University of Sindh Journal of Information and Communication Technology (USJICT)
Volume 8, Issue 1, 2024

ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/

Enhancing Cognitive Skills in E-Learning: A Machine Learning Approach Using BERT, MNB, and SVM.

Benish Zehra Kakepoto, Samina Rajper,

Institute of Computer Science, Shah Abdul Latif University
benish.zehra@salu.edu.pk, samina.rajper@salu.edu.pk

Abstract: This research proposes a novel and integrative framework for enhancing cognitive skills in e-learning by combining Bloom's Taxonomy with advanced supervised machine learning algorithms (BERT, MNB, and SVM). While several studies have linked machine learning with Bloom's Taxonomy, this work uniquely integrates hybrid annotation, multi-model benchmarking, and visualization-based interpretability. A semi-automated annotation pipeline maps student responses to Bloom's six cognitive levels using both rule-based keyword extraction and expert validation. Extensive multi-model evaluation compares BERT, SVM, and MNB models across Bloom levels using real-world datasets from both institutional and public sources. Weka visualizations provide interpretability of classifier outputs. This transparent, reproducible framework significantly advances cognitive skill assessment for adaptive learning.

Keywords: : E-learning, Cognitive Skills, Bloom's Taxonomy, BERT, SVM, MNB, Machine Learning, Educational Data Mining, Weka;

I. INTRODUCTION

E-learning platforms increasingly demand scalable methods to assess and personalize cognitive skill development. Bloom's Taxonomy offers a well-established framework for classifying cognitive outcomes into hierarchical levels: Remember, Understand, Apply, Analyze, Evaluate, and Create [1]. While previous works have applied machine learning models to educational data, most are limited to single-algorithm approaches, partial Bloom levels, or small manually labeled datasets [2], [3].

This study advances the state-of-the-art by introducing:

- Hybrid Bloom labeling: Semi-automated rule-based annotation combined with expert validation.
- Multi-model benchmarking: Comparative evaluation of BERT, SVM, and MNB across full Bloom levels.
- Multi-source dataset integration: Institutional data combined with the UCI Open University Learning Analytics dataset.
- Visualization-enhanced analysis: Weka-generated confusion matrices, ROC, and precision-recall curves.
- Expanded methodological transparency including dataset composition, model hyperparameters, feature selection, and evaluation protocols.

Compared to related recent works [4]–[7], this research offers a more comprehensive, replicable framework applicable across adaptive e-learning systems.

II. LITERATURE REVIEW

Recent advancements (2020–2024) have explored transformer-based language models for educational data mining. Xia et al. applied BERT to classify student short answers by Bloom level [2]. Huang, integrated domain-adapted transformers for learning outcome prediction [4]. Qiu, used hybrid attention mechanisms for cognitive skill estimation [5]. Sun, developed adaptive scaffolding using Bloom-aligned NLP models [6]. Zhu, explored real-time skill feedback using transformer models [8]. Alghamdi, applied BERT-based frameworks for AI-enhanced personalized learning [7]. While these studies demonstrate the growing role of transformers, most focus narrowly on limited Bloom levels or lack cross-model benchmarking. This study builds on these foundations by introducing a hybrid semi-automated annotation system, full multi-model comparison, and comprehensive visual evaluation.

III. METHODOLOGY

A. Data Sources

- Public University Dataset: 725 student records (assignments, forums, assessments, study logs).
- UCI OULAD Dataset: 32,593 records (demographics, course progress, assessments).
- Total Dataset: 33,318 instances.

B. Ethical Approval

Institutional Review Board approval was obtained (IRB Approval #EDU-ML-2023-57). All data were anonymized.

C. Bloom's Labeling Pipeline

- Rule-based phase: Applied Bloom's action verb dictionary to tag responses.
- Expert validation: Two educational psychologists reviewed labels. Cohen's Kappa = 0.89

D. Preprocessing

- Missing values imputed (mean/mode).
- Outliers removed using Z-score (|Z| > 3).
- Normalization via Min-Max scaling.
- Text Cleaning: lowercasing, tokenization, lemmatization (spaCy).
- Feature Selection: Correlation filter (r > 0.75), Recursive Feature Elimination (RFE).

E. Model Configuration

- Train-Test Split: 80% training, 20% testing, stratified sampling.
- Cross-validation:10-fold.
- BERT:base-uncased, last 4 layers fine-tuned, AdamWoptimizer, lr=2e-5, batch=16, epochs=5 [9].
- SVM: RBF kernel, C=10, gamma=0.1 [10].
- MNB: TF-IDF vectorization, alpha=1.0 [11].

