Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An Efficient Modified ID3 Decision Tree Algorithm for Data Mining Using S-T Entropy

Arvind Kumar
IT Group, CSIR-NGRI,
Hyderabad, India

Shiva Soni
CSE, JNTU-ATRI,
Hyderabad, India

G. Rama Seshagiri
IT Group, CSIR-NGRI,
Hyderabad, India

***Abstract**—Data mining is the process of automatically discovering useful information in large data repositories. Data mining techniques deployed to scour large databases in order to find novel and useful patterns that might otherwise remain unknown. The classification techniques help to learn a model from a set of training data and to classify a test data well into one of the classes. It can be described as supervised learning algorithm as it assigns class labels to data objects based on the relationship between the data items with a pre-defined class label. Decision tree is an important method for both induction research and data mining, which is mainly used for model classification and prediction. A very simple decision tree approach which is ID3 algorithm proposed by Quinlan. ID3 uses Information Gain as Splitting criteria. In proposed work we have used the ID3 decision tree algorithm of data mining along with combining the S-T entropy* [4, 5] *instead of Shannon entropy. By computing information we set particular property from the data take as root of tree, also sub-root by repeating the process continually, and finally build the most optimized tree. We have also compared the results with the H-C entropy [11] of single parameter 'α' .The formation of decision tree by using the above modified approach may help in taking the better decision to analyze the data like weather data, medical data, and seismic data collected from different resources.*

***Keywords:** Data mining, decision Tree, Shannon entropy, S-T entropy, H-C entropy, ID3 algorithm.*

I. INTRODUCTION

The last decade has experienced a revolution in information availability and exchange via the internet. In the same spirit, more and more businesses and organizations began to collect data related to their own operations. While the database technologists have been seeking efficient means of storing, retrieving and manipulating data, the machine learning community has focussed on developing techniques for learning and acquiring knowledge from the data. At times the data can be considered to be a gold mine for strategic planning for research and development in this area which is often referred to as *Data Mining* (DM) and *Knowledge Discovery in Databases* (KDD) [1, 2, 6].

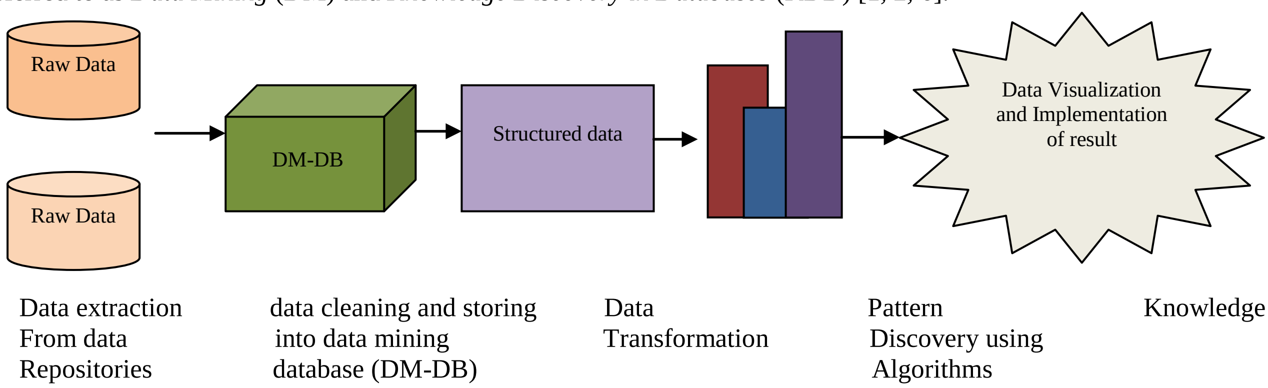

Fig.1: Data mining steps

At present, the decision tree has become an important data mining method. The basic learning approach of decision tree is greedy algorithm, which use the recursive top-down approach of decision tree structure [10]. Quin lan in 1979 put forward a well-known ID3 algorithm, which is the most widely used algorithm in decision tree [3]. But that algorithm has a defect of tending to use attributes with many values. To select the most useful attribute using classification technique, we present a metric termed as information gain and to catch an optimal way to classify set, we need to minimize the depth of the tree [9]. Aiming at the shortcomings of the ID3 algorithm, in the paper, we need to define information gain exactly, and need to deliberate entropy. ID3 decision tree use the Shannon entropy to calculate the information gain contained by data, which helps to make decision tree. This paper is organized as follow: Section2 covers background and related work, section3 discuss about Issue in id3 algorithm, and proposed modified id3 algorithm using Sharma-Taneja Entropy, in section4, we explain in detail about implementation of proposed algorithm and result analysis. Finally section5 covers the conclusion and future work.

