

# BUILDING A CLASSIFIER EMPLOYING PRISM ALGORITHM WITH FUZZY LOGIC

Ishrat Nahar Farhana, A.H.M. Sajedul Hoque , Rashed Mustafa and Mohammad  
Sanaullah Chowdhury

Department of Computer Science and Engineering, University of Chittagong, Chittagong,  
Bangladesh

## **ABSTRACT**

*Classification in data mining is receiving immense interest in recent times. As the knowledge is based on historical data, classifications of data are essential for discovering the knowledge. To decrease the classification complexity, the quantitative attributes of data need splitting. But the splitting using the classical logic is less accurate. This can be overcome by the use of fuzzy logic. This paper illustrates how to build up the classification rules using the fuzzy logic. The fuzzy classifier is built on by using the prism decision tree algorithm. This classifier produces more realistic results than the classical one. The effectiveness of this method is justified over a sample dataset.*

## **KEYWORDS**

*Data Mining, Classification, Prism algorithm, Fuzzy logic , Membership Degree*

## **1. INTRODUCTION**

Data are big power for an organization to make better decisions. All knowledge of any organization is hidden in its historical data. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. We have a lot of raw data around us which has no meaning or use until they are given into a form that is useful and familiar to human. This helpful form is knowledge which can be derived from data and information. Thus data mining is necessary to discover that latent knowledge. Data mining has different tasks. The primary data mining tasks are : Classification, Clustering, Regression, Summarization, Dependency Modeling, Change and Deviation Detection [1]. Among those tasks, classification assigns items in a collection to target classes and accurately predicts the target classes for each case in the data [2] . Generally classification task generates a set of rules from training data set for future prediction and decision tree is a widely used technique for that purpose. There are many specific decision-tree algorithms, Prism algorithm is one of them. Prism algorithm is a rule based algorithm that induces modular rules using ‘separate and conquer’ approach [3]. Training data set may have both or either categorical or numerical attribute. For categorical attributes classification model can be built easily. But if the data set has any numerical attribute, it should be converted into categorical attribute by splitting the numeric range through classical logic or fuzzy logic. Fuzzy logic refers to better performance due to the overestimation and under estimation problem

of classical logic which has been shown in this paper. Moreover, classical classifier assigns one item in one and only one class, where fuzzy classifier supports multiple existence of one item among different classes with some membership values. This paper describes the way of building fuzzy classifier employing fuzzy logic on prism decision tree approach.

The rest of the paper is organized as follows. Section 2 provides a summary of related works. Classification using prism algorithm is discussed in section 3. Section 4 illustrates classical Prism algorithm. Generating rules using Fuzzy Prism classifier is presented in section 5. Section 6 illustrates the experimental result. Finally, section 7 concludes the paper.

## **2. RELATED WORK**

The main objective of this research is to propose a new way of building classifiers using fuzzy logic. We apply prism algorithm to classify training data set in fuzzy logic system. We focus on building classifiers or rules using fuzzy logic which gives more accurate result than classical logic system.

Some author proposed learning algorithm generates fuzzy rules from “soft” instances, which differ from conventional instances in that they have class membership values [4].

On the other hand, some author proposed fuzzy decision tree induction method, which is based on the reduction of classification ambiguity with fuzzy evidence [5]. Cognitive uncertainties involved in classification problems are explicitly represented, measured, and incorporated into the knowledge induction process according to them.

Methods of fuzzy k-means is used to overcome the problem of class overlap but their usefulness maybe reduced when data sets are large and when the data include artefacts introduced by the derivation of landform attributes from gridded digital elevation models. ‘High-resolution landform classification using fuzzy k-means’ presents ways to overcome these limitations using spatial sampling methods, statistical modelling of the derived stream topology, and fuzzy k-means using the Distance metric [6].

Fuzzy decision trees represent classification knowledge more naturally to the way of human thinking and are more robust in tolerating imprecise, conflict, and missing information.

