AI POWERED RUBRIC GENERATION & STUDENT ANSWER EVALUATION SYSTEM

Project Reference No.: *49S_BE_0031*

College : Dayananda Sagar College of Engineering, Bengaluru
Branch : Department of Artificial Intelligence and Machine Learning
Guide : Prof. Anupama Vijaykumar
Students : Ms. Disha B L
Mr. Shrihari Krishna Kasbe
Mr. Suryanshu Dhargave
Mr. Udayaraj

Keywords:

AI-based automated grading, rubric generation, criterion-based evaluation, Outcome-Based Education (OBE), Natural Language Processing (NLP), transformer models, Large Language Models (llms), FLAN-T5, BERT, deberta, semantic analysis, contextual understanding, text classification, personalized feedback, explainable AI, ROUGE score, cosine similarity, Mean Absolute Error (MAE), accuracy, fastapi, REST apis, Docker, scalable ML pipeline, JSON structuring, learning analytics, Bloom's taxonomy, course outcomes, human-in-the-loop validation

Introduction:

The rapid growth of digital education and the adoption of Outcome-Based Education (OBE) have changed how we assess student performance in schools today. Traditional evaluation methods depend on manual grading, which often has drawbacks such as subjectivity, inconsistency, delayed feedback, and increased workload for teachers, especially in large classrooms. These issues show the need for smart, scalable, and efficient assessment systems that can provide fairness, transparency, and timely feedback.

In recent years, Artificial Intelligence (AI) and Natural Language Processing (NLP) have become powerful tools that can automate tasks involving language understanding and decision-making. Using these advancements, automated grading systems have drawn attention for their potential to improve the evaluation process. However, most current systems rely on fixed rubrics or simple keyword-based methods, which do not work well for assessing descriptive answers that require deeper understanding and context.

To tackle these issues, this project introduces an AI-powered rubric-generating student assessment evaluation system. This system uses transformer-based models to create structured evaluation rubrics based on the question at hand and assesses student responses at the criterion level. By using models like FLAN-T5 for generating rubrics and BERT/DeBERTa for evaluating answers, the system provides a more thorough and context-aware grading process.

The proposed framework has a two-stage evaluation method. In the first stage, it generates a detailed rubric that aligns with learning goals. In the second stage, it evaluates student responses based on each rubric criterion. This method improves grading consistency and accuracy while allowing for meaningful, personalized feedback. Also, the system can handle various types of assessments, such as descriptive answers, quizzes, and programming assignments, making it versatile for different educational situations.

Overall, this project aims to create a scalable, transparent, and efficient automated grading solution that lightens the faculty's workload, improves evaluation quality, and meets modern educational standards. By combining AI-driven rubric generation with criterion-based evaluation, the system helps advance intelligent educational technologies and supports better learning outcomes.

Objectives:

1. Develop an AI-powered automated student assessment system.
2. Design and implement dynamic rubric generation using NLP and Large Language Models.
3. Perform criterion-based evaluation of student responses using transformer-based models.
4. Support multiple assessment types such as essays, quizzes, and programming assignments.
5. Generate personalized and meaningful feedback for students.
6. Ensure fairness, transparency, and consistency in the evaluation process.
7. Build a scalable and efficient system architecture for real-time grading.
8. Incorporate a human-in-the-loop mechanism for validation and continuous improvement.
9. Evaluate system performance using standard evaluation metrics.
10. Provide learning analytics and dashboards for monitoring student performance.

Methodology

The proposed system implements an AI-based automated grading framework that combines dynamic rubric generation with criterion-level evaluation of student responses. The methodology follows a structured process that includes rubric creation, answer evaluation, score aggregation, and system deployment. This approach ensures scalable and clear academic assessment.

1. Data Preparation

The system uses datasets that include assessment questions, rubrics defined by instructors, and student responses. The input data is prepared with Natural Language Processing techniques like tokenization and text normalization. The processed data supports both rubric creation and answer evaluation tasks.

2. Rubric Generation Module

A transformer-based sequence-to-sequence model automatically generates evaluation rubrics. The model uses the assessment question as input and produces multiple evaluation criteria with corresponding marks. The rubric generation process involves:

• Preprocessing the input question

• Generating rubric criteria using the model

• Structuring the output in a standardized JSON format

This module allows for the dynamic creation of rubrics tailored to each question, ensuring they meet learning objectives.

