Theoretical Study of Decision Tree Algorithms to Identify Pivotal Factors for Performance Improvement: A Review

Pooja Gulati
Research Scholar
Dept. of CS& IT
THE IIS University,
Rajasthan, India

Amita Sharma, PhD
Asst. Professor
Dept. of CS& IT
THE IIS University,
Rajasthan, India

Manish Gupta, PhD
Project Director (OSD)
(Former), Dept. of Science
&Tech., Govt. of Rajasthan,
Rajasthan, India

ABSTRACT

Decision tree is a data mining technique used for the classification and forecasting of the data. It is the supervised learning algorithm that follows the greedy approach and works in a top down manner. Decision tree uses white box model approach and classifies the data in a hierarchical structure. It makes data easy to represent and understand. It can handle a large database and works well with both numerical and categorical variables. A variety of decision tree algorithms are proposed in the literature like ID3 (Iterative Dichotomiser 3), C4.5 (successor of ID3), CART (Classification and Regression tree), CHAID (Chi-squared Automatic Interaction Detector). These algorithms have specific mechanisms based on certain criteria's. The study of these criteria are important and requisite for analysis of DT algorithms. The aim of this paper is to identify and inspect these vital criteria's or factors of DT algorithms. The major contribution of this review is to provide a path to select a specific factor for improvement of DT algorithm as per requirement or problem.

Keywords

Data Mining, Decision Tree Technique, Decision Tree Methodology, Decision Tree Factors.

1. INTRODUCTION

Decision tree is a technique for representation of data in hierarchical manner. There are two approaches that decision tree follows like top down or bottom up. As name reveals, its purpose is to help in decision making process. There are various algorithms that work for decision tree like ID3, C4.5, C5, CART, CHAID etc.

A decision tree is a graph that uses a branching method to illustrate every possible outcome of a problem. The tree structure shows how one choice leads to the next, and the use of branches indicates that each option is mutually exclusive. A decision tree can be used to clarify and find an answer to a complex problem. Decision tree can work on any kind of dataset and can handle conditional information by dividing the dataset into sub groups. These sub groups are further considered as an individual dataset for detail processing in decision tree methodology.

There are various applications of decision tree are available such as education, health care, finance etc. and many software packages provide implementations of decision tree algorithms for ex. Salford Systems CART (which licensed the proprietary code of the original CART authors), IBM SPSS Modeler, Rapid Miner, SAS Enterprise Miner, Mat lab, R (an open source software environment for statistical computing which includes several CART implementations such as rpart, party and random Forest packages), Weka (a free and open-source data mining suite, contains many decision tree algorithms), Orange (a free data mining software suite, which includes the tree module orng Tree), KNIME, Microsoft SQL Server, and scikit-learn (a free and open-source machine learning library for the Python programming language).[10]

There are various types of Decision tree algorithms are available in the field of data mining but the major drawback of decision trees technique is that it behave differently with distinct types of data sets. Some data are fitted well with algorithms and some are not. That's why there is a continuous enhancement needed in the field of decision tree generation. This paper will assist the user to analyze and select a specific factor for understanding their functioning and performance to make any kind of modifications according to requirement or problem.

The organization of this paper is as follows. Section 2, describes concept of decision tree and types of decision tree, Section 3 explains basic decision tree algorithms. The basic decision tree factors and their detail study are presented in section 4. The paper ends with conclusion and contribution described in section 5.

2. CONCEPT OF DECISION TREE

The Decision Tree method is remarkable data mining technique [9]. It is a flowchart like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes or class distributions [7]

Decision Tree signify rules that can be manifest by users used in databases for knowledge discovery. They are used to predict the relationship of elements to different classes or categories by looking at the values that correspond to their attributes or predictor variables.

In data mining, a decision tree techniques are employed for either classification & prediction or classification & regression. Two types of decision can be generated in tree pattern: classification tree and regression tree shown in Table 1.

