Subjective Evaluation of Management Information System Using Latent Dirichlet Allocation Approach in Machine Learning

Samundiswary S
Department of information Tecnology
Assistant Professor
SIESGST
Navi Mumbai
samundiswarys@sies.edu.in

Ayush Dodake
Department of information Tecnology
Student
SIESGST
Navi Mumbai
ayushdit120@gst.sies.edu.in

Kirti Ekhande
Department of information Tecnology
student
SIESGST
Navi Mumbai
kirtieit120@gst.sies.edu.in

Pratik Mohite
Department of information Tecnology
Student
SIESGST
Navi Mumbai
pratikmit120@gst.sies.edu.in

***Abstract*—Subjective Evaluation in education is a difficult task, especially in technical fields like Management Information Systems (MIS). This study presents a thorough method that uses Natural language processing (NLP) and machine learning (ML) approaches to expedite the assessment process. Our main goal is to streamline the assessment of Management Information Systems (MIS) papers, which is an engineering subject. We adopted Topic Modelling to identify key themes within our dataset, employing the Latent Dirichlet Allocation (LDA) algorithm. By doing so, we can identify underlying themes and offer a systematic interpretation of the information included in student submissions. We added model responses to our system to improve the assessment process even more. Using the cosine similarity NLP approach, we compared these model answers with student answers to measure the semantic similarity between the two. This comparison study was an essential part of our methodology for evaluation. Our novel hybrid approach combines the dataset's intrinsic scores with similarity scores derived from comparing the responses of the model answers and the student's answers. With the use of this dual-scoring system, MIS papers can be evaluated with greater accuracy and nuance, giving a comprehensive perspective of the student's performance. With a focus on the difficulties associated with subjective evaluation in technical fields, this research adds to the continuing discussion on the integration of ML and NLP in educational assessments. Our proposed hybrid methodology demonstrates its effectiveness in providing an objective and insightful evaluation of MIS papers, with implications for further applications in subjective assessments across diverse educational domains.**

Keywords—Subjective Evaluation, Management Information Systems, Machine Learning, Natural Language Processing, Topic Modeling, Latent Dirichlet Allocation (LDA), Cosine Similarity, Hybrid Approach.

I. INTRODUCTION

One important area of study in education is the subjective question-and-answer method of evaluating a student's performance and skills. Subjective tests present a unique set of difficulties for learners and teachers, largely because of their intrinsic complexity brought about by the subjective nature of answers. Subjective answers necessitate a careful analysis of each word, which burdens the mental health, fatigue, and objectivity of human graders significantly more than objective answers, which are readily assessed by machines. Having said that, there are a few more crucial distinctions between subjective and objective responses. They also carry a lot more context and demand a lot more concentration and objectivity from the teacher grading them. They also carry a lot more context and demand a lot more concentration and objectivity from the teacher grading them. Hence, there is a strong need to move toward automated systems because it takes a lot of time and resources to manually evaluate subjective responses. With the increasing accessibility of automated objective assessments, we are concentrating on creating a system that can navigate the complexities of subjective responses.

Human judgment is subject to emotional variation, which might affect the assessment's quality. When evaluation is carried out by computers using intelligent procedures, marking is uniform because every student employs the same inference engine. Numerous studies have been conducted on subjective answer evaluation in one form or another. This study counts the noun phrases that appear in the documents, compares keywords in the responses to the text, maps the text's context to the solution's context, and calculates the degree of similarity between different words, texts, and even documents. With this project, we aim to improve the effectiveness and integrity of subjective paper assessments, which will ultimately lead to a more trustworthy and efficient process for educational evaluation.

Natural Language Processing (NLP) has become a transformative science at the nexus of linguistics and artificial intelligence, enabling machines to understand, interpret, and produce human language. Natural language processing (NLP) has great potential to transform the evaluation of subjective papers in the field of education. This process has always been labor-intensive and time-consuming. The textual data can be compared using a range of techniques, including idea graphs, ontologies, latent semantic structures, and document similarity. Similarity, the presence of keywords, structure, and language are among the evaluation factors that determine the final score [1] [2]. Although this subject has been discussed previously [3][4] and [5], there is still room for improvement, some of which are addressed in this study.

