ISSN: 2278-4632
Vol-11 Issue-02 2021

Online Exam Using AI

1. N.Shilpa,assistant professor,CSE,Sri Indu Institute of Engineering&Technology(SIIET), Sheriguda,Ibrahimpatnam,Hydarabad

2. B.Rajeswari,assistant professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

3.B.S.Swapna Shanthi, assistant professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

4.Vagicherla Venkata Saikarthik,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

5.Sai Vinay Martha,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

6.Sriramoju Uday Kiran,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

***Abstract*—** Every year educational institutes conduct various examinations, which include institutional and non-institutional competitive exams. Now a day's online tests and examinations are becoming popular to reduce the burden of the examination evaluation process. The online exams include either objective or multiple-choice questions. Nevertheless, the exams include only objective or multiple-choice questions. However, subjective-based questions and answers are not involved due to the evaluation process complexity and efficiency of the evaluation process. An automatic answer checker application that checks the written answers and marks the weightage similar to a human being is more helpful in the current modern era is necessary. Hence, the software applications built to check subjective answers may be more useful for allocating marks to the user after verifying the answers for online examination. This type of tool/application/system has the challenge of having an abundant resource database, including questions, corresponding answers, and the marks allocated to the corresponding answers. At the same time system need to verify the answers provided by the users by checking the template answers and the answers provided by the user. However, Artificial Intelligence is required to identify the core element of the answers while allocating marks. Hence, an Artificial Intelligence-based answer verifier is proposed to do the job of examiner/evaluator automatically. As a result of this artificial intelligence-based answer verifier, the evaluator's time and energy can be conserved, with improved work efficiency.

I. INTRODUCTION

The Online Examination is beneficial to users as in the present day, and the online exams are based on objective questions and exams are getting digitized all over the world. In this scenario, exam questions can even be based on subjective answers. Meaning that the traditional pen-paper based tests are replaced by computer-based tests that have proven to be both: (i)more consistent in allocating marks and (ii)faster than teachers correcting papers [1,11,12]. The traditional exam usually consisted of subjective answers, which were not the best way of grading the student's perception of the subject. Because sometimes, examiners get bored by checking many answer sheets, and there may be an increase in the false evaluation. Hence, the Artificial Intelligence-based Answer Verifier is required to grade the student after he/she has solved the question paper. However, the system reduces the workload by automating the manual checking process also. An automatic subjective question's marking is a key technology in the network test system. To solve this problem, Artificial Intelligence (AI) based Answer Verifier (AV) is required to analyze like a grading teacher and think while reviewing subjective questions [3]. In various institutes, the results are declared after time because teachers take a long time to evaluate the subjective papers. As a bundle of answers, booklets need to be evaluated and each booklet may contain the answer in a different manner, which requires a longer duration. Hence, then AI-based Answer Verifier can come into role [4], based on the nearness theory of fuzzy mathematics [2]. It does the automatic scoring of subjective questions through specific reference values. In the investigation process, a decision-based scoring algorithm is identified as an efficient one [6]. The system based on the AI will save time and effort of humans. The online examination makes an interesting working environment by giving faster access. The existing system mainly calculates the score based on one or two parameters like keywords, or synonyms. But, coming to the "Artificial Intelligence-based Answer Verifier", it calculates the score of the student by combining various parameters such as keywords, question specific things along with the proper grammar, which in terms provides a more accurate score. The artificial intelligence-based approach will continue to grow, and ultimately provide a full breadth of services and benefits to the students/teachers to have an efficient grading system

