IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ....
Email:editoriijcs@ipasj.org
Volume 5, Issue 8, August 2017
ISSN 2321-5992

Towards Discriminant Function Analysis based Classification

Ajit Kumar Jain[1], Dr. C. K. Jha[2]

[1]Department of Computer Science, Banasthali Vidyapith

[2]Professor, Department of Computer Science, Banasthali Vidyapith

ABSTRACT

The applications of data mining in the field of education are growing fast. This is due to availability of large volume of data and the need of transforming such data into valuable information and knowledge. Educational Data Mining is an area of data mining where lots of researchers are carrying out their work on various issues like enrollment management, performance evaluation, feedback system to support instructors, designing courseware, study of student's learning behavior, and dropout analysis. The proposed research covers the issue of student's dropout in computer science courses of higher education by designing a classification model. This classification model can be used to provide prior information about the dropout so that appropriate steps can be taken for reducing the dropout ratio. Machine learning is one of the disciplines in data mining that plays a key role in predictive data analysis. In this research, we have combined a machine learning method called Classification with the statistical technique known as discriminant function analysis for constructing the classification model. The accuracy of the dropout classifier has been calculated using the measure of classification success. In the last, the results of designed dropout classifier have been compared with the results of Naïve Bayesian classification method. We have used the students' data, collected from computer science courses in a university and SPSS, and WEKA as software tools for the experimental purpose.

Keywords: Educational Data Mining, Discriminant Function Analysis, Training Data, Test Data, Classification, Classifier Accuracy, k-fold cross-validation.

1. INTRODUCTION

Evolutionary development in the area of information technology made the accessibility of data easier for data analysis. A huge amount of data is generated in the areas like finance, retail, medical, and educational organizations. Easier access and the upcoming requirements for transforming such data into valuable information and knowledge emerged in the field of data mining. In simple words, data mining is the process of extracting the useful information from huge amounts of data [10]. In a more comprehensive manner it can be defined as a "process of nontrivial extraction of implicit, previously unknown and potentially useful information (such as knowledge rules, constraints, and regularities) from data in databases [21]". Data mining functionalities such as classification, clustering, frequent pattern mining, association rule mining, and outlier analysis makes it different from the straightforward information retrieval using SQL commands and statistical methods. It is an interdisciplinary field consisting of methods, tools and techniques from other disciplines like information retrieval, database systems, machine learning, statistics, and data visualization.

Data mining in the field of education is a rising area where many researchers are doing lots of work in resolving the challenging issues of education such as enrollment management, prediction of dropouts, and performance prediction by applying data mining tools and methods [31]. Extraction of information and knowledge by applying data mining techniques on educational data leads to the Educational Data Mining, one of the most preferred research areas for the researcher working in the field of computer science. It has been defined by educational data mining community as: "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, and the settings which they learn in [6][22][23]." The major reason behind the significance of Educational Data Mining is that a huge amount of data has been collected by higher educational institutes for years. But this data is never put in the form so that it can be used for the betterment of the students as well as the institutes. To date, educational institutes are data-rich but information poor. This data is growing exponentially every year. The size of data repositories have been exceeded so far that it has become beyond the scope of human's ability to explore this data without the use of powerful data analysis tools. The final situation is that data stored in data repositories has converted into data archives that are almost never visited. By applying the data mining techniques on the unutilized data stored in data archives; patterns, associations, or relationships among all this data can yield crucial information about the educational system, standard of courses, and infrastructural facilities in the institute. This crucial information can be converted into knowledge about historical patterns as well as future trends such as summary information on students' data which may then be analyzed in light of promotional efforts to provide knowledge of students' interest in general and for a specific course. The knowledge may then be used by the Educational Institute in determining the courses or branches success rate, rating of institute based on the infrastructural facilities and educational system, and the reasons concerning with dropouts.

This research aims to develop a classification model for dropouts in computer science courses of higher education institute using a statistical technique named discriminant function analysis. Classification is a data mining function that is used to classify the records into one of the class among set of predefined classes [13]. Classification is an example of supervised learning because the classifier is developed using class labeled training data. The data for the study collected from the computer science students studying in a university, contains information about the student's socio-economic status and the features concerning with the study environment of the university. The results of the study show the main reasons for dropouts and present a classification model based on discriminant analysis that can provide the prior information about the dropouts. The accuracy of the classifier is calculated using measures of classification success and then compared with the classification results of Naïve Bayesian Classification [16].

