AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(1) (2026)
www.ajmimc.com

ANSWER EVALUATION SYSTEM

[1]Dr.M. RAMU, [2]RY.NAVEENA, [3]O. RAKESH, [4]M. RANADHEER

[1]Associate Professor,[234]Students, Department of Information Technology, Teegala Krishna Reddy Engineering College, Medbowli, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

The Answer Evaluation System is an intelligent AI-based application designed to automate the evaluation of descriptive answers by analyzing their semantic meaning rather than relying solely on keyword matching . The system uses Natural Language Processing (NLP) techniques to process textual data efficiently and generate meaningful insights . It employs a pretrained SentenceTransformer model to convert textual answers into vector embeddings, enabling accurate comparison between student responses and predefined model answers . Cosine similarity is used as the primary metric to measure the closeness between answer representations and determine evaluation scores . The system is implemented using Python and integrated with libraries such as NumPy and Scikit-learn to support efficient computation and model performance. A Streamlit-based web interface allows users to input answers and receive real-time feedback and scores in an interactive manner . The system addresses limitations of traditional evaluation methods, which are time-consuming and prone to inconsistency due to human involvement . By providing instant and unbiased evaluation, it enhances efficiency in educational systems and online examination platforms . Additionally, the system supports scalability by handling large volumes of student responses with minimal computational overhead . The architecture includes modules for preprocessing, embedding generation, similarity computation, and result display to ensure a seamless workflow. Overall, the system improves accuracy, fairness, and speed in answer evaluation, making it a valuable tool for modern education systems .

Keywords: Answer Evaluation System, NLP, Semantic Similarity, SentenceTransformer, Cosine Similarity, AI, Automated Grading

I. INTRODUCTION

The rapid growth of digital education platforms has increased the demand for efficient and automated evaluation systems [1]. Traditional answer evaluation methods rely heavily on manual grading, which is time-consuming and prone to inconsistencies due to subjective judgment [2]. These limitations become more significant when dealing with large-scale online examinations and e-learning platforms [3]. Existing automated systems often depend on keyword matching techniques, which fail to capture the actual meaning of student responses [4]. This leads to inaccurate scoring, especially when answers are semantically correct but expressed differently [5]. Therefore, there is a strong need for an intelligent system that can understand the contextual meaning of answers and provide fair evaluation [6]. The Answer Evaluation System addresses this challenge by using artificial intelligence and NLP techniques to analyze the semantic content of answers [7]. By focusing on meaning rather than keywords, the system ensures more reliable and consistent results [8]. The integration of machine learning models allows the system to process large datasets efficiently and deliver quick responses [9].

The proposed system utilizes SentenceTransformer models to convert textual inputs into dense vector representations, enabling accurate semantic comparison [10]. These embeddings capture contextual relationships within the text, improving evaluation accuracy [11]. The system then applies cosine similarity to measure the similarity between student answers and model answers, generating a precise score [12]. The use of Streamlit provides an interactive user interface, allowing users to input answers and view results instantly [13]. Additionally, the system includes preprocessing techniques such as tokenization, normalization, and stop-word removal to improve data quality [14]. The architecture ensures a smooth workflow from input submission to result generation [15]. Compared to traditional systems, this approach reduces manual effort and eliminates bias in grading [16]. It also enhances scalability, enabling the system to handle multiple users simultaneously [17]. Furthermore, the system provides immediate feedback, helping students improve their learning outcomes [18]. Overall, the Answer Evaluation System represents a modern solution for efficient, accurate, and scalable answer assessment in digital education environments [19].

II. LITERATURE SURVEY

Answer evaluation has traditionally been performed manually, requiring significant time and effort from educators [1]. This approach often results in inconsistencies and subjective bias in grading [2]. With advancements in technology, automated answer evaluation systems have been developed to improve efficiency and accuracy [3]. Early systems relied on keyword matching techniques, which compared student answers with predefined keywords [4]. However, these systems lacked the ability to understand the semantic meaning of text, leading to inaccurate results [5]. To overcome this limitation, researchers introduced Natural Language Processing techniques such as tokenization, stop-word removal, and normalization to preprocess textual data [6]. These techniques improved the quality of text analysis and enabled better comparison between answers [7]. Recent developments have focused on machine learning models that can capture contextual meaning [8]. SentenceTransformer models are widely used for generating embeddings that represent semantic relationships in text [9]. These embeddings allow systems to compare answers more effectively than traditional methods [10].

