Asian Journal of Empirical Research 3(1): 74-86

Asian Journal of Empirical Research

journal homepage: http://aessweb.com/journal-detail.php?id=5004

DECISION OF THE CLASSIFICATION OF STUDENTS ON THE BASIS OF THEIR ACADEMIC GRADES

Muhammad Azam[1]
Muhammad Aslam[2]

ABSTRACT

This paper presents the classification of students studying at Forman Christian College University Lahore using some commonly applied classification techniques namely classification trees (CT) using Gini function as splitting rule, linear discriminant analysis (LDA) and multinomial logistic regression (MLR). As per rule set by the university, each student is getting a final grade or label (taken as response variable) after passing through number of performances (taken as predictors). Motive behind the classification of students on the basis of their academic achievements is to build a model using the available information in favor of each student and predict the class labels for the new comers or fresh students.

Key Words: Classification, Classification Techniques, Gini Function, Bootstrap Samples.

INTRODUCTION

University education is considered as most important in any country to achieve the goals and to run it for the progress. In the advanced countries, much attention is paid to university education and research. For example, in USA, the quality of the education is very good. It is very important for the universities to set a good environment and offer advance courses to produce quality students from the universities. Therefore, the grading criteria reflect the quality of education in the university. Normally, the grading is done on the basis of percentage marks obtained by a student. The grading and good GPA is also very important for the student. For example, in Pakistan, at least 3.0 GPA out of 4.0 is considered good for the higher studies. So, good GPA depicts good student and present that the student has a good command on the courses that he has taken during the previous study. Therefore, to classify the student according to their academic GPA has been a good area for the researchers. A lot of work has been done in the area, including for example, (Jonassen

[1]Department of Statistics, Forman Christian College University Lahore 54000, Pakistan **E-mail:**mazam72@yahoo.com

[2]Department of Statistics, Forman Christian College University Lahore 54000, Pakistan **E-mail:**aslam_ravian@hotmail.com

and Grabowski, 1993; Hartley, 1998; Duckworth and Seligman, 2005; Rayner, 2007; Dickerson et al. 2008; Cassidy and Eachus, 2000 a, b; Peterson et al. 2009; Naderi et al. 2009a, b; Sheard, 2009). More recently, (Cassidy, 2012) explored factors of students of academic achievement in higher education.

The purpose of this paper is to classify the students on the basis of their grade. We collected the grading data of statistics students of Stat-100 course form the Forman Christian College University, Lahore. We will present the classification of the students on the basis of decision tree analysis. In this paper, we have applied the classification tree technique to classify the students of Stat-100 studying at Forman Christian University Lahore (Pakistan). The students were classified into categories/grades A, B, C, D and F (A is representing superior while F a failing student) on the basis of their academic performance in various areas including assignments, quizzes, midterm examination, a project, class participation and final term examination. Basically, we have two objectives while performing classification i.e. classification of students and prediction of grade for new comers. These two objectives can easily be achieved by:

- Building a classification tree using the scores of students obtained in various areas as predictors. The constructed tree will classify the students into number of classes as A, B, C, D and F.
- Predicting class label of fresh students who are interested to enroll for the Stat-100. Just looking at the trend of scores in the constructed classification tree, class label can be predicted. This is done by assuming their scores equal to as splitting points used in the constructed tree.

METHODOLOGY

To classify the academic grades of Forman Christian College students, we have implemented three well known classification techniques. These techniques include classification trees (CT), multinomial logistic regression (MLR) and linear discriminant analysis (LDA). Here, we will describe classification trees in detail while others have been implemented for comparison purposes.

Classification

A data set $D=\{X, Y\}$, where X is a set of p predictors may be discrete, continuous or both and Y is a response variable taking values either continuous or having categories. The first case is considered as regression problem while second case is referred to as classification problem. Here we will only focus on classification problem. Classification is the problem of assigning data objects (individuals) to one of several pre-assumed classes on the basis of information collected from all the data objects. Several techniques are available to handle classification problems. Commonly applied techniques under current scenario are decision trees (Breiman et al. 1984), linear discriminate analysis (Fisher, 1936), neural networks (Ripley, 1996), support vector machines, logistic regression etc.

