

# Rule-Based Classification using Multi-Soft Set Theory



Santhosh Kottam, Varghese Paul

**Abstract---** One of the important data mining functionality is classification. Presently, different methods exist for implementing classification. Rule-based classification using decision tree induction method is a conventional and simple method for identifying an unknown class of a given object. This method has a set of demerits and to remove these demerits, we depend on a soft computing tool which is known as soft set theory. One branch of soft set theory is called - multi soft theory- and it has a wide range of applications in the area of classification. We made a certain alteration in the rule-based classification using decision tree induction method by applying multi soft set theory. These changes will simplify the difficulties of the rule-based classification using decision tree induction method. The first two sections of this research work discuss introduction and preliminaries. In the remaining sections, the authors describe the multi soft set theory and its applications in rule base classification. Lastly, the paper finishes with a new algorithm, which the research scholars implemented as software using python programming. The suggested work experts can use in data mining industry. It has massive use in the fields of business, agriculture, health, education and many more.

**Keywords:** : Soft Computing, Soft set, Multi soft set, Decision tree induction, Information gain, classification, Rule-based classification.

## I. INTRODUCTION

Data mining is not a single task; it consists of a set of different operations known as Data mining functionalities. Association analysis, clustering, outlier analysis, and classification are popular functionalities widely used in the information mining world. Classification is the process of predicting the identity of an object whose class label is unknown. Classification is a two-step process. In the first step, we construct a model from the training data set and in the second step; the derived model is used for declaring the identity of an unknown object. In the model construction process, data mining experts are deriving or using an algorithm for predicting class labels.

Currently, algorithms like - decision tree induction, Bayesian classification, Rule-based classification and classification by backpropagation are using for classification. Another branch of computing is emerged and continuing its growth by simplifying the implementation complexity of different functionalities of data mining. Collectively this computing is known as soft computing. It consists of following branches – fuzzy set, rough set, genetic algorithms, and soft set. Rule-based classification using decision tree induction is the simplest method for predicting the class of an unknown entity. Now a day this technique is very popular, because it does not need a domain knowledge and parameter setting. It has certain demerits. One of the well known soft computing tools is soft set theory, which is invented by Russian mathematician D. Molodtsov [1]. To overcome the difficulties of rule-based classification using decision tree induction method, we propose an alternative method with the help of multi-soft set theory

## II. PRELIMINARIES

### A. Classification

Classification is consists of two phases. In the first phase a classifier is develop explaining a predefined set of data objects. This phase is known as training phase and here a classification method develops the classifier by studying a training set made up of database records and their related class labels. A record,  $\mathbf{A}$ , is described by an n-dimensional attribute vector,  $\mathbf{A} = (a_1, a_2, \dots, a_n)$ , representing n values made on the record from n database attributes, respectively,  $X_1, X_2, \dots, X_n$ . Each record,  $\mathbf{A}$ , is assumed to belong to a class called class label attribute. The class label attribute is discrete-valued and unordered. It is categorical in that each value serves as a category or class. The individual records building up the training set are called as training tuples and are chosen from the dataset under study. In the second phase, the model is chosen for classification. Initially, the predictive accuracy of the model is estimated using a test data set, made up of test records and their connected class labels. These records are arbitrarily elected from the common database. They are not dependent of the training records, meaning that they are not used to build the model. The correctness of a model on a given test set is the percentage of test set records that are correctly classified by the model. The related class label of each test record is compared with the experienced classifier's class prediction for that record. If the correctness of the classifier is considered adequate, the model can be used to classify future data records for which the class label is not known [2].

Revised Manuscript Received on February 13, 2020.

\* Correspondence Author

**Santhosh Kottam\***, Research Scholar, Research and Development Centre, Bharathiyar University, Coimbatore, India. E-mail: sankottam@gmail.com.

**Dr. Varghese Paul**, Professor & HOD, Department of Information Technology, Cochin University of Science & Technology, Thrikkakara, Kochi, India. E-mail: [vp.iticusat@gmail.com](mailto:vp.iticusat@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

**B. Decision tree induction method**

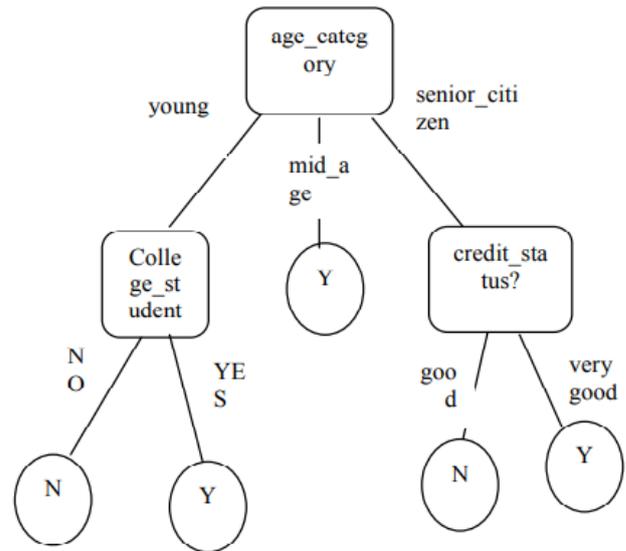
Decision tree induction is the generation of decision trees from class labelled training records. A decision tree keeps a tree structure, where each non leaf node denotes an examination on an attribute, each division shows a result of the examination, and each leaf node gives a class label. The uppermost node in a decision tree is the root node. Consider a record,  $Y$ , for which the related class label is not known, the attribute values of the record are evaluated against the decision tree. A path is derived from the root to a leaf node, which shows the class result for that record. In a certain situation, decision trees can easily be migrated to classification rules.

