The Comparison of Gini and Twoing Algorithms in Terms of Predictive Ability and Misclassification Cost in Data Mining: An Empirical Study

Murat Kayri[1], İsmail Kayri*[2]

[1]Department of Computer Engineering, Faculty of Engineering, MuşAlparslan University, 49100 Muş, Turkey

[2]Department of Electric Education, Faculty of Technical Education, Batman University, 72050 Batman, Turkey

Abstract**—***The classification tree is commonly used in data mining for investigating interaction among predictors, particularly. The splitting rule and the decision trees technique employ algorithms that are largely based on statistical and probability methods. Splitting procedure is the most important phase of classification tree training. The aim of this study is to compare Gini and Twoing splitting rules in terms of misclassification cost, obtained the optimal balanced trees and the importance of independent variables. This study shows that the results obtained using the Twoing criterion, as it yields a tree that is much more equally balanced than the tree obtained with the Gini criterion. Misclassification rate was slightly different for the two methods (19% using Twoing criterion and 21,2% for the Gini).Using Twoing splitting rule gets more importance level independent variables and the improvement values are higher than the Gini algorithm. All things being considered, the good performance of the Twoing splitting in this study combined with its robustness to get high classification accuracy, tree structure and the importance of independent variables.*

Keywords *—association rules, classification, data mining, parameter estimation, statistical learning.*

I. INTRODUCTION

Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules [1, 2]. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for future use. Data Mining is about solving problems by analyzing data already present in databases [3-5].

The analysis of Classification tree is one of the main techniques used in Data Mining. This technique is used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on predictor variables. Classification trees are a popular tool in applied statistics because their heuristic search approach based on impurity reduction is easy to understand and the interpretation of the output is straightforward [6]. Classification trees readily lend themselves to being displayed graphically, helping to make them easier to interpret than they would be ifonly a strict numerical interpretation were possible [7].

Classification and Regression Trees (CART) were first introduced by Breiman et al. [8]. CART is a binary decision tree algorithm capable of processing both continuous (Regression Trees) and/or categorical (Classification Trees) predictor and/or target variables. Furthermore, decision tree model results provide clear information on the importance of significant factors for prediction or classification [9]. CART algorithm works recursively: it partitions data into two subsets to make the records in each subset more homogeneous than the previous/alternative subsets; the two subsets are then split again until the homogeneity criterion or some other time-based stopping criteria are satisfied. The same predictor variable may be used several times in the process of growing the decision tree. The ultimate aim of splitting is to determine the right variable associated with the right threshold to maximize the homogeneity of the subgroups/branches [10].

Various criteria have been proposed for selecting the variable used for splitting the data in creating classification trees [11]. Gini and Twoing criterions are known as impurity measurements which are commonly used in classification trees. When using classification trees, the variable used at every node to split the tree affects the performance of the decision tree [12]. The problem of selecting the splitting variable is therefore not trivial. After the variable has been selected the value of the variable that gives the best split is then selected. The objective of classification is to allocate individuals to the correct population with the minimum classification error. Using classification trees this is done so as to arrive at the different classes with the least number of splits with the least misclassification errors.

The model in the classification tree is obtained by recursively partitioning the data space and fitting a simple prediction model within each partition and the partitioning is concluded visually. Classification Trees are designed for dependent variables that take a finite number of unordered values, with prediction error measured in terms of misclassification cost [13]. Essentially a classification tree consists of three steps. The steps are tree “growing”, “pruning tree” and “obtaining optimal tree” by calculating misclassification and complexity costs.

Recursively partitioning the dependent variable, which is the basic process of tree growing, depends on maximizing purity in two child nodes. The process of classification tree depends on training sample in terms of constructing the decision tree. The training sample involves a set of N cases. The denotation of *j* point one of the classes and it is written the training sample as follow:

$$L=\{(X_k, jk)\}_{k=1}^{N} \tag{1}$$

In the construction process of the decision tree, firstly, a split (t) and a set of associated cases (Lt) are constituted. If the all individuals (or any cases) have the same characteristic features then the cases belong to the same subclass (j), by this way it doesn’t need more work to construct this node. In this manner, it is declared that "t" is a leaf of the tree. If individuals in Lt belong more than one subclasses then it ought to be queried that “is Xi<=c?” in order to split the individuals in (Lt) into two parts. If the answer of the question is “yes” then t will be associated to the left child. But the answer is “no” then t will be associated to the right child.

To interpret the quality of any split, it ought to be used i(t) that i(t) measures the impurity of all the individuals in Lt. If all the individuals belong to just one class then it is seen that t is a pure node and in this case i(t) is equal to zero. Failing that, the maximum impurity will be occurred.

