# **This is title of the article**

**Given-Name Surname1, Given-Name Surname 2, Given-Name Surname\*3**

1author's affiliation, Include Department, Institutional address, Country  
2Co-author's affiliation, Include Department, Institutional address, Country

\*3Co-author's affiliation, Include Department, Institutional address, Country

Corresponding author\* : email@aa.ac.kr

## Abstract

It should be <300 words. Mention your findings in the form of statements along with the conclusive data of statistical importance; Mention how your findings are unique and novel; how your findings are in consensus with the existing values/ reports or how different are they from the already reported findings. Highlight how your results are helpful in adding more value to the existing reports.

**Keywords:** 5-6 words, Normal, Drawn from title, Word representing the work.

## Introduction

Data mining helps to extract the original and the valuable data from the large amount of dataset. Data mining can be implemented in different areas such as Fraud detection, Medical, Education, Banking, Marketing and Telecommunications. Feature selection is a process to pick a group of features as subset that are identically suitable for investigation and for future predication by removing the unrelated or redundant features. The ultimate objective of feature selection process is to increase the predictive accuracy and reducing complexity of learner results. In the universities or in academic institutions, it’s very difficult to predict the frailer or dropout students in early stage[1][2].

## Literature review

**2.1 Review A**

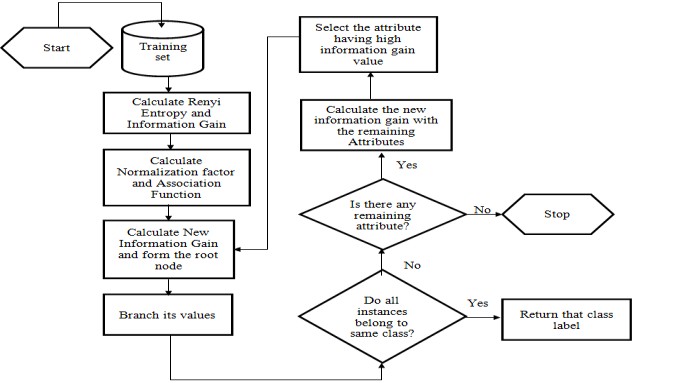
Retention in college courses.

**2.2 Review B**

Many studies included a wide range of potential predictors, including personality factors, intelligence and aptitude tests, academic achievement, previous college achievements, and demographic data and some of these factors seemed to be stronger than others, however there is no consistent agreement among different studies[3]-[5].

## Proposed Work

J.Ross Quinlan proposed the Iterative Dichotomized 3 (ID3 algorithm) in the year 1979 which is used to build the decision tree using information theory. Top down approach with no backtracking is used to build the model in the decision tree algorithm. Information gain is used to determine which attribute will best decide the target data classification. The traditional ID3 algorithm is improved by using Renyi entropy, Information gain and Association Function in this work. This combination is used as a new criterion to construct the decision tree and to predict the dropout of the university students. Initially Renyi entropy is determined using which the Information gain is calculated. This value is kept as the old gain for every attribute[6].



**Figure 1. Design of Improved Decision tree algorithm for Educational Data mining**

Calculate Renyientropy using the formula

, α≥0 and α ≠1.

Here X is a discrete random variable with possible outcomes 1, 2…n. α is the order and when it equals to 1 it is Shannon entropy. A completely homogeneous sample has the entropy of 0.Equally divided sample has the entropy of 1.

**Table 1: Initial Set of features used for the experimentation**

|  |  |
| --- | --- |
| Residence  Family Type  Family Annual Income  Father’s Education  Mother’s Education  Father’s occupation  Mother’s occupation  College Location of student  Student grade/percentage in High School (10th )  Student grade/percentage in Senior Secondary (10th )  Course Admitted  Admission type  Satisfaction with Course | Like this University  Educational system of University  Infrastructure of university  Extra-curriculum activities in university  Entertainment in university  Time for self study  Placement Status  Participate in extra curriculum activity  Teacher Student relationship  Family Problem  Home Sickness  Campus Environment  Change of Goal  Adjustment Problem  Enrolled in other universities |

## Conclusion

This paper proposed an improved decision tree algorithm for prediction of dropout student. The objective of this work is to develop an improved decision algorithm that enhances the ability to form decision trees and thereby to prove that the classification accuracy of improved decision algorithm on educational dataset is greater. A new decision tree model is to be constructed by using Renyi entropy for calculating the information gain and the association function will be used which determines the relative degree between the given attribute and class C. Experimental results will prove that improved decision tree algorithm will provide better prediction accuracy on student dropout data than that of traditional classification algorithms.

## Acknowledge

This research was supported by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea(NRF).

## References

[1] Kimio T, Natarajan G, Hideki A, Taichi K, Nanao K, “Higher involvement of subtelomere regions for chromosome rearrangements in leukemia and lymphoma and in irradiated leukemic cell line”, Indian Journal of Science and Technology, 2012 April, 5(1), pp. 1801-11.

[2] Cunningham C H, A laboratory guide in virology, 6th edn. Burgess Publication Company, 1973.

[3] Sathishkumar E, Varatharajan M, “Microbiology of Indian desert. In: Ecology and vegetation of Indian desert. D.N.Sen (ed.), Agro Botanical Publ.: India. 1990, pp. 83-105.

[4] Varatharajan M, Rao B S, Anjaria K B, Unny V K P, Thyagarajan S, Radiotoxicity of sulfur-35, Proceedings of 10th NSRP, India, 1993, pp. 257-58.