# PREDICTING STUDENTS PERFORMANCE THROUGH INTERNET USAGE BY ARTIFICIAL NEURAL NETWORK

RABEYA KHATUN RINKY

This Thesis Report submitted in partial fulfillment of the requirements
for the award of the degree of
Master of Engineering in Information and Communication Engineering

Department of Information and Communication Engineering
NOAKHALI SCIENCE AND TECHNOLOGY UNIVERSITY

JANUARY 2021

## SUPERVISOR'S DECLARATION

We hereby declare that we have checked this project and in our opinion, this project is adequate in terms of scope and quality for the award of the degree of Master of Engineering in Department of Information and Communication Engineering of Noakhali Science and Technology University.

____________________________________

Name of Supervisor: DR. MD. ASHIKUR RAHMAN KHAN

Position: PROFESSOR & CHAIRMAN

Date:

____________________________________

Name of Co-Supervisor: ZAYED-US-SALEHIN

Position: ASSISTANT PROFESSOR

Date:

## STUDENT'S DECLARATION

I hereby declare that the work in this project is my own except for quotations and summaries which have been duly acknowledged. This project report has not been accepted for any degree and is not concurrently submitted for the award of other degrees.

**Name: RABEYA KHATUN RINKY**
**Roll No: BKH1411MSC108F**
**Date:**

## ACKNOWLEDGEMENTS

I am grateful and would like to express my sincere gratitude to my supervisor Professor Dr. Md. Ashikur Rahman Khan for his germinal ideas, invaluable guidance, continuous encouragement, and constant support in making this research possible. He has always impressed me with his outstanding professional conduct, his strong conviction for science, and his belief that a Msc program is only the start of a life-long learning experience. I appreciate his consistent support from the first day I applied to the graduate program to these concluding moments. I am truly grateful for his progressive vision about my training in science, his tolerance of my naive mistakes, and his commitment to my future carrier. I also would like to express very special thanks to my co-supervisor Zayed-Us-Salehin for his suggestions and co-operations throughout the study. I also sincerely thanks you for the time spent proofreading and correcting my many mistakes.

My sincere thanks go to all my lab mates and members of the staff of the Information and Communication Engineering Department, NSTU, who helped me in many ways and made my stay at NSTU pleasant and unforgettable.

I acknowledge my sincere indebtedness and gratitude to my parents for their love, dream, and sacrifice throughout my life. Special thanks should be given to my committee members. I would like to acknowledge their comments and suggestions which were crucial for the successful completion of this study.

## ABSTRACT

Students academic performance primarily depends on their internet usage for various activity purpose. Predicting students academic success is critical for educational institutions. Because strategic programs can be planned in improving or maintaining students' performance during the period of their studies. The estimation in this study is done by a multilayer perceptron neural network model. 18 parameters which contain students internet usage data captured by a questionnaire chosen as input layer parameters. Hidden nodes will be determined experimentally. The output level values which define success of the students. The Back-Propagation algorithm is used for training of ANN. The mean squares of the errors are used as a performance (error) function with its goal set to zero. In conclusion, the application done with comparison with other ML algorithms. It is recommended that a research with same parameters would be better results with higher participation. In this paper we perform our experiment for school, college and university students which helps us to indicate  them separately.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| NN | Neural Network |
| KNN | K-Nearest Neighbor |
| SVM | Support Vector Machine |
| MLP | Multilayer Perceptron |
| TP | True Positive |
| FP | False Positive |
| TN | True Negative |
| FN | False Negative |
| ERR | Error Rate |
| ACC | Accuracy |
| PREC | Precision |

# CHAPTER 1

# INTRODUCTION

## 1.1 INTRODUCTION

The art of the using artificial neural network in predicting student performance is characterized by the application of Artificial Intelligence techniques. It seems natural to use analogies when making decision or prediction, as by definition they contain information about how people have behaved in similar situation in the past. We may explain human behavior by assuring that decisions are made by analogy (logic form reasoning) with previous experiences, hence students' early performance in tertiary institutions can be used to predict the success rate.

Internet plays a fundamental role in organizations and society. The basic fact justifies the information revolution that has taken place across the globe in recent times. The Internet has linked thousands of nations and enterprises across the world. Hence, the world which appears physically large has been made to be too small by Internet and refers to as "global village". Academic and non-academic education represents one of the key roles in the process of a country modernization and achieving competitive advantage. The process of education and development of human resources is becoming increasingly important concept. It can be said that the new social elite is coming from academic circles with the aim to stabilize and improve the overall state of the economy and society. It is used to determine students who would be expected do well. The quality of students enlisted into any university influences the research and training level within the university, and furthermore, has an extensive effect, as these students may become a master planners in all sorts of the economy. Such factors as score of subject, number of credits passed, student cumulative grade point average of freshman year and gender etc. were then used as input variables. In this research, supervised learning algorithms are used in order to make a system for early prediction of the success of students based on their success on the exams on the year of their studies.

It should be noted that this module of predicting student performance will help in identifying which student possibly will succeed. In view of the fact that students play important part in the university customs, student performance has long been a key issue for all kinds of educational organizations. So far, the adoption of information technology makes student's information management easy and efficient. In accordance with the purpose of the study, when the created artificial neural network is examined to check which parameters provide the highest accuracy score and the weights between the nodes on the network, which independent variable contributes to prediction of student final performance at what level is examined?

## 1.2 PROBLEM STATEMENT

As many researchers have turned their attention to predicting students' performance researchers in this field manage to innovative solutions to improve performance Thus, students' performance assessment may not be viewed as being somewhat separate from learning process. It is a continuous and an integral part of learning processes. By revealing what students what they need to learn, it enables teachers to build on existing knowledge and pro-vide appropriate scaffolding. If such information is timely and specific, it can serve as a valuable feedback to both teachers and students so that it will improve student performance. This section will discuss the results analysis of the recent works in predicting students' performance. The result on prediction accuracy is depending on the attributes or features that were used during the prediction process. Neural Network method gave the highest prediction accuracy because of the influence from main attributes. The main advantage of Neural Network is the ability to capture nonlinear relationships easily. It is also referred as adaptive system due to its ability to readily update the historical data like a human brain. Students play important part in the university customs, student performance has long been a key issue for all kinds of educational organizations. Early prediction of students' success could help teachers to identify students that have potential for advanced courses and also students who need the additional education in order to improve their knowledge. On the other side, developed system could be useful for students, so that they could estimate their future success in their studies on the basis of their learning habits, their work and grades. This could help them to identify on time whether they need an additional effort in order to achieve the success they want. Generally, performance predictions based on student scores are made, but overall students' learning management system is not focused. This study carried out the desired system with artificial neural networks. Prediction of student performance is one of the most important subjects of educational data mining. Artificial neural networks are seen to be an effective tool in predicting student performance in e-learning environments. Hence this study is to tackle the problem of academic failure by seeking ways to make the process more effective. Specifically the study seeks to investigate the possibility of using neural network model to forecast the performance of a student before graduating the student. Data from all sector of education are gathered for evaluating i.e. school, college and university. So the evaluated statements are of great use in learning.

## 1.3 OBJECTIVES OF THE STUDY

i. The objective of the study is to create, develop and optimize an artificial neural network for prediction of student final performance in online learning environments.
ii. To studying and identifying the variables that are used in analyzing performance.
iii. The final performance of the students was estimated from the students' behaviors of navigating in the e-learning environment.
iv. To reduce the problem of academic failure by seeking ways to make the process more effective, efficient and reliable.
v. Generally, performance predictions based on student scores are made, but overall students' learning management system is not focused. This study carried out the desired system with artificial neural networks.

## 1.4 SCOPE OF THE STUDY

Educational organization are one of the important parts of our society and plays a vital role for growth and development of any nation. Educational data analyzing is an emerging research area that deals with the development of methods to explore data originating in an educational context. Predicting students' performance becomes more challenging due to the large volume of data in educational databases. Most of the time institutions need observe student's performance for their future development. Despite the importance of predicting student's performance, collecting such data from the real-world is a challenging task. For predicting the student's performance at first we collect student's relevant data. Here we use 700 student's data for the research which is taken from various school, college and university. 300 from Noakhali science and Technology University, 200 from 2 colleges situated at maijdee, Noakhali, and other 200 from three school situated at Sonapur, Noakhali. We collect these data from the students by offline survey. In this survey we consider 20 factors. In these factors, we regard as not only academic data but also others personal information such as family, living area, academic result, social media interaction and mostly internet usage etc. After collecting data these are analyzed with algorithmic methods. The following hardware, software requirements, programming language and algorithm are used for analysis.

**Hardware Requirement:** Computer (Intel core i5, windows 10/Ubuntu operating system)

**Software Requirement**: Anaconda Navigator, Jupyter notebook,Spyder, Google colab

**Programming Language:** python.

**Algorithms:** Back propagation, KNN, SVM.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 INTRODUCTION

Substantial studies based on neural networks have been conducted on data from schools, colleges, and distance-education courses, aiming at predict student achievement. When the studies aimed at estimating student performance are examined, it is noteworthy that the variables that have a certain percentage effect on performance are often used. Some of these reviewed studies are mentioned below.

## 2.2 RELATED WORKS

Oladokun, V.O., et al 2008 developed and tested various factors that may likely influence the performance of a student were identified. Such factors as ordinary level subjects' scores and subjects' combination, matriculation examination scores, age on admission, parental background, types and location of secondary school attended and gender, among others, were then used as input variables for the ANN model. A model based on the Multilayer Perceptron Topology was developed and trained using data spanning five generations of graduates from an Engineering Department of University of Ibadan, Nigeria's first University.Oancea, B., et al 2013 used a neural network to predict the students' results measured by the grade point average in the first year of study. For this purpose we used a sample of 1000 students from "Nicolae Titulescu" University of Bucharest from the last three graduates' generations, 800 being used for training the network and 200 for testing the network. The neural network was a multilayer perceptron (MLP) with one input layer, two hidden layers and one output layer and it was trained using a version of the resilient backpropagation algorithm. The input data were the students profile at the time of enrolling at the university including information about the student age, the GPA at high school graduation, the gap between high school graduation and higher education enrolling. After training the network we obtained MSE of about 1.7%. The ability to predict students' results is of great help for the university management in order to take early action to avoid the phenomenon of leaving education.Livieris et al., 2012 describes the implementation of a user-friendly software tool based on neural network classifiers for predicting the student's performance in the course of "Mathematics" of the first year of Lyceum. The authors of the paper compared many training algorithms namely the Broyden-FletcherGoldfarb-Shanno (BFGS), the Levenberg-Marquardt (LM), the Resilient Backpropagation (Rprop) and the modified spectral Perry (MSP). The performances of the neural network classifier were also compared with other classifiers such as Bayesian networks and support vector machines and the authors proved the better results of the neural network classifier.Data registered by Moodle were used in Calvo-Flores et al., 2006 for predicting students' marks. The authors used a RBF neural network and information logged by Moodle regarding the number and types of educational resources accesses to predict the students' marks at a discipline. They obtained an accuracy of prediction of about 75-80%. The model developed in this paper shows that it is possible to predict those students with problems to pass a course and to give professors an indication to pay more attention on those students that probably will fail passing a course.In El Moucary et al., 2011 the authors present a use Neural Networks and Data Clustering based method designed to predict students' GPA according to their foreign language performance for those students who study in a foreign language. In a second stage the students are grouped in well-defined cluster for further advising. They obtained a maximum error of prediction less than 10% for GPA.Oladokun et al., 2008 used a neural network in order to determine those factors that influence students' performance. They classified students in three classes according to their results. The accuracy of prediction obtained by the authors of the paper was about 74%.Karamouzis et al., 2008 used a three-layer perceptron network trained by back propagation to predict student graduation rate. The network model built by the authors had an accuracy of 70.27% for successful graduates and 66.29% for unsuccessful graduates.

Neural networks were also used in Naik et al., 2004 to predict MBA student success. The authors classified applicants to MBA program into successful and marginal student pools based on undergraduate GPA, undergraduate major, age, GMAT score using a neural network with three layers. They obtained overall prediction accuracy for their model of about 89%. To assess the ability of the neural network for classification of student, the authors compared the results obtained using the neural network with a logit and probit regression model. The overall rate of accuracy in prediction of the logit model was about 73% while the accuracy of the profit was 73.37%. Ibrahim Z., et al 2007 defined three predictive models have been developed using SAS Enterprise Miner that are, artificial neural network, decision tree and linear regression. The result of this study shows that all of the three models produce more than 80% accuracy. It also shows that artificial neural network outperforms the other two models. Rajeswari, K., et al 2014 evaluated students' performance and some attributes are selected which generate rules by means of association rule mining.. Artificial neural network checks accuracy of the results. A Multi-Layer Perceptron Neural Network is employed for selection of interesting features using 10 – fold cross validation.The artificial neural network selects 5 out of 8 attributes based on the accuracy obtained for correctly classified data. It is observed that in association rule mining important rules are generated using these selected attributes. The Experiment is conducted using Weka and real time data set available in the college premises. Shahiri, A.M. et al 2015 proposed a systematical literature review on predicting student performance by using data mining techniques to improve students achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict students performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a students data. We could actually improve students achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions. Siri, A., 2015. aims at contribute to the continuing debates on the possibilities how to reduce the student failure and improve educational processes with the help of data mining techniques, in particular of the artificial neural networks. The population consists of 810 students enrolled for the first time in a health care professions degree course at the University of Genoa in the academic year 2008-09. The research is based on the analysis of data and information originating from primary sources:

administrative data related to the careers of students; statistical data collected during the research through an ad hoc survey; data derived from telephone interviews with students who had not completed the enrolment in the subsequent years. The neural network correctly predicted 84 percent of the cases of group 1,81 percent of the cases of group 2, and 76 percent of the cases belonging to group of dropouts. The application of the artificial neural network model can offer a valid tool to design educational interventions to deliver to those who score high in the level of risk.Mueen, A., et al 2016 analyze students' academic performance based on their academic record and forum participation. Educational Data Mining (EDM) is an emerging tool for academic intervention. The educational institutions can use EDU for extensive analysis of students’ characteristics. In this study, we have collected students’ data from two undergraduate courses. Three different data mining classification algorithms (Naïve Bayes, Neural Network, and Decision Tree) were used on the dataset. The prediction performance of three classifiers are measured and compared. It was observed that Naïve Bayes classifier outperforms other two classifiers by achieving overall prediction accuracy of 86%. This study will help teachers to improve student academic performance. Lau, E.T., et al 2019 presents an approach with both conventional statistical analysis and neural network modelling/prediction of students’ performance. Conventional statistical evaluations are used to identify the factors that likely affect the students’ performance. The neural network is modelled with 11 input variables, two layers of hidden neurons, and one output layer. Levenberg–Marquardt algorithm is employed as the backpropagation training rule. The performance of neural network model is evaluated through the error performance, regression, error histogram, confusion matrix and area under the receiver operating characteristics curve. Overall, the neural network model has achieved a good prediction accuracy of 84.8%, along with limitations. Aydoğdu, Ş., 2020. carried out with artificial neural networks, performance predictions based on student scores are generally made, but students’ use of learning management system is not focused. In this study, performances of 3518 university students, who studying and actively participating in a learning management system, were tried to be predicted by artificial neural networks in terms of gender, content score, time spent on the content, number of entries to content, homework score, number of attendance to live sessions, total time spent in live sessions, number of attendance to archived courses and total time spent in archived courses variables. Since it is difficult to interpret how much input variables in artificial neural networks contribute to predicting output variables, these networks are called black boxes. Also, in this study the amount of contribution of input variables on the prediction of output variable was also examined. The artificial neural network created as a result of the study makes a prediction with an accuracy of 80.47%. Finally, it was found that the variables of number of attendance to the live classes, the number of attendance to archived courses and the time spent in the content contributed most to the prediction of the output variable. Ahmad, Z. et al 2018 Predict the chance of students being at risk (AR) or not 'Not at risk' (NAR) with respect to their degree. Population of study consisted of all students of social sciences studying in 4th semester and they enrolled in 2007 session of BS and MA/MSc program at University of Gujrat Hafiz Hayat Campus. By using stratified sampling with proportional allocation method, a sample of 300 students was selected. We have used Multilayer Perception Neural Network Model to predict the chance of students being at risk (AR) or not 'Not at risk' (NAR) with respect to their degree on the basis of CGPA at the end of 2nd semester, Study time. Previous degree marks, Home environment, Study habits Learning skills, Hardworking and Academic interaction. In classifying the students at risk/not at risk, we could achieve a rate of correct classification of over 95% in training sample and over 85% in holdout sample. The estimated models can be used to predict the students being at risk or not with respect to their degree. Iyanda, A.R., et al 2018 obtained result show that although Multilayer Perceptron had prediction accuracy of 75%, Generalized Regression Neural Network had a better accuracy. The response time of Generalized Regression Neural Network (0.016sec) was faster than Multilayer Perceptron (0.03sec) and its memory consumption size (5kb) lower than that of Multilayer Perceptron (8kb). The simulated models were further compared with t-test method using a confidence interval of 95%. The attained t-test result from p-value (0.6854) suggests acceptance of null hypothesis, which shows that there is no significant difference between the predicted Grade Point Average and the actual Grade Point Average. The findings therefore reveal that the overall performance of Generalized Regression Neural Network outperforms the Multilayer Perceptron model with an accuracy of 95%. The study concluded that Generalized

Regression Neural Network model which was simulated and with 95 % accuracy could be deployed by educationists to predict students' academic performance using single performance factor. In Asogwa, O.C. et al 2015 a Soft computing technique, which artificial neural network is part of, has been recognized as attractive alternatives to the standard, well-established hard computing paradigms. The main objective of this work is to evaluate the accuracy of the Artificial Neural Network Model Architecture developed under a statistical programme (Matlab 2009) as a classifier, using its Mean Correct Classification Rate CCR (%), among others is: to report the order of predictor's significance to the model programmed. Kotsiantis, S., et al 2010 proposed therefore a better proposal is an online ensemble of classifiers that combines an incremental version of Naive Bayes, the 1-NN and the WINNOW algorithms using the voting methodology. Among other significant conclusions it was found that the proposed algorithm is the most appropriate to be used for the construction of a software support tool. Karamouzis, S.T. et al 2008 showed declining student graduation rates is a significant and growing problem in higher education. Students are dropping out from colleges for a variety of reasons and school administrators are scrambling to increase graduation rates. Predicting student graduation is of great value to schools and an enormous potential utility for targeted intervention. Considering the promising behavior of Artificial Neural Networks (ANNs) as classifiers led us into the development, training, and testing of an ANN for predicting student graduation outcomes. The network was developed model as a three-layered perceptron and was trained using the backpropagation principles. The average predictability rate for the training and test sets were 77% and 68%, respectively. Kotsiantis, S., et al 2004 explained the ability to predict a student's performance could be useful in a great number of different ways associated with university-level distance learning. Students' key demographic characteristics and their marks on a few written assignments can constitute the training set for a supervised machine learning algorithm. The learning algorithm could then be able to predict the performance of new students, thus becoming a useful tool for identifying predicted poor performers. The scope of this work is to compare some of the state of the art learning algorithms. Two experiments have been conducted with six algorithms, which were trained using data sets provided by the Hellenic Open University. Among other significant conclusions, it was found that the Naïve Bayes algorithm is the most appropriate to be used for the construction of a software support tool, has more than satisfactory accuracy, its overall sensitivity is extremely satisfactory, and is the easiest algorithm to implement. Kalejaye, B.A., et al 2015 focuses on the use of artificial neural network (ANN) model for predicting students' academic performance in a University System, based on the previous datasets. The domain used in the study consists of sixty (60) students in the Department of Computer and Information Science, Tai Solarin University of Education in Ogun State, who have completed four academic sessions from the university. The codes were written and executed using MATLAB format. The students' CGPAs from first year through their third year were used as the inputs to train the ANN models constructed using nntool and the Final Grades (CGPA) served as a target output. The output predicted by the networks is expressed in-line with the current grading system of the case study. CGPA values simulated by the network are compared with the actual final CGPA to determine the efficacy of each of the three feed-forward neural networks used. Test data evaluations showed that the ANN model is able to predict correctly, the final grade of students with 91.7% accuracy. Abass et al., 2011 applied another technique of Artificial Intelligence (AI) i.e., case-base reasoning (CBR) to predict student academic performance based on the previous datasets using 20 students in the Department of Computer Science, TASUED as the study domain. The high correlation coefficient observed between the actual student CGPA and the CBR prediction also proved the efficiency of AI techniques in this type of assignment. Lykourentzou et al., 2009 predicted the final grades of students in e-learning courses with multiple feed-forward neural networks using self-designed multiple-choice test data of students of as input. In this study Rastrollo-Guerrero, J.L., et al 2020 almost 70 papers were analyzed to show different modern techniques widely applied for predicting students' performance, together with the objectives they must reach in this field. These techniques and methods, which pertain to the area of Artificial Intelligence, are mainly Machine Learning, Collaborative Filtering, Recommender Systems, and Artificial Neural Networks, among others.Konar, D., et al 2020 provides a vivid explanation about dataset preparation and reports outcome on grade prediction. A hybrid fuzzy neural network classifier has been employed to target model students' academic profiles, relying on inexact information. In the decision tree method, the property of each generated node is evaluated based on information gain theory and attributed with the highest information gain in terms of entropy reduction in the level of maximum. The SVM is introduced as a supervised classifier targeting to solve linear and nonlinear classification. The GA is employed to leverage the SVM parameters for a fixed kernel, and a data driver kernel construction is introduced. Different accuracy values are generated by the different kernel types, and the polynomial type kernel reported the highest accuracy.Waheed, H., et al 2020 show the proposed model to achieve a classification accuracy of 84%–93%. We show that a deep artificial neural network outperforms the baseline logistic regression and support vector machine models. While logistic regression achieves an accuracy of 79.82%–85.60%, the support vector machine achieves 79.95%–89.14%. Aligned with the existing studies - our findings demonstrate the inclusion of legacy data and assessment-related data to impact the model significantly. Students interested in accessing the content of the previous lectures are observed to demonstrate better performance. The study intends to assist institutes in formulating a necessary framework for pedagogical support, facilitating higher education decision-making process towards sustainable education. Al-Shehri, H., et al 2017 Forecasting the performance of students can be useful in taking early precautions, instant actions, or selecting a student that is fit for a certain task. The need to explore better models to achieve better performance cannot be overemphasized. Most of earlier work on the same dataset used K-Nearest Neighbor algorithm and achieved low results, while Support Vector Machine algorithm was rarely used, which happens to be a very popular and powerful prediction technique. To ensure better comparison, we applied both Support Vector Machine algorithm and K-Nearest Neighbor algorithm on the dataset to predict the student's grade and then compared their accuracy. Empirical studies outcome indicated that Support Vector Machine achieved slightly better results with correlation coefficient of 0.96, while the K-Nearest Neighbor achieved correlation coefficient of 0.95. Bhutto, E.S., et al 2020 introduces students' academic performance prediction model that uses supervised type of machine learning algorithms like support vector machine and logistic regression. The results supported by various experiments using different technologies are compared and it is showed that sequential minimal optimization algorithm outperforms by achieving improved accuracy as compared to logistic regression. And the knowledge found through this research can help educational institutes to predict the future behavior of students so that they can categorize their performance into good or bad. The objective is not just to predict future performance of students but also provide the best technique for finding the most impactful features to work on like teacher's performance, student's motivation that will eventually decrease the student's dropout ratio, a multi-class support vector machine (MSVM) predictor was built in this study. The performance of the MSVM predictor was examined using educational dataset of students from the University of Lagos, Nigeria. Findings from our experiment show that the MSVM with K-fold (K=7) cross validation adequately predicted the performances of students across all categories. Tsai, C.F. et al 2008 investigate the performance of a single classifier as the baseline classifier to compare with multiple classifiers and diversified multiple classifiers by using neural networks based on three datasets. By comparing with the single classifier as the benchmark in terms of average prediction accuracy, the multiple classifiers only perform better in one of the three datasets. The diversified multiple classifiers trained by not only different classifier parameters but also different sets of training data perform worse in all datasets. However, for the Type I and Type II errors, there is no exact winner. We suggest that it is better to consider these three classifier architectures to make the optimal financial decision. Agarwal, S., et al 2017 presents a method for predicting spot prices using techniques of artificial neural network. It also discusses the proposed methodology in detail and shows the experimental evaluation and results on various instances of Amazon Elastic Compute Cloud (EC2). For most of the instances, percentage error varies from 1% to 8.6% using the recurrent neural network technique. The proposed technique reduces the percentage of average error significantly in-comparison to other technique.In this paper Poudel, P. et al 2017 presents solar power output prediction using long shortterm memory (LSTM) network of artificial neural network. The neural network is applied to model solar power data obtained from the experimental site in the Hae-Nam, Korea. LSTM neural network with one input node, two hidden layers and one output node is used to model a day solar power data. The results obtained from the comparison of LSTM neural network and moving average indicate that the LSTM neural network approach is reasonable for short term solar power prediction.

## 2.3 SUMMARY

Recently, many researchers have turned their attention to predicting students' performance. They have contributed to the related literature. By and large, researchers in this field manage to advocate novel and smart solutions to improve performance Thus, students' performance assessment may not be viewed as being somewhat separate from learning process. It is a continuous and an integral part of learning processes. By revealing what students what they need to learn, it enables teachers to build on existing knowledge and pro-vide appropriate scaffolding. If such information is timely and specific, it can serve as a valuable feedback to both teachers and students so that it will improve student performance. This section will discuss the results analysis of the recent works in predicting students' performance. The result on prediction accuracy is depending on the attributes or features that were used during the prediction process. Neural Network method gave the highest prediction accuracy because of the influence from main attributes. The main advantage of Neural Network is the ability to capture nonlinear relationships easily. It is also referred as adaptive system due to its ability to readily update the historical data like a human brain. So, the model always functions beyond the knowledge base. In addition, the strength of neural network is the ability to learn from a limited set of data. Next is Support Vector Machine with the performance accuracy around. Based on the analysis, psychometric factors are the most suitable attributes in predicting students' performance. e. Conversely, K-Nearest Neighbor showed high accuracy with the combination of three attributes, which are internal assessment, CGPA and extra-curricular activities in predicting students' performance. When comparing with the other two methods, which are Decision Tree and Nave Bayes the accuracy results are lower than K-Nearest Neighbor method. From the literature review we see that all works about performance prediction is only in one sector in which our work reflects in three sector school, college and university. We collect data and perform analysis in every sector of studies which has a great impact on our educational system. Generally, performance predictions based on student scores are made, but overall students' learning management system is not focused.

# CHAPTER 3

# METHODOLOGY

## 3.1 INTRODUCTION

In this research artificial neural network, k nearest neighbor and support vector machine is used for analysis. This section involves these three algorithms working principle and algorithm at how they works. Artificial neural network is used to predict the performance of students. Artificial neural network model is evaluated with required algorithm. After that KNN and SVM machine learning algorithms are used for testing accuracy in accordance with ANN to find which model works best in calculating performance. Various performance evaluation metrics are also applied for data assessment e.g. confusion metrics, correlation matrix, accuracy, precision, recall etc. which helps our research consequently. In the following section 3.2 overview of the methodology is described, the overall process is shown in a graph figure and then step by step processing is also showed along with algorithmic view. Then section 3.3 the data collection and preprocessing method that include collecting, preprocessing, encoding and standardizing of the dataset. In section 3.4 the models (ANN, KNN and SVM) are developed along with their algorithms and code shown in appendix. At last in section 3.5 the performance evaluation metrics are described with their educational form such as confusion matrix, accuracy, precision, recall etc.

## 3.2 OVERVIEW OF THE METHODOLOGY

The overview of the system is shown in Figure 1. It depicts the overall idea how the prediction will happen. First of all, we collect data from educational institutions. After collecting the data, we preprocess it for the data mining tool. In this case, we use Jupiter notebook 6.0.3, spider and Google colab in which multi paradigm programming language python is coded and data mining operations are conducted. Using this notebook, we organized the data. Then we design a supervised neural network which trains the data and extracts information from the data. In our experimental case, we predict the student's performance. We also use Levenberg Marquardt back propagation algorithm to train the neural network. In this system, we predict student's performance and then we find out the accuracy percentage. Each part of our propose system is described in the following subsections.