To describe i(t), it is introduced p(j|t) and an individual in Lt belongs to subclass j. Note that

$$\sum_j p(j|t) = 1 \tag{2}$$

The impurity of the Gini which is a typical measure can be expressed as follow;

$$i(t) = iG(t) = 1 - \sum_j p(j|t)^2 \tag{3}$$

If $p(j|t)$ are all equal completely then all the individuals belong to one class.To observe a split's change in terms of impurity, it is used the probabilities pL and pR. In this case an individual is assigned in tL or tR.And it is computed as follow:

$$\Delta i = i(t) - pLi(t_L) - pRi(t_R) \tag{4}$$

$\Delta i \geq 0$means that a split will never increase the totalimpurity.It can be seen obviously that Δi≥0 and it means that increasing the total impurity of a split will never be occurred. Applying the Gini impurity function to maximization problem;

$$\Delta i = i(t) - pLi(tL) - pRi(tR)) \tag{5}$$

It will be obtained the following change of impurity measure;

$$\Delta i(t) = -\sum_{j=1}^{J} p^2(j|t_p) + Pl\sum_{j=1}^{J} p^2(j|t_l) + Pr\sum_{j=1}^{J} p^2(j|t_r) \tag{6}$$

(see more at http://www.ams.org/samplings/feature-column/fc-2014-12#sthash.zN2OOc2d.dpuf)

Unlike Gini algorithm, Twoing rule will search for two classes that will make up together more than 50% of the data [14]. Twoing splitting rule will maximize the following change-of impurity measure:

$$\Delta i(t) = \frac{P_L P_R}{4}\left[\sum_{j=1}^{J} |p(j|t_L) - p(j|t_R)|\right]^2 \tag{7}$$

Although Twoing splitting algorithm allows to build more balanced trees, this algorithm works slower than Gini algorithm. Gini algorithm will search in learning sample for the largest class and isolate it from the rest of the data. Gini works well for noisy data. The Twoing criterion is often a superior performer on multi-class targets as well as on inherently difficult-to-predict (e.g. noisy) binary targets [15].

As specified before, the main purpose of tree growing is to partition the class into homogeneous subclasses. During the process of tree growing, partitioning the dependent variable to attain optimal purity in each node. By this way, a balanced or a saturated tree will be obtained [16]. In other words, the balanced/saturated tree yields the best fit for the model. Thus, to build a classification model, the data set is generally partitioned into two subsets. One of the subsets is used for training sample and the second subset is used for testing process [17]. Essentially, the purpose of using the training (learning) sample is to split nodes. Besides, the main aim of using testing sample is to compare the cost of misclassification [18]. Higher misclassification will be resulted if the structure of the tree is overly large. So that, overly large tree is “pruned” in the second step in order to reduce the misclassification cost. Removing branches which cause high misclassification is aimed in the pruning process. By this way, it can be obtained a simpler tree by removing increasingly nodes. Fundamentally, the process of the pruning is based on a complexity parameter that this parameter is identified as a cost function. And the tree size is a measure of how much additional accuracy is added to the tree to warrant the additional complexity [19-21]. The pruning process aims to cut off the nodes (or branches) resulting in high misclassification costs. A misclassification cost Cji defines the relative cost of misclassifying. If the calculation of Cji = 0 then there will not be seen any cost in classifying. To minimize the misclassification cost of all the individuals;

$$L(t); M(j) = \sum_i C_{ji}\, p(i|t) \tag{8}$$

The description of the misclassification cost at node t aims to minimize:

$$r(t) = \sum_i C_{j(t)i} p(i|t) \tag{9}$$

To evaluate the quality of a tree, the total misclassification cost of T or a leaf of T should be considered and the influence of t to the total misclassification can be computed as:

$$R(t) = r(t)p(t) \tag{10}$$

In Equation 10, $p(t)$ defines the probability of a class that this class makes unsuccessfully the tree to t.

The last step (optimal tree) is defined as that tree in the pruned sequence that achieves minimum cost on test data. Because test misclassification cost measurement is subject to sampling error, uncertainty always remains regarding which tree in the pruning sequence is optimal. BFOS (Breiman, Friedman, Olshen, and Stone) algorithm recommend selecting the “1 SE” tree that is the smallest tree with an estimated cost within1 standard error of the minimum cost (or “0 SE”) tree [15].

II. MATERIAL AND METHOD

A. Material (data set)

The data set was composed of 2247 university students. There are 12 variables (11 predictors and 1 response) in the model. The score of response variable was obtained from Research Anxiety Scale. The scale of Research Anxiety is composed of 12 items and each item is measured with using five-level Likertscala (1: Strongly disagree 2: Disagree 3: Uncertain 4: Agree 5: Strongly agree). The Research Anxiety Scale is developed by Büyüköztürk [22] and the aim of the scale is to measure university students' research anxiety level. The response is a binary variable (1: high level anxiety and 2: low level anxiety). While the high score shows a high level of research anxiety of the students, the low score shows a low level of research anxiety. The reliability of Research Anxiety was examined with Cronbach Alfa and the value of Cronbach Alfa was 0.828. This value shows that the reliability of the Research Anxiety Scale is good.

In this study, 11 predictors which they were asked to the students via prepared questionnaire were used. Some of predictors are nominal (dichotomous or multinomial), some of them are ordinal and the rest is continuous variables. The predictors in the model are "gender"," the department", "the graduation of the branch from high school (social, science, linguistics etc.)", "current transcript score", "the place of the growing up (village, city, small town)", "the degree of mother's education (primary, middle, high school, university, etc.)", "the degree of father's education (primary, middle, high school, university, etc.)", "the satisfaction from department", "the frequency of reading newspaper", "the section being read from newspaper" and lastly " the number of the books having been read".

B. Method

As mentioned above, the aim of this study is to compare Gini and Twoing splitting rules. During testing these rules; the misclassification cost, pruned tree to avoid overfitting, obtained the optimal trees and the independent variable importance will be compared in terms of efficiency and predictive ability. The rules of Gini and Twoing algorithms were discussed in introduction section.

