**Mini Project Report on**



# **Academy Performance Prediction**



**Submitted in partial fulfilment of the requirement for the award of the degree**

**of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Abhishek Singh** **2017447**

***Under the Mentorship of***

**Ms. Meenakshi Maindola**

**Designation**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**July-2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Academic performance Prediction”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Ms. Meenakshi Maindola, Designation**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Abhishek Singh 2017447

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction |  |
| Chapter 2 | Literature Survey |  |
| Chapter 3 | Methodology |  |
| Chapter 4 | Result and Discussion |  |
| Chapter 5 | Conclusion and Future Work |  |
|  | References |  |

**Chapter 1**

**Introduction**

Educational systems need innovative ways to improve quality of education to achieve the best results and decrease the failure rate. Educational Data Mining (EDM) has boomed in the educational systems recently as it enables to analyze and predict student performance so that measures can be taken in advance. Due to lack of prediction accuracy, improper attribute analysis, and insufficient datasets, the educational systems are facing difficulties and challenges exist to effectively benefit from EDM. In order to improve the prediction process, a thorough study of literature and selection of the best prediction technique is very important. The main objective of this paper is to present a comparative study of various recently used data mining techniques, classification algorithms, their impact on datasets as well as the prediction attribute’s result in a clear and concise way. The paper also identifies the best attributes that will help in predicting the student performance in an efficient way.

* 1. **Introduction**

Improvement of student performance and enhancement of quality of education is of utmost importance for all educational institutes. To provide quality education to learners, deep analysis of previous records of the learners can play a vital role.

EDM involves analysis and improvement in the predication methods of student performance. Based on prediction results, if the student needs are fulfilled timely then the overall result and performance will increase year by year. For the purpose of performance analysis and prediction, important attributes and previous records of students are gathered. Subsequently, various data mining techniques and classification algorithms are applied to get deeper insights and predictions.

The purpose of EDM is to reduce the failure rate, improve the educational system and analyze the attributes, which are of key importance and consider the student success and performance. Moreover, it enables us to develop useful predictive models for the performance prediction. It does not only help to immediately take steps for betterment of at risk students but also provides information and insights for the next year planning of education process. In recent years, various data mining techniques and classification algorithms have been used such as Naïve Bayes, Decision tree, neural networks, outlier’s detections and advanced statistical techniques. These techniques are applied on the student data in order to get information, to help in decision support systems, and pattern extracting etc. Commonly the student’s academic performance is measured by previous CGPA but there are various other important attributes that affect the overall performance of the student. Recently various empirical and statistical based researches have been conducted on student’s dataset.

**Chapter 2**

**Literature Survey**

In various studies on EDM, e-learning systems have been successfully analysed (Lara et al., [2014](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR30)). Some studies have also classified educational data (Chakraborty et al., [2016](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR13)), while some have tried to predict student performance (Fernandes et al., [2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR18)).

Asif et al. ([2017](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR4)) focused on two aspects of the performance of undergraduate students using DM methods. The first aspect is to predict the academic achievements of students at the end of a four-year study program. The second one is to examine the development of students and combine them with predictive results. He divided the students into two parts as low achievement and high achievement groups. He have found that it is important for the educators to focus on a small number of courses indicating particularly good or poor performance in order to offer timely warnings, support underperforming students and offer advice and opportunities to high-performing students. Cruz-Jesus et al. ([2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR15)) predicted student academic performance with 16 demographics such as age, gender, class attendance, internet access, computer possession, and the number of courses taken. Random forest, logistic regression, k-nearest neighbours and support vector machines, which are among the machine learning methods, were able to predict students’ performance with accuracy ranging from 50 to 81%.

Fernandes et al. ([2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR18)) developed a model with the demographic characteristics of the students and the achievement grades obtained from the in-term activities. In that study, students' academic achievement was predicted with classification models based on Gradient Boosting Machine (GBM). The results showed that the best qualities for estimating achievement scores were the previous year's achievement scores and unattendance. The authors found that demographic characteristics such as neighbourhood, school and age information were also potential indicators of success or failure. In addition, he argued that this model could guide the development of new policies to prevent failure. Similarly, by using the student data requested during registration and environmental factors, Hoffait and Schyns ([2017](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR24)) determined the students with the potential to fail. He found that students with potential difficulties could be classified more precisely by using DM methods. Moreover, their approach makes it possible to rank the students by levels of risk. Rebai et al. ([2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR37)) proposed a machine learning-based model to identify the key factors affecting academic performance of schools and to determine the relationship between these factors. He concluded that the regression trees showed that the most important factors associated with higher performance were school size, competition, class size, parental pressure, and gender proportions. In addition, according to the random forest algorithm results, the school size and the percentage of girls had a powerful impact on the predictive accuracy of the model.

