ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

Automatic Essay Scoring with Context-based Analysis with Cohesion and Coherence

Winarsih[1]
Gunadarma University
Depok, Indonesia

Adang Suhendra[2]
Gunadarma University
Depok Indonesia

Ana Kurniawati[3]
Gunadarma University
Depok, Indonesia

Abstract:- Automatic Essay Scoring (AES) with context-based analysis with cohesion and coherence aims to develop a model that can assess essays automatically or by translating language diversity and student understanding. AES with context-based analysis using methods and applications based on Natural Language Processing (NLP) and the Machine Learning Framework (MLF), apart from being able to provide essay answers automatically, can also assess student understanding. Student or student understanding is the value obtained from answering questions according to the level of understanding that comes from the answer. By using the concepts of cohesion and coherence in the essay assessment system, the teacher can assess the quality of the answers obtained.

The context-based essay assessment system was built to facilitate and speed up the process of assessing essay exam answers, to obtain standards and consistency in essay assessment according to the diversity of answers and the diversity of assessors if they have more than one subject. An essay exam is a learning evaluation given in the form of essay questions which have more varied answers than multiple choice questions. These variations in answers create difficulties for lecturers or teaching staff in assessing answers.

Keywords:- *AES; Cohesion ; Coheresion; NLP; Machine Learning.*

I. INTRODUCTION

Along with the rapid progress of the world of information technology today, the world of education has also experienced a shift in meaning where carrying out learning activities can now be done through websites called E-Learning. Electronic learning (E-Learning) allows the learning process to be carried out en masse with a large number of students and provides the advantage of saving classroom space and teaching staff and can be carried out remotely. E-Learning can be used as a place to share learning materials, a place to give students assignments, a place to practice questions or exams and a place for discussions.

The introduction of automated essay scoring or often called Automated Essay Scoring is an innovative attempt to reduce essay scrutiny and eliminate scoring bias and discrepancies. AES emerged as standalone computer software or distributed services that evaluate and grade written prose (Shermis & Burstein, 2003). The goal of AES is to overcome the time, cost, and reliability issues of manual essay grading. It should be made clear that AES is not intended to completely replace human graders but to be used as part of low-stakes classroom assessments to assist teachers' essay grading routines. On the other hand, AES can be adopted in large-scale high-stakes assessments for the purpose of increasing reliability, where AES serves as an additional assessor for cross-checking. The introduction of auto-grading essays or AES is an innovative attempt to reduce essay checking and eliminate grading bias and discrepancies. Automated Essay Scoring (AES) emerged as standalone computer software or distributed services that evaluate and grade written prose (Shermis & Burstein, 2003). Automatic essay assessment (AES) is part of artificial intelligence which is combined with machine learning tasks in natural language processing. In research, automatic essay assessments, it must have a high accuracy value, the data that can be used for the essay assessment model has a high heterogeneity value, because it is proportional to the type of writer, diversity of language and understanding, compared to the limited amount of data collection.

Automatic essay research in case studies using Indonesian, often uses the Latent Semantic Analysis (LSA) method or similarity of term frequencies. in an essay research sentence is automatically required to vary statements within the same meaning or content, therefore there is a need for context-based analysis using linguistic analysis processes, such as cohesion and coherence. Cohesion is the relationship between propositions that are stated explicitly by grammatical and semantic elements in the sentences that form discourse. Cohesion is a formal aspect in grammar, especially in the syntactic organization consisting of sentences in order to produce complete speech. In his analysis, In a cohesion statement, there are three meanings consisting of semantics, morphology, grammatical and phonological. in the work function of cohesion it can mean the relationship between words and sentences, in this concept cohesion can be used to

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

understand meanings that are difficult for a system, especially in automatic essay assessment. Meanwhile, coherence is the harmony between lingual units in text or speech. Coherence also means a semantic relationship between elements.

II. LITERATURE REVIEW

A. Instructional Media

Learning media also has benefits. Azhar Arsyad (2011: 25-26) reveals several practical benefits from using learning media in the teaching and learning process as follows:

- Clarify the presentation of messages and information.
- Increase and direct children's attention so that it can create motivation to learn, more direct interaction between students and their environment, and the possibility for students to learn independently according to their abilities and interests.
- Provide students with a common experience regarding events in their environment.
- Overcoming the limitations of the senses, space and time.

➢ *Nana Sudjono and Ahmad Rival (2010: 2) Detail the Benefits of Teaching media in the Student Learning Process as follows:*

- Teaching will attract more students' attention so that it can foster motivation to learn.
- The meaning of teaching materials will be clearer so that it can be better understood by students, and enable students to master the teaching objectives better.
- Teaching methods will be more varied, not just verbal communication through verbal regulations by the teacher, so that students don't get bored and teachers don't run out of energy, especially if the teacher teaches every class hour.
- Students do more learning activities, because they not only listen to the teacher's explanations, but also other activities such as observing, doing, demonstrating and so on.

