

Machine Learning based Slow Learner Prediction in Educational Sector

BY

Nurun Nahar Ela

ID: 171-15-9412

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

Supervised By

Nusrat Jahan

Sr.Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Jerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JUNE 2021

APPROVAL

This Project titled “**Machine Learning based Slow Learner Prediction in Educational Sector**”, submitted by Nurun Nahar Ela, ID No: 171-15-9412 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 03 June, 2021.

BOARD OF EXAMINERS

Chairman



Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Internal Examiner

Nazmun Nessa Moon

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Internal Examiner

Aniruddha Rakshit

Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



External Examiner

Dr. Mohammad Shorif Uddin

Professor

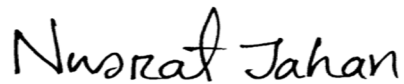
Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Nusrat Jahan, Sr.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:



Nusrat Jahan
Sr.Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Nurun Nahar Ela
ID: 171-15-9412
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Nusrat Jahan**, Sr.Lecturer, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest with supportive instructions helped us in the field of deep learning based research, finally we completed our work on “*Machine Learning*”. 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 **Dr. Touhid Bhuiyan**, Professor and Head, Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his valuable support and advice to finish our project and also heartiest thanks to other faculty member and the staff of department of CSE, Daffodil International University.

At last, again we want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This research is a result of hard work and all those inspirations and assistance.

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

ABSTRACT

The educational sector has been proved to be a sector where improvement measures are a must to improve the system, the curriculum which changes in every few years , student's state in order to cope with the changing curriculum every now and then, how much a student has been able to achieve academically, why are they falling behind etc. Education is said to be the backbone of a nation, if the students are falling behind that means the nation is falling behind. So, the first and foremost duty of the education ministry, education ministry related personnel's, teachers is to improve the malfunctioning educational system and come up with a system which can help the students truly in a sense of letting them achieve their dreams. Our purpose of the study is to predict the slow learners in the university level which is the last paddock of their study life and the step where they must acquire skills to face the professional life. We have acquired genuine university student information. The information has been acquired from the CSE (computer science and engineering) department students. In order to achieve the prediction and accuracy outcomes, Machine learning classification based algorithms will be applied.

TABLE OF CONTENTS

CONTENTS	PAGE
Approval Page	ii-iii
Declaration	iv
Acknowledgement	v
Abstract	vi
List of Tables	ix
List of Figures	x-xi
 CHAPTER 1: INTRODUCTION	 1- 5
1.1 Introduction	1-2
1.2 Motivation	2
1.3 Rationale of the Study	3
1.4 Research Questions	3
1.5 Expected Output	3-4
1.6 Project Management and Finance	4
1.7 Report Layout	4-5
 CHAPTER 2: BACKGROUND	 6-12
2.1 Preliminaries/Terminologies	6
2.2 Related Works	6-11
2.3 Comparative Analysis and Summary	11
2.4 Scope of the Problem	11-12
2.5 Challenges	12
 CHAPTER 3: RESEARCH METHODOLOGY	 13-20
3.1 Research Subject and Instrumentation	13
3.2 Data Collection Procedure/Dataset Utilized	13-16
3.3 Statistical Analysis	16
3.4 Proposed Methodology/Applied Mechanism	17-20
3.5 Implementation Requirements	20

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	21-33
4.1 Experimental Setup	21
4.2 Experimental Results & Analysis	21-30
4.3 Discussion	31-33
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	34-36
5.1 Impact on Society	34
5.2 Impact on Environment	34
5.3 Ethical Aspects	34-35
5.4 Sustainability Plan	35
CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE	36-37
6.1 Summary of the Study	36
6.2 Conclusions	36
6.3 Implication for Further Study	36-37
REFERENCES	38-39
APPENDICES	40
PLAGIARISM REPORT	41

LIST OF TABLES

TABLES	PAGE NO
3.2.1.Data Variables	14-16
3.4.1.High Potential Variables in Ascending Order	18-19
4.2.1.Average Class Confusion Matrix	28
4.2.2 Fast Class Confusion Matrix	29
4.2.3.Slow Class Confusion Matrix	29
4.3.1.Classification Report of all the Algorithms used	32-33

LIST OF FIGURES

FIGURES	PAGE NO
3.4.1. Flowchart of our respective study	17
3.4.2. High Potential Data Categories in Bar chart	18
4.2.1. Decision Tree Accuracy	21
4.2.2. Confusion Matrix of Decision Tree	21
4.2.3. Classification Report of Decision Tree Algorithm	22
4.2.4. Random Forest Accuracy	22
4.2.5. Confusion Matrix of Random Forest Algorithm	22
4.2.6. Classification Report of Random Forest Algorithm	23
4.2.7. GaussianNB Accuracy	23
4.2.8. Confusion matrix of GaussianNB Algorithm	23
4.2.9. Classification Report of GaussianNB Algorithm	24
4.2.10. BernoulliNB Accuracy	24
4.2.11. Confusion matrix of BernoulliNB Algorithm	24
4.2.12. Classification Report of BernoulliNB Algorithm	25

4.2.13. MultinomialNB Accuracy	25
4.2.14. Confusion matrix of MultinomialNB Algorithm	25
4.2.15. Classification Report of MultinomialNB Algorithm	26
4.2.16. K Nearest Neighbor Accuracy	26
4.2.17. Confusion matrix of K Nearest Neighbor Algorithm	26
4.2.18. Classification Report of K Nearest Neighbor Algorithm	27
4.2.19. SVM Accuracy	27
4.2.20. Confusion matrix of SVM Algorithm	27
4.2.21. Classification Report of SVM Algorithm	28
4.2.22. Algorithm Accuracies	30
4.2.23. Comparison of Algorithm Accuracies	30
A1. Sample Dataset	40

CHAPTER 1

INTRODUCTION

1.1 Introduction

Slow learners has always come to be a stumble in the path of success of any educational institution or more menacing in the life of a slow learning student himself. In a third world country like ours, the slow learning problem of students is not taken seriously rather than helping them this problem of theirs, it often becomes an issue of mockery in the context of our educational/social environment which leads them in the path of grave depression. In the university level, this problem happens to be intimidating for the slow learners as they have to start their professional life soon after they finish their bachelors. The university level is for acquiring skills to be better in a particular field in order to have a job and be able to take care of their families as many students come from a background where they are the only potential member of their family to receive higher education. Our approach will be to find out the potential slow learners in order to get them the assistance of their respective professors to help them improve their situation and be considered as an equal and competent student along with the others. The world is facing a world-wide pandemic right-now which is the virus called Corona. It affected the whole world so vastly that the entire world came to a stand-still, every sector was affected. The socio-economic structure broke down in almost every country. This pandemic made the educational institutes to open the online learning era. The slow learners were more affected in this sudden change of learning environment.

Dima Alhadidi et al. described in their study that they have apprehended the utilization of collective interchange that appears in machine-learning so as to achieve a new illustrated movement specified as collective learning forms a structure. Collective learning inexact and on the progress in particular should enable novices and demonstrators to regularly notify each other regarding what their notions are , discovering or undergoing. The methods used for this paper is a combination of the learning atmosphere through learning commodities and the social atmosphere through social commodities. The conclusion of this paper is, collective learning should be the

reason of encouragement for additional opportunities for both novices and demonstrators in terms of expanding participation in various online activities while on the progress, amplifying their evaluative thinking capabilities by exposing them to different experiences and last but not the least supporting anywhere and any-time social learning [6].

