DATA CLUSTERING USING MAX-MAX ROUGHNESS AND ITS APPLICATION TO CLUSTER PATIENTS SUSPECTED HEART DISEASE

MOHD AMIROL REDZUAN BIN MAT ROFI

UNIVERSITI MALAYSIA PAHANG

BORANGPENGESAHANSTATUSTESIS

JUDUL: Data Clustering Using Max-Max Roughness And Its Application to Cluster Patients Suspected Heart Disease

SESIPENGAJIAN: 2011/2012

Saya: MOHD AMIROL REDZUAN BIN MAT ROFI
(HURUFBESAR)

Mengaku membenarkan tesis (Projek Sarjana Muda /Sarjana /Doktor Falsafah)* ini disimpan di Perpustakaan Universiti Malaysia Pahang dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hak milik Universiti Malaysia Pahang
2. Perpustakaan Universiti Malaysia Pahang dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. **Sila tandakan (4)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/ badan di mana penyelidikan dijalankan)

☐ TIDAK TERHAD

Disahkan oleh

(TANDATANGAN PENULIS)	(TANDATANGAN PENYELIA)
Alamat Tetap:	Nama penyelia:
NO. 22, LORONG BALOK BARU 27,	
PAKR BALOK, 26100,	
KUANTAN	
Tarikh: 11 JUNE 2012	Tarikh:

CATATAN: *Potong yang tidak berkenaan.
**Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/ organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT atau TERHAD.
*** Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana secara penyelidikan, atau diserta sibagi pengajian secara

DATA CLUSTERING USING MAX-MAX ROUGHNESS AND ITS APPLICATION TO CLUSTER PATIENTS SUSPECTED HEART DISEASE

MOHD AMIROL REDZUAN BIN MAT ROFI

A thesis submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Computer Science (Software Engineering)

FACULTY OF COMPUTER SYSTEM & SOFTWARE ENGINEERING

UNIVERSITI MALAYSIA PAHANG

JUNE 2012

SUPERVISOR'S DECLARATION

"I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of the Degree in Computer Science (Software Engineering)"

Signature : …………………………………………..

Supervisor : TUTUT HERAWAN

Date :

STUDENT'S DECLARATION

I hereby declare that this thesis entitled *"Data Clustering Using Max-Max Roughness And Its Application To Cluster Patients Suspected Heart Disease"* is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : …………………………………………..

Name : MOHD AMIROL REDZUAN BIN MAT ROFI

Date :

DEDICATION

To my gracious God, Allah SWT

To my beloved father and mother and also my lovely siblings for always being there for me through my days and nights.

To my very concern lecturer,Dr. Tutut Herawan, thank you for your guidance, advices and support in completing the task.

My supportive friends. Thanks for your support and cooperation.

"May Allah bless yours"

Sincerely

Mohd Amirol Redzuan bin Mat Rofi

ACKNOWLEDGMENT

Alhamdulillah, I am very grateful and thankfulness to Allah SWT for giving me strength, patience and ability to complete this thesis project. Without His help, this thesis would not be possible for me to complete.

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Tutut Herawan for his germinal ideas, priceless guidance, continuous encouragement and constant support in making this research possible. All his teaching and efforts will always be remembered and appreciated.

To my beloved family, especially to my father, Mat Rofi bin Mat Adam and my mother, Jamisah binti Jusoh for giving me their blessings, unconditional love, financial support and also understand on my conditions that I can't visit them many often.

Last but not least, my sincere gratitude extends to all individuals who have contributed either directly or indirectly in ensuring the success of this thesis.

ABSTRACT

Nowadays, there are many technique to clustering large-scale data. One of the technique to clustering data is using the Rough Set Theory.The objective of this paper is to present the process of Data Clustering Using Maximum-Maximum Roughness and its application to cluster patients suspected heart disease. It is based on clustering techniques based on rough set theory name Max-Max Roughness to describes and employed regarding to solve a classification problem of heart disease patients.

ABSTRAK

Dewasa ini, terdapat banyak cara yang digunakan untuk mengklusterkan data yang bersaiz besar. Salah satu caranya adalah dengan menggunakan Teori Rough Set. Objektif bagi projek ini adalah untuk memberi pendedahan tentang proses Mengklusterkan Data Menggunakan Teknik Maximum-Maximum Roughness dan Applikasi Teknik ini terhadap Pesakit yang Disyaki Menghidap Penyakit Jantung. Teknik ini berdasarkan teknik mengkelaskan yang diambil daripada Teori Rough Set bernama Max-Max Roughness yang digunakan untuk menyelesaikan masalah mengkelaskan pesakit yang menghidap penyakit jantung.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

This chapter briefly discuss on the overview of this research. It contains five parts. The first part is background; follow by the problem statement. After that, are the motivation followed by the scopes. Next are the objectives where the project research's goal is determined. Lastly is the thesis organization which briefly describes the structure of this thesis.

1.1 BACKGROUND

The most well-known branch of data mining is knowledge discovery, also known as Knowledge Discovery from Databases (KDD). Just as many other forms of knowledge discovery it creates abstractions of the input data. The knowledge obtained through the process may become additional data that can be used for further usage and discovery. [1]

KDD algorithms can be classified into three general areas: classificatory, association, and sequencing. Classificatory algorithms partition input data into disjoint groups. The results of such classification might be represented as a decision tree or a set of characteristic rules as from ID3 or KID3. Association algorithms find, from transaction records, sets of items that appear together in sufficient frequency to merit attention. Sequencing algorithms find items or events that are related in time, such as events A and B usually being followed by C. [2]

KDD exhibits four main characteristics. The first one is high-level language. Discovered knowledge is represented in a high-level language. It need not be directly used by humans, but its expression should be understandable by human users. The second one is accuracy. Discoveries accurately portray the contents of the database. The extent to which this portrayal is imperfect is expressed by measures of certainty. The next one is interesting results. Discovered knowledge is interesting according to user-defined biases. In particular, being interesting implies that patterns are novel and potentially useful, and the discovery process is nontrivial. The last one is efficiency. The discovery process is efficient. Running times for large-sized databases are predictable and acceptable. [3]

Data mining is a step in the KDD process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the data. [4] The term data mining has been mostly used by statisticians, data analysts, and the management information systems (MIS) communities. It has also gained popularity in the database field. The earliest uses of the term come from statistics and the usage in most cases was associated with negative connotations of blind exploration of data without a priori hypotheses to verify. [5]

Two common data mining techniques for finding hidden patterns in data are clustering and classification analyses. Although classification and clustering are often mentioned in the same breath, they are different analytical approaches. Classification is a different technique than clustering. Classification is similar to clustering in that it also segments customer records into distinct segments called classes. But unlike clustering, a classification analysis requires that the end-user/analyst know ahead of time how classes are defined. For example, classes can be defined to represent the likelihood that a customer defaults on a loan (Yes/No). It is necessary that each record in the dataset used to build the classifier already have a value for the attribute used to define classes. Because each record has a value for the attribute used to define the classes, and because the end-user decides on the attribute to use, classification is much less exploratory than clustering. The objective of a classifier is not to explore the data to discover interesting segments, but rather to decide how new records should be classified. Clustering is an automated process to group related records together. Related records are grouped together on the basis of having similar values for attributes. This approach of segmenting the database via clustering analysis is often used as an exploratory technique because it is not necessary for the end-user/analyst to specify ahead of time how records should be related together. In fact, the objective of the analysis is often to discover segments or clusters, and then examine the attributes and values that define the clusters or segments. As such, interesting and surprising ways of grouping customers together can become apparent, and this in turn can be used to drive marketing and promotion strategies to target specific types of customers. [6] Heart disease is an umbrella term for a number of different diseases which affect the heart such as arrhythmia, myocardial ischemia, and myocardial infarction. It is also one of the leading causes of death in the world. [7] Because there are many possible conditions that follow under the umbrella of heart disease, the related symptoms are numerous. [8] Few symptoms are more alarming than chest pain. In the minds of many people, chest pain equals heart pain. And while many other conditions can cause chest pain, cardiac disease is so common - and so dangerous - that the symptom of chest pain should never be dismissed out of hand as being insignificant. "Chest pain" is an imprecise term. It is often used to describe any pain, pressure, squeezing, choking, numbness or any other discomfort in the chest, neck, or upper abdomen, and is often associated with pain in the jaw, head, or arms. It can last from less than a second to days or weeks, can occur frequently or rarely, and can occur sporadically or predictably. This description of chest pain is obviously very vague, and as you might expect, many medical conditions aside from heart disease can produce symptoms like this.

Palpitations, an unusual awareness of the heartbeat, are an extremely common symptom. Most people who complain of palpitations describe them either as "skips" in the heartbeat (that is, a pause, often followed by a particularly strong beat,) or as periods of rapid and/or irregular heartbeats. Most people with palpitations have some type of cardiac arrhythmia -- abnormal heart rhythms. There are many types of arrhythmias, and almost all can cause palpitations, but the most common causes of palpitations are premature atrial complexes (PACs), premature ventricular complexes (PVCs), episodes of atrial fibrillation, and episodes of supraventricular tachycardia (SVT). Unfortunately, on occasion, palpitations can signal a more dangerous heart arrhythmia, such as ventricular tachycardia.

Episodes of light-headedness or dizziness can have many causes, including anaemia (low blood count) and other blood disorders, dehydration, viral illnesses, prolonged bed rest, diabetes, thyroid disease, gastrointestinal disturbances, liver disease, kidney disease, vascular disease, neurological disorders, dysautonomias, vasovagal episodes, heart failure and cardiac arrhythmias. Because so many different conditions can produce these symptoms, anybody experiencing episodes of light-headedness or dizziness ought to have a thorough and complete examination by a physician. And since disorders of so many organ systems can cause these symptoms, a good general internist or family doctor may be the best place to start.

Syncope is a sudden and temporary loss of consciousness, or fainting. It is a common symptom - most people pass out at least once in their lives - and often does not indicate a serious medical problem. However, sometimes syncope indicates a dangerous or even life-threatening condition, so when syncope occurs it is important to figure out the cause. The causes of syncope can be grouped into four major categories: neurologic, metabolic, vasomotor and cardiac. Of these, only cardiac syncope commonly leads to sudden death.

Fatigue, lethargy or somnolence (daytime sleepiness) is very common symptoms. Fatigue or lethargy can be thought of as an inability to continue functioning at one's normal levels. Somnolence implies, in addition, that one either craves sleep - or worse, finds oneself suddenly asleep, a condition known as narcolepsy - during the daytime. While fatigue and lethargy can be symptoms of heart disease (particularly, of heart failure), these common and non-specific symptoms can also be due to disorders of virtually any other organ system in the body. Similar to light-headedness and dizziness, individuals with fatigue and lethargy need a good general medical evaluation in order to begin pinning down a specific cause. Somnolence is often caused by nocturnal sleep disorders such as sleep apnea, restless leg syndrome or insomnia. All these sleep disturbances, however, are more common in patients with heart disease.

Shortness of breath is most often a symptom of cardiac or pulmonary (lung) disorders. Heart failure and coronary artery disease frequently produce shortness of breath. Patients with heart failure commonly experience shortness of breath with exertion, or when lying flat on their backs. They also can suddenly wake up at night gasping for breath, a condition known as paroxysmal nocturnal dyspnea. Other cardiac conditions such as valvular heart disease or pericardial disease can produce this symptom, as can cardiac arrhythmias. Numerous lung conditions can produce shortness of breath including asthma, emphysema, bronchitis, pneumonia, or pleural effusion (a fluid accumulation between the lung and chest wall). Shortness of breath is almost always a sign of a significant medical problem, and should always be evaluated by a doctor.

According to the research of Jonathan R. Carapetis, it is estimated that there were a minimum of 15.6 million people in the world with rheumatic heart disease, with 282 000 new cases each year and 233 000 resultant deaths each year; however, we also noted that the estimates of the number of cases in school-aged children in China (176 500) and Asia Other (102 000; Asia excluding South-Central Asia and China) were based on very few studies, none of which used echocardiography to confirm the presence of rheumatic heart disease lesions. Moreover, 5 of the 6 studies included in the Asia Other estimate came from 1 country, the Philippines. [9]

1.2 PROBLEM STATEMENT

In this research project, we need to figure out the suitable technique of clustering set to be use and how to apply it in the grouping of patients with heart disease. We also need to see the data clustering that is suitable to this research project.

Many techniques have been introduced to make grouping or clustering data attributes. For example, fuzzy set, soft set and rough set. In this research project, the technique that will be implementing is the rough set. The rough set is the most suitable type of clustering technique because the technique can deal with the multi-valued data which is required by this research.

1.3 SCOPES

The scopes for this research:

a. The clustering uses max-max roughness technique.
b. Clustering patients with heart disease.

1.4 OBJECTIVES

The objectives for this research:

a. To clustering the patients with heart disease using the techniques of rough set.
b. To apply the rough set clustering technique into a real life case.

1.5 THESIS ORGANIZATION

The rest of this paper is organized as follows. Section 2 describes the notion of information system (databases). Section 3 describes the theory of rough set.

Section 4 describes the dataset, modeling process and rough set-based decision making using maximal supported objects by parameters. Section 5 describes the results from an application of rough set theory for decision making and grouping patients suspected Influenza-Like Illness (ILI) following by discussion. Finally, the conclusion of this work is described in section 6.

CHAPTER 2

LITERATURE REVIEW

This chapter briefly discusses about the literature review of this research using the maximum-maximum roughness technique. There are seven main sections in this chapter. The first main section is introduction of this chapter. Then, the next main section describes the concept. After that, the manual system of the project will be discussed. Next, there are two main sections which discuss several technologies and techniques separately. The next main section discusses the existing system while the last main section reviews the methodologies used to develop game.

2.1 HEART DISEASE

This section firstly presents a description and symptoms of heart disease. Further, information of heart disease in the world, Asia and Malaysia also presented.

2.1.1 Heart Disease Descriptions

The heart is the organ that pumps blood, with its life-giving oxygen and nutrients, to all tissues of the body. If the pumping action of the heart becomes inefficient, vital organs like the brain and kidneys suffer. And if the heart stops working altogether, death occurs within minutes. Life itself is completely dependent on the efficient operation of the heart. [10]

There are many kinds of heart disease, and they can affect the heart in several ways. But the ultimate problem with all varieties of heart disease is that, in one way or another, they can disrupt the vital pumping action of the heart.

2.1.2 Heart Disease Symptoms

Because there are many possible conditions that follow under the umbrella of heart disease, the related symptoms are numerous. [11]

Few symptoms are more alarming than chest pain. In the minds of many people, chest pain equals heart pain. And while many other conditions can cause chest pain, cardiac disease is so common - and so dangerous - that the symptom of chest pain should never be dismissed out of hand as being insignificant. "Chest pain" is an imprecise term. It is often used to describe any pain, pressure, squeezing, choking, numbness or any other discomfort in the chest, neck, or upper abdomen, and is often associated with pain in the jaw, head, or arms. It can last from less than a second to days or weeks, can occur frequently or rarely, and can occur sporadically or predictably. This description of chest pain is obviously very vague, and as you might expect, many medical conditions aside from heart disease can produce symptoms like this.

Palpitations, an unusual awareness of the heartbeat, are an extremely common symptom. Most people who complain of palpitations describe them either as "skips" in the heartbeat (that is, a pause, often followed by a particularly strong beat,) or as periods of rapid and/or irregular heartbeats. Most people with palpitations have some type of cardiac arrhythmia -- abnormal heart rhythms. There are many types of arrhythmias, and almost all can cause palpitations, but the most common causes of palpitations are premature atrial complexes (PACs), premature ventricular complexes (PVCs), episodes of atrial fibrillation, and episodes of supraventricular tachycardia (SVT). Unfortunately, on occasion, palpitations can signal a more dangerous heart arrhythmia, such as ventricular tachycardia.

Episodes of light-headedness or dizziness can have many causes, including anaemia (low blood count) and other blood disorders, dehydration, viral illnesses, prolonged bed rest, diabetes, thyroid disease, gastrointestinal disturbances, liver disease, kidney disease, vascular disease, neurological disorders, dysautonomias, vasovagal episodes, heart failure and cardiac arrhythmias. Because so many different conditions can produce these symptoms, anybody experiencing episodes of light-headedness or dizziness ought to have a thorough and complete examination by a physician. And since disorders of so many organ systems can cause these symptoms, a good general internist or family doctor may be the best place to start.

Syncope is a sudden and temporary loss of consciousness, or fainting. It is a common symptom - most people pass out at least once in their lives - and often does not indicate a serious medical problem. However, sometimes syncope indicates a dangerous or even life-threatening condition, so when syncope occurs it is important to figure out the cause. The causes of syncope can be grouped into four major categories: neurologic, metabolic, vasomotor and cardiac. Of these, only cardiac syncope commonly leads to sudden death.

Fatigue, lethargy or somnolence (daytime sleepiness) is very common symptoms. Fatigue or lethargy can be thought of as an inability to continue functioning at one's normal levels. Somnolence implies, in addition, that one either craves sleep - or worse, finds oneself suddenly asleep, a condition known as narcolepsy - during the daytime. While fatigue and lethargy can be symptoms of heart disease (particularly, of heart failure), these common and non-specific symptoms can also be due to disorders of virtually any other organ system in the body. Similar to light-headedness and dizziness, individuals with fatigue and lethargy need a good general medical evaluation in order to begin pinning down a specific cause. Somnolence is often caused by nocturnal sleep disorders such as sleep apnea, restless leg syndrome or insomnia. All these sleep disturbances, however, are more common in patients with heart disease.

Shortness of breath is most often a symptom of cardiac or pulmonary (lung) disorders. Heart failure and coronary artery disease frequently produce shortness of breath. Patients with heart failure commonly experience shortness of breath with exertion, or when lying flat on their backs. They also can suddenly wake up at night gasping for breath, a condition known as paroxysmal nocturnal dyspnea. Other cardiac conditions such as valvular heart disease or pericardial disease can produce this symptom, as can cardiac arrhythmias. Numerous lung conditions can produce shortness of breath including asthma, emphysema, bronchitis, pneumonia, or pleural effusion (a fluid accumulation between the lung and chest wall). Shortness of breath is almost always a sign of a significant medical problem, and should always be evaluated by a doctor.

2.1.3 Heart Disease in the World

If you said cardiovascular (CV) disease these days, most people would have a look of fear on their faces. CV diseases are those of the heart and blood vessel system, and thoughts of coronary heart disease, a heart attack, high blood pressure, stroke, angina (chest pain), or rheumatic heart disease would send people scurrying to medical professionals looking for a cure. And indeed it should, as according to the WHO, CV diseases are ranked as the number one killers in the world claiming an estimated 17 million lives annually. [12] It is estimated that every one in three people around the world dies due to stroke or heart attack. WHO estimates that if no action is taken to improve CV health and current trends such as changes of lifestyle, lack of exercise, stress, and smoking continue, 25 percent more of healthy life years are likely to be lost to CV disease globally by 2020.

According to the research of Jonathan R. Carapetis, it is estimated that there were a minimum of 15.6 million people in the world with rheumatic heart disease, with 282 000 new cases each year and 233 000 resultant deaths each year. [13]

2.1.4 Heart Disease in Asia

We also noted that the estimates of the number of cases in school-aged children in China (176 500) and Asia Other (102 000; Asia excluding South-Central Asia and China) were based on very few studies, none of which used echocardiography to confirm the presence of rheumatic heart disease lesions. Moreover, 5 of the 6 studies included in the Asia Other estimate came from 1 country, the Philippines [14].

Many people go through life not knowing that they may be susceptible to CVD. Lifestyle changes, lack of exercise, stress, and smoking are responsible for the increase of risk factors leading to the development of CV diseases such as hypertension, hyperlipidemia, diabetes, and obesity. For example, in Singapore, the national health survey conducted in 2001 found that approximately 14.1 percent of its citizens aged 65 and above had high blood cholesterol levels, 32.6 percent were hypertensive, and 64.4 percent of those aged 70 and above had completely sedentary lifestyles. If left unchecked and no steps are taken to rectify the spiralling conditions, it could lead to heart attacks or strokes. However, most often, people do not realize that they may get CV diseases because the risk factors and symptoms associated with such diseases are also associated with ageing and other diseases. [15]

The standard course of therapy for the treatment of CV diseases and its related diseases such as hypertension, diabetes, and hyperlipidemia include classes of medications such as diuretics, ace inhibitors, angiotension II receptor blockers, beta blockers, alpha blockers, calcium channel blockers, vasodilators, statins, and antiplatelets. These medications are usually used in combinations with each other to get the optimal results.

2.1.5 Heart Disease in Malaysia

While in the 1960s and 1970s, deaths due to CVD were levelling off in the United States, Asia was experiencing a different scenario altogether. Where in the early 1960s and 1970s, communicable diseases posed the greatest threat in this region, especially for Malaysia, the late 1990s and early 2000 saw an increasing trend of lifestyle killer diseases such as heart disease, cancer, and stroke leading the way. The rapid pace of development has led to the changing pattern of diseases as countries underwent economic development. [16]

The number of CV disease cases in Malaysia has increased to 14 percent in five years from 96,000 in 1995 to 110,000 in 2000. It is the leading cause of death in the country claiming a third of all its patients. In 2001, approximately 20 percent of all deaths at the Ministry of Heath hospitals were due to heart attacks and strokes. Two thirds of these deaths were due to heart diseases and the rest to strokes. In fact, it is estimated that 40,000 new stroke cases are recorded annually in Malaysia. Figure 1 depicts the major causes of death in Malaysia in 2001.

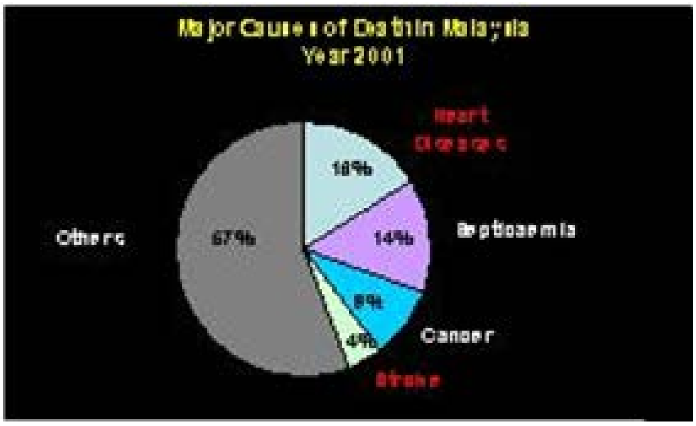


Figure 2.1: Major causes of Death in Malaysia

However, there are certain patients whose conditions do not respond to medications. In the last decade or so, in line with Malaysia's drive to become leader in healthcare, both countries have initiated and made technological breakthrough in other methods of treating CV diseases, namely in treating heart failure patients. Heart transplants are now considered a treatment option after initial immunological obstacles. Drug-coated stents, which revolutionized the cardiology world in the early 1990s, are being increasingly used in angioplasty surgeries. The National Heart Institute, the premier cardiology centre in Malaysia, has used almost half of the 1,050 Cypher stents since its launch in May 2002. Singapore has also introduced cutting edge strategies such as implantable mechanical assist devices, ventricular reduction and remodeling surgery, and cardiac resynchronization therapy.

The costs of these treatments do not come cheap. In Singapore, the cost of treating cardiovascular diseases is estimated at about $64 million (Singapore $110 million) a year. The Malaysian Ministry of Health spends approximately $2.6 million (RM10 million) annually just on the use of statins in primary prevention of atherosclerosis. Additionally, the drug-coated stent, Cypher costs $2,632 (RM10,000) per stent. Patients with CV diseases are usually on life long treatment and may not be able to afford such high costs of medications. Thus, most often than not, unless it is a life and death situation, physicians usually leave the choice of medications to patients while providing all the pros and cons of each. For example, between taking an aspirin or Plavix, which are anti-platelets for the prevention of stroke, according to most physicians in Malaysia, patients opt to take aspirin as it is 10 to 20 times cheaper than Plavix. [17]

2.2 KNOWLEDGE DISCOVERY IN DATABASES

This section firstly presents definitions of Knowledge Discovery in Databases (KDD). Further, information of KDD processes and definitions related to KDD processes. Finally, the last sub-section presents the applications of KDD in computer science field.

2.2.1 Definitions of KDD

The most well-known branch of data mining is knowledge discovery, also known as Knowledge Discovery from Databases (KDD). Just as many other forms of knowledge discovery it creates abstractions of the input data. The knowledge obtained through the process may become additional data that can be used for further usage and discovery. [18]

KDD algorithms can be classified into three general areas: classificatory, association, and sequencing. Classificatory algorithms partition input data into disjoint groups. The results of such classification might be represented as a decision tree or a set of characteristic rules as from ID3 or KID3. Association algorithms find, from transaction records, sets of items that appear together in sufficient frequency to merit attention. Sequencing algorithms find items or events that are related in time, such as events A and B usually being followed by C.

KDD exhibits four main characteristics. The first one is high-level language. Discovered knowledge is represented in a high-level language. It need not be directly used by humans, but its expression should be understandable by human users. The second one is accuracy. Discoveries accurately portray the contents of the database. The extent to which this portrayal is imperfect is expressed by measures of certainty. The next one is interesting results. Discovered knowledge is interesting according to user-defined biases. In particular, being interesting implies that patterns are novel and potentially useful, and the discovery process is nontrivial. The last one is efficiency. The discovery process is efficient. Running times for large-sized databases are predictable and acceptable.

2.2.2 KDD Processes

The term Knowledge Discovery in Databases, or KDD for short, refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods. It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems, and data visualization.

The unifying goal of the KDD process is to extract knowledge from data in the context of large databases. It does this by using data mining methods (algorithms) to extract (identify) what is deemed knowledge, according to the specifications of measures and thresholds, using a database along with any required pre-processing, sub-sampling, and transformations of that database.

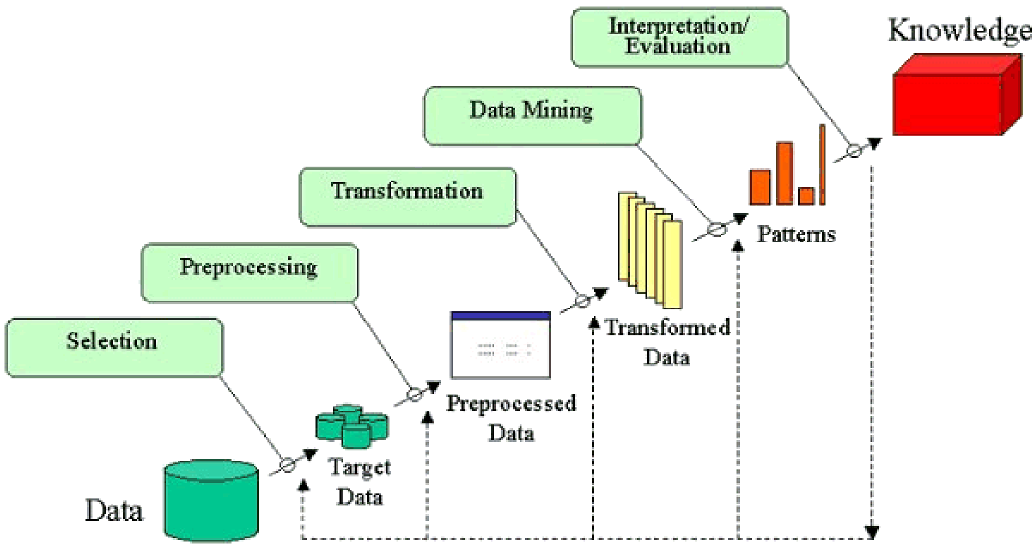


Figure 2.2: An Outline of the Steps of the KDD Process

The overall process of finding and interpreting patterns from data involves the repeated application of the following steps. First, developing an understanding of the application domain, the relevant prior knowledge and the goals of the end-user. Second, creating a target data set by selecting a data set or focusing on a subset of variables, or data samples, on which discovery is to be performed. Third, data cleaning and pre-processing involving the removal of noise or outliers, collecting necessary information to model or account for noise, strategies for handling missing data fields and accounting for time sequence information and known changes. The forth step is about data reduction and projection. It is to find useful features to represent the data depending on the goal of the task and using dimensionality reduction or transformation methods to reduce the effective number of variables under consideration or to find invariant representations for the data. Fifth step is choosing the data mining task by deciding whether the goal of the KDD process is classification, regression, clustering, etc. Next, choosing the data mining algorithm(s) by selecting method(s) to be used for searching for patterns in the data, deciding which models and parameters may be appropriate and matching a particular data mining method with the overall criteria of the KDD process. Seventh, data mining. Searching for patterns of interest in a particular representational form or a set of such representations as classification rules or trees, regression, clustering, and so forth. Then, interpreting mined patterns and also consolidating discovered knowledge.

The terms knowledge discovery and data mining are distinct. KDD refers to the overall process of discovering useful knowledge from data. It involves the evaluation and possibly interpretation of the patterns to make the decision of what qualifies as knowledge. It also includes the choice of encoding schemes, pre-processing, sampling, and projections of the data prior to the data mining step.

Data mining refers to the application of algorithms for extracting patterns from data without the additional steps of the KDD process.

2.2.3 Definitions related to the KDD Processes

Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

Table 2.1: Terms related to KDD process

Terms	**Descriptions**
Data	A set of facts, F.
Pattern	An expression E in a language L describing facts in a subset F_E of F.
Process	KDD is a *multi-step process* involving data preparation, pattern searching, knowledge evaluation, and refinement with iteration after modification.
Valid	Discovered patterns should be true on new data with some degree of certainty. Generalize to the future (other data).
Novel	Patterns must be novel (should not be previously known).
Useful	Actionable; patterns should potentially lead to some useful actions.
Understandable	The process should lead to human insight. Patterns must be made understandable in order to facilitate a better understanding of the underlying data.

Interestingness is an overall measure of pattern value, combining validity, novelty, usefulness, and simplicity.

2.2.1 Application of KDD in Computer Science Fields

Online data continues to grow at an explosive pace, due to the Internet and the widespread use of database technology. This phenomenon has created an immense opportunity and need for methodologies of Knowledge Discovery and Data Mining (KDD). KDD is an interdisciplinary area focusing upon building automated techniques for extracting useful knowledge from data. Research in this area draws principally upon methods from statistics, data management, pattern recognition, and machine learning, to deliver advanced techniques for business intelligence. IBM Research has been at the forefront of this exciting new area from the very beginning. Key advances in robust and scalable data mining techniques, methods for fast pattern detection from very large databases, text and Web mining, as well as innovative business intelligence applications, have come from our worldwide research laboratories. [19]

An area of particular focus in our KDD research has been on high performance, scalable data mining techniques for large-scale databases and data repositories. IBM's early lead in this area was established by our invention of association rule and sequential patterns technology for efficiently detecting patterns in large-scale databases. These and other technologies for scalable and parallel data mining developed as part of this project provided the original basis and impetus for IBM's flagship data mining products. This theme continues in recent research activities that include automatic subspace clustering, discovery-driven exploration of OLAP data cubes, and fast techniques for pre-computing and maintaining OLAP data.

Another area of investigation has been focusing on predictive data mining algorithms, systems, and solutions. One long-term effort has centred on rule-based predictive modelling and its integration into data mining frameworks. A recent effort has resulted in new data mining middleware for rule-based probabilistic estimation, which combines machine learning with principles from statistical learning theory and data management for scalable predictive modelling of massive data sets. This technology has been embedded in innovative business intelligence applications for areas such as insurance risk management and retail targeted marketing. Related research continues in areas such as rare-event predictive mining, robust feature selection, ensemble-based and regularization methods for predictive modelling, and support vector machines.

The exploration of machine learning and statistical techniques for new KDD methods and solutions continues to grow across all our research laboratories. These include research activities in text categorization, information extraction from document collections and the Web, item recommendation and personalization, and event mining for systems and network management, and business process insight discovery. We continue to emphasize research in KDD techniques for handling massive amounts of data, resulting in new approaches for clustering, predictive modelling, and frequent pattern detection, as well as for integration of KDD methodologies into database middleware. Many of the data mining techniques that were originally developed and tuned for structured data (e.g. associations, classification, clustering) are also now being increasingly used and refined for projects in natural language processing and knowledge management.

A rapidly growing area of KDD research is in applications to Internet, mobile, and pervasive computing solutions. Some of our projects are currently developing advanced business intelligence components for B2C personalization and B2B commerce. We are also beginning to investigate uses for KDD in applications for e-marketplaces, mobile commerce, as well as information mining from multimedia on the Web.

