Predict Quality of Life of Patients with Coronary Heart Disease in Related to Risk Factors Using Data Mining Techniques

Saparamadu D.D.N.A.

2019

Predict Quality of Life of Patients with Coronary Heart Disease in Related to Risk Factors Using Data Mining Techniques

A dissertation submitted for the Degree of Master of Computer Science

Saparamadu D.D.N.A
University of Colombo School of Computing
2019

DECLARATION

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name : **Saparamadu D.D.N.A.**
Registration Number: **2016MCS101**
Index Number : **16441017**

Student's Signature: ……………………… Date: 25/09/2019

This is to certify that this thesis is based on the work of **Ms. Saparamadu D.D.N.A.** under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:
Supervisor Name: **Dr. M. G. Noel A. S. Fernando**

Supervisor's Signature: ……………………… Date: 25/09/2019

ACKNOWLEDGEMENT

As a student of the Master of Computer Science, University of Colombo School of Computing, Sri Lanka, I would like to express my gratitude to the University Of Colombo School Of Computing and also the people who helped me to complete this project throughout the final year successfully.

It is my great pleasure to express my special thanks of gratitude to my supervisor Dr. M. G. Noel A. S. Fernando, Senior Lecturer, University of Colombo School of Computing who gave me the golden opportunity to conduct this wonderful project on the topic of Predict quality of life of patients with coronary heart disease in related to risk factors using Data Mining Techniques. Thank you sir for guiding me from the beginning towards the end of the project with such attention and by spending your valuable time.

Secondly I would like to thank my parents who helped me a lot by encouraging and standing beside me the whole time. A special thanks goes to my husband Dr. Kalana Rathnayake, Medical Officer, Coronary Care Unit, District General Hospital, Gampaha, who was a main advisor for the project and he guided me on handling medical related data and theoretical, clinical medical knowledge.

I would like to express my gratitude towards Director, District General Hospital, Gampaha and Dr. Janath P. Gunasinghe, Consultant Cardiologist, Coronary Care Unit, District General Hospital, Gampaha for authorizing the collection of data from Cardiology Clinic. My heartiest thank goes to all the Doctors of Coronary Care Unit for supporting me with their valuable consultations and Nursing staff, Minor staff for helping me throughout the process of data collection.

Finally my special gratitude towards the most important resource persons, the patients of Cardiology Clinic, District General Hospital, Gampaha for providing their confidential and sensible details for this project.

ABSTRACT

There are numerous types of diseases identified worldwide and those are still emerging. New generations as well as older generations are suffering from these illnesses, which could be disastrous which could take lives sometimes. This situation occurs due to reasons such as lesser knowledge on the domain and inappropriate life styles. The Coronary Heart Disease (CHD) is one of the leading diseases, which diagnosed in many Sri Lankans. Nowadays, many number of individuals are spotted with this disease and number of patients suffers and live with the disease throughout their life. This study was proposed to find out the state of patients following CHD. The current situation of the patients, their weaknesses and expectations will be addressed and to find out whether they spend a quality life or not. Responsible institutions such as Hospitals, Medical Centers generate and use huge amount of data related to these diseases daily. These data could be used to educate illness and find solutions to heal or to control them in proper manner.

Factors affecting on CHD and suitable data mining techniques were identified through a thorough study on the existing researches in the Sri Lankan context and globally. Based on the factors identified the data set was collected from a government hospital in Sri Lanka and sent through the process of data mining in order to discover knowledge by using the identified data mining techniques.

A set of strong patterns and relationships among the data items were found and finally through a thorough discussion of the results found, the conclusion was developed as the patients following CHD has a good quality of life as they have a good mental and physical state.

Contents

List of Figures

List of Abbreviations

CD	Cerebrovascular Disease
CHD	Coronary Heart Disease
CVD	Cardio Vascular Disease
GDP	Gross Domestic Product
HDL	High Density Lipoprotein
HRQOL	Health Related Quality Of Life
LDL	Low Density Lipoprotein
PCI	Percutaneous Coronary Intervention
QOL	Quality Of Life
RHD	Rheumatic Heart disease
SVM	Support Vector Machine
WAC	Weighted Associative Classifier
WHO	World Health Organization

Chapter 1

INTRODUCTION

1.1 Problem Domain

According to World Health Organization (WHO), Cardio Vascular Diseases (CVDs) account the lives of 17.7 million every year, 31% of all global deaths. More than 75% of CVD deaths occur in low-income and middle-income countries whereas 80% of all CVD deaths are due to Heart attacks and Strokes [1]. CVDs are disorders of the heart and blood vessels and include Coronary Heart Disease (CHD), Cerebrovascular Disease (CD), Rheumatic Heart Disease (RHD) and other conditions. Among them CHD plays a major role.

As major triggering factors of Coronary heart disease the modifiable risk factors are tobacco smoking and the harmful use of alcohol, unhealthy diet, physical inactivity, Stress, living style and non-communicable diseases namely Diabetes mellitus, Hypertension and Hyperlipidemia. Apart from the modifiable risk factors, age, gender and first degree family history (parents of the patient) have been identified as the non-modifiable risk factors which mainly influence on the Coronary heart disease [1]. Therefor appropriate optimization of above mentioned risk factors may reduce the ratio of occurrences of Coronary heart disease.

According to the Sri Lanka state of the economy 2016, Sri Lanka is a middle-income country where the total health expenditure is 3.5% of Gross Domestic Product (GDP) in 2014 [2]. Over decades repeatedly, Sri Lankan health care system runs with a limited number of resources (Human resource/Non-human resource), with an unbearable number of patients compared to total health expenditure. Hence the requirement of managing the available resources efficiently has become a serious matter.

The quality of life of patients depend on the properties like lifespan, number of complications and recovery speed to the regular life style which anonymously determine well-being of individuals and the overall society [3]. Quality of life implies less number (occurrences) of complications, longer life span and recovery speed (time duration) to a regular life style [4]. Hence a proper guidance for the patient is a major requirement.

Thus all these problems indicate the need of a solution for providing guidance to mitigate the CVD considering the non-communicable risk factors and modifiable risk factors with related to the Sri Lankan context. The background study was conducted to cater aspects such as identifying risk factors related to the Coronary Heart Disease and to identify correlations between the identified factors. Furthermore evaluate the factors affecting quality of life with related to Health and identifying the scope and the boundaries of the domain. The methodology to be followed has to be constructed based on the existing plans trailed by the other studies. Finding out the tools and techniques is another challenge to be address in the background study.

1.2 Problem

The main challenge in the healthcare sector is to provide a quality service at an affordable cost. The patient management happens based on the doctor's experience and the intuition thus human errors can be exist and sometimes poor clinical decisions can lead to disastrous consequences with unwanted biases, errors and excessive medical costs. This process finally affects the quality of service provided to patient [4].

Patient's life style with the disease completely transformed in to a new life due to many reasons as changes in day to day activities, changed mentality, changing expenses on drugs, medical tests and treatments. This change entirely affects the rest of the life of the patient. Most of the patients are elders who might not have a proper and sufficient income and sometimes they might depend on their children. Ultimately this problem throws the patient life in to an uncertain situation [5].

1.3 Motivation

A huge amount of complex and voluminous data is produced by healthcare transactions each day[1]. These data could be converted in to useful information which is valuable in decision-making which is a critical aspect of health care sector. Generating the information by processing and analyzing have become complicated tasks as far as the collection of data size is huge. However the decision making is the base for resource management, cost management and clinical practice management thus each of these are interrelated. As an example effective treatment helps to reduce treatment costs.

Finding a suitable mechanism to manage this complex and voluminous data in the health care is an interesting task since, it would give solutions for many problems in clinical decision support such as reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome [4].

The main concern is to provide guidance for patients to prevent from the complications following CHD which will lead to a quality life.

1.4 The exact computer science problem

Management of the patient happens based on the doctor's decisions, other than using the knowledge rich data hidden in the huge data set generated. This messy data set should be cleaned, well organized in order to make use of it. This data can be categories in to two types such as Patient centric data, resource management data [4].

The field of Information technology has the ability to store and manipulate huge data in different extents. Data mining or Knowledge discovery is the process of discovering hidden useful knowledge from massive databases. This process demonstrates a well-defined structure combined with different techniques.

The main computer science related concern of this research is to find out suitable data mining techniques needed to extract the correct knowledge. Thus based on the risk factors of the patient a research process is needed to find out the precise data mining tools or techniques which will output the level of quality of life of the patient [4].

1.5 Objectives

- Identify the risk factors related to quality of life of CHD patients with related to Sri Lankan environment using literature review
- Investigate how these risk factors effect on the individuals using data mining process
- Propose solutions to mitigate risk factors leading for a quality life by evaluating the patterns

1.6 Scope

Most of the previous studies on this domain have identified a set of risk factors. But most of these studies have done in overseas countries. Thus the first challenge is to relate these identified risk factors in to Sri Lankan context.

Elements such as environmental conditions, economical status, educational background, life style and habits of the Sri Lankans have to be taken under the consideration when undertaking such a mapping.

Other than the globally identified risk factors, there can be special risk factors conventional to Sri Lankans (Such as special hobbies). Hence the next step is identifying those special risk factors.

Related data and information are collected from hospital staff, Coronary Care Unit (CCU), Cardiology clinic, Clinical records inventory at District General hospital, Gampaha, Sri Lanka.

The goal is to determine and predict quality of life of patients who had coronary heart disease in related to modifiable risk factors, age category and gender.

1.7 Research contribution

The target data set is collected through a series of face to face interviews with the doctors as well as patients who currently follow up monthly in the hospital clinic and by using a well-defined questionnaire for both the parties. The arrangements have been made to collect data with the permission of the Director and the consultant cardiologist, district general hospital, Gamaha.

Thereby the collected data is categorized based on the modifiable risk factors which were identified as major triggering factors for coronary heart disease with age and gender of the each individual. Almost all the data containing major risk factors as well as the lifestyle risk factors were collected through clinical records and patients feedbacks.

The collected data are sent through a chain of analysis (the process of knowledge discovery) in a define time frame and built up hidden and significant patterns and relationships. Initially to identify the patterns of the data, the data is sent through the processes Cleaning & Integration, Selection & Transformation and Data Mining. Finally the Identified patterns are sent through the process of Evaluation & Presentation in order to achieve the Knowledge.

1.8 Remaining Chapters

The chapter literature review describe about the background of the domain thoroughly and the main focus will occupied by existing similar researches found during the study. The scope and context of each project, areas covered, techniques and tools used will be described in detail. The chapter on methodology describe the phases of the process or the approach to be followed to reach the goal stepwise. Thus the target data, analyzing techniques and tools are defined clearly. The chapter on evaluation and results is used to indicate hypothesis constructed. The results such as graphs, models and statistical values are illustrated throughout this chapter and furthermore each result is interpreted in detail for a complete and accurate conclusion. The chapter on conclusion and future work describes the ultimate result and proposed work for future are listed down in this chapter.

Chapter 2

LITERATURE REVIEW

2.1 Coronary Heart Disease

Finding the risk factors related to the Coronary Heart Disease with the context of Sri Lankans has done by physicians from early nineties [6]. One of the articles which were conducted over thirteen years has named the risk factors as Hypertension, Smoking and Cholesterol. They have found that Sri Lankans have high hypertension compared to Americans.

High salt intake and coconut usage, less use of fruits and vegetables, irregular Exercise have been identified as the major reasons for the above fact. The research results show that even though the Japanese have a good health condition when they migrate in to a different country like America the health condition changes in a negative manner. Hence the diet and Lifestyle (exercise) matters. The health condition changes due to the combination of the factors and they say since the people in Japan has a lower level of cholesterol the smoking causes little risk to the heart when compared to people in South Asia. According to community hypertension evaluation clinic programs' data analysis, they have identified that the dense of CHD in Sri Lankans are prominent when compared to Americans. One of the possible causative factors could be high daily dietary salt intake with the average of 8.7g. Hence prevalence of hypertension and salt intake has a linear acceleration according to this study.

With compared to western population Sri Lankans have lower levels of mean high density lipoprotein (HDL) cholesterol levels which were identified as a protective factor for prevalence of CHD. This can be due to Sri Lankan diet or genetic relationships. Reducing coconut fat in the diet has been suggested as an effective tip of lowering the blood cholesterol level. It was proven by the Framinhgam study[7] that, Japanese and other populations with low plasma cholesterol levels tend to have higher rates of CHD after migrating to USA confirming that the low incidence is not primarily genetic among them. The noble price winning research [8] by Brown and Goldstein says elevated Low density lipoprotein (LDL) itself be a sufficient cause of CHD. More overly Lipid Research Clinic Coronary Primary Prevention Trial proves that reducing cholesterol levels concomitantly reduces the risk of CHD. Furthermore abnormal blood lipids such as low HDL in cholesterol raised LDL and triglycerides as the major cause of mortality due to CVD[9].

It is found to have current smokers have a threefold increase in CHD comparable with men who had never smoked according to the British regional heart study. Even if the cigarette smoking is a considerable risk factor, incidence of risk factors for CHD varies from country to country even salt intake and blood cholesterol levels are at a lower level. This study has considered the countries Japan and Yugoslavia where the blood cholesterol level is at a lower level hence the smoking cause little risk to the heart than Sri Lanka. Many sociocultural factors which direct the community norms, attitudes and believe towards smoking have been identified as high prevalence of tobacco smoking.

Finally they have concluded that the largely modifiable sociocultural factors such as improving nutrition, elimination of smoking, control of hypertension and blood cholesterol levels, promotion of a healthy life style including graded exercise and controlling obesity as well as political change are the protective factors for CHD. There is a correlation between the socio economy and CVD since the studies have proved that if the lower level of socio-economy position is there the patient is vulnerable to CVD and the mortality level gets high hence they die earlier [9].

Apart from the above identified risk factors of CHD one of the studies have identified the age, gender, chest pain, patient history, family history, cholesterol, diabetes, blood pressure, fasting blood sugar, resting ECG, Maximum heart rate as some of the reliable indicators which they suggest to use for the prediction [4].

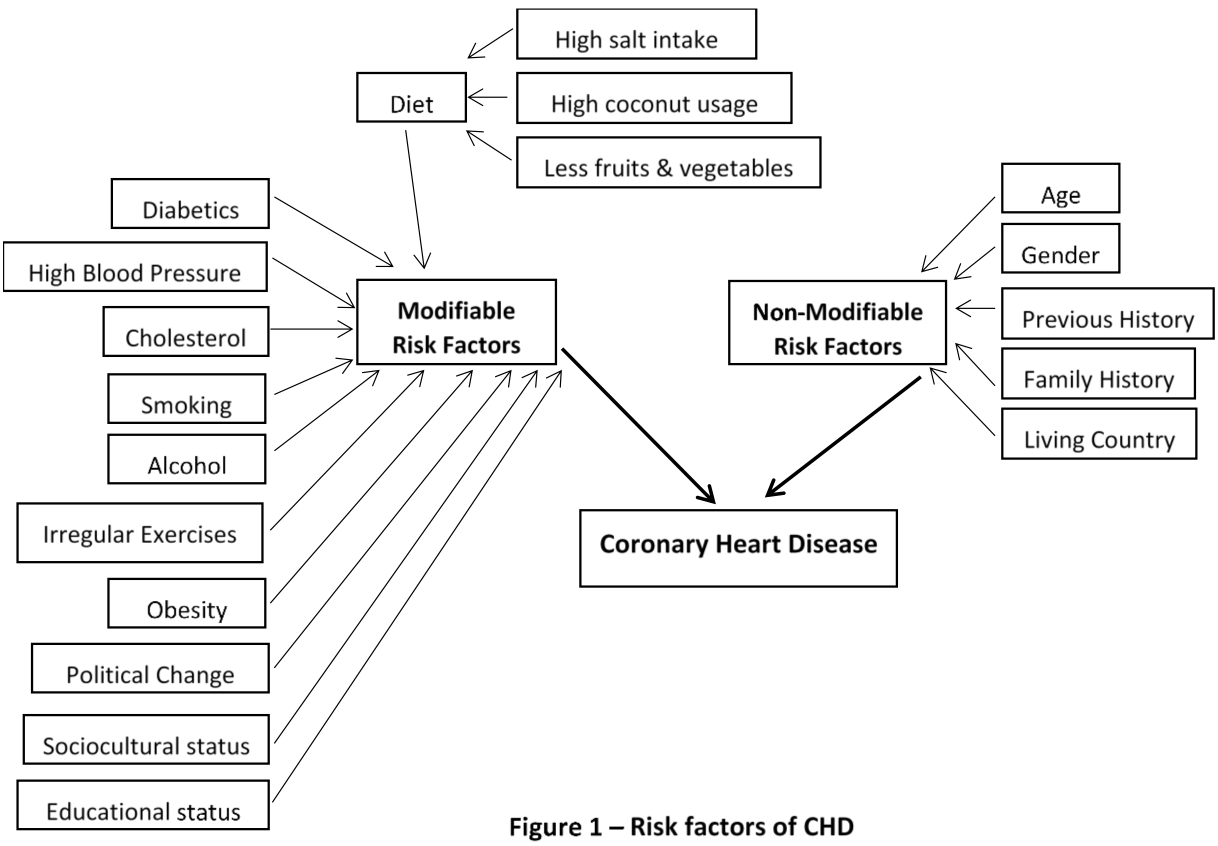


Figure 1 – Risk factors of CHD

The Figure 1 demonstrates the identified risk factors for CHD and their correlations found from the above discussed studies.

2.2 Health Related Quality of Life

The topic of maintaining and improving Quality Of Life (QOL) is one of the leading ongoing discussions done by many researches. A research conducted by TURKEY University proves that the CHD has association with the QOL with related to health (HRQOL) [10]. Socio-demographic characteristics, Disease and treatment specific variable and Support system during the disease course had been taken as the areas to be covered using the questionnaire. They have found gender as a less determinant for QOL. Marital status, educational status, social & economic status, living area (Urban or Rural), smoking or alcohol, medical treatments and heart attacks have been identified as the facts for QOL.

Even though bypass surgery/ Percutaneous Coronary Intervention (PCI) is proved to be not affected to the domain, since they effect on the mental health the fact cannot be ignored. Anxiety and depression, difficult in daily works are the reasons for considering the heart attack patients as a fact. Medical treatments are more toward the patient mental health since they feel better and safe when the number of medical treatments arises. Hence the above study declares HRQOL as the effect of the disease and the treatment perceived by the patient.

