Cluster Based Classification of Students' Responses to Algebraic Questions

Okesh Ankireddypalli, Mouhitha A, Saranya Gujjula, Roshni M Balakrishnan, Peeta Basa Pati*
Dept of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
m_roshni@blr.amrita.edu
*0000-0003-2376-4591

***Abstract*—This study introduces an automated system for evaluating algebraic type questions using advanced machine learning techniques. The traditional method of manual grading is time taking and inconsistent, prompting the development of a model leveraging MathBERT, a specialized language model for mathematical text. Our methodology encompasses data preparation, embedding and feature extraction, clustering and dimensionality reduction, similarity analysis, and comprehensive validation and performance evaluation. MathBERT embeddings convert textual answers into high-dimensional numerical arrays, which are then processed using t-SNE for dimensionality reduction and K-means for clustering. Rule based approach i.e, Cosine similarity is used to measure the alignment between student answers and key solutions, categorizing them into correct, partially correct, and incorrect based on predefined thresholds. The model achieved an accuracy of 71% compared to human graders, demonstrating its potential to streamline the grading process and ensure consistency. This research contributes to automated educational assessment by providing a scalable, reliable solution for grading algebraic questions.**

***Index Terms*—MathBert Embeddings, Automated grading, Algebraic Solution, Statistical Method.**

I. Introduction

Autograding offers significant advantages for teachers and students alike. Marking work by hand is subject to time constraints and also to errors which hinders the fairness of evaluation[1]. It ensures fairness by providing consistent and unbiased evaluation across all student submissions, addressing a common issue where some teachers award full marks only for correct final answers, overlooking partial correctness and the problem-solving process. Autograding models can assign marks for partially correct solutions, thereby recognizing students understanding and effort even when the final answer is incorrect. This approach not only enhances the fairness of grading but also reduces the workload on teachers, freeing up valuable time that can be used towards personalized instruction and other educational activities. Consequently, the integration of autograding systems in educational settings fosters a more equitable and efficient learning environment.

Autograding is beneficial for algebraic questions due to the structured and logical nature of algebra. Algebraic problems often have well-defined steps and intermediate solutions that can be systematically evaluated. Traditional manual grading can be subjective, especially when it comes in giving partial marks for intermediate steps. This not only ensures fairer grading but also helps students understand their mistakes and learn more effectively. Additionally, algebra forms the foundation for many advanced topics in mathematics and science, making accurate and consistent grading crucial for building a solid understanding of these subjects.

The project employing machine learning techniques aims to develop a dependable system for autograding algebraic questions. The model has been explored with different types of algebraic questions using MathBERT. Our methodology encompasses several essential stages: pre-processing, embedding and feature extraction, clustering and dimensionality reduction, similarity computation, and validation using performance metrics. The process of data preparation consisted of assembling training datasets in mathematical format and in LaTeX format and providing with correctness labels for each answer. MathBERT embeddings converted these text-based answers into high-dimensional numerical arrays, which allowed to include semantic details of the mathematical hints [2]. t-SNE was as a tool for dimensionality reduction, thereby getting retained the locally structured data, and K-means cluster was utilized for forming the data based on questions and correctness label. The cosine similarity analysis function was used to make comparisons between student answers and key solutions by dividing them into correct, partially correct and incorrect sets based on specific levels of similarity.

The research primarily answers the following questions:

RQ1: Can cluster based classification helps in grading the mathematical responses more accurately when compared to the rule based approach?

RQ2: How well does MathBERT trained on Latex codes understand the semantics of mathematical expressions?

