RASHI 4 (1) : 17-25 (2019)

An application of Boosted Classification and Regression Trees (CART) in agricultural ergonomics

Samir Barman, Ramasubramanian V. and Mrinmoy Ray

ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110 012

Received : ; Revised : ; Accepted :

ABSTRACT

Classification and Regression Trees (CART) is a decision tree based approach widely used for classification when dependent variable is categorical and for regression when dependent variable is continuous. The advantages of this approach are that they are non-parametric, data-driven, can handle outliers, suitable when there are interactions between the independent variables. In spite of several advantages, some of the drawbacks of CART approach are low model accuracy, high prediction variance and model overfitting. Boosted CART is one of the popular approaches for dealing with this problem in which multiple trees are trained sequentially based on information from previously grown trees. In this study, two approaches conventional CARTand Boosted CART have been compared for classification problem. Empirical results based on simulation as well as real datasets uncover that Boosted CART performed best in terms of classification accuracy over conventional CART.

Keywords : Boosting; Classification and Regression Trees;Total accuracy rate.

1. INTRODUCTION

Now a days there is an increasing demands for development of a reliable decision rule that can be used to classify new observations into some predefined categories. In certain specific situations, the existing traditional statistical methods are not suitable in the sense that these are cumbersome and has less utility to handle these classification problems. There are many possible reasons for these difficulties.Firstly, there are generally many possible "predictor" variables which makes the task of variable selection difficult. Traditional statistical methods are poorly suited for this sort of multiple comparison. Secondly, the distribution of predictor variables are not well defined. Generally many variables in any field are not normally distributed and different groups of subjects may have markedly different degrees of variation or variance. Thirdly, complex interactions or patterns may exist in the data. For example, the value of one variable (e.g., age) may substantially affect the importance of another variable (e.g., weight). These types of interactions are generally difficult to model and virtually impossible to model when the number of interactions and variables becomes substantial. Fourthly, the results of traditional methods may be difficult to use. For example, a multivariate logistic regression model yields a probability for different classes of the dependent variable, which can be calculated using the regression coefficients and the values of the explanatory variable. But the practitioners generally do not think in terms of probability but, rather in terms of categories, such as "presence" versus "absence." Regardless of the statistical methodology being used, the creation of a decision rule requires a relatively large dataset. Following the above reasons, in recent times, there has been an increasing interest in the use of Classification and Regression Tree (CART) analysis. It is a tree-building technique which is different from traditional data analysis methods. In many studies it is found that CART analysis performs quite effective for creating decision rules as well or better than the traditional methods. In addition, CART is often used to uncover complex interactions between predictors which may be difficult or impossible using traditional multivariate techniques. It is now possible to perform a CART analysis with a simple understanding of each of the multiple steps involved in its procedure. Classification tree methods such as CART are convenient way to produce a prediction rule from a set of observations described in terms of a vector of features and a response value. Use of CART has greatly increased in popularity during the recent years. Tree based decision methods i.e. CARTs are statistical systems that mine data to predict or classify future observations based on a set of decision rules and are sometimes called rule induction methods because the reasoning process behind them is clearly evident when browsing the trees.The CART methodology have been applied in several areas such as agriculture, medicine, forestry, natural resources management etc. as alternatives to the conventional approaches such as discriminant function method, multiple linear regression, logistic regression etc. Analytic results demonstrated that CART outperforms the conventional approaches in terms of scoring accuracy and misclassification rate.

E-mail: samir.iasri15@gmail.com

CART has some advantages which are as follows: (Lewis, 2000)

- Has no assumption requirement i.e. distribution assumption free.
- Easy to explore and to take decision when the data used is complex and multivariate.
- The result of the analysis is simple and easy to classify new data efficiently and easy to interpret.

Though CART has many advantages over the conventional approaches, it has some weakness that the result is not stable i.e. a small changes in training data will change the prediction result of the yielded trees and gives different results (Sutton, 2005). This causes low model accuracy and high prediction variance. Also CART suffers on model overfitting problems. To solve these, there are some alternative methods that handles the problems of CART analysis by improving model accuracy and for obtaining prediction variance.