2. BACKGROUND

2.1 Related Work

Educational data mining has become as a new promising research community because of increasing interest of researchers in the area of data mining and education systems. Romero and Ventura [5] made a survey about the applications of data mining for web-based education systems. The results of the survey shows that data mining techniques such as classification and clustering, outlier analysis, pattern recognition, text mining, statistics and visualization are very effective, efficient and useful for mining web-based education systems. Data Mining is an interdisciplinary area which can also be applied in the field of education making EDM as new emerging research area. The researchers mentioned that a big amount of data is coming from various sources at different granularity levels concerning to education. Lots of different types of issues are there in this field and these issues require different forms of data mining for resolving them [4].

Researchers have proposed many student performance classification models. These models have used commonly known methods for classification such as decision tree, and Bayesian network [17][33]. But the researchers have just focused on prediction analysis without mentioning the methodologies used for feature selection. Feature selection is an important step in the process of classification and the selection of appropriate features for the classification directly affects the accuracy and efficiency of the classifier. In addition to this it will also reduce the complexity of the results. Due to major role in data mining and analysis, it has become a dynamic and productive area of research in the field of pattern recognition, machine learning, and statistics [2][20]. The primary goal of feature selection is to pick a subset of information parameters by removing out features that are unimportant or have no information for prediction. Making the ranking of attributes decides the significance of any individual attribute. Ranking strategies depend on information theory, statistics, or some functions of classifier's outputs [32].

A study was carried out on the students' retention and success rate for the students of mathematics and computer at the Open University College in UK. In this study, most critical factors for pass, fail or dropout were found using logistic regression. These factors were recognized as first assignment marks, the number of math courses that student studied in the past 2 years, the level of the course, the credit points of the course and the student's occupation group[25]. Lin Chang conducted a study on college enrollments. He showed that how extracted patterns may help the administrators to indentify the group of students that may be targeted for admission. This information would likewise benefit recruitment, advertising, communication policies, scholarship payment, program assessment, and numerous regions of institutional approach and practices [12]. In the IEEE international Conference on Information Reuse and Integration, 2006, Aksenova et al. [26] published a paper in which they proposed predictive models based on support vector machines and rule-based prediction. The goal of this study was to predict total enrollment. The sorts of information used during the mining procedure incorporate educational cost and charges, family income, secondary school performance etc. Support vector machines provided the prediction results, which are then utilized by a tool called "Cubist" to produce straightforward rule-based predictive models.

Tissera et al. conducted a research by utilizing the data repository containing the students' performance data collected from the various ICT educational institutes in Sri Lanka. The focus was to find out the association between the undergraduate subjects in the syllabi, by applying the association rule mining. The strength of relationships between two associated subjects was determined using correlation coefficient. The results found provided the numerous bits of information and knowledge for the decision making regarding the quality of educational programmes and design of syllabi for various courses[18]. The academic performance of the students at under graduate and post graduate level, studying at different educational institutes was predicted by Nguyen et al. [19] using the classification methods of decision tree and Bayesian network. The objectives of this study were to provide support for the identification and assistance of failing students as well to determine the scholarship for suitable candidates. They compared the prediction accuracy of both the classification techniques and found that the decision tree based method provides more accurate results than the Bayesian networks.

Web based education systems are exponentially growing due to their independency on location and specific type of equipment requirements. C. Romero et al. conducted a survey on the applications of data mining techniques for the management of online courses by using the Moodle system as a case study [8]. Moodle is an open source online e-learning framework that provides the support to the users to develop their own e-learning courses and projects. The vast amount of data is generated due to its web based nature. The authors in this study have indicated that how this data can be accumulated and analyzed by using the different data mining techniques. The results and useful patterns that are extracted after analysis indicates that how the student's learning and online system can be improved. A research was conducted by S.K. D'mello et al. in the year of 2008 about the learning behavior of the students [29]. They used an intelligent tutoring system named AutoTutor. This system is capable of holding conversation in natural language. The results of the research showed that in discriminating boredom, confusion, flow, frustration, and neutral, machine learning classifiers were reasonably successful. Cesar Vialardi et al. proposed a recommendation system. The principle thought behind their work was the issues raised by college students in selecting the right choice in connection to their academic schedule in light of available data. In this connection, their work proposed the use of recommendation system in view of the data mining techniques that give backing to the students to better pick course to enroll on, having as experience of past students with comparable academic achievements as basis. For this reason, they analyzed real data from the students that had taken the same courses before, utilizing techniques, for example, decision trees to extract patterns, and rules that will be utilized as backing for choice taking in the itinerary of a specific university career [9].

