SUBJECTIVE ANSWERS EVALUATION USING ML & NLP

Prof. Satish C. Cholke[1], Shrutika Holkar[2]

**1 Asst. Professor Department of Information technology, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India.*

**2 student Department of Information technology, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India.*

---***---

Abstract:- Every year, educational boards and universities conduct exams in offline mode, where a large number of students take subjective type exams. Evaluating such a vast number of papers manually requires significant effort. The quality of evaluation can also vary depending on the evaluator's mood, making the process inconsistent. Moreover, the manual evaluation is both lengthy and time-consuming.

On the other hand, competitive and entrance exams typically consist of objective or multiple-choice questions (MCQs). These exams are evaluated using machines, making the process straightforward. This method also conserves resources and minimizes human interaction, resulting in error-free evaluations. While multiple systems are available for evaluating objective (MCQ) type questions, there is a lack of provision for assessing subjective (descriptive) type questions.

Automating the evaluation process for descriptive answers would be extremely beneficial for educational institutions, as it would enable efficient and accurate assessment of students' exam answer sheets.

Key Words: *Subjective type exams, manual evaluation, evaluation consistency, time-consuming, objective type questions, multiple-choice questions (MCQs), automated evaluation, educational institutions, efficient assessment, accurate assessment, error-free evaluations, evaluation systems, descriptive answers, resource conservation, human interaction minimization*

1.INTRODUCTION

The assessment of students' academic performance is a critical component of the educational process, traditionally conducted through examinations that include both subjective and objective questions. With the onset of the Covid-19 pandemic, educational institutions worldwide have shifted towards virtual modes of learning and assessment. In this new scenario, the manual evaluation of subjective answers has become increasingly cumbersome and impractical. questions (MCQs), these systems work by comparing responses to pre-defined correct answers and are primarily useful for competitive or objective-type exams.

Objective examinations, particularly those involving multiple-choice questions (MCQs), benefit from automated evaluation systems that compare student responses to pre-defined correct answers. These systems are highly efficient, providing quick and error-free results, and are extensively used in competitive exams. However, their application is limited to objective assessments.

Subjective examinations remain the cornerstone of university and board-level assessments. These exams require students to provide descriptive answers, which are essential for measuring the depth of their understanding and knowledge acquisition. Moderators evaluate these answers to assign marks, a process that is inherently subjective and influenced by the evaluator's interpretation, mood, and potential biases. This manual evaluation process is not only labor-intensive and time-consuming but also susceptible to inconsistencies, which can affect the fairness and accuracy of the assessment. as hand outlines. Gesture recognition techniques are then utilized to recognize specific hand gestures.

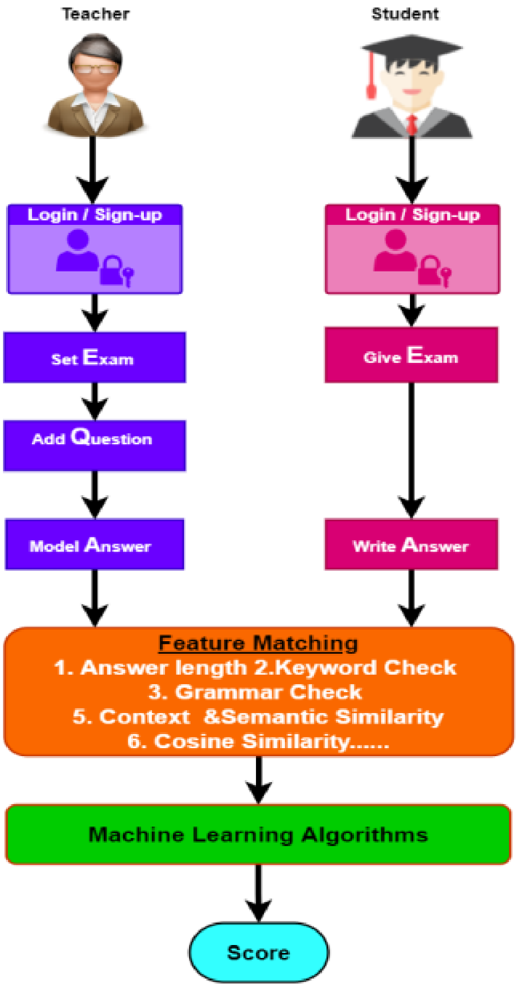
HandVibe addresses the growing demand for intuitive human-computer interaction methods, particularly in scenarios where traditional input modalities such as keyboards or touchscreens may be impractical or cumbersome. Through its innovative approach, HandVibe enhances user experience by offering a natural and intuitive means of interaction, fostering a more seamless and efficient computing environment. The project not only showcases the capabilities of computer vision and machine learning in gesture recognition but also underscores their potential applications in developing user-centric interfaces and interactive systems. HandVibe exemplifies a fusion of advanced technologies aimed at bridging the gap between humans and machines, paving the way for more intuitive and immersive computing experiences.