**Table 1: Class Labeled Records**

| TI<br>D | colle<br>ge_st<br>udent | credit_sta<br>tus | income<br>_info | age_cate<br>gory   | purcha<br>se_co<br>mputer |
|---------|-------------------------|-------------------|-----------------|--------------------|---------------------------|
| 1       | NO                      | good              | good            | young              | N                         |
| 2       | NO                      | very_good         | good            | young              | N                         |
| 3       | NO                      | good              | good            | mid_age            | Y                         |
| 4       | NO                      | good              | average         | senior_ci<br>tizen | Y                         |
| 5       | YES                     | good              | poor            | senior_ci<br>tizen | Y                         |
| 6       | YES                     | very_good         | poor            | senior_ci<br>tizen | N                         |
| 7       | YES                     | very_good         | poor            | mid_age            | Y                         |
| 8       | NO                      | good              | average         | young              | N                         |
| 9       | YES                     | good              | poor            | young              | Y                         |
| 10      | YES                     | good              | average         | senior_ci<br>tizen | Y                         |
| 11      | YES                     | very_good         | average         | young              | Y                         |
| 12      | NO                      | very_good         | average         | mid_age            | Y                         |
| 13      | YES                     | good              | good            | mid_age            | Y                         |
| 14      | NO                      | very_good         | average         | senior_ci<br>tizen | N                         |

The development of decision tree model does not require any specific knowledge or attribute setting, and therefore it is appropriate for elaborated knowledge discovery. Decision trees can handle high dimensional data. Their illustration of acquired knowledge in tree form is sensitive and generally easy to understand by humans. The learning and classification steps of decision tree induction are simple and fast. In general, decision tree techniques have good accuracy. During tree development, attribute selection procedures are used to select the attribute that best divides records into different class classes. Decision tree developed from table 1 is given in fig 1.

An attribute selection process includes searching for choosing the splitting criterion that “excellent” divides a given data partition,  $DT$ , of class-labeled training records into individual class labels. If we were to split  $DT$  into smaller divisions according to the outcomes of the splitting criterion, then each partition would be pure. The attribute selection process gives a ranking for each attribute defining the given training records. The tree node formed for partition  $DT$  is identified with the splitting criterion, branches are developed for each result of the criterion, and the records are divided orderly. Three well known attribute selection process are—*information gain*, *gain ratio*, and *gini index*.



**Fig. 1 Final constructed decision tree.**

Decision tree induction has certain limitations [3]

1. A minute updating in the dataset can affect a big alteration in the arrangement of the decision tree causing unsteadiness.
2. For a decision tree computation can set out far more difficult compared to other methods.
3. Decision tree often take large time to build up the model.
4. Decision tree preparation is comparatively expensive as difficulty and time taken is more.
5. Decision Tree method is insufficient for applying regression and predicting continuous values.

**C. Information gain**

This attribute evaluation procedure is used information theory contributed by Claude Shannon, which calculated the worth or “information content” of messages. Let node  $M$  stand for or hold the records of partition  $DT$ . The element with the maximum information gain is selected as the splitting element for node  $M$ . This element reduces the information required to classify the records in the ensuing partitions and shows the least randomness or “impurity” in these divisions. This method reduces the anticipated number of experiments required to classify a given record and ensure that a simple tree is generated.

The anticipated information required to classify a record in  $DT$  is shown by

$$Info(DT) = - \sum_{j=1}^n qi \log_2(qi)$$

here  $qi$  is the probability that a random record in  $DT$  attach to class  $Xi$  and is determined by  $|Xi, DT|/|DT|$ . Here information is encoded in bits format.  $Info(DT)$  stands for average amount of information required to recognize the class label of a record in  $DT$ .  $Info(DT)$  is also known as the entropy of  $DT$ . Next, we partition the records in  $DT$  on the element  $R$  having  $u$  different values,  $\{r1, r2, \dots, ru\}$ , these values are collected from the training data. If  $R$  is discrete-valued, these values match directly to the  $u$  outcomes of a test on  $R$ . Element  $R$  can be used to split  $DT$  into  $u$  partitions or subsets,  $\{DT1, DT2, \dots, DTu\}$ , where  $DTj$  includes those records in  $DT$  that have outcome  $rj$  of  $R$ .



