Team 3: Dan Goldstein, Manisha Gupta, Teenaz Ralhan, Jiayi Xu

Project Proposal: Automated Essay Scoring System

1. Project Description

In this project, an automated essay scoring system is built to score essays as human expert graders. Because of the subjectivity and the fact that process is time consuming, methods of checking student's grasp of course material have become uncreative and now include multiple-choice or fill-in-blank tests. These limit the ability of students to express their understanding freely and emphatically. If the manual grading process can be relieved by an efficient and accurate automated grading system, essays and other creative methods could be frequently used to indicate academic achievements. We are introducing a model that forms word representations by learning the extent to which specific words contribute to the text's score and predict a final essay grade on a marking scale. Our model will typically use a large range of textual features that correspond to different properties of text, such as grammar, vocabulary, style, topic relevance, and discourse coherence and cohesion.

2. Dataset Description

Parameter	Mean	Standard Error	Median	Mode	Standard D	Sample Va	Kurtosis	Skewness	Range	Minimum	Maximum	Sum	Count
score	0.442028	0.002609051	0.5	0.6	0.297203	0.08833	0.082022	-0.44709	1.5	-0.5	1	5735.75	12976
Sentsense	12.85621	0.09805932	9	6	11.17059	124.782	4.12059	1.756662	96	0	96	166835	12977
Syllable	304.4283	2.108792424	224	108	240.2265	57708.79	1.573518	1.382231	1583	1	1584	3950566	12977
Words	222.8323	1.542563057	164	133	175.7236	30878.78	1.942147	1.463374	1061	2	1063	2891695	12977
FleschIndex	63.61719	0.382603012	69.52	68.1	43.58485	1899.639	113.1457	-9.07845	1077.61	-958.42	119.19	825560.3	12977
Readibility	13.24609	0.180281639	10	9.3	20.53708	421.7716	128.6746	9.890753	515.8	-4	511.8	171894.5	12977

A detailed description of the columns contained within the dataset can be found at: https://www.kaggle.com/c/asap-aes/data. The table above shows the summary statistics based on the parameters from the dataset.

3. Delegation of Tasks

Name	Tasks
Dan Goldstein and Manisha Gupta	Find a dataset to work on Write Project Proposal Data Exploration Find and code suitable machine learning techniques (first half) Make Presentation and Write Final Report
Teenaz Ralhan and Jiayi Xu	Find a dataset to work on Write Project Proposal Data Cleaning Find and code suitable machine learning techniques (second half) Make Presentation and Write Final Report

4. Potential Methodologies

Some of the methodologies that we could use for this project are: KNN, Neural Networks, Long Short Term Memory, and Clustering, which are types of Supervised Learning and Unsupervised Learning utilizing the TensorFlow and Keras packages that come with Python.