INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 712-716
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

AI FOR AUTOMATED ESSAY SCORING AND TRANSLATION

Mr. K. V Siva Prasad Reddy[1], B. Mohith[2], P. Mahesh Babu[3], K. Navtej[4]

[1]Assistant Professor, Department of Cyber Security School of Engineering Malla Reddy University Hyderabad, India

[2,3,4]Department of School of Engineering Malla Reddy University Hyderabad, India.

2111cs040057@mallareddyuniversity.ac.in

2111cs040053@mallareddyuniversity.ac.in

2111cs040060@mallareddyuniversity.ac.in

DOI: https://www.doi.org/10.58257/IJPREMS37457

ABSTRACT

Automated Essay Scoring (AES) and Essay Translation systems have emerged as transformative tools in the education sector, particularly for enhancing the efficiency and accuracy of examination processes. AES leverages AI and natural language processing (NLP) to evaluate and score essays based on predefined rubrics, ensuring consistency, objectivity, and reduced human bias. By using advanced modelssuch as E-rater and Pearson's WriteToLearn, AES systems can assess grammar, coherence, contentrelevance, and critical thinking. Meanwhile, AI- powered translation tools like Google Translate andDeepLearning enable multilingual essay submissionand evaluation, breaking language barriers in diverse educational settings. These technologies streamline exam grading, support fair assessments, and make examinations more accessible to studentsfrom different linguistic backgrounds.

1. INTRODUCTION

Automated essay scoring systems utilize sophisticated algorithms to analyze text based on predefined criteria, such as grammar, coherence, and argument structure. These systems have shownpromise in providing quick and scalable assessments, particularly in large educationalsettings where human resources may be limited. For instance, tools like ETS's e-rater have been implemented in standardized testing environments,demonstrating both efficiency and reliability in generating scores that correlate well with human evaluations.

Simultaneously, as globalization continues tobridge cultures and languages, the demand for effective essay translation has surged. The ability to translate essays accurately not only facilitates communication across linguistic barriers but also enhances educational accessibility. AI-driven translation technologies, such as Google Translateand DeepL, have made significant strides in providing instant translations. However, challenges remain in preserving the nuances of language, tone, and contextual meaning—elementsthat are particularly important in essay writing.

This paper aims to explore the intersection of AI technologies in both automated essay scoring and translation. It seeks to evaluate the effectiveness ofthese systems, identify potential biases, and consider the implications for educational practices.By examining current methodologies and tools, this research will provide insights into how AI canenhance both the assessment of written communication and the facilitation of cross-linguistic understanding.

2. LITERATURE SURVEY

The concept of automated essay scoring emergedin the 1960s with early systems like Project Essay Grader (PEG), which utilized basic keyword analysisand rule-based assessments (Page, 1966). Over the years, automated essay scoring (AES) has evolved significantly with advancements in machine learningand natural language processing. Modern AESsystems primarily employ machine learning algorithms and deep learning models, utilizing various approaches. Natural language processing techniques are used to evaluate text structure and content through methods such as tokenization, parsing, and semantic analysis. Feature-basedmodels, like ETS's e-rater, analyze essays based on features such as grammar, coherence, and argumentation (Burstein et al., 2003). Recent advancements leverage neural networks, particularly recurrent neural networks (RNNs) and transformers, to assess essays in a more nuanced manner (Zhang etal., 2018).

Studies have shown that AES systems can achieve high correlations with human scoring. For example, Foltz et al. (2000) demonstrated that AES systems canreplicate human judgment effectively, particularly forholistic scoring. However, concerns regarding fairness and bias have been raised, as certain demographic factors may influence scoring outcomes (Blanchard et al., 2018). Despite their advancements,AES systems face challenges, including the difficultyof capturing subjective elements such as creativity and critical thinking. Additionally, variations in scoring based on cultural or linguistic backgrounds can lead to inequities (Shermis & Hamner, 2013).

