STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING ALGORITHM BY

MD. AMIR HAMJA
ID: 181-15-2064

ARAFAT UDDIN
ID: 181-15-2032

MD.NUR YOUSUF
ID: 181-15-1750

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

Supervised By

Saima Afrin

Lecturer

Department of CSE
Daffodil International University

Co-Supervised By

Md.Sabab Zulfiker

Lecturer (Senior Scale)

Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This Project titled "**Student Performance Prediction Using Machine Learning Algorithm**", submitted by Md.Amir Hamja, Arafat Uddin and Md.Nur Yousuf 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 17 January 2022.
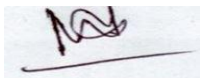
<u>BOARD OF EXAMINERS</u>

_____   **Chairman**
Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Narayan Ranjan Chakraborty   **Internal Examiner**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

_____   **Internal Examiner**
Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University
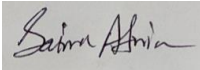
_____   **External Examiner**
Dr. Mohammad Shorif Uddin
Professor
Department of Computer Science and Engineering
Jahangirnagar University

Daffodil International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Saima Afrin, Lecturer, 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:

Saima Afrin
Lecturer
Department of CSE
Daffodil International University
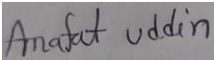
Co-Supervised by:

Md.Sabab Zulfiker
Lecturer (Senior Scale)
Department of CSE
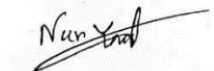Daffodil International University

Submitted by:

Md.Amir Hamja
ID: 181-15-2064
Department of CSE
Daffodil International University

Arafat Uddin
ID: 181-15-2032
Department of CSE
Daffodil International University

Md.Nur Yousuf
ID: 181-15-1750
Department of CSE
Daffodil International University

Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Saima Afrin**, **Lecturer**, **Department of CSE Daffodil International University, Dhaka.** Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" 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 our heartiest gratitude to our honorable 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

In our Research-Based Project, we are predicting student performance of students using a Machine Learning Algorithm. We have collected necessary data from the students through a Google form. We asked them several questions to them like their average CGPA, academic results, study hours, attendance percentage, etc. And using the machine learning algorithm we will predict that the student is good or bad according to their class performance, academic results, and overall CGPA. After collecting data through Google form we cleaned these data and also removed unnecessary data. We have selected some attributes and for that reason, we applied different algorithms. And we will apply different of algorithm operations and we will analyze the results. which one's accuracy is better. After analyze we can predict why students are performing good results badly why they are performing badly. For this project, we use Google co-lab. Through that, we are predicting the student's performance and also the student is good or bad. Besides we can predict also students' upcoming CGPA, the probability of going for higher education.

TABLE OF CONTENTS

Daffodil International University

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1.1 Introduction

Machine learning is the method, which a machine learns and predicts data from a data set on its own and then understands how to improve the data. Machine learning is currently being used in various research works and plays a different role in predicting something. We are no exception. We will use the machine learning algorithm to find a possible result based on the students' academic results, class performance, semester results, attendance percentage, etc., which will tell the students the possible results later. We will discuss this in more detail later.

1.2 Motivation

Our motivation is to explain to those who do bad results in university that their results are getting worse. They have to get out of this and do their results well. Many times students do not realize that their results are getting worse. Through our project, they understand their results and their classroom performance. This will allow them to improve themselves and increase their performance. Some students are often depressed with poor results, do not spend time studying, have other reasons including family problems. From our research, we can find out who is a good student or a bad student. So how can he be a good student with his performance is our main goal.

1.3 Rationale of the Study:

Through this research, we will use machine learning algorithms. We will use different branches of machine learning in our project. We will use different algorithms of machine learning in this study. Through algorithm implication, we will get the output by analyzing our datasets.

1.4 Research Related Questions:

How helpful our research will be for students?

What type of data is needed for our research?

How we will implement this research in future?

1.5 Research Output:

Our expected output is which students are ahead and who is behind in terms of studies. There will be an output based on which we can select and verify the students. And he does well based on verification. And based on their semester results, class performance, we will give a possible SGPA prediction. Some students who will go abroad for higher education in the future, we will also predict them based on their studies and their results and class performance.

1.6 Project Management:

Firstly, we have decided to do a Performance Prediction of Student Using Machine Learning Algorithm which is Research-Based project. Through this project, we will be able to pedicure the results of a student studying in a university. Had to face it because it turns out that almost all of our students have had a hard time putting us together to fill out this form to motivate them to provide us with exactly what they need and to fill out this Google form to help us with our research work. Then we clean our data in various training methods such as mapping and mapping and level in coding then we will call Google will give our results by applying different machine learning algorithms. We had a lot of trouble collecting data for the Covid-19 situation. It took us about 15 days to complete the data. The three of us worked hard on this. We have spent huge labor moments. For collect data with Google Forms link through Facebook Instagram WhatsApp etc. We did not spend any money to do this project, only our labor cost.

CHAPTER 2

2.1 Literature Review

Havan Agrawal, Harshil Mavani [1] their study shows that they used machine learning to make predictions on their past results according to the subject. They used the KNN algorithm. They tested the data train through this study. They worked with demographic data and used 5 algorithms. They have an average of 8 accuracies 296, an accuracy rating of 80.48%.

