A MACHINE LEARNING APPROACH FOR DESCRIPTIVE ANSWER EVALUATION USING COSINE SIMILARITY

SANDEEP REBBA[1]

1.Student, Dept. of Computer Science and Engineering (CSE), Usha Rama College of Engineering and Technology. Near Gannavaram Ungutur Mandalam, Telaprolu, Andhra Pradesh.

Dr. S.M ROY CHOUDRI[2]

2. Professor, Dept. of Computer Science and Engineering (CSE), Usha Rama College of Engineering and Technology. Near Gannavaram Ungutur Mandalam, Telaprolu, Andhra Pradesh

Dr. BEJAGAM PRAVEEN KUMAR[3]

3.M. Tech Coordinator Dept. of Computer Science and Engineering (CSE), Usha Rama College of Engineering and Technology. Near Gannavaram Ungutur Mandalam, Telaprolu, Andhra Pradesh.

ABSTRACT

Subjective paper evaluation is a complex and time-consuming task when performed manually. Challenges such as limited understanding and variability in interpreting student responses make automated evaluation using Artificial Intelligence (AI) difficult. Although several approaches have been proposed to assess descriptive answers, many rely on traditional techniques such as keyword matching or word frequency counts, which fail to capture the true semantic meaning of responses. Additionally, the lack of well-curated datasets further restricts the effectiveness of such systems. This paper proposes a novel approach for automated evaluation of descriptive answers by leveraging machine learning and natural language processing techniques. The system incorporates tools such as WordNet, Word2Vec, Word Mover's Distance (WMD), Cosine Similarity, Multinomial Naive Bayes (MNB), and Term Frequency–Inverse Document Frequency (TF-IDF). Student answers are evaluated by comparing them with model answers and predefined keywords, and a machine learning model is trained to predict grades accordingly. Experimental results indicate that WMD outperforms cosine similarity in capturing semantic similarity between texts. The proposed system achieves an accuracy of 88% without the MNB model, and the inclusion of MNB further reduces the error rate by 1.3%. With sufficient training data, the system has the potential to function as a reliable standalone solution for automated answer evaluation.

1.INTRODUCTION

Subjective questions play a crucial role in assessing a student's understanding, analytical ability, and depth of knowledge in an open-ended manner. Unlike objective questions, subjective answers are not constrained by fixed responses; students are free to express their understanding in their own words and style. As a result, these answers are typically longer, more detailed, and rich in context. However, this flexibility also makes evaluation more challenging, as it requires significant time, concentration, and objectivity from the evaluator. Automating the evaluation of subjective answers using computers is a complex task, primarily due to the inherent ambiguity of natural language. Before analysis, several preprocessing steps such as data cleaning, normalization, and tokenization must be performed. Once processed, textual data can be analyzed using techniques like document similarity measures, latent semantic analysis, concept graphs, and ontologies.

The final score is typically determined based on multiple factors, including semantic similarity, keyword presence, answer structure, and language quality. Although several approaches have been proposed to address this problem, there is still considerable scope for improvement. Subjective evaluations are often perceived as difficult and stressful for both students and teachers due to the importance of context and interpretation. Manual evaluation requires the examiner to carefully assess each response, where factors such as fatigue, bias, and mental effort can influence the final outcome.

In contrast, objective answer evaluation is straightforward and can be easily automated, as responses are fixed and require simple matching techniques.

However, subjective answers vary significantly in length, vocabulary, and expression. The use of synonyms, paraphrasing, and abbreviations further increases the complexity of automated evaluation. Therefore, developing an efficient and reliable system for evaluating subjective answers is essential to reduce human effort, improve consistency, and enhance the overall assessment process.

2.LITERATURE REVIEW

The evaluation of descriptive answers using automated systems has gained significant attention with advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP). Researchers have proposed various techniques to address the challenges of subjective answer evaluation, particularly focusing on semantic understanding and similarity measurement.

Early approaches relied on **keyword matching** and statistical methods such as Term Frequency–Inverse Document Frequency (TF-IDF). Gerard Salton and Christopher Buckley (1988) introduced term-weighting approaches for text retrieval, which laid the foundation for TF-IDF-based evaluation. However, these methods were limited in capturing contextual meaning.

To improve semantic understanding, lexical databases such as WordNet were introduced. George A. Miller (1995) developed WordNet, which enables identification of relationships like synonyms and hypernyms. While this improved semantic evaluation, it still struggled with complex sentence structures.

Machine learning techniques further enhanced automated evaluation systems. Andrew McCallum and Kamal Nigam (1998) explored the use of **Naive Bayes** for text classification, demonstrating its effectiveness in handling textual data. Similarly, Thorsten Joachims (1998) applied **Support Vector Machines (SVM)** for text categorization, improving classification accuracy.

With the advancement of deep learning, **word embedding techniques** gained popularity. Tomas Mikolov et al. (2013) introduced **Word2Vec**, which represents words in vector space and captures semantic relationships more effectively. Building on this, Matt Kusner et al. (2015) proposed **Word Mover's Distance (WMD)**, a method that measures semantic distance between documents using word embeddings, providing highly accurate similarity measures.

Another widely used technique is **cosine similarity**, which calculates similarity between text vectors based on the cosine of the angle between them. Christopher D. Manning et al. (2008) highlighted its effectiveness in information retrieval and text analysis, especially when combined with TF-IDF.

