Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj

Predicting The Impact of ICT on University Students' Academic Performance Using Machine Learning

By

Md. Shamim Parvej
ID: 16CSE004

Department of Computer Science & Engineering

Bangabandhu Sheikh Mujibur Rahman Science and Technology University

Predicting The Impact of ICT on University Students' Academic Performance Using Machine Learning

Course Name: Project/Thesis

Course Code: CSE488

Md. Shamim Parvej
ID: 16CSE004

Department of Computer Science & Engineering

Bangabandhu Sheikh Mujibur Rahman Science and Technology University

Predicting The Impact of ICT on University Students' Academic Performance Using Machine Learning

This thesis is submitted in the partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering (B.Sc. Engg. in CSE)

By

Md. Shamim Parvej

ID: 16CSE004

Supervised by

Nasif Ahmed

Lecturer,

Department of Computer Science & Engineering

Bangabandhu Sheikh Mujibur Rahman Science & Technology University

The thesis titled "**Predicting The Impact of ICT on University Students' Academic Performance Using Machine Learning**" submitted by Roll No: 16CSE004 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering (B.Sc. Engg. in CSE) in Bangabandhu Sheikh   Mujibur Rahman Science and Technology University.

1. Nasif Ahmed
    Department of Computer Science & Engineering
    Bangabandhu Sheikh Mujibur Rahman Science & Technology University

Supervisor

2.

Examiner

3.

Examiner

_____   _____
Signature of the Candidate   Signature of the Supervisor
Date: Date:

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGMENT

At first, I would like to express my gratitude to almighty Allah who has created me and given me the strength to conduct this thesis. Many gratitude goes to my honorable mentor and supervisor Nasif Ahmed, who has given me a lot of instructions to conduct this thesis. His continuous and wise supervision gives me good directions to do the thesis perfectly. He gives me many guidelines on which issues in my thesis need to be focused on. I'm very much grateful to my thesis supervisor for his wise directions. So, I want to give heartfelt thanks to my supervisor for his co-operation and guidelines. While doing this thesis I have got much guidance, encouragement, and inspiration. Moreover, my gratitude goes to those who helped me a lot in developing this thesis from many aspects.

Md. Shamim Parvej

February,2022

ABSTRACT

Nowadays, Information and Communication Technology plays a vital role in our daily life. It is a part and parcel of our daily activities. We are all surrounded by Information and Communication Technology (ICT) and can't deny its contribution to our life. The usage of ICT is growing exponentially and day by day our students are going ahead using various technologies including the internet, social media, and various kinds of entertainment-related media. In this situation, we need to show the impact of information and communication technology on the students. As said earlier, we are all surrounded by ICT and students are not out of the boundary of this. There are positives as well as negatives impacts of ICT on the students. We are going to show that how much the students are being hampered or benefitted by using these technologies in their academic performance. The academic performance will be evaluated as bad, good, better, and best based on some input factors which has been collected from university students. The performance measurement will be conducted through the students' acquired GPA in their examination. After completing this thesis, we will be able to make decisions whether the students are lagging behind or not in their academic performance using various technologies including the internet, social media, and various kinds of online activities

CHAPTER 1

INTRODUCTION

In this modern era, we are all confined by the Information and Communication Technology web and we can't pass a moment without the technologies. From dawn to dusk we are surrounded by the technology web. Day by day, the usage of ICT is growing exponentially.

1.1 Introduction

Nowadays, Students are getting involved in many kinds of online networks. Internet is so much important in our life but we should keep into our mind that there are positive and negative impacts of Information and Communication Technology (ICT). It does no doubt that the influences of ICT are getting larger day by day. If you look in our near environment that how much the students are using a mobile phone, laptop and various kinds of technologies then it will be more clear that how much we are getting confined in the technology web. Primarily, the main responsibilities of the students should be attentive to study and academic activities. But if we look closely that a lot of students are passing a lot of time behind playing the online game, freelancing, and other activities. Freelancing is not bad but at first, as a student, it is more needed to focus on study than freelancing. A lot of students are passing time online without academic purposes and at last, in the examination there occurs a bad academic result. So, in my thesis, I tried to demonstrate that how much they are using ICT in their academic arena and using these how much performance they are achieving in their academic result. I will use some machine learning algorithms including Support Vector Machine (SVM), Naïve Bayes Classifier algorithm, Logistic Regression, Random Forest, and K-Nearest Neighbor algorithm for data training and testing. My machine learning algorithm will evaluate the performance level of the students based on some input factors, the performance level is classified as bad, good, better, and best. For doing this, I have used supervised machine learning and Jupyter as the editor for data mining tools to mine the data from the dataset.

1.2 Background and Present State

As said earlier, Information and Communication Technology has a great impact on our daily life. A lot of students are using these technologies. So, as a student, the main responsibility should be attentive to study. How much the students are using the internet for preparing his class lecture, learning any complicated matters for their academic purposes and if the students fall in any academic problems how frequent they are taking help from the internet. Moreover, the dependency on the internet is increasing but how much they are being benefitted using the internet, we are going to show that. At present, technologies are everywhere and it is so much difficult to concentrate on studying for our students for achieving the best academic outcome. So, in this time we need to develop a model that can evaluate the performance level of students based on some input factors. So, it is a crying need to measure the students' academic performance based on their usage of the internet and daily study time using the internet and passing time in social media. It is general that using social media and doing any online activities without academic purpose results in bad academic performance.

1.3 Motivation and Aims

There are very few articles related to this topic and they are not implemented using a machine learning algorithm. To demonstrate the student's academic performance we tried to conduct the thesis. We also evaluated the academic performance that classified as good, better, best, and bad performance. My main motivation is that how much the students are getting benefitted or hampered academically, I will just show it.

1.4 Objectives and Specific Aims

As the range of Information and Communication Technology is increasing day by day and the students are getting involved more in this section, so we need to demonstrate how much the students are being benefitted or hampered by technologies. The main objectives are

- Predict the academic performance.
- How much the students are being benefitted using Information and Communication Technologies.

- How frequent the students are taking help from the internet.