Ahmad and Shahzadi, ([2018](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR1)) proposed a machine learning-based model to find an answer to the question whether students were at risk regarding their academic performance. Using the students' learning skills, study habits, and academic interaction features, they made a prediction with a classification accuracy of 85%. The researchers concluded that the model they proposed could be used to determine academically unsuccessful student. Musso et al., ([2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR33)) proposed a machine learning model based on learning strategies, perception of social support, motivation, socio-demographics, health condition, and academic performance characteristics. With this model, he predicted the academic performance and dropouts. He concluded that the predictive variable with the highest effect on predicting GPA was learning strategies while the variable with the greatest effect on determining dropouts was background information.

Waheed et al., ([2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR44)) designed a model with artificial neural networks on students' records related to their navigation through the LMS. The results showed that demographics and student clickstream activities had a significant impact on student performance. Students who navigated through courses performed higher. Students' participation in the learning environment had nothing to do with their performance. However, he concluded that the deep learning model could be an important tool in the early prediction of student performance. Xu et al. ([2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR47)) determined the relationship between the internet usage behaviors of university students and their academic performance and he predicted students’ performance with machine learning methods. The model he proposed predicted students' academic performance at a high level of accuracy. The results suggested that Internet connection frequency features were positively correlated with academic performance, whereas Internet traffic volume features were negatively correlated with academic performance. In addition, he concluded that internet usage features had an important role on students' academic performance. Bernacki et al. ([2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR9)) tried to find out whether the log records in the learning management system alone would be sufficient to predict achievement. He concluded that the behaviour-based prediction model successfully predicted 75% of those who would need to repeat a course. He also stated that, with this model, students who might be unsuccessful in the subsequent semesters could be identified and supported. Burgos et al. ([2018](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR10)) predicted the achievement grades that the students might get in the subsequent semesters and designed a tool for students who were likely to fail. He found that the number of unsuccessful students decreased by 14% compared to previous years. A comparative analysis of studies predicting the academic achievement grades using machine learning methods.

A review of previous research that aimed to predict academic achievement indicates that researchers have applied a range of machine learning algorithms, including multiple, probit and logistic regression, neural networks, and C4.5 and J48 decision trees. However, random forests (Zabriskie et al., [2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR48)), genetic programming (Xing et al., [2015](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR46)), and Naive Bayes algorithms (Ornelas & Ordonez, [2017](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR35)) were used in recent studies. The prediction accuracy of these models reaches very high levels.

Prediction accuracy of student academic performance requires an deep understanding of the factors and features that impact student results and the achievement of student (Alshanqiti & Namoun, [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR2)). For this purpose, Hellas et al. ([2018](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR23)) reviewed 357 articles on student performance detailing the impact of 29 features. These features were mainly related to psychomotor skills such as course and pre-course performance, student participation, student demographics such as gender, high school performance, and self-regulation. However, the dropout rates were mainly influenced by student motivation, habits, social and financial issues, lack of progress, and career transitions.

The literature review suggests that, it is a necessity to improve the quality of education by predicting the academic performance of the students and supporting those who are in the risk group. In the literature, the prediction of academic performance was made with many and various variables, various digital traces left by students on the internet (browsing, lesson time, percentage of participation) (Fernandes et al., [2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR18); Rubin et al., [2010](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR39); Waheed et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR44); Xu et al., [2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR47)) and students demographic characteristics (gender, age, economic status, number of courses attended, internet access, etc.) (Bernacki et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR9); Rizvi et al., [2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR38); García-González & Skrita, [2019](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR20); Rebai et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR37); Cruz-Jesus et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR15); Aydemir, [2017](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR5)), learning skills, study approaches, study habits (Ahmad & Shahzadi, [2018](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR1)), learning strategies, social support perception, motivation, socio-demography, health form, academic performance characteristics (Costa-Mendes et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR14); Gök, [2017](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR21); Kılınç, [2015](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR28); Musso et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR33)), homework, projects, quizzes (Kardaş & Güvenir, [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR26)), etc. In almost all models developed in such studies, prediction accuracy is ranging from 70 to 95%. Hovewer, collecting and processing such a variety of data both takes a lot of time and requires expert knowledge. Similarly, Hoffait and Schyns ([2017](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR24)) suggested that collecting so many data is difficult and socio-economic data are unnecessary. Moreover, these demographic or socio-economic data may not always give the right idea of preventing failure (Bernacki et al., [2020](https://slejournal.springeropen.com/articles/10.1186/s40561-022-00192-z#ref-CR9)).