By using two main approaches for data mining i.e. predictive and descriptive, Fadzilah and Abdullah introduced the results of analysis on enrollment data of Sebha University in Libya. As a descriptive task, cluster analysis was used to group the data into clusters. As a predictive analysis, three data mining techniques Neural Network, Decision Tree, and Logistic Regression were used. After evaluating the results obtained from these techniques, it was found that Neural Networks has the highest accuracy in terms of classification [28]. Ramaswami and Bhaskaran presented a study for identifying slow learners and study the significant factors influencing their academic performance. They used the dataset of 772 students from five different schools of Tamilnadu. Using this dataset they developed CHAID based prediction model. By using this model, it was found that attributes like marks obtained, instruction medium, school location, area of living and secondary education type were the significant factors for the student's performance in higher education [15].

One of the most common techniques for representing the frequent patterns in the data set is the association rule mining. Prediction models can also be represented using association rules. There is a common assertion that representations, for example, 'if-then' rules are more justifiable than others. In the wake of an extensive survey in machine learning and data mining literature and additionally observational tests for users, J. Huysmans et al. also inferred that if-then rules are the most common and preferred method for model representation. The understandability of the found models is measured using the number of rules and conditions within the rules [1][11]. Study performed by Natek Srecko and Zwilling Moti explored the success rate of the students in higher educational institutes (HEIs). It has been observed that methods of data mining can be applied effectively on large data sets. But the students' dataset in higher education institutes is limited and small in size comparatively. This study showed that by applying desktop data mining tools like EXCEL and WEKA on the students' dataset for evaluating the students success rate can provide the useful information without going for heavy investment. The results of the study also proved that the small size of students' dataset did not limit the applicability of data mining tools [30]. The objective of the paper written by Cristo´bal Romero et al. was to provide the applications of association rule mining for improving the standard of quizzes and courses. After preprocessing the quiz data, proposed algorithm uses grammar-guided genetic programming for extracting the interesting rules satisfying objective and subjective measures for interestingness threshold. Experiments were carried out on an AI Moodle's course. Based on the results, quizzes and courses were updated [7]. X. Wanli et al. used Genetic Programming, an advanced modeling technique for the design of prediction model for the performance of students participating in "Computer Supported Collaborative Learning" (CSCL) course. The results of the study showed that GP based algorithms are more easily interpretable and optimized as compared to other traditional modeling approaches for prediction [34]. It was shown that there is higher rate of dropout for online course than traditional course by R. S. Baker et al. [24]. In this study, the authors predicted the student success and failure at very early stages. The data source was "Soomo Learning Environment". Logical regression, based prediction model was applied to identify the poor performer students.

Carlos Márquez-Vera et al. proposed a classifier to solve the serious issue of the early dropout in schools. They proposed a Genetic Programming based Interpretable Classification Rule Mining (ICRM) algorithm. The proposed model was capable of predicting student's dropout within the 4-6 weeks of admission in the course and reliable enough to be used in an early warning system (EWS)[3]. Data collected during the admission process of adult learners in a Singapore University from various sources included features like demographics, student profiles, and their academic and profession experiences was used to build the prediction models for predicting their academic performance. This analysis done by S. Chong et al. [27] was used to select the best-fit model and from the evaluation it was found that CHAID decision tree is the best among them in terms of prediction accuracy.

2.2 Tools and Techniques

In the proposed research, we have used discriminant function analysis and Naïve Bayesian classification methods for dropout analysis and SPSS and WEKA as a software tools for supporting these methods, respectively.

Classification

The process of constructing a prediction model for predicting categorical values is called classification. There are two phases in the classification process:

I. **The Learning Phase:** It is also called training phase. In this phase, a classifier is designed using classification algorithm. For the design of classifier a training dataset is used.

II. **The Classification Phase:** It is also called testing phase. In this phase, the accuracy of the designed classifier is tested using the test data. If the accuracy is considerably acceptable, the designed classifier is used for the classification of the real time data.

