Analysis and Metaphysics
ISSN: 1584-8574, e-ISSN: 2471-0849
Vol 20(1), 2021
pp. 38–54

A Comparative Analysis on Classifying Model Problems Using Revised Bloom's Taxonomy with SVM and K-NN Techniques

Sharanagouda N. Patil[1], RajShekhar S Heera[2], RajKumar Bainoor[3]

[1]Govt. Polytechnic for Women, Kalaburagi,

[2]Govt. Polytechnic, Kalaburagi

[3]PDA College of Engineering, Kalaburagi

Abstract

Effective education involves more than rote learning—it develops students' reasoning and creativity. The Revised Bloom's Taxonomy (RBT) offers a structured way to assess cognitive levels, ranging from recalling facts to generating new ideas. In this study, we automate the classification of exam questions into Bloom's levels using machine learning—specifically, Support Vector Machines (SVM) and K-Nearest Neighbors (K-NN). We incorporate preprocessing steps such as grammar validation and subject-relevance checking, and leverage action verbs to map questions to cognitive levels. Our work is grounded in the findings of several studies, which indicate that SVM often surpasses K-NN in classification tasks [1–6]. We believe our approach can assist educators in designing more balanced and thought-provoking assessments.

I. Introduction

In a broader sense, education is a process of shaping learner behavior through planning, teaching**,** and assessment. Among these, assessment plays a pivotal role in determining conceptual growth and developing cognitive skills. The comparative study presented in this work focuses on applying machine learning classification frameworks to categorize questions according to RBT, aiming to achieve improved accuracy and enhance the overall effectiveness of assessments.

Within the education system, the question bank plays a vital role in assessment. It not only stores a large pool of questions but also supports their organization, contextual placement, and classification. Properly managed question banks allow instructors to design effective evaluations by selecting unique subsets of questions that target different skill levels. They also enable the analysis of student performance based on responses and doubts raised during classroom interactions.

Revised Bloom's Taxonomy (RBT)**,** proposed by Krathwohl, serves as a widely accepted framework for categorizing questions into six hierarchical cognitive skill levels, ranging from lower to higher complexity. These levels emphasize the learner's thinking ability and cognitive processes, making Bloom's Taxonomy an essential tool for educators in preparing assessments, designing curricula, and setting course objectives.

To classify questions systematically, machine learning methods such as Support Vector Machines (SVM**)** and K-Nearest Neighbors (K-NN**)** can be applied. Before classification, question text undergoes preprocessing, after which classifiers are trained on labeled datasets, with expert guidance defining the cognitive categories. Once trained, these models can automatically assign new questions to the appropriate category. This approach facilitates more accurate and scalable classification of assessment material.

II. Related Work

Several studies have attempted to automate the classification of questions according to Bloom's Taxonomy.

A hybrid ensemble of classifiers was introduced by Abduljabbar and Omar [1] to improve the accuracy of Bloom's taxonomy classification. Their approach demonstrated that combining multiple algorithms significantly outperforms standalone classifiers, setting a foundation for future hybrid systems.

According to Osadi et al. [2], reliability in Bloom's taxonomy classification can be enhanced by using an ensemble majority-voting approach. Their framework addressed dataset imbalance and ambiguity, demonstrating the strength of ensemble learning in educational text mining.

The challenge of STEM-specific assessment was tackled by Jayakodi et al. [3] through the development of an automatic classifier. By integrating linguistic features with Bloom's taxonomy, their system achieved promising results, proving automation feasible in engineering contexts.

Haris and Omar [4] adopted a rule-based NLP strategy for Bloom's taxonomy classification. The use of handcrafted syntactic rules provided transparency, though scalability limitations highlighted the trade-off between interpretability and performance.

In their analysis of Syrian geography textbooks, Soudan and Sleeman [5] found that most questions targeted lower-order thinking skills. This imbalance exposed curriculum weaknesses that hinder critical thinking development among students.

A study by Abdelrahman [6] on Jordanian English textbooks revealed a disproportionate focus on recall and comprehension. With evaluation and creation underrepresented, the findings questioned alignment with modern educational objectives.

To enhance feature representation, Mohammed and Omar [7] combined modified TF-IDF with word2vec embeddings. Their hybrid feature-engineering method boosted classification robustness across varied datasets, highlighting the benefit of blending statistical and semantic features.

Supervised machine learning was leveraged by Jain et al. [8] to classify exam questions according to Bloom's taxonomy. Benchmark tests validated the framework's effectiveness, offering empirical support for ML-driven educational quality assurance.

Aninditya et al. [9] demonstrated that lightweight models can be highly effective by proposing a TF-IDF + Naïve Bayes pipeline. Their approach was both computationally efficient and accurate for large-scale educational datasets.

Magas et al. [10] extended Bloom's taxonomy into surgical education, showing that structured intraoperative questioning nurtures higher-order skills. Their work proved taxonomy's relevance beyond classroom-based learning.

A comprehensive review was conducted by Makhlouf et al. [11], who compared rule-based, statistical, and deep learning methods for taxonomy-based classification. They identified machine learning as the most scalable direction, offering a roadmap for further research.

Zhang et al. [12] applied machine learning to large-scale computing exam datasets, achieving promising results, particularly for lower-order Bloom's categories. Their work bridged the gap between computer science education and cognitive assessment.

With the rise of deep learning, Huang et al. [13] proposed a scalable framework for automatic question classification. Their model achieved adaptability across multiple subjects, supporting real-time smart classroom applications.

Patil [14] examined a mechanical engineering course question paper and discovered a strong bias toward knowledge-level items. The findings suggested insufficient evaluation of higher-order skills, highlighting the need for instructional reform.

A mathematical perspective was introduced by Voskoglou [15], who modeled Bloom's taxonomy levels using Markov chains. This probabilistic framework provided insights into learning transitions and added rigor to assessment methods.

Das et al. [16] addressed cognitive complexity detection through supervised multi-class text classification. Their method improved recognition of higher-order skills, advancing computational measurement of learning outcomes.

By applying Bloom's revised taxonomy, Sami and Arumugam [17] evaluated student learning skills. Their descriptive study highlighted mismatches between intended outcomes and observed performance, calling for targeted pedagogical interventions.

The imbalance between lower- and higher-order skills was exposed by Şanli [18] in Turkish geography coursebooks. Findings revealed insufficient promotion of creativity and critical thinking, reinforcing taxonomy's role in curriculum reform.

Meissner et al. [19] benchmarked multiple machine learning techniques for automatic annotation of Bloom's taxonomy levels in e-assessment items. Their results highlighted the superior performance of deep learning models, strengthening adaptive e-learning systems.

An overreliance on recall-based items was observed by Pugh and Gates [20] in physics exam questions. They recommended inclusion of more application and synthesis-level items to foster balanced assessment in STEM.