- Are they passing so much time on social media without academic purposes?

- Doing so many internet-related activities result in bad academic result.

My main aims are to find out the academic performance through Grade Point Average (GPA) and there are four main target factors and these will be evaluated based on some input factors. The machine learning algorithm will find out the target factor through the usage of supervised machine learning algorithm techniques.

CHAPTER 2

LITERATURE REVIEW

Our modern technology always tries to maximize the utilization of Information and Communication Technology. ICT not only has a positive effect on our students but also a negative effect. There are several papers related to our topic and they tried to find out the students' academic performance.

1.**Using Machine Learning to Predict Student Performance by Murat Pojon**

Three different machine learning methods were used in this thesis. They are linear regression, decision trees, and naïve Bayes classification. The aim is to predict if the student has passed the exam or not by looking some input factors.

Table 2.1:Classification Report of Different Algorithms

Method	Accuracy (%)	Precision (%)	Recall (%)	F-measure
Linear regression	77.5	76.1	97.0	0.854
Decision trees	77.6	76.8	95.5	0.853
Naïve Bayes classification	75.5	76.3	92.4	0.838

Md. Shamim Talulder[1] has shown in his paper-The Impact of ICT on Students' Performance: A Case Study on Undergraduate University Students. The paper showed the academic use of it, non-academic and social use of it and he also showed the correlation between the variables of its addiction and students' performance.

Mohammad Aman Ullah[2] showed in his paper- Impact of ICT on Students' academic performance that the basic information profile of students based on Gender, Internet browse regularly, Browsing period(Daily), use of technology, and average CGPA. He also showed the frequency of each attribute as well as a percentage of the corresponding attribute.

Md. Shamimul Islam[3] showed in his paper named as-The Impact of ICT on Students' Performance: A Case Study of ASA University Bangladesh. He has explored the reasons for students not having access to ICT facilities and also explored how students spend their time while using ICT.

Ishtiaq Hussain[4] showed in his paper the advantage of ICT in education, ICT in school, Teaching through ICT. He showed the students' performance level by statistical analysis and calculated mean and variance.

CHAPTER 3

METHODOLOGY

To conduct the thesis, we need a structural dataset. A dataset with good quality can give us a good accuracy of the model. At first, we need to collect a dataset from the education institute and from there we will get raw data. After getting raw data, we need to preprocess the data. After completing preprocessing, we need to convert it into numerical values.
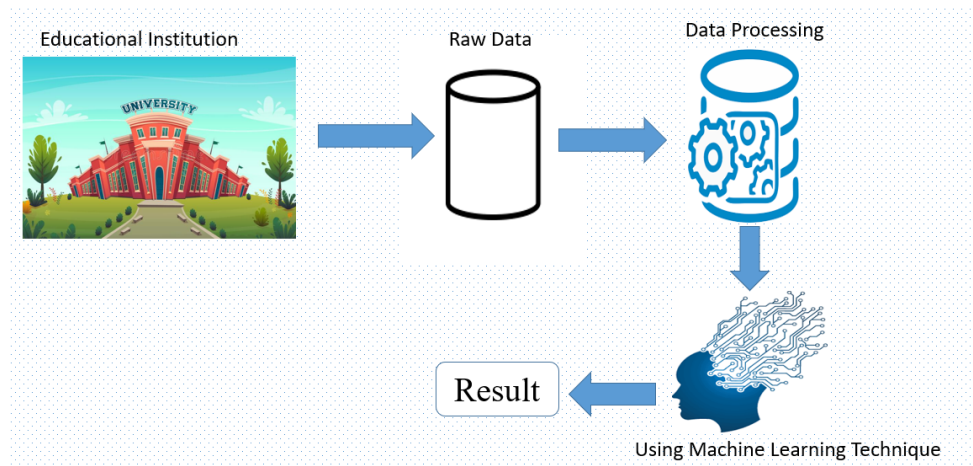
3.1 System Overview

Fig-3.1: System Overview

The system overview describes the overall methodology of the system and how the data will be collected from the field. After getting the raw data, we will use some preprocessing methods making it suitable for machine learning. We will use a supervised machine learning algorithm for our thesis. We will also see the model and various algorithms that which algorithm gives better accuracy. We need a lot of datasets of good quality. However, a good and level dataset can increase the accuracy of any model. We should keep into our mind that we have collected so much data from the field and make it suitable for machine learning.

3.2 Data Collection

Data collection is the main factor of any research. To achieve a good and quality model we should develop a questionnaire and survey to collect data, in collecting the dataset, we have given a survey questionnaire to university students. They have filled it and finally, we collected raw data from them.
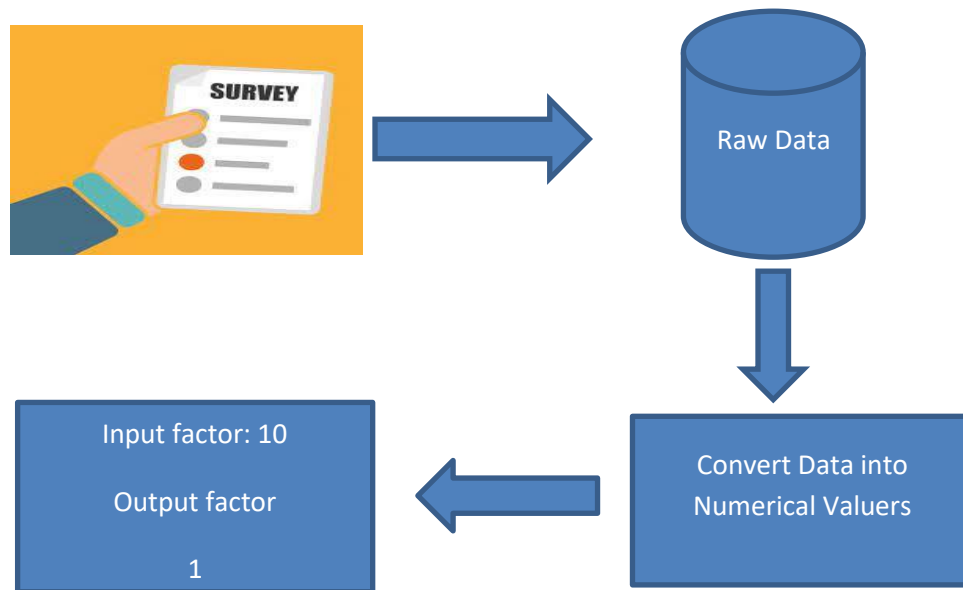
Fig-3.2: Data Pre-processing

We tried to collect data as much as possible to get a better model. We have asked the following questionnaire: Daily assignment time using the internet, how much time are passing daily on the internet for academic purposes, how frequent the students are taking help from the internet, how much time are passing in social media without study purpose, how much time are passing before the exam. in social media, how much time playing the online game and before exam how much time is passing, freelancing time, and finally the previous academic result.