Decision Trees

In the data mining and machine learning research, classification and regression trees collectively named as decision trees, is an ideal and most suitable choice for the analysis of complex and high dimensionality datasets containing number of continuous and discontinuous predictor variables. The major difference between two types of trees is of response variable as described above. These trees are widely used as a source of generating classification rules because of utilizing simple but very strong tree algorithm called **TDIDT**, which is the abbreviation of *Top-Down Induction of Decision Trees*. This technique produces the decision rules in the form of tree structure. The tree structure is generated by repeating splitting process on the values of predictors. This process is named as *recursive partitioning* (Bramer, 2007).

A classification tree is a simple, non parametric model having a tree-based structure. Such type of tree is used to learn a classification model which is further used to predict the values of response variable given the values of predictor variables. So, there are two basic purposes of classification trees like other decision trees. First to build the classification trees using available datasets, in which each data object belongs to one and only one class. Second is to use the constructed classification trees to predict class labels of unseen or newly admitted data objects. Therefore, the main objective of classification trees is to provide binary or multiple tree structure of a given dataset and make predictions for future objects using the built-in tree structure.

Tree Construction

Construction of classification tree follows algorithm which generates simple binary nodes from root node to terminal nodes and finally one gets a binary tree structure and uses it to classify the objects with respect to their classes. It was introduced by well known American scientists Breiman, Friedman, Olshen and Stone in 1984 (Breiman et al. 1984) and implemented in CART by (Salford, 1995).

The algorithmic steps to elaborate construction of a classification tree provided in (Azam et al., 2009c) are stated below.

Algorithm 1: Algorithmic steps to generate a classification tree

Input: Let D be the training data, X is the set of predictors, p be the number of predictors, and S_p be the set of distinct points in p^{th} predictor.

1. Start with the top node i.e. $X_t = X$
2. For each new node t
3. **for** *(p* in 1 : *P) {*
4. **for** *(s* in 1 : S_p*) {*

5. Measurement of $g(x, s, t)$ and selecting one which maximizes i.e. $g^p(x^*, s^*, t^*)$.
6. }
7. $G[p] = g^p(x^*, s^*, t^*)$.
8. }
9. $G^* = maximum\ (G)$; it provides best splitting predictor x^* and its value s^* at node t^*.
10. **if** *(stopping condition is* TRUE*) {*
11. Declare node t as a terminal node and designate it with a class label.
12. }
13. **else** {
14. Generate two descendent nodes t_L and t_R with associated subsets that depends on the replies to the condition: Is $x^* \leq s^*$ and $x^* > s^*$ respectively
15. }

Output: A classification tree

Misclassification of Data Objects

In general, one can divide the error rates committed by any classification technique into two types i.e., training error rate and generalization error rate. Training error rate or apparent error rate is computed by dividing the number of misclassified data objects to total number of data objects in the training data.

Let m_i be the possible number of misclassified data objects in the i^{th} node of the constructed classification tree and N is the total number of data objects in the training data. Then the training error rate is given by $\text{Training Error Rate } (TER) = \frac{\text{total no of misclassified data objects in a classification tree}}{\text{total number of data objects in training data}}$

$$TER = \frac{\sum_i m_i}{N}.$$

Similarly, a confusion matrix is constructed for a three class problem to measure the generalization error rate (GER). This table can easily be extended for more number of classes.

