IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15178

AUTOMATED ESSAY GRADING USING MACHINE LEARNING

Nayana N K[1], Dr. Madhu H.K[2]

Department of MCA, BIT, Bangalore, India,

Professor and head, Department of MCA, BIT[1,2]

Abstract: The increasing volume of written assessments in education has made manual essay evaluation time-consuming, subjective, and inconsistent. To address these challenges, this paper presents an Automated Essay Grading System that uses Natural Language Processing (NLP) and Machine Learning techniques to evaluate essays efficiently and fairly. The proposed system combines TF-IDF features and BERT-based embeddings to capture both statistical and contextual meaning of text, and employs a Ridge Regression model for accurate score prediction. In addition to grading, the system provides detailed feedback including grammar checking, readability analysis, sentiment evaluation, and bias detection, helping students improve their writing skills. A user-friendly web interface developed using Streamlit allows users to submit essays and view results instantly. The system also supports multilingual essay grading for regional languages such as English, Kannada, Telugu, and Tamil, making it more inclusive and accessible. Experimental results show that the system delivers consistent, scalable, and objective evaluation while significantly reducing grading effort. This work demonstrates the potential of AI-driven assessment tools in enhancing modern educational practices.

Keywords: Automated Essay Grading, Natural Language Processing, Machine Learning, BERT, TF-IDF, Ridge Regression, Multilingual Essay Evaluation, Educational Technology, AI-Based Assessment

1. INTRODUCTION

Essay evaluation plays a vital role in the education system, as it helps assess a student's understanding, critical thinking, and writing skills. However, traditional manual grading methods are time-consuming, subjective, and often inconsistent, especially when handling a large number of essays. With the growing use of digital learning platforms, the need for an efficient and reliable automated evaluation system has become increasingly important.

Advancements in Artificial Intelligence and Natural Language Processing (NLP) provide powerful tools for addressing these challenges. An Automated Essay Grading System uses machine learning models to analyze written responses and generate scores along with meaningful feedback. By combining techniques such as TF-IDF feature extraction and BERT-based contextual understanding, the system can evaluate essays more accurately and consistently. The inclusion of multilingual essay grading further enhances accessibility, allowing students to submit essays in regional languages like Kannada, Telugu, Tamil, and English. This approach improves fairness, reduces evaluation time, and supports modern, scalable educational assessment systems.

1.1 Project Description

The Automated Essay Grading System is an AI-based web application that automatically evaluates student essays and provides meaningful feedback. Using Natural Language Processing and Machine Learning, it analyzes essay content and context through a hybrid of TF-IDF features and BERT embeddings. The system supports multilingual grading in English, Kannada, Telugu, and Tamil, offering scores, grammar checks, readability analysis, sentiment evaluation, and structured feedback via a user-friendly interface.

1.2 Motivation

Manual essay grading is time-consuming, subjective, and challenging in large classrooms or online platforms. This project aims to automate grading while ensuring fairness and accuracy, reducing teacher workload, providing consistent evaluation, and supporting students from diverse linguistic backgrounds.

2. LITERATURE SURVEY

Ramesh and Sanampudi (2022) reviewed automated essay scoring systems using AI and NLP. They found deep learning models are widely used, but evaluating essay coherence and content relevance remains challenging.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15178

Borade and Netak (2020) analyzed traditional and modern AES methods. Their study shows automated grading reduces teacher workload but often lacks detailed feedback and adaptability to diverse essay topics.

Misgna et al. (2024) explored deep learning models, including LSTM, CNN, and transformers, for essay scoring. They reported improved grading accuracy but limited interpretability and actionable insights for learners.

Ibrahim et al. (2022) proposed an AES system using NLP preprocessing with bidirectional LSTM. Their approach enhances grading consistency and reduces evaluator bias, tested on the ASAP dataset.

Ghanta (2019) presented a feature-based AES system using Random Forest. By extracting grammar, vocabulary, and sentence structure features, the system achieved high agreement with human graders.

2.1 Existing System vs Proposed System

Existing Essay Grading System

Traditional essay evaluation relies heavily on manual grading by teachers or examiners. Essays are read and scored based on predefined rubrics, which makes the process time-consuming and subjective. Large volumes of essays, especially in online courses or competitive exams, create delays in feedback and inconsistencies in scoring. Some existing automated systems use rule-based or surface-level features like word count, spelling checks, or basic grammar rules. While they can detect minor language errors, they fail to assess semantic meaning, content relevance, or logical flow. Most commercial tools focus only on final scores without providing detailed feedback, making them less useful for educational improvement.

