IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"SUBJECTIVE ANSWER EVALUATION USING ML AND NLP"

D.R.Gatne[1] ,Samiksha Gangurde[2], Yamini Rathod[3], Yukta Dhake[4], Shreya Guraddi[5]
Students, Maratha Vidya Prasarak Samaj's, Karmaveer Baburao Ganpatrao Thakare College Of Engineering, Udoji Maratha Boarding Campus, Gangapur Road, Nashik, Maharashtra, India.

Abstract: Evaluating subjective answers presents significant challenges for teachers. They must meticulously examine each word to assign scores,considering factors such as their own mental state, fatigue, and objectivity.While machines can easily assess Subjective answers by matching them with concise responses provided by students.Addressing these Challenges of assessment ,this system is being proposed.A previous proposed system for subjective answer evaluation using ML incorporates advanced NLP techniques. The system addresses challenges like handling ambiguity, domain adaptation, and bias mitigation offering an automated approach for fair subjective answer assessments.The proposed system have different modules like Preprocessing Module,Evaluation Module,Result Predicting Module and Similarity Measurement Module.The answers will be checked on basis of the Answer length, Keyword Check, Grammar check,Context and Semantic Similarity and Cosine Similarity against model answer. The software will take copy of the answer as an input and then after the preprocessing step it will extract the test of the answer. This text will again go through processing to build a model of keywords and feature sets. Model answer sets and keywords categorized as mentioned will be the input as well. Classifier will then, based on the training will give marks to the answers will be the final output.

Keywords: NLP, Contextual similarities, Semantic Analysis,ML,WordNet,Cosine similarity.

1.INTRODUCTION

The assessment of students' performance and abilities through subjective questions and answers is a critical area of interest in education.Unlike objective questions, subjective answers allow students the freedom to express their thoughts and understanding without being restricted by constraints. These answers tend to be longer, carrying richer context and energy, distinguishing them from their objective counterparts. However, evaluating subjective answers presents significant challenges for teachers. They must meticulously examine each word to assign scores, considering factors such as their own mental state, fatigue, and objectivity. Consequently, subjective exams become complex and intimidating for both students and teachers. Given the complexities involved, it is imperative to develop an efficient system for evaluating subjective answers. While machines can easily assess objective answers by matching them with concise one-word responses provided by students, subjective answers pose greater difficulties. They exhibit variations in length and encompass a wide range of vocabulary.Moreover, individuals often employ synonyms and convenient abbreviations, further complicating the evaluation process.

Addressing these challenges through the development of effective systems and techniques for evaluating subjective answers is of utmost importance in the realm of education and assessment. Such advancements have the potential to enhance the accuracy, fairness, and efficiency of evaluating students' understanding and performance in open-ended assessments. In this paper, we explore a machine learning and natural language processing-based approach for subjective answers evaluation. Our work is based on natural languages processing techniques such as tokenization, lemmatization, text representing techniques such as TF-IDF, Bag of Words, word2vec, similarity measuring techniques such as cosine similarity, and word mover's distance, classification techniques such as multinomial Naive Bayes. We use different evaluation measures such as F1-score, Accuracy, and Recall to evaluate the performance of various models against each other. We also discuss various techniques used in the past for subjective answers evaluation or text similarity evaluation in general.

2.METHODOLOGY

Data Collection : To train and test the proposed model, there is a need for a huge amount of subjective question answers, but there is no publicly available labeled subjective question answers to the best of our knowledge. In this work, we created a dataset of 400 Questions and Answers with respective marks. For generating dataset, the important thing is to target those websites and blogs where subjective questions and answers exist. Sklearn model is used for splitting data into training and testing data.

Data Annotation: Data annotaion is a vital step in natural language processing and text analysis, involving the removal of irrelevant or unwanted information from text data. This includes tasks like eliminating punctuation, converting all text to lowercase, and removing any other noise or special characters that could hinder analysis. Keywords help decide whether a student has mentioned relevant information in their subjective answers or not.

