ABSTRACT

This study analyses the rule extraction of the cardiovascular database using classification algorithm. Heart disease is one of the most common causes of death in the western world and increasing so in Cyprus.

We constructed a user-friendly data mining application based on the decision tree classification algorithm, so rules can be extracted from the decision tree models. Building the decision tree various splitting criteria methods were used. Using the database, we compared the performance of each splitting criteria method. The extraction of rules is done from the decision tree model with the highest performance.

Our goal is to try and reduce this high number of casualties, by identifying the main causes of heart attack.

RULE EXTRACTION OF CARDIOVASCULAR DATABASE USING DECISION TREES

DEMETRA HADJIPANAYI

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science at the
University of Cyprus

Recommended for Acceptance
by the Department of Computer Science
June, 2009

APPROVAL PAGE

Master of Science Thesis

RULE EXTRACTION OF CARDIOVASCULAR DATABASE USING DECISION TREES

Presented by
Demetra Hadjipanayi

Research Supervisor_____*Constantinos Pattichis*_____
Research Supervisor's Name

Committee Member_____ *Christos N. Schizas*_____
Committee Member's Name

Committee Member_____ *Chris Christodoulou* _____
Committee Member's Name

University of Cyprus
June, 2009

ACKNOWLEDGEMENTS

There are numerous persons whose assistance and support over the last two years contributed significantly to the successful completion of this work.

Foremost, I would like to sincerely thank my supervisor Constantinos Pattichis for his invaluable help in completing this work. He always found time for me in his busy schedule and aided in every step of my analysis. Mr. Minas Karaolis' contribution was also of the outmost importance. This study would not have been successful without Mr. Minas' help. Also, I would like to thank the cardiologist Dr Joseph Mantiris Senior Cardiologist at the Department of Cardiology at the Pafos General Hospital for helping us in forming our cardiovascular database.

On a personal note I would like to thank my family and friends for giving me a supporting boost whenever I needed it. Especially my parents, Nicos and Chrystalla Hadjipanayi, not one day did pass without them reminding me of the deadline of my thesis and of the importance of completing my master degree.

Last but not least, I am grateful to my fiancé, Ares Neophytou, for his persistent motivation as well as for assisting with the housework, allowing me to devote my time and energy on my thesis.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1: Introduction

1.1. Brief summary of the problem

Heart disease is one of the most common causes of death in the western world and increasing so in Cyprus [1]. Myocardial infarction, commonly known as heart attack, occurs when blood flow to a section of heart muscle becomes blocked. If the flow of blood is not restored quickly, the section of heart muscle becomes damaged from lack of oxygen and begins to die [4].

Classical symptoms of cardiac episode are sudden chest pain, shortness of breath, nausea, vomiting, palpitations, sweating, and anxiety. Women may experience fewer typical symptoms than men, most commonly shortness of breath, weakness, a feeling of indigestion, and fatigue. Approximately one quarter of all myocardial infarctions is silent, without chest pain or other symptoms. A heart attack is a medical emergency, and people experiencing chest pain are advised to alert their emergency medical services, because prompt treatment can be crucial to survival [3].

In Cyprus, cardiac patients are increasing each year. In 2007, approximately 64 thousand people visited the public hospitals in Cyprus and 14% of these patients had a disease of the circulatory system (cardiology). Roughly 8 thousand patients were diagnosed with a disease of the circulatory system. Most of these patients improved without surgery, by pharmaceutical treatment. A small percentage of the cardiac cases recovered completely, while another percentage had no change in their condition. Unfortunately, in 2007, 560 patients died of a cardiac disease, seven of whom died during a cardiovascular operation [5].

In order to try and reduce this high number of casualties we try to identify the main causes of heart attack. This is done by studying all the cardiac cases that occurred in a public hospital in Cyprus.

1.2. Data Mining introduction

Data analysis has been conducted for centuries in order to produce useful information for research and other purposes. However, data size and complexity are increasing exponentially, making it impossible to do a manual data analysis. That is why mankind started using the power of computer technology. Computer technology aids in capturing, managing and storing data. However, the captured data needs to be converted into information and subsequently to knowledge to become useful. Data mining is the process of using computers to implement methodologies and techniques, to find new and remarkable patterns from data and therefore extract knowledge [8].

The first step of the process is pre-processing of the raw data. Initially a target data set must be gathered. The target data must be big enough to hold the findings while remaining summarized enough to be mined in an acceptable timeframe. After the data set is built, it's cleaned. Noise and missing values are purged or supplemented. The clean data set is then condensed into feature vectors, n-dimensional vectors of numerical features that represent observations in a summarised fashion. The feature vectors are divided into two sets, the "training set" and the "testing set". The training set is used to "train" the data model that the data mining algorithm generates, while the testing set is used to verify the accuracy of the model.

The next step is to set up the data mining algorithm [9]. There are six types of algorithms:

- Classification – Arranges the data into predefined groups.
- Clustering – Arranges the data into non predefined groups.
- Regression – Attempts to find a function which models the data with the least error.
- Segmentation – Divides data into groups, or clusters, of items that have similar properties.
- Association rule learning – Searches for relationships between variables.
- Sequence analysis – Summarizes frequent sequences or episodes in data.

Different algorithms can perform the same task, but each algorithm produces different results (see Table 1). Mining models can predict values, produce summaries of data, and find hidden correlations. The following table gives us which algorithms to use for specific tasks [10].

Table 1: Data Mining Algorithm

Task	Type of data mining algorithms to use
Predicting a discrete attribute.	Classification, clustering
Predicting a continuous attribute.	Classification, regression
Predicting a sequence.	Sequence analysis
Finding groups of common items in transactions.	Association, classification
Finding groups of similar items.	Clustering, sequence analysis

The final step is interpretation of the results. The model that is produced by the data mining algorithm is evaluated by applying it to the "testing set" data. The resulting output is compared to the desired output. If the resulting output do not meet the desired, then it is necessary to reevaluate and change the preprocessing and data mining algorithm. If the output does meet the desired standards, then we use the model to generate knowledge.

Many businesses are looking for a way to make profitable and useful decisions based on observation data. In medical science researchers extract information from the massive raw data that is collected from hospitals and clinics. Examples of such data include drug side effects, hospital cost, genetic sequence and many more. In finance, stock market can be predicted by analyzing the history of a stock. Marketing strategies can be generated by studying the past purchasing history. In general extracted information can provide guidance to the future in any area of business.

1.3. Objectives of the study

The objective of this study is to create a data mining application based on the decision tree classification algorithm based on different splitting criteria method. We construct a user-friendly application that can generate rules extracted from the decision tree models. The following splitting criteria methods were used for building the decision trees: Information Gain, Gain Ratio, Gini Gain, Distance Measure and Likelihood Ratio Chi-Squared statistics. The algorithm uses the training data set to build a data mining model and then it uses the testing data set to validate the model.

Our training and testing data sets were collected by cardiologist Dr Joseph Mantiris, Senior Cardiologist at the Department of Cardiology at the Pafos General Hospital. Medical records were collected from each cardiac patient that includes age, sex, smoking habits, blood test results, etc. This data helps generate a mining model that can predict whether the patient will have Myocardial Infarction (MI), or Percutaneous Coronary Intervention (PCI), or Coronary Artery Bypass Surgery (CABG).

1.4. Review of thesis

Chapter 2 describes the Decision Tree algorithm and the different splitting criteria methods that can be used in the algorithm. The evaluation procedure that the application uses is also explained.

Chapter 3 describes the data that was collected by the hospital. Furthermore, we illustrate the methodology that we use in order to get a correct mining model for the cardiac patients.

Chapter 4 shows all the results of our study. The results show the decision tree and the extracted rules. From there, the most important factors are derived and interesting combination of the factors.

Chapter 5 discusses the results of the study. Based on other articles, we are verified the results of this study.

Chapter 6 is the conclusion of the study. It also outlines suggestions for future work.

Chapter 7 is the bibliography that is used in our research.

Appendix A includes the manual of the application.

Appendix B includes the full detail decision trees of the results.

Appendix C includes a comparison study with Loucia Papaconstantinou's study.

Chapter 2: Decision Tree Algorithms

2.1. Decision Tree Algorithm

Decision tree algorithm produces a decision tree as a predictive model which maps all the observations of the study and concludes the class output value (prediction). Decision tree algorithms are popular due to their simplicity and transparency [12].

The decision tree has decision nodes and leaf nodes. The decision node contains the attribute that was selected to split the tree to the next level. The leaf node contains the class output value.

The algorithm begins with the creation of the training data set. The training data set is a table of observations. Each observation holds a vector of values for a number of attributes and the class output value. Next, the training data set, attribute list and the selected splitting criteria method goes through a recursive method that builds the decision tree [6].

The decision tree algorithm is given below.

Algorithm Generate Decision Tree:

Input:

- Training data set D, which is a set of training observations and their associated class value
- Attribute list A, the set of candidate attributes
- Selected splitting criteria method

Output: A decision tree

Method:

(1) *Create a node N*

(2) *If all observations in the training data set have the same class output value C,* ***then*** *return N as a leaf node labeled with C*

(3) *If attribute list is empty,* ***then*** *return N as leaf node labeled with majority class output value in training data set*

(4) *Apply* ***selected splitting criteria method*** *to training data set in order to* ***find*** *the 'best' splitting criterion attribute*

(5) *Label node N with the splitting criterion attribute*

(6) *Remove the splitting criterion attribute from the attribute list*

(7) ***For each*** *value j in the splitting criterion attribute*

- *Let D_j be the observations in training data set satisfying attribute value j*
- *If D_j is empty (no observations),* ***then*** *attach a leaf node labeled with the majority class output value to node N*
- ***Else*** *attach the node returned by* ***Generate Decision Tree*** *(D_j, attribute list, selected splitting criteria method) to node N*

(8) ***End for***

(9) *Return node N*

2.2. Splitting Criteria Methods

The following splitting criteria methods are implemented in the application:

2.2.1. Information Gain

Information gain is based on Claude Shannon's work on information theory, which calculates the value of messages. InfoGain of an attribute A is used to select the best splitting criterion attribute. The highest InfoGain is selected to build the decision tree [11].

$$InfoGain(A) = Info(D) - Info_A(D) \quad \text{(Eq 2.1)}$$

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$
$$p_i = possibility(class\_i\_in\_dataset\_D) \quad \text{(Eq 2.2)}$$
$$m => all\_class\_values$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|Dj|}{|D|} \times Info(D_j) \quad \text{(Eq 2.3)}$$

$$|D_j| = number\_of\_observations\_with\_attribute\_value\_j\_in\_dataset\_D$$
$$|D| = total\_number\_of\_observations\_in\_dataset\_D$$
$$D_j = sub\_dataset\_of\_D\_that\_contains\_attribute\_value\_j$$
$$v => all\_attribute\_values$$

Although information gain is usually a good measure for deciding the relevance of an attribute, it is not perfect. A problem occurs when information gain is applied to attributes that can take on a large number of distinct values. When that happens, then Gain ratio is used instead.

2.2.2. Gain Ratio

Gain ratio biases the decision tree against considering attributes with a large number of distinct values. So it solves the drawback of information gain [7].

$$GainRatio(A) = \frac{InfoGain(A)}{SplitInfo_A(D)} \quad \text{(Eq 2.4)}$$

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right) \quad \text{(Eq 2.5)}$$

v => $all\_attribute\_values$

2.2.3. Gini Gain

The Gini Gain is used in CART algorithm [7].

$$GiniGain(D) = Gini(D) - \sum_{j=1}^{v} p_j \times Gini(D_j) \quad \text{(Eq 2.6)}$$

v => $all\_attribute\_values$

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \quad \text{(Eq 2.7)}$$

m => all_class_values

2.2.4. Distance Measure

Distance measure, like Gain ratio, normalizes the gini index [7].

$$DM(A) = \frac{Gini(D)}{-\sum_{j=1}^{v}\sum_{i=1}^{m} p_{ij} \times \log_2(p_{ij})} \quad \text{(Eq 2.8)}$$

v => $all\_attribute\_values$

m => all_class_values

2.2.5. Likelihood Ratio Chi-squared statistics

The Likelihood ratio Chi-squared statistics is useful for measuring the statistical significance of the information gain criterion [7].

$$G^2(A, D) = 2 \times \ln(2) \times |D| \times InfoGain(A) \qquad \text{(Eq 2.9)}$$

2.3. Pruning

In decision tree algorithm, described in chapter 2.1, you can grow each branch of the tree just deeply enough to perfectly classify the training examples. While this is sometimes a reasonable strategy, in fact it can lead to difficulties when there is noise in the data, or when the number of training examples is too small to produce a representative sample of the true target function. In either of these cases, this simple algorithm can produce trees that over-fit the training examples. Over-fitting is a significant practical difficulty for decision tree learning and many other learning methods. Therefore, pruning is implemented to avoid over-fitting [26].

There are two ways to prevent the over-fitting of a decision tree [24]:

a) Bottom-up pruning algorithm [23] has two phases: In phase one, the growth phase, a very deep tree is constructed. In phase two, the pruning phase, this tree is cut back to avoid over-fitting the training data.

b) Top-down pruning algorithm [25] has interleaved the two phases: Stopping criteria are calculated during tree growth to inhibit further construction of parts of the tree when appropriate.

We choose the bottom-up pruning algorithm using Laplace error estimation. While the decision tree is built and a leaf node is created, then the Laplace error is estimated [24, 25].

$$E(D) = \frac{N - n + k - 1}{N + k}$$
$$D => dataset$$
$$C => class\_value - majority\_class\_in\_D \qquad \text{(Eq 2.10)}$$
$$k => number\_of\_class\_values$$
$$N => number\_of\_observations\_in\_D$$
$$n => number\_of\_observations\_that\_has\_the\_class\_value\_C$$

As the algorithm returns to the root node, the error of the leaf node is passed to the father node. The father node calculates the total error of all of its children and its own error. If the father's error is less than the total error of the children, then the father node is pruned and replace by a leaf node with the majority class value. If the father's error is greater than the total error of the children, then no more pruning is done to the path and the returned error is zero.

Example:

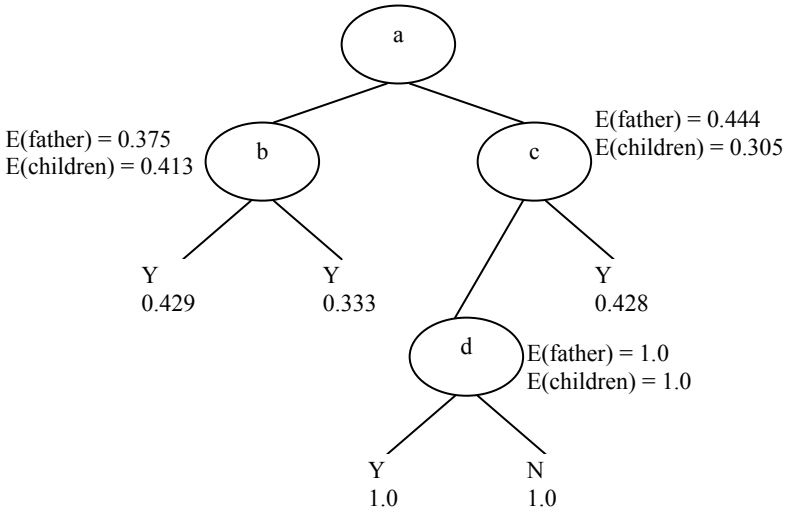


Figure 1: Pruning Example

Node b has 4 observations with the class value Y and 2 observations with the class value N. Therefore, its error is $E(b) = \frac{N - n + k - 1}{N + k} = \frac{6 - 4 + 2 - 1}{6 + 2} = \frac{3}{8} = 0.375$. Its' children has their own errors, 0.429 and 0.333. The total error of the children is $E(b's\_children) = (\frac{5}{6}) \times 0.429 + (\frac{1}{6}) \times 0.333 = 0.413$. The E(b) is less than E(b's children), therefore node b is pruned and replaced by a leaf node with the majority class value.

Node d has 1 observation with the class value Y and 1 observation with the class value N. Therefore, its error is $E(d) = \frac{N - n + k - 1}{N + k} = \frac{2 - 1 + 2 - 1}{2 + 2} = \frac{4}{4} = 1.0$. Its' children has their own errors, 1.0 and 1.0. The total error of the children is $E(d's\_children) = (\frac{1}{2}) \times 1.0 + (\frac{1}{2}) \times 1.0 = 1.0$. Since the E(d) and E(d's children) are equal, then there is no pruning. Therefore in the same path, no more pruning will be made and node d returns zero as error estimation to node c.

Node c has 4 observations with the class value Y and 3 observations with the class value N. Therefore, its error is $E(c) = \frac{N - n + k - 1}{N + k} = \frac{7 - 4 + 2 - 1}{7 + 2} = \frac{3}{9} = 0.444$. Its' children has their own errors, 0.0 and 0.428. The total error of the children is $E(c's\_children) = (\frac{2}{7}) \times 0.0 + (\frac{5}{7}) \times 0.428 = 0.305$. E(c) has more error than its children, so the nodes are not pruned.

2.4. Rule Extraction

Once the decision tree is built, then the association rules are extracted. For each path of the tree until a lead node is an association rule.

For example,

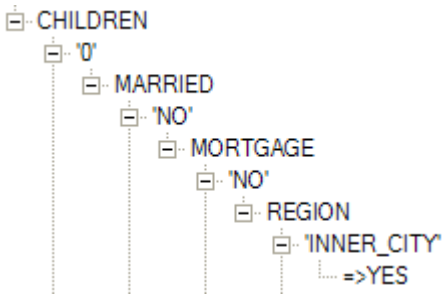


Figure 2: Decision Tree

Rule:

Children(0) AND Married(No) AND Mortgage(No) AND Region(Inner City) => CLASS(Yes)

2.5. Measures

When all the rules are extracted, then each rule calculates the statistics measures for both data sets (training and testing). (A=>B)

Support:

Support calculates the probability of the rule to be true in a given database [13]. The values vary from 0 to 1.

$$Support = P(AB)$$
$$= \frac{number\_of\_observations\_that\_satify\_A\_and\_B}{total\_number\_of\_observations} \quad \text{(Eq.2.11)}$$

Confidence:

Confidence calculates the probability that the class value is predicted when the left part of the rule (A) is satisfied. The values vary from 0 to 1 [13].

$$Confidence = P(B \mid A) = \frac{P(AB)}{P(A)} \quad \text{(Eq.2.12)}$$

Coverage:

Coverage shows the percentage of the observations that satisfy the left part of the rule (A) [13].

$$Coverage = P(A) \quad \text{(Eq.2.13)}$$

Prevalence:

Prevalence shows the percentage of the observations that satisfy the right part of the rule, which is the class output value [13].

$$\text{Pr}evalence = P(B) \quad \text{(Eq.2.14)}$$

Recall:

Recall is the probability that the left part of the rule is true, if the right part of the rule is satisfied [13].

$$\text{Re}call = P(A \mid B) = \frac{P(AB)}{P(B)} \quad \text{(Eq.2.15)}$$

Specificity:

Specificity is the probability that the right part of the rule is false, if the left part of the rule is not satisfied [13].

$$Specificity = P(\neg B \mid \neg A) = \frac{P(\neg A \neg B)}{P(\neg A)} \quad \text{(Eq.2.16)}$$

Accuracy:

Accuracy is the addition of the probability that the rule is true and the probability that the rule is false. If the value equals to one (1), then the left part of the rule and right part of the rule are dependent. Else they are independent [13].

$$Accuracy = P(AB) + P(\neg A \neg B) \quad \text{(Eq.2.17)}$$

Lift:

Lift equals to the confidence divided by the percentage of all the observations of the rule. The measure is independent from coverage. The values vary from 0 to infinity (∞). When the value is close to one (1), then the two parts of the rule are independent so the rule is not interesting. When the value equals to zero (0), then the rule is not important. Else if the $P(B|A)$ is one (1), then the rule is interesting [13].

$$Lift = \frac{P(B \mid A)}{P(B)} = \frac{P(AB)}{P(A)P(B)} \quad \text{(Eq.2.18)}$$

Leverage:

This measure is an importance measure of the rule, because it involves the confidence and the coverage. It is the percentage of the observations that cover the left part of the rule and the right part, over the result that is expected if the two parts are independent. The values vary from -1 to 1. Values close to zero (0) show a powerful independent between the two parts of the rule. Values close to one (1), point that the rule is important [13].

$$Leverage = P(B \mid A) - P(A)P(B) = \frac{P(AB)}{P(A)} - P(A)P(B) \quad \text{(Eq.2.19)}$$

Added Value [13]:

$$Added\_Value = P(B \mid A) - P(B) = \frac{P(AB)}{P(A)} - P(B) = \frac{P(AB) - P(A)P(B)}{P(A)} \quad \text{(Eq.2.20)}$$

Conviction [13]:

$$Conviction = \frac{P(A)P(\neg B)}{P(A\neg B)} \quad \text{(Eq.2.21)}$$

Odds Ratio [13]:

$$Odds\_Ratio = \frac{P(AB)P(\neg A\neg B)}{P(A\neg B)P(\neg AB)} \quad \text{(Eq.2.22)}$$

2.6. Evaluation

Once the decision tree is build and the rule are extracted, then the evaluation is conducted. In the beginning, the observations were split to two data sets, the training data set and the testing data set. The training data set is used to build the tree. Therefore the testing data set is used to evaluate our results. The percentage split is 70% - 30%.

For each observation in the testing data set, we observe the class output value. If the rule that were extracted from the decision tree, predict correctly the class output value, then the observation is correctly classified. If the rule predicts wrongly, then the observation is incorrectly classified. Else if there is no rule to predict the class output value, then the observation is unclassified.

Using the same process, the confusion matrix is created. The confusion matrix is a table that compares the expected class output value (training) with the found class output value (testing) [13].

Table 2: Confusion Matrix

		Observed Class	
Expected Class		C_1	Not C_1
	C_1	True Positive (TP)	False Positive (FP)
	Not C_1	False Negative (FN)	True Negative (TN)

Final, by using the confusion matrix, the accuracy measures for each class output value are calculated [14].

TP-Rate:

$$TP - Rate(C_1) = \frac{TP}{TP + FN} \quad \text{(Eq.2.23)}$$

Recall:

$$\text{Re}call(C_1) = \frac{TP}{TP + FN} \quad \text{(Eq.2.24)}$$

FP-Rate:

$$FP - Rate(C_1) = \frac{FP}{FP + TN} \quad \text{(Eq.2.25)}$$

Precision:

$$\text{Pr}ecision(C_1) = \frac{TP}{TP + FP} \quad \text{(Eq.2.26)}$$

F-Measure:

$$F - Measure(C_1) = \frac{2 * TP}{2 * TP + FP + FN} \quad \text{(Eq.2.27)}$$

Sensitivity:

$$Sensitivity(C_1) = \frac{TP}{TP + FP} \quad \text{(Eq.2.28)}$$

Specificity:

$$Specificity(C_1) = \frac{TN}{TN + FN} \quad \text{(Eq.2.29)}$$

2.7. Example

The training data set are observations from an optician store about contact lenses (WEKA data sample - contact-lenses.arff).

The attributes are:

1. AGE - young, prePresbyopic, presbyopic
2. SPECTACLEPRESCRIP - myope, hypermetrope
3. ASTIGMATISM - no, yes
4. TEARPRODRATE - reduced, normal

The class is the type of contact lenses that suit the customer (soft, hard, none).

Table 3: Example Data Set

No.	AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
1	young	hypermetrope	yes	normal	hard
2	young	myope	yes	normal	hard
3	prePresbyopic	myope	yes	normal	hard
4	presbyopic	myope	yes	normal	hard
5	young	hypermetrope	no	reduced	none
6	prePresbyopic	hypermetrope	no	reduced	none
7	presbyopic	hypermetrope	no	reduced	none
8	young	myope	no	reduced	none
9	prePresbyopic	myope	no	reduced	none
10	presbyopic	myope	no	reduced	none
11	presbyopic	myope	no	normal	none
12	prePresbyopic	hypermetrope	yes	normal	none
13	presbyopic	hypermetrope	yes	normal	none
14	young	hypermetrope	yes	reduced	none
15	prePresbyopic	hypermetrope	yes	reduced	none
16	presbyopic	hypermetrope	yes	reduced	none
17	young	myope	yes	reduced	none
18	prePresbyopic	myope	yes	reduced	none
19	presbyopic	myope	yes	reduced	none
20	young	hypermetrope	no	normal	soft
21	prePresbyopic	hypermetrope	no	normal	soft
22	presbyopic	hypermetrope	no	normal	soft
23	young	myope	no	normal	soft
24	prePresbyopic	myope	no	normal	soft

1st Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- False, so continue with (3)

(3) If attribute list is empty

- False, so continue with (4)

(4) Apply selected splitting criteria method to training data set in order to find the 'best' splitting criterion attribute

Table 4: 1st Iteration Splitting Criteria Measures

Attribute	Information Gain	Gain Ratio	Gini Gain	Likelihood ratio Chi-squared statistics	Distance Measure
AGE	0.03939	0.02485	0.01736	1.31076	0.00604
SPECTACLEPRESCRIP	0.03951	0.03951	0.01041	1.31456	0.00455
ASTIGMATISM	0.37700	0.37700	0.07291	12.54336	0.03741
TEARPRODRATE	**0.54879**	**0.54879**	**0.21180**	**18.25899**	**0.11917**

- The attribute with the maximum splitting criteria method is

  TEARPRODRATE

(5) Label node N with the splitting criterion attribute TEARPRODRATE

(6) Remove the splitting criterion attribute from the attribute list

(7) a) For each value j in the splitting criterion attribute - **reduced**

- Let D(**reduced)** be the observations in training data set satisfying attribute value **reduced**

Table 5: 1st Iteration D(reduction)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	no	reduced	none
prePresbyopic	hypermetrope	no	reduced	none
presbyopic	hypermetrope	no	reduced	none
young	myope	no	reduced	none
prePresbyopic	myope	no	reduced	none
presbyopic	myope	no	reduced	none
young	hypermetrope	yes	reduced	none
prePresbyopic	hypermetrope	yes	reduced	none
presbyopic	hypermetrope	yes	reduced	none
young	myope	yes	reduced	none
prePresbyopic	myope	yes	reduced	none
presbyopic	myope	yes	reduced	none

- If D(**reduced)** is empty (no observations)
  - False, continue go to **2nd Iteration**.

(7) b) For each value j in the splitting criterion attribute - **normal**

- Let D(**normal)** be the observations in training data set satisfying attribute value **normal**

Table 6: 1st Iteration D(normal)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard
young	myope	yes	normal	hard
prePresbyopic	myope	yes	normal	hard
presbyopic	myope	yes	normal	hard
presbyopic	myope	no	normal	none
prePresbyopic	hypermetrope	yes	normal	none
presbyopic	hypermetrope	yes	normal	none
prePresbyopic	myope	no	normal	soft
young	myope	no	normal	soft
young	hypermetrope	no	normal	Soft
prePresbyopic	hypermetrope	no	normal	Soft
presbyopic	hypermetrope	no	normal	Soft

- If D(**normal)** is empty (no observations)
  - o False, continue go to **3rd Iteration**.

