Implementation of ID3 Algorithm

N D

Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Implementation of ID3 Algorithm

Rupali Bhardwaj , Sonia Vatta
CSE, Bahra Uuniversity
India

Abstract— Inductive learning is the learning that is based on induction. In inductive learningDecision tree algorithms are very famous. For the appropriate classification of the objects with the given attributes inductive methods use these algorithms basically. These algorithms are very important in the classification of the objects. That is why many of these algorithms are used in the intelligent systems as well. In this paper the ID3 decision tree learning algorithm is implemented with the help of an example which includes the training set of two weeks. The basic calculations are used to calculate the classification related to the training set used.The resultant of the work will be the classified decision tree and the decision rules. The algorithm is implemented in the java language.

Keywords— inductive, classification, ID3, decision learning, algorithm

I. INTRODUCTION

In the human history, people had used various technologies to model themselves. There are many evidences of this from ancient China, Egypt, and Greece that bears the witness to the universality of this. Each new technology has been use to build intelligent agents or models of mind. Clockwork, hydraulics, telephone switching systems, holograms, analog computers, and digital computers have all been proposed both as technological metaphors for intelligence and as mechanisms for modelling mind. Hobbes (1588-1679), who has been described by Haugeland (1985), as the "Grandfather of AI," supported the position that thinking was the symbolic reasoning like talking out loud or working out an answer with pen and paper. Symbolic operations became more definite with the development of computers. The first general-purpose computer designed (but not built until 1991, at the Science Museum of London) was the **Analytical Engine** by Babbage (1792-1871). In the early part of the 20th century, there was much work done on understanding computation. Several models of computation were proposed, including the Turing machine by Alan Turing (1912-1954), a theoretical machine that writes symbols on an infinitely long tape, and the lambda calculus of Church (1903-1995), which is a mathematical formalism for rewriting formulas. There was a large body of work on **expert systems,** during the 1970s and 1980s.The aim was to apprehend the knowledge of an expert in some realm so that a computer could carry out expert tasks. For example, **DENDRAL** [Buchanan and Feigenbaum (1978)], developed from 1965 to 1983 in the field of organic chemistry, proposed plausible structures for new organic compounds. **MYCIN** [Buchanan and Shortliffe (1984)], developed from 1972 to 1980, diagnosed infectious diseases of the blood, prescribed antimicrobial therapy, and explained its reasoning. The 1970s and 1980s were also a period when AI reasoning became widespread in languages such as **Prolog** [Colmerauer and Roussel (1996)][Kowalski (1988)]. During the 1990s and the 2000s there was great growth in the sub disciplines of AI such as perception, probabilistic and decision-theoretic reasoning, planning, embodied systems, machine learning, and many other fields. There are many of the more work has been going on till today on the advancement of the already work done ,and on the new concepts of the AI. ID3, Iterative Dichotomiser 3 is a decision tree learning algorithm which is used for the classification of the objects with the iterative inductive approach. In this algorithm the top to down approach is used. The top node is called as the root node and others are the leaf nodes. So it's a traversing from root node to leaf nodes. Each node requires some test on the attributes which decide the level of the leaf nodes. These decision trees are mostly used for the decision making purpose [8].

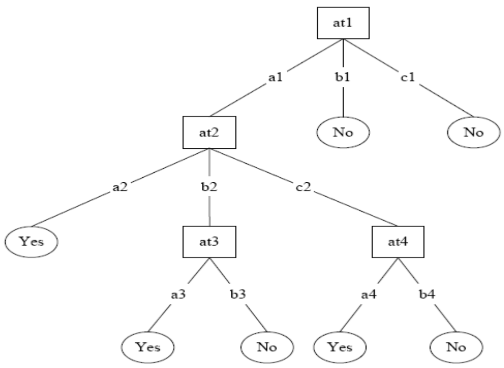

Figure 1: Decision tree

Decision tree learning is a procedure for calculating the target value having discrete function. The function that has been learned is symbolized by a decision tree. For the inductive inference the decision tree learning is one of the most commonly and broadly used methods which are practical in nature [1][3].