These partitions would correspond to the branches grown from node  $M$ . After dividing the entire records base on the attribute  $R$ , next we decide, volume of information that we need to reach at exact solution. This amount is determined by

$$InfoR(DT) = \sum_{j=1}^u \frac{|DT_j|}{|DT|} \times Info(DT_j)$$

The term  $\frac{|DT_j|}{|DT|}$  denotes the influence of the  $j^{th}$  division.  $InfoR(DT)$  is the anticipated information needed to distinguish a record from  $DT$  based on the division by  $R$ .

Information gain is calculated as

$$Gain(R) = Info(DT) - InfoR(DT)$$

It is defined as the difference between the original information requirement and the new requirement.

#### D. Rule-based classification

In rule-based classifiers, the developed model is presented as a set of IF-THEN rules. An IF-THEN rule syntax is

$$IF \text{ exp THEN class\_label.}$$

The “IF”-part of a rule is known as the antecedent and the “THEN”-part is the consequent. In the rule antecedent, the condition consists of one or more *attribute tests* that are logically ANDed. The rule’s consequent contains a class prediction. A rule  $RL$  can be qualified by its coverage and accuracy. Given a record,  $A$ , from a class labeled data set,  $DT$ , let  $N_{covers}$  be the number of records covered by  $RL$ ;  $N_{correct}$  be the number of records correctly classified by  $RL$ ; and  $|DT|$  be the number of records in  $DT$ . We can define the coverage and accuracy of  $RL$  as

$$coverage(RL) = \frac{N_{covers}}{|DT|}$$

$$accuracy(RL) = \frac{N_{correct}}{N_{covers}}$$

Rule-based classifier; we can build in two ways-Rule extraction from decision tree and rule induction using a sequential covering algorithm. First method is an accepted technique of classification. Even though decision trees are well known for its accuracy, but it is very difficult to understand. The IF-THEN rules are easier for data mining experts to understand, mainly a situation where decision tree is very large in size. According to this method, one rule is extracted for each path from the root node to a leaf node. Each splitting criterion along a specified path is logically ANDed to form the rule antecedent (“IF” part). The leaf node holds the class result, forming the rule consequent (“THEN” part).

From the fig 1, following rules are derived using the decision tree method.

**RL1:** IF age\_category = *young* AND college\_student = *NO* THEN purchase\_computer = *N*

**RL2:** IF age\_category = *young* AND college\_student = *YES* THEN purchase\_computer = *Y*

**RL3:** IF age\_category = *mid\_age* THEN purchase\_computer = *Y*

**RL4:** IF age\_category = *senior\_citizen* AND credit\_status = *very\_good* THEN purchase\_computer = *Y*

**RL5:** IF age\_category = *senior\_citizen* AND credit\_status = *good* THEN purchase\_computer = *N*

Consider rule  $RL2$ , it has coverage of 2 out of 14 records. It can correctly classify both records. Therefore, *coverage* ( $RL2$ ) =  $2/14 = 14:28\%$  and *accuracy* ( $RL2$ ) =  $2/2 = 100\%$ .

### III. MULTI-SOFT SET THEORY

Soft computing is not a single task; it is a group of concepts that executes unanimously to cope with real-world problems. It has a well defined data execution capability for managing problems having uncertainty, ambiguity, and approximate reasoning [4]. In the data mining world following soft computing theories - genetic algorithms, neural networks, fuzzy sets, and rough sets - are commonly used for retrieving relevant data from huge repositories. All these theories have certain drawbacks, as given out in [1]. Causes for these drawbacks is an insufficiency of the parameterization tool of these theories. Well known Russian mathematician, Molodtsov contributed a new mathematical concept called soft set theory. It has the capability to dealing uncertainties and free from the above difficulties.

#### A. Definition and Preliminaries

##### Definition 1 (Soft Set)

A pair  $(H, Y)$  is called a soft set over  $Q$ , where  $H$  is a mapping given by

$$H : Y \rightarrow P(Q)$$

In other words, a soft set over  $Q$  is a parameterized family of subsets of the universe  $Q$ .

Every set  $H(\varepsilon), \varepsilon \in Y$  may be treated as the set of  $\varepsilon$ -elements of the soft set  $(G, Y)$ . A soft set is not a set.

**Proposition 1.** If  $(H, Y)$  is a soft set over the universe  $S$ , then  $(H, Y)$  is a binary valued information system  $Q = (S, A, V_{\{0,1\}}, h)$ .