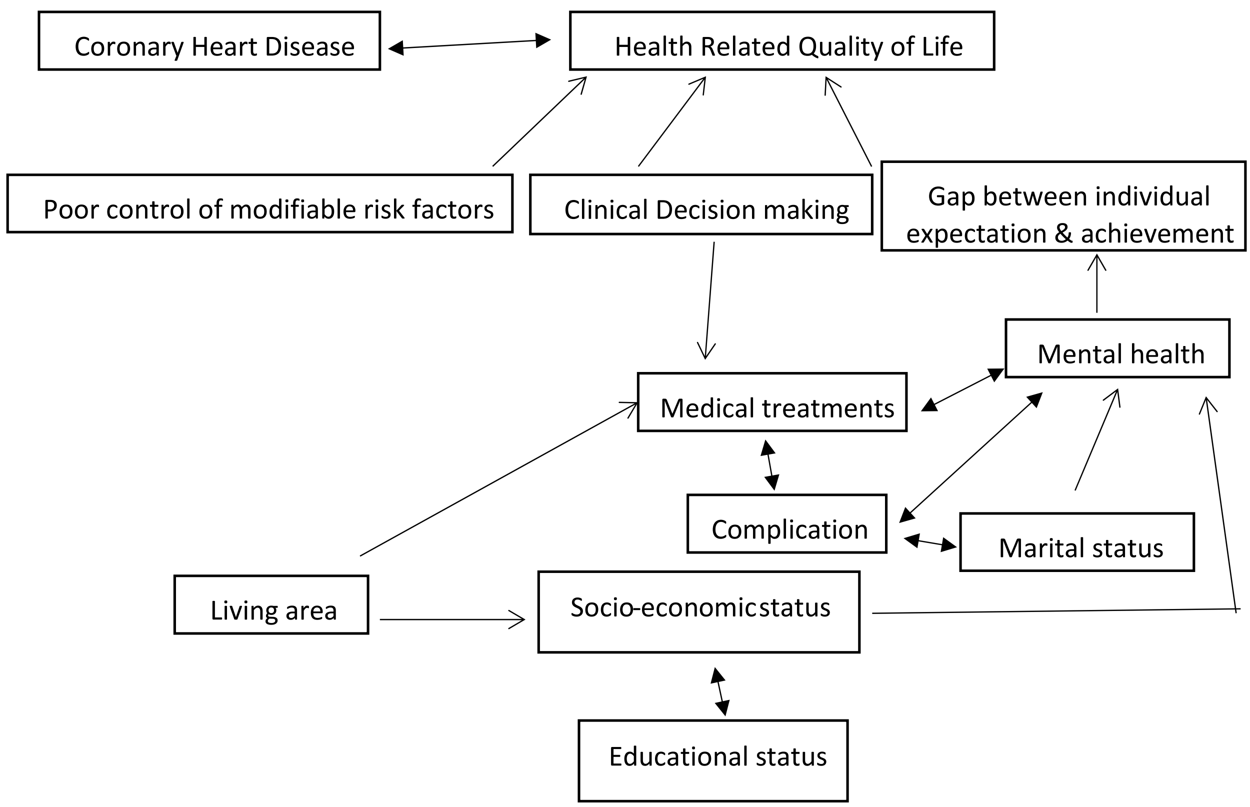


Figure 2: Factors for HRQL

One of the studies describes the significance of cancer disease clinical data and the corresponding impact on the clinical practice with regards to QOL [11]. Data from clinical practices are important resources for these types of studies since they can be interpreted and transformed in to clinical knowledge thus it is useful in enhancing medical-decision making, such as deciding how the patient is treated and so on. They confirms that those decisions are directly associated with QOL (HRQOL).

This study describes the HRQOL as the gap between the individual's expectation and the real achievement or, the aspects affected by the health care. The suggestions contain instructions such as informing the patient about the treatment possible long term and short term complications. There is variety of tools to measure HRQOL. A combination of physical, social and emotional functioning measurements, symptoms measurements and side effects for both disease and treatments, is used in tools.

Finding causes and relate them with the data mining techniques is one of the complicated task and it has done by many researches. One of the journal describes the importance of addressing the matter of CHD and have identified the most of the deaths occur in low and middle income countries [4].

A study [12] interprets the term of QOL as a physical, emotional and social well-being of individuals and they suggests that the symptoms of angina and heart failure, limited exercise capacity of the symptoms, the physical weakness caused and emotional stress associated with the chronic stress as the factors of QOL with related to the CHD hence HRQOL. Furthermore they highlight the need of data collection from the patients live with the disease because the sickness disturbs all the aspects of the patient's life.

2.3 Data Mining Techniques

Several data mining techniques have been used by previous researches in different domains to get positive outcome. A discussion on usage and selection of various data mining techniques as follows.

One of the studies [4] discusses about these types of techniques by reviewing each of them and shows the performance and accuracy between them. Basically they confirm the Decision Tree and Support Vector Machine (SVM) techniques are more accurate than other methods. Since heart diseases' critical situation is the accuracy of the technique use for the prediction becomes a significant factor.

Only few of the algorithms have been taken under considerations when carrying out the study hence analysis of more algorithms and techniques might affect the ultimate outcome of the above study.

The initial step of the process is data collection where the primary source is the patient who suffer from the disease [11]. Methods such as Interviews, well organized surveys can be used for this purpose but the questionnaire should be built up within the boundaries of the domain based on the target sample population. Global/ generic questionnaires or domain specific questionnaires should be used based on the requirement. The difference is the global questionnaires are more towards collecting data across different population and regions/

physical locations while the domain specific questionnaires more towards collecting information of the domain of the study.

A research conducted by Maharaja Surajmal institute of technology, New Delhi, India has come up with early heart disease prediction system using data mining classification modeling techniques [13] namely, Decision Trees, naïve Bayes and Neural network, weighted association Apriori algorithm and MAFIA algorithm in heart disease prediction. The main source of data collection was medical profiles such as age, gender, blood pressure and blood sugar with a total of 909 records from Cleveland heart disease data base. They say it can predict the likelihood of patients getting heart diseases using the above mentioned medical profiles. Different algorithms together with fourteen identified attributes has been used for the prediction.

Department of Computer Science & Engineering Integral University, Lucknow, India has come up with a research study [14] which is focused on early heart disease detection based on major risk factors using hybrid technique for an accurate prediction. The hybrid system implemented uses the global optimization advantage of genetic algorithm for initialization of neural network weights. They have categorized risk factors into two parts namely, major risk factors (diabetes, hypertension, high blood cholesterol) and lifestyle risk factors (eating habits, physical inactivity, smoking, alcohol intake, obesity). Data from 50 people were surveyed which were collected by American Heart Association. The system indicates whether the patient has the risk of heart disease or not.

Department of Industrial Engineering and Management Yuan Ze University with Ming Chi University of technology, Taiwan [15] has published a research paper on using data mining techniques for multi-disease prediction modeling of hypertension and hyperlipidemia by common risk factors. This paper proposes a two-phase analysis procedure to simultaneously predict hypertension and hyperlipidemia. First they have selected the common risk factors for both the diseases using voting principle. They have used multivariate adaptive regression splines (MARS) method to construct a multiple predictive model with 2048 data sets. The proposed multi disease prediction method has a classification accuracy rate of 93.07%.

A survey [16] on heart disease prediction system using data mining techniques by school of computer science and engineering, Bharathidasan University, Tiruchirappalli, India provides a survey of current techniques of knowledge discovery in data bases. The classifiers like Naïve Bayes, Neural Network, SVM and Weighted Associative Classifier (WAC) algorithms have used to predict the presence of heart disease more accurately. They have identified as the WAC classifier provides the accurate result compared to other techniques.

Almost all the existing research studies were conducted for the early prediction of heart disease whereas in this research study the focus is towards predicting the quality of life of a patient following a heart disease, which will affect the individual mentally, physically as well as socially thus identified risk factors and correlations will be used for the study.

Chapter 3

METHODOLOGY

In the literature review several factors have been identified which were known to affect for the Quality of Life of an individual. Namely,

- Marital Status
- Educational Status
- Socioeconomic Status
- Living area
- Smoking, Alcohol and Drugs
- Medical Treatments
- Coronary Heart Disease (Heart Attacks)

These factors were defined by the researchers in the global context where as in Sri Lanka several adjustments have to be made accordingly when considering above factors.

$\mathbf{H_0}$**:** The patients with Coronary Heart Disease has a lower quality of life

$\mathbf{H_1}$**:** The patients with Coronary Heart Disease has a higher quality of life

The above hypothesis is proposed due to number of reasons and those factors or the reasons are tested throughout this chapter. The following are the reasons to build up a positive hypothesis.

- Controlled diet
- Ended bad habits
- Monthly clinics and follow-ups
- More care from others
- Controlled daily works

3.1 Knowledge Discovery Process

Most of the existing researchers follow the Knowledge Discovery Process or Data mining process to find the Knowledge out of the collected huge amount of data.

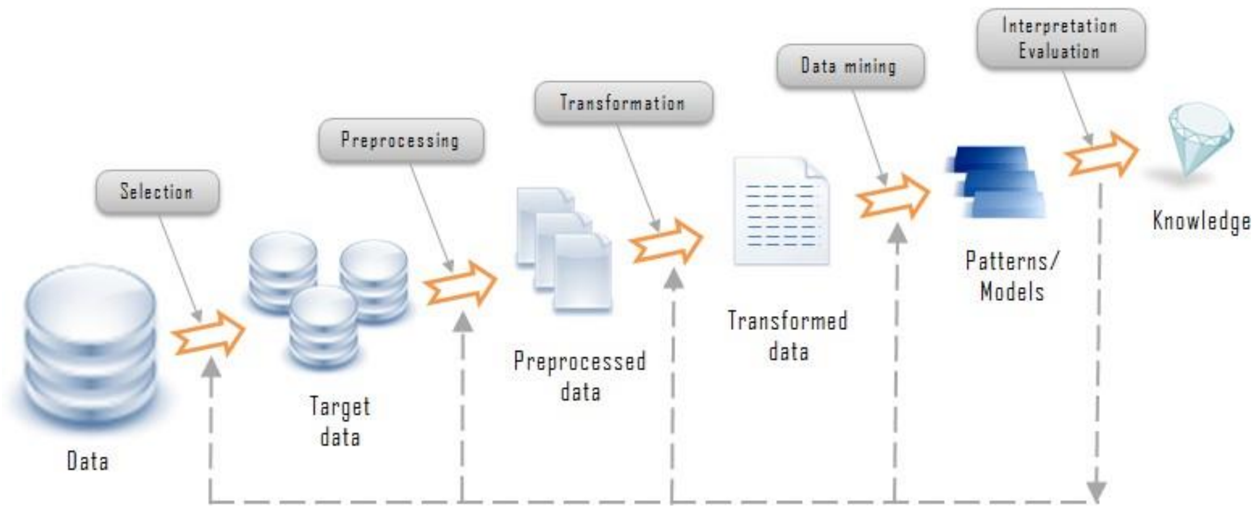


Figure 3: Knowledge Discovery Process

The Figure:3 illustrates the process for knowledge discovery includes data gathering, filtering (extracting) the required data, processing them to a target collecting type, identifying the relationships and patterns and finally interpreting them to build the knowledge [4]. A verity of algorithms and techniques have been developed and used in this process such as Neural Networks , Nearest Neighbor method ,Clustering, Regression, Decision Trees, Artificial Intelligence, Classification, Association Rules and Genetic Algorithm. Finding and applying appropriate techniques for the data collection is a real challenge since the behavior of each technique or the algorithm is different from each other thus an analysis is a need.

3.2 Data Collection

The target data set is collected through a series of face to face interviews with the doctors as well as patients who will currently follow up monthly in the hospital clinic or by using a well-defined.

Questionnaire for both the parties. Furthermore the clinical records of patients who passed away following admission to the hospital planned to be collected from the hospital. The arrangements have been made to collect data with the permission of the Director and the consultant cardiologist, district general hospital, Gampaha.

Thereby the collected data will be categorized based on the modifiable risk factors which were identified as major triggering factors for coronary heart disease with age and gender of the each individual. Almost all the data containing major risk factors as well as the lifestyle risk factors are to be collected through clinical records and patients feedbacks.

Related data and information to be collected from hospital staff, Coronary Care Unit (CCU), Cardiology clinic, Clinical records inventory at District General hospital, Gampaha. All the sample data and information collected will be assumed true and correct.

The goal is to determine and predict quality of life of patients who had coronary heart disease in related to modifiable risk factors, age category and gender.

Expect to collect minimum of 700 data sets from the District General hospital, Gampaha. Data collection process and evaluation will be categorized into four parts based on the time frame.

Historical data will be collected and analyzed to build up the base for the system. The verification will be based on Doctors feed backs, patients information and clinical records.

3.2.1 Questionnaire

Development of the questionnaire is a critical task since all the identified risk factors should be covered and the size of the document should be small as much as possible and easy to answers.

The following main points were considered when constructing the questionnaire.

- General Information
  - Age
  - Gender
  - Marital status
  - Religion

- Socio- economic status
  - Education Level
  - Occupation
  - Income & expense status

- Family

- Expectation and Achievements

- Complications

- Diet
  - Salt usage
  - Coconut usage
  - Fruit & Vegetable usage

- Exercises

- Smoking
- Alcohol Usage
- Drugs Usage
- History of the disease

The document was translated in to Sinhala language since the target sample common language is the same.

3.3 Target Data

The proposed study focuses on surveying patients with CHD where most of them are from different education levels, complications and different environments hence the development of the Survey was a critical task since the target population is from different environments.

The survey was developed by using questions representing categorical variables. A Categorical Variable is a variable type defined in analytic context which will consist a set of different categories whereas the Categorical data are the counts for each of those categories. Based on the ordering of the categories the Categorical variable can be separated in to two types of measurements scales namely Ordinal or Nominal. Data such as education level falls in to the Ordinal type since the education levels are under natural order and if the data do not inherit any ordering they comes under nominal measurement scale.

Ease of collation and categorization were the main reasons to select Categorical Variables and Data for the study because the target data was opinions of the patients. Human nature is finding an easy method to perform a task rather than hardly doing it. The questionnaire was constructed using Categorical Variables to make it easy to answer just by ticking one of the categories under each questions. Only less number of answers were asked to be completed by writing and they were also short answers. In the other hand if the audience was asked to fill out the survey without providing no options or variable the analytical stage will be problematic since the answers will be so diverse and they could not be converted in to statistics.

A set of related risk factors were identified during the background study which can be consider as the backbone of the survey to collect data under the subject. The risk factors were grouped in to two different sections such as risk factors of CHD and risk factors of Health related Quality of Life.

Under the above concerns the survey was built with 62 questions including details such as personal, family, day to day living standards and the illness of the participants. Due to time limitation and to prevent human bias since the participants are not fit enough, most of the questions were created considering the categorical variables.

Verification of diagnosis and management were collected from the Doctors. Around 150 data was collected per week from cardiology clinic District General Hospital, Gampaha. Ultimately around 500 data were collected. Each data will be verified by feedback from the Doctors prior to creating the dataset out of it.

3.4 Data Preprocessing and Transformation

Starting with Data Preprocessing, the data collected using surveys were stored in to digital format using SPSS tool. Even though the survey consist of 62 questions, the real target data set was formed as 85 attributes. In the selection process several new attributes were formed based on existing attributes to represent a better outcome from the data. The Body Mass Index (BMI) of the patient was calculated and categorized in to three categories accordingly based on the collected Weight and Height measurements and a variable named Bad Habits was created by referring to the facts such as smoking, usage of alcohol and drugs.

A simple table was included in the survey in order to collect data of past medical illnesses such as Diabetes, Hypertension, Hyperlipidemia, Stroke and CHD. Associated columns representing each illnesses namely duration of the illness, type of medications, compliance of treatment, family history were formed as separate categorical attributes. The collected feedback of the patients from doctors were used to construct a categorical variable named Diagnosis, by summarizing all the variances of CHD types and patients with diseases other than CHD.

Sometimes the categories exist on the survey needed to be revised by adding more or removing categories in order to cover up all the possible scenarios of each answers.

When entering the survey result to the digital format, set of missing values were identified and they were denoted by -1 label, as they cannot handled by replacing suitable values over them. Even though some filters were tried on the missing values, but the generated outcome was less accurate to replace the missing values by.

Occasionally conversion of collected numerical data in to categorical scales and vice versa happened. Finally all the variables other than Height and Weight of the patients were created as categorical variables. The tool used for this process provides functionalities to use different types of filters for attributes conversion.

The causes such as controlled Diet, ended Drug, Alcohol and Smoking habits, participation on regular monthly clinics, improved care from others for the patients and reduced daily work were considered in the hypothesis development and acceptance of the hypothesis or rejection will be done based on the analysis done in the whole process.

After creating the data sheets the next phase was to find out the relationships in between different variables of the data set.

Attribute selection feature was really helpful in this step which could be used to find out the relationships among variables. Set of attribute evaluators together with set of search methods were available to use with the mode types full training set or 10 fold cross validation method. The attribute selection with 10 fold cross validation was used in the study because the relationships from one variable to all other variables were offered with percentages, where the significance of the relationships could be measured and selected variables with most significant relationships.

The factors and their proved relationships found throughout the literature review phase were considered in the context of quality of life of the patients when extracting relationships within variables. Basically Marital status, Education status, Social and economic status, Living area, Smoking and alcohol, Medical treatments and Heart attacks are the identified factors related to Quality of Life. Variable sets under each of these factors were applied on the algorithms to identify the types of relationships between them. The same relationship was tested under different algorithms to come up with the most accurate relationships with considering the significance of each outputs.

The set of identified patterns/ models are listed together and finally the knowledge is to be extracted out of them. The relationships/ patterns identified by the analysis were applied on the found factors of Quality of Life thus these patterns will be verified to identify their impact on Quality of Life in positive or negative manner.

3.5 Data Mining Techniques

The Knowledge discovery process includes Preprocessing, Transformation, Data mining and Evaluation techniques, which have been identified in the background study. These steps has to be followed in order to reach the knowledge from the raw data set collected through the survey.

The data mining tasks involve two different forms namely Descriptive data mining and Predictive data mining. Descriptive data mining provides information to understand what is happening inside the data without using predetermined data whereas the Predictive data mining allows the user to submit records with unknown field values, and this mechanism will guess the unknown values based on previous patterns discovered from the data set. The Clustering model and Association rule mining comes under Descriptive data mining and the Classification model and Prediction model comes under Predictive data mining.

The measures of confidence, certainty or the trustworthiness will be focused and decision making will happen based on these values of any algorithm output used.

Any data set consist of Discrete or Nominal data values are defined as Classification problems on the other hand Continuous or Numeric data values falls in to Regression problem category. Since the data set used for this study mainly focuses for Nominal data values the research problem falls on to the Classification problem. Hence corresponding algorithms will be applied on the data set accordingly.

The previous studies have used different tools and algorithms for the process and out of those different data mining algorithms, mainly three techniques can ben identify as most accurate and responsive algorithms. Those are namely Association Rules, Decision Tree Algorithms and Naïve Bayes.

When searching for relationships among variables Association Rule learning is a powerful method. It is a rule based machine learning method for identifying strong rules/ frequencies using some measure of interestingness. The tool proposed to use for the analysis has the capability for applying this machine learning mechanism on the dataset. This mechanism is a Dependency model which helps to un-patch hidden patterns in the data. The variable identification process takes place based on the appearance frequency of the variables. This methodology comes under exploratory analysis and they might be used to help predict behavior of the data. Even though this process would create a set of rules, the important rules should be filtered out by referring to the highest percentage of confidence.