Given these challenges, there is a pressing need to develop an automated evaluation system for subjective answers. Leveraging advancements in machine learning (ML) and natural language processing (NLP), this research aims to create a system that can reliably and efficiently evaluate descriptive answers. Such a system would standardize the evaluation process, reduce the workload on educators, and ensure a fairer assessment of student performance.

2. LITERATURE SURVEY

1. A model to evaluate subjective Answer papers using Semi Automated Evaluation technique is created. For that first they create Question base which contains question type, sub type, question and marks. Then Answer base is created with model answer. Evaluated answer is mapped using hash index which referred as question number. The student answer is evaluated by considering semantic meaning and length of the sentence.
2. A model to evaluate subjective Answer papers using Semi Automated Evaluation technique is created. For that first they create Question base which contains question type, sub type, question and marks. Then Answer base is created with model answer. Evaluated answer is mapped using hash index which referred as question number. The student answer is evaluated by considering semantic meaning and length of the sentence.
3. -assessment system developed to checks the answer sheet of the student and provides marks to the same. The system consists of an algorithm that compares the student's answer against three reference answers given by three different faculties and the answer with most close results and with highest precision is taken into consideration and marks are allocated accordingly. Algorithm based on TFIDF, Grammar check, WMD, cosine and Jaccard Similarity. Both the answers need not be exactly the same or word to word. This approach can be a quick and easy way for the examiners by reducing their workload.
4. The proposed system receives solution sets from admin and student's answers. Then stop words are removed from them in order to generate keywords. After keyword generation it checks for similarity and by calculating similarity it also checks the relation of the keywords along with sentences with the sentences in the dataset which finds the exact similarity and correctness of the sentence with the datasets. If the sentences match with the datasets it generates similarity score as per the overlapping percentage. It also checks the synonyms and similar words before relating the keywords in order to increase the accuracy of the overlap. Data duplication technique is used to compare the previous answers submitted by students and on the basis of uniqueness of answers, grades are generated.
5.] Developed an algorithm which will evaluate theoretical answers and give marks according to the keyword matching which will reduce manual work and saves time with faster result evaluation. A person should collect the answer copy from the student and scan it. The machine will take the image as input and will evaluate the answer based on the length of the answer and important keywords covered Which are specified by the teacher with each answer which is to be evaluated.
6. The handwritten text image, given as input to the handwriting conversion module is extracted and converted to machine encoded text using Optical Character Recognition (OCR) Algorithm. The machine encoded text is given as input to the evaluation module. Evaluate module evaluates the answer based on grammatical meaning of the sentences; number of keywords matched and gives marks as output.

3.PROPOSED SYSTEM

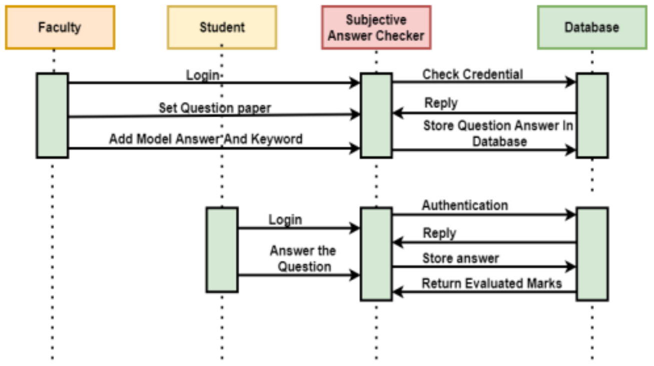


We proposed a system with two sections: one for teachers and one for students. Both teachers and students have login pages, and if not registered, they can create an account. In the teacher section, teachers can set exams by creating a test with a name, date, and time, and add questions with marks and model answers. In the student section, students need to join a class to take the exam, which will then become visible and start and end according to a timer. After completing the exam, students must submit their answers. The answers are then evaluated against the model answers based on criteria such as answer length, keyword presence, grammar, context, semantic similarity, and cosine similarity. These criteria values are processed through a machine learning model to generate accurate scores for each answer. Teachers can manually adjust the marks if the algorithmically assigned scores are deemed inaccurate.