Table 1. Types of Decision tree:

Classification tree	Regression Tree
It classifies dataset into different classes.	It classifies dataset for some range of real numbers.

Mainly Used for categorical target variables.	Mainly used for numeric and continuous target variables.
Use information gain for splitting dataset	Use standard deviation method for splitting dataset
Calculate Entropy or Gini Index for splitting to find homogeneity in dataset.	Calculate lowest standard square error for splitting dataset.

2.1. Creation of DT & Why DT

DT creation is a peculiar process and encapsulated following key points:

- How to select the value of each node?

- What are the splitting criteria used?

- When to use decision rules.

- Where to stop (Leaf nodes).

Decision tree works well for classification and prediction models. While classifying the dataset decision tree depends on response variables or resulting variable shown in figure 1. If response variable has two categories, then standard classification tree works well here. If response variable has multiple categories, then C4.5 classification tree algorithm. At the time of prediction of the dataset if response variable is continuous and also if there is any linear relationship between predictors and response variables then use regression tree algorithm and for non-linear relationship use C4.5 tree implementation.
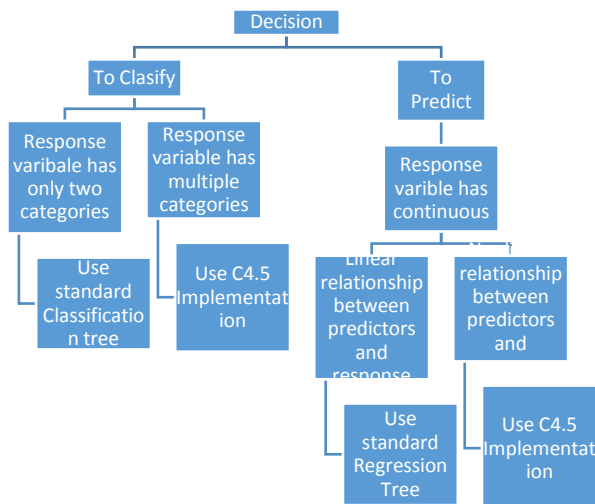

Figure 1: Decision Tree for Classification and Prediction

2.2 Limitations of Decision Tree Techniques

Although, DT is comprehensive approach, but it suffers with several shortcomings. The major limitations are:

- Modification in decision tree is not an easy task.

- If output variable or target variable is not defined appropriately then inappropriate result can be found.

- Decision tree with many branches can result in more complex and time consuming process.

- Large trees are difficult to interpret and also result in high cost.

- Can't give appropriate result for continuous values.

3. DECISION TREE ALGORITHMS

DT algorithms construct decision tree from a dataset. These tree are created in two ways top-down and bottom-up. There are various top–down decision trees algorithms such as ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), CART (Breiman et al., 1984). Some consist of two conceptual phases: growing and pruning (C4.5 and CART) and some algorithms perform only the growing phase [8]. After producing tree in top-down manner pruning occur in bottom-up manner by using pruning methods or algorithms to solve over-fitting.

3.1. Id3

ID3 (invented by J. Ross Quinlan in 1979) is a machine learning algorithm or mathematical algorithm for constructing decision tree. It constructs tree in a top down manner without backtracking. It classifies the dataset by using entropy and information gain to select root node for classification. ID3 can make fasted and short decision tree. Limitation of ID3 is for small sample, data may be over-fitted or over-classified. It tests only one attribute at a time. Generally, Complexity issues areas where classifying continuous data. It doesn't support pruning and can't handle missing values.

3.2. C4.5

C4.5 is an extension of ID3, also developed by Quinlan. It is designed to handle noisy and missing data better than ID3. Pre pruning and post pruning of decision tree is possible. It can handle attributes that have continuous values. In order to handle continuous attributes, C4.5 creates a threshold and then splits the list into those whose attribute value is above the threshold and those that are less than or equal to it. C4.5 allows attribute values to be marked as "?" where Handling training data with missing attribute values. Missing attribute values are simply not used in gain and entropy calculations. [2] Decision rules can be applied to modify variable or training dataset. The basic limitation of c4.5 is that it gives very poor results when large distinct values are used in both continuous and discrete attributes [11].

