



USAGE OF DATAMINING TECHNIQUES FOR PREDICTION OF STUDENTS ACADEMIC PERFORMANCE

¹Elayaraja.N, ²Mythili.K

¹M.Phil Scholar, ²Assistant Professor

Department of Computer Science and Applications, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram

ABSTRACT

Few studies had exposed that data mining techniques can be practical by higher education institutions or universities in determining student failure/success rate so that administration of students' enrollment at the opening of the academic year, support students before they reached possibility of failure, efficient resource exploitation and cost minimization, serving and advising administrative officers to be victorious in administration and decision making. In the most up to date research work, an effort has been completed to charge the applicability of data mining technology to forecast the possibility of student success/failure in the college by using some set of variables/attributes that were measured significant by special literatures. However there are several data mining procedure, the investigator done experimentation by classification only. The amount of the scenarios for experimentation were also very few. Thus the proposed work applies "Cluster Analysis" technique for analyzing the student's success and failure prediction.

I. INTRODUCTION

The amount of data stored in educational databases is increasing, such as large data and to find hidden relationships between variables using different data mining techniques developed and used. The scope of this research is to predict the student result, factors that affect the performance of the students. Two different sources, mark list and questionnaires were used from the higher education college, to predict their mark at the end of the year. Cluster analysis and Naïve Bayes techniques can be used to identify those students with the earlier prediction.

II. RELATED WORK

1. Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)

He have been examining ways to leverage good video games to assess and support important student competencies, especially those that are not optimally measured by traditional assessment formats. The term "stealth assessment" refers to the process of embedding assessments deeply and invisibly into the gaming environment.

2. Dr.Vuda, S. & Capt. Genetu, Y. (2012). Improving Academic Performance of Students of Defence University Based on Data warehousing and Data mining.

The student academic performance in Defence University College is of great concern to the higher technical education managements, where several factors may affect the performance. The student academic performance in engineering during their first year at university is a turning point in their educational path and usually encroaches on their general point average in a decisive manner. The students evaluation factors like class quizzes mid and final exam assignment are studied. It is recommended that all these correlated information should be conveyed to the class teacher before the conduction of final exam. This study will help the teachers to reduce the drop out ratio to a significant



level and improve the performance of students. concepts of data mining like k-Means clustering, Decision tree Techniques, to help in enhancing the quality of the higher technical educational system by evaluating student data to study the main attributes that may affect the performance of student in courses.

3. Vandamme, J.P., Meskens, N., & Superby, J.F. (2007). Predicting Academic Performance by Data Mining Methods.

This paper compares the accuracy of decision tree and Bayesian network algorithms for predicting the academic performance of undergraduate and postgraduate students at two very different academic institutes: Can Tho University (CTU), a large national university in Viet Nam; and the Asian Institute of Technology (AIT), a small international postgraduate institute in Thailand that draws students from 86 different countries. Although the diversity of these two student populations is very different, the data-mining tools were able to achieve similar levels of accuracy for predicting student performance.

III. EXISTING SYSTEM

The existing system classification rule generation process is based on the decision tree and Bayes as a classification technique.

Analysis is done by using WEKA 3.7 application software. The research results offer a helpful and constructive recommendations to the academic planners in universities of learning to enhance their decision making process. This will also aid in the curriculum structure and modification in order to improve students' academic performance. Students able to decide about their field of study before they are enrolled in specific field of study based on the previous experience taken from the research-findings.

The research findings indicated that EHEECE (Ethiopian Higher Education Entrance Certificate Examination) result, Sex, Number of students in a class, number of courses given in a semester, and field of study are the major factors affecting the student performances.

IV. PROPOSED SYSTEM

The proposed system is the most up to date research work; an effort has been completed to charge the applicability of data mining technology to forecast the possibility of student success/failure in the college by using some set of variables/attributes that were measured significant by special literatures. For a number of other variables, in education field; particularly student presentation against healthiness associated problems, financial source troubles, family background, academic schedule and measurement method, skill set of lecturers and much more, is to investigate our work, the effect of those variables to build models with better accuracy and performance than the models built in this research work.

However there are several data mining procedure, the investigator done experimentation by classification only. The amount of the scenarios for experimentation were also very few. I would like to apply "Cluster Analysis" technique for analyzing the student's success and failure prediction.

V. MINING TECHNIQUE

Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. It is about solving problems by analyzing data already present in large quantities of data in order to discover meaningful patterns and rules.

DATA SET

The data collection for the study has been accomplished. These data were obtained from the various colleges, comprising student success and failure prediction. The literature show many risk factors that can be used to predict the students success/ failure in higher education, such as age transfer, stress in life, health related problem, etc.