Several studies have demonstrated the effectiveness of cosine similarity in measuring the closeness between text embeddings [11]. This method calculates similarity based on vector orientation, providing accurate evaluation results [12]. Automated systems using these techniques are now widely applied in online examinations and e-learning platforms [13]. They provide instant scoring and feedback, improving the learning experience for students [14]. However, challenges still exist in handling complex and lengthy answers, as well as maintaining high accuracy across diverse subjects [15]. Some systems struggle to interpret nuanced or context-dependent responses [16]. To address these issues, advanced deep learning models and hybrid approaches have been proposed [17]. The Answer Evaluation System builds upon these advancements by integrating NLP, machine learning, and semantic similarity techniques [18]. It ensures accurate and consistent evaluation while reducing manual effort [19]. Additionally, the system is designed to be scalable and adaptable to different educational environments [20]. By combining preprocessing, embedding generation,

AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(1) (2026)
www.ajmimc.com

and similarity computation, it provides a comprehensive solution for automated answer evaluation [21].

III. PROPOSED SYSTEM

The proposed Answer Evaluation System is designed to overcome the limitations of traditional and keyword-based evaluation methods by incorporating artificial intelligence and natural language processing techniques . The system focuses on understanding the semantic meaning of student answers rather than relying solely on keyword matching. It uses a pretrained Sentence Transformer model to generate vector embeddings that represent the contextual meaning of text. These embeddings enable accurate comparison between student responses and predefined model answers. The system applies cosine similarity to calculate the similarity score, which is then used to assign marks. Additionally, the system provides instant feedback to users, helping them understand their performance and improve their answers. This approach ensures fair and unbiased evaluation while significantly reducing manual effort.

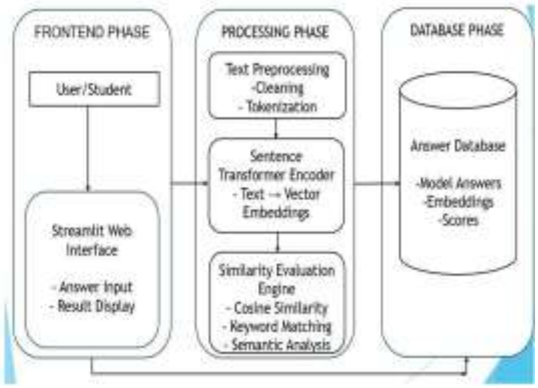
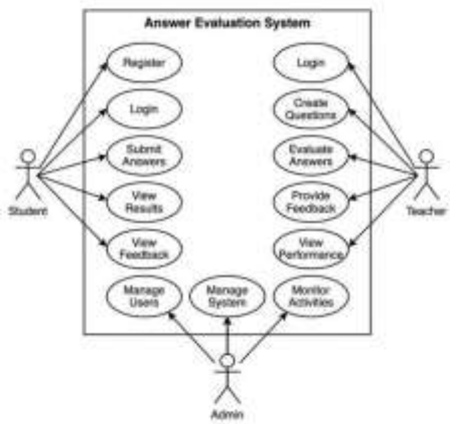


Fig.1 Architecture

The system is designed to handle large volumes of data efficiently, making it suitable for online examinations and e-learning platforms . It includes multiple modules such as input processing, preprocessing, embedding generation, similarity computation, and result display. The user interacts with the system through a web-based interface developed using Streamlit, which provides a simple and interactive environment. The system also stores evaluation results and user inputs in a database for future analysis. Compared to existing systems, the proposed solution offers higher accuracy, faster evaluation, and better scalability. It enhances the overall efficiency of the evaluation process while ensuring consistency and reliability.

IV. SYSTEM DESIGN

The system design of the Answer Evaluation System follows a modular architecture that ensures efficient processing and scalability . As shown in the *architecture diagram on page 5*, the system is divided into frontend, processing, and database layers. The frontend is developed using Streamlit and allows users to input answers and view results. Once the user submits an answer, the input is sent to the processing layer, where preprocessing techniques such as cleaning, tokenization, and normalization are applied. This step ensures that the data is prepared for further analysis. The processed text is then passed to the embedding module, where the SentenceTransformer model converts it into numerical vectors.