Based on the benefits that have been described, it can be concluded that the use of learning media in the teaching and learning process can direct students' attention so that it creates motivation to learn and the material taught will be clearer, quickly understood so that it can improve student achievement.

In this research, learning media is useful for evaluating learning which is given in the form of essay questions which have more varied answers than multiple choice questions".

B. Cosine Similarity

Cosine Similarity is a measure of the similarity between two vectors in a dimensional space. Cosine similarity is used in positive space, where the results are limited between the values 0 and 1. Two vectors are said to be similar when the value of cosine similarity is 1. Each word/term (term) is assumed to be a different dimension and the document is marked with a vector where the value each dimension corresponds to how many terms appear in the document. The following is the cosine similarity formula, (Ariantini et al., 2016):

$$Similarity = cos(\theta) = \frac{A.B}{\|A\| .\|B\|} = \frac{\sum_{i=1}^{n} A_i . B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} . \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

From the formula above, if it is related to information retrieval, A is the weight of each term in document A, and B is the weight of each term in document B. Similarity measurements can be done by comparing document 1 with document 2, then the system will calculate the similarity value. Ai .Bi is the value obtained from term A and term B, then the two values are added together, then the value $Ai2$, all the values of term document A, all the values are raised to the power of two, as well as for term $Bi2$, all the values obtained are raised to the power of two, then all the values obtained are added up so that A similarity value can be obtained from the two documents.

C. Cohesion and Coheresion

Cohesion use of relationships between various elements that express conflict by using the conjunction but or yet, concessive by using the conjunction although or more, purpose by using the conjunction so or so, advantages by using the conjunction even or even, exceptions by using the conjunction unless.

Coherence and Coherence in use can have different meanings but the words replaced and substituted must refer to a reference that is considered the same. Replacement of forms refers to the same group. Another replacement in discourse is replacement through metaphor. Cohesion formation can also be done with lexical relationships. Cohesion is also established by showing part-whole relationships.

Discourse is cohesive if there is conformity of language form to the co-text (situation within the language) as opposed to the context (situation outside the language). This means that a mismatch between the form of language between the co-text and the context will produce a text that is not cohesive. Cohesion in a paragraph is the attraction between sentences in a paragraph, so that they do not contradict each other, but appear to unite to support the main idea or main idea of a paragraph. Paragraphs that meet these criteria are called cohesive paragraphs

Discourse is composed of descriptions of words that form sentences which are then arranged into paragraphs which are interconnected to form a unified whole. In the discourse there is a topic, while the sentences that are related to each other explain the topic.

The integrity of discourse can be determined by cohesiveness (integration of form) and coherence (continuity of meaning). So, the integrity of a discourse can be seen from the relationship between cohesion (form) and coherence (meaning) of the utterance. The relationship between cohesion (form) can be seen explicitly or clearly, while the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

relationship between coherence (meaning) cannot be seen explicitly. In other words, the cohesiveness of the form of speech does not guarantee the coherent meaning of the discourse. A discourse sometimes has cohesion but no coherence, making it difficult to understand the discourse.

Cohesion in discourse does not only state the relationship of external forms, but what is important is that (good cohesion) implies coherence, namely the semantic relationship that underlies the discourse. So the cohesiveness of a discourse does not guarantee that the discourse is good in terms of coherence or sequence of meaning but explicitly (real) cohesiveness or the integrated form of a discourse can illustrate that the discourse also has a coherent meaning (coherence). And good cohesion is one that implies coherence in the discourse so that there is an integration of form and a sequence of meaning in the discourse.

➢ *The Concept of Cohesion in Discourse can be Seen as follows*

$$dw_{ik} = tf_{ik} * (ln(n) - ln(df_k) + 1)$$

- tfik is the word frequency of term k in document i
- i=document
- dfk = number of documents in the reference corpus
- n = total number of documents in the reference corpus
- In this method the clause weight is simply the sum of the tf idf weights of the words in the clause

➢ *Meanwhile, the Concept of Coherence is as follows:*

Coherence works in the form of similarity values in the form of a matrix. This matrix describes the number of similar meanings/similarities using coherence classes. Cohesion is often contrasted with coherence, as it relates to the level structuring of multi-sentence texts in terms of the relationships between sentences and clauses. This can determine the student's understanding value\

$$SCC(C) = \begin{cases} 0 \\ 1 \\ MMIC(C) \\ \frac{k(k-1)[MMAC(C)+2MMIC(C)+l(l-1)AAC(C)+2lkAMC(C)}{3k(k-1)+l(l-1)+2lk} \end{cases}$$

