ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Sena ÖZTÜRK

AN EXPERIMENTAL EVALUATION FOR FUZZY DECISION TREES

DEPARTMENT OF COMPUTER ENGINEERING

ADANA, 2015

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

AN EXPERIMENTAL EVALUATION FOR FUZZY DECISION TREES

Sena ÖZTÜRK

MSc THESIS

DEPARTMENT OF COMPUTER ENGINEERING

We certify that the thesis titled above was reviewed and approved for the award of degree of the Master of Science by the board of jury on 26/01/2015.

……	……..	……..
Assoc. Prof. Dr. S. Ayşe ÖZEL	Assoc. Prof. Dr. Mustafa ORAL	Assoc. Prof. Dr. Mutlu AVCI
SUPERVISOR	MEMBER	MEMBER

This MSc Thesis is written at the Department of Institute of Natural And Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Mustafa GÖK
Director
Institute of Natural and Applied Sciences

Not: The usage of the presented specific declerations, tables, figures, and photographs either in this thesis or in any other reference without citiation is subject to "The law of Arts and Intellectual Products" number of 5846 of Turkish Republic

ABSTRACT

MSc THESIS

AN EXPERIMENTAL EVALUATION FOR FUZZY DECISION TREES

Sena ÖZTÜRK

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF COMPUTER ENGINEERING

Supervisor : Assoc. Prof. Dr. Selma Ayşe ÖZEL
Year: 2015, Pages: 109
Jury : Assoc. Prof. Dr. Selma Ayşe ÖZEL
: Assoc. Prof. Dr. Mustafa ORAL
: Assoc. Prof. Dr. Mutlu AVCI

In this study, fuzzy decision tree which uses fuzzified data by using triangular or trapezoidal membership functions was developed and compared with classical decision tree. Basic and fuzzy version of splitting criteria were used. Moreover, effect of the linguistic terms that were used for fuzzified data was studied in this thesis. To achieve this aim, two different fuzzy datasets were obtained; winner linguistic term and all linguistic terms and compared with each other. The aim of this study is to improve classification performance of fuzzy decision trees and test the effect of the fuzzification of the dataset on the decision tree induction. The proposed fuzzy and classical decision tree methods based on ID3 decision tree algorithm was experimented on 18 datasets from UCI Machine Learning Repository. The experimental results of this study showed that, fuzzy decision tree is more successful than classical decision tree.

Key Words: Fuzzy decision tree, Fuzzy set theory, Fuzzy splitting criteria, Membership functions, Linguistic Terms

ÖZ

YÜKSEK LİSANS TEZİ

BULANIK KARAR AĞAÇLARI İÇİN BİR DENEYSEL DEĞERLENDİRME

Sena ÖZTÜRK

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Danışman : Doç. Dr. Selma Ayşe ÖZEL
Yıl: 2015, Sayfa: 109
Juri : Doç. Dr. Selma Ayşe ÖZEL
: Doç. Dr. Mustafa ORAL
: Doç. Dr. Mutlu AVCI

Bu çalışmada, üçgen ve yamuk üyelik fonksiyonları kullanılarak bulanıklaştırılan bulanık veri ile bulanık karar ağacı geliştirilip klasik karar ağacı ile karşılaştırılmıştır. Ayırma kriterlerinin temel ve bulanık versiyonları kullanılmıştır. Bu tezde ayrıca, bulanık veriler için kullanılan sözel terimlerin etkisi çalışılmıştır. Bunun için, kazanan sözel terim ve tüm sözel terimler olmak üzere iki farklı bulanık veri kümesi elde edilmiş ve birbirleriyle karşılaştırılmıştır. Bu çalışmanın amacı, bulanık karar ağaçlarının sınıflama performansını geliştirmek ve karar ağacı oluşumunda veri kümesi bulanıklaştırmanın etkisini test etmektir. ID3 karar ağacı algoritması temel alınarak, önerilen bulanık ve temel karar ağacı yöntemleri UCI makine öğrenme veri havuzundan alınan 18 veri kümesi ile test edilmiştir. Deneysel sonuçlar, bulanık veri kullanılarak elde edilen ID3 karar ağacının klasik karar ağacından daha başarılı olduğunu göstermiştir.

Anahtar Kelimeler: Bulanık karar ağaçları, Bulanık küme teorisi, Bulanık ayırma kriterleri, Üyelik fonksiyonları, Sözel terimler

ACKNOWLEDGEMENTS

I would like to thank my advisor Assoc. Prof. Dr. S. Ayşe Özel for her continuous and unconditional support, invaluable guidance and encouragement. I especially appreciate for her patience and motivation.

I would like to thank each and every member of the evaluation committee for their guidance.

Special thanks to my dear friends İhsan Demiray and Melis Özyıldırım for their patience and endless support.

I also want to thank my family for their support and encouragements for my life and career.

CONTENTS **PAGE**

ÖZ I
ABSTRACT II
ACKNOWLEDGEMENTS III
CONTENTS IV
LIST OF TABLES… VII
LIST OF FIGURES XI
l. INTRODUCTION 1
2. PREVIOUS WORK 5
2.1. Decision Trees 5
2.2. Fuzzy Decision Trees 6
3. MATERIAL AND METHOD 13
3.1. Material 13
3.1.1. Classical Decision Trees 13
3.1.2. Fuzzy Set Theory 17
3.1.3. Membership Functions 18
3.1.3.1. Triangular Membership Function 19
3.1.3.2. Trapezoidal Membership Function 21
3.1.4. Fuzzy Decision Trees 22
3.1.5. Splitting Criteria 24
3.1.5.1. Information Gain 24
3.1.5.2. Gain Ratio 27
3.1.5.3. Gini Index 28
3.1.5.4. Best Split Method for Numerical Data 30
3.1.6. Fuzzy Form of Splitting Criteria 32
3.1.6.1. Fuzzy Information Gain 32
3.1.6.2. Fuzzy Gain Ratio 33
3.1.6.3. Fuzzy Gini Index 34
3.1.7. Performance Measures 35
3.1.7.1. Rule Performance 35

3.1.7.2. Measuring the Performance of the Classification Model 36
3.1.8. Weka Data Mining Tool 38
3.1.8.1. Weka Data Formats 39
3.1.9. Datasets 40
3.2. Method 42
3.2.1. Data Fuzzification 43
3.2.2. Inducing Decision Tree with Non-Fuzzy Data Method 49
3.2.2.1. ID3 with Best Split Method 49
3.2.3. Inducing Decision Trees With Fuzzy Data Methods 51
3.2.3.1. ID3 with Fuzzy Data and Basic Splitting Criterion 52
3.2.3.2. ID3 with Fuzzy Data and Fuzzy Form of Splitting Criterion 54
3.2.4. Extracting Classification Rules from Fuzzy Decision Tree 57
3.2.5. Applying Fuzzy Rules for Classification 58
4. RESEARCH AND DISCUSSION 61
4.1. Effect of Data Fuzzification 61
4.2. Effect of Using Fuzzy Decision Tree on Classification Performance 66
4.3. Effect of LinguisticTerms 70
4.4. Comparison of Information Gain, Gain Ratio and Gini Index 86
4.5. Comparison of Fuzzy Information Gain, Fuzzy Gain Ratio and Fuzzy Gini Index 88
4.6. Comparison of Triangular and Trapezoidal Fuzzy Membership Functions 93
4.7. Comparison of Rule Selection Methods 96
4.8. Experimental Results of the Weka Classification Tool 98
5. CONCLUSION 103
REFERENCES 105
BIOGRAPHY 109

LIST OF TABLES **PAGE**

Table 3.1. Weather Dataset 15
Table 3.2. Values of the "Temperature" Attribute of Numerical Weather Dataset and Their Class Labels 31
Table 3.3. A 2x2 Confusion Matrix 36
Table 3.4. A Confusion Matrix for More Than Two Class 37
Table 3.5. UCI datasets used in the experiments 41
Table 3.6. Partitioned Datasets 42
Table 3.7. A Small Car Type Dataset 47
Table 3.8. Fuzzified Small Car Type Dataset Using Linguistic Terms Having Maximum Membership Value (a) with Triangular Membership Function (b) with Trapezoidal Membership Function 48
Table 3.9. Fuzzified Small Car Type Dataset Using All Linguistic Terms According to Triangular Membership Function 49
Table 3.10. Fuzzified Small Car Type Dataset Using All Linguistic Terms According to Trapezoidal Membership Function 49
Table 3.11. Test Dataset for the Numerical Weather Dataset 60
Table 3.12. Fuzzified Test Dataset Presented in Table 3.11 Using Triangular Membership Function 60
Table 4.1. Accuracy of the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions 62
Table 4.2. F-measure of the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions 63
Table 4.3. Number of Rules of the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions 64

Table 4.4. Training Time in Seconds for the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions 65
Table 4.5. Test Time in Seconds for the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions 66
Table 4.6. Accuracy of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions 68
Table 4.7. F-measure of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions 68
Table 4.8. Experimental Results in Number of Rules of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions 69
Table 4.9. Experimental Results in Training Time in Seconds of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions 69
Table 4.10. Experimental Results in Test Time in Seconds of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions 70
Table 4.11. Classification Accuracy of "Heart Statlog" Dataset for All and Single Linguistic Terms 72
Table 4.12. Classification Accuracy of "Mammographic Masses" Dataset for All and Single Linguistic Terms 72
Table 4.13. Classification Accuracy of "Breast Cancer" Dataset for All and Single Linguistic Terms 73
Table 4.14. Classification Accuracy of "Diabetes" Dataset for All and Single Linguistic Terms 73

Table 4.15. Classification Accuracy of "Hepatitis" Dataset for All and Single Linguistic Terms 74
Table 4.16. Classification Accuracy of "Spect Heart" Dataset for All and Single Linguistic Terms 74
Table 4.17. Classification Accuracy of "Yeast" Dataset for All and Single Linguistic Terms 75
Table 4.18. Classification Accuracy of "Vertebral Column 2C" Dataset for All and Single Linguistic Terms 75
Table 4.19. Classification Accuracy of "Vertebral Column 3C" Dataset for All and Single Linguistic Terms 76
Table 4.20. Classification Accuracy of "Ecoli" Dataset for All and Single Linguistic Terms 76
Table 4.21. Classification Accuracy of "Balance Scale" Dataset for All and Single Linguistic Terms 77
Table 4.22. Classification Accuracy of "Thyroid" Dataset for All and Single Linguistic Terms 77
Table 4.23. Classification Accuracy of "LD Bupa" Dataset for All and Single Linguistic Terms 78
Table 4.24. Classification Accuracy of "Iris" Dataset for All and Single Linguistic Terms 78
Table 4.25. Classification Accuracy of "Glass" Dataset for All and Single Linguistic Terms 79
Table 4.26. Classification Accuracy of "Monk1" Dataset for All and Single Linguistic Terms 79
Table 4.27. Classification Accuracy of "Monk2" Dataset for All and Single Linguistic Terms 80
Table 4.28. Classification Accuracy of "Monk3" Dataset for All and Single Linguistic Terms 80
Table 4.29. F-Measure of "ID3 with Fuzzified Data, Basic, and Fuzzified Splitting Criteria" Method with All Linguistic Terms 81

Table 4.30. Number of Rules for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method with All Linguistic Terms..... 83
Table 4.31. Training Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method with All Linguistic Terms..... 83
Table 4.32. Test Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method with All Linguistic Terms..... 84
Table 4.33. Number of Rules for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with All Linguistic Terms..... 84
Table 4.34. Training Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with All Linguistic Terms..... 85
Table 4.35. Test Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with All Linguistic Terms..... 85
Table 4.36. Accuracy of J48 Algorithm in Weka Classification Tool..... 99
Table 4.37. Accuracy of ID3 Algorithm in Weka Classification Tool 100

LIST OF FIGURES **PAGE**

Figure 1.1. A Classical Decision Tree 3
Figure 1.2. Decision Tree with Fuzzified Data 3
Figure 3.1. An Example of Decision Tree of The Weather Dataset 16
Figure 3.2. An Example of Complex Decision Tree of the Weather Dataset 16
Figure 3.3. Sharp-edged Membership Function for "Tall" Attribute 17
Figure 3.4. Fuzzy Membership Function for "Tall" Attribute 18
Figure 3.5. An Example of Linguistic Terms of the Pressure Attribute 19
Figure 3.6. Triangular Membership Function 20
Figure 3.7. An Example of the Triangular Membership Function in the Age Attribute 21
Figure 3.8. Trapezoidal Membership Function 21
Figure 3.9. Trapezoidal Membership Function of Income 22
Figure 3.10. Temperature Attribute with Cut-point Values 32
Figure 3.11. The GUI of the Weka 39
Figure 3.12. Example of arff File Format in Weka 40
Figure 3.13. Triangular Membership Function 43
Figure 3.14. Trapezoidal Membership Function 45
Figure 3.15. Membership Function for Examinee's Score 46
Figure 3.16. Sharp Training Part of the Basic Decision Tree Induction Using "ID3 with Best Split Method" 50
Figure 3.17. Decision Tree for Small Car Type Dataset Using Best Split Method with Information Gain Measure 51
Figure 3.18. Training Part of the Decision Tree Induction Using "ID3 with Fuzzy Data and Basic Splitting Criterion" 52
Figure 3.19. FDT using Fuzzy Information Gain for Dataset in Table 3.8. (a) 53
Figure 3.20. FDT using Fuzzy Information Gain for Dataset in Table 3.9 i.e., with All Linguistic Terms of Samples in the Dataset 54
Figure 3.21. Training Part of the Decision Tree Induction Using "ID3 with Fuzzy Data and Fuzzy Form of Splitting Criterion" 55

Figure 3.22. FDT using Fuzzy Information Gain for Dataset in Table 3.8. (a)..... 56
Figure 3.23. FDT using Fuzzy Information Gain for Dataset in Table3.9 i.e., 63 with All Linguistic Terms of Samples in the Dataset 57
Figure 3.24. Test part of the Induction Algorithm 59
Figure 3.25. Test Results of the Test Weather Dataset 60
Figure 4.1. Accuracy of the "ID3 with Fuzzified Data and Basic Splitting Criteria" Method 87
Figure 4.2. Number of Rules Learned by "ID3 with Fuzzified Data and Basic Splitting Criteria" Method 87
Figure 4.3. Training Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method 88
Figure 4.4. Test Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method 88
Figure 4.5. Accuracy of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method 89
Figure 4.6. Number of Rules Learned by "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method 90
Figure 4.7. Training Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method 90
Figure 4.8. Test Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method 91
Figure 4.9. Accuracy of Basic and Fuzzy Version of Info Gain 91
Figure 4.10. Accuracy of Basic and Fuzzy Version of Gain Ratio 92
Figure 4.11. Accuracy of Basic and Fuzzy Version of Gini Index 93
Figure 4.12. Classification Accuracy for Triangular and Trapezoidal Membership Functions..... 94
Figure 4.13. Number of Rules Obtained From Decision Tree Using Basic Information Gain for Triangular and Trapezoidal Membership Functions..... 95

Figure 4.14. Training Time in Seconds of "ID3 with Fuzzified Data and Basic Splitting Criteria" Method for Triangular and Trapezoidal Membership Functions 95
Figure 4.15. Test Time in Seconds of "ID3 with Fuzzified Data and Basic Splitting Criteria" Method for Triangular and Trapezoidal Membership Functions 96
Figure 4.16. Accuracy of Test Types for All Linguistic Terms and Info Gain 97
Figure 4.17. Accuracy of Test Types for All Linguistic Terms and Gain Ratio 97
Figure 4.18. Accuracy of Test Types for All Linguistic Terms and Gini Index 98

1. INTRODUCTION

Developments of the technology have led to increase in the amount of data. As a result of this increase in the volume of data, need for computers and databases have risen. Since data become more complex the problem of data classification occurs. High-dimensional data having thousands of attributes complicate classification further.

Pattern classification can be defined as assigning class labels to data such that, similar data has the same class label. Neural networks and machine learning algorithms provide efficient solutions for pattern classification. For instance, Multilayer perceptrons (MLP), radial basis functions (RBF), probabilistic neural networks (PNN), self-organizing maps (SOM), decision trees, support vector machines (SVM), k-nearest neighbor neural network (kNN), Bayesian neural networks are some of these methods used for classification. On the other hand, there are unsupervised algorithms such as k-means, k-medoids, Gaussian mixtures for clustering. Unlike supervised learning, unsupervised learning is used for clustering which is a categorization task of unlabeled data into groups.

In machine learning, classification is a task for data analysis that can help understanding large data. It builds a model and identifies a set of categories that can be named as classes from large data. Then, new data whose class descriptions are not known can be categorized with this learned model. For example we can build a classification model to categorize people who are at high, medium or low risk for a bank loan. Classification is an instance of supervised machine learning technique. In the supervised learning methods, the goal is to build a model from the training set to predict new examples (Rokach, Maimon, 2008; Han, Kamber, 2006).

Since decision trees are one of the most popular approaches in machine learning, they are used for data classification (Janikow, 1998; Lee, Lee, and Lee-kwang 1999; Janikow, 1996). A decision tree is built by partitioning the dataset recursively until whole data is completely partitioned and data are classified by using rules generated with the decision tree. Generation of understandable classification rules and low computational costs of classification are the strengths of decision trees.

On the other hand, for classical decision trees, continuous data should be made discrete before using it. So, classical decision trees are sensitive to small changes in attribute values. Most popular decision trees are ID3 and C4.5 (Janikow, 1998; Lee, Lee, and Lee-kwang 1999; Quinlan 1986). Decision trees have been used successfully in many different areas, such as medical diagnosis, plant classification and customer marketing strategies. A decision tree can be used in classification by starting from the root node of the tree and moving through it until a leaf node denoting the class of the instance. Since, data used in real world can be fuzzy or uncertain form, decision trees should be able to give accurate results with such data. For this purpose, fuzzy decision trees are created (Umano, Okamoto, Hatono, Tamura, Kawachi, Umedzu, and Kinoshita, 1994; Wang, Chen, Qian, and Ye, 2000). Fuzzy decision trees are extension of classical decision trees and they are more robust with incorrect, noisy and uncertain data (Janikow, 1998; Peng, Flach, 2001). According to (Chiang, and Hsu, 1996; Maher, Clair, 1993; Marsala, 2009; Peng, Flach, 2001), fuzzy decision trees have better performance than classical decision trees.

A classical decision tree which is built by dicretization of continuous numerical data is shown in Figure 1.1. Decision tree which is built using fuzzified data is presented in Figure 1.2. According to Figure 1.1, if x_1 is 81 and x_2 is 21, test data is classified as class c_3. If x_2 is changed with small value, for example x_1 is 81 and x_2 is 19, class becomes c_1. This is the disadvantage of the classical decision trees. In contrast, fuzzy decision trees classify samples with possibilities. Test data is classified according to probabilities of belonging to the classes c_1, c_2, c_3. As a result, fuzzy decision trees are more robust than classical decision trees.

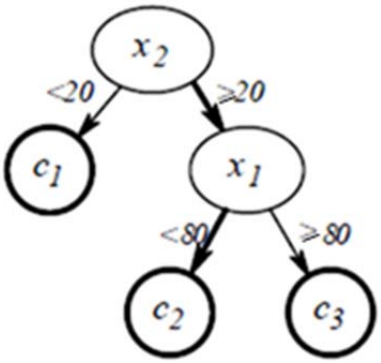

Figure 1.1. A Classical Decision Tree (Peng, Flach, 2001)

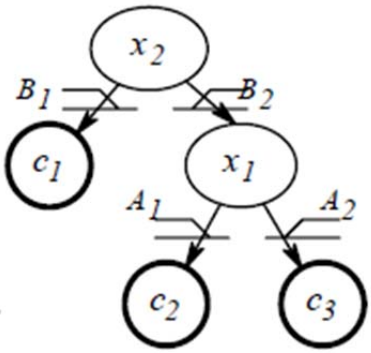

Figure 1.2. A Decision Tree with Fuzzified Data (Peng, Flach, 2001)

In this thesis, a detailed evaluation of fuzzy decision trees was performed. To do that classical decision tree and two forms of fuzzy decision tree were implemented to compare with each other. For fuzzy decision trees, numerical attributes were fuzzified by using triangular and trapezoidal membership functions. Information gain, gain ratio and gini index measures were used to select best attribute, to split the training dataset and to induct the decision tree. Additionally fuzzy forms of splitting criteria which were computed by using membership values of the attributes were used. Also, effect of the linguistic terms that were used for fuzzified data was studied in this thesis. To do that, two different fuzzy datasets were obtained. One of the fuzzy datasets contained linguistic terms having the maximum membership value and the other fuzzy dataset contained all linguistic terms that have greater than zero membership for an element. All methods are tested on various datasets from UCI Repository such as Heart Statlog, Iris, etc.

The contributions of this thesis can be summarized as follows:

i. Classical decision trees were compared with fuzzy decision trees which use only fuzzified data, and both fuzzified data and fuzzified splitting criteria such as fuzzy information gain, fuzzy gain ratio, and fuzzy gini index.
ii. Performance of triangular and trapezoidal fuzzy membership functions were compared for fuzzy decision tree construction.
iii. Performance of different data fuzzification methods which uses winner linguistic term, or all linguistic terms were compared.
iv. Performance of different rule selection methods that works if more than one rule is correct for one test data were compared.
v. ID3 and J48 algorithm in Weka Machine Learning Tool were compared with our fuzzy decision tree algorithm.

In the next section we give previous study on fuzzy decision trees, in Section 3 the algorithms and the datasets used in the experiments are presented. The results of the experiments and the discussions are given in detail in Section 4. Finally Section 5 concludes the study.

2. PREVIOUS WORKS

This section includes introductory information about decision trees and previous studies about fuzzy decision trees. More detail about fuzzy and non-fuzzy decision tree induction are given in Section 3.

2.1. Decision Trees

There are huge volumes of data today, so knowledge extraction is very important. In machine learning, classification techniques such as decision trees are one of the most important techniques for understanding large data.

Decision trees are one of the common approaches for learning from examples and they are widely used in machine learning for data classification (Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Janikow, 1996). Decision trees are used for classification because of their flexibility and clarity. They are based on partitioning the dataset recursively and data are classified by generating rules from the learned decision tree. Decision trees represent set of rules and rules obtained from tree to classify new samples. Generation of understandable classification rules are advantage of the decision trees. On the other hand, decision trees are sensitive to small changes in the attribute values and this may cause wrong classification (Yuan, Shaw, 1995). Decision trees were popularized by Quinlan (Quinlan, 1986). One of the most popular decision tree algorithm is ID3 (Quinlan, 1986), which stands for Interactive Dichotomizer 3. This technique has been used successfully in many different areas, such as diseases, plant classification. ID3 builds the decision tree from symbolic data for classifying the data and is generally not suitable for numerical values. If continuous attributes are involved, they must be discretized by dividing into several numerical intervals before using in the algorithm (Peng, Flach, 2001). More details about the ID3 algorithm and classical decision trees are presented in Section 3.1.

2.2. Fuzzy Decision Trees

Data can be fuzzy or uncertain form in real world. Decision trees should be able to give accurate results with such data. To make decision trees more flexible and deal with uncertain data, fuzzy decision trees have been proposed by many researchers (Abu-halaweh, Harrison, 2009; Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Umano, Okamoto, Hatono, Tamura, Kawachi, Umedzu, and Kinoshita, 1994 ; Yuan, Shaw, 1995; Wang, Chen, Qian, and Ye, 2000; Wang, Tien-chin, Lee, and Hsien-da, 2006; Janikow, 1996; Mitra, Konwar, Pal, 2002; Lee, Sun, Yang, 2003). The fuzzy decision tree induction process is similar to classical decision tree induction process. It builts the decision tree by dividing the dataset recursively with the best attribute selected by the information gain value. Fuzzy decision tree induction process generally consists fuzzification of the training data, induction of the fuzzy decision tree, extraction of fuzzy rules from the tree, and the classification processes. Fuzzy decision trees are more robust with incorrect, noisy, uncertain data, and for fuzzy decision trees ratio of wrong classification is lower than classical decision trees (Janikow, 1998; Peng, Flach, 2001). According to (Maher, Clair, 1993; Chiang, and Hsu, 1996; Peng, Flach, 2001; Marsala, 2009), fuzzy decision trees had better performance than classical decision trees.

Because of the fact that classical decision trees are not successful with uncertain data, in (Maher, Clair, 1993) fuzzy decision trees were constructed by UR-ID3 that was applied to uncertain data and several experiments were conducted. UR-ID3 was extension of the classical ID3 algorithm which was combined with fuzzy logic. In this method, uncertain data was defined by triangular membership function. In the experiments, Iris, Thyroid, Breiman datasets were used from UCI Repository and UR-ID3 algorithm was better performance than ID3.

Fuzzy decision tree was constructed by integrating decision tree and fuzzy classifiers (Chiang, and Hsu, 1996). Fuzzy classification tree (FCT) algorithm integrates the fuzzy classifiers with decision trees. Golf, Monks' Problem that is Monk1, Monk2, and Monk3, and Ionosphere data sets were used from UCI Repository. Results obtained from those 5 datasets were compared with classical

decision tree. In general, obtained results with fuzzy decision trees are more successful (Chiang, and Hsu, 1996).

A study about uncertain data was presented in (Peng, Flach, 2001) in which an application was conducted for machine fault diagnosis. Classical decision trees are sensitive to noisy data. To overcome this problem, an alternative method called soft discretization was presented. This method was based on fuzzy set theory. All samples were sorted and cut points were produced and then fuzzified. An experiment was done to show the effectiveness of the soft discretization method. Results were compared with classical decision tree using 80 samples for training and 40 instances for test. All data were correctly classified with fuzzy decision tree (Peng, Flach, 2001).

Another fuzzy decision tree algorithm was presented in (Wang, Borgelt, 2004). The aim of this study was to generate a comprehensible classification model. Information gain, and information gain ratio were used as the information measures. Also, modifications for these measures were presented for missing values and it was suggested that threshold for the information measures should be used to control the complexity of the decision tree. Additionaly, three pruning methods were presented to optimize the fuzzy rules. Five datasets from UCI Machine Learning Repository were used for experiments. Results were compared with C4.5, neural network training, and neuro fuzzy classification (NEFCLASS) which coupled neural networks and fuzzy systems. According to the results, comprehensible classifiers were obtained.

In (Janikow, 1998; Wang, Tien-chin, Lee, and Hsien-da, 2006; Cintra, Monard, and Camargo, 2012) a modified version of the decision tree was presented to generate fuzzy decision tree. Fuzzy set theory and fuzzy logic were discussed. Information theory was used to select the best attribute and split the dataset to construct a decision tree. Triangular membership function was used to fuzzify the numerical data because of its simplicity, easy comprehension and computational efficiency (Wang, Tien-chin, Lee, and Hsien-da, 2006). Results revealed that integration of both fuzzy theory and information gain make classification tasks too difficult, however it can be an alternative for classification. In (Cintra, Monard, and

Camargo, 2012), fuzzy systems based on fuzzy logic and fuzzy set theory were combined with decision tree. The FuzzyDT that is fuzzy decision tree based on C4.5 algorithm was presented in this study. 16 datasets were used from UCI Machine Learning Repository to compare C4.5 and FuzzyDT algorithm. According to the results, FuzzyDT has smaller error for 10 datasets and less number of rules than C4.5.

Also studies in medical field using fuzzy decision trees were resulted in success by (Marsala, 2009). The aims of this study were detection of diseases and making predictions to prevent patients. In this study INDANA (Individual data analysis of antihypertensive intervention) data set was used. Experiments were conducted on both classical and fuzzy decision trees and results of fuzzy decision tree were found more successful (Marsala, 2009). In (Levashenko, Zaitseva, 2012), three types of fuzzy decision tree that are non-ordered, ordered, and stable trees were presented. Fuzzy ID3 algorithm was used to learn non-ordered fuzzy decision tree. The ordered and stable fuzzy decision trees were build based on cumulative information estimations. The cumulative information estimates allow defining criterion of expanded attributes selection to induct fuzzy decision tree. The proposed approach was implemented based on medical problem benchmark with real clinical data for breast cancer diagnosis. In (Liu, Pedrycz, 2005), a new algorithmic framework for building fuzzy sets with membership functions and Axiomatic Fuzzy Set (logic) theory (AFS) were proposed. Also, fuzzy decision trees in this framework was presented.

