iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4463

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Prediction Model for Student Dropout Analysis using Data Mining Algorithms

R. Preethi[1], S. Priyadharshini[2], E. Balraj[3]
[1, 2]UG Student, [3]Assistant Professor, Information Technology, M. Kumarasamy College of Engineering, Karur

Abstract: Online instruction oftentimes manages the high rate of dropout amid a course. There is a colossal measure of chronicled data put away in on the web. Consequently, it winds up important to extend effective technique to predicate the understudies in danger of dropping out. In this paper, we manage the issue by proposing prescient models to give instructive framework the obligation to recognize understudies whom are in as far as possible. Subsequently, this examination work recommend a model which can naturally perceive whether the understudy will proceed with their investigation or drop their examination utilizing arrangement procedure dependent on choice tree. The dropout chance components are recognized by utilizing the information mining properties, for example, visit designs, connections, comparability measures, affiliations rule mining,

I. INTRODUCTION

There are many research thinks about important to e-learning field that exhibits the diverse method for applying AI procedures for different instructive purposes. The principle goal of these examinations are anticipating the rate of dropouts or in danger understudies in remote courses by inspecting the log information gathered from different Learning Management Systems.(LMS).

The examination led by Kotsiantis is one of the underlying investigations which researched use of AI strategies in separation learning for dropout expectation. In this examination, time-invariant and time-changing information were incorporated and absolutely six AI systems was utilized, which are Decision Trees, Neural Networks, Naïve Bayes calculation, Instance-Based Learning Algorithms, Logistic Regression and Support Vector Machines. This investigation was made out of two test stages, preparing and testing. Amid these stages, number of properties was expanded well ordered. For instance, while as it were statistic information was incorporated into the initial step, information from the main up close and personal gathering was included the subsequent stage. Six calculations were tried for each these ensuing advances and afterward they were looked at. The vital finish of this investigation is that Naïve Bayes calculation is effective in the expectation of dropouts; it predicts with 83% precision.

II. LITERATURE REVIEW

Bharadwaj and Pal [1], proposed the novel methodology utilizing the choice tree strategy for order to assess the understudies execution. This contextual investigation is to decide the information that portrays understudies' execution in end semester examination. This investigation was very helpful to recognize the dropout's understudy in prior stage and understudies who need extraordinary consideration and enable the guide to take prior regard for the understudies.

El-Halees [2], anticipated a straightforward contextual investigation that usea instructive information mining to break down conduct of understudies learning. The target of his investigation is to indicate how helpful information mining can be utilized in advanced education to improve understudy's execution. They connected systems of information mining to reveal pertinent data from extensive database, for example, affiliation tenets and order rules utilizing choice tree, grouping and exception examination.

Fadzilah and Abdullah [3], They connected information mining methods to enlistment information. Illustrative Analysis and prescient Analysis approaches were utilized. To gather the information into groups dependent on their likenesses, bunch investigation is utilized. For prescient examination, Neural Network, Logistic relapse, and the Decision Tree have been utilized. In the wake of assessing these procedures, Neural Networks classifier was found to give the most elevated outcomes in term of order precision.

Ramasubramanian et.al.[4], he proposed foresee parts of advanced education understudies. In this paper they break down that one of the greatest difficulties that advanced education faces today is anticipating the conduct of understudies. Organizations might want to know, something about the exhibitions of the understudies amass astute. He proposed an issue to examine the exhibitions of the understudies when the substantial information base of Students data framework (SIS) is given. For the most part understudies' issues will be characterized into various examples dependent on the dimension of understudies like ordinary, normal and underneath normal. In this paper we endeavor to examine SIS database utilizing harsh set hypothesis to anticipate the fate of understudies.

©IJRASET: All Rights are Reserved

III. TECHNIQUE

Achievement rate of any instructive foundation can be dissected by knowing the purposes behind dropout understudy. In this examination, understudy data on different parameters was gathered through Machine Learning archive by Predicting the understudies dropout status whether they intrigued to proceed with their investigation or not, needs heaps of parameters, for example, individual subtleties, scholarly subtleties , family foundation, social, natural, and so on factors are fundamental data for the successful forecast of properties . Since the present investigation is in connection to arrange and relapsing the different quantitative and subjective components to recognize the reasons for dropout in the point of view of information revelation and information mining. To achieve the above goals the accompanying advances were pursued (Fig.1)

A. Preparation of Information

The dataset utilized for this investigation was set up from the UCI Machine Learning Repository. The information traits incorporate understudy grades, statistic, social and school related highlights) and it was gathered by utilizing school reports and surveys. Two datasets are given with respect to the execution in two particular subjects: Mathematics and Portuguese language. In the Two datasets were displayed under paired/five-level grouping and relapse assignments.