Faieza Chowdhury also stated in a paper how classrooms in Bangladesh should be flipped in order to be prepared for the online learning environment. By utilizing the blended learning tools, HEIs in Bangladesh can achieve radical improvements in education quality as well as in the accessibility and cost-effectiveness of learning programs. Moreover, any innovative educational reform will be successful only when it is fully accepted and adopted by all the key stakeholders: students, parents, teachers, academic administrators, researchers and policy makers. First, learning outcomes (LO) and appropriate assessment techniques (AT) should be set clearly. Matching the LO and AT to have appropriate teaching technological aids using Bloom's Digital Taxonomy of Learning. Then the transformation of a classroom from a traditional one to a flipped one will take place [9]. We will apply classification based machine learning algorithms to predict the slow learners.

1.2 Motivation

We have witnessed many of our fellow classmates face problems regarding class lessons or to even understand the teacher's lectures. Understanding some course topics also became the reason of struggle for many students. After witnessing this problem for a long time and when this problem begin to rise after the switch of physical learning environment to online learning environment, the students of this sort were in grave distress. Many students were also found who backed out of this new online learning environment because they could not deal with these problems related to their studies and there are other reasons also. That is where we found our motivation to work in this study which will classify slow, average, fast learners and detection of slow learners will be very easy.

1.3 Rational of the Study

In this analysis, the dataset we have chosen with data categories which define various information about student's and this vast amount of information helps us classifying the slow, average and fast learners with the help of ml(machine learning) classification algorithms. This study is crucial for the students who are lagging behind in this present learning scenario. In this generation, being a slow learner is the worst possible thing that could happen to a student because this is a very competitive era which will not consider any lacking in any respective area of work. By detecting slow learners, they can be brought under a special assistance program which can prove to be benefitted towards the university, the student himself and also for the nation.

1.4 Research Questions

To complete this study, the journey has been very challenging. Collection of the desired data that we wanted has been a tough process, then carrying out the procedures in machine learning has also been a part which was tricky as the other course of action. The research questions which can rise are -

- Can we collect this immense amount of information of students for the purpose of machine learning analysis?
- Is it feasible enough to pre-process the raw data utilizing machine learning perspective?
- Is it possible to improve education system by detecting slow learners using this approach?

1.5 Expected Outcome

In this section, we will portray the outcomes of our respective study.

- Machine learning classification algorithms will classify students in the following slow, average and fast learners.
- Detection of slow learners will be in an efficient way.

- Assistance can be provided to the slow learners to help them to come out of their shells and fight against their problems.

1.6 Project Management and Finance

Project management for our study was a hectic routine. In this online environment of study, we had to wait at-least 6-7 months just to fully achieve our entire dataset as our targeted amount of data was to get information of 1000 CSE department students. After achieving the data set, the pre-process stage of the study took nearly 10-15 days as the dataset contains an immense amount of data. When the Pre-processing stage was completed the feature selection took 1-2 days and applying the algorithms along with confusion matrix, classification report, accuracy and prediction procedure was done in the next 7-10 days. As project finance, we needed a laptop or personal computer with the right configuration which will cost around 18-25 thousand or much higher.

1.7 Report Layout

Chapter one illustrates the objective, motivation, rationale of the study, questions regarding the analysis, and anticipated consequence , the structure of this report is outlined in this segment.

Chapter two lay outs the information about how things have been done in this area of research before, comparison between our study and the other related studies, scopes of this respective problem and the challenges that we faced during the collection of dataset and implantation of the study.

Chapter three portrays conceptualization of the research subject, instruments and environment needed to apply the algorithms. How the Dataset is collected and pre-processed precisely is also represented along with the statistical analysis of the study. Proposed work-flow of the analysis and requirements for fulfilling implementation of the analysis is also explained.

Chapter four provides with the detailed experimental outputs and analysis with the use of figures and tables to understand the study to its core.

Chapter five is the section where the impacts of this study are explained briefly. Impacts on the society and environment are discussed with an overview of ethical aspects and this study's sustainability plan for the long run.

Chapter six describes the summary of this study, conclusion and implications for the future research.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

Slow-learners, educational sector, machine learning, classification algorithms.

2.2 Related Works

Slow learner prediction has been a world-wide research field for research scientists all over the world. The main motivation behind the selection of this particular topic for researchers world-wide has been lead to one sole purpose which is the welfare of the students and helping them to be better at studies. Slow learning is not a problem which only the developing countries face, in the flourished countries it has also been identified as an intimidating problem for the socio-economic structure of the society. Hence to solve this issue, many researchers has proposed many ways around the world.

Loannis E.livieris et al. proposed an algorithm to forecast student's performance early and accurately in order to improve their learning strategy to achieve a better performance in class. For forecasting early and accurate student's academic outcome, semi-supervised algorithm and ensemble techniques were applied. The base classification algorithms were Naive Bayes, C4.5 decision tree algorithm, Sequential Minimum Optimization (SMO), Multilayer Perceptron (MLP), 3-NN and J Rip. These algorithms were compared with the SSL algorithm specifically Self-training, Co-training and Tri-training in order to have the best possible accuracy which is effective. All classification accuracies were evaluated with the procedure scaling 10-fold cross-validation where the features was into individual pleats distributed evenly with all the grades inside the data set. The labeled data was evaluated in different portions such as 10%, 20% and 30% to measure the impact. The future work could be done with more categories of attributes the researchers added [1].

Amirah Mohamed Shahiria et al. has analyzed that the chosen attributes has an enormous influence on the quest of perceiving student's performance. Classification algorithms has different outcomes based on high potential attributes with every algorithm that has been

enforced in the study. The enforced classification algorithms were Neural Network with the elevated accuracy (98%), Decision tree (91%), Support Vector Machine (83%) which is same as the Naïve Bayes and K-Nearest Neighbor with the truncated accuracy (76%). The attributes which are most impactful has been recognized as the internal and external assessment for the highest accuracy gainer Neural Networks. Students CGPA was the most effective attribute for the Decision Tree algorithm. For the Support Vector Machine, psychometric variables were identified as suitable. Naïve Bayes technique used CGPA, high school background, scholarship, student demographic, social network interaction etc attributes [2].

Parneet Kaur et al. describes identification of the slow learners among students and exhibiting it by a probabilistic data mining model utilizing classification based algorithms. Genuine student's data is extracted and purification of the preferred potential variables is achieved by WEKA which is an Open Source Tool. After applying methods such as SMO, J48, Multilayer Perception, REP-Tree and Naïve Bayes the altercations are generated upon on all classification algorithms. The contrast of all five classifiers is concluded to detect accuracy along with fore-most performing classification algorithm amid all. The Multilayer Perception has come out with the best accuracy of 75% with the F-measure of 82%. The future prospect can be integrating data mining approaches with E-learning methods to find out the accuracy on different data-sets. Various fields are also eligible for this kind of research such as medicine, sports depending on the availability of enormous data-bases [3].

