Volume 7, Issue 3, March 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey of Various Tree Based Classification Techniques

S. Surekha*
Department of CSE, JNTUK-University College of Engineering Vizianagaram,
Andhra Pradesh, India

DOI: 10.23956/ijarcsse/V7I3/0105

Abstract— This paper gives a survey of various decision tree based classification techniques for the classification of data.

Keywords— Classification; Decision Tree Classification;

I. INTRODUCTION

In real world, the facts are represented as an Information System [1] consisting of conditional attributes and decision attributes. The set of conditional attributes are the features that characterize the fact and the decision attribute describes the respective facts' category. Classification[2] is a supervised learning technique, which is used to predict the target class of an unknown record and this process involves two phases. The first phase in classification is building a classification model from the available set of facts. As the classification process is a supervised learning process, the collection of known facts used to build a classification model is known as the training data. The second phase uses the built model to classify an unknown record. There are two approaches to build a classifier: one is White box approach and the other is Black Box approach. In white box approach [2], the classifier reveals the details of how an unknown record has been classified. Whereas the Black box approach doesn't reveals the details about how it classifies an unknown record. As the white box classification model reveals the details of the classification process, it is being widely applied successfully in many applications like Weather forecasting, Disease diagnosis, Image classification e.t.c.

The rest of the paper is organized as follows, Section II explains the decision tree classification model, Section III gives the basic algorithm for inducing a decision tree and further discusses the evolution of various tree based classification algorithms. Finally, Section IV concludes the paper.

II. DECISION TREE BASED CLASSIFICATION

Decision Tree Classification [3] is one of the widely used white box classification technique. Decision Tree classification technique uses a Tree like structure in building the classification model. The generated decision tree consists of three kinds of nodes namely the Root node, Internal nodes and Leaf nodes. The Root node is the origin of the decision tree without any incoming edges but with outgoing edges. Internal nodes are the nodes with exactly one incoming edge and one or more outgoing edges. Leaf nodes are the nodes with no outgoing edges but with exactly one incoming edge. The label of the Root and Internal nodes represents the names of the conditional attributes of the given training data and the label of the Leaf nodes represents the decision class. The decision class of an unknown record is classified by submitting the conditional features of the unknown record to the obtained decision tree by traversing the tree starting from the root node and based on the outcome; traverse the path to arrive at a Leaf node. Finally, the label of the leaf node at which the path is terminated will be taken as the class of the given unknown record. Fig. 1 shows a sample decision tree for Shapes dataset.

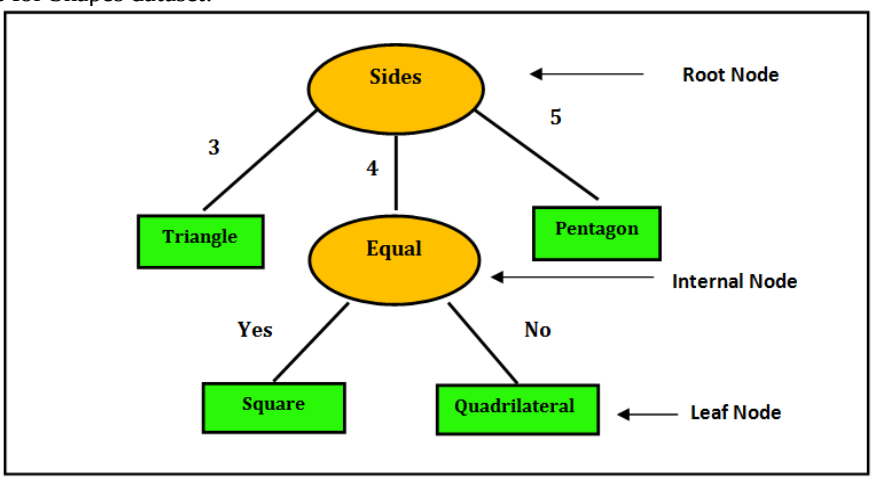


Fig. 1 An example of Decision Tree

The performance of any decision tree classification technique is measured in terms of the number of leaf nodes generated, total number of nodes including the internal nodes and leaf nodes called as the tree size, and the ability of the decision tree to classify an unknown record called as the Prediction accuracy.