**Proof.** Let  $(H, Y)$  be a soft set over the universe  $S$ , we define a mapping

$$H = \{h_1, h_2, \dots, h_n\},$$

Where

$$h_i : S \rightarrow V_i \text{ and } h_i(x) = \begin{cases} 1 & x \in H(y_i) \\ 0 & x \notin H(y_i) \end{cases}, \text{ for } 1 \leq i \leq n$$

Hence, if  $A=Y, V = \bigcup_{ei \in A} V_{ei}$  where

$V_{ei} = \{0,1\}$ , then a soft set  $(H, Y)$  can be considered as a

binary-valued information system  $Q = (S, A, V_{\{0,1\}}, h)$ .

Commonly, soft set is a relationship from parameter to the crisp subset of universal set. This relationship categorizes the objects into two classes yes or no. In the real world problems, a parameter may have a collection of values.

## Rule-Based Classification using Multi-Soft Set Theory

For example, the Zoology degree of candidate can be categorized into three values; best, average and worst. Here, every parameter decides a partition of the universal set which contains more than two disjoint subsets. To implement this, experts describe a multi-valued information system to  $n$  binary-valued information systems based on each parameter, where each soft set can be defined [5].

Consider a multi-valued information system  $Q = (S, A, V, h)$ , where  $S$  is a non-empty finite set of objects,  $A$  is a non-empty finite set of attributes,  $V = \bigcup_{a \in A} V_a$ ,  $V_a$  is the domain of attribute  $a$  which has multi value ( $|V_a| \geq 3$ ) and  $h: S \times A \rightarrow V$  is a total function such that  $h(s, a) \in V_a$ , for every  $(s, a) \in S \times A$ , called information function, there are  $(|A|)$  number of binary-valued information systems  $Q = (S, A, V_{\{0,1\}}, h)$  can be generated, where  $|A|$  represents the cardinality of  $A$ . The  $|A|$  size of binary-valued information systems describes multi-soft sets of the multi-valued information systems.

Consider a multi-valued information system  $Q = (S, A, V, h)$  and  $Q^i = (S, a_i, V_{a_i}, h)$ ,  $i=1, 2, \dots, |A|$  be the  $|A|$  binary-valued information systems. From proposition 1.

$$Q = (S, A, V, h) = \begin{cases} Q^1 = (S, a_1, V_{\{0,1\}}, h) & \iff (H, a_1) \\ Q^2 = (S, a_2, V_{\{0,1\}}, h) & \iff (H, a_2) \\ \vdots & \vdots \\ Q^{|A|} = (S, a_{|A|}, V_{\{0,1\}}, h) & \iff (H, a_{|A|}) \end{cases}$$

$$= ((H, a_1), (H, a_2), \dots, (H, a_{|A|}))$$

We can conclude  $(H, Y) = ((H, a_1), (H, a_2), \dots, (H, a_{|A|}))$  as a multi soft set over universe  $S$  representing a multi-valued information system  $Q = (S, A, V, f)$ .

### IV. MULTI SOFT SET IN CLASSIFICATION

Multi soft set for table 1 is given below

$$(H, Y) = ((H, \text{college\_student}), (H, \text{credit\_status}), (H, \text{income\_info}), (H, \text{age\_category}), (H, \text{purchase\_computer})).$$

Its representation in multi soft set is

$$(H, Y) = \left( \begin{array}{l} \{\text{YES}=\{5,6,7,9,10,11,13\}, \text{NO}=\{1,2,3,4,8,12,14\}\} \\ \{\text{good}=\{1,3,4,5,8,9,10,13\}, \text{very\_good}=\{2,6,7,11,12,14\}\} \\ \{\text{good}=\{1,2,3,13\}, \text{average}=\{4,8,10,11,12,14\}, \text{poor}=\{5,6,7,9\}\} \\ \{\text{young}=\{1,2,8,9,11\}, \text{mid\_age}=\{3,7,12,13\}, \text{senior\_citizen}=\{4,5,6,10,14\}\} \\ \{Y=\{3,4,5,7,9,10,11,12,13\}, N=\{1,2,6,8,14\}\} \end{array} \right)$$

**Fig. 2 Multi soft set representation of table 1.**

Using multi soft set theory, following binary valued information systems are generated.

$$\begin{aligned} (H, \text{college\_student}) &= \{\text{YES}=\{5,6,7,9,10,11,13\}, \text{NO}=\{1,2,3,4,8,12,14\}\} \\ (H, \text{credit\_status}) &= \{\text{good}=\{1,3,4,5,8,9,10,13\}, \text{very\_good}=\{2,6,7,11,12,14\}\} \\ (H, \text{income\_info}) &= \{\text{good}=\{1,2,3,13\}, \text{average}=\{4,8,10,11,12,14\}, \text{poor}=\{5,6,7,9\}\} \\ (H, \text{age\_category}) &= \{\text{young}=\{1,2,8,9,11\}, \text{mid\_age}=\{3,7,12,13\}, \text{senior\_citizen}=\{4,5,6,10,14\}\} \\ (H, \text{purchase\_computer}) &= \{Y=\{3,4,5,7,9,10,11,12,13\}, N=\{1,2,6,8,14\}\} \end{aligned}$$

Commonly, a soft set  $(H, Y)$  over  $S$  can be treated as a partition type of the universe  $S$ . We can define  $H: Y \rightarrow P(S)$  as the division  $\{H(y), S \setminus H(y)\}$  for every  $y \in Y$ .