- Scc = similarity class coherence
- 0 = student does not understand
- 1 = student understands
- MMIC = measurement of understanding between values 0 and 1
- i= document
- K= number of equal Coherence

III. RESEARCH METHOD

➢ *A General Overview of the Research can be Seen in the following Picture*

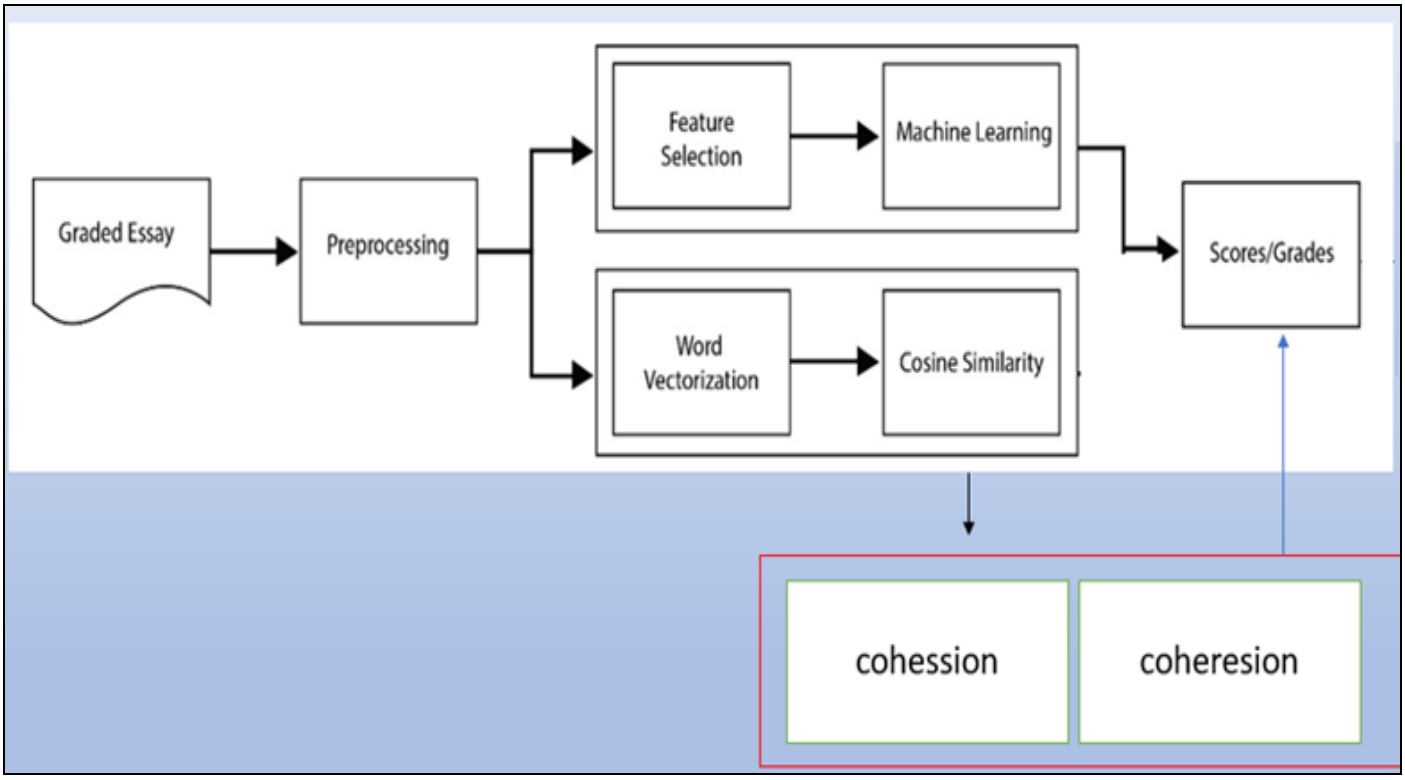

Fig 1 Overview Research

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

Figure 1 shows the research sequence after preprocessing and obtaining a corpus or answer reference which will be stored in the database. There are two steps, namely comparing the Machine Learning concept. The Machine Learning concept is the basis for training data to be compared with the test concept carried out using cosine similarity. To determine students' understanding scores, it is necessary to assess essay answers sequentially and continuously using the concepts of cohesion and coherence.

The concept of cohesion works after the results of word vector and cosine similarity are in the form of similarity values in the form of a matrix. This matrix describes the direct interaction between the AES method and the question completeness attribute, and the matrix is formed based on the assumption that the set of data types accessed by the method is an intersection of the set of data types in the method parameters and class attributes. In general, this representation also allows the representation of different levels of hierarchical structure. for linguistic analysis because words are structured and can be grouped into phrases, into clauses and sentences. Obviously such a representation can represent cohesion.