Dorina Kabakchieve [2] they have used data mining algorithms based on their data, student performance through data mining for student performance, the neural network in their data. The accuracy of their pre-processed data is 8%, 63%.

Anal Acharya, Devadatta Sinha [3] in their research, the students who will get scholarships and those who will fail have been predicted. They have used Decision Tree AI in this research. They used different instances and they found that less than half got F grades. They will use genetic algorithms.

Karani Kardaş, Altay Güvenir [4] In their paper, they made predictions based on their quiz on the programming language course, as well as all class performances. They have used linear support vector algorithms with logistic regression to do this work. They need to be further improved.

Erkan Er [5] In this paper, they have used different algorithms for the level of performance of the students through which they get some good output. They have determined the accuracy of this result. The result of which is better is the better accuracy of the algorithm. They used KNN and K Star to get better results.

Amirah Mohamed Shahiri, Wahida Hussain,Nur Aini abdul rashid [6] This research paper gives the students a prediction of their own performance using machine learning algorithm. This project has been done among the higher students because at this stage many of the results are bad due to various reasons. Many fall at this age due to family, environmental or personal reasons. They are concerned about this and hope to do something for them which is why their team members are interested in doing this research based project of student performance prediction.

Arwa Batoaq,Olatunj [7] Classification is a very common method for classifying data that they have used in this research paper. The decision tree algorithm has been used. The data is basically collected from the academic office where the names of 6 major departments are

Daffodil International University

mentioned. The model pattern informs the students about the performance. The results of this research paper are 2% got good grade and rest of them got bad grades, out of every 100 students.

Huda al-shehri [8] In this research paper, 392 data samples have been used to create 2 prediction model patterns for the final exam of the students. Kn and svm are used here. It's main purpose is to improve student performance based on the features of the datasets collected .Here,SVM algorithm is better than K99.

Abir Hussain Robert Keight Wasiq khan Raghad Alshabandar [9] This research paper predicts the performance of students through online classes. Through various innovations, such as online classes, presentations, exams, etc., students can acquire skills. Through this research paper, the success rate of students online can be observed. Here, the predictions have been made based on the class lectures and the class performances of the students. Two sets of studies have been done by observing regression and classification. The gap between the traditional offline class and the online class's seam performance is or may be the subject of very skillful research. It also features the ability to predict students' assessment grade models for future research.

Jabeen sultana, usha rani,farquad[10] In this research paper, research has been done on the backward education system of the country and the way out of it using the data mining method. At present, the education system is getting worse day by day. If it can be well predicted then some reduction is possible, It has been observed that Mlp,Dlt,Dmt,Rf have given good results. The accuracy of all, Decision trees and Random Forest is 99.45%, 99.61%, and 100% respectively. Which gave better results than SVM, Ibk, lazy Lwl.

O, Obagbuwa, I.C [11] In this research paper, the academic performance prediction of students has been made using a k-mean Clusturing algorithm. In addition to this, the research done through the data collection of 69 students here will play a big role in the education system of future students. The clustering algorithm is providing the Student Performance Progression to take it to a higher level.

Fishera Berhanu, Addisalem Abera [12]   This research paper is an academic record performance prediction for students .They use data mining algorithms. Their entire dataset of their papers are divided into two parts Pre-processing tasks like cleaning, integration, etc. have been done before applying the data mining algorithm as they are required. Data from 2 types of Data Source 1 and Data Source 2 are integrated and prepared for preprocessing. After pre-processing, a data mining algorithm is applied. Then we analyze based on the results obtained.

Elaf abu amrich,thair hamtini,ibrahim aljarah [13] In this research paper, practical features of students have been worked out and the dataset has been collected. Various studies study the development of educational standards but rarely discuss their mental health and behavior Students are divided into 3 levels.

Murat pojon [14] This research paper is based on student activity in class . This thesis is basically done using linear regression, decision tree algorithm, and naive Bayes classifier algorithm. Linear regression creates a definite set between a dependent variable and an independent variable. Obtained g.p.a. If 3+ then yes otherwise it will be no. Also, fail if 15+ years otherwise pass.

2.2 Research Summary:

Our goal is Student Performance using Machine Learning. There has been a lot of work like this in the past. They have done it through different types. They have used different algorithms including data mining machine learning algorithms. Targeted on that with that data they have given different types of pedicures at any time they have seen what they are passing on to pass on any subject. We have seen that they have collected data offline and online. They have used it in various operations because they have not been accounted for. They have used different techniques and different algorithms. We tried our best to get good accuracy. We collected data online through Google form which was very challenging for us. We have analyzed their research papers and seen that many of them have used different types of volume-based software. Through this software, they have shown these works through different songs. The work that we will do is we have been collecting data through that data and we are working on how we can make our work better with ideas from the work that was related to this.