Boosting is one such approach that is used to improve the predictions resulting from a decision tree. Like CART, Boosting can be applied for both classification as well as regression problems. Boosting also work ina way that trees are grown sequentially, each tree is grown using information from previous grown trees.In boosting each tree is fitted on a modified version of original data set. This boosting is very useful when we have a lot of dataset and thetree is to be very complex. The advantage of boosting is that it overcomes the drawbacks of CART by using boosting technique to combine several decision trees additively in order to obtain a lower variance prediction.

In this study, mainfocus is on classification problem. The performances of these three methods are compared by anempirical study.

1. Materials and Methods

1.1. Data description

The data for the present study has been taken in the area of agricultural ergonomics obtained from Division of Agricultural Engineering, IARI, New Delhi collected during 2007-08. The variable considered as the dependent variable (Y) is dichotomous i.e. “presence” or “absence” of discomfort for the farm labourers during agricultural field operation. The set of qualitative explanatory variables are: modes of operation (X_1) and percent aerobic capacity of the farm labourers (X_2), each having two levels. The variable X_2 has two levels labelled as low and high viz., less than or equal to 35% and greater than 35% of aerobic capacity of the farm labourers respectively. The dataset consists of 405 observations available for the study. In this dataset, broadly two levels of modes of operation viz. predominantly foot operated (e.g. Bicycle, Stepper, Pedal etc.) and other mode of operation (e.g. Flywheel, Rocking etc.) are considered. The quantitative explanatory variables are: load given to farm machinery (X_3), difference between working and resting heart rates (X_4) and oxygen consumption at the time of farm operation (X_5).

Table.1: Variables used in the study

Variable	Information	Category
Y	Discomfort for the farm labourers	0= Absence 1= Presence
Qualitative		
X_1	Modes of operation	0= Predominantly foot operated 1= Hand driven
X_2	Aerobic capacity of the farm labourers	0= Low (≤ 35%) 1= High (≥ 35%)
Quantitative		
X3	Load given to farm machinery	
X4	Difference between working and resting heart rates	
X5	Oxygen consumption at the time of farm operation	

1.1. Classification and Regression trees (CART)

Classification and regression trees (CART), a nonparametric statistical procedure introduced by Breinmanet al. (1984) isa classifying method done by using decision tree technique. It is an umbrella term used for the following type of decision trees:Classification trees: where the target variable is categorical and nominal data and the tree is used to identify the class within which a target variable is likely to fall.Regression tree: where the target variable is continuous and tree is used to predict its value.The CART algorithm is structured as a sequence of questions, the answers to which determine what the next question, if any should be. The result of these questions are a tree like structure where the ends are terminal nodes at which point there is no more questions.In order to grow a classification tree, one need to know how to choose the conditions for splitting at each node, which criterion is to be used to split a parent node into its two daughter nodes, how to decide when a node becomes a terminal node and how to assign a class to a terminal node (Ratihet al., 2018).

The algorithm involved in CART is recursive partitioning algorithm. There are three steps in CART algorithm which described subsequently.

i. Constructing classification trees:

There are three different steps in constructing classification trees. A) First step is to select thesplitter. Selection of splitter is done based on one independent variable. For categorical variable, suppose that a particular categorical variable is defined by *L* distinct categories (levels) then there are'c'distinct splits for *L* level. For ordinal data, the number of possible splits (say *L*) at a given node is one fewer than the number of its distinctly observed values i.e. (*L* – 1). There is no best method for selecting the splitter. The commonly employed method used to select best splitter is Gini index. Other methods that are popularly used are Information index, node impurity function, Towing index and Entropy index. The form of Gini index is

