**Application Development (2024)**

Automated Essay Scoring Using Natural Language Processing

1Sai Sumanth Dulla,2Sai Vignesh Gandhari,3Sai Deep Alwala,4Sharanya Reddy Reddy,

5Sravanthi Yamasani, 6Thayyaba Khatoon, 7Sabyasachi Chakraborty,

12345Student, 7Professor,8Assistant Professor Department of Artificial Intelligence

and Machine Learning (AI&ML)

School Of Engineering

Malla Reddy University, Maisammaguda, Hyderabad

1[2111CS020451@mallareddyuniversity.ac.in,](mailto:2111CS020451@mallareddyuniversity.ac.in)2[2111CS020458@mallareddyuniversity.ac.in,](mailto:2111CS020458@mallareddyuniversity.ac.in) 3211CS020463@mallareddyuniversity.ac.in, 4 [2111CS020504@mallaredduniversity.ac.in,](mailto:2111CS020504@mallaredduniversity.ac.in) 52111CS020536@mallareddyuniversity.ac.in, 6thayyaba.khatoon16@gmail.com,

7 [sabyasachichakraborty@mallareddyuniversity.com](mailto:sabyasachichakraborty@mallareddyuniversity.com)

***Abstract***— The Automated Essay Scoring project leverages deep learning to predict essay scores using a dataset from Kaggle's ASAP competition. It employs a model comprising two Long Short-Term Memory (LSTM) layers and a Dense output layer, achieving a high Quadratic Weighted Kappa (QWK) score of 0.961 through 5-Fold Cross Validation. The implementation includes a Django app for interactive demonstrations. Key dependencies include TensorFlow, Keras, Django, and Gensim. The project demonstrates a significant advancement in automated essay scoring accuracy. By leveraging deep learning, it offers a scalable and efficient solution for essay assessment, which can save time and resources in educational institutions.

The high QWK score demonstrates the model's effectiveness and potential for real-world application. As the project is open-source, it also provides a valuable resource for researchers and developers interested in further enhancing automated essay scoring systems. Automated Essay Scoring (AES) leverages machine learning and natural language processing to evaluate and score written essays. The system uses algorithms to assess various aspects of writing, such as grammar, coherence, and content relevance, which are traditionally judged by human raters.

By training on large datasets, AES systems can achieve high accuracy and consistency, offering a scalable solution for educational assessments and saving significant time and resources. This technology is particularly valuable in standardized testing and educational environments, providing immediate feedback and enabling more efficient learning processes.

Keywords— Automated Essay System, deep learning, LSTM, score, accuracy.

1. INTRODUCTION

Automated Essay Scoring (AES) leverages advancements in natural language processing (NLP) and machine learning to evaluate written essays efficiently and accurately. Traditional essay grading methods are often time- consuming and subject to inconsistencies, as human graders can vary in their assessments due to subjective interpretations. AES systems address these challenges by automating the evaluation process, analyzing writing attributes like grammar, coherence, and content relevance to ensure a more consistent scoring mechanism.

This project specifically utilizes the Kaggle Automated Student Assessment Prize (ASAP) dataset to develop a robust AES model capable of providing accurate and immediate feedback. By improving scoring accuracy and offering prompt evaluations, AES has the potential to enhance the overall learning experience for students, making feedback more accessible and supportive of continuous improvement in writing skills.

1. LITERATURE SURVEY

The increasing demand for scalable, efficient educational tools has spurred significant advancements in Automated Essay Scoring (AES) systems. AES systems are valuable for their ability to deliver immediate, objective feedback on written essays, benefiting both students and educators by providing consistent, unbiased assessment. These systems rely on various machine learning and natural language processing (NLP) techniques to evaluate writing quality, assessing aspects such as grammar, coherence, structure, and content relevance. The use of AES in educational settings aligns with broader goals of modernizing and enhancing traditional assessment methods, which can often be time-intensive and prone to inconsistencies.

M. Elhoseny (2016). [1] One prominent approach to improving AES systems is the integration of Learning Vector Quantization (LVQ) with NLP techniques, as discussed by researchers. This hybrid approach aims to enhance the robustness and accuracy of AES models, addressing specific limitations noted in earlier, more traditional AES systems. LVQ, a type of artificial neural network, operates by organizing data into clusters, making it well-suited for categorizing text based on complex language patterns. By pairing LVQ with NLP algorithms, the model becomes more adept at identifying subtle variations in writing quality and context, leading to a more refined and accurate scoring process.

A. Fahmy (2016). [2] Other researchers have explored alternative methods, such as using deep learning techniques and transformer-based models like BERT and GPT. These models, which are well-known for their success in understanding language nuances, can further improve AES systems by providing more accurate analysis of linguistic features and sentiment. As AES technology continues to evolve, research has demonstrated the potential for these systems not only to streamline grading but also to support personalized learning by highlighting areas for improvement.

