A Report

on

Applying Regression Algorithm   
on HCI Quiz score data

By:

Manasi Kattel  
CE 3rd year

## Introduction

HCI quiz score data is the data collected from moodle system of the students of Department of Computer Science and Engineering, Kathmandu University of the course Human Computer Interaction (HCI). The data initially consisted of the attributes : Surname, first name, institution, department, email address, state, started on, completed on, time taken, grade and grade for each question.

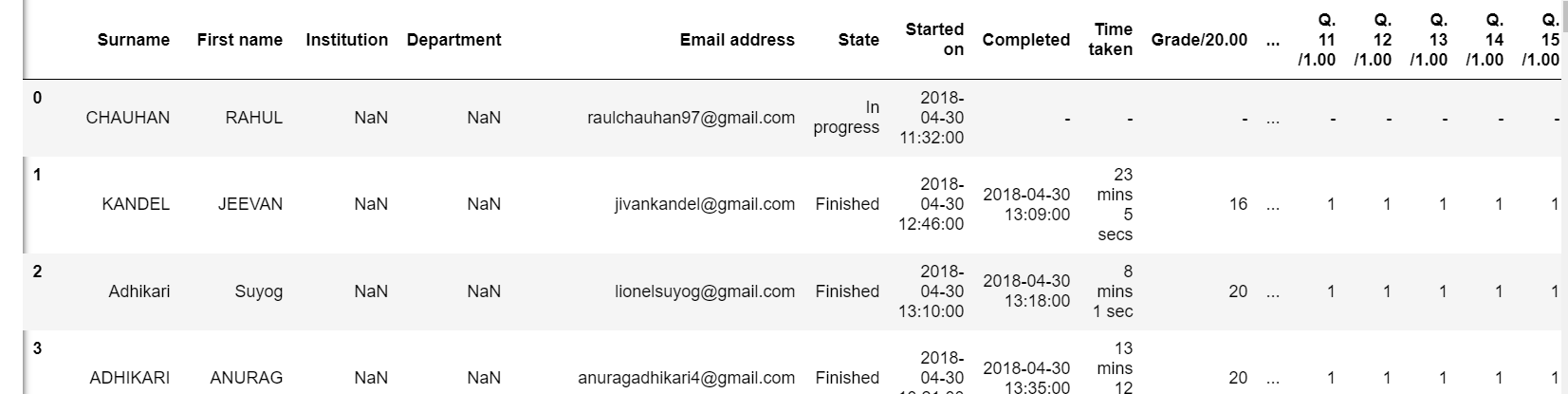


Fig 1: Quiz-score data

Methods

To develop a predictive model, linear regression was applied in the above data using library scikit learn.

To apply regression, we need dependent and independent variables. Here, in the above dataset, we are trying to see if the grade scored by a student depends on time taken to complete the quiz. So, we set time taken as an independent variable and grade scored as dependent variable.

The time feature is of format “23 mins 5 secs” as shown in above table. But we need numeric value in order to apply regression, so we convert all times to minutes as:

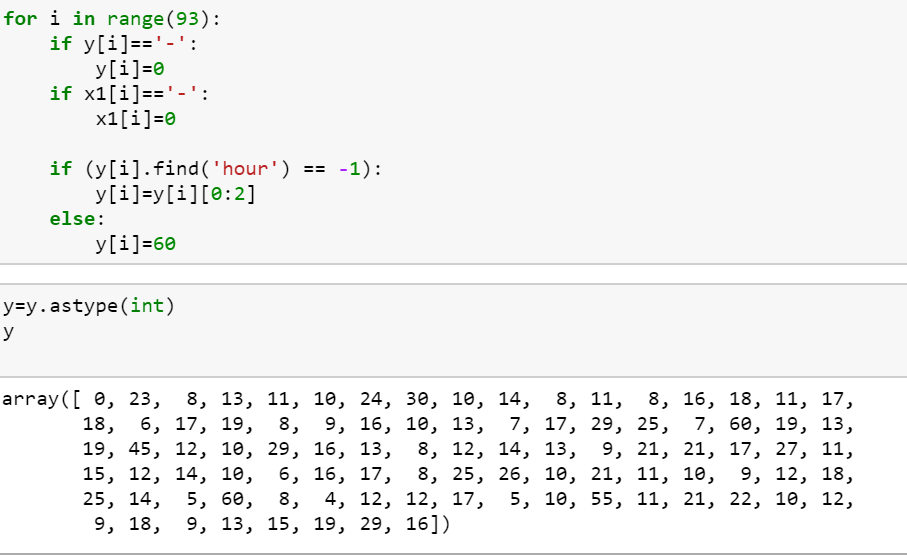


Fig 2: Extracting minutes from string

The data is now split into train set and test set. The training set is fitted to the model. The testing set is for prediction and error calculation.



Fig 3: Splitting data into training and testing set

Finally, linear regression is applied.

## Result

The result of applying linear regression are :-

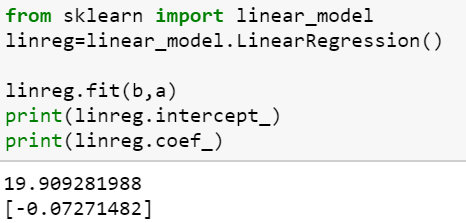
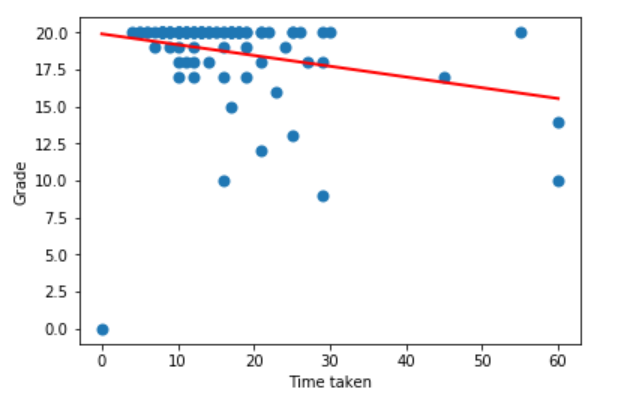


Fig 4: Result of linear regression

From the above figure we see that the intercept of the predicted line is 19.909281988 and the coefficient is -0.07271482. This is a line with the equation y=-0.07271482x+19.909281988. We can plot this line along with the actual points to see what our prediction looks like.

  
Fig 5: Linear model developed

To check the error in prediction, the training set is used as:-

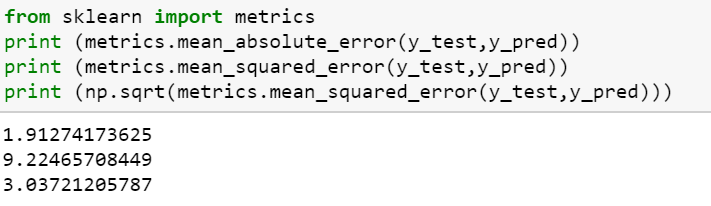


Fig 6: Error in prediction

## Discussion

From our prediction, it can be concluded that grade scored has an inverse relation with time taken to complete the quiz as the coefficient is negative. It can be interpreted that students who take less time to complete the quiz score better which could mean that they are intelligent enough to complete the quiz correctly in short time. But our prediction is not very accurate as there is error in prediction. The mean absolute error is 1.91274173625. This means that in every prediction of score, the predicted score could be greater or smaller by 1.9.