International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Ai Using Automatic Subjective Paper Evaluation

T. Naga Venkata Durga[1], G. Siva[2], A.M.N.S. Rajesh[3], K. Sampangi[4], K.Manikanta[5], M. Pujitha[6]

[1]assistant Professor, Cse Department,Vsm College Of Engineering, Ramachandrapuram, Dr.B.R. Ambedkar Konaseema Dt, Andhra Pradesh-533255
[2,3,4,5,6]students Cse Department,Vsm College Of Engineering, Ramachandrapuram, Dr.B.R. Ambedkar Konaseema Dt, Andhra Pradesh-533255

ABSTRACT
Evaluating subjective papers by hand is a difficult and time-consuming undertaking. When utilising artificial intelligence (AI) to analyse subjective articles, there are important issues related to inadequate comprehension and acceptance of findings. There have been several attempts to use computer science to grade students' responses. To do this, the majority of the job, however, makes use of conventional counts or certain terms. Additionally, vetted data sets are also lacking. In order to automatically evaluate descriptive responses, this paper suggests a novel method that makes use of a variety of machine learning, natural language processing, and toolkits, including Wordnet, Word2vec, word mover's distance (WMD), cosine similarity, multinomial naive bayes (MNB), and term frequency-inverse document frequency (TF-IDF). Answers are assessed using keywords and solution statements, as well as a machine learning algorithm.

I INTRODUCTION

Subjective questions demand lengthy, time-consuming answers, making evaluation challenging. Computers struggle due to language ambiguity, necessitating preprocessing steps like data cleaning and tokenization. Various techniques like document similarity and concept graphs are used for evaluation, but room for improvement exists. Subjective exams are daunting due to contextual analysis required for scoring, making automation desirable. Unlike objective questions, subjective ones vary in length and vocabulary usage, posing a challenge for machine evaluation. Despite efforts, issues like semantic context loss and costly training persist. This paper proposes a machine learning and NLP approach for subjective answer evaluation, employing techniques like TF-IDF and cosine similarity. Limitations include synonyms, varying answer lengths, and random sentence order in existing studies. The proposed method involves evaluating answers based on provided keywords and solution similarity, followed by model training for autonomous evaluation.

.

II LITERATURE REVIEW

Various approaches have been proposed for the evaluation of subjective questions in the literature. Hu and Xia utilized Latent Semantic Indexing (LSI) to assess subjective questions, effectively reducing issues with synonyms and polysemy, achieving a 5% difference in grading compared to teacher assessments. Kusner et al. introduced Word Mover's Distance (WMD) as a novel concept for measuring text dissimilarity, resulting in reduced error rates and faster classification speeds, showing improvements

of 2 to 5 times. Kim et al. proposed a method based on lexico-semantic patterns (LSP) for grading short descriptive answers, demonstrating superior performance over existing systems by 0.137. Oghbaie and Zanjireh developed a pair-wise document similarity measure (PDSM), outperforming other measures by
0.08 recall. Orkphol and Yang utilized word2vec and cosine similarity for sentence similarity assessment, achieving a performance of 50.9% with the inclusion of probability of sense distribution. Additionally, Xia et al. combined word2vec with legal documents, enhancing accuracy by
0.2 over Bag of Words, potentially further improving with specialized training. Wagh and Anand proposed a multi-criteria decision-making approach for legal document similarity, obtaining an F1-score of up to 0.8, surpassing traditional methods like TF-IDF. Alian and Awajan explored various factors affecting sentence similarity using different embedding models, achieving 0.87 recall and 0.782 precision with pre-trained embeddings. Muangprathub et al. introduced plagiarism detection using formal concept analysis (FCA), achieving an impressive 94% accuracy. Additionally, Jain and Lobiyal proposed a novel approach based on concept graphs for subjective question evaluation, while Montes et al. discussed techniques for similarity in concept graphs and information retrieval. Bahel and Thomas presented an architecture for subjective question evaluation, although it showed a slightly higher error compared to Jaccard's similarity approach, it struggled with non-textual data.

III EXISTING SYSTEM

The idea of automatically evaluating responses is not new; in fact, it has been researched and developed for nearly 20 years. procedures in place based on evaluation of objective answers. Numerous studies have been conducted on the subject of evaluating subjective answers in one way or another. These studies have included comparing texts, words, and even documents for similarity, identifying the context of a text and mapping it to the context of a solution, counting noun-phrases in documents, matching keywords in answers, and more. We investigate a method for evaluating subjective responses that is based on machine learning and natural language processing. Tokenization, lemmatization, text representation techniques like TF-IDF, Bag of Words, and word2vec, and similarity measuring techniques like cosine form the foundation of our work in natural language processing.