Even though there are set of association rule learning algorithms available in the tool basically Apriori algorithms was used for the patterns extracting and this algorithm cannot be used with quantitative variables. Apriori is one of the earliest and most fundamental algorithm for generating association rules. This pioneered the use of support for pruning the item set and controlling the exponential growth of candidate items set. The minimum support value can be set and only the frequency items set which are greater than or equal to the minimum support will be considered in the process.

All the interesting rules found were noted with significance and sent through the Classification analysis phase in order prove the patterns with different Classification algorithms. Classification Analysis retrieves important information about the data using classes concept. This mechanism is somewhat similar to Clustering mechanism which uses Segment data records. Based on the manner of the classification technique different algorithms can be used. Mainly One Rule algorithm, Decision tree algorithm and the Naïve Bayes algorithm were used for this study.

The simple and an accurate classification algorithm named One Rule or OneR is used in the study to generate simple relationships of the data set. This generates one rule for each predictor in the data set and then selects the rule with the smallest total error. The rules produced are simple to interpret.

Decision Tree algorithm is useful for problems in which the goal is to make broad categorical classification or prediction and even though this algorithm can be applied on quantitative variables it is not useful, but it will fit when the questions identified a priori.

The other main classification technique is Naïve Bayes algorithm which is a simple probabilistic classifier by applying Bayes' theorem with the (naïve) independence assumptions between the features/ predictors. This classifier is simple, straightforward, powerful, easy to build and useful for very large datasets. Other than these characteristics this model performs very well and this is widely used because it often outperforms more sophisticated classification method.

Hence above algorithms will be basically used for knowledge discovery process in this research project.

3.6 Tools

The Data analysis is a major part of the research when it comes to Data Mining. Data set creation, Statistical analysis and analysis using data mining algorithms is to be done under this section. The following tools are used in this research project.

3.6.1. Weka

This software application is a collection of tools developed including set of machine learning algorithms for the application of Data Mining. It contains the visualization tools and algorithms for data preparation, classification, regression, clustering, association rules mining, and visualization [17].

This is an open source software with functionality to apply algorithms directly to data set or to call from our own java code. This application is widely used by modern researchers as it can process big data and deep learning. Due to GUI of the application the software is more user friendly [18].

3.6.2. IBM SPSS Software

This is a software application containing tools such as statistical analysis, machine learning algorithms, bid data processing and text analysis [19].

Chapter 4

EVALUATION AND RESULTS

The evaluation is done based on the hypothesis designed and the identified factors are mainly addressed, tested and verified during the study and finally all the results are demonstrated.

Basically the classification mechanisms such as Decision Tree technique, One Rule technique and Naïve Bayes theorem as well as Association Rule mining are used for the analysis. Each of these results are compared with each other in order to increase the accuracy and the efficiency of the output.

This data set consists of 482 instances and 84 attributes. Two attributes namely height and weight considered as numerical attributes and rest of other attributes are under nominal category. In the Classification phase the data set is divided in to two groups as Training data set and the Test data set automatically by the tool Weka. The division is done based on a percentage of 66 for Training data and remaining set for Test data set.

4.1 Analysis

The section Analysis will demonstrate all the important relationships and statistics obtained during the data analysis. To represent the missing values of the data set the global constant -1 is used in the data set. Even though it could be replaced by a suitable value suggested by a filter since the integrity was considered the missing value did not replaced.

Each of these results will be explained and discussed in the section 4.2.

Marital Status

One of the main factors identified which affects the QOL directly is marital status. This happens due to number of reasons such as the partner is there for help when needed, mental health, children. Following are the variables considered under marital status factor.

Partner	– the partner of the patient is currently lives with patient or not,
Love_Affection	– the patient obtain love and affection from someone or not,
Fortunate	– the patient thinks that he/she fortunate or not
Member_count	– total member count of the family
Dependent	– patients' dependency status (independent, depend on children, depend on parents, children depend on)
Help_fromOthers	– the patient has help from someone or not

Association Rules

```
Instances:    482
Attributes:   3
              Partner
              Love_Affection
              Fortunate
=== Associator model (full training set) ===

Apriori
=======
Best rules found:
 1. Partner=Lives with Fortunate=Yes 318 ==> Love_Affection=Yes 316    <conf:(0.99)> lift:(1.05)
    lev:(0.03) [15] conv:(5.94)
 2. Fortunate=Yes 419 ==> Love_Affection=Yes 412    <conf:(0.98)> lift:(1.04) lev:(0.03) [16] conv:(2.93)
 3. Partner=Lives with 362 ==> Love_Affection=Yes 347    <conf:(0.96)> lift:(1.02) lev:(0.01) [5]
    conv:(1.27)
 4. Partner=Lives alone Fortunate=Yes 90 ==> Love_Affection=Yes 86    <conf:(0.96)> lift:(1.01) lev:(0)
    [1] conv:(1.01)
 5. Partner=Lives alone Love_Affection=Yes 94 ==> Fortunate=Yes 86    <conf:(0.91)> lift:(1.05)
    lev:(0.01) [4] conv:(1.37)
 6. Partner=Lives with Love_Affection=Yes 347 ==> Fortunate=Yes 316    <conf:(0.91)> lift:(1.05)
    lev:(0.03) [14] conv:(1.42)
 7. Love_Affection=Yes 455 ==> Fortunate=Yes 412    <conf:(0.91)> lift:(1.04) lev:(0.03) [16] conv:(1.35)
 8. Partner=Lives alone 104 ==> Love_Affection=Yes 94    <conf:(0.9)> lift:(0.96) lev:(-0.01) [-4]
    conv:(0.53)
```

Figure 4: Model by Association Rules with Partner, Love & Affection and Fortunate

One Rule

Instances: 482
Attributes: 3
Partner
Love_Affection
Fortunate
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Love_Affection:
Yes -> Yes
-1 -> -1
No -> No
(432/482 instances correct)

=== Summary ===

Correctly Classified Instances 143 87.1951 %
Incorrectly Classified Instances 21 12.8049 %

=== Confusion Matrix ===

```
  a   b   c   <-- classified as
 133  0   1 |  a = Yes
  0   3   0 |  b = -1
 20   0   7 |  c = No
```

Decision Tree

Attributes: 3
Partner
Love_Affection
Fortunate
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 unpruned tree

Love_Affection = Yes: Yes (455.0/43.0)
Love_Affection = -1: -1 (6.0)
Love_Affection = No: No (21.0/7.0)

=== Summary ===

Correctly Classified Instances 139 84.7561 %
Incorrectly Classified Instances 25 15.2439 %

=== Confusion Matrix ===

```
  a   b   c   <-- classified as
 133  0   1 |  a = Yes
  0   3   0 |  b = -1
 24   0   3 |  c = No
```

Figure 5: Models by One Rule & Decision Tree with Partner, Love & Affection and Fortunate

Naïve Bayes

```
Attributes:   3
          Partner
          Love_Affection
          Fortunate
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===
Naive Bayes Classifier

             Class
Attribute         Yes     -1     No
                 (0.87) (0.01) (0.12)
=====================================
Partner
  Lives with     319.0    6.0   40.0
  Lives alone     91.0    2.0   14.0
  -1              11.0    1.0    6.0
  [total]        421.0    9.0   60.0

Love_Affection
  Yes            413.0    1.0   44.0
  -1               1.0    7.0    1.0
  No               8.0    1.0   15.0
  [total]        422.0    9.0   60.0

=== Summary ===

Correctly Classified Instances        143           87.1951 %
Incorrectly Classified Instances       21           12.8049 %

=== Confusion Matrix ===

   a    b   c   <-- classified as
 133   0   1 |   a = Yes
   0   3   0 |   b = -1
  20   0   7 |   c = No
```

Figure 6: Model by Naive Bayes with Partner, Love & Affection and Fortunate

One Rule

Attributes: 2
Partner
Love_Affection
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Partner:
Lives with -> Yes
Lives alone -> Yes
-1 -> Yes
(455/482 instances correct)

=== Summary ===

Correctly Classified Instances 153 93.2927 %
Incorrectly Classified Instances 11 6.7073 %

=== Confusion Matrix ===

```
  a   b  c   <-- classified as
153   0  0 |  a = Yes
  3   0  0 |  b = -1
  8   0  0 |  c = No
```

Decision Tree

Attributes: 2
Partner
Love_Affection
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 unpruned tree

: Yes (482.0/27.0)

Number of Leaves : 1
Size of the tree : 1

=== Summary ===

Correctly Classified Instances 153 93.2927 %
Incorrectly Classified Instances 11 6.7073 %

=== Confusion Matrix ===

```
  a   b  c   <-- classified as
153   0  0 |  a = Yes
  3   0  0 |  b = -1
  8   0  0 |  c = No
```

Figure 7: Models by One Rule & Decision Tree with Partner, Love & Affection

Naïve Bayes

```
Attributes:   2
          Partner
          Love_Affection
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

             Class
Attribute         Yes     -1     No
                (0.94)  (0.01) (0.05)
===================================
Partner
  Lives with      348.0    6.0   11.0
  Lives alone      95.0    2.0   10.0
  -1               14.0    1.0    3.0
  [total]         457.0    9.0   24.0

=== Summary ===

Correctly Classified Instances        153             93.2927 %
Incorrectly Classified Instances       11              6.7073 %

=== Confusion Matrix ===

   a   b   c   <-- classified as
 153   0   0 |   a = Yes
   3   0   0 |   b = -1
   8   0   0 |   c = No
```

Figure 8: Model by Naive Bayes with Partner, Love & Affection

Association Rules

Attributes: 3
Member_count
Dependent
Help_fromOthers
=== Associator model (full training set) ===

Apriori
=======

Best rules found:

1. *Member_count=more than 5 Dependent=Depend on children 82 ==> Help_fromOthers=Help available 79 <conf:(0.96)> lift:(1.04) lev:(0.01) [2] conv:(1.45)*
2. *Member_count=3 to 4 Dependent=Children depend on 78 ==> Help_fromOthers=Help available 74 <conf:(0.95)> lift:(1.02) lev:(0) [1] conv:(1.1)*
3. *Dependent=Depend on children 212 ==> Help_fromOthers=Help available 200 <conf:(0.94)> lift:(1.01) lev:(0.01) [2] conv:(1.15)*
4. *Member_count=3 to 4 226 ==> Help_fromOthers=Help available 213 <conf:(0.94)> lift:(1.01) lev:(0.01) [2] conv:(1.14)*
5. *Dependent=Children depend on 115 ==> Help_fromOthers=Help available 108 <conf:(0.94)> lift:(1.01) lev:(0) [1] conv:(1.01)*
6. *Member_count=more than 5 149 ==> Help_fromOthers=Help available 138 <conf:(0.93)> lift:(1) lev:(-0) [0] conv:(0.88)*
7. *Member_count=3 to 4 Dependent=Depend on children 94 ==> Help_fromOthers=Help available 87 <conf:(0.93)> lift:(1) lev:(-0) [0] conv:(0.83)*
8. *Dependent=Independent 143 ==> Help_fromOthers=Help available 130 <conf:(0.91)> lift:(0.98) lev:(-0.01) [-2] conv:(0.72)*
9. *Member_count=1 to 2 102 ==> Help_fromOthers=Help available 92 <conf:(0.9)> lift:(0.97) lev:(-0.01) [-2] conv:(0.65)*

Figure 9: Model by Association Rules with Member Count, Dependent & Help from others

One Rule

Attributes: 3
Member_count
Dependent
Help_fromOthers
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Dependent:
Independent -> Help available
Depend on children -> Help available
-1 -> Help available
Children depend on -> Help available
(449/482 instances correct)

=== Summary ===

Correctly Classified Instances 147 89.6341 %
Incorrectly Classified Instances 17 10.3659 %

=== Confusion Matrix ===

```
  a   b   c   <-- classified as
 147   0   0 |   a = Help available
   3   0   0 |   b = -1
  14   0   0 |   c = Help not available
```

Decision Tree

Attributes: 3
Member_count
Dependent
Help_fromOthers
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 unpruned tree

```
------------------
Member_count = 3 to 4: Help available
                                        (226.0/13.0)
Member_count = 1 to 2: Help available
                                        (102.0/10.0)
Member_count = -1: Help available
                                        (5.0)
Member_count = more than 5
| Dependent = Independent: Help available
                                        (34.0/4.0)
| Dependent = Depend on children: Help available
                                        (82.0/3.0)
|   Dependent = -1: Help available (2.0/1.0)
|   Dependent = Children depend on: Help available
                                (30.0/2.0)
Number of Leaves  :        1
Size of the tree :   1
```

=== Summary ===

Correctly Classified Instances 147 89.6341 %
Incorrectly Classified Instances 17 10.3659 %

=== Confusion Matrix ===

```
  a   b   c   <-- classified as
 147   0   0 |   a = Help available
   3   0   0 |   b = -1
  14   0   0 |   c = Help not available
```

Figure 10: Models by One Rule, Decision Tree with Member Count, Dependent, Help from others

Naïve Bayes

```
Attributes:   3
              Member_count
              Dependent
              Help_fromOthers
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

                          Class
Attribute                 Help available      -1       Help not available
                          (0.93)             (0.01)       (0.06)
=================================================================================
Member_count
  3 to 4                     14.0            5.0          10.0
  1 to 2                     93.0            3.0           9.0
  -1                          6.0            1.0           1.0
  more than 5               139.0            1.0          12.0
  [total]                   452.0           10.0          32.0

Dependent
  Independent               131.0            4.0          11.0
  Depend on children        201.0            3.0          11.0
  -1                         11.0            1.0           2.0
  Children depend on        109.0            2.0           7.0
  Depent on parents           1.0            1.0           2.0
  [total]                   453.0           11.0          33.0

=== Summary ===

Correctly Classified Instances        147           89.6341 %
Incorrectly Classified Instances       17           10.3659 %

=== Confusion Matrix ===

   a   b   c   <-- classified as
 147   0   0 |   a = Help available
   3   0   0 |   b = -1
  14   0   0 |   c = Help not available
```

Figure 11: Model by Naïve Bayes with Member Count, Dependent & Help from others

Education and Socioeconomic status

The factor Education did not contribute any significant value to the data set and could not find significant relationships. Hence the following variables considered.

Religion – the religion of the patient
ReligionBelivesVsLife – patient believe on the religion or not
Living_Status_Sufficiency – patient lives happily, sadly or normally
Fortunate – patients is fortunate or not
Monthly_Income – the rough monthly income of the patient
Currently_Working – patients is currently working or not
Dependent – dependency status (independent, depend on children, children depend on, depend on parents)
IncomeVsExpenditure_Sufficiency – patient has a sufficient income to bear living expenses

One Rule

```
Instances:    482
Attributes:   2
          Religion
          ReligionBelievesVsLife
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

Religion:
          Buddhist-> Needed
          -1        -> Needed
          Catholic -> Needed
          Islam     -> Needed
          Hindu     -> Needed
          Other     -> Needed
(444/482 instances correct)

=== Summary ===

Correctly Classified Instances   148     90.2439 %
Incorrectly Classified Instances   16      9.7561 %

=== Confusion Matrix ===

  a   b   <-- classified as
 148   0 |   a = Needed
  16   0 |   b = Not Needed
```

Decision Tree

```
Attributes:   2
          Religion
          ReligionBelievesVsLife
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 unpruned tree
------------------
: Needed (482.0/38.0)

Number of Leaves  :         1

Size of the tree :   1


=== Summary ===

Correctly Classified Instances   148   90.2439 %
Incorrectly Classified Instances   16     9.7561 %

=== Confusion Matrix ===

  a   b   <-- classified as
 148   0 |   a = Needed
  16   0 |   b = Not Needed
```

Figure 12: Models by One Rule and Decision Tree with Religion and Religion Believes

Naïve Bayes

```
Attributes:   2
          Religion
          ReligionBelievesVsLife
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

              Class
Attribute        Needed   Not Needed
                 (0.92)     (0.08)
====================================
Religion
  Buddhist       397.0      37.0
  -1              12.0       3.0
  Catholic        28.0       1.0
  Islam            8.0       1.0
  Hindu            3.0       1.0
  Other            2.0       1.0
  [total]        450.0      44.0

=== Summary ===

Correctly Classified Instances       148         90.2439 %
Incorrectly Classified Instances      16          9.7561 %

=== Confusion Matrix ===

   a   b   <-- classified as
 148   0 |   a = Needed
  16   0 |   b = Not Needed
```

Figure 13: Model by Naïve Byes with Religion and Religion Believes

Association Rules

=== Associator model (full training set) ===
Apriori
=======

Best rules found:

1. *Living_Status_Sufficiency=Happy ReligionBelievesVsLife=Needed 150 ==> Fortunate=Yes 145 <conf:(0.97)> lift:(1.11) lev:(0.03) [14] conv:(3.27)*
2. *Living_Status_Sufficiency=Normal Fortunate=Yes 240 ==> ReligionBelievesVsLife=Needed 230 <conf:(0.96)> lift:(1.04) lev:(0.02) [8] conv:(1.72)*
3. *Living_Status_Sufficiency=Happy 162 ==> Fortunate=Yes 155 <conf:(0.96)> lift:(1.1) lev:(0.03) [14] conv:(2.65)*
4. *Fortunate=Yes 419 ==> ReligionBelievesVsLife=Needed 394 <conf:(0.94)> lift:(1.02) lev:(0.02) [8] conv:(1.27)*
5. *Living_Status_Sufficiency=Normal 278 ==> ReligionBelievesVsLife=Needed 261 <conf:(0.94)> lift:(1.02) lev:(0.01) [4] conv:(1.22)*
6. *Living_Status_Sufficiency=Happy Fortunate=Yes 155 ==> ReligionBelievesVsLife=Needed 145 <conf:(0.94)> lift:(1.02) lev:(0) [2] conv:(1.11)*
7. *Living_Status_Sufficiency=Happy 162 ==> ReligionBelievesVsLife=Needed 150 <conf:(0.93)> lift:(1.01) lev:(0) [0] conv:(0.98)*

Figure 14: Model by Association Rules with Living Status Sufficiency, Fortune, Religion Believes

Selected attribute

Name: Monthly_Income
Missing: 0 (0%) Distinct: 5
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Rs. 15,000-30,000	72	72.0
2	Rs.0-5000	295	295.0
3	Rs. 5000- 15,000	82	82.0
4	Rs 30,000- 50,000	31	31.0
5	> Rs. 50,000	2	2.0

