

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts*

**Jyoti Namdeo<sup>1</sup>**Department of Computer Applications  
Bharati Vidyapeeth University, IMED  
Pune – India**Naveenkumar Jayakumar<sup>2</sup>**Department of Computer Applications  
Bharati Vidyapeeth University, College of Engineering  
Pune – India

*Abstract: Data being generated in the academic domain and educational perspective are increasing in an exponential rate. There exists many data some are relevant and some are irrelevant. The knowledge extraction from these data will yield wanted and unwanted information. The challenging task is to extract wanted and relevant information and knowledge from these huge set of data. Over the time of research in the field of data mining, rapid growth is seen in Rough set theory and its applications. This paper discusses the basic approach and concepts of the rough set theory in the field of academic domain for the performance prediction of students in course works.*

*Keywords: Information Systems, knowledge discovery, Rough set theory, redundant attributes, Decision Table, Reducts, Predictions.*

### I. INTRODUCTION

As improvements in data mining is persistent and rapid usage of rough set theory is spreading across multiple disciplines, predicting the students' performance in academics can also be done through generating the reducts and core. In the process of prediction the rough set theory is implemented as the heart of a knowledge discovery process. Therefore the students are classified in two categories one is pass and other is fail.

Data mining by traditional methods has some limitations like it cannot properly handle the missing values, requires detailed information about the data, and cannot deal with uncertainty or vagueness in any information domain. However, these limitations can be overcome by methods based on Rough Set Theory (Pawlak, 1991). Knowledge discovery is the process of defining and extracting the inexplicit, previously unknown and possible useful information from data.

Rough set has applications in artificial intelligence, approximate reasoning, machine learning, pattern recognition, knowledge discovery, data mining, and expert system. Applications of rough set theory in different domains has been reported however it has not been much explored on educational domain so we planned to use rough set in educational data mining.

Due to rapid advancement in the field of information technology, the amount of data storage is increasing exponentially in all the fields like medical, retail, banking, agriculture, spatial etc. It has also touched the field of Education. These huge databases of education contain a wealth of data and constitute a potential goldmine of valuable information. Data mining in educational data offers comprehensive characteristics analysis of students.

The main emphasis in higher education is to improve the result of students, uplifting the quality of education imparted; taking out the precautionary measures for those students whose current result is not satisfactory. Higher institutions have to meet out the needs of students through the academic staff, supporting staff and management.

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data [Witten and Frank 1999]. Though the data mining has got sufficient attention from industries warehouses, banking, medicals, retails, fraud detection etc., in dealing with educational database it is still in its

beginning. [www.educationaldatamining.org](http://www.educationaldatamining.org), defines educational data mining as: “Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.” Baker [2008] classified work in educational data mining as follows:

1. Prediction
  - a. Classification
  - b. Regression
  - c. Density estimation
2. Clustering
3. Relationship mining
  - a. Association rule mining
  - b. Correlation mining
  - c. Sequential pattern mining
  - d. Causal data mining
  - e. Distillation of data for human judgment Discovery with models

## II. RELATED WORK

Vialardi et. al. (2009) developed recommendation system based on data mining techniques to help students to take decisions on their academic itineraries, to choose course, based on experience of previous students with similar academic achievements [3]. However Sahay and Mehta (2010) developed a software system to assist higher education in assessing and predicting key issues related to student success. Software used data mining algorithm and quality tools such as quality function deployment to study and predict issues related to enrollment management, dropout rate, and time to degree and suggest ways to improve courses and programs[4]. Ayesha et. al. (2010) studied K-means clustering algorithms to discover knowledge from education data mining. They recommended that all correlated information of class quiz, mid & final exam should be conveyed so that dropout ratio can be reduced and student performance can be improved [6]. Quadril and Kalyankar (2010) used data mining in predicting drop out feature of students. They used decision tree techniques to choose the best prediction and analysis for direct or indirect intervention from teacher and management [7].

