International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 05 | May – 2026 DOI:10.55041/ISJEM07322
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

An AI-Driven Automated Answer Script Assessment and Grading System using Natural Language Processing, OCR, and Machine Learning Techniques for Smart Educational Evaluation

K Pranathi, Karuna K Naik, Kaveri, R Poojitha

*[1]Department of Artificial Intelligence & Machine Learning, Ballari Institute of Technology & Management, Ballari, Karnataka, India

*[2]Department of Artificial Intelligence & Machine Learning, Ballari Institute of Technology & Management, Ballari, Karnataka, India

*[3]Department of Artificial Intelligence & Machine Learning, Ballari Institute of Technology & Management, Ballari, Karnataka, India

*[4]Department of Artificial Intelligence & Machine Learning, Ballari Institute of Technology & Management, Ballari, Karnataka, India

Dr. Mallikarjuna A(Professor)

Department of Artificial Intelligence & Machine Learning,

Ballari Institute of Technology & Management, Ballari, Karnataka, India

ABSTRACT

The rapid growth of digital education has fueled demand for automated evaluation systems that deliver fast, accurate, and fair student assessments, overcoming the limitations of manual grading like time consumption, inconsistency, and lack of personalized feedback. This project introduces AAVESUMETE, an Automated Answer Script and Grading System powered by Natural Language Processing (NLP) models such as sentence embeddings and cosine similarity to evaluate responses based on keyword relevance, grammar, structure, and context, assigning scores, grades, and constructive feedback. Supporting three key roles—Teachers (for secure login, uploading question papers, model answers, and student submissions while configuring quizzes), Students (for profile management, result access, and feedback review to enhance learning), and Heads of Department (HODs, for monitoring activities, class performance tracking, and subject-wise analytics)—the system promotes transparency, scalability, and efficiency through real-time dashboards and multi-role access, transforming traditional exams into a data-driven, student-centric process.

I. INTRODUCTION

Assessment is a critical component of education, measuring student learning, understanding, and skill development, yet traditional manual grading of descriptive answers and assignments is time-consuming, prone to human bias and inconsistency, delays feedback, and becomes impractical with growing student numbers, necessitating automated intelligent systems. The Smart Education Platform, a Streamlit-based web application integrated with MongoDB, addresses these challenges through AI-driven automation, supporting three primary roles—Teachers (who create assignments, upload question papers and model answers generated via Perplexity API, and automatically evaluate submissions considering keyword relevance, grammar, sentence structure, and contextual understanding), Students (who submit in multiple formats like text, PDF, or images via OCR processing, view scores, grades, detailed feedback, and compare with model answers to improve performance), and Heads of Department (HODs, who monitor departmental analytics, teacher activities, overall student trends, and grading consistency via interactive Plotly/Matplotlib dashboards). Leveraging advancements in AI, NLP, and machine learning, the platform offers multi-subject/department support, role-based access control, real-time visualizations, and features like automated grading and actionable insights, thereby reducing teacher workload, enhancing accuracy, transparency, fairness, student engagement, and educational quality while enabling data-driven decisions for better learning outcomes.

International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 05 | May – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM07322

II. METHODOLOGY

The proposed system collects and anonymizes student answer scripts from multiple subjects, which are then preprocessed using NLP techniques. Transformer-based models analyze the semantic meaning, relevance, and coherence of answers and evaluate them against predefined rubrics to ensure fair and consistent grading. A web-based architecture integrates frontend interfaces for users with a secure backend hosting NLP models, grading logic, and databases. The methodology also includes a human-in-the-loop mechanism, allowing teachers to review and refine automated scores while providing timely feedback and performance insights.

