Smart Paper Assessment System

Dr. S. B. Chaudhari
Dept. Computer Engineering
JSPM's JSCOE Pune, India
shankarchaudhari@jspmjscoe.edu.in

Asmita Mote
Dept. Computer Engineering
JSPM's JSCOE Pune, India
asmitamote9411@gmail.com

Sonal Sawant
Dept. Computer Engineering
JSPM's JSCOE Pune, India
sawantsonal308@gmail.com

Kunal Shinde
Dept. Computer Engineering
JSPM's JSCOE Pune, India
kdshinde2003@gmail.com

Abstract –

The increasing demand for rapid and accurate assessment in large-scale examinations has driven the development of automated grading systems. This paper introduces a Smart Paper Assessment System that leverages Optical Character Recognition (OCR), Natural Language Processing (NLP), and Large Language Models (LLMs) to convert physical exam papers into digital formats, analyze student responses, and autonomously assign scores. The system operates by first applying OCR to digitize handwritten or printed text on exam papers, enabling the identification and organization of student answers. NLP techniques then align these responses with their corresponding questions, while LLMs assess the content by comparing it to a predefined answer key. Using keyword matching, sentence structure, and semantic analysis, the model assigns marks based on the relevance and accuracy of each response. This approach accelerates grading, minimizes human errors and biases, and ensures consistent and fair evaluations. The scalability of the system allows it to manage exams of varying sizes, making it suitable for educational institutions ranging from schools to universities. The proposed Smart Paper Assessment System streamlines the grading process, enhances precision, and provides timely feedback, contributing to a more efficient educational assessment framework.

keywords- Optical Character Recognition (OCR), Natural Language Processing (NLP), Large Language Models (LLM), Automated Grading, Paper-Based Assessment, Keyword Matching, Machine Learning in Education, Text Recognition.

I. INTRODUCTION

A. Background on Smart Paper Assessment Systems

The demand for efficient and accurate grading of large-scale examinations has driven the development of automated grading solutions. Traditional manual grading is time-consuming and prone to human error, leading to delays and potential biases in the evaluation process. A Smart Paper Assessment System represents a shift from conventional grading methods, utilizing advanced technologies like Optical Character Recognition (OCR), Natural Language Processing (NLP), and Large Language Models (LLMs) to automate the assessment of student performance. By digitizing and analyzing handwritten or printed exam papers, this system can process answers rapidly, matching them against pre-defined answer keys to ensure fairness and consistency. The adoption of such automated systems introduces unique challenges, including maintaining grading accuracy and handling variations in handwriting and question structure.

B. Role of Advanced Technologies

OCR, NLP, and LLMs are pivotal in addressing the challenges of automating exam assessments. OCR converts physical text from exam sheets into digital data, making it machine-readable, while NLP structures and aligns the extracted text to their respective questions. LLMs then evaluate the responses by analyzing content, structure, and semantic meaning, assigning scores based on predefined criteria. Each technology must balance speed, accuracy, and computational efficiency to

facilitate real-time processing. Ensuring scalability and the ability to handle diverse question types, including descriptive and open-ended responses, is critical for the widespread application of these systems in educational settings.

II. LITERATURE SURVEY

Gupta, A., & Malhotra, S. [1] proposed Automated essay scoring using machine learning and natural language processing. This paper explores a machine learning-based framework for automated essay scoring, a system designed to assess written responses in a standardized, efficient, and fair manner. By leveraging natural language processing (NLP) and machine learning algorithms, the framework analyzes essays based on criteria like content quality, coherence, grammar, and relevance. It integrates NLP techniques such as word embedding and sentiment analysis to derive meaningful patterns from text data, which are then processed by machine learning algorithms to produce scores aligned with human grading standards