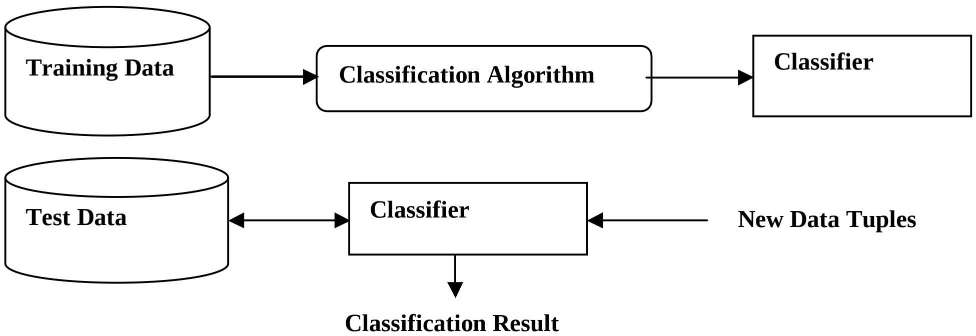

Figure 1 Two Phases of Classification Process

Discriminant Function Analysis (DA)

Discriminant Function Analysis is a statistical technique used for prediction. It is similar to regression analysis which is also used for prediction. However, in case of regression analysis the dependent variables are limited to interval variables to predict numerical values. But in real world scenario, many of the dependent variables are categorical such as whether a customer will 'buy' or 'not buy' the product, whether a student will 'dropout' or 'not' from the course. Here, DA can be used for prediction. In DA, a linear equation is determined for prediction.

$$D = a_1V_1 + a_2V_2 + \dots\dots\dots + b \quad (1)$$

Here, $V_1, V_2,$ .... are the grouping variables and represent the respondents' scores,
D is the discriminant function showing discriminant score for the grouping variable,
$a_1, a_2,$ .... are discriminant coefficients or weight for grouping variables, and b is a constant

The number of discriminant functions in discriminant analysis depends on the number of classes or groups in a dependent variable. For example, a dependent variable having 2 groups will have one discriminant function; a dependent variable having 3 groups will have two discriminant functions and so on.

Naïve Bayesian Classification

Naïve Bayesian Classification is a supervised learning data mining technique. It is a statistical approach using the theory of probability. It uses the Bayes' theorem to predict class membership for a given tuple. There is an assumption behind classification using Naive Bayesian method. The assumption is that the attributes of a tuple are independent of each other. According to Bayes' theorem:

$$P(C_i|T) = \frac{P(T|C_i)P(C_i)}{P(T)} \quad (2)$$

Here, T is a tuple, represented by a vector, $T = (t_1, t_2, ..., t_n)$,
Where, $t_1, t_2, ..., t_n$ are n measurements for a tuple having n attributes.
C_i is a class
$P(C_i | T)$ is the posterior probability of class C_i conditioned on T.
$P(C_i)$ is the prior probability of class C_i.
$P(T | C_i)$ is the posterior probability of T conditioned on C_i.
$P(T)$ is the prior probability of T.

Statistical Package for the Social Sciences (SPSS)

SPSS is a popular and most widely used software package developed by the IBM. It a statistical package for data analysis in the area of social science. As an input, it can take data from various file formats. It analyzes the data using the statistical tools such as descriptive statistics, regression, correlation, classification, and displays the results in the form tabulated reports, charts and plots of dispersions. The application range of SPSS is wide that incorporates higher education, finance, banking, telecommunications, healthcare, insurance, retail, consumer packaged goods, and market research.

Waikato Environment for Knowledge Analysis (WEKA)

WEKA is amongst the most popular and broadly used machine learning and data mining software toolkit [14]. It supports diverse type of data mining algorithms which are written in JAVA. It is open source software developed at "University of Waikato" in New Zealand and freely available under GNU general public license. It provides wide range of tools for data preprocessing, attribute selection, association rule mining, classification and regression analysis, clustering, and data visualization.

3. METHODOLOGY

Data Collection

Data was collected by distributing the questionnaire to the students of computer science belonging to BCA and B.Tech courses in the design of classification model for prediction of dropout. Total number of 310 responses was received from the students. The complete dataset of 310 students was divided in two datasets. First dataset containing 240 tuples was used for learning purpose and second dataset of 70 tuples for testing purpose. The variable 'DROPOUT' in the dataset is the grouping variable having two values 'Yes' and 'No'.

Selection of Best Attributes for Prediction

We have used forward stepwise analysis method for selection of most appropriate predictor variables for developing dropout model. This stepwise selection method has been applied using Wilks' Lambda method with F-value criteria. Wilk's Lambda value ranges from 0 (total discrimination) to 1 (no discrimination). The lower the Wilks' Lambda value, the higher significance is ensured in discrimination of group values.

IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ....
Email:editoriijcs@ipasj.org
Volume 5, Issue 8, August 2017
ISSN 2321-5992

The Significance of Discriminant Function

The significance of the discriminant function (D) using the predictor variables is tested using Eigenvalue, Cannonical Correlation, and Chi-Square tests. Eigenvalue shows the proportion of variance explained by the model. The Eigenvalue must be more than one for a good model. The bigger the Eigenvalue means the more discriminating power of discriminant function. The value of Canonical Correlation parameter reflects the association between grouping variable and the discriminant function. When we take the square of canonical correlation, it tells the percentage of variance explained by the discriminant function in predicting the grouping variable. The chi-square test specifies the measure of importance for the variables in prediction. The smaller the p-value of the chi-square test the stronger the discriminant function separates the groups well.

Determination of Linear Equation for Discriminant Analysis

After selection of most appropriate predictor variables and testing their significance in discrimination of the values of grouping variable, the linear equation for discriminant analysis is derived by putting the values of discriminant coefficients associated with each selected predictor variable and a constant in equation (i). These values have been calculated with the support of SPSS. The values of discriminant coefficients are the weights assigned to predictor variables.

Classification using Linear Equation for Discriminant Analysis

The tuples are classified using newly derived discriminant analysis equation in the following steps:
(1) Compute the discriminant scores (D) for all the records using discriminant analysis linear equation.
(2) Compute the mean of discriminant scores called group centroids for the both groups of DROPOUT i.e. for ‘Yes’ and ‘No’.
(3) Find out the mean centroid by computing the mean of group centroids calculated in step (2). This mean will work as a cut-off value for the classification.
(4) Take the record of student for which the class value of DROPOUT is unknown. Calculate its discriminant score.
(5) Compare discriminant score with the value of mean centroid. If it is less than mean centroid, it means this record belongs to ‘No’ group otherwise it belongs to ‘Yes’ group.

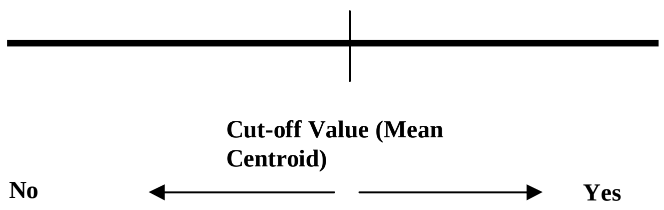

Figure 2 Classification using Discriminant Analysis

4. RESULTS AND DISCUSSIONS

Attribute subset selection from the list of attributes plays a very important role in the classification accuracy as well classifier’s efficiency. After applying the stepwise attribute selection method, the following set of attributes including 9 variables was obtained:

Table 1: Variables Filtered for Prediction Model

S. No.	Name of Variable	Value
1.	Residence Type	Urban(1)/Rural(2)
2.	Type of Family	Nuclear(1)/Joint(2)
3.	Stress	No(1)/Financial(2)/Illness(3)/ Other(4)
4.	Participation in Extra-curriculum Activities	Yes(1)/No(2)
5.	Family Problem	Yes(1)/No(2)
6.	Home Sickness	Yes(1)/No(2)

7.	Adjustment with Campus Environment	Yes(1)/No(2)
8.	Goal Changed	Yes(1)/No(2)
9.	Satisfied with the Selected Course	Highly Satisfied(1)/Satisfied(2)/ Not Satisfied(3)

The significance of the discriminant function (D) using the predictor variables in Table 1 is tested using Eigenvalue, Canonical Correlation, and Chi-Square tests. Table 2 presents the summary of testing parameters. The value of Wilk's Lambda is 0.221 which is less than 1 and nearer to 0. The Eigenvalue value is 3.526 that is greater than 1. The value 0.883 of Canonical Correlation Coefficient suggests that the discriminant function explains 78% (i.e. square of 0.883) of the variation in the grouping variable. In case of Ch-square test, the hypothesis is "The function has no discriminating ability." The value of Chi-Square is 352.543 and the value of p is 0.001 which is less than the chosen significance level 0.05 i.e. $p < 0.05$; it reflects that the hypothesis is rejected. Rejection of hypothesis means that the discriminant function separates the groups well. The values of the test parameters in significance of discriminant function show that the derived discriminant function will be a good discriminator for the discriminating between the values of class variable DROPOUT.

