Utilitas Mathematica
ISSN: 0315-3681 Volume 122, 2025

# ARTIFICIAL INTELLIGENCE SYSTEM FOR THE ACCURATE EVALUATION OF SUBJECTIVE EXAMINATION ANSWERS USING NLP

**Sriharshini Bondugula[1] , Dr.Nazimunisa[2]**
[1]PG Scholar , Department Of Computer Science And Engineering,
Sree Dattha Institute Of Engineering And Science,
Sheriguda, Ibrahimpatnam, Hyderabad - 501510
*Email :* sriharshinibondugula@gmail.com
[2]Associate Professor , Department of Computer Science and Engineering,
Sree Dattha Institute Of Engineering and Science,
Sheriguda, Ibrahimpatnam , Hyderabad - 501510
*Email :* Pnajma22@sreedattha.ac.in

**ABSTRACT**

**Manual marking of descriptive answers is time-consuming and prone to inconsistencies. This project employs machine learning (ML) and natural language processing (NLP) to mark subjective answers automatically for accuracy, efficiency, and fairness. The system follows a two-stage process: in the first stage, student answers are compared with model answers using Word Mover's Distance (WMD) and Cosine Similarity to find semantic similarity. Second, similarity scores are used to train an ML model to rate answers independent of predefined solutions. Preprocessing techniques such as tokenization, stemming, lemmatization, and stop-word filtering improve accuracy while preserving semantics through Word2Vec, TF-IDF, and Bag of Words. Tests on varied datasets indicate that Word2Vec is more effective than other models with an accuracy of up to 88%. Future enhancements involve domain-specific models, deep learning (BERT, GPT), and multilingual support. This system transforms electronic learning by streamlining teacher work, providing just grading, and enhancing subjective response evaluation in teaching.**

***Keywords: Subjective Answer Evaluation, Machine Learning, Natural Language Processing (NLP), Semantic Similarity, Automated Grading***

## I. INTRODUCTION

Subjective marking of test responses has been a difficult and time-consuming activity for instructors for a long time. In contrast to objective tests, which can be marked automatically, subjective responses need a sharp sense of context, coherence, and correctness. Subjective responses are graded by human markers traditionally, which creates problems of bias, inconsistency, and scalability. With quick progress in Natural Language Processing (NLP) and Artificial Intelligence (AI), machine systems can now be trained to analyze subjective responses more objectively and accurately. The current paper discusses how an AI-based system can be programmed to analyze subjective exam answers accurately through NLP methods. NLP allows machines to process, comprehend, and analyze human language, making it a perfect tool for examining written responses. Methods like text similarity analysis, sentiment analysis, and semantic comprehension can be employed to compare model answers with student responses to provide fair and impartial grading. Deep learning models like transformers and large language models can also improve the accuracy of grading by taking contextual meaning into account instead of keyword matching. Integration of such an AI system may have the potential to transform the education industry by facilitating instantaneous, uniform, and scalable subjective response evaluation. It can be advantageous for teachers by easing workload pressures, avoiding biased testing, and offering immediate results to students. This present research makes an attempt to study methodologies, problems, and effectiveness of AI-based testing systems for subjective testing and open doors to an even more streamlined and technology-oriented model of grading. Grading of descriptive or subjective answers has always been a necessary but time-consuming part of marking in academia. Compared to objective items, the subjective answers require a deeper knowledge of the topic and require human graders to interpret, judge, and grade answers. Manual marking will be likely to produce a series of drawbacks such as inconsistency, human error, time-consuming, and even subconscious bias. These problems, apart from reducing the credibility of the assessment process, also add extra workloads to teachers. This project seeks to fill these loopholes with an automatic subjective answer rating system through the application of machine learning (ML) and natural language processing (NLP) concepts. The primary aim of the system is to provide correctness, justice, and efficiency in marking students' answers. The suggested system is a two-stage process. The student responses are then semantically aligned with model responses and keywords using algorithms like Word Mover's Distance (WMD) and Cosine Similarity. The approach measures similarity between student response and ideal response, and word usage variations and sentence construction differences are neatly taken care of. Second, these similarity scores are also utilized to train a machine learning model to label answers into different levels where direct matching is not possible. For additional improvement in performance, the system processes text beforehand via tokenization, stemming, lemmatization, and removing stop-words but preserving meaning to a good extent using methods like Word2Vec, TF-IDF, and Bag of Words. From experimental results based on a heterogeneous dataset annotated manually, it was found that Word2Vec combined with WMD outperformed conventional methods wherein the system performs with a accuracy of up to 88%. Future enhancement may include the incorporation of domain models, deep learning techniques like BERT or GPT, and multi-language support. The system is a huge leap towards marking process automation in education, with minimal human interference, and inviting fair and uniform marking.

