesia," in *Int. Conf. Inf. Knowl. Manag.*, 1996, pp. 560–563.