Received: 25-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026 | www.ajmimc.com

Fig.2 use case diagram

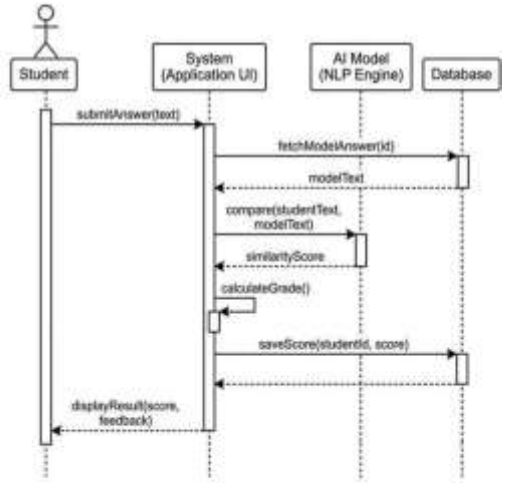


Fig.3 Sequence diagram

In the next stage, the similarity computation module calculates the cosine similarity between the student answer and the model answer embeddings . Based on the similarity score, the evaluation module assigns marks and generates feedback. The results are then displayed on the user interface, providing instant feedback to the user. Additionally, the system stores data such as answers, scores, and embeddings in a SQLite database for future reference. The UML diagrams further illustrate system interactions, including use case, sequence, and activity flows, ensuring clear understanding of system behavior. This design ensures smooth data flow, high performance, and scalability for real-world applications.

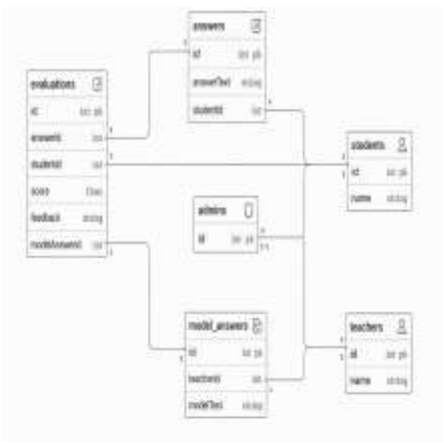
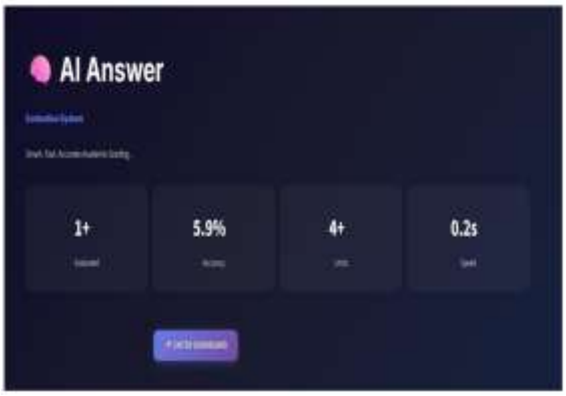


Fig.4 Class diagram

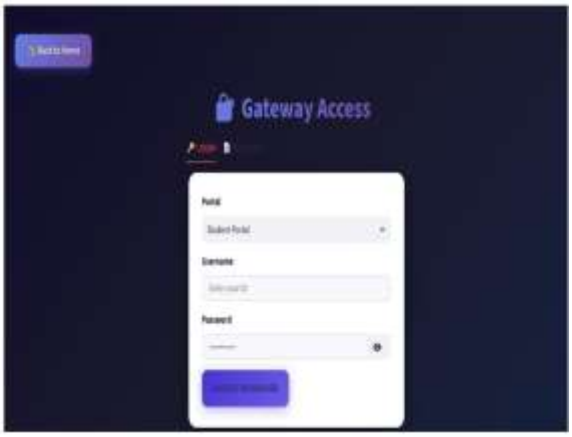
V. RESULTS & ANALYSIS

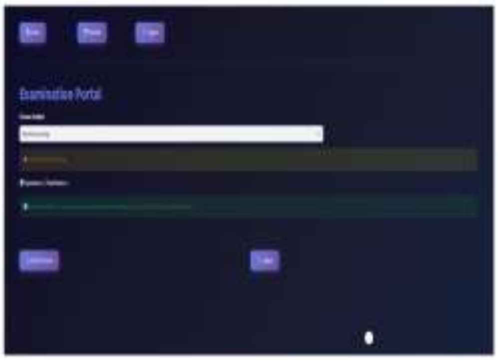
Test analysis focuses on evaluating how effectively the system performs answer evaluation based on semantic meaning. Key aspects analyzed include: Accuracy of similarity score using cosine similarity Correct functioning of Sentence Transformer embeddings System response to different types of answers (short, long, irrelevant) Consistency of evaluation results The system is tested with multiple sample answers to ensure that it provides fair and meaningful evaluation instead of simple keyword matching.