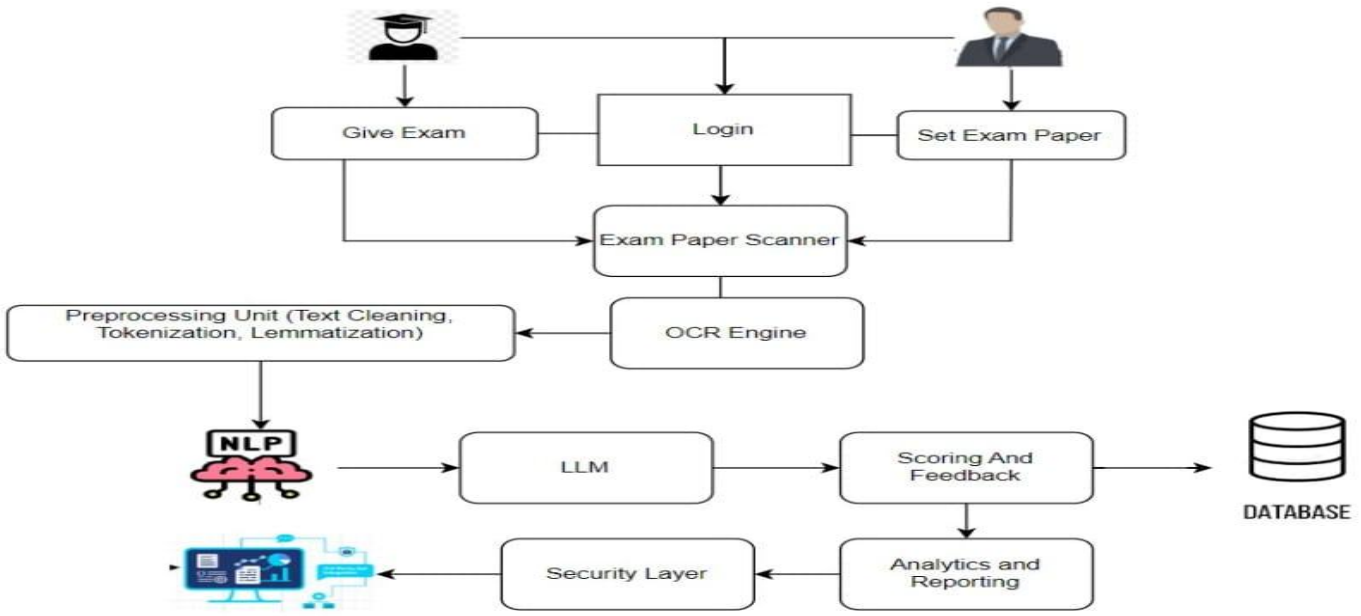
Patel, R., & Bhushan, P.[6] presented Enhanced Grading System for Subjective Responses Using NLP and OCR Technologies (2022) In this paper, the authors present a grading system designed to evaluate handwritten subjective answers by integrating Optical Character Recognition (OCR) and Natural Language Processing (NLP) technologies. The system extracts text from scanned answer sheets and applies NLP techniques to assess the answers. It combines semantic analysis with keyword matching to ensure that the responses align with expected content, grammar, and coherence. The system's main advantage is its ability to handle diverse handwriting styles, making it applicable across educational levels.

Hu and Xia [4] proposed a Latent Semantic Indexing approach for the assessment of subjective questions online. The answers were presented in TF-IDF embedding matrices, and then Singular Value Decomposition (SVD) was applied to the term-document matrix, which formed a semantic space of vectors. LSI played the role of reducing problems with synonym and polysemy. At last, the Similarity between answers was calculated using cosine similarity. Dataset consisted of 35classes and 850 instances marked by teachers, and the results showed a 5%difference in grading done by teacher and the proposed system.

Rahman, T., & Li, M.[8] (2022). Automated feedback generation for open-ended student responses using NLP. *Journal of Educational Data Mining, 14(2), 45-60.* The authors of this paper investigate a system designed to automatically generate personalized feedback for students based on their responses to open-ended questions. Using Natural Language Processing (NLP), the system evaluates the content, grammar, and structure of answers, providing constructive feedback that helps students improve. The system has been shown to enhance learning outcomes by offering immediate, tailored feedback, which is especially beneficial in large-scale education settings.

III. METHODOLOGY

The methodology for developing the automated exam evaluation system begins with defining exam types and performance metrics to meet educational standards. Next, exam samples are collected, digitized using OCR, and preprocessed for consistency. A large language model (LLM) is then fine-tuned with annotated data to understand context, grade responses, and generate feedback. The model is validated on test sets to ensure accuracy, and the system is integrated into a secure, user-friendly platform for deployment. Finally, pilot testing and iterative refinement based on user feedback enhance the system's reliability and usability in real-world settings.



A. Architecture Model

Fig. 3.1 Architecture Diagram

The proposed system for automated exam evaluation integrates several key components to streamline the process of assessing student submissions. Initially, the **Exam Paper Scanner** captures digital images of exam papers, which are then processed by an **Optical Character Recognition (OCR) Engine**. The OCR engine converts these images into editable text, enabling further analysis. The extracted text undergoes preprocessing, where it is cleaned, tokenized, and lemmatized to create a standardized input for analysis. Following this, a **Large Language Model (LLM)**, equipped with advanced natural language processing (NLP) capabilities, interprets and evaluates the textual content. A dedicated **Security Layer** ensures data protection, maintaining confidentiality and integrity throughout the analysis.

Once processed, the system generates scores and targeted feedback through an automated **Scoring Module**. This module assesses each response based on predefined criteria, allowing for consistent and objective grading. Additionally, a **Feedback System** provides students with constructive feedback, highlighting areas of strength and identifying potential areas for improvement. To support administrative needs, an **Analytics and Reporting** module generates detailed performance reports, enabling educators to analyze trends and track student progress. All results, feedback, and reports are stored in a secure **Database** for easy access and long-term retention. The system also includes a user interface that allows students to log in and take exams, while teachers can log in to set exams, review results, and access performance data. This comprehensive architecture, combining OCR and NLP technologies with secure data management, offers an efficient and scalable solution for automated exam assessment.