Measures that are used to select the attribute which partitions the datasets to construct the decision tree are important for induction of the decision tree. In this thesis information gain, gain ratio, gini index and fuzzy versions of them are used as information measures and compared with each other. In (Yuan, Shaw, 1995) a fuzzy decision tree induction method which reduces classification ambiguity with fuzzy evidence was presented. Training data were fuzzified using triangular membership function. Cluster centers obtained using Kohonen's feature map (Kohonen, 1989) were used to represent triangular membership function. Small dataset was used to show efficiency of the proposed fuzzy decision tree induction method.

In (Wang, Chen, Qian, and Ye, 2000) two optimization principles of fuzzy decision trees were presented. Minimum total number and average depth of leaves were aimed by using fuzzy entropy and classification ambiguity. Also a new algorithm called Merging Branches, in short MB, was proposed to construct fuzzy decision tree. This new algorithm has decreased the number of branches, and increased the classification accuracy.

A modified version of the fuzzy ID3 algorithm that integrates information gain and classification ambiguity was introduced in (Abu-halaweh, Harrison, 2009). In the experiments, seven datasets from the UCI Repository were used and it was found that the proposed method was more successful than the original ID3 on a wide range of datasets. Also the fuzzy decision tree induction software tool was presented in (Abu-halaweh, Harrison, 2009).

A new measure, extended from classification ambiguity for fuzzy decision tree induction was proposed in (Marsala, 2012). Three measures that are entropy of fuzzy events, classification ambiguity, and extended classification ambiguity were compared using medical data. According to the results, the proposed measure has better accuracy and smaller size of obtained FDT, in average, than other measures.

An extended heuristic algorithm to build the Fuzzy ID3 was proposed in (Li, Lv, Zhang, Guo, 2010). The minimization information theory and mutual information entropy were used to avoid selecting the redundancy attributes for fuzzy decision tree induction. Several datasets were used to test the extended heuristic algorithm and compared to the Fuzzy ID3. Experimental results showed that the proposed method to build the Fuzzy ID3 improve the efficiency, simplicity and generalization capability.

Information gain was the commonly used measure in fuzzy decision trees (Quinlan, 1986; Wang, Tien-chin, Lee, and Hsien-da, 2006). Fuzzy decision tree method based on fuzzy set theory and information theory was proposed in (Wang, Tien-chin, Lee, and Hsien-da, 2006). Entropy was used to calculate the information gain. Experimental results showed that proposed method make classification tasks too difficult, but it can be an alternative for classification.Fuzzy information gain was used to construct fuzzy decision trees in (Abu-halaweh, Harrison, 2009; Chiang, and

Hsu, 1996; Umano, Okamoto, Hatono, Tamura, Kawachi, Umedzu, and Kinoshita, 1994; Yuan, Shaw, 1995; Wang, Chen, Qian, and Ye, 2000; Mitra, Konwar, Pal, 2002; Chen, Shie, 2009). Membership values of the attributes were used to compute fuzzy information gain. In (Umano, Okamoto, Hatono, Tamura, Kawachi, Umedzu, and Kinoshita, 1994) the fuzzy ID3 algorithm was used to construct the fuzzy decision tree from numerical data using fuzzy sets defined by user. The proposed method is similar to classical decision tree but it used fuzzy information gain to select the attribute. The fuzzy ID3 algorithm was applied to diagnosis of potential transformers which contain oil. According to the results, proposed method can be used to generate fuzzy rules from a set of numerical data but it has disadvantage about the number of fuzzy rules. In (Chen, Shie, 2009) the class degree, in short CD, was used to compute fuzzy information gain. A new method for constructing membership functions of a numeric feature and for classifying test instances was developed based on the proposed fuzzy information gain. The proposed method was tested on six different datasets from the UCI Machine Learning Repository. According to the results, proposed method based on fuzzy infomation gain had higher average classification accuracy rates than C4.5, naive bayes, and sequential minimal optimization (SMO) methods.A fuzzy decision tree algorithm was proposed in (Chandra, Varghese, 2009) who used gini index to learn decision tree. The proposed method was called G-FDT, and its performance was compared with gini index based crisp decision tree. 14 real life datasets from the UCI Machine Learning Repository were used for the experiments. According to the results, G-FDT algorithm is more successful than gini index based crisp decision tree in terms of accuracy and the size of tree.

In (Abu-halaweh, Harrison, 2010) features of a new freeware fuzzy decision tree tool (FDT) for supervised classification was presented. An improved version of FID3 was implemented in FDT which has four different variations of FID3 that use fuzzy information gain, classification ambiguity, fuzzy version of gini index; and integrated fuzzy information gain and classification ambiguity. Proposed fuzzy decision tree tool was applied to 8 datasets from UCI Repository. Experimental results showed that FDT had classification results better or similar than other

classification tools and versions of FDT implementation was produced the same or better classification results with lower number of rules.

3. MATERIAL AND METHOD

3.1. Material

This section includes explanations about classical decision trees, fuzzy sets, membership functions, fuzzy decision tree, splitting criteria and fuzzy versions of them, datasets used in the experiments, and Weka classification environment which were used in this thesis.

3.1.1. Classical Decision Trees

Decision tree algorithms are widely used in machine learning and applied in many real world applications for classification. Several methods have been proposed to construct a decision tree such as ID3, C4.5 (Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Quinlan, 1986) and one of the most popular decision tree algorithm is ID3, which stands for Interactive Dichotomizer 3 and it is proposed by Quinlan (Quinlan, 1986).

ID3 algorithm is applied to a set of data partitioned as traning and test data. It is based on iteration of the induction process. ID3 algorithm uses information gain based on entropy measure and applies this measure to all attributes. Finally, one attribute having maximum information gain or minimum entropy is selected as the splitting attribute for all dataset (Quinlan, 1986; Janikow, 1998; Mitra, Konwar, Pal, 2002).

ID3 algorithm can be described in detail as follows (Quinlan, 1986; Janikow, 1998; Mitra, Konwar, Pal, 2002):

1. Compute the entropy of all attributes in the dataset by using the following equation:

$$\text{Entropy} = -\sum_{i=1}^{c}(p_i \times \log_2 p_i) \tag{3.1}$$

where p_i is the probability of class i in the dataset and c is the number of class in the dataset.

2. Select the attribute having minimum entropy as the root node of the decision tree.
3. Build sub nodes from branches of the root node.
4. Repeat steps 1 through 3, until termination condition is met.

Conditions for stopping partitioning are:

1. All samples belong to the same class.
2. There are no remaining attributes for partitioning.
3. There are no samples left.

After building the decision tree, a new data is classified with a decision tree that is built in the induction process. So rules are created from the tree. Induction of rules starts from the root node. Each path of branches from root to leaf can be converted into an "*IF… THEN…*" rule.

Let $A=\{A_1,\ldots,A_k\}$ be a set of attributes, $T(A_k)=\{T_1,\ldots,Ts_k\}$ be a set of samples that are used for one attribute and $C=\{C_1,\ldots, C_j\}$ be a set of classes. A classification rule having the condition part representing the attributes at the branches as the rule antecedent, and the conclusion part representing the class at the leaf as the rule consequent is IF (A_1 is Ti_1) AND …… (A_k is Ti_k) THEN (C is C_j) (Janikow, 1998; Yuan, Shaw, 1995).

Table 3.1 shows a small training dataset named "Weather Dataset" that has four attributes and the class label column (Quinlan, 1986).

Four attributes of the "Weather Dataset" are as follows:

i. Outlook attribute has values {Sunny, Overcast, Rainy}
ii. Temperature attribute has values {Hot, Mild, Cool}
iii. Humidity attribute has values {High, Normal}
iv. Windy attribute has values {True, False}

Class attribute of the "Weather Dataset" has values Play = {Yes, No}

Table 3.1. Weather Dataset (Quinlan, 1986)

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

An example decision tree of the "Weather Dataset" is shown in Figure 3.1. Leaves of the decision tree are the class names "Yes" or "No" as shown in the figure. "Outlook" is the root node having 3 branches that are the values of the attribute. Internal nodes "Humidity" and "Windy" also have two branches since they have two attribute values. Decision trees represent a set of rules for classification a new sample. For example according to the decision tree given in Figure 3.1, if "Outlook" attribute has value "Overcast", the sample is classified as "Yes".

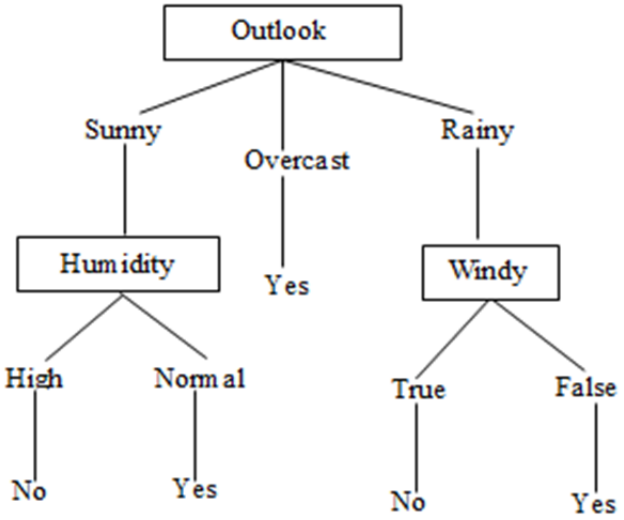

Figure 3.1. An Example of Decision Tree of theWeather Dataset (Quinlan, 1986)

Several different decision trees can be inducted from the same dataset. Another decision tree for the same "Weather Dataset" is shown in Figure 3.2. It is more complex decision tree than the one shown in Figure 3.1.

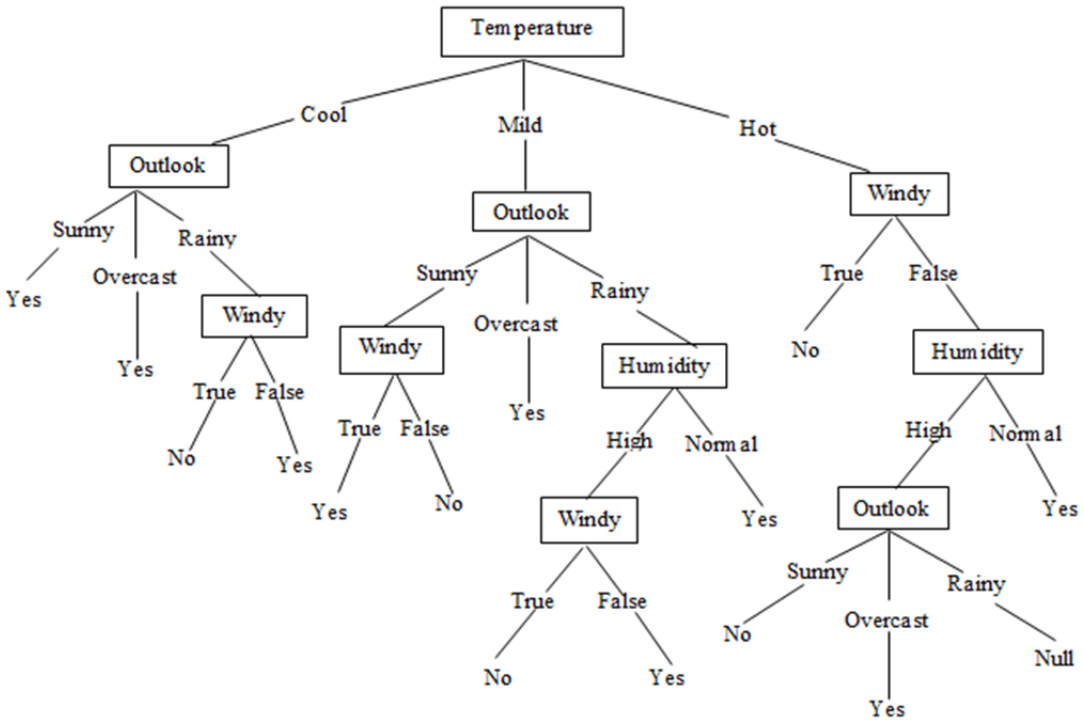

Figure 3.2. An Example of Complex Decision Tree of the Weather Dataset (Quinlan, 1986)

In classical decision tree algorithms, if the dataset used consists of the continuous attribute values, they must be discretized before using in the algorithm. Attribute values can be partitioned into several intervals using a set of cut points that have sharp edges. Because of the sharp edges this method has sensitivity with imprecise samples and in the real world problems where vague and imprecise data exists (Peng, Flach, 2001; Wang, Tien-chin, Lee, and Hsien-da, 2006). So for this type of samples, fuzzy decision trees are used as described in the next section.

3.1.2. Fuzzy Set Theory

Fuzzy set theory was formalized by Zadeh (Zadeh, 1965) at the University of California in 1965 to deal with uncertainty, and inexact data. Fuzzy set theory is also known as possibility theory. Fuzzy sets are extension of the classical set theory. In classical set theory, a value either belongs to a particular set or not. However, in the real world there are many values that cannot be classified with one set exactly. In fuzzy set theory, each element is represented by a membership function $\mu_A(x)$ and has a membership degree in the interval [0..1]. Membership degree refers to probability of belonging to a set for an element (Janikow, 1998; Wang, Tien-chin, Lee, and Hsien-da, 2006; Rokach, Maimon, 2008; Janikow, 1996). The set of "tall man" example is shown below. In Figure 3.3, membership function with sharp-edge is shown for "Tall" attribute. Sharp-edges in the figure mean that the person is tall or not. For example; the man having 5' height is not a tall man, while the man who has 6' height is a tall man (Riberio, 2015).

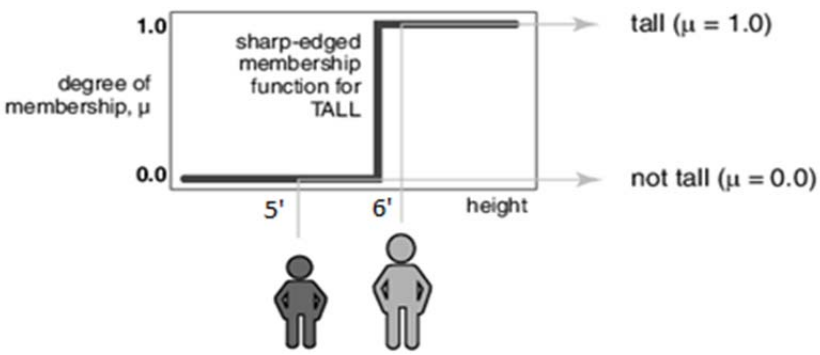

Figure 3.3. Sharp-edged Membership Function for "Tall" Attribute (Riberio, 2015)

On the contrary to sharp edges, fuzzy sets are more appropriate to the real world. For example some people are very tall, while others are just tall. As shown in Figure 3.4, the man who has 5' height is not very tall person and he has 0.3 membership value for tall set, and the other man whose height is 6' has 0.95 membership value for tall set and he is a tall person. So, both men belong to the set of "tall man" with different membership degrees (Riberio, 2015).

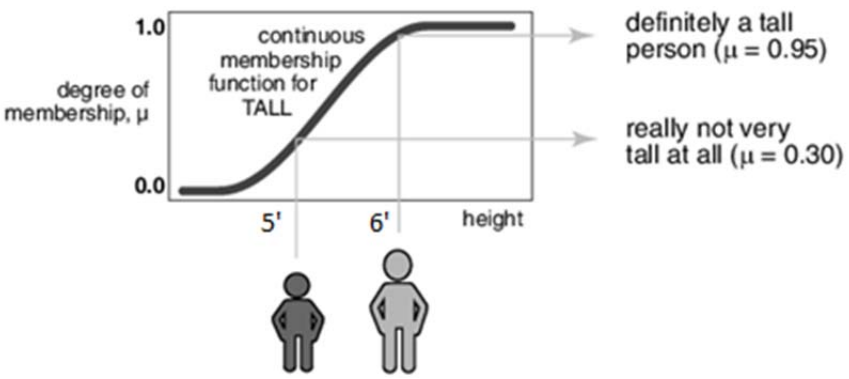

Figure 3.4. Fuzzy Membership Function for "Tall" Attribute (Riberio, 2015)

3.1.3. Membership Functions

Membership function takes values in the interval [0..1] and indicates the degree of membership of an element to a set. $\mu_A(x)=0$ means that x is not certainly a member of set A and $\mu_A(x)=1$ means that x is certainly a member of A (Rokach, Maimon, 2008). Membership values represented with the fuzzy linguistic variables also named as the fuzzy term. For example, "pressure" is a continuous variable and five linguistic terms, such as "weak", "low", "medium", "powerful" and "high", can be used when it becomes a fuzzy variable. This is illustrated in Figure 3.5 where "pressure" variable u belongs to "weak" and "low" domains with different degrees.

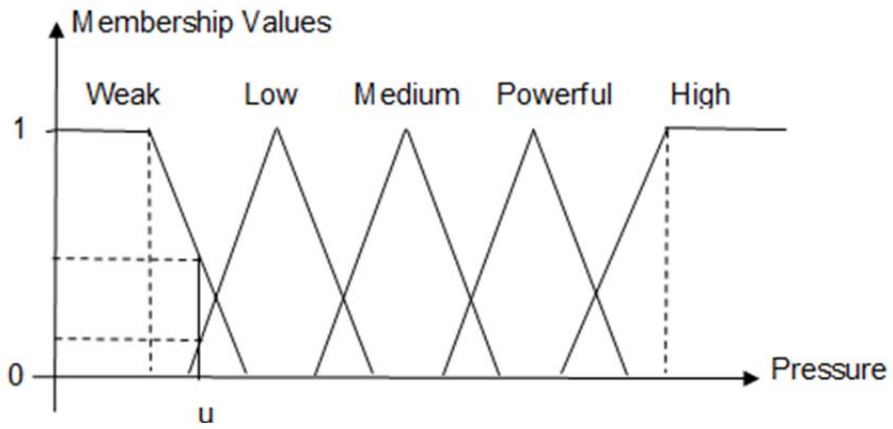

Figure 3.5. An Example of Linguistic Terms of the Pressure Attribute

Commonly used fuzzy membership functions are: triangular, and trapezoidal (Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Yuan, Shaw, 1995; Wang, Tien-chin, Lee, and Hsien-da, 2006; Rokach, Maimon, 2008; Mitra, Konwar, Pal, 2002; Au, Chan, Wong, 2006; Kuwajima, Nojima, Ishibuchi, 2008). In this study, triangular and trapezoidal membership functions are used to fuzzify the numerical data in the datasets.

3.1.3.1. Triangular Membership Function

Triangular membership function is one of the most commonly used membership function. It is described by the three corners as shown in Figure 3.6.

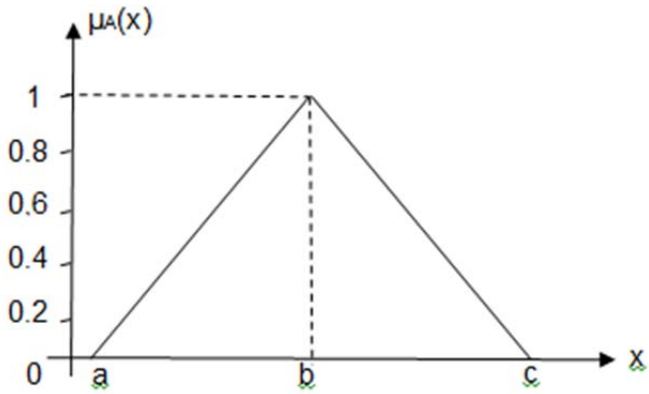

Figure 3.6. Triangular Membership Function (Wang, Tien-chin, Lee, and Hsien-da, 2006)

Triangular membership function is computed as follows:

$$\mu_F(x,a,b,c) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases} \tag{3.2}$$

where x is the sample of the attribute in the dataset and a, b, c represent the x coordinates of the three vertices of $\mu_A(x)$ in a fuzzy set A (a: lower boundary and c: upper boundary where membership degree is zero, b: the center where membership degree is 1) (Yuan, Shaw, 1995; Wang, Tien-chin, Lee, and Hsien-da, 2006; Mitra, Konwar, Pal, 2002).

An example of the triangular membership function is shown in Figure 3.7 in which it is illustrated that “age” attribute is defined as “young”, “early adulthood”, “middle-aged” and “old age”. For example, someone who is younger than 10, belongs to the “young” class with 1 membership degree. If age is between 10 and 30, age belongs to the “young” and “early adulthood” domains with different degrees. If someone who is older than 30, age is not certainly a member of the “young” class, this age belongs to one of the “early adulthood”, “middle-aged” and “old age” classes.

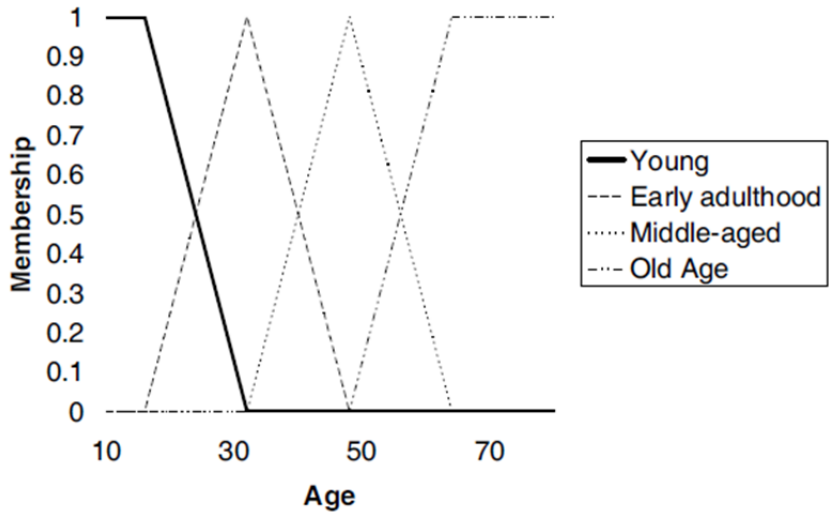

Figure 3.7. An Example of the Triangular Membership Function in the Age Attribute (Rokach, Maimon, 2008)

3.1.3.2. Trapezoidal Membership Function

Trapezoidal membership function is described by the four corners as shown in Figure 3.8.

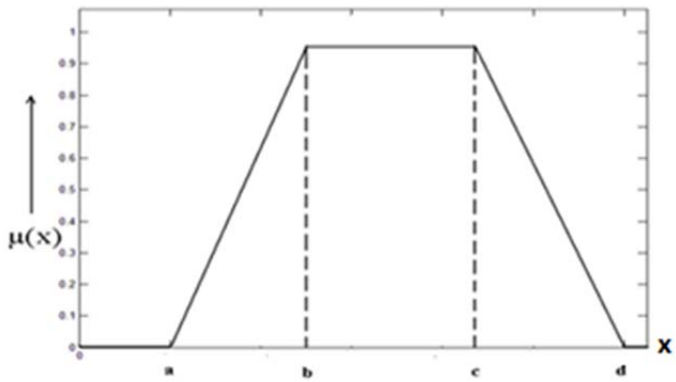

Figure 3.8. Trapezoidal Membership Function

Trapezoidal membership function is computed as follows:

$$\mu_F(x,a,b,c,d) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b < x < c \\ \dfrac{d-x}{d-c}, & c \le x \le d \\ 0, & x \ge 0 \end{cases} \tag{3.3}$$

where a, b, c and d represent the x coordinates of the four vertices of $\mu_A(x)$ in a fuzzy set A (a: lower boundary and d: upper boundary where membership degree is zero, b and c: the center where membership degree is 1) (Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Au, Chan, Wong, 2006; Kuwajima, Nojima, Ishibuchi, 2008).

An example of trapezoidal membership function of the "income" attribute defined as three linguistic terms, such as "low", "medium" and "high" is illustrated in Figure 3.9 (Wang, Tien-chin, Lee, and Hsien-da, 2006; Han, Kamber, 2006). If income is 15.000, it means that income is "low" and if income is 25.000, it belongs to "medium" class.

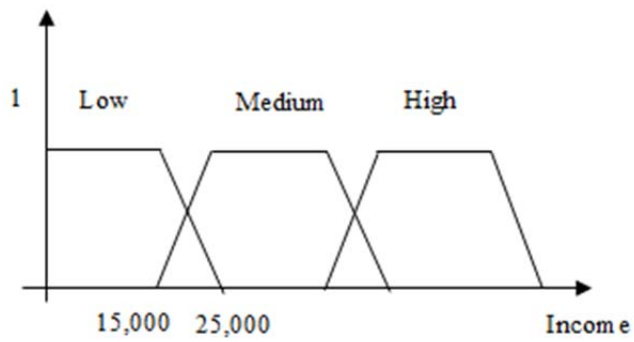

Figure 3.9. Trapezoidal Membership Function of Income (Janikow, 1998)

3.1.4. Fuzzy Decision Trees

Fuzzy decision trees are the extension of classical decision trees. It has become popular solution for several industrial problems having incorrect, ambiguous, noise and missing data (Janikow, 1998). In fuzzy decision trees, each node is a subset of a universal set. All leaf nodes are fuzzy part of universal set. Due to fuzzy reasoning this method is robust and provides generalization in classification.

It indicates probability of belonging to a class of data. Degrees of accuracy of belonging to a class are between 0 and 1 (Peng, Flach, 2001).

Differences between classical and fuzzy decision trees are as follows (Lee, Lee, and Lee-kwang, 1999; Peng, Flach, 2001; Wang, Chen, Qian, and Ye, 2000; Wang, Tien-chin, Lee, and Hsien-da, 2006);

- In fuzzy method, data can be assigned to more than one class, while in classical method; each data is assigned to a single class. Because, probabilities of belonging to a class is achieved with fuzzy method.
- In classical method, a path from root to a leaf represents a product rule. In contrast to classical method, fuzzy method represents a fuzzy rule with degree of accuracy.
- In classical method, nodes are subsets of crisp set. In fuzzy method, nodes are fuzzy subsets.
- Fuzzy decision trees give more successful results with ambiguity, incorrect data etc. Obtained results are closer to human thinking.
- In classical decision trees, a small change of value in data effects result. In other words, classical method is sensitive to noise and possibilities of incorrect classification are high.

The fuzzy decision tree induction consists of the following steps (Abu-halaweh, Harrison, 2009; Janikow, 1998; Yuan, Shaw, 1995; Wang, Tien-chin, Lee, and Hsien-da, 2006):

a) Data fuzzification
b) Inducing a fuzzy decision tree
c) Extracting classification rules from fuzzy decision tree
d) Applying fuzzy rules for classification

For fuzzy decision tree, continuous features are represented by fuzzy sets before the induction process. So data fuzzification is usually applied to numerical

data. A certain numeric attribute needs to be fuzzified into linguistic terms before it can be used in the algorithm. The fuzzification process can be performed manually by experts or it can be derived automatically. Also fuzzy membership functions can be used. Commonly used membership functions are triangular and trapezoidal membership functions (Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Wang, Tien-chin, Lee, and Hsien-da, 2006).

Fuzzy decision tree algorithm is similar to ID3 algorithm. Fuzzy decision tree building procedure partitions the dataset recursively based on the values of the selected attribute. To select the attribute to split the dataset, attribute selection measure also named as splitting criterion is used. Commonly used splitting criteria are information gain, gain ratio, and gini index (Abu-halaweh, Harrison, 2009; Chiang, and Hsu, 1996; Janikow, 1998; Lee, Lee, and Lee-kwang, 1999; Maher, Clair, 1993; Marsala, 2009; Peng, Flach, 2001; Umano, Okamoto, Hatono, Tamura, Kawachi, Umedzu, and Kinoshita, 1994; Abu-halaweh, Harrison, 2010; Han, Kamber, 2006). Also in this thesis fuzzy format of the splitting criteria are used: fuzzy information gain, fuzzy gain ratio, fuzzy gini index and they are explained in the next section.

3.1.5. Splitting Criteria

3.1.5.1. Information Gain

Information gain uses entropy measure. Entropy, measures the quality of samples. Large entropy means that data set is impure. If the entropy is 0, dataset is totally pure that all samples belong to the same class (Lee, Lee, and Lee-kwang, 1999; Wang, Tien-chin, Lee, and Hsien-da, 2006; Rokach, Maimon, 2008).

Entropy is calculated as in equation 3.1. Attribute having maximum information gain is selected. Hence information gain is based on decrease in entropy. Information gain for an attribute is computed as follows:

$$Information\ Gain = Entropy(Dataset) - Entropy(Attribute) \quad (3.4)$$

where *Entropy(Attribute)* is computed as follows:

$$Entropy(Attribute) = -\sum_{v \in Values(A)} \frac{|S_v|}{|S|} \times Entropy(S_v) \quad (3.5)$$

where $|S_v|$ is the number of samples in the subset S_v and $|S|$ is the number of all samples in the dataset S.