II. Literature Survey

Balakrishnan et al. [1] reports on automatic grading of quadratic equation problems utilizing Google's T5 Model

979-8-3503-8875-6/24/$31.00 ©2024 IEEE

to generate embeddings. The research tells both traditional and ensembled machine learning models, along with deep learning models like LSTM and BiLSTM. The fine-tuned T5 model significantly improves performance, reducing error by 70% and increasing the R^2 value to 97%. The written exam is the most commonly used method for testing students performance, but many traditional forms of assessment can be time-consuming and based upon the teacher's opinion. Sanuvala et al., in [2] suggested the Handwritten Answer Evaluation System (HAES) which uses conducting OCR analysis on the answers and the application of machine learning to grade the answers thus saving as much time as when using the conventional methods but with a similar or even better degree of effectiveness of precision. Likewise, a model is presented by Dodia et al. [3] using the BERT technique and similarity scores to rank descriptive answers with 91% accuracy. Teachers should therefore try to achieve an improvement on the structural aspect of the students' knowledge of algebra. This involves the use of instructional practices that education methods developed that help to explain how symbolic equations work in real life situations. In their work, Humberstone et al. [4] assert that such an approach is vital when conducting research on 'Profiles of Algebraic Competence' and the critical aspect of improving the students' performance in of arithmetic and algebraic word problems. Murtaza et al. [5] note that datasets are analyzed by clustering techniques by type of solutions, and using artificial intelligence in clustering students by levels of understanding and attempted learning modes to present educational content and assessments for the relevant cluster.

Spectral clustering falling under the algebraic graph theory is what being used in the solution of the clustering problem due to its efficiency. These aspects such as the construction of similarity matrices and the selection of the eigenvectors have been described by Jia et al. [6] giving important pointers for future studies. Bernius et al. [7] also propose a system coined CoFee that has a mechanism of automated feedback in textual solutions that cuts the instructor's work by grading the solutions with the help of a machine and then clusters the similar solutions through hierarchical clustering. According to Ahuja et al. [8], when information is grouped into clusters by using clustering algorithms for pooling and sorting, their way of enhancing the effectiveness of a model to truly make accurate prediction or classification is actually verified. Wu et al. [9] further analysing the clustering and classification issues continue the previous discussion on the instance retrieval efficiency and its effectiveness, gave an understanding towards information organisation directions and constructing answers for multi-parametric information search. Jing et al. [10] have developed a new technique in the text clustering that can combine the text corpora and ontologies with the vector space models to give a new definition of term similarities as opposed to the traditional term-based models. Thomas et al. [11] mentioned in the area of text classification a method to combine important similarity measures with semi supervised learning approaches where centroids are set through semi supervised clustering to increase the accuracy. Lange et al. [12], propose a cluster stability measure designed to check validity of cluster models, this is different from the problem of identifying the best number of clusters by checking how cluster solutions behave in consequence of different runs, unknown to each other. An appealing model that integrate agglomerative hierarchical clustering with Gaussian mixture and Bayesian information criterion for model selection and enhances the detection of overlapping clusters was introduced by Fraley et al. [13].

Following this, Shirafuji et al. [14] build upon this even further with their HAES tool that uses OCR and cosine similarity measures for text extraction and grading from scanned answer sheets. These advancements are a noble improvement to educational technology since they tackle challenges such as evaluating subjectivity of answers to questions, and provide accurate replacements of conventional methods. Also, Abidi et al. [15] establish that newer models including InstructGPT and ChatGPT are less affected by small modifications in programming problems formulation stressing on the significance of this format.

Automated assessment of students' answers is one of the most important tasks of the modern educational evaluation field whose purpose is to minimize the time needed for evaluation and the number of people involved in this process without the loss of evaluation quality. In recent studies Krithika et al. [16] include a new method that takes into consideration the use of two new measures, which are Semantic and Graph alignment features along with basic elements like Two-class Averaged Perceptron and Two-class Support Vector Machine where in case of application of Random Projection in the proposed method shows promising results. In regards to the smart classroom solutions, Ani R et al. [17] propose IoT-based solutions that manage to achieve the minimal energy consumption depending on the occupancy detected through sensors or cameras. Further, Narayanan et al. [18] have worked on assessment quality by tapping the possibilities of the unsupervised learning to capture inter-item correlation, and this has greatly improved the question bank calibration. In addition, Divya et al. [19] have also adopted pre-trained embedding models to grade short answers where the results are good using regression model of predicting scores from the metrics like RMSE and Pearson correlation. Anand et al. [20] have used machine learning techniques including CNN and Random Forest to improve Pedagogy in the education of the difficulties like dyslexia and dyscalculia using machine learning techniques like CNN and Random Forest.

