Automated Essay Scoring and Feedback Generation – Detailed Report

1. Introduction

This project aims to develop an advanced Automated Essay Scoring (AES) system integrating classical natural language processing, machine learning, and state-of-the-art deep learning models. The objective is not only to predict essay scores accurately but also to generate detailed, personalized feedback that can guide students in improving their writing.

2. Dataset and Preparation

The dataset contained essay text along with holistic human-assigned scores. Data cleansing involved handling missing values, removing formatting inconsistencies, and ensuring uniform text encoding. A train–validation split ensured proper evaluation and prevented data leakage. Initial exploratory analysis helped understand score distributions, essay lengths, and variability in writing quality.

3. Linguistic Feature Engineering

A comprehensive set of linguistic features was extracted using spaCy, nltk, and textstat:

- Tokenization and part-of-speech analysis

- Syntactic dependency parsing

- Readability indices (Flesch-Kincaid, Gunning Fog Index)

- Lexical richness measurements (TTR, Hapax Legomena)

- Sentence-level metrics (average sentence length, complexity)

- Grammar and fluency indicators

These features provided interpretability and complemented deep-learning models by highlighting measurable writing attributes.

4. ULMFiT Modeling (AWD-LSTM)

ULMFiT was leveraged to build a language model fine-tuned on the essay corpus:

- The language model stage enabled the network to learn domain-specific writing patterns.

- The classifier head was trained to map text representations to essay scores.

- Training curves showed steady convergence, and evaluation metrics demonstrated robust performance.

ULMFiT served as a strong baseline due to its adaptability and efficient fine-tuning mechanism.

5. Transformer-Based Scoring with BERT

A BERT-based regression/classification model was implemented for deeper contextual understanding:

- Essays were tokenized with a modern Transformer tokenizer.

- A pre-trained BERT model was fine-tuned specifically for essay scoring.

- Evaluation used Cohen Kappa, a standard metric in AES research.

Results indicated that BERT outperformed ULMFiT significantly due to its bidirectional contextual encoding, capturing nuanced writing characteristics and semantic coherence.

6. Feedback Generation System

A hybrid feedback module was created:

- It analyzed linguistic features such as clarity, coherence, grammar, vocabulary strength, and structural organization.

- Combined with predicted scores, it produced structured feedback including:

* Strengths of the essay

* Areas needing improvement

* Specific actionable recommendations

The system bridges the gap between automated scoring and meaningful instructional assistance.

7. Final Evaluation and Model Comparison

Both models were compared on accuracy, consistency, and error rates:

- ULMFiT offered strong text representation learning with good accuracy.

- BERT demonstrated superior generalization with lower MAE and higher Cohen Kappa.

- The comparison table highlighted metric improvements across dimensions.

8. Interactive Essay Grader

A user-facing function was included:

- Accepts a custom essay as user input

- Generates predicted score instantly

- Produces tailored writing feedback

This transforms the notebook into a practical educational tool for real-world usage.

Conclusion

This assignment successfully demonstrates the development of a comprehensive AES pipeline: from data preparation and linguistic analysis to deep-learning modeling, evaluation, and interactive feedback delivery. The integration of classical NLP with modern AI models showcases a sophisticated, effective approach to automated writing assessment.