Information gain is calculated as in follows (Han, Kamber, 2006):

1. The entropy is calculated for all dataset
2. The entropy is calculated for all attributes
    - For this step, entropy is calculated for each value of the attribute.
    - Results are added proportionally to get the total entropy.
3. The information gain is calculated for all attributes according to equation 3.4.
4. Attribute having maximum information gain (minimum entropy) is selected

As an example, information gain is calculated for outlook attribute from the weather dataset given in Table 3.1 (Wang, Tien-chin, Lee, and Hsien-da, 2006; Han, Kamber, 2006) as follows:

1. All data set has 14 samples and class labels of 9 of them are “yes”, 4 of them are “no”. So, the entropy of the dataset is calculated as follows:
   Entropy (Play) = – (9/14× $\log_2$ (9/14))–(5/14 ×$\log_2$ (5/14)) = 0.940
2. Now, let’s compute entropy of the Outlook attribute. This attribute has three different values: Sunny, Overcast and Rainy.

a. For Outlook = "Sunny", class labels are as follows: 3 of the 5 samples are labeled as "yes" and 2 of the 5 samples are labeled as "no". So, the entropy for Outlook="Sunny" is equal to

Entropy (Outlook = Sunny) = $-(3/5 \times \log_2 (3/5)) - (2/5 \times \log_2 (2/5)) = 0.971$

b. For Outlook = "Overcast", class labels are as follows: 4 of the 4 samples are labeled as "yes" and there is no sample for label "no". So, the entropy for Outlook="Overcast" is equal to
Entropy (Outlook = Overcast) = $-(4/4 \times \log_2 (4/4)) - (0/4 \times \log_2 (0/4)) = 0$

c. For Outlook = "Rainy", 2 of the 5 samples are labeled as "yes" and 3 of them are labeled as "no". So, the entropy for Outlook="Rainy" is equal to
Entropy (Outlook = Rainy) = $-(2/5 \times \log_2 (2/5)) - (3/5 \times \log_2 (3/5)) = 0.971$

d. Total entropy for Outlook attribute is computed as follows:

Entropy (Outlook) = $5/14 \times 0.971 + 4/14 \times 0 + 5/14 \times 0.971 = 0.694$

3. Now, information gain of Outlook can be calculated as follows:
Information Gain (Outlook) = $0.940 - 0.694 = 0.246$

Information gains of the other attributes are computed similarly and they are listed below:

- Information Gain(Temperature) = 0.029
- Information Gain(Humidity) = 0.151
- Information Gain(Windy) = 0.048

After the information gain is calculated for all the attributes, an attribute having the highest information gain is selected. So, "Outlook" is selected as the root node to build the tree.

3.1.5.2. Gain Ratio

Gain ratio is a normalized form of the information gain (Rokach, Maimon, 2008; Han, Kamber, 2006) using a "split information" value and it is computed as follows:

$$Gain\ Ratio(A) = \frac{Information\ Gain\ (A)}{Split_{Info}(A)} \tag{3.6}$$

Gain ratio is maximum when split info is minimum. If split info is zero, gain ratio is not defined. $Split_{Info}(A)$ is computed as follows:

$$Split_{Info}(A) = -\sum_{i=1}^{v}\left(\frac{|S_i|}{|S|} \times log_2 \frac{|S_i|}{|S|}\right) \tag{3.7}$$

where v is the number of branches of the attribute and $|S_i|$ is the number of the samples in the each branches.

To calculate gain ratio:

i. Information gain of attribute A is calculated firstly.
ii. Then it is divided by the split info.

The attribute having maximum gain ratio is selected as best useful attribute.
Example: gain ratio computation for "Outlook" attribute is as follows:
Information Gain(Outlook) = 0.246 (as calculated before)

$Split_{Info}$(Outlook) = – (5/14 × log (5/14)) –(4/14 × log (4/14)) –(5/14 × log (5/14)) = 1.577

Gain ratio (Outlook) = 0.246 / 1.577 = 0.156

3.1.5.3. Gini Index

Gini index measures impurity of dataset *S* as in the following formula (Rokach, Maimon, 2008; Abu-halaweh, Harrison, 2010; Han, Kamber, 2006).

$$Gini(S) = 1 - \sum_{i=1}^{c} {p_i}^2 \tag{3.8}$$

where p_i is the probability of class *i* in the dataset and *c* is the number of class in the dataset. *S* is the dataset and p_i is the proportion of class *i* in the set *S*. The gini index value of an attribute *A* is also named as reduction impurity and it is computed as follows:

$$\Delta Gini(A) = Gini(S) - Gini_A(S) \tag{3.9}$$

where *A* is the attribute in the dataset *S*. Gini index of an attribute *A* is computed as follows:

$$Gini_A(S) = \sum_{i=1}^{v} \frac{|S_i|}{|S|} Gini(S_i) \tag{3.10}$$

where *v* is the number of distinct values of the attribute, $|S_i|$ is the number of samples in the subset S_i and $|S|$ is the number of all samples in the dataset *S*.

To calculate the gini index of the attribute:

1. First of all, gini index for the dataset *S* is calculated.
2. Then, the gini index is calculated for all attributes

a. For this step, gini index is calculated for each different value of the attribute.
b. Gini index computations for each value of the attribute are added proportionally to get the total gini index for the attribute.

3. Finally, reduction in impurity is computed for each attribute.

The attribute having minimum gini index is selected for splitting. In other words, the attribute maximizing the reduction in impurity is the best attribute.

As an example gini index calculation for "Outlook" attribute from weather dataset is as follows:

1. All data set has 14 samples and class label of 9 of them are "yes", and 4 of them are "no". So gini index of dataset is computed as follows:
   Gini(Play) = $1 - [(9/14)^2 + (5/14)^2] = 0.459$
2. Now, let's compute gini of the "Outlook" attribute. This attribute has three values: "Sunny", "Overcast" and "Rainy".

   a. For "Outlook" = "Sunny", 3 of the 5 samples are labeled as"yes" and 2 of the 5 samples are labeled as "no".
      Gini (Outlook = sunny) = $1 - [(3/5)^2 + (2/5)^2] = 0.48$
   b. For Outlook = "Overcast", 4 of the 4 samples are labeled as "yes" and there is no sample labeled as "no".
      Gini(Outlook = Overcast) = $1 - [(4/4)^2 + (0/4)^2] = 0$
   c. For Outlook = "Rainy", 2 of the 5 samples are labeled as"yes" and 3 of the 5 samples are labeled as "no".
      Gini (Outlook = Rainy) = $1 - [(2/5)^2 + (3/5)^2] = 0.48$
   d. Total entropy for outlook attribute is a computed as follows:
      Gini(Outlook) = 5/14 * 0.48 + 4/14 * 0 + 5/14 * 0.48 = 0.171

3. Reduction of impurity of "Outlook" is computed as follows:

ΔGini(Outlook) = 0.459 – 0.171= 0.288

3.1.5.4. Best Split Method for Numerical Data

If the dataset contains only numerical attributes, best split method can be used to discretize a continuous numeric valued attribute. It is used to find best split point of the attribute *A* using all data in the attribute *A* and applied similarly to all numeric attributes in the dataset *D*. Best split point is also named as cut point or mid-point (Han, Kamber, 2006).

To find the best split point of the attribute in the dataset, the following algorithm can be used (Han, Kamber, 2006):

1. Firstly, all samples of the attribute *A* are sorted in ascending order.
    - Assume that *D* is the dataset, $\{a_1, a_2,\ldots,a_i, a_j,\ldots,a_v\}$are sorted values of attribute *A* of dataset *D* and *v* is the number of samples in the attribute *A*.
2. Then, all cut points arecomputed and *v*-1 different cut points are obtained for attribute *A*. Cut points are computed as follows:
    a. The average of each consecutive data values a_i and a_j is computed as in equation 3.11:

$$Cut\ point = \frac{a_i + a_j}{2} \tag{3.11}$$

    b. Now, for each cut point, there are two conditions:
        i. *D*1 is the first subset of the attribute *A* that has values less than or equal to cut-point.

ii. $D2$ is the second subset of the attribute A that has values greater than cut-point.

c. Then, information gain is computed for each subset $D1$ and $D2$ to choose the one with the best information gain. Information gain is computed for two subsets $D1$ and $D2$ as follows:

$$Info\ Gain(D, A) = \frac{|D_1|}{|D|} Entropy(D_1) + \frac{|D_2|}{|D|} Entropy(D_2) \tag{3.12}$$

d. The cut point having minimum information gain also named as expected information requirement is selected as the best splitting point for attribute A.

3. First two steps are repeated for all attributes in the dataset D and the attribute having the best information gain is selected to form the decision tree.

As an example, there is a numerical attribute namely "temperature" of the weather dataset (Witten, Frank, 2005) and the values of this attribute in ascending order are given below:

Table 3.2. Values of the "Temperature" Attribute of Numerical Weather Dataset and Their Class Labels

Values	64	65	68	69	70	71	72	72	75	75	80	81	83	85
Class Label	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	No

The computed cut-point values for the temperature attribute are presented in Figure 3.10.

Values	64	65	68	69	70	71	72	72	75	75	80	81	83	85
Cut-point	64.5	66.5	68.5	69.5	70.5	71.5	72	73.5	75	77.5	80.5	82	84	

Figure 3.10. Temperature Attribute with Cut-point Values

1. As an example,the cut-point between values 71 and 72 is (71+72)/2=71.5
2. For the condition "Temperature ≤ 71.5", class labels are as follows: 4 of the 6 samples are labeled as "yes" and 2 of the 6 samples are labeled as "no".
3. For the condition "Temperature >71.5", 5 of the 8 samples are labeled as "yes" and 3 of the 8 samples are labeled as "no".
4. Information gain for the cut point 71.5 is computed as below:

$\text{Information_Gain}_{71.5}$ ([4, 2], [5, 3]) = 6/14 × Entropy ([4, 2]) + 8/14 × Entropy ([5, 3]) = 0.939

3.1.6. Fuzzy Form of Splitting Criteria

In this study, the fuzzy forms of splitting criteria are used and compared to the basic forms of splitting criteria. Membership values of the samples in the fuzzy data are used to compute fuzzy forms of splitting criteria. These fuzzy measures are fuzzy information gain, fuzzy gain ratio and fuzzy gini index and in the following, details of the fuzzy form of splitting criteria are defined.

3.1.6.1. Fuzzy Information Gain

The fuzzy form of information gain *FIG*(*S*, *A*) (Chen, Shie, 2009) of a fuzzy feature *A* is defined as follows:

$$FIG(S,A) = -\sum_{i=1}^{c}(p_i \, log_2 \, p_i) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} FE(S_v) \tag{3.13}$$

where $|S_v|$ is the number of samples in the subset S_v and $|S|$ is the number of all samples in the dataset S and p_i is the proportion of class i in the set.

The fuzzy entropy $FE(S_v)$ (Chen, Shie, 2009) of a subset of training instances is defined as in equation 3.14.

$$FE(S_v) = -\sum_{i=1}^{c}(CD_c(v) \, log_2 \, CD_c(v)) \tag{3.14}$$

where $CD_c(v)$ is the class degree measure which denotes the probability of the training instance v belonging to a class c.

The class degree $CDc(v)$ (Chen, Shie, 2009) is computed as follows:

$$CD_c(v) = \frac{\sum_{x \epsilon A_c} \mu_v(x)}{\sum_{x \epsilon A} \mu_v(x)} \tag{3.15}$$

where $\mu_v(x)$ is the membership value of sample x of attribute A belonging to the fuzzy set v, and $\mu_v(x) \in [0,1]$. c denotes the class, and A_c denotes the set of the values of the feature A of the subset S_v of the training instances belonging to the class c.

3.1.6.2. Fuzzy Gain Ratio

In this thesis, fuzzy gain ratio is a normalized form of the fuzzy information gain using a "Split Information" value. It is calculated according to equation 3.16.

$$Fuzzy\ Gain\ Ratio(A) = \frac{FIG\ (S,A)}{Split_{Info}(A)} \tag{3.16}$$

where *FIG*(S, *A*) denotes the fuzzy information gain of attribute *A* in the dataset *S* and it is computed as in equation 3.13, the $Split_{Info}(A)$ is computed as follows:

$$Split_{Info}(A) = -\sum_{i=1}^{v}\left(\frac{|S_i|}{|S|} \times log_2 \frac{|S_i|}{|S|}\right) \tag{3.17}$$

where *v* is the number of branches of the attribute and $|S_i|$ is the number of the samples in the each branches.

3.1.6.3. Fuzzy Gini Index

In this thesis, fuzzy form of the gini index is obtained by using class degree measure presented in equation 3.15. The fuzzy gini index of attribute *A* in the dataset *S* is formulated as follows:

$$\Delta Fuzzy\ Gini(A) = 1 - \sum_{i=1}^{c} {p_i}^2 - Fuzzy\ Gini_A(S) \tag{3.18}$$

where p_i is the proportion of class *i* in the set *S* and $FuzzyGini_A(S)$ denotes the fuzzy gini index measures of the attribute *A*.

$$Fuzzy\ Gini_A(S) = \sum_{i=1}^{v} \frac{|S_v|}{|S|} Fuzzy\ Gini(S_v) \tag{3.19}$$

where *v* denotes the each branch of the attribute in the dataset *S*, $|S_v|$ is the number of samples in the subset S_v and $|S|$ is the number of all samples in the dataset *S*.

$$Fuzzy\ Gini(S_v) = 1 - \sum_{v=1}^{c} CD_c(v)^2 \tag{3.20}$$

where $CD_c(v)$ is the class degree measure denoted the probability of the training instance v belonging to a class c and it is computed as in equation 3.15.

3.1.7. Performance Measures

This section includes explanation for coverage, accuracy, and F-measure (Han, Kamber, 2006) which were used to evaluate performance of rule and classification model.

3.1.7.1. Rule Performance Measures

Coverage and accuracy measures are used to obtain quality of the classification rules. In this thesis, these measures are used to select rules if there are more than one rule satisfying the same test data (Han, Kamber, 2006).

Coverage: Coverage of the rule is the ratio of number of the test instances satisfied by the rule to the number of all instances in the test dataset (Han, Kamber, 2006).

$$Coverage(R) = \frac{n_{covers}}{|D|} \tag{3.21}$$

where R is the one of the rules, $|D|$ is the number of instances in the test dataset and n_{covers} is the number of instance satisfied by rule R.

Accuracy: Accuracy of the rule $R$is the ratio of number of correctly classified test instances by the rule R to the number of test instances satisfied by the rule R (Han, Kamber, 2006).

$$Accuracy(R) = \frac{n_{correct}}{n_{covers}} \tag{3.22}$$

where R is the one of the rules, n_{covers} is number of instances satisfied by R, and $n_{correct}$ is the number of instances correctly classified by R.

3.1.7.2. Measuring the Performance of the Classification Model

To measure the performance of the model obtained during the training process; true positive rate, false positive rate, precision, recall and F-measure values are used. To compute these values a confusion matrix which is also named as contingency table is used. A confusion matrix contains both actual and predicted class labels of test datasets. A confusion matrix (Han, Kamber, 2006) for two class classifier is shown in Table 3.3.

Table 3.3. A 2x2 Confusion Matrix (Han, Kamber, 2006)

		Predicted class label	
		Positive	Negative
Known class label	Positive	TP	FN
	Negative	FP	TN

- TP is the true positive case that means the actual class is positive and the predicted class is also positive.
- FN is the false negative case that means the actual class is positive and the predicted class is negative.
- FP is the false positive case that means the actual class is negative and the predicted class is positive.
- TN is the true negative case that means the actual class is negative and the predicted class is negative.

Table 3.4 shows the confusion matrix for a dataset having three classes. All diagonal elements in Table 3.4 show the correct number of classification for each class and other elements show the wrong classification. *TP* in Table 3.4 represents the correct number of the classification. For example, TP_A shows the correct number of the classification for *A* class. In Table 3.4, *E* represents the error. For example, E_{AB}

represents the wrong number of the classification for class *A*. It means that test samples have *A* class label but they are predicted as *B* class label.

Table 3.4. A Confusion Matrix for More Than Two Class (Witten, Frank, 2005)

		Predicted class label		
		A	B	C
Known class label	A	TP_A	E_{AB}	E_{AC}
	B	E_{BA}	TP_B	E_{BC}
	C	E_{CA}	E_{CB}	TP_C

Precision: Precision is the proportion of the correctly predicted positive cases to all predicted positive cases (Han, Kamber, 2006). Precision is calculated by using equation 3.23.

$$Precision = \frac{TP}{TP + FP} \tag{3.23}$$

For the precision of the class *A* in Table 3.3 it is calculated as following:

$$Precision_A = \frac{TP_A}{TP_A + E_{BA} + E_{CA}} \tag{3.24}$$

Recall: Recall is the proportion of actual positives which are correctly identified. Recall is also called as true positive rate or sensitivity (Han, Kamber, 2006), and it is calculated using equation 3.25.

$$Recall = Sensitivity = True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{3.25}$$

For the recall of the class *A* in Table 3.3 is calculated as follows:

$$Recall_A = \frac{TP_A}{TP_A + E_{AB} + E_{AC}} \tag{3.26}$$

False Positive Rate: False positive rate (Han, Kamber, 2006) is the proportion of negative cases which are incorrectly classified as positive cases, calculated as follows:

$$False\ Positive\ Rate = \frac{FP}{FP + TP} \tag{3.27}$$

For the false positive rate of the class *A* in Table 3.3 is calculated as follows:

$$False\ Positive\ Rate_A = \frac{E_{BA} + E_{CA}}{E_{BA} + E_{CA} + TP_A} \tag{3.28}$$

F_Measure: F_Measure (Han, Kamber, 2006) is the harmonic average of the precision and the recall that is calculated as follows:

$$F_Measure = 2 * \frac{precision * recall}{precision + recall} \tag{3.29}$$

3.1.8. Weka Data Mining Tool

Weka (http://www.cs.waikato.ac.nz/ml/weka) is a machine learning software written in Java for data mining tasks that contain data preprocessing, classification, clustering, regression, association rule mining, and feature selection. Weka was developed at the University of Waikato in New Zealand and it is short form of Waikato Environment for Knowledge Analysis.

Figure 3.11. The GUI of the Weka

3.1.8.1. Weka Data Formats

Attribute Relation File Format (ARFF) is the default file type used in Weka but Comma Separated Values (CSV) in text file is also supported. Additionally data can be read from a database.

ARFF files have two parts. The first part is Header information that contains relation and attribute declarations. @RELATION line defines the name of the relation. @Attribute declaration contains name and type of the attribute. The attribute type can be numeric (real or integer), string, nominal, and date. The second part of the arff file is data part that contains instances in the dataset. Each row in the data section represents an instance that contains attribute values separated with commas.

An example of arff file format of Iris dataset is shown in Figure 3.12. The part of the Iris dataset from UCI Repository presented in the figure has 8 instances, 3 classes named as iris-setosa, iris-versicolor, and iris-virginica; and 4 attributes which contain numerical samples named as sepal length, sepal width, petal length, and petal width. Each class in this dataset refers the type of iris plant. In our study, arff file format was used to form train and test datasets with Weka.

```
@RELATION iris

@ATTRIBUTE sepallength REAL
@ATTRIBUTE sepalwidth REAL
@ATTRIBUTE petallengthREAL
@ATTRIBUTE petalwidth REAL
@ATTRIBUTE class {Iris-setosa, Iris-versicolor, Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
6.3,2.5,5.0,1.9,Iris-virginica
6.5,3.0,5.2,2.0,Iris-virginica
6.2,3.4,5.4,2.3,Iris-virginica
5.9,3.0,5.1,1.8,Iris-virginica
```

Figure 3.12. Example of arff File Format in Weka

3.1.9. Datasets

In this thesis 18 datasets were used from UCI Machine Learning Repository. The datasets used are presented in Table 3.5.

Table 3.5. UCI datasets used in the experiments

DataSet	**# of instances**	**# of attributes**	**# of classes**
Heart Statlog	270	13	2
Mammographic Masses	961	5	2
Breast Cancer	683	9	2
Diabetes	768	8	2
Hepatitis	155	19	2
Spect Heart	267	22	2
Yeast	1484	8	10
Vertebral Column 2C	310	6	2
Vertebral Column 3C	310	6	3
Ecoli	336	8	8
Balance Scale	625	4	3
Thyroid	215	5	3
LD Bupa	345	6	2
Iris	150	4	3
Glass	214	9	7
Monk1	556	6	2
Monk2	601	6	2
Monk3	554	6	2

Each dataset were partitioned into the training and the test sets. 75% of each dataset were used for training process to build the model, and the remaining 25% part of the dataset were used for testing to validate the model which were built from the training data. In this study, to partition all the datasets for training and test phases, Weka Data Mining Tool was used.

To partition a dataset as training and test sets, following steps were applied:

1. The dataset file was converted to arff format to work with Weka.
2. All data in the arff file were distributed randomly with “Randomize” filter in Weka.
3. 75% of the data set were selected for the training by using “RemovePercentage” in Weka and remaining 25% of the data set was selected for the test phase.

The number of instances in the training and the test datasets are given in Table 3.6.

Table 3.6. Partitioned Datasets

DataSet	**total # of instances**	**# of instances for training**	**# of instances for testing**
Heart Statlog	270	202	68
Mammographic Masses	961	721	240
Breast Cancer	683	512	171
Diabetes	768	576	192
Hepatitis	155	116	39
Spect Heart	267	200	67
Yeast	1484	1113	371
Vertebral Column 2C	310	232	78
Vertebral Column 3C	310	232	78
Ecoli	336	252	84
Balance Scale	625	469	156
Thyroid	215	161	54
LD Bupa	345	259	86
Iris	150	112	38
Glass	214	160	54
Monk1	556	417	139
Monk2	601	451	150
Monk3	554	415	139

If the dataset has missing values, average value of the corresponding attribute are used for filling the missing value. If the datasets are already partitioned as training and test data, the partitions are merged and then the partitioning steps as mentioned above are applied.

3.2. Method

In this thesis, fuzzy ID3 and basic ID3 algorithms were implemented and compared with each other to show the effect of fuzzification of data and decision tree on the classification performance. One method for classical decision tree and two forms of fuzzy decision tree methods were prepared. Java programming languagewas

used with Netbeans IDE. Both of these implementations include training and test phases.

Methods used in this study are explained in the subsequent sections which are data fuzzification, inducing a fuzzy decision tree, extracting classification rules from fuzzy decision tree, applying fuzzy rules for classification, and inducing classical decision tree using ID3.

3.2.1. Data Fuzzification

Fuzzification process is usually applied to numerical data. In this thesis, triangular and trapezoidal membership functions are used to fuzzify the numerical data and three linguistic terms are used for each numerical data. Three linguistic terms are obtained by fuzzification using both triangular and trapezoidal membership functions and they are defined as “low”, “medium”, “high”.

For each linguistic term, there is a triangular membership function and also a trapezoidal membership function, as explained in detail in the below.

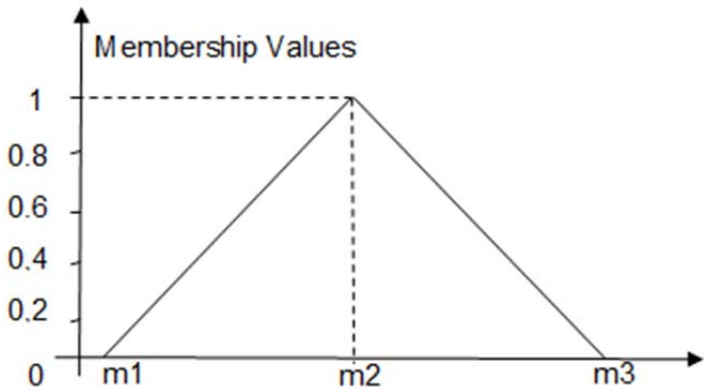

Figure 3.13. Triangular Membership Function

An example of triangular membership function can be shown in Figure 3.13. Membership values are obtained with triangular membership function (Yuan, Shaw, 1995; Rokach, Maimon, 2008; Mitra, Konwar, Pal, 2002) as follows: Let x be the sample value of the attribute A in the dataset S, $M=\{m_1,\ldots,m_k\}$ denotes the set of k

centers of the membership functions. Three centers are used to define triangular membership function, so value of *k* is three.

The first linguistic term *v*1 is "low" and it is computed by the following membership function:

$$\mu_{v1}(x) = \begin{cases} 1 & x \leq m_1 \\ \dfrac{m_2 - x}{m_2 - m_1} & m_1 < x < m_2 \\ 0 & x \geq m_2 \end{cases} \tag{3.30}$$

The second linguistic term *v*2 is "medium" and it is computed by the following membership function:

$$\mu_{v2}(x) = \begin{cases} 0 & x \leq m_1 \\ \dfrac{x - m_1}{m_2 - m_1} & m_1 < x \leq m_2 \\ \dfrac{m_3 - x}{m_3 - m_2} & m_2 < x < m_3 \\ 0 & x \geq m_3 \end{cases} \tag{3.31}$$

Finally the last linguistic term *v*3 is "high" and it is computed by the following membership function:

$$\mu_{v3}(x) = \begin{cases} 0 & x \leq m_2 \\ \dfrac{x - m_2}{m_3 - m_2} & m_2 < x < m_3 \\ 1 & x \geq m_3 \end{cases} \tag{3.32}$$

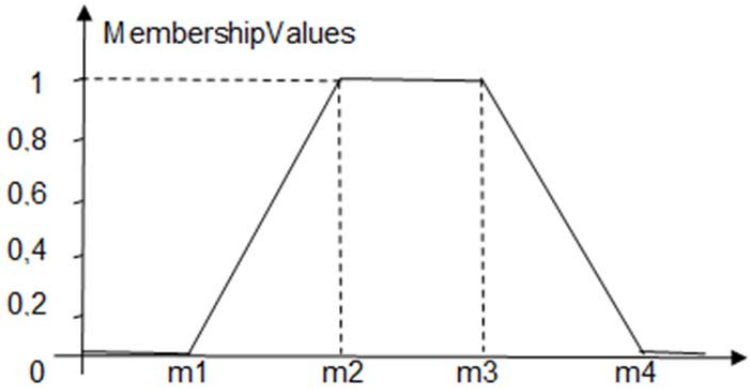

Figure 3.14. Trapezoidal Membership Function

An example of trapezoidal membership function is shown in Figure 3.14. Membership values are obtained with trapezoidal membership function (Au, Chan, Wong, 2006) as follows. Let x be the sample value of the attribute A in the dataset S, $M=\{m_1,\ldots,m_k\}$ denotes the set of k centers of the membership functions. Four centers are used to define trapezoidal membership function, so value of k is four.

For the first linguistic term $v1$ which is "low", the following membership function is used:

$$\mu_{v1}(x) = \begin{cases} 1 & x \leq m_1 \\ \dfrac{m_2 - x}{m_2 - m_1} & m_1 < x < m_2 \\ 0 & x \geq m_2 \end{cases} \tag{3.33}$$

For the second linguistic term $v2$ which is "medium", the following membership function is used:

$$\mu_{v2}(x) = \begin{cases} 0 & x \leq m_1 \\ \dfrac{x - m_1}{m_2 - m_1} & m_1 < x < m_2 \\ 1 & m_2 \leq x \leq m_3 \\ \dfrac{m_4 - x}{m_4 - m_3} & m_3 < x < m_4 \\ 0 & x \geq m_4 \end{cases} \tag{3.34}$$

For the last linguistic term *v*3 which is "high", the following membership function is used:

$$\mu_{v3}(x) = \begin{cases} 0 & x \leq m_3 \\ \dfrac{x - m_3}{m_4 - m_3} & m_3 < x < m_4 \\ 1 & x \geq m_4 \end{cases} \tag{3.35}$$

Center values are required in order to calculate membership degrees of the numerical data. Three centers are used for triangular membership function, and four centers are used for trapezoidal membership function. In this case, in order to determine centers, following formula is used (Yuan, Shaw, 1995):

$$m_i = \min\{x, x \in X\} + (\max\{x, x \in X\} - \min\{x, x \in X\}) * \frac{i-1}{k-1} \tag{3.36}$$

where $M = \{m_i, i = 1,\ldots,k\}$ denotes the centers of the membership functions and k is the number of centers. $\min\{x, x \in X\}$ represents the minimum value of the attribute and $\max\{x, x \in X\}$ represents the maximum value of the attribute.