(8) End for

2nd Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(none)**

Table 7: 2nd Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	no	reduced	none
prePresbyopic	hypermetrope	no	reduced	none
presbyopic	hypermetrope	no	reduced	none
young	myope	no	reduced	none
prePresbyopic	myope	no	reduced	none
presbyopic	myope	no	reduced	none
young	hypermetrope	yes	reduced	none
prePresbyopic	hypermetrope	yes	reduced	none
presbyopic	hypermetrope	yes	reduced	none
young	myope	yes	reduced	none
prePresbyopic	myope	yes	reduced	none
presbyopic	myope	yes	reduced	none

3rd Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- False, so continue with (3)

(3) If attribute list is empty

- False, so continue with (4)

(4) Apply selected splitting criteria method to training data set in order to find the 'best' splitting criterion attribute

Table 8: 3rd Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard
young	myope	yes	normal	hard
prePresbyopic	myope	yes	normal	hard
presbyopic	myope	yes	normal	hard
presbyopic	myope	no	normal	none
prePresbyopic	hypermetrope	yes	normal	none
presbyopic	hypermetrope	yes	normal	none
prePresbyopic	myope	no	normal	soft
young	myope	no	normal	soft
young	hypermetrope	no	normal	Soft
prePresbyopic	hypermetrope	no	normal	Soft
presbyopic	hypermetrope	no	normal	Soft

Table 9: 3rd Iteration Splitting Criteria Measures

Attribute	Information Gain	Gain Ratio	Gini Gain	Likelihood ratio Chi-squared statistics	Distance Measure
AGE	0.22125	0.13959	0.06944	3.68064	0.02379
SPECTACLEPRESCRIP	0.09543	0.09543	0.04166	1.58764	0.01694
ASTIGMATISM	**0.77042**	**0.77042**	**0.29166**	**12.81644**	**0.16347**

- The attribute with the maximum splitting criteria method is **ASTIGMATISM**

(5) Label node N with the splitting criterion attribute **ASTIGMATISM**

(6) Remove the splitting criterion attribute from the attribute list

(7) a) For each value j in the splitting criterion attribute - **yes**

- Let D(**yes)** be the observations in training data set satisfying attribute value **yes**

Table 10: 3rd Iteration D(yes)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard
young	myope	yes	normal	hard
prePresbyopic	myope	yes	normal	hard
presbyopic	myope	yes	normal	hard
prePresbyopic	hypermetrope	yes	normal	none
presbyopic	hypermetrope	yes	normal	none

- If D(**yes)** is empty (no observations)
  - False, continue go to **4th Iteration**.

(7) b) For each value j in the splitting criterion attribute - **no**

- Let D(**no)** be the observations in training data set satisfying attribute value **no**

Table 11: 3rd Iteration D(no)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	myope	no	normal	none
prePresbyopic	myope	no	normal	soft
young	myope	no	normal	soft
young	hypermetrope	no	normal	soft
prePresbyopic	hypermetrope	no	normal	soft
presbyopic	hypermetrope	no	normal	soft

- If D(**no)** is empty (no observations)
  - False, continue go to **5th Iteration**.

(8) End for

4th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- False, so continue with (3)

(3) If attribute list is empty

- False, so continue with (4)

(4) Apply selected splitting criteria method to training data set in order to find the 'best' splitting criterion attribute

Table 12: 4th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard
young	myope	yes	normal	hard
prePresbyopic	myope	yes	normal	hard
presbyopic	myope	yes	normal	hard
prePresbyopic	hypermetrope	yes	normal	none
presbyopic	hypermetrope	yes	normal	none

Table 13: 4th Iteration Splitting Criteria Measures

Attribute	Information Gain	Gain Ratio	Gini Gain	Likelihood ratio Chi-squared statistics	Distance Measure
AGE	0.25162	0.15876	0.11111	2.09299	0.04934
SPECTACLEPRESCRIP	**0.45914**	**0.45914**	**0.22222**	**3.81908**	**0.15229**

- The attribute with the maximum splitting criteria method is **SPECTACLEPRESCRIP**

(5) Label node N with the splitting criterion attribute **SPECTACLEPRESCRIP**

(6) Remove the splitting criterion attribute from the attribute list

(7) a) For each value j in the splitting criterion attribute - **myope**

- Let D(**myope)** be the observations in training data set satisfying attribute value **myope**

Table 14: 4th Iteration D(myope)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	myope	yes	normal	hard
prePresbyopic	myope	yes	normal	hard
presbyopic	myope	yes	normal	hard

- If D(**myope)** is empty (no observations)
  - o False, continue go to **6th Iteration**.

(7) b) For each value j in the splitting criterion attribute - **hypermetrope**

- Let D(**hypermetrope)** be the observations in training data set satisfying attribute value **hypermetrope**

Table 15: 4th Iteration D(hypermetrope)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard
prePresbyopic	hypermetrope	yes	normal	none
presbyopic	hypermetrope	yes	normal	none

- If D(**hypermetrope)** is empty (no observations)

- False, continue go to **7th Iteration**.

(8) End for

5th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- False, so continue with (3)

(3) If attribute list is empty

- False, so continue with (4)

(4) Apply selected splitting criteria method to training data set in order to find the 'best' splitting criterion attribute

Table 16: 5th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	myope	no	normal	none
prePresbyopic	myope	no	normal	soft
young	myope	no	normal	soft
young	hypermetrope	no	normal	soft
prePresbyopic	hypermetrope	no	normal	soft
presbyopic	hypermetrope	no	normal	soft

Table 17: 5th Iteration Splitting Criteria Measures

Attribute	Information Gain	Gain Ratio	Gini Gain	Likelihood ratio Chi-squared statistics	Distance Measure
AGE	**0.31668**	**0.19980**	**0.11111**	**2.63414**	**0.05792**
SPECTACLEPRESCRIP	0.19087	0.19087	0.05555	1.58764	0.03807

- The attribute with the maximum splitting criteria method is **AGE**

(5) Label node N with the splitting criterion attribute **AGE**

(6) Remove the splitting criterion attribute from the attribute list

(7) a) For each value j in the splitting criterion attribute - **young**

- Let D(**young)** be the observations in training data set satisfying attribute value **young**

Table 18: 5th Iteration D(young)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	myope	no	normal	soft
young	hypermetrope	no	normal	soft

- If D(**young)** is empty (no observations)
  - False, continue go to **8th Iteration**.

(7) b) For each value j in the splitting criterion attribute - **prePresbyopic**

- Let D(**prePresbyopic)** be the observations in training data set satisfying attribute value **prePresbyopic**

Table 19: 5th Iteration D(prePrebyopic)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
prePresbyopic	myope	no	normal	soft
prePresbyopic	hypermetrope	no	normal	soft

- If D(**prePresbyopic)** is empty (no observations)
  - False, continue go to **9th Iteration**.

(7) c) For each value j in the splitting criterion attribute - **presbyopic**

- Let D(**presbyopic)** be the observations in training data set satisfying attribute value **presbyopic**

Table 20: 5th Iteration D(presbyopic)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	myope	no	normal	none
presbyopic	hypermetrope	no	normal	soft

- If D(**presbyopic)** is empty (no observations)
  - False, continue go to **10th Iteration**.

(8) End for

6th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(hard)**

Table 21: 6th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	myope	yes	normal	hard
prePresbyopic	myope	yes	normal	hard
presbyopic	myope	yes	normal	hard

7th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- False, so continue with (3)

(3) If attribute list is empty

- False, so continue with (4)

(4) Apply selected splitting criteria method to training data set in order to find the 'best' splitting criterion attribute

Table 22: 7th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard
prePresbyopic	hypermetrope	yes	normal	none
presbyopic	hypermetrope	yes	normal	none

Table 23: 7th Iteration Splitting Criteria Measures

Attribute	Information Gain	Gain Ratio	Gini Gain	Likelihood ratio Chi-squared statistics	Distance Measure
AGE	0.38997	0.24604	0.44444	3.24372	0.28041

- The attribute with the maximum splitting criteria method is **AGE**

(5) Label node N with the splitting criterion attribute **AGE**

(6) Remove the splitting criterion attribute from the attribute list

(7) a) For each value j in the splitting criterion attribute - **young**

- Let D(**young)** be the observations in training data set satisfying attribute value **young**

Table 24: 7th Iteration D(young)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard

- If D(**young)** is empty (no observations)
  - False, continue go to **11th Iteration**.

(7) b) For each value j in the splitting criterion attribute - **prePresbyopic**

- Let D(**prePresbyopic)** be the observations in training data set satisfying attribute value **prePresbyopic**

Table 25: 7th Iteration D(prePresbyopic)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
prePresbyopic	hypermetrope	yes	normal	none

- If D(**prePresbyopic)** is empty (no observations)
  - False, continue go to **12th Iteration**.

(7) c) For each value j in the splitting criterion attribute - **presbyopic**

- Let D(**presbyopic)** be the observations in training data set satisfying attribute value **presbyopic**

Table 26: 7th Iteration D(presbyopic)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	hypermetrope	yes	normal	none

- If D(**presbyopic)** is empty (no observations)
  - False, continue go to **13th Iteration**.

(8) End for

8th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(soft)**

Table 27: 8th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	myope	no	normal	soft
young	hypermetrope	no	normal	soft

9th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(soft)**

Table 28: 9th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
prePresbyopic	myope	no	normal	soft
prePresbyopic	hypermetrope	no	normal	soft

10th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- False, so continue with (3)

(3) If attribute list is empty

- False, so continue with (4)

(4) Apply selected splitting criteria method to training data set in order to find the 'best' splitting criterion attribute

Table 29: 10th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	myope	no	normal	none
presbyopic	hypermetrope	no	normal	soft

Table 30: 10th Iteration Splitting Criteria Measures

Attribute	Information Gain	Gain Ratio	Gini Gain	Likelihood ratio Chi-squared statistics	Distance Measure
SPECTACLEPRESCRIP	1	1	0.5	2.77258	0.5

- The attribute with the maximum splitting criteria method is **SPECTACLEPRESCRIP**

(5) Label node N with the splitting criterion attribute **SPECTACLEPRESCRIP**

(6) Remove the splitting criterion attribute from the attribute list

(7) a) For each value j in the splitting criterion attribute - **myope**

- Let D(**myope)** be the observations in training data set satisfying attribute value **myope**

Table 31: 10th Iteration D(myope)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	myope	no	normal	none

- If D(**myope)** is empty (no observations)
  - False, continue go to **14th Iteration**.

(7) b) For each value j in the splitting criterion attribute - **hypermetrope**

- Let D(**hypermetrope)** be the observations in training data set satisfying attribute value **hypermetrope**

Table 32: 10th Iteration D(hypermetrope)

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	hypermetrope	no	normal	soft

- If D(**hypermetrope)** is empty (no observations)
  - False, continue go to **15th Iteration**.

(8) End for

11th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(hard)**

Table 33: 11th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
young	hypermetrope	yes	normal	hard

12th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(none)**

Table 34: 12th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
prePresbyopic	hypermetrope	yes	normal	none

13th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(none)**

Table 35: 13th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	hypermetrope	yes	normal	none

14th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(none)**

Table 36: 14th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	myope	no	normal	none

15th Iteration:

(1) Create a node N

(2) If all observations in the training data set have the same class output value C

- True, return N as a leaf node labelled with **CLASS(soft)**

Table 37: 15th Iteration D

AGE	SPECTACLEPRESCRIP	ASTIGMATISM	TEARPRODRATE	CLASS
presbyopic	hypermetrope	no	normal	soft

Finally, the decision tree algorithm returns the following tree:

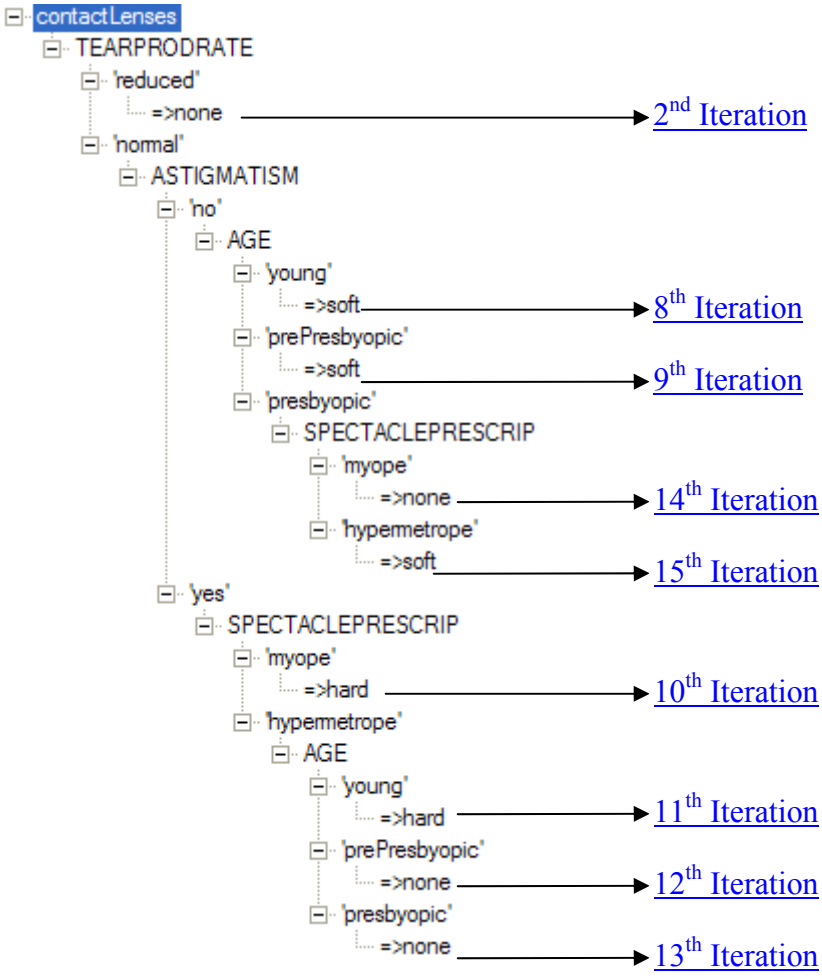


Figure 3: Example Decision Tree

The rules are extracted from the decision tree.

1. TEARPRODRATE=reduced => CLASS(none)

2. AGE=young AND ASTIGMATISM=no AND TEARPRODRATE=normal => CLASS(soft)

3. AGE=prePresbyopic AND ASTIGMATISM=no AND TEARPRODRATE=normal => CLASS(soft)

4. AGE=presbyopic AND SPECTACLEPRESCRIP=myope AND ASTIGMATISM=no AND TEARPRODRATE=normal => CLASS(none)

5. AGE=presbyopic AND SPECTACLEPRESCRIP=hypermetrope AND ASTIGMATISM=no AND TEARPRODRATE=normal => CLASS(soft)

6. SPECTACLEPRESCRIP=myope AND ASTIGMATISM=yes AND TEARPRODRATE=normal => CLASS(hard)

7. AGE=young AND SPECTACLEPRESCRIP=hypermetrope AND ASTIGMATISM=yes AND TEARPRODRATE=normal => CLASS(hard)

8. AGE=prePresbyopic AND SPECTACLEPRESCRIP=hypermetrope AND ASTIGMATISM=yes AND TEARPRODRATE=normal => CLASS(none)

9. AGE=presbyopic AND SPECTACLEPRESCRIP=hypermetrope AND ASTIGMATISM=yes AND TEARPRODRATE=normal => CLASS(none)

Training data set and testing data set are the same; therefore the statistic measures for both data sets are the same.

Table 38: Example Statistics Measures part 1

Rule No	Support	Confidence	Coverage	Prevalence	Recall	Specificity
1	0.500	1.000	0.500	0.625	0.800	0.750
2	0.083	1.000	0.083	0.208	0.400	1.000
3	0.083	1.000	0.083	0.208	0.400	1.000
4	0.042	1.000	0.042	0.625	0.067	0.000
5	0.042	1.000	0.042	0.208	0.200	1.000
6	0.125	1.000	0.125	0.167	0.750	1.000
7	0.042	1.000	0.042	0.167	0.250	1.000
8	0.042	1.000	0.042	0.625	0.067	0.000
9	0.042	1.000	0.042	0.625	0.067	0.000

Table 39: Example Statistics Measures part 2

Rule No	Accuracy	Lift	Leverage	Added value	Conviction	Odds Ratio
1	0.875	1.600	0.688	0.375	0.000	0.000
2	0.250	4.800	0.983	0.792	0.000	0.000
3	0.250	4.800	0.983	0.792	0.000	0.000
4	0.042	1.600	0.974	0.375	0.000	0.000
5	0.125	4.800	0.991	0.792	0.000	0.000
6	0.250	6.000	0.979	0.833	0.000	0.000
7	0.125	6.000	0.993	0.833	0.000	0.000
8	0.042	1.600	0.974	0.375	0.000	0.000
9	0.042	1.600	0.974	0.375	0.000	0.000

For each observation in the testing data set, we validate with the extracted rules. Each observation has an observed class output value and each rule has the expected class output value. When the testing data set is validated, then confusion matrix is created. The confusion matrix is a table that holds in the rows the expected class output values and in the columns the observed class output values (Table 2).

For the example, the confusion matrix is displayed below:

Table 40: Example Confusion Matrix

		Observed Class		
		none	soft	hard
Expected Class	none	15	0	0
	soft	0	5	0
	hard	0	0	4

For each class output value, the accuracy table is calculated using the confusion matrix table.

Table 41: Example Accuracy Measures

Class	TP-Rate	Recall	FP-Rate	Precision	F-Measure	Sensitivity	Specificity
none	1	1	0	1	1	1	1
soft	1	1	0	1	1	1	1
hard	1	1	0	1	1	1	1

Chapter 3: Methodology

3.1. Database

The database resulted from a protocol that was developed by Dr. Joseph Mantiris, Senior Cardiologist, Department of Cardiology, in Pafos General Hospital. For four years, Dr. Mantiris collected three hundred patients each year. In the database there are some fields that did not have any values or had in very few observations. Although there are some fields that are not need in the analysis, because they do not offer any new knowledge, for example the field smoking after the episode. Thus we resulted to the fields that are reported in sub-chapter 3.2, which we coded based the instructions of doctors and the International and European specifications in sub-chapter 3.3. The initial database contained 1200 cardiovascular patients that had suffered or were operated by the following:

a) Myocardial Infarction

b) Percutaneous Coronary Intervention

c) Coronary Artery Bypass Surgery

The table that follows shows a general and brief of the attributes in the database.

Table 42: Attribute Description

ATTRIBUTE	DESCRIPTION
MI Myocardial Infarction	This indicates whether the patient has suffered from a heart attack. Possible values are: 1. Yes (Y) 2. No (N)
PCI Percutaneous Coronary Intervention	Shows whether the patient was treated with a percutaneous coronary intervention surgery. Possible values are: 3. Yes (Y) 4. No (N)
CABG Coronary Artery Bypass Surgery	Shows whether the patient was treated with a coronary artery bypass surgery. Possible values are: 5. Yes (Y)

	6. No (N)
AGE	Represents the age of the patient.
GEN	Represents the gender of the patient. Possible values are: 7. Male (M) 8. Female (F)
W	Represents the weight of the patient.
H	Represents the height of the patient.
BMI Body Mass Index	Body Mass Index is calculated by dividing the height with the weight.
AS Active Smoker	Illustrates if the patient is an active smoker. Possible values are: 9. Yes (Y) 10. No (N)
PS Passive Smoker	Illustrates if the patient is a passive smoker. Possible values are: 11. Yes (Y) 12. No (N)
S-R Stop – Restart smoking	Illustrates if the patient stopped and restarted smoking. Possible values are: 13. Yes (Y) 14. No (N)
EX-SM Ex-smoker	Illustrates if the patient is an ex-smoker. Possible values are: 15. Yes (Y) 16. No (N)
POS FH Positive Family History	Presents the family history of the patient. Possible values are: 17. Yes (Y) – someone from the family has a cardiac disease 18. No (N)
HT Hypertension	Hypertension, also referred to as high blood pressure, is a medical condition in which the blood pressure is chronically elevated [16]. Shows if the patient suffers from hypertension. Possible values are: 19. Yes (Y) 20. No (N)
DM Diabetes Mellitus	Diabetes Mellitus, often referred to simply as diabetes, is a syndrome of disordered metabolism, usually due to a combination of hereditary and environmental causes, resulting in abnormally high blood sugar levels [17]. Diabetes develops due to a diminished production of insulin (in *type 1*) or resistance to its effects (in *type 2* and *gestational*) [18]. Shows if the patient suffers from diabetes. Possible values are: 21. Yes (Y) 22. No (N)

STAT Stress	Illustrates whether the patient has stress. Possible values are: 23. Yes (Y) 24. No (N)
EXER Exercise	Illustrates whether the patient exercises. Possible values are: 25. Yes (Y) 26. No (N)
HR Heart Rate	Shows the heart rate of the patient.
SBP Systolic Blood Pressure	Systolic pressure is peak pressure in the arteries, which occurs near the end of the cardiac cycle when the ventricles are contracting [12].
DBP Diastolic Blood Pressure	Diastolic pressure is minimum pressure in the arteries, which occurs near the beginning of the cardiac cycle when the ventricles are filled with blood.
TC Total Cholesterol	Cholesterol is a lipidic, waxy alcohol found in the cell membranes and transported in the blood plasma of all animals. It is an essential component of mammalian cell membranes where it is required to establish proper membrane permeability and fluidity.
HDL High Density Lipoprotein	High-density lipoprotein is one of the five major groups of lipoproteins which enable lipids like cholesterol and triglycerides to be transported within the water based blood stream. In healthy individuals, about thirty percent of blood cholesterol is carried by HDL [19].
LDL Low Density Lipoprotein	Low-density lipoprotein is a type of lipoprotein that transports cholesterol and triglycerides from the liver to peripheral tissues.
TG Triacylglyceride	Triacylglyceride is a glyceride in which the glycerol is esterified with three fatty acids [20].
GLU Glucosamine	Glucosamine is an amino sugar and a prominent precursor in the biochemical synthesis of glycosylated proteins and lipids [21].
UA Uric Acid	Uric acid is produced by xanthine oxidase from xanthine and hypoxanthine, which in turn are produced from purine. Uric acid is more toxic to tissues than either xanthine or hypoxanthine [12].
FIBR Fibrinogen	Fibrinogen is a protein produced by the liver. This protein helps stop bleeding by helping blood clots to form [22].
CRP C-Reactive Protein	C-reactive protein is a protein found in the blood in response to inflammation [12].

3.2. Study for selecting attributes

For the selecting of the attributes the following requirements took under consideration:

a) Speciality doctors gave the guidelines to what will be analyzed in this study. So the selection of the factors that should be studied is done.
b) Attributes that are enclosed by other attributes are not included, for example height and weight that is enclosed by Body Mass Index.
c) Factors that have many missing values and there was not chance of regaining the values were removed.

3.3. Pre-processing

The first step of pre-processing is to filling the missing values. Then we code the attributes so the attributes have few and abbreviated values.

3.3.1. Filling in the missing values

In the database there were many feature vectors (Chapter 1.2) that contain missing values. Firstly, we checked the written report of the patients. Some values were completed in the report, but did not go into the database. During the check some values in the database were corrected as well.

3.3.2. Coding the attributes

The coding of the attributes was based on the instructions of the doctors and the International and European specifications. Following is a table that contains the attributes with the possible values.

Table 43: Attributes and possible values

No.	Attribute Name	Possible values
Clinical factors		
1	AGE	34 – 85
2	SEX	Female, Male
3	SMBEF	Yes, No
4	SBP	20 – 150 mmHg
5	DBP	30 – 120 mmHg
6	FH	Yes, No
7	HT	Yes, No
8	DM	Yes, No
Biochemical factors		
9	TC	100 – 500 mg/dL
10	HDL	20 – 100 mg/dL
11	LDL	20 – 200 mg/dL
12	TG	100 – 300 mg/dL
13	GLU	50 – 200 mg/dL

For each attribute, the following coding was made.

