

# Superiority of Rotation Forest Machine Learning Algorithm in Prediction of Students' Performance

Manish Kumar

Researcher, Banaras Hindu University  
Department of Computer Science  
Institute of Science, Varanasi-221005

## ABSTRACT

Now-a-days the volume of educational data stored in educational database is increasing rapidly and these databases comprise hidden information for improvement of students' performance. Therefore we need computational methods to study the data available in the educational field and bring out the hidden knowledge from it. In this study, the experiments were conducted for the prediction task of educational data obtained from UCI Machine Learning repository using the five machine learning algorithms. The feature selected is used for training and testing of each classifier individually with ten-fold cross validation. The results obtained show that the ROF classifier outperforms other classifiers in terms of Area under the ROC curve (AUC), accuracy and MCC respectively.

## General Terms

Datasets, Demographics features, School related features, Features during study, UCI Machine Learning repository, Predicting protocol, Rotation Forest algorithm, Principle Component Analysis (PCA), Feature Extraction, Base classifier, WEKA, Sensitivity, Specificity, Accuracy, MCC, AUC

## Keywords

Accuracy, Machine learning, Rotation forest, students' performance, UCI Machine Learning repository

## 1. INTRODUCTION

Education is a significant element for accomplishing a long-term economic progress. The precise prediction of student performance could have numerous valuable applications for progressive outcomes of the learning process and central to improvements in learning principle. This prediction could function as a primary cautioning of future low academic performance and guide interventions that could prove helpful for such students. Similarly, being able to recognize the importance of different prevailing variables that influence performance for all and for each category of performance level, would be a significant influence to improve the methodology of teaching and better comprehend the learning processes. An early prediction of the students' academic performance is a key for various reasons: to obtain an idea of the students' level of learning, to get a grip of information on the methodology of teaching, to take decision on success or failure of students in the course registered, to enlighten low performance students to put in additional effort in the course to overwhelmed their weaknesses. Increasing students' withholding or persistence is a long term goal of all academic institutes. Typically, the first-year students are at greatest risk of dropping out from the study. Students' early experience in course satisfaction plays an important role in keeping a high retention rate. One of the intrinsic goals of the model is to retain students in the course by finding the effect of the family

activities, work activities and classroom environment on their performance.

Alternatively, the attention in Business Intelligence (BI)/Data Mining (DM) [1], arose due to the developments of Information Technology, prominent to an exponential evolution of commercial and organizational databases. All this data holds significant information, such as fashions and patterns, which can be important to improve decision making and promote success. Also, the volume of educational data stored in educational database increasing rapidly. Until now, human professionals and experts are limited in finding hidden pattern from data. Hence, the alternative is to use computational methods to investigate the raw data and mine exciting information for the decision-maker.

**Behrouz Minaei-Bidgoli et al.** [2] designed, implement, and evaluate many pattern classifiers and compare their performance on an online course dataset and also used combination of multiple classifiers that leads to a substantial enhancement in classification performance. **Romero C, S. Ventura** [3] surveys the application of data mining to traditional educational systems, particular web-based courses, well-known learning content management systems, and adaptive and intelligent web-based educational systems. **Hsia et al.** [4] used machine learning methodology to investigate the course preferences and course completion rates of enrolees in extension education courses at a university in Taiwan by applying three data mining algorithms: Decision Tree, Link Analysis, and Decision Forest. Decision tree was used to find enrolee course preferences, Link Analysis found the correlation between course category and enrolee profession, and Decision Forest found the probability of enrolees completing preferred courses. **El-Alfy and Abdel-Aal** [5] suggested a methodology that uses abductive network modelling to automatically recognise the most-informative subset of test items. They trained the predictor with a training dataset of 1,500 cases (examinees) and 45 test items. The proposed approach automatically selected only 12 items which classified an evaluation population of 500 cases with 91% accuracy. **Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás** [6] compared different data mining methods and techniques for classifying students based on their Moodle usage data and the final marks obtained in their respective courses. **V.O. Oladokun, A.T. Adebajo and O.E. Charles-Owaba** [7] developed a model based on the Multilayer Perceptron and trained the model on the dataset of five generations of graduates from an Engineering Department of University of Ibadan, Nigeria's first University. **Lykourentzou et al.** [8] proposed a dropout prediction method for e-learning courses using three widespread machine learning techniques feed-forward neural networks, support vector machines and probabilistic ensemble simplified fuzzy ARTMAP and detailed student data. **Jantan et al.** [9] try to discover the classification approach for