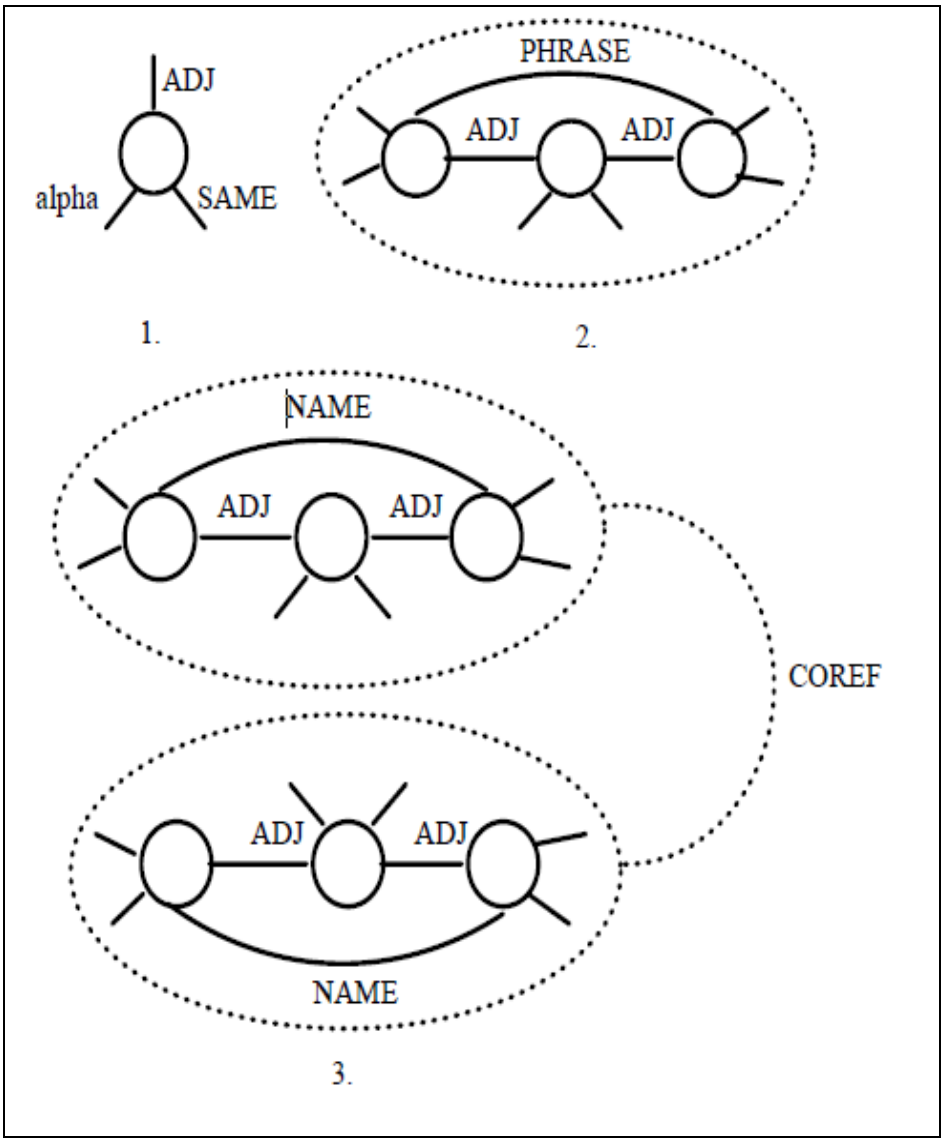

Fig 2 Cohesion Dan Cohesion Concept

In the Coherence concept, Coherence works after the results of the word vector and cosine similarity are in the form of similarity values in the form of a matrix. This matrix describes the number of similar meanings/similarities using coherence classes. Cohesion is often contrasted with coherence, as it relates to the level structuring of multi-sentence texts in terms of the relationships between sentences and clauses. This can determine the student's understanding value.