## **3. CLASSIFICATION: A TASK OF DATA MINING**

Data mining using classification goes through supervised learning approach where labelled training data are used. The goal of classification is to accurately predict the target class for each case in the data. Classification consists of two steps: building the classifier and using classifier for classification [7]. Building the classifier phase trains the machine according to the given classes in training data set. Then the second phase which is the testing phase tests the input data from testing data set and put them into a class using classifier formed in first phase.

Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown. Classification models are tested by comparing the predicted values to known target values in a set of test data. Different techniques of data mining are Decision Tree, Naive Bayes method, Support Vector Machine, Neural network, Kernel

Estimation. Different algorithms are used to produce modular classification rules to construct regular decision tree. A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute, each branch represents the outcome of the test and each leaf node represents a class label where decision is taken after computing all attributes. The paths from root to leaf represent classification rules. Prism algorithms generate modular classification rules that cannot necessarily be represented in the form of a decision tree and provide higher classification accuracy [3].

Classification can be classified into two categories: Classical Classification and Fuzzy Classification. Training data set may have vagueness which makes confusion to classify. That is it seems an item may belong to different classes with some percentage. In that scenario, fuzzy classification is used where an item can be classified into more than one class with some membership degree.

#### 4. CLASSICAL PRISM ALGORITHM

There are two general approaches to the induction of classification rules, the 'divide and conquer' approach, also known as TDIDT and the 'separate and conquer' approach. 'Divide and conquer' induces classification rules in the intermediate representation of a decision tree. 'Separate and conquer' induces a set of IF..THEN rules. However the most notable development using the 'separate and conquer' approach is the Prism family of algorithms [8].

The prism algorithm was introduced by Cendrowska in 1987. The aim is to induce modular classification rules directly from the training set. The algorithm assumes that all the attributes are categorical. When there are continuous attributes they can first be converted to categorical one. Alternatively the algorithm can be extended to deal with continuous attributes. Prism uses the 'take the first rule fires' conflict resolution strategy when the resulting rules are applied to the unseen data, so it is important that as far as possible the most important rules are generated first.

The algorithm generates the rules concluding each of the possible classes in turn. Each rule is generated term by term with each term of the form 'attribute = value'. The attribute/value pair added at each step is chosen to maximize the probability of the target 'outcome class' [9]. The basic Prism algorithm is shown in Algorithm 1 [9].

##### Algorithm 1 (Classical Prism Algorithm)

**Input:** A training dataset with  $n$  classes  $C_i, i = 1, 2, 3, \dots, n$

**Output:** Generated rules for all classes

**Method:** The rules are generated in the following steps:

1. For each class  $C_i$  start with the complete training set each time
2. Compute the probability of each attribute/value pair for the class,  $C_i$
3. Select the pair with the largest probability and create a subset of the training set comprising all the instances with the selected attribute/value combination for each class,  $C_i$
4. Repeat steps 2 and 3 for this subset until a subset is reached that contain only instances of  $C_i$ .
5. The rule is induced by the conjunction of all the attribute/value pairs selected.
6. Remove all instances covered by this rule from the training set.
7. Repeat step 2 through 6 until all instances of  $C_i$  have been removed
8. Go to step 1 until all classes are examined.

## 5. BUILDING CLASSIFIER USING FUZZY PRISM ALGORITHM

Researchers training data set may consist of categorical and numeric attributes. Building classification model on numeric attribute is more complex than on categorical attribute. The domain of numeric attribute is huge which increases the branch nodes of decision tree and consequently increases the time complexity in testing phase. For example if the cardinality of any numeric attribute is  $n$ , the space complexity and the time complexity of that attribute is  $O(n)$ . This problem refers to construct bins on the domain of that attribute. If the number of bins of any numerical attribute is 3, then the complexity would be  $O(n/3)$ . Constructing bins impose another problem named vagueness. Some values may behave to be associated to more than one bin with different membership degree. This problem can be solved by using fuzzy logic during the construction of bins which ultimately implies fuzzy classifier. This section describes the process of constructing fuzzy classifier step by step applying on Table 1.