academic talent forecasting in higher education institutions. **Muslihah Wook et. al.** [10] compared two classification techniques the Artificial Neural Network (ANN) and the combination of clustering and decision tree for predicting and classifying students' academic performance. **M. M. Quadri and N.V. Kalyankar** [11] applies decision tree technique of data mining in predicting the drop out of students. **Sajadin Sembiring et. al.** [12] used smooth Support Vector Machine (SSVM) classification and kernel k-means clustering techniques to study the student academic performance. **Mohammed M. Abu Tair and Alaa M. El-Halees** [13] applied data mining procedures to determine association, classification, clustering and outlier detection rules to improve graduate students' performance. **Indriana Hidayah, Adhitya Erna Permanasari and Ning Ratwastuti** [14] developed classification model by applying neuro fuzzy concept; a combination of fuzzy's IF-THEN rules and neural networks to predict the students' academic performance. **Pauziah Mohd Arsad, Norlida Buniyamin and Jamalul-lail Ab Mayan** [15] used Artificial Neural Network (ANN) model in predicting academic performance of engineering students. **D. Magdalene Delighta Angeline** [16] used Apriori algorithm that is extracting the set of rules, and analyses the known data to classify the student based on their performance in academics using assignment, internal assessment tests, attendance etc., which benefits to predict the performance of the student generated on the pattern taken out from the educational database. **Cristobal Romero, Manuel-Ignacio Lopez, Jose-Maria Luna and Sebastian Ventura** [17] recommends the usage of different data mining methods for refining prediction of students' final performance beginning from participation indicators in both quantitative, qualitative and social network forums. **Tjioe Marvin Christian and Mewati Ayub** [18] used NBTree algorithm and students' personal data, education data, admission data, and academic data to predict the performance of students. **M. Mayilvaganan and D. Kalpanadevi** [19] compared the performance of C4.5 algorithm, AODE, Naïve Bayesian classifier algorithm and Multi Label K-Nearest Neighbour algorithm to discover the suitable accuracy of classification algorithm and decision tree algorithm to analyse the students' performance. **Elakia,, Gayathri, Aarthi and Naren J** [20] considered to validate that numerous data mining techniques which comprises classification, can be used in educational databases to recommend career choices for the high school students and also to predict the possible intense behavior among the students by including additional parameters other than academic details. **Geraldine Gray, Colm McGuinness and Philip Owende** [21] considered application of six classification models to recognise college students at risk of failing in the first year of study by using data collected from three student associates in the academic years 2010 through 2012 **Raheela Asif used et al.** [22] used socio-economic or demographic features and pre-university marks for prediction of the graduation performance in 4th year at university. He used decision tree, 1-nn, neural networks and naïve Bayes as a classifying protocol, but naïve Bayes outperformed with accuracy of 83.65%. **Glyn Hughes and Chelsea Dobbins** [23] attempt to predict learner performance in massive open online courses (MOCCs) using data mining techniques.

So, due to the large volume of data in educational databases and challenge of prediction of students' performance computational method is essential. In this study I tried to find the machine learning algorithm that outperform the existing ones for the domain considered.

