IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A study on- Evaluation of Theoretical answers using Machine Learning and Generation of Model answer using Generative Artificial Intelligence

Prof. Sampada A. Kulkarni, Reva Joshi, Saloni Ghayal, Sayali Deshpande, Rohan Shinde

Professor, Student ,Student ,Student ,Student

Information Technology ,PES Modern College of Engineering, Pune, Pune, India

Abstract— The paper proposes a unique approach for evaluating theoretical answers using machine learning and generative artificial intelligence. The goal is to assess long, descriptive, and handwritten answers. Since theoretical paper writing reflects a student's ability to express taught concepts, it is crucial to assess them accurately. However, paper checking is time-consuming, and it's observed that each and every answer is not thoroughly reviewed. Factors such as answer length, grammatical errors, and word similarity may be overlooked by human evaluators, resulting in inaccurate marking. To address this, a system is proposed that utilizes modern tools like machine learning and natural language processing. The system takes input from answer sheets and uses Optical Character Recognition (OCR) to convert handwritten text to digital format. It then processes both the submitted answer and a model answer, either provided by the evaluator or generated by a trained chatbot, into numerical formats or vectors. The similarity of the two is then compared using cosine similarity to derive a result. While the concept of such a system has been explored previously, this paper presents a novel approach using different techniques and methodologies.

Index Terms - Artificial intelligence, generative artificial intelligence, machine learning model, chatbot

I. INTRODUCTION

Checking handwritten answer sheets is a crucial but tedious task across all academic courses. Currently, manual evaluation is used for almost all types of studies, regardless of whether they are at the school or college level. Despite providing teachers with a model answer, there is no definitive method for grading papers.

Nearly every institution and university relies on manual evaluation methods to assess student results. However, it is impractical to thoroughly read and evaluate every paper. Therefore, there is a need for a standardized and efficient method for this task. Answers to theoretical questions, such as those worth 2, 3, 5, or 10 marks, vary in length and content.

Given the prevalence of automation in various fields, it can likewise be applied to paper evaluation in colleges affiliated with a university, thus aiding in accurate
student assessment as per the length and content of the answer.

As automation can be seen in every field nowadays, it can be used for paper evaluation in education as well. This system can be implemented on a larger scale for a university to set a common standard of checking for the colleges affiliated with it. So this system can be very useful for proper grading of students' papers.

II..BACKGROUND

When answering questions, students may provide descriptive responses that are required to be evaluated in image format. These responses often contain a range of keywords and may consist of one or more sentences, as per the question and the student's writing style. Such responses tend to contain synonyms more frequently than ideal, so preprocessing is necessary. Text documents must undergo preprocessing to ready them for machine processing. This involves several Natural Language Processing (NLP) techniques, including Case Folding, Tokenization, Lemmatization, Stop-word Removal, stemming, and Parts of Speech (POS) Tagging [2]. Maximum marks should be provided for answers, and the answer length should be kept fixed. Model answers for questions can also be generated using Generative Artificial Intelligence (AI), by specifying conditions like the length and content of the answer and the question. The generated model answer can be reviewed by the faculty for use as the model answer. Alternatively, the model answer and keywords can be manually inputted into the model. Then the similarity between the model answer and the human-written answer sheet can be checked using cosine similarity. The final score can be determined based on language, structure, and the percentage of similarity between the model answers and human-written answers.