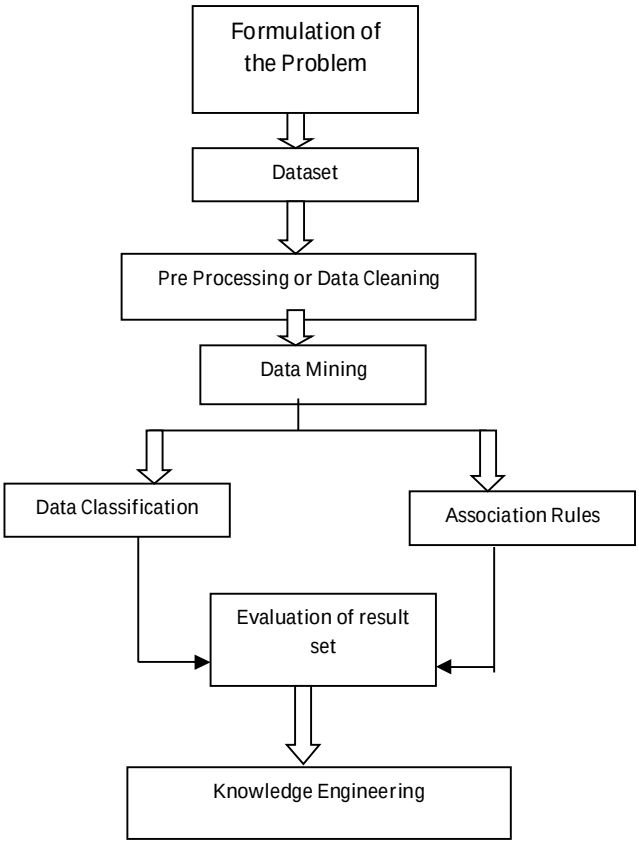


Figure 1 - Work Methodology

After accumulation of information, the dataset was set up to apply the information mining systems. Prior to utilization of endorsed demonstrate, information preprocessing was connected to quantify the quality and reasonableness of information. For this, evacuate missing qualities; smoothing uproarious information, choice of applicable characteristic from database or expelling insignificant properties, recognizing or expel anomaly esteems from informational collection, and settling irregularities of information. A portion of the insignificant parameters was expelled from database, for example, ID, age, date of birth, class, conjugal status, condition of residence, first language, religion, the sexual orientation field containing just a single esteem female, since college concerns just female understudies, the conjugal status field containing one esteem unmarried, and so on. An evaluation scale is utilized for assessment of understudy execution at school. "Evaluation An" understudies are viewed as the individuals who have a rate more noteworthy

©IJRASET: All Rights are Reserved

Table 1 – Variables in dataset

VARIABLES	DESCRIPTION	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}

Than 8.5, "Evaluation B"- in the range somewhere in the range of 7.5 and 8.5, "Evaluation C"- in the range somewhere in the range of 6.5 and 7.5, and "Grade D" in the range beneath 6.5. A four dimension scale is utilized in the family yearly pay. "VHigh" yearly pay are viewed as the individuals who have salary more prominent than 6 lakhs, "High" yearly pay extend between 4 lakhs and 6 lakhs, "Medium" yearly pay run between 2 lakhs and 4 lakhs and "Low" yearly pay in the range beneath 2 lakhs. A straight out target variable "Dropout status" is built dependent on the perspective on respondents; it has two conceivable qualities "Yes" (understudies who are totally chosen to pull back from their course) and "No" (understudies who are need to proceed with their investigation).

The last dataset utilized for the investigation contains 220 occurrences (183 in the "No" class and 37 are in "Yes" classification) each portrayed with 34 qualities (1 yield and 33 input factors), ostensible and numeric. The examination is restricted to the understudy information for undergrad. At long last, the pre-prepared information were changed into a reasonable configuration to apply information mining procedures.