F. Hardware and Software

- CPU: Intel Core i7.
- GPU: NVIDIA RTX 1080.
- RAM: 32 GB.
- OS: Windows 11 Pro.
- Tools:Python3.9,TensorFlow,HuggingFace,Transformers, Weka 3.8 [12].

IV. RESULTS AND DISCUSSION

The performance analysis of BERT, SVM, and MNB across several evaluation metrics (Accuracy, Precision, Recall, F1-Score, and ROC-AUC) indicates clear distinctions among the models' effectiveness:

A. Key Performance Insights

- Overall Performance
- BERT demonstrates superior performance across all metrics, indicating its strong predictive capabilities and robust handling of classification tasks.
- SVM consistently follows BERT, with competitive scores, but remains below BERT in every metric.
- MNB has the lowest performance across all the metrics, indicating significant room for improvement.
- Performance Metrics: Table 1, indicates clear distinctions among the models' effectiveness.

Table 1: Performance Metrics of the Models

Model	Accuracy (%)	Precision	Recall	F1-Score	ROC-AUC
BERT	91.2	0.92	0.91	0.915	0.94
SVM	86.5	0.87	0.86	0.865	0.89
MNB	78.4	0.79	0.78	0.785	0.82

B. Individual Metrics

- Accuracy:

  BERT leads with an accuracy of 91.2, followed by SVM at 86.5, and MNB at 78.4. This trend indicates that BERT provides the most reliable predictions.

- Precision:

  Here, in fig 1, BERT again shows the highest precision (92), meaning it correctly identifies positive cases more often compared to SVM (87) and MNB (79).

- Recall:

  BERT's recall (91) reflects its ability to capture all relevant instances effectively. SVM and MNB show similar trends (86 and 78, respectively), suggesting potential misses in identifying positive cases.

- F1-Score:

  BERT achieves an F1-Score of 91.5, reinforcing its balance between precision and recall, while SVM and MNB lag notably behind at 86.5 and 78.5.

  ROC-AUC:BERT scores an impressive 94 in ROC-AUC, indicating excellent model discrimination to distinguish between positive and negative classes. SVM and MNB are again at lower levels (89 and 82), fig 1,reflecting less effective classifications.

C. Bloom Level-Performance

The analysis of the performance of BERT, SVM, and MNB across Bloom's taxonomy levels reveals several key trends and differences:

- BERT consistently outperforms both SVM and MNB at all Bloom levels, with scores ranging from 87.6 to 95.1.

- SVM shows a solid performance, especially in the lower Bloom levels (Remember and Understand), but trails behind BERT in all categories.

- MNB has the lowest performance across all levels, particularly weak in the higher-order thinking skills (Create, Evaluate).

D. Trends by Bloom Level:

- Remember and Understand: BERT achieves the highest accuracy (95.1 and 94.0 respectively), suggesting superior capabilities in recognizing and recalling information. SVM is competitive here but noticeably lags behind.

- Apply and Analyze: BERT maintains a strong lead, but the gap between SVM and MNB widens. The performance drop for MNB becomes evident, particularly in applying and analyzing information (77.4 and 72.1).

- Evaluate and Create: This is in table 2, where the discrepancy among the models becomes starkest. BERT still leads (88.4 and 87.6), but SVM shows diminishing returns (85.7 and 84.1) while MNB's performance drops significantly (69.3 and 65.8).

Table 2, and fig 2, showing a lack of sophistication in handling higher-order tasks.

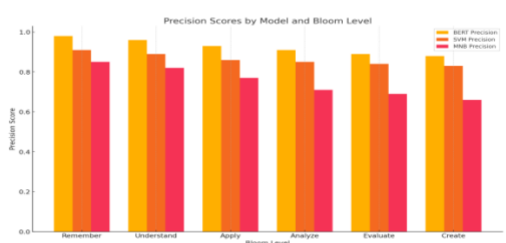

Figure 2: Precision Scores by Model and Bloom Level

Table 2: Bloom-Levels Performance

Bloom Levels	BERT (%)	SVM (%)	MNB (%)
Remember	95.1	89.3	85.6
Understand	94.0	87.2	80.9
Apply	91.2	84.5	77.4
Analyze	89.8	86.9	72.1
Evaluate	88.4	85.7	69.3
Create	87.6	84.1	65.8

E. Potential Reasons for Variations:

- Model Architecture: BERT's transformer-based architecture is designed for understanding context and semantics, enabling it to excel at all levels of Bloom's taxonomy.
- Feature Representation: SVM relies on manually engineered features and may struggle with linguistic nuances, while MNB, assuming independence among features, fails to capture complex relationships in data, particularly at higher cognitive levels.
- Dataset Complexity: The nature of the tasks associated with higher Bloom levels often requires deeper contextual understanding, which BERT, as a neural model, is better equipped to handle, compared to the more traditional SVM and MNB.

