

# Forecast for Performance Improvement of Graduate Students utilizing Decision Trees Algorithms

Chirag Sarna  
K.K.W.I.E.E.R,  
University of Pune,  
Nasik, India

Rahul Sangale  
K.K.W.I.E.E.R,  
University of Pune,  
Nasik, India

Ravinder Kumar  
K.K.W.I.E.E.R,  
University of Pune,  
Nasik, India

Sushil Mahajan  
K.K.W.I.E.E.R,  
University of Pune,  
Nasik, India

**Abstract**— Understudy execution in college courses is of incredible concern to the advanced education where a few components may influence the execution. This paper is an endeavor to apply the data mining methods, especially arrangement, to help in upgrading the nature of the higher instructive framework by assessing understudy information to study the principle characteristics that may influence the understudy execution in courses. The arrangement standard era procedure is taking into account the Decision tree as an arrangement system where the created tenets are contemplated and assessed. A framework that encourages the utilization of the created guidelines is manufactured which permits understudies to foresee the last review in a course under study.

**Keywords**— Data Mining, Classification, Decision Trees, Student Data, Higher Education

## 1. INTRODUCTION

Data mining ideas and systems can be connected in different fields like showcasing, medication, land, client relationship administration, building, web mining and so on. Instructive data mining is another rising strategy of data mining that can be connected on the information identified with the field of training. There are expanding exploration engages in utilizing data mining in training.

This new rising field, called Instructive Data Mining. Instructive Data Mining uses numerous procedures, for example, Decision Trees, Neural Networks, Naïve Bayes, K-Nearest neighbour, and numerous others. Utilizing these procedures numerous sorts of information can be found, for example, affiliation standards, arrangements and grouping. The found learning can be utilized for expectation in regards to enlistment of understudies in a specific course, expectation about understudy's execution etc. In a University results general execution of an understudy is dictated by inward evaluation and outer exam. Inner evaluation is made on the bases of an understudy's task imprints, class test, lab work, participation past year evaluation and his/her inclusion in additional educational module exercises. While in the meantime outside appraisal of a understudy taking into account imprints scored in end of the year test. In this paper we make forecast about fall flat and pass proportion of understudies in light of last, most decisive test. Examination assumes an imperative part in any understudy's life. The imprints acquired by the understudy in the examination choose his future. Hence it gets to be fundamental to

anticipate whether the understudy will pass or fizzle in the examination. In the event that the expectation says that an understudy has a tendency to fall flat in the examination before the examination at that point additional endeavours can be taken to enhance his studies and help him to pass the examination

## 2. DECISION TREE

A Decision tree is a stream diagram like tree structure, where each inward hub is meant by rectangles, and leaf hubs are meant by ovals. All inside hubs have two or more kid hubs. All inner hubs contain parts, which test the estimation of an outflow of the traits. Circular segments from an inside hub to its kids are named with unique results of the test. Every leaf hub has a class name connected with it. Decision tree are ordinarily utilized for picking up data for the reason for choice -making. Decision tree begins with a root hub on which it is for clients to take activities. From this hub, clients part every hub recursively as indicated by Decision tree learning calculation. The last result is a Decision tree in which every limb speaks to a conceivable situation of choice and its result. The three broadly utilized Decision tree learning calculations are: ID3, J48 and CART.

### ID3

This is a Decision tree calculation presented in 1986 by Quinlan Ross [1]. It is in view of Hunts calculation. The tree is built in two stages. The two stages are tree building and pruning. ID3 uses information expansion measure to pick the part trademark. It simply recognizes absolute qualities in building a tree model. It doesn't give exact result when there is problem. To empty the commotion preprocessing technique must be used. To manufacture Decision tree, information expansion is figured for each in addition each quality and select the trademark with the most lifted information increment to dole out as a root center point. Name the trademark as a root center and the possible estimations of the quality are addressed as bends. By then all possible result cases are attempted to check whether they are falling under the same class or not. In case all the samples are falling under the same class, the center point is identified with single class name, for the most part pick the part credit to portray the events. Interminable qualities can be dealt with using the ID3 figuring by discretizing or direct, by considering the qualities to find the best part point by taking an edge on the attribute values. ID3 does not help pruning.

## CART

CART [1] stands for Classification and Regression Trees introduced by Breiman. It is also based on Hunt's algorithm. CART handles both categorical and continuous attributes to build a decision tree. It handles missing values. CART uses Gini Index as an attribute selection measure to build a Decision tree. Unlike ID3 and C4.5 algorithms, CART produces binary splits. Hence, it produces binary trees. Gini Index measure does not use probabilistic assumptions like ID3, J48. CART uses cost complexity pruning to remove the unreliable branches from the decision tree to improve the accuracy.

