IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.7, Special Issue.2, Jan 2019 E-ISSN: 2347-2693

Different Classification Technique using Analysis of Student Academic Dataset

N. Umarani[1*], R. Rajakumar [2]

[1]Department of Computer Science, MaruthuPandiyar College, Thanjavur, India
[2]Dept. of Computer Science, PG and Research Department, Marudupandiyar College, Thanjavur, India

Available online at: www.ijcseonline.org

***Abstract*—**Data mining methods are executed in numerous associations as a standard technique for breaking down the vast volume of accessible data, removing valuable data and information to help the real basic leadership forms. Data mining can be connected to wide assortment of utilizations in the instructive division to improve the execution of understudies and additionally the status of the instructive foundations. Instructive data mining is quickly creating as a key method in the examination of data produced in the instructive space. The point of this examination displays an investigation of each semester consequences of UG certificate understudies utilizing data mining strategy. This research work thinks about the outcome characterization algorithms. The correlation is finished utilizing the estimation of precision and estimations of Error Rate. This research work likewise demonstrates what algorithm is most reasonable for anticipating the execution of the understudies among the chose algorithms. The examination work is finished by considering different kinds of algorithm like choice tree algorithm, rule based algorithm, Bayesian algorithm and function based algorithms. This nonexclusive novel methodology can be reached out to different trains too.

***Keywords*—**Data mining, Classification, Data collection

I. INTRODUCTION

Advanced education has picked up significance manifolds in the previous couple of decades. The higher instructive establishments are compelled to update its extension and items as a result of the private investment. The controller of administrative body has put a few rules as to foundation, personnel and different assets. New advancements are being created in the field of data the board and investigation because of expansive supply of data being available in a few organizations, including both private and open. The principle point of the systems of data mining is to find covered up and irrelevant connections inside the data having differing qualities. Different methods of data mining are being utilized in various fields including the instructive condition. An extremely promising territory to accomplish this goal is the use of Data Mining (DM). Truth be told, order is a standout amongst the most accommodating DM work in instruction.

Data mining has been executed well in the business applications, however its utilization in advanced education and higher learning establishments is still moderately new. In the area of instruction, instructive data mining ends up being a developing practice which is extremely later and its training is biased to recognize and remove new and profitable information from the data. The point is to determine issues of research regions of training and enhance the entire instructive process utilizing different measurable methods, machine getting the hang of programming (MLP) and data mining algorithms. Instructive data Mining (EDM) is a thriving practice that can be utilized for investigation and representation of data, forecast of understudy execution, understudy demonstrating, gathering of understudies and so forth.

Instructive Data Mining is centeredaround creating strategies to investigate the exceptional and progressively substantial dataset which touches base from instructive sources and further utilizing those techniques to comprehend the understudies and the earth in which they learn betterly. Instructive Data Mining (EDM) is the procedure to change over crude data from training frameworks to advantageous data which can be further be utilized by guardians, educators, instructive designers, other instructive researchers and understudies.

In Educational Data Mining, the Student's execution in scholastic accomplishment is the significant worries in the universities.The expanding of understudies going to college has built up the enthusiasm for recognizing variable to anticipate scholarly execution. In advanced education, the issue of forecast and clarification of scholarly execution and an examination to distinguish the key pointers to the scholastic achievement and perseverance of understudies are critical.

II. Related Work

Romero and Ventura, covering the research endeavors in the region somewhere in the range of 1995 and 2005 in Education area, and by Baker and Yacef for the period after 2005 in Education space. Luan examines in the potential utilizations of data mining in advanced education and clarifies how data mining spares assets while augmenting productivity in scholastics. Understanding understudy types and focused on showcasing dependent on data mining models are the research subjects of a few papers. The usage of prescient displaying for amplifying understudy enlistment and maintenance is exhibited in the investigation of Noel-Levitz. These issues are likewise talked about by DeLongetal. The advancement of enlistment forecast models dependent on understudy confirmations data by applying distinctive data mining techniques is the research focal point of Nandeshwar and Chaudhari. Dekkeretal. center around anticipating understudies drop out. Kovacicin utilizes data mining strategies (highlight determination and grouping trees) to
investigate the socio-statistic factors (age, sex, ethnicity, education, work status, and incapacity) and study condition (course program and course hinder) that may impact tirelessness or dropout of understudies.
Ramaswami and Bhaskarancenter around creating prescient data mining model to recognize the moderate students and concentrate the impact of the predominant factors on their scholastic execution, utilizing the prevalent CHAID choice tree algorithm. Yuetal investigate understudy maintenance by utilizing grouping trees, Multivariate Adaptive Regression Splines (MARS), and neural systems. Cortez and Silva attempt to anticipate understudy disappointment by applying and looking at four data mining algorithms − Decision Tree, Arbitrary Forest, Neural Network and Support Vector Machine. Kotsiantiset al. apply five order algorithms (Decision Tree, Perceptron-based Learning, Bayesian Net, Instance Based Learning and Rule-learning) to foresee the execution of software engineering understudies from separation learning.

