Automated Question Classification Based on Bloom's Taxonomy

Varsha Naik[1], Mangesh Bedekar[2], Anuj Apte[3], Vedang Atgur[4]
[1, 2, 3, 4]Department of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University
[1]varsha.powar@mitwpu.edu.in
[2]mangesh.bedekar@mitwpu.edu.in
[3]1032211241@ mitwpu.edu.in
[4]1032211207@ mitwpu.edu.in

***Abstract*— This project attempts to revolutionize the process of question paper generation by automating the classification of questions according to revised Bloom's Taxonomy, to enhance the efficiency in evaluating cognitive levels, providing teachers with a valuable tool to create exam papers. Its value lies in its potential to revolutionize the process of making exam papers easily and in a short amount of time. The scope of our research project extends to personalized learning experiences and adaptive learning methodologies, which can be achieved by significantly increasing the amount of data to be processed, further refinement of the model and adding real-time feedback for continuous improvement. We have created this system especially for practical implications to simplify the exam paper assessment process, as well as for social implications as it could help to contribute to the ever-advancing education industry by promoting inclusive learning environments and develop higher order thinking skills among students.**

***Keywords*— Bloom's Taxonomy; Machine Learning; Natural Language Processing; Education**

ICTIEE* Track*: Assessment of Effective Teaching

ICTIEE Sub-Track: Assessment for Learning: Empowering Students through effective Assessment Practices

I. INTRODUCTION

Education, being a journey of continuous learning and evaluation, is tied to the important task of classifying problems according to Revised Bloom's taxonomy A et al. (2019); K et al. (2017) which divides skills into six levels that can be easily understood by higher considerations M & Omar (2018). But manually categorizing questions using Bloom's Taxonomy is laborious, arbitrary, and unreliable. This paper offers an automated approach that uses machine learning techniques like Random Forest and natural language processing (NLP) tools like RAKE and SpaCy to overcome these difficulties and achieve 90% classification accuracy. The implemented method improves assessment creation consistency and speed while staying in line with cognitive goals. It encourages balanced, individualized learning experiences and helps the development of higher-order thinking skills by focusing on a wide range of cognitive levels Our automated classification system's strength is its capacity to make test design a quick, effective procedure that works with the surroundings N et al. (2022). Future directions and limitations to increase system flexibility and computing efficiency are also examined.

II. BLOOM'S TAXONOMY

Bloom's Taxonomy was created by Benjamin Bloom in the 1950s to provide a language teachers, professors, scientists and examiners to discuss and communicate learning and assessment Jain et al. (2019). The illustration below describes each level of intelligence, from higher thinking to lower thinking Uma et al. (2017). The six levels of Revised Bloom's Taxonomy—"Remember," "Understand," "Apply," "Analyse," "Evaluate," and "Create"—are shown in a pyramid in Fig. 1 and are placed hierarchically according to the order of thinking skills.

Varsha Naik
Dr. Vishwanath Karad MIT World Peace University
varsha.powar@mitwpu.edu.in.

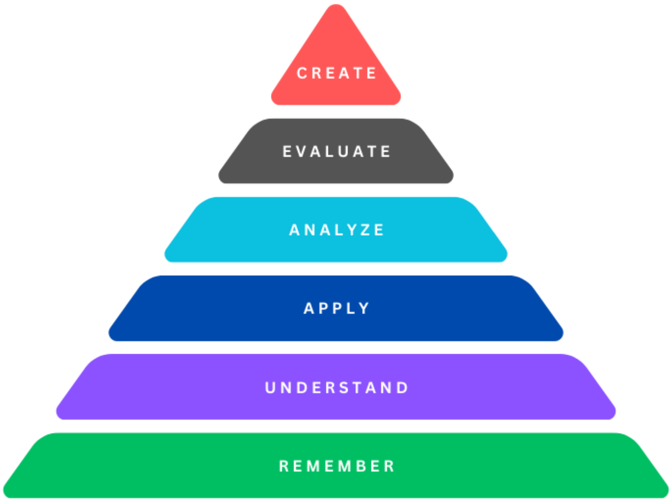

Fig. 1. Pyramid diagram of Revised Bloom's Taxonomy levels

Significance of Bloom's Taxonomy

By assisting teachers in matching learning objectives to the appropriate levels, Bloom's Taxonomy helps enhance assessment. For instance, teachers may evaluate students at the higher level of Bloom's taxonomy while teaching a higher-level course, while they may evaluate students at the lower level when teaching a lower-level course. The learning objectives of a course should create a dialogue so that teachers and students can understand the true purpose behind them Uma et al. (2017) R (2024) K et al. (2015). Furthermore, it can provide a framework for cognitive behaviors that can be used to assess the difficulty of a test paper and simplify or complicate the level of difficulty in the process A et al. (2019).