Through the lens of Bloom's taxonomy, Ginting et al. [21] analyzed Indonesian English worksheets. The study found a dominance of comprehension-level questions, limiting students' exposure to analysis and evaluation.

Zorluoğlu and Güven [22] evaluated fifth-grade science exam papers and discovered deficiencies in evaluation and creation levels. Their work highlighted the urgent need to align curricula with holistic learning goals.

The influence of feature weighting on performance was investigated by Sangodiah et al. [23], who tested TF, IDF, and hybrid weighting schemes. Results confirmed that weighting strategies play a critical role in classification accuracy.

Waheed et al. [24] advanced the field by proposing BloomNet, a transformer-based architecture for Bloom's taxonomy classification. Their model achieved state-of-the-art results, signaling a shift toward transformer-driven NLP for education.

III. Background

A. Revised Bloom's Taxonomy (RBT)

The original Bloom's Taxonomy, introduced in 1956, classified learning objectives into a hierarchical model that emphasized the progression of cognitive skills from lower-order to higher-order thinking. While revolutionary for its time, the framework focused primarily on knowledge acquisition and did not fully capture the dynamic and evolving nature of learning in modern educational contexts.

In 2001, Anderson and Krathwohl revised Bloom's framework, introducing a more flexible and multidimensional model that better aligned with contemporary teaching and learning practices. The revised taxonomy reorganized the hierarchy of cognitive processes into six levels: Remember, Understand, Apply, Analyze, Evaluate, and Create**.** This shift emphasized action-oriented verbs and placed "Create" at the highest level, reflecting the growing importance of innovation, problem-solving, and original thought in education.

Another key innovation of the revised model was the introduction of the Knowledge Dimension, which expanded the scope of learning beyond factual recall. This dimension includes Factual Knowledge (basic elements and terminology), Conceptual Knowledge (interrelationships and theories), Procedural Knowledge (methods, algorithms, and techniques), and Metacognitive Knowledge (awareness of one's own learning strategies). This two-dimensional structure enables educators to design more precise and measurable learning objectives.

The Cognitive Process Dimension focuses on how learners engage with knowledge, while the Knowledge Dimension emphasizes what learners need to know. Together, these two axes create a matrix that provides educators with a comprehensive framework for curriculum design, instructional planning, and assessment. For instance, students may be required to "analyze procedural knowledge" in a laboratory setting or "create conceptual knowledge" through project-based assignments.

Revised Bloom's Taxonomy has become an essential tool in Outcome-Based Education (OBE)**,** where learning outcomes are defined in terms of measurable skills and competencies. It assists educators in aligning Program Outcomes (POs), Course Outcomes (COs), and Program-Specific Outcomes (PSOs)**,** ensuring that students develop both foundational knowledge and higher-order thinking capabilities.

In research and higher education, the revised taxonomy plays a vital role in structuring teaching-learning processes, evaluation methods, and accreditation frameworks such as NBA and NAAC. By mapping learning objectives to specific cognitive and knowledge dimensions, educators can design balanced assessments that evaluate students' abilities across multiple levels, from basic recall to critical evaluation and creative application.

Overall, Revised Bloom's Taxonomy provides a dynamic and versatile model that supports the development of 21st-century skills, fostering critical thinking, innovation, and lifelong learning. Its wide adoption across disciplines highlights its value in bridging traditional pedagogy with modern educational needs.

B. Support Vector Machines (SVM)

Support Vector Machines (SVM) are a class of supervised machine learning algorithms introduced by Vapnik and colleagues in the early 1990s. They are primarily used for classification and regression tasks, with a focus on maximizing the separation between different classes of data. Unlike traditional linear classifiers, SVM aims to construct an optimal hyperplane that not only separates data points but also maximizes the margin between the closest points of different classes, known as support vectors.

The key principle of SVM is the maximization of the margin, which is the distance between the separating hyperplane and the nearest data points from each class. By maximizing this margin, SVM achieves better generalization performance, reducing the risk of misclassification on unseen data. This margin-based approach distinguishes SVM from other machine learning models that often minimize classification error without explicit emphasis on generalization.

For non-linearly separable data, SVM employs the concept of kernel functions, which map input data into higher-dimensional feature spaces where linear separation becomes feasible. Commonly used kernels include the linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel. This "kernel trick" allows SVM to handle complex decision boundaries efficiently without explicitly computing transformations into higher dimensions.

Mathematically, SVM solves a convex optimization problem that guarantees a unique global solution, ensuring model stability. The optimization involves minimizing a cost function subject to constraints defined by the data points and the margin. The trade-off between margin maximization and misclassification tolerance is controlled by the regularization parameter (C), while kernel-specific parameters such as gamma (γ) influence the decision boundary's complexity.

SVM has been widely adopted across various domains, including bioinformatics (gene expression analysis, protein classification), computer vision (image recognition, facial detection), natural language processing (text classification, sentiment analysis), and engineering (fault diagnosis, signal processing). Its robustness in high-dimensional spaces and ability to handle sparse datasets make it a preferred choice in many applications.

Despite its advantages, SVM has limitations such as high computational complexity for large datasets and sensitivity to parameter tuning. In practice, careful selection of kernels and hyperparameters through methods like cross-validation is necessary to achieve optimal performance. Moreover, recent advancements in deep learning have reduced the dominance of SVM in large-scale image and speech recognition tasks, though SVM remains highly relevant for structured and medium-sized datasets.

In conclusion, Support Vector Machines (SVMs) represent a robust and mathematically rigorous approach to both classification and regression tasks. By maximizing the margin between data classes, leveraging kernel functions for flexible nonlinear mapping, and ensuring strong generalization, SVMs have become a cornerstone in the evolution of machine learning methods. As shown in Fig. 1, SVMs construct an optimal hyperplane supported by critical data points—referred to as support vectors—and can be extended through kernel transformations into higher-dimensional feature spaces for handling complex, non-linearly separable data. Despite the increasing dominance of deep learning models, SVMs continue to serve as a benchmark technique and remain a reliable alternative for many real-world applications.

.