Table 2: Summary of Significant Tests for Discriminant Function

Function	1
Wilks' Lambda	0.221
Eigenvalue	3.526
Canonical Correlation	0.883
Chi-Square	352.543
Degree of Freedom (df)	9
p-value	0.000

The Table 3 exhibits the values of unstandardized discriminant function coefficients. These coefficients are used to construct a discriminant function.

Table 3: Unstandardized Canonical Discriminant Function Coefficients

	Function
	1
Residence Type	-0.720
Type of Family	0.666
Stress	1.068
Participation in Extra-curriculum Activities	1.282
Family Problem	-0.591
Home Sickness	-0.887
Adjustment with Campus Environment	-1.098
Goal Changed	-0.871
Satisfied with the Selected Course	0.555
(Constant)	1.667

Using the discriminant function coefficients in Table 3, the model for dropout in the linear discriminant analysis equation form is derived as:

*D = (-0.720 * Residence Type) + (0.666 * Type of Family) + (1.068 * Stress) + (1.282 * Participation in Extra-curriculum activities) + (-0.591 * Family Problem) + (-0.887 * Home Sickness) + (-1.098 * Adjustment with Campus Environment) + (-0.871 * Goal Changed) + (0.555 * Satisfied with the Selected Course) + 1.667* (3)

The group centroids for the values of grouping variable DROPOUT are calculated by taking the mean of discriminant scores for the value of 'Yes' and 'No'. The Table 4 shows the values of group centroids.

Table 4: Function at Group Centroids

	Function
Will You Dropout or Not?	**1**
Yes	4.120
No	-0.849

The mean of group centroids is 1.635. This value is the cut-off value for classifying the data tuples.

Histograms shown in Figure 3 present the distribution of discriminant scores for each group. The graph reflects that the overlapping between the two histograms is minimal. It implies that the derived discriminant function is good for making the discrimination between the values of grouping variable dropout.

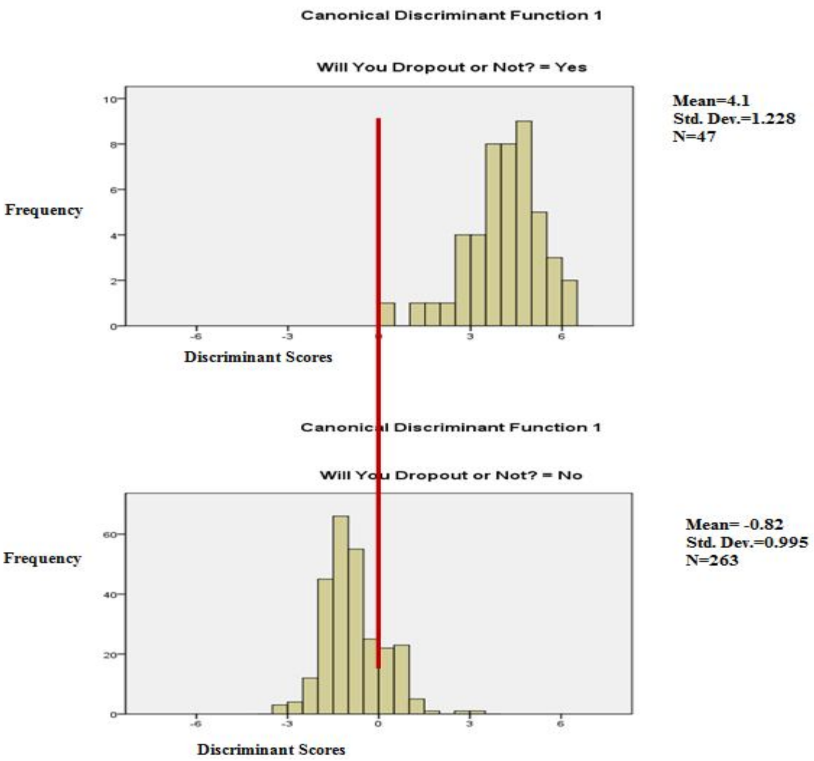

Figure 3 Histograms Showing the Distribution of Discriminant Scores for the Groups ‘Yes’ and ‘No’

The accuracy of the developed discriminant based classifier is compared with the Naïve Bayesian classification method. We have used the WEKA as a software toolkit, for classifying the records using Naïve Bayesian method. Table 5 shows the comparison of correctly classified cases on training data using k-fold cross-validation and test data. The selected value of k is 10.