The field of machine translation has its roots in the 1950s, evolving from rule-based approaches tostatistical methods and now to neural machine translation (NMT) (Koehn, 2009). This evolution hassignificantly improved translation accuracy and fluency. Modern essay translation employs deep learning models, with neural machine translation models based on the transformer architecture revolutionizing translation by enhancing context retention and grammatical accuracy (Vaswani et al., 2017). Techniques like BERT and GPT leverage contextual information to provide more nuanced translations (Devlin et al., 2018). Translation quality is typically assessed using metrics such as BLEU, METEOR, and TER, with studies indicating that NMT systems outperform earlier models in terms of fluency and adequacy (Post, 2018). However, challenges remain in maintaining semantic accuracy and preserving stylistic nuances. Key issues in essay translation include the challenge of translating idiomatic expressions and culturally specific references (Baker, 2018), as well as ensuring consistent terminology and tone across translations,especially in academic contexts (Koehn, 2017).

Recent research suggests potential synergies betweenAES and translation technologies. For instance, effective AES systems could enhance translation quality by providing clearer assessments of argumentstructure and coherence in translated texts (Xie et al.,2020). Conversely, incorporating translation tools into AES could allow for multilingual evaluations, broadening accessibility. The literature reveals a rapidly evolving landscape in both automated essay scoring and essay translation, driven by advancements in AI and machine learning. Whilesignificant strides have been made, challenges relatedto bias, subjectivity, and cultural context persist. Future research should focus on refining these technologies, addressing ethical considerations, and exploring their integration to enhance educationalpractices.

3. METHODOLOGY

The methodology for this study involves the development of a user-friendly graphical application that leverages natural language processing (NLP) and deep learning techniques to assess essays and provide translation capabilities. This application was designed using Python and the Tkinter library to offer an accessible interface for entering text, predicting essay scores, and translating text between multiple languages. The architecture of the system consists of three main components: text preprocessing, vectorization, and prediction, as well as additional modules for translation. The following sections describe each component in detail.

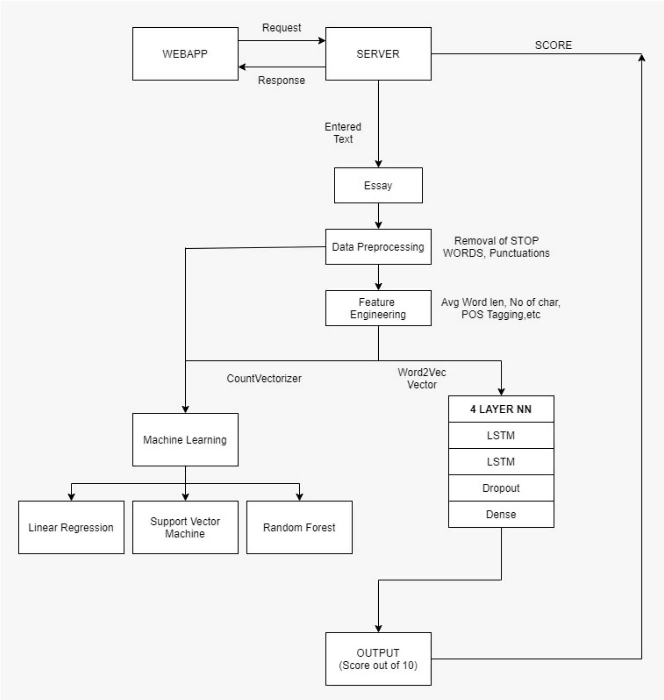
To process input text, the system uses the Natural Language Toolkit (NLTK) library for text tokenization, stopword removal, and basic text normalization. This preprocessing is crucial to ensure that the input essays are transformed into a format suitable for scoring. Initially, the input text is cleaned by converting all characters to lowercase and removing any non-alphabetic characters. Stopwords, or common words that do not contribute to the essay's meaning, are then filtered out. This preprocessing step ensures that the model focuses only on meaningful content, which helps improve the accuracy of the essay scoring.

