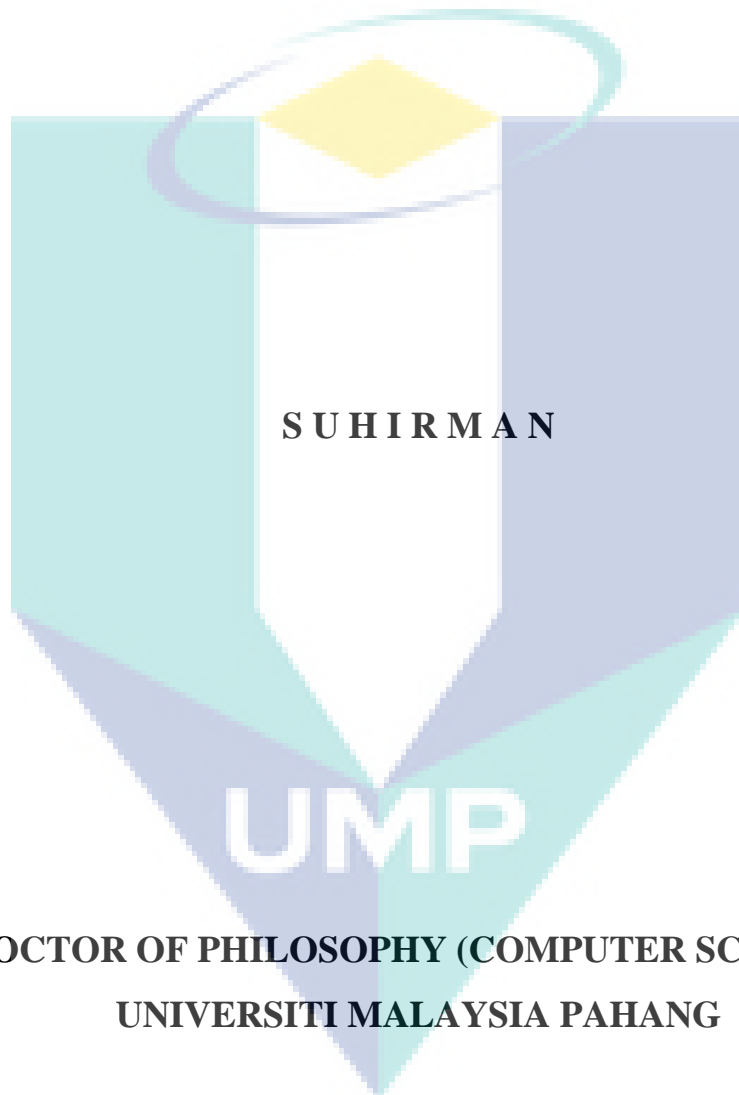
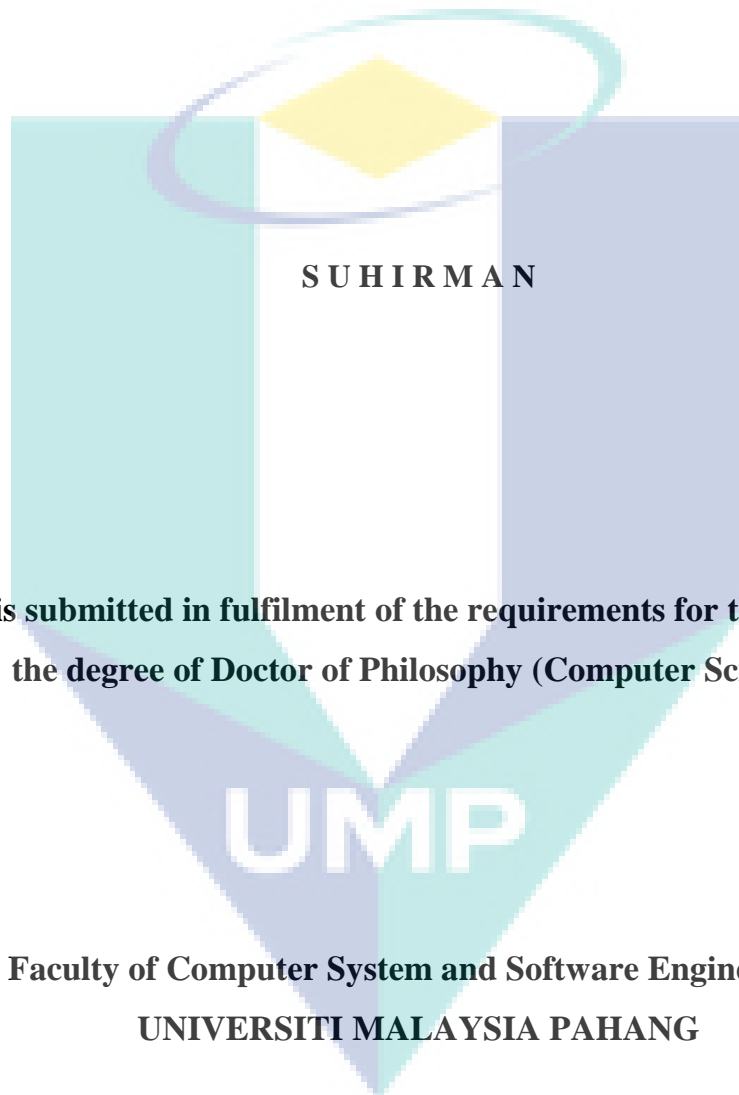


**A NEW SOFT SET-BASED TECHNIQUE FOR CLUSTERING
ATTRIBUTE SELECTION IN EDUCATIONAL
DATA MINING**



**DOCTOR OF PHILOSOPHY (COMPUTER SCIENCE)
UNIVERSITI MALAYSIA PAHANG**

**A NEW SOFT SET-BASED TECHNIQUE FOR CLUSTERING
ATTRIBUTE SELECTION IN EDUCATIONAL
DATA MINING**



**Thesis submitted in fulfilment of the requirements for the award of
the degree of Doctor of Philosophy (Computer Science)**

**Faculty of Computer System and Software Engineering
UNIVERSITI MALAYSIA PAHANG**

JANUARY 2016

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's full name : Suhirman
Date of birth : 14th April 1971
Title : A New Soft Set-based Technique for Clustering Attribute Selection
In Education Data Mining
Academic Session : 2012-2015

I declare that this thesis is classified as:

- ☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
- ☐ **RESTRICTED** (Contains restricted information as specified by the organization where research was done)*
- ☐ **OPEN ACCESS** I agree that my thesis to be published as online open access (Full text)

I acknowledge that Universiti Malaysia Pahang reserve the right as follows:

1. The Thesis is the Property of University Malaysia Pahang
2. The Library of University Malaysia Pahang has the right to make copies for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified By:

(Student's Signature)

(Signature of Supervisor)

A1739252

Prof. Dr. Jasni Mohamad Zain

New IC / Passport Number

Name of Supervisor

Date :

Date :

NOTES :*If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.

SUPERVISOR'S DECLARATION

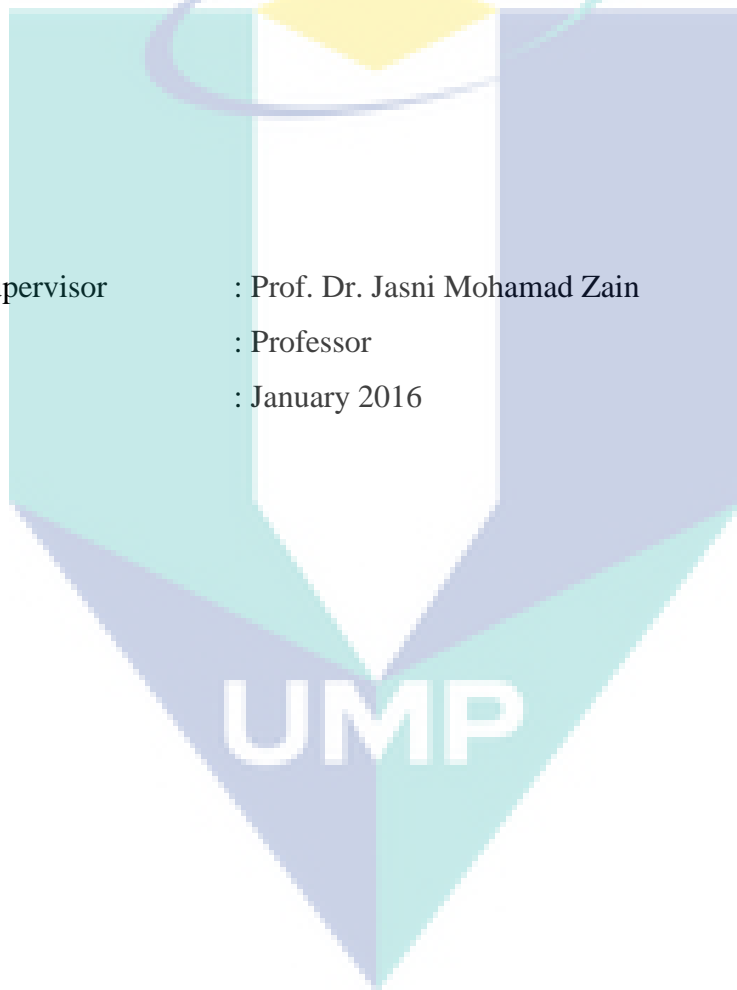
I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Doctor of Philosophy (Computer Science).

Signature

Name of Supervisor : Prof. Dr. Jasni Mohamad Zain

Position : Professor

Date : January 2016



STUDENT'S DECLARATION

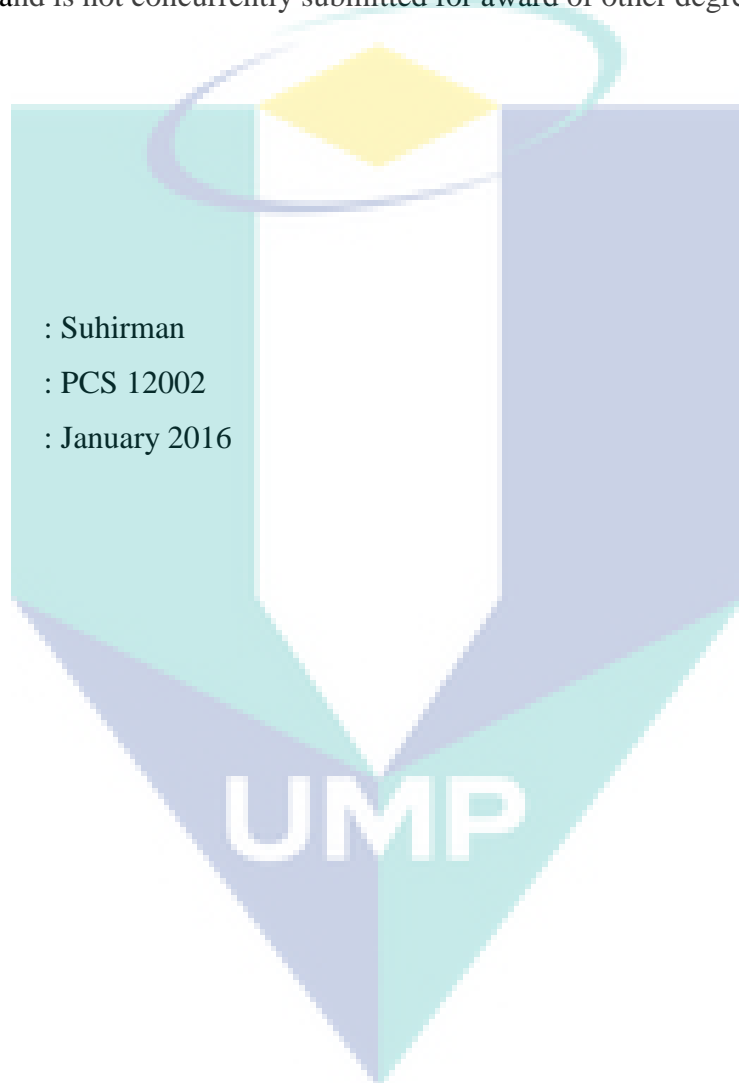
I hereby declare that the work in this thesis is my own except for the quotations and summaries which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree.

Signature

Name : Suhirman

ID Number : PCS 12002

Date : January 2016



ACKNOWLEDGEMENT

First of all, I would like to thank Allah S.W.T for giving me the strength and knowledge to complete this thesis and for blessing me with many great people who have been my greatest support in life. Shalawat and salam for the prophet Muhammad S.A.W.