The maximum tree depth was selected as up to five and the minimum number of cases was 100 for parent node and the minimum number of cases was 50 for child node. The minimum change in improvement was referred as 0.0001. Impurity measures were Gini and Twoing which was aimed for this study. In Gini, splits are found that maximize the homogeneity of child nodes with respect to the value of the dependent variable. By using Twoing, categories of the dependent variable are grouped into two subclasses. Splits are found that best separate the two groups in Twoing. The splits are all significant at α=5%.

The structures of the trees were avoided for overfitting problem for both Gini and Twoing rules. On the other hand, after the tree was grown to its full depth, pruning trims the tree downed to the smallest subtree that had an acceptable risk value. To select the subtree with the smallest risk (to avoid overfitting), maximum difference in risk (in standard errors) was entered as zero.

One of the major topics for classification tree is validation rule. The validation rule is based on estimating risk and basically there are three kinds of validation. These are re-substitution, cross-validation and split-sample validation. Hill and Lewicki [23] and Abdelrahman and Abdel-Hady [24] gave the following measures for risk estimates:

a. Re-substitution estimate: The estimation of Re-substitution is the proportion of individuals or cases which are misclassified by the splitter rule constructed from the data set. Risk estimation using re-substitution is the easiest method, but it usually underestimates the true risk.

b. Test sample estimate: In this method, The total number of all the individuals/cases is partitioned into two subsamples which are Z_1, and Z_2. The estimation of the test sample is the proportion of individuals/cases in the subsample Z_2 which is misclassified by the splitter constitued from the subsample Z_1. It is a useful way when the data set has a large size for partitioning.

c. V-fold cross-validation measure:The total number of individuals/cases are partitioned into ν subsamples Z_1, Z_2…,Z_v. And the subsamples of Z_1, Z_2…,$Z_v$has generally almost equal sizes. V-fold cross validation measure is the proportion of individuals/cases in the subsample Z which are misclassified by the splitter algorithm constitued from the subsample Z- Z_v. It is useful when the data set is too small for partitioning.

Regarding the above information, Test Sample Estimate method was preferred for this empirical study because of large enough for partitioning.67% of the sample was proportioned for training sample and the rest of the sample (33%) was used for testing sample. Random assignment was used to the estimate risk via Test Sample Estimate. The result of the Test Sample Estimate was confirmed that the value of the risk estimate was the lowest among the other methods (Re-substitution and V-fold cross-validation measure). The detailed results will be given at Results and Discussion section.

The prior probabilities were obtained from the training sample (empirical priors). It is known that the prior probabilities are estimates of the overall relative frequency for each category of the dependent variable prior to knowing anything about the values of the independent (predictor) variables. If the training dataset is a random sample prior probabilities should be obtained from training sample [25]. A membership assigns as follow:

$$\pi_j = \frac{N_j}{N} \tag{11}$$

for obtaining from the training sample. Whereas, using equal across categories method assigns objects as follow:

$$\pi_1 = \pi_2 = \frac{1}{2} \tag{12}$$

in terms of membership. The general criterion for assigning classes to nodes is given as follows:

Let:

$c(j|i)$: Cost of classifying i as j.

$\pi(i)$: Prior probability of i.

$N(i)$: Number of observations in category i in dataset.

$N_i(t)$: Number of observations in category i in node.

Node is class i, if:

$$\frac{c(j|i)\pi(i)N_{i(t)}}{c(i|j)\pi(j)N_{j(t)}} > \frac{N_i}{N_j} \tag{13}$$

for all values of j [24, 26].

III. Results and Discussion

In this section, it was tested different classification tree induction algorithms in order to identify the best-performing tree structure to predict the anxiety status of Research. Gini and Twoing algorithms were applied to the dataset by defining equal misclassification costs, test sample estimate, obtained prior probabilities from the training sample. And finally, obtained tree structures from these algorithms (Gini, Twoing) were scrutinized not only in terms of the depth of the relationship between the predictors and the dependent variable but also the level of the independent variable importance.

Normally, the dependent variable which is the score of Research Anxiety was a continuous variable. It was thought that does a student have an anxiety or not? So, the dependent variable was transformed to a binary type (high anxiety or low anxiety). Two-Step Cluster analysis was used to convert continuous variable to binary form. The findings of the cluster analysis were given in Table I:

TABLE I

THE FINDINGS OF THE CLUSTER ANALYSIS FOR THE ANXIETY SCORE

Clusters	Size	Mean	Std. Deviation
1	971 students (43,2%)	48,23 (high anxiety)	5,06
2	1276 students (56,8%)	33,17 (low anxiety)	4,38

The distribution of the Cluster 1 (high anxiety) and the Cluster 2 (low anxiety) have been shown in Table I.

After processing of the cluster analysis, the main aim of the study was modeled and the estimation of the relationship between the independent variables and the dependent variable was examined by using Gini and Twoing rules separately (comparatively) in terms of performance and predictive ability.