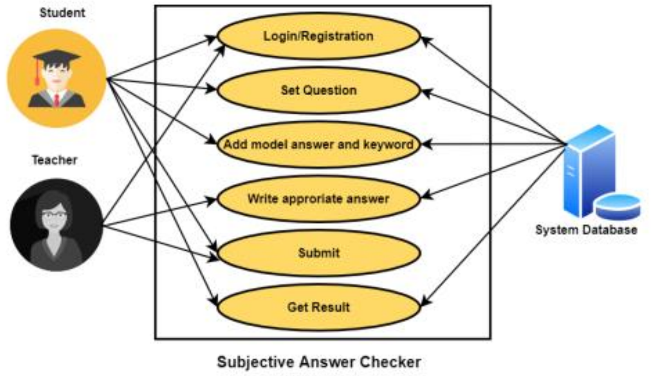
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

3.1 UML DIAGRAM

Sequence Diagram:



Use Case Diagram:



4. NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a field of AI that enables computers to comprehend, interpret, and manipulate human languages like English or Hindi to analyze and derive meaning. NLP assists developers in organizing and structuring knowledge to perform tasks such as translation, summarization, named entity recognition, relationship extraction, speech recognition, and topic segmentation**.**

1. **Stemming:** Stemming is the process of reducing words to their base or root form by removing prefixes and suffixes. This involves mapping a group of words to the same stem, even if the stem itself is not a valid word in the language. Essentially, stemming reduces words to their base form, which may not be a proper lemma.
2. **Lemmatization**: Lemmatization involves analyzing the morphological structure of words to find their proper base form, or lemma, which has a dictionary meaning. Unlike stemming, lemmatization is based on linguistic principles and provides a more accurate root form of the word. Essentially, stemming reduces words to their base form, which may not be a proper lemma.
3. **Stopwords**: Stop words are the most frequently used words in any language. For text analysis and building NLP models, these words are often removed as they do not significantly contribute to the meaning of the document.

5. WEIGHTING MODULE FOR EVALUATION OF ANSWER

5.1 Answer Length: In the current system, students may receive full marks by merely including keywords, regardless of answer length. This makes answer length an important factor, as students could write all keywords and grammatically correct short sentences. Such answers would receive full marks for keywords and grammar but fewer marks for brevity.

5.2 Keyword Matching: Keywords are crucial for evaluating whether a student has covered all key concepts in their answer. If the student's answer contains all the correct keywords, maximum marks are awarded. However, if some keywords differ from those in the model answer, marks are deducted accordingly.

5.3 Contextual Similarity: We have developed algorithms to assess the similarity between two answers. If the answers are not contextually similar, the algorithm assigns a score of 0; otherwise, it provides a percentage of similarity. We reviewed various Python libraries, noting that some yield high similarity ratios even when the answers are not contextually alike.

5.4 Semantic Similarity: For answers that are contextually similar, we measure similarity with respect to contextual meaning using an API.

5.5 Contradiction: Previous systems struggled to detect contradictory sentences because they removed crucial stop words (like "not," "never," "doesn't") during preprocessing and did not account for key antonyms. We use a POS tagger to identify contradictory sentences in student answers compared to the model answer.

5.6 Grammar Check: To identify grammatical mistakes and spelling errors, our system uses the Grammar Text gear API. When text is sent to this API, it returns the number of grammatical errors present.

5.7 Cosine Similarity: Cosine similarity is a standard method to measure the similarity between documents regardless of their size. It is represented as the dot product of two text vectors. For instance, vectors A and B represent the word vectors of the texts being compared.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

6. MACHINE LEARNING ALGORITHMS

6.1 Naïve Bayes: The Naïve Bayes algorithm is a supervised learning method based on Bayes' theorem, commonly used for classification problems. It is particularly effective in text classification tasks involving high-dimensional training datasets. The Naïve Bayes Classifier is one of the simplest and most efficient classification algorithms, enabling the creation of fast machine learning models that can make quick predictions. Being a probabilistic classifier, it bases its predictions on the probability of an event occurring.
The formula for Bayes' theorem is as follows:

- P(A|B) is the posterior probability, representing the probability of the hypothesis given the observed event B.
- $P(B|A)$P(B|A) is the likelihood probability, indicating the probability of the evidence given that the hypothesis is true.
- $P(A)$P(A) is the prior probability, reflecting the initial probability of the hypothesis before observing the evidence.
- $P(B)$P(B) is the marginal probability, representing the probability of the evidence.