I wish to express my sincere and deep gratitude to my supervisor, Prof. Dr. Jasni Mohamad Zain for her guidance, help and encouragement during the three year period of my Ph.D program. Her invaluable help of constructive comments and suggestions throughout the simulation and thesis work have contributed to the success of this research and also my to co-supervisor

Deepest gratitude goes to my parents, wife and my daughters, chancellor and staff of Universiti Malaysia Pahang, staff of University Technology of Yogyakarta, and all the comrades that are unable to be mentioned one by one. Brothers who have directly or indirectly contributed to this research. Your kindness means a lot to me.

The author realizes that in the preparation of this research proposal there are still weaknesses that need to be strengthened and shortcomings that need to be completed. Therefore, humbly I suggest that the necessary input, corrections and suggestions given will help to strengthen my weaknesses and complement the shortage.

The logo of Universiti Malaysia Pahang (UMP) is a large, stylized 'V' shape composed of several overlapping triangles in shades of teal and light blue. The letters 'UMP' are prominently displayed in white, bold, sans-serif font across the center of the 'V' shape.

UMP

ABSTRACT

Determining the best clustering attribute is an essential process in data clustering, since this task is a relatively simple and efficient for attributes-based data clustering. Five well-known rough and soft sets-based techniques for selecting a clustering attribute respectively TR, MMR, MDA, NSS, and MAR have been proposed. MAR technique achieves better computational time than that the four other aforesaid approaches. However, in reviewing MAR, execution time is still an outstanding issue, due to iteration processes in determining the relative attribute. This research proposes an alternative soft set-based technique for selecting a clustering attribute, named Maximum Degree of Domination in Soft set theory (MDDS). In this technique, the notion of multi-soft sets is firstly described. Secondly, the domination of soft sets and its degree are defined. Finally, the maximum degree of domination is used to determine the best clustering attribute. The proposed technique is examined through eighteen UCI benchmark machine learning datasets and compared with the results obtained with that of MAR. The results show that MDDS technique achieves fairly well in reducing computation time and outperforms MAR technique up to 43.99%. Furthermore, MDDS has a good scalability, i.e. the executing time of the technique tends to increase linearly as the data sizes are increased. While the accuracy of eight data sets which have a class attributes has increased 3.23%. Furthermore, the proposed MDDS technique was used to solve real world clustering problem in Educational Data Mining. The data sets were taken from a survey on a few courses at the Information Engineering and the Architecture Departments of the University Technology of Yogyakarta (UTY) Indonesia during the last 4 years. The dominant attribute of dataset assessment were determined using MDDS technique, due to its increased efficiency and accuracy, so decisions can be made faster and accurately.

ABSTRAK

Menentukan sifat kelompok terbaik adalah satu proses yang penting dalam pengelompokan data, kerana tugas ini adalah agak mudah dan cekap bagi sifat-sifat berasaskan data kelompok. Lima teknik berasaskan set terkenal kasar dan lembut untuk memilih satu sifat kelompok masing-masing adalah TR, MMR, MDA, NSS dan MAR telah dicadangkan. Teknik MAR mencapai masa pengiraan yang lebih baik daripada empat pendekatan lain yang tersebut di atas. Walau bagaimanapun, dalam mengkaji MAR, masa pengiraan masih merupakan masalah yang besar, kerana proses lelaran dalam menentukan sifat relatif. Penyelidikan ini mencadangkan set teknik alternatif yang lembut untuk memilih atribut kelompok, yang dinamakan Maximum Degree of Domination in Soft set theory (MDDS). Dalam teknik ini, pertama konsep multi soft set diterangkan. Kedua, penguasaan set lembut dan darjahnya ditakrifkan. Akhir sekali, tahap maksimum penguasaan digunakan untuk menentukan sifat kelompok terbaik. Teknik yang dicadangkan itu diperiksa melalui lapan belas mesin UCI penanda aras dataset pembelajaran dan keputusan yang diperolehi dibandingkan dengan MAR. Hasil kajian menunjukkan bahawa pencapaian teknik MDDS agak baik dalam mengurangkan masa pengiraan dan melebihi teknik MAR sehingga 43.99%. Tambahan pula, MDDS mempunyai skala yang baik, iaitu masa untuk melaksanakan teknik tersebut meningkat secara linear dengan saiz data. Malahan ketepatan lapan set data yang bersifat kelas telah meningkat 3.23%. Tambahan pula, teknik MDDS yang dicadangkan telah digunakan untuk menyelesaikan masalah kelompok dunia sebenar dalam perlombongan data pendidikan. Set data diambil daripada kajian oleh beberapa kursus di Kejuruteraan Maklumat dan Jabatan Seni Bina Universiti Teknologi Yogyakarta (UTY) Indonesia dalam tempoh 4 tahun yang lalu. Sifat dominan penilaian dataset ditentukan menggunakan teknik MDDS. Oleh kerana peningkatan kecekapan dan ketepatan, keputusan boleh dibuat dengan lebih cepat dan tepat.

TABLE OF CONTENTS

	Page
SUPERVISOR’S DECLARATION	iii
STUDENT’S DECLARATION	iv
ACKNOWLEDGEMENT	v
ABSTRACT	vi
ABSTRAK	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF SYMBOLS	xv
LIST OF ABBREVIATIONS	xvi
 CHAPTER 1 INTRODUCTION	
1.1. Knowledge Discovery in Database	1
1.1.1. Data Mining	4
1.1.2. Classification	5
1.1.3. Clustering	5
1.1.4. Association Rule Mining	6
1.2. Problem Statement	8
1.3. Research Objectives and Scope	9
1.4. Thesis Organization	9
 CHAPTER 2 LITERATURE REVIEW	
2.1. Knowledge Discovery in Databases	11
2.2. Data Mining	11
2.3. Clustering	12
2.3.1. The Basic Steps of The Clustering Process	13
2.3.2. Data Types	15
2.3.3. Objective Function	16
2.3.4. Membership	17
2.3.5. Categorization of Clustering Methods	17
2.4. Algorithms for Categorical Data Clustering	19

2.4.1. K-Mode	20
2.4.2. Squeezer	21
2.4.3. LIMBO	23
2.4.4. ROCK	24
2.4.5. QROCK	26
2.4.6. CACTUS	26
2.4.7. COOLCAT	27
2.4.8. CLOPE	29
2.4.9. CLICK	30
2.5. Genetic Clustering Algorithms	31
2.6. Rough Set-based Clustering Attribute Selection Techniques	32
2.6.1. Total Roughness (TR) Technique	33
2.6.2. Min-Min Roughness (MMR) Technique	34
2.6.3. Maximum Dependency Attribute (MDA) Technique	37
2.7. Soft Set Theory	38
2.7.1. Information Systems	38
2.7.2. Definition of Soft Set Theory	39
2.7.3. Multi Soft Sets	42
2.7.4. Applications of Soft Set Theory	44
2.8. Soft Set-based Clustering Attribute Selection Techniques	44
2.8.1. A Novel Soft Set (NSS) Technique	44
2.8.2. Maximum Attribute Relative (MAR) Technique	45
2.9. Educational Data Mining	47
2.10. Discussion	50

CHAPTER 3 PROPOSED METHOD

3.1. Soft Set Based Technique	52
3.2. Maximum Attribute Relative (MAR) Technique	54
3.2.1. The MAR Technique	55
3.2.2. Complexity of MAR	57
3.3. Maximum Degree of Domination in Soft Set Theory (MDDS) Technique	58
3.3.1. Domination in Multi Soft Sets	58
3.3.2. The MDDS Technique	59
3.3.3. Complexity of MDDS	62
3.4. Summary	63

CHAPTER 4 EXPERIMENTAL RESULTS

4.1. Experimental Design	64
4.1.1. Data Sets	64
4.1.2. Language and Platform for Implementation	69
4.2. Performance Analysis	69
4.2.1. Execution Time	70
4.2.2. Scalability	71
4.2.3. Accuracy	73
4.2.4. Iterations	74
4.3. Comparison between MAR and MDDS Techniques	76
4.4. Summary	77

CHAPTER 5. APPLICATIONS OF MDDS TECHNIQUE FOR EDUCATIONAL DATA MINING

5.1. Data Sets of Assessment	78
5.2. Data Description	79
5.3. Result and Discussion	87
5.4. Execution Time on Data Sets Assessment	107
5.5. Summary	108

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

6.1. Contributions and Limitations	109
6.2. Recommendations	111
6.3. Summary	112

REFERENCES	113
-------------------	-----

APPENDIX A LIST OF PUBLICATIONS	119
--	-----

APPENDIX B PROGRAMMING CODE	120
------------------------------------	-----

APPENDIX C1 MAR TECHNIQUE CALCULATION EXAMPLE	130
--	-----

APPENDIX C2 MDDS TECHNIQUE CALCULATION EXAMPLE	145
---	-----

LIST OF TABLES

Table No.	Title	Page
2.1	An Information System	39
2.2	Tabular representation of Soft Set	41
2.3.	Summary of clustering techniques	50
3.1	Multi-valued information system	53
3.2	Support value of MAR technique of all soft set	56
3.3	Overall max and min support from the example	57
3.4	Domination degree of all soft sets	61
4.1	The summary information about the seventeen data sets	69
4.2	Execution time of comparison results	70
4.3	Accuracy comparison results	74
4.4	Number of Iterations comparison results	75
5.1	Data sets of Assessment	78
5.2	Transformation data assessment into categorical data	79
5.3	Assessment of Algorithm course	79
5.4	Assessment of Software Engineering course	80
5.5	Assessment of System Security course	81
5.6	Assessment of File System course	81
5.7	Assessment of Artificial intelligence course	82
5.8	Assessment of Databases course	82
5.9	Assessment of Architectural Design course	83
5.10	Assessment of Infrastructure Design Studio course	84
5.11	Assessment of Architecture Studio course	85
5.12	Assessment of Contextual Studio course	86
5.13	Matrix results from Algorithm course 2011	87
5.14	Matrix results from Algorithm course 2012	88
5.15	Matrix results from Algorithm course 2013	89

5.16	Matrix results from Software Engineering course 2011	90
5.17	Matrix results from Software Engineering course 2012	90
5.18	Matrix results from Software Engineering course 2013	91
5.19	Matrix results from System Security course 2011	92
5.20	Matrix results from System Security course 2012	93
5.21	Matrix results from System Security course 2013	93
5.22	Matrix results from File System course 2011	94
5.23	Matrix results from File System course 2012	95
5.24	Matrix results from File System course 2013	96
5.25	Matrix results from Artificial Intelligence course 2014	96
5.26	Matrix results from Databases course 2014	97
5.27	Matrix results from Architectural Design course 2011	98
5.28	Matrix results from Architectural Design course 2012	99
5.29	Matrix results from Architectural Design course 2013	100
5.30	Matrix results from Infrastructure Design Studio course 2014	102
5.31	Matrix results from Architecture Studio course 2014	104
5.32	Matrix results from Contextual Studio course 2014	105
5.33	Execution time of comparison results on datasets assessment	108



UMP

LIST OF FIGURES

Figure No.	Title	Page
1.1	Relation between KDD, data mining, and clustering	1
1.2	Overview of the steps constituting the KDD process	2
1.3	Research area position	3
1.4.	Data mining and database technology	4
2.1	Fuzzy c-means algorithm	17
2.2	The Squeezer algorithm	22
2.3	The ROCK clustering algorithm	25
2.4	The QROCK algorithm	26
2.5	Incremental step of COOLCAT algorithm	28
2.6	The sketch of the CLOPE algorithm	30
2.7	Basic steps in Gas	32
2.8	The MMR algorithm	36
2.9	The MDA algorithm	38
2.10	A decomposition of a multi-valued information system	43
2.11	The NSS algorithm	45
2.12	The MAR algorithm	46
2.13	The cycle of applying data mining in educational systems	49
3.1	Multi-soft set composition from dataset	53
3.2	The MDDS algorithm	60
3.3	Result of clustering	63
4.1	The scalability of MAR and MDDS to the number of instances	72
4.2	The scalability of MAR and MDDS to the number of attributes	72
4.3	The scalability of MAR and MDDS to the number of data size	73
4.4	The accuracy of MAR and MDDS	74
4.5	Comparison of iteration number MAR and MDDS	76
5.1	User interface of the developed MDDS software	87
5.2	Clustering visualization of students on algorithm course 2011	88
5.3	Clustering visualization of students on algorithm course 2012	89
5.4	Clustering visualization of students on algorithm course 2013	89
5.5	Clustering visualization of students on software engineering course 2011	90
5.6	Clustering visualization of students on software engineering course 2012	91
5.7	Clustering visualization of students on software engineering course 2013	92

