Real-time Evaluation of Descriptive Answer Using NLP and Machine Learning

Chetan Bapu Sonawane
Computer Engineering
MET's Institute Of Engineering Nashik, India
sonawanechetan2002@gmail.com

Omkar Bhausaheb Dighe
Computer Engineering
MET's Institute Of Engineering Nashik, India
omkardighe291@gmail.com

Om Shivaji Kshirsagar
Computer Engineering
MET's Institute Of Engineering Nashik, India
omkshirsagar055@gmail.com

Shrutika Shrikant Khairnar
Computer Engineering
MET's Institute Of Engineering Nashik, India
shrutikakhairnar999@gmail.com

Abstract: The process of evaluating descriptive answers to gauge student performance poses significant challenges. Typically, this assessment is conducted manually, which can introduce subjective biases. The outcome may be influenced by factors such as the authority's nature, mood, and the student-teacher relationship. Furthermore, manual analysis is time-consuming and demanding. To address these issues, this proposed system aims to automate the analysis process by leveraging technologies such as Machine Learning, Natural Language Processing (NLP), and Deep Linguistic Analysis (DLA). By employing these techniques, we can extract pertinent information from textual answers and measure the similarities between the extracted summaries and the correct answers. This automated approach eliminates the need for manual evaluation and allows for efficient scoring. In addition to traditional techniques, this system will incorporate advanced methods like the use of Sentence Transformers. Sentence Transformers are models that encode sentences or text snippets into fixed-dimensional vectors, enabling accurate semantic similarity comparisons. By integrating Sentence Transformers into our framework, we can enhance the accuracy and effectiveness of the analysis process. When evaluating student answers, the system considers various factors, including spelling errors, grammatical errors, and diverse similarity measures, to determine the scoring marks. By considering multiple aspects, we aim to provide a comprehensive and fair assessment. Language processing, a subset of artificial intelligence, focuses on enabling effective communication and interaction between human languages and machines. By leveraging advancements in this field, we can develop a robust and efficient system for evaluating descriptive.

.Keywords: Machine Learning, Deep Linguistic Analysis, Sentence Transformer, Natural Language Processing

I. INTRODUCTION

In the current education system, entrance exams in various fields predominantly rely on objective tests. However, objective tests alone may not be sufficient to accurately assess students' knowledge. Students

are solely evaluated based on the answers they have marked, which can lead to two scenarios. Either the marked answer is genuinely known by the student, or it is merely an assumed answer. In such cases, it becomes challenging to determine whether a student's success is a result of true intelligence or mere luck. Some students may have a partial understanding of the topic but lack comprehensive knowledge. In these situations, descriptive answers play a vital role in evaluating the students' depth of knowledge.

However, evaluating descriptive answers manually poses a significant burden on faculty members, given the variations in students' responses. To address this challenge, we propose an online evaluation system for descriptive-type answers. The proposed approach incorporates the concept of semantic similarity to extract meaning from the diverse range of student answers.

Under this approach, faculty members are required to provide their answers, including a set of compulsory expected keywords. To streamline the analysis, the student answers undergo processes such as pruning and stemming to reduce their size. Subsequently, they are converted into vectors and matrices for further analysis. The marking process is based on the presence and relevance of the expected keywords within the answers.

To support the analysis, the system extracts mandatory text from a database using techniques such as Term Frequency/Inverse Document Frequency (TF/IDF). To measure similarity effectively, Sentence Transformers are employed as a powerful technique within this framework. Sentence Transformers encode sentences or text snippets into fixed-dimensional vectors, allowing for accurate semantic similarity comparisons. By incorporating Sentence Transformers, the system enhances the accuracy and effectiveness of the semantic analysis process.

By combining these techniques, the proposed system can provide a comprehensive evaluation of descriptive answers. It alleviates the burden of manual checking for faculty members and offers a more efficient and objective assessment method.

The rest of the paper is organized as follows. Section II discusses the related work that has been conducted on the topic, providing an overview of previous research to establish existing knowledge and identify gaps for the current study.

Section III focuses on the problem statement, clearly defining the specific issue or challenge that the research aims to address.

Section IV, the existing system and its limitations are discussed, highlighting the shortcomings or drawbacks that need to be addressed for further improvement. Section V introduces the proposed system and its architecture, presenting a new approach or system designed to overcome the limitations discussed earlier.