A. Result of Ginisplitting

Gini rule was used to split into groups based on values of all independent variables. The tree diagram (Figure 1) shows tree construction based on the entire sample of 2247 cases, Test Sample Estimate 0.05 adjustment of the probabilities, a minimum parent node size of 100, a minimum child nodes size of 50 and equal misclassification costs. 67% of thesample was proportioned for the training sample and the rest of the sample (33%) was used for the testing sample. Random assignment was used to estimate the risk via Test Sample Estimate. Lastly, to construct the sub-tree with thesmallest risk (to avoid overfitting), maximum difference in risk (in standard errors) was entered as zero.

There are totally 11 nodes that consist of 6 terminal nodes and the first node placed in the tree is root node. The first independent variable (department) splitted the root node into two child nodes (Primary School Education, Science Education, Turkish Education, etc.); (Divinity Education, Theology, Literature, Philosophy, Sociology, etc.). The improvement value of this classification is 0,183, which is significant at α = 5%. The second classifier was "the number of the books having been read" which was splitted by Node 2. "the number of the books having been read " independent variable was splitted into two sub-child nodes (Node 5 and Node 6): Less than between 50 and 75 books and upper than 50-75 books with 0,007 value of improvement. Node 1 which was splitted by Department discriminator node was splitted into two child nodes (Node 3 and Node 4; transcript score): less than 3,025 and upper than 3.025, with 0,001 improvement value. Node 3 (transcript) was splitted into two child nodes (department): "Primary School Education, Science Education, Turkish Education, Social Science Education (Node 7)" and "Preschool Education, Maths Science (Node 8)", with 0,001 improvement value. Node 6 was splitted into two sub-child nodes (the section being readfrom newspaper): "News, Scientific article (Node 9)" and "Sport, Magazine, Fortune Prediction, others (Node 10)", with 0,003 improvement value. Table II shows the prior probabilities of the clusters.

TABLE II
PRIOR PROBABILITIES

Research Anxiety	Prior Probability
1	0,426
2	0,574

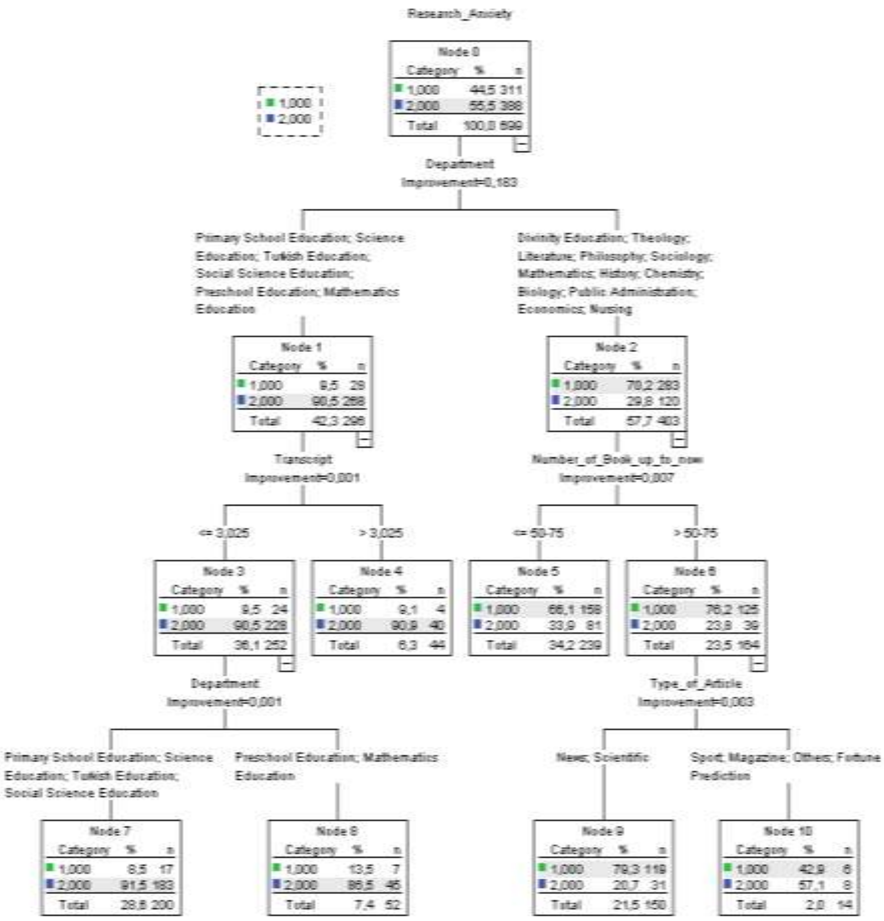

Fig.1. Classification tree with Gini splitting

As mentioned before, the prior probabilities were obtained from the training sample. The classificationaccuracy and the classification risk were given in Table III and Table IV below.

TABLE III
RISK ESTIMATES OF THE TRAINING AND THE TEST SAMPLES OF GINI SPLITTING

Sample	Estimate	Std.Error
Training	0,212	0,010
Test	0,212	0,015

TABLE IV
THE PREDICTION OF THE CLASSIFICATION ACCURACY FOR GINI SPLITTING

Sample	Observed	Predicted		
	Observed	1	2	Percent Correct
Training	1	611	49	92,6%
	2	279	609	68,6%
	Overall Percentage	**57,5%**	**42,5%**	**78,8%**
Test	1	283	28	91,0%
	2	120	268	69,1%
	Overall Percentage	**57,7%**	**42,3%**	**78,8%**

The misclassification rate (risk Test Sample Estimate-split sample) is 0,212 for training sample (with 0,010 standard error) and 0,212 for test sample with 0,015 standard error. The training sample of the classification accuracy prediction is 78,8% (overall percentage), with 92,6% of sensitivity and 68,6% of specificity. The test sample of the classification accuracy prediction is 78,8% (overall percentage), with 91,0% of sensitivity and 69,1% of specificity.