## 2. MATERIAL AND METHODS

### 2.1 Dataset

For study, the dataset has been downloaded [Cortez, 2008] from the UCI Machine Learning Repository. In this dataset, the data was assimilated into two datasets associated to Mathematics having 395 instances and the Portuguese language having 649 instances. The features selected of the dataset is listed in table 1. The set of the attributes can be divided into three groups: Demographic Features, School related Features and Features during study. The data was collected during the 2005- 2006 school year from two public schools, from the Alentejo region of Portugal.

**Table 1: Selected Features of students used in experiment**

Demographic Features	
Sex	Student's sex (binary: female or male)
age	Student's age (numeric: from 15 to 22)
address	Student's home address type (binary: urban or rural)
Pstatus	Parent's cohabitation status (binary: living together or apart)
Medu	Mother's education (numeric: from 0 to 4 <sup>a</sup> )
Mjob	Mother's job (nominal <sup>b</sup> )
Fedu	Father's education (numeric: from 0 to 4a)
Fjob	Father's job (nominalb)
guardian	Student's guardian (nominal: mother, father or other)
famsize	Family size (binary: ≤ 3 or > 3)
famrel	Quality of family relationships (numeric: from 1 – very bad to 5 – excellent)
School related Features	
School	Student's school (binary: Gabriel Pereira or Mousinho da Silveira)
reason	reason to choose this school (nominal: close to home, school reputation, course preference or other)
traveltime	home to school travel time (numeric: 1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour).
Features during study	
studytime	Weekly study time (numeric: 1 – < 2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – > 10 hours)
failures	Number of past class failures (numeric: n if 1 ≤ n < 3, else 4)
schoolsup	Extra educational school support (binary: yes or no)
famsup	Family educational support (binary: yes or no)
activities	Extra-curricular activities (binary: yes or no)
paidclass	Extra paid classes (binary: yes or no)
internet	Internet access at home (binary: yes or no)
nursery	Attended nursery school (binary: yes or no)
higher	Wants to take higher education (binary: yes or no)
romantic	With a romantic relationship (binary: yes or no)
freetime	Free time after school (numeric: from 1 – very low to 5 – very high)
goout	Going out with friends (numeric: from 1 – very low to 5 – very high)
Walc	Weekend alcohol consumption (numeric: from 1 – very low to 5 – very high)
Dalc	Workday alcohol consumption (numeric: from 1 – very low to 5 – very high)
health	Current health status (numeric: from 1 – very

	bad to 5 – very good)
absences	Number of school absences (numeric: from 0 to 93)
G1	First period grade (numeric: from 0 to 20)
G2	Second period grade (numeric: from 0 to 20)
G3	Final grade (numeric: from 0 to 20)
Result	Binary: pass or fail

## 2.2 Predicting Protocol

We experimented with five better performing machine learning algorithms and giving below a brief description about relatively the best performing classifier Rotation Forest. The Rotation Forest classifier chooses decision tree as base classifier and Principle Component Analysis (PCA) as the feature extraction method. The main factor for the outperformance of the rotation forest is that the transformation matrix used to evaluate the extracted features.

**Rotation Forest Algorithm:** Let  $X$  be the training set kernel vectors and  $Y$  be the class labels of the corresponding input kernel vector and  $F$  be the feature set. Assume that there are  $N$  numbers of training instances with  $n$  features, and then  $X$  will become an  $N$  by  $n$  matrix. Let  $Y$  take values from set of class labels  $\{\omega_1, \dots, \omega_j\}$  denoted by  $\omega$ . The feature set of dataset is assumed to be partitioned into  $K$  subsets and the decision trees numbers of Rotation Forest algorithm is to be  $L$  with notation of  $\{D_1, \dots, D_L\}$ . The data used in training of base classifier is created with a randomly split  $K$  feature set [24, 25].

