A Mini Project Report on

Predicting Student Performance using Regression Analysis

Submitted in partial fulfilment of the requirements for the award of the degree of

**Bachelor of Engineering**

in

**Computer Engineering**

by

**Janhavi Anap (19102043)**

**Prathamesh Hambar (19102001)**

**Tejas Sheth (19102026)**

**Het Patel (19102005)**

Under the Guidance of

**Dr Pravin Adivarekar**

****

**Department of Computer Engineering**

A. P. Shah Institute of Technology G.B.Road, Kasarvadavli,

Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

**Academic Year 2020-2022**

## 

## 

## 

## Approval Sheet

This Mini Project Report entitled ***“Predicting Student Performance using Regression Analysis”*** submitted by ***“Janhavi Anap” (19102043), Prathamesh Hambar” (19102001), “Tejas Sheth” (19102026), “Het Patel” (19102005)*** is approved for the partial fulfilment of the requirement for the award of the degree of ***Bachelor of Engineering*** in ***Computer Engineering*** from ***University of Mumbai.***

Prof. Sachin H. Malave Dr. Pravin Adivarekar

Head of Department Guide

(Computer Engineering) (Computer Engineering)

Place: A. P. Shah Institute of Technology, Thane   
Date: 9th November 2021

## CERTIFICATE

This is to certify that the mini project entitled ***“Predicting Student Performance using Regression Analysis”*** submitted by ***“Janhavi Anap” (19102043), “Prathamesh Hambar” (19102001), “Tejas Sheth” (19102026), “Het Patel” (19102005)*** for the partial fulfilment of the requirement for the award of a degree ***Bachelor of Engineering*** in ***Computer Engineering,*** to the University of Mumbai, is a bonafide work carried out during the academic year 2020-2021.

(Dr Pravin Adivarekar)

Guide

Prof. Sachin H. Malave Dr. Uttam D.Kolekar

Head of Department Principal

(Computer Engineering)

External Examiner(s)

1.

2.

Place: A. P. Shah Institute of Technology, Thane   
Date: 9th November 2021

## Declaration

We declare that this written submission represents our ideas in our own words and where others’ ideas or words have been included, we have adequately cited and referenced the sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

———————————————

(Signature)

———————————————

(Janhavi Anap 19102043)

(Prathamesh Hambar 19102001)

(Tejas Sheth 19102026)

(Het Patel 19102005)

Date: 9th November 2021

## Abstract

The project focuses on predicting the final grade by analyzing the performance of the student. The prediction will be based on educational factors like cumulative GPA, personal factors like the number of hours spent by a student for studying and social factors such as the economical and educational background of their family and many more factors that may affect the GPA of the student.

It is important to find patterns in the student performance to be able to provide the necessary, accurate and timely diagnosis to the student. It also serves as a basic criterion for institutions to monitor the quality of education provided.

The Conventional Statistical Analysis and Artificial Neural Network prediction approaches are necessary for the prediction. Conventional statistical evaluations help in identifying the multiple factors that affect student performance. With these factors as input variables, an Artificial Neural Network is modelled. Artificial Neural Network helps in analyzing large datasets which are not easily simplified through conventional statistical techniques. It also helps to detect non-linear relationships between dependent and independent factors.

With the confirmation of the outputs from conventional statistical analysis, the training and testing of the model will be done for accurate prediction of the student performance. The performance of this neural network model is evaluated through various techniques. In this project one of the major techniques used for analysis is regression.