Farah Khaliq et al. stated in the study that students who are slow learners and the main reason behind it is their sensory and visual impairment. Perceptible neural alteration of optical information is one of the subsidiary element for their under achievement. Methods which has been used for this study is mostly psychological evaluation of their classroom behavioral or academic problems and standard IQ test has also been conducted to reach at a conclusion. The outcomes in the current analysis is indicative towards the fact that earlier indication of deprivation concerning optical performance at a initial phase of optical processing in the slow learners leads to one of the subsidiary reasons of the students incapability to produce a satisfactory outcome [4].

Sapan H Mankad proposes that this study targets on developing advanced education along with the co-operation of data mining algorithms. Insistence is on analyzing explicit variables regarding students and anticipation of their learning behavior being fast, medium or slow learners according to their pro-activeness in academically concerning issues. Exploratory outcomes regarding this analysis which use methods Naïve Bayes, Decision tree(C4.5) and Nearest Neighbor algorithm displays that decision tree produces comparatively fine result contrasted to Naïve Bayes and Nearest Neighbor techniques. This study additionally outlines few techniques to look for ways student attention to stay attached with classroom sessions and complete the course with full involvement. In future, they want to pursue this study by studying a larger dataset which can more efficiently describe a student's performance or behavior [5].

V L N Sastry et al. discusses in their paper about constructive as well as dynamic learning processes in instructing throughout several procedures. In the present situation, teaching in classroom through conventional methods such as lesson and discussion turns the students into submissive learners. Initiation process of minds of the students of several groups such as (fast, medium and slow learners) constructive learning methods are utilized in almost every academically inclined establishments. Methods which are used in this paper are- field trip strategy, Audio-Visual aid, Problem based learning, Project based learning, Role Play, Demonstration, Think Pair Share, Jigsaw Activity, KWL chart, Critical Debate, Student Seminar method, ICT learning ,Case Study, Poster Presentation etc. The above cited instructing techniques are interpreted in the classroom will be implemented for teaching and learning process. An exchange of view session is organized for the students to give their authentic assessment on these student enlightenment techniques. The best rated methods will be executed to achieve constructive-learning of the students [8].

Ali Akbar Arjmandnia and Keivan Kakabaraee expressed in their study about the fact which is viewpoint of parents with regard to their children who are slow learners in regular academic institutes. The factors which were most significant are the connection between parent's viewpoint and their age, how many friends the children's have and their respective grades. Other issues or factors were not so much of an influence on the study or we can say they were not that much significant. There were also some factors like parents

dissatisfaction because of inferior facilities in classrooms along with school-size. Parents did sanction teacher's encouraging as well as sympathetic connection towards their children. The method which was used in this case was regression analysis. Inclusive education means or inclusive education is the concept that the learning environment of the higher study institutions should change into a structure which supports or meets the learning needs of every student present in the environment. Inclusive education works with normalization principal, it means all the categories of students as per their learning ability (higher, medium or slow learners) should benefit the same rights or opportunities as per the students who are thought to be excellent. The facilities of this normalized educational environment should be planned by exact requirements, support of parents, teachers and class-mates [11].

Bernard Ugalde et al. describes in the study that the purpose of this paper is to look into as well as determination of the reasons regarding academic collapse and probation of students. There are several factors or variables on which the study was founded upon such as gender, semester offerings, level, student transcripts based on course classification etc. The significant findings were students most commonly who had failures in Math and English courses in IT, Engineering and Business Departments. This study was also able to determine on which courses help should be provided to students to achieve the performance which they intend to do. The identification of these courses also help the college to take necessary steps to make a strategy which will be more efficient in improving student's academic performance. The technique which was used was a matrix with a margin of error of 3% and confidence level of 95% from the total population of 350 students who were in probation during 2017-2018. The recommendation of researchers for future was to pass on the findings of this study to each department and improve their part [12].

Sangeeta et al uses Naïve Bayes which has an accuracy of 89.984% and the confusion matrix accuracy is 85%, J48 model which has the accuracy of 89.4569% and the confusion matrix 89.5%, k-NN model has the accuracy of 100% out of all 313 instances used throughout the classification and detection of slow learners. E .Balraj et al suggests a predictive model that will provide an educational system with the duty to identify students whom are in the dropout limit. The future plan is to study on vast repository of dropout

students along with additional data mining methods to regulate correspondence as well as connection between numerous components [13, 14].

Lorenz Kemper et al describes two procedures such as Decision tree and Logistic Regression to anticipate student dropout. The models are evaluated on the basis of data available at all universities without need of specific collection, examination data etc. Mayra Alban et al narrates University dropout is a worldwide difficulty which influences adversely on universities throughout the whole world along with outcomes such as reduced revenue for the university, reduced enrolment as well as financial loses for the state which funds the studies, constitutes a social problem for students, their families and society in common. In here, dataset has been categorically divided in five dimensions. They are personal, academic, economic, social and institutional. The algorithms which were applied are Neural Network classifier, Support vector machine, Decision tree classifier, A priori algorithm, K-Nearest neighbor classifier, Naïve Bayes , Logistical regression, Lineal regression, Discriminant analysis. Carlos Márquez-Vera et al outlines the conventional techniques for the anticipation of the student dropout which utilizes the entire data convenient at the course ending along with classification methods. As soon as possible the prediction can be made, the related personnel can start working to provide particular aid to students at risk of dropout as well as correlating to student's performance, viewpoint and behavior promptly [15, 16, 17].

Wanli Xing et al portrays the purpose of this study is to build an dropout prediction model based on other temporal prediction mechanisms or models which are mostly the traditional ones versus deep learning algorithm. Michalis Xenos et al narrates the purpose of this paper is to find out the rate of dropout and causes among the students in the course "Informatics" by using the university-level academic techniques of remote learning. Methods which were applied are statistical information of the students through student's registry, Tutor's records, Questionnaire based survey, telephone interview etc [18, 20].

Vrushali Mhetre et al provides an insight on how EDM plays a significant part in the prediction of student's educational accomplishment. The methods which has been used in this paper are J48, Naïve Bayes, Random Tree and ZeroR using Weka which is an open source tool. Sharayu N. Bonde et al lay outs a review paper containing other research

paper, EDM techniques and ml algorithms which has been used in these papers. Jae Young Chung et al provides the proceeding with the purpose of creating an early warning system which will help to know which students are in the verge of dropout and come up with a solution to help them. The random forest technique has been used to provide the result of the students who are very likely to drop out from the education program. The variables which were mostly effective in this study were “school register change” and the “reason of dropout” [7, 10, 19].

2.3 Comparative Analysis and Summary

We can see several machine learning classification algorithms has been used previously to solve this kind of issues in other research papers. As the number of EDM (Educational Data Mining) related work in the university level of our country is much lesser than it should be, our main motivation to work on this research area became more and more prominent over time. This slow learning issue is much bigger than anyone can anticipate at their first encounter with this research field. To be able to outline a much stronger structure of our educational background this kind of research work is the need of the moment. It can play a crucial part in making the educational sector much stronger as ever before. After much conceptualization of this research field and how previous researchers has done their research work, we selected Decision Tree , Random Forest , SVM (Support Vector Machine) , K Nearest Neighbor, Gaussian Naïve Bayes, Bernoulli Naïve Bayes and Multinomial Naïve Bayes classification algorithms to classify and predict slow learners. For the variables we have used several categories which were new in this field of research such as Problem in particular subject, Teacher’s attitude towards students, Attentive but not understanding, Student’s nature, Passionate about something else, Eager to be better at studies, Forcefully studying just to get a degree, Not necessary subject and responsiveness in class. As our dataset contained a number of non-identical categories of data, comparison with other researches was not attainable.

