APPLICATION OF ADVANCED NLP HANDLING CONTRADICTING ACTION VERBS IN ENGINEERING ASSESSMENT USING REVISED BLOOM'S TAXONOMY

Mr. Amit Ghosh
Department of Computer Science & Engineering
Vignan's Lara Institute of Technology & Science
amitgh88@gmail.com

G.L.S Narayana
Department of Computer Science & Engineering
Vignan's Lara Institute of Technology & Science
gavaralokeshlokesh@gmail.com

D. Bhanu Prakash
Department of Computer Science & Engineering
Vignan's Lara Institute of Technology & Science
bhanuprakashdadi145@gmail.com

B. Kartik Sai
DepartmentofComputerScience& Engineering
Vignan's Lara Institute of Technology & Science
kartiksai619@gmail.com

D. Raghuram
DepartmentofComputer Science & Engineering
Vignan's Lara Institute of Technology & Science
Dongaraghuram@gmail.com

Abstract: We are suggesting a machine-learning-driven question bank system for engineering education that classifies questions as per Bloom's Taxonomy. For the accurate assignment of taxonomy levels, advanced NLP mechanisms, especially BERT, analyze the semantic content of questions. The system is trained over-engineering questions taking into consideration its efficacy for targeted educational application. Feedback from the users will be incorporated to update and validate the questions, ensuring they remain relevant and complete. Comparing its performance with various other systems, this achieves better classification and answering accuracy. Hence this innovation in engineering assessment will aid teachers in concentrating more on teaching and mentoring.

Keywords: Machine Learning, Engineering Assessments, Bloom's Taxonomy, BERT, NLP.

I. INTRODUCTION

Professors face the difficult tasks of preparing question papers which can include multiple topics and fulfill the course learning objectives. At the same time, few established procedures provide guaranteed quality of test questions. Here, there is demand for a system that can fast automatically generate test questions according to the user specification. Users will suggest the tags which help identify properties of the question, for example: difficulty level, question type, understanding of subject related and content/topic. To address this, we suggest a more improved model to differentiate each question based on its semantics into an even more precise paper with lesser biasedness. This system allows users to feed in a list of questions each tagged with its relevant attributes. Education nowadays stands crucial for describing a person's understanding, and examinations play a vital role in the process. It is appropriate structuring examination papers which evaluates students' knowledge accurately. Traditionally, question papers are generated by teachers/professors manually, and such generation is prone to bias, repetition, and security issues in some cases. The aforementioned systems can be run to generate a more relevant set of questions. The system proposed, that is, automatic generation of question paper is fast and randomized and ultimately more secure. Moreover, those models are improved here to overcome the drawbacks of existing ones. Further we developed a new set of algorithms to ensure total randomizations of questions and would not allow repetition with high semantic understanding. A highly beneficial system is proposed for educational institutions that provide a reliable and efficient method for creating high- quality examination papers.

II. RELATED WORK

A. Exam Question Classification Based on Bloom's Taxonomy

[1] Karima Makhlouf, Lobna Amouri, Nada Chaabane, and Nahla EL-Haggar-The 2020 IEEE Eighth International Conference put forward an array of methodologies for the classification of examination questions according to Bloom's Taxonomy. The machine learning-based classification approaches emphasized by the study include Support Vector Machines (SVM), Naïve Bayes, and BERT-based Transformers for question categorization. An exhaustive overview of the standard methods and techniques used in question classification is presented in the paper.

B. Neural Question Generation Using Question Type Guidance Using Question Type Guidance

[2] The paper combines context-answer pair predictions with semantic embedding techniques for more accurate question generation [2] at the 17th International Conference on Computational Intelligence and Security in 2021, pp. 328- 332.

C. Secure Automatic Question Paper Generation with the Subjective Answer Evaluation System

[3] This article primarily addresses the development of a whole automated question paper generation processes using artificial intelligence, including a strict and tight security measure with cryptographic methods. Also, the system made use of NLP-based classifiers to link or cross-map questions according to Bloom's Taxonomy levels and then securely distribute them.