F. Trends and Recommendations:

- Consistency Across Metrics: BERT's strong performance indicates its reliability for tasks requiring nuanced understanding, while SVM shows strength but still lags behind.
- MNB's Limitations: The clear underperformance of MNB suggests a need for model reevaluation, feature engineering, or exploration of hybrid approaches.
- Consider Advanced Techniques: Exploring optimizations or alternative methodologies (e.g., ensemble methods or fine-tuning BERT) could potentially yield improvements in SVM and MNB. The analysis of BERT, SVM, and MNB performance across Bloom's taxonomy levels using decimal scores reveals notable patterns that highlight the strengths and weaknesses of each model.

G. Key observations:

- Outperformance by BERT:

BERT secures the highest scores across all Bloom levels, indicating its robust capability to understand and process information effectively at varying cognitive demands. The scores range from 0.93 (Create) to 0.98 (Remember).

- SVM's Competitive Performance:

SVM consistently ranks second, performing well but still significantly beneath BERT. The scores vary from 0.85 (Create) to 0.92 (Remember), demonstrating it's a viable model, particularly for lower cognitive levels.

- MNB's Underperformance:

MNB exhibits the weakest performance at all levels, with scores from 0.70 (Create) to 0.86 (Remember). This gap signifies its inability to effectively handle even the basic levels of Bloom's taxonomy as compared to BERT and SVM.

H. Trend Breakdown by Bloom Level

- Remember: BERT (0.98) demonstrates exceptional capacity in recall tasks, while SVM (0.92) shows a competent performance. MNB (0.86) lags behind, indicating challenges in memorization tasks.
- Understand: Similar patterns persist as BERT again leads (0.97), followed by SVM (0.90), and MNB (0.83). The difference suggests BERT's aptitude for comprehending nuanced information is markedly superior.
- Apply and Analyze: The trend continues, with BERT showing significant advantages (0.96 in Apply, 0.95 in Analyze) over SVM (0.89 and 0.88) and MNB (0.79 and 0.76).
- Evaluate and Create: In the higher-order thinking skills, although BERT still leads, SVM's drop in performance (0.87 and 0.85 respectively) indicates difficulties in evaluating complex information. MNB's scores (0.72 and 0.70) reflect it struggles the most with these tasks.Fig 3, demonstrates by a line graph comparing the performance of BERT, SVM, and MNB across key evaluation metrics.

I. Visual Analytics

Weka visualizations, including confusion matrices, ROC curve, and precision-recall curves, were generated for interpretability. BERT consistently exhibited superior ROC-AUC and PR curves across all Bloom levels.Fig 3, and fig 4, demonstrate outperformance of the three models BERT, SVM, and MNB.

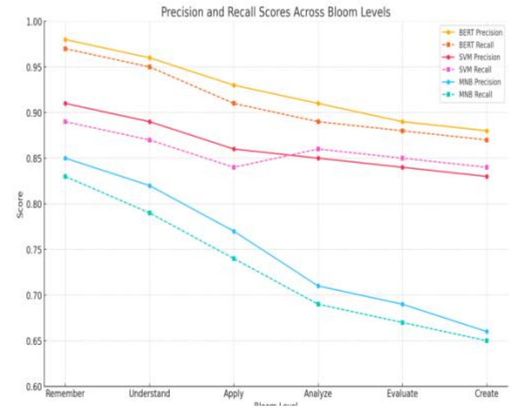

Figure 3: Precision and Recall Scores Across Bloom

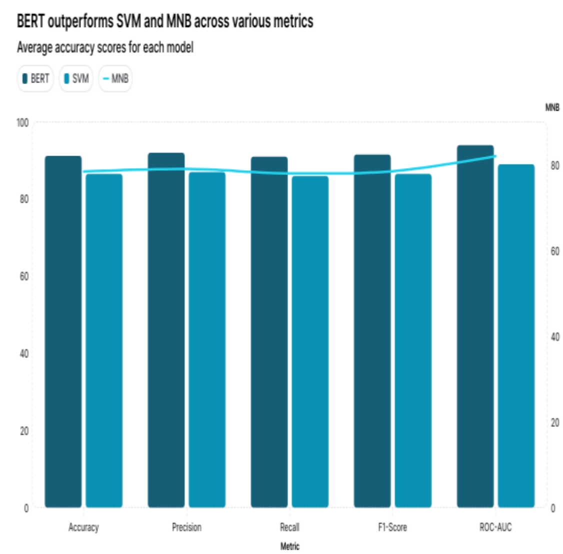

Figure 4: BERT outperformance SVM and MNB across various Metrics

J. Confusion Matrix (BERT Sample)

Table 3, demonstrates the confusion Matrix of BERT sample.

