A REVIEW: CLASSIFICATION BASED DECISION TREE INDUCTION

Shabana Pathan[1] , Sanjeev Kumar Sharma[2]

[1]Department of Computer Science & Engineering, Oriental University ,Indore, (M.P), India,sbn.pathan@gmail.com

[2]Department of Computer Science & Engineering , Oriental Institute of Science and Technology, Bhopal, (M.P), India, spd50020@gmail.com

Abstract- Due to their higher accuracy and performance, decision tree forests are quite popular. The decision tree is used to draw conclusions from the provided dataset, as suggested by its name. The idea behind the decision tree is that it aids in choosing suitable characteristics for subdividing the tree into portions. The splitting method utilised is ID3. The decision tree originated in the statistics area before being enhanced by several researchers in the fields of data mining, pattern recognition, and machine learning for the purpose of prediction. Algorithms like Decision Trees (ID3, C4.5, C5.0, and CART), Support Vector Machine (SVM), Neural Networks, Linear Regression, K-Nearest Neighbor (KNN), and others can be used to resolve issues relating to classification, regression, clustering, and optimization. It is crucial that the forecast be accurate. In this study, each method is described in detail and its efficiency and precision are compared. On the basis of datasets, a decision tree is built, and several decision tree induction parameters and prediction efficiency measures are presented.

Keywords: Machine Learning, Supervised, Classification, Decision Tree, Induction

DOI Number: 10.48047/nq.2022.20.19.NQ99235 ***NeuroQuantology2022;20(19): 2764-2769***

1. INTRODUCTION

Nowadays, technology has developed a lot, especially in the field of Machine Learning (ML), in which computers learn and behave like humans by feeding data and knowledge without being specifically programmed. Machine learning algorithms use statistical techniques to “learn" information directly from the data without a model equation.

Machine learning uses approach styles: Supervised learning that trains a model to predict future impacts of known input and output data and unsupervised learning that detects hidden trends in input data or intrinsic structures. Classification technique is a machine learning task that constructs models for class distribution using a set of predictive attributes characterized instances. Classification organizes the information into classes by using predetermined class labels. Using the training set data called class labels; the classification algorithms build a model named classifier. Classifier is used to predict unclassified object class labels within the testing information. Since a single classifier can only perform so well, ensemble classification based on ensemble learning is employed with many models to increase accuracy and achieve superior performance (Yan & Gobel, 2004). The many Classification techniques include Logistic Regression, k Nearest Neighbor (KNN), Support Vector Machine (SVM), Neural Network, Naive Bayes, and Decision Tree.

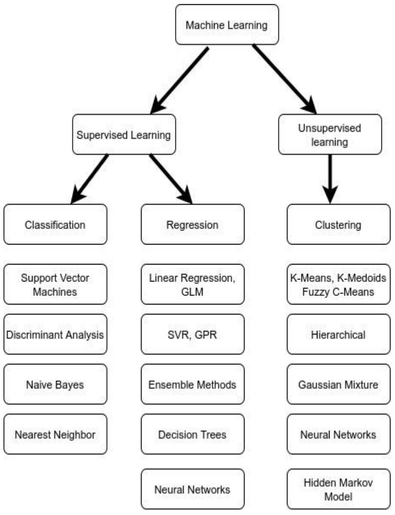


Figure 1. Types of Machine Learning

In this work, a thorough analysis of the most recent and effective methods used by researchers to study decision trees in various branches of machine learning is presented. Additionally, a summary is provided of how the decision tree is constructed the method's specifics, including the datasets used, algorithms/approaches used, and results obtained. Additionally, this analysis emphasized the most widely applied strategies and the most accurate techniques attained.

II.LITERATURE REVIEW

This section comprised of the research work carried out so far in the related field and briefly described in subsequent paragraphs.

Breimen (2001) [2] describes the hypothetical circumstances for haphazard forests, determining the split by the haphazard collection of features at each node. Ensembles of trees are generated using random vectors to enhance classification accuracy. Bagging trees are grown using unsystematic selection after the training set. For the haphazard split selection, the split is randomly chosen from among the best K splits at any node. Another technique is to select the training set of the examples within the training set from an irregular set of weights. Bagging can be used to enhance accuracy by using random features, also as estimates strength and correlation.