©IJRASET: All Rights are Reserved

B. Data Analysis Techniques

After appropriate accumulation, information mining grouping calculations and choice tree approach were utilized to foresee understudy dropout rates and reasons for dropout in starting phase of their examination either previously or after fruition of their first year of their investigation program. Characterization model will be executed by utilizing R instrument. Increasingly number of classifiers accessible in R however ID3 was utilized to actualize this contextual analysis. Trait critical examination was completed to rank the characteristics by importance utilizing data gain. Relationship based Feature choice (CFS) utilizing BFS procedure and Discriminant examination were utilized to rank and choose the characteristics that are generally helpful. A portion of the affiliation rule mining procedure was utilized to find connection between irrelevant factors in vast database.

1) *Correlation-Based Feature Selection(CFS):* Highlight choice is the procedure to choose a subset of info information examination and future expectation by taking out insignificant data in indicator characteristics. It lessens the intricacy and increment the prescient data. In proposed strategy, the CFS approach is utilized to distinguish include subset which is exceptionally connected with class and least related with traits by utilizing best first inquiry technique. Best First Search strategy starts with void arrangement of highlights and creates all conceivable single component developments. The subset with most noteworthy assessment is picked and extended in a similar way by including single highlights. On the off chance that growing a subset results in no improvement, the hunt back to the following best unexpanded subset and proceeds from that point.

$$M_i = \frac{Kr_{cf}}{\sqrt{K+K(K-1)r_{ff}}}$$

Where Mi is the heuristic "merit" of highlight subset S containing K highlights, r_cfis the mean component class relationship and r_ff is the normal element include between connection. **3.2.2ID3 (Iterative Dichotomizer 3)**

ID3 (Iterative Dichotomizer 3) calculation is concocted by J. Ross Quinlan in 1979. It is utilized for building the choice tree utilizing data hypothesis. It manufactures the choice tree from best down methodology with no backtracking. Data Gain is utilized to choose the best characteristic for order. **Calculation:**

figure grouping entropy.
For all qualities
Ascertain data gain for each property
Select the characteristic with most elevated Information gain
Expel the trait
End

2) *Entropy:* Entropy proportion of vulnerability about a wellspring of message. It lies between 0 to 1.When entropy is 1 implies dataset is homogenous. Entropy is determined by recipe:

$$E(S) = \sum_{j=1}^{c} -P_j \ log_2 P_j$$

Where E(S) is the Entropy of S, Pj is the likelihood of S having a place with class j.

3) *Information Gain:* It gauges the normal decrease in entropy. ID3 figures the Gain everything being equal, and select the one with most noteworthy increase.

$$G(S,A) = E(S) - \sum_{V \epsilon Variables(A)} \frac{|S_V|}{|S|} E(S_V)$$

Where G(S,A) is Information gain ,Sv is the subset of S for which the property A has esteem v, values(A) is the arrangement of every single imaginable incentive for trait A.

C. Evaluating Performance

To assess the execution of grouping rule F-Measure, review, exactness techniques are utilized.

Genuine Positive = $\frac{TP}{N}$

False Negative = $\frac{FP}{N}$

Review = $\frac{TP}{TP+FN}$

Exactness = $\frac{TP}{TP+FP}$

F-measure = $\frac{2* \text{Recall} * \text{Precison}}{Recall+Precision}$

©IJRASET: All Rights are Reserved

IV. RESULT & DISCUSSION

The accompanying table speaks to the statistic data's, for example, Age, Category, Martial Students, Resedential Status, Mother Tongue, Relegion, and Family Type of about understudies.

Table2 - Demographic Details of Student

Demographic factors	Particulars	Frequency (no. of students)	Percent
Age	Below 18	70	31.8
	18-20	131	59.5
	20 & Above	19	8.6
Category	General	172	78.2
	OBC	45	20.5
	SC	3	1.4
Conjugal Status	Unmarried	220	100
Private Status	Urban Rural	169	76.8
		51	23.2
Mother Tongue	Hindi Others	209	95
		11	5
Religion	Hindu	204	92.7
	Jainism	7	3.18
	Sikh	7	3.18
	Muslim	2	0.9
Family Type	Nuclear Joint	115	52.3
		105	47.7

The accompanying diagram speaks to the Educational subtleties of guardians

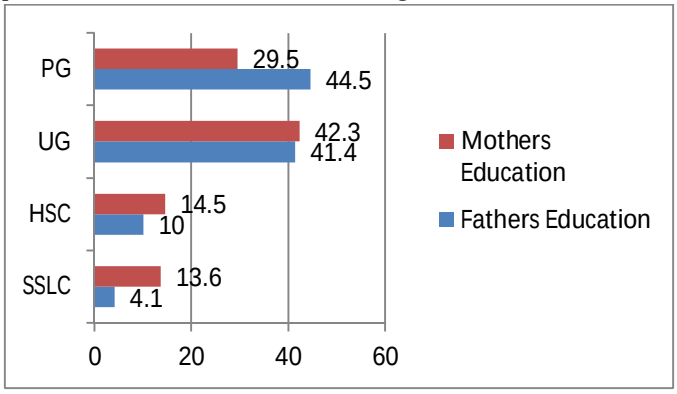


Figure 2 – Educational Details of Parents

The accompanying table speaks to the dropout rates because of the course picked.