Apply information gain method for finding best splitting criterion. Following information gain values are calculated for each attribute of the data set  $DT$ .

$$\begin{aligned} \text{Gain}(\text{age\_category}) &= 0.246 \text{ bits}, \text{Gain}(\text{income\_info}) = 0.029 \text{ bits}, \\ \text{Gain}(\text{college\_student}) &= 0.151 \text{ bits and} \\ \text{Gain}(\text{credit\_status}) &= 0.048. \end{aligned}$$

Here  $\text{age\_category}$  has highest information gain, hence it is selected as a splitting element.

**Definition 2.** Let  $(H, Y) = ((H, a_i): i=1, 2, \dots, |A|)$  be a multi-soft set over  $S$  representing a multi-valued information system  $Q = (S, A, V, h)$ . The AND operation between  $(H, a_i)$  and  $(H, a_j)$  is defined as

$$(H, a_i) \text{ AND } (H, a_j) = (H, a_i \times a_j),$$

where

$$F(Va_i, Va_j) = H(Va_i) \cap G(Va_j), \forall (Va_i, Va_j) \in a_i \times a_j, \text{ for } 1 \leq i, j \leq |A|.$$

Use above definition and use  $\text{age\_category}$  as first splitting criterion

$$\begin{aligned} (H, \text{age\_category}) &= \{\text{young}=\{1,2,8,9,11\}, \text{mid\_age}=\{3,7,12,13\}, \\ &\text{senior\_citizen}=\{4,5,6,10,14\}\} \\ &\text{and} \\ (H, \text{purchase\_computer}) &= \{Y=\{3,4,5,7,9,10,11,12,13\}, \\ &N=\{1,2,6,8,14\}\} \end{aligned}$$

According to multi soft set AND operation

$$(H, \text{age\_category}) \text{ AND } (H, \text{purchase\_computer}) = (H, \text{age\_category} \times \text{purchase\_computer})$$

$$\left( \begin{array}{l} (\text{young}, Y) = \{9, 11\}, (\text{young}, N) = \{1, 2, 8\}, \\ (\text{mid\_age}, Y) = \{3, 7, 12, 13\}, (\text{mid\_age}, N) = \emptyset \\ (\text{senior\_citizen}, Y) = \{4, 5, 10\}, (\text{senior\_citizen}, N) = \{6, 14\} \end{array} \right)$$

**Fig. 3 Multi soft set AND operation result on age\_category and purchase\_computer.**

The records are divided into three partitions, records belonging into the division for age\_category=mid\_age all posses to the same class purchase\_computer=Y. Partitioning of this branch is end here and leaf node is labeled with “Y”. Select college\_student as next splitting criterion

(H,college\_student)=  
{ YES={ 5,6,7,9,10,11,13 },NO={ 1,2,3,4,8,12,14 } }  
(H, age\_category × purchase\_computer)  
AND (H,college\_student)=

$$\left( \begin{array}{l} (\text{young, YES, Y})=\{9,11\}, (\text{young, NO, Y})= \emptyset \\ (\text{young, YES, N})= \emptyset, (\text{young, No, N})=\{1,2,8\} \end{array} \right)$$

**Fig. 4 Multi soft set AND operation result on age\_category, college\_student and purchase\_computer.**

Proceeding branch is no need to partition recursively. All of the records in this branch are correctly classified. Next we consider the attribute credit\_status.

(H,credit\_status)={ good={ 1,3,4,5,8,9,10,13 },very\_good={ 2, 6,7,11,12,14 } }

(H, age\_category × purchase\_computer)AND(H,credit\_status)=

$$\left( \begin{array}{l} (\text{senior_citizen, good, Y})=\{4,5,10\}, (\text{senior_citizen, very_good, Y})= \emptyset \\ (\text{senior_citizen, good, N})= \emptyset, (\text{senior_citizen, very_good, N})= \{6,14\} \end{array} \right)$$

**Fig. 5 Multi soft set AND operation result on age\_category, credit\_status and purchase\_computer.**

Using next splitting criterion “credit\_status”, above partition completely classified all the records.

Final classification result using multi-soft set is

$$(R,X)= \left( \begin{array}{l} (\text{young, YES, Y})=\{9,11\}, \\ (\text{young, No, N})=\{1,2,8\} \\ (\text{mid_age, Y})=\{3,7,12,13\} \\ (\text{senior_citizen, good, Y})=\{4,5,10\}, \\ (\text{senior_citizen, very_good, N})= \{6,14\} \end{array} \right)$$

**Fig. 6 Multi soft set based classification result.**

From the multi-soft set(R,X) following classification rules are derived, these rules are same to the rules derived from decision tree induction method.

