

# ANALYSIS OF STUDENTS DROPOUT FORECASTING USING DATA MINING

Suganya S<sup>1</sup>, Dr.V.Narayani<sup>2</sup>

<sup>1</sup>Research Scholar, Bharathiar University, Coimbatore (India)

<sup>2</sup>Director i/c- MCA, Karpagam College of Engineering, Coimbatore (India)

## ABSTRACT

*Education is the backbone of any country and it is very important to improve the educational strength of the country. There are various methods and challenges on the way, use of technologies like computers, smart rooms, projectors, and eBooks. But these resources are useful only when we know which student needs which type of resource or, in other words, if we are able to predict the results of students, we can improve results and decrease drop out ratio. In our research, we used Data Mining in education to improve the results of the schools. As we know large amount of data is stored in educational database, so in order to get required data & to find the hidden relationship, different data mining techniques are developed & used. There are varieties of popular data mining task within the educational data mining e.g. classification, clustering, outlier detection, association rule, prediction etc. How each of data mining tasks can be applied to education system is explained. To predict the failure of students is a complex task, as it requires large number of the data to be handled. For which the record of students, their each and every activities, academic related information need to maintain. Based on this information, it will be easier to predict the student's failure by applying data mining algorithms on it. The final objective of this paper is to detect the failure of students as early as possible to prevent them from dropping out and improve their academic performance. The outcomes are compared, the models with best results can be shown and the students who are at risk of failure can be provided with the guidance.*

**Keywords:** Data Mining, Educational Data Mining, Decision Trees, Induction Rules Rebalancing Data, Classification Algorithm.

## **I. INTRODUCTION**

Many educational organizations and school administrations today, leave no stone unturned to improve their student's academic performance. In which the marks obtained by the student in the examination decide his/her future. They want to increase the number of student's getting passed in the yearly academics. The reason for this is to develop the best quality of the education process in their institute, to maintain the brand name of the organization and to educate students in a better way. In order to increase the number of students getting passed, the students that may get failed in that particular year in academics need to find firstly. This project basically aims to foretell the student's failure beforehand, so that some measures can be taken to avoid the student's failure in future. To predict the failure of students is a complex task, as it requires large number of the data to be handled. For which the record of students, their each and every activities, academic related information need to maintain. Based on this information, it will be easier to predict the student's failure by applying data mining algorithms on it. The quality is the major key performance factor in higher educational system. The acquirement of quality in higher educational system must be planned, monitored, and controlled in each and every education processes with the main purpose of improving the efficiency of students. The indicator of the quality weakness of the educational system is the large number of students that drop out. By the way, to predict the number of drop out students and factors affecting the drop out situation must use the effective processes.

## **II. RELATED WORK**

Data mining techniques have been successfully used to enhance various aspects of educational quality of higher educational system. Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, and M. Inayat Khan [1] used data mining technique named k-means clustering applied to analyze student's learning behavior that will help the teachers to reduce the drop out ratio to a significant level and improve the performance of students. Sajadin Sembiring et al. [2] studied to apply the kernel method as data mining techniques to analyze the relationships between students' behavioral and their success then they developed the model of student performance predictors which can help to predict the successful student by employing psychometric factors as variables predictors. Xie Wu et al. [3] Used data mining technique with data of undergraduates have to be stored in database or data warehouse with the capacity increasing. The method is carried out by decision tree algorithms.

## **III. DATA MINING IN EDUCATION**

Data mining is a powerful tool for academic intervention. Through data mining, a university could, for example, predict with 85 percent accuracy which students will or will not graduate. The university could

use this information to concentrate academic assistance on those students most at risk. In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: Classification, categorization, estimation, and visualization. Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay." Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates. Education institutions can use classification, for example, for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success. The researchers then used a replacement method that looked at educational outcomes in combination with lengths of study. Defining educational outcomes is easier said than done. Enough time must pass to conclude that a student has reached a certain milestone. Dropping out is also an outcome by itself. Further work was conducted to determine length of study, which required decisions on how to deal with "stopouts," students who left school and later returned.

#### **IV. DROP OUT**

Graduation, especially timely graduation is an increasingly important policy issue[4]. College graduates earn twice as much as high school graduates and six times as much as college dropouts[5]. In addition to the financial rewards, the spouses of college graduates are more educated and their children do better in schools and colleges. Graduation rates are considered as one of the institutional effectiveness[6]. Students drop out due to different reasons; academic trouble, academic preferences, marriage (girls) and their financial position.

- I. Students are unable to get into the major they prefer when they matriculate and therefore they find it difficult to carry on with the course and may leave the institution due to academic trouble.
- II. Students also dropout due to academic preferences. Generally, students choose majors offering the greatest stream of future earnings.
- III. In Indian society, girls are expected to get married at the age of 18 and they may drop out when they are married.
- IV. Financial position of the students plays an important role in drop out percentage.