The decision tree learning algorithms are mainly used because of the three reasons :

1. Decision tree is a good infer from the particular cases that are unobserved instance.
2. The calculations in these methods are efficient and are proportional to the instances that are observed.
3. At the final, the decision tree which is produced is easily understood by the human. [2][3]

II. Rationale Of Basic Id3

A. CHOOSING ATTRIBUTES AND ID3

The order in which attributes are chosen determines how complicated the tree is.ID3 uses information theory to determine the most informative attribute.A measure of the information content of a message is the inverse of the probability of receiving the message**:**

$$\textbf{\textit{information1}}\ \boldsymbol{(M) = 1/}\textbf{\textit{probability}}\ \boldsymbol{(M)}$$

Taking logs (base 2) makes information correspond to the number of bits required to encode a message:

$$\textbf{\textit{information}}\ \boldsymbol{(M) = -\log_2 (}\textbf{\textit{probability}}\ \boldsymbol{(M))}$$

B. INFORMATION

The information content of a message should be related to the degree of surprise in receiving the message.Messages with a high probability of arrival are not as informative as messages with low probability.Learning aims to predict accurately i.e. reduce surprise.Probabilities are multiplied to get the probability of two or more things both/all happening.Taking logarithms of the probabilities allows information to be added instead of multiplied

C. ENTROPY

Different messages have different probabilities of arrival.Overall level of uncertainty (termed entropy) is:

$$-\Sigma_i P_i \log_2 P_i$$

Frequency can be used as a probability estimate.e.g. if there are 5 positive examples and 3 negative examples in a node the estimated probability of positive is 5/8 = 0.625.

D. INFORMATION AND LEARNING

We can think of learning as building many-to-one mappings between input and output.Learning tries to reduce theinformation content of the inputs by mapping them to fewer outputs.Hence we try to minimize entropy.The simplest mapping is to map everything to one output.We seek a trade-off between accuracy and simplicity.

E. SPLITTING CRITERION

Work out entropy based on distribution of classes.Trying splitting on each attribute.Work out expected information gain for each attribute.Choose best attribute.

ID3 is simple decision tree learning algorithm which uses the greedy top to down search to build the tree which will decide the decision rules. For this there is a requirement for some mathematical concepts. The two concepts which are basically involved in ID3 are ENTROPY and INFORMATION GAIN.

.

F.RULES OF CLASSIFYING.

If the entropy of the attribute is 0, it is a homogeneous node and there is no need to classify further. If the entropy of the attribute is 1, it is a heterogeneous node and there is a need to classify further.