Word embedding: Word embedding is a technique that transforms words into numerical vectors, enabling mathematical operations and facilitating machine learning models in processing text data. Text documents need to be processed and made ready for the machine; this step is called preprocessing and involves various natural language techniques such as Tokenization, Stopword Removal, Parts of Speech Tagging, Lemmatization, Stemming, Case Folding. Tokenization is the process of separating data into smaller parts, such as paragraphs, sentences, words, and characters. In this work, we tokenize data into sentences and words based on white spaces and period signs.

Cosine Similarity: Cosine similarity is a widely used measure in natural language processing and text analysis to determine the similarity between two vectors. It quantifies the cosine of the angle between the vectors, with a value of 1 indicating complete similarity and a value of 0 indicating no similarity. Cosine similarity is particularly useful when comparing text documents or word embeddings, as it considers the direction rather than the magnitude of the vectors, making it robust to differences in vector lengths. The jacquard similarity is the ratio of intersection to union regarding mutual words. It finds the union of two texts and finds their intersecting terms

Similarity Score: The similarity score quantifies the degree of similarity between two pieces of text by utilizing cosine similarity on their respective word embeddings. By comparing the angles between the vectors, the similarity score indicates the level of similarity between the texts. A higher similarity score suggests a greater resemblance between the two pieces of text, while a lower score signifies less similarity.

3.LITERATURE SURVEY

3.1 Subjective Answer Evaluation System : Exams and assignments play a crucial role in determining the overall academic performance of the students and foster their cognitive learning. This paper provides an outlook to test the degree ofstudent's learning, by evaluating their answers. Our system uses concepts of natural language processing and Machine learning to achieve the goal. The proposed system uses the techniques of natural language processing for preprocessing the text and then using machine learning algorithms for evaluation of the text and assigning the accurate grades. The developed system is comprised of three stages Preprocessing, Semantic Extraction, Classification and Grading. The proposed dataset is taken as the input for the training phase. The preprocessing phase converts the answers in the dataset into a set of index to glove vectors corresponding to each sentence

3.2 ASSESS :Automated Subjective Answer Evaluation using Semantic learning: In this paper, we propose 'ASSESS', a system where the evaluation of subjective answers for an examination becomes easier and convenient. In this paper, we directed our research to propose a system that gives features like full-length subjective tests, automated subjective answer evaluation using natural language processing and semantic learning, auto-generated feedback for self-improvement of the students, visual statistics for both teacher and student after each test, text-to-speech speech-to-text accessibility options and a fully functional hands-free mode for the specially-abled students with disabilities like sluggish typing, poor eyesight, and amputated hands. Since everything will be automated from the evaluation of the answers to providing feedback, there will be minimal stress on the assessors.

3.3 Machine Learning based Automatic Answer Checker Imitating Human Way of Answer Checking: Competitive exams are usually of mcq types and due to this they need to be conducted on computer screens as well as evaluated on them. But apart from competitive exams , computers cannot be used to carry out subjective exams like boards exam . This brings in the need of Artificial Intelligence in our online exam systems. If artificial intelligence gets implemented in online exam conduction systems , then it will be a great help in checking subjective answers as well. Another advantage of this would be the speed and accuracy with which the results of the exams would be produced. Our proposed system would be designed in such a way that it will give marks in a similar way as of a human. This system will hence be of great use to educational institutions .

3.4 Intelligent Short Answer Assessment using Machine Learning: Handwritten text images are processed through OCR to convert them into machine-readable text. This text is then fed into the evaluation module, which assesses it based on sentence grammar, the number of matched keywords, and provides a mark as the output. The proposed system is designed to evaluate answers for five students providing five different answers. The standard answer is stored in the database with the keywords, meaning and the description of that answer. Then each answer is evaluated by matching the keywords as well as its synonyms with the standard answer. It will also check the grammar and spellings of the words. After the evaluation, the answer is graded depending on the correctness of it.