B. Comparison between snippets of question papers without and with Bloom's Taxonomy Levels

A question paper without Bloom's Taxonomy levels is an older method of assessing a student's depth of knowledge and learning, primarily focusing on memorization, where questions are generally simpler and based around recall of facts and definitions, thereby limiting the students' ability to have a profound understanding of a subject. Furthermore, questions might only belong to a narrow level of cognitive levels, which can hinder critical thinking and the students' ability to apply whatever they have learnt K et al. (2015). This creates an unbalanced assessment where the paper might inadvertently favor rote learning over analysis, synthesis and evaluation. On the other hand, with reference to Fig. 1, a paper with Bloom's Taxonomy levels, questions are made to target a wide variety of cognitive levels, encouraging analysis, application and evaluation of knowledge, enabling a more holistic method of learning A & R (2022). Furthermore, it distributes points across various cognitive levels, providing flexibility and ensuring that students are rewarded for demonstration of various cognitive skills Jain et al. (2019).

C. Motivation and Drawbacks of the Old Method

The motivation for automating question classification lies in providing educators with a tool to streamline the assessment process, make it efficient and accurate, gain valuable insights and create well-balanced question papers. The manual creation of question papers and susceptibility of human assessment to biases often leads to imbalances in the distribution of cognitive skills. Human assessors may bring their own perspectives, experiences, and inherent biases to the classification process, potentially leading to imbalances or inaccuracies in the distribution of questions across cognitive levels. S & A (2023) Educators and examiners face the challenge of dedicating significant time and to consistently and accurately assess and categorize questions based on cognitive complexity, which can be a labor-intensive and tedious task S & A (2023) Deena & K (2023) N et al. (2022).

D. Role of Educators in this System

While automation can help with classification and other tasks, teachers are still crucial in these areas since they make sure that the questions are in line with Bloom's Taxonomy's suitable cognitive levels. Educators must establish clear learning goals for their students and use these objectives to guide the selection of cognitive levels for questions. For example, teaching a foundational course might emphasize Remembering and Understanding, while advanced courses might target Analyzing or Creating. Furthermore, while automation can streamline processes, educators must integrate these tools effectively, with validation being an important task, where machine-suggested classifications should be reviewed and refined to ensure they align with course objectives. Overall, educators should ensure that assessments cover a variety of cognitive levels, fostering holistic learning. For development of a specific skill, educators should also emphasize their development through targeted questions. This synergy between educator expertise and automation enhances the quality of education and supports diverse learning methodologies.

III. DATASET DESCRIPTION

The dataset we used was a combination of 3 datasets, taken from Gani & Sangodiah (2023), having 3849 entries in total, each row corresponding to a specific question and its associated question type according to Bloom's Taxonomy. The dataset contains 2 columns, namely Questions and Question Types. Questions (QUESTIONS) are exam questions, ranging in complexity, catering to various cognitive levels defined by Bloom's Taxonomy, and span all possible academic subjects. They are the basis on which the model will train itself and give accurate predictions. The question types column uses Revised Bloom's Taxonomy to categorize each question into its appropriate type. This categorization spans the six cognitive levels, namely Remembering, Understanding, Applying, Analysing, Evaluating, and Creating.

IV. TOOLS USED

The project uses a combination of advanced tools and libraries to accomplish its objectives, integrating natural language processing (NLP) K et al. (2015) G & D (2023) A et al. (2023) and machine learning (ML) techniques. We utilized SpaCy for natural language processing tasks, including tokenization A et al. (2019), lemmatization, tagging and sentence parsing, and RAKE for keyword extraction Deena & K (2023). Regex library was used to remove special characters. We implemented tools like TF-IDF vectorizer A et al. (2019) Deena & K (2023), TrainTestSplit, Random Forest Classifier etc. in our model for training, testing and making predictions.