The normalized importance of the independent variables in the classification was given in Table V and Figure 2.

TABLE V
THE IMPORTANCE OF THE INDEPENDENT VARIABLES

Independent Variable	Importance	Normalized Importance
Department	0,184	100%
The graduation of the branch from high school	0,013	6,8%
The number of the books having been read	0,007	3,7%
The satisfaction from department	0,005	2,8%
The section being read from newspaper	0,005	2,6%
The degree of mother's education	0,002	1,3%
Current transcript score	0,002	1,2%
The degree of father's education	0,001	0,6%
Gender	0,000	0,1%
The frequency of reading newspaper	2,156E-6	0,0%

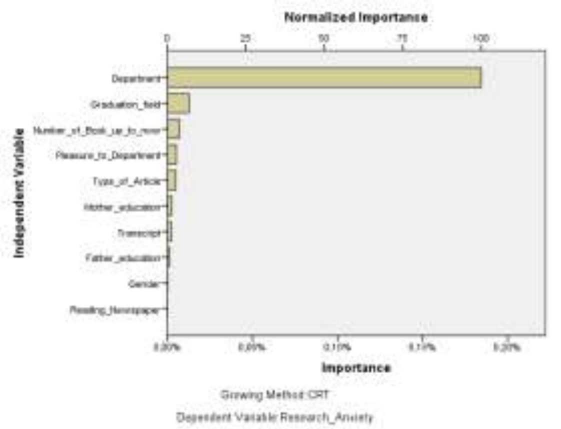

Fig.2.The importance of the independent variables

The most importance independent variable is "department", with 0,184 importance value. Although Gini measure hasn't met "the graduation of the branch from high school" independent variable as a splitter, Table V shows that the predictor of "the graduation of the branch from high school" has an importance effect on the dependent variable, with 0,013 importance level. However, the importance level of the transcript score predictor is low but this predictor is obtained as a splitter by using Gini measure. "The number of the books having been read" independent variable has been attained not only as a splitter but important (with 0,007 value of importance). The other independent variables can be interpreted with examining Table V.

B. Result of Twoingsplitting

Splits are found that maximize the homogeneity of child nodes with respect to the value of the dependent variable in Gini measure. Unlike Gini, categories of dependent variable are grouped into two classes and splits are found that best separate the two groups by using Twoing measure. The prior probabilities, the validation rule (Test Sample Estimate-split sample) and the other adjustments which were set for Gini, all these adjustments were the same for Twoing measure too. The findings of the Gini and the Twoing algorithms will be compared in terms of tree constructions, prior probabilities, misclassification costs, independent variables importance. By this way, this study will reveal the performance of Gini and Twoing and this study also will illuminate researchers in order to select better one (Gini or Twoing) for their researches.

The tree construction via Twoing splitting algorithm was shown in Figure 3. Unlike the findings of the Gini, the Twoing yielded 7 nodes that consisted of 4 nodes and the first node placed in the tree is root node.

Similarly, the first independent variable is "department" of the students. And this independent variable splitted the root node into two child nodes (Primary School Education, Science Education, Turkish Education, etc.); (Divinity Education, Theology, Literature, Philosophy, Sociology, etc.). The improvement value of this classification is 0,340 and apparently this value is bigger than the Gini's finding (0,183). The second classifier is "the number of the books having been read" which is splitted by Node 2, with 0,016 improvement value. This value is 0,007 for the Gini's growth tree and apparently the Twoing's improvement is bigger than the Gini's. The Node 4 which is yielded by Node 2 constituted two child nodes, with 0,008 improvement value.

Comparing Gini's and Twoing's tree construction, it can clearly be observed that the constructions of the two trees are so different from each other in terms of importance of independent variables and improvement values. Table VI shows the prior probabilities of the clusters.

TABLE VI
PRIOR PROBABILITIES

Research Anxiety	Prior Probability
1	0,434
2	0,566

Like the Gini, prior probabilities were obtained from the training sample in the Twoing process. The classification accuracy and the classification risk were given in Table VII and Table VIII below.

Like the Gini, prior probabilities were obtained from the training sample in the Twoing process. The classification accuracy and the classification risk were given in Table VII and Table VIII below.

