ሀዋሳ ዩኒቨርሲቲ
HAWASSA UNIVERSITY

INSTITUTE OF TECHNOLOGY

FACULTY OF INFORMATICS

MSC PROGRAM IN COMPUTER SCIENCE

Academic Performance Prediction Model for Teacher's Training Colleges Using Machine learning Approach

A Thesis Submitted to Hawassa University, Institute of Technology, Faculty of Informatics, Department of Computer science in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science

By: Firehiwot Getachew

Advisor: Dr. Dereje Teferi

Co. Advisor: Ato Mulat Sheferaw

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY

EXAMINERS' APPROVAL SHEET

We, the undersigned, members of the Board of Examiners of the final open defense by **Firehiwot Getachew** and evaluated his/her thesis entitled titled "**Academic Performance Prediction Model for Teachers' Training Colleges Using Machine Learning Approach**", and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science.

______________________	________________	______________________
Name of the Chairperson	Signature	Date
Dreje Teferi (phD)	________________	April, 2020
Name of Major Advisor	Signature	Date
______________________	________________	______________________
Name of Internal Examiner	Signature	Date
Wondwosen Mulugeta (PhD)	________________	April, 2020
Name of External Examiner	Signature	Date
______________________	________________	______________________
SGS Approval	Signature	Date

Final approval and acceptance of thesis is contingent upon the submission of the final copy of the thesis to the school of graduate studies (SGS) through the department/School Graduate committee (DGC/SGS) of the candidate's department.

Stamp of SGS

Date______________________

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY

ADVISORS' APPROVAL SHEET

This is to certify that the thesis entitled titled "**Academic Performance Prediction Model for Teachers' Training Colleges Using Machine Learning Approach**", Submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science, the graduate program of the Department of Computer Science, has been carried out by **Firehiwot Getachew**. Id. No: PGCSR/008/10 under my /our supervision. Therefore, I/WE recommend that the student has fulfilled the requirements and hence hereby can submit the thesis to the department.

Dreje Teferi (phD)	[signature]	April, 2020
Name of Major Advisor	Signature	Date
Mulat Sheferaw		
Name of Co-Advisor	Signature	Date

DECLARATION

I, Firehiwot Getachew, declare that this thesis work titled "**Academic Performance Prediction Model for Teachers' Training Colleges Using Machine Learning Approach**" along with the experiment and the result presented in it is my original work. This work is not submitted in whole or in part for any publication and references are made to the work of others. Other sources are acknowledged by citation by giving an explicit reference.

I have undertaken the study independently with the guidance and support of my research advisor.

Date: ________________________

Signature____________________

Firehiwot Getachew

This thesis has been done under my guidance and submitted for examination with my approval as a university advisor.

Date: __April, 2020________

Signature____________________

Dr. Dereje Teferi (PhD)

Place and Date of submission:

Hawassa University

June, 2020

Abstract

Data mining is the process of extracting novel or previously unknown information from a large amount of data. The purpose of this study is to develop an academic performance prediction model and identifying the factors that affect academic performance of college student using data mining techniques. The data used for this study are 1023 active students from HCTE in 2018/19 academic year. For the consumption of this research, both primary and secondary data was used. Primary data such as age, gender, previous high school, department, library usage, study hours, sport interest, mother education, father education, time spent in social media, family support and economic status of family is collected by means of questionnaire. Secondary data was obtained from the HCTE registrar office.

The prediction model was developed using multilayer perceptron (MLP) classification algorithm, Naive Bayes and J48 and correlation based feature selection (CFS) is applied to identify the predictive attributes of academic performance. Finally, Multilayer perceptron, Naive Bayes and J48 is compared using the same dataset. According to the result of the experiments, Multilayer perceptron using all attributes with test method of 10-fold cross validation and accuracy 60.6% gives better result compared to Naive Bayes, J48 and MLP after applying attribute selection. The study findings also showed that sex of the student, total courses credit hours taken by the students, study hours, assignment performance and library usage of the students are identified as a significant factor affecting academic performance. WEKA 3.8.1 tool was used for data mining process.

Acknowledgements

I am thankful to all those who, in one way or another, contributed to the successful completion of this research. My sincere gratitude goes to my Advisor Dr. DerejeTeferi, for his review of my work as well his encouragement and essential guidance on this thesis.

I would like to thank the staff and all the students of Hawassa College of teacher education, for completing the research questionnaire and providing academic related valid information in this regard.

Finally, for my kids Mama and Nishu, sorry for being even grumpier than normal! And for my Husband Gech, thanks for all your support, it was a great relief to know that you were willing to provide management of our household activities while I completed my work.

May God bless all!

Firehiwot Getachew

February 2020

Table of Contents

List of Abbreviations

AP	Academic Performance
AREF	Attribute Relation File Format
BFS	Best First Search
CFS	Correlation Based Feature Selection
CGPA	Commutative Grade Point Average
EGSECE	Ethiopian General School Education Certificate Examination
GPA	Grade Point Average
HCTE	Hawassa College of Teacher Education
LR	Learning Rate
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
WEKA	Waikato Environment for Knowledge Analysis

List of Figures

List of Tables

Chapter One

1. Introduction

Prediction is a statement that forecast the occurrence of a future event. Most often the event to be predicted is in the future, but predictive modeling can be applied to any type of unknown event, regardless of when it occurred. In machine learning, classification is identifying to which of a set of categories belongs, on the basis of a training set of data containing instances whose category membership is known. An algorithm that implements classification is known as a classifier. Classifier performs mathematical function implemented by a classification algorithm, which maps input data to a category (Ramageri, 2011).Prediction models can use one or more classifiers in trying to determine the class of a set of data. This thesis proposes a student academic performance (AP) prediction model which can be used by teachers' training colleges in Ethiopia.

Data mining is the extraction of hidden predictive information from large databases. It is a powerful new technology with great potential to help institutions and companies focus on the most important information in their data warehouses.

Mining of data in educational environment is called educational data mining. It is an emerging research field and this area of research is gaining popularity because of its potentials to educational institutes. Educational Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process (El-Halees, 2009) and AP of students.

A survey that target university students and collect multiple personal, social, and academic data related to them reveal that the student's performance is not totally dependent on their academic efforts, in spite, there are many other factors that have equal to greater influences (Saa, 2016). This study can help universities to perform data mining tasks on their students' data regularly to find out interesting results and patterns which can help both the university as well as the students in many ways.

A study conducted on predicting first year university students' academic success that involves 3301 students from Adama University (Olani, 2017). In this study, preparatory class average score, university entrance exam results and aptitude scores were used as predictors of university first semester GPA. The result shows that the three variables in combination accounted for 17 percent of the variance in students' university first semester GPA.

Furthermore, preparatory class average score was found to be the only significant predictor of university first semester GPA.

On the other hand, a study conducted by (Yoseph, 2013) on students who joined Kotobe University College in 2011/12 academic year to assess the degree to which preparatory class average score and university entrance exam result predict first year college performance. The results indicated that preparatory class average score and university entrance exam result appeared to be valid predictors of first year college CGPA and jointly accounted for 33.70 percent of the variation in college performance. The Preparatory class average score was found to be a more valid predictor of first year college CGPA than university entrance exam result.

A researcher (Yigermal, 2017) also conducted a study to investigate the determinant factors affecting the AP of regular undergraduate students of ArbaMinch University (AMU) Chamo campus students by Moges Endalemaw. This study used Pearson product moment correlation statistical tool and econometrics data analysis (OLS regression) method to identify relationship between gender difference, university entrance exam, studying hours and AP (CGPA). This study is conducted on 100 students using both primary and secondary data. The result obtained by this study verifies that there is significant relationship between students former academic back ground, studying hours, and student’s behavior on taking of alcoholic drug and chat on AP of students.

1.1. Statement of the problem

Teachers training colleges use high school grade point average, Ethiopian general education certificate examination (EGSECE), and college entrance exam scores as major criteria to admit students.

In spite of the fact that there are many students that obtain high scores in both college entry qualifications and college CGPA, there are also a significant number of students that score high results in entry qualifications, but incapable of getting pass marks in teacher training colleges and vice verse. This shows that the success of AP of students is dependent on factors other than their previous academic result. However, entry qualification may be one among many other factors that predict future academic achievement there are also many other factors that should be investigated and have equal to greater influences on students AP. In Ethiopia, there is only small number of research evidences produced on this topic specifically focused on using personal, social and academic factors as a predictor of College student’s AP. The predictive validity of personal, environmental, psychological, social and academic factors should be investigated and evaluated exhaustively to identify the attributes that contributes for AP of college student.

The objective of this research is therefore to produce AP prediction model that can be used by teacher training colleges and identify relationship between students' academic, personal and social factors, with the college students' academic performance.

1.2. Research Question

The following research questions will be answered in this study:

1. Which attributes contribute important to the prediction of college students AP?
2. How can college students AP be predicted using machine learning tools?

1.3. Objectives

1.3.1. General Objectives

The general objective of this research is to develop academic performance prediction model using multilayer perceptron algorithm.

1.3.2. Specific Objectives

The following specific objectives are formulated to achieve the general objectives of this research

- Review conceptual and empirical literature to understand the research area and identify research gaps
- Identify most determinant factor affecting AP of students.
- Produce AP prediction model using multilayer perceptron algorithm.
- Evaluate the performance of the model using model evaluation metrics.
- Compare the performance of the model by MLP with model developed by J48 Decision tree and Naïve Bayes algorithms.

1.4. Scope of the thesis

The study is conducted at Hawassa College of teacher education. Using a sample of second and third year students selected from all departments at HCTE. The content covered in this thesis were student related factors such as socio-economic status, school background and academic related factors which are considered as academic performance predictors of college students. This study involved only active students within the period of academic year 2018/2019.

1.5. Significance of the thesis

The result of this thesis can be valuable in many ways. First, the findings may guide the Ministry of Education in identifying AP predictors of students attending teacher training colleges and it may help them to review college admission criteria. Second, the finding might helps college officials, teachers and students of teacher training colleges to identify those students having high risk of being dismissed or dropped out from the college. Further, the findings might guide the teachers' training colleges to enhance their efforts in guiding the students encounter the problem. It also helps students with high risk of getting poor result in course to work hard to decrease the risk.

1.6. Organization of the Study

The research report is organized into five Chapters. The first Chapter presented the Introduction, the statement of the problem and research questions, significance of the study, scope of the study and organization of the thesis.

In Chapter two theoretical foundations and review of related study is presented. This chapter covers important issues related to the academic performance and data mining technology concepts and techniques.

Chapter three describes the research design. It describes the study area, research design, sampling methods, sample size, methods of data collection, preprocessing. In Chapter four, model learning, best model selection and evaluation result of the model is described and finally Chapter five provides conclusion and recommendation of the study.

Chapter Two

2. Theoretical foundation and literature review

2.1. Theoretical foundation

Data Mining Definitions and Techniques

Data mining, also known as Knowledge Discovery in Database, refers to extracting or “mining" of knowledge from huge sets of data. In other words, data mining can be defined as the procedure of mining knowledge from data. The information or knowledge extracted can be used for any of the following applications: market analysis, fraud detection, customer retention, and production control and science exploration (Chen et al., 1996). The essential difference between the data mining and the traditional data analysis is that the data mining is to mine information and discover knowledge on the premise of no clear assumption.

Data mining uses already build tools to get out useful hidden patterns, trends, and predictions. Various algorithms and techniques like Classification, Clustering, Regression, Artificial Neural Networks, Association Rules, Attribute selection methods, Decision Trees, Nearest Neighbor method etc., are used for knowledge discovery from databases(Rehman, 2017)These techniques and methods in data mining are discussed as follows:

A. Classification

Classification is one of the most commonly used data mining technique, which contains a set of pre-classified or labeled examples to develop a model that can classify the population of records at large (Rehman, 2017).Most commonly used Classification approach employs decision tree or neural network-based classification algorithms. Developing the prediction model process involves data collection, preprocessing, model learning and classification. In model learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the prediction model or classification rules. If the accuracy is acceptable the rules can be applied to the new dataset. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm that encodes these parameters into a model called a classifier.

B. Artificial Neural Networks

A neural network is made up of a set of connected input and output units, input unit receive input cells performing some transformations on it and generating output for further processing. ANN can provide us a powerful pattern recognition and pattern classification. According to (Kalyani, 2016) artificial neural network can work with non-linear, complex, imprecise and noisy data. During the learning phase, the network learns by changing weights to be able to predict the correct class labels of the input data. Neural networks have a remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. ANN techniques are well suited for continuous valued inputs and outputs. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

Artificial Neural Networks depend on number of layers, number of neurons in input layer, number of neurons in hidden layer, momentum, learning rate, number of training iterations that are required to obtain the best result, transfer function used for hidden and output layer, training algorithm and learning function used(2019, June 4).

Multilayer perceptron

Multilayer perceptron algorithms (Voicu, 2008) are one of the most popular and commonly used artificial neural network structures. Multilayer perceptron is a feedforward neural network with one or more hidden layers between input and output layer. In Feed forward data flows in one direction from input to output layer. This type of network is trained with the backpropagation learning algorithm. MLPs are widely used for pattern classification, recognition, prediction and approximation. Multilayer Perceptron can solve problems which are not linearly separable (Kalyani, 2016). It is suitable for predicting a classification function which sets the example determined by the attribute values into one or more classes.

Multilayer perceptron is often applied to supervised learning problems they train on a set of input-output pairs and learn to model the correlation (or dependencies) between those inputs and outputs. Training involves adjusting the parameters, or the weights and biases, of the model in order to minimize error. Backpropagation is used to make those weigh and bias adjustments relative to the error.

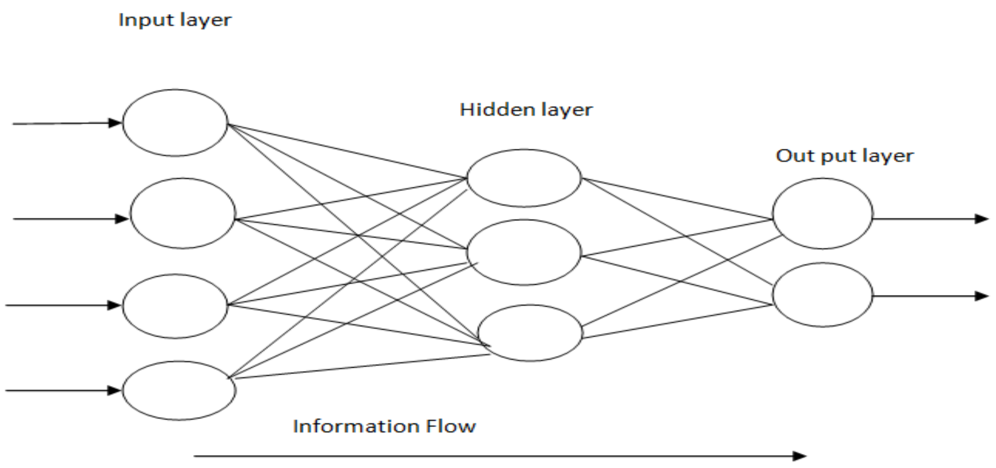


Figure 1: Feed forward neural Network

C. Decision Tree

Decision tree learning algorithm is a supervised machine learning algorithm. It is a hierarchical tree structure that attempts to predict the value of a target variable based on a sequence of decisions set of rules generated from the tree (Murphy, 2017). In short, given a data of attributes together with its classes or labels, a decision tree produces a sequence of rules that can be used to recognize the class. In a decision tree structure each node represents an attribute, each link represents a rule and each leaf represents an outcome/class or category. Decision tree uses simple representation with high prediction accuracy (Anuradha & Velmurugan, 2015).

D. Naive Bayes

Naïve Bayes is a simple probabilistic machine learning algorithm based on applying Bayes' theorem from Bayesian statistics or based on conditional probability theory with strong independence assumptions. It is used in a wide variety of classification tasks (Rehman, 2017).Typical applications of Naïve Bayes classifier include filtering spam, classifying documents, sentiment prediction; performance prediction etc. Naive Bayes classifier assumes that the presence or absence of a particular feature of a class is unrelated to the presence or absence of any other feature.

The advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters necessary for classification and it provides high prediction accuracy (Anuradha & Velmurugan, 2015). Naive Bayes classifier does not consider dependency among attributes and it may not be applicable in a many real world problems.

E. Attribute Selection

Irrelevant or partially relevant features can negatively impact model performance (Brownlee, 2016). Feature selection methods are used for identifying most relevant futures from the dataset. Applying feature selection methods helps to reduce over fitting, improve accuracy and reduce Training Time. Feature selection has two parts (Brownlee, 2016):

Feature selection is divided into two parts:

- ✓ Attribute Evaluator
- ✓ Search Method.

The attribute evaluator is the technique that evaluates each attribute based on the context of the output variable. The search method is the technique to navigate different combinations of attributes in the dataset in order to arrive on a short list of chosen features.

F. Application of data mining

In today's information world, there is a huge amount of data is available; this data is of no use until it is processed in to useful and meaningful information. Extraction of information is not the only process that can be performed on stored data in data mining. It also involves other processes such as data cleaning, data integration, data transformation, Pattern evaluation and data presentation. Once all these processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, educational data mining etc.

G. Educational data mining

Educational Data Mining (Cheng, 2017) is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those data mining methods to better understand students, and the settings which they learn in.

According to (Romero & Ventura, 2010), the primary applications of educational data mining are:

- ✓ Analysis and visualization of educational data
- ✓ Providing feedback for supporting instructors
- ✓ Providing recommendations for students
- ✓ Predicting student performance
- ✓ Modeling student

- ✓ Detecting undesirable student behavior
- ✓ Grouping students
- ✓ Analysis of students social network behavior
- ✓ Developing concept maps
- ✓ Constructing courseware
- ✓ Planning and scheduling educational activities

H. Academic performance

Academic performance is the short or long-term educational goals that are achieved by student, teacher or institution. It can also be described as learning achievement of students who meet performance standards prescribed by course work. Academic performance can also be defined as students' reporting of past semester CGPA/GPA and their expected GPA for the current semester. The grade point average or GPA is now used by most institutions as a convenient summary measure of the academic performance of their students. The GPA is a better measurement because it provides a greater insight into the relative level of performance of individuals and different group of students.

I. Factors affecting academic performance

Students academic achievement is affected by a number of factor including gender, age, teaching faculty, students schooling, father/guardian social economic status, residential area of students, medium of instructions in schools, type of school and daily study hour. Students face a number of academic challenges in college, including finding time to study, understanding course content and maintaining a high degree of motivation. Along with meeting these challenges, students often struggle to balance academic demands with work, personal responsibilities and social experiences. Major challenges are content understanding, test anxiety, balancing study time and lack of motivations.

Some of the factors that contribute the academic performance of students considered in while preparing the questionnaire are described below.

Socioeconomic Status

According to (Raychaudhuri et al., 2010),Socio-economic factors like attendance in the class, family income, and mother's and father's education, teacher-student ratio, presence of trained teacher in school, sex of student and distance of school are also affected the performance of the students. Wealth and status are an umbrella for many issues that affect performance, including suffering chronic stress, having little to no homework help available at home and having obligations around the house or at a part-time job that may curtail study time. Literacy and development issues can begin to affect children before they are even of school age, putting them at an immediate disadvantage.