Because Natural Language is by its very nature ambiguous, it is difficult to evaluate machine responses to such questions. To prepare for analysis, the method involves multiple preparation processes, including tokenization and data cleansing [6]. Our study is based on natural language processing techniques and similarity metrics such as cosine similarity and word mover's distance [6], as well as text representation methods including TF-IDF, Bag of Words, and word2vec. Similarly, the presence of keywords, structure, and language are some of the factors that go into determining the final score.[6]

Although this challenge has been addressed in previous attempts, there is still much room for improvement, as this paper discusses. We examined various methods used in the past to investigate subjective answer evaluation and text similarity assessment [7][8]. We do, however, point out several important drawbacks when handling subjective responses. One prevalent challenge is the existence of synonyms in the studies [], leading to potential inconsistencies. These studies also frequently show a wide range of potential answer lengths, which adds even more complexity to the assessment procedure. These drawbacks highlight the necessity of improving current strategies and creating more reliable techniques for evaluating subjective answers. Various evaluation metrics, including F1-score, Accuracy, and Recall, are employed to assess how well different models perform in comparison to one another [9].

The objective is to contribute to a more accurate and efficient assessment paradigm in educational settings by improving evaluation efficiency and offering a nuanced understanding of the content and structure of responses.

II. RELATED WORK

A thorough analysis of previous studies in the field was necessary to comprehend the terrain of subjective answer evaluation. A careful analysis of a range of research papers shed light on the difficulties and complexities involved in grading subjective answers.We were inspired to investigate subjective answer evaluation as a research topic because we saw a gap in current methods and the possibility for creative solutions to these problems. The recent developments in machine learning and natural language processing (NLP) influenced the choice to concentrate on automated evaluation techniques using sophisticated NLP methodologies. In keeping with the overarching objective of improving the effectiveness and equity of educational evaluations, we aimed to add to the continuing conversation about automating subjective assessments.

The evaluation of subjective responses has always been difficult, which has led to the investigation of numerous approaches during the previous fifteen years. Numerous methods, such as Natural Language Processing, Topic Modeling, Latent Dirichlet Allocation (LDA), Latent Semantic Analysis, Generalized Latent Semantic Analysis, Bayes theory, and K-nearest neighbor, have been tested to solve the issue. The three main categories of these approaches are classification techniques, clustering techniques, and natural language processing techniques. For instance, Landauer developed the Intelligent Essay Evaluator in 2003[18], which produced results with an accuracy of 60–90% by using latent semantic analysis. Using probabilistic LSA [19], Kakkonen improved this and produced an automatic essay evaluator. By focusing on vectors, generalized LSA [20] expands on the method and produces more accurate results.

In addition to these techniques, we investigated other classification strategies such the maximum entropy, K-nearest neighbor, Bayes theorem, and others. For instance, Rudner's 2002 implementation of the Bayes theorem yielded an accuracy rate of 80% [15]. The K-nearest neighbor clustering method randomly selects cluster heads and builds clusters based on distances from those heads, with a 76% accuracy rate [16]. Crater, a tool that employs Maximum Entropy, achieved an 80% accuracy rate for short answers when compared to human graders.[17] Another approach considered was BLEU (bilingual evaluation understudy), an algorithm designed to assess the quality of machine-translated text. However, BLEU did not do well when used to evaluate individual sentences; it received a score of only 50% [21].

On the other hand, Topic Modeling describes an effective natural language processing (NLP) method used to identify relevant topics and patterns within a sizable set of arbitrary responses. Finding underlying themes in textual data is the main goal since it enables a more in-depth comprehension of the content. This method helps to classify and pinpoint popular topics, terms, or ideas that appear in subjective answers. Topic modeling in this project is a sophisticated approach that goes beyond mere keyword matching. It dives deep into the semantic relationships and contextual nuances present in subjective responses, contributing to a more nuanced and accurate evaluation of these answers. The study also examined topic modeling techniques, which are ways to find topics in a group of texts. Finding hidden thematic structures in large datasets is made easier with the use of topic modeling. It consists of methods such as Non-Negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA). Assuming that documents are mixtures of topics and topics are mixtures of words, LDA is a probabilistic model. In contrast, NMF factorsizes the provided document-term matrix into two lower-dimensional matrices that correspond to the distributions of the document-topic and topic-term. These topic modeling techniques offer insightful information for additional analysis and interpretation, fostering a more nuanced understanding of the underlying themes in textual data [24].

