MACHINE LEARNING APPROACH TO PREDICT SGPA AND CGPA

BY

MD. TANVIR AHMED

ID:171-15-8801

AND

MD. REFAT HASSAN HREDOY

ID:161-15-7512

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Ms. Nazmun Nessa Moon
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Raja Tariqul Hasan Tusher
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

2nd January 2021

APPROVAL

This Project/internship titled **"Machine Learning Approach To Predict SGPA & CGPA"**, submitted by **MD. Tanvir Ahmed**, ID No: **171-15-8801**, and **MD. Refat Hassan Hedoy**, ID No: **161-15-7512** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **28.01.21.**

BOARD OF EXAMINERS

_____

Dr. Touhid Bhuiyan   **Chairman**

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

_____

Dr. Fizar Ahmed   **Internal Examiner**

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Farah Sharmin

Ms. Farah Sharmin   **Internal Examiner**

Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Dr. Mohammad Shorif Uddin   **External Examiner**

Professor

Department of Computer Science and Engineering

Jahangirnagar University
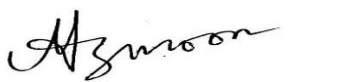
DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Nazmun Nessa Moon, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.
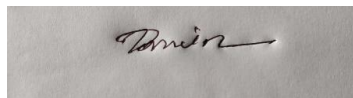
Supervised by: **Co- Supervised by:**
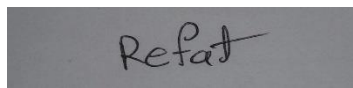
Ms. Nazmun Nessa Moon   **Mr. Raja Tariqul Hasan Tusher**
Assistant Professor   Senior Lecturer
Department of CSE   Department of CSE
Daffodil International University   Daffodil International University

Submitted by:

MD. Tanvir Ahmed
ID:171-15-8801
Department of CSE
Daffodil International University

MD. Refat Hassan Hredoy
ID.161-15-7512
Department of CSE

ACKNOWLEDGEMENT

At first, we express my heartiest thanks and gratefulness to Almighty Allah for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound my indebtedness to Ms. **Nazmun Nessa Moon, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Machine Learning interest of our supervisor in the field of AI to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express my heartiest gratitude to the Almighty Allah and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Prediction of the SGPA & CGPA is a helpful thing to the university students. From this project students can easily get the prediction about their final result. So, the students will able to prepare them for a good result. Students passes time by doing many other activities. Student use Facebook, Instagram, Twitter etc. social media. They play mobile games, listen music and do some other activities. So, they passed many times with these activities. So, if a student passes his/her time with doing all those activities much, she will not to be able to make a good result. So, from our project students can get a routine or rule about study and their other activities. Also, there will be a predict result for students by their activities. Today, we can see the use of machine learning with Python everywhere. Following this, we built here a smart SGPA & CGPA prediction project and which are the effects on students. We use here Naïve Byes algorithm for prediction the results. Naïve Byes is a simple algorithm but most powerful algorithm for prediction. It's also a machine learning algorithm. So, Students will get a prediction of their final exam result. They, can prepare them to make a good result by following the routine of our SGPA & CGPA prediction project.

TABLE OF CONTENTS

PAGE NO

CONTENTS

CHAPTER

vi

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

TABLE NAME   **PAGE NO.**

CHAPTER 1
INTRODUCTION

1.1 Introduction

Nowadays, the world goes so faster. This is the era of modernization & competition. In the education system there is too much competition between students. Basically, in the universities there is the educational system is different from school or college. In universities there has some curriculums or activities for students. In the universities there has, attendance mark, quiz marks, assignment marks, presentation marks, project marks, mid-term marks, & final exam marks. So, if the students achieve the good marks in all segments then they have to study more and sincerely. Students pass their time by doing all other activities. That's are also important for a human but if students pass their all times or more times using social media or something like that, then they will not be able to make a good result. That's why we built a project that is called, Prediction Of The Final Exam SGPA & CGPA. In our project , we tried to make a good routine for students. As a result, students can follow the routine and make a good preparation for exam to get a good result. Also, we collected data from different kinds of university students. There are some students who got good marks in all the segments and there also those students who did not get the enough good marks and also there are some students who got worst marks in all the segments. If a student got less mark than his/her expectations but he/she wants to make a good result then he/she will use our project for the prediction SGPA. He/she will get suggestion from our project that how he/she would take preparation for a good result. In our project we used machine learning/ Deep learning.

1.2 Motivation

We are doing this project for the benefit of the university students to predict their SGPA & CGPA easily. That why, students will have the advantage of making good results. When, a student can predict their final SGPA & CGPA by using our project they will be able to take good preparation for a good result**.**

1.3 Rationale of the Study

There are lots of attempts to predict the SGPA & CGPA for students. But using data mining is the most efficient and good decision. Data mining or data classifier can show the accurate output with less training data. So, this is an interesting thing that we can use easily.

The uses of machine learning is everywhere and it's very reliable to predict anything. So, we can see how the capacity to predict SGPA & CGPA for students. So we used data mining in machine learning to predict the final exam SGPA & CGPA.

1.4 Research Questions

- Does more use of social media impact academic result positively, negatively or neutrally?
- Does university students use all the day social media the most?
- How can students prepare them for a good result?
- How can student manage their routine for study?
- Does the predicted result is helpful to achieve that?

1.5 Expected Output

Expected outcome of this research- based project:

- ➢ Students will see how they can get good results.
- ➢ There will be a daily routine to get good results.
- ➢ There will be a predicted targeted SGPA/CGPA which helps to take a good preparation.

1.6 Report Layout

In **chapter 1:** In this chapter we can discuss as regards the Introduction, Motivation then Objectives and also discuss as regards the Expected Outcome and so on is discussed briefly. In a word, chapter 1 is elaboration of introduction of this project.

In **chapter 2:** In this chapter we had presented about the primary condition of our work and talk over jurisprudence in the project. We also present in regard to the related works on this area which were studied are showed. Their findings and limitations are summarized and hence the scope and challenges of the research are also mentioned.

In **chapter 3:** In this chapter we talked about research methodology discusses research subject and instrumentation, data collection procedure, statistical analysis and implementation requirements.

In **chapter 4:** experimental results and discussion Experimental Results and Descriptive Analysis are presented.

In **chapter 5:** In this section we displayed the performance of the entire project. Presents a short conclusion, and list of references.

CHAPTER. 2
BACKGROUND

2.1 Introduction

Technology makes our life so fast and easy where we want everything in our touch. In this chapter, we will discuss related research or project about data classifier which is related to data mining. Data mining id primary foundation of machine learning. Data mining mainly using for extract more accurate data. For any kind of prediction data mining is more important. In the first section we will discuss about previous related work, then in the second section we will show the outcome or a summary of my study of the related work and then we will discuss about the benefits and challenges that we face to do this project.

2.2 Related Works

In a research done by Lubna Mahmoud Abu Zohair[1] who made a project Prediction of Student's performance by modelling small dataset size. Prediction of student's performance is an urgent desire at this time. So, in this situation they think to help at-risk students and assure their retention, providing the excellent learning resources and improving the university's ranking and reputation. But that was not so easy to be achieved for startup to mid-sized universities, especially those which are specialized in graduate and post graduate programs, and have small students' records for analysis the students result. Also, in most researches that were aimed to classify or predict, researchers used to spend much efforts just to extract the important indicators that could be more useful in constructing reasonable accurate predictive models. And the finally they will look at the selected features while training the dataset on different machine learning algorithms, like in (Comendador, Rabago, & Tanguilig, 2016; Mueen, Zafar, & Manzoor, 2016).

Navin venkat (Indian Institute Of Science) & Sahaj Srivastava (Birla Institute of Technology & Science)[2] are made a project on Predicting Student Grades using Machine Learning. They predict students grade point with machine learning. Also the used data classifier as data mining. They model the prediction of final Grade as a classification problem. Each grade is assumed to be a class. Hence, They have 9 classes (A through E, and NC), since those denote the students who took the course till completion. For these tests they ignore the class withdrawn (W) because most of the data is missing for such records. Since the dataset is very small (197 records for scores, after removal of they as mentioned above), they perform a stratified 5-fold cross-validation for each of the classifiers to observe and compare their stability. The stratification aids in handling the class imbalance problem. They have used Decision Tree, Naive Bayes, SVM (with linear, RBF and sigmoid kernels).

Khurum Nazir Junejo and Eman Eman [3] both are made a project on Grade Prediction Using Supervised Machine Learning Techniques. They have made a first attempt at predicting the final grade of students in undergraduate courses based on their performance in the same course prior to the final exam and the type of the course and the teacher teaching that course. After applying various data preprocessing and machine learning algorithm on 2500 course records, they have achieved quite a remarkable prediction performance (varying between 73-96 % accuracy). They have the following useful conclusions: Grade of more than 96% students can be accurately predicted even without the result of the final exam. This implies that the performance of students in various instruments before the final exam is highly correlated to their performance in the final exam. Another interesting point is that the F grade is the most easiest to predict followed by C grade, while B+ is the most difficult grade to predict. After all the number of students whose grade slip because of their performance on the final exam is more than those who improve it because of final exam and finally rule induction classifier outperforms ID3 classifier on student performance prediction.

2.3 Comparative Analysis and Summary

With an extensive research over relevant papers and projects, we have come to the decision of using Naïve Byes Classifier because, this classifier is the best for predict the probability of different class based on various attributes.

•.It is very simple and easy to implement. It doesn't require as much training data. It handles both continuous and discrete data. It is highly adaptable with the number of predictors and data points.

• Naïve Byes Classifier is the most accurate among others data classification algorithm which is most simple algorithm.