3.3 Cart

CART states for classification and regression tree. Classification is referred to when the target variables categorical, whereas regression is used for continuous values of the target variable (infinite number of values) [1]. It produces only binary decision trees. CART use Gini Index to select attribute for making decision tree just like entropy used by ID3 and C4.5. By changing the splitting value (increase complexity) cart ay give unstable tree. It split dataset by only one variable.

3.4 Chaid

Chi-Squared Automatic Interaction Detection (CHAID)is developed by Kass in 1980.CHAID can be used for prediction (in a similar fashion to regression analysis, this version of CHAID being originally known as XAID) as well as classification, and for detection of interaction between variables [6].CHAID can only work with categorical data. This algorithm uses Chi-Squared test between response or target variable (dependent) and input variables (independent) and on the basis of result of Chi–Square test it chooses the

best splitting variable from the input variables. Limitations of CHAID algorithm is, it can only work with categorical variables and can't handle pruning methods and cross validation.

There are various kind of decision tree algorithms that can generate different kind of decision tree such as classification tree, regression tree, decision tree forest etc. One can differentiate these algorithms by focusing on some important factors of a decision tree. For example, ID3 and C4.5 generate decision tree using some information-theoretic measure such as entropy, information gain, information gain ratio and CART and CHAID generate decision tree by using statistical method such as gini index, chi-square test.

4. DECISION TREE FACTORS

There are different criteria's and factors that are responsible for difference in working of various DT algorithms. This section outlines these factors. *Before discussing those factor, it is essential to learn about basic of decision tree creations.* There are four fundamental points related to generation on decision tree:

- The *splitting criterion* i.e. how "variance" is calculated. It works on the basis of which is the best attribute to split. There are different techniques are used by distinct algorithms on the basis of their datasets for ex. Entropy, Information Gain, Information Gain Ratio, Gini Index etc.

- *Model generation* i.e. the data types of attributes taken for model building whether it builds models for *regression* (continuous variables, e.g., a score) as well as *classification* (discrete variables, e.g., a class label)

- *Over-fitting* i.e. technique to eliminate/reduce over-fitting i.e. it used to reduce misclassification of data by using pruning methods.

- *Incomplete data i.e. whether algorithm* work well with unknown and noisy data. How it treats unknown values.

4.1 Factors

Decision tree is a systematic approach where information seeking and inference drawing are imperative tasks. The DT algorithm captures data and mine it by employing educative id of classification

4.1.1 Entropy

Entropy is the quantitative measure of disorder in a system [5]. It is used to calculate to find homogeneity in the dataset to divide dataset into several classes. If resulting class contain similar data, then entropy will be zero and if resulting class equally divide into two datasets then entropy will be one. On the other hand, it measures the impurity of dataset that means higher the entropy value result in more information content. Entropy is used to measure how informative is a node. [3]

$$Entropy = \sum_{i=1}^{c} p_i \log p_i$$

Entropy used for:

- When node belongs to only one class, then entropy will become zero.

- When disorder of dataset is high or classes are equally divided then entropy will be maximal.

- Helps in making decision at several stages.

4.1.2 Information Gain

Information Gain is used to measure purity or homogeneity in the dataset. Information gain describes how data can be distributed in respect of response variable. If information gain is high, then that variable is more informative and can be consider for root node. Information gain split the dataset based on an attribute that result in lesser entropy. In decision tree information gain find out the highest information gain attribute which divide the dataset into more homogeneous classes or sub-dataset.