2.3 Research Challenges:

At first, the challenge we had was to collect data. We had to go to the students to collect the data. And they've been told to fill out a Google survey form and shares with their circle of friends. But, not everyone fills out the form. It was challenging. For this, we went to the hostel of the varsity and went from room to room, and collected data from the students. We had to get a lot of speed in doing this Many have given wrong data and some wrong instances that we have had to remove and restore. We started working at Google Co-Lab but had to implement the Anaconda Jupiter Notebook due to some technical glitches. We have faced the above problems to do this work.

CHAPTER 3

3.1 Introduction:

Our purpose for this research is to know the condition of a student based on their class performance. To get an idea about their performance based on their result. A possible number to know how many people can go abroad for higher education. For this, we have collected data from university students. We have cleaned and processed this data. We will describe the details in more detail later.

Figure 3.1.1: Google survey form

Enter Your H.S.C GPA *

Your answer

(!) This is a required question

Your Current Semester *

Choose ▼

Your Current S.G.P.A *

Your answer

Your Overall C.G.P.A *

○ 2.00-3.00

○ 3.00-3.50

○ 3.50-4.00

Figure 3.1.2: Google survey form

Your Attendence Percentage *

○ 0-30%

○ 30-50%

○ 50-75%

○ 75-100%

Time You Spend In Studying *

○ 0-2 Hours

○ 2-3 Hours

○ 3-4 Hours

○ 4-6 Hours

○ Above 6 Hours

Does Your Course Teacher Provide Full Course Materials To You *

○ Yes

○ No

Figure 3.1.3: Google survey form

How Much Do You Understand About Your Course Material *

○ 0-40%

○ 40-60%

○ Above 60%

Do You Face Any Problem During Studying *

○ Yes

○ No

Do you have any publications in any conference paper, journal or anywhere based on research area? *

○ Yes

○ No

How Many Publications You Have?

○ 1-2

○ 2-4

○ Above 4

○ No Publications.

Figure 3.1.4: Google survey form

Daffodil International University

Do You Properly Engage Your Assignments,Quizes,Lab-Reports & Presentations? *

○ Yes

○ No

Do You Feel Any Pressure In Your Study? *

○ Yes

○ No

If Yes,Please Specify Your Problem-

○ Family Issue

○ Pandemic Situation

○ Anxiety

○ Depression

Do You Face Any Financial Problem? *

○ Yes

○ No

Figure 3.1.5: Google survey form

Daffodil International University

How Much Time Do You Spend In Social Media? *

○ Above 15 Hours

○ 11-15 Hours

○ 5-10 Hours

○ 1-5 Hours

○ None of the Above
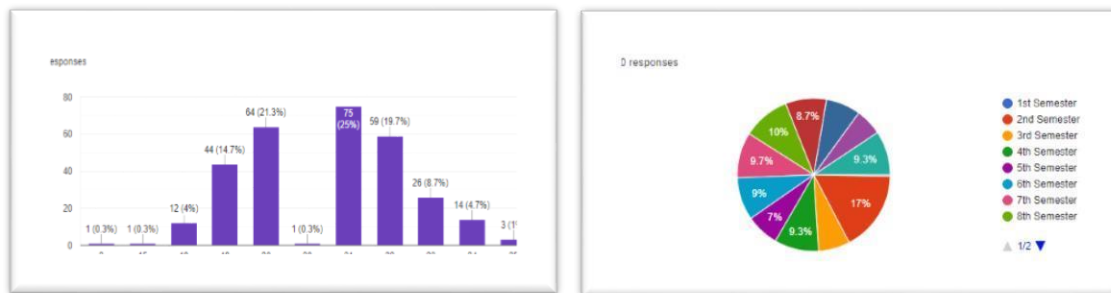
How Much Time You Spent in Games(PC/Mobile Games) *