$$I(t)=\sum_{i\neq j} p(i/t)p(j/t) \tag{2.2.1}$$

where $I(t)$ is the heteroscedasticity index of *t* node, $p(i|t)$ is proportion of class *i* in *t* node and $p(j|t)$ is the proportion of class *j* in node *t*. In the two-class case, this reduces to $I(t) = 2p\ (1\text{-}p)$ where we set $p = p(1|t)$. Then selecting the best splitterto generate classification tree is done by using Goodness of split criteria. Goodness of split *s* in *t* node is a base value and it is given by the decreasing of heteroscedasticity of a class by splitting the parent node t into daughter nodes t_L and t_R as

$$\phi(s,t)=\Delta I(s,t)=I(t)-p_L I(t_L)-p_R I(t_R) \tag{2.2.2}$$

where $\phi(s,t)$ is the Goodness of split value, p_L is the proportion of left node observation, p_R is the proportion of right node observation, $I(t_L)$ and $I(t_R)$ are heteroscedasticities of left and right nodes respectively. The best split is one that has the largest $\phi(s,t)$ value among the splitters since it has the capacity to decrease heterogeneity at most.According1y, the best splitter will show the important variables of classification trees. The variable score indicates the contribution of each variable for generation of classification tree and best splitter shows the biggest variable score.

Table.2: Splitter variation

Data scale	Ordinal	Nominal	Continuous
Number of samples	*N*	*N*	*n*
Number of levels	*L*	*L*	-
Splitter variation	*L-1*	$2^L - 1$	*n-1*

B) Second step is to determine the terminal node. A node will be terminal node if a node contains only one observation or reach minimum number of observations or if it reaches certain level of depth. C) Last step of CART analysis is class labelling for each terminal node. It is needed to characterize the classification result of each class based on response variable. One criteria is $p(j_0|t)=\max_j p(j|t)=\max_j \frac{N_j(t)}{N(t)}$, where j_0 is the class level for *t* terminal node. $N_j(t)$ is the number of observations of *j* class in *t* node and $N(t)$ is the total number of observations in *t* terminal node.

ii. Pruning:

The classification tree which we get is a maximal classification tree and it is of big size. The bigger the size of classification tree causes problem either due to overfitting or underfitting. So to avoid these we need an optimum classification trees. A better approach is to let the tree grow to saturation and then prune off branches until the tree is of optimum size. Pruned tree is a sub-tree of original tree. Pruning is done by cutting the node without decreasing the accuracy so that the size of the tree is not so complex and accuracy is still good. Now how to prune a tree is a crucial part of the process. There are many different ways to prune a tree.

The pruning algorithm is as follows: (Breimanet al., 1984)

1. Grow a large tree, say, $T_{\max}$, where we keep splitting until the nodeseach contain fewer than $n_{\min}$ observations.
2. Compute an estimate of $R(t)$ at each node $t \in T_{\max}$.
3. Prune $T_{\max}$ upwards towards its root node so that at each stage of pruning, the estimate of $R(T)$ is minimized.

To know which tree is good to prune, commonly used method is cost complexity method.This method uses complexity parameter denoted by α, and its value increases as long as pruning process is happening. The form used to calculate cost complexity function in T sub-trees from maximum classification tree ($T_{\max}$) with is

$$R_c(T) = R(T) + \alpha\left|\tilde{T}\right| \tag{2.2.3}$$

where, $R_c(T)$ is cost complexity measure or T^{th} tree complexity at α error value, $R(T)$ is resubtitution estimate or T^{th} tree error classification, α is complexity parameter and $\left|\tilde{T}\right|$ is number of terminal nodes on T^{th} tree.

iii. Selecting optimal classification trees :

The classification tree with too big size will cause bigger value of cost complexity. So we need to select an optimum tree which provides simple structure and small error. Choice of the best subtree depends upon having a good estimate of them is classification rate $R(T)$ corresponding to the sub-tree T. There are two estimation methods that can be used to select an optimum tree (Breiman et al., 1984):

a. Test Sample Estimation :