3.5 Online Subjective Answer Verifying System using NLP: This automation of descriptive answer evaluation process would be helpful for various universities and academic institution to efficiently handle the assessment of exam answer sheets of students. Our objective is to design a Subjective Answer Evaluation Model for the automatic evaluation of the multiple sentence subjective answer. This paper provides an outlook to test the degree ofstudent's learning, by evaluating their answers. Our system uses concepts of natural language processing and Machine learning to achieve the goal. The proposed system uses the techniques of natural language processing for preprocessing the text and then using machine learning algorithms for evaluation of the text and assigning the accurate grades

4.PROPOSED SYSTEM

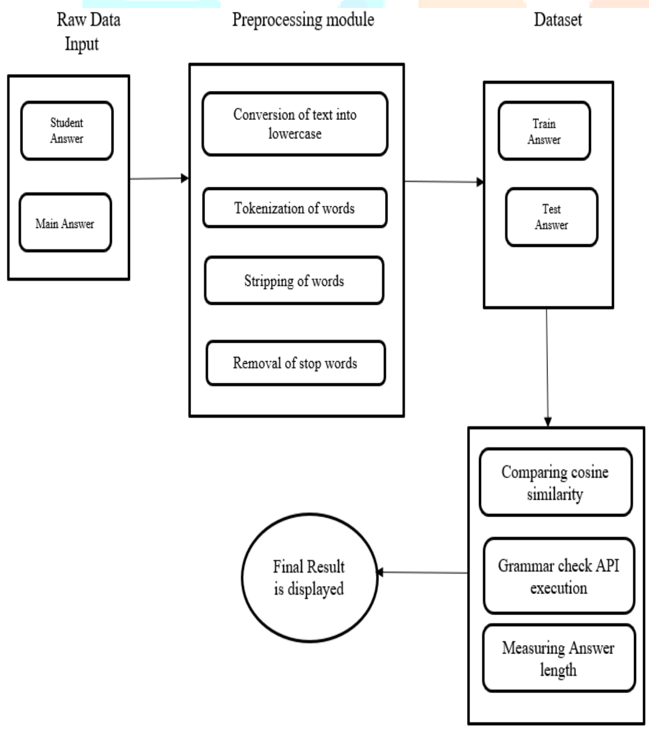

Fig-1:System Architecture

The system architecture for evaluating subjective answers integrates machine learning and natural language processing (NLP) techniques to ensure accurate assessments. The architecture begins with raw input data consisting of the main answer and the student's answer.

A.PREPROCESSING MODULE

After taking inputs from the user, both the solution and the answer go through some preprocessing steps, which involve tokenization, stemming, lemmatization, stop words removal, case folding, finding, and attaching synonyms to the text. However, stop words are removed before passing to a machine learning model such as Multinomial Naive Bayes because they hinder the machine's ability to learn the patterns.

B. EVALUATION MODULE

These preprocessing steps prepare the data for further analysis.This dataset is trained and tested with using Sklearn model for splitting dataset into training,testing and validation.By transforming the data into word embeddings, the architecture facilitates meaningful comparison and evaluation. The preprocessed and word-embedded data is stored for easy retrieval and comparison during the evaluation process. When evaluating the student's answer, the architecture utilizes cosine similarity, a metric that measures the similarity between two vectors. In this case, the vectors represent the student's answer and the main answer. A higher cosine similarity score indicates a closer match between the two answers.

C. SIMILARITY MEASUREMENT MODULE

This module consists of WDM and Cosine Similarity functions which take two sentences or word vectors and return their Similarity. WDM tells us the dissimilarity while Cosine Similarity measures Similarity. Our approach uses both of these similarity measures one at a time and compares the results at the end.

D. RESULT PREDICTING MODULE

Result Predicting Module is the core of this work. We will have the overall score calculated by our module using either WDM or Cosine Similarity while considering the maximum matched solution/answer sentence pairs. This result can be compared to an actual score or fed into a machine learning model to be trained.

5.RESULTS

The proposed system's advantage is that it can provide more consistent and unbiased evaluation results with the accuracy of 83 % by, freeing up time and effort for teachers to focus on other academic endeavors. Additionally, the system calculates score sand provides results quickly.