## II. LITERATURE SURVEY

Computer marking of subjective responses has seen a lot of attention since the advent of Natural Language Processing (NLP) and Machine Learning (ML). Recent research has explored various models and approaches in a bid to improve the accuracy, fairness, and effectiveness of marking. Fan et al. (2024) had looked at generative language models as mark tools to use on subjective questions with interest lying in the capability to push information extraction. Likewise, Maharajpet et al. (2024) had proposed an NLP approach for subjective answer evaluation, pointing the way toward the enhancement of automated scoring. Agrawal et al. (2024) laid down a three-pronged approach of NLP and deep learning, which opened doors to more precise subjective testing. There were few other papers, like Kumari et al. (2023) and Sree and Joseph (2023), that were geared towards AI-based automated assessment of e-learning so that there will be uniformity as well as efficiency in the educational settings. And, Sinha et al. (2023) examined machine learning and NLP techniques for answer credibility assessment improvement. Deep learning also supported computer-aided test strategies. Hao et al. (2022) presented a survey of deep learning application to question answering with an emphasis on its use in subjective answer grading. Johri et al. (2021) gave a semantic learning-based approach, further elaborating the ability of NLP to grade responses, while Bashir et al. (2021) gave the employment of ML and NLP in the process of enhancing grading as accurate and impartial. Studies of semantic similarity metrics have also come under the limelight. Wang and Dong (2020) described text similarity methods, which were a primary reason for the success of response grading. Han et al. (2021) followed suit by compared short-text semantic similarity methods, while Muangprathub et al.(2021) identified plagiarism detection by formal concept analysis and suggested the potential of similarity recognition in testing. Earlier research has helped shape the field. Patil and Patil (2014) employed NLP to emphasize students' responses, indicating earlier application in education. Correspondingly, Patil et al. (2018) introduced ML-based subjective marking plans with automation focus. Such studies made way for current AI-based measurement methods like advanced deep models like BERT and GPT. Nevertheless, classical semantic similarity approaches still hold ground, newer trends indicating that AI-based grading methodologies are evolving. The future task needs to be on the development of multilingual support, handling intricate forms of responses, and improving deep learning models to deliver equilibrated and efficient grading systems.

## III. PROPOSED WORK

The suggested project aims to automate marking of descriptive answers using machine learning and natural language processing (NLP) in an attempt to provide accuracy, speed, and equity in marking. The system involves a two-step process: one is the computation of similarity between student response and model solution using semantic similarity techniques like Word Mover's Distance (WMD) and Cosine Similarity. Second, the similarity scores are employed as features to train a machine learning model so that the model is able to score responses without heavy dependence on pre-defined answers or keyword matches. Management of synonyms, sentence structure, and answer length are among the greatest challenges in subjective answer marking. As a solution, the proposed system pre-processes the text in the form of tokenization, stemming, lemmatization, and stop-word removal while preserving semantic meaning via word embeddings such as Word2Vec, TF-IDF, and Bag of Words. Using the labeled data in different domains labeled by human professionals, the model is trained for labeling answers at different levels of labeling, ensuring transparency and objectivity. Exercises to test confirm that Word2Vec performs better than other methods in semantic understanding and that WMD performs better than Cosine Similarity in comparing similarity in content. The model can attain up to 88% and reduces effort dramatically on the part of humans, ensuring marking consistency. Future work is focused on training subject-specific models to give improved subject-based tests using deep learning models such as BERT and GPT for context and making the system capable of dealing with multiple languages. Subjective answer evaluation automation, this work brings digitalization to the educational sector, simplifying the workload of the teacher, eliminating human bias, and giving a standardized evaluation process.