Latent Dirichlet Allocation (LDA) is the most commonly used topic modeling method across a wide number of technical fields. Latent Dirichlet Allocation, or LDA, operates under the premise that every document can be modeled as a combination of distinct topics, each of which is a combination of different words. The model considers that each topic has a probability of containing a certain word, and that there is an underlying probability that a document will be about a given topic. This makes it easier to find the hidden themes or subjects in a group of documents [22]. By employing rigorous training and evaluation processes, the study achieved an 87% coherence between LDA-generated topics and human judgment, demonstrating the reliability of the method [23]. The importance of supervised training and assessment in improving the validity and interpretability of LDA-generated topics is emphasized in the paper. The results imply that by reflecting multiple themes in texts, extracting new themes not highlighted by human coders, and minimizing human bias, LDA can provide benefits over manual content analysis [23].

Therefore, as topic modeling can reveal hidden themes and patterns within a collection of documents, we decided to use it as our method for evaluating subjective answers. By using topic modeling, and more specifically the Latent Dirichlet Allocation (LDA) algorithm, we can examine how topics are distributed among different documents. LDA is appropriate for comprehending the underlying structure of subjective responses because it assumes that documents are mixtures of topics and that topics are mixtures of words. To achieve our objective of automating the evaluation of descriptive answers, LDA's probabilistic nature makes it a useful tool for capturing the subtleties and complexity of human language.

To improve the assessment of subjective responses, we utilize a variety of Natural Language Processing (NLP) methods to gauge how similar student and model responses are. NLP, a discipline at the nexus of linguistics and computer science, gives us the tools to computationally analyze and comprehend human language. Our methodology aims to reduce the discrepancy between student responses and the model by employing a variety of natural language processing (NLP) techniques, such as tokenization, stop word removal, parts-of-speech tagging, lemmatization, stemming, case folding, cosine similarity, word movers' distance, TF-IDF, and others. This comprehensive evaluation process outperforms traditional methods. The use of natural language processing (NLP) not only makes it easier to comprehend textual content more deeply, but it also advances the more general objective of automating and improving the subjective answer evaluation.

Prominent publications presented a novel method of formal concept analysis (FCA)-based plagiarism detection. The groundwork for comprehending the nuances of textual similarity was established by this work.[12] We also examined word embeddings and their use in determining text similarity. When measuring the dissimilarity between two text documents, the incorporation of Natural Language Processing (NLP) techniques—more specifically, Word Movers Distance—became apparent as a major factor[13]. Additionally, the application of the word2vec method outperformed conventional word embedding methods, highlighting the significance of semantics in automated assessments [14]. This study shed light on the current shortcomings in handling subjective responses, highlighting issues with synonyms and differences in response durations as the two main roadblocks. We saw agreement on the significance of tokenization in dividing text into meaningful chunks for later analysis. Furthermore, it was found that eliminating stop words was an essential preprocessing step to reduce noise and enhance the quality of textual data. Also, it has been observed that lemmatization—which is the process of breaking down words into their dictionary or base form—improves the interpretability and precision of NLP models. All things considered, these methods are important for streamlining NLP pipelines and deriving valuable insights from textual data in a variety of contexts and applications.

We made a strategic decision to concentrate on Management Information Systems (MIS) for our project because of the crucial part that MIS plays in contemporary organizational dynamics. Information technology, human resources, and business processes work together harmoniously to form MIS, which serves as a catalyst for gathering, storing, and processing data [10]. Producing data-driven insights to support managerial decision-making is its main objective. Organizations are depending more and more on MIS in today's competitive environment to help them manage the intricacies of data and gain insightful knowledge. MIS as the Key to Streamlined Decision-Making: A Research-Driven Decision. MIS, in this context, emerges as the linchpin for achieving the highest quality decisions across all levels of management.[11]

It is essential to educational institutions because it covers a wide range of information management and decision-making activities. Every company that wants to stay competitive in the market must have a management information system (MIS).[10]

III. PROPOSED SYSTEM

The proposed system consists of the following modules: results predicting, result predicting, preprocessing, model training, similarity assessment, and data collecting and annotation. The initial inputs that are acquired from the user are answers and solutions. Figure 1 shows our system architecture.

The solution acts as a standard by which the replies from students are measured and assessed. It represents the perfect, thorough response to a prompt or inquiry, carefully created by the instructor or evaluator. When generating the solution, all important terms, ideas, and situations related to the query are carefully included, guaranteeing a thorough portrayal of the desired material. Clarity and coherence are achieved by clearly outlining each important topic element in different lines or paragraphs within the solution. The solution serves as the basis for evaluation, establishing the benchmark for judging the caliber and thoroughness of students' responses and directing the grading procedure in the direction of uniformity and objectivity.