Table 3 – Dropout Vs Course Opted

Degree	Dropout Due to Course		
	No	Yes	Total
BCA	97	31	128
B.Tech	86	6	92
Total	183	37	220

The Following table speaks to the understudies are drop out because of the family reasons.

©IJRASET: All Rights are Reserved

Table 4 –

Dropout	Family Problem		Total
	No	Yes	
No	142	41	183
Yes	18	19	37
Total	160	60	220

Dropout Vs Family Problem

The Following table speaks to the understudies are drop out because of abhorrence of grounds condition.

Table 4 – Dropout Vs Campus Environment

Dropout	Like Campus Environment		Total
	No	Yes	
No	168	15	183
Yes	19	18	37
Total	187	33	220

V. CONCLUSION

The principle reason for the examination was to explore the central point causing the dropout of understudies in undergrad ICT Courses at Residential University. In light of the audit of the related writing, essential inquiries were defined (see Annex-I) to demonstrate the idea of expected connections among different parameters considered in this investigation. To confirm the expressed suppositions, the Study had utilized diverse strategies and procedures. Specifically, the examination was led taking 220 examples from first year of BCA and B.Tech understudies of Computer office at Residential University. Information were gathered in pre-planned arrangement which was given over to the understudy's alongside directions. The produced data will be very valuable for the executives of college to create arrangements and techniques for better arranging and usage of instructive program and framework under quantifiable condition to expand the enrolment rate in University and to take powerful choice to lessen understudy dropout.