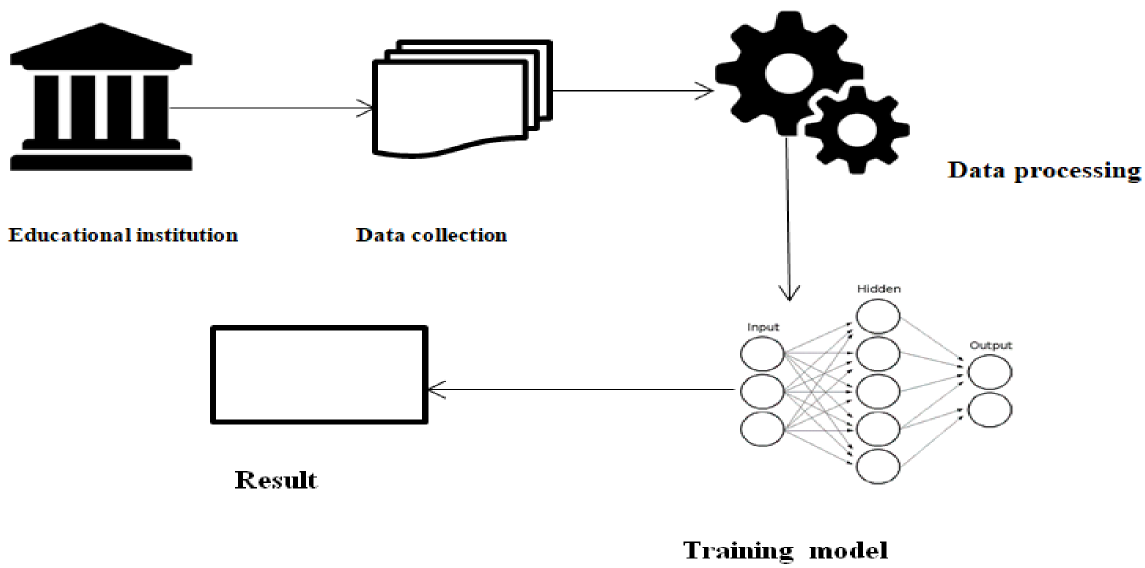



**Figure 3.1:** System Overview

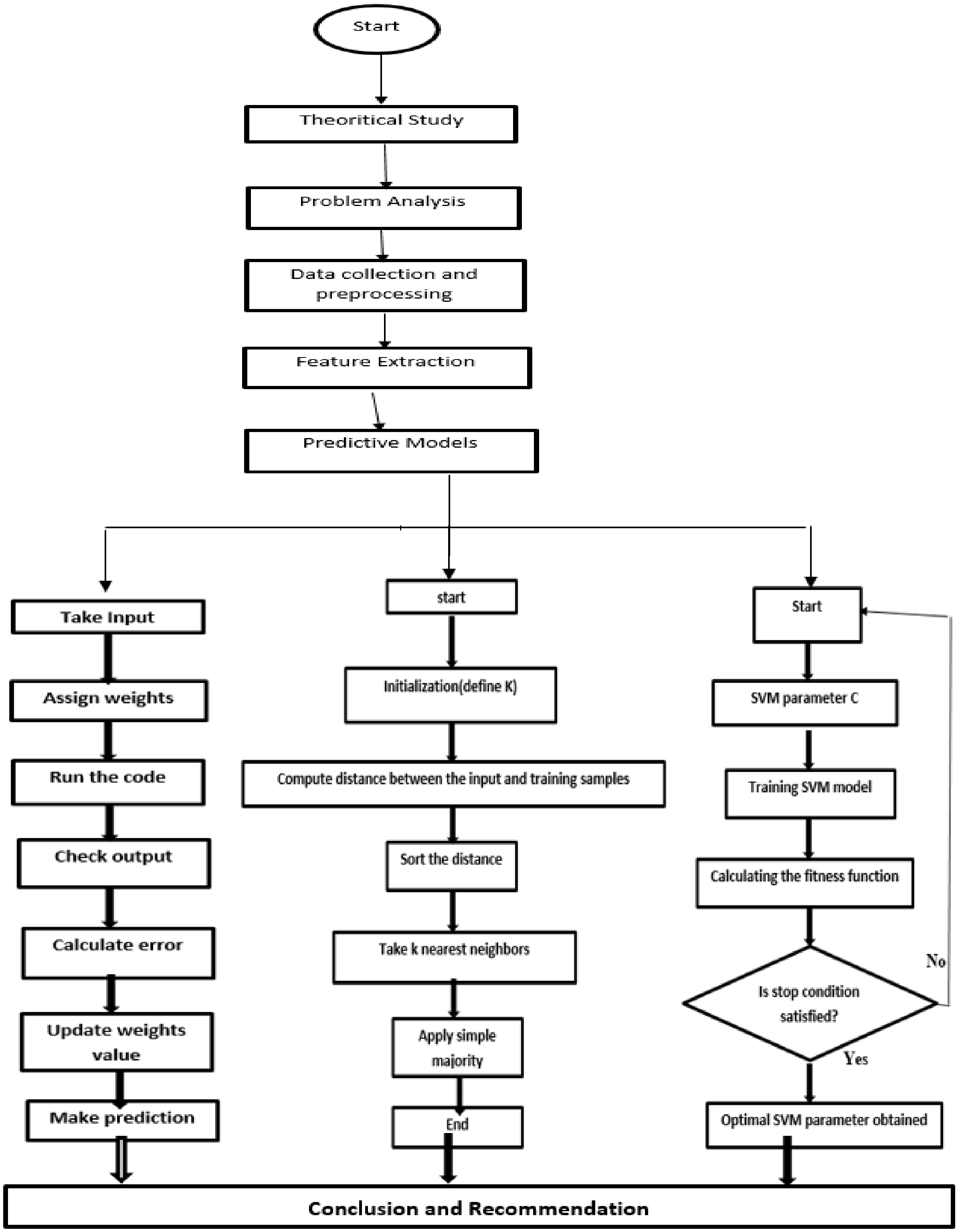



**Figure 3.2:** Work Flow Diagram

## 3.3 DATA COLLECTION AND PREPROCESSING

### 3.3.1 Parameters

Student's performance is especially significant for institutions because its raking depends on student's quality. Most of the time institutions need observe student's performance for their future development. Despite the importance of predicting student's performance, collecting such data from the real-world is a challenging task. For predicting the student's performance at first we collect student's relevant data. Here we use 700 student's data for the research which is taken from various school, college and university. We collect these data from the students by offline survey. In this survey we consider 20 factors. In these factors, we regard as not only academic data but also others personal information such as family, living area, social media interaction and mostly internet usage etc. Questionnaire from survey are given below.

**Table 3.1:** Dataset Description

| Variable Name | Type | Value |
|---|---|---|
| Gender | Nominal | Male, Female |
| Age | Numeric | 14-27 |
| Popular Website | Nominal | Google, You Tube, Whatsapp, Messenger, Instagram, Twitter, Gmail, Others |
| Skillness | Nominal | Good, Average, Better<br>Not at All |
| Medium | Nominal | Laptop, Mobile, Desktop<br>Laptop and Mobile |
| Location | Nominal | Home, Library, University<br>Cyber, School, College |
| Household Internet Facilities | Nominal | Connected, Not Connected |
| Browse Time | Nominal | 6am-12pm, 12pm-6pm<br>6pm-12am, 12am-6am |
| Browsing Status | Nominal | Daily, Weekly, Monthly |

| | | |
|---|---|---|
| Residence | Nominal | Village, Town, Remote |
| Total Internet Use | Nominal | 0_5, 5_10, 10_15<br>15_20, 20_24, Rarely |
| Time Spend in Academic | Nominal | 0_5, 5_10, 10_15<br>15_20, 20_24, Rarely |
| Purpose of Use | Nominal | Social Sites, File Sharing<br>News, Group Study<br>Music, Shopping |
| Years of Internet Use | Nominal | >1, 1 To 5, 5 To10, 10< |
| Browsing Purpose | Nominal | Academic, Non-Academic |
| Online Course | Nominal | Yes, No |
| Webinar | Nominal | Yes, No |
| Class Performance | Nominal | Good, Average<br>Satisfactory, Not Satisfactory |
| Obstacles | Nominal | Slow, Payment, Power Failure<br>Poor Computer Skill, Lack of<br>Computer, Cost of Computer<br>Equipment |
| Result | Numeric | From 2.50 to 4.00 |

### 3.3.2 Data Collection

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. Primary data are collected from two school situated in Sonapur, Noakhali. Three college student’s data are also collected for reporting. A considerable amount of data on students’ performance is collected at Noakhali science and Technology University. Here we collect 200 school data, 200 college data and 300 university data. A total of 700 data is collected during the session.

### 3.3.3 Data Preprocessing

In recent years, education institute collect a vast amount of data. As the volume of data increase day by day the relationship underneath the data is becoming more challenging. Hence, we need to filter out some unnecessary data. For this research, we used Anaconda navigator as the data mining software for filtering these unnecessary data. It works with numerical value that is why we convert the data into different numerical values. Then we divide the twenty factors in two parts. One part is the input part which uses nineteen factors among these twenty factors and remaining one factor is called target value. This target value is the most essential part of the data because the neural network will find the pattern of the data and predict the result based on the target value. Figure 2 shows the steps of data reprocessing.

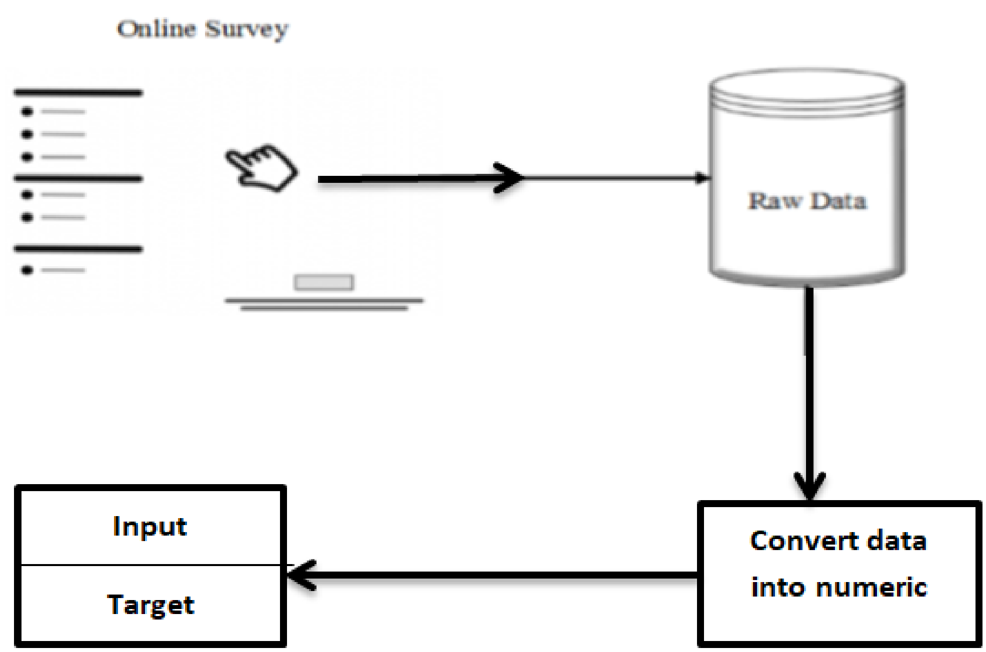



**Figure 3.3:** Preprocess the data

### 3.3.4 Encoding Data

Machine learning algorithms don’t work so well with processing raw data. Before we can feed such data to an algorithm, we must preprocess it. In other words, we must apply some transformations on it. With data preprocessing, we convert raw data into a clean data set. At first, we have to import various libraries of python, then insert the csv file of data. The file contains values of categorical data, which is also known as the raw data set. We have to convert them to numeric data as they are in category. There have some missing values. To handle these missing values a library known as Simple Imputer is used which replace the NaN values with the most frequent value. If we look at the csv file there we have some data in categorical form and some are numeric. Therefore, we want all of our data to be numerical form. We want it to be some continuous variable. These are called dummy variables. Label Encoder is used to label the data. X have 19 independent variables that are converted into dummy variables and y has 1 dependent variable that is already in categorical format. After preprocessing here is the processed data. Here we also handle missing variables using most frequent values. The code shown in appendix section A.

| gender | age | Popular website | proficiency | medium | location | household internet facilities | browsing time | browsing status | residence | total internet use | time spend in academic | purpose of use | Years of internet use | Browsing purpose | Online course | Webinar | Class performance | obstacles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 14 | 4 | 1 | 2 | 0 | 0 | 4 | 0 | 1 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 1 | 5 |
| 0 | 14 | 4 | 2 | 2 | 0 | 1 | 4 | 0 | 1 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 5 |
| 0 | 14 | 0 | 2 | 2 | 0 | 1 | 4 | 0 | 1 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 1 | 5 |
| 1 | 15 | 4 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 1 | 2 |
| 1 | 14 | 4 | 2 | 1 | 2 | 0 | 4 | 0 | 1 | 3 | 3 | 3 | 2 | 1 | 0 | 0 | 1 | 5 |
| 1 | 16 | 0 | 2 | 2 | 0 | 1 | 4 | 0 | 1 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 14 | 4 | 2 | 2 | 0 | 1 | 1 | 2 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 0 |
| 0 | 14 | 0 | 0 | 2 | 0 | 1 | 3 | 0 | 1 | 0 | 3 | 6 | 2 | 0 | 0 | 0 | 0 | 4 |
| 0 | 15 | 4 | 2 | 2 | 0 | 1 | 4 | 0 | 1 | 0 | 3 | 3 | 2 | 1 | 0 | 0 | 1 | 5 |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 0 | 15 | 4 | 2 | 2 | 0 | 1 | 4 | 0 | 1 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 1 | 5 |
| 1 | 15 | 4 | 2 | 2 | 0 | 1 | 4 | 2 | 1 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 1 | 5 |
| 1 | | | | | | | | | | | | | | | | | | |

### 3.3.5 Standardizing Data

Standardization and Scaling Data mean all data is to be on the same scale, so that it is not like some data is in the thousand scale and some is to be 1 to 10 scale. It improves prediction capability of our model. In standardization feature scaling is the most important part of data processing. We see in our dataset that some value is very high and some are very low. This will cause some issues in our machinery model. To solve the problem we set all values on the same scale, that is called standardization.

$$X_{stand} = \frac{\boldsymbol{X - mean(X)}}{\boldsymbol{Standard\ Deviation(X)}} \qquad 3.1$$

Here standard scalar import from sklearn library. The required code is in appendix section. After standardization data lies between -1 to +1 that shows below

```
In [63]: x_test

Out[63]: array([[ 3.23270501,  0.68074565, -0.26726124, ..., -1.81265393,
                 -0.47380354, -0.18569534],
                [-0.69896325,  0.68074565,  3.74165739, ...,  0.55167728,
                 -0.47380354, -0.18569534],
                [-0.69896325, -1.46897745, -0.26726124, ...,  0.55167728,
                 -0.47380354, -0.18569534],
                ...,
                [-0.69896325,  0.68074565, -0.26726124, ..., -1.81265393,
                 -0.47380354, -0.18569534],
                [ 0.61159284,  0.68074565, -0.26726124, ...,  0.55167728,
                 -0.47380354, -0.18569534],
                [ 0.61159284,  0.68074565, -0.26726124, ...,  0.55167728,
                 -0.47380354, -0.18569534]])

In [64]: x_train

Out[64]: array([[ 0.50411838, -1.25462109, -0.19316685, ...,  0.83373974,
                 -0.36066785, -0.2123977 ],
                [-0.84342882,  0.7970534 , -0.19316685, ...,  0.83373974,
                 -0.36066785, -0.2123977 ],
                [-0.84342882,  0.7970534 , -0.19316685, ..., -1.19941506,
                 -0.36066785, -0.2123977 ],
                ...,
                [ 0.50411838, -1.25462109, -0.19316685, ..., -1.19941506,
                 -0.36066785, -0.2123977 ],
                [-0.84342882,  0.7970534 , -0.19316685, ...,  0.83373974,
                 -0.36066785, -0.2123977 ],
```

### 3.3.6 Data Analyzing

*(a) Describing the dataset*

Using the method describe (), we can find out parameters like count, mean, std, min, 25%, 50%, 75% and max of the data.

```
In [31]: student_data.describe()
Out[31]:
```

| | Age | Target | male | others | whatsapp | yahoo | you tube | b |
|---|---|---|---|---|---|---|---|---|
| **count** | 199.000000 | 199.000000 | 199.000000 | 199.000000 | 199.000000 | 199.000000 | 199.000000 | 199.00 |
| **mean** | 14.597990 | 0.296482 | 0.633166 | 0.045226 | 0.070352 | 0.010050 | 0.658291 | 0.09 |
| **std** | 0.751561 | 0.457858 | 0.483156 | 0.208324 | 0.256384 | 0.099997 | 0.475479 | 0.28 |
| **min** | 14.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 |
| **25%** | 14.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.00 |
| **50%** | 14.000000 | 0.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 1.000000 | 0.00 |
| **75%** | 15.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 1.000000 | 0.00 |
| **max** | 17.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.00 |

8 rows × 39 columns

*(b) Shape of the dataset*

Simply, the shape tuple will give us the dimensions of the dataset.

```
In [32]: student_data.shape
Out[32]: (199, 20)
```

*(c) Extracting data from the dataset*

Now if we want only the first ten rows from the dataset, we can call the head() method on it. To this, we can pass it the argument 10.

In [3]:
```
student_data.head(10)
```

Out[3]:

| | Gender | Age | Popular website | proficiency | Medium | Location | household internet facilities | Browse time | browsing status | Residence | Total internet use | Time spend in academic | Purpose of use | Years of internet use | Browsing purpose | Online course |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | female | 14 | you tube | better | mobile | home | connected | 6pm-12am | daily | town | 0_5 | 0_5 | music | >1 | acdemic | no |
| **1** | female | 14 | you tube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | gaming | >1 | acdemic | no |
| **2** | female | 14 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no |
| **3** | male | 15 | you tube | better | mobile | home | not connect | 6am-12am | daily | town | 0_5 | 0_5 | music | 1 to 5 | non-academic | no |
| **4** | male | 14 | you tube | good | laptop and mobile | school | connected | 6pm-12am | daily | town | rarely | rarely | music | >1 | non-academic | no |
| **5** | male | 16 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no |
| **6** | male | 14 | you tube | good | mobile | home | not connect | 12pm-6pm | weekly | village | 0_5 | 0_5 | music | 1 to 5 | acdemic | no |
| **7** | female | 14 | google | average | mobile | home | not connect | 6am-12pm | daily | town | 0_5 | rarely | social sites | >1 | acdemic | no |

## 3.4 DEVELOPED MODEL

This is most important phase which includes model building for prediction of performance. In this we have implemented three machine learning algorithms for prediction. These algorithms include Artificial Neural Network, Support Vector Classifier and K-Nearest Neighbor. First discussing about the neural network and then it will be compared with other two algorithms.

### 3.4.1 Artificial Neural Network Model

An Artificial Neural Network (ANN) is an arithmetical model that is motivated by the organization and/or functional feature of biological neural networks. A neural network contains an interrelated set of artificial neurons, and it processes information using a connectionist form to computation. As a general rule an ANN is an adaptive system that adjusts its structure based on external or internal information that runs through the network during the learning process. Recent neural networks are non-linear numerical data modeling tools. They are usually used to model intricate relationships among inputs and outputs or to uncover patterns in data. ANN has been applied in numerous applications with considerable attainment. Neurons are often grouped into layers. Layers are groups of neurons that perform similar functions. There are three types of layers. The input layer is the layer of neurons that receive input from the user program. The layer of neurons that send data to the user program is the output layer. Between the input layer and output layer are hidden layers. Hidden layer neurons are only connected only to other neurons and never directly interact with the user program. The input and output layers are not just there as interface points. Every neuron in a neural network has the opportunity to affect processing. Processing can occur at any layer in the neural network. Not every neural network has this many layers. The hidden layer is optional. The input and output layers are required, but it is possible to have on layer act as both an input and output layer. ANN learning can be either supervised or unsupervised. Supervised training is accomplished by giving the neural network a set of sample data along with the anticipated outputs from each of these samples. Supervised training is the most common form of neural network training. As supervised training proceeds the neural network is taken through several iterations, or epochs, until the actual output of the neural network matches the anticipated output, with a reasonably small error. Each epoch is one pass through the training samples. Unsupervised training is similar to supervised training except that no anticipated outputs are provided.

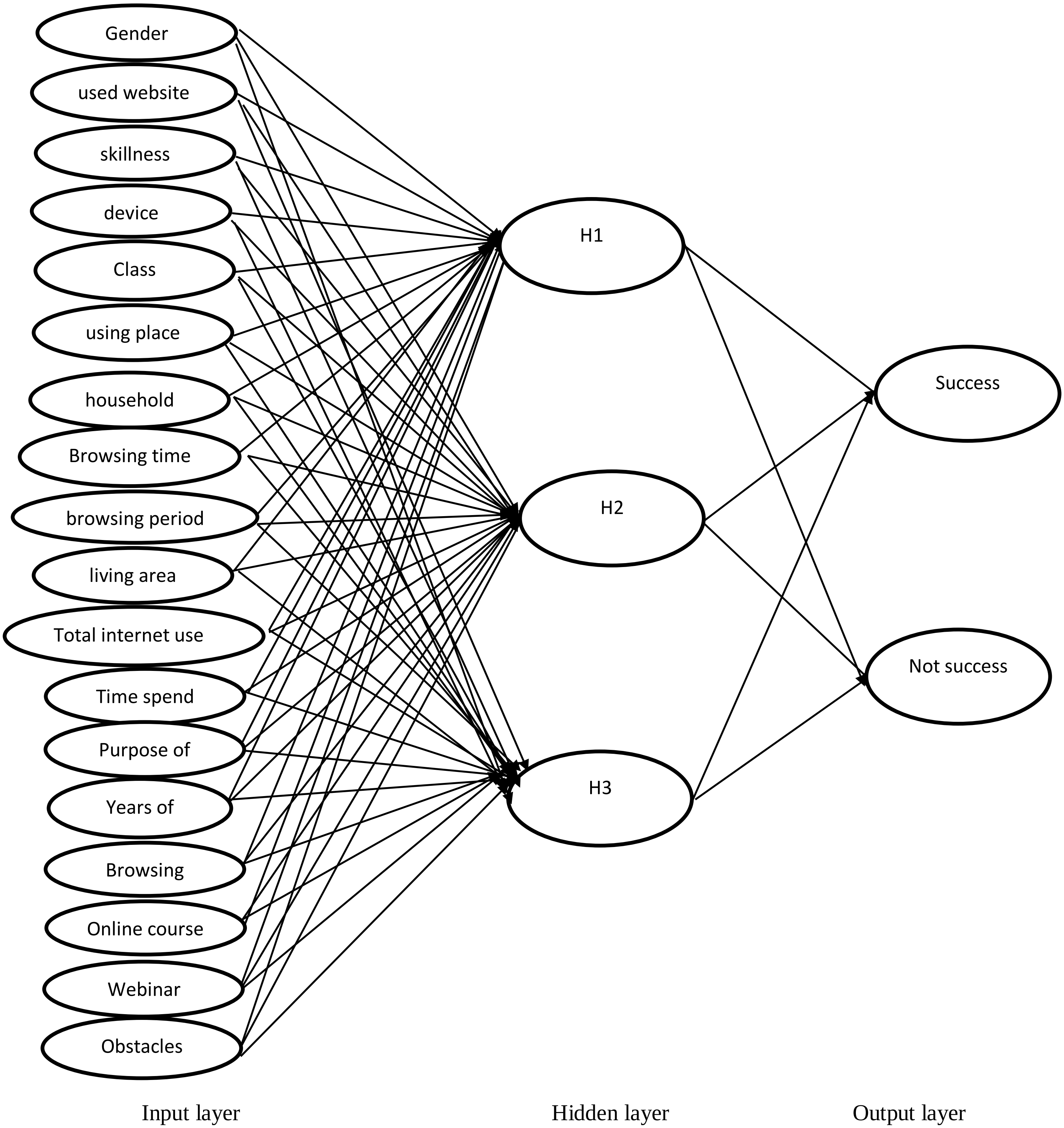



**Figure 3.4:** Schematic view of artificial neuron network

*(a) Model Description*

Steps involved in the evaluation of a neural network:

A neural network executes in 2 steps:

*Feed forward:*

On a feed forward neural network, we have a set of input features and some random weights. Notice that in this case, we are taking random weights that we will optimize using backward propagation.

*Backpropagation:*

During back propagation, we calculate the error between predicted output and target output and then use an algorithm (gradient descent) to update the weight values.

*Summarizing an Artificial Neural Network:*

Take inputs
Add bias (if required)
Assign random weights to input features
Run the code for training.
Find the error in prediction.
Update the weight by gradient descent algorithm.
Repeat the training phase with updated weights.
Make predictions.

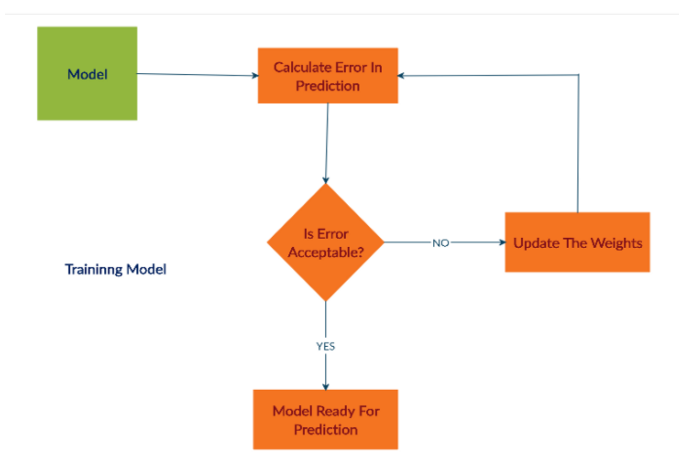



**Figure 3.5:** Training process

*(b)Training Process*

At the time of training process we have to give each input a weight, which can be a positive or negative number. An input with a large positive weight or a large negative weight, have a strong effect on the neuron's output. Before starting, we set each weight to a random number. Then begin the training process:

1. Take the inputs from a training set example, adjust them by the weights, and pass them through a special formula to calculate the neuron's output.
2. Calculate the error, which is the difference between the neuron's output and the desired output in the training set example.
3. Depending on the direction of the error, adjust the weights slightly.
4. Repeat this process 10, 000 times.

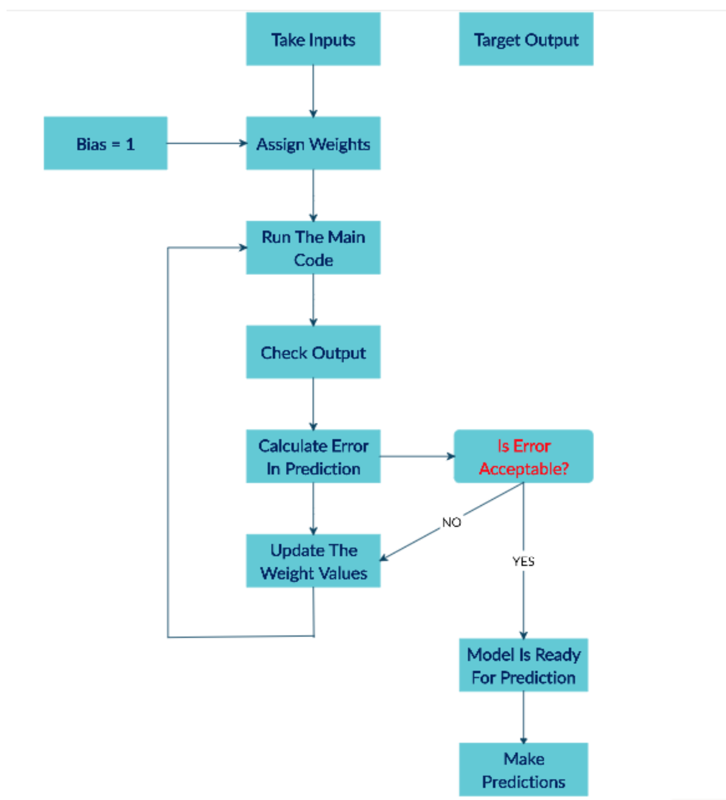



**Figure 3.6:** Flow chart for a simple neural network

Eventually the weights of the neuron will reach an optimum for the training set. If we allow the neuron to think about a new situation, that follows the same pattern, it should make a good prediction.

*(c)Neuron'sOutput*

To define the formula for calculating the neuron's output, first we take the weighted sum of the neuron's inputs, which is:

$\sum \mathit{weight}_i.\mathit{input}_i = \mathit{weight1.input1} + \mathit{weight2.input2} + \mathit{weight3.input3} + \ldots + \mathit{weighti.inputi}$

Next normalize this, so the result is between 0 and 1. For this, use a mathematically convenient function, called the sigmoid function:

$$\frac{1}{1+e^{-x}} \tag{3.2}$$

If plotted on a graph, the sigmoid function draws an S shaped curve.

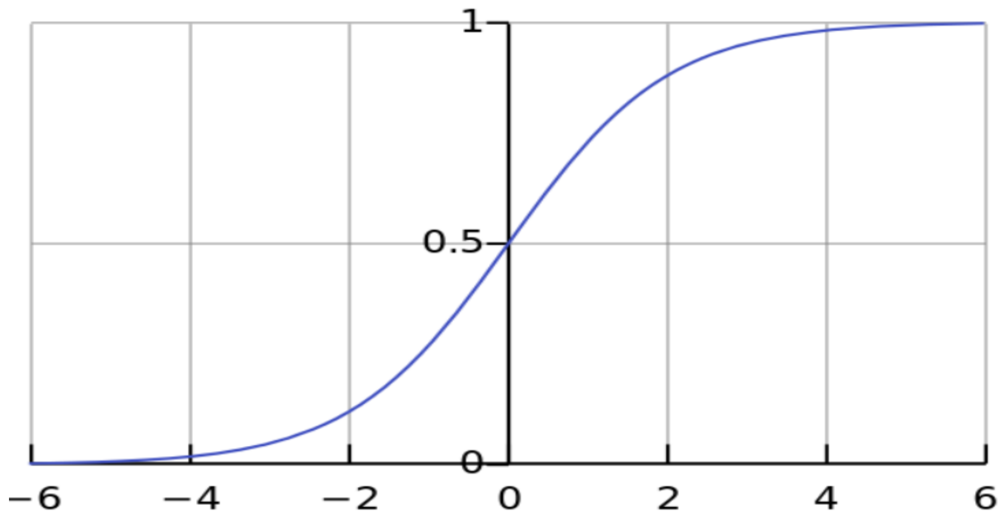



**Figure 3.7:** Sigmoid function curve

So by substituting the first equation into the second, the final formula for the output of the neuron is:

$$\frac{1}{1+e^{-(\sum weighti.inputi)}} \quad (3.3)$$

You might have noticed that we're not using a minimum firing threshold, to keep things simple.

*Formula for adjusting the weights*

During the training cycle we adjust the weights. To adjust the weights we can use the "Error Weighted Derivative" formula:

Adjust weights by = error. SigmoidCurveGradient(output) . input (3.4)

**Gradient Descent**

Gradient Descent is a machine learning algorithm that operates iteratively to find the optimal values for its parameters. It takes into account, user-defined learning rate, and initial parameter values.