Column Name	Data Type	Allow Nulls
name	varchar(50)	<input type="checkbox"/>
fathername	varchar(50)	<input type="checkbox"/>
gender	varchar(50)	<input type="checkbox"/>
dob	varchar(50)	<input type="checkbox"/>
marital_status	varchar(150)	<input type="checkbox"/>
department	varchar(50)	<input type="checkbox"/>
classification	varchar(50)	<input type="checkbox"/>
transfer	varchar(50)	<input type="checkbox"/>
grades	varchar(50)	<input type="checkbox"/>
stress	varchar(50)	<input type="checkbox"/>
brother	varchar(50)	<input type="checkbox"/>
sister	varchar(50)	<input type="checkbox"/>
HRP	varchar(150)	<input type="checkbox"/>
FOccupation	varchar(50)	<input type="checkbox"/>
financial_problem	varchar(150)	<input type="checkbox"/>
income	varchar(50)	<input type="checkbox"/>
enthused	varchar(50)	<input type="checkbox"/>
right_clg_university	varchar(50)	<input type="checkbox"/>
handle_clg_costs	varchar(50)	<input type="checkbox"/>
personal_troubles	varchar(50)	<input type="checkbox"/>

CLUSTERING ANALYSIS

Clustering group's similar data together into clusters. Clustering is an unsupervised operation. It is used where we wish to find groupings of similar records in our data without any preconditions as to what similarity may involve. Clustering is used to identify interesting groups in a customer base that may not have been recognized before. For example, it can be used to identify similarities in customers' telephone usage etc...

NAÏVE BAYES

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

VI. METHODOLOGY

The data is collected and preprocessed to reduce the noise data. Then Clustering and Naïve Bayes techniques are applied with the factors considered that affect the students performance. Finally prediction is carried out.

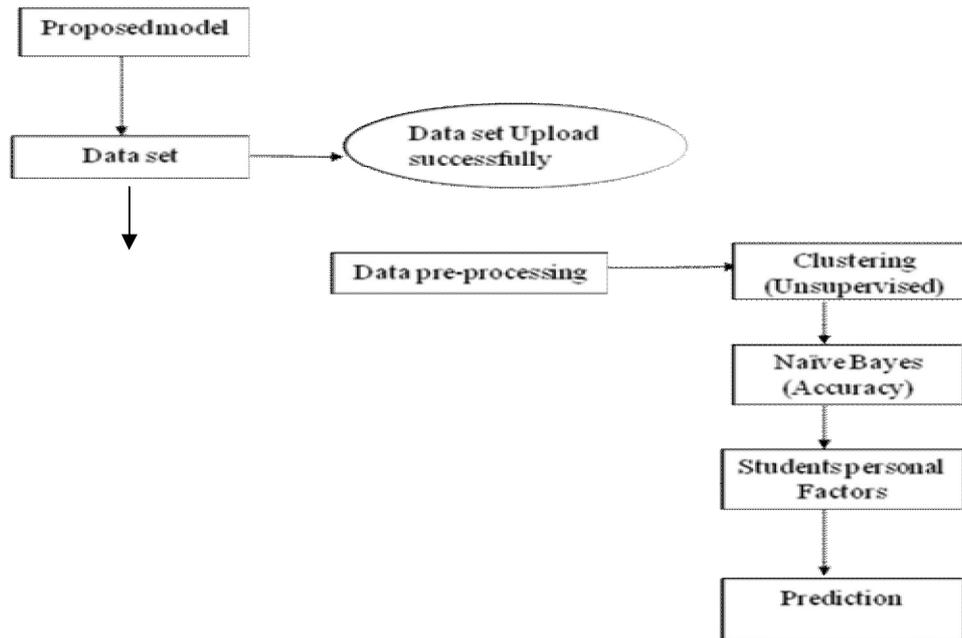


Figure 1 Proposed Architecture

VII. RESULT AND DISCUSSION

Cluster data analysis is basically tracking to findings for specific questions change over time. Once a benchmark is established, to determine whether and numbers shift. Suppose the satisfaction rate for discussion was 50% three years ago, 55% two years ago, 65% last year, and 75% this year.

The experiment has been performed using Cluster analysis and naïve bayes algorithm with several data mining classification techniques SSFPS(Student Success/Failure prediction system) produce more accuracy by considering the factors used for prediction. The resultant is predicted based on the factor which has high probability.