II. LITERATURE SURVEY

Organizations/educational institutes always depend on the grading system through examinations. However, most of the examinations are objective. These systems or any other such system are more advantageous in terms of saving resources but failed to include subjective questions [1, 9, 10]. This paper attempted to evaluate the descriptive answer. The evaluation is done through graphical comparison with a standard answer. A subjective answer verifier was proposed [2] by allotting the marks according to the percentage of accuracy present in the answer for different users providing three different answers. The system should have a database that includes questions, corresponding answers and the marks allocated to the corresponding answers. At the same time system need to verify the answers provided by the users by checking the template answers and the answers provided by the user. However, Artificial Intelligence is required to identify the core element of the answers while allocating marks. The system used a part-to-speech tagger to recognize the user answers. The answers were purely ranked based on the keyword similarities to heuristic metrics. The application achieved 70% efficiency as it was not able to consider mathematical formulas, brief description, examples and problems with the identification of statement formation. Another system [8] was designed for analyzing the subjective answers using fuzzy logic states. However, the system missed verifying the grammar in the sentence and performance evaluation. Work was done on a similar ground [3], which provided the design based on 1:1 string matching from the user answers to the database answers. This kind of system is useful to start but not design an efficient answer verifies. A similar system is proposed [1] to add a grammar verifier. The system was quite approachable with the discussions made, but no traces of implementation and verification of system efficiency. By considering the works done in the past years, can come to the time being a conclusion that artificially based answer verifiers are suitable to define for the subjective answers. However, in most of the works, only 1:1 keyword matching was done and neglected to identify the synonyms words present in the answers. Hence, thought of designing and developing an Artificial Intelligence-based answer verifier to automatically do the job of evaluator for objective and subjective answers with the standard answer can be stored in the database [8, 3] with descriptions and keywords. Then the AI can evaluate each answer by matching the keywords or its synonyms with the standard answer. The system can also be designed to verify the sentence formation through Grammarly tools and assess more weightage for grading the answer. The artificial intelligence-based answer verifier can evaluate the answers if all the credentials are satisfied.

 ISSN: 2278-4632
Vol-11 Issue-02 2021

III. PROPOSED SYSTEM

Fig. 1 shows the general block diagram of the Answer Verifying System. It consists of 4 Blocks, namely: User Answer extractor unit, Answer verifier unit, and Result set unit.
i)User Answer Extractor Unit: User Answer Extractor is a unit that extracts the keywords from the standard answer with the synonyms.

ii)Answer Verifier Unit: This unit consists of two separate modules, namely: Cosine Similarity module and Text Gears Grammar API. Cosine Similarity says about the two relations between the related strings with the angle of magnitude. After that, the extracted keywords checked with the array of the student's answer.
Text Gears grammar API allows the integration of language processing technologies into any product. No matter what it is, mobile application or complex enterprise system. The output of this API is either 0 or 1 based on different input values. The API produces 1 as output if the grammar is entirely perfect, whereas the API produces 0 as output if there is any fault in the sentence..

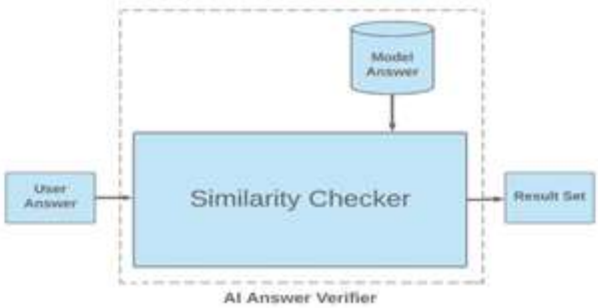

Figure 1: General Block Diagram Answer Verifying System

iii)Result Set Unit: The Result Set Unit consists of 3 attributes viz. "Keywords", "Grammar", "QST"(Question Specific Terms). So, the values of "keywords" and "qst" attributes defined as:

• The value of keywords ranges from 1 to 6 where 1 is for Excellent and 6 is for Very Poor.
• The values of "grammar" attribute ranges from 0 which is for Improper and 1 which is for Proper.
• The value of "class" ranges from 1 to 9, 1 being the best and 9 being the worst. After the calculation of results, the final output is generated that reflects the final score including grammar, keywords and class values of a particular answer.