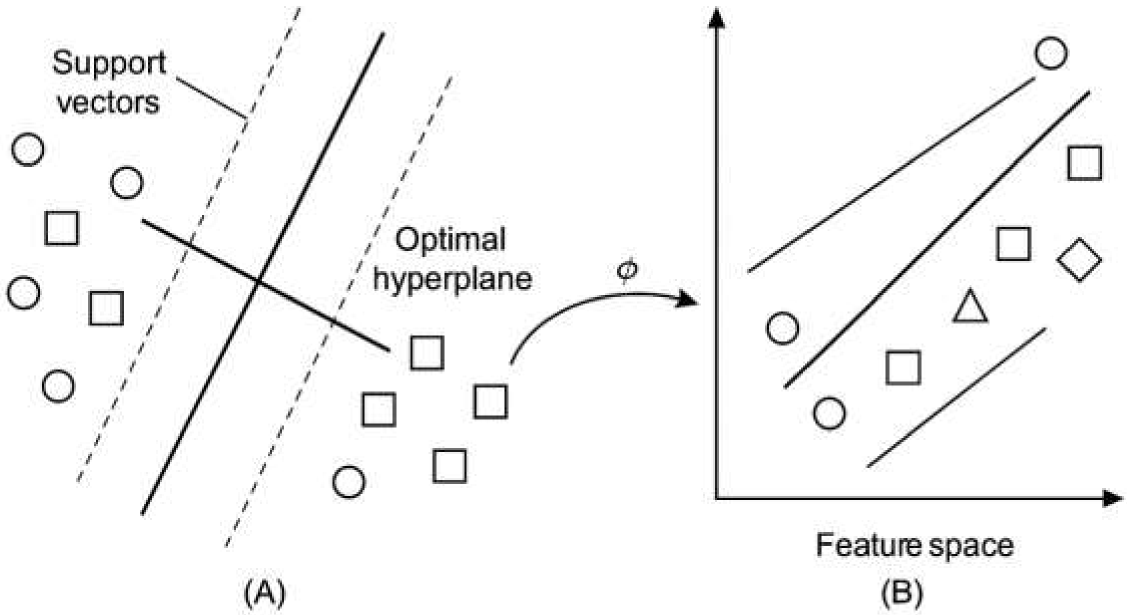

Fig. 1. Support Vector Machine: (A) Optimal separating hyperplane with maximum margin defined by support vectors; (B) Kernel mapping (φ) transforming input data into a higher-dimensional feature space for linear separation.

C. K-Nearest Neighbors (K-NN)

K-Nearest Neighbors (K-NN) is one of the simplest and most widely used supervised machine learning algorithms. Introduced by Fix and Hodges in the 1950s, K-NN is a non-parametric and instance-based learning method, meaning it does not assume an underlying probability distribution for the data and makes predictions based on stored examples rather than building an explicit model during training. Its simplicity, interpretability, and effectiveness in various domains have made it a popular choice for classification and regression tasks.

The fundamental principle of K-NN is based on the idea that similar instances exist in close proximity within a feature space. For a given test sample, the algorithm identifies the k closest training samples—its “neighbors”—using a distance metric such as Euclidean, Manhattan, or Minkowski distance. The majority class (for classification) or the average value (for regression) among these neighbors determines the output.

The choice of k plays a critical role in K-NN’s performance. A small k (e.g., k=1) makes the model highly sensitive to noise and outliers, while a large k smooths decision boundaries but may overlook finer distinctions between classes. Typically, k is chosen through cross-validation to balance bias and variance effectively. Additionally, the distance metric used can influence classification accuracy, with Euclidean distance being most common in continuous feature spaces.

K-NN has been widely applied across domains such as pattern recognition, image classification, recommendation systems, medical diagnosis, and anomaly detection. For instance, in medical imaging, K-NN can classify tumor cells based on shape and intensity features, while in recommendation systems, it helps suggest products by finding users with similar profiles. Its flexibility allows it to adapt to both binary and multi-class classification problems.

Despite its advantages, K-NN has limitations. It is computationally expensive at prediction time because the algorithm must compute distances between the query sample and all training samples. It is also sensitive to irrelevant or redundant features, which can distort distance calculations. Dimensionality reduction techniques such as Principal Component Analysis (PCA) are often employed to address the "curse of dimensionality," where performance degrades in high-dimensional spaces.

Another challenge with K-NN is class imbalance, where the majority class may dominate the decision boundary. Weighted K-NN variants mitigate this by assigning greater influence to nearer neighbors. Furthermore, indexing techniques like KD-trees or Ball-trees can reduce computational costs, making K-NN more scalable for larger datasets.

In conclusion, the k-nearest neighbors (K-NN) algorithm remains a foundational method in machine learning, appreciated for its simplicity, transparency, and effectiveness in small to medium-sized datasets. As shown in Fig. 1, K-NN classifies new data points based on the majority class among their nearest neighbors, with decision boundaries varying according to the chosen value of kkk. While more advanced models such as deep learning provide higher accuracy for large-scale and complex tasks, K-NN continues to be relevant as a baseline classifier and in applications where interpretability and straightforward implementation are essential.

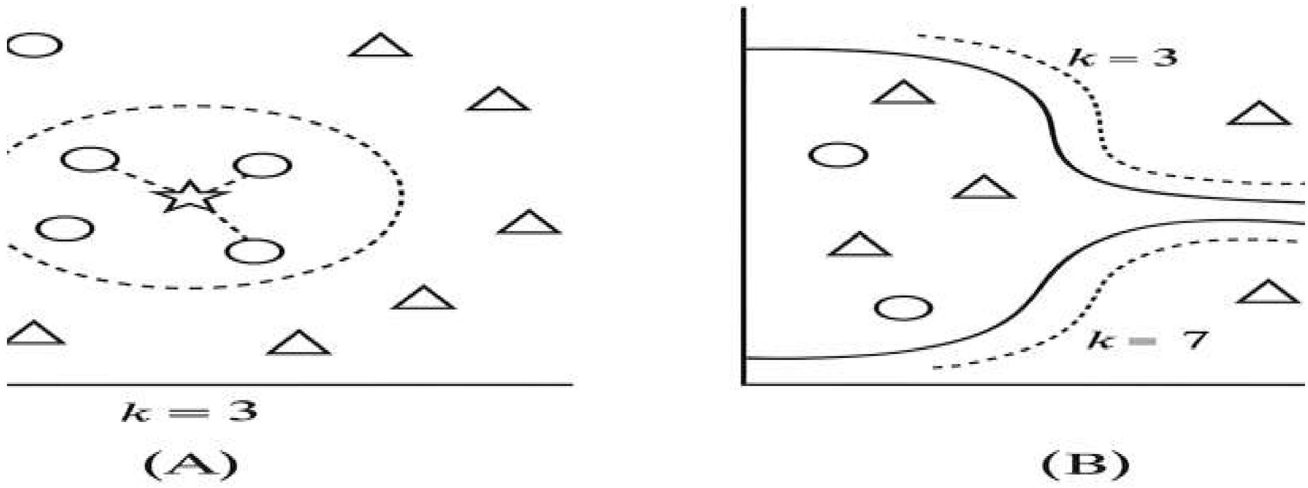

Fig. 2. K-Nearest Neighbors Algorithm: (A) Query point classification based on its three nearest neighbors; (B) Decision boundary variation for k = 3 and k = 7, showing the effect of parameter choice on classification outcomes.