III. Literature Review

The first work on ID3 was done by J.R Quinlan in 1986.He summarizes an approach tosynthesizing decision trees that has been used in a variety of systems, and it describes one such system, ID3, in detail. Results from recent studies show ways in which the methodology can be modified to dealwith information that is noisy and/or incomplete. A reported shortcoming of the basic algorithm isdiscussed and two means of overcoming it are compared. He concludes with illustrations of currentresearch directions. [4]. Data mining techniques basically use the ID3 algorithm as it's the basic algorithm of classification. Some reusable components were identified from ID3 algorithm by Milija Suknovic (et.al). They propose a generic decision tree framework that supports reusable components design. The proposed generic decision tree framework consists of several sub-problems which were recognized by analysing well-known decision tree induction algorithms, namely ID3, C4.5, CART, CHAID, QUEST, GUIDE, CRUISE, and CTREE. They identified reusable components in these algorithms as well as in several of their partial improvements that can be used as solutions for sub-problems in the generic decision tree framework. The identified components can now be used outside the algorithm they originate from. Combining reusable components allows the replication of original algorithms, their modification but also the creation of new decision tree induction algorithms. Every original algorithm can outperform other algorithms under specific conditions but can also perform poorly when these conditions change. Reusable components allow exchanging of solutions from various algorithms and fast design of new algorithms. They offer a generic framework for component-based algorithms design that enhances understanding, testing and usability of decision tree algorithm parts. They combined the reusable components so as to allow the replication of original algorithm. They also found out that their modification will create a new induction algorithm [14]. An interesting implementation or the use of ID3 was done in the field of networks security. Victor H.Goreia (et.al) use ID3 to web attack detection. Decision tree learning algorithms have been successfully used in knowledge discovery. They use induction in order to provide an appropriate classification of objects in terms of their attributes, inferring decision tree rules. Author reports on the use of ID3 to Web attack detection. Even though simple, ID3 is sufficient to put apart a number of Web attacks, including a large proportion of their variants. It also surpasses existing methods: it portrays a higher true-positive detection rate and a lower false-positive one. The ID3 output classification rules that are easy to read and so computer officers are more likely to grasp the root of an attack, as well as extending the capabilities of the classifier.[9]. Another researcher was Sonika Tiwari who used the improved ID3 for detecting Network Anomalies with horizontal portioning based decision tree. She first apply different clustering algorithms and after that apply horizontal portioning decision tree and then check the network anomalies from the decision tree. She found the comparative analysis of different clustering algorithms and existing id3 based decision tree. [10]. Data mining techniques basically use the ID3 algorithm as it's the basic algorithm of classification. In the medical field ID3 were mainly used for the data mining.Ruijuan Hu used the ID3 algorithm for retrieving the data for the breast cancer which is carried out for the primarily predicting the relationship between the recurrence and other attributes of breast cancer. Detailed elaborations are presented for the idea on ID3 algorithm of Decision Tree. An improved method called Improved ID3 algorithm that can improve the speed of generation is brought forward owing to the disadvantages of ID3 algorithm. Moreover, based on Improved ID3 algorithm, data mining for breast-cancers is carried out for primarily predicting the relationship between recurrence and other attributes of breast cancer by making use of SQL Server 2005 Analysis Services. Results prove the effectiveness of Decision Tree in medical data mining which provide physicians with diagnostic assistance. [7]. Prediction of common diseases in mobile phones and television is also done with the use of ID3 by L.Sathish Kumar and A.Padmapriya. They propose method phone which helps the people to know about the diseases to avoid the death rate and diseases affected people count. The data mining has become a unique tool in analyzing data from different perspective and converting it into useful and meaningful information. They said that we have a lot of known diseases and unknown diseases around the world. The healthcare has big challenge to predict the kind of disease and the solution for that disease. In India illiteracy rate is high, so that most of the people are scared about these diseases become of thesis ignorance. Hence they may take wrong decision regarding the disease that they have been affected problem. Considering this serious issue we have used data mining as a tool to overcome this issue. We have already created the prediction for common disease. They processed implementing of mobile phone and television because all category people can used easily find and predicted what kind of disease through television and mobile phones. [13]. Mary Slocum implemented the ID3 for data mining in the medical research. She modifies the large amount of data and transforms the data into information which can be used to make a decision. In this study, she reviewed the initial data mining task of collecting and cleaning the data prior to the use of the data with the algorithm. For ID3, the two key concepts are Entropy (measurement of uncertainty) and Information Gain (measurement of purity). Using these parameters, she created a top-down tree that she can later traverse breadth-first to make a decision given a new data set. Though the ID3 algorithm worked well with the non-contiguous data that she created and has the advantage that it generates a smaller depth decision tree, she would want to evaluate this algorithm with a larger and more complicated data set. Also, she might want to consider evaluating different types of decision trees along with clustering algorithms to determine if there is a better approach for the medical industry specifically for determination of the risk of heart disease. In fact, since ID3 was first developed, there is now an improved version called C4.5. Lastly, there are vast amounts of data available and there are many ways it can be manipulated. So, using these algorithms is an iterative process where processes are always being improved (such as when new attributes are added for considerations – for example, there may other factors for risk of heart disease such as weight or family history). For the medical industry, these decisions can determine if a patient is a high risk for heart disease along with making conclusions as to what insurance coverage a company should give a person based on this risk [11].

The implementation and evaluation of ID3 was done by some authors on different examples, like Anand Bahety implemented the algorithm on the "Play Tennis" database. He classified whether the weather is suitable for playing tennis or not?. Decision tree algorithms are a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree. These kinds of algorithms are famous in inductive learning and have been successfully applied to a broad range of tasks. He examines the decision tree learning algorithm – ID3 against nominal and continuous attributes extend it to handle missing value. Experiments to evaluate the performance of the algorithm with continuous valued attributes and missing attribute values reveal that ID3 does not give acceptable results for continuous valued attributes and works well in certain data sets with missing values. [8].

T.Y HSU 662096093 did the survey on the simple "loan application" dataset. He classified whether a person will get a house loan or not?