Information Gain = Entropy of a Target Variable (Entropy before Split) - Entropy of a Target Variable in Respect to an Input Variable (Entropy after Split)

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i)$$

with:

{S1, …., Si, …., Sn} = partition of S according to value of attribute A

n = number of attribute A

|Si| = number of cases in the partition Si

|S| = total number of cases in S

4.1.3 Information Gain Ratio

Information gain doesn't fit for high branches attributes.info gain ratio reduce the bias of info gain. The resulting gain ratio value is large when data is equally divided and it's small when data belongs to one class. While calculating gain ratio it also takes number and size of branches of an attribute. It removes the bias of info gain by taking the intrinsic information of split of an attribute.

$$IntrinsicInfo(S, A) = - \sum \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

If intrinsic value of an attribute is high, then that attribute is less informative or less important. Less intrinsic values of an attribute make attribute more important.

$$GainRatio(S, A) = \frac{Gain(S, A)}{IntrinsicInfo(S, A)}$$

Figure 2. Theoretical study of pivotal factors of a decision tree

S. No.	Factors	Variables	Algorithms	Methods	Splitting Criteria	Type of Tree	Missing Value	Noisy Data	Analysis	Performance	Pruning Support
1	Entropy	Continuous Discrete Categorical	ID3 and C4.5	Exploratory Analysis (Impurity Measure)	Grouping of Classes that include up to 50% of the Data	---	---	---	---	Little slower to compute	---
2	Information Gain	Categorical, discrete or Nominal	ID3 and C4.5	Use Entropy for Analysis	Split dataset into large number of partitions	Consider Binary Split	Can't handle missing values	Can't handle noisy Data	Biased towards multivalued attributes	---	Cannot handle Over Fitting
3	Information Gain Ratio	Numerical Continuous Discrete and Nominal	C4.5	Use Information Gain and split information value or Intrinsic Information for Analysis	Prefer unbalanced splits in which one partition is much smaller than the other	Consider Multivalued Split	Can handle Missing Values	Can handle noisy Data	Reduce Biased towards multivalued attributes	---	Reduce Over Fitting by using Pruning Methods
4	Gini Index	Discrete Numerical continuous	CART	Use Weighted average of each branch index (Impurity Measure)	Grouping of largest class from other classes (Twoing Criteria)	Consider Binary Split	Can handle Missing Values	Can handle noisy Data	Biased towards multivalued attributes	Little Faster to compute	Reduce Over Fitting by using Pruning Methods

4.1.4 Gini Index

Gini index is used by CART. It is an impurity measure of dataset. It's an alternative of Information Gain.

$$Gini(S) = 1 - \sum p_i^2$$

Gini index is an impurity-based criterion that measures the divergences between the probability distributions of the target attribute's values. [4]

$$Gini(S,A) = \sum_i \frac{|S_i|}{|S|}.Gini(S_i)$$

(Average gini index in respect of attribute A)

4.2 Study of Decision Tree Factors

Figure 2 shows the study of different pivotal factors of decision tree algorithms in tabular form.

We can see from above table that Entropy is the most basic factor that can measure impurity of the data set for classification task. On other hand Gini Index also measure impurity of the dataset, difference is that entropy works on attribute that occur in classes and Gini is for continuous attributes. Entropy use exploratory Analysis that help in find groups of classes that makes up to 50% of the data and Gini minimize misclassification that helps to find largest class from other classes. Most of the decision tree algorithm use entropy for calculating other factors like information gain and information gain ratio. Entropy used by ID3 and C4.5 decision tree algorithms and Gini Index used by CART Decision Tree algorithms. Other basic factors are information Gain (IG) and Information Gain Ratio. IG used for small number of distinct values but not for continuous values and information gain ratio used for large number of distinct value also for continuous variable. Algorithm that use IG (information Gain) can't handle Noisy and Missing data but with Information Gain Ratio it can possible to handle those data by using pruning methods. IG and

Gini index consider binary split and IG ratio consider multivalued split.