Learning Facilities

Students' performance is significantly correlated with satisfaction with academic environment and the facilities of library, computer lab and etc. In the institution with regard to background variables, a positive effect of high school performance and school achievement found no statistical evidence of significant association between family income level and academic performance of the student. Students who are actively engage in the learning process are observed to have a positive correlation with the CGP (Roberts & Sampson, 2011).

A study (Raychaudhuri et al., 2010) effort from student and the proper use of the facilities provided by the institution to the student, a good match between students' learning style and are positively affect the student's performance, held the view that student performances are linked with use of library and level of their parental education. The use of the library positively affected the student performance. The academic environment is the effective variable for students and has positive relationship with fathers' education and grade level.

Proper Guidance

Students' academic accomplishments and activities, perceptions of their coping strategies and positive attributions, and background characteristics i.e., family income, parents' level of education, guidance from parents and number of negative situations in the home were indirectly related to their composite scores, through academic achievement in high school (Noble et al., 2006). The students face a lot of problems in developing positive study attitudes and study habits. Guidance is of the factor through which a student can improve his study attitudes and study habits and is directly proportional to academic achievement. The students who are properly guided by their parents have performed well in the exams. The guidance from the teacher also affects the student performance. According to (Hussain, 2003) guidance from the teacher also affects the student performance.

Student Motivation & Engagement

When a student enters a college, they become self-determined with their lives than ever before. Academically, college students are more motivated and engaged in their learning activities aiming to obtain good academic achievements. Students' academic achievement is claimed to influence by students motivation and engagement (Allen et al. 2008).

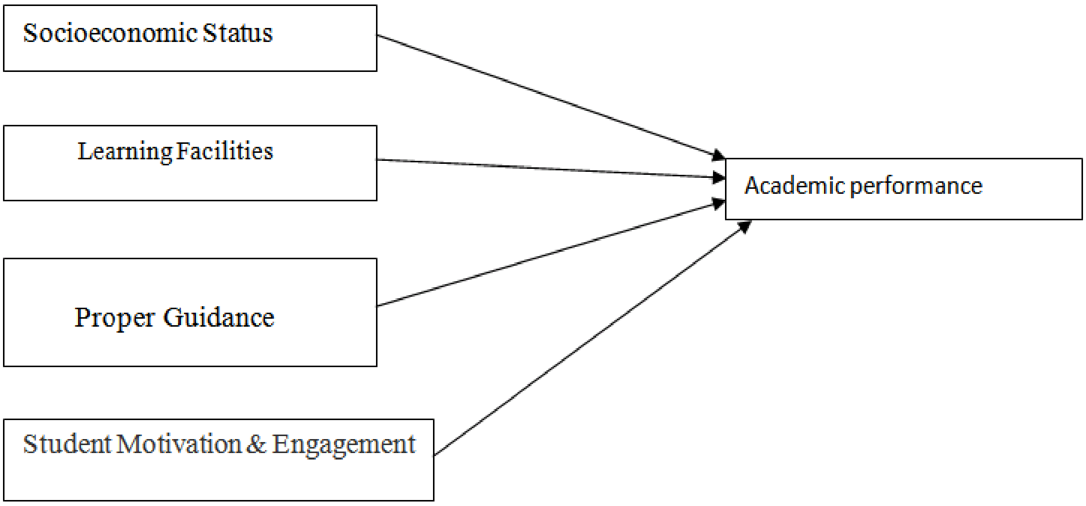


Figure 2: Factors that contribute academic performance of students

2.1. Literature review

In this section, research works related to student academic performance prediction are presented. The study by (Akessa & Dhufera, 2015) conducted to examine different factors affecting academic performance of student in higher institution for the case of Rift Valley University, Jimma campus in Ethiopia. The data is collected from 24 students selected using simple random sampling method. Student related attributes such as age, gender, religion, mother education, father education; family job status and economic status of family are used to collect data by means of questionnaire. Chi-square test of association and regressions was applied to investigate the effect of different factors on students' academic performance. According to the result of the study, teachers have a great role in fostering positive or negative attitude to achievements of students and there is significant academic performance deference between students who are enrolled in the department of Accounting, Health officer, management and Nursing. There is no identified relation between student's achievement and Sex of students.

A study conducted to predict academic performance of students based on academic and personal attributes using artificial neural network by (Sebastian, 2016) use Multilayer Perceptron Neural Network for the implementation of prediction strategy. The data was collected from grade 8 and 9 students. A dataset of 300 students were used for the evaluation. The attributes are academic details of student and personal attributes related to the personal details of student that affects the study and performance of student**.** The academic attributes selected are: Interest of study, Unit test mark, Assignment mark, Attendance s, extracurricular activities and Residence. The personal attributes selected are parent's education and family status. Association rule mining is used to identify the most important attributes. The concepts of genetic algorithm are applied to the result to obtain better performance. According to the result of the research Artificial Neural Network with k-fold cross validation gives the most accurate result than basic training method and training after association rule mining.

The research conducted by (Yohannes & Ahmed, 2018) designed an application to assist higher education institutions to predict their students' academic performance at an early stage before graduation and decrease students' dropout. This study collected data from Student Information System of Hawassa University School of Computer Science. The dataset of 134 undergraduate degree students who graduated from the university in the year 2015, 2016 and 2017 which consisted of 52, 39 and 43 students respectively and who managed to reach the final at semester eight. The performance of the students was measured based on cumulative grade point average at eighth semester. To predict academic performance Neural Network, Support Vector Regression, and Linear Regression were used by the researchers. The students' course scores for core and non-core courses from the first semester to the sixth semester are used as predictor variables for predicting the final cumulative grade point A8 upon graduation. The result of comparisons of the experiments done using the above three algorithm show that which approaches performs better than others. Support vector machine and linear regression methods performed better than Neural Network. Finally, the study used the models from support vector machine and linear regression methods for designing an application to do the prediction task.

The study conducted by (Baharin et al., 2015) discovers the most influencing factors that lead to the success and failure of academic performance. A questionnaire having21items was developed for the study and distributed to all undergraduates' students at UiTM Johor branch. Several factors influencing academic achievement of students discussed in the research paper .It consist of family related characteristics, self efficacy attributes, and university features. The Cumulative GPA of each respondent will be collected at the end of the semester. The collected data was analyzed using exploratory factor analysis and Pearson correlation of Statistical Package for Social Sciences (SPSS).The study findings showed that significant factors affecting students' academic achievement are family characteristics and university features but self-efficacy was not identified as a significant factor for affecting academic performance.

The study (Amjad et al, 2020) aimed to predict the students' academic performance based on a new dataset extracted from a student information system of a private university in the United Arab of Emirates. The extracted dataset includes 34 attributes and 56,000 records related to students' information are obtained from the system. The empirical results indicated that the Random Forest algorithm was the most appropriate data mining technique used to predict the students' academic performance. It is also revealed that the most important attributes that have a direct effect on the students' academic performance are belonged to four main categories, namely students' demographics, student previous performance information, course and instructor information, and student general information. The evidence from this study would assist the higher educational institutions by allowing the instructors and students to identify the weaknesses and factors affecting the students' performance, and act as an early warning system for predicting the students' failures and low academic performance.

Moreover, in the study conducted by (Mushtaq & Khan, 2012) on private colleges in Rawalpindi and Islamabad that attempted to apply different classification techniques to an educational dataset to identify factors affecting college students' performance. This study was done to explore the important factors that affect the students' performance in intermediate examination is linked with students'. Primary data collected from 175 students using questionnaire and secondary data is obtained from student profile on the bases of information. The result shown that Communication, learning facilities and proper guidance shows the positively affect the student performance and the family stress shows the negative impact on the student performance. The study indicated that the communication the most important attribute that affects the student performance.

The study (Affendey et al., 2010) that analyzed attributes for predicting a higher learning institution's bachelor of computer science students' academic performance. The data set for this study comprised of 2427 number of student records and 396 attributes of registered students. Naive Bayes, decision trees or function were applied on the data set and cross-validation with 10 folds was used to evaluate the prediction accuracy. The results revealed that ranking of courses that has significant impact on predicting the students' overall academic results and experiments showed that Naïve Bayes, AODE and RBFNetwork classifiers scored the highest percentage of prediction accuracy of 95.29%.

The paper (Huang & Fang, 2013) compared two neural network models Multilayer perceptron and Radial basis function with two other algorithms Multiple linear regression, Support vector machine for the prediction of academic performance of students in an introductory engineering course. Amongst the four specified classifiers used for prediction, Support vector machine outperformed the other models with an accuracy of 89% and therefore identified as the most superior model for the prediction. Another study (Arora & Saini, 2013) proposed a fuzzy probabilistic neural network model designed in MATLAB environment for predicting students' academic performance to classify them based on their similar characteristics. The findings demonstrate that fuzzy probabilistic neural network model gave an average classification accuracy of 98.56% and went further to reveal that the proposed model was achieved at a very minimal training time which is in line with the conclusion found in (Undavia et al., 2013).

The study (Amoo et al, 2018) adopted feed-forward artificial neural network to establish and identify the complex nonlinear relationship that exist between cognitive and psychological variables that influence the academic performance of secondary school students. The study considered a total of 120 students' dataset in five science subjects were extracted from the results of 2015 West African Examination Council. This data was obtained from four randomly selected secondary schools in Ibadan North Local Government Area of Oyo state. Results from the study showed that artificial neural network is efficient at clustering students into different categories according to their predicted level of performance.

Various studies have been conducted to predict AP of the students, but most of the researches are done in the universities on small size of dataset (Yadav, & Pal, 2012; Yigermal, 2017). Because many reasons studies conducted on university students may not work for college students. For example all universities in Ethiopia have dormitories but teachers' training colleges has no dormitory. So, because of above specified reason and others studies are expected to consider college students separately.

Many variables are to be considered to identify the factors affecting quality of academic achievement. The quality of teachers graduating from HCTE is decreasing from time to time, students graduating from the college are unable to pass qualification exam given by the south region. This creates a growing interest in identifying the factors predicting academic performance of students in the college. The results of this thesis may above specified gap by identifying the determinant factors affecting AP of the college students.

Chapter Three

3. Design and Development

The purpose of this study is to develop academic performance prediction model using multilayer perceptron algorithm in WEKA and identify the factors that contributes to the AP of students in teacher training colleges.

3.1. Description of the study area

Hawassa College of Teacher Education (HCTE) is one of the oldest teachers' training colleges in Ethiopia established in 1969 E.C. It is located near Hawassa tabor Preparatory School and in front of menbo site condominium. HCTE is the only governmental source of diploma holder teacher for sidama zone and Hawassa city. Students of the college come from all Hawassa city and Sidama zone high schools. To join the college students are required to have 2.00 average point in EGSECE and pass marks in high school GPA, and pass entrance exam given by the college.

3.2. Research design

In order to achieve the objectives of the study, the research follows the Design science research approach, using both qualitative and quantitative data. In design science research methodology there are six steps that should be followed by the researcher. The steps are problem identification and motivation, objective definition, design and development, evaluation, demonstration and communication (Peffers et al., 2006). Through those steps, this research examines the predictive validity of sex, department, previous school type, assignment performance, library usage, school facility usage, environmental motivation, study hours, co-curricular activity participation, sport interest, home location ,financial problem, family support, family education, time spend in social media and family economic status and develop AP prediction model using WEKA as a tool.

3.3. The Proposed Work

The proposed approach was implemented using the following strategy. First, dataset is collected and different data cleaning tasks are applied on the data and then numeric attributes of the dataset are discretized by supervised discretizer. After discretization numbers of features in the dataset were reduced to six from twenty five by CFS evaluator based on best first search method. The dataset having 6 selected attributes and the original dataset having all 25 attributes are classified by using MLP, Naïve Bayes and J48. These models are then evaluated with different performance measures like accuracy, precision, Recall, F-measure and mean absolute error. Finally, the performance of the MLP is compared with Naïve Bayes and Decision tree (J48) algorithm using above specified performance measures. The architectural diagram of proposed work is shown in Fig 3.

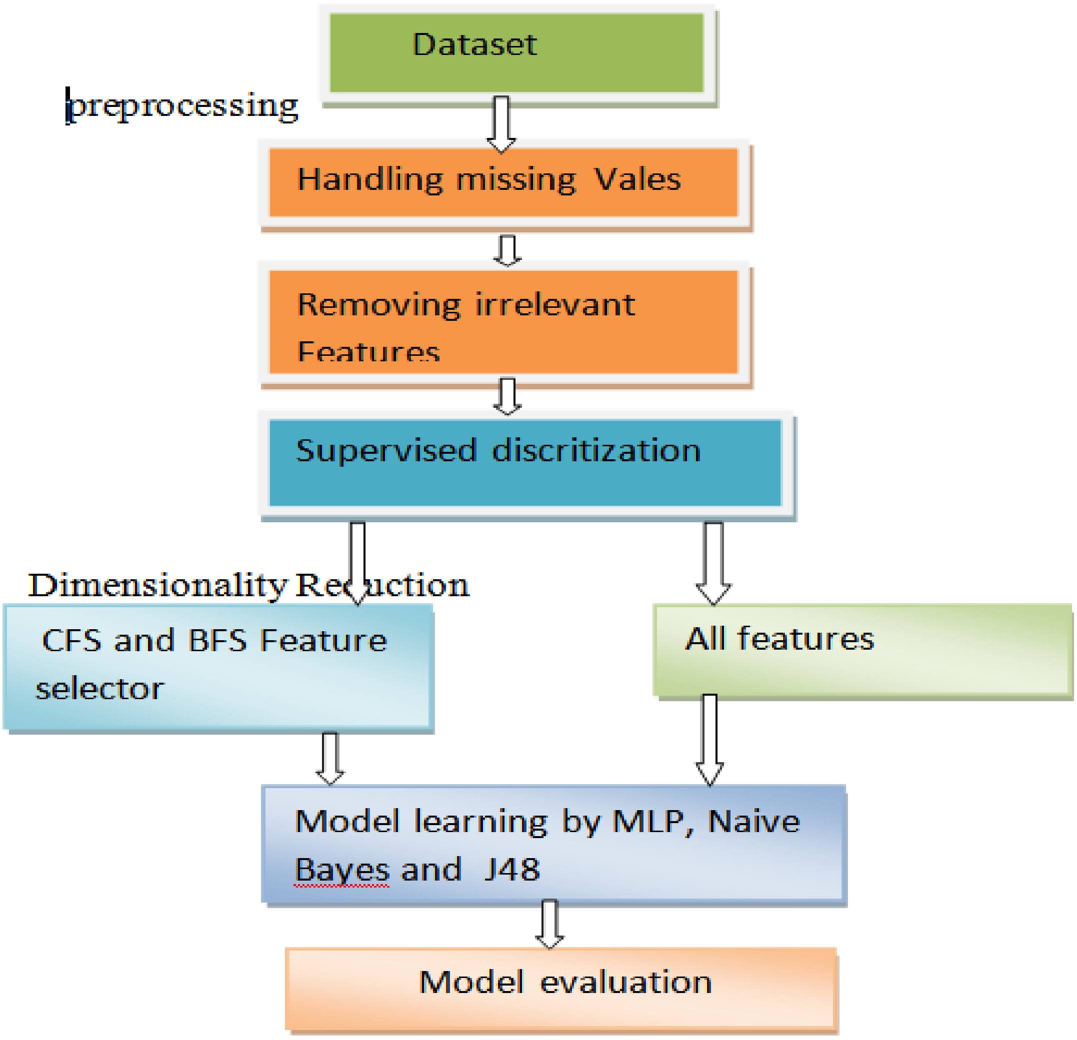


Figure 3 Architectural diagram of proposed work

The following section explains how each of specified process/steps in the proposed work is performed. The sections also explain methods, algorithm and tools used in each step of the proposed work.

3.3.1. Dataset

3.3.1.1. Sample size and sampling procedure

The target population of this study is around 6000 active students at HCTE in 2018/19 academic year. Actually, it is impossible to collect data from the whole population because of time available to the researcher and large size of current students at HCTE. Group stratified sampling method was used to select the samples for thesis. By applying the specified sampling method 26 groups of students were selected from total of 127 groups. Means this study considered 1023out of around 6000 students currently taking courses in HCTE in regular, weekend and evening programs.

3.3.1.2. Data Types and Sources

Both primary and secondary data are collected for this research. Primary data which is known as first hand data are collected directly from the students through questionnaires distributed to the respondents and Secondary data are obtained from the college's registrar office.

3.3.1.3. Study Variables

The dependent variable is the variable being tested and measured in an experiment (Mcleod, 2019).The dependent variable of this study is academic performance of the students. Academic performance of the students was measured by the students CGPA/ Status. The independent variable is the variable(s) that the experimenter changes or controls and is assumed to have a direct effect on the dependent variable (Mcleod, 2019).The independent variables of this study include sex, department, previous school type, assignment performance, library usage, school facility usage, environmental motivation, study hours, co-curricular activity participation, sport interest, home location ,financial problem, family support, family education, time spend in social media and family economic status.

3.3.1.4. Data collection

Data was collected from 14 different department and 26 group students selected using group stratified sampling. The departments are, English, Mathematics and environmental science, Preschool, Amharic, Biology, Adult education, Chemistry, Physics, Mathematics, Social study, Civic, SidamaAfoo, Chemistry laboratory and Biology laboratory. The questionnaire was distributed to 26 groups of selected students in college. In total, 1023(724 regular, 223 weekend, and 76 evening) students filled the questionnaire. The questionnaire constitutes personal and academic related attributes that are assumed to have an effect on the college students' academic achievement. The collected data was organized in Microsoft Excel worksheet and each student record contains different features.

The complete description of attributes used for this study is described in table 1.