An Implementation of ID3 byWei Peng (et.al) describes that Decision tree learning algorithm has been successfully used in expert systems in capturing knowledge. The main task performed in these systems is using inductive methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules. He examines the decision tree learning algorithm ID3 and implements this algorithm using Java programming. He first implements basic ID3 in which he dealt with the target function that has discrete output values. He also extends the domain of ID3 to real-valued output, such as numeric data and discrete outcome rather than simply Boolean value. The Java applet provided at last section offers a simulation of decision-tree learning algorithm in various situations. Some shortcomings are discussed in this project as well. Kumar Ashok (et.al) did the same on the "census 2011 of India" dataset. He classified it by using ID3 algorithm. The concept of information theory is applied in the field of data mining. As in the algorithms of data mining, the classification is an essential step, using an information theoretic measure in ID3 algorithm, one of the key algorithms of decision tree algorithms, they have discussed the different steps of the development of decision tree so that the best classification criteria can be developed which is helpful in making good decisions. From the data under consideration having a set of values, a property on the basis of calculation is selected as the root of the tree and the process is repeated to develop complete decision tree. Further, this method is applied to the data of Census-2011 of India to get some values worth in improving or implementing a policy and to select right policy for right people Another research implemented the same to find out whether a person is sunburnt or not? [12].

IV. Objectives Of The Research Work

The main objectives of the research work here are

- First, to construct the decision tree until the appropriate classification is reached.
- Another is to generate the decision rules for the problem.

V. Results And Discussion

Using ID3 algorithm we need to decide if the weatheris amenable to play CRICKET.Training data collected over two weeks.

DAY	WEATHER	TEMP	HUMIDITY	WIND	PLAY
1	Sunny	85	85	week	No
2	Sunny	80	90	Strong	No
3	Cloudy	83	78	Week	Yes
4	Rainy	70	96	Week	Yes
5	Rainy	68	80	Week	Yes
6	Rainy	65	70	Strong	No
7	Cloudy	64	65	Strong	Yes
8	Sunny	72	95	Week	No
9	Sunny	69	70	Week	Yes
10	Rainy	75	80	Week	Yes
11	Sunny	75	70	Strong	Yes
12	Cloudy	72	90	Strong	Yes
13	Cloudy	81	75	Week	Yes
14	Rainy	71	85	Strong	No

LEARNING SET

In the above example, two attributes, the temperature and humidity have continuous ranges; ID3 requires them to be discrete like hot, medium, cold, high, and normal. Table below indicates the acceptable values.

ATTRIBUTE	POSSIBLE VALUES		
WEATHER	Sunny	Cloudy	Rainy
Temperature	Hot	Medium	Cold
Humidity	High	Normal	-
Wind	Strong	Week	-
Class	Play	No play	-
Decision	N(negative)	P(positive)	-

Assign discrete values to the attributes Partition the continuous attribute values to make them discrete, following the key mentioned below.

Temperature	Hot (H) 80 o 85	Medium (M) 70 to 75	Cold (C) 64 to 69
Humidity	High (H) 81 to 96	Normal (N) 65 to 80	
Class	Yes (Y) Play	No (N) no play	

Day	Weather	Temp	Humidity	Wind	Play
1	Sunny	85Hot	85 High	week	No
2	Sunny	80Hot	90 High	Strong	No
3	Cloudy	83Hot	78 High	Week	Yes
4	Rainy	70Medium	96 High	Week	Yes
5	Rainy	68Cold	80 Normal	Week	Yes
6	Rainy	65Cold	70 Normal	Strong	No
7	Cloudy	64Cold	65 Normal	Strong	Yes
8	Sunny	72Medium	95 High	Week	No
9	Sunny	69Cold	70 Normal	Week	Yes
10	Rainy	75Medium	80 Normal	Week	Yes
11	Sunny	75Medium	70 Normal	Strong	Yes
12	Cloudy	72Medium	90 High	Strong	Yes
13	Cloudy	81Hot	75 Normal	Week	Yes
14	Rainy	71Medium	85 High	Strong	No

4.6 STEP BY STEP CALCULATIONS:

STEP 1:“example” set s

The set s of 14 examples with 9 yes and 5 no then

Entropy (S) = -(9/14) Log2 (9/14) – (5/14) Log2 (5/14) = 0.940

STEP 2: Attribute weather

Weather value can be sunny, cloudy, and rainy.

Weather =sunny is of occurrence 5

Weather = cloudy is of occurrences 4

Weather = rainy is of occurrences 5

Weather = sunny, 2 of the examples are ‘yes’ and 3 are ‘no’

Weather = cloudy, 4 of the examples are ‘yes’ and 0 are ‘no’

Weather = rainy, 3 of the examples are ‘yes’ and 2 are ‘no’

Entropy (Ssunny) = -(2/5)xlog2(2/5) – (3/5)xlog2(3/5) = 0.970950

Entropy (Scloudy) = -(4/4)xlog2 (4/4) – (0/4) xlog2 (0/4) = 0

Entropy (Ssunny) = -(3/5)xlog2 (3/5) – (2/5) xlog2 (2/5) = 0.970950

Gain (S, weather) = Entropy (S) – (5/14) x Entropy (S sunny)
- (4/14) x Entropy (Scloudy)
- (5/14) x Entropy (Srainy)

= 0.940 – (5/14) x 0.97095059 – (4/14) x 0 – (5/14) x 0. 97095059
= 0.940 – 0.34676 – 0 – 0.34676
= 0.246

STEP 3: Attribute temperature

Temp value can be hot, medium or cold.