II. BACKGROUND AND RELATED WORK

In this section we have discuss background and related work which is needed to understand the working functionality of the existing ID3 algorithm. In related we have discussed the work which already has been done by various researcher and students in this area.

A. Basic About ID3 Algorithm

The main ideas behind the ID3 algorithm are [8]:

- Each non-leaf node of a decision tree corresponds to an input attribute, and each arc to a possible value of that attribute. A leaf node corresponds to the expected value of the output attribute when the input attributes are described by the path from the root node to that leaf node.
- In a "good" decision tree, each non-leaf node should correspond to the input attribute which is the most informative about the output attribute amongst all the input attributes not yet considered in the path from the root node to that node. This is because we would like to predict the output attribute using the smallest possible number of questions on average.
- Entropy is used to determine how informative a particular input attribute is about the output attribute for a subset of the training data. Entropy is a measure of uncertainty in communication systems introduced by Shannon (1948).

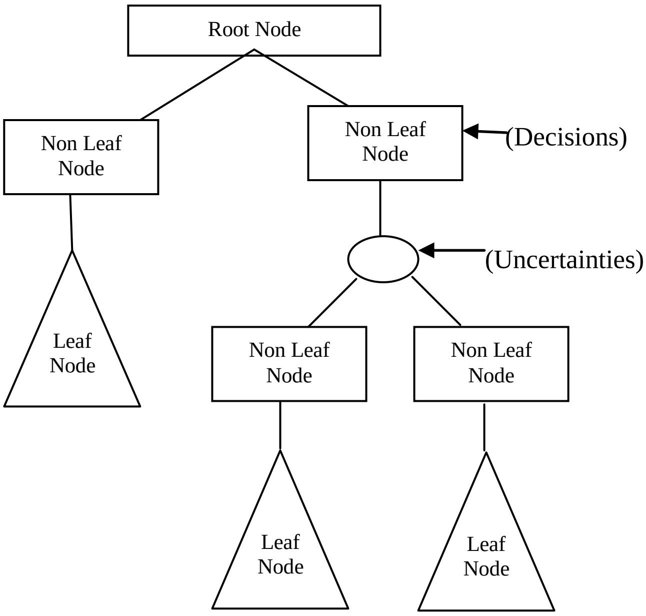

Fig2: Decision tree structure

B. Information Gain and Entropy

The main concepts used in ID3 algorithm are Information Gain and Entropy to select the attribute that is most useful for classifying the given sets. Entropy is defined as the measure of the amount of uncertainty in the (data) set. (i.e. entropy characterizes the data set). Let S is a set including *s* number of data samples whose type attribute can take m potential different values corresponding to m different types of C_i i (1,2,3,4,...,m). Assume that s_i is the sample number of C_i. Then the required amount of information to classify a given data is (Entropy is calculated by)[8]:

$$I(s_1,s_2,s_3,s_4,.........S_m)= - \sum_{i=1}^{n} p_i \log(p_i) \quad (1)$$

Where $p_i=S_{ij}/|S_j|$ is the probability that any subset of data samples belong to categories C_i.

Suppose that A is a property which has *n* different values $\{a_1, a_2, a_3, a_4,....,a_n\}$. Using the property of A, S can be divided into n number of subsets $\{S_1,S_2,S_3,.......,S_n\}$, in which Sj contain data samples whose attribute A are equal to aj in S set. If property A is selected as the property for test, that is, used to make partitions for current sample set, suppose that *ij S* is a sample set of type C_i in subset S_i , the required information entropy is

$$E(A)=\sum_{j=1}^{n} \frac{S_{1j}+S_{3j}+S_{3j}+S_{4j}+\cdots.......+S_{mj}}{s} \; I(S_{1j}, S_{2j} S_{3j} S_{4j} \ldots\ldots\ldots\ldots\ldots S_{mj} \quad (2)$$