Table 1: List of selected attributes and possible values

Attributes	Description	Possible values
SEX	The gender of the student	{Female, Male}
CH	Amount of total course credit hours taken by the students	Integer (0, 1, 2, 3.....)
PHST	Type of previous high school attended	{Private, Public}
PRO	The type of program	{Day, Evening, weekend}
YEAR	Number of years the student Spend in the college.	{One, Two, Three}
DEP	The students department	{Amharic, English, Biology,}
ATT	Percentage of classes attended	Integer (0, 1, 2, 3.....)
BBR	Book borrowing rate From the school library.	{Once a week, Twice a month, Never or almost never, Few times a year}
PIA	Performance in doing assignments.	{Always , Often, Sometimes, Rarely Copy from friends}
SFUR	Students school facility usage, Like laboratory...	{Once a week, Twice a month , Never or almost never, Few times a year}
ASHW	Average study hours per week.	Integer (0, 1, 2, 3.....)
RPCA	Rate of Participation in Co-curricular activity.	{Low, Average , Good}
RISA	Rate of interest of students in sport activities	{Low, Average, Good}
EMIS	Environmental motivation for Study	{Yes , No}
LL	Attention during lecture time.	{ Always, Often, Sometimes, Rarely , Never}
TI	Teaching interest	{ Yes , No}
HL	Home distance from the college	{Yes , No}
FP	Student's Financial problem experience.	{Always, Often, Sometimes, Rarely, Never}
FS	Financial support from family and relatives	{Always, Often, Sometimes, Rarely, Never}
MD	Destruction of the student's mobile phone while he/she is in	{Always, Often, Sometimes, Rarely, Never}

Attributes	Description	Possible values
	study	
SMT	Average Time spent in social medias per day	{0m',0-10m' ,10m- 30m' , 1-2 hrs, 2-3hrs, 3-4 hrs, 4-5hrs, 5+hrs }
ME	Education level of the students' mother.	{Primary school(1-6 grades), Secondary school(7-12) grades, Diploma, Degree, Masters Degree, Other specify______}
FE	Education level of the students' Father.	{Primary school(1-6 grades), Secondary school(7-12) grades, Diploma, Degree, Masters Degree, Other specify______}
FES	Economic status of the students family	{Upper class, Medium class, Lower class}
CGPA	Previous semester Cumulative grade point average	1.85-4.00
CLASS	The student class / current status of the student	{GreatDistinction–>=3.5,Distinction->=3.25, Good–>2.6,satisfactory->=2.00 and Warning<2.00}

3.3.1.5. Background of the respondents

This Section describes the background of the respondents according to gender, department, year level, and program type.

3.3.1.5.1. Respondents by gender

The following table shows distribution of respondents by gender.

Table 2: Distribution of respondents by gender

Gender	Frequency	Percent
Female	279	27.3%
Male	744	72.7%
Total	1023	100%

Table 2 shows, out of the total of 1023 respondents, the male students had the highest representation of about 72.7% and the female respondents had less representation of 27.3%.

3.3.1.5.2. Respondents by department

The following table shows distribution of respondents based on their departments.

Table 3: Distribution of respondents by department

Department	Frequency	Percent
English	89	8.7%

Department	Frequency	Percent
Mathematics and environmental science	120	11.7%
Pre-school	49	4.9%
Amharic	83	8.1%
Biology	76	7.4%
Adult education	49	4.9%
Chemistry	77	7.5%
Physics	85	8.3%
Mathematics	81	7.9%
Social study	79	7.7%
Civic	41	4.0%
SidamaAfoo	87	8.5%
Chemistry laboratory	48	4.7%
Biology laboratory	59	5.8%

3.3.1.5.3. Respondents by year of study

The following table shows distribution of respondents by year of study.

Table 4: Distribution of respondents by year of study

Year	frequency	Percent
Two (2^{nd} year)	405	39.6%
Three (3^{rd} year)	618	60.4%
Total	1023	100%

Table 4 shows from a total of 1023 students 618 are third year students or around two third of the respondents are 3^{rd} year students.

3.3.1.5.4. Respondents by type of program

The following table shows distribution of respondents based on study program.

Table 5: Distribution of respondents by program type

Program	Frequency	Percent
Regular	724	70.8%
Evening	76	7.4%
Weekend	223	21.8%
Total	1023	100%

Table 5 shows out of total of 1023 respondents, the regular students had the highest representation of about 70.8%and the weekend and evening students had less representation of 21.8% and 7.4% respectively.

3.3.2. Preprocessing

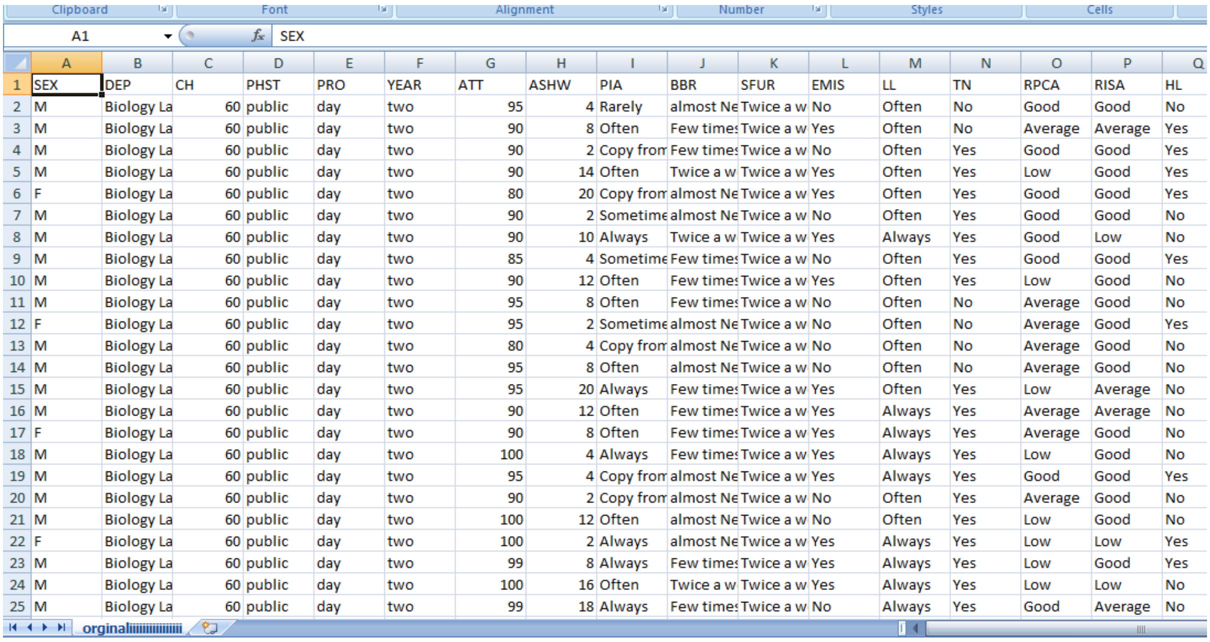
Raw data are highly exposed for noise, inconsistency and missing values. Before applying DM algorithms, it is necessary to carry out some pre-processing tasks such as cleaning, integration, reduction and variable transformation. Data pre-processing helps to improve the quality of data and the mining result (Alasadi, 2017).It also helps to reduce confusion during learning and to get better input data for data mining techniques. Irrelevant attributes, attribute that have same value and instances that are missing important values were removed from the dataset. During this processes instances are reduced from the dataset and the numbers of instances become less than the instances in the original dataset.

3.3.2.1. Data cleaning

Data cleaning is the processes of finding missing values, smoothing noisy data, recognize outliers and correct inconsistencies (Alasadi, 2017). Unclean data lead to poor output, so before the actual data mining processes, the dataset must pass through data cleaning stage.

At this stage the questionnaires were checked for completeness and some questionnaires with too many unfilled values were selected and discarded. The dataset obtained from the college registrar office contains name, identification number of the students and record of dismissed student from the college. The specified attributes are not relevant for predicting AP of the students, so they were identified and removed from the dataset by the researcher.

Figure3 shows part of the dataset in excel format prepared for next stage after data cleaning.



	SEX	DEP	CH	PHST	PRO	YEAR	ATT	ASHW	PIA	BBR	SFUR	EMIS	LL	TN	RPCA	RISA	HL
2	M	Biology La	60	public	day	two	95	4	Rarely	almost Ne	Twice a w	No	Often	No	Good	Good	No
3	M	Biology La	60	public	day	two	90	8	Often	Few time:	Twice a w	Yes	Often	No	Average	Average	Yes
4	M	Biology La	60	public	day	two	90	2	Copy from	Few time:	Twice a w	No	Often	Yes	Good	Good	Yes
5	M	Biology La	60	public	day	two	90	14	Often	Twice a w	Twice a w	Yes	Often	Yes	Low	Good	Yes
6	F	Biology La	60	public	day	two	80	20	Copy from	almost Ne	Twice a w	Yes	Often	Yes	Good	Good	Yes
7	M	Biology La	60	public	day	two	90	2	Sometime	almost Ne	Twice a w	No	Often	Yes	Good	Good	No
8	M	Biology La	60	public	day	two	90	10	Always	Twice a w	Twice a w	Yes	Always	Yes	Good	Low	No
9	M	Biology La	60	public	day	two	85	4	Sometime	Few time:	Twice a w	No	Often	Yes	Good	Good	Yes
10	M	Biology La	60	public	day	two	90	12	Often	Few time:	Twice a w	Yes	Often	Yes	Low	Good	No
11	M	Biology La	60	public	day	two	95	8	Often	Few time:	Twice a w	No	Often	No	Average	Good	No
12	F	Biology La	60	public	day	two	95	2	Sometime	almost Ne	Twice a w	No	Often	No	Average	Good	Yes
13	M	Biology La	60	public	day	two	80	4	Copy from	almost Ne	Twice a w	No	Often	No	Average	Good	No
14	M	Biology La	60	public	day	two	95	8	Often	almost Ne	Twice a w	No	Often	No	Average	Good	No
15	M	Biology La	60	public	day	two	95	20	Always	Few time:	Twice a w	Yes	Often	Yes	Low	Average	No
16	M	Biology La	60	public	day	two	90	12	Often	Few time:	Twice a w	Yes	Always	Yes	Average	Average	No
17	F	Biology La	60	public	day	two	90	8	Often	Few time:	Twice a w	Yes	Always	Yes	Average	Good	No
18	M	Biology La	60	public	day	two	100	4	Always	Few time:	Twice a w	Yes	Always	Yes	Low	Good	No
19	M	Biology La	60	public	day	two	95	4	Copy from	almost Ne	Twice a w	Yes	Always	Yes	Good	Good	Yes
20	M	Biology La	60	public	day	two	90	2	Copy from	almost Ne	Twice a w	No	Often	Yes	Average	Good	No
21	M	Biology La	60	public	day	two	100	12	Often	almost Ne	Twice a w	No	Often	Yes	Low	Good	No
22	F	Biology La	60	public	day	two	100	2	Always	almost Ne	Twice a w	Yes	Always	Yes	Low	Low	Yes
23	M	Biology La	60	public	day	two	99	8	Always	Few time:	Twice a w	Yes	Always	Yes	Low	Good	Yes
24	M	Biology La	60	public	day	two	100	16	Often	Twice a w	Twice a w	Yes	Always	Yes	Low	Low	No
25	M	Biology La	60	public	day	two	99	18	Always	Few time:	Twice a w	No	Always	Yes	Good	Average	No

Figure 4 Dataset in excel after data cleaning

3.3.2.2. Data Discretization

Discrete attributes are those that describe a category, called nominal attributes. Data discretization is performed to replace raw values of numeric attribute by ranges (interval levels) or conceptual level. Concept hierarchies allow the mining of data at multiple level of abstraction and are a powerful tool for data mining. For this study, Total credit hour, Average study hours per week and Average percentage of classes attended by the students are collected as numeric attribute. To discretize those attributes CSV data format of the data set is imported to WEKA and then the Discretization is done by one of the filter tools in the Software called supervisedDescritizer. SupervisedDescretizer sorts the values of a feature, and then attempts to find interval boundaries such that each interval has a strong majority of one particular class. The method is constrained to form intervals of some minimal size in order to avoid having intervals with very few instances.

Supervised discretizer is selected because of its power to consider class level of attributes while discretizing and work with different types of attributes like binary attributes, empty nominal attributes, numeric attributes, string attributes etc. The discretization process is screen captured as shown in figure 4.

The dataset has 25 attributes and contains 1000 instances.

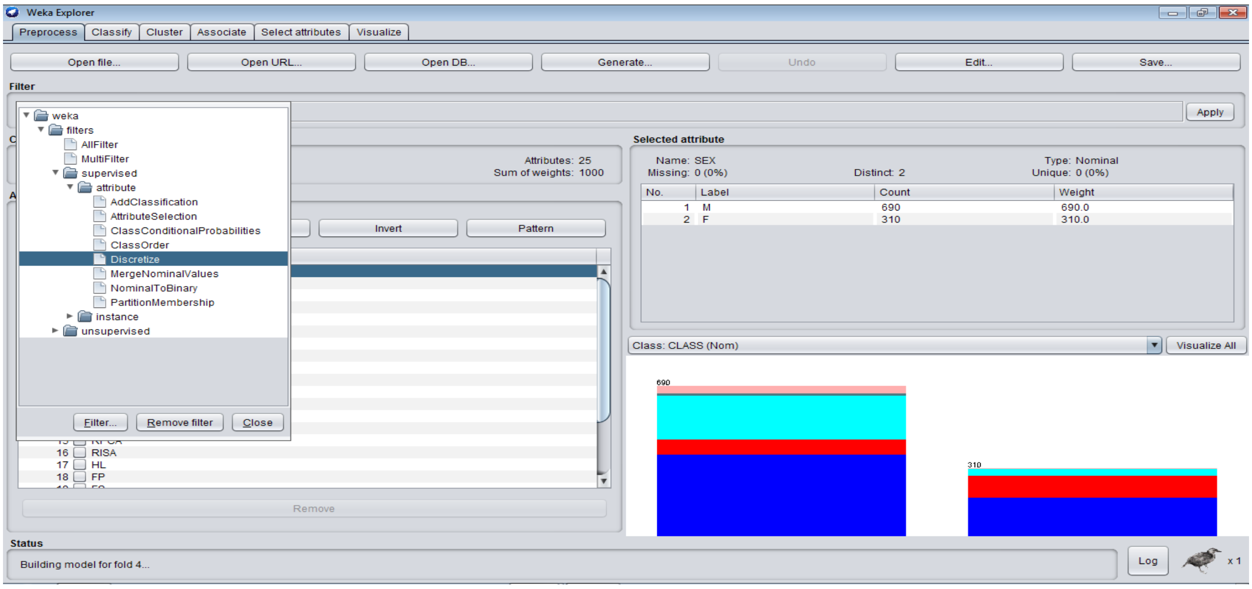


Figure 5 Preprocessor setup

3.3.3. Feature Selection

Feature selection is the process of selecting attributes that contribute most to our redaction variable or output in which we are interested in. Irrelevant or partially relevant features can negatively impact model performance. Feature selection methods used to reduce over fitting, improve accuracy and reduce Training Time**.**

Feature selection is divided into two parts: Attribute Evaluator and Search Method. The attribute evaluator is the technique that evaluates each attribute based on the context of the output variable. The search method is the technique to navigate different combinations of attributes in the dataset in order to arrive on a short list of chosen features (Brownlee, 2016). In this work CFS is used as an attribute evaluator and BFS as searching methods.

Feature Evaluation

A Correlation-based Feature Selector (Hall & Smith, 1999; Wang et al., 2014) is popular technique for selecting the most relevant attributes in the dataset that have strong correlation with the class. CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.CFS is a fully automatic algorithm or it does not require the user to specify the number of features to be selected. The correlation feature selection (CFS) measure evaluates subsets of features on the basis of the following hypothesis: "Good feature subsets contain features highly correlated with the classification, yet uncorrelated to each other".

The following equation gives the merit of a feature subset *S* consisting of *k* features and class c, CFS defined relevance of features subset by using Pearson's correlation equation

$$M_S = \overline{kr_{cf}} / \sqrt[2]{k + (k-1)r_{ff}\overline{r_{cf}}} \qquad (3.1)$$

Where M_s is relevance of feature subset, $\overline{r_{cf}}$ is the average linear correlation coefficient between these features and classes and $\overline{r_{ff}}$ is the average linear correlation coefficient among different features.

Best first Search

The implementation of CFS used in the experiments described in this thesis applies best first search strategy.BFS(Hall, 1999) is a search algorithm which explores attributes by expanding the most promising feature chosen according to a specified rule. It can start with either no features or with all features. In the former, the search progresses forward through the search space adding single features; in the latter the search moves backward through the search space deleting single features. To prevent the best first search from exploring the entire feature subset search space, a stopping criterion is imposed. Normally this involves limiting the number of fully expanded subset and that results in no improvement over the current best subset.

The BFS algorithm uses the following steps:

1. Begin with the OPEN list containing the start state, the CLOSED list empty, and BEST ←start state.
2. Let s = arg max e(x) (get the state from OPEN with the highest evaluation).
3. Remove s from OPEN and add to CLOSED.
4. If e(s) ≥ e (BEST), then BEST ← s.
5. For each child t of s that is not in the OPEN or CLOSED list, evaluate and add to OPEN.
6. If BEST changed in the last set of expansions, goto 2.
7. Return BEST

3.3.4. Model learning tools and algorithms

3.3.4.1. WEKA environment

The researcher used WEKA 3.8.1 tool to develop the model using machine learning algorithm. WEKA stands for Waikato Environment for Knowledge Analysis (Hall et al., 2009). It was developed by the University of Waikato, New Zealand. WEKA supports many data mining tasks such as classification, clustering, regression and feature selection. The data formats supported by WEKA are ARFF, CSV, C4.5 and binary. Alternatively, it is possible to import from URL or an SQL database. WEKA is open source software issued under the General Public License.

3.3.4.1. Multilayer perceptron

Multilayer perceptron (MLP) is the most popular architecture for neural networks. The network consists of a set of sensory elements that make up the input layer, one or more hidden layers of processing elements, and the output layer of the processing elements (Witten, Frank, & Hall, 2011).where each neuron is connected to all the neurons from the previous layer.MLP is especially suitable for approximating a classification function to know relationship between input and output attributes.

The Multilayer perceptron algorithm is found in the weka.classify.functions tab and has many parameters for configuration. This paper used SYNOPSIS that represents a classifier that uses backward Propagations learning method to classify instances. This network can be built by hand, created by an algorithm or both. The nodes in this network are all sigmoid. (Except for when the class is numeric in which case the output nodes become linear units without threshold).

The MLP parameters that can be configuring in WEKA are ("Department of Computer Science: University of Waikato," 2016):

- ✓ **Seed**: - used to initialize the random number generator. Random numbers are used for setting the initial weights of the connections between nodes, and for shuffling the training data.
- ✓ **Momentum**:– is the value that is applied to the weights during updating
- ✓ **nominalToBinaryFilter** :– represents the filter that will preprocess the instances. This could help to improve performance if there are nominal attributes in the dataset.
- ✓ **hiddenLayers** :- This defines the number hidden layers of the neural network and the number of neurons from each layer. This is a list of positive integer numbers, one for each hidden layer, comma separated. To specify no hidden layers, you need to put a single 0 here, and this will only be used if auto build is set.

There are also wildcard values 'a' = (attributes + classes) / 2, 'i' = attributes, 'o' = classes,

't' = attributes + classes. We have used the ‘a’ wildcard.