**RL1:** IF age\_category = young AND college\_student = NO THEN purchase\_computer = N

**RL2:** IF age\_category = young AND college\_student = YES THEN purchase\_computer = Y

**RL3:** IF age\_category = mid\_age THEN purchase\_computer = Y

**RL4:** IF age\_category = senior\_citizen AND credit\_status = very\_good THEN purchase\_computer = Y

**RL5:** IF age\_category = senior\_citizen AND credit\_status = good THEN purchase\_computer = N

Coverage and accuracy of the above rules are given below

**Table 2:Rule Coverage And Accuracy**

| COVERAGE      | ACCURACY   |
|---------------|------------|
| 2/14 = 14.28% | 2/2 = 100% |
| 3/14 = 21.43% | 3/3 = 100% |
| 4/14 = 28.57% | 4/4 = 100% |
| 3/14 = 21.43% | 3/3 = 100% |
| 2/14 = 14.28% | 2/2 = 100% |

Soft set (H,purchase\_computer)={ Y={3,4,5,7,9,10,11,12,13},N={1,2,6,8,14} } is proved.

Algorithm: Develop classification rules from a multi-soft set, which is derived from the training records of the data set DT.

Input:

Data set, DT, which is a set of training records and their related class labels;

Attr\_list, the set of attributes;

Attribute selection method, a method to decide the splitting attribute that “best” divides the data records into separate classes. This measure consists of a dividing attribute and, probably, either a partition point or partitioning subset.

Output: A multi soft set which consists of class labels of records.

Method:

- (1) if records in DT are all of the same class, CL then
- (2) return end the algorithm and labeled all records with the class CL;
- (3) if attribute list is blank then
- (4) return labeled with the majority class in DT; // majority voting
- (5) create a multi soft set for the given data set;
- (6) apply Attribute selection method-(DT, attr\_list) to find the “best” partitioning criterion;
- (7) select attribute split\_attr as best splitting criterion;
- (8) if splitting attribute is discrete-valued and multi soft sets(H, split\_attr) for different values of split\_attr is generate
- (9) Multi soft set (H,attr\_CL) for attribute which determines class label is generate
- (10) Do multi soft set –logical AND -operation on soft sets generated in steps 8 and 9
- (11) attr\_list → attr\_list – split\_attr; // remove splitting attribute
- (12) for each outcome j of split\_attr does steps 7 to 11 until class labels of all records are known
- (13)Generate final multi-softset result (R, X)
- (14) Derive classification rules from the soft set (R,X)

**Fig 7. Rule-based classification algorithm using multi soft set theory**

V. EXPERIMENT RESULT

In this section, we assess the new soft set based classification algorithm with one popular method- Decision tree induction. The suggested method is implemented in the data sets[6] and [7]. The suggested method is typed in python programming language. These methods are executed repeatedly on a processor Intel Core i5-6200U CPU. The RAM is 4 GB and the OS is Windows 7 Professional.

A. Weather Dataset

Weather dataset is consists of five attributes- Outlook, Temperature, Humidity, Windy, and Play foot ball. These attribute are used for determining whether to play football or not and are classified into independent and dependent variables. The independent variables are Outlook, Temperature, Humidity, and Wind. The dependent variable is whether to play football or not. The attributes, temperature and humidity are considered as numerical. Remaining attributes are categorical, they cannot be ordered. Based on the training data set, we find a set of rules to identify what values of outlook, temperature, humidity and wind, decide whether or not to play foot ball. Table representing given dataset is given below.

Table 3:Weather Data Set

| TI D | Outlook  | Temperature | Humidity | Windy | Play ball |
|------|----------|-------------|----------|-------|-----------|
| 1    | Rainy    | Hot         | High     | FALSE | No        |
| 2    | Rainy    | Hot         | High     | TRUE  | No        |
| 3    | Overcast | Hot         | High     | FALSE | Yes       |
| 4    | Sunny    | Mild        | High     | FALSE | Yes       |
| 5    | Sunny    | Cool        | Normal   | FALSE | Yes       |
| 6    | Sunny    | Cool        | Normal   | TRUE  | No        |
| 7    | Overcast | Cool        | Normal   | TRUE  | Yes       |
| 8    | Rainy    | Mild        | High     | FALSE | No        |
| 9    | Rainy    | Cool        | Normal   | FALSE | Yes       |
| 10   | Sunny    | Mild        | Normal   | FALSE | Yes       |
| 11   | Rainy    | Mild        | Normal   | TRUE  | Yes       |
| 12   | Overcast | Mild        | High     | TRUE  | Yes       |
| 13   | Overcast | Hot         | Normal   | FALSE | Yes       |
| 14   | Sunny    | Mild        | High     | TRUE  | No        |

We start the process by determining the parent node for the proposed work. Apply information gain theory for finding best splitting criterion and collect nodes, which are arranged in decreasing order of information gain. Following information gain values are calculated for above dataset. Gain (outlook) =0.247 bits, gain(Temperature)=0.029 bits gain(Humidity)=0.152 bits and gain(windy)=0.048. Here outlook has highest information gain, hence it is selected as a splitting element.