Table 1. The Training Dataset

<b>P_ID</b>	<b>Age</b>	<b>SpecRx</b>	<b>Astigmatism (angle in degree)</b>	<b>Tear Production Rate(%)</b>	<b>Class</b>
1	Young	Myope	85	41	no contact lenses
2	Young	Myope	80	78	soft contact lenses
3	Young	Myope	95	45	no contact lenses
4	Young	Myope	100	71	hard contact lenses
5	Young	Hypermetrope	130	60	no contact lenses
6	Young	Hypermetrope	120	61	hard contact lenses
7	Pre-presbyopic	Myope	60	56	no contact lenses
8	Pre-presbyopic	Myope	51	64	soft contact lenses
9	Pre-presbyopic	Hypermetrope	75	50	no contact lenses
10	Pre-presbyopic	Hypermetrope	90	73	soft contact lenses
11	Pre-presbyopic	Hypermetrope	115	44	no contact lenses
12	Pre-presbyopic	Hypermetrope	125	66	no contact lenses
13	Presbyopic	Myope	105	53	no contact lenses
14	Presbyopic	Myope	122	80	hard contact lenses
15	Presbyopic	Hypermetrope	88	57	no contact lenses
16	Presbyopic	Hypermetrope	110	47	no contact lenses
17	Presbyopic	Hypermetrope	100	75	no contact lenses

## 5.1 Fuzzy Prism Algorithm

Fuzzy logic is the multi-valued logic in which truth values of variables may have any real number between 0 and 1, where for classical logic it would be either 0 or 1 [10]. For example, Age = {(young,1), (young, .5), (middle aged, .5), (middle age,1), (middle aged, .1), (old, .4), (old, .7), (old,1)}, where every Cartesian product consists of linguistic variables and membership degree. The cardinality of the fuzzy set is computed by summing all membership values. So the cardinality of Age is (1+.5+.5+1+.1+.4+.7+1) or 5.2. Similarly |young|, |middle aged| and |old| are 1.5, 1.6 and 2.1 respectively. The probability in Fuzzy Prism algorithm is calculated using this cardinality shown in equation (1).

$$P_{L \in A}(L) = \frac{\text{Total Cardinality of } L}{\text{Total Cardinality of } A} \quad (1)$$

Here L is a fuzzy linguistic term and A is the set of all linguistic terms. For instance, the probability of young using eq. 1 will be

$$P_{\text{young} \in \text{Age}}(\text{young}) = \frac{1.5}{5.2} = 0.29$$

The modified prism algorithm using fuzzy logic is shown in Algorithm 2.

### Algorithm 2 (Fuzzy Prism Algorithm)

**Input:** A training dataset with n classes,  $C_i, i = 1, 2, 3, \dots, n$

**Output:** Fuzzy Classifier

**Method:** The rules are generated in the following steps:

1. Map the given training data set into the fuzzy training set by employing fuzzy logic on numeric attribute.
2. For each class  $C_i$  start with the complete training set each time.
3. Compute the probability of each attribute/value pair using equation 1 for the class,  $C_i$
4. Select the pair with the largest probability and create a subset of the training set comprising all the instances with the selected attribute/value combination for each class,  $C_i$
5. Repeat steps 3 and 4 for this subset until a subset contains only instances of  $C_i$  or covers all the attributes.
6. The rule is induced by the conjunction of all the attribute/value pairs selected.
7. Remove all instances covered by this rule from the training set.
8. Repeat step 3 through 7 until all instances of  $C_i$  have been removed
9. Go to step 2 until all classes are examined.