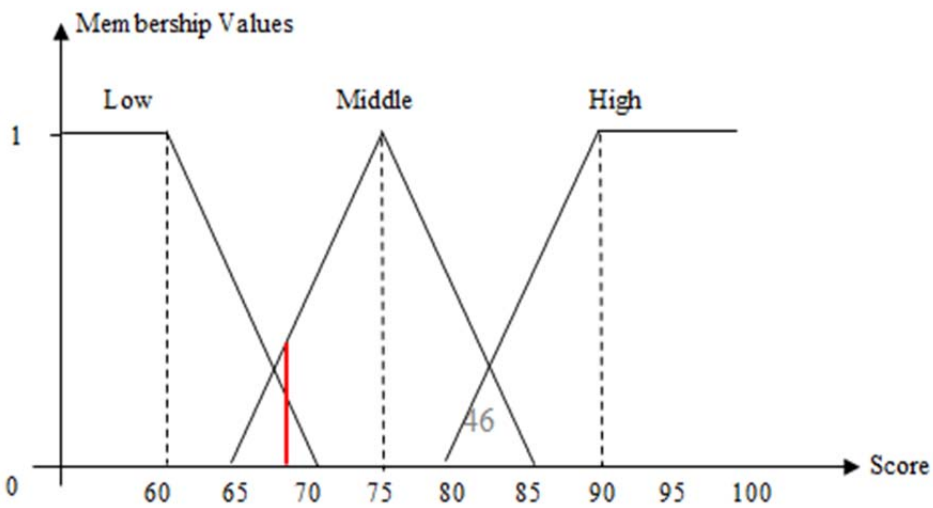

Figure 3.15. Membership Function for Examinee's Score (Wang, Tien-chin, Lee, and Hsien-da, 2006)

Figure 3.15. illustrates triangular membership function for examinee's score. If score is 68, it belongs to "low" and "middle" domain with different degrees as it appears in Figure 3.15 with red line.

In this thesis, two methods are used about membership values:

- Firstly, if an element is a member of more than one domain, linguistic term having the maximum membership value is used. Table 3.8 (a) and (b) show the fuzzy form of car type dataset which is presented in Table 3.7, using linguistic term that has maximum membership value with triangular and trapezoidal membership functions.
- Secondly, all linguistic terms that have greater than zero membership for an element are used. Table 3.9 and Table 3.10 show the fuzzy form of car type dataset with all linguistic terms.

Table 3.7. A Small Car Type Dataset (Lee, Sun, Yang, 2003)

Tuple Id	**Height**	**Weight**	**Length**	**Class**
1	2,2	2,8	4	P
2	3,2	4,4	16	N
3	3,8	10	6	P
4	3	18	7	N
5	3	25	6	N
6	3,8	5,1	6	P
7	3	17	14	N
8	3,4	19	13	N

For example, Height attribute in Table 3.7, has 3 centers for triangular membership function and they are computed with equation 3.30, 3.31, and 3.32 as 2.2, 3.0, 3.8. Membership values are computed for each samples of the Height attribute. Membership values for three linguistic terms when numeric value is equal to 3.2 for the Height attribute are shown below:

- $\mu_{low}(3.2) = 0$
- $\mu_{medium}(3.2) = 0.75$
- $\mu_{high}(3.2) = 0.25$

According to the membership values listed above, the numeric value 3.2 for the Height attribute belongs to "medium" and "high" domains with different degrees and sample has maximum membership value for "medium" domain. If all linguistic terms are used for fuzzification, "medium" and "high" are used and they can be shown in Table 3.9. Otherwise,"medium" is used and it can be shown in Table 3.8.(a).

Table 3.8. Fuzzified Small Car Type Dataset Using Linguistic Terms Having Maximum Membership Value a) with Triangular Membership Function b) with Trapezoidal Membership Function

Tuple Id	**Height**	**Weight**	**Lenght**	**Class**
1	Low	Low	Low	P
2	Medium	Low	High	N
3	High	Medium	Low	P
4	Medium	Medium	Low	N
5	Medium	High	Low	N
6	High	Low	Low	P
7	Medium	Medium	High	N
8	Medium	Medium	Medium	N

Tuple Id	**Height**	**Weight**	**Lenght**	**Class**
1	Low	Low	Low	P
2	Low	Low	High	N
3	High	Low	Low	P
4	Low	High	Low	N
5	Low	High	Low	N
6	High	Low	Low	P
7	Low	Low	High	N
8	High	High	High	N

For triangular membership function, if all linguistic terms of the small car type dataset are used, fuzzified small car type dataset is shown in Table 3.9.

Table 3.9. Fuzzified Small Car Type Dataset Using All Linguistic Terms According to Triangular Membership Function

Tuple Id	Height	Weight	Length	Class
1	low	low	low	P
2	medium,high	low,medium	high	N
3	high	low,medium	low,medium	P
4	medium	medium,high	low,medium	N
5	medium	high	low,medium	N
6	high	low,medium	low,medium	P
7	medium	medium,high	medium,high	N
8	medium,high	medium,high	medium,high	N

For trapezoidal membership function, if all linguistic terms of the small car type dataset are used, fuzzified small car type dataset is shown in Table 3.10.

Table 3.10. Fuzzified Small Car Type Dataset Using All Linguistic Terms According to Trapezoidal Membership Function

Tuple Id	Height	Weight	Length	Class
1	low	low	low	P
2	medium	low,medium	high	N
3	high	low,medium	low,medium	P
4	medium	medium,high	low,medium	N
5	medium	high	low,medium	N
6	high	low,medium	low,medium	P
7	medium	medium	medium,high	N
8	medium,high	medium,high	medium,high	N

3.2.2. Inducing Decision Tree with Non-Fuzzy Data

In this method, decision tree is induced by using the ID3 algorithm. All datasets are used without fuzzification and this method was employed to make comparison with other methods that are explained in Section 3.2.3.

3.2.2.1. ID3 with Best Split Method

Best split method is used with basic ID3 algorithm to induce the decision tree and this method is applied to the numerical datasets. There is no fuzzification process

for used datasets. Also it includes information gain, gain ratio and gini index to split the dataset.

Figure 3.6 shows the graphical user interface of the application program that were developed in this thesis to train the basic non-fuzzy decision tree using the ID3 algorithm. This application contains splitting criteria to find best split point named as information gain, gain ratio, gini index for decision tree induction, and it has delimiter options for the dataset. First step of the decision tree induction algorithm is to upload the dataset. Then a splitting criterion is choosen and finally algorithm is started with "Start" button as shown in Figure 3.16.

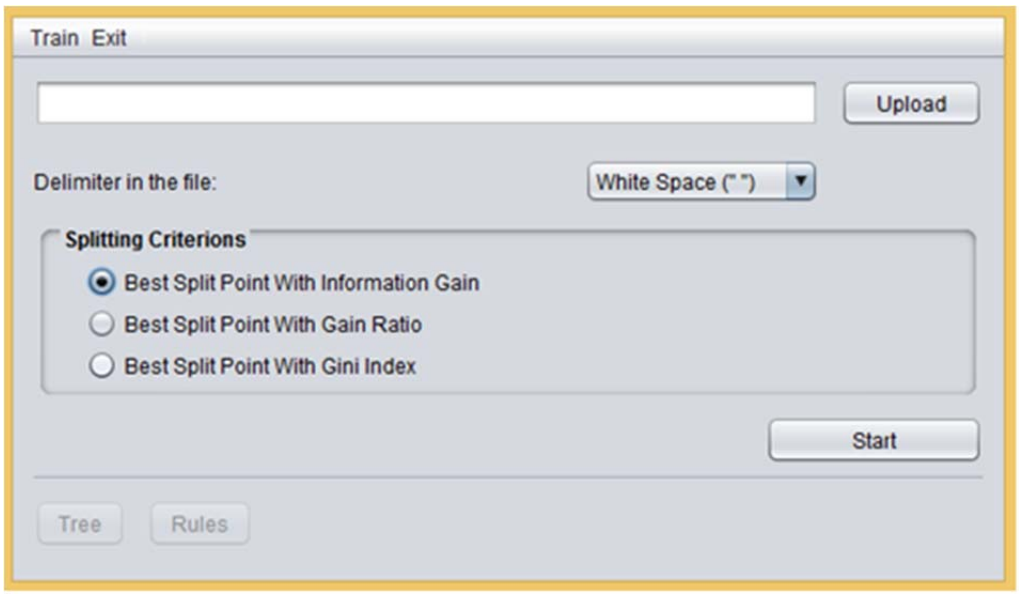

Figure 3.16. Training Part of the Basic Decision Tree Induction Using "ID3 with Best Split Method"

We used the algorithm to generate the basic decision tree:

1. Create a new decision tree with a single root node.
2. For all attributes, compute one of the information gain, gain ratio or gini index to select the best attribute and the best splitting point.
    a. For the attribute, calculate one of the three measures for all possible split points.

b. Select the split point having the best measure value as best split point for the attribute.
c. For all attribute in the dataset do the same steps to obtain the best split points to split the dataset.
d. Select the attribute having the maximum information gain, gain ratio, or gini index using its best split point.

3. Use the selected attribute with its best splitting point to build the decision tree.
4. For each branches of the selected attribute, repeat the steps from 2 to 3 recursively until termination condition is met.

An example decision tree for the small car type dataset in Table 3.7that is built with best split point method is shown in Figure 3.17.

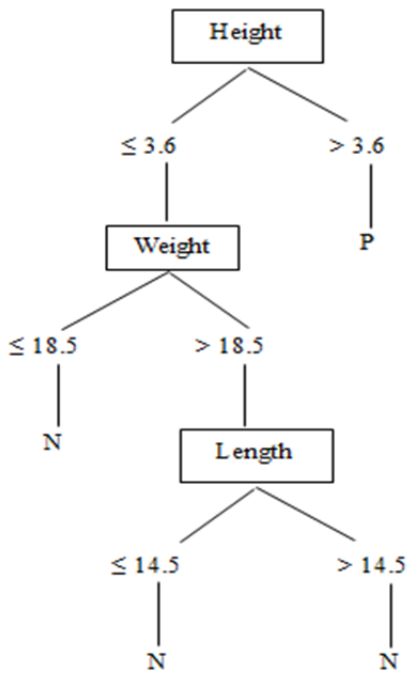

Figure 3.17. Decision Tree for Small Car Type Dataset Using Best Split Method with Information Gain Measure

3.2.3. Inducing Decision Trees with Fuzzy Data Methods

Two different methods were applied to the datasets that were from UCI Machine Learning Repository shown in Table 3.5. At first, basic ID3 algorithm were

applied to fuzzified datasets. In this method, basic splitting criteria, information gain, gain ratio, and gini index were used. After that fuzzy format of the information gain, gain ratio, and gini index were applied to the basic ID3 algorithm. Then, the results were compared. These methods are presented in the next subsections.

3.2.3.1. ID3 with Fuzzy Data and Basic Splitting Criterion

In this method, basic ID3 algorithm is used to construct the decision tree. First of all the dataset is fuzzified before it is used in the algorithm. We have two options of the membership functions that are triangular and trapezoidal and three options for attribute selection that are information gain, gain ratio and gini index as mentioned in Section 3. Figure 3.18 shows the graphical user interface of the application developed for this method. This application contains splitting criteria as information gain, gain ratio, gini index; and membership functions as triangular and trapezoidal and also it has delimiter options for the dataset. First step of decision tree induction algorithm is the uploading of the dataset. Then a splitting criterion and a membership function are choosen and finally algorithm is started with “Start” button as shown in Figure 3.18.

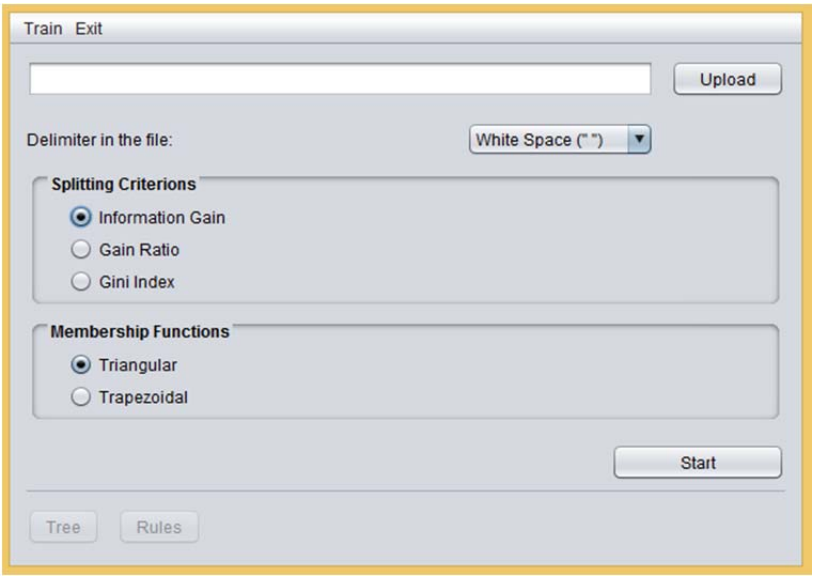

Figure 3.18. Training Part of the Decision Tree Induction Using “ID3 with Fuzzy Data and Basic Splitting Criterion”

The steps of the algorithm used in this application is described below:

1. Fuzzify the training data.
2. Create a new fuzzy decision tree with a single root node.
3. For all attributes, compute one of the splitting criteria(i.e., information gain, gain ratio, gini index) that is selected by the user.
    - If all linguistic terms obtained from fuzzification of the data are used, all terms are used in the calculations of the splitting criterion.
4. Select the attribute having the maximum attribute selection criterion.
5. Divide the fuzzy data into fuzzy subsets with the selected attribute, and for each subset generate a new node and a branch.
6. If all the data in a branch belongs to the same class, generate a leaf node and assign this class label to it and then link this leaf node to the branch.
7. If the data set in a branch needs to split (e.g., data samples are in different classes) but there is no remaining attribute, generate a leaf node and assign the majority class label to it and then link this leaf node to the branch.
8. For each branches of the selected attribute, repeat steps 3 to 7 recursively.

A decision tree with fuzzified data shown in Table 3.8.(a) is presented in Figure 3.19. Information gain is used to select attributes as splitting criterion to build decision tree and triangular membership function is used to fuzzify the dataset.

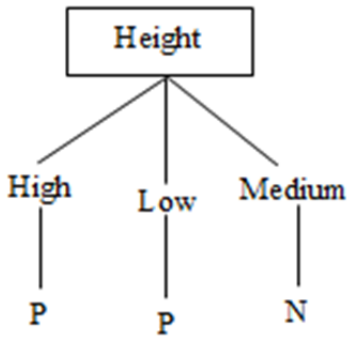

Figure 3.19. FDT using Information Gain for Dataset in Table 3.8.(a)

If all linguistic terms for a sample in the attributes are used to form the decision tree, the resulting tree is presented in Figure 3.20.

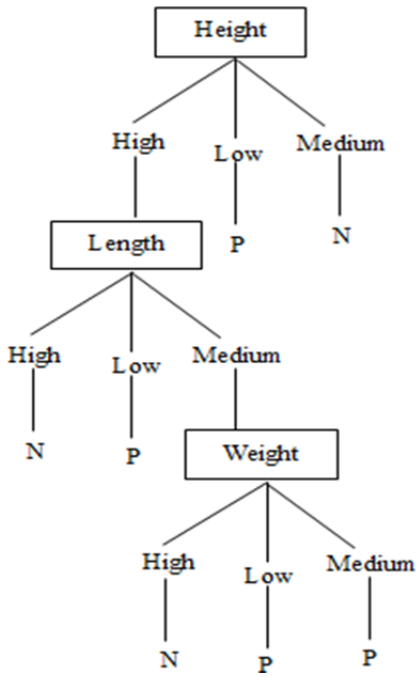

Figure 3.20. FDT using Information Gain for Dataset in Table 3.9 i.e., with All Linguistic Terms of Samples in the Dataset

3.2.3.2. ID3 With Fuzzy Data and Fuzzy Form of Splitting Criterion

Basic ID3 algorithm with fuzzy attribute selection criterion is used in this method. We have three options to determine the attributes, that are fuzzy information gain, fuzzy gini index, and fuzzy gain ratio to split the fuzzy dataset (Abu-halaweh, Harrison, 2010; Chen, Shie, 2009). Membership values of attributes obtained by fuzzification are used to compute the fuzzy attribute selection criteria. To compute the fuzzy splitting criteria, we used two forms of data fuzzification. In the first method, we chose the linguistic term having the maximum membership value. In the second method, we took all linguistic terms having non-zero membership values. These membership values were used in fuzzy splitting criteria computations. Figure 3.21 shows the graphical user interface of fuzzy decision tree induction with fuzzy splitting criteria namely, fuzzy information gain, fuzzy gain ratio, and fuzzy gini

index. Membership functions are triangular and trapezoidal, and the UI allows to choose delimiter options for the dataset. First step of the decision tree induction algorithm is uploading of the dataset. Then a splitting criterion and a membership function are chosen and finally algorithm is started with "Start" button as shown in Figure 3.21.

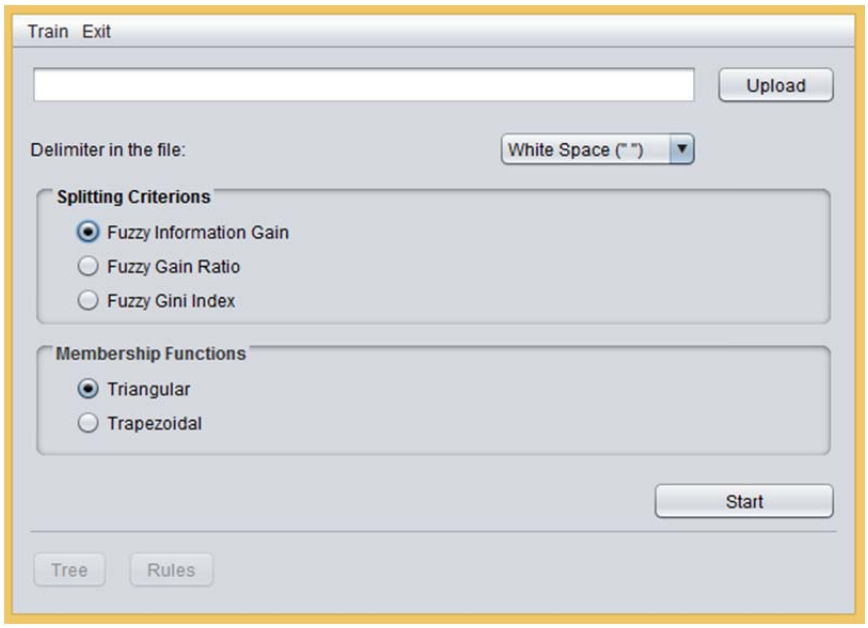

Figure 3.21. Training Part of the Decision Tree Induction Using "ID3 with Fuzzy Data and Fuzzy Form of Splitting Criterion"

The steps of the algorithm is as follows:

1. Fuzzify the training data.
2. Create a new fuzzy decision tree with a single root node.
3. For all attributes, compute one of the fuzzy splitting criteria that is selected by user. Membership values of samples in the attribute are used in the calculations.
    - If all linguistic terms obtained from the fuzzification of the dataset are used, membership values of all terms are used to compute fuzzy splitting criteria.
4. Select the attribute having maximum fuzzy splitting criterion.

5. Partition the fuzzy data into fuzzy subsets using the selected attribute, and for each subset generate a new node and a branch.
6. If all data in a branch belongs to the same class, assign this class name to the branch as a label.
7. If the dataset in a branch needs to split but there is no remaining attribute, assign the majority class label to the branch.
8. For each branch of the selected attribute, repeat steps 3 to 7 recursively.

A fuzzy decision tree for fuzzified data in Table 3.8.(a) is presented in Figure 3.22. Fuzzy information gain is used to select attributes and triangular membership function is used to fuzzify the dataset. Also membership values of the samples in the dataset are used in the computations of the fuzzy splitting criteria.

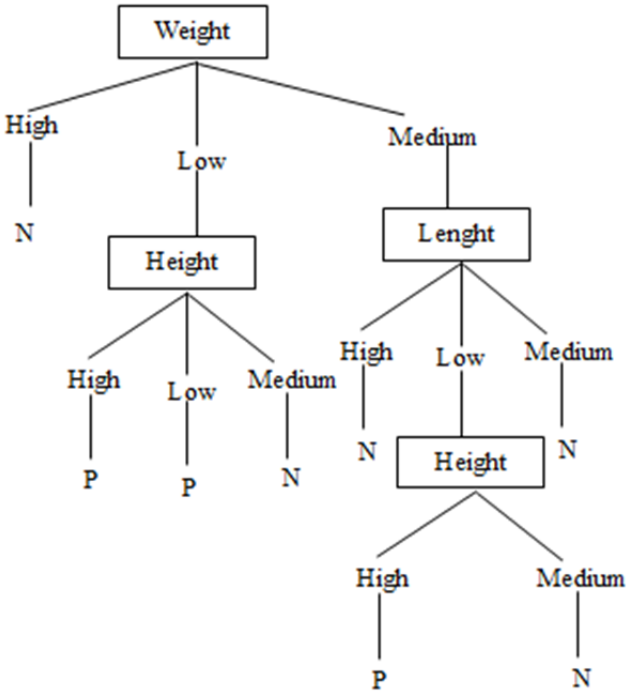

Figure 3.22. FDT using Fuzzy Information Gain for Dataset in Table 3.8. (a)

If all linguistic terms of samples in the fuzzified dataset are used, the fuzzy decision tree with fuzzy splitting criteria is presented in Figure 3.23 for the dataset in Table 3.9.

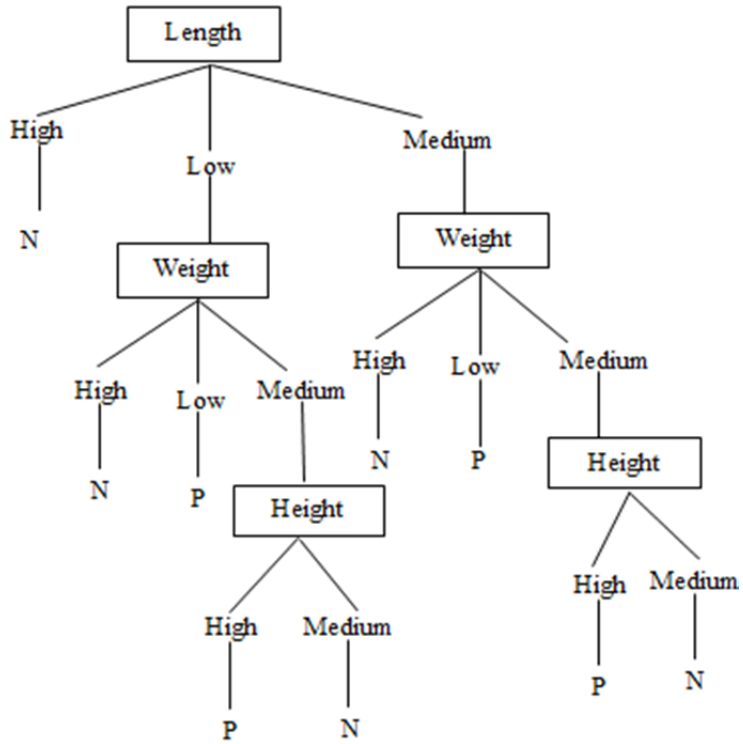

Figure 3.23. FDT using Fuzzy Information Gain for Dataset in Table 3.9 i.e., with All Linguistic Terms of Samples in the Dataset

3.2.4. Extracting Classification Rules From Fuzzy Decision Tree

Classification rules are generated from the decision tree and these rules are used to assign class labels to new samples in the test dataset to measure the performance of the decision tree. Each path of the branches from root to any leaf is converted into a rule to classify the test dataset. Starting from the root node, each node is visited until a leaf node is reached. All the paths from root to leaves form classification rules. The rules extracted from the decision tree that is learned from the Weather Dataset in Figure 3.1 are listed below. As an example, according to the tree,“Outlook=Overcast”then the class is “Yes”.

1) IF Outlook = Overcast THEN Play = Yes
2) IF Outlook = Rainy AND Windy = False THEN Play = Yes
3) IF Outlook = Rainy AND Windy = True THEN Play = No
4) IF Outlook = Sunny AND Humidity = High THEN Play = No

5) IF Outlook = Sunny AND Humidity = Low THEN Play = Yes
6) IF Outlook = Sunny AND Humidity = Medium THEN Play = No

3.2.5. Applying Fuzzy Rules for Classification

After the training process, the new samples can be classified with the rules obtained during the training phase. In order to classify the test dataset, the following steps are applied:

1. If one of the "ID3 with Fuzzy Data and Basic Attribute Selection Criterion" or "ID3 with Fuzzy Data and Fuzzy Form of Attribute Selection Criterion" is selected, the test dataset is fuzzified before it is used.
2. If "ID3 with Best Split Method" is selected, the test dataset is not fuzzified before it is used.
3. For each test tuple, all rules are investigated and the rules which are satisfied by the test data are determined. Then, accuracy and coverage measures are compared for each chosen rule.

If a test tuple is classified by more than one rules, one of the rules is selected for this test data. The process of selecting one of the rules has four options in our implementation. These options are explained as below:

1. If class distributions of the rules which classify the test data are equal, class of the rule having the maximum coverage×accuracy value is selected for the test tuple. If class distributions of the rules are not equal, the majority class label is selected. This option is named as "Test 1".
2. Class of the rule having maximum accuracy value is selected for test data. This option is named as "Test 2".
3. Class of the rule having maximum coverage value is selected for test data. This option is named as "Test 3".

4. Class of the rule having maximum coverage×accuracy value is selected for test data. This option is named as “Test 4”. The difference between “Test 1” and “Test 4” is that class distributions of the rules are only important for “Test 1” option.

Figure 3.24 illustrates the graphical user interface of the test part of the decision tree induction algorithms. This application contains membership functions as triangular and trapezoidal functions and delimiter options for the dataset. First step of testing phase is the upload of the test dataset. Then a membership function is choosen and finally algorithm is started with “Start” button as shown in Figure 3.24.

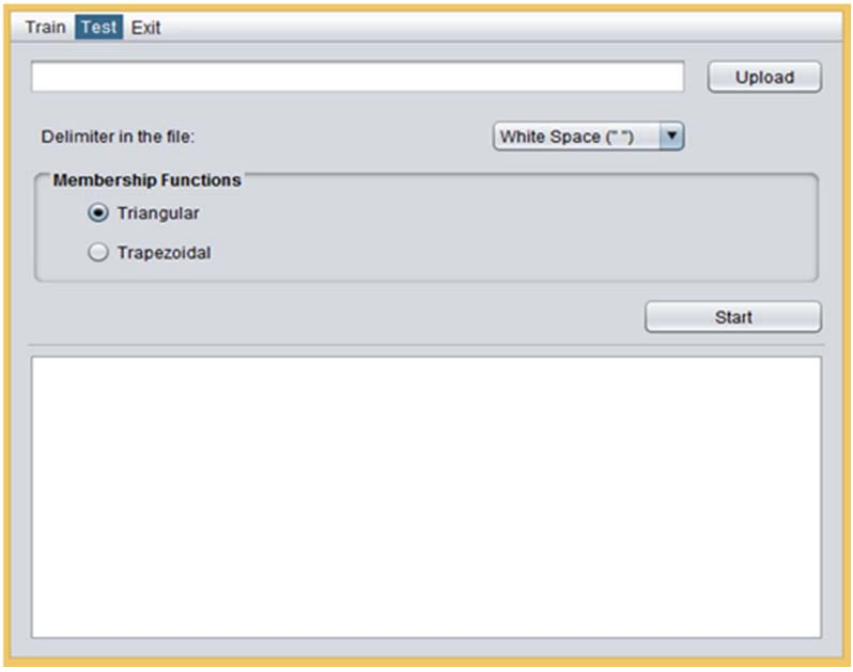

Figure 3.24. Test part of the Induction Algorithm

As an example, let’s assume that, the test dataset presented in Table 3.11 is used to test the learned decision tree. The test dataset has 4 samples in Numerical Weather Dataset.

Table 3.11. Test Dataset for the Numerical Weather Dataset

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
Sunny	85	70	False	No
Sunny	80	90	True	No
Rainy	70	96	False	Yes
Overcast	70	78	True	Yes

Since triangular membership function is used to build the decision tree, the test samples are fuzzified with the triangular membership function. Fuzzified test dataset is shown in Table 3.12.