Such use of property A on the current branch node corresponding set partitioning samples obtained information gain is

$$Gain(A)= I(s_1,s_2,s_3,s_4,.........S_m) - E(A) \quad (3)$$

In the decision tree method, information gain approach is generally used to determine suitable property for each node of a generated decision tree. Thus, we can select the attribute with the highest information gain in other words attributes with minimum entropy as the test attribute of current node. In this way, the information needed to classify the training sample subset obtained from later on partitioning will be the smallest. That is to say, the use of this property to partition the sample set contained in current node will make the mixture degree of different types for all generated sample subsets reduce to a minimum. The purpose of this ordering is to create small decision trees so that records can be identified after only a few decision tree splitting and match a hoped for plainness of the process of decision making.

C. Related Work
Chen Jin, Luo De-lin and Mu Fen-xiang, presented an association function in” An Improved ID3 algorithm” [3].In this paper to overcome the shortcoming of ID3 is inclining to choose attributes with many values, by combing the *association function* with the existing ID3algorithm.The result of the proposed algorithm by author show that the proposed algorithm can overcome ID3’s shortcoming effectively and get more reasonable and effective rules.
Nishant Mathur, Sumit Kumar, Santosh Kumar, and Rajni Jindal, presented “The Base Strategy for ID3 Algorithm of Data Mining Using Havrda and Charvat Entropy Based on Decision Tree” [9]. In This paper author introduces the use of ID3 algorithm of decision tree and use Havrda and Charvat Entropy instead of Shannon Entropy. By computing information author set particular property from taken data as root of tree, also sub-root by repeating the process continually, to finally build the most optimized tree.This paper introduces the use of ID3 algorithm of decision tree. Author used Havrda and Charvat Entropy Instead of Shannon Entropy.

III. ISSUE IN ID3 ALGORITHM , AND PROPOSED MODIFIED ID3 ALGORITHM USING SHARMA-TANEJA ENTROPY

A. Issue in ID3 Algorithm
The simple ID3 algorithm can have difficulties when an input attribute has many possible values, because Gain(X, T) tends to favour attributes which have a large number of values. To select the most useful attribute using classification technique, we present a metric termed as information gain and to catch an optimal way to classify set, we need to minimize the depth of the tree. In direction to define information gain exactly, we need to deliberate entropy. ID3 decision tree use the Shannon entropy to calculate the information gain contained by data, which helps to make decision tree. Shannon entropy result outcome are rather complex, have more numbers of node and leaf and decision rules. Thus it makes the decision making process time consuming.To minimize these problems, new algorithm proposed by modifying ID3 algorithm using Sharma-Taneja entropy instead of Shannon entropy.

B. Definition and Role of Sharma and Taneja’s Entropy in ID3 algorithm
Let P = ($p1$, $p2$ …pi); pi≥0 be a discrete probability distribution, p denotes the probability mass function and degree *α-entropy, β-entropy* is its inherent parameter. Then Sharma and Taneja’s gave the general entropy measure by formula shown under [4, 5]

$$H_{\alpha,\beta}\,(p) = \frac{1}{2^{1-\beta}-2^{1-\alpha}} \sum_{i=1}^{n} (p_i^{\beta} - p_i^{\alpha}), \alpha \neq \beta, \alpha, \beta > 0 \qquad (4)$$

This formula calculates Entropy. It has similar properties as that of Shannon entropy, but it contains additional parameter ‘α ‘and ‘β’ which can be used to make it more or less sensitive to the shape of probability distributions. If it has large positive value this measure is more sensitive to events that occur often, while for large negative values of ’α’ and ‘β’, it is more sensitive to the events which happen seldom. To avoid deduced solution in decision tree making process, Sharma and Taneja’s entropy based ID3 algorithm is proposed which may give good solution in reasonable time.

C. Proposed ID3 algorithm
ID3 algorithm traverses possible decision making space using top down greedy strategy, and never trace back and reconsider previous selections [3]. Information gain is exactly the matrices for selecting the best attribute in each step of the construction of decision tree in ID3 algorithm [7, 8].