2.3 DATA CLUSTERING

This section firstly presents a description and symptoms of diabetics. Further, information of diabetics in the world, Asia and Malaysia also presented. Finally, the last sub-section presents information of patient having pre-diabetes.

2.3.1 Definitions of Data Clustering

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorial, and differences in assumptions and contexts in different communities has made the transfer of useful generic concepts and methodologies slow to occur.

Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, document retrieval, image segmentation, and pattern classification. However, in many such problems, there is little prior information (e.g., statistical models) available about the data, and the decision-maker must make as few assumptions about the data as possible. It is under these restrictions that clustering methodology is particularly appropriate for the exploration of interrelationships among the data points to make an assessment (perhaps preliminary) of their structure.

The term “clustering” is used in several research communities to describe methods for grouping of unlabeled data. These communities have different terminologies and assumptions for the components of the clustering process and the context in which clustering is used.

2.3.2 Classification vs. Clustering

Classification is defined as the task to learn to assign instances to predefined classes while clustering is defined as the task to learn a classification from the data. This means that, in classification, no predefined classification is required. [20] Both of them are the central concept of pattern recognition, and both of them are used as important knowledge discovery tools in modern machine learning process. Classification involves assigning input data into one or more pre-specified classes based on extraction of significant features or attributes and the processing or analysis of these attributes. Classification requires supervised learning, or also called as learning by example. Supervised learning is a process where the system attempts to find concept descriptions for classes that are together with pre-classified examples. It assumes that the data are labeled or defined. Figure 3 explains the classification where the data or observation (O) is mapped to the classes (C) with a predefined patterns or descriptions (P).

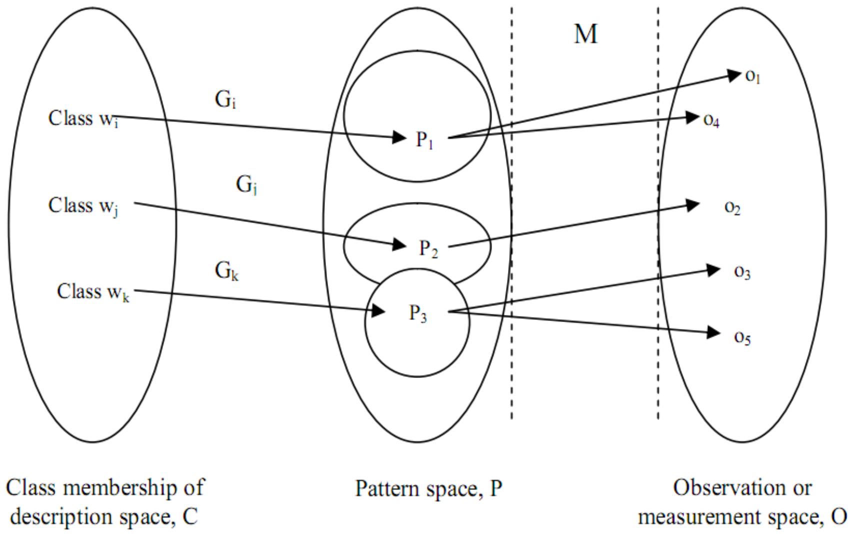


Figure 2.3: Classification [22]

However, in many cases, there are no clear structures in the data where the classes are not defined. There are no obvious number of classes and the relationship between the data's attributes or characteristics are not clear. Besides, when dealing with everyday's data, over time, the characteristics or attributes of the class-specific pattern can change continuously. Therefore, clustering is used to solve this. Clustering involves automatic identification of groups of similar objects or patterns. This is done by maximizing the inter-group similarity and minimizing the intra-group similarity, and as a result, a number of clusters would form on the measurement or observation space. Then the data can be easily recognized and be assigned to the clusters suitable label or feature description. This is shown in the figure below.

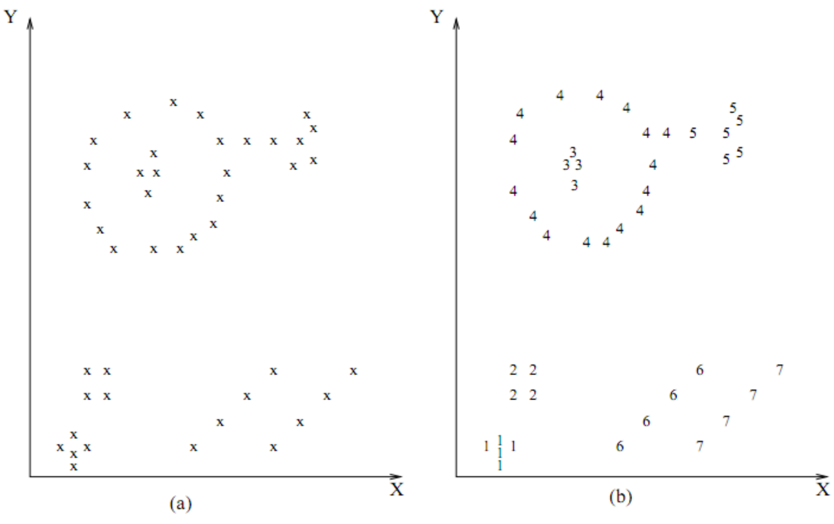


Figure 2.4: Clustering [23]

Clustering requires unsupervised data learning where the task is only directed to search the data for interesting associations, and attempts to group elements by postulating class descriptions for sufficient many classes to cover all items in the data [21, 22].

2.3.3 Clustering Technique

Different approaches to clustering data can be described with the help of the hierarchy (other taxonometric representations of clustering methodology are possible; ours is based on the discussion in Jain and Dubes [1988]). At the top level, there is a distinction between hierarchical and partitional approaches (hierarchical methods produce a nested series of partitions, while partitional methods produce only one).

The taxonomy must be supplemented by a discussion of cross-cutting issues that may (in principle) affect all of the different approaches regardless of their placement in the taxonomy.

Agglomerative *vs.* divisive: This aspect relates to algorithmic structure and operation. An agglomerative approach begins with each pattern in a distinct (singleton) cluster, and successively merges clusters together until a stopping criterion is satisfied. A divisive method begins with all patterns in a single cluster and performs splitting until a stopping criterion is met.

Monothetic *vs.* polythetic: This aspect relates to the sequential or simultaneous use of features in the clustering process. Most algorithms are polythetic; that is, all features enter into the computation of distances between patterns, and decisions are based on those distances. A simple monothetic algorithm reported in Anderberg [1973] considers features sequentially to divide the given collection of patterns. This is illustrated in Figure 8. Here, the collection is divided into two groups using feature $x1$; the vertical broken line V is the separating line.

Each of these clusters is further divided independently using feature $x2$, as depicted by the broken lines $H1$ and $H2$. The major problem with this algorithm is that it generates $2d$ clusters where d is the dimensionality of the patterns. For large values of d (d . 100 is typical in information retrieval applications [Salton 1991]), the number of clusters generated by this algorithm is so large that the data set is divided into uninterestingly small and fragmented clusters.

Hard *vs.* fuzzy: A hard clustering algorithm allocates each pattern to a single cluster during its operation and in its output. A fuzzy clustering method assigns degrees of membership in several clusters to each input pattern. A fuzzy clustering can be converted to a hard clustering by assigning each pattern to the cluster with the largest measure of membership.

Deterministic *vs.* stochastic: This issue is most relevant to partitional approaches designed to optimize a squared error function. This optimization can be accomplished using traditional techniques or through a random search of the state space consisting of all possible labelling.

Incremental *vs.* non-incremental: This issue arises when the pattern set to be clustered is large, and constraints on execution time or memory space affect the architecture of the algorithm. The early history of clustering methodology does not contain many examples of clustering algorithms designed to work with large data sets, but the advent of data mining has fostered the development of clustering algorithms that minimize the number of scans through the pattern set, reduce the number of patterns examined during execution, or reduce the size of data structures used in the algorithm’s operations.

A cogent observation in Jain and Dubes [1988] is that the specification of an algorithm for clustering usually leaves considerable flexibilty in implementation.

2.3.4 Clustering on Numerical Dataset

Clustering analysis consist of polythetic methods that use all attributes, aqqlomerative methods that begin with all cases being singleton cluster and in divisive methods which initially all cases are placed in one cluster. [24] Different numerical attributes were standardized by dividing the attribute values by the corresponding attributes standard deviation. Cluster formation was started by computing a distance matrix between every cluster. By using agglomerative discretization method, new clusters were formed by merging two existing clusters that were the closest to each other. When such a pair was founded (cluster b and), they were fused to form a new cluster d. Definability of Sets (let $B \subseteq A$), a union of some intersections of attributes-value pair blocks, attributes are members of B, will be called a B-locally definable set.

2.3.5 Clustering on Categorical Dataset

In dealing with the raw categorical data, one of the difficulties is to resolve the problem of similarity measure, for the nature of categorical data is the non-numerical so that Euclidean distance extensively-used in numerical data processing cannot be employed directly. However, rough set theory can be applied to clustering analysis; and the clustering data set is mapped as the decision table through introducing a decision attribute.

There has been work in the area of applying rough set theory to the process of selecting a clustering attribute including the works of (Mazlack, He, Zhu and Coppock, 2000).

2.3.6 Application of Clustering Technique

Cluster analysis is a major tool in a number of applications in many fields of business and science. Hereby, we summarize the basic directions in which clustering are used (Theodoridis and Koutroubas, 1999):

• ***Data reduction.***

Cluster analysis can contribute in compression of the information included in data. In several cases, the amount of available data is very large and its processing becomes very demanding. Clustering can be used to partition data set into a number of "interesting" clusters. Then, instead of processing the data set as an entity, we adopt the representatives of the defined clusters in our process. Thus, data compression is achieved.

• ***Hypothesis generation.***

Cluster analysis is used here in order to infer some hypotheses concerning the data. For instance we may find in a retail database that there are two significant groups of customers based on their age and the time of purchases. Then, we may infer some hypotheses for the data, that it, "young people go shopping in the evening", "old people go shopping in the morning".

• ***Hypothesis testing.***

In this case, the cluster analysis is used for the verification of the validity of a specific hypothesis. For example, we consider the following hypothesis:

"*Young people go shopping in the evening*".

One way to verify whether this is true is to apply cluster analysis to a representative set of stores. Suppose that each store is represented by its customer's details (age, job etc) and the time of transactions. If, after applying cluster analysis, a cluster that corresponds to "*young people buy in the evening*" is formed, then the hypothesis is supported by cluster analysis.

• *Prediction based on groups.*

Cluster analysis is applied to the data set and the resulting clusters are characterized by the features of the patterns that belong to these clusters.
Then, unknown patterns can be classified into specified clusters based on their similarity to the clusters' features. Useful knowledge related to our data can be extracted.

Assume, for example, that the cluster analysis is applied to a data set concerning patients infected by the same disease. The result is a number of clusters of patients, according to their reaction to specific drugs. Then for a new patient, we identify the cluster in which he/she can be classified and based on this decision his/her medication can be made.

More specifically, some typical applications of the clustering are in the following fields (Han and Kamber, 2001):

• *Business.*

In business, clustering may help marketers discover significant groups in their customers' database and characterize them based on purchasing patterns.

• ***Biology.***

In biology, it can be used to define taxonomies, categorize genes with similar functionality and gain insights into structures inherent in populations.

• ***Spatial data analysis.***

Due to the huge amounts of spatial data that may be obtained from satellite images, medical equipment, Geographical Information Systems (GIS), image database exploration etc., it is expensive and difficult for the users to examine spatial data in detail. Clustering may help to automate the process of analysing and understanding spatial data. It is used to identify and extract interesting characteristics and patterns that may exist in large spatial databases.

• ***Web mining.***

In this case, clustering is used to discover significant groups of documents on the Web huge collection of semi-structured documents. This classification of Web documents assists in information discovery.

In general terms, clustering may serve as a pre-processing step for other algorithms, such as classification, which would then operate on the detected clusters.

2.4 ROUGH SET THEORY

Rough set theory is a new mathematical approach to imperfect knowledge. The problem of imperfect knowledge, tackled for a long time by philosophers, logicians and mathematicians, has become also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge.

The most successful one is, no doubt, fuzzy set theory proposed by Zadeh (Zadeh, 1965). Rough set theory (Pawlak, 1982) presents still another attempt at this problem. This theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition.

The rough set approach provides efficient algorithms for finding hidden patterns in data, minimal sets of data (data reduction), evaluating significance of data, and generating sets of decision rules from data. This approach is easy to understand, offers straightforward interpretation of obtained results, most of its algorithms are particularly suited for parallel processing.

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information. For example, if objects are patients suffering from a certain disease, symptoms of the disease form information about patients. Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The *indiscernibility relation* generated in this way is the mathematical basis of rough set theory.

Any set of all indiscernible (similar) objects is called an *elementary set* (*neighborhood*), and forms a basic *granule* (*atom*) of knowledge about the universe. Any union of elementary sets is referred to as *crisp* (*precise*) set - otherwise the set is *rough* (*imprecise, vague*).

Consequently each rough set has *boundary-line cases,* i.e., objects which cannot with certainty be classified either as members of the set or of its complement. Obviously crisp sets have no boundary-line elements at all. This means that boundary-line cases cannot be properly classified by employing the available knowledge.

Vague concepts, in contrast to precise concepts, cannot be characterized in terms of information about their elements. Therefore, in the proposed approach, we assume that any vague concept is replaced by a pair of precise concepts, called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contains all objects which possibly belong to the concept. The difference between the upper and the lower approximation constitutes *the boundary region* of the vague concept. Approximations are two basic operations in rough set theory. The observation that vague concepts should have non-empty boundary was made by Gottlob Frege (Frege, 1905).

Vague complex concept approximation and reasoning about vague concepts by means of such approximations become critical for numerous applications related to multiagent systems (such as web mining, e-commerce, monitoring, security and rescue tasks in multiagent systems, cooperative problem solving, intelligent smart sensor fusion, human-computer interfaces, telemedicine, soft views of databases for specific customers). Further development of such methods for approximate reasoning by intelligent agents (Russell & Norvig, 2003) based on databases and knowledge bases is needed.

2.4.1 Rough Set: An Approach to Vagueness

The very successful technique for rough set methods was Boolean reasoning [25]. The idea of Boolean reasoning is based on construction for a given problem *P* a corresponding Boolean function *fP* with the following property: the solutions for the problem P can be decoded from prime implicants of the Boolean function fP. It is worth to mention that to solve real-life problems it is necessary to deal with Boolean functions having a large number of variables.

2.4.2 History of Rough Set

Rough Set Theory can deal with inexact, uncertain, and vague datasets (Walczak & Massart, 1999). Both Fuzzy Set Theory and Rough Set Theory are used with the indiscernibility relation and perceptible knowledge. The major difference between them is that Rough Set Theory does not need a membership function; thus, it can avoid pre-assumption and one-sided information analysis. A detailed discussion of Rough Set Theory can be found in Walczak and Massart (1999). Rough Set Theory was developed by Pawlak (1982, 1984, 2004). It has been applied to the management of a number of the issues, including: medical diagnosis, engineering reliability, expert systems, empirical study of materials data (Jackson, Leclair, Ohmer, Ziarko, & Al-kamhwi, 1996), machine diagnosis (Zhai, Khoo, & Fok, 2002), business failure prediction (Beynon & Peel, 2001; Dimitras, Slowinski, Susmaga, & Zopounidis, 1999), activity-based travel modeling (Witlox & Tindemans, 2004), travel demand analysis (Goh & Law, 2003), solving linear programs (Azibi & Vanderpooten, 2002), data mining (Li & Wang, 2004; Hu, Chen, & Tzeng, 2003; Chan, 1998), and a-RST (Quafafou, 2000). Another paper discusses the preference-order of attribute criteria needed to extend the original Rough Set Theory, such as sorting, choice and ranking problem (Greco, Matarazzo, & Slowinski, 2001). The Rough Set method is useful for exploring data patterns because of its ability to search through a multi-dimensional data space and determine the relative importance of each attribute with respect to its output. Rough Set Theory in analyzing the attributes of combination values for the insurance market

2.4.3 Fuzzy Set

Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets were introduced simultaneously by Lotfi A. Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition — an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval [0, 1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1. In fuzzy set theory, classical bivalent sets are usually called *crisp* sets. The fuzzy set theory can be used in a wide range of domains in which information is incomplete or imprecise, such as bioinformatics.

Fuzzy sets can be applied, for example, to the field of genealogical research. When an individual is searching in vital records such as birth records for possible ancestors, the researcher must contend with a number of issues that could be encapsulated in a membership function. Looking for an ancestor named John Henry Pittman, who you *think* was born in (probably eastern) Tennessee circa 1853 (based on statements of his age in later censuses, and a marriage record in Knoxville), what is the likelihood that a particular birth record for "John Pittman" is *your* John Pittman? What about a record in a different part of Tennessee for "J.H. Pittman" in 1851? (It has been suggested by Thayer Watkins that Zadeh's ethnicity is an example of a fuzzy set)

2.4.4 Relation between Rough Set Theory and Fuzzy Set Theory

The theory, which was developed by Pawlak (1982), is a rule-based decision-making technique that can handle crisp datasets and fuzzy datasets without the need for a pre-assumption membership function, which fuzzy theory requires. It can also deal with uncertain, vague, and imperceptible data. Until now, analysis of the attributes of combination values using Rough Set Theory has only been addressed by a few papers. In this study, a questionnaire with single-choice and multi-choice answers is used to apply Rough Set Theory to investigate the relationship between a single value and a combination of values of the attributes. Based on expert knowledge, the value class of the questions with multi-choice answers is reclassified in order to simplify the value complexity, which is useful in the decision-making procedure.

Rough Set Theory in analyzing the attributes of combination values for the insurance market. Many attempts have been made to combine theories of fuzzy sets and rough sets in order to have an algebra which is both an extension and a deviation of classical set algebra. [26, 27] One may introduce additional set-theoretic operators in the theory of fuzzy sets, or use graded binary relations in the theory of rough sets. [28] Our comparisons of fuzzy sets and rough sets lead to further interesting results. The notion of interval fuzzy sets have been introduced for both modal logic based fuzzy sets and set-oriented view of rough sets.

2.4.5 Applications of Rough Set

Fault Diagnosis using Rough Sets Theory

- Diagnosis of a valve fault for a multi-cylinder diesel engine
- Rough Sets Theory is used to analyze the decision table composed of attributes extracted from the vibration signals

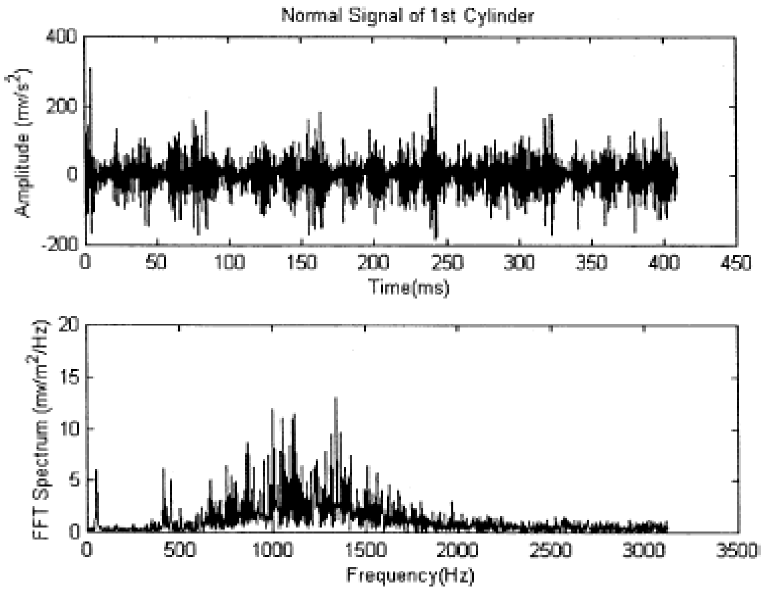


Figure 2.5: Normal state in Rough Set Theory

$$BN_B(X) = \overline{B}X - \underline{B}X$$

$$\underline{\gamma}_B(X) = \frac{|\underline{B}X|}{|U|} \quad \text{and} \quad \overline{\gamma}_B(X) = \frac{|\overline{B}X|}{|U|}.$$

$$\alpha_B(\Omega) = \frac{\Sigma \operatorname{card}(\underline{B}X_i)}{\Sigma \operatorname{card}(\overline{B}X_i)}$$

$$\gamma_B(\Omega) = \frac{\Sigma \operatorname{card}(\underline{B}X_i)}{\operatorname{card}(U)}$$

- Investigates the safety aspects of computer software in safety-critical applications
- Assessment of software safety using qualitative evaluations
- Use of checklists to collect data on software quality
- Waterfall model
  - Project Planning
  - Specification of requirements
  - Design
  - Implementation and integration
  - Verification and validation
  - Operation and maintenance

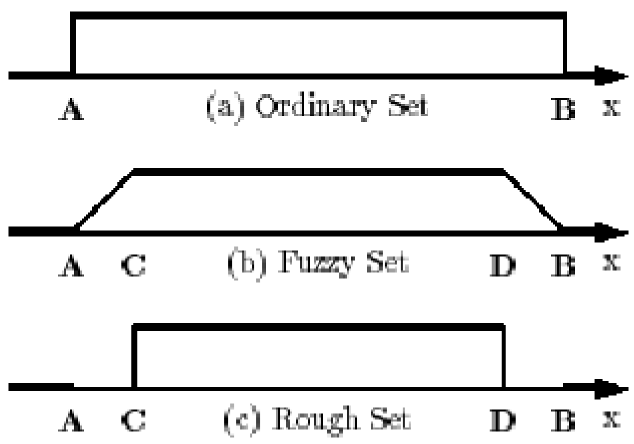


Figure 2.6: Illustration of the Notion of a Rough Set.

Economic and Financial Prediction using Rough Sets Model

- Applications of Rough Sets model in economic and financial prediction
- Emphasis on main areas of business failure prediction, database marketing and financial investment.

Table 2.2: Rough set models and its corresponding application.

Rough sets models	Business failure prediction	Database marketing	Financial investment
RSES			Bazan et al. (1994) Baltzersen (1996)
LERS		Poel (1998)	
DataLogic	Szladow and Mills (1993)	Mills (1993) Mrozek and Skabek (1998)	Ziarko et al. (1993) Golan (1995) Golan and Edwards (1993) Ruggiero (1994a,b,c) Skalko (1996) Lin and Tremba (2000)
TRANCE		Eiben et al. (1998) Kowalczyk and Slisser (1997) Kowalczyk and Piasta (1998) Kowalczyk (1998a)	
ProbRough		Poel (1998) Poel and Piasta (1998)	
Dominance relation	Greco et al. (1998)		
RoughDas and ProFit	Slowinski and Zopounidis (1994, 1995) Dimitras et al. (1999) Slowinski et al. (1997) Slowinski et al. (1999)		Susmaga et al. (1997)
Hybrid model	Ahn et al. (2000) Hashemi et al. (1998)		

2.5 ROUGH CLUSTERING

A rough cluster is defined in a similar manner to a rough set. The lower approximation of a rough cluster contains objects that only belong to that cluster. The upper approximation of a rough cluster contains objects in the cluster which are also members of other clusters. [29] The advantage of using rough sets is that, unlike other techniques, rough set theory does not require any prior information about the data such as apriori probability in statistics and a membership function in fuzzy set theory.

2.5.1 Application of Rough Set in Data Clustering

The theory of rough set, purposed by Pawlak [30], is an extension of set theory for the study of intelligent system characterized by insufficient and incomplete information. Using the concepts of lower and upper approximations in rough set theory, knowledge hidden in information systems may be unraveled and expressed in the form of decision rules. The conventional rough set model is based upon indiscernability relation. [31] On the basis of the presented approach a program has been developed (Fila & Wilk, 1983) which computes lower and upper approximations of sets, checks whether a set of attributes is dependent or independent, computers reducts of a set of attributes and computers accuracy of approximation. As an example, the program has been used for medical data analysis. A file of 150 patients suffering from heart disease seen in one of the hospitals in Warsaw was used as a data base. All patients have been divided by experts into six classes corresponding to their health status. With every patient seven items of information (attributes) were associated. For the sake of simplicity attributes were numbered 1,2,3,4,5,6,and 7, and their domains were $V_1 = V_2 = V_3 = V_4 = V_5 = \{0,1,2\} V_6 = \{0,1,2,3,4\}\, and\, V_7 = \{0,1,2,3\}$. The problem was to find the description of each class in terms of data available for dent, find reducts for each class and computer accuracy of descriptions. [32]

2.5.2 Rough Set Theory in Categorical Data Clustering

Data clustering is a popular data analysis task that involves the distribution of ‘unannotated’ data (i.e. with no a priori class information), in an inductive manner, into a finite sets of categories or cluster such that data items within a cluster are similar in some respect and unlike those from other cluster. [33] If one regards data as an underlying quantitative statement about a system’s behavior-either human or engineered-within a particular environment, then exploratory data clustering algorithms attempt to learn the topology of the data by analyzing the inherent similarities and differences of the individual data items in the untagged data set. Notwithstanding the efficacy of traditional data clustering techniques, it can be argued that the outcome of a data clustering task does not necessarily explicate the intrinsic relationship between the various attributes of the dataset. Given an un-annotated dataset satisfying the above assumption, we first partition it into *k* cluster, where each cluster comprises data-vectors with similar inherent characteristic. Note that the data clustering task is carried out with no a priori knowledge about the intrinsic class structure-i.e. how the data is inherently partitioned into distinct clusters. This is the phase 1 then next phase is data discretisation. The motivation for thus phase is driven by the fact the ordinal or continuous valued attributes are proven to be rather unsuitable for the extraction of concise symbolic rules. Last phase is symbolic rule discovery. We use rough set approximation-an interesting alternative to a variety of symbolic rule methods to derive symbolic rules that explain the inherent dependencies, attribute significance and structural characteristic of the annotated and clustered data-set.

CHAPTER 3

METHODOLOGY

This chapter briefly discusses about the model and method of data clustering based on Rough Set Theory (RST). The method so-called Max-Max Roughness is presented details here. There are three main sections in this chapter. The first section is rough set theory, which describes about information system, indiscernibility relations, approximation space and set approximations. The second section describes Max-Max Roughness (MMR) technique, together with an example of the application of the technique on a heart disease dataset. The last section describes the object splitting model.

3.1 ROUGH SET THEORY

The problem of imprecise knowledge has been tackled for a long time by mathematicians. Recently it became a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imprecise knowledge. The most successful one is, no doubt, the fuzzy set theory proposed by Zadeh (Zadeh, 1965). The basic tools of the theory are possibility measures. There is extensive literature on fuzzy logic with also discusses some of the problem with this theory. The basic problem of fuzzy set theory is the determination of the grade of membership of the value of possibility (Busse, 1998).

In the 1980's, Pawlak introduced rough set theory to deal this problem (Pawlak, 1982). Similarly to rough set theory it is not an alternative to classical set theory but it is embedded in it. Fuzzy and rough sets theories are not competitive, but complementary to each other (Pawlak and Skowron, 2007; Pawlak, 1985). Rough set theory has attracted attention to many researchers and practitioners all over the world, who contributed essentially to its development and applications. The original goal of the rough set theory is induction of approximations of concepts. The idea consists of approximation of a subset by a pair of two precise concepts called the *lower approximation* and *upper approximation*. Intuitively, the lower approximation of a set consists of all elements that surely belong to the set, whereas the upper approximation of the set constitutes of all elements that possibly belong to the set. The difference of the upper approximation and the lower approximation is a *boundary region*. It consists of all elements that cannot be classified uniquely to the set or its complement, by employing available knowledge. Thus any rough set, in contrast to a crisp set, has a non-empty boundary region. Motivation for rough set theory has come from the need to represent a subset of a universe in terms of equivalence classes of a partition of the universe. In this chapter, the basic concept of rough set theory in terms of data is presented.

3.1.1 Information Systems

Data are often presented as a table, columns of which are labeled by *attributes*, rows by *objects* of interest and entries of the table are *attribute values*. By an *information system*, we mean a 4-tuple (quadruple) $S = (U, A, V, f)$, where U is a non-empty finite set of objects, $A$is a non-empty finite set of attributes, $V = \bigcup_{a \in A} V_a$, V_a is the domain (value set) of attribute *a*, $f : U \times A \rightarrow V$ is a total function such that $f(u,a) \in V_a$, for every $(u,a) \in U \times A$, called information (knowledge) function. An information system is also called a knowledge representation systems or an attribute-valued system and can be intuitively expressed in terms of an information table (refer to Table 3.1).

Table 3.1: An information system

U	a_1	a_2	$\cdots$	a_k	$\cdots$	$a_{\|A\|}$
u_1	$f(u_1, a_1)$	$f(u_1, a_2)$	$\cdots$	$f(u_1, a_k)$	$\cdots$	$f(u_1, a_{\|A\|})$
u_2	$f(u_2, a_1)$	$f(u_2, a_2)$	$\cdots$	$f(u_2, a_k)$	$\cdots$	$f(u_2, a_{\|A\|})$
u_3	$f(u_3, a_1)$	$f(u_3, a_2)$	$\cdots$	$f(u_3, a_k)$	$\cdots$	$f(u_3, a_{\|A\|})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	$f(u_{\|U\|}, a_1)$	$f(u_{\|U\|}, a_2)$	$\cdots$	$f(u_{\|U\|}, a_k)$	$\cdots$	$f(u_{\|U\|}, a_{\|A\|})$

In many applications, there is an outcome of classification that is known. This *a posteriori* knowledge is expressed by one (or more) distinguished attribute called decision attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A *decision system* is an information system of the form $D = (U, A = C \bigcup D, V, f)$, where D is the set of *decision attributes* and $C \bigcap D = \phi$. The elements of C are called *condition attributes*. A simple example of decision system is given in Table 3.2.

Suppose that data about 10 patients is given, as shown in Table 3.2.

Table 3.2: A heart disease decision system

Patient/ Attribute	age	sex	cp	rbp	cho	fbs	thal	thalach	exia	oldpeak
1	60.0	1.0	1.0	140.0	230.0	1.0	6.0	150.0	2.0	2.3
2	60.0	1.0	4.0	160.0	280.0	2.0	3.0	100.0	1.0	1.5
3	60.0	1.0	4.0	120.0	220.0	2.0	7.0	120.0	1.0	2.6
4	30.0	1.0	3.0	130.0	250.0	2.0	3.0	180.0	2.0	3.5
5	40.0	2.0	2.0	130.0	200.0	2.0	3.0	170.0	2.0	1.4
6	50.0	1.0	2.0	120.0	230.0	2.0	3.0	170.0	2.0	0.8
7	60.0	2.0	4.0	140.0	260.0	2.0	3.0	160.0	2.0	3.6
8	50.0	2.0	4.0	120.0	350.0	2.0	3.0	160.0	1.0	0.6
9	60.0	1.0	4.0	130.0	250.0	2.0	7.0	140.0	2.0	1.4
10	50.0	1.0	4.0	140.0	200.0	1.0	7.0	150.0	1.0	3.1

The following values are obtained from Table 3.2,

$$U = \{1,2,3,4,5,6,7,8,9,10\},$$

$$A = \{\text{age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak}\},$$

$$V_{\text{age}} = \{30, 40, 50, 60\},$$

$$V_{\text{sex}} = \{1, 2\}$$

$$V_{\text{cp}} = \{1, 2, 3, 4\},$$

$$V_{\text{rbp}} = \{120, 130, 140, 160\},$$

$$V_{\text{cho}} = \{200, 220, 230, 250, 260, 280, 350\},$$

$$V_{\text{fbs}} = \{1, 2\},$$

$$V_{\text{thal}} = \{3, 6, 7\},$$

$$V_{\text{thalach}} = \{100, 120, 140, 150, 160, 170, 180\},$$

$$V_{\text{exia}} = \{1, 2\},$$

$$V_{\text{oldpeak}} = \{0.6, 0.8, 1.4, 1.5, 2.3, 2.6, 3.1, 3.5, 3.6\}.$$

A relational database may be considered as an information system in which rows are labeled by the objects (entities), columns are labeled by attributes and the entry in row *u* and column *a* has the value $f(u,a)$. It is noted that each map

$$f(u,a): U \times A \rightarrow V \text{ is a tuple } t_i = \left(f(u_i, a_1), f(u_i, a_2), f(u_i, a_3), \cdots, f\left(u_i, a_{|A|}\right)\right),$$

$$\text{for } 1 \le i \le |U|, \text{ where } |X| \text{ is the cardinality of } X.$$

Note that the tuple t is not necessarily associated with entity uniquely. In an information table, two distinct entities could have the same tuple representation (duplicated/redundant tuple), which is *not permissible* in relational databases. Thus, the concepts in information systems are a generalization of the same concepts in relational databases.