Table 3.12. Fuzzified Test Dataset Presented in Table 3.11 Using Triangular Membership Function

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
Sunny	High	Low	False	No
Sunny	Medium	High	True	No
Rainy	Low	High	False	Yes
Overcast	Low	Medium	True	Yes

Test results of the fuzzified test dataset are shown in Figure 3.25. According to the figure, the model built in the training part is successfull at 75% rate, in other words it can classify correctly 3 samples out of every 4 test samples. Also it contains tp rate, fp rate, precision, recall, f-measure and confusion matrix detailed in the measuring the performance of the classification model section.

```
Time taken to build model: 0,219 seconds.
Time taken to test model on training data: 0,14 seconds.

****Summary****

Correctly Classified Instance        3        75%
Incorrectly Classified Instances     1        25%
Total Number of Instances            4

****Detailed Accuracy By Class****

                    TP RATE   FP RATE   Precision   Recall   F-Measure
Class
                    1         0,5       0,5         1        0,667
No
                    0,667     0         1           0,667    0,8
Yes
 Weighted Ave.      0,833     0,25      0,75        0,833    0,733

****Confusion Matrix****

No      Yes     <-- classified as
 1       0      | No
 1       2      | Yes
```

Figure 3.25. Test results of the test weather dataset

4. RESULTS AND DISCUSSION

Three methods which are used to build decision trees in this thesis were applied to 18 datasets selected from UCI Machine Learning Repository and compared to each other. Datasets used are given in Table 3.5. These datasets were partitioned as test and train sets before used in the experiments and the details about the partitioned datasets arepresented in Table 3.6.

All methods were implemented with Java programming language under the Netbeans environment. The proposed method were prepared and tested under Microsoft Windows 7 operating system. The hardware used in the experiments had 4 GB of RAM and Intel Core 2 Duo 2.53 GHz processor.

The first method which was implemented in this thesis is the basic ID3 algorithm. It was used as a baseline method to compare with the fuzzified versions. Best split point method was used to discretize numerical attributes in the basic ID3 method.

Other two methods were used to test the effect of the fuzzification of the dataset on the decision tree induction. For these methods, attributes in the datasets were fuzzified before using in the implementation. ID3 algorithm was used to build decision tree with fuzzified datasets.

All methods contain two main parts: training and test. A decision tree as a model was built in the training part. In the test part new samples were classified with the learned model which was built in the training part. So test part shows the success of the model on the test datasets.

Experimental results which are obtained with these methods by using 18 datasets are presented in the following subsections.

4.1. Effect of Data Fuzzification

Fuzzified and discrete data were used to indicate the effect of the data fuzzification on the classification. To make a comparison between “ID3 with fuzzy data and basic splitting criteria” and “ID3 with best split point”, same datasets and

basic splitting criteria were used. Datasets had numerical values. These samples were fuzzified before using in "ID3 with fuzzy data and basic splitting criteria" method. This means that numerical values were converted to linguistic terms by fuzzification process which uses triangular and trapezoidal membership functions. On the other hand, numerical values were discretized to numerical intervals in "ID3 with best split point" method. Table 4.1 shows accuracy of classification with the two methods. In this experiment, triangular and trapezoidal membership functions were used to fuzzify numerical data. The best accuracy values for each datasets are written in the boldface in Table 4.1 to make comparison between methods. In Table 4.1, IG, GR, and GI stand for Information Gain, Gain Ratio, and Gini Index, respectively.

Table 4.1. Accuracy of the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions

Datasets	**ID3 with best split point**			**ID3 with triangular fuzzified data**			**ID3 with trapezoidal fuzzified data**		
	IG	**GR**	**GI**	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**
HeartStatlog	58.82	58.52	64.71	66.18	**72.06**	42.65	**72.06**	70.59	66.18
MammographicM.	50.00	51.67	50.00	80.00	80.00	82.08	**82.50**	**82.50**	**82.50**
BreastCancer	35.09	35.09	35.09	**94.74**	93.57	90.64	93.57	94.15	**94.74**
Diabetes	62.50	55.73	63.02	61.46	61.98	60.94	**67.71**	**67.71**	66.67
Hepatitis	**84.62**	58.97	64.10	71.80	64.10	58.97	76.92	71.80	58.97
SpectHeart	68.66	71.64	71.64	77.61	**79.11**	**79.11**	77.61	**79.11**	**79.11**
Yeast	31.81	31.53	34.23	**43.67**	**43.67**	42.86	36.39	36.66	36.93
VertebralCol 2C	39.02	39.02	37.81	57.32	54.88	54.88	68.29	69.51	**70.73**
VertebralCol 3C	56.41	**62.82**	55.13	56.41	58.97	57.69	41.03	41.03	51.28
Ecoli	27.38	10.71	23.81	72.62	67.86	64.29	71.43	71.43	**73.81**
BalanceScale	74.36	66.67	**78.21**	73.72	73.72	73.72	67.95	67.95	67.95
Thyroid	81.48	24.07	77.78	90.59	**90.74**	88.89	88.89	88.89	88.89
LD Bupa	67.44	56.98	**72.09**	47.67	46.51	43.02	58.14	58.14	58.14
Iris	**94.74**	**94.74**	**94.74**	92.11	92.11	89.47	65.79	65.79	65.79
Glass	57.41	**62.96**	53.70	38.89	24.07	22.22	27.78	25.93	20.37
Monk1	48.20	48.20	48.20	**98.56**	**98.56**	95.68	66.91	66.91	66.91
Monk2	68.67	68.67	68.67	78.00	77.33	**79.33**	60.00	60.00	60.00
Monk3	81.30	81.30	81.30	93.53	93.53	92.09	**96.40**	**96.40**	**96.40**

According to Table 4.1, applying ID3 decision tree algorithm to fuzzified data is more succesful than classical decision tree but it has disadvantage about the

number of rules which are obtained from the tree and used to classify new data. Table 4.2 shows the number of rules learned by the "ID3 with best split point" and "ID3 with fuzzified data and basic splitting criteria" methods. The less number of rules for each datasets are written in boldface in Table 4.2 to make comparison between the methods. According to the results shown in the Table 4.2, number of rules obtained from fuzzy decision trees are more than number of rules obtained from classical decision trees.

Table 4.2. F-measure of the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions

Datasets	**ID3 with best split point**			**ID3 with triangular fuzzified data**			**ID3 with trapezoidal fuzzified data**		
	IG	**GR**	**GI**	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**
HeartStatlog	59.5	59.1	65.2	66.2	**72.5**	42.6	72.2	70.3	65.6
MammographicM.	39.3	42.7	39.3	79.8	79.8	81.9	**82.5**	**82.5**	**82.5**
BreastCancer	18.2	18.2	18.2	**94.7**	93.6	90.7	93.5	94.1	**94.7**
Diabetes	60.9	54.8	60.0	59.4	60.1	58.7	**66.7**	**66.7**	65.6
Hepatitis	**83.2**	61.5	64.1	70.0	62.1	56.6	76.5	71.0	59.0
SpectHeart	67.6	71.2	69.8	77.0	**78.8**	78.0	77.0	**78.8**	78.0
Yeast	15.7	15.3	25.6	42.2	**42.3**	41.2	32.4	32.6	32.8
VertebralCol 2C	38.2	38.2	36.5	59.3	56.8	56.8	62.4	63.6	**64.8**
VertebralCol 3C	54.2	**58.3**	54.1	47.8	49.7	48.8	29.8	29.8	37.7
Ecoli	26.8	10.3	22.3	69.3	63.5	60.6	71.1	71.1	**73.9**
BalanceScale	70.3	61.5	**73.9**	71.1	71.1	71.1	64.9	64.9	64.9
Thyroid	74.7	17.4	76.6	**92.7**	90.8	89.0	87.7	87.7	87.7
LD Bupa	67.5	52.2	**71.8**	40.3	38.9	40.3	48.9	48.9	48.9
Iris	**94.8**	**94.8**	**94.8**	92.0	92.0	89.4	66.1	66.1	66.1
Glass	54.4	**58.8**	49.1	36.6	20.4	19.1	26.3	23.1	18.5
Monk1	31.4	31.4	31.4	**98.6**	**98.6**	95.7	66.9	66.9	66.9
Monk2	55.9	55.9	55.9	78.2	77.3	**79.3**	58.2	58.2	58.2
Monk3	80.9	80.9	80.9	93.5	93.5	92.1	**96.4**	**96.4**	**96.4**

Table 4.2 shows f-measure of classification with the two methods that are classical and fuzzy decision trees. In this experiment, triangular and trapezoidal membership functions were used to fuzzify numerical data. The best f-measure values for each datasets are written in the boldface in Table 4.2 to make comparison between methods. According to the results presented in Table 4.2, f-measure and

accuracy values of the decision trees have almost same values. So, fuzzy decison tree is more succesful than classical decision tree in terms of f-measure.

Table 4.3. Number of Rules for the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions

Datasets	**ID3 with best split point**			**ID3 with triangular fuzzified data**			**ID3 with trapezoidal fuzzified data**		
	IG	**GR**	**GI**	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**
HeartStatlog	**37**	40	38	58	60	92	60	60	70
MammographicM.	27	**19**	27	50	50	58	21	20	20
BreastCancer	38	30	28	56	56	78	23	**22**	33
Diabetes	54	**14**	66	107	108	114	30	29	32
Hepatitis	24	**20**	22	24	23	47	30	29	37
SpectHeart	47	48	**41**	47	48	75	47	48	75
Yeast	11	**6**	16	87	84	89	20	18	19
VertebralCol 2C	20	14	29	25	24	27	8	7	9
VertebralCol 3C	33	16	29	28	27	25	10	**9**	**9**
Ecoli	25	**18**	24	36	39	39	22	21	21
BalanceScale	15	16	16	63	63	69	**14**	**14**	16
Thyroid	15	**6**	13	12	14	15	7	**6**	**6**
LD Bupa	30	**8**	39	39	40	40	9	9	9
Iris	7	7	7	**6**	**6**	12	**7**	**7**	8
Glass	36	25	43	33	34	34	**14**	**14**	15
Monk1	**12**	**12**	**12**	63	63	125	36	36	36
Monk2	**42**	**42**	**42**	193	184	188	59	59	60
Monk3	16	**15**	17	37	35	61	17	16	19

Best split point method finds the best split point of an attribute *A* by using all samples and all possible split points in attribute *A*. So, this method includes too many mathematical compuations and it takes long time to discretize data and then build the model in the training part. Training time in seconds for both methods are shown in Table 4.4. According to the results given in Table 4.4, "ID3 with best split point" method takes longer time than "ID3 with fuzzy data and basic splitting criteria" method. But "ID3 with best split point" method has less number of rules as shown in Table 4.3 and because of this, test time in seconds for the "ID3 with best split point" method are less than the other method. These results are shown in Table 4.5.

According to the results presented in Table 4.1, 4.2, 4.3, 4.4, and 4.5 when numerical data is fuzzified by both of the triangular and trapezoidal membership functions, both accuracy, and training time performances are better with respect to discretization by best split point method. Only the number of rules that are learned when the best split point discretization method is used becomes less than that of discretization by fuzzification. This only reduces test time slowly.

Performance evaluation of information gain, gain ratio, and gini index methods are presented in Section 4.4.

Table 4.4. Training Time in Seconds for the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions

	ID3 with best split point			**ID3 with triangular fuzzified data**			**ID3 with trapezoidal fuzzified data**		
Datasets	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**
HeartStatlog	186.33	192.34	150.95	2.10	1.97	2.52	2.05	1.34	**1.01**
MamM.	94.24	72.22	1056.16	2.57	2.54	2.29	2.97	2.03	**1.99**
BreastCancer	74.98	72.09	67.77	2.71	2.75	3.12	2.52	**2.45**	2.54
Diabetes	1842.4	1309.8	2049.5	4.14	4.16	4.19	3.82	3.03	**2.99**
Hepatitis	79.35	115.97	99.39	**0.87**	0.97	1.42	1.24	1.11	1.22
SpectHeart	56.93	54.82	65.60	**3.77**	3.86	4.71	4.56	3.83	4.71
Yeast	4430.7	3949.9	2809.5	11.12	10.00	10.07	12.56	8.73	**8.48**
VertebralCol 2C	612.96	660.59	720	0.69	0.65	0.61	0.78	0.59	**0.56**
VertebralCol 3C	415.74	397.25	406.86	0.7	0.66	0.64	0.97	0.59	**0.56**
Ecoli	155.37	153.89	141.80	0.85	0.86	**0.81**	1.12	0.80	0.74
BalanceScale	12.21	11.47	12.27	1.03	**1.02**	1.04	1.17	0.80	0.82
Thyroid	207.91	59.38	157.82	0.28	0.29	0.26	0.52	**0.25**	0.25
LD Bupa	742.89	189.94	309.34	0.75	0.66	0.65	0.79	**0.50**	0.90
Iris	17.86	19.75	23.26	0.14	**0.14**	0.15	0.27	0.16	0.21
Glass	263.72	340.13	334.79	0.86	0.77	0.62	0.82	0.54	**0.53**
Monk1	16.20	6.05	5.73	**1.28**	1.29	1.58	1.72	**1.28**	**1.28**
Monk2	19.68	24.29	16.75	2.14	2.29	1.94	2.19	1.72	**1.67**
Monk3	10.48	6.61	7.44	1.57	1.08	1.33	1.46	1.06	**1.04**

Table 4.5. Test Time in Seconds for the "ID3 with Best Split Point" and "ID3 with Fuzzified Data and Basic Splitting Criteria" Classifications Methods Using Triangular and Trapezoidal Membership Functions

	ID3 with best split point			**ID3 with triangular fuzzifieddata**			**ID3 with trapezoidal fuzzifieddata**		
Datasets	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**	**IG**	**GR**	**GI**
HeartStatlog	0.56	**0.09**	**0.09**	0.27	0.24	0.26	0.24	0.48	0.27
MammographicM.	0.32	0.64	0.93	0.35	0.38	0.33	0.46	**0.29**	0.34
BreastCancer	0.16	0.29	0.07	0.74	0.35	0.39	0.51	0.32	0.34
Diabetes	0.85	0.71	0.71	0.51	0.5	0.53	0.50	**0.40**	0.41
Hepatitis	0.13	0.06	**0.06**	0.23	0.21	0.23	0.31	0.23	0.30
SpectHeart	0.3	0.12	**0.04**	0.35	0.35	0.44	0.54	0.37	0.41
Yeast	2.74	3.7	**0.53**	1.13	1.17	1.20	1.23	0.96	0.95
VertebralCol 2C	2.55	2.36	6.14	0.19	0.18	**0.14**	0.26	0.16	0.15
VertebralCol 3C	0.17	0.41	**0.07**	0.14	0.14	0.14	0.23	0.15	0.20
Ecoli	0.15	0.41	**0.07**	0.14	0.23	0.15	0.24	0.16	0.17
BalanceScale	0.39	0.13	**0.08**	0.17	0.21	0.17	0.24	0.16	0.16
Thyroid	0.55	**0.07**	0.09	0.09	0.09	0.09	0.2	0.10	0.10
LD Bupa	0.64	0.43	0.54	0.14	0.14	**0.13**	0.22	0.13	0.14
Iris	0.12	0.10	0.09	**0.08**	0.08	0.08	0.16	0.11	0.11
Glass	0.12	0.59	**0.11**	0.15	0.18	0.15	0.24	0.19	0.14
Monk1	0.32	0.15	**0.12**	0.20	0.20	0.25	0.31	0.21	0.21
Monk2	0.18	0.22	**0.13**	0.34	0.38	0.34	0.35	0.28	0.24
Monk3	0.29	**0.04**	0.06	0.27	0.18	0.21	0.30	0.19	0.13

4.2. Effect of Using Fuzzy Decision Tree on Classification Performance

Performance of the fuzzy decision tree with fuzzified data and fuzzy splitting criteria is evaluated in this section. Numerical values in the datasets were fuzzified before train and test phases. In this method, fuzzy splitting criteria were used and they are named as fuzzy information gain, fuzzy gain ratio, and fuzzy gini index. Membership degrees of the numerical values were used to compute fuzzy splitting criteria which are explained in detail in Section 3.1.6.

In Table 4.6, accuracy of classification that are performed by the "ID3 with fuzzified data and fuzzy splitting criteria" method is presented. As membership function for fuzzification, we employed triangular and trapezoidal membership functions. According to the results in Table 4.6, "ID3 with fuzzified data and fuzzy splitting criteria" method is more successful than "ID3 with best split" method as

presented in Table 4.1. For "ID3 with fuzzified data and fuzzy splitting criteria" method, number of rules obtained from decision tree are more than "ID3 with best split" method given in Table 4.3 in the previous section but it takes less time in seconds than best split method for train and test phases. For "ID3 with fuzzified data and fuzzy splitting criteria" method, the number of rules learned are shown in Table 4.8, time required to train and test the method are given in Table 4.9, 4.10.

According to the results presented in Table 4.1, and 4.6, classification accuracy of ID3 with fuzzy splitting criteria does not much effect on the classification performance. For some datasets, higher performance is obtained. For example, mammographic masses dataset has 80.00% classification accuracy for gain ratio and triangular membership function. This classification accuracy rose to 81.67% for fuzzy gain ratio. Breast cancer dataset has 93.57% accuracy for information gain and trapezoidal membership function and it has 94.15% accuracy for fuzzy information gain and trapezoidal membership function. In the decision tree built with fuzzy splitting criteria, number of rules that are learned are greater and it takes longer time in seconds for train and test phases than basic splitting criteria.

Table 4.7 shows f-measure of classification that are performed by the "ID3 with fuzzified data and fuzzy splitting criteria" method. As membership function for fuzzification, we employed triangular and trapezoidal membership functions. The best f-measure values for each datasets are written in the boldface. According to the results presented in Table 4.7, f-measure and accuracy presented in Table 4.6 values of the decision trees have almost same values.

Table 4.6. Accuracy of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions

Datasets	**Triangular Membership**			**Trapezoidal Membership**		
	F-IG	**F-GR**	**F-GI**	**F-IG**	**F-GR**	**F-GI**
HeartStatlog	30.88	39.71	44.12	61.77	64.71	**66.18**
MammographicM	80.83	81.67	82.08	**82.50**	**82.50**	**82.50**
BreastCancer	90.06	85.97	89.47	94.15	**94.74**	**94.74**
Diabetes	61.98	60.42	60.42	**67.19**	66.67	66.67
Hepatitis	51.28	53.85	56.41	56.41	**66.67**	58.97
SpectHeart	71.64	**79.12**	79.11	71.64	79.11	79.11
Yeast	39.62	39.89	**42.86**	36.93	36.93	36.93
VertebralCol 2C	56.10	54.88	54.88	**70.73**	**70.73**	**70.73**
VertebralCol 3C	**57.69**	**57.69**	**57.69**	51.28	51.28	51.28
Ecoli	65.48	66.67	64.29	71.43	71.43	**73.81**
BalanceScale	**73.72**	**73.72**	**73.72**	67.95	67.95	67.95
Thyroid	**90.59**	88.89	88.89	88.89	88.89	88.89
LD Bupa	44.19	44.19	38.37	56.98	**58.14**	**58.14**
Iris	**89.47**	**89.47**	**89.47**	65.79	65.79	65.79
Glass	31.48	24.07	**35.19**	20.37	20.37	20.37
Monk1	83.45	82.73	**95.68**	66.91	66.91	66.91
Monk2	**80.67**	80.00	79.33	60.00	60.00	60.00
Monk3	74.82	75.54	92.09	**96.40**	**96.40**	**96.40**

Table 4.7. F-measure of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions

Datasets	**Triangular Membership**			**Trapezoidal Membership**		
	F-IG	**F-GR**	**F-GI**	**F-IG**	**F-GR**	**F-GI**
HeartStatlog	31.1	40.2	44.1	61.5	64.7	**65.6**
MammographicM	80.6	81.5	81.9	**82.5**	**82.5**	**82.5**
BreastCancer	90.1	85.9	89.5	94.1	**94.7**	**94.7**
Diabetes	59.5	58.2	58.2	**65.8**	65.6	65.6
Hepatitis	49.3	53.1	54.1	55.4	**66.4**	59.0
SpectHeart	73.1	**79.1**	**79.1**	73.1	**79.1**	**79.1**
Yeast	38.1	38.4	**41.3**	32.8	32.8	32.8
VertebralCol 2C	57.8	56.8	56.8	**64.8**	**64.8**	**64.8**
VertebralCol 3C	**48.8**	**48.8**	**48.8**	37.7	37.7	37.7
Ecoli	62.0	63.0	61.1	71.6	71.6	**73.9**
BalanceScale	**71.1**	**71.1**	**71.1**	64.9	64.9	64.9
Thyroid	**92.7**	89.0	89.0	87.7	87.7	87.7
LD Bupa	40.7	39.8	34.6	46.8	**48.9**	**48.9**
Iris	**89.4**	**89.4**	**89.4**	66.1	66.1	66.1
Glass	29.8	21.8	**31.3**	18.5	18.5	18.5
Monk1	83.5	82.7	**95.7**	66.9	66.9	66.9
Monk2	**80.5**	79.9	79.3	58.2	58.2	58.2
Monk3	74.8	75.5	92.1	**96.4**	**96.4**	**96.4**

Table 4.8. Number of Rules of the for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions

Datasets	**Triangular Membership**			**Trapezoidal Membership**		
	F-IG	**F-GR**	**F-GI**	**F-IG**	**F-GR**	**F-GI**
Heart Statlog	156	131	91	98	87	**70**
MammographicM	62	60	59	21	21	**20**
Breast Cancer	100	96	77	**40**	41	44
Diabetes	134	130	116	33	34	**32**
Hepatitis	83	72	48	88	52	**37**
Spect Heart	111	101	**75**	111	101	**75**
Yeast	92	89	89	**19**	**19**	**19**
Vertebral Col 2C	26	27	27	**8**	9	9
Vertebral Col 3C	27	27	25	**9**	**9**	**9**
Ecoli	42	42	40	22	22	**21**
Balance Scale	69	69	69	**14**	15	16
Thyroid	17	14	15	**6**	**6**	**6**
LD Bupa	42	45	42	10	**9**	**9**
Iris	13	12	12	10	9	**8**
Glass	38	38	34	16	16	**15**
Monk1	229	194	125	64	56	**36**
Monk2	199	193	188	63	62	**60**
Monk3	240	216	65	52	33	**19**

Table 4.9. Training Time in Seconds for the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions

Datasets	**Triangular Membership**			**Trapezoidal Membership**		
	F-IG	**F-GR**	**F-GI**	**F-IG**	**F-GR**	**F-GI**
Heart Statlog	3.47	2.92	2.6	2.22	1.93	**1.50**
MammographicM	3.25	2.39	2.38	2.13	**2.1**	2.13
Breast Cancer	3.56	3.48	**2.55**	2.75	2.71	2.65
Diabetes	5.93	4.67	4.52	3.12	**3.06**	3.08
Hepatitis	2.38	1.76	**0.91**	2.17	1.37	1.31
Spect Heart	10.34	5.23	**2.41**	9.53	5.36	2.5
Yeast	14	10.51	10.55	8.76	**8.67**	8.74
Vertebral Col 2C	1.07	0.70	**0.43**	0.54	0.64	0.53
Vertebral Col 3C	1.08	0.70	0.69	**0.53**	0.6	0.54
Ecoli	1.19	0.90	0.93	0.76	0.77	**0.74**
Balance Scale	1.47	1.08	**0.36**	0.84	0.89	0.85
Thyroid	0.45	0.30	0.29	0.28	**0.26**	**0.26**
LD Bupa	1.04	0.72	0.86	0.53	0.54	**0.52**
Iris	0.27	0.17	0.16	0.21	**0.15**	0.16
Glass	0.68	0.68	0.62	0.62	0.55	**0.45**
Monk1	2.78	1.9	1.71	1.52	1.5	**1.3**
Monk2	1.76	1.66	**1.20**	1.79	1.66	1.7
Monk3	1.83	1.53	**0.71**	1.49	1.20	1.07

Table 4.10. Experimental Results in Test Time in Seconds of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with Triangular and Trapezoidal Membership Functions

Datasets	**Triangular Membership**			**Trapezoidal Membership**		
	F-IG	**F-GR**	**F-GI**	**F-IG**	**F-GR**	**F-GI**
Heart Statlog	0.29	0.31	0.30	0.28	0.27	**0.20**
MammographicM	0.54	0.35	0.34	0.3	**0.29**	**0.29**
Breast Cancer	0.62	0.42	0.4	0.35	0.36	**0.34**
Diabetes	0.79	0.57	0.58	**0.40**	0.46	0.41
Hepatitis	0.35	0.28	0.3	**0.26**	0.32	0.29
Spect Heart	0.67	0.42	**0.23**	0.41	0.43	0.30
Yeast	1.71	1.22	1.2	1.01	**0.99**	**0.99**
Vertebral Col 2C	0.24	0.2	0.19	0.15	**0.14**	0.19
Vertebral Col 3C	0.28	0.15	0.14	0.18	**0.13**	**0.13**
Ecoli	0.24	0.17	0.17	0.16	**0.15**	0.16
Balance Scale	0.30	0.18	**0.11**	0.17	0.15	0.15
Thyroid	0.17	**0.10**	0.1	0.11	0.11	0.15
LD Bupa	0.24	0.15	0.20	0.16	**0.13**	**0.13**
Iris	0.11	**0.08**	**0.08**	0.09	0.09	**0.08**
Glass	0.16	0.16	0.13	0.20	0.15	**0.12**
Monk1	0.37	0.37	0.27	0.23	**0.22**	0.23
Monk2	0.3	0.32	0.46	**0.25**	**0.25**	0.28
Monk3	0.31	0.25	0.19	0.23	0.21	**0.19**

Performance comparison of fuzzy versions of splitting criteria are presented in section 4.5.

4.3. Effect of LinguisticTerms

In this thesis, datasets were fuzzified before learning the decision tree, and we obtained two different set of fuzzified data which can be explained as follows: in the first method if an element is a member of more than one fuzzy set, the linguistic term having the maximum membership value is chosen to fuzzify the data. The second method, on the other hand, uses all linguistic terms that have greater than zero membership for an element. Experimental results obtained when all linguistic terms of the elements are used during the decision tree induction are explained in this

section. The results given in the previous sections are belong to the fuzzification process which uses single linguistic term for each element.

For all linguistic terms which are obtained by using triangular or trapezoidal membership functions, experimental results in terms of accuracy for the "ID3 with fuzzy data and basic splitting criterion" and "ID3 with fuzzy data and fuzzy splitting criterion" methods are presented in the next tables. Experimental results of single linguistic terms which are obtained by using triangular or trapezoidal membership functions are compared with all linguistic terms. In the following tables, basic splitting criteria and fuzzy version of basic splitting criteria are compared with each other. If all linguistic terms are used to apply a rule for a test sample we used four rule selection methods that are "Test 1", "Test 2", "Test 3", and "Test 4". On the other hand, if one linguistic terms is used, same result is obtained for all rule selection methods. "T1", T2", "T3", and "T4" are short form of "Test 1", "Test 2", "Test 3", and "Test 4" and they are detailed in method section.

According to the results presented in Tables 4.11 - 4.28 using all linguistic terms yields better classification performance with respect to using single linguistic term. Generally, for triangular and trapezoidal membership functions, results of 13 datasets out of 18 datasets for all linguistic terms have the best accuracy values. But single linguistic term is more successful for fuzzified Yeast, Thyroid, Iris, Monk 1, and Monk 3 datasets by using triangular membership function and for fuzzified Thyroid, LD Bupa, and Monk 3 datasets by using trapezoidal membership function.