Table 44: Attribute Coding

No.	Attribute Name	Coding			
Clinical factors					
1	AGE	1 (34-50)	2 (51-60)	3 (61-70)	4 (71-85)
2	SEX	F (Female)	M (Male)		
3	SMBEF	Y (Yes)	N (No)		
4	SBP	L (<90)	N (90-120)	H (>120)	
5	DBP	L (<60)	N (60-80)	H (>80)	
6	FH	Y (Yes)	N (No)		
7	HT	Y (Yes)	N (No)		
8	DM	Y (Yes)	N (No)		
Biochemical factors					
9	TC	D (<200)	N (201-240)	H (>240)	
10	HDL				
	Women	L (<50)	M (50-60)	H (>60)	
	Men	L (<40)	M (40-60)	H (>60)	
11	LDL	N (<130)	H (131-160)	D (>160)	
12	TG	N (<150)	H (151-200)	D (>200)	
13	GLU	N (<=110)	H (>110)		

3.4. Statistical analysis of the attributes

Table 45: Statistical Analysis per class value

ATTRIBUTES			CLASSES					
NAME	VALUE	TOTAL	CABG		MI		PCI	
			Y	N	Y	N	Y	N
AGE	1	36	6	30	32	4	24	12
	2	103	35	68	80	23	53	50
	3	139	59	80	94	45	59	80
	4	91	52	39	59	32	24	67
SEX	F	50	21	29	34	16	20	30
	M	319	131	188	231	88	140	179
SMBEF	N	126	72	54	37	89	49	77
	Y	243	54	189	176	67	111	132
TC	D	251	111	140	178	73	107	144
	N	90	32	58	66	24	46	44
	H	28	9	19	21	7	7	21
HDL	L	157	70	87	114	43	67	90
	M	193	74	119	140	53	86	107
	H	19	8	11	11	8	7	12
LDL	N	284	123	161	200	84	125	159
	H	59	21	38	22	37	26	33
	D	26	8	18	21	5	9	17
TG	N	201	83	118	144	57	81	120
	H	79	33	46	59	20	34	45
	D	89	36	53	62	27	45	44
GLU	N	189	77	112	139	50	88	101
	H	180	75	105	126	54	72	108
SBP	L	31	8	23	24	7	8	23
	N	106	41	65	86	20	45	61
	H	232	103	129	155	77	107	125
DBP	L	6	3	3	4	2	1	5
	N	235	97	138	172	63	113	122
	H	128	52	76	89	39	46	82
FH	N	209	84	125	153	56	75	134
	Y	160	68	92	112	48	85	75
HT	N	133	45	88	102	31	64	69
	Y	236	107	129	163	73	96	140
DM	N	246	101	145	193	53	131	115
	Y	105	51	54	72	33	29	76
TOTAL		369	152	217	265	104	160	209