2.4 Scope of the Problem

After studying this area of research and related research work, we began to understand that

this kind of research work is much less in number in the perspective of our country and also with the university going students where this kind of research is much demanded for the sake of the students and for them to be able to being competent in their next step which is the job life. Detection of the slow learners in university level is a much needed area of research for the welfare of the students along with the country.

2.5 Challenges

The main challenges regarding this study was to manage enough data for a smooth output or we can say precise and more prominent output. After collecting the data, the main concern was too label the data in the categories they belong to which was a tough situation for this amount of data. Then the pre-processing stage which was the section that needed to be done very carefully in case of getting precise and better output.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The research area of our respective study analysis was Educational Data Mining (EDM), to gain ideas about how Educational Data Mining actually works in the field of our study analysis which was to classify slow, average and fast learners and also detecting slow learners among the students. The conceptual ideas we gained from EDM were not only for implementation of the classification algorithms but also the model structure, collection of categorized data which were mandatory for the implementation of the algorithms and also to get the desired result, pre-processing the raw data which was collected and of course training the model as well. On the other hand, Instrumentation is the part where the technologies we used were solely responsible and along with it are the methods to implement our model. We used Jupyter Notebook for the training and testing, implementation and accuracy checking purpose. The packages of python which were used such as – pandas, scikit-learn, numpy, sea born etc.

3.2 Data Collection Procedure / Data Set Utilized

To initiate our respective study, we used a database of Computer Science and Engineering (CSE) department students of Daffodil International University.

The database contained 1000 student data in the categories sought by us. With the help of Google Form we collected data from our CSE department students.

In the database (prepared by us), there are 29 different categories or labels and 1 Class (Learner) category which is divided into slow, average and fast. Some of the categories are- Grades, Study – hour, university environment, problem in any particular subjects and more. A table containing all the data variables is given below-

TABLE 3.2.1- DATA VARIABLES

VARIABLE NAME	DESCRIPTION	DOMAIN
1.GENDER	STUDENT'S GENDER	{M,F}
2.AGE	AGE OF STUDENT	{<18,18-20,>20}
3.W-F	WHERE ARE THEY FROM	{DHAKA,OUTSIDE OF DHAKA}
4.GRADES	GRADE OF STUDENTS	{A+,A,A-,B+,B,B-,C+,C,D}
5.S-HOUR	STUDY HOUR	{ 1-2,2-3,OTHER}
6.PSP	PROBLEM IN PARTICULAR SUBJECT	{SUBJECT NAME,NONE}
7.T-ATT	TEACHER'S ATTITUDE	{FRIENDLY,STRICT}
8.NETACS	STUDENT HAVING NET ACCESS	{YES,NO}
9.COM-HM	MOBILE/COMPUTER AT HOME	{YES,NO}
10.ATTEN-NU	ATTENTIVE BUT NOT UNDERSTANDING	{SUBJECT NAME,NONE}
11.EXT-CA	EXTRA CURRICULAR-ACTIVITIES	{YES,NO}
12.S-NAT	STUDENT'S NATURE	{INTROVERT,EXTROVERT}
13.S-CHO	SECTOR CHOICE	{ON YOUR OWN,FAMILY}

14.L-UNI	LIKE UNIVERSITY	{YES,NO}
15.PAS-ASE	PASSIONATE ABOUT SOMETHING ELSE	{YES,NO}
16.L-DEL	LECTURE DELIVERY	{ENGLISH,BANGLA}
17.VOSI	VISUAL OR SENSORY IMPAIRMENT	{YES,NO}
18.EBS	EAGER TO BE BETTER AT STUDIES	{YES,NO}
19.H-ISSUE	HEALTH ISSUES	{YES,NO}
20.FORCE-D	FORCEFULLY STUDYING TO GET A DEGREE	{YES,NO}
21.FAM-SUP	FAMILY SUPPORT	{YES,NO}
22.FAM-TYPE	FAMILY TYPE	{NUCLEAR,JOINT}
23.FAM-TYPE(2)	FAMILY TYPE(2)	{CONSERVATIVE,OPEN-MINDED}
24.AN-INC	ANNUAL INCOME	{LOW,MEDIUM,HIGH,VERY-HIGH}
25.UES	UNIVERSITY EDUCATION SYSTEM	{EXCELLENT,GOOD,POOR}
26.N-NEC	NOT NECESSARY SUBJECT	{SUBJECT NAME,NONE}
27.FAM-P	FAMILY PROBLEMS	{YES,NO}

28.S-MED	STUDY MEDIUM	{ENGLISH,BANGLA}
29.RESPONSE	RESPONSIVE IN CLASS OR NOT	{YES,NO}
30.Class(LEARNER)	KIND OF LEARNER	{AVERAGE,FAST,SLOW}

After labeling the data into desired categories, we pre-processed the raw data into numerical format. For the pre-processing method we have used Label Encoder Classifier.

3.3 Statistical Analysis

Statistical Analysis of the study is definitely one of the key points in carrying out the model precisely. To have a precise model, first of all our data needs to be precise, the more carefully and precisely we pre-process our data, the model will be as precise as that. The database with our desired categories needs to be pre-processed first, then the feature selection will take place, after that we will label our data into the train-test set, algorithms will be applied accordingly along with confusion matrix, classification report, accuracy and prediction method.

3.4 Proposed Methodology/Applied Mechanism

In this section, the methodology we have proposed for our analysis is represented. The methodology will describe how the workflow of our study will proceed.

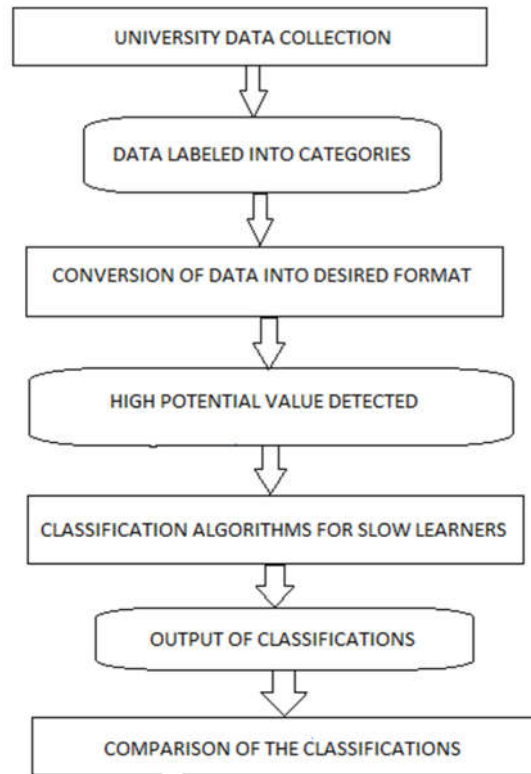


Fig 3.4.1: Flowchart of our respective study

After the pre-processing stage, we will apply Feature selection method called Extra Trees Classifier which we will import from scikit learn, this Classifier will show us which are the most important categories of data . Then, to find out the top 15 categories we will use the nlargest method.

