Mini Project II – Students Performance

Project description
The data was collected from a fully online nine-week-long course on machine learning, hosted on the online learning management system Moodle. The goal of this project is to use two different supervised learning approaches (of your choice) to predict students' final grade in an online course.

Data description
The dataset contained anonymized information relating to 107 enrolled students. The data included students' grades (from 3 mini projects, 3 quizzes and 3 peer reviews and the final overall grade) as well as the course logs. The deadline for the three mini projects fell within weeks 3, 5 and 8 of the course, whereas the deadline for the quizzes fell within weeks 2, 4 and 8.

- Status0: course / lectures / content related (Course module viewed, Course viewed, Course activity completion updated, Course module instance list viewed, Content page viewed, Lesson started, Lesson resumed, Lesson restarted, Lesson ended)
- Status1: assignment related (Quiz attempt reviewed, Quiz attempt submitted, Quiz attempt summary viewed, Quiz attempt viewed, Quiz attempt started, Question answered, Question viewed, Submission re-assessed, Submission assessed, Submission updated, Submission created, Submission viewed)
- Status2: grade related (Grade user report viewed, Grade overview report viewed, User graded, Grade deleted, User profile viewed, Recent activity viewed, User report viewed, Course user report viewed, Outline report viewed)
- Status3: forum related (Post updated, Post created, Discussion created, Some content has been posted, Discussion viewed)
- 9 grades (Week2_Quiz1, Week3_MP1, ... Week7_MP3)
- 36 logs (Week1_Stat0, Week1_Stat1, Week1_Stat2, Week1_Stat3, ... Week9_Stat0, Week9_Stat1, Week9_Stat2, Week9_Stat3)

Task description

Perform the following tasks and describe them in your report.

Step1- Data Processing

Are there any missing values? What are your features? Are you keeping every feature?

Step 2 – Data split: Training and Test sets

Divide your data in an appropriate proportion to train and test dataset.

Step 3 – Model Training

Train two models to predict students' final grade using two different approaches of your choice. How accurately does your models predict the students' final grade and how do these models compare against each other? Which one is better? Could you explain why?

Step 4 – Performance Evaluation

How are your models performing? Do you need to change anything in your model? Visualize the performance of your models.

Step 5 – Important features

What are the three most important features in predicting students' final grade?

Reporting

Write a scientific report which includes:
- Introduction (What is the problem you are solving?)
- Data processing (What can you tell about the data? What features do you choose and what is your method of choice?)
- Data analysis (How did you visualise the data and how do you interpret the data visualisation? What are the interesting observations you could report?)
- Conclusion (What were the "scientific" bottlenecks? How did you overcome them?)