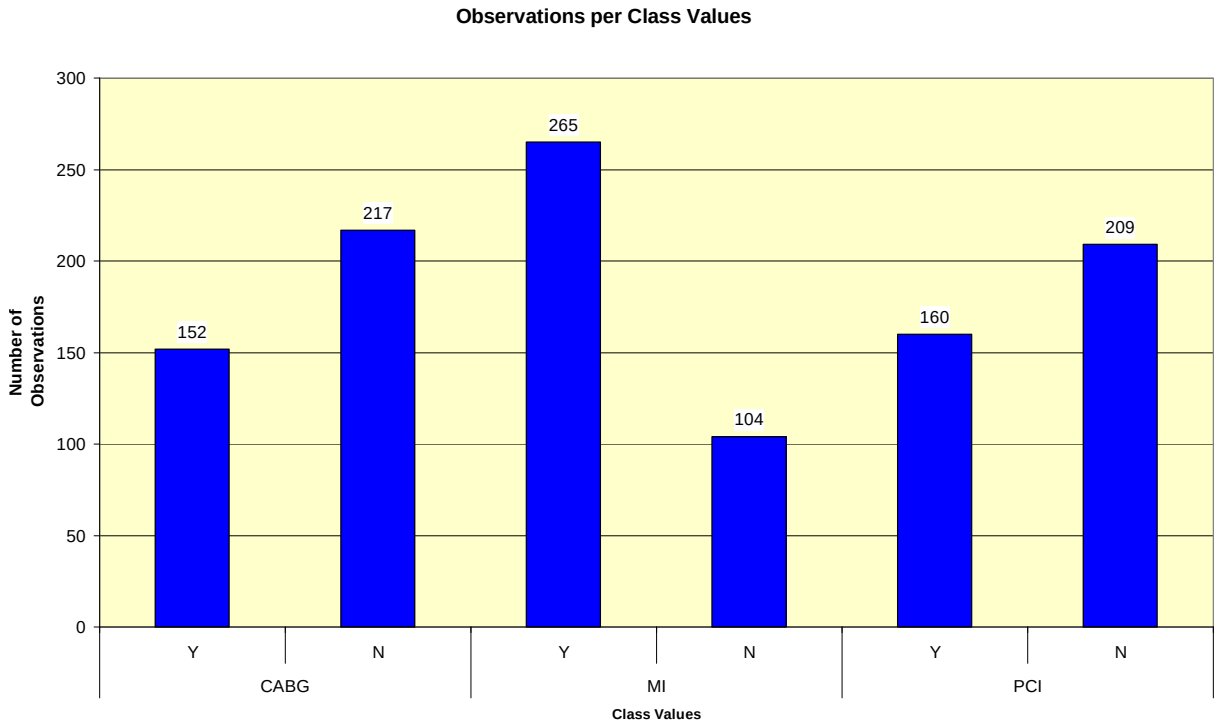


Figure 4: Chart - Observations per Class Values

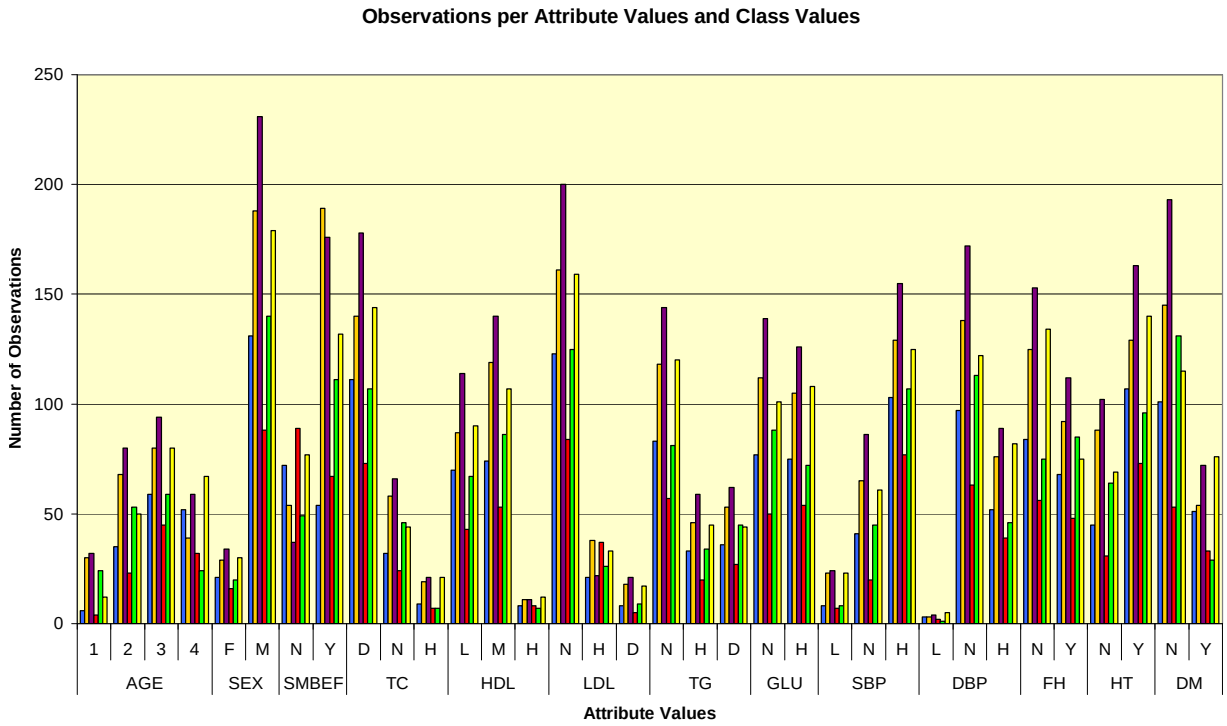


Figure 5: Chart - Observations per Attribute Values and Class Values

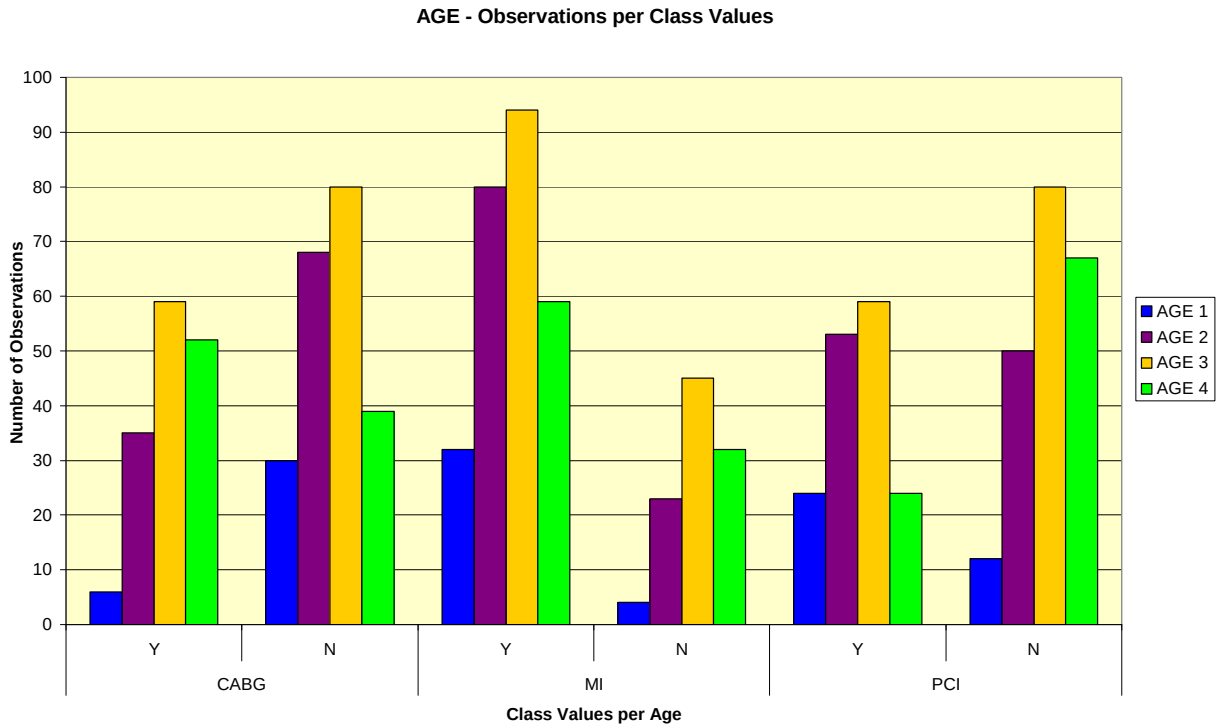


Figure 6: Chart - Observations per Class Values per Age Values

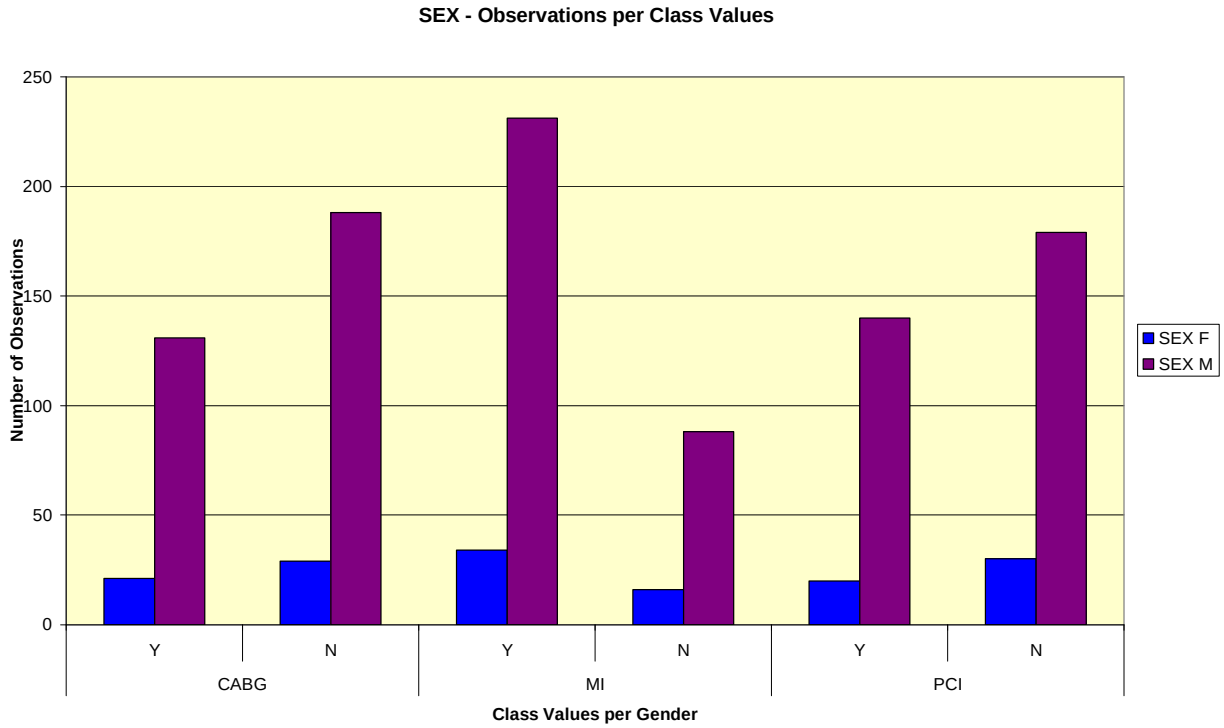


Figure 7: Chart - Observations per Class Values per Sex Values

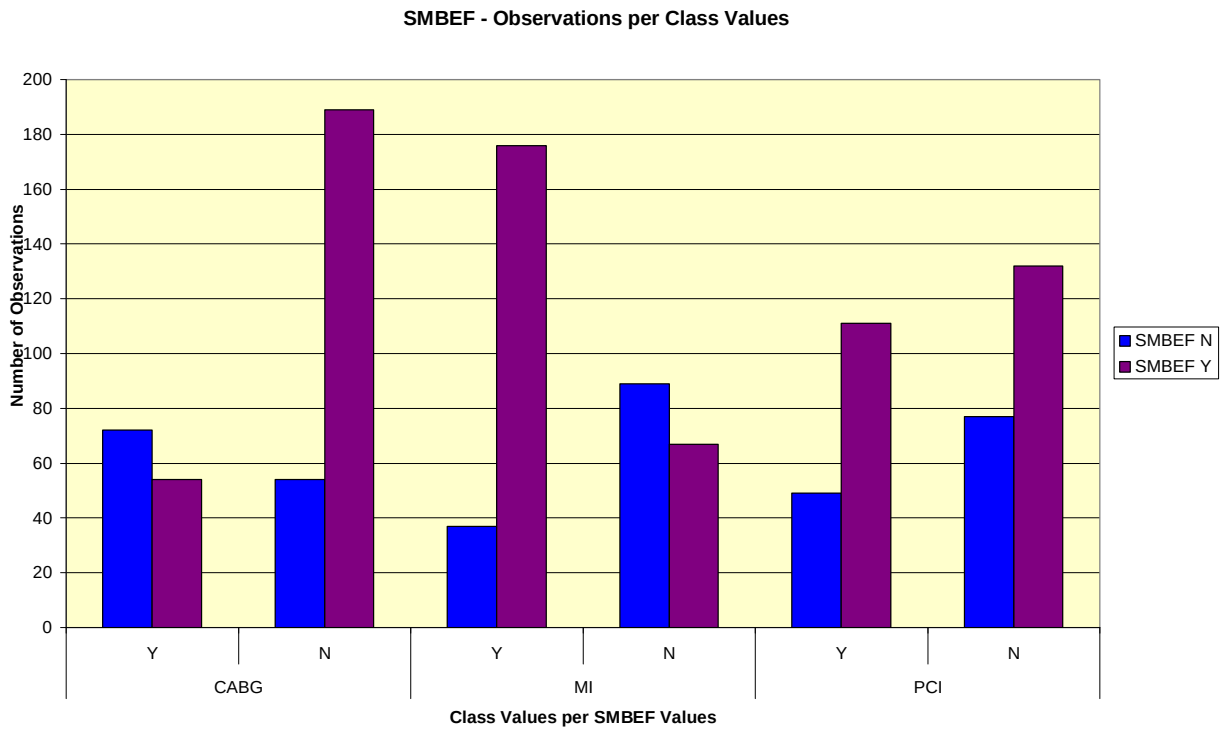


Figure 8: Chart - Observations per Class Values per SMBEF Values

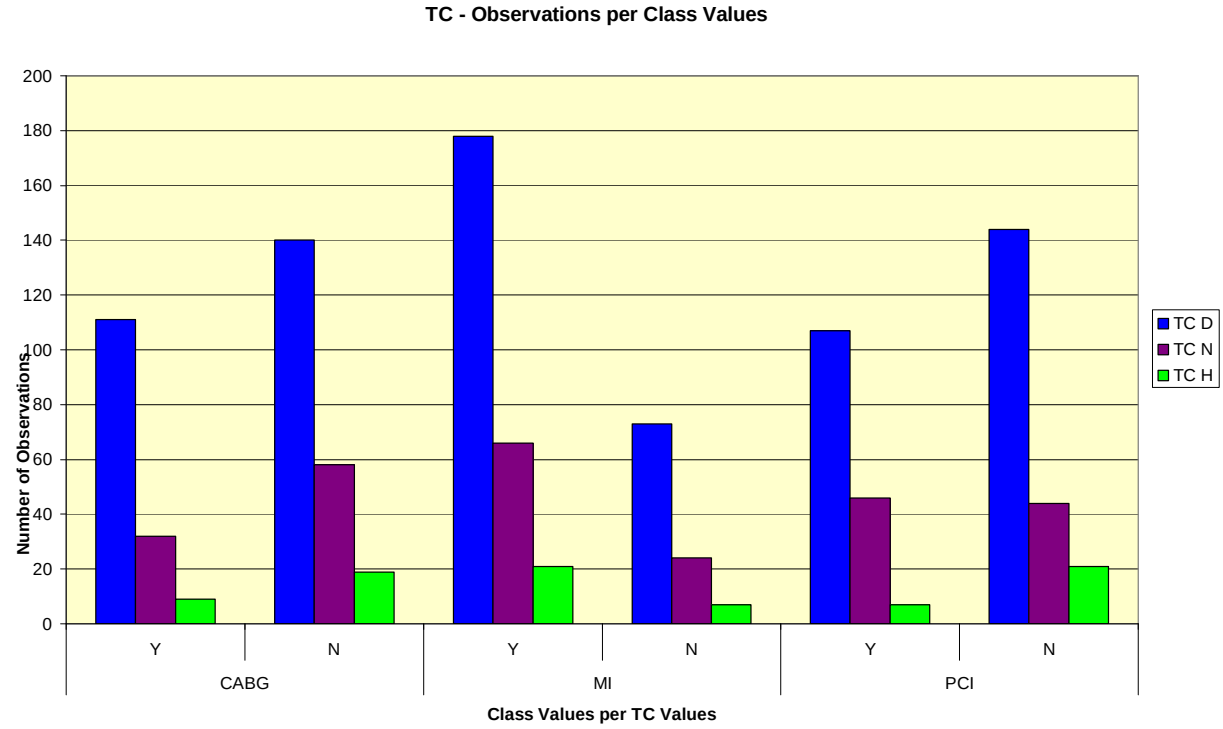


Figure 9: Chart - Observations per Class Values per TC Values

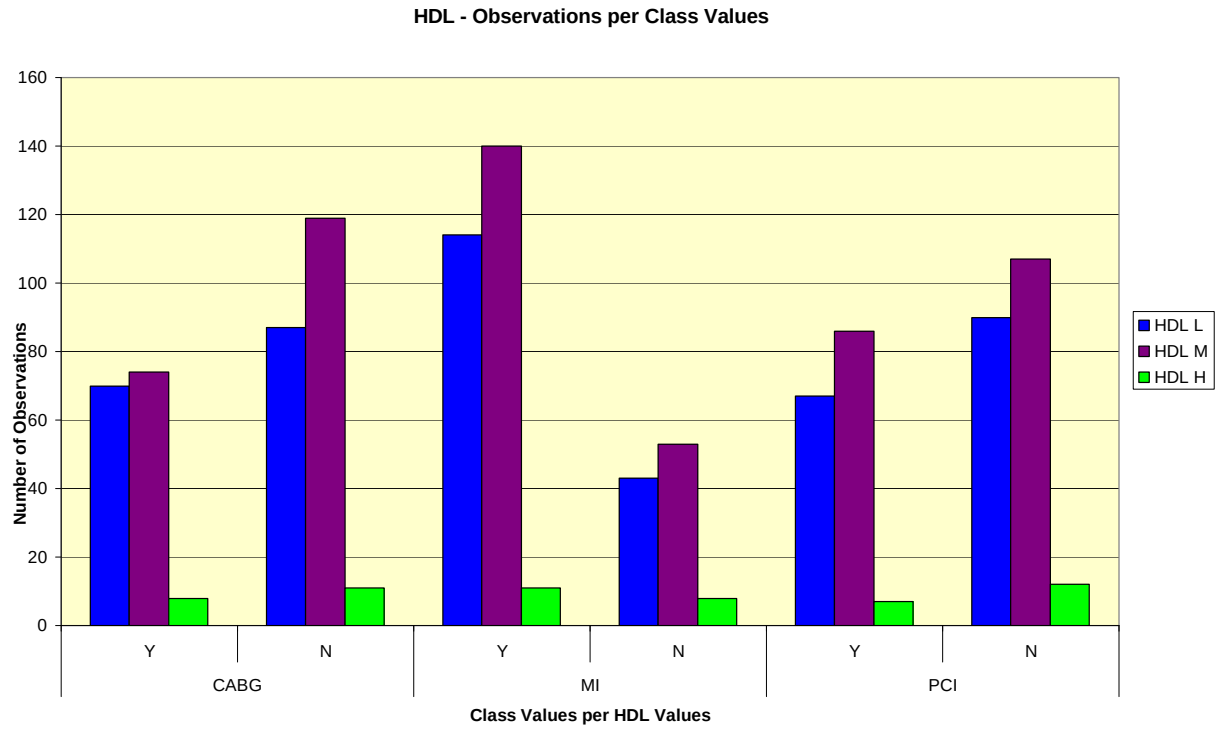


Figure 10: Chart - Observations per Class Values per HDL Values

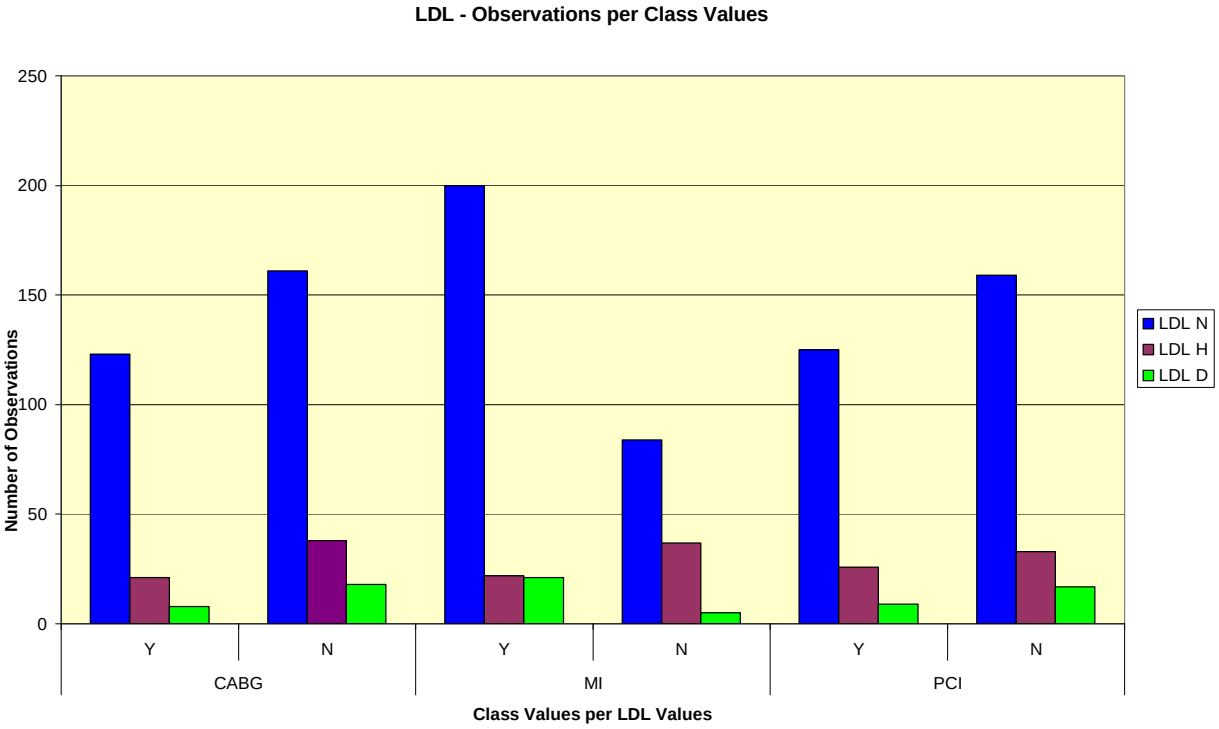


Figure 11: Chart - Observations per Class Values per LDL Values

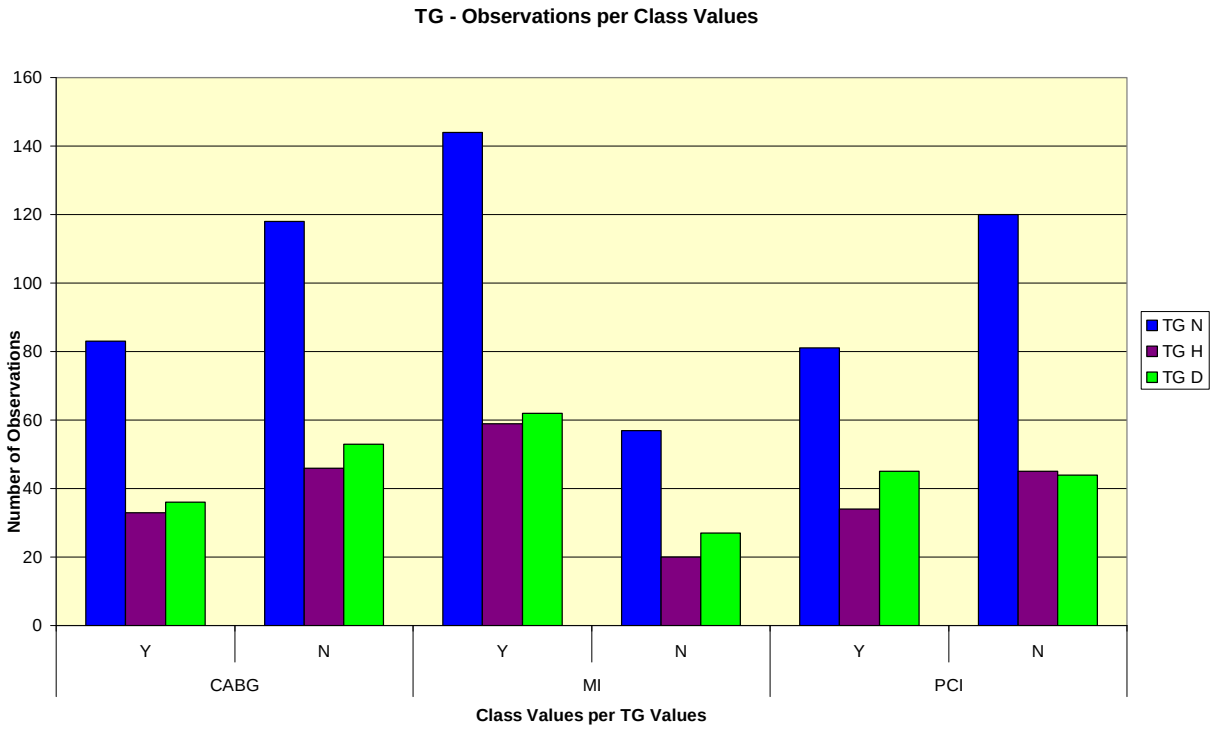


Figure 12: Chart - Observations per Class Values per TG Values

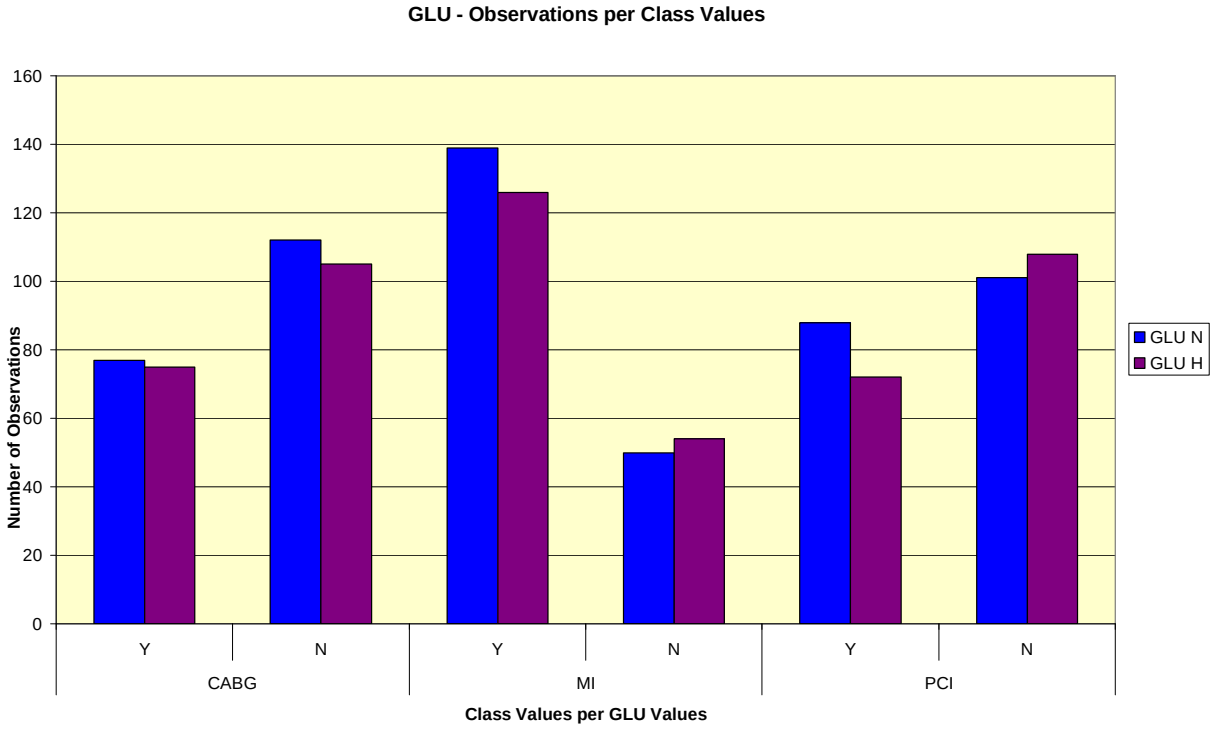


Figure 13: Chart - Observations per Class Values per GLU Values

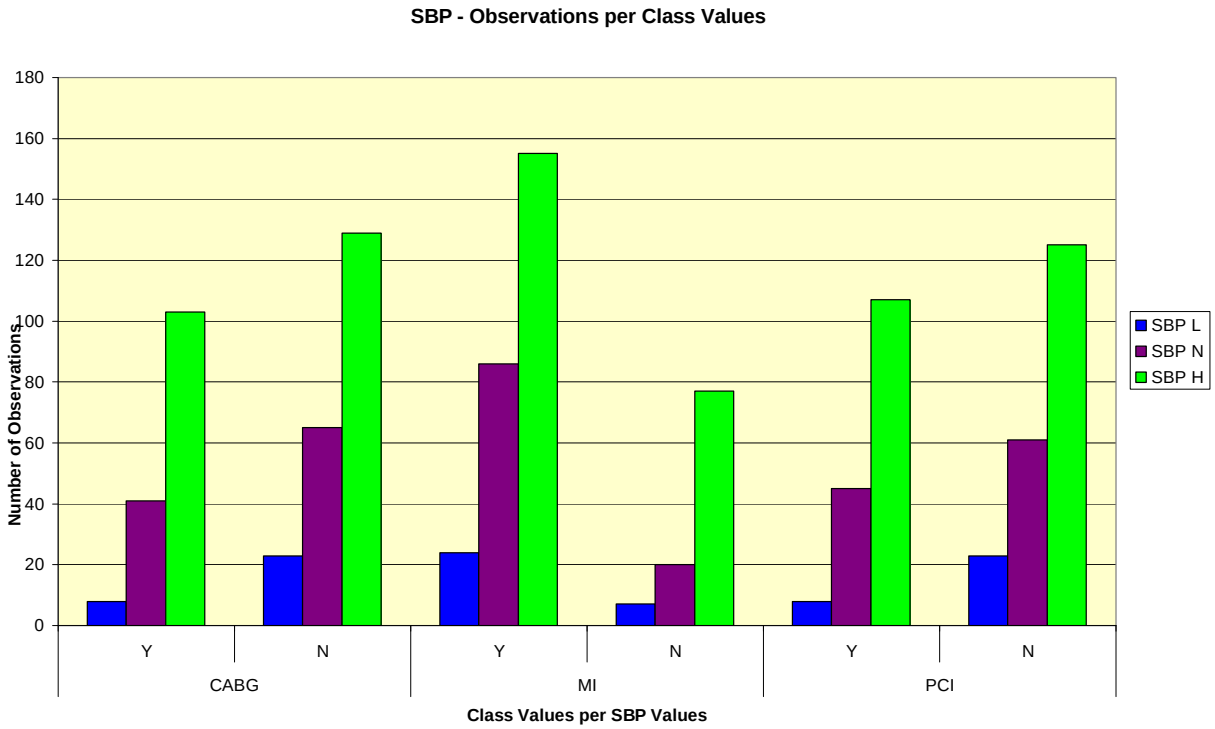


Figure 14: Chart - Observations per Class Values per SBP Values

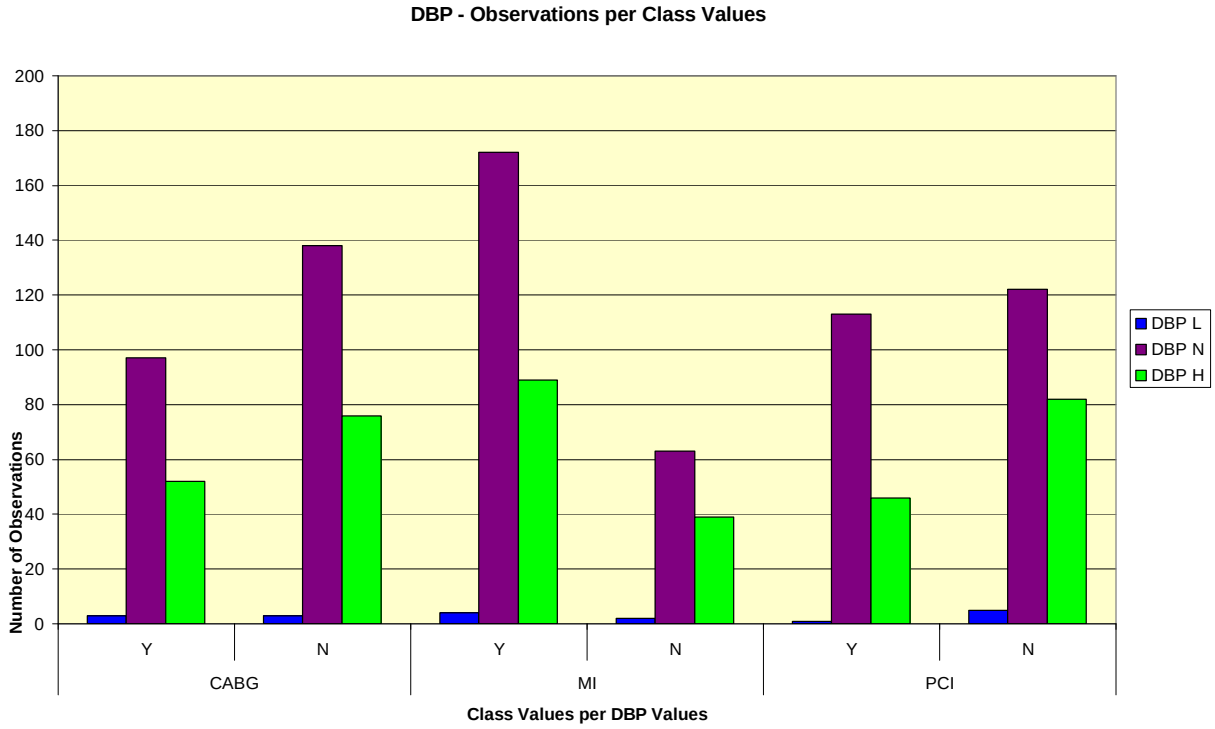


Figure 15: Chart - Observations per Class Values per DBP Values

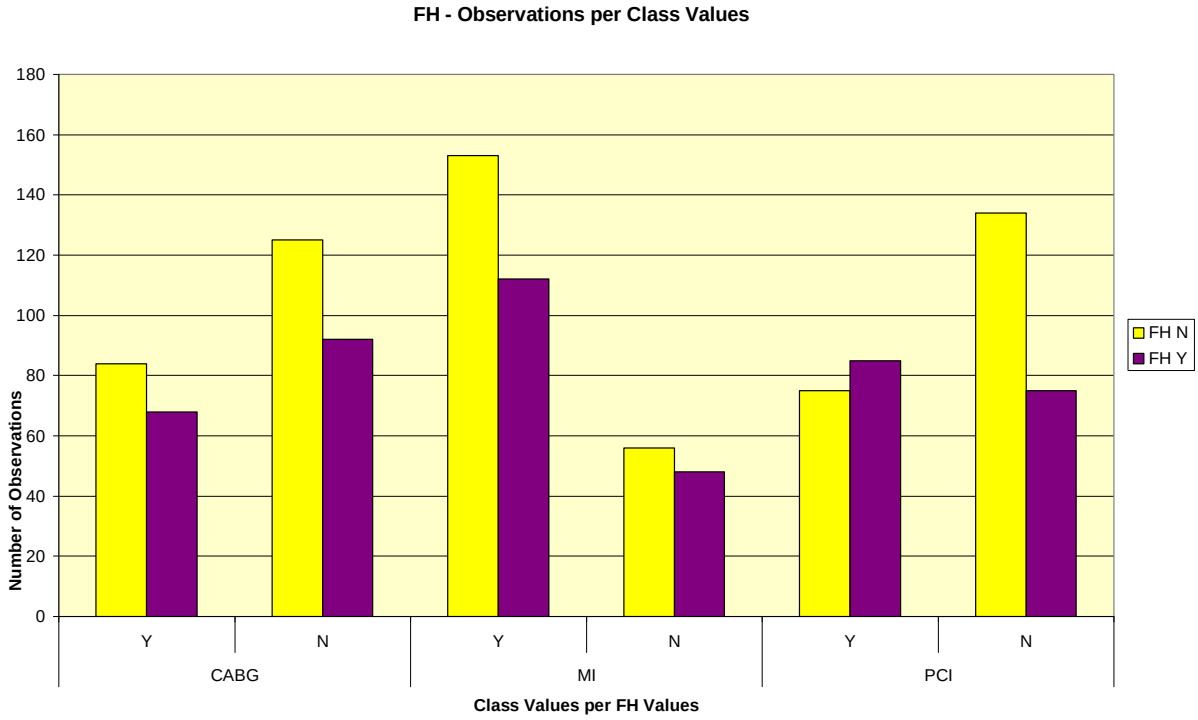


Figure 16: Chart - Observations per Class Values per FH Values

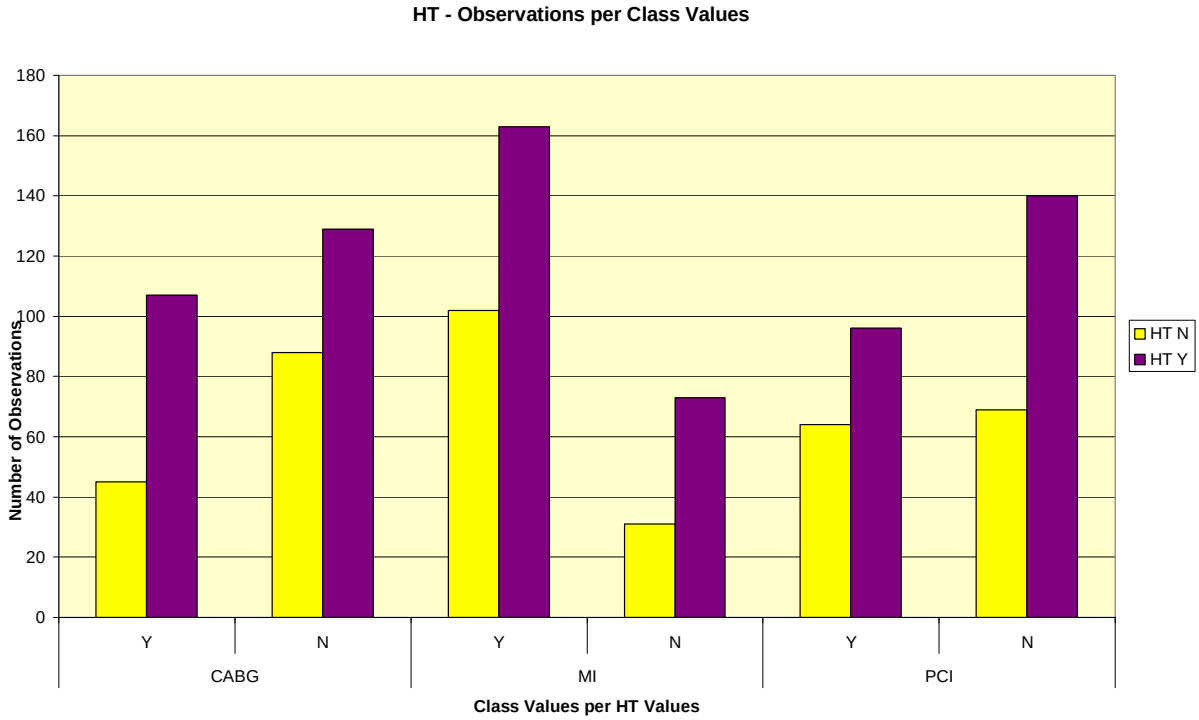


Figure 17: Chart - Observations per Class Values per HT Values

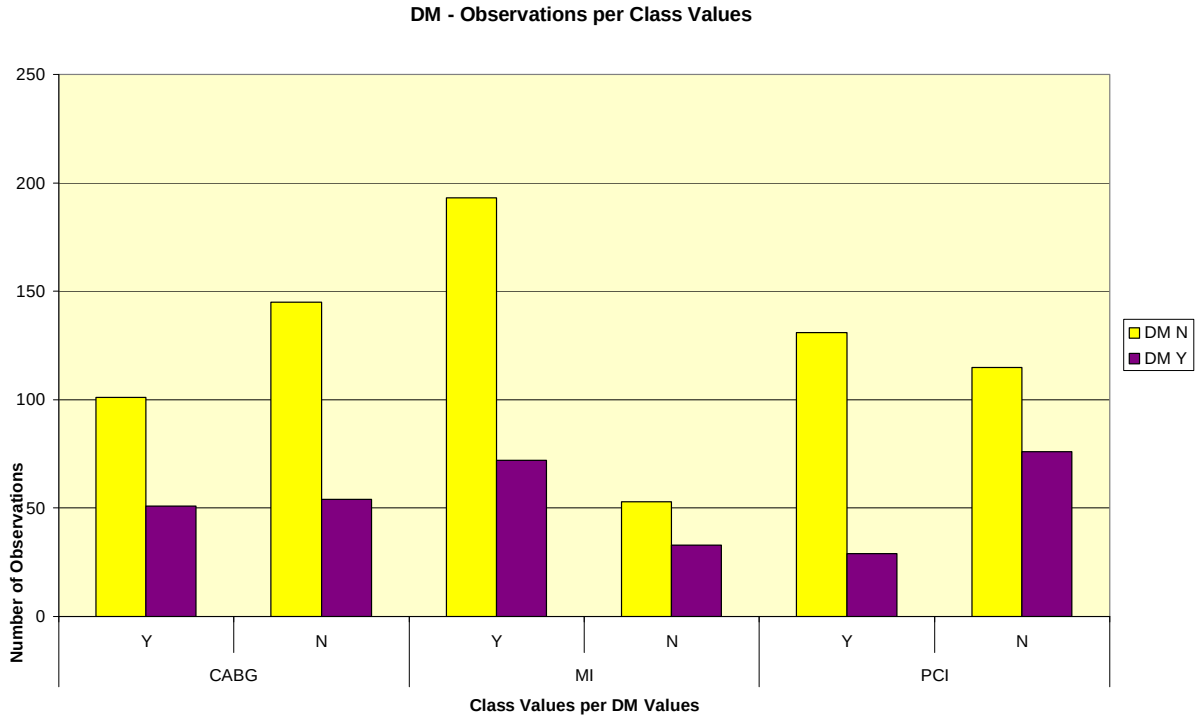


Figure 18: Chart - Observations per Class Values per DM Values

Chapter 4: Results

After preprocessing, we split the data to three data sets. All data sets have the same following attributes:

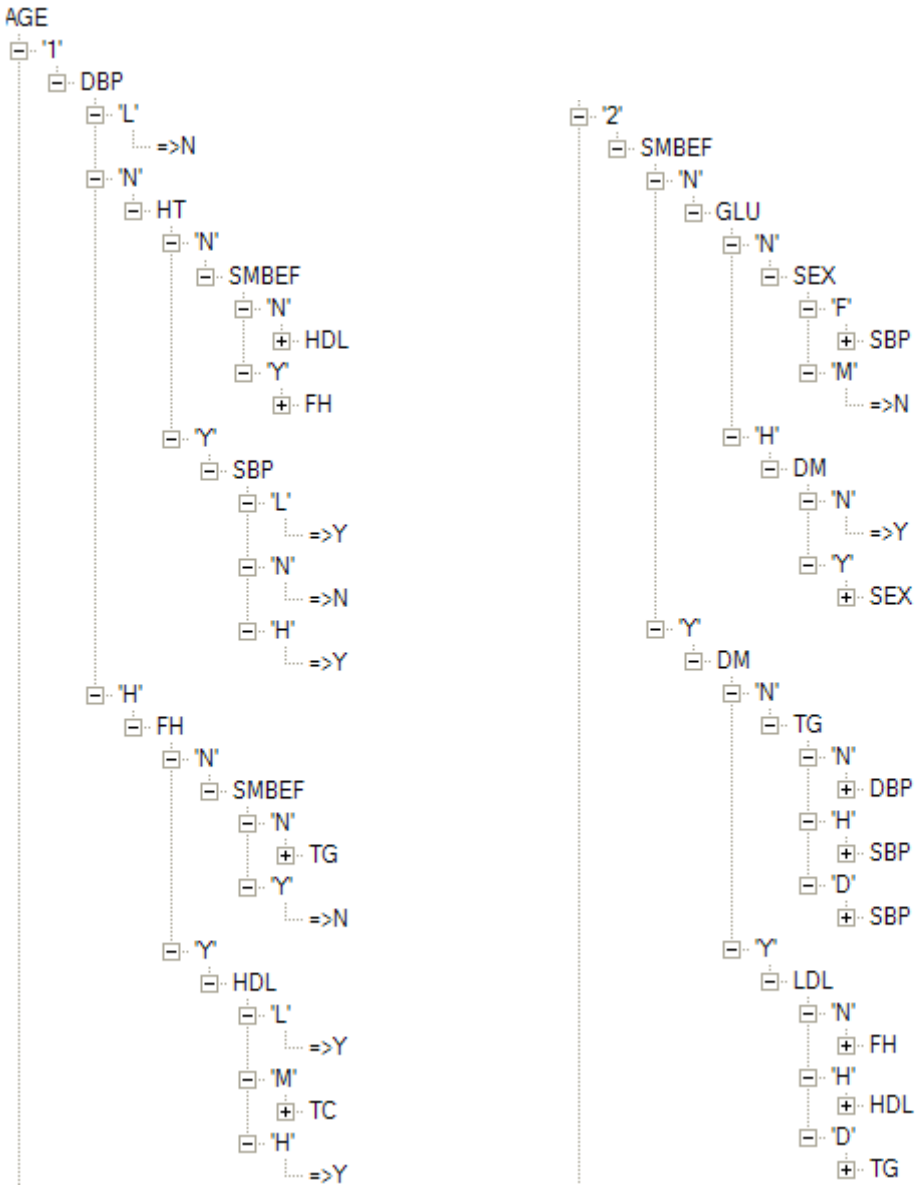
- AGE – values {1,2,3,4}
- SEX – values {F,M}
- SMBEF – values {N,Y}
- TC – values {N,D,H}
- HDL – values {L,M,H}
- LDL – values {N,H,D}
- TG – values {N,H,D}
- GLU – values {N,H}
- SBP – values {L,N,H}
- DBP – values {L,N,H}
- FH – values {N,Y}
- HT – values {N,Y}
- DM – values {N,Y}

The only difference is the output class. The following models were generated:

1. MI - Myocardial Infarction model – values {N,Y}
2. CABG - Coronary Artery Bypass Surgery model – values {N,Y}
3. PCI - Coronary Intervention model – values {N,Y}
4. Multiple classes models with MI, CABG and PCI

For each model, all the splitting criteria were studied. The decision tree with the best performance was generated. Finally, we generate the model using the best splitting criteria method with all three output classes in a single data set.

If we generate un-pruned decision trees, then the model will be very large to analyze. In the example below, an un-pruned decision tree is given. It is so big, that we could not put completely in the study, whereas Fig. 20 gives a pruned decision tree.



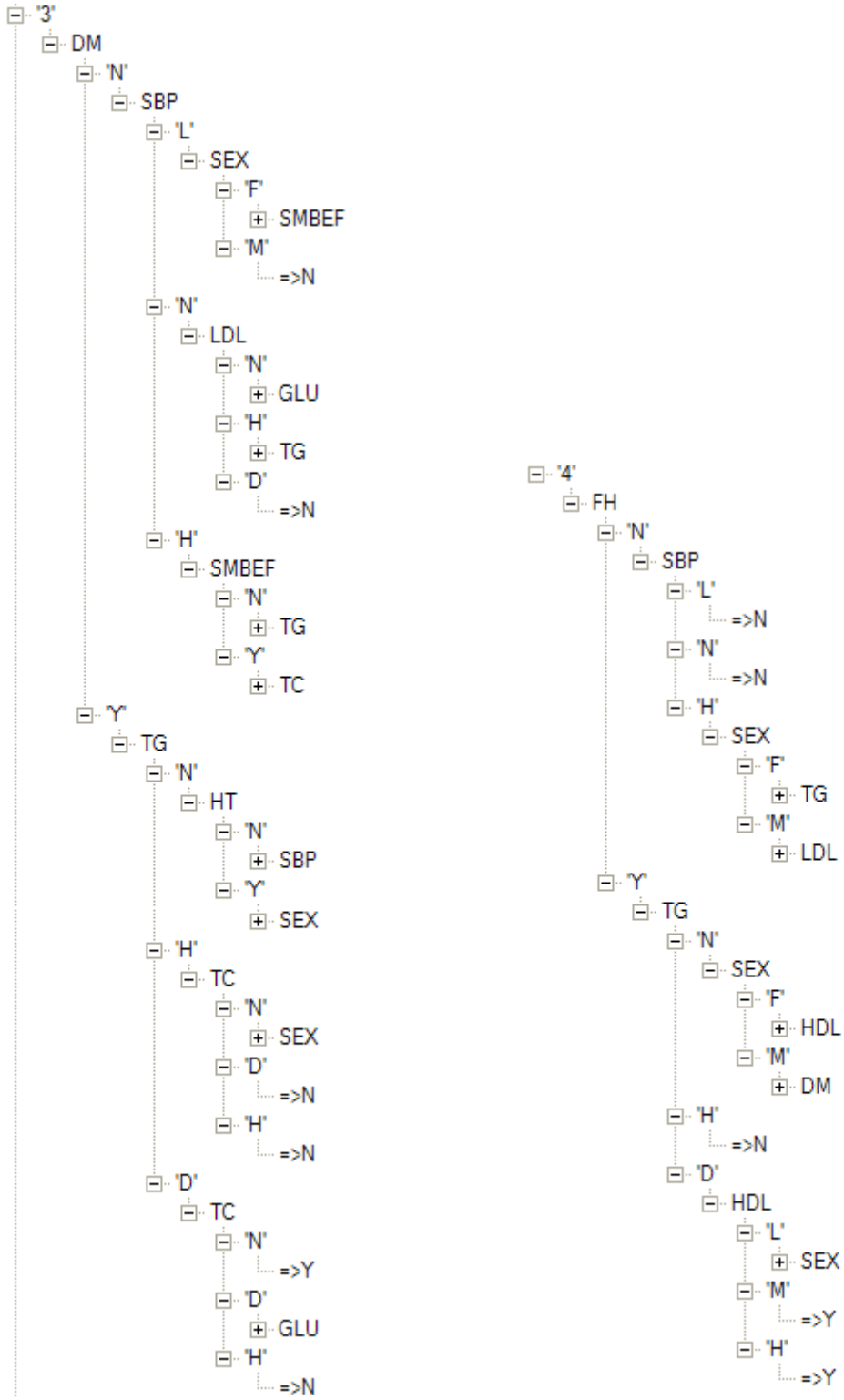


Figure 19: Un-pruned Decision Tree

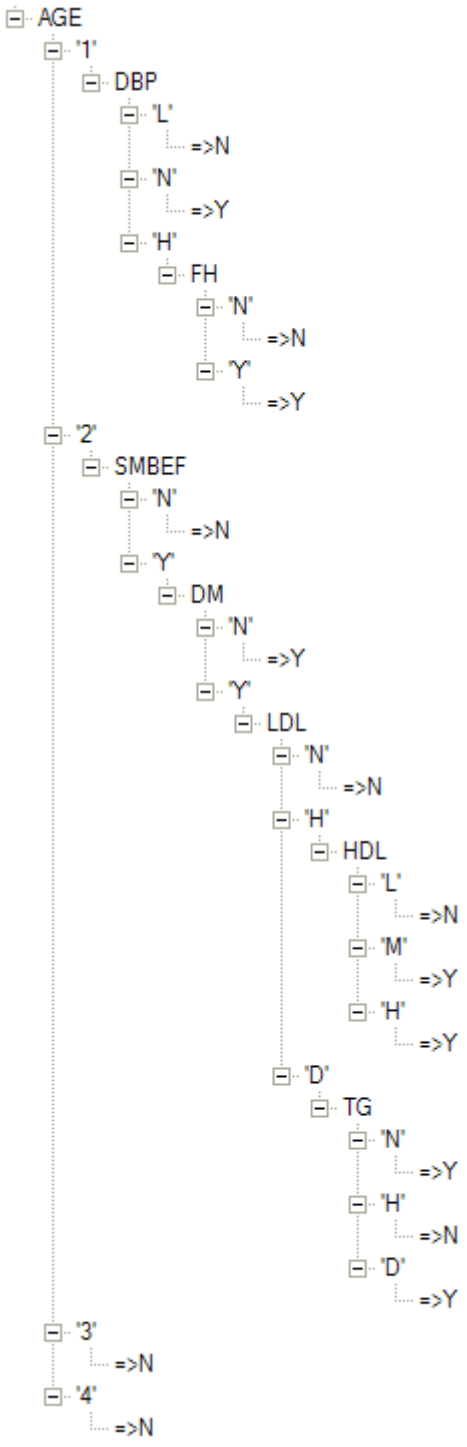


Figure 20: Pruned Decision Tree

4.1 MI - Myocardial Infarction model

We conducted ten runs for each splitting criterion method. In each run, we used a split percentage of 70%-30% and the pruning method. The following table contains the corrected classified instances percentage for each run for each splitting criterion method.