V. METHODOLOGY

First, we wanted to determine which classification algorithm would work best for our dataset's Revised Bloom's Taxonomy-based question classification. Since the MultinomialNB classification algorithm is a probabilistic learning technique based on the Bayes Theorem and is frequently employed in natural language processing, we first tested it. It's said to be particularly useful in classifying samples of text. The accuracy of this algorithm on the dataset was low (54%), although the precision of each question type was very high, with the recall and f1 score being very low, hence we had to turn to preprocessing the data first. We then applied preprocessing techniques which included porter-stemmer K et al. (2015) Haris & Omar (2012), text cleaning function to convert text to lower-case Haris & Omar (2012), tokenizing Haris & Omar (2012) using TF-IDF Vectorizer. We applied the MultinomialNB classification algorithm N et al. (2022) A & R (2022) on this preprocessed data to achieve an accuracy of 70%, with precision decreased slightly and recall, f1 score showing a slight increase. However, this result was not adequate, so we searched for another algorithm. We then tried using the SpaCy library for preprocessing in hopes to achieve a higher accuracy. This library is used in advanced NLP and Deep Learning to build/extract information from a sample of text. It's also used to preprocess data. With this we achieved an accuracy of 73%, with precision further decreasing and recall, f1 score increasing slightly. Next, we tried using RAKE for preprocessing Deena & K (2023), which can help identify important phrases in a paragraph. We achieved an accuracy of 75%, with precision further decreasing and recall, f1 score increasing slightly. However, we felt that the accuracy could be further improved, so we made adjustments to the TF-IDF vectorizer parameters and used English stopwords to filter out common words. We also used RAKE to extract keywords, SpaCy for tokenization and lemmatization and Regex to remove special characters. Next, we used multiple classifiers including TF-IDF Vectorizer and Support Vector Machine, among which the one with the best result will be used. This approach led to an accuracy of 85%. After experimenting on different Machine Learning Techniques like K-Nearest Neighbor N et al. (2022) A & R (2022) A et al. (2023), Logistic Regression, Linear Discriminant Analysis (LDA) R (2024), Support Vector Machine (SVM) N et al. (2022) A & R (2022); A et al. (2023), Decision Tree A et al. (2023), and finally, Random Forest, we discovered that Random Forest Classifier had the highest accuracy of 90% (Fig. 8, Table II). According to the classification report generated (Fig. 3) in the final iteration, the precision was not affected much, but the recall and the f1 score were finally within the accepted threshold that we set. Soon after we created a GUI for this model as shown in Fig. 4, Fig. 5 and Fig. 6 using Python's tkinter library to simulate a simple system that teachers or examiners can use to create a model question paper.

VI. SYSTEM ARCHITECTURE

As shown in Fig. 2 initially, preprocessing of the raw data takes place using RAKE, SpaCy and Regex. The preprocessed output is then trained and tested. Parameters are vectorized and tuned using the TF-IDF Vectorizer, and the data is classified using Random Forest Classifier, after which the model is ready for making predictions.

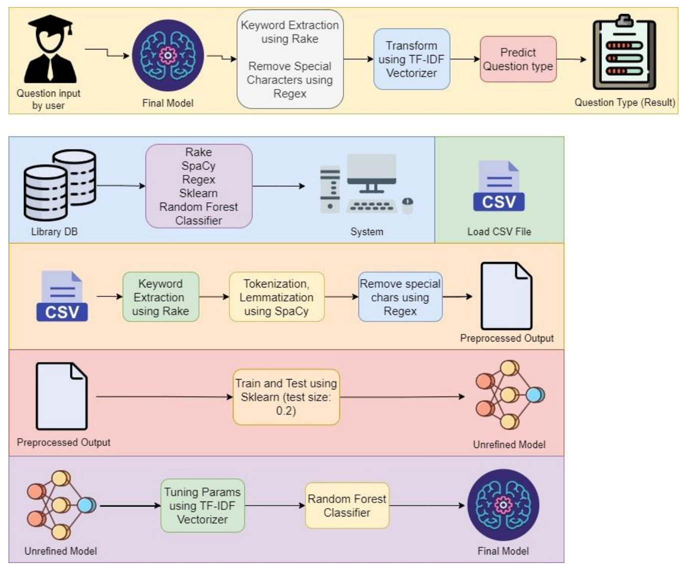

Fig. 2. System Architecture Diagram of the proposed work

The user uses the model by typing a question as input, after which it gets preprocessed using RAKE, SpaCy and Regex, transformed using TF-IDF Vectorizer and finally the cognitive level of the question that is predicted is displayed to the user.

