

Students' Dropout Risk Assessment in Undergraduate Courses of ICT at Residential University – A Case Study

Sweta Rai
M.Tech(CS)
Banasthali University
Jaipur, Rajasthan

Ajit Kumar Jain
Assistant Professor
Banasthali University
Jaipur, Rajasthan

ABSTRACT

The present case study describes the results of the educational data mining aimed at predicting the undergraduate courses of computer science (BCA and B.Tech.) students' instant dropout or after first semester. For this purpose, the preliminary data of 220 students collected randomly in prescheduled format on personal interview to find out dropout rates and dropout reasons. The simple and intuitive classifiers (decision trees) ID3 and J48 were used in this paper. The main reason recorded for dropout of students at this residential university were personal factor viz; illness & homesickness, Educational factors viz; learning problems & difficult courses, change of Institution with present goal and low placement rate and institutional factors such as campus environment, too many rules in hostel life and poor entertainment facilities. The information generated will be useful for better planning and implementation of educational program and infrastructure under measurable condition to increase the enrollment rate of students in ICT courses at this university.

Keywords: KDD, classification, dropout, educational data mining (EDM), decision tree, prediction, ICT (Information & Communication Technology)

1. INTRODUCTION

Education is one of the social factors whereby gender disparity is reflected. The number and proportion of educated females is very low. As the grade level of education increases, the number of female students starts to decline. Consequently, higher education remains the level of learning where females are less represented both as students and staff [1]. The very few women that are fortunate enough to join higher learning institutions can be characterized by lower academic performance and higher forced withdrawal. Female education in India has got momentum after independence. There has been slow development in technical fields, but during the recent past, the continuation of female students in technical as well as in almost every field has envisaged higher status, still due to several factors there has been comparatively lesser percentage of female education as per national statistics conducted by HRD, Govt. of India. Hence, to develop a strong base of interest among female population in the country to prosecute higher education breaking the barrier of gender psychology towards which Banasthali University has played significant role. Since inception in 1935, the university has played major role in educating girl students from primary level to higher studies being undertaken from students coming from different parts of the country. As per need of the present education and employment scenario, the university has restructured its traditional educational system and introduced professional courses in the field of Engineering, Law, Management, Information & Communication Technology, Applied and pure science and Teacher's Education at undergraduate and postgraduate level. However

the reasons for still shortfall in number by female students are a matter of study from various socio-economic considerations which is of great significance being undertaken both from academic and national interest under present study. Since Banasthali University is promoting mainly female education, it is more appropriate to conduct such study at this university where female students from composite strata come for high aspiration.

The present study describes the experiments and the results from a data mining techniques for the students of Information and Communication Technology, Banasthali University, Rajasthan to assist the student dropout program on campus. The development of machine learning algorithms in recent years has enabled a large number of successful data mining projects in various application domains in science, engineering, and business.

The main objective of this paper is to study the student's dropout risk assessment and causes of dropout at undergraduate level of ICT courses using ID3 and J48 decision tree methods. Information like grade in High School, grade in Senior Secondary, student's family income, qualification of parent's etc. were collected from the student's residing in university campus, to predict list of students who need special attention.

2. EDUCATIONAL DATA MINING

Data mining, often known as Knowledge Discovery in Databases (KDD), refers to digging or mining knowledge from huge amount of data. The data mining defined as "the non trivial process of identifying valid, novel, previously unknown, potentially useful information, and ultimately understandable patterns from data in database" [2].

Educational organizations are one of the important parts of our society and playing a vital role for growth and development of any nation. Educational data mining is the application of data mining. The Educational Data Mining community defines EDM as follows: 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 in which they learn.

EDM include analysis (evaluation/ exploration) of educational processes including admission, alumni relations, course selections predicting drop out student, student's success rate, course success rate, performance evaluation of student, learning behavior of students, list of course taken by the student, when the student selected or changed his or her academic major, finding which tasks, courses etc are difficult/easy for which student's, finding elective courses often taken by student's.

2.1. Classification

Classification is called supervised learning. Data classification is a two step process. In first step, a model is built and in second step the model is used for classification. There are many classification methods such as Naive Bayes and Bayesian Belief Networks, Decision Tree, Neural Network etc. But Decision Tree is the basic and popular technique for classification. In educational data mining, one may predict his/her final grade, predict whether student will graduate or not. The decision tree is used to represent logical rules.