The AES system works by collaborating with the cohesion and coherence system as a value system that comes from the wording of the students' own answers.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

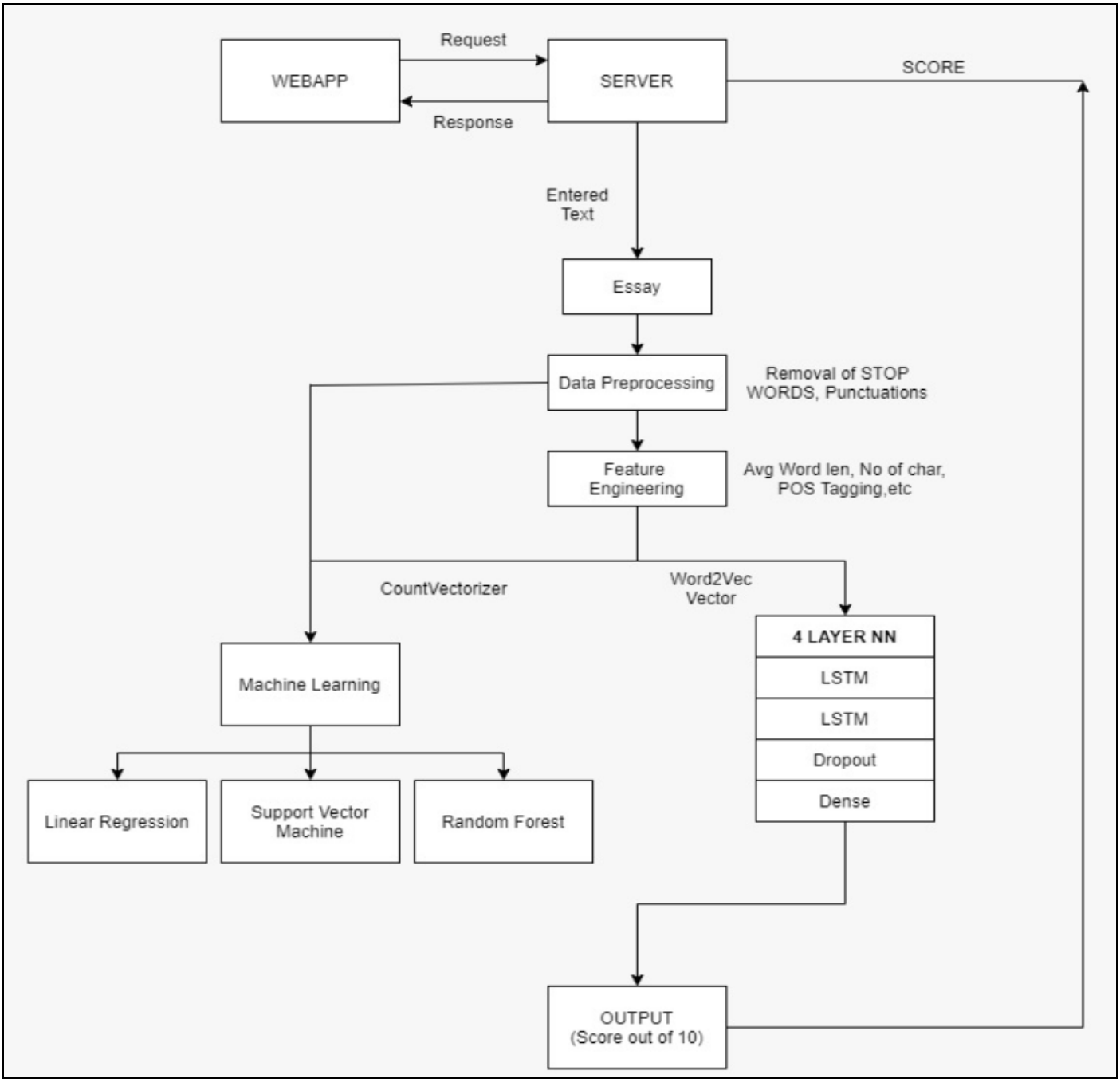

Fig 3 Application Overview

In Figure 3 the AES application is in the form of a web application that has a server to carry out word processing from data input, data collection to the output score which is the student's understanding value.

In this research, this evaluation method compares the average scores marked manually by human raters and the scores generated automatically by the proposed model to test its accuracy. Prediction results are classified into three classes,

➢ *Quadrated Weighted Kappa (QWK)*

Quadratic weighted kappa will find the agreement between the human evaluation score and the system evaluation score and produce a value ranging from 0 to 1, Kappa can range from -1 (no agreement) to +1 (perfect agreement).

- When k = 0, student answers in AES are not very consistent and complete
- When k = negative, student answers are very inconsistent and incomplete.
- When k is positive, students' answers are very consistent and complete.