3.3 Data Processing

Data processing is very much important making it suitable for machine learning, that's why we need to transform the dataset for machine training and testing. We have a considerable number of a dataset and using these we are going to develop a better model. In our questionnaire, we have 10 attributes and 1 target attribute. We have classified it as a non-academic feature questionnaire such as gender and social media related questionnaire and finally academic feature questionnaire. In processing the data we have followed some strategy including data cleaning, data addition, data reduction, and data transformation. We have also handled the missing values data by applying some strategies. If we find any missing value, replace the value by finding the mean of the corresponding attribute or if the data size becomes so large we can just delete the corresponding row from the dataset.

3.4 Description of Dataset

As mentioned earlier, we have divided the dataset into two categories one is independent factors and another is dependents factor which is our target factor.

1. Students daily assignment time

Generally, an attentive student makes a good academic result. If a student does his daily assignment given by the teacher, he surely will be able to achieve a good academic result.

2. Taking help from the internet

If a student falls into any academic-related subject problems then how frequent he is taking help from the internet, it is a good practice to search for any academic problems on

Internet. Because, in this modern era, everything is at our hand and with one click we can found a lot of problems solution in youtube, google or any trusted side.

3. Daily study time using the internet

The Internet gives us a lot of problems solutions related to the academic arena. Taking help from the internet while stuck in any academic problems gives us a better solution. From this

attribute, we will be able to know how much the students are dependent on the internet while stuck in any academic-related problems.

4. Before the exam. how much time are studying

Before an exam, it is a must to be attentive in the study and just before the examination, the time is very much important. So, this attribute will show the study time of a particular student.

5. How much time pass in social media without study purpose

Social media is an important platform both for communication and passing time. But, it can be very much dangerous if students pass their valuable time on social media without the study purpose.

6. Before examination how much time pass in social media

Social media captures a lot of valuable time. Students are passing their valuable time on social media and in this era it is natural. But before the examination, passing their valuable time on social media results in bad academic results. So, before the academic examination, it is a bad habit to pass the valuable time on social media.

7. How much time playing an online game:

Playing online games has now become a dangerous practice for our students. Even during the exam. students are given priority in playing the online game instead of studying. So, it brings about a bad academic result if students get addicted to playing the online game.

8. How much time playing the online game before the exam. time

During the examination time, it is not expected from students that they will play the online game instead of studying. So, being addicted to any online game may bring about a bad academic result.

9. Freelancing time

Freelancing is a good idea for being solvent economically and sometimes doing outsourcing is a normal activity. It eradicates the unemployment problem. But as a student getting involved excessively in freelancing is not a good idea. So, instead of doing academic study doing income from the internet is not a good practice and results in bad academic performance.

10. Previous academic result

We also collect a particular student's academic performance to know how great he/she is performing through his acquired GPA. As we are doing machine learning-related research, we need to know how he is doing well based on the input factors. As we will predict the performance level of a student including good, better, best, and bad performance, we need to know the previous GPA of students.

Input factors	
Factors Name	**Value**
Daily assignment time	Hours (Numerical Value from 1-3)
Taking help from internet	Hours (Numerical value from 1-3)
Daily study time	Hours (Numerical value from 0-5)
Daily study time before exam	Hours (Numerical values from 0-4)
Social media time	Hours (Numerical values from 0-3)
Before exam social media time	Hours (Numerical values from 0-3)
Daily online game	Hours (Numerical values from 0-3)
Before exam online game	Hours (Numerical values from 0-3)
Freelancing time	Hours (Numerical values from 0-3)
Previous Academic Result	Bad(1),good(2),better(3),best(4)

Fig-3.3 Input factor

Target Factor	
Factor Name	**Value**
Academic Performance through GPA	Bad(1) Good(2) Better(3) Best(4)

Fig-3.3 Input factor

CHAPTER 4

IMPLEMETATION DETAILS

To implement a system for research, a qualitative dataset are required. In prediction based research, there are some independent attribute and a target attribute. Target attribute is dependent on independent attribute. Independent attribute should be a relationship with target factor. In this section, we describe the visualization of various factor and also describe the implementation of the methodology.
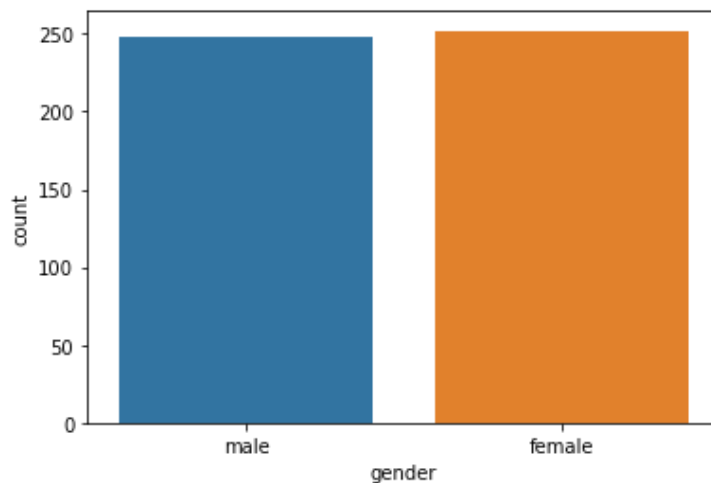
4.1 Data Visualization

Figure-4.1: Gender Feature Visualization

Data visualization is an important preprocessing task, which used graphical representation to simplify and understand complex data. Visualization techniques have been recently used to visualize online learning aspects. Instructors can utilize the graphical representations to understand their learners better and become aware of what is occurring in the distance classes. This research visualizes the current data set using Anaconda Navigator tool. As shown in Figure-4.1, the data set is visualized based on gender feature into 252 male and 248 female students. To learn the dataset for machine learning purposes, we have let male as 1 and female as 2.

