



Supporting Students with Data

By: Linpeng Chen, Donald Dinerman, Emily Du, Alan Zhu
Project Advisor: Peter Freeman



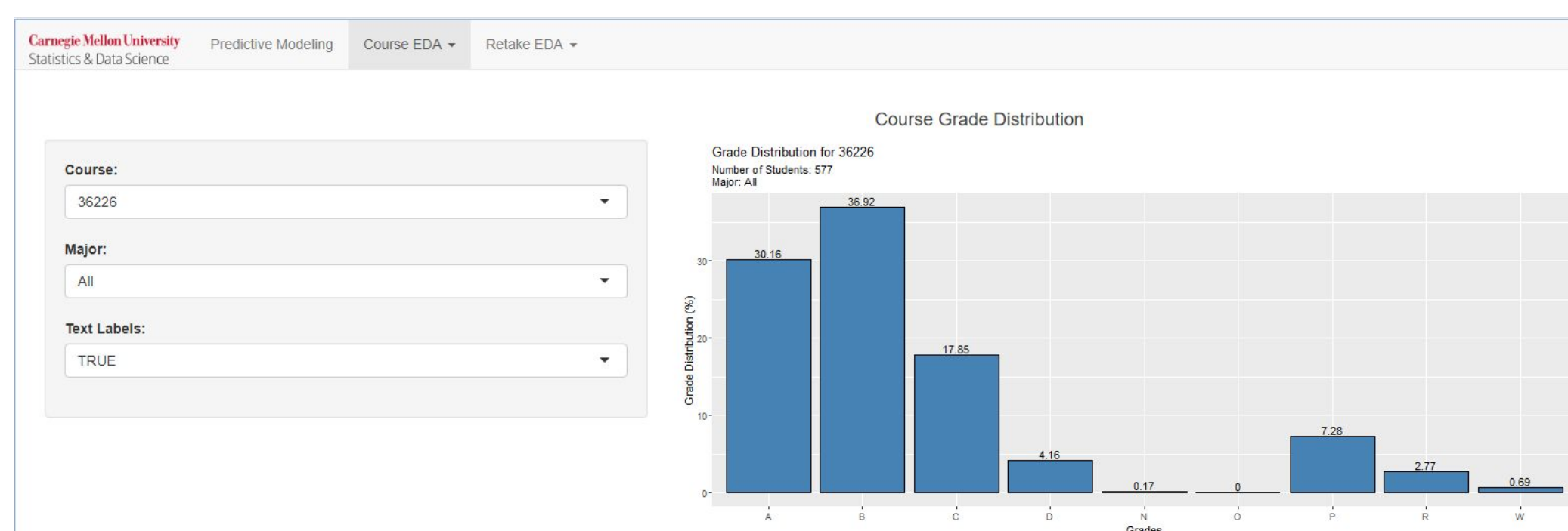
Carnegie Mellon University
Statistics & Data Science

Introduction

Students oftentimes wonder about past students' performances in a course when deciding which courses to take. Similarly, the advisors are interested in assessing whether a course is suitable for a particular student. To help address these concerns, our team built a Shiny App where advisors can examine past course performances and predict students' future performances in courses they are interested in taking.

Data

We use anonymized course performance data for Statistics & Data Science graduates from 2016 to 2020. These data include a student identifier, class level, major information, courses taken, and final course grades. Our exploratory data analysis includes course grade distributions, comparative course performance, and course retake performance. Although the app's interactive features allow exhaustive toggling for majors and courses, to demonstrate the features of our Shiny app, we focus upon 36-226, *Introduction to Statistical Inference*.



We can display the grade distribution for 36-226 grouped by specific majors to compare performance across different majors.



We can visualize the association in course performance across courses. Here, we display the grade association between 36-226 and its prerequisite, 36-225.



We can display the last grade in a retaken course versus the first grade. Here, we see that student performance noticeably improves when 36-226 is retaken.