3.1.2 Indiscernibility Relation

From Table 3.2, it is noted that no atrribute are indiscernible (or similar or indistinguishable). The starting point of rough set theory is the indiscernibility relation, which is generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. Therefore, generally, we are unable to deal with single object. Nevertheless, we have to consider clusters of indiscernible objects. The following definition precisely defines the notion of indiscernibility relation between two objects.

Definition 2.1. *Let $S = (U, A, V, f)$ be an information system and let B be any subset of A. Two elements $x, y \in U$ are said to be B-indiscernible (indiscernible by the set of attribute $B \subseteq A$ in S) if and only if $f(x, a) = f(y, a)$, for every $a \in B$.*

Obviously, every subset of *A* induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute *B*, denoted by $IND(B)$, is an equivalence relation. It is well known that, an equivalence relation induces unique partition. The partition of *U* induced by $IND(B)$ in $S = (U, A, V, f)$ denoted by U / B and the equivalence class in the partition U / B containing $x \in U$, denoted by $[x]_B$.

Studies of rough set theory may be divided into two class, representing the set-oriented (constructive) and operator-oriented (descriptive) views. They produce extension of crisp set theory (Yao, 1996; Yao, 1998; Yao, 2001). In this work, rough set theory is presented from the point of view of a constructive approach.

3.1.3 Approximation Space

Let $S = (U, A, V, f)$ be an information system, let *B* be any subset of *A* and $IND(B)$ is an indiscernibility relation generated by *B* on *U*.

Definition 2.2. *An ordered pair* $AS = (U, IND(B))$ *is called a (Pawlak) approximation space.*

Let $x \in U$, the equivalence class of U containing x with respect to R is denoted by $[x]_B$. The family of definable sets, i.e. finite union of arbitrary equivalence classes in partition $U / IND(B)$ in AS, denoted by $DEF(AS)$ is a Boolean algebra (Pawlak, 1982). Thus, an approximation space defines unique topological space, called a *quasi-discrete* (*clopen*) *topological space* (Herawan and Mat Deris, 2009a). Given arbitrary subset $X \subseteq U$, X may not be presented as union of some equivalence classes in U. In other means that a subset X cannot be described precisely in AS. Thus, a subset X may be characterized by a pair of its approximations, called lower and upper approximations. It is here that the notion of rough set emerges.

3.1.4 Set Approximations

The indiscernibility relation will be used to define set approximations that are the basic concepts of rough set theory. The notions of lower and upper approximations of a set can be defined as follows.

Definition 2.3.*Let* $S = (U, A, V, f)$ *be an information system, let B be any subset of A and let X be any subset of U. The B-lower approximation of X, denoted by* $\underline{B}(X)$ *and B-upper approximations of X, denoted by* $\overline{B}(X)$*, respectively, are defined by*

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\} \text{ and } \overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \phi\}.$$

From Definition 2.3, the following interpretations are obtained

a. The *lower approximation* of a set X with respect to B is the set of all objects, which can be for *certain* classified as X using B (are certainly X in view of B).

b. The *upper approximation* of a set X with respect to B is the set of all objects which can be *possibly* classified as X using B (are possibly X in view of B).

Hence, with respect to arbitrary subset $X \subseteq U$, the universe U can be divided into three disjoint regions using the lower and upper approximations

a. The *positive region* $\mathrm{POS}_B(X) = \underline{B}(X)$, i.e., the set of all objects, which can be for *certain* classified as Xusing$B$ (are *certainly*X with respect to B).

b. The *boundary region* $\mathrm{BND}_B(X) = \overline{B}(X) - \underline{B}(X)$, i.e., the set of all objects, which can be classified neither as X nor as not-Xusing B.

c. The *negative region* $\mathrm{NEG}_B(X) = U - \overline{B}(X)$, i.e., the set of all objects, which can be for *certain* classified as not-Xusing B (are *certainly* not-X with respect to B).

These notions of lower and upper approximations can be shown clearly as in Figure 3.1.

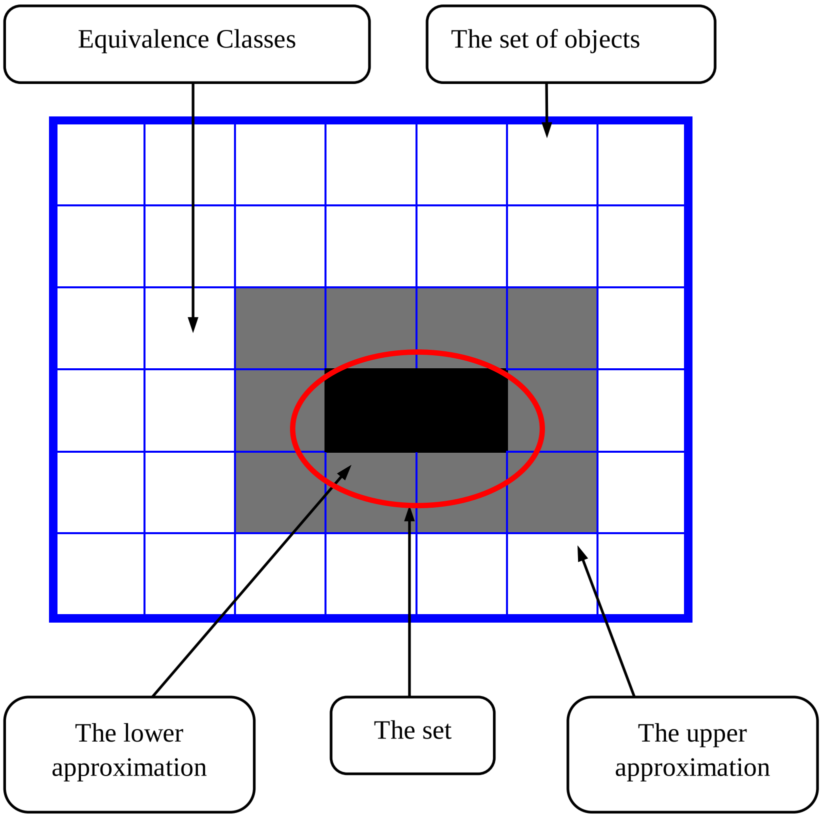


Figure 3.1: Set approximations

From Figure 3.1, three disjoint regions are given as follows

a. The positive region

b. The boundary region

c. The negative region

It is easily seen that the lower and the upper approximations of a set, respectively, are *interior* and *closure* operations in a quasi discrete topology generated by the indiscernibility relation [Herawan and Mat Deris, 2009e].

The accuracy of approximation (accuracy of roughness) of any subset $X \subseteq U$ with respect to $B \subseteq A$, denoted $\alpha_B(X)$ is measured by

$$\alpha_B(X) = \frac{\left|\underline{B}(X)\right|}{\left|\overline{B}(X)\right|}, \qquad (2.1)$$

where $|X|$ denotes the cardinality of *X*. For empty set ϕ, it is defined that $\alpha_B(\phi) = 1$ (Pawlak and Skowron, 2007). Obviously, $0 \le \alpha_B(X) \le 1$. If X is a union of some equivalence classes of *U*, then $\alpha_B(X) = 1$. Thus, the set X is *crisp* (precise) with respect to *B*. And, if *X* is not a union of some equivalence classes of *U*, then $\alpha_B(X) < 1$. Thus, the set X is *rough* (imprecise) with respect to B (Pawlak and Skowron, 2007). This means that the higher of accuracy of approximation of any subset $X \subseteq U$ is the more precise (the less imprecise) of itself.

Example 3.2. Let us depict above notions by examples referring to Table 3.2. Consider the set $U = \{1,2,3,4,5,6,7,8,9,10\}$ and the set of attributes $V_{\mathrm{age}} = \{30, 40, 50, 60\}$. The partition of *U* induced by $IND(sex)$ is given by

$$U / IND(age) = \{\{4\}, \{5\}, \{6,8,10\}, \{1,2,3,7,9\}\}.$$

The corresponding lower approximation and upper approximation of X are as follows

$$\underline{X(age=60)}=\phi \text{ and } \overline{X(age=60)}=\{1,4,6,9,10\}.$$

Thus, concept "Decision" is imprecise (rough). For this case, the accuracy of approximation is given as

$$\alpha_C(X)=\frac{0}{5}.$$

It means that the concept "Decision" can be characterized partially employing attributes age, sex, cp, rbp, cho, fbs, thal, thalach, exia and oldpeak .

3.2 MAX-MAX ROUGHNESS

In this section, a technique for selecting a clustering attribute based on rough set theory is presented. Parmar *et al.* proposes a technique called Max-Max Roughness (MMR) which can handle datasets with multi-valued attributes **[Parmar]**. This section, however, will be presenting the technique MMR to select the clustering attributes.

3.2.1 Selecting a Clustering Attributes

To find meaningful clusters from a dataset, clustering attribute is conducted so that attributes within the clusters made will have a high correlation or high interdependence to each other while the attributes in other clusters are less correlated or more independent.

3.2.2 Model for Selecting a Clustering Attributes

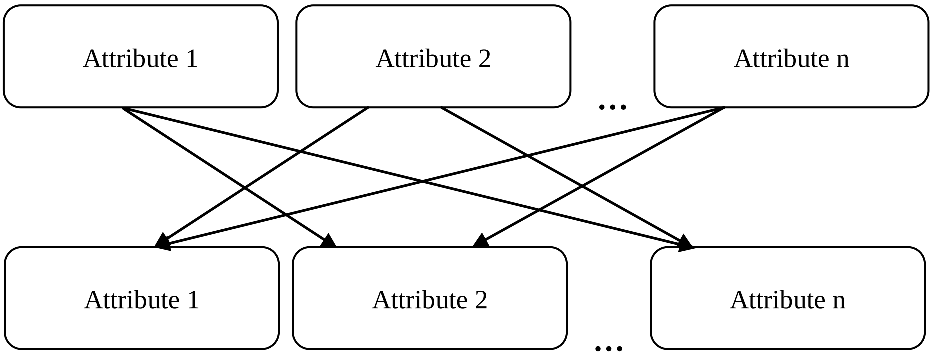


Figure 3.2: A model for selecting a clustering attribute

3.2.3 Max-Max Roughness Technique

The following Table shows step-by-step to calculate Max-Max Roughness.

Table 3.3: Step-by-step Max-Max Roughness

Step	Max-Max Roughness
1	Given data set
2	Each attribute in data set considered as a candidate attribute to partition
3	Determine equivalence classes of attribute-value pairs
4	Determine lower approximation of each equivalence classes in attribute a_i w.r.t. to attribute a_j, $i \neq j$
5	Determine upper approximation of each equivalence classes in attribute a_i w.r.t. to attribute a_j, $i \neq j$
6	Calculate roughness of each equivalence classes in attribute a_i w.r.t. to attribute a_j, $i \neq j$

7	Calculate mean roughness of attribute a_i w.r.t. to attribute a_j, $i \neq j$
8	Calculate maximum roughness a_i w.r.t. to all attribute a_j, $i \neq j$
9	If there are two greatest value of mean roughness, calculate maximum roughness relative to the second, third greater maximum roughness until the tie is broken
10	Selecting a clustering attribute

3.2.4 Example

The following example shows a calculation resultof Max-Max Roughness through an information system taken from **[Parmar et al, 2007]**.

Table 3.4: An information system of heart disease in MMR **[Parmar]**

Patient/ Attribute	age	sex	cp	rbp	cho	fbs	thal	thalach	exia	oldpeak
1	60.0	1.0	1.0	140.0	230.0	1.0	6.0	150.0	2.0	2.3
2	60.0	1.0	4.0	160.0	280.0	2.0	3.0	100.0	1.0	1.5
3	60.0	1.0	4.0	120.0	220.0	2.0	7.0	120.0	1.0	2.6
4	30.0	1.0	3.0	130.0	250.0	2.0	3.0	180.0	2.0	3.5
5	40.0	2.0	2.0	130.0	200.0	2.0	3.0	170.0	2.0	1.4
6	50.0	1.0	2.0	120.0	230.0	2.0	3.0	170.0	2.0	0.8
7	60.0	2.0	4.0	140.0	260.0	2.0	3.0	160.0	2.0	3.6
8	50.0	2.0	4.0	120.0	350.0	2.0	3.0	160.0	1.0	0.6
9	60.0	1.0	4.0	130.0	250.0	2.0	7.0	140.0	2.0	1.4
10	50.0	1.0	4.0	140.0	200.0	1.0	7.0	150.0	1.0	3.1

As an information system, from Table 3.4, we have

$$U = \{1,2,3,4,5,6,7,8,9,10\},$$

$$A = \{\text{age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak}\},$$

$$V_{\text{age}} = \{30, 40, 50, 60\},$$

$$V_{\text{sex}} = \{1, 2\}$$

$$V_{\text{cp}} = \{1, 2, 3, 4\},$$

$$V_{\text{rbp}} = \{120, 130, 140, 160\},$$

$$V_{\text{cho}} = \{200, 220, 230, 250, 260, 280, 350\},$$

$$V_{\text{fbs}} = \{1, 2\},$$

$$V_{\text{thal}} = \{3, 6, 7\},$$

$$V_{\text{thalach}} = \{100, 120, 140, 150, 160, 170, 180\},$$

$$V_{\text{exia}} = \{1, 2\},$$

$$V_{\text{oldpeak}} = \{0.6, 0.8, 1.4, 1.5, 2.3, 2.6, 3.1, 3.5, 3.6\}.$$

Ten partitions of *U* generated by indiscernibility relation of singleton attribute are;

a. $X(age = 30) = \{4\}$, $X(age = 40) = \{5\}$, $X(age = 50) = \{6,8,10\}$, $X(age = 60) = \{1,2,3,7,9\}$,

$$U / IND(age) = \{\{4\}, \{5\}, \{6,8,10\}, \{1,2,3,7,9\}\}.$$

b. $X(sex = 1) = \{1,2,3,4,6,9,10\}$, $X(sex = 2) = \{5,7,8\}$,

$$U / IND(sex) = \{\{1,2,3,4,6,9,10\}, \{5,7,8\}\}.$$

c. $X(cp = 1) = \{1\}$, $X(cp = 2) = \{5,6\}$, $X(cp = 3) = \{4\}$, $X(cp = 4) = \{2,3,7,8,9,10\}$,

$$U / IND(cp) = \{\{1\}, \{5,6\}, \{4\}, \{2,3,7,8,9,10\}\}.$$

d. $X(rbp = 120) = \{3,6,8\}$, $X(rbp = 130) = \{4,5,9\}$, $X(rbp = 140) = \{1,7,10\}$, $X(rbp = 160) = \{2\}$,

$$U / IND(rbp) = \{\{3,6,8\}, \{4,5,9\}, \{1,7,10\}, \{2\}\}.$$

e. $X(cho = 200) = \{5{,}10\}$, $X(cho = 220) = \{3\}$, $X(cho = 230) = \{1{,}6\}$,

$X(cho = 250) = \{4{,}9\}$, $X(cho = 260) = \{7\}$, $X(cho = 280) = \{2\}$,

$X(cho = 350) = \{8\}$,

$$U / IND(a_5) = \{\{5{,}10\}, \{3\}, \{1{,}6\}, \{4{,}9\}, \{7\}, \{2\}, \{8\}\}.$$

f. , $X(fbs = 1) = \{1{,}10\}$ $X(fbs = 2) = \{2{,}3{,}4{,}5{,}6{,}7{,}8{,}9\}$,

$$U / IND(fbs) = \{\{1{,}10\}, \{2{,}3{,}4{,}5{,}6{,}7{,}8{,}9\}\}.$$

g. $X(thal = 3) = \{2{,}4{,}5{,}6{,}7{,}8\}$, $X(thal = 6) = \{1\}$, $X(thal = 7) = \{3{,}9{,}10\}$,

$$U / IND(thal) = \{\{2{,}4{,}5{,}6{,}7{,}8\}, \{1\}, \{3{,}9{,}10\}\}.$$

h. $X(thalach = 100) = \{2\}$, $X(thalach = 120) = \{3\}$, $X(thalach = 140) = \{9\}$,

$X(thalach = 150) = \{1{,}10\}$, $X(thalach = 160) = \{7{,}8\}$,

$X(thalach = 170) = \{5{,}6\}$, $X(thalach = 180) = \{4\}$,

$$U / IND(thalach) = \{\{2\}, \{3\}, \{9\}, \{1{,}10\}, \{7{,}8\}, \{5{,}6\}, \{4\}\}.$$

i. $X(exia = 1) = \{2{,}3{,}8{,}10\}$, $X(exia = 2) = \{1{,}4{,}5{,}6{,}7{,}9\}$,

$$U / IND(exia) = \{\{2{,}3{,}8{,}10\}, \{1{,}4{,}5{,}6{,}7{,}9\}\}.$$

j. $X(oldpeak = 0.6) = \{8\}$, $X(oldpeak = 0.8) = \{6\}$, $X(oldpeak = 1.4) = \{5{,}9\}$,

$X(oldpeak = 1.5) = \{2\}$, $X(oldpeak = 2.3) = \{1\}$, $X(oldpeak = 2.6) = \{3\}$,

$X(oldpeak = 3.1) = \{10\}$, $X(oldpeak = 3.5) = \{4\}$, $X(oldpeak = 3.6) = \{7\}$,

$$U / IND(fbs) = \{\{8\}, \{6\}, \{5{,}9\}, \{2\}, \{1\}, \{3\}, \{10\}, \{4\}, \{7\}\}.$$

Calculation of Roughness and Mean Roughness

First, we determine of upper and lower approximations of singleton attribute with respect to other different singleton attribute. Then we calculate the roughness and the mean roughness of each attribute.

a. **Attribute** *age*

For attribute *age*, it is clear that $\left|V(age)\right| = 4$.

The roughness and the mean roughness on *age* with respect to $A,\quad A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to *sex*

The lower and upper approximations are

$$\underline{X(age=30)} = \phi \text{ and } \overline{X(age=30)} = \{1,2,3,4,6,9,10\},$$

$$\underline{X(age=40)} = \phi \text{ and } \overline{X(age=40)} = \{5,7,8\},$$

$$\underline{X(age=50)} = \phi \text{ and } \overline{X(age=50)} = \{1,2,3,4,5,6,7,8,9,10\},$$

$\underline{X}(age=60)=\phi$ and $\overline{X}(age=60)=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rsex(X|age=30)=\frac{0}{7}=0,$$

$$Rsex(X|age=40)=\frac{0}{3}=0,$$

$$Rsex(X|age=50)=\frac{0}{10}=0,$$

$$Rsex(X|age=60)=\frac{0}{10}=0.$$

Mean roughness

$$Rough_{sex}(age)=\frac{0+0+0+0}{4}=0.$$

2) With respect to cp

The lower and upper approximations are

$\underline{X}(age=30)=\{4\}$ and $\overline{X}(age=30)=\{4\}$,

$\underline{X}(age=40)=\phi$ and $\overline{X}(age=40)=\{5,6\}$,

$\underline{X}(age=50)=\phi$ and $\overline{X}(age=50)=\{2,3,5,6,7,8,9,10\}$,

$\underline{X}(age=60)=\{1\}$ and $\overline{X}(age=60)=\{1,2,3,7,8,9,10\}$.

Roughness

$$Rcp(X|age=30)=\frac{1}{1}=1,$$

$$Rcp(X|age=40)=\frac{0}{2}=0,$$

$$Rcp(X|age=50)=\frac{0}{8}=0,$$

$$Rcp(X|age=60)=\frac{1}{7}=0.1429.$$

Mean roughness

$$Rough_{cp}(age)=\frac{1+0+0+0.1429}{4}=0.2857.$$

3) With respect to rbp

The lower and upper approximations are

$\underline{X(age=30)}=\phi$ and $\overline{X(age=30)}=\{4,5,9\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{4,5,9\}$,

$\underline{X(age=50)}=\phi$ and $\overline{X(age=50)}=\{1,3,6,7,8,10\}$,

$\underline{X(age=60)}=\{2\}$ and $\overline{X(age=60)}=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rrbp(X|age=30)=\frac{0}{3}=0,$$

$$Rrbp(X|age=40)=\frac{0}{3}=0,$$

$$Rrbp(X|age=50)=\frac{0}{6}=0.$$

$$Rrbp(X|age=60)=\frac{1}{10}=0.1.$$

Mean roughness

$$Rough_{rbp}(age)=\frac{0+0+0+0.1}{4}=0.025.$$

4) With respect to cho

The lower and upper approximations are

$\underline{X(age=30)}=\phi$ and $\overline{X(age=30)}=\{4,9\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{5,10\}$,

$\underline{X(age=50)}=\{8\}$ and $\overline{X(age=50)}=\{1,5,6,8,10\}$,

$\underline{X(age=60)}=\{2,3,7\}$ and $\overline{X(age=60)}=\{1,2,3,4,6,7,9\}$.

Roughness

$$Rcho(X|age=30)=\frac{0}{2}=0,$$

$$Rcho(X|age=40)=\frac{0}{2}=0,$$

$$Rcho(X|age=50)=\frac{1}{5}=0.2\,.$$

$$Rcho(X|age=60)=\frac{3}{7}=0.4286\,.$$

Mean roughness

$$Rough_{cho}(age)=\frac{0+0+0.2+0.4286}{4}=0.1571\,.$$

5) With respect to fbs

The lower and upper approximations are

$\underline{X(age=30)}=\phi$ and $\overline{X(age=30)}=\{2,3,4,5,6,7,8,9\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{2,3,4,5,6,7,8,9\}$,

$\underline{X(age=50)}=\phi$ and $\overline{X(age=50)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(age=60)}=\phi$ and $\overline{X(age=60)}=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rfbs(X|age=30)=\frac{0}{8}=0\,,$$

$$Rfbs(X|age=40)=\frac{0}{8}=0\,,$$

$$Rfbs(X|age=50)=\frac{0}{10}=0\,,$$

$$Rfbs(X|age=60)=\frac{0}{10}=0\,.$$

Mean roughness

$$Rough_{fbs}(age)=\frac{0+0+0+0}{4}=0\,.$$

6) With respect to $thal$

The lower and upper approximations are

$\underline{X(age=30)}=\phi$ and $\overline{X(age=30)}=\{2,4,5,6,7,8\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{2,4,5,6,7,8\}$,

$\underline{X(age=50)}=\phi$ and $\overline{X(age=50)}=\{2,3,4,5,6,7,8,9,10\}$,

$\underline{X(age=60)}=\{1\}$ and $\overline{X(age=60)}=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rfbs(X|age=30)=\frac{0}{6}=0,$$

$$Rfbs(X|age=40)=\frac{0}{6}=0,$$

$$Rfbs(X|age=50)=\frac{0}{9}=0,$$

$$Rfbs(X|age=60)=\frac{1}{10}=0.1$$

Mean roughness

$$Rough_{fbs}(age)=\frac{0+0+0+0.1}{4}=0.025\,.$$

7) With respect to *thalach*

The lower and upper approximations are

$\underline{X(age=30)}=\{4\}$ and $\overline{X(age=30)}=\{4\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{5,6\}$,

$\underline{X(age=50)}=\phi$ and $\overline{X(age=50)}=\{1,5,6,7,8,10\}$,

$\underline{X(age=60)}=\{2,3,9\}$ and $\overline{X(age=60)}=\{1,2,3,7,8,9,10\}$.

Roughness

$$Rfbs(X|age=30)=\frac{1}{1}=1,$$

$$Rfbs(X|age=40)=\frac{0}{2}=0,$$

$$Rfbs(X|age=50)=\frac{0}{6}=0,$$

$$Rfbs(X|age=60)=\frac{3}{7}=0.4286\,.$$

Mean roughness

$$Rough_{fbs}(age)=\frac{1+0+0+0.4286}{4}=0.3571\,.$$

8) With respect to *exia*

The lower and upper approximations are

$\underline{X(age=30)}=\phi$ and $\overline{X(age=30)}=\{1,4,5,6,7,9\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{1,4,5,6,7,9\}$,

$\underline{X(age=50)}=\phi$ and $\overline{X(age=50)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(age=60)}=\phi$ and $\overline{X(age=60)}=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rfbs(X|age=30)=\frac{0}{6}=0,$$

$$Rfbs(X|age=40)=\frac{0}{6}=0,$$

$$Rfbs(X|age=50)=\frac{0}{10}=0,$$

$$Rfbs(X|age=60)=\frac{0}{10}=0.$$

Mean roughness

$$Rough_{fbs}(age)=\frac{0+0+0+0}{4}=0.$$

9) With respect to *oldpeak*

The lower and upper approximations are

$\underline{X(age=30)}=\{4\}$ and $\overline{X(age=30)}=\{4\}$,

$\underline{X(age=40)}=\phi$ and $\overline{X(age=40)}=\{5,9\}$,

$\underline{X(age=50)}=\{6,8,10\}$ and $\overline{X(age=50)}=\{6,8,10\}$,

$\underline{X(age=60)}=\{1,2,3,7\}$ and $\overline{X(age=60)}=\{1,2,3,5,7,9\}$.

Roughness

$$Rfbs(X|age=30)=\frac{1}{1}=1,$$

$$Rfbs(X|age=40)=\frac{0}{2}=0,$$

$$Rfbs(X|age = 50) = \frac{3}{3} = 1,$$

$$Rfbs(X|age = 60) = \frac{4}{6} = 0.6667.$$

Mean roughness

$$Rough_{fbs}(age) = \frac{1 + 0 + 1 + 0.6667}{4} = 0.6667.$$

b. **Attribute** *sex*

For attribute sex, it is clear that $|V(sex)| = 2$.

The roughness and the mean roughness on sex with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(sex=1)} = \{4\}$ and $\overline{X(sex=1)} = \{1,2,3,4,6,7,8,9,10\}$,

$\underline{X(sex=2)} = \{5\}$ and $\overline{X(sex=2)} = \{1,2,3,5,6,7,8,9,10\}$.

Roughness

$$Rage(X|sex = 1) = \frac{1}{9} = 0.1111,$$

$$Rage(X|sex = 2) = \frac{1}{9} = 0.1111.$$

Mean roughness

$$Rough_{age}(sex) = \frac{0.1111 + 0.1111}{2} = 0.1111.$$

2) With respect to cp

The lower and upper approximations are

$\underline{X(sex=1)} = \{1,4\}$ and $\overline{X(sex=1)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(sex=2)} = \phi$ and $\overline{X(sex=2)} = \{2,3,5,6,7,8,9,10\}$.

Roughness

$$Rcp(X|sex=1)=\frac{2}{10}=0.2,$$

$$Rcp(X|sex=2)=\frac{0}{8}=0.$$

Mean roughness

$$Rough_{cp}(sex)=\frac{0.2+0}{2}=0.1.$$

3) With respect to rbp

The lower and upper approximations are

$\underline{X(sex=1)}=\{2\}$ and $\overline{X(sex=1)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(sex=2)}=\phi$ and $\overline{X(sex=2)}=\{1,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rrbp(X|sex=1)=\frac{1}{10}=0.1,$$

$$Rrbp(X|sex=2)=\frac{0}{9}=0.$$

Mean roughness

$$Rough_{rbp}(sex)=\frac{0.1+0}{2}=0.05.$$

4) With respect to cho

The lower and upper approximations are

$\underline{X(sex=1)}=\{1,2,3,4,6,9\}$ and $\overline{X(sex=1)}=\{1,2,3,4,5,6,9,10\}$,

$\underline{X(sex=2)}=\{7,8\}$ and $\overline{X(sex=2)}=\{5,7,8,10\}$.

Roughness

$$Rcho(X|sex=1)=\frac{6}{8}=0.75,$$

$$Rcho(X|sex=2)=\frac{2}{4}=0.5.$$

Mean roughness

$$Rough_{cho}(sex) = \frac{0.75 + 0.5}{2} = 0.625.$$

5) With respect to *fbs*

The lower and upper approximations are

$\underline{X(sex=1)} = \{1,10\}$ and $\overline{X(sex=1)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(sex=2)} = \phi$ and $\overline{X(sex=2)} = \{2,3,4,5,6,7,8,9\}$.

Roughness

$$Rfbs(X|sex=1) = \frac{2}{10} = 0.2,$$

$$Rfbs(X|sex=2) = \frac{0}{8} = 0.$$

Mean roughness

$$Rough_{fbs}(sex) = \frac{0.2 + 0}{2} = 0.1.$$

6) With respect to *thal*

The lower and upper approximations are

$\underline{X(sex=1)} = \{1,3,9,10\}$ and $\overline{X(sex=1)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(sex=2)} = \phi$ and $\overline{X(sex=2)} = \{2,4,5,6,7,8\}$.

Roughness

$$Rthal(X|sex=1) = \frac{4}{10} = 0.4,$$

$$Rthal(X|sex=2) = \frac{0}{6} = 0.$$

Mean roughness

$$Rough_{thal}(sex) = \frac{0.4 + 0}{2} = 0.2.$$

7) With respect to *thalach*

The lower and upper approximations are

$\underline{X(sex=1)} = \{1,2,3,4,9,10\}$ and $\overline{X(sex=1)} = \{1,2,3,4,5,6,9,10\}$,

$\underline{X(sex=2)} = \{7,8\}$ and $\overline{X(sex=2)} = \{5,6,7,8\}$.

Roughness

$$Rthalach\left(X\,|\,sex=1\right)=\frac{6}{8}=0.75,$$

$$Rthalach\left(X\,|\,sex=2\right)=\frac{2}{4}=0.5.$$

Mean roughness

$$Rough_{thalach}\left(sex\right)=\frac{0.75+0.5}{2}=0.625.$$

8) With respect to $exia$

The lower and upper approximations are

$$\underline{X\left(sex=1\right)}=\phi \text{ and } \overline{X\left(sex=1\right)}=\{1,2,3,4,5,6,7,8,9,10\},$$

$$\underline{X\left(sex=2\right)}=\phi \text{ and } \overline{X\left(sex=2\right)}=\{1,2,3,4,5,6,7,8,9,10\}.$$

Roughness

$$\operatorname{Re}xia\left(X\,|\,sex=1\right)=\frac{0}{10}=0,$$

$$\operatorname{Re}xia\left(X\,|\,sex=2\right)=\frac{0}{10}=0.$$

Mean roughness

$$Rough_{exia}\left(sex\right)=\frac{0+0}{2}=0.$$

9) With respect to $oldpeak$

The lower and upper approximations are

$$\underline{X\left(sex=1\right)}=\{1,2,3,4,6,10\} \text{ and } \overline{X\left(sex=1\right)}=\{1,2,3,4,5,6,9,10\},$$

$$\underline{X\left(sex=2\right)}=\{7,8\} \text{ and } \overline{X\left(sex=2\right)}=\{5,7,8,9\}.$$

Roughness

$$Roldpeak\left(X\,|\,sex=1\right)=\frac{6}{8}=0.75,$$

$$Roldpeak\left(X\,|\,sex=2\right)=\frac{2}{4}=0.5.$$

Mean roughness

$$Rough_{oldpeak}(sex) = \frac{0.75 + 0.5}{2} = 0.625\,.$$

c. **Attribute** cp

For attribute cp, it is clear that $|V(cp)| = 4$.

The roughness and the mean roughness on cp with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(cp=1)} = \phi$ and $\overline{X(cp=1)} = \{1,2,3,7,9\}$,

$\underline{X(cp=2)} = \{5\}$ and $\overline{X(cp=2)} = \{5,6,8,10\}$,

$\underline{X(cp=3)} = \{4\}$ and $\overline{X(cp=3)} = \{4\}$,

$\underline{X(cp=4)} = \phi$ and $\overline{X(cp=4)} = \{1,2,3,6,7,8,9,10\}$.