III. DECISION TREE INDUCTION

The sequence of steps to induce a decision tree is shown in Fig 2.

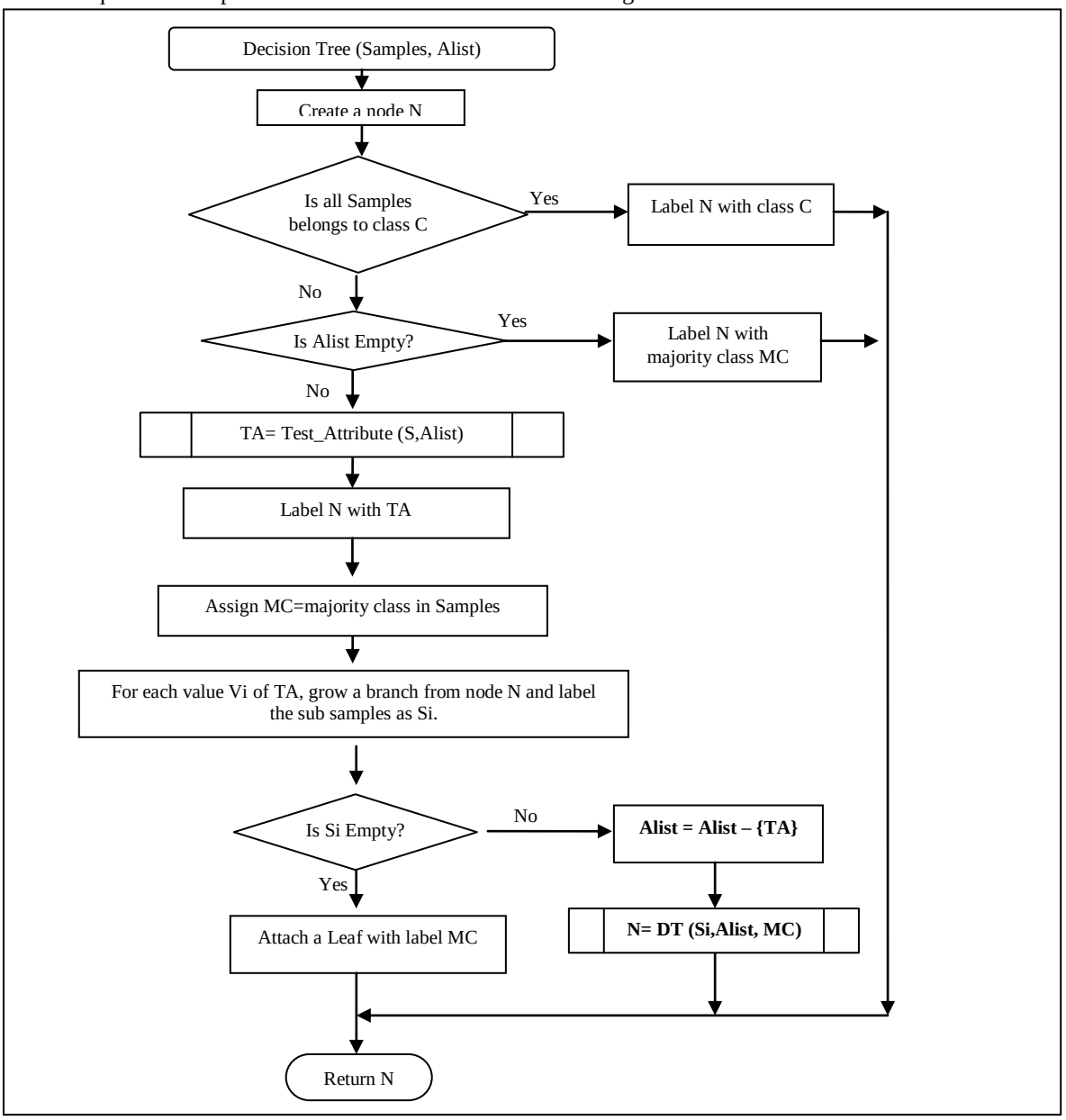


Fig. 2 Decision Tree induction process

In the process of inducing a decision tree, one of the attribute in the attributes list should be selected as the splitting attribute and based on the possible values of the splitting attribute the tree induction process continues. The efficiency of the decision tree induction algorithm depends on selecting the best attribute as the splitting attribute. There exist several measures like Entropy, Information-gain, Gain-ratio, Gini-index, Chi-square, etc. This section discusses various Decision tree classification algorithms, and all these algorithms use different measures to select the splitting attribute.

