STUDENT PERFORMANCE EVALUVATOR USING DECISION TREE

Donamol Thomas

[4]*Asst. Professor of Christ College Puliyanmala, Kerala*

Abstract-Analyses data mining methods and techniques to student's data to construct a predictive model for student's performance prediction on final examination. The research on the educational fields that involves Data Mining techniques is rapidly increasing. Applying Data Mining methods in an educational background are known as Educational Data Mining.Initially collected the marks of series mark of one subject to predict the final grade in that subject. The internal marks are converted into 100 (percentage) to have accurate results. The model provides expected percentage of final examination in particular subject. It also helps students to know how many marks in the internal examination are required to get particular grade.

Keywords-Decision tree, prediction

I.INTRODUCTION

Classification techniques are using for education data mining. Educational data mining applied many techniques are K- nearest neighbour, decision tree, Naïve Bayes, Neural network, Fuzzy, Genetic and other techniques are applied in the field. These methods do not have the ability to reveal useful hidden information.

Predicting student performance in advance can help students and their teacher to keep track of progress of a student and thereby take action. In this model student performance is accessed by considering the internal exam marks to predict the final grade. Linear regression which is a statistical measure is used to predict the final grade of student.

Decision tree which is a classification technique is used to compare the student performance based on the training data set.

II.LITERATURE REVIEW

Students Performance Prediction Using Decision Tree Technique

The major objective of the proposed methodology is to build the classification model that classifies a students' performance. The classifiers, has been built by combining the Standard for Data Mining that includes student performance and finally application of data mining techniques which is classification in present study.
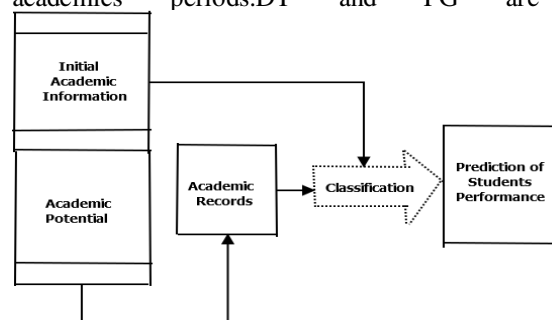
In other words, using this Decision tree algorithm, we wanted to be able to guide student towards achievement of good score that we felt they would enjoy doing.Tree-based methods classify instances by sorting the instances down the tree from the root to some leaf node. Each node in the tree specifies a test of some attribute of the instance and each branch descending from that node corresponds to one of the possible values for this instance.

The decision trees generated by C4.5 algorithm can be used for classification. At each node of the tree, C4.5 algorithm chooses the attribute of the data that most effectively splits its set of samples into subsets enriched into one class. Thenormalized information gain is chosen as splitting criterion. The attribute with the highest normalized information gain is chosen to make the decision in the algorithm. The C4.5 algorithm then persist on the smaller sub lists.

We also using Generalized Sequential Pattern mining algorithm for predicting the student's performance as pass or fail. Once the student is found at the risk of failure he/she can be provided guidance for improving performance.

Student performance prediction using DT algorithm and Fuzzy algorithm

This section introduces the student classification models to predict the student performance due to low academics performance that uses the student's initial academic information and the academic records of current academics periods.DT and FG are used.

Decision Tree Method for Predicting Students Academic Performance

The data collected was for a course CSC214 (Data Structures) which is a 2nd year course for NCE in computer science in Nigerian Colleges of Education. The course was chosen because of the familiarity of the researchers with the course and the gradesconsidered. A questionnaire was distributed to students to collect data about the other factors considered in the prediction such as the student's financial strength. The data collected from the result sheet was stored into SPSS Version 20 for analysis.

Table. Factors of the Prediction

Factor	Description	Attributes	SPSS value
Score	Students grade	>40 and <40	0 – 100
Status	Students status	Pass, Fail	1,2
Gender	Students gender	Male, female	1,2
Finance	Financial strength	Low, Medium, High	1,2,3
Motivation	Attitude to learning	Low, Moderate, High	1,2,3

The factors that are considered in predicting the performance of students in this study are:

- Scores: this is the total scores obtained by the students in the previous session of the course under consideration. The score's attributes are within the range of >40 and <40. The scores are however entered into SPSS as obtained from the results sheets (i.e. 0 – 100). For the purpose of this study,the classification was reduced from the six class classification of scores i.e. A, B, C, D, E, F to a two-class classification of Pass and Fail in order to reduce the margin of error for the research as it regards to the prediction

- Status: Status refers to the remark obtained for the scores of the students, i.e. „Pass" or „fail". In which case a score less than 40 is entered as fail while a score greater or equal to 40 is entered as pass. The status is coded as 1 or 2; where 1 stands for pass and 2 stands for fail.

- Gender: Gender is used to classify the students as either male or female. This factor is quite important because it helps to know the effect of gender of the students on predicting. This is in agreement with who established that gender plays a great role in the performance of female

- Students in a large public Turkish University. Though gender is considered as better predictor if we consider age and gender in theacademic performance of students.

