Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 9 (Sep -2023)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2024.v8.i9.pp1-8

A Method for Vibration Testing Decision Tree-Based Classification Systems.

To Cite this Article

B.Siva, Ruhi." A Method for Vibration Testing Decision Tree-Based Classification Systems"
Journal of Science and Technology, Vol. 08, Issue 09 - Sep 2023, pp1-8

Article Info

Received: 28-08-2023 ***Revised: 05 -08-2023*** ***Accepted: 22-08-2023*** ***Published: 7-09-2023***

***Abstract*—** "Without any intervention from a person, computers are capable of "learning" new things by analyzing data in various ways (training and testing) and making conclusions. One application of ML is decision trees. Many diverse disciplines make use of decision tree techniques. These algorithms have a wide variety of potential applications, including search engines, text extraction, and companies that provide medical certifications. Decision tree algorithms that are both accurate and affordable are now at our fingertips. Whenever a choice is necessary, it is critical to know what the best choice is. We present three decision tree algorithms in this study: ID3, C4.5, and CART. We use tools like WEKA, ML, and DT.

I. EXPLORING THE DECISION TREE

Classification is the process of assigning things to categories, and it has many different uses.

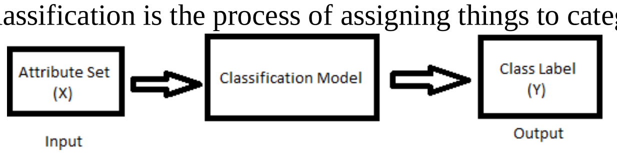

Fig. 1: Classification of mapping attribute set (X) to its classlabel (Y)

Decision Diagram

Trunk, branches, and leaves are the typical components of a tree. Decision Trees follow the same pattern. A tree's trunk, branches, and foliage give this structure its distinctive appearance. Attribute testing occurs at each leaf node [3, 4]. The test results are shown at the leaf node and continued down the branch. As its name implies, the root node is the initial node in a tree and acts as the biological parent to all the other nodes. According to [4], a "node" represents a quality or attribute, a "branch" a choice or rule, and a "leaf" a result, whether continuous or categorical. Decision trees provide quick data collecting and accurate conclusion drawing since they are designed around human thought processes. With the goal of processing a single result at each leaf, we want to build such a tree for all the data.

CONNECTED RESEARCH ON THE DECISION TREE

Decision Tree is straightforward as it attempts to simulate human decision-making process. Issues that persist independent of the data type (continuous or discrete). An example of a Decision Tree is shown here [15].

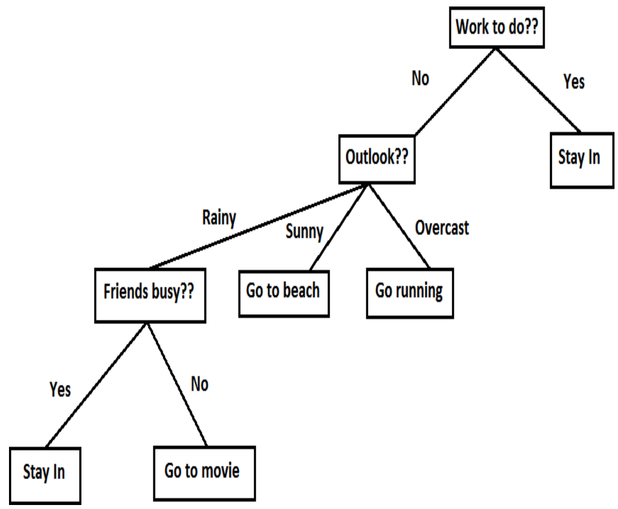

Fig. 2: Example of Decision Tree on what to do whendifferent situations occur in weather.

Splitting is instantly terminated if any data is deemed useless. Finding specific tests is more effective than trying to optimize the tree overall. Bear in mind that the data set only gives categorical information, and that the ID3 technique can only be simulated using the WEKA tool, when you analyze the properties of Decision Tree. Under ID3's simulation conditions, continuous data collecting is not possible. A number of parallels exist between CART and C4.5. features identical to those of ID3. C4.5 and CART are similar in that both may use continuous data sets as input for simulation purposes [11], but there is one key distinction.

Table-1: Characteristics of DT.

Decision Tree Algorithm	**Data Types**	**Numerical Data Splitting Method**	**Possib leTool**
CHAID	Categorical	N/A	SPSS answ ertree
ID3	Categorical	No Restriction	WEKA
C4.5	Categorical, Numerical	No Restriction	WEKA
CART	Categorical, Numerical	Binary Splits	CART 5.0

A One way to organize all the potential outcomes and the actions needed to get to each one is using a decision tree [12]. One of Decision Tree's strengths is how transparent and honest it is.You also

Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 9 (Sep -2023)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2024.v8.i9.pp1-8

get to choose the most biased and understandable nature, which is a huge plus. Both its classification and comprehension are within my capabilities. Able to effortlessly process info that is either continuous or discrete. The decision tree only needs to be able to segment features and filter variables [19]. Despite its impact on performance, non-linear makes no adjustments to the decision tree's parameters.