Class: Monthly_Income (Nom)
Visualize All

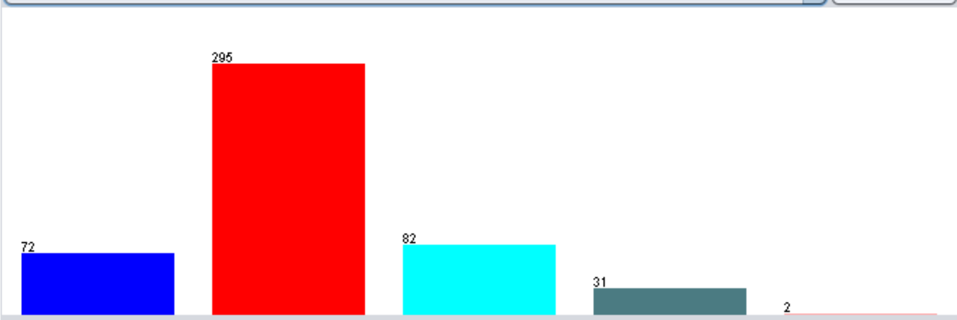
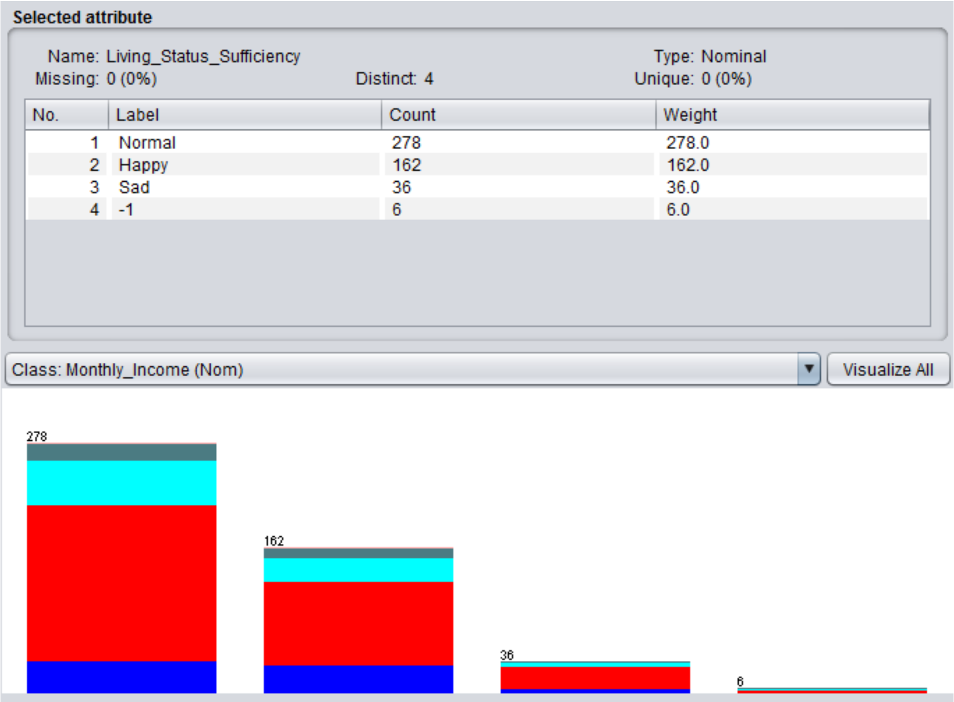


Figure 15: Monthly Income

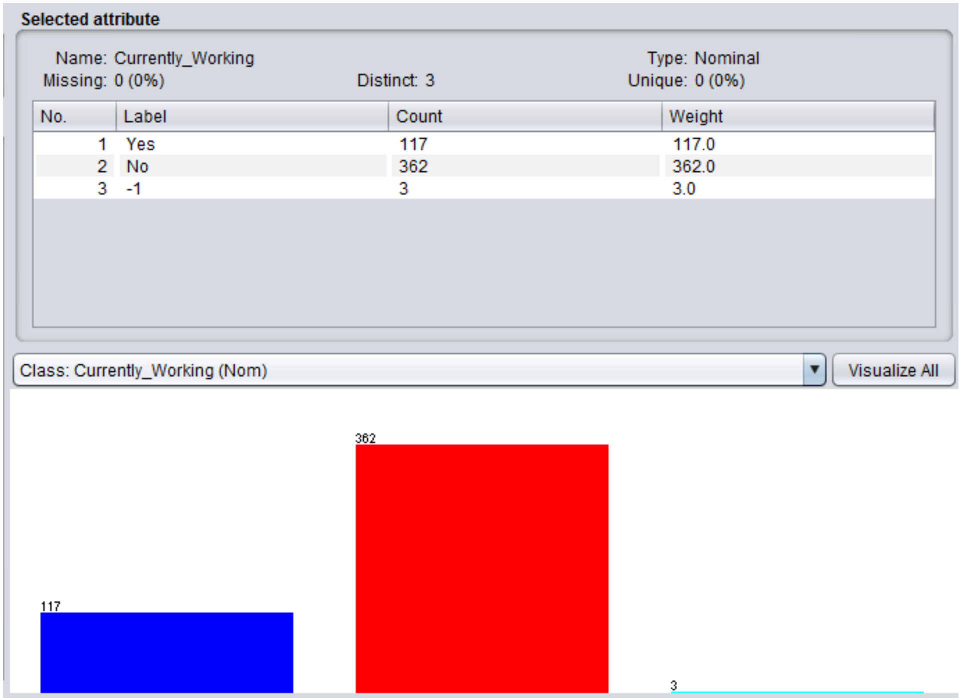


Selected attribute

Name: Living_Status_Sufficiency
Type: Nominal
Missing: 0 (0%)
Distinct: 4
Unique: 0 (0%)

No.	Label	Count	Weight
1	Normal	278	278.0
2	Happy	162	162.0
3	Sad	36	36.0
4	-1	6	6.0

Figure 16: Monthly Income vs Living Status Sufficiency



Selected attribute

Name: Currently_Working
Type: Nominal
Missing: 0 (0%)
Distinct: 3
Unique: 0 (0%)

No.	Label	Count	Weight
1	Yes	117	117.0
2	No	362	362.0
3	-1	3	3.0

Figure 17: Current Work Status

One Rule

```
Attributes:   2
          Dependent
          IncomeVsExpenditure_Sufficiency
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

Dependent:
           Independent        -> Sufficient
           Depend on children -> Not Sufficient
           -1           -> Sufficient
           Children depend on -> Sufficient
           Depend on parents -> Not Sufficient
(285/482 instances correct)

=== Summary ===

Correctly Classified Instances     94    57.3171 %
Incorrectly Classified Instances  70    42.6829 %

=== Confusion Matrix ===

  a    b    c  d   <-- classified as
 58  27  0   0 |  a = Sufficient
 30  36  0   0 |  b = Not Sufficient
  4   3    0   0 |  c = Very Sufficient
  5   1    0   0 |  d = -1
```

Decision Tree

```
Attributes:   2
          Dependent
          IncomeVsExpenditure_Sufficiency
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree
------------------
Dependent = Independent: Sufficient
                                             (143.0/56.0)
Dependent = Depend on children: Not Sufficient
                                             (212.0/96.0)
Dependent = -1: Sufficient  (11.0/3.0)
Dependent = Children depend on: Sufficient
                                             (115.0/42.0)
Dependent = Depend on parents: Not Sufficient
                                             (1.0)

Number of Leaves  :          5
Size of the tree :   6

=== Summary ===

Correctly Classified Instances        94      57.3171 %
Incorrectly Classified Instances       70      42.6829 %
=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 58 27  0  0 |  a = Sufficient
 30 36  0  0 |  b = Not Sufficient
  4  3  0  0 |  c = Very Sufficient
  5  1  0  0 |  d = -1
```

Figure 18: Models by One Rule, Decision Tree with Dependent, IncomeVSExpenditure Sufficiency

Naïve Bayes

```
Attributes:   2
              Dependent
              IncomeVsExpenditure_Sufficiency
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

                         Class
Attribute                Sufficient  Not Sufficient Very Sufficient       -1
                           (0.52)       (0.42)         (0.04)           (0.02)
==================================================================================
Dependent
  Independent               88.0         46.0           8.0              5.0
  Depend on children        86.0        117.0           8.0              5.0
  -1                         9.0          4.0           1.0              1.0
  Children depend on        74.0         37.0           6.0              2.0
  Depent on parents          1.0          2.0           1.0              1.0
  [total]                  258.0        206.0          24.0             14.0


=== Summary ===

Correctly Classified Instances          94          57.3171 %
Incorrectly Classified Instances        70          42.6829 %

=== Confusion Matrix ===

  a   b  c  d   <-- classified as
 58 27  0  0 |  a = Sufficient
 30 36  0  0 |  b = Not Sufficient
  4  3  0  0 |  c = Very Sufficient
  5  1  0  0 |  d = -1
```

Figure 19: Model by Naïve Byes with Dependent and Income vs Expenditure Sufficiency

Diet

The factor Diet was identified as one of the main factor affects to Quality of Life and under this section the analysis results are presented to find the affection of diet to the patient following CHD. The Member_Count variable was used since the data provided by the patients represent the consumption of food for the whole family.

Fruit_Usage	– how frequent fruits are used
Vegetable_Usage	– how frequent vegetable are used
Member_Count	– the number of members in the family
Oil_PerMonth	– amount of cooking oil used per month
Salt_ PerMonth	– amount of salt used per month
CookingMilk_PerDay	– amount of cooking milk used per day
BMI	– Body Mass Index (healthy underweight, over weight)

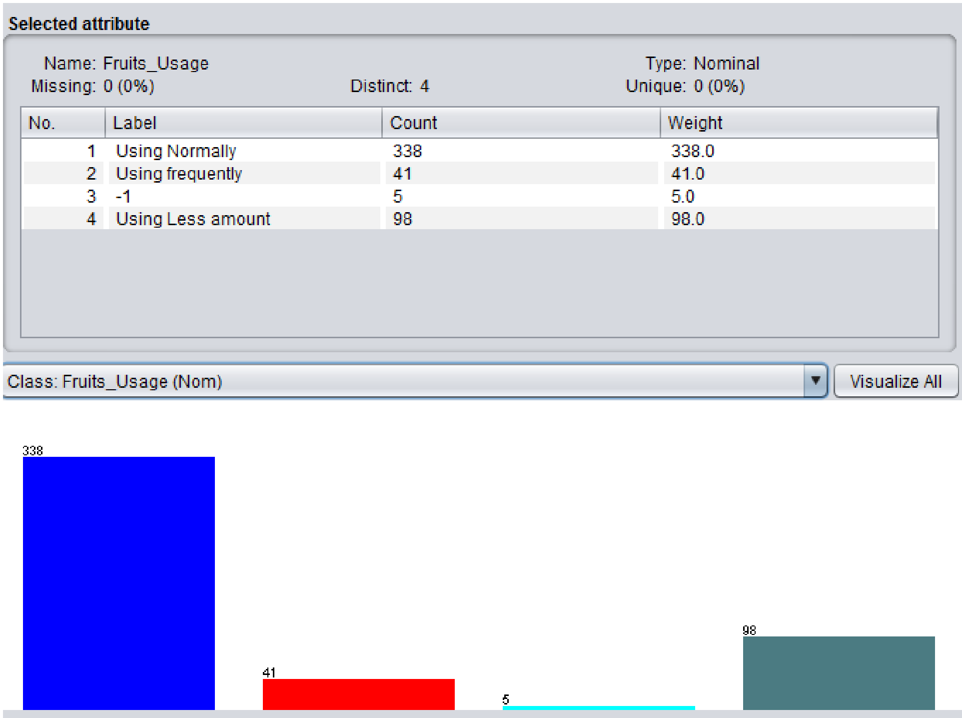


Figure 20: Fruit Usage

Selected attribute

Name: Vegetable_Usage
Missing: 0 (0%)
Distinct: 4
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Using frequently	179	179.0
2	Using Normally	279	279.0
3	-1	5	5.0
4	Using Less amount	19	19.0

Class: Vegetable_Usage (Nom) | Visualize All

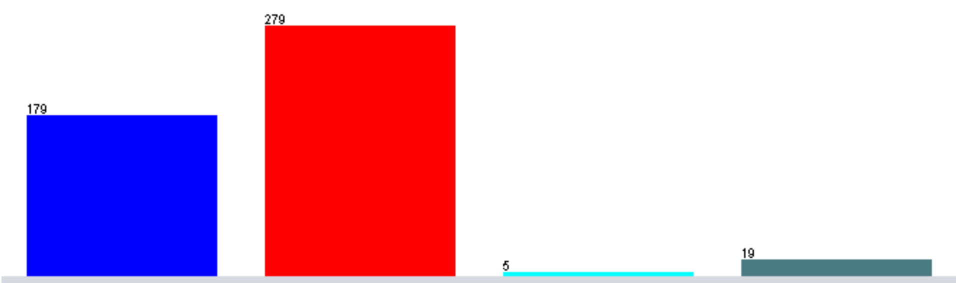


Figure 21: Vegetable Usage

Selected attribute

Name: Member_count
Missing: 0 (0%)
Distinct: 4
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	3 to 4	226	226.0
2	1 to 2	102	102.0
3	-1	5	5.0
4	more than 5	149	149.0

Class: Member_count (Nom) | Visualize All

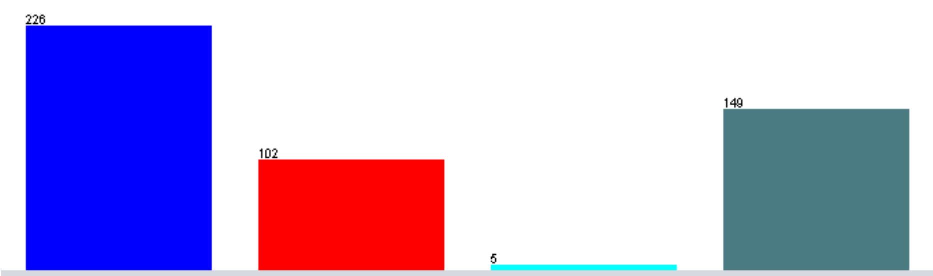


Figure 22: Member Count

Selected attribute

Name: Oil_PerMonth | Type: Nominal
Missing: 0 (0%) | Distinct: 5 | Unique: 0 (0%)

No.	Label	Count	Weight
1	Half a Bottle	80	80.0
2	Quarter Bottle	50	50.0
3	One Bottle	161	161.0
4	> One Bottle	154	154.0
5	-1	37	37.0

Class: Member_count (Nom) | Visualize All

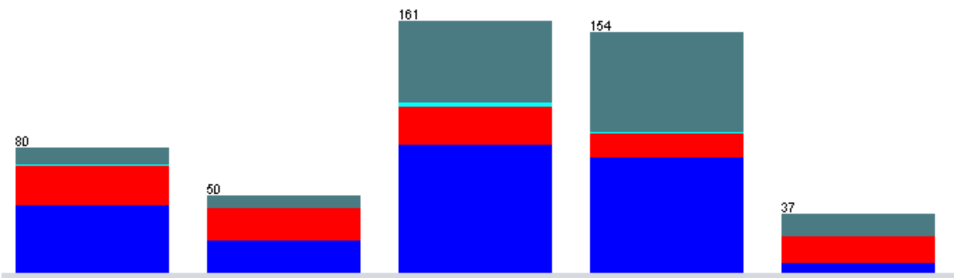


Figure 23: Oil per Month

Selected attribute

Name: Salt_PerMonth | Type: Nominal
Missing: 0 (0%) | Distinct: 5 | Unique: 0 (0%)

No.	Label	Count	Weight
1	< 500g	159	159.0
2	500g - 1kg	220	220.0
3	1kg - 2kg	60	60.0
4	> 2kg	20	20.0
5	-1	23	23.0

Class: Member_count (Nom) | Visualize All

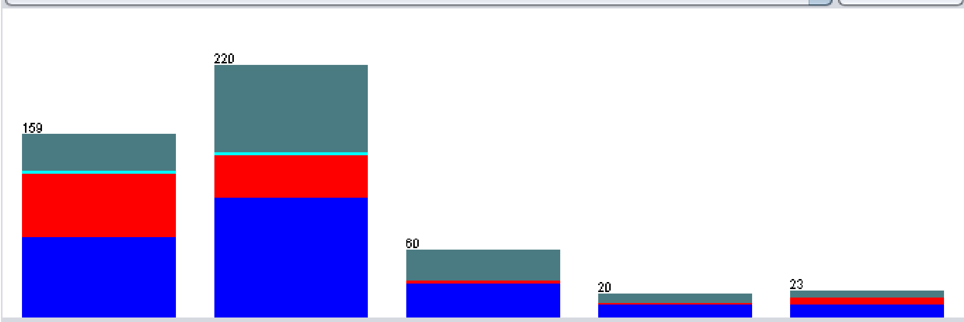


Figure 24: Salt per Month

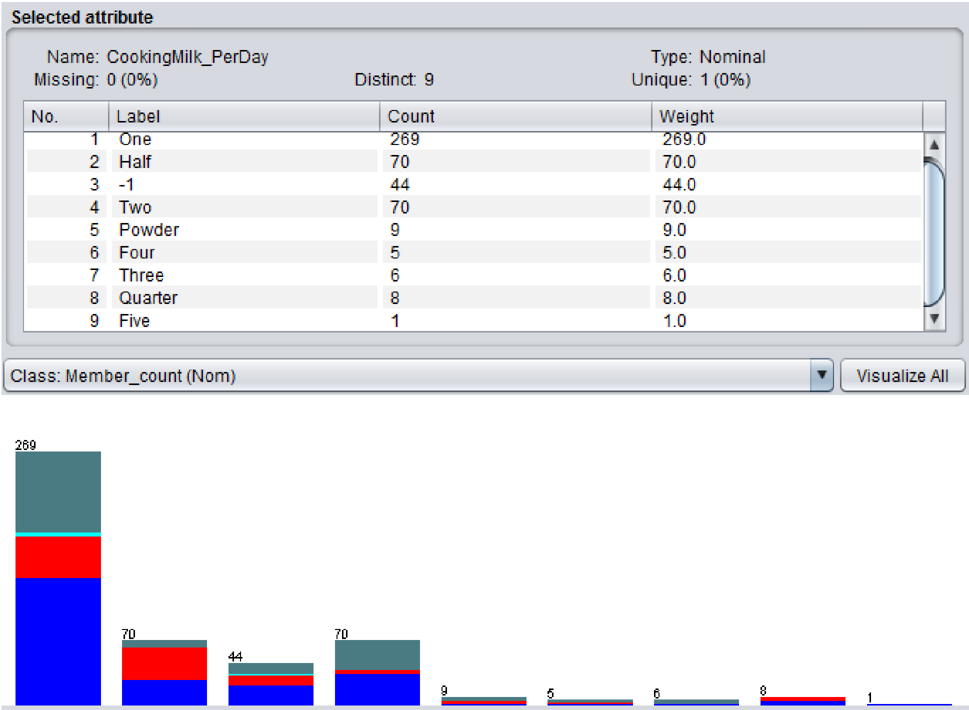


Figure 25: Cooking milk per Month

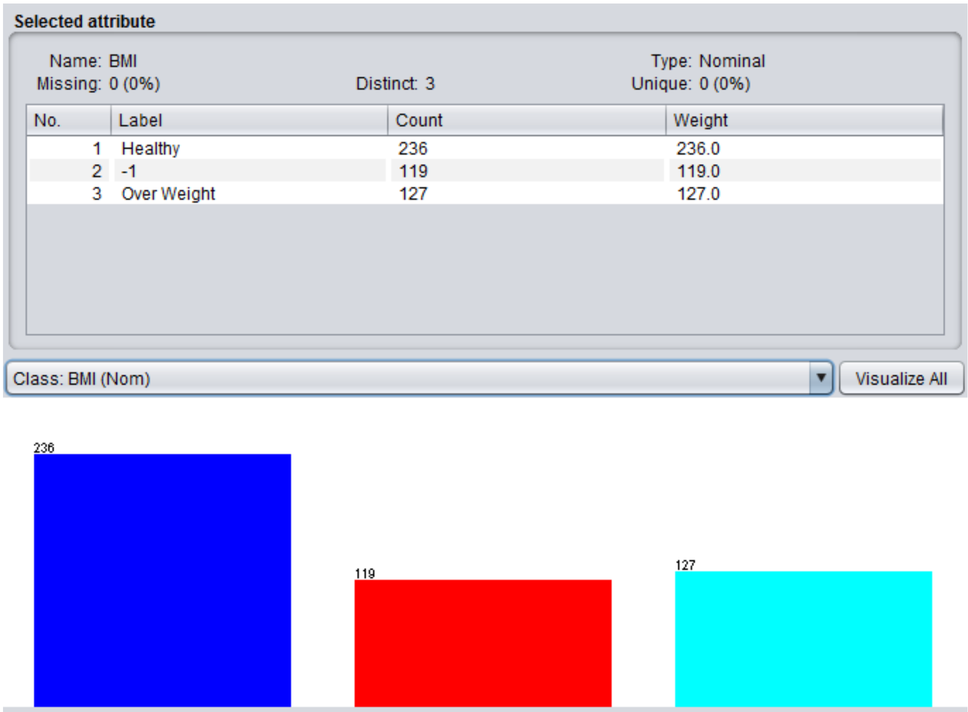


Figure 26: Body Mass Index

Smoking, Alcohol and Drugs

The patients' usage habits on smoking, alcohol and drugs were collected and analyzed. The current habit details and the habit details before occurring CHD were collected. Gender was considered since there was a significant relationship between these habits and the gender of the patient.