TABLE VII
THE PREDICTION OF THE CLASSIFICATION ACCURACY FOR TWOING CRITERION

Sample	**Observed**	**Predicted**		
	Observed	1	2	Percent Correct
Training	1	606	60	91,0%
	2	281	588	67,7%
	Overall Percentage	**57,8%**	**42,2%**	**77,8%**
Test	1	288	17	94,4%
	2	118	289	71,0%
	Overall Percentage	**57,0%**	**43,0%**	**81,0%**

TABLE VIII
TWOING'S RISK ESTIMATES OF THE TRAINING AND THE TEST SAMPLES

Sample	**Estimate**	**Std.Error**
Training	0,222	0,011
Test	0,190	0,015

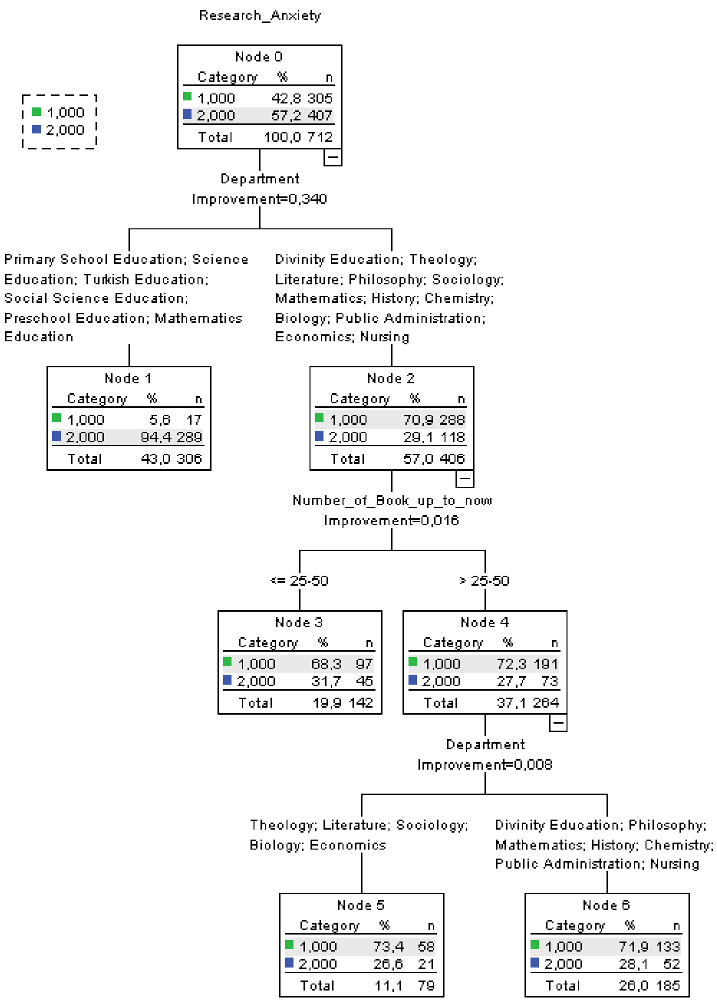

Fig.3. Classification tree with Twoing splitting

The misclassification rate (risk Test Sample Estimate-split sample) is 0,222 for the training sample (with 0,011standard error) and 0,190 for the test sample with 0,015 standard error. The training sample of the classification accuracy prediction is 77,8% (overall percentage), with 91,0% of sensitivity and 67,7% of specificity. The test sample of the classification accuracy prediction is 81% (overall percentage), with 94,4% of sensitivity and 71,0% of specificity. Comparing Table III-VI and Table IV-VII, it will be noticed that the result of the Twoing (especially for Test Sample) is more accuracy than the Gini's result in terms of classification accuracy and the risk estimate. On the other hand, the error rate of the Twoing criterion is lower than the Gini's. It means that the findings of Twoing are more robust and reliable in terms of classification accuracy. By this way, it can be said that the tree construction of the Twoing splitting is more reliable and realistic than the Gini's. The normalized importance of the independent variables in the classification (Twoing's criterion) was given in Table IX and Figure 4.

TABLE IX
THE IMPORTANCE OF THE INDEPENDENT VARIABLES

Independent Variable	Importance	Normalized Importance
Department	0,350	100%
The graduation of the branch from high school	0,031	9,0%
The number of the books having been read	0,016	4,6%
The satisfaction from department	0,007	2,0%
The section being read from newspaper	0,004	1,3%
The degree of mother's education	0,003	0,8%
Current transcript score	0,001	0,4%

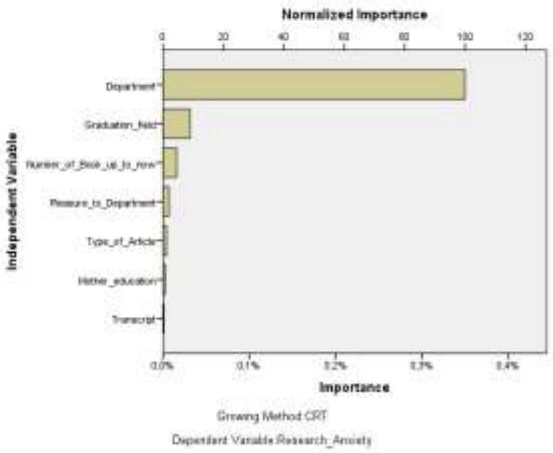

Fig. 4. The importance of the independent variables

The most importance independent variable was "department", with 0,350 importance value. Although Twoing measure hasn't met "the graduation of the branch from high school" independent variable as a splitter, Table IX shows that the predictor of the graduation of the branch from high school has an importance on the dependent variable, with 0,031 importance level. "The number of the books having been read" independent variable has been attained not only as a splitter but also important (with 0,016 value of importance). The other independent variables can be interpreted with examining Table IX.