➢ *Mean Square Error*

Mean square error (MSE) measures the average squared error, that is, the average squared difference between human judgment and the system-generated score. MSE will always only give positive numbers.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

➢ *Pearson Correlation Coefficient*

Pearson Correlation Coefficient (PCC) finds the correlation coefficient between two variables. This will give three values (0, 1, 1). "0" represents the human value and the system score are unrelated. "1" represents an increase in two scores. "– 1" describes a negative relationship between two scores.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

IV. RESULT

➢ *The Results of Cohesion and Coherence begin with Detecting Student Answers, with the Following Example:*

Table 1 Example of Assumptions using Cohesion and Coherence

	Input	Proses	Output	Keyboard	Disk
Q1	0.8	0.9	0.5	0.1	0.02
Q2	0.1	0.1	0.02	1.2	0.75

Based on the assumption that: A text (document) consists of several topics. A topic consists of a collection of words. Topic Modeling Algorithms are mathematical/statistical models used to infer what topics better represent the data. For simplicity, a topic can be described as a collection of words, such as ['input', 'process', 'output'] and ['keyboard',' disk'], but in practice, what the algorithm does is assign each word in the vocabulary a grade within a particular topic. The words with the highest scores can be considered as actual participants of a topic.

➢ *For Example, a Student's Answer Contains the following Words.*

- Student 1: input, output, keyboard, disk (answers understand)
- Student 2: data, information, network. (answer does not understand)

The flow of the cohesion concept compares student answers on one topic or question with a corpus reference or existing answers, these two concepts are combined in the form of metrics. If there are many similarities between the topic and reference, the cohesion value will be high (close to 1).

➢ *After the Cohesion Process Continues with Cosine Similarity, or Finding the Value of each Metric.*

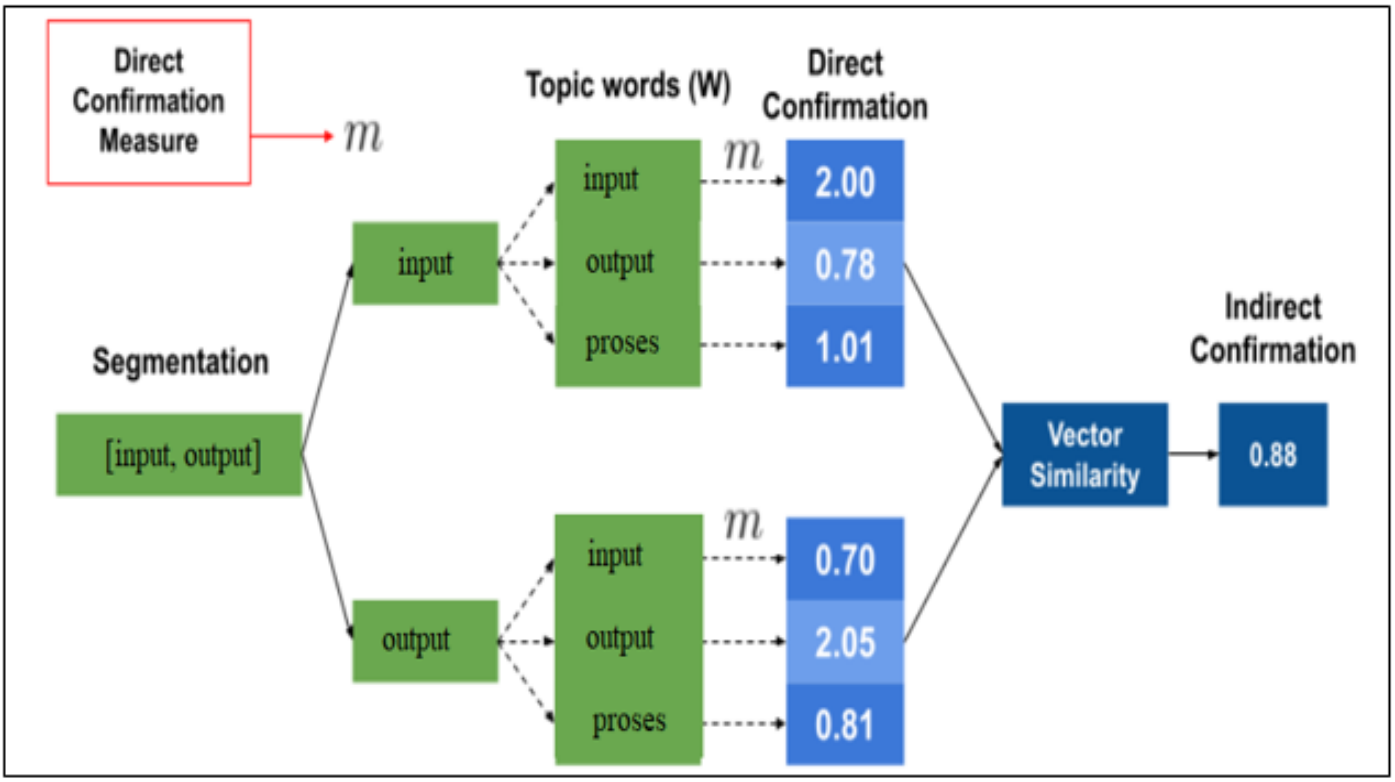

Fig 4 Cosine Similarity

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

In Figure 4, the segmentation results of student answers can be explained, divided into 2 main parts in word topics. This word topic has a vector value or percentage of word similarities. The final process is looking for the average value. This average will be the input for the coherence value.