2.2. Decision Tree

A decision tree is a flow –chart like tree structure where each node denotes a test on an attribute value and each branch represents result of the test. The tree leaves represents classes. Decision tree is a model that is both predictive and descriptive. Decision tree generation consist of two phases: Tree Construction & Tree Pruning. In this case study, ID3 and J48 decision tree algorithm are used.

2.2.1 ID3 (Iterative Dichotomise 3): This algorithm was invented by Ross Quinlan used to generate a decision tree from a dataset by employing a top- down, greedy search. First calculate the information gain for every attribute. The attribute with highest information gain or greatest entropy reduction is chosen as the test attribute.

2.2.2 J48: This algorithm is based upon Ross Quinlan C4.5 algorithm. It splits the attribute value into two partitions based on the selected threshold as one child and the remaining as another child. J48 uses a gain ratio as an attribute selection measure. This algorithm can handle both discrete and continuous attribute. It also handles missing attribute values. At first, calculate the gain ratio of each attribute to maximize the root node of tree. J48 uses pessimistic pruning to remove unnecessary branches in the decision tree to improve the accuracy of classification.

3. REVIEW OF LITERATURE

In recent past, various studies have been carried out in connection to study the students' academic achievement, dropout and organizational performance using application of data mining methods by number of authors in the area of traditional and professional education particularly in information science, management and social Sciences.

In this regard, Thai Nghe et.al [13] has compared Decision Tree and Bayesian Network to predict students GPA for undergraduate and postgraduate from two different institutes. The results indicate that the Decision Tree outperformed Bayesian Network in all classes.

Ian et.al [3] conducted a case study using EDM to warn the students at risk before their final exam using behavior of failure students. Similar type of study was conducted by Romero et.al [4] in Moodle course management system. Polpinij [17] predicted final grade of students using data acquired from Web based system. Beikzadeh and Delavari [7] give knowledge to use data mining method in higher education. On the other hand, Waiyamai [6] developed new curricula to select an appropriate major for engineering student using data mining methods.

Kotsiantis et.al [8] conducted a comparative study of different classifiers to predict students drop-out in the middle of a course using different parameter of 350 students. It was found that the Naive Bayes and Neural Network were the best classifiers to predict about 80% of drop-outs. Based on the

findings it was observed that Naïve Baves model is best suited for small data set in comparison to other methods.

The data mining classification techniques was applied by Al-Radaideh et al. [9] study the factors that affect the student performance in course to improve the quality of higher education system.

Hijazi and Naqvi [10], studied the student performance on 300 students using educational data mining. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students family income, students mother's age and mother's education are significantly related with student performance" was framed. Outcome indicate that Mother's education and family income highly correlated with the academic performance.

A case study on educational data mining was conducted by Kovacic [11] predict student's success using enrollment data. The CHAID and CART algorithms were applied on dataset to classifying the students (successful and unsuccessful). The accuracy CART was greater than CHAID.

Shannaq et al. [12], applied the classification technique to predict the numbers of enrolled students by evaluating academic data from enrolled students to study the main attributes that may affect the students' loyalty (number of enrolled students). The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated using different evaluation methods. It allows the University management to prepare necessary resources for the new enrolled students and indicates at an early stage which type of students will potentially be enrolled and what areas to concentrate upon in higher education systems for support.

A study on factor analysis was conducted by Walters and Soyibo [14], conducted a study of 305 students of Jamaican high school to study the level of performance on five integrated science process skills with performance linked to gender, grade level, school location, school type, student type, and socio-economic background (SEB). The attribute student's academic performance and the nature of the school have positive significant relationship.

Yadav et.al [15], conducted a study on university students to predict the performance at the end of the semester using ID3, C4.5 and CART and shows that CART is the best algorithm for classification of data.

The student' performance was evaluated using decision tree method for classification by Bharadwaj and Pal [16]. The objective of this study is to extract knowledge that describes students' performance in end semester examination. They used students' data from the students' previous database including Attendance, Class test, Seminar and Assignment marks. This study helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising. Further, they conducted the study on student performance based on selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayesian classification method on 17 attributes, it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mothers' qualification, students other habit, family annual income and students' family status were highly correlated with the student academic performance. A similar type of work was conducted by Pandey and Pal [5], to study the performance of students.

Using association rule they find the interestingness of student in opting class teaching language.

Shaeela Ayesha et.al [18] used k-means clustering data mining techniques to analyze student's learning activities. The information generated is helpful for instructor as well as for students to improve performance.