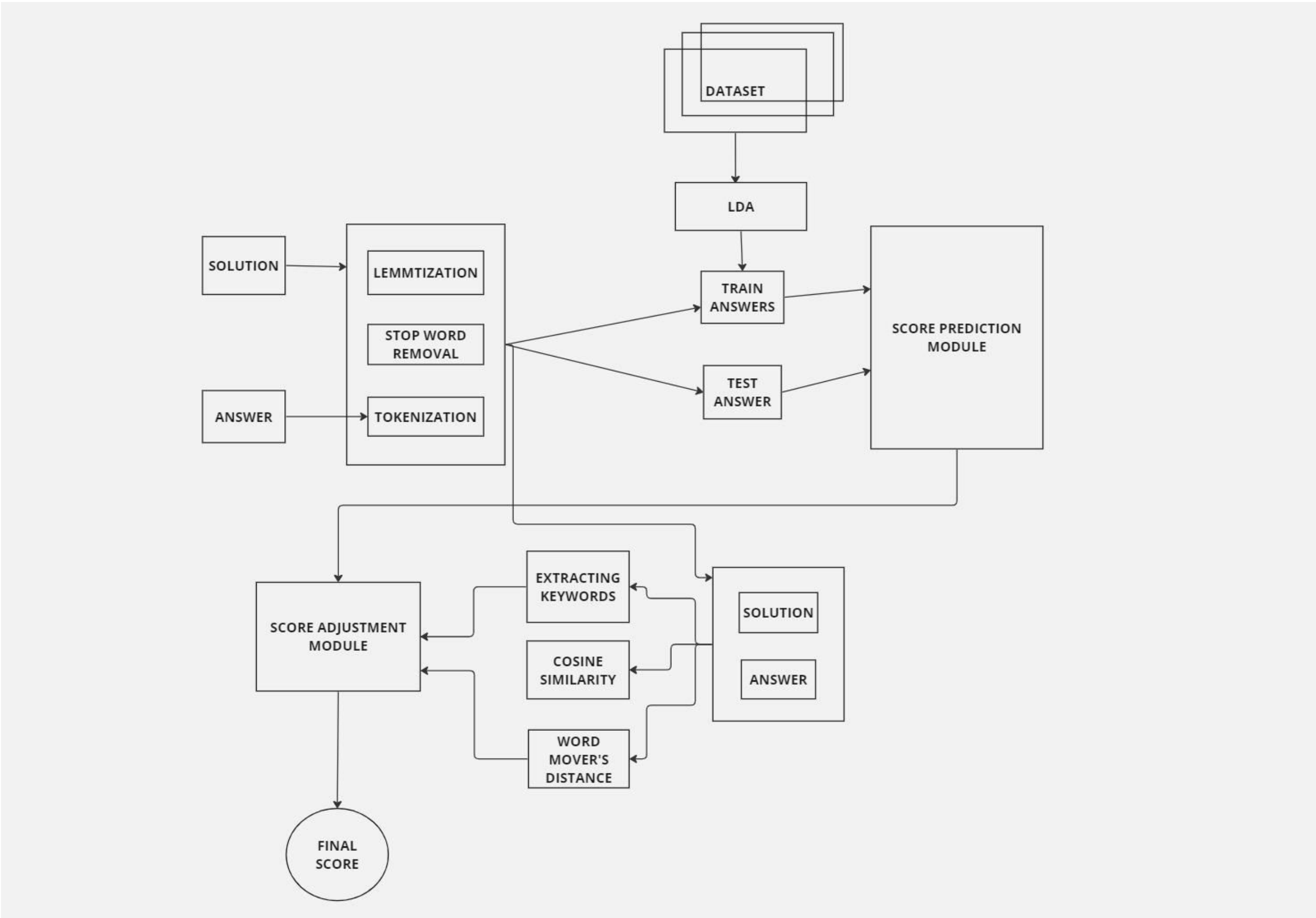

Fig. 1.*System Architecture*

The answer serves as the foundation for evaluation in the assessment process and reflects the student's subjective interpretation and understanding of the given question. The length of the response, which usually ranges from one to several sentences, is determined by the difficulty of the topic and the student's unique writing style. In contrast to the solution, the answer represents the student's comprehension and articulation of the subject matter and might not include every term or concept covered. It frequently uses synonyms or different terms than the solution, therefore processing entails close attention to semantic details. As the primary focus of evaluation, the answer provides insights into the student's grasp of the subject matter, their ability to articulate ideas, and their aptitude for critical thinking and analysis. Therefore, thorough examination and interpretation of the answer are essential in gauging the depth and quality of the student's response.