III. Data Collection Methodology

There are different strategies are utilized to gather the data viewing the students, for example, we have arranged inquiries in google spreadsheet and shared it among the students of different organizations. We likewise have arranged survey in hardcopy and shared it to the students to gather the information. We additionally have arranged a site alongside the poll to gather the information from the establishments. By utilizing these different systems we have gathered around 3600 understudy's information that covers the data like understudy's statistic, scholastic and learning conduct.

IV. Used Tools and Technology

During this exploration examination we have utilized WEKA and SPSS instruments. WEKA is open source information mining examination apparatus. We have utilized this instrument to break down different characterization algorithms and to analyze the consequence of these algorithms. We additionally have utilized SPSS measurable device to locate the most impact parameters on the understudy's execution improvement among the gathered parameters.

V. Proposed Model

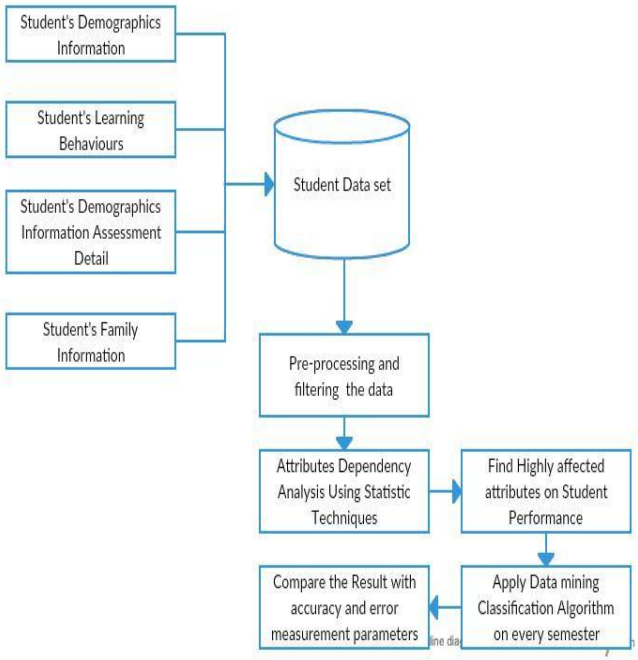

Figure 1. Research Model

Step-wise procedure for Implementation of Model:
Step 1: Collect the student information (Demographic Information, Academic Information, learning Behavioral Information).
Step 2: Perform the information pre-preparing and change.
Step 3: Apply Statistical systems for finding exceptionally influenced parameters on understudies' execution.
Step 4: Apply different information mining methods (Classification, Clustering, and Association) on understudy informational collection.
Step 5: Find the Most Optimized Model and create the learning.

Objective of Research Model:
The result from this examination is required to be utilized for distinguishing the components impacting students' scholastic execution. In expansion, the expectation model could be utilized by the board to structure unique program for the "remarkable" and the "low" achievers for every degree

program. In along these lines, students who are required to do well could be pushed to get the phenomenal dimension. On the other hand, students who are relied upon to be low achievers could be helped to increase better evaluations upon graduation. This is to guarantee the nature of alumni is another continue or advancement in a positive direction.