*(d) Model Implementation*

- Start with initial values.
- Calculate cost.
- Update values using the update function.
- Returns minimized cost for our cost function

Generally, what we do is, we find the formula that gives us the optimal values for our parameter. However, in this algorithm, it finds the value by itself!

***Formula for the Gradient Descent algorithm***

$$x = x - lr * \frac{d}{dx} f(x) \quad (3.5)$$

where,

x = input

F(x) = output based on x

Lr = learning rate

We are going to update our weight with this algorithm. First of all, we need to find the derivative **f(X).**

***Derivation of the formula used in single layer perceptron***

***Phase 1:***

$$\frac{\partial Error}{\partial w} = \frac{\partial Error}{\partial outo} * \frac{\partial outo}{\partial ino} * \frac{\partial ino}{\partial w} \quad (3.6)$$

**i.** ***Finding the first derivative:***

$$\frac{\partial Error}{\partial outo} = \text{output-target} \quad (3.7)$$

*ii.* ***Finding the second derivative:***

$$\frac{\partial outo}{\partial ino} = \text{outo(1-outo)} \quad (3.8)$$

*iii.* ***Finding the third derivative:***

$$\frac{\partial ino}{\partial w} = \text{input for the output layer=outh} \quad (3.9)$$

***Derivation of the formula used in multilayer perceptron***

***Phase 2:***

In phase-1, we find the updated weight for the output layer. In the second phase, we need to find the updated weights for the hidden layer. Hence, find how the change in hidden weight affects the change in error value.

Represented as:

$$\frac{\partial Error}{\partial wh} = \frac{\partial Error}{\partial outh} * \frac{\partial outh}{\partial inh} * \frac{\partial inh}{\partial wh} \quad (3.10)$$

**i.** ***Finding the first derivative*:**

Here we are going to use the chain rule to find the derivative.

$$\frac{\partial Error}{\partial outh} = \frac{\partial Error}{\partial ino} * \frac{\partial ino}{\partial outh} \quad (3.11)$$

Using the chain rule again.

$$\frac{\partial Error}{\partial ino} = \frac{\partial Error}{\partial outo} * \frac{\partial outo}{\partial ino} \quad (3.12)$$

The step below is similar to what we did in the first part of single layer perceptron

$$\frac{\partial ino}{\partial outh}\text{=output weight=wo} \tag{3.13}$$

ii. ***Finding the second derivative:***

$$\frac{\partial outh}{\partial inh}\text{=outh(1-outh)} \tag{3.14}$$

iii. ***Finding the third derivative:***

$$\frac{\partial inh}{\partial wh} = \text{input value} \tag{3.15}$$

***The back-propagation training algorithm***

**Step 1: Initialization**

All the weights and threshold levels of the network are random numbers uniformly distributed inside a small range. The weight initialization is done on a neuron-by-neuron basis.

**Step 2: Activation**

Activate the back-propagation neural network by applying inputs $x_1(p)$, $x_2(p)$,…, $x_n(p)$ and desired outputs $y_{d,1}(p)$, $y_{d,2}(p)$,…, $y_{d,n}(p)$.

(a) Calculate the actual outputs of the neurons in the hidden layer:

$$y_j\text{(p)=sigmoid }[\textstyle\sum_{i=1}^{n} x_i(p) \cdot w_{ij}(p) - \theta_j] \tag{3.16}$$

Where n is the number of inputs of neuron j in the hidden layer, and sigmoid is the Sigmoid activation function.

(b) Calculate the actual outputs of the neurons in the output layer:

$$y_k(p) = sigmoid[\textstyle\sum_{j=1}^{m} x_{jk}(p) \cdot w_{jk}(p) = \theta_k] \tag{3.17}$$

Where m is the number of inputs of neuron k in the output layer.

**Step 3: Weight training**

Update the weights in the back-propagation network propagating backward the errors associated with output neurons.

(a) Calculate the error gradient for the neurons in the output layer:

$$\delta_k(p) = y_k(p) \cdot [1 - y_k(p)] \cdot e_k(p) \qquad (3.18)$$

Where $e_k(p) = y_{d,k}(p) - y_k(p)$

Calculate the weight corrections:

$$\Delta w_{jk}(p) = \alpha \cdot y_j(p) \cdot \delta_k(p) \qquad (3.19)$$

Update the weights at the output neurons:

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \qquad (3.20)$$

(b) Calculate the error gradient for the neurons in the hidden layer:

$$\delta_j(p) = y_j(p) \cdot [1 - y_j(p)] \cdot \sum_{k=1}^{l} \delta_k(p) w_{jk}(p) \qquad (3.21)$$

Calculate the weight corrections:

$$\Delta w_{ij}(P) = \alpha \cdot x_i(P) \cdot \delta_j(P) \qquad (3.22)$$

Update the weights at the hidden neurons:

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p) \qquad (3.23)$$

**Step 4: Iteration**

Increase iteration p by one till 10,000 and go back to Step 2 and repeat the process until the selected error criterion is below 0.001.

### 3.4.2 K-Nearest Neighbor Model

- K-Nearest Neighbor (KNN) is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

*(a) Working Principle*

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:

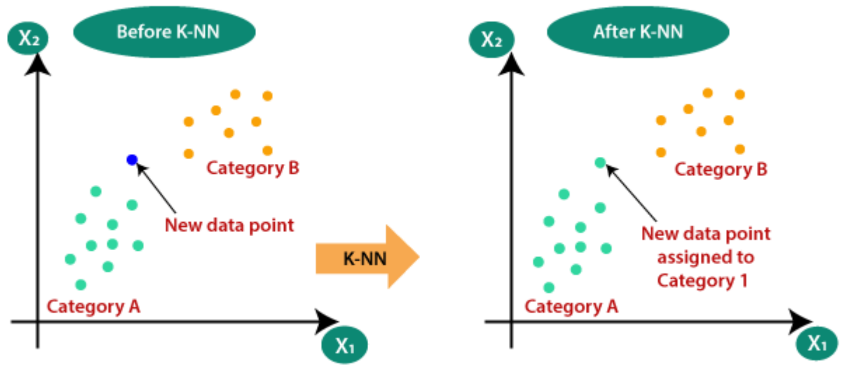



**Figure 3.8:** KNN with a class or category

Suppose we have a new data point and we need to put it in the required category. Consider the below image:

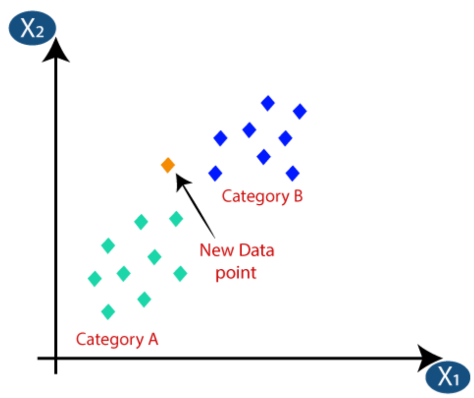



**Figure 3.9**: KNN with a new data point

Firstly, we will choose the number of neighbors, so we will choose the k=5.Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:

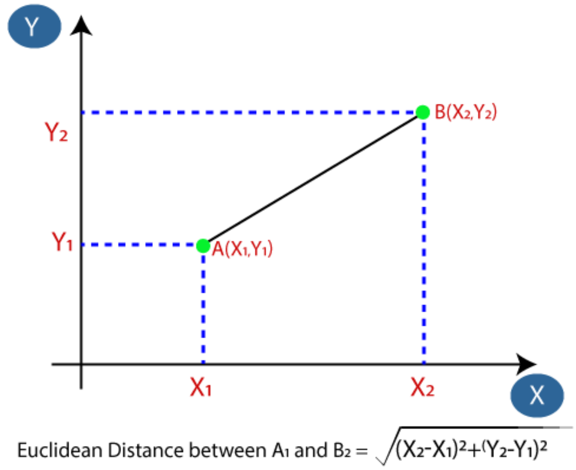



**Figure 3.10:** Eucledian distance in KNN algorithm

By calculating the Euclidean distance we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the below image:

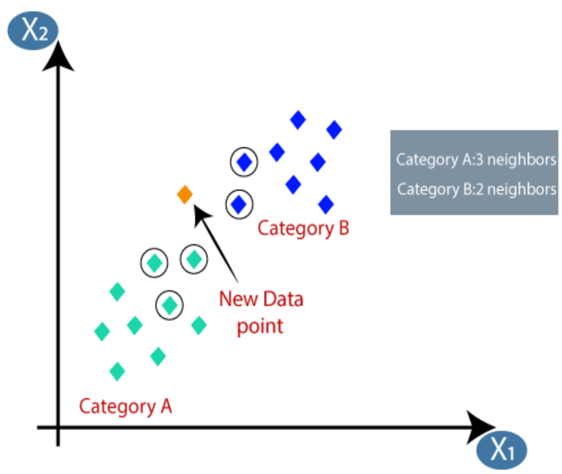



**Figure 3.11:** A new data point close to a class

As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

*(b) Selecting K value*

- There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.
- A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
- Large values for K are good, but it may find some difficulties.

*(c) Algorithm*

The K-NN working can be explained on the basis of the below algorithm:

- Step-1: Select the number K of the neighbors
- Step-2: Calculate the Euclidean distance of K number of neighbors
- Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
- Step-4: Among these k neighbors, count the number of the data points in each category.
- Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
- Step-6: Our model is ready.

### 3.4.3 Support Vector Machine Model

SVM is a famous supervised machine learning algorithm used for classification as well as regression algorithms. However, mostly it is preferred for classification algorithms. It basically separates different target classes in a hyperplane in n-dimensional or multidimensional space.

The main motive of the SVM is to create the best decision boundary that can separate two or more classes(with maximum margin) so that we can correctly put new data points in the correct class.

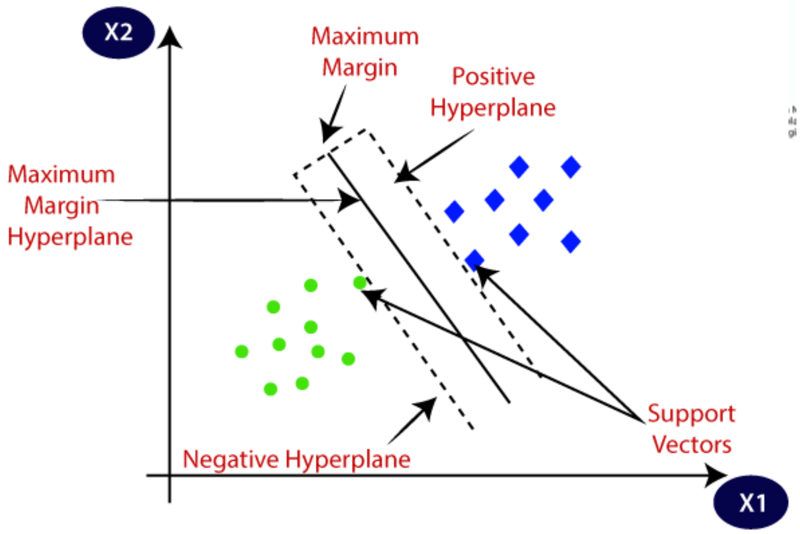



**Fig 3.12:** Support vector machine

## Algorithm

Although the SVM can be applied to various optimization problems such as regression, the classic problem is that of data classification. The basic idea is shown in figure . The data points are identified as being positive or negative, and the problem is to find a hyperplane that separates the data points by a maximal margin. The above figure shows the 2-dimensional case where the data points are linearly separable. The mathematics of the problem to be solved is the following:

$$\min_{w,b} \frac{1}{2}\|w\|,$$

$$\text{s.t} \quad y_i = +1 \rightarrow \vec{w} \cdot \vec{x_i} + b \geq +1$$

$$y_i = -1 \rightarrow \vec{w} \cdot \vec{x_i} - b \leq -1 \qquad \text{(i)}$$

$$\text{s.t} \quad y_i(\vec{w} \cdot \vec{x_i} + \text{b}) \geq 1, \forall_i$$

The identification of the each data point xi is yi, which can take a value of +1 or (representing positive or negative respectively). The solution hyper-plane is the following:

u $=\vec{w} \cdot \vec{x} + b$(2)

The scalar b is also termed the bias. A standard method to solve this problem is to apply the theory of Lagrange to convert it to a dual Lagrangian problem. The dual problem is the following:

$$\min_{\alpha} ¥(\vec{\alpha}) = \min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \left(\vec{x_i} \cdot \vec{x_j}\right) \alpha_i \alpha_j - \sum_{i=1}^{N} \alpha_i \qquad \text{(ii)}$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \qquad \forall_i$$

The variables αi are the Lagrangian multipliers for corresponding data point xi.

## 3.5 PERFORMANCE ANALYSIS

Performance analysis involves systematic observations to enhance performance that improves decision making of data and gives feedback to our system. It is the process of studying or evaluating the performance of a particular scenario in comparison of the objective which to be achieved. For analyses we use confusion matrix, accuracy, precision, recall and f1 score

### *(a) Confusion Matrix*

A Confusion matrix is an N × N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making. For a binary classification problem, we would have a 2×2 matrix as shown below with 4 values:

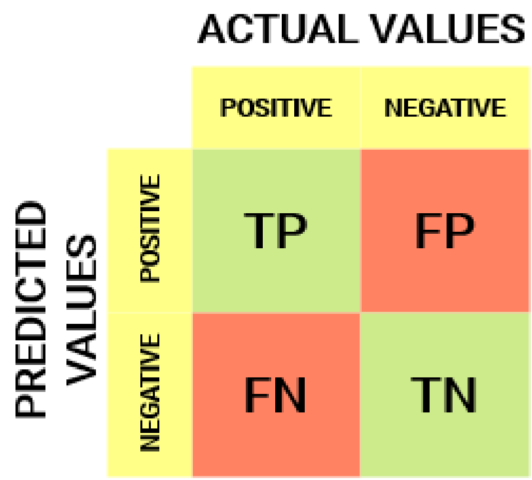



**Figure 3.13:** Confusion Matrix

The target variable has two values: Positive or Negative The columns represent the actual values of the target variable The rows represent the predicted values of the target variable

*True Positive (TP) :*The predicted value matches the actual valueThe actual value was positive and the model predicted a positive value.

*True Negative (TN) :*The predicted value matches the actual value. That means the actual value was negative and the model predicted a negative value

*False Positive (FP) – Type 1 error:* The predicted value was falsely predicted. That means the actual value was negative but the model predicted a positive value. Also known as the Type 1 error.

*False Negative (FN) – Type 2 error:* The predicted value was falsely predicted**.** That meansthe actual value was positive but the model predicted a negative value**.** Also known as the Type 2 error.

*(b) Correlation Matrix*

The Correlation matrix is an important data analysis metric that is computed to summarize data to understand the relationship between various variables and make decisions accordingly.It is also an important pre-processing step in Machine Learning pipelines to calculate and analyze the correlation matrix where dimensionality reduction is desired on a high-dimension data.

We mentioned how each cell in the correlation matrix is a ‘correlation coefficient ‘between the two variables corresponding to the row and column of the cell. Each row and column represents a variable, and each value in this matrix is the correlation coefficient between the variables represented by the corresponding row and column.

*(c) Accuracy*

In confusion matrix the accuracy is the ratio of correct prediction to the total predictions made. It is often presented as a percentage by multiplying the result by 100. The numerical value of accuracy represents the proportion of true positive results (both true positive and true negative) in the selected population. The formula shown as below

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+TN+FN}} \quad \text{(i)}$$

*(d) Precision*

Precision tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. Precision is a useful metric in cases where False Positive is a higher concern than False Negatives.

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad \textbf{(ii)}$$

*(e) Recall*

Recall tells us how many of the actual positive cases we were able to predict correctly with our model. Recall is a useful metric in cases where False Negative trumps False Positive.

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{(iii)}$$

*(f) F1-Score*

In practice, when we try to increase the precision of our model, the recall goes down, and vice-versa. The F1-score captures both the trends in a single value:F1-score is a harmonic mean of Precision and Recall, and so it gives a combined idea about these two metrics. It is maximum when Precision is equal to Recall. But there is a catch here. The interpretability of the F1-score is poor.

$$\text{F1 Score} = \frac{2}{\frac{1}{Recall}+\frac{1}{Precision}} \qquad \text{(iv)}$$

# CHAPTER 4

# RESULT AND DISCUSSION

## 4.1 INTRODUCTION

In this section, we used different regression and classification algorithms for calculation of performance evaluation metrics. Results are produced from all of the used algorithms. This section presents the numerical analysis performed with neural network, k nearest neighbor and support vector machine. Compare the results obtained by various performance metrics and finally conclude the obtained results by comparison. We evaluate data by bar chart, correlation matrix, confusion matrix etc.Training accuracy and validation accuracy, training loss and validation loss graph is also depicted from the neural network figure to show if the model is overfitted or underfitted. After performing various calculation the final result is obtained by the algorithms output. In section 4.2 the data assessment based on bar chart and correlation chart is showed. In the next section 4.3 performance analysis which include confusion metrics, accuracy and validation graph, training and validation graph for ANN , error rate and k value, k neighbor classifier score for KNN and confusion metrics for SVM is explained with required graph. And in the section 4.4 evaluation for distinct methods are showed at tabular form. We apply these evaluation for school, college and university level students. The following section described the results briefly.

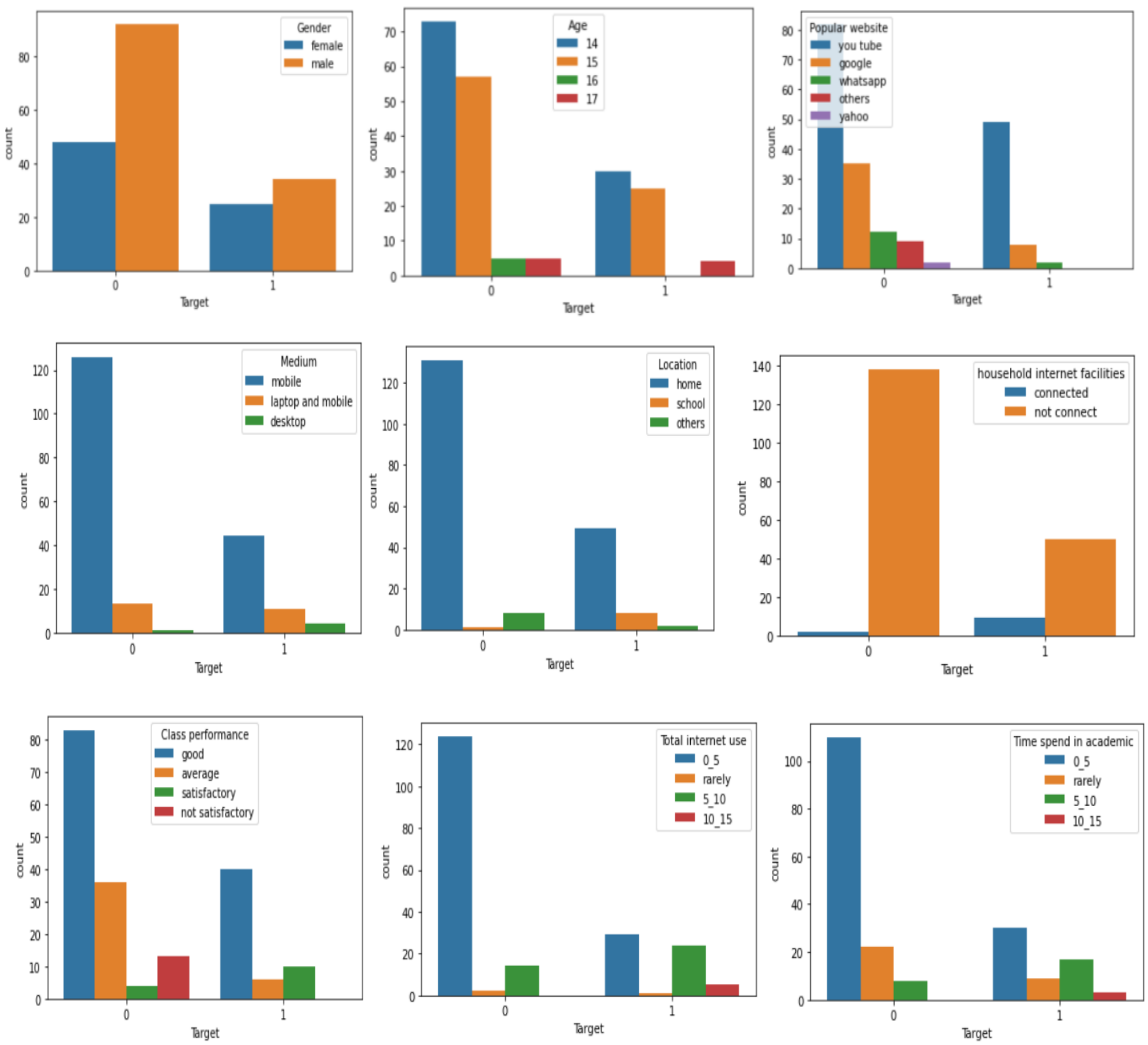
## 4.2 DATA ASSESSMENT

Assessment of data is done by measuring particular features of the data to see if they meet defined standards. It is the process of statistically evaluating data in order to determine whether they meet the required quality. To visualize data as plots and charts to learn more about it, we can use pandas with Matplotlib. We will discuss two kinds of plots- univariate and multivariate. A univariate plot suggests we’re only examining one variable. On the other hand,a multivariate analysis examines more than two variables. For two variables, we call it bivariate.

### 4.2.1 Data Assessment for School

*(a) Based on Bar Chart*

Since bar chart group data into bins and give us an idea of how many observations each bin holds, this is a good way to visualize data for ML. The shapes of the bins tell us whether an attribute is Gaussian, skewed, or has an exponential distribution. It also hints us about outliers.



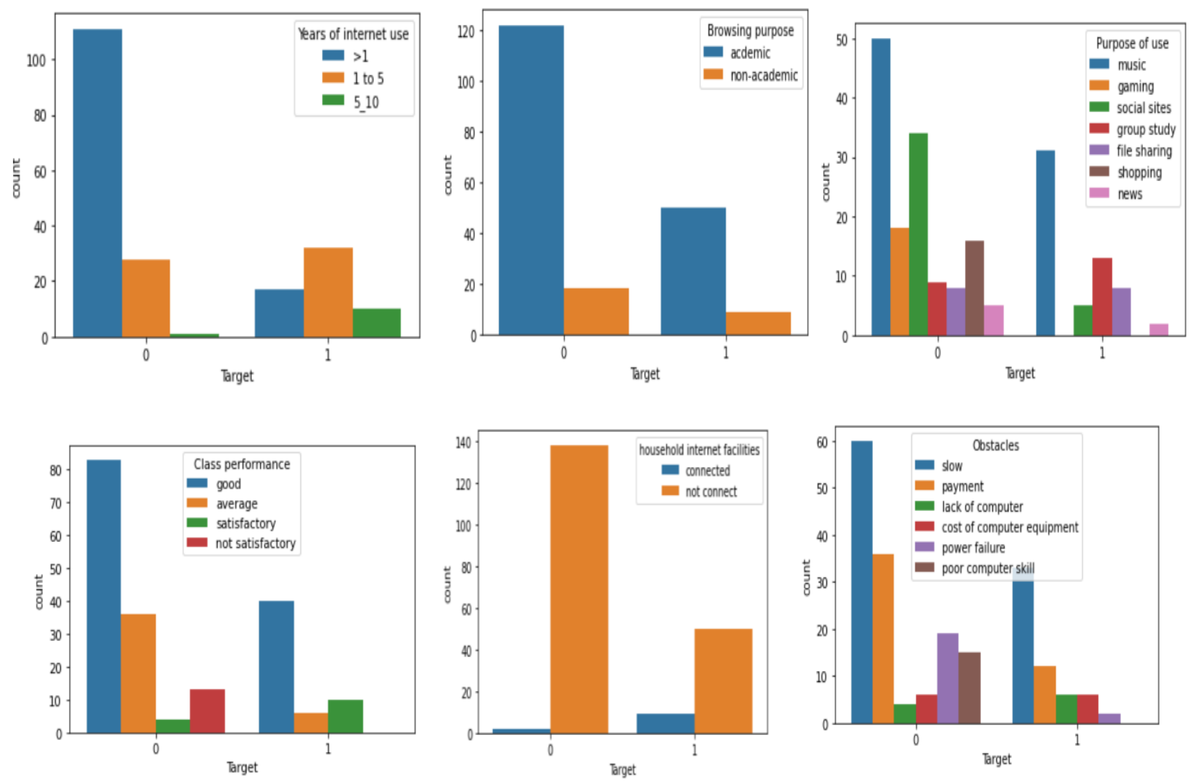



**Fig 4.1:** Visualization of school data in terms of target variable

*(b) Based on Correlation Graph*

A correlation matrix is a table showing correlation coefficients between variables, each cell in the table shows the correlation between two variables. Correlation plot denotes how changes between two variables relate. Two variables that change in the same direction are positively correlated. A change in opposite directions implies negative correlation. Herewe plot a correlation matrix for school data.The positive signed values are positively correlated and negative signed values are negatively correlated to the target variable. From this figure we can determine the variables which have a good impact on final outcome and thus analyses will also improved.The correlation matrix plot shows that the variable popular website,skillness,browsing time, browsing status,years of internet use and online course has positive effect on the performance of sudents that is the target variable.