**Keywords:** Statistics, Analysis, Academics, Regression, Neural Networks

## Contents

[**Introduction**](#_ycdiz52rt6pm) **1**

[Problem Definition](#_t3v94tb24gt5) 1

[Objective](#_5bl7s8996jcd) 1

[Scope](#_n48y7y3aagcy) 1

[**Technology Stack**](#_j8utk4hi5281) **2**

[**Benefits and Applications**](#_wadbpq3jhr4x) **3**

[3.1 Benefits for Society](#_jk9arp4s78xe) 3

[3.2 Applications](#_p3ipjmlw0nkm) 3

[**Project Design**](#_mwzxbqy6k8r3) **4**

[4.1 Proposed System](#_74dn7x23hs07) 4

[4.2 Flow of Modules](#_5wobyg1phjc2) 4

[4.2.1 Prediction Model](#_w90jay7da65x) 4

[4.2.2 User Interface](#_hird9os17c7) 4

[4.3 Data Flow Diagram](#_dvrt1ip2kyvc) 5

[4.3.1 Level 0](#_xt81jucm828p) 5

[4.3.2 Level 1](#_ft7um5cgle7l) 5

[4.3.2.1 Prediction Model](#_o4vmwdkknygf) 5

[4.3.2.2 User Interface](#_2zqr05r8qwio) 6

[**Implementation**](#_s1n9m46yhupy) **7**

[**Annexure A**](#_9pstnbpe52gi) **9**

[6.1 Gantt Chart](#_3znysh7) 9

[**Bibliography**](#_ngck8m6umkya) **10**

[**Appendices**](#_2et92p0) **11**

[Appendix A](#_o08ue6nf8trm) 11

[Appendix B](#_xawvsnamshq9) 12

## List of Figures

4.1 Flow of Modules . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

4.2 Level 0 DFD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

4.3 Level 1 Prediction Model DFD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

4.4 Level 1 User Interface DFD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

5.1 Input Snippet . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

5.2 Output Snippet . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

6.1 Gantt Chart . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

## List of Abbreviations

DFD: Data Flow Diagram

## Chapter 1 Introduction

## 

### Problem Definition

To design an artificial neural network model that predicts the marks obtained by a student in their final semester via regression. The model considers factors such as previous academic record, attendance, socioeconomic background, etc.

### Objective

* The main purpose of the project is to analyze and improve upon the student results.
* The program will help professors to keep track of their weak students, who require extra attention.
* This project is a means which will eventually help the student and the professors to learn, analyze and grow from their past mistakes.
* Improve the success rate of schools/colleges, while monitoring students personally and having a clear picture of their goals and targets.

### Scope

The project will include the prediction of grades based on various factors. This will be implemented by training the data set on various machine learning models and selecting the best model that gives the maximum accuracy.

## 

## Chapter 2 Technology Stack

* **Python:** It is a General Programming Language used to code the model. It involves the usage of several libraries such as numpy, pandas, scikit-learn, etc.
* **VS Code/Anaconda:** Integrated Development Environment to run, train and test the ML model
* **Git and Github:** A Version Control System used for collaboration
* **Google Analytics/Tableau/Power BI:** Data Visualization Tool to visualize the dataset and gain some insights
* **Teamgantt:** Project Management Tool to improve team coordination and file sharing using Gantt chart.

## Chapter 3 Benefits and Applications

### 3.1 Benefits for Society

* The program will help professors to keep track of their weak students who require extra attention in their subject, thus bringing out the best in everyone.
* It will help the student, the professor and overall the society to learn, analyze and grow from their past mistakes.
* Predicting the overall success rate/ passing percentage of schools/colleges, in a way enhancing their reputation.

### 3.2 Applications

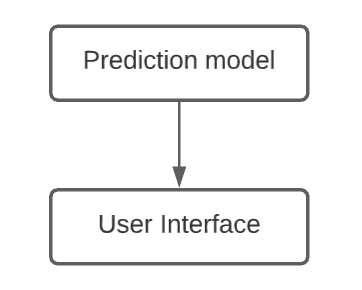
* It will enhance the Student Learning Experience,
* It can be implemented in any academic institution to monitor and analyze students,
* The algorithm, with the right parameters, can also be used to analyze employees in organizations and can be used for their welfare as well,
* Regression is also used in sports prediction for the improvement of players and coaches.

## Chapter 4 Project Design

### 4.1 Proposed System

In this project, student performance will be predicted based on various affecting factors. To predict the performance, first, data will be collected, cleaned, and processed. Then the data will be split into training and testing datasets. Various machine learning models like linear regression and decision tree classifier will be trained on the training dataset. The accuracy of the various models will be compared and the model giving the best accuracy will be selected. A web app will be deployed on Streamlit that will provide the interface to the user. The user will provide input and based on the trained model the output grade will be generated.

### 4.2 Flow of Modules

  
Fig 4.1 Flow of Modules

#### 4.2.1 Prediction Model

This module involves making the prediction model which includes processing the data, training and testing various models and saving the trained model.

#### 

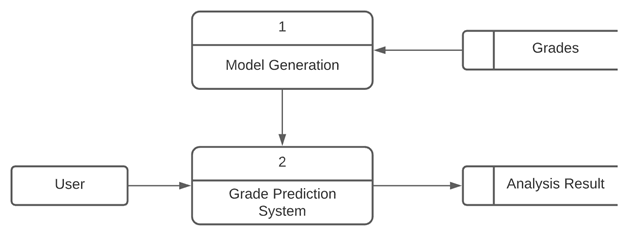
#### 4.2.2 User Interface

This module involves making an interface for the user to interact with the system