- ✓ **validationThreshold**- Used to terminate the learning process. The value here dictates how many times in a row the validation set error can get worse before training is terminated.
- ✓ **GUI**- Brings up a graphic user interface. This will allow the pausing and altering of the neural network during training. Can add a node, create new connections between nodes, and remove a connection or a node. If this option is activated, then the network is automatically paused at the beginning and the user can reconfigure the network. Once the network configuration is done, it will pause again and either waits to be accepted or trained more. If the GUI is not set the network will not require any interaction.
- ✓ **normalizeAttributes**- This will normalize the attributes. This could help to improve the performance of the network. This is not reliant when we have numeric classes. This will also normalize nominal attributes as well (after they have been run through the nominal to binary filter if that is in use) so that the nominal values are between -1 and 1
- ✓ **numDecimalPlaces**- The number of decimal places to be used for the output of numbers in the model.
- ✓ **batchSize**- The preferred number of vector instances kept in cache once if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- ✓ **decay**- This will cause the learning rate to decrease. This will divide the starting learning rate by the epoch number, to determine what the current learning rate should be. This may help to stop the network from diverging from the target output, as well as improve general performance. Note that the decaying learning rate will not be shown in the GUI, only the original learning rate. If the learning rate is changed in the GUI, this is treated as the starting learning rate.
- ✓ **validationSetSize**- The percentage size of the validation set. The training will continue until the error on the validation set has been consistently getting worse, or if the training time is reached.
- ✓ **trainingTime**- The number of epochs to train through. If the validation set is nonzero then it can terminate the network learning.

- ✓ **Debug**- If set to true, classifier may output additional info to the console.
- ✓ **autoBuild**- Adds and connects up hidden layers in the network.
- ✓ **normalizeNumericClass**- This will normalize the class if it is numeric. This could help improve performance of the network, It normalizes the class to the range [-1, 1]. Note that this is done only internally; the output will be scaled back to the original range.
- ✓ **learningRate**-- The amount the weights are updated.
- ✓ **doNotCheckCapabilities**- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime).

reset- This will allow the network to reset with a lower learning rate. If the network diverges from the answer this will automatically reset the network with a lower learning rate and begin training again. This option is only available if the GUI is not set. Note that if the network diverges but isn't allowed to reset it will fail the training process and return an error message.

3.3.5. Model Evaluation and Evaluation metrics

Evaluation metrics are used to evaluate the performance efficiency of the classification algorithm. Accuracy, precision, recall, f-measure, True Rate, Negative Rate etc. are some of the classification algorithm evaluation metrics in WEKA. The following evaluation measures are used for this research: Classification Accuracy, precision, Recall, F-measure and mean square error. Each evaluation metrics is explained as follows.

Confusion matrix

A confusion matrix is a table layout that is used to describe the performance of a classification algorithm. Confusion matrix is used to visualize the performance of the classifier that is trained in a supervised learning. It also shows the actual and predicted values/ labels from classification problem and used to derive different performance measures such as accuracy, precision, recall, f-measure etc.

Table 6: Truth tale confusion matrix

		Prediction outcome	
		Negative	Positive
Actual value	Negative	True negative/TN	False positive/FP
	Positive	False negative/FN	True positive/TP

There are 4 important terms:

- ✓ True Positives: predicted as positive and the actual results were also positive.
- ✓ True Negatives: predicted as negative and the actual results were also negative.
- ✓ False Positives: predicted as positive and the actual results were also negative.
- ✓ False Negatives: predicted as negative and the actual results were also positive.

Classification Accuracy

Classification Accuracy is the ratio of the number correct prediction out of all prediction or it is the measure of how close a predictive value is to the actual value (Tiwari, 2019).Often presented as a percentage value and it retrieves the percent of correctly classified instances out of 100%. It is the ratio of number of correct predictions to the total number of input samples.

$$\text{Accuraccey} = \frac{\text{Nmber of correct pridiction}}{\text{total no of pridiction}}$$

$$\text{Accuracey} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (3.2)}$$

Precision

Precision is measure of accurateness or correctness of the model, which is the measure of the percentage of instances in test dataset that are labeled as positive, are actually positive (Tiwari, 2019) or it is a fraction of relevant instances among the retrieved instances. It the number of correct positive results divided by the number of positive results predicted by the classifier.

$$precition = \frac{True\ Posatives}{Total\ posative\ pridiction}$$

$$Precition = \frac{TP}{TP+FP} \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (3.3)}$$

Recall

Recall is the measure of completeness refers to the percentage of total relevant results correctly classified by the algorithm (Tiwari, 2019) or it is the fraction of the total amount of relevant instances that were actually retrieved. It is the number of correct positive results divided by the number of all relevant samples or all samples that should have been identified as positive.

$$Recall = \frac{TruePosatives}{trueposatives + Falsenegatives}$$

$$Recall = \frac{TP}{TP+FN} \quad (3.4)$$

F-measure

F- Measure (Tiwari, 2019) is a measure of a test's accuracy. It is a metric which takes into account both precision and recall, and maximizing F-measure increase the quality of the model or leads to a better model performance. It is the harmonic mean of precision and recall. F-measure is a useful metric if we are looking for a balance between precision and recall and there's an uneven class distribution. F-measure is stable of measuring performance of a model with unbalanced number of classes.

$$F - Measure = 2 \times \frac{precision \times Recall}{Precision+Recall} \quad (3.5)$$

Mean Absolute Error (MAE)

The Mean absolute error (Agri, 2019) measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

$$MAE = \frac{1}{n} \times \sum_{i=0}^{n} |o_i - p_i| \quad (3.6)$$

Where n is the number of errors O_i is the actual value and p_i is predicted value.

3.3.5.1. Cross validation

Cross validation is one of the techniques used to select training and test set in data mining algorithms. In cross-validation method, the original dataset are randomly divided into mutually exclusive and equal-sized subsets. For each subset, the classifier is trained with the union of all other subsets and tested on that subset (Kohavi, 1995).If the original dataset is segmented into k subsets, then the method is called k-fold cross validation. The cross validation method ensures that each record from the original dataset is used the same number of times for both training and testing. In cross validation method, the total accuracy is obtained by averaging the accuracies for all k runs.

To validate the prediction model, the 10-fold cross-validation was used for this thesis. The advantage of this method is that all observations are used for both training and validation, it avoids over fitting and generalizes the model to an unknown dataset (Cawley & Talbot, 2010).The 10 fold cross valuation works by dividing the dataset into 10 parts, nine of which were for training and the rest for testing. The training process was repeated 10 times and then the performance of the model was computed.

Chapter Four

4. Experimentation and Result Discussion

4.1. Model learning and model selection experiments

The model learning experiments have been performed in the WEKA framework using Multilayer perceptron algorithm as described earlier. Initially the dataset is in CSV format after loading the dataset in to WEKA, the dataset format is converted to ARFF.

The steps followed to learn and save the finalize model are

1. Open the WEKA GUI Chooser.
2. Click the "Explorer" button to open the WEKA Explorer interface.
3. Click "open file" to Load the dataset,
4. Click the "Classify" tab to open up the classifier.
5. Select the classifier, configure and then click start.
6. Right click on the "Result list" and click "save model" from the right click menu
7. Select a location and enter a filename, click the "Save" button.

The following three sections of experiments (A, B, C) are conducted with different objectives, experiment A is conducted to check weather data discritization helps to improve model performance or not, experiment B conducted to select important feature and experiment C is Model learning and model selection experiment and six experiments under experiment C are conducted to select best performing model by MLP classification algorithm with different possible combinations of LR and Momentum.

Experiments A: Performance of MLP before and after data discretization

The objective of this experiment is to measure the performance of MLP algorithms before and after discretizing the dataset. It is conducted to check weather data discretization helps to improve performance of MLP algorithm or not. The First experiment is conducted before discretizing numeric attributes and the second experiment is conducted after discretizing the numeric attributes. Weka.Filter.Supervised.Attribute.Discretize is used to discretize the dataset. The results of the two experiments are screen captured as shown in figure 6.

```
=== Summary ===

Correctly Classified Instances         575               57.5    %
Incorrectly Classified Instances       425               42.5    %
Kappa statistic                          0.2798
Mean absolute error                      0.1767
Root mean squared error                  0.3768
Relative absolute error                 72.0664 %
Root relative squared error            107.7128 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.717    0.461    0.656      0.717    0.685      0.260    0.679     0.697     satisfactory
                 0.435    0.104    0.463      0.435    0.448      0.340    0.771     0.467     warning
                 0.451    0.145    0.486      0.451    0.468      0.314    0.738     0.450     Good
                 0.000    0.005    0.000      0.000    0.000      -0.007   0.727     0.024     Greate Distinction
                 0.029    0.017    0.059      0.029    0.038      0.017    0.676     0.086     Distinction
Weighted Avg.    0.575    0.306    0.555      0.575    0.564      0.275    0.709     0.572

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 395  74  74   2   6 |   a = satisfactory
  85  74  11   0   0 |   b = warning
 107  11 105   2   8 |   c = Good
   3   0   6   0   2 |   d = Greate Distinction
  12   1  20   1   1 |   e = Distinction
```

Figure 6: Classification by MLP before data Discretization

```
=== Summary ===

Correctly Classified Instances         584               58.4    %
Incorrectly Classified Instances       416               41.6    %
Kappa statistic                          0.2898
Mean absolute error                      0.1745
Root mean squared error                  0.3759
Relative absolute error                 71.1816 %
Root relative squared error            107.4566 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.739    0.468    0.660      0.739    0.697      0.277    0.682     0.694     satisfactory
                 0.453    0.089    0.510      0.453    0.480      0.382    0.764     0.464     warning
                 0.421    0.140    0.478      0.421    0.447      0.294    0.768     0.480     Good
                 0.000    0.004    0.000      0.000    0.000      -0.007   0.741     0.039     Greate Distinction
                 0.057    0.022    0.087      0.057    0.069      0.043    0.653     0.063     Distinction
Weighted Avg.    0.584    0.306    0.565      0.584    0.572      0.288    0.716     0.576

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 407  63  73   2   6 |   a = satisfactory
  87  77   6   0   0 |   b = warning
 112  11  98   2  10 |   c = Good
   3   0   3   0   5 |   d = Greate Distinction
   8   0  25   0   2 |   e = Distinction
```

Figure 7: Classification by MLP after data Discretization

Result discussion

As the experimental result indicates, classification accuracy of MLP classifier before discretizing numeric attributes is 57.5% and after discretizing the numeric attributes 58.4%, so it is clearly shown that the classification accuracy of MLP classifier is improved after discretizing the dataset. Moreover, all the evaluation metrics such as Precision, Recall, F-measure and MAE show improved performance of MLP classifier after descretization. Therefore, all experiments following this are conducted using discretizeddataset.

Experiment B: Attribute selection using CFS and visualizing attributes relation with the class.

The following experiment is conducted to identify attribute that has strong correlation with the class attribute using CFS feature selection. From weka.Selectattributes section,CfsSubsetEvalevaluatoris used with BestFirst search method to select the most important attributes in predicting AP of the students. Feature selection experiment is screen captured as shown in figure 8.

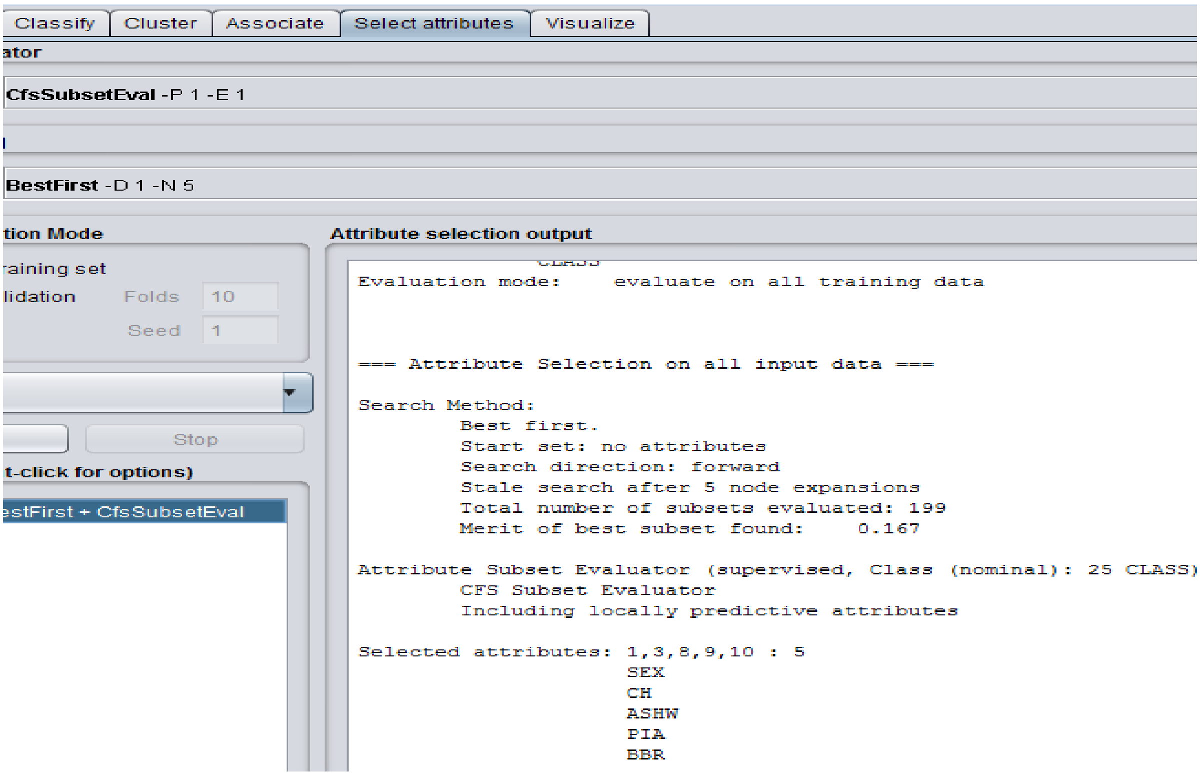


Figure 8: Attribute selection using CFS with BFS

The above figure shows the top five attributes have strong correlation with class attribute. That is, these five are the most important attributes for predicting AP.

Following this, the weka.visualize tab is used to identify the distribution of instances based on their features relative to the class. It helps to identify positive or negative relationship that exists between the class and selected attributes. The distribution of instances based on the identified features such as, SEX, CH, ASHW, PIA and BBR with respect to the class attribute is visualized, screen captured (see figures 9,10,11,12,and 13) and discussed below.

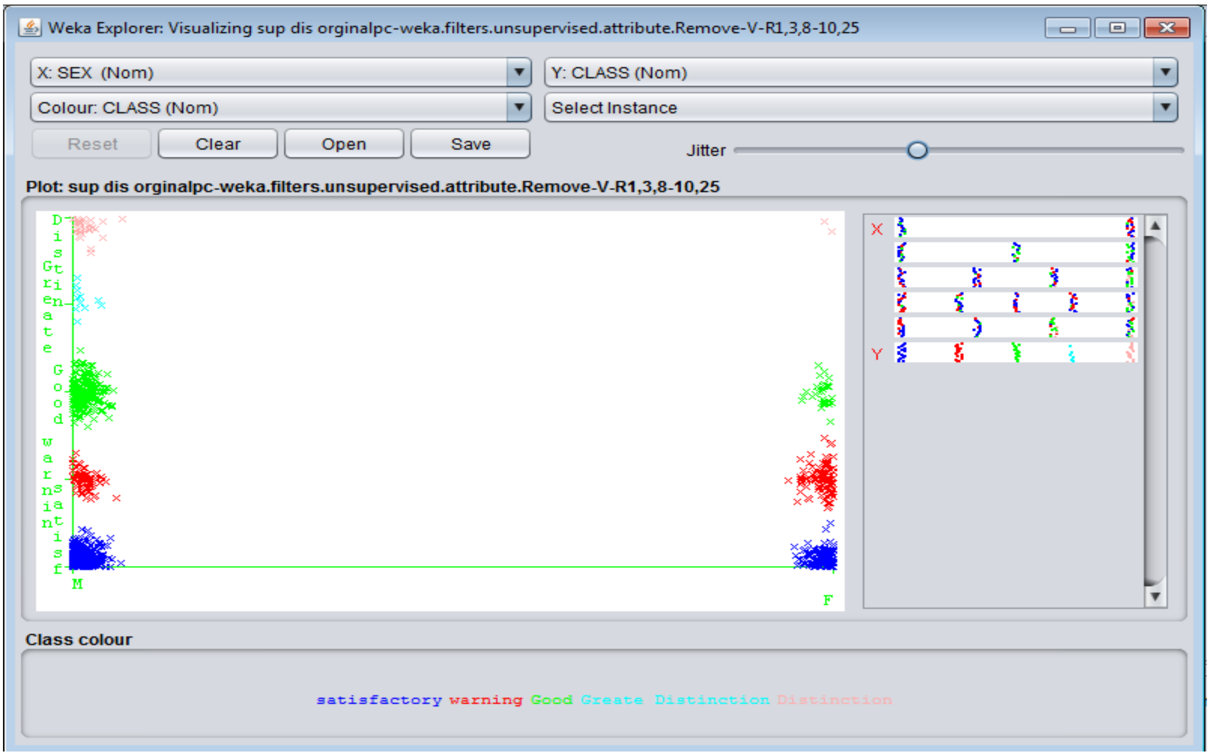


Figure 9: Distribution of instances based on sex with respect to the class

The above figure shows that most of female students had obtained class value of warning and satisfactory and only few female students had class of good, distinction and great distinction. This indicates good or bad performance of the student is highly dependent on the sex of the student, female students had high risk of getting poor result than male students. This finding contradicts with the findings obtained by (Akessa & Dhufera, 2015) and in line with the findings obtained by (Yigermal, 2017).

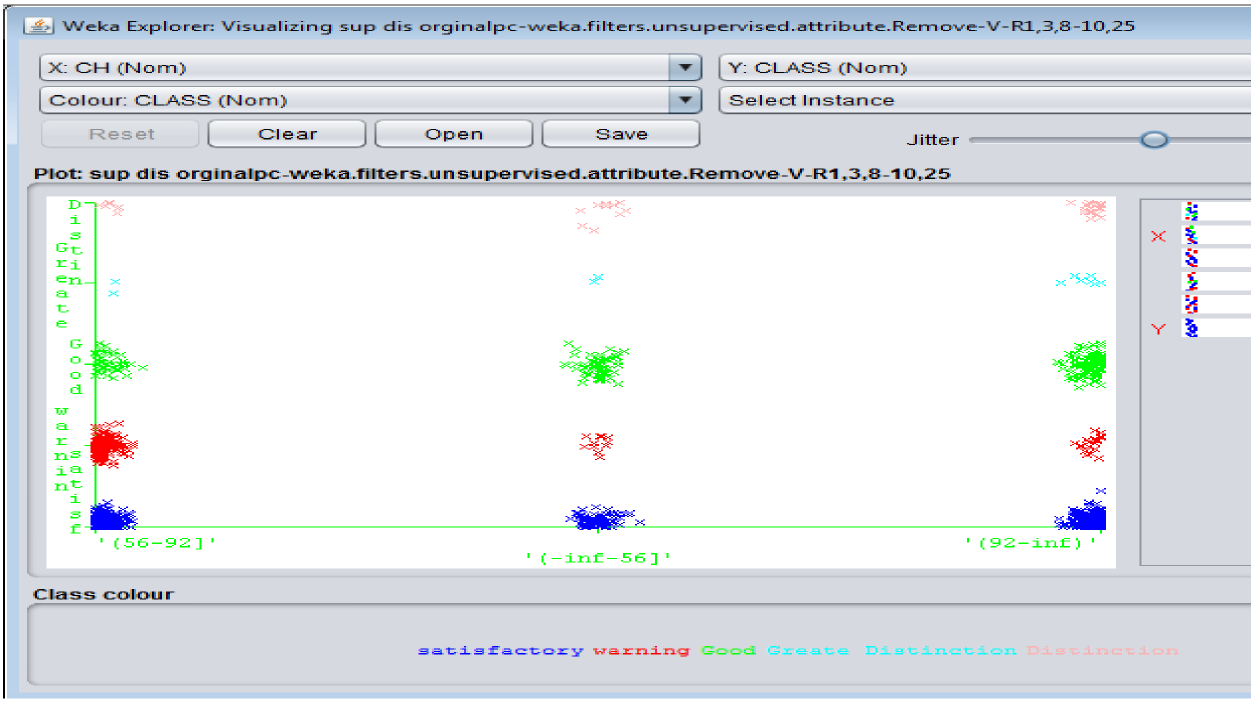


Figure 10: Distribution of instances based on total credit hours with respect to the class

The above figure shows the distribution of instances based on amount of total courses credit hours taken by the students. Most of the students' obtained better result at the beginning and end of the study. The large distribution of instances under warning class is obtained at the middle of the study or after taking courses of total credit course greater than 56 to 92.