Step 1: [select data set from the database, find number of attribute and total number of instance of dataset]
// Where n is total number of instance. And A is selected attribute
Step 2: [Calculate entropy of each attribute using Sharama-Taneja entropy calculation method]:

$$H_{\alpha,\beta}\,(p) = \frac{1}{2^{1-\beta} - 2^{1-\alpha}} \sum_{i=1}^{n} (p_i^{\beta} - p_i^{\alpha}), \alpha \neq \beta, \alpha, \beta > 0$$

Where H $_{\alpha,\beta}$ (p) is calculated entropy and is *α-entropy, β-entropy* is its inherent parameter..
Step 3: [Calculate the information Gains of all attributes.]
Gain (A) = Info (D) – InfoA(D)
//where Gain (A) tells us how much would be gained by branching on A attribute
Step 4: [split tree which has maximum value of information gain]
// set child node according the maximum information gain values
Step 5: For each child of the root Node, apply algorithm recursively until reach node that has entropy zero or reach leaf node.
Step 6: Display generated final optimal Decision tree.

IV. IMPLEMENTATION OF PROPOSED ALGORITHM AND RESULT ANALYSIS
To build a decision tree, we needed to decide which will be the root node of the tree. To find such node we have to calculate the needed information from the available sample data. To do this we divided our sampled data into category to build a decision tree. The sampled data is shown below

TABLE I

Sl. NO.	Freight Charges	Product Price	Product Weight	Delivery Period	Customer Category
1.	100-1000	<100	100-500	<5	L
2.	>1000	<100	>500	>20	L
3.	>1000	<100	100-500	5-20	L
4.	>1000	<100	>500	5-20	L
5.	100-1000	<100	>500	<5	N
6.	100-1000	100-2000	100-500	>20	L
7.	100-1000	>2000	<100	5-20	N
8.	100-1000	<100	<100	5-20	N
9.	100-1000	100-2000	<100	>20	N
10.	100-1000	<100	100-500	>20	L
11.	<100	>2000	100-500	5-20	N
12.	<100	<100	<100	5-20	N
13	<100	<100	<100	5-20	L
14	<100	<100	<100	<5	N
15	<100	>2000	<100	5-20	L
16	<100	<100	<100	<5	N
17	<100	100-2000	<100	<5	N
18	<100	100-2000	<100	<5	N
19	<100	<100	<100	<5	N

The summarized data of customer dispatch information in a section period (one month) from an information system database of a 3PL, which including 19 items in this sample data set. In this example [13], all sample data is divided by Customer category (CC) into two classes, which are Login_Customer (L) and Normal_Customer (N) respectively, and has four properties: Freight charges, Product price, Product weight, and Delivery period. On the one hand, the summarizing data is integrated data from different sections and different consignment nodes. On the other hand, it is the process of generalizing the sample data, namely, the low level data are substituted by high level convenient to data mining. The values of these four properties are: Freight charges (<100,100~1000, >1000); Product price (<100, 100~2000, <2000); Product Weight (<100 kg, 100 kg~500 kg, >500 kg); Delivery Period (<5, 5~20, >20) days. The meanings of these properties are: The freight charges is paid by customer for the transport cost; the Product price is Transportation Company bring the money of the goods from the receiver to dispatcher; the product weight is measured by kilogram; the Delivery period is the sum times during the summarized period.

A. Step by step Manual Calculation to evaluate Entropy & Information Gain

Step 1: We can calculate needed information by taking probability of customer category here we L class have 8 items and N has 11 items. Therefore, needed information (Entropy) of taken sample by putting α=0.25, and β=0.65 in Sharma and Taneja formula is shown below

Assuming α=0.25, and β=0.65

The needed information (Entropy) will be $I(p)=[\frac{-(\sum_{i=1}^{n} p_i^{\propto} - \sum_{i=1}^{n} p_i^{\beta})}{2^{1-\beta}-2^{1-\propto}}]$

$I(8,11)= (2^{1-.65}-2^{1-.25})^{-1}*[-(([\frac{8}{19}]^{.25} + [\frac{11}{19}]^{.25}) - ([\frac{8}{19}]^{.65} + [\frac{11}{19}]^{.65}))] = .0.999$