Table-1. A 3 × 3 confusion matrix

True Class	**Predicted Class**			**Total**
	1	***2***	***3***	
1	n_{11}	n_{12}	n_{13}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	$n_{2.}$
3	n_{31}	n_{32}	n_{33}	$n_{3.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	N

$$\text{Generalization Error Rate } (GER) = 1 - \frac{\text{correctly classified data objects in confusion matrix}}{\text{total number of data objects in confusion matrix}},$$

$$GER = 1 - \frac{n_{11} + n_{22} + n_{33}}{N}.$$

It is well documented that a good classifier must not only classify the training data well, but also, it must accurately classify the data objects which are new or unseen to this fitted classifier. In other words, a good classifier must have low apparent as well as generalization error rates. More detailed discussion on this issue can be found in (Tan et al. 2006; Bramer, 2007).

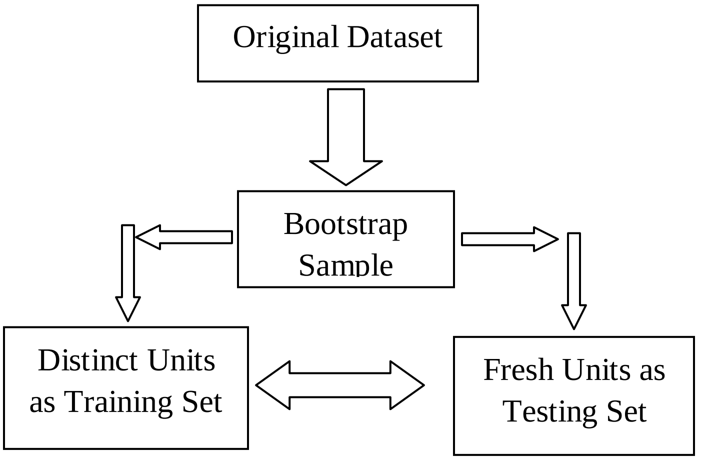
BOOTSTRAP SAMPLES

The bootstrap which is computer-based technique used to select random samples was introduced in 1979 by Efron for the estimation of any estimator. It is described in detail in (Efron and Tibshirani, 1993; Tan et al. 2006 and Merler and Furlanello, 1997) described the bootstrap strategy used in data mining problems as under:

1. A bootstrap sample selects N (i.e. size of the original dataset) data objects from the original data set D.
2. The sampling units (data objects) are drawn from the population using the "with replacement" criteria, that means once a data object is selected, there is a possibility to be selected again, so the probability of any data object to be selected in a sample is $1 - [1 - \frac{1}{N}] \cong 0.632$ for $N > 40$. As a result, for N sufficiently large, bootstrap replicates have a tendency to include $0.632N$ proportion of original data objects, while remaining $0.368N$ proportion is replaced by duplicate data objects.
3. Step1 and step2 are repeated B times and true error rate is estimated over the data objects not included in B bootstrap samples.

Selected data objects in the bootstrap samples formulate the training set that are used to construct a classifier, while data objects have not been selected by the bootstrap sample formulate the testing set (see Figure 1). Therefore, on average in B bootstrap samples, approximately 63% of the data objects are chosen as training and remaining 37% as testing set (Breiman, 1996a; Pan, 1999)

Figure-1. Bootstrap strategy



After splitting the random samples into training and testing parts, one has to estimate the training as well as testing misclassification rates. In practical problems, B bootstrap samples D_b^{tr} (training sets) are constructed from the original dataset. For each of them a true misclassification rate is computed over the data objects D_b^{ts} (testing sets) not included in the training set D_b^{tr} (Merler and Furlanello, 1997). The bootstrap estimate of true misclassification rate for B bootstrap samples is computed as

$$\bar{R}_{Boot}^{ts} = \frac{1}{B}\sum_{b=1}^{B} R^{ts}(T)_b$$

where $R^{ts}(T)_b$ is the true average misclassification rate of the b^{th} bootstrap sample and $\bar{R}_{Boot}^{ts}$ is the overall true misclassification rate for B constructed classification trees.

Algorithm 2: Classification of student as per their secured academic grade using classification trees.