Part I: Decision-Based Algorithms Determining the "Best" way to split an attribute between two categories is possible using a decision tree approach. We need a consistent criteria for creating the splits if we want the partitions at each branch to be as pure as possible.

		datasets. The technique called “PRUNNING”, solves the problem of over- filtering [9].
C5.0	Improved version ofthe C4.5	C5.0 allows to whether estimate missing values as a function of other attributes or apportions the case statistically among the results [13].
CHAID (CHi-square Automatic Interaction Detector) [6]	Predates the original ID3 implementation.	For a nominal scaled variable, this type of decision tree is used. The technique detects the dependent variable from the categorized variables of a dataset [3, 11].
MARS (multi-adaptive	Used to find the best split.	In order to achieve the best

Table- 2: Decision tree algorithms

Algorithm name	**Classification**	**Description**
CART (Classification and Regression Trees)	Uses Gini Index as a metric.	By applying numeric splitting, we can construct the tree based on CART [4].
ID3 (Iterative Dichotomiser 3)	Uses Entropy function and Information gain as metrics.	The only concern with the discrete values. Therefore, continuous dataset must be classified within the discrete data set [5].
C4.5	The improved version on ID 3	Deals with both discrete as well as a continuous dataset. Also, it can handle the incomplete
regression splines)		split, we can use the regression tree based on MARS [2, 10].

Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 9 (Sep -2023)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2024.v8.i9.pp1-8

I. METRICS

Depending on the settings of the splitting attribute, several subsets of the training data are formed. Until every instance in a subset belongs to the same class in any Decision Tree, the technique iteratively iterates [6].

Table- 3: Splitting Criteria

Metrics	Equation
InformationGain	$Information\ Gain = I(p,n) = \left(\frac{-p}{p+n}\right)\log_2\left(\frac{p}{p+n}\right) - \left(\frac{n}{n+p}\right)\log_2\left(\frac{n}{p+n}\right)$
Gain Ratio	Gain Ratio=I(p,n)-E(A) I(p,n)= Information beforesplitting E(A)= Information after splitting
Gini Index	$Gini\ Index, G = \left(\frac{1}{2n^2\mu}\right)\sum_{j=1}^{m}\sum_{k=1}^{m} n_j n_k \lvert y_j - y_k \rvert$

The fundamental issue with Information Gain is its bias towards features that include several variables [6]. An example of unfair data partitioning would be a child node with an unusually large record count in comparison to the rest. An increased Gain Ratio is favored [7, 12]. When there are more than two groups in the data, the credibility of the Gini Index is weakened. Here are several problems with the splitting criteria [15].The fundamental issue with Information Gain is its bias towards features that include several variables [6]. An example of unfair data partitioning would be a child node with an unusually large record count in comparison to the rest. An increased Gain Ratio is favored [7, 12]. When there are more than two groups in the data, the credibility of the Gini Index is weakened. Here are several problems with the splitting criteria [15]. Information Gain prioritizes multivariate characteristics over univariate ones, which is a big problem [6]. In an unfair data partition, one of the child nodes contains a disproportionately high number of records compared to the others. Advantage is given to greater gain ratios [7, 12]. The Gini Index is rendered useless since more categories are included in the data. The following issues often arise while trying to divide criteria: [15].
When elements of a set are very closely packed together, we say that the set is exact. Measuring accurately involves taking an average of the amounts that have been measured and comparing it to the actual value of the variable. The only way to measure quantities with more than two terms is to use data points collected from several measurements of the same quantity [13].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

TP = True positive, TN = True Negative FP = False Positive, FN = False Negative

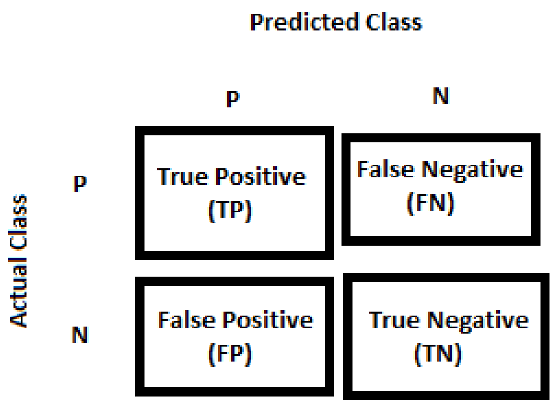

Fig. 3: Confusion Matrix sample in Decision Tree.