The study concerns predicting students’ academic achievement using grades only, no demographic characteristics and no socio-economic data. This study aimed to develop a new model based on machine learning algorithms to predict the final exam grades of undergraduate students taking their midterm exam grades, Faculty and Department of the students.

For this purpose, classification algorithms with the highest performance in predicting students’ academic achievement were determined by using machine learning classification algorithms. The reason for choosing the Turkish Language-I course was that it is a compulsory course that all students enrolled in the university must take. Using this model, students’ final exam grades were predicted. These models will enable the development of pedagogical interventions and new policies to improve students' academic performance. In this way, the number of potentially unsuccessful students can be reduced following the assessments made after each midterm.

**Chapter 3**

**Methodology**

The purpose of this paper is to find out the mostly used attributes which are used for prediction of performance and determine which algorithm and parameters are best to improve the prediction mechanism in educational system.

We take dataset from Kaggle.

Now read the data into a Dataframe.

categorical\_cols **=** df**.**select\_dtypes(include**=**'object')**.**columns

To achieve the project’s aims, quantitative simulation research methods were conducted as suggested in the framework phases shown in Fig. [1](https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-019-0160-3#Fig1). In these phases the dataset will be prepared to be passed through visualization and clustering techniques, i.e. like heat map and hierarchical clustering, to extract the top correlated indicators. Then, the indicators will be used in different classification algorithms and the most accurate model will be the chosen for predicting student performance in dissertation projects and all courses grades. In between, and before the classification models’ evaluation phase, the datasets will pass through a pre-processing (cleansing, missing data imputation, …) stage to make it ready for the analysis phase. That will be more detailed in the following sections.

### **Participants and datasets**

In this study, the records of fifty graduated students in one master’s program were collected from the administration department. These records include students’ ID, age, bachelor degree name, bachelor degree accumulated grade, courses taken during their master’s study with their grades and instructors name of each course. Table [1](https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-019-0160-3#Tab1) shows the list of the main used attributes, their datatypes, and other related details. From that records, 2 datasets were created to answer the research questions and Table [2](https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-019-0160-3#Tab2) illustrates the descriptive statistics of that sets. These records were provided after to comply with the university’s data privacy obligations requirements and the replacement of students’ IDs and instructors’ names with other unique identifiers.

### **Data Analysis & Procedures**

**for** i **in** categorical\_cols:

print(df[i]**.**unique())

## **How many of the students take preparatory classes**

count\_test **=** df['test preparation course']**.**value\_counts()

labels **=** df['test preparation course']**.**value\_counts()**.**index

plt**.**figure(figsize**=** (6,6))

plt**.**pie(count\_test,labels**=**labels,autopct**=**'%1.1f%%')

plt**.**legend(labels)

plt**.**show()

A pie chart with numbers and a black background

Description automatically generated

## **Student performance in subjects based on gender**

sns**.**scatterplot(x**=**df['average\_score'],y**=**df['math score'],hue**=**df['gender'])

A screen shot of a graph

Description automatically generated

**Decision Tree**

Decision Tree classifier is the regression model which is represented in the form of tree structure. The purpose of Decision Tree classifier is to breakdown the dataset into smaller subset. The tree consists of decision nodes and leaf nodes.In our proposed architecture the attribute which delivers maximum information will act as a decision node. The node which is present as the top most of the decision node acts as a predictor which is called as root node. The node which cannot be further divided is known as leaf node. The steps involved in the decision tree are specified below:

• Process 1: Start the root.

• Process 2: Perform the test.

• Process 3: Follow the edges corresponding to the outcome.

• Process 4: go to step 2 until reaches leaf node.

• Process 5: Predict the outcome associated with the leaf.

**K-Nearest Neighbour**

K-Nearest Neighbour is one of the basic and essential classification algorithms in machine learning. It is non-parametric and makes any underlying assumptions about the distribution of data. The steps involved the KNN is listed below:

• File the training data in a sample points array.

• The Euclidean distance measure.

• Make the least distance range available.

