



DATA MINING CLASSIFICATION TECHNIQUES TO PREDICT THE PLACEMENT OF A STUDENT BASED ON ACADEMIC DETAILS

Dr.K.Sarojini* **Mrs.C.MariaSheeba****

**Asst.Professor and Head of the Department, Department of Computer Science,LRG Government Arts College for Women,Tirupur.*

***Guest Lecturer,Department of Computer Science,Government Arts College,Udhagamandalam.*

Abstract

Data mining is the system of extracting the needed information from large amount of data. Data mining is an effective device for educational intervention. There is an increasing research in the usage of data mining in education. One of the new research groups of application includes studying the students overall performance. This paper deals with a comparative study of various data mining techniques for the overall performance analysis of the educational facts and test which set of the algorithm is most beneficial for classifying student's performance. This will help the students to perceive themselves and prepare accordingly in an efficient way.

Keywords-EDU, Students performance, Classification Techniques, Prediction.

I. Introduction

Majority of students in higher education join in a course where they can find a secure job. Therefore taking a wise profession choice concerning the placement after finishing a specific course is critical in a student's existence. An academic group consists of a huge quantity of student records. Therefore locating patterns and traits on this large quantity of records is not a difficult task. In the field of education, teachers or professionals always classify their students based on knowledge, motivation and conduct. Assessing examination result is also a classification method, where the mark is decided according to certain assessment criteria. For this study the student's data about their previous and current academic records have collected and then applied different classification algorithm for analyzing the academic performance which will be useful for training and placement.

1.1 Educational Data Mining

Mining in educational environment is called Educational Data Mining. Educational Data Mining is involved in developing new methods to discover knowledge from the educational database and might used for decision making in educational machine. Data mining has various techniques such as Classification, Clustering, Prediction, Association Rules, Decision Trees, Neural networks and many others. Predicting the performance of the student's placement is a challenging task. Educational data mining uses many techniques such as Decision making trees, Naive Bayes, K-means, Multilayer perception and many others. Using these methods, many kinds of useful information can be predicted. For this study the students data about their previous and current academic records have collected and then applied different classification algorithm for analyzing the academic performance which will be useful for Training and placement.

II. Related Works

The author [1] has used ID3 and C4.5 algorithms to evaluate student's performance. The intention behind this research is to traverse the data mining techniques which are used for the improvement of student's performance. Italso identifies the best for the current environment. They stated that the



behavior and performance of both the algorithms differed. Among the various classifications the performance of C4.5 stamps its superiority.

The data set used for [2] is obtained from VBS Purvachal University,Janupur(Uttar Pradesh).The details of the students were collected under the following category: Students Sex, MCA Results, Seminar Performance, Lab Work, communication skill and graduation background. The algorithms used for classification is Naive Bayes, Multilayer Perceptron and J4.8.The result of this study states that the Naïve bayes classification is the best algorithm with anaccuracy of 85.15% and it has lowest average error at0.028% compared to others.

This research paper [3] observed the students data from Engineering Institute with the following fields such as S.No,Name,Branch,S.S.L.C.,H.Sc,B.ETech passing percentage and final grade. The different classification algorithms were applied for analysis the student’s academic performance for training and placements.The data of 50 students usedin this study.The result obtained from the various data mining algorithms through BayesNet,Naïve Bayes,Multilayer Perceptron,IBI,Decision table and PART Classification. The result concludes that IBI Classifier is most suitable method for this type of student dataset.

This paper[4] discussed the use of classification algorithms in educational data. Classification algorithms are applied on previous year computer application students data.J48,REPTREE,Naive Bayes,BayesNet and Multilayer Perceptron algorithms were used. This predictive model provides way to predict whether a new student will place or not. The experimental results show that the prediction rates are not uniform among algorithms. The range varies from 72-80%.Bayesian Classifiers perform very well in Predicting the placement.

The author [5] investigates the accuracy of NaiveBayes Simple, MultiLayer Perception, SMO, J48, REPTree, techniques for predicting student performance in the courses, to predict their placement at the end of the Final semester. The attributes such as marks in English, Maths, Programming language, Practical marks were collected. The seven semesters Maths, Programming, Practical marks are converted as a single Maths, Programming and Practical marks. After applying the classification techniques, they accurately predict the performance of the students and helped to find out the weak student. This will help to improve the performance of such students in the early stage. From the results it is proven that Multilayer Perception algorithm is most appropriate for predicting student performance.