Despite these advancements, challenges such as ambiguity in natural language, lack of standardized datasets, and difficulty in evaluating long descriptive answers still persist.

Therefore, hybrid approaches combining multiple techniques—such as cosine similarity, WMD, and machine learning models—have been proposed to improve performance. This work builds upon prior research by integrating multiple NLP and machine learning techniques to develop a more robust and efficient system for automated descriptive answer evaluation.

3.METHODOLOGY

3.1 SYSTEM ANALYSIS & EXISTING SYSTEM

In the current educational environment, the evaluation of descriptive or subjective answers is primarily performed manually by teachers or examiners. This traditional method requires evaluators to read each answer thoroughly and assign marks based on their understanding, experience, and judgment.

While this approach allows flexibility in assessing students' knowledge and expression, it heavily depends on human interpretation.

Manual evaluation is widely used due to the lack of fully reliable automated systems capable of understanding natural language effectively. Although some basic digital tools exist, they mostly rely on keyword matching or simple scoring techniques, which are insufficient for evaluating complex, context-rich answers.

3.2 LIMITATIONS OF EXISTING SYSTEM

- **Subjectivity and Bias:**
  Manual evaluation often varies between evaluators, leading to inconsistent grading and potential bias.
- **Time-Consuming:**
  Evaluating descriptive answers requires significant time, especially with large numbers of students.
- **Scalability Issues:**
  Handling large-scale examinations becomes difficult due to dependency on human evaluators.
- **Limited Feedback:**
  Due to workload and time constraints, students may not receive detailed feedback on their answers.
- **Resource Intensive:**
  Requires a large number of teachers and evaluators, increasing operational costs.
- **Inconsistency in Evaluation:**
  Factors like fatigue, stress, and human error can affect grading accuracy.

3.3 PROPOSED SYSTEM

The proposed system, **"A Descriptive Answer Evaluation System Using Cosine Similarity Technique,"** introduces an automated and intelligent approach to evaluating subjective answers.

The system utilizes Natural Language Processing (NLP) and Machine Learning techniques to compare student responses with model answers.

The core of the system is the **cosine similarity technique**, which measures the similarity between student answers and reference answers based on their semantic content. This ensures that evaluation is based on meaning rather than exact word matching.

The system also incorporates additional techniques such as TF-IDF, Word2Vec, and Word Mover's Distance (WMD) to enhance accuracy. By automating the evaluation process, the system minimizes human effort and ensures consistency in grading.

Furthermore, the system is implemented as a web-based application, making it accessible and adaptable to modern digital education platforms. It also provides visual feedback (graphs/charts) to help students better understand their performance.

3.4 ADVANTAGES OF PROPOSED SYSTEM

- **Objective Evaluation:**
  Eliminates human bias by using mathematical similarity measures.
- **Time Efficiency:**
  Reduces evaluation time significantly compared to manual grading.
- **High Scalability:**
  Capable of handling large volumes of student responses efficiently.
- **Consistent Results:**
  Ensures uniform grading standards for all students.
- **Enhanced Feedback:**
  Provides graphical and analytical insights into student performance.

🔹 1. Cosine Similarity Formula

This is the **main formula used in your paper**.

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where:

- A = Student answer vector
- B = Model answer vector
- $A \cdot B$ = Dot product
- $||A||, ||B||$ = Magnitudes

👉 Output range: **0 to 1**

- 1 → Highly similar
- 0 → No similarity

🔹 2. TF-IDF Formula

Used to convert text into numerical vectors.

Term Frequency (TF):

$$TF(t,d) = \frac{\text{Number of times term } t \text{ appears in document}}{\text{Total number of terms in document}}$$

Inverse Document Frequency (IDF):

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

TF-IDF:

$$TF\text{-}IDF = TF(t,d) \times IDF(t)$$

Where:

- N = Total documents
- $df(t)$ = Number of documents containing term

🔹 3. Word Mover's Distance (WMD)

Measures semantic distance between texts using word embeddings.

$$WMD(D_1, D_2) = \min \sum_{i,j} T_{ij} \cdot c(i,j)$$

Where:

- T_{ij} = Flow between words
- $c(i,j)$ = Distance between word vectors

🔹 4. Multinomial Naive Bayes (MNB)

Used for classification / scoring.

$$P(c|d) = \frac{P(c) \cdot P(d|c)}{P(d)}$$

👉 Simplified (for text classification):

$$P(c|d) \propto P(c) \prod_{i=1}^{n} P(w_i|c)$$

Where:

- c = Class (score/grade)
- d = Document (answer)

🔹 5. Word2Vec Concept Formula (Skip-gram)

$$P(w_o|w_i) = \frac{\exp(v'_{w_o} \cdot v_{w_i})}{\sum_{w=1}^{V} \exp(v'_w \cdot v_{w_i})}$$