IV. Methodology

Step 1: Dataset Collection and Preprocessing – The initial stage involves compiling a sufficiently large dataset to enable an effective comparative study. In this work, a dataset of 1,000 questions was collected from multiple sources including textbooks, university question papers, pedagogical activities, and online Q&A platforms. For this study, the domain selected was Engineering and Mathematics Subjects. Alongside the dataset, a catalog of action verbs corresponding to the cognitive skill dimensions of the Revised Bloom's Taxonomy was created. The original list of verbs was drawn from Anderson and Krathwohl's framework and extended with synonyms and additional verbs available in the public domain. Furthermore, subject-specific keywords were extracted from the textbook index, which was considered as the "mini world" of the domain. During preprocessing, all stop words, punctuation, and irrelevant tokens were removed from the questions.

Step 2: Grammar Validation – The curated dataset was then subjected to a grammar-checking process to eliminate syntactically incorrect and ambiguous questions. This was implemented using the Ginger API, an English grammar-checking tool. As a result, only grammatically valid sentences were retained for further analysis.

Step 3: Context Verification – The next stage involved verifying whether a question was relevant to the chosen subject domain. While some questions may be grammatically correct, they may fall outside the scope of the *Operating Systems* context. A question was considered "in-context" if it contained at least one keyword from the subject-specific index derived from the textbook. This filtering step ensured that the final dataset strictly represented the target domain, yielding a refined question bank.

Step 4: Classification by Revised Bloom's Taxonomy – After validation, the filtered questions were classified into one of the six cognitive levels of Revised Bloom's Taxonomy: Remember, Understand, Apply, Analyze, Evaluate, and Create. Each level was assigned a numerical value ranging from 1 Remember to 6 Create. The classification process involved tokenizing each question after removing punctuation and matching the extracted words with the verb/keyword list. If a match was found, the question was assigned to the corresponding Bloom's level.

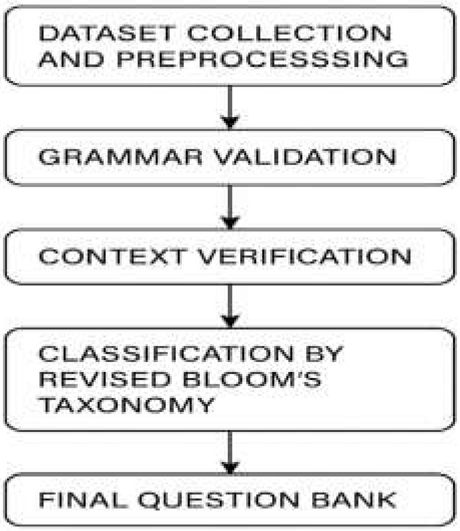

Fig. 3. Methodology flowchart: Dataset collection and preprocessing → Grammar validation → Context verification → Classification by Revised Bloom's Taxonomy → Final question bank.

Ambiguity Resolution – Certain verbs are inherently ambiguous, as they appear in more than one cognitive level. For instance, the verb *choose* may belong to both *Evaluate* and *Create*. In such cases, the question was assigned to the higher cognitive level (e.g., *Create*) under the assumption that answering such a question requires advanced reasoning skills.

Iterative Process – This entire methodology was applied iteratively across the question bank. If a question failed the grammar check in Step 2, it was immediately skipped and the next question was processed. Similarly, if a question was determined to be out-of-context in Step 3, it was excluded. The process of checking grammar, verifying context, and classifying by Bloom's level continued until all questions in the dataset had been processed. This ensured that only valid, context-specific, and pedagogically meaningful questions were categorized within the framework of Revised Bloom's Taxonomy.

V. Results and Discussion

This section presents the results obtained by applying the proposed methodology. The outputs include sample key verbs and their synonyms for each Bloom's level as shown in Table 1, a sample keyword list for in-context checking as described in Table 2, and the set of test questions with corresponding outputs illustrated in Table 3. During execution, the minimum output is a single column ("approved" or "not approved") generated during the grammar-checking phase. The maximum output consists of three columns—"approved," "in context," and "Revised Bloom's taxonomy level." Questions yielding only one column are discarded, as they indicate non-approval. Similarly, questions with two columns (approved but not in the context of the textbook) are also excluded from Bloom's level assignment. Any questions that are either not approved or not contextually valid are omitted from further analysis. Consequently, the final output consists of validated questions that are grammatically correct, contextually relevant, and assigned to their respective Revised Bloom's Taxonomy levels, with readability scores included. It is important to note that only theoretical questions are considered in this study, with numerical problems excluded, and the keyword set for in-context checking is limited to a single textbook, though it can be expanded to multiple sources.

Table 1: Sample Key Verb List for Cognitive Skill Domain

Bloom's Level	Key Verbs	Synonyms / Related Terms
Remember	Define, List, Recall, State	Identify, Recognize, Repeat
Understand	Explain, Summarize, Classify, Describe	Interpret, Paraphrase, Discuss
Apply	Implement, Solve, Use, Demonstrate	Execute, Practice, Illustrate
Analyze	Differentiate, Compare, Examine, Organize	Break down, Categorize, Contrast
Evaluate	Judge, Critique, Recommend, Justify	Assess, Argue, Validate
Create	Design, Develop, Construct, Formulate	Generate, Compose, Propose

Table 2: Output columns for Context checking

Output Columns	Meaning	System Decision	Next Step
1 Column (Approved / Not Approved)	Grammar check result only	If *Not Approved* → Question discarded	No further evaluation
2 Column (Approved + In/Not in Context)	Question is grammatically correct but may not align with textbook context	If *Not in Context* → Question discarded	No Bloom's level assignment
3 Column (Approved + In Context + Bloom's Level)	Question is grammatically correct, contextually relevant, and assigned a Bloom's taxonomy level	Accepted as valid	Used for further analysis & classification

Table 3: Sample Test question and output

0	Justify the use of UART/SPI/I2C over state mac...	Evaluate	Remember,Apply,Evaluate
1	Apply the concept of GPIO and timers to solve ...	Apply	Apply
2	List the key features of low-power techniques ...	Remember	Remember
3	Design a circuit using PLC programming to achi...	Understand	
4	Justify the use of SNR and noise figure over Q...	Evaluate	Apply,Evaluate
5	Compare interrupts with state machines in Embe...	Analyze	Remember,Analyze
6	Compare transducers with DMM calibration in El...	Analyze	Analyze
7	Define op-amp applications.	Remember	Remember
8	Analyze the effect of varying parameters on QA...	Analyze	Analyze
9	Analyze the effect of varying parameters on co...	Analyze	Analyze

The distribution of questions across different Bloom's taxonomy levels is illustrated in Fig. 4**,** where "Analyze" emerges as the most frequently assigned level, followed by "Remember" and "Evaluate." A more detailed representation of the matched levels for the same set of questions is shown in Fig. 5, highlighting instances where multiple Bloom's levels are associated with a single question (e.g., *Remember, Apply, Evaluate*). To further analyze the alignment between rule-based classification and matched levels, a heatmap is presented in Fig. 6**,** which visualizes the overlap and intensity of correspondence across different Bloom's levels. Together, these results validate the robustness of the methodology in categorizing theoretical questions with consistency and contextual relevance.

