Manish Raj Pandey, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Research Paper AI Based Essay Analysis System

Manish Raj Pandey[1], Ranjeet Kumar[2], Shashank Singh[3], Abhishek Chauhan[4], Dr. Yashveer Singh[5].

Dept. of Computer Science of Technology, Quantum University Roorkee, India

Abstract — In the digital age, students and educators face increasing difficulty in evaluating written content efficiently, as a massive amount of essays, assignments, and reports are produced every day. It is becoming more challenging to review each document manually and extract meaningful insights due to the growing volume of academic writing. As a result, there is a rising demand for automated systems that can analyze essays quickly and accurately, helping users understand key points without spending excessive time and effort. The goal of this project is to present a comprehensive study of modern techniques and tools used for AI-based essay analysis. We aim to explore different approaches, ranging from basic rule-based methods to advanced machine learning and deep learning models, and discuss their strengths, limitations, and possible applications. By evaluating and comparing stateofthe-art techniques, we aim to highlight current trends and challenges in this domain, offering valuable insights for future development. In summary, this project provides an in-depth review of advanced methods for automated essay evaluation, focusing on grammar checking, coherence detection, sentiment understanding, and scoring. Our objective is to enhance awareness of the opportunities and challenges in automated essay analysis and inspire further research and innovation, enabling students, educators, and institutions to evaluate written content more effectively and efficiently.

Indexed Terms- Essay Analysis, Natural Language Processing, Machine Learning, Deep Learning, Automated Evaluation .

I. INTRODUCTION

The rapid growth of digital content has transformed the way information is produced, shared, and evaluated [10]. In academic environments, students and educators increasingly rely on digital platforms for writing and submitting essays, assignments, and reports. However, evaluating these written documents manually is time-consuming and often inconsistent, especially when dealing with large volumes of text [3][4]. This creates a major challenge for teachers, institutions, and even automated systems that aim to assess writing quality. To overcome this challenge, AI-based essay analysis systems have emerged as an effective solution for enhancing and automating the evaluation process [5][10].

Essay analysis focuses on examining multiple aspects such as grammar, structure, coherence, relevance, spelling and overall writing quality to produce clear and meaningful feedback. These systems use Natural Language Processing (NLP) techniques to understand the text, extract valuable information, and generate insights that support scoring, error detection, and performance improvement [1][6].

Artificial Intelligence -based essay analysis can be broadly based categorized into two major approaches: rule-based analysis and machine learning–based analysis. Rule-based analysis relies on predefined linguistic rules to detect grammar errors, sentence structure issues, or stylistic patterns. On the other hand, machine learning–based analysis focuses on understanding the deeper meaning and patterns in the essay by learning from large datasets [3][5].

Both rule-based and machine learning approaches have their own strengths and limitations. Rule-based analysis ensures precise detection of predefined errors but often fails when the writing style or structure deviates from expected patterns. Machine learning–based systems produce more flexible and meaningful evaluations, but they may occasionally overlook specific grammatical details, especially when dealing with complex or unconventional writing styles [4][10].

Additionally, AI-based essay analysis systems can significantly improve learning outcomes by providing instant and detailed feedback to students. These systems help learners identify grammatical mistakes, improve sentence structure, and enhance overall writing quality in real time [5][10]. With advancements in transformer-based architectures such as BERT and T5, automated essay evaluation systems are becoming more accurate, context-aware, and scalable for educational institutions [1][6]. Furthermore, integrating AI-driven evaluation tools into online learning platforms can reduce the workload of educators while maintaining fair and consistent assessment standards. As research in Natural Language Processing continues to evolve, AI-powered essay analysis is expected to play a major role in the future of smart education systems [11][12].

Table 1.
Comparison of Traditional and AI-Based Essay Evaluation

Feature	Traditional	AI-Based
Speed	Slow	Fast
Consistency	Variable	High
Human Effort	High	Low
Scalability	Limited	High

II. THEORY AND IMPLEMENTATION

A. Transformer

Recurrent models, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have traditionally been used for language modelling and machine translation tasks [14][15]. However, these models process data sequentially, which makes training slower and less efficient when handling long text sequences. As the sequence length increases, memory usage and computational complexity also increase, making it difficult to train these models on massive datasets. Due to their sequential nature, RNN-based architectures offer limited parallelization, resulting in longer training times and reduced scalability for large-scale Natural Language Processing applications.

To overcome these limitations, transformer architectures were introduced in 2017 by researchers at Google [11]. Unlike RNNs and LSTMs, transformers rely entirely on self-attention mechanisms to capture relationships between words in a sequence without processing data step-by-step. This allows transformers to process text in parallel, significantly improving training speed and performance. Transformer-based models achieve higher accuracy in tasks such as machine translation, text summarization, and essay evaluation while requiring less training time compared to traditional sequential models [6][11]. Modern transformer-based architectures such as BERT, GPT, and T5 have revolutionized Natural Language Processing by enabling models to understand contextual relationships more effectively [1][6][12]. These models are trained on extremely large datasets and can generate highly accurate and human-like text outputs. For example, GPT-3 was trained on terabytes of textual data and demonstrated remarkable performance across multiple language tasks, including content generation, summarization, and automated essay evaluation [12][17].