VI. USED PARAMETERS ANALYSIS

Table 1: Used student's parameter in research work

ATTRIBUTES	DATA TYPE	POSSIBLE VALUES
Gen	Nominal	Male, female
Percentagehsc	Nominal	Poor, average, good, very_good, excellent
Stream	Nominal	Commerce, science
F_annual_income	Nominal	Low, average, middle, high, very high
F_qualification	Categorical	No formal education, primary, ssce, 1st degree, 2nd degree, phd
F_occupation	Categorical	Unemployed, government worker, private, self employed
M_qulification	Categorical	No formal education, primary, ssce, 1st degree, 2nd degree, phd
M_occupation	Categorical	Unemployed, government worker, private, self employed
No_of_sublings	Categorical	One, two, three, four
Overall_attendance	Nominal	Poor, average, good, very_good, excellent
W_l_h	Nominal	Poor, average, good, very_good, excellent
W_li_u	Nominal	Poor, average, good, very_good, excellent
D_re_h	Nominal	Poor, average, good, very_good, excellent
E_w_l_u_h	Nominal	Poor, average, good, very_good, excellent
Internal_marks	Nominal	Poor, average, good, very_good, excellent
Assignment_ marks	Nominal	Poor, average, good, very_good, excellent
Participation_extra_curriculum	Nominal	Poor, average, good, very_good, excellent
Practical_knowledge	Nominal	Poor, average, good, very_good, excellent
Theory_marks	Nominal	Poor, average, good, very_good, excellent
Internet_uses_learning	Nominal	Poor, average, good, very_good, excellent
Previous_sem_marks	Nominal	Poor, average, good, very_good, excellent
Subject Name	Nominal	Subject Name
Internal_Th_Marks	Nominal	Poor, average, good, very_good, excellent
Internal_Pr_Marks	Nominal	Poor, average, good, very_good, excellent
External_Th_Marks	Nominal	Poor, average, good, very_good, excellent
External_Pr_Marks	Nominal	Poor, average, good, very_good, excellent
Subject_Attendance	Nominal	Poor, average, good, very_good, excellent
Subject_Faculty_Performanace	Nominal	Poor, average, good, very_good, excellent
Subject Result	Nominal	Poor, average, good, very_good, excellent
Semester_wise_result	Nominal	Poor, average, good, very_good, excellent

VII. SPSS EXPERIMENTAL ANALYSIS

In the following table we have found the coefficient table after performing the statistical analysis into the SPSS tool.

Table 2: Coefficients of used variables

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.237	.075		3.149	.002
Gen	-.022	.015	-.010	-1.465	.143
Percentagehsc	-.009	.005	-.011	-1.648	.000
Stream	.018	.016	.008	1.129	.259
F_annual_income	-.019	.005	-.026	-3.717	.000
FQ	.001	.005	.001	.186	.853
FP	.005	.006	.005	.798	.425
MQ	.012	.008	.015	1.589	.112
MP	-.025	.014	-.017	-1.736	.083
NOS	.095	.011	.057	8.301	.000
Overall_attendance	.207	.010	.183	21.058	.000
W_L_H	-.013	.005	-.018	-2.626	.004
W_Li_U	-.007	.005	-.010	-1.412	.158
D_Re_H	.001	.005	.002	.270	.787
E_W_L_U_H	.008	.005	.010	1.498	.134
INTERNAL_MARKS	.265	.009	.249	29.749	.000
ASSIGNMENT_MARKS	-.013	.007	-.014	-1.820	.000
PATICIPATION_EXTRA_CURRICULAM	.002	.008	.001	.196	.000
PRACTICAL_KNOLEDGE	.167	.013	.187	12.765	.000
THEORY_MARKS	.021	.013	.022	1.592	.000
INTERNET_USES_LEARNI NG	-.248	.016	-.114	-15.145	.003
PREVIOUS_SEM_MARKS	.390	.010	.422	37.925	.000

A multiple regression was rushed to predict Sixth_sem_result from autonomous variables. These variables measurably essentially predicted Sixth_sem_result, $F(21, 3537) = 863.946$, $p < .0005$, So, hold to those variables whose significate level is < 0.0005 and expel those variables whose importance level is > 0.0005 from the model.

Selected highly affected parameters on student's performance after SPSS analysis:

Table 3: Highly affected parameters on student's performance

Percentage HSC	Assignment_Marks
F_Annual_Income	Practical_Knowledge
Weekly Lab Hours	Theory_Marks
Overall Attendance	Internet_Uses_Learning
Internal_Marks	Previous_Semester_Marks
Participation Extra Curriculum	No.ofSublings

VIII. WEKA EXPERIMENTAL RESULT

In this weka experimental analysis we have used various classification algorithms like J48, Bayes Net, Decision stump, Logistic Regression, Multi-layer perception, Naïve Bayes, One R, Rep Tree, and sequential minimal optimization . After that we have compared these algorithm using the WEKA tool. The semester wise comparative result is described as per the following in table.