Roughness

$$Rage(X|cp=1) = \frac{0}{5} = 0\,,$$

$$Rage(X|cp=2) = \frac{1}{4} = 0.25\,,$$

$$Rage(X|cp=3) = \frac{1}{1} = 1\,,$$

$$Rage(X|cp=4) = \frac{0}{8} = 0\,.$$

Mean roughness

$$Rough_{age}(cp) = \frac{0 + 0.25 + 1 + 0}{4} = 0.3125\,.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X(cp=1)} = \phi$ and $\overline{X(cp=1)} = \{1,2,3,4,6,9,10\}$,

$\underline{X(cp=2)} = \phi$ and $\overline{X(cp=2)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(cp=3)=\phi$ and $\overline{X}(cp=3)=\{1,2,3,4,6,9,10\}$,

$\underline{X}(cp=4)=\phi$ and $\overline{X}(cp=4)=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rsex(X|cp=1)=\frac{0}{7}=0,$$

$$Rsex(X|cp=2)=\frac{0}{10}=0,$$

$$Rsex(X|cp=3)=\frac{0}{7}=0,$$

$$Rsex(X|cp=4)=\frac{0}{10}=0.$$

Mean roughness

$$Rough_{sex}(cp)=\frac{0+0+0+0}{4}=0.$$

3) With respect to rbp

The lower and upper approximations are

$\underline{X}(cp=1)=\phi$ and $\overline{X}(cp=1)=\{1,7,10\}$,

$\underline{X}(cp=2)=\phi$ and $\overline{X}(cp=2)=\{3,4,5,6,8,9\}$,

$\underline{X}(cp=3)=\phi$ and $\overline{X}(cp=3)=\{4,5,9\}$,

$\underline{X}(cp=4)=\{2\}$ and $\overline{X}(cp=4)=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rrbp(X|cp=1)=\frac{0}{3}=0,$$

$$Rrbp(X|cp=2)=\frac{0}{6}=0,$$

$$Rrbp(X|cp=3)=\frac{0}{3}=0,$$

$$Rrbp(X|cp=4)=\frac{1}{10}=0.1.$$

Mean roughness

$$Rough_{rbp}(cp) = \frac{0+0+0+0.1}{4} = 0.025.$$

4) With respect to cho

The lower and upper approximations are

$\underline{X(cp=1)} = \phi$ and $\overline{X(cp=1)} = \{1,6\}$,

$\underline{X(cp=2)} = \phi$ and $\overline{X(cp=2)} = \{1,5,6,10\}$,

$\underline{X(cp=3)} = \phi$ and $\overline{X(cp=3)} = \{4,9\}$,

$\underline{X(cp=4)} = \{2,3,7,8\}$ and $\overline{X(cp=4)} = \{2,3,4,5,7,8,9,10\}$.

Roughness

$$Rcho(X|cp=1) = \frac{0}{2} = 0,$$

$$Rcho(X|cp=2) = \frac{0}{4} = 0,$$

$$Rcho(X|cp=3) = \frac{0}{2} = 0,$$

$$Rcho(X|cp=4) = \frac{4}{8} = 0.5.$$

Mean roughness

$$Rough_{cho}(cp) = \frac{0+0+0+0.5}{4} = 0.125.$$

5) With respect to fbs

The lower and upper approximations are

$\underline{X(cp=1)} = \phi$ and $\overline{X(cp=1)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(cp=2)} = \phi$ and $\overline{X(cp=2)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(cp=3)} = \phi$ and $\overline{X(cp=3)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(cp=4)} = \phi$ and $\overline{X(cp=4)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rfbs(X|cp=1) = \frac{0}{2} = 0,$$

$$Rfbs(X|cp=2)=\frac{0}{8}=0,$$

$$Rfbs(X|cp=3)=\frac{0}{8}=0,$$

$$Rfbs(X|cp=4)=\frac{0}{10}=0.$$

Mean roughness

$$Rough_{fbs}(cp)=\frac{0+0+0+0}{4}=0.$$

6) With respect to *thal*

The lower and upper approximations are

$\underline{X(cp=1)}=\{1\}$ and $\overline{X(cp=1)}=\{1\}$,

$\underline{X(cp=2)}=\phi$ and $\overline{X(cp=2)}=\{2,4,5,6,7,8\}$,

$\underline{X(cp=3)}=\phi$ and $\overline{X(cp=3)}=\{2,4,5,6,7,8\}$,

$\underline{X(cp=4)}=\{3,9,10\}$ and $\overline{X(cp=4)}=\{2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rthal(X|cp=1)=\frac{1}{1}=1,$$

$$Rthal(X|cp=2)=\frac{0}{6}=0,$$

$$Rthal(X|cp=3)=\frac{0}{6}=0,$$

$$Rthal(X|cp=4)=\frac{3}{9}=0.3333.$$

Mean roughness

$$Rough_{thal}(cp)=\frac{1+0+0+0.3333}{4}=0.3333.$$

7) With respect to *thalach*

The lower and upper approximations are

$\underline{X(cp=1)}=\phi$ and $\overline{X(cp=1)}=\{1,10\}$,

$\underline{X(cp=2)}=\{5,6\}$ and $\overline{X(cp=2)}=\{5,6\}$,

$\underline{X}(cp=3)=\{4\}$ and $\overline{X}(cp=3)=\{4\}$,

$\underline{X}(cp=4)=\{2,3,7,8,9\}$ and $\overline{X}(cp=4)=\{1,2,3,7,8,9,10\}$.

Roughness

$$Rthalach(X|cp=1)=\frac{0}{2}=0,$$

$$Rthalach(X|cp=2)=\frac{2}{2}=1,$$

$$Rthalach(X|cp=3)=\frac{1}{1}=1,$$

$$Rthalach(X|cp=4)=\frac{5}{7}=0.7143.$$

Mean roughness

$$Rough_{thalach}(cp)=\frac{0+1+1+0.7143}{4}=0.6786.$$

8) With respect to $exia$

The lower and upper approximations are

$\underline{X}(cp=1)=\phi$ and $\overline{X}(cp=1)=\{1,4,5,6,7,9\}$,

$\underline{X}(cp=2)=\phi$ and $\overline{X}(cp=2)=\{1,4,5,6,7,9\}$,

$\underline{X}(cp=3)=\phi$ and $\overline{X}(cp=3)=\{1,4,5,6,7,9\}$,

$\underline{X}(cp=4)=\{2,3,8,10\}$ and $\overline{X}(cp=4)=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$\mathrm{Re}\,xia(X|cp=1)=\frac{0}{6}=0,$$

$$\mathrm{Re}\,xia(X|cp=2)=\frac{0}{6}=0,$$

$$\mathrm{Re}\,xia(X|cp=3)=\frac{0}{6}=0,$$

$$\mathrm{Re}\,xia(X|cp=4)=\frac{4}{10}=0.4.$$

Mean roughness

$$Rough_{exia}(cp) = \frac{0+0+0+0.4}{4} = 0.1.$$

9) With respect to *oldpeak*

The lower and upper approximations are

$\underline{X(cp=1)} = \{1\}$ and $\overline{X(cp=1)} = \{1\}$,

$\underline{X(cp=2)} = \{6\}$ and $\overline{X(cp=2)} = \{5,6,9\}$,

$\underline{X(cp=3)} = \{4\}$ and $\overline{X(cp=3)} = \{4\}$,

$\underline{X(cp=4)} = \{2,3,7,8,10\}$ and $\overline{X(cp=4)} = \{2,3,5,7,8,9,10\}$.

Roughness

$$Roldpeak(X|cp=1) = \frac{1}{1} = 1,$$

$$Roldpeak(X|cp=2) = \frac{1}{3} = 0.3333,$$

$$Roldpeak(X|cp=3) = \frac{1}{1} = 1,$$

$$Roldpeak(X|cp=4) = \frac{5}{7} = 0.7143.$$

Mean roughness

$$Rough_{oldpeak}(cp) = \frac{0+0.3333+1+0.7143}{4} = 0.7619.$$

d. **Attribute** *rbp*

For attribute rbp, it is clear that $|V(rbp)| = 4$.

The roughness and the mean roughness on rbp with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to *age*

The lower and upper approximations are

$\underline{X(rbp=120)} = \phi$ and $\overline{X(rbp=120)} = \{1,2,3,6,7,8,9,10\}$,

$\underline{X}(rbp = 130) = \{4,5\}$ and $\overline{X}(rbp = 130) = \{1,2,3,4,5,7,9\}$,

$\underline{X}(rbp = 140) = \phi$ and $\overline{X}(rbp = 140) = \{1,2,3,6,7,8,9,10\}$,

$\underline{X}(rbp = 160) = \phi$ and $\overline{X}(rbp = 160) = \{1,2,3,7,9\}$.

Roughness

$$Rage(X|rbp = 120) = \frac{0}{8} = 0,$$

$$Rage(X|rbp = 130) = \frac{2}{7} = 0.2857,$$

$$Rage(X|rbp = 140) = \frac{0}{8} = 0,$$

$$Rage(X|rbp = 160) = \frac{0}{5} = 0.$$

Mean roughness

$$Rough_{age}(rbp) = \frac{0 + 0.2857 + 0 + 0}{4} = 0.07143.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X}(rbp = 120) = \phi$ and $\overline{X}(rbp = 120) = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(rbp = 130) = \phi$ and $\overline{X}(rbp = 130) = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(rbp = 140) = \phi$ and $\overline{X}(rbp = 140) = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(rbp = 160) = \phi$ and $\overline{X}(rbp = 160) = \{1,2,3,4,6,9,10\}$.

Roughness

$$Rsex(X|rbp = 120) = \frac{0}{10} = 0,$$

$$Rsex(X|rbp = 130) = \frac{0}{10} = 0,$$

$$Rsex(X|rbp = 140) = \frac{0}{10} = 0,$$

$$Rsex(X|rbp = 160) = \frac{0}{7} = 0.$$

Mean roughness

$$Rough_{cp}(rbp) = \frac{0+0+0+0}{4} = 0\,.$$

3) With respect to cp

The lower and upper approximations are

$\underline{X(rbp=120)} = \phi$ and $\overline{X(rbp=120)} = \{2,3,5,6,7,8,9,10\}$,

$\underline{X(rbp=130)} = \{4\}$ and $\overline{X(rbp=130)} = \{2,3,4,5,6,7,8,9,10\}$,

$\underline{X(rbp=140)} = \{1\}$ and $\overline{X(rbp=140)} = \{1,2,3,7,8,9,10\}$,

$\underline{X(rbp=160)} = \phi$ and $\overline{X(rbp=160)} = \{2,3,7,8,9,10\}$.

Roughness

$$Rcp(X|rbp=120) = \frac{0}{8} = 0\,,$$

$$Rcp(X|rbp=130) = \frac{1}{9} = 0.1111\,,$$

$$Rcp(X|rbp=140) = \frac{1}{7} = 0.1429\,,$$

$$Rcp(X|rbp=160) = \frac{0}{6} = 0\,.$$

Mean roughness

$$Rough_{cp}(rbp) = \frac{0+0.1111+0.1429+0}{4} = 0.0635\,.$$

4) With respect to cho

The lower and upper approximations are

$\underline{X(rbp=120)} = \{3,8\}$ and $\overline{X(rbp=120)} = \{1,3,6,8\}$,

$\underline{X(rbp=130)} = \{4,9\}$ and $\overline{X(rbp=130)} = \{4,5,9,10\}$,

$\underline{X(rbp=140)} = \{7\}$ and $\overline{X(rbp=140)} = \{1,5,6,7,10\}$,

$\underline{X(rbp=160)} = \{2\}$ and $\overline{X(rbp=160)} = \{2\}$.

Roughness

$$Rcho(X|rbp=120) = \frac{2}{4} = 0.5\,,$$

$$Rcho\left(X\middle|rbp=130\right)=\frac{2}{4}=0.5,$$

$$Rcho\left(X\middle|rbp=140\right)=\frac{1}{5}=0.2,$$

$$Rcho\left(X\middle|rbp=160\right)=\frac{1}{1}=1.$$

Mean roughness

$$Rough_{cho}(rbp)=\frac{0.5+0.5+0.2+1}{4}=0.55.$$

5) With respect to fbs

The lower and upper approximations are

$\underline{X(rbp=120)}=\phi$ and $\overline{X(rbp=120)}=\{2,3,4,5,6,7,8,9\}$,

$\underline{X(rbp=130)}=\phi$ and $\overline{X(rbp=130)}=\{2,3,4,5,6,7,8,9\}$,

$\underline{X(rbp=140)}=\{1,10\}$ and $\overline{X(rbp=140)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(rbp=160)}=\phi$ and $\overline{X(rbp=160)}=\{2,3,4,5,6,7,8,9\}$.

Roughness

$$Rfbs\left(X\middle|rbp=120\right)=\frac{0}{8}=0,$$

$$Rfbs\left(X\middle|rbp=130\right)=\frac{0}{8}=0,$$

$$Rfbs\left(X\middle|rbp=140\right)=\frac{2}{10}=0.2,$$

$$Rfbs\left(X\middle|rbp=160\right)=\frac{0}{8}=0.$$

Mean roughness

$$Rough_{fbs}(rbp)=\frac{0+0+0.2+0}{4}=0.05.$$

6) With respect to $thal$

The lower and upper approximations are

$\underline{X(rbp=120)}=\phi$ and $\overline{X(rbp=120)}=\{2,3,4,5,6,7,8,9,10\}$,

$\underline{X(rbp=130)}=\phi$ and $\overline{X(rbp=130)}=\{2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(rbp=140)=\{1\}$ and $\overline{X}(rbp=140)=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(rbp=160)=\phi$ and $\overline{X}(rbp=160)=\{2,4,5,6,7,8\}$.

Roughness

$$Rthal(X|rbp=120)=\frac{0}{9}=0,$$

$$Rthal(X|rbp=130)=\frac{0}{9}=0,$$

$$Rthal(X|rbp=140)=\frac{1}{10}=0.1,$$

$$Rthal(X|rbp=160)=\frac{0}{6}=0.$$

Mean roughness

$$Rough_{thal}(rbp)=\frac{0+0+0.1+0}{4}=0.025.$$

7) With respect to *thalach*

The lower and upper approximations are

$\underline{X}(rbp=120)=\{3\}$ and $\overline{X}(rbp=120)=\{2,5,6,7,8\}$,

$\underline{X}(rbp=130)=\{4,9\}$ and $\overline{X}(rbp=130)=\{4,5,6,9\}$,

$\underline{X}(rbp=140)=\{1,10\}$ and $\overline{X}(rbp=140)=\{1,7,8,10\}$,

$\underline{X}(rbp=160)=\{2\}$ and $\overline{X}(rbp=160)=\{2\}$.

Roughness

$$Rthalach(X|rbp=120)=\frac{1}{5}=0.2,$$

$$Rthalach(X|rbp=130)=\frac{2}{4}=0.5,$$

$$Rthalach(X|rbp=140)=\frac{2}{4}=0.5,$$

$$Rthalach(X|rbp=160)=\frac{1}{1}=1.$$

Mean roughness

$$Rough_{thalach}(rbp)=\frac{0.2+0.5+0.5+1}{4}=0.55.$$

8) With respect to $exia$

The lower and upper approximations are

$\underline{X(rbp=120)}=\phi$ and $\overline{X(rbp=120)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(rbp=130)}=\phi$ and $\overline{X(rbp=130)}=\{1,4,5,6,7,9\}$,

$\underline{X(rbp=140)}=\phi$ and $\overline{X(rbp=140)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(rbp=160)}=\phi$ and $\overline{X(rbp=160)}=\{2,3,8,10\}$.

Roughness

$$\mathrm{R}exia(X|rbp=120)=\frac{0}{10}=0,$$

$$\mathrm{R}exia(X|rbp=130)=\frac{0}{6}=0,$$

$$\mathrm{R}exia(X|rbp=140)=\frac{0}{10}=0,$$

$$\mathrm{R}exia(X|rbp=160)=\frac{0}{4}=0.$$

Mean roughness

$$Rough_{exia}(rbp)=\frac{0+0+0+0}{4}=0.$$

9) With respect to $oldpeak$

The lower and upper approximations are

$\underline{X(rbp=120)}=\{3,6,8\}$ and $\overline{X(rbp=120)}=\{3,6,8\}$,

$\underline{X(rbp=130)}=\{4,5,9\}$ and $\overline{X(rbp=130)}=\{4,5,9\}$,

$\underline{X(rbp=140)}=\{1,7,10\}$ and $\overline{X(rbp=140)}=\{1,7,10\}$,

$\underline{X(rbp=160)}=\{2\}$ and $\overline{X(rbp=160)}=\{2\}$.

Roughness

$$Roldpeak(X|rbp=120)=\frac{3}{3}=1,$$

$$Roldpeak\left(X\,|rbp=130\right)=\frac{3}{3}=1,$$

$$Roldpeak\left(X\,|rbp=140\right)=\frac{3}{3}=1,$$

$$Roldpeak\left(X\,|rbp=160\right)=\frac{1}{1}=1.$$

Mean roughness

$$Rough_{oldpeak}(rbp)=\frac{1+1+1+1}{4}=1.$$

e. **Attribute** cho

For attribute cho, it is clear that $|V(cho)|=7$.

The roughness and the mean roughness on cho with respect to A, $A=age,sex,cp,rbp,cho,fbs,thal,thalach,exia,oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(cho=200)}=\{5\}$ and $\overline{X(cho=200)}=\{5,6,8,10\}$,

$\underline{X(cho=220)}=\phi$ and $\overline{X(cho=220)}=\{1,2,3,7,9\}$,

$\underline{X(cho=230)}=\phi$ and $\overline{X(cho=230)}=\{1,2,3,6,7,8,9,10\}$,

$\underline{X(cho=250)}=\{4\}$ and $\overline{X(cho=250)}=\{1,2,3,4,7,9\}$,

$\underline{X(cho=260)}=\phi$ and $\overline{X(cho=260)}=\{1,2,3,7,9\}$,

$\underline{X(cho=280)}=\phi$ and $\overline{X(cho=280)}=\{1,2,3,7,9\}$,

$\underline{X(cho=350)}=\phi$ and $\overline{X(cho=350)}=\{6,8,10\}$.

Roughness

$$Rage\left(X\,|cho=200\right)=\frac{1}{4}=0.25,$$

$$Rage\left(X\,|cho=220\right)=\frac{0}{5}=0,$$

$$Rage(X|cho=230)=\frac{0}{8}=0,$$

$$Rage(X|cho=250)=\frac{1}{6}=0.1667,$$

$$Rage(X|cho=260)=\frac{0}{5}=0,$$

$$Rage(X|cho=280)=\frac{0}{5}=0,$$

$$Rage(X|cho=350)=\frac{0}{3}=0.$$

Mean roughness

$$Rough_{age}(cho)=\frac{0.25+0+0+0.1667+0+0+0}{7}=0.05953.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X(cho=200)}=\phi$ and $\overline{X(cho=200)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(cho=220)}=\phi$ and $\overline{X(cho=220)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(cho=230)}=\phi$ and $\overline{X(cho=230)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(cho=250)}=\phi$ and $\overline{X(cho=250)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(cho=260)}=\phi$ and $\overline{X(cho=260)}=\{5,7,8\}$,

$\underline{X(cho=280)}=\phi$ and $\overline{X(cho=280)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(cho=350)}=\phi$ and $\overline{X(cho=350)}=\{5,7,8\}$.

Roughness

$$Rsex(X|cho=200)=\frac{0}{10}=0,$$

$$Rsex(X|cho=220)=\frac{0}{7}=0,$$

$$Rsex(X|cho=230)=\frac{0}{7}=0,$$

$$Rsex\left(X\middle|cho=250\right)=\frac{0}{7}=0,$$

$$Rsex\left(X\middle|cho=260\right)=\frac{0}{3}=0,$$

$$Rsex\left(X\middle|cho=280\right)=\frac{0}{7}=0,$$

$$Rsex\left(X\middle|cho=350\right)=\frac{0}{3}=0.$$

Mean roughness

$$Rough_{cp}(cho)=\frac{0+0+0+0+0+0+0}{7}=0.$$

3) With respect to cp

The lower and upper approximations are

$\underline{X(cho=200)}=\phi$ and $\overline{X(cho=200)}=\{2,3,5,6,7,8,9,10\}$,

$\underline{X(cho=220)}=\phi$ and $\overline{X(cho=220)}=\{2,3,7,8,9,10\}$,

$\underline{X(cho=230)}=\{1\}$ and $\overline{X(cho=230)}=\{1,5,6\}$,

$\underline{X(cho=250)}=\{4\}$ and $\overline{X(cho=250)}=\{2,3,4,7,8,9,10\}$,

$\underline{X(cho=260)}=\phi$ and $\overline{X(cho=260)}=\{2,3,7,8,9,10\}$,

$\underline{X(cho=280)}=\phi$ and $\overline{X(cho=280)}=\{2,3,7,8,9,10\}$,

$\underline{X(cho=350)}=\phi$ and $\overline{X(cho=350)}=\{2,3,7,8,9,10\}$.

Roughness

$$Rcp\left(X\middle|cho=200\right)=\frac{0}{8}=0,$$

$$Rcp\left(X\middle|cho=220\right)=\frac{0}{6}=0,$$

$$Rcp\left(X\middle|cho=230\right)=\frac{1}{3}=0.3333,$$

$$Rcp\left(X\middle|cho=250\right)=\frac{1}{7}=0.1429,$$

$$Rcp\left(X\middle|cho=260\right)=\frac{0}{6}=0,$$

$$Rcp(X|cho = 280) = \frac{0}{6} = 0,$$

$$Rcp(X|cho = 350) = \frac{0}{6} = 0.$$

Mean roughness

$$Rough_{cp}(cho) = \frac{0+0+0.3333+0.1429+0+0+0}{7} = 0.06803.$$

4) With respect to rbp

The lower and upper approximations are

$\underline{X(cho=200)} = \phi$ and $\overline{X(cho=200)} = \{1,4,5,7,9,10\}$,

$\underline{X(cho=220)} = \phi$ and $\overline{X(cho=220)} = \{3,6,8\}$,

$\underline{X(cho=230)} = \phi$ and $\overline{X(cho=230)} = \{1,3,6,7,8,10\}$,

$\underline{X(cho=250)} = \phi$ and $\overline{X(cho=250)} = \{4,5,9\}$,

$\underline{X(cho=260)} = \phi$ and $\overline{X(cho=260)} = \{1,7,10\}$,

$\underline{X(cho=280)} = \{2\}$ and $\overline{X(cho=280)} = \{2\}$,

$\underline{X(cho=350)} = \phi$ and $\overline{X(cho=350)} = \{3,6,8\}$.

Roughness

$$Rrbp(X|cho = 200) = \frac{0}{6} = 0,$$

$$Rrbp(X|cho = 220) = \frac{0}{3} = 0,$$

$$Rrbp(X|cho = 230) = \frac{0}{6} = 0,$$

$$Rrbp(X|cho = 250) = \frac{0}{3} = 0,$$

$$Rrbp(X|cho = 260) = \frac{0}{3} = 0,$$

$$Rrbp(X|cho = 280) = \frac{1}{1} = 1,$$

$$Rrbp(X|cho = 350) = \frac{0}{3} = 0.$$

Mean roughness

$$Rough_{rbp}(cho) = \frac{0+0+0+0+0+1+0}{7} = 0.1429.$$

5) With respect to fbs

The lower and upper approximations are

$\underline{X(cho=200)} = \phi$ and $\overline{X(cho=200)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(cho=220)} = \phi$ and $\overline{X(cho=220)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(cho=230)} = \phi$ and $\overline{X(cho=230)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(cho=250)} = \phi$ and $\overline{X(cho=250)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(cho=260)} = \phi$ and $\overline{X(cho=260)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(cho=280)} = \phi$ and $\overline{X(cho=280)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(cho=350)} = \phi$ and $\overline{X(cho=350)} = \{2,3,4,5,6,7,8,9\}$.

Roughness

$$Rfbs(X|cho=200) = \frac{0}{10} = 0,$$

$$Rfbs(X|cho=220) = \frac{0}{8} = 0,$$

$$Rfbs(X|cho=230) = \frac{0}{10} = 0,$$

$$Rfbs(X|cho=250) = \frac{0}{8} = 0,$$

$$Rfbs(X|cho=260) = \frac{0}{8} = 0,$$

$$Rfbs(X|cho=280) = \frac{0}{8} = 0,$$

$$Rfbs(X|cho=350) = \frac{0}{8} = 0.$$

Mean roughness

$$Rough_{fbs}(cho) = \frac{0+0+0+0+0+0+0}{7} = 0.$$

6) With respect to $thal$

The lower and upper approximations are

$$\underline{X}(cho=200)=\phi \text{ and } \overline{X}(cho=200)=\{2,3,4,5,6,7,8,9,10\},$$

$$\underline{X}(cho=220)=\phi \text{ and } \overline{X}(cho=220)=\{3,9,10\},$$

$$\underline{X}(cho=230)=\{1\} \text{ and } \overline{X}(cho=230)=\{1,2,4,5,6,7,8\},$$

$$\underline{X}(cho=250)=\phi \text{ and } \overline{X}(cho=250)=\{2,3,4,5,6,7,8,9,10\},$$

$$\underline{X}(cho=260)=\phi \text{ and } \overline{X}(cho=260)=\{2,4,5,6,7,8\},$$

$$\underline{X}(cho=280)=\phi \text{ and } \overline{X}(cho=280)=\{2,4,5,6,7,8\},$$

$$\underline{X}(cho=350)=\phi \text{ and } \overline{X}(cho=350)=\{2,4,5,6,7,8\}.$$

Roughness

$$Rthal(X|cho=200)=\frac{0}{9}=0,$$

$$Rthal(X|cho=220)=\frac{0}{3}=0,$$

$$Rthal(X|cho=230)=\frac{1}{7}=0.1429,$$

$$Rthal(X|cho=250)=\frac{0}{9}=0,$$

$$Rthal(X|cho=260)=\frac{0}{6}=0,$$

$$Rthal(X|cho=280)=\frac{0}{6}=0,$$

$$Rthal(X|cho=350)=\frac{0}{6}=0.$$

Mean roughness

$$Rough_{thal}(cho)=\frac{0+0+0.1429+0+0+0+0}{7}=0.02041.$$

7) With respect to $thalach$

The lower and upper approximations are

$\underline{X}(cho=200)=\phi$ and $\overline{X}(cho=200)=\{1,5,6,10\}$,

$\underline{X}(cho=220)=\{3\}$ and $\overline{X}(cho=220)=\{3\}$,

$\underline{X}(cho=230)=\phi$ and $\overline{X}(cho=230)=\{1,5,6,10\}$,

$\underline{X}(cho=250)=\{4,9\}$ and $\overline{X}(cho=250)=\{4,9\}$,

$\underline{X}(cho=260)=\phi$ and $\overline{X}(cho=260)=\{7,8\}$,

$\underline{X}(cho=280)=\{2\}$ and $\overline{X}(cho=280)=\{2\}$,

$\underline{X}(cho=350)=\phi$ and $\overline{X}(cho=350)=\{7,8\}$.

Roughness

$$Rthalach(X|cho=200)=\frac{0}{4}=0,$$

$$Rthalach(X|cho=220)=\frac{1}{1}=1,$$

$$Rthalach(X|cho=230)=\frac{0}{4}=0,$$

$$Rthalach(X|cho=250)=\frac{2}{2}=1,$$

$$Rthalach(X|cho=260)=\frac{0}{2}=0,$$

$$Rthalach(X|cho=280)=\frac{1}{1}=1,$$

$$Rthalach(X|cho=350)=\frac{0}{2}=0.$$

Mean roughness

$$Rough_{thalach}(cho)=\frac{0+1+0+1+0+1+0}{7}=0.4286.$$

8) With respect to *exia*

The lower and upper approximations are

$\underline{X}(cho=200)=\phi$ and $\overline{X}(cho=200)=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(cho=220)=\phi$ and $\overline{X}(cho=220)=\{2,3,8,10\}$,

$\underline{X}(cho=230)=\phi$ and $\overline{X}(cho=230)=\{1,4,5,6,7,9\}$,

$\underline{X}(cho=250)=\phi$ and $\overline{X}(cho=250)=\{1,4,5,6,7,9\}$,

$\underline{X}(cho=260)=\phi$ and $\overline{X}(cho=260)=\{1,4,5,6,7,9\}$,

$\underline{X}(cho=280)=\phi$ and $\overline{X}(cho=280)=\{2,3,8,10\}$,

$\underline{X}(cho=350)=\phi$ and $\overline{X}(cho=350)=\{2,3,8,10\}$.

Roughness

$$\operatorname{Re}xia(X|cho=200)=\frac{0}{10}=0,$$

$$\operatorname{Re}xia(X|cho=220)=\frac{0}{4}=0,$$

$$\operatorname{Re}xia(X|cho=230)=\frac{0}{6}=0,$$

$$\operatorname{Re}xia(X|cho=250)=\frac{0}{6}=0,$$

$$\operatorname{Re}xia(X|cho=260)=\frac{0}{6}=0,$$

$$\operatorname{Re}xia(X|cho=280)=\frac{0}{4}=0,$$

$$\operatorname{Re}xia(X|cho=350)=\frac{0}{4}=0.$$

Mean roughness

$$Rough_{exia}(cho)=\frac{0+0+0+0+0+0+0}{7}=0.$$

9) With respect to $oldpeak$

The lower and upper approximations are

$\underline{X}(cho=200)=\{10\}$ and $\overline{X}(cho=200)=\{5,9,10\}$,

$\underline{X}(cho=220)=\{3\}$ and $\overline{X}(cho=220)=\{3\}$,

$\underline{X}(cho=230)=\{1,6\}$ and $\overline{X}(cho=230)=\{1,6\}$,

$\underline{X}(cho=250)=\{4\}$ and $\overline{X}(cho=250)=\{4,5,9\}$,

$\underline{X}(cho=260)=\{7\}$ and $\overline{X}(cho=260)=\{7\}$,

$\underline{X}(cho=280)=\{2\}$ and $\overline{X}(cho=280)=\{2\}$,

$\underline{X}(cho = 350) = \{8\}$ and $\overline{X}(cho = 350) = \{8\}$.

Roughness

$$Roldpeak(X|cho = 200) = \frac{1}{3} = 0.3333,$$

$$Roldpeak(X|cho = 220) = \frac{1}{1} = 1,$$

$$Roldpeak(X|cho = 230) = \frac{2}{2} = 1,$$

$$Roldpeak(X|cho = 250) = \frac{1}{3} = 0.3333,$$

$$Roldpeak(X|cho = 260) = \frac{1}{1} = 1,$$

$$Roldpeak(X|cho = 280) = \frac{1}{1} = 1,$$

$$Roldpeak(X|cho = 350) = \frac{1}{1} = 1.$$

Mean roughness

$$Rough_{oldpeak}(cho) = \frac{0.3333 + 1 + 1 + 0.3333 + 1 + 1 + 1}{7} = 0.8095.$$

f. **Attribute** *fbs*

For attribute fbs, it is clear that $|V(fbs)| = 2$.

The roughness and the mean roughness on fbs with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X}(fbs = 1) = \phi$ and $\overline{X}(fbs = 1) = \{1,2,3,6,7,8,9,10\}$,

$\underline{X}(fbs = 2) = \{4,5\}$ and $\overline{X}(fbs = 2) = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rage\left(X \mid fbs=1\right)=\frac{0}{8}=0\,,$$

$$Rage\left(X \mid fbs=2\right)=\frac{2}{10}=0.2\,.$$

Mean roughness

$$Rough_{age}\left(fbs\right)=\frac{0+0.2}{2}=0.1\,.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X\left(fbs=1\right)}=\phi$ and $\overline{X\left(fbs=1\right)}=\{1,2,3,4,6,9,10\}$,

$\underline{X\left(fbs=2\right)}=\{5,7,8\}$ and $\overline{X\left(fbs=2\right)}=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rsex\left(X \mid fbs=1\right)=\frac{0}{7}=0\,,$$

$$Rsex\left(X \mid fbs=2\right)=\frac{3}{10}=0.3\,.$$

Mean roughness

$$Rough_{cp}\left(fbs\right)=\frac{0+0.3}{2}=0.15\,.$$

3) With respect to cp

The lower and upper approximations are

$\underline{X\left(fbs=1\right)}=\{1\}$ and $\overline{X\left(fbs=1\right)}=\{1,2,3,7,8,9,10\}$,

$\underline{X\left(fbs=2\right)}=\{4,5,6\}$ and $\overline{X\left(fbs=2\right)}=\{2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rcp\left(X \mid fbs=1\right)=\frac{1}{7}=0.1429\,,$$

$$Rcp\left(X \mid fbs=2\right)=\frac{3}{9}=0.3333\,.$$

Mean roughness

$$Rough_{cp}\left(fbs\right)=\frac{0.1429+0.3333}{2}=0.2381\,.$$

4) With respect to rbp

The lower and upper approximations are

$\underline{X(fbs=1)} = \phi$ and $\overline{X(fbs=1)} = \{1,7,10\}$,

$\underline{X(fbs=2)} = \{2,3,4,5,6,8,9\}$ and $\overline{X(fbs=2)} = \{1,2,3,4,5,6,7,8,9,10\}$

Roughness

$$Rrbp(X \mid fbs=1) = \frac{0}{3} = 0,$$

$$Rrbp(X \mid fbs=2) = \frac{7}{10} = 0.7.$$

Mean roughness

$$Rough_{rbp}(fbs) = \frac{0+0.7}{2} = 0.35.$$

5) With respect to cho

The lower and upper approximations are

$\underline{X(fbs=1)} = \phi$ and $\overline{X(fbs=1)} = \{1,5,6,10\}$,

$\underline{X(fbs=2)} = \{2,3,4,7,8,9\}$ and $\overline{X(fbs=2)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rcho(X \mid fbs=1) = \frac{0}{4} = 0,$$

$$Rcho(X \mid fbs=2) = \frac{6}{10} = 0.6.$$

Mean roughness

$$Rough_{cho}(fbs) = \frac{0+0.6}{2} = 0.3.$$

6) With respect to $thal$

The lower and upper approximations are

$\underline{X(fbs=1)} = \{1\}$ and $\overline{X(fbs=1)} = \{1,3,9,10\}$,

$\underline{X(fbs=2)} = \{2,4,5,6,7,8\}$ and $\overline{X(fbs=2)} = \{2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rthal\left(X \middle| fbs = 1\right) = \frac{1}{4} = 0.25,$$

$$Rthal\left(X \middle| fbs = 2\right) = \frac{6}{9} = 0.6667.$$

Mean roughness

$$Rough_{thal}\left(fbs\right) = \frac{0.25 + 0.6667}{2} = 0.4583.$$

7) With respect to *thalach*

The lower and upper approximations are

$\underline{X\left(fbs = 1\right)} = \{1{,}10\}$ and $\overline{X\left(fbs = 1\right)} = \{1{,}10\}$,

$\underline{X\left(fbs = 2\right)} = \{2,3,4,5,6,7,8,9\}$ and $\overline{X\left(fbs = 2\right)} = \{2,3,4,5,6,7,8,9\}$.