It is important to understand the determinants of successful and timely degree completion. Most studies of student departure focus on the characteristics of students as determinants of success. The study considers the features such as gender, attendance, previous semester grade, parent education, parent income, scholarship, first child, and part time job.

Parental income: Is an important determinant of the demand for education. Students from higher-income families are less likely to have to drop out to work to finance their education and are most likely to have aspirations that promote persistence. Empirical studies indicate a strong positive correlation between family income and other family background measures on educational attainment enrollment, persistence and graduation[7,8].

Parental education: Plays an important role. Children of college graduates fare well in their exams and are less likely to drop out. A student's previous semester grade and attendance are also included in the study. Grades and attendance may have some tangible value that can be used for future educational and career mobility. Grades may also be considered as an indication of realized academic potential. Financial aid/scholarship plays an important role in higher education by lowering the costs of attendance. The study measures the effect of financial aid/scholarship on student departure. The study also investigates about other information such as whether the student is the 'first child' in the family and he/she is doing part time job to support the family. Both these variables are expected to be positively correlated with graduation.

## V.DROPOUT FORECATSING

### DROPOUT FORECASTING

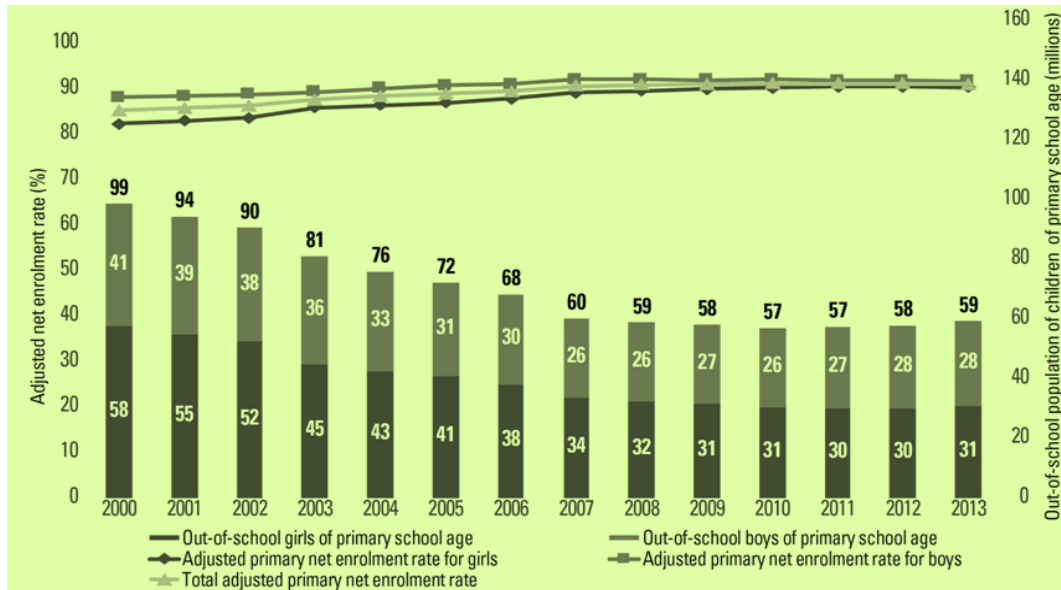
**Drop out** means leaving a school or college for practical reasons, necessities, or disillusionment with the system

Dropout rate forecasting						
year	count (x)	primary (y)	$x - x'$	$y - y'$	$sq(x - x')$	$(x - x')(y - y')$
1990-00	1	40.3	-5.5	9.2575	30.25	-50.91625
2000-01	2	40.7	-4.5	9.6575	20.25	-43.45875
2001-02	3	39	-3.5	7.9575	12.25	-27.85125
2002-03	4	34.89	-2.5	3.8475	6.25	-9.61875
2003-04	5	31.47	-1.5	0.4275	2.25	-0.64125
2004-05	6	29	-0.5	-2.0425	0.25	1.02125
2005-06	7	25.67	0.5	-5.3725	0.25	-2.68625
2006-07	8	25.6	1.5	-5.4425	2.25	-8.16375
2007-08	9	25.09	2.5	-5.9525	6.25	-14.88125
2008-09	10	24.93	3.5	-6.1125	12.25	-21.39975
2009-10	11	28.86	4.5	-2.1825	20.25	-9.82125
2010-11	12	27	5.5	-4.0425	30.25	-22.23375
mean	6.5	31.0425		sum	143	-210.645
Forecasting						
Years	Dropout rate					
2011 - 12	21.46772727					
2012 - 13	19.99468531					
2013 - 14	18.52164336					

## VI.ANALYSIS

The analyzer component incorporates a number of machine learning methods for automatically analyzing the data in the log database. In addition to getting a better insight into the underlying relationship in the data, this also allows for prediction and classification of future sessions. Many machine. Learning methods provide their output in an intelligible, human readable form. For instance, methods for generating decision trees from data such as C4.5 [9], allow for a tree shaped representation of the learning results. The aim of using the decision tree method was to characterize the students motivation in terms of the other attributes that are automatically generated from the log

data, in order to provide an abstracted view on the underlying data as well as to allow for predicting motivational aspects in other students.