• Naïve Byes is easy to use and convenient for working on further development.

• It is highly adaptable/flexible quick to execute, and typically out performs other algorithms..

• We have received a great result and high accuracy by using Naïve Byes Classifier.

We decided to use Naïve Byes Classifier as the main classification model Naïve Byes classifier is first to train and fast to classify. It is not sensitive to irrelevant features. It handles real and discrete data. Also it handles streaming data well. It evaluates results on test data. We have remained careful cannot about bad accuracy and output. By using dimensions, we have got good accuracy range in order to find the best comparison possible

2.4 Scope of the Problem

The main focus of this research work is making a system that can predict the final exam result of the students. There will be some scope of problems are like, such as, students can think over their midterm result is good, so they don't need to take much preparation for final exam. So It would be harmful for those students.

2.5 Challenges

1) Data Collection

Since there are available multiple factors related to prediction results , data collection had faced one of the major difficulties for this research project. When we started collecting data locally it became too much difficult because of the pandemic situation of our country. For more accuracy we need more train data. So we decided to collect data online. Through survey questionnaire, we could collect data as much as we could.

Model selection

Model selection is an most important part and also one of the hardest parts of a research project. The success of any research relies on Data set and Model selection. If we can choose the model correctly, there should be a positive outcome soon. On the other hand choosing a bad one may lead to failure as a sure fact. For this, we have tested some models with our test data in search of the most suitable one for our research project. But from that entire search, we have come to the conclusion that Naïve Byes Classifier algorithm as the easiest and the best one for us. When we have found that it provides an easy to use library, we understood that the classifier work would be more-easy. So, we finalized this model for our use in our project.

2) Data labeling

Data labeling is also a most important part for makes code so faster. It was the most important part of our work. Because it makes code more-faster to run and less time consuming for data training. We use Label Encoder in Python for labeling our data.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Research methodology used to identify select, process and analyze information. Now in this chapter we will discuss & describe about our research methodology. Moreover, tools for the research project, data collection, research topic, pre-processing, processing, statistical analysis, and its implementation will be discussed in this session.

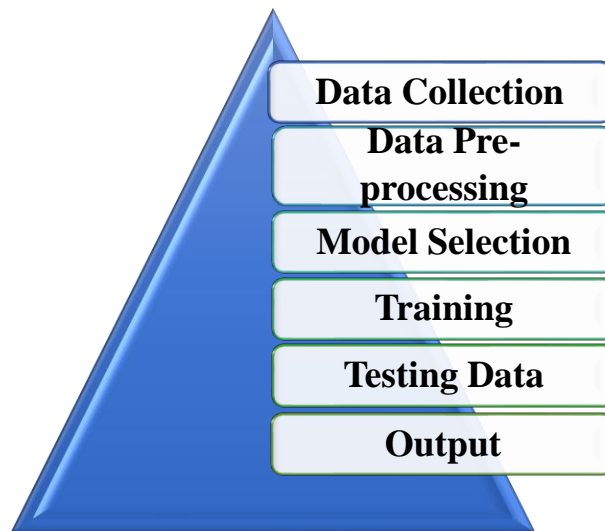The steps of research methodology is given below at fig 3.1:

Fig 3.1: Research Methodology at a glance.

3.2 Research Subject and Instrumentation

Data is the main part of the research project. It is a very critical part for a researcher to find out perfect data and perfect algorithm or model for our research work. We also needs to study about related research papers. Then we need to make several decisions:

1. Which data should be collected?
2. That collected data are okay?
3. How each data should be organized?
4. How each data should be labeled?

3.3 Data Collection Procedure

In survey, we used a questionnaire as dataset. These data are mainly selected according to the effects of mobile use on a student's academic result, health condition, social relationship, and psychological status.

1) Data Pre-processing

Data pre-processing is a processing that means to the pre phase of processing datasets. Generally raw data sets are not able to perform operations and generate expected outcome. As a result, data pre-processing is required. Also it is considered to be one of the most important parts of research. In this phase, we have collected more than 500 surveys. Here, to predict the final exam result.

2)Data Organizing

Data organizing is a system of organize data. So, for organizing data, we have tested and trained the data and saved them in two folders. We have also used validation folder to check train data validation in data organizing.

Labeling Data:

At first we have to do label the data. Its called data labeling. We labeled the dataset with label encoder with python. Fig 3.2 shows the data labeling by Data Encoder.

```
from sklearn.preprocessing import LabelEncoder

fc=['quiz marks','midterm marks','assignment marks','presentation marks',
    'CGPA','SGPA','academic study','spend online','spend on mobile','spend on pc',
    'spend on social media','spend online for study','leisure time',
    'most of time on mobile','most of time on pc','study routinely','regular class',
    'note in class','listen carefully in class','study before the exam','help of a teacher/classmate']
X=df[fc]
```

Fig 3.2: Data Labeling by Data Encoder

After labeling data we got numeric dataset. Given below with Fig 3.3:

	quiz marks	midterm marks	assignment marks	presentation marks	CGPA	academic study	spend online	spend on mobile	spend on pc	spend on social media	spend online for study	leisure time	most of time on mobile	most of time on pc	study routinely	regular class	note in class	listen carefully in class	study before the exam
0	10	12	4	5	3.22	3	8	5	3	3	1	1	Social media	Social media	No	Yes	Yes	Yes	Yes
1	12	20	5	6	3.20	1	1	1	2	1	1	1	Social media	Other	No	Yes	Yes	Yes	Yes
2	8	16	4	5	3.39	1	6	7	3	3	1	2	Social media	Study	No	Yes	No	Yes	Yes
3	9	18	4	6	3.45	1	7	8	9	6	3	5	Social media	Social media	No	Yes	No	Yes	Yes
4	11	17	3	6	2.80	2	5	5	7	3	2	4	Social media	Social media	No	Yes	No	Yes	No
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
175	12	21	5	5	3.28	2	3	1	3	4	3	3	Study	Social media	Yes	No	Yes	Yes	Yes
176	9	15	5	6	3.23	3	2	3	2	3	2	3	Study	Study	Yes	No	No	No	Yes
177	12	21	3	4	3.66	3	6	3	3	2	1	2	Study	Study	Yes	Yes	Yes	Yes	No
178	12	22	4	5	3.88	4	2	3	2	2	3	1	Study	Study	Yes	No	No	Yes	Yes
179	12	20	5	6	3.72	2	5	3	2	3	2	1	Study	Study	Yes	Yes	Yes	Yes	Yes

180 rows × 20 columns

Fig 3.3: Numeric Dataset

4)Data Storing

Where data can be stored it's called data store So, we have to store data. Here we store all the data in Google drive because it makes our work easier. In Google drive, we can use those online stored data in our project we haves saved all the data as CSV files. The advantage of google drive that data cannot be lost.

3.4 Statistical Analysis

Table 3.1 The trained data amount 504:

Current Semester	Student Amount
1,2	31,39
3,4	41,43
5,6	45,40
7,8	38,43
9,10	48,42
11,12	41,53

3.5 Implementation Requirements

• Python 3.8

Python 3.8 is a Python version which is a high-level programming language. For machine learning related research most of the researchers use it to do their research. Python is a highly recommended programming language for Artificial Intelligence (AI) based work and it is very popular among new generation's programmers because it is very simple, easy to learn and understandable.

• Google CoLab

For python programming language Google CoLab is a free to use open source distributor. Here we can work online through our browser as well as through notebook. The main benefit of Google CoLab is it provides us free online virtual GPU access. That's very helpful to us.

• Hardware/Software Requirements

1. Operating System (Windows 7 or above).
2. Web Browser (Preferably chrome).
3. Hard Disk (Minimum 4 GB).
4. Ram (More than 4 GB).

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

In this chapter for our model and code implementation we have collected the data first. The procedure is given below:

• We are working with prediction of impacts on social media using, paly games, leisure time spend, study time, attendance marks, assignment marks, quiz marks, presentation marks, midterm marks etc. So, we had to collect student data on these topics.

• Taking survey we have spent more time. After all in this pandemic situation it is so tough to collect data from students in physically.

• We collected data from online by using Google Form.

• After labelling data, they become perfect for further use.

• After that, we have converted data to numeric type.

• After that, we took the normalized and finalized data so that we can begin the training.

• Then we have preprocessed our data.

4.2. Model Summary

Model summery is an important part of experimental segment. We used a sequential model with some parameters such as data, label, missing, weight, silent, feature_names, feature_types, nthread. We used all of those parameters in our research. These have provided accuracy in our work. Even, by researching on other such research works, we have found that using all the parameters gives the best & accurate results with the model. We then, have taken the opportunity so that we can optimize the model with best accuracy and output from it.

4.3 Experimental Results & Analysis

From our collected data we have run our dataset to create a model. It provided us with the desired the output. Then we created each features in dataset.