SYSTEM ARCHITECTURE

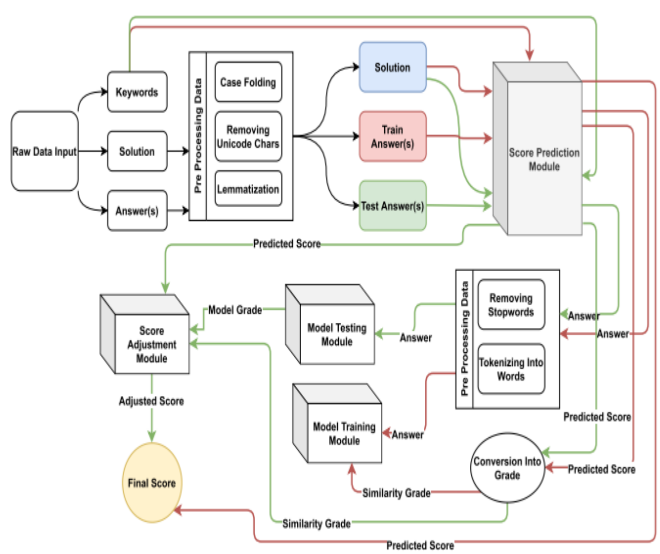


Figure 1: Overall System Architecture of Descriptive Answer Evaluation System

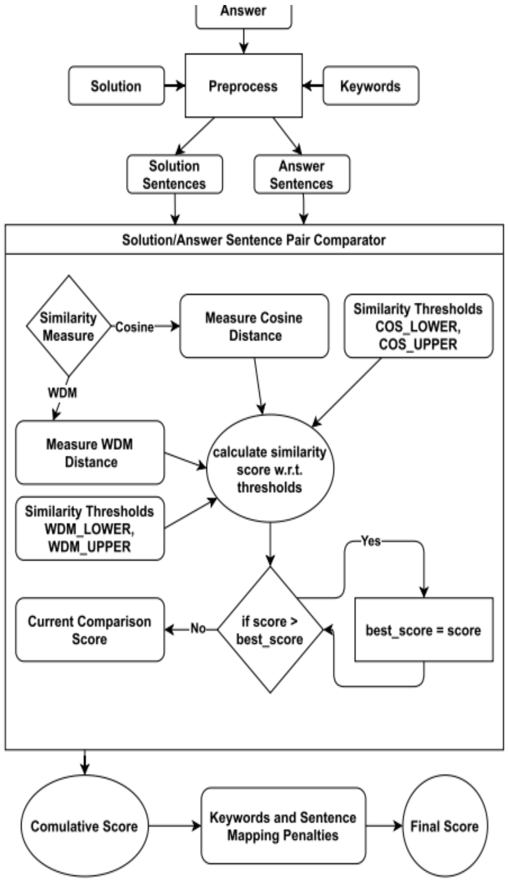


Figure 2: Data Flow Diagram of the Proposed System

4.IMPLEMENTATION

MODULES

1. User Authentication and Management

This module handles secure login and user management within the system. It supports multiple roles such as administrators, teachers, and students. Each user is granted access based on their role, ensuring data security and controlled system usage. The module also manages user registration, login validation, and session handling.

2. Answer Submission Interface

This module provides a user-friendly platform for students to submit their descriptive answers. It allows users to enter or upload responses efficiently. The submitted answers are stored in the database and are readily available for evaluation. The interface ensures smooth interaction and minimizes errors during submission.

3. Automated Evaluation Engine

This is the core module of the system. It evaluates descriptive answers using Natural Language Processing (NLP) and machine learning techniques. The system applies cosine similarity to compare student answers with model answers and determine their semantic similarity. Additional techniques such as TF-IDF and Word2Vec may be used to enhance accuracy. Based on this analysis, scores are generated automatically.

4. Result Presentation and Visualization

After evaluation, this module displays the results to users in a clear and understandable format. It provides score summaries along with graphical representations such as charts or similarity scores. This helps students and teachers quickly interpret performance and outcomes.

5. Feedback and Reporting

This module enables teachers to provide detailed feedback on student responses. It highlights strengths, weaknesses, and areas for improvement. Additionally, it generates reports for administrators, offering insights into system usage, student performance, and overall evaluation trends.

HARDWARE REQUIREMENTS

S. No.	Component	Minimum (Required for Execution)	My System (Development)
1	System	Pentium IV 2.2 GHz	Intel i3 Processor (5th Gen)
2	Hard Disk	20 GB	500 GB
3	RAM	1 GB	4 GB

SOFTWARE REQUIREMENTS

S. No.	Component	Specification
1	Operating System	Windows 10 / Windows 11
2	Development Software	Python 3.10
3	Programming Language	Python
4	Domain	Natural Language Processing & Machine Learning
5	IDE	Visual Studio Code
6	Front-End Technologies	HTML5, CSS3, JavaScript
7	Back-End Framework	Django
8	Database Language	SQL
9	Database (RDBMS)	MySQL
10	Database Software	WAMP / XAMPP Server
11	Web Server	Django Development Server
12	Design / Modelling Tool	Rational Rose

DESCRIPTIVE ANSWER EVALUATION:

Main:

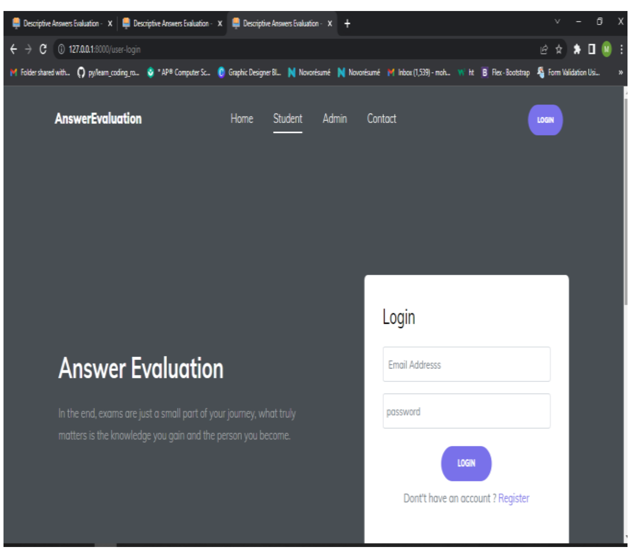


Figure 3: Landing Page of Answer Evaluation Web Application

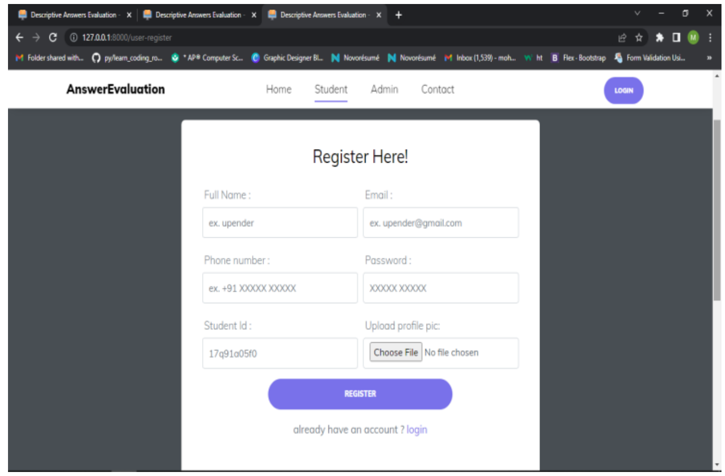


Figure 4: User Registration Page of Answer Evaluation System

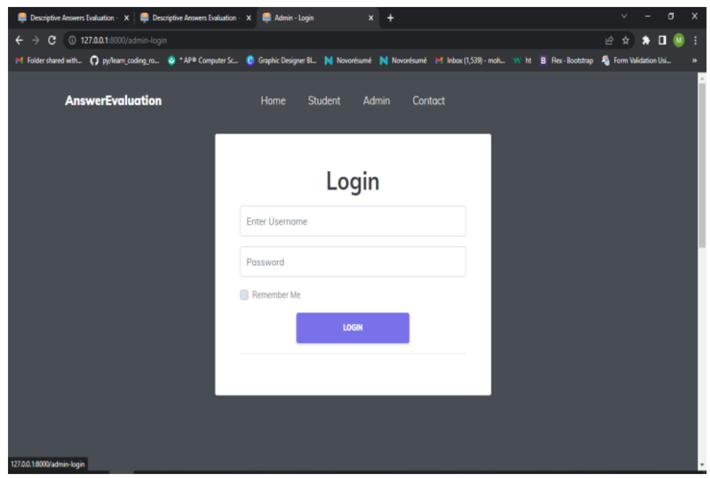
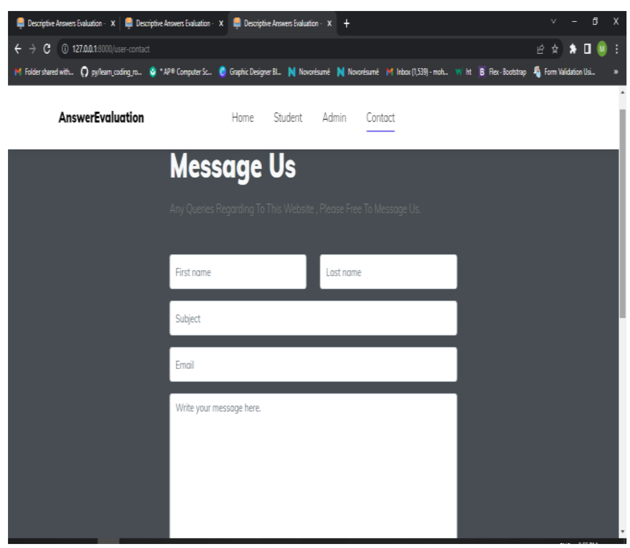
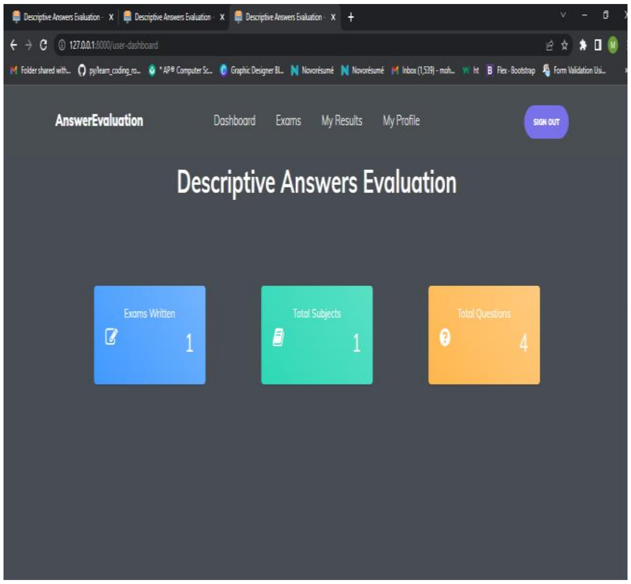
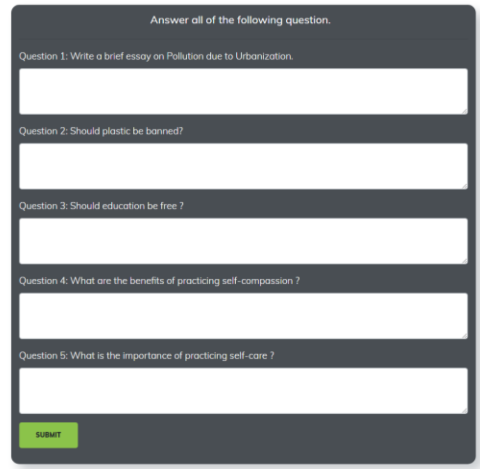