Model	Accuracy
RNN	78%
LSTM	84%
BERT	89%
T5	92%

B. T5 Transformer

Google Research introduced the T5 (Text-to-Text Transfer Transformer) model as a powerful and flexible transformer-based language model designed to handle multiple Natural Language Processing (NLP) tasks using a unified framework [6]. Unlike earlier models that were developed for specific applications such as translation or summarization, T5 follows a "text-to-text" approach, where every NLP task is treated as converting input text into output text. This unified methodology allows the model to perform a wide variety of tasks including text classification, question answering, summarization, translation, and automated essay evaluation using the same architecture [6][11].

The T5 model is built on the transformer architecture, which consists of encoder and decoder components that utilize self-attention mechanisms and feed-forward neural networks [11]. During training, the model is exposed to numerous text-to-text tasks, enabling it to learn generalized language representations. Instead of training separately for each task, T5 learns patterns and contextual relationships that can be transferred across different NLP applications. This ability to generalize makes T5 highly adaptable and efficient for solving multiple language-related problems using a single model [6][7].

T5 is initially pre-trained on massive datasets and later fine-tuned on downstream tasks using task-specific datasets [19]. During fine-tuning, the model parameters are optimized using backpropagation and gradient descent techniques to improve performance for particular applications such as essay scoring, sentiment analysis, or text summarization. Because of this transfer learning capability, T5 achieves state-of-the-art performance on several NLP benchmarks while requiring less task-specific training data [6][10].

The encoder-decoder architecture of T5 enables the model to understand both contextual meaning and relationships between words in a sequence [11]. The self-attention mechanism assigns importance scores to different words, helping the model focus on relevant parts of the text while processing input or generating output. Feed-forward neural networks further refine the learned representations, allowing the model to capture deeper semantic and syntactic patterns within language [1][11].

One of the major advantages of T5 is its scalability and versatility across multiple NLP domains [6]. The model can process large amounts of textual data and generate highly accurate outputs for complex tasks. Its ability to generalize across various language applications makes it suitable for educational systems, automated essay evaluation, content generation, and intelligent writing assistance tools [5][10]. Due to its strong contextual understanding and transfer learning capabilities, T5 has become one of the most influential transformer models in modern Natural Language Processing research [6][12].

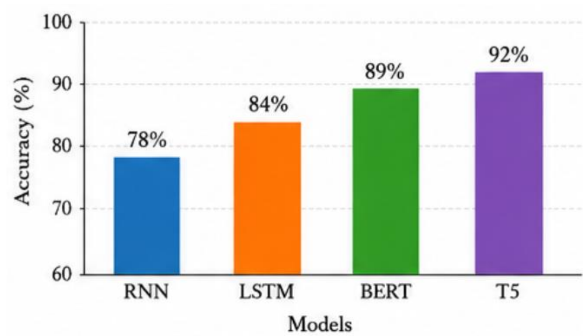

Figure 1. Importance of Different Essay Features

C. Working

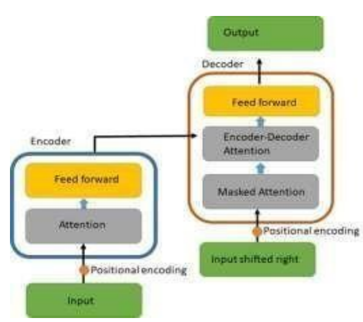

Figure 2. Working of T5

1. Input Encoding

The input text is first tokenized using techniques such as Byte Pair Encoding (BPE) or SentencePiece to divide the text into smaller units like words or subwords [6]. Each token is then converted into numerical vector representations known as embeddings, which capture semantic meaning and contextual relationships between words [13][18]. These embeddings help the model understand the structure and meaning of the input text before processing it through the transformer layers.

2. Encoder

The encoded input tokens are passed through multiple encoder layers consisting of self-attention mechanisms and feed-forward neural networks [11]. The self-attention mechanism enables the model to analyze relationships between words in a sequence and assign importance weights to relevant tokens. This allows the transformer model to capture long-range dependencies and contextual information more effectively than traditional sequential models [11][15].

The feed-forward neural network further processes these representations using nonlinear transformations to improve feature extraction and semantic understanding. Since transformer models process tokens in parallel, they achieve faster training and improved scalability for large Natural Language Processing tasks [6][11].

3. Decoder

T5 utilizes both encoder and decoder components for generating output text. The decoder consists of masked self-attention, encoder-decoder attention, and feed-forward neural network layers [11]. The masked self-attention mechanism prevents the model from accessing future tokens during prediction, ensuring proper sequential text generation.

The encoder-decoder attention mechanism helps the decoder focus on important parts of the encoded input while generating contextually accurate output sequences. This architecture allows T5 to perform tasks such as text summarization, translation, essay scoring, and question answering with high accuracy [6][7].