Table 46: MI - Splitting Criteria Method Correctly Classified Instances Percentage per run

Splitting Criteria per run	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10
Information Gain	54.46%	70.54%	68.75%	58.04%	63.39%	65.18%	69.64%	57.14%	66.07%	50.00%
Gain Ratio	58.04%	66.96%	64.29%	59.82%	63.39%	60.71%	65.18%	59.82%	64.29%	67.86%
Chi-squared	66.96%	59.82%	58.93%	54.46%	56.25%	68.75%	60.71%	66.96%	59.82%	65.18%
Gini Gain	64.29%	73.21%	64.29%	66.96%	67.86%	65.18%	66.96%	61.61%	63.39%	65.18%
Distance Measure	57.14%	66.96%	65.18%	61.61%	56.25%	69.64%	66.96%	56.25%	70.54%	63.39%

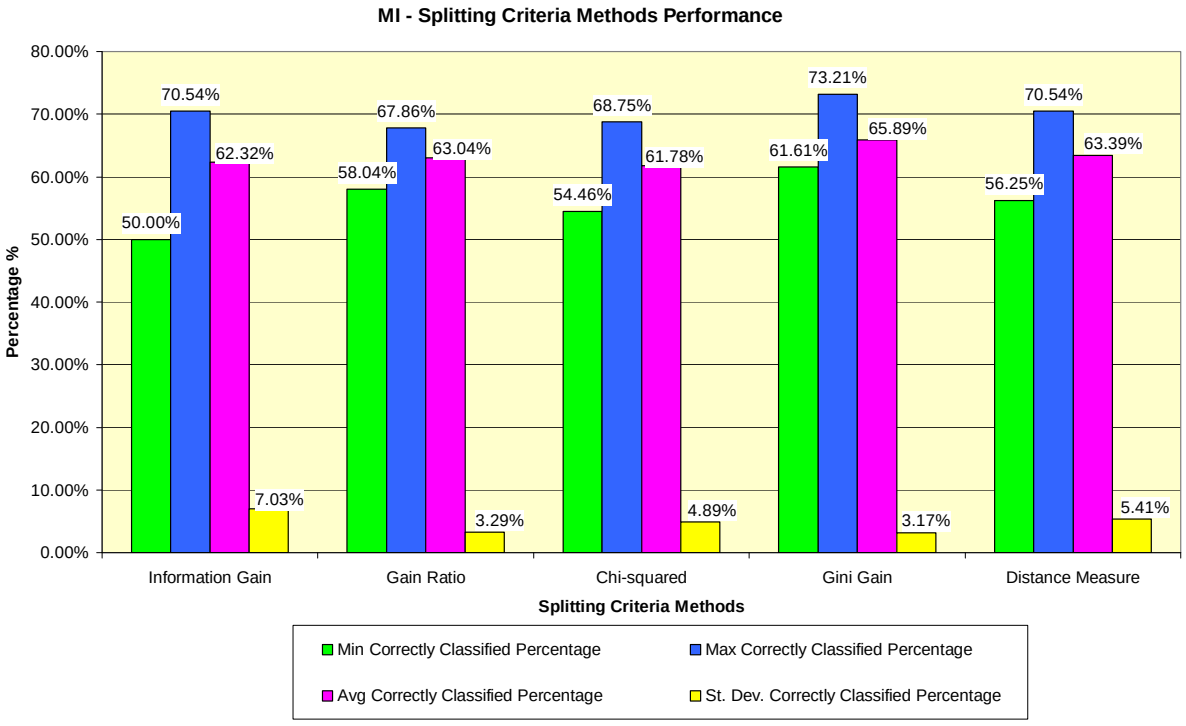


Figure 21: MI - Splitting Criteria Performance Analysis

From the above results, we observed that the splitting criterion with the highest average correctly classified instances percentage is Gini Gain.

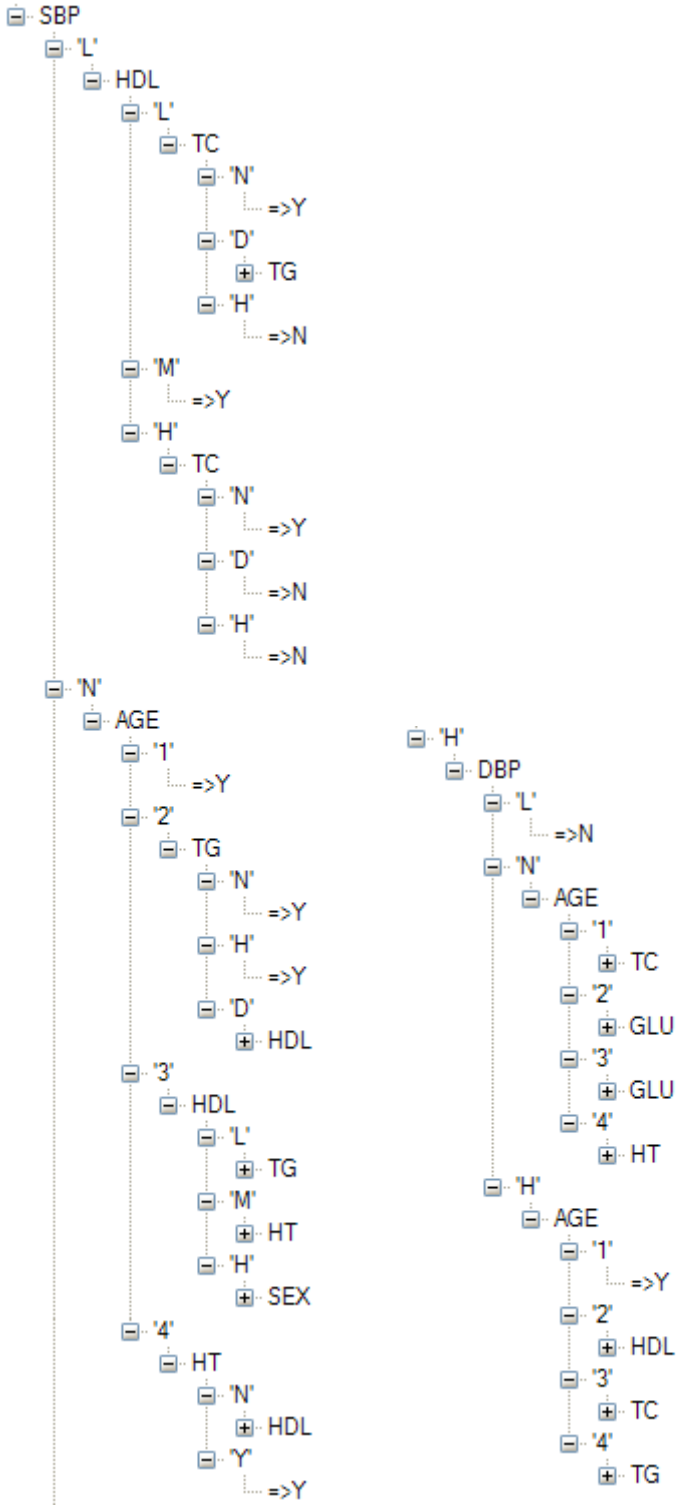


Figure 22: Decision Tree - Run #2 with Gini Gain

Figure 22 gives the decision tree of the second run when it was build with the Gini Gain. The second run had the highest correctly classified percentage from all the Gini Gain's runs.

The attributes that were significant in the decision tree were:

1. SBP
2. DBP
3. AGE
4. HDL
5. HT
6. TC
7. TG

Furthermore, rules extracted from the decision tree are given in Table 67 in Appendix B.

Table 47: Rules with the best measures

AGE	SEX	SMBEF	TC	HDL	LDL	TG	GLU	SBP	DBP	FH	HT	DM	MI	# Attr.	Support	Confidence	Accuracy	Lift
1								N					Y	2	0.027	1	0.241	1.4
				M				L					Y	2	0.062	0.778	0.17	1.089
2						H		N					Y	3	0.018	1	0.17	1.4
2				L		D		N					N	4	0.009	1	0.152	3.5
			N	L				L					Y	3	0.009	1	0.143	1.4
			D	L		N		L					N	4	0.009	0.333	0.098	1.167
2						N		N					Y	3	0.045	1	0.098	1.4
1								H	H				Y	3	0.018	1	0.071	1.4
4								N			Y		Y	3	0.027	0.5	0.062	0.7
3			H					H	H				Y	4	0.018	1	0.062	1.4
3			D	M				H	H				N	5	0.027	1	0.054	3.5
3				M		N		N			N		Y	5	0.018	0.667	0.045	0.933
2				L				H	H	Y			Y	5	0.027	1	0.045	1.4
3				M				N			Y		Y	4	0.036	1	0.045	1.4
2				M				H	H				Y	4	0.027	0.75	0.036	1.05
4						H		H	N		N		N	5	0.009	1	0.036	3.5
2				L		D	H	H	H	N			N	7	0.009	1	0.036	3.5
3			D	L		H		H	H				N	6	0.009	1	0.036	3.5
			D	L		H		L					Y	4	0.009	1	0.036	1.4
2				L			N	H	N				Y	5	0.027	1	0.036	1.4
4						D		H	H				Y	4	0.027	1	0.036	1.4
3							N	H	N	N			Y	5	0.027	0.75	0.027	1.05
4				M			H	N			N		Y	5	0.018	0.667	0.027	0.933
2				M		H	N	H	N		Y		Y	7	0.018	1	0.027	1.4
3		N		L		N		N				Y	Y	6	0.009	1	0.018	1.4
2		Y	N	M			H	H	N				Y	7	0.018	1	0.018	1.4

3	M		N					H	H	N			N	6	0.009	0.5	0.009	1.75
4		Y	D			N	H	H	H			Y	Y	8	0.009	0.5	0.009	0.7
4					N			H	N		Y	Y	Y	6	0.009	0.333	0.009	0.467
2	M	Y	D	L	N	D	H	H	N	Y			N	11	0.009	1	0.009	3.5
2		Y	D	M			H	H	N	Y			N	8	0.009	1	0.009	3.5
3			D			H	N	H	N	Y	Y		N	8	0.009	1	0.009	3.5
3			D	L		N		H	H	N			N	7	0.009	1	0.009	3.5
4		Y	D			N	H	H	H			N	N	8	0.009	1	0.009	3.5
4		Y			N	N		H	N		Y	N	N	8	0.009	0.5	0.009	1.75
3				L		N		N				N	Y	5	0.009	1	0.009	1.4
2				M			N	H	N		N		Y	6	0.009	1	0.009	1.4
2	M	Y	D	L	N	H	H	H	N				Y	10	0.009	1	0.009	1.4
2	M	Y	D	L	N	D	H	H	N	N			Y	11	0.009	1	0.009	1.4
3			N			N	H	H	N				Y	6	0.009	1	0.009	1.4
3			D	L		N	H	H	N		Y		Y	8	0.009	1	0.009	1.4
3	M		D	L		D	H	H	N	N			Y	9	0.009	1	0.009	1.4
3	M	Y	D	L	N	D	H	H	N	Y	Y	N	Y	13	0.009	1	0.009	1.4
4	M	N		M	N	N		H	N	Y	Y	N	Y	11	0.009	1	0.009	1.4
2	M	Y	N	L		H		H	H	N		N	Y	10	0.009	1	0.009	1.4
3			D	L		N		H	H	Y			Y	7	0.009	1	0.009	1.4

1st Observation:

There are no rules with the attribute SEX and the values Female. This occurs because most of the cardiac cases happen to men.

2nd Observation:

This applies for patients between 61-70 years old, have dangerous levels of TC, have low levels of HDL, have normal levels of TG, have high blood pressure.

Table 48: MI - 2nd Observation

AGE	TC	HDL	TG	SBP	DBP	FH	MI
3	D	L	N	H	H	Y	Y
3	D	L	N	H	H	N	N

The above rules show that when the FH attribute is changed from No to Yes, then the outcome is changed. When the patient does not have family history, then the patient will not have a heart attack. If the patient has family history, then the patient will have a heart attack.

3rd Observation:

This applies for patients over 71 years old, smokes, have dangerous levels of TC, have normal levels of TG, have high levels of GLU and have high blood pressure.

Table 49: MI - 3rd Observation

AGE	SMBEF	TC	TG	GLU	SBP	DBP	DM	MI
4	Y	D	N	H	H	H	Y	Y
4	Y	D	N	H	H	H	N	N

The above rules show that when the DM attribute is changed from No to Yes, then the outcome is changed. When the patient does not have diabetes, then the patient will not have a heart attack. If the patient has diabetes, then the patient will have a heart attack.

4th Observation:

This applies for patients that have dangerous levels of TC, have low levels of HDL and have low blood pressure.

Table 50: MI - 4th Observation

TC	HDL	TG	SBP	MI
D	L	N	L	N
D	L	H	L	Y

The above rules show that when the TG attribute is changed from Normal to High, then the outcome is changed. When the patient has normal levels of TG, then the patient will not have a heart attack. If the patient has high levels of TG, then the patient will have a heart attack.

5th Observation:

This applies for patients between 61-70 years old, have dangerous levels of TC, have low levels of HDL and have high blood pressure.

Table 51: MI - 5th Observation

AGE	TC	HDL	TG	SBP	DBP	MI
3	D	L	H	H	H	N
3	D	L	D	H	H	Y

The above rules show that when the TG attribute is changed from High to Dangerous, then the outcome is changed. When the patient has high levels of TG, then the patient will not have a heart attack. If the patient has dangerous levels of TG, then the patient will have a heart attack.

4.2 CABG - Coronary Artery Bypass Surgery model

We conducted ten runs for each splitting criterion method. In each run, we used a split percentage of 70%-30% and the pruning method. The following table contains the corrected classified instances percentage for each run for each splitting criterion method.

Table 52: CABG - Splitting Criteria Method Correctly Classified Instances Percentage per run

Splitting Criteria	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10
Information Gain	51.79%	59.82%	58.04%	54.46%	55.36%	52.68%	56.25%	52.68%	57.14%	56.25%
Gain Ratio	57.14%	52.68%	63.39%	52.68%	55.36%	51.79%	52.68%	50.00%	59.82%	50.89%
Chi-squared	52.68%	47.32%	55.36%	58.93%	62.50%	56.25%	50.89%	58.04%	55.36%	54.46%
Gini Gain	52.68%	53.57%	50.89%	51.79%	57.14%	56.25%	45.54%	54.46%	52.68%	56.25%
Distance Measure	55.36%	51.79%	60.71%	58.04%	60.71%	48.21%	58.93%	50.89%	56.25%	57.14%

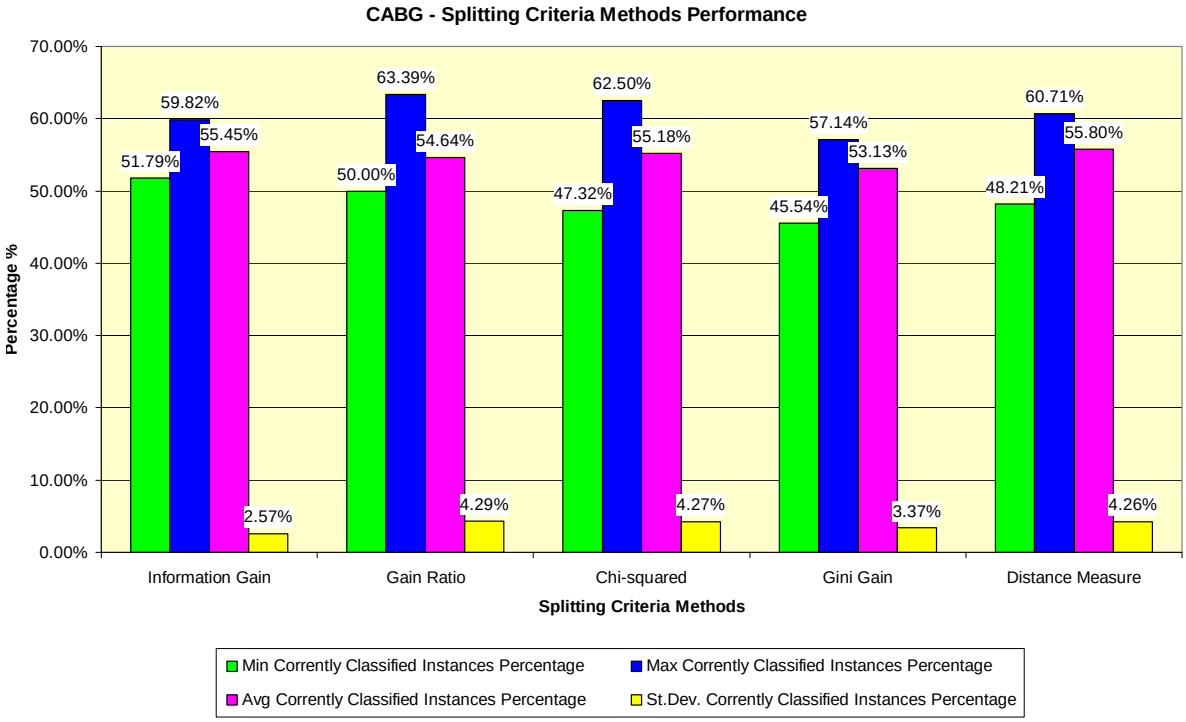


Figure 23: CABG - Splitting Criteria Performance Analysis

From the above results, we observed that the splitting criterion with the highest average correctly classified instances percentage is Distance Measure.

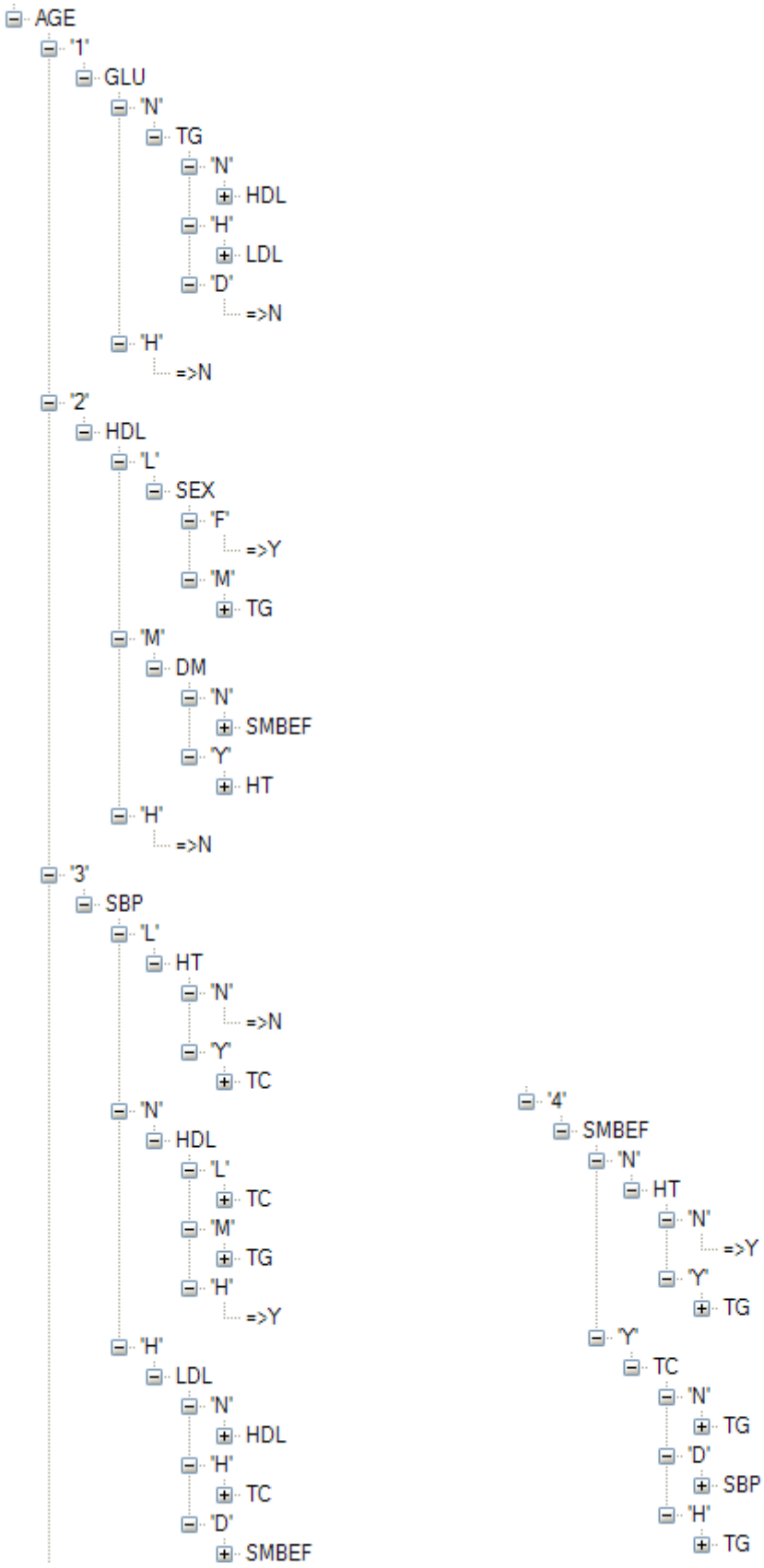


Figure 24: Decision Tree – Run #3 with Distance Measure

The above figure is the decision tree of the third run when it was build with the Distance Measure. The third run had the highest correctly classified percentage from all the Distance Measure's runs.

The attributes that were significant in the decision tree were:

1. AGE
2. GLU
3. HDL
4. SBP
5. SMBEF
6. TG
7. SEX
8. DM
9. HT
10. LDL

Furthermore, rules extracted from the decision tree are given in Table 68 in Appendix B.

Table 53: Rules with the best measures

AGE	SEX	SMBEF	TC	HDL	LDL	TG	GLU	SBP	DBP	FH	HT	DM	CABG	# Attr.	Support	Confidence	Accuracy	Lift
1		N		M		N	N						N	5	0.009	1	0.062	1.697
1					N	H	N						N	4	0.009	1	0.045	1.697
2		N		M			H	H				N	N	6	0.009	1	0.009	1.697
2		Y		M						Y	Y	N	N	6	0.009	1	0.009	1.697
2				M							N	Y	N	4	0.009	1	0.071	1.697
3				M		D	N	N					N	5	0.009	1	0.045	1.697
3		N		L	N	N		H		N			N	7	0.009	1	0.009	1.697
3		N		L	N	H		H		N			N	7	0.009	1	0.009	1.697
3		N		M	N	D		H					N	6	0.009	1	0.009	1.697
3		Y	N	M	N			H					N	6	0.009	1	0.009	1.697
3		Y	D	M	N		H	H		N			N	8	0.009	1	0.009	1.697
2	M	Y	N	L		H							N	6	0.018	1	0.045	1.697
2	M	Y	D	L		H	H	H					N	8	0.018	1	0.018	1.697
2	M			L	N	D	N	H					N	7	0.018	1	0.018	1.697
3		Y			D			H					N	4	0.018	1	0.062	1.697
2	M			L		N			H				N	5	0.027	1	0.036	1.697
2	M			L	N	D	H	H		Y			Y	8	0.009	1	0.009	2.435
2			D	M							Y	Y	Y	5	0.009	1	0.027	2.435
4		Y	D	M			N	N					Y	6	0.009	1	0.027	2.435
2			N	M	N						Y	Y	Y	6	0.018	1	0.018	2.435
3		Y		L	N			H		N			Y	6	0.036	1	0.036	2.435
4		N									N		Y	3	0.036	1	0.205	2.435
1							H						N	2	0.027	0.75	0.214	1.273
4		N				D	H				Y		Y	5	0.027	0.75	0.062	1.826
4		N				N			N		Y		N	5	0.045	0.714	0.045	1.212
1						D	N						N	3	0.018	0.667	0.152	1.131
3		N		M		N		N					N	5	0.018	0.667	0.045	1.131
3				H	N			H					N	4	0.018	0.667	0.045	1.131
2		Y		M						N		N	N	5	0.036	0.667	0.036	1.131

4		Y	D					H			Y		Y	5	0.036	0.571	0.045	1.391
2	M			L		D		N					N	5	0.009	0.5	0.027	0.848
2		N		M				N				N	N	5	0.009	0.5	0.036	0.848
3								L			N		N	3	0.009	0.5	0.223	0.848
4		Y	D				H	N					N	5	0.009	0.5	0.009	0.848
3				L	N	N		H		Y		N	N	7	0.009	0.333	0.009	0.566
3		Y		M		N	N	N					Y	6	0.009	0.333	0.009	0.812
3		N		M	N	N		H					Y	6	0.009	0.333	0.027	0.812
3			D	L				N					Y	5	0.009	0.2	0.062	0.487

1st Observation:

There are no rules with the attribute SEX and the values Female. This occurs because most of the cardiac cases happen to men.

2nd Observation:

This applies for patients between 51-60 years old, is male, have low levels of HDL, have normal levels of LDL, have dangerous levels of TG and have high blood pressure.

Table 54: CABG – 2nd Observation

AGE	SEX	HDL	LDL	TG	GLU	SBP	FH	CABG
2	M	L	N	D	N	H		N
2	M	L	N	D	H	H	Y	Y

The above rules show that when the GLU and FH attributes are changed, then the outcome is changed. When the patient has normal levels of GLU, then the patient will not have a CABG surgery. If the patient has family history and has high levels of GLU, then the patient will have the CABG surgery.

3rd Observation:

This applies for patients over 71 years old, does not smoke and have hypertension (HT).

Table 55: CABG - 3rd Observation

AGE	SMBEF	TG	GLU	DBP	HT	CABG
4	N	D	H		Y	Y
4	N	N		N	Y	N

The above rules show that when the TG, GLU and DBP attributes are changed, then the outcome is changed. When the patient has normal levels of TG and DBP, then the patient will not have a CABG surgery. If the patient has dangerous levels of TG and high levels of GLU, then the patient will have the CABG surgery.

4th Observation:

In my analysis, I found that patients, which are between 61-70 years old, have dangerous levels of TC, have low levels of HDL and have normal blood pressure; from 5 cases only one of them will have the CABG surgery.

Table 56: CABG - 4th Observation

AGE	TC	HDL	SBP	CABG	A. Count	S. Count
3	D	L	N	Y	5	1

5th Observation:

In my analysis, I found that patients, which are over 71 years old, smoke, have dangerous levels of TC, have high blood pressure and have hypertension (HT); from 7 cases only four of them will have the CABG surgery.

Table 57: CABG - 5th Observation

AGE	SMBEF	TC	SBP	HT	CABG	A. Count	S. Count
4	Y	D	H	Y	Y	7	4

4.3 PCI - Coronary Intervention model

We conducted ten runs for each splitting criterion method. In each run, we used a split percentage of 70%-30% and the pruning method. The following table contains the corrected classified instances percentage for each run for each splitting criterion method.