Table 4 : Semester wise time taken to build the model by different classifiers

Semesters	**J48**	**BN**	**DS**	**LS**	**MLP**	**NB**	**1R**	**RT**	**SMO**
Sem I	0.02	0.8	0.9	2.41	56.43	0.09	0.18	0.11	2.25
Sem II	0.0502	0.1202	0.9202	3.4302	66.4502	0.1002	0.1402	0.1202	1.2702
Sem III	0.0625	0.1325	0.9325	3.4425	66.4625	0.1125	0.1525	0.1325	1.2825
Sem IV	0.059	0.1299	0.9299	3.4399	66.4599	0.1099	0.1499	0.1299	1.2799
Sem V	0.0573	0.1273	0.9273	3.4373	66.4573	0.1073	0.1473	0.1273	1.2773
Sem VI	0.03	0.1	0.9	3.41	66.43	0.08	0.12	0.1	1.25
Mean Value	0.046	0.234	0.9183	3.26165	64.781	0.0999	0.14832	0.11998	1.4349

Table 5: Semester wise correctly classified instance by different classifiers

Semesters	**J48**	**BN**	**DS**	**LS**	**MLP**	**NB**	**1R**	**RT**	**SMO**
Sem I	99.4342	97.4227	60.8764	97.5289	87.157	97.4827	78.29	97.391	87.1
Sem II	99.12	96.4265	69.8764	97.3141	92.2976	97.3984	82.29	95.348	92.3257
Sem III	99.129	97.4365	71.8864	97.8241	93.8976	98.8984	84.29	96.448	93.5257
Sem IV	99.2695	97.5365	71.9864	98.0241	93.9076	98.9184	84.59	97.438	94.0257
Sem V	99.325	97.6565	72.0264	98.4241	94.7076	98.9284	84.75	97.458	94.1257
Sem VI	99.07	98.4827	59.8764	98.5389	89.157	98.4827	79.29	98.391	89.1
Value	99.224	97.4935	67.7547	97.942	91.854	98.3515	82.25	97.079	91.700

Table 6: Semester wise in correctly classified instance by different classifiers

Semesters	**J48**	**BN**	**DS**	**LS**	**MLP**	**NB**	**1R**	**RT**	**SMO**
Sem I	0.3658	2.5173	39.1236	2.4611	12.8429	2.517	12.708	2.2082	12.899
Sem II	0.88	3.5735	30.1236	2.6859	7.7024	2.601	17.7081	4.652	7.6743
Sem III	0.871	2.5635	28.1136	2.1759	6.1024	1.101	15.7081	3.552	6.4743
Sem IV	0.7305	2.4635	28.0136	1.9759	6.0924	1.081	15.4081	2.562	5.9743
Sem V	0.6743	2.3435	27.9736	1.5759	5.2924	1.071	15.2481	2.542	5.8743
Sem VI	0.92	1.5173	40.1236	1.4611	0.8429	1.517	20.708	1.2082	0.899
Mean Value	0.7402	2.4964	32.245	2.0559	6.479233	1.648	16.2481	2.7874	6.6325

Table 7: Semester wise kappa statistics rate by different classifiers

Semesters	**J48**	**BN**	**DS**	**LS**	**MLP**	**NB**	**1R**	**RT**	**SMO**
Sem I	0.9752	0.9695	0.296	0.9680	0.9686	0.9695	0.6175	0.9737	0.97
Sem II	0.9489	0.9219	0.65646	0.9308	0.8806	0.93168	0.7806	0.9111	0.88
Sem III	0.9489	0.9320	0.67656	0.9359	0.8966	0.94668	0.8006	0.9221	0.89
Sem IV	0.9503	0.9330	0.67756	0.9379	0.8967	0.94688	0.8036	0.9320	0.89
Sem V	0.9509	0.9342	0.67796	0.9419	0.904	0.94698	0.805	0.9322	0.89
Mean Value	0.9590	0.94506	0.5634	0.9474	0.9210	0.9535	0.7541	0.94085	0.92

Table 8: Semester wise Mean Absulate Error by different classifiers (MAE)

Semesters	**J48**	**BN**	**DS**	**LS**	**MLP**	**NB**	**1R**	**RT**	**SMO**

Sem I	0.0173	0.021	0.299	0.0256	0.1146	0.021	0.092	0.0256	0.14
Sem II	0.0093	0.013	0.201	0.0076	0.0066	0.013	0.084	0.0076	0.2424
Sem III	0.0103	0.014	0.202	0.0086	0.0076	0.014	0.085	0.0086	0.2434
Sem IV	0.0113	0.015	0.203	0.0096	0.0086	0.015	0.086	0.0096	0.2444
Sem V	0.0123	0.016	0.204	0.0106	0.0096	0.016	0.087	0.0106	0.2454
Sem VI	0.005	0.011	0.199	0.0066	0.0069	0.011	0.082	0.0056	0.24
Mean Value	0.0109	0.0155	0.2189	0.011433	0.02565	0.0158	0.086	0.01126	0.2259

Table 9: Semester Root Mean Squared Error Rate by different classifiers (RMSE)