Section VI delves into the methodology employed in the study, providing a detailed explanation of the steps or techniques used for research, including experiments, data collection, and analysis methods. It also discusses the properties of the methodology, such as effectiveness or efficiency.

Section VII, the conclusion is presented, summarizing the main findings and conclusions drawn from the study. This section also highlights the contributions of the research and may suggest potential avenues for future work.

Section VIII lists the references used throughout the paper, providing credibility to the study by acknowledging the sources consulted during the research process.

II. LITERATURE SURVEY

Previous research on evaluating student descriptive answers using Natural Language Processing (NLP) has primarily focused on keyword comparison rather than considering grammatical mistakes. Several professors and researchers have contributed to this area of study, each with their unique approaches and limitations.

Ms. Shweta M. Patil and Prof. Ms. Sonal Patil researched descriptive answer checking using NLP. This research paper discusses the use of Computer Assisted Assessment (CAA) in evaluating student learning. CAA systems have traditionally focused on objective-type questions, but they fail to assess higher-order thinking skills required in descriptive answers. The paper proposes a new system that evaluates descriptive answers by considering the collective meaning of multiple sentences. The system utilizes techniques such as statistical matching, information extraction, and full natural language processing. It aims to provide accurate grading, feedback, and improvement opportunities for students. The proposed system addresses the limitations of existing CAA systems and focuses on evaluating student performance at a higher level of Bloom's taxonomy [1].

Prof. S.P. Raut, Prof. S.D. Chaudhari and their team developed this paper and this paper presents a method for automatic answer script analysis using natural language processing (NLP) techniques. The paper discusses four similarity measures: Cosine, Jaccard, Bigram, and Synonym. Automatic analysis of answer scripts has proven to be beneficial and, in some cases, produces results consistent with manually assigned marks. The paper also highlights the motivation behind the automated analysis, its contributions, and the organization of the paper. It provides background information on subjective answer analysis techniques, including statistical techniques, information extraction techniques, and full natural language processing (NLP) . The technical background section explains various NLP techniques such as tokenization, stop word removal, parts of speech tagging, lemmatization, stemming, and case folding. It also introduces word embedding techniques like Bag of Words (BoW) and Word2Vec. The paper concludes by mentioning the importance of automated evaluation systems and their potential to improve efficiency and reduce human errors in the educational sector [3].

Prof. P. Sinha, Prof. A. Kaul, Prof. S. Bharadia, and Dr. S. Rathi have developed this and this research paper proposes an automated answer evaluation system using machine learning to reduce the time and manpower required for manual evaluation. The system uses OCR, backpropagation algorithm, ReLU, ANN, CNN, RNN, and CRNN. It aims to evaluate theory-based answers by matching keywords and providing equal distribution of marks. The proposed methodology involves scanning the answer sheet, extracting keywords using OCR, and calculating marks based on keyword matching and answer length. The system is compared to manual evaluation and shows promising results with improved efficiency and accuracy [2].

Mr. N. Dave, Mr. H. Mistry, and Mr. J. P. Verma, students of MCA, explored subjective answer checking using the Occurrence matrix. The research paper explores various methods for text comparison and similarity identification, including Latent Semantic Analysis (LSA), TF-

IDF, Cosine Similarity, Euclidean distance, and Manhattan distance. It discusses the use of an occurrence matrix to represent word frequencies and explains preprocessing steps like data extraction, stopword removal, and stemming. The proposed system employs a block diagram and suggests the use of Jortho, a Java API, for spell-checking. It aims to calculate marks for answers, identify spelling mistakes, determine similarity indices, and calculate final marks based on question grading.

The paper outlines the steps for text comparison: selecting the source file, preprocessing it, creating occurrence matrices for the source file and student's answer sheet, and comparing them using distance measurement algorithms such as Euclidean distance, Manhattan distance, and a custom method proposed by the authors.

An experimental case study is presented, demonstrating the process with sample content, occurrence matrices, and percentage similarity calculations using different distance measurement methods [4].