Proposed Automated Essay Grading System

The proposed system uses machine learning and NLP to evaluate essays efficiently and consistently. Essays are preprocessed using cleaning, tokenization, normalization, and lemmatization. Features are extracted using TF-IDF and BERT embeddings, capturing both word importance and semantic relationships. A Ridge Regression model predicts essay scores based on these features. In addition to scoring, the system provides real-time feedback on grammar, readability, and sentiment. A web interface built with Streamlit allows easy essay submission, while a Flask backend manages processing and result generation. This approach reduces human bias, saves time, scales for large volumes of essays, and offers actionable feedback for learners.

SYSTEM ARCHITECTURE DIAGRAM

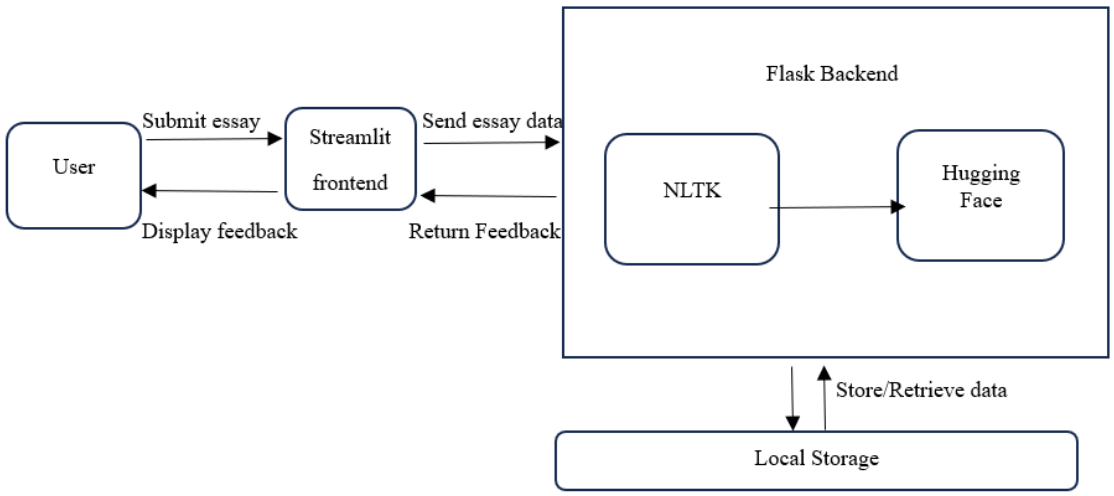

Fig.1 System Architecture Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15178

