Abstract

This project presents an automated essay evaluation system designed to assess essays using a hybrid approach that combines machine learning and rule-based natural language processing (NLP). The system employs a Random Forest Regressor trained on a TF-IDF representation of an IELTS writing dataset to predict overall essay scores. In addition to the model prediction, the system analyzes four critical dimensions of writing quality: grammar, structure, coherence, and vocabulary. Grammar is evaluated using syntactic dependencies via spaCy; structure is assessed through sentence complexity and average sentence length; coherence is measured by transition word usage, topic consistency, and paragraphing; and vocabulary is analyzed through word richness and frequency. Each component contributes to a composite scoring algorithm with weighted contributions, resulting in a final score reflecting both automated prediction and linguistic feature analysis. The tool includes a command-line interface allowing users to input essays manually or via file upload for detailed feedback and scoring.