4.3.1 Correlations for each Features

When we run dataset we get a graph for creations for each features in dataset. In the Fig 4.1:
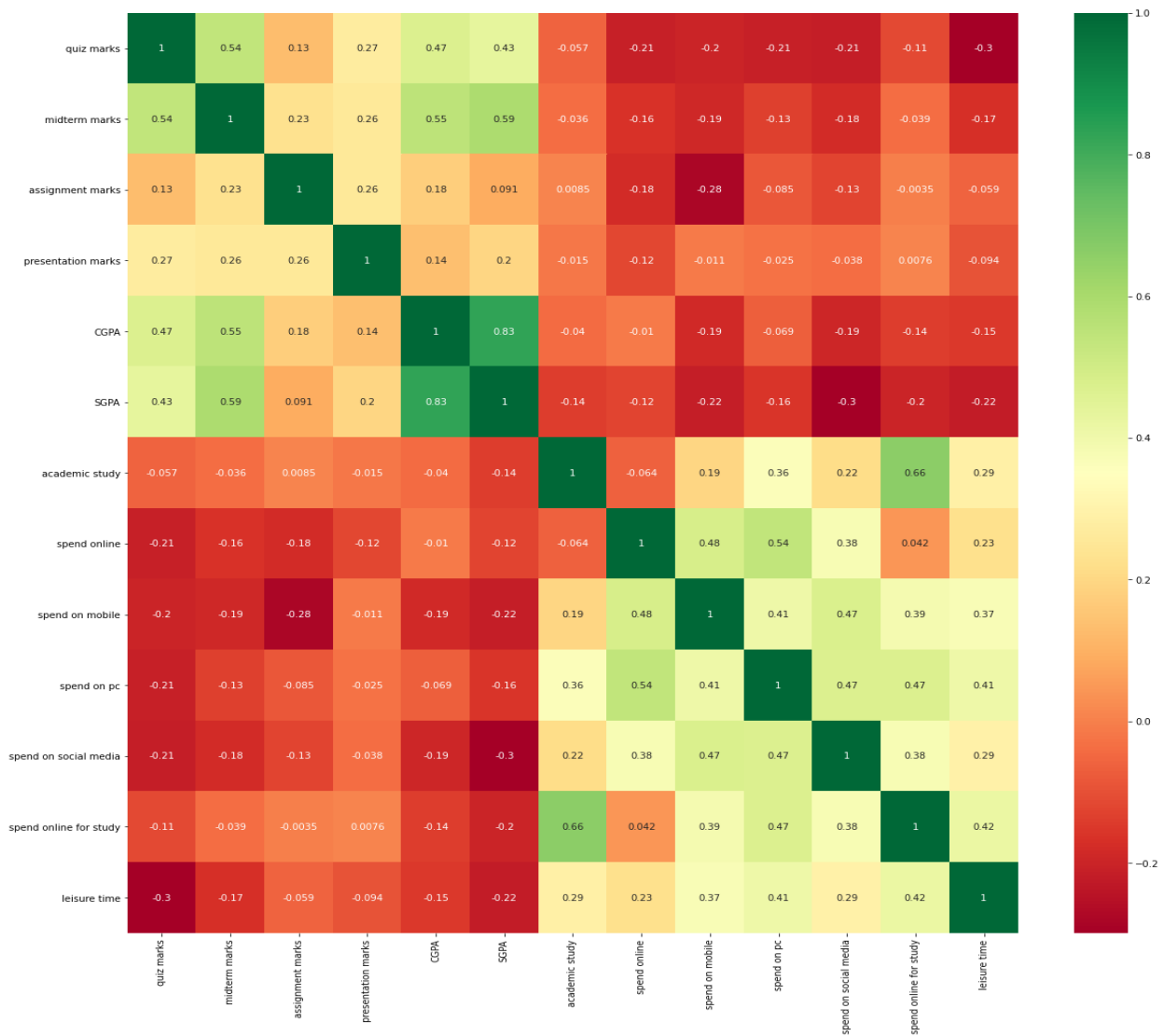
Fig 4.1: Correlations For Each features In Dataset.

Here we can see in the figure 4.1, here is all features are correlated. From the beginning we got, quiz marks, mid-term marks, assignment marks, presentations marks and then from these marks we can predict a SGPA for the students. But here also some activities those will effect in the result. Those are, academic study, spend time on mobile, spend time on online, spend time on mobile, spend time on pc, spend time on social media, spend time for online study, spend leisure time. The result is dependent on those activities also.

4.3.2 Effect Of Activities

There are so many activities in student daily life. They do study, time pass with social media, video game playing, leisure time, online study etc. All these activities have to a routine. That's why we created this project for help the student make their routine and target a good SGPA or CGPA.

4.3.2.1 Assignment

Assignment is a very important task of all subjects. It is mandatory to submit for make a good result. It carries 5 marks out of 100. So, its not so neglected topic. Teachers give a topic to the students and give instructions for write an assignment. After that students do their assignment work carefully. In an assignment there would be some important topics, some important objectives, case study about the main topic etc. If students do this assignment good, then it is helpful to get a good result. The graph is showing at Fig 4.2:
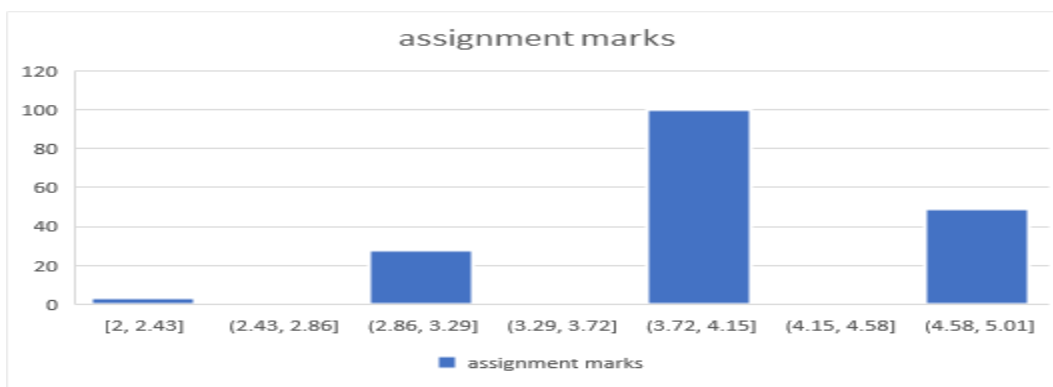
Fig 4.2: Effect of assignment

4.3.2.2 Quiz

In a semester there are another important test. That is quiz. Quiz is a plus point to make a good SGPA or CGPA. If students do well in the quiz, then they will be able to make good SGPA. Fig 4.3 shows that.
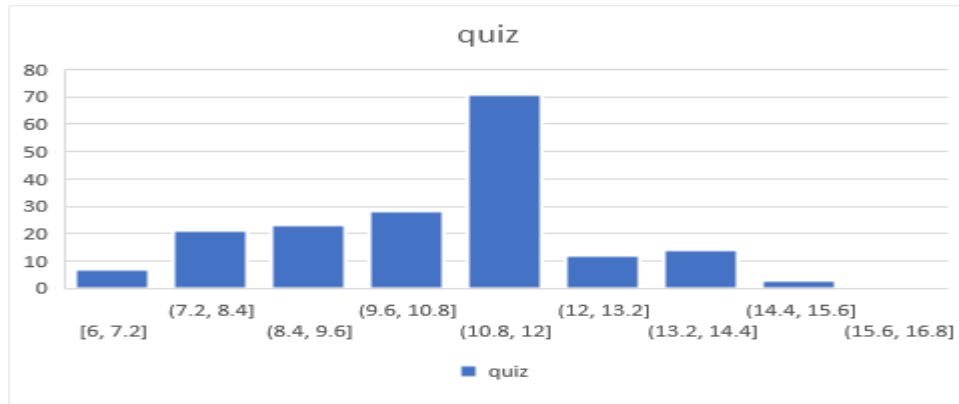
Fig 4.3: Effect of quiz

So, here is the graph of quiz effect.

4.3.2.3 Presentation

Presentation is a good way to make a well communication skill.Students who can make presentation and able to give presentation in front of audience, they are very confident. Presentation creates confidence level up. Every students must have this skill. Presentation segment carries 7 marks out of 100 for each subjects or courses. Here is the graph for effect of presentation. Fig 4.4 shows that.
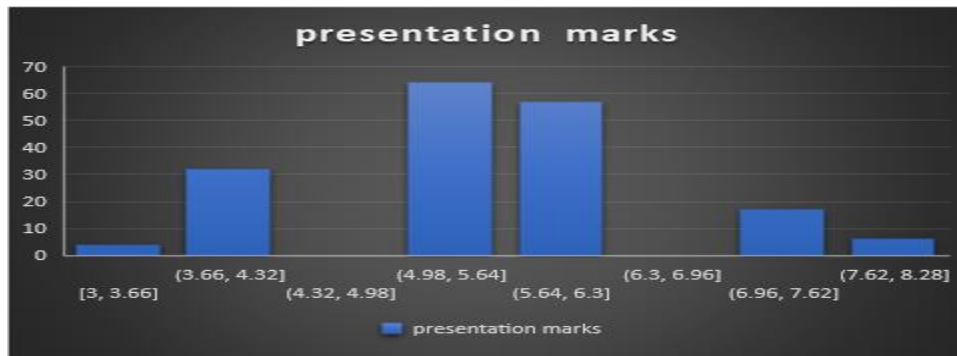
Fig 4.4: Effect of presentation

4.3.2.4 Mid Term Exam

Mid-term exam is a big part of get marks out of 100. In the mid term there will be 25 marks on each subjects. So, if students get good marks in the mid term exam also, so it is very easy for those students to make a good sgpa and cgpa. In our project we suggest students how to make a good result in the final exam. Students will get a good suggestion from our project. Here is the effect of mid-term exam graph at Fig 4.5.
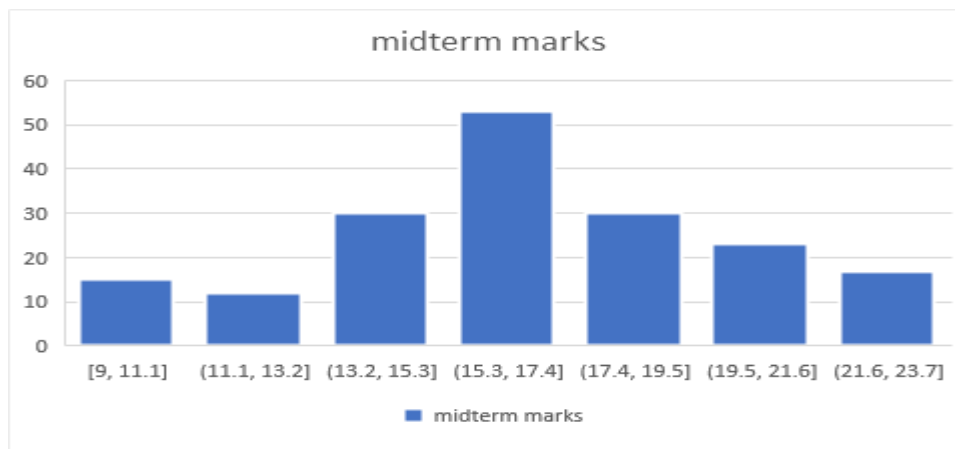
Fig 4.5: Effect of mid-term exam

So, we can see from the graph, the effect of mid term exam. Who will get good marks in the mid-term they will be able to make a good result in the final exam easily.