This method is used when the sample size is too large. In this method first the data set is divided into two parts i.e. training and testing data. Training data is used to make the form of the tree and testing data is used to estimate total error proportion of test sample estimate. The formula is given by

$$R^{ts}(T_t) = \frac{1}{N_2}\sum_{y_n, j_n \in L_2} X\left[d(y_n) \neq j_n\right] \tag{2.2.4}$$

where $R^{ts}(T_t)$ is the total error proportion of test sample estimate, N_2 is the number of observations of training data and $X\left(d(y_n) \neq j_n\right)$ is 0 if the statement in bracket is false, 1 when statement is true. To estimate total error proportion from this method, we can select optimum tree (T_t) with $R^{ts^*}(T_t) = \min R^{ts}(T_t)$.

b. Cross Validation Estimation :

This is in general also called *V-fold Cross validation Estimation*. This method is used when the data sample is small enough. The procedure is to divide the number of observations randomly to independently and same sized V fold. From this, one fold is used as training data to form a classification tree while other folds act as testing data. The formula used to do this V-fold cross validation estimation is given by

$$R^{cv}(T_t) = \frac{1}{V}\sum_{v=1}^{V} R^{cv}\left(T_t^{(v)}\right) \tag{2.2.5}$$

where $R^{cv}(T_t)$ is the total error proportion of V-fold cross validation estimation and V is the number of folds used. To estimate total error proportion from this method we can select optimum tree (T_t) with $R^{cv^*}(T_t) = \min R^{cv}(T_t)$.

2.3. Boosted CART

Boosting is one of the ensemble techniques in machine learning and widely used in classification and regression analysis. The idea of boosting method is to improve the weak learners sequentially and to increase the accuracy of

performance with a combined model. In literature, several boosting algorithms for solving regression and classification problems under various loss functions like Gradient boosting, AdaBoost (Adaptive Boost), XGBoost and others are available. One of the earliest and most popular boosting algorithms is Adaboost algorithm developed by Freund and Schapire (1997). Friedman et al. (2000) had developed variants of this Adaboost algorithm and traced it likes to logistic regression model. Later on Friedman (2001 and 2002) had proposed boosting can be think as a function-approximation problem and solved it by using forward gradient descent technique.

Considering the results of Friedman et al. (2000) and Dopkeet al. (2017), the output variable has been modelled as a binary variate, say, $y_{t+h} \in \{0,1\}$, where y_t is output value at time index t, $(t = 1,2,...)$ and h is the forecast horizon. The output variable is modelled with leading indicators, $\mathbf{x}_t = (x_{t,1}, x_{t,2}, x_{t,3}, ...)$, by a function $F(\mathbf{x}_t)$. It is used to minimize to expected loss function $\phi(.)$. Consider an exponential loss function (Dopke et al., 2017) as

$$\phi(F) = E\exp(-\tilde{y}_{t+h}F(\mathbf{x}_t)) \tag{2.3.1}$$

where $\tilde{y}_{t+h} = 2y_{t+h} - 1$ such that $\tilde{y}_{t+h} \in \{-1,1\}$ and E denotes the conditional expectation operator. The loss function $\phi(F)$ increases when $\tilde{y}_{t+h}$ and have a different sign and decreases when and have same sign. Let the conditional probability be , then equation (2.3.1) can be expressed as

$$\phi(F) = P(\tilde{y}_{t+h} = 1|\mathbf{x}_t) * \exp(-(1) * F(\mathbf{x}_t)) + P(\tilde{y}_{t+h} = -1|\mathbf{x}_t) * \exp(-(-1) * F(\mathbf{x}_t))$$

$$\Rightarrow \phi(F) = P(\tilde{y}_{t+h} = 1|\mathbf{x}_t) * \exp(-F(\mathbf{x}_t)) + P(\tilde{y}_{t+h} = -1|\mathbf{x}_t) * \exp(F(\mathbf{x}_t)) \tag{2.3.2}$$

as $\tilde{y}_{t+h}$ takes the value 1 and -1 and the loss function is minimized by defining function $F(\mathbf{x}_t)$ as a half of log-odd ratio as

$$F(\mathbf{x}_t) = \frac{1}{2}\log\frac{P(\tilde{y}_{t+h} = 1|\mathbf{x}_t)}{P(\tilde{y}_{t+h} = -1|\mathbf{x}_t)} \tag{2.3.3}$$

This Log-odds ratio is used to estimate the function $F(\mathbf{x}_t)$ using unconditional probability. However this unconditional probability is a crude measure of conditional probability. So boosting techniques show how this measure can be explained.