**Linear Regression**

Logical regression is the classification technique which handles with the threshold value. The following arguments are used for determining the threshold value:

• Low Precision / High Recall

• High Precision / Low Recall

Based on the number of categories the Logical regression can be classified as:

• Binomial

• Multinomial

• Ordinal

**RandomForestRegressor()**

**from** sklearn.model\_selection **import** train\_test\_split

x\_train,x\_test,y\_train,y\_test**=**train\_test\_split(x,y,test\_size**=**0.25,random\_state**=**0)

**from** sklearn.ensemble **import** RandomForestRegressor

model**=**RandomForestRegressor()

model**.**fit(x\_train,y\_train)

predictions**=**model**.**predict(x\_test)

predictions

array([68.79666667, 76.92333333, 45.62 , 68.15666667, 74.2 ,

72.84 , 69.98 , 43.94666667, 86.62333333, 38.32333333,

51.66666667, 64.09333333, 82.11333333, 80.75 , 53.95333333,

39.44666667, 53.21333333, 86.6 , 53.53666667, 79.46333333,

62.72333333, 55.41 , 76.30333333, 60.72 , 47.86333333,

64.44333333, 58.64333333, 48.48666667, 63.12 , 81.99 ,

79.56666667, 89.26666667, 89.39666667, 56.02333333, 54.82 ,

83.54 , 96.96666667, 59.00666667, 71.69 , 72.07333333,

68.51333333, 56.74333333, 70.16666667, 71.72333333, 78.52333333,

58.29333333, 85.23666667, 61.61666667, 84.01 , 58.07333333,

39.22 , 62.33 , 48.96333333, 57.07 , 39.89666667,

70.21666667, 52.62333333, 64.60333333, 78.02333333, 60.94 ,

67.69333333, 77.35333333, 50.26666667, 71.78 , 65.61666667,

64.83333333, 71.80666667, 51.45 , 67.01 , 55.23666667,

75.44333333, 84.96 , 74.18333333, 65.45333333, 57.83 ,

65.22 , 72.27 , 70.11 , 29.57 , 76.66333333,

78.26666667, 85.80333333, 68.66 , 83.59 , 67.64666667,

75.59666667, 64.70666667, 90.53333333, 71.13 , 56.57666667,

71.93666667, 61.65 , 61.11666667, 60.44333333, 78.95666667,

71.91666667, 51.52333333, 68.20666667, 62.38666667, 64.45333333,

43.61 , 86.94666667, 63.83666667, 43.82333333, 76.65666667,

75.38333333, 68.55666667, 59.66333333, 55.17666667, 85.27666667,

68.21 , 76.20333333, 36.26333333, 53.67 , 67.77333333,

71.96666667, 80.51333333, 77.09 , 80.89333333, 59.39333333,

52.90666667, 78.50333333, 91.73666667, 81.22 , 74.68 ,

58.46 , 73.18333333, 74.62333333, 71.59666667, 62.49666667,

65.92666667, 66.33 , 23.56 , 82. , 84.04666667,

60.53666667, 65.05 , 64.51 , 69.15 , 82.69666667,

71.55333333, 66.74666667, 56.98333333, 60.1 , 76.64 ,

38.22 , 41.65333333, 44.90333333, 81. , 87.34 ,

50.32333333, 72.08666667, 95.38666667, 70.5 , 75.54666667,

80.03666667, 51.53333333, 55.79333333, 55.53666667, 51.51333333,

91.30666667, 50.25333333, 51.61333333, 92.43 , 65.37333333,

76.32 , 76.14333333, 69.37 , 58.87333333, 62.89666667,

57.05 , 65.97 , 33.11 , 65.05666667, 85.51333333,

68.03666667, 68.99 , 52.42666667, 96.26333333, 78.9 ,

65.37333333, 83.33 , 67.90333333, 67.51666667, 78.64 ,

67.91666667, 96.81666667, 99.49333333, 73.26333333, 47.55333333,

51.07333333, 60.45666667, 50.38666667, 76.82333333, 58.35 ,

60.96333333, 83.52333333, 81.44666667, 69.54666667, 66.01333333,

75.86 , 78.30333333, 57.18333333, 92.44 , 68.73333333,

51.44 , 77.98333333, 82.24333333, 73.88 , 69.49 ,

54.21333333, 91.04 , 57.14 , 63.54 , 37.03666667,

69.68 , 80.60333333, 63.27666667, 56.31666667, 72.22333333,

76.33666667, 66.79333333, 60.91 , 73.23666667, 64.08333333,

78.83666667, 87.9 , 60.1 , 44.36666667, 63.62666667,

44.63333333, 63.92 , 54.51333333, 48.64333333, 59.13333333,

66.12 , 70.63666667, 67.21 , 69.99333333, 58.92 ,

77.54333333, 51.32333333, 79.96666667, 58.87333333, 75.51333333,

72.00333333, 61.57666667, 92.32666667, 61.98666667, 84.99333333])

**Chapter 4**

**Result and Discussion**

Early student performance prediction can help universities to provide timely actions, like planning for appropriate training to improve students’ success rate. Exploring educational data can certainly help in achieving the desired educational goals. By applying EDM techniques, it is possible to develop prediction models to improve student success. However, using data mining techniques can be daunting and challenging for non-technical persons. Despite the many dedicated software’s, this is still not a straightforward process, involving many decisions. This study presents a clear set of guidelines to follow for using EDM for success prediction. The study was limited to undergraduate level, however the same principles can be easily adapted to graduate level. It has been prepared for those people who are novice in data mining, machine learning or artificial intelligence.