Table 5: Comparison of Classifiers

Classification Method	Test Data	10-fold cross-validation
Discriminant Analysis	0.97	0.98
Naïve Bayesian	0.96	0.98

The obtained results show that accuracy of the developed discriminant analysis based classifier on test data is 97% which is higher than 95% in case of Naïve Bayesian method. The results indicate that the developed discriminant analysis based classifier for dropout produces high percentage of accuracy.

5. CONCLUSION AND FUTURE SCOPE

The prediction of dropout is very helpful in many contexts. This prior knowledge about student's view regarding the dropout can be used to find out the reasons for dropout as well as by providing the appropriate counseling to the student; dropout ratio can be marginally reduced. This research has proposed a prediction model for dropout using discriminant analysis for the computer science students of higher educational institute. The final outcomes of the research show the reasons for dropout and produce a model for predicting its value.

The proposed research targets the students of the university pursing their study on regular basis for the analysis and designing of the model for the prediction of dropout. In present scenario, web-based online learning is getting more and more popularity among the students. Lots of students are registering in web-based courses. Since vast amount of students' profile data is generated, statistical, data mining, and other knowledge discovery techniques can be applied on this data for making the dropout analysis.

REFERENCES

[1]. A. A. Freitas, D. C. Wieser, and R. Apweiter, "On the Importance of Comprehensible Classification Models for Protein Function Prediction," IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 2010, vol. 7(1), pp. 172–182.
[2]. Alan Miller, "Subset Selection in Regression," CRC Press, 2nd Ed., 2002.
[3]. C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. M. Fardoun, and S. Ventura, "Early Dropout Prediction using Data Mining: A Case Study with High School Students," Expert Systems, 2016, vol. 33(1), pp. 107-124.
[4]. C. Romero and S. Ventura, "Data Mining in Education," Copyright © John Wiley & Sons, Inc, 2013, vol. 3(1), pp. 12-27.
[5]. C. Romero and S. Ventura, "Educational Data Mining: A Survey from 1995 to 2005," Expert Systems with Applications, an International Journal, ELSEVIER, 2007, pp. 135-146.
[6]. C. Romero and S. Ventura., "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, November, 2010, 40(6), pp. 601-618.
[7]. C. Romero, A. Zafra, J. M. Luna, and S. Ventura, "Association Rule Mining using Genetic Programming to Provide Feedback to Instructors from Multiple-Choice Quiz Data," Expert Systems, © Wiley Publishing Ltd, May 2013, vol. 30(2), pp. 162-172.
[8]. C. Romero, S. Ventura, and E. Garcia, "Data Mining in Course Management Systems: MOODLE Case Study and Tutorial", Elsevier Science, Computers & Education, August 2008, 51(1), pp. 368–384.
[9]. C. Vialardi, J. Bravo, L. Shafti, and A. Ortigosa, "Recommendation in Higher Education using Data Mining Techniques," in Proc. 2nd International Conference on Educational Data Mining, Cordoba, Spain, , July 1-3, 2009, pp. 190-199.
[10]. J. Han and M. Kamber, "Data Mining: Concepts and Techniques," 2nd edition, The Morgan Kaufmann Series in Data Management Systems, 2006.
[11]. J. Huysmans, K. Dejaeger, C. Mues, J. Vanthienen, and B. Baesens et al., "An Empirical Evaluation of the Comprehensibility of Decision Table, Tree and Rule Based Predictive Models," Decision Support Systems, ELSEVIER, 2011, vol. 51(1), pp. 141–154.
[12]. Lin Chang, "Applying Data Mining to Predict College Admissions Yield: A Case Study," John Wiley & Sons, Inc., Fall 2006, Issue 131, pp. 53-68.
[13]. M. Cope, H. Baker, R. Fisk, J. Gorby, and R. Foster, "Prediction of Student Performance on the Comprehensive Osteopathic Medical Examination Level Based on Admission Data and Course Performance," Journal of the American Osteopathic Association, 2001, vol. 101(2), pp. 84 – 90.
[14]. M. Hall, E. Frank, G. Holmes, B. Pfahringer , P. Reutemann, Ian H. Witten, "The WEKA Data Mining Software: an Update," ACM SIGKDD Explorations Newsletter, ACM New York, NY, USA, June 2009, vol. 11(1).
[15]. M. Ramaswami and R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining," International Journal of Computer Science, IJCSI, 2010, vol. 7(1), pp. 10-18.
[16]. M. Ramaswami and R. Bhaskaran, "Student Performance Forecasting: A Study of Classification Models," Intelligent Computing Models, Narosa Publishing House, New Delhi, 2009, pp. 38-45.
[17]. M. Ramaswami and R. Bhaskaran, "Student Performance Forecasting: A Study of Classification Models," Intelligent Computting Models, Narosa Publishing House, New Delhi, 2009, pp. 38-45.