draw.io

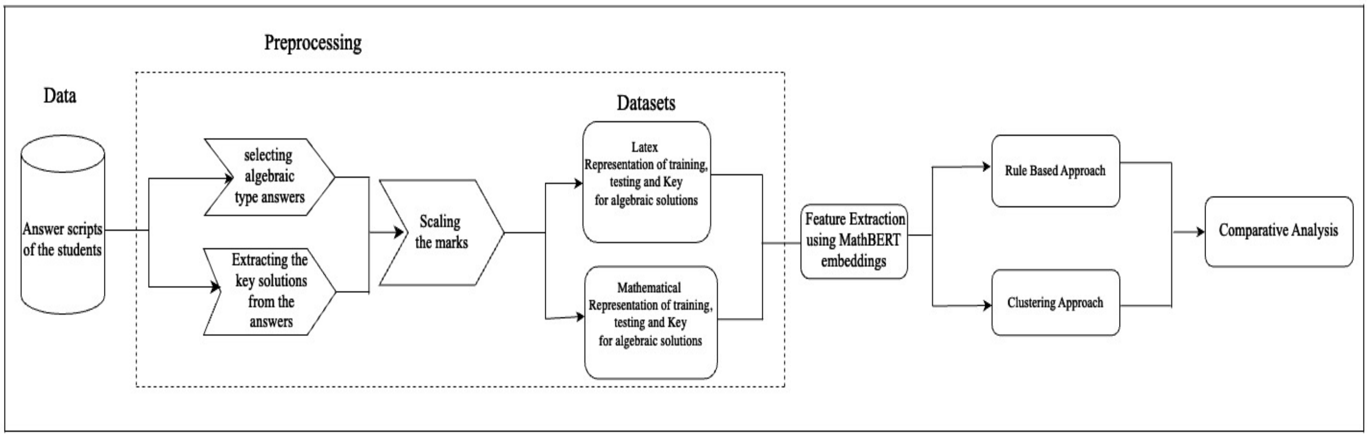


Fig. 1: System Architecture

III. Methodology

The methodology of this project involved several comprehensive steps to ensure the accurate and effective prediction of the correctness of answers based on their mathematical and LaTeX representations. The process included data preparation, embedding and feature extraction, clustering and dimensionality reduction, similarity analysis, and validation and performance evaluation. The detailed working is depicted in Fig. 1.

A. Dataset

The dataset has been prepared manually from the answer scripts of 21 different students, focusing solely on algebraic type questions drawn from all the scripts. This segregation resulted in 6 different algebraic questions. The maximum marks for the questions were derived from the answer sheet of a student whose total marks were highest among all evaluated scripts, and the maximum marks for all 6 individual questions were scaled to 5. Separate Excel sheets were created for classifying the datasets as mentioned in Table I. The testing dataset contains of 26 algebraic solutions, which was created manually based on the key solutions. These 26 solutions cover six different algebraic questions, including correct, partial, and incorrect answers.

TABLE I: Representation of Datasets

Name of the excel	Representation
Latex codes of algebraic answers of 21 answer sripts (training set)	D1
Latex code for 26 solutions (testing dataset)	D2
Mathematical algebraic solutions of 21 answer scripts (training set)	D3
Mathematical algebraic solutions of 26 solutions (testing set)	D4
Latex code of key answers for the 6 algebraic questions	latexkey
Mathematical algebraic solutions of key answers of 6 questions	mathematicalkey

B. Embeddings and Feature Extraction

The answers were captured via screenshots and processed using the Mathpix snipping tool to obtain LaTeX code for the solutions.

For clarity and ease of understanding, specific names were assigned to each Excel sheet, as detailed in Table I, which contains LaTeX and mathematical representations of solutions respectively.

To convert the textual answers into a numerical format suitable for computational analysis, MathBERT embeddings were utilized. This process transformed the answers into numerical arrays, resulting in dimensions of 126×385 for the training data and 26×385 for the additional data. MathBERT, a specialized model for mathematical text, provided a robust representation of the answers in a high-dimensional space, capturing the semantic nuances of the mathematical and LaTeX content effectively.

