IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automated Subjective Answer Evaluation UsingNLP

[1]Gaurang Kudale, [2]Nishant Mali, [3]Nachiket Suryawanshi, [4]Mukesh Bansode, [5] Prof. Richa Agarwal

[1,2,3,4]UG Student, [5]Head of Department of Information Technology Engineering,Trinity College of Engineering and Research, Pune, Maharashtra, India

Abstract: This study has been undertaken to investigate the determinants of Natural Language Processing (NLP) is one of the important issues of concern in giving computers the ability to understand text and speech in much the same way human beings can. NLP can be used in subjective answer evaluation in various ways. One of the most common approaches is to use NLP techniques to automatically score the quality of a written response based on its language features, such as grammar, syntax, vocabulary, and coherence. This can be done by training machine learning algorithms on a large dataset of human-scored essays or short answer responses, using the language features mentioned above as input features, and the corresponding scores as target values. The trained model can then be used to automatically score new responses based on their language features. Another approach is to use NLP techniques to analyze the content and structure of the response, in order to identify key concepts and arguments and assess their relevance and coherence with the question prompt. This can be done by using techniques such as topic modelling, sentiment analysis, and text classification. Overall, NLP can be a powerful tool for subjective answer evaluation, as it can help to improve the efficiency and consistency of the grading process, while also providing valuable insights into the language and reasoning skills of the students.

***Index Terms* - Subjective Answer Evaluation, Natural Language Processing**

I. INTRODUCTION

The manual system for the evaluation of Subjective Answers for technical subjects involves a lot of time and effort of the evaluator. Subjective answers have various parameters upon which they can be evaluated such as the question specific content and writing style. Evaluating subjective answers is a hard task to Perform. When a human being evaluates anything, the quality of evaluation may vary along with the emotions of the person. Evaluation through computers using intelligent techniques ensures uniformity in marking as the same inference mechanism is used for all the students.

Our Proposed System uses machine learning and NLP to solve this problem. Our Algorithm performs tasks like Grammar Check, Check Answer Length, Keyword Extraction, Conceptual Similarity, Semantic Similarity, Contradiction Similarity, Cosine similarity. In the present scenario, manual evaluation of subjective answers is a hectic task. There are multiple systems which can evaluate objective type or MCQ questions quickly. These techniques are evaluated in machines itself after providing a pre-defined correct answer. But it helps only in competitive or objective type exam evaluation.

In this study, we have developed a novel approach of evaluation of subjective answers using different NLP algorithms. Our proposed approach consists of two main processes: Pre-processing of text files followed by its analysis and Evaluation of Students' answers on the basis of analysis. The main advantage of our proposed approach is that it evaluates the answers more efficiently with reducing the human error.

II. LITERATURE REVIEW

A model is created for evaluating subjective responses using a semi-automatic evaluation technique. First, it creates a question base that contains the question type, subtype, question mark, and tags. An answer base is then created with a model answer. The evaluated answer is mapped using a hash index, referred to as the question number. The student's answer is evaluated by considering the semantic meaning and length of the sentence.[1]

The proposed system is designed to evaluate the answers of five students giving five different answers. The standard answer is stored in the database with keywords, meaning, and description of the given answer. Then each answer is evaluated by matching keywords and their synonyms with the standard answer. It also checks the grammar and spelling of words. After evaluation, the answer is scored depending on its correctness. [2]

An E-assessment system was developed to check student's answer sheets and provide the same marks. The system consists of an algorithm that compares the student's answer with three reference answers given by three different faculties, taking into account the answer with the closest results and with the highest accuracy and assigning marks accordingly. An algorithm based on TFIDF, grammar checking, WMD, cosine and Jaccard similarity. Both answers may not be exactly the same or word for word. This approach can be a quick and easy way for examiners to reduce their workload.[3]

The proposed system accepts solution sets from administrator and student responses. Then the stop words are removed from them to generate keywords. After generating the keywords, it checks the similarity, and by calculating the similarity, it also checks the relationship of the keywords along with the sentences with the sentences in the data set, thereby finding the exact similarity and correctness of the sentence with the data sets. If the sentences match the datasets, it generates a similarity score based on the percentage of overlap. It also checks for synonyms and similar words before matching keywords to increase overlap accuracy. The data duplication technique is used to compare the answers entered by the students and grades are generated based on the uniqueness of the answers.[4]

