**Student Writing Score Prediction using Linear Regression**

**Rohith Raj Boge**

**Problem Statement and Task Definition:**It is critical to evaluate and comprehend the aspects that affect a student's academic performance in the context of education, especially their writing abilities. The challenge is to create a predictive model that can forecast a student's writing grade based on a variety of characteristics and traits, providing important insights into the variables influencing writing performance. To improve student outcomes, this model seeks to be a useful tool for educators, managers, and policymakers when making judgment calls and implementing interventions. The task at hand involves building a predictive model using input features as gender, ethnicity, parental level of education, Lunch type, test preparation course and math, reading score to predict student writing scores as output.

**Dataset:**The dataset is from Royce kimmons: Understanding digital participation divides:(<http://roycekimmons.com/tools/generated_data/exams>). The dataset consists of the following features: It has 1000 rows and 8 columns. The first step in building the system is preprocessing of the data, which involves handling the missing data, encoding categorical variables, and scaling or normalizing numerical features. The next step involves identifying which elements are relevant for predicting student writing using techniques like correlation analysis. Then the linear regression model is implemented using the selected features. The data set is split into a training set and testing set for model training and evaluation. The trained model is used to make predictions on testing data to estimate the student writing score.

**Evaluation metric:**choosing a metric that places equal weight on all errors may be more appropriate for this dataset, rather than minimizing the overall error. To Evaluate the model’s performance on the testing data the evaluation metrics such as mean squared error, mean absolute error, or R-squared error are used. These evaluation metrics work well for regression problems.

**Challenges:**1. Handling categorical data (gender, ethnicity, parental education) effectively.  
2. using correlation techniques to identify relevant features for predicting student writing score.3. Balancing model complexity with predictive accuracy.

**Related Works:**  
“Predicting Student performance: An Application of Data Mining Methods with an Educational Web-Based System” by Romero and ventura (2007)   
"Predicting Student Performance in Higher Education: A Systematic Literature Review" by Samarakoon and Fernando (2019)   
"Predicting Student Success Using a Simple Neural Network" by Dong et al. (2017)