Tibshirani and Hastie (2007) [10] proposed the prime margin classifier to create an alternative tree with high-dimensional features for the sorting of more than two classes at each split.The margin-tree method planned seems compatible to high-dimensional problems with more than two classes having an evaluation, precision competitive multiclass support vector machines and nearby centroid methods, and provides a hierarchical grouping of the classes.

Mease et al. (2007) [9] presented an algorithm in combination with Over/under sampling & Jittering of the records that use AdaBoost. Classifiers gatherings are used from a countless network to make class probabilities to compare a number of approaches across through simulated and real data sets.

Fierens et al. (2010) [3] address the benefits & drawbacks of the six importance probability trees pruning criteria. An Experimental analysis of the comparative analysis of these pruning criteria is conducted and a randomization assisted pruning criterion performs perfectly due to 6 its extreme data features. Probability trees and decision trees are first analyzed by building and pruning a large tree. The Pruning is achieved by moving down the tree, thus removing subtrees that do not apply.

Hapfelmeier and Ulm (2012) investigate possible solutions for data containing missing values by imputation procedures, complete case scrutiny, and new recommended importance measure. Random Forests used for data prediction and interpretation have the ability to handle high-dimensional details, missing values, complex interactions, and collinearity indirectly. For the analysis of missing data, the predictive accuracy of Random Forests has been explored. The Hapfelmeier [4] [5] tests were compatible with and without the imputation of missing values and only showed minor variations between the values [5].

Kocsis et al.(2013) [8] projected multiple sequences of weak hypotheses, organized in a boosting tree and using some randomized process to add new weak hypotheses to promising tree nodes. With the theoretical results the efficiency of the base boosting algorithmic rule is achieved.

Amasyali and Ersoy (2013) [1] proposed the extended space forest method and compared it with the original ensemble algorithm. A decision tree is made, including all the original features and haphazard groupings, showing better results. Thus, by using all the input features and random selection, it provides more diversity and individual precision, but because of the additional options it takes more training time than the original algorithms. Generating less complex base learners reduces testing time.

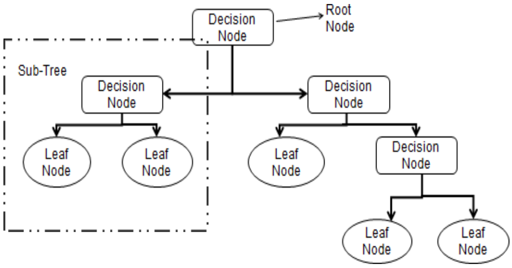
To overcome the problem of nonlinear decision rules Johnson and Zhang (2014) [6] suggested a framework for studying decision forests via greedy search using the underlying 7 forest structure. In many of the data sets higher precision and smaller models are obtained using this approach. The decision tree is a common method for easy handling of heterogeneous data easily when different features come from different sources. The boosted tree approach is considered as an additive model over multiple decision trees to improve predictive efficiency, resulting in a function as a decision forest.

III. DECISION TREE

3.1 Decision Tree Structure

Figure 1 typically illustrates a decision tree. Every internal node evaluates an attribute. A value for each branch is corresponding to an attribute. Each leaf node designates a category. The root node is the very first node. Leaf Nodes are the branches that come from the root node.

The decision forest (DF) model for both classification and regression, comprises of a huge collection of decision trees T, was first given by Breiman [1]. Decision forest is capable of handling dense, i.e. non-linear



relationships and various interactions among them, which is the reason that
it is popular for forecasting or speculations with the complex datasets [28].

Figure 1. Structure of Decision Tree

The optimal tree is a tree with leaf nodes and decision nodes. There are at least two branches on a decision node.There are at least two branches of a decision node. A classification or judgment is shown by the leaf nodes. On leaf nodes, we can't do further splitting.