In this paper, they proposed an algorithm for detecting document files' plagiarism, for this they used formal concept analysis (FCA) with the presented concept similarity. The proposed similarity measures employ concept approximation using frequency of the formal concepts and those were mathematically proven to be formal similarity metrics. The source documents were processed and retrieved with the proposed algorithm to visualize performance of the proposed similarity measure in document plagiarism detection by implemented web applications. Also, in the last format, the presented system applies services provided by Google. The proposed system was demonstrated to be efficient and effective with a case study of news and different academic documents. The experiments were evaluated from two aspects: efficiency tests by type of document and its effectiveness test regarding correctness. The results show that (1) their proposed system can detect document types .docx, .pdf, and .txt as designed and (2) their proposed system can detect plagiarized documents with an average accuracy of 94.01%.[5]

In this research they dealt with E-examinations which were randomly designed and they propose an E-assessment system which can be used for subjective questions. This system assesses answers to subjective questions by finding a ratio which matches for the keywords in instructor and student answers. The matching ratio is achieved based on document and semantic similarity. Their assessment system is composed of four modules: pre-processing, keyword expansion, matching, and grading. They used a survey and case study in the research design to validate their proposed system. Their examination assessment system will help instructors to save time, costs, and resources, while increasing efficiency and improving the productivity of exam setting and assessments.[6]

In this paper they present the Word Mover's Distance (WMD), a novel distance function between text documents. The work is based on recent results in word embeddings that learn semantically meaningful representations for words from local co-occurrences in sentences. The WMD distance measures the dissimilarity between two text documents as the minimum amount of distance that the embedded words of one document need to "travel" to reach the embedded words of another document. We show that this distance metric can be cast as an instance of the Earth Mover's Distance, a well studied transportation problem for which several highly efficient solvers have been developed. Our metric has no hyperparameters and is straight-forward to implement. Further, we demonstrate on eight real world document classification data sets, in comparison with seven state-of-the-art baselines, that the WMD metric leads to unprecedented low k-nearest neighbor document classification error rates.[7]

III. RESEARCH AND METHODOLOGY

The propose work is to focus on the hybrid approach of algorithms for Subjective Answer Evaluation. Proposed model consists of two phases i.e., Pre-processing and Evaluation. In the proposed model various NLP algorithms and some inbuilt NLP functions are going to use, NLP for classification of feature extraction to classify the text. The first section of the proposed model is consisting of conversion, stripping and preprocessing of the text. The very first step is selecting the testing text file. After getting the text extract the required data.

The second section is consisting of Evaluation. There are many features available for subjective text evaluation but this system will be concentrating on cosine similarity, contradiction algorithm, semantic similarity, conceptual similarity, keyword matching techniques etc. After getting all tasks done our system will evaluate the answer text and give results.

Step 1: Select the Student Answer key text file in one tab. Step 2: Select the Master Answer key text file in another tab. Step 3: Pre-processing of both the sets is initiated.
Step 4: Conversion of text into lowercase.
Step 5: Tokenization of words.

Step 6: Removal of stop words.
Step 7: Stripping of words.

Step 8: Pre-processing of both sets is ended.

Step 9: Evaluation model is initiated.

Step10: Cosine similarity, Contradiction algo, Semantic Similarity algo, Conceptual Similarity algo, Keyword matching, Jacquard similarity algorithm are executed one by one.

Step11: Grammar Check API is executed for test text file.

Step12: The length of test text is measured.

Step13: According to the Pre-processing and Evaluation the result marks of student are generated.

Step14: Finally, result is displayed.