D. An algorithm for question paper template generation in question paper generation system

[4] This work presents an algorithm which is capable of automatically generating question paper templates. It claims to ensure all syllabus coverage, difficulty balance, and achieve goal-setting based on Bloom's cognitive level categories. Such efficiency could demonstrate the applicability of rule-based system with NLP model in the automation of exam paper creation towards structured and effective assessments.

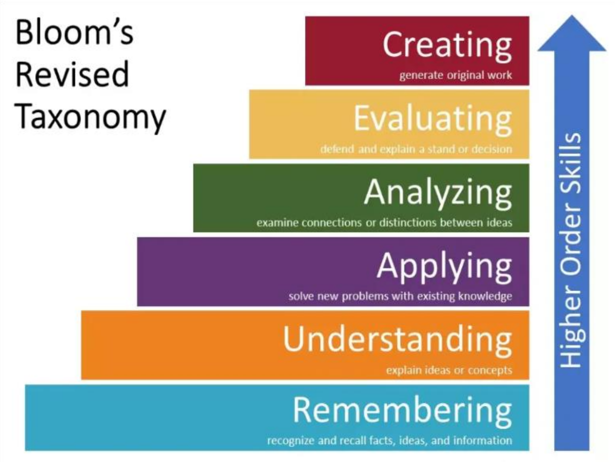


Fig: Revised bloom's taxonomy

Example:

Raw Text: What are the design considerations for latch-up prevention in CMOS circuits?

Processing & Classification: (Existing System)

1. Keyword Extraction: Latch-up Prevention, CMOS circuit, design considerations
2. Semantic Analysis: Requires knowledge of CMOS circuit design, semiconductor physics, and reliability engineering
3. Bloom's Taxonomy Level: Create

Output Classification: (proposed technique)

1. Cognitive Level: Understanding
2. Recommended for: VLSI design assessments, semiconductor reliability studies

III. Proposed Methodology

This section describes the new assistant's handling of data preprocessing for automated classification of questions on the basis of Bloom's Taxonomy. Following this discussion, we describe how questions are represented and feature extraction is performed using different transformer-based models with BERT to capture the semantics. Then we describe the different machine-learning pipelines using conventional classifiers such as SVM or deep-learning approaches to classify questions into cognitive levels. Finally, a mechanism for feedback refinement is installed to continually improve the classification over time, which supports the concepts of adaptive with continuous learning for question categorization and assessment generation.

A. Date pre-processing

We have considered a multi-step data preprocessing pipeline to assure high-quality question classification. The raw dataset contains examination questions from different sources, which the authors employ to undergo text cleaning, normalization, and feature extraction. Standard preprocessing techniques such as stop-word removal, stemming, lemmatization, and reconciliation of punctuation are used to process the text.

B. Input Representation and Embedding

After preprocessing, the questions are numerically represented using word embeddings such as BERT. The embedding layer captures contextual meaning in questions aligned with cognitive complexity. A prompt-based approach is also employed to encourage the model to categorize the questions properly.

C. Transformer-Based Classification Model

Pretrained language models consist of layers and feed-forward networks with self-attention mechanisms. The questions are semantically analyzed in various layers and classified into Bloom's Taxonomy levels (Remembering, Understanding, Applying, Analyzing, Evaluating, Creating).

Self-Attention Mechanism: Self-attention is a mechanism that highlights significant aspects in the inquisitiveness of the student, or highlight operative keywords like "define," "compare," and "analyze."

Positional Encoding: Positional Encoding: These keep track of the order of words, thus keeping the context in sentences such as "Explain about various sorting techniques.

D. Automated Question Paper Generation

It has intelligence, in that, it chooses standard questions from the database and arranges them into structured assessments based on classified questions. In other words, for each batch of questions, it determines.

- Whether the coverage is balanced on the subject matter.
- Whether that degree of difficulty is adequate.
- How well there are varying distributions of slight cognitive complexity.