Table 58: PCI - Splitting Criteria Method Correctly Classified Instances Percentage per run

Splitting Criteria	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10
Information Gain	58.56%	64.86%	65.77%	54.05%	62.16%	62.16%	55.86%	53.15%	63.96%	58.56%
Gain Ratio	64.86%	58.56%	61.26%	63.06%	58.56%	58.56%	59.46%	60.36%	58.56%	60.36%
Chi-squared	60.36%	61.26%	63.06%	54.95%	54.05%	60.36%	63.96%	60.36%	55.86%	58.56%
Gini Gain	55.86%	55.86%	63.06%	62.16%	57.66%	58.56%	54.05%	57.66%	72.07%	55.86%
Distance Measure	64.86%	61.26%	62.16%	59.46%	57.66%	48.65%	60.36%	58.56%	61.26%	65.77%

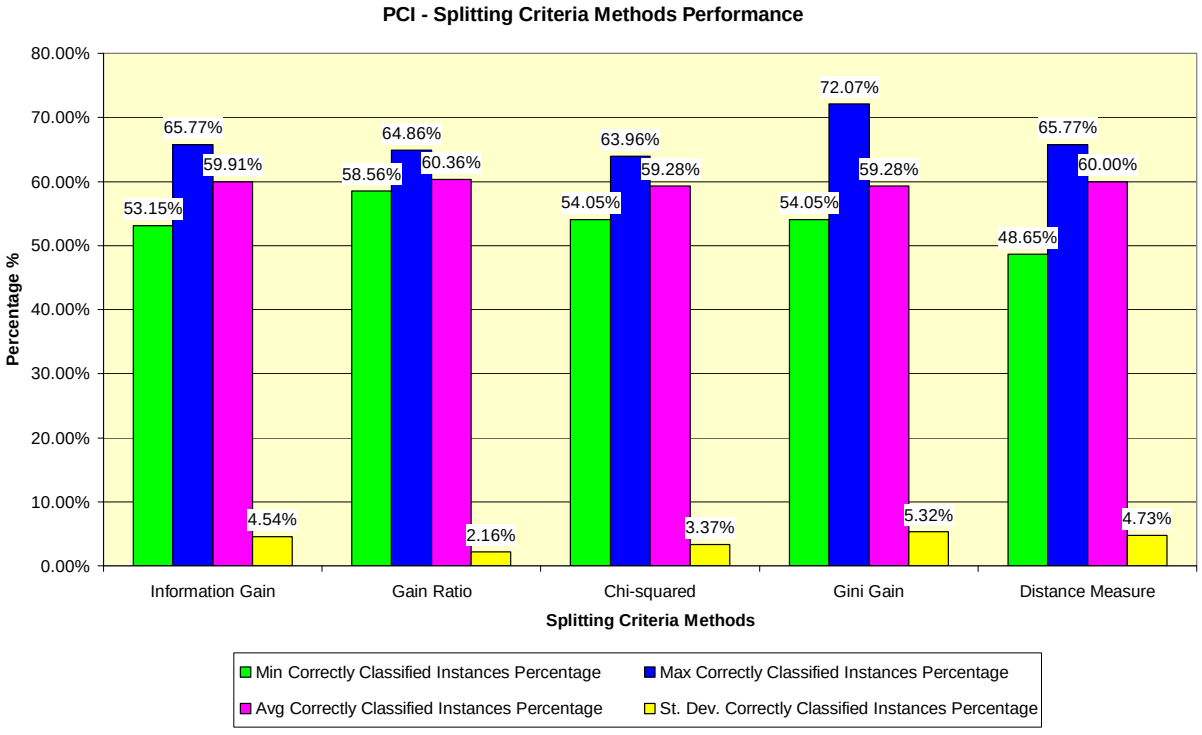


Figure 25: PCI - Splitting Criteria Performance Analysis

From the above results, we observed that the splitting criterion with the highest average correctly classified instances percentage is Gain Ratio.

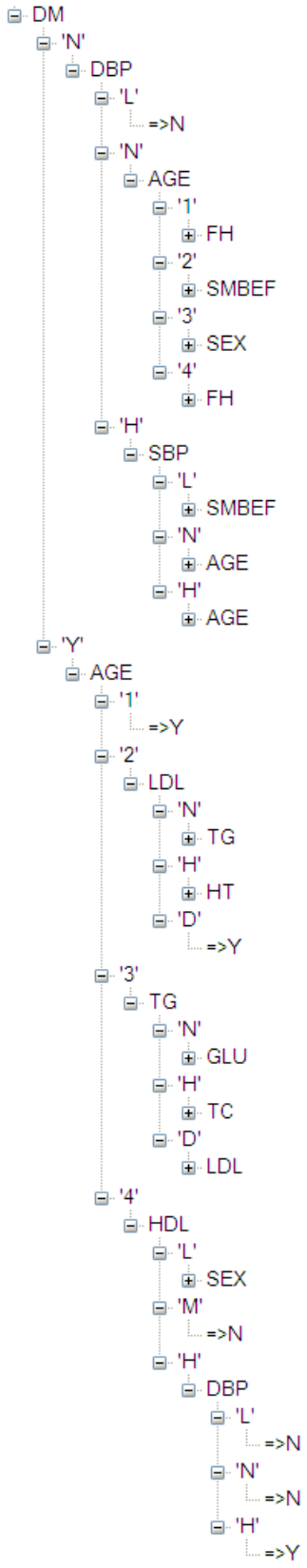


Figure 26: Decision Tree – Run #1 with Gain Ratio

The above figure is the decision tree of the first run when it was build with the Gain Ratio. The first run had the highest correctly classified percentage from all the Gain Ratio's runs.

The attributes that were significant in the decision tree were:

1. DM
2. DBP
3. AGE
4. SBP
5. LDL
6. TG
7. HDL

Furthermore, rules extracted from the decision tree are given in Table 69 in Appendix B.

Table 59: Rules with the best measures

AGE	SEX	SMBEF	TC	HDL	LDL	TG	GLU	SBP	DBP	FH	HT	DM	PCI	# Attr.	Support	Confidence	Accuracy	Lift
2	M	Y	D			H	N	N	N			N	N	9	0.009	1	0.009	1.762
3	M	N			H				N			N	N	6	0.009	1	0.009	1.762
3	M	Y	D	L	N			H	N	N		N	N	10	0.009	1	0.009	1.762
3	M	Y	H					N	N			N	N	7	0.009	1	0.009	1.762
1	M	Y		M				L	H			N	N	7	0.009	1	0.018	1.762
4	M	Y		M				L	H			N	N	7	0.009	1	0.018	1.762
2				L	H						Y	Y	N	5	0.009	1	0.063	1.762
3				M		N	N					Y	N	5	0.009	1	0.027	1.762
3			D			H						Y	N	4	0.009	1	0.063	1.762
4	M			L				H	H	Y		Y	N	7	0.009	1	0.018	1.762
4				H					N			Y	N	4	0.009	1	0.108	1.762
3	M	Y		M				L	H			N	N	7	0.018	1	0.027	1.762
4		Y						H	H			N	N	5	0.018	1	0.018	1.762
2					N	N						Y	N	4	0.018	1	0.054	1.762
2		N					N		N			N	N	5	0.027	1	0.045	1.762
3				L		N	H					Y	N	5	0.027	1	0.081	1.762
2	M	Y	D	M		N			N	N	N	N	Y	10	0.009	1	0.009	2.312
2	M	Y	N			H			N			N	Y	7	0.009	1	0.027	2.312
3	M	Y	N	M	N	H			N			N	Y	9	0.009	1	0.009	2.312
3	M	Y	D	M	N	H		N	N	Y	Y	N	Y	12	0.009	1	0.009	2.312
3	M	Y	D		N	N		H	N	Y		N	Y	10	0.009	1	0.009	2.312
2		Y	D			H		H	H			N	Y	7	0.009	1	0.009	2.312
3	F							H	H			N	Y	5	0.009	1	0.027	2.312
1												Y	Y	2	0.009	1	0.351	2.312
1		Y							N	N	N	N	Y	6	0.036	1	0.072	2.312
2		Y				N		H	H			N	Y	6	0.045	1	0.045	2.312
4				M								Y	N	3	0.027	0.75	0.18	1.321
4	M								N	N		N	N	5	0.054	0.75	0.054	1.321
2	M	Y				D			N			N	Y	6	0.027	0.75	0.045	1.734

3	M	N		M	N				N			N	Y	7	0.027	0.75	0.027	1.734
									L			N	N	2	0.018	0.667	0.09	1.175
3	M						H	H	H			N	N	6	0.018	0.667	0.018	1.175
1									N	Y		N	Y	4	0.018	0.667	0.054	1.542
3				M		N	H					Y	N	5	0.027	0.6	0.09	1.057
4						N			N	Y		N	Y	5	0.027	0.6	0.027	1.388
4	F					N			N	N		N	N	6	0.009	0.5	0.027	0.881
1							N	H	H			N	N	5	0.009	0.5	0.009	0.881
2					N	H				Y		Y	N	5	0.009	0.5	0.018	0.881
2					N	D						Y	N	4	0.009	0.5	0.018	0.881
3					N	D						Y	Y	4	0.009	0.333	0.018	0.771

1st Observation:

This applies for patients that have normal levels of LDL, have dangerous levels of TG and have diabetes (DM).

Table 60: PCI - 1st Observation

AGE	LDL	TG	DM	PCI
2	N	D	Y	N
3	N	D	Y	Y

The above rules show that when the AGE attribute is changed, then the outcome is changed. When the patient is between 51-60 years old, then the patient will not have a PCI surgery. If the patient is between 61-70 years old, then the patient will have the PCI surgery.

2nd Observation:

This applies for patients that are between 61-70 years old, have high levels of HDL, have normal levels of TG, have high levels of GLU and have diabetes (DM).

Table 61: PCI - 2nd Observation

AGE	SEX	HDL	TG	GLU	DM	PCI
3	F	H	N	H	Y	N
3	M	H	N	H	Y	Y

The above rules show that when the SEX attribute is changed, then the outcome is changed. When the patient is female, then the patient will not have a PCI surgery. If the patient is male, then the patient will have the PCI surgery.

3rd Observation:

This applies for patients that have normal levels of LDL, have dangerous levels of TG and have diabetes (DM).

Table 62: PCI - 3rd Observation

AGE	SMBEF	HDL	LDL	HT	DM	PCI
2	N	M	H	Y	Y	N
2	Y	M	H	Y	Y	Y

The above rules show that when the SMBEF attribute is changed, then the outcome is changed. When the patient does not smoke, then the patient will not have a PCI surgery. If the patient smokes, then the patient will have the PCI surgery.

4th Observation:

This applies for patients that are over 71 years old, are male, have low levels of HDL, have high blood pressure, do not have family history and have diabetes (DM).

AGE	SEX	HDL	SBP	DBP	TC	FH	DM	PCI
4	M	L	H	H	N	N	Y	N
4	M	L	H	H	H	N	Y	Y
4	M	L	H	H	D	N	Y	Y

The above rules show that when the TC attribute is changed, then the outcome is changed. When the patient has normal levels of TC, then the patient will not have a PCI surgery. If the patient has high or dangerous levels of TC, then the patient will have the PCI surgery.

4.4 Multiple classes model with MI, CABG and PCI

We conducted four runs for each splitting criterion method. Each run had different number of class output values. Below, we are explaining the coding for each run.

3 class output values:

- A – the patient had a heart attack but no surgeries (MI=Y, CABG or PCI=N)
- B – the patient did not have a heart attack but had surgeries (MI=N, CABG or PCI=Y)
- C – the patient had a heart attack and had surgeries (MI=Y, CABG or PCI=Y)

4 class output values:

- A – the patient had a heart attack but no surgeries (MI=Y, CABG or PCI=N)
- B – the patient did not have a heart attack but had surgeries (MI=N, CABG or PCI=Y)
- C – the patient had a heart attack and had surgeries (MI=Y, CABG or PCI=Y)
- D – the patient did not have a heart attack and did not have surgeries (MI=N, CABG or PCI=N)

5 class output values:

- A – the patient had a heart attack but no surgeries (MI=Y, CABG or PCI=N)
- B – the patient did not have a heart attack but had PCI surgery (MI=N, CABG=N, PCI=Y)
- C – the patient did not have a heart attack but had CABG surgery (MI=N, CABG=Y, PCI=N)
- D – the patient had a heart attack and had PCI surgery (MI=Y, CABG=N, PCI=Y)
- E – the patient had a heart attack and had CABG surgery (MI=Y, CABG=Y, PCI=N)

6 class output values:

- A – the patient had a heart attack but no surgeries (MI=Y, CABG=N, PCI=N)

- B – the patient did not have a heart attack but had PCI surgery (MI=N, CABG=N, PCI=Y)
- C – the patient did not have a heart attack but had CABG surgery (MI=N, CABG=Y, PCI=N)
- D – the patient had a heart attack and had PCI surgery (MI=Y, CABG=N, PCI=Y)
- E – the patient had a heart attack and had CABG surgery (MI=Y, CABG=Y, PCI=N)
- FGH – the patient had CABG and PCI surgeries (CABG=Y, PCI=Y)

8 class output values:

- A – the patient had a heart attack but no surgeries (MI=Y, CABG or PCI=N)
- B – the patient did not have a heart attack but had PCI surgery(MI=N, CABG=N, PCI=Y)
- C – the patient did not have a heart attack but had CABG surgery (MI=N, CABG=Y, PCI=N)
- D – the patient had a heart attack and had PCI surgery (MI=Y, CABG=N, PCI=Y)
- E – the patient had a heart attack and had CABG surgery (MI=Y, CABG=Y, PCI=N)
- F – the patient did not have a heart attack but had CABG and PCI surgeries (MI=N, CABG=Y, PCI=Y)
- G – the patient had a heart attack and had CABG and PCI surgeries (MI=Y, CABG=Y, PCI=Y)
- H – the patient did not have a heart attack or surgeries (MI=N, CABG=N, PCI=N)

In each run, we used the 100% of training data set for testing and the pruning method. The following table contains the corrected classified instances percentage for each run for each splitting criterion method.

Table 63: All classes - Splitting Criteria Method Correctly Classified Instances Percentage per run

Splitting Criteria	3 classes	4 classes	5 classes	6 classes	8 classes
Information Gain	64.56%	56.91%	57.23%	95.39%	92.41%
Gain Ratio	62.36%	56.37%	55.16%	94.04%	91.06%
Chi-squared	64.56%	56.91%	56.23%	95.39%	92.41%
Gini Gain	60.71%	56.10%	51.92%	95.12%	93.50%
Distance Measure	62.36%	57.45%	54.57%	94.85%	93.22%

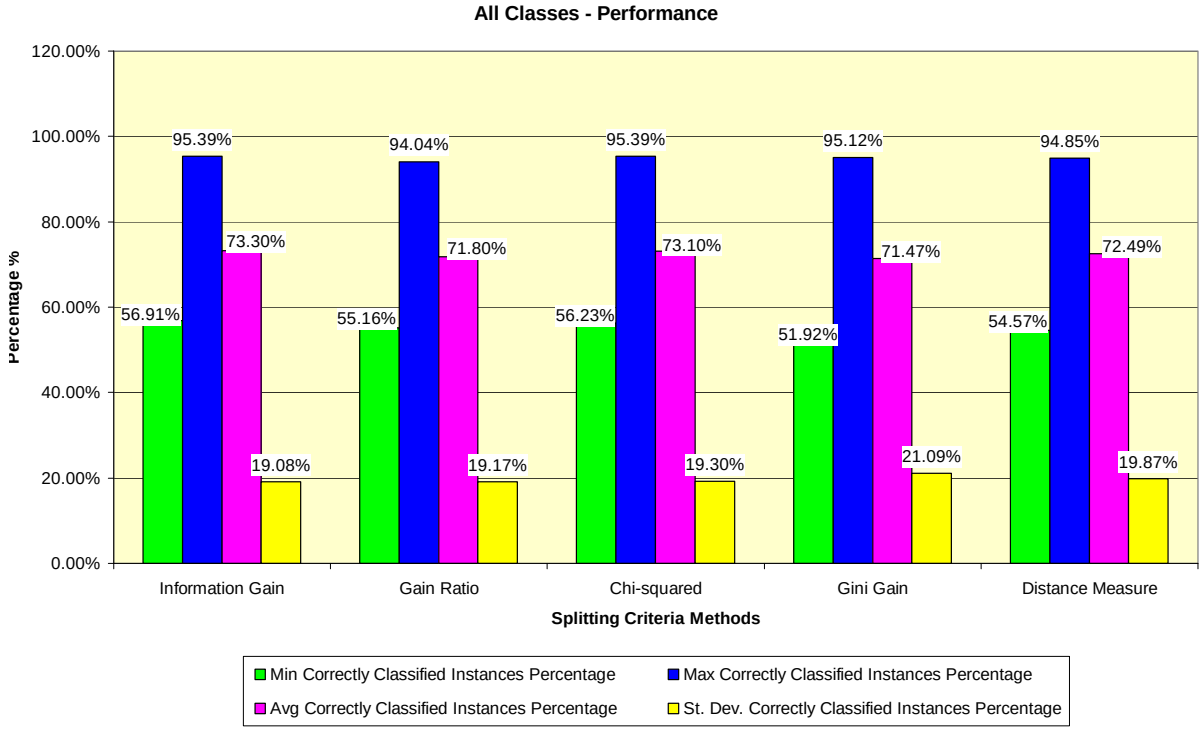


Figure 27: All classes - Splitting Criteria Performance Analysis

From the above results, we observed that the splitting criterion with the highest average correctly classified instances percentage is Information Gain.

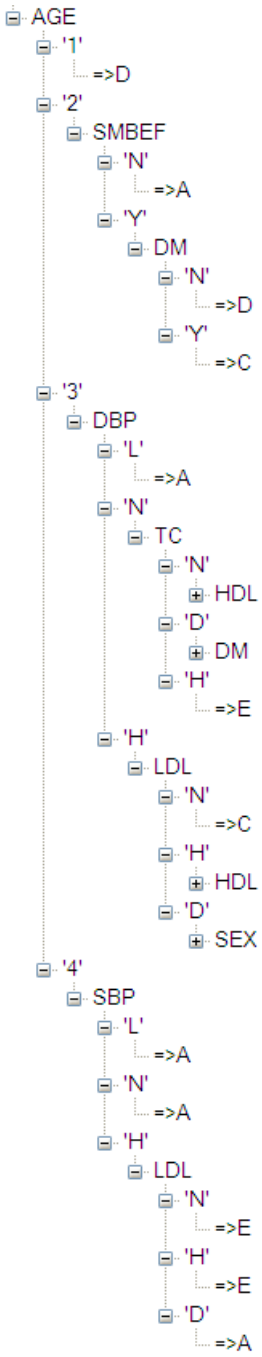


Figure 28: Decision Tree – Run with 6 classes with Information Gain

The above figure is the decision tree of the run with 6 classes when it was build with the Information Gain. The run had the highest correctly classified percentage from all the Information Gain's runs.

The attributes that were significant in the decision tree were:

1. AGE
2. SMBEF
3. DBP
4. SBP
5. DM
6. TC
7. LDL

Furthermore, rules extracted from the decision tree are given in Table 70 in Appendix B.

Table 64: Rules with best measures

AGE	SEX	SMBEF	TC	HDL	LDL	TG	GLU	SBP	DBP	FH	HT	DM	6 Classes	# Attr.	Support	Confidence	Accuracy	Lift
3									L				A	2	0.003	1	0.485	4.792
3		N	D	L		N			N		Y	N	A	8	0.003	1	0.003	4.792
3		Y	D	L		H		N	N			N	A	8	0.003	1	0.003	4.792
3			N	L		N	H		N				A	6	0.005	1	0.016	4.792
3			N	M		N	H	N	N				A	7	0.005	1	0.025	4.792
3	M		H		D				H				A	5	0.011	1	0.035	4.792
3		N	D			H		N	N			N	B	7	0.003	1	0.011	10.25
3			D			D			N		N	Y	B	6	0.003	1	0.024	10.25
3				L	H				H				B	4	0.003	1	0.192	10.25
3			N	L		D			N				C	5	0.003	1	0.060	6.710
3	M		N		D				H				C	5	0.003	1	0.030	6.710
3			D			H			N			Y	C	5	0.005	1	0.043	6.710
3			N	L		H			N				D	5	0.003	1	0.060	3.728
3			N	M		N	H	H	N				D	7	0.003	1	0.016	3.728
3			N	H				H	N				D	5	0.003	1	0.047	3.728
3		Y	D			H		H	N	Y		N	D	8	0.003	1	0.003	3.728
3	F		D			D			N			N	D	6	0.003	1	0.011	3.728
3	F		D			D			N		Y	Y	D	7	0.003	1	0.008	3.728
3	M		D	M		D			N		Y	Y	D	8	0.003	1	0.003	3.728
3				H	H				H				D	4	0.003	1	0.244	3.728
3	F				D				H				D	4	0.003	1	0.230	3.728
3			D	L		N	N		N			Y	D	7	0.005	1	0.008	3.728
3		N	D			H		H	N	N		N	D	8	0.008	1	0.008	3.728
3	F		D			N		H	N		N	N	E	8	0.003	1	0.003	5.125
3			D	M		N	N		N			Y	E	7	0.003	1	0.005	5.125
3	M		D	L		D			N		Y	Y	E	8	0.005	1	0.005	5.125
3		Y	D			H		H	N	N		N	E	8	0.008	1	0.008	5.125
3			N	L		N	N		N				FGH	6	0.003	1	0.014	12.3
3			N	M	H	D			N				FGH	6	0.003	1	0.060	12.3

3	F		D			N		N	N		N	N	FGH	8	0.003	1	0.014	12.3
3		Y	D	M		H		N	N			N	FGH	8	0.003	1	0.008	12.3
3		N	D			H		H	N	Y		N	FGH	8	0.003	1	0.003	12.3
3			N	M		N	N		N				FGH	6	0.005	1	0.035	12.3
4								L					A	2	0.011	0.8	0.564	3.834
3	M		D			N			N		N	N	D	7	0.016	0.75	0.019	2.795
3			D	L		N	H		N			Y	E	7	0.008	0.75	0.019	3.844
3			D	M		N	H		N			Y	A	7	0.005	0.667	0.019	3.195
3			N	M	N	D			N				B	6	0.005	0.667	0.008	6.833
3			N	H				N	N				C	5	0.005	0.667	0.128	4.473
3				M	H				H				A	4	0.014	0.625	0.141	2.995
3		N	D	M		N			N		Y	N	D	8	0.008	0.6	0.008	2.236
1													D	1	0.054	0.556	0.743	2.071
2		N											A	2	0.030	0.5	0.363	2.396
3	M		D			D			N			N	B	6	0.008	0.5	0.008	5.125
3			N	M		H			N				D	5	0.005	0.5	0.065	1.864
4					H			H					E	3	0.008	0.5	0.206	2.563
3		Y	D			N			N		Y	N	E	7	0.011	0.444	0.011	2.278
2		Y										N	D	3	0.073	0.435	0.144	1.623
3			H						N				E	3	0.008	0.428	0.157	2.197
4					D			H					A	3	0.005	0.4	0.209	1.917
3					N				H				C	3	0.035	0.394	0.092	2.643
4								N					A	2	0.016	0.316	0.442	1.514
2		Y										Y	C	3	0.016	0.316	0.173	2.119
4					N			H					E	3	0.041	0.268	0.113	1.373

Table 65: Run with 6 classes with Information Gain - Confusion Matrix

CLASS	A	B	C	D	E	FGH
A	41	1	9	15	11	0
B	2	8	6	13	7	0
C	9	0	25	3	18	0
D	7	1	8	72	11	0
E	12	2	8	15	35	0
FGH	2	0	3	11	7	7

Table 66: Run with 6 classes with Information Gain - Accuracy per class

TP-Rate	FP-Rate	Precision	Recall	F-Measure	Sensitivity	Specificity	CLASS
0.532	0.179	0.562	0.532	0.547	0.562	0.803	A
0.222	0.022	0.667	0.222	0.333	0.667	0.865	B
0.455	0.173	0.424	0.455	0.439	0.424	0.845	C
0.727	0.329	0.558	0.727	0.632	0.558	0.811	D
0.486	0.261	0.393	0.486	0.435	0.393	0.805	E
0.233	0.000	1.000	0.233	0.378	1.000	0.887	FGH

Chapter 5: Discussion

Clinical and statistical studies have identified several factors that increase the risk of a heart attack. Major risks factors are those that research has shown significantly increase of cardiovascular disease, but their significance and frequency have not yet been determined. Therefore, they are called contributing risk factors.

Our study has proven that the contributing risk factors affect the risk of cardiovascular diseases. Some of them can be modified, treated or controlled, and some can not. The more risk factors you have, the greater your chance of developing a heart disease. Also, the greater the level of each risk factor, the greater the risk. In Figure 29, the American Heart Association is illustrating the ABCs of preventing heart diseases [27].

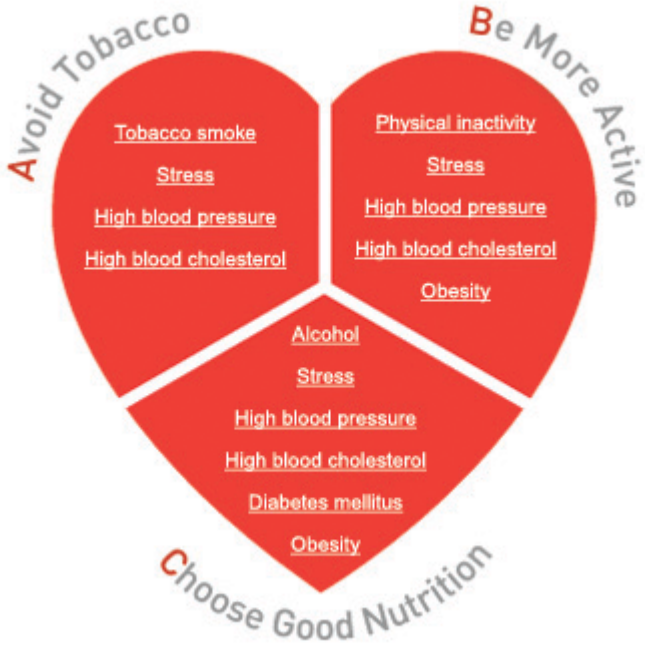


Figure 29: ABCs of Preventing Heart Disease, Stroke and Heart Attack

Contributing Risk Factors [28, 29]:

- Increasing age - Over eighty percent of people who die of heart disease is 65 or older. (PCI – 1st Observation)
- Male sex (gender) - Men have a greater risk of heart attack than women do. (PCI – 2nd Observation)
- Heredity - Most people with a strong family history of heart disease have one or more other risk factors. (CABG – 2nd Observation, MI – 2nd Observation)
- Smoking - Smoking is a risk factor for sudden cardiac death in patients with heart disease; smokers have about twice the risk of non-smokers. Exposure to other people's smoke increases the risk of heart disease even for non-smokers.(PCI – 3rd Observation)
- High blood cholesterol - As blood cholesterol rises, so does risk of heart disease. When other risk factors are present, this risk increases even more. (PCI – 4th Observation)
- High blood pressure - High blood pressure increases the heart's workload, causing the heart to thicken and become stiffer. When high blood pressure exists with obesity, smoking, high blood cholesterol levels or diabetes, the risk of heart attack or stroke increases several times.(CABG – 3rd Observation)
- Physical inactivity - An inactive lifestyle is a risk factor for coronary heart disease.
- Obesity and overweight - People who have excess body fat are more likely to develop heart disease and stroke even if they have no other risk factors.
- Diabetes mellitus — Diabetes seriously increases your risk of developing cardiovascular disease. Even when glucose (blood sugar) levels are under control, diabetes increases the risk of heart disease and stroke, but the risks are even

greater if blood sugar is not well controlled. (MI – 3rd Observation, PCI – 2nd and 3rd Observation)

- Stress - Some scientists have noted a relationship between heart disease risk and stress in a person's life.
- Alcohol - Drinking too much alcohol can raise blood pressure, cause heart failure and lead to stroke. It can contribute to high triglycerides, cancer and other diseases, and produce irregular heartbeats. It contributes to obesity, alcoholism, suicide and accidents. (MI – 4th and 5th Observation)

We compared our results with the results of the master thesis of Loucia Papaconstantinou. Loucia's thesis used the same cardiovascular database. However Loucia extracted her rules using a different algorithm. She used Apriori association algorithm.

