

# An Efficient Model for Predicting Student Dropout using Data Mining and Machine Learning Techniques

Mercy Paul Selvan, Nagubadi. Navadurga, Nimmagadda. Lakshmi Prasanna

*Abstract— Education could be an important resource that has to lean to all or any kids. one in all the largest assets of the longer term generation cloud is alleged because the education that's given to the youngsters. Most of the youngsters aren't ready to continue their education because of many reasons. The prediction of student dropout plays a very important role in characteristic the scholars World Health Organization are on the sting of being a dropout from their education. whereas predicting this, we will simply try and solve their issues and create them continue their education. during this paper, we've planned a model for predicting the scholars can get born out or not mistreatment many machine learning techniques. we have a tendency to create use of decision trees that make a call mistreatment many factors. the choice of the prediction involves crucial wherever many knowledge attributes are used for prediction like correlations, similarity measures, frequent patterns, and associations rule mining. The planned work is evaluated mistreatment numerous parameters and is well-tried to figure expeditiously in predicting the dropout students compared with alternative.*

**Keywords—**Education, Classification, Accuracy, Decision Trees, Prediction, Dropout, Machine Learning

## I. INTRODUCTION

Education is one amongst the foremost vital things that has to be to all or any the scholars of the approaching generation. the whole growth of the country depends chiefly on the student's education. however there are several students UN agency are thought capable get dropout from their education thanks to many reasons. a number of the explanations being family issues, money problems and lots of a lot of. There exists a desire for predicting the quantity of scholars UN agency are possible to induce dropout of their education. This prediction would be helpful to induce to understand the precise reasons behind their call in might conjointly facilitate other countries create selections to know what might be done to eradicate that.

The expansion of computer science has born to several new technologies out of that the favored ones are being Machine Learning Approaches and Deep Learning Techniques. Machine learning techniques are wide employed in various applications like medical image analysis [1], golem path coming up with [2], flood detection

during a explicit town or space [3] and land cowl classification [4]. Machine learning technique could be a method of learning a particular task with none human intervention and rising the performance solely by the continual learning process [20,21]. the training is of 2 types: supervised learning [5], wherever the labeling is given for the options of the coaching dataset and unattended learning [6], wherever no labels are given and therefore the system has to label the options of the dataset. a number of the classifiers used preponderantly in machine learning techniques are SVM classifiers [7], call Trees [8], Naive mathematician [9] and supply Regression [10]. once larger datasets are getting used in an application then the utilization is most well-liked for feature extraction because it produces a lot of correct results [11]. Use of ANNs is wide referred to as Deep Learning Approaches because the neural network learns every and each layer very deeply and uses the output of a layer as the input of the following layer. ANNs is a classifier resembling the perform of a biological vegetative cell having various layers connected to every alternative through weights [12]. ANNs are scientific discipline structures which will solve any drawback through learned examples instead of pre-specified algorithms [13].

Machine Learning Approaches finds distinctive importance within the field of education. during this paper, we've projected a model victimisation machine learning techniques that might predict the coed dropout and facilitate United States of America to investigate an answer for the prediction. a call tree is employed for creating the choices for the prediction and numerous knowledge attributes are thought-about for the identical. remainder of the paper consists of some connected works projected by numerous researchers, the methodologies and algorithms employed in the projected technique and experimental results.

## II. RELATED WORK

El-Halees [14] has designed a straightforward case study that uses academic data processing to investigate the varied behavior of scholars learning. the target of his study is to point out how helpful data processing will be employed in educational activity to enhance student's performance. They applied techniques info} mining to uncover big databases like association rules and classification rules employing a call tree, bunch and outlier analysis. Bharadwaj and Pal [15] planned the novel approach exploitation the choice tree methodology for classification to judge the student's performance. This case study is to work out the information

**Revised Manuscript Received on July 18, 2019.**

**Dr. Mercy Paul Selvan**, School of Computing, Sathyabama Institute Of Science And Technology, Chennai, Tamilnadu, India.(email: mercypaulselvan@gmail.com)

**Nagubadi. Navadurga**, School of Computing, Sathyabama Institute Of Science And Technology, Chennai, Tamilnadu, India.(email: navadurga481@gmail.com)

**Nimmagadda. Lakshmi Prasanna**, School of Computing, Sathyabama Institute Of Science And Technology, Chennai, Tamilnadu, India.(email: nimmagaddapvl@gmail.com)