[18]. M. Tissera, R. Athauda, H. C. Fernando, “Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining,” IEEE International Conference on Information Automation, 2006, pp. 57-62.
[19]. N. Thai Nghe, P. Haddawy, and P. Janecek , “A Comparative Analysis of Techniques for Predicting Academic Performance,” In Proc. of the 37th ASEE/IEEE Frontiers in Education Conference, 2007, pp. 7-12.
[20]. P. Mitra, C. A. Murthy, and S. K. Pal, “Unsupervised Feature Selection using Feature Similarity,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, vol. 24(3), pp. 301–312.
[21]. Piatetsky-Shapiro and W. J. Frawley, “Knowledge Discovery in Databases,” AAAI-MIT Press, Menlo Park, California, 1992, vol. 13(2), pp. 57-70.
[22]. R. Baker, "Data Mining for Education," International Encyclopedia of Education (3rd edition), Oxford, UK: Elsevier, 2010, vol. 7, pp. 112-118.
[23]. R. Baker. and K. Yacef , “The State of Educational Data Mining in 2009: A Review and Future Visions, Journal of Educational Data Mining,” 2009, vol. 1(1), pp. 3– 17.
[24]. R. S. Baker, D. Lindrum, J. Mary, and D. Perkowski, “Analyzing Early At-Risk Factors in Higher Education e-Learning Courses,” in Proc. of the 8th International Conference on Educational Data Mining, Madrid-Spain 26-29 June, 2015, pp. 150-155.
[25]. R. Woodman, “Investigation of Factors that Influence Student Retention and Success Rate on Open University Courses in the East Anglia Region,” M.Sc. Dissertation, Sheffield Hallam University, UK, 2001.
[26]. S. Aksenova, D. Zhanf, and M. Lu, “Enrollment Prediction Through Data Mining,” in Proc. IEEE International Conf. on Information Reuse and Integration, Hawaii, USA, Sept. 2006, pp. 510-515.
[27]. S. Chong, M. Mak, and W. M. Loh, "Data Mining Applications with the Admission Data of Adult Learners in Higher Education: a Pilot Study ," International Journal of Management in Education, 2016, vol. 10(2), pp 354-359.
[28]. S. Fadzilah and M. Abdoulha, “Uncovering Hidden Information Within University’s Student Enrollment Data Using Data Mining,” In Proc. of the Third Asia International Conference on Modelling & Simulation Conference, IEEE Computer Society, Washington, DC, USA, 2009, pp. 413-418.
[29]. S. K. D'mello, S. D. Craig, A. W. Witherspoon, B. T. MacDaniel, and A. C. Graesser, “Automatic Detection of Learner’s Affect from Conversational Cues,” User Modeling and User- Adapted Interaction, ACM Digital DL, 2008, vol. 18(1), pp. 45-80.
[30]. Srecko Natek and Moti Zwilling, "Student Data Mining Solution–Knowledge Management System Related to Higher Education Institutions," Expert Systems with Applications, © 2014 Elsevier Ltd.2014, vol. 41(14), pp. 6400--6407.
[31]. T. Calders and M. Pechenizkiy, “Introduction to the Special Section on Educational Data Mining," ACM SIGKDD Explorations Newsletter, 2012, vol. 13(2), pp. 3–6.
[32]. W. Duch, T. Winiarski, J. Biesiada, and A. Kachel, “Feature Ranking, Selection and Discretization,” Int. Conf. on Artificial Neural Networks (ICANN) and Int. Conf. on Neural Information Processing (ICONIP), 2003, pp. 251 – 254.
[33]. W. R. Veitch, “Identifying Characteristics of High School Dropouts: Data Mining with a Decision Tree Model,” Paper Presented at Annual Meeting of the American Educational Research Association, San Diego, CA, 2004.
[34]. X. Walni, G. Rui, P. Eva, and G. Sean , "Participation-Based Student Final Performance Prediction Model Through Interpretable Genetic Programming: Integrating Learning Analytics, Educational Data Mining and Theory," Computers in Human Behavior, ©Elsevier Ltd., 2015, vol.47, pp. 168-181.