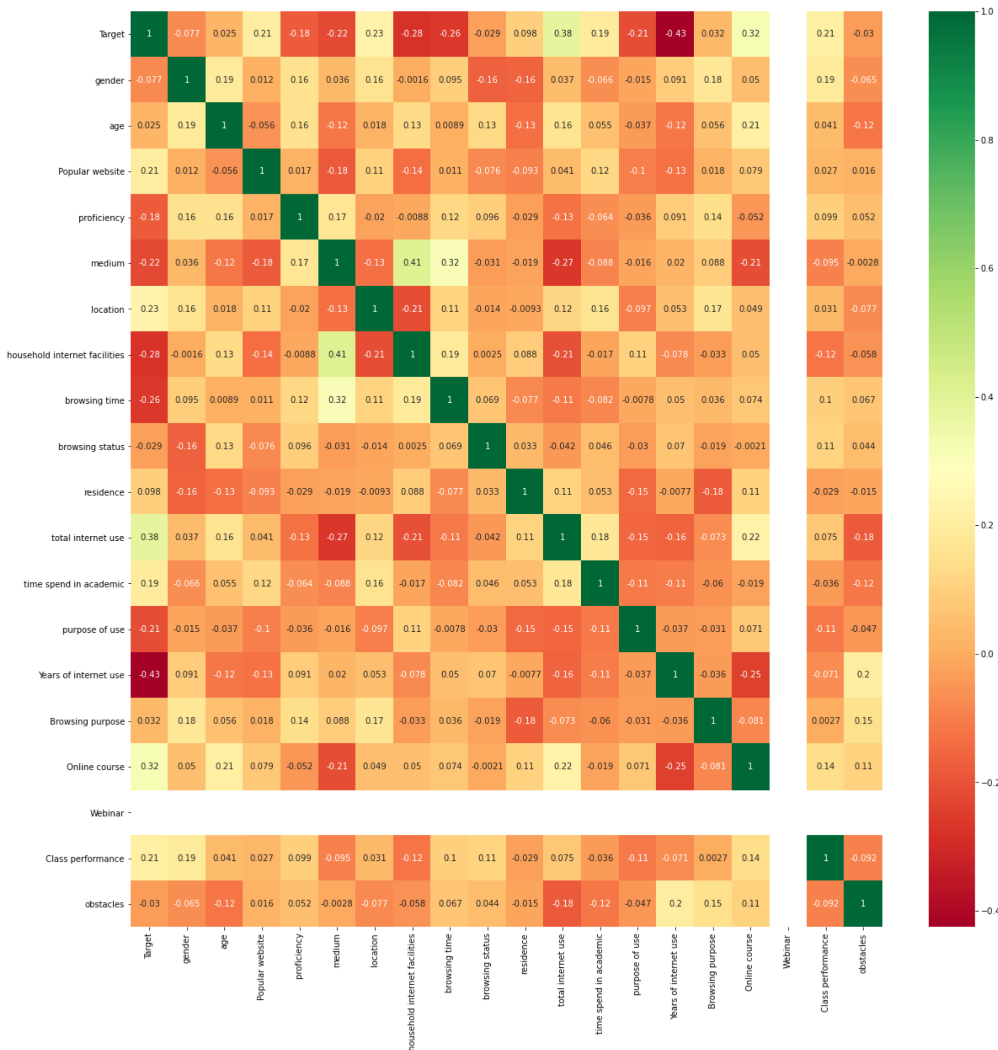
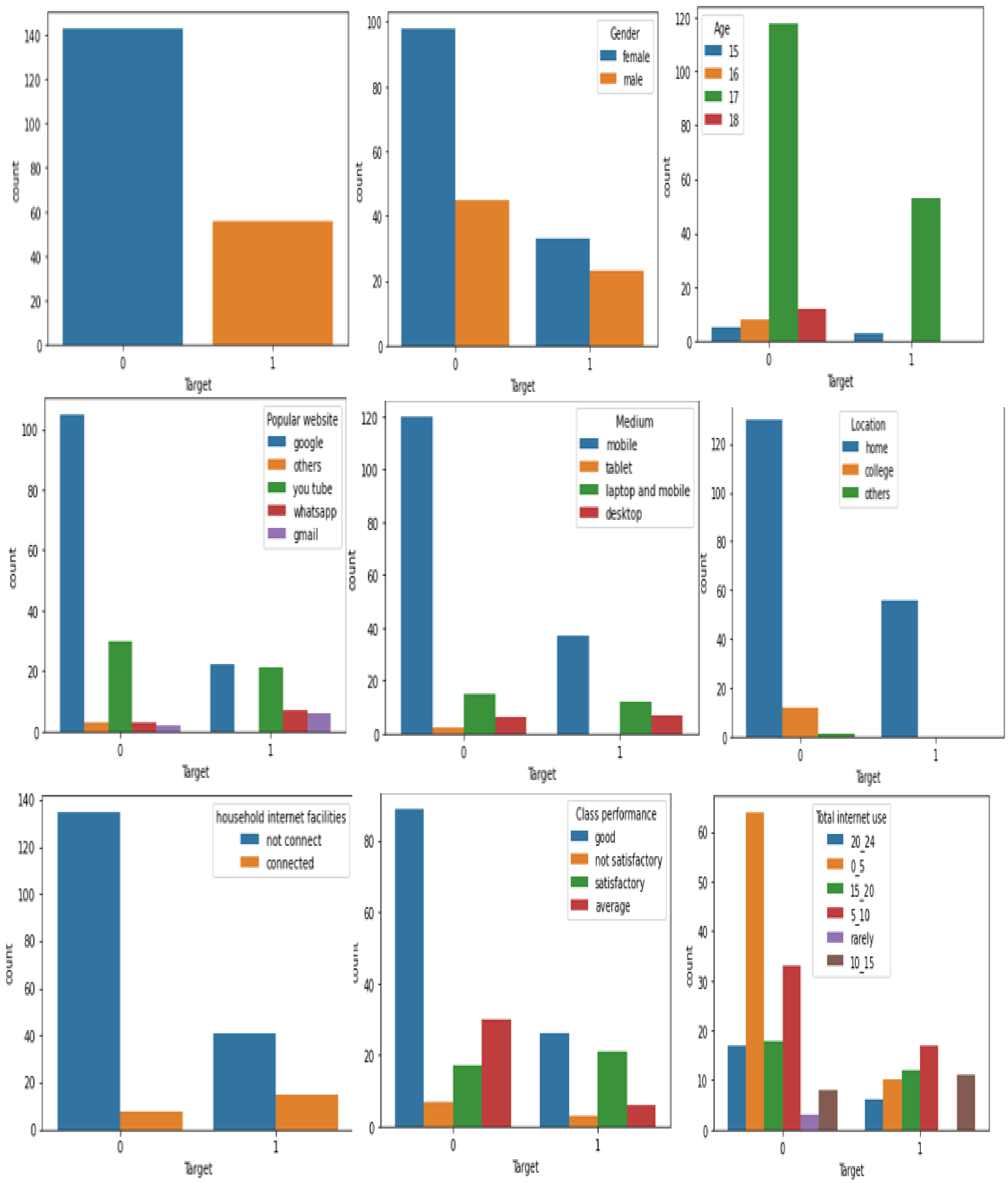



**Fig 4.2:** Correlation Matrix plot for school data

### 4.2.2 Data Assessment for College

*(a) Based on Bar Chart*

Since bar chart group data into bins and give us an idea of how many observations each bin holds, this is a good way to visualize data for ML. The shapes of the bins tell us whether an attribute is Gaussian, skewed, or has an exponential distribution. It also hints us about outliers.


count
Target
Gender
female
male
Age
15
16
17
18
Popular website
google
others
you tube
whatsapp
gmail
Medium
mobile
tablet
laptop and mobile
desktop
Location
home
college
others
household internet facilities
not connect
connected
Class performance
good
not satisfactory
satisfactory
average
Total internet use
20_24
0_5
15_20
5_10
rarely
10_15

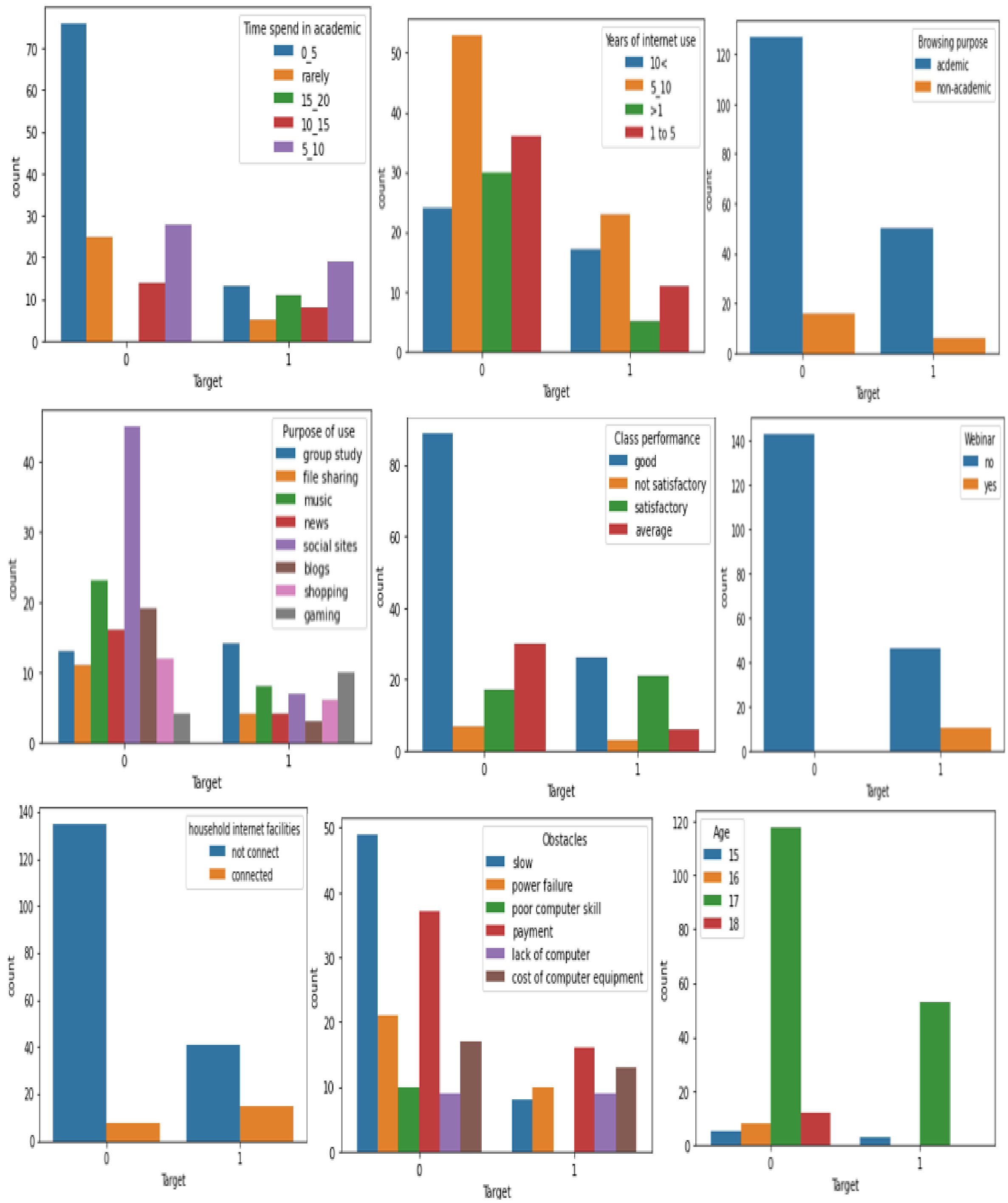



**Fig 4.3:** Visualization of college data in terms of target variable

*(b) Based on Correlation Graph*

A correlation matrix is a table showing correlation coefficients between variables, each cell in the table shows the correlation between two variables. Correlation plot denotes how changes between two variables relate. Two variables that change in the same direction are positively correlated. A change in opposite directions implies negative correlation. Here we plot a correlation matrix for college data.The positive signed values are positively correlated and negative signed values are negatively correlated to the target variable. From this figure we can determine the variables which have a good impact on final outcome and thus analyses will also improved.The correlation matrix plot shows that the variable popular website,skillness,browsing time, browsing status,years of internet use and online course has positive effect on the performance of sudents that is the target variable.

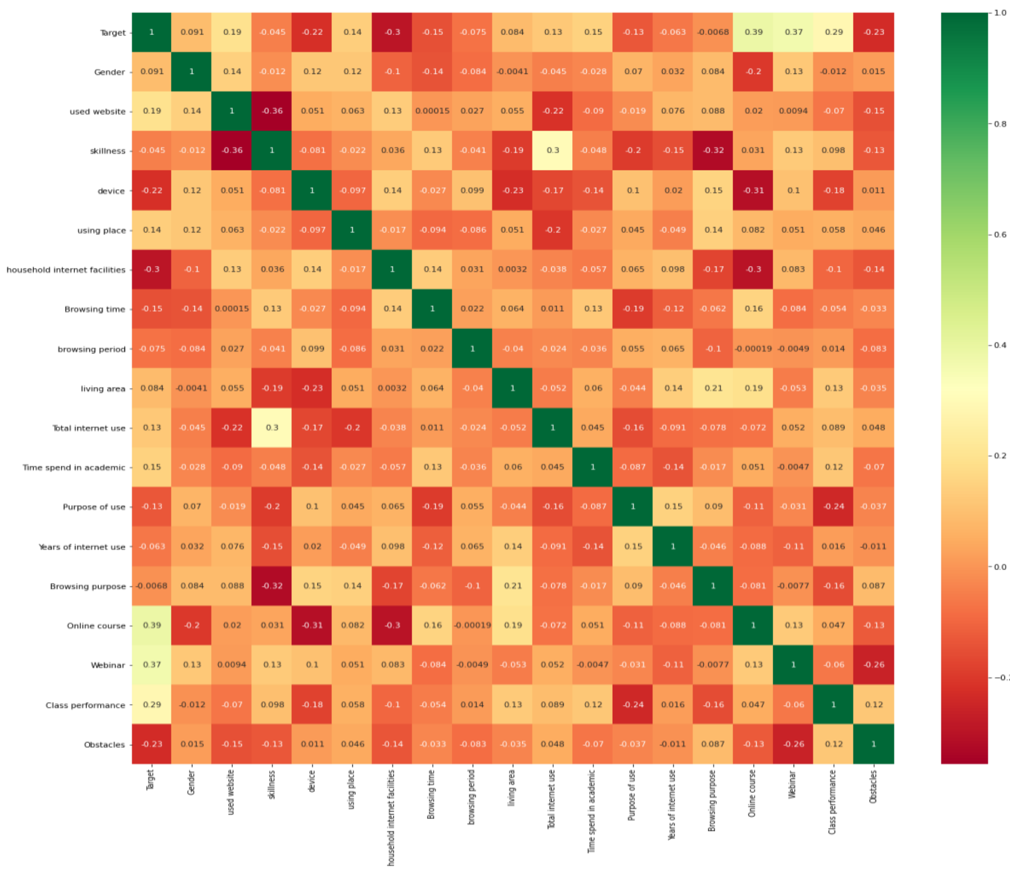
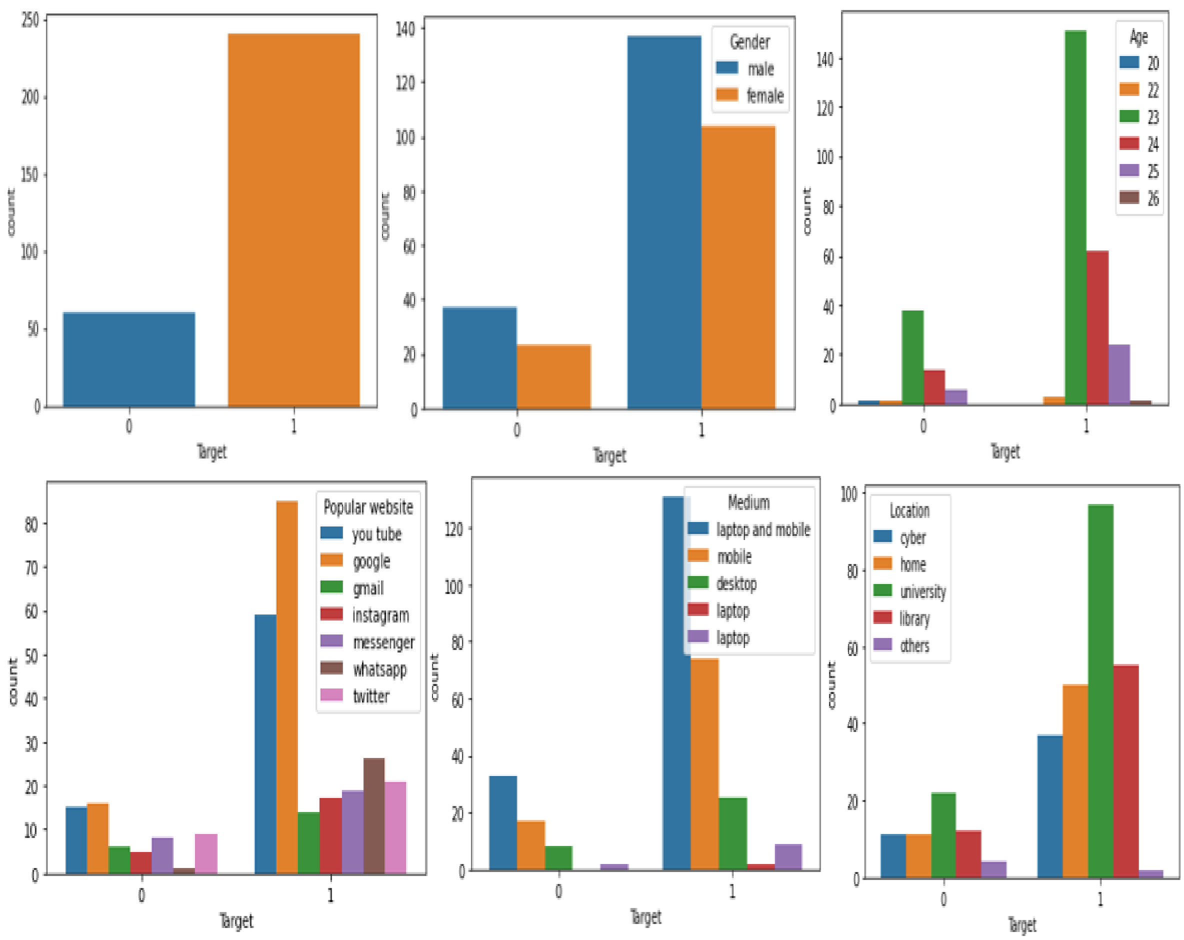


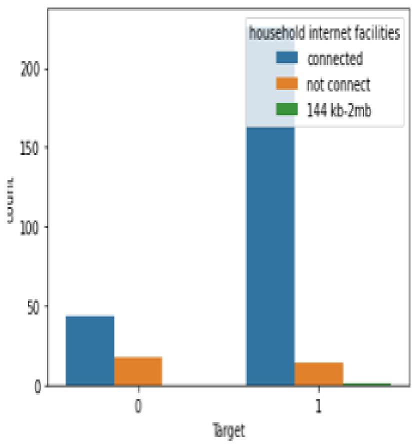

**Fig 4.4:** Correlation Matrix plot for college data

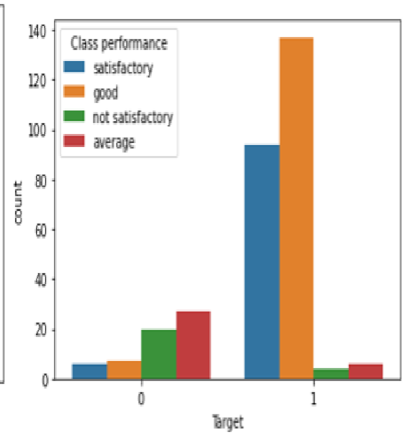
### 4.2.1. Data Assessment for university

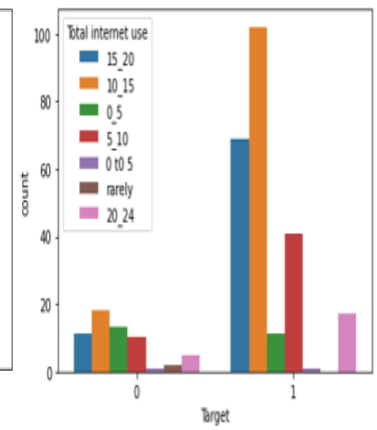
*(a) Based on Bar Chart*

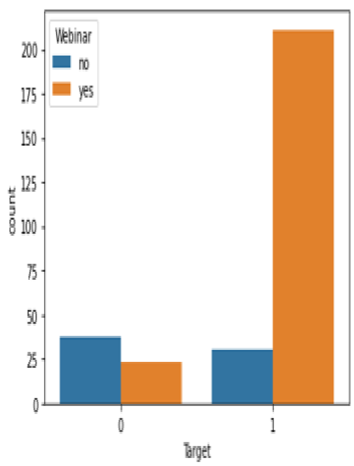
Since bar chart group data into bins and give us an idea of how many observations each bin holds, this is a good way to visualize data for ML. The shapes of the bins tell us whether an attribute is Gaussian, skewed, or has an exponential distribution. It also hints us about outliers.

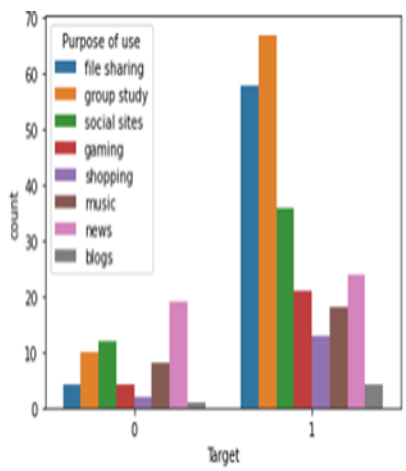


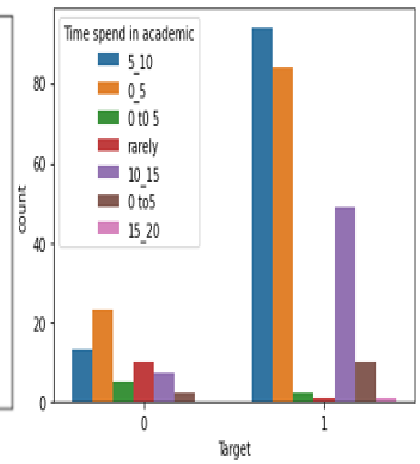

household internet facilities
connected
not connect
144 kb-2mb
Target

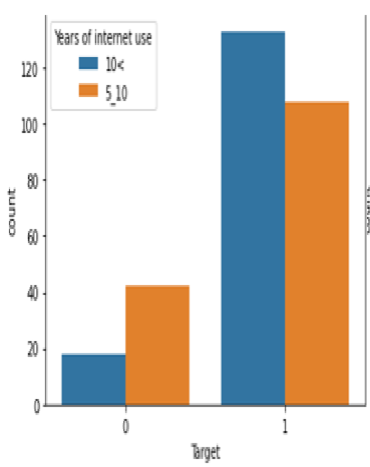

Class performance
satisfactory
good
not satisfactory
average
count
Target

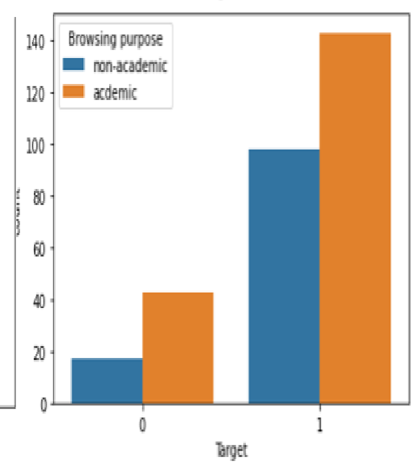

Total internet use
15_20
10_15
0_5
5_10
0 t0 5
rarely
20_24
count
Target

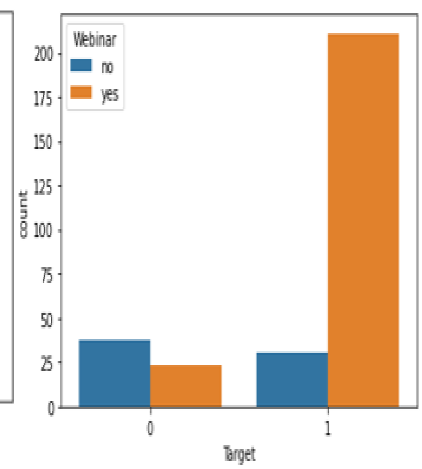

Webinar
no
yes
count
Target


Purpose of use
file sharing
group study
social sites
gaming
shopping
music
news
blogs
count
Target


Time spend in academic
5_10
0_5
0 t0 5
rarely
10_15
0 to5
15_20
count
Target


Years of internet use
10<
5_10
count
Target


Browsing purpose
non-academic
acdemic
Target


Webinar
no
yes
count
Target

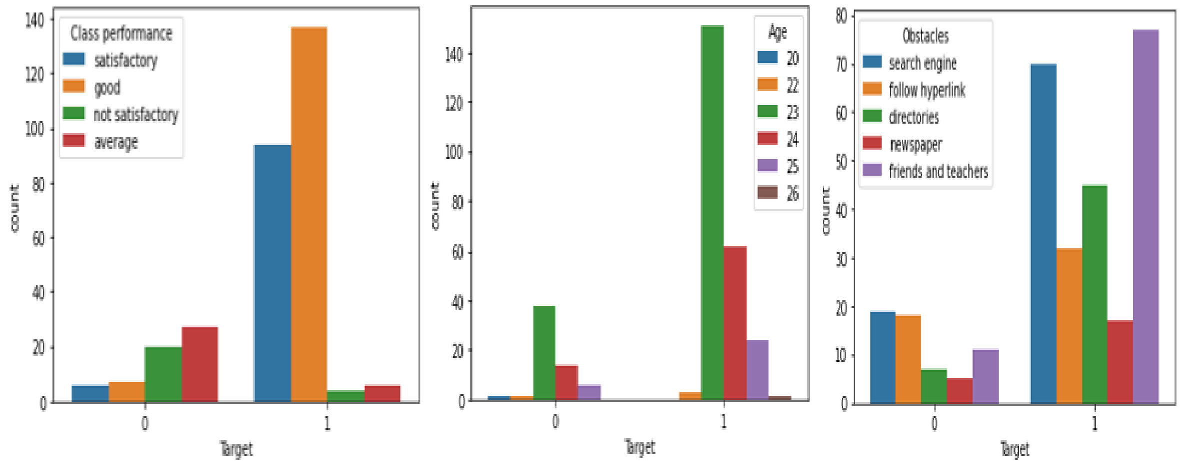



**Figure 4.4:** Visualization of university data in terms of target variable

*(b) Based on Correlation Graph*

A correlation matrix is a table showing correlation coefficients between variables, each cell in the table shows the correlation between two variables. Correlation plot denotes how changes between two variables relate. Two variables that change in the same direction are positively correlated. A change in opposite directions implies negative correlation. Here we plot a correlation matrix for university data.The positive signed values are positively correlated and negative signed values are negatively correlated to the target variable. From this figure we can determine the variables which have a good impact on final outcome and thus analyses will also improved.The correlation matrix plot shows that the variable popular website,skillness,browsing time, browsing status,years of internet use and online course has positive effect on the performance of sudents that is the target variable.

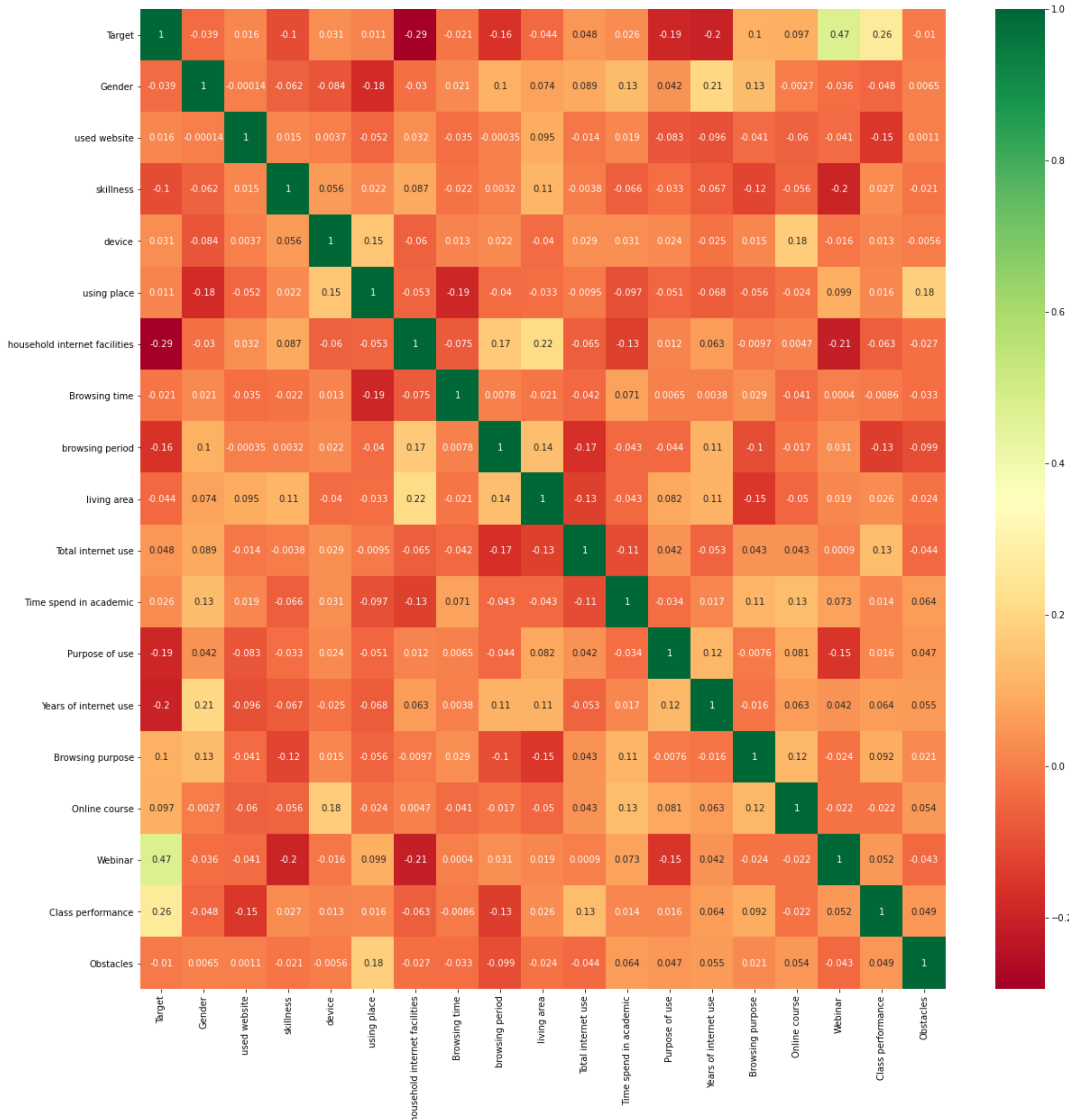



**Fig 4.5**: Correlation matrix plot of university data

## 4.3 PERFORMANCE ANALYSIS

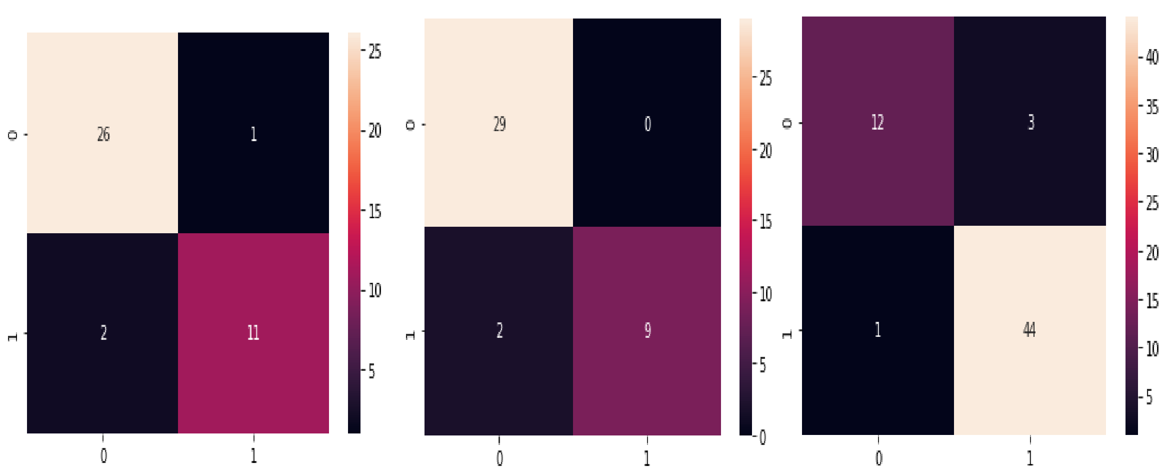
In this section performance analysis which include confusion metrics, accuracy and validation graph, training and validation graph for ANN, error rate and k value, k neighbor classifier score for KNN and confusion metrics for SVM is explained with their associated graph. The matrix compares the actual target values with those predicted by the machine learning model. Visualizing the training loss vs. validation loss or training accuracy vs.

validation accuracy over a number of epochs is a good way to determine if the model has been sufficiently trained. This is important so that the model is not undertrained and not overtrained. Training loss is measured during each epoch while validation loss is measured after each epoch. In k nearest neighbor, k is the quantity of how many nearest values we need to consider for prediction in terms of distance. The distance is calculated by Eucledian or Manhattan distance. Random selection of k value does not gives us better accuracy. So that we have to derive the k value. For derivation we use a simple technique by which we can determine for which k value the error rate is reduced or increased.