C. Clustering and Dimensionality Reduction

The project advanced to the crucial step of clustering and dimensionality reduction, aimed at organizing the high-dimensional data into meaningful clusters and simplifying the data for analysis.

Dimensionality Reduction: The t-SNE (t-distributed Stochastic Neighbor Embedding) technique was employed for dimensionality reduction. t-SNE is particularly well-suited for visualizing high-dimensional data by minimizing the divergence between pairwise similarities of the input objects in the high-dimensional space and the corresponding low-dimensional points in the embedding. Applying t-SNE to the high-dimensional embeddings obtained from MathBERT transformed the data into a two-dimensional space, making it more manageable and facilitating visualization. This step was essential as it helped in preserving the local structure of the data, ensuring that similar points in the high-dimensional space remained close to each other in the reduced space.

Clustering: After reducing the dimensions using t-SNE, K-means clustering was employed to group the data. K-means is a widely-used clustering algorithm that partitions the data into k clusters, where each data point belongs to the cluster with the nearest mean. The number of clusters k was varied from 2 to 6 to explore different levels of clustering granularity. This range allowed for a detailed exploration of how the data could be grouped at various levels of specificity.

The clustering process was conducted based on two main criteria:

- **Question-based Clustering:** The aim was to ensure that answers related to the same question were grouped together. This approach helped in identifying patterns and similarities within answers to the same question, providing insights into commonalities and variations in student responses.
- **Correctness-based Clustering:** The aim was to group answers according to their correctness labels (incorrect, partially correct, and correct). This step was crucial for understanding the distribution of correctness within the data and for validating the effectiveness of the embedding and similarity measures.

D. Evaluation Metrics

Several evaluation metrics as shown in Table II were computed to assess the quality of the clusters:

TABLE II: Evaluation Metrics for Different K Values

K value	Silhouette Score	Davies-Bouldin Score	Calinski-Harabasz Score	Adjusted Rand Score
2	0.26	1.18	32.12	0.10
3	0.30	1.31	32.88	0.16
4	0.33	1.51	35.20	0.18
5	0.31	1.63	32.74	0.23
6	0.29	1.68	30.21	0.25

- **Silhouette Score:** Measures how similar an object is to its own cluster compared to other clusters, providing insight into the cohesion and separation of the clusters.
- **Davies-Bouldin Score:** Evaluates the average similarity ratio of each cluster with its most similar cluster, indicating the cluster compactness and separation.
- **Calinski-Harabasz Score:** Assesses the ratio of the sum of between-cluster dispersion and within-cluster dispersion, reflecting the cluster dispersion.
- **Adjusted Rand Index (ARI):** Measures the similarity between the true labels and the clustering results, providing a normalized score to compare the clustering performance.

These metrics were evaluated for the original data, after applying SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalances, and after performing PCA (Principal Component Analysis) for dimensionality reduction. Through this analysis, PCA with SMOTE appears to be less effective for clustering compared to the t-SNE dimensionality reduction technique.

E. Cosine Similarity and Accuracy Prediction

In the similarity analysis phase, cosine similarity matrices were computed to measure the alignment between the training data and the key answers. Cosine similarity evaluates the cosine of the angle between two vectors, indicating their similarity. This was particularly useful for comparing high-dimensional embeddings from MathBERT.

A cosine similarity matrix was computed for D1 versus latexkey, resulting in a 126×6 matrix. Each row represented a training answer, and each column represented one of the six key answers. The values in this matrix ranged from -1 to 1, where 1 indicated identical vectors, 0 indicated no similarity, and -1 indicated opposite vectors.

To classify the correctness of the answers, specific thresholds were established. They are as follows:

- **Incorrect Answers:** Scores below 0.5. This threshold indicated that the student's answer was substantially different from the key answer.
- **Partially Correct Answers:** Scores between 0.5 and 0.7. This range represented answers that shared some similarities with the key answers but were not entirely accurate.
- **Correct Answers:** Scores above 0.7. This threshold indicated a high degree of similarity, suggesting the answer was accurate and complete.