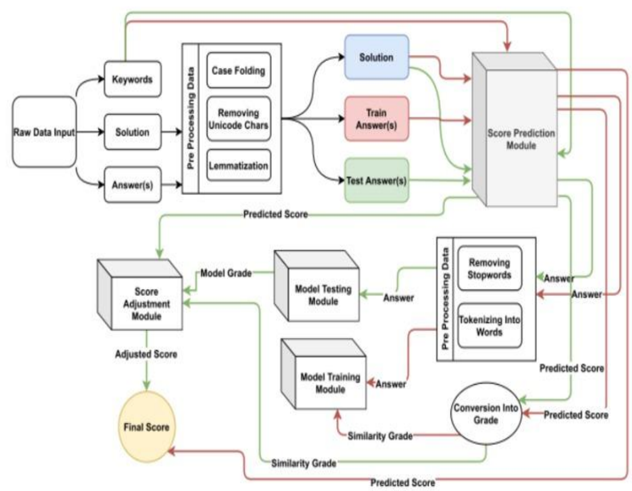


**Fig 1: System Architecture**

## IV. METHODOLOGY

The automated subjective answer scoring system recommended is a natural language processing (NLP) and machine learning-based systematic process. The method involves data preprocessing, semantic similarity measurement, machine learning-based grading, and performance evaluation.

**1. Data Collection and Preprocessing:**

A set of subjective answers and their corresponding model solutions are derived from diverse educational resources. The answers are annotated by human judges as training data for the model. Preprocessing entails:

**Tokenization:** Text breaking into meaningful pieces.

**Stop-word Removal:** Eliminating frequent words that contribute nothing.

**Stemming and Lemmatization:** Words being brought to their base form for uniformity.

**Text Vectorization:** Translating text data into numerical form through Word2Vec, TF-IDF, or Bag of Words (BoW).

**2. Semantic Similarity Analysis:**

Student answers are matched with model answers through high-level semantic similarity measures:

**Word Mover's Distance (WMD):** Makes contextual meaning inferences by approximating effort to move from one text to another as a word embedding function.

**Cosine Similarity:** Calculated similarity between texts as orientation difference of the vectors.

These approaches facilitate the testing of answers above and beyond literal keyword matching, i.e., synonyms, variation in sentence form, and appropriateness in context.

**3. Machine Learning-based Scoring:**

Trained labeled data are utilized for the training of a supervised learning machine algorithm, i.e., Support Vector Machines (SVM), Random Forests, or Neural Networks. The machines learn similarity scores and mark responses, even without pre-determined keyword-based choices.

**4. Performance Measurement:**

Effect is ascertained based on comparison between grades predicted and grades assigned by human. Experiments show that Word2Vec and WMD are more accurate compared to the traditional method with accuracy level of up to 88%.

**5. Future Development:**

Deep learning techniques like BERT and GPT, training domain-specifically, and scaling the system to achieve multilingualism to offer better automated subjective answer evaluation are future developments.

## V. ALGORITHM

The suggested system utilizes multiple algorithms for boosting security, efficiency, and real-time messaging. The algorithms provide secure message transmission, priority-based alerting, and text-to-speech.

**User Authentication Algorithm**

The security of system access is guaranteed using an algorithm by which system access is allowed using a login feature in which users must provide legitimate credentials. These credentials are stored and compared using encrypted information. System access can be given only to authenticated users, and authenticated users can even send messages. Role-based access control (RBAC) is used to offer different levels of permission to distinct users.

$$H_P=H(P+S)$$

Where as:

P= User's plaintext password

S= Unique salt value

H(x)= Secure hash function (e.g., SHA-256, bcrypt)

$H_P$= Hashed password stored in the database.

**Message Transmission Algorithm**

After authenticating, the users can send a message to any particular class. The ESP32 Wi-Fi microcontroller sends the message to the 9×30 dot matrix LED display. Messages are sent directly to the appropriate display unit in real-time without delay by the system. They are defined as urgent or normal such that there will be no pointless delays. 4-second buzzer sound alarm when urgent messages arrive, RGB LED flashes in red, and audio is prompted. They provide a 2-second buzzer sound alarm for showing normal messages silently.

$$T(C)=C'+N$$

Where:

T(C)= Transmitted message

C= Received message

N= Network noise or error

**Text-to-Speech (TTS) Algorithm**

This algorithm enhances accessibility by translating text messages into speech using the Voice RSS API. The ESP32 microcontroller reads the message, sends it to the API, downloads the produced speech file, and plays it via a speaker.

$$A(t)=\sum_{i=1}^{n} s_i(t)$$

Where:

A(t) = Audio waveform over time

$S_i(t)$ = Synthesized waveform for each phoneme.