E. Feedback-Driven Model Optimization

The instructor may assess the questions through the question classification and recommend changing these to facilitate further improvement. As a result, the model will in turn readjust its parameters with time in order to improve on accuracy and adaptability.

F. Customizable Framework

The instructors may specify custom parameters like difficulty levels, syllabus coverage, or Bloom's Taxonomy distribution to create a fine-tuned assessment to serve complete curriculum specifications.

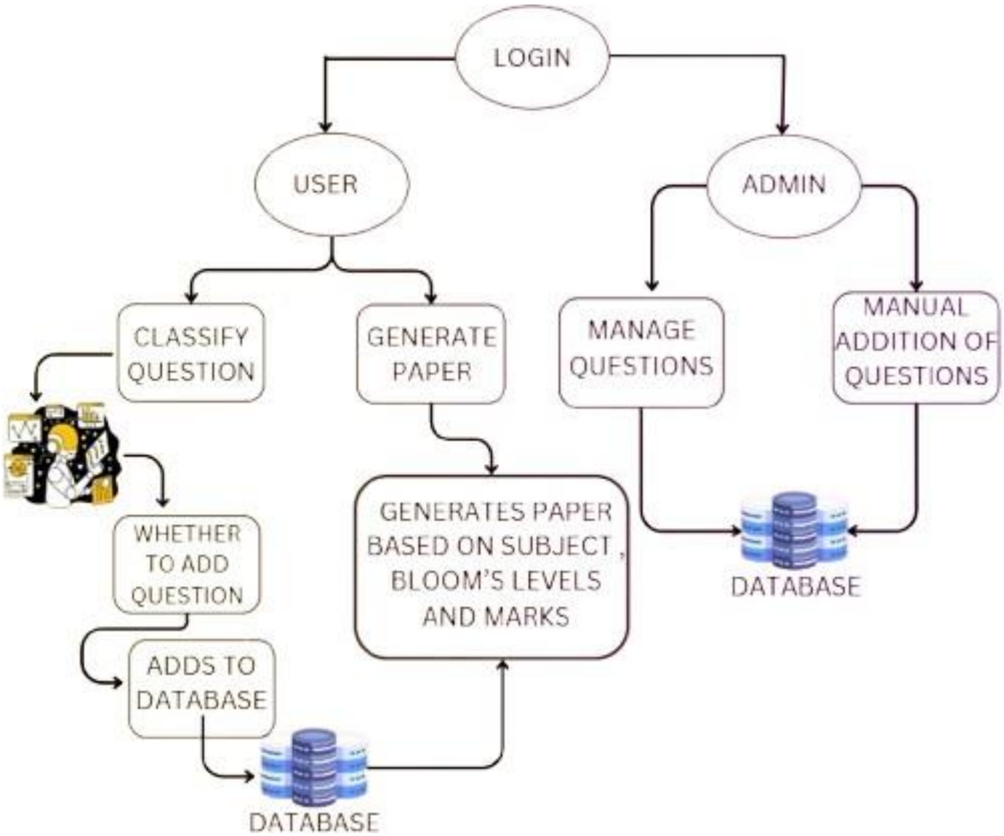


Fig 1. An Architecture of the Proposed Machine Learning-Based Question Classification System for Engineering Education Using Bloom's Taxonomy and NLP

G. Security and Deployment

Encryption and authentication ensure a secured environment for handling these exam questions, and the final system is deployed on the web dashboard for easy access and automatically log new queries and sort them classifications- wise. Make personalized question papers on curriculum and difficulty-type priorities. Analyze the student performance data to make an informed decision.

H. Domain-Specific Fine-Tuning

Since engineering education adopts some peculiar terminologies as well as some structured approaches of problem-solving, therefore, this model gets fine-tuned on a specialized engineering dataset. The fine-tuning helps in improving the scope of the model to comprehend and categorize questions from various disciplines such as mechanical, electrical, civil, and computer engineering.