Comparing the results of Twoing and Gini criterions in terms of the independent variable importance; Twoing splitting method can attain higher normalized importance percentage than the Gini's method. Having the higher importance percentage is meant the higher reliability and productivity. Overall, the performance of the Twoing is better than the Gini in terms of the predictive ability, the risk estimate and the robust tree construction.

IV. CONCLUSION

Splitting methods in classification tree are critical to reliable results and should be evaluated separately to determine optimum performance of classification methods [27]. Also, splitting procedure is the most important phase of classification tree training [28]. The Gini and the Twoing splitting methods were applied to research anxiety data of 2247 students in Turkey. The dependent variable has a binary form of research anxiety score {1= high anxiety, 2= low anxiety}. The two splitting methods produced different classifiers. However, misclassification rate was slightly different for the two methods (19% using Twoing criterion and 21,2% for the Gini). Especially, the importance of independent variables was so different. Using Twoing splitting rule gets more importance level and the values of improvement level are higher than Gini algorithm.

The main strength of tree-structured classification is that it provides understanding and insight of the data [29]. More insight can be obtained by using Twoing splitting criteria when compared with the Gini. Shih [30] tested families of splitting criteria (compared Chi-squared, Entropy and Gini criteria) and the results of Chi-squared and Entropy were slightly better than Gini in terms of accuracy of classification and strength of tree-structured. In this study, the accuracy of classification tree produced by Twoing is even better than that obtained by the Gini criterion.

Zambon et al. [31] compared the performance of splitting rule on image processing. The result of this study indicated that the Gini splitting is more appropriate rule than Twoing splitting rule for image classification. For image classification, the result hasn't proved our findings. According to our findings, Twoing splitting rule is better than the Gini in terms of classification accuracy, tree structure and importance level of independent variable. But for image processing, Zambon's et al. [31] study proved that the Gini is better than Twoing rule in terms of classification, the opposite to our findings.

In a theoretical study of the Gini and Twoing splitting rules, Breiman [32] concludes that all criteria should produce similar results when the number of values of the dependent variable is small. Hamza and Larocque [33] examined an empirical comparison of Twoing and Gini rules and this study indicated that the difference between the Gini and Twoing rules was negligible. Even so, the performance of Twoing was slightly better than the Gini rule [33]. Another empirical study shows that the results obtained using the Twoing criterion, as it yields a tree that is much more equally balanced than the tree obtained with the Gini criterion. This observation can be attributed to the fact that the Twoing criterion seeks for splits that are roughly equal in size and Twoing is better for multi-class dependent variables than Gini [34].

In general, exclude some special situations (such as image processing studies), a lot of studies [15, 35, 36] show that the Twoing splitting criteria is better than the Gini in terms of some aspects (such as misclassification error). Although the Twoing splitting rule allows us to build more balanced trees, this algorithm works more slowly than the Gini rule; for example, if the total number of classes is equal to C, then we will have 2C−1 possible splits [36].

The splitting rule and the decision trees technique employ algorithms that are largely based on statistical and probability methods. The splitting rule is essentially the heart of the transformation process from Data to Information [37]. To reveal the relationship between dependent and independent variables, choosing splitting rule is too vital. To conclude, all things being considered, the good performance of the Twoing splitting in this study combined with its robustness to get high classification accuracy, balanced tree structure and the importance of independent variables.

REFERENCES

[1] J. Nikita and V.Srivastasa,"Data Mining Techniques: A Survey Paper," International Journal of Research in Engineering and Technology, vol. 2, pp. 116-119, 2013.

[2] B.K. Baradwaj and S. Pal,"Mining Educational Data to Analyze Students' Performance," International Journal of Advanced Computer Science and Applications, vol. 2, pp. 63-69, 2011.

[3] Z. Joseph,"Data Mining as a Civic Duty – Online Public Prisoners' Registration Systems,"International Journal on Social Media MMM: Monitoring, Measurement, and Mining, vol. 1, pp. 84-96, 2010.

[4] T. Waheed, R.B. Bonnell, S.O. Prasher and E. Paulet, "Measuring Performance in Precision Agriculture: Cart—A Decision Tree Approach,"Agricultural Water Management, vol. 84, pp. 173-185, 2006.

[5] J. Gehrke, R. Ramakrishnan and V. Ganti,"Rainforest-a Framework for Fast Decision Tree Construction of Large Datasets," Data Mining and Knowledge Discovery, vol. 4, pp. 127-162, 2000.

[6] C. Strobl, A.L. Boulesteix andT. Augustin,"Unbiased Split Selection for Classification Trees Based on theGini Index,"Computational Statistics & Data Analysis, vol. 52, pp. 483-501, 2007.

[7] B.D. Ripley, "Pattern Recognition and Neural Networks", pp. 403-410,Cambridge University Press, 1996.

[8] L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone,"Classification and Regression Trees",pp. 22-34,Chapman & Hall/CRC, 1984.

[9] G.K.F. Tso and K.K.W. Yau,"Predicting Electricity Energy Consumption: A Comparison of Regression Analysis, Decision Tree and Neural Networks,"Energy, vol. 32, pp. 1761-1768, 2007.

[10] E. Bastı, C. Kuzey andD. Delen,"Analyzing Initial Public Offerings' Short-Term Performance Using Decision Trees and Svms,"Decision Support Systems, vol. 73, pp. 15-27, 2015.