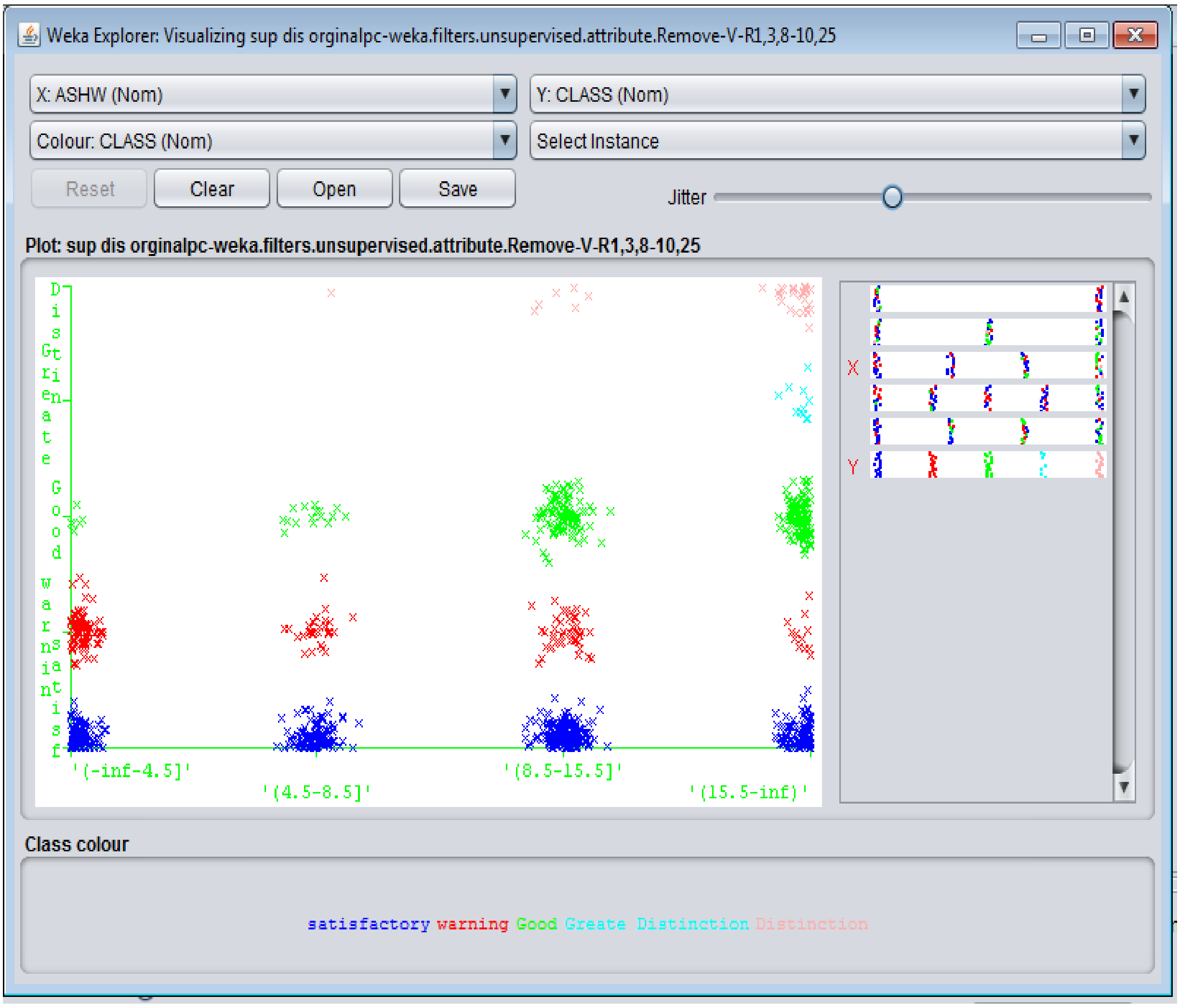


Figure 11: Distribution of instances based on Average study hours per week with respect to the class

The above figure shows the distribution of instances based on amount of students' average study hours per week. The figure clearly shows that the AP is directly related to amount of time the student spend in studying courses. Students reading for more than 15.5 hours per week on average had higher probability of getting distinction and great distinction. This indicates good or bad performance of the student depends on the amount of time he/she spent in studying courses per week. This finding is in line with the findings obtained by (Yigermal, 2017).

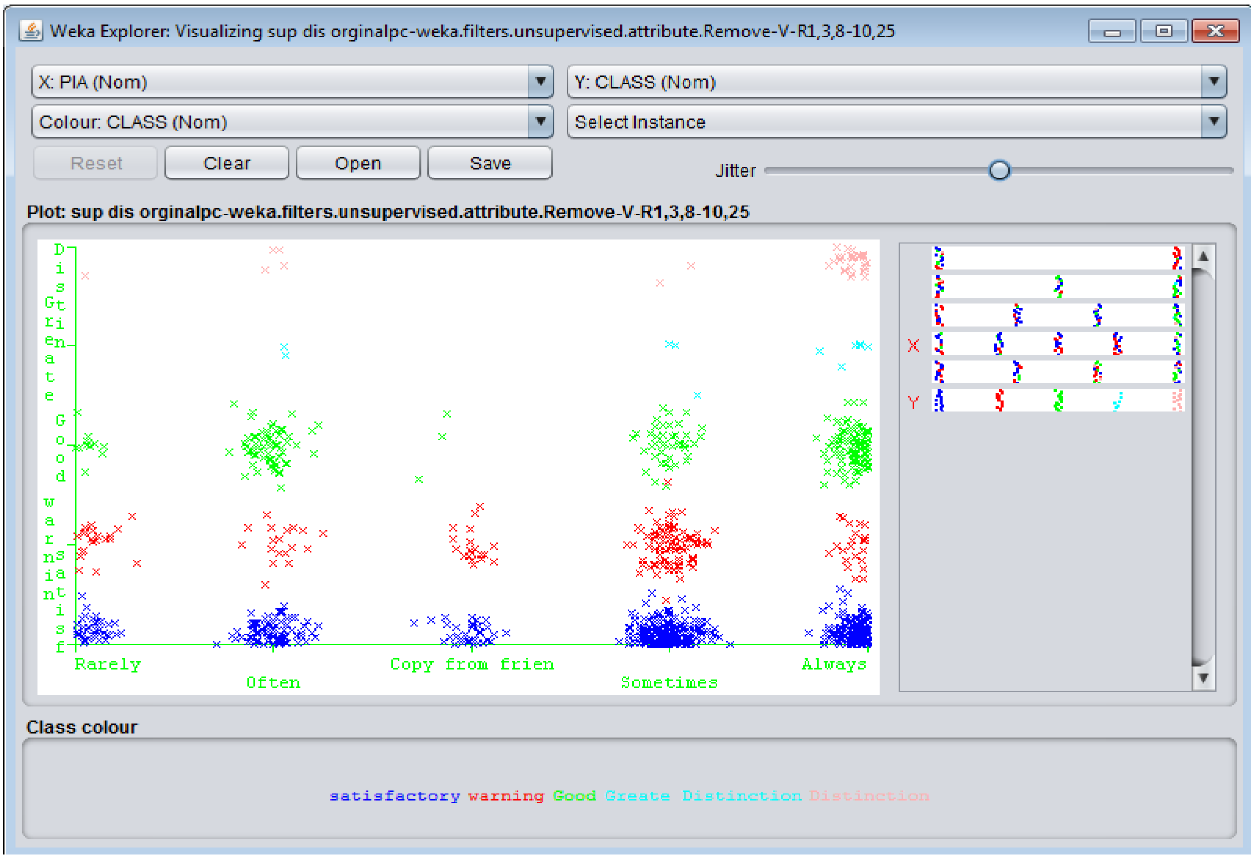


Figure 12: Distribution of instances based on Assignment performance with respect to the class

The forth important attribute is students' assignment performance. The distribution of instances based on their performance in doing assignments indicate that most of the students who can do their assignments regularly had higher probability of getting better academic achievement than others. This result shown that doing assignments regularly helps to improve college students academic achievement.

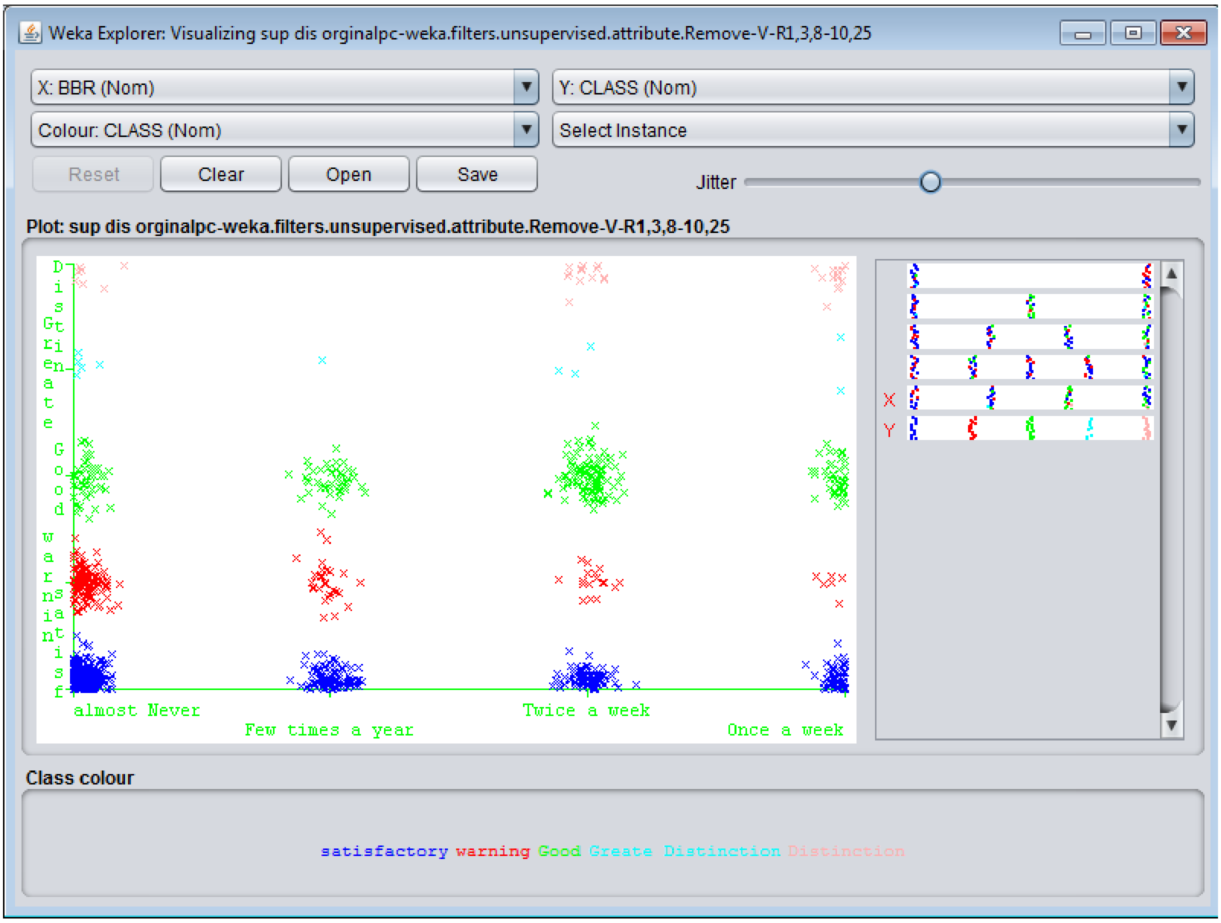


Figure 13: Distribution of instances based on Book borrowing rate with respect to the class

The above figure shows distribution of instances based on book borrowing rate (BBR) with respect to the class attribute. From the above figure we can observe that students who borrow books from library once a week and twice a month had higher probability of getting good, Distinction and great distinction class status. Most of the students who never borrow books from library categorized under warning and satisfactory class label. This finding show borrowing books from library regularly helps to improve AP of the students.

Based on the result of the above feature selection experiment and discussions, the first research question can be answered. Attributes that contributes more in predicting AP of college students are sex of the student, Amount of courses credit hours taken by the students, Average study hours per week, Assignment performance and Book borrowing rate.

Experiment C: Performance of MLP before and after attribute selection

The following six experiments are conducted to select best performing MLP with 10-fold cross validation test method by changing the learning rate and momentum. The first three experiments are conducted using all 25 attributes and the later three are conducted using the 6 selected attributes. Through the process of model learning, the researcher performed many experiments by configuring learning rate and momentum of MLP, but only experiments that cause significant changes are reported here. The parameters considered for reporting the result of the following experiments are learning rate and momentum, classification accuracy, precision, recall, F-measure, root mean absolute error and training time.

Experiment one

The first experiment is conducted to evaluate the performance of a MLP classifier using all the 25attributes, the values of learning rate and momentum are 0.2 and 0.9respectively. The result of the experiment is screen captured as shown in figure 14.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         497               49.7    %
Incorrectly Classified Instances       503               50.3    %
Kappa statistic                          0.1578
Mean absolute error                      0.2015
Root mean squared error                  0.4401
Relative absolute error                 82.2069 %
Root relative squared error            125.7993 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.731    0.555    0.618      0.731    0.670      0.185    0.613     0.631     satisfactory
                 0.047    0.008    0.533      0.047    0.086      0.119    0.572     0.255     warning
                 0.356    0.111    0.494      0.356    0.414      0.277    0.685     0.424     Good
                 0.273    0.163    0.018      0.273    0.034      0.031    0.559     0.045     Greate Distinction
                 0.000    0.001    0.000      0.000    0.000      -0.006   0.583     0.051     Distinction
Weighted Avg.    0.497    0.335    0.547      0.497    0.481      0.187    0.621     0.492

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 403   6  55  86   1 |   a = satisfactory
 130   8   8  24   0 |   b = warning
 105   0  83  45   0 |   c = Good
   3   0   5   3   0 |   d = Greate Distinction
  11   1  17   6   0 |   e = Distinction
```

Figure 14: Multilayer perceptron with LR=0.2 and M=0.9

Experiment two

The second experiment is conducted to evaluate the performance of a MLP using 25 attributes by changing the values of learning rate and momentum into 0.4 and 0.7 respectively. The result of the experiment is screen captured as shown in figure 15.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         589               58.9    %
Incorrectly Classified Instances       411               41.1    %
Kappa statistic                          0.2974
Mean absolute error                      0.1662
Root mean squared error                  0.3935
Relative absolute error                 67.8011 %
Root relative squared error            112.4825 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.739    0.474    0.656      0.739    0.695      0.271    0.661     0.666     satisfactory
                 0.365    0.082    0.477      0.365    0.413      0.316    0.735     0.393     warning
                 0.489    0.133    0.528      0.489    0.508      0.366    0.764     0.497     Good
                 0.000    0.007    0.000      0.000    0.000      -0.009   0.668     0.029     Greate Distinction
                 0.171    0.022    0.222      0.171    0.194      0.170    0.708     0.112     Distinction
Weighted Avg.    0.589    0.307    0.574      0.589    0.578      0.294    0.699     0.554

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 407  62  74   2   6 |   a = satisfactory
  99  62   7   1   1 |   b = warning
  97   5 114   4  13 |   c = Good
   5   0   5   0   1 |   d = Greate Distinction
  12   1  16   0   6 |   e = Distinction
```

Figure 15: Multilayer perceptron with LR=0.4 and M= 0.7

Experiment three

The third experiment is conducted to evaluate the performance of a MLP, the values of learning rate and momentum into 0.5 and 0.3 respectively. All the 25 attributes are used for this experiment; the result of the experiment is screen captured as follows (see figure 16).

```
Classifier output

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         606               60.6    %
Incorrectly Classified Instances       394               39.4    %
Kappa statistic                          0.3347
Mean absolute error                      0.1632
Root mean squared error                  0.3649
Relative absolute error                 66.5525 %
Root relative squared error            104.3047 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.737    0.430    0.678      0.737    0.706      0.312    0.701     0.704     satisfactory
                 0.482    0.077    0.562      0.482    0.519      0.431    0.766     0.509     warning
                 0.489    0.142    0.511      0.489    0.500      0.353    0.760     0.474     Good
                 0.000    0.006    0.000      0.000    0.000      -0.008   0.746     0.049     Greate Distinction
                 0.114    0.023    0.154      0.114    0.131      0.106    0.758     0.116     Distinction
Weighted Avg.    0.606    0.284    0.593      0.606    0.598      0.331    0.728     0.589

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 406  57  78   3   7 |   a = satisfactory
  82  82   5   0   1 |   b = warning
  99   5 114   2  13 |   c = Good
   3   0   7   0   1 |   d = Greate Distinction
   9   2  19   1   4 |   e = Distinction
```

Figure 16: Multilayer perceptron with LR=0.5 and M= 0.3

Experiment Four

The fourth experiment is conducted to evaluate the performance of a MLP classifier the values of learning rate and momentum are 0.1 and 0.4 respectively. This experiment is conducted using 6selected attributes. The result of the experiment is screen captured as shown in figure 17.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         566               56.6    %
Incorrectly Classified Instances       434               43.4    %
Kappa statistic                          0.2371
Mean absolute error                      0.196
Root mean squared error                  0.3409
Relative absolute error                 79.9616 %
Root relative squared error             97.4413 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.733    0.550    0.621      0.733    0.672      0.191    0.645     0.674     satisfactory
                 0.329    0.076    0.471      0.329    0.388      0.294    0.779     0.403     warning
                 0.451    0.154    0.471      0.451    0.461      0.301    0.766     0.453     Good
                 0.000    0.000    0.000      0.000    0.000      0.000    0.788     0.031     Greate Distinction
                 0.029    0.006    0.143      0.029    0.048      0.049    0.831     0.184     Distinction
Weighted Avg.    0.566    0.352    0.537      0.566    0.545      0.227    0.704     0.552