### 4.3.1 Artificial Neural Network

(a) *Confusion Matrix*

A Confusion matrix is an N × N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making. For a binary classification problem, we would have a 2×2 matrix as shown below with 4 values:



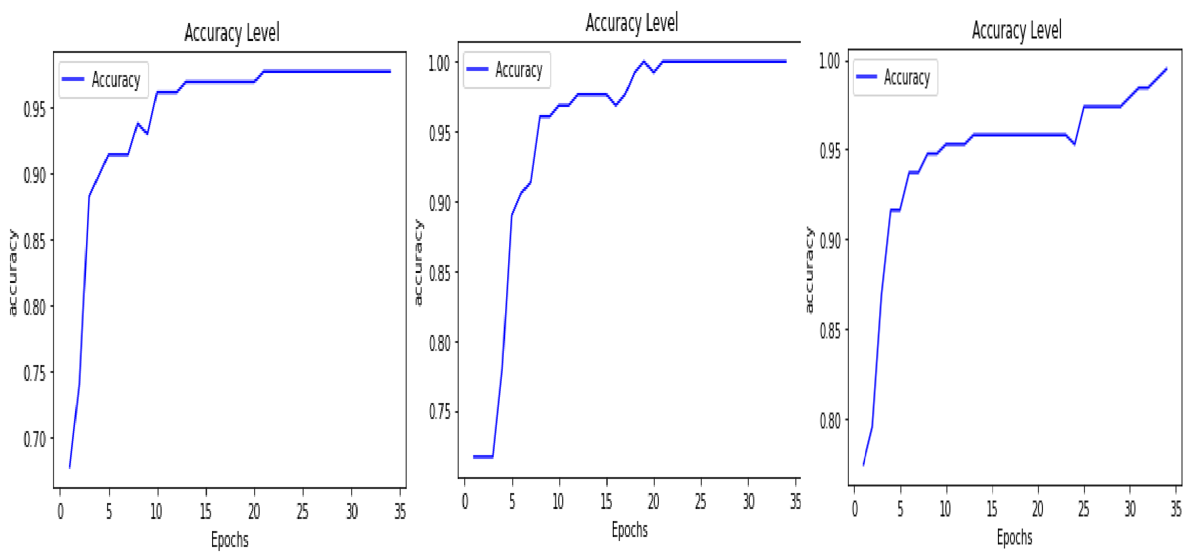

(i) School (ii) College (iii) University

**Fig 4.6:** Confusion Matrix for Neural Network algorithm

In the first figure for the true positive(TP) value is 15 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 4 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 1 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value. In the second figure for the true positive(TP) value is 17 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 3 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 0 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value. In the third figure for the true positive(TP) value is 1 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 26 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 3 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value.

### *(b) Accuracy and Validation graph*

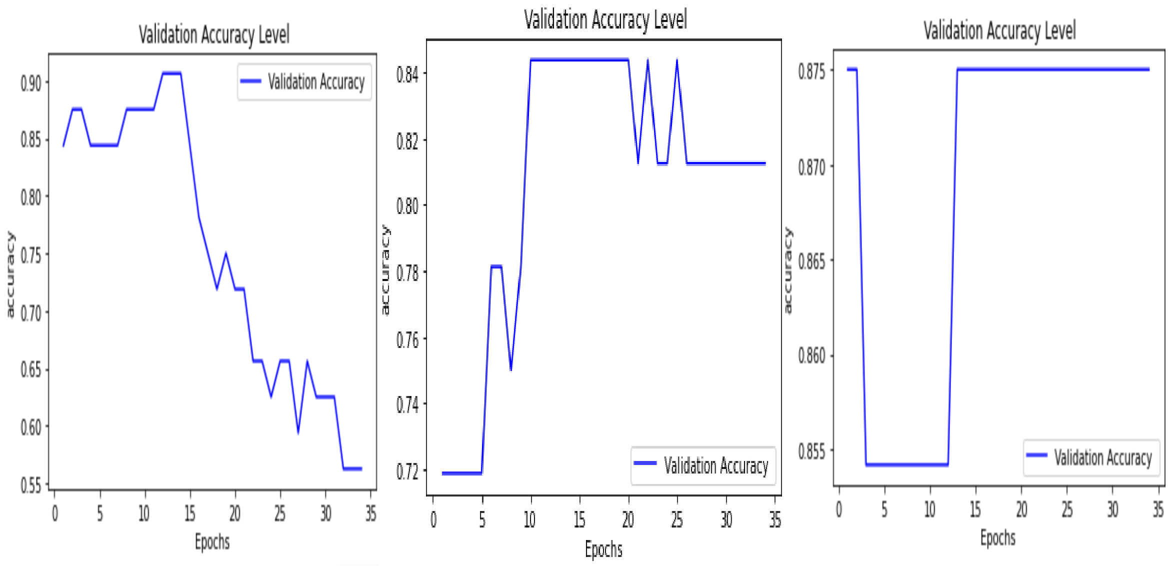
One of the primary difficulties in any machine learning approach is to make the model generalized so that it is good in predicting reasonable results with the new data and not just on the data it has already been trained on. Visualizing the training loss vs. validation loss or training accuracy vs. validation accuracy over a number of epochs is a good way to determine if the model has been sufficiently trained. This is important so that the model is not undertrained and not overtrained such that it starts memorizing the training data which will, in turn, reduce its ability to predict accurately.




**(i) School** **(ii) College** **(iii) University**

**Fig 4.7:** Accuracy graph for Neural Network algorithm

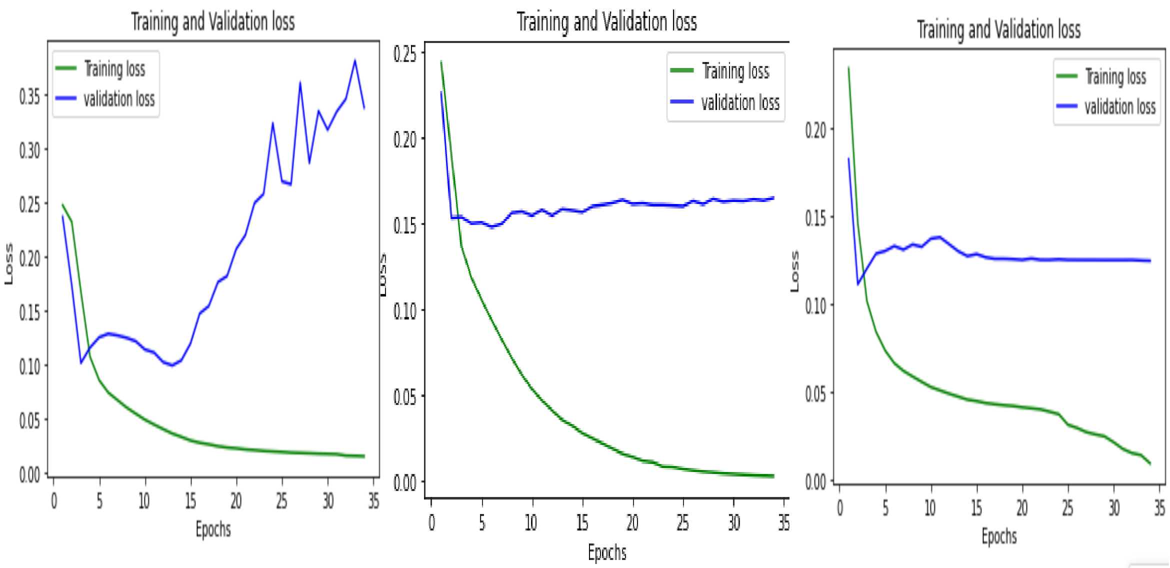
In the above figure the highest levelof accuracy for school data collection is 0.9950 e.g 99%. The accuracy level for college data is .1000 e.g 100%. And the accuracy level for university is .9971 e.g 99%.Here we count the epoch for 1 to 35 and get various accuracy level for different data. The final accuracy we got from school data is 91%, college data is 100% and university data is 98%




**(i) School** **(ii) College** **(iii) University**

**Figure 4.8**: Validation accuracy graph for Neural Network algorithm

Now from the validation accuracy graph we see the accuracy level for validation data for school data and the highest accuracy we got from epochs is .9062 e.g 90%, college is .8525 e.g 85% , and university is .8750 e.g 87%. The count is done for 35 epochs.



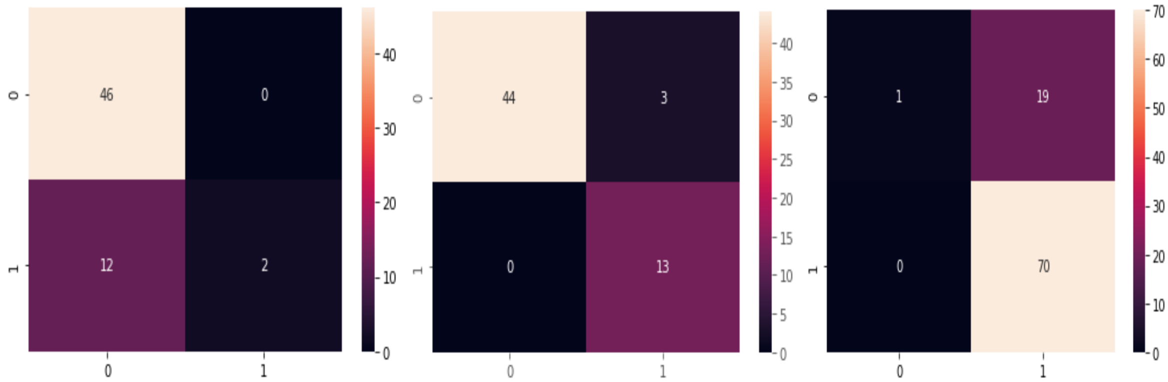

**(i) School** **(ii) College** **(iii) University**

**Fig 4.9:** Training and validation loss graph for Neural Network algorithm

At the most basic level, a loss function quantifies how "good" or "bad" a given predictor is at classifying the input data points in a dataset. The smaller the loss, the better a job the classifier is at modeling the relationship between the input data and the output targets. By modeling the training data too closely, our model loses the ability to generalize. Training loss is measured during each epoch while validation loss is measured after each epoch. Validation metrics are computed over the validation set only once the current training epoch is completed. Here we see the training loss for school after 34 epochs is 0.0431 and validation loss is .3380. Next for college the training loss is 7.6417e and validation loss is .1669 and again for university the training loss after 34 epochs is .0177 and validation loss is .1244

### 4.3.2 K Nearest Neighbor

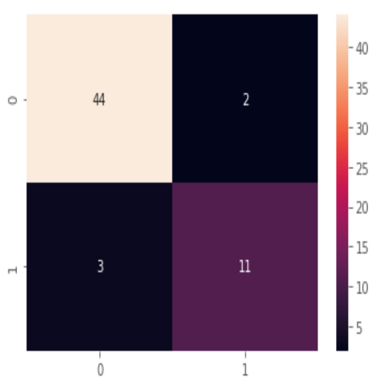
*(a) Confusion Matrix*



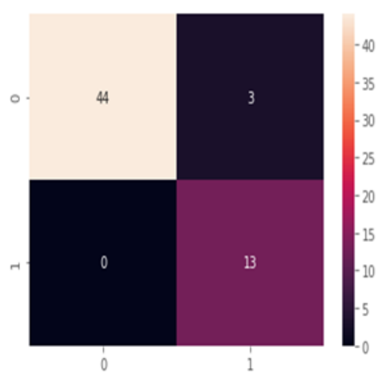

**(i) School** **(ii) College** **(iii) University**

**Fig 4.10:** Confusion Matrix of KNN algorithm for K=20

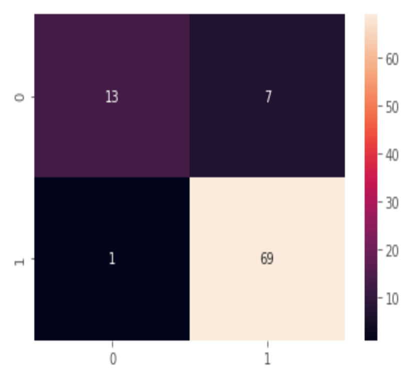
In the first figure for k=20 the true positive(TP) value is 46 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 2 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 0 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 12 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value. In the second figure for k=20 the true positive(TP) value is 44 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 13 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 3 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value. In the third figure for k=20 the true positive(TP) value is 1 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 70 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 19 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value.




**(i) School**




**(ii) College**




**(iii) University**

**Fig 4.11:** Confusion Matrix of KNN algorithm for K=1

In the first figure for k=1 the true positive(TP) value is 44 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 11 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 2 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 3 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value. And in the second figure for k=1 the true positive(TP) value is 44 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 13 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 3 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value. And in the third figure for k=1 the true positive(TP) value is 13 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 69 which means The predicted value matches the actual value. The actual value was negative and the model predicted a negative value.The false positive(FP) value is 7 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 1 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value *(b) Error rate and k value*

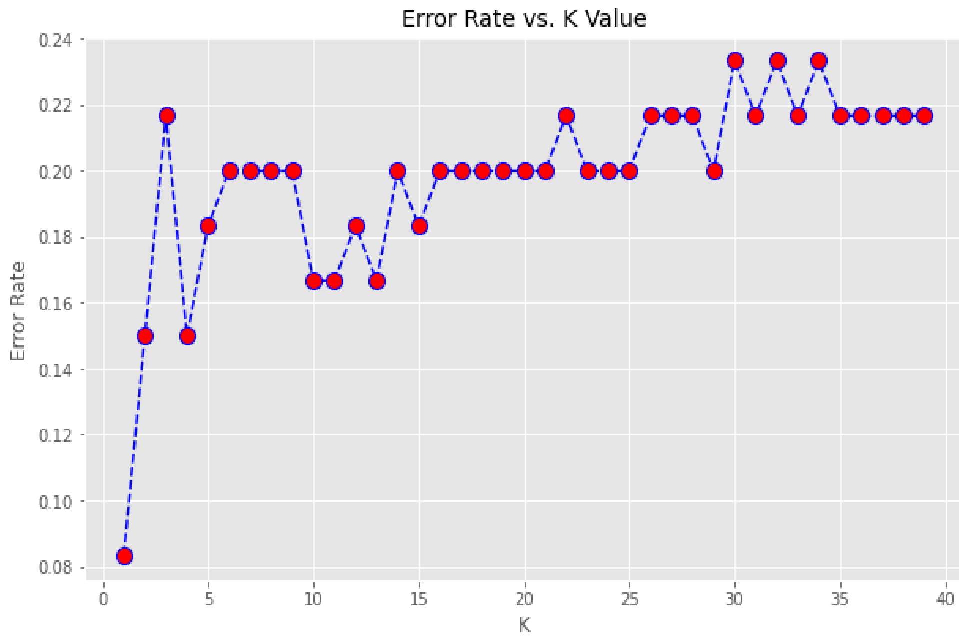



**Fig 4.12:** Error Rate vs. K Value for school

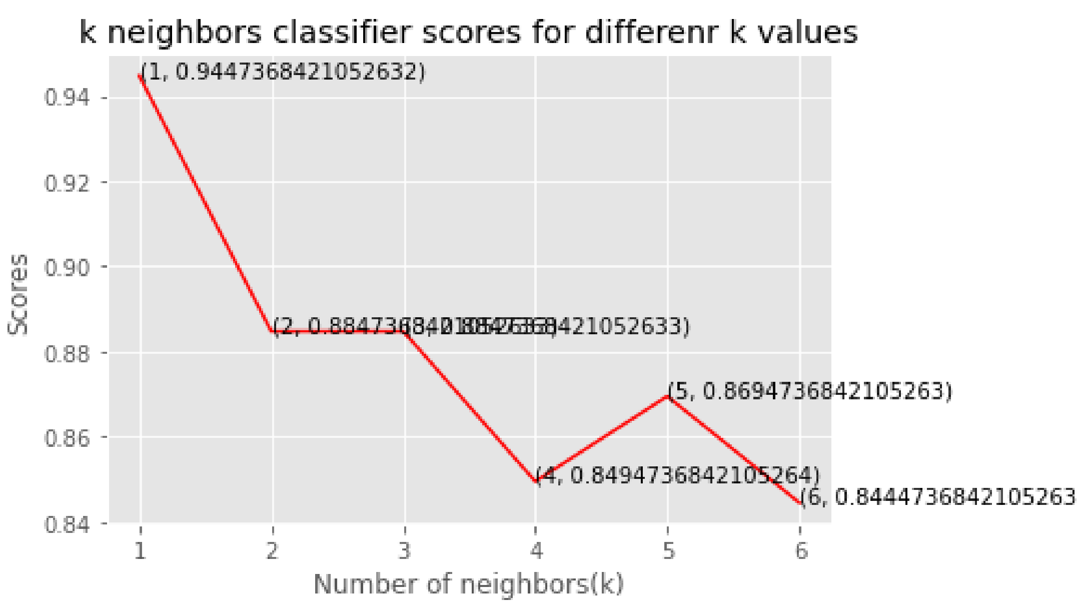



**Fig 4.13**: k neighbors classifier scores at different k values for school

In k nearest neighbor, k is the quantity of how many nearest values we need to consider for prediction in terms of distance. The distance is calculated by Eucledian or Manhattan distance.For example if we consider k=5,then 5 nearest neighbors are selected by eucledian distance. Random selection of k value does not gives us better accuracy. So that we have to derive the k value. For derivation we use a simple technique by doing a number of iteration. For choosing k value we run a loop from 1 to 40, these values are going to consider as k value. Here the error rate for different k values are obtained then we try to plot the figure and see where the error rate is quite stable. Now we can see from the first figure that for k=1 it obtain stability, that is after 1 the error line is going higher, for k=1 error rate in in between 0.08 to 0.10. For further understanding we can see the second figure that the score for k=1 is higher than the others. Though it gives us good accuracy but it is the problem of underfitting for considering 1 nearest neighbor.

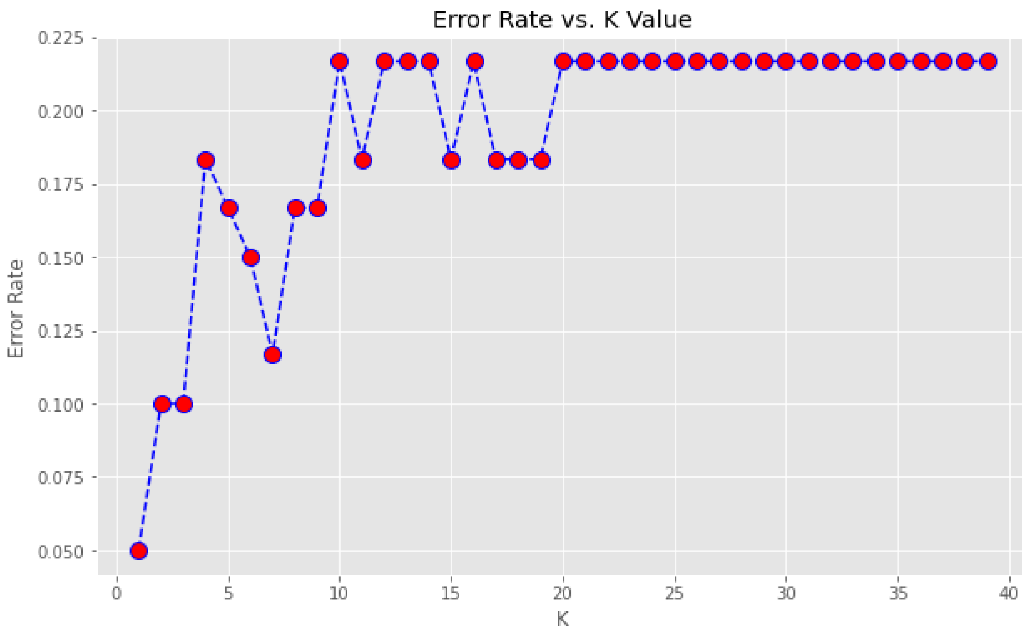



**Fig 4.14:** Error Rate vs. K Value for college

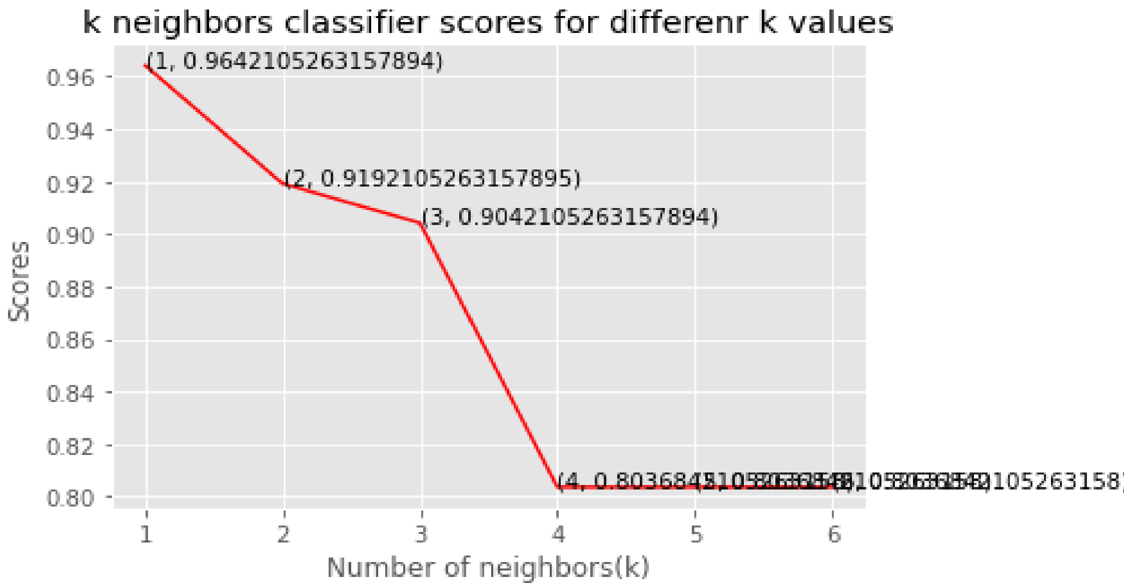



**Fig 4.15:** k neighbors classifier scores at different k values for college

In k nearest neighbor, k is the quantity of how many nearest values we need to consider for prediction in terms of distance. The distance is calculated by Eucledian or Manhattan distance.For example if we consider k=5,then 5 nearest neighbors are selected by eucledian distance. Random selection of k value does not gives us better accuracy. So that we have to derive the k value. For derivation we use a simple technique by doing a number of iteration. For choosing k value we run a loop from 1 to 40, these values are going to consider as k value. Here the error rate for different k values are obtained then we try to plot the figure and see where the error rate is quite stable. Now we can see from the first figure that for k=1 it obtain stability, that is after 1 the error line is going higher, for k=1 error rate in in between 0.050 to 0.075. For further understanding we can see the second figure that the score for k=1 is higher than the others. Though it gives us good accuracy but it is the problem of underfitting for considering 1 nearest neighbor.

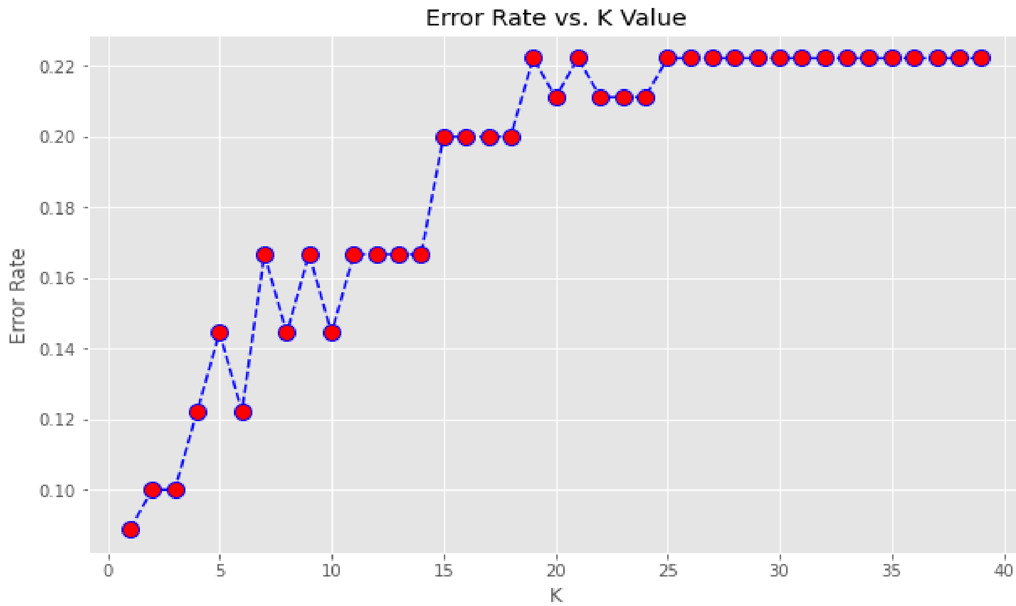



**Fig 4.16:** Error Rate vs. K Value for university

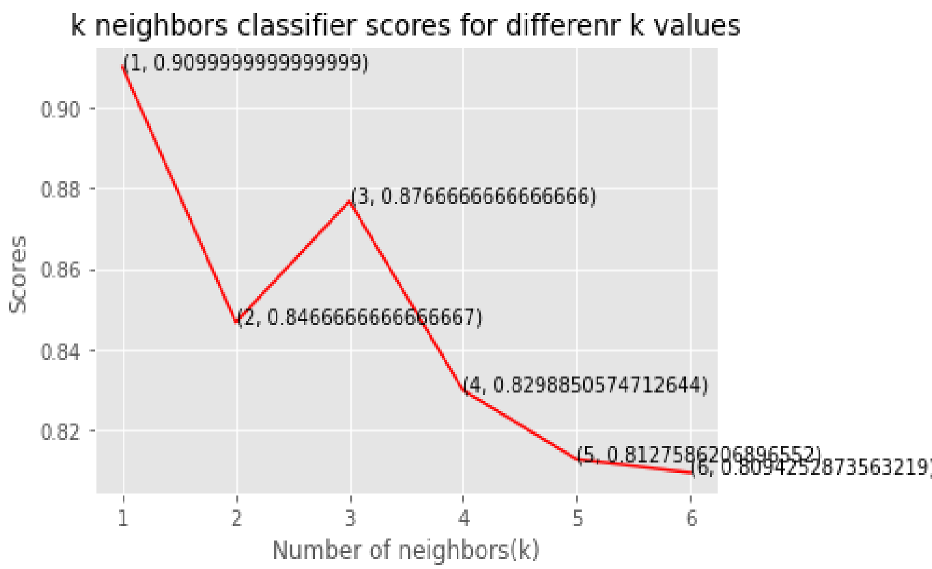
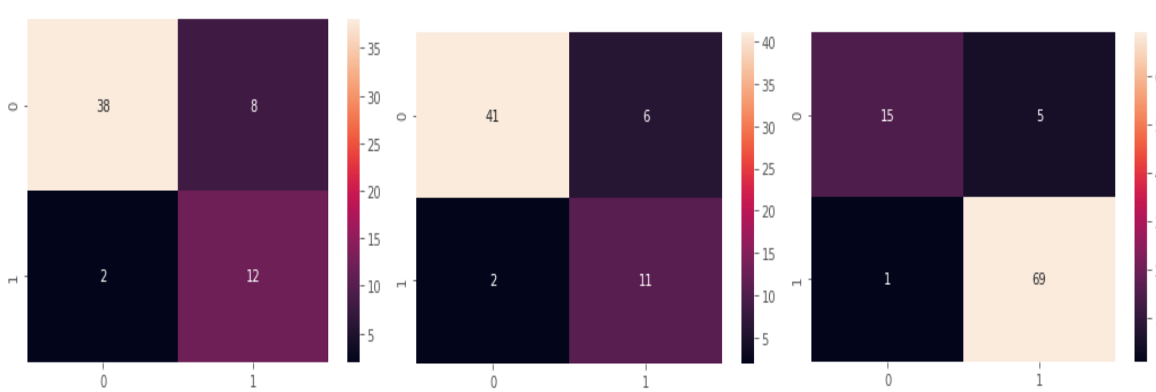



**Fig 4.17:** K neighbors classifier scores at different k values for university

In k nearest neighbor, k is the quantity of how many nearest values we need to consider for prediction in terms of distance. The distance is calculated by Eucledian or Manhattan distance.For example if we consider k=5,then 5 nearest neighbors are selected by eucledian distance. Random selection of k value does not gives us better accuracy. So that we have to derive the k value. For derivation we use a simple technique by doing a number of iteration. For choosing k value we run a loop from 1 to 40, these values are going to consider as k value. Here the error rate for different k values are obtained then we try to plot the figure and see where the error rate is quite stable. Now we can see from the first figure that for k=1 it obtain stability, that is after 1 the error line is going higher, for k=1 error rate in in between 0.00 to 0.10. For further understanding we can see the second figure that the score for k=1 is higher than the others. Though it gives us good accuracy but it is the problem of underfitting for considering 1 nearest neighbor.

### 4.3.3 Support Vector Machine



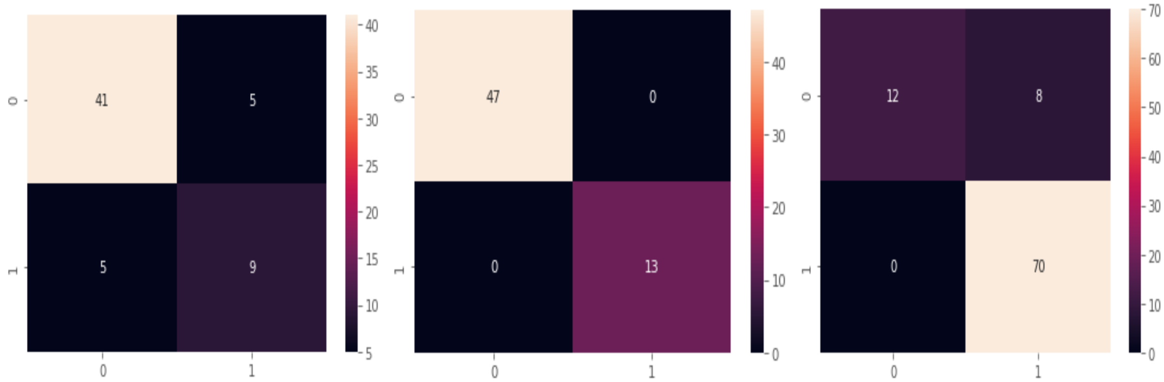

**(i) School** **(ii) College** **(iii) University**

**Fig 4.18:** Confusion Matrix for SVM with Linear kernel

In the first figure for linear kernel the true positive(TP) value is 38 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 12 which means The predicted value matches the actual value.The false positive(FP) value is 8 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 2 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value.