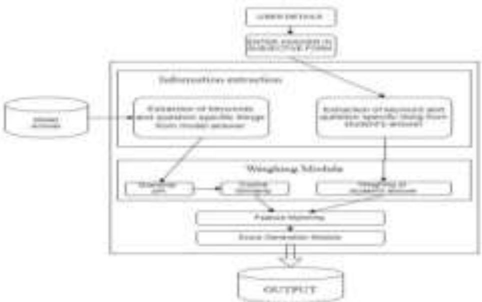

Figure 2: Artificial Intelligence-based Answer Verifier

A. Proposed Artificial Intelligence-based Answer Verifier System

Fig. 2 shows the proposed system. The proposed system is categorized into two main modules namely: Information Extraction module and Weighing Module. Information Extract module: User Answer Extractor is a unit where the user submitted answer is processed through the extraction of keywords from the standard answers with the synonyms. Weighing Module: This unit contains three parts, namely: Cosine Similarity module, Text Gears Grammar API and Results Calculation

B. Answer Verifier Architecture

Fig. 3 shows the proposed system architecture. It consists of two modules: Module 1 consists of Grammar API, and Module 2 includes the cosine similarity algorithm. In the following section, the function of each module is explained in detail. Grammar API: A Grammar Application Programming Interface (API) is generally designed for mobile-based applications, web-based systems, database systems, computer hardware, or software library. In today's life, having Text Gears API, Grammarly API, which provides a good source for our usage. Under this module, Text Gears API is used for the output verification. These APIs produce either zero or one based on different input values. The API produces one as output if the input string has less than the defined number of errors, whereas the API produces zero as output if there are more than a defined number of faults in the sentence. The Text Gears API is free to use for students. The Text Gears API here is a python API client used for checking English grammar in the answers. Cosine Similarity: Cosine Similarity is used to measure the similarity between two non-zero vectors which are the inner product space. The measure is the cosine of the angle between the two vectors i.e., 0° is 1, and less than 1 for any angle in the interval (0, π) radians. Here, the two vectors with the same orientation will have a cosine similarity of 1, and two vectors oriented at 90° relative to each other will have a similarity of 0, and two vectors opposed have a similarity of -1, independent of their magnitude. In this process, the Euclidean dot product is used for text matching, the two vectors of the documents. In the proposed system, two input strings are considered, where the first one being the user answer and the second one being the model answer stored in the database. Then both the strings are converted into vectors and vectors are then processed and a corresponding keyword score will be generated.

C. Working Parameters

For evaluation of the subjective answers in the proposed system, 3 Parameters such as keywords, QST and Grammar are considered. Keywords: Keywords are the main strength of the system, which is matched using the Cosine Similarity. Then the evaluation by converting the text into vectors. If the value is in between 0 to 1, it will be converted into numeric form (i.e. 0 to 100) to get the value for each keyword from 1-6. A lower value is chosen for representing the accuracy of an answer. The system assigns values between 1 and 6 with 1 being the best score and 6 being the worst.

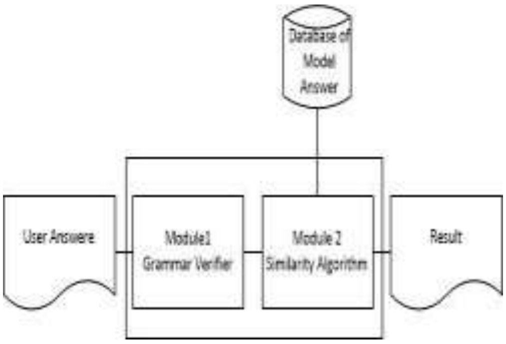

Figure 3: Answer Verifier Architecture

Question Specific Things (QST): Fuzzy Logic is used to give the value of QST. The python module named Fuzzy Wuzzy is used to calculate the QST score of an answer. The set_token_ratio () function is used. Fuzzy Wuzzy token sort ratio raw score is a measure of the string's similarity as an integer in the range of [0, 100]. For two strings X and Y, the score is obtained by splitting the two strings into tokens and then sorting them. The score is the fuzzy-wuzzy ratio for the transformed strings that float in the range [0, 1]. The score is resultant of the raw score divided by 100. Grammar: Grammar is used to form the structure of a sentence. While evaluating, there is a likelihood of writing keywords without the correct sentence formation. However, for more weightage of marks allotment, keywords with correct sentence formation are crucial. Hence, the combination of grammar checking with keywords is essential for the quality of the answers.