4.3.2.5 Effect Of Semester
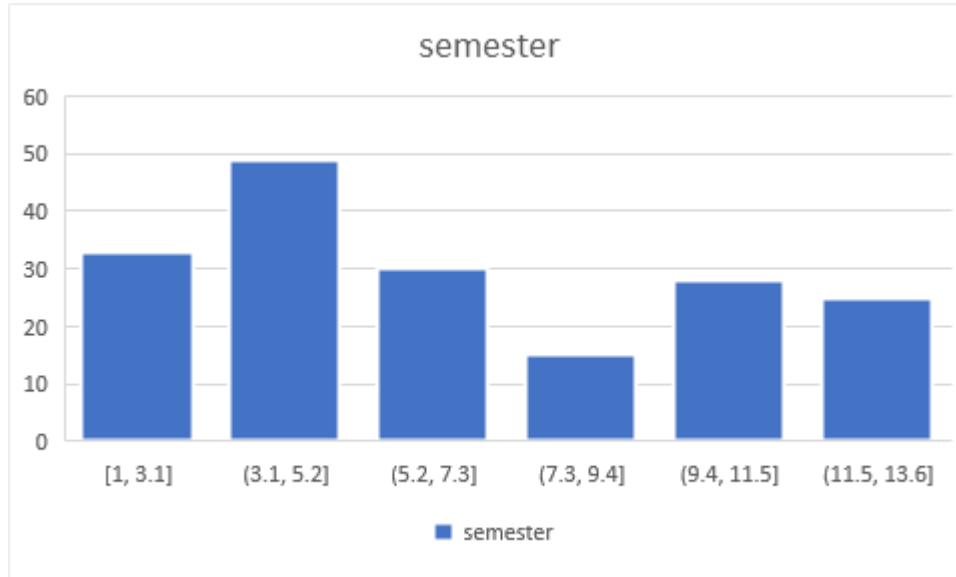
Fig 4.6: Effect of semester

Here, at the Fig 4.6 we can see from the graph, here is the effect of semester on results. The graph shows different types of results with semester.

4.3.2.6 Effect Of SGPA

After all these we got the effect of SGPA graph. At Fig 4.7 we can see this graph:
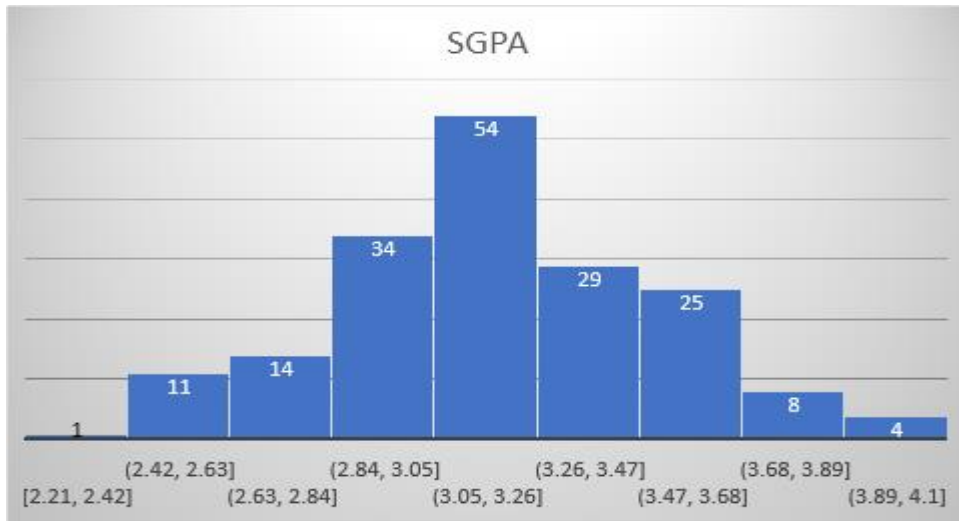
Fig 4.7: Effect Of SGPA

So, this is the graph for effect of SGPA.

4.3.2.7 Effect Of CGPA

Here we got, both SGPA & CGPA from this graph. The graph is given below at Fig 4.8
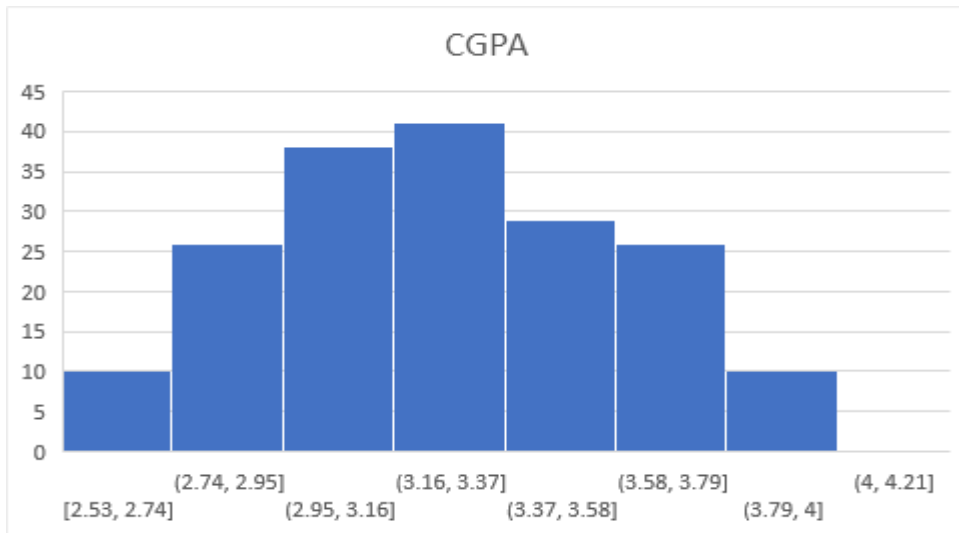
Fig 4.8: Effect Of CGPA

Here we can see the CGPA chart.

4.3.2.8 Linear Operation

In this study we need best accuracy. We have run different types of algorithm to know which one give a better output. That's why, at WEKA we have run linear regression model and logistic regression algorithm at python.