Although Naïve Bayes is widely used and generally effective for classification, it provides less accuracy for certain models.

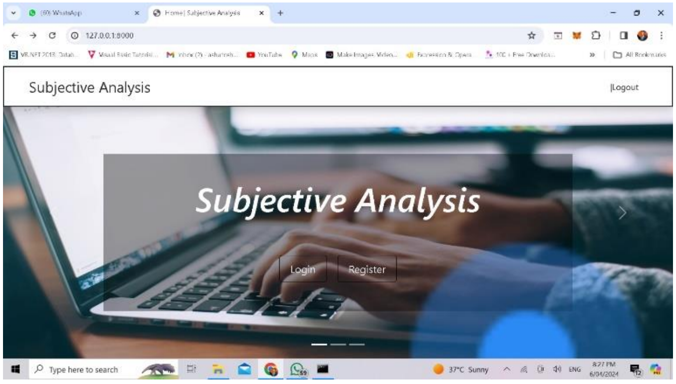
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

6.2 Decision Tree Classification: Decision Tree is a supervised learning method used for both classification and regression problems, though it is primarily used for classification tasks. It is structured as a tree, where internal nodes represent dataset features, branches represent decision rules, and leaf nodes represent outcomes. In a decision tree, there are two types of nodes: Decision Nodes and Leaf Nodes. Decision Nodes are used to make decisions and have multiple branches, while Leaf Nodes represent the outcomes of these decisions and do not have further branches. Decisions or tests are based on the features of the dataset.

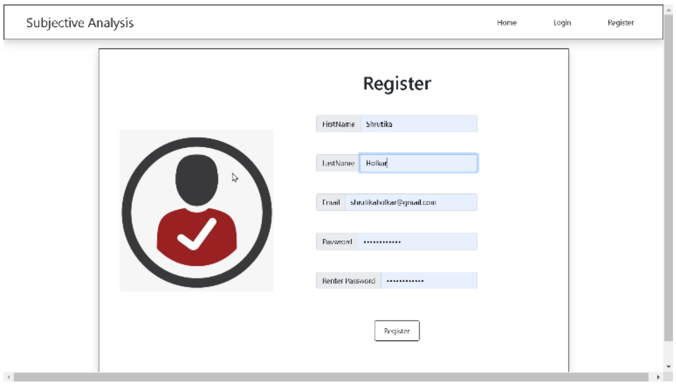
A decision tree provides a graphical representation of all possible solutions to a problem or decision based on given conditions. It is called a decision tree because it starts with a root node and expands into branches, forming a tree-like structure. The CART algorithm, which stands for Classification and Regression Tree, is used to build the tree. The decision tree operates by asking a question and splitting the tree into subtrees based on the answer (Yes/No).

7. IMPLEMENTATION

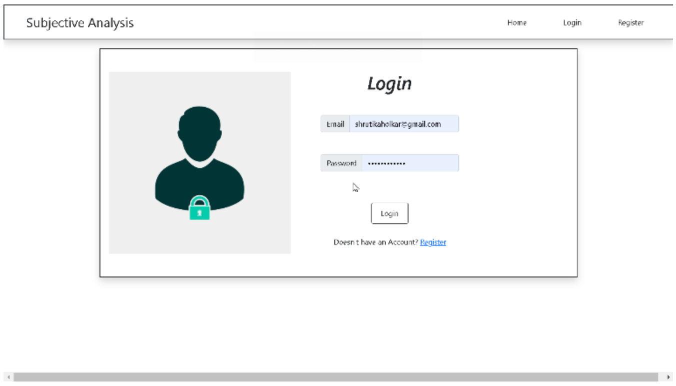
1. Home Page :



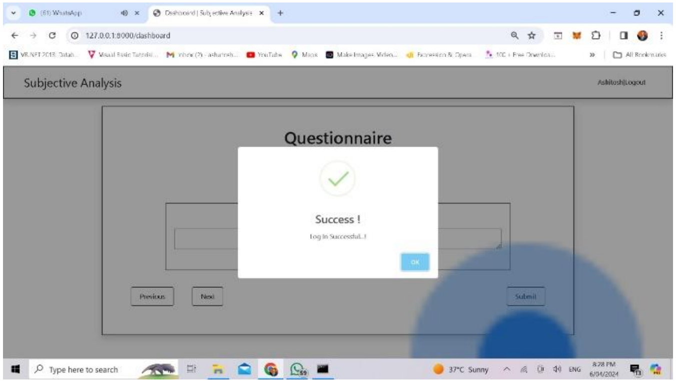
2. **User/Student Registration:** The registration page allows new users (teachers and students) to create an account by providing necessary details such as name, email, and password.