Input: Grades_data, Number of trees to be constructed (not), Length of Grades_data (ldata)
for(h in 1:not) {# "not" for number of trees

- s = sort(unique(sample(1:ldata, ldata, replace=TRUE))); ls <- length(s); d[h] <- ldata-ls
- ds.tr[h] <- list(rpart(Grade ~., data = Grades_data, subset = s, method = c("class"), parms = list(split = c("gini"))))
- opt[h] <- which.min(ds.tr[[h]]$cptable[,"xerror"])
- cp[h] <- ds.tr[[h]]$cptable[opt[h], "CP"]
- # pruned tree
- pds.tr[h] <- list(prune(ds.tr[[h]], cp=cp[h])) # choose value of cp at min xerror
- pred_class <- predict(ds.tr[[h]], Grades_data[-s,], type="class")

- true_class <- Grades_data[-s, "Grade"]   # prediction for out of bag data objects
- tabl <- table(pred_class, true_class)
- misc1[h]=(1-sum(diag(tabl))/d[h])*100
- tab=table(predict(pds.tr[[h]], Grades_data[-s,], type="class"), Grades_data[-s, "Grade"]); table1[h] <- list(tab)   # prediction for out of bag data objects
- misc[h]=(1-sum(diag(tab))/d[h])*100

  }
- wmisc <- which.min(misc)
- ds.tr_min <- ds.tr[[wmisc]]
- pds.tr_min <- pds.tr[[wmisc]]; misc_pds <- misc[wmisc]; misc_ds <- misc1[wmisc]
- print("Misclassification Rate For Pruned Classification Tree"); print(misc_pds)
- plot(pds.tr_min, uniform=TRUE, margin=0.1, branch=0.5, compress=TRUE,
- main="Pruned Classification Trees of FCC Grades")
- text(pds.tr_min, use.n=TRUE, all=TRUE, cex=.8)

Output: Selection and plot of the final tree with minimum value of misclassification (generalization error).

Algorithm 3: Classification of student as per their secured academic grade using Multinomial Logistic Regression.

Input: Grades_data, Number of treesto be constructed (not), Length of Grades_data (ldata)
for(h in 1:not){# "not" for number of trees

- s = sort(unique(sample(1:ldata, ldata, replace=TRUE))); ls[h] <- length(s); d[h] <- ldata-ls[h]
- ds.tr[h] <- list(multinom(Grade ~., data= Grades_data, subset = s))
- pred_class=predict(ds.tr[[h]], Grades_data [-s,], type="class")
- true_class= Grades_data [-s, "Grade"]   # prediction for out of bag data objects
- table=table(pred_class,true_class)
- misc[h]=(1-sum(diag(table))/d[h])*100; table1[h] <- list(table)
- wmisc <- which.min(misc); Training_size <- ls[wmisc]
- ds.tr_min <- ds.tr[[wmisc]]; print(ds.tr_min)
- misc_ds <- misc[wmisc]
- print(table1[[wmisc]])
- print("Misclassification Rate For Logistic Model"); print(misc_ds)

Output: Selection of final model with minimum value of misclassification (generalization error).

Algorithm 4: Classification of student as per their secured academic grade using Linear Discriminant Analysis.

Input: Grades_data, Number of trees to be constructed (not), Length of Grades_data (ldata)
for (h in 1:not){# "not" for number of trees

- s = sort(unique(sample(1:ldata, ldata, replace=TRUE))); ls[h] <- length(s); d[h] <- ldata-ls[h]
- ds.tr[h] <- list(lda(Grade ~., data= Grades_data, subset = s))
- pred_class[h]=list(predict(ds.tr[[h]], data[-s,], type="class"))
- true_class=data[-s, "Grade"]   # prediction for out of bag data objects
- tabl <- table(pred_class[[h]][[1]], true_class)
- misc[h]=(1-sum(diag(tabl))/d[h])*100; table1[h] <- list(tabl)*}*
- wmisc <- which.min(misc); Training_size <- ls[wmisc]
- ds.tr_min <- ds.tr[[wmisc]]
- misc_ds <- misc[wmisc]
- print(ds.tr_min); print(table1[[wmisc]]);
- print(d[[wmisc]]);
- print("Misclassification Rate For Linear Discriminant Analysis "); print(misc_ds)

Output: Selection of final model with minimum value of misclassification (generalization error).