4.3.2.9 Linear Regression

At WEKA we used linear regression. Linear regression model show the CGPA and SGPA accuracy 87%. So, we got 87% accuracy on SGPA and CGPA. The following figure 4.9 & 4.10 Shows in the bellow:
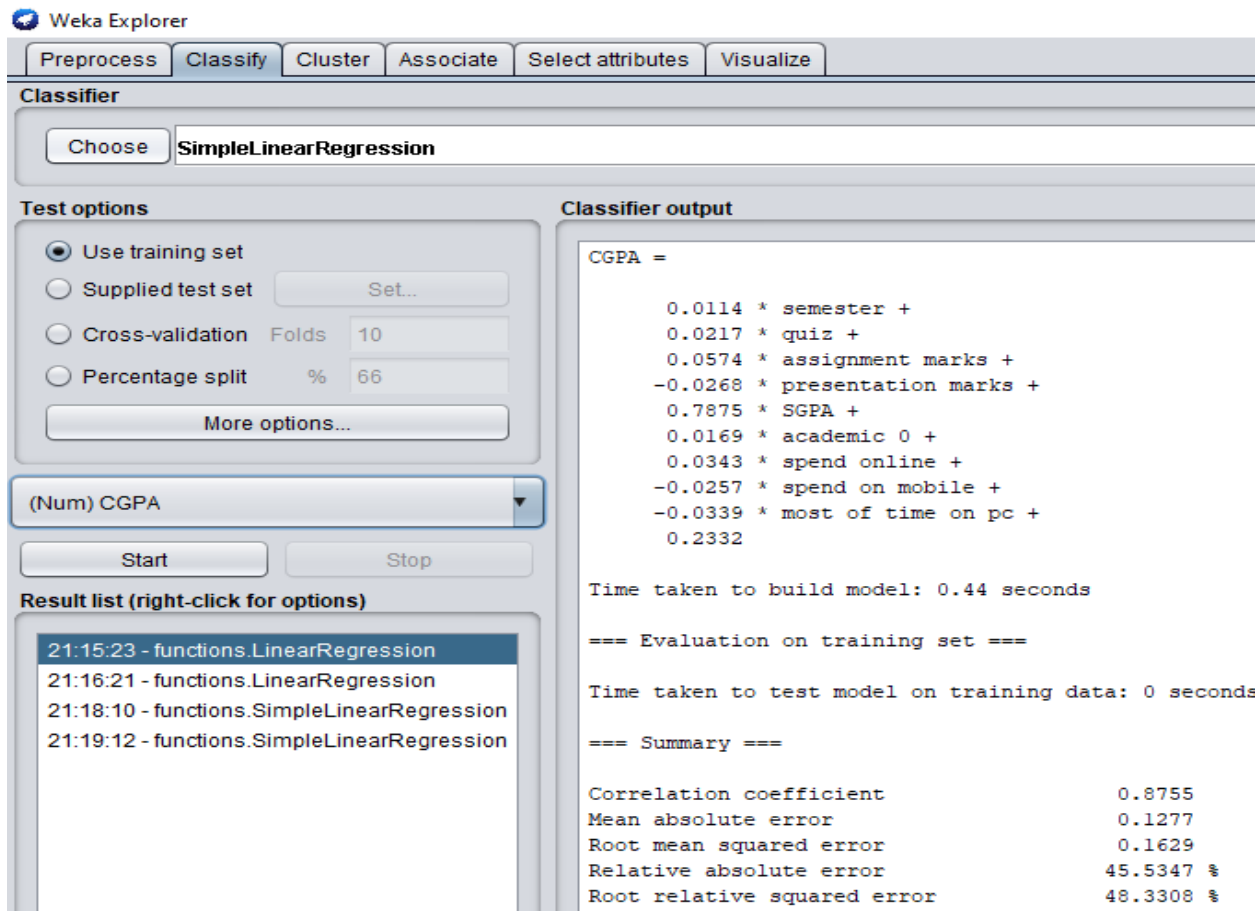
Fig 4.9: Linear regression (CGPA)
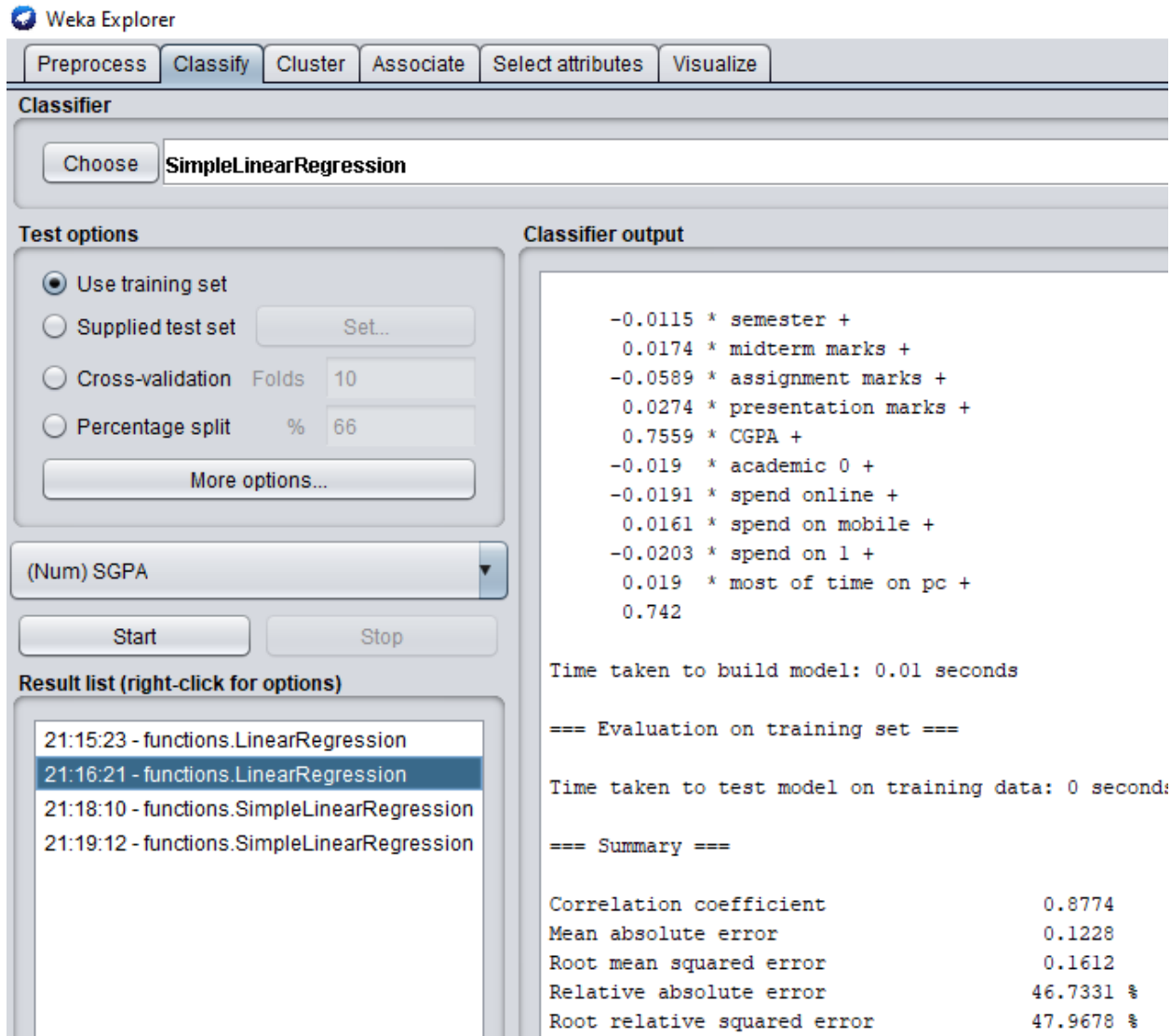
Fig 4.10: Linear regression (SGPA)

4.3.2.10 Simple Linear Regression

Here, We used simple linear regression and we got 83% accuracy on CGPA and SGPA.

The following figure 4.11 & 4.12 Shows in the bellow:

Fig 4.11: Simple Linear Regression (CGPA)
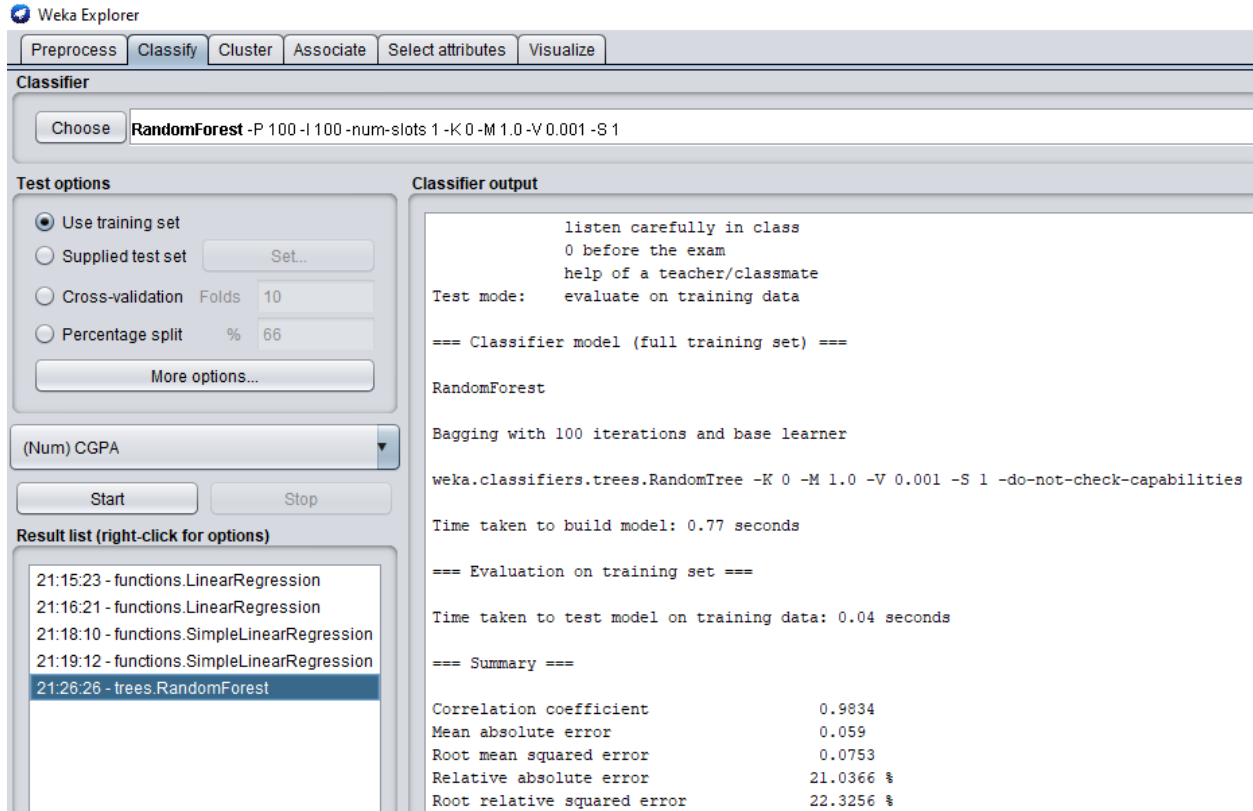
Fig 4.12: Simple Linear Regression (SGPA)