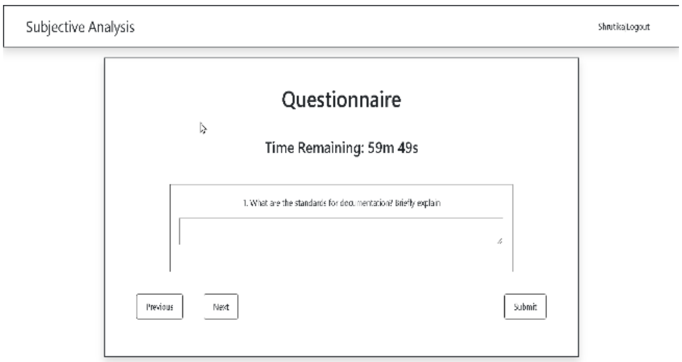
3. **Login Page:** The login page allows registered users to log in by entering their email and password.



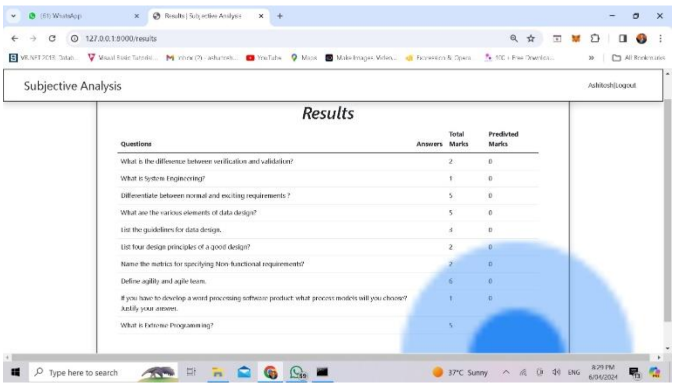
4. **Login Success:** Upon successful login, users are redirected to their respective dashboards (teacher or student), where they can perform relevant actions such as creating exams or joining classes.



5. Questionnaire Page:



6. Exam Analysis Result: After an exam is completed, the system evaluates the answers using NLP and machine learning techniques. The results are then displayed to the student and teacher.



CONCLUSION

We proposed the Subjective Answer Checker System using NLP and Machine Learning (Pariksha Software), which will be beneficial for online exams at universities, schools, and colleges. During the COVID-19 pandemic, most educational institutes conducted their exams online, but these exams typically included only multiple-choice questions. Our Subjective Answer Evaluation software assigns marks to subjective questions based on answer length, keyword matching, grammar check, cosine similarity, and contextual similarity when compared to the model answer provided by the faculty. Additionally, we developed an algorithm to identify contradicting statements between the student's answer and the model answer. This system ensures that even if a student's answer does not match the model answer word for word, it can still be accurately evaluated.

ACKNOWLEDGEMENT

We express our deep gratitude to our project guide, Asst. Prof. Satish C. Cholke, whose guidance, timely assistance, and encouragement have been instrumental in the development of our project. We also extend our sincere thanks to all the lecturers who provided valuable advice and support, making our journey possible.

REFERENCE

[1] Chhanda Roy, Chitrita Chaudhuri, "Case Based Modeling of Answer Points to Expedite Semi-Automated Evaluation of Subjective Papers", in Proc. Int. Conf. IEEE 8th International Advance Computing Conference (IACC), 2018, pp. 85-9.
[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-Assessment System", 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 Prince Sinha, Sharad Bharadia, Ayush Kaul, Dr. Sheetal Rathi ,"Answer Evaluation Using Machine Learning" Conference-McGraw-Hill Publications March 2018
[6] Rosy Salomi Victoria D, Viola Grace Vinitha P, Sathya R," Intelligent Short Answer Assessment using Machine Learning" International Journal of Engineering

and Advanced Technology (IJEAT) , Volume-9 Issue-4, April 2020.
[7 Sakshi Berad, Pratiksha Jaybhaye, Sakshi Jawale,"AI Answer Verifier", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue01; Jan 2019.