Preprocessing is done on both the student's answer and the correct model solution before the assessment to guarantee uniformity and standardization in the analytical procedure. Preprocessing entails improving the text's organization and substance to make it ready for analysis. Preprocessing serves the purpose of eliminating any unnecessary components from the text that can distort the evaluation's findings. With preprocessing methods like tokenization, lemmatization, and stop word elimination, we want to reduce the amount of time spent on assessment and concentrate only on the answers' relevant content. This guarantees an impartial and equitable assessment, enabling to appropriately evaluate students' answers according to their understanding and expression of the material.

Lemmatization is the first stage in preparing the answer as well as the model solutions. This method seeks to standardize and maintain uniformity throughout the analytical process by breaking words down to their most basic or root form. Variations resulting from different word forms are reduced by normalizing the text by lemmatization, which makes it possible to compare the solution and the student's response fairly while evaluating them. To provide a consistent evaluation, terms such as "running," "ran," and "runs" can all be lemmatized to the base form "run." This procedure reduces semantic disparities caused by word variants, making it possible to assess student answers more precisely.

After lemmatization, stop word removal is used to remove terms that are often used but have little meaning from the student's answer as well as the solution. Stopwords like "the," "is," and "are" are common in natural language but don't really add anything to the text's actual meaning. Eliminating these stopwords improves the assessment process's clarity and relevance by shifting the analysis's attention to the answer's significant substance. By eliminating extraneous data and noise from the analysis, this stage allows us to focus on the most important components of the student's response as well as the solution.

Tokenization, the last preprocessing step in our project, divides the text into discrete words or tokens. This procedure makes it easier to analyze the text in detail and see how each word contributes to the solution. Tokenizing the text makes the analysis easier to handle and makes it possible to find important phrases, ideas, and trends in both the student's response and the solution. Tokenization also establishes the framework for additional analysis and feature extraction, supplying the essential underpinning for later assessment methods. The material is divided into distinct parts by tokenization, making it possible to conduct a thorough and methodical evaluation of students' replies.

The project's initial focus is on gathering a diverse dataset, which primarily consists of MIS questions. We compiled a large dataset of various Management Information Systems (MIS) documents to aid in the training and testing of our proposed assessment model. The lack of publicly accessible labeled subjective question-answer corpora relevant to the MIS domain led to the construction of this dataset. Identifying this vacuum, we set out to gather a diverse and comprehensive set of subjective question-answer pairs about management information systems. The main goal was to compile information from reliable sources where MIS-related conversations and questions are common, such as academic forums, learning websites, and scientific publications. We ensured a wide representation of issues and viewpoints within the MIS domain by methodically gathering subjective question-answer pairs from a variety of sources using web crawling techniques. Essential metadata, including question ID, subject, and answer, are included with every entry in the dataset to facilitate thorough analysis and assessment of student replies. Our goal in carefully selecting this dataset was to give a reliable tool for training and confirming the suggested assessment model, which would improve the MIS paper assessment procedure.

We used topic modeling in our methodology as a critical first step in deciphering the underlying organization of our MIS dataset. The Latent Dirichlet Allocation (LDA) algorithm was utilized in order to find latent themes or topics that were present in the MIS questions. To put this into practice, we first preprocess the dataset to make sure the text data is consistent and uniform. The dataset is then modeled using LDA, where the algorithm allocates a distribution over a fixed number of topics and a distribution over words to each document (MIS question). This gives us insights into the main concepts discussed in the subject and enables us to identify the latent topics hidden in the MIS questions. LDA is used for a number of purposes. First off, by grouping MIS questions into logical topics, it aids in dataset structuring and makes a more methodical analysis easier. Second, LDA helps us to accurately assess subjective answers by allowing us to capture the underlying concepts and relationships within the MIS domain. We make sure that our evaluation process is informed by a thorough comprehension of the MIS subject matter by training the model on the LDA algorithm. The dataset is trained and validated using standard machine learning methods. We separated the dataset into training and testing sets in order to evaluate the performance of the LDA model. Eighty percent of the dataset is used for training, while twenty percent is used for testing. While the training set is used to train the model on the latent subjects found in the MIS questions, the testing set is used to assess the model's capacity to adapt to new data. Our objective in putting ourselves through this demanding training and testing procedure is to develop a framework that reliably and accurately evaluates subjective responses in the MIS domain.