Drug_Using	– drugs are currently using or not
Drug_Used	– drugs were previously used or not
Smoking	– currently smoking or not
Smoked	– previously smoked or not
Alcohol_Using	– alcohol is currently using or not
Alcohol_Used	– alcohol was previously used or not
Gender	– gender of the patient
Bad Habits	– if any of the above habits have used or using then Yes, No otherwise

Association Rules

Attributes: 2
Drugs_Using
Drugs_Used
=== Associator model (full training set) ===

Apriori
=======
Best rules found:

1. Drugs_Used=No 463 ==> Drugs_Using=No 454 <conf:(0.98)> lift:(1) lev:(0) [0] conv:(0.96)
2. Drugs_Using=No 472 ==> Drugs_Used=No 454 <conf:(0.96)> lift:(1) lev:(0) [0] conv:(0.98)

Figure 27: Model by Association Rules with Drug Using and Drug Used

Association Rules

Attributes: 2
Smoking
Smoked
=== Associator model (full training set) ===

Apriori
=======
Best rules found:

1. Smoked=Yes 140 ==> Smoking=No 136 <conf:(0.97)> lift:(1.05) lev:(0.01) [6] conv:(2.09)
2. Smoked=No 342 ==> Smoking=No 310 <conf:(0.91)> lift:(0.98) lev:(-0.01) [-6] conv:(0.77)

Figure 28: Model by Association Rules with Smoking and Smoked

Association Rules

Attributes: 2
Alcohol_Used
Alcohol_Using
=== Associator model (full training set) ===

Apriori
=======
Best rules found:

1. *Alcohol_Used=Yes 100 ==> Alcohol_Using=No 96 <conf:(0.96)> lift:(1.05) lev:(0.01) [4] conv:(1.66)*
2. *Alcohol_Used=No 382 ==> Alcohol_Using=No 346 <conf:(0.91)> lift:(0.99) lev:(-0.01) [-4] conv:(0.86)*

Figure 29: Model by Association Rules with Alcohol Using and Alcohol Used

Association Rule

Attributes: 2
Gender
Bad Habits
=== Associator model (full training set) ===
Apriori
=======
Best rules found:

1. *Bad Habits=Have 168 ==> Gender=Male 161 <conf:(0.96)> lift:(1.74) lev:(0.14) [68] conv:(9.45)*

Figure 30: Model by Association Rules with Gender and Bad Habits

One Rule

Attributes: 2
Gender
Bad Habits
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Gender:
Male -> Have
Female -> No
-1 -> Have
(350/482 instances correct)

=== Summary ===

Correctly Classified Instances 123 75 %
Incorrectly Classified Instances 41 25 %

=== Confusion Matrix ===

```
 a   b   c   <-- classified as
72   0  26 |  a = No
 8   0   3 |  b = -1
 4   0  51 |  c = Have
```

Decision Tree

Attributes: 2
Gender
Bad Habits
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

Gender = Male: Have (265.0/104.0)
Gender = Female: No (216.0/28.0)
Gender = -1: Have (1.0)

Number of Leaves : 3
Size of the tree : 4

=== Summary ===

Correctly Classified Instances 123 75 %
Incorrectly Classified Instances 41 25 %

=== Confusion Matrix ===

```
 a  b  c   <-- classified as
72  0 26 |  a = No
 8  0  3 |  b = -1
 4  0 51 |  c = Have
```

Figure 31: Models by One Rule and Decision Tree with Gender and Bad Habits

Medical Treatments

The factor medical treatment was found as a factor affecting for the both mental and physical health. All the types of related diseases treatment details were collected and analyzed with compared to the mind status of the patient.

Treatment_Sufficiency – patient is satisfied with the treatments provided by the institution or not
Diabetes_Treatment_Taken – patient is following diabetes treatments properly or not
Hyperlipideamia_Treatment_Taken – patient is following hyperlipidemia treatments properly or not
Hypertension_Treatment_Taken – patient is following hypertension treatments properly or not
Stroke_Treatment_Taken – patient is following hypertension treatments properly or not
MI_Treatment_Taken – patient is following MI treatments properly or not
MI_Diagnosis – type of the MI (criticalness)
Living_Status_ Sufficiency – patient lives happily, sadly or normally
Health_Status_ Sufficiency – patients' health status
Situation_Symptoms – complications
TimeTaken_Symptomfree – time taken to resolve complication