4.3.2.11 Random Forest (CGPA & SGPA)

In classification on WEKA we use tree and choose Random Forest tree and show the correlation coefficient for CGPA rate is 98.34%. The following figure 4.13 shows in bellow:
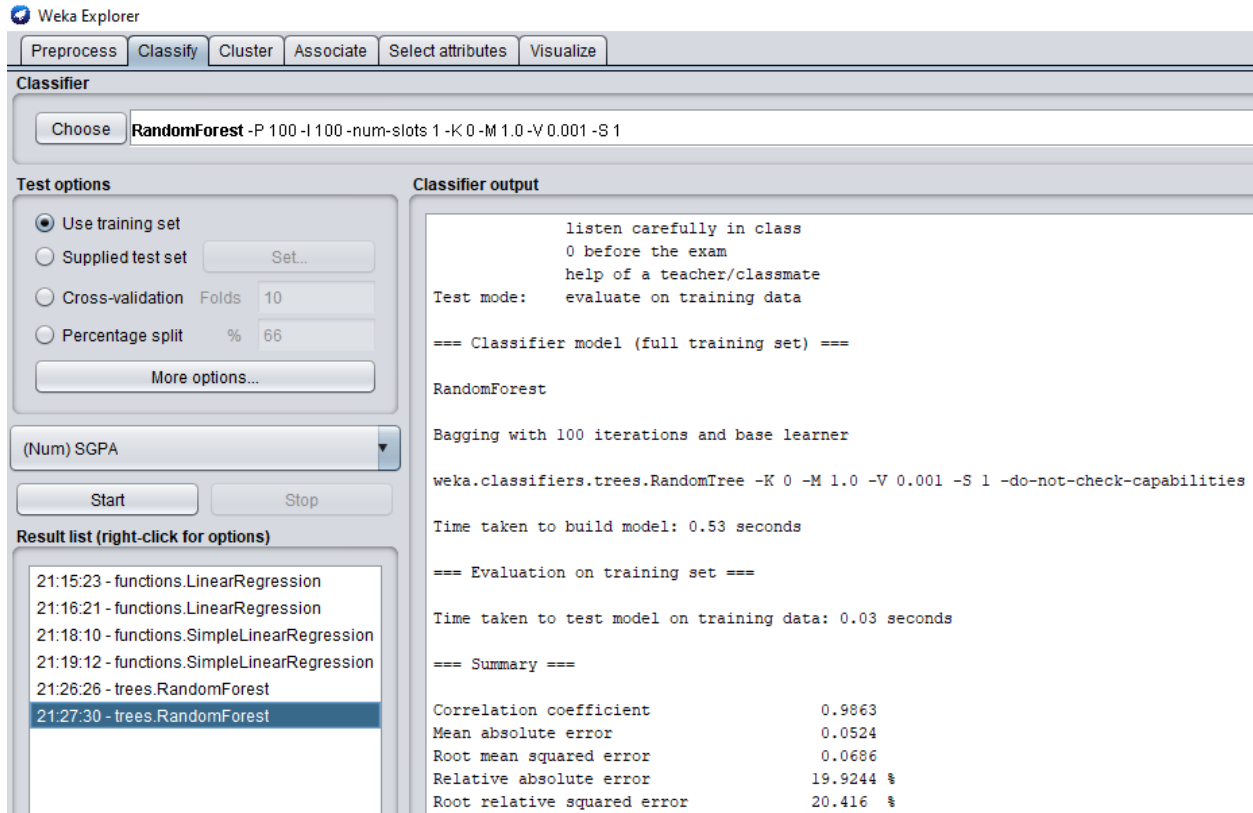
Fig 4.13: Random Forest (CGPA)

In classification on WEKA we use tree and choose Random Forest tree and show the correlation coefficient for SGPA rate is 98.63%. The following figure 4.14 shows in bellow:

Fig 4.14: Random Forest (SGPA)

4.3.2.12 Random Tree (CGPA & SGPA)

We take another tree and choose Random tree and show the correlation coefficient for CGPA rate is 99.99%. The following figure 4.15 shows in bellow:
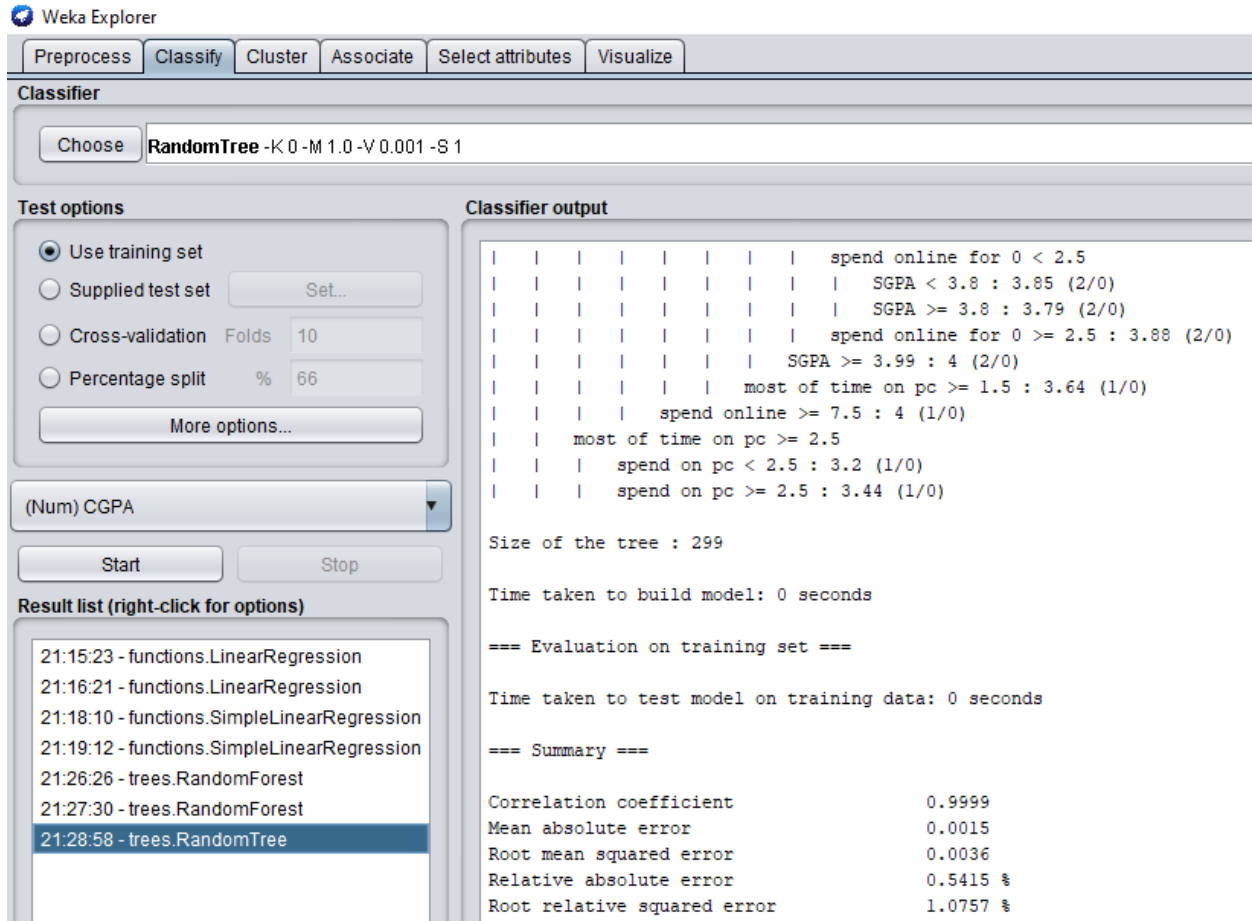
Fig 4.15: Random Tree (CGPA)

We take another tree and choose Random tree and show the correlation coefficient for SGPA rate is 1. The following figure 4.16 shows in bellow:
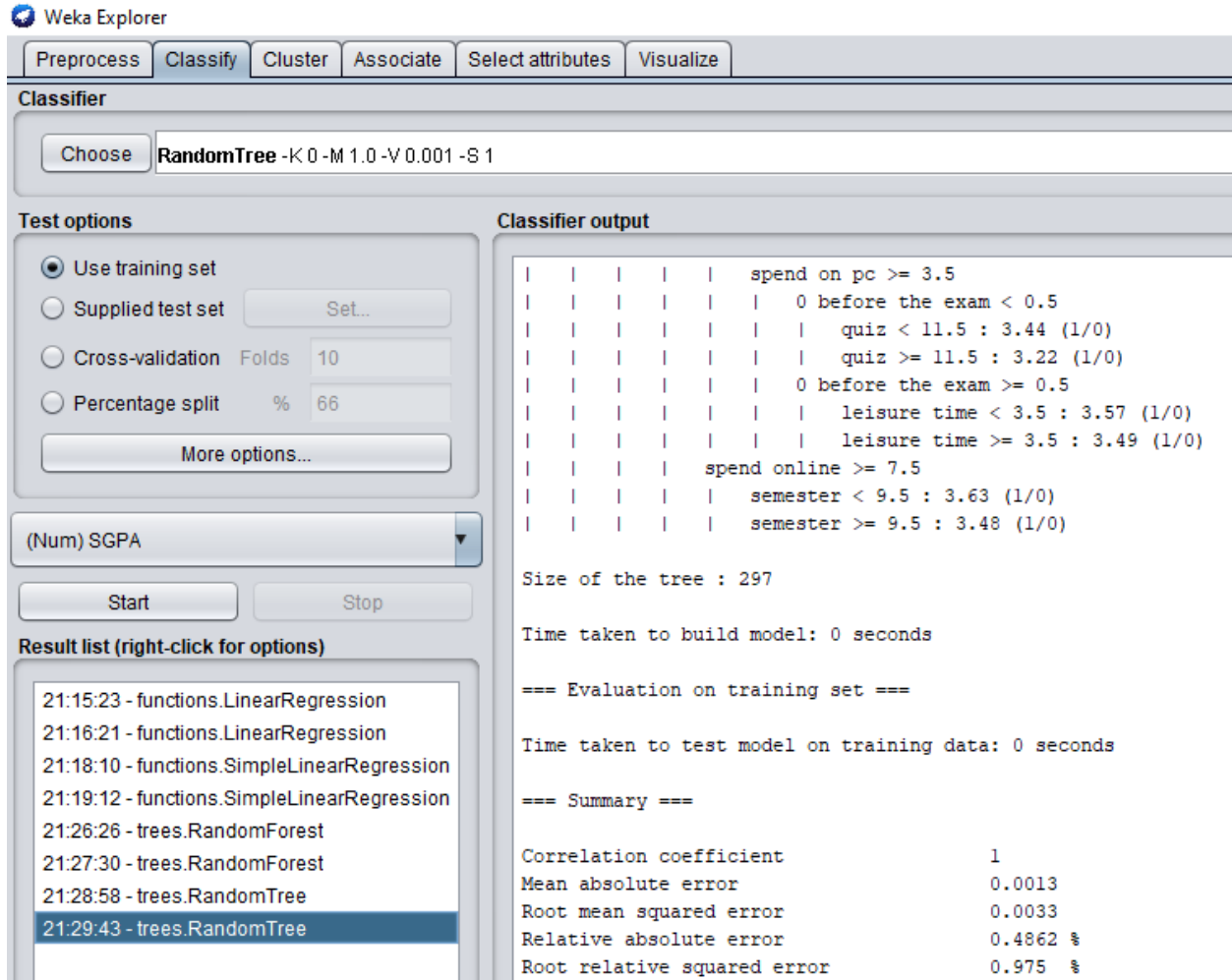
Fig 4.16: Random Tree (SGPA)