○ Above 15+ Hours

○ 10-15 Hours

○ 5-10 Hours

○ 1-5 Hours

○ None of the Above

Submit   Clear form

Never submit passwords through Google Forms.
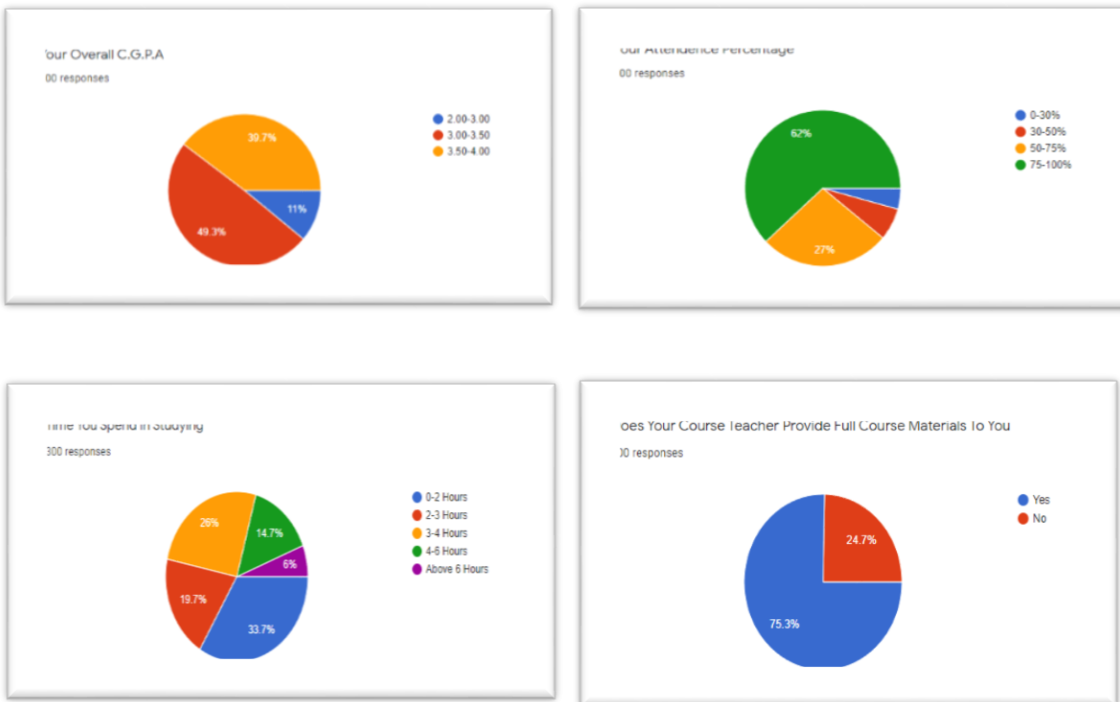
Figure 3.1.6: Google survey form

Daffodil International University

Fig.3.1.7 Graph and pie Chart of data

Do you have any publications in any conference paper, journal or anywhere based on research area?

300 responses

- Yes
- No

80.3%

19.7%

How Many Publications You Have?

298 responses

- 1-2
- 2-4
- Above 4
- No Publications

80.5%

12.1%

Do You Properly Engage Your Assignments,Quizes,Lab-Reports & Presentations?

299 responses

- Yes
- No

82.9%

17.1%

Do You Feel Any Pressure In Your Study?

300 responses

- Yes
- No

50.7%

49.3%

Figure 3.1.8: Pie Chart of data

If Yes,Please Specify Your Problem-

195 responses

- Family Issue
- Pandemic Situation
- Anxiety
- Depression
- Pandamic Situation

12.8%

27.2%

40.5%

19%

Do You Face Any Financial Problem?

300 responses

- Yes
- No

53.7%

46.3%

Daffodil International University
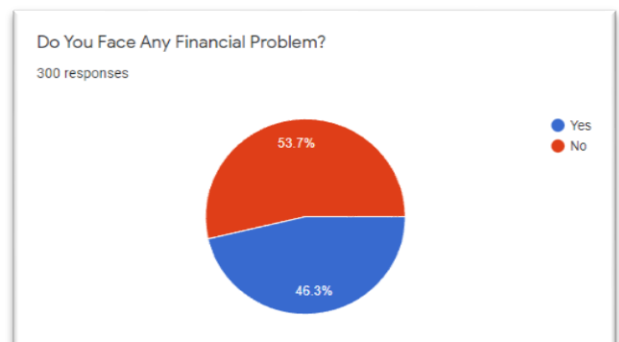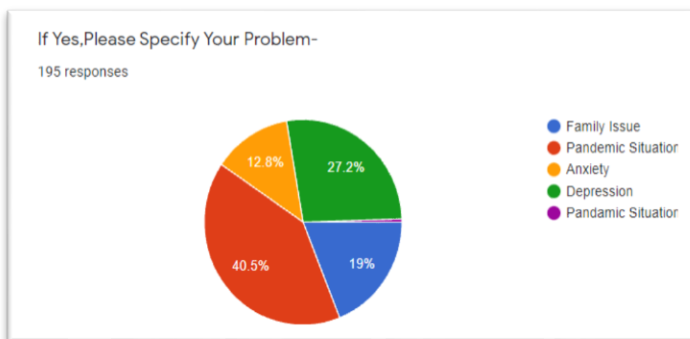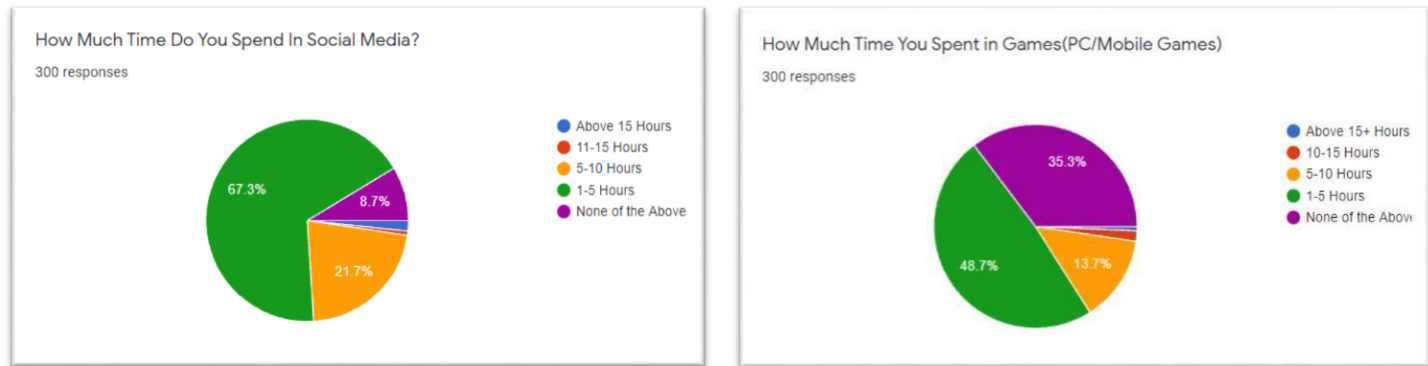
Figure 3.1.9: Pie Chart of data