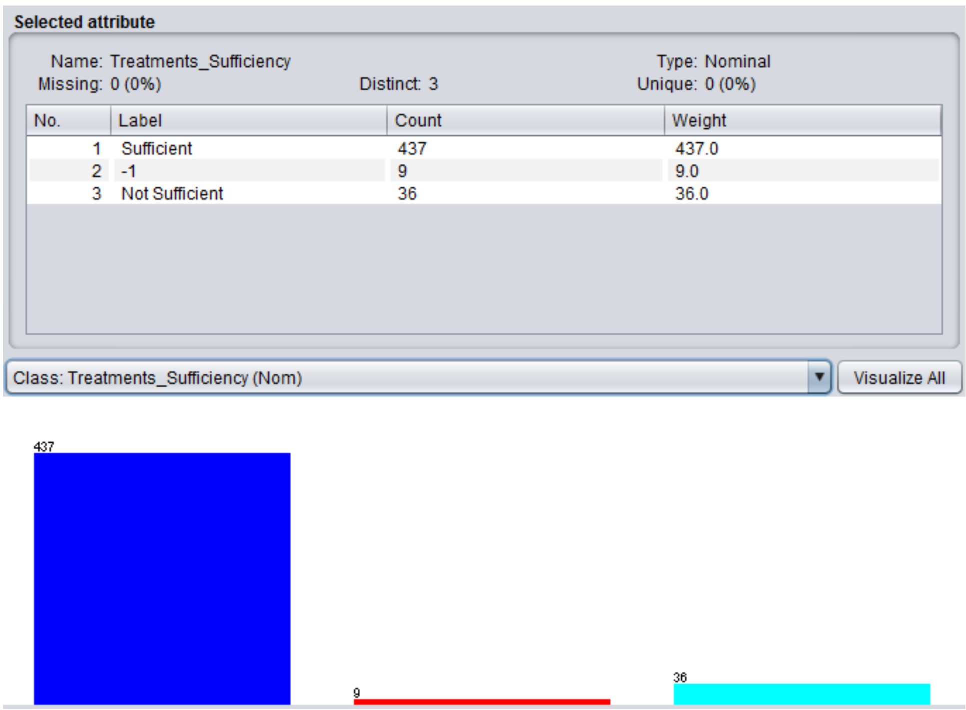


Figure 32: Treatment Sufficiency Histogram

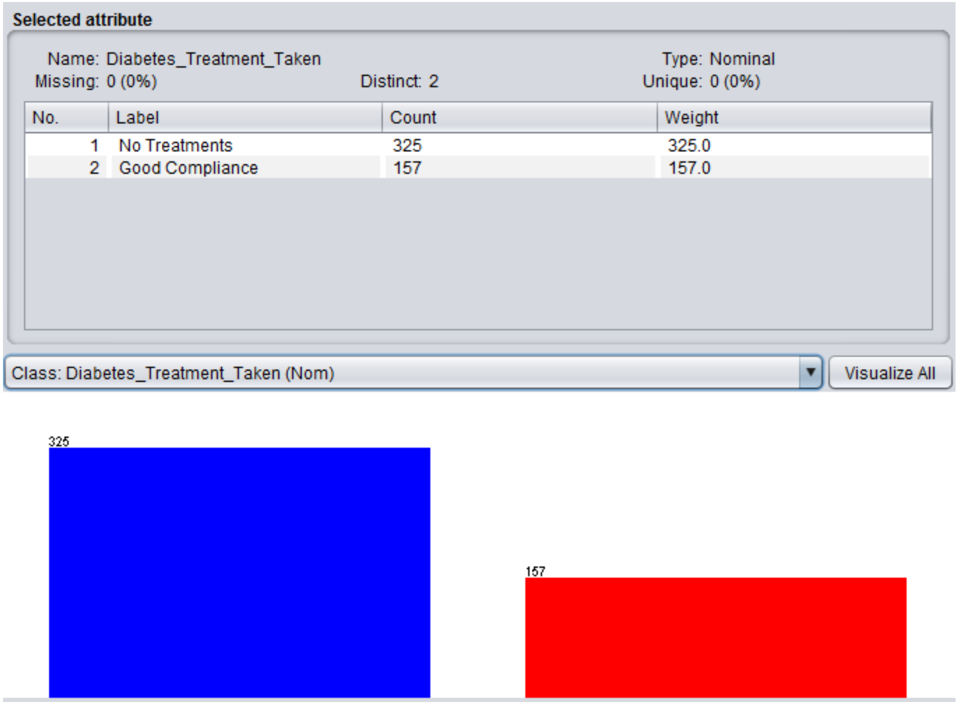


Figure 33: Diabetic Treatment Taken Histogram

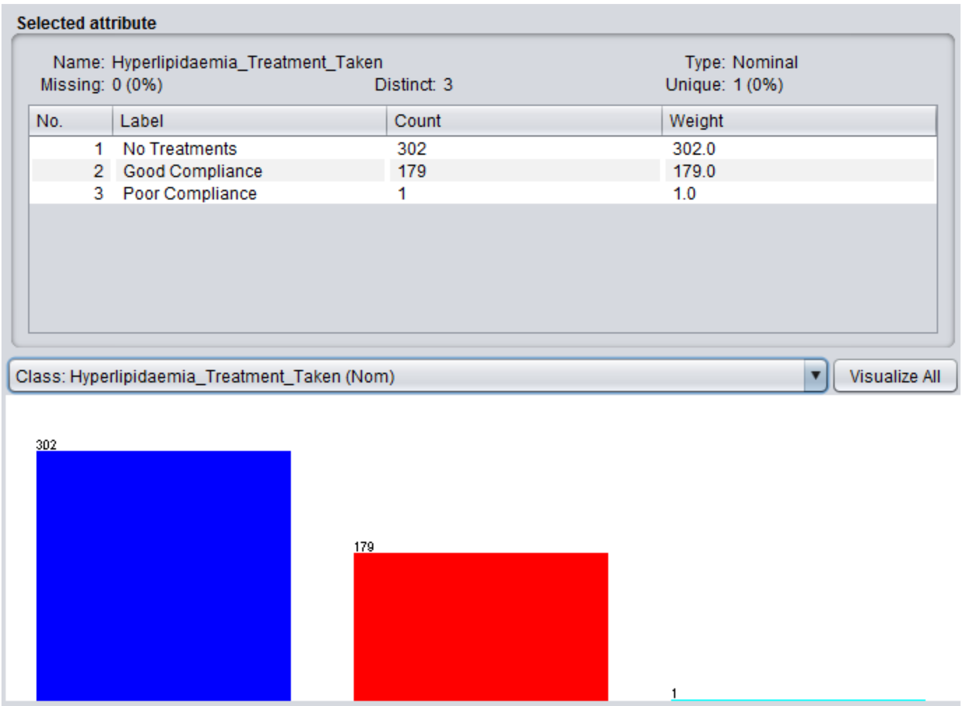


Figure 34: Hyperlipidemia Treatment Taken Histogram

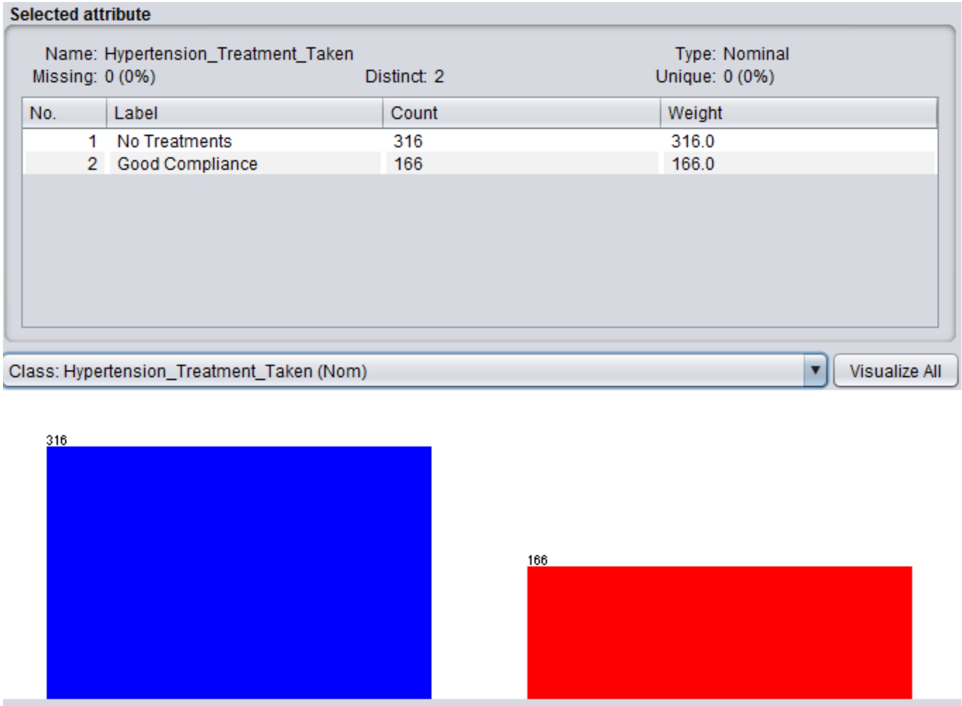


Figure 35: Hypertension Treatment Taken Histogram

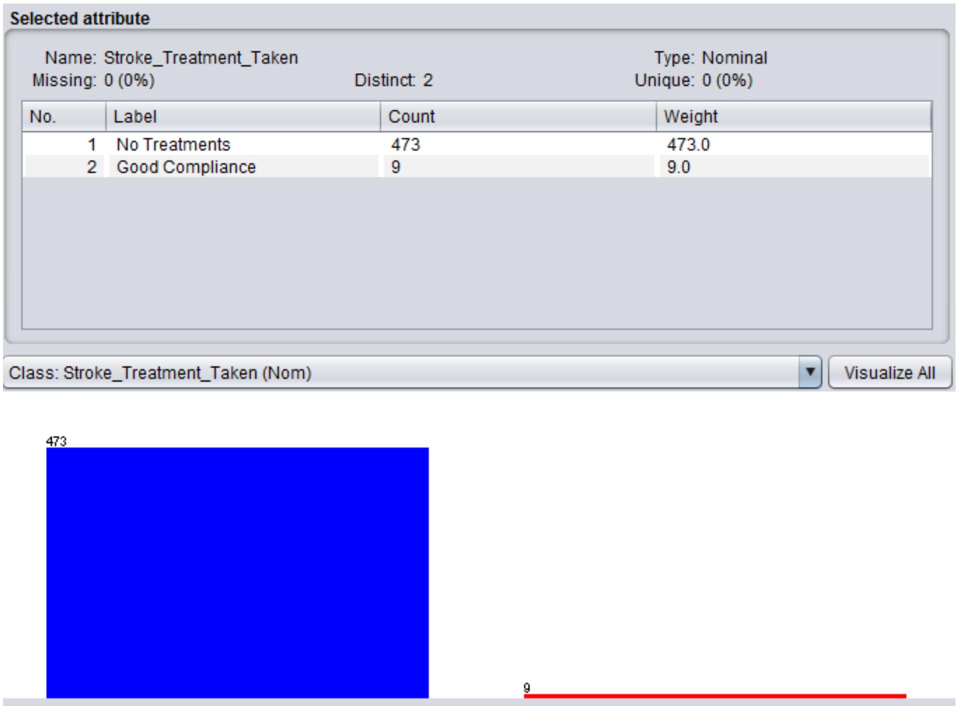


Figure 36: Stroke Treatments Taken Histogram

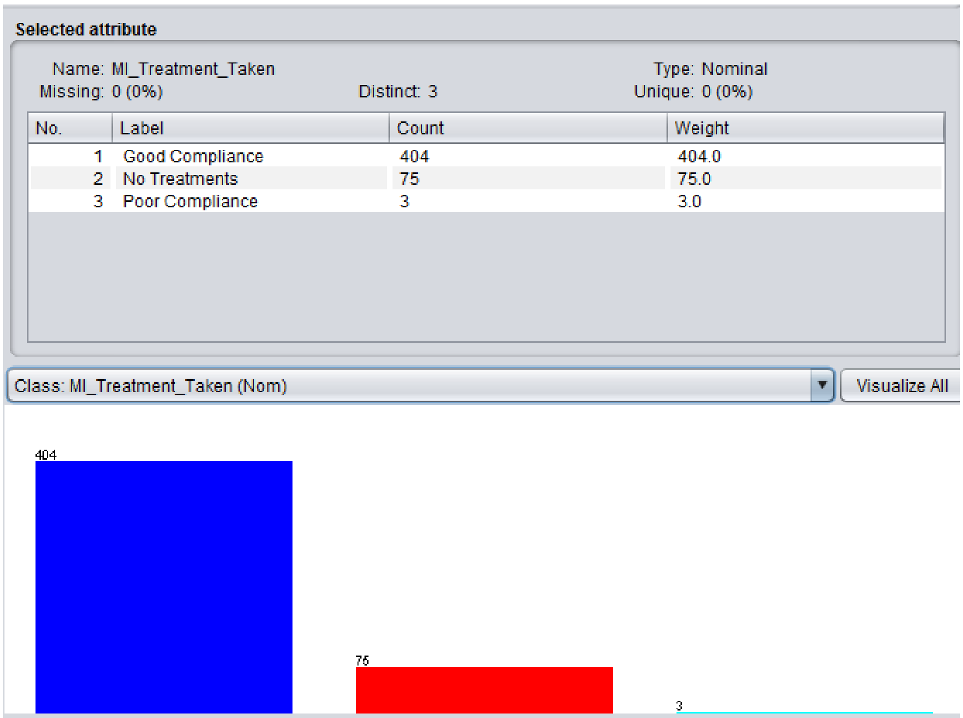


Figure 37: MI Treatment Taken

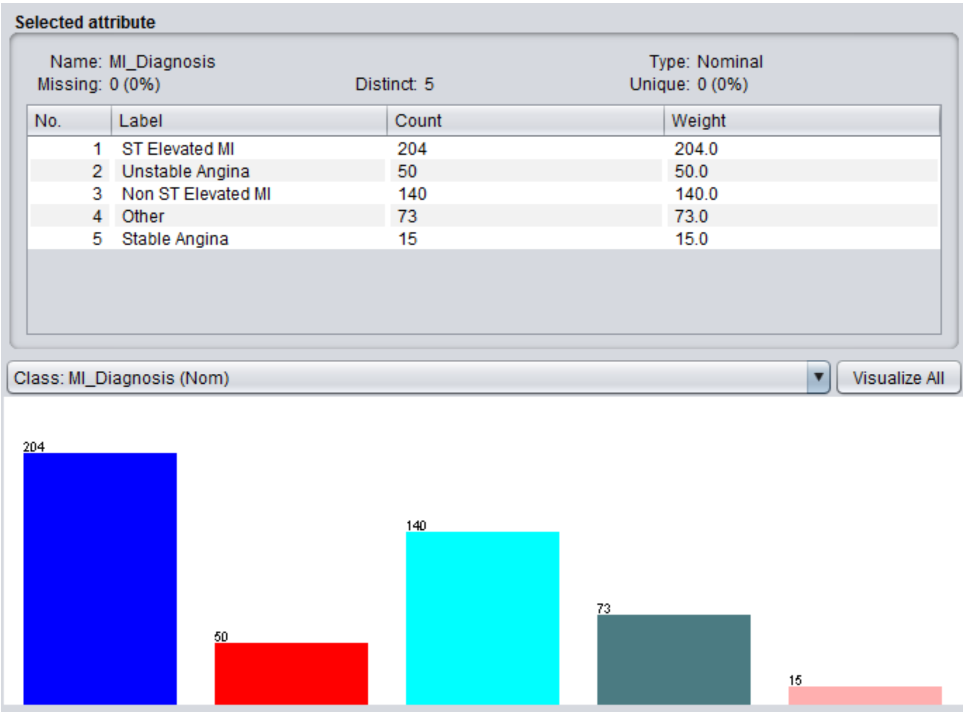


Figure 38: MI Diagnosis

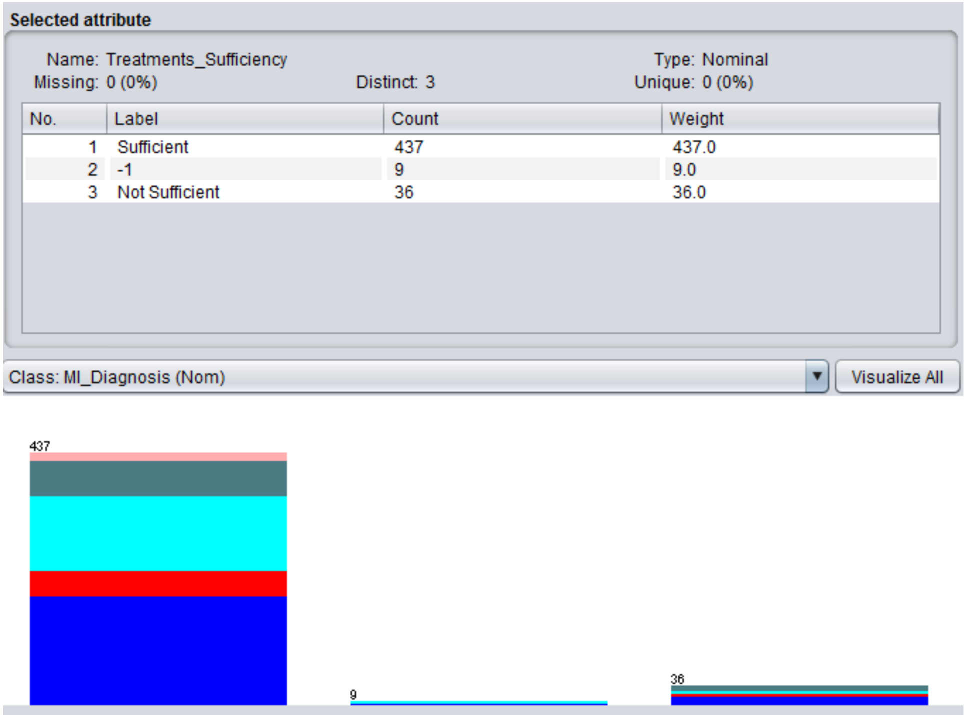


Figure 39: Treatments Sufficiency vs MI Diagnosis

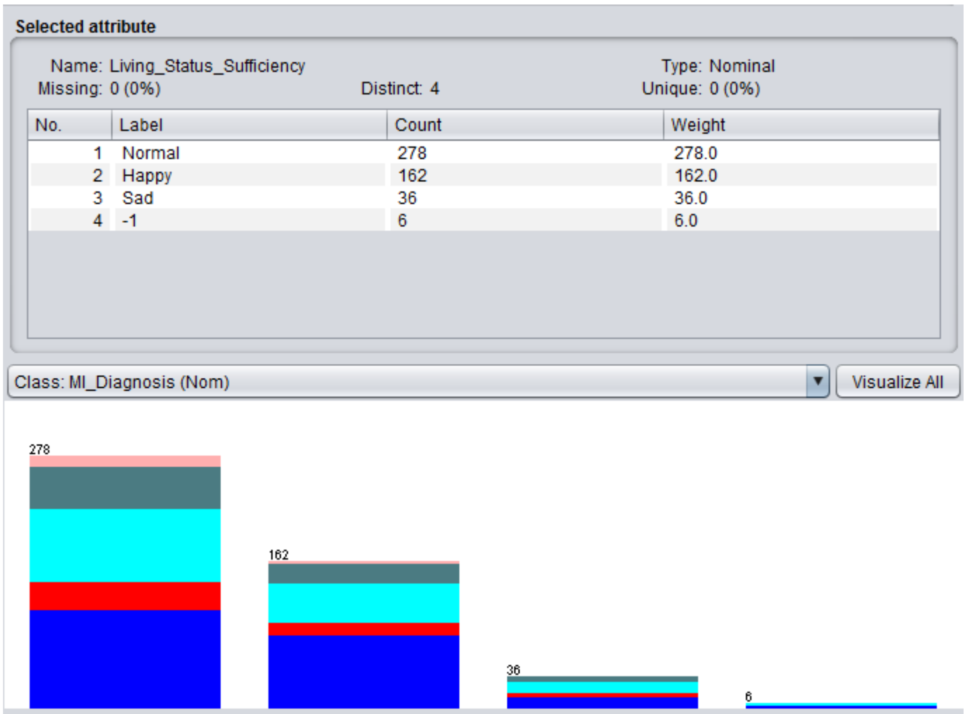


Figure 40: Happiness or Sadness Status vs MI Diagnosis

Selected attribute

Name: Health_Status_Sufficiency　　Type: Nominal
Missing: 0 (0%)　　Distinct: 6　　Unique: 0 (0%)

No.	Label	Count	Weight
1	Weak	92	92.0
2	Normal	316	316.0
3	Very Weak	19	19.0
4	Good	46	46.0
5	Very Good	6	6.0
6	-1	3	3.0

Class: MI_Diagnosis (Nom)　　Visualize All

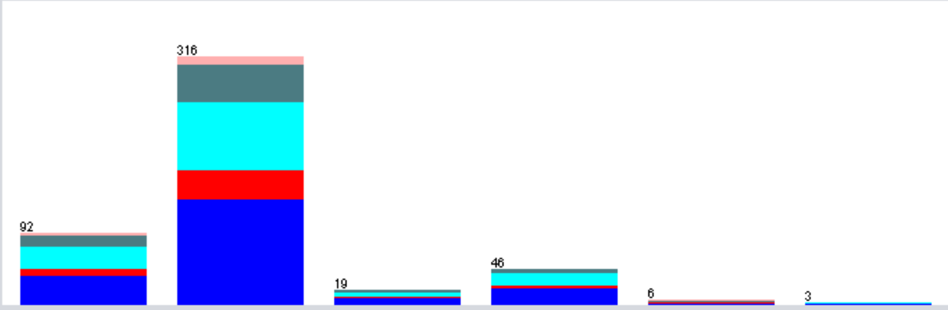


Figure 41: Health Status vs MI Diagnosis

Selected attribute

Name: Situation_Symptoms　　Type: Nominal
Missing: 0 (0%)　　Distinct: 6　　Unique: 0 (0%)

No.	Label	Count	Weight
1	Climbing steps or Hill	237	237.0
2	-1	8	8.0
3	Walking normally on a Flat Surf...	75	75.0
4	At rest	21	21.0
5	Not at all	60	60.0
6	Walking speedly on a Flat Surfa...	81	81.0

Class: MI_Diagnosis (Nom)　　Visualize All

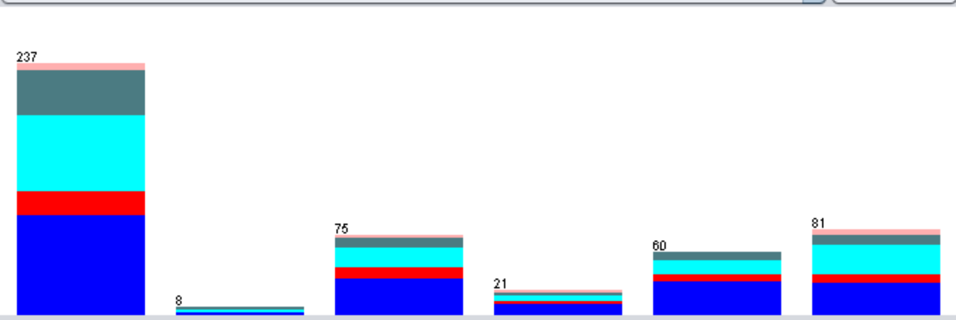
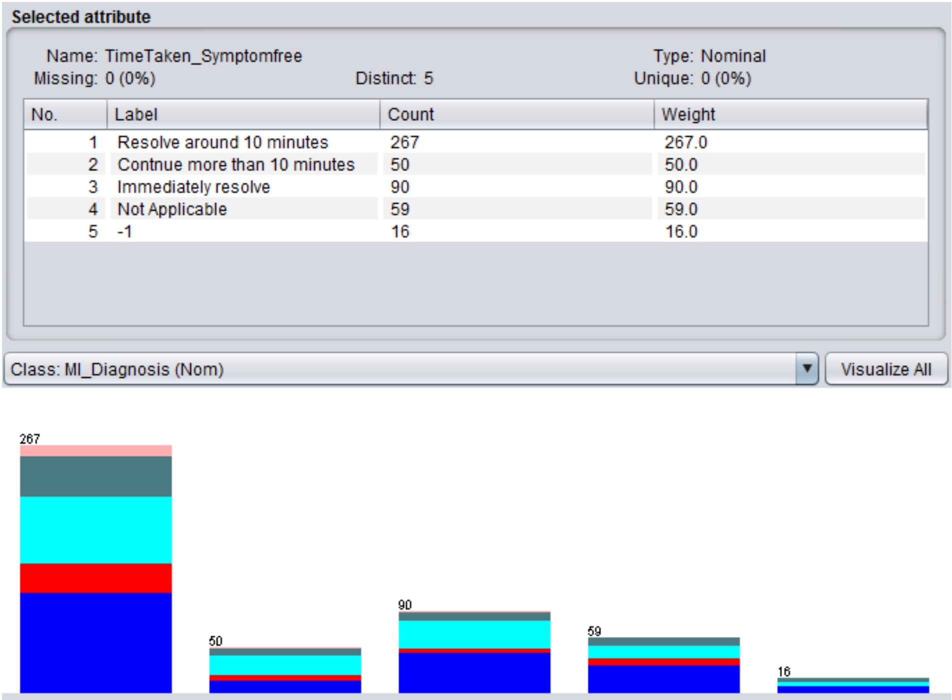


Figure 42: Situation of Complications vs MI Diagnosis



No.	Label	Count	Weight
1	Resolve around 10 minutes	267	267.0
2	Contnue more than 10 minutes	50	50.0
3	Immediately resolve	90	90.0
4	Not Applicable	59	59.0
5	-1	16	16.0

Figure 43: Time taken for resolving complication vs MI Diagnosis

Association Rule

Attributes: 2
Treatments_Sufficiency
MI_Diagnosis
=== Associator model (full training set) ===

Apriori
=======
Best rules found:

1. *MI_Diagnosis=Non ST Elevated MI 140 ==> Treatments_Sufficiency=Sufficient 129 <conf:(0.92)> lift:(1.02) lev:(0) [2] conv:(1.09)*
2. *MI_Diagnosis=ST Elevated MI 204 ==> Treatments_Sufficiency=Sufficient 187 <conf:(0.92)> lift:(1.01) lev:(0) [2] conv:(1.06)*

Figure 44: Model by Association Rules with Treatments Sufficiency and MI Diagnosis

```
Association Rule

Attributes:   2
          Treatments_Sufficiency
          Fortunate
=== Associator model (full training set) ===

Apriori
=======

Best rules found:

  1.  Fortunate=Yes 419 ==> Treatments_Sufficiency=Sufficient 397
                                          <conf:(0.95)> lift:(1.05) lev:(0.04) [17] conv:(1.7)
  2.  Treatments_Sufficiency=Sufficient 437 ==> Fortunate=Yes 397
                                          <conf:(0.91)> lift:(1.05) lev:(0.04) [17] conv:(1.39)
```

Figure 45: Model by Association Rules with Treatments Sufficiency and Fortunate

4.2 Discussions

The results found will be discussed by referring to each and every graph and statistical values illustrated in the section 4.1.

Marital Status

With correspondence to Marital Status globally the majority falls in to Nuclear Family category but in Sri Lanka most are under Extended Family. Elementary or Nuclear family refers to a family group consist of two parents and dependent children and the Extended family extend beyond the Nuclear family as it consist of relations such as aunts, uncles, grandparents and cousins all lives in the same house hold.

Relationship between each member of Extended family differs from Nuclear family due to set of reasons such as mutual understanding, care for each other, bond within the family and more. Normally the number of members of an Extended family is more compared to Nuclear family. But when the children turn in to 18 they will leave the parents' house and try to build their own family and that factor will directly affect on the family. Hence apart from the Marital Status factor this study will focus on the children, dependents and their bond between each other.

The following three variables namely Partner (describes whether the partner of the patient is living with the patient or not), Love and affection (if the patients has love and affection from someone they will mark as yes, no otherwise) and Fortunate (whether the patient think he/she is fortunate or not) are considered under the marital status factor.

In the Figure 4, the rule number 2 shows that 419 patients believe that they are fortunate if then they have someone to give love and affection for them with a number 412. This rule has a confidence interval of 0.98. The rule 7 describes that 455 patients believe they have love and affection. If then they are fortunate with a number of 412 and with confidence interval 0.91.

Furthermore the rule number 3 shows that 362 patients have their partners with them, if then 347 believes they have love and affection with a confidence interval 0.96.Rule number 8 shows that even though 104 patients do not have their partners with them, 94 of them have love and affection with the confidence 0.90. The rules 3 and 8 proves that patients have love and affection irrespective whether the partner lives with them or without them.

The rules 5 describes that out of 94 patients who lives alone (without partner) and have love and affection from someone, 86 of them are fortunate with confidence interval 0.91. The rule 6 describes that out of 347 patients who lives with their partner and have love and affection from someone, 316 of them think that they are fortunate with confidence interval 0.91.

Considering the results of above both rules a conclusion can be made as significant number of patients believe that they are fortunate when someone is available to give love and affection for them irrespective of whether the partner lives with or without them.

The above mentioned rules are proved furthermore by generating models using the techniques such as One Rule, Decision Tree and Naïve Bayes respectively.
The result illustrated in Figure 5, from the One-Rule technique furthermore proves the result from the association rule set discussed in the previous step. The above result shows that 432 patients believe they are fortunate when they have love and affection and they are not fortunate when they do not have love and affection in their life with a percentage of 87.2% correctly classified instances.

The result from the decision tree technique can be described as follows. 455 patients who believe that they have love and affection a less number of patients think that they are not fortunate with a number of 43 people which is insignificant. The percentage of correctly classified instances is 84.75%.

The Naïve Bayes technique result illustrated in Figure 6 shows that 319 patients whom partner lives with them believe they are fortunate while 40 of them believes that they are not with a ratio of 319:40 respectively. 91 patients whom their partner is not living with them believes that they are fortunate while 14 of them believes that they are not with the above same ratio. 413 patients those who have love and affection from someone are fortunate while 44 of them are not. This result has an amount of 87.2% correctly classified instances.

This result is again and again proved using One Rule, Decision Tree and Naïve Bayes techniques as follows.

The result from One Rule technique and Decision tree technique illustrated in Figure 7 shows that the factor, partner living with them or not is not linked with love and affection with a percentage of correctly classified instances 93.3%.

With the same percentage of correctly classified instances, the Figure 8th result shows that 348 patients whom their partner lives with, believe they have love and affection while 11 are not. This is insignificant. 95 patients whom their partner is not available thinks they have love and affection while 10 of them are not.

As the factor partner is not affected anymore for love and affection or fortunate, a thorough assessment was needed to elaborate to find the source of the love and affection they get from.

The three factors member count of the family, if the patient is a dependent or not and if he/she has help from others consider in to matter accordingly. Children and close relatives falls into member count of the family and the member count is calculated by including the patient as well. Dependent factor explains whether the patient is depend on others, other family members depend on the patient or if the patient lives independently. The factor help from other denotes that the patient have or do not have help from other when required.

In the Figure 9, the 3rd and 5th rules shows that in any case such as if the patient is a dependent on the children and the children depend on the patient, the patient have help from others with a significant ratio and with the confidence interval 0.94. The rule 8 describe that even though the patient is independent, still the help is available when needed with the confidence interval 0.91.

All the above association rules generated, demonstrates that irrespective of the scenario, help from others is available under any condition with the significant confidence interval amount of at least 0.90. Hence there is insignificant relationship in between the dependency and the help availability.

The 1st rule shows a relationship when member count is more than 5, help from others available with a confidence interval of 0.96, which is the highest. The rule number 4 describes when the member count is 3 to 4 the help from others available with a confidence interval of 0.94. The rule number 9 illustrates when member count is 1 to 2, help from other available with a confidence interval of 0.90.