Prof. U. Hasanah, Prof. A.E. Permanasari, Prof. S.S. Kusumawardani, and Prof. F.S. Pribadi conducted a case study on various research conducted in the Information Extraction (IE) era . The research paper discusses the use of automatic grading systems for short-answer essays in education. The authors explore different evaluation models, focusing on information extraction techniques. These techniques involve matching facts between student answers and teacher answers using methods such as parse tree matching, regular expression matching, boolean phrase matching, syntactic pattern matching, syntactic-semantic pattern matching, semantic word matching, and LRS representation matching. The paper presents various studies conducted in the field, describing the data sets used and the text pre-processing techniques applied. It also evaluates the performance of the grading models by comparing them to human raters. The evaluation metrics commonly used include accuracy agreement, kappa, and Pearson correlation. Overall, the paper highlights the advantages and limitations of automatic short-answer grading systems and discusses possible future research directions in this area. The information extraction techniques show promising results in terms of accuracy and efficiency [5].

Prof R. Devika, Prof. S. Vairavasundaram, Prof. C. S. J. Mahenthar, Prof. V. Varadarajan, And Ketan Kotecha conduct a case study on Bert Model . This paper proposes a model called Semkey-BERT for extracting key phrases from Twitter data using a sentence transformer with BERT. Key phrase extraction from Twitter content is a challenging task, and the proposed model aims to improve the performance by leveraging deep learning algorithms and automatic feature extraction. The model uses BERT embedding and three different sentence transformer models to extract key phrases from tweets. The key phrases are then ranked using a rank aggregation method. Experimental results show that the Semkey-BERT model outperforms existing models with an accuracy of 86% [6].

In summary, previous research efforts in evaluating student descriptive answers using NLP have primarily concentrated on keyword comparison, often neglecting the consideration of grammar and semantic meaning. Each study presented different approaches and limitations, highlighting the ongoing challenges in developing comprehensive and accurate automated evaluation systems for subjective answers.

III. PROBLEM DEFINITION

Existing commercial assessment tools available on the market are limited to evaluating objective-type questions such as multiple-choice questions or short one-line free text responses. These tools primarily assess students' knowledge at the lower level of Bloom's taxonomy of educational objectives, failing to capture their performance at higher cognitive levels. Additionally, these systems cannot identify and address spelling and grammatical errors made by students in their responses.

The problem at hand is the need for an automated assessment system that can effectively evaluate descriptive answers and provide accurate grading based on the semantic analysis of the responses. The objective is to overcome the limitations of manual answer checking, which often results in partial marking for answers with similar semantics and context.

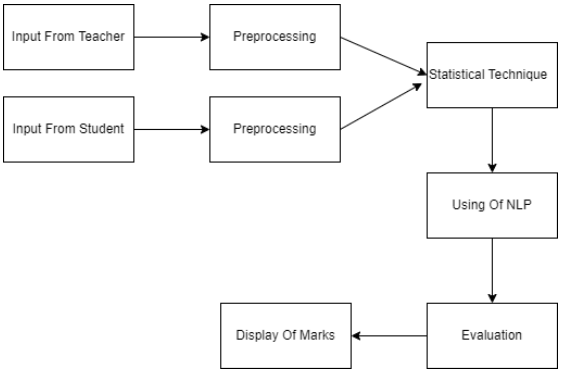
To tackle the issue of manual answer checking which may lead to partial marking for the descriptive answer with the same semantics and context, by building a platform that will understand the descriptive answer and allocate marks according to the comparative semantic analysis.

IV. PROPOSED SYSTEM

The proposed system offers a user-friendly interface for both teachers and students to facilitate the evaluation of descriptive answers. Upon login, teachers have the flexibility to select the questions that will be included in the students' exams. They can easily add new questions or remove existing ones as needed. The system also provides teachers with access to the history records of students' scores, enabling them to track and analyze students' performance over time.

Once students log in, they are presented with the set of questions selected by the teacher for the test. Students can provide their answers in the form of descriptive text. To evaluate these answers, the system utilizes the state-of-the-art machine learning model known as the sentence transformer Bert. This model employs natural language processing techniques to analyze and compare the model answer entered by the teacher with the answer given by the student. Based on this comparison, the system generates a score for each answer, providing immediate feedback on their performance.

The system offers several advantages. Firstly, it streamlines the evaluation process, saving valuable time for teachers who would otherwise have to manually assess and grade each descriptive answer. Additionally, it enhances consistency in grading by utilizing a standardized evaluation model. By storing the test records, the system facilitates easy access to historical data for performance tracking and analysis. Furthermore, it promotes fairness and transparency in the evaluation process by eliminating subjective biases that may arise in manual grading.