## 5.2 Constructing Bins

Due to the complexity of constructing decision tree on quantitative attributes: Astigmatism and Tear Production Rate, these attributes must be split using either classical logic or fuzzy logic. The constructed bins for Astigmatism is {Yes, No} and Tear Production Rate are is {Reduced, Normal}. The range for Astigmatism is from 51 to 130 and for Tear Production Rate is from 41 to 80. If the intervals for Yes and No bins are from 51 to 90 and from 91 to 130, according to the

classical logic each value under same interval has the same membership value that means either 0 or 1. For example, the membership degrees of 52 under Yes and No is 1 and 0 respectively. Again for 92 it is vice versa. Here though the difference between 52 and 91 is 39, 52 and 91 both belong to Yes. On the other hand, in spite of being 92 very closer to 91, it is associated to other interval No. These are known as overestimation and underestimation problem in classical logic. In order to remove these problems, fuzzy logic is applied. It needs to construct bins. This paper addresses S-shaped membership function for No and Reduced and Z-shaped membership function for Yes and Normal, which are shown in Figure 1.

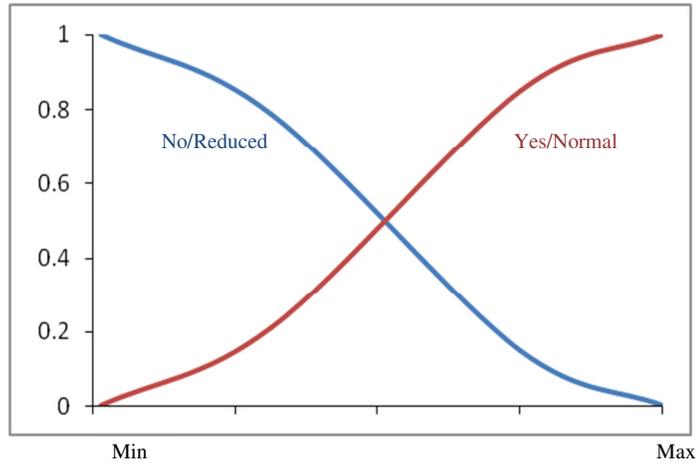


Figure 1. S-shaped and Z-shaped membership function for No/Reduced and Yes/Normal

The mathematical equations for S and Z-shaped membership functions are also shown in (2) and (3) respectively [11].

$$f(x)_s = \frac{1}{2} + \frac{1}{2} \cos \left( \frac{hb - x}{hb - lb} \right) \Pi \quad (2)$$

$$f(x)_z = \frac{1}{2} + \frac{1}{2} \cos \left( \frac{x - lb}{hb - lb} \right) \Pi \quad (3)$$

The classes are considered as:

- Hard contact lens- Class 1(C1),
- Soft contact lens- Class 2 (C2)
- No contact lens- Class 3(C3).

### 5.3 Constructing Mapping Table

Using equation (2) and equation (3) on astigmatism and tear production the mapping Table 2 has been constructed. Though astigmatism and tear production rate both have 2 intervals each, the mapping table get 4 pairs for each data or each patient. Thus the total dataset of Table 1 which holds 17 patient's data is converted to the mapping table which holds 68 records, 4 instances for each patient. Here the mapping table with 4 mapped records is shown partially in Table 2. The

membership degree of a class is computed by taking the minimum membership value from all fields of that record. In this way, the membership degrees for C3 in mapping table are 0.608, 0.392 and 0.

Table 2. The Partial Mapping Table

P_ID	R_ID	Age	SpecRx	Astigmatism		Tear Production Rate		Class (Membership value)
				No	Yes	Reduced	Normal	
1	1	Young	Myope	0.608		1		C3 (0.608)
	2	Young	Myope		0.392	1		C3 (0.392)
	3	Young	Myope	0.608			0	C3 (0)
	4	Young	Myope		0.392		0	C3 (0)

### 5.4 Generating Classification Rules

The steps from 2 to 9 of Algorithm 2 are now applied on Table 2 for each class. Here the demonstration of constructing Rule 1 for class 1 has been shown. At first the probabilities of all values of attributes are calculated using equation (1) is shown in Table 3.