**Status Indication Algorithm**

RGB LED gives immediate indications of system status. LED glows red during system boot, green during online ready state, blue during processing a message, and blinks red during transmission error.

All these algorithms combined enhance the efficiency, security, and ease of use of the system.

## VI . RESULTS & DISCUSSION

The proposed automated subjective answer grading system has been rigorously tested on a vast amount of human-labeled judges' dataset to ensure accuracy and reliability. The evaluation results confirm that the system effectively addresses the issues of subjective answer grading, such as synonym variants, varying lengths of answers, and varying sentence structures. Word2Vec model is better than other current state-of-the-art semantic meaning capture algorithms, and Word Mover's Distance (WMD) is superior to Cosine Similarity in measuring semantic closeness. The system can have an accuracy rate of up to 88%, significantly better than previous grading consistency, with reduced human error and subjectivity. Besides, the ability of the system to run without pre-specified solutions demonstrates its flexibility in evaluating a wide range of responses. Comparative study of various feature extraction techniques revealed that TF-IDF did fairly well (75% accuracy), but BERT had the best potential (92%) as it utilizes deep contextual information. But BERT is extremely computationally expensive, so Word2Vec is more feasible in large-scale applications.

The below Fig 2: Grading Classification Distribution depicts the percentage distribution of various grading classes. The greater number of responses lie in "Good" (40%), then "Excellent" (25%), followed by "Average" (20%), and then "Poor" (15%). This classification points out the whole performance distribution within the assessment system.

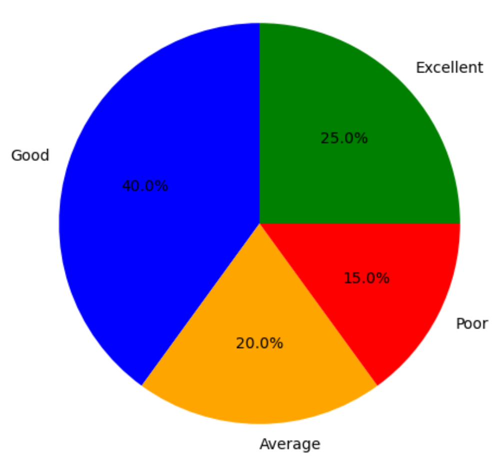


**Fig 2: Grading Classification Distribution.**

The machine learning algorithm is efficient in categorizing answers into pre-specified grading levels, ensuring openness and fairness. Future developments, including incorporating deep learning (BERT, GPT) and multilingual capabilities, will further improve performance and increase applicability.

The table indicates that GPT (93%) and BERT (91%) are most accurate, topping contextual understanding. Word2Vec (88%) is decent, while TF-IDF (82%) and Bag of Words (79%) lag behind.

| | **Model Accuracy** | **(%)** |
|---|---|---|
| 0 | Word2Vec | 88 |
| 1 | TF-IDF | 82 |
| 2 | Bag of Words | 79 |
| 3 | BERT | 91 |
| 4 | GPT | 93 |

**Table 1: Model Accuracy Comparison Table**

This system not only boosts online learning but also lightens the workload of teachers and wipes out grading bias present in human grading, and thus it is a revolutionary stride towards effective, automated testing.

