ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.6, 3821–3826
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.6.61 Spec Issiue on ICAML-2020

PREDICTION OF STUDENT DROPOUT USING ENHANCED MACHINE LEARNING ALGORITHM

G. SALLAN[1] AND S. BEHAL

ABSTRACT. In this paper, an AI based expert system is proposed for predicting the reasons for drop out students in the college using data mining classification technique. The proposed system give benefit to the students, parents, teachers, academicians, society and owners to analyze the performance of the students and to impart the quality of education in the educational institutions and make decision regarding the improvements. As part of the work, machine learning based Decision Stump, NDTREE and Enhanced Machine Learning algorithms (EMLA) are used. The proposed algorithms are evaluated using benchmarked Students dropout dataset. The reporting classification results shows that EMLA produce accuracy of 78.31%, NDTREE of 70.02% and Decision Stump of 30.71%. In the future works, the authors shall evaluate other data mining techniques for improving student performance.

1. INTRODUCTION

Training is a fundamental factor for building up a nation. Scholastic accomplishment is the part assessment in instruction the executives. A consultant framework could assist the understudy with getting the exhibition better since the quantity of resigned or dropped understudies out is less. Learning the board

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 68T05, 68T37.

Key words and phrases. dropout prediction, decision dtump, NDTREE, EMLA (Enhanced Machine Learning Algorithm).

frameworks produce a lot of information, where information revelation is conceivable utilizing information mining procedures, [4].

The previous forecast of dropout understudy is testing task in the advanced education. Information examination is one approach to downsize the pace of dropout understudies and increment the enlistment pace of understudies in the school. It is actuality that understudy dropout regularly in the primary year of graduation. Dropout in private school is brought about by scholastic, family and individual reasons, grounds condition and foundation of school and fluctuates relying upon the instructive framework embraced by the school, [7]. The primary point of this paper is to plan an order model utilizing choice tree acceptance calculation and classifier rules to anticipate whether understudy will graduate or not utilizing the noteworthy information. In this paper, Decision Stump, NDTREE, Enhanced Machine Learning Algorithm choice tree calculations is utilized to plan a theoretical model.

Information Mining (DM) or Knowledge Discovery in Database (KDD) is a procedure to discover examples and connection from huge informational collections with the different strategies, for example, AI, measurements, and database framework. Grouping systems are totally founded on AI. These strategies order each dataset into predefined classes. To order information in database some numerical methods like neural system, choice trees, measurements and straight writing computer programs are utilized. We attempt to comprehend the issue of characterization with a genuine model. Give us a chance to take a case of College where various understudies are examining. With the assistance of order systems we can foresee about those understudies who may have instructive dropout in not so distant future. We can likewise order distinctive understudy as per their exhibition in their examination, [2,5]. There are many machine learning based classification algorithms and case studies about student dropouts are available in the literature [1,3,6,8]. Out of many machine learning based algorithms, we propose to use the following algorithms in this work.

2. Decision Tree based Machine Learning Methods

- Decision Stump: A choice stump is an AI model containing a one-level decision tree. A decision stump makes a desire subject to the estimation of just a singular information feature. Some of the time they are

additionally called 1-rules. Contingent upon the sort of the information highlight, a few varieties are conceivable. For ostensible highlights, one may gather a stump which contains a leaf for each possible component value or a stump with the two leaves, one of which identifies with some picked grouping, and the other leaf to the different orders. For twofold features these two plans are vague.

- NB Tree: The calculation is like the old style recursive parceling plans, then again, actually the leaf hubs made are Naive-Bayes categorizers rather than hubs anticipating a solitary class. A limit for constant traits is picked utilizing the standard entropy minimization strategy, as is accomplished for choice trees.
- Enhanced Machine Learning Algorithm (EMLA) : An Enhanced Machine Learning Algorithm compute the closeness of an explicit component in a class to the approximation of some other component in terms of proximity. The model is give as:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)},$$

where $P(c|x) = P(x_1|c) \times P(x_1|c) \times ....P(x_n|c)$. $P(c|x)$ is the posterior probability of class (c, target) given predictor (x, attributes), $P(c)$ is the prior probability of class, $P(x|c)$ is the likelihood which is the probability of predictor given class, $P(x)$ is the prior probability of predictor.

3. Tools Used

We used IDE of NetBeans for Java for implementing the proposed work. Further, we used an AI based data mining tool called WEKA. It contains a collection of visualization tools and algorithms for data analysis and predictive modeling together with graphical user interfaces for easy access to this functionality.

4. Results and Discussion

We used students dropout dataset to evaluate the performance of proposed algorithms using proposed approach. This dataset consists of 5 attributes namely: Student name, Roll No., Course, Semester, Reason. It has approximately 5 attributes and 407 instances. The proposed approach is shown in figure 1.