II. DESCRIPTION OF THE DATASET

The car dataset is used in this investigation. Running this dataset through the CART, ID3, and C4.5 decision tree algorithms. Here is how this dataset is defined.The car database is composed of two components. Automotive Technology and Popularity. The vehicle's acceptability is affected by a number of elements, including its purchase price and the cost to operate it. The amount of trunk space, safety features, expected passenger capacity, number of doors, and overall size of the passageway all have a role in crashworthiness.Of these, 1728 are examples. Six characteristics It is useless to have an attribute if it is not present.Excellence in Personality Traits:

Attribute	**Attribute Values**
buying	v-high, high, med, low
maint	v-high, high, med, low
doors	2, 3, 4, 5-more
persons	2, 4, more
lug_boot	small, med, big
safety	low, med, high

Class Distribution (Number of instances per class):

Class	**N**	**N [%]**

Unacc	1210	70.023%
Acc	384	22.222%
good	69	3.993%
v–good	65	3.762%

II. EXPERIMENT The experiment is being replicated using WEKA. The WEKA package of machine learning algorithms could be useful for data mining projects. Weka provides resources for pre-analysis data cleaning, organization, regressing, grouping, connection discovery, and visualization. Weka is freely available, open-source software that operates under the GNU Public License. Additionally, it may be used to create novel methods for machine learning. The algorithms may be called directly from Java code or executed on a dataset [18].

Table- 4: Theoretical results

Algorithm	**AttributeType**	**Missing Value**	**Pruning Strategy**	**Outlier Detection**
ID3	Only categorical values	No	No	Susceptible to outlier
CART	Categorical and Numerical both	Yes	Cost complexity pruning isused	Can handle
C4.5	Categoricaland Numericalboth	Yes	Error based pruning isused	Susceptible to outlier

In this study, three decision tree algorithms—ID3, C4.5, and CART—are tested on the same datasets to see how they do. Table [17] below provides a concise summary of the three approaches' results regarding runtime and accuracy. The splitting Criteria column details the algorithm's division strategy for performance improvement. In the attribute type column, you can see what sorts of data the algorithm can handle. The algorithm's performance may be evaluated by checking whether the Missing Value field is filled in.

Table- 5: Practical results

Algorithm	**Time Taken (Seconds)**	**Accuracy (%)**	**Precision**
ID3	0.02	89.35	0.964
CART	0.5	97.11	0.972
C4.5	0.06	92.36	0.924

The following table shows the realistic output of three algorithms: ID3, C4.5, and CART. The execution time for CART, ID3, and C4.5 is 0.5, 0.02, and 0.06 seconds, respectively. When compared to ID3, CART has the slowest execution time. In spite of being the slowest method, CART delivers the most precise results, making it the best option. It seems that CART is the best algorithm out of the three that were considered, according to the data presented in the table above.

Confusion Matrix:

```
=== Confusion Matrix ===

   a   b   c   d   <-- classified as
  35 363  27   3 |   a = vhigh
 361   4  60   6 |   b = high
 267  54  11 100 |   c = med
 237  41 107  47 |   d = low
```

Fig. 2 – Confusion matrix for ID3

```
=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 341  64  27   0 |   a = vhigh
 348  24  46  14 |   b = high
 261  37  48  86 |   c = med
 231  23  84  94 |   d = low
```

Fig. 3 – Confusion matrix for C4.5

```
=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 360  61  11   0 |   a = vhigh
 341  44  39   8 |   b = high
 268  41  57  66 |   c = med
 246  20  73  93 |   d = low
```

Fig. 4 – Confusion matrix for CART

III. CONCLUSION

Our decision tree methods of choice for this dataset were CART, ID3, and C4.5. For reliability, speed, and precision, decision trees are the way to go. The recommendation system is heavily relied upon by users to discover valuable material. After much discussion, the article's writers have settled on the conclusion that, when tested on this dataset, CART achieves the highest levels of accuracy and precision among decision tree methods.

REFERENCE

Yes, for research papers [1]. You sadden me, my love. A survey of the research on multi-label learning algorithms. Oregon State University in Corvallis, Oregon, on December 18, 2010. [2]. Akcayol Yuldiz O., Utku A., and Uke Karacan Utku. Decision-Tree-Based Recommendation System with Implicit Relevance Feedback Implementation. The citation is from the Journal of Safety and Health, volume 10, issue 12, pages 1367–14. Last updated on December 1, 2015. Contributed to by Gershman, Meisels, Lüke, Rokach, Schclar, and Sturm [3]. A Recommender System Based on Decision Trees. Article published in Volume 17, Issue 3, by the International Institute of Cognitive Science on June 3, 2010. I used "Jadhav SD" and "Channe HP" as references. A collaborative filtering decision tree classifier is an effective recommendation system. Published in 2016 by the International Journal of Engineering Research and Technology, volume 3, pages 2113–2118.

Everyone from Nürnberger and Genzmehr to Langer and Beel wrote a word or two here. We provide Docear's approach of recommending scholarly articles in this article. Presented at the ACM/IEEE-CS 2013 Digital Libraries Conference on July 22, 2013. The ACM, an organization for modern languages. Zhou Xiang and Jiang S. Decision tree learning using a similarity split-off criterion. As of 2012 Aug;7(8):1775-82, the paper was published in JSW. Mathuria, N., Bhargava, N., Bhargava, R., and Sharma, G. (2013). An investigation into data mining decision trees using the j48 algorithm. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, was released in June 2013 (IJARCSEE 2013). several approaches to serial decision tree categorization and how to compare them. [8] Article by Myanwu MN and Shiva SG published in the Journal of International Computer Science and Security, volume 3, issue 3, pages 230–140, June 2009.