3.2 Data Analysis

After data collection, we analyzed our data set. Our data set includes the nominal data type and numeric data type. In our data sheet no garbage value allowed. We have created the form in an easy way for students to understand. The attributes we have used in this work are given below in the table:

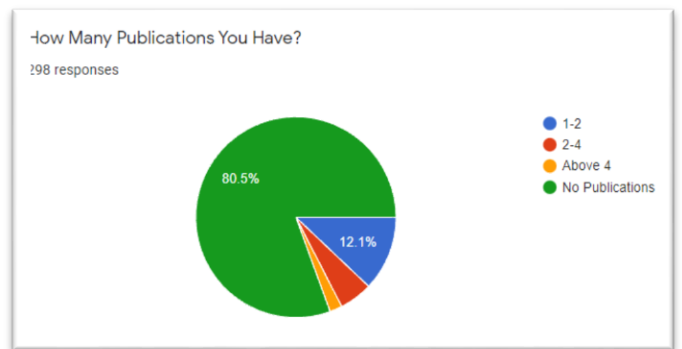| Serial Number	Attribute Name	Data type
1	Age	Numeric
2	S.S.C Grade	Nominal and Numeric
3	H.S.C Grade	Nominal and Numeric
4	Current Semester	Nominal
5	Current S.G.P.A	Nominal and Numeric

Daffodil International University

6	Overall C.G.P.A	Nominal and Numeric
7	Attendance Percentages	Nominal and Numeric
8	Study Time	Nominal and Numeric
9	Providing Course Material	Nominal and Numeric
10	Understanding Course Material	Nominal and Numeric
11	Problem during Studying	Nominal and Numeric
12	Publications	Nominal
13	Publication Quantity	Nominal and Numeric
14	Class Activity	Nominal and Numeric
15	Pressure During Study	Nominal
16	Pressure Type	Nominal
17	Financial problem	Nominal
18	Spending Time in social media	Nominal and Numeric
19	Gaming Time	Nominal and Numeric

Table 3.2.1: Attribute

3.3 Data Collection Agenda:

We created an online-based Google form. Where they were asked different types of questions is based on our research. We do not have any personal data here. As a student, we have figured out what subjects depend on a student's performance and we have created such questions in the form of Google forms and we have selected attributes. We have 24 attributes and 320 instances. We look at their age, their SSC, HSC GPA, their semester results, their average CGPA, how much time they study each day, their class performance rate, whether they have a research paper or publication. How much time do they spend on social media, How much time do they play online games, what is their attendance in class, what % of their course do they understand, how many class performances including regular quizzes, assignments are these questions asked in Google form.

Daffodil International University

3.4 Algorithm Description:

We have applied seven algorithm Logistic Regression, RF, KNN, SVM, AdaBoost algorithm, Multinomial Naive Bayes algorithm, Gradient boosting etc. Among them top 5 algorithm has batter accuracy they are LR, KNN, SVM, AdaBoost algorithm.

3.5 LR

Logistic regression is a method by which we can learn about the relationship of an independent variable of a data set.

3.6 RF

We can compare this algorithm with a forest. Just as there are many trees in a forest, this algorithm gives a perfect prediction by running between different decision trees.

CHAPTER 4
Result and Discussion

We have used seven algorithms in our research. In which we checked the accuracy and saw that the maximum accuracy comes in the logistic regression and AdaBoost algorithm. We have applied the Logistic Regression algorithm. We have used logistic regression classifies. We got 72% accuracy in logistic regression. After that, we have used the Random forest algorithm. We have imported a random forest classifier. Here we got 72% accuracy. Which is equal to logistic regression. After that, we have used the SVM algorithm. We have imported a random forest classifier. Here we got 70% accuracy. Which is less than logistic regression and random forest. After that, we have used the svm algorithm. We have imported a random forest classifier. Here we got 71% accuracy. Which is less than logistic regression and random forest but greater than SVM. After that we have used multinomial naïve bayes algorithm. We have imported a random forest classifier. Here we got 66% accuracy. Which is less then logistic regression and random forest but greater than svm. After that we have used AdaBoost algorithm. We have imported a random forest classifier. Here we got 72% accuracy. Which is less then logistic regression, multinomial naïve bayes and random forest and equal to svm.After that, we have used Gradient Boosting algorithm. We have imported a random forest classifier. Here we got 68% accuracy. Which is less then logistic regression, multinomial naïve Bayes and random forest and greater than SVM.