The training set for classifier  $D_i$  is handled in three steps:

- (i) As a first step,  $F$  is divided into  $K$  feature sets randomly with each subset of  $M = n/K$  number of features.
- (ii) In this second step, let  $F_{ij}$  denote the  $j^{\text{th}}$  subset of features to train classifier  $D_i$  and  $X_{ij}$  be the set of data for  $F_{ij}$ , being subset of features. A nonempty random subset is drawn from  $X_{ij}$  and then with bootstrap to form a new training set the 75% of this training data is selected as  $X'$ . A linear transformation is operated on  $X'$  to generate the coefficients of matrix  $C_{ij}$ . Each matrix  $X'$  has size of  $M \times 1$  and the coefficients of this matrix are  $a_{ij}^{(1)} \dots a_{ij}^{(M_j)}$
- (iii) In this last step, having obtained the coefficients of matrix  $C_{ij}$ , a sparse rotation matrix  $R_i$  then formed as given in (4)

$$R_i = \begin{bmatrix} a_{i1}^{(1)} \dots a_{i1}^{(M_1)} & [0] & \dots & [0] \\ [0] & a_{i2}^{(1)} \dots a_{i2}^{(M_2)} & \dots & [0] \\ \dots & \dots & \dots & \dots \\ [0] & [0] & \dots & a_{iK}^{(1)} \dots a_{iK}^{(M_K)} \end{bmatrix}$$

Here at this point, the columns of  $R_i$  are rearranged with respect to the original feature set and the new rotation matrix is represented as  $R_i^a$ . The transformed training set for classifier  $D_i$  will become  $XR_i^a$ . By means of this approach, the classifiers are provided with parallel training.

While the classification phase is evaluated for a given instance  $x$ , let the probability of this instance being classified by classifier  $D_i$  to one of classes is denoted with  $d_{ij}(xR_i^a)$ . From this point, the confidence of a class is calculated by Eq.

(1), and  $x$  is assigned to a class with the largest confidence calculated.

$$\mu_j(x) = \frac{1}{L} \sum_{i=1}^L d_{ij}(xR_i^a), j = 1 \dots c \quad (1)$$

The Rotation Forest algorithm applies Principal Component Analysis (PCA) transformation to each  $K$  subset to determine principal components that are expected to preserve the variability of information in the data. By means of  $K$  axis rotations, the new features for base classifier are formed. The Rotation approach in this method serves the ensemble with accuracy and diversity. In traditional the Rotation Forest algorithm, decision trees are chosen for rotation task, because of their sensitivity to rotation of the feature axes. And hence the name 'forest' is inspired from this scheme.

## 3. RESULTS AND DISCUSSION

For testing our proposed method the experiments were conducted for classifying the students for the label pass or fail by separately applying five classifiers machine learning algorithms namely: Rotation forest (ROF), NaiveBayes, Multilayer Perceptron (MLPC), Radial Basis Function (RBFClassifier), Sequential Minimum Optimisation (SMO), using Weka 3.7.12 [26]. The classification performances of the classifiers were analysed with respect to the standard performance parameters, namely: Accuracy, Specificity, Sensitivity, Precision, Receiver Operating Characteristic (ROC) Area [27], Matthew's Correlation Coefficient (MCC) besides time taken for training (learning). The formula for calculating these parameters are given below:

$$\text{Sensitivity} = \frac{tp}{tp + fn} * 100$$

$$\text{Specificity} = \frac{tn}{tn + fp} * 100$$

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + tn + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{MCC} = \frac{(tp * tn) - (fp * fn)}{\sqrt{(tp + fn) * (tn + fp) * (tp + fp) * (tn + fn)}}$$

Where

$tp$  is the number of true positives,

$tn$  is the number of true negatives,

$fp$  is the number of false positives and

$fn$  is the number of false negatives.

The table 2 shows the values of Sensitivity, Specificity, Accuracy, Precision, MCC, AUC performance metrics besides their training time for all the five classifiers separately for the chosen dataset.