4. METHODOLOGY

Success percentage rate of any institute can be improved by knowing the reasons for dropout student. In present study, information on various parameters was collected through a structured questionnaire on personal interview basis from a composite sample of 220 students of undergraduate courses (BCA and B.Tech) in computer science / information technology of University. Predicting the students dropout status whether they continue to their study or not, needs lots of parameters such as personal, academic record, social, environmental, etc. variables are necessitated for the effective prediction.

Since the present study is in relation to classify the various quantitative and qualitative factors to study the causes of dropout which belongs to the process of knowledge discovery and data mining. This information will be helpful for the management to reduce the dropout rate in campus. In order to achieve the above mentioned objectives the following steps were followed:

4.1 Data Preparation: The data used in this study was prepared from Banasthali University through structured questionnaire on personal interview. The questionnaire has been constructed based on theoretical and empirical grounds about factor affecting student's performance and causes of dropout. The questionnaire included socio-demographic indicators (Age, Date of birth, Geographical location, Marital status, Parents education, Parents occupation and Annual income), Educational factors (Performance in High school, Senior Secondary School , Location of Schooling, Type of Examination Board, Medium of Study etc.), Parental Attitudes, Causes of dropout, and Institutional factors, etc. Data was collected from 1st year undergraduate (BCA and B.Tech) students. The data format is presented in Table 1.

Table 1: Variables related to student

Variables	Description of variables	Possible values
AGE	Age	{<18, 18-20, >20}
RES	Residence	{Rural, Urban}
N_STATE	Native state	Categorical
FTYPE	Family type	{Nuclear, Joint}
ANN	Annual Income	{Low, Medium, High, VHigh}
FEDU	Father's education	{Illiterate, Sec, HSec, UG, PG}
MEDU	Mother's education	{Illiterate, Sec, HSec, UG, PG}
FOCC	Father's occupation	{Govt. service, Pvt. Service, Business, Agriculture, Retried, NA}
MOCC	Mother's occupation	{Govt. service, Pvt. Service, Business, HWife, NA}
SSCG	10 th Grade	{A=90-100%,

		B=80-89%, C=70-79%, D=60-69%, E=less than 60% }
HSCG	12 th Grade	{A=90-100%, B=80-89%, C=70-79%, D=60-69%, E=less than 60% }
S_LOC	location of school	{village, town city}
MED	Medium of schooling	{Hindi, English}
HSC_STREAM	Stream in higher secondary	{Math, Bio, Com, Arts, Arts(Math)}
C_ADMITTED	Enrolled in course	{B.Tech, BCA}
A_TYPE	Admission type	{Entrance, Merit, Management}
SAT_LEVEL	Student's satisfaction with course	{V.Satisfied, Satisfied, Not V.Satisfied, Not Satisfied}
C_SYLL	Syllabus of course	{V.satisfactory, satisfactory, Balanced, difficult, V.difficult, Lengthy}
Uni_EXPENSES	Expenses in university	{own_income, Loan, Both}
STRESS	Any type of stress in family	{No, Financial, illness, Other}
ULIK	Like University	{Yes, No}
UES	University Education sys.	{Excellent, V.Good, Good, Poor, V.Poor}
UINF	University infrastructure	{Excellent, V.Good, Good, Poor, V.Poor}
CURR	Extracurricular in university	{Excellent, V.Good, Good, Poor, V.Poor}
ENTER	Availability of entertainment in campus	{Excellent, V.Good, Good, Poor, V.Poor}
Self study	Time spare for study	{<1 hr, 2-3 hrs, 4-5 hrs, >6 hrs}
PAR_CURR	Participation in extracurricular activity	{Yes, No}
PLAC	Placement status	{below avg, avg, good, V.good, Excellent}
DROP	Withdraw from course	{Yes, Not Sure, No}

4.2 Data Selection and Transformation

After collection of data, the dataset was prepared to apply the data mining techniques. Before application of prescribed model, data preprocessing was applied to measure the quality and suitability of data. In this step only those attributes were selected which were needed for data mining. For this, remove missing values; smoothing noisy data, selection of relevant attribute from database or removing irrelevant attributes, identifying or remove outlier values from data set, and resolving inconsistencies of data. Finally, the pre-processed

data were transformed into a suitable format to apply data mining techniques.

4.3 Data Analysis Techniques: In this study, one quantitative and the other qualitative data analysis techniques have to be employed using statistical methods and data mining methods.