Table 3. Probability of 68 instances

Attribute/Value Pair	Total Membership Value for C1	Total Membership Value of all 68 instances	Probability
Age =Young	8	24	0.33
Age = Pre-presbyopic	0	24	0
Age = presbyopic	4	20	0.2
SpecRx= Myope	8	32	0.25
SpecRx= Hypermetrope	4	36	0.11
Astig=No	0.76	13.416	0.06
Astig=Yes	5.24	20.584	0.255
Tear= Reduced	1.212	17.544	0.07
Tear=Normal	4.788	16.456	0.291

Since the highest probability is 0.33 for the attribute value pair (Age =Young), it will be the first term for Rule 1 is:

Rule 1 : (Age =Young) ^ ..... => C1

Subset S1 is created by picking those records whose has (Age =Young) pair over class C1 and the constructed subset S1 = { 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 }. All the probability for S1 is shown in Table 4.

Table 4 . Probability of 24 instances (S1)

Attribute/Value Pair	Total Membership Value for C1	Total Membership Value of all 68 instances	Probability
SpecRx= myope	4	16	0.125
SpecRx= Hypermetrope	4	8	0.5
Astig=No	0.71	4.154	0.17
Astig=Yes	3.29	7.846	0.42
Tear= Reduced	1.212	6.212	0.19
Tear = Normal	2.788	5.788	0.48

Here the highest probability is 0.5 for the attribute value pair ( SpecRx = Hypermetrope ) and this is the 2<sup>nd</sup> term for Rule 1 :

Rule 1 : ( Age =Young ) ^ ( SpecRx= Hypermetrope ) ^..... => C1

Again subset S2 is created by selecting those records which have the pair (Age =Young) and (SpecRx= Hypermetrope) under class C1 and S2 ={17,18,19,20,21,22,23,24}. Like Table 4, Table 5 has been formed on subset S2.

Table 5. Probability of 8 instances (S2)

Attribute/Value Pair	Total Membership Value for C1	Total Membership Value of all 68 instances	Probability
Astig=No	0.039	0.039	1
Astig=Yes	1.922	3.922	0.49
Tear = Reduced	0.96	2	0.48
Tear = Normal	1.04	2	0.52

Similarly since the attribute value pair (Astig=No) has the highest probability 1 and is the 3<sup>rd</sup> term of Rule 1 is :

Rule 1:(Age =Young )^ (SpecRx= Hypermetrope) ^ (Astig=No).....=>C1

The constructed new subset S3 with (Age =Young), ( SpecRx = Hypermetrope ) and (Astig=No) will be {17,19,21,23}. The next constructed probability over subset S3 is shown in Table 6.

Table 6. Probability of 4 instances (S3)

Attribute/Value Pair	Total Membership Value for C1	Total Membership Value of all 68 instances	Probability
Tear = Reduced	0.48	1	0.48
Tear = Normal	0.52	1	0.52

From the Table 6, it is shown that the next attribute/value pair will be (Tear=Normal) which probability is 0.52 is the 4<sup>th</sup> term of Rule 1. The final generated Rule 1 is :

Rule 1: (Age =Young)^(SpecRx=Hypermetrope)^(Astig=No) ^ (Tear=Normal) => C1

Then all records with pairs (Age =Young), (SpecRx= Hypermetrope), (Astig=No) and (Tear=Normal) are deleted to generate rule 2.

In this manner, all the rules for class 1 (C1) is generated until all the instances labelled by class 1 (C1) are deleted from the mapping table. Similarly rules for class 2 (C2) and class 3 (C3) are generated.

## 6. EXPERIMENTAL RESULT

According to the procedure described so far, the fuzzy classifier that has been found is shown in Figure 2.