A. ID3 Decision Tree Classification Algorithm

ID3 decision tree classification algorithm was proposed by Quinlan. ID3 uses the statistical Entropy [3] based measure called the Information-Gain [3] as the measure to select the splitting attribute.

Definition:

For a given InformationSystem IS=(U,{C∪D}), for an A∈C with '*n*' distinct values represented as $\{v_1, v_2, \ldots, v_n\}$ and D with '*m*' distinct classes. Then the Entropy and Information-Gain of attribute 'A' are defined as,

$$Entropy(U) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \times log_2^{\frac{|S_i|}{|S|}} \quad (1)$$

$$Entropy(A) = \sum_{j=1}^{n} \frac{|s_{1j}|+|s_{2j}|+\cdots+|s_{nj}|}{|S|} * Entropy(S_j) \quad (2)$$

$$Information - Gain(A) = Entropy(U) - Entropy(A) \quad (3)$$

Where, S_i represents the facts that belongs to decision class i,

S_{ij} represents the facts that belongs to decision class i and having value V_j for attribute A,

S represents the universe of facts.

In the process of inducing ID3 decision tree, the attribute with highest Information-Gain value will be selected as the splitting attribute.

1) Example:

To explain the ID3 decision tree generation algorithm, a sample FLU dataset with 8 objects is taken into consideration. The FLU dataset is consisting of two conditional attributes namely Headache(H) and Temperature(T). There are two decision classes, which helps to predict whether the patient is suffering with FLU or not suffering with FLU. The sample dataset is given in TABLE I.

Table I Sample FLU Dataset

U	Feature Set		
	HeadAche	**Temperature**	**Decision**
1	Yes	Normal	No-Flu
2	Yes	High	Flu
3	Yes	Very-High	Flu
4	No	Normal	No-Flu
5	No	High	No-Flu
6	*No*	*Very-High*	*Flu*
7	No	High	No-Flu
8	*No*	*Very-High*	*No-Flu*

Sample size = 8

Number of decision classes = 2 i.e., Flu or No-Flu

Now, the Entropy of attribute Headache can be calculated using (3) as follows,

$$Information - Gain(H) = Entropy(U) - Entropy(H)$$

The Entropy of attribute Headache can be calculated using (1) and (2) as follows,

$$Entropy(U) = -\left(\frac{3}{8}\right) log_2 \left(\frac{3}{8}\right) - \left(\frac{5}{8}\right) log_2 \left(\frac{5}{8}\right) = 0.954434$$

The number of possible values of Headache are 2 namely yes and no, then

$$Entropy(H) = \frac{3}{8} Entropy(H = yes) + \frac{5}{8} Entropy(H = no)$$

$$Entropy(H = yes) = -\left(\frac{2}{3}\right) log_2 \left(\frac{2}{3}\right) - \left(\frac{1}{3}\right) log_2 \left(\frac{1}{3}\right) = 0.344361$$

$$Entropy(H = no) = -\left(\frac{1}{5}\right) log_2 \left(\frac{1}{5}\right) - \left(\frac{4}{5}\right) log_2 \left(\frac{4}{5}\right) = 0.451205$$

$$Entropy(H) = \frac{3}{8} Entropy(H = yes) + \frac{5}{8} Entropy(H = no)$$

$$= \frac{3}{8} * 0.344361 + \frac{5}{8} * 0.451205 = 0.795566$$

Therefore, $Information - Gain(H) = 0.158868$

Similarly calculating the Information-Gain for the Temperature and is obtained as,

$$Information - Gain(T) = 0.265712$$

The attribute *Temperature* is having the highest Information-Gain value and hence, it is selected as the root node. Then inducing decision tree according to the values of Temperature, the tree at root level is obtained as shown in Fig 3 and the final decision tree is obtained as shown in Fig 4.