Step 2: Attribute Fright Charges

$E\ (Fright\ Charges) = -(2^{1-.65}-2^{1-.25})^{-1}*[\frac{9}{19}*[([\frac{2}{9}]^{.25} + [\frac{7}{9}]^{.25}) - ([\frac{2}{9}]^{.65} + [\frac{7}{9}]^{.65})] + \frac{7}{19}*[([\frac{3}{7}]^{.25} + [\frac{4}{7}]^{.25}) - ([\frac{3}{7}]^{.65} + [\frac{4}{7}]^{.65})] + \frac{3}{19}*[[\frac{3}{3}]^{.25} - ([\frac{3}{3}]^{.65})]] = .833$

Gain (I, Fright Charges) = I (S1, S2) – E (Freight Charges) =.999-.833=.166

Step3: Attribute "Product Price"

$E\ (Product\ Price) = -(2^{1-.65}-2^{1-.25})^{-1}*[\frac{12}{19}*[([\frac{6}{12}]^{.25} + [\frac{6}{12}]^{.25}) - ([\frac{6}{12}]^{.65} + [\frac{6}{12}]^{.65})] + \frac{4}{19}*[([\frac{1}{4}]^{.25} + [\frac{3}{4}]^{.25}) - ([\frac{1}{4}]^{.65} + [\frac{3}{4}]^{.65})] + \frac{3}{19}*[([\frac{2}{3}]^{.25} + [\frac{1}{3}]^{.25}) - ([\frac{2}{3}]^{.65} + [\frac{1}{3}]^{.65})]] = .996$

Gain (I, Product Price) = I (S1, S2) - E (Product price) =.999-.996=.002

Step4: Attribute "Product Weight"

$E\ (Product\ Weight) = -(2^{1-.65}-2^{1-.25})^{-1}*[\frac{11}{19}*[([\frac{2}{11}]^{.25} + [\frac{9}{11}]^{.25}) - ([\frac{2}{11}]^{.65} + [\frac{9}{11}]^{.65})] + \frac{5}{19}*[\left([\frac{4}{5}]^{.25} + [\frac{1}{5}]^{.25}\right) - ([\frac{4}{5}]^{.65} + [\frac{1}{5}]^{.65})] + \frac{3}{19}*[([\frac{2}{3}]^{.25} + [\frac{1}{3}]^{.25}) - ([\frac{2}{3}]^{.65} + [\frac{1}{3}]^{.65})]] = .977$

Gain (I, Product Weight) = I (S1,S2) – E (Product Weight)= .999-.97=.021

Step 5: Attribute "Delivery Period"

E (Delivery Period) = - $(2^{1-.65}-2^{1-.25})^{-1}$ $*[\frac{7}{19}*[(\left[\frac{1}{7}\right]^{.25}+\left[\frac{6}{7}\right]^{.25})-(\left[\frac{1}{7}\right]^{.65}+\left[\frac{6}{7}\right]^{.65})]+\frac{8}{19}*[(\left[\frac{4}{8}\right]^{.25}+\left[\frac{4}{8}\right]^{.25})-(\left[\frac{4}{8}\right]^{.65}+\left[\frac{4}{8}\right]^{.65})]+\frac{4}{19}*[(\left[\frac{3}{4}\right]^{.25}+\left[\frac{1}{4}\right]^{.25})-(\left[\frac{3}{4}\right]^{.65}+\left[\frac{1}{4}\right]^{.65})]]$ =.981

Gain (I, Delivery Period) = I (S1,S2) – E(Delivery Period) = .999-.981 =.017

Step 6: Summary of results is:

Entropy	= 0.999
Gain (I, Fright Charges)	**= 0.166**
Gain (I, Product Price)	= 0.002
Gain (I, Product Weight)	= 0.021
Gain (I, Delivery Period)	= 0.017

Step 7: Find which attribute is the root node

Gain (I, Fright Charges)= 0.166 is Highest. Therefore, “Fright Charges” attribute is the decision attribute in the root node. “Fright Charges” root node has three possible values <100, 100-1000, >1000.

Step8: Now considering table where freight fee is <100, from main table.