4.3.2.13 Decision Table (CGPA & SGPA)

We take another tree and choose Decision Table and show the same correlation coefficient rate for CGPA. The rate is 81.82%. The following figure 4.17 shows in bellow:
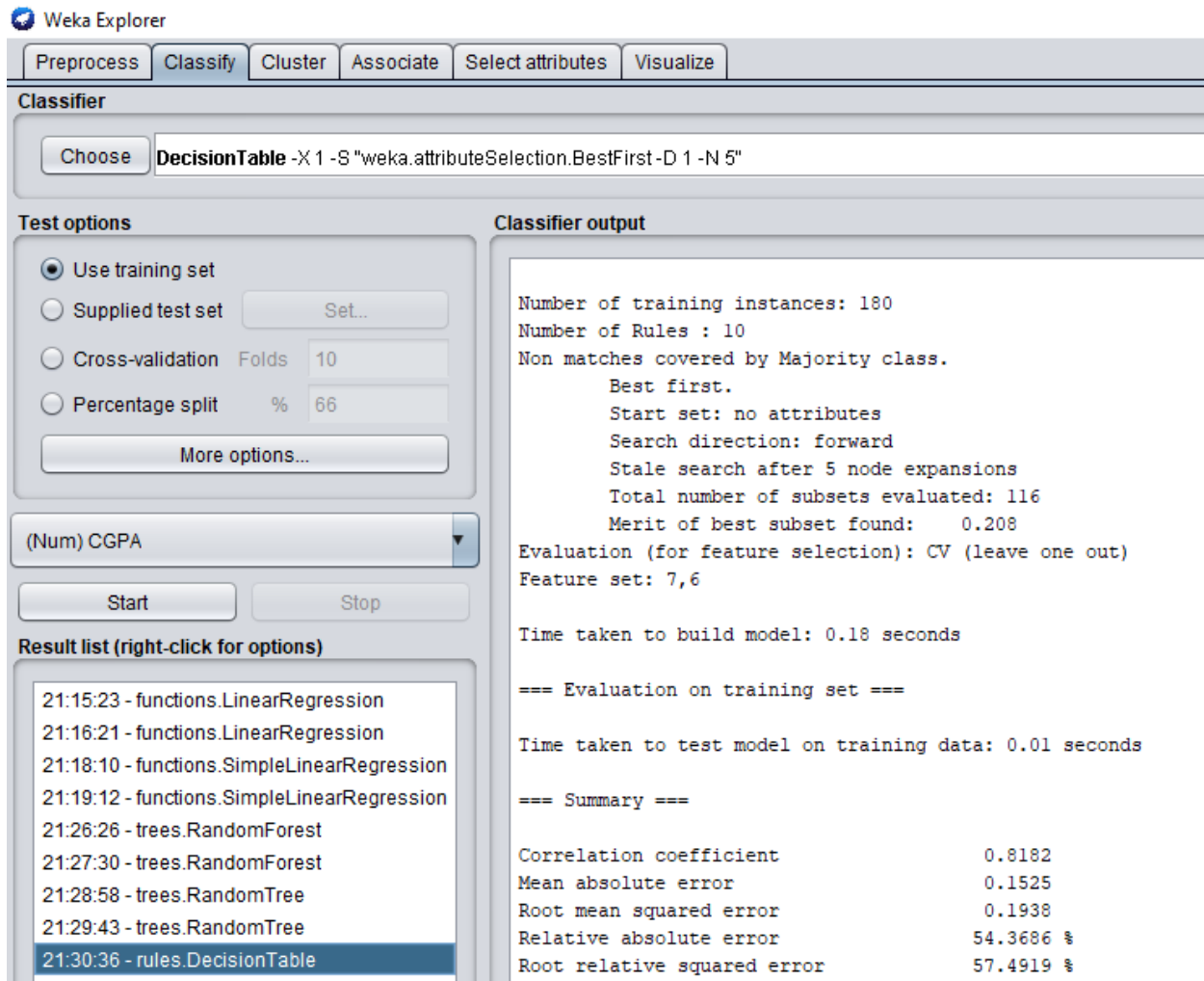
Fig 4.17: Decision Table CGPA

We take another tree and choose Decision Table and show the same correlation coefficient rate for SGPA. The rate is 84.82%. The following figure 4.18 shows in bellow:
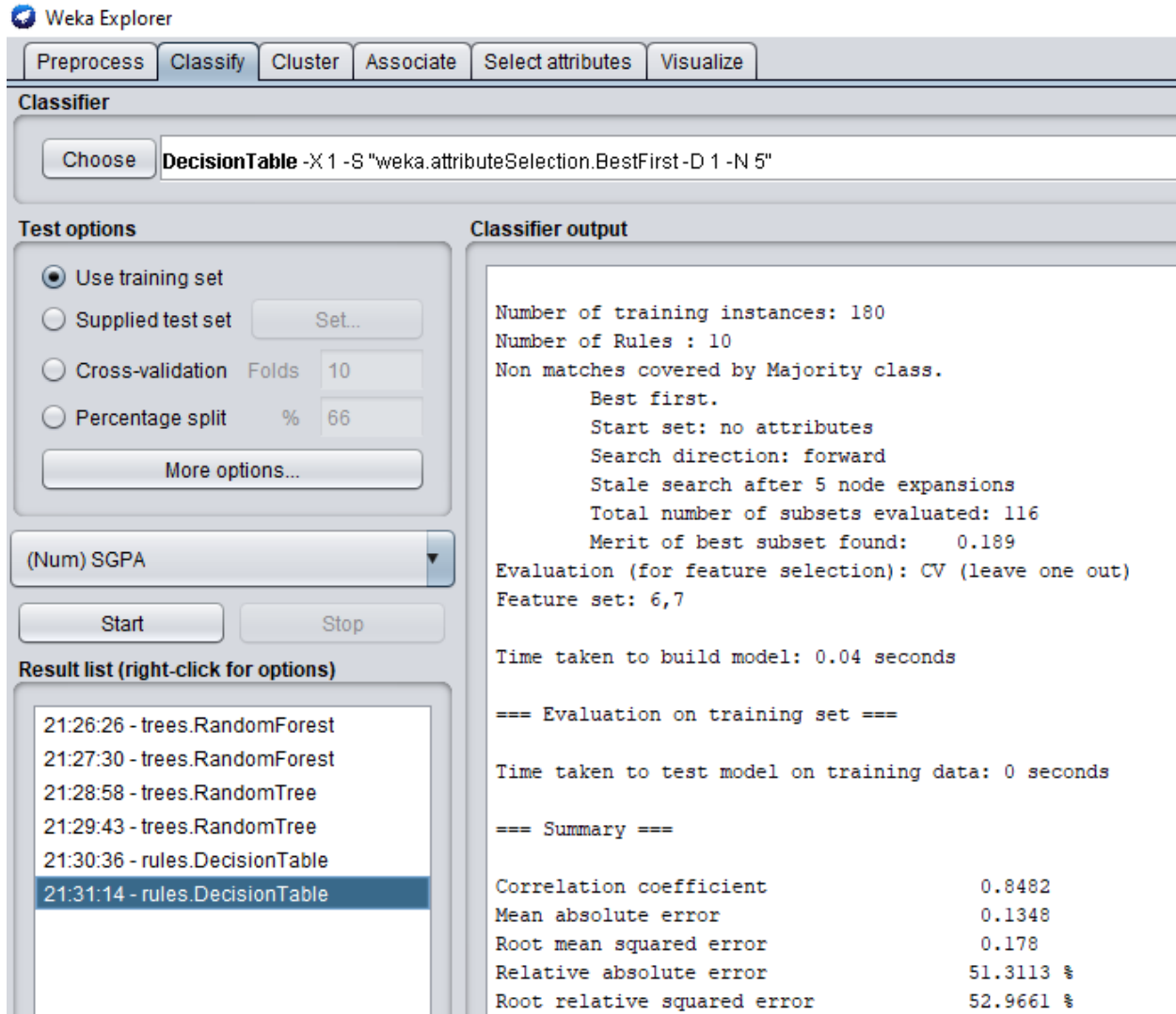
Fig 4.18: Decision Table (SGPA)