Data Description

The data about the academic grades of students studying Basic Statistics (Course code: Stat-100) in various sections of FC College University Lahore for three semesters Spring 2011, Summer 2011 and Fall 2011 has been taken. The total number of students listed in the data is 289. The final grade achieved by any student is based on assessment of following tasks and activities:

1. Class Participation (CP)
2. Project (Proj)
3. Quizzes (Quiz)
4. Assignments (Ass)
5. Mid Term Examination (Mid)
6. Final Term Examination (Final)

Overall, the academic grades have been divided into 11 categories starting from A (Superior) to F (Failing) as per university policy. But for the sack of simplicity and to avoid too many classes, the academic grades have been resettled. After this new setting, we divided the academic grades into five major categories by merging A^- to A, B^+ and B^- to B, C^+ and C^- to C and D^+ to D (see table 2). Grading is based on 4 quizzes (worth 10% of the final grade), 4 assignments (worth 10% of the final grade), 1 case study (worth 10% of the final grade), class participation (worth 5% of the final grade), midterm examination (worth 25% of the final grade) and a final examination (worth 40% of the final grade).

Table-1. The Grading Criteria

Grades	Quality Pts	Numerical Value	Meanings
A	4.00	93-100	Superior
A^-	3.70	90-92	
B^+	3.30	87-89	
B	3.00	83-86	Good
B^-	2.70	80-82	
C^+	2.3	77-79	
C	2.00	73-76	Satisfactory
C^-	1.70	70-72	
D^+	1.30	67-69	
D	1.00	60-66	Passing
F	0.00	≤ 59	Failing

Table-2. Adjusted Academic Grades

Grades	A	A^-	B^+	B	B^-	C^+	C	C^-	D^+	D	F
Adjusted Grades	A		B			C			D		F
Meaning	Superior		Good			Satisfactory			Passing		Failing

Statistical Analysis and Discussion

We have implemented three well known classification techniques namely classification trees, linear discriminant analysis and multinomial logistic regression in **R** software to classify the Academic Grades Data of student of FC College/ University, Lahore. All three classification techniques used in this analysis are available in R under the libraries rpart (recursive partitioning) using “Gini” as a node splitting function, lda (linear discriminant analysis) and nnet (neural networks) respectively. For the application of each technique, bootstrap samples have been generated and used for construction as well as prediction purposes. Overall 5000 bootstrap samples have been generated for each type of classification method and finally a single tree/ model that provides minimum generalization error (true misclassification rate) is chosen. For details see algorithm 2 to 4. The confusion matrix obtained after applying each classification procedure is presented in Table 3 to 5. The diagonal values in each confusion matrix are representing correctly classified data objects while off-diagonal values are representing incorrectly classified or misclassified data objects. A classification tree selected after certain optimization process is shown in Figure 2. It was built by using the bootstrap sample containing 186 data objects in the training set. The remaining 103 data objects (testing set) called out of bag are used for prediction purposes. So predicted class labels are obtained. At this stage, a confusion matrix is formed by using the original as well as predicted class labels of data objects contained in the testing set. Similar procedure is adopted for all three classification methods and misclassification rates presented in Table 6 are obtained. In Table 3, 13 out of 103 data objects were misclassified by the classification tree, so misclassification rate is $13/103 \times 100 = 12.62\%$ In Table 4, 9 out of 109 data objects were misclassified by linear discriminant analysis, so misclassification rate remained $9/109 \times 100 = 8.26\%$. Similarly, results are obtained in Table 5, where only 2 out of 103 data objects were misclassified by multinomial logistic regression. Therefore, misclassification rate remained $2/103 \times 100 = 1.94\%$ which is minimum as compared to other classification methods. We are interested in classification trees because of their attractive tree based structure which is more appealing to non-statisticians. The classification trees ability to predict class label of an individual by putting the information of each predictor used in the construction process creates the difference between other modeling techniques. A classification tree can easily be interpreted as compared to LDA and MLR which are more complex because of their mathematical structure.