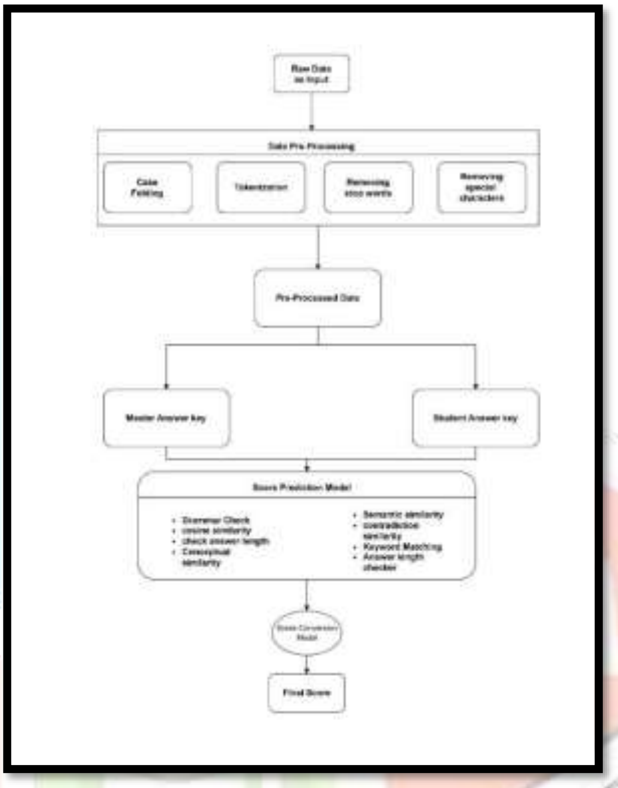

Figure No.1.0 System Architecture

IV. RESULTS AND DISCUSSION

The manual evaluation system of subjective answers for technical subjects can be a time-consuming and tedious process for evaluators. Evaluating subjective responses is challenging and can be influenced by human emotions, resulting in varied evaluation quality. To address this issue, the proposed system uses machine learning and NLP techniques to evaluate subjective answers more efficiently and with reduced human error. The system performs tasks such as grammar checks, answer length checks, keywordextraction, conceptual, semantic, and contradiction similarity, and cosine similarity.

The proposed system's advantage is that it can provide more consistent and unbiased evaluation results with the accuracy of 90.3% by, freeing up time and effort for teachers to focus on other academic endeavors. Additionally, the system calculates scoresand provides results quickly. Previous related work has explored semi-automatic evaluation techniques by creating question and answer bases and mapping evaluated answers using hash indexing with the accuracy of 88%. These techniques consider the semantic meaning and lengthof the sentence to evaluate the student's answer. Overall, the proposed system offers a novel approach to evaluating subjective answers using machine learning and NLP, whichcan save time, reduce human error, and provide more consistent and unbiased evaluation results.

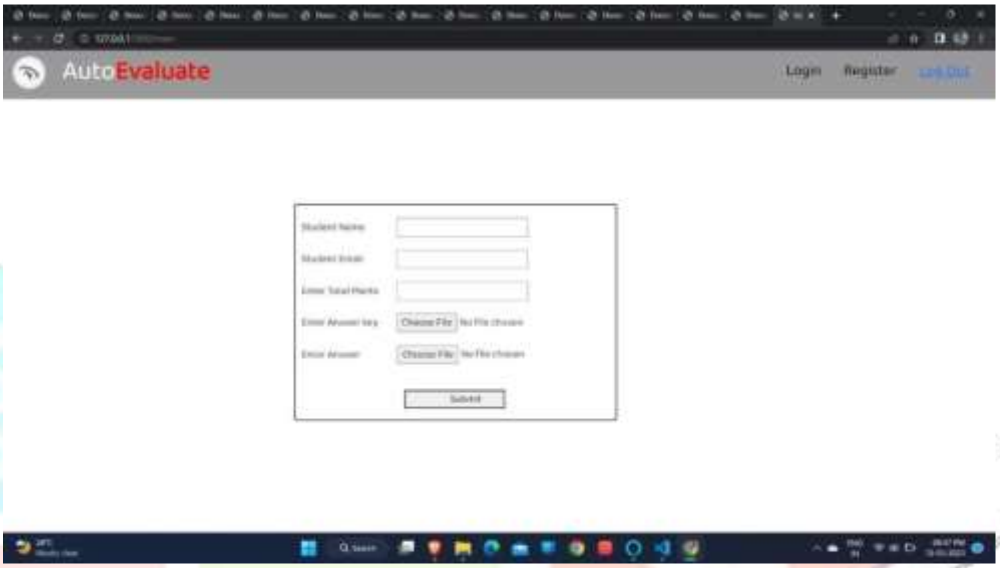

Fig.1.1 Output

Fig. 1.2 Output

V. MATHEMATICAL MODEL

PRE-PROCESSING

Preprocessing in Natural Language Processing (NLP) involves the steps taken to clean and prepare text data before it can be used foranalysis or modeling. This includes tasks such as tokenization, stop word removal, stemming/lemmatization, normalization etc.

Pre-process(text) = stripped

Where, **text** is the input string and stripped is the list of preprocessed tokens returned by the function.
The text preprocessing steps performed by this function can be expressed mathematically as follows:

Step1: Case_folding:text_lower=text.lower()

Step2: Tokenization:tokens=word_tokenize(text_lower)

Step 3 :Removing stop words: tokens_filtered = [token for token in tokens if token.lower() not in stop_words]
where stop_words is a set of predefined stop words in English.

Step4:Removing punctuation and special characters: table =str.maketrans('', '', string.punctuation), stripped = [token.translate(table) for token in tokens_filtered], and stripped = [token for token in stripped if token.isalpha() or token.isdigit()].