4.3.1 Statistical methods: The data collected were analyzed using SPSS statistical software to measure the quality of data based on mean, standard deviation, t-test, and linear regression analysis.

4.3.2 Data mining methods: After proper collection, scrutiny and transformation of data using appropriate measures, Data mining classification and decision tree approach were applied to predict student dropout rates and dropout causes in early stage of their study either before or after completion of their first year of their study program.

4.4 Implementation of Classification Model

In this study, the data mining software WEKA tool was used for analysis of data in database. WEKA is open source software that implements a large collection of machine learning algorithms and is widely used in data mining application. It is created by researchers at university of Waikato in New Zealand. It is platform independent and portable because it is implemented in java programming language. From the above data, student_drop.arff file is created. This file was loaded into WEKA Explorer to generate a classification model, using decision tree method. There are several techniques in WEKA tool such as ID3, C4.5, CART and J48 etc to generate a decision tree. 10-fold cross validation was selected as our approach.

5. RESULTS AND DISCUSSION

The dataset collected from 220 students was analyzed using frequency distribution and it was found that dropout rate is less than 20% which is mainly due to personal and institutional reasons. Particularly, the first semester students of B.Tech in Computer Science / Information Technology were interested to change the programme and Institution in better prospective. Further, Data set was classified using ID3 and J48 decision tree algorithms. A confusion matrix was constructed for ID3 and J48 to study the actual and predicted class. The dropout data set was classified in three groups such as Yes (students who have completely decided to dropout), Not Sure (students who likely to be dropout) and No (students not interested to dropout) and based on these three groups, 3x3 confusion matrix for J48 and ID3 decision tree algorithm was constructed (Table 2). The correctly classified instance is the sum of the diagonal in matrix whereas all others figures are incorrectly classified. On the other hand the accuracy percentages for dropout for ID3 and J48 decision tree algorithms using 10-fold cross validation presented in Table 3 shows the highest accuracy percent 79.55 for J48 followed by ID3 (67.73). It indicates that J48 decision tree algorithm is the best classifier as compare to ID3 decision tree algorithm for present dataset to predict the student's dropout status.

Table 2: Confusion matrix of ID3

Actual class	Predicted class			
		A	B	C
	A (No)	131	18	9
	B (Yes)	21	10	5
	C (Not sure)	8	1	8
Total				220

Confusion matrix of J48

Actual class	Predicted class			
		A	B	C
	A (No)	154	7	5
	B (Yes)	22	14	1
	C (Not sure)	10	0	7
Total				220

Table 3: Accuracy percentage for dropout

Algorithm	Correctly classified instances	Incorrectly classified instances
ID3	67.7273%	28.1818%
J48	79.5455%	20.4545%