5.8	Clustering visualization of students on system security course 2011	92
5.9	Clustering visualization of students on system security course 2012	93
5.10	Clustering visualization of students on system security course 2013	94
5.11	Clustering visualization of students on file system course 2011	95
5.12	Clustering visualization of students on file system course 2012	95
5.13	Clustering visualization of students on file system course 2013	96
5.14	Clustering visualization of students on artificial intelligence course 2014	97
5.15	Clustering visualization of students on databases course 2014	98
5.16	Clustering visualization of students on architectural design course 2011	99
5.17	Clustering visualization of students on architectural design course 2012	100
5.18	Clustering visualization of students on architectural design course 2013	101
5.19	Clustering visualization of students on infrastructure design studio course 2013	104
5.20	Clustering visualization of students on architecture studio course 2014	104
5.21	Clustering visualization of students on contextual studio course 2014	107
5.22	Comparison of execution time MAR and MDDS on datasets assessment	107



UMP

LIST OF SYMBOLS

$()$	value function
$/$	division
$\{ , \}$	set brackets
$ $	the absolute value
$ $	Norm
$=$	Equal
\neq	not equal
\emptyset	Empty Association
\in	set membership
\notin	not the elements of the set
Σ	summation - sum of all values in range of series
\cap	Intersection
\cup	Union
\leq	less than or equal to
\subseteq	Subset
\rightarrow	implication
A	Attribute
f	Function
O	Time Complexity
S	Set
U	Object
V	Value
\times	Cartesian product
δ	delta function

LIST OF ABBREVIATIONS



ARM	Association Rule Mining
DM	Data Mining
EDM	Educational Data Mining
GAs	Genetic algorithms
KDD	Knowledge Discovery in Database
LIMBO	ScaLable InforMation BOttleneck
MAR	Maximum Attribute Relative
MDA	Maximum Dependency Attribute
MDDS	Maximum Degree of Domination in Soft Set Theory
MMR	Min-Min Roughness
NSS	Novel Soft Set
OLAP	On Line Analytical Processing
OLTP	On Line Transaction Processing
ROCK	A Robust Clustering Algorithm for Categorical Attributes
STIRR	Sieving Through Iterated Relational Reinforcement
TR	Total Roughness

CHAPTER 1

INTRODUCTION

1.1. Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) and data mining have been attracting a significant amount of attention from research, industries, and media of late. Data mining is part of KDD. While one branch of data mining is the clustering. The relationship between KDD, data mining, and clustering can be shown in Figure 1.1.

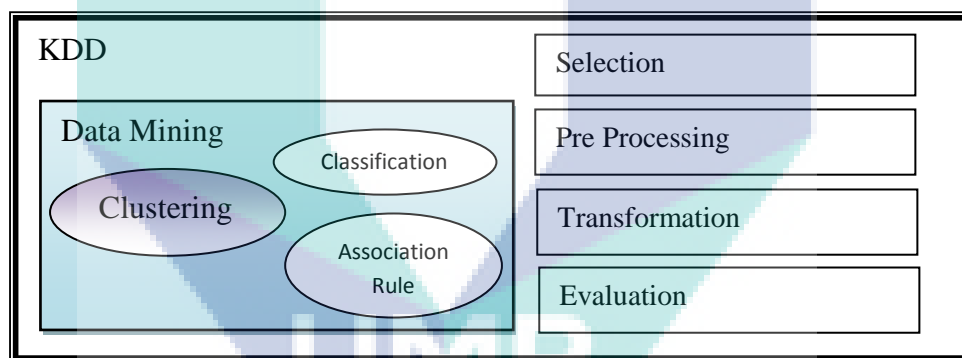


Figure 1.1. Relation between KDD, data mining, and clustering

KDD is the non trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad *et.al.*, 1996). KDD is a multi-step process that turns raw data into useful knowledge (Bagga and Singh, 2011). KDD can also be defined as the entire process of converting raw data into useful knowledge which consists of a series of phase transformations, such as data preprocessing and postprocessing. Briefly KDD is the process of discovering which is deemed as useful knowledge from the data set. Knowledge Discovery in Databases brings together current research on the exciting problem of discovering useful and interesting knowledge in

databases. While data mining is one of the overall process of the KDD process. The KDD process is shown in Figure1.2.

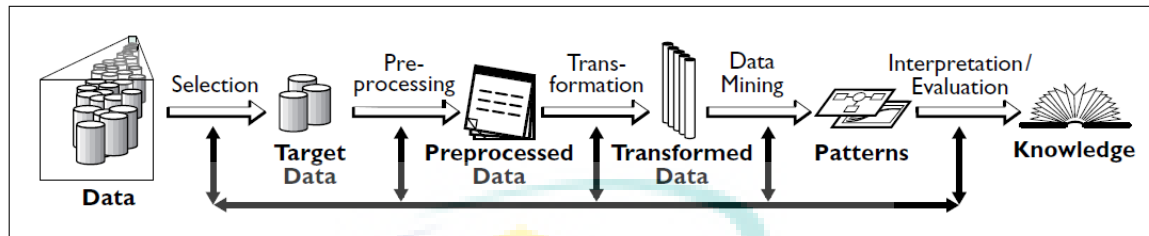


Figure 1.2. Overview of the steps constituting the KDD process (Fayyad *et.al*, 1996)

The KDD process is an interactive and iterative (with many decisions made by the user), involving numerous steps as follow :

- a. Learning the application domain: which includes relevant prior knowledge and the goals of the application.
- b. Creating a target dataset: which includes selecting a dataset or focusing on a subset of variables or data samples on which discovery is to be performed.
- c. Data cleaning and preprocessing: which includes basic operations, such as removing noise or outliers if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, and accounting for time sequence information and known changes, as well as deciding DBMS issues.
- d. Data reduction and projection: which includes finding 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.
- e. Choosing the function of data mining: which includes deciding the purpose of the model derived by the data mining algorithm.
- f. Choosing the data mining algorithm(s): which includes selecting method(s) to be used for searching for patterns in the data, such as deciding which models and parameters may be appropriate (e.g., models for categorical data are different from models on vectors over reals) and matching a particular data mining method with the overall criteria of the KDD process.

- g. Data mining: which includes searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression, clustering, sequence modeling, dependency, and line analysis.
- h. Interpretation: which includes interpreting the discovered patterns and possibly returning to any of the previous steps, as well as possible visualization of the extracted patterns, removing redundant or irrelevant patterns, and translating.
- i. Using discovered knowledge: which includes incorporating this knowledge into the performance system, taking actions based on the knowledge, or simply documenting it and reporting it to interested parties, as well as checking for and resolving potential conflicts with previously believed (or extracted) knowledge.

However, there is a risk of a loss of values and the rich information contained on this database unless a good technique is used to extract the existing knowledge (Chen *et al*, 1996; Fayyad, 1996).

The research area focuses on clustering on the cluster dimension with a low dimensional data. In this research used data categories, based on soft set theory approach. The flow of research area is shown in Figure 1.3.

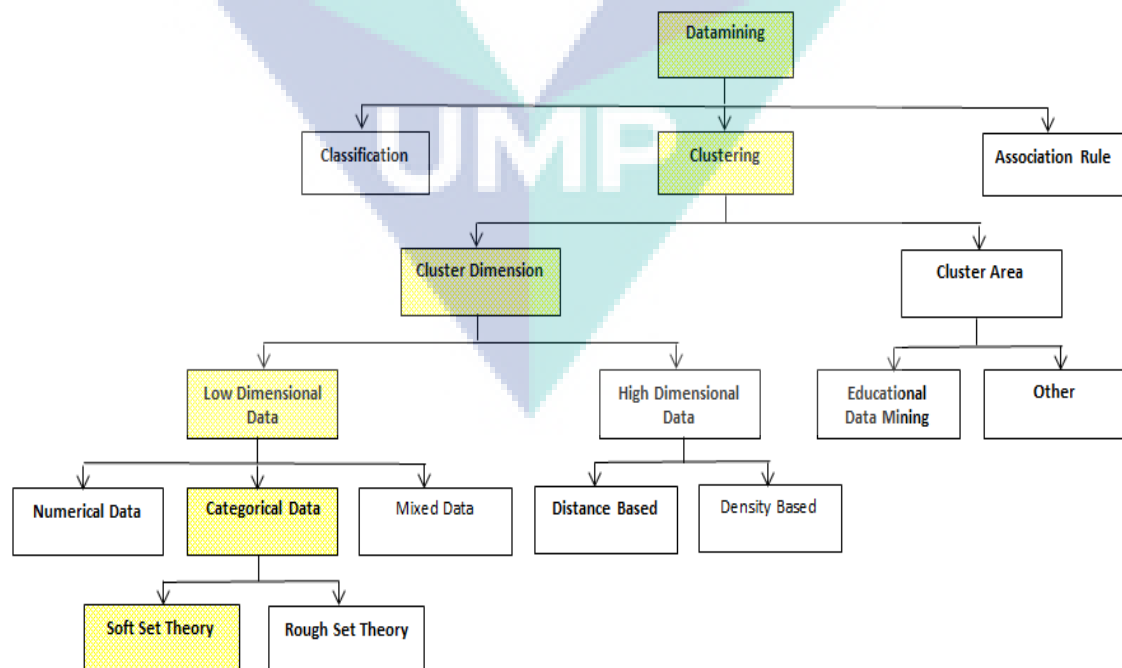


Figure 1.3. Research area position

1.1.1. Data Mining

Data mining is referred to the process of extracting hidden and useful information in large data repositories (Tan *et.al.*, 2005). The existence of data mining with the data explosion problem (bigdata) has recently occurred in many organizations (purchasing data, sales data, customer data, transaction data, etc.). Almost all the data was entered using a computer application used to handle the day-to-day transactions that are mostly On Line Transaction Processing (OLTP). Imagine how many transactions had been entered into by a Carefour hypermarket or some sort of credit card transactions from a bank in a day and imagine how big the size of their data, if it has been running a few years later. The data warehouse supports On Line Analytical Processing (OLAP), the functional and performance requirements of which are quite different from those of the on line transaction processing applications traditionally supported by the operational databases. Data warehouses provide OLAP tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data mining (Saagari *et.al.*, 2013).

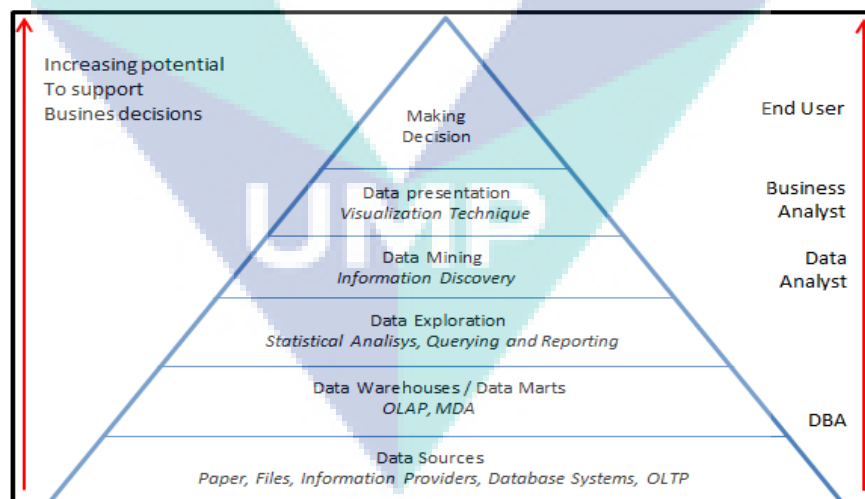


Figure 1.4: Data mining and other database technology

Figure 1.4 shows data warehouse technology is used to do OLAP, while data mining is used to perform information discovery, that the information is intended for a data analyst and business analyst (with added visuals of course). In practice, data mining also retrieves data from the data warehouse. The application of data mining is

more specific and the given OLAP database is not the only field of science that affects data mining. This areas of science which has enriched data mining are such as : information science, high performance computing, visualization, machine learning, statistics, neural networks, mathematical modeling, information retrieval, and pattern recognition. Even the image processing is also used in order to perform data mining to given the data image / spatial.

Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large databases. Concretely, four of the core data mining task are described below :

1.1.2. Classification

Classification is one of the most common learning model in data mining (Ahmed, 2004; Berry and Linoff, 2004). The goal is to build a model to predict the behavior of something through the database records which are classified into a number of standard classes based on certain criteria (Ahmed, 2004; Chen *et.al*, 2003). Equipment used for classification are neural networks, decision trees, and if-then-else rules. Classification process is usually divided into two phases: learning and test. In the learning phase, most of the data that has been known to form a class of data is fed to the model estimates. Test phase models are tested with most other data to determine its accuracy. With sufficient accuracy this model can be used to predict the class of data that is not known yet.

1.1.3. Clustering

Data clustering is one of the basic tools available, to understand the structure of the data set (Mesakar and Chaudhari, 2012). The process of grouping a set of physical or abstract objects into classes of similar objects is known as clustering. Clustering

techniques play an important role in machine learning, data mining, information retrieval, web analytics, marketing, medical diagnostics, and pattern recognition. Clustering is often called unsupervised learning task because there is no class that shows the value of a prior clustering given from the data sample, which is the case in supervised learning. General definition of clustering could be "the process of organizing objects into groups whose members are similar in some ways". Therefore, the cluster is a collection of data objects that are similar to each other in a same and distinct clusters with objects in other clusters.

Clustering is a dynamic area of research on data mining. Many clustering algorithms are constantly being developed. The selection algorithm clustering depends on the types of data available and the purpose of an application (Mesakar and Chaudhari, 2012). Clustering technique can be categorized into :

- a. Hierarchical
- b. Partitional
- c. Density Based
- d. Grid Based

Many techniques available for clustering cannot be directly applied to categorical data clustering, in which the values are different and do not have a particular order. Data category has a different structure from numerical data. Function of distance in the numerical data may not be applicable to categorical data. The use of technique to classify numerical data can not be applied to categorical data.

1.1.4. Association Rule Mining

Association Rule Mining is one of the popular techniques used in data mining. Positive association rules are very useful in correlation analysis and decision making processes (Abdullah *et.al*, 2011). Examples of associative rule of analysis are in supermarket purchases which are able to identify how likely a customer buys bread along with milk. From the knowledge, the supermarket owner can adjust the placement of the goods or design a marketing campaign using a combination of discount coupons for certain items. Associative rule can be defined by two parameters, namely the

percentage of combined support items in the database and confidence in that the strong relationship between the items in associative rules.

Clustering can be performed on the data that has some attributes that are mapped as a multidimensional space. Many clustering algorithms require a distance function to measure the similarity between data. Methods are also needed for the normalization of various attributes of the data. Some categories of clustering algorithms which are known are partitioning methods in which the user must specify the desired number of partitions. The data are tested for inclusion on one partition. Other methods have long been the hierarchy method that is divided in two. The combination of bottom-up cluster and small clusters has become larger and resulted in top-down splitting large clusters into smaller clusters.

Cluster dimension can be divided into two, namely the low-dimensional data and high-dimensional Data. The term 'dimension' is used to refer to the independent variables that were expressed in tuples, while 'variant' refers to the dependent variable (Wong and Bergeron, 1997). As a concept, low-dimensional data limit is 1 to 3 dimensional data. In general, discussion of four or more dimensions is a very complex problem. Characterization of high-dimensional data set is not an easy thing. As a guide, the data dimension is distinguished as follows, 'low' for dimensions 1 to 3, 'medium' for dimensions 4 to 9, and 'high' for the dimension 10 to the top, however this is only an option. Low-dimensional Data consists of three types of data, ie the categorical data, numerical data, and mixed data. The data types can be classified as follows:

a. **Categorical Data**

The categorical data are qualitative data so as to be analyzed using mathematical formulas / statistics need to be given a code number. Analysis of mathematics / statistics used are based on the results of counting in each category. Categorical data are called nonmetric data or data that is not the result of measurement. Data categories can be distinguished, namely:

1. **Nominal categorical**, i.e. data that can not be stated categorically, that the category is one better than the other categories or in other words, a category that does not have a particular order. Example: Men-women, purple-blue, and others.

Because it does not have a particular order, it may be the category "men" is coded "0" and "women" is coded as "1" or vice versa.

2. **Ordinal categorical**, i.e. categorical data that have a certain order, but the "distance" between category is equally difficult or otherwise. Example: The tool is categorized in a "good", "medium", "broken". Because it has a sequence, then if the "broken" is coded "1", then the next sequence is the "medium" which is coded "2", and the "good" category is coded as "3" or vice versa. Coding sequence above can not be exchanged at random, as it would be in the order category.

b. Numerical Data

Numerical data is the data that is the result of the measurement if the exact measurement data produces pure metric data.

1.2. Problems Statement

Many clustering techniques are used to partition large data sets. Several clustering techniques were based on rough set theory and soft set theory. However, existing techniques have various limitations, namely a long execution time and low accuracy.

Evaluation of clustering technique for categorical data is a difficult job. Currently, the most widely used criteria for evaluation includes accuracy and efficiency of clustering. Clustering accuracy measures the quality of the clustering technique. A higher value of the clustering indicates better clustering results (He et.al., 2008). Clustering efficiency is measured by the running time of the technique. The length of execution time, the low efficiency. (Guha et.al., 2000; Cristofor and Simoici, 2002; Parmar et.al., 2007; Deng et.al., 2010) have shown that the clustering technique has low accuracy, while having a higher efficiency compared to direct optimization-based techniques. On the contrary, direct optimization-based techniques have high accuracy, while the grouping has a low efficiency compared with a heuristic technique. Therefore, a new clustering technique for categorical data with high efficiency and high

classification accuracy is required. Therefore, this research proposes three research questions:

- a. How to study and analyze previous techniques and to propose a new technique ?.
- b. How to develop an attribute-oriented clustering technique for categorical data that improves efficiency and higher accuracy ?
- c. How to develop categorical data clustering technique that can be applied in educational data mining ?.

1.3. Research Objectives and Scope

This research has the following objectives :

- a. To study and develop an attribute-oriented clustering technique on the data categories that improves efficiency and higher accuracy.
- b. To apply the proposed technique in educational data mining.

The scope of this research is to study and analyze several clustering attribute selection techniques for categorical data. Based on the disadvantages of previous techniques, it should be proposed a new technique be introduced that has better performance. The execution time and accuracy of the proposed technique will be compared with previous technique in seventeen UCI benchmark datasets machine learning. Finally, the new technique will be applied in educational data mining

1.4. Thesis Organization

The rest of this thesis is organized as follows :

Chapter-2 : Basic knowledge of clustering techniques that are discussed in this chapter. Some clustering technique and its development are also described. The final section of this chapter describes educational data mining

Chapter-3 : An alternative clustering technique for categorical data is discussed. In this chapter, the proposed technique is the Maximum Degree of Domination in

Soft Set Theory (MDDS). This technique uses a soft set theory. Comparison with the MAR technique is also described.

Chapter-4 : An explanation of experiments are conducted to evaluate the clustering performance, clustering efficiency and accuracy of MAR and MDDS algorithm. In this chapter, the experiment and the results are described and comparison is made between MAR and MDDS techniques in terms of clustering efficiency and accuracy. The chapter is structured as follow :

- a. Introduces the experimental design, data sets, language, technique used for comparison, and platform for implementation
- b. Describes performance analysis, comparison of MAR and MDDS techniques, execution time, scalability, accuracy results clustering, and iterations.
- c. Describes the comparison between MAR and MDDS techniques.

Chapter-5 : The MDDS application to clustering dataset student assessment is described in this chapter. The real data description to be used in experiments, the presentation of results, and discussion are also described in this chapter.

Chapter-6 : This chapter presents the conclusions, recommendations, and future work for the research.

CHAPTER 2

LITERATURE REVIEW

2.1. Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) is an exploratory analysis and modeling of large data repositories that process organized, valid, new, useful, and understandable patterns of large data and complex (Maimon, O. and Rokach, L, 2010). The KDD is also an area of common interest to the researchers in the discovery engine, statistics, databases, knowledge acquisition, machine learning, data visualization, high performance computing, and knowledge-based systems.

The collection of data is increasing worldwide, so there is an urgent need of a technique, method, and algorithm to help researchers, analysts, decision makers and managers in extracting useful patterns of the rapidly growing volume of data. KDD has evolved from the interaction and cooperation among the various fields such as machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge representation and knowledge acquisition for intelligent systems. KDD's main idea is to find knowledge from raw data, or to find interpretations and abstractions which are better than previously known. This causes significant research on the process of knowledge discovery, especially KDD (Fayyad *et al*, 1996; Atkinson-Abutridy *et al*, 2004).

2.2. Data Mining

Every day in this world transactions occurs and the collection of data is in very large quantities. Analyzing these data is an important requirement. Data mining can

meet this requirement by providing the tools to discover knowledge from data and data mining can be seen as a result of the natural evolution of information.

Data mining is a truly interdisciplinary subject. Data mining can be defined in various technology in real ways (Han et al., 2011). The mining is a term that clearly characterizes the process of finding a small set of valuable nuggets of many raw materials. In addition, many other terms have the same meaning as data mining for example, knowledge of the datamining, knowledge extraction, data analysis / pattern. The knowledge discovery from data, or KDD, while others view data mining as an important step in the process of knowledge discovery. Knowledge discovery process is shown in Figure1.2.

In general, data mining can be applied to all types of data. Data mining is widely used in database application, data warehouse, and transactional data. Data mining can also be applied to other forms of the data for example, the data stream, the sequence of data, graphs or data networks, spatial data, text data, and multimedia data, and all order data types that appears.

2.3. Clustering

Cluster analysis is an important data analysis method used to group sets of data with similar characteristics. It has been used in many areas such as gene data processing (Jiang *etal*, 2004), transactional data processing (Giannotti,2002), decision support (Mathieu and Gibson, 2004), and radar signals processing (Haimovet *al*, 1989). Most previous clustering algorithms focus on data characterized by numerical attributes, for which natural geometric distances between the objects to be clustered are available. Recently, much attention have been put on categorical data clustering (Yanto *et al*, 2011; Yanto *et al*, 2012; Herawan, 2012), where data objects are made up of non-numerical attributes. Clustering categorical data presents a special challenge due to the absence of inherently ordered property of the data.

By using clustering, dense regions can be identified, the patterns of distribution and overall interesting linkages can be performed between data attributes. Data mining effort focuses on the discovery of methods to cluster in large databases effectively and

efficiently. The number of clustering approaches makes it difficult to determine the quality of a universal size. However, some things to note, are the input parameters that do not complicate the user. The cluster results can be analyzed, and its scalability can increase the size and dimensions of the dataset record. Previous work on inducing clustering dimensions has focused on producing multiple clusterings of a dataset, and can be broadly divided into two categories :

- a. **Semi-supervised methods.** These methods are semi-supervised in the sense that one of the clusterings is provided (by the human) as input, and the goal is to produce another clustering that is distinctively different from the given one. For instance, (Gondek and Hofmann, 2004) approach learns a non-redundant clustering that maximizes the conditional mutual information $I(C; Y | Z)$, where C , Y and Z denotes the clustering to be learned, the relevant features and the known clustering. It turns out to be difficult to implement, since it requires modeling of the joint distribution of the cluster labels and the relevant features. On the other hand, (Davidson *et al.*, 2007) first learnt a distance metric D_C from the original clustering C , and then reversed the transformation of D_C using the Moore-Penrose pseudo-inverse to get the new distance metric D'_C , which is used to produce a new clustering.
- b. **Unsupervised methods.** Here, each of the possible clustering is produced without using any labeled data. Meta clustering (Caruana *et al.*, 2006) is an approach that produces multiple clusterings of a dataset by running k-means multiple times, each time with a random selection of seeds and a random weighting of features. Its goal is to present each local minimum found by k-means as a possible clustering. This approach has two weaknesses. First, many of these local minimum are qualitatively poor. Second, k-means tends to produce similar clusterings regardless of the number of times it is run.

2.3.1. The Basic Steps of The Clustering Process

The process of clustering may lead to different partitions of a data set, depending on the specific criteria used for clustering. Thus, there is a need of preprocessing before

the user assumes the task of grouping a set of data. The basic steps for developing a clustering process can be presented as follows (Fayyad *et al.*, 1996):

- a. **Feature selection.** The goal is to select properly the features on which clustering is to be performed so as to encode as much information as possible concerning the task of interest. Thus, preprocessing of data may be necessary prior to their utilization in clustering task.
- b. **Clustering algorithm.** This step refers to the choice of an algorithm that results in the definition of a good clustering scheme for a data set. Aproximity measure and a clustering criterion mainly characterize a clustering algorithm as well as its efficiency to define a clustering scheme that fits the data set.
 - i. **Proximity measure** is a measure that quantifies how “similar” two data points (i.e. feature vectors) are. In many cases, it should be ensured that all the selected features contribute equally to the computation of the proximity measure and there are no features that dominate others.
 - ii. **Clustering criterion.** There is a Need to define criteria for clustering, which can be expressed by a function of cost or some other type of rule. The calculation of the expected type of cluster occurs in the data set. Thus, it can be defined "good" clustering criteria, which leads to a suitable partition data set.
- c. **Validation of the results.** The correctness of clustering algorithm results is verified using appropriate criteria and techniques. Since clustering algorithms define clusters that are not known a priori, irrespective of the clustering methods, the final partition of data requires some kind of evaluation in most applications (Rezaee *et al.*, 1998).
- d. **Interpretation of the results.** In many cases, the experts in the application area have to integrate the clustering results with other experimental evidence and analysis in order to draw the right conclusion.