Algorithm	Correctly Classified Instances	Absolute Mean Error
NaïveBayesSimple	83.1933	0.1852
Multilayer Perception	87.395	0.2002
SMO	84.0336	0.1597
J48	84.8739	0.2553
REPTree	84.8739	0.231

Table 2.1: Comparison of classification algorithms for students evaluation database

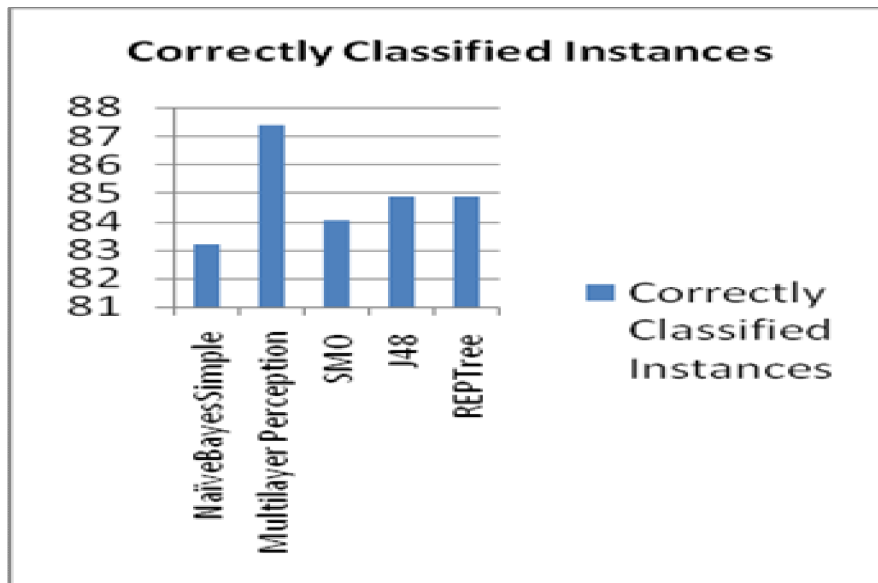


Fig2.1 Graphical representation of algorithms based on correctly classified instances

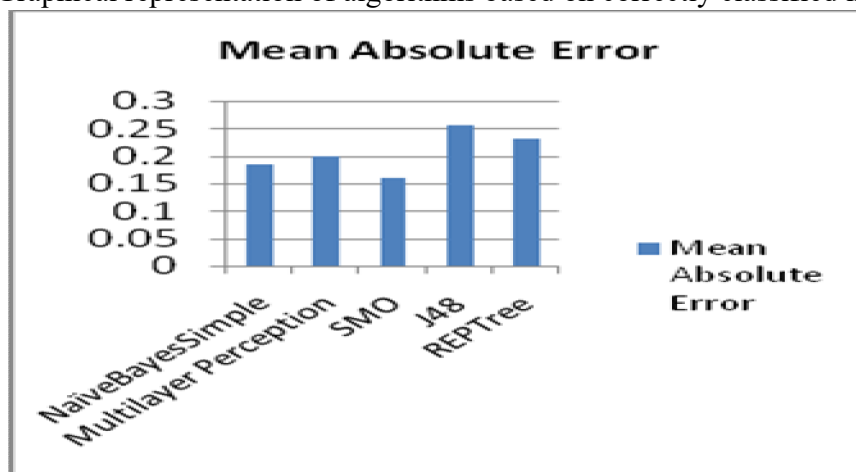


Fig 2.2 Graphical representation of algorithms based on mean absolute error

This paper [6] makes use of existing ML strategies so that it will expect students' performance in a distance learning system. It compares a number of art learning algorithms to find out which algorithm is more suitable not only to find the students performance appropriately, however also for use as an academic supporting device for tutors. Two experiments were executed with six algorithms. With the assist of machine-learning methods, the tutors are in a role to recognize which of their students will successfully complete a course with sufficiently accurate precision. The analysis of six algorithms has verified sufficient proof that the Naive Bayes is the maximum suitable algorithm for the construction of a software program tool.

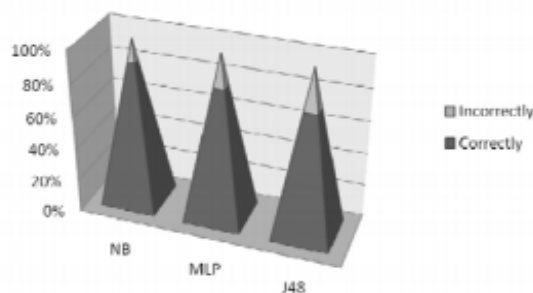
This reference[7] investigated, the classification techniques to predict the final grade of students especially the decision tree method ID3 is used. The research work is carried out with the following attributes: Department, seminar performance HSD, Labtest grade, assessment, attendance, and final grade mark. This study is used to assist the student's to improve the overall performance and also to



perceive those students who wanted special interest to reduce failing ration and taking appropriate motion at right time.

This paper[8] collected the details of the students from various colleges under the category of academic and personal skills. This data were used to analyze and predict the placement of the students by applying classification techniques such as decision tree algorithm (J48), K-Nearest Neighbor (IBk) and Naïve Bayes using one of the Data Mining tools known as WEKA. All the three classifier models were compared, and it was observed that the highest accuracy was shown by KNN classifier. This system is found to be more reliable and can be used to predict if the student can be placed or not.