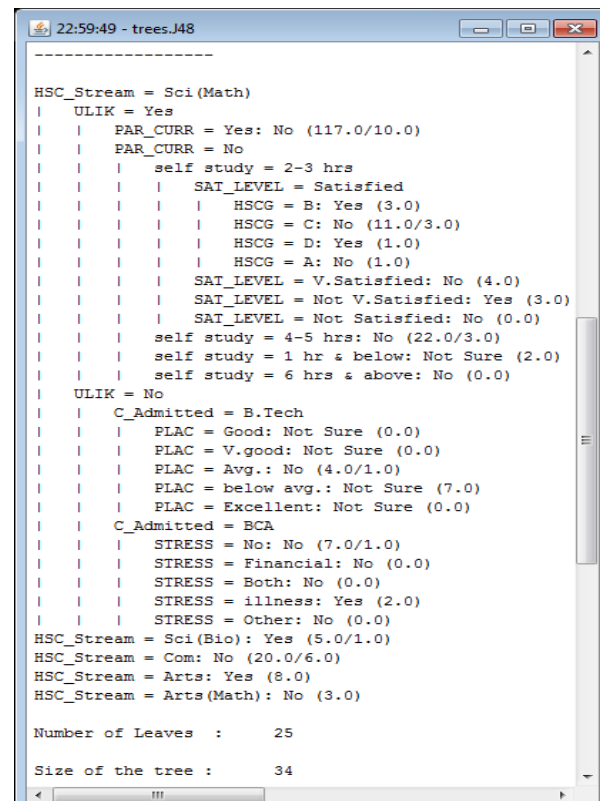


Figure 1: A decision tree generated by J48 algorithm

Classifier rules

IF HSC_Stream=Sci(Bio) AND C_Admitted= BCA THEN Dropout=Yes (5.0/1.0)
IF HSC_Stream=Arts AND C_Admitted= BCA THEN Dropout=Yes (8.0)
IF HSC_Stream=Arts(Math) AND C_Admitted= BCA THEN Dropout=No (3.0)
IF HSC_Stream=Commerce AND C_Admitted= BCA THEN Dropout=No (20.0/6.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=Yes THEN Dropout=No (117.0/10.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= Satisfied AND HSCG=A THEN Dropout=No (1.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= Satisfied AND HSCG=B THEN Dropout=Yes (3.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= Satisfied AND HSCG=C THEN Dropout=No (11.0/3.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= Satisfied AND HSCG=D THEN Dropout=Yes (1.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= V.Satisfied THEN Dropout=No (4.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= Not V.Satisfied THEN Dropout=Yes (3.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=2-3hrs. AND SAT_LEVEL= Not Satisfied THEN Dropout=No (0.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=4-5hrs. THEN Dropout=No (22.0/3.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=1hr. & below THEN Dropout=Not sure (2.0)
IF HSC_Stream=Sci(Math) AND ULIK= Yes AND PAR_CURR=No AND Self Study=6 hrs & above THEN Dropout=No (0.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= BCA AND STRESS=No THEN Dropout=No (7.0/1.0)

IF HSC_Stream=Sci(Math) AND C_Admitted= BCA AND STRESS=Financial THEN Dropout=No (0.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= BCA AND STRESS=illness THEN Dropout=Yes (2.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= BCA AND STRESS=other THEN Dropout=No (0.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= B.Tech AND PLAC=Good THEN Dropout=Not sure (0.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= B.Tech AND PLAC=V.Good THEN Dropout=Not sure (0.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= B.Tech AND PLAC=Avg THEN Dropout=No (4.0/1.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= B.Tech AND PLAC=below avg. THEN Dropout=Not sure (7.0)
IF HSC_Stream=Sci(Math) AND C_Admitted= B.Tech AND PLAC=Excellent THEN Dropout=Not sure (0.0)

6. CAUSES OF DROPOUT

The data collected from 220 students was analyzed to study the frequency distribution against each factor of those students who have completely decided to drop out during the course of study programme. The result of each causing factor was listed in Table 4. The highest dropout reasons were personal factor (28%) and Institutional factors ranged between 10 to 17 %. Whereas few students likely to dropout due to home sickness and adjustment problems etc.

Table 4: Cause of dropout

Reasons	Yes (Students who have completely decided to dropout)	
	Number	percentage
Illness	14	6.36
Family Problem	60	27.27
Home sickness	11	5.00
Marriage	0	0.00
Campus environment	22	10.00
Too many Rules	38	17.27
Low Placement rate	26	11.82
Course were difficult	23	10.45
Course fee is too high	23	10.45
Learning Problem	23	10.45
Change of personal goal	9	4.09
Employment	1	0.45
Peer/friend Problems	2	0.91
Hectic Schedule	4	1.82
Adjustment problem	7	3.18
enrolled for other institute	11	5.00

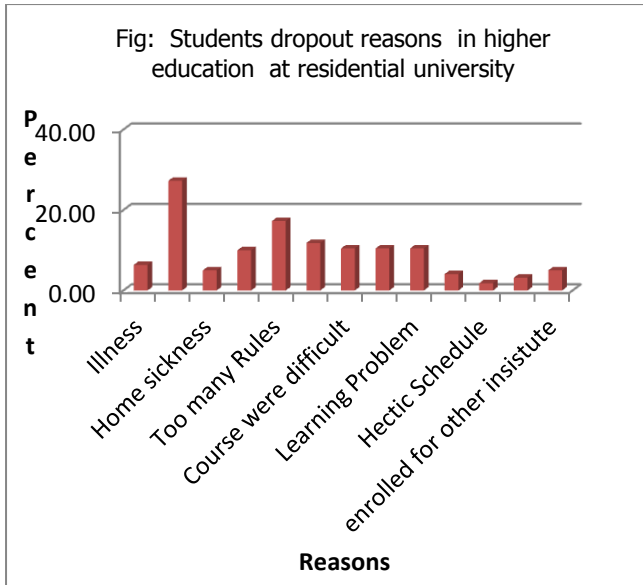


Figure 2: Student's dropout reasons at residential university

The graphical presentation (Figure 2) also shows trend against each causing factors. The highest contribution was recorded by family problem followed by too many rules imposed by the university and low placement rate etc.