No	Classifier Name	Accuracy	Precision	Recall	Fi-Score	Support
01	Logistic regression	72%	71%	59%	65%	34%
			73%	83%	78%	46%
02	Random Forest	71%	66%	68%	67%	34%
			76%	74%	75%	46%
03	K nearest neighbor	71%	76%	47%	58%	34%

			69%	89%	78%	46%
04	SVM	70%	67%	59%	62%	34%
			72%	78%	75%	46%
05	Multinomial Naive Bayes	66%	62%	53%	57%	34%
			69%	76%	72%	46%
06	AdaBoost	72%	73%	56%	63%	34%
			72%	85%	78%	46%
07	Gradient Boosting	69%	65%	59%	62%	34%
			71%	76%	74%	46%

Table 4.1.1: Accuracy

```
In [220]:  # Logistic regression
           logRegression=LogisticRegression()
           logRegression.fit(X_train,y_train)

           y_predictionLogistic=logRegression.predict(X_test)

           print("Accuracy: ",metrics.accuracy_score(y_test, y_predictionLogistic))
           print("\nConfusion Matrix:")
           print(metrics.confusion_matrix(y_test,y_predictionLogistic))
           print("\nReport for Logistic: ")
           print(metrics.classification_report(y_test,y_predictionLogistic))

           Accuracy:  0.725
```

We have applied Logistic Regression algorithm. We have used logistic regression classifies. We got 72% accuracy in logistic regression.

```
In [221]:  # Random forest
           from sklearn.ensemble import RandomForestClassifier

           RndmForestClassification=RandomForestClassifier(random_state=10,n_estimators=20)
           RndmForestClassification.fit(X_train,y_train)
           y_predictionRndmForest=RndmForestClassification.predict(X_test)
           print("Accuracy: ")
           print(metrics.accuracy_score(y_test, y_predictionRndmForest))
           print("\nConfusion Matrix")
           print(metrics.confusion_matrix(y_test,y_predictionRndmForest))

           print("\nReport for Random Forest: ")
           print(metrics.classification_report(y_test,y_predictionRndmForest))

           Accuracy:
           0.7125
```

After that we have used Random forest algorithm. We have imported random forest classifier. Here we got 72% accuracy. Which is equal to logistic regression.

```
In [223]: # svm
          from sklearn import svm

          svmclassifier = svm.SVC(kernel='linear')
          svmclassifier.fit(X_train, y_train)
          y_pred = svmclassifier.predict(X_test)

          print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
          print("\nConfusion Matrix")
          print(metrics.confusion_matrix(y_test,y_pred))
          print("\nReport for SVM: ")
          print(metrics.classification_report(y_test,y_pred))

          Accuracy: 0.7
```

After that we have used svm algorithm. We have imported random forest classifier. Here we got 70% accuracy. Which is less then logistic regression and random forest.

```
In [222]: # KNN
          from sklearn.neighbors import KNeighborsClassifier

          knnClassification = KNeighborsClassifier(n_neighbors=60)
          knnClassification.fit(X_train, y_train)
          y_pred = knnClassification.predict(X_test)

          print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
          print("\nConfusion Matrix")
          print(metrics.confusion_matrix(y_test,y_pred))
          print("\nReport for KNN: ")
          print(metrics.classification_report(y_test,y_pred))

          Accuracy: 0.7125
```

After that we have used svm algorithm. We have imported random forest classifier. Here we got 71% accuracy. Which is less then logistic regression and random forest but greater than svm.

Daffodil International University

```
In [225]: #multinomial naive bayes
          from sklearn.naive_bayes import MultinomialNB
          multinomial = MultinomialNB()
          multinomial.fit(X_train, y_train)
          y_pred = multinomial.predict(X_test)

          print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
          print("\nConfusion Matrix")
          print(metrics.confusion_matrix(y_test,y_pred))
          print("\nReport for Multinomial Naive Bayes: ")
          print(metrics.classification_report(y_test,y_pred))

          Accuracy: 0.6625
```

After that we have used multinomial naïve bayes algorithm.We have imported random forest classifier. Here we got 66% accuracy.Which is less then logistic regression and random forest but greater than svm.

```
In [226]: #AdaBoost
          from sklearn.ensemble import AdaBoostClassifier
          adaBoodt = AdaBoostClassifier(n_estimators=100, random_state=0)
          adaBoodt.fit(X_train, y_train)
          y_pred = adaBoodt.predict(X_test)

          print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
          print("\nConfusion Matrix")
          print(metrics.confusion_matrix(y_test,y_pred))
          print("\nReport for AdaBoodt: ")
          print(metrics.classification_report(y_test,y_pred))

          Accuracy: 0.725
```

After that we have used AdaBoost algorithm. We have imported random forest classifier. Here we got 72% accuracy. Which is less then logistic regression, multinomial naïve bayes and random forest and equal to svm.

```
In [224]: # Gradient Boosting
          from sklearn.ensemble import GradientBoostingClassifier
          gradientboosting = GradientBoostingClassifier(random_state=1)
          gradientboosting.fit(X_train, y_train)
          y_pred = gradientboosting.predict(X_test)

          print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
          print("\nConfusion Matrix")
          print(metrics.confusion_matrix(y_test,y_pred))
          print("\nReport for Gradient Boosting: ")
          print(metrics.classification_report(y_test,y_pred))

          Accuracy: 0.6875
```