## III. APPROACH

Namdeo V. et. al. (2010) collected student’s data from engineering & applied four different classification methods & classifies students based on their final grades and applied four classification methods on student data i.e. Decision tree (ID3), Multilayer Perceptron, Decision Table and Naïve Bayes Network Classification method. They concluded that ID3 Classifier is most suitable method for this type of student dataset [8]. The same result was also proved by Cristina Oprea and Marian Zaharia (2011) whose results indicated that the ID3 algorithm and Random Forest algorithm provided the highest accuracy and correctly classified instances [9]. Mohamad Farhan et.al. (2010) focused the relationship between academic factor and personality characteristic towards programming performance [10]. Ramasubramanian et. al. (2009) proposed a concept map for each student and staff. This map finds the result of the subjects and also recommends a sequence of remedial teaching. Narli (2010) [15] gave detailed analysis of quantitative and categorical data using rough set theory and concluded that rough set approach can be applied to analyze educational research data for the investigation of attitudes, behaviors or beliefs could reveal more comprehensive information about the data[14]. Narli and Ozelik (2010) reported rough sets to discover meaningful information

from large data sets. They have incorporated rough set theory in topology instruction for qualitative data analysis and concluded that students' incorrect or partially correct ideas affect their learning [16]. Ahmad et. al. (2010) used ROSETTA tool for classification, proposed an integrated Felder Silverman learning style model by analyzing student's preferences while using e-learning system developed using Moodle[17]. JansiRani and Bhaskaran (2010) developed algorithm based on RST to implement in the data collected from various college students to identify the dominant attributes such as constant care, affection, motivation, parent – student relationship, no comparison, not so strict, frequent tests, friendly behavior[18]. Edin Osmanbegović \*, Mirza Suljić \*\*(2012) used three supervised classification algorithm NB, J48 and MLP on educational data. They concluded that Naïve Bayes classifier outperforms in prediction decision tree and neural network methods.

Data mining by traditional methods has some limitations like it cannot properly handle the missing values, requires detailed information about the data, and cannot deal with uncertainty or vagueness in any information domain. However, these limitations can be overcome by methods based on Rough Set Theory (Pawlak, 1991).

In the present paper rough set theory has been used for finding the important attributes called as reduct.

#### IV. DATA COLLECTION

Masters in Computer Applications (MCA) is a professional program of three years having six semesters. Each semester has 7 subjects and thus the total number of subjects including compulsory and elective is about 40. Secondary data is collected from university exam section.

Preliminary data of 53 students was taken from the batch of 2007. Final marks to students were given out of 100. The 100 marks were divided into ratio of 80:20 in which, 80 marks were allotted for theory paper whereas 20 marks were kept for internal assessment.

Initially more than 20 attributes have been collected. Out of these, 12 conditional attributes and one class attribute have been chosen because teaching has to be centred on these main subjects and have to be taught theoretically. The attributes along with their descriptions and possible values are presented in Table 1.

Table 1: Description of Attributes

Attribute Code	Description
101	Elementary Algorithm
103	Procedure Oriented Programming
201	Data Structure
202	Operating System
203	Data Base Management System
301	Software Engineering
302	Computer Networking
303	Object Oriented Programming
401	UML
402	Unix
501	Software Project Management
502	Artificial Intelligence

#### V. DATA PREPROCESSING

The data is stored in excel sheet. No backlog student data is taken as we want to concentrate on students who had cleared the exam in one stroke only. In this study we mainly concentrated on the theoretical exams conducted by the university and not on the internal assessment as we are interested only in the study of performance of students in the theoretical exam. So we removed the internal marks of 20 from the data collected.

### A. Discretization

The process of partitioning continuous variables into categories is usually termed as discretization. Discrete values have important roles in data mining and knowledge discovery. They are about intervals of numbers which are more concise to represent and specify, easier to use and comprehend as they are closer to a knowledge-level representation than continuous values. Discretization of real value attributes (features) is an important pre-processing task in data mining for classification problems (Chmielewski and Grzymala-Busse, 1994; Dougherty et al., 1995; Nguyen and Skowron, 1995; Nguyen, 1998; Liu et al., 2002). Dougherty et al. (1995) reported that discretization makes learning faster.

In this research a 10-point grading system is used to discretise the range of marks into groups as shown in the following table 2

Table 2: Marks Grade point mapping

MARKS	GRADE POINT
[75-100]	10
[70-74.9]	9
[65-69.9]	8
[60-64.9]	6
[55-59.9]	7
[50-54.9]	5.5
[45-49.9]	5.0
[40-44.9]	4.5
[00-39.9]	0

This 10-point scale will be easy to handle the selected data as this will discretise the continuous range of 0-100 marks into 9 discrete intervals.