Multi soft set for table 3 is given below

{Rainy={1,2,8,9,11},overcast={3,7,12,13}, sunny={4,5,6,10,14}}  
 {hot={1,2,3,13},mild={4,8,10,11,12,14},cool={5,6,7,9}}  
 {normal={5,6,7,9,10,11,13},high={1,2,3,4,8,12,14}}  
 {false={1,3,4,5,8,9,10,13},true={2,6,7,11,12,14}}

Fig. 8 Multi soft set based classification result.

Apply proposed approach for rule base classification on derived multi soft set. We will get following classification rules.

- RL1:** IF outlook = *rainy* AND humidity = high THEN play\_ball = *No*
- RL2:** IF outlook = *rainy* AND humidity = *normal* THEN play\_ball = *Yes*
- RL3:** IF outlook = *overcast* THEN play\_ball = *Yes*
- RL4:** IF outlook = *sunny* AND windy = *false* THEN play\_ball = *Yes*
- RL5:** IF outlook = *sunny* AND windy = *false* THEN play\_ball = *No*

The identical collection of classification rules are generated from both decision tree induction and proposed multi soft set based approaches. The processing time of our suggested approach-multi soft set is .00249 s. The comparison of the response time of traditional decision tree induction and proposed multi soft set-based classification rule production approaches is given in the following Fig 9.

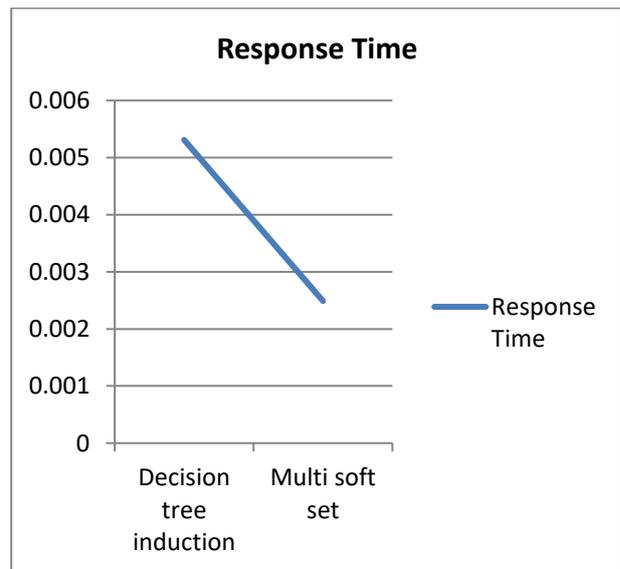


Fig 9. The comparison of executing time

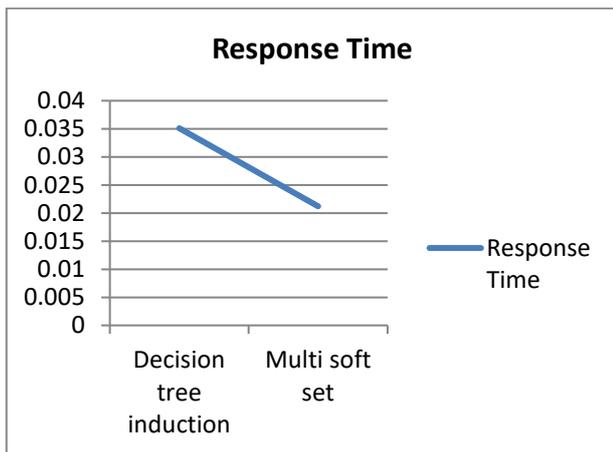
From the above fig 9 , the response time improvement of Multi Soft set with decision tree induction method is given below.

**Table 4: Response Time Improvement of Multi-Softset**

| Method                         | Response time improvement of Multi-Softset |
|--------------------------------|--|
| Decision tree induction method | 53.10%                                     |

**B. Car Evaluation Dataset**

Car Evaluation Dataset was resulting from a simple hierarchical decision model originally developed for the demonstration of DEX (M. Bohanec, V. Rajkovic: Expert system for decision making). The model evaluates cars using seven input attributes: buying, maint, doors, persons, lug\_boot, safety and class. Here we performed rule-based classification using the decision tree induction method and proposed multi soft set classification method. This contains searching for a set of classification rules that frequently happen in different transactions in order to understand unknown class of a car. It includes 1728 tuples and used in the processing of above two algorithms for mining classification rules. The training data set has a size of 392 tuples, which are extracted from the given data set. Response time of Multi soft set is better than the conventional decision tree induction method. The comparison chart is given below in figure 10.