# AN EFFICIENT MODEL FOR PREDICTING STUDENT DROPOUT USING DATA MINING AND MACHINE LEARNING TECHNIQUES

that describes students' performance within the finish semester examination. This study was quite helpful to spot the dropout's student in AN earlier stage and students United Nations agency want special care and permit the mentor to require earlier attention to the scholars. Fadzilah and Abdullah [17], applied data processing techniques to incoming knowledge. Descriptive Analysis and prophetic Analysis cluster the information into clusters supported their similarities, cluster analysis is employed. For prophetic analysis, Neural Network, supplying regression, and also the call Tree are used. once evaluating these techniques, Neural Networks classifier was found to grant the best leads to term of classification accuracy. [18,19] planned predicting aspects of upper education students. during this paper, they analyze that one amongst the most important challenges that educational activity faces nowadays is predicting the behavior of scholars. establishments would really like to grasp, one thing regarding the performances of the student's cluster wise. He planned a controversy to research the performances of the scholars once the big info of the scholars data system (SIS) is given. Generally, students' issues are going to be classified into completely different patterns supported the extent of scholars like traditional, average and below average.

### III. PROPOSED WORK

The rate of the self-made establishment may be proverbial only the explanations for chucking up the sponge of scholars are known by the management and necessary actions are taken to scale back the amount of dropout student. during this projected work, the varied parameters related to the scholars were surveyed for applying machine learning techniques to predict whether or not the coed is probably going to induce dropout of the establishment. a number of the parameters that were thought-about were their family background, money financial gain of the family, tutorial details, social activities, environmental activities of the coed and much additional. Fig 1. shows the varied steps concerned in predicting the coed dropout.

#### Preparation of Data:

During this step, the info is ready by taking a survey of all the scholars within the establishment and asking them to fill the shape with all the mandatory details. during this work, we've used a dataset using machine learning repository that would use for the prediction. a number of the info attributes that are used for data processing are the coed background, grades obtained, money standing of the oldsters and additional. This information is then used for any classification victimisation machine learning techniques.

#### Classification:

Classification is one among the foremost important processes of the whole prediction methodology. In this, we tend to are victimisation call trees. call trees comprises the varied tree that's used for creating a choice. In this, a choice is created whether or not the coed can continue his education or will drop out of his education. the choice trees are used because the classifier to come to a decision by victimisation numerous parameters that are given to the classifier[16]

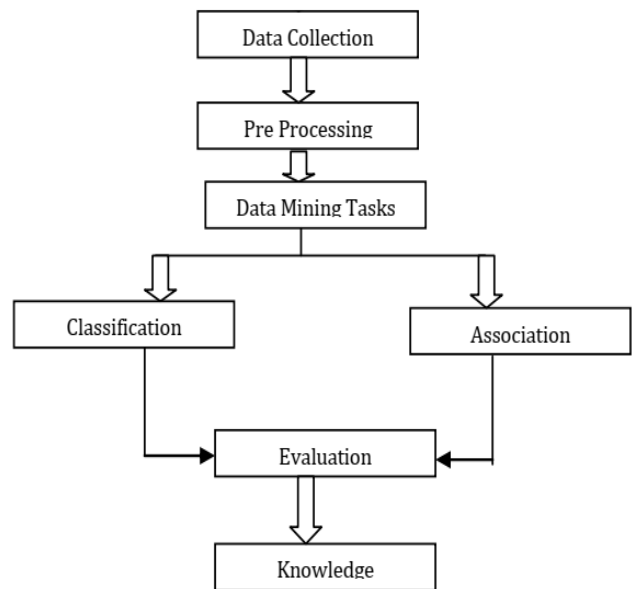


Fig. 1 Workflow of Classification of images

### IV. EXPERIMENTAL RESULTS

The experiments are performed on the MATLAB R2017b version. The computations are performed using image Classification Learner Toolbox that is readily available in MATLAB. First, the input image from the scanner database is given to the tool. Fig. 2 gives the educational details of the parents of the students. Fig. 3 gives the total number of dropout with respect to the course opted by the students and Fig. 4 gives the dropout of students due to family problems. The accuracy is observed with 97.69% of specificities. The prediction was done using various parapets considered for each and every student.