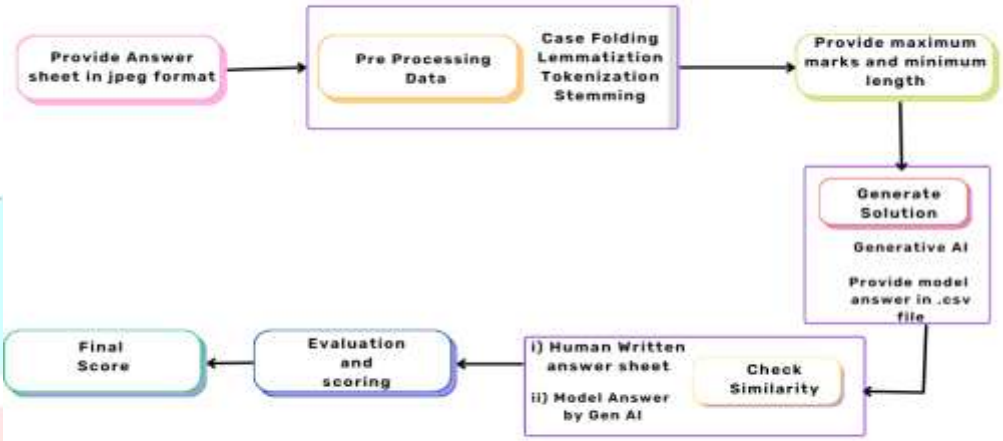

Fig.1.Block Diagram explaning the process

III. LITERATURE REVIEW

The manual approach for evaluating subjective responses for scientific fields requires the evaluator to invest a significant amount of time and resources. However, computers are currently used to evaluate multiple-choice questions. Subjective responses judged on number of factors, according to content and writing style of the inquiry. Evaluation of subjective responses is an essential responsibility. When a human analyses anything, the evaluation's quality may change depending on the person's emotions also.

Here, methods for automatically assessing descriptive answers using various machine learning (ML) techniques and natural language processing (NLP) techniques was proposed by the authors [4]. The study incorporates Word2vec, multinomial naive Bayes (MNB), cosine similarity, Word Mover's Distance (WMD), WordNet, and term frequency – inverse document frequency (TF-IDF) to evaluate answers based on statements and keywords. The approach builds a machine learning system to evaluate answer grades, with WMD demonstrating superior performance compared to cosine similarity. The model achieves 88% accuracy without MNB, while incorporating MNB reduces the error rate by 1.3%.

In this research, the authors [2] introduced methodologies for the automatic evaluation of student descriptive answers, marking a significant advancement in educational assessment methodologies. By combining LDA (Latent Dirichlet Allocation) for thematic coverage with T5(Text-To-Text Transfer Transformer) for semantic understanding, this model provides a flexible and all-encompassing approach to evaluate student responses across different subjects, achieving 91% accuracy, 91% precision, 92% recall, and a 91% F1-score. These outcomes calculate model's efficiency and dependability in assessing student performance.

In the research paper of [3], a web interface is created using Django, machine learning and NLP techniques. In this system, there are three types of logins for students, teachers and admins. According to schedule, to arrange a test, a teacher has to initiate the process by ready to style your paper; use the generating a question paper and giving a model answer. The students are then able to see and attempt the test either by typing answers into the system or submitting their handwritten answer sheets in virtual format. It also uses a pretrained model namely bert-base-nli-mean-tokens by Google as a feature extractor. The score derived by cosine similarity is multiplied by 100 to generate the final score, as cosine similarity ranges from -1 to 1. However, the system is useful for grading short theoretical answers. The system was developed with an intension to solve the dilemma of conducting subjective tests during covid times.

Integrating generative AI to produce model answers in automated subjective grading systems can significantly advance the field. Generative AI can augment the training datasets with diverse and nuanced responses according to the syllabus of universities, enhancing the system's ability to precisely evaluate varied student answers. By providing high-quality, contextually relevant model answers, generative AI can refine the grading process, leading to more precise and reliable evaluations. Generative AI reduce biases in grading by generating a broad spectrum of model responses, thereby enabling the system to fairly evaluate diverse expressions of similar ideas. Its adaptability across different subject areas allows for the creation of tailored model answers, improving the system's generalization capabilities across disciplines. Additionally, the iterative nature of generative AI facilitates continuous improvement of the grading system, ensuring it evolves and remains effective over time.

Moreover, the automation of model answer generation can increase resource efficiency, freeing educators from the labor-intensive task of manually creating examples. This, in turn, enables a greater focus on refining and enhancing the grading algorithms. Overall, using the generative AI in subjective grading research, present great potency for developing more accurate, fair, and adaptable automated grading systems.

IV. PROPOSED METHODOLOGY

The proposed methodology will help us understand detailed information regarding the subject of our study beginning with developing a chatbot for deriving answers to final calculation of the result.

1) **Developing a generative AI model :** A foundational generative artificial intelligence model can be created using python. The first step to create the model is to prepare the data. The essential data regarding to any subject topic should be systematically gathered and processed. This data should be complete, precise and structured, without any duplicates This data will act as a base data file for the model.

2) Training the model : The base data file will be used to train and validate the model. It includes processes like tokenization, building a model card. Then the model will be fine-tuned and ready to be deployed. The model is build to use as a chatbot for the ease of communicating answers to the evaluators.

