

Prediction of Intellectual and Personality Traits of Students using Machine learning and Deep learning Algorithms

S.Thanganila¹, Mrs.P.Tamilzhchelvi

¹ Second year PG Computer Science Students

² Associate Professor,

^{1,2} Department of Computer Science, Ayya Nadar Janaki Ammal College, Sivakasi.

Email : ¹thanganila54@gmail.com, ²tamilindhu@rediffmail.com

Abstract : Nowadays, Educational data mining and deep learning are tried by the educators to deal with the academic challenges of the students. The main objective of this project is to predict the student's intellectual performance and personality behaviour using machine learning algorithms. Once the assessment is known, it can further be used to improve their academic performances. Separate questionnaires need to be prepared to build a model based on set of questions answered by the students. The model which is regression model will be constructed based on the train data and the model's accuracy is checked by the test data. Entire process is done using two algorithms, regression algorithms (simple and multilinear) and deep learning algorithm and the results are used to identify the strengths and weaknesses of any student at the entry level so that he would be monitored and guided to become a successful performer.

Keywords: Intellectual and personality trait data, Multilinear and Linear Regression algorithms, Google Cloud, Google Form, Keras regression algorithm

1. INTRODUCTION

Student unrest, Student retention, unemployability threaten the entire education system in our country and force the government to go for Accreditation, Ranking for Innovation etc. These exercises would bring fruits only if quality is ensured at the students and teachers level. This project is aimed to identify the academic and personal strengths and weaknesses of the students at the entry level itself so that their contribution to the they enrolled would be maximum. The required outcome is obtained using simple, multilinear regression algorithms and then using deep learning algorithm also. We use [7] Google cloud in order to work with huge volume of dataset. Google cloud allows to deploy our workload on virtual machines. Google Forms is a tool that allows collecting information from users via a personalized survey or quiz.

The response is collected and downloaded in csv format.

2. OVERVIEW OF THE PAPER

This paper uses a data set that contains a lot of information about the students' intellectual and personality related data. The data help us to identify student's personality and intellectual ability. This knowledge can be applied for better performance of students.

Problem Description

To improve the academic performance and personality of the students and analyze the accuracy of machine learning and deep learning algorithms.

3. ALGORITHMS USED

There are a number of prediction algorithms that have been proposed by several researchers in the field of

machine learning and deep learning. This research work deals with the prediction based regression algorithm and convolutional network based keras regression algorithm.

A) The Linear and Multilinear Algorithms

Linear and Multilinear algorithms are one of the simplest

supervised learning algorithms that solve prediction problem. Linear regression is used to predict the value of an outcome variable Y based on one or more input predictor variables X . The aim is to establish a linear relationship (a mathematical formula) between the predictor variable(s) and the response variable, so that, we can use this formula to estimate the value of the response Y , when only the predictors (X s) values are known. The aim of linear regression is to model a continuous variable Y as a mathematical function of one or more X variable(s), so that we can use this regression model to predict the Y when only the X is known. This mathematical equation can be generalized as follows:

$$Y = \beta_1 + \beta_2 X + \epsilon$$

where, β_1 is the intercept and β_2 is the slope. Collectively, they are called regression coefficients. ϵ is the error term, the part of Y the regression model is unable to explain.

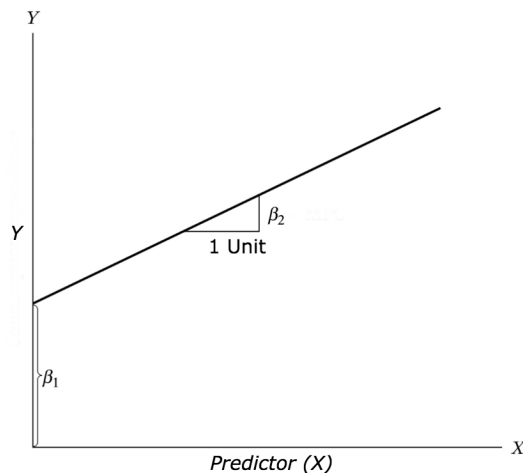


Figure 1 :Regression Diagram

Multiple linear regression is an extension of simple linear regression used to predict an outcome variable (y) on the basis of multiple distinct predictor variables (x). With three predictor variables (x), the prediction of y is expressed by the following equation:

$$y = b_0 + b_1*x_1 + b_2*x_2 + b_3*x_3$$

The “b” values are called the regression weights (or beta coefficients). They measure the association between the predictor variable and the outcome. “b_j” can be interpreted as the average effect on y of a one unit increase in “x_j”, holding all other predictors fixed.