The following idea of boosting fits an additive model using a better loss function for classification that breaks down the function estimation problems into a series of simple problems by stipulating that $F(\mathbf{x}_t)$ can be expressed as the sum of simpler functions, $T(\mathbf{x}_t)$ as

$$F(\mathbf{x}_t) = \sum_{m=1}^{M} T_m(\mathbf{x}_t) \tag{2.3.4}$$

where, m denotes the index of weak learner, M is the upper bound on the functions considered in the simulation study. Here $F(\mathbf{x}_t)$ is called the strong learner and $T_m(\mathbf{x}_t)$ is the weak learner. The algorithm which is followed is the gradient-descent boosting algorithm to estimate the weak learners in a forward stagewise manner as

1. Initialize the algorithm: initializing weight w_t and $F_0(\mathbf{x}_t) = \frac{1}{2}\log\frac{P(\tilde{y}_{t+h} = 1)}{P(\tilde{y}_{t+h} = -1)}$
2. Define some upper bound M for the number of weak learners.
3. For $m = 1$ *to* M,
   a. Compute the negative gradient vector given by

      $$z_{t.m} = -\partial\phi(F)/\partial F(\mathbf{x}_t) = \tilde{y}_{t+h}\exp(-\tilde{y}_{t+h}F(\mathbf{x}_t))$$

      Note : taking partial differentiation with respect to $F(\mathbf{x}_t)$ rest will be constant.

      Hence the conditional expectation will not apply.

   b. Fit a weak learner, $T_m(\mathbf{x}_t)$, to the negative gradient vector.

   c. Update the function estimate, $F_m(\mathbf{x}_t)$, by adding the weak learner $T_m(\mathbf{x}_t)$ to $F_{m-1}(\mathbf{x}_t)$.

   d. Equipped with the new function estimate, go back to the step (a).

4. At $m = M$: Compute $F_M(\mathbf{x}_t)$ as the sum of weak learners, $T_m(\mathbf{x}_t)$, $m = 1, ..., M$

In our study, we used classification tree as a weak learner. Finally, the resulting stochastic gradient descent boosting algorithm requires only a sample, a subset from the data without replacement before fitting a weak learner. Only the sample data are then used to estimate the next weak learner.

3. Illustration

Accuracy measure :

To determine classification accuracy, Total Accuracy Rate is used. Total Accuracy Rate is the proportion of observations which predict as success given by

$$Total\ Accuracy\ Rate = \frac{total\ succes\ prediction}{total\ prediction} \qquad (4.2.1)$$

3.1. CART analysis

The classification method used in this study is CART.The selection of splitter is done by Gini index methodand the best splitter selection is done on the basis of Goodness of split value. In this study we have taken two sets of data, for first data set 70% data points from main data set as train data and rest 30% data as test data and for second 60% data points from main data set as train data and rest 40% data as test data. Since the number of sample data in this study is relatively small we used 10-fold cross validation estimation to select optimum classification tree.The first step of CART analysis is to form maximum classification tree.First the important splitter is selected based on Gini index value. On the basis of this, we have taken X3 variable as primary splitter or as parent node as X3 variable has maximum contribution. After determining the parent node we continued the splitting process as mentioned above to form the maximal classification tree. Based on this analysis we continued to get optimum classification trees with six terminal nodes.

Table.3: Relative importance of different variables (in %)

X3	X4	X5	X2	X1
39	25	25	11	1

Pruning and selecting optimum classification tree:

After forming maximum classification trees we have seen if the trees needed pruning or not. Main purpose of pruning is avoiding underfitting and overfitting. Pruning is done by using minimumrelative error and 10-folds cross validation estimate. Optimum classification trees is a tree withminimum relative error.