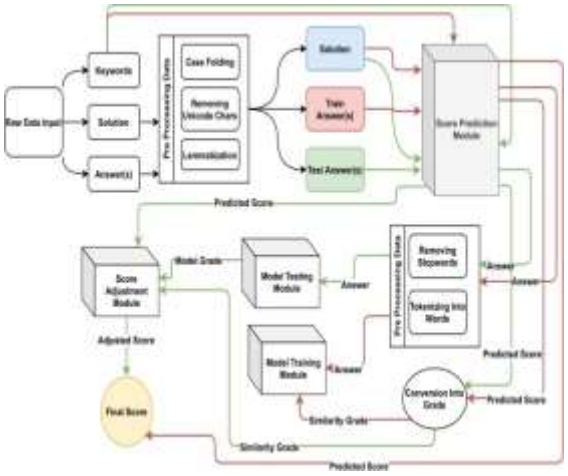
IV PROPOSED SYSTEM

This research suggests an automated method for assessing descriptive question replies that makes use of natural language processing and machine learning. It approaches the problem's solution in two steps. Initially, the solutions and supplied keywords are used to analyze the answers using a variety of similarity-based methodologies, including word mover's distance.The outcomes of this stage are then used to train a model so that it can assess responses without the assistance of keywords or solutions. For instance, the following is a valid response to the subjective question, "What is the capital city of Pakistan and what is it famous for?": "Islamabad is the capital city of Pakistan and it is famous for mountain scenery."
Prior to assessing the student's response to the inquiry, the inquiry, the response, and The system receives a few keywords that are necessary for the answer (in this case, "Islamabad" and "mountain scenery"). It then compares the student's response to the modal answer, taking the context into consideration, and determines whether or not any keywords are present. In other words, a student who answers, "Karachi is the capital of Pakistan and is famous for mountain scenery," may receive 50% of

the possible points; "Islamabad and mountain scenery" may receive 30% since the main keywords are present even though the context is missing; and "Islamabad is the capital and its famous for mountain scenery" may receive 100% since it corresponds to the correct answer and meets the requirements for both contextual similarities and keyword presence.

V SYSTEM ARCHITECTURE

The conceptual design of a software system, including all of its elements, structures, and interactions, is referred to as system architecture. It entails specifying the structure, capabilities, and actions of the system to satisfy predetermined standards. A basic system architecture is shown here:



FULL NATURAL LANGUAGE PROCESSING

In order to parse the text and determine its semantic meaning, these strategies make use of natural language tools . The ultimate score can then be determined by comparing that meaning with the meaning derived from the solution. Preprocessing, the process of processing text documents to prepare them for machine treating, includes a number of natural language techniques including tokenization, stopword removal, parts of speech tagging, lemmatization, stemming, and case folding. Below is a quick explanation of a few of these methods. Tree similarity was measured using Natural Language Processing.

A TOKENIZATION

The practice of breaking up data into smaller units, such words, sentences, paragraphs, and characters, is known as tokenization.

B STOPWORDS REMOVAL

The vocabulary of Natural Language is very large, and the majority of its components are used to make things easier for people to understand, such "the," "on," "is," "in," and so forth. These words don't really play any function in the majority of machine learning tasks

C PARTS OF SPEECH TAGGING

The process of assigning a word in the data to its associated part of speech, such as an adjective, verb, adverb, or noun, is known as parts of speech tagging.

D LEMMATIZATION

Natural language words can take on a variety of forms, including as several tense variations. For

instance, the terms "go," "going," Although they all derive from the same root word, "go," "went" has distinct forms. Lemmatization is the technique of eliminating every word to their original forms in the dataset.

E CASE FOLDING

Words in the natural language frequently duplicate each other depending on their case, appearing in multiple cases. To ensure that the computer can understand every word the same way, it is customary to reduce all the data to the same case, which is often lower case.

F RESULT PREDICTING MODULE

We now know the total score that our module determined by taking into account the maximum matched solution answer sentence pairs and applying either WDM or Cosine Similarity. This outcome can be fed into a machine learning model to be trained, or it can be compared to an actual score.

G COSINE SIMILARITY

Cosine similarity is a measure of Similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. A cosine angle between two vectors is measured, and its value lies between 0 and 1, 1 representing a full match.

H WORD MOVER'S DISTANCE

The minimum distance that the embedded words of one document must "travel" in order to reach the embedded words of another document
is known as the WMD distance, and it is used to quantify the dissimilarity of two text documents.

VI FUTURE RECOMMENDATIONS

Camera Access: Integrate camera access functionality to allow students to upload handwritten answers or diagrams, expanding the types of questions that can be evaluated automatically.

Time Limit: Implement a time limit feature for answering questions, encouraging students to manage their time effectively during exams and providing a more realistic evaluation environment.

Consider Every Correct Answer: Instead of relying solely on trained answers, explore techniques for dynamically identifying correct answers based on patterns and similarities in student responses. This would make the evaluation system more flexible and adaptable to diverse student responses.

Adaptive Scoring: Implement adaptive scoring mechanisms that adjust scoring criteria based on the difficulty level of questions or the overall performance of students in a particular exam.

VII CONCLUSION

This study presented a novel method for evaluating subjective responses that is based on methods from natural language processing and machine learning. There are two suggested score prediction systems that yield scores more accurate. Different criteria of similarity and dissimilarity are examined, and other metrics like the existence of the keyword and the
percentage mapping of sentences are employed to address the anomalous instances of semantically imprecise responses. The findings of the experiment indicate that, on average, the word2vec approach outperforms conventional word embedding techniques because it preserves semantics.

REFERENCES

1. J. Wang and Y. Dong, ‘‘Measurement of textsimilarity: A survey,’’Information, vol. 11, no. 9, p. 421, Aug. 2020.

2. M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, ‘‘A survey on the techniques, applications, and performance of short text semantic similarity,’’ Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021.
3. P. Patil, S. Patil, V. Miniyar, and A. Bandal, ‘‘Subjective answer evaluation using machine learning,’’ Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 1–13, 2018.