

A Question Selection Strategy for Early Warning Systems

By Victor Zeng, Dr. Nouhad Rizk

UNIVERSITY of
HOUSTON

DEPARTMENT OF COMPUTER SCIENCE

Introduction

Early warning systems are systems to identify students at risk of failing a course. These systems allow educators to perform targeted interventions, thereby increasing student retention. One of the primary ways in which an early warning system collects data about soft factors is by identifying worthy students attributes. By selecting the optimal subset of questions to ask, a university can save both time and resources while increasing student engagement with the early warning system.

The purpose of this study is two fold : first to find out whether implementing an early warning system based on non academic students attributes provides a decent level of prediction accuracy of student progress; and second whether applying some data analytics techniques can produce optimal attributes.

Methodology

Step 1

Develop question set about soft academic factors
Establish baseline level of prediction accuracy with previous grade data



Step 2

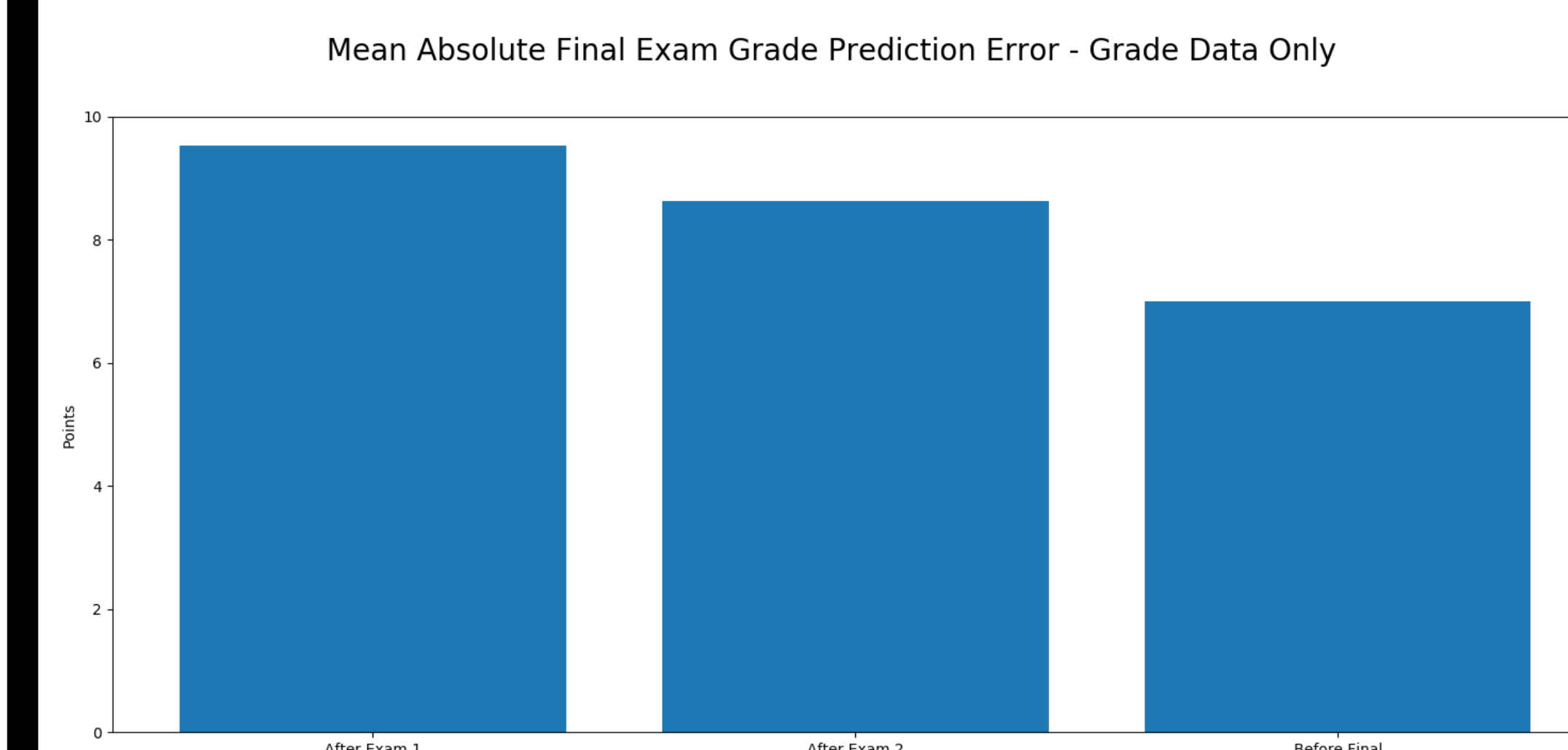
Send out survey to fall Data Structures class with questions from developed question set.
Combine survey responses with assignment grades to form dataset