5.1. System Architecture of Proposed System

METHODOLOGY

1. Data Pre-processing
2. Data Extraction
3. Natural language processing
4. Sentence Transformer
5. Cosine similarity

Data Pre-processing

Data pre-processing is a vital step in machine learning model development that involves preparing raw data to make it suitable for analysis. In many cases, the data we encounter may be unclean or unstructured. Therefore, it is necessary to clean and format the data before performing any operations on it. This process is known as data pre-processing and is essential for ensuring the quality and reliability of the machine learning model.

Data Extraction

Data extraction is the process of gathering or retrieving various types of data from different sources, even if they are poorly organized or unstructured [5]. The purpose of data extraction is to consolidate, process, and refine the data so that it can be stored in a centralized location for further transformation. These locations can be on-site, in the cloud, or a combination of both. Data extraction enables organizations to effectively manage and utilize their data by making it accessible and usable for analysis and decision-making.

Natural language processing

Natural Language Processing combines computational linguistics with statistical, machine learning, and deep learning models. This interdisciplinary field enables computers to process and understand human language in the form of text or voice data, including the intended meaning and sentiment expressed by the speaker or writer. NLP has various applications, such as text classification, sentiment analysis, language translation [3], and Chabot's, which greatly enhance human-computer interaction and enable automated language-based tasks.

Sentence Transformer

Sentence transformation is the process of altering the structure or form of a sentence while preserving its original meaning. It involves modifying the sentence through various grammatical transformations, such as changing the word order, using different verb forms, or employing different sentence structures. Sentence transformation is commonly used in language tasks such as paraphrasing, summarization, and language generation. It allows for expressing the same idea in different ways, enhancing clarity and variety in communication.

Cosine similarity

Cosine similarity is a metric used to determine the similarity between data objects, regardless of their size. It is particularly useful in measuring the similarity between two sentences. In Python, cosine similarity can be used to calculate the similarity between vectors representing the sentences. By treating data objects in a dataset as vectors, cosine similarity quantifies the similarity

based on the angle between the vectors, providing a numerical measure of their resemblance.

V. CONCLUSION

The proposed system aims to overcome the challenges of an existing system for evaluating descriptive answers. This Proposed system presents the proposed system for evaluating descriptive answers and offers a user-friendly interface for teachers and students, streamlining the evaluation process and providing prompt feedback. Utilizing the sentence transformer Bert machine learning model, it ensures standardized and objective grading. The system's advantages include time-saving for teachers, enhanced consistency in grading, and fair evaluation for students. With the ability to track performance records and provide valuable insights, the system empowers both teachers and students in their educational journey. Overall, the implementation of this system has the potential to revolutionize the evaluation of descriptive answers and improve the learning experience.

VII. REFERENCES

[1] Ms. Shweta M. Patil, Prof. Ms. Sonal Patil March - 2014, "Evaluating Student Descriptive Answers Using Natural Language Processing" International Journal of Engineering Research & Technology (IJERT).

[2] Prof. Prince Sinha, Dr. Sheetal Rathi March - 2018, "Answer Evaluation Using Machine Learning" Thakur College of Engineering and Technology, Mumbai.

[3] Prof. Sumedha P Raut, Siddhesh D Chaudhari March – 2022, "Automatic Evaluation of Descriptive Answers Using NLP and Machine Learning" International Journal of Advanced Research in Science, Communication and Technology (IJARSCT).

[4] Mr. N. Dave, Mr. H. Mistry, Mr. J. P. Verma 2017, "Text Data Analysis: Computer Aided Automated Assessment System" Institute of Technology, Nirma University, Ahmedabad.

[5] Prof. U. Hasanah, Prof. A. E. Permanasari, Prof. F. S. Pribadi, 2016 "A Review of an Information Extraction Technique Approach for Automatic Short Answer Grading", STMIK Amikom Purwokerto, Indonesia.

[6] Prof R. Devika, Prof. S. Vairavasundaram, Prof. C. S. J. Mahenthar, Prof. V. Varadarajan, and Ketan Kotecha, 2021 "A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data" Science and Engineering Research Board (SERB), Department of Science and Technology, India