Just-in-Time Intervention

Hosuk Choi, Matt Yang, Mia Zhao, Attush Dhakal,
Aashay Amin, Saurabh Chatterjee, Rachel Paul

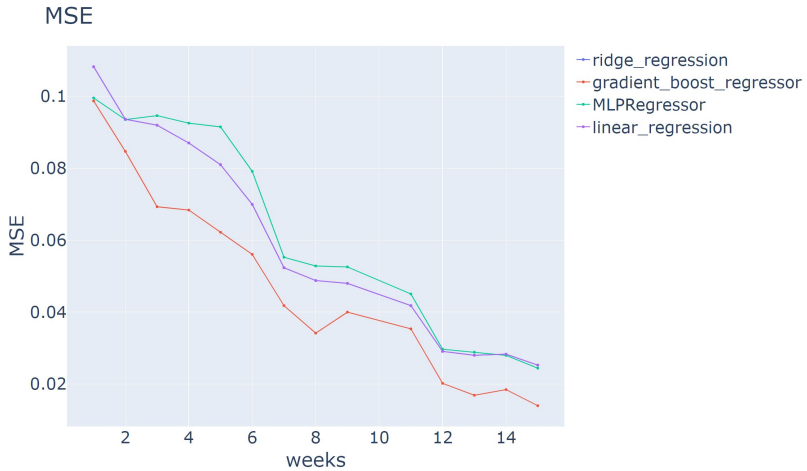
Introduction

Background of Project

- Project inspired by an analysis conducted by Gardner and Brooks (UMich) in 2018, who laid the framework for attempting to predict student success in MOOCs (https://arxiv.org/abs/1711.06349)
  - In their paper, they used raw data sources from many MOOCs platforms.
    - Clickstream data
    - Assignment data
    - Forum data
  - We have access to many similar datasets via Georgia Tech's numerous online courses
- Our team's ultimate goal is to not only gain insight into what drives student success, but also to be able to easily present our results to professors and students on a class-by-class basis

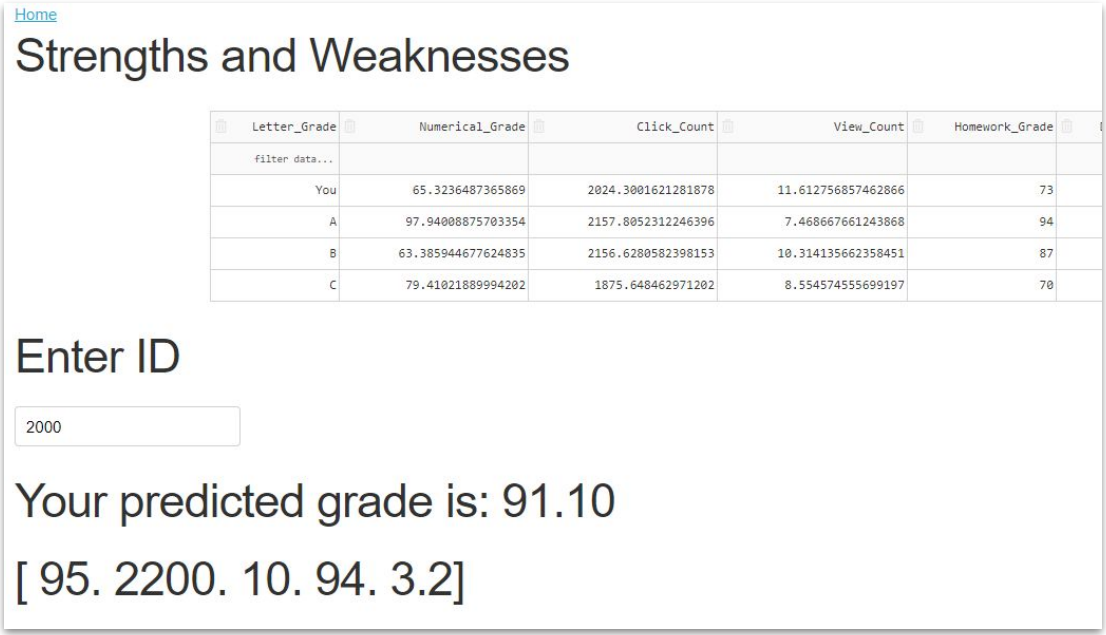
Past Semester Progress (1)

- Created models to predict final student grades for ISYE 6501. Best performing model achieved an MAE of 15 pts at the midterm.



Past Semesters Progress (2)

- Created proof-of-concept web-application with prototype student-facing features
- Uses dummy data



This Semester Goals

- Split into two-teams, simultaneously working
- Model-training team
  - Many improvements can be made to previous model training
  - A lot of cool ML techniques to explore, such as feature engineering, exploring new datasets, hyperparameter tuning, trying new algorithms
  - Working on extracting further insights from new data
- Web-development team
  - Turning the bare-bones application into a complete web-app
  - Connect to database, integrate model prediction
  - Evaluate best way to deliver results

Team Members

- Saurabh Chatterjee - Model Training
- Rachel Paul - Model Training
- Huimin Zhao - Model Training
- Aashay Amin - Web Dev (back end)
- Attush Dhakal - Web Dev (back end)
- Hosuk Choi - Web Dev (front end)
- Matthew Yang - Mixed (Project Manager)

Data Modeling Team

Previous Work Evaluation & Plan for this Semester

- Understand the problem space- Input Features & Target Output
- Evaluated & mapped data used in the previous semesters
- Got an understanding of what features were considered and the model performance.

Plan for this semester

- Explore a different data set to get an understanding of how various data sources can help JITI
- Evaluate **click-stream input features** and build a model to **predict percentage of course completed**.

Data Wrangling & Feature Engineering

- Explore the new data with clickstream input features provided in the excel format.
- Use python to clean data to remove null values and any outliers.
- Use python(or other tools) to find feature significance and decide on the best features to use.
- Convert the final feature to a dataframe for model development to predict percentage of course completed.
- Compare features from the past data and the new data to find a common trend in features that relates to JITI

Model Development & Other Goals

- Build models and use cross validation to evaluate performance.
- Tune hyperparameters using Grid Search
- Visualize the model predictions.

Other Goal(If model development is complete with the new data)

- Improve model performance of the previous semester data using more features and trying other algorithm
- Expose the existing model as REST API for consumption by App Team.

Web Development Team

Front-end

- Hook up with data in back-end
- Clean up presentation
- Clean up display of data/predictions
- Logins(?)
- Reformat look/style
- Restructure code (not just one app.py)

Back-end: What have we done so far?

- Created a basic SQL file to quickly initialize Postgres DB with dummy data
- Created online instances of PostgreSQL and Mongo databases (through ElephantSQL and MongoDB Cloud) with dummy data
- Learned how to connect to these databases with python libraries (psycopg2 and pymongo) and perform simple queries
- Created functions to abstract out querying to get a student's data using a student_id parameter

Back-end: Next Steps

What we are working on now:

- Creating a local dummy model to mimic taking in parameters and outputting a value
- Write a function that takes in a student_id and model and outputs a value that can be used on the frontend
- Change data in dummy DBs to be closer to actual DBs

Our goals for the semester:

- Connect web app with C21U DB
- Display student under risk factors and indicators
- Write functions to assist with query as needed by the Front-end
- Build email notification that triggers when specified criteria is met

Thank you!!