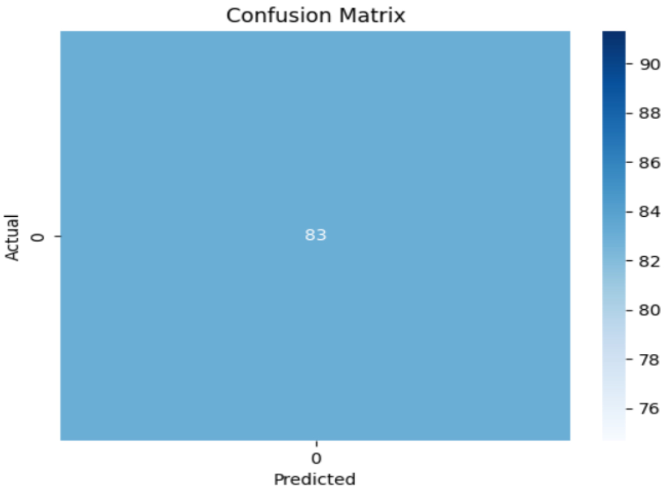

Fig-2:Confusion Matrix

Train data was used to calculate initial scores from the score prediction modules and train the machine learning model. Afterward, testing data was fed to the system one by one, updating the machine learning model.

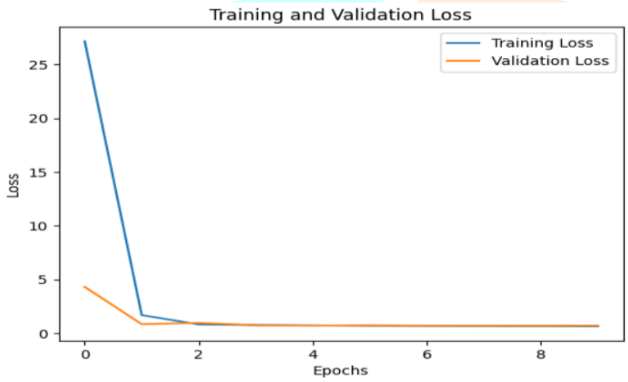

Fig-3:Training and Validation metrics

The results are obtained using cosine similarity and word mover's distance combined with a Naive Bayes model. Both the approaches with and without the model produced results in under a minute at Google Colab. The results are as follows. The score prediction module is working fairly accurately, achieving 83% accuracy. This much accuracy is significant because of word2vec in this case, and it can capture the semantic meaning of answers so well that it gives us very well Similarity among answers. Furthermore, if word2vec lacks inconsistent answers, keyword mapping and unmapped sentences thresholds still give a satisfactory score to the answers

6.CONCLUSION

Subjective Answer evaluation system,the proposed system works with the same factors which an actual human being considers while evaluation such as length of the answer, presence of keywords, and context of key-words.During the project all current requirements and possibilities were taken into account. 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.Further improvement by taking feedback from all the stakeholders such as students and teachers can improve the system meticulously.

7.ACKNOWLEDGEMENT

We express our heartfelt gratitude to all those who contributed to the realization of the "Subjective Answer Evaluation using Ml and Nlp" project. This initiative would not have been possible without the dedication, expertise, and collaborative efforts of our guide Ms. D. R. Gatne. This project stands as a testament to the collective effort and dedication of everyone involved. We look forward to the positive impact Online Evaluation system will have on the Subjective Anwers Grading System.

8.REFERENCES

[1]J. Wang and Y. Dong, "Measurement of text similarity: A survey," Inf., vol. 11, no. 9, p. 421, 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," Concurr. Comput. Pract. Exp., vol. 33, no. 5, 2021.

[3]M. S. M. Patil and M. S. Patil, "Evaluating student descriptive answers using natural language processing," International Journal of Engineering Research Technology (IJERT), vol. 3, no. 3, pp. 1716–1718, 2014.

[4]P. Patil, S. Patil, V. Miniyar, and A. Bandal, "Subjective answer evaluation using machine learning," International Journal of Pure and Applied Mathematics, vol. 118, no. 24, pp. 1–13, 2018.

[5]J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat, "Document plagiarism detection using a new concept similarity in formal concept analysis," Journal of Applied Mathematics, vol. 2021, 2021.