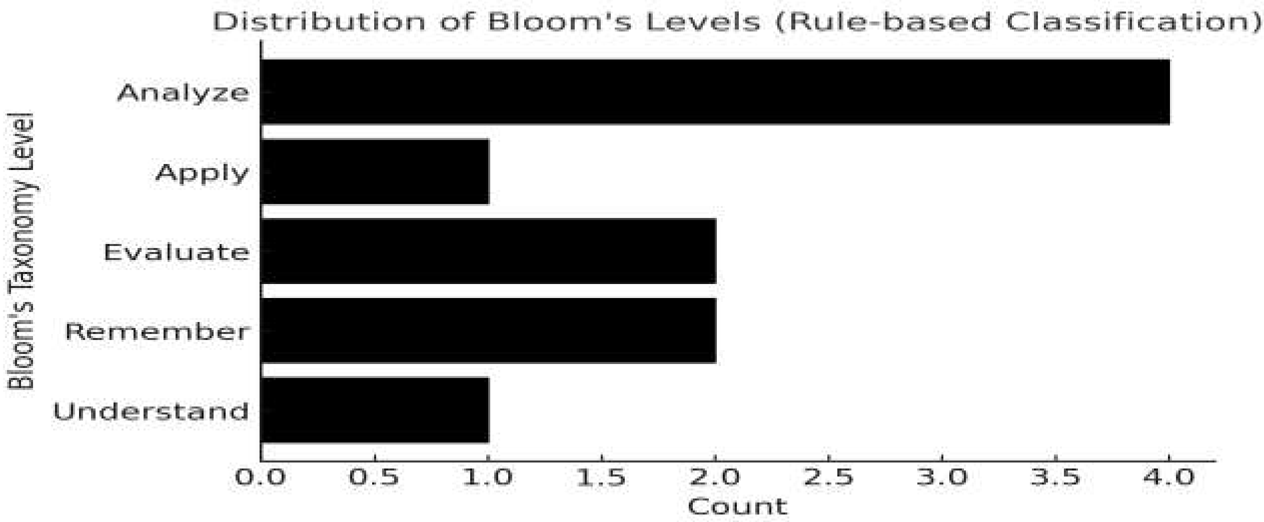

Fig. 4. Distribution of questions across Bloom's taxonomy levels based on rule-based classification.

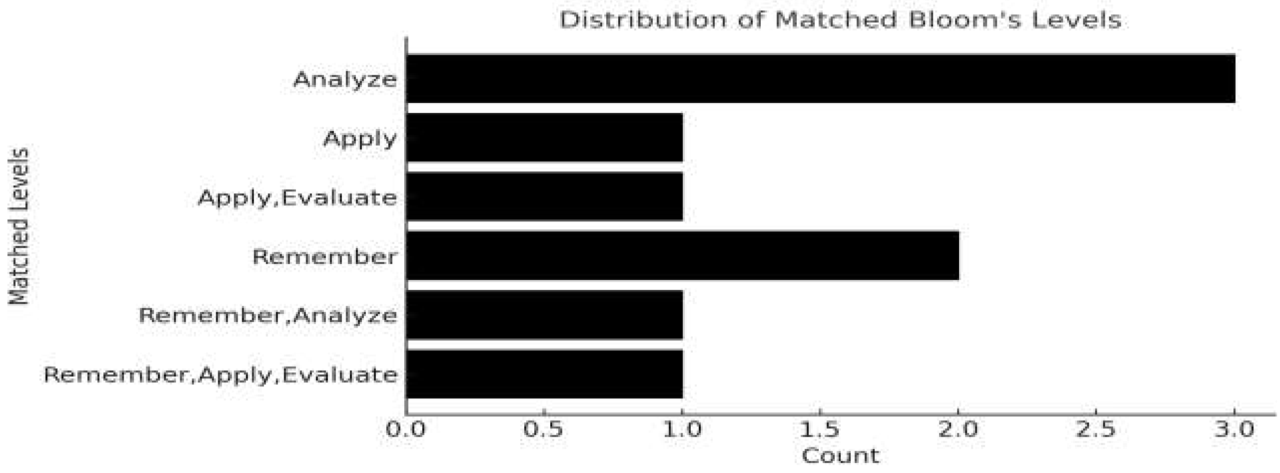

Fig. 5. Matched Bloom's taxonomy levels showing single and multiple-level associations for questions

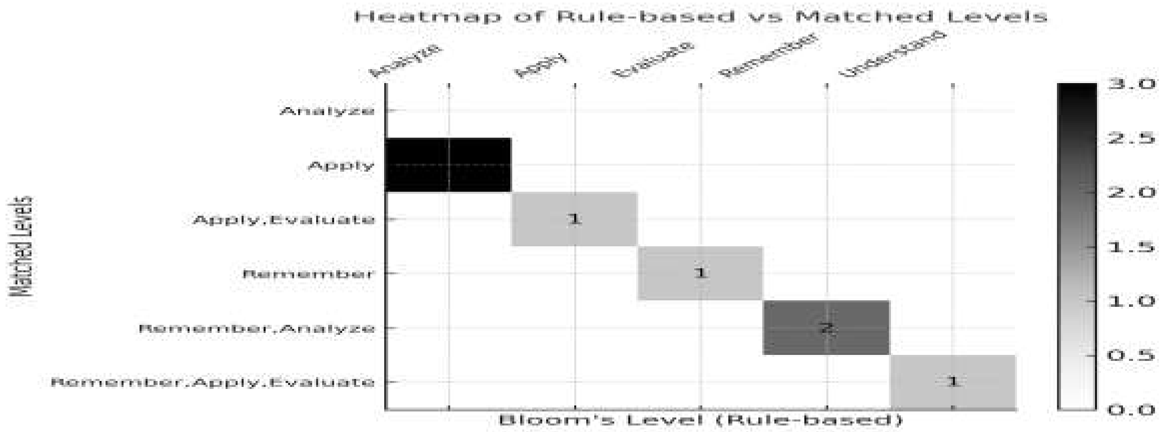

Fig. 6. Heatmap illustrating the correspondence between rule-based Bloom's level classification and matched levels.

The effectiveness of the Support Vector Machine (SVM) and the k-Nearest Neighbor (K-NN) classifiers is evaluated using four standard measures: Accuracy, Precision, Recall, and F-measure. These metrics are commonly employed in classification tasks to assess the overall performance of a system.