**Fig 10. The comparison of executing time**

In the second example, the response time of Multi-Softset is considerably better in comparison with decision tree induction method. The process time of Multi-Softset is .0212s. Its response time improvement is given below.

**Table 5: Response Time Improvement Of Multi-Softset**

| Method                         | Response time improvement of Multi-Softset |
|--------------------------------|--|
| Decision tree induction method | 39.60%                                     |

In both experiments multi softest method is produced a good response time.

**VI. CONCLUSION AND FUTURE WORK**

Soft set theory is considered as a one of the major soft computing tool in knowledge discovery process. It has a good number of application data mining. Soft set theory is well known for its reliability and accuracy. Other soft computing

tools like Fuzzy set and Rough set have certain limitations while it handling incomplete data. Soft set theory is overcoming these difficulties. One of the data mining functionality is classification. Currently a handsome number of algorithms exist for classification. Among them rule-based classification using decision tree induction method is widely popular in data mining industry. We tried to bring out the capabilities of soft set theory in classification. In our proposed work, firstly, we studied the rule-based classification using decision tree induction. We experienced the merits and demerits of decision tree induction method.

Decision tree induction is a unique method for rule-based classification. It has to implement different steps for finishing its task. Construction of decision tree and its processing is more complex and time consuming. Development of model is consuming large time. A small change in the data set will affect entire structure of decision tree. These limitations we experienced in our research. To remove these difficulties, we have work out the new method rule-based classification using multi soft set theory; it could to decrease the demerits of rule-based classification using decision tree induction method. We implemented both methods in Python language, using the data sets weather and Car evaluation. At the end of the programming, we compared the result of both implementations. Rule-based classification using decision tree induction method is produce a good response time.

In future, we would like to examine opportunities of soft set theory in different data mining functionalities like outlier analysis and clustering.

**REFERENCES**

1. D. Molodtsov, "Soft set theory- First results", An International Journal of Computers & mathematics with applications", Elsevier, Volume 37, Issue 4-5, Page 19-31, February-March 1999.
2. Jiawei Han, Micheline Kamber, "Data Mining: Concept and Techniques", Elsevier, second edition, 2006.
3. <https://medium.com/@dhiraj8899/top-5-advantages-and-disadvantages-of-decision-tree-algorithm-428ebd199d9a>
4. Sushmita Mitra, Sankar K. Pal, Pabitra Mitra "Data Mining in Soft Computing Framework: A Survey", IEEE transactions on neural networks, vol. 13, no. 1, January 2002.
5. [Tutut Herawan, Mustafa Mat Deris "On Multi-soft Sets Construction in Information Systems", 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009, Proceedings (pp.101-110)
6. Weather data set, <https://github.com/sgaim/decision-tree-induction-id3/blob/master/baseball.csv>
7. Car evaluation Data Set, <https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

**AUTHORS PROFILE**



**Mr. Santhosh Kottam**, completed his Master of Computer Applications (MCA) from Madras University, Tamilnadu and B.Sc mathematics degree from M G University, Kerala. He is currently pursuing PhD in computer science at Bharatiyar University, Coimbatore. His research area is data mining and has more than 19 years of teaching experience, which includes UG and PG. He has been serving Federal Institute of Science and Technology (FISAT), Angamaly, Kerala as HOD & Assistant Professor (Senior Grade) in the Department of Computer Applications since May 2008.



## Rule-Based Classification using Multi-Soft Set Theory

During this period of time, he has taught many subjects including Programming Languages, System Analysis and Design, Operating Systems, Object Oriented Modeling and design and Data Mining. He has published research papers in the International Journals, National and International Conferences. Mr. Santhosh received the best teacher award of the FISAT College during the year 2017 for his overall contribution.



**Dr. Varghese Paul**, is completed B.Sc (Engg) in Electrical Engineering from Kerala University, M.Tech in Electronics and Ph.D in Computer Science from Cochin University of Science and Technology. Research Supervisor of Cochin University of Science and Technology, M G University Kottayam, Anna Technical University Chennai, Bharathiar University Coimbatore, Bharathidasan University Trichy and Karpagam

University Coimbatore. Under the guidance, 29 research scholars had already completed research studies and degree awarded. Research areas are Data Security using Cryptography, Data Compression, Data Mining, Image Processing and E\_Governance. Developed TDMRC Coding System for character representation in computer systems and encryption system using this unique coding system. Published many research papers in international as well as national journals and a text book also. Earlier worked as Industrial Engineer with O/E/N India Ltd Cochin, Communication Engineer with KSE Board, SCADA Engineer in Saudi Electricity Department, Head of IT Department CUSAT and Dean (CS, IT and Research) in Toc H Institute of Science and Technology. Certified Software Test Manager, Ministry of Information Technology, Govt of India. Life Member, Information System Audit and Control Association, USA (ISACA), Indian Society for Technical Education, India (ISTE) and National Geographic Society, USA.