CONCLUDING REMARKS

In the current paper, we classified the students of Forman Christian College University Lahore on the basis of their academic achievements during the semester. A data set consists of 289 students was considered and three well known classification methods including classification trees, linear discriminant analysis and multinomial logistic regression were applied. The bootstrap samples were taken to construct (using training set) and predict (using testing set) the class labels of the fresh or unseen data objects. Finally, a single tree/model is chosen which provided minimum misclassification rate for fresh data objects. Although, classification trees which provides tree-based structure is comparatively more attractive than other mathematical models. But multinomial logistic regression out performed and correctly classified over 98% of the fresh data objects. Similarly, classification trees and linear discriminant analysis correctly classified approximately 87% and 92% of the fresh data objects respectively.

Table-3. Confusion matrix generated for classification tree

	Predicted Class					
	Class	**A**	**B**	**C**	**D**	**F**
True Class	**A**	**15**	1	0	0	0
	B	4	**21**	0	0	0
	C	0	2	**11**	2	1
	D	0	0	0	**26**	3
	F	0	0	0	0	**17**

Table-4. Confusion matrix generated for linear discriminant analysis

	Predicted Class					
	Class	**A**	**B**	**C**	**D**	**F**
True Class	**A**	**21**	1	0	0	0
	B	0	**29**	1	0	0
	C	0	2	**11**	2	0
	D	0	0	0	**23**	3
	F	0	0	0	0	**16**

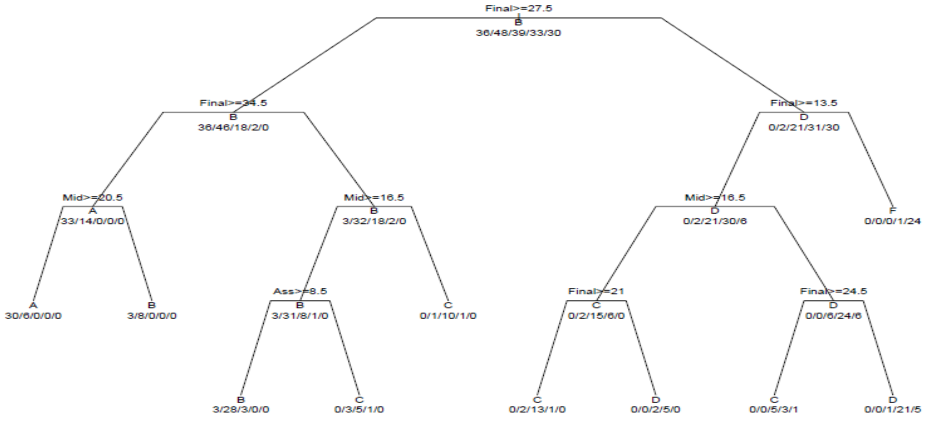
Table-5. Confusion matrix generated for multinomial logistic regression

	Predicted Class					
	Class	**A**	**B**	**C**	**D**	**F**
True Class	**A**	**22**	0	0	0	0
	B	1	**22**	0	0	0
	C	0	0	**17**	0	0
	D	0	0	0	**21**	1
	F	0	0	0	0	**19**

Table-6. Training/ Testing size and misclassification rates of the final selected model

Method	Training Size	Testing Size	Misclassification Rate (%age)
CT	186	103	12.62
LDA	180	109	8.26
MLR	186	103	1.94

Figure-2. Classification tree of academic grades achieved by students at Forman Christian College/ University, Lahore, Pakistan.