4.2.1 Working Criteria of DT factors

This section describe DT factors by an example to see the working criteria of all the factors. Table 1 shows "weather data set", will used to explain the functioning of these factors.

Table 1: weather Dataset

Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Factors that we are taking in account are Entropy, information gain, information gain ratio and gini index. **"Entropy"** is used to calculate homogeneity of the dataset. If dataset is completely homogeneous the entropy will be zero else if dataset is equally divided that entropy will be one as shown in figure 3.
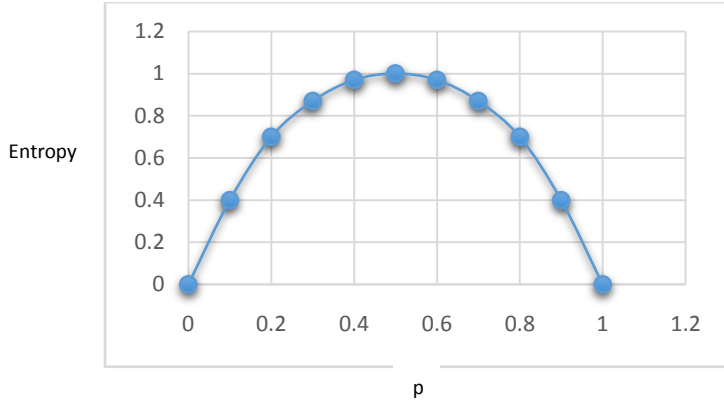
$$Entropy = -0.5\, log_2 p - 0.5\, log_2 0.5 = 1$$

Figure 3. Entropy Graph

For decision tree construction first calculate entropy of target variable i.e. "Play Golf". To calculate entropy, formula use probability of classes occurs in that attribute as shown in figure4. Calculate entropy of the target variable i.e. play golf. This is also referred as entropy before split, *Entropy (T)*

$$E(S) = \sum_{i=1}^{c} -p_i\, log_2\, p_i$$

Play Golf	
Yes	No
9	5

$Entropy(PlayGolf) = Entropy(5,9)$

$$= Entropy(0.36, 0.64)$$

$$= -(0.36\, log_2 0.36) - (0.64\, log_2 0.64)$$

Figure 4. Entropy Result of Play Golf

Than calculate entropy of target attribute with respect to other attribute or predictors to find homogeneity of attribute with respect to target attribute as shown in figure 5. This is referred as entropy after split, *Entropy (T,X)*.

$$E(T,X) = \sum_{c \in X} P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

$E(PlayGolf, Outlook)$
$$= P(Sunny) * E(3,2) + P(Overcast)$$
$$* E(4,0) + P(Rainy)$$
$$* E(2,3)$$

$$= \left(\frac{5}{14}\right) * 0.971 + \left(\frac{4}{14}\right) * 0.0 + \left(\frac{5}{14}\right)$$
$$* 0.971$$
$$= 0.693$$

Figure 5. Entropy after spilt "Outlook"

"Information gain or Gain" uses entropy to find out most homogeneous branches. To calculate information gain entropy before split is subtracted from the entropy after split. At the time of construction of decision tree dataset can be split on any attribute. To select best attribute to split, information gain value is used. Attribute which have highest information gain, select as a root node.ID3 use information gain to construct a decision tree.

$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain = 0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain = 0.048			

Figure 6. Entropy after split "Outlook, Temp. Humidity, Windy"

As shown in figure 6, Outlook attribute has the highest information gain and will select as a decision attribute in the root node of the tree shown in figure7.
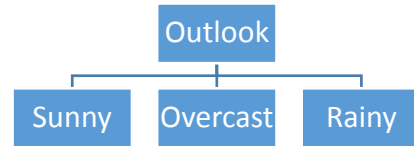
Figure 7. Outlook as root node

Dataset is now divided into sub dataset according to the branches of root node and same criteria will run in recursive manner for sub datasets till entropy become zero. A branch with entropy "zero" is a leaf node and branch with entropy more that "zero" needs further splitting. For example, sub dataset for branch "sunny" where entropy will be more that zero need further split into sub branches see figure8 and for branch "overcast" entropy will be zero because resulting variable is completely homogeneous or belonging to only one class of target variable i.e. "YES" set as a leaf node see figure 9.