Semesters	**J48**	**BN**	**DS**	**LS**	**MLP**	**NB**	**1R**	**RT**	**SMO**
Sem I	0.0541	0.0667	0.2162	0.0634	0.0607	0.0672	0.1878	0.168	0.216
Sem II	0.0841	0.0967	0.3362	0.0934	0.0707	0.0972	0.3078	0.088	0.3364
Sem III	0.0941	0.1067	0.3462	0.1034	0.0807	0.1072	0.3178	0.098	0.3464
Sem IV	0.1041	0.1167	0.3562	0.1134	0.0907	0.1172	0.3278	0.106	0.3564
Sem V	0.1141	0.1267	0.3662	0.1234	0.1007	0.1272	0.3378	0.118	0.3664
Sem VI	0.05	0.0767	0.3162	0.0734	0.0607	0.0772	0.2878	0.068	0.316
Mean Value	0.0834	0.0983	0.3228	0.0950	0.0773	0.0988	0.2944	0.108	0.3229

We have done analysis utilizing the weka tool. In this examination investigation we have connected different arrangement calculations into the WEKA tool like J48, Bayes Net, Decision stump, Logistic Regression, Multi layer observation, Naïve Bayes, One R, Rep Tree, and successive negligible advancement and getting the semester insightful execution of characterized calculation in particular to the utilized precision estimated and blunder estimated parameters. In this examination we have utilized exactness estimated parameters like time taken to fabricate the model, accurately ordered examples and mistakenly grouped occurrences. In this examination we have utilized mistake estimation parameters like kappa measurements, mean outright blunder (MAE) and Root mean square error(RMSE). After the investigation we reasoned that among all the characterization calculations, J48 calculation gives the most elevated precise outcome and it has the least mistake rate. It additionally requires the less investment to manufacture the model. Thus, we presumed that J48 gives the most noteworthy exact calculation.

IX. CONCLUSION

Assessment of students‟ execution and holding the standard of training is an essential issue in all the educational foundations. Information mining techniques are frequently actualized for breaking down accessible information and extracting Information and learning to help basic leadership. In this exploration paper connected diverse order algorithms of information digging those are utilized for improvement of an information digging model for expectations of exhibitions of students, based on their own statistic and scholarly data. This examination is finished by WEKA device. Results are as Accuracy of the classifiers and Error Rate of the classifiers. Theories produced results are thought about and watch what algorithm is ideal for this sorts of dataset. As a result of perception seen that in both the model J48 algorithm gives the higher precision and Lower Error rate.This examination work is finished by considering just chosen sorts of algorithm this work is extended by choosing other algorithm moreover. This work is utilized in instruction domain however this conventional novel methodology can be stretched out to different trains moreover.

REFERENCES

[1]. Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M, Mining Student Data Using Decision Trees‟, The 2006 International Arab Conference on Information Technology (ACIT'2006) – Conference Proceedings 2006.

[2]. Ayesha, S. , Mustafa, T. , Sattar, A. and Khan, I. Data Mining Model for Higher Education System‟, European Journal of Scientific Research, vol. 43, no. 1, pp. 24-29. 2010.
[3]. Baradwaj, B. and Pal, S. Mining Educational Data to Analyze Student s‟ Performance‟, International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63-69.2011.
[4]. Chandra, E. and Nandhini, K. Knowledge Mining from Student Data‟, European Journal of Scientific Research, vol. 47, no. 1, pp. 156-163.2010.
[5]. El-Halees, A. Mining Students Data to Analyze Learning Behavior: A Case Study‟, The 2008 international Arab Conference of Information Technology (ACIT2008) – Conference Proceedings, University of Sfax, Tunisia, Dec 15-18. 2018.
[6]. Han, J. and Kamber, M. Data Mining: Concepts and Techniques, 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor 2016.
[7]. Kumar, V. and Chadha, A. An Empirical Study of the Applications of Data Mining Techniques in Higher Education‟, International Journal of Advanced Computer Science and Applications, vol. 2, no. 3, pp. 80-84.2011.
[8]. Mansur, M. O. Sap, M. and Noor, M. Outlier Detection Technique in Data Mining: A Research Perspective‟, In Postgraduate Annual Research Seminar.2005.
[9]. Romero, C. and Ventura, S. Educational data Mining: A Survey from 1995 to 2005‟, Expert Systems with Applications (33), pp. 135-146.2007.
[10]. Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, “Mining student data using decision trees”, International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.
[11]. U. K. Pandey, and S. Pal, “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, 2011.
[12]. Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, “Data mining model for higher education system”, European Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010.