This research uses the "Assignment Time" feature to show the influence of such feature on student's performance. As shown in Figure-4.2, the data set is visualized based on study time feature in hours into 501 students. From this figure, we can see that the intervals of the assignment time is divided into three ranges 0-1 hour, 1-2 hour and 2+ hours daily. In

our dataset, maximum students give their daily assignment time for 0-1 hour. A very few students spend in assignment time in 2+ hours.
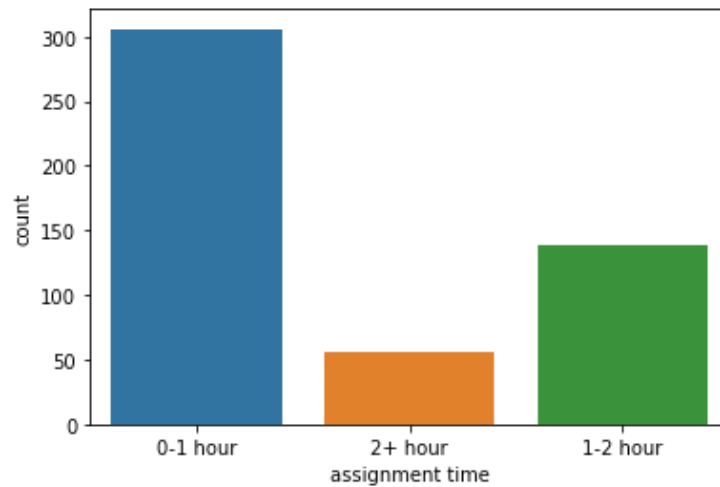

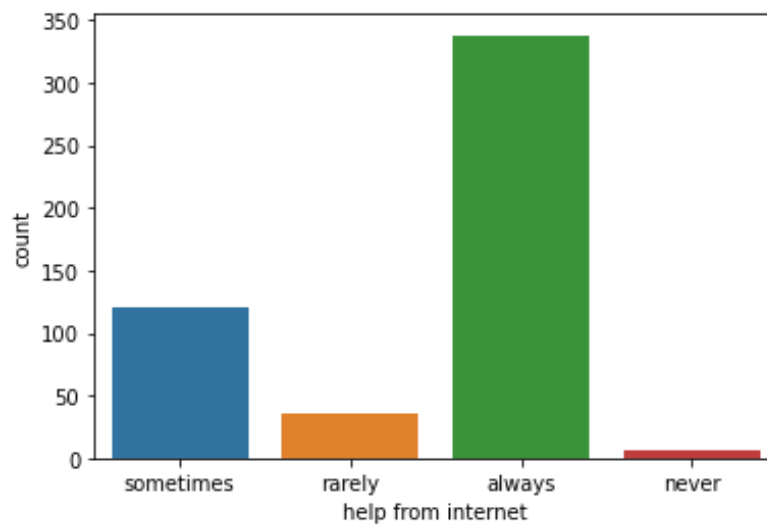Figure-4.2: Assignment Time Feature Visualization

Figure-4.3: Help From Internet Feature Visualization

In figure-4.3, the data set is visualized based on help from internet to demonstrate how frequent students are taking help from internet. 0 indicates students never take help from internet, only follow teacher lecture, 1 indicates very rare tale help, 2 indicates sometimes take help and 3 indicates students always take help from internet. The highest frequency of 'help from internet features' is 3 that means students always take help from internet. There are also some students who take very rare help from internet. On the other hand, a few student are there who sometime

take help from internet. This research paper works with the students daily study time feature shown in figure 4.4. Students who study on the regular basis can have higher possibility for better academic performance.
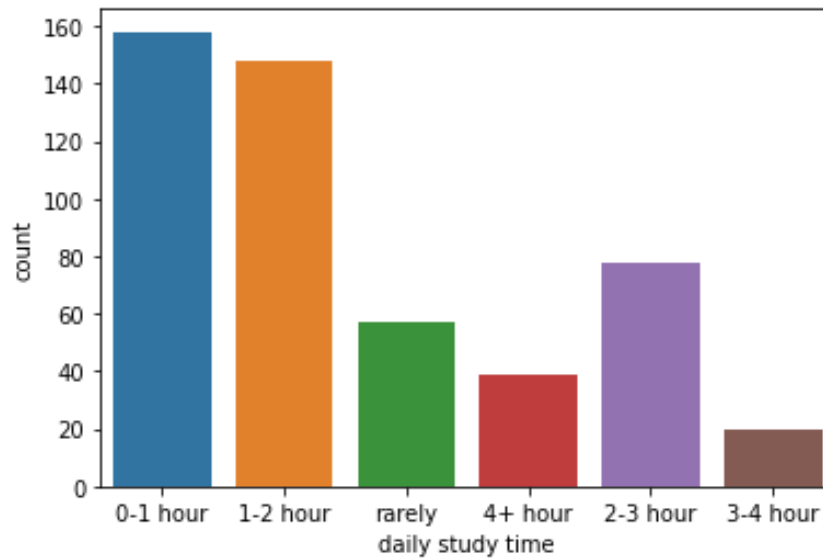