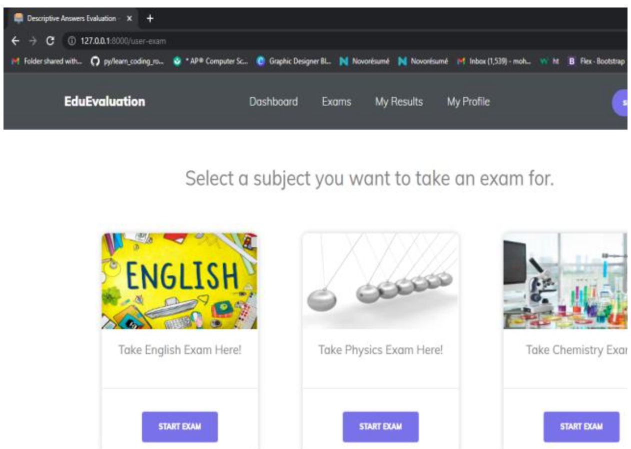
Figure 5: Front-End Design of Login Page

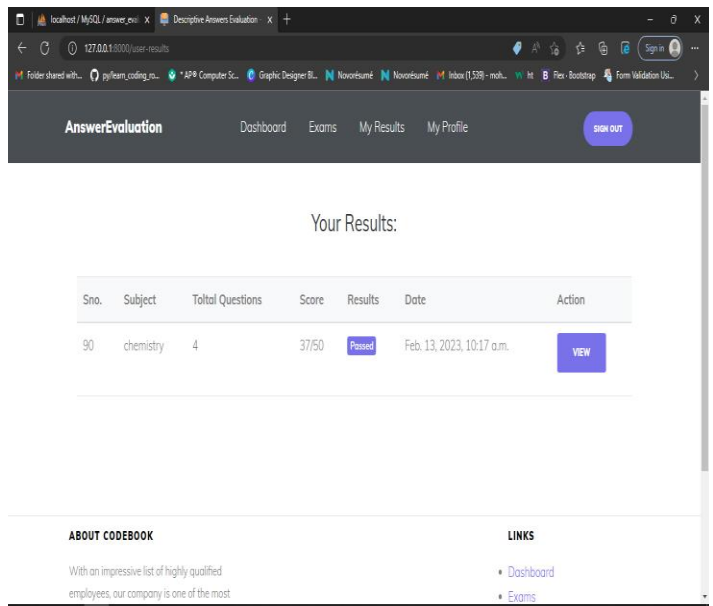


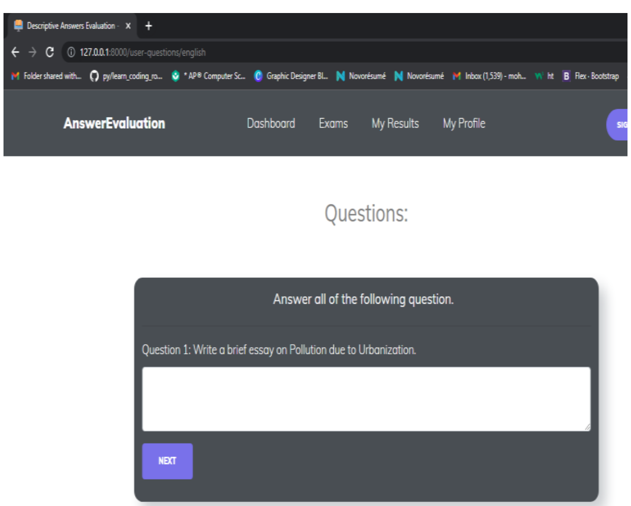
User:

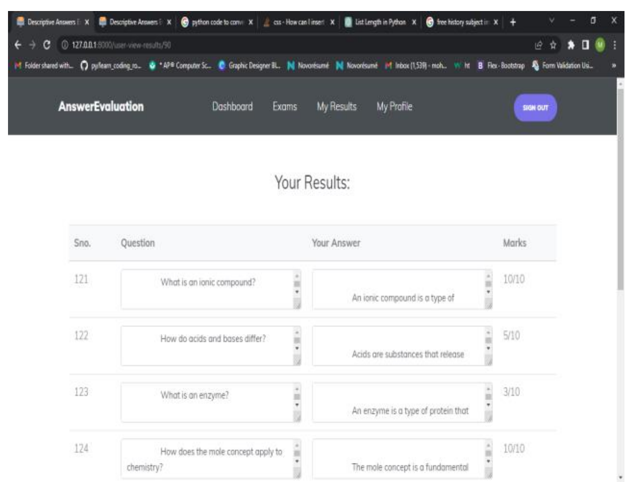

AnswerEvaluation
Dashboard
Exams
My Results
My Profile
SIGN OUT
Descriptive Answers Evaluation
Exams Written
1
Total Subjects
1
Total Questions
4

