**Project Proposal**

**Automated Essay Grading using Natural Language Processing Techniques**

Team Members: 1. Sai Phani Teja Chilukuri

2. Sai Yashwanth Reddy Gujjula

3. Naveen Bolla

4. Sumanth Dasari

# Project Background

According to Shermis et al. (2013), Automated Essay Grading (AEG) is the process of evaluating and scoring essays using Natural Language Processing (NLP) techniques. The use of AEG systems is becoming increasingly popular in education and assessment as it allows for efficient and objective grading of essays (Yannakoudakis et al., 2011). Traditionally, grading essays has been a time-consuming and subjective task, with the grading process often varying from one grader to another. The use of AEG systems can significantly reduce the time and resources required for grading essays, as well as provide more consistent and objective results.

The development of AEG systems requires the use of NLP techniques, such as text classification, text summarization, and sentiment analysis. These techniques allow for the extraction of relevant features from the text, which can be used to evaluate the quality and content of the essays. The use of machine learning algorithms can then be employed to train the AEG systems to accurately grade essays based on these features.

There has been extensive research on the use of AEG systems in education and assessment, with various approaches and algorithms proposed for achieving accurate and reliable grading (Liu et al., 2019; Yannakoudakis et al., 2011; Shermis et al., 2010; Taghipour & Ng, 2016). However, despite the potential benefits of AEG systems, there are still concerns about the reliability and validity of the grading produced by these systems (Liu et al., 2019). Shermis et al. (2010) argue that AEG systems cannot accurately capture the complexity and nuances of human language, and therefore cannot provide accurate and fair evaluations of essays.

## Project Proposal

Despite the success of automated essay grading, there is still room for improvement. One limitation of current approaches is that they are often based on hand-crafted features, which may not capture all the relevant information in the essays. Deep learning approaches, which can learn representations directly from the text, have shown promise in other NLP tasks and may be able to improve the accuracy of automated essay grading.

In this project, we propose to use deep learning techniques for automated essay grading. Specifically, we will use a neural network model that can learn representations of the essays and predict their corresponding scores (Liu et al., 2019). Our goal is to achieve a high level of agreement with human graders while also improving the efficiency and scalability of the grading process.

## Project Aims and Objectives

The primary aim of this project is to develop an automated essay grading system that can accurately and efficiently grade essays using natural language processing techniques. To achieve this aim, we have the following objectives:

* Preprocess the ASAP-AES dataset to extract the essays and corresponding scores.
* Develop and train a machine learning algorithm that can accurately grade essays based on relevant features extracted from the text.
* Implement NLP techniques such as text classification, text summarization, and sentiment analysis to extract relevant features from the essays.
* Evaluate the accuracy and reliability of the developed AEG system using various metrics such as Mean Absolute Error (MAE) and correlation coefficients.
* Compare the results of the developed AEG system with human graders to evaluate its effectiveness.

## Rationale and Goals of the Project

Automated essay grading has the potential to revolutionize the education system by providing faster and more consistent feedback to students. It can also reduce the workload of human graders and make the grading process more efficient and scalable. However, the accuracy of automated essay grading is crucial for its success. If the system is not accurate, it may produce misleading feedback to students, which can harm their learning outcomes.

The rationale for this project is to address the limitations of traditional essay grading methods, which are time-consuming and subjective. The Proposed AEG system using NLP techniques, will provide a more accurate, reliable, efficient and objective method of grading essays.

Our goal is to develop an automated essay grading system that achieves a high level of accuracy while also improving the efficiency and scalability of the grading process. We believe that deep learning techniques can help us achieve this goal by enabling the system to learn representations of the essays directly from the text.

## Motivation

The motivation for this project is to develop an AEG system using NLP techniques that can accurately and efficiently grade essays. The use of such a system can significantly reduce the time and resources required for grading essays, provide a more consistent and objective results.

## Significance

The significance of this project lies in the potential benefits that an accurate and reliable AEG system can provide to education and assessment (Mehmood et al., 2019). Such a system can help educators and examiners to evaluate student performance more efficiently and objectively, while also providing students with more detailed feedback on their writing skills.

## Features

* The proposed AEG system will have the following features:
* The ability to accurately and efficiently grade essays using NLP techniques.
* The ability to extract relevant features from the text, such as grammar, syntax, and semantics.
* The ability to provide detailed feedback to students on their writing skills.
* The ability to improve the efficiency and consistency of grading essays in education and assessment.

## Project Dataset

For this project, we will use the essay with human graded scores available in Kaggle through: <https://www.kaggle.com/c/asap-aes/data>. The dataset contains over 12,000 essays written by students from grades 7-12, along with their corresponding scores given by human graders. The dataset also includes the prompts, which are the essay topics, and the rubrics used for grading the essays. Each essay is associated with a prompt and a corresponding rubric, which specifies the criteria for grading the essay.

The ASAP-AES dataset is widely used for research on automated essay grading. Various research studies have been conducted using this dataset, and it has been shown that automated essay grading can achieve comparable accuracy to human graders. One study showed that a machine learning model achieved an agreement rate of 91% with human graders on a subset of the dataset (Shermis et al., 2010). Another study compared the performance of various machine learning models on the dataset and found that the best model achieved a correlation of 0.86 with human graders (Taghipour & Ng, 2016).

The dataset is preprocessed and cleaned, and each essay is represented as a bag-of-words vector. We will further preprocess the dataset to convert the bag-of-words vectors into word embedding’s, which are dense representations of the words in the essays that capture their semantic meanings. We will then use these word embedding’s as inputs to our deep learning model.

# References

Shermis, M. D., Burstein, J., Elliot, N., Miel, L., & Foltz, P. (2010). Automated essay scoring: Writing assessment and instruction. Routledge.

Taghipour, K., & Ng, H. T. (2016). A neural approach to automated essay scoring. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 188-197).

Liu, X., Li, X., & Li, X. (2019). Automated essay scoring using machine learning techniques. Journal of Ambient Intelligence and Humanized Computing, 10(8), 3199-3213.

Mehmood, F., & Siddiqi, I. (2019). Automated essay scoring system based on deep learning. IEEE Access, 7, 166225-166239.

Shermis, M. D., & Burstein, J. (Eds.). (2013). Handbook of automated essay evaluation: current applications and new directions. Routledge.

Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011). A new dataset and method for automatically grading ESOL texts. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers (pp. 180-184).