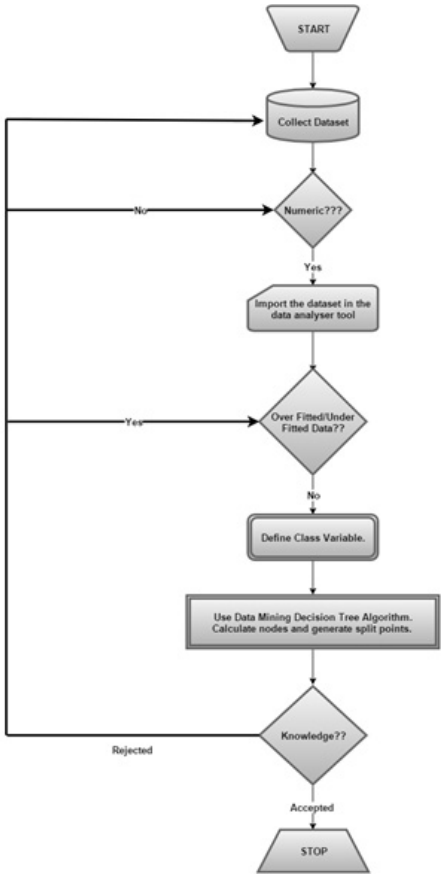

FIGURE 1. Proposed Research Methodology

The results of comparison of Decision Stump, NDTREE and EMLA Algorithms are shown in Figure 2 and Table 1. It has been observed that Decision Stump algorithm correctly classified 30.71% , NDTREE classified 70.02% and EMLA classified 78.37 % of instances. Whereas these three algorithms produced error rate of 99.58%, 81.23% and 81.64% respectively. So, it is concluded that EMLA algorithm gave best results as compared to Decision Stump and NDTREE algorithms in predicting the dropout rate of students.

5. CONCLUSION

In this research, an AI based expert system is proposed for predicting the drop out rate of students in a college using data mining classification techniques. The

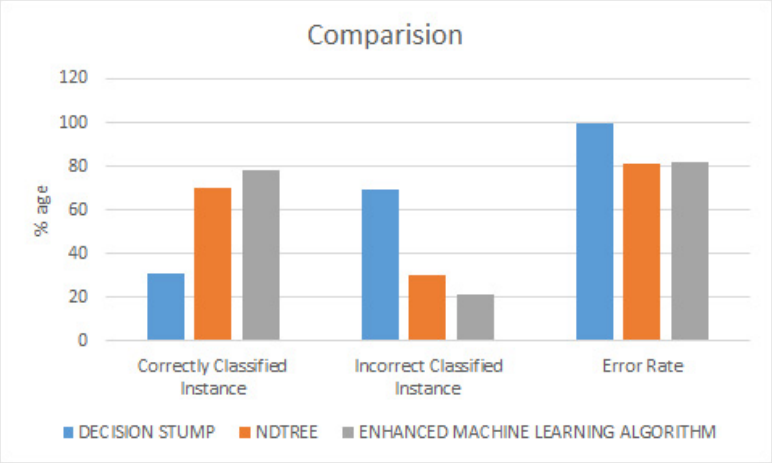

FIGURE 2. Comparison of Decision Stump, NDTREE and EMLA Algorithms

TABLE 1. Comparison of Decision Stump, NDTREE and EMLA Algorithms

Parameter	Decision Stump	NDTREE	EMLA
Correctly Classified Instance	30.71	70.02	78.37
Incorrect Classified Instance	69.28	29.97	21.62
Error Rate	99.58	81.23	81.64

system may give benefit to the students, parents, teachers, Academicians, society and owners to analyze the performance of the students and to impart the quality of education in the educational institutions and make decision regarding the improvements. The classification results illustrated that the EMLA is the best algorithm with detection rate of 78.31% as compared to NDTREE and Decision Stump algorithms. The proposed approach is validated using benchmarked Students dropout dataset. The future works is to study on large database of dropout student at the university using other data mining techniques such as Naive Bayes, classification and association in order to determine similarities and relationship between multiple factors.

REFERENCES

[1] L. AULUCK, N. VELAGAPUDI, J. BLUMENSTOCK, J. WEST: *Predicting student dropout in higher education*, 2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications, New York, (2016), 16–20.

[2] N. DENGEN, E. BUDIMAN, M. WATI, U. HAIRAH: *Student Academic Evaluation using Naïve Bayes Classifier Algorithm*, 2nd IEEE East Indonesia Conference on Computer and Information Technology (EIConCIT), (2018), 104–107.

[3] A. KASHYAP, A. NAYAK: *Different machine learning models to predict dropouts in MOOCs*, IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), (2018), 80–85.

[4] N. KETUI, W. WISOMKA, K. HOMJUN: *Using Classification Data Mining Techniques for Students Performance Prediction*, IEEE Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON), (2019), 359–363.

[5] K. LIMSATHITWONG, K. TIWATTHANONT, T. YATSUNGNOEN: *Dropout prediction system to reduce discontinue study rate of information technology students*, IEEE 5th International Conference on Business and Industrial Research (ICBIR), (2018), 110–114.

[6] C. MÁRQUEZ-VERA, A. CANO, A. Y. M. NOAMAN H. MOUSA, S. VENTURA: *Early dropout prediction using data mining: a case study with high school students*, Expert System J. , **33**(1) (2016), 107–124.

[7] O. SUKHBAATAR, K. OGATA AND T. USAGAWA: *Mining Educational Data to Predict Academic Dropouts: a Case Study in Blended Learning Course*, TENCON 2018 IEEE Region 10 Conference, (2018), 2205–2208.

[8] T. THILAGARAJ AND N. SENGOTTAIYAN: *A Review of Educational Data Mining in Higher Education System*, Proceedings of the 2nd International Conference on Research in Intelligent and Computing in Engineering, RICE, (2017), 349–358.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SHAHEED BHAGAT SINGH STATE TECHNICAL CAMPUS, FEROZEPUR
IKG PUNJAB TECHNICAL UNIVERSITY, KAPURTHALA, INDIA
Email address: gautamsallan269@gmail.com

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
SHAHEED BHAGAT SINGH STATE TECHNICAL CAMPUS, FEROZEPUR
IKG PUNJAB TECHNICAL UNIVERSITY, KAPURTHALA, INDIA
Email address: sunnybehal.sbs@gmail.com