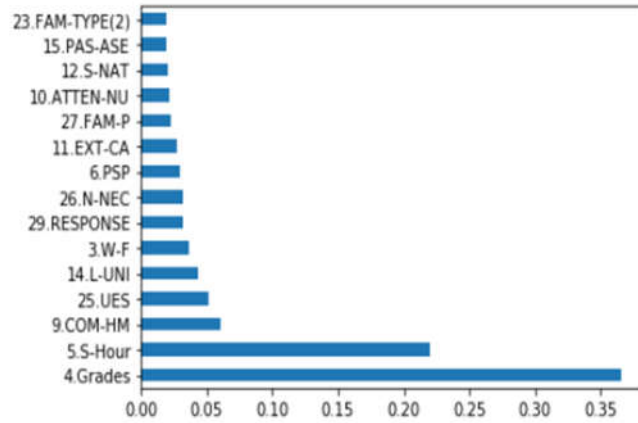


Fig 3.4.2: High Potential Data Categories in Bar chart

TABLE 3.4.1- HIGH POTENTIAL VARIABLES IN ASCENDING ORDER

VARIABLE NAME	DESCRIPTION	DOMAIN
1.GRADES	GRADE OF STUDENTS	{A+,A,A-,B+,B,B-,C+,C,D}
2.S-HOUR	STUDY HOUR	{ 1-2,2-3,OTHER}
3.COM-HM	MOBILE/COMPUTER AT HOME	{YES,NO}
4. UES	UNIVERSITY EDUCATION SYSTEM	{EXCELLENT,GOOD,POOR}
5.L-UNI	LIKE UNIVERSITY	{YES,NO}

6.W-F	WHERE ARE THEY FROM	{DHAKA,OUTSIDE OF DHAKA}
7.N-NEC	NOT NECESSARY SUBJECT	{SUBJECT NAME,NONE}
8.RESPONSE	RESPONSIVE IN CLASS OR NOT	{YES,NO}
9.PSP	PROBLEM IN PARTICULAR SUBJECT	{SUBJECT NAME,NONE}
10.EXT-CA	EXTRA CURRICULAR- ACTIVITIES	{YES,NO}
11.FAM-P	FAMILY PROBLEMS	{YES,NO}
12.ATTEN-NU	ATTENTIVE BUT NOT UNDERSTANDING	{SUBJECT NAME,NONE}
13.PAS-ASE	PASSIONATE ABOUT SOMETHING ELSE	{YES,NO}
14.S-NAT	STUDENT'S NATURE	{INTROVERT,EXTROVERT}
15.FAM-TYPE(2)	FAMILY TYPE(2)	{CONSERVATIVE,OPEN-MINDED}

After sorting the data categories which carries the most significance, we labeled the data in train-test set then application of Classification techniques were concluded.

1. Decision Tree,
2. Random Forest,
3. Gaussian Naïve Bayes,
4. Bernoulli Naïve Bayes,
5. Multinomial Naïve Bayes,
6. K-Nearest Neighbors,
7. SVM (Support Vector Machine).

We also applied confusion matrix for every classification algorithm, classification report method accuracy and prediction methods.

3.5 Implementation Requirements

To carry out the implementation procedure we will need our required environment.

The requirements:

Hardware/Software Essentials

- Hard Disk (minimum 500 GB)
- Operating System (Windows 10)
- Ram (Minimum 4 GB)

Developing Mechanism

- Python Environment
- Jupyter Notebook

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

In this section, the setup was built in the python environment called Jupyter Notebook to carry out the development procedures. The pre-processed dataset along with the high potential variables are ready to be applied with the classification algorithms, confusion matrix, classification report, accuracy check and prediction procedures in the Jupyter Notebook.

4.2 Experimental Results & Analysis

- **DECISION TREE**

First of all, we applied the Decision Tree Algorithm. We created an object called dtf then we fitted the data into xtrain and ytrain. The accuracy of Decision Tree Classifier was -

```
dtf.score(xtest,ytest)  
0.99
```

Fig 4.2.1: Decision Tree Accuracy.

The confusion matrix of Decision Tree algorithm –

```
confusion_matrix(ytest,pred)  
array([[ 60,   2,   0],  
       [  0,  93,   0],  
       [  1,   0, 144]], dtype=int64)
```

Fig 4.2.2: Confusion Matrix of Decision Tree Algorithm.

Classification report of Decision Tree algorithm –

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	0.98	0.97	0.98	62
Fast	0.98	1.00	0.99	93
Slow	1.00	0.99	1.00	145
accuracy			0.99	300
macro avg	0.99	0.99	0.99	300
weighted avg	0.99	0.99	0.99	300

Fig 4.2.3: Classification Report of Decision Tree Algorithm.

- **RANDOM FOREST**

The accuracy of Random Forest algorithm-

```
Rclf.score(xtest,ytest)
```

```
0.9766666666666667
```

Fig 4.2.4: Random Forest Accuracy.

Confusion matrix of Random Forest algorithm–

```
confusion_matrix(ytest,pred)
```

```
array([[ 55,   1,   6],  
       [  0,  93,   0],  
       [  0,   0, 145]], dtype=int64)
```

Fig 4.2.5: Confusion Matrix of Random Forest Algorithm.

Random Forest Classification report –

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	1.00	0.89	0.94	62
Fast	0.99	1.00	0.99	93
Slow	0.96	1.00	0.98	145
accuracy			0.98	300
macro avg	0.98	0.96	0.97	300
weighted avg	0.98	0.98	0.98	300

Fig 4.2.6: Classification Report of Random Forest Algorithm.

- **GAUSSIAN-NB**

Accuracy of GaussianNB-

```
GNB.score(xtest,ytest)
```

0.81

Fig 4.2.7: GaussianNB Accuracy.

Confusion matrix of GaussianNB Algorithm –

```
confusion_matrix(ytest,pred)
```

```
array([[ 35,  16,  11],
       [  1,  74,  18],
       [  5,   6, 134]], dtype=int64)
```

Fig 4.2.8: Confusion matrix of GaussianNB Algorithm.

Classification Report of GaussianNB Algorithm -

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	0.85	0.56	0.68	62
Fast	0.77	0.80	0.78	93
Slow	0.82	0.92	0.87	145
accuracy			0.81	300
macro avg	0.82	0.76	0.78	300
weighted avg	0.81	0.81	0.80	300

Fig 4.2.9: Classification Report of GaussianNB Algorithm.

- **BERNOULLI-NB**

The accuracy of BernoulliNB-

```
BNG.score(xtest,ytest)
```

```
0.6433333333333333
```

Fig 4.2.10: BernoulliNB Accuracy.

Confusion matrix of BernoulliNB Algorithm -

```
confusion_matrix(ytest,pred)
```

```
array([[ 13,  18,  31],  
       [  6,  47,  40],  
       [  1,  11, 133]], dtype=int64)
```

Fig 4.2.11: Confusion matrix of BernoulliNB Algorithm.