## J48

The J48 Decision tree [13] classifier takes after the accompanying basic calculation. With a specific end goal to characterize another thing, it first needs to make a decision tree in view of the characteristic estimations of the accessible preparing information. Along these lines, at whatever point it experiences an arrangement of things (preparing set) it recognizes the quality that segregates the different occasions generally unmistakably. This gimmick that has the capacity let us know most about the information cases so we can group them the best is said to have the most information gain. Presently, among the possible estimations of this feature, if there is any value for which there is no ambiguity, that is, for which the information occasions falling inside its class have the same value for the target variable, then we end that branch and relegate to it the target value that we have acquired.

For the other cases, we then look for another attribute that gives us the highest information gain. Hence we continue in this manner until we either get a clear decision of what combination of attributes gives us a particular target value, or we run out of attributes. In the event that we run out of attributes, or if we cannot get an unambiguous result from the available information, we assign this branch a target value that the majority of the items under this branch possess.

Now that we have the decision tree, we follow the order of attribute selection as we have obtained for the tree. By checking all the respective attributes and their values with those seen in the decision tree model, we can assign or predict the target value of this new instance.

## 3. RELATED WORK

Nguyen et al. [4] thought about the precision of Decision tree and Bayesian system calculations for foreseeing the scholastic execution of undergrad and postgraduate undergrads at two altogether different scholastic establishments. These forecasts are most valuable for recognizing and helping fizzling undergrads, and better focus grants. Therefore, the Decision tree classifier given better precision in correlation with the Bayesian system classifier. Al-Radaideh et al. [5] proposed to utilize data mining order methods to upgrade the nature of the higher instructive framework by assessing undergrads' information that may influence the undergrads' execution in courses. They utilized the CRISP structure for data mining to mine undergrads' connected scholastic information. An

arrangement model was fabricated utilizing the Decision tree system. They utilized three diverse order systems ID3, C4.5 and the NaïveBayes. The outcomes demonstrated that the Decision tree model would be wise to forecast precision than the different models. Accordingly, a framework was fabricated to encourage the use of the produced standards that undergrads need to foresee the last grade in the C++ college class. Cesar et al. [6] proposed the utilization of a suggestion framework in light of data mining methods to help undergrads to make choices identified with their scholastic track. The framework gave help for undergrads to better pick what number of and which courses to enlist on. Therefore, the creators added to a framework that is proficient to anticipate the disappointment or accomplishment of an undergrad in any course utilizing a classifier got from the examination of a set of recorded information identified with the scholarly field of different undergrads who took the same course previously. Muslihan et al. [7] have analyzed two data mining methods which are: Artificial Neural Network and the mix of bunching and Decision tree characterization systems for foreseeing and characterizing undergrad's scholastic execution. Undergrads' information were gathered from the information of the National Safeguard University of Malaysia (NDUM). Subsequently, the procedure that gives exact forecast and arrangement was picked as the best model. Utilizing the proposed model, the design that impacts the undergrad's scholastic execution was distinguished. Han and Kamber [8] depicts data mining programming that permit the clients to investigate information from diverse measurements, arrange it and abridge the connections which are distinguished amid the mining procedure. Bharadwaj and Pal [9] directed study on the undergrad execution based by selecting 300 undergrads from 5 diverse degree school directing BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By method for Bayesian grouping technique on 17 qualities, it was discovered that the elements like undergrads' evaluation in senior auxiliary exam, living area, medium of educating, mother's capability, undergrads other propensity, family yearly pay and undergrad's family status were exceedingly connected with the undergrad scholastic execution. Pandey and Pal [10] directed study on the undergrad execution based by selecting 600 undergrads from diverse schools of Dr. R. M. L. Awadh University, Faizabad, India. By method for Bayes Classification on classification, dialect and foundation capability, it was discovered that whether new comer undergrads will entertainer or not. Ramaswami and Bhaskaran [11] have developed a prescient model called CHAID with 7-class reaction variable by utilizing very affecting prescient variables acquired through peculiarity choice in order to assess the scholastic accomplishment of undergrads at higher optional schools in India. Information were gathered from distinctive schools of Tamil nadu, 772 undergrads' records were utilized for CHAID expectation model development. Therefore, set of guidelines were removed from the CHAID expectation model and the productivity was found. The exactness of the present model was contrasted and different models and it has been discovered to be tasteful. Shannaq et al. [12], connected the

characterization as data mining system to anticipate the quantities of enlisted understudies by assessing scholarly information from enlisted understudies to study the fundamental characteristics that may influence the understudies' faithfulness (number of selected understudies). The extricated order guidelines are based on the Decision tree as an arrangement technique, the extricated arrangement guidelines are contemplated and assessed utilizing distinctive assessment routines. It permits the University administration to get ready vital assets for the new enlisted understudies and demonstrates at an early stage which kind of understudies will possibly be selected and what zones to pack upon in advanced education frameworks for backing.