Roughness

$$Rthalach\left(X \middle| fbs = 1\right) = \frac{2}{2} = 1,$$

$$Rthalach\left(X \middle| fbs = 2\right) = \frac{8}{8} = 1.$$

Mean roughness

$$Rough_{thalach}\left(fbs\right) = \frac{1+1}{2} = 1.$$

8) With respect to *exia*

The lower and upper approximations are

$\underline{X\left(fbs = 1\right)} = \phi$ and $\overline{X\left(fbs = 1\right)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X\left(fbs = 2\right)} = \phi$ and $\overline{X\left(fbs = 2\right)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$\mathrm{Re}\,xia\left(X \middle| fbs = 1\right) = \frac{0}{10} = 0,$$

$$\mathrm{Re}\,xia\left(X \middle| fbs = 2\right) = \frac{0}{10} = 0.$$

Mean roughness

$$Rough_{exia}\left(fbs\right) = \frac{0+0}{2} = 0.$$

9) With respect to *oldpeak*

The lower and upper approximations are

$\underline{X(fbs=1)}=\{1,10\}$ and $\overline{X(fbs=1)}=\{1,10\}$,

$\underline{X(fbs=2)}=\{2,3,4,5,6,7,8,9\}$ and $\overline{X(fbs=2)}=\{2,3,4,5,6,7,8,9\}$.

Roughness

$$Roldpeak(X|fbs=1)=\frac{2}{2}=1,$$

$$Roldpeak(X|fbs=2)=\frac{8}{8}=1.$$

Mean roughness

$$Rough_{oldpeak}(fbs)=\frac{1+1}{2}=1.$$

g. **Attribute** *thal*

For attribute $thal$, it is clear that $|V(thal)|=3$.

The roughness and the mean roughness on $thal$ with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(thal=3)}=\{4,5\}$ and $\overline{X(thal=3)}=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(thal=6)}=\phi$ and $\overline{X(thal=6)}=\{1,2,3,7,9\}$,

$\underline{X(thal=7)}=\phi$ and $\overline{X(thal=7)}=\{1,2,3,6,7,8,9,10\}$.

Roughness

$$Rage(X|thal=3)=\frac{2}{10}=0.2,$$

$$Rage(X|thal=6)=\frac{0}{5}=0,$$

$$Rage(X|thal=7)=\frac{0}{8}=0.$$

Mean roughness

$$Rough_{age}(thal) = \frac{0.2+0+0}{3} = 0.06667.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X(thal=3)} = \{5,7,8\}$ and $\overline{X(thal=3)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(thal=6)} = \phi$ and $\overline{X(thal=6)} = \{1,2,3,4,6,9,10\}$,

$\underline{X(thal=7)} = \phi$ and $\overline{X(thal=7)} = \{1,2,3,4,6,9,10\}$.

Roughness

$$Rsex(X|thal=3) = \frac{3}{10} = 0.3,$$

$$Rsex(X|thal=6) = \frac{0}{7} = 0,$$

$$Rsex(X|thal=7) = \frac{0}{7} = 0.$$

Mean roughness

$$Rough_{sex}(thal) = \frac{0.3+0+0}{3} = 0.1.$$

3) With respect to cp

The lower and upper approximations are

$\underline{X(thal=3)} = \{4,5,6\}$ and $\overline{X(thal=3)} = \{2,3,4,5,6,7,8,9,10\}$,

$\underline{X(thal=6)} = \{1\}$ and $\overline{X(thal=6)} = \{1\}$,

$\underline{X(thal=7)} = \phi$ and $\overline{X(thal=7)} = \{2,3,7,8,9,10\}$.

Roughness

$$Rcp(X|thal=3) = \frac{3}{9} = 0.3333,$$

$$Rcp(X|thal=6) = \frac{1}{1} = 1,$$

$$Rcp(X|thal=7) = \frac{0}{6} = 0.$$

Mean roughness

$$Rough_{cp}(thal) = \frac{0.3333 + 1 + 0}{3} = 0.4444.$$

4) With respect to rbp

The lower and upper approximations are

$\underline{X(thal = 3)} = \{2\}$ and $\overline{X(thal = 3)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(thal = 6)} = \phi$ and $\overline{X(thal = 6)} = \{1,7,10\}$,

$\underline{X(thal = 7)} = \phi$ and $\overline{X(thal = 7)} = \{1,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rrbp(X|thal = 3) = \frac{1}{10} = 0.1,$$

$$Rrbp(X|thal = 6) = \frac{0}{3} = 0,$$

$$Rrbp(X|thal = 7) = \frac{0}{9} = 0.$$

Mean roughness

$$Rough_{rbp}(thal) = \frac{0.1 + 0 + 0}{3} = 0.03333.$$

5) With respect to cho

The lower and upper approximations are

$\underline{X(thal = 3)} = \{2,7,8\}$ and $\overline{X(thal = 3)} = \{1,2,4,5,6,7,8,9,10\}$,

$\underline{X(thal = 6)} = \phi$ and $\overline{X(thal = 6)} = \{1,6\}$,

$\underline{X(thal = 7)} = \{3\}$ and $\overline{X(thal = 7)} = \{3,4,5,9,10\}$.

Roughness

$$Rcho(X|thal = 3) = \frac{3}{9} = 0.3333,$$

$$Rcho(X|thal = 6) = \frac{0}{2} = 0,$$

$$Rcho(X|thal = 7) = \frac{1}{5} = 0.2.$$

Mean roughness

$$Rough_{cho}(thal) = \frac{0.3333 + 0 + 0.2}{3} = 0.1778\,.$$

6) With respect to fbs

The lower and upper approximations are

$\underline{X(thal = 3)} = \phi$ and $\overline{X(thal = 3)} = \{2,3,4,5,6,7,8,9\}$,

$\underline{X(thal = 6)} = \phi$ and $\overline{X(thal = 6)} = \{1,10\}$,

$\underline{X(thal = 7)} = \phi$ and $\overline{X(thal = 7)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rfbs(X|thal = 3) = \frac{0}{8} = 0\,,$$

$$Rfbs(X|thal = 6) = \frac{0}{2} = 0\,,$$

$$Rfbs(X|thal = 7) = \frac{0}{10} = 0\,.$$

Mean roughness

$$Rough_{fbs}(thal) = \frac{0 + 0 + 0}{3} = 0\,.$$

7) With respect to $thalach$

The lower and upper approximations are

$\underline{X(thal = 3)} = \{2,4,5,6,7,8\}$ and $\overline{X(thal = 3)} = \{2,4,5,6,7,8\}$,

$\underline{X(thal = 6)} = \phi$ and $\overline{X(thal = 6)} = \{1,10\}$,

$\underline{X(thal = 7)} = \{3,9\}$ and $\overline{X(thal = 7)} = \{1,3,9,10\}$.

Roughness

$$Rthalach(X|thal = 3) = \frac{6}{6} = 1\,,$$

$$Rthalach(X|thal = 6) = \frac{0}{2} = 0\,,$$

$$Rthalach(X|thal = 7) = \frac{2}{4} = 0.5\,.$$

Mean roughness

$$Rough_{thalach}(thal) = \frac{1+0+0.5}{3} = 0.5.$$

8) With respect to *exia*

The lower and upper approximations are

$\underline{X}(thal=3) = \phi$ and $\overline{X}(thal=3) = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(thal=6) = \phi$ and $\overline{X}(thal=6) = \{1,4,5,6,7,9\}$,

$\underline{X}(thal=7) = \phi$ and $\overline{X}(thal=7) = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$\mathrm{Re}\,xia(X|thal=3) = \frac{0}{10} = 0,$$

$$\mathrm{Re}\,xia(X|thal=6) = \frac{0}{6} = 0,$$

$$\mathrm{Re}\,xia(X|thal=7) = \frac{0}{10} = 0.$$

Mean roughness

$$Rough_{exia}(thal) = \frac{0+0+0}{3} = 0.$$

9) With respect to *oldpeak*

The lower and upper approximations are

$\underline{X}(thal=3) = \{2,4,6,7,8\}$ and $\overline{X}(thal=3) = \{2,4,5,6,7,8,9\}$,

$\underline{X}(thal=6) = \{1\}$ and $\overline{X}(thal=6) = \{1\}$,

$\underline{X}(thal=7) = \{3,10\}$ and $\overline{X}(thal=7) = \{3,5,9,10\}$.

Roughness

$$Roldpeak(X|thal=3) = \frac{5}{7} = 0.7143,$$

$$Roldpeak(X|thal=6) = \frac{1}{1} = 1,$$

$$Roldpeak(X|thal=7) = \frac{2}{4} = 0.5.$$

Mean roughness

$$Rough_{oldpeak}(thal) = \frac{0.7143 + 1 + 0.5}{3} = 0.7381.$$

h. **Attribute** *thalach*

For attribute $thalach$, it is clear that $|V(thalach)| = 7$.

The roughness and the mean roughness on $thalach$ with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(thalach = 100)} = \phi$ and $\overline{X(thalach = 100)} = \{1,2,3,7,9\}$,

$\underline{X(thalach = 120)} = \phi$ and $\overline{X(thalach = 120)} = \{1,2,3,7,9\}$,

$\underline{X(thalach = 140)} = \phi$ and $\overline{X(thalach = 140)} = \{1,2,3,7,9\}$,

$\underline{X(thalach = 150)} = \phi$ and $\overline{X(thalach = 150)} = \{1,2,3,6,7,8,9,10\}$,

$\underline{X(thalach = 160)} = \phi$ and $\overline{X(thalach = 160)} = \{1,2,3,6,7,8,9,10\}$,

$\underline{X(thalach = 170)} = \{5\}$ and $\overline{X(thalach = 170)} = \{5,6,8,10\}$,

$\underline{X(thalach = 180)} = \{4\}$ and $\overline{X(thalach = 180)} = \{4\}$.

Roughness

$$Rage(X|thalach = 100) = \frac{0}{5} = 0,$$

$$Rage(X|thalach = 120) = \frac{0}{5} = 0,$$

$$Rage(X|thalach = 140) = \frac{0}{5} = 0,$$

$$Rage(X|thalach = 150) = \frac{0}{8} = 0,$$

$$Rage(X|thalach = 160) = \frac{0}{8} = 0,$$

$$Rage(X|thalach = 170) = \frac{1}{4} = 0.25,$$

$$Rage(X|thalach=180)=\frac{1}{1}=1.$$

Mean roughness

$$Rough_{age}(thalach)=\frac{0+0+0+0+0+0.25+1}{7}=0.1786.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X(thalach=100)}=\phi$ and $\overline{X(thalach=100)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(thalach=120)}=\phi$ and $\overline{X(thalach=120)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(thalach=140)}=\phi$ and $\overline{X(thalach=140)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(thalach=150)}=\phi$ and $\overline{X(thalach=150)}=\{1,2,3,4,6,9,10\}$,

$\underline{X(thalach=160)}=\phi$ and $\overline{X(thalach=160)}=\{5,7,8\}$,

$\underline{X(thalach=170)}=\phi$ and $\overline{X(thalach=170)}=\{1,2,3,4,5,6,7,8,9,10\}$

$\underline{X(thalach=180)}=\phi$ and $\overline{X(thalach=180)}=\{1,2,3,4,6,9,10\}$.

Roughness

$$Rsex(X|thalach=100)=\frac{0}{7}=0,$$

$$Rsex(X|thalach=120)=\frac{0}{7}=0,$$

$$Rsex(X|thalach=140)=\frac{0}{7}=0,$$

$$Rsex(X|thalach=150)=\frac{0}{7}=0,$$

$$Rsex(X|thalach=160)=\frac{0}{3}=0,$$

$$Rsex(X|thalach=170)=\frac{0}{10}=0,$$

$$Rsex(X|thalach=180)=\frac{0}{7}=0.$$

Mean roughness

$$Rough_{cp}(thalach) = \frac{0+0+0+0+0+0+0}{7} = 0 .$$

3) With respect to cp

The lower and upper approximations are

$\underline{X(thalach=100)} = \phi$ and $\overline{X(thalach=100)} = \{2,3,7,8,9,10\}$,

$\underline{X(thalach=120)} = \phi$ and $\overline{X(thalach=120)} = \{2,3,7,8,9,10\}$,

$\underline{X(thalach=140)} = \phi$ and $\overline{X(thalach=140)} = \{2,3,7,8,9,10\}$,

$\underline{X(thalach=150)} = \{1\}$ and $\overline{X(thalach=150)} = \{1,2,3,7,8,9,10\}$,

$\underline{X(thalach=160)} = \phi$ and $\overline{X(thalach=160)} = \{2,3,7,8,9,10\}$,

$\underline{X(thalach=170)} = \{5,6\}$ and $\overline{X(thalach=170)} = \{5,6\}$,

$\underline{X(thalach=180)} = \{4\}$ and $\overline{X(thalach=180)} = \{4\}$.

Roughness

$$Rcp(X|thalach=100) = \frac{0}{6} = 0,$$

$$Rcp(X|thalach=120) = \frac{0}{6} = 0,$$

$$Rcp(X|thalach=140) = \frac{0}{6} = 0,$$

$$Rcp(X|thalach=150) = \frac{1}{7} = 0.1429,$$

$$Rcp(X|thalach=160) = \frac{0}{6} = 0,$$

$$Rcp(X|thalach=170) = \frac{2}{2} = 1,$$

$$Rcp(X|thalach=180) = \frac{1}{1} = 1.$$

Mean roughness

$$Rough_{cp}(thalach) = \frac{0+0+0+0.1429+0+1+1}{7} = 0.3061.$$

4) With respect to rbp

The lower and upper approximations are

$$\underline{X(thalach=100)}=\{2\} \text{ and } \overline{X(thalach=100)}=\{2\},$$

$$\underline{X(thalach=120)}=\phi \text{ and } \overline{X(thalach=120)}=\{3,6,8\},$$

$$\underline{X(thalach=140)}=\phi \text{ and } \overline{X(thalach=140)}=\{4,5,9\},$$

$$\underline{X(thalach=150)}=\phi \text{ and } \overline{X(thalach=150)}=\{1,7,10\},$$

$$\underline{X(thalach=160)}=\phi \text{ and } \overline{X(thalach=160)}=\{1,3,6,7,8,10\},$$

$$\underline{X(thalach=170)}=\phi \text{ and } \overline{X(thalach=170)}=\{3,4,5,6,8,9\},$$

$$\underline{X(thalach=180)}=\phi \text{ and } \overline{X(thalach=180)}=\{4,5,9\}.$$

Roughness

$$Rrbp(X|thalach=100)=\frac{1}{1}=1,$$

$$Rrbp(X|thalach=120)=\frac{0}{3}=0,$$

$$Rrbp(X|thalach=140)=\frac{0}{3}=0,$$

$$Rrbp(X|thalach=150)=\frac{0}{3}=0,$$

$$Rrbp(X|thalach=160)=\frac{0}{6}=0,$$

$$Rrbp(X|thalach=170)=\frac{0}{6}=0,$$

$$Rrbp(X|thalach=180)=\frac{0}{3}=0.$$

Mean roughness

$$Rough_{rbp}(thalach)=\frac{1+0+0+0+0+0+0}{7}=0.1429.$$

5) With respect to cho

The lower and upper approximations are

$$\underline{X(thalach=100)}=\{2\} \text{ and } \overline{X(thalach=100)}=\{2\},$$

$\underline{X}(thalach=120)=\{3\}$ and $\overline{X}(thalach=120)=\{3\}$,

$\underline{X}(thalach=140)=\phi$ and $\overline{X}(thalach=140)=\{4,9\}$,

$\underline{X}(thalach=150)=\phi$ and $\overline{X}(thalach=150)=\{1,5,6,10\}$,

$\underline{X}(thalach=160)=\{7,8\}$ and $\overline{X}(thalach=160)=\{7,8\}$,

$\underline{X}(thalach=170)=\phi$ and $\overline{X}(thalach=170)=\{1,5,6,10\}$,

$\underline{X}(thalach=180)=\phi$ and $\overline{X}(thalach=180)=\{4,9\}$.

Roughness

$$Rcho(X|thalach=100)=\frac{1}{1}=1,$$

$$Rcho(X|thalach=120)=\frac{1}{1}=1,$$

$$Rcho(X|thalach=140)=\frac{0}{2}=0,$$

$$Rcho(X|thalach=150)=\frac{0}{4}=0,$$

$$Rcho(X|thalach=160)=\frac{2}{2}=1,$$

$$Rcho(X|thalach=170)=\frac{0}{4}=0,$$

$$Rcho(X|thalach=180)=\frac{0}{2}=0.$$

Mean roughness

$$Rough_{cho}(thalach)=\frac{1+1+0+0+1+0+0}{7}=0.4286.$$

6) With respect to fbs

The lower and upper approximations are

$\underline{X}(thalach=100)=\phi$ and $\overline{X}(thalach=100)=\{2,3,4,5,6,7,8,9\}$,

$\underline{X}(thalach=120)=\phi$ and $\overline{X}(thalach=120)=\{2,3,4,5,6,7,8,9\}$,

$\underline{X}(thalach=140)=\phi$ and $\overline{X}(thalach=140)=\{2,3,4,5,6,7,8,9\}$,

$\underline{X}(thalach=150)=\{1,10\}$ and $\overline{X}(thalach=150)=\{1,10\}$,

$\underline{X}(thalach=160)=\phi$ and $\overline{X}(thalach=160)=\{2,3,4,5,6,7,8,9\}$,

$\underline{X}(thalach=170)=\phi$ and $\overline{X}(thalach=170)=\{2,3,4,5,6,7,8,9\}$,

$\underline{X}(thalach=180)=\phi$ and $\overline{X}(thalach=180)=\{2,3,4,5,6,7,8,9\}$.

Roughness

$$Rfbs(X|thalach=100)=\frac{0}{2}=0,$$

$$Rfbs(X|thalach=120)=\frac{0}{2}=0,$$

$$Rfbs(X|thalach=140)=\frac{0}{2}=0,$$

$$Rfbs(X|thalach=150)=\frac{2}{2}=1,$$

$$Rfbs(X|thalach=160)=\frac{0}{8}=0,$$

$$Rfbs(X|thalach=170)=\frac{0}{8}=0,$$

$$Rfbs(X|thalach=180)=\frac{0}{8}=0.$$

Mean roughness

$$Rough_{fbs}(thalach)=\frac{0+0+0+1+0+0+0}{7}=0.1429.$$

7) With respect to $thal$

The lower and upper approximations are

$\underline{X}(thalach=100)=\phi$ and $\overline{X}(thalach=100)=\{2,4,5,6,7,8\}$,

$\underline{X}(thalach=120)=\phi$ and $\overline{X}(thalach=120)=\{3,9,10\}$,

$\underline{X}(thalach=140)=\phi$ and $\overline{X}(thalach=140)=\{3,9,10\}$,

$\underline{X}(thalach=150)=\{1\}$ and $\overline{X}(thalach=150)=\{1,3,9,10\}$,

$\underline{X}(thalach=160)=\phi$ and $\overline{X}(thalach=160)=\{2,4,5,6,7,8\}$,

$\underline{X}(thalach=170)=\phi$ and $\overline{X}(thalach=170)=\{2,4,5,6,7,8\}$,

$\underline{X}(thalach=180)=\phi$ and $\overline{X}(thalach=180)=\{2,4,5,6,7,8\}$.

Roughness

$$Rthal(X|thalach=100)=\frac{0}{6}=0,$$

$$Rthal(X|thalach=120)=\frac{0}{3}=0,$$

$$Rthal(X|thalach=140)=\frac{0}{3}=0,$$

$$Rthal(X|thalach=150)=\frac{1}{4}=0.25,$$

$$Rthal(X|thalach=160)=\frac{0}{6}=0,$$

$$Rthal(X|thalach=170)=\frac{0}{6}=0,$$

$$Rthal(X|thalach=180)=\frac{0}{6}=0.$$

Mean roughness

$$Rough_{thal}(thalach)=\frac{0+0+0+0.25+0+0+0}{7}=0.03571.$$

8) With respect to $exia$

The lower and upper approximations are

$\underline{X(thalach=100)}=\phi$ and $\overline{X(thalach=100)}=\{2,3,8,10\}$,

$\underline{X(thalach=120)}=\phi$ and $\overline{X(thalach=120)}=\{2,3,8,10\}$,

$\underline{X(thalach=140)}=\phi$ and $\overline{X(thalach=140)}=\{1,4,5,6,7,9\}$,

$\underline{X(thalach=150)}=\phi$ and $\overline{X(thalach=150)}=\{1,2,3,4,5,6,7,8,9,10\}$

$\underline{X(thalach=160)}=\phi$ and $\overline{X(thalach=160)}=\{1,2,3,4,5,6,7,8,9,10\}$

$\underline{X(thalach=170)}=\phi$ and $\overline{X(thalach=170)}=\{1,4,5,6,7,9\}$,

$\underline{X(thalach=180)}=\phi$ and $\overline{X(thalach=180)}=\{1,4,5,6,7,9\}$.

Roughness

$$\mathrm{Re}\,xia(X|thalach=100)=\frac{0}{4}=0,$$

$$\operatorname{Re}xia\left(X\left|thalach=120\right.\right)=\frac{0}{4}=0\,,$$

$$\operatorname{Re}xia\left(X\left|thalach=140\right.\right)=\frac{0}{6}=0\,,$$

$$\operatorname{Re}xia\left(X\left|thalach=150\right.\right)=\frac{0}{10}=0\,,$$

$$\operatorname{Re}xia\left(X\left|thalach=160\right.\right)=\frac{0}{10}=0\,,$$

$$\operatorname{Re}xia\left(X\left|thalach=170\right.\right)=\frac{0}{6}=0\,,$$

$$\operatorname{Re}xia\left(X\left|thalach=180\right.\right)=\frac{0}{6}=0\,.$$

Mean roughness

$$Rough_{exia}(thalach)=\frac{0+0+0+0+0+0+0}{7}=0\,.$$

9) With respect to $oldpeak$

The lower and upper approximations are

$\underline{X(thalach=100)}=\{2\}$ and $\overline{X(thalach=100)}=\{2\}$,

$\underline{X(thalach=120)}=\{3\}$ and $\overline{X(thalach=120)}=\{3\}$,

$\underline{X(thalach=140)}=\phi$ and $\overline{X(thalach=140)}=\{5,9\}$,

$\underline{X(thalach=150)}=\{1,10\}$ and $\overline{X(thalach=150)}=\{1,10\}$,

$\underline{X(thalach=160)}=\{7,8\}$ and $\overline{X(thalach=160)}=\{7,8\}$,

$\underline{X(thalach=170)}=\{6\}$ and $\overline{X(thalach=170)}=\{5,6,9\}$,

$\underline{X(thalach=180)}=\{4\}$ and $\overline{X(thalach=180)}=\{4\}$.

Roughness

$$Roldpeak\left(X\left|thalach=100\right.\right)=\frac{1}{1}=1\,,$$

$$Roldpeak\left(X\left|thalach=120\right.\right)=\frac{1}{1}=1\,,$$

$$Roldpeak\left(X\left|thalach=140\right.\right)=\frac{0}{2}=0\,,$$

$$Roldpeak(X|thalach = 150) = \frac{2}{2} = 1,$$

$$Roldpeak(X|thalach = 160) = \frac{2}{2} = 1,$$

$$Roldpeak(X|thalach = 170) = \frac{1}{3} = 0.3333,$$

$$Roldpeak(X|thalach = 180) = \frac{1}{1} = 1.$$

Mean roughness

$$Rough_{oldpeak}(thalach) = \frac{1+1+0+1+1+0.3333+1}{7} = 0.7619.$$

i. **Attribute** $exia$

For attribute $exia$, it is clear that $|V(exia)| = 2$.

The roughness and the mean roughness on $exia$ with respect to A, $A = age, sex, cp, rbp, cho, fbs, thal, thalach, exia, oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(exia = 1)} = \phi$ and $\overline{X(exia = 1)} = \{1,2,3,6,7,8,9,10\}$,

$\underline{X(exia = 2)} = \{4,5\}$ and $\overline{X(exia = 2)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rage(X|exia = 1) = \frac{0}{8} = 0,$$

$$Rage(X|exia = 2) = \frac{2}{10} = 0.2.$$

Mean roughness

$$Rough_{age}(exia) = \frac{0+0.2}{2} = 0.1.$$

2) With respect to sex

The lower and upper approximations are

$\underline{X}(exia=1)=\phi$ and $\overline{X}(exia=1)=\{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X}(exia=2)=\phi$ and $\overline{X}(exia=2)=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rsex(X|exia=1)=\frac{0}{10}=0,$$

$$Rsex(X|exia=2)=\frac{0}{10}=0.$$

Mean roughness

$$Rough_{cp}(exia)=\frac{0+0}{2}=0.$$

3) With respect to cp

The lower and upper approximations are

$\underline{X}(exia=1)=\phi$ and $\overline{X}(exia=1)=\{2,3,7,8,9,10\}$,

$\underline{X}(exia=2)=\{1,4,5,6\}$ and $\overline{X}(exia=2)=\{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rcp(X|exia=1)=\frac{0}{6}=0,$$

$$Rcp(X|exia=2)=\frac{4}{10}=0.4.$$

Mean roughness

$$Rough_{cp}(exia)=\frac{0+0.4}{2}=0.2.$$

4) With respect to rbp

The lower and upper approximations are

$\underline{X}(exia=1)=\{2\}$ and $\overline{X}(exia=1)=\{1,2,3,6,7,8,10\}$,

$\underline{X}(exia=2)=\{4,5,9\}$ and $\overline{X}(exia=2)=\{1,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rrbp(X|exia=1)=\frac{1}{7}=0.1429,$$

$$Rrbp(X|exia=2)=\frac{3}{9}=0.3333.$$

Mean roughness

$$Rough_{rbp}(exia) = \frac{0.1429 + 0.3333}{2} = 0.2381.$$

5) With respect to cho

The lower and upper approximations are

$\underline{X(exia = 1)} = \{2,3,8\}$ and $\overline{X(exia = 1)} = \{2,3,5,8,10\}$,

$\underline{X(exia = 2)} = \{1,4,6,7,9\}$ and $\overline{X(exia = 2)} = \{1,4,5,6,7,9,10\}$.

Roughness

$$Rcho(X|exia = 1) = \frac{3}{5} = 0.6,$$

$$Rcho(X|exia = 2) = \frac{5}{7} = 0.7143.$$

Mean roughness

$$Rough_{cho}(exia) = \frac{0.6 + 0.7143}{2} = 0.6571.$$

6) With respect to fbs

The lower and upper approximations are

$\underline{X(exia = 1)} = \phi$ and $\overline{X(exia = 1)} = \{1,2,3,4,5,6,7,8,9,10\}$,

$\underline{X(exia = 2)} = \phi$ and $\overline{X(exia = 2)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rfbs(X|exia = 1) = \frac{0}{10} = 0,$$

$$Rfbs(X|exia = 2) = \frac{0}{10} = 0.$$

Mean roughness

$$Rough_{fbs}(exia) = \frac{0 + 0}{2} = 0.$$

7) With respect to $thal$

The lower and upper approximations are

$\underline{X(exia = 1)} = \phi$ and $\overline{X(exia = 1)} = \{2,3,4,5,6,7,8,9,10\}$,

$\underline{X(exia = 2)} = \{1\}$ and $\overline{X(exia = 2)} = \{1,2,3,4,5,6,7,8,9,10\}$.

Roughness

$$Rthal\left(X\,\middle|\,exia=1\right)=\frac{0}{9}=0,$$

$$Rthal\left(X\,\middle|\,exia=2\right)=\frac{1}{10}=0.1.$$

Mean roughness

$$Rough_{thal}\left(exia\right)=\frac{0+0.1}{2}=0.05.$$

8) With respect to *thalach*

The lower and upper approximations are

$\underline{X\left(exia=1\right)}=\{2,3\}$ and $\overline{X\left(exia=1\right)}=\{1,2,3,7,8,10\}$,

$\underline{X\left(exia=2\right)}=\{4,5,6,9\}$ and $\overline{X\left(exia=2\right)}=\{1,4,5,6,7,8,9,10\}$.

Roughness

$$Rthalach\left(X\,\middle|\,exia=1\right)=\frac{2}{6}=0.3333,$$

$$Rthalach\left(X\,\middle|\,exia=2\right)=\frac{4}{8}=0.5.$$

Mean roughness

$$Rough_{thalach}\left(exia\right)=\frac{0.3333+0.5}{2}=0.4167.$$

9) With respect to *oldpeak*

The lower and upper approximations are

$\underline{X\left(exia=1\right)}=\{2,3,8,10\}$ and $\overline{X\left(exia=1\right)}=\{2,3,8,10\}$,

$\underline{X\left(exia=2\right)}=\{1,4,5,6,7,9\}$ and $\overline{X\left(exia=2\right)}=\{1,4,5,6,7,9\}$.

Roughness

$$Roldpeak\left(X\,\middle|\,exia=1\right)=\frac{4}{4}=1,$$

$$Roldpeak\left(X\,\middle|\,exia=2\right)=\frac{6}{6}=1.$$

Mean roughness

$$Rough_{oldpeak}(exia)=\frac{1+1}{2}=1.$$

j. **Attribute** *oldpeak*

For attribute $oldpeak$, it is clear that $|V(oldpeak)|=9$.

The roughness and the mean roughness on $oldpeak$ with respect to $A,\quad A=age,sex,cp,rbp,cho,\ fbs,thal,thalach,exia,oldpeak$ is calculated as the following.

1) With respect to age

The lower and upper approximations are

$\underline{X(oldpeak=0.6)}=\phi$ and $\overline{X(oldpeak=0.6)}=\{6,8,10\}$,

$\underline{X(oldpeak=0.8)}=\phi$ and $\overline{X(oldpeak=0.8)}=\{6,8,10\}$,

$\underline{X(oldpeak=1.4)}=\{5\}$ and $\overline{X(oldpeak=1.4)}=\{1,2,3,5,7,9\}$,

$\underline{X(oldpeak=1.5)}=\phi$ and $\overline{X(oldpeak=1.5)}=\{1,2,3,7,9\}$,

$\underline{X(oldpeak=2.3)}=\phi$ and $\overline{X(oldpeak=2.3)}=\{1,2,3,7,9\}$,

$\underline{X(oldpeak=2.6)}=\phi$ and $\overline{X(oldpeak=2.6)}=\{1,2,3,7,9\}$,

$\underline{X(oldpeak=3.1)}=\phi$ and $\overline{X(oldpeak=3.1)}=\{6,8,10\}$,

$\underline{X(oldpeak=3.5)}=\{4\}$ and $\overline{X(oldpeak=3.5)}=\{4\}$,

$\underline{X(oldpeak=3.6)}=\phi$ and $\overline{X(oldpeak=3.6)}=\{1,2,3,7,9\}$.