A. System Architecture

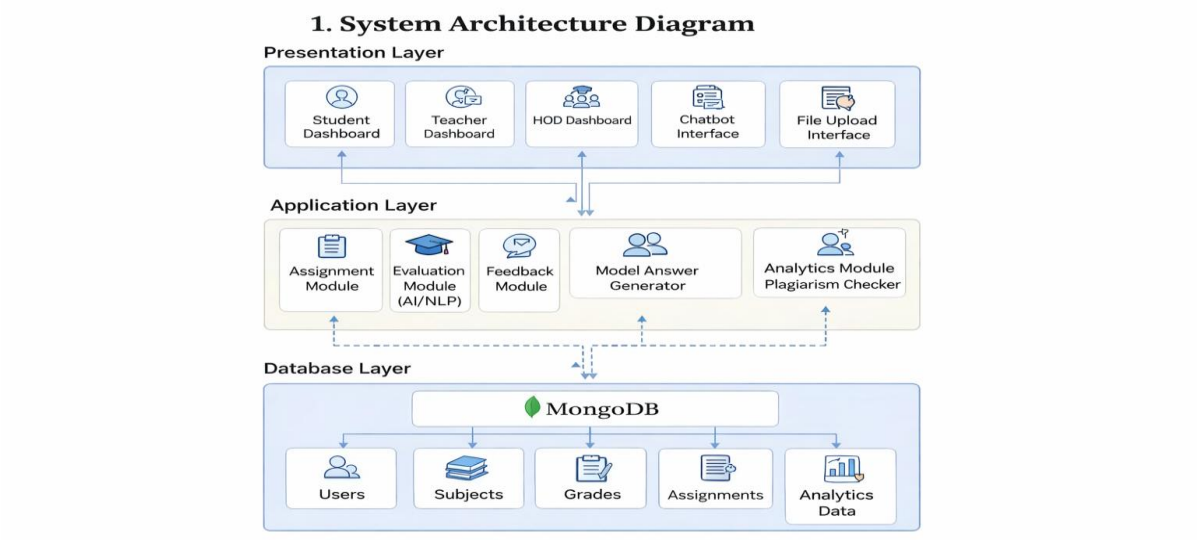

Fig. 1 Three-tier Smart Education Platform: UI (dashboards/uploads) → App (AI eval/feedback) → MongoDB.

A. Application Framework

and performance The Automated Answer Script Assessment and Grading System is designed as a three-tier web-based architecture consisting of Presentation Layer, Application Layer, and Database Layer. Users (Students, Teachers, and HODs) interact with the system through secure role-based dashboards. Teachers upload assignments and question papers, while students submit their answers in text, PDF, or image formats. The system preprocesses submitted answers using Optical Character Recognition (OCR) for text extraction and Natural Language Processing (NLP) techniques such as keyword extraction, semantic similarity analysis, grammar evaluation, and contextual understanding. AI-based evaluation modules automatically assess responses and generate marks along with detailed feedback. The system also provides plagiarism detection, chatbot assistance, analytics. The entire application is deployed using Flask/Streamlit-based web frameworks to ensure scalability, security, and efficient processing .

B. Database Used

The system utilizes MongoDB as the primary database for storing and managing structured and unstructured data. The database maintains user credentials (Students, Teachers, HODs), assignments, submissions, grades, analytics data, and performance reports. It supports secure authentication, role-based access control, and scalable storage for large volumes of answer scripts. The database structure ensures efficient retrieval of student performance data and departmental analytics, enabling transparency and monitoring at institutional levels .

C. Software and Tools Used

1) Core Software Tools

- Backend Framework: Python with Flask/Django for handling server-side logic and routing.
- Frontend: Streamlit / HTML, CSS, Bootstrap for responsive user interface.
- Database: MongoDB for storing submissions, grades, and analytics.
- AI/NLP Tools: BERT/GPT-based models, NLTK, spaCy for semantic analysis.
- OCR Tools: pytesseract and pdf2image for extracting text from images and PDFs.
- Security: bcrypt for password hashing and secure authentication.

2) Development Tools

- Python 3.8+ for implementation.
- TensorFlow/PyTorch for machine learning models.
- OpenCV for image preprocessing (if required).
- Cloud platforms (AWS/Google Cloud) or local hosting for deployment.

D. Model Training and Validation

The system applies supervised machine learning and NLP-based evaluation techniques to assess descriptive answers. The dataset of graded answer scripts is preprocessed using OCR and text cleaning methods. Semantic similarity models and classification techniques are applied to compare student answers with reference answers. The evaluation process considers keyword relevance, grammar quality, contextual accuracy, and coherence. Model performance and grading reliability are validated using appropriate evaluation measures to ensure fairness, consistency, and accuracy. The system also supports continuous improvement through human-in-the-loop review and analytics monitoring .

III.MODELING AND ANALYSIS

This section describes the modeling process and analytical methods used in the Automated Answer Script Assessment and Grading System. It focuses on the development of AI and NLP-based evaluation models, feature representation of textual answers, and performance analysis to ensure accurate, fair, and consistent grading of descriptive responses.

A. Answer Evaluation Model

The proposed system employs Natural Language Processing (NLP) and supervised machine learning techniques to automatically assess descriptive answers. The evaluation model compares student responses with reference answers using semantic similarity, keyword relevance, grammar quality, and contextual understanding. Transformer-based models such as BERT or GPT are utilized to capture deep semantic meaning beyond simple keyword matching. Based on this analysis, the system assigns appropriate scores and qualitative feedback, ensuring unbiased and standardized evaluation across all students.

