CSCE 606 Software Engineering, Fall 2020

**Team MarsshMellow: Iteration 2 Project Report**

# Customer Meetings

* November 2, 2020 from 5:00 pm to 7:00 pm – Progress tracking and Feedback:

The customer informed us about their preference to focus on implementing a decent machine learning model over the GUI and web application.

* November 9, 2020 from 5:00 pm to 7:00 pm – Feedback for obtained results

The prediction results of our ML model for grade V was presented. The customer was satisfied with the performance.

Further the remaining scope of tasks for the project was discussed:

1. Compare the performance of our ML model with the existing human prediction.
2. Provide a certification for schools using the scores predicted from our ML model.
3. Form bubble groups - categorizing kids based on their projected progress.

Amongst these only (i.) is a deliverable for this semester.

* Both the meetings were held over Zoom.

# Implementation of User Stories

A total of 2 user stories were implemented in this iteration.

## Data Preprocessing in ML Model

🔗 [[Pivotal Story](https://www.pivotaltracker.com/story/show/175383847)] [[Source Code](https://colab.research.google.com/drive/1ivdVUUG8yn-VUsrAZQpsFm5LGkwuwN7T?usp=sharing)]

User Story:

Preprocess the student data mock test scores and final test scores for the machine learning models to predict the student scores.

Details:

* The data was available as a set of separate csv files. Moreover, the schema for these files is different for each grade.
* The data had multiple types of scores which had to be understood from the information that the customer had given us.
* First the data was joined to get a single consolidated table with all the features and the scores to predict.
* Then the redundant features were dropped, missing data was imputed, and the data was processed for using in a ML model.

Status: In Progress

The development has been done for grade V.

## Grade V model

🔗 [[Pivotal Story](https://www.pivotaltracker.com/story/show/175631923)] [[Source Code](https://github.com/harshit08173/CSCE-606-TeamMarsshMellow/tree/main/DistrictSchoolPerformancePrediction/Machine%20Learning)]

User Story:

Develop individual models for 5th grade math and reading

Details:

* Using the preprocessed data, a basic ML model was developed to predict the final test scores.
* First, a Linear regression model was trained on the data.
* Then, we trained a random forest classifier using grid search for hyper-parameter tuning. We also used leave one out cross validation.
* The results were then converted in a presentable format along with calculating the derived results (Raw Score, Approaches, Meets and Masters categories) and delivered to the customer.

Results:

* Please find the file [“Results\_Grade5\_Predictions\_Final.xlsx”](Fall%202020/I2/Results_Grade5_Predictions_Final.xlsx) in the submission tarfile. This contains the predictions results for Grade V students for Maths and Reading subjects.

Status: Done

# User Interface

No new user interfaces were developed in this iteration.

# Submission related links:

* Pivotal Tracker

Our user stories have been entered into our public Pivotal Tracker project, which can be found here: <https://www.pivotaltracker.com/n/projects/2469696>

* Version Control

The GitHub repository being used for this project can be found here: <https://github.com/harshit08173/CSCE-606-TeamMarsshMellow.git>