VII. RESULTS

For every input question, the model produced predictions for the different kinds of Bloom's Taxonomy questions. The accuracy metric measured the proportion of questions that were correctly classified out of all the questions. In the final version of the model, we were able to attain 90% accuracy. For every type of Bloom's Taxonomy question, precision, recall, and F1-score were displayed in a comprehensive classification report.

A. Preprocessing

Fig.7 represents all the successive iterations of the preprocessing techniques used in the model (columns) and how each iteration affects all the levels of Revised Bloom's Taxonomy (rows). Starting with the first one which is using

MultinomialNB, the precision is unusually high (except for Understand and Apply) probably due to over fitting, and the recall, f1- scores are very low for all the levels. The precision gradually lowers to accepted levels as more iterations were made, and the recall, f1-score improves significantly. Table I shows the accuracy (columns) of each successive iteration (rows) in raw values and their corresponding percentages according to preprocessing techniques used. The accuracy increases with each iteration. When RAKE and SpaCy were used, we got the highest accuracy of 90% among all other methods.

TABLE I
ALL PREPROCESSING ALGORITHMS AND THEIR RESPECTIVE ACCURACIES

Algorithm	Accuracy	In %
MultinomialNB	0.54025	54%
MulitnomialNB with Preprocessing	0.70010	70%
SpaCy	0.73336	73%
RAKE	0.75467	75%
RAKE and SpaCy (final Iteration)	0.9	90%

B. Machine Learning Techniques

Every iteration of the machine learning techniques applied to the model (columns) is depicted in Fig. 8, along with the ways in which each iteration altered the Revised Bloom's Taxonomy at each level (rows). The accuracy of each machine learning method we applied is displayed in Table II along with the corresponding percentages. Following the last iteration, we experimented with these machine learning techniques; of these, Random Forest produced the best accuracy. The classification report of the last iteration, produced by the sklearn.metrics library, is displayed in Fig. 3 (using Random Forest for predictions and RAKE and SpaCy for preprocessing). The f1-score, recall, and precision are all within acceptable bounds. Despite this, the Understand level has exceptionally high support (326).

TABLE II
MACHINE LEARNING ALGORITHMS AND THEIR ACCURACIES IN %

ML Techniques	Accuracy
K-Nearest Neighbor	59.9%
Logistic Regression	78.9%
Linear Differential Analyzer	80.9%
Support Vector Machine	87.85%
Decision Tree	89.5%
Random Forest	90%

C. GUI

Fig. 4, the first window of the GUI asks the user the number of questions they want to enter, the user clicks “Submit Questions” and consequently the second window of the GUI opens. Fig. 5 opens and contains the number of question entries based on the number of questions the user entered in the previous window. After entering all the questions, user has to click on the “Generate Question Paper” button below which takes us to the third window. Fig. 6 generates a question paper with the first column containing the questions entered in the second window, and the second column containing the corresponding predicted levels for each question.