In the second figure for linear kernel the true positive(TP) value is 41 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 11 which means The predicted value matches the actual value.The false positive(FP) value is 6 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 2 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value.

In the third figure for linear kernel the true positive(TP) value is 15 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 69 which means The predicted value matches the actual value.The false positive(FP) value is 5 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 1 which means The predicted value was falsely predicted. The actual value was positive but the model predicted a negative value.



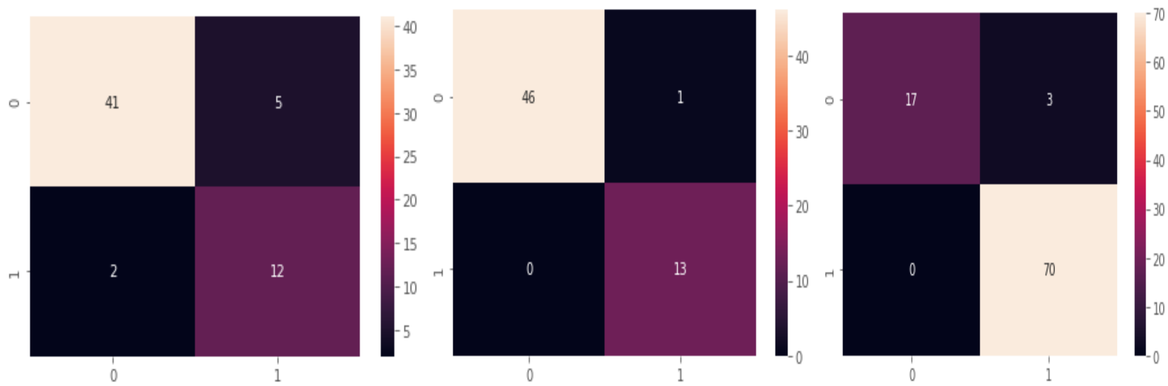

**(i) School** **(ii) College** **(iii) University**

**Fig 4.19:** Confusion Matrix for SVM with Polynomial kernel

In the first figure for polynomial kernel the true positive(TP) value is 41 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 9 which means The predicted value matches the actual value. The false positive(FP) value is 5 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 5 which means the predicted value was falsely predicted.

In the second figure for polynomial kernel the true positive(TP) value is 47 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 13 which means The predicted value matches the actual value. The false positive(FP) value is 0 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 0 which means the predicted value was falsely predicted.

In the third figure for polynomial kernel the true positive(TP) value is 12 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 70 which means The predicted value matches the actual value. The false positive(FP) value is 8 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 0 which means the predicted value was falsely predicted.



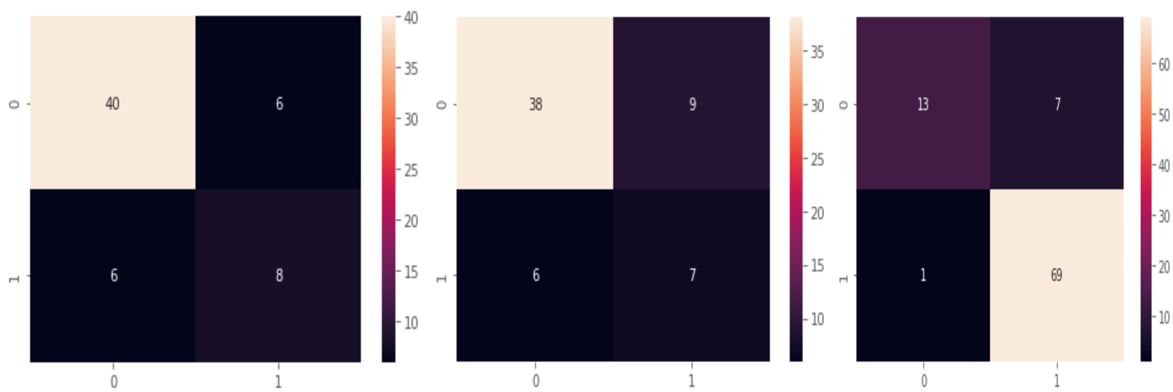

**(i) School** **(ii) College** **(iii) University**

**Fig 4.20:** Confusion Matrix for SVM with RBF kernel

In the first figure for RBF kernel the true positive(TP) value is 41 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 12 which means The predicted value matches the actual value. The false positive(FP) value is 5 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 2 which means The predicted value was falsely predicted.

In the second figure for sigmoid kernel the true positive (TP) value is 46 which means the predicted value matches the actual value.. The true negative (TN) value is 13 which means The predicted value matches the actual value. The false positive (FP) value is 1 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 0 which means the predicted value was falsely predicted.

In the third figure for RBF kernel the true positive(TP) value is 17 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 70 which means The predicted value matches the actual value. The false positive(FP) value is 3 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 0 which means The predicted value was falsely predicted.




**(i) School** **(ii) College** **(iii) University**

**Fig 4.21:** Confusion Matrix for SVM with Sigmoid kernel

In the first figure for sigmoid kernel the true positive(TP) value is 40 which means the predicted value matches the actual value.. The true negative(TN) value is 8 which means The predicted value matches the actual value. The false positive(FP) value is 6 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 6 which means the predicted value was falsely predicted.

In the second figure for RBF kernel the true positive (TP) value is 38 which means the predicted value matches the actual value. The actual value was positive and the model predicted a positive value. The true negative(TN) value is 7 which means The predicted value matches the actual value. The false positive(FP) value is 9 this is also known as type 1 error which means the predicted value was falsely predicted. The false negative(FN) value also known as Type 2 error is 6 which means The predicted value was falsely predicted.

In the third figure for sigmoid kernel the true positive(TP) value is 13 which means the predicted value matches the actual value.. The true negative(TN) value is 69 which means The predicted value matches the actual value. The false positive(FP) value is 7 this is also known as type 1 error which means the predicted value was falsely predicted. The actual value was negative but the model predicted a positive value. The false negative(FN) value also known as Type 2 error is 1 which means the predicted value was falsely predicted.

## 4.4 EVALUATION FOR DISTINCT METHODS

In our experiment, we evaluated the prediction results by evaluation experiments according to four values: Negative), FP (False Positive) and FN (False Negative) and calculated Precision, Recall,F-measure and Accuracy in each lesson. These are evaluated for artificial neural network(ANN), k nearest neighbor(KNN) and for support vector machine(SVM) as follows.

***(a)*** *Performance Evaluation Metrics For School Data*

At first we evaluated performance metrics for school data in a tabular form, which is shown in table 4.1. Here accuracy, precision, recall and F1 score is showed at percent

**Table 4.1:** Performance Evaluation Metrics For School Data

| **Evaluation Metrics** | **ANN** | **KNN** | | **SVM** | | | |
|---|---|---|---|---|---|---|---|
| | | **K=20** | **K=1** | **Linear Kernel** | **Polynomial Kernel** | **RBF kernel** | **Sigmoid Kernel** |
| Accuracy(%) | 91 | 80 | 92 | 83 | 83 | 88 | 80 |
| Precision(%) | 93 | 79 | 94 | 95 | 89 | 95 | 87 |
| Recall(%) | 96 | 100 | 96 | 83 | 89 | 89 | 87 |
| F1 Score(%) | 95 | 88 | 95 | 88 | 89 | 92 | 87 |

In multilayer perceptron neural network the accuracy is calculated from the equation (TP+TN)/(TP+FP+TN+FN) which is 95%.The TP(true positive), TN(true negative), FP(false positive), FN(false negative) values are obtained from the confusion matrix.The equation TP/(TP+FP) is used for precision calculation. The precision value is 100%. It tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. Precision is a useful metric in cases where False Positive is a higher concern than False Negatives. The equation TP/(TP+FN) for recall value is 94% which tells us how many of the actual positive cases we were able to predict correctly with our model. Recall is a useful metric in cases where false negative trumps false positive. And the equation for finding the value of F1 score is calculated by 2/((1/recall)+(1/precision)) is 97%. In practice, when we try to increase the precision of our model, the recall goes down, and vice-versa. The F1-score captures both the trends in a single value. F1-score is a harmonic mean of Precision and Recall, and so it gives a combined idea about these two metrics. The interpretability of the F1-score is poor. Now in the same manner the other two algorithms(KNN and SVM) predicted their evaluation metrices which is shown in the table.

***(b)*** *Performance Evaluation Metrics For College Data*

Here we evaluated performance metrics for college data in a tabular form, which is shown in table 4.2. Accuracy, precision, recall and F1 score is showed at percentage.

**Table 4.2:** Performance Evaluation Metrics for College Data

| **Evaluation Metrics** | **ANN** | **KNN** | | **SVM** | | | |
|---|---|---|---|---|---|---|---|
| | | **K=20** | **K=1** | **Linear Kernel** | **Polynomial Kernel** | **RBF kernel** | **Sigmoid Kernel** |
| **Accuracy(%)** | **100** | **78** | **95** | **87** | **100** | **98** | **75** |
| **Precision(%)** | **94** | **78** | **100** | **95** | **100** | **100** | **86** |
| **Recall(%)** | **100** | **100** | **94** | **87** | **95** | **98** | **81** |
| **F1 Score(%)** | **97** | **88** | **97** | **91** | **100** | **99** | **84** |

In multilayer perceptron the accuracy is calculated from the equation (TP+TN)/(TP+FP+TN+FN) which is 100%.The TP(true positive), TN(true negative), FP(false positive), FN(false negative) values are obtained from the confusion matrix.The equation TP/(TP+FP) is used for precision calculation. The precision value is 94%. It tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. Precision is a useful metric in cases where False Positive is a higher concern than False Negatives. The equation TP/(TP+FN) for recall value is 100% which tells us how many of the actual positive cases we were able to predict correctly with our model. Recall is a useful metric in cases where false negative trumps false positive. And the equation for finding the value of F1 score is calculated by 2/((1/recall)+(1/precision)) is 97%. In practice, when we try to increase the precision of our model, the recall goes down, and vice-versa. F1-score is a harmonic mean of Precision and Recall, and so it gives a combined idea about these two metrics. It is maximum when Precision is equal to Recall. But there is a catch here. The interpretability of the F1-score is poor. Now in the same manner the other two algorithms(KNN and SVM) predicted their evaluation metrices which is shown in the table.

***(c)*** *Performance Evaluation Metrics For University Data*

Here we evaluated performance metrics for university data in a tabular form, which is shown in table 4.3. Accuracy, precision, recall and F1 score is showed at percentage.

**Table 4.3:** Performance Evaluation Metrics for University Data

| **Evaluation Metrics** | **ANN** | **KNN** | | **SVM** | | | |
|---|---|---|---|---|---|---|---|
| | | **K=20** | **K=1** | **Linear Kernel** | **Polynomial Kernel** | **RBF kernel** | **Sigmoid Kernel** |
| **Accuracy(%)** | **98** | **79** | **91** | **93** | **91** | **97** | **91** |
| **Precision(%)** | **92** | **100** | **93** | **94** | **100** | **100** | **93** |
| **Recall(%)** | **80** | **5** | **65** | **75** | **60** | **85** | **65** |
| **F1 Score(%)** | **86** | **10** | **76** | **83** | **75** | **92** | **76** |

In multilayer perceptron the accuracy is calculated from the equation (TP+TN)/(TP+FP+TN+FN) which is 90%.The TP(true positive), TN(true negative), FP(false positive), FN(false negative) values are obtained from the confusion matrix. The equation TP/(TP+FP) is used for precision calculation. The precision value is 100%. It tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. Precision is a useful metric in cases where False Positive is a higher concern than False Negatives. The equation TP/(TP+FN) for recall value is 25% which tells us how many of the actual positive cases we were able to predict correctly with our model. Recall is a useful metric in cases where false negative trumps false positive. And the equation for finding the value of F1 score is calculated by 2/((1/recall)+(1/precision)) is 40%. In practice, when we try to increase the precision of our model, the recall goes down, and vice-versa. The F1-score captures both the trends in a single value. F1-score is a harmonic mean of Precision and Recall, and so it gives a combined idea about these two metrics. It is maximum when Precision is equal to Recall. But there is a catch here. The interpretability of the F1-score is poor. Now in the same manner the other two algorithms (KNN and SVM) predicted their evaluation metrics which is shown in the table.

## 4.5 CONCLUSION

From the above discussion we can conclude that an indicative number of tests of the system proved that ANN has comparatively a good rate in predicting the performance of students’ success. The parameter included in each students profile will achieve the model higher predictability. The other algorithms such as KNN and SVM are also showed a significant result in terms of predicting performance. Designed system with neural network is universal and can be applied to various institutions using only basic technical indicators as input data.

# CHAPTER 5

# CONCLUSION

## 5.1 INTRODUCTION

An artificial Neural Network model for predicting student success is presented in this study. The model use multilayer back propagation algorithm for training and testing. The factors for the model will be obtained from student questionnaire records. The model will tested and show overall result.

## 5.2 CONCLUSION

From the diminution of previous works about students' performance prediction, we try to construct a model that will be helpful for estimation of success rate of students' in an e-learning environment. To solve the problem of identifying the students who have a poor academic performance in school, college and university level three classification models have been built to predict the performance of students. Three machine learning techniques(k nearest neighbor and support vector machine) and multilayer perceptron network have been used. The models have been compared to one another using the performance evaluation metrics. For school students multilayer perceptron neural network the accuracy gained is which is 95%. The precision value is 100%. It tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. The recall value is 94% which tells us how many of the actual positive cases we were able to predict correctly with our model. The value of F1 score is calculated 97%. In practice, when we try to increase the precision of our model, the recall goes down. The F1-score captures both the trends in a single value. As the other two algorithms (KNN and SVM) predicted their evaluation metrics which is shown in the result and discussion section table 4.1. For college students in multilayer perceptron neural network the accuracy gained is which is 100%. The precision value is 94%. It tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. The recall value is 100% which tells us how many of the actual positive cases we were able to predict correctly with our model. The value of F1 score is calculated 97%. In practice, when we try to increase the precision of our model, the recall goes down. The F1-score captures both the trends in a single value. As the other two algorithms (KNN and SVM) predicted their evaluation metrics which is shown in the result and discussion section table 4.2. For university students in multilayer perceptron neural network the accuracy gained is which is 98%. The precision value is 92%. It tells us how many of the correctly predicted cases actually turned out to be positive. This would determine whether our model is reliable or not. The recall value is 80% which tells us how many of the actual positive cases we were able to predict correctly with our model. The value of F1 score is calculated 86%. In practice, when we try to increase the precision of our model, the recall goes down. The F1-score captures both the trends in a single value. As the other two algorithms (KNN and SVM) predicted their evaluation metrics which is shown in the result and discussion section table 4.3. This study showed the potential of artificial neural network for estimating student performance.

## 5.3 FUTURE SCOPE

- Development and implementation of a fully automated prediction system.
- Real time implementation of the model.
- Experimentation with different modelling approach. For example, homogenous classifiers and Genetic Algorithm.
- Validation of the model using other higher and further educational institutions

**REFERENCES**

[1] Oladokun, V.O., Adebanjo, A.T. and Charles-Owaba, O.E., 2008. Predicting students academic performance using artificial neural network: A case study of an engineering course.

[2] Oancea, B., Dragoescu, R. and Ciucu, S., 2013. Predicting students' results in higher education using a neural network.

[3] Livieris, I.E., Drakopoulou, K., Pintelas, P., 2012. Predicting students' performance using artificial neural networks, Proceedings of the 8th Pan-Hellenic Conference "Information and Communication Technology in Education", Available at : http://hcicte2012.uth.gr/main/sites/default/files/proc/Proceedings/LivierisEtAl.pdf, 17.01.2013.

[4] Calvo-Flores, M.D, Galindo, E.G., Pegalajar Jiménez, M.C. and Piñeiro, O.P., 2006. Predicting students' marks from Moodle logs using neural network models, A. Méndez-Vilas, A. Solano Martín, J.A. Mesa González and J. Mesa González (eds.) Current Developments in Technology-Assisted Education, pp. 586-590.

[5] Heaton, J., 2011. Programming Neural Networks with Encog3 in Java, 2nd Edition, Heaton Research

[6] El Moucary, C., Khair, M., Zakhem, W., 2011. Improving Student's Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education, The Research Bulletin of Jordan ACM, vol II, pp. 27-34

[7] Oladokun, V.O., Adebanjo, A.T., and Charles-Owaba, O.E., 2008. Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course, The Pacific Journal of Science and Technology, Volume 9, Number 1, pp. 72-79.

[8] Naik, B., Ragothaman, S., 2004. Using Neural Network to Predict MBA Student Success, College Student Journal, Vol. 38, No. 1. pp. 1-4.

[9] M. Riedmiller and H. Braun, 1993. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In Proceedings of the IEEE International Conference on Neural Networks, pages 586–591. IEEE Press.

[10] Ibrahim, Z. and Rusli, D., 2007, September. Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In *21st Annual SAS Malaysia Forum, 5th September*.

[11] Borkar, S. and Rajeswari, K., 2014. Attributes selection for predicting students' academic performance using education data mining and artificial neural network. *International Journal of Computer Applications*, *86*(10).

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

[13] Siri, A., 2015. Predicting students' dropout at university using artificial neural networks. *Italian Journal of Sociology of Education*, *7*(2).

[14] Mueen, A., Zafar, B. and Manzoor, U., 2016. Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, *8*(11), p.36.

[15] Lau, E.T., Sun, L. and Yang, Q., 2019. Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, *1*(9), p.982.

[16] Sikder, M.F., Uddin, M.J. and Halder, S., 2016, May. Predicting students yearly performance using neural network: A case study of BSMRSTU. In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)* (pp. 524-529). IEEE.

[17] Aydoğdu, Ş., 2020. Predicting student final performance using artificial neural networks in online learning environments. *Education and Information Technologies*, *25*(3), pp.1913-1927.

[18] Ahmad, Z. and Shahzadi, E., 2018. Prediction of Students' Academic Performance Using Artificial Neural Network. *Bulletin of Education and Research*, *40*(3), pp.157-164.

[19] Iyanda, A.R., Ninan, O.D., Ajayi, A.O. and Anyabolu, O.G., 2018. Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models. *International Journal of Modern Education & Computer Science*, *10*(6).

[20] Asogwa, O.C. and Oladugba, A.V., 2015. Of students academic performance rates using artificial neural networks (ANNs). *American Journal of Applied Mathematics and Statistics*, *3*(4), pp.151-155.

[21] Kotsiantis, S., Patriarcheas, K. and Xenos, M., 2010. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems*, *23*(6), pp.529-535.

[22] El-Sebakhy, E. A., Hadi, A. S. & Faisal, K. A. 2007 Iterative Least Squares Functional Networks Classifier. IEEE Transactions on Neural Networks, Vol.18, No.3, pp. 844-850.

[23] Diaz, A. L. 2003 Personal, family, and academic factors affecting low achievement in secondary school. Electronic Journal of Research in Educational Psychology and psychopadology, Vol. 1. No 1, pp. 43-66.

[24] Karamouzis, S.T. and Vrettos, A., 2008, October. An artificial neural network for predicting student graduation outcomes. In *Proceedings of the World Congress on Engineering and Computer Science* (pp. 991-994).

[25] Kotsiantis, S., Pierrakeas, C. and Pintelas, P., 2004. PREDICTING STUDENTS'PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES. *Applied Artificial Intelligence*, *18*(5), pp.411-426.

[26] Kalejaye, B.A., Folorunso, O. and Usman, O.L., 2015. Predicting students' grade scores using training functions of artificial neural network. *Journal of Natural Sciences Engineering and Technology*, *14*(1), pp.25-42.

[27] Abass,O., Oyekanlu ,E.A., Alaba, O.B, Longe. O.B., 2011. Forecasting Student Academic Performance using Case-Base Reasoning. International conference on ICT for Africa. 23-26th march, 2010 Otta Nigeria, Pp. 105-112

[28] Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., Loumos, V., 2009. Early and Dynamic Student Achievement Prediction in E-Learning Courses

Using Neural Networks. Journal of the American Society for Information Science and Technology, 60(2):372-380.

[29] Yegnanarayana, B., 2012. Artificial Neural Networks. PHI Learning Private Limited, M-97, Connaught Circus, New Delhi-110001, ISBN978-81-203-1253-1.

[30] Kokinov, B., 2003. Analogy in decisionmaking social interaction and emergent rationality. Behavioral and Sciences.26, 167-168.

[31] Rastrollo-Guerrero, J.L., Gómez-Pulido, J.A. and Durán-Domínguez, A., 2020. Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Applied Sciences*, *10*(3), p.1042.

[32] Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., Alhiyafi, J. and Olatunji, S.O., 2017, April. Student performance prediction using support vector machine and k-nearest neighbor. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1-4). IEEE.

[33] Taruna, S. and Pandey, M., 2014, February. An empirical analysis of classification techniques for predicting academic performance. In *2014 IEEE International Advance Computing Conference (IACC)* (pp. 523-528). IEEE.

[34] Kurniadi, D., Abdurachman, E., Warnars, H.L.H.S. and Suparta, W., 2018, November. The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm. In *IOP Conf. Ser. Mater. Sci. Eng* (Vol. 434, No. 1, p. 012039).

[35] Amra, I.A.A. and Maghari, A.Y., 2017, May. Students performance prediction using KNN and Naïve Bayesian. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 909-913). IEEE.

[36] Ko, C.Y. and Leu, F.Y., 2020. Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques. *IEEE Transactions on Education*.

[37] Pandey, M. and Taruna, S., 2014. A multi-level classification model pertaining to the student's academic performance prediction. *International Journal of Advances in Engineering & Technology*, *7*(4), p.1329.

[38] Lin, A., Wu, Q., Heidari, A.A., Xu, Y., Chen, H., Geng, W. and Li, C., 2019. Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-Nearest neighbor classifier. *Ieee Access*, *7*, pp.67235-67248.

[39] Zulfiker, M.S., Kabir, N., Al Amin Biswas, P.C. and Rahman, M.M., Predicting Students' Performance of the Private Universities of Bangladesh using Machine Learning Approaches.

[40] Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J. and Garcia, P., 2019. Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Sciences*.

[41] Konar, D., Pradhan, R., Dey, T., Sapkota, T. and Rai, P., 2020. Predicting Students' Grades Using CART, ID3, and Multiclass SVM Optimized by the Genetic Algorithm (GA): A Case Study. *Recent Advances in Hybrid Metaheuristics for Data Clustering*, pp.85-99.

[42] Waheed, H., Hassan, S.U., Aljohani, N.R., Hardman, J., Alelyani, S. and Nawaz, R., 2020. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, *104*, p.106189.

[43] Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., Alhiyafi, J. and Olatunji, S.O., 2017, April. Student performance prediction using support vector machine and k-nearest neighbor. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1-4). IEEE.

[44] Bhutto, E.S., Siddiqui, I.F., Arain, Q.A. and Anwar, M., 2020, February. Predicting Students' Academic Performance Through Supervised Machine Learning. In *2020 International Conference on Information Science and Communication Technology (ICISCT)* (pp. 1-6). IEEE.

[45] Asogbon, M.G., Samuel, O.W., Omisore, M.O. and Ojokoh, B.A., 2016. A multi-class support vector machine approach for students academic performance prediction. *Int. J. of Multidisciplinary and Current research*, *4*.

[46] Tsai, C.F. and Wu, J.W., 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications*, *34*(4), pp.2639-2649.

[47] Agarwal, S., Mishra, A.K. and Yadav, D.K., 2017. Forecasting price of amazon spot instances using neural networks. *Int. J. Appl. Eng. Res*, *12*(20), pp.10276-10283.

[48] Daranda, A., 2016. Neural network approach to predict marine traffic. *Trans. Balt. J. Mod. Comput*, *4*(3), p.483.

[49] Poudel, P. and Jang, B., 2017. Solar power prediction using deep learning. *Adv. Sci. Technol. Lett*, *146*, pp.148-151.

[50] Loy, J., 2019. *Neural Network Projects with Python: The ultimate guide to using Python to explore the true power of neural networks through six projects*. Packt Publishing Ltd.

## APPENDIX A

***Code implemented in analysis***

### Single layer perceptron

```
# Import required libraries:

import numpy as np

import pandas as pd

df = pd.read_csv("school1.csv")

array_input=df.to_numpy()

inputs = array_input

print (inputs.shape)

print (inputs)

# Define target output:

target_output =
np.array([[1,0,1,1,1,0,1,0,1,0,1,1,0,0,1,0,1,0,1,0,1,1,0,0,0,1,0,0,0,1,0,0,0,0,1,0,0,1,1,0,0,0,1,
1,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,1,0,1,0,0,0,0,0,0,1,1,0,0,0,0,0,1,0,1,0,0,0,0,0,0,1,0,0,0,0,
0,1,0,0,0,0,0,0,1,0,0,1,0,1,0,1,1,1,0,0,1,0,1,0,1,1,0,1,1,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,1,0,
0,0,0,0,0,1,0,0,0

]])

# Reshaping our target output into vector:

target_output = target_output.reshape(150,1)

print(target_output.shape)

print (target_output)

# Define weights:

weights = 2 * np.random.random((18,1)) - 1

print(weights.shape)

print (weights)
```

```
# Bias weight:
bias = 0.3
# Learning Rate:
lr = 0.05
# Sigmoid function:
def sigmoid(x):
    return 1/(1+np.exp(-x))
# Derivative of sigmoid function:
def sigmoid_der(x):
    return sigmoid(x)*(1-sigmoid(x))
for epoch in range(10000):
#feedforward
#Input for output layer
    output_in = np.dot(inputs, weights) + bias
#Output from output layer
    output_out = sigmoid(output_in)
    #Backpropogation
#Calculating error
    error = output_out - target_output
    print(error)
#Going with the formula:
    x = error.sum()
    #print(x)
#Calculating intermediate derivatives:
 derror_douto = output_out - target_output
    douto_dino = sigmoid_der(output_out)
    dino_dwo = inputs
```

```
#Calculating Final derivative
    derror_dwo = np.dot(inputs.T,derror_douto * douto_dino)
#Updating the weights values:
    weights = weights - lr * derror_dwo
```

**Multi layer perceptron**

```
import numpy as np
import pandas as pd
# Define input features :
df = pd.read_csv("school1.csv")
np_array=df.to_numpy()
input_features = np_array
print (input_features.shape)
print (input_features)
# Define target output :
target_output =
np.array([[1,0,1,1,1,0,1,0,1,0,1,1,0,0,1,0,1,0,1,0,1,1,0,0,0,1,0,0,0,1,0,0,0,0,1,0,0,1,1,0,0,0,1,
1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,1,0,1,0,0,0,0,0,0,0,1,1,0,0,0,0,0,1,0,1,0,0,0,0,0,0,1,0,0,0,0,
0,1,0,0,0,0,0,0,1,0,0,1,0,1,0,1,1,1,0,0,1,0,1,0,1,1,0,1,1,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,1,0,
0,0,0,0,0,1,0,0,0
target_output = target_output.reshape(150,1)
print(target_output.shape)
print (target_output)
# Define weights :
weight_hidden = np.random.rand(18,10)
weight_output = np.random.rand(10,1)
# Learning Rate :
lr = 0.05
# Sigmoid function :
```

```
def sigmoid(x):
    return 1/(1+np.exp(-x))
# Derivative of sigmoid function :
def sigmoid_der(x):
    return sigmoid(x)*(1-sigmoid(x))
for epoch in range(10000):
 # Input for hidden layer :
    hidden_in = np.dot(input_features, weight_hidden)
 # Output from hidden layer :
    hidden_out = sigmoid(hidden_in)
 # Input for output layer :
    output_in = np.dot(hidden_out, weight_output)
 # Output from output layer :
    output_out = sigmoid(output_in)
#=======================Phase1==========================
 # Calculating Mean Squared Error :
    print ("Mean Squared Error")
    error_out = ((1 / 2) * (np.power((output_out - target_output), 2)))
    print(error_out.sum()
 # Derivatives for phase 1 :
    derror_douto = output_out - target_output
    douto_dino = sigmoid_der(output_in)
    dino_dwo = hidden_out
 # Final Derivative
    derror_dwo = np.dot(dino_dwo.T, derror_douto * douto_dino)
 #=========================Phase2=======================
  # Derivatives for phase 2 :
```

```
    derror_dino = derror_douto * douto_dino
    dino_douth = weight_output
    derror_douth = np.dot(derror_dino , dino_douth.T)
    douth_dinh = sigmoid_der(hidden_in)
    dinh_dwh = input_features
  # Final Derivative
    derror_wh = np.dot(dinh_dwh.T, douth_dinh * derror_douth)
  # Update Weights
    weight_hidden = weight_hidden - lr * derror_wh
    weight_output = weight_output - lr * derror_dwo
    # Final values of weights in the hidden layer
print ("hidden layer weight")
print (weight_hidden)
# Final values of weights in the output layer
print ("output layer weight")
print (weight_output)
#Taking inputs :
df = pd.read_csv("school2.csv")
test_data=df.to_numpy()
#1st step :
result1 = np.dot(test_data, weight_hidden)
#2nd step :
result2 = sigmoid(result1)
#3rd step :
result3 = np.dot(result2,weight_output)
#4th step :
result4 = sigmoid(result3)
```

print(result4)

