Brad Pawlow, Sebastian Reid, Noah Slosberg

5/15/23 - 5/19/23

STAT 390

Progress Report 5

**Goals from Last Week:**

* Task 1, Continue working on Kaggle project, compensate for overfitting
* Task 2, Better understand what educational metrics are required for project
  + Literature analysis of issue, understand what current research says about topic
* Task 3, Remove new issues with row duplication found in data
  + Combine into multi-factor columns
* Task 4, Make early prediction model so we can understand what factors play the most into the availability of educational programs
* Task 5, Prepare for MCMF progress presentation in two weeks

**Progress/Accomplishments**:

* Completed Task 1, with slightly improved public score of 8.28852 and improvement on overfitting test results
* Made some progress on Task 2, we will provide analysis of various test scores, including ACT, and assess different academic categories within test scores to determine types of programs
* Completed Task 3 with multi-factored columns
* Made some progress on Task 4, we decided to focus more on inferential models to determine which areas could benefit from increased program availability
* Began Task 5, started developing and mapping out our slide orders for the presentation

**Problems/Challenges**:

* We are still in the process of understanding which educational metrics would be best to understand the problems in the CPS system. Many metrics can be biased and are unreliable, so we are reading the necessary literature to better understand this problem.
* Our Kaggle score seems to have plateaued and we are struggling to find new ways to improve the accuracy of our models
* Our ETL process for external datasets is taking a while to complete and integrate, such as integrating additional test score data and cleaning that data.

**Plan/Next Steps**:

* Continue working on Kaggle project to find alternative methods for improving the overall score (better PCA, feature selection, different ensemble variations)
* Better understand what educational metrics are required for project
  + Literature analysis of issue, understand what current research says about topic
* Prepare for MCMF progress presentation next week by adding more slides
  + Create more comprehensive visualizations
* Select additional appropriate evaluation metrics for Kaggle competition and robust validation techniques