The fact that the help available with the confidence interval 0.90, we can assume that the patient may satisfied with the help from the partner or in addition to that the patient may getting help from people other than his/ her family.

Furthermore the identified association rules have been generalized using the recent three data mining techniques as earlier.

The result predicted by the One Rule algorithms shows that the help available from others irrespective of the dependency with 449 correct instances with the correctly classified instances percentage of 89.63%. Result from the Decision tree shows that in any circumstance help is available from others with the same correctly classified instance percentage. The Naïve Bayes algorithm's result can be interpreted as when the member count of the family is more than five help is available with a significant ration of 139:12. A significant number 453 patients believe that they have help available irrespective of the dependency, whereas 33 says they don't. The correctly classified instance percentage is 89.6%.

As shown in Figure 10 and 11, a major amount of patients have help available from the family and when the number of members in the family increases the confidence of this relationship rises. Furthermore irrespective of the dependency the help does not vary.

Education and Socioeconomic Status

The Socioeconomic Status is measured from House hold income, Earners education and Occupation. Even though the Educational status is mentioned as a factor of Quality Of Life of an individual, in this study the Educational status will be considered as a part of Socioeconomic Status. Apart from these factors Religious believes will also be considered as a part of Socioeconomic Status.

The Figure 12 result presented by One Rule technique, came as patients need their religion to attain success in their life irrespective of the type of the Religion (Buddhist, Catholic, Hindu, Islam and Other) with a percentage of 90.24% correctly classified instances. The same relationship is proved by the decision tree technique result with the same correctly classified instances percentage value.

In Figure 13, result output y the Naïve Bayes technique, 450 patients whom majority are Buddhist believe that, they need religion to attain success in their life where as 44 of them not, with 90.24% percentage of correctly classified instances.

Furthermore this study is planned to elaborate the relationship between needfulness of the religion for the betterment of their life and living status, whether they are happy, normal or unhappy with their lives. Association rules were generated accordingly considering the above mentioned factors.

In the Figure 14, the rule number 3 shows that higher number of patients who thinks they live happily are fortunate, 155 out of 162 with a confidence interval 0.96. The rule number 4 shows when they are fortunate they need religious believes for their lives with a number of 394 out of 419 with a confidence interval 0.94. According to 7th rule when the living status is happy they need religious need for their lives with a number of 450 out of 452 with a confidence interval of 0.93. the rule number 6 further illustrate the combined link between all three attributes when living status is happy and when they are fortunate they need religious believes for their lives with a value of 445 out of 455 with a confidence interval 0.94.

The factor fortunate or not which is linked to mental health, is directly related to quality of life of patients with CHD as proved previously. According to the above mentioned rule set most of the patients think that, when they are fortunate they need religious believes for their lives. And also most of them who think they live happily, needs religious believes for their lives as well. Furthermore it was proved that irrespective of the type of the religion, a significant number of patients thinks that they need a religion for their lives. Hence this study proves that there is a relationship in between religion and fortunate status and living status which directly affects mental health of an individual. A significant number of patients states that they need a religion to believe.

The figure 15 Monthly income shows that majority of the patients are having a monthly income less than Rs.5000. Out of 482 instances there are 295 instances of them. Furthermore 82 instances are marked as they are having an income of Rs.5000 to Rs.15000. Rest of all the instances are marked as they get an income more than Rs.30,000 which is minority.

The Figure 16 with the graph Living status sufficiency out reveal that their living status as Normal, or Happy with a significant majority of patients, where very less number falls into the living status Sad.

The Figure 17 with the graph, current work states elaborate that significant amount of patients are currently not engaged in a job. Which means they are engaged in usual day to day activities which they are free from physical exhaustion of doing a job, as well as less complications. Ultimately leading them for a better physical health of the individuals.

In the Figure 18, the result by One Rule technique illustrates that patients have a sufficient income for the expenditure when they are independent or children depend on them. And income for the expenditure is not sufficient when they depend on children or parents with a percentage of 57.31% correctly classified instances. The same result is interpreted and proved by the result by decision tree technique with same percentage of correctly classified instances.

In the Figure 19 the Naïve Bayes technique expresses the income for expenditure of the patient as sufficient when they are independent and when children depend on them with ratios of 88:46 and 74:37 respectively. Their income for expenditure is not sufficient when they depend on children with a ratio of 117:86 with a percentage of 57.31% correctly classified instances.

Diet

The data was collected to find out the diet habits, amounts and their healthiness level, thus the member count of the family has to be compared when discussing the amounts that the patients use in the daily diet. Hence the factors such as the member count of the family, fruit usage per meal, vegetable usage per meal, salt usage per month, oil usage per month and finally coconut milk and different types of cooking oil usage per month will be taken in to matter when analyzing the diet habits of patients.

As mentioned in the Figure 20, a major amount of participants use fruit for their meal normally or frequently. This amount is significant since the total number of patients who use fruit normally and frequently is 379 out of 482 instances. Only 98 participants uses less amount of fruit for their meal. The histogram plotted in the figure demonstrates the statistics mentioned and it's clearly defines that they are trying to build up their health by using a proper diet.

The Figure 21 illustrates the usage of fruits for the day to day meal of the participants. Only an insignificant number of patients are using vegetables less for their meal and when compared to fruit usage, here the condition is at a good level. Only 19 patients out of 482 are using vegetable less for the meal. Both in the histogram and statistical table, it shows that, the vegetable usage is in a frequent or normal state with majority of the full data set. Around 458 instances under these two category out of 482 instances, with a percentage of 93.77%. Hence the diet of the patients seems to be good when referring to statistics illustrated.

The Figure 22 illustrates the distribution of the member count of families of patients. A minor number of patients are having one to two members in their family with including the patient himself/ herself. That is 102 instances out of 482, which is 21.16% from the whole data set. Most of the families have more than three members in their family. All together 355 instances have marked as, there are more than three members in their family and when comparing diets the study should consider this matter since most families have more people for the meal. Specially salt usage, oil usage and cooking milk usage should be evaluated by taking member count under consideration.

The histograms used for rest of the figures will be combined with member count. Hence the member count histogram will be needed when interpreting the histograms under salt usage, oil usage and cooking milk usage. The following colors and their representations should be taken under consideration.

RED : 1 to 2 members in the family

BLUE : 3 to 4 members in the family

GREY : Above 5 members in the family

LIGHT BLUE : Unmarked

According to the statistics shown in the Figure 23, 80 patients, 50 patients and 161 patients uses half a bottle, quarter a bottle and one bottle oil monthly, respectively. That is a total of 291 out of 482 with a percentage 60.37% which is comparatively higher amount. And only 154 number of patients are using more than one oil bottle per month which is with a percentage of 31.95% from the total data set.

It can assume that the patients use less amount of oil for their meal since 73.65% percent patients of total data set have more than 3 member in their family and 60.37% of patients from the total set are using one bottle or less amount of oil per month. Which is a good trend.

The Figure 24 shows that, in this data set 379 patients from total data set uses salt less than one kilogram per month, with a percentage of 78.63%. The first and second bars of the histogram illustrated in the figure 11, can be interpreted as follows. In the 2nd bar, there are more number of families with more than 5 members when compared to second bar. That amount is more than twice as the amount in the 1st bar. Which means more number of families with more than 5 members in their family uses 0.5 kg to 1 kg amount of salt per month.

It's a good trend that only less number of patients are using more than 1kg salt per month where as significant number of patients are using less amount of salt for their meal.

The Figure 25 illustrates cooking milk usage per day by the whole family. The statistics describe that more number of patients are using less than one coconut per day. One coconut, half a coconut and Quarter a coconut users are the majority in the histogram with the percentage of 71.99%. Even though there are 82 families who use two or more coconut per day, 70 out of them are using only two coconuts. Furthermore it is illustrated that even though there are many members in the family (at least 3 members), many families are trended to use less amount of cooking milk per day. Out of the total data set only 9 participants have denoted that they are using powdered cooking milk for their meal and that is a good practice as it is a processed food and natural food is the healthiest for the health.

The statistics shown in the Figure 26 can be discussed as follows. The height and weight data was collected from every participant of the survey and the Body Mass Index (BMI) could be calculated by applying the standard equation on the data. And finally the calculated values of BMI was used to interpret the healthiness of the patient. This was done by labelling three states as Healthy, Under Weight and Over Weight. Based on the values calculated, each patient were labelled and the figure 13 illustrates those statistics. It clearly shows that there are no underweight patients can be seen in the data and 236 patients are under healthy category. And only a percentage of 26.35% of patients are under over weight category.

It could found that there are no or very less number of underweight patients could be seen in the data set and a significant number from the total number of participants are in the healthy category. That 48.96% patients. Since this study is focused on patients with CHD, it’s please to see that many of them are in the healthy category.

Smoking, Alcohol and Drugs

Previous studies have proved that use of Cigarettes, Alcohol and Drugs have harmful effects which directly affects the Quality of life of an individual and the nearby people.

Throughout the data gathering phase it could see that higher number of patients are not currently using Cigarettes, Alcohol or Drugs. The Association rule technique was used to generate relationships between Smoking, Alcohol and Drug usage.

In order to find out current and previous smoking habits this study used two types of variables namely smoking and smoked. The same types of variables were created in order to assess alcohol usage as well as drug usage.

The Association rules generated in the Figures 27, 28 and 28 the following relationships are constructed. Significant number of patients who smoked earlier have stopped smoking following CHD. Out of 140 patients who smoked 136 are not smoking currently with a confidence interval of 0.97.96 patients out of 100 patients who were using alcohol have stopped using alcohol with a confidence interval 0.96 which is significant. It could found that majority of patients are not using or used drugs in their life.

Furthermore the association rules were made accordingly to evaluate the involvement of the gender of patients with CHD regarding Bad habits (Smoking, Alcohol, Drugs). The Figure 30 represents those rules generated. A new attribute was created by referring six existing attributes namely Smoking, Smoked, Alcohol Using, Alcohol Used, Drug Using and Drug Used. If any patient have used or using any of the above mentioned bad habits the new attribute will be labelled as Have and if not labelled as No.

It is found to have a link between gender and the bad habits when applying association rule technique on the new attribute. There were 168 patients under Bad Habits attribute labelled as Have, whom majority were males and they are 161 in number with a confidence interval of 0.96.

In the Figure 31, One Rule and Decision tree algorithms were applied on the same set of attributes to clarify the relationship found in the result by Association rule technique. Hence that rule was proved by the result interpreted.

The result generated by the One Rule technique describes that if the patients’ gender is male then he has bad habits and no otherwise. This rule is supported by 75% of correctly classified instances which is a significant value. The same relationship can be seen in the result output by decision tree technique with the same correctly classified instance percentage.

Medical Treatments

Medical treatments are found to be an affected area for the domain hence the data related to medical treatments provided were collected and analyzed. The following figures illustrates the statistics related to medical treatments in the point of view of the patients as well as the medical officers.

Treatments Sufficiency, Diabetic Treatment Taken, Hyperlipidemia Treatment Taken, MI Treatment Taken and MI Diagnosis attributes were considered when evaluating treatments provided for the patients.

Apart from those set of attributes Treatments Sufficiency, Living Status Sufficiency, Health Status Sufficiency, Situation Symptoms and Time Taken Symptom free attributes were compared against to MI Diagnosis in order find out participants' opinion compared to their criticalness of the disease.

The Figure 32 illustrates the patients' opinions on the treatments they receive. And a significant amount of patients says that they are satisfied with the treatments provided by the hospital with 90.66% out of whole data set.

The statistical values illustrated in the Figure 33 shows that there are only 157 patients who diagnosed with diabetes, which is 32.57% from the total data set. Furthermore these patients have marked that they receive treatments regularly.

The Figure 34 shows that there are 180 patients who are diagnosed with Hyperlipidemia and it could see that a greater number of them, follow the treatments for hyperlipidemia properly. That is 179 patients out of the set of 180 instances. Which is a good value.

Figure 35 shows a result which is similar to the result Figures 33 and 34. There are no patients diagnosed with a percentage of 65.83% significant majority and the minor amount of patients with the disease have good compliance of following medications.

As in the Figure 36, a significant amount of patients are not diagnosed with Stroke. Only 9 persons out of 482 set is diagnosed with the disease.

Figure 37 shows the statistics MI Diagnosis. Only 3 out of 407 set is marked as irregular treatments followed. Which is a good value.

According to the Medical officer's opinion, the diagnosis types namely ST Elevated MI and Non ST Elevated MI fall in to the severe CHD types. As mentioned in the statistics there are 204 patients with ST Elevated MI and 140 of Non ST Elevated MI. that is 71.34% of the total data set. Hence a considerable amount of participants have sever type of CHD

The Figure 38 to 43 will illustrate the statistical values for each attribute used against the MI Diagnosis types identified in order to see the relationships in between critical/ non critical diagnosis types and the patient's opinions on their health status. Hence the following colors will be used to discuss the statistics.

BLUE : ST Elevated MI

RED : Unstable Angina

LIGHT BLUE : Non ST Elevated MI

GREY : Other

LIGHT PINK : Stable Angina

As the Figure 39 shows, many patients bare the opinion that, they receive proper treatments irrespective of the severity of the diagnosis. Among 437 patients the majority of the first bar, represents Blue and Light blue color which indicate they had ST Elevated MI and Non ST Elevated MI which are the most severe types.

Significant amount of patients think that they receive proper medical treatments and care from the hospital which lead to a good mental health of the patients.

Even though there are significant sever patients in the data set, the Figure 40 illustrates that most of the patients are living happily or normally. And in Figure 41 it can see that there are only less number of patients who think they are weak or very weak patients in the data set. Hence the severity of the disease is not affected to the patients' mentality.

The situation of complications such as when at rest, Walking normally on a flat surface and walking speedily on a flat surface are the severe complications respectively. Most of the participants falls in to less complication category since significant number of them feel complications when climbing steps or hills and some of them do not face any complication at all. This statistics are listed in the Figure 42.

Figure 43 shows the relationship in between MI Diagnosis and complication resolving time. The result shows that if the complication is immediately resolved or resolved around 10 minutes is comparatively good state and significant number of patients falls into that category.

Figure 44 and 45 has listed down the possible relationships among the MI Diagnosis, Treatment sufficiency and Fortune variables. The relationships constructed prove that, even though with the most critical two types of diagnosis the patients are with the idea of that they receive proper treatments. This relationship supports with a 0.92 confidence which is significant. Furthermore a huge amount of instances out of the whole data set think they are receiving sufficient treatments and they are fortunate, vice versa.

Chapter 5

CONCLUSION AND FUTURE WORK

Based on the all the Association Rules listed Figure 4, it can conclude that a significant amount of patients think that they are fortunate when they have love and affection, and vice versa. And irrespective of the partner, most of the patients believe that they have love and affection from someone and if then, they are fortunate.

The elaborated results set in the Figure 5 and 6 using the three techniques proves that a significant amount of patients think that they are fortunate when they have love and affection and fact of partner is out of the domain and there will be someone to offer love and affection.

Results demonstrated in the Figure 7 and 8 are also proving that a significant amount of patients think that they have someone to give love and affection irrespective of their partner lives with them or not. Hence the partner factor is not affected anymore.

Based on the Association Rules constructed in the Figure 9 it can state that, significant amount of patients have help available from the family and when the number of members in the family increases the confidence of this relationship rises. Furthermore irrespective of the dependency the help does not vary. A new relationship specifically for Sri Lankan context have been identified as people are more generous to offer help to others when they are in need of help.

According to Figure 10 and 11, the availability of help from others is a serious factor which directly affect an individuals' both physical and mental health. The availability of someone to offer love and affection and the mental status about fortunate or not, are directly link with mental health of an individual. And it can conclude that a significant amount of patients have help available but when the number of members in the family increases the percentage as well as the ratio of help availability increases.

Based on the result interpreted in Figures 12 and 13 it can state that, whatever the religion, significant number of participants thinks that they need religious believes in succeeding their life hence they follow their religion.

The Figure 14 denotes that whatever the religion, many participants thinks that they need religious believes in for the betterment for their life. And a significant number of participants who live happily and believe in their religion thinks they are fortunate with a high number of confidence. Hence their mental status is at an optimum condition.

According to the Figures 15 and 16 it is proved that a higher amount of patients have a less income. But very less number of participants are in the state sad. And significant number of people are Normal or Happy with their lives. Therefor irrespective of the income they get generally, the living status of the patients are maintained as Happy or Normal which directly denotes patients' mental wellbeing.

The result from the figures 16, 17 and 18 can be interpreted as follows. If the individuals are independent, or the children depend on them then they have sufficient income to bare expenditure. But the individuals depend on their children have an insufficient income. Even though more than half of the patients are currently not working, most of these patients thinks that they have sufficient money to handle their expenses.

The diet factor was monitored by referring the Figures 20 to 26.According to the statistical values mentioned in the figures, it is highlighted that salt usage, oil usage and the cooking milk usage for their meal is in a lower level, which is a healthier practice. Even though there are many patients with higher number of members in the family, they use similar, less amount of those ingredients. And even though these statistical data was collected from patients, it can see that majority of them are in the healthy state in the BMI categorization.

According to results interpreted based on the Figures 27, 28 and 29, almost all the patients who have used Alcohol or Cigarettes has stopped using them following the illness. And it could see that almost all the patients have not or is not using Drugs. Ultimately leading to ended bad habits by improving physical wellbeing of them.

When going through the data set, it could see that most of the female patients are not using or used any of these bad habits mentioned. Hence it could prove that among the addicted people the male patients has a significant higher count compared to females. This conclusion is supported by the Figures 30 and 31.

By referring to the results interpreted based on the statistics shown in Figure 32 it can state that a significant number of patients are satisfied with the treatments received from the hospital and that leads to both good mental health and physical health of the patient.

The data is collected from the monthly clinic of Cardiology Clinic shows that almost all the patients have a good compliance on the monthly follow ups on their illnesses. These data could be referred from the Figures 33 to 37 where the main non-communicable diseases statistics illustrated.

The following conclusion is constructed by referring the result illustrated in Figures 38 to 43. Almost all the patients thinks that they receive proper treatments for their diseases. Even though most of them are having severe type of CHD their mental health is at a good condition and most of them shows less complications and good state on recovering complications.

The following theories can be listed down based on the each conclusion constructed thought the analysis of the study.

- Most of the participants are in a good mental state because of they think that,
  - They are fortunate,
  - They have care form others,
  - They receives love and affection from others
  - They think that they are not sad
  - They believe in religious activities affect their health
  - They receive medical treatments properly

- Most of Patients have a good physical states due to,
  - They are not working currently and have rest
  - They have less complications

They receive medical treatments continuously
They have a controlled diet
They have abandoned bad habits from their lives

With considering all the above factors and results which were interpreted and proved using different data mining techniques emphasizes validity of the hypothesis.

Hence based on the discovered knowledge the null hypothesis H_0**:** *The patients with Coronary Heart Disease has a lower quality of life* is disproved and ignored.

The alternative hypothesis H_1**:** *The patients with Coronary Heart Disease has a higher quality of life* is proved and accepted.

Due to above factors proved the patients with CHD has a good physical and mental health which leads to a higher quality of life.