Classification Report of BernoulliNB Algorithm–

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	0.65	0.21	0.32	62
Fast	0.62	0.51	0.56	93
Slow	0.65	0.92	0.76	145
accuracy			0.64	300
macro avg	0.64	0.54	0.55	300
weighted avg	0.64	0.64	0.61	300

Fig 4.2.12: Classification Report of BernoulliNB Algorithm.

- **MULTINOMIAL-NB**

The accuracy of MultinomialNB–

```
MNB.score(xtest,ytest)
```

0.69

Fig 4.2.13: MultinomialNB Accuracy.

Confusion matrix of MultinomialNB Algorithm–

```
confusion_matrix(ytest,pred)
```

```
array([[ 12,  36,  14],  
       [ 12,  61,  20],  
       [ 11,   0, 134]], dtype=int64)
```

Fig 4.2.14: Confusion matrix of MultinomialNB Algorithm.

Classification Report of MultinomialNB Algorithm –

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	0.34	0.19	0.25	62
Fast	0.63	0.66	0.64	93
Slow	0.80	0.92	0.86	145
accuracy			0.69	300
macro avg	0.59	0.59	0.58	300
weighted avg	0.65	0.69	0.66	300

Fig 4.2.15: Classification Report of MultinomialNB Algorithm.

- **K NEAREST NEIGHBOR**

The accuracy of K Nearest Neighbor-

```
KN.score(xtest,ytest)
```

```
0.67
```

Fig 4.2.16: K Nearest Neighbor Accuracy.

Confusion matrix of K Nearest Neighbor -

```
confusion_matrix(ytest,pred)
```

```
array([[ 15,  23,  24],
       [ 14,  49,  30],
       [  4,   4, 137]], dtype=int64)
```

Fig 4.2.17: Confusion matrix of K Nearest Neighbor Algorithm.

Classification Report of K Nearest Neighbor Algorithm -

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	0.45	0.24	0.32	62
Fast	0.64	0.53	0.58	93
Slow	0.72	0.94	0.82	145
accuracy			0.67	300
macro avg	0.61	0.57	0.57	300
weighted avg	0.64	0.67	0.64	300

Fig 4.2.18: Classification Report of K Nearest Neighbor Algorithm.

- **SVM**

The accuracy of SVM-

```
SV.score(xtest,ytest)
```

```
0.7333333333333333
```

Fig 4.2.19: SVM Accuracy.

Confusion matrix of SVM algorithm-

```
confusion_matrix(ytest,pred)
```

```
array([[ 19,  29,  14],  
       [ 21,  61,  11],  
       [  4,   1, 140]], dtype=int64)
```

Fig 4.2.20: Confusion matrix of SVM algorithm.

Classification Report of SVM algorithm-

```
print(classification_report(ytest,pred))
```

	precision	recall	f1-score	support
Average	0.43	0.31	0.36	62
Fast	0.67	0.66	0.66	93
Slow	0.85	0.97	0.90	145
accuracy			0.73	300
macro avg	0.65	0.64	0.64	300
weighted avg	0.71	0.73	0.72	300

Fig 4.2.21: Classification Report of SVM algorithm.

Confusion Matrix –

Confusion Matrix happens to be one of the performance computation techniques in favor of ML classification algorithms. In here, the matrix output can be two or more classes. This matrix contains 4 distinct amalgamation of anticipated class and genuine class.

In our study analysis, it is a 3*3 performance matrix.

For Average Class, the matrix is –

TABLE 4.2.1- AVERAGE CLASS CONFUSION MATRIX

		Predicted		
		Average	Fast	Slow
Actual	Average	TP	FN	FN
	Fast	FP	TN	TN
	Slow	FP	TN	TN

For Fast Class, the matrix is –

TABLE 4.2.2- FAST CLASS CONFUSION MATRIX

		Predicted		
		Average	Fast	Slow
Actual	Average	TN	FP	TN
	Fast	FN	TP	FN
	Slow	TN	FP	TN

For Slow Class, the matrix is –

TABLE 4.2.3- SLOW CLASS CONFUSION MATRIX

		Predicted		
		Average	Fast	Slow
Actual	Average	TN	TN	FP
	Fast	TN	TN	FP
	Slow	FN	FN	TP

Comparison of the Algorithms -

To detect which machine learning algorithm has the best accuracy, we will have a comparison between the Classifiers and the scores are given below –

	Name	Score
0	Decision_Tree	0.986667
1	Random_Forest	0.973333
2	Gaussian_Naive_Bayes	0.810000
3	Bernoulli_Naive_Bayes	0.643333
4	Multinomial_Naive_Bayes	0.690000
5	Nearest_Neighbors	0.670000
6	SVM	0.733333

Fig 4.2.22: Algorithm Accuracies.

Comparison of classifiers with the help of Bar chart-

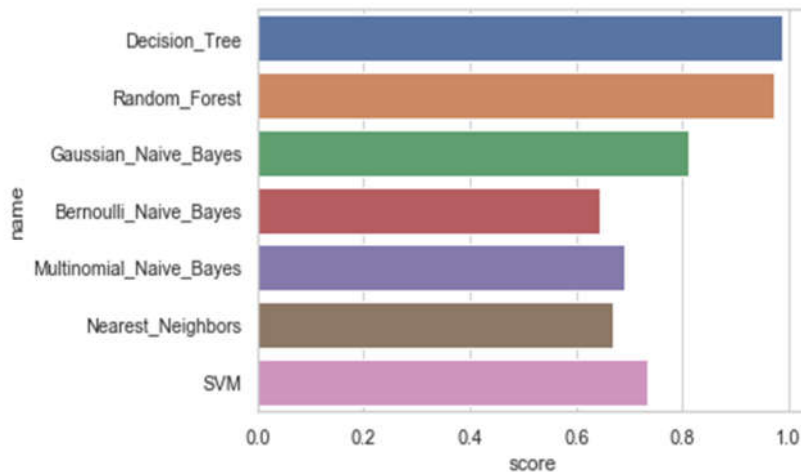


Fig 4.2.23: Comparison of Algorithm Accuracies.

We can see Decision Tree Classifier has the best accuracy 98.66%, Random Forest Classifier has the second best accuracy 97.33%, Gaussian Naïve Bayes has 81.00%, SVM Classifier has 73.33%. Multinomial Naïve Bayes, K Nearest Neighbor and Bernoulli Naïve Bayes has the least accuracy accordingly – 69.00%, 67.00% and 64.33%.

4.3 Discussion

For Confusion Matrix Accuracy the equation is –

The number of classes we correctly predicted out of all the classes is called the Accuracy. Accuracy should be high as much as possible.

$$\text{Accuracy} = (TN/TP) / (TN+FP+FN+TP)$$

For Precision –

Number of classes that we have predicted correctly actually how many of them are correct, this is called Precision.

$$\text{Precision} = TP / (TP+FP)$$

For Recall –

Recall minimizes false positives in the dataset.

$$\text{Recall} = TP / (TP+FN)$$

For F1-Score –

F1-score helps us to calculate the models Precision as well as Recall simultaneously.