#### 4. CLASSIFICATION

This area depicts the building of the grouping model. By and large, information grouping is a two-stage process. In the first step, which is known as the learning step, a model that depicts a foreordained set of classes or ideas is assembled by examining a set of preparing database examples. Every case is expected to have a place with a predefined class. In the second step, the model is tried utilizing an alternate information set that is utilized to gauge the classification exactness of the model. On the off chance that the exactness of the model is viewed as satisfactory, the model can be utilized to characterize future information cases for which the class name is most certainly not known. Toward the end, the model goes about as a classifier in the choice making procedure. There are a few procedures that can be utilized for characterization, for example, Decision tree, Bayesian routines, tenet based calculations, and Neural Networks. Decision tree classifiers are truly mainstream strategies in light of the fact that the development of tree does not require any area master information or parameter setting, and is proper for exploratory information revelation. Decision tree can create a model with principles that are intelligible and interpretable. Decision tree has the favorable circumstances of simple translation and understanding for leaders to contrast and their space information for acceptance and supports their choice. Some of Decision tree classifiers are C4.5/ID3/CART, NBTree, what are more others.

#### 5. DATA COLLECTION

In this step just those fields were chosen which were needed for data mining. The information is gathered from the general understudies. Our target is to utilize the Examination information of the understudy. The information is put away in a database: MS Excel; however it can likewise be utilized with Oracle, Access, Interbase, any database supporting ODBC associations and others. We have utilized MS Excel expectations in light of the fact that is the world's most famous database. Since the data mining programming used to produce affiliation principles acknowledges information just in arff form, we have initially changed over the information on MS Excel expectations record into comma separated values format and afterward to arff group.

#### 6. CONCLUSION

One of the data mining strategies i.e., order is an intriguing subject to the scientists as it is precisely and productively orders the information for learning revelation. Decision trees are so famous on the grounds that they create order rules that are anything but difficult to translate than other arrangement systems. Much of the time utilized Decision tree classifiers are contemplated and the analyses are led to discover the best classifier for forecast of understudy's execution. Decision trees model is effectively distinguishing the understudies who are liable to fall flat. These understudies can be considered for legitimate advising in order to enhance their outcome. This finding is a preparatory research around there and we think it is a decent beginning stage for specialists in the locale to make an examination track identified with utilizing data mining to upgrade school/college training. This examination ought to be further upgraded as a future work by considering information from a few other college incorporating private college in different urban communities in India and gather more cases to manufacture the model. Other properties could likewise be added to the information set for further upgrading the created model. Moreover, some other grouping models could be tried in this space.

#### REFERENCES

1. J. R. Quinlan, "Introduction of decision tree", *Journal of Machine learning*, : pp. 81-106, 1986.
2. J. R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, Inc, 1992.
3. Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009.
4. Nguyen N., Paul J., and Peter H., A Comparative Analysis of Techniques for Predicting Academic Performance. In *Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference*. pp. 7-12, 2007.
5. Al-Radaideh Q., Al-Shawakfa E., and Al-Najjar M., Mining Student Data using Decision Trees, In *Proceedings of the International Arab Conference on Information Technology (ACIT'2006)*, Yarmouk University, Jordan, 2006
6. Cesar V., Javier B., liela S., and Alvaro O., Recommendation in Higher Education Using Data Mining Techniques, In *Proceedings of the Educational Data Mining Conference*, 2009.
7. Muslihah W., Yuhanim Y., Norshahriah W., Mohd Rizal M., Nor Fatimah A., and Hoo Y. S., Predicting NDUM Student's Academic Performance Using Data Mining Techniques, In *Proceedings of the Second International Conference on Computer and Electrical Engineering*, IEEE Computer society, 2009
8. J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.
9. B.K. Bharadwaj and S. Pal. "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, 2011.
10. U . K. Pandey, and S. Pal, “Data Mining: A prediction of performer or underperformer using classification”, (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.
  11. Ramaswami M., and Bhaskaran R., CHAID Based Performance Prediction Model in Educational Data Mining, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, 2010.
  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. J48:<http://www.d.umn.edu/~padhy005/Chapter5.html>