Table 4.11. Classification Accuracy of "Heart Statlog" Dataset for All and Single Linguistic Terms

Heart Statlog	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	77.94	**80.88**	77.94	79.41	73.53	75.00	70.59	75.00	66.18	30.88
Trapezoidal	**79.41**	77.94	77.94	77.94	67.65	70.59	64.71	70.59	72.06	61.77
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	76.47	79.41	77.94	77.94	82.35	**83.82**	75.00	**83.82**	72.06	39.71
Trapezoidal	73.53	73.53	72.06	73.53	**77.94**	70.59	66.18	69.11	70.59	64.71
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	73.53	**79.41**	76.47	77.94	73.53	**79.41**	76.47	77.94	42.65	44.12
Trapezoidal	75.00	**77.94**	73.53	**77.94**	75.00	**77.94**	73.53	**77.94**	66.18	66.18

Table 4.12. Classification Accuracy of "Mammographic Masses" Dataset for All and Single Linguistic Terms

Mamm. Masses	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**82.08**	80.42	80.42	80.42	80.83	78.33	79.17	77.50	80.00	80.83
Trapezoidal	81.67	79.58	79.17	79.58	**82.50**	82.50	81.25	82.08	**82.50**	**82.50**
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**81.67**	80.42	80.42	80.42	80.83	80.83	80.83	80.41	80.00	**81.67**
Trapezoidal	81.67	79.58	79.17	79.58	**82.50**	**82.50**	81.25	82.08	**82.50**	**82.50**
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	81.67	81.25	**82.50**	**82.50**	81.67	80.83	81.67	82.08	82.08	82.08
Trapezoidal	81.67	80.00	79.17	79.58	**82.50**	**82.50**	80.83	82.08	**82.50**	**82.50**

Table 4.13. Classification Accuracy of "Breast Cancer" Dataset for All and Single Linguistic Terms

Breast Cancer	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**96.49**	94.74	94.74	94.74	96.49	95.32	94.74	94.74	94.74	90.06
Trapezoidal	**97.66**	95.91	96.50	95.91	97.08	95.91	95.32	95.91	93.57	94.15
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**95.91**	94.74	94.74	94.74	96.49	96.49	95.32	96.49	93.57	85.97
Trapezoidal	**97.66**	95.91	96.49	95.91	97.08	95.91	95.32	95.32	94.15	94.74
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**96.49**	95.32	94.15	94.15	**96.49**	95.32	95.32	95.32	90.64	89.47
Trapezoidal	**97.08**	95.91	95.32	95.32	**97.08**	95.91	95.32	95.32	94.74	94.74

Table 4.14. Classification Accuracy of "Diabetes" Dataset for All and Single Linguistic Terms

Diabetes	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**65.10**	63.54	63.02	63.54	64.58	64.06	63.02	64.06	61.46	61.98
Trapezoidal	65.63	64.06	63.02	64.06	65.10	65.10	63.02	64.58	**67.71**	67.19
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	65.10	63.54	63.02	63.54	64.58	**65.63**	63.02	64.06	61.98	60.42
Trapezoidal	65.63	64.06	63.02	64.06	65.10	64.58	63.02	64.06	**67.71**	66.67
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**65.10**	63.54	63.02	63.54	65.10	65.10	63.02	63.54	60.94	60.42
Trapezoidal	**65.63**	64.06	63.54	64.06	66.67	67.19	64.06	65.10	66.67	66.67

Table 4.15. Classification Accuracy of "Hepatitis" Dataset for All and Single Linguistic Terms

Hepatitis	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	76.92	76.92	76.92	76.92	**82.05**	**82.05**	79.49	**82.05**	71.80	51.28
Trapezoidal	76.92	76.92	76.92	76.92	74.36	**84.62**	76.92	**84.62**	76.92	56.41
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	66.67	66.67	66.67	66.67	**87.18**	**87.18**	84.62	**87.18**	64.10	53.85
Trapezoidal	66.67	66.67	66.67	66.67	**76.92**	**76.92**	74.36	**76.92**	71.80	66.67
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	66.67	**69.23**	**69.23**	**69.23**	66.67	**69.23**	**69.23**	**69.23**	58.97	56.41
Trapezoidal	**76.92**	**76.92**	**76.92**	**76.92**	**76.92**	**76.92**	**76.92**	**76.92**	58.97	58.97

Table 4.16. Classification Accuracy of "Spect heart" Dataset for All and Single Linguistic Terms

Spect Heart	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**77.61**	**77.61**	**77.61**	**77.61**	71.64	71.64	71.64	71.64	**77.61**	71.64
Trapezoidal	**77.61**	**77.61**	**77.61**	**77.61**	71.64	71.64	71.64	71.64	**77.61**	71.64
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11
Trapezoidal	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11
Trapezoidal	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11	79.11

Table 4.17. Classification Accuracy of "Yeast" Dataset for All and Single Linguistic Terms

Yeast	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	37.47	33.69	33.69	33.69	38.00	33.69	33.69	33.69	**43.67**	39.62
Trapezoidal	**40.97**	34.50	33.96	33.96	40.70	35.04	34.50	34.50	36.39	36.92
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	38.81	33.42	33.69	33.69	38.81	33.69	33.69	33.69	**43.67**	39.89
Trapezoidal	40.97	34.50	33.96	33.96	**41.24**	37.47	34.23	34.23	36.66	36.93
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	37.74	33.69	33.69	33.69	37.74	33.69	33.69	33.69	**42.86**	**42.86**
Trapezoidal	40.97	34.23	33.96	33.96	**41.51**	39.89	33.96	35.04	36.93	36.93

Table 4.18. Classification Accuracy of "Vertebral Column 2C" Dataset for All and Single Linguistic Terms

Vertebral Column 2C	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	63.42	**75.61**	**75.61**	**75.61**	62.20	39.02	39.02	39.02	57.32	56.10
Trapezoidal	52.44	**74.39**	**74.39**	**74.39**	50.00	57.32	50.00	57.32	68.29	70.73
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	63.42	**75.61**	**75.61**	**75.61**	63.42	59.76	40.24	40.24	54.88	54.88
Trapezoidal	52.44	**74.39**	**74.39**	**74.39**	50.00	57.32	50.00	57.32	69.51	70.73
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	63.42	**75.61**	**75.61**	**75.61**	63.42	58.54	40.24	40.24	54.88	54.88
Trapezoidal	50.00	**71.95**	**71.95**	**71.95**	48.78	58.54	48.78	58.54	70.73	70.73

Table 4.19. Classification Accuracy of "Vertebral Column 3C" Dataset for All and Single Linguistic Terms

Vertebral Column 3C	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	60.26	57.69	52.56	52.56	**64.10**	51.28	51.28	51.28	56.41	57.69
Trapezoidal	**60.26**	55.13	55.13	55.13	58.97	55.13	53.85	53.85	41.03	51.28
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	60.26	61.54	52.56	52.56	**61.54**	52.56	52.56	52.56	58.97	57.69
Trapezoidal	**60.26**	55.13	55.13	55.13	58.97	55.13	53.85	52.56	41.03	51.28
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**61.54**	53.85	53.85	53.85	**61.54**	52.56	52.56	52.56	57.69	57.69
Trapezoidal	**58.97**	53.85	53.85	53.85	**58.97**	53.85	53.85	53.85	51.28	51.28

Table 4.20. Classification Accuracy of "Ecoli" Dataset for All and Single Linguistic Terms

Ecoli	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**75.00**	50.00	50.00	50.00	**75.00**	61.91	30.95	30.95	72.62	65.48
Trapezoidal	71.43	52.38	50.00	51.19	**72.62**	66.67	65.48	69.05	71.43	71.43
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**75.00**	50.00	50.00	50.00	**75.00**	61.91	30.95	29.76	67.86	66.67
Trapezoidal	71.43	52.38	50.00	51.19	**72.62**	60.71	52.38	53.57	71.43	71.43
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**75.00**	50.00	50.00	50.00	**75.00**	55.95	30.95	29.76	64.29	64.29
Trapezoidal	70.24	52.38	50.00	51.19	71.43	54.76	52.38	53.57	**73.81**	**73.81**

Table 4.21. Classification Accuracy of "Balance Scale" Dataset for All and Single Linguistic Terms

Balance Scale	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	85.90	82.05	76.92	78.85	**86.54**	83.33	75.64	77.56	73.72	73.72
Trapezoidal	85.90	82.05	76.92	78.85	**86.54**	82.69	75.00	78.85	67.95	67.95
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	85.90	82.05	76.92	78.85	**86.54**	84.62	76.92	79.49	73.72	73.72
Trapezoidal	85.90	82.05	76.92	78.85	**86.54**	83.97	76.28	80.77	67.95	67.95
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**86.54**	82.69	75.00	78.85	85.90	83.33	75.64	78.21	73.72	73.72
Trapezoidal	**86.54**	82.69	75.00	78.85	85.90	84.62	75.00	78.21	67.95	67.95

Table 4.22. Classification Accuracy of "Thyroid" Dataset for All and Single Linguistic Terms

Thyroid	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	81.48	87.04	81.48	81.48	81.48	83.33	81.48	83.33	**90.59**	**90.59**
Trapezoidal	83.33	83.33	81.48	81.48	83.33	87.04	81.48	81.48	**88.89**	**88.89**
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	81.48	87.04	81.48	81.48	83.33	83.33	81.48	83.33	**90.74**	**88.89**
Trapezoidal	83.33	83.33	81.48	81.48	83.33	87.04	81.48	81.48	**88.89**	**88.89**
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	83.33	81.48	81.48	81.48	83.33	81.48	81.48	81.48	**88.89**	**88.89**
Trapezoidal	85.19	81.48	81.48	81.48	85.19	83.33	81.48	81.48	**88.89**	**88.89**

Table 4.23. Classification Accuracy of "LD Bupa" Dataset for All and Single Linguistic Terms

LD Bupa	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	47.67	44.19
Trapezoidal	55.81	56.98	56.98	56.98	55.81	55.81	55.81	55.81	**58.14**	56.98
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	46.51	44.19
Trapezoidal	55.81	56.98	56.98	56.98	55.81	55.81	55.81	55.81	**58.14**	**58.14**
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	**55.81**	43.02	38.37
Trapezoidal	55.81	56.98	56.98	56.98	55.81	56.98	56.98	56.98	**58.14**	**58.14**

Table 4.24. Classification Accuracy of "Iris" Dataset for All and Single Linguistic Terms

Iris	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	71.05	78.95	55.26	65.79	71.05	76.32	55.26	55.26	**92.11**	89.47
Trapezoidal	92.10	89.47	73.68	81.58	**94.74**	89.47	55.26	86.84	65.79	65.79
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	71.05	78.95	55.26	65.79	71.05	68.42	55.26	55.26	**92.11**	89.47
Trapezoidal	92.10	89.47	73.68	**94.74**	**94.74**	89.47	65.79	86.84	65.79	65.79
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	71.05	73.68	55.26	55.26	71.05	73.68	55.26	60.53	**89.47**	**89.47**
Trapezoidal	**92.10**	78.95	55.26	63.16	89.47	73.68	57.90	52.63	65.79	65.79

Table 4.25. Classification Accuracy of "Glass" Dataset for All and Single Linguistic Terms

Glass	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	35.19	42.59	40.74	44.44	35.19	**46.30**	**46.30**	**46.30**	38.89	31.48
Trapezoidal	44.44	48.15	42.59	46.30	**50.00**	48.15	46.30	48.15	27.78	20.37
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	33.33	**55.56**	44.44	48.15	35.19	51.85	46.30	50.00	24.07	24.07
Trapezoidal	42.59	**46.30**	44.44	44.44	**46.30**	**46.30**	42.59	**46.30**	25.93	20.37
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	33.33	48.15	38.89	44.44	33.33	**51.85**	38.89	**51.85**	22.22	35.19
Trapezoidal	42.59	**44.44**	42.59	42.59	42.59	**44.44**	**44.44**	**44.44**	20.37	20.37

Table 4.26. Classification Accuracy of "Monk1" Dataset for All and Single Linguistic Terms

Monk1	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	87.05	87.05	87.05	87.05	73.38	76.98	71.22	73.38	**98.56**	83.45
Trapezoidal	**98.56**	**98.56**	**98.56**	**98.56**	83.45	83.45	83.45	83.45	66.91	66.91
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	87.05	87.05	87.05	87.05	71.94	73.38	68.35	71.94	**98.56**	82.73
Trapezoidal	**98.56**	**98.56**	**98.56**	**98.56**	82.73	82.73	82.73	82.73	66.91	66.91
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	77.70	80.58	73.38	77.70	77.70	80.58	73.38	77.70	**95.68**	**95.68**
Trapezoidal	**95.68**	**95.68**	**95.68**	**95.68**	**95.68**	**95.68**	**95.68**	**95.68**	66.91	66.91

Table 4.27. Classification Accuracy of "Monk2" Dataset for All and Single Linguistic Terms

Monk2	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	82.67	**85.33**	78.00	82.67	**85.33**	**85.33**	82.67	**85.33**	78.00	80.67
Trapezoidal	78.00	78.00	78.00	78.00	**80.67**	**80.67**	**80.67**	**80.67**	60.00	60.00
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	82.67	**85.33**	78.00	82.67	83.33	**85.33**	79.33	83.33	77.33	80.00
Trapezoidal	77.33	77.33	77.33	77.33	**80.00**	**80.00**	**80.00**	**80.00**	60.00	60.00
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	82.67	**85.33**	78.00	82.67	82.67	**85.33**	76.67	82.67	79.33	79.33
Trapezoidal	**79.33**	**79.33**	**79.33**	**79.33**	**79.33**	**79.33**	**79.33**	**79.33**	60.00	60.00

Table 4.28. Classification Accuracy of "Monk3" Dataset for All and Single Linguistic Terms

Monk3	All Linguistic Term								One Linguistic Term	
	IG				F-IG				IG	F-IG
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	81.30	81.30	81.30	81.30	69.07	70.50	66.91	69.07	**93.53**	74.82
Trapezoidal	93.53	93.53	93.53	93.53	74.82	74.82	74.82	74.82	**96.40**	**96.40**
	GR				F-GR				GR	F-GR
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	81.30	81.30	80.58	81.30	70.50	70.50	64.75	70.50	**93.53**	**75.54**
Trapezoidal	93.53	93.53	93.53	93.53	74.82	74.82	74.82	74.82	**96.40**	**96.40**
	GI				F-GI				GI	F-GI
	T1	T2	T3	T4	T1	T2	T3	T4		
Triangular	74.10	74.10	67.63	74.10	82.67	85.33	76.67	82.67	**92.09**	**92.09**
Trapezoidal	92.09	92.09	92.09	92.09	92.09	92.09	92.09	92.09	**96.40**	**96.40**

Table 4.29. F-Measure of "ID3 with Fuzzified Data, Basic, and Fuzzified Splitting Criteria" Method with All Linguistic Terms

Datasets	**Triangular Membership**						**Trapezoidal Membership**					
	IG	GR	GI	F-IG	F-GR	F-GI	IG	GR	GI	F-IG	F-GR	F-GI
HeartStatlog	78.2	76.7	74.1	73.5	**82.7**	74.1	79.2	73.1	74.8	67.6	78.2	75.1
Mammog.M.	82.1	81.7	81.7	80.8	80.8	81.7	81.7	81.7	81.7	**82.5**	**82.5**	**82.5**
BreastCancer	96.5	95.9	96.5	96.5	95.9	96.5	**97.7**	**97.7**	97.1	97.1	**97.7**	97.1
Diabetes	54.1	54.1	54.1	53.8	53.8	54.1	57.7	57.7	57.7	56.7	56.7	**59.5**
Hepatitis	76.5	67.2	66.0	78.2	**87.0**	67.2	76.5	67.2	76.3	70.1	77.3	75.5
SpectHeart	77.0	78.8	78.0	73.1	**79.1**	**79.1**	77.0	78.8	78.0	73.1	**79.1**	**79.1**
Yeast	28.2	28.7	28.2	27.9	28.7	28.2	34.0	34.0	34.0	32.1	34.0	**34.8**
Verteb.Col 2C	**65.4**	**65.4**	**65.4**	64.4	**65.4**	**65.4**	56.1	56.1	53.5	53.6	53.6	52.3
Verteb.Col 3C	53.3	53.3	54.5	**58.1**	54.5	54.5	54.4	54.4	53.2	53.7	53.7	53.7
Ecoli	69.0	69.0	69.0	69.0	69.0	69.0	68.8	68.8	67.8	**70.7**	69.8	68.8
BalanceScale	81.3	81.3	**81.9**	**81.9**	**81.9**	81.3	81.3	81.3	**81.9**	**81.9**	**81.9**	81.3
Thyroid	77.3	77.3	80.2	77.3	80.2	80.2	80.2	80.2	**82.8**	80.2	80.2	**82.8**
LD Bupa	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**	**41.2**
Iris	69.7	69.7	69.7	69.7	69.7	69.7	92.2	92.2	92.2	**94.7**	**94.7**	88.7
Glass	30.6	28.6	28.6	30.8	30.8	28.6	40.0	38.0	37.9	**43.1**	42.1	37.9
Monk1	86.8	86.8	77.6	72.9	71.4	77.6	**98.6**	**98.6**	95.7	83.4	82.7	95.7
Monk2	82.9	82.9	82.9	**85.6**	83.5	82.9	78.2	77.3	79.3	80.4	79.9	79.3
Monk3	80.7	80.7	73.8	68.7	70.0	73.8	**93.5**	**93.5**	92.1	74.8	74.8	92.1

Table 4.29 shows f-measure values of classification with fuzzy decision trees that use basic and fuzzy splitting criteria. In this experiment, triangular and trapezoidal membership functions were used to fuzzify numerical data, and results presented in Table 4.29 shows only "Test1". The best f-measure values for each datasets are written in the boldface. According to the results, f-measure presented in Table 4.29 and accuracy presented in Tables 4.11 - 4.28 have almost same results. There is no remarkable differences between two measures.

Table 4.30 shows experimental results in terms of number of rules obtained when "ID3 with fuzzified data and basic splitting criteria" method with all linguistic terms which are obtained by using triangular or trapezoidal membership functions. The best results are written in boldface in the table. Fuzzified data with all linguistic terms have disadvantages on classification. The number of rules for all linguistic terms is more than the number of rules obtained by using single linguistic terms which is shown in Table 4.3. So decision tree with single linguistic term takes less

time for train and test phases as shown in Table 4.4 and Table 4.5 respectively than that of all linguistic terms. For all linguistic terms and basic splitting criteria, training and test time in seconds for triangular and trapezoidal membership functions are given in Table 4.31 and 4.32 respectively. In addition, for single linguistic term and fuzzy splitting criteria, number of rules given in Table 4.7 are less than number of rules learned from decision tree which uses all linguistic terms which is presented in Table 4.33. So training and test parts take less time as presented in Table 4.8 and 4.9. For all linguistic terms and fuzzy splitting criteria, training and test times in seconds for triangular and trapezoidal membership functions are given in Table 4.34 and 4.35 respectively.

When all linguistic terms that have greater than zero membership for an element are used, a lot of computations need to be done to learn a decision tree. So, for all linguistic terms and both basic and fuzzy splitting criteria, training and test part takes longer time than single linguistic term.

According to the results presented in Table 4.30, number of rules obtained from fuzzy decision tree which uses trapezoidal membership function to fuzzify the datasets are less than number of rules obtained from fuzzy decision tree which employs triangular membership functions. For “ID3 with fuzzified data and fuzzy splitting criteria” method using all linguistic terms, number of rules obtained from fuzzy decision tree with trapezoidal membership function are less than that of triangular membership function, as presented in Table 4.33.

Table 4.30. Number of Rules for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method with All Linguistic Terms

Datasets	Triangular Membership			Trapezoidal Membership		
	IG	GR	GI	IG	GR	GI
Heart Statlog	638	634	1410	**144**	145	348
MammographicM.	143	145	137	93	99	**92**
Breast Cancer	2335	2369	2747	**263**	264	432
Diabetes	2312	2370	2430	509	**492**	552
Hepatitis	65	62	268	58	**55**	167
Spect Heart	**47**	48	75	**47**	48	75
Yeast	814	816	864	216	**215**	224
Vertebral Col 2C	327	335	319	**117**	**117**	144
Vertebral Col 3C	258	277	283	**121**	123	146
Ecoli	499	503	568	172	**170**	215
Balance Scale	**77**	**77**	**77**	**77**	**77**	**77**
Thyroid	79	83	79	**57**	**57**	77
LD Bupa	334	328	352	155	**152**	163
Iris	54	55	57	**19**	**19**	23
Glass	1610	1590	2102	242	**236**	302
Monk1	81	81	209	**63**	**63**	125
Monk2	191	189	191	193	**184**	188
Monk3	125	125	224	37	**35**	61

Table 4.31. Training Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method with All Linguistic Terms

Datasets	Triangular Membership			Trapezoidal Membership		
	IG	GR	GI	IG	GR	GI
Heart Statlog	5.86	5.12	7.55	**2.53**	2.57	4.97
MammographicM.	4.56	3.77	3.58	3.31	2.97	**2.82**
Breast Cancer	24.95	24.71	27.17	6.27	**5.88**	7.92
Diabetes	35.59	35.45	33.12	10.36	10.21	**10.16**
Hepatitis	1.43	1.13	3.60	1.27	**1.12**	2.89
Spect Heart	**3.76**	3.99	4.85	4.04	4.05	5.06
Yeast	**3.26**	30.01	28.41	17.47	15.58	15.59
Vertebral Col 2C	3.24	3.09	2.47	1.54	**1.45**	1.51
Vertebral Col 3C	3.08	2.75	2.41	1.64	**1.59**	1.64
Ecoli	5.07	4.77	4.36	2.03	**1.84**	2.02
Balance Scale	1.75	1.38	**1.34**	1.41	1.39	1.36
Thyroid	0.88	0.74	0.65	0.59	**0.54**	0.59
LD Bupa	2.88	2.92	2.70	1.94	1.87	**1.84**
Iris	0.37	0.26	0.26	0.21	**0.17**	0.25
Glass	8.64	8.74	9.00	1.79	**1.77**	1.81
Monk1	1.93	1.45	1.92	1.61	**1.32**	1.61
Monk2	2.88	2.18	2.09	2.41	2.13	**2.03**
Monk3	2.18	1.69	2.00	1.38	**1.11**	1.33

Table 4.32. Test Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method with All Linguistic Terms

Datasets	Triangular Membership			Trapezoidal Membership		
	IG	GR	GI	IG	GR	GI
Heart Statlog	0.86	0.68	1.35	0.37	**0.36**	0.46
Mammographic M.	0.76	0.52	0.54	0.49	0.46	**0.42**
Breast Cancer	4.00	3.78	4.53	0.78	**0.66**	0.96
Diabetes	5.16	4.92	5.39	1.35	**1.21**	1.38
Hepatitis	0.71	**0.25**	0.37	0.28	**0.25**	0.34
Spect Heart	0.61	0.40	0.40	**0.37**	0.38	0.41
Yeast	5.01	4.46	4.66	2.08	**1.85**	1.89
Vertebral Col 2C	0.58	0.45	0.46	0.25	**0.22**	0.24
Vertebral Col 3C	0.46	0.41	0.39	0.24	**0.21**	0.22
Ecoli	0.63	0.55	0.56	0.31	**0.27**	0.31
Balance Scale	0.34	0.22	**0.21**	0.22	0.22	0.22
Thyroid	0.27	0.14	0.14	0.14	**0.12**	0.13
LD Bupa	0.51	0.41	0.44	0.28	**0.24**	0.25
Iris	0.13	0.10	0.10	0.09	0.09	**0.08**
Glass	1.07	1.13	1.43	0.31	**0.27**	0.37
Monk1	0.35	0.23	0.34	0.32	**0.22**	0.27
Monk2	0.50	**0.34**	0.35	0.40	**0.34**	0.36
Monk3	0.44	0.27	0.36	0.27	0.25	**0.22**

Table 4.33. Number of Rules for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with All Linguistic Terms

Datasets	Triangular Membership			Trapezoidal Membership		
	F-IG	F-GR	F-GI	F-IG	F-GR	F-GI
Heart Statlog	5437	4018	1414	1220	979	**348**
MammographicM.	162	154	137	109	106	**92**
Breast Cancer	3634	3170	2747	576	570	**432**
Diabetes	2662	2482	2430	669	626	**552**
Hepatitis	4217	2620	273	1024	755	**167**
Spect Heart	111	101	**75**	111	101	**75**
Yeast	926	866	864	314	282	**224**
Vertebral Col 2C	317	317	319	**135**	**135**	144
Vertebral Col 3C	269	287	283	**139**	**139**	146
Ecoli	602	574	568	246	236	**215**
Balance Scale	81	**77**	**77**	81	**77**	**77**
Thyroid	91	79	79	**62**	63	77
LD Bupa	357	342	352	**163**	**163**	**163**
Iris	59	59	57	**23**	**23**	**23**
Glass	2765	2592	2102	406	408	**302**
Monk1	305	305	209	229	194	**125**
Monk2	258	258	191	199	193	**188**
Monk3	305	303	224	241	208	**61**

Table 4.34. Training Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with All Linguistic Terms

Datasets	Triangular Membership			Trapezoidal Membership		
	F-IG	F-GR	F-GI	F-IG	F-GR	F-GI
Heart Statlog	58.20	36.84	8.17	14.54	11.33	**5.14**
MammographicM.	4.56	3.99	3.95	3.17	3.03	**2.97**
Breast Cancer	39.41	32.41	28.62	9.95	9.19	**8.37**
Diabetes	43.73	40.61	41.41	13.38	12.08	**11.20**
Hepatitis	51.03	20.87	4.00	13.24	9.00	**3.06**
Spect Heart	9.37	5.21	**4.70**	9.46	5.30	4.77
Yeast	56.68	32.17	33.71	23.90	16.97	**16.04**
Vertebral Col 2C	3.42	3.34	3.65	**1.66**	1.70	1.83
Vertebral Col 3C	3.34	3.29	3.61	**1.69**	1.73	1.91
Ecoli	7.45	5.30	5.72	3.04	2.29	**2.21**
Balance Scale	1.32	**1.29**	1.32	1.44	1.41	1.50
Thyroid	1.05	0.96	1.07	**0.70**	**0.70**	0.79
LD Bupa	3.51	3.42	3.74	2.07	**2.03**	2.30
Iris	0.34	0.35	0.38	**0.22**	**0.22**	0.23
Glass	16.22	12.38	10.91	3.91	2.73	**2.24**
Monk1	2.51	2.44	2.06	1.99	1.80	**1.62**
Monk2	2.54	2.45	2.07	2.01	2.04	**2.00**
Monk3	2.49	2.42	2.05	2.09	1.87	**1.35**

Table 4.35. Test Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method with All Linguistic Terms

Datasets	Triangular Membership			Trapezoidal Membership		
	F-IG	F-GR	F-GI	F-IG	F-GR	F-GI
Heart Statlog	4.82	3.60	1.30	1.08	0.92	**0.45**
Mammographic M.	0.59	0.58	0.50	0.43	0.42	**0.40**
Breast Cancer	6.44	5.80	4.45	1.26	1.27	**0.93**
Diabetes	5.81	5.51	5.24	1.70	1.60	**1.30**
Hepatitis	2.72	1.95	0.36	0.82	0.70	**0.31**
Spect Heart	0.40	0.40	**0.38**	0.40	0.42	0.39
Yeast	5.05	4.50	4.47	2.14	2.05	**1.81**
Vertebral Col 2C	0.39	0.41	0.42	0.24	0.24	**0.22**
Vertebral Col 3C	0.33	0.36	0.36	0.23	0.23	**0.21**
Ecoli	0.53	0.57	0.52	0.29	**0.28**	0.31
Balance Scale	0.20	**0.19**	**0.19**	0.21	0.21	0.20
Thyroid	0.14	0.13	0.13	**0.12**	**0.12**	**0.12**
LD Bupa	0.42	0.41	0.42	0.26	0.26	**0.24**
Iris	0.09	0.09	0.09	**0.08**	0.11	**0.08**
Glass	1.95	1.93	1.38	0.41	0.40	**0.31**
Monk1	0.43	0.43	0.34	0.36	0.34	**0.27**
Monk2	0.42	0.41	0.39	0.36	0.58	**0.35**
Monk3	0.43	0.42	0.35	0.36	0.34	**0.33**

4.4. Comparison of Information Gain, Gain Ratio and Gini Index

In this thesis, information gain, gain ratio, and gini index measures are used to select the best attributes during the decision tree induction. In this section we compare these three measure in terms of classification accuracy, number of rules learned, training and test times.

Experimental results which compares accuracy of the "ID3 with fuzzy data and basic splitting criteria" method are shown in Table 4.1. For the results presented in this table, triangular and trapezoidal membership function, and linguistic term having the maximum membership value are used for fuzzification of the data. Figure 4.1 is a graphical representation of the results presented in Table 4.1. According to the results given in Figure 4.1, information gain has best accuracy value for 10 datasets out of 18 datasets, gain ratio is more successful for 9 datasets, and gini index has the best accuracy value for 3 datasets out of 18 datasets. So information gain is more successful than other splitting criteria. Figure 4.2 shows the number of rules learned when information gain, gain ratio, and gini index are used. According to the results, number of rules learned for fuzzy decision tree which is built by using information gain are less than gain ratio and gini index for 12 datasets out of 18 datasets. Training and test times of the decision trees when different splitting criteria are used are compared in Figure 4.3, and 4.4 respectively. According to these figures, although information gain yields less number of rules in the learned decision tree, there is no remarkable difference in the training and test time of the splitting criteria.