The result of the student depends on the conditional attributes. The final result of student is declared by evaluating the average marks scored at the completion of the program. These marks contain only the total of theoretical marks out of 80.

The result constitutes the decision attribute. Grades from Fl (fail) to D (distinction) are used which discretise the result in the manner as shown below in table. Discretization of Decision Attribute (Final results)

Table 3: Marks Grade Mapping

MARKS	GRADES	EXPLANATION
[70-100]	O	D, Distinction
[60-69.99]	A+	F, First class
[55-59.99]	B+	HS, Higher Second class
[50-54.99]	B	S, Second class
[40-49.99]	C	P, Pass class
[00-39.99]	F	Fl, Fail class

Using the above criteria, all the subject marks along with the result are discretised.

### B. Feature Selection

Since the collected attributes may have some irrelevant attributes that may degrade the performance of the classification model, a feature selection approach is used to select the most appropriate set of features. Rough set theory (RST) can be used as such a tool to discover data dependencies and to reduce the number of attributes contained in a data set using the data alone and no additional information.

For this purpose the Rough set based Rosetta toolkit is used and the Johnson Reducer algorithm has been applied which gave one reduct. This is a simple greedy heuristic algorithm that is often applied to discernibility functions to find a single reduct (Øhrn, 1999). Reducts found by this process are generally of a size close to the minimal.

Applying this algorithm on the pre-processed data gave the set of most important attributes as 101, 201 and 303.

### C. Classification

The ability to predict/classify a student's performance is very important in educational institutions. For the automatic classification of instances model is generated with the help of classification algorithm. Classification algorithms used in the study are J48, MLP, Naïve Bayes and Random Forest. All these fall under the category of supervised algorithm. This technique uses training the samples of the data for learning. The model thus obtained usually achieve very high accuracy on the training samples, however on test data, it may perform poorly. Therefore validation method is used to test its generalization accuracy. Generally leave-one-out is used as a validation method in which all the data records except one are used to train the classification model. Left out one data record in turn is used as the test data. The trained model is then applied to the left out data. In the present study, we used 10 fold cross validation method in which training data is divided into 10 equal parts, 9 parts are used for training and 1 part is used for testing. In general, there will be some correctly and wrongly classified subjects after 10 runs. The average error rates are then calculated on 10 runs.

### D. Decision Tree

Decision trees automatically generate rules, which are conditional statements that reveal the logic used to build the tree. A decision tree is a set of conditions organized in a hierarchical structure [25]. It is a predictive model in which an instance is classified by following the path of satisfied conditions from the root of the tree until reaching a leaf, which will correspond to a class label. A decision tree can easily be converted to a set of classification rules. Some of the most well-known decision tree algorithms are C4.5 [25] and CART [4]. In weka J48 is the implementation of C4.5

### E. Naïve Bayes

Naive Bayes: Naive Bayes uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data.

### F. MLP

A neural network is a parallel distributed processing network used for rule induction and modelled over neurons in the brain. It consists of interconnected processing elements called nodes or neurons that work together to produce an output function. Examples of neural network algorithms are multilayer perceptron (with conjugate gradient-based training) [22]. As the name suggests MLP network consists of a set of sensory elements that make up the input layer, one or more hidden layers of processing elements, and the output layer of the processing elements (Witten and Frank).

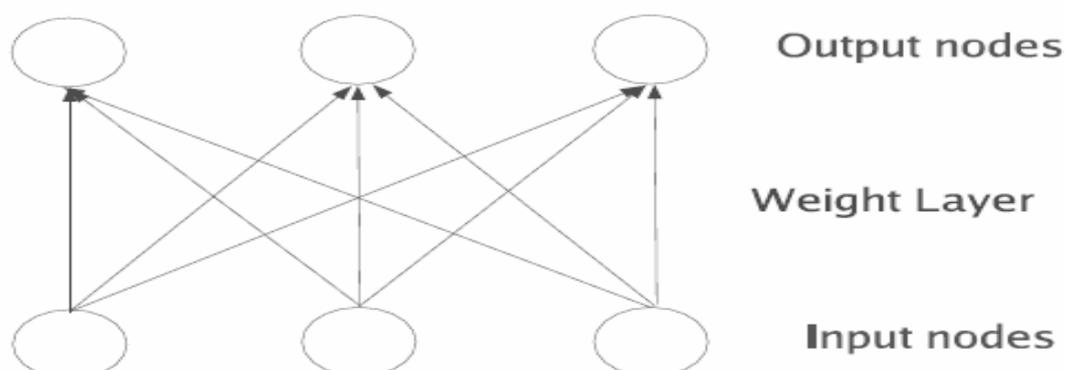


Figure 1: Neural network Layer

### G. Random Forest

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Leo Breiman, 2001). Using a random selection of features to split each node yields error rates that compare favorably to Adaboost (Freund and Schapire[1996]), but are more robust with respect to noise.