Predictions were made by applying these thresholds to the cosine similarity scores. For example, for the first question, the first column of the first 21 rows of the matrix was used. This systematic approach was repeated for each question, ensuring each training answer was compared against the relevant key answer, similarly it was also done for the testing dataset.

The accuracy of these predictions was validated by comparing them with the actual labels in the dataset. Accuracy for each question was calculated individually, and an overall average accuracy was computed.

F. Validation and Visualization

To validate the clustering results, the test data was plotted on the training data to check whether the question-specific test data clustered near its respective question cluster. This step was crucial for ensuring that the clustering was meaningful and that the embeddings accurately represented the underlying data structures. Similarly, the clustering based on correctness was validated to verify that incorrect, partially correct, and correct answers were appropriately grouped.

G. Performance Metrics

The overall performance of the methodology was assessed by computing accuracy metrics for each question and averaging the results across all questions. This evaluation was conducted for both LaTeX and mathematical representations in both training and testing phases. The objective was to ensure that the methodology provided reliable predictions across various scenarios. In addition to traditional accuracy metrics, a more comprehensive evaluation was performed to assess the clustering accuracy based on question and correctness. This involved computing the Adjusted Rand Index (ARI) for question-based clustering and devising a novel method for correctness-based clustering accuracy.

For correctness-based clustering accuracy, the following procedure was followed:

- **Centroid Calculation:** Centroids were computed for each cluster based on correctness labels (incorrect, partially correct, and correct).
- **Distance Calculation:** The distances from each point in the cluster to its centroid were calculated.
- **Nearest Points Selection:** The n nearest points to each centroid were selected, where n is the total number of points in that cluster.
- **Class Label Comparison:** For each selected point, the correctness label was compared with the correctness label of the centroid. If the correctness labels did not match, the point was considered misclassified.
- **Accuracy Calculation:** The accuracy of each cluster was calculated as the proportion of points correctly classified relative to the total number of points in that cluster.
- **Overall Accuracy:** The accuracy of each cluster was then summed, and the sum was divided by the total number of clusters to obtain the overall correctness-based clustering accuracy.

With the use of cosine similarity and well-defined thresholds, the project accurately predicted the correctness of answers, offering valuable insights into how closely student answers matched the key solutions. The methodology's robustness was demonstrated by its consistent performance across both LaTeX and mathematical representations, ensuring reliable and insightful results for educational assessments.

IV. Result and Discussion

The evaluation of correctness prediction, answer-based clustering, question prediction, and question-based clustering was conducted on training and testing datasets for latex and mathematical solutions. Table III illustrates the accuracy results for correctness prediction using cosine similarity and machine learning-based clustering.

A. Research Question 1 (RQ1):

Notably, cosine similarity achieved accuracies of 65.8% and 72.2% on the training datasets (D1 and D3) for latex and mathematical solutions, respectively. However, on the testing datasets (D2 and D4), the accuracies dropped to 35% and 71.1%, indicating potential overfitting in LaTeX representation of answers (Table III).

TABLE III: Accuracy Results for All Approaches

Methods	**Accuracy (%)**			
	D1	**D2**	**D3**	**D4**
Correctness Prediction using Cosine Similarity	65.8	35.0	72.2	71.1
Correctness-based clustering using Machine learning	86.8	35.0	87.2	38.3
Question Prediction using Cosine Similarity	70.6	88.4	90.4	92.3
Question-based clustering using Machine learning	87.03	94.4	92.2	85.5

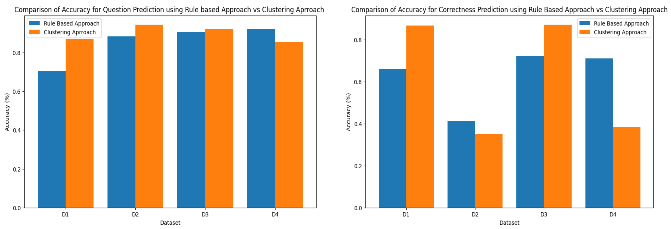
Similarly, for correctness-based clustering using machine learning, high accuracies were achieved on training datasets. From the results, it was observed that the accuracies on the training datasets D1 and D3 were 86. 8% and 87. 2% respectively while the accuracies on the testing datasets D2 and D4 were 35. 0% and 38. 3% respectively showing that there was over fitting. These results postulate that the correct form of robustness is necessary in order to achieve good performance on new data. Especially, the drastic reduction in the accuracy measured by correctness-based clustering means that overfitting is a primary problem on clustering tasks.