**Table 2: Performance of five classifiers for the Mathematics subject dataset**

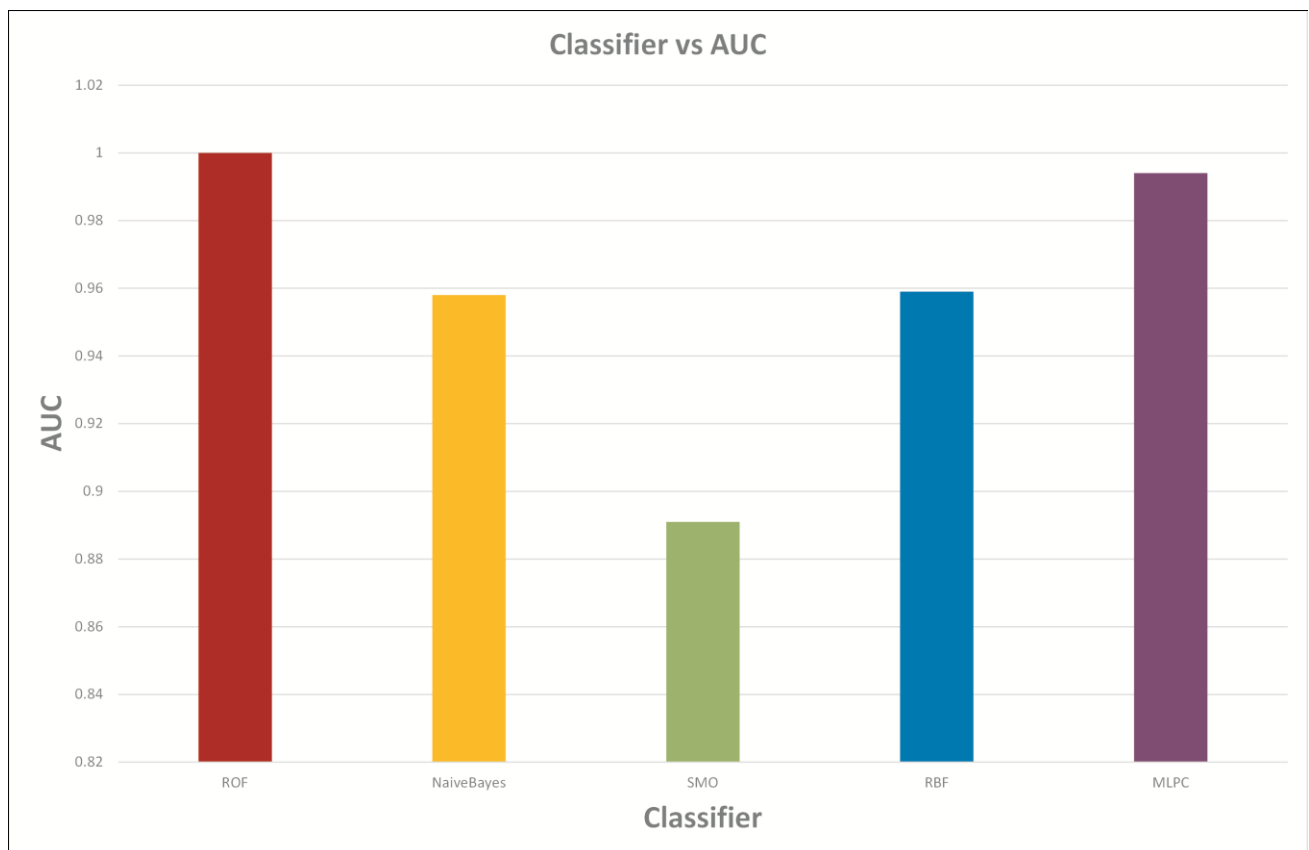
Classifiers	Sensitivity	Specificity	Accuracy	Precision	MCC	AUC	Training Time(in secs)
ROF	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Naive Bayes	0.92	0.86	0.90	0.90	0.78	0.96	1.00
SMO	0.97	0.81	0.92	0.92	0.82	0.89	0.06
RBF	0.96	0.74	0.89	0.89	0.74	0.96	0.81
MLPC	0.97	0.94	0.96	0.96	0.91	0.95	1.16