3. SYSTEM DESIGN

Data flow diagram

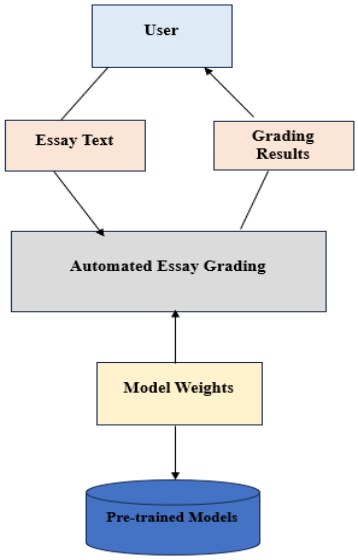

Fig. 1 Level 0 Data Flow Diagram

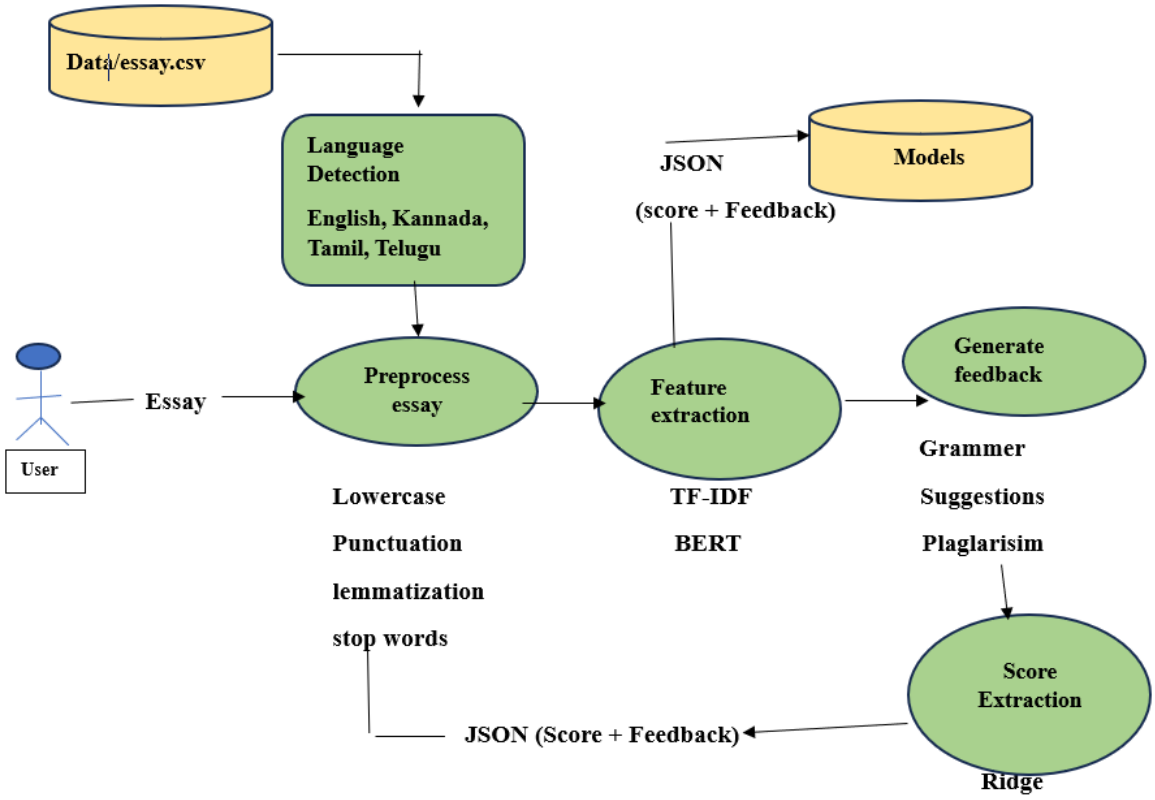

Fig. 3 Level 1 Data Flow Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15178

4. IMPLEMENTATION DETAILS

The system is implemented using a three-layer architecture:

Frontend Layer

- Developed using Streamlit for an interactive and user-friendly interface
- Allows students to submit essays and select the language (English, Kannada, Telugu, Tamil)
- Displays instant scores, detailed feedback, and visualizations of grammar, readability, and sentiment
- Communicates with the backend via REST API

Processing & ML Layer

- Text preprocessing: cleaning, tokenization, normalization, and lemmatization
- Feature extraction using TF-IDF and BERT embeddings to capture content and context
- Ridge Regression model predicts essay scores based on extracted features
- Performs grammar checking, readability analysis, and sentiment evaluation

Backend Layer

- Built using Flask to handle essay submissions, API requests, and scoring
- Pandas and NumPy manage datasets and feature matrices
- Stores essay metadata, results, and feedback logs
- Supports modular expansion for multi-language grading and future integration with Learning Management Systems (LMS)

4.1 System Modules and Workflow

System Modules

User Module

- Essay submission through web interface
- Language selection (English, Kannada, Telugu, Tamil)
- View overall score and detailed feedback
- Download PDF/JSON report

Admin Module

- Admin authentication
- Upload and manage essay datasets
- Train and update ML models
- Monitor system performance and review reports

Processing & ML Module

- Text preprocessing (cleaning, tokenization, normalization, lemmatization)
- Feature extraction using TF-IDF and BERT embeddings
- Score prediction using Ridge Regression

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15178

- Grammar, readability, sentiment, and plagiarism analysis

Visualization Module

- Displays score distribution, sentiment analysis, bias detection, and progress tracking

Workflow

- Admin uploads essay datasets and configures grading parameters
- User submits an essay through the web interface
- System preprocesses essay and extracts features
- ML model predicts score and generates detailed feedback
- Feedback and visualizations are displayed to the user
- User can download a detailed report
- Admin monitors performance and updates models as needed

4.2 Testing Overview

Unit Testing

- Tested individual modules: text preprocessing, feature extraction, scoring, grammar, and feedback.

Integration Testing

- Verified smooth flow from essay submission → ML processing → feedback display.

System Testing

- End-to-end testing with real essays.
- Checked multi-language support, scoring, grammar, readability, and feedback.

Performance Testing

- Tested processing time for essays of different lengths.
- Ensured fast and stable results.

Security Testing

- Checked input validation to prevent errors or misuse.
- Protected admin functions and ML models.

User Acceptance Testing (UAT)

- Students submitted essays and reviewed feedback.
- Confirmed interface is clear and feedback is useful.