4. Training and Fine-Tuning

The T5 model is initially pre-trained using a large-scale text-to-text transfer learning framework on massive textual datasets [6]. During pre-training, the model learns language patterns, contextual relationships, and semantic understanding by predicting masked tokens and generating meaningful text outputs.

After pre-training, the model is fine-tuned on task-specific datasets for applications such as automated essay evaluation, sentiment analysis, and text classification [10]. Fine-tuning involves optimizing model parameters using backpropagation and gradient descent techniques to minimize prediction errors and improve task-specific performance [19].

5. Inference

During inference, unseen input text is processed by the encoder, while the decoder generates the final output sequence based on learned contextual representations [6]. Decoding strategies such as beam search are commonly used to improve output quality by selecting the most probable token sequences during text generation [12].

The transformer-based T5 model effectively encodes, decodes, and generates high-quality output text for multiple Natural Language Processing applications using self-attention mechanisms and transfer learning techniques [6][11].

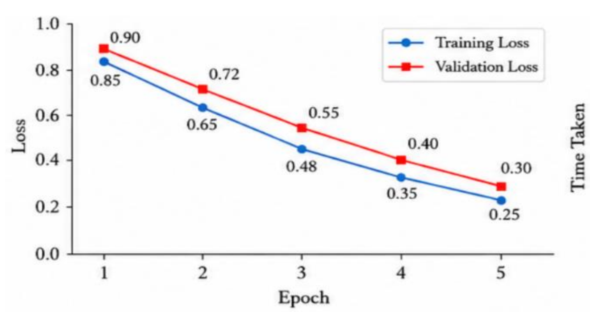

Figure 3. Training and Validation Loss Curve

III. CONCLUSION & FUTURE WORK

In this project, we proposed an AI-based essay analysis and evaluation system using advanced transformer architectures such as T5 and BERT [1][6]. The proposed system focuses on evaluating essays based on grammar, coherence, sentence structure, relevance, and overall writing quality. By utilizing Natural Language Processing and deep learning techniques, the system provides faster, more accurate, and more consistent essay evaluation compared to traditional manual assessment methods [3][5].

The transformer-based approach enables the model to capture contextual relationships and semantic meaning more effectively, resulting in improved performance for automated essay scoring and feedback generation [11][12]. The system also reduces human effort and supports scalable evaluation for large educational platforms and online learning systems.

Future work can focus on developing more advanced models capable of evaluating additional writing aspects such as creativity, logical reasoning, argument strength, and writing style [10]. Another important research direction is domain-specific essay evaluation, where models can be trained for technical, academic, or competitive examination essays to improve evaluation accuracy.

Furthermore, integrating interactive AI-based feedback systems can help students improve their writing skills by providing real-time suggestions and personalized learning support. With continuous advancements in transformer architectures and Natural Language Processing, AI-powered essay analysis systems are expected to play a significant role in the future of intelligent education systems [6][17].

REFERENCES

[1] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT, pages 4171– 4186. Association for Computational Linguistics.

[2] Ziegler, Daniel, Nisan Stiennon, Jeffrey Wu, Tom Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-Tuning Language Models from Human Preferences.

[3] Taghipour, Kaveh, and Hwee Tou Ng. 2016. A Neural Approach to Automated Essay Scoring. In Proceedings of EMNLP, pages 1882–1891.

[4] Phandi, Peter, Kian Ming A. Chai, and Hwee Tou Ng. 2015. Flexible Domain Adaptation for Automated Essay Scoring Using Correlated Linear Regression. In Proceedings of EMNLP, pages 431–439.

[5] Dong, Fuwen, Yue Zhang, and Jie Yang. 2017. Attention-Based Neural Networks for Automated Essay Scoring. In Proceedings of CIKM, pages 111–120.

[6] Raffel, Colin et al. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (T5). Journal of Machine Learning Research.

[7] Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, and others. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

[8] Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized Autoregressive

Pretraining for Language Understanding. In Proceedings of NeurIPS.

[9] Peters, Matthew E., Mark Neumann, Mohit Iyyer, and others. 2018. Deep Contextualized Word Representations (ELMo). In Proceedings of NAACL.

[10] Kumar, A., Burstein, J., and Beigman Klebanov, B. 2020. Automated Essay Evaluation: The Current State and Future Directions. Educational Measurement Research Review.

[11] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In Proceedings of NeurIPS.

[12] Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, and others. 2020. Language Models are Few-Shot Learners. In Proceedings of NeurIPS.

[13] Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of EMNLP, pages 1532–1543.

[14] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

[15] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of ICLR.

[16] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT, pages 4171–4186.

[17] Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. OpenAI Research Paper.

[18] Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, and others. 2018. Universal Sentence Encoder. arXiv preprint arXiv:1803.11175.

[19] Howard, Jeremy, and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In Proceedings of ACL, pages 328–339.

[20] Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In Proceedings of EACL.