### 4.3 Data Flow Diagram

#### 

#### 4.3.1 Level 0

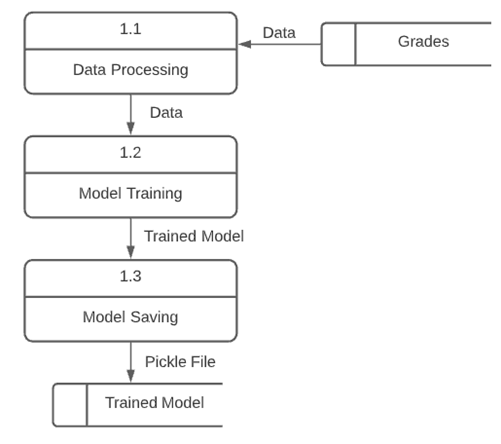
  
Fig 4.2 Level 0 DFD

#### 

#### 4.3.2 Level 1

##### 

##### 4.3.2.1 Prediction Model

  
Fig 4.3 Level 1 Prediction model DFD

##### 

##### 

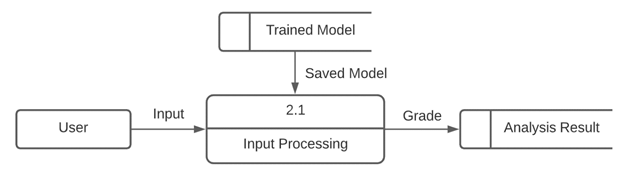
##### 

##### 

##### 

##### 

##### 4.3.2.2 User Interface

  
Fig 4.4 Level 1 User Interface DFD

## Chapter 5 Implementation

Input:

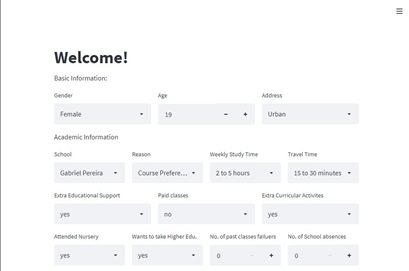
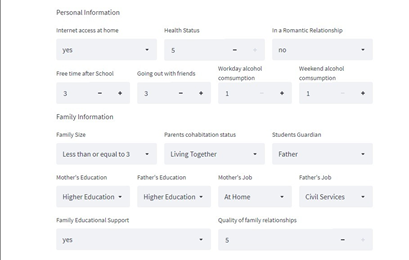
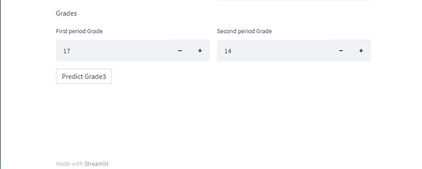
  
  


Fig 5.1 Input Snippet

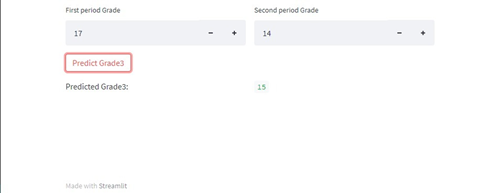


Fig 5.2 Output Snippet

## Chapter 6 Annexure A

### 6.1 Gantt Chart

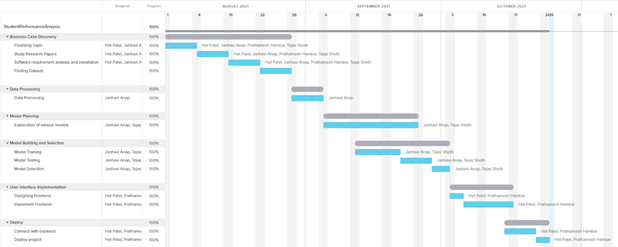


Fig 6.1 Gantt Chart

## 

## Bibliography

1. Huang, Shaobo, and Ning Fang. “Regression Models For Predicting Student Academic Performance In An Engineering Dynamics Course.” ASEE PEER Document Repository, 2010, <https://peer.asee.org/15759>.
2. Sujatha, G. & S, Sindhu & Savaridassan, P.. (2018). Predicting students' performance using personalized analytics. International Journal of Pure and Applied Mathematics. 119. 229-237.
3. Hussain, S., Gaftandzhieva, S., Maniruzzaman, M. et al. Regression analysis of student academic performance using deep learning. Educ Inf Technol 26, 783– 798 (2021). <https://doi.org/10.1007/s10639-020-10241-0>
4. Lau, E.T., Sun, L. & Yang, Q. Modelling, prediction and classification of student academic performance using artificial neural networks. SN Appl. Sci. 1, 982 (2019). <https://doi.org/10.1007/s42452-019-0884-7>
5. Abu Zohair, L.M. Prediction of Student’s performance by modelling small dataset size. Int J Educ Technol High Educ 16, 27 (2019). <https://doi.org/10.1186/s41239-019-0160-3>