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 404  59  87   0   1 |   a = satisfactory
 110  56   4   0   0 |   b = warning
 120   4 105   0   4 |   c = Good
   4   0   6   0   1 |   d = Greate Distinction
  13   0  21   0   1 |   e = Distinction
```

Figure 17: Multilayer perceptron with LR=0.1 and M= 0.4 using 6 selected attributes

Experiment five

This experiment is conducted to evaluate the performance of MLP using 6selected attributes with attribute selection section. The values of learning rate and momentum are 0.1 and 0.9 respectively. The result of the experiment is screen captured as shown in figure 18.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         588               58.8    %
Incorrectly Classified Instances       412               41.2    %
Kappa statistic                          0.2535
Mean absolute error                      0.1872
Root mean squared error                  0.3401
Relative absolute error                 76.337  %
Root relative squared error             97.2238 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.799    0.584    0.627      0.799    0.702      0.234    0.659     0.673     satisfactory
                 0.253    0.055    0.483      0.253    0.332      0.261    0.760     0.424     warning
                 0.438    0.123    0.520      0.438    0.476      0.336    0.774     0.499     Good
                 0.000    0.000    0.000      0.000    0.000      0.000    0.748     0.024     Greate Distinction
                 0.086    0.010    0.231      0.086    0.125      0.122    0.807     0.207     Distinction
Weighted Avg.    0.588    0.360    0.557      0.588    0.559      0.256    0.709     0.567

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 440  43  65   0   3 |   a = satisfactory
 123  43   4   0   0 |   b = warning
 121   3 102   0   7 |   c = Good
   5   0   6   0   0 |   d = Greate Distinction
  13   0  19   0   3 |   e = Distinction
```

Figure 18: Multilayer perceptron with LR=0.1 and M= 0.9 using 6 selected attributes

Experiment six

The last experiment is conducted to evaluate the performance of a MLP classifier; values of learning rate and momentum are 0.4 and 0.8 respectively. Only the 6 selected attributes are considered. The result of the experiment is screen captured as shown in figure 19.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         597               59.7    %
Incorrectly Classified Instances       403               40.3    %
Kappa statistic                          0.2708
Mean absolute error                      0.1884
Root mean squared error                  0.3475
Relative absolute error                 76.8407 %
Root relative squared error             99.3471 %
Total Number of Instances             1000

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.806    0.584    0.629      0.806    0.706      0.243    0.649     0.666     satisfactory
                 0.341    0.064    0.523      0.341    0.413      0.332    0.767     0.398     warning
                 0.386    0.100    0.539      0.386    0.450      0.324    0.748     0.490     Good
                 0.000    0.000    0.000      0.000    0.000      0.000    0.644     0.019     Greate Distinction
                 0.143    0.011    0.313      0.143    0.196      0.193    0.796     0.186     Distinction
Weighted Avg.    0.597    0.356    0.572      0.597    0.571      0.272    0.697     0.555

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 444  49  55   0   3 |   a = satisfactory
 110  58   2   0   0 |   b = warning
 131   4  90   0   8 |   c = Good
   6   0   5   0   0 |   d = Greate Distinction
  15   0  15   0   5 |   e = Distinction
```

Figure 19: Multilayer perceptron with LR=0.4 and M= 0.8 using 6 selected attributes

The following table (table 7) summarizes the performance results obtained from all the above six experiments.

Table 7: Experimental results

No	Multilayer perceptron with 10 fold cross validation	Learning rate	Momentum	precision	Recall	F-measure	MAE	Accuracy	Training time
1	25 attributes	0.2	0.9	0.547	0.497	0.481	0.2015	49.7%	35.58
2	25 attributes	0.4	0.7	0.574	0.589	0.578	0.1662	58.9%	35.54
3	25 attributes	0.5	0.3	0.593	0.606	0.598	0.1632	60.6%	37.47
4	Selected 6 attributes	0.1	0.4	0.537	0.566	0.545	0.196	56.6%	2.9
5	Selected 6 attributes	0.1	0.9	0.557	0.588	0.559	0.1872	58.8%	2.29
6	Selected 6 attributes	0.4	0.8	0.572	0.592	0.571	0.1884	59.7%	2.73

Result Discussion

Table 7 summarizes the performance of the MLP algorithm executed using the dataset having 25 attributes and the dataset having only the selected 6 attributes by applying CFS method.

In the first three experiments, where all available attributes were used, the MLP classifier at the learning rate and momentum of 0.5 and 0.3 respectively outperformed MLP classifier at every deferent possible combinations of learning rate and momentum when looking at accuracy, precision, Recall, F-measure and MAE. Classification accuracy of the classifier is 60.60 %. That is the algorithm has classified around 60 of the instances to the correct class out of 100 resulting in 60.6 % successful classifications. Total training time taken by the algorithm is 37.47 seconds.

When looking at last three experiments, where 6 selected attributes (SEX, CH, ASHW, PIA, BBR and CLASS attribute) were used, the MLP classifier at the learning rate and momentum of 0.4 and 0.8 respectively, MLP classifier outperformed at every deferent possible combinations of learning rate and momentum when looking at accuracy, precision, Recall, F-measure and MAE. The classifier has classification accuracy of 59.7 %. It classified around 59 of the instances to the correct class out of 100 and the total training time taken by MLP algorithm is 2.73 seconds.

When considering Performance of the two MLP classifiers that perform better using all attributes and 6 selected attributes, in terms of accuracy, precision, recall, F-measure and MAE, the former had better performance in terms of all specified performance metrics at learning rate and momentum of 0.5 and 0.3 respectively. MLP classifier using all attributes has better performance than MLP algorithm using 6 selected attributes. This shows attributes which were not selected by CFS are also important for predicting AP of the students. The model developed using all the 25 attributes classify more instances to the correct class than the model developed using the 6 selected attributes.

When training time is considered, MLP algorithm using all available attributes and classification accuracy of 60.6%took 37.4 seconds and MLP using the 6 selected attributes took 2.73 seconds to train the model. This clearly shows that MLP using all attributes took more training time to build the model as compared with the time taken to build the model by MLP using 6 selected attributes. Attribute selection using CFS does not improve performance of the classifier it only reduces training time, but training time is not the main metric for evaluating the performance of prediction model.

Following the above result discussion, the best performing model is identified. The best model is the one developed using all available attributes that has a classification accuracy of 60.6. Thus, this model is used for predicting AP of students in teacher training colleges. The next step is model comparison. In model comparison, the models having better performance using all available attributes and using 6 selected numbers of attributes are compared with the models developed using Naive Bayes and Decision tree (J48) using same dataset.

Finally, the model developed using all the 25 attributes and has classification accuracy of 60.06% is saved with file name of “APMODEL”. The model saving process is screen captured as shown in figure 20.

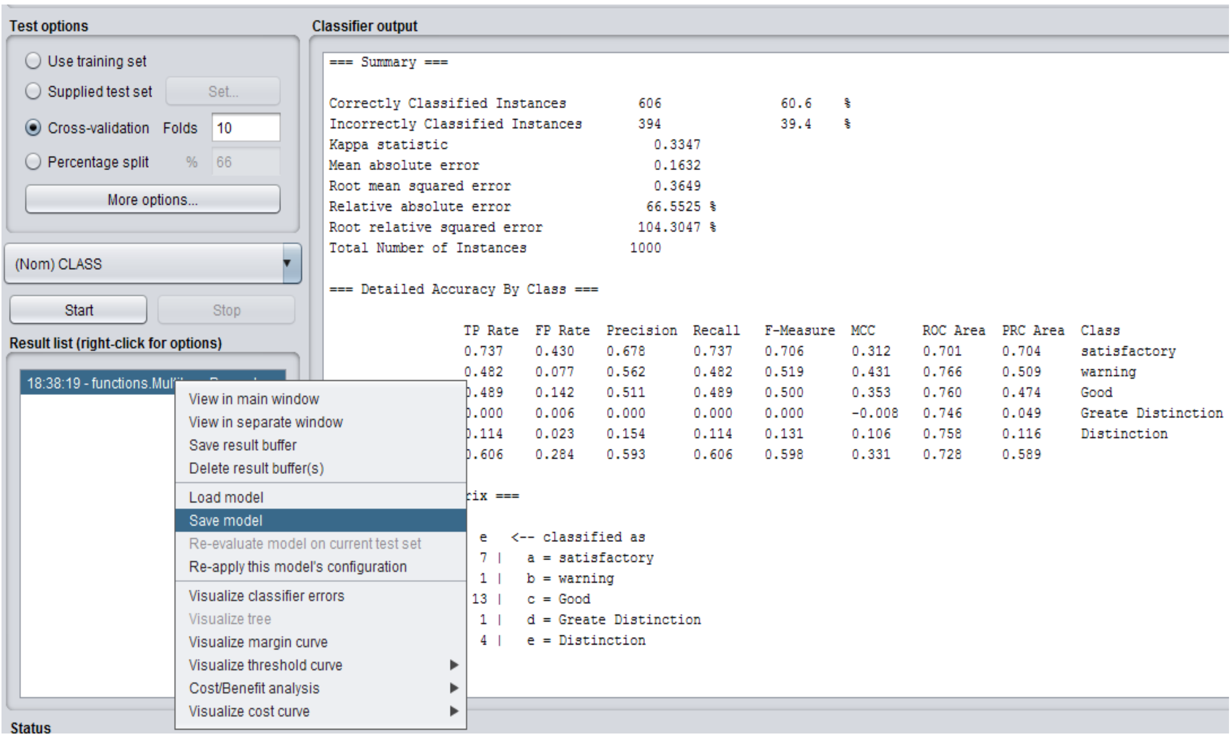


Figure 20 : Model saving setup

4.2. Model evaluation experiments

This section discusses the performance evaluation results of three machine learning algorithms, Multilayer perceptron, Naïve Bayes and J48 using same dataset prepared to predict academic performance. Performance evaluation is used to evaluate a specific classification or clustering algorithm so that the one with the best performance can be selected. Different algorithms have different performance level on same data. One algorithm can be a best choice for some dataset and other can have the worst performance evaluation on some other data sets. The performance of one classier differs from another due to several reasons such as the size of the dataset, quality of dataset, presence of outliers, presence of noise, presence of redundant attributes, over fitting and under fitting(Moustafa, Nassef, & Salah, 2019). All listed reasons show classifiers performance depends greatly on the characteristics of the data to be classified. Therefore, pre-processing methods are applied to increase the quality of the dataset. This plays a great role to improve the performance of the classifier.

In this thesis, model evaluation is done using five most commonly used evaluation metrics. These are: classification accuracy, precision, recall, F-measure and mean absolute error. To perform the model evaluation *weka.Analys.experment.performtest* component of WEKA that is designed to compare classification algorithms performance is used. The performance of MLP classifier is compared with respect to J48 and Naïve Bayes algorithms. Two algorithms are selected for comparison because of J48 decision tree and Naïve Bayes classification algorithms are fast and provides high prediction accuracy in predicting student's academic performance (Mayilvaganan, & Kalpanadevi, 2014; AkinsolaJ et al,.2017).

Experiments

The following two experiments are conducted to compare the performance of MLP classifiers with Naïve Bayes and Decision tree (J48) using all the 25 attributes and 6 selected attributes. Classification accuracy, Precision, Recall, F-measure and mean absolute error are used to compare the performance of the three classification algorithms.

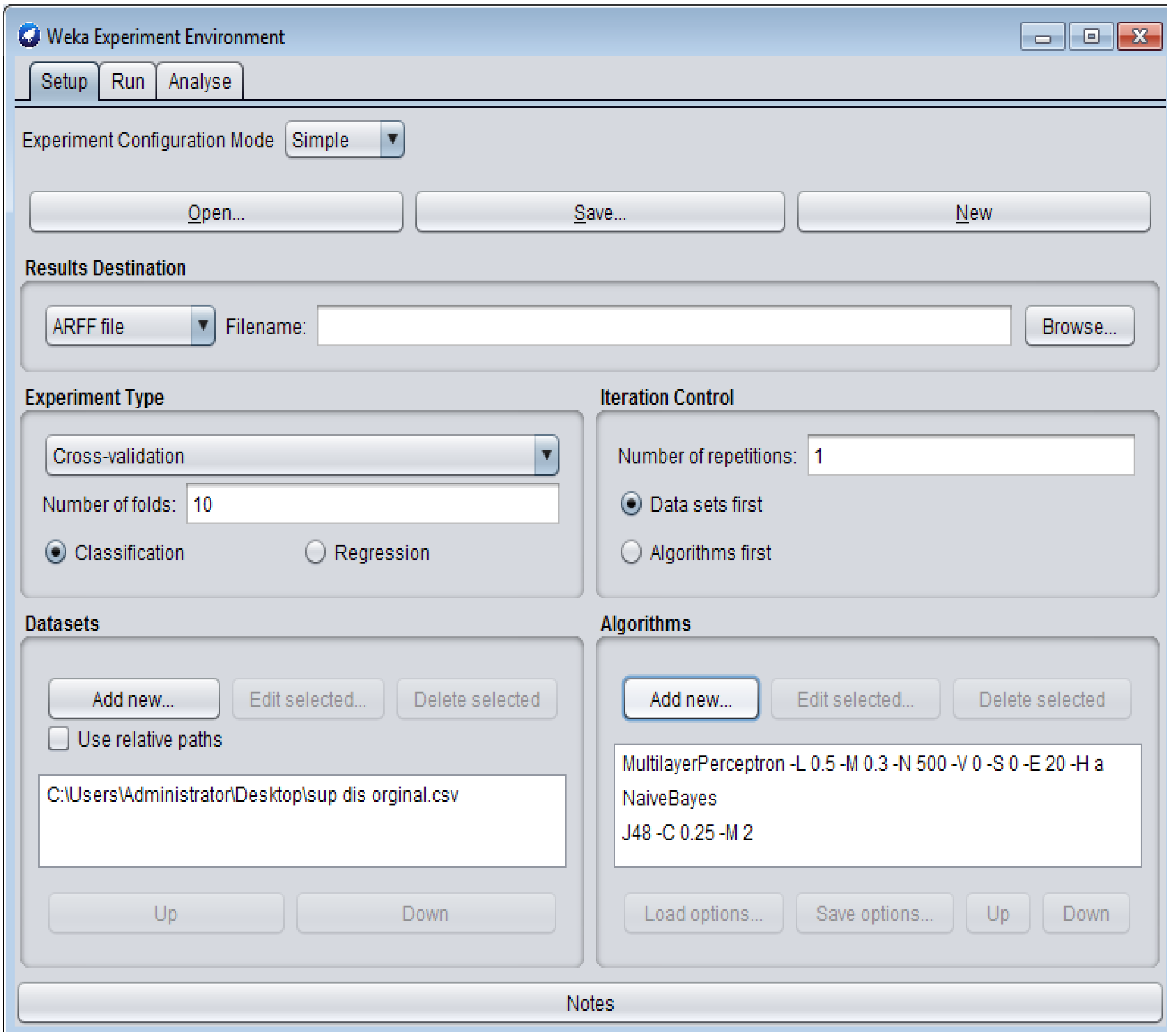


Figure 21: Model evaluation setup

Experiment one

The first experiment is conducted to compare the performance of MLP classifier with Naive Bayes and Decision tree (J48). For MLP, the values of learning rate and momentum are set to 0.5 and 0.3 respectively. The comparison is made using all the 25 attributes.

Experiment two

In the second experiment the values of learning rate and momentum for MLP classifier are 0.4 and 0.8 respectively. The performance of MLP is compared with Naive Bayes and Decision tree (J48) classifier using 6 selected attributes (SEX, CH, ASPW, PIA, BBR and CLASS attribute) in attribute selection section. The result of model comparison experiment is screen captured as follows.

Table 8 shows the values obtained From above two experiments by comparing MLP, NaiveBayes and J48 using various performance evaluation parameters. The first row represents results from first experiment and second row for second experiment.

Table 8: performance of three data mining algorithms

10 Fold cross validation	**Multilayer perceptron**					**Naïve Bayes**					**J48**				
Performance measures	Precision	Recall	F-measure	Accuracy in %	MAE	Precision	Recall	F-Measure	Accuracy %	MAE	Precision	Recall	F-Measure	MAE	Accuracy %
All 25 attributes	0.60	0.61	0.60	60.6	0.16	0.57	0.54	0.54	53.6	0.2	0.53	0.57	0.54	0.2	56.8
Selected 6 attributes	0.57	0.60	0.56	59.7	0.19	0.58	0.59	0.58	59.1	0.2	0.57	0.59	0.54	0.21	58.7

Result discussion

In the first experiment, where all available attributes were used, the MLP classifier outperformed Naive Bayes and J48 decision tree when looking at accuracy. Classification accuracy of the MLP classifier is 60.60 %. That is, the algorithm has classified around 60 of the instances to the correct class out of 100 resulting 60.6 % successful classifications. Classification accuracy of Naive Bayes and J48 decision tree are 53.6 % and 56.8 % respectively. When comparing accuracy in the second experiment, where 6 selected attributes were used, and MLP classifier outperform Naive Bayes and J48 decision tree. MLP classifier classified 59 of the instances to the correct class out of 100 and resulting 59.7 % successful classifications. Naive Bayes and J48 algorithms had classification accuracy of 59.1% and 58.7%respectively. From the above discussion for both cases, MLP can be said to have better classification accuracy than Naive Bayes and J48.

In terms of other performance measure such as average Precision; when all available attributes were used, the values of precision for MLP, Naive Bayes and J48 are 0.60, 0.57 and 0.53respectively. Clearly, from the result, MLP has the highest precision value than Naive Bayes and J48 Decision tree. That is, the lowest numbers of false positive errors were committed by MLP classifier and J48 decision tree clearly does the worst with maximum number of false positives relative to the two classifiers mentioned above. When six selected attributes were used, precision of MLP, Naive Bayes and j48 are 0.57, 0.58 and 0.57

respectively. From the result it is observed that MLP and J48 classifiers performed the same with equal number of highest false positives, but Naive Bayes clearly performed with lowest number of false positives relative to the other two classifiers. The result shows, in terms of precision, MLP outperformed Naive Bayes and J48 classifier when all available attributes were used and Naive Bayes outperformed MLP and J48 when six selected attributes were used.

When looking at Recall, where all available attributes were used, recall values for MLP, Naive Bayes and J48 decision tree, are 0.61, 0.54 and 0.57 respectively. Clearly from the result, MLP has the highest True positive but Naive Bayes and J48 have low True positive rate. That is, there are large positive examples misclassified as the negative class by Naive Bayes and J48 than MLP. When 6selected attributes were used, the values of recall for MLP, Naive Bayes and J48 are 0.60, 0.59 and 0.59 respectively. This value shows MLP has highest true positive rate than Naive Bayes and J48 classifiers. Naive Bayes and J48 classifiers perform the same with equal number of positive examples misclassified as the negative class. The result shows that for both cases MLP outperformed Naive Bayes and J48 when looking at Recall.

When looking at F-measure of the above two experiments, where all the 25 attributes were used, F-measure of MLP, Naive Bayes and J48 are 0.60, 0.54 and 0.54 respectively. The MLP classifier has highest value for F-measure which ensures that both Precision and Recall are reasonably high but Naive Bayes and J48 Decision Tree has same value of F-measure.For the case where six selected attributes were used the values of F-measure for MLP, Naive Bayes and J48 are 0.56, 0.58 and o.54 respectively. This value shows Naïve Bayes has highest value of F-measure which ensures that both Precision and Recall are reasonably high but MLP and J48 has lowest values. Therefore, in terms of F-measure MLP outperformed Naive Bayes and J48 where all available attributes were used and Naive Bayes outperformed MLP and J48 where six selected attributes were used.

When MAE is considered, where all available attributes were used the values of MAE for MLP, Naive Bayes and J48 are 0.16, 0.2 and 0.2 respectively. The result clearly shown that at an average lowest number of errors committed by MLP classifier than Naive Bayes and J48 Decision. When six selected attributes were used, MAE of MLP, Naive Bayes and j48 are 0.19, 02 and 0.21 respectively. From the result, it is observed that MLP committed lower number of errors on average than Naive Bayes and J48 Decision tree. However, Naïve Bayes and J48 committed the same number of errors on average.MAE results show that MLP outperformed Naive Bayes and J48 for both cases.

Therefore, based on the above result discussion it is confirmed that the model generated by MLP classification algorithm outperformed Naïve Bayes and J48 where all attributes were used. But for the case where 6 selected attributes were used all the three classifiers performs relatively the same.

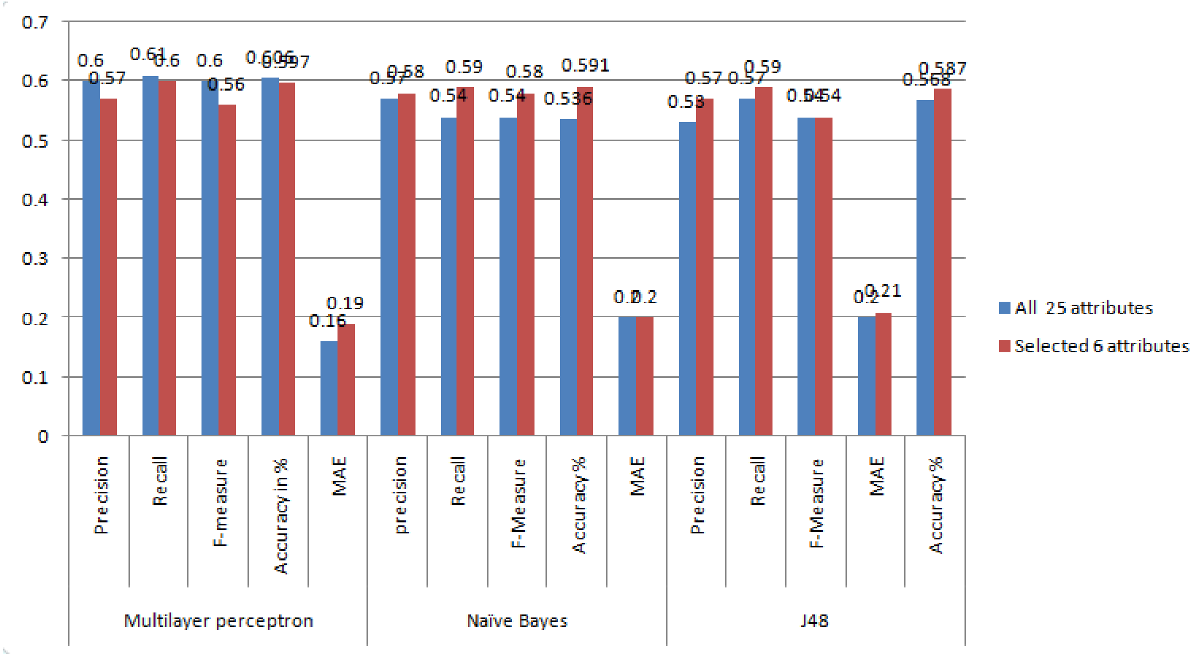


Figure 22: graph representation of performance evaluation of MLP, Naive Bayes and J48

The following graph (see figure 23) visualizes the performance of MLP, Naive Bayes and J48 Decision Tree classifiers, where all available attributes and 6 selected attributes were used. The graph clearly show that MLP classifier had better performance where all available attributes were used and Bayes and J48 Decision Tree performed better when 6 selected attributes were used. From this result we can concluded that MLP algorithm is suitable for a dataset having large number of attributes.

The following tree is obtained by visualizing the result of J48.From the tree, the first and most important attribute to predict AP of the students is Average study hours per week (ASHW) or reading habit of the students.

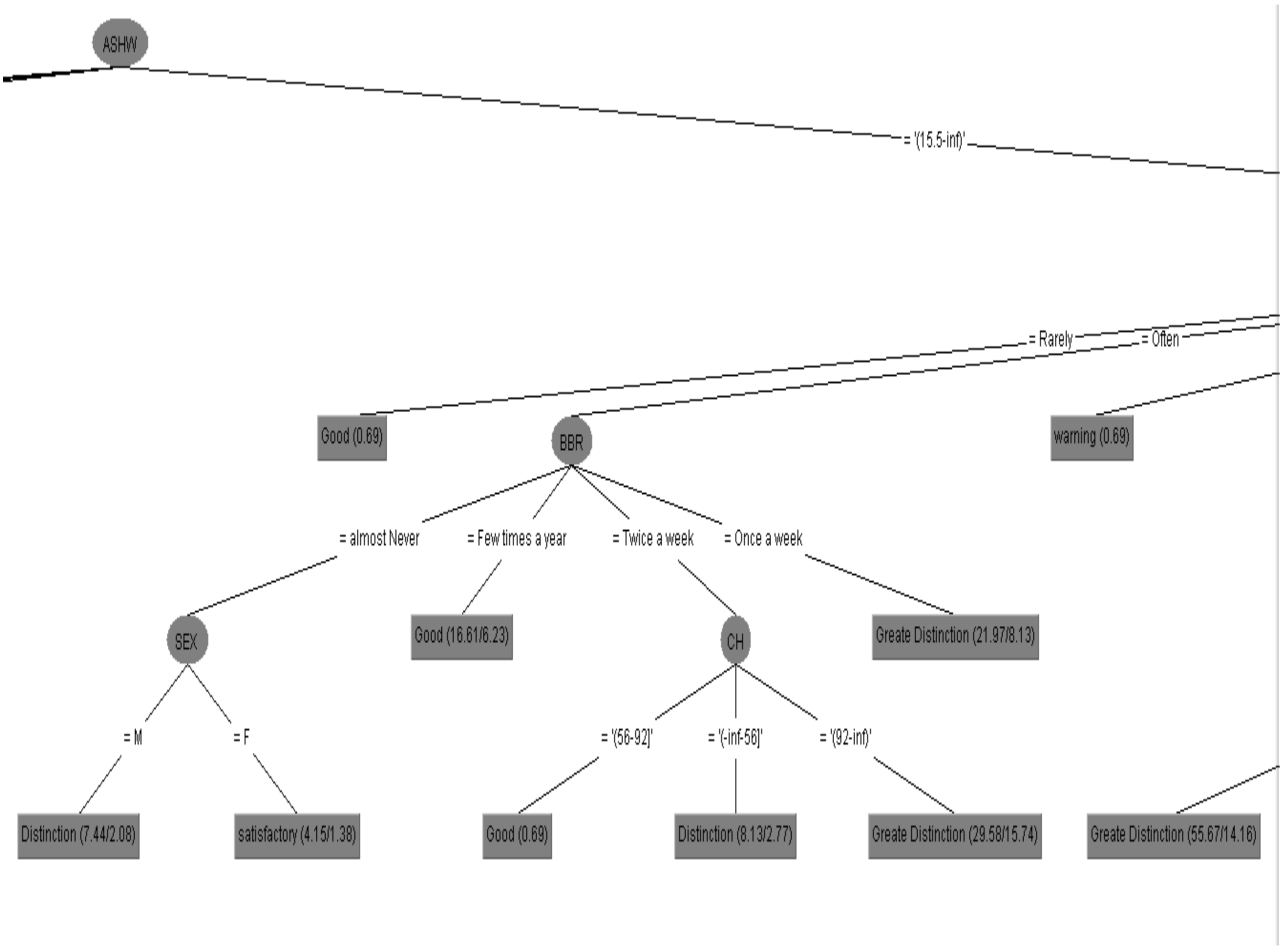


Figure 23: tree by J48 using 6 attributes

A decision rule is a simple IF-THEN statement consisting of a condition and a prediction. Decision rules follow a general structure: IF the conditions are met THEN make a certain prediction and it is the most interpretable prediction model.

Some of the decision rules generated from the above tree are listed and explained below:

IF (ASHW = (0-8.5] then

Return (status= satisfactory)

Else if (ASHW = (8.5-15.5] and BBR = "almost Never" or " Few times a year")

Return (status= Satisfactory)

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week" and PIA = "Rarely" or "Often")

Return (status= Good)

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week" and PIA = "Copy from friends ")

Return (status=Satisfactory)

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week” and PIA = “Sometimes" and CH = (56-92])

Return (status= warning)

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week” and PIA = “Always" and CH = (56-92])

Return (status= Distinction)

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week” and PIA = “Always" and CH = (0-56))

Return (status= Good)

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week” and PIA = “Always" and CH >92 and SEX="M")

Return (status= "Distinction")

Else if (ASHW = (8.5-15.5] and BBR = "Twice a week” and PIA = “Always" and CH >92 and SEX="F")

Return (status= "Satisfactory")

Else (ASHW = (8.5-15.5] and BBR = "Once a week")

Return (status= "Satisfactory")

According to above specified rules all students who reads 8.5 hours or less per week or has higher probability of getting status of satisfactory. A student who reads 8.5 to 15.5 hours and never borrow books or borrow books few times a year also will get Satisfactory. All rules generating from the tree are interpretative in to meaningful form.

By using the above decision rule it is possible to predict AP of the college student. This thesis used java application software to develop prototype that will going to be used tested by the students to know their academic success or failure.

The Java code for the prototype is:

```
import java.awt.GridLayout;

import javax.swing.*;

import java.awt.*;
```

```
import java.awt.event.*;
class Aperformance extends JFrame implements ActionListener
{
  JButton SUBMIT;
  JPanel panel;
  JComboBox combo,combo1,combo2,combo3,combo4;
  JLabel label1,label2, label3, label4,label5;
     Aperformance()
     {
   label3 = new JLabel();
   label3.setText("Select Your sex:");
   String[] items2 = {"select one", "Male","Female"};
   combo2 = new JComboBox(items2);
   label1 = new JLabel();
   label1.setText("Approximate Your Average study hours per week?");
   String[] items = {"select one", "less than 4.5 hours","between 4.5 and 8.5 hours","between
         8.5 and 15.5 hours","above 15.5 hours"};
   combo = new JComboBox(items);
   label2 = new JLabel();
   label2.setText("Select the Total course credit hours taken");
   String[] items1 = {"select one", "less than 56 hours","between 56 and 92 hours","above 92
hours"};
   combo1 = new JComboBox(items1);
   label4 = new JLabel();
   label4.setText("How often do you borrow books from your school laibrary?");
   String[] items3 = {"select one","Once a week","Twice a month","Few times a year
","Never almost never"};
```

```
combo3 = new JComboBox(items3);
label5 = new JLabel();
label5.setText("How often do you do your assignments regularly?");
String[] items4 ={"select one", "Always","Often","Sometimes","Rarely","copy from friends"};
combo4 = new JComboBox(items4);
SUBMIT=new JButton("SUBMIT");
panel=new JPanel(new GridLayout(0,1));
panel.add(label3);
panel.add(combo2);
panel.add(label1);
panel.add(combo);
panel.add(label2);
panel.add(combo1);
panel.add(label4);
panel.add(combo3);
panel.add(label5);
panel.add(combo4);
panel.add(SUBMIT);
add(panel,BorderLayout.CENTER);
SUBMIT.addActionListener(this);
setTitle("ACADEMIC PERFORMANCE PERIDICTOR FOR TEACHER TRAINING COLLEGE STUDENTS");
  }
@Override
public void actionPerformed(ActionEvent ae)
{
```

```
   Object ASHW;
    ASHW= combo.getSelectedItem();
    Object CH;
    CH = combo1.getSelectedItem();
    Object sex;
    sex = combo2.getSelectedItem();
    Object BBR;
    BBR = combo3.getSelectedItem();
    Object PIA;
    PIA= combo4.getSelectedItem();
 if ( ASHW.equals("less than 4.5 hours"))
 {
JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
      "Message",JOptionPane.INFORMATION_MESSAGE);
 }
 else if(ASHW.equals("between 4.5 and 8.5 hours"))
 {
 JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6,
Satisfactory",
      "Message",JOptionPane.INFORMATION_MESSAGE);
 }
 else if(ASHW.equals("between 8.5 and 15.5 hours"))
{
  if(BBR.equals("almost Never")&& BBR.equals("Few times a year"))
  {
    JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6,
Satisfactory",
```

```
 "Message",JOptionPane.INFORMATION_MESSAGE);
}
else if(BBR.equals("Twice a week"))
  {
  if(PIA.equals("Rarely")&& PIA.equals("Often"))
  {
  JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25 , Good",
   "Message",JOptionPane.INFORMATION_MESSAGE);
  }
  else if(PIA.equals("Copy from friends"))
  {
  JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
   "Message",JOptionPane.INFORMATION_MESSAGE);
  }
  else if(PIA.equals("Sometimes"))
  {
     if(CH.equals("between 56 and 92 hours"))
          {
        JOptionPane.showMessageDialog(this, "your GPA will lessthan 2, Warning",
         "Message",JOptionPane.INFORMATION_MESSAGE);
          }
     else
       {
  JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
```

```
"Message",JOptionPane.INFORMATION_MESSAGE);
      }
    }
    else
    {
      if(CH.equals("between 56 and 92 hours"))
        {
        JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
        "Message",JOptionPane.INFORMATION_MESSAGE);
        }
      else if(CH.equals("less than 56 hours"))
        {
    JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25 , Good",
    "Message",JOptionPane.INFORMATION_MESSAGE);
        }
      else
       {
        if(sex.equals("Male"))
          {
          JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
             "Message",JOptionPane.INFORMATION_MESSAGE);
          }
          else
           {
```

```
        JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
          "Message",JOptionPane.INFORMATION_MESSAGE);


                  }
      }
      }
      }
       else
      {
  JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
          "Message",JOptionPane.INFORMATION_MESSAGE);
      }
   }
else
{
 if(PIA.equals("Rarely"))
 {
  JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25 , Good",
       "Message",JOptionPane.INFORMATION_MESSAGE);
 }
 else if(PIA.equals("Often"))
 {
    if(BBR.equals("almost Never"))
    {
       if(sex.equals("Male"))
```

```
{
 JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
   "Message",JOptionPane.INFORMATION_MESSAGE);
}
else
{
 JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
  "Message",JOptionPane.INFORMATION_MESSAGE);
}
}
else if(BBR.equals("Few times a year"))
{
JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25 , Good",
 "Message",JOptionPane.INFORMATION_MESSAGE);
}
else if(BBR.equals("Twice a week"))
{
 if(CH.equals("between 56 and 92 hours"))
 {
 JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25 , Good",
 "Message",JOptionPane.INFORMATION_MESSAGE);
 }
 else if(CH.equals("less than 56 hours"))
 {
```

```
        JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
            "Message",JOptionPane.INFORMATION_MESSAGE);
        }
        else
        {
        JOptionPane.showMessageDialog(this," Your GPA will be above 3.5,Great Distinction",
            "Message",JOptionPane.INFORMATION_MESSAGE);
        }
    }
    else
    {
  JOptionPane.showMessageDialog(this," Your GPA will be above 3.5,Great Distinction",
            "Message",JOptionPane.INFORMATION_MESSAGE);
    }
 }
 else if(PIA.equals("Copy from friends"))
 {
 JOptionPane.showMessageDialog(this, "your GPA will lessthan 2, Warning",
             "Message",JOptionPane.INFORMATION_MESSAGE);
 }
 else if(PIA.equals("Sometimes"))
 {
    if(sex.equals("Male"))
    {
        if(BBR.equals("almost Never"))
```

```
        {
        JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
            "Message",JOptionPane.INFORMATION_MESSAGE);
        }
        else if(BBR.equals("Twice a week"))
        {
        JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25, Good",
           "Message",JOptionPane.INFORMATION_MESSAGE);
        }
        else
        {
        JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
           "Message",JOptionPane.INFORMATION_MESSAGE);
        }
    }
    else
    {
     JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
           "Message",JOptionPane.INFORMATION_MESSAGE);
    }
  }
  else if(ASHW.equals("above 15.5 hours"))
  {
```

```
if(BBR.equals("almost Never"))
  {
   if(CH.equals("between 56 and 92 hours"))
   {
   JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
        "Message",JOptionPane.INFORMATION_MESSAGE);
   }
   else if(CH.equals("less than 56 hours"))
   {
      if(sex.equals("Male"))
      {
       JOptionPane.showMessageDialog(this," Your GPA will be above 3.5, Great Distinction",
        "Message",JOptionPane.INFORMATION_MESSAGE);
      }
      else
      {
     JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
      "Message",JOptionPane.INFORMATION_MESSAGE);
      }
   }
  else
   {
    if(sex.equals("Male"))
    {
```

```
        JOptionPane.showMessageDialog(this," Your GPA will be above 3.5, Great Distinction",
            "Message",JOptionPane.INFORMATION_MESSAGE);
         }
         else
         {
        JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
           "Message",JOptionPane.INFORMATION_MESSAGE);
         }
        }
  }
    else if(BBR.equals("Few times a year"))
    {
       if(CH.equals("between 56 and 92 hours"))
       {
      JOptionPane.showMessageDialog(this," Your GPA will be above 3.5, Great Distinction",
            "Message",JOptionPane.INFORMATION_MESSAGE);
       }
       else if(CH.equals("less than 56 hours"))
       {
      JOptionPane.showMessageDialog(this," Your GPA will be between 2.6 and 3.25, Good",
           "Message",JOptionPane.INFORMATION_MESSAGE);
       }
       else
```

```
        {
        JOptionPane.showMessageDialog(this," Your GPA will be between 2 and 2.6, Satisfactory",
            "Message",JOptionPane.INFORMATION_MESSAGE);
        }
    }
  else
      {
   JOptionPane.showMessageDialog(this," Your GPA will be between 3.25 and 3.5, Distinction",
             "Message",JOptionPane.INFORMATION_MESSAGE);
    }
  }
else
 {
  JOptionPane.showMessageDialog(this," Please complete the form",
   "Message",JOptionPane.ERROR_MESSAGE);
        }
  }
  }
}
public class AperformanceDemo
       {
public static void main(String arg[])
      {
   try
   {
```