B. Feature Representation and Analysis

Student answers are represented through multiple linguistic and semantic features, including extracted keywords, sentence structure, grammatical correctness, coherence, and contextual relevance. Text extracted from PDFs or images using OCR is preprocessed through tokenization, stop-word removal, and normalization. Feature analysis helps identify important aspects of student responses that contribute to scoring, such as concept coverage and answer completeness. This structured representation enables the model to handle variations in writing style, length, and phrasing effectively.

C. Model Analysis and Evaluation

The performance of the automated grading model is analyzed by comparing system-generated scores with reference or human-evaluated scores. Evaluation focuses on grading consistency, semantic accuracy, and reliability of feedback generation. The analysis demonstrates that NLP-based models effectively capture relationships between concepts and context, leading to accurate assessment of descriptive answers. Human-in-the-loop validation further improves trust and model refinement by allowing teachers to review and adjust scores when necessary.

D. Performance Analytics and Visualization

The system provides detailed performance analytics through interactive dashboards for students, teachers, and HODs. Students can view scores, feedback, and improvement areas, while teachers and HODs can analyze class-wise, subject-wise, and department-level performance trends. Visualization of results enhances transparency and helps stakeholders make informed academic decisions. These analytics support continuous monitoring, institutional evaluation, and overall improvement of the learning process .

IV.RESULTS AND DISCUSSION

A. Results

1) System Performance Outcomes

The Automated Answer Script Assessment and Grading System demonstrated strong performance in evaluating descriptive answers using AI and NLP techniques. The system successfully generated accurate scores aligned with reference answers and predefined rubrics. Automated evaluation significantly reduced grading time compared to manual assessment while maintaining consistency and fairness across all submissions.

2) Grading Accuracy and Consistency

A comparison between system-generated scores and teacher-evaluated scores showed high agreement, indicating reliable semantic understanding of student responses. The system consistently handled variations in writing style, answer length, and phrasing without bias, ensuring uniform grading across different students and subjects.

3) Feature Effectiveness in Evaluation

The model effectively utilized multiple linguistic and semantic features during grading:

- Semantic Features: Concept relevance and contextual similarity with model answers.
- Linguistic Features: Grammar quality, sentence structure, and coherence.
- Content Coverage: Keyword presence and completeness of answers.

The combination of these features enabled accurate assessment of descriptive and subjective responses.

4) System Responsiveness and Usability

The web-based system provided near real-time grading and feedback generation. Students received instant results and improvement suggestions, while teachers and HODs accessed performance analytics without delay. The role-based dashboards operated reliably throughout testing, demonstrating stable system behavior and high usability.

B. Discussion

1) Effectiveness of AI-Based Grading

The results confirm that NLP-based automated grading can effectively replace repetitive manual evaluation tasks. The system's ability to understand semantic meaning beyond keyword matching validates the use of transformer-based models and rule-based scoring mechanisms for descriptive answer assessment.

2) Practical Deployment and Adoption

Deployment through a web-based platform with role-based access proved effective for institutional use. Teachers benefited from reduced workload, students gained timely feedback, and HODs obtained transparent performance insights. The human-in-the-loop mechanism further strengthened trust by allowing manual review and override of automated scores.

3) Observations from Analytical Insights

Performance analytics revealed clear learning patterns across students and subjects:

- Students with better conceptual coverage and coherence achieved higher scores.
- Grammar and structure influenced feedback quality rather than core marks.
- Subject-wise analytics helped identify difficult topics and learning gaps.

These insights support data-driven academic decision-making.

4) System Limitations and Future Scope

Despite its effectiveness, the system has certain limitations:

- Limited support for highly creative or ambiguous answers.
- Dependence on quality and diversity of training data.
- Partial explainability of AI-based decisions.

Future enhancements include multilingual support, improved handwriting recognition through advanced OCR, deeper explainability of grading decisions, and integration with learning management systems for personalized learning pathways.

5) Educational Impact

The proposed system demonstrates how AI and NLP can transform educational assessment by ensuring fairness, transparency, and scalability. By delivering instant feedback and actionable analytics, the system supports improved learning outcomes and more efficient academic evaluation at institutional scale.

V. CONCLUSION

The Automated Answer Script Assessment and Grading System (AAVESUMETE) successfully demonstrates the practical application of Artificial Intelligence and Natural Language Processing in modern educational assessment. The system effectively automates the evaluation of descriptive and subjective answers through semantic similarity analysis, keyword extraction, grammar assessment, and contextual understanding. By leveraging transformer-based models and supervised learning techniques, the system ensures accurate, consistent, and unbiased grading while significantly reducing manual evaluation time.