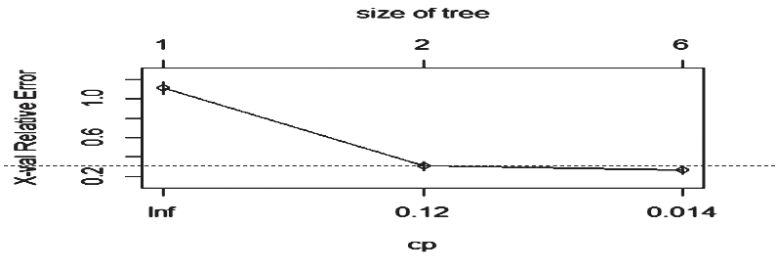

Fig.1: Graphical representation complexity parameter, relative error vs tree size.

Fig.1 shows the relative error plot of each pruning process. Since the tree innitially formed is optimum tree, it did not needed pruning.

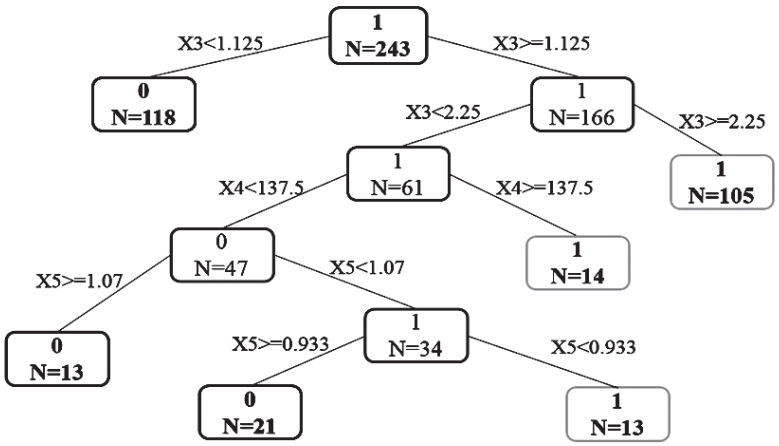

Fig.2: Optimum classification tree.

Table.4: Terminal node class labelling

Class	Terminal nodes	Total patients	Number of Absent	Number of Present
Absent	1	118	109	9
	2	13	9	4
	3	21	13	8
Present	4	13	3	10
	5	14	1	13
	6	105	0	105

CART analysis accuracy :

The data set was divided in a 70:30 ratio for training and testing purposes that means out of 405 observations,284 observations were used for training and rest 121 observations for testing. Based on Table.5, training data yielded Total Accuracy Rate 91.19% that means 91.19% successfully predicted on optimum tree and for testing total data accuracy rate was 85.95%. And for the second data set divided in a 60:40 ratio, for training and testing purpose that means out of 405 observations 243 observations are used for training and rest 162 observations for testing. Based on Table5, training data yielded Total Accuracy Rate 90.12% and for testing total data accuracy rate was 85.80%.

Table.5: CART analysis

Prediction for 70:30					
	Observation	Absent	Present	Total	Total Accuracy rate
Training data	Absent	131	21	152	91.19 %
	Present	4	128	132	
Testing data	Absent	58	13	71	85.95 %
	Present	4	46	50	
Prediction for 60:40					
	Observation	Absent	Present	Total	Total Accuracy rate
Training data	Absent	106	13	119	90.13 %
	Present	11	113	124	
Testing data	Absent	67	10	77	85.80 %
	Present	13	72	85	

3.2. Boosted CART analysis:

For Boosted CART analysis, the same data set was used.We used the R programming environment with the add-on package ''gbm'' for statistical computing for our empirical analysis, for estimating the Boosted CART model. The shrinkage parameter or learning rate wasassumed as the value $\lambda = 0.001$. We simulated this process 2500 times to make statistical inference, using 10-fold cross-validation to determine the optimal weak learners in every simulation run. In this respect, we fixed the maximum number of weak learners to M = 2500. From this analysis, we can find relative importance of leading indicators. The relative importance of a leading indicator Boosted tree is obtained by averaging across weak learners (Friedman, 2001). In this study, average across simulation has been runs.