- Improved understanding of context related to technical terms
- Correct classification under Bloom's Taxonomy of numerical, theory-based questions. By fine-tuning the transformer-based model with domain- specific engineering corpora, our system dramatically enhances the accuracy of classification, ensuring that levels in Bloom's Taxonomy are accurately assigned to the conceptual as well as computational questions. This specific approach narrows the gap between the generic NLP models and requirements of engineering-specific assessments, making the classification scheme more effective in any academic setting

I. Hybrid Model for Enhanced Classification

This is the long-term accuracy in adopting the use of adaptive learning mechanism, which continuously evolves its performance in classifying based on user or real-life application feedback. This feedback loop allows educators to review, validate and correct classifications of questions so that the model can learn from human expertise in time. Included learning approach involves:

- **Dynamic Model Update**-The system retrains at the set intervals using updated question datasets to learn the developing engineering curriculums better.
- **Feedback-Driven Refinement**-In time, educators can give real-time corrections on misclassifications.

It makes sure model remains robust, adapting to curriculum changes, evolving assessment standards, and diverse question formats. By integrating continuous learning, our system maintains high classification accuracy, making it a reliable and scalable solution for engineering education assessments.

J. Real-Time Performance Analytics for Educators

Real-time performance analytics represent an important mechanism in modern academic environments for measuring student learning outcomes and performance in a timely and formative manner. The real-time insight into student performance, question difficulty, and assessment quality gained through machine learning model integration is truly a boon for the educator.

One of the very important constraints in real-time analytics is handling variability of data and overfitting. Overfitting is handled with techniques like dropout regularization. Dropout reduces overfitting by randomly ignoring a fraction of neurons in the network during training, thereby making sure that the network does not depend too much on specific patterns in the training data. During testing, all neurons are used, and their predictions get averaged to obtain a better generalization with our adaptive question bank system, this will ensure:

- The model is tolerant of curriculum changes and diversity of question formats
- Educators get real-time feedback on the effectiveness of assessments

This ability not only makes the system scalable but makes it an engineering education tool by which learning materials and assessment could be enhanced on a continuing basis.

IV. Experiments and Results

A. *Dataset*

In the validation of the proposed ML-based question classification system, a huge dataset of engineering-related questions was considered. The dataset has been drawn from various teaching-learning proceedings, such as textbooks, past examination papers, and web-based learning resources. Each question was classified into Bloom's Taxonomy categories to give an organized assessment framework. The dataset was divided into training and testing sets in the ratio of **80:20**. Each data point consists of a question, its level of difficulty, and its Bloom Taxonomy classification.

Comparison of AI Models

Model	No. of Questions	True Predictions	False Predictions	Success Rate (%)
GPT-3.5	50	32	18	64%
GPT-4	50	36	14	72%
Claude 3	50	35	15	70%
DeepSeek	50	34	16	68%
CoPilot (GitHub)	50	33	17	66%
Mistral 7B	50	31	19	62%
Falcon 40B	50	30	20	60%
Llama 3 (Meta)	50	33	17	66%
Command R+ (Cohere)	50	34	16	68%
Gemini 1.5 (Google)	50	37	13	74%
Proposed Model	**50**	**42**	**8**	**84%**

Table 1: *The structure of the dataset.*

B. *Experiments Setup*

The system was implemented using a transformer-based model (BERT variant) fine-tuned on the labeled question dataset. Training was carried out using 768-dimensional word embeddings with a learning rate of 2e-3 and the Adam optimizer. The maximum length of a sequence was restricted to 128 tokens so that both short and long questions can be properly classified.