The higher decision node in a tree that connects to the strongest predictor is known as the root node. Decision trees can handle both category and numerical data. By using a set of fundamental decision rules to categorize a large number of records into smaller sets, the Decision Tree is a hierarchical tree structure that can be used. A decision tree model is a collection of rules for segmenting a big diverse population into more predictable, homogenous, or exclusive groupings.

3.2. Types of Decision Tree Algorithms

1. Classification and Regression Tree (CART): Depending on the dependent variable, this dynamic learning technique may generate both a regression tree and a classification tree. Decision tree models for classification are those in which the goal values are discrete. A finite or countably infinite collection of values includes an age, size, or other discrete value. Regression models are generally numerical models with continuous values for the goal values. Floating-point variables are continuous variables. Together, these two models are known as CART. The Gini Index is the classification matrix used by CART.
2. Iterative Dichotomiser 3 (ID3): This algorithm chooses the characteristic to utilise to categorise the current subset of data using information gain. Information gain is determined iteratively for each tier of the tree for the remaining knowledge.The learning systems notion that E.B. Hunt, J., and Marin presented was expanded upon by this algorithm.
3.C4.5: This algorithm is the replacement for the ID3 algorithm. The categorization attribute is determined by this algorithm using either data gain or the Gain ratio. Since it can manage both continuous and missing attribute data, it is a straightforward improvement over the ID3 method. A greedy top-down approach is used by ID3 and C4.5 to construct decision trees.

Table II. Parameter Comparisonof Decision tree algorithm

Algorithms	Type of Data	Speed	Pruning	Missing Values	Measure	Procedure
ID3	Categorical	Low	No	Can't deal with	Entropy info-gain	Top-down decision tree construction
C4.5	Continuous and Categorical	Faster than ID3	Pre-Pruning	Can't deal with	Entropy info-gain	Top-down decision tree construction
C5.0	Continuous and Categorical, dates, times, timestamps	High	Pre-Pruning	Can dealwith	Entropy info-gain	Top-down decision tree construction
CART	Continuous and nominal attributes data	Average	Post pruning	Can dealwith	Gini diversity index	Constructs binary Decision tree

3.3 Working of Decision Tree

A decision tree uses a collection of rules that advance from its root hub in order to forecast the class for a given dataset. By comparing the values of the root property with those of the reference attribute, it goes from the branch to the node. In order to transmit them to the next node, a set of rules checks the attribute's value once again with the opposing sub-nodes. Up until the phase exceeds the leaf node of the tree, it is preserved.Using the following metrics, performance may be fully understood in terms of methodology:

1. To begin the tree containing the complete dataset, use the root node.
2. Pick out a quality characteristic from the dataset.
3. Divide the root node into several subgroups to incorporate the attribute value.
4. The quality characteristic is developed.
5. Create new decision trees iteratively using the subsets of the dataset generated in step 3 as a starting point. Continue using this technique up until it becomes impossible to further categorise the nodes, at which point you should declare a leaf node to be the last node.

The decision tree is constructed by taking into account the entropy criteria and Information gain to determine where to split the data and which feature must be assigned Root node.Higher the entropy lower is the purity of the node and such node is assigned as root node.

Step by step Calculations of Entropy & Information gain to construct Decision treee:
Step 1: “example” set RID
The set RID of 14 examples with 9 yes and 5 no then
Entropy (RID) = − (9/14) log2 (9 /14) - (5/14)log2 (5/14) = 0.940

Step 2: Attribute Income
Income value can be high, medium, and low
Income = high is of occurrence 4, 2 of the examples are “yes” and 2 are “no”
Income = medium is of occurrence 6, 4 of the examples are “yes” and 2 are “no”
Income = low is of occurrence 4, 3 of the examples are “yes” and 1 are “no”

Entropy (RID_Income_high) = - (2/4) x log2 (2/4) – (2/4) x log2 (2/4) = 1
Entropy (RID_Income_medium) = - (4/6) x log2 (4/6) – (2/6) x log2 (2/6) = 0.9173
Entropy (RID_Income_low) = - (3/4) x log2 (3/4) – (1/4) x log2(1/4) = 0.8112