7. CONCLUSION

This paper demonstrated the integration of statistics and data mining techniques for the purpose of educational data mining. During analysis of dropout students, many attributes have been examined, and some of them are found effective on the prediction. Attribute HSC_Stream (stream in higher secondary) was the strongest attribute, ULIK (Like university), with slight effect of PAR_CURR (participation in extracurricular activity), Self Study (time spare for study), C_Admitted (enrolled in course B.Tech or BCA) and Placement status etc.

Result indicates that J48 decision tree algorithm is best classifier compare to ID3. This study will also work to identify those students which needed special attention to reduce drop-out rate. The information generated will be useful for better planning and implementation of educational program and infrastructure under measurable condition to increase the enrollment rate of students in ICT courses at this university.

8. REFERENCE

- [1] Yeshimebrat Merasha, Alemayehu Bishaw and Firew Tegegne (2013), Factors Affecting Female Students' Academic Achievement at Bahir Dar University, *Journal of International Cooperation in Education*, Vol.15 No.3 pp.135 – 148
- [2] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, R (1996). "The KDD Process for Extracting Useful Knowledge from Volumes of Data," *Communications of the ACM*, (39:11), pp.27-34.
- [3] Ian, H.W., and Eibe, F. (2005), "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations," California: Morgan Kaufmann,
- [4] Romero, C., Ventura, S. and Garcia, E. (2008), "Data Mining in Course Management Systems: Moodle Case Study and Tutorial," *Computers & Education*, vol. 51, no. 1. pp. 368-384.
- [5] Pandey U.K., and Pal S. (2011), "A Data mining view on class room teaching language", (IJCSI) *International Journal of Computer Science Issue*, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814.
- [6] Waiyamai, K. (2003), "Improving Quality of Graduate Students by Data Mining," Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Thailand.
- [7] Beikzadeh, M. R. and Delavari, N. (2005), "A New Analysis Model for Data Mining Processes in Higher Educational Systems," On the proceedings of the 6th Information Technology Based Higher Education and Training 7-9 July
- [8] Kotsiantis S. Pierrakeas C. and Pintelas P. (2004), "Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques", *Applied Artificial Intelligence*, Vol. 18, No. 5, pp. 411-426.
- [9] Al-Radaideh, Q. A., Al-Shawakfa, E.M. and Al-Najjar, M.I. (2006) Mining Student Using Decision Trees, *The International Arab Conference on Information Technology*.
- [10] Hijazi S.T. and Naqvi R.S.M.M (2006), "Factors Affecting Student's Performance: A Case of Private Colleges", *Bangladesh e-Journal of Sociology*, Vol. 3, No. 1.
- [11] Kovacic Z.J. (2010), "Early prediction of student success: Mining student enrollment data", *Proceedings of Informing Science & IT Education Conference*.
- [12] Shannaq, B. , Rafael, Y. and Alexandro, V. (2010) 'Student Relationship in Higher Education Using Data Mining Techniques', *Global Journal of Computer Science and Technology*, vol. 10, no. 11, pp. 54-59.
- [13] Thai Nghe, N., Janecek, P. and Haddawy, P., (2007) "A Comparative Analysis of Techniques for Predicting Academic Performance," *ASEE/IEEE Frontiers in Education Conference*,
- [14] Walters Y.B., and Soyibo K. (2001) "An Analysis of High School Students' Performance on Five Integrated Science Process Skills", *Research in Science & Technical Education*, Vol. 19, No. 2, pp.133 – 145.
- [15] Yadav S.K., Bharadwaj B.K. and Pal S. (2013), "Data Mining Applications: A comparative study for Predicting Student's Performance", *International Journal of Innovative Technology and Creative Engineering (IJITCE)*, Vol. 1, No. 12, pp. 13-19.
- [16] Bharadwaj B.K. and Pal S. (2011), "Data Mining: A prediction for performance improvement using classification", *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 9, No. 4, pp. 136-140.
- [17] Polpinij, J. (2002), "The Probabilistic Models Approach for Analysis the Factors Affecting of Car Insurance Risk," M.S. thesis, Department of Computer Science, Kasetsart University, Thailand.
- [18] Shaeela Ayesha and Tasleem Mustafa. (2010). Data Mining Model for Higher Education System. *European Journal of Scientific Research [Online]*. 43(1). pp.24-29.