1. Accuracy

Accuracy represents the ability of the classifier to correctly assign questions to their proper Bloom's taxonomy category. It is defined as the ratio of the number of correctly classified questions (True Positives + True Negatives) to the total number of questions (True Positives + True Negatives + False Positives + False Negatives).

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \qquad \text{-- (1)}$$

2. Precision

Precision measures the proportion of questions that were correctly classified among all the questions predicted for a given category. It is expressed as the ratio of True Positives to the sum of True Positives and False Positives.

Precision = TP / (TP + FP) -- (2)

3. Recall

Recall, also known as sensitivity, reflects the proportion of questions correctly identified among all the questions that actually belong to a given category. It is calculated as the ratio of True Positives to the sum of True Positives and False Negatives.

Recall = TP / (TP + FN) -- (3)

4. F-measure

The F-measure, also called the F1-score, combines Precision and Recall into a single metric by computing their harmonic mean. This provides a balanced evaluation of a classifier's effectiveness when both false positives and false negatives are of concern.

F1-score = 2 × (Precision × Recall) / (Precision + Recall) -- (4)

For multi-class classification, these values are calculated separately for each class and then averaged across all classes to obtain the macro-average, which ensures that each class contributes equally to the overall evaluation.

The terminology used in these measures is as follows:

• True Positive (TP): Questions correctly classified into their actual category.

• False Positive (FP): Questions incorrectly assigned to a category they do not belong to.

• False Negative (FN): Questions that belong to a category but were not correctly classified into it.

• True Negative (TN): Questions correctly identified as not belonging to a given category.

For multi-class classification, these measures are computed for each class individually and then averaged across all classes to obtain the macro-averaged scores. This provides an unbiased measure of classifier performance irrespective of class imbalance.

- True Positive (TP): The set of questions correctly assigned to their actual category.
- False Positive (FP): The set of questions incorrectly assigned to a category they do not belong to.
- False Negative (FN): The set of questions that were not assigned to their correct category.
- True Negative (TN): The set of questions correctly identified as not belonging to a category.

The experimental evaluation demonstrates the comparative performance of SVM and KNN classifiers under different validation strategies. As shown in **Table IV**, using 5-fold cross-validation, SVM achieves superior results with an accuracy of 0.980 and an F1-score of 0.956, significantly outperforming KNN (k=5k=5k=5), which records lower values across all metrics. A similar trend is observed in **Table V**, where under a 70/30 train–test split, SVM attains an accuracy of 0.938 and an F1-score of 0.915, while KNN lags behind with 0.750 accuracy and 0.543 F1-score. Finally, in **Table VI**, when the training and testing sets are identical, SVM reaches perfect performance (1.000 for both accuracy and F1-score), whereas KNN, although improved, still records comparatively lower values. These results consistently highlight the robustness and generalization capability of SVM over KNN in the proposed framework.

Table4: 5-Fold Cross-Validation (Macro-Averaged Results)

Model	Accuracy	Precision	Recall	F1-Score
KNN_k5	0.725	0.507	0.547	0.504
SVM	0.980	0.952	0.960	0.956

Table 5: Case-1 (70/30 Train/Test Split)

Model	Accuracy	F1-Score
SVM	0.938	0.915
KNN_k5	0.750	0.543

Table 6: Case-2 (Train = Test)

Model	Accuracy	F1-Score
SVM	1.000	1.000
KNN_k5	0.902	0.856

The confusion matrices provide detailed insights into the classification performance of SVM and KNN across different experimental cases. As shown in Fig. 7**,** the SVM model in Case-1 achieves strong diagonal dominance, correctly classifying most instances with minimal misclassifications. In contrast, Fig. 8 highlights that KNN (k=5) misclassifies certain "Remember" and "Apply" instances as "Understand," indicating weaker separation boundaries. For Case-2, Fig. 9 demonstrates that SVM achieves near-perfect classification, with all Bloom's taxonomy categories accurately identified. Meanwhile, Fig. 10 shows that although KNN improves under Case-2, it still exhibits confusion between "Remember" and "Understand" levels. These results further confirm the superior consistency and reliability of SVM compared to KNN in Bloom's taxonomy-based question classification.

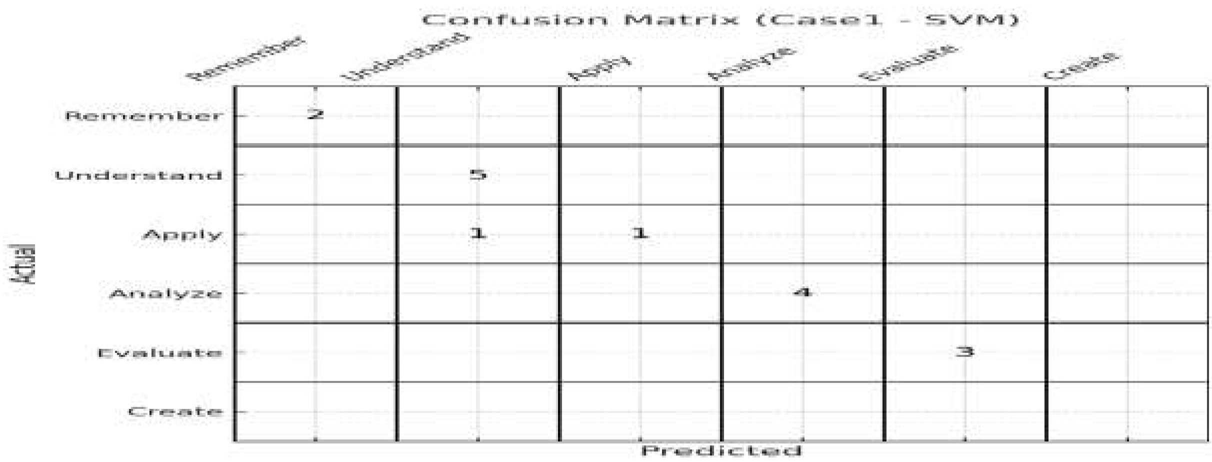

Fig.7. Confusion matrix results for the test and score data of SVM (Case 1)