COSINE SIMILARITY ALGORITHM

Cosine similarity is commonly used to measure the similarity between two text documents or sentences.
Given two text documents or sentences, we can represent them as vectors in a high-dimensional space, where each dimension corresponds to a word or term in the vocabulary. The entries in the vector represent the frequency or weight of each word or term inthe respective document or sentence.

To calculate the cosine similarity between two text documents or sentences, we use the following formula,
Step 1: denominator1 = sqrt(reduce(lambda x, y: x + y, map(lambda w: word_dict1[w]**2, word_dict1.keys())))
Step 2: denominator2 = sqrt(reduce(lambda x, y: x + y, map(lambda w: word_dict2[w]**2, word_dict2.keys())))
Step 3 : denominator = denominator1 * denominator2

CONTRADICTION

In NLP, detecting contradictions is important for a variety of tasks, including information extraction, question answering, and text classification. It involves comparing multiple statements or pieces of text to identify inconsistencies or conflicting information, while others rely on knowledge bases or ontologies to detect contradictions. Overall, the detection of contradictions is an important task in NLP, as it helps to ensure the accuracy and consistency of natural language processing applications.

SEMANTIC SIMILARITY

Semantic similarity refers to the degree of relatedness or similarity between two pieces of text based on the meaning of thewords and concepts used in them. It involves understanding the underlying context and meaning of the text, rather than just matching individual words or phrases.

CONCEPTUAL SIMILARITY

Conceptual similarity refers to the degree of similarity between two or more concepts or ideas based on their underlying meaning or essence. It involves understanding the similarities and differences between the key characteristics and attributes of each concept or idea, and comparing them to determine how closely related they are.

KEY-WORD MATCHING

Keyword matching is a technique used to identify the presence of specific keywords or phrases in a given piece of text. It involves comparing the text with a pre-defined list of keywords and phrases to find matches

VI. CONCLUSION

The project titled "Automated Subject Evaluation Using Natural Language Processing" was deeply studied and analyzed for code design and implementation. It was done under the guidance of an experienced project guide. During the project, all current requirements and possibilities were taken into account. It is a platform that provides seamless.

subjective assessment of answers. The system provides an easy way to evaluate the answers written by the student against the solutions provided by the experienced teachers and saves the time of the teachers in manually checking the lengthy subjective answers. The project has a very broad scope for the future. The project can be updated in the near future as soon as there is a requirement for the same as it is very flexible in terms of expansion.

VII. REFERENCES

[1] Farrukh Bashir, Hamza Arshad, Abdul Rehman Javed,Natalia ryvinska, “Subjective Answer Evaluation Using Machine Learning and Natural Language Processing”, 10.1109/ACCESS.2021.3130902, IEEE Access
[2] Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande, “Subjective Answer Evaluation System”,International Journal of Engineering Science and Computing, April 2017 Volume 7 Issue No.4..
[3] Saloni Kadam, Priyanka Tarachandani, Prajakta Vetaln and Charusheela Nehete,”AI Based E-AssessmentSystem”, EasyChair Preprint ,March 18, 2020
[4] Vishal Bhonsle, Priya Sapkal, Dipesh Mukadam, Prof. Vinit Raut, “An Adaptive Approach for Subjective Answer Evaluation” VIVA-Tech International Journal for Research and Innovation Volume 1, Issue 2 (2019).
[5] J. Muangprathub, S. Kajornkasirat, and A.Wanichsombat, “Document plagiarism detection using a new conceptsimilarity in formal concept analysis,”Journal of Applied Mathematics, vol. 2021, 2021.
[6] X. Hu and H. Xia, “Automated assessment system forsubjective questions based on lsi,” in 2010 Third InternationalSymposium on Intelligent InformationTechnology and Security Informatics, pp. 250–254, IEEE,2010

[7] M. Kusner, Y . Sun, N. Kolkin, and K. Weinberger,From word embeddings to document distances,” in Internationalconference on machine learning, pp. 957

[8] C. Xia, T. He,W. Li, Z. Qin, and Z. Zou, “Similarityanalysis of law documents based on word2vec,” in 2019 IEEE 19th International Conference on Software Quality,Reliability and Security Companion (QRS-C), pp. 354– 357, IEEE,2019.

[9] H. Mittal and M. S. Devi, “Subjective evaluation: Acomparison of several statistical techniques,” Appl. Artif. Intell.,vol. 32, no. 1, pp. 85–95, 2018.

[10] A. Ceterone and M. Chang, “Automarking: Automatic assessment of open questions,” in ICALT 2010, 10th IEEE International Conference on Advanced Learning Technologies, Sousse, Tunisia, 5-7 July 2010,pp. 143–147, IEEEComputer Society, 2010