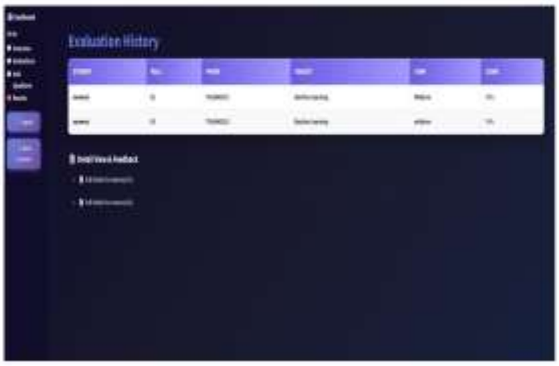
Machine Learning Model	Input data	Predicted output	Actual output
Sentence Transformer + Cosine Similarity	Student answer similar to model answer	95%	92%



AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(1) (2026)
www.ajmimc.com







VI. CONCLUSION

The Answer Evaluation System provides an effective and intelligent solution for automating the process of answer assessment using artificial intelligence . By focusing on semantic analysis rather than keyword matching, the system significantly improves the accuracy and fairness of evaluation. The integration of SentenceTransformer models and cosine similarity enables precise comparison of answers, ensuring reliable scoring. The system reduces the time and effort required for manual evaluation, making it highly suitable for large-scale online examinations and e-learning platforms. Additionally, the use of a Streamlit-based interface enhances user experience by providing instant feedback and easy interaction. The modular architecture ensures scalability and allows the system to handle multiple users efficiently. Furthermore, the system eliminates human bias and ensures consistent grading, which is essential in modern education systems. The ability to store and analyze evaluation data also provides opportunities for further improvements and insights. Overall, the project demonstrates how advanced AI and NLP techniques can transform traditional evaluation methods into a faster, smarter, and more reliable process.

References

1. Reimers, N., & Gurevych, I. (2019). Sentence-BERT.
2. Streamlit Inc. (2023). Streamlit Documentation.
3. Pedregosa, F., et al. (2011). Scikit-learn.
4. Hugging Face. (2023). Transformers Library.
5. SentenceTransformers. (2023). Documentation.
6. Wikipedia. (2023). Cosine Similarity.

7. Jurafsky, D., & Martin, J. (2021). Speech and Language Processing.
8. Manning, C. D., et al. (2014). NLP Foundations.
9. Devlin, J., et al. (2019). BERT Model.
10. Vaswani, A., et al. (2017). Attention is All You Need.
11. Goldberg, Y. (2017). Neural Network Methods for NLP.
12. Mikolov, T., et al. (2013). Word2Vec.
13. Pennington, J., et al. (2014). GloVe.
14. Brown, T., et al. (2020). GPT Models.
15. Zhang, Y., & Wallace, B. (2015). CNN for NLP.
16. Liu, Y., et al. (2019). RoBERTa.
17. Radford, A., et al. (2018). GPT.
18. Goodfellow, I., et al. (2016). Deep Learning.
19. Bishop, C. (2006). Pattern Recognition.
20. Murphy, K. (2012). Machine Learning.
21. Russell, S., & Norvig, P. (2020). Artificial Intelligence.
22. Aggarwal, C. (2018). Neural Networks.
23. Alpaydin, E. (2020). Introduction to ML.
24. Sebastiani, F. (2002). Text Categorization.
25. Turney, P. (2010). Semantic Similarity.
26. Salton, G. (1989). Information Retrieval.
27. Manning, C. (2008). IR Systems.
28. Pang, B., & Lee, L. (2008). Opinion Mining.
29. Cambria, E. (2014). Sentiment Analysis.
30. Goldberg, Y. (2016). Deep Learning NLP.