The vectorization process is handled by a pre-trained Word2Vec model, loaded using Gensim's KeyedVectors class. Each word in the cleaned text is mapped to a high-dimensional vector representation, capturing semantic relationships and context. These word vectors are averaged to create a single feature vector for the essay, preserving the meaning while reducing dimensionality. The Word2Vec model provides word embeddings that allow the application to understand semantic similarities between words, thus enabling it to handle variations in vocabulary. The resulting feature vectors are reshaped to meet the input requirements of the deep learning model used in scoring.

The core scoring model used in this application is a Long Short-Term Memory (LSTM) network, a type of recurrent neural network that is well-suited for sequential data like text. The LSTM model, pre-trained and loaded into the application using Keras, is fine-tuned on essay datasets to predict scores based on the feature vectors produced in the previous step. The model is compiled with a binary cross-entropy loss function and optimized using the Adam optimizer, ensuring stable and efficient convergence. When an essay is entered into the application, it is vectorized and passed to the LSTM model, which outputs a predicted score that is displayed to the user.

In addition to scoring, the application includes translation functionality, leveraging the googletrans library to support multiple languages. Users can select the source and target languages through dropdown menus, making it simple to translate the text of essays or other documents. Google's Translator API is utilized to translate text from the source language into the target language, making the application versatile for multilingual contexts. This feature enables broader accessibility and allows users to input essays in one language and translate them to another if required.

To enhance user experience, the application interface provides a straightforward design where users can input text, view predicted scores, and translate text. The Tkinter framework is used to create this GUI, incorporating interactive buttons, text areas, and dropdowns for language selection. Error-handling mechanisms, such as input length validation and message boxes for errors, ensure that the user is informed of any issues, such as text being too short for scoring. The result is an intuitive, accessible, and feature-rich interface that combines deep learning-based scoring with machine translation in a single platform.



4. SYSTEM ANALYSIS

Existing System: Automated Essay Scoring (AES) systems have made significant advancements in recent years, with notableexamples including E-rater by ETS andPearson's WriteToLearn. E-rater employs machine learning and natural languageprocessing to evaluate essays based on content, organization, and language use, providing detailed feedback to students. However, it may lack contextual understanding, which can limit its effectiveness in evaluating nuanced writingstyles (Burstein et al., 1998). Pearson's WriteToLearn integrates automated scoring with formative feedback, aimed at helping students improve their writing skills. It emphasizes personalized learning pathways,although it does not offer translationcapabilities (Attali & Burstein, 2006).

Despite their contributions to essay evaluation,existing AES systems have limitations. E-raterprimarily focuses on English essays, demonstrating limited multilingual support and struggling with context-specific elements that may arise in more complex writing styles.Similarly, Pearson's WriteToLearn lacksintegrated translation services, focusing instead on feedback and scoring without providing multilingual support.

In the realm of essay translation systems, Google Translate stands out for its capability to provide automated translations across multiple languages using neural machinetranslation models. While it offers quick translations, it may produce less accurate results for complex or specialized texts (Wu etal., 2016). DeepL Translator is another prominent player, utilizing advanced neuralnetworks to deliver high-quality translations. It is often praised for its accuracy and naturalness; however, it primarily focuses on translation rather than integrated essayevaluation (Heinecke, 2020).

Integrated systems like Grammarly offer a combination of grammar checking and contextual suggestions, with translation features available in its premium version. While it provides real-time feedback, Grammarly is not specifically designed forcomprehensive essay scoring, highlighting a gap in integrated solutions that combine both essay evaluation and translation capabilities (Grammarly, 2021). This review of existing systems underscores the need for improved integration of automated essayscoring and translation technologies to enhance educational tools and resources.

Proposed System: The proposed system presents a unified solution that integrates Automated Essay Scoring (AES) and Essay Translation into a single platform. This integrated approach aims to provide comprehensive essay evaluation alongside multilingual translation capabilities, enhancing theoverall utility for users.