2.3.2. Data types

The data is a record of a collection of facts. In every day usage data means a statement that is accepted as it is. This statement is the result of measurements or observations of a variable whose shape can be a number, word, or image. The data which were processed by cluster analysis may come from a variety of different applications. Each data object described has many attributes, that the contents of each object is a multidimensional vector. In general, the value of the attribute can be classified as follows:

a. Binary

Binary data is a type of data that is represented or displayed in the binary numeral system. Binary data is the only category of data that can be directly understood and executed by a computer. It is numerically represented by a combination of zeros and ones. Binary data is considered the native data/language of a computer and it interacts with the lowest abstraction layer of its hardware. This type of data is produced whenever a process is performed on a computer. The application requesting the process sends instructions in a high-level language that is ultimately converted into binary data to be executed or sent to the processor. All processes, regardless of their type, are converted into binary form before execution.

b. Numerical

Numeric data types used in a variable or constant is to store the value in the form of numbers or figures. Numeric data can be discrete data or continuous data. Discrete data is data obtained from the results of calculating (not measured). For example, there are 20 table numbers, number of people are 12,30 pieces of the building, and so on. Data continuum (continuous) is the data obtained from the measurements, i.e. the amount of data that can occupy all the values that exist between the two points. Such data have the characteristics of a continuity between a number by another number.

c. Categorical

Categorical data is the data that consists of only the small number of values. It is the part of data observed and its contents can be sorted into categories. This type of data

can also be put into different groups. Categorical data is information that is divided into groups. Each categorical data can be set in a variety of groups for example race, sex and religion.

d. Ordinal

Ordinal data are often used in nonparametric statistics. These data is in addition to having the attributes and ratings. Containing a given number of levels. It is used to sort objects from the lowest to the highest, or vice versa. This measure does not give the absolute value of the object, but only the first rank. If there is a set of objects that are numbered, from 1 to, for example, rank 1, 2, 3, 4, 5, and so on, when expressed in scale, the distances between data that are not equal to each other.

e. Mixed

In the database, the object is described as a mixture or combination of data types. A database can contain various types of data. There are the four listed above, both numeric and non-numeric. The data can be a number, word, image, and sound.

2.3.3. Objective Function

Clustering is the problem of discovering “meaningful” groups in given data. Many algorithms try to achieve this by minimizing a certain quality function. The objective of clustering is then stated as a discrete optimization problem. Given a data set $X_n = \{X_1, \dots, X_n\}$ and a clustering quality function Q_n , the ideal clustering algorithm should take into account all possible partitions of the data set and output the one that minimizes Q_n . The implicit understanding is that the “best” clustering can be any partition out of the set of all possible partitions of the data set. The algorithmic challenge is to construct an algorithm which is able to find this clustering. This approach is called "discrete optimization approach for clustering". A related effect occurs for spectral clustering, which is trying to minimize relaxation. It has been proved that under certain conditions the solution of the relaxed problem converges to the limited sample of some grouping boundaries. However, it has been assumed that the clustering does not necessarily limit the optimizer of the objective function. So in both cases, the consistency of results is an improvement : algorithms provably converges to

the limit of the minimizer, respectively. The same results apply to a large number of alternative objective function used for clustering (Luxburg *et al.*, 2007).

2.3.4. Membership

Usually, clustering algorithms assume that an object is only a single cluster member only, Though in many cases an object can be a member of another cluster, resulting in overlapping. This happens so some objects that may have multiple memberships, so the uncertainty is quite high. To resolve this problem the theory of fuzzy algorithm can be a solution. Clustering with fuzzy logic continues to grow because in general the data can not be strictly separated into cluster, but it has a degree of membership, which is valued between 0 and 1 to group (Hoppner *et al.*, 2004). An example of the FCM algorithm is shown in Figure 2.1.

1. *Determining the number of groups to be formed ($2 \leq c \leq n$)*
2. *Initialize the initial membership function matrix U^0 , usually chosen at random.*
3. *Calculating the value of the center of the group to-k by using equation 3.*
4. *Calculate the degree of membership that the new matrix U^{t+1} with equation 1. Comparing the value of membership in the matrix U , if $\|U^t - U^{t+1}\| < \varepsilon$ then it is and convergent iteration is stopped, where ε is a specified threshold value. If $\|U^t - U^{t+1}\| \geq \varepsilon$, then return to step 4. Threshold value is a number small positive near zero, 0.00001 (10^{-5}).*

Figure 2.1. Fuzzy c-means algorithm

2.3.5. Categorization of Clustering Methods

Many clustering methods have been developed, each of which uses a different induction principle. Fraley and Raftery (1998) suggest dividing the clustering methods into two main groups: hierarchical and partitioning methods. Han and Kamber (2011) suggest categorizing the methods into additional three main categories: density-based methods, model-based clustering and grid based methods.

- a. Hierarchical clustering :These methods construct the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion. These methods can be subdivided as follows :
 - i. Agglomerative hierarchical clustering-Each object initially represents a cluster of its own. Then clusters are successively merged until the desired cluster structure is obtained.
 - ii. Divisive hierarchical clustering-All objects initially belong to one cluster. Then the cluster is divided into sub-clusters, which are successively divided into their own sub-clusters. This process continues until the desired cluster structure is obtained.

The results represent the hierarchical method nested on clusters of objects and the degree of similarity where the clustering changes. A clustering of the data objects are obtained by cutting the dendrogram at the desired similarity level. The merging or division of clusters is performed according to some similarity measure, chosen so as to optimize some criterion (such as a sum of squares). The hierarchical clustering methods could be further divided according to the manner that the similarity measure is calculated (Jain *et al.*, 1999).

- b. Partitional clustering : Partitioning methods relocate instances by moving them from one cluster to another, starting from an initial partitioning. Such methods typically require that the number of clusters will be pre-set by the user. To achieve global optimality in partitioned-based clustering, an exhaustive enumeration process of all possible partitions is required. Because this is not feasible, certain greedy heuristics are used in the form of iterative optimization. Namely, a relocation method iteratively relocates points between the k clusters.
- c. Density based clustering :Density-based methods assume that the points that belong to each cluster are drawn from a specific probability distribution (Banfield and Raftery, 1993). The overall distribution of the data is assumed to be a mixture of several distributions. The aim of these methods is to identify the clusters and their distribution parameters. These methods are designed for discovering clusters of arbitrary shape which are not necessarily convex. The idea is to to cluster growth given for density (number of objects or data points) in the neighborhood which

exceeds some threshold. Namely, the neighborhood of a given radius has to contain at least a minimum number of objects. When each cluster is characterized by a local mode or maxima of the density function, these methods are called mode-seeking.

- d. Grid based clustering :These methods partition the space into a finite number of cells that form a grid structure on which all of the operations for clustering are performed. The main advantage of the approach is its fast processing time (Han and Kamber, 2011).

2.4. Algorithms For Categorical Data Clustering

K-means clustering is a very popular technique for partitioning large data sets with numerical attributes. K-means clustering algorithm is a partition-based clustering method which is simple and unsupervised. As a basic partitioning algorithm which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. K in k-means is the number of clusters to be included to the algorithm. This will be repeated. K-means algorithm is simple and fast. K- Means algorithm is only applicable to numeric data and does not work with categorical data. The process is formulated as follows (Huang, 1998) :

$$\text{Minimise} \quad P(W, Q) = \sum_{i=1}^k \sum_{l=1}^n w_{i,l} d(X_i, Q_l) \quad (2.1)$$

$$\begin{aligned} \text{Subject to} \quad & \sum_{l=1}^k w_{i,l} = 1, \quad 1 \leq i \leq n \\ & w_{i,l} \in \{0,1\}, \quad 1 \leq i \leq n, 1 \leq l \leq k \end{aligned} \quad (2.2)$$

where W is an $n \times k$ partition matrix, $Q = (Q_1, Q_2, \dots, Q_k)$ is a set of objects in the same object domain, and $d(., .)$ is the squared Euclidean distance between two objects. Problem P can be solved by iteratively solving the following two problems:

- a. Problem $P1$: Fix $Q = \hat{Q}$ and solve the reduced problem $P(W, \hat{Q})$
- b. Problem $P2$: Fix $W = \hat{W}$ and solve the reduced problem $P(\hat{W}, Q)$

Problem $P1$ is solved by

$$\begin{aligned} w_{i,j} &= 1 \text{ if } d(X_i, Q_j) \leq d(X_i, Q_t), \text{ for } 1 \leq t \leq k \\ w_{i,t} &= 0 \text{ for } t \neq j \end{aligned} \quad (2.3)$$

and problem $P2$ is solved by

$$q_{j,s} = \frac{\sum_{i=1}^n w_{i,j} x_{i,s}}{\sum_{i=1}^n w_{i,j}} \quad (2.4)$$

For $1 \leq j \leq k$, and $1 \leq s \leq d$

2.4.1. K-Mode

K-mode, expands the paradigm of algorithm for k-means to categorical domains. K-mode (Huang, 1997) extend k-means and introduce new differences of size for the data category. Dissimilarity measure between two object is calculated as the number of attributes whose value does not match. Then, the algorithm for K-mode replace cluster mode by using the method of frequency based on renewal mode in the process of clustering to minimize the cost function clustering. K-mode generates local optimal solutions based on the initial mode and order of objects in the collection of data. K-mode should be run multiple times with different initial value mode to test the stability of clustering solutions.

a. Dissimilarity measure

Let X, Y be two categorical objects described by m categorical attributes. The dissimilarity measure between X and Y can be defined by the total mismatches of the corresponding attribute categories of the two objects. The smaller the number of mismatches is, the more similar are the two objects. This measure is often referred to as simple matching (Kaufman and Rousseeuw, 1990). Formally,

$$d_l(X, Y) = \sum_{j=1}^m \delta(x_j y_j) \quad (2.5)$$

where

$$\delta(x_j y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases} \quad (2.6)$$

b. Mode of a set

Let X be a set of categorical objects described by categorical attributes, A_1, A_2, \dots, A_m . A mode of $X = \{X_1, X_2, \dots, X_n\}$ is a vector $Q = [q_1, q_2, \dots, q_m]$ that minimises

$$D(X, Q) = \sum_{i=1}^n d_1(X_i, Q) \quad (2.7)$$

Here, Q is not necessarily an element of X .

c. The k-modes algorithm

When (2.5) is used as the dissimilarity measure for categorical objects, the cost function (2.1) becomes

$$P(W, Q) = \sum_{l=1}^k \sum_{i=1}^n \sum_{j=1}^m w_{i,l} \delta(x_{i,j}, q_{i,j}) \quad (2.8)$$

where

$$w_{i,l} \in W \text{ and } Q_l = [q_{l,1}, q_{l,2}, \dots, q_{l,m}] \in Q$$

To minimise the cost function the basic k -means algorithm can be modified by using the simple matching dissimilarity measure to solve $P1$, using modes for clusters instead of means and selecting modes according to Theorem 1 to solve $P2$. The basic algorithm needs to calculate the total cost of P with respect to the entire dataset every time when a new Q or W is obtained. To make the computation more efficiently used, the following algorithm in practice.

- a. Select k initial modes, one for each cluster.
- b. Allocate objects in a cluster with nearest mode. Update the mode of the cluster after each allocation.
- c. After all objects have been allocated to clusters, retest the dissimilarity of objects against the current modes. If an object is found such that its nearest mode belongs to another cluster rather than its current one, reallocate the object to that cluster and update the modes of both clusters.
- d. Repeat 3, until no object has changed clusters, after a full cycle test of the whole data set.