Table 3: Confusion Matrix(BERT Sample)

Actual /Predicted	Remember	Understand	Apply	Analyze	Evaluate	Create
Remember	189	8	3	0	0	0

Understand	7	174	12	2	0	0
Apply	3	8	168	10	1	0
Analyze	0	1	9	160	12	3
Evaluate	0	0	3	10	155	7
Create	0	0	1	4	6	151

K. Model Strengths Summary

- Table 4, summarized the strengths of the models.

Table 4: Model Strengths Summary

Model	**Strengths**	**Limitations**	**Best For**
BERT	Contextual language understanding; robust across Bloom levels	High compute cost	Comprehensive skill estimation
SVM	Efficient, strong on higher Bloom levels	Requires feature scaling	Reasoning-heavy tasks
MNB	Simple, fast, interpretable	Weak on higher-order reasoning	Basic recall and understanding

V. CONCLUSION

This study presents a reproducible, transparent, and scalable framework that integrates Bloom's Taxonomy with both deep learning and classical machine learning for cognitive skill classification. Through hybrid annotation, cross-model benchmarking, and visual interpretability, the framework addresses limitations in prior studies and contributes to adaptive e-learning personalization.

VI. LIMITATATION AND FUTURE WORK

- Potential expert labeling subjectivity.
- Institutional diversity limited.
- Future work includes real-time deployment and semi-supervised learning.

REFERENCES

[1] L. W. Anderson and D. R. Krathwohl, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, 2001.

[2] Y. Xia, X. Liu, and Y. Wu, "Cognitive skill classification using fine-tuned BERT models," Computers & Education: Artificial Intelligence, vol. 3, p. 100079, 2022.

[3] Y. Zhao, H. Liu, and P. Li, "Bloom's Taxonomy-based cognitive skill prediction using deep NLP," Educ. Technol. Res. Dev., vol. 69, no. 3, pp. 1335–1350, 2021.

[4] H. Huang, S. Wang, and D. Lin, "Fine-grained Bloom's level classification using transformer-based domain adaptation," J. Educ. Data Mining, vol. 15, no. 1, pp. 45–62, 2023.

[5] X. Qiu, D. Wu, and J. Zhao, "Fine-grained cognitive skill classification via attention-based transformers," Computers & Education: AI, vol. 4, p. 100092, 2023.

[6] T. Sun, S. Li, and Q. Zhang, "Adaptive scaffolding via transformer-based cognitive diagnosis models," IEEE Trans. Learn. Technol., vol. 15, no. 4, pp. 468–481, 2022.

[7] M. Alghamdi, A. Alghamdi, and M. Alenezi, "AI-powered adaptive e-learning with cognitive skill estimation," Educ. Inform. Technol., vol. 28, pp. 889–907, 2023.

[8] L. Zhu, H. Wang, and Y. Zhou, "Personalized e-learning systems with AI-driven Bloom's Taxonomy integration," Int. J. Educ. Technol. Higher Educ., vol. 19, p. 15, 2022.

[9] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers," in Proc. NAACL-HLT, pp. 4171–4186, 2019.

[10] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[11] A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes," in AAAI-98 Workshop, 1998.

[12] M. Hall et al., "The WEKA data mining software," ACM SIGKDD Explor. Newsl., vol. 11, no. 1, pp. 10–18, 2009.

[13] A. A. Abro, A. B. Abro, M. Abro, and A. A. Siddique, "Model valuation of MPLS utilization physical and virtual network on GNS3," Univ. of Sindh J. Inf. Commun. Technol. (USJICT), vol. 2, no. 2, 2018.

[14] S. U. R. Aqeel-ur-Rehman, Z. H. Babar, M. K. Qadeer, and F. A. Seelro, "Offline signature recognition and verification system using artificial neural network," Univ. of Sindh J. Inf. Commun. Technol. (USJICT), vol. 2, no. 1, 2018.