Acknowledgements

The authors are deeply thankful to the editor and the reviewers for their valuable suggestions to improve the quality of the paper.

REFERENCES

Azam M., Ostermann A. and Pfeiffer, K. P. (2009c) New developments in classification trees, Unpublished Ph.D thesis, University of Innsbruck: Austria.

Bramer, M. (2007) Principles of data mining, Springer-Verlag: London, UK.

Breiman, L. (1996a) Bagging predictors, Machine Learning, Vol. **24**, pp.123-140.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984) Classification and regression trees, Belmont, CA: Wadsworth International Group.

Cassidy, S. (2012) Exploring individual differences as determining factors in student academic achievement in higher education. Studies in Higher Education, Vol. **37**, No. 7, pp. 793-810.

Cassidy, S. and Eachus. P. (2002a) Development of the computer self-efficacy (CSE) scale: Investigating the relationship between CSE, gender and experience with computers. Journal of Educational Computing Research, Vol. **26**, No. 2, pp.133-53.

Cassidy, S. and Eachus, P. (2002b) The development of the general academic self-efficacy scale (GASE), Paper presented at the British Psychological Society Annual Conference, March 13-16, in Blackpool, UK.

Dickerson-Mayes, S., Calhoun, S. L., Bixler, E. O. and Zimmerman, D.N. (2008) IQ and neurop-psychological predictors of academic achievement. Learning and Individual Differences, Vol.**19**, No.2, pp.238-41.

Duckworth, A.L., and Seligman, M. E. P. (2005) Self-discipline outdoes IQ in predicting academic performance of adolescents. Psychological Science, Vol. **16**, No.12, pp.939-44.

Efron, B. and Tibshirani, R. J. (1993) An Introduction to the bootstrap, Chapman and Hall: London.

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics, Vol. **7**, No. 2, pp.179-188.

Hartley, J. (1998) Learning and studying: A research perspective, Routledge: London.

Jonassen, D. H. and Grabowski, B. L. (1993) Handbook of individual differences, learning, and instruction, Hillsdale, NJ: Lawrence Erlbaum Associates.

Merler, S. and Furlanello, C. (1997) Selection of tree-based classifiers with the bootstrap 632 + rule. Biometrical Journal, Vol. **39**, No. 3, pp. 369-382.

Naderi, H., Abdullah, R.,Tengku, H. Sharir, J. and Kumar, V. (2009a) Creativity, age and gender as predictors of academic achievement among undergraduate students. Journal of American Science, Vol. **5**, No. 5, pp.101-112.

Naderi, H., Abdullah, R., Tengku, H. Sharir, J. and Kumar, V. (2009b) Intelligence and gender as predictors of academic achievement among undergraduate students, European Journal of Social Sciences, Vol. **7**, No. 2, pp.199-207.

Pan, W. (1999) Shrinking classification trees for bootstrap aggregation. Pattern Recognition Letters, Vol. **20,** pp. 961-965.

Peterson, E. R., Rayner, S. G. and Armstrong, S.J. (2009) Researching the psychology of cognitive style and learning style: Is there really a future? Learning and Individual Differences, Vol. **19**, No. 4, pp. 518-523.

Rayner, S. (2007) A teaching elixir or best fit pedagogy? Do learning styles matter? Support for Learning, Vol. **22**, No. 1, pp. 24-30.

Ripley B. D. (1996) Pattern Recognition and Neural Networks, Cambridge: New York.

Salford, S. (1995) CART, San Diego, California: USA.

Sheard, M. (2009) Hardiness commitment, gender and age differentiate university academic performance. British Journal of Educational Psychology, Vol. **79**, No. 1, pp. 189-204.

Tan P., Steinbach M. and Kumar V. (2006) Introduction to Data Mining. Addison Wesley: New York.

Therneau, T. M. and Atkinson, E. J. (2009) Package rpart, CRAN Repository, Department of Statistics and Mathematics, WU Wien: Austria.

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S, Fourth Edition, Springer: New York.