

Mini Project 2 – Students Performance

Hektor Dahlberg (41685)

October 14, 2024

Contents

1	Introduction	3
2	Data Processing	3
2.1	What can you tell about the data?	3
2.2	What features do you choose and what is your method of choice? .	3
3	Data analysis	4
4	Conclusion: What were the bottlenecks? How did you overcome them?	4

1 Introduction

The aim of the mini project is to predict the students final grade from the data set given.

2 Data Processing

2.1 What can you tell about the data?

The data can be divide in to two categories 1.quizzes ,mini project and peer reviews 2. course logs on how the student has been active on the course web page. I decided to look more into the correlation between the final grade and activity on the course web page. From my plots of cumulative mean for the different stats you can see that there is a correlation in between being active on the course page but the difference between grade 3,4,5 is small while a grade of 0 or 2 should be easier to identify.

2.2 What features do you choose and what is your method of choice?

The features i chose where the features with an correlation higher than 0.5 and drop the feature "Week8_Total", I tried using only the data from Status0, Status1,Status2 and Status3 but noticed that the models had an very low accuracy so decide to also use the data from quizzes, mini projects and peer reviews.

3 Data analysis

I used two confusion matrices one for each model that was trained to see what the grade vs predicted grades are. From this you can observe that both models are capable of predicting the grade 0 with a accuracy of 100% but the rest of the grades are harder for the models to predict. This was as i expected as from the cumulative mean plots that where done earlier there is a much bigger difference between the grade 0 and the rest of the grades.

4 Conclusion: What were the bottlenecks? How did you overcome them?

Bottlenecks for me was to decide on the features to keep and use for the models, I still would have liked to only use the logs to try to predict the grade but the small amount of data makes it hard to predict anything which is not a grade 0 from only the logs.