The sensitivity indicates the ability of the classifier to identify positive instances correctly, the specificity indicates the ability of the classifier to identify negative instances correctly and accuracy indicates the percentage of correct classification of both positive class as well as negative class instances. The ROF performs better than other classifiers with sensitivity,

specificity and accuracy values 1.00, 1.00 and 1.00 respectively for Mathematics and 0.998, 1.000 and 0.998 respectively for Portuguese subject.

The Mathews Correlation Coefficient (MCC) is another important parameter to evaluate the performance of the binary class classifiers. A coefficient of +1 represents a perfect classification, 0 an average random classification and -1 an inverse classification. It can be observed from the table 2 that, that classifier having high value of accuracy performance parameter for a particular family also have high MCC. In our experiment the MCC value we achieved is 1.00 for ROF for Mathematics and 0.994 for Portuguese subject.

The area under ROC curve (AUC) is an important statistical property to compare the overall relative performance of the classifiers. AUC can take values from 0 to 1. The value 0 for the worst case, 0.5 for random ranking and 1 indicates the best classification as the classifier has ranked all positive examples above all negative example. The figure 1 and figure 2 shows that AUC value of ROF classifier is greater than other classifier for our considered dataset equals to 1.00 for Mathematics and 0.999 for Portuguese subject.



**Fig. 1: AUC for different classifiers for Mathematics subject**

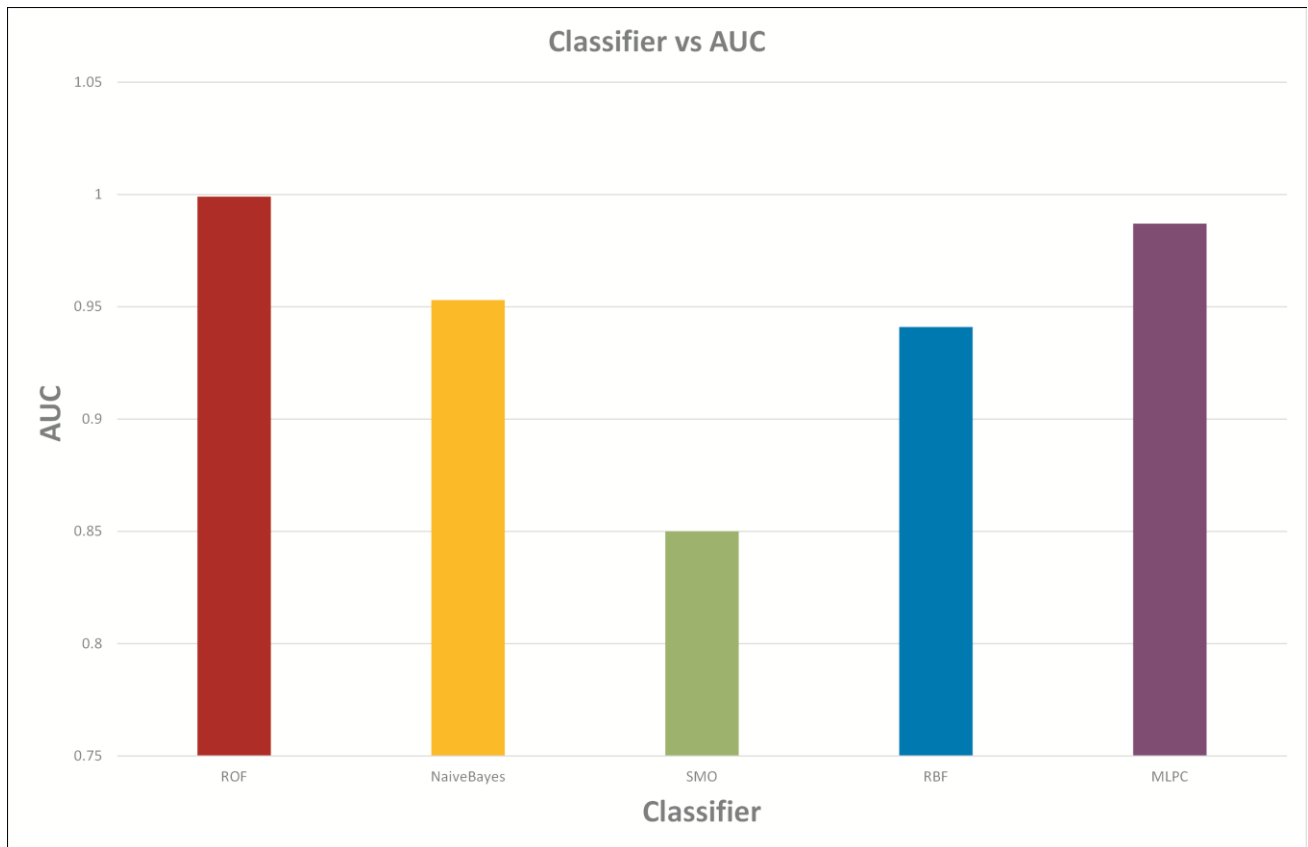


Fig. 2: AUC for different classifiers for Portuguese subject

#### 4. CONCLUSION AND FUTURE WORK

Predicting student's performance is regularly advantageous to assist the educators and learners in improving their learning and teaching process. We have compared the performance of five classifiers (including SVM, which was reported as the better performing classifier by the previous studies) in the prediction of students' performance. The experimental results of our proposed method have demonstrated that ROF has produced superior classification performance in terms of classification accuracy, AUC and MCC respectively for both of the subject Mathematics