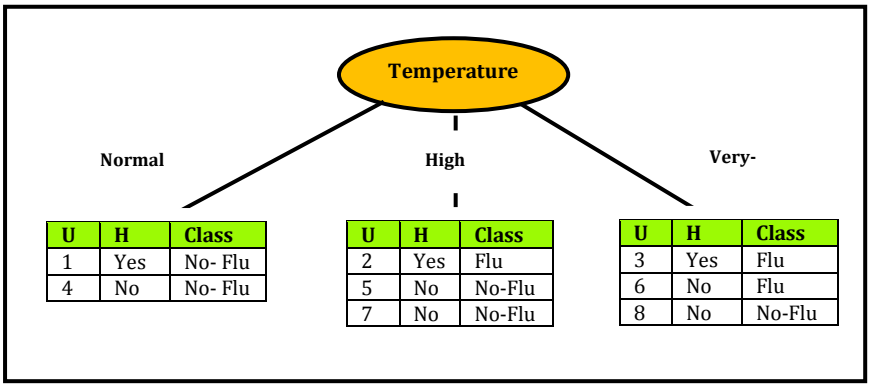


Fig. 3 Partial Decision Tree induced by ID3 algorithm

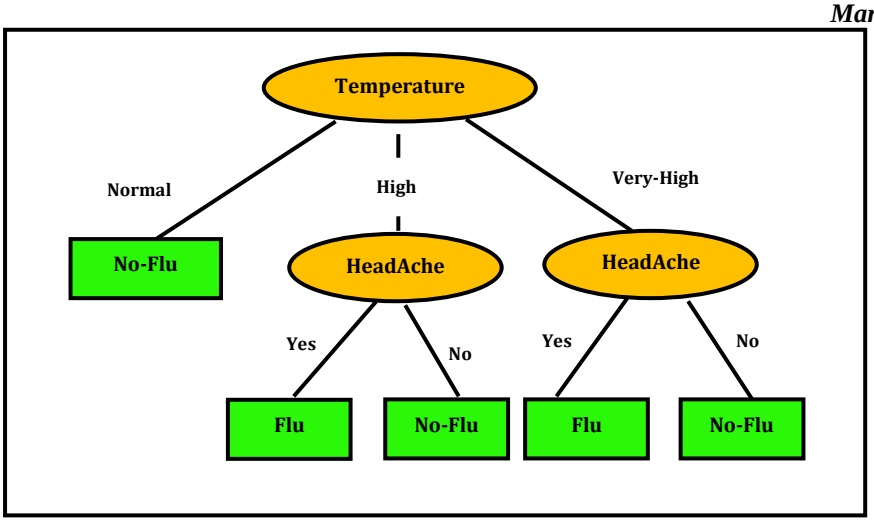


Fig. 4 final Decision Tree induced by ID3

The number of leaves nodes generated by the ID3 algorithm for the data in TABLE I is 5 and the tree size is 8.

B. C4.5 Decision Tree Classification Algorithm

C4.5 decision tree classification algorithm was proposed by Quinlan, is an improved version of ID3, which uses Gain-Ratio[4] as the measure to select the splitting attribute. The Gain-Ratio value of an attribute A is defined as,

$$Split - Information(A) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log_2 \left(\frac{|S_i|}{|S|}\right) \quad (4)$$

Where, Si represents the number of facts having value 'vi' for the attribute A.

$$Gain - Ratio(A) = \frac{Information-Gain(A)}{Split-Information(A)} \quad (5)$$

In the process of inducing C4.5 decision tree, the attribute with highest Gain-Ratio value will be selected as the splitting attribute.

Example:

The Gain-Ratio of attribute Headache can be calculated using (5) as follows,

$$Gain - Ratio(Headache) = \frac{Information-Gain(Headache)}{Split-Information(Headache)}$$

From the previous section 3.2, the Information-Gain (Headache) is 0.158868, and the Split-Information(Headache) can be obtained using (4) as follows,

$$Split - Information(Headache) = -\left(\frac{3}{8}\right) log_2 \left(\frac{3}{8}\right) - \left(\frac{5}{8}\right) log_2 \left(\frac{5}{8}\right) = 0.954434$$

$$Split - Information(Temperature) = -\left(\frac{2}{8}\right) log_2 \left(\frac{2}{8}\right) - \left(\frac{3}{8}\right) log_2 \left(\frac{3}{8}\right) - \left(\frac{3}{8}\right) log_2 \left(\frac{3}{8}\right) = 0.954434$$