Boosted CART analysis accuracy :

Table.6 shows the accuracy level of Boosted CART analysis for the training and testing data. Based on Table.6, it can be seen that for first data set, the accuracy for training data reacheda value of 95.42 % and for testing data reached a value of 87.60 %. And for second data set, the accuracy for training data reacheda value of 95.29 % and for testing data reached a value of 87.76 %.

Table.6: Boosted CART accuracy

Total Accuracy rate (%)	Boosted CART for 70:30	Boosted CART for 60:40
Training data	95.42	95.29
Testing data	87.60	87.76

3.3. Comparison of Accuracy of CART and Boosted CART Analysis:

Table.7 shows the information of accuracy of training testing data for Discomfort for the farm labourers during agricultural field operationusing CART and Boosted CART analysis. Based Table.7 we have found that for the first data set, the change in total accuracyrate for Boosted CART than conventional CART is as much as 4.63 % for training data and 2 % for testing data. Also for second data set, the change in total accuracy rate for Boosted CART than conventional CART is as much as 5.73 % for training data and 2.28 % for testing data.So we can conclude that the accuracy of the classification for Boosted CART is higher than that of CART analysis for both training and testing data. It implies that Boosted CART can classifies the observations more accurately than conventional CART.

Taboble. 7 : Comparison of accuracy

	Trtaining : Testing data (70:30)		Trtaining:Testing data(60:40)	
Total Accuracy rate (%)	CART	Boosted CART	CART	Boosted CART
Training data	91.19	95.42	90.13	95.29
Testing data	85.95	87.60	85.80	87.76

4. Conclusions:

Based on the analysis section we can conclude that:

1. CART analysis yields an optimum classification tree with 6 terminal nodes which 3 terminal nodes belong to class Absence of discomfort for the farm labourersand 3 terminal nodes belong to class Presence of discomfort for the farm labourers during agricultural field operation. For both the data sets, importance variable in classification tree is truly classified with accuracy of 91.19 % for training data and 85.95% for testing data and 90.73 % for train data and 85.80 % for test data respectively.
2. Boosted CART results in increasing the accuracy percentage, for first set of data, to 95.42% for training data and 87.60% for testing data and for second set of data, to 95.29% for training data and 87.76% for testing data.
3. The empirical result shows that Boosted CART outperforms conventional CART.

REFERENCES

Breiman, L., Friedman,J., Stone, C.J., and Olshen, R.A. (1984). *Classification and Regression Trees*. 1st Edn., Chapman Hall, New York, London.

Döpke, J., Fritsche, U. and Pierdzioch, C. (2017). Predicting recessions with Boosted regression trees. *International Journal of Forecasting*, ***33***(4), 745-759.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, **55**(1), 119-139.

Friedman, J., Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, ***28***(2), 337-407.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, **29**(5), 1189-1232.

Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, ***38***(4), 367-378.

Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis, *In Annual Meeting of the Society of Academic Emergency Medicine*. San Fransisco, California.

Ratih, I. D., Wulandari, S. P., Otok, B. W. and Nurai, Y. A. (2018). Comparison of CART and CART ARCING to Classify Household Food Security of Tuberculosis Patients in Surabaya. *Journal of Mathematics and Statistics*, **14**, 141-150.

Robinzonov, N., Tutz, G. and Hothorn, T. (2012). Boosting techniques for nonlinear time series models. *AStA Advances in Statistical Analysis*, **96**(1), 99-122.

Sadhu, S. K.,Ramasubramanian, V., Rai, A. and Kumar, A. (2014). Decision Tree Based Models for Classification in Agricultural Ergonomics. *Statistics and Applications*, **12**, 21-33.

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of Statistics*, **24**, 303-329.