(Age =Young)^(SpecRx = Hypermetrope)^(Tear=Reduced)^(Astig=No)	=> C <sub>1</sub> (0.039) , C <sub>3</sub> (0)
(Age =Young)^(SpecRx = Hypermetrope)^(Tear=Reduced)^(Astig=Yes)	=> C <sub>1</sub> (0.480) , C <sub>3</sub> (0.520)
(Age =Young)^(SpecRx = Hypermetrope)^(Tear=Normal)^(Astig=No)	=> C <sub>1</sub> (0.039) , C <sub>3</sub> (0)
(Age =Young)^(SpecRx = Hypermetrope)^(Tear=Normal)^(Astig=Yes)	=> C <sub>1</sub> (0.520) , C <sub>3</sub> (0.480)
(Age =Young)^(SpecRx = Myope)^(Tear=Reduced)^(Astig=No)	=> C <sub>1</sub> (0.126) , C <sub>2</sub> (0.006) , C <sub>3</sub> (0.510)
(Age =Young)^(SpecRx = Myope)^(Tear=Reduced)^(Astig=Yes)	=> C <sub>1</sub> (0.126), C <sub>2</sub> (0.006), C <sub>3</sub> (0.491)
(Age =Young)^(SpecRx =Myope)^(Tear=Normal)^(Astig=Yes)	=> C <sub>1</sub> (0.316), C <sub>2</sub> (0.703), C <sub>3</sub> (0.013)
(Age =Young)^(SpecRx = Myope)^(Tear=Normal)^(Astig=No)	=> C <sub>1</sub> (0.684), C <sub>2</sub> (0.297), C <sub>3</sub> (0.013)
(Age = Presbyopic)^(SpecRx= Myope)^(Tear=Reduced)^(Astig=No)	=> C <sub>1</sub> (0), C <sub>3</sub> (0.227)
(Age = Presbyopic)^(SpecRx=Myope)^(Tear=Reduced)^(Astig=Yes)	=> C <sub>1</sub> (0), C <sub>3</sub> (0.773)
(Age = Presbyopic)^(SpecRx=Myope)^(Tear=Normal)^(Astig=Yes)	=> C <sub>1</sub> (0.025) , C <sub>3</sub> (0.216)
(Age = Presbyopic)^(SpecRx= Myope)^(Tear=Normal)^(Astig=No)	=> C <sub>1</sub> (0.025) , C <sub>3</sub> (0.216)
(Age = Pre-presbyopic)^(SpecRx= Hypermetrope)^(Tear=Reduced)^(Astig=No)	=> C <sub>2</sub> (0.077), C <sub>3</sub> (0.295)
(Age = Pre-presbyopic)^(SpecRx= Hypermetrope)^(Tear=Reduced)^(Astig=Yes)	=> C <sub>2</sub> (0.077) , C <sub>3</sub> (0.470)
(Age = Pre-presbyopic)^(SpecRx= Hypermetrope)^(Tear=Normal)^(Astig=No)	=> C <sub>2</sub> (0.510) , C <sub>3</sub> (0.050)
(Age = Pre-presbyopic)^(SpecRx= Hypermetrope)^(Tear=Normal)^(Astig=Yes)	=> C <sub>2</sub> (0.490) , C <sub>3</sub> (0.285)
(Age = Pre-presbyopic)^(SpecRx= Myope)^(Tear=Reduced)^(Astig=No)	=> C <sub>2</sub> (0.361) , C <sub>3</sub> (0.677)
(Age = Pre-presbyopic)^(SpecRx=Myope)^(Tear=Reduced)^(Astig=Yes)	=> C <sub>2</sub> (0) , C <sub>3</sub> (0.032)
(Age = Pre-presbyopic)^(SpecRx=Myope)^(Tear=Normal)^(Astig=Yes)	=> C <sub>2</sub> (0.639) , C <sub>3</sub> (0.323)
(Age = Pre-presbyopic)^(SpecRx= Myope)^(Tear=Normal)^(Astig=No)	=> C <sub>2</sub> (0) , C <sub>3</sub> (0.032)
(Age = Presbyopic)^(SpecRx= Hypermetrope)^(Tear=Reduced)^(Astig=No)	=> C <sub>3</sub> (0.247)
(Age = Presbyopic)^(SpecRx= Hypermetrope)^(Tear=Reduced)^(Astig=Yes)	=> C <sub>3</sub> (0.447)
(Age = Presbyopic)^(SpecRx= Hypermetrope)^(Tear=Normal)^(Astig=Yes)	=> C <sub>3</sub> (0.245)
(Age = Presbyopic)^(SpecRx= Hypermetrope)^(Tear=Normal)^(Astig=No)	=> C <sub>3</sub> (0.367)

Figure 2. Fuzzy Classifier

Here total 24 rules are generated for the training dataset of table 1 using fuzzy prism algorithm. Some rule contains more than 1 class with a membership degree, which is not possible in classical classifier. This is the beauty of fuzzy classifier. Here, if same rule has the different membership degree in same class, the membership degree is aggregated by taking average values.