3) Providing an answer sheet as input: One has to scan the paper, and then the system will split the answer using optical character recognition (OCR) based on the keyword in the answer. The input will be in the form of a PDF(Portable Document Format) containing images of the answer sheet. By using the Pytesseract library of Python, the written text will be converted into digital text and get stored in a new text file. In this step, the handwritten answer was converted into a digital format [9].

4) Extracting and preprocessing: Optical character recognition (OCR) is the electronic or mechanical conversion of images of typed, handwritten, or printed text into machine-encoded text [7]. Natural Language Processing (NLP) was used to clean the extracted text. It also enables computers to comprehend, interpret, and modify human languages such as English or Hindi to study and determine their meaning. Other techniques include cleaning, removing punctuation, tokenization, padding [7]. The machine cannot accurately evaluate results from long answers. Therefore, we have summarized this text. Summarization is the process of creating a short and accurate summary of a longer text. We must implement a deep learning algorithm such as an Artificial Neural Network (ANN) [8].

Tokenization- To break down a text into concepts and their relationships, information extraction approaches rely on extracting a structure or a pattern from the text [8].

Stemming - Every language has some sort of formal grammar and that the vocabulary is created by always keeping those grammatical rules in mind. Stemming is a technique for reducing words to their stems. For instance, stemming terminating characters, stemming plurals into singulars (words into word) and so forth [8].

Lemmatization- Human language words can take on a variety of forms, including many tenses. For instance, the terms ”go,” ”going,” and ”went” all have different forms but come from the same root word, ”go.” This method boils down each term in the dataset to its simplest form [8].

Padding- Naturally there will be sentences which would be different in length. However, all neural networks require to have inputs with the same size. For this purpose, padding is done [8].

Stopword Removal- Natural Language has most of the words such as ”the,” ”in,” ”on,” and ”is”. Most machine learning projects involve little to no use of these phrases, and they may even make the process more difficult. Therefore, their removal becomes necessary for system to evaluate accurately [8].

Case Folding – All the words are converted from upper case to lower case to maintain uniformity in the text [2].

Bag of words and Term Frequency–inverse document frequency – The frequency of words, individually and sentence-wise, is counted and stored for calculating similarity [2].Word2Vec, a neural network, was applied to detect synonyms for keywords [2]. Another essential criterion to be considered when determining whether a student has covered every crucial keypoint in his response is keyword matching. If the keyword matches, the maximum marks are given as per the keyword [8]. All of these preprocessing steps are performed on both the model and input answers.

5) Comparing the answers - When the derived model answer and answers written by student are converted into vectors by embedding, the similarity between them can be easily calculated by taking the dot product. It directly produces a number that denotes the similarity between the two. Answer having contextual similarity and match with keywords will be getting higher marks [2]. The cosine similarity is a most useful method to compare two text documents [2].

6) Calculation of Result - The final marks will be calculated based on the similarity between the model answer and the student's answer. We will use cosine similarity to measure the match of keywords. Depending on the degree of similarity, a specific range of marks will be assigned. The final score will be determined according to these predefined criteria, and a corresponding grade will be given based on the mark range.