Temp = hot is of occurrences 4

Temp = medium is of occurrences 6

Temp = cold is of occurrences 4

Temp =hot, 2 of the examples are ‘yes’ and 2 are ‘no’

Temp =medium, 4 of the examples are ‘yes’ and 2 are ‘no’

Temp =cold, 3 of the examples are ‘yes’ and 1 are ‘no’

Entropy (Shot) = - (2/4) x log2 (2/4) – (2/4) x log2 (2/4) = - 0.99999999

Entropy (Smedium) = **-** (4/6) x log2 (4/6) – (2/6) x log2 (2/6) = **-** 0.91829583
Entropy (Scold) = **-** (3/4) x log2 (3/4) – (1/4) x log2 (1/4) = **-** 0.81127812

Gain (S, Temp) = Entropy (S) – (4/14) x Entropy (Shot)
- (6/14) x Entropy (Smedium)
- (4/14) x Entropy (Scold)

= 0.940 – (4/14) x 0.99999 – (6/14) x 0.91829583 – (4/14) x 0.81127812
= 0.940 – 0.2857142 **-** 0.393555 – 0.2317937
= 0.0289366072

STEP 4: Attribute Humidity
Humidity value can be high, normal.
Humidity = high is of occurrences 7
Humidity = normal is of occurrences 7
Humidity = high, 3 of the examples are 'yes' and 4 are 'no'
Humidity = normal, 6 of the examples are 'yes' and 1 are 'no'
Entropy (Shigh) = **-** (3/7) x log2 (3/7) – (4/7) x log2 (4/7) = **-** 0.9852281
Entropy (Snormal) = **-** (6/7) x log2 (6/7) – (1/7) x log2 (1/7) = **-** 0.5916727

Gain (S, Temp) = Entropy (S) – (7/14) x Entropy (Shigh)
- (7/14) x Entropy (Snormal)

= 0.940 – (7/14) x 0.9852281 – (7/14) x 0. 5916727
= 0.940 –0.49261405 **-** 0.29583635
= 0.1515496

STEP 5: Attribute Wind
Windvalue can be weak or strong.
Wind = weak is of occurrences 8
Wind = strong is of occurrences 6
Wind = weak, 6 of the examples are 'yes' and 2 are 'no'
Wind = strong, 3 of the examples are 'yes' and 3 are 'no'

Entropy (Sweak) = **-** (6/8) x log2 (6/8) – (2/8) x log2 (2/8) = **-** 0.811
Entropy (S strong) = **-** (3/6) x log2 (3/6) – (3/6) x log2 (3/6) = **-** 1.00

Gain (S, Temp) = Entropy (S) – (8/14) x Entropy (S weak)
- (6/14) x Entropy (S strong)

= 0.940 – (8/14) x 0.811 – (6/14) x 1.00
= 0.048

STEP 6: Summary of results are:

Entropy (S)	= 0.940
Gain (S, weather)	=***0.246***
Gain (S, Temp)	= 0.0289366072
Gain (S, Humidity)	= 0.1515496
Gain (S, Wind)	= 0.048

STEP 7: Find which attribute is the root node
Gain (S, weather) = 0.246 is highest.
Therefore, "weather" attribute is the decision attribute in the root node.
"Weather" as root node has three possible values – sunny, cloudy, rain.

STEP 8: Find which attribute is next decision node.
"Weather" has three possible values.
So root node has three branches (sunny, cloudy, rain).
S_{sunny} = {D1, D2, D8, D9, D11} = 5 "examples". D represents "days".
With weather = sunny
Gain(S_{sunny}, Humidity) = 0.970

- Gain(S_{sunny}, Temperature) = 0.570

- Gain(S_{sunny}, Wind) = 0.019

Humidity has the highest gain; therefore, it is used as the decision node.

STEP 9: This process goes on until all data is classified perfectly or we run out of attributes.