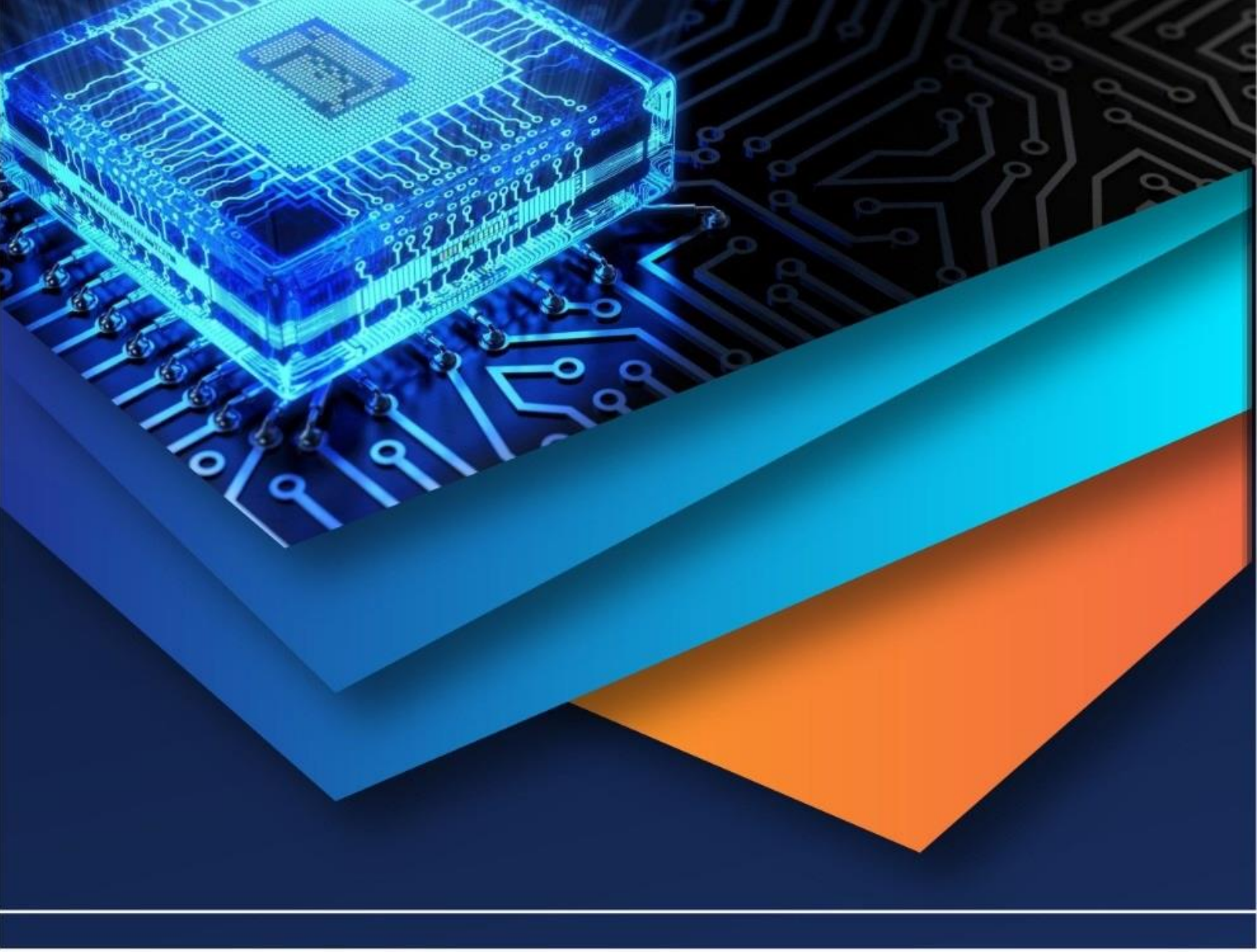
VI. FUTURE WORK

Future work is to contemplate on huge database of dropout understudy at the college utilizing other information mining methods, for example, Logistic Regression, Clustering and Neural Network so as to decide likenesses and connection between different elements.

REFERENCES

[1] Alaa el-Halees (2009) Mining Students Data to Analyze e-Learning Behavior: A Case Study.
[2] Bharadwaj B.K. what's more, Pal S. "Mining Educational Data to Analyze Students? Performance", International Journal of Advance Computer Science and Applications (IJACSA), Vol. 2, No. 6, pp. 63-69.
[3] Fadzilah S. what's more, Abdoulha M. (2009), "Revealing Hidden Information Within University's Student Enrollment Data Using Data Mining", In Proceedings of the Third Asia International Conference on Modeling and Simulation Conference, IEEE PC society.
[4] Ramasubramanian, P. Iyakutti, K. what's more, Thangavelu, P. (2009) "Upgraded Data Mining Analysis in Higher Educational System Using Rough Set Theory", African Journal of Mathematics and Computer Science Research Vol. 2(9), pp. 184-188.
[5] Agrawal, R. what's more, Srikant, R. 1994. Quick calculations for mining affiliation rules. In Proc. twentieth Int. Conf. Large Data Bases, VLDB, J. B. Bocca, M. Jarke, and C. Zaniolo, Eds. Morgan Kaufmann, 487-499.
[6] Al-Radaideh, Q. An., Al-Shawakfa, E. M., and Al-Najjar, M. I. (2006). Mining understudy information utilizing choice trees. In the Proceedings of the 2006 International Arab Conference on Information Technology (ACIT'2006).
[7] Ayesha, S. , Mustafa, T. , Sattar, A. what's more, Khan, I. (2010) "Information Mining Model for Higher Education System", European Journal of Scientific Research, vol. 43, no. 1, pp. 24-29.
[8] Bharadwaj B.K. what's more, Pal S. "Mining Educational Data to Analyze Students Performance", International Journal of Advance Computer Science and Applications (IJACSA), Vol. 2, No. 6, pp. 63-69.
[9] Blazenka Divjak. "Forecast of scholastic execution utilizing discriminant examination", Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, june 22-25,2009 pp. 225-229
[10] Boero, G., Laureti, T., and Naylor, R. (2005). An econometric investigation of understudy withdrawal and movement in post-change Italian colleges. Centro Ricerche Economiche Nord Sud - CRENoS Working Paper 2005/04.
[11] Dr.S.Chitra,Mrs.G.Mohanaprabha "Plan and Development of an effective various leveled approach for multi mark protein work predicition" ,Biomedical Research Special Issue.India 2017 PP-S370-379..
[12] Mr.E.Balraj, Mrs.D.Maalini (2018) “A Survey On Predicting Student Dropout Analysis Using Data Mining Algorithms”, International Journal of Pure and Applied Mathematics,Vol 118 issue8 pg.no621-626.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)