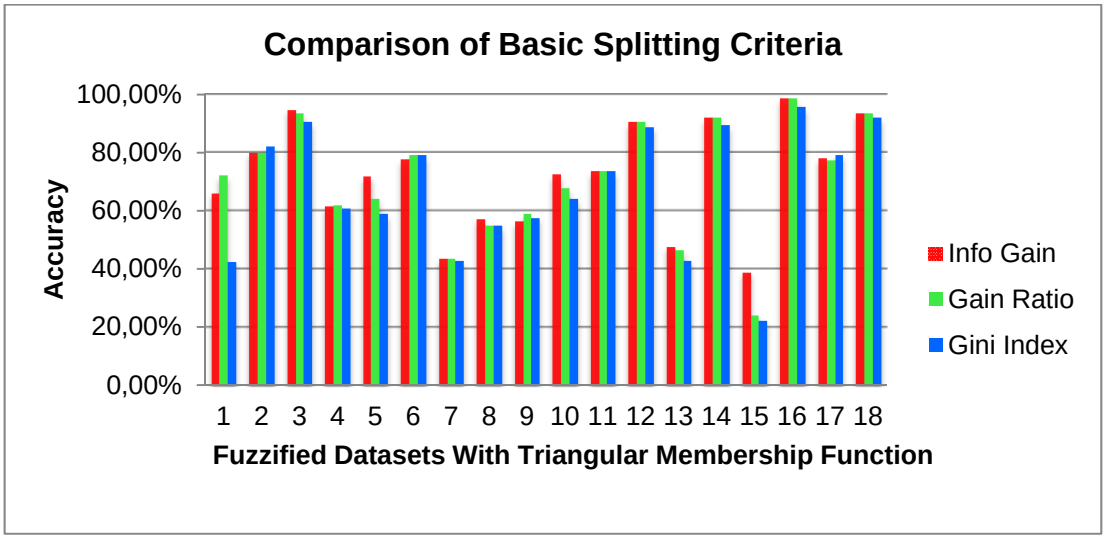

Figure 4.1. Accuracy of the "ID3 with Fuzzified Data and Basic Splitting Criteria" Method

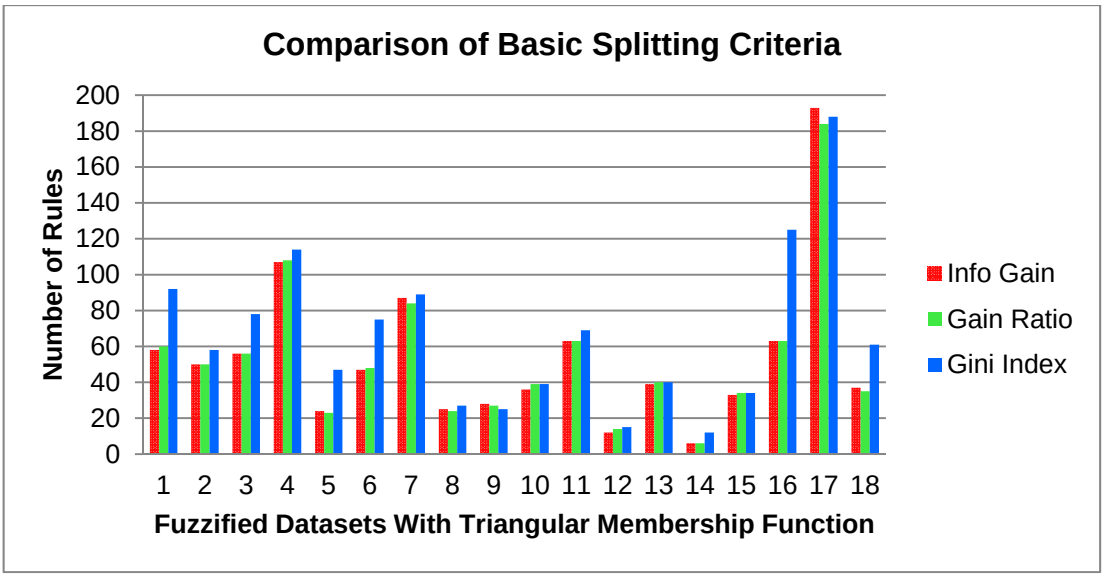

Figure 4.2. Number of Rules Learned by "ID3 with Fuzzified Data and Basic Splitting Criteria" Method

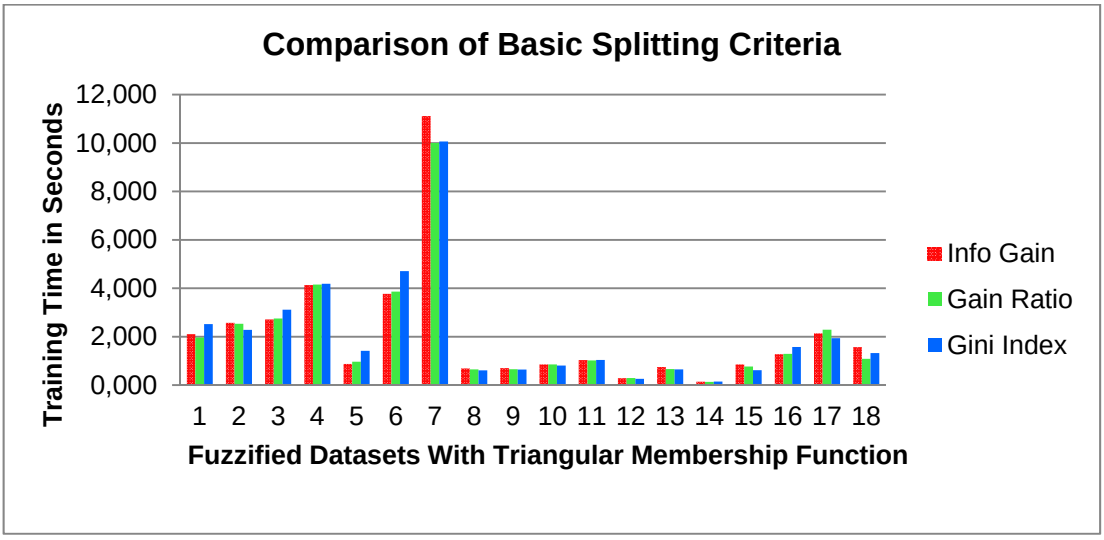

Figure 4.3. Training Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method

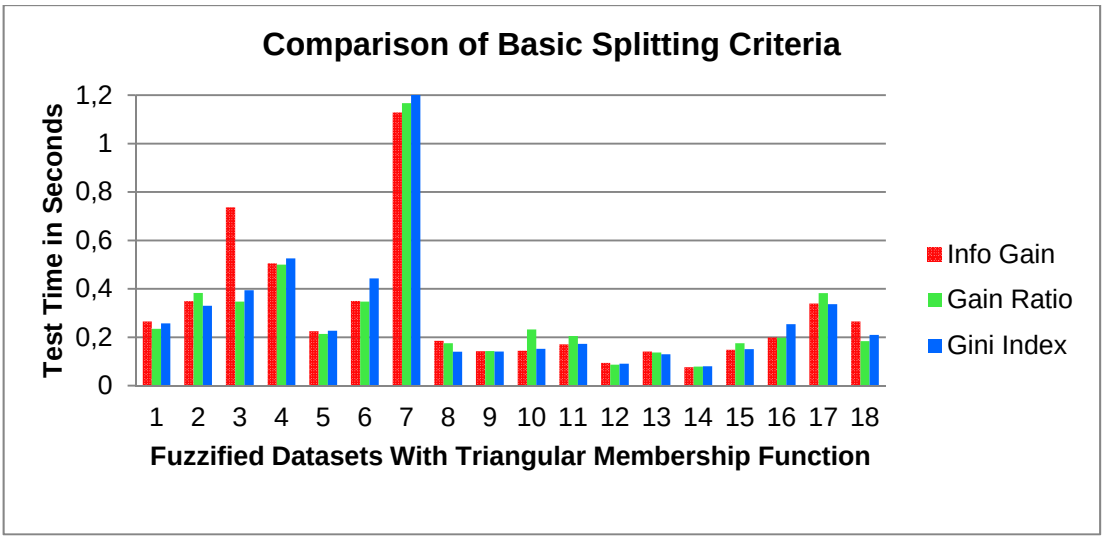

Figure 4.4. Test Time in Seconds for "ID3 with Fuzzified Data and Basic Splitting Criteria" Method

4.5. Comparison of Fuzzy Information Gain, Fuzzy Gain Ratio and Fuzzy Gini Index

In this thesis, fuzzy format of the basic splitting criteria that are information gain, gain ratio, and gini index are also used to select best attributes for fuzzy decision tree induction. The fuzzy versions of these measures are named as fuzzy information gain, fuzzy gain ratio, and fuzzy gini index. During the computation of

basic splitting criteria, proportion of samples for each class is used; for fuzzy splitting criteria on the other hands, membership values of the samples are taken for computations. Accuracy of the "ID3 with fuzzy data and fuzzy splitting criteria" method is presented in Table 4.5 in which, triangular membership function is used to fuzzify the data. Figure 4.5 shows the results presented in Table 4.5 as a graphical form.

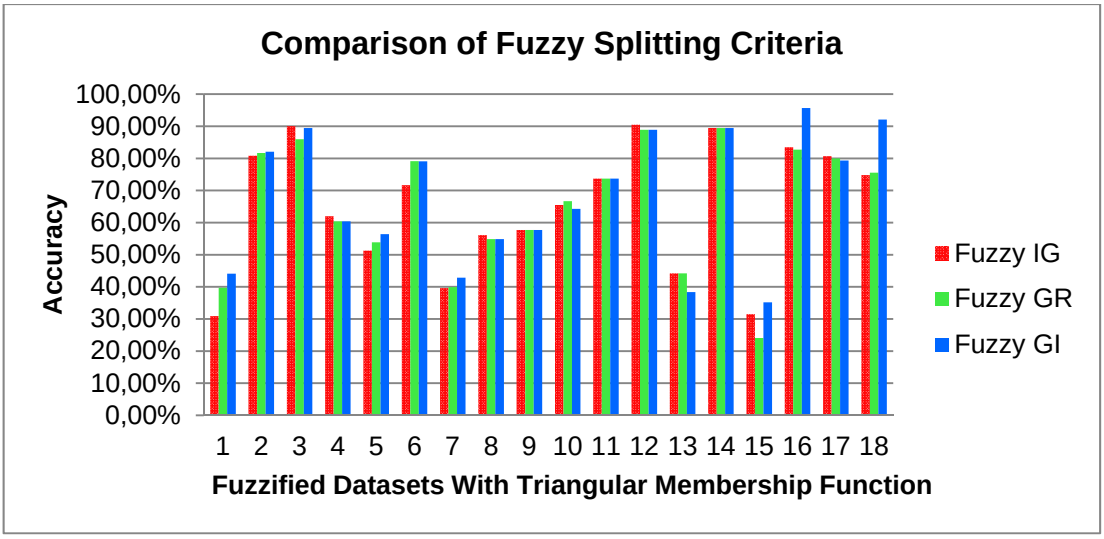

Figure 4.5. Accuracy of the "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method

According to Figure 4.5, fuzzy information gain has best accuracy value in 9 datasets out of 18 datasets. Fuzzy gain ratio is more successful in 6 datasets and fuzzy gini index has best accuracy value in 11 datasets. So fuzzy version of gini index is more successful in terms of classification accuracy than other fuzzy versions of splitting criteria additionaly, fuzzy gini index has less number of rules in 17 datasets out of 18 datasets. Only for 3 and 4 datasets fuzzy information gain and fuzzy gain ratio have lower number of rules respectively. The number of rules learned by "ID3 with fuzzified data and fuzzy splitting criteria" method are presented in Figure 4.6, and Table 4.7.

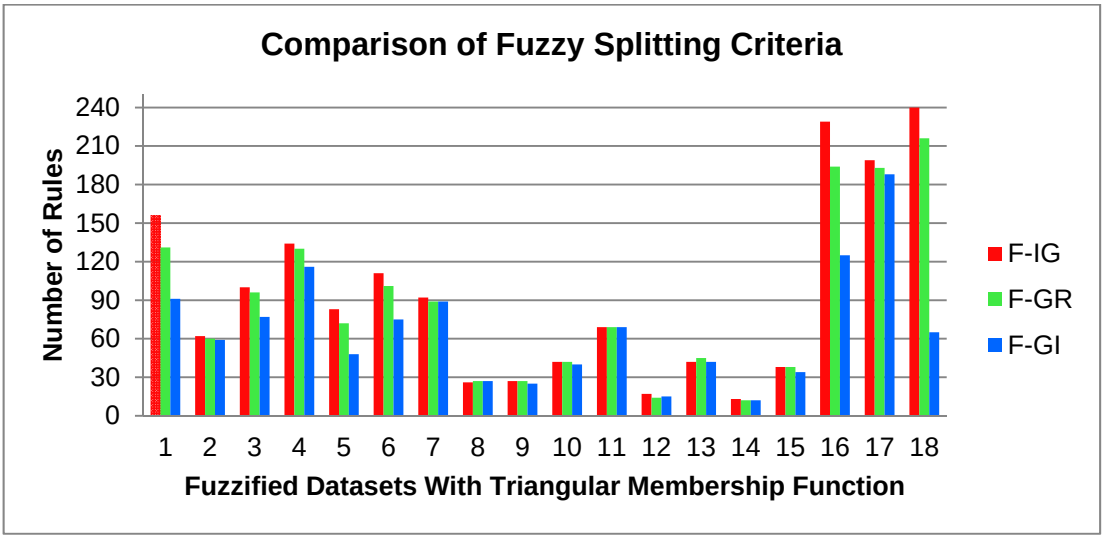

Figure 4.6. Number of Rules Learned by "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method

Figure 4.7 and 4.8 show training and test times in seconds for "ID3 with fuzzified data and fuzzy splitting criteria" method. These figures shows that fuzzy decision tree which is built by using fuzzy gini index takes less time for training and test in 13 datasets out of 18 datasets.

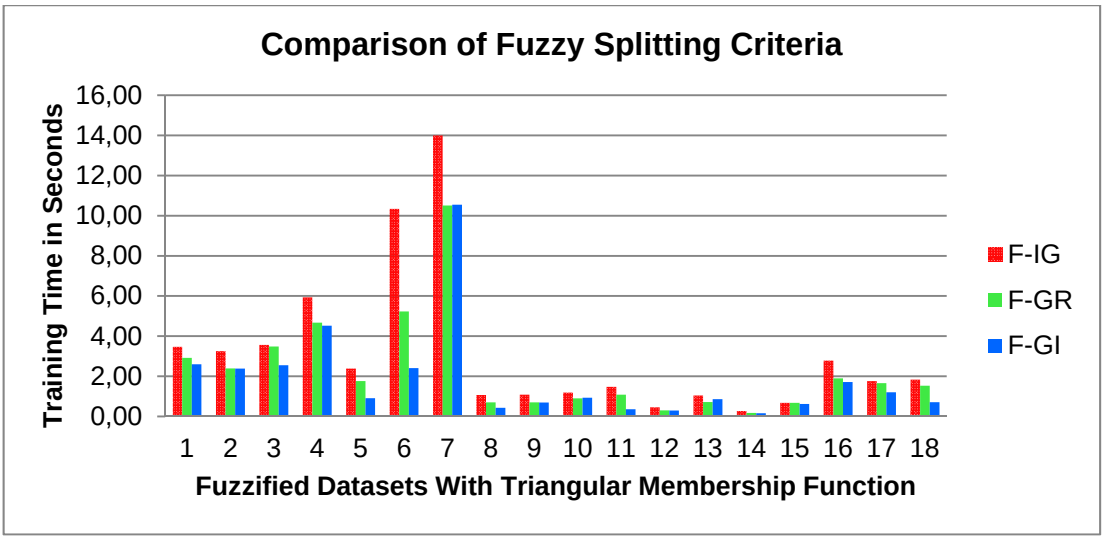

Figure 4.7. Training Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method

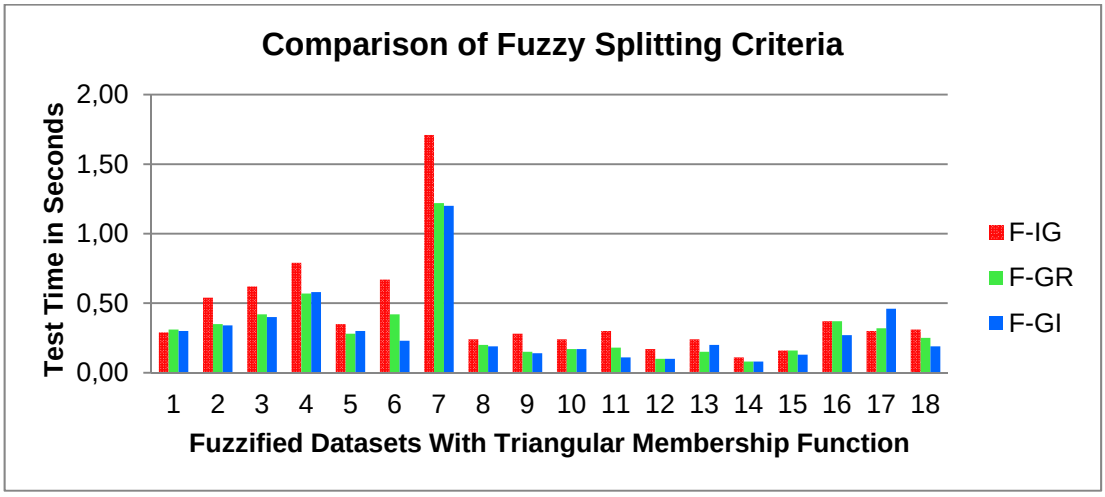

Figure 4.8. Test Time in Seconds for "ID3 with Fuzzified Data and Fuzzy Splitting Criteria" Method

According to the classification accuracy, number of rules learned, training and test time in seconds presented in this section, for most of the datasets, fuzzy version of the gini index is more successful than other fuzzy splitting criteria.

Figure 4.9 shows experimental results of classification accuracy of basic and fuzzy version of information gain. Triangular membership function is used to fuzzified datasets in this experiments presented in this figure. Fuzzy decision tree which is built by using information gain has the best performance in 14 datasets, on the other hand fuzzy information gain has the best performance in 6 datasets. So basic information gain is more successful than fuzzy version of this.

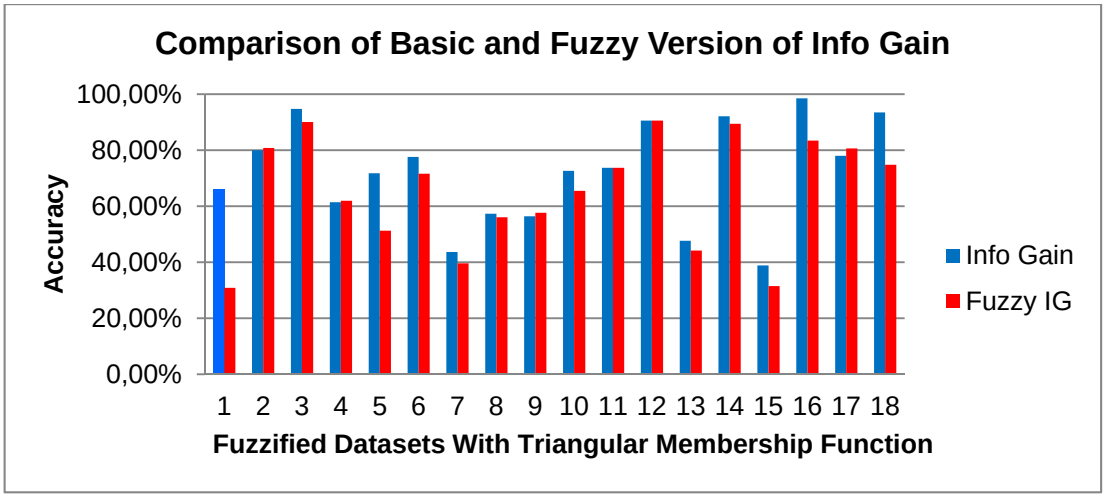

Figure 4.9. Accuracy of Basic and Fuzzy Version of Info Gain

Figure 4.10 shows experimental results of classification accuracy of basic and fuzzy version of gain ratio. Triangular membership function is used to fuzzified datasets in this experiments presented this figure. Fuzzy decision tree which is built by using gain ratio has the best performance in 16 datasets, on the other hand fuzzy gain ratio has the best performance in 6 datasets. So basic gain ratio is more successful than fuzzy version of this.

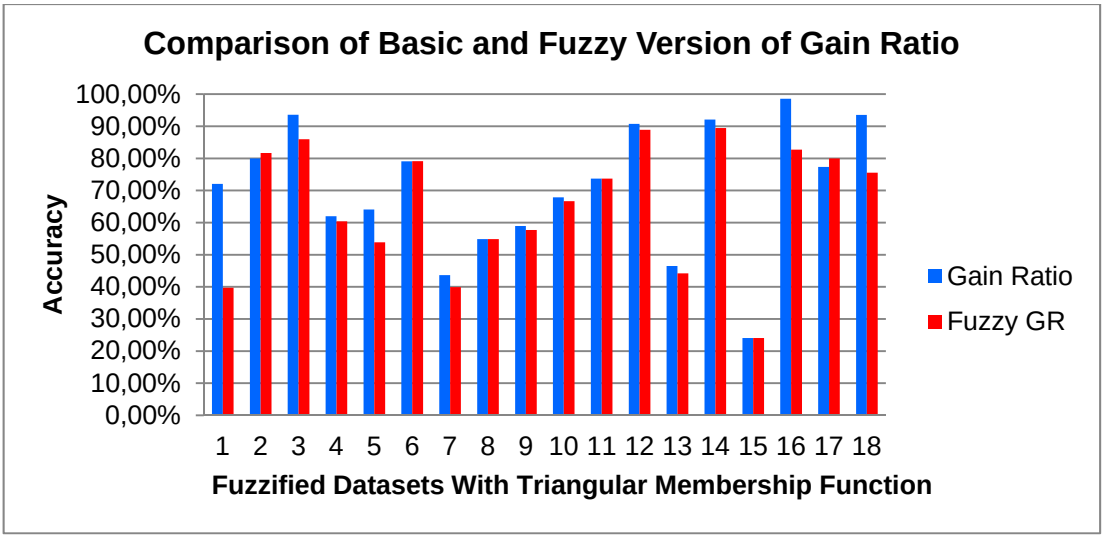

Figure 4.10. Accuracy of Basic and Fuzzy Version of Gain Ratio

Figure 4.11 shows experimental results of classification accuracy of basic and fuzzy version of gini index. Triangular membership function is used to fuzzified datasets in this experiments presented in this figure. Fuzzy decision tree which is built by using gini index has the best performance in 16 datasets, on the other hand fuzzy gini index has the best performance in 15 datasets. So basic gini index and fuzzy version of this have almost same performance of classification.

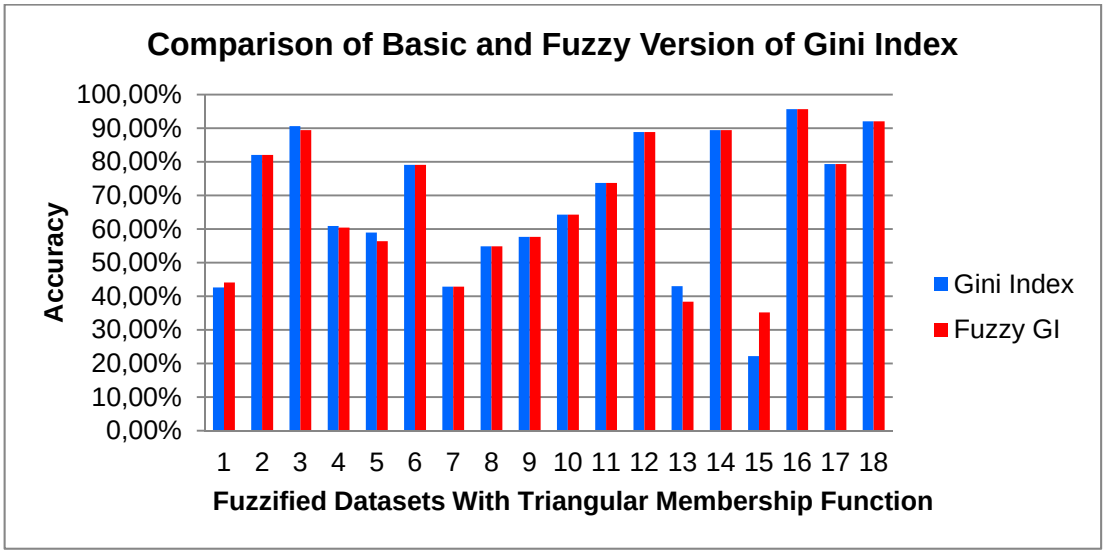

Figure 4.11. Accuracy of Basic and Fuzzy Version of Gini Index

4.6. Comparison of Triangular and Trapezoidal Fuzzy Membership Functions

In this section, the effect of triangular and trapezoidal membership functions on classification performance are compared. To fuzzify numerical values, commonly used triangular and trapezoidal membership functions are used in this thesis. For these experiments information gain is used for decision tree induction.

Accuracy of "ID3 with fuzzified data and basic splitting criteria" method which uses triangular and trapezoidal membership functions is shown in Figure 4.12. According to the results presented in Figure 4.12, "ID3 with fuzzified data and basic splitting criteria" method with information gain and triangular membership function has higher classification accuracy in 11 datasets. On the other hand, trapezoidal membership function has best accuracy value in 8 datasets. But the number of rules obtained from decision tree which is built by using trapezoidal membership function are less than triangular membership function in 15 datasets and number of rules for triangular membership function are less in 4 datasets. Number of rules learned by "ID3 with fuzzified data and basic splitting criteria" method which uses triangular and trapezoidal membership functions is shown in Figure 4.13. In other words, triangular membership function is more successful for accuracy but number of rules obtained from decision tree which is built by using triangular membership function

are more than trapezoidal membership function. So training with the trapezoidal membership function is faster than triangular membership function but triangular membership function is faster for test part and these results are presented in Figures 4.14 and 4.15.

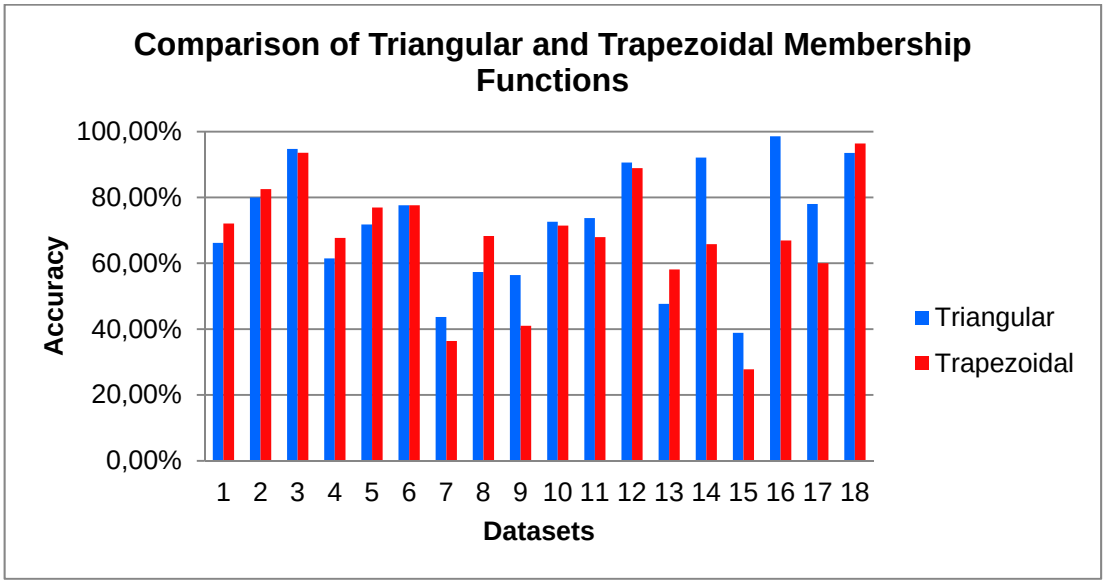

Figure 4.12. Classification Accuracy forTriangular and Trapezoidal Membership Functions

Although, trapezoidal membership function yields less number of rules, training and test times with trapezoidal membership function are greater than the triangular membership function, due to the computational cost of trapezoidal membership function.