- Finance: this refers to the financial status of students, this factor is important because it affects the stability of the students. Finance is an avenue through which the students are able to settle their bill. Students without adequate financial strength are affected adversely. The attributes of the factor include, High, Medium mad Low are they are coded in SPSS as 1, 2 and 3.

- Motivation: this refers to a factor that pushes a student to work hard or serves as reinforcement for the student; such as studying in groups, rewards and prices etc. the attributes are high, moderate and low. These are coded as 1, 2 and 3; 1 for low, 2 for moderate and 3 for high.

In building the predictive decision tree using the SPSS program based on the criteria in Table, the SPSS predictive approach used is the CHAID approach. The dependent variable is Score which implies that the study is concerned with predicting the score of the student using the previous score. The factors that influence that scores are listed under the independent variables as status finance and gender while the variable chosen as the influence variable is the motivation.

A. A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques

This study consists of three steps. The steps are data cleaning, feature selection and classification analysis. Data cleaning is important step that must be conducted before analysing the data to identify missing value, noisy and inconsistent data. The second step is Feature Selection. Feature selection algorithm can determine as the combination of a search technique to propose new feature subsets.The objective of feature selection is to select a subset of input attributes by eliminating features, which give less predictive information. This step is necessary to be applied in order to define the high influencing variables.Anuradha and Velmurugan as well as Ramaswami and Bhaskaran used several methods to conduct the feature selection in their study, and the five popular filter feature subset methods with Rank search or Greedy search method are Correlation-based Attribute evaluation, Gain-Ratio Attribute evaluation, Information-Gain Attribute evaluation, Relief Attribute evaluation, and Symmetrical Uncertainty Attribute. Those methods were also used in this current study. The last step was conducting the classification analysis using Bayesian Network and Decision Tree. The student's performance parameter that is used in this study is whether the student will be DO or not. The performance parameter to compare both algorithm that used in this study is Accuracy Rate. For the purpose of conducting the experiments, WEKA which is freely open source data mining software was used to implement the algorithm.

B. Student final grade prediction based on linear regression

In this examination, we connected linear regression based model to foresee the last grade for understudy specifically subject. The model is prepared utilizing signs of existing understudy in one subject. In this research the X variable is considered as average of internal examination marks. Moreover, the imprints are changed over into rates to increase appropriate precision of the framework.

In the present system the grade are converted in the non-numeric forms which are converted into numerical form as depicted in Table.

Grade	Grade Point range	Equivalent Grade Point
AA	>=80	80
AB	>=75 and <80	75
BB	>=70 and <75	70
BC	>=65 and <70	65
CC	>=60 and <65	60
CD	>=55 and <60	55
DD	>=50 and <55	50

To check the co-relational between the variable X and Y the co-relation coefficient r is calculated based on the following equations:

$$Z_x = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

$$Z_y = \frac{y - \mu_y}{\sigma_y} \tag{2}$$

$$r = \frac{Z_x Z_y}{n-1} \tag{3}$$

III.CONCLUSION

Scholarly achievement of understudies of any expert Institution has turned into the serious issue for the administration An early examination of understudies in danger of poor execution enables the Administration to make auspicious move to improve their presentation through additional instructing and advising.This paper focuses on analysis student academic performance by using advantage of data mining techniques model. Predicting student performance in a specific subject based on their performance of test result components during the performance by applying the C4.5 decision tree and linear regression.

IV.FUTURE WORK

Future work include applying data mining techniques on an expanded data set which consider extracurricular activities and other vocational courses completed by students, which we believe may have a significant impact on the overall performance of the students.

REFERENCES

[1] Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in HigherEducation," *IEEE Access,*Volume*:* 4, 2016.

[2] TriptiMishra,Dr. DharminderKumar,Dr. SangeetaGupta,"Mining Students' Data for Performance Prediction," *in fourth International Conference on Advanced Computing & Communication Technologies,*2014.

[3] Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat," Predicting IT Employability Using Data Mining Techniques," *in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC),* 2016.

[4] BipinBihariJayasingh,"A Data Mining Approach to Inquiry Based Inductive Learning Practice in Engineering Education," *in IEEE 6th International Conference on Advanced Computing,* 2016.

[5] S. M. Merchán,"Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic,"*IEEE Latin America Transactions,* vol. 14, no. 6, June 2016.

[6] KonstantinaChrysafiadi and Maria Virvou," Fuzzy Logic for adaptive instruction in an e-learning environment for computer programming," *IEEE Transactions on Fuzzy Systems, Volume:* 23, Issue: 1, Feb. 2015.

[7] M. Mayilvaganan, D. Kalpanadevi," Comparison of Classification Techniques for predicting the performance of Students Academic Environment," *in International Conference on Communication and Network Technologies (ICCNT),* 2014.

[8] Mahesh Gadhavi et al "Student final grade prediction based on linear regression" *Indian Journal of Computer Science and Engineering (IJCSE),*2017