Regression Testing

- Re-tested system after updates to ensure stability.

Compatibility Testing

- Tested across browsers, operating systems, and supported languages.

Functional Testing

- Verified essay submission, language selection, scoring, and feedback display.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15178

Input Validation Testing

- Checked empty essays, very short essays, unsupported characters, and mixed-language input.

Usability Testing

- Ensured interface is intuitive and feedback is readable.

Model Validation Testing

- Verified predicted scores are consistent and fair.

Error Handling Testing

- Checked system handles errors without crashing.

5. RESULTS AND DISCUSSION

The Automated Essay Grading System was successfully implemented and tested with essays in English, Kannada, Telugu, and Tamil. Students could submit essays through the web interface and receive instant evaluation, including scores, grammar and spelling feedback, readability analysis, sentiment evaluation, and plagiarism detection. Using TF-IDF and BERT embeddings with Ridge Regression, the system provided accurate and consistent grading across topics and languages.

The system also offered detailed, downloadable reports with suggestions for improvement and visualizations of key metrics. Performance testing showed fast processing even for longer essays, while input validation and error handling ensured stability. Compared to manual grading, the system reduced evaluation time, removed human bias, and provided actionable feedback, enhancing students' learning experience.

6. CONCLUSION

The Automated Essay Grading System demonstrates the effective use of Artificial Intelligence and Natural Language Processing to automate essay evaluation, combining TF-IDF features with BERT embeddings for accurate and reliable scoring. The system provides detailed feedback, including grammar corrections, readability, sentiment analysis, and bias detection, helping students understand their strengths and areas for improvement while saving educators time. Built with a modular architecture using Flask and Streamlit, it ensures ease of use, scalability, and consistent performance across multiple essays. Overall, this project highlights how AI-driven tools can enhance efficiency, fairness, and learning outcomes in modern education, laying a foundation for future enhancements such as multilingual support and LMS integration.

7. FUTURE WORK

In the future, the system can offer AI-powered personalized feedback, support voice and handwriting submissions, and provide dashboards to track student progress over time. Mobile access and offline functionality can make it more accessible, while gamification features can increase student engagement. Explainable AI grading and integration with learning platforms will further enhance transparency and streamline the assessment process.

REFERENCES

[1] R. Ramesh and P. Sanampudi, "Automated Essay Scoring: A Systematic Literature Review," *International Journal of Computer Applications*, 2022.
[2] S. Borade and N. Netak, "Automated Grading of Essays: A Review," *International Journal of Advanced Research in Computer Science and Software Engineering*, 2020.
[3] B. Misgna, A. Teshome, and M. Tesfaye, "Deep Learning-Based Automated Essay Scoring and Feedback Generation," *International Journal of Educational Technology in Higher Education*, 2024.
[4] S. Ibrahim, R. Ahmed, and T. Khan, "Improved Automated Essay Grading Using NLP and Deep Learning," *Journal of Emerging Technologies and Innovative Research*, 2022.
[5] V. Ghanta, "Automated Essay Evaluation Using Machine Learning," *International Journal of Computer Science and Information Security*, 2019.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15178

[6] N. Taghipour and H. T. Ng, "A Neural Approach to Automated Essay Scoring," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016.
[7] D. Alikaniotis, H. Yannakoudakis, and T. Rei, "Neural Automated Essay Scoring Using Word Embeddings," *Computational Linguistics*, 2016.
[8] M. Shermis and J. Burstein, *Handbook of Automated Essay Evaluation*, Routledge, 2013.
[9] H. Yannakoudakis, T. Briscoe, and B. Medlock, "Automatic Essay Grading for Language Learners," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 2011.
[10] P. W. Page, "Project Essay Grade (PEG)," *Journal of Educational Measurement*, 1966.
[11] Y. Lu and A. Cutumisu, "Deep Learning for Automated Feedback in Essay Scoring," *International Journal of Artificial Intelligence in Education*, 2021.
[12] R. Kumar, S. Sharma, and P. Singh, "NLP Techniques for Enhancing Automated Essay Scoring," *International Journal of Advanced Research in Computer Science and Communication Engineering*, 2024.
[13] A. Mamidala, R. P. Rao, and S. K. Reddy, "AI-Driven Essay Grading Machine," *International Journal of Computer Applications*, 2025.
[14] L. Dong and D. Zhang, "Automatic Features for Essay Scoring," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016.
[15] Y. Ke and H. T. Ng, "Automated Essay Scoring with Deep Neural Models," *Journal of Educational Data Mining*, 2019.