2.4.2. Squeezer

As a single-pass algorithm, Squeezer (He *et al.*, 2002) utilises the prespecified similarity threshold to determine which of the existing group (or a new cluster) data point under examination is given. Squeezer algorithm used to cluster categorical data clustering gives greater weight to the value of the attribute in the calculation of

similarity. An algorithm for clustering categorical data, Squeezer combines both efficiency and quality of clustering results. The algorithm is suitable for clustering data streams, in a given sequence of points. The objective is to maintain consistently good clustering of the sequence, using a small amount of memory and time. The algorithm does not require the number of desired clusters as an input parameter. This is very important since the user usually does not know this number in advance. The only parameter to be specified is the value of similarity between the tuple and cluster, which incorporates the user's exception that how closely the tuples in a cluster should be. Squeezer algorithm is shown in Figure 2.2.

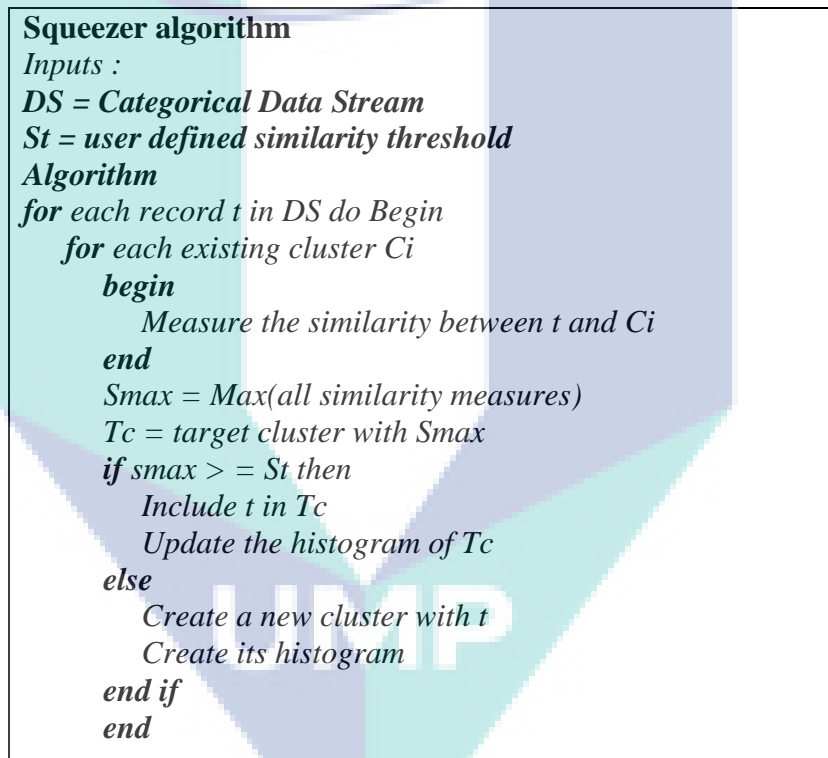


Figure 2.2. The Squeezer algorithm

The time and complexities of the Squeezer algorithm depend on the size of dataset and the number attributes.

2.4.3. LIMBO

The ScaLable InforMation Bottleneck (LIMBO) (Andritsos *et al.*, 2004) is a hierarchical clustering algorithm using the Information Bottleneck (IB) framework for determining the distance measure for categorical tuples. LIMBO has the advantage that it can produce clusterings of different sizes in a single execution. The IB framework is used to define a distance measure for categorical tuples. LIMBO handles large data sets to produce a summary of memory models which are limited to the data. LIMBO algorithm proceeds in four stages. In the first stage, the original artifacts encapsulate into a set \mathbf{S} of SA s. In the second stage, Agglomerative Information Bottleneck algorithms used in the \mathbf{S} to produce a series of clusterings of SA s with a decrease in cardinality. The third stage is to transform clusterings decomposition of the original artifact sets. Finally, phase 4 decomposition chooses one of these as the final result.

- a. Creation of the Summary Artifacts. In this phase, original artifacts are read one by one. The first artifact \mathbf{a}_1 is converted into the summary artifact $SA(\{\mathbf{a}_1\})$, whose probability vector is equal to that of \mathbf{a}_1 . For each subsequent artifact \mathbf{a}_i , Calculation of the distance to any existing SA . The distance between the original artifacts and artifacts summary $SA(\{\mathbf{S}_j\})\mathbf{a}_i$ is the loss of information that would be imposed if the \mathbf{a}_i is added to \mathbf{S}_j .
- b. Application of the AIB algorithm. Phase 1 replaces the original artifact set with a much smaller set of SA s. In the second phase, the algorithm employs the AIB algorithm to cluster the set SA s. The input to the algorithm is a set of AIB conditional probability distribution of all SA s created in the first phase. This phase creates many artifacts clusterings summaries, one for each value between 2 and $|\mathbf{S}|$. Note that the AIB algorithm is applied only once, with intermediate clusterings recorded at each step.
- c. Associating the original artifact clusters. Phase 2 generates $|\mathbf{S}|-1$ clusterings, each containing a number that represents the probability distribution of its cluster. In the third stage, scanning over a series of original artifacts and each of them is assigned to the closest cluster, with respect to the distance D_{KL} , in all clusterings.
- d. Determining the number of clusters. In order to choose the appropriate number of clusters k , Research decomposition made in Phase 3 for the rising value of k

ranging from 2. Let C_k be the grouping of clusters k and C_{k+1} a grouping clusters. If the cluster group in C_k reflect inherent in the data, then C_{k+1} cluster must contain the same, except that one of them has been split into two (because the number of clusters should increase).

2.4.4. ROCK

ROCK :A Robust Clustering Algorithm for Categorical Attributes (Guha *et al.*, 2000) is an agglomerative hierarchical clustering algorithm that explores the concept of a link to the data by category attributes. Traditional clustering algorithm for clustering categorical data using a distance function. Distance measure does not lead to high-quality clusters when classifying categorical data. The algorithm considers rock environment of the individual pairs points. ROCK algorithm begins by assigning each tuple to a separated cluster, and then repeatedly merge clusters according to the closeness between clusters. The closeness between clusters is defined as the sum of the number of "links" between all pairs of tuples, where the number of "links" is calculated as the number of common neighbors between two tuples. The basic idea of ROCK is based on the following definition.

- a. **Neighbor** : An object's neighbors are those objects that are considerably similar to it. Let $sim(O_i, O_j)$ be similarity function that normalizes and captures the closeness between the pair of objects O_i and O_j . The similarity function assumes values between 0 and 1. Given a threshold θ (between 0 and 1), a pair of objects O_i and O_j are defined as neighbours if $sim(O_i, O_j) \geq \theta$.
- b. **Link** : The $link(O_i, O_j)$ between the object is defined as the number of common neighbors between O_i and O_j . ROCK attempts to maximize the sum of $link(O_q, O_r)$ for pairs of objects O_q and O_r belonging to single cluster and at the same time to minimizes the sum of the $link(O_q, O_r)$ for object pairs belonging to other cluster.
- c. **Link between clusters** : It is a summation of links of all pairs, where each pair is formed by taking one object from each cluster.

- d. **Goodness measure** : The goodness measure between two clusters are the result obtained after dividing the number of cross-links between the clusters by the expected number of cross-links between the cluster.

ROCK's hierarchical clustering algorithm is presented in Figure 2.3. It accepts as input the set S of n sampled points to be clustered (that are drawn randomly from the original data set), and the number of desired clusters k . The procedure begins by computing the number of links between pairs of points. Initially, each point is a separate cluster. For each cluster i , An algorithm builds a local heap $q[i]$ and maintains the heap during the execution of the algorithm. $q[i]$ contains every cluster j such that $link[i; j]$ is non-zero. The clusters j in $q[i]$ are ordered in the decreasing order of the goodness measure with respect to i , $q[i; j]$.

```

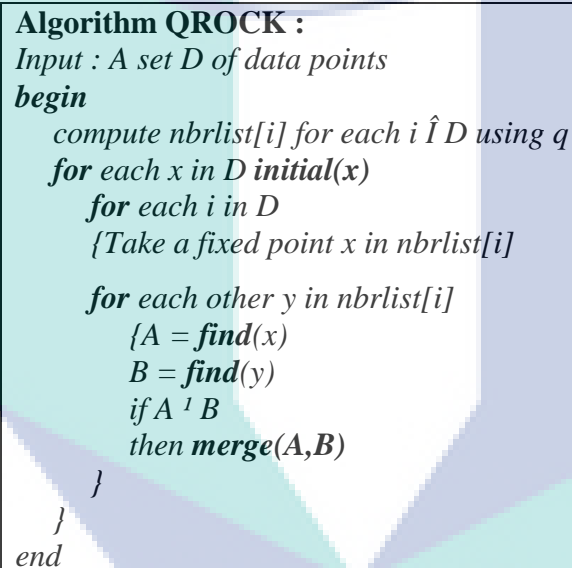
procedure cluster(S; k)
begin
  link := compute links(S)
  for each  $s \in S$  do
     $q[s] := \text{build local heap}(\text{link}; s)$ 
   $Q := \text{build global heap}(S; q)$ 
  while  $\text{size}(Q) > k$  do {
     $u := \text{extract max}(Q)$ 
     $v := \text{max}(q[u])$ 
    delete( $Q; v$ )
     $w := \text{merge}(u; v)$ 
    for each  $x \in q[u] \cup q[v]$  do {
       $\text{link}[x; w] := \text{link}[x; u] + \text{link}[x; v]$ 
      delete( $q[x]; u$ ); delete( $q[x]; v$ )
       $\text{insert}(q[x]; w; g(x; w)); \text{insert}(q[w]; x; g(x; w))$ 
      update( $Q; x; q[x]$ )
    }
     $\text{insert}(Q; w; q[w])$ 
    deallocate( $q[u]$ ); deallocate( $q[v]$ )
  }
end

```

Figure 2.3. The ROCK clustering algorithm

2.4.5. QROCK

The next generation is QROCK (Dutta *et al.*, 2005) algorithms that compute clusters by determining the connected components of the graph. This leads to a very efficient method to obtain clusters and provides a drastic reduction of the computing time ROCK algorithm. QROCK more practical algorithm to determine the similarity threshold rather than specifying the desired number of clusters a priori. QROCK algorithm also detects the outliers in this process and discusses the new similarity measure for categorical attributes. Figure 2.4 shows the QROCK algorithm.



```

Algorithm QROCK :
Input : A set  $D$  of data points
begin
  compute  $nbrlist[i]$  for each  $i \in D$  using  $q$ 
  for each  $x$  in  $D$  initial( $x$ )
    for each  $i$  in  $D$ 
      {Take a fixed point  $x$  in  $nbrlist[i]$ 
      for each other  $y$  in  $nbrlist[i]$ 
        { $A = find(x)$ 
         $B = find(y)$ 
        if  $A \neq B$ 
          then merge( $A, B$ )
        }
      }
    }
  end

```

Figure 2.4. The QROCK algorithm.

2.4.6. CACTUS

CACTUS (Ganti *et al.*, 1999) is a fast summarization-based algorithm for clustering categorical data that consists of three phases: summarization, clustering, and validation. At this stage of summarization, the summary information is computed from the dataset. In the clustering stage, the summary information is used to find a set of cluster candidates. In the validation phase, the actual set of clusters is determined from the set of cluster candidates. The main idea behind CACTUS is that the summary of the entire dataset is sufficient to calculate a set of "candidate" clusters which can then be

validated to determine the actual set of clusters. CACTUS consists of three phases : summarization, clustering, and validation. At this stage of summarization, the summary information is computed from the dataset. In the clustering stage, a summary of information is to find a set of cluster candidates. In the validation phase, the actual determination of the cluster are sets of candidate clusters. Hypothetical examples are used throughout the paper to illustrate the successive phases in the algorithm.

2.4.7. COOLCAT

COOLCAT (Barbara *et al.*, 2002) is a method that uses the concept of entropy group records. COOLCAT is an additional algorithm that aims to minimize the expected entropy of the cluster. Given a set of clusters, COOLCAT will put the next point in the cluster which minimizes the overall expected entropy. COOLCAT act gradually, and it will able to cluster at any new points without having to reprocess the entire set. Therefore, COOLCAT is suited to cluster data streams. An entropy-based algorithm, COOLCAT, consists of two steps: initialization and incremental step.