Figure-4.4: Daily Study Time Feature Visualization

From above figure, we can see that maximum number of students study daily in the '1' range that means 0-1 hour and in '2' that means 1-2 hours. A very few students are in the range of '4' that means students rarely study daily 3-4 hours.
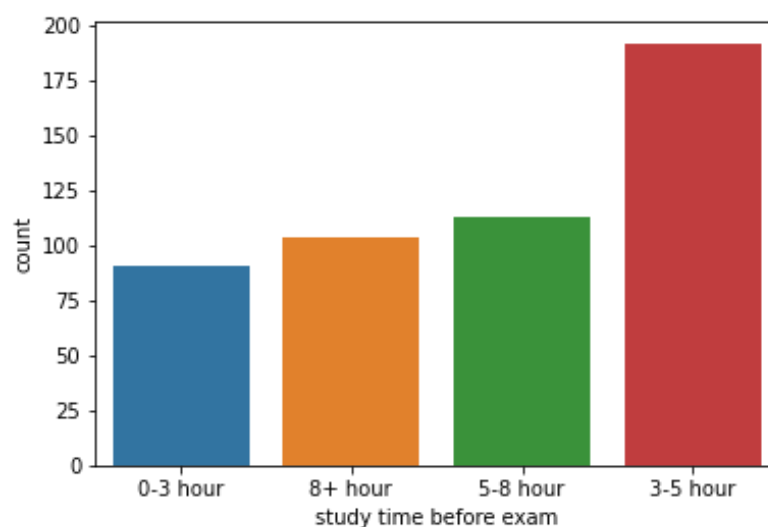

Figure-4.5: Study Time Before Exam Feature Visualization

As shown in Figure-4.5, the data set is visualized based on study time before exam feature on 501 students. We have let 0 when students study before exam 0-3 hours , 2 when study 3-5 hours , 3 when study 5-8 hours and 4 when study 8+ hours before exam. As we see that from graph, maximum number of   students study before exam 3-5 hours.

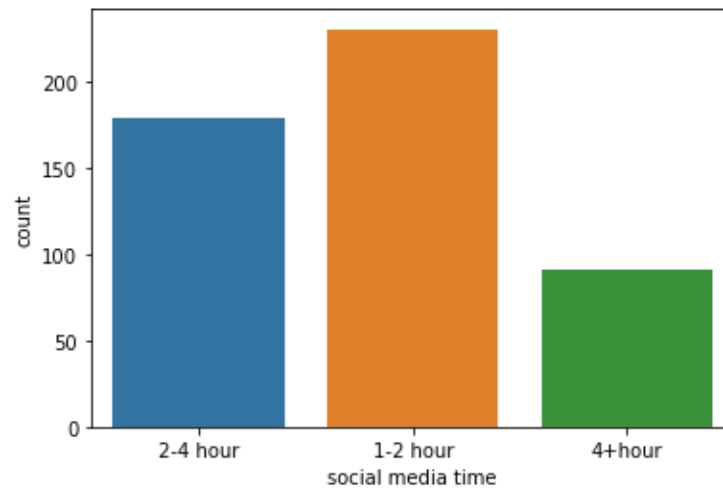Figure-4.6: Social Media Time Feature Visualization

From the figure 4.6, we can see that most of the students pass in social media 1-2 hour and 2-4 hour range.
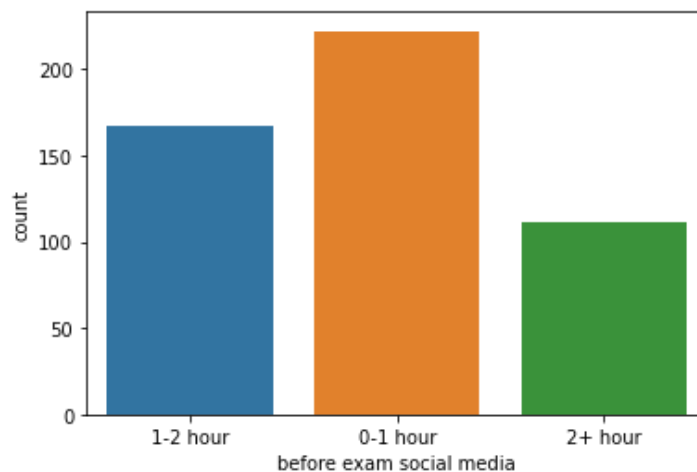
Figure-4.7: Before Exam Social Media Usage Time Feature

From the above figure 4.7, we can see that, most of the students use social media only 0-1 hour before examination and the minimum usage of social media before exam is 2+ hour.

Figure 4-8: Playing Online Game Feature Visualization

As we can see in the above figure that most of the students do mot play online game and some students play online game 0-2 hour, 2-4 hour and 4+ hour.

Figure-4.9: Freelancing Time Feature Visualization

As we can see that most of the students do not have freelancing profession. But if they do it, they pass 4+ hour , 0-2 hour and 2-4 hour.

4.2 Data Cleaning

Data cleaning is one of the main preprocessing tasks, is applied on this data set to remove irrelevant items and missing values. The data set co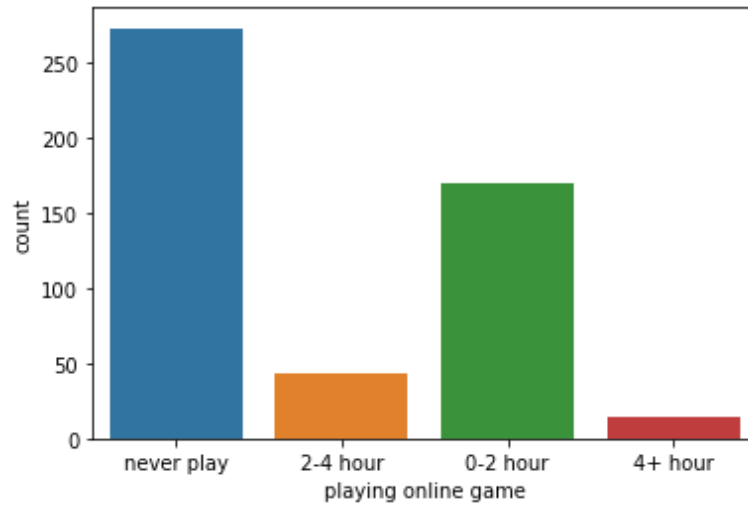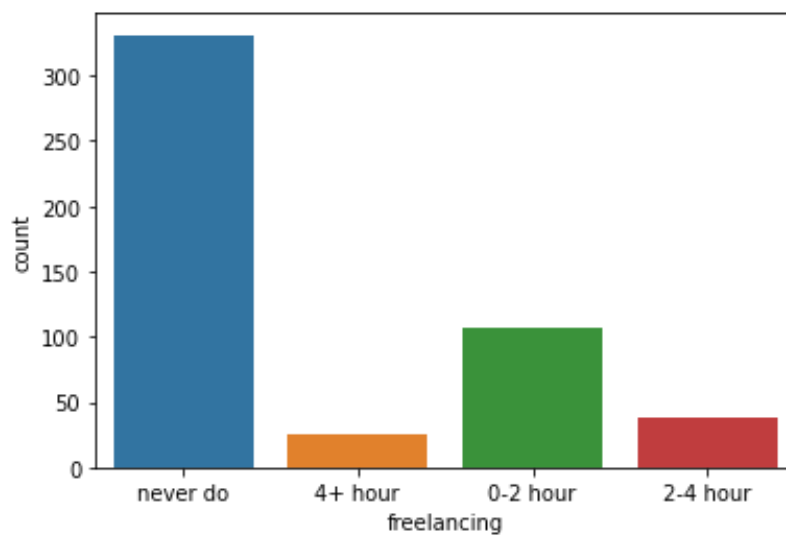ntains 10 missing values in various features from 511 records, the records with missing values are removed from the data set, and the data set after cleaning becomes 501 records.