From her results, she found the most significant factors for the three events: MI, CABG and PCI.

MI:

1. SEX
2. LDL
3. **TC**
4. DM
5. **DBP**
6. SMBEF
7. **HT**

CABG:

1. **SEX**
2. **LDL**
3. **DM**
4. TC
5. **SMBEF**
6. DBP
7. **HT**
8. **GLU**
9. FH
10. **SBP**

PCI:

1. SEX
2. **LDL**
3. TC
4. **DM**
5. SMBEF
6. HT
7. FH

Comparing Loucia's results with our results, for the MI event only 3 factors are common, for the CABG event only 7 factors are common and for PCI event only 2 factors are common.

Chapter 6: Conclusions and Future Work

6.1 Conclusions

In our study we have analyzed and tested different splitting criteria methods. In our first model, that contains the observations for the event of a heart attack (MI), we found that the best splitting criteria method was Gini Gain. In our second model, that contains the observations for the event of a Percutaneous Coronary Intervention surgery (PCI), we discovered that the best splitting criteria method was Distance Measure (which is the normalized form of Gini Gain). In our third model, that contains the observations for the event of a Coronary Artery Bypass Surgery (CABG), we have uncovered that the best splitting criteria method was Gain Ratio (which is the normalized form of Information Gain). Finally, when we united all the three models and created multiple classes models, we found that the best splitting criteria method was Information Gain. Therefore, we conclude that each method works differently on different models.

Studying the results of the classification algorithms, we were able to find the contributing risk factors to cardiovascular diseases. Using the decision tree model, doctors can identify the risk factors that might contribute or cause an event. In order to prevent heart attack event, the cardiologist can control the risk factors of the patient that can be modified.

6.2 Future Work

Our future work is to create a user-friendly application that the cardiologist can use to generate the rules that apply for a specific patient. The cardiologist can input the patient's information (blood results, age, smoking habits, etc) and the application will produce the prediction of the patient. Furthermore, the application can offer suggestions to which risk factors should be controlled by the patient to lower risk of a heart attack.

Bibliography

[1] M. Russell, The Biggest Cause Of Death In The Western World, Ezine @rticles®, June 2006, http://ezinearticles.com/?The-Biggest-Cause-Of-Death-In-The-Western-World!&id=225059, Last Accessed April 2009.

[2] D Lee, D Kulick, J Marks, *Heart Attack (Myocardial Infarction)*, MedicineNet.com, November 28, 2006.

[3] John G. Canto, Robert J. Goldberg, Mary M. Hand, Robert O. Bonow, George Sopko, Carl J. Pepine, and Terry Long, *Symptom Presentation of Women With Acute Coronary Syndromes: Myth vs Reality*, Arch Intern Med, Dec 10/24, 2007, Vol. 167, p. 2405 – 2413.

[4] M Wijesinghe, K Perrin, A Ranchord, M Simmonds, M Weatherall and R Beasley, *Routine use of oxygen in the treatment of myocardial infarction: systematic review*, Heart, Feb 2009, Vol. 95, p. 198 - 202.

[5] Ministry of Health, *Cyprus: Annual Report 2007*, May 2008

[6] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Second Edition.

[7] L. Rokach, O. Maimon, *Data Mining with Decision Trees: Theory and Applications*, Series in Machine Perception Artificial Intelligence, Vol. 69.

[8] Kantardzic, Mehmed, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, 2003

[9] F. Usama, G. Piatetsky-Shapiro, P. Smyth, *From Data Mining to Knowledge Discovery in Databases*, 1996

[10] *Data Mining Algorithms*, Analysis Services - Data Mining, SQL Server 2008 Books Online

[11] S. Kullback, R.A. Leibler, *On Information and Sufficiency, The Annals of Mathematical Statistics*, Vol. 22, p. 79–86, 1951

[12] http://www.wikipedia.org (Keywords: Myocardial Infarction, Percutaneous coronary intervention, Coronary artery bypass surgery, Data Mining) Last Accessed May 2009

[13] L. Geng, H. J. Hamilton, *Interestingness Measures for Data Mining: A Survey*, ACM Comput. Surv., Vol. 38, No. 3, 2006, p. 1-32.

[14] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Chapman & Hall/CRC, 2000.

[15] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, Second Edition, 2005.

[16] A. Maton, J. Hopkins, C. W. McLaughlin, S. Johnson, M. Q. Warner, D. LaHart, J. D. Wright, *Human Biology and Health*, Englewood Cliffs, New Jersey, USA, Prentice Hall, 1993.

[17] L. M. Tierney, S. J. McPhee, M. A. Papadakis, *Current medical Diagnosis & Treatment*, International edition, New York, Lange Medical Books/McGraw-Hill, 2002, pp. 1203–1215.

[18] *Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications*, Department of Noncommunicable Disease Surveillance, 1999.

[19] *LDL and HDL Cholesterol: What's Bad and What's Good*, American Heart.

[20] *Nomenclature of Lipids*, IUPAC-IUB Commission on Biochemical Nomenclature (CBN).

[21] D. Horton, J.D. Wander, *The Carbohydrates*, Vol IB, New York: Academic Press, 1980, pp. 727–728.

[22] http://www.nlm.nih.gov/medlinepluss/ency/article/003650.htm, Last Accessed April 2009

[23] M. Mehta, J. Rissanen, R. Agrawal, *MDL-based decision tree pruning*, In Proc. of KDD, 1995, p. 216-221.

[24] J. Gehrke, R. Ramakrishnan, V. Ganti, *RainForest - A Framework for Fast Decision Tree Construction of Large Datasets*, In VLDB, 1998, p. 416-427, Morgan Kaufmann

[25] R.Rastogi, K.Shim, *PUBLIC: A Decision Tree Classifier that Integrates Pruning and Building*, In Proc. of VLDB, 1998, 4(4):315-344.

[26] H. Hamilton, E. Gurak, L. Findlater, and W. Olive, *Decision Trees*, 2002, http://www.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/4_dtrees1.html, Last Accessed April 2009 .

[27] *ABCs of Preventing Heart Disease, Stroke and Heart Attack*, American Heart Association, 2009.

[28] *Myocardial Infarction (Heart Attack)*, British Heart Foundation, 2008.

[29] *Risk Factors and Coronary Heart Disease*, American Heart Association, 2009.

Appendix A

In this appendix, we are describing the application. When the application is started, the first screen is the input screen. By clicking on the button 'Open file', we can select any ARFF file as our database file.

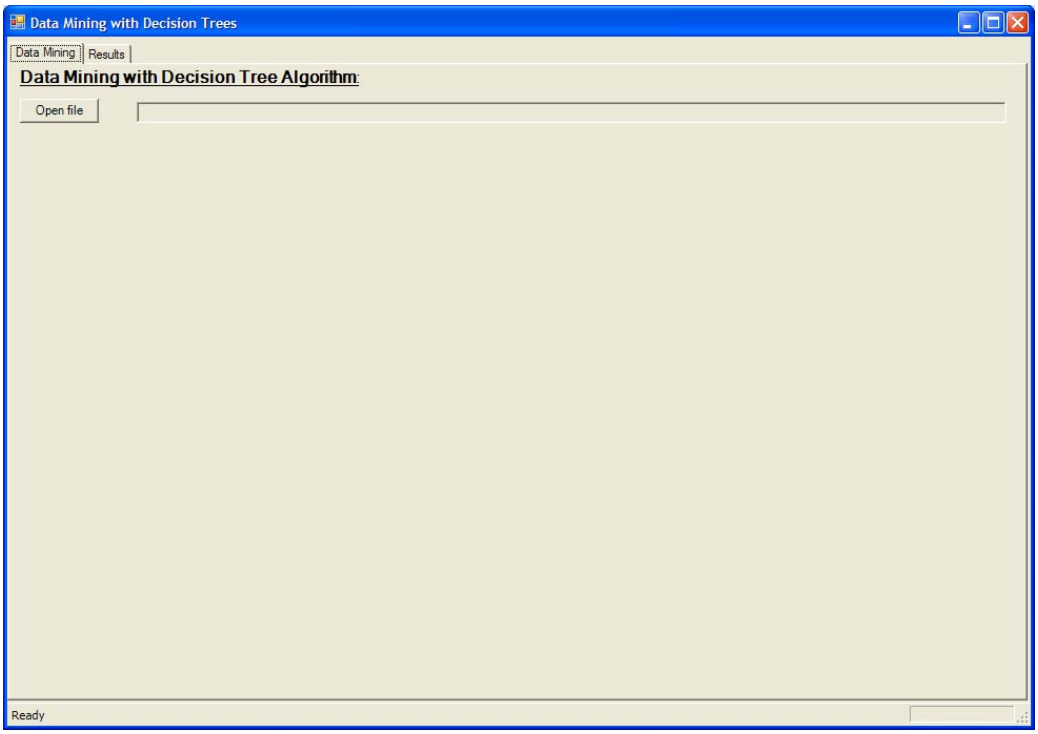


Figure 30: Appendix A - Input Screen

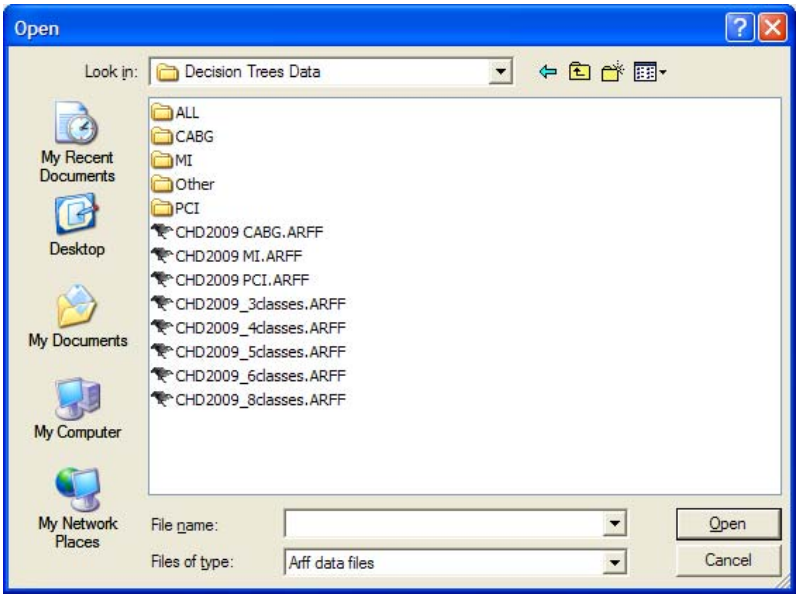


Figure 31: Appendix A - Open file screen

When the database file is loaded, then the screen will show you the following:

1. data values table
2. attributes information table
3. splitting criteria methods options
4. statistics measures table
5. selecting single class value option
6. pruning option
7. testing method option

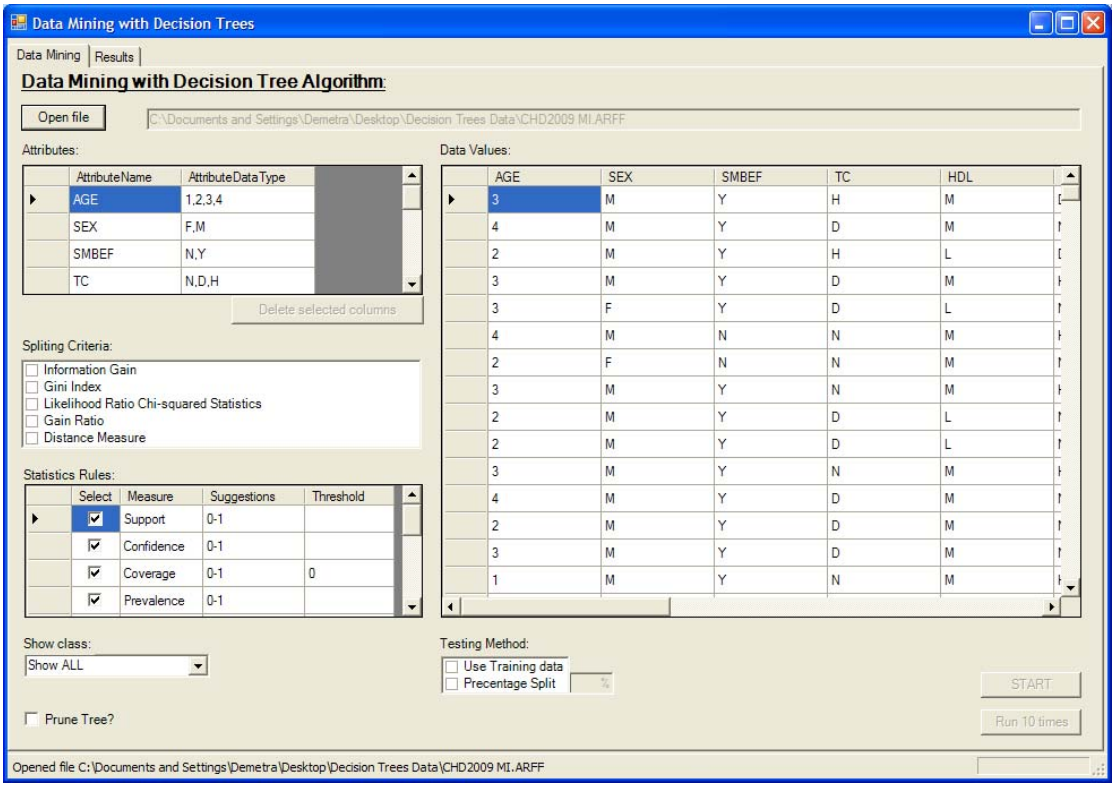


Figure 32: Appendix A – Options

You can select attributes, except the class, and remove them from the data set. You can only select one splitting criteria method. You can select any statistic measure for the rules and set any threshold, in order to filter the rules. You can select to show all the class output values of select a single class values. You can use the pruning method. You can test the extracted rules by using the training data set or by creating the testing data set from the percentage split.

Once you configure the data mining run, then click on the button “Start” for a single execution or the button “Run 10 times” for a 10-times execution.

When you click on the button “Run 10 times” for a 10-times execution, the application will run for 10 times and save each run to log files in the same directory with the input database file.

When you click on the button “Start”, then the next tab with the results appears.

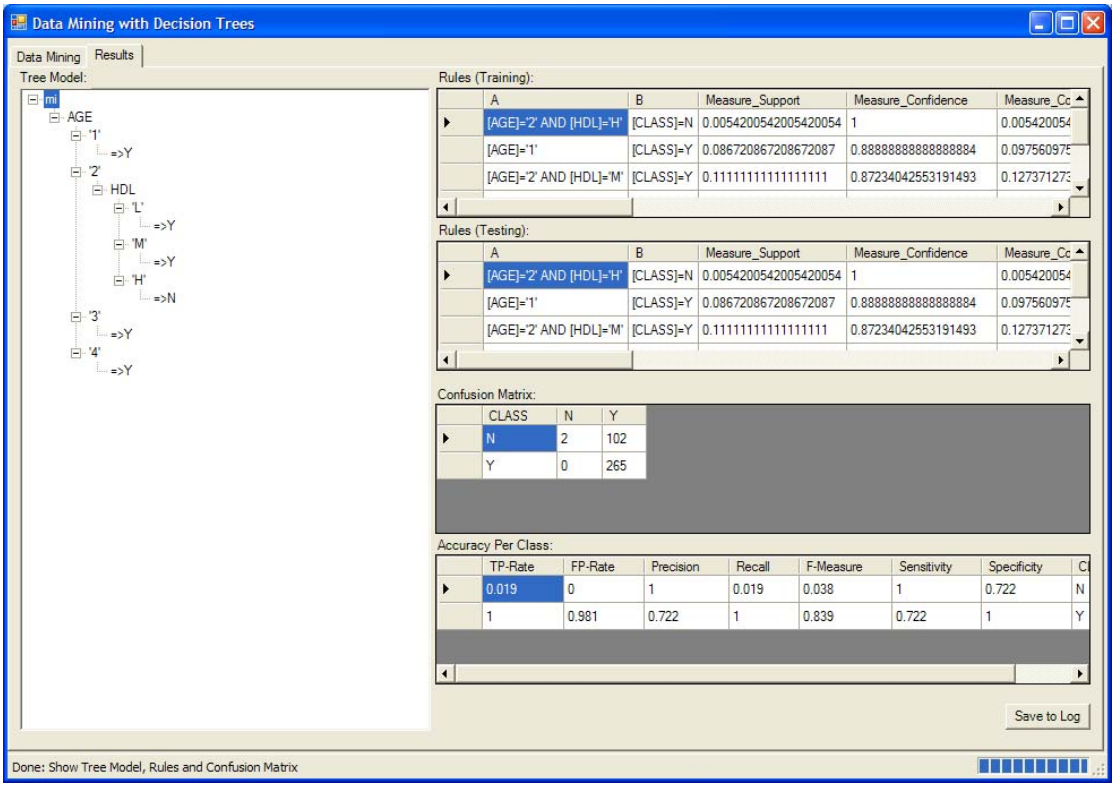


Figure 33: Appendix A - Result Screen

In the Result screen, it shows the decision tree on the left. On the right side, the rules are extracted from the training data set and from the testing data set. The confusion matrix and the accuracy per class are calculated using the testing data set.

You can save the results to a log file, when you click on the button “Save to Log”.

In order to return to the Option Screen, click on the tab “Data Mining”.

Appendix B

Table 67: Rules from Run #2 with Gini Gain

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10	Level 11	Level 12	Level 13	Level 14
SBP='L'	HDL='L'	TC='N'	Y										
SBP='L'	HDL='L'	TC='D'	TG='N'	N									
SBP='L'	HDL='L'	TC='D'	TG='H'	Y									
SBP='L'	HDL='L'	TC='D'	TG='D'	Y									
SBP='L'	HDL='L'	TC='H'	N										
SBP='L'	HDL='M'	Y											
SBP='L'	HDL='H'	TC='N'	Y										
SBP='L'	HDL='H'	TC='D'	N										
SBP='L'	HDL='H'	TC='H'	N										
SBP='N'	AGE='1'	Y											
SBP='N'	AGE='2'	TG='N'	Y										
SBP='N'	AGE='2'	TG='H'	Y										
SBP='N'	AGE='2'	TG='D'	HDL='L'	N									
SBP='N'	AGE='2'	TG='D'	HDL='M'	Y									
SBP='N'	AGE='2'	TG='D'	HDL='H'	Y									
SBP='N'	AGE='3'	HDL='L'	TG='N'	DM='N'	Y								
SBP='N'	AGE='3'	HDL='L'	TG='N'	DM='Y'	SMBEF='N'	Y							
SBP='N'	AGE='3'	HDL='L'	TG='N'	DM='Y'	SMBEF='Y'	N							
SBP='N'	AGE='3'	HDL='L'	TG='H'	Y									
SBP='N'	AGE='3'	HDL='L'	TG='D'	SEX='F'	Y								
SBP='N'	AGE='3'	HDL='L'	TG='D'	SEX='M'	N								
SBP='N'	AGE='3'	HDL='M'	HT='N'	TG='N'	Y								
SBP='N'	AGE='3'	HDL='M'	HT='N'	TG='H'	N								
SBP='N'	AGE='3'	HDL='M'	HT='N'	TG='D'	TC='N'	N							
SBP='N'	AGE='3'	HDL='M'	HT='N'	TG='D'	TC='D'	Y							
SBP='N'	AGE='3'	HDL='M'	HT='N'	TG='D'	TC='H'	Y							

SBP='N'	AGE='3'	HDL='M'	HT='Y'	Y									
SBP='N'	AGE='3'	HDL='H'	SEX='F'	N									
SBP='N'	AGE='3'	HDL='H'	SEX='M'	GLU='N'	Y								
SBP='N'	AGE='3'	HDL='H'	SEX='M'	GLU='H'	N								
SBP='N'	AGE='4'	HT='N'	HDL='L'	SEX='F'	N								
SBP='N'	AGE='4'	HT='N'	HDL='L'	SEX='M'	Y								
SBP='N'	AGE='4'	HT='N'	HDL='M'	GLU='N'	N								
SBP='N'	AGE='4'	HT='N'	HDL='M'	GLU='H'	Y								
SBP='N'	AGE='4'	HT='N'	HDL='H'	Y									
SBP='N'	AGE='4'	HT='Y'	Y										
SBP='H'	DBP='L'	N											
SBP='H'	DBP='N'	AGE='1'	TC='N'	N									
SBP='H'	DBP='N'	AGE='1'	TC='D'	Y									
SBP='H'	DBP='N'	AGE='1'	TC='H'	Y									
SBP='H'	DBP='N'	AGE='2'	GLU='N'	HDL='L'	Y								
SBP='H'	DBP='N'	AGE='2'	GLU='N'	HDL='M'	HT='N'	Y							
SBP='H'	DBP='N'	AGE='2'	GLU='N'	HDL='M'	HT='Y'	TG='N'	N						
SBP='H'	DBP='N'	AGE='2'	GLU='N'	HDL='M'	HT='Y'	TG='H'	Y						
SBP='H'	DBP='N'	AGE='2'	GLU='N'	HDL='M'	HT='Y'	TG='D'	Y						
SBP='H'	DBP='N'	AGE='2'	GLU='N'	HDL='H'	Y								
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='N'	N							
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='F'	Y						
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='N'	Y					
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='N'	FH='N'	N		
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='N'	FH='Y'	DM='N'	HT='N'	Y
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='N'	FH='Y'	DM='N'	HT='Y'	Y
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='N'	FH='Y'	DM='Y'	Y	
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='H'	Y			
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='D'	FH='N'	Y		
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='D'	FH='Y'	N		
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='N'	TG='H'	Y			

SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='D'	SEX='M'	SMBEF='Y'	LDL='D'	Y				
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='L'	TC='H'	N							
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='M'	SMBEF='N'	Y							
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='M'	SMBEF='Y'	TC='N'	Y						
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='M'	SMBEF='Y'	TC='D'	FH='N'	Y					
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='M'	SMBEF='Y'	TC='D'	FH='Y'	N					
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='M'	SMBEF='Y'	TC='H'	Y						
SBP='H'	DBP='N'	AGE='2'	GLU='H'	HDL='H'	Y								
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='N'	Y								
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='Y'	HT='N'	Y							
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='Y'	HT='Y'	TC='N'	Y						
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='Y'	HT='Y'	TC='D'	TG='N'	N					
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='Y'	HT='Y'	TC='D'	TG='H'	N					
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='Y'	HT='Y'	TC='D'	TG='D'	Y					
SBP='H'	DBP='N'	AGE='3'	GLU='N'	FH='Y'	HT='Y'	TC='H'	Y						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='N'	Y							
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='D'	HT='N'	N						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='D'	HT='Y'	HDL='L'	Y					
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='D'	HT='Y'	HDL='M'	DM='N'	Y				
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='D'	HT='Y'	HDL='M'	DM='Y'	N				
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='D'	HT='Y'	HDL='H'	Y					
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='N'	TC='H'	Y							
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='H'	DM='N'	Y							
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='H'	DM='Y'	SEX='F'	Y						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='H'	DM='Y'	SEX='M'	N						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='N'	N						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='F'	Y					
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='N'	Y				
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='Y'	SMBEF='N'	Y			
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='Y'	SMBEF='Y'	LDL='N'	HT='N'	Y	
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='Y'	SMBEF='Y'	LDL='N'	HT='Y'	DM='N'	Y

SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='Y'	SMBEF='Y'	LDL='N'	HT='Y'	DM='Y'	Y
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='Y'	SMBEF='Y'	LDL='H'	Y		
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='D'	SEX='M'	FH='Y'	SMBEF='Y'	LDL='D'	Y		
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='H'	Y						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='M'	N						
SBP='H'	DBP='N'	AGE='3'	GLU='H'	TG='D'	HDL='L'	TC='H'	N						
SBP='H'	DBP='N'	AGE='4'	HT='N'	TG='N'	N								
SBP='H'	DBP='N'	AGE='4'	HT='N'	TG='H'	N								
SBP='H'	DBP='N'	AGE='4'	HT='N'	TG='D'	Y								
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='L'	FH='N'	Y			
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='L'	FH='Y'	N			
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='M'	SEX='F'	Y			
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='M'	SEX='M'	FH='N'	N		
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='M'	SEX='M'	FH='Y'	Y		
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='H'	SEX='F'	N			
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='N'	HDL='H'	SEX='M'	Y			
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='H'	Y					
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='N'	TG='D'	Y					
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='N'	SMBEF='Y'	TG='N'	N					
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='N'	DM='Y'	Y							
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='H'	N								
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='D'	DM='N'	Y							
SBP='H'	DBP='N'	AGE='4'	HT='Y'	LDL='D'	DM='Y'	N							
SBP='H'	DBP='H'	AGE='1'	Y										
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='N'	N							
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='F'	Y					
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='N'	Y				
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='N'	Y			
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	GLU='N'	Y	
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	GLU='H'	HT='N'	Y
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	GLU='H'	HT='Y'	Y

SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='D'	LDL='H'	Y		
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='D'	LDL='D'	Y		
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='N'	SEX='M'	SMBEF='Y'	TC='H'	Y			
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='H'	DM='Y'	Y						
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='D'	GLU='N'	Y						
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='N'	TG='D'	GLU='H'	N						
SBP='H'	DBP='H'	AGE='2'	HDL='L'	FH='Y'	Y								
SBP='H'	DBP='H'	AGE='2'	HDL='M'	Y									
SBP='H'	DBP='H'	AGE='2'	HDL='H'	N									
SBP='H'	DBP='H'	AGE='3'	TC='N'	FH='N'	SEX='F'	Y							
SBP='H'	DBP='H'	AGE='3'	TC='N'	FH='N'	SEX='M'	N							
SBP='H'	DBP='H'	AGE='3'	TC='N'	FH='Y'	Y								
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='L'	TG='N'	FH='N'	N						
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='L'	TG='N'	FH='Y'	Y						
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='L'	TG='H'	N							
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='L'	TG='D'	Y							
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='M'	N								
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='H'	SEX='F'	N							
SBP='H'	DBP='H'	AGE='3'	TC='D'	HDL='H'	SEX='M'	Y							
SBP='H'	DBP='H'	AGE='3'	TC='H'	Y									
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='N'	HT='N'	Y							
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='N'	HT='Y'	N							
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='H'	SMBEF='N'	Y							
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='H'	SMBEF='Y'	TC='N'	Y						
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='H'	SMBEF='Y'	TC='D'	DM='N'	N					
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='H'	SMBEF='Y'	TC='D'	DM='Y'	Y					
SBP='H'	DBP='H'	AGE='4'	TG='N'	GLU='H'	SMBEF='Y'	TC='H'	Y						
SBP='H'	DBP='H'	AGE='4'	TG='H'	Y									
SBP='H'	DBP='H'	AGE='4'	TG='D'	Y									