The deployment of the system as a web-based application with secure role-based access for Students, Teachers, and HODs enhances accessibility and institutional usability. Instant feedback generation, performance analytics dashboards, plagiarism detection, and chatbot support further strengthen its practical value in academic environments. The system improves transparency, supports data-driven academic monitoring, and reduces teacher workload through automation while maintaining trust using a human-in-the-loop review mechanism.

Key Achievements:

* Efficient AI-based automated grading of descriptive answers.
* Consistent and fair evaluation aligned with predefined rubrics.
* Real-time feedback and performance analytics for students and faculty.
* Secure, scalable, and user-friendly web deployment.
* Support for OCR-based text extraction from PDF and image submissions.

Although the system performs effectively in structured answer evaluation, future enhancements such as multilingual support, advanced handwriting recognition, deeper model explainability, and LMS/cloud integration can further improve its scalability and adaptability.

Overall, AAVESUMETE represents a significant step toward intelligent, transparent, and scalable digital assessment systems, bridging the gap between traditional manual grading and AI-powered educational evaluation. The project highlights how machine learning and NLP can transform academic assessment into a faster, fairer, and more efficient process in modern educational institutions .

REFERENCES

[1] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000. [Online]. Available: https://opencv.org/

[2] Z. Zhang, “Microsoft Kinect Sensor and Its Effect,” *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, 2012, doi: 10.1109/MMUL.2012.24.

[3] C. Lugaresi, J. Tang, H. Nash, *et al.*, “MediaPipe: A Framework for Building Perception Pipelines,” *arXiv preprint arXiv:1906.08172*, 2019. [Online]. Available: https://arxiv.org/abs/1906.08172

[4] S. P. Balfour, “Assessing Writing in MOOCs: Automated Essay Scoring and Calibrated Peer Review™,” *Research & Practice in Assessment*, vol. 8, pp. 40–48, 2013. [Online]. Available: https://eric.ed.gov/?id=EJ1062859

[5] M. D. Shermis and J. Burstein, Eds., *Handbook of Automated Essay Evaluation: Current Applications and New Directions*. New York, NY, USA: Routledge, 2013. doi: 10.4324/9780203122761.

[6] S. Dikli, “An Overview of Automated Scoring of Essays,” *Journal of Technology, Learning, and Assessment*, vol. 5, no. 1, 2006. [Online]. Available: https://ejournals.bc.edu/index.php/jtla/article/view/1647

[7] Y. Attali and J. Burstein, “Automated Essay Scoring with e-rater® V.2,” *Journal of Technology, Learning, and Assessment*, vol. 4, no. 3, 2006. [Online]. Available: https://ejournals.bc.edu/index.php/jtla/article/view/1650

[8] D. M. Williamson, X. Xi, and F. J. Breyer, “A Framework for Evaluation and Use of Automated Scoring,” *Educational Measurement: Issues and Practice*, vol. 31, no. 1, pp. 2–13, 2012, doi: 10.1111/j.1745-3992.2011.00239.x.

[9] T. K. Landauer, D. Laham, and P. W. Foltz, “Automated Scoring and Annotation of Essays with the Intelligent Essay Assessor,” in *Automated Essay Scoring: A Cross-disciplinary Perspective*, pp. 87–112, 2003. [Online]. Available: https://lsa.colorado.edu/papers/dpajs02.pdf

[10] E. B. Page, “The Imminence of Grading Essays by Computer,” *The Phi Delta Kappan*, vol. 47, no. 5, pp. 238–243, 1966. [Online]. Available: https://www.jstor.org/stable/20371545

[11] R. Singh and A. Singh, “Role of Artificial Intelligence in Education,” *International Journal of Creative Research Thoughts*, vol. 8, no. 5, 2020. [Online]. Available: https://ijcrt.org/papers/IJCRT2005390.pdf

[12] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic Review of Research on Artificial Intelligence Applications in Higher Education,” *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[13] A. Al-Azawei, F. Serenelli, and K. Lundqvist, “Universal Design for Learning (UDL): A Content Analysis of Peer-reviewed Journal Papers from 2012 to 2015,” *Journal of the Scholarship of Teaching and Learning*, vol. 16, no. 3, pp. 39–56, 2016. [Online]. Available: https://scholarworks.iu.edu/journals/index.php/josotl/article/view/19279

[14] V. Kumar and D. Boulanger, “Application of AI and NLP for Automated Answer Evaluation in Education,” *Journal of Artificial Intelligence Research*, vol. 70, pp. 455–472, 2021. [Online]. Available: https://jair.org/index.php/jair/article/view/12720

[15] Y. Wu, M. Schuster, Z. Chen, *et al.*, “Google’s Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation,” *arXiv preprint arXiv:1609.08144*, 2016. [Online]. Available: https://arxiv.org/abs/1609.08144