A variety of factors have been investigated in the literature related to its impact on predicting students ‘academic success which was measured as academic achievement, as our investigation showed that prior-academic achievement, student demographics, e-learning activity, psychological attributes, are the most common factors reported. In terms of prediction techniques, many algorithms have been applied to predict student success under the classification technique.

Moreover, a six stages framework is proposed, and each stage is presented in detail. While technical background is kept to a minimum, as this not the scope of this study, all possible design and implementation decisions are covered, along with best practices compiled from the relevant literature.

It is an important implication of this review that educators and non-proficient users are encouraged to applied EDM techniques for undergraduate students from any discipline (e.g. social sciences). While reported findings are based on the literature (e.g. potential definition of academic success, features to measure it, important factors), any available additional data can easily be included in the analysis, including faculty data (e.g. competence, criteria of recruitment, academic qualifications) may be to discover new determinants.

**import** seaborn **as** sns

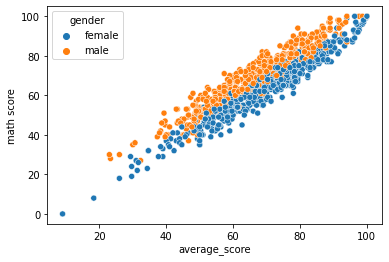
sns**.**countplot(df['gender'])

**A blue and orange rectangles

Description automatically generated**

## Student performance in subjects based on gender

sns**.**scatterplot(x**=**df['average\_score'],y**=**df['math score'],hue**=**df['gender'])

****

sns**.**scatterplot(x**=**df['average\_score'],y**=**df['reading score'],hue**=**df['gender'])

**A screen shot of a graph

Description automatically generated**

**Chapter 5**

**Conclusion and Future Work**

The prediction of student’s performance in advance is very important issue. We concluded after deep studies that various datasets of student provides different results with different attributes. This is the reason that the results are vary with different evaluation measures like accuracy, precision and geometric mean. We concluded after these studies that every data mining approach and algorithm results are varied according to the dataset and variable attribute used for prediction. However, if we use the decision tree algorithms, ADTree, JRip, Ridor, logistic regression and neural network approach, according to our requirements these algorithms provide extra ordinary accurate results for future prediction and help in the betterment of education system. In this way, we can improve the prediction methods and performance of education system. This research will implement in future by use of real datasets of Fast University and take the student’s attributes. In order to determine the effects of best predictive algorithm (Decision Tree/Neural Network) and other techniques will evaluate by statistical and empirical studies. In this way, we can compare the results of students with previous semester results. The best techniques are selected based on accuracy. We will find more efficient techniques based on other execution measure like recall and other in future.

**References**

[1]. Kabakchieva, D., Predicting student performance by using data mining methods for classification. Cybernetics and information technologies, 2013. 13(1): p. 61-72.

[2]. Osmanbegović, E. and M. Suljić, Data mining approach for predicting student performance. Economic Review, 2012. 10(1).

[3]. Käser, T., N.R. Hallinen, and D.L. Schwartz. Modeling exploration strategies to predict student performance within a learning environment and beyond. in Proceedings of the Seventh International Learning Analytics & Knowledge Conference. 2017. ACM.

[4]. Marquez-Vera, C., C.R. Morales, and S.V. Soto, Predicting school failure and dropout by using data mining techniques. Tecnologias del Aprendizaje, IEEE Revista Iberoamericana de, 2013. 8(1): p. 7-14.

[5]. Rovira, S., E. Puertas, and L. Igual, Data-driven system to predict academic grades and dropout. PloS one, 2017. 12(2): p. e0171207.

[6]. Ahmed, A.B.E.D. and I.S. Elaraby, Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, 2014. 2(2): p. 43-47.

[7]. Tamhane, A., et al. Predicting student risks through longitudinal analysis. in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014. ACM.

[8]. Daud, A., et al. Predicting Student Performance using Advanced Learning Analytics. in Proceedings of the 26th International Conference on World Wide Web Companion. 2017. International World Wide Web Conferences Steering Committee.