Table Information of Customer Dispatch Goods for freight fee (<100)

Sl. NO.	Freight Charges	Product Price	Product Weight	Delivery Period	Customer Category
11	<100	>2000	100-500	5-20	N
12	<100	<100	<100	5-20	N
13	<100	<100	<100	5-20	L
14	<100	<100	<100	<5	N
15	<100	>2000	<100	5-20	L
16	<100	<100	<100	<5	N
17	<100	100-2000	<100	<5	N
18	<100	100-2000	<100	<5	N
19	<100	<100	<100	<5	N

So, here L=2, N= 7 we calculated needed information

I (2, 7) = $(2^{1-.65}-2^{1-.25})^{-1}*[-((\left[\frac{2}{9}\right]^{.25}+\left[\frac{7}{9}\right]^{.25})-(\left[\frac{2}{9}\right]^{.65}+\left[\frac{7}{9}\right]^{.65}))]$= .982

Corresponding anticipated information of different properties are: E (Fright Charges) = .852

E (product Price) = $(2^{1-.65}-2^{1-.25})^{-1}*[\frac{5}{9}*[(\left[\frac{1}{5}\right]^{.25}+\left[\frac{4}{5}\right]^{.25})-(\left[\frac{1}{5}\right]^{.65}+\left[\frac{4}{5}\right]^{.65})]+\frac{2}{9}*[([0]^{.25}+[2/2]^{.25})-([0]^{.65}+\left[\frac{2}{2}\right]^{.65})]+\frac{2}{9}*[(\left[\frac{1}{2}\right]^{.25}+\left[\frac{1}{2}\right]^{.25})-(\left[\frac{1}{2}\right]^{.65}+\left[\frac{1}{2}\right]^{.65})]]$= .765

E (Product Weight) = $(2^{1-.65}-2^{1-.25})^{-1}*[\frac{8}{9}*[(\left[\frac{2}{8}\right]^{.25}+\left[\frac{6}{8}\right]^{.25})-(\left[\frac{2}{8}\right]^{.65}+\left[\frac{6}{8}\right]^{.65})]+0+0]$=.877

E (Delivery Period) = $(2^{1-.65}-2^{1-.25})^{-1}*[\frac{5}{9}*[([0]^{.25}+\left[\frac{5}{5}\right]^{.25})-([0]^{.65}+\left[\frac{5}{5}\right]^{.65})]+\frac{4}{19}*[\left[\frac{2}{4}\right]^{.25}+\left[\frac{2}{4}\right]^{.25}-(\left[\frac{2}{4}\right]^{.65}+\left[\frac{2}{4}\right]^{.65})]+0]$ =.444

From the above information gain and entropy, we can calculate the gain for fright Charges<100

- Gain (Freight Charges) = I (S1, S2) - E (Freight Charges) =.982-.982 =0.00
- Gain (Product Price) = I (S1, S2) - E (Product Price) =.982-.7650 =0.217
- Gain (Product Weight) = I (S1, S2) - E (Product Weight) = .982-.877 =0.104
- Gain (Delivery Period) = I (S1, S2) - E (Delivery Period) = .982-..444 =0.538

Here Information gain of Time is maximum, therefore *Delivery Period* will be SUBROOT under Freight charges (<100) root.

Step9: This process will continue till all the data become classified or we run out of attributes. Final Decision tree will be