Roughness

$$Rage(X|oldpeak=0.6)=\frac{0}{3}=0,$$

$$Rage(X|oldpeak=0.8)=\frac{0}{3}=0,$$

$$Rage(X|oldpeak=1.4)=\frac{1}{6}=01667,$$

$$Rage(X|oldpeak=1.5)=\frac{0}{5}=0,$$

$$Rage\left(X\left|oldpeak = 2.3\right.\right) = \frac{0}{5} = 0,$$

$$Rage\left(X\left|oldpeak = 2.6\right.\right) = \frac{0}{5} = 0,$$

$$Rage\left(X\left|oldpeak = 3.1\right.\right) = \frac{0}{3} = 0,$$

$$Rage\left(X\left|oldpeak = 3.5\right.\right) = \frac{1}{1} = 1,$$

$$Rage\left(X\left|oldpeak = 3.6\right.\right) = \frac{0}{5} = 0.$$

Mean roughness

$$Rough_{age}\left(oldpeak\right) = \frac{0+0+0.1667+0+0+0+0+1+0}{9} = 0.1296.$$

2) With respect to *sex*

The lower and upper approximations are

$\underline{X\left(oldpeak = 0.6\right)} = \phi$ and $\overline{X\left(oldpeak = 0.6\right)} = \{5,7,8\}$,

$\underline{X\left(oldpeak = 0.8\right)} = \phi$ and $\overline{X\left(oldpeak = 0.8\right)} = \{1,2,3,4,6,9,10\}$,

$\underline{X\left(oldpeak = 1.4\right)} = \phi$ and $\overline{X\left(oldpeak = 1.4\right)} = \{1,2,3,4,5,6,7,8,9,10\}$

$\underline{X\left(oldpeak = 1.5\right)} = \phi$ and $\overline{X\left(oldpeak = 1.5\right)} = \{1,2,3,4,6,9,10\}$,

$\underline{X\left(oldpeak = 2.3\right)} = \phi$ and $\overline{X\left(oldpeak = 2.3\right)} = \{1,2,3,4,6,9,10\}$,

$\underline{X\left(oldpeak = 2.6\right)} = \phi$ and $\overline{X\left(oldpeak = 2.6\right)} = \{1,2,3,4,6,9,10\}$,

$\underline{X\left(oldpeak = 3.1\right)} = \phi$ and $\overline{X\left(oldpeak = 3.1\right)} = \{1,2,3,4,6,9,10\}$,

$\underline{X\left(oldpeak = 3.5\right)} = \phi$ and $\overline{X\left(oldpeak = 3.5\right)} = \{1,2,3,4,6,9,10\}$,

$\underline{X\left(oldpeak = 3.6\right)} = \phi$ and $\overline{X\left(oldpeak = 3.6\right)} = \{5,7,8\}$.

Roughness

$$Rsex\left(X\left|oldpeak = 0.6\right.\right) = \frac{0}{3} = 0,$$

$$Rsex\left(X\left|oldpeak = 0.8\right.\right) = \frac{0}{7} = 0,$$

$$Rsex\left(X\middle|oldpeak = 1.4\right) = \frac{0}{10} = 0,$$

$$Rsex\left(X\middle|oldpeak = 1.5\right) = \frac{0}{7} = 0,$$

$$Rsex\left(X\middle|oldpeak = 2.3\right) = \frac{0}{7} = 0,$$

$$Rsex\left(X\middle|oldpeak = 2.6\right) = \frac{0}{7} = 0,$$

$$Rsex\left(X\middle|oldpeak = 3.1\right) = \frac{0}{7} = 0,$$

$$Rsex\left(X\middle|oldpeak = 3.5\right) = \frac{0}{7} = 0,$$

$$Rsex\left(X\middle|oldpeak = 3.6\right) = \frac{0}{3} = 0.$$

Mean roughness

$$Rough_{sex}(oldpeak) = \frac{0+0+0+0+0+0+0+0+0}{9} = 0.$$

3) With respect to cp

The lower and upper approximations are

$\underline{X(oldpeak = 0.6)} = \phi$ and $\overline{X(oldpeak = 0.6)} = \{2,3,7,8,9,10\}$,

$\underline{X(oldpeak = 0.8)} = \phi$ and $\overline{X(oldpeak = 0.8)} = \{5,6\}$,

$\underline{X(oldpeak = 1.4)} = \phi$ and $\overline{X(oldpeak = 1.4)} = \{2,3,5,6,7,8,9,10\}$,

$\underline{X(oldpeak = 1.5)} = \phi$ and $\overline{X(oldpeak = 1.5)} = \{2,3,7,8,9,10\}$,

$\underline{X(oldpeak = 2.3)} = \{1\}$ and $\overline{X(oldpeak = 2.3)} = \{1\}$,

$\underline{X(oldpeak = 2.6)} = \phi$ and $\overline{X(oldpeak = 2.6)} = \{2,3,7,8,9,10\}$,

$\underline{X(oldpeak = 3.1)} = \phi$ and $\overline{X(oldpeak = 3.1)} = \{2,3,7,8,9,10\}$,

$\underline{X(oldpeak = 3.5)} = \{4\}$ and $\overline{X(oldpeak = 3.5)} = \{4\}$,

$\underline{X(oldpeak = 3.6)} = \phi$ and $\overline{X(oldpeak = 3.6)} = \{2,3,7,8,9,10\}$.

Roughness

$$Rcp(X|oldpeak = 0.6) = \frac{0}{6} = 0,$$

$$Rcp(X|oldpeak = 0.8) = \frac{0}{2} = 0,$$

$$Rcp(X|oldpeak = 1.4) = \frac{0}{8} = 0,$$

$$Rcp(X|oldpeak = 1.5) = \frac{0}{6} = 0,$$

$$Rcp(X|oldpeak = 2.3) = \frac{1}{1} = 1,$$

$$Rcp(X|oldpeak = 2.6) = \frac{0}{6} = 0,$$

$$Rcp(X|oldpeak = 3.1) = \frac{0}{6} = 0,$$

$$Rcp(X|oldpeak = 3.5) = \frac{1}{1} = 1,$$

$$Rcp(X|oldpeak = 3.6) = \frac{0}{6} = 0.$$

Mean roughness

$$Rough_{cp}(oldpeak) = \frac{0+0+0+0+1+0+0+1+0}{9} = 0.2222.$$

4) With respect to rbp

The lower and upper approximations are

$$\underline{X(oldpeak = 0.6)} = \phi \text{ and } \overline{X(oldpeak = 0.6)} = \{3,6,8\},$$

$$\underline{X(oldpeak = 0.8)} = \phi \text{ and } \overline{X(oldpeak = 0.8)} = \{3,6,8\},$$

$$\underline{X(oldpeak = 1.4)} = \phi \text{ and } \overline{X(oldpeak = 1.4)} = \{4,5,9\},$$

$$\underline{X(oldpeak = 1.5)} = \{2\} \text{ and } \overline{X(oldpeak = 1.5)} = \{2\},$$

$$\underline{X(oldpeak = 2.3)} = \phi \text{ and } \overline{X(oldpeak = 2.3)} = \{1,7,10\},$$

$$\underline{X(oldpeak = 2.6)} = \phi \text{ and } \overline{X(oldpeak = 2.6)} = \{3,6,8\},$$

$$\underline{X(oldpeak = 3.1)} = \phi \text{ and } \overline{X(oldpeak = 3.1)} = \{1,7,10\},$$

$\underline{X}(oldpeak = 3.5) = \phi$ and $\overline{X}(oldpeak = 3.5) = \{4,5,9\}$,

$\underline{X}(oldpeak = 3.6) = \phi$ and $\overline{X}(oldpeak = 3.6) = \{1,7,10\}$.

Roughness

$$Rrbp(X|oldpeak = 0.6) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 0.8) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 1.4) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 1.5) = \frac{1}{1} = 1,$$

$$Rrbp(X|oldpeak = 2.3) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 2.6) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 3.1) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 3.5) = \frac{0}{3} = 0,$$

$$Rrbp(X|oldpeak = 3.6) = \frac{0}{3} = 0.$$

Mean roughness

$$Rough_{rbp}(oldpeak) = \frac{0+0+0+1+0+0+0+0+0}{9} = 0.1111.$$

5) With respect to *cho*

The lower and upper approximations are

$\underline{X}(oldpeak = 0.6) = \{8\}$ and $\overline{X}(oldpeak = 0.6) = \{8\}$,

$\underline{X}(oldpeak = 0.8) = \phi$ and $\overline{X}(oldpeak = 0.8) = \{1,6\}$,

$\underline{X}(oldpeak = 1.4) = \phi$ and $\overline{X}(oldpeak = 1.4) = \{4,5,9,10\}$,

$\underline{X}(oldpeak = 1.5) = \{2\}$ and $\overline{X}(oldpeak = 1.5) = \{2\}$,

$\underline{X}(oldpeak = 2.3) = \phi$ and $\overline{X}(oldpeak = 2.3) = \{1,6\}$,

$\underline{X}(oldpeak = 2.6) = \{3\}$ and $\overline{X}(oldpeak = 2.6) = \{3\}$,

$\underline{X}(oldpeak = 3.1) = \phi$ and $\overline{X}(oldpeak = 3.1) = \{5,10\}$,

$\underline{X}(oldpeak = 3.5) = \phi$ and $\overline{X}(oldpeak = 3.5) = \{4,9\}$,

$\underline{X}(oldpeak = 3.6) = \{7\}$ and $\overline{X}(oldpeak = 3.6) = \{7\}$.

Roughness

$$Rcho(X|oldpeak = 0.6) = \frac{1}{1} = 1,$$

$$Rcho(X|oldpeak = 0.8) = \frac{0}{2} = 0,$$

$$Rcho(X|oldpeak = 1.4) = \frac{0}{4} = 0,$$

$$Rcho(X|oldpeak = 1.5) = \frac{1}{1} = 1,$$

$$Rcho(X|oldpeak = 2.3) = \frac{0}{2} = 0,$$

$$Rcho(X|oldpeak = 2.6) = \frac{1}{1} = 1,$$

$$Rcho(X|oldpeak = 3.1) = \frac{0}{2} = 0,$$

$$Rcho(X|oldpeak = 3.5) = \frac{0}{2} = 0,$$

$$Rcho(X|oldpeak = 3.6) = \frac{1}{1} = 1.$$

Mean roughness

$$Rough_{cho}(oldpeak) = \frac{1+0+0+1+0+1+0+0+1}{9} = 0.4444.$$

6) With respect to fbs

The lower and upper approximations are

$\underline{X}(oldpeak = 0.6) = \phi$ and $\overline{X}(oldpeak = 0.6) = \{2,3,4,5,6,7,8,9\}$,

$\underline{X}(oldpeak = 0.8) = \phi$ and $\overline{X}(oldpeak = 0.8) = \{2,3,4,5,6,7,8,9\}$,

$\underline{X}(oldpeak = 1.4) = \phi$ and $\overline{X}(oldpeak = 1.4) = \{2,3,4,5,6,7,8,9\}$,

$\underline{X}(oldpeak = 1.5) = \phi$ and $\overline{X}(oldpeak = 1.5) = \{2,3,4,5,6,7,8,9\}$,

$\underline{X}(oldpeak = 2.3) = \phi$ and $\overline{X}(oldpeak = 2.3) = \{1,10\}$,

$\underline{X}(oldpeak = 2.6) = \phi$ and $\overline{X}(oldpeak = 2.6) = \{2,3,4,5,6,7,8,9\}$,

$\underline{X}(oldpeak = 3.1) = \phi$ and $\overline{X}(oldpeak = 3.1) = \{1,10\}$,

$\underline{X}(oldpeak = 3.5) = \phi$ and $\overline{X}(oldpeak = 3.5) = \{2,3,4,5,6,7,8,9\}$,

$\underline{X}(oldpeak = 3.6) = \phi$ and $\overline{X}(oldpeak = 3.6) = \{2,3,4,5,6,7,8,9\}$.

Roughness

$$Rfbs(X|oldpeak = 0.6) = \frac{0}{8} = 0,$$

$$Rfbs(X|oldpeak = 0.8) = \frac{0}{8} = 0,$$

$$Rfbs(X|oldpeak = 1.4) = \frac{0}{8} = 0,$$

$$Rfbs(X|oldpeak = 1.5) = \frac{0}{8} = 0,$$

$$Rfbs(X|oldpeak = 2.3) = \frac{0}{2} = 0,$$

$$Rfbs(X|oldpeak = 2.6) = \frac{0}{8} = 0,$$

$$Rfbs(X|oldpeak = 3.1) = \frac{0}{2} = 0,$$

$$Rfbs(X|oldpeak = 3.5) = \frac{0}{8} = 0,$$

$$Rfbs(X|oldpeak = 3.6) = \frac{0}{8} = 0.$$

Mean roughness

$$Rough_{fbs}(oldpeak) = \frac{0+0+0+0+0+0+0+0+0}{9} = 0.$$

7) With respect to *thal*

The lower and upper approximations are

$\underline{X}(oldpeak = 0.6) = \phi$ and $\overline{X}(oldpeak = 0.6) = \{2,4,5,6,7,8\}$,

$$\underline{X}(oldpeak = 0.8) = \phi \text{ and } \overline{X}(oldpeak = 0.8) = \{2,4,5,6,7,8\},$$

$$\underline{X}(oldpeak = 1.4) = \phi \text{ and } \overline{X}(oldpeak = 1.4) = \{2,3,4,5,6,7,8,9,10\},$$

$$\underline{X}(oldpeak = 1.5) = \phi \text{ and } \overline{X}(oldpeak = 1.5) = \{2,4,5,6,7,8\},$$

$$\underline{X}(oldpeak = 2.3) = \{1\} \text{ and } \overline{X}(oldpeak = 2.3) = \{1\},$$

$$\underline{X}(oldpeak = 2.6) = \phi \text{ and } \overline{X}(oldpeak = 2.6) = \{3,9,10\},$$

$$\underline{X}(oldpeak = 3.1) = \phi \text{ and } \overline{X}(oldpeak = 3.1) = \{3,9,10\},$$

$$\underline{X}(oldpeak = 3.5) = \phi \text{ and } \overline{X}(oldpeak = 3.5) = \{2,4,5,6,7,8\},$$

$$\underline{X}(oldpeak = 3.6) = \phi \text{ and } \overline{X}(oldpeak = 3.6) = \{2,4,5,6,7,8\}.$$

Roughness

$$Rthal(X|oldpeak = 0.6) = \frac{0}{6} = 0,$$

$$Rthal(X|oldpeak = 0.8) = \frac{0}{6} = 0,$$

$$Rthal(X|oldpeak = 1.4) = \frac{0}{9} = 0,$$

$$Rthal(X|oldpeak = 1.5) = \frac{0}{6} = 0,$$

$$Rthal(X|oldpeak = 2.3) = \frac{1}{1} = 1,$$

$$Rthal(X|oldpeak = 2.6) = \frac{0}{3} = 0,$$

$$Rthal(X|oldpeak = 3.1) = \frac{0}{3} = 0,$$

$$Rthal(X|oldpeak = 3.5) = \frac{0}{6} = 0,$$

$$Rthal(X|oldpeak = 3.6) = \frac{0}{6} = 0.$$

Mean roughness

$$Rough_{thal}(oldpeak) = \frac{0+0+0+0+1+0+0+0+0}{9} = 0.1111.$$

8) With respect to *thalach*

The lower and upper approximations are

$\underline{X}(oldpeak = 0.6) = \phi$ and $\overline{X}(oldpeak = 0.6) = \{7,8\}$,

$\underline{X}(oldpeak = 0.8) = \phi$ and $\overline{X}(oldpeak = 0.8) = \{7,8\}$,

$\underline{X}(oldpeak = 1.4) = \{9\}$ and $\overline{X}(oldpeak = 1.4) = \{5,6,9\}$,

$\underline{X}(oldpeak = 1.5) = \{2\}$ and $\overline{X}(oldpeak = 1.5) = \{2\}$,

$\underline{X}(oldpeak = 2.3) = \phi$ and $\overline{X}(oldpeak = 2.3) = \{1,10\}$,

$\underline{X}(oldpeak = 2.6) = \{3\}$ and $\overline{X}(oldpeak = 2.6) = \{3\}$,

$\underline{X}(oldpeak = 3.1) = \phi$ and $\overline{X}(oldpeak = 3.1) = \{1,10\}$,

$\underline{X}(oldpeak = 3.5) = \{4\}$ and $\overline{X}(oldpeak = 3.5) = \{4\}$,

$\underline{X}(oldpeak = 3.6) = \phi$ and $\overline{X}(oldpeak = 3.6) = \{7,8\}$.

Roughness

$$Rthalach(X|oldpeak = 0.6) = \frac{0}{2} = 0,$$

$$Rthalach(X|oldpeak = 0.8) = \frac{0}{2} = 0,$$

$$Rthalach(X|oldpeak = 1.4) = \frac{1}{3} = 0.3333,$$

$$Rthalach(X|oldpeak = 1.5) = \frac{1}{1} = 1,$$

$$Rthalach(X|oldpeak = 2.3) = \frac{0}{2} = 0,$$

$$Rthalach(X|oldpeak = 2.6) = \frac{1}{1} = 1,$$

$$Rthalach(X|oldpeak = 3.1) = \frac{0}{2} = 0,$$

$$Rthalach(X|oldpeak = 3.5) = \frac{1}{1} = 1,$$

$$Rthalach(X|oldpeak = 3.6) = \frac{0}{2} = 0.$$

Mean roughness

$$Rough_{thalach}\left(oldpeak\right)=\frac{0+0+0.3333+1+0+1+0+1+0}{9}=0.3704.$$

9) With respect to $exia$

The lower and upper approximations are

$\underline{X\left(oldpeak=0.6\right)}=\phi$ and $\overline{X\left(oldpeak=0.6\right)}=\{2,3,8,10\}$,

$\underline{X\left(oldpeak=0.8\right)}=\phi$ and $\overline{X\left(oldpeak=0.8\right)}=\{1,4,5,6,7,9\}$,

$\underline{X\left(oldpeak=1.4\right)}=\phi$ and $\overline{X\left(oldpeak=1.4\right)}=\{1,4,5,6,7,9\}$,

$\underline{X\left(oldpeak=1.5\right)}=\phi$ and $\overline{X\left(oldpeak=1.5\right)}=\{2,3,8,10\}$,

$\underline{X\left(oldpeak=2.3\right)}=\phi$ and $\overline{X\left(oldpeak=2.3\right)}=\{1,4,5,6,7,9\}$,

$\underline{X\left(oldpeak=2.6\right)}=\phi$ and $\overline{X\left(oldpeak=2.6\right)}=\{2,3,8,10\}$,

$\underline{X\left(oldpeak=3.1\right)}=\phi$ and $\overline{X\left(oldpeak=3.1\right)}=\{2,3,8,10\}$,

$\underline{X\left(oldpeak=3.5\right)}=\phi$ and $\overline{X\left(oldpeak=3.5\right)}=\{1,4,5,6,7,9\}$,

$\underline{X\left(oldpeak=3.6\right)}=\phi$ and $\overline{X\left(oldpeak=3.6\right)}=\{1,4,5,6,7,9\}$.

Roughness

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=0.6\right)=\frac{0}{4}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=0.8\right)=\frac{0}{6}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=1.4\right)=\frac{0}{6}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=1.5\right)=\frac{0}{4}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=2.3\right)=\frac{0}{6}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=2.6\right)=\frac{0}{4}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=3.1\right)=\frac{0}{4}=0,$$

$$\mathrm{Re}\,xia\left(X\middle|oldpeak=3.5\right)=\frac{0}{6}=0,$$

$$\operatorname{Re}xia\left(X \,\middle|\, oldpeak = 3.6\right) = \frac{0}{6} = 0.$$

Mean roughness

$$Rough_{exia}\left(oldpeak\right) = \frac{0+0+0+0+0+0+0+0+0}{9} = 0.$$

Calculation of Max Roughness on Each Attribute

Table 3.5: Calculation of the Max Roughness on each attribute

Attribute	**Mean Roughness**									**Max Roughness**
age	*sex*	*cp*	*rbp*	*cho*	*fbs*	*thal*	*thalach*	*exia*	*oldpeak*	0.6667
	0	0.2857	0.025	0.1571	0	0.025	0.3571	0	0.6667	
sex	*age*	*cp*	*rbp*	*cho*	*fbs*	*thal*	*thalach*	*exia*	*oldpeak*	0.625
	0.1111	0.1	0.05	0.625	0.1	0.2	0.625	0	0.625	
cp	*age*	*sex*	*rbp*	*cho*	*fbs*	*thal*	*thalach*	*exia*	*oldpeak*	0.7619
	0.3125	0	0.025	0.125	0	0.3333	0.6786	0.1	0.7619	
rbp	*age*	*sex*	*cp*	*cho*	*fbs*	*thal*	*thalach*	*exia*	*oldpeak*	**I=1**
	0.07143	0	0.0635	0.55	0.05	0.025	0.55	0	1	**II=0.55**
cho	*age*	*sex*	*cp*	*rbp*	*fbs*	*thal*	*thalach*	*exia*	*oldpeak*	0.8095
	0.05953	0	0.06803	0.1429	0	0.02041	0.4286	0	0.8095	
fbs	*age*	*sex*	*cp*	*rbp*	*cho*	*thal*	*thalach*	*exia*	*oldpeak*	**I=1**
	0.1	0.15	0.2381	0.35	0.3	0.4583	1	0	1	**II=1**
thal	*age*	*sex*	*cp*	*rbp*	*cho*	*fbs*	*thalach*	*exia*	*oldpeak*	0.7381
	0.06667	0.1	0.4444	0.03333	0.1778	0	0.5	0	0.7381	
thalach	*age*	*sex*	*cp*	*rbp*	*cho*	*fbs*	*thal*	*exia*	*oldpeak*	0.7619
	0.1786	0	0.3061	0.1429	0.4286	0.1429	0.03571	0	0.7619	
exia	*age*	*sex*	*cp*	*rbp*	*cho*	*fbs*	*thal*	*thalach*	*oldpeak*	**I=1**

	0.1	0	0.2	0.2381	0.6571	0	0.05	0.4167	1	**II=0.6571**
oldpeak	*age*	*sex*	*cp*	*rbp*	*cho*	*fbs*	*thal*	*thalach*	*exia*	0.4444
	0.1296	0	0.2222	0.1111	0.4444	0	0.1111	0.3704	0	

3.3 OBJECT SPLITTING MODEL

From Table 3.5, we can calculate the Max-Max Roughness of all attributes

Table 3.6: Max-Max Roughness

	Max Roughness										**Max-Max Roughness**
Attribute	*age*	*sex*	*cp*	*rbp*	*cho*	*fbs*	*thal*	*thalach*	*exia*	*oldpeak*	**1**
	0.6667	0.625	0.7619	**I=1** **II=0.55**	0.8095	**I=1** **II=1**	0.7381	0.7619	**I=1** **II=0.6571**	0.4444	

3.3.1 A Clustering Attribute with Max-Max Roughness is Found

Table 5 shows the calculation and illustrates that the attribute *rbp* and *fbs* have the same Max Roughness. It is recommended to look at the next highest Max Roughness inside the attributes that are tied and so on until the tie is broken. In the Table 5, the second Max Roughness corresponding to attribute *fbs* is higher than that of *rbp*. Therefore, attribute *fbs* is selected as a clustering attribute and binary splitting is conducted.

3.3.2 A Splitting Point Attribute *fbs* is Determined

The splitting set should include the attribute value which has maximum roughness. Taking a look at Table 3.4:

$X(fbs=2)$ has overall maximum roughness with respect to A, $(A = sex, \cdots, oldpeak)$ comparing to $X(fbs=1)$. Thus, splitting on $X(fbs=2)$ versus $X(fbs=1)$ is chosen. The partition at this stage can be represented as a tree and is shown in figure 3.3.

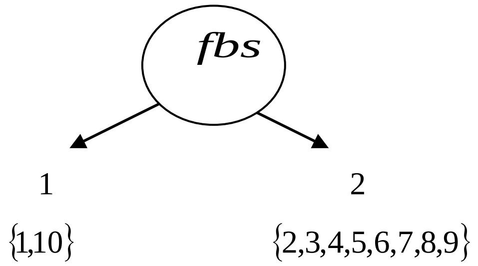


Figure 3.3: Splitting attributes

3.4 CLUSTER PURITY

Cluster purity is defined as the percentage of the dominant class label in each cluster [34]. Therefore, it can be said that, the higher the number of dominant class label in each cluster, the higher the level of the cluster's purity. Cluster purity is used as the determination of the accuracy or quality of the clusters made. However, clustering which is an unsupervised learning process does not deal with labeled or predefined dataset. Therefore, assumptions are made where all the samples are predicted to be members of the actual dominant class for that cluster [35].

Mathematically, cluster purity is defined as follows [36]:

$$Purity(i) = \frac{the\ number\ of\ data\ occuring\ in\ both\ (i)th\ cluster\ under\ given\ threshold}{the\ number\ of\ data\ in\ the\ data\ set}$$

$$Overall\ Purity = \frac{\sum_{i=1}^{n} Purity(i)}{n}$$

Where n is the total number of cluster. Generally, the larger the purity value is, the better the clustering is.

CHAPTER 4

RESULT AND DISCUSSION

This chapter briefly elaborates on the expected results and the discussion of this project. The fourth chapter consists of two sections where the first parts explain the implementation and the second part explain about the dataset used in order to achieve the result. Then the expected result will be discussed further in this chapter.

4.1 IMPLEMENTATION

The data clustering technique of Maximum-maximum Roughness of Rough Set will be implemented using the programming tool of Visual Basic (VB). The system will firstly ask the user to browse for the dataset with .xls file-formatted to be used in the data clustering program. Then, the system will show the table of dataset that the user has input in the browser box. The system first can calculate the lower and upper class approximations based on the attributes selected by the user. After that, the system can calculate the mean roughness as well as the maximum roughness of attributes. The way the calculation for the maximum roughness work is by analyzing the inserted data set. Then, each of the attribute in the data set will be considered as a candidate attribute to partition. After that, the system will determine the mean roughness of attribute-value pairs followed by determining the highest maximum roughness of attribute a_i on attribute a_j where I is not equal to j. Finally, based on the result of the maximum roughness for each attribute, the system will select a clustering attribute.

4.2 DATASET

Firstly, use the small datasets of heart disease patient with the criteria or attributes. The data set contains 10 attributes. The attributes are: age, sex, cp rbp, cho, fbs, thal, thalach, exia and oldpeak. Secondly, is a real dataset of heart disease patient with the attributes taken on “Internet”. The dataset consists of few patients with the attributes with several attributes used to determine the decision. From the proposed system of Data Clustering using Maximum-Maximum Roughness, it is expected that the system can select the best attributes that can be used to cluster the data based on its attributes, by making a fast and accurate calculation of each of the attributes degree so that the maximum dependency of the attributes can be determined. The result is important as it is the based on which attribute that is the most suitable to be selected to cluster the data. If the calculations are accurate, the best kind of clustering can be accomplished by using the most suitable attribute. By having this system, it is also expected that user can easily cluster a data using the maximum roughness technique easily. The calculation work will be faster as the system will do it for the user especially with large datasets. Since this system will make desicion based on the data inserted by the user, there are certain cases that cannot be solved by this system. This system is not perfected yet as the method of

Data Clustering using Maximum-Maximum Roughness might require some improvement later.

4.3 INTERFACES

The interface will shown when the users start running the software. It is the first interface of the software. The first interface consist one part for viewing the table of datasets and two tabs which is for viewing calculation part and result part respectively. The figure 4.1 for calculations tab and figure 4.2 for results tab are shown below.

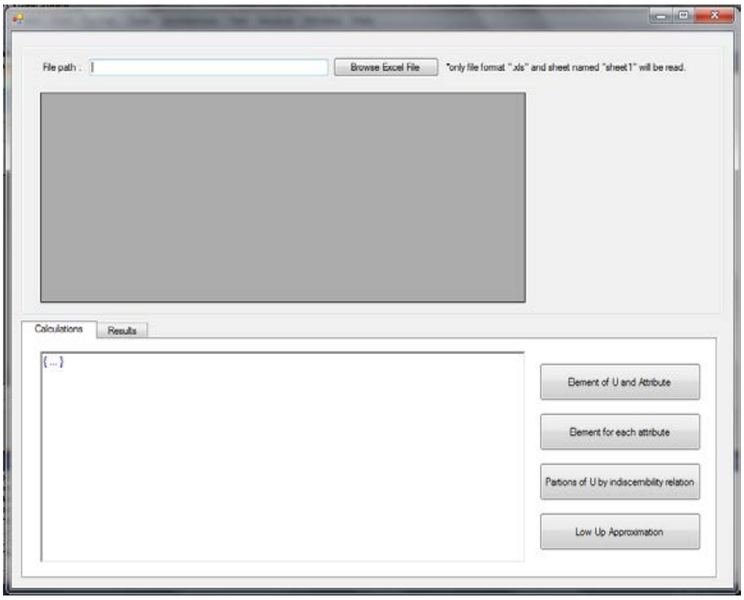


Figure 4.1: First main interface (Calculations tab).

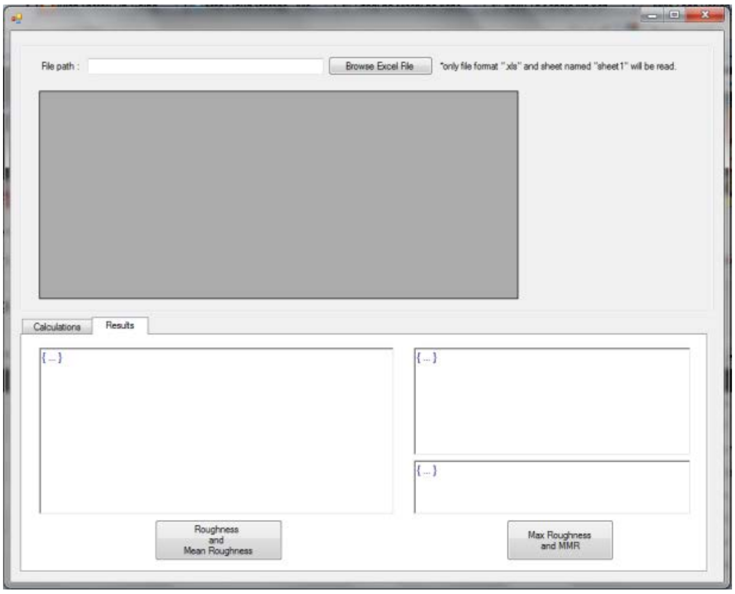


Figure 4.2: Second main interface (Results tab).

The button ‘Browse excel file’ will need the user to browse the excel file with .xls format only and sheet named “sheet1”. After that, the application system will display the table of dataset from the excel file into the provided space below the button ‘Browse excel file’.

From the figure 4.3, the user need to browse the excel file before the application can view the table of datasets.