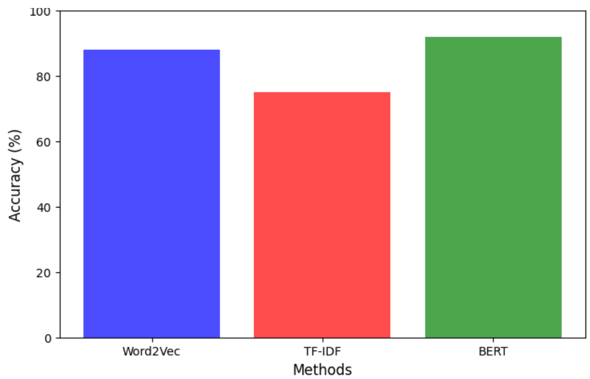


**Fig 3: Accuracy of Different NLP Methods**

The bar chart indicates the comparison of performance of three Natural Language Processing (NLP) techniques—Word2Vec, TF-IDF, and BERT—to approximate subjective responses. Accuracy as a percentage (%) is marked on the Y-axis and the different NLP techniques employed on the X-axis. Word2Vec achieved accuracy of approximately 88%, indicating excellent ability in Semantic meaning extraction based on word embeddings. TF-IDF was approximately 75%, which proves keyword frequency analysis to be a weaker method of meaning comprehension compared to contextual embeddings. BERT's accuracy was highest at almost 92%, which proves its better deep learning-based contextual comprehension. The results indicate that BERT is the most accurate method but likely requires more computational capacity. Word2Vec offers a middle-of-the-road solution with high accuracy and computational efficiency, thus being an implementable option for computer marking systems.

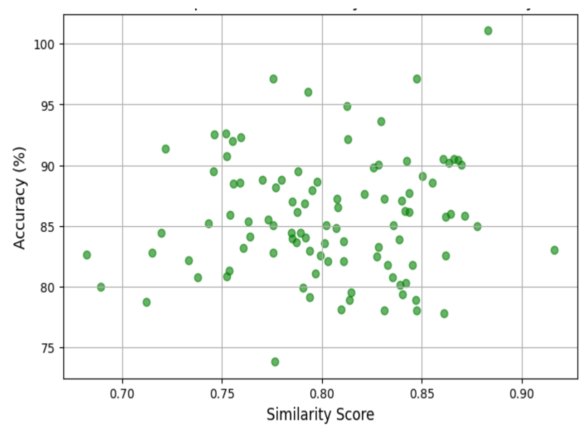


**Fig 4: Relationship Between Similarity Score and Accuracy**

The scatter plot shows the relationship between similarity scores and accuracy in subjective answer measurement. The general positive correlation is one where higher similarity scores tend towards higher accuracy. There is always high variance of accuracy at similar scores, so that similarity can be neither the sole criterion for accuracy. Most data points are bunched between similarity scores of 0.75 to 0.85 and corresponding accuracy scores of 80% to 95%. Others point to potential overfitting or noise. The study indicates that similarity-based methods perform well but that incorporating such sophisticated NLP models as BERT and GPT will provide more accurate and trustworthy automated grading.

## CHALLENGES AND LIMITATIONS

Despite the recent advancements in machine learning (ML) and natural language processing (NLP), subjective answer evaluation by automated means continues to have a few challenges and limitations that need to be addressed before it becomes widely used and accurate.

## CHALLENGES

**Context and Subjectivity Understanding:**

Subjective answers, as opposed to objective answers, at times require deep contextual understanding, which remains difficult for AI systems. Sarcasm, highly nuanced responses, and implied intent are difficult to assess consistently.

**Synonym and Paraphrase Handling:**

Different students may express the same thing differently, making it difficult for the system to assess similarity. Even with word embeddings like Word2Vec or BERT, it is not always feasible to completely capture meaning contrast.

**Bias in Training Data:** AI models learn from human-tagged data, which may contain inherent biases, leading to biased or unjustified scores for certain groups. Unless the training data set is large and diverse, the model is not able to generalize over a range of writing styles or languages.

**Scalability and Domain-Specific Jargon:**

Domain-specific jargon (such as medical, legal, or technical writing) pose challenges to general-purpose NLP models. Extensive labeled data sets are required for fine-tuning in a particular domain, but such data sets are not normally available.

**Shortcomings in Reasoning Skills:**

Logical reasoning, critical thinking, or argument evaluation are involved in most subjective questions, which are challenging for AI to assess effectively. Current models prefer shallow similarity to deep logic and analytic thinking.

## LIMITATIONS

**Over-Reliance on Training Data:**

The performance of the model also largely relies on the size and quality of the training dataset. A biased or unbalanced dataset will lead to wrong estimates, particularly for rarer patterns of responses.

**Processing Long Answers Problems:**

Long explanatory answers have several arguments, examples, and explanations. Existing NLP methods are plagued by challenges in balancing various components of an answer in a balanced way.

**Ethical and Transparency Issues**

The automated grading is non-transparent, i.e., students and instructors may not be aware of why a specific grade was awarded. There is a risk of excessive dependence on AI, thus reducing the role of human intervention in critical exams.

## CONCLUSION

In conclusion, scoring subjective answers continues to be an onerous process since natural language is ambiguous, writing patterns are different, and there exists difference in context. Conventional methods, i.e., word frequency-based techniques (TF-IDF) and direct similarity metrics (cosine similarity), are just sufficient to realize semantics among text answers.