After that we have used Gradient Boosting algorithm. We have imported random forest classifier. Here we got 68% accuracy. Which is less then logistic regression, multinomial nb and rf and greater than svm.
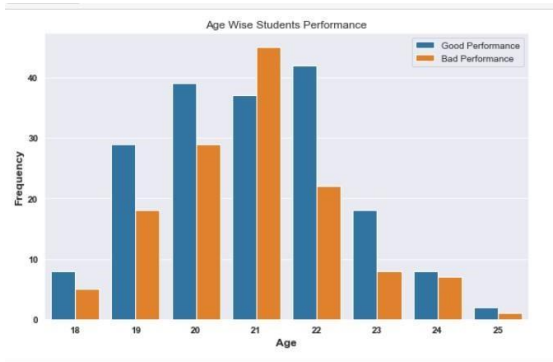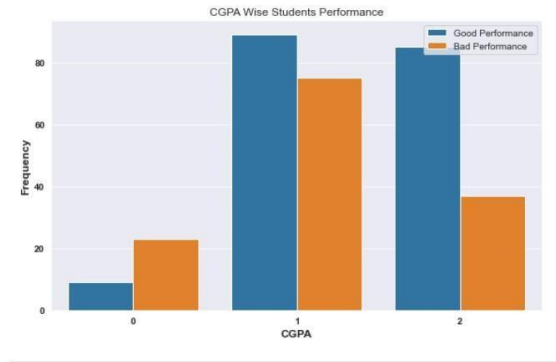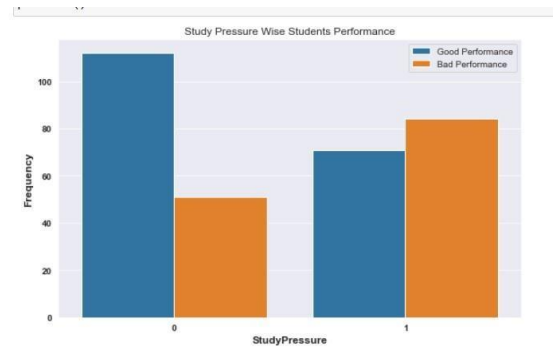
Daffodil International University

FIG.**4.1**

FIG.**4.2**

FIG.**4.3**
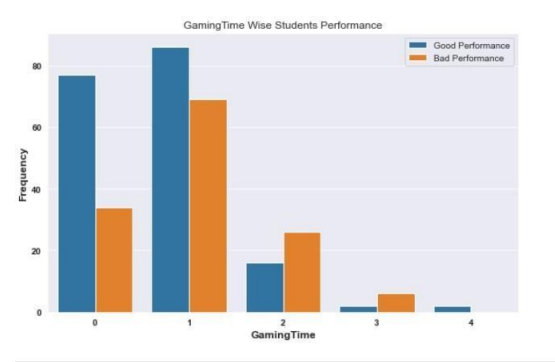
FIG.**4.4**
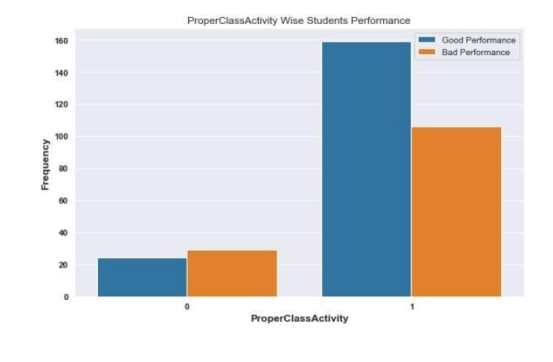
FIG.**4.5**
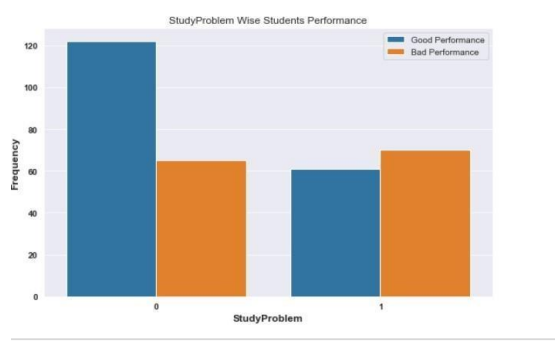
FIG.**4.6**

Fig4: Student performance in different scenarios

Daffodil International University

Fig 4.1 shows that age wise student performance. Here is the age-based performance of the students which shows that the 18-year-old students gave good performance compared to the bad one. This graph shows that the worst performance was of 21-year-old students. The best performance was of 22-year-old students.

Fig 4.2 This graph shows the performance of students in the CGPA sequence. It is seen that for those whose CGPA is 3+ their performance is good and those whose 3.50+ are also good. But those with a CGPA below 3 perform relatively poorly.

Fig 4.3 This graph shows how the study pressure falls on a student and how this pressure falls on a student's performance. This graph shows that those who do not have study pressure have very good performance whereas those who have high pressure have relatively low academic performance.