### Implementation of Prediction Model

To develop prediction model, many techniques are used such as Neural Networks, Decision Trees, Association Rule Mining, Nearest Neighbor Method, Clustering and Classification. Classification is one of the most commonly used technique for prediction especially when more accuracy needed. Classification prediction based on class label having successful and failure scenarios. In this study, we used the two Decision Tree algorithms, Random Forest and J48graft versus two Bayes classification algorithms Naive Bayes and Bayesian Logistic Regression. Some details of proposed classification algorithms are following

#### 1. Decision Tree Approach

Decision Tree is all about splitting data into segments called branches, shaped as parent child relation. These branches form an upward down Tree that originates with top of the node called root node. Decision Tree also deals with both continuous and categorical data. Decision Tree is influential and wide spread method for prediction. The charm of Decision Trees is because of its comparison with Neural Networks, rules are drawn in Decision Trees that are very easy to understand and interpret. In our research, Random Forest and J48graft algorithms are used to calculate the accuracy of our model.

#### 2. Bayes Approach

Bayes rule is basically core of Bayesian inference method, used to update the probability estimation for a hypothesis as additional [evidence](#). Bayesian updating is dynamic throughout statistics, and in [mathematical statistics](#), its role is very important. Especially to predict a target class, Naive Bayes classification technique is preferred. It works with Bayesian theorem to calculate the probabilities. Bayesian results are often proved more accurate comparatively. Naive Bayes method is good for

a number of causes. Its narrowing is easy, complex iterations are not needed, that shows that it can be applied to bulky data. Its interpretation is very easy, that's why a non-technical users can understand. In our research, Naive Bayes and Bayesian Logistic Regression algorithms were used to calculate the accuracy of our model.

## **VII.CONCLUSION**

This study proposes to foretell the student's academic failure using the algorithms of data mining techniques. The algorithms are applied on huge collection of dataset and the results are obtained, through which the failure can be predicted. This information is more useful for the teachers

and principal of the organization, so that they can make proper arrangements and facilities to increase the capability of students and reduce/prevent the failure of students in academics years. These experiments can show almost expected results in context with economic, educational or sociological characteristics that may be helpful in foretelling low academic performance. This study introduced the data mining approach to modeling drop out feature and some implementation of this approach. The key to gaining a competitive advantage in the educational industry is found in recognizing that student databases, if properly managed, analyzed and exploited, are unique, valuable assets. Data mining uses predictive modeling, database segmentation, market basket analysis and combinations to more quickly answer questions with greater accuracy. New strategies can be developed and implemented enabling the educational institutions to transform a wealth of information into a wealth of predictability, stability and profits.

## **REFERENCES**

- [1]. S. Ayesha, T. Mustafa, A.R. Sattar, and M.I. Khan, Data Mining Model for Higher Education System, European Journal of Scientific Research, Vol.43, No.1, 2010, pp.24-29.
- [2] S. Sembiring, M. Zarlis, D. Hartama, R. S and E. Wani, Prediction of Student Academic Performance By an Application of Data Mining Techniques.Proceedings of International Conference on Management and Artificial Intelligence, 2011, pp.110-114.
- [3] X. Wu, H. Zhang and H. Zhang, Study of Comprehensive Evaluation Method of Undergraduates Based on Data Mining, Proceedings of International Conference on Intelligent Computing and Integrated Systems, pp 541-543.
- [4]. DesJardins, S.L., D.A. Ahlburg and B.P. McCall, 2002. A temporal investigation of factors related to timely degree completion. J. Higher Education, 73:555-581.
- [5]. Murphy, K and F. Welch, 1993. Inequality and relative wages. Ameri. Economic review, 83: 104-109.
- [6]. Murtaugh, P.A., L.D. Burns and J. Schuster, 1999 Predicting the retention of university students. Higher Education, 4: 355-357.
- [7]. Quinlan, J.R., 1983. Induction of Machine learning Machine learning, 1:81-106.
- [8] .Han, J. and M. Kambar, 2003. Data mining: Concepts and techniques. Morgan Kaufmann Publishers, New Delhi.
- [9]. Quinlan, R. (1993). C4.5 : Programs for Machine Larning, Morgan Kaufmann Publishers, San Mateo, CA