and Portuguese included in the dataset. It was also observed that few classifiers have yielded poor classification accuracy like SMO and RBF. This problem will be investigated in our future study by (i) Exploring all possible combination of various different types of input features and different machine learning algorithms, (ii) By deal with various factors that affects prediction performance (such as class imbalance, incomplete learning etc.) for improving the prediction accuracy and finally identifying the exact cause (through checking very high similarity by generating human interpretable rules through PART algorithm. In future I am also planning to develop a web based tool on our discovered algorithm which will be helpful in prediction of students' in risk of failure.

#### 5. REFERENCES

- [1] Turban, E.; Sharda, R.; Aronson, J.; & King D. (2007). Business Intelligence, A Managerial Approach (2nd ed.). Prentice-Hall.
- [2] B. M. Bidgoli, D. Kashy, G. Kortemeyer, & W. Punch (2003). Predicting student performance: An application of data mining methods with the educational web-based system lon-capa, in: Proceedings of ASEE/IEEE frontiers in education conference.
- [3] Romero C, Ventura S (2007) Educational data mining: a survey from 1995 to 2005. *Expert Syst Appl* 33(1):135–146.
- [4] Hsia T.,C., Shie A.,J., & Chen L.,C. (2008) Course planning of extension education to meet market demand by using data mining techniques—an example of Chinkuo technology university in Taiwan. *Expert Syst Appl* 34:596–602.
- [5] El-A. E., Abdel-A. R. (2008) Construction and analysis of educational tests using abductive machine learning. *Comput Educ* 51:1–16.
- [6] C. Romero, S. Ventura, P. G. Espejo, C. Herv. (2008). Data mining algorithms to classify students, in: *Educational Data Mining*.
- [7] V. Oladokun, A. Adebajo, O. Charles-Owaba. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course, *The Pacific Journal of Science and Technology* 9 (1) 72–79.
- [8] Lykourantzou I, Giannoukos I, Nikolopoulos V, Mpardis G, Loumos V (2009) Dropout prediction in e-learning courses through the combination of machine learning techniques. *Comput and Educ* 53:950–965.
- [9] Jantan H , Hamdan AR, Othman ZA (2010) Classification and prediction of academic talent using data mining techniques, *KES 2010, Part I, LNAI 6276*, pp 491–500.