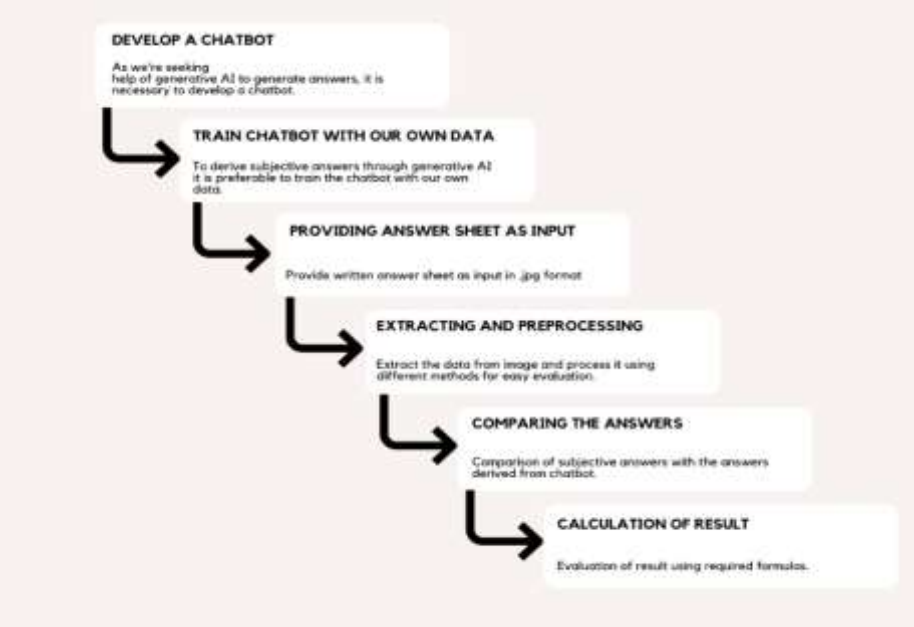

Fig.2. Flowchart explaning the process

V. CONCLUSION

We believe that this system will be more efficient. The advancements outlined in the abovementioned study can be implemented, and the system can be tested with real papers. This method of evaluation has high capability, as it has not been extensively explored. Implementing this system would result in fewer errors in result generation and omit the need for re-evaluating papers. This will help restore students’ trust and faith in paper evaluations and highlight the importance of subjective assessments. It will also help faculty, as it will reduce their duty to check papers and ease the task. As technology continues to advance, the system will have a tremendous scope for growth.

REFERENCES

[1] Sneha G, Shreya Shree G, Tina Babu, Rekha R Nair "Evaluation of Subjective Answers using Machine Learning" 3rd International conference on issues and challenges in intelligent computing techniques (ICICT) 2022

[2] Muhammad Farrukh Bashir, Hamza Arshad Abdul Rehman Javed, Natalia Kryvinska and Shahab S. Band "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing" IEEE Access volume 9, 2021

[3] Rishabh Kothari, Burhanuddin Rangwala, Kush Patel "Automatic Subjective Answer Grading Software using Machine Learning" Proceedings of the 7th International Conference on Trends in Electronics and Informatics (ICOEI 2023)

[4] Lalitha Manasa Chandrapati, Ch. Koteswara Rao "Descriptive Answers Evaluation using Natural Language Processing approaches" IEEE Access volume 11, 2023

[5] Mandada Samemi, Tirumala Sai Hareesha, Gudluru Venkata Siva Sai, Pavan Kumar, Nalluri Pramod, S.nahida "Automatic Answer Evaluation Using Machine Learning" Dogo Rangsang Research Journal UGC Care Group I Journal Vol-09 Issue-01 No. 01 : 2022

[6] Sangeeta Mangesh, Prateek Maheshwari, Aditi Upadhyaya "Subjective Answer Script Evaluation using Natural Language Processing" Journal Of Emerging Technologies And Innovative Research (JETIR) August 2022, Volume 9, Issue 8

[7] Prince Sinha, Sharad Bharadia, Ayush Kaul, Dr. Sheetal Rathi "Answer Evaluation Using Machine Learning" Conference: McGraw-Hill Publications

[8] Mrs. Dr. Meenakshi A.Thalor, Sejal Khopade, Sakshi Shinde, Mayuri Garad, Vipin Kumar Singh, Shreyas Kumbhar "Handwriting to text converter web application" International Research Journal of Modernization in Engineering Technology and Science Volume:05/Issue:05/May-2023

[9] Shreya Saloni Verma, Abdullah Sarguroh, Jyotsana Rawat "Identification of Text Similarity Based On Context" International Research Journal of Engineering and Technology (IRJET) Volume: 08 Issue: 04 Apr 2021

[10] Nicholas Gahman and Vinayak Elangovan "A Comparison of Document Similarity Algorithms" International Journal of Artificial Intelligence and Applications (IJAIA), Vol.14, No.2, March 2023

[11] Gaurang Kudale, 2Nishant Mali, 3Nachiket Suryawanshi, 4Mukesh Bansode, 5 Prof. Richa Agarwal "Automated Subjective Answer Evaluation Using NLP" 2023 IJCRT | Volume 11, Issue 5 May 2023 | ISSN: 2320-2882

[12] Vijay Kumari, Prachi Godbole and Yashvardhan Sharma "Automatic Subjective Answer Evaluation" 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)

[14] Sapana Sachin Baheti "Study of Subjective Answer Evaluation using Natural Language Processing and Machine Learning" International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 4, Issue 4, March 2024

[15] Raghunath Dey, Rakesh Chandra Balabantaray, Surajit Mohanty, Debabrata Singh, Marimuthu Karuppiah, Debabrata Samanta "Approach for Preprocessing in Offline Optical Character Recognition (OCR)" 2022 Interdisciplinary Research in Technology and Management (IRTM).