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

TABLE 4.3.1- CLASSIFICATION REPORT OF ALL THE ALGORITHMS USED

Classification Algorithms	Class	TP Rate	FP Rate	Precision	Recall	F1-Score	Support
Decision Tree	Average	60	1	0.98	0.97	0.98	62
	Fast	93	2	0.98	1.00	0.99	93
	Slow	144	0	1.00	0.99	1.00	145
Random Forest	Average	55	0	1.00	0.89	0.94	62
	Fast	93	1	0.99	1.00	0.99	93
	Slow	145	6	0.96	1.00	0.98	145
Gaussian Naïve Bayes	Average	35	6	0.85	0.56	0.68	62
	Fast	74	22	0.77	0.80	0.78	93
	Slow	134	29	0.82	0.92	0.87	145
Bernoulli Naïve Bayes	Average	13	7	0.65	0.21	0.32	62
	Fast	47	29	0.62	0.51	0.56	93
	Slow	133	71	0.65	0.92	0.76	145
Multinomial Naïve Bayes	Average	12	23	0.34	0.19	0.25	62
	Fast	61	36	0.63	0.66	0.64	93
	Slow	134	34	0.80	0.92	0.86	145

K Nearest Neighbor	Average	15	18	0.45	0.24	0.32	62
	Fast	49	27	0.64	0.53	0.58	93
	Slow	137	54	0.72	0.94	0.82	145
Support Vector Machine	Average	19	25	0.43	0.31	0.36	62
	Fast	61	30	0.67	0.66	0.66	93
	Slow	140	25	0.85	0.97	0.90	145

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

As our country, considered as an emerging developing country, trying to be better at every aspect is a goal for our country. Education sector is considered one of the most important part of the structure of any country. If the literacy rate is better for any country that portrays that even if the country is financially behind, it has the potential to be better. To increase that literacy rate, we need our students to be very serious regarding their studies as their studies does not only affect themselves also the nation. As per our study analysis, our model detects the slow learners who can be given special assistance to be better at studies which accordingly help our literacy rate to grow day by day.

5.2 Impact on Environment

In this section, we will discuss about the impact on the study environment of our respective research. The goal of our study is mainly detecting the slow learners among the university level students who can be given assistance to get better at studies and achieve a good result which will not only benefit the student herself/himself also the university to get a good position and overall the country. This research of ours will generate a constructive effect upon the study environment the learners have and will encourage themselves to be better at studies.

5.3 Ethical Aspects

On the ethical ground, our study is very prominent. It is our ethical right to be better at our studies which is not only beneficial towards us but also the welfare of the country. If the education sector is getting better that means the structure of the country is getting stronger from within. A better educational sector means not only the students of the country but the country itself is entirely getting benefitted just at the cost of quality education in numerous sectors. If the base of the country is quality ensured which is education, the rest of the

sectors fall into their places easily and an educated person can differentiate between what is good and what is not when they are taking decisions as well.

5.4 Sustainability Plan

Sustainability plan of our study is to take our study on a more prominent and precise level. With more and more important data categories and significant number of machine learning algorithms which will give this study a more accurate output. We can extend this analysis in other fields also such as Career detection for students which will provide them with an insight of their careers and will also help them to choose their career wisely based on the output of the study.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

Overall summary of our study is we have chosen several machine learning algorithms along with confusion matrix, classification report, accuracy and prediction methods to be applied into the dataset containing various important data categories regarding the students which will help us to classify the students into three classes (slow, average, fast) and detect slow learners by which assistance can be provided to them and they can better at their studies.

6.2 Conclusions

The Decision tree Classifier has the accuracy of 98.66% which is the best among the seven classifiers, Random Forest has 97.33% which is the second best accuracy. After Decision Tree and Random Forest Classifier, Gaussian Naïve Bayes with accuracy 81.00% and Support Vector Machine with accuracy 73.33% comes accordingly. Multinomial Naïve Bayes, K Nearest Neighbor and Bernoulli Naïve Bayes possesses least accuracy accordingly – 69.00%, 67.00% and 64.33%. With these machine learning techniques and the selected high potential variables, we can definitely detect the slow learners who are getting behind for various reasons and also having problems with adaptation of the new learning environment. With this analysis, we can come forward in the educational system to fight the unspoken barrier which is the slow learning problem. This barrier has already caused us a lot as a nation because our education sector has not proven to be better nor it has taken steps by which actually help can be provided to the students to be better at their studies.

6.3 Implication for Further Study

In the future, with many more machine learning techniques and vast database with more and more data and data categories, we can work on this study more and better results can be provided. This sort of research can be done in several other sectors like Career Prediction of Students, Prediction of which course they are falling behind and many other. The

possibilities are vast in this area of research. Educational data mining is a sector of research which is flourishing day by day. The students are at the 35 most benefitted position if the respective analysis of studies regarding educational data mining take place. It can prove to be tremendously helpful not only for the students but the institution and as a whole for the entire nation. The quality of education should never be compromised and more and more studies in this sector should be done in order to have a better future of educational sector.

REFERENCES

- [1] Ioannis E. Iivieris, Vassilis Tampakas, Niki Kiriakiou, Tassas Mirkropoulos and Panagiotis Pintelas, “Forecasting Students Performance Using An Ensemble SSL Algorithm”, In: Springer Nature Switzerland AG – 2019, M. Tsitoridou et al. (Eds): TECH-EDU 2018, CCIS 993, pp. 566-581, 2019.
- [2] Amirah Mohamed Shahiria, Wahidah Husaina, Nur'aini Abdul Rashida, “A Review on Predicting Students Performance using Data Mining Techniques”, In: The Third Information Systems International Conference published by Elsevier, 2015.
- [3] Parneet Kaur a, Manpreet Singh, Gurpreet Singh Josan, “Classification and Prediction based Data Mining Algorithms to Predict Slow Learners in Education Sector”, In: 3rd International Conference on Recent Trends in Computing published by Elsevier, 2015.
- [4] Farah Khaliq, Yumnam Anjana And Neelam Vaney “Visual Evoked Potential Study In Slow Learners” Indian J Physiol Pharmacol; 53 (4): 341–346, 2009.
- [5] Sapan H Mankad “Predicting learning behaviour of students Strategies for making the course journey interesting”. In: IEEE 10th International Conference on Intelligent Systems and Control, 2016.
- [6] Dima Alhadidi, Zakaria Maamar, and Farhi Marir “Social Learning on the Move - A Research Roadmap”. In: IEEE International Conference on Web and Open Access to Learning, 2014.
- [7] Vrushali Mhetre, Prof. Mayura Nagar “Classification based data mining algorithms to predict slow, average and fast learners in educational system using Weka”. In: IEEE International Conference on Computing Methodologies and Communication (ICCMC), 2017.
- [8] V L N Sastry, K Srinivasa Rao, Nekkanti Venkata Rao, Prof Paul Clee and G Reena Kumari “Effective and Active Learning in Classroom Teaching through Various Methods”. In: IEEE 4th International Conference on MOOCs, Innovation and Technology in Education, 2016.
- [9] Faieza Chowdhury “Blended learning: how to flip the classroom at HEIs in Bangladesh?” Published in Emerald Insight, 2019.
- [10] Sharayu N. Bondel and D. K. Kirange2 “Educational Data Mining Survey for Predicting Student’s Academic Performance”, In - Springer Nature Switzerland AG 2020.
- [11] Ali Akbar Arjmandnia, Keivan Kakabaraee “THE INVESTIGATION OF PARENT’S ATTITUDE TOWARD INCLUSIVE EDUCATION FOR SLOW LEARNERS” published in iconte in 2011.
- [12] Bernard Ugalde, Yashir Ambula, Allan Salburo “Courses Affecting the Academic Performance of Students in Salalah College of Technology” in *International Journal of Computer Applications (0975 – 8887) Volume 178 – No. 11, May 2019*.