IV. RESULTS

A. **OCR Accuracy**: The OCR system's ability to accurately extract text is essential, but it can struggle with messy handwriting or unclear text, which may lead to errors in grading. Comparing OCR results with manual extraction highlights where OCR improves efficiency but also where it may need further tuning.

B. **LLM Grading Consistency**: The LLM's grading accuracy was compared to human

grading across different question types. While it performed well on straightforward answers, some discrepancies appeared in subjective answers, suggesting areas where the model could benefit from adjustments.

C. **Handling Subjective Answers**: The system flagged certain subjective responses for manual review, allowing it to focus human graders only on the complex cases. This hybrid approach is effective but highlights that subjective answers still need human oversight to ensure fair grading.

D. **Generation of Scorecards for Students and Teachers**: The system automatically generates detailed scorecards, providing students with scores and personalized feedback based on topic coverage, grammar, and overall performance. Teachers receive a summarized report with scores, highlighting areas where students performed well and where improvement is needed, thus aiding both students' learning and teachers' assessment efforts.

- **Score prediction using cosine similarity with model suggestion.**
  The following table shows difference in errors resulting from the machine learning model correction. It shows that the models accuracy decreased by 1.54% when using cosine similarity along with classification models. Cosine similarity paired with a machine learning model yields 86% accuracy for this short dataset. Following Figure shows the comparison of accuracy obtained from various combinations.

Human Score	Error Without Model	Error With Model
46	13	0.5
46	13	0.5
60	18	30.5
60	18	30.5
55	9	3.5
55	24	11.5
27	19	6.5
0	13	25.5
77	27	14.5
27	1	13.5

Fig. Score prediction using cosine similarity with model suggestion.

V. CONCLUSION

The Smart Paper Assessment System, leveraging OCR, NLP, and LLMs, represents a significant advancement in automating and optimizing the grading process for written exams and assignments. By converting handwritten or printed responses into a structured digital format and analyzing them with advanced NLP techniques, this system enhances the accuracy, efficiency, and consistency of grading, particularly for open-ended and subjective responses. Beyond reducing reliance on manual grading, this approach democratizes access to high-quality, objective assessment, enabling educators to deliver timely and personalized feedback, while freeing up resources to focus on deeper engagement with student learning.

VI. FUTURE WORK

- Incorporation of Multilingual Capabilities: Expanding the system to support multiple languages would facilitate its application in diverse educational settings worldwide, accommodating students and educators from varied linguistic backgrounds and promoting inclusivity.
- Enhanced Oral Assessment Integration: By incorporating speech-to-text conversion capabilities alongside the existing OCR setup, the system could be extended to evaluate spoken responses in oral exams, contributing to a more comprehensive assessment framework adaptable to multiple exam types.
- Adaptive Learning Analytics: Integrating advanced analytics would enable tracking of individual student performance over time, highlighting areas for improvement and recommending tailored learning resources based on each student's strengths and areas of need, thereby supporting personalized learning pathways.
- Increased Accessibility with Mobile Integration:
  Developing a mobile-friendly version of the system would allow both students and instructors to access assessments and

feedback on mobile devices, enhancing the usability and accessibility of the system for on-the-go interaction..

- Real-Time Grading and Feedback: Implementing real-time grading and feedback mechanisms would provide students with immediate insights into their performance, fostering an interactive and engaging learning environment that encourages continuous improvement.

.

VII. REFERENCES

[1] Gupta, A., & Malhotra, S. (2023). Automated essay scoring using machine learning and natural language processing. *International Journal of Artificial Intelligence in Education*, 32(4), 567-589.

[2] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, A survey on the techniques, applications, and performance of short text semantic similarity, Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021.

[3] M. S. M. Patil and M. S. Patil, Evaluating Student descriptive answers using natural language processing, Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 17161718, 2014.

[4] X.HuandH.Xia, Automated assessment system for subjective questions based on LSI, inProc.3rdInt. Symp.Intell. Inf. Technol. Secur. Informat., Apr. 2010, pp. 250254.

[5] P. Patil, S. Patil, V. Miniyar, and A. Bandal, Subjective answer evalua tion using machine learning, Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 113, 2018

[6] Patel, R., & Bhushan, P. (2022). Enhanced grading system for subjective responses using NLP and OCR technologies. *Journal of Educational Technology & Society*, 25(3), 222-235.

[7] S. Afzal, M. Asim, A. R. Javed, M. O. Beg, and T. Baker, URLdeep Detect: A deep learning approach for detecting malicious URLs using semantic vector models, J. Netw. Syst. Manage., vol. 29, no. 3, pp. 127, Mar. 2021.

[8] Rahman, T., & Li, M. (2022). Automated feedback generation for open-ended student responses using NLP. *Journal of Educational Data Mining*, 14(2), 45-60.

[9] L. A. Cutrone and M. Chang, Automarking: Automatic assessment of open questions, in Proc. 10th IEEE Int. Conf. Adv. Learn. Technol., Sousse, Tunisia, Jul. 2010, pp. 143147

[10] G.Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, Atwo-stage text feature selection algorithm for improving text classi cation, Tech. Rep., 2021.