3. Criterion-Based Answer Evaluation

The generated rubric evaluates student responses in detail. A transformer-based classification model assesses each rubric criterion. The evaluation process includes:

• Combining the question, rubric criterion, and student answer

• Converting inputs into token embeddings

• Examining semantic and contextual relationships

• Predicting scores for each criterion

This method ensures a precise and context-aware evaluation of descriptive answers.

4. Multi-Type Assessment Evaluation

The system supports various types of assessments:

• Descriptive answers are evaluated with NLP-based transformer models

• Objective questions are evaluated using rule-based methods and machine learning techniques

• Programming assignments are evaluated through code analysis models and execution in a secure environment

This capability allows the system to handle a range of academic evaluation scenarios.

5. Score Aggregation and Feedback Generation

The system adds up the scores for each rubric criterion to compute the final grade. It generates a structured evaluation report that clearly displays scores for each criterion. Additionally, the system provides basic feedback by identifying strengths and areas for improvement based on rubric performance.

6. System Architecture and Deployment

The system operates on a scalable backend architecture. APIs are created to manage communication between different modules. The deployment framework supports:

• Model inference for rubric generation and evaluation

• Processing user inputs

• Generating and delivering evaluation results

Containerization techniques ensure efficient and scalable execution.

7. Performance Evaluation

The system's performance is assessed using standard metrics:

• The quality of rubric generation is measured with similarity-based metrics

• The performance of answer evaluation is assessed using accuracy and error-based metrics

The system is also evaluated for response time and scalability to support real-time usage.

8. Overall Workflow

The complete workflow of the system is as follows:

1. The instructor provides the assessment question
2. The system generates a structured rubric
3. The student submits their response
4. The system evaluates the response for each rubric criterion
5. Scores are aggregated to compute the final result
6. Evaluation results and feedback are generated and delivered

Result and Conclusion:

In summary, by combining dynamic rubric generation with criterion-based answer assessment, the suggested AI-powered student assessment system effectively automates the evaluation process. The system accurately assesses student responses and produces structured, context-aware rubrics. The findings show that while the answer evaluation model achieves satisfactory scoring performance with dependable accuracy and low error rates, the rubric generation model generates coherent evaluation criteria.

The system's versatility is increased by its ability to handle a variety of assessments, such as programming assignments, quizzes, and descriptive answers. Furthermore, the framework improves consistency by lowering human bias in grading and guarantees transparency through criterion-level scoring. The system's deployment demonstrates a respectable response time, which qualifies it for real-time academic settings. Overall, the suggested approach greatly lessens the amount of manual grading work while enhancing student assessment's effectiveness, equity, and scalability—thereby promoting outcome-based education.

References:

1. A. Vaswani et al., “Attention Is All You Need,” *Advances in Neural Information Processing Systems*, 2017.
2. J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL-HLT*, 2019.
3. C. Raffel et al., “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer,” *Journal of Machine Learning Research*, 2020.

4. H. Chung et al., “Scaling Instruction-Finetuned Language Models,” 2022.
5. Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” 2019.
6. P. He et al., “DeBERTa: Decoding-Enhanced BERT with Disentangled Attention,” 2020.
7. V. Sanh et al., “DistilBERT: A Distilled Version of BERT,” 2019.
8. V. Kumar et al., “LLM-Based Automated Grading of Programming Assignments,” *IEEE FIE*, 2023.
9. Y. Wang et al., “RubricGPT: Generating Scoring Rubrics with Large Language Models,” 2024.
10. L. Yan et al., “Automated Grading and Feedback Generation Using Large Language Models,” *IEEE Transactions on Learning Technologies*, 2024.

Future Scope:

The future scope of this project includes:

1. Improving the creation of feedback by using more sophisticated Large Language Models for thorough and customised explanations.
2. Using larger, domain-specific datasets and bias mitigation strategies to improve model accuracy and fairness.
3. Connecting the system to actual Learning Management Systems (LMS) for widespread implementation and instantaneous academic use.