4.3 Features Selection

Feature selection is a fundamental task in data preprocessing area. The objective of feature selection process is to select an appropriate subset of features which can efficiently describe the input data, reduces the dimensionality of feature space and removes redundant and irrelevant data [13]. This process can play an important role in improving the data quality therefore the performance of the learning algorithm. Feature selection methods are categorized into wrapper-based and filter-based methods. Filter method is searching for the minimum set of relevant features while ignoring the rest. It uses variable ranking techniques to rank the features where the highly ranked features are selected and applied to the learning algorithm. Different feature ranking techniques have been proposed for feature evaluations such as chi-square test.
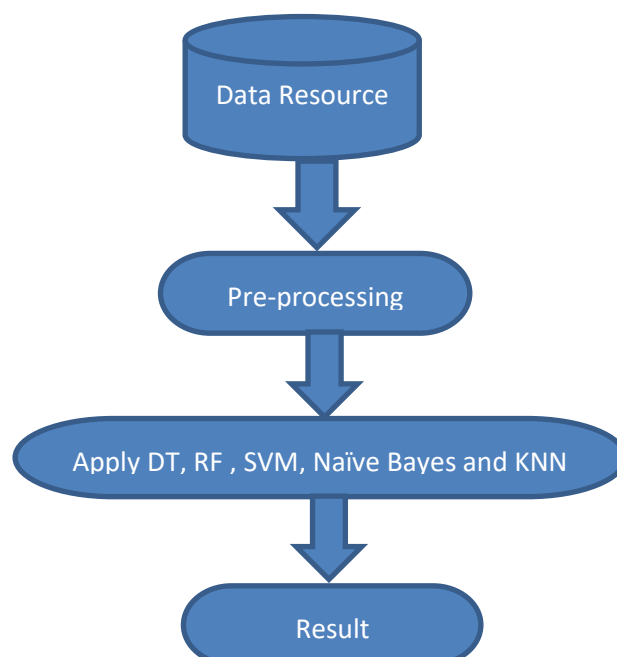
4.4 Model Evaluation

Figure-4.10: Student's Academic Performance Evaluation Model

As we are conducting a machine learning model, we need to know which type of machine learning model we are running. We are evaluating the students' academic performances based on some input factors and our output label is a classification based. Our machine learning model is supervised learning model, that's why we need to select algorithms which will yield to best accuracy for our model. Decision Tree, Random Forest, Support Vector Machine , Naïve Bayes and KNN are supervised learning model and we are conducting our machine learning based on these supervised algorithms.

4.5 Methodology

In this paper, we introduce a student's performance model using ensemble methods. Ensemble methods is a learning approach that combines multiple models to solve a problem. In contrast to traditional learning approaches which train data by one learning model, ensemble methods try to train data using a set of models, then combine them to take a vote on their results. The predictions made by ensembles are usually more accurate than predictions made by a single model. The aim of such approach is to provide an accurate evaluation for the features that may have an impact on student's academic success. Figure-4.10 shows the main steps in the proposed methodology.

4.6 Basic Concept of Decision Tree (DT)

Figure-4.11: Decision Tree Structure
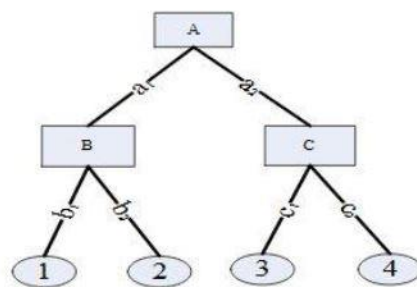
Decision tree is a tree structure which looks similar as flow chart. Every node in the tree represents the test of an attribute, every branch represents the output of the test, and every leaf means a class or distribution of classes. The top node is root node, from root node to one leaf consists a classification rule. So decision tree is easy to be transferred to classification rules. It

has many algorithms, but the main idea is using from top to bottom induction method. And the most important part is choosing which attribute to be the node as well as the evaluation of whether the tree is correct. As Figure-4.8 shows, it's a very simple decision tree. A, B, C represents different attributes separately in one data set and each branch like a_1, a_2, b_1, b_2, c_1, c_2 represents the value of split attribute. Leaf node 1, 2, 3, 4 represents the class of decision attribute in each sample set. There are mainly two steps in decision tree, build a decision tree and do pruning. The thought of building decision tree is called CLS. We have a data set S, the attributes set is A, decision attributes set is D and the whole process is as follow:

Step-1: Make S be the root node, if all data in S belongs to the same class, turn node to leaf node.

Step-2: Otherwise choose one attribute a epsilon A and divide nodes according to different values of attribute a. S has the number of m lower layer nodes, branches represent the situation of different values of a.

