

Design of Academic Performance Prediction System Using Multi-Layer Perceptron

G. Jayanthi¹ and Dr.V.Ramesh²

*¹M.Phil (CS) Research Scholar, ²Assistant Professor
Department of Computer Science and Applications
Sri Chandrasekharendra Saraswathi Viswa Maha Vidyalaya University
Kanchipuram – 631 561, Tamil Nadu
jayashanthiguna@gmail.com, rameshvpothy@gmail.com*

ABSTRACT

The amount of data stored in educational database increasing rapidly. These databases contain the useful information for predict of students performance. Educational data mining is used to study the data available in the educational field and bring out the hidden knowledge from it. Factors other than intelligence which affects the academic performance of the students were analyzed in this study. The main objective of this study is to use data mining methodologies to predict academic student's performance based on the student profile. The classification algorithms Naive bayes, Multilayer perceptron, SMO, J48, REPTree are applied on the student's data to predict their performance. Our prediction model predicts the marks scored by a student based on various socio economic environmental conditions of the students. The study revealed that the Multi-Layer Perceptron is more accurate than the other algorithms. Our model will helps the weaker students to predict their marks in advance to take better action to improve their standard to get more marks.

Keywords: Classification Techniques, Decision tree, Multi-Layer perceptron, Neural Networks, REPTree.

I. INTRODUCTION

Data Mining is defined as extracting the information from the huge set of data. Education is a key factor for achieving a long-term economic progress. Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using

those methods to better understand students. Traditional issues such as enrolment management and placement opportunities motivate the higher education institutions to search for better solutions. The primary data were collected from the higher school students. The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, to pass information on to others. In this study we collected primary data from school students through questionnaire method. A total of 200 datasets of the year 2014 from four different schools in Kanchipuram district of Tamil nadu were collected. We present the results based on characteristics such as scalability, accuracy to identify their characteristics in a DataMining tool-WEKA. The classification algorithms Naive bayes, Multilayer perceptron, SMO, J48, REPTree are applied on the student's data to predict their performance.

1.1 STUDY OBJECTIVES

The main objective of this study is to use data mining methodologies to predict academic performance of high school (SSLC) student based on the student profile. Data mining provides many tasks that could be used to study the student performance.

- To predict the SSLC marks using classification algorithm.
- To compare different classification algorithms.
- To study the relationships among different factors deciding student performance.
- To design a model to help and identify the weaker students to help educational institution to take decision.

1.2 STATEMENT OF THE PROBLEM

Most data-based modeling studies are performed in a particular application domain. Hence, domain-specific knowledge and experience are usually necessary in order to come up with a meaningful problem statement. Many application studies tend to focus on the data-mining technique at the expense of a clear problem statement. This study is mainly focusing on predicting the grade of high school (SSLC) student's performance. The first step requires the combined expertise of an application domain and a data mining model. In practice, it usually means a close interaction between the data-mining expert and the application expert.

1.3 SCOPE OF THE STUDY

The Scope of the Study is to compare the algorithms in classifications methods, namely Naive Bayes, Multilayerperceptron, SMO,J48, REPTree, This paper deals with predict the SSLC student performance based on the classification techniques. To predict the SSLC student final grade using classification algorithms and neural ware.

2. LITERATURE REVIEW

Abeer Badr El Din Ahmed, *et al.*[1], made an attempt to use data mining methodologies to study student's performance. In this paper, the classification task is used to predict the final grade of students and as there are many approaches that are used for data classification, the decision tree (ID3) method is used. Sohil Pandya *et al.* [2] tried to examine and investigate various methods of classifications like Decision

Trees (DT), Naïve Bayes (NB), k-Nearest Neighbour (kNN), Feed Forward Neural Networks (FFNN) and Support Vector Machine (SVM) to identify the best fit methods among them for the University domain. All the above mentioned algorithms were implemented using WEKA, an Open Source Software which consists of a collection of machine learning algorithms for data mining tasks. P.V.Praveen Sundar [3] made an attempt to predict the student academic performance and make a comparative study on bayesian network classifiers, through that we compute which classifier predicts more students when compared to other classifiers. In this paper, student's information like Previous Semester Performance, Attendance, Seminar , Assignment marks, Internal marks, and whether the student has attend any Co-curricular Activities are collected from students to predict the performance at the end of the semester examination.