Therefore,

$$Gain - Ratio(Headache) = \frac{0.158868}{0.954434} = 0.166452$$

And

$$\mathbf{Gain - Ratio(Temperature) = \frac{0.265712}{1.561278} = 0.170189}$$

The attribute Temperature is having highest Information-Gain and hence, Temperature is selected as the root node. The final decision tree induced by C4.5 algorithm is obtained as shown in Fig 4.

C. Rough Set Theory based Decision Tree Classification Algorithm

Rough Set Theory (RST)[5] is an intelligent mathematical tool introduced by Z.Pawlak to deal with uncertainties in data. Wei[6] incorporated the concepts of RST in the process of inducing decision trees. The basic concepts of RST can be found in the literature[5,6,9,10].

Rough Set Theory (RST)[5] is an intelligent mathematical tool introduced by Z.Pawlak to deal with uncertainties in data. Wei[6] incorporated the concepts of RST in the process of inducing decision trees. The basic concepts of RST can be found in the literature[5,6,9,10].

In RST based approach of decision tree classification technique, the attribute with highest size of Explicit region is selected as the splitting attribute. For an Information system defined in section 3.1, the definition for explicit region of an attribute C with respect to the decision attribute D is,

$$ER_C(D) = \bigcup_{D_i \in D^*} \underline{C}(D_i) \quad (6)$$

Where, $\underline{C}(D_i)$ is the Lower approximation[6] of D_i with respect to C

1) Example:

The equivalence classes generated by the attributes Headache(H),Temperature(T) and Decision(D) is,

$\frac{U}{H}$ = {{1,2,3},{4,5,6,7,8}}

$\frac{U}{T}$ = {{1,4}, {2,5,7}, {3,6,8}}

$\frac{U}{D}$ = {{2,3,6}, {1,4,5,7,8}} i.e., D_1={2,3,6} D_2={1,4,5,7,8}

The Explicit regions of the attribute Headache(H) and Temperature(T) can be calculated using (6) as follows,

$$ER_H(D) = \underline{H}(D_1) \bigcup \underline{H}(D_2)$$

The lower-approximation of D with respect to H is obtained as,

$$\underline{H}(D_1) = \Phi \quad \text{and} \quad \underline{H}(D_2) = \Phi.$$

Therefore, $ER_H(D) = \Phi$

The lower-approximation of D with respect to T is obtained as,

$$ER_T(D) = \underline{T}(D_1) \bigcup T(D_2)$$

$\underline{T}(D_1) = \Phi$ and $\underline{T}(D_2)$ = {1,4} since {1,4} ⊆ {1,4,5,7,8}

Therefore, $ER_T(D) = \{1,4\}$

The size of the explicit region of Temperature is highest and hence the decision tree induced by the RST approach is also same as that of the tree shown in Fig 4.

D. Variable Precision Rough Set Model based Decision Tree Classification Algorithm

Variable Precision Rough Set Model (VPRSM)[8] is an enhanced version of RST which allows some degree of misclassification. The misclassification error rate [9] is represented by β and the range of β is in between 0 and 0.5. The basic concepts of VPRSM can be found in the literature [8,9].

In VPRSM based approach of decision tree classification, the attribute with maximum size of Variable Precision Explicit region will be selected as the splitting attribute. The Variable Precision Explicit Region for an attribute C with respect to a decision D can be defined as,

$$\beta - ER_C(D) = \cup_{D_i \in D} \underline{C}_\beta(D_i) \qquad (6)$$

Where, $\underline{C}_\beta(D_i)$ is the β-Lower approximation[8] of D_i with respect to C.