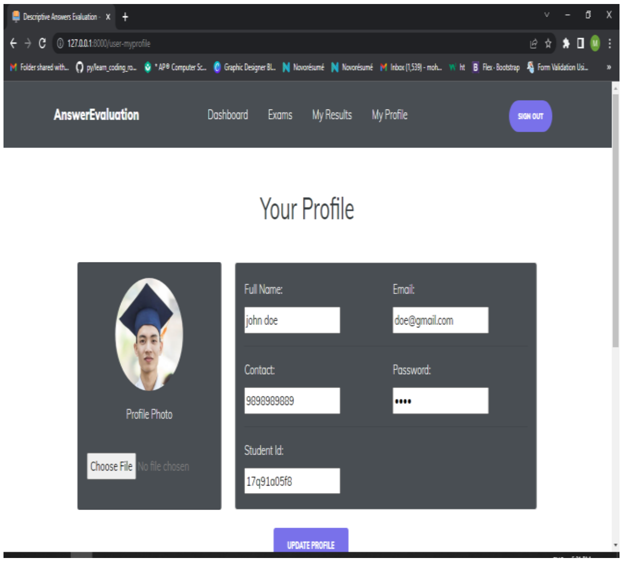

Answer all of the following question.
Question 1: Write a brief essay on Pollution due to Urbanization.
Question 2: Should plastic be banned?
Question 3: Should education be free ?
Question 4: What are the benefits of practicing self-compassion ?
Question 5: What is the importance of practicing self-care ?
SUBMIT


EduEvaluation
Dashboard
Exams
My Results
My Profile
Select a subject you want to take an exam for.
ENGLISH
Take English Exam Here!
Take Physics Exam Here!
START EXAM
START EXAM
START EXAM

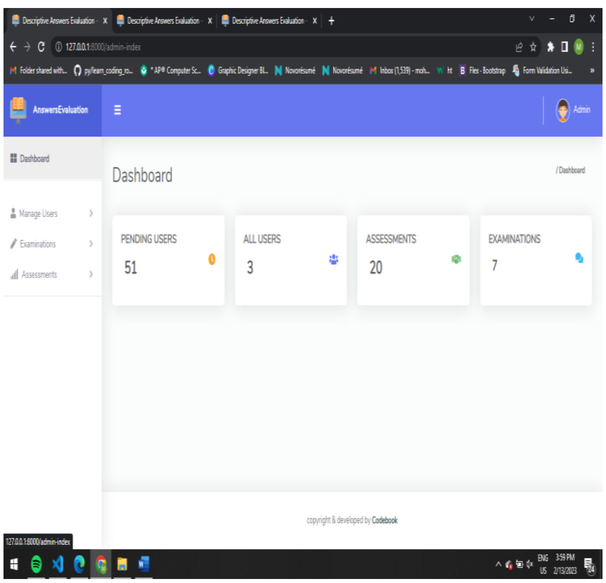

AnswerEvaluation
Dashboard
Exams
My Results
My Profile
SIGN OUT
Your Results:
Sno.
Subject
Toltal Questions
Score
Results
Date
Action
90
chemistry
4
37/50
Passed
Feb. 13, 2023, 10:17 a.m.
VIEW
ABOUT CODEBOOK
With an impressive list of highly qualified
employees, our company is one of the most
LINKS
Dashboard
Exams

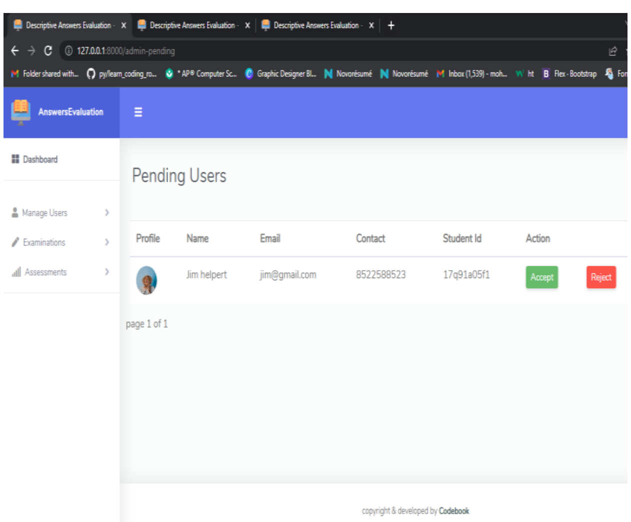

AnswerEvaluation
Dashboard
Exams
My Results
My Profile
Questions:
Answer all of the following question.
Question 1: Write a brief essay on Pollution due to Urbanization.
NEXT

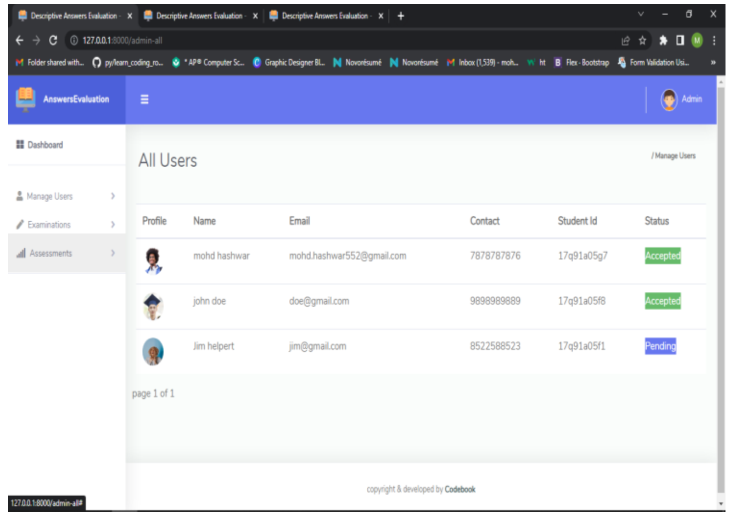

AnswerEvaluation
Dashboard
Exams
My Results
My Profile
SIGN OUT
Your Results:
Sno.
Question
Your Answer
Marks
121
What is an ionic compound?
An ionic compound is a type of
10/10
122
How do acids and bases differ?
Acids are substances that release
5/10
123
What is an enzyme?
An enzyme is a type of protein that
3/10
124
How does the mole concept apply to chemistry?
The mole concept is a fundamental
10/10