Step-3: Induct step 1 and step 2 for m branch nodes.

Step-4: If attributes in one node belong to same class or there is no node to divide, then stop.

The most two important things in decision tree are: 1 how to decide best split node? 2 when to stop splitting? Because real data can't be pure. There must be data attributes miss, data inaccurate, noise these situations, which will result in over fit. Over fit may lower the accuracy of the classification and prediction of decision tree and increase the complexity of tree structure. So after building a tree, we also need to pruning.

4.7 Basic Concept of K-Nearest Neighbor (KNN)

The K-Nearest-Neighbors (KNN) method of classification is one of the simplest methods in machine learning,and a great way to introduce yourself to machine learning and classification in general. At its most basic level, it is essentially classification by finding the most similar data points in the training data, and making an educated guess based on their classifications. Although very simple to understand and implement, this method has seen wide application in many domains, such as in recommendation systems, semantic searching, and anomaly detection. As we would need to in any machine learning problem, we must first find a way to represent data points as feature vectors. A feature vector is our mathematical representation of data, and since the desired characteristics of our data may not be inherently numerical, preprocessing and feature-engineering may be required in order to create these vectors. Given data with N unique features, the feature vector would be a vector of length N, where entry I of

the vector represents that data point's value for feature I. Each feature vector can thus be thought of as a point in R^N. Now, unlike most other methods of classification, KNN falls under lazy learning, which means that there is no explicit training phase before classification. Instead, any attempts to generalize or abstract the data is made upon classification. While this does mean that we can immediately begin classifying once we have our data, there are some inherent problems with this type of algorithm. We must be able to keep the entire training set in memory unless we apply some type of reduction to the data-set, and performing classifications can be computationally expensive as the algorithm parse through all data points for each classification.

Euclidean distance is probably the one that you are most familiar with; it is essentially the magnitude of the vector obtained by subtracting the training data point from the point to be classified.

$$E(x, y) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2}$$

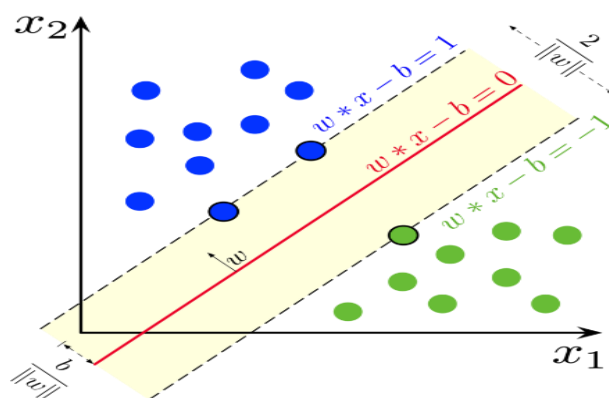## 4.8 Linear Support Vector Machine (SVM)

Figure-4.12: Linear Support Vector Machine

 In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as

belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. In Figure-4.12, we have shown the linear representation of support vector machine.

 When data is unlabeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support-vector clustering algorithm, created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications.

4.8 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
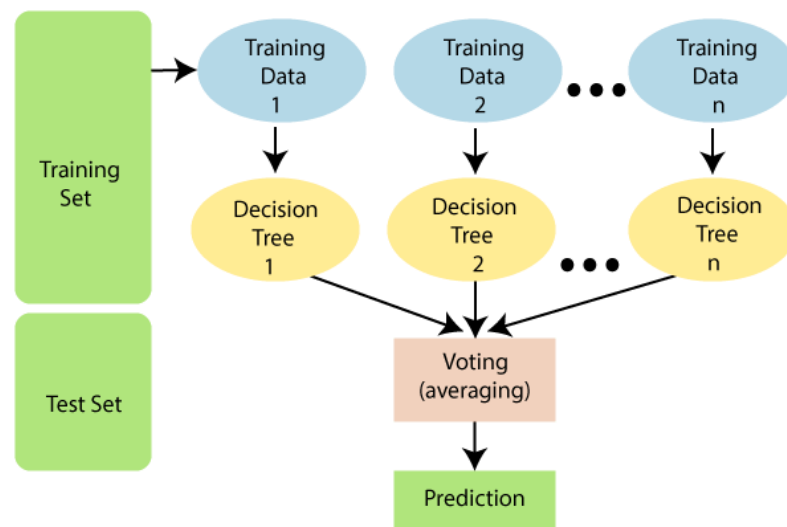
Figure 4.13- Random Forest Algorithm Working Procedure   [5]

Why use Random Forest

Below are some points that explain why we should use the Random Forest algorithm:

- o It takes less training time as compared to other algorithms.
- o It predicts output with high accuracy, even for the large dataset it runs efficiently.
- o It can also maintain accuracy when a large proportion of data is missing.

Working Procedure:

Fig 4.14- Random Forest Working Procedure Example   [6]

During the training phase, each decision tree produces a prediction result, and when a new data point occurs, then based on the majority of results, the Random Forest classifier predicts the final decision.

CHAPTER 5

EXPERIMENTAL RESULT AND DISCUSSION

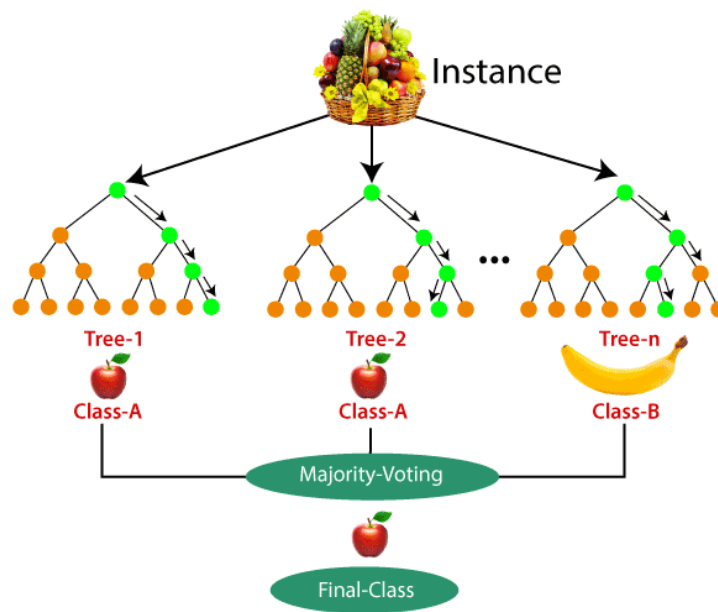In any type of research, experimental result is very essential. All the researches want to reach the highest accuracy level according to their work. This accuracy level may be difference by using various algorithm and methodology. The researches select the algorithm and methodology which give the best accuracy level for the corresponding research.