Temp	Humidity	Windy	Play Golf
Mild	High	False	Yes
Cool	Normal	False	Yes
Mild	Normal	False	Yes
Cool	Normal	True	No
Mild	High	True	No
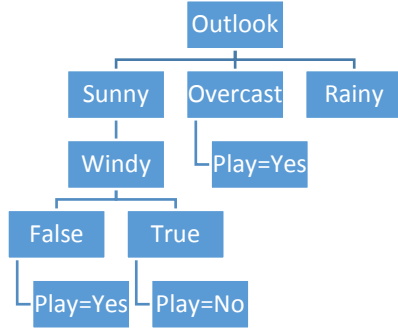
Figure 8. Sub dataset "Sunny"

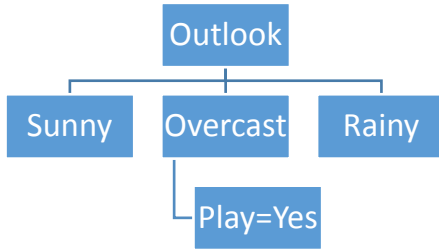| Temp	Humidity	Windy	Play Golf
Hot	High	False	Yes
Cool	Normal	True	Yes
Mild	High	True	Yes
Hot	Normal	False	Yes
Hot	High	False	Yes

Figure 9. Sub dataset "Overcast"

Drawback of "information gain" is that it is biased towards test with many outcomes and sometime produce one per partition and it cannot split numerical attributes. To overcome this drawbacks, ratio of split information value or intrinsic value is used with information gain value and new formula referred as "**Information Gain Ratio**" (used in C4.5 decision Tree algorithm). Split information give value that generated from partition of dataset according to distinct classes. High split information shows that all partitions are of almost same size. Low split information value shows few partitions hold most of the tuples of the dataset.

$$SplitInfo(p, test) = -\sum_{j=1}^{n} p'\left(\frac{j}{p}\right) * \log\left(P'\left(\frac{j}{p}\right)\right)$$

Gain ratio, is defined as follows:

$$GainRatio(p, T) = \frac{Gain(p, T)}{SplitInfo(p, T)}$$

The figure10 shows the resulting value of information gain ratio. Attribute which have highest gain ration will set as root node.

Outlook		Temperature	
Info:	0.693	Info:	0.911
Gain: 0.940-0.693	0.247	Gain: 0.940-0.911	0.029
Split info: info([5,4,5])	1.577	Split info: info([4,6,4])	1.557
Gain ratio: 0.247/1.577	0.157	Gain ratio: 0.029/1.557	0.019
Humidity		Windy	
Info:	0.788	Info:	0.892
Gain: 0.940-0.788	0.152	Gain: 0.940-0.892	0.048
Split info: info([7,7])	1.000	Split info: info([8,6])	0.985
Gain ratio: 0.152/1	0.152	Gain ratio: 0.048/0.985	0.049

Figure 10. Information Gain, Split Info and Gain Ratio Result

For regression tree the formula **"Gini Index"** is used.it consider binary split.