B) Keras regression algorithm

Keras is an API used for running high-level neural networks. In deep learning, Keras regressor algorithm is used to build a model using train data and the trained model will be fit to our data using fit(). The Keras regressor algorithm is based upon convolutional neural network. A convolutional neural network (CNN, or ConvNet) is a class of deep neural networks. A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, RELU layer.

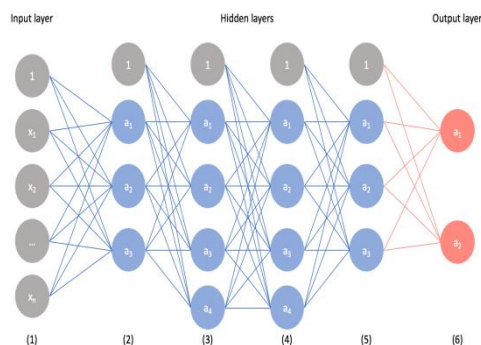


Figure 2 :Architecture of Convolutional neural network

4. IMPLEMENTATION

Before applying regression algorithm, the data will be split into train and test data. The train data is used to create a model and test data is used to predict the

new outcome. The model is built using lm(). After creating the model, the accuracy is predicted using forecast package. In Deep learning, keras regression algorithm is used to build a deep learning model. The deep learning model is based on convolutional neural network. It has input layer, multiple hidden layers and output layer. In deep neural network, the compile function is used to compile the model. The fit() is used to validate our data using the test data. The accuracy is found using fit().

A) Properties of the dataset

Separate intellectual and personality trait questions are framed through google form and sent to the students and their responses are downloaded in csv format. The few of the properties available in our dataset are as follows.

Intellectual dataset

Dataset size :106KB

Name	Name of the student
Scl/Cldistance	distance of scl or college
Studyhour	hour to spend study
Language	languages known
Parentsencourage	parents encourage to student
Furtherstudy	Further study of student
Target	Outcome of dataset

Personality dataset

Dataset size: 264KB

Gender	Gender of the student (Male, Female, Others).
Initiateconversation	If students start the conversation in any situation.
Socialactivity	The student's interest to any social activity.
Favpersonality	Anyone favorite personality of the student.
Favcolor	Favorite color of the student.
Target	A personality trait has 5 outcome values.

B) Platform preparation

- Step 1 : create an account in google cloud platform using gmail id. (for large data set)
- Step 2 : create Rstudio server pro instance.
- Step 3 : After creating the Instance, deploy the instance using deployment manager.
- Step 4 : store the intellectual and personality trait dataset using rstudiocssh.

Step 5: Use linear and multilinear regression in Rstudio. Login with username and password.
 Step 6: [8] Install keras in Rstudio instance using the following commands.

Devtool::install_github("/rstudio/keras")
Library(keras)
Install_Keras()

Step 7: In Rstudio, install the required packages.

Install.packages("")

The above function is used to install the packages.

Step 8: Retrieve data from rstudiossh using path of data storage.

Step 9: Split data into train and test using caTools package.

Step 10: Use lm() to create a linear and multilinear model.

Step 11: Plot the output using ggplot map package

Step 12: [9] Then finally find the accuracy using accuracy() available in forecast package.

Step 13: After finding the accuracy, implement the problem in deep learning using keras regression algorithm

Step 14: Declare a keras model, and build a model using model(), and fit model to our test data.

5. RESULTS

Personality trait outcome:

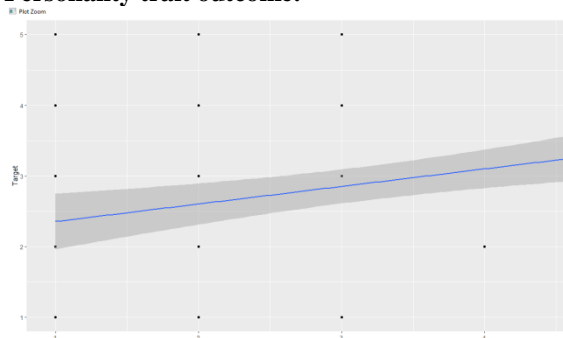


Figure 3 :Outcome of Simple linear regression

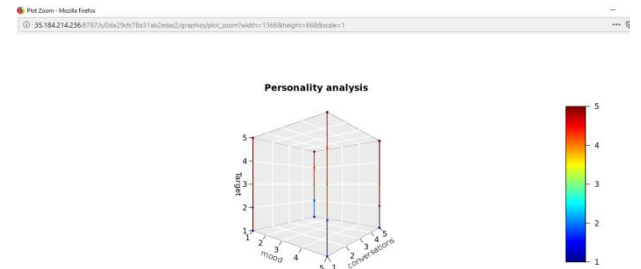


Figure 4 :Outcome of multilinear regression

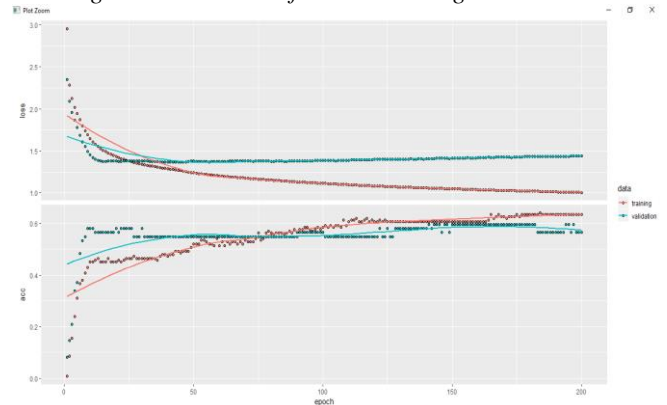


Figure 5:Outcome of deep learning

Intellectual trait outcome :

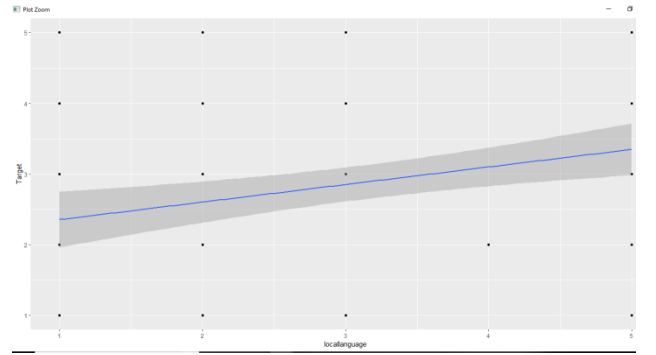


Figure 6 :Outcome of simple linear regression

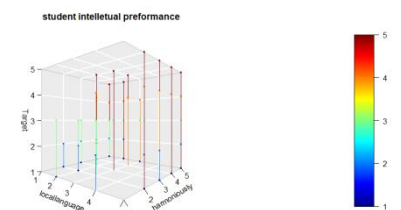


Figure 7 :Outcome of multilinear regression

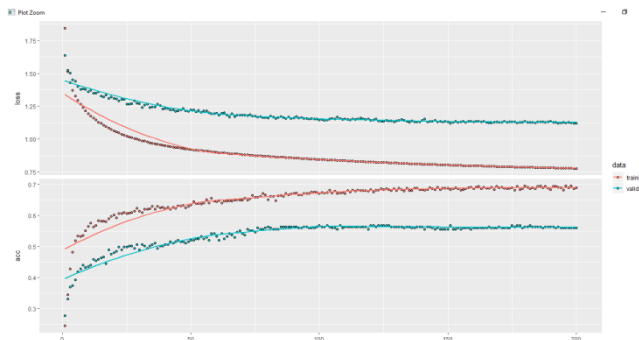


Figure 8 :Outcome of Deep learning

6. CONCLUSION

A deep learning and machine learning model are proposed in this project for predicting the students' personality trait and intellectual performance. In this project, I've analyzed the personality and intellectual performances of the students using linear, multilinear algorithms and found the accuracy. The machine learning algorithm has 45% accuracy of the outcome. Then, I've applied deep learning kerasregressor algorithm and found the accuracy of the outcome is 61%. I strongly believe that my work can be utilised in all academic institutions to identify the weaks and strengths of the students so that the weaknesses can be overcome and strengths can be further enhanced and exposed in different occasions so that every student would become a successful one.

REFERENCES

- [1] Bendangnuksung and Dr. Prabu P (2008) "Students' Performance Prediction Using Deep Neural Network" International Journal of Applied Engineering
- [2] Brunner, Lori A., (2010) "The role of personality traits and The Big Five in predicting future success of interior design students".
- [3] "Google Cloud Console" <https://console.cloud.google.com/compute/instances?project=mineral-highway-237014&authuser=5&instancessize=50&duration=PT1H>
- [4] "Datacamp" <https://www.datacamp.com/community/tutorials/keras-r-deep-learning>
- [5] "R-Blogger" <https://www.r-bloggers.com/convolutional-neural-networks-in-r/>
- [6] "datanovia" <https://www.datanovia.com/en/lessons/>
- [7] "sites.google" <https://sites.google.com/site/>
- [8] "CRAN" <https://cran.r-project.org/web/packages/>
- [9] "Medium" <https://medium.com/deeplearning>