Gain (Income,RID) = Entropy (RID) – (4/14) x Entropy (RID_Income_high) - (6/14) x Entropy (RID_Income_medium)- (4/14)Entropy (RID_Income_low)
= 0.940 – (4/14) x1 – (6/14) x 0.9173 – (4/14) x 0.8112
= 0.0293

Step 3:Attribute age
Age value can be youth,middle_aged,senior
age = youth is of occurrence 5, 2 of the examples are “yes” and 3 are “no”
age = middle_aged is of occurrence 4, 4 of the examples are “yes” and 0 are “no”
age = senior is of occurrence 5, 3 of the examples are “yes” and 2 are “no”

Entropy (RID_age_youth) = - (2/5) x log2 (2/5) – (3/5) x log2 (3/5) = 0.971
Entropy (RID_age_middle-aged) = - (4/4) x log2 (4/4) – (0/4) x log2 (0/4) = 0

Entropy (RID_age_senior) = - (3/5) x log2 (3/5) – (2/5) x log2(2/5) = 0.971

Gain (age,RID) = Entropy (RID) – (5/14) x Entropy (RID_age_youth) - (4/14) x Entropy (RID_age_middle_aged)- (5/14)Entropy (RID_age_senior)
= 0.940 – (5/14) x0.971 – (4/14) x 0 + (5/14) x0.971
= 0.246
similarily,
Gain (Loan_rating, RID) = 0.048

Table 2. Dataset for Buys_Computer

RID	age	income	Employee	Loan_rating	Buys_Computer
1	youth	high	no	poor	no
2	youth	high	no	good	no
3	middle_aged	high	no	poor	yes
4	senior	medium	no	poor	yes
5	senior	low	yes	poor	yes
6	senior	low	yes	good	no
7	middle_aged	low	yes	good	yes
8	youth	medium	no	poor	no
9	youth	low	yes	poor	yes
10	senior	medium	yes	poor	yes
11	youth	medium	yes	good	yes
12	middle_aged	medium	no	good	yes
13	middle_aged	high	yes	poor	yes
14	senior	medium	no	good	no

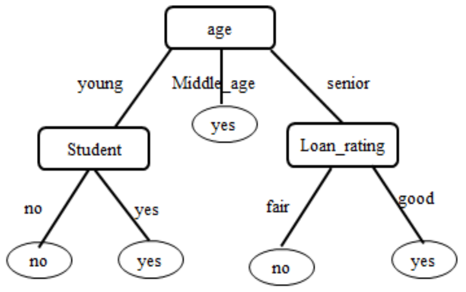


Figure 4 . Decision tree for concept buys_computer

IV. PROPOSED METHODOLOGY

In the classification, the decision forest is mostly used for finding out the probability of predictions, which falls into one of the 'C' possible group and these outcomes is used for the prediction or forecasting of events like weather forecasting, Stock prediction and many more.

Various methodologies have been proposed for the construction of decision forest and collecting outcomes from the individual learner till date but their scope is very limited. In the proposed research work, firstly, various decision forest methodologies will be investigated with a special focus on the aggregation of tree-based forecast.

In the Figure 5,We suggested two methods for building the decision tree forest. The first method is rotation decision tree forest, which will build the tree with exceptional accuracy and variety. Therefore, the goal is to use the transformation approach to shape the information to each node to more space in order to ascertain the optimal split at this node.