**KNN Code:**

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=20)
knn.fit(x_train, y_train)
pred = knn.predict(x_test)
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.model_selection import cross_val_score
print('WITH K=20')
print('\n')
print(confusion_matrix(y_test,pred))
print(classification_report(y_test,pred))
import seaborn as sns
cm=confusion_matrix(y_test,pred)
sns.heatmap(cm,annot=True)
plt.savefig('h.png')
error_rate = []
for i in range(1,40):
 knn = KNeighborsClassifier(n_neighbors=i)
 knn.fit(x_train,y_train)
 pred_i = knn.predict(x_test)
 error_rate.append(np.mean(pred_i != y_test))
import matplotlib.pyplot as plt
plt.style.use('ggplot')
plt.figure(figsize=(10,6))
plt.plot(range(1,40),error_rate,color='blue', linestyle='dashed', marker='o',
```

```
markerfacecolor='red', markersize=10)
plt.title('Error Rate vs. K Value')
plt.xlabel('K')
plt.ylabel('Error Rate')
from sklearn.model_selection import cross_val_score
knn_scores=[]
for k in range(1,7):
knn_classifier=KNeighborsClassifier(n_neighbors=k)
score=cross_val_score(knn_classifier,x,y,cv=10)
knn_scores.append(score.mean())
plt.plot([k for k in range(1,7)],knn_scores,color='red')
for i in range (1,7):
plt.text(i,knn_scores[i-1],(i,knn_scores[i-1]))
plt.xticks([i for i in range(1,7)])
plt.xlabel('Number of neighbors(k)')
plt.ylabel('Scores')
plt.title('k neighbors classifier scores for differenr k values')
knn = KNeighborsClassifier(n_neighbors=1)


knn.fit(x_train,y_train)
pred = knn.predict(x_test)
print('WITH K=1')
print('\n')
print(confusion_matrix(y_test,pred))
print('\n')
print(classification_report(y_test,pred))
```

**SVM Code:**

```
from sklearn.svm import SVC
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.metrics import confusion_matrix
```

**Linear kernel**

```
classifier = SVC(kernel = 'linear',C=100)
classifier.fit(x_train, y_train)
pred = classifier.predict(x_test)
print(confusion_matrix(y_test,pred))
import seaborn as sns
cm=confusion_matrix(y_test,pred)
sns.heatmap(cm,annot=True)
plt.savefig('h.png')
print(classification_report(y_test,pred))
```

**Polynomial kernel**

```
svclassifier = SVC(kernel='poly',C=100)
svclassifier.fit(x_train, y_train)
pred = svclassifier.predict(x_test)
print(confusion_matrix(y_test,pred))
print(classification_report(y_test,pred))
import seaborn as sns
cm=confusion_matrix(y_test,pred)
sns.heatmap(cm,annot=True)
plt.savefig('h.png')
```

**RBF/Gaussian Kernel**

```
svclassifier = SVC(kernel='rbf',C=100)
```

```
svclassifier.fit(x_train, y_train)
pred = svclassifier.predict(x_test)
print(confusion_matrix(y_test,pred))
print(classification_report(y_test,pred))
import seaborn as sns
cm=confusion_matrix(y_test,pred)
sns.heatmap(cm,annot=True)
plt.savefig('h.png')
```

**Sigmoid Kernel**