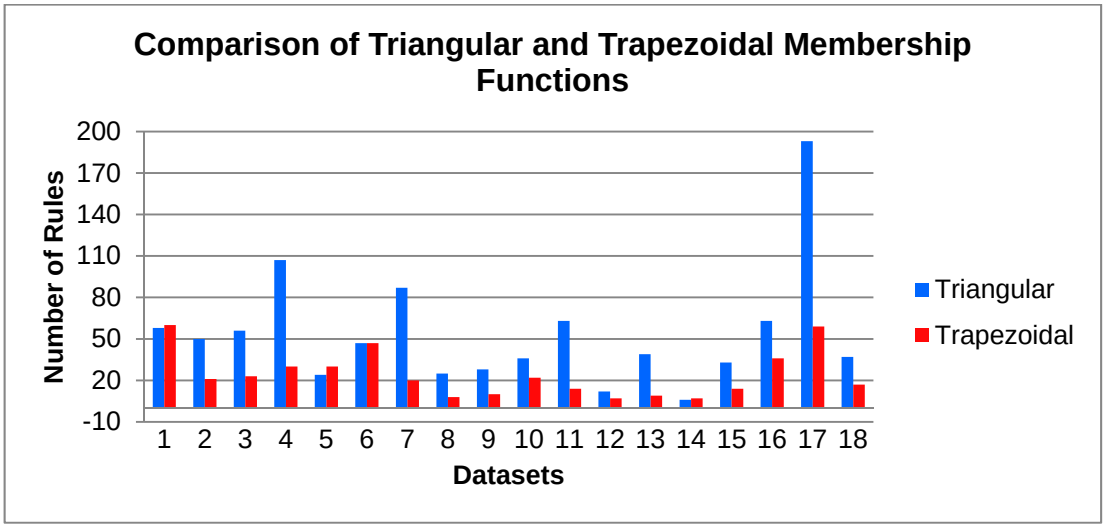

Figure 4.13. Number of Rules Obtained From Decision Tree Using Basic Information Gain for Triangular and Trapezoidal Membership Functions

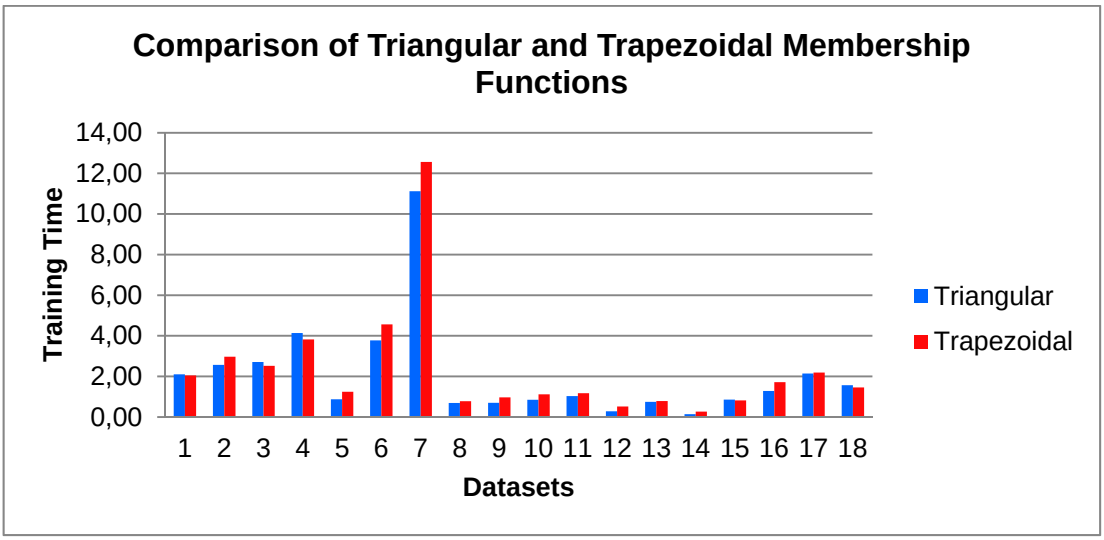

Figure 4.14. Training Time in Seconds of "ID3 with Fuzzified Data and Basic Splitting Criteria" Method for Triangular and Trapezoidal Membership Functions

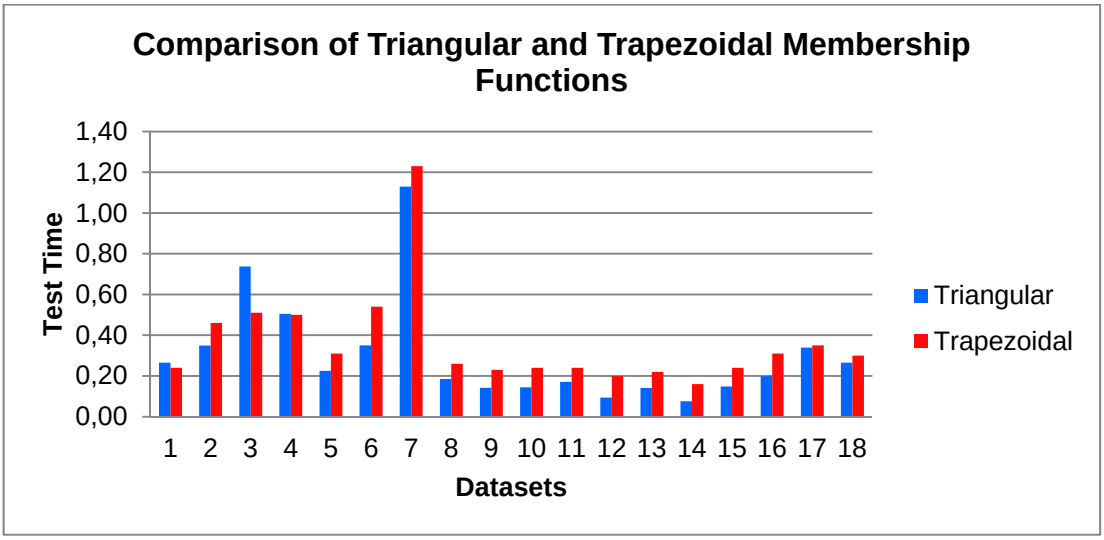

Figure 4.15. Test Time in Seconds of "ID3 with Fuzzified Data and Basic Splitting Criteria" Method for Triangular and Trapezoidal Membership Functions

4.7. Comparison of Rule Selection Methods

In this thesis, there are four different rule selection process named as "Test 1", "Test 2", "Test 3", "Test 4" and they are explained in Section 3.2.5. For decision trees which is built by using fuzzified data, if more than one rules classify the same test tuple, one of the rules is selected. Experimental results of the rule selection methods are shown in Figure 4.16, 4.17, and 4.18 for information gain, gain ratio, and gini index criteria respectively.

According to the results presented in Figure 4.16, "Test 1" has best accuracy value for 13 datasets out of 18 datasets. "Test 2" has best accuracy value for 11 datasets. "Test 3" and "Test 4" are successful for 6 datasets. So "Test 1" is more successful than other test types for all linguistic terms and information gain. Figure 4.17 shows that "Test 1" and "Test 2" has best accuracy value for 12 datasets. "Test 3" has best accuracy value for 5 datasets. "Test 4" is successful for 6 datasets. So "Test 1" and "Test 2" has the same performance and they are more successful than other test types for all linguistic terms and gain ratio. Figure 4.18 shows that "Test 1" has best accuracy value for 11 datasets. "Test 2" has best accuracy value for 10 datasets. "Test 3" has best accuracy value for 5 datasets. "Test 4" is successful for 6

datasets. So "Test 1" is more successful than other test types for all linguistic terms and gini index.

According to the results presented in Figure 4.16, 4.17, and 4.18, "Test 1" and "Test 2" have generally better performance in terms of classification accuracy.

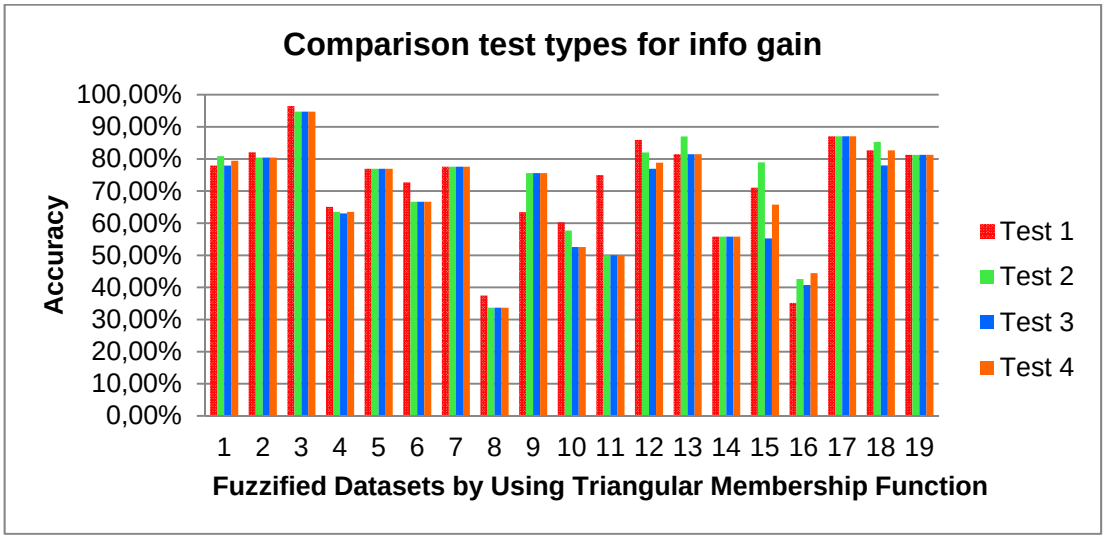

Figure 4.16. Accuracy of Test Types for All Linguistic Terms and Information Gain

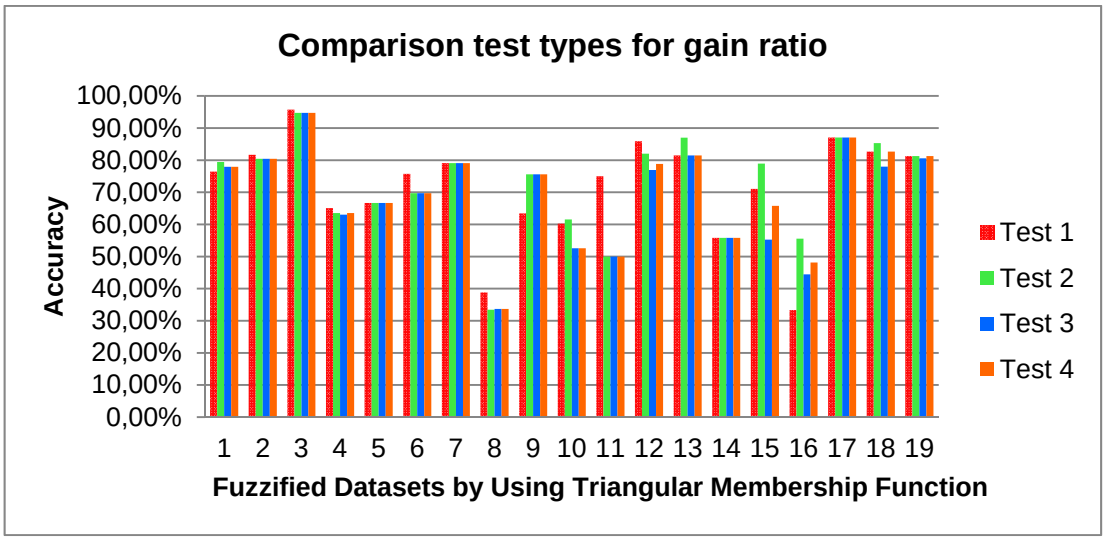

Figure 4.17. Accuracy of Test Types for All Linguistic Terms and Gain Ratio

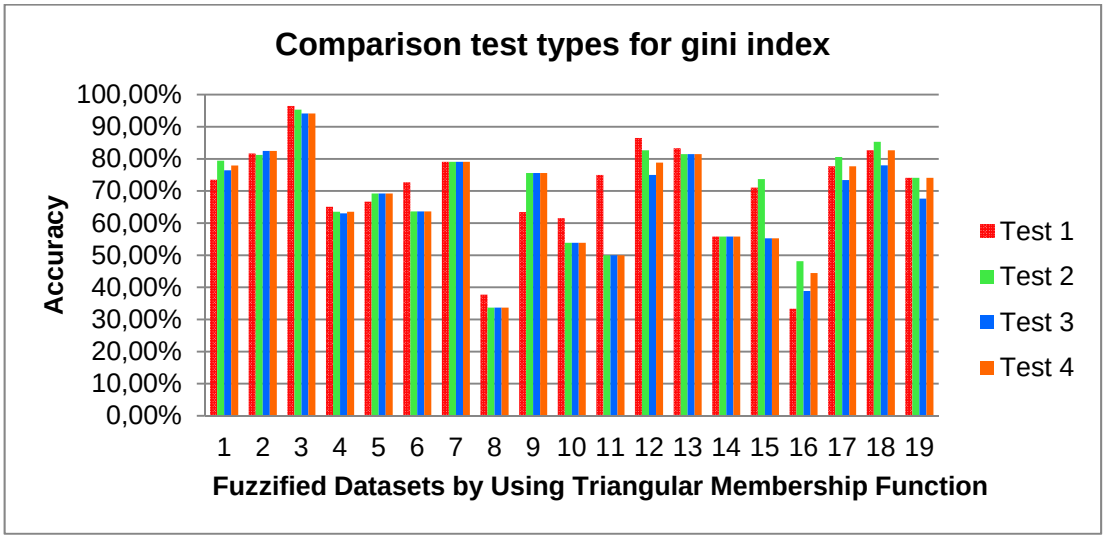

Figure 4.18. Accuracy of Test Types for All Linguistic Terms and Gini Index

4.8. Experimental Results of the Weka Classification Tool

In this thesis, J48 (Witten, Frank, 2005) and ID3 algorithm in Weka Data Mining Tool is used to test the same datasets and to compare with our methods. J48 is the Weka implementation of C4.5 decision tree algorithm, and it has pruned and unpruned versions. ID3 algorithm in Weka is an unpruned decision tree based on ID3 algorithm. Same datasets are classified with J48 and ID3 algorithm and experimental results in terms of accuracy and number of rules for J48 and ID3 algorithm are shown in Table 4.36 and Table 4.37.

Table 4.36. Accuracy of J48 Algorithm in Weka Classification Tool

Datasets	(%) Accuracy	# of Rules
Heart Statlog	77.94	28
Mammographic Masses	78.75	29
Breast Cancer	97.08	19
Diabetes	68.23	8
Hepatitis	84.62	15
Spect Heart	83.58	21
Yeast	56.60	181
Vertebral Column 2C	41.03	2
Vertebral Column 3C	82.05	11
Ecoli	84.53	11
Balance Scale	75.00	50
Thyroid	94.44	8
LD Bupa	65.12	26
Iris	94.74	5
Glass	61.11	20
Monk1	100	8
Monk2	88.67	36
Monk3	99.28	5

According to the results presented in Table 4.36, number of rules obtained from J48 in Weka are less than other decision tree methods presented in this thesis. If linguistic term having the maximum membership value is used, J48 has higher classification accuracy than “ID3 with fuzzy data and basic splitting criterion” and “ID3 with fuzzy data and fuzzy splitting criterion” methods. If all linguistic terms are used in the “ID3 with fuzzy data and basic splitting criterion” and “ID3 with fuzzy data and fuzzy splitting criterion” methods, higher classification accuracy is obtained for some datasets. For example, classification accuracy of the “Heart Statlog” dataset is 77.94% with J48. According to our experimental results which are obtained from fuzzy decision tree with basic splitting criteria and all linguistic terms shown in Table 4.11, classification accuracy of the “Heart Statlog” dataset is 80.88% when information gain, triangular membership function, and “Test 2” are used and is 79.41% when information gain, trapezoidal membership function and “Test 1” are employed. Classification accuracy of the “Balance Scale” dataset is 75.50% with J48. According to our experimental results obtained from fuzzy decision tree with

fuzzy splitting criteria and all linguistic terms shown in Table 4.21, classification accuracy of the "Balance Scale" dataset is 86.54% for fuzzy information gain, both triangular and trapezoidal membership functions, and "Test 1".

Table 4.37. Accuracy of ID3 Algorithm in Weka Classification Tool

	Triangular Membership Function		**Trapezoidal Membership Function**	
Datasets	**Accuracy**	**# of Rules**	**Accuracy**	**# of Rules**
Heart Statlog	72.06	89	73.53	119
Mammographic M.	80.42	63	82.92	41
Breast Cancer	95.32	71	93.57	45
Diabetes	63.54	167	67.71	57
Hepatitis	71.80	39	76.92	59
Spect Heart	76.12	93	76.12	93
Yeast	43.40	143	38.01	39
Vertebral Col 2C	63.42	39	71.95	15
Vertebral Col 3C	58.97	39	56.41	19
Ecoli	63.10	29	73.81	43
Balance Scale	74.36	63	67.95	27
Thyroid	92.59	15	88.89	13
LD Bupa	53.49	55	58.14	17
Iris	92.11	9	65.79	13
Glass	53.70	51	46.30	27
Monk1	98.56	67	70.50	63
Monk2	78.00	253	61.33	111
Monk3	94.96	43	96.40	33

According to the results presented in Table 4.37, accuracy results of our ID3 with fuzzified data by using triangular membership function, basic splitting criteria and single linguistic term are better than results of ID3 in Weka for 9 datasets, and our method with trapezoidal membership function has best accuracy value for 10 datasets. Accuracy values of ID3 with fuzzified data by using triangular membership function, fuzzy splitting criteria and single linguistic term are better than results of Weka for 4 datasets and accuracy values of same method with trapezoidal membership function are better than results of Weka for 8 datasets. If single linguistic term is used, ID3 with fuzzified data and basic splitting criteria method and ID3 in Weka have almost the same rate of success. Number of rules learned by basic

splitting criteria and single linguistic term are less than number of rules learned by Weka. But ID3 in Weka is more successful than ID3 with fuzzified data and fuzzy splitting criteria method. Number of rules learned by ID3 in Weka are better than fuzzy splitting criteria and triangular membership function but number of rules learned by fuzzy splitting criteria and trapezoidal membership function are lower than ID3 in Weka. If all linguistic terms are used, our all methods are more successful for accuracy than ID3 method in Weka and number of rules learned by Weka are better than our methods. Since C4.5 (J48) is more successful algorithm than ID3 in general, J48 has better performance than ID3 as expected. However for Heart Statlog and Balance Scale datasets, our fuzzy ID3 algorithms have better classification accuracy than J48 algorithm. If all linguistic terms are used, our fuzzy ID3 algorithms yield better classification accuracy than ID3 algorithm in Weka.

5. CONCLUSIONS

In this thesis we have developed classical and fuzzy decision tree based on ID3 algorithm which uses fuzzified data which are obtained by using triangular or trapezoidal membership functions; basic and fuzzified splitting criteria such as fuzzy information gain, fuzzy gain ratio, and fuzzy gini index, and compared with each other. Also, we have used two different set of fuzzified data which are named as single linguistic term having the maximum membership value if an element is a member of more than one fuzzy set and all linguistic terms that have greater than zero membership for an element to test of the effect the linguistic terms. Experimental results show that applying ID3 decision tree algorithm to fuzzified data with basic and fuzzified splitting criteria is more successful than classical decision tree. Fuzzy decision tree using fuzzified splitting criteria does not much effect on the classification performance. Number of rules learned by fuzzified splitting criteria are greater and it takes long time in seconds for train and test phases than basic splitting criteria. So, fuzzy decision tree using fuzzified splitting criteria needs to be optimized. According to the results of the fuzzy decision tree using basic splitting criteria, information gain is more successful and has less number of rules. But there is no remarkable difference in the training and test time of the splitting criteria. According to the classification accuracy, number of rules learned, training and test time in seconds, fuzzy gini index is more successful than other fuzzified splitting criteria.

Fuzzy decision tree using all lingustic terms have better classification performance with respect to using single linguistic term. But all linguistic terms have disadvantage about the number of rules. Since number of rules for all linguistic terms are more than the number of rules obtained by using single linguistic terms, decision tree with single linguistic term takes less time for train and test phases. When all linguistic terms are used, a lot of computations need to be done to learn a decision tree. So, decision tree induction with all linguistic terms for both basic and fuzzified splitting criteria, training and test part takes longer time than single linguistic term.

In addition, we examined performance of triangular and trapezoidal membership functions. According to the experimental results, number of rules learned by decision tree using trapezoidal membership function is less than triangular membership function. So training with the trapezoidal membership function is faster than triangular membership function but triangular membership function is faster for test part. But triangular membership function is more successful for accuracy than trapezoidal membership function.

In this thesis, there are four different rule selection process named as “Test 1”, “Test 2”, “Test 3”, and “Test 4”. “Test 1” is more successful than other test types for all linguistic terms and information gain.“Test 1” and “Test 2” has the same performance and they are more successful than other test types for all linguistic terms and gain ratio. “Test 1” is more successful than other test types for all linguistic terms and gini index. So “Test 1” is generally more successful rule selection process.

Also, we have compared our results with results of Weka Classification Tool for the same datasets. According to the results of J48 algorithm, Weka has higher classification accuracy than our fuzzy decision tree with basic and fuzzified splitting criteria for single linguistic term. For all linguistic terms, our fuzzy decision tree with basic and fuzzified splitting criteria has best classification accuracy for some datasets. According to the results of ID3 algorithm in Weka, if single linguistic term is used, our fuzzy decision tree based on ID3 using basic splitting criteria and ID3 in Weka have almost the same rate of success. Number of rules learned by basic splitting criteria are less than number of rules learned by Weka. But ID3 in Weka is more successful than our fuzzy decision tree based on ID3 using fuzzified splitting criteria. Number of rules learned by ID3 in Weka are better than fuzzy splitting criteria and triangular membership function but number of rules learned by fuzzy splitting criteria and trapezoidal membership function are lower than ID3 in Weka. If all linguistic terms are used, our all methods are more successful for accuracy than ID3 method in Weka and number of rules learned by Weka are better than our methods.

REFERENCES

ABU-HALAWEH, N. M., HARRISON, R. W., 2009. Practical Fuzzy Decision Trees, Proceeding of IEEE Symposium on Computational Intelligence and Data Mining (CIDM09), pp. 211-216.

ABU-HALAWEH, N. M., HARRISON, R. W., 2010. An Improved Fuzzy Decision Tree Induction Tool, IEEE

AU, W. H., CHAN, K. C. C., WONG, A. K. C., 2006. A Fuzzy Approach to Partitioning Continuous Attributes for Classification, in IEEE Transactions On Knowledge and Data Engineering, Vol. 18.

CHANDRA, B., and VARGHESE, P. P., 2009. Fuzzifying Gini Index Based Decision Trees, Expert Systems with Applications, Vol. 36, pp. 8549-8559.

CHEN, S. M., SHIE, J. D., 2009. Fuzzy Classification Systems Based on Fuzzy Information Gain Measures, Expert Systems with Applications, pp. 4517-4522.

CHIANG, I. J., and HSU, J. Y. J., 1996. Integration of Fuzzy Classifiers with Decision Trees, in Proc. Asian Fuzzy Syst. Symp., pp. 266–271.

HAN, J., KAMBER, M., 2006. Data Mining Concepts and Techniques Second Edition, Morgan Kaufmann Publishers, San Francisco, 743p.

JANIKOW, C. Z., 1998. Fuzzy Decision Trees: Issues and Methods, IEEE Trans. On Man, Systems and Cybernetics, Vol. 28, Issue 1, pp. 1-14.

JANIKOW, C. Z., 1996. Exemplar Learning in Fuzzy Decision Trees, Proc. of 5th IEEE Int, Conf on Fuzzy Systems, New Orleans, Vol 2, pp. 1500-1505.

KOHONEN, T., 1989. Self-Organization and Associative Memory, Springer, Berlin.

KUWAJIMA, I., NOJIMA, Y., ISHIBUCHI, H., 2008. Effects of Constructing Fuzzy Discretization from Crisp Discretization for Rule-Based Classifiers, Artificial Life and Robotics, Vol 13, Issue 1, pp. 294-297.

LEE, K., LEE, J., and LEE-KWANG, H., 1999. A Fuzzy Decision Tree Induction Method for Fuzzy Data, Proc. IEEE Conf. on Fuzzy Systems, FUZZ-IEEE 99, Seoul, Vol. 1, pp. 16-25.

LEE, J. W. T., SUN, J., YANG, L. Z., 2003. A Fuzzy Matching Method of Fuzzy Decision Trees, Int. Conf. on Machine Learning and Cybernetics, Vol. 3, Issue 2, pp. 1569-1573.

LEVASHENKO, V., ZAITSEVA, E., 2012. Fuzzy Decision Trees in Medical Decision Making Support System, Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 213-219.

LI, H., LV, G., ZHANG, S., GUO, Z., 2010. Using Mutual Information for Fuzzy Decision Tree Generation, Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, pp. 327-331.

LIU, X., PEDRYCZ, W., 2005. The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic, Science Direct, Vol. 7, Issue 1, pp. 325-342.

MAHER, P. E., CLAIR, D. ST., 1993. Uncertain Reasoning in an ID3 Machine Learning Framework, Fuzzy Systems, 7-12 vol. 1.

MARSALA, C., 2009. A fuzzy decision tree based approach to characterize Medicaldata, Fuzzy Systems, pp. 1332-1337.

MARSALA, C., 2012. Gradual Fuzzy Decision Tree to Help Medical Diagnosis, IEEE International Conference on Fuzzy Systems, pp. 1-6.

MITRA, S., KONWAR, K. M., PAL, S. K., 2002. Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation, in IEEE Transactions on Systems, Vol 32, pp. 328-339.

PENG, Y., FLACH, P., 2001. Soft Discretization to Enhance the Continuous Decision Tree Induction, Integrating Aspects of Data Mining, Decision Support and Meta Learning. Christophe Giraud-Carrier, Nada Lavrac, Steve Moyle, (eds.), pp. 109–118.

QUINLAN, J. R., 1986. Induction of Decision Trees, Machine Learning, Vol. 1, pp. 81-106.

RIBERIO, P. F., 2015. Getting Started with Fuzzy Logic, https://www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzysets.htm

ROKACH, L.,MAIMON, O., 2008. Data Mining With Decision Trees Theory and Applications, World Scientific Publishing Co. Pte. Ltd., Singapore, 243p.

UCI Machine Learning Repository - http://archive.ics.uci.edu/ml/datasets

UMANO, M., OKAMOTO, H., HATONO, I., TAMURA, H., KAWACHI, F., UMEDZU, S., AND KINOSHITA, J., 1994. Fuzzy Decision Trees by Fuzzy ID3 Algorithm and Its Application to Diagnosis Systems, in Proc. 3rd IEEE Conf. on Fuzzy Systems, Orlando, Vol. 3, pp. 2113-2118.

WANG, X., BORGELT, C., 2004. Information Measures in Fuzzy Decision Trees, Proc. of the IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 85-90.

WANG, X., CHEN, B., QIAN, G., and YE, F., 2000. On the Optimization of Fuzzy Decision Trees, Fuzzy Sets Syst., vol.112, pp. 117–125.

WANG, TIEN-CHIN, LEE, HSIEN-DA, 2006. Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy, World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA, pp. 306-311.

WEKA, Data Mining Software in Java. Available at: http://www.cs.waikato.ac.nz/ml/weka

WITTEN, H. I., FRANK, E., 2005. Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, San Francisco, 558p.

YUAN, Y., SHAW, M. J., 1995. Induction of Fuzzy Decision Trees, Fuzzy Sets and Systems, Vol. 69, Issue 2, pp. 125-139.

ZADEH, L. A., 1965. Fuzzy Sets, Information and Control, Vol. 8, pp. 338-353.

BIOGRAPHY

Sena Öztürk was born in Osmaniye, in 1987. She completed her elementary education at Cebelibereket İlköğretim Okulu. She graduated from Osmaniye Anadolu Lisesi, in 2005. Then she completed university education at Department of Computer Engineering, Cukurova University in 2010. Since 2012, she has been working in a software company in Ankara.