Result based on Cluster analysis

Income	50,000	5
Enthused	Enthusiastic	Enthusi
Right college or university	Confident	Conf
Handle college costs	Difficult	Diffic
Personal troubles	Preoccupied	Preocc

Time taken to build model (full training data) : 0.01 seco

=== Model and evaluation on training set ===

Clustered Instances

0	147 (65%)
1	79 (35%)

Result based on Naive Bayes

```

Correctly Classified Instances      173           86.5 %
Incorrectly Classified Instances    27           13.5 %
Kappa statistic                    0.7401
Mean absolute error                 0.1485
Root mean squared error             0.2644
Relative absolute error             43.0631 %
Root relative squared error         63.8099 %
Total Number of Instances          200
Ignored Class Unknown Instances     26
  
```

=== Detailed Accuracy By Class ===

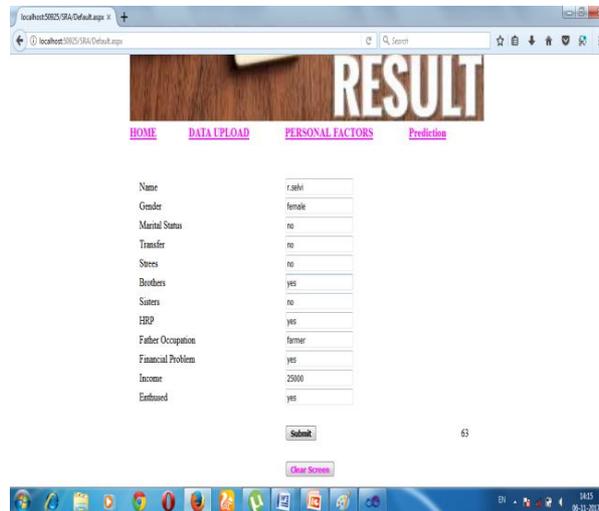
```

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC
          0.847   0.118   0.874     0.847   0.860     0.730
          0.879   0.129   0.870     0.879   0.874     0.750
          1.000   0.010   0.600     1.000   0.750     0.771
Weighted Avg.  0.865   0.122   0.868     0.865   0.866     0.741
  
```

Comparison Of Clustery Analysis And Naïve Bayes

Techniques/Me Assures	Correctly Classified Instances	Mis-Classified Instances	Overall Accuracy
Cluster analysis	147	79	65
Naïve Bayes	173	27	86.5

Prediction



CONCLUSION AND FUTURE WORK

The number of the scenarios for experimentation were also very few. But the other data mining techniques which were not tested by the researcher might reveal important patterns in relation to factors affecting student success/failure. The experiment has been successfully performed with several data mining classification techniques and it is found that the Naive Bayes algorithm gives a better performance over the supplied data set with the accuracy of 86.5%. Thus the proposed method is used for prediction of student success/failure in the higher education by considering the factors which has



high probability. This methodology is focussed on data collection based on questionnaire method from the common students. Hence it is limited to the knowledge of the individual. In future the student success/failure can be predicted in higher education. An experiment can be extended with more distinctive attributes to get more an 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.

REFERENCE

- [1] Dekker, G., Pechenizkiy, M., and Vleeshouwers, J. (2009). Predicting students drop out: A case study. Proceedings of the 2nd International Conference on Educational Data Mining, EDM'09, 41-50.
- [2] Dr.Vuda, S. & Capt. Genetu, Y. (2012). Improving Academic Performance of Students of Defence University Based on Data warehousing and Data mining. August 08/2004, Available at: <http://www.dmu.edu.et>
- [3] Thai, N. (2007). A Comparative Analysis Of Techniques For Predicting Academic Performance. In Proceedings of 37th conf. on ASEE/IEE Frontiers in Education.
- [4] Vandamme, J.P., Meskens, N., & Superby, J.F. (2007). Predicting Academic Performance by Data Mining Methods.
- [5] Carlos Marquez-Vera, Cristobal Romero Morales, and Sebastian Ventura Soto, "Predicting school failure and dropout by using data mining techniques". IEEE Journal of Latin-American Learning Technologies, Vol. 8, No. 1, February 2013.
- [6] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Syst. Appl., vol. 33, no. 1, pp. 135-146, 2007.
- [7] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans.Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 6, pp. 601-618, Nov. 2010.
- [8] Oyelade, O. J , Oladipupo, O. O, Obagbuwa, I. C, " Application of k-Means Clustering algorithm for prediction of students' academic performance," International Journal of Computer Science and Information Security, vol. 7, no. 1, 2010.
- [9] Dr. Vuda Sreenivasarao, Capt. Genetu Yohannes, " Improving Academic Performance of Students of Defence University Based on Data Warehousing and Data Mining," Global Journal of Computer Science and Technology, Vol 12, Issue 2, Version 1.0, January 2012.
- [10] V.P. Bresfelean, "Analysis and Predictions on Students' Behaviour using Decision trees in WEKA Environment", Proceedings of the ITI 2007 29th Int Conf. on Information Technology Interfaces, June 25-28, 2007.