Initialization step "bootstraps" algorithm , finding a suitable set of clusters of samples S , taken from the data set ($|S| \ll N$), where N is the size of the entire data set. First of all k is searched the most "different" from the record set by maximizing the entropy of the sample pairs which were chosen at minimum points. Two search points ps_1, ps_2 maximizes $E(ps_1, ps_2)$ and places them in two separate groups ($C1, C2$), marking a record (this takes $O(|S|^2)$). From there, proceed in stages, ie, to find records, that need to put in the j -th cluster, marked ps_j point chosen that maximizes $Min_i =_{1, \dots, j-1} (E(ps_i, ps_j))$. After initialization, the remaining records from the data set is processed (the rest of the sample outside the sample point) gradually, finding clusters that are suitable for each record. This is done by calculating the entropy which is expected to result from putting points in each cluster and choosing cluster entropy as the expected minimum. Further on additional steps are required to bring the buffer memory to the main points and grouping them one by one. Coolcat procedure is shown in Figure 2.5.

1. Given an initial set of clusters $\check{C} = C_1, \dots, C_k$
2. Bring points to memory from disk and for each point p do
 3. For $i = 1, \dots, k$
 4. Tentatively place p in C_i and compute $\bar{E}(\check{C}_i)$ where \check{C}_i denotes the clustering obtained by placing p in cluster C_i
 5. Let $j = \text{argmini}(\bar{E}(\check{C}_i))$
 6. Place p in C_j
7. Until all points have been placed in some cluster

Figure 2.5: Incremental step of COOLCAT algorithm

(Gibson *et al.*, 2000) proposed an algorithm called **STIRR** (Sieving Through Iterated Relational Reinforcement), a general method of spectral graph partitioning for categorical data. Shake is an iterative approach, which maps the data categories for non-linear dynamic systems. In the convergent dynamic system, the categorical data can be clustered.

The database is represented as a graph where each distinct value in the domain of each attribute is represented by a weighted node. Thus if there are N attributes and the domain size of i – *th* attribute is d_i . Then the number of nodes in the graph is $\sum_i d_i$. For each tuple in the database, an edge represents a set of nodes which participate in that tuple. Thus a tuple is collection of nodes, one from each attribute type. The set of weights of all the nodes define a configuration of this structure. The initial weights of all the nodes can be either assigned uniformly or randomly or by focusing technique. STIRR iteratively changes the configuration by updating the weight of any node. The new weight of the node is calculated based on a combiner function, which combines the weights of other nodes participating in any tuple with given node for which the weight is to be updated.

Thus it moves from one configuration to the other till it reaches a stable point, called as basin. The convergence is dependent on the combiner function. Analyzing the stability is hard for any arbitrary combiner function. However, for simple combination function like sum of multiplication, the system definitely converges to a fixed point. It is easy to see that for categorical attributes, the values which are related through

common tuples influence each other during weight modification. Thus one does not require any similarity metric to be defined for categorical attributes. Interestingly, in order to cluster the set of tuples, STIRR maintains multiple copies of weights. When the fixed point is reached, the weight in one or more of the basins isolate into two groups of attribute values on each attribute. The first with large positive weights and second with small negative weights. The nodes with large positive weights and second with small negative weights are grouped to determine the cluster. These groups correspond intuitively to projections of clusters on the attribute. However, the automatic identification of such sets of closely related attribute values from their weights requires a non-trivial post-processing step; such a post-processing step was not addressed in their work. Moreover, the post-processing step will also determine which ‘cluster’ are output. The underlying idea of STIRR is unique but it may be hard to analyze the stability of the system for any useful combiner function.

2.4.8. CLOPE

CLOPE (Yang *et al.*, 2002) is a clustering algorithm for categorical data. Clope proposal is based on the intuitive idea of increasing the ratio of high-to-width histogram cluster. The general idea is the repulsion parameter that controls the tightness of transactions in a cluster, and thus the number of clusters generated. The simple idea behind Clope is that if make fast, scalable, and memory savings of clustering large, sparse transactional databases with high dimensional. CLOPE uses a global criterion function that tries to increase the intra-cluster overlapping of transaction items by increasing the height-to-width ratio of the cluster histogram. Moreover, it generalizes the idea by introducing a parameter to tightness of the cluster. Different number of clusters can be obtained by varying parameter. A larger height-to-width ratio of the cluster histogram means better intra-cluster similarity. And, the global criterion function is defined using the geometric properties of the cluster histogram. Clope algorithm is shown in Figure 2.6.

```

/* Phrase 1 - Initialization */
1: while not end of the database file
2:   read the next transaction  $\langle t, \text{unknown} \rangle$ ;
3:   put  $t$  in an existing cluster or a new cluster  $C_i$  that maximize profit;
4:   write  $\langle t, i \rangle$  back to database;
/* Phrase 2 - Iteration */
5: repeat
6:   rewind the database file;
7:   moved = false;
8:   while not end of the database file
9:     read  $\langle t, i \rangle$ ;
10:    move  $t$  to an existing cluster or new cluster  $C_j$  that maximize profit;
11:    if  $C_i \neq C_j$  then
12:      write  $\langle t, j \rangle$ ;
13:      moved = true;
14: until not moved;

```

Figure 2.6. The sketch of the CLOPE algorithm

2.4.9. CLICK

CLICK (Zaki *et al.*, 2005) finds clusters in categorical datasets based on a search for k -partite maximal cliques. CLICK selective vertical expansion approach is to ensure a complete search, there is no valid clusters that are missed. It also incorporates cliques that overlap to report more meaningful clusters. It does not impose any constraint domain and it is scalable to high dimensions. CLICK are exceeded for high dimensional datasets. CLICKS algorithm are for mining categorical (subspace) clusters. The main contributions are: i) A novel formalization of categorical datasets as k -partite graphs, where clusters correspond to k -partite cliques after post-processing. ii) A selective vertical expansion approach to guarantee a complete search; overlapping cliques are merged to report more meaningful clusters. iii) CLICKS outperforms existing approaches by over an order of magnitude. It can mine subspace clusters and scales extremely well for high dimensions.

Clustering is a dynamic field of research in data mining. It is able to find highly correlated areas into desirable objects as the data set grows. A variety of data clustering algorithms for categorical data is mentioned. Various data clustering technique has its

own advantages and disadvantages. The drawback lies in the accuracy and efficiency is still low.

2.5. Genetic Clustering Algorithms

Genetic algorithms (GAs) (Filho *et al.*, 1994) are randomized search and optimization techniques guided by the principles of evolution and natural genetics, having a large amount of implicit parallelism. GAs perform search in complex, large and multimodal landscapes, and provide near-optimal solutions for objective or fitness function of an optimization problem.

In GAs, the parameters of the search space are encoded in the form of strings (called chromosomes). A collection of such strings is called a population. Initially, a random population is created, which represents different points in the search space. An objective and fitness function is associated with each string that represents the degree of goodness of the string. Based on the principle of survival of the fittest, a few of the strings are selected and each is assigned to a number of copies that go into the mating pool. Biologically inspired operators like cross over and mutation are applied on these strings to yield a new generation of strings. The process of selection, crossover and mutation continues for a fixed number of generations or till a termination condition is satisfied. An excellent survey of GAs along with the programming structure used can be found in Ref.(Filho *et al.*, 1994).

The searching capability of GAs has been used in this technique. The clustering metric that has been adopted is the sum of the Euclidean distances of the points from their respective cluster centres. Mathematically, the clustering metric M for the K clusters C_1, C_2, \dots, C_K is given by :

$$M(C_1, C_2, \dots, C_K) = \sum_{i=1}^k \sum_{x_j \in C_i} \|X_j - Z_i\| \quad (2.9)$$

The basic steps of GAs, which are also followed in the GA-clustering algorithm, are shown in Figure 2.7.


```

Begin
1.  $t = 0$ 
2. initialize population  $P(t)$ 
3. compute  $p(t)$ 
4.  $t = t + 1$ 
5. if termination criterion achieved go to step 10
6. select  $P(t)$  from  $P(t-1)$ 
7. crossover  $P(t)$ 
8. mutate  $P(t)$ 
9. go to step 3
10. output best and stop
end

```

Figure. 2.7. Basic steps in GAs

It has been proven (Bhandari *et al.*, 1996) that an elitist model of GAs will definitely provide the optimal string as the number of iterations that goes to infinity, provided the probability of going from any population to the one containing the optimal string is greater than zero. Note that this has been proved for nonzero mutation probability values and is independent of the probability of crossover. However, since the rate of convergence to the optimal string will definitely depend on these parameters, a proper choice of these values is imperative for the good performance of the algorithm. Note that the mutation operator as used in this article also allows nonzero probability of going from any string to any other string.

Therefore, the GA-clustering algorithm will also provide the optimal clusters as the number of iterations that goes to infinity. Such a formal theoretical proof is currently being developed that will effectively serve as a theoretical proof of the optimality of the clusters provided by the GA-clustering algorithm. However, it is imperative to once again realize that for practical purposes a proper choice of the genetic parameters, which may possibly be kept adaptive, is crucial for a good performance of the algorithm.

2.6. Rough Set-based Clustering Attribute Selection Techniques

Currently, there are two measurement approaches that are based on the theory of rough set in clustering attribute selection. The first approach is based on the roughness of the attribute, and the second one is based on the attribute dependencies. In this

section, two techniques of the first approach, i.e. Total Roughness (TR) and Min-Min Roughness (MMR), and a technique of second approach called Maximum Dependency of Attribute (MDA) will be reviewed.

2.6.1. Total Roughness (TR) Technique

Total Roughness technique, which is proposed by (Mazlack *et al.*, 2000), has three main steps. The first one is to determine the roughness of each attribute by exploiting the relationship between lower and upper approximation which is defined in equations (1) and (2), respectively. Let $S = (U, A, V, f)$, an information system as described in sub-section 2.1, that $a_i \in A$ has k -different values, say β_k , $k = 1, 2, \dots, n$. Let $X(a_i = \beta_k)$, $k = 1, 2, \dots, n$ be a subset of the objects having k -different values of attribute a_i . The roughness of the set $X(a_i = \beta_k)$, $k = 1, 2, \dots, n$ with respect to a_j , where $i \neq j$, denoted by $\text{TR_Rough}_{a_j}(X | a_i = \beta_k)$, is defined by

$$\text{TR_Rough}_{a_j}(X | a_i = \beta_k) = \frac{|X_{a_j}(a_i = \beta_k)|}{|\overline{X}_{a_j}(a_i = \beta_k)|}, k = 1, 2, \dots, n \quad (2.10)$$

where, $|X_{a_j}(a_i = \beta_k)|$ and $|\overline{X}_{a_j}(a_i = \beta_k)|$ are, respectively, the cardinality of lower and upper approximations of a subset of the objects having k -different values of attribute a_i with respect to attribute a_j , where $i \neq j$.

The next step, the mean roughness of attribute $a_i \in A$ with respect to attribute $a_j \in A$, where $i \neq j$, denoted by $\text{TR_MeanRough}_{a_j}(a_i)$ is evaluated as follows:

$$\text{TR_MeanRough}_{a_j}(a_i) = \frac{\sum_{j=1}^{|V(a_i)|} R(X | a_i = \beta_k)}{|V(a_i)|} \quad (2.11)$$

where $V(a_i)$ is the set of values of attribute $a_i \in A$. Finally, the total roughness of attribute $a_i \in A$, with respect to attribute $a_j \in A$, where $i \neq j$ denoted as $TR(a_i)$ is obtained by the following formula:

$$TR(a_i) = \frac{\sum_{j=1}^{|A|} TR_MeanRough_{a_j}(a_i)}{|A| - 1} \quad (2.12)$$

Mazlack *et al.* suggested that the highest of TR is the best selection for a clustering attribute.

2.6.2. Min-Min Roughness (MMR) Technique

MMR, proposed by (Parmar *et al.*, 2007), is another rough set-based technique. The value of roughness of MMR technique is given below:

$$\begin{aligned} MMR_Rough_{a_j}(X|a_i = \beta_k) &= 1 - \frac{|X_{a_j}(a_i = \beta_k)|}{|X_{a_j}(a_i = \beta_k)|} \\ &= 1 - TR_Rough_{a_j}(X | a_i = \beta_k), \end{aligned} \quad (2.13)$$

It is clear that MMR technique uses Marczewski-Steinhausmetric (Lin, 1997) to measure the roughness of the set $X(a_i = \beta_k)$, $k = 1, 2, \dots, n$, with respect to a_j , where $i \neq j$. The value of the mean roughness of MMR technique is given as follows:

$$MMR_Mean\ Rough_{a_j}(a_i) = \frac{\sum_{k=1}^{|V(a_i)|} MMR_Rough_{a_j}(X|a_i = \beta_k)}{|V(a_i)|} \quad (2.14)$$

From equations (2.13) and (2.14), it is easily seen that the value of the mean roughness of MMR technique is an opposite of the value of the mean roughness of TR.

$$\begin{aligned}
\text{MMR_MeanRough}_{a_j}(a_i) &= \frac{\sum_{k=1}^{|V(a_i)|} (1 - \text{TR_Rough}_{a_j}(X|a_i = \beta_k))}{|V(a_i)|} \\
&= \frac{\sum_{k=1}^{|V(a_i)|} 1 - \sum_{k=1}^{|V(a_i)|} \text{TR_Rough}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|} \\
&= \frac{|V(a_i)|}{|V(a_i)|} - \frac{\sum_{k=1}^{|V(a_i)|} \text{TR_Rough}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|} \\
&= 1 - \text{TR_MeanRough}_{a_j}(a_i),
\end{aligned}$$

for $i \neq j$.