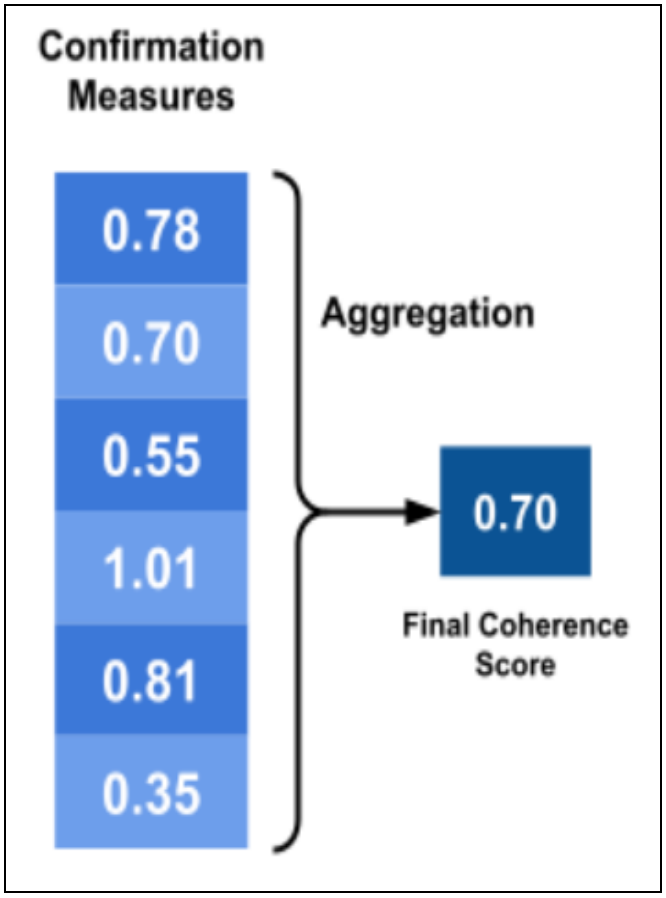

Fig 5 Final Coherence Score

After obtaining cohense, coherence and cosine similarity, the next step is to assess accuracy

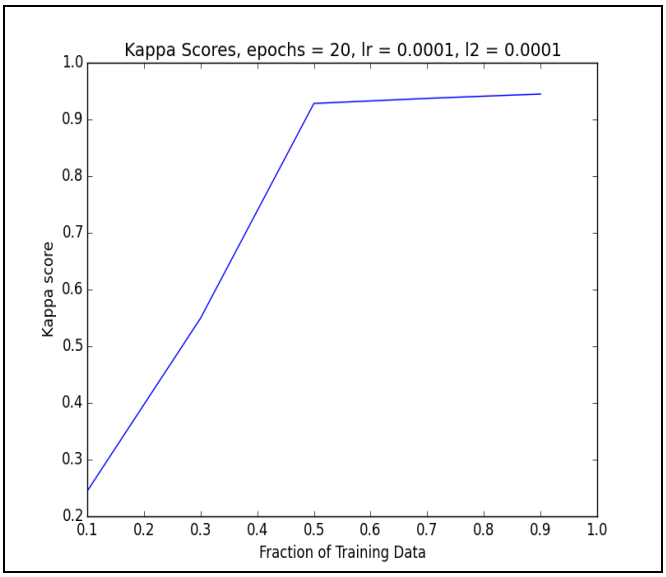

Fig 6 Kappa Result

In figure 6 is the result of QWK with a kappa score value of less than 5 percent of the percentage of errors in training, this can be concluded that the system is running well and has high accuracy, reaching 0.9 or 90%.

PCC (Pearson Correlation Coefficient) Results. This test is based on the value of the relationship between the essay answers from the user or students and the test answers from the teacher. This relationship is to compare the results of this research concept.

Table 2 PCC Result

Student	PCC	Information
1	-1	do not understand
2	0.89	very understanding
3	0.54	understand
4	0.89	understand

From table 2 are the results using a statistical program, by entering the results of cohesion and coherence, these results are entered as input in SPSS, and the results show that the answers from students are in accordance with the research results, namely student 1 with the statement that he does not understand, student 2 really understands , and students 3 and 4 with understanding

V. CONCLUSION

In this research, automatic essay scoring has been carried out using cohesion and coherence techniques, as well as cosine similarity. The results show that research can automatically assess answers along with students' understanding scores. The accuracy value from QWK shows that the system shows an error of less than 5 percent, meaning that the system can be used widely.