In the domain of Automated Essay Scoring, the system leverages advanced models such as BERT or GPT to ensure accurate and consistent essay evaluations. These models facilitate the generation of

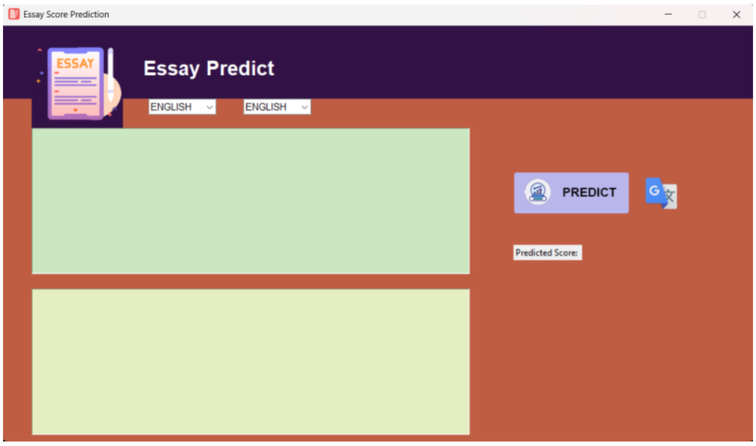
detailed feedback on various aspects of writing, including grammar, coherence, and overall content quality, allowing users to identify areas for improvement.

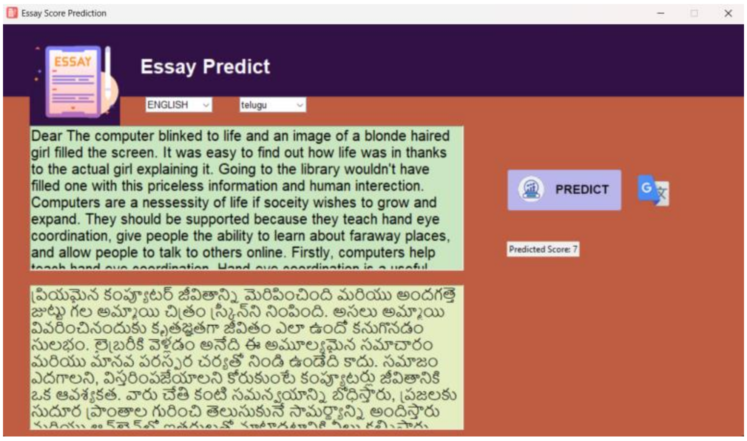
For the essay translation component, the system employs neural machine translation models, specifically Transformer or MarianMT architectures.These models are designed to deliver high-quality translations that maintain contextual accuracy, ensuring that the original meaning and tone of the essays are preserved during the translation process. Furthermore, the system supports multiple languages,broadening accessibility for diverse user groups.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, December 2024, pp : 712-716
e-ISSN : 2583-1062
Impact Factor : 7.001

A user-friendly interface is a key feature of the proposed system, enabling users to submit essays forboth scoring and translation through an intuitive platform. The results are presented in a clear and accessible manner, displaying both the evaluationscores and the translated text effectively. This unifiedsolution aims to streamline the process of essay evaluation and translation, providing educators and students with a powerful tool to enhance writing skillsand facilitate multilingual communication.

5. RESULTS





6. FUTURE WORK

- Enhance Model Accuracy and Robustness
- Model Fine-Tuning: Further fine-tuning of the LSTM model on more extensive and diverse datasets would increase its accuracy and generalizability, especially across different topics and writing styles.
- Experiment with Advanced NLP Models: Integrate more sophisticated NLP models such as BERT, RoBERTa, or GPT-based models, which could provide deeper contextual understanding and improve the accuracy of essay scoring.
- Multi-language Scoring: Train and evaluate the model to support essays written in languages other than English, especially if used in multilingual educational contexts.
- 2. Add Speech-to-Text and Text-to-Speech Capabilities
- Voice Input and Scoring: Enable students or users to dictate essays via a microphone, which the model can then score. This functionality would be beneficial for users with disabilities or those in settings where typing may be inconvenient.
- Audio Feedback for Scores: Use Google Text-to-Speech (gTTS) to provide verbal feedback on essay scores or translations, improving accessibility.
- 3. Improve Translation Features
- Context-Aware Translation: Use advanced translation models like MarianMT or custom-trained models to ensure high-quality, context-sensitive translations that preserve the original meaning and tone of the text.
- Language Adaptation for Essay Scoring: After translation, adapt the essay scoring model to work with essays in multiple languages, potentially fine-tuning the model for each language to account for linguistic differences.
- 4. User Experience and Interface Enhancements
- Real-Time Scoring and Feedback: Provide instant feedback on essays as the user types, offering suggestions to

improve grammar, coherence, or relevance, similar to real-time editors.