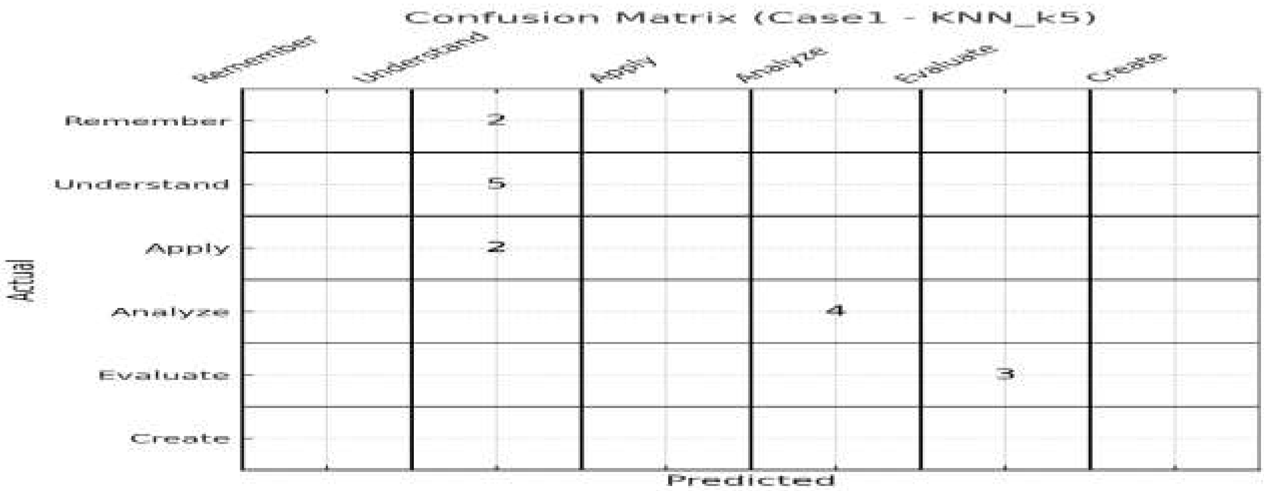

Fig.8. Confusion matrix results for the test and score data of KNN (Case 1)

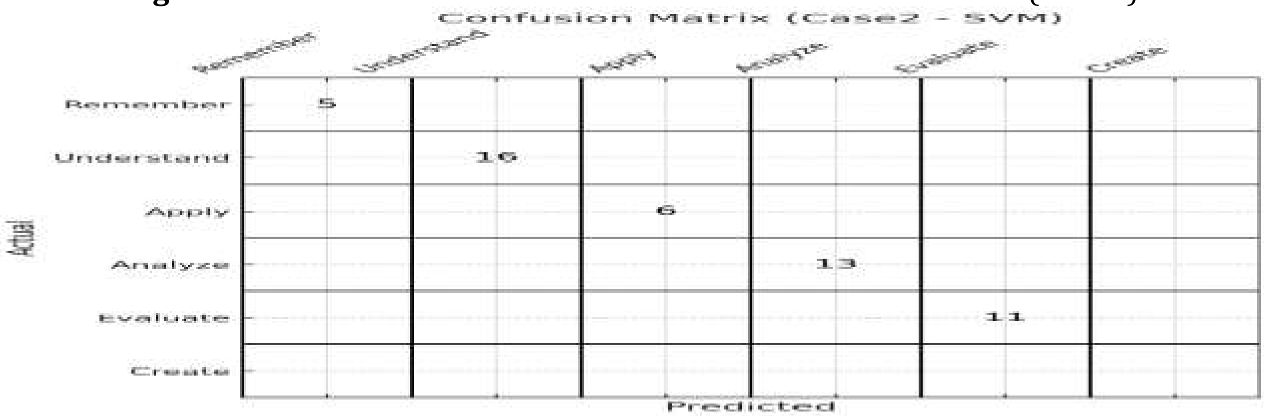

Fig.9. Confusion matrix results for the test and score data of SVM (Case 2)

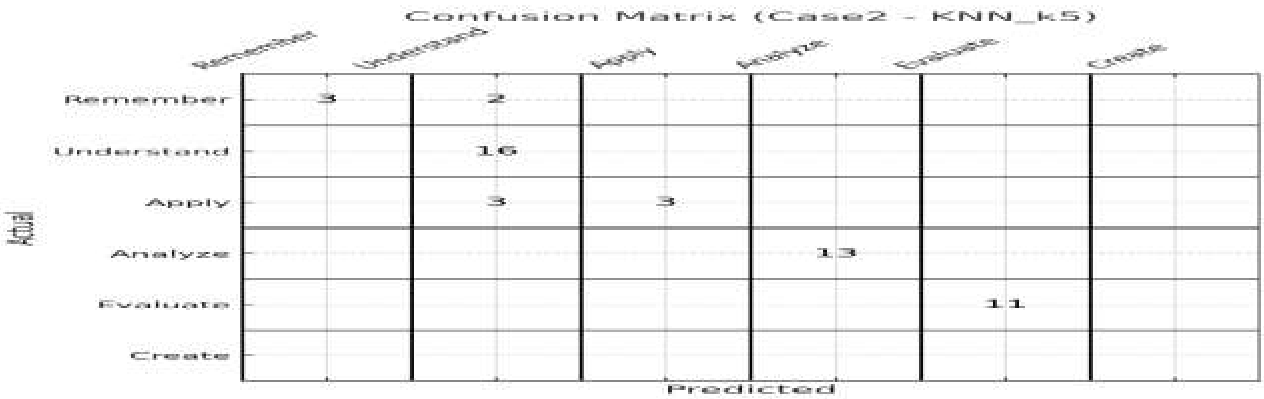

Fig.10. Confusion matrix results for the test and score data of KNN (Case 2)

Across all metrics, SVM consistently outperformed K-NN—a finding that aligns with prior studies emphasizing the strength of SVM in this domain.

VI. Future Work

Future research will focus on expanding the dataset with a larger and more diverse set of subject-specific questions to improve generalizability. Advanced natural language processing techniques such as word embeddings, semantic similarity measures, and transformer-based models (e.g., BERT, GPT) will be explored to enhance feature representation. In addition, hybrid classification frameworks that combine rule-based and machine learning approaches will be developed to improve accuracy and interpretability. The integration of this system into adaptive learning platforms and automated assessment tools will also be investigated to provide real-time support for educators and learners.

VII. Conclusion

In this work, an automated framework for classifying educational questions into the Revised Bloom's Taxonomy levels was presented using Support Vector Machine (SVM) and *k*-Nearest Neighbor (K-NN) classifiers. The system incorporated grammar validation, in-context keyword analysis, and verb-based feature extraction to ensure that only relevant and properly structured questions were processed. Performance was evaluated using accuracy, precision, recall, and F-measure, with macro-averaging applied to handle the multi-class nature of the problem.

Experimental results demonstrated that the SVM classifier consistently outperformed the K-NN model across all evaluation scenarios. SVM achieved high accuracy and F1-scores in both the 70/30 train–test split and 5-fold cross-validation, confirming its robustness in handling variations across multiple Bloom's categories. On the other hand, K-NN showed moderate performance, indicating its limitations in effectively capturing the discriminative features of educational text.

The findings confirm that SVM is a more effective choice for question classification tasks within the Revised Bloom's framework, ensuring better reliability in supporting automated assessment and instructional design. Future work will focus on expanding the dataset with diverse subjects, incorporating advanced natural language processing techniques such as word embeddings and transformer-based models, and exploring hybrid classifiers to further enhance classification accuracy and generalization

References

[1] M. Mohammed and N. Omar, "Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec," *PLOS ONE*, vol. 15, no. 3, pp. 1–21, Mar. 2020, doi: 10.1371/journal.pone.0230442.