The LDA model is applied to evaluate the answer's relevance to the latent subjects found in the dataset using the preprocessed student response as input. Predicted scores are the result of the score prediction module, which offers a preliminary assessment of the student's performance based on subject alignment. The predicted score functions as an initial evaluation of the student's performance and can be further modified or improved upon in light of additional factors or considerations, such as the outcomes of natural language processing methods or the educators' subjective assessment. All things considered; the score prediction module is essential to automating the assessment procedure.

Now, we compare the student's response with the model solution in our suggested system using a variety of Natural Language Processing (NLP) techniques to provide a similarity score. The purpose of this comparison is to evaluate the student's answer to the anticipated solution in terms of semantic alignment and relevance. First, we use cosine similarity, which computes the cosine of the angle between the vector representations of the two replies to determine how similar their textual content is. We also use the word movers' distance (WMD) approach, which measures the degree of dissimilarity between word distributions in the two replies in order to quantify the semantic similarity. Additionally, we use the RAKE (Rapid Automatic Keyword Extraction) algorithm to extract keywords or key phrases from both replies, giving each phrase a likelihood value based on its frequency and relevance in the text. Thus, we are able to derive a complete similarity score that indicates the degree of semantic alignment and relevance between the student's response and the model solution by combining the outcomes of several NLP approaches. This approach facilitates a nuanced evaluation of the student's response, enhancing the accuracy and effectiveness of the assessment process.

The score adjustment module stores the similarity score that was determined from the comparison using natural language processing (NLP) techniques such as cosine similarity, word movers' distance, and keyword extraction using the RAKE algorithm.