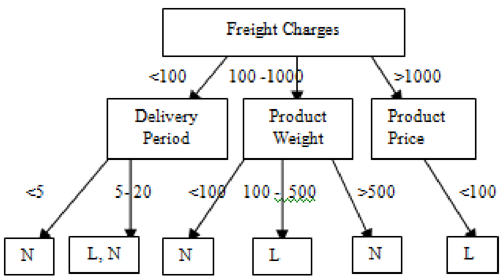

Fig4. Final Decision tree for Customer Dispatch Goods

B. Code Output After The Execution of the Modified ID3 Algorithm for the Given Input Data Set

```
Building table and reading entries.
enter the '0' if u want to calculate H-C entropy  or '1' to calculate 'S-T entropy
1
Finding the property having highest information gain.
Information Gain : 0.999185
        column 0    Entropy 0.833693    Info Gain : 0.165492
        column 1    Entropy 0.996689    Info Gain : 0.002496
        column 2    Entropy 0.977750    Info Gain : 0.021435
        column 3    Entropy 0.981813    Info Gain : 0.017372
       Best Information Gain for column 0
Sub root node 0
Finding the property having highest information gain.
Information Gain : 0.982755
        column 0    Entropy 0.982755    Info Gain : 0.000000
        column 1    Entropy 0.765461    Info Gain : 0.217293
        column 2    Entropy 0.877766    Info Gain : 0.104989
        column 3    Entropy 0.444444    Info Gain : 0.538310
       Best Information Gain for column 3
Sub root node 1
Finding the property having highest information gain.
Information Gain : 0.999338
        column 0    Entropy 0.999338    Info Gain : 0.000000
        column 1    Entropy 0.857143    Info Gain : 0.142195
        column 2    Entropy 0.000000    Info Gain : 0.999338
        column 3    Entropy 0.712449    Info Gain : 0.286890
       Best Information Gain for column 2
Sub root node 2
Finding the property having highest information gain.
Information Gain : 0.000000
        column 0    Entropy 0.000000    Info Gain : 0.000000
        column 1    Entropy 0.000000    Info Gain : 0.000000
        column 2    Entropy 0.000000    Info Gain : 0.000000
        column 3    Entropy 0.000000    Info Gain : 0.000000
       Best Information Gain for column 1
root node:
        column 0 Number of child nodes 3
Sub nodes
       Sub Node 0 column 3 Number of leaf nodes 3
       Leaf Nodes
              Leaf Node 0 Only N
              Leaf Node 1 Both L & N
              Leaf Node 2 None
       Sub Node 1 column 2 Number of leaf nodes 3
       Leaf Nodes
              Leaf Node 0 Only N
              Leaf Node 1 Only L
              Leaf Node 2 Only N
       Sub Node 2 column 1 Number of leaf nodes 3
       Leaf Nodes
              Leaf Node 0 Only L
              Leaf Node 1 None
          Leaf Node 2 None
Building table and reading entries.
enter the '0' if u wanna to calculate H-C entropy
 or '1' to calculate 'S-T entropy
0
Finding the property having highest information gain.
Information Gain : 0.903763
        column 0    Entropy 0.728497    Info Gain : 0.175266
        column 1    Entropy 0.892819    Info Gain : 0.010944
        column 2    Entropy 0.821574    Info Gain : 0.082189
        column 3    Entropy 0.845198    Info Gain : 0.058565
```



```
        Best Information Gain for column 0
Sub root node 0
Finding the property having highest information gain.
Information Gain : 0.834258
        column 0    Entropy 0.834258    Info Gain : 0.000000
        column 1    Entropy 0.657184    Info Gain : 0.177073
        column 2    Entropy 0.755806    Info Gain : 0.078452
        column 3    Entropy 0.404025    Info Gain : 0.430233
        Best Information Gain for column 3
Sub root node 1
Finding the property having highest information gain.
Information Gain : 0.904731
        column 0    Entropy 0.904731    Info Gain : 0.000000
        column 1    Entropy 0.779192    Info Gain : 0.125539
        column 2    Entropy 0.000000    Info Gain : 0.904731
        column 3    Entropy 0.638838    Info Gain : 0.265893
        Best Information Gain for column 2
Sub root node 2
Finding the property having highest information gain.
Information Gain : 0.000000
        column 0    Entropy 0.000000    Info Gain : 0.000000
        column 1    Entropy 0.000000    Info Gain : 0.000000
        column 2    Entropy 0.000000    Info Gain : 0.000000
        column 3    Entropy 0.000000    Info Gain : 0.000000
        Best Information Gain for column 1
root node:
        column 0 Nr of child nodes 3
Sub nodes
        Sub Node 0 coloumn 3 Nr of leaf nodes 3
        Leaf Nodes
                Leaf Node 0 Only N
                Leaf Node 1 Both L & N
                Leaf Node 2 None
        Sub Node 1 coloumn 2 Nr of leaf nodes 3
        Leaf Nodes
                Leaf Node 0 Only N
                Leaf Node 1 Only L
                Leaf Node 2 Only N
        Sub Node 2 coloumn 1 Nr of leaf nodes 3
        Leaf Nodes
                Leaf Node 0 Only L
                Leaf Node 1 None
                Leaf Node 2 None
```