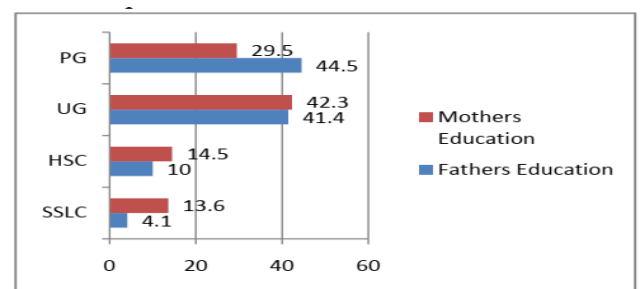


Fig. 2 Educational Details of parents

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

Fig. 3 Drop out Vs Course Opted



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

Fig. 4 Drop out Vs Family Problem

## V CONCLUSION

Education could be a vital want that must incline to any or all the scholars. There are varied issues baby-faced by several and this happens as a reason for his or her dropouts. during this paper, we've got planned a prognosticative model for predicting the quantity of dropouts in their education because of varied reasons. during this model, we tend to created use of decision trees to come to a decision the explanations and predict the dropout rate. The planned work is evaluated exploitation varied parameters and is verified to figure with efficiency in predicting the dropout students compared with alternative existing models.

## VI REFERENCES

1. Rahman, M. M., Bhattacharya, P., & Desai, B. C. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Transactions on Information Technology in Biomedicine*, 11(1), 58-69.
2. Morales, M., Tapia, L., Pearce, R., Rodriguez, S., & Amato, N. M. (2004). A machine learning approach for feature-sensitive motion planning. In *Algorithmic Foundations of Robotics VI*(pp. 361-376). Springer, Berlin, Heidelberg.
3. Ireland, G., Volpi, M., & Petropoulos, G. P. (2015). Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: a case study from a Mediterranean flood. *Remote sensing*, 7(3), 3372-3399.
4. Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4), 725-749.
5. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3-24.
6. Cheriyyadat, A. M. (2014). Unsupervised feature learning for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 439-451.
7. Bauer, S., Nolte, L. P., & Reyes, M. (2011, September). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In the *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 354-361). Springer, Berlin, Heidelberg.
8. Liu, K., Tong, M., Xie, S., & Zeng, Z. (2014, August). Fusing decision trees based on genetic programming for classification of microarray datasets. In *International Conference on Intelligent Computing* (pp. 126-134). Springer, Cham.
9. Dumitru, D. (2009). Prediction of recurrent events in breast cancer using the Naive Bayesian classification. *Annals of the University of Craiova-Mathematics and Computer Science Series*, 36(2), 92-96.

10. Chou, Y. H., Tiu, C. M., Hung, G. S., Wu, S. C., Chang, T. Y., & Chiang, H. K. (2001). Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis. *Ultrasound in medicine & biology*, 27(11), 1493-1498.
11. Sarhan, A. M. (2009). Cancer classification based on microarray gene expression data using DCT and ANN. *Journal of Theoretical & Applied Information Technology*, 6(2).
12. Steven K. Rogers, Dennis W. Ruck, Matthew Kabrisky, *Artificial Neural Networks for early detection and diagnosis of cancer*, Elsevier Scientific Publishers, pp. 79-83, 1994.
13. Md. BadrulAlamMiah, Mohammad Abu Yousuf, *Detection of Lung Cancer from CT Image Using Image Processing and Neural Network*, IEEE, In Proceedings of 2nd Int'l Conference on Electrical Engineering and Information & Communication Technology, 2015.
14. Alaa el-Halees (2009) Mining Students Data to Analyze e-Learning Behavior: A Case Study.
15. Bharadwaj B.K. and 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.
16. Gladence, L. Mary, M. Karthi, and V. Maria Anu. "A statistical comparison of logistic regression and different Bayes classification methods for machine learning." *ARN Journal of Engineering and Applied Sciences* 10, no. 14 (2015): 5947-5953.
17. Fadzilah S. and Abdoulha M. (2009), "Uncovering Hidden Information Within University's Student Enrollment Data Using Data Mining", In Proceedings of the Third Asia International Conference on Modelling & Simulation Conference, IEEE computer society.
18. L.MaryGladence,Karthi,M.Ravi, T "A novel technique for Multi-class ordinal regression-APDC" *IndianJournal of Science and Technology* Vol.9/ No.10/March 2016/1-5
19. Ramasubramanian, P. Iyakutti, K. and Thangavelu, P. (2009) "Enhanced 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.
20. Kulkarni R V, Patil S H, Subhashini R "An overview of learning in data streams with label scarcity, " *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016, 2017*
21. Sapphire Paul J, Brumancia E, and Monisha GS., *Comparison of Effort Estimation Techniques Using Decision Table with Regression Analysis.*, *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, Vol 7/ Issue 4/July-2016.