Example:

Assume the misclassification error β=0.2. The variable precision explicit region for the attribute Headache(H) for the given β can be calculated using (6) as,

$$\beta - ER_H(D) = \underline{H}_\beta(D_1) \bigcup \underline{H}_\beta(D_2)$$

$\frac{U}{H}$ = {{1,2,3},{4,5,6,7,8}} i.e., H_1={1,2,3} and H_2={4,5,6,7,8}

$\frac{U}{D}$ = {{2,3,6}, {1,4,5,7,8}} i.e., D_1={2,3,6} and D_2={1,4,5,7,8}

The Relative classification error[8] for the set H_1 with respect to D_1 is,

C (H_1, D_1) = 1- $|H_1 \cap D_1| / |H_1|$ = 1-2/3 = 0.33

and C (H_1, D_2) =1-1/5 = 0.8

Either the relative classification error of the set H_1 or H_2 is less than the allowable degree of misclassification. So, $\underline{H}_\beta(D_1) = \Phi$

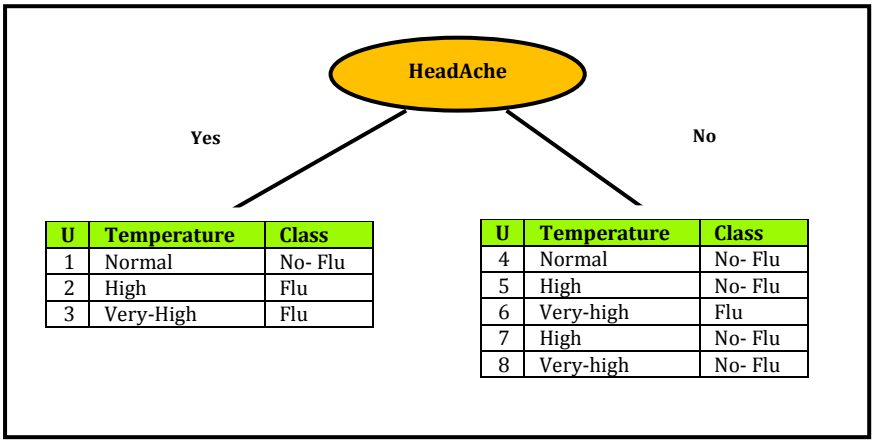
Similarly, C (H_2, D_1) =1-1/3=0.66 **C (H_2, D_2) =1-4/5=0.2 (0.2 ≤ β)**

Therefore, $\underline{H}_\beta(D_2)$ = {H_2} = {4,5,6,7,8}

The variable precision explicit region of H is obtained as $\beta - ER_H(D) = \{4,5,6,7,8\}$

The variable precision explicit region of T is $\beta - ER_T(D) = \{1,4\}$

The VPER of attribute H is highest and hence H is selected as the root node and the partial decision tree generated by VPRSM approach is shown in Fig 5, and the final decision tree is shown in Fig 6.



U	Temperature	Class
1	Normal	No- Flu
2	High	Flu
3	Very-High	Flu

U	Temperature	Class
4	Normal	No- Flu
5	High	No- Flu
6	Very-high	Flu
7	High	No- Flu
8	Very-high	No- Flu

Fig. 5 Partial Decision Tree induced by VPRSM approach

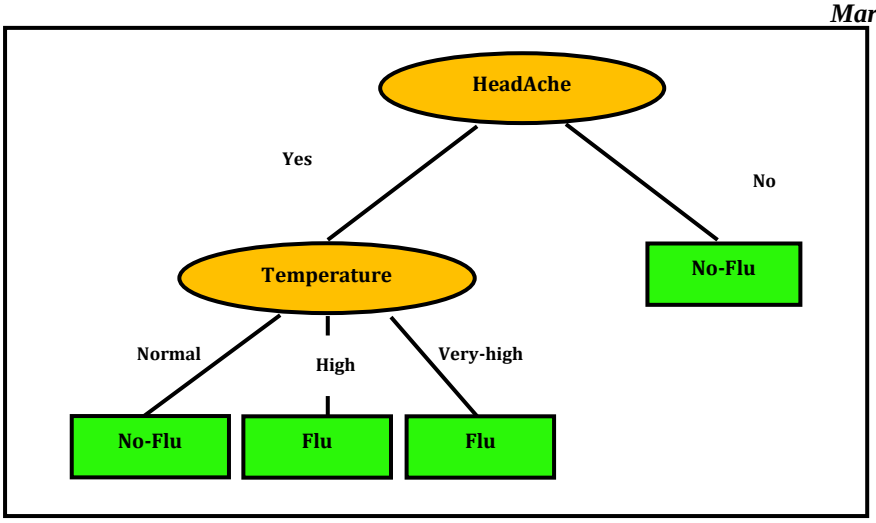


Fig. 6 Final Decision Tree induced by VPRSM approach

The number of leaf nodes generated by the tree induced using VPRSM approach is 4 and the tree size is 6.