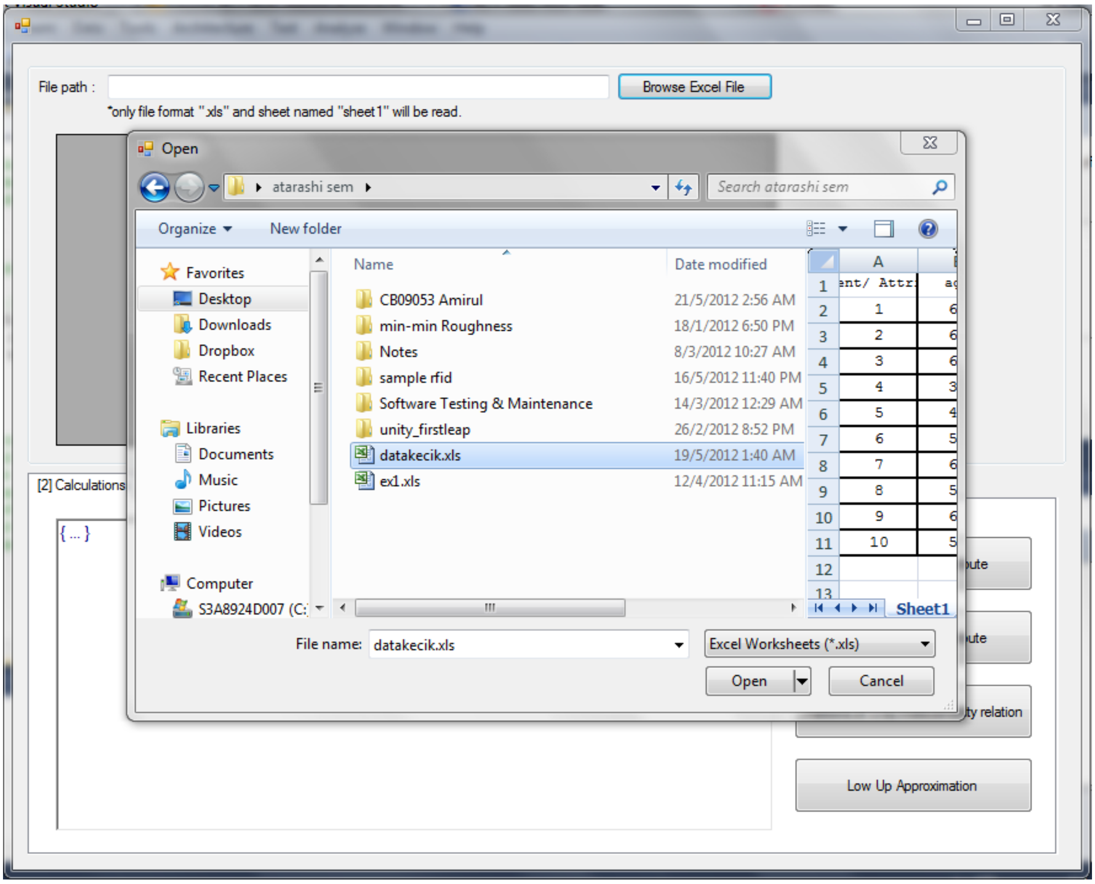
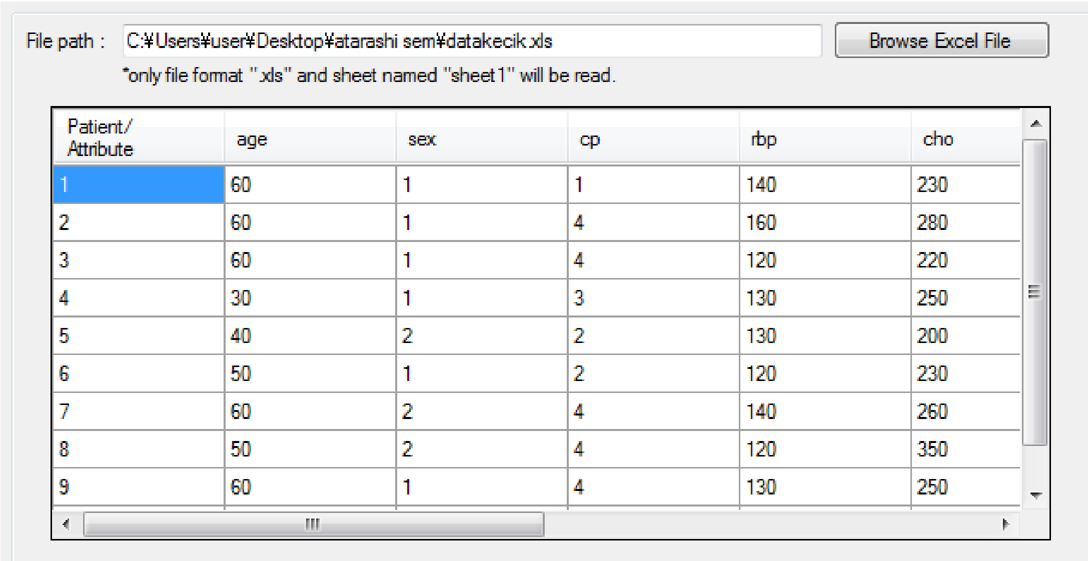


Figure 4.3: Browse the excel file.

The application then will show the table of datasets after browsing the excel file into the spaced area provided at below of the browsing part as shown in figure 4.4.



Patient/ Attribute	age	sex	cp	rbp	cho
1	60	1	1	140	230
2	60	1	4	160	280
3	60	1	4	120	220
4	30	1	3	130	250
5	40	2	2	130	200
6	50	1	2	120	230
7	60	2	4	140	260
8	50	2	4	120	350
9	60	1	4	130	250

Figure 4.4: Dataset of imported excel file

Then, there are two tabs where the first tab is the calculations tab and the second tab is the results tab. The calculation tab will show to the user the set of element of U and Attributes, element for each attributes, partitions of U by indiscernbility relation; and show the calculated lower approximation and upper approximation for each attributes. The button 'Element of U and Attribute' clicked by user will show the list of elements in the U and list of attributes as shown in figure 4.5.

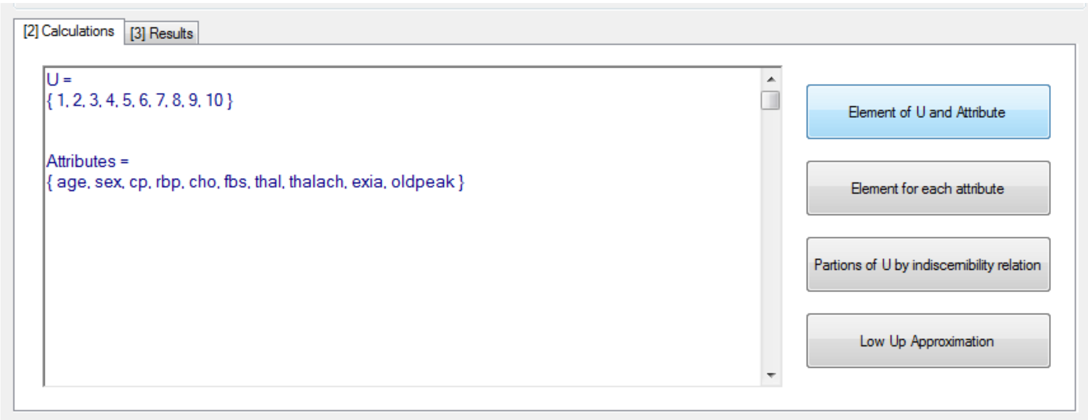


Figure 4.5: Element of U and Attribute

The next button after 'Element of U and Attribute' button is the 'Element for each attribute'. The button will show the list of elements in each of the attributes to the user as shown in figure 4.6.

Figure 4.6: Element for each attribute

The third button in calculation tab which is the ‘Partitions of U by indiscernbility relation’ button will show the clustered partitions by using the element in U and with the list of attributes. The partitions are shown in figure 4.7.

Figure 4.7: Partitions of U indiscernibility relation

The last button in the calculation tab which is the ‘Low Up Approximation’ is for showing the calculated lower approximation and upper approximation of each attribute with respect to the other attribute and the calculated lower and upper approximation is shown as the figure 4.8.

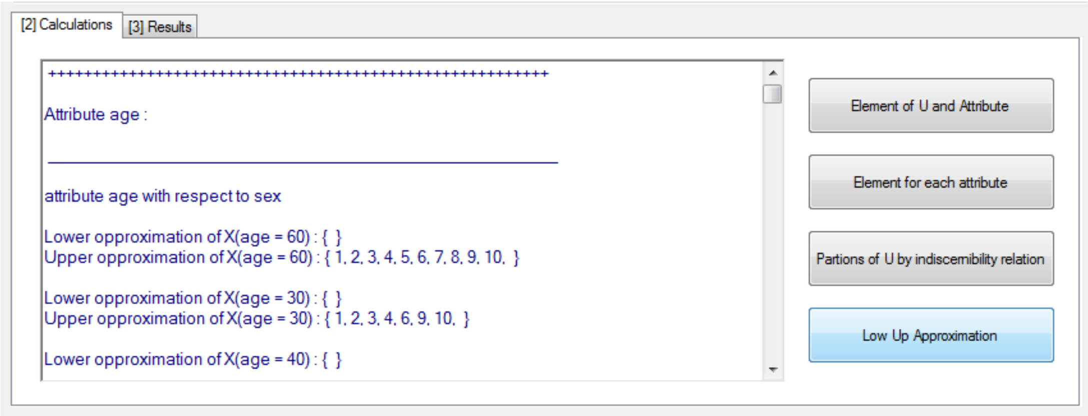


Figure 4.8: Calculation of Lower and Upper Approximation

The second tab called the '[3] Results' tab consist of two buttons which is the 'Roughness and Mean Roughness' button and 'Max Roughness and MMR' button. The button 'Roughness and Mean Roughness' will show the user the calculated equation to get the roughness of the attributes and mean of the roughness. The second button gives the user the highest value from the mean of roughness of each attributes which is called max roughness and also the application will show the highest value from the max roughness. The final highest value is known as maximum-maximum roughness. The following figure 4.9 show the action made by the buttons.

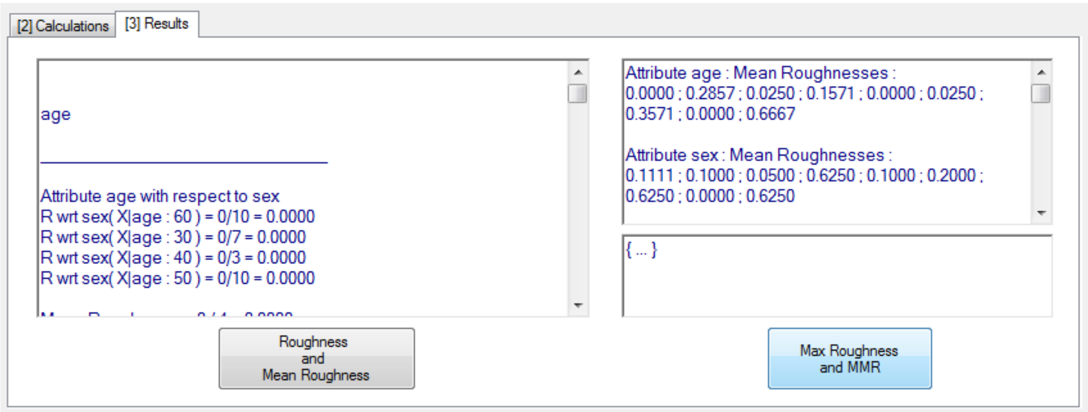


Figure 4.9: Result of Roughnesses

CHAPTER 5

CONCLUSION

Clustering is the problem of identifying the distribution of patterns and intrinsic correlation in large data set by partitioning the data points into similar classes. Data clustering is a common technique for statistical data analysis, which is used in many fields including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the classification of similar object into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait often proximity according to some similarity. In this project, a clustering algorithm using rough set theory has been presented. For calculating the highest value of maximum roughness of attributes of patient suspected heart diseases, the Maximum-maximum Roughness technique based on Rough Set Theory is introduced in for this project. This theory was produced by Pawlak in 1981. The first thing for the technique is to find the lower approximation and the upper approximation of the dataset based on the attributes in the dataset. Next is to find the average roughness of the attributes and after that we will find the maximum roughness of attributes by calculating the degree for each of the element. Finally, with the maximum roughness, we can get the highest maximum roughness of the attributes. The clustering system will be developed using Microsoft Visual Basic 2010 express as a programming language.

REFERENCES

[1] http://en.wikipedia.org/wiki/Knowledge_discovery.

[2] Carter C. L. and Hamilton H. J., Efficient Attribute-Oriented Generalization for Knowledge Discovery from Large Databases, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 10, No. 2, March/April, 1998.

[3] Frawley W. J., Piatetsky-Shapiro G., and Matheus C. J., Knowledge Discovery in Databases: An Overview, *AI Magazine*, Vol. 13, No. 3, 1992.

[4] Fayyad U., Piatetsky-Shapiro G., Smyth P., Knowledge Discovery and Data Mining: Towards a Unifying Framework, 1996.

[5] Fayyad U., Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases, 1997.

[6] Colet E., Clustering and Classification: Data Mining Approaches, http://www.taborcommunications.com/dsstar/00/0704/101861.html.

[7] Kim K., Choi H., Design of a Clinical Knowledge Base for Heart Disease Detection, 2007.

[8] Fogoros R. N., M.D., Key Symptoms of Heart Disease, *About.com Health's Disease and Condition*, 2009.

[9] Carapetis J. R., Rheumatic Heart Disease in Asia, 2008.

[10] Fogoros R. N., M.D., What is Heart Disease?, *About.com Health's Disease and Condition*, 2009.

[11] Fogoros R. N., M.D., Key Symptoms of Heart Disease, *About.com Health's Disease and Condition*, 2009.

[12] http://www.frost.com/prod/servlet/market-insight-top.pag?docid=18817766.

[13] Carapetis J. R., Rheumatic Heart Disease in Asia, 2008.

[14] Carapetis J. R., Rheumatic Heart Disease in Asia, 2008.

[15] http://www.frost.com/prod/servlet/market-insight-top.pag?docid=18817766.

[16] http://www.frost.com/prod/servlet/market-insight-top.pag?docid=18817766.

[17] http://www.sign.ac.uk/guidelines/fulltext/95/section2.html.

[18] Reference: Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, *Advances in Knowledge Discovery and Data Mining,* AAAI Press / The MIT Press, Menlo Park, CA, 1996, pp.1-34.

[19] http://www.research.ibm.com/compsci/kdd/brochure2000.html.

[20] http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/lecture13_4up.pdf.

[21] Cattral R., Oppacher F., Deugo, D., Supervised and unsupervised data mining with an evolutionary algorithm, *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, Vol.2, No., pp.767-774.

[23] Cheong C. W., Clustering and Predicting Stock Prices Using Advanced Wavelet Methods and Neural Networks – a study on KLSE.

[24] Jain A. K., Murty M. N., Flynn P.J., Data Clustering: A Review, *ACM Computing Surveys (CSUR)*, Vol. 31, Issue 3, pp. 264-323.

[25] Jerzy W. Grzymala-Busse, Mining Numerical Data- A rough Set Approach, *Institute of Computer Science, Polish Academy of Sciences 01-237 Warsaw*, Poland.

[26] Brown F.M., *Boolean Reasoning*, Kluwer Academic Publishers, Dordrecht, 1990.

[27] Dubois D., Prade H., Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems, 17:191-209 (1990).

[28] Nakamura A., Gao, J.M. On a KTB-modal fuzzy logic, Fuzzy Sets and Systems, 45:327-334 (1992).

[29] Yao Y.Y., Combination of rough and fuzzy sets based on _-level sets, in: Lin T.Y. and Cercone N. (Eds.), Rough Sets and Data Mining: Analysis for Imprecise Data, Kluwer Academic Publishers, Boston, pp. 301-321, 1997.

[30] Voges K.E., Pope N.K.L., Brown M.R., Cluster analysis of marketing data examining online shopping orientation: a comparison of k-means and rough clustering approaches, in: H.A. Abbass, R.A. Sarker, C.S. Newton (Eds.), Heuristics and Optimization for Knowledge Discovery, Idea Group Publishing, Hershey, 2002, pp. 207–224.

[31] Pawlak Z., Rough sets, *International Journal of Computer and Information Science* 11, 341-356, (1982).

[32] W, -Z. Wu, W,-X. Zhang and H. –Z. Li, Knowledge acquisition in incomplete fuzzy information system via the rough set approach, Expert System 20 (5), 280-286, (2003).

[33] Pawlak Z., Rough classification, *Institute of Computer Science, Polish Academy of Science, P.O. Box 22, 00-901 Warsaw, PKiN, Poland*, 6 May 1983.

[34] Abidi S. S. R., Meng Hoe K., Goh A., Analyzing Data Clusters: A Rough Set Approach to Extract Cluster-Defining Symbolic Rules.

[35] Aggarwal C. C. A Human-Computer Interactive Method for Projected Clustering. *IEEE Transacions on Knowledge and Data Engineering*, 2004, 16(4), pp. 448-460.

[36] Cluster Purity Metric, http://www.gabormelli.com/RKB/Cluster_Purity_Metric, Retrieved December 15, 2011.

[37] Yanto I. T. R., Herawan T. and Deris M. M., A Framework of Rough Clustering for Web Transactions, *Advances In Intelligent Information and Database Systems, Studies in Computational Intelligence*, 2010, vol 283/2010, pp.265-277.

APPENDIX

SOURCE CODE

```
'   declaring"
Public Class frmStart


    Dim UInd As String
    Dim attlowUpCount(-1, -1, -1, -1) As Integer
    Dim attVariable(-1, -1) As String
    Dim wholeMaxValue As Integer
    Dim maxAttVar(-1) As Integer
    Dim meanRoughness(-1, -1) As Double
    Dim maxR(-1, -1) As Double
    Dim trying As Integer

'   browse and show dataset"
    Private Sub btnBrowse_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnBrowse.Click

        odf1.InitialDirectory = "::{20D04FE0-3AEA-1069-A2D8-08002B30309D}"
        odf1.ShowDialog()
        txtDataLink.Text = odf1.FileName.ToString

        Dim dt As New DataTable()
        Dim connStr As String = "Provider=Microsoft.Jet.OLEDB.4.0;" & _
                          "Data Source=" + txtDataLink.Text + ";" & _
                          "Extended Properties=Excel 8.0;"
        Dim sqlStr As String = "SELECT * FROM [SHEET1$]"
        Dim conn As New OleDb.OleDbDataAdapter(sqlStr, connStr)
```

```
        conn.Fill(dt)
        conn.Dispose()
        dgv1.DataSource = dt

        ReDim attVariable(dgv1.Columns.Count - 1, 1)
        ReDim maxAttVar(dgv1.Columns.Count - 1)
        ReDim meanRoughness(dgv1.Columns.Count - 1, dgv1.Columns.Count - 1)
        ReDim maxR(dgv1.Rows.Count - 1, dgv1.Rows.Count - 1)

'    ##########  listing element of U :  ##########
        rtbResult.Text = "U = " & vbNewLine & "{ "
        For a = 0 To dgv1.Rows.Count - 2
            rtbResult.Text = rtbResult.Text & CStr(dgv1.Item(0, a).Value) & ", "
        Next
        rtbResult.Text = rtbResult.Text & CStr(dgv1.Item(0, dgv1.Rows.Count - 1).Value)
& " }"

        rtbResult.Text = rtbResult.Text & vbNewLine & vbNewLine & vbNewLine

'    ##########  listing attributes :    ##########
        rtbResult.Text = rtbResult.Text & "Attributes = " & vbNewLine & "{ "
        For a = 1 To dgv1.Columns.Count - 2
            rtbResult.Text = rtbResult.Text & CStr(dgv1.Columns(a).HeaderText) & ", "
        Next
        rtbResult.Text = rtbResult.Text & CStr(dgv1.Columns(dgv1.Columns.Count -
1).HeaderText) & " }"

    End Sub

    Private Sub btnCalcLowUp_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnCalcLowUp.Click


        Dim a, a2, b, b1, b2, e1, e2, count, count2, lowTestCount, tag, tag2 As Integer
        wholeMaxValue = 1


'    tagging all items to "True"
        For e1 = 1 To dgv1.Columns.Count - 1
            For e2 = 0 To dgv1.Rows.Count - 1
                dgv1.Item(e1, e2).Tag = "True"
            Next
        Next
```

```
' fnding the max number variable for each attribute in column (a)
' tagging 1st occurence of each variable item as "True", other as "False"
        For a = 1 To dgv1.Columns.Count - 1
            maxAttVar(a - 1) = 1
            attVariable(a - 1, (maxAttVar(a - 1) - 1)) = dgv1.Item(a, 0).Value
            For b1 = 1 To dgv1.Rows.Count - 1
                For b2 = 0 To (b1 - 1)
                    If dgv1.Item(a, b1).Value = dgv1.Item(a, b2).Value Then
                        dgv1.Item(a, b1).Tag = "False"
                    End If
                Next
                If dgv1.Item(a, b1).Tag <> "False" Then
                    maxAttVar(a - 1) = maxAttVar(a - 1) + 1
                    If maxAttVar(a - 1) > wholeMaxValue Then
                        wholeMaxValue = maxAttVar(a - 1)
                        ReDim Preserve attVariable(dgv1.Columns.Count - 1,
wholeMaxValue)
                    End If
                    attVariable(a - 1, (maxAttVar(a - 1) - 1)) = dgv1.Item(a, b1).Value
                End If
            Next
            ReDim Preserve attlowUpCount(dgv1.Columns.Count - 1, dgv1.Columns.Count - 1,
2, wholeMaxValue)



'   finding the total number of variable in an attribute column
            tag = 1
            For b1 = 0 To dgv1.Rows.Count - 1
                If dgv1.Item(a, b1).Tag = True Then
                    dgv1.Item(a, b1).Tag = CStr(tag)
                    tag = tag + 1
                End If
            Next

' tagging each attribute value as 1, 2, ..n to represent each attribute variable
            For tag = 1 To maxAttVar(a - 1)
                For b1 = 0 To dgv1.Rows.Count - 1
                    If dgv1.Item(a, b1).Tag = CStr(tag) Then
                        For b2 = (b1 + 1) To dgv1.Rows.Count - 1
                            If dgv1.Item(a, b2).Value = dgv1.Item(a, b1).Value And
dgv1.Item(a, b2).Tag <> CStr(tag) Then
                                dgv1.Item(a, b2).Tag = CStr(tag)
                            End If
                        Next
```

```vb
                    End If
                Next
            Next

        Next



        rtbResult.Text = " ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++" &
vbNewLine
        For a = 1 To dgv1.Columns.Count - 1

            rtbResult.Text = rtbResult.Text & vbNewLine & "Attribute " &
dgv1.Columns(a).HeaderText & " :" & vbNewLine & vbNewLine

' finding the lower & upper with respect to other attributes
            rtbResult.Text = rtbResult.Text & "
__________________________________________________________" & vbNewLine
            For a2 = 1 To dgv1.Columns.Count - 1
                If a <> a2 Then

                    rtbResult.Text = rtbResult.Text & vbNewLine & "attribute " &
dgv1.Columns(a).HeaderText & " with respect to " & dgv1.Columns(a2).HeaderText &
vbNewLine

                    For tag = 1 To maxAttVar(a - 1)

'    counting the number of occurence for each attribute variable
                        count = 0
                        For b1 = 0 To dgv1.Rows.Count - 1
                            If dgv1.Item(a, b1).Tag = CStr(tag) Then
                                count = count + 1
                            End If
                        Next

'    tagging all item in 1st column (or U) as "None"
                        For b1 = 0 To dgv1.Rows.Count - 1
                            dgv1.Item(0, b1).Tag = "None"
                        Next

'    comparing each set in (a) to each set in (a2) to find out the lower and upper
approximation
                        For tag2 = 1 To maxAttVar(a2 - 1)


'    counting the number of occurence for each attribute variable in column (a2)
```

```
                            count2 = 0
                            For b1 = 0 To dgv1.Rows.Count - 1
                                If dgv1.Item(a2, b1).Tag = CStr(tag2) Then
                                    count2 = count2 + 1
                                End If
                            Next

                            lowTestCount = 0

'   comparing to set to find out the number/count of the same value
                            For e1 = 0 To dgv1.Rows.Count - 1
                                If dgv1.Item(a, e1).Tag = CStr(tag) Then
                                    For e2 = 0 To dgv1.Rows.Count - 1
                                        If dgv1.Item(a2, e2).Tag = CStr(tag2) And
dgv1.Item(0, e2).Value = dgv1.Item(0, e1).Value Then
                                            lowTestCount = lowTestCount + 1
                                        End If
                                    Next
                                End If
                            Next

'    tagging as "low" if the number/count of the same value is the same as the number of
occurence of the respective attribute variable
                            If lowTestCount = count2 Then
                                For e2 = 0 To dgv1.Rows.Count - 1
                                    If dgv1.Item(a2, e2).Tag = CStr(tag2) Then
                                        dgv1.Item(0, e2).Tag = "low"
                                    End If
                                Next
                            End If

'    tagging as "up" or "low&up" for each item in the set where there exist the
occurence of the same value
                            For e1 = 0 To dgv1.Rows.Count - 1
                                If dgv1.Item(a, e1).Tag = CStr(tag) Then
                                    For e2 = 0 To dgv1.Rows.Count - 1
                                        If dgv1.Item(a2, e2).Tag = CStr(tag2) And
dgv1.Item(0, e1).Value = dgv1.Item(0, e2).Value Then
                                            For b = 0 To dgv1.Rows.Count - 1
                                                If dgv1.Item(a2, b).Tag = CStr(tag2) And
dgv1.Item(0, b).Tag = "None" Then
                                                    dgv1.Item(0, b).Tag = "up"
                                                ElseIf dgv1.Item(a2, b).Tag = CStr(tag2)
And dgv1.Item(0, b).Tag = "low" Then
                                                    dgv1.Item(0, b).Tag = "low&up"
                                                End If
```

```
                                        Next
                                    End If
                                Next
                            End If
                        Next

                    Next

'   writing the low and up approx in the richtextbox
                    rtbResult.Text = rtbResult.Text & vbNewLine & "Lower
approximation of X(" & dgv1.Columns(a).HeaderText & " = " & attVariable(a - 1, tag - 1)
& ") : { "
                    attlowUpCount(a - 1, a2 - 1, 0, tag - 1) = 0
                    For e1 = 0 To dgv1.Rows.Count - 1
                        If dgv1.Item(0, e1).Tag = "low&up" Then
                            rtbResult.Text = rtbResult.Text & dgv1.Item(0, e1).Value
& ", "
                            attlowUpCount(a - 1, a2 - 1, 0, tag - 1) =
attlowUpCount(a - 1, a2 - 1, 0, tag - 1) + 1
                        End If
                    Next
                    rtbResult.Text = rtbResult.Text & " }" & vbNewLine
                    rtbResult.Text = rtbResult.Text & "Upper approximation of X(" &
dgv1.Columns(a).HeaderText & " = " & attVariable(a - 1, tag - 1) & ") : { "
                    attlowUpCount(a - 1, a2 - 1, 1, tag - 1) = 0
                    For e1 = 0 To dgv1.Rows.Count - 1
                        If dgv1.Item(0, e1).Tag = "up" Or dgv1.Item(0, e1).Tag =
"low&up" Then
                            rtbResult.Text = rtbResult.Text & dgv1.Item(0, e1).Value
& ", "
                            attlowUpCount(a - 1, a2 - 1, 1, tag - 1) =
attlowUpCount(a - 1, a2 - 1, 1, tag - 1) + 1
                        End If
                    Next
                    rtbResult.Text = rtbResult.Text & " }" & vbNewLine

                Next


                rtbResult.Text = rtbResult.Text & vbNewLine & "
________________________________________________________" & vbNewLine
            End If

        Next
        rtbResult.Text = rtbResult.Text & vbNewLine & "
++++++++++++++++++++++++++++++++++++++++++++++++++++++++" & vbNewLine
```

```vb
        Next
        TabPage1.Enabled = True

    End Sub

    Private Sub btnEOfAtt_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnEOfAtt.Click

        Dim maxAttValue, count As Integer
'   ##########  listing element of each attributes : ##########

        For a = 1 To dgv1.Columns.Count - 1
            For b1 = 1 To dgv1.Rows.Count - 1
                dgv1.Item(a, b1).Tag = "True"
            Next
        Next

        rtbResult.Text = ""
        For a = 1 To dgv1.Columns.Count - 1
            rtbResult.Text = rtbResult.Text & "V of " & CStr(dgv1.Columns(a).HeaderText)
& " : { " & dgv1.Item(a, 0).Value & ", "
            maxAttValue = 0
            count = 1
            For b1 = 1 To dgv1.Rows.Count - 1
                For b2 = 0 To (b1 - 1)
                    If dgv1.Item(a, b1).Value = dgv1.Item(a, b2).Value Then
                        dgv1.Item(a, b1).Tag = "False"
                    End If
                Next
                If dgv1.Item(a, b1).Tag <> "False" Then
                    maxAttValue = maxAttValue + 1
                End If
            Next
            For b1 = 1 To dgv1.Rows.Count - 1
                If dgv1.Item(a, b1).Tag <> "False" And maxAttValue <> count Then
                    rtbResult.Text = rtbResult.Text & dgv1.Item(a, b1).Value & ", "
                    count = count + 1
                ElseIf dgv1.Item(a, b1).Tag <> "False" And maxAttValue = count Then
                    rtbResult.Text = rtbResult.Text & dgv1.Item(a, b1).Value & " }" &
vbNewLine & vbNewLine
                End If
            Next
        Next

    End Sub
```

```
    Private Sub Button1_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnUInd.Click

        Dim maxAttValue As Integer

' ########## finding U/IND for each attribute : ##########
' tagging all items to "True"
        For a = 1 To dgv1.Columns.Count - 1
            For b1 = 1 To dgv1.Rows.Count - 1
                dgv1.Item(a, b1).Tag = "True"
            Next
        Next


        rtbResult.Text = " ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++" &
vbNewLine & vbNewLine
        For a = 1 To dgv1.Columns.Count - 1

            rtbResult.Text = rtbResult.Text & "Attribute " & dgv1.Columns(a).HeaderText
& vbNewLine & vbNewLine & vbNewLine
            maxAttValue = 1

' fnding the max number variable for each attribute
            For b1 = 1 To dgv1.Rows.Count - 1
                For b2 = 0 To (b1 - 1)
                    If dgv1.Item(a, b1).Value = dgv1.Item(a, b2).Value Then
                        dgv1.Item(a, b1).Tag = "False"
                    End If
                Next
                If dgv1.Item(a, b1).Tag <> "False" Then
                    maxAttValue = maxAttValue + 1
                    trying = maxAttValue
                End If
            Next

'change item tag back to "True"
            For c1 = 0 To dgv1.Rows.Count - 1
                dgv1.Item(a, c1).Tag = "True"
            Next

'write in richtextbox the (X=att) and U/IND equation for all attribute
            UInd = ""
            For b = 1 To maxAttValue
                For c1 = 0 To dgv1.Rows.Count - 1
                    If dgv1.Item(a, c1).Tag = "True" Then
```

```
                        rtbResult.Text = rtbResult.Text & "X(" &
dgv1.Columns(a).HeaderText & " = " & dgv1.Item(a, c1).Value & ") =  { "
                        UInd = UInd & "{ "
                        For c2 = c1 To dgv1.Rows.Count - 1
                            If dgv1.Item(a, c2).Value = dgv1.Item(a, c1).Value Then
                                UInd = UInd & dgv1.Item(0, c2).Value
                                rtbResult.Text = rtbResult.Text & dgv1.Item(0, c2).Value
& ", "
                                dgv1.Item(a, c2).Tag = "False"
                            End If
                        Next
                        UInd = UInd & " }, " & vbNewLine
                        rtbResult.Text = rtbResult.Text & "}" & vbNewLine & vbNewLine
                    End If
                Next
                b = b + 1
            Next
            rtbResult.Text = rtbResult.Text & "U/IND(" & dgv1.Columns(a).HeaderText &
")" & " = { " & vbNewLine & UInd & " }" & vbNewLine & vbNewLine
            rtbResult.Text = rtbResult.Text & "
++++++++++++++++++++++++++++++++++++++++++++++++++++++++" & vbNewLine & vbNewLine
        Next

    End Sub

    Private Sub btnUandA_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnUandA.Click

'   ##########  listing element of U :  ##########
        rtbResult.Text = "U = " & vbNewLine & "{ "
        For a = 0 To dgv1.Rows.Count - 2
            rtbResult.Text = rtbResult.Text & CStr(dgv1.Item(0, a).Value) & ", "
        Next
        rtbResult.Text = rtbResult.Text & CStr(dgv1.Item(0, dgv1.Rows.Count - 1).Value)
& " }"

        rtbResult.Text = rtbResult.Text & vbNewLine & vbNewLine & vbNewLine

'   ##########  listing attributes :    ##########
        rtbResult.Text = rtbResult.Text & "Attributes = " & vbNewLine & "{ "
        For a = 1 To dgv1.Columns.Count - 2
            rtbResult.Text = rtbResult.Text & CStr(dgv1.Columns(a).HeaderText) & ", "
        Next
        rtbResult.Text = rtbResult.Text & CStr(dgv1.Columns(dgv1.Columns.Count -
1).HeaderText) & " }"
```

```
    End Sub

    Private Sub btnR_Click(ByVal sender As System.Object, ByVal e As System.EventArgs)
Handles btnR.Click

        Dim roughness(dgv1.Columns.Count - 1, dgv1.Columns.Count - 1, wholeMaxValue) As
Double
        Dim totalRoughness(dgv1.Columns.Count - 1, dgv1.Columns.Count - 1) As Double
        rtbRes2.Text = "++++++++++++++++++++++++++++++++" & vbNewLine
        For a = 1 To dgv1.Columns.Count - 1
            rtbRes2.Text = rtbRes2.Text & vbNewLine & vbNewLine &
dgv1.Columns(a).HeaderText & vbNewLine & vbNewLine
            rtbRes2.Text = rtbRes2.Text & "________________________________" & vbNewLine
            For a2 = 1 To dgv1.Columns.Count - 1
                totalRoughness(a - 1, a2 - 1) = 0
                If a <> a2 Then
                    rtbRes2.Text = rtbRes2.Text & vbNewLine & "Attribute " &
dgv1.Columns(a).HeaderText & " with respect to " & dgv1.Columns(a2).HeaderText &
vbNewLine
                    For b = 1 To maxAttVar(a - 1)
                        roughness(a - 1, a2 - 1, b - 1) = (attlowUpCount(a - 1, a2 - 1,
0, b - 1) / attlowUpCount(a - 1, a2 - 1, 1, b - 1))
                        rtbRes2.Text = rtbRes2.Text & "R wrt " &
dgv1.Columns(a2).HeaderText & "( X|" & dgv1.Columns(a).HeaderText & " : " &
attVariable(a - 1, b - 1) & " ) = " & CStr(attlowUpCount(a - 1, a2 - 1, 0, b - 1)) & "/"
& CStr(attlowUpCount(a - 1, a2 - 1, 1, b - 1)) & " = " & Format(roughness(a - 1, a2 - 1,
b - 1), "0.0000") & vbNewLine
                        totalRoughness(a - 1, a2 - 1) = totalRoughness(a - 1, a2 - 1) +
roughness(a - 1, a2 - 1, b - 1)
                    Next
                    meanRoughness(a - 1, a2 - 1) = totalRoughness(a - 1, a2 - 1) /
maxAttVar(a - 1)
                    maxR(a - 1, a2 - 1) = totalRoughness(a - 1, a2 - 1) / maxAttVar(a -
1)
                    rtbRes2.Text = rtbRes2.Text & vbNewLine & "Mean Roughness  = " &
CStr(totalRoughness(a - 1, a2 - 1)) & " / " & CStr(maxAttVar(a - 1)) & " = " &
Format(meanRoughness(a - 1, a2 - 1), "0.0000")
                    rtbRes2.Text = rtbRes2.Text & vbNewLine &
"________________________________" & vbNewLine
                ElseIf a = a2 Then
                    maxR(a - 1, a2 - 1) = 0
                End If
            Next
```

```
            rtbRes2.Text = rtbRes2.Text & vbNewLine & "++++++++++++++++++++++++++++++++"
& vbNewLine
        Next

    End Sub

    Private Sub btnMMR_Click(ByVal sender As System.Object, ByVal e As System.EventArgs)
Handles btnMMR.Click

        Dim mmr, temp, biggest As Double
        Dim first As Boolean
        Dim one As String = "false"
        Dim maxCount(dgv1.Columns.Count - 1), maxColumn(1) As Integer

'write n txtbox mean roughnesses for each attribute.
        rtbRes3.Text = ""
        For a = 1 To dgv1.Columns.Count - 1
            rtbRes3.Text = rtbRes3.Text & "Attribute " & dgv1.Columns(a).HeaderText &
" : Mean Roughnesses : " & vbNewLine
            first = True
            For a2 = 1 To dgv1.Columns.Count - 1
                If a <> a2 Then
                    If first = True Then
'maxR(0, a - 1) = meanRoughness(a - 1, a2 - 1)
                        rtbRes3.Text = rtbRes3.Text & ""
                        first = False
                    Else
                        rtbRes3.Text = rtbRes3.Text & " ; "
                    End If
                    rtbRes3.Text = rtbRes3.Text & Format(meanRoughness(a - 1, a2 - 1),
"0.0000")
                End If
            Next
            rtbRes3.Text = rtbRes3.Text & vbNewLine & vbNewLine
        Next

'sorting from the biggest mean roughness to the lowest
        For a = 1 To dgv1.Columns.Count - 1
            For a2 = 1 To dgv1.Columns.Count - 1
                biggest = a2
                For b = a2 + 1 To dgv1.Columns.Count - 1
                    If maxR(a - 1, b - 1) > maxR(a - 1, biggest - 1) Then
                        biggest = b
                    End If
                Next
                temp = maxR(a - 1, a2 - 1)
```

```
                maxR(a - 1, a2 - 1) = maxR(a - 1, biggest - 1)
                maxR(a - 1, biggest - 1) = temp
            Next
        Next

        rtbRes3.Text = rtbRes3.Text & "++++++++++++++++++++++++++++++++++" & vbNewLine &
vbNewLine & "Maximum mean roughnesses = " & vbNewLine

        first = True
        For a = 1 To dgv1.Columns.Count - 1
            If first = True Then
                rtbRes3.Text = rtbRes3.Text & ""
                first = False
            Else
                rtbRes3.Text = rtbRes3.Text & " ; "
            End If
            rtbRes3.Text = rtbRes3.Text & Format(maxR(a - 1, 0), "0.0000")
        Next

        mmr = 0

' find the max out of the max mean roughnesses
        For a = 1 To dgv1.Columns.Count - 1
            If maxR(a - 1, 0) > mmr Then
                mmr = maxR(a - 1, 0)
                maxColumn(0) = a
            End If
        Next

'count how many occurence of the mmr
        maxCount(0) = 0
        For a = 1 To dgv1.Columns.Count - 1
            If mmr = maxR(a - 1, 0) Then
                maxCount(0) = maxCount(0) + 1
                ReDim Preserve maxColumn(maxCount(0))
                maxColumn(maxCount(0) - 1) = a
            End If
        Next

' in case of 2 or more same mmr
        For a2 = 2 To dgv1.Columns.Count - 1
            If maxCount(a2 - 2) > 1 Then
                mmr = 0
'find mmr from the other value
                For b = 1 To maxCount(a2 - 2)
                    If maxR(maxColumn(b - 1) - 1, a2 - 1) > mmr Then
```

```
                        rtbRes2.Text = rtbRes2.Text & maxColumn(b - 1) & " | "
                        mmr = maxR(maxColumn(b - 1) - 1, a2 - 1)
                    End If
                Next

'count how many occurence of the mmr from the other value
                maxCount(a2 - 1) = 0
                For b = 1 To maxCount(a2 - 2)
                    If mmr = maxR(maxColumn(b - 1) - 1, a2 - 1) Then
                        maxCount(a2 - 1) = maxCount(a2 - 1) + 1
                        maxColumn(maxCount(a2 - 1) - 1) = maxColumn(b - 1)
                    End If
                Next
                ReDim Preserve maxColumn(maxCount(a2 - 1))
            End If
        Next

        rtbFinal.Text = "MMR = " & Format(mmr, "0.0000") & vbNewLine & "Splitting
Attribute = " & dgv1.Columns(maxColumn(0)).HeaderText


''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
'change item tag back to "True"
        For c1 = 0 To dgv1.Rows.Count - 1
            dgv1.Item(maxColumn(0), c1).Tag = "True"
        Next
        For b = 1 To trying 'maxAttValue
            For c1 = 0 To dgv1.Rows.Count - 1
                If dgv1.Item(maxColumn(0), c1).Tag = "True" Then
                    rtbFinal.Text = rtbFinal.Text & vbNewLine & "X(" &
dgv1.Columns(maxColumn(0)).HeaderText & " = " & dgv1.Item(maxColumn(0), c1).Value & ") =
{ "
                    UInd = UInd & "{ "
                    For c2 = c1 To dgv1.Rows.Count - 1
                        If dgv1.Item(maxColumn(0), c2).Value = dgv1.Item(maxColumn(0),
c1).Value Then
                            UInd = UInd & dgv1.Item(0, c2).Value & ", "
                            rtbFinal.Text = rtbFinal.Text & dgv1.Item(0, c2).Value & ","
                            dgv1.Item(maxColumn(0), c2).Tag = "False"
                        End If
                    Next
                    UInd = UInd & " }, " & vbNewLine
                    rtbFinal.Text = rtbFinal.Text & "}"
                End If
            Next
```

```
            rtbResult.Text = rtbResult.Text & "U/IND(" &
dgv1.Columns(maxColumn(0)).HeaderText & ")" & " = { " & vbNewLine & UInd & " }" &
vbNewLine & vbNewLine

        Next
'''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
    End Sub

    Private Sub frmStart_Load(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles MyBase.Load

    End Sub
End Class
```