In contrast, the 3rd and 4th rows of Table III represents the accuracy results for question prediction and question-based clustering respectively. Both cosine similarity and machine learning-based clustering demonstrated higher accuracies on the testing datasets compared to the training datasets, indicating better generalization. Notably, machine learning-based clustering outperformed cosine similarity in question-based clustering across all datasets, showcasing its ability to generalize well to unseen data.

The Fig. 2 compare rule-based and clustering approaches for question and correctness prediction across four datasets (D1-D4). Clustering approach outperforms rule-based methods in question prediction. Whereas, for correctness prediction, the clustering approach shows higher accuracy in D1 and D3 (training datasets), but for testing datasets(D2 and D4) the rule-based approach performs better than clustering approach.

Research Question 2 (RQ2):

The Fig. 3 compare LaTeX representation and mathematical representation for correctness and question-based prediction using rule-based and clustering approaches. In correctness prediction, mathematical representation outperforms LaTeX in both rule-based and clustering approaches. For question

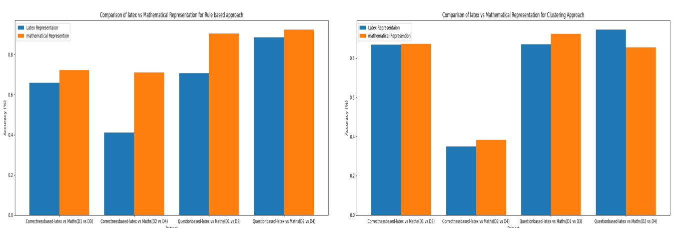


(a) Question Prediction

(b) Correctness Prediction

Fig. 2: Comparison of accuracies of Rule based Approach vs Clustering Aprroach for 2a and 2b

prediction, LaTeX representation shows higher accuracy using clustering approach, but fails to perform with the rule-based approach. Overall, we can say that mathematical representation outperforms latex representation when using the rule-based approach. For the clustering approach, both representations show almost similar results. However, for datasets such as D1, D2, and D3, the mathematical representation is slightly better than the LaTeX representation, whereas for D4, the LaTeX representation is better than mathematical representation.



(a) Rule-based Approach

(b) Clustering Approach

Fig. 3: Comparison of accuracies of latex Representation vs Mathematical Representation using 3a and 3b

Hence, our dataset contains 6 algebraic questions, for each question there are 21 answers, which results in a total of 126 answers, and those 126 answers are considered as training dataset. In response to that, there are only 26 different answers available in the testing dataset. This situation shows an imbalance between the number of questions and the variety of answers available, which had contributed to overfitting. Also the number of answers are very less with respective to the number of questions present. During training, instead of all the possible variations of the answers, a limited set of discrete answers is provided to the model for training, thus it may not handle new answers so well. This may lead to over fitting of the model as it tends to learn specific answers and not every possible way of the answers.

V. Challenges

- Converting the manual answer scripts into MathBERT embeddings.
- Extracting the exact algebraic equations of answers and filling them in excel.

VI. Conclusion

In addressing the challenges of manual evaluation in education, the integration of automated assessment systems like MathBERT-based models and machine learning-based clustering offers promising solutions. The resource-intensive nature of manual grading, coupled with the complexities of evaluating descriptive responses, necessitates more efficient approaches. Using MathBERT's language skills, automated systems make grading easier for teachers and able to handle more work without getting overwhelmed. However, the efficiency of these automated systems must be rigorously evaluated. The comparison of correctness prediction, answer-based clustering, question prediction, and question-based clustering using cosine similarity and machine learning techniques highlights both the potential and limitations of these methods. While cosine similarity shows promise, particularly in correctness prediction, issues such as overfitting become apparent when applied to testing datasets. In contrast, machine learning-based clustering demonstrates better generalization, especially in question-related tasks. Overall, the combination of automated assessment systems with advanced techniques like machine learning offers a pathway to more efficient and accurate educational evaluation. By embracing these technologies, educational institutions can optimize resources, improve assessment accuracy, and ultimately enhance educational outcomes for students.