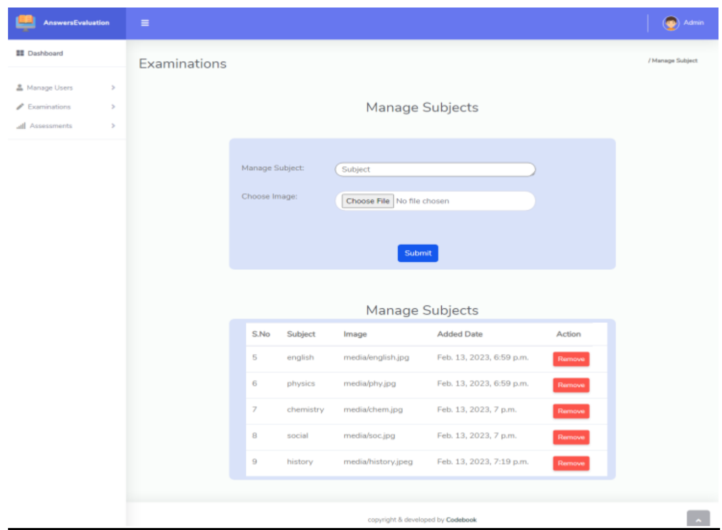

AnswerEvaluation
Dashboard
Exams
My Results
My Profile
SIGN OUT
Your Profile
Profile Photo
Choose File No file chosen
Full Name:
john doe
Email:
doe@gmail.com
Contact:
9898989889
Password:
Student Id:
17q91a05f8
UPDATE PROFILE

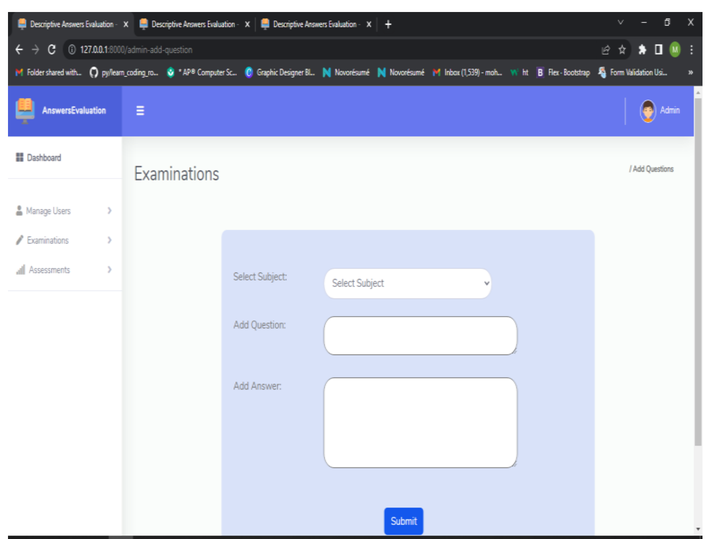
Admin:

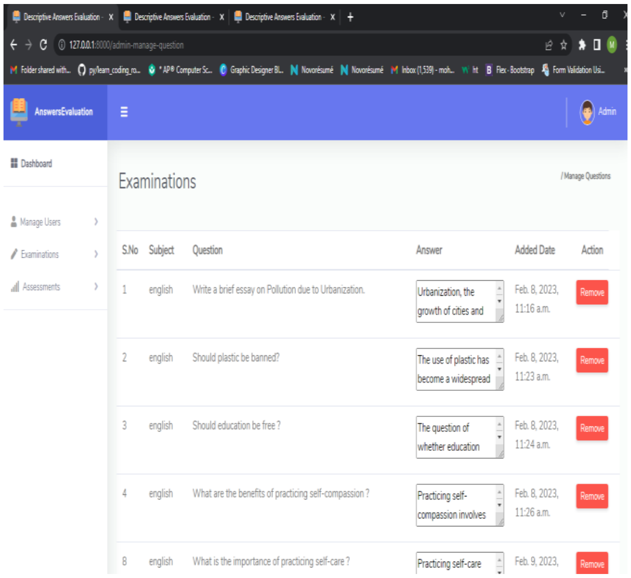

AnswersEvaluation
Admin
Dashboard
Manage Users
Examinations
Assessments
Dashboard
PENDING USERS
51
ALL USERS
3
ASSESSMENTS
20
EXAMINATIONS
7
copyright & developed by Codebook