Shaeela Ayesha *et al* [4] used Bayesian Classification Method as a data mining technique and concluded that students grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habits, family annual income and students family status were highly correlated with the student academic performance. Umamaheswari *et al* [5] made an attempt to provide proper placement facilities to the student. For this reason they categorize the student based on their skill level. Skill level is ranked in the form of CGPA grade taking into account end semester marks and also based skill test. The institution allows the eligible students to attend, the recruitment process based on the companies criteria. The eligible students were categorized using clustering technique. This makes easier to select and reduce time as well. Yadav *et al* [6] conducted a study using classification tree to predict student academic performance using students' gender, admission type, previous schools marks, medium of teaching, location of living, accommodation type, father's qualification, mother's qualification, father's occupation, mother's occupation, family annual income and so on. In their study, they achieved around 62.22%, 62.22% and 67.77% overall prediction accuracy using ID3, CART and C4.5 decision tree algorithms respectively.

3.1 EDUCATIONAL DATA MINING

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students. In educational system predicting students academic performance, classification of students performance in computer programming course according to learning style, students living area ,students food habit, investing the similarity and difference between school.

3.2 Proposed Model

In this study author collected recent real world data from various school students. Four schools were randomly selected from Kanchipuram district, Tamil Nadu. A sample of 200 students was taken from a group of schools. The primary data was collected using a questionnaire. Which include questions (i.e. with predefined options) related to several personal, socio-economic, psychological and school related variables that were expected to affect student performance. The data set used in this

study was obtained from a student's data used in one of the educational institutions. Initially size of the data is 200 records. In this step data stored in different tables was joined in a single table after joining process errors were removed.

4.1 METHODOLOGY

Classification analysis is the organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a *training set* where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects. Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. In our case study we used classification methods to represent logical rules of student final grade.

PREDICTION: There are two major types of predictions: one can either try to predict some unavailable data values or pending trends, or predict a class label for some data. The latter is tied to classification. Once a classification model is built based on a training set, the class label of an object can be foreseen based on the attribute values of the object and the attribute values of the classes. Prediction is however more often referred to the forecast of missing numerical values, or increase/decrease trends in time related data. The major idea is to use a large number of past values to consider probable future values.

WEKA TOOL: The WEKA workbench is a collection of machine learning algorithms and data pre-processing tools. Weka is a collection of machine learning algorithms for data mining tasks. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. The weka GUI chooser launches the weka's graphical environment which has six buttons: Simple CLI, Explorer, Experimenter, Knowledge Flow, ARFF viewer and log. The basic idea of classification algorithm is to construct the decision tree employing a top-down, search through the given sets to test each attribute at every tree node.

WEKA has a very flexible combination of search and evaluation methods for the dataset's attributes. Search methods include Best-first, Ranker, Genetic-search, etc. Evaluation measures include InformationGain, GainRatio, ReliefF, etc. For this research Selecting attributes using ChiSquareAttributeEval and Ranker method for getting the high potential chi square values for that we get the highest accuracy to predict the performance of the SSLC students grade. The ChiSquaredAttributeEval and ranker is such a function. Given a data table that contains attributes and class of the attributes, we can measure homogeneity of the table based on the classes. To determine the best attribute by using the ChiSquaredAttributeEval and Ranker

method. The index used to measure degree of impurity is Entropy. In this study author use ChiSquared AttributeEval and for this the search method is Ranker. Then select the top 18 attributes from those 46 attributes for this study.

5. RESULTS AND DISCUSSIONS

The result of classification techniques and multilayer perceptron classifier predicts the majority of class in training data. It predicts the mean for numeric value and mode for nominal class. The confusion matrix is given which gives the accuracy of solution to the classification problem. The Chi square test is used to testing the hypothesis and there is any difference between the attributes of student related variables. The classification matrix predicts the accurate students grade. The Naïve bayes, MLP, SMO, J48, REPTree classification matrix are given below.

Table 5.1: Prediction Performance of Classification Algorithms

Algorithms/Grades	NAÏVE BAYES	MLP	SMO	J48	REPTREE	NEURAL WARE
O	75.00	87.50	75.08	75.00	62.50	75.00
A	58.33	83.33	50.00	58.33	58.33	75.00
B	50.00	58.33	58.33	33.33	50.00	58.33
C	57.14	71.42	42.85	71.42	57.14	42.85
D	33.33	66.66	33.33	66.00	33.33	66.66
E	00.00	00.00	00.00	00.00	00.00	00.00

Table 5.2: Comparisons of Algorithms

	NAÏVEBAYES	MLP	SMO	J48	REPTREE
Training set	68.66%	98%	92.66%	78%	59%
Testing set	60%	76%	58%	56%	56%
Accuracy	64.33%	87%	75.33%	67%	57.5%

From the results it is proven that Multilayer Perception (MLP) algorithm is most appropriate for predicting student performance. MLP gives 62.12% prediction which is relatively higher than other algorithms.