REFERENCES

[1]. Cutrone, Maiga Chang and Kinshuk, “Auto-Assessor:Computerized Assessment System for Marking Student’s Short-Answers Automatically”, IEEE International Conference on Technology for Education, 2011

[2]. Thomas N. T., Ashwini Kumar, and Bijlani, “Automatic Answer Assessment in LMS (Learning Management Systems) using Latent Semantic Analysis”, Procedia Computer Science , 2015

[3]. Ms. Shweta M. Patil and Prof. Ms. Soanl Patil, “Evaluating Student Descriptive Answers Using Natural Language Processing”, International Journal of Engineering Research & Technology (IJERT), Vol. 3, Maret, 2014

[4]. Senthil Kumaran and A. Sankar, “Towards an Automated System for Short-Answer Assessment Using Ontology Mapping”, International Arab Journal of Technology, Vol.4, January, 2015

[5]. Emad Fawzi Al-Shalabi, “An Automated System for Essay Scoring of Online Exams in Arabic based on Stemming Techniques and Levenshtein Edit Operations”, International Journal of Computer Science Issues (ICJSI)”, Vol. 13, September, 2016

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY200

[6]. Darwish SM, Mohamed SK (2020) Automated essay evaluation based on fusion of fuzzy ontology and latent semantic analysis. In: Hassanien A, Azar A, Gaber T, Bhatnagar RF, Tolba M (eds) The International Conference on Advanced Machine Learning Technologies and Applications
[7]. Dasgupta T, Naskar A, Dey L, Saha R (2018) Augmenting textual qualitative features in deep convolution recurrent neural network for automatic essay scoring. In: Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications p 93–102
[8]. Ding Y, et al. (2020) "Don't take "nswvtnvakgxpm" for an answer–The surprising vulnerability of automatic content scoring systems to adversarial input." In: Proceedings of the 28th International Conference on Computational Linguistics
[9]. Ajetunmobi SA, Daramola O (2017) Ontology-based information extraction for subject-focussed automatic essay evaluation. In: 2017 International Conference on Computing Networking and Informatics (ICCNI) p 1–6. IEEE
[10]. Alva-Manchego F, et al. (2019) EASSE: Easier Automatic Sentence Simplification Evaluation." ArXiv abs/1908.04567 (2019): n. pag
[11]. Chen M, Li X (2018) "Relevance-Based Automated Essay Scoring via Hierarchical Recurrent Model. In: 2018 International Conference on Asian Language Processing (IALP), Bandung, Indonesia, 2018, 378–383, doi: https:// doi. org/ 10. 1109/ IALP. 2018. 86292 56
[12]. Chen Z, Zhou Y (2019) "Research on Automatic Essay Scoring of Composition Based on CNN and OR. In: 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, p 13–18, doi: https:// doi. org/ 10. 1109/ ICAIBD. 2019. 88370 07
[13]. Contreras JO, Hilles SM, Abubakar ZB (2018) Automated essay scoring with ontology based on text mining and NLTK tools. In: 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), 1-6
[14]. D. Grimes and M. Warschauer, "Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing
[15]. Evaluation," Journal of Technology, Learning, and Assessment, vol. 8, no. 6, Mar. 2010, publisher: Technology and
[16]. Assessment Study Collaborative. [Online]. Available: https://eric.ed.gov/?id=EJ882522
[17]. Y. Attali and J. Burstein, "Automated essay scoring with e-rater® v.2," The Journal of Technology, Learning and
[18]. Assessment, vol. 4, no. 3, Feb. 2006. [Online]. Available: https://ejournals.bc.edu/index.php/jtla/article/view/1650
[19]. Y. Tay, M. Phan, L. A. Tuan, and S. C. Hui, "Skipflow: Incorporating neural coherence features for end-to-end
[20]. [automatic text scoring," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, 2018.
[21]. P. Wangkriangkri, C. Viboonlarp, A. T. Rutherford, and E. Chuangsuwanich, "A comparative study of pretrained language models for automated essay scoring with adversarial inputs," in 2020 IEEE REGION 10 CONFERENCE (TENCON), 2020, pp. 875–880.
[22]. K. Taghipour and H. T. Ng, "A neural approach to automated essay scoring," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: Association for Computational
[23]. Linguistics, Nov. 2016, pp. 1882–1891. [Online]. Available: https://www.aclweb.org/anthology/D16-1193
[24]. Z. Ke and V. Ng, "Automated essay scoring: A survey of the state of the art," in Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19. International Joint Conferences on Artificial