Classification Results for Placement



The growth and the development of the country depend upon the quality of the education given to the students. Professional education plays a vital role in higher education. Data mining techniques were used to extract hidden knowledge in educational data that are used to predict the future of the students based on their academic and personal skills. The objective of this paper[9] is to predict about the knowledge of the students those who are willing to join the Masters of Computer Application course.

Databases are stored in the educational institutions by collecting huge data based on students academic as well as their extracurricular activities. This information was used to predict their performance in the forth coming semester also in their placements. This paper [10] analyzed the machine leaning algorithms such as K-Nearest neighbor methods and Naïve Bayes to predict the employability opportunity of the students as 'yes' or 'No'. Prediction is observed on the basis of concept level programming, personality development, mathematics, and advanced technical courses.

[11] A nation's economy is valued by the higher education offered by that nation. Predicting student employability based on their existing data can help the management to know who are at risk of unemployment and also to the students where they can improve themselves. This paper used data mining classification techniques and concluded that Empathy, Drive and Stress Management abilities are one of the major emotional feelings that affect the employment of the students.

This paper [12] provided a solution for predicting student's academic performance of first year bachelor students in computer science courses. Nearly 8 years of data were collected for this framework. Decision Tree, Naïve Bayes, and Rule Based classification techniques were applied to find the success level of the students. The authors observed that the Rule Based classification had the best classification accuracy.



III. Classification Algorithms

Classification is the most generally carried out statistics mining technique based on machine learning. Basically, classification is to classify each item in a set of data into one of the predefined sets of classes or groups. This method regularly employs the decision tree or neural network-based type algorithms. The data classification technique includes learning and classification. In learning the information are analyzed by different types of classifying algorithms to find the accuracy.

C4.5

The C4.5 algorithm is used in Data Mining as a Decision Tree. Advantages of C4.5 compared to different Decision Tree structures are the algorithm inherently employs one time Pass Pruning Process to mitigate over fitting. It has the ability of handling both discrete and continuous data and C4.5 can cope with the problem of incomplete facts very well.

Naïve Bayes Algorithm

Naive Bayes classifiers may be trained very successfully in a supervised gaining knowledge of placing. Naïve Bayes classifiers are quite scalable. Naive Bayes model is simple to construct and especially beneficial for very large data set.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Naïve bayes theorem converts the data set into a frequency table, Creates Likelihood table by finding the probabilities then use Naive Bayesian equation to calculate the posterior probability for each class. The elegance with the very best posterior probability is the final results of prediction.

IV. Conclusion

By analyzing the students performance using classification algorithm helps to identify the dropouts earlier and to take appropriate action to improve them. It also helps the faculty to identify the students those who need special attention.

References

1. C.Anuradha and T.VelMurugan "A Data Mining based Survey on student Performance Evaluation System" Classification Model of Prediction for Placement of Students Classification Model of Prediction for Placement of Students .
2. Ajay Kumar Pal and Saurabh pal "Classification model of Prediction for students placement of students"
3. Samrat Singh, Dr. Vikesh Kumar "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques" IJCSET Feb 2013



4. K.Manikandan,S.Sivakumar,M.Ahokvel “ Classification Model for Predicting Campus Placement performance class using data Mining Technique”,International Journal of Advance Research in Science and Engineering,March-18.
5. V.Ramesh,P.Parkavi,P.Yasodha “Performance Analysis of Data Mining Techniques or Placement Change Prediction “ International Journal of Scientific &Engineering Research Volume 2,Issue 8,August-2011
6. S. Kotsiantis, C. Pierrakeas, and P. Pintelas, —Prediction of Student’s Performance in Distance Learning Using Machine Learning Techniques, Applied Artificial Intelligence, Vol. 18, No. 5, 2004, pp. 411-426.
7. AbeerBadr 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.
8. Sumaira Asif, Tabrez Nafis, “Students Placement Prediction Using Classification Techniques,” International Journal of Computer Sciences and Engineering, Vol.7, Issue.4, pp.289-293, 2019
9. Neelam Naik and Seema Purohit,“ Prediction of Final Result and Placement of Students using Classification Algorithm”, International Journal of
10. G.Vadivu*, K.Sornalakshmi,“Applying Machine Learning Algorithms for Student Employability Prediction Using R”, Int. J. Pharm. Sci. Rev. Res., 43(1), March - April 2017; Article No. 11, Pages: 38-41 ISSN 0976 – 044X
11. Tripti Mishra , Dharminder Kumar, Sangeeta Gupta “Students employability prediction model through data mining”International Journal of Applied Engineering Research ISSN 0973-4562 Volume
12. Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, “The Prediction of Students’ Academic Performance Using Classification Data Mining Techniques”, Applied Mathematical Sciences, Vol. 9, 2015, no. 129, 6415 – 6426.