Table: 5.3 Comparison between WEKA and Neural ware

	MLP	NEURAL WARE
Training Set	98 %	76 %
Testing Set	76 %	66 %
Accuracy	87 %	71 %

We conclude our analysis by comparing the generalization accuracy of the neural networks that were trained with other well-known classifiers such as decision tree, Bayesian networks, classification rules and support vector machines. In order to evaluate the performance of MLP, we have applied the same dataset to the neural network learning algorithm. We found that both MLP and neural network classifies the testing data with same accuracy. But MLP is taking more accuracy for training set than neural network. Result shows that student living area, medium of instruction, parents education, school area, interest in sports these factors more affect the performance of the students. The empirical results show that we can produce short but accurate prediction list for the student by applying the predictive models to the records of incoming new students. This study will also work to identify those students which needed special attention.

5.1 HIGH SCHOOL (SSLC) MARKS PREDICTION SYSTEM

A high school marks prediction system is developed using .NET framework. In this model we used classification rules using Multi-Layer Perceptron is used to predict the expected grade of a student for the given different social and economic factors pertaining to the student.

The screenshot shows a web application titled "SSLC MARKS PREDICTION SYSTEM". The form contains the following fields and values:

INTROB	ENGINEER	WATCHY	ONE	TUTORIN	ONE
BLOODGROUP	AB-	STUDY HOME	FOUR	MEDIUM	ENGLISH
TYPE SC	PRIVATE	TRANSPORT	BIYCLE	COMPUTER	YES
GRPSTUDY		STUFASIZE	FIVE	FOODHAB	NOWAYS
NETACS	TWO	CAREHOME	MOTHER	PARENT	BOTH
FRI/SUB	WATHY	COMMUNITY	HIC	INTSPORT	NO

Below the form is a "SUBMIT" button and a "PREDICTED SSLC GRADE" field showing the value "A". At the bottom, there is a copyright notice: "© 405-988-8490-4058-5100-490C-500-500-5100-988Eshaw 200" and "Developed By: Jayanthi".

Figure:5.1 SSLC Marks Prediction System

This model is used to help and identify the weaker SSLC students. The results states that the prediction has helped the weaker students to improve and betterment in the result. This model will help the institute to accurately predict the performance of the students, to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time.

6. CONCLUSION

In this paper, classification method is used on student's database to predict the student's performance on the basis of student's database. This study will help the student's to improve the student's performance, to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time. Multi-Layer perceptron gives more classification accuracy for predicting the

student's grade. For future work, the experiment can be extended with more distinctive attributes to get more accurate results, useful to improve the students learning outcomes. Also, experiments could be done using other data mining algorithms to get a broader approach, and more valuable and accurate outputs. Some different software may be utilized while at the same time various factors will be used.

REFERENCES

1. Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby ,“Data Mining: A prediction for Student's Performance Using Classification Method”, *World Journal of Computer Application and Technology* 2(2): 43-47, 2014
2. Sohil Pandya &Dr. Paresh V. Virparia, “Comparing the Applications of Various Algorithms of Classification Technique of Data Mining in an Indian University to Uncover Hidden Patterns”, *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, Issue 5, May 2013 ISSN: 2277 128X.
3. P.V.Praveen Sundar , “A Comparative Study for Predicting Student's Academic Performance using Bayesian network Classifier”,*IOSR Journal of Engineering (IOSRJEN)* e-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 2 (Feb. 2013).
4. Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, “Data mining model for higher education system”, *European Journal of Scientific Research*, Vol.43, No.1, pp.24-29, 2010.
5. Umamaheswari. K and S. Niraimathi, “A Study on Student Data Analysis Using Data Mining Techniques”, *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, Issue 8, August 2013 ISSN: 2277 128X.
6. K. Yadev & Pal., S. 2012. *Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification*, *World of Computer Science and Information Technology (WCSIT)*, 2(2), 51-56.