[2] D. A. Abduljabbar and N. Omar, "Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination," *Journal of Theoretical and Applied Information Technology*, vol. 78, no. 3, pp. 447–455, Aug. 2015.

[3] J. Zhang, C. Wong, N. Giacaman, and A. Luxton-Reilly, "Automated classification of computing education questions using Bloom's taxonomy," in *Proc. Australasian Computing Education Conf.*, Feb. 2021, pp. 58–65, doi: 10.1145/3441636.3442305.

[4] K. A. Osadi, M. Fernando, and W. V. Welgama, "Ensemble classifier based approach for classification of examination questions into Bloom's taxonomy cognitive levels," *Int. J. Comput. Appl.*, vol. 162, no. 4, pp. 1–6, Mar. 2017. [Online].

[5] M. Jain, R. Beniwal, A. Ghosh, T. Grover, and U. Tyagi, "Classifying question papers with Bloom's taxonomy using machine learning techniques," in *Proc. Int. Conf. Advances in Intelligent Systems and Computing*, Springer, 2019, pp. 399–408, doi: 10.1007/978-981-13-9942-8_38.

[6] A. Anindditya, M. A. Hasibuan, and E. Sutoyo, "Text mining approach using TF-IDF and Naive Bayes for classification of exam questions based on cognitive level of Bloom's taxonomy," in *Proc. Int. Conf. Internet of Things: Systems, Management and Security (IOTAIS)*, Nov. 2019, pp. 112–117, doi: 10.1109/IOTAIS47347.2019.8980428.

[7] C. P. Magas, L. D. Gruppen, M. Barrett, P. H. Dedhia, and G. Sandhu, "Intraoperative questioning to advance higher-order thinking," *Am. J. Surg.*, vol. 213, no. 2, pp. 222–226, Feb. 2017, doi: 10.1016/j.amjsurg.2016.08.027.

[8] K. Makhlouf, L. Amouri, N. Chaabane, and N. El-Haggar, "Exam questions classification based on Bloom's taxonomy: Approaches and techniques," in *Proc. Int. Conf. Computer and Information Sciences (ICCIS)*, Oct. 2020, doi: 10.1109/ICCIS49240.2020.9257698.

[9] K. Jayakodi, M. Bandara, and I. Perera, "An automatic classifier for exam questions in engineering: A process for Bloom's taxonomy," in *Proc. IEEE Int. Conf. Teaching, Assessment, and Learning for Engineering (TALE)*, Dec. 2015, pp. 195–202, doi: 10.1109/TALE.2015.7386043.

[10] S. S. Haris and N. Omar, "A rule-based approach in Bloom's taxonomy question classification through natural language processing," in *Proc. Int. Conf. Computing and Convergence Technology (ICCCT)*, Dec. 2012, pp. 410–414. [Online].

[11] J. Huang *et al.*, "Automatic classroom question classification based on Bloom's taxonomy," in *Proc. ACM Conf. Education*, 2021, doi: 10.1145/3498765.3498771.

[12] K. Soudan and J. Sleeman, "An evaluative analytical study of the questions in natural geography textbook assigned for the first secondary class in Syrian Arab Republic schools according to Bloom's cognitive taxonomy," Jan. 2009. [Online].

[13] M. Abdelrahman, "An analysis of the tenth grade English language textbooks questions in Jordan based on the revised edition of Bloom's taxonomy," *J. Educ. Pract.*, vol. 5, no. 18, pp. 139–151, Jan. 2014. [Online].

[14] A. Sangodiah, Y. T. Fui, L. E. Heng, N. A. Jalil, R. K. Ayyasamy, and K. H. Meian, "A comparative analysis on term weighting in exam question classification," in *Proc. Int. Symp. Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Oct. 2021, doi: 10.1109/ISMSIT52890.2021.9604639.

[15] A. M. Patil, "Assessment of basic mechanical engineering course question paper using Bloom's taxonomy," *Int. J. Adv. Res. Innov. Ideas Educ.*, vol. 3, no. 5, pp. 745–753, Jan. 2017.

[16] A. Waheed, M. Goyal, N. Mittal, D. Gupta, A. Khanna, and M. Sharma, "BloomNet: A robust transformer-based model for Bloom's learning outcome classification," *arXiv preprint arXiv:2108.07249*, Aug. 2021.

[17] M. Gr. Voskoglou, "A Markov chain application on the levels of the Bloom's taxonomy of learning objectives," *Am. J. Educ. Res.*, vol. 7, no. 3, pp. 294–298, Mar. 2019, doi: 10.12691/education-7-3-15.

[18] S. Das, S. K. Das Mandal, A. Basu, and A. Basu, "Identification of cognitive learning complexity of assessment questions using multi-class text classification," *Contemporary Educational Technology*, vol. 12, no. 2, Jun. 2020, doi: 10.30935/cedtech/8341.

[19] J. C. A. Sami and U. Arumugam, "A descriptive analysis of students learning skills using Bloom's revised taxonomy," *J. Comput. Sci.*, vol. 16, no. 2, pp. 183–193, Feb. 2020, doi: 10.3844/jcssp.2020.183.193.

[20] C. Şanli, "Investigation of question types in high school geography coursebooks and their analysis in accordance with the revised Bloom's taxonomy," *Eurasian J. Educ. Res.*, vol. 28, no. 2, pp. 111–127, Dec. 2019.

[21] R. Meissner, D. Jenatschke, and A. Thor, "Evaluation of approaches for automatic e-assessment item annotation with levels of Bloom's taxonomy," in *Proc. Int. Conf. Technology Enhanced Learning (EC-TEL)*, Springer, Cham, 2020, pp. 57–69, doi: 10.1007/978-3-030-66906-5_6.

[22] S. L. Pugh and J. Gates, "The application of Bloom's taxonomy to higher education examination questions in physics," *New Directions in the Teaching of Physical Sciences*, no. 16, May 2021, doi: 10.29311/ndtps.v0i16.3674.

[23] R. C. Ginting, I. Hartoyo, and A. T. Saragih, "An analysis of reading questions in students' worksheet for grade eight based on Bloom's taxonomy," *REGISTER J. English Lang. Teaching FBS-Unimed*, vol. 8, no. 1, Aug. 2020.

[24] S. L. Zorluoğlu and Ç. Güven, "Analysis of 5th grade science learning outcomes and exam questions according to revised Bloom taxonomy," *J. Electron. Imaging*, vol. 6, no. 1, pp. 58–69, Mar. 2020, doi: 10.5296/jei.v6i1.16197.