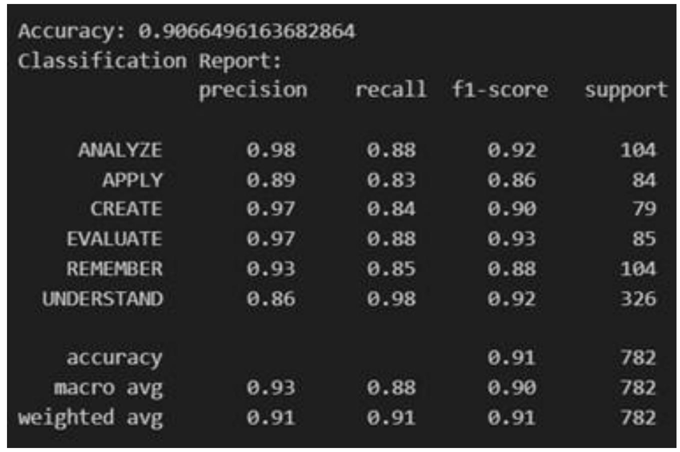

Fig. 3. Classification Report of the final iteration

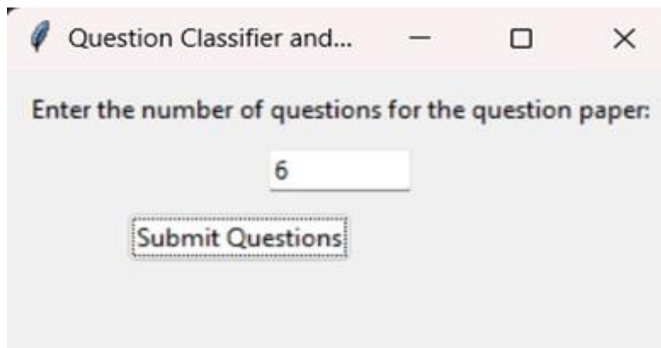

Fig. 4. First window of the GUI

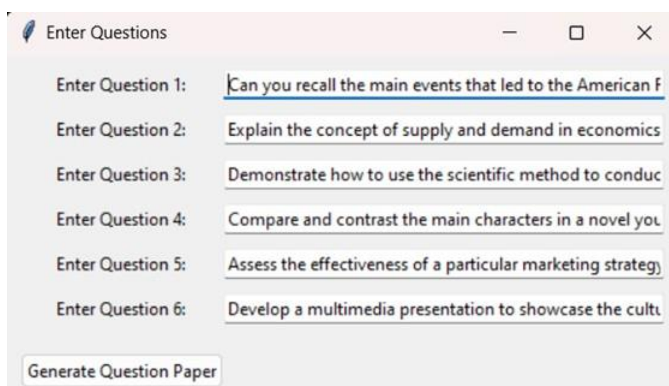

Fig. 5. Second window of the GUI

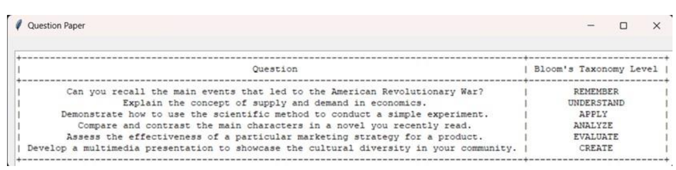

Fig. 6. Third window of the GUI

ALGORITHM	MultinomialNB			MultinomialNB with Pre-processing			Spacy			Rake-NLTK			Final Iteration		
QUESTION_LEVEL	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
REMEMBER	1	0.1	0.19	0.98	0.37	0.54	0.99	0.38	0.55	1	0.42	0.59	0.93	0.85	0.89
UNDERSTAND	0.49	1	0.66	0.59	1	0.74	0.62	1	0.77	0.64	1	0.78	0.87	0.96	0.91
APPLY	0.95	0.23	0.37	0.98	0.5	0.66	0.99	0.55	0.71	0.99	0.6	0.74	0.9	0.81	0.85
ANALYZE	1	0.12	0.21	1	0.46	0.63	1	0.52	0.68	1	0.55	0.71	0.98	0.88	0.93
EVALUATE	1	0.26	0.41	0.99	0.6	0.75	0.99	0.66	0.79	0.98	0.7	0.81	0.93	0.86	0.9
CREATE	1	0.2	0.34	0.97	0.46	0.63	0.97	0.59	0.73	0.98	0.62	0.76	0.89	0.84	0.86

Fig. 7. All preprocessing iterations, from start to final iteration

TECHNIQUE	K-Nearest Neighbor			Logistic Regresion			Linear Differential Analyzer		
QUESTION LEVEL	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
REMEMBER	0.6	0.62	0.61	0.92	0.64	0.76	0.74	0.76	0.75
UNDERSTAND	0.56	0.9	0.69	0.7	0.99	0.82	0.86	0.88	0.87
APPLY	0.91	0.25	0.39	0.84	0.63	0.72	0.7	0.76	0.73
ANALYZE	0.69	0.17	0.28	1	0.59	0.74	0.84	0.78	0.81
EVALUATE	0.72	0.74	0.73	0.96	0.78	0.86	0.79	0.75	0.77
CREATE	0.73	0.1	0.18	0.91	0.61	0.73	0.78	0.73	0.76

TECHNIQUE	Support Vector Machine			Decision Tree			Random Forest		
QUESTION LEVEL	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
REMEMBER	0.85	0.79	0.82	0.9	0.82	0.86	0.93	0.85	0.88
UNDERSTAND	0.88	0.95	0.92	0.88	0.95	0.92	0.86	0.98	0.92
APPLY	0.75	0.87	0.81	0.82	0.81	0.81	0.89	0.83	0.86
ANALYZE	0.92	0.84	0.87	0.97	0.88	0.91	0.98	0.88	0.92
EVALUATE	0.96	0.81	0.88	0.96	0.86	0.91	0.97	0.88	0.93
CREATE	0.94	0.84	0.89	0.97	0.89	0.9	0.97	0.84	0.9