VI. Conclusion

The studies and their implementation conducted here conclude that the decision tree learning algorithm ID3 works well on any classification problems having dataset with the discrete values [8]. Related to the research work it concludes that thus the classification tree built using ID3 algorithm is shown below. It tells if the weather was amenable to play cricket?

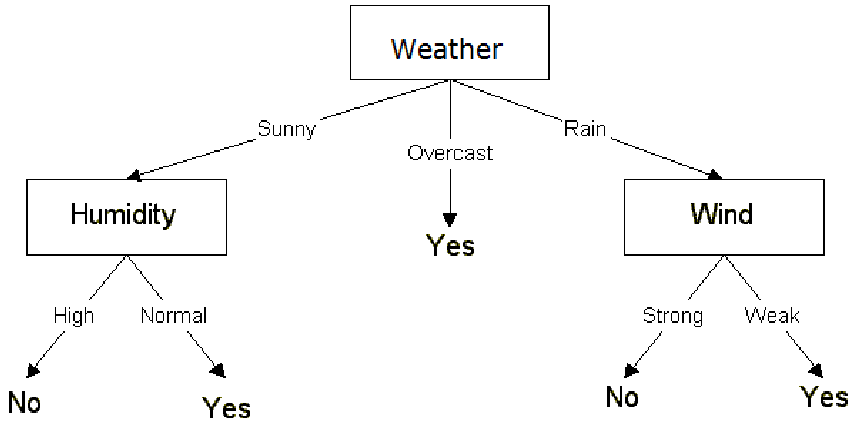

Figure 2: Final decision tree

The final rules are

1) If the weather is sunny and humidity is normal then play cricket but if humidity is high do not play cricket.
2) If the weather is overcast then play cricket.
3) If the weather is rainy and the wind is weak then play cricket but if wind is strong do not play cricket

References

[1] Tom M. Mitchell, (1997),*Machine Learning*, Singapore, McGraw-Hill.

[2] Paul E. Utgoff and Carla E. Brodley, (1990), *'An Incremental Methodfor Finding Multivariate Splits for Decision Trees'*, Machine Learning:Proceedings of the Seventh International Conference, (pp.58),Palo Alto, CA: Morgan Kaufmann.

[3] Wei Peng, Juhua Chen and Haiping Zhou, of ID3,*' An Implementation Decision Tree Learning Algorithm'*,University of New South Wales, School of Computer Science & Engineering,Sydney, NSW 2032, Australia.

[4] Quinlan, J.R. 1986, *Induction of Decision trees*, Machine Learning.

[5] http://www.cs.cornell.edu/Courses/cs578/2003fa/missing_featsel_lecture.ppt

[6] http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm

[7] Ruijuan Hu,(2011),*' Medical Data Mining Based on Decision Tree Algorithm',* Computer and Information Science,Vol. 4, No. 5; September 2011, www.ccsenet.org/cis.

[8] Anand Bahety,*' Extension and Evaluation of ID3 – Decision Tree Algorithm'*. University of Maryland, College Park.

[9] Vıctor.H.Garcıa, Raul Monroy and MaricelaQuintana,' Web Attack Detection Using ID3'. http://homepage.cem.itesm.mx/raulm/pub/ifip06.pdf.

[10] Sonika Tiwari and Prof. Roopali Soni,*' Horizontal partitioning ID3 algorithm A new approach of detecting network anomalies using decision tree'*, International Journal of Engineering Research & Technology (IJERT)Vol. 1 Issue 7, September – 2012.

[11] Mary Slocum,',*Decision making using ID3*,RIVIER ACADEMIC JOURNAL, VOLUME 8, NUMBER 2, FALL 2012

[12] Kumar Ashok, Taneja H C, Chitkara Ashok K and Kumar Vikas,*' Classification of Census Using Information Theoretic Measure Based ID3 Algorithm'* . Int. Journal of Math. Analysis, Vol. 6, 2012, no. 51, 2511 – 2518.

[13] L. Sathish Kumar and A. Padmapriya,*' Prediction for Common Disease using ID3 Algorithm in Mobile Phone and Television'*. International Journal of Computer Applications (0975 – 8887) Volume 50 – No.4, July 2012

[14] M. Suknovic et al(2011).*' Reusable components in decision tree induction Algorithms'*.

[15] Herbert Schildt, *Java 2: The Complete Reference*,5th edition, McGraw-Hill/Osborne

[16] http://www.wikipedia.com/

[17] http://www.google.com/