IV. Results

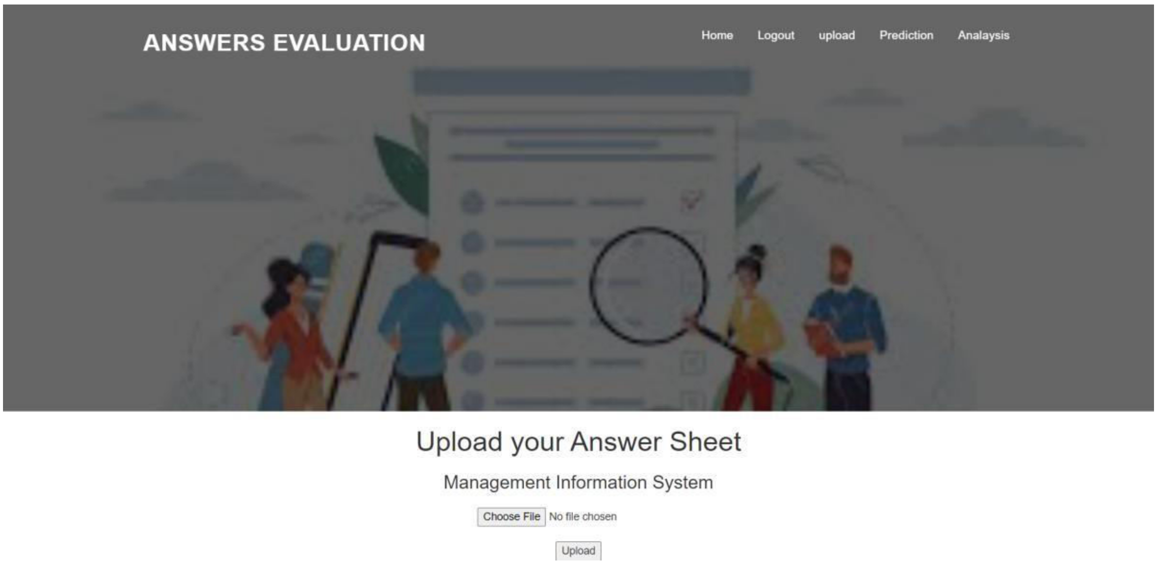

Fig 2. *The student has to upload the answer*

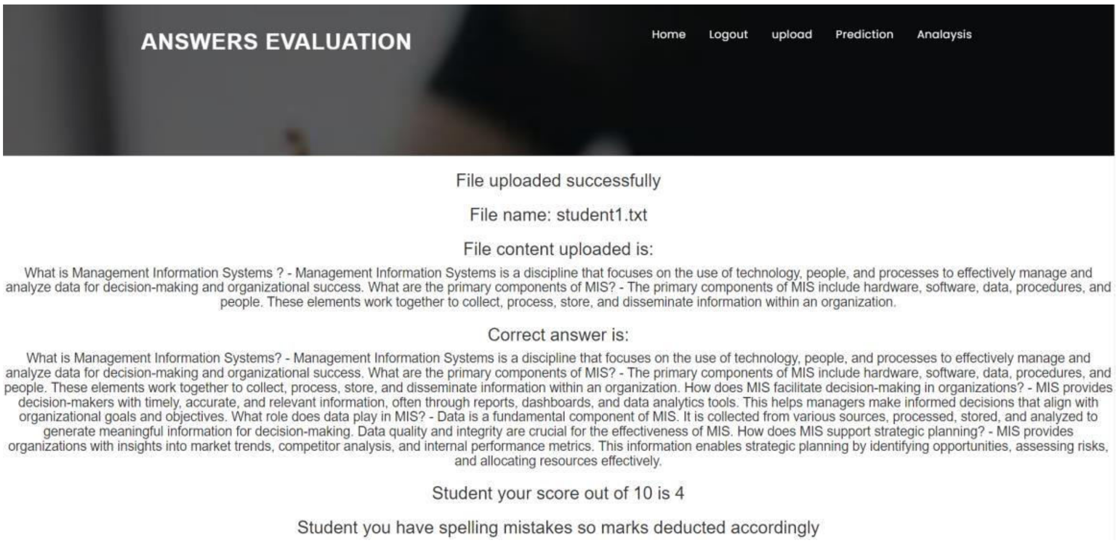

Fig 3. *Score will be given to the student*

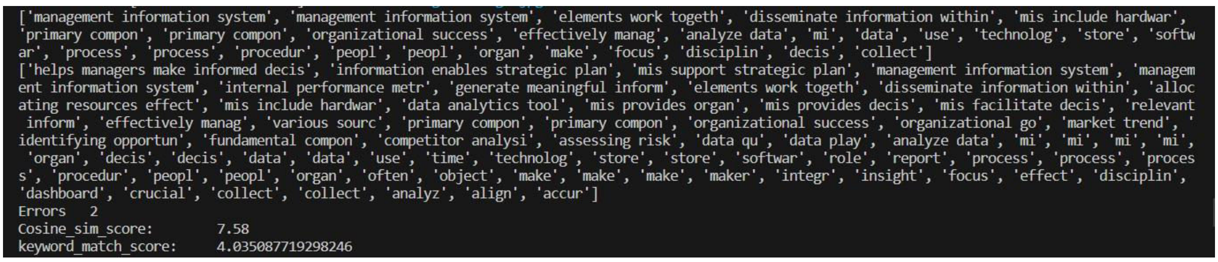

Fig 4. *Keywords extracted from student answer and reference answer*

The evaluation of our proposed automated evaluation system yielded promising outcomes, with an overall accuracy rate of 89% achieved across a diverse set of Management Information Systems (MIS) papers. We evaluated the system's performance through extensive testing and validation by contrasting the projected scores for a representative portion of the dataset with manually awarded ground truth values. Through this procedure, the accuracy and dependability of our system's evaluation of student replies was guaranteed. Our system's user-friendly interface, which is intended to make student participation easy, is one of its primary benefits. When they submit their responses in the form of text documents, the hybrid model automatically assesses the information and assigns a score between one and ten. Both teachers and students will save time and effort by not having to grade assignments by hand thanks to this simplified approach.

The system's user-friendly interface is seen in Figure 2, which also shows how easy it is for students to upload their answer keys. Students may get instant feedback on their performance by looking at Figure 3, which shows the score that our hybrid model predicted. Keywords along with grammar will be extracted from the student answer and reference answer which will assist in grading as shown in Fig 5. Our goal is to provide students with quick and accurate assessments by combining machine learning.

models and advanced natural language processing techniques. This will improve their learning experience and encourage ongoing development.

Furthermore, a more complicated assessment of the degree of alignment between the two may be made by contrasting the keywords taken from the student's response with those from the reference source. This thorough methodology guarantees that the evaluation procedure takes into consideration the particular concepts and themes included in both papers, in addition to semantic similarities.

By using this approach, we hoped to create a strong evaluation mechanism that would be able to record several facets of students' replies and yield a final score that would fairly represent the caliber and relevance of their solutions in relation to the assignment.