Fig. 8. All ML Techniques, from start to final iteration

CONCLUSIONS

In conclusion, we made an effort to tackle the problem of automatically classifying questions using Bloom's Taxonomy, which has a big influence on how assessments are used in education. The developed system exhibits the capacity to precisely classify questions into cognitive levels by utilizing machine learning and natural language processing techniques. The integration of preprocessing methods, TF-IDF vectorization, and a Random Forest classifier contributes to the creation of a robust and adaptive tool for educators. The project’s success is evident in the achieved accuracy and detailed insights provided by the classification report. The system not only streamlines the assessment process but also encourages educators to adapt their instructional strategies based on the cognitive demands of questions. However, this experiment was not without its challenges, which included the significant amount of time required for the code to execute and the arduous efforts taken to find the appropriate Machine Learning algorithm. Furthermore, the significance of this automated classification system extends to fostering higher-order thinking skills and enhancing the overall learning experience for students. As technology continues to play a pivotal role in education, the automated question classification system presented here stands as a testament to the potential for innovation in educational assessment. The knowledge gained from this project serves as a basis for additional investigation and improvement, with the ultimate objective of continuously enhancing and customizing learning opportunities for students.

FUTURE SCOPE

We could extend the scope of this project by improving computational time by implementing knowledge graph using Spacy, NLTK, PyTorch as well as TensorFlow, instead of sequential access of arrays, using pandas. Furthermore, we aim to create a robust system that assists the teachers in making a question paper by recommending similar type of lower or higher – level questions.

REFERENCES

A, A, M., & E (2019). Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom’s Taxonomy. 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS).

A & R (2022). Primary and High School Question Classification based on Bloom’s Taxonomy. 2022 10th International Conference on Information and Communication Technology (ICoICT).

A, S, & Singh, S. (2023). A Systematic Review of Question Classification Techniques Based on Bloom’s Taxonomy. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT).

Deena, G. & K (2023). Keyword Extraction Using Latent Semantic Analysis For Question Generation. Journal of Applied Science and Engineering, 26(4).

G & D (2023). Accurate Recasting of Giant Text into Charts Using Rapid Automatic Keyword Extraction Algorithm in Comparison with Bag of Words Algorithm. 2023 6th International Conference on

Contemporary Computing and Informatics (IC3I). IEEE.

Gani, M. O. & Sangodiah, A. (2023). Exam Question Datasets.

Haris, S. S. & Omar, N. (2012). A rule-based approach in Bloom's Taxonomy question classification through natural language processing. In 7th International Conference on Computing and Convergence Technology (ICCCT), (pp. 410–414).

Jain, M., R, A, Grover, T., & U (2019). Classifying Question Papers with Bloom's Taxonomy Using Machine Learning Techniques," Communications in Computer and Information Science. Singapore: Springer.

K, M, & I (2015). An automatic classifier for exam questions in Engineering: A process for Bloom's taxonomy. 2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE).

K, S., A, & S, I. (2017). Outcome based predictive analysis of automatic question paper using data mining. 2017 2nd International Conference on Communication and Electronics Systems (ICCES).

M & Omar, N. (2018). Question Classification Based on Bloom's Taxonomy Using Enhanced TF-IDF. International Journal on Advanced Science, Engineering and Information Technology, 8(4-2), 1679–1685.

N, K, & M (2022). Automatic Question Classifier. 2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA).

R (2024). Multi-Agent System for Students Cognitive Assessment in E-Learning Environment. IEEE Access, 12, 15458– 15467.

S, L. & A, F. (2023). Comparative Analysis of Term Weighting Methods for Question Classification in Bloom Taxonomy Using Machine Learning Approach. 2023 IEEE International Conference on Computing (ICOCO).

Uma, S., Thenmozhi, R., & Hansda (2017). Analysis on Cognitive Thinking of an Assessment System Using Revised Bloom's Taxonomy. In 5th IEEE International Conference on MOOCs, Innovation and Technology in Education (MITE), (pp. 152–159).