- [10] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, H. Y. Seong. (2009) Predicting ndum student's academic performance using data mining techniques, in: Computer and Electrical Engineering, ICCEE'09, Second International Conference on, Vol. 2, IEEE, 2009, pp. 357–361.
- [11] M. M. Quadri, N. Kalyankar. (2010) Drop out feature of student data for academic performance using decision tree techniques, *Global Journal of Computer Science and Technology* 10 (2).
- [12] S. Sembiring, M. Zarlis, D. Hartama, S. Ramlina, E. Wani. (2011). Prediction of student academic performance by an application of data mining techniques, in: International Conference on Management and Artificial Intelligence IPEDR, Vol. 6, pp. 110–114.
- [13] M. M. A. Tair, A. M. El-Halees [2012] Mining educational data to improve students' performance: a case study, *International Journal of Information* 2 (2).
- [14] I. Hidayah, A. E. Permanasari, N. Ratwastuti.( 2013) Student classification for academic performance prediction using neuro fuzzy in a conventional classroom, in: Information Technology and Electrical Engineering (ICITEE), International Conference on, IEEE, 2013, pp. 221–225.
- [15] P. M. Arsad, N. Buniyamin, J.-I. A. Manan (2013) A neural network students' performance prediction model (nnsppm), in: Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on, IEEE, pp. 1–5.
- [16] D. M. D. Angeline. (2013) Association rule generation for student performance analysis using Apriori algorithm, *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)* 1 (1) p12–16.
- [17] C. Romero, M.-I. Lopez, J.-M. Luna, S. Ventura (2013). Predicting students' final performance from participation in on-line discussion forums, *Computers & Education* 68 458–472.
- [18] T. M. Christian, M. Ayub. (2014) Exploration of classification using nbtree for predicting students' performance, in: *Data and Software Engineering (ICODSE)*, 2014 International Conference on, IEEE, pp. 1–6.
- [19] M. Mayilvaganan, D. Kalpanadevi. (2014) Comparison of classification techniques for predicting the performance of students' academic environment, in: *Communication and Network Technologies (ICCNT)*, International Conference on, IEEE, pp. 113–118.
- [20] G. Elakia, N. J. Aarthi, (2014). Application of data mining in educational database for predicting behavioural patterns of the students, Elakia et al./ (IICSIT) *International Journal of Computer Science and Information Technologies* 5 (3) 4649–4652.
- [21] G. Gray, C. McGuinness, P. Owende. (2014) An application of classification models to predict learner progression in tertiary education, in: *Advance Computing Conference (IACC)*, 2014 IEEE International, IEEE, 2014, pp. 549–5154.
- [22] Raheela A., Agathe M., & Mahmood K. P. (2015). Predicting Student Academic Performance at Degree Level: A Case Study. *I.J. Intelligent Systems and Applications*, 01, 49-61
- [23] Glyn, H., & Chelsea, D. (2015). The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs), *Research and Practice in Technology Enhanced Learning*.
- [24] Juan, J., Rodriguez, L., & Kuncheva I. (2006). Rotation forest: a new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1619–30.
- [25] Ludmila I., Kuncheva, Jaun J., Rodriguez, (2007). An experimental study on rotation forest ensembles. *Proceedings of the 7th International Conference on Multiple Classifier Systems*, Springer-Verlag, Prague, Czech Republic, 459–68.
- [26] .Witten I., H., & Ian, H., (2011). *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems.
- [27] Tom Fawcett, (2003). *ROC graphs: Notes and practical considerations for data mining researchers*. Technical report, HP Laboratories