C. . Results Comparison of H-C Entropy and S-T Entropy

TABLE 8: COMPARISON CHART FOR DIFFERENT VALUE FOR DIFFERENT TYPE OF ENTROPY

Sl. No.	**H-C Entropy For (α=.25) Information Gain**	**S-T Entropy For α=.25, β=.65 Information Gain**	**S-T Entropy For α=.25, β=.85 Information Gain**
1.class-column having highest information gain.	0.903763	0.999185	0.996159
Class-column 0 Max. Information gain	0.175266	0.165492	0.182238
2.class-column having highest information gain for subtable.	0.834258	0.982755	0.942939
Class-column3 Max. Information gain	0.430233	0.444444	0.498495
3. class-column having highest information gain. (for sub table2)	0.904731	0.999338	0.996864
Class-column2 Max. Information gain	0.904731	0.999338	0.996864

4. class-column having highest information gain.(for sub table3)	0.000000	0.000000	0.000000
Class-column1 Max. Information gain	0.000000	0.000000	0.000000

V. CONCLUSION AND FUTURE WORK

Data mining is a broad area that integrates techniques from several fields, for the analysis of large volumes of data. Data mining can use for classifying customer data. This will help us to give more value to customer by increasing their information, also help in providing high quality services to them by understanding them. The decision tree tells what customers want the most. In this paper ID3algorithm is used with modification. Instead of using Shannon Entropy, Sharma and Taneja Entropy has been used to find the information of different properties which is used as the node of decision tree. This modification has reduced the size of tree as well as decreased the rules, which will help to understand customer characteristics by which company growth and profit can be increased. As a result for lower value of alpha (α) =0.25, and beta (β)=0.65, and 0.85 and other value ($\alpha<1$, $\beta<1$) tree is small and less complex as compared to the use of Shannon entropy and H-C Entropy. To conclude, we can say that if we want to get less number of node and leaf in a tree and to make it more effective and less complex, we can use the Shrama and Taneja entropy instead of Shannon entropy and Havrda and Charvat entropy. The value of both alpha (α), and beta (β) less than one will give decision tree with less number of nodes.In this research, the value of alpha (α) has been put as 0.25 and the value of beta(β) has been put as 0.65,and 0.85, but for further research we can take varying values of alpha (α) and beta(β)which may give different trees. In future work, Instead of using Sharma-Taneja entropy, different entropies can also be used for further research.

REFERENCES

[1] Introduction to Data Mining Pang-Ning Tan, Michigan State University, Michael Steinbach, Vipin Kumar, University of Minnesota.

[2] Fayyad, U., Data Mining and Knowledge Discovery: Making Sense out of Data, IEEE Expert, Oct. 20-25, 1996

[3] Chen Jin, Luo De-lin, Mu Fen-xiang "An Improved ID3 Decision Tree Algorithm, "Proceedings of IEEE 978-1-4244-3521-0/09, 2009, 4th International Conference on Computer Science & Education.

[4] Ishwar Singh, "On Sharma-Taneja's Entropy of Type (α, β)," 19/11/1977

[5] B.D. Sharma and I.J. Taneja, I.J.: Entropy of type (α, β) and other generalized measures in information theory. Metrika, 22 (1975), 205-215

[6] http://en.wikipedia.org/wiki/Data_mining (data mining application)

[7] Building Classification Models ID3 and C4.5 algorithm with detail explanation and example:http://wwww.cis.temple.edu/~cis587/reading/id3-c45.html

[8] Wei Peng, Juhua Chen and Haiping Zhou Project of Comp 9417: Machine Learning," An Implementation of ID3 --Decision Tree Learning Algorithm" From web.arch.usyd.edu.au/wpeng/DecisionTree2.pdf Retrieved date: May 13, 2009.

[9] Nishant Mathur, Sumit Kumar, Santosh Kumar, and Rajni Jindal," The Base Strategy for ID3 Algorithm of Data Mining Using Havrda and Charvat Entropy Based on Decision Tree;" International Journal of Information and Electronics Engineering, Vol. 2, No. 2, March 2012

[10] Rahul A. Patil, Prashant G. Ahire, Pramod. D. Patil, Avinash L. Golande," A Modified Approach to Construct Decision Tree in Data Mining Classification," ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 1, July 2012

[11] M. Havrda and F. Charvat, "Quantification method of classification processes: concept of structural alpha-entropy," Kybernetica, vol. 3, pp. 30–35, 1967