D. Score Computation

The dataset has 3 attributes viz. "Keywords", "Grammar", "QST"(Question Specific Terms). The grammar parameters computation is purely based on the considered grammar API. The left-out parameters such as "keywords" and "qst" attributes considered to be computed as

ISSN: 2278-4632
Vol-11 Issue-02 2021

excellent, very good, good, Ok, poor and very poor. Excellent (1): If after comparison a value one is generated, this implies that all the keywords and Question Specific Things were found in the answer. It also means that it is one of the best possible matches with the model answer.

Very good (2): If after comparison a value two is generated, this implies that all the keywords and Question Specific Things were found not in the answer, but only a few were missing. This number of missing keywords and QST depend on the specified number of keywords and QST for the answer. It also means that it is a very good match with the model answer.

Good (3): If after comparison a value three is generated, this implies that more than 70% of keywords and Question Specific Things were found in the answer. It also means that it is a good match with the model answer.

Ok (4): If after comparison a value four is generated, this implies that more than 50% of keywords and Question Specific Things were found in the answer. It also means that it is an acceptable match with the model answer.

Poor (5): If after comparison a value five is generated, this implies that more than 30% of the keywords and Question Specific Things were found in the answer. It also means that it is not an acceptable match with the model answer.

Very poor (6): If after comparison a value six is generated, this implies that more than 10% of the keywords and Question Specific Things were found in the answer. It also means that it is one of the worst possible matches with the model answer. In the same, the values of "grammar" attributes are defined as improper and proper. The following are the respective definitions:

Improper (0): This implies that there are many grammatical mistakes in the answer. It has more than a specified number of mistakes.

Proper (1): This implies that there are few grammatical mistakes in the answer. It has less than the specified number of mistakes. The value of "class" ranges from 1 to 9. The class values are generated based on the parameters mentioned above. All of them combined are used to predict the class value.

IV. RESULT ANALYSIS

A test case was steered to evaluate the recite of the proposed system. The system was considered for the efficiency of the application and the accuracy test. In [1, 3], a similar system was developed, but they did not
mention any clue regarding the system's testing or performance analysis. Hence, some parameters, such as
efficiency, complexity, and reliability measures are found to evaluate the proposed system.

After verifying different methodologies, different parameters were found to evaluate the efficiency of different implementations of AI-based answer verifier:

1) For small to medium length answers the efficiency of this system is around 80%.
2) This system is based on a keyword approach hence the efficiency of this system is around 60-70%.
3) This system is based on keyword matching and length of the answer; hence, its efficiency is around 80-90%.

The following kind of test cases was considered for testing the points mentioned above. The students were asked to register to the developed system through login credentials for some 2 courses. The system was loaded with the expected subjective answers and the registered students were asked to answer for the two courses. Based on the answers provided by the students, the system was tested. While testing, the effect of
extra word evaluation is also done. The consequence of adding 50, 100, 250, and 450 new words in the answer was conducted. At the same time, the effect of different lengths of about 5, 10, 20 and 50 words was also conducted.

ISSN: 2278-4632
Vol-11 Issue-02 2021

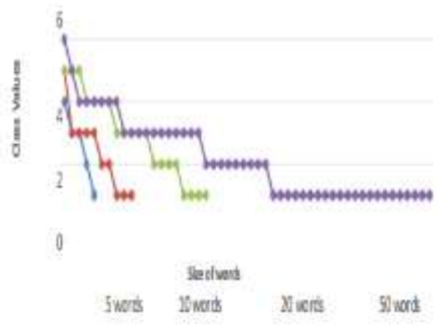

Figure 4: Class values that are obtained after testing the system with different sizes of answers

The Fig.4 describes the class values obtained after testing the AI Answer Verifier system with different sizes of answers, which is the second type of test cases developed to test the system. The answers were of different lengths, starting from 5 words to 50 words. The answers given as inputs were a mix of model answers and user inputs. The model answers were given as input for testing to analyze the working and correctness of the system, while user answers helped in analyzing the efficiency of the system. After the result analysis, it is observed that as the size of answers increases, the efficiency also increases i.e., even if the answer contains all the keywords but if the length of the answer is short, the system will not return perfect results. It implies that the proposed system becomes more efficient as the answers' size increases. Hence, can be concluded that the AI Answer Verifier is more efficient for extensive subjective answers.