The overall comparison of the performance of the decision tree classifier induced by ID3, C4.5, RST, and VPRSM approaches is shown in Fig 7.

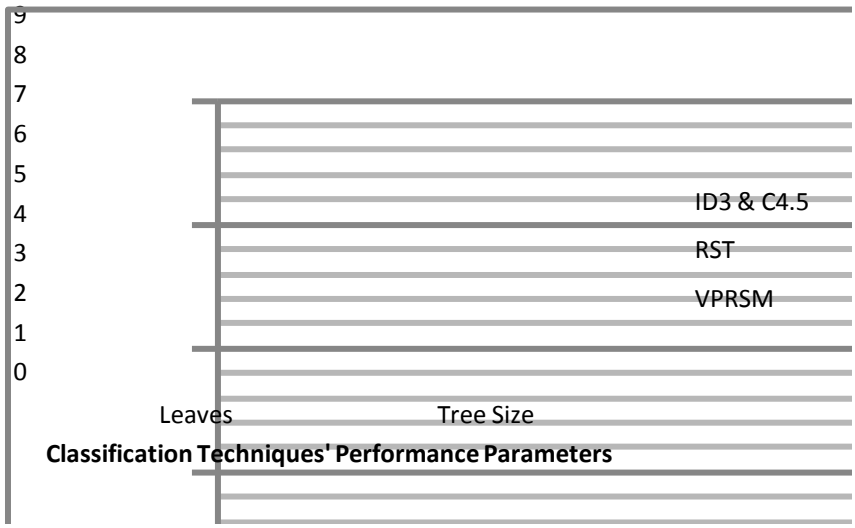


Fig. 7 Comparison of various decision tree induction algorithms

From Fig 7, it is observed that the tree induced by the VPRSM approach is generating trees with minimum nodes i.e., the number of conditions required to arrive at a decision is less when compared to the other approaches.

IV. CONCLUSIONS

In this paper, various algorithms for inducing decision trees are explained for a sample FLU dataset. From the worked out examples, it is very clear that decision tree induction based on VPRSM approach is generating very optimal decision trees from uncertain data.

REFERENCES

[1] Z Pawlak, "Rough Set approach to Knowledge based Decision-Support", *European Journal of Operational Research*, pp.48-57, 1997.

[2] Rokach, L., & Maimon, O, *Datamining with decision trees: theory and applications*. World Scientific Publishing Co. Pte. Ltd., Singapore. 2008.

[3] J R Quinlan, "Introduction of Decision Trees ", *Machine Learning 3*, 1986, pp. 81-106.

[4] Tom M. Mitchell, "*Machine Learning*", Singapore, McGraw-Hill, 1997.

[5] Pawlak, Z, Rough set approach to knowledge-based decision support. *European Journal of Operational* Research, 1997,99(1), pp.48-57.

[6] Jin Mao Wei," Rough Set Based Approach to Selection of Node", *International Journal of Computational Cognition*, Vol. 1, 2003, No. 2, pp. 25–40.

[7] J M Wei, S Q Wang, M Y Wang, J P You, and D Y Liu, " Rough Set based approach for inducing decision trees", *Knowledge-Based Systems*, Vol.20, 2007, pp.695-702.

[8] W.Ziarko, Variable Precision Rough Set Model", *Journal of Computer and System Sciences* Vol.46, 1993, pp.39-59.

[9] Jin-Mao Wei, Ming-Yang Wang, Jun-Ping You, "VPRSM Based Decision Tree Classifier", *Computing and Informatics*, Vol. 26, 2007, 663–677

[10] Surekha S, Jaya Suma G, "Comparison of Feature Selection Techniques for Thyroid disease", *In proceedings of ICICMT'2015*, 2015, pp.20-26.

[11] Asuncion, A., & Newman, D.J. *UCI Machine Learning Repository,* 2007, [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science.