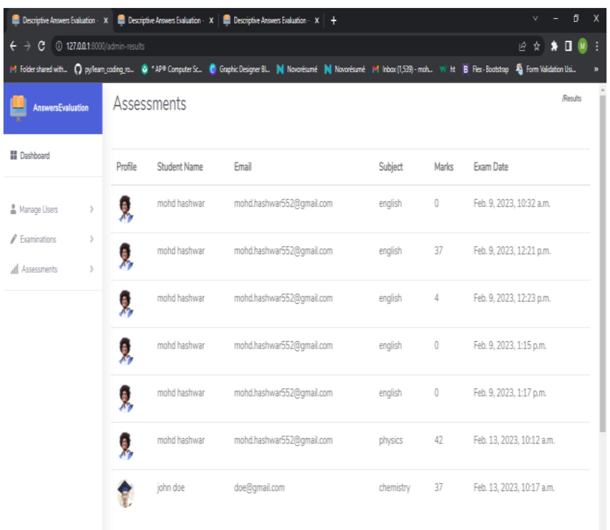

AnswersEvaluation
Dashboard
Manage Users
Examinations
Assessments
Pending Users
Profile
Name
Email
Contact
Student Id
Action
Jim helpert
jim@gmail.com
8522588523
17q91a05f1
Accept
Reject
page 1 of 1
copyright & developed by Codebook

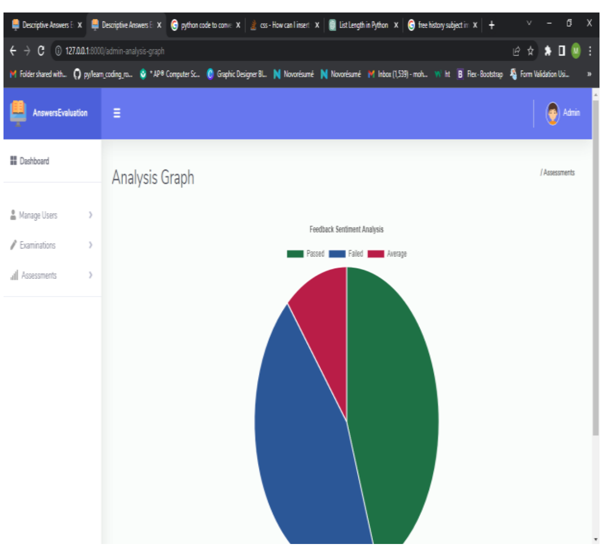

AnswersEvaluation
Admin
Dashboard
Manage Users
Examinations
Assessments
All Users
/ Manage Users
Profile
Name
Email
Contact
Student Id
Status
mohd hashwar
mohd.hashwar552@gmail.com
7878787876
17q91a05g7
Accepted
john doe
doe@gmail.com
9898989889
17q91a05f8
Accepted
Jim helpert
jim@gmail.com
8522588523
17q91a05f1
Pending
page 1 of 1
copyright & developed by Codebook


AnswersEvaluation
Admin
Examinations
Manage Subjects
Manage Subject:
Choose Image:
Choose File No file chosen
Submit
Manage Subjects
S.No
Subject
Image
Added Date
Action
5
english
media/english.jpg
Feb. 13, 2023, 6:59 p.m.
Remove
6
physics
media/phy.jpg
Feb. 13, 2023, 6:59 p.m.
Remove
7
chemistry
media/chem.jpg
Feb. 13, 2023, 7 p.m.
Remove
8
social
media/soc.jpg
Feb. 13, 2023, 7 p.m.
Remove
9
history
media/history.jpeg
Feb. 13, 2023, 7:19 p.m.
Remove


AnswersEvaluation
Admin
Dashboard
Manage Users
Examinations
Assessments
Examinations
/ Add Questions
Select Subject:
Select Subject
Add Question:
Add Answer:
Submit







CONCLUSION

This paper presents an automated descriptive answer evaluation system using cosine similarity and advanced natural language processing techniques. The proposed approach addresses the limitations of traditional keyword-based evaluation methods by incorporating semantic understanding through tools such as Word2Vec and Word Mover's Distance (WMD). By comparing student responses with model answers and extracting meaningful features, the system provides a more accurate and consistent evaluation process. The experimental results demonstrate that WMD performs better than cosine similarity in capturing contextual similarity between answers. The system achieves an accuracy of 88% without using the Multinomial Naive Bayes (MNB) model, and further improvement is observed with a reduction in error rate by 1.3% when MNB is applied. These results indicate that combining multiple techniques enhances the overall performance and reliability of the evaluation system. In conclusion, the proposed model reduces manual effort, ensures consistency in grading, and shows strong potential for real-world academic applications. With the availability of larger and well-structured datasets, the system can be further improved and deployed as a standalone automated evaluation tool in educational institutions.

REFERENCES

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2013.

[2] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and

Phrases and their Compositionality," *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[3] K. W. Church and P. Hanks, "Word Association Norms, Mutual Information, and Lexicography," *Computational Linguistics*, vol. 16, no. 1, pp. 22–29, 1990.

[4] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[5] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From Word Embeddings to Document Distances," *Proceedings of the International Conference on Machine Learning (ICML)*, 2015.

[6] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[7] S. Bird, E. Loper, and E. Klein, *Natural Language Processing with Python*, O'Reilly Media, 2009.

[8] J. B. Lovins, "Development of a Stemming Algorithm," *Mechanical Translation and Computational Linguistics*, vol. 11, pp. 22–31, 1968.

[9] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proceedings of ECML*, 1998.

[10] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI-98 Workshop on Learning for Text Categorization*, 1998.

[11] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.

[12] Z. Harris, "Distributional Structure," *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.