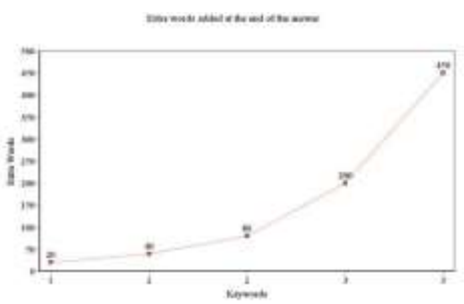

Figure 5: Effect of additional words

Fig. 5 describes the influence of the extra words added at the end of the answer in our considered question and answer set. While testing the proposed system with different sizes of answers viz: 50 words, 100 words, 250 words and 450 words are considered. After the result analysis can be analyzed that as the size of answers increases, the value of the keyword also increases, which means that the less the extra words are added in the answer, the more efficient the system's evaluation. It implies that the proposed system is more efficient for answers that are point to point.

V. CONCLUSION

The proposed work is robust due to the subjective answer verification by allocating the appropriate weights. The system is also verified for adding new words and is not affected by the weightage. The system also has scope for future developments in the system. Hence, any grammar checking can be changed based on the standard requirements. Further, the second module is designed and tested for the 'Cosine Similarity' algorithm. The module can be tested by verifying any other algorithm which can give a more robust and effective answer verifier and grading system.

REFERENCES

[1] Ashutosh Shinde, C. Nishit Nirbhavane, Sharda Mahajan, Vikas Katkar, Supriya Chaudhary, "AI Answer Verifier," Int. Journal of Engineering Research in Computer Science and Engineering (IJERCSE), Volume 5, Issue 3, March 2018.

[2] Asmita Dhokrat, Gite Hanumant R, C. Namrata Mahender, "Assessment of Answers: Online Subjective Examination, In Proc. of the Workshop on Question Answering for Complex Domains," Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India, pp. 47-56, 2014.

[3] Merien Mathew, Ankit Chavan, Siddharth Baikar, "Online Subjective Answer Checker," Int. Journal of Scientific & Engineering Research, Volume 8, Issue 2, 2017.

[4] Sakshi Berad, Prakash Jaybhaye, Sakshi Jawale, "AI Answer Verifier, Int. Research

Journal of Engineering and Technology," Volume 06, Issue 01, 2019.

[5] Jacob Perkins, "Text Processing with Python," Packt Publishing Ltd, 2014.

[6] John Paul Mueller and Luca Massaron, "Machine Learning for Dummies," John Wiley & Sons; 2016.

[7] Nish Tahir Blog, "String matching using cosine similarity algorithm," HackerNews.com,2015. https://blog.nishtahir.com/2015/09/19/fuzzystring-matching-using-cosine-similarity

[8] Xia Yaowen and Li Zhiping Lv Saidong and Tang Guohua, "The Design and Implemention of Subjective Questions Automatic Scoring Algorithm in Intelligent Turtoring System," In 2nd Int. Symposium on Computer, Communication, Control and Automation (ISCCCA-13), Atlantis Press, Paris, France, 2013.

[9] Jacob, I. Jeena. "Performance Evaluation of Caps-Net Based Multitask Learning Architecture for Text Classification." Journal of Artificial Intelligence 2, no. 01 (2020): 1-10.

[10] Vijayakumar, T., and Mr R. Vinothkanna. "Capsule Network on Font Style Classification." Journal of Artificial Intelligence 2, no. 02 (2020): 64-76.

[11] P. Pisat, D. Modi, S. Rewagad, G. Sawant and D. Chaturvedi, "Question paper generator and answer verifier," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, 2017, pp. 1074-1077, doi: 10.1109/ICECDS.2017.8389603.

[12] Yogish, Deepa, T. N. Manjunath, and Ravindra S. Hegadi. "Survey on trends and methods of an intelligent answering system." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 346-353. IEEE, 2017.