- [13] Sangeeta.K1, G.V.S.S.Naveen Babu2, Madhuri.G3 “Classification and Prediction of Slow Learners Using Machine Learning Algorithms” published in *International Journal of Computer Trends and Technology (IJCTT) – Volume 68 Issue 2 – Feb 2020*.
- [14] E. Balraj, D. Maalini, “A Survey on Predicting Student Dropout Analysis Using Data Mining Algorithms”, In : *International Journal of Pure and Applied Mathematics*. Volume 118 No.8 2018, 621-627 ISSN: 1311-8080 (printed version).
- [15] Lorenz Kemper, Gerrit Vorhoff & Berthold U. Wigger : “Predicting student dropout: A machine learning approach”, *European Journal of Higher Education*,2020.
- [16] Mayra Alban and David Mauricio “Predicting University Dropout through Data Mining: A Systematic Literature”. In: *Indian Journal of Science and Technology*, *January 2019*.
- [17] Carlos Márquez-Vera , Alberto Cano , Cristobal Romero , Amin Yousef Mohammad Noaman , Habib Mousa Fardoun and Sebastian Ventura “Early dropout prediction using data mining: a case study with high school students ”. In: *Wiley Publishing Ltd*,2015.
- [18] Wanli Xing¹ and Dongping Du² “Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention” in *Journal of Educational Computing Research* 0(0) 1–24 @ The Author(s) 2018.
- [19] Jae Young Chung, Sunbok Lee “Dropout early warning systems for high school students using machine learning”, 2018.
- [20] Michalis Xenos a,* , Christos Pierrakeas a, Panagiotis Pintelas b “A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University” published in *Elsevier*,2002.

APPENDICES

Abbreviations

1. ML – Machine Learning
2. EDM – Educational Data Mining
3. CSE – Computer Science and Engineering
4. SVM – Support Vector Machine
5. NB – Naïve Bayes

Sample Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Chittagon	A	other	None	Yes	None	Yes	Extrovert	Yes	Yes	Conservat	Good	None	Yes	Yes	Fast
Chittagon	A-	other	Algorithm	Yes	Data struc	No	Introvert	Yes	Yes	Conservat	Good	Art of livir	Yes	Yes	Fast
Dhaka	A	1-2 hours	None	Yes	None	Yes	Extrovert	No	Yes	Conservat	poor	None	Yes	Yes	Fast
Khulna	A-	other	None	Yes	None	Yes	Extrovert	Yes	Yes	Open-min	Excellent	None	Yes	Yes	Fast
Rajshahi	A	2-3 hours	None	Yes	None	No	Introvert	Yes	Yes	Open-min	Excellent	None	Yes	No	Average
Chittagon	B+	1-2 hours	Mathema	Yes	Mathema	Yes	Extrovert	No	No	Conservat	Good	None	Yes	Yes	Average
Dhaka	B+	2-3 hours	None	Yes	None	No	Extrovert	Yes	No	Conservat	Good	None	Yes	No	Average
Rangpur	A-	other	None	Yes	None	No	Introvert	Yes	No	Conservat	Good	None	Yes	No	Fast
Dhaka	A	1-2 hours	Algorithm	Yes	Algorithm	Yes	Introvert	Yes	Yes	Open-min	Good	Art of livir	Yes	Yes	Fast
Rajshahi	A	2-3 hours	None	Yes	Digital Ele	Yes	Introvert	Yes	Yes	Open-min	Good	Art of livir	Yes	Yes	Average
Rangpur	B+	other	Microproc	Yes	None	No	Introvert	Yes	Yes	Conservat	Excellent	None	Yes	Yes	Fast
Rangpur	A-	2-3 hours	None	Yes	None	Yes	Extrovert	Yes	Yes	Open-min	Good	Art of livir	No	Yes	Average
Chittagon	C+	other	None	Yes	None	No	Extrovert	No	Yes	Open-min	poor	None	No	No	Slow
Chittagon	D	other	None	Yes	None	Yes	Extrovert	No	Yes	Open-min	Good	None	No	Yes	Slow
Rajshahi	C	2-3 hours	None	Yes	None	Yes	Introvert	No	Yes	Conservat	Good	None	Yes	Yes	Slow
Rangpur	D	other	None	Yes	None	No	Extrovert	No	Yes	Conservat	Good	None	No	Yes	Slow
Dhaka	C+	1-2 hours	C	Yes	C	No	Extrovert	No	Yes	Conservat	Good	Art of livir	No	Yes	Slow
Dhaka	D	other	Algorithm	Yes	Algorithm	No	Introvert	No	Yes	Conservat	Good	None	Yes	Yes	Slow
Dhaka	A	1-2 hours	None	Yes	None	Yes	Introvert	No	No	Open-min	poor	None	No	No	Fast
Dhaka	B-	1-2 hours	Algorithm	Yes	Algorithm	No	Extrovert	Yes	Yes	Conservat	poor	None	Yes	Yes	Slow
Dhaka	C+	other	C, Algorithm	Yes	C	Yes	Introvert	Yes	Yes	Conservat	Good	None	No	No	Slow
Dhaka	C	other	Microproc	Yes	Microproc	No	Introvert	No	Yes	Open-min	Good	Economic	No	Yes	Slow
Dhaka	C+	other	Algorithm	Yes	Algorithm	No	Introvert	No	Yes	Conservat	Good	None	No	Yes	Slow

Fig A1: Sample Dataset

Plagiarism Report

Turnitin Originality Report	
Processed on: 30-Apr-2021 03:51 +06	
ID: 1573784037	
Word Count: 7494	
Submitted: 1	
Ela By Nusrat Jahan	
Similarity Index 13%	Similarity by Source Internet Sources: 10% Publications: 8% Student Papers: 7%

2% match (Internet from 25-Mar-2021) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5250/162-15-7955%20%287_%29.pdf?isAllowed=y&sequence=1
2% match (student papers from 20-Jan-2021) Submitted to Daffodil International University on 2021-01-20
1% match (Internet from 24-Jun-2019) https://link.springer.com/content/pdf/10.1007%2F978-3-030-20954-4.pdf
< 1% match (Internet from 04-Aug-2020) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/4092/P15438%20%286_%29.pdf?isAllowed=y&sequence=1
< 1% match (Internet from 07-Apr-2021) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5444/192-25-773%20%2823%25%29.pdf?isAllowed=y&sequence=1
< 1% match (Internet from 02-Apr-2021) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5208/162-15-7927%20%2818_%29.pdf?isAllowed=y&sequence=1
< 1% match (Internet from 12-Feb-2021) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5180/162-15-7911%20%2810_%29.pdf?isAllowed=y&sequence=1