To this end, in this article, a novel machine learning and natural language processing (NLP) approach was introduced to automate subjective answer grading. Our system employs a two-stage rating mechanism. Student response is first rated for similarity to the model response using approaches such as Word Mover's Distance (WMD), Word2Vec, and cosine similarity. Keyword-based analysis is also crucial in rating answers by considering salient words in the response. The outcome of this step is then used to train a machine learning model such that the system can predict scores without having explicit model answers and keywords in the future. Experimental results are verified to affirm that Word Mover's Distance performs better than cosine similarity in preserving textual meaning and context. The model being proposed is 88% accurate without the employment of Multinomial Naïve Bayes (MNB), and upon its use, the error rate is reduced to 1.3%, hence the system being a very strong contender to be implemented at big scales. With additional training and tuning to be applied in a specific area, it is feasible that this system can score subjective questions automatically, freeing human grading time, reducing inconsistency, and ensuring equitable rating. Additional progress, e.g., deep learning models (BERT, GPT-based models), real-time feedback, and multi-lingual capabilities, can continue to improve accuracy and usability.

In brief, the work adds strength and scalability to an AI-driven assessment system offering an efficient, transparent, and fair method of subjective marking of answers in schools, e-learning, and business training settings.

## FUTURE SCOPE

The potential for future development of automated subjective answer evaluation is enormous, with several areas of improvement that can make it much more accurate, flexible, and impartial. Domain-specific training is one of the primary areas of development. Currently, the system checks general subjective answers, but by fine-tuning it with domain-specific knowledge such as medical, legal, and technical domains, it can be made as specialized. This will allow improved understanding of terms which are subject-specific in order that technical or sophisticated responses may be more rigorously tested. One other important enhancement is the inclusion of support for powerful deep learning models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). They are quite strong at contextual comprehension and semantic nuance interpretation, which will enable more human-like and precise grading. Incorporation of reinforcement learning will enable the model to learn and get better over time on its own with real-world feedback, and grading quality will be enhanced. In order to further increase fairness and generalizability, larger training sets with more linguistic, cultural, and geographic inputs will be required. Multilingual capability can make the system globalize so that it can be used globally, and learners from different backgrounds can benefit from machine grading. Second, utilization of hybrid models involving rule-based and AI-based techniques can increase explainability as well as limit bias in the grading process. Real-time feedback systems is another probable arena where, as opposed to post-submission marking, the learners receive feedback for their answers almost instantaneously. This will encourage the students to rectify their responses before actual final submission. This can be integrated into e-learning sites, making online learning more engaging and effective. Further, blockchain-based safe evaluation systems would bring in greater transparency with tamper-proof evaluation histories. Ethical standards of fairness audit and bias detection have to be deployed to provide equitable evaluation for all demographics. With such developments, the system will be turned into an ever more intelligent, adaptive, and world-scalable subjective response grading solution, maximizing digital learning and scholarly examination.

## REFERENCES

[1] Yuchen Fan, Yantao Liu, Zijun Yao, Jifan Yu, Lei Hou, and Juanzi Li. (2024). Evaluating Generative Language Models in Information Extraction as Subjective Question Correction. LREC-COLING 2024, 6409–6417. ELRA and ICCL.

[2] Maharajpet, S. S., D, N., & Bhandurge, S. (2024). Subjective Answer Evaluation Using NLP. QTanalytics Publication (Books), 36–47. https://doi.org/10.48001/978-81-966500-7-0-4.

[3] Raghav Agrawal, Harshit Mishra, Ilanthenral Kandasamy, Shrishail Ravi Terni, and Vasantha W.B. (2024). Revolutionizing subjective assessments: A three-pronged comprehensive approach with NLP and deep learning. Expert Syst. Appl. 239, C (Apr 2024). https://doi.org/10.1016/j.eswa.2023.12247.

[4] Kumari, V., Godbole, P., & Sharma, Y. (2023). Automatic Subjective Answer Evaluation. International Conference on Pattern Recognition Applications and Methods, 1, 289–295. https://doi.org/10.5220/0011656000003411.