Table 68: Rules from Run #3 with Distance Measure

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10	Level 11	Level 12	Level 13	Level 14
AGE='1'	GLU='N'	TG='N'	HDL='L'	Y									
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='N'	N								
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='Y'	FH='N'	SBP='L'	N						
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='Y'	FH='N'	SBP='N'	N						
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='Y'	FH='N'	SBP='H'	DBP='L'	Y					
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='Y'	FH='N'	SBP='H'	DBP='N'	Y					
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='Y'	FH='N'	SBP='H'	DBP='H'	N					
AGE='1'	GLU='N'	TG='N'	HDL='M'	SMBEF='Y'	FH='Y'	Y							
AGE='1'	GLU='N'	TG='N'	HDL='H'	N									
AGE='1'	GLU='N'	TG='H'	LDL='N'	N									
AGE='1'	GLU='N'	TG='H'	LDL='H'	Y									
AGE='1'	GLU='N'	TG='H'	LDL='D'	N									
AGE='1'	GLU='N'	TG='D'	N										
AGE='1'	GLU='H'	N											
AGE='2'	HDL='L'	SEX='F'	Y										
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='L'	Y								
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='N'	SBP='L'	Y					
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='N'	SBP='N'	N					
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='N'	SBP='H'	Y					
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='L'	N					
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='N'	Y			
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='D'	LDL='N'	HT='N'	DM='N'	Y
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='D'	LDL='N'	HT='N'	DM='Y'	Y
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='D'	LDL='N'	HT='Y'	Y	
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='D'	LDL='H'	Y		
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='D'	LDL='D'	Y		
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='N'	TC='H'	Y			
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='N'	FH='Y'	N				

AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='N'	SMBEF='Y'	SBP='H'	N					
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='N'	GLU='H'	N							
AGE='2'	HDL='L'	SEX='M'	TG='N'	DBP='H'	N								
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='N'	N								
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='N'	N							
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='L'	N						
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='N'	Y						
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='N'	DBP='L'	Y			
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='N'	DBP='N'	FH='N'	Y		
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='N'	DBP='N'	FH='Y'	HT='N'	Y	
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='N'	DBP='N'	FH='Y'	HT='Y'	DM='N'	Y
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='N'	DBP='N'	FH='Y'	HT='Y'	DM='Y'	Y
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='N'	DBP='H'	Y			
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='H'	Y				
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='N'	LDL='D'	Y				
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='D'	SBP='H'	GLU='H'	N					
AGE='2'	HDL='L'	SEX='M'	TG='H'	SMBEF='Y'	TC='H'	Y							
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='L'	Y								
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='N'	N								
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='N'	GLU='N'	N						
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='N'	GLU='H'	FH='N'	DBP='L'	Y				
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='N'	GLU='H'	FH='N'	DBP='N'	DM='N'	Y			
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='N'	GLU='H'	FH='N'	DBP='N'	DM='Y'	N			
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='N'	GLU='H'	FH='N'	DBP='H'	Y				
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='N'	GLU='H'	FH='Y'	Y					
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='H'	GLU='N'	Y						
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='H'	GLU='H'	DM='N'	N					
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='H'	GLU='H'	DM='Y'	Y					
AGE='2'	HDL='L'	SEX='M'	TG='D'	SBP='H'	LDL='D'	N							
AGE='2'	HDL='M'	DM='N'	SMBEF='N'	SBP='L'	N								

AGE='2'	HDL='M'	DM='N'	SMBEF='N'	SBP='N'	N								
AGE='2'	HDL='M'	DM='N'	SMBEF='N'	SBP='H'	GLU='N'	Y							
AGE='2'	HDL='M'	DM='N'	SMBEF='N'	SBP='H'	GLU='H'	N							
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='N'	N								
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='N'	GLU='N'	TG='N'	N					
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='N'	GLU='N'	TG='H'	LDL='N'	Y				
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='N'	GLU='N'	TG='H'	LDL='H'	N				
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='N'	GLU='N'	TG='H'	LDL='D'	Y				
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='N'	GLU='N'	TG='D'	N					
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='N'	GLU='H'	Y						
AGE='2'	HDL='M'	DM='N'	SMBEF='Y'	FH='Y'	HT='Y'	N							
AGE='2'	HDL='M'	DM='Y'	HT='N'	N									
AGE='2'	HDL='M'	DM='Y'	HT='Y'	TC='N'	LDL='N'	Y							
AGE='2'	HDL='M'	DM='Y'	HT='Y'	TC='N'	LDL='H'	N							
AGE='2'	HDL='M'	DM='Y'	HT='Y'	TC='N'	LDL='D'	Y							
AGE='2'	HDL='M'	DM='Y'	HT='Y'	TC='D'	Y								
AGE='2'	HDL='M'	DM='Y'	HT='Y'	TC='H'	N								
AGE='2'	HDL='H'	N											
AGE='3'	SBP='L'	HT='N'	N										
AGE='3'	SBP='L'	HT='Y'	TC='N'	Y									
AGE='3'	SBP='L'	HT='Y'	TC='D'	SEX='F'	Y								
AGE='3'	SBP='L'	HT='Y'	TC='D'	SEX='M'	N								
AGE='3'	SBP='L'	HT='Y'	TC='H'	Y									
AGE='3'	SBP='N'	HDL='L'	TC='N'	N									
AGE='3'	SBP='N'	HDL='L'	TC='D'	Y									
AGE='3'	SBP='N'	HDL='L'	TC='H'	Y									
AGE='3'	SBP='N'	HDL='M'	TG='N'	SMBEF='N'	N								
AGE='3'	SBP='N'	HDL='M'	TG='N'	SMBEF='Y'	GLU='N'	Y							
AGE='3'	SBP='N'	HDL='M'	TG='N'	SMBEF='Y'	GLU='H'	N							
AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='N'	SMBEF='N'	N							

AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='N'	SMBEF='Y'	TC='N'	GLU='N'	Y					
AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='N'	SMBEF='Y'	TC='N'	GLU='H'	N					
AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='N'	SMBEF='Y'	TC='D'	Y						
AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='N'	SMBEF='Y'	TC='H'	Y						
AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='H'	Y								
AGE='3'	SBP='N'	HDL='M'	TG='H'	LDL='D'	Y								
AGE='3'	SBP='N'	HDL='M'	TG='D'	GLU='N'	N								
AGE='3'	SBP='N'	HDL='M'	TG='D'	GLU='H'	Y								
AGE='3'	SBP='N'	HDL='H'	Y										
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='N'	SMBEF='N'	TG='N'	N						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='N'	SMBEF='N'	TG='H'	N						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='N'	SMBEF='N'	TG='D'	Y						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='N'	SMBEF='Y'	Y							
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='N'	N						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='H'	Y						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='F'	N					
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='N'	Y				
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='N'	Y			
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='D'	GLU='N'	Y		
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='D'	GLU='H'	DBP='L'	Y	
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='D'	GLU='H'	DBP='N'	HT='N'	Y
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='D'	GLU='H'	DBP='N'	HT='Y'	Y
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='D'	GLU='H'	DBP='H'	Y	
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='N'	TG='D'	SEX='M'	SMBEF='Y'	TC='H'	Y			
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='Y'	TG='N'	Y						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='Y'	TG='H'	N						
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='Y'	TG='D'	GLU='N'	Y					
AGE='3'	SBP='H'	LDL='N'	HDL='L'	FH='Y'	DM='Y'	TG='D'	GLU='H'	N					
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='N'	TG='N'	Y							
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='N'	TG='H'	Y							

AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='N'	TG='D'	N							
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='N'	N							
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='D'	FH='N'	GLU='N'	Y					
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='D'	FH='N'	GLU='H'	N					
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='D'	FH='Y'	GLU='N'	N					
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='D'	FH='Y'	GLU='H'	DBP='L'	Y				
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='D'	FH='Y'	GLU='H'	DBP='N'	N				
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='D'	FH='Y'	GLU='H'	DBP='H'	Y				
AGE='3'	SBP='H'	LDL='N'	HDL='M'	SMBEF='Y'	TC='H'	N							
AGE='3'	SBP='H'	LDL='N'	HDL='H'	N									
AGE='3'	SBP='H'	LDL='H'	TC='N'	N									
AGE='3'	SBP='H'	LDL='H'	TC='D'	N									
AGE='3'	SBP='H'	LDL='H'	TC='H'	Y									
AGE='3'	SBP='H'	LDL='D'	SMBEF='N'	Y									
AGE='3'	SBP='H'	LDL='D'	SMBEF='Y'	N									
AGE='4'	SMBEF='N'	HT='N'	Y										
AGE='4'	SMBEF='N'	HT='Y'	TG='N'	DBP='L'	N								
AGE='4'	SMBEF='N'	HT='Y'	TG='N'	DBP='N'	N								
AGE='4'	SMBEF='N'	HT='Y'	TG='N'	DBP='H'	SEX='F'	DM='N'	Y						
AGE='4'	SMBEF='N'	HT='Y'	TG='N'	DBP='H'	SEX='F'	DM='Y'	N						
AGE='4'	SMBEF='N'	HT='Y'	TG='N'	DBP='H'	SEX='M'	N							
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='L'	Y								
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='N'	Y								
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='H'	FH='N'	N							
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='H'	FH='Y'	Y							
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='D'	GLU='N'	DBP='L'	Y						
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='D'	GLU='N'	DBP='N'	Y						
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='D'	GLU='N'	DBP='H'	N						
AGE='4'	SMBEF='N'	HT='Y'	TG='H'	SBP='D'	GLU='H'	Y							
AGE='4'	SMBEF='Y'	TC='N'	TG='N'	Y									

AGE='4'	SMBEF='Y'	TC='N'	TG='H'	FH='N'	N								
AGE='4'	SMBEF='Y'	TC='N'	TG='H'	FH='Y'	Y								
AGE='4'	SMBEF='Y'	TC='N'	TG='D'	Y									
AGE='4'	SMBEF='Y'	TC='D'	SBP='L'	DM='N'	N								
AGE='4'	SMBEF='Y'	TC='D'	SBP='L'	DM='Y'	Y								
AGE='4'	SMBEF='Y'	TC='D'	SBP='N'	GLU='N'	HDL='L'	N							
AGE='4'	SMBEF='Y'	TC='D'	SBP='N'	GLU='N'	HDL='M'	Y							
AGE='4'	SMBEF='Y'	TC='D'	SBP='N'	GLU='N'	HDL='H'	Y							
AGE='4'	SMBEF='Y'	TC='D'	SBP='N'	GLU='H'	N								
AGE='4'	SMBEF='Y'	TC='D'	SBP='H'	HT='N'	HDL='L'	Y							
AGE='4'	SMBEF='Y'	TC='D'	SBP='H'	HT='N'	HDL='M'	TG='N'	N						
AGE='4'	SMBEF='Y'	TC='D'	SBP='H'	HT='N'	HDL='M'	TG='H'	Y						
AGE='4'	SMBEF='Y'	TC='D'	SBP='H'	HT='N'	HDL='M'	TG='D'	Y						
AGE='4'	SMBEF='Y'	TC='D'	SBP='H'	HT='N'	HDL='H'	Y							
AGE='4'	SMBEF='Y'	TC='D'	SBP='H'	HT='Y'	Y								
AGE='4'	SMBEF='Y'	TC='H'	TG='N'	N									
AGE='4'	SMBEF='Y'	TC='H'	TG='H'	Y									
AGE='4'	SMBEF='Y'	TC='H'	TG='D'	N									

Table 69: Rules from Run #1 with Gain Ratio

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10	Level 11	Level 12	Level 13	Level 14
DM='N'	DBP='L'	N											
DM='N'	DBP='N'	AGE='1'	FH='N'	HT='N'	SMBEF='N'	N							
DM='N'	DBP='N'	AGE='1'	FH='N'	HT='N'	SMBEF='Y'	Y							
DM='N'	DBP='N'	AGE='1'	FH='N'	HT='Y'	N								
DM='N'	DBP='N'	AGE='1'	FH='Y'	Y									
DM='N'	DBP='N'	AGE='2'	SMBEF='N'	GLU='N'	N								
DM='N'	DBP='N'	AGE='2'	SMBEF='N'	GLU='H'	Y								
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='F'	Y								
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='L'	Y				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='N'	TC='N'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='N'	TC='D'	LDL='N'	HT='N'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='N'	TC='D'	LDL='N'	HT='Y'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='N'	TC='D'	LDL='H'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='N'	TC='D'	LDL='D'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='N'	TC='H'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='N'	FH='Y'	Y			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='N'	SBP='H'	Y				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='N'	Y				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='N'	Y			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='D'	LDL='N'	SBP='L'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='D'	LDL='N'	SBP='N'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='D'	LDL='N'	SBP='H'	HT='N'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='D'	LDL='N'	SBP='H'	HT='Y'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='D'	LDL='H'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='D'	LDL='D'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='L'	GLU='H'	FH='Y'	TC='H'	Y			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='N'	TC='N'	N				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='N'	TC='D'	FH='N'	Y			

DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='N'	TC='D'	FH='Y'	N			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='N'	TC='H'	Y				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='N'	Y				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='N'	GLU='N'	SBP='L'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='N'	GLU='N'	SBP='N'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='N'	GLU='N'	SBP='H'	FH='N'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='N'	GLU='N'	SBP='H'	FH='Y'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='N'	GLU='H'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='H'	Y			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='D'	LDL='D'	Y			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='M'	HT='Y'	TC='H'	Y				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='N'	HDL='H'	Y						
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='N'	Y						
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='L'	N				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='N'	N				
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='L'	LDL='N'	FH='N'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='L'	LDL='N'	FH='Y'	HT='N'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='L'	LDL='N'	FH='Y'	HT='Y'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='L'	LDL='H'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='L'	LDL='D'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='M'	LDL='N'	FH='N'	HT='N'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='M'	LDL='N'	FH='N'	HT='Y'	Y
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='M'	LDL='N'	FH='Y'	Y	
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='M'	LDL='H'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='M'	LDL='D'	Y		
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='N'	SBP='H'	HDL='H'	Y			
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='D'	GLU='H'	Y					
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='H'	TC='H'	Y						
DM='N'	DBP='N'	AGE='2'	SMBEF='Y'	SEX='M'	TG='D'	Y							
DM='N'	DBP='N'	AGE='3'	SEX='F'	GLU='N'	N								

DM='N'	DBP='N'	AGE='3'	SEX='F'	GLU='H'	SMBEF='N'	Y							
DM='N'	DBP='N'	AGE='3'	SEX='F'	GLU='H'	SMBEF='Y'	N							
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='N'	HDL='L'	HT='N'	Y					
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='N'	HDL='L'	HT='Y'	TG='N'	N				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='N'	HDL='L'	HT='Y'	TG='H'	Y				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='N'	HDL='L'	HT='Y'	TG='D'	Y				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='N'	HDL='M'	Y						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='N'	HDL='H'	Y						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='H'	N							
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='N'	LDL='D'	Y							
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='N'	Y						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='L'	Y					
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='N'	Y				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='H'	GLU='N'	SBP='L'	Y		
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='H'	GLU='N'	SBP='N'	Y		
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='H'	GLU='N'	SBP='H'	FH='N'	HT='N'	Y
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='H'	GLU='N'	SBP='H'	FH='N'	HT='Y'	Y
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='H'	GLU='N'	SBP='H'	FH='Y'	Y	
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='H'	GLU='H'	Y			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='M'	LDL='D'	Y				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='H'	HDL='H'	Y					
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='N'	TG='D'	Y						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='N'	SBP='L'	N				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='N'	SBP='N'	Y				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='N'	SBP='H'	HDL='L'	N			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='N'	SBP='H'	HDL='M'	N			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='N'	SBP='H'	HDL='H'	Y			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='L'	Y				
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='L'	N			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='M'	HT='N'	N		

DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='M'	HT='Y'	TG='N'	GLU='N'	Y
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='M'	HT='Y'	TG='N'	GLU='H'	Y
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='M'	HT='Y'	TG='H'	Y	
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='M'	HT='Y'	TG='D'	Y	
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='N'	HDL='H'	N			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='N'	Y			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='H'	Y			
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='D'	HDL='L'	GLU='N'	N	
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='D'	HDL='L'	GLU='H'	HT='N'	N
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='D'	HDL='L'	GLU='H'	HT='Y'	N
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='D'	HDL='M'	N		
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='N'	FH='Y'	SBP='H'	TG='D'	HDL='H'	N		
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='H'	N						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='D'	LDL='D'	N						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='H'	SBP='L'	Y						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='H'	SBP='N'	N						
DM='N'	DBP='N'	AGE='3'	SEX='M'	SMBEF='Y'	TC='H'	SBP='H'	Y						
DM='N'	DBP='N'	AGE='4'	FH='N'	SEX='F'	TG='N'	N							
DM='N'	DBP='N'	AGE='4'	FH='N'	SEX='F'	TG='H'	Y							
DM='N'	DBP='N'	AGE='4'	FH='N'	SEX='F'	TG='D'	N							
DM='N'	DBP='N'	AGE='4'	FH='N'	SEX='M'	N								
DM='N'	DBP='N'	AGE='4'	FH='Y'	TG='N'	Y								
DM='N'	DBP='N'	AGE='4'	FH='Y'	TG='H'	N								
DM='N'	DBP='N'	AGE='4'	FH='Y'	TG='D'	HDL='L'	N							
DM='N'	DBP='N'	AGE='4'	FH='Y'	TG='D'	HDL='M'	Y							
DM='N'	DBP='N'	AGE='4'	FH='Y'	TG='D'	HDL='H'	Y							
DM='N'	DBP='H'	SBP='L'	SMBEF='N'	AGE='1'	Y								
DM='N'	DBP='H'	SBP='L'	SMBEF='N'	AGE='2'	Y								
DM='N'	DBP='H'	SBP='L'	SMBEF='N'	AGE='3'	Y								
DM='N'	DBP='H'	SBP='L'	SMBEF='N'	AGE='4'	N								

DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='F'	Y								
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='L'	N							
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='M'	AGE='1'	N						
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='M'	AGE='2'	TC='N'	N					
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='M'	AGE='2'	TC='D'	Y					
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='M'	AGE='2'	TC='H'	Y					
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='M'	AGE='3'	N						
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='M'	AGE='4'	N						
DM='N'	DBP='H'	SBP='L'	SMBEF='Y'	SEX='M'	HDL='H'	N							
DM='N'	DBP='H'	SBP='N'	AGE='1'	Y									
DM='N'	DBP='H'	SBP='N'	AGE='2'	Y									
DM='N'	DBP='H'	SBP='N'	AGE='3'	Y									
DM='N'	DBP='H'	SBP='N'	AGE='4'	N									
DM='N'	DBP='H'	SBP='H'	AGE='1'	GLU='N'	N								
DM='N'	DBP='H'	SBP='H'	AGE='1'	GLU='H'	Y								
DM='N'	DBP='H'	SBP='H'	AGE='2'	SMBEF='N'	N								
DM='N'	DBP='H'	SBP='H'	AGE='2'	SMBEF='Y'	TG='N'	Y							
DM='N'	DBP='H'	SBP='H'	AGE='2'	SMBEF='Y'	TG='H'	TC='N'	N						
DM='N'	DBP='H'	SBP='H'	AGE='2'	SMBEF='Y'	TG='H'	TC='D'	Y						
DM='N'	DBP='H'	SBP='H'	AGE='2'	SMBEF='Y'	TG='H'	TC='H'	Y						
DM='N'	DBP='H'	SBP='H'	AGE='2'	SMBEF='Y'	TG='D'	Y							
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='F'	Y								
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='L'	N						
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='M'	LDL='N'	FH='N'	N				
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='M'	LDL='N'	FH='Y'	TC='N'	N			
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='M'	LDL='N'	FH='Y'	TC='D'	Y			
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='M'	LDL='N'	FH='Y'	TC='H'	Y			
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='M'	LDL='H'	Y					
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='M'	LDL='D'	Y					
DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='N'	HDL='H'	Y						

DM='N'	DBP='H'	SBP='H'	AGE='3'	SEX='M'	GLU='H'	N							
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='N'	GLU='N'	Y							
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='N'	GLU='H'	FH='N'	Y						
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='N'	GLU='H'	FH='Y'	SEX='F'	N					
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='N'	GLU='H'	FH='Y'	SEX='M'	TG='N'	Y				
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='N'	GLU='H'	FH='Y'	SEX='M'	TG='H'	N				
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='N'	GLU='H'	FH='Y'	SEX='M'	TG='D'	Y				
DM='N'	DBP='H'	SBP='H'	AGE='4'	SMBEF='Y'	N								
DM='Y'	AGE='1'	Y											
DM='Y'	AGE='2'	LDL='N'	TG='N'	N									
DM='Y'	AGE='2'	LDL='N'	TG='H'	FH='N'	Y								
DM='Y'	AGE='2'	LDL='N'	TG='H'	FH='Y'	N								
DM='Y'	AGE='2'	LDL='N'	TG='D'	N									
DM='Y'	AGE='2'	LDL='H'	HT='N'	Y									
DM='Y'	AGE='2'	LDL='H'	HT='Y'	HDL='L'	N								
DM='Y'	AGE='2'	LDL='H'	HT='Y'	HDL='M'	SMBEF='N'	N							
DM='Y'	AGE='2'	LDL='H'	HT='Y'	HDL='M'	SMBEF='Y'	Y							
DM='Y'	AGE='2'	LDL='H'	HT='Y'	HDL='H'	N								
DM='Y'	AGE='2'	LDL='D'	Y										
DM='Y'	AGE='3'	TG='N'	GLU='N'	HDL='L'	Y								
DM='Y'	AGE='3'	TG='N'	GLU='N'	HDL='M'	N								
DM='Y'	AGE='3'	TG='N'	GLU='N'	HDL='H'	N								
DM='Y'	AGE='3'	TG='N'	GLU='H'	HDL='L'	N								
DM='Y'	AGE='3'	TG='N'	GLU='H'	HDL='M'	N								
DM='Y'	AGE='3'	TG='N'	GLU='H'	HDL='H'	SEX='F'	N							
DM='Y'	AGE='3'	TG='N'	GLU='H'	HDL='H'	SEX='M'	Y							
DM='Y'	AGE='3'	TG='H'	TC='N'	DBP='L'	Y								
DM='Y'	AGE='3'	TG='H'	TC='N'	DBP='N'	Y								
DM='Y'	AGE='3'	TG='H'	TC='N'	DBP='H'	N								
DM='Y'	AGE='3'	TG='H'	TC='D'	N									

DM='Y'	AGE='3'	TG='H'	TC='H'	N									
DM='Y'	AGE='3'	TG='D'	LDL='N'	Y									
DM='Y'	AGE='3'	TG='D'	LDL='H'	Y									
DM='Y'	AGE='3'	TG='D'	LDL='D'	N									
DM='Y'	AGE='4'	HDL='L'	SEX='F'	N									
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='L'	N								
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='N'	Y								
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='H'	DBP='L'	N							
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='H'	DBP='N'	N							
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='H'	DBP='H'	FH='N'	TC='N'	N					
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='H'	DBP='H'	FH='N'	TC='D'	Y					
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='H'	DBP='H'	FH='N'	TC='H'	Y					
DM='Y'	AGE='4'	HDL='L'	SEX='M'	SBP='H'	DBP='H'	FH='Y'	N						
DM='Y'	AGE='4'	HDL='M'	N										
DM='Y'	AGE='4'	HDL='H'	DBP='L'	N									
DM='Y'	AGE='4'	HDL='H'	DBP='N'	N									
DM='Y'	AGE='4'	HDL='H'	DBP='H'	Y									

Table 70: Rules from Run with 6 classes with Information Gain

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
AGE='1'	D							
AGE='2'	SMBEF='N'	A						
AGE='2'	SMBEF='Y'	DM='N'	D					
AGE='2'	SMBEF='Y'	DM='Y'	C					
AGE='3'	DBP='L'	A						
AGE='3'	DBP='N'	TC='N'	HDL='L'	TG='N'	GLU='N'	FGH		
AGE='3'	DBP='N'	TC='N'	HDL='L'	TG='N'	GLU='H'	A		
AGE='3'	DBP='N'	TC='N'	HDL='L'	TG='H'	D			
AGE='3'	DBP='N'	TC='N'	HDL='L'	TG='D'	C			
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='N'	GLU='N'	FGH		
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='N'	GLU='H'	SBP='L'	A	
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='N'	GLU='H'	SBP='N'	A	
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='N'	GLU='H'	SBP='H'	D	
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='H'	D			
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='D'	LDL='N'	B		
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='D'	LDL='H'	FGH		
AGE='3'	DBP='N'	TC='N'	HDL='M'	TG='D'	LDL='D'	B		
AGE='3'	DBP='N'	TC='N'	HDL='H'	SBP='L'	C			
AGE='3'	DBP='N'	TC='N'	HDL='H'	SBP='N'	C			
AGE='3'	DBP='N'	TC='N'	HDL='H'	SBP='H'	D			
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='N'	SEX='F'	SBP='L'	FGH
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='N'	SEX='F'	SBP='N'	FGH
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='N'	SEX='F'	SBP='H'	E
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='N'	SEX='M'	D	
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='Y'	SMBEF='N'	HDL='L'	A
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='Y'	SMBEF='N'	HDL='M'	D
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='Y'	SMBEF='N'	HDL='H'	D
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='N'	HT='Y'	SMBEF='Y'	E	

AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='L'	D		
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='N'	SMBEF='N'	B	
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='N'	SMBEF='Y'	HDL='L'	A
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='N'	SMBEF='Y'	HDL='M'	FGH
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='N'	SMBEF='Y'	HDL='H'	FGH
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='H'	SMBEF='N'	FH='N'	D
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='H'	SMBEF='N'	FH='Y'	FGH
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='H'	SMBEF='Y'	FH='N'	E
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='H'	SBP='H'	SMBEF='Y'	FH='Y'	D
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='D'	SEX='F'	D		
AGE='3'	DBP='N'	TC='D'	DM='N'	TG='D'	SEX='M'	B		
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='N'	GLU='N'	HDL='L'	D	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='N'	GLU='N'	HDL='M'	E	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='N'	GLU='N'	HDL='H'	D	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='N'	GLU='H'	HDL='L'	E	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='N'	GLU='H'	HDL='M'	A	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='N'	GLU='H'	HDL='H'	E	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='H'	C			
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='D'	HT='N'	B		
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='D'	HT='Y'	SEX='F'	D	
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='D'	HT='Y'	SEX='M'	HDL='L'	E
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='D'	HT='Y'	SEX='M'	HDL='M'	D
AGE='3'	DBP='N'	TC='D'	DM='Y'	TG='D'	HT='Y'	SEX='M'	HDL='H'	E
AGE='3'	DBP='N'	TC='H'	E					
AGE='3'	DBP='H'	LDL='N'	C					
AGE='3'	DBP='H'	LDL='H'	HDL='L'	B				
AGE='3'	DBP='H'	LDL='H'	HDL='M'	A				
AGE='3'	DBP='H'	LDL='H'	HDL='H'	D				
AGE='3'	DBP='H'	LDL='D'	SEX='F'	D				
AGE='3'	DBP='H'	LDL='D'	SEX='M'	TC='N'	C			

AGE='3'	DBP='H'	LDL='D'	SEX='M'	TC='D'	A			
AGE='3'	DBP='H'	LDL='D'	SEX='M'	TC='H'	A			
AGE='4'	SBP='L'	A						
AGE='4'	SBP='N'	A						
AGE='4'	SBP='H'	LDL='N'	E					
AGE='4'	SBP='H'	LDL='H'	E					
AGE='4'	SBP='H'	LDL='D'	A					