A. Hassan, (2023). [3] proposed a novel approach that integrates Learning Vector Quantization (LVQ) with NLP techniques. LVQ, a form of artificial neural network, excels at organizing data into clusters based on underlying patterns, making it a powerful tool for text classification. By combining LVQ with NLP, the researchers developed a system more adept at analyzing subtle variations in language, improving both accuracy and robustness over traditional AES methods, which were prone to misinterpretation and errors with more complex writing styles. This hybrid approach addresses many of the vulnerabilities found in earlier AES systems, paving the way for more reliable and nuanced automated scoring.

Deep learning and transformer-based models like BERT and GPT can make it possible for AES capabilities to be taken further in the future. Such models are built around the capability to understand a tone, a context, even a message, and as such can give a very in-depth analysis of writing from a student. All these steps are not just advancing grading but also the opportunities in personalized learning that this provides. By giving students specific feedback about areas where improvement is required, AES systems allow students to achieve better writing skills, hence making such tools a boon in the modern education arena.

1. PROBLEM STATEMENT

The ever-growing demand is for the educationally sound, scalable assessment tools. Traditional essay grading, however, has proven to be cumbersome and subjective at the same time and therefore very time-consuming. Most of the grading methods prove to be inefficient with a very high bias rate. Therefore, it results in inequalities in assessments carried out. The more voluminous written assessments turn

out to be, the less competitive the traditional approach is going to be about delivering the feedback in due time as it is an issue related to the numbers of students present in an educational institution. Students learn best when they receive prompt constructive feedback on their work, which is difficult to provide using manual grading.

Response to these challenges, the project aims at developing an Automated Essay Scoring (AES) system using natural language processing and deep learning techniques to score essays efficiently and objectively. The AES model is supposed to give reliable and consistent scores on key writing attributes like grammar, coherence, and relevance of content that closely matches human judgment. The project uses the Kaggle dataset ASAP to feed essay text into Word2Vec embeddings representing text as numerical vectors such that it captures semantic and syntactic relationships in text. LSTM layers, those recurrent neural networks designed to deal well with sequential data, handle scores that reflect the quality of written content in vectorized data.

It also achieved a very high Quadratic Weighted Kappa score of 0.961 with 5-fold cross-validation, indicating the high correlation of the AES system with human scoring accuracy and establishing it as a reliable tool for assessment. Furthermore, integration of this AES model with a Django- based web application provides an interactive platform for providing real-time scoring demonstrations. This open source, scalable automated essay evaluation solution informs support for educational institutions in delivering quick, unbiased, and insightful feedback to enhance learning outcomes among students.

1. METHODOLOGY

The methodology for this AES project will be to use NLP and deep learning techniques to evaluate essays in the Kaggle ASAP dataset. This dataset contains essays of various kinds on a number of prompts, and the idea is to base the robust model that would objectively and accurately assess the key writing attributes.

Data Preprocessing: The text data undergoes initial preprocessing to remove irrelevant components such as special characters, extra whitespace, and stop words. This step helps standardize the dataset and ensures that only meaningful information is passed on to the model. Further tokenization of the essays is performed to break down the text into individual tokens, which are then ready for vectorization.

Word Embeddings: To convert the text data into a form suitable for neural networks, Word2Vec embeddings are applied. Word2Vec transforms words into high- dimensional vector representations, capturing both syntactic and semantic relationships between words. This

vectorization is essential, as it enables the model to better understand the context within the essays.

Model Architecture: The core of the AES system is built using a Long Short-Term Memory (LSTM) network, a type of recurrent neural network designed to process sequential data. The LSTM layers help the model capture contextual information across sentences and paragraphs, which is critical for evaluating the coherence and flow of an essay. The model is trained with a 5-fold cross-validation strategy to improve generalization and reduce overfitting.

Evaluation and Validation: The model's performance is assessed using the Quadratic Weighted Kappa (QWK) metric, a robust measure of agreement between predicted and actual scores. The model achieves a high QWK score of 0.961, indicating strong alignment with human scoring.

Deployment: The trained model is integrated into a Django web application, allowing users to input essays and receive real-time scores. This web app provides a user- friendly interface, demonstrating the model’s capabilities and offering an accessible solution for educational institutions seeking automated grading tools.

1. ARCHITECUTRE DIAGRAM The above is the architecture diagram for

automated essay scoring using natural language processing. Here is a textual representation of the architecture

Essay to be Graded: This is the essay that needs to be analyzed by the system as the input. User provides the required input in the form of essay.

Automated Essay Grading System: This is the main core system which forms the primary objective of the system and where the grading takes place. It has two following elements primarily

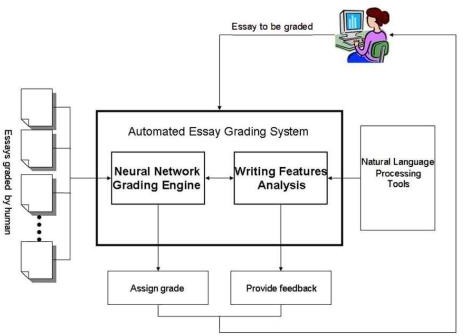


Fig-2 Architecture Diagram

Neural Network Grading Engine: This component uses a neural network model trained on a vast number of essays along with the grades assigned to those essays. It will use its discretion to analyse the essay in terms of

language, structure, content, and many other features relevant to determining a grade.