5.1 Environment

We ran the experiments on the PC containing 8GB of RAM, 5 Intel cores (2.67GHz each). For our experiments, we used Anaconda Navigator to evaluate the proposed classification models and comparisons. We learnt the dataset with five algorithms. These are SVM, NB, DT, KNN and RD. We have got different type results across of the different algorithm.

5.2 Evaluation Measures

In our experiments, we use four common different measures for the evaluation of the classification quality: Accuracy, Precision, Recall and F1-Score. Measures calculated using Table-5.1, which shows classification confusion matrix based on the Equations 1, 2, 3 and 4, respectively.

Table-5.1: Confusion Matrix

		Detected	
		Positive	**Negative**
Actual	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Accuracy is the proportion of the total number of predictions where correctly calculated. Precision is the ratio of the correctly classified cases to the total number of misclassified cases and correctly classified cases. Recall is the ratio of correctly classified cases to the total number of unclassified cases and correctly classified cases. In addition, we used the F measure to combine the recall and precision which is considered a good indicator of the relationship between them.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1} = \frac{Precision + Recall}{Precision + Recall} \tag{4}$$

5.3 Classification Results

Finally,we applied machine learning techniques on all of data features and the results were shown in Table-5.2

Table-5.2: Classification Results by using different algorithm

Evaluation Measure	Decision Tree	Random Forest	SVM
Accuracy	0.84	0.80	0.64
Recall	0.84	0.93	0.65
Precision	0.81	0.89	0.67
F1-Score	0.82	0.91	0.65

The table is all about the classification report of various kinds of supervised learning algorithms. Different algorithms yields different accuracy of our model.

5.4 Confusion Matrix for DT, RD and SVM

Table-5.3: Confusion Matrix with Using Decision Tree

		Actual			
		Bad	**Good**	**Better**	**Best**
Predicted	**Bad**	22	1	2	2
	Good	1	40	2	0
	Better	4	0	22	3
	Best	0	0	4	21

Table-5.4: Confusion Matrix with Using Random Forest

		Actual			
		Bad	**Good**	**Better**	**Best**
Predicted	**Bad**	22	1	1	3
	Good	1	40	0	2
	Better	2	2	20	6
	Best	1	2	3	19

Table-5.5: Confusion Matrix with Using SVM

		Actual			
		Bad	**Good**	**Better**	**Best**
Predicted	**Bad**	15	10	1	1
	Good	1	20	4	0
	Better	4	9	13	0
	Best	5	2	3	33

In Table-5.3, we have shown the confusion matrix for Decision Tree algorithm. This confusion matrix is more accurate than others algorithm. In this technique, the accuracy is 84%.

In Table-5.4, the confusion matrix with using Random Forest is shown. In this technique, the accuracy is 80%.

In Table-5.7, we have shown the confusion matrix of using the SVM algorithm. The accuracy is 64%.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

Every research paper has its way to acquire the best result. But there may have some lacking. No research paper is out of mistakes. The machine can't become human and it is not possible. So, it is never possible to acquire 100% accuracy. But using the machine learning technique, we can predict a considerable output.

6.1 Conclusion

There may be mistakes and lacking. But, we can reduce the mistakes and lacking using many strategies. As we are working on the impact of information and communication technology on university students' academic performance, there may be mistakes for predicting desired performance level of a particular student due to our dataset faults. But, we will try to achieve better accuracy and students' performance level by applying various machine learning strategies. We know that the more we enlarge the dataset size, the more we get the best prediction from our model. But each machine learning algorithm has its strategy for finding out the result. So, we need to apply various algorithms to predict our outcome. However,we will find out the various supervised machine learning algorithm's accuracy and compare the accuracy level to which one gives the best prediction.

6.2 Future scope

There is no doubt that from dawn to dusk we are all surrounded by Information and Communication Technology (ICT). The main responsibilities of the students are to give attention to study and make the best use of the internet in their academic field. Using social media for a long duration is normal but the students should make sure that they are not using various kinds of social media excessively. We have tried our best to show it in our research that students should not use various online gaming platforms and entertainment platforms excessively. In our experiment, we have used a considerable dataset. If the dataset becomes larger, we can achieve better accuracy from the model. We know that no model is out of errors and every research has limitations. What we can do in this perspective is that we can reduce our limitations and develop a better model to achieve the best accuracy. In the future, we will try to enlarge the dataset and try to focus on the questionnaire. Primarily, setting up a perfect

questionnaire is somewhat difficult. In the future, we will try to analyze the survey questionnaire. As the perfect questionnaire is not possible to make survey but the improvement on the survey questionnaire is possible. We will also keep the focus on the algorithm which can be perfect more achieve more accurate prediction. So, in the future, we will keep more focus on collecting the dataset and making the survey questionnaire more consistent and perfect that will become relevant to our research topic.

REFERENCES

[1] Md. Shamim Talukder, "The Impact of ICT on Students' Performance: A Case Study of Undergraduate University Students".

[2] Mohammad Aman Ullah, "Impact of ICT on Students' Academic Performance: Applying Association Rule Mining and Structured Equation Modeling".

[3] Md. Shamim Islam, "The Impact of ICT on Students' Performance: A Case Study of ASA University Bangladesh".

[4] Ishtiaq Hussain, "Effect of Information and Communication Technology (ICT) on Student's Academic Achievement".

[5] Random Forest Algorithm Working Procedure. Website: https://javatpoint.com

[6] Example of Random Forest Algorithm. Website: https://javatpoint.com