```
  Aperformance frame=new Aperformance();
  frame.setSize(600,500);
  frame.setVisible(true);
  }
  catch(Exception e)
 {JOptionPane.showMessageDialog(null, e.getMessage());}
  }
}
```

The following figure 24 shows interface of the prototype.

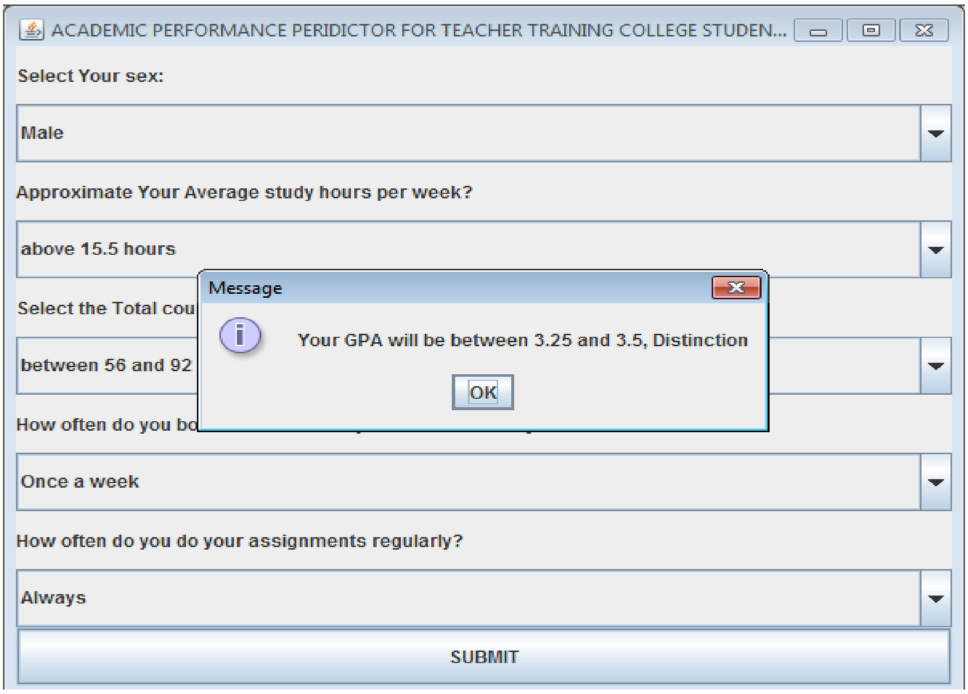


Figure 24: Academic performance prediction application

Students who wants to know their future academic performance is required to fill the form by providing all the necessary information and click submit. If the information is enough to predict the AP of the student the system show the performance otherwise the system asks the student to fill all information required for prediction.

Chapter Five

5. Conclusion and Recommendation

5.1. Conclusion

From the study, the researcher observed that Educational data mining is the field that can assists teacher training colleges to improve institutional effectiveness and student academic achievement. On the other hand, this research showed how useful data mining can be used in teacher training colleges particularly to predict students' performance using academic related factors and their academic record. The study was conducted with purpose of identifying the factors that affect students' AP and contribute support to the students to take actions that can prevent students' dismissal, and to improve student learning abilities to graduate with good academic point.

The MLP classification algorithm was implemented using WEKA tool to develop the AP prediction model. The dataset was generated by collecting data from the students and additional data extracted from the college registrar office. The database has 25 attributes and 1000 instances. The result of attribute selection using correlation based feature (CFS) method reveals that sex of the student, Total courses credit hours; average study hours per week, performing in assignment and book borrowing rate are attributes that determine the AP of student. Further, MLP classification algorithm is compared with Naive Bayes and J48 decision tree algorithms. The result of model comparison proved that MLP classifier outperformed Naïve Bayes and J48 Decision tree where all available attributes were used. Thus, the Model developed using MLP classification algorithm is selected to better predict students' future academic performance. The results presented here help the teacher training colleges by allowing the instructors and students of the college in identifying AP predictors and it can also act as an early warning system for predicting the students' low academic achievement.

5.2. Recommendation

The study requires further research because of its two main limitations. First limitation is the data collection: for this study data was collected from only current active students of 2018/19 academic year. Future studies are encouraged to include more Current and historical records in order to improve the prediction of students' performance. Second, the items on research questionnaire covered only small number of attributes for data collection. Including more questions help to identify many different attributes that negatively or positively affects AP of students. So, further researches are encouraged to include more attributes in order to identify unidentified factors that contribute to academic achievement of the college student.

References

Affendey, L.S., Paris, I.H., Mustapha, N., Sulaiman, M. N., &Muda, Z. (2010). Ranking of influencing factors in predicting students' academic performance. Information Technology Journal, 9: 832-837.DOI: 10.3923/itj.2010.832.837

Agri, M.(2019).Root Mean Square Error Calculator. Retrieved from Agrimetsoft.comwebsite:https://agrimetsoft.com/calculators/Root%20Mean%20Square%20Error.

Akessa, G. M., & Dhufera, A. G. (2015).Factors That Influences Students Academic Performance: A Case of Rift Valley University, Jimma, Ethiopia. Journal of Education and Practice, Vol.6, No.22.

Alasadi, S.(2017).Review of data processing technique in data mining. Journal of engineering and applied science, 12(16).4102-4107

Allen, J., Robbins, S., Casillas, A., & Oh, I.(2008). Third-year college retention and transfer: effects of academic performance, motivation, and social connectedness. Research in Higher Education, 49(7), 647–664.

Amjad, A. S., Mostafa, A. and Khaled, S.(2020).Mining Student Information System Records to Predict Students’ Academic Performance. pp. 229–239.

Amoo, M. A., Alaba, O. B., & Usman, O. L. (2018). Predictive modeling and analysis of academic performance of secondary school students: Artificial Neural Network approach. International Journal of Science and Technology Education Research, 9(1), 1–8.

Anuradha, C., & Velmurugan, T.(2015).A Comparative Analysis on the Evaluation of Classification Algorithms in The Prediction of Students Performance. Indian Journal of Science And Technology, 8(15).

Arora, N., & Saini, J. R. (2013). A Fuzzy Probabilistic Neural Network For Student's Academic Performance Prediction. International Journal of Innovative Research in Science, Engineering and Technology, 2(9):4425-4432.

Baharin, S. H., Othman, R. &Azizan, N.(2015).STATISTICAL ANALYSIS ON THE DETERMINANTS OF STUDENTS' ACADEMIC ACHIEVEMENT: A STUDY IN UITM JOHOR.

Brownlee, J. (2019, August 22). How To Perform Feature Selection With Machine Learning Data In Weka. Retrieved from Machine Learning Mastery website: https://machinelearningmastery.com/perform-feature-selection-machine-learning-data-weka/

Cawley, G, C. & Talbot, N L, C. (2010).On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research 11.2079-2107

Chen, M, Han, J., & Yu, P. (1996). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866–883.

Cheng, J. (2017). Data-Mining Research In Education.

El-Halees, A. M. (2009). Mining students' data to analyze e-Learning behavior: A Case Study.

Endalamaw Yigermal, M. (2017).The Determinants of Academic Performance of under Graduate Students: In the Case of Arba Minch University Chamo Campus.Advances in Sciences and Humanities, 3(4), 35.

F.Y, O., AkinsolaJ.E., T., Awodele, O., HinmikaiyeJ., O., Olakanmi, O., & Akinjobi, J. (2017).Supervised Machine Learning Algorithms: Classification And Comparison.

Hall, M. A. (2011).Data Mining. (Third Edition).

Hall, M.A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009).

The WEKA data mining software: an update. SIGKDD Explorations, 11, 10-18.

Hall, M. A., Smith, L. A. (1999). Feature Selection for Machine Learning: Comparing a Correlation Based Filter Approach to the Wrapper. In: FLAIRS conference, 235-239.

Hussain, A. (2003).Effect Of Guidance Services on Study Attitudes, Study Habits and Academic Achievement Of Secondary School Students.

Huang, S., & Fang, N. (2013). Predicting Student Academic Performance In An Engineering Dynamics Course: A Comparison Of Four Types Of Predictive Mathematical Models. Computers & Education, 61, 133–145

Kalyani, F. (2016).Data mining with Artificial Neural Network: A perspective

Kirmani, N. S, & Siddiquah, A. (2008). Identification and Analysis of Factors Affecting Students Achievement In Higher Education.2nd International Conference On Assessing Quality In Higher Education.

Kohavi, R. (1995). A Study Of Cross-Validation And Bootstrap For Accuracy Estimation And Model Selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, Vol. 2, Montreal, 20-25 August 1995, 1137-1145.

Kokemuller, N. (2019). Examples of Academic Challenges In College. Retrieved From Education. seattlepi. comwebsite: https://education.seattlepi.com/examples-academic-challenges-college-1289.html

Mayilvaganan, M., & Kalpanadevi, D. (2014).Comparison Of Classification Techniques for Predicting the Performance of Students' Academic Environment. International Conference on Communication and Network Technologies, 113-118.

Mcleod, S.(2019).Independent, Dependent, and Extraneous Variables | Simply Psychology. Retrieved June 8, 2019, from Simplypsychology.org website:

Moustafa, M., Nassef, M., & Salah, A.(2019).Categorization of Factors Affecting Classification Algorithms Selection. SSRN Electronic Journal.

https://doi.org/10.2139/ssrn.3438184

Murphy, D. (2017). Decision Trees for the Beginner.

Mushtaq, I. & Khan, S. N.(2012).Factors Affecting Students' Academic Performance. Volume 12 Issue 9 Version 1.0.

Noble, J., Roberts, W.L., & Sawyer, R.(2006).Student Achievement, Behavior, Perceptions, and Other Factors Affecting ACT Scores. ACT Research Report Series, 2006-1.

Undavia, J. N., Dolia, P.M., & Shah, N. (2013).Prediction of Graduate Students for Master Degree based on Their Past Performance using Decision Tree in Weka Environment.

Olani, A. (2017).Predicting First Year University Students' Academic Success. Electronic Journal of Research in Education Psychology, 7(19).

Osmanbegovi, E. & Suljić, M. (2012) .DATA mining approach for predicting student performance. Journal of Economics and Business, Vol. X, Issue 1.

Peffers, K., Tuunanen, T., Gengler, E. C., Rossi, M., Hui, w., Virtanen, v. & Bragg,J.(2006). The design science research process: A model for producing and presenting information systems research.

Ramageri, B.M. (2011). Data Mining Techniques And Applications. Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305

Rehman, N.(2017).Data Mining Techniques Methods Algorithms and Tools. International Journal of Computer Science and Mobile Computing, Vol.6 Issue, pg. 227-231

Raychaudhuri, A. D., Manojit, S. S. &Majundra, B. G. (2010). Factors affecting Student's academic performance: A case study in agartala municipal concial area. Bangladesh e-journal of sociology , vol.7, Number.2.

Roberts, K. L., & Sampson, P. M.(2011). School Board Member Professional Development and Effects On Student Achievement. International Journal of Educational Management, 25(7), 701–713.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601–618.

Saa, A.A. (2016).Educational Data Mining & Students' Performance Prediction. International Journal of Advanced Computer Science and Applications, Vol.7. No. 5.

Sebastian, S. (2016). Performance Evaluation By Artificial Neural Network Using WEKA. International Research Journal of Engineering and Technology. Volume: 03 Issue: 03.

Tiwari, S. (2019, October 21). Complete Guide to Machine Learning Evaluation Metrics. Retrieved February 16, 2020, from Medium website: https://medium.com/analytics-vidhya/complete-guide-to-machine-learning-evaluation-metrics-615c2864d916

Voicu, H. (2008). The cerebellum: An incomplete multilayer perceptron? Neurocomputing, 72(1–3), 592–599. https://doi.org/10.1016/j.neucom.2007.11.013

Witten, I. H., Frank, E., & Hall, M. A. (2011).Data mining : practical machine learning tools and techniques. Amsterdam: Elsevier.

Yadav, S.K., & Pal, S. (2012). Data Mining Application in Enrollment Management: A Case Study.

Yoseph, S. (2013).Preparatory Class Average Score and University Entrance Exam Result as Predictors of College Performance: The Case of 2011/2012 Kotobe University College Degree Program Students.

Yigermal, M.E. (2017). Determinant of Academic Performance of Under Graduate Students: In the Cause of Arba Minch University Chamo Campus.

Yohannes, E., & Ahmed, S. (2018). Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study. International Journal of Computer Applications, 180(40), 39–47.

Appendix I: Sample Students record from registrar

The data is copped from excel.

Hawassa Collage of Teacher Education)									
Stream :Biology Lab									
(Regular Program)									
Id No	Full Name	Sex	Tch	Cgp	Cgpa	Status			
BILR/864/09	AbenezerTesfaye	M	108	259	2.40	pass	V.G.Distinction	1	3.9
BILR/865/09	AdimasBoje	M	108	263	2.44	pass	G.Distinction	2	3.5
BILR/866/09	AndualemGezahegn	M	108	292	2.70	pass	Distinction	0	3.25
BILR/867/09	AsheduHabibSemano	M	108	235	2.18	pass	Pass	31	1.96
BILR/868/09	AsmamawuAbebeDawit	M	108	267	2.47	pass	Warning	1	1.95-1.90
BILR/869/09	BereketBeyeneBezabih	M	108	250	2.31	pass	Re-admision	0	1.89-1.85
BILR/870/09	BirhanBizualemTilahun	M	108	246	2.28	pass	Dismisal	1	1.84
BILR/871/09	BrihanuChefk	M	108	260	2.41	pass	Total	36	
BILR/872/09	DenekeWorku	M	108	331	3.06	pass			
BILR/874/09	ElfineshLankamo	F	108	228	2.11	pass			
BILR/875/09	EsayasMekuria	M	108	245	2.27	pass	SUMMERY		
BILR/876/09	FevenBefkaduTegegn	F	108	339	3.14	pass	M	27	
BILR/877/09	GeneneYilmaDingato	M	108	244	2.26	pass	F	9	
BILR/878/09	GuyeKusse	M	108	254	2.35	pass	Total	36	
BILR/881/09	KalkidanBerako	M	108	254	2.35	pass			
BILR/882/09	KarayuKayeso	M	108	389	3.60	Great distniction			
BILR/885/09	ManonsaKakamo	M	108	314	2.91	pass			
BILR/887/09	MariyamChanna	F	108	242	2.24	pass			
BILR/888/09	MarkosJugisaHamaro	M	108	245	2.27	pass			
BILR/889/09	MedhanitDemirewWorku	F	108	197	1.82	Dismisal			
BILR/890/09	MindayeYetera	M	108	244	2.26	pass			
BILR/891/09	MulukenAderaFayiso	M	108	333	3.08	pass			
BILR/892/09	SamiroAnatoDebesa	M	108	291	2.69	pass			
BILR/893/09	ShemsuHusenNunencho	M	108	273	2.53	pass			
BILR/894/09	ShibruBogaleBorche	M	108	250	2.31	pass			
BILR/896/09	SisayTilahun	M	108	259	2.40	pass			
BILR/898/09	TameneHammayaGammada	M	108	250	2.31	pass			
BILR/899/09	TamenechTafeseBulke	F	108	207	1.92	Warning			
BILR/900/09	TayeTageleGimbo	M	108	336	3.11	pass			
BILR/901/09	Terechatadeledangiso	M	108	304	2.81	pass			
BILR/902/09	TizitawBogaleAdamo	M	108	284	2.63	pass			
BILR/903/09	TokasheDoketeriSalfa	M	108	293	2.71	pass			
BILR/904/09	WorkneshGeremu	F	108	416	3.85	Very great distniction			
BILR/906/09	YemisrachAshagreAsefa	F	108	245	2.27	pass			
BILR/907/09	ZenawitAbaynehTacho	F	108	223	2.06	pass			
BILR/1240/08	MarishetTakeleAbdi	F	108	390	3.61	Great distniction			

AppendixII: Questionnaire

HAWASSA UNIVERSITY

FACULTY OF INFORMATICS

DEPARTEMENT OF COMPUTER SCIENCE

Dear respondent,

Thank you very much for your willingness to take time to respond to this research questionnaire. The study is being conducted by a postgraduate student at Hawassa University, Faculty of informatics, and Department of computer science.

It is all about questions pertaining to predictors of Academic performance in teacher training colleges. To this end, it intends to gather information from Hawassa College of teacher education students.

The participation is fully on voluntary basis, and your accurate and frank responses are imperative for the successful accomplishment of the study.

Please be assured that your responses will be treated in a strictly confidential manner, and the results will be used only for the purpose of this research, presented only at aggregate level without focusing on individual student.

Thank you in advance.

FirehiwotGetachew

Please circle a response or fill in the blank SPACES FOR each question. When you are done, return the questionnaire to representative from your class.

Your ID.NO ____________________

No	Academic, and personal parameters	
1.	Your Gender	A. Female B. Male
2.	Previous high school attended?	A. private B. public
3.	Do you take day, Evening or weekend classes?	A. Day B. Evening C. Weekend
4.	Year level	A. 1st Year B. 2nd Year C. 3rd Year
5.	What is your Department of study in College?	________________
6.	What percent of classes do you generally attended?	________________%
7.	Approximate your average study hours per week.	_________________
8.	How often do you do your assignments regularly?	A. Always B. often C. sometimes D. rarely E. copy from friends
9.	How often do you borrow books from your school library?	A. once a week B. twice a month C. Never or almost never D. few times a year
10.	How well do you use the learning facilities provided by the college (computer library, phy lab...)	A. once a week B. twice a month C. Never or almost never D. few times a year
11.	Are your environment motivating to improve your studies	A. Yes B. No
12.	Do you listen attentively to the lecture of your teacher?	A. Always B. often C. sometimes D. rarely E. Never
13.	Do you have an interest of being a teacher?	A. Yes B. No
14.	Rate your participation in any extra Curricular activities in the college?	A. Low B. average C. good
15.	Rate your interest in sport activities?	A. Low B. average C. good
16.	Do your home is far from the college?	A. Yes B. No
17.	Do you experience financial problem?	A. Always B. often C. sometimes D. Rarely E. Never
18.	Do you get any financial support from your family or relatives?	A. Always B. often C. sometimes D. Rarely E. Never
19.	Does your mobile phone destruct you while studying?	A. Always B. often C. sometimes

		D rarely E. Never
20.	Do you use social media networks?	A. Yes B. No
21.	How long do you spend on social networking cites during a day?	A. 10m' B. 30m' C. 1-2 hrs D.2-3 hrs E.3-4 hrsF.4-5hrs G. 5+hrs
22.	Your family educational status	Mother ___________Father _________ A. Primary school (1-6 grades) B. Secondary school (7-12) grades C. Diploma D. Degree E. Master's Degree F. Other specify____________________
23.	Your family economic status	A. Upper class B. Medium class C. lower class
24.	Your current college cumulative Grade point average(CGPA)	A. 2.00-2.2 B. 2.2-2.4 C. 2.4-2.6 D. 2.6-2.8 E. 3.0-3.2 F. 3.2-3.4 G. 3.4-3.6 G. 3.6-3.8 G. 3.8-4.0