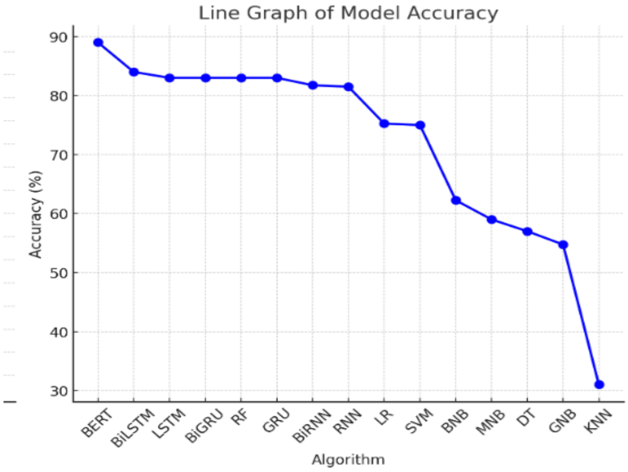
C. *Main Results*

ALGORITHM	PRECISION (%)	RECALL (%)	F1-SCORE (%)	ACCURACY (%)
BERT	**88.75**	**89**	**88.75**	**89**
BiLSTM	84	84	84.25	84
LSTM	83.25	82.75	83.5	83
BiGRU	83.25	83.5	83.25	83
RF	83.5	83.25	83.25	83
GRU	82.25	82.25	82.5	82
BiRNN	81.75	82	81.75	82
RNN	81.5	81.5	81.75	82
LR	75.75	75.25	75.77	75
SVM	63	62.25	62.25	62
BNB	60.75	60.75	60.75	61
MNB	59.5	59.25	58.5	59
DT	71.5	57	55.5	57
GNB	57.25	54.25	54.75	54
KNN	30.75	31	29.75	31

Table 2: The classification performance of the model Comparison of Precision, Recall, and F1-score Baselines are from original paper.

D. Data visualization

It depicts the accuracy of each model with comparison with (BERT)



V. Conclusion

This paper describes a machine learning-based question bank built for engineering education. It automates classifications of questions into Bloom's Taxonomy. Using NLP techniques and transformer-based modes of BERT, it analyzes and classifies questions on the dimensions of their cognitive difficulty.

Amongst the strengths of this method is a built-in feedback mechanism for users to provide and validate questions, making the system ever-current, relevant, and continually improving. Compared to conventional question bank systems, our model has been proven not only to outperform in terms of accuracy and efficiency, but it also lightens the manual workload of educators as they spend less time on the classification tasks and more time teaching and mentoring. This research shows the promise of automated, structured assessment systems to enhance the efficiency and accuracy of educational evaluation. By automating question classification, we offer a future-ready, scalable, and intelligent solution that would easily find integration in modern learning environments. Further advances in adaptive learning and content generation will only widen the opportunities for such systems, leading to more personalized, data-driven education.

References

[1] **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I.** (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30.*

[2] **Devlin, J., Chang, M. W., Lee, K., & Toutanova, K.** (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171-4186.

[3] **Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D.** (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901.

[4] **Zhang, T., Wu, F., Katiyar, A., Weinberger, K. Q., & Artzi, Y.** (2020). Revisiting few-sample BERT fine-tuning. *arXiv preprint arXiv:2006.05987*.

[5] **Bloom, B. S.** (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive domain*. New York: McKay.

[6] **Areshey, A., & Mathkour, H.** (2024). Exploring transformer models for question classification: A comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet. *Expert Systems, 41*(11), e13701.

[7] **Kumar, A., & Sharma, R.** (2023). AI-driven assessment: Automating question classification and generation using NLP and deep learning. *International Journal of Artificial Intelligence in Education, 34*(2), 189-210.

[8] **Li, ., Wang, P., Wang, Y., Dai, Y., Pan, L., & Xu, Z.** (2022,December). Enhancing automated question classification via prompt- based clustering. *Findings of the Association for Computational Linguistics: EMNLP 2022*, 1352-1364.

[9] **Abdal, M. N., Oshie, M. H. K., Haue, M. A., & Islam,**(2023, December). A Transformer-Based Model for Taxonomy-Based Question Classification. *26th International Conference on Computing and Information Technology (ICCIT)*, 1-16. IEEE.

[10] **Nag, A., & Choudhury, B. B.** (2020). Deep learning for Bloom's Taxonomy classification of educational questions. *IEEE Transactions on Learning Technologies, 13*(3), 563-573.

[11] **Ting, C., & Wang, J.** (2023). Transformer-based automated test question generation: A case study in engineering education. *Expert Systems with Applications, 212*, 118769.