## Appendices

### Appendix A

**Python Download and Installation**

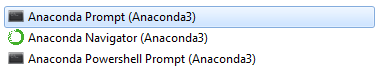
1. Visit the official website and go to <https://www.python.org/downloads/>. Click the Download button.
2. Once we click the download button, it might ask for a location to save the file. Select an appropriate location and then proceed towards the installation.
3. Double Click the downloaded .exe file and select the Add Python to PATH checkbox below to ensure it is automatically added to the Windows Environment variable. Else we have to do it later on manually. Once the box is checked, click on Install Now.
4. At the time of installation of python, the pop-up will show like the installation is in progress here.
5. Once the setup is complete, we will get a message like this. Click on the Close button to finish the installation of python.
6. Once Python is installed, go to the Windows search bar and type Python, and we will find a desktop app called Python 3.7 (32-bit). Click on that and a command prompt will open.

### Appendix B

**Jupyter Notebook Download and Installation**

There are mainly two ways to Install Jupyter Notebook,

Following are the steps to Install:

1. Directly using Python
   1. Open Command Prompt (.cmd) and give Administrative Access to it.
   2. Then open the directory where your python is installed. If you have already added Python to your environment variable, then there is no need to look for the Python folder in the directory.
   3. For installing, we are going to use the pip command. And for installation pip should be upgraded. If you have the latest pip version, then directly move to the next step. Or use the below command to upgrade your pip.  
      **python –m pip install –upgrade pip**
   4. After going into the Scripts folder > run the below cmd command: ‘pip install jupyter.’
   5. Then it will start downloading the Jupyter, and it will install the Jupyter notebook.
   6. After completion, let’s run the Jupyter notebook, using the following command: jupyter notebook.’
   7. After this, it will open the Jupyter notebook in your default browser.
   8. And you can also open the Jupyter notebook using the given link below: ‘<http://localhost:8888/tree>’.
   9. It is mandatory to start the Jupyter in the command prompt then, and only then you will be able to access it in your browser.
2. Using Anaconda
   1. The First step is to download Anaconda. Use - <https://www.anaconda.com/products/individual>
   2. Download the appropriate version and OS.
   3. Anaconda also provides an inbuilt python version. So also check the latest version of python.
   4. And start installing the software.
   5. After installation, you will see new software options, as shown below:
   6. Open the Anaconda Prompt terminal. It works the same as the cmd command prompt.
   7. But it works specifically for the installation of python and python dependent libraries.
   8. To install anything in the anaconda, we are going to use the ‘conda’ keyword.
   9. For installation of Jupyter, we are going to use the following conda command: ‘**conda install Jupyter**’
   10. It will start installing the Jupyter notebook, and if it is already installed, then there is no need for installation.
   11. To run the notebook, use the following command in the Anaconda prompt ‘**jupyter notebook.**’
   12. This will open your Jupyter notebook in your default browser.
   13. You can also open the Jupyter notebook using the following link: [**‘http://localhost:8888/tree**](http://localhost:8888/tree)**.’**
   14. But it is mandatory to run the Jupyter notebook command in the Anaconda prompt; without it, you will not be able to access the notebook. Once you run that command, you can access Jupyter notebook from anywhere using the above link.
   15. In the above link, we can see that the Jupyter notebook runs on the 8888 port number in your system.
   16. Using command prompt/Anaconda prompt, use control + c-command for windows to stop the Jupyter server.
   17. Direct from the Jupyter notebook. Check the quit button on the top of the right-hand side. This will stop the Jupyter server.

## 

## 

## Acknowledgement

We have great pleasure in presenting the mini project report on **Predicting Student Performance using Regression Analysis.** We take this opportunity to express our sincere thanks towards our guide **Dr Pravin Adivarekar** Department of Computer Engineering, APSIT, Thane for providing the technical guidelines and suggestions regarding the line of work. We would like to express our gratitude for his constant encouragement, support and guidance through the development of the project.

We thank **Prof. Sachin Malave** Head of Department, Computer Engineering, APSIT for his encouragement during the progress meeting and for providing guidelines to write this report.

We also thank the entire staff of APSIT for their invaluable help rendered during this work. We wish to express our deep gratitude to all our colleagues of APSIT for their encouragement.

**Student Name1: Janhavi Anap   
Student ID1: 19102043**

**Student Name2: Prathamesh Hambar   
Student ID2: 19102001**

**Student Name3: Tejas Sheth   
Student ID3: 19102026**

**Student Name4: Het Patel   
Student ID4: 19102005**