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

Subset which give minimum Gini index is selected as splitting attribute. For example, first compute the gini index of target attribute "play Golf"

$$Gini(play\ golf) = 1 - \left\{\left(\frac{9}{14}\right)^2 + \left(\frac{5}{14}\right)^2\right\} = 0.459$$

Now calculate Gini with respect to attribute temperature which have three classes i.e. cool, mild and hot. To compute gini index value pick first partition that belongs to classes (cool, Mild) which include 10 tuples and second partition belongs to class (hot) which have 4 tuples.

$$Gini_{temperature\ \in\{cold,mild\ \}}(D)$$
$$= \frac{10}{14}Gini(D_1) + \frac{4}{14}Gini(D_2)$$
$$= \frac{10}{14}\left\{1 - \left(\frac{6}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right\} + \frac{4}{14}\left\{1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2\right\}$$
$$= 0.450$$
$$= Gini_{temperature\ \in\{hot\}}(D)$$

Like above calculate Gini Index value for remaining partitions:

$$Gini_{\{cool,hot\}and\ \{mild\ \}}(D) = 0.315$$
$$Gini_{\{mild,hot\}and\ \{cool\ \}}(D) = 0.300$$

Therefore, the best binary split for attribute temperature is on {mild, hot} and (cool).

Factors explained above are used by standard decision tree algorithms ID3, C4.5, CART. They give satisfied result but they also have some limitations i.e. Information gain is biased towards multivalued attribute, Gain ratio sometime create unbalanced split like one partition is much smaller than other, Gini index has difficulties when number of classes is large etc. To understand these problems, this paper focuses DT factor which will help user to understand the working criteria of these factor for further analysis or modifications.

5. CONCLUSION

This paper shows some of the key factor of decision tree algorithm i.e. Entropy, Information Gain, Information Gain Ratio and Gini Index. These factors help user to analyze how decision tree works. Entropy is an important factor that can work with both continuous and discrete variables. It is the basic and essential factor used to calculate Information Gain and information Gain Ratio incorporated in ID3 and C4.5 decision tree algorithms. It helps to generate both binary and multivalued tree form. It is used to maximize mutual information and further involved in classification task while Gini Index used by CART is to minimize misclassification probability and further involved in regression analysis.

Sections explained above, shows how the basic and pivotal factors of decision tree algorithms works which is involved in decision tree generation. This paper assist a researcher to recognize these factors to select and focus on specific part of decision tree algorithm for further modifications and improvements.

6. REFERENCES

[1] Chady El Moucary,2011, "Data Mining for Engineering Schools Predicting Students' Performance and Enrollment in Masters Programs", International Journal of Advanced Computer Science and Applications, Louaize (NDU), Vol. 2, No. 10, 2011

[2] N. Suneethaet.et.,2010., MODIFIED GINI INDEX CLASSIFICATION: A CASE STUDY OF HEART DISEASE DATASET, International Journal on Computer Science and Engineering,Vol. 02, No. 06, 2010, 1959-1965

[3] Kusrini and Sri Hartati,2007," Implementation of C4.5 algorithms to evaluate the cancellation possibility of new student applicants at stmikamikomyogyakarta." Proceedings of the International Conference on Electrical Engineering and Informatics InstitutTeknologi Bandung, Indonesia June 17-19, 2007, 623-626, ISBN 978-979-16338-0-2

[4] LiorRokach and OdedMaimon, DECISION TREES, Chapter 9, online source: http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf

[5] http://physics.about.com/od/glossary/g/entropy.htm

[6] http://en.wikipedia.org/wiki/CHAID

[7] H. Jiawei and K. Micheline, Data Mining: Concepts and Techniques, vol. 2, Morgan Kaufmann Publishers, 2006.

[8] LiorRokach and OdedMaimon, DECISION TREES, Chapter 9, Data Mining and Knowledge Discovery Handbook online source: http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf

[9] Florin Gorunescu, Data Mining: Concepts, Models and Techniques, Intelligent Systems Reference Library, Vol. 12, Springer Publication, 2011.

[10] https://en.wikipedia.org/wiki/Decision_tree_learning#Im plementations

[11] S.Santhosh Kumar, Dr.E.Ramaraj, Modified C4.5 Algorithm with Improved Information Entropy and Gain Ratio International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 9, September – 2013