4.3.2.14 Logistic Regression using Visualizing Confusion Matrix

Use a confusion matrix to visualize the results of the model in the form of a confusion matrix using numpy, seaborn and matplotlib. Here, figure 4.19 show the matrix value.
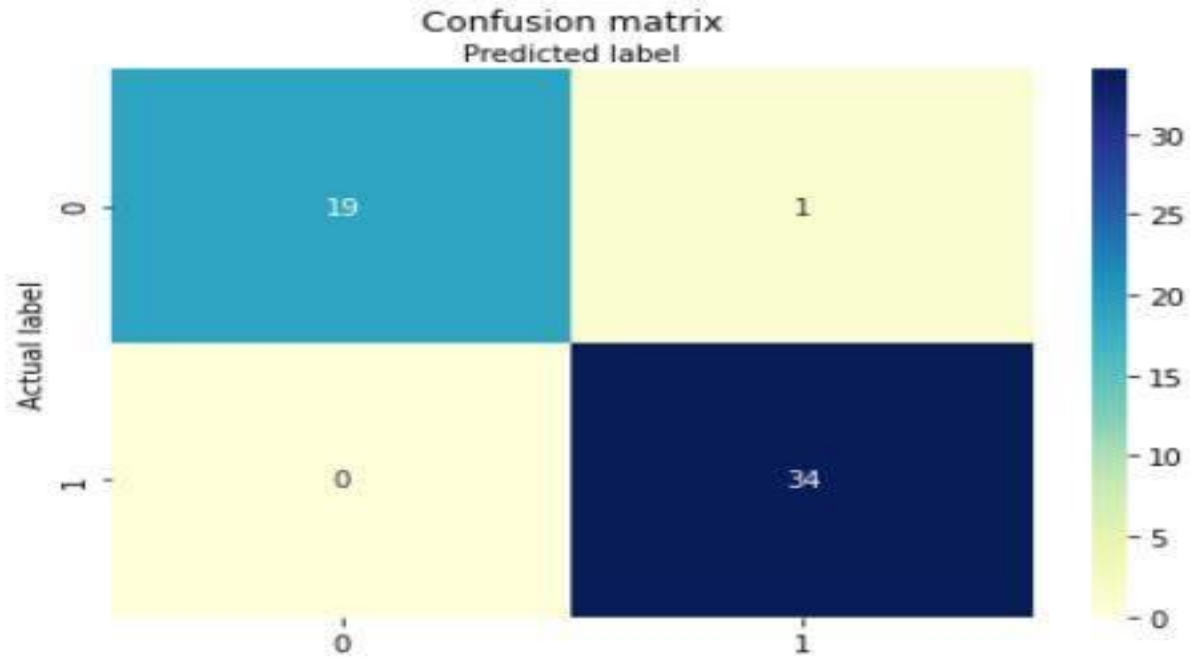
Fig 4.19: Confusion Matrix

4.3.3 Effect Of Result

From all those research we can say that, the effect of result will be depend how student manage their time. That means, how many times they spend on which activities. If students give more time to study then they will get good marks and able to be make a good result. But if students give more time to other activities like social media, video gaming, leisure time, then they not be able to get a good marks. As a results , the won't be able to make a good result. So it's very important to follow the routine for the students for make a good result.

4.4 Discussion

Here In this chapter we discussed about the results of the survey of performance and the results of the survey include analysis of experimental results, and also expressive analysis, this research questions. We get effective responses that makes research to become helpful the output of this expressive and frequency analysis. Also Logistic Regression gave the best result, the accuracy is (0.9814814814814815) 98.981%.

```
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

Accuracy: 0.9814814814814815
```

Fig 4.20: Final Accuracy

CHAPTER 5

SUMMARY AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

We have developed this research project to predict the final exam SGPA & CGPA of students and also discussed about the spending time in various ways of students, so how they spend their time and what result they could get we predict that with probability. The project has proved that the best data classifier algorithm for its research is Naive Byes classifier. Then we have collected all the required data from university students. Those data sets are collected from many university students and online surveys. After that, we sincerely applied data pre-processing rules so that we could make them perfect for the systems environment. Obviously datasets need to be trained so that those data can meet the data handling purpose. In our research works, some of the dataset are trained and some of the dataset are tested.

5.2 Conclusions

At this modern age, technology is so advanced. We all are probably technology dependent in our daily life. But it has some advantages & disadvantages at the same time. Controlling big data is a very complex problem in the technological field. Nowadays, the world goes so faster. This is the era of modernization & competition. In the education system there is too much competition between students. Basically, in the universities there is the educational system is different from school or college. In universities there has some curriculums or activities for students. Prediction of the SGPA or CGPA is very helpful project for students. We created this project using machine learning.

5.3 Recommendations

There are so many development steps can be happened by big data analysis like finding out impacts of other activities of students without study. The future of this world depends on Artificial Intelligence (AI).

In Artificial Intelligence, machine learning is the most important part. So, using machine learning to solve our problems is critical to the advancement of today's inventions for developing system.. In machine learning, data classifier impacts more important role. Data classifier must needed in the machine learning. Data classifier may change the entire concept about the earlier understandings regarding the device.

5.4 Implication for Further Study

- We will work with more data to make our research more efficient and convenient.
- We will include more categories of data to spread our research areas.
- We will build an open source working platform with a huge amounts of data.

REFERENCES

[1] Lubna Mahmoud Abu Zohair- "Prediction of Student's performance by modelling small dataset size", 2019.

[2] Navin Venkat & Sahaj Srivastava– "Predicting Student Grades using Machine Learning", Nov - 2018.

[3] Khurum Nazir Junejo & Eman Eman– "Grade Prediction Using Supervised Machine Learning Techniques". – 2014.

[4] Prepare organizational data in Workplace Analytics ...docs.microsoft.com

[5] How to Write a Research Methodology in Four Steps - Scribbrwww.scribbr.com

APPENDIX
RESEARCH REFLECTION 1

We are doing this project for the benefit of the university students to predict their SGPA & CGPA easily. That why, students will have the advantage of making good results. When, a student can predict their final SGPA & CGPA by using our project they will be able to take good preparation for a good result. During project activities, we faced several problems. But three problems ware major among them. one is selecting best algorithm, creating survey question and another is data collecting. Before working with Naïve Byes algorithm we tried many ways to solve our problem and we failed to get best output. When we get to create survey question we face some problem. One of them is to create question and select question. To collecting data, we faced too much difficulty. Because locally data collection is time consuming and hard for student. So, we have to collect some data from online sources. And after a long time and a lot of attempts and hard work we got succeed.

Our data collection question given below.

Data collection for predicting of the final exam SGPA and CGPA result by machine learning.

* Required

Your current semester? (আপনার বর্তমান সেমিস্টার?) *

```
Your answer
```

What is the average of your quiz? (আপনার কুইজের গড় কত?) *

```
Your answer
```

How many marks did you get in the midterm? (মিডটার্মে আপনি কত নম্বর পেয়েছেন?) *

```
Your answer
```

How many marks did you get in the assignment? (অ্যাসাইনমেন্টে আপনি কত নম্বর পেয়েছেন?) *

1
2
3
4
5
don't know

How many marks did you get in the presentation? (উপস্থাপনায় আপনি কত নম্বর পেয়েছেন?) *

1
2
3
4
5
6
7

don't know

What is your current CGPA? (আপনার বর্তমান সিজিপিএ কত?) *

Your answer

What is your current SGPA? (আপনার বর্তমান এসজিপিএ কত?) *

Your answer

How many hours do you study academically every day? (প্রতিদিন কত ঘণ্টা একাডেমিক পড়াশোনা করেন?) *

Your answer

How many hours do you spend online every day? (প্রতিদিন অনলাইনে কত ঘন্টা অতিবাহিত করেন?) *

Your answer

How many hours do you spend on mobile every day? (প্রতিদিন মোবাইলে কত ঘন্টা অতিবাহিত করেন?) *

Your answer

How many hours do you spend on the laptop/computer every day? (প্রতিদিন ল্যাপটপ/কম্পিউটার এ কত ঘন্টা অতিবাহিত করেন?) *

Your answer

How many hours do you spend every day on social media? (সামাজিক যোগাযোগ মাধ্যমে প্রতিদিন কত ঘন্টা অতিবাহিত করেন?) *

How many hours do you spend online each day for academic study? (একাডেমিক পড়াশোনার জন্য প্রতিদিন কত ঘন্টা অনলাইনে অতিবাহিত করেন ?) *

How many hours of leisure time do you spend every day? (প্রতিদিন কত ঘণ্টা অবসর সময় কাটান?) *

[ ]
Your answer

What do you spend most of your time on mobile? (মোবাইলে কোন কাজে আপনি বেশি সময় অতিবাহিত করেন?) *

Study
Social media
Gaming
Other

What do you spend most of your time on the laptop/computer? (ল্যাপটপ /কম্পিউটার এ কোন কাজে আপনি বেশি সময় অতিবাহিত করেন?) *

Study
Social media
Gaming
Other

Do you study routinely every day? (আপনি কি প্রতিদিন রুটিন অনুযায়ী পড়াশোনা করেন?) *

Yes
No

Do you attend classes regularly? (আপনি কি ক্লাসে নিয়মিত?) *

Yes
No

Do you take notes in class? (আপনি কি ক্লাসে নোট করেন?) *

Yes
No

Do you listen carefully to the teacher in class? (আপনি কি ক্লাসে শিক্ষকের কথা মনোযোগ দিয়ে শোনেন?) *

Yes
No

Do you just study the day before the exam? (আপনি কি শুধু পরীক্ষার আগের দিন পড়াশোনা করেন?) *

Yes
No

Do you take the help of a teacher/classmate if you do not understand any subject of academic study? (একাডেমিক পড়াশোনার কোন বিষয় না বুঝলে আপনি কি শিক্ষক/সহপাঠীর সাহায্য নেন?) *

Yes
No

Submit

Machine Learning Approach to Predict SGPA and CGPA