## VI. RESULT

All these algorithms were trained on the same data of 2007 batch having 51 students. A software tool Weka is used for this purpose. After training these algorithm were tested on 2008 batch. A given table is showing the comparison of classification accuracy of these algorithms on the training as well as test data.

Reducts are 101,201,303, and classification accuracy is as follows

Table 4: Reducts accuracy on various algorithms

ALGORITHM	TRAINING	CROSS VALIDATION	PERCENTAGE SPLIT	TEST ON 2008 DATA
Naïve bayes	84.31%	66.66%	38.29%	31.57%
Multilayer perceptron	96.07%	58.82%	58.82%	26.31%
J48	78.43%	56.86%	58.82%	17.54%
Random forest	100%	56.86%	58.82%	24.56%

From the table we can see that classification accuracy of these algorithms is very high for training, however it gets reduced for the cross validation. It further reduces when checked on the test data. Out of these algorithms Naïve Bayes is giving good classification accuracy.

## VII. CONCLUSION

In educational data of college students, final result classifies the student. On the basis of previous student result we can predict the future student result using J48, NB, MLP and random forest; however classification accuracy is not very high. Although NB is giving more correct classification accuracy compared to rest three.

## References

1. Ceglar, J.F Roddick. "Association mining". ACM Computing Surveys, 38:2, pp. 1-42, 2006.
2. Chmielewski and Grzymala-Busse Global discretization of continuous attributes as pre-processing for machine learning. In Third International Workshop on Rough Sets and Soft Computing 1994, pp. 294–301.
3. Dougherty et al., Supervised and unsupervised discretization of continuous features. In Proc. Twelfth International Conference on Machine Learning. Los Altos, CA: Morgan Kaufmann, 1995 pp. 194– 202.
4. Han, J., Hu, X., Lin, T. Y. (2004). Feature Subset Selection Based on Relative Dependency between Attributes. Rough Sets and Current Trends in Computing: 4th International Conference, RSCTC 2004, Uppsala, Sweden, June 1-5, pp. 176–185.
5. Jensen R. and Shen Q (2005). Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches, IEEE Transactions on Knowledge And Data Engineering, Vol. 17, No. 1.
6. Kotsiantis S. and Kanellopoulos. D. "Association Rules Mining" A Recent Overview. GESTS Int. Transactions on Computer Science and Engineering, Vol. 32 (1), pp. 71-82, 2006.
7. Liu H. et al., Discretization: An Enabling Technique, Data Mining and Knowledge Discovery, 6, 393–423, 2002. Kluwer Academic Publishers, the Netherlands.
8. Mehmed Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms 2003 John Wiley & Sons publishers.
9. Nguyen H. S. and Skowron, Quantization of real value attributes: rough set and Boolean reasoning approach. Proceedings of the Second Joint Annual Conference on Information Sciences, pp. 34-37, Wrightsville Beach, NC, September 1995.

10. Nguyen, H.S. Discretization problem for rough sets methods. Proceedings of the First International Conference on Rough Sets and Current Trends in Computing, pp. 545-552, Warsaw, Poland, June 1998, Springer-Verlag.
11. Pawlak, Z. (1991). Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishing, Dordrecht.
12. Skowron, A., and Rauszer, C. (1992). The discernibility matrices and functions in information systems. Intelligent Decision Support, Kluwer Academic Publishers, Dordrecht, pp. 331–362.
13. Starzyk, J. A.et.al. (2000). A Mathematical Foundation for Improved Reduct Generation in Information Systems. Journal of Knowledge and Information Systems, Vol. 2, No. 2, pp.131-146.
14. Chaobo H. and Qimai C. (2011): Rough set analysis model for correlation between courses and its application. Computer Engineering and Applications, 47(27), 233-235