

SQUAD Datathon

Additional Information

April 3, 2024



THE UNIVERSITY
of ADELAIDE

1 Feature Information

- sex – Male or Female.
- age – Age of the student.
- address – Address of the student. It is either “Urban” or “Rural”.
- family_size – Rough idea of the number of members in the family. It is either “Greater than 3” or “Less than 3”.
- parents_together – Whether the student’s parents live “Together” or “Apart”.
- mother_job – Mother’s job.
- father_job – Father’s job.
- guardian - “Mother”, “Father”, or “Other”.
- travel_time – Time it takes to travel to uni in minutes.
- study_time – Time that the student spends on studying. Here’s the mapping for this feature:
 - 1: Less than 2 hours
 - 2: 2 to 5 hours
 - 3: 5 to 10 hours
 - 4: More than 10 hours
- failed_classes - Number of classes failed last year.
- school_support – Whether the student receives additional education support from the school or not.
- extra_curricular – Whether the student is involved in extra-curricular activities or not.
- want_higher – Whether the student wants to go for higher education or not.
- internet – Whether the student has internet access at home or not.
- romantic_rel – Whether the student is in a romantic relationship or not.
- family_rel – The quality of student’s relationship with family.

- `free_time` – Whether the student has free time after school or not.
- `go_out` – Whether the student goes out with friends or not.
- `workday_alcohol` – Whether the student consumes alcohol on working days or not.
- `weekend_alcohol` – Whether the student consumes alcohol on the weekend or not.
- `absences` – The number of absences taken by the student.
- `grade` – The grade earned by the student.

Additionally, any feature having unique values same as the list [1, 2, 3, 4, 5] represents a scale where 1 means “Very low” and 5 means “Very high”.

2 Additional Information

- You must submit both the report and code file (Python notebooks are preferable) by 3rd April, 11:59 pm.
- The code needs to have a code block at the end for evaluation. This code block should use the model to predict on the test data, calculate the RMSE score and print the score. Here’s a quick example in python:

```
predictions = model.predict(X_test)
# calculate_rmse is a placeholder
rmse = calculate_rmse(predictions, y_test)
print(f\The RMSE score is: {rmse}\)
```

- Along with the evaluation function, you must provide a preprocessing function that performs all the preprocessing that you’ve done on the test dataset, be it removing features, removing certain values, changing data types, encoding data, scaling data, and so on, your preprocessing function should convert raw data (like the train.csv file) into processed data ready for modelling. This function will return the preprocessed data. Write this function just before the evaluate function. This is required for R as well.
- Make sure you add comments to your code since that’ll influence the points you receive for your code.

- Do not keep the data file in the same directory as your code. Create a folder called "data" and keep the "train.csv" file in there. In your code, remember to use the absolute path to the file.
- All of the above points are to be followed irrespective of the language (Python or R) you use. If the model that you trained has a predict method,