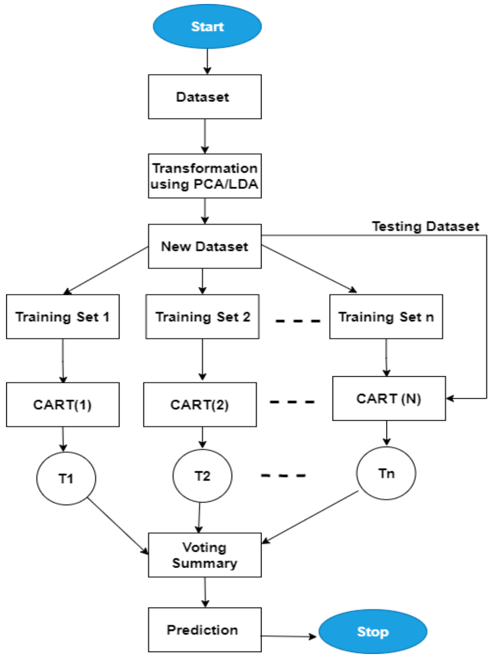


Figure 5 . Transformation based Improved Decision Tree

In the second approach Principal Component Analysis and Linear Discriminate Analysis techniques will be used to reduce the unwanted and undesired dimensions. In our proposed research work, the ensembles will be known as PDTF and LDTF i.e., Principal Component Analysis based Decision Tree Forest and Linear Discriminate Analysis based Decision Tree Forest respectively. Both LDA and PCA have exhibited improved performance in the areas of pattern detection, image processing, and ensemble analysis. Here, we suggested a unique method of combining Rotation Decision Tree Forests with the previously proposed Decision Tree Forests as an ensemble.

V.CONCLUSION

Researchers have a lot of options to pursue research in machine learning algorithms. It is an essential resource for discovering novel design links in huge datasets. Machine learning has improved several aspects of our life and various sectors by providing effective means to increase effectiveness.Decision tree methods like ID3 and C4.5 have been the most often used classification algorithms throughout time. To provide researchers with the work that has already been done and establish the groundwork for further study, this report studied and examined numerous of these improvements that are required. Future studies might look at how these enhanced algorithms have been used by academics and how they have performed in real-world circumstances. Furthermore, there hasn't been much research on the use of evolutionary algorithms for the feature selection. More research is needed in this area because the performance of the algorithms may be greatly enhanced by good feature selection in huge datasets and missing data issues.

References

[1] AMASYALI,M., AND ERSOY,O.2013.Classifier Ensembles with the Extended Space Forest, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26(3).

[2] BREIMAN,L.2001. Random Forests, *Machine Learning,* Vol. 5(1), pp. 5-32.

[3] FIERENS,D., RAMON,J., BLOCKEE,H., AND BRUYNOOGHE,M.2010.A comparison of pruning criteria for probability trees, *Journal of Machine Learning and Research*, pp.251-285.

[4] HAPFELMEIER,A., AND ULM,K.2012 Random Forest variable importance with missing data, Technical Report Number 121 ,Department of Statistics, University of Munich , [Online], Available: http://www.stat.un -muenchen.de.

[5] HAPFELMEIER,A., AND ULM,K.2013.Variable selection with Random Forests for missing data, Technical Report Number 137,Department of Statistics, University of Munich,15, [Online], Available: http://www.stat.uni -muenchen.de

[6] JOHNSON,R., AND ZHANG,T.2014.Learning Nonlinear Functions Using Regularized Greedy Forest, *IEEE Transactions on Pattern Analysis and Machine Intelligenc*e, Vol. 36(5).

[7] JIJO,B.T.,AND ABDULAZEEZ,A.M. 2021.Classification Based on Decision Tree Algorithm for Machine Learning, *Journal of Applied Science and Technology,*Trends,Vol.02, No.01, pp.20-28

[8] KOCSIS,L., GYORGY,A AND BAN.A.2013.Boosting Tree: parallel selection of weak learners in boosting, with application to ranking, *Journal of Machine Learning Research*.

[9] MEASE,D., WYNER,A., AND BUJA,A.. 2007. Boosted Classification Trees and Class Probability/ Quantile Estimation, *Journal of Machine Learning Research*, Vol. 8, pp. 409-439.

[10] TIBSHIRANI, R., AND HASTIE, T.2007.Margin Trees for High-dimensional Classification, *Journal of Machine Learning Research* ,Vol. 8, pp.637-652.

[11] YAN,J., ZHANG,Z., XIE,L., AND ZHU,Z.2019.A Unified Framework for Decision Tree on Continuous Attributes, *IEEE Access*, Vol.7.