Step 3

Perform mutual information analysis on collected data.
Perform offline experiments on collected data to explore mutual information and PCA based approaches to question selection

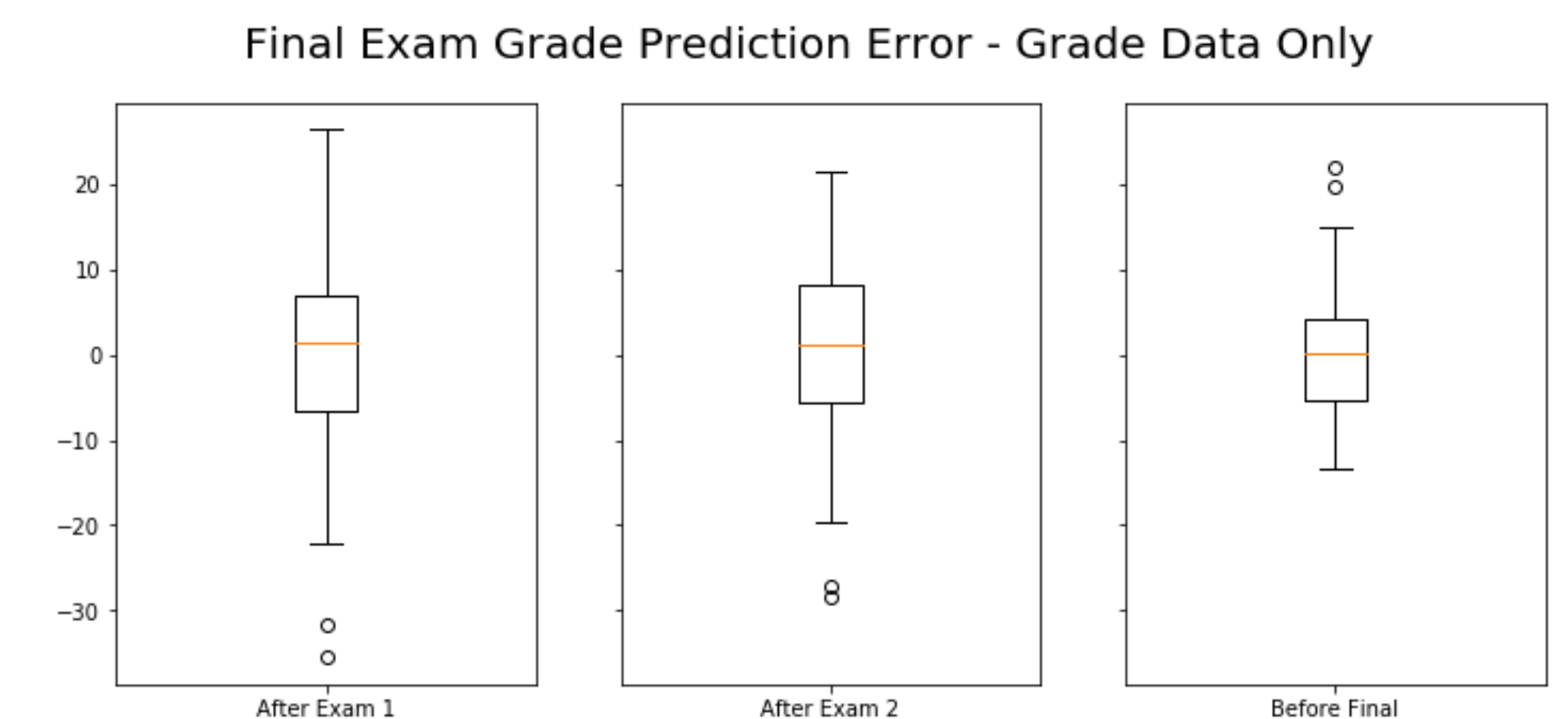
Baseline Results



On average, as students progress through the course, the grade predictions become increasingly accurate. When you include assignments between Exam 1 and the final exam, the mean absolute grade prediction error decrease from 9.5 to 7.0 points.

There does not appear to be a significant bias in the grade predictions.

While the grade predictions for most students was fairly accurate (within 10 points), there were several students whose prediction was off by more than 20 points. This was true even when the algorithm had grades for all assignments before the final.



Discussion and Future Work

Despite only covering a third of the material in the Data Structures course, students' grades on assignments up to Exam 1 were a surprisingly decent predictor of their final exam grade. This suggests that grade data also conveys information about factors beyond the student's knowledge. Likewise, the high variance in the prediction accuracy may be simply a result of the small dataset size, or it could also suggest the existence of some additional factor which the grade-only model did not account for. This study has demonstrated the need for a focused questionnaire identifying attributes which affect student performance. Future work is needed to evaluate the effect of such attributes on an early warning system or on a reliable recommender