TURNITIN ORIGINALITY REPORT

Turnitin Originality Report

check draft by MOHD AMIROL REDZUAN MAT ROFI

From thesis (PSM2)

- Processed on 21-May-2012 06:17 MYT
- ID: 250586539
- Word Count: 16625

Similarity Index

43%

Similarity by Source

Internet Sources:	36%
Publications:	28%
Student Papers:	22%

sources:

1

3% match (Internet from 11/24/06)

http://www.healthcare.frost.com/prod/servlet/market-insight-top.pag?docid=18817766

2

3% match (Internet from 7/20/10)

http://www.aaa7.org/newsletters/Service%20Coordination/2010%20Service%20Coordination/SC%2002%2010.pdf

3

2% match (Internet from 3/15/10)

http://www.arbuz.com/showthread.php?t=55380

4

2% match (Internet)

http://www.ticam.utexas.edu/~zhangyf/paper/validity_survey.pdf

5

2% match (Internet from 11/5/02)

http://www.research.ibm.com/compsci/kdd/brochure.html

6

2% match (Internet from 11/2/02)

http://www.cs.uregina.ca/~hamilton/courses/831/notes/kdd/1_kdd.html

7

2% match (Internet from 1/23/07)

http://eprints.iisc.ernet.in/archive/00000273/01/p264-jain.pdf

8

2% match (student papers from 04/29/11)

Submitted to King Saud University on 2011-04-29

9

1% match (Internet from 8/30/10)

http://dblab.mgt.ncu.edu.tw/%E6%95%99%E6%9D%90/2010%20DM/34.pdf

10

1% match (publications)

Shyng, J.Y.. "Rough Set Theory in analyzing the attributes of combination values for the insurance market", Expert Systems With Applications, 200701

11

1% match (Internet from 4/19/12)

http://www.subnetwork.sexzoznamka.info/p-Fuzzy_set

12

1% match (Internet from 4/8/09)

http://circ.ahajournals.org/cgi/content/extract/118/25/2748

13

1% match (Internet)

http://www.cis.gsu.edu/~ghubona/NeuralInternet/p264-jain.pdf

14

1% match (Internet from 11/4/02)

http://www.tgc.com/dsstar/00/0704/101861.html

15

1% match (publications)

Norhalina Senan. "Rough Set Approach for Attributes Selection of Traditional Malay Musical Instruments Sounds Classification", Communications in Computer and Information Science, 2011

16

1% match (publications)

C.L. Carter. "Efficient attribute-oriented generalization for knowledge discovery from large databases", IEEE Transactions on Knowledge and Data Engineering, 1998

17

1% match (Internet from 9/16/05)

http://logic.mimuw.edu.pl/prace/1999/nd2/pawlak_skowron_Bull_IRSS.ps

18

1% match (publications)

Pawlak, Z.. "Rough classification", International Journal of Man-Machine Studies, 198405

19

1% match (student papers from 02/10/12)

Submitted to The Hong Kong Polytechnic University on 2012-02-10

20

1% match (publications)

Syed Sibte Raza Abidi. "Analyzing Data Clusters: A Rough Sets Approach to Extract Cluster-Defining Symbolic Rules", Lecture Notes in Computer Science, 2001

21

1% match (publications)

Shangzhi Wu. "An Algorithm for Clustering Data Based on Rough Set Theory", 2008 International Symposium on Information Science and Engineering, 12/2008

22

1% match (Internet from 7/13/10)

http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/frawley92.pdf

23

< 1% match (publications)

Kumar, P.. "Rough clustering of sequential data", Data & Knowledge Engineering, 200711

24

< 1% match (publications)

Tutut Herawan. "A Direct Proof of Every Rough Set is a Soft Set", 2009 Third Asia International Conference on Modelling & Simulation, 05/2009

25

< 1% match (Internet from 7/14/10)

http://www2.cs.uregina.ca/~yyao/PAPERS/is_comparison.pdf

26

< 1% match (publications)

JAIN, A.K. MURTY, M.N. FLYNN, P.J.. "Data Clustering: A Review.", ACM Computing Surveys, Sept 1999 Issue

27

< 1% match (Internet from 9/14/09)

http://www.cs.laurentian.ca/julia/rough/ApplicationsOfRST.ppt

28

< 1% match (student papers from 08/30/07)

Submitted to New Art College on 2007-08-30

29

< 1% match (Internet from 3/9/12)

http://westernct.alwaysbestcare.com/monthlynewsletter/monthly-newsletter-february-2011.aspx

30

< 1% match (publications)

Zdzisław Pawlak. "Rough Set Theory", Wiley Encyclopedia of Computer Science and Engineering, 12/14/2007

31

< 1% match (Internet from 6/3/10)

http://logic.mimuw.edu.pl/publikacje/ZP-AS-InfSci07-1.pdf

32

< 1% match (publications)

Tutut Herawan. "Rough Set Theory for Topological Spaces in Information Systems", 2009 Third Asia International Conference on Modelling & Simulation, 05/2009

33

< 1% match (Internet)

http://lrd.yahooapis.com/_ylc=X3oDMTVndGl1ZjV1BF9TAzIwMjMxNTI3MDIEYXBwaWQDTHJlazRUTFYzNEdRVjYwVDFRYVlHeC5xMDYuMHVja2pJb3dfYzJFV3NGejhWZzVHX2xkQjRPX1YweDZPdVNOME9zVjg2a0I2BGNsaWVudANib3NzBHNlcnZpY2UDQk9TUwRzbGsDdGl0bGUEc3JjcHZpZANlX1VMc0VnZUF1MWFyRENITTdQbmdLdzlKbS5UVGt2YVJZQUFCSmk1/SIG=13tv36moh/**http%3A//amrita.edu/rtcsp/papers/Hall-4/S24/3%2520A%2520Financial%2520Trading%2520System%2520using%2520ROUGH_SET_PAPER.pdf

34

< 1% match (Internet from 8/27/10)

http://sist.swjtu.edu.cn/imc/itw06/rskt2008/six.pdf

35

< 1% match (Internet from 6/3/10)

http://logic.mimuw.edu.pl/publikacje/AS-art-Eservice-Book06.pdf

36

< 1% match (publications)

Herawan, Tutut. "Rough Set Approach for Categorical Data Clustering", International Journal of Database Theory & Application/20054270, 20100301

37

< 1% match (Internet from 4/11/10)

http://adultmatchfirm.com/intgroups/mi33226563_95421/rhverriberri/acrlist.html?dcb=adultmatchfirm

38

< 1% match (publications)

Andrzej Skowron. "Approximate Reasoning in MAS: Rough Set Approach", 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, 12/2006

39

< 1% match (student papers from 10/04/06)

Submitted to (school name not available) on 2006-10-04

40

< 1% match (publications)

Herawan, T.. "A rough set approach for selecting clustering attribute", Knowledge-Based Systems, 201004

41

< 1% match (Internet from 1/12/10)

http://www.icis.ntu.edu.sg/scs-ijit/1203/1203_15.pdf

42

< 1% match (Internet from 7/12/10)

http://tribals.com/browsemembers.cfm?pagenumber=11&start=251&c=78

43

< 1% match (Internet from 9/13/10)

http://www.sersc.org/journals/IJDTA/vol3_no2/5.pdf

44

< 1% match (publications)

Choochart Haruechaiyasak. "A web-page recommender system via a data mining framework and the Semantic Web concept", International Journal of Computer Applications in Technology, 2006

45

< 1% match (Internet from 3/18/10)

http://www.adgam.com.pl/aj/16.pdf

46

< 1% match (publications)

U. Fayyad. "Data mining and knowledge discovery in databases: implications for scientific databases", Proceedings Ninth International Conference on Scientific and Statistical Database Management (Cat No 97TB100150) SSDM-97, 1997

47

< 1% match (Internet from 8/31/11)

http://icarzcu3.gov.in/book_cover/Concepts_2nd_Edn/10Chapter%209.pdf

48

< 1% match (Internet from 2/18/12)

http://www.enotes.com/topic/Knowledge_discovery

49

< 1% match (Internet from 5/21/11)

http://sci2s.ugr.es/keel/pdf/specific/congreso/ieee31.pdf

50

< 1% match (publications)

Wei-Zhi Wu. "Knowledge Acquisition in Incomplete Information Systems Based on Variable Precision Rough Set Model", 2005 International Conference on Machine Learning and Cybernetics, 2005

51

< 1% match (publications)

Tutut Herawan. "Soft Decision Making for Patients Suspected Influenza", Lecture Notes in Computer Science, 2010

52

< 1% match (publications)

Jerzy W. Grzymala-Busse. "Knowledge acquisition under uncertainty — a rough set approach", Journal of Intelligent and Robotic Systems, 1988

53

< 1% match (publications)

Awang Mohd Isa. "Dominance-Based Soft Set Approach in Decision-Making Analysis", Lecture Notes in Computer Science, 2011

54

< 1% match (Internet from 11/3/11)

http://www.cholestrol.info/otherresources/browse/31/Default.aspx?cthelium.com/topic/5395-heart-disease

55

< 1% match (publications)

Jerzy W. Grzymala-Busse. "Increasing Incompleteness of Data Sets—A Strategy for Inducing Better Rule Sets", Studies in Computational Intelligence, 2010

56

< 1% match (publications)

K.S. Trivedi. "Analysis of software cost models with rejuvenation", Proceedings Fifth IEEE International Symposium on High Assurance Systems Engineering (HASE 2000) HASE-00, 2000

57

< 1% match (student papers from 12/06/08)

Submitted to Boston University on 2008-12-06

58

< 1% match (Internet from 11/5/10)

http://grammars.grlmc.com/GRLMC/reports/rep29.doc

59

< 1% match (publications)

Iwan Tri Riyadi Yanto. "A Framework of Rough Clustering for Web Transactions", Studies in Computational Intelligence, 2010

60

< 1% match (Internet from 12/25/09)

http://www.forums.ffshrine.org/forumdisplay.php?s=&f=2&daysprune=-1&order=desc&sort=views

61

< 1% match (Internet from 11/20/06)

http://www2.ceri.go.jp/eng/its-win/pdf/12e.pdf

62

< 1% match (publications)

Ho-Jin Choi. "Design of a Clinical Knowledge Base for Heart Disease Detection", 7th IEEE International Conference on Computer and Information Technology (CIT 2007), 10/2007

63

< 1% match (Internet from 7/3/09)

http://www.dbs.informatik.uni-muenchen.de/~breunig/HomepageResearch/research.html

64

< 1% match (publications)

Shi, F.. "Employing rough sets and association rule mining in KANSEI knowledge extraction", Information Sciences, 20120801

65

< 1% match (publications)

Norhalina Senan. "Rough Set Theory for Feature Ranking of Traditional Malay Musical Instruments Sounds Dataset", Communications in Computer and Information Science, 2011

66

< 1% match (student papers from 01/31/11)

Submitted to Rochester Institute of Technology on 2011-01-31

67

< 1% match (Internet from 11/15/06)

http://aaai.org/AITopics/assets/PDF/AIMag17-03-2-article.pdf

68

< 1% match (publications)

"Preparing Your Media in iTunes", Taking Your iPod touch to the Max, 2008

69

< 1% match (Internet from 4/12/10)

http://www.ischool.drexel.edu/faculty/thu/IJGCRSIS%20V%201%20N%201%202009%20-%20Final%20pdf.pdf

70

< 1% match (Internet from 2/5/09)

http://www2.upb.de/Verwaltung/sachgebiete/3.1/47.07%20BA%20Europ%C3%A4ische%20Studien.pdf

71

< 1% match (publications)

Salvatore Greco. "Rough Sets and Gradual Decision Rules", Lecture Notes in Computer Science, 2003

72

< 1% match (publications)

Yang, X.. "Credible rules in incomplete decision system based on descriptors", Knowledge-Based Systems, 200901

73

< 1% match (student papers from 09/01/10)

Submitted to Universiti Teknologi Malaysia on 2010-09-01

74

< 1% match (Internet from 6/23/09)

http://www.linkedin.com/pub/javier-iribarne/4/402/88

75

< 1% match (Internet from 4/16/10)

http://www.iaeng.org/publication/WCE2007/WCE2007_pp243-248.pdf

76

< 1% match (Internet from 7/20/06)

http://www.cirm.ca.gov/transcripts/pdf/2005/02-03-05.pdf

77

< 1% match (Internet from 12/9/11)

http://www.tabspace.com/wherever-i-may-roam-live-intro-guitar-tabs-metallica.html

78

< 1% match (publications)

Jerzy W. Grzymala-Busse. "Three Approaches to Missing Attribute Values: A Rough Set Perspective", Studies in Computational Intelligence, 2008

79

< 1% match (publications)

Andrew C. Steer. "Prevention and treatment of rheumatic heart disease in the developing world", Nature Reviews Cardiology, 09/15/2009

80

< 1% match (Internet from 6/3/10)

http://www2.cs.uregina.ca/~yyao/PAPERS/is_belief.pdf

81

< 1% match (publications)

Herawan, Tutut. "Soft Set Theoretic Approach for Dimensionality Reduction", International Journal of Database Theory & Application/20054270, 20100601

82

< 1% match (Internet from 3/24/12)

http://ethesys.lib.mcu.edu.tw/ETD-db/ETD-search-c/view_etd?URN=etd-0824109-192932

83

< 1% match (Internet)

http://informatik.uni-trier.de/~ley/db/conf/ida/ida2001.html

84

< 1% match (publications)

Sang C. Suh. "Extraction of Meaningful Rules in a Medical Database", 2008 Seventh International Conference on Machine Learning and Applications, 12/2008

85

< 1% match (publications)

Tutut Herawan. "Soft Set Theoretic Approach for Dimensionality Reduction", Communications in Computer and Information Science, 2009

86

< 1% match (publications)

Liping An. "Binary Relations as a Basis for Rule Induction in Presence of Quantitative Attributes", Journal of Computers/1796203X, 20100301

87

< 1% match (publications)

D.K. Christodoulakis. "State synchronization in local-interaction networks is robust with respect to time delays", Proceedings of the 44th IEEE Conference on Decision and Control, 2005

88

< 1% match (Internet from 5/29/09)

http://ie.fulton.asu.edu/files/shared/workingpapers/MMRSupplier_IEEEonEM.pdf

89

< 1% match (publications)

Liu, G.. "The algebraic structures of generalized rough set theory", Information Sciences, 20081101

90

< 1% match (Internet from 5/1/10)

http://biomed.umit.at/upload/mscarbeti_mosl_fullversion.pdf

91

< 1% match (Internet from 1/9/09)

http://dblab.mgt.ncu.edu.tw/Main/%E8%AB%96%E6%96%87%E5%88%86%E9%A1%9E/%E9%A1%9E%E5%88%A5/Cluster%20Analysis.asp

92

< 1% match (Internet from 4/19/09)

http://www.saildart.org/prog/CSR/USE_CSR/.html/000037?36,218855

93

< 1% match (publications)

Osmar R. Zaïane. "On Data Clustering Analysis: Scalability, Constraints, and Validation", Lecture Notes in Computer Science, 2006

94

< 1% match (student papers from 05/18/10)

Submitted to Universiti Teknologi Petronas on 2010-05-18

95

< 1% match (Internet from 9/5/10)

http://www2.cs.uregina.ca/~yyao/PAPERS/book_generalization.pdf

96

< 1% match (Internet)

http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/rs/wp/KMITL/data/bangkok/96-02/bg960216.txt

97

< 1% match (Internet from 1/28/12)

http://www.must.edu.my/index.php?option=com_content&view=article&id=75&Itemid=95

98

< 1% match (Internet from 11/5/03)

http://l1.lamsade.dauphine.fr/mcda/biblio/Author/PAWLAK-Z.html

99

< 1% match (publications)

Lustiana Pratiwi. "An Empirical Study of Density and Distribution Functions for Ant Swarm Optimized Rough Reducts", Communications in Computer and Information Science, 2011

100

< 1% match (publications)

Charu C. Aggarwal. "A Framework for Projected Clustering of High Dimensional Data Streams", (CrossRef) NoVol ISBN 9780120884698, 2004

101

< 1% match (publications)

Tutut Herawan. "Matrices Representation of Multi Soft-Sets and Its Application", Lecture Notes in Computer Science, 2010

102

< 1% match (publications)

Jin S. Lee. "Parameter conditions and least squares identification of single-input single-output convex fuzzy system", SICE Annual Conference 2007, 09/2007

103

< 1% match (publications)

Emma Chavez–Mora. "A Real Time Multiagent Information Systems Architecture for Actively Monitoring Chronic Diseases", Information Systems Development, 2011

104

< 1% match (Internet from 5/14/03)

http://www6.wind.jp/gun-tama/Kyouiku/Tamashokaiho.htm

105

< 1% match (Internet from 11/26/08)

http://www.irandoc.ac.ir/etela-art/14/14-1&2-2.htm

106

< 1% match (publications)

Degang, C.. "Rough approximations on a complete completely distributive lattice with applications to generalized rough sets", Information Sciences, 20060703

107

< 1% match (publications)

Chao Deng. "The BSC Alarm Management System Based on Rough Set Theory in Mobile Communication", 2011 Seventh International Conference on Computational Intelligence and Security, 12/2011

108

< 1% match (Internet from 10/27/11)

http://biology.mcgill.ca/faculty/bell/articles/86.KaltzBell_2002_Evolution56.pdf

109

< 1% match (Internet from 9/29/09)

http://downloads.hindawi.com/journals/ahci/2008/719291.pdf

110

< 1% match (Internet from 3/29/03)

http://www.sugardvd.com/browse/?feature=1&feature_id=18

111

< 1% match (Internet from 5/1/03)

http://www.ucheba.com/met_rus/k_upravobraz/k_vnutrikon/prof_7.htm

112

< 1% match (publications)

Andrzej Skowron. "Rough-Granular Computing", Handbook of Granular Computing, 07/07/2008

113

< 1% match (publications)

Zdzis Pawlak. "Rough Sets and Conflict Analysis", Studies in Computational Intelligence, 2007

114

< 1% match (publications)

Hong Liu. "Research of performance evaluation of course design based on rough set theory", 2009 4th International Conference on Computer Science & Education, 07/2009

115

< 1% match (publications)

Jerzy W. Grzymala-Busse. "Increasing Data Set Incompleteness May Improve Rule Set Quality", Communications in Computer and Information Science, 2009

116

< 1% match (student papers from 12/11/08)

Submitted to Franklin University on 2008-12-11

117

< 1% match (student papers from 10/09/07)

Submitted to University of South Dakota on 2007-10-09

118

< 1% match (Internet from 8/17/07)

http://drtc.isibang.ac.in/~aparajita/project/project%20appu/Project/Chapter2.doc

119

< 1% match (Internet from 5/21/11)

http://fuzzylab.et.ntust.edu.tw/e_papers.html

120

< 1% match (Internet)

http://www.certifiedsociety.com/registry/comics/set_score_detail.asp?SetTypeID=63

121

< 1% match (Internet from 6/25/06)

http://nomadman.blogchina.com/blog/4409365.html

122

< 1% match (Internet from 8/27/10)

http://www.telefilmworld.com/tag/amici-8/

123

< 1% match (publications)

Xiaoming Xu. "Feature Selection via Modified RSBRA for SVM Classifiers", Proceedings of the 2005 American Control Conference 2005, 2005

124

< 1% match (publications)

Kalla, P.. "Strategies for solving the Boolean satisfiability problem using binary decision diagrams", Journal of Systems Architecture, 200106

125

< 1% match (Internet from 12/3/10)

http://heartdisease.about.com/?once=true&

126

< 1% match (Internet from 12/26/06)

http://noikosar.hu/game/stat.phtml?year=x0405&lea=whunf&code=n1af1093

127

< 1% match (Internet)

http://www.nazpei.com/HTMLobj-809/C__CHURCH_Role_of_Laity_SurveyResults_1_Age_Under_30_Frequency_Distributions.pdf

128

< 1% match (publications)

Dai, J.. "Approximations and uncertainty measures in incomplete information systems", Information Sciences, 20120901

129

< 1% match (publications)

Jiadong Ren. "A Weighted Subspace Clustering Algorithm in High-Dimensional Data Streams", 2009 Fourth International Conference on Innovative Computing Information and Control (ICICIC), 12/2009

130

< 1% match (Internet from 9/16/05)

http://logic.mimuw.edu.pl/prace/1999/D5/Tutor06_09.ps

131

< 1% match (Internet from 12/18/06)

http://cs836.cinatech.com/class_paper/skowron.pdf

132

< 1% match (Internet from 3/19/10)

http://bcam.berkeley.edu/ARCHIVE/theses/Ning_Ma_Dissertation_Spring_2009_final.pdf

133

< 1% match (Internet from 10/23/10)

http://www.yangsky.org/ijcc/pdf/ijcc33a.pdf

134

< 1% match (Internet from 5/17/10)

http://www.cs.uky.edu/~goldsmit/cdm/DMCDSS.pdf

135

< 1% match (Internet)

http://www.transport-ticket.org.uk/pdf/LTTH.pdf

136

< 1% match (publications)

Hongwu Qin. "A Soft Set Model on Information System and Its Application in Clustering Attribute Selection", Communications in Computer and Information Science, 2011

137

< 1% match (publications)

El-Monsef, M.E. Abd Abu-Donia, H.M. Mare. "Special approach to near set theory.(Research Article)(Report)", Mathematical Problems in Engineering, Annual 2011 Issue

138

< 1% match (publications)

C. I. Faustino Agreira. "The Rough Set Theory applied to a set of the new severity indices", 2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, 06/2010

139

< 1% match (publications)

Wei-Zhi Wu. "Connections between rough set theory and Dempster-Shafer theory of evidence", International Journal of General Systems, 1/1/2002

140

< 1% match (student papers from 11/07/11)

Submitted to University of South Alabama on 2011-11-07

141

< 1% match (Internet from 4/16/12)

http://heartdisease.about.com/od/coronaryarterydisease/a/heartsymptoms.htm

142

< 1% match (Internet from 10/7/10)

http://www.db-net.aueb.gr/index.php/corporate/content/download/394/1790/version/1/file/HBV_JIIS01.pdf

143

< 1% match (Internet from 3/19/10)

http://eprints.utm.my/4066/1/MasurahMohamadMCD2006TTT.pdf

144

< 1% match (Internet from 4/27/11)

http://www.slideshare.net/Tommy96/generalized-densitybased-clustering-for-spatial-data-mining

145

< 1% match (Internet)

http://www.cprn.ca/docs/work/cec_e.pdf

146

< 1% match (Internet)

http://www.crashdaviscentral.com/html/team22b.html

147

< 1% match (Internet from 1/14/07)

http://www.stes.es/traslados/pvalencia/cv_codis_castello.pdf

148

< 1% match (publications)

Qiang Shen. "Rough sets, their extensions and applications", International Journal of Automation and Computing, 07/2007

149

< 1% match (publications)

J. Zalewski. "Software safety analysis using rough sets", Proceedings IEEE Southeastcon 98 Engineering for a New Era SECON-98, 1998

150

< 1% match (publications)

Liu, D.. "Probabilistic model criteria with decision-theoretic rough sets", Information Sciences, 20110901

151

< 1% match (publications)

Guoyin Wang. "Research on Rough Set Theory and Applications in China", Lecture Notes in Computer Science, 2008

152

< 1% match (publications)

Youness, E.A.. "Characterizing solutions of rough programming problems", European Journal of Operational Research, 20060201

153

< 1% match (publications)

Kok Meng Hoe. "Symbolic exposition of medical data-sets: a data mining workbench to inductively derive data-defining symbolic rules", Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002) CBMS-02, 2002

154

< 1% match (publications)

Shyng, J.Y.. "Using FSBT technique with Rough Set Theory for personal investment portfolio analysis", European Journal of Operational Research, 20100301

155

< 1% match (publications)

Siouffi, A.M.. "Computer-assisted sample clean-up liquid chromatography from thin-layer chromatographic data", Journal of Chromatography A, 1991

156

< 1% match (publications)

Andrzej Skowron. "Rough Sets in Perception-Based Computing", Lecture Notes in Computer Science, 2005

157

< 1% match (publications)

Greco, S.. "Rough sets theory for multicriteria decision analysis", European Journal of Operational Research, 20010215

158

< 1% match (publications)

Ming He. "Research on Attribute Reduction Using Rough Neighborhood Model", 2008 International Seminar on Business and Information Management, 12/2008

159

< 1% match (publications)

B. Davvaz. "Rough approximations in a general approximation space and their fundamental properties", International Journal of General Systems, 6/2008

160

< 1% match (publications)

Xu, W.-H.. "Measuring roughness of generalized rough sets induced by a covering", Fuzzy Sets and Systems, 20071116

161

< 1% match (publications)

Qian, Y.. "Interval ordered information systems", Computers and Mathematics with Applications, 200810

162

< 1% match (publications)

"SIME DARBY WITHDRAWS FROM IJN PROJECT.", BERNAMA The Malaysian National News Agen, Jan 6 2009 Issue