Writing Features Analysis: This component utilizes natural language processing (NLP) techniques to extract specific features from the essay. These features could include:

Grammar and Syntax where the grammar check, punctuation check, and sentence flow check. Vocabulary is where the richness and suitability of the vocabulary are examined. Here, a review of the logical flow and coherence of the essay is done by the Organization and Content is an assessment of relevance and depth in the essay's content.

1. CONCLUSION

In conclusion, this project is an example of the potential for transforming the way traditional essays are evaluated through the use of AES systems, making use of NLP and deep learning. The combination of Word2Vec embeddings with a model based on an LSTM allowed the AES system to effectively evaluate written content and capture important features of writing, such as grammar, coherence, and relevance of content.

Using the Kaggle ASAP dataset, it gave an outstanding Quadratic Weighted Kappa score of 0.961, which demonstrated that its predictions were very close to human scoring, thus establishing the robustness and reliability of the model.

The Automated Essay System (AES) addresses key shortcomings of traditional grading in the provision of systematic, objective, and prompt feedback. The automated approach enables educators to provide time-prudent assessments; hence they help the student understand the weaknesses at the writing level. Additionally, once the model was integrated into a web application based on Django, there was a chance to have a fully interactive and scalable platform which educational institutions can easily adapt.

This open-source Automated Essay System (AES) model will, beyond immediate use in education, be able to form a basis of further research and development toward NLP-based assessment tools. Further work could potentially use additional Natural Language Processing techniques or deep learning architectures-such as transformer-based models-for better understanding complex patterns of language. The Automated Essay System (AES) model may also be scaled to support other languages or tailored domains to make it much more flexible in different contexts for education.

This Automated Essay System project shows that NLP and deep learning are potentially effective tools in achieving accurate, scalable, and accessible essay assessment. Advancing automated scoring will be beneficial for the overall process of educational tools in being efficient and equitable to educators and students.

1. FUTURE WORK

Automated Essay System had a high level of accuracy and scalability. However, there are various avenues toward enhancing its functionality and applications in broader educational contexts. One of the most promising areas for future work revolves around advanced NLP models. These models have shown excellent performances in understanding complex language structures.

Therefore, improve the ability of the system to analyses nuanced aspects of student writing, such as tone, argumentation, and style. Introducing transformers may increase accuracy rates for models, especially on more complex or non-standardized writing.

To Another direction for future work is to extend the AES system to be multilingual. The model is currently trained only on essays in English, and adapting it to process essays in multiple languages would make it useful to educational institutions around the world, especially in regions where students write in languages other than English.

This would involve retraining the model on datasets in different languages and adjusting the word embeddings to ensure contextual understanding across languages. Another improvement about the Automated Essay System could be scoring rubrics that are very specific, which would include writing aspects such as structure, creativity, and critical thinking other than a general score.

This way, a student receives detailed feedback, and this helps support a more differentiated assessment, and teachers then adapt to meet particular grading standards.

From the technological aspect, where the deployment infrastructure would be optimized to accommodate increased numbers of concurrent users, Automated Essay System would be quite suitable for large educational institutions. Utilizing services through the cloud with the optimization of the model for real-time scoring will make the system highly scalable in institutions where usage is deeply demanded.

To sum up, extensive user-testing with teachers and students will be done to guide extra refinements of the system so that Automated Essay System will correspond to actual needs in its implementation.

By looking for these improvements, future development will allow the model a higher level of accuracy flexibility, and accessibility to provide a significant addition to Automated Essay System (AES) into modern education.

1. REFERENCES
2. Handbook of Automated Essay Evaluation: Current Applications and New Directions.:

Shermis, M. D., & Burstein, J. (Eds.). (2013).

1. An automated essay scoring system using Learning Vector Quantization and natural language processing:

Elhoseny, M., Hassan, A., & Fahmy, A. (2016).

1. Efficient estimation of word representations in vector space:

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013).

1. Article - A new dataset and method for automatically grading ESOL texts:

Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011).

[5] M. Elhoseny, A. Hassan, and A. Fahmy (2016). (LVQ) Learning Vector Quantization with NLP techniques, highlighting vulnerabilities in AES systems.

[6] Ben Taskar, Carlos Guestrin, and Daphne Koller (2003). AES research, covering feature extraction, machine learning algorithms, and evaluation metrics.

[7] Lili Jiang and Nathan S. Netanyahu (2017). Explore the use of recurrent neural networks for AES.

[8] Tim Althoff, Cynthia Rudin, and David Williamson (2019) discuss bias and fairness issues in AES systems.