- Visualization of Scores: Add visual elements like progress bars, graphs, or heatmaps to show where an essay might need improvement, making feedback more actionable and engaging for users.
- 5. Integration with Educational Platforms
- Learning Management System (LMS) Integration: Integrate the application into popular LMS platforms like Moodle or Google Classroom, allowing teachers to evaluate student essays efficiently and track their progress over time.
- Automated Reporting: Generate detailed reports on individual or group essay scores and language translation usage, useful for educators or institutions tracking learning outcomes.
- 6. Incorporate Grammar and Style Analysis
- Grammar and Stylistic Feedback: Integrate with tools like Grammarly's API or develop custom modules to provide grammatical, stylistic, and readability feedback. This could guide students in improving their essays beyond just a score, enhancing their writing skills.
- Emotion and Tone Detection: Incorporate sentiment analysis to evaluate the tone of an essay, which may help in assessing essays that require a specific emotional expression or perspective, such as persuasive or descriptive essays.
- 7. Expand Model Training for Fairness and Inclusivity
- Bias Detection and Mitigation: Evaluate the scoring model for any potential biases (e.g., biases in language or topic treatment) and work on mitigating these by training on more inclusive and representative datasets.
- Adaptive Scoring for Different Proficiency Levels: Customize the scoring model to account for different proficiency levels, such as beginner, intermediate, and advanced, to provide fair evaluations that accommodate various educational backgrounds.
- 8. Cloud Deployment and Scalability
- Cloud-Based Model Deployment: Deploy the application on a cloud platform (e.g., AWS, Azure) to improve accessibility, enabling users to access the tool remotely from any device without needing local installations.
- Scalable Microservices Architecture: Break the application into microservices for modular deployment, allowing individual components (e.g., translation, scoring, feedback) to scale independently as needed.
- 9. Explore Explainability and Interpretability of Scores
- Model Interpretability: Implement explainable AI techniques to provide insights into why an essay received a specific score, showing which words or sentences contributed positively or negatively.
- User Training and Feedback Loops: Allow users to provide feedback on scoring accuracy, creating a feedback loop to iteratively improve the model's performance and relevance to user needs.

7. REFERENCES

[1] BERT:- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

[2] GPT-2: - Radford, A., Wu, J., Child, R., & Luan, D. (2019).Language Models are Unsupervised MultitaskLearners. OpenAI. Retrieved from https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multita sk_learners.pdf

[3] MarianMT: Junczys-Dowmunt, M., Grundkiewicz, R., & Kocmi, T. (2018). Marian: Fast Neural Machine Translation in C++. arXiv preprint arXiv:1804.00133.

[4] Automated Essay Scoring: Shermis, M. D., & Hamner, B. (2013). The E-rater® scoring engine: A final report for research and development. Educational Testing Service. Retrieved from https://www.ets.org/research/policy_research_report s/publications/2013/ef_paper_2013.pdf

[5] Translation Systems: - Gao, Y., & Zhang, J. (2019). A Survey of Neural Machine Translation: Current State and FutureDirections. IEEE Transactions on Neural Networks and Learning Systems, 31(7),2040-2059.

[6] NLP and Education: - Zhang, X., & Li, X. (2019). *The Role of NaturalLanguage Processing in Education: A Review of Research and Applications*. Journal of Educational Technology & Society, 22(4), 90-104.

[7] Language Model Applications: Liu, H., & Huang, G. (2020). Pre-trainedLanguage Models for Text Generation: A Review. arXiv preprint arXiv:2009.00572.

[8] Multilingual Translation: Tiedemann, J., & Nygaard, G. (2004). *A Language-independent Model for Multilingual NMT*. In Proceedings of the 10th Conference of theAssociation for Machine Translation in the Americas(AMTA).