[11] E. Muchai and L.Odongo,"Comparison of Crisp and Fuzzy Classification Trees Using Gini Index Impurity Measure on Simulated Data," European Scientific Journal, vol. 10, pp. 130-134, 2014.

[12] Z. Yua, F. Haghighat, B.J. Fung and H.Yashino,"A Decision Tree Method For Building Energy Demand Modeling,"Energy and Buildings, vol. 42, pp. 1637-1646, 2010.

[13] W.Y. Loh,"Classification and Regression Trees,"WIREs Data Mining Knowledge Discovery, vol. 1, pp. 14-23, 2011.

[14] R. Timofeev,"Classification and Regression Trees (Cart) Theory And Applications", MSc, Center of Applied Statistics and Economics Humboldt University, Berlin, Germany, 2004.

[15] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S Yu, Z.H. Zhou, M. Steinbach, D.J. Hand and D. Steinberg,"Top 10 Algorithms in Data Mining,"Knowledge and Information Systems, vol. 14, pp. 1-37, 2008.

[16] Y.H. Cho, J.K. Kim andS.H. Kim,"A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction,"Expert Systems with Applicaitons, vol. 23, pp. 329-342, 2002.

[17] I.H. Witten and E. Frank, "Weka Machine Learning Algorithms in Java, in Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", pp. 122-132, Morgan Kaufmann Publishers, 2000.

[18] L.Y. Chang andH.H. Wang,"Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques,"Accident Analysis & Prevention, vol 38, pp. 1019-1027, 2006.

[19] PJ. Bradford, C. Kunz, R. Kohavi, C. Brunk and C.E. Brodley,"Pruning Decision Trees with Misclassification Costs,"Lecture Notes in Computer Science, vol. 1398, pp. 131-136,1998.

[20] F. Esposito, D. Malerba andG.A. Semeraro,"Comparative Analysis of Methods for Pruning Decision Trees,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 476-491, 1997.

[21] M. Bramer,"Using J-Pruning to reduce overfitting in classification trees," Knowledge-Based Systems, vol. 15, pp. 301-308, 2002.

[22] Ş. Büyüköztürk,"The Development of Research Anxiety," Journal of Education Management, vol. 3, pp. 453-464, 1996.

[23] T. Hill and P.Lewicki, "Statistics: Methods and Applications, A Comprehensive Reference for Science, Industry, and Data Mining", USA,Statasoft Inc., 2006.

[24] A.I. Abdelrahman andD.H. Abdel-Hady,"Classification of InsolventSmall Businesses in Egypt by Some Running Cost Variables: A Decision Tree Approach," Applied Mathematical Sciences, vol. 5, pp. 421-440, 2011.

[25] W.Y. Loh,"Classification and Regression Tree Methods", pp. 315-323, Wiley, 2008.

[26] R.J. Lewis,"An introduction to classification and regression tree (cart) analysis,"*Annual Meeting of the Society for Academic Emergent Medicine*, San Francisco,May 22-25, pp. 1-14, 2000.

[27] M. Pal and P.M. Mather,"An Assessment of the Effectiveness of Decision Tree Methods for Land Cover

Classification,"Remote Sensing of Environment, vol. 86, pp. 554-565, 2003.
[28] M.H. Moattar, M.M. Homayounpour and D. Zabihzadeh, "Persian text normalization using classification tree and support vector machine," Information and Communication Technologies (ICTTA'2006), Damascus, April 24-28, pp. 1308-1311, 2006.
[29] D.J. Hand,"Construction and Assessment of Classification Rules", pp. 79-85, John Wiley, 1997.
[30] Y.S. Shih,"Families of Splitting Criteria for Classification Trees,"Statistics and Computing, vol. 9, pp. 309-316, 1999.
[31] M. Zambon, R. Lawrence,A. Bunn andS. Powell,"Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis,"Photogrammetric Engineering and Remote Sensing, vol. 72, pp. 25-30, 2006.
[32] L. Breiman,"Technical Note: Some Properties of Splitting Criteria,"Machine Learning, vol. 24, pp. 41-47, 1996.
[33] K. Hamza andD. Larocque,"An Empirical Comparison of Ensemble Methods Based on Classification Trees,"Journal of Statistical Computation and Simulation, vol. 75, pp. 629-643, 2006.
[34] G. Martens, H.D. Meyer, B.D.Baets, M. Leman, M. Lesaffre andJ.P. Martens,"Tree-Based Versus Distance-Based Key Recognition in Musical Audio," Soft Computing, vol. 9, pp. 565-574, 2005.
[35] C. Conversano and F.Mola, Detecting Subset of Classifiers for Multi-attribute response prediction, In: Palumbo F, Lauro CN, Greenacre MJ, editors. Data Analysis and ClassificationStudies in Classification, Data Analysis, and Knowledge Organization. Berlin, Germany: Springer Berlin Heidelberg Publisher, 2010, pp. 233-240.
[36] U. Boryczka and J. Kozak,"Enhancing the Effectiveness of Ant Colony Decision Tree Algorithms By Co-Learning,"Applied Soft Computing, vol. 30, pp. 166-178, 2015.
[37] V. Chiew,"A software engineering cognitive knowledge discovery framework," *Computer Society, Proceedings of the First IEEE International Conference on Cognitive Informatics (ICCI'02)*, Calgary,August 18-20, pp. 163-172, 2002.