Predictive Modeling

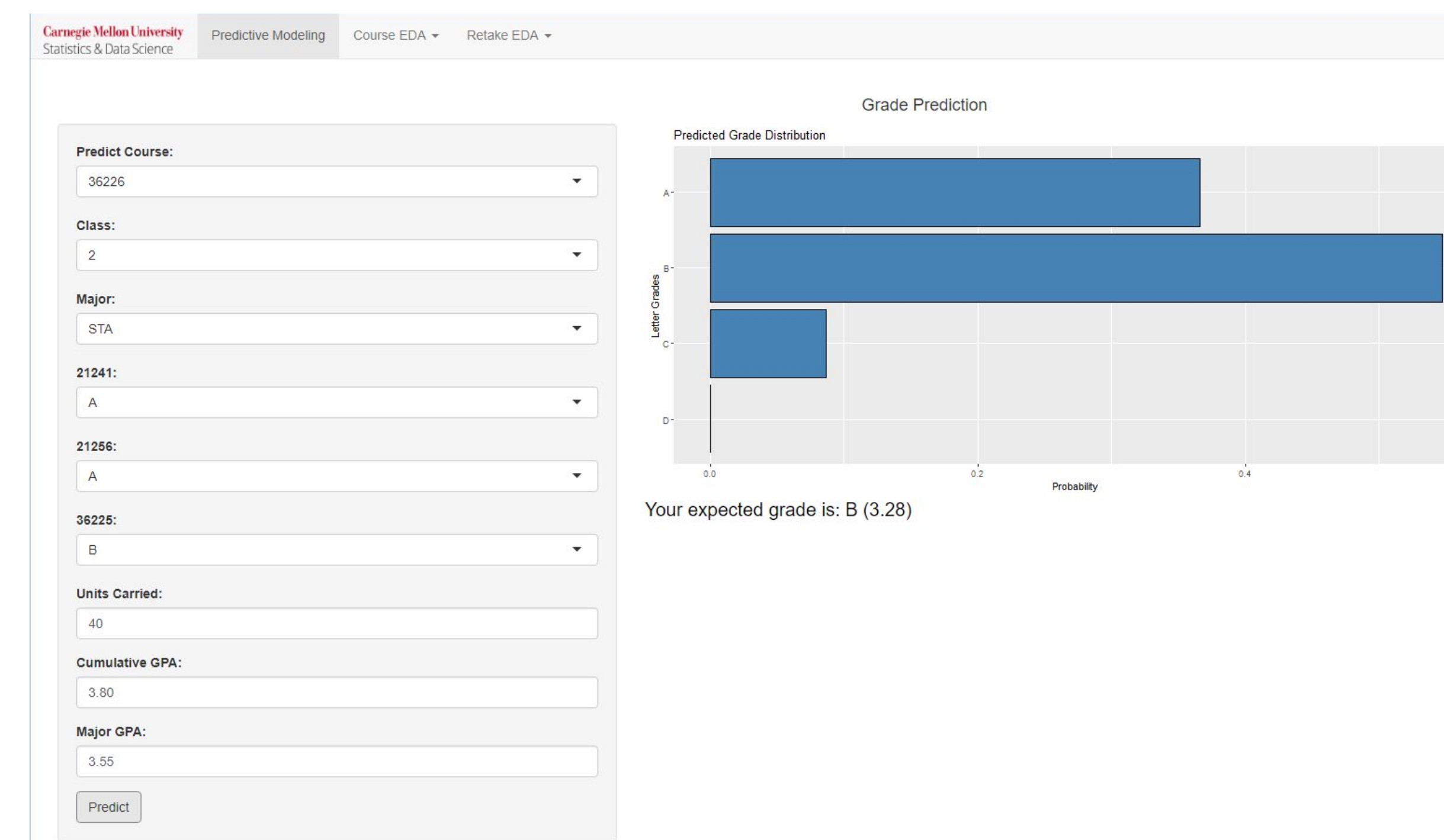
We utilize random forest models to predict future performance across six departmental courses using students' academic history information. Our model outputs a probability distribution of course performance along with a most likely course outcome.

Predictor Variables

- Class Year
- Major (Stat, EconStat, StatML)
- Major GPA
- Units Carried
- 2 or 3 Prerequisite Courses
- Cumulative GPA

Response Variable

- Departmental Courses (15-122, 15-351, 36-225, 36-226, 36-401, 36-402)



On the left side, users select the predicted course and input the explanatory academic information.

After clicking the predict button, the grade distribution graph with an expected average grade are output.

Conclusion and Limitations

We successfully designed a Shiny App with predictive modeling and EDA capabilities. Aside from being an addition to advisors' toolkit, our predictive models allow advisors to discern impactful factors associated with course performance. Furthermore, advisors can update our predictive models and data visualizations with new student data to get the most up-to-date insights on the Statistics and Data Science cohort.

Our data contain mostly StatML students, allowing us to create predictive models tailored to the StatML curriculum (e.g., 15-122, 15-351). However, we do not have the sufficient sample size to build models for Stat- or EconStat-specific courses. Also, the numbers of grades that are C and below are relatively small, resulting in overly optimistic predictive models that favor A's and B's over lesser grades.