## 7. CONCLUSION

In classical prism classifier, every rule belongs to one class only that means a rule either belongs C1 or C2 or C3. So the classical result is not fully accurate. Because in real world, a single rule cannot contain 100% characteristics of the features of a class. So it is not realistic. Because of overestimation and underestimation problem testing data or patient (according to this paper) contains some features of other classes but gives only one class as output, though it belongs more than one class.

In this paper, Fuzzy classifier has been built using Prism algorithm successfully. It shows that a single rule can contain more than one class with individual membership degree. So it gives more accurate result than classical classification. It gives more realistic result than classical classification by overcoming overestimation and underestimation problems of classical classification. This fuzzy classifier has big application in industry. For example, bank authority wants to give some incentive to the customers by classifying them into very committed, committed and non committed customer. Some customer's behavior is very confusing. Authority cannot classify them in one class. It seems to the authority that those customers belong to multiple classes with different percentage. In this circumstance, fuzzy classification is perfect. The total calculation discussed in this paper has some difficulties. Fuzzy Prism classifier needs huge calculation which increases time and space complexity. One can extend this work on Big Data by inventing parallel algorithm on distributed computing platform such as hadoop or spark.

## REFERENCES

- [1] Fayyad, Piatetsky-Shapiro, Smyth , (1996) , “From Data Mining to Knowledge Discovery in Databases”. AI Magazine.
- [2] Bea Jiawei Han, Micheline Kamber and Jian Pei, (2012) , Data Mining Concepts and Techniques , 3rd edition , Morgan Kaufmann, Waltham, USA .
- [3] Frederic Stahl and Max Bramer, (2013) “Random Prism: A noise tolerant Alternative to Random Forests ” , Expert Systems, Vol. 31, Issue 5, pp .411-420.
- [4] Ching-Hung Wang , Jau-Fu Liu , Tzung-Pei Hong and Shian-Shyong , (1999) “A fuzzy inductive learning strategy for modular rules”, Elsevier Science B.V. , Fuzzy Sets and Systems , Vol. 103, Issue 1, pp 91-105.
- [5] Yufei Yuan and Michael J. Shaw , (1995) , “Induction of fuzzy decision trees”, Elsevier Science B.V. , Fuzzy Sets and Systems , Vol. 69, Issue 2, pp 125-139.
- [6] P.A. Burrough, F.M. van Gaans, R.A. MacMillan , (2000), “High-resolution landform classification using fuzzy k-means”, Elsevier Science B.V. , Fuzzy Sets and Systems , Vol. 113, Issue 1, pp 37-52.
- [7] Data Mining-Classification & Prediction, Tutorials Point, [online], [http://www.tutorialspoint.com/data\\_mining/dm\\_classification\\_prediction.html](http://www.tutorialspoint.com/data_mining/dm_classification_prediction.html)(Accessed 27 February, 2017) .
- [8] Frederic Stahl and Max Bramer , (2011), “Random Prism: An Alternative to Random Forests ” , Research and Development in Intelligent Systems XXVIII , Springer.

- [9] Max Bramer ,(2001), Principles of Data Mining , Springer, Berlin, Germany.
- [10] Claudio Moraga , (2005), “Introduction to Fuzzy Logic”, Facta Universitatis, Series: Electronics and Energetics, vol. 18, pp. 319-328 .
- [11] Tassnim Manami Zaman, A. H. M. Sajedul Hoque, Md. Al-Amin Bhuiyan,(2011) “Knowledge Discovery and Intelligent Query Employing Fuzzy Logic”, in Proc. ICMCS , Chennai, India..