REFERENCES

[1] J. Wang and Y. Dong, "Measurement of text similarity: A survey," Inf., vol. 11, no. 9, p. 421, 2020. - A survey

[2] 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," Concurr. Comput. Pract. Exp., vol. 33, no. 5, 2021.

[3] M. S. M. Patil and M. S. Patil, "Evaluating student descriptive answers using natural language processing," International Journal of Engineering Research & Technology (IJERT), vol. 3, no. 3, pp. 1716–1718, 2014.

[4] P. Patil, S. Patil, V. Miniyar, and A. Bandal, "Subjective answer evaluation using machine learning," International Journal of Pure and Applied Mathematics, vol. 118, no. 24, pp. 1–13, 2018.

[5] J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat, "Document plagiarism detection using a new concept similarity in formal concept analysis," Journal of Applied Mathematics, vol. 2021, 2021.

[6] MUHAMMADFARRUKHBASHIR, HAMZA ARSHAD, ABDUL REHMAN JAVED NATALIA KRYVINSKA and SHAHAB S. BAND, "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing"

[7] H. Mittal and M. S. Devi, "Subjective evaluation: A comparison of several statistical techniques," Appl. Artif. Intell., vol. 32, no. 1, pp. 85–95, 2018

[8] L. A. Cutrone and M. Chang, "Automarking: Automatic assessment of open questions," in ICALT 2010, 10th IEEE International Conference on Advanced Learning Technologies, Sousse, Tunisia, 5-7 July 2010, pp. 143–147, IEEE Computer Society, 2010.

[9] Fang, Zhao & Zhang, Qiang & Tang, Xiaoan & Wang, Anning & Baron, Claude. (2020). An implicit opinion analysis model based on feature-based implicit opinion patterns. Artificial Intelligence Review. 53. 10.1007/s10462-019-09801-9.

[10] Al Moaiad, Yazeed & Ts,. (2023). A Review Of The Effectiveness Of Management Information System In Decision Making. 10.47750/pnr.2023.14.S02.155.

[11] Kiradoo, Giriraj, A Study on Management Information Systems Role and Adoption in Managerial Decision Making (2020). International Journal of Management (IJM), 11 (3), 2020, pp. 114–121., Available at SSRN: https://ssrn.com/abstract=3577432

[12] Piyush Patil, Sachin Patil, Vaibhav Miniyar, Amol Bandal , "Subjective Answer Evaluation Using Machine Learning", May 2018.

[13] Abhishek Girkar, Mohit khambayat, Ajay Waghmare, Supriya Chaudhary, "Subjective Answer Evaluation using Natural Language Processing and Machine Learning", Apr. 2021

[14] MUHAMMADFARRUKHBASHIR, HAMZA ARSHAD, ABDUL REHMAN JAVED NATALIA KRYVINSKA ,AND SHAHAB S. BAND, "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing.", 2021.

[15] L. Rudner and T. Liang, Automated essay scoring using Bayes theorem, J. Technol. Learn. ,vol. 1, no. 2, 2002.

[16] L. Bin, L. Jun, Y. Jian-Min, and Z. Qiao-Ming, Automated essay scoring using the KNN algorithm,Proc. - Int. Conf. Comput. Sci. Softw. Eng. CSSE 2008, vol. 1, pp. 735738, 2008.

[17] C. Leacock and M. Chodorow, C-rater: Automated scoring of short-answer questions, Comput.Hum., vol. 37, no. 4, pp. 389405, 2003.

[18] T. K. Landauer, Automatic Essay Assessment, Assess. Educ. Princ. Policy Pract., vol. 10, no. 3, pp.295308, 2003.

[19] T. Kakkonen, N. Myller, E. Sutinen, and J. Timonen, Comparison of dimension reduction methods for automated essay grading, Educ. Technol. Soc., vol. 11, no. 3, pp. 275288, 2008.

[20] M. Islam, Automated Essay Scoring Using Generalized, in Proceesings of 13th InternationalConference on Computer and Information Technology (ICCIT 2010), 2010.

[21] F. Noorbehbahani and a. a. Kardan, The automatic assessment of free text answers using a modified BLEU algorithm, Comput. Educ., vol. 56, no. 2, pp. 337345, 2011.

[22] Blei DM, Ng AY, Jordan MI: Latent Dirichlet Allocation. Journal of Machine Learning Research. 2003, 3: 993-1022.