References

[1] Balakrishnan RM, Pati PB, Singh RP, Santhanalakshmi S, Kumar P. "Fine-Tuned T5 For Auto-Grading Of Quadratic Equation Problems". Procedia Computer Science. 2024 Jan 1;235:2178-86.

[2] S. Dodia, V. Spoorthy. and T. Chandak, "Machine Learning-based Automated System for Subjective Answer Evaluation", 2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2023.

[3] G. Sanuvala and S. S. Fatima, "A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques", 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 1049-1054.

[4] Judi Humberstone, Robert A. Reeve, "Profiles of algebraic competence", Volume 18, Issue 4, August 2008, Pages 354-367.

[5] M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani and M. Usman, "AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions", in IEEE Access, vol. 10, pp. 81323-81342, 2022.

[6] Hongjie Jia, Shifei Ding, Xinzheng Xu & Ru Nie, "The latest research progress on spectral clustering", Volume 24, pages 1477–1486, (2014).

[7] Jan Philip Bernius, Stephan Krusche, Bernd Bruegge, "Machine learning based feedback on textual student answers in large courses", Volume 3, 2022.

[8] Ahuja, R., Chug, A., Gupta, S., Ahuja, P., Kohli, S., "Classification and Clustering Algorithms of Machine Learning with their Applications". In: Yang, XS., He, XS. (eds) Nature-Inspired Computation in Data Mining and Machine Learning. Studies in Computational Intelligence, vol 855.

[9] Mingfang Wu, Michael Fuller, Ross Wilkinson, "Using clustering and classification approaches in interactive retrieval", Information Processing & Management.

[10] Jing, L., Ng, M.K. & Huang, J.Z. "Knowledge-based vector space model for text clustering". Knowl Inf Syst 25, 35–55 (2010).

[11] Anisha Mariam Thomas, M.G. Resmipriya, "An Efficient Text Classification Scheme Using Clustering", Procedia Technology, Volume 24, 2016.

[12] Tilman Lange, Volker Roth, Mikio L. Braun, Joachim M. Buhmann, "Stability-Based Validation of Clustering Solutions. Neural Comput", 2004.

[13] C. Fraley, A. E. Raftery, "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis", The Computer Journal, Volume 41, Issue 8, 1998, Pages 578–588.

[14] Atsushi Shirafuji, Yutaka Watanobe, "Exploring the Robustness of Large Language Models for Solving Programming Problems", 26 Jun 2023.

[15] SMR Abidi, Hussain M, Xu Y, Zhang W. "Prediction of Confusion Attempting Algebra Homework in an Intelligent Tutoring System through Machine Learning Techniques for Educational Sustainable Development"Sustainability, 2019.

[16] R. Krithika and Jayasree Narayanan, "Learning to Grade Short Answers Using Machine Learning Techniques", in Proceedings of the Third International Symposium on Women in Computing and Informatics, New York, NY, USA, 2015.

[17] Ani R., Krishna S., H, A., and U, A., "An Approach Towards Building an IoT Based Smart Classroom", in International Conference on Advances in Computing, Communications and Informatics (ICACCI) Cite this publication, 2018.

[18] S. Narayanan, V. S. Kommuri, S. N. Subramanian and K. Bijlani, "Question bank calibration using unsupervised learning of assessment performance metrics", 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017.

[19] A. Divya, V. Haridas and J. Narayanan, "Automation of Short Answer Grading Techniques: Comparative Study using Deep Learning Techniques", 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023.

[20] N. D. Anand, S. L. K. Das and R. V R, "Analysis and Early Diagnosing Tool for Learning Disability using Machine Learning Models", 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023.