Fig 4.4 This graph shows that those who do not play the game at once have a much better reading performance, and those who play more than 5 hours have worse performance.

Fig 4.5 By this graph class activity is meant; it is seen here that those who have good class activity have better performance whereas those who have less class activity have less performance.

Fig 4.6 This graph of study problems shows that the performance of those who have a problem is relatively poor. Those who do not have a study problem.

CHAPTER 5
RECOMMENDATIONS & CONCLUSIONS

5.1 Upcoming Work

We have planned a very big future for this research, we will also create an app in the future, where one will get the output after inputting the data in question periodically. I will add something new using ML and Python so that more accuracy will come later and acceptance will also increase. After all, there are a lot of plans for this project that we, our team members, will take further in the coming days.

5.1 Conclusions

Finally, by the grace of God, we have successfully completed the project. Among our various algorithms, the accuracy of the LR algorithm and AdaBoost has been relatively high. Here, using graphs from place to place, various topics of the students have been boiled down very well and at the same time, it has been analyzed. This will allow us to know in advance about the rules of the speed of students' performance. Which will be very useful for us in the future for the new generation.

REFERENCES

[1] Agrawal, H. and Mavani, H., 2015. Student performance prediction using machine learning. International Journal of Engineering Research and Technology, 4(03), pp.111-113.

[2] Kabakchieva, D., 2012. Student performance prediction by using data mining classification algorithms. International journal of computer science and management research, 1(4), pp.686-690.

[3] Acharya, A. and Sinha, D., 2014. Early prediction of students performance using machine learning techniques. International Journal of Computer Applications, 107(1).

[4] Kardaş, K. and GÜVENİR, H.A., 2020. Kısa Sınavların, Ödevlerin ve Projelerin Dönem Sonu Sınavına Olan Etkilerinin Farklı Makine Öğrenmesi Teknikleri ile Araştırılması. EMO Bilimsel Dergi, 10(1), pp.2127.

[5] Er, E., 2012. Identifying at-risk students using machine learning techniques: A case study with is 100. International Journal of Machine Learning and Computing, 2(4), p.476.

[6] Shahiri, A.M. and Husain, W., 2015. A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72, pp.414-422.

[7] Oyelade, O.J., Oladipupo, O.O. and Obagbuwa, I.C., 2010. Application of k Means Clustering algorithm for prediction of Students Academic Performance. arXiv preprint arXiv:1002.2425.

[8] Amrieh, E.A., Hamtini, T. and Aljarah, I., 2015, November. Preprocessing and analyzing educational data set using X-API for improving student's performance. In 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) (pp. 1-5). IEEE.

[9] Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., Alhiyafi, J. and Olatunji, S.O., 2017, April. Student performance prediction using support vector machine and k-nearest neighbor. In 2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE) (pp. 14). IEEE.

[10] Alshabandar, R., Hussain, A., Keight, R. and Khan, W., 2020, July. Students Performance Prediction in Online Courses Using Machine Learning Algorithms. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

[11] Berhanu, F. and Abera, A., 2015. Students' performance prediction based on their academic record. International Journal of Computer Applications, 131(5), pp.09758887.

[12] Sultana, J., Usha, M. and Farquad, M.A.H., 2019. Student's performance prediction using deep learning and data mining methods. Int. J. Recent Technol. Eng.(IJRTE)(1S4), pp.1018-1021.

[13] Pojon, M., 2017. Using machine learning to predict student performance (Master's thesis).

[14] Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V.V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C. and Liao, S.N., 2018, July. Predicting academic performance: a systematic literature review. In Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education (pp. 175199).

PLAGIARISM REPORT

STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING ALGORITHM

ORIGINALITY REPORT

22% SIMILARITY INDEX	**20**% INTERNET SOURCES	**10**% PUBLICATIONS	**18**% STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	8%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	eprints.utar.edu.my Internet Source	2%
4	Submitted to University of Leicester Student Paper	1%
5	www.testmagzine.biz Internet Source	1%
6	Submitted to University of Derby Student Paper	1%
7	journals.covenantuniversity.edu.ng Internet Source	1%
8	Submitted to University of East London Student Paper	1%
9	Submitted to De Montfort University Student Paper	1%

10 S. Anthony Philomen Raj, Vidyaathulasiraman. "Prioritization of e-learners activities using principal component analysis method", International Journal of Information Technology, 2021
Publication — 1%

11 assets.researchsquare.com
Internet Source — 1%

12 sersc.org
Internet Source — <1%

13 igbr.org
Internet Source — <1%

14 tesi.cab.unipd.it
Internet Source — <1%

15 e-space.mmu.ac.uk
Internet Source — <1%

16 d-nb.info
Internet Source — <1%

17 researchspace.ukzn.ac.za
Internet Source — <1%

18 Leena H. Alamri, Ranim S. Almuslim, Mona S. Alotibi, Dana K. Alkadi, Irfan Ullah Khan, Nida Aslam. "Predicting Student Academic Performance using Support Vector Machine and Random Forest", 2020 3rd International — <1%