```
svclassifier = SVC(kernel='sigmoid',C=100)
svclassifier.fit(x_train, y_train)
pred = svclassifier.predict(x_test)
print(confusion_matrix(y_test,pred))
print(classification_report(y_test,pred))
import seaborn as sns
cm=confusion_matrix(y_test,pred)
sns.heatmap(cm,annot=True)
plt.savefig('h.png')
```

# APPENDIX B

## *Data used for analysis*

### *School Data*

| Gender | Age | Popular website | proficiency | Medium | Location | household internet facilities | Browse time | browsing status | Residence | Total internet use | Time spend in academic | Purpose of use | Years of internet use | Browsing purpose | Online course | Webinar | Class performance | Obstacles | Target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| female | 14 | youtube | better | mobile | home | connected | 6pm-12am | daily | town | 0_5 | 0_5 | music | >1 | acdemic | no | no | good | slow | 1 |
| female | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | gaming | >1 | acdemic | no | no | average | slow | 0 |
| female | 14 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | good | slow | 1 |
| male | 15 | youtube | better | mobile | home | not connect | 6am-12am | daily | town | 0_5 | 0_5 | music | 1 to 5 | non-academic | no | no | good | payment | 1 |
| male | 14 | youtube | good | laptop and mobile | school | connected | 6pm-12am | daily | town | rarely | rarely | music | >1 | non-academic | no | no | good | slow | 1 |
| male | 16 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | average | lack of computer | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 12pm-6pm | weekly | village | 0_5 | 0_5 | music | 1 to 5 | acdemic | no | no | satisfactory | cost of computer equipment | 1 |
| female | 14 | google | average | mobile | home | not connect | 6am-12pm | daily | town | 0_5 | rarely | social sites | >1 | acdemic | no | no | average | power failure | 0 |
| female | 15 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | rarely | music | >1 | non-academic | no | no | good | slow | 1 |
| fem | 15 | others | good | mobile | hom | not connect | 6pm-12 | weekly | town | 0_5 | 0_5 | social sit | 1 to 5 | acdemic | no | no | average | poor computer | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ale | | | | | e | | am | | | | | es | | | | | | skill | |
| male | 14 | others | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | satisfactory | payment | 0 |
| female | 14 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | not satisfactory | power failure | 0 |
| female | 14 | youtube | good | mobile | home | not connect | 6pm-12am | monthly | town | 0_5 | 0_5 | gaming | >1 | acdemic | no | no | good | slow | 0 |
| female | 14 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | village | 0_5 | 0_5 | music | >1 | acdemic | no | no | average | slow | 0 |
| female | 14 | google | good | mobile | home | not connect | 6pm-12am | monthly | village | 5_10 | 0_5 | music | >1 | acdemic | no | no | average | slow | 0 |
| male | 17 | google | good | mobile | home | not connect | 6am-12am | weekly | town | 5_10 | 5_10 | file sharing | 5_10 | acdemic | no | no | good | lack of computer | 1 |
| male | 14 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | >1 | non-academic | no | no | good | slow | 0 |
| male | 17 | youtube | good | laptop and mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | average | cost of computer equipment | 0 |
| female | 14 | youtube | average | mobile | home | not connect | 6pm-12am | daily | village | 5_10 | 0_5 | group study | >1 | acdemic | no | no | good | slow | 1 |
| female | 14 | google | average | mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | average | power failure | 0 |
| female | 14 | youtube | better | mobile | home | connected | 6pm-12am | weekly | town | 0_5 | 0_5 | music | >1 | acdemic | no | no | good | slow | 1 |
| female | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | gaming | >1 | acdemic | no | no | average | slow | 0 |
| fe | 14 | google | good | mobile | ho | not conne | 6pm- | daily | tow | 0_5 | 0_5 | social | 1 to 5 | acdemi | no | no | good | slow | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | | | | | me | ct | 12am | | n | | | sites | | c | | | | | |
| male | 15 | youtube | better | mobile | home | not connect | 6am-12am | weekly | town | 0_5 | 0_5 | music | 1 to 5 | non-academic | no | no | good | payment | 1 |
| male | 14 | youtube | good | laptop and mobile | school | connected | 6pm-12am | daily | town | 0_5 | 0_5 | music | >1 | non-academic | no | no | good | slow | 1 |
| male | 16 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | average | lack of computer | 0 |
| male | 16 | others | not at all | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | rarely | shopping | >1 | acdemic | no | no | not satisfactory | payment | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 12pm-6pm | daily | village | 0_5 | 0_5 | music | 1 to 5 | acdemic | no | no | satisfactory | cost of computer equipment | 1 |
| female | 14 | google | average | mobile | home | not connect | 6am-12am | daily | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | average | power failure | 0 |
| female | 15 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | music | >1 | non-academic | no | no | good | slow | 1 |
| male | 14 | youtube | better | laptop and mobile | home | not connect | 12am-6am | daily | town | 10_15 | 5_10 | file sharing | 1 to 5 | acdemic | no | no | average | slow | 1 |
| male | 15 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | music | >1 | acdemic | no | no | good | payment | 0 |
| male | 14 | whatsapp | average | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | news | >1 | acdemic | no | no | good | cost of computer equipment | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | shopping | >1 | acdemic | no | no | good | payment | 0 |
| male | 15 | youtube | good | mobile | others | not connect | 6pm-12am | weekly | town | 0_5 | rarely | gaming | >1 | acdemic | no | no | good | power failure | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 14 | google | good | mobile | home | not connect | 6am-12pm | daily | village | 0_5 | 0_5 | gaming | >1 | non-academic | no | no | good | poor computer skill | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 5_10 | group study | 1 to 5 | acdemic | no | no | average | slow | 1 |
| male | 15 | others | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | average | poor computer skill | 0 |
| male | 14 | others | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | rarely | social sites | >1 | acdemic | no | no | satisfactory | payment | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | 1 to 5 | acdemic | no | no | not satisfactory | power failure | 0 |
| male | 15 | youtube | better | laptop and mobile | home | not connect | 6pm-12am | weekly | town | 10_15 | 10_15 | social sites | 1 to 5 | acdemic | yes | no | good | slow | 1 |
| male | 14 | youtube | good | mobile | school | not connect | 6pm-12am | weekly | town | 0_5 | 5_10 | music | >1 | acdemic | no | no | good | lack of computer | 1 |
| male | 15 | whatsapp | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | rarely | music | 1 to 5 | acdemic | no | no | average | payment | 0 |
| male | 15 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 5_10 | social sites | 1 to 5 | acdemic | no | no | average | payment | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | file sharing | 1 to 5 | non-academic | no | no | average | slow | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 5_10 | group study | >1 | acdemic | no | no | average | payment | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | shopping | >1 | acdemic | no | no | average | poor computer skill | 0 |
| male | 14 | youtube | better | mobile | home | not connect | 6pm-12a | daily | town | 5_10 | 5_10 | group study | 1 to 5 | acdemic | no | no | good | payment | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | m | | | | | | | | | | | | |
| male | 14 | google | good | mobile | home | not connect | 6pm-12am | daily | village | 5_10 | 0_5 | music | >1 | acdemic | no | no | average | slow | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | 10_15 | file sharing | 1 to 5 | acdemic | no | no | good | payment | 1 |
| male | 15 | youtube | not at all | mobile | home | not connect | 6pm-12am | weekly | town | 5_10 | 0_5 | gaming | >1 | acdemic | no | no | not satisfactory | power failure | 0 |
| male | 14 | yahoo | good | mobile | home | not connect | 6pm-12am | daily | village | 5_10 | 0_5 | group study | >1 | acdemic | no | no | average | lack of computer | 0 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | rarely | rarely | shopping | 1 to 5 | non-academic | no | no | good | cost of computer equipment | 0 |
| male | 15 | youtube | average | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | news | >1 | acdemic | no | no | average | payment | 0 |
| male | 15 | others | good | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | 0_5 | gaming | >1 | acdemic | no | no | not satisfactory | payment | 0 |
| male | 17 | youtube | not at all | mobile | home | not connect | 6pm-12am | weekly | remote | 0_5 | 0_5 | music | >1 | non-academic | no | no | good | slow | 0 |
| male | 15 | whatsapp | better | mobile | school | not connect | 6pm-12am | daily | town | 5_10 | 5_10 | group study | 5_10 | acdemic | no | no | average | payment | 1 |
| male | 14 | youtube | good | mobile | home | not connect | 6pm-12am | daily | remote | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | good | slow | 0 |
| male | 15 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | music | 5_10 | acdemic | no | no | good | slow | 1 |
| male | 15 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | music | 1 to 5 | non-academic | no | no | average | slow | 1 |

***College Data***

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| female | 17 | youtube | average | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | good | slow | 0 |
| female | 17 | youtube | average | mobile | home | not connect | 6pm-12am | weekly | village | 0_5 | rarely | social sites | >1 | acdemic | yes | no | average | payment | 1 |
| female | 17 | google | average | mobile | home | not connect | 12pm-6pm | weekly | town | 0_5 | 5_10 | social sites | 5_10 | acdemic | no | no | satisfactory | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | news | 5_10 | acdemic | no | no | good | slow | 0 |
| female | 17 | youtube | better | desktop | home | not connect | 6pm-12am | daily | town | 10_15 | 5_10 | news | 10< | acdemic | yes | no | average | cost of computer equipment | 1 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | social sites | 1to5 | acdemic | no | no | good | slow | 0 |
| female | 17 | youtube | better | laptop and mobile | home | not connect | 6pm-12am | weekly | town | 15_20 | 15_20 | music | 5_10 | acdemic | no | no | good | lack of computer | 1 |
| female | 17 | whatsapp | good | mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 5_10 | group study | 5_10 | acdemic | no | no | satisfactory | payment | 1 |
| female | 17 | youtube | average | mobile | home | not connect | 12pm-6pm | daily | village | 0_5 | 0_5 | social sites | >1 | non-academic | no | no | not satisfactory | payment | 0 |
| female | 17 | gmail | good | desktop | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | music | 10< | acdemic | yes | no | not satisfactory | payment | 1 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 10_15 | 0_5 | social sites | >1 | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | remote | 0_5 | 0_5 | group study | 1to5 | acdemic | no | no | good | slow | 0 |
| fe | 1 | yo | be | mobil | h | not | 6p | d | to | 1 | 0 | blo | > | acde | n | n | good | poor | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 6 | utube | tter | e | ome | connect | m-12am | aily | wn | 5_20 | _5 | gs | 1 | mic | o | o | | computer skill | |
| female | 17 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 10_15 | music | 1to5 | acdemic | no | no | good | lack of computer | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 10_15 | rarely | news | 1to5 | acdemic | yes | no | good | cost of computer equipment | 0 |
| female | 17 | google | average | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | 5_10 | shopping | 1to5 | non-academic | no | no | average | power failure | 0 |
| male | 17 | google | good | mobile | home | connected | 12pm-6pm | daily | town | 20_24 | 0_5 | gaming | 1to5 | acdemic | no | no | satisfactory | power failure | 1 |
| male | 17 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 10_15 | gaming | 1to5 | acdemic | no | yes | good | cost of computer equipment | 1 |
| female | 17 | youtube | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | rarely | blogs | 1to5 | acdemic | yes | no | good | slow | 0 |
| male | 17 | google | average | mobile | home | not connect | 6pm-12am | daily | village | 10_15 | 0_5 | social sites | 1to5 | non-academic | no | no | average | slow | 0 |
| male | 17 | gmail | good | mobile | home | connected | 6pm-12am | daily | town | 15_20 | 5_10 | group study | 1to5 | acdemic | no | no | good | power failure | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 10_15 | 5_10 | news | 1to5 | acdemic | no | yes | good | slow | 1 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 20_24 | 0_5 | group study | 10< | acdemic | no | no | good | slow | 0 |
| male | 17 | others | good | mobile | college | not connect | 6pm-12am | daily | town | 20_24 | rarely | file sharing | 5_10 | acdemic | no | no | good | slow | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| female | 17 | google | good | tablet | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | group study | 5_10 | acdemic | no | no | not satisfactory | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 0_5 | music | 5_10 | acdemic | no | no | good | power failure | 0 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | rarely | news | 10< | acdemic | no | no | good | slow | 0 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | social sites | 5_10 | acdemic | no | no | good | poor computer skill | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | rarely | social sites | 10< | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 15_20 | blogs | 5_10 | acdemic | no | no | satisfactory | payment | 1 |
| male | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | town | 5_10 | 10_15 | shopping | 5_10 | acdemic | no | yes | good | lack of computer | 1 |
| female | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | daily | village | 5_10 | 0_5 | gaming | 5_10 | acdemic | yes | no | satisfactory | power failure | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 10_15 | social sites | 10< | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 5_10 | file sharing | 5_10 | acdemic | no | no | good | slow | 0 |
| female | 17 | youtube | average | mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | satisfactory | slow | 0 |
| female | 17 | youtube | average | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | good | slow | 0 |
| male | 17 | youtube | average | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | 0_5 | social sites | >1 | acdemic | yes | no | average | payment | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 17 | google | average | mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | 5_10 | social sites | 5_10 | acdemic | no | no | satisfactory | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 5_10 | 0_5 | news | 5_10 | acdemic | no | no | good | slow | 0 |
| female | 17 | youtube | better | desktop | home | not connect | 6pm-12am | daily | town | 10_15 | 5_10 | news | 10< | acdemic | yes | no | average | cost of computer equipment | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | social sites | 1to5 | acdemic | no | no | good | slow | 0 |
| male | 17 | youtube | better | laptop and mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 15_20 | music | 5_10 | acdemic | no | no | good | lack of computer | 1 |
| female | 17 | whatsapp | good | mobile | home | not connect | 6pm-12am | weekly | town | 15_20 | 5_10 | group study | 5_10 | acdemic | no | no | satisfactory | payment | 1 |
| female | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | weekly | village | 5_10 | 0_5 | gaming | 5_10 | acdemic | yes | no | satisfactory | power failure | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | rarely | social sites | 10< | acdemic | no | no | good | cost of computer equipment | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | shopping | 10< | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | remote | 5_10 | 0_5 | social sites | 5_10 | acdemic | no | no | average | payment | 0 |
| male | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | weekly | town | 5_10 | 0_5 | file sharing | 5_10 | acdemic | no | no | good | payment | 0 |
| male | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | rarely | group study | 5_10 | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | music | 10< | acdemic | no | no | average | power failure | 0 |
| mal | 16 | googl | good | mobile | ho | not con | 6pm- | dail | villa | 0_ | 0_ | blogs | >1 | acdemic | no | no | average | payment | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e | | e | | | me | nect | 12am | y | ge | 5 | 5 | | | | | | | | |
| female | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | daily | town | 0_5 | rarely | social sites | >1 | acdemic | no | no | average | poor computer skill | 0 |
| female | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | good | payment | 0 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 10_15 | social sites | 10< | acdemic | no | no | good | payment | 0 |
| female | 17 | whatsapp | good | mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 5_10 | group study | 5_10 | acdemic | no | no | satisfactory | payment | 1 |
| male | 17 | youtube | average | mobile | home | not connect | 12pm-6pm | weekly | village | 0_5 | 0_5 | social sites | >1 | non-academic | no | no | not satisfactory | payment | 0 |
| female | 17 | gmail | good | desktop | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | music | 10< | acdemic | yes | no | not satisfactory | payment | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 10_15 | rarely | social sites | >1 | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | remote | 0_5 | 0_5 | group study | 1to5 | acdemic | no | no | good | slow | 0 |
| male | 16 | youtube | better | mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 0_5 | blogs | >1 | acdemic | no | no | good | poor computer skill | 0 |
| male | 17 | whatsapp | good | mobile | college | not connect | 6pm-12am | daily | town | 5_10 | 5_10 | news | 10< | acdemic | no | no | average | lack of computer | 0 |
| male | 17 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 10_15 | 15_20 | group study | 1to5 | acdemic | no | no | satisfactory | cost of computer equipment | 1 |
| female | 18 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 10_15 | 0_5 | music | >1 | acdemic | no | no | good | lack of computer | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 17 | google | good | mobile | home | connected | 6pm-12am | daily | town | 15_20 | rarely | shopping | 5_10 | acdemic | no | no | good | power failure | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | rarely | rarely | social sites | 10< | non-academic | no | no | average | cost of computer equipment | 0 |
| female | 17 | google | good | mobile | college | not connect | 6pm-12am | daily | town | 15_20 | 0_5 | social sites | 5_10 | acdemic | no | no | good | cost of computer equipment | 0 |
| female | 16 | google | good | laptop and mobile | home | not connect | 12pm-6pm | daily | town | 5_10 | 0_5 | shopping | 1to5 | acdemic | no | no | good | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 10_15 | 0_5 | news | 1to5 | acdemic | yes | no | good | cost of computer equipment | 0 |
| female | 17 | google | average | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | rarely | shopping | 1to5 | non-academic | no | no | average | power failure | 0 |
| male | 17 | google | good | mobile | home | connected | 12pm-6pm | daily | town | 20_24 | rarely | gaming | 1to5 | acdemic | no | no | satisfactory | power failure | 1 |
| male | 17 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 10_15 | gaming | 1to5 | acdemic | no | yes | good | cost of computer equipment | 1 |
| female | 17 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | blogs | 1to5 | acdemic | yes | no | good | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 0_5 | music | 5_10 | acdemic | no | no | good | power failure | 0 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | news | 10< | acdemic | no | no | good | slow | 0 |
| male | 17 | google | good | mobile | home | connected | 6pm-12am | weekly | town | 5_10 | 0_5 | social sites | 5_10 | acdemic | no | no | good | poor computer skill | 0 |
| femal | 17 | google | good | mobile | hom | not connec | 6pm-12a | daily | town | 0_5 | 0_5 | social site | 10< | acdemic | no | no | good | payment | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e | | | | | e | t | m | | | | | s | | | | | | | |
| female | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 15_20 | blogs | 5_10 | acdemic | no | no | satisfactory | payment | 1 |
| male | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | town | 5_10 | 10_15 | shopping | 5_10 | acdemic | no | yes | good | lack of computer | 1 |
| female | 17 | google | good | laptop and mobile | home | connected | 6pm-12am | daily | village | 5_10 | 0_5 | gaming | 5_10 | acdemic | yes | no | satisfactory | power failure | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 10_15 | social sites | 10< | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 5_10 | file sharing | 5_10 | acdemic | no | no | good | slow | 0 |
| female | 17 | you tube | average | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | rarely | social sites | >1 | acdemic | no | no | satisfactory | slow | 0 |
| female | 17 | you tube | average | mobile | home | connected | 6pm-12am | weekly | town | 0_5 | 0_5 | social sites | >1 | acdemic | no | no | good | slow | 0 |
| male | 17 | you tube | average | mobile | home | connected | 6pm-12am | daily | village | 0_5 | 0_5 | social sites | >1 | acdemic | yes | no | average | payment | 1 |
| male | 17 | google | average | mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | 5_10 | social sites | 5_10 | acdemic | no | no | satisfactory | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | news | 5_10 | acdemic | no | no | good | slow | 0 |
| female | 17 | you tube | better | desktop | home | not connect | 6pm-12am | daily | town | 10_15 | 5_10 | news | 10< | acdemic | yes | no | average | cost of computer equipment | 1 |
| male | 17 | google | good | mobile | home | not connect | 6pm-12am | weekly | town | 5_10 | 0_5 | social sites | 1to5 | acdemic | no | no | good | slow | 0 |
| male | 17 | you tu | better | laptop and mobil | hom | not connec | 6pm-12a | daily | town | 15_ | 15_ | music | 5_1 | acdemic | no | no | good | lack of computer | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | be | | e | e | t | m | | | 20 | 20 | | 0 | | | | | | |
| female | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | remote | 5_10 | 0_5 | social sites | 5_10 | acdemic | no | no | average | payment | 0 |
| male | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | daily | town | 5_10 | 0_5 | file sharing | 5_10 | acdemic | no | no | good | payment | 0 |
| male | 17 | google | good | mobile | home | not connect | 12pm-6pm | daily | town | 0_5 | rarely | group study | 5_10 | acdemic | no | no | good | payment | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 0_5 | music | 10< | acdemic | no | no | average | power failure | 0 |
| male | 16 | google | good | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | 0_5 | blogs | >1 | acdemic | no | no | average | payment | 0 |
| female | 17 | google | good | laptop and mobile | home | not connect | 6pm-12am | weekly | town | 0_5 | rarely | social sites | >1 | acdemic | no | no | average | poor computer skill | 0 |
| female | 17 | google | good | mobile | home | connected | 6pm-12am | weekly | town | 10_15 | 0_5 | news | 1to5 | acdemic | yes | no | good | cost of computer equipment | 0 |
| female | 17 | google | average | mobile | home | not connect | 6pm-12am | daily | village | 0_5 | 5_10 | shopping | 1to5 | non-academic | no | no | average | power failure | 0 |
| male | 17 | google | good | mobile | home | connected | 12pm-6pm | daily | town | 20_24 | rarely | gaming | 1to5 | acdemic | no | no | satisfactory | power failure | 1 |
| male | 17 | youtube | good | mobile | home | not connect | 6pm-12am | daily | town | 0_5 | 10_15 | gaming | 10< | non-academic | no | yes | good | cost of computer equipment | 1 |
| female | 17 | youtube | good | mobile | home | connected | 6pm-12am | daily | town | 0_5 | 0_5 | blogs | 1to5 | non-academic | yes | no | good | slow | 0 |
| female | 17 | google | good | mobile | home | not connect | 6pm-12am | daily | town | 15_20 | 0_5 | music | >1 | non-academic | no | no | good | power failure | 0 |

### *University Data*

| Gender | Age | Popular website | proficiency | Medium | Location | household internet facilities | Browse time | browsing status | Residence | Total internet use | Time spend in academic | Purpose of use | Years of internet use | Browsing purpose | Online course | Webinar | Class performance | obstacles | Result | Target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 24 | youtube | good | laptop and mobile | cyber | connected | 6pm-12am | daily | town | 15_20 | 5_10 | file sharing | 10< | non-academic | yes | no | satisfactory | search engine | 3.52 | 1 |
| male | 20 | youtube | good | laptop and mobile | home | connected | 6pm-12am | weekly | town | 10_15 | 0_5 | group study | 10< | non-academic | no | no | good | follow hyperlink | 3.2 | 0 |
| male | 23 | youtube | better | laptop and mobile | university | connected | 6pm-12am | daily | town | 10_15 | 0_5 | social sites | 10< | non-academic | yes | no | not satisfactory | follow hyperlink | 3.41 | 0 |
| male | 22 | google | better | laptop and mobile | cyber | connected | 12am-6am | daily | village | 15_20 | 5_10 | group study | 10< | non-academic | no | no | good | directories | 3.55 | 1 |
| male | 22 | youtube | better | mobile | home | connected | 6pm-12am | daily | town | 10_15 | 0_5 | social sites | 10< | non-academic | yes | yes | good | newspaper | 3.75 | 1 |
| male | 23 | google | good | laptop and mobile | home | not connect | 12am-6am | daily | town | 15_20 | 0_5 | social sites | 10< | acdemic | yes | no | good | directories | 3.89 | 1 |
| male | 22 | gmail | better | laptop and mobile | library | connected | 6pm-12am | daily | town | 10_15 | 5_10 | group study | 10< | acdemic | yes | no | satisfactory | search engine | 3.5 | 1 |
| male | 22 | gmail | good | desktop | home | not connect | 6pm-12am | weekly | village | 0_5 | 0_5 | gaming | 5_10 | acdemic | yes | no | not satisfactory | search engine | 3 | 0 |
| male | 23 | gmail | good | laptop and mobile | home | connected | 6pm-12am | daily | town | 10_15 | 0_5 | shopping | 10< | acdemic | yes | no | good | directories | 3.72 | 1 |
| male | 23 | youtube | average | mobile | home | connected | 12am-6am | daily | town | 5_10 | 0_5 | social sites | 5_10 | acdemic | yes | yes | good | directories | 3.58 | 1 |
| mal | 23 | instagram | good | laptop and | librar | not connect | 6am-12 | monthly | town | 0 t0 5 | 0 t0 5 | shoppin | 5_10 | acdemic | yes | no | average | friends and | 3.4 | 0 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e | | | | mobile | y | | pm | | | | | g | | | | | | teachers | 5 | |
| female | 23 | messenger | average | mobile | university | connected | 6am-12pm | daily | remote | rarely | rarely | music | 5_10 | non-academic | yes | no | average | friends and teachers | 3.29 | 0 |
| male | 23 | youtube | good | mobile | cyber | connected | 6am-12pm | weekly | village | 10_15 | 0_5 | filesharing | 10< | acdemic | yes | yes | good | follow hyperlink | 3.78 | 1 |
| female | 24 | messenger | better | laptop and mobile | library | connected | 12am-6am | daily | town | 10_15 | 0 to5 | news | 5_10 | acdemic | no | yes | good | friends and teachers | 3.82 | 1 |
| male | 24 | youtube | good | laptop and mobile | university | connected | 12am-6am | daily | town | 10_15 | 5_10 | groupstudy | 10< | acdemic | yes | yes | average | directories | 3.25 | 0 |
| female | 23 | whatsapp | good | laptop and mobile | university | connected | 12am-6am | daily | town | 0_5 | 0_5 | gaming | 5_10 | acdemic | yes | yes | average | search engine | 3 | 0 |
| female | 24 | twitter | not at all | laptop and mobile | cyber | connected | 12am-6am | weekly | town | 20_24 | 10_15 | gaming | 10< | non-academic | yes | yes | good | search engine | 3.66 | 1 |
| female | 23 | whatsapp | average | mobile | cyber | connected | 6pm-12am | daily | town | 15_20 | 5_10 | groupstudy | 5_10 | non-academic | yes | no | satisfactory | search engine | 3.84 | 1 |
| female | 24 | youtube | average | laptop and mobile | cyber | connected | 6pm-12am | daily | village | 15_20 | 5_10 | filesharing | 10< | acdemic | yes | yes | satisfactory | newspaper | 3.82 | 1 |
| female | 24 | whatsapp | better | laptop and mobile | cyber | connected | 12pm-6pm | daily | village | 10_15 | 5_10 | news | 5_10 | non-academic | no | yes | satisfactory | directories | 3.5 | 1 |
| male | 23 | youtube | better | desktop | university | not connect | 12pm-6pm | daily | village | 5_10 | 0_5 | music | 10< | non-academic | yes | yes | good | directories | 3.47 | 1 |
| fem | 24 | twitter | better | laptop and | cyber | connected | 6pm-12 | daily | town | 0_5 | 0_5 | fileshari | 5_10 | acdemic | yes | yes | not satisfac | newspaper | 3.3 | 0 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ale | | | | mobile | | | am | | | | | ng | | | | | tory | | 3 | |
| male | 23 | twitter | better | desktop | library | 144 kb-2mb | 12am-6am | daily | town | 5_10 | 0_5 | music | 10< | non-academic | no | no | good | friends and teachers | 4 | 1 |
| male | 23 | youtube | good | desktop | cyber | connected | 6am-12pm | daily | town | 20_24 | 10_15 | filesharing | 5_10 | acdemic | yes | yes | good | follow hyperlink | 3.69 | 1 |
| male | 23 | google | good | mobile | home | connected | 6am-12pm | daily | town | 5_10 | 0_5 | music | 10< | acdemic | yes | yes | good | friends and teachers | 3.85 | 1 |
| male | 23 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | news | 10< | non-academic | yes | yes | good | follow hyperlink | 3.72 | 1 |
| male | 23 | whatsapp | good | mobile | library | connected | 12am-6am | daily | town | 5_10 | 0_5 | music | 10< | non-academic | no | yes | satisfactory | directories | 3.59 | 1 |
| male | 23 | twitter | better | laptop and mobile | university | connected | 6am-12pm | daily | town | 5_10 | 0_5 | group study | 10< | acdemic | yes | yes | satisfactory | directories | 3.48 | 1 |
| male | 23 | whatsapp | better | laptop and mobile | library | connected | 12pm-6pm | daily | town | 5_10 | 0_5 | gaming | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.67 | 1 |
| male | 24 | youtube | not at all | laptop and mobile | cyber | not connect | 6am-12pm | weekly | village | 15_20 | 5_10 | group study | 5_10 | acdemic | no | no | average | follow hyperlink | 3.28 | 0 |
| male | 23 | messenger | good | mobile | cyber | connected | 6am-12pm | daily | town | 10_15 | 5_10 | social sites | 10< | non-academic | yes | yes | good | follow hyperlink | 3.77 | 1 |
| female | 23 | instagram | better | mobile | home | connected | 12pm-6pm | daily | town | 5_10 | 0_5 | shopping | 10< | acdemic | yes | yes | good | directories | 3.55 | 1 |
| male | 23 | google | better | desktop | university | connected | 12pm-6pm | daily | village | 10_15 | 0_5 | blogs | 5_10 | non-academic | yes | yes | good | newspaper | 3.58 | 1 |
| ma | 23 | twitte | bett | mobile | uni | connected | 12a | daily | tow | 20_24 | 15_20 | news | 10< | acdemi | yes | ye | satisfac | directori | 3. | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| le | | r | er | | versity | | m-6am | | n | | | | | c | | s | tory | es | 61 | |
| female | 25 | instagram | good | desktop | university | connected | 6am-12pm | daily | town | 5_10 | 0_5 | music | 10< | non-academic | yes | yes | satisfactory | search engine | 3.8 | 1 |
| male | 23 | youtube | not at all | mobile | others | not connect | 12pm-6pm | daily | village | 0_5 | 0_5 | social sites | 10< | acdemic | no | no | average | newspaper | 3.4 | 0 |
| female | 23 | twitter | good | mobile | university | connected | 12am-6am | daily | town | 10_15 | 5_10 | social sites | 10< | acdemic | yes | yes | good | friends and teachers | 3.96 | 1 |
| male | 25 | youtube | good | laptop and mobile | library | connected | 6pm-12am | daily | town | 5_10 | 0_5 | social sites | 10< | non-academic | yes | yes | good | search engine | 3.92 | 1 |
| male | 23 | whatsapp | better | laptop and mobile | others | connected | 6am-12pm | daily | town | 10_15 | 0_5 | group study | 10< | acdemic | no | yes | satisfactory | directories | 3.86 | 1 |
| female | 23 | instagram | average | desktop | library | connected | 6pm-12am | daily | remote | 10_15 | 5_10 | gaming | 5_10 | acdemic | no | no | not satisfactory | follow hyperlink | 3.27 | 0 |
| female | 24 | messenger | better | laptop and mobile | university | connected | 6am-12pm | daily | town | 5_10 | 0_5 | news | 5_10 | acdemic | yes | yes | good | directories | 3.93 | 1 |
| female | 23 | whatsapp | good | laptop and mobile | university | connected | 6pm-12am | daily | town | 5_10 | 0_5 | shopping | 10< | acdemic | yes | yes | good | newspaper | 3.56 | 1 |
| male | 23 | google | good | laptop and mobile | home | connected | 6am-12pm | daily | town | 10_15 | 5_10 | social sites | 5_10 | acdemic | no | yes | satisfactory | search engine | 3.82 | 1 |
| female | 23 | youtube | not at all | mobile | university | not connect | 6pm-12am | weekly | village | 0_5 | 0_5 | shopping | 5_10 | non-academic | yes | no | average | friends and teachers | 3.25 | 0 |
| male | 23 | messenger | good | laptop and mo | home | connected | 6pm-12a | daily | village | 10_15 | 5_10 | group study | 10< | acdemic | yes | yes | good | friends and teac | 3.99 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | bile | | | m | | | | | | | | | | | hers | | |
| male | 23 | whatsapp | better | laptop and mobile | library | connected | 6am-12pm | daily | town | 10_15 | 5_10 | file sharing | 10< | acdemic | yes | yes | good | search engine | 3.39 | 1 |
| male | 23 | messenger | better | mobile | library | connected | 6am-12pm | daily | town | 10_15 | 5_10 | social sites | 10< | non-academic | yes | yes | good | search engine | 3.78 | 1 |
| female | 23 | messenger | better | laptop and mobile | library | connected | 6am-12pm | daily | town | 15_20 | 10_15 | file sharing | 10< | acdemic | no | no | good | search engine | 3.57 | 1 |
| female | 23 | google | good | mobile | cyber | connected | 6pm-12am | daily | village | 15_20 | 5_10 | file sharing | 5_10 | non-academic | yes | yes | good | search engine | 3.55 | 1 |
| male | 23 | youtube | average | mobile | others | connected | 6pm-12am | daily | town | 15_20 | 5_10 | file sharing | 5_10 | acdemic | no | yes | satisfactory | search engine | 3.28 | 1 |
| male | 23 | youtube | average | laptop and mobile | library | connected | 6am-12pm | daily | town | 15_20 | 5_10 | social sites | 5_10 | acdemic | yes | yes | good | friends and teachers | 3.9 | 1 |
| male | 23 | whatsapp | average | laptop and mobile | home | connected | 6pm-12am | daily | village | 15_20 | 5_10 | social sites | 5_10 | acdemic | yes | yes | good | follow hyperlink | 3.65 | 1 |
| female | 23 | instagram | good | desktop | home | connected | 6pm-12am | daily | town | 15_20 | 5_10 | news | 10< | acdemic | yes | yes | good | follow hyperlink | 3.77 | 1 |
| female | 23 | youtube | good | laptop and mobile | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | music | 5_10 | acdemic | yes | yes | satisfactory | newspaper | 3.68 | 1 |
| male | 23 | youtube | good | mobile | library | connected | 6am-12pm | daily | town | 5_10 | 5_10 | shopping | 10< | non-academic | yes | yes | satisfactory | directories | 3.54 | 1 |
| male | 23 | youtube | good | desktop | cyber | connected | 6pm-12am | daily | village | 15_20 | 5_10 | blogs | 5_10 | acdemic | yes | yes | average | directories | 3.58 | 1 |
| fem | 23 | google | good | laptop and | other | connected | 6pm-12 | daily | town | 15_20 | 0_5 | music | 10< | acdemic | no | no | not satisfac | directories | 3.2 | 0 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ale | | | | mobile | s | | am | | | | | | | | | | tory | | 9 | |
| male | 23 | google | better | mobile | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | social sites | 5_10 | acdemic | yes | no | average | follow hyperlink | 3 | 0 |
| male | 23 | google | better | laptop and mobile | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | social sites | 5_10 | non-academic | yes | yes | satisfactory | search engine | 3.95 | 1 |
| male | 23 | google | better | laptop and mobile | home | connected | 12pm-6pm | daily | town | 5_10 | 0_5 | social sites | 5_10 | acdemic | yes | yes | satisfactory | friends and teachers | 3.81 | 1 |
| male | 23 | google | better | mobile | home | connected | 6pm-12am | daily | town | 5_10 | 0_5 | group study | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.59 | 1 |
| male | 23 | youtube | better | mobile | library | connected | 6pm-12am | daily | village | 5_10 | 0_5 | group study | 10< | acdemic | yes | yes | good | friends and teachers | 3.48 | 1 |
| male | 23 | messenger | better | mobile | home | connected | 6pm-12am | daily | town | 5_10 | 0_5 | social sites | 5_10 | non-academic | no | no | good | search engine | 3.28 | 0 |
| male | 23 | google | better | laptop and mobile | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | group study | 5_10 | acdemic | yes | yes | good | search engine | 3.77 | 1 |
| female | 23 | google | good | mobile | cyber | connected | 6pm-12am | daily | town | 10_15 | 10_15 | group study | 10< | acdemic | yes | yes | good | directories | 3.91 | 1 |
| female | 24 | google | good | mobile | home | connected | 6pm-12am | daily | town | 15_20 | 5_10 | news | 10< | non-academic | yes | yes | good | friends and teachers | 3.37 | 1 |
| female | 24 | google | good | laptop and mobile | university | connected | 6pm-12am | daily | town | 15_20 | 10_15 | file sharing | 10< | acdemic | no | yes | good | friends and teachers | 3.78 | 1 |
| male | 24 | google | good | mobile | library | connected | 6pm-12am | daily | town | 15_20 | 5_10 | group study | 10< | non-academic | yes | yes | good | friends and teachers | 3.5 | 1 |
| fem | 25 | google | good | laptop and | librar | connected | 6pm-12 | daily | town | 10_15 | 5_10 | group stu | 10< | non-aca | yes | no | good | friends and | 3.4 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ale | | | | mobile | y | | am | | | | | dy | | demic | | | | teachers | 6 | |
| female | 25 | google | good | mobile | library | connected | 6pm-12am | daily | town | 20_24 | 10_15 | group study | 5_10 | acdemic | no | yes | satisfactory | follow hyperlink | 3.59 | 1 |
| male | 25 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | gaming | 5_10 | non-academic | no | yes | satisfactory | search engine | 3.62 | 1 |
| female | 25 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | group study | 10< | non-academic | yes | no | satisfactory | search engine | 3.55 | 1 |
| female | 25 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 15_20 | 10_15 | social sites | 5_10 | acdemic | yes | no | satisfactory | search engine | 3.52 | 1 |
| male | 25 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | blogs | 5_10 | non-academic | no | yes | average | friends and teachers | 3.3 | 0 |
| male | 25 | youtube | good | laptop and mobile | university | connected | 12am-6am | daily | town | 15_20 | 10_15 | group study | 10< | non-academic | yes | yes | satisfactory | follow hyperlink | 3.48 | 1 |
| male | 24 | youtube | good | laptop and mobile | university | connected | 6am-12pm | daily | town | 10_15 | 5_10 | gaming | 5_10 | non-academic | yes | no | good | friends and teachers | 3.79 | 1 |
| female | 23 | google | good | laptop and mobile | university | connected | 12am-6am | daily | town | 20_24 | 10_15 | group study | 10< | non-academic | yes | yes | good | follow hyperlink | 3.76 | 1 |
| female | 24 | google | better | mobile | university | connected | 6pm-12am | daily | town | 10_15 | 5_10 | group study | 10< | acdemic | yes | yes | good | follow hyperlink | 3.56 | 1 |
| male | 24 | google | better | laptop and mobile | university | not connect | 6pm-12am | weekly | village | 15_20 | 10_15 | news | 5_10 | non-academic | no | yes | not satisfactory | search engine | 3.2 | 0 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| female | 23 | whatsapp | better | laptop and mobile | university | connected | 12pm-6pm | daily | village | 10_15 | 5_10 | gaming | 5_10 | acdemic | yes | yes | good | search engine | 3.85 | 1 |
| male | 23 | twitter | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | group study | 10< | non-academic | yes | no | good | follow hyperlink | 3.88 | 1 |
| female | 23 | google | better | laptop and mobile | university | connected | 6pm-12am | daily | village | 10_15 | 5_10 | social sites | 10< | non-academic | yes | yes | good | friends and teachers | 3.67 | 1 |
| female | 23 | messenger | better | mobile | university | connected | 6am-12pm | daily | village | 15_20 | 10_15 | shopping | 10< | acdemic | yes | yes | good | friends and teachers | 3.53 | 1 |
| female | 23 | google | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | social sites | 10< | acdemic | no | yes | good | friends and teachers | 3.45 | 1 |
| male | 23 | instagram | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 20_24 | 10_15 | group study | 5_10 | acdemic | no | yes | satisfactory | search engine | 3.52 | 1 |
| male | 23 | youtube | better | laptop and mobile | university | not connect | 12am-6am | daily | village | 10_15 | 5_10 | news | 5_10 | acdemic | yes | no | average | search engine | 3.1 | 0 |
| male | 23 | youtube | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | file sharing | 10< | acdemic | yes | yes | good | search engine | 3.59 | 1 |
| male | 23 | youtube | better | mobile | home | connected | 12pm-6pm | daily | town | 15_20 | 5_10 | group study | 10< | acdemic | yes | yes | good | friends and teachers | 3.68 | 1 |
| male | 23 | google | good | laptop and mobile | home | connected | 6am-12pm | daily | town | 15_20 | 10_15 | file sharing | 10< | acdemic | yes | yes | good | friends and teachers | 3.73 | 1 |
| male | 23 | google | better | laptop and mobile | university | connected | 6pm-12am | daily | village | 20_24 | 10_15 | file sharing | 5_10 | acdemic | yes | yes | satisfactory | search engine | 3.89 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 23 | google | good | mobile | university | connected | 6pm-12am | daily | town | 15_20 | 10_15 | group study | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.54 | 1 |
| female | 23 | google | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 15_20 | 10_15 | gaming | 10< | acdemic | yes | yes | satisfactory | search engine | 3.4 | 1 |
| female | 23 | google | good | mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | file sharing | 10< | acdemic | yes | yes | satisfactory | friends and teachers | 3.66 | 1 |
| male | 23 | google | better | laptop and mobile | university | connected | 6pm-12am | weekly | town | 10_15 | 5_10 | news | 5_10 | acdemic | yes | yes | average | search engine | 2.99 | 0 |
| male | 24 | google | good | mobile | university | connected | 6am-12pm | daily | town | 15_20 | 5_10 | file sharing | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.95 | 1 |
| male | 24 | google | better | mobile | home | connected | 12am-6am | daily | town | 10_15 | 0_5 | file sharing | 10< | non-academic | no | no | satisfactory | follow hyperlink | 3.86 | 1 |
| female | 24 | youtube | good | mobile | home | connected | 12pm-6pm | daily | village | 10_15 | 0_5 | news | 10< | acdemic | no | no | satisfactory | search engine | 3.75 | 1 |
| female | 24 | youtube | better | mobile | university | connected | 12pm-6pm | daily | town | 15_20 | 0_5 | group study | 10< | acdemic | yes | yes | good | search engine | 3.9 | 1 |
| female | 24 | youtube | good | laptop and mobile | university | not connect | 12am-6am | daily | village | 0_5 | 0_5 | group study | 5_10 | acdemic | yes | yes | not satisfactory | follow hyperlink | 2.8 | 0 |
| female | 24 | google | better | mobile | university | connected | 12am-6am | daily | town | 15_20 | 10_15 | gaming | 5_10 | acdemic | yes | yes | good | follow hyperlink | 3.66 | 1 |
| male | 25 | youtube | better | laptop and mobile | universi | connected | 12am-6am | daily | town | 15_20 | 10_15 | group study | 10< | acdemic | yes | yes | good | search engine | 3.98 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | ty | | | | | | | | | | | | | | | |
| female | 25 | youtube | better | mobile | library | connected | 6am-12pm | daily | town | 5_10 | 0_5 | news | 10< | non-academic | no | yes | average | follow hyperlink | 3.3 | 0 |
| female | 25 | google | better | laptop and mobile | university | connected | 6am-12pm | daily | town | 15_20 | 10_15 | file sharing | 10< | acdemic | yes | yes | good | search engine | 3.59 | 1 |
| male | 25 | google | better | mobile | library | connected | 6am-12pm | daily | town | 10_15 | 0_5 | file sharing | 5_10 | acdemic | yes | yes | good | friends and teachers | 3.93 | 1 |
| female | 25 | google | better | laptop and mobile | university | connected | 6am-12pm | daily | town | rarely | rarely | file sharing | 5_10 | acdemic | yes | yes | average | friends and teachers | 3.23 | 0 |
| female | 25 | youtube | good | mobile | home | connected | 6pm-12am | daily | town | 5_10 | 0_5 | file sharing | 10< | acdemic | yes | no | good | friends and teachers | 3.65 | 1 |
| female | 25 | youtube | better | laptop and mobile | university | connected | 6pm-12am | daily | town | 10_15 | 5_10 | news | 10< | non-academic | yes | yes | good | follow hyperlink | 3.81 | 1 |
| female | 25 | youtube | better | laptop and mobile | library | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | file sharing | 5_10 | acdemic | no | yes | good | follow hyperlink | 3.72 | 1 |
| male | 25 | youtube | better | mobile | university | connected | 12pm-6pm | daily | town | 5_10 | 0_5 | news | 5_10 | acdemic | yes | no | good | search engine | 3.55 | 1 |
| female | 25 | google | good | laptop and mobile | library | connected | 6pm-12am | daily | town | 10_15 | 5_10 | social sites | 10< | acdemic | yes | yes | satisfactory | friends and teachers | 3.48 | 1 |
| male | 25 | google | good | laptop and mobile | university | connected | 12am-6am | daily | town | 5_10 | 0_5 | social sites | 5_10 | acdemic | no | yes | satisfactory | friends and teachers | 3.93 | 1 |
| fem | 25 | youtube | good | mobile | librar | connected | 12am- | daily | town | 10_15 | 0_5 | group stu | 5_10 | acdemic | yes | yes | good | search engi | 3.6 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ale | | | | | y | | 6am | | | | | dy | | | | | | ne | 7 | |
| female | 25 | google | better | laptop and mobile | university | not connect | 12am-6am | daily | village | 5_10 | 0_5 | social sites | 10< | acdemic | yes | no | not satisfactory | search engine | 3.25 | 0 |
| female | 24 | youtube | average | mobile | university | connected | 6pm-12am | daily | town | 15_20 | 10_15 | group study | 10< | acdemic | yes | yes | good | friends and teachers | 3.84 | 1 |
| female | 24 | twitter | better | laptop and mobile | university | connected | 6pm-12am | daily | town | 10_15 | 0_5 | group study | 5_10 | acdemic | yes | yes | good | search engine | 3.95 | 1 |
| female | 23 | google | better | mobile | university | connected | 12am-6am | weekly | town | 5_10 | 0_5 | news | 5_10 | non-academic | yes | no | not satisfactory | friends and teachers | 3 | 0 |
| female | 24 | google | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 15_20 | 10_15 | file sharing | 10< | acdemic | no | yes | good | search engine | 3.54 | 1 |
| female | 24 | messenger | better | mobile | university | connected | 12am-6am | daily | town | 10_15 | 0_5 | news | 10< | acdemic | no | yes | good | friends and teachers | 3.78 | 1 |
| female | 24 | google | better | mobile | university | connected | 12pm-6pm | daily | town | 5_10 | 0_5 | news | 10< | non-academic | no | yes | good | search engine | 3.88 | 1 |
| male | 23 | google | good | mobile | home | connected | 12pm-6pm | daily | town | 15_20 | 10_15 | news | 5_10 | acdemic | yes | yes | satisfactory | friends and teachers | 3.81 | 1 |
| female | 24 | google | good | laptop and mobile | library | connected | 12am-6am | daily | town | 10_15 | 0_5 | group study | 5_10 | acdemic | yes | yes | satisfactory | search engine | 3.69 | 1 |
| female | 24 | google | good | laptop and mobile | university | connected | 6pm-12am | daily | town | 15_20 | 10_15 | gaming | 10< | acdemic | no | yes | satisfactory | friends and teachers | 3.95 | 1 |
| ma | 24 | youtub | good | mobile | uni | connected | 6pm- | daily | tow | 5_10 | 0_5 | filesh | 10< | non- | yes | no | good | friends | 3. | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| le | | e | | | versity | | 12am | | n | | | aring | | academic | | | | and teachers | 52 | |
| male | 24 | youtube | good | mobile | library | not connect | 6pm-12am | daily | vilage | 10_15 | 0_5 | file sharing | 10< | acdemic | yes | no | average | search engine | 3.4 | 0 |
| female | 23 | instagram | better | laptop and mobile | library | not connect | 6am-12pm | daily | town | 5_10 | 0_5 | file sharing | 10< | non-academic | yes | yes | good | friends and teachers | 3.48 | 1 |
| male | 23 | messenger | good | laptop and mobile | library | connected | 6am-12pm | daily | town | 20_24 | 10_15 | news | 5_10 | acdemic | yes | no | satisfactory | friends and teachers | 3.3 | 0 |
| male | 23 | twitter | average | laptop | library | connected | 6pm-12am | daily | town | 10_15 | 5_10 | file sharing | 10< | non-academic | yes | yes | not satisfactory | follow hyperlink | 3.93 | 1 |
| female | 23 | twitter | good | laptop and mobile | cyber | connected | 12am-6am | daily | vilage | 15_20 | 0 to5 | news | 5_10 | acdemic | yes | no | good | directories | 2.99 | 0 |
| male | 23 | twitter | not at all | laptop and mobile | university | connected | 6am-12pm | daily | vilage | 10_15 | 0 to5 | music | 5_10 | acdemic | yes | yes | average | newspaper | 3.55 | 1 |
| male | 23 | messenger | better | laptop and mobile | cyber | connected | 6pm-12am | weekly | vilage | 10_15 | 5_10 | news | 5_10 | acdemic | no | yes | satisfactory | directories | 3.69 | 1 |
| male | 24 | gmail | good | laptop and mobile | university | connected | 12am-6am | weekly | vilage | 0_5 | 0_5 | group study | 5_10 | acdemic | yes | yes | good | directories | 3.81 | 1 |
| female | 24 | youtube | good | laptop and mobile | home | connected | 6am-12pm | daily | vilage | 20_24 | 10_15 | social sites | 10< | acdemic | no | yes | satisfactory | search engine | 3.54 | 1 |
| male | 24 | whatsapp | average | laptop | library | connected | 6pm-12am | daily | vilage | 15_20 | 5_10 | social sites | 5_10 | acdemic | yes | yes | good | directories | 3.59 | 1 |
| male | 23 | gmail | average | laptop | cyber | connected | 12am-6am | daily | town | 15_20 | 5_10 | shopping | 5_10 | acdemic | yes | yes | good | search engine | 3.76 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| male | 23 | google | average | desktop | cyber | connected | 6pm-12am | daily | town | 10_15 | 5_10 | gaming | 10< | non-academic | yes | yes | good | directories | 3.52 | 1 |
| male | 23 | google | good | desktop | cyber | connected | 12pm-6pm | daily | town | 5_10 | 0_5 | social sites | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.66 | 1 |
| female | 23 | google | average | mobile | university | connected | 12pm-6pm | daily | town | 0_5 | 0_5 | social sites | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.85 | 1 |
| male | 24 | google | not at all | mobile | home | not connect | 12pm-6pm | daily | village | 5_10 | rarely | music | 10< | acdemic | no | no | not satisfactory | follow hyperlink | 3.29 | 0 |
| male | 23 | instagram | better | mobile | home | not connect | 12am-6am | daily | town | 20_24 | 10_15 | file sharing | 5_10 | non-academic | no | yes | satisfactory | directories | 3.53 | 1 |
| male | 23 | instagram | good | desktop | cyber | connected | 6am-12pm | daily | town | 15_20 | 5_10 | file sharing | 10< | non-academic | yes | yes | satisfactory | newspaper | 3.86 | 1 |
| female | 23 | messenger | average | desktop | cyber | connected | 12pm-6pm | daily | town | 10_15 | 0 to5 | group study | 10< | non-academic | yes | yes | good | directories | 3.63 | 1 |
| male | 23 | twitter | not at all | desktop | university | connected | 12am-6am | daily | remote | 5_10 | 0_5 | group study | 10< | acdemic | yes | no | average | search engine | 3.15 | 0 |
| male | 23 | youtube | average | laptop and mobile | cyber | connected | 6pm-12am | daily | town | 0_5 | 0_5 | file sharing | 5_10 | non-academic | yes | yes | good | newspaper | 3.84 | 1 |
| female | 23 | whatsapp | good | desktop | university | connected | 6am-12pm | daily | village | 15_20 | 0_5 | group study | 5_10 | acdemic | yes | yes | satisfactory | search engine | 3.76 | 1 |
| female | 23 | whatsapp | better | laptop and mobile | university | connected | 12am-6am | daily | town | 15_20 | 5_10 | group study | 10< | non-academic | no | yes | satisfactory | search engine | 3.7 | 1 |
| fema | 23 | google | good | laptop and mo | univer | connected | 12pm-6p | daily | village | 10_15 | 0_5 | group study | 5_10 | acdemic | yes | yes | satisfactory | directories | 3.62 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| le | | | | bile | sity | | m | | | | | | | | | | | | | |
| female | 23 | gmail | average | laptop and mobile | university | connected | 6am-12pm | daily | town | 10_15 | 0_5 | shopping | 10< | acdemic | no | yes | good | search engine | 3.55 | 1 |
| male | 24 | gmail | good | desktop | library | connected | 12am-6am | daily | village | 15_20 | rarely | social sites | 5_10 | acdemic | yes | yes | satisfactory | search engine | 3.33 | 0 |
| male | 23 | messenger | better | mobile | home | connected | 6pm-12am | daily | town | 5_10 | rarely | social sites | 5_10 | non-academic | yes | no | good | search engine | 3 | 0 |
| male | 23 | google | better | laptop and mobile | home | connected | 6pm-12am | daily | town | 10_15 | 5_10 | group study | 5_10 | acdemic | yes | yes | good | search engine | 3.58 | 1 |
| female | 23 | google | good | mobile | cyber | connected | 6pm-12am | daily | town | 10_15 | 10_15 | group study | 10< | acdemic | yes | yes | good | directories | 3.73 | 1 |
| female | 24 | google | good | mobile | home | connected | 6pm-12am | daily | town | 15_20 | 5_10 | news | 10< | non-academic | yes | yes | good | friends and teachers | 3.88 | 1 |
| female | 24 | google | good | laptop and mobile | university | connected | 6pm-12am | daily | town | 15_20 | 10_15 | file sharing | 10< | acdemic | no | yes | good | friends and teachers | 3.85 | 1 |
| male | 24 | google | good | mobile | library | connected | 6pm-12am | daily | town | 15_20 | 5_10 | group study | 10< | non-academic | no | yes | good | friends and teachers | 3.38 | 1 |
| female | 25 | google | good | laptop and mobile | library | connected | 6pm-12am | daily | town | 10_15 | 5_10 | group study | 10< | non-academic | yes | no | good | friends and teachers | 3.54 | 1 |
| female | 25 | google | good | mobile | library | connected | 6pm-12am | daily | town | 20_24 | 10_15 | group study | 5_10 | acdemic | no | yes | satisfactory | follow hyperlink | 3.62 | 1 |
| male | 26 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | gaming | 5_10 | non-academic | yes | yes | satisfactory | search engine | 3.97 | 1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| female | 25 | youtube | good | laptop and mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | group study | 10< | non-academic | yes | no | satisfactory | search engine | 3.89 | 1 |
| male | 23 | google | good | mobile | university | connected | 6pm-12am | daily | town | 15_20 | 10_15 | group study | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.57 | 1 |
| female | 23 | google | better | laptop and mobile | university | connected | 12pm-6pm | daily | town | 15_20 | 10_15 | gaming | 10< | acdemic | no | yes | satisfactory | search engine | 3.83 | 1 |
| female | 23 | google | good | mobile | university | connected | 12pm-6pm | daily | town | 10_15 | 5_10 | file sharing | 10< | acdemic | yes | yes | satisfactory | friends and teachers | 3.78 | 1 |
| male | 23 | google | better | laptop and mobile | university | connected | 6pm-12am | weekly | town | 10_15 | 5_10 | news | 5_10 | acdemic | no | yes | average | search engine | 3.34 | 0 |
| male | 24 | google | good | mobile | university | connected | 6am-12pm | daily | town | 15_20 | 5_10 | file sharing | 5_10 | non-academic | yes | yes | satisfactory | friends and teachers | 3.57 | 1 |
| male | 24 | google | better | mobile | home | connected | 12am-6am | daily | town | 10_15 | 0_5 | file sharing | 10< | non-academic | yes | no | satisfactory | follow hyperlink | 3.68 | 1 |
| female | 24 | youtube | good | mobile | home | connected | 12pm-6pm | daily | village | 10_15 | 0_5 | news | 10< | acdemic | yes | no | satisfactory | search engine | 3.88 | 1 |
| female | 23 | messenger | average | mobile | university | connected | 6am-12pm | daily | remote | 0_5 | 0 t0 5 | music | 5_10 | non-academic | yes | no | average | friends and teachers | 2.99 | 0 |
| male | 23 | youtube | good | mobile | cyber | connected | 6am-12pm | weekly | village | 10_15 | 0_5 | file sharing | 10< | acdemic | yes | yes | good | follow hyperlink | 3.7 | 1 |
| female | 23 | whatsapp | better | laptop and mobile | university | connected | 12am-6am | monthly | town | 15_20 | 10_15 | file sharing | 10< | acdemic | no | yes | good | directories | 3.89 | 1 |