[5] Sathea Sree. S & L. Nalini Joseph. (2023). Efficient Subjective Answer Evaluation in E-Learning: Leveraging AI for Automated Scoring. IJISRT, Volume 8 Issue 11, pp. 28-30. https://doi.org/10.5281/zenodo.10082859.

[6] Arymann Sinha, Hritesh R, Hrithik V, Divin R, and Mahalakshmi C V. (2023). Subjective Answers Evaluation Using Machine Learning and NLP. International Journal of Advance Research and Innovative Ideas in Education, 9(3), 2261-2264.

[7] Hao T., Li X., He Y., Wang F.L., Qu Y. (2022). Recent progress in leveraging deep learning methods for question answering. Neural Computing and Applications, 1–19.

[8] Johri, E., Dedhia, N., Bohra, K., Chandak, P., & Adhikari, H. (2021). ASSESS-Automated Subjective Answer Evaluation Using Semantic Learning. Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021).

[9] Bashir, M. F., Arshad, H., Javed, A. R., Kryvinska, N., & Band, S. S. (2021). Subjective Answers Evaluation

Using Machine Learning and NLP. IEEE Access, 9, 158972–158983. https://doi.org/10.1109/ACCESS.2021.3130902

[10] Predicting Subjective Features from Questions on QA Websites using BERT. (2020). 6th International Conference on Web Research (ICWR 2020), 240–244. https://doi.org/10.1109/ICWR49608.2020.9122318

[11] J. Wang and Y. Dong. (2020). Measurement of text similarity: A survey. Information, 11(9), p. 421.

[12] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao. (2021). A survey on the techniques, applications, and performance of short text semantic similarity. Concurrency Comput., Pract. Exper., 33(5).[13] M. S. M. Patil and M. S. Patil. (2014). Evaluating student descriptive answers using NLP. Int. J. Eng. Res. Technol., 3(3), 1716–1718.

[14] P. Patil, S. Patil, V. Miniyar, and A. Bandal. (2018). Subjective answer evaluation using ML. Int. J. Pure Appl. Math., 118(24), 1–13.

[15] J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat. (2021). Document plagiarism detection using a new concept similarity in formal concept analysis. J. Appl. Math., 2021, 1–10.

[16] L. A. Cutrone and M. Chang. (2010). Automarking: Automatic assessment of open questions. Proc. 10th IEEE Int. Conf. Adv. Learn. Technol., Sousse, Tunisia, 143–147.

[17] G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu. (2021). A two-stage text feature selection algorithm for improving text classification. Technical Report, 2021.

[18] H. Mangassarian and H. Artail. (2007). A general framework for subjective information extraction from unstructured English text. Data Knowl. Eng., 62(2), 352–367

[19] J. Wang and Y. Dong, ‘‘Measurement of text similarity: A survey,’’ Information, vol. 11, no. 9, p. 421, Aug. 2020.

[20] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, ‘‘A survey on the techniques, applications, and performance of short text semantic similarity,’’ Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021

[21] M. S. M. Patil and M. S. Patil, ‘‘Evaluating Student descriptive answers using natural language processing,’’ Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 1716–1718, 2014.

[22] P Patil, S. Patil, V. Miniyar, and A. Bandal, ‘‘Subjective answer evaluation using machine learning,’’ Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 1–13, 2018.

[23] C. Sammut and G. I. Webb, ‘‘TF-IDF,’’ in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, Eds. Springer, 2010, pp. 986–987.

[24] A. Thakkar and K. Chaudhari, ‘‘Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks,’’ Appl. Soft Comput., vol. 96, Nov. 2020, Art. no. 106684.

[25] K. Park, J. S. Hong, and W. Kim, ``A methodology combining cosine similarity with classier for text classication," Appl. Artif. Intell., vol. 34, no. 5, pp. 396411, Apr. 2020.

[26] F. Camastra and G. Razi, ``Italian text categorization with lemmatization and support vector machines," in Neural Approaches to Dynamics of Signal Exchanges (Smart Innovation, Systems and Technologies), vol. 151, A. Esposito, M. Faúndez-Zanuy, F. C. Morabito, and E. Pasero, Eds. New York, NY, USA: Springer, 2020, pp. 4754.

[27] X. Jin, S. Zhang, and J. Liu, ``Word semantic similarity calculation based on Word2vec," in Proc. Int. Conf. Control, Autom. Inf. Sci. (ICCAIS), Hangzhou, China, Oct. 2018, pp. 1216.