This study can be enhanced by considering patients form different geographical areas in Sri Lanka due to many reasons such as cultural differences, climate differences and much more factors.

REFERENCES

[1] "Cardiovascular disease," World Health Organization, [Online]. Available: http://www.who.int/cardiovascular_diseases/en/.

[2] "Economy of Sri Lanka," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Economy_of_Sri_Lanka.

[3] "Elders Homes in Sri Lanka," The Island, [Online]. Available: http://www.island.lk/index.php?page_cat=article-details&page=article-details&code_title=35745.

[4] "Quality of life (healthcare)," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Quality_of_life_(healthcare).

[5] "Data mining techniques to find out heart diseases: an overview," *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 4, no. III, pp. 80-84, 2016.

[6] S. MENDIS, "Major Risk Factors of Coronary Heart Disease in Sri Lankans," *Journal of the Ceylon College of Physicians.,* pp. 17-34, 1991.

[7] G. M. Oppenheimer, "Framingham Heart Study: The First 20 Years," *Progress in cardiovascular diseases,* pp. 55-61, 2010.

[8] M. S. BROWN and J. L. GOLDSTEIN, "A RECEPTOR-MEDIATED PATHWAY FOR CHOLESTEROL HOMEOSTASIS".

[9] S. U. B. Tennakoon, "Cardiovascular risk factors and predicted risk of cardiovascular disease among Sri Lankans living in Kandy, Sri Lanka and Oslo, Norway," 2012.

[10] T. DURMAZ, Ö. ÖZDEMİR, B. A. ÖZDEMİR, T. KELEŞ, N. A. BAYRAM and E. BOZKURT, "Factors affecting quality of life in patients with coronary heart disease," *Turkish Journal of Medical Sciences,* pp. 343-351, 2009 .

[11] K. Bjordal, "Impact of quality of life measurement in daily clinical practice," *Annals of Oncology 15,* 2004.

[12] "Quality of life in patients with coronary heart disease-I: Assessment," *Health and Quality of Life Outcomes (BioMed Central),* vol. 16, 2013.

[13] A. Methaila, P. Kansal, H. Arya and P. Kumar, "Ealy heart disease prediction using data mining techniques," *Conference: Fourth International Conference on Computational Science, Engineering and Information Technology*, pp. 53-59, 2014.

[14] S. U. Amin, K. Agarwal and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," *IEEE Conference on Information & Communication Technologies*, pp. 1227-1231, 2013.

[15] C.-D. Chang, C.-C. Wangb and B. C.Jianga, "Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors," *Expert Systems with Applications,* vol. 38, no. 5, pp. 5507-5513, 2011.

[16] S. Gnanasoundhari, G. Visalatchi and M. Balamurugan, "A survey on heart disease prediction system using data mining techniques," *International Journal of Computer Science and Mobile Applications,* vol. 2, no. 2, pp. 72-77, 2014.

[17] "Weka (machine learning)," [Online]. Available: https://en.wikipedia.org/wiki/Weka_(machine_learning).

[18] "Weka 3: Data Mining Software in Java," Waikato, [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/.

[19] "IBM SPSS software," IBM, [Online]. Available: https://www.ibm.com/analytics/spss-statistics-software.

[20] "Wikipedia," [Online]. Available: https://en.m.wikipedis.org/wiki/Categorical_variabl.

APPENDIX A Questionnaire

ඔබ සපයන මෙම තොරතුරු වල රහසිගත භාවය ආරක්ෂා කෙරෙන අතර එම තොරතුරු අධ්‍යාපන කටයුතු සඳහා පමණක් භාවිතා කරන බව දන්වා සිටින්නෙමි.

අදාළ කොටුව තුල x සළකුණ යෙදීමෙන් සහ කඩ ඉරිමත ලිවීමෙන් පිළිතුරු සපයන්න

1. **ඔබ ජීවත් වන්නේ,** නාගරික ප්‍රදේශයක ☐ අර්ධ නාගරික ප්‍රදේශයක ☐ ග්‍රාමීය ප්‍රදේශයක ☐
2. **ස්ත්‍රී පුරුෂ භාවය :-** ස්ත්‍රී ☐ පුරුෂ ☐
3. **ඔබ කාන්තාවක් නම් ඔබගේ මාසික ඔසප් වීම සම්පූර්ණයෙන් නැවතුන (ආර්තවහරණය වූ) අයෙක් ද?** ඔව් ☐ නැත ☐
4. **වයස :-** 18ට අඩු ☐ 19ත් -40 අතර ☐ 41ත්-60අතර ☐ 60ත්-70අතර ☐ 70ට වැඩි ☐
5. **ඔබගේ ආගම,** බෞද්ධ ☐ කතෝලික ☐ ඉස්ලාම් ☐ හින්දු ☐ වෙනත් ☐
6. **ඔබ,** විවාහකයි ☐ අවිවාහකයි ☐
7. **ඔබගේ සහකරුවා/සහකරිය ඔබ සමග,** සිටියි ☐ නොසිටියි ☐
8. **ඔබගේ උපරිම අධ්‍යාපන සුදුසුකම කුමක් ද?**
   පාසල් අධ්‍යාපනයක් නොලැබුවෙමි ☐ ප්‍රාථමික අධ්‍යාපනය ලැබුවෙමි ☐ 8 වසර දක්වා අධ්‍යාපනය ලැබුවෙමි ☐
   සාමාන්‍ය පෙළ දක්වා අධ්‍යාපනය ලැබුවෙමි ☐ උසස් පෙළ දක්වා අධ්‍යාපනය ලැබුවෙමි ☐ උපාධිධාරියෙක්මි ☐
9. **ඔබගේ උස** :- අඩි .... අඟල් .... **ඔබගේ බර** :- .... kg
10. **ඔබගේ නිවසේ ඔබත් ඇතුලත්ව කොපමණ සාමාජිකයන් සංඛ්‍යාවක් සිටී ද?**
    1ත් 2ත් අතර ☐ 3ත් 4ත් අතර ☐ 5ට වැඩි ☐
11. **ඔබ,** දරුවන් ගෙන් යැපේ ☐ දරුවන් ඔබගෙන් යැපේ ☐ ස්වාධීනයි ☐
12. **ඔබගේ රැකියාව කුමක් ද?** ....
13. **ඔබ දැනට රැකියාවේ නිරත වෙනවා ද?** ඔව් ☐ නැත ☐
14. **ඔබ මාසික ආදායම,**
    රු.0 – රු.5,000 ☐ රු.5,000 – රු.15,000 ☐ රු.15,000- රු.30,000 ☐
    රු.30,000- රු.50000 ☐ රු.50000 ට වැඩි ☐
15. **ඔබ දුම් පානය කරනවා ද?** ඔව් ☐ නැත ☐
16. **ඔබ දුම් පානය කරන්නේ නම් දුම් පානය ආරම්භ කර දැනට කොපමණ කාලයක් ද?**
    දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ අවුරුදු 5ක් ☐
    පමණ අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐
17. **ඔබ දුම් පානය කර නැවතු අයෙක් ද?** ඔව් ☐ නැත ☐
18. **ඔබ දුම් පානය කර නැවතු අයෙක් නම් දුම් පානය නවතා දැනට කොපමණ කාලයක් ද?**
    දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ අවුරුදු 5ක් පමණ ☐
    අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐
19. **ඔබ දුම් පානය කරන්නේ හෝ දුම් පානය කලේ නම් දිනකට කොපමණ ප්‍රමාණයක් දුම් පානය කරන්නේ ද?**
    දුම් වැටි 2ක් පමණ ☐ දුම් වැටි 5ක් පමණ ☐ දුම් වැටි 10 ක් පමණ ☐ දුම් වැටි 10ට වැඩි ☐

20. භාවිතා කරනු ලබන වර්ගය :- බීඩි ☐ සිගරට් ☐ දෙවර්ගයම ☐

21. ඔබ දැනට මත්පැන් පානය කරනවා ද? ඔව් ☐ නැත ☐

22. ඔබ දැනට මත්පැන් පානය කරන්නේ නම් ඒ කොපමණ කාලයක සිට ද?

දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ අවුරුදු 5ක් පමණ ☐

අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐

23. භාවිතා කරනු ලබන වර්ගය

බොතල් කරන ලද දේශීය ☐ බොතල් කරන ලද විදේශීය ☐ බොතල් නොකරන ලද (කසිප්පු) ☐

24. ඔබ මත්පැන් පානය කර නැවතූ අයෙක් ද? ඔව් ☐ නැත ☐

25. ඔබ මත්පැන් පානය කර නැවතූ අයෙක් නම් ඒ කොපමණ කාලයක සිට ද?

දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ අවුරුදු 5ක් පමණ ☐

අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐

26. ඔබ මත්පැන් පානය කරන්නේ හෝ පානය කලේ නම් දිනකට කොපමණ මත්පැන් ප්‍රමානයක් භාවිතා කරන්නේ ද?

බෝතල් කාලට අඩුවෙන් ☐ බෝතල් කාලක් ☐ බෝතල් බාගයක් ☐ බෝතලයක් ☐ බෝතලයට වැඩි ☐

27. ඔබ මත්ද්‍රව්‍ය භාවිතා කරනවා ද? ඔව් ☐ නැත ☐

28. ඔබ මත්ද්‍රව්‍ය භාවිතා කරන්නේ නම් ඒ කොපමණ කාලයක සිට ද?

දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ අවුරුදු 5ක් පමණ ☐

අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐

29. භාවිතා කරනු ලබන වර්ග, ගංජා ☐ හෙරොයින් ☐ අබිං ☐ ඇෂ් ☐ බාබුල් ☐

30. ඔබ මත්ද්‍රව්‍ය භාවිතා කර නැවතූ අයෙක් ද? ඔව් ☐ නැත ☐

31. ඔබ මත්ද්‍රව්‍ය භාවිතා කර නැවතූ අයෙක් නම් ඒ කොපමණ කාලයක සිට ද?

දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ ඇවුරුදු 5ක් පමණ ☐

අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐

32. ඔබ මත්ද්‍රව්‍ය භාවිතා කරන්නේ හෝ භාවිතා කලේ නම් දිනකට කොපමණ ප්‍රමානයක් භාවිතා කරන්නේ ද?

ග්‍රෑම් 1ට අඩු ☐ ග්‍රෑම් 1ක් පමණ ☐ ග්‍රෑම් 2ක් පමණ ☐

ග්‍රෑම් 3ක් පමණ ☐ ග්‍රෑම් 4ක් පමණ ☐ ග්‍රෑම් 5ක් පමණ ☐ ග්‍රෑම් 5ට වැඩි ☐

33. ඔබගේ නිවසේ ආහාර සැකසීම සඳහා සාමාන්‍යයෙන් මසකට කොපමණ ප්‍රමානයක් තෙල් භාවිතා කරනවා ද?

බෝතල් කාලයි ☐ බෝතල් භාගයයි ☐ බෝතලයයි ☐ බෝතලයට වැඩි ☐

34. ඔබ මේ අතරින් භාවිතා කරන්නේ කුමන තෙල් වර්ගය වර්ග ද ?

පොල් තෙල් ☐ එළවලු තෙල් ☐ සූරියකාන්ත තෙල් ☐ ඔලිව් තෙල් ☐ වෙනත් ☐

35. ඔබ එම තෙල් ලබා ගන්නේ කෙසේද ?

නිවසේදී පිළියෙල කරන ලද තෙල් ☐ බෝතල් නොකල තෙල් ☐ බෝතල් කල තෙල් ☐

36. එම තෙල් ලබා ගන්නේ කුමන ස්ථානයකින් ද? නිවසේදී පිළියෙල කිරීමෙන් ☐ සති පොළෙන් ☐

තොග හෝ සිල්ලර කඩයකින් ☐ සුපිරි වෙළඳසැලකින් ☐

37. ඔබගේ නිවසේ ආහාර සැකසීම සඳහා යොදා ගන්නේ

දිය කිරි පමණක් ☐ මිටි කිරි පමණක් ☐ දිය කිරි, මිටි කිරි යන දෙවර්ගයම ☐

පිටි කල පොල් කිරි පමණක් ☐ පිටි කල, ස්වභාවික පොල් කිරි යන දෙවර්ගයම ☐

38. දිනකට ආහාර පිසීමට කොපමණ පොල් ප්‍රමාණයක් අවශ්‍යද? ගෙඩි / මේස හැඳි ගනන (පිටි කල පොල් කිරි නම්)

....

39. ඔබගේ නිවසේ ආහාර සැකසීම සඳහා සාමාන්‍යයෙන් මසකට කොපමණ ප්‍රමානයක් ලුනු භාවිතා කරනවා ද?

500g ට අඩු ☐ 500g සිට 1kg දක්වා ☐ 1kg සිට 2kg දක්වා ☐ 2kg ට වැඩි ☐

40. ඔබ මේ අතරින් භාවිතා කරන්නේ කුමන ලුනු වර්ගය/වර්ග ද?

අයඩින් සහිත පැකට් නොකල ලුනු ☐ අයඩින් සහිත පැකට් කල ලුනු ☐
අයඩින් රහිත පැකට් නොකල ලුනු ☐ අයඩින් රහිත පැකට් කල ලුනු ☐

41. ඔබගේ ආහාර වේලට එළවළු

වැඩියෙන් භාවිතා කරයි ☐ සාමාන්‍ය ලෙස භාවිතා කරයි ☐ අඩුවෙන් භාවිතා කරයි ☐

42. ඔබගේ ආහාර වේලට පළතුරු

වැඩියෙන් භාවිතා කරයි ☐ සාමාන්‍ය ලෙස භාවිතා කරයි ☐ අඩුවෙන් භාවිතා කරයි ☐

43. ඔබ සතියට ව්‍යායාම වල නිරත වන දින ගනන

නිරත නොවේ ☐ 1 ත් 2ත් අතර ☐ 3 ත් 5ත් අතර ☐ 5ට වැඩි ☐

44. ඔබ ඔබගේ සෞඛ්‍ය තත්වය ගැන පහදන්නේ කෙසේ ද?

ඉතා දුර්වලයි ☐ දුර්වලයි ☐ සාමාන්‍යයි ☐ හොඳයි ☐ ඉතා හොඳයි ☐

45. ඔබ ජීවිතය සතුටින් ගතකරනවා ☐ සාමාන්‍ය ලෙස ගතකරනවා ☐ දුකින් ගතකරනවා ☐

46. ඔබගේ ශාරීරික අපහසුතා නිසා වැඩකටයුතු කරගැනීමේ අපහසුවක් නැත ☐ තරමක් ඇත ☐ දැඩිව ඇත ☐

47. ඔබගේ ආදායම ඔබගේ වියදම් පිරිමහ ගැනීම සඳහා

ඉතා ප්‍රමාණවත් ☐ සාමාන්‍යයෙන් ප්‍රමාණවත් ☐ ප්‍රමාණවත් නැත ☐

48. ඔබට ලැබෙන ප්‍රතිකාර පිළිබඳව, සෑහීමකට පත්වේ ☐ සෑහීමකට පත්නොවේ ☐

49. ඔබගේ නිදහස බුක්ති විඳීමට ඔබගෙ සෞඛ්‍ය ප්‍රමාණවත්, ප්‍රමාණවත් වේ ☐ ප්‍රමාණවත් නොවේ ☐

50. අවශ්‍ය වූ විටෙක අසල් වැසියන් පවුලේ සාමාජිකයන් හිත මිතුරන් ඔබට, සහය වනු ඇත ☐ සහය නොවනු ඇත ☐

51. ඔබට ආදරය සහ කරුණාව දක්වන්න අයෙකු, සිටී ☐ නොසිටියි ☐

52. ඔබ, වාසනාවන්තයි ☐ අවාසනාවන්තයි ☐

53. ඔබට අවශ්‍ය දෑ ලබා ගැනීමට තනිවම වියදම්, දැරිය හැකියි ☐ දැරිය නොහැකියි ☐

54. වගකීම් හා යුතුකම් නිසා ඔබගේ නිදහස, සීමා වී ඇත ☐ සීමා වී නැත ☐

55. ආගමික කටයුතු සහ විශ්වාස ඔබගේ ජීවිතය සාර්ථක කරගැනීමට, ඉවහල් වේ ☐ ඉවහල් නොවේ ☐

56. පහත වගුව පුරවන්න

රෝගය	කොපමණ කලක සිට ද (අවුරුදු)	බටහිර ප්‍රතිකාර ගනී	ආයුර්වේද ප්‍රතිකාර ගනී	වෙනත් ප්‍රතිකාර ගනී	ප්‍රතිකාර කිසිවක් නොගනී	නොකඩවා ප්‍රතිකාර ගනී	මවට පියාට හෝ සහෝදර සහෝදරියන්ට තිබේ
දියවැඩියාව							
අධිරුධිර පීඩනය							
කොලෙස්ට්‍රෝල්							
හෘදයාබාධ							
අංශභාගය							
වෙනත්							

57. **ඔබ මාසිකව ඖෂධ ලබා ගන්නේ,** රෝහලෙන් ☐ රාජ්‍ය ඔසුසල ☐ පෞද්ගලික ෆාමසියෙන් ☐

58. **පහත අවස්ථාවන්හිදී ඔබට පපුවේ වේදනාව හෝ මහන්සිය ඇතිවේ ද?**

කන්දක් හෝ පඩිපෙලක් නැගීමේදී ☐
වේගයෙන් තැනිතලා පොළොවක ඇවිදීමේදී ☐
සාමාන්‍ය වේගයෙන් තැනිතලා පොළොවක ඇවිදීමේදී ☐
විවේකීව සිටිය දී ☐
කිසිසේත් ඇති නොවේ ☐

59. **එවැනි පපුවේ වේදනාවක් හෝ මහන්සියන් ඇති වූ විට ඔබ**

නවතියි ☐
වේගය අඩාල කරයි ☐
ක්‍රියාවේ වෙනසක් සිදු නොකරයි ☐

60. **එවැනි අවස්ථාවක ඔබ නතරවූ විට හෝ වේගය අඩාල කල විට වේදනාව හෝ මහන්සිය**

ක්ෂණිකව සම්පූර්ණයෙන්ම නැතිවී යයි ☐
විනාඩි දහයක් පමණ ඇතුලත නැතිවී යයි ☐
විනාඩි දහයකට වැඩිය පවතී ☐

61. **ඔබ පහත සඳහන් ප්‍රතිකාර ලබා ගත්/ලබා ගන්නා අයෙක් ද?**

ඖෂධ පමණක් ☐
ඇන්ජියොග්‍රෑම් + ස්ටෙන්ට් ☐
බයිපාස් සැත්කම ☐

62. **ඔබ හෘදයාබාධ වැළඳී කොපමණ කලක් ද ?**

දින කිහිපයක් ☐ මාස කිහිපයක් ☐ අවුරුදු 1ත් 2ත් අතර කාලයක් ☐ ඇවුරුදු 5ක් පමණ ☐

අවුරුදු 10ක් පමණ ☐ අවුරුදු 15ක් පමණ ☐ අවුරුදු 15ට වැඩි කාලයක සිටය ☐