Let m be the attributes, Min-Roughness (MR) of attribute a_i with respect to all a_j , where $i \neq j$ is defined as

$$\text{MinR}(a_i) = \min\{\text{MMR_MeanRough}_{a_1}(a_i), \dots, \text{MMR_MeanRough}_{a_m}(a_i)\}$$

where $1 \leq i, j \leq m$.

The minimum of the Min-Roughness or known as Min-Min-Roughness, which is denoted by MMR, is formulated as follows:

$$\text{MMR} = \min(\text{MinR}(a_1), \dots, \text{MinR}(a_i), \dots, \text{MinR}(a_{|A|})) \quad (2.15)$$

where A is the number of attribute in the information system, $a_i \in A$, $1 \leq i \leq |A|$. Parmar *et al.* suggested that the least mean roughness is the best selection of a clustering attribute. The details of the MMR algorithm is shown in Figure 2.8.

```

Procedure MMR(U, k)
Begin
  Set current number of cluster CNC = 1
  Set ParentNode = U
  Label 1:
  If CNC < k and CNC ≠ 1 then
    ParentNode = ProcParentNode (CNC)
  End if
  // Clustering the ParentNode
  For each  $a_i \in A$  ( $i = 1$  to  $n$ , where  $n$  is the number of attributes in  $A$ )
    Determine  $[x_i]_{Ind(a_i)}$ 
    For each  $a_j \in A$  ( $j = 1$  to  $n$ , where  $n$  is the number of attributes in  $A$ ,  $j \neq i$ )
      Calculate  $Rough_{a_j}(a_i)$ 
    Next
    Min-Roughness ( $a_i$ ) = Min ( $Rough_{a_j}(a_i)$ )
  Next
  Set Min-Min-Roughness = Min (Min-Roughness ( $a_i$ )),  $i = 1, \dots, n$ 
  Determine splitting attribute  $a_i$  corresponding to the Min-Min-Roughness
  Do binary split on the splitting attribute  $a_i$ 
  CNC = the number of leaf nodes
  Go to Label 1
End

ProcParentNode (CNC)
Begin
  Set  $i = 1$ 
  Do until  $i < CNC$ 
    Size ( $i$ ) = Count (Set of Elements in Cluster  $i$ )
     $i = i + 1$ 
  Loop
  Determine Max (Size ( $i$ ))
  Return (Set of Elements in cluster  $i$ ) corresponding to Max (Size ( $i$ ))
End

```

Figure 2.8. The MMR algorithm

The accuracy of clusters was used as a measure to test the quality of the clusters. The accuracy of a cluster is defined as :

$$\text{Accuracy}(i) = \frac{\text{the number of data occurring in both the } i\text{th cluster and its corresponding class}}{\text{the number of data in the data set}} \quad (2.16)$$

The overall accuracy is defined as

$$\text{Over all accuracy} = \frac{\sum_{i=1}^{n \text{ of Cluster}} \text{Purity}(i)}{n \text{ of Cluster}} \quad (2.17)$$

2.6.3. Maximum Dependency Attribute (MDA) Technique

Maximum Dependency of Attributes, which is proposed by (Herawan *et al.*, 2010), disseminate the attributes of information system by using the attribute dependencies. MDA has a four steps algorithm which starts with a search for indiscernibility relation. Subsequently, attribute dependency for each attribute in the information system is calculated. If given two attributes, $a_i, a_j \in A$, the dependency degree (in k) of attribute a_i on attribute a_j denoted by $a_j \Rightarrow_k a_i$ is calculated using the following equation:

$$k_{a_j}(a_i) = \frac{\sum_{X \in (U/a_i)} |a_j(X)|}{|U|} \quad (2.18)$$

The highest dependency degree (Max-Dependency), which is denoted as MD for each $a_i \in A$, $1 \leq i \leq |A|$, is selected in the following step. If given m attributes, then, MD of attribute $a_i, a_j \in A$ is defined as

$$MD(a_i) = \max(k_{a_i}(a_i), \dots, k_{a_j}(a_i), \dots, k_{a_m}(a_i)) \quad (2.19)$$

where $a_i \neq a_j$ and $1 \leq i, j \leq m$. Lastly, MDA technique suggested that the highest MD is selected as a clustering attribute using equation (2.20). If there are more than one attribute that had shared the highest k value, then, the next highest k in each attribute will be used and compared until the tie is broken.

$$MDA = \max(MD(a_1), \dots, MD(a_i), \dots, MD(a_m)) \quad (2.20)$$

In reviewing the above three techniques to handle large datasets, even with increasing computing capabilities, the complexity of computation is still an outstanding issues. From the analysis, it has been found that the two techniques, i.e., MMR and MDA, have suffered with ‘unpredicted iteration’. If there are n objects in the data-set, by taking the worst case scenarios, this situation can trigger n^2 of processing time

which can increase the computational complexity when involve large data-sets. Although TR does not face the same problems as discussed above, TR initially has insufficient computational time as discussed in (Herawan *et al.*, 2010). Therefore, there is a need for a technique in clustering attribute selection to further improve the computational complexity. The details of the MDA algorithm is shown in Figure 2.9.

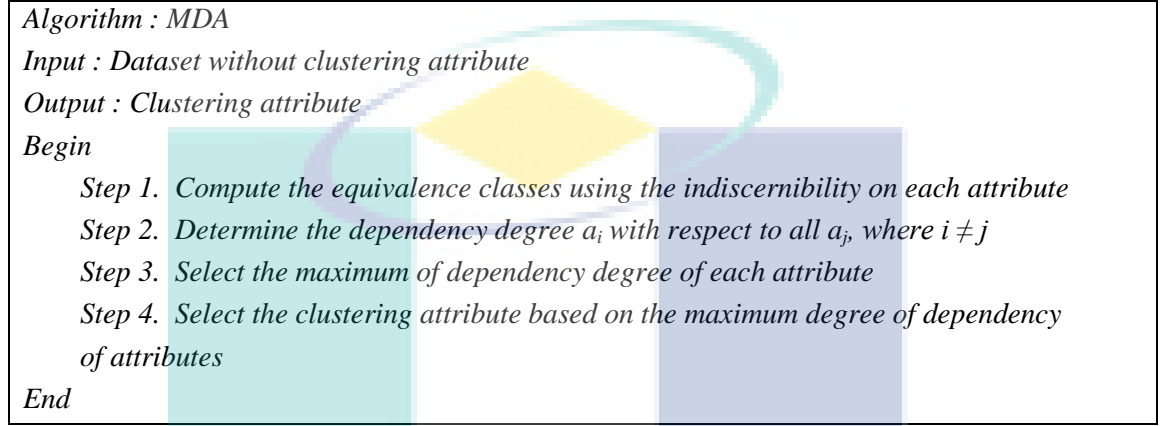


Figure 2.9. The MDA algorithm

2.7. Soft Set Theory

2.7.1. Information Systems

An *information system* as defined in (Pawlak, 2007) is a 4-tuple (quadruple) $S = (U, A, V, f)$, where $U = \{u_0, u_1, \dots, u_{|U|-1}, u_{|U|}\}$ is a non-empty finite set of objects, $A = \{a_0, a_1, \dots, a_{|A|-1}, a_{|A|}\}$ is a non-empty finite set of attributes, $V = \bigcup_{e_i \in A} V_{e_i}$, where V_a is the domain (value set) of attribute a , $f : U \times A \rightarrow V$ is an information function, such that, $f(x, a) \in V_a$, for every $(x, a) \in U \times A$, is called information (knowledge) function.

The amount of information generated by an application can be valuable for data mining purposes (Kong *et al.*, 2014). An information system can be intuitively expressed in terms of an information table such as in Table 2.1.

Table 2.1. An information system

U/A	a_1	a_2	\dots	a_k	\dots	$a_{ A }$
u_1	$f(u_1, a_1)$	$f(u_1, a_2)$	\dots	$f(u_1, a_k)$	\dots	$f(u_1, a_{ A })$
u_2	$f(u_2, a_1)$	$f(u_2, a_2)$	\dots	$f(u_2, a_k)$	\dots	$f(u_2, a_{ A })$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
$u_{ U }$	$f(u_{ U }, a_1)$	$f(u_{ U }, a_2)$	\dots	$f(u_{ U }, a_k)$	\dots	$f(u_{ U }, a_{ A })$

In many applications, there is an outcome of classification that is known. This 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 $D = (U, A \cup \{d\}, V, f)$, where $d \notin A$ is the *decision attribute*.

2.7.2. Definition of Soft Set Theory

Throughout this section, a set U refers to a non-empty initial universe, E is a set of parameters describing objects in U , $P(U)$ is the power set of U and $A \subseteq E$.

Definition 2.1: See (Molodtsov, 1999) A pair (F, A) is called a soft set over U , where F is a mapping given by

$$F : A \rightarrow P(U) \quad (2.21)$$

In other words, a soft set (F, A) over U is a parameterized family (subset) of the universe U . For $\alpha \in A$, $F(\alpha)$ may be considered as the set of α -elements of the soft set $F(A)$ or the set α -approximate elements of the soft set $F(A)$. Clearly, a soft set is not a (crisp) set.

Example : Let a universe $U = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\}$ be a set of candidates and a set of parameters $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ be a set of soft skills which stand for the parameters “communicative”, “critical thinking”, “team work”, “information management”, “entrepreneurship”, “leadership” and “moral”, respectively. Consider F

be a mapping of E into the set of all subsets of the set U as $F(e_1) = \{c_1, c_2, c_4, c_5\}$,
 $F(e_2) = \{c_3, c_8, c_9\}$, $F(e_3) = \{c_6, c_9, c_{10}\}$, $F(e_4) = \{c_2, c_3, c_4, c_5, c_8\}$,
 $F(e_5) = \{c_2, c_5, c_6, c_7, c_8, c_9, c_{10}\}$, $F(e_6) = \{c_6, c_9, c_{10}\}$ and $F(e_7) = \{c_6, c_9, c_{10}\}$. Now
 consider a soft set (F, E) , which describes the “capabilities of the candidate for hire”.
 According to the data collected, the soft set (F, E) is given by

$$(F, E) = \left\{ \begin{array}{l} \text{communicative} = \{c_1, c_2, c_4, c_5\}, \\ \text{critical thinking} = \{c_3, c_8, c_9\}, \\ \text{team work} = \{c_6, c_9, c_{10}\}, \\ \text{information management} = \{c_2, c_3, c_4, c_5, c_8\}, \\ \text{entrepreneurship} = \{c_2, c_5, c_6, c_7, c_8, c_9, c_{10}\}, \\ \text{leadership} = \{c_6, c_9, c_{10}\}, \\ \text{moral} = \{c_6, c_9, c_{10}\} \end{array} \right\}$$

Obviously, the soft set (F, E) is not a crisp set and (F, E) is a parameterized family $\{F(e_i), i = 1, 2, 3, \dots, 7\}$ of subsets of the set U that have two parts of approximation: predicate (p) and value (v). For example, for the approximation “moral = $\{c_6, c_9, c_{10}\}$ ”, p is moral and $v = \{c_6, c_9, c_{10}\}$.

Definition 2.2. : Let $S = (U, A, V_{\{0,1\}}, f)$ be an information system. If $V_a = \{0,1\}$, for every $a \in A$, then $S = (U, A, V_{\{0,1\}}, f)$ is called a Boolean-valued information system.

Proposition 1: See (Herawan, 2011) Each soft set can be considered as a Boolean-valued information system.

Proof: Let (F, E) be a soft-set over the universe U , $S = (U, A, V, f)$ be an information system. Obviously, the universe U in (F, E) can be considered as the universe U , the parameter set E may be considered as the attributes A . Then, the information function, f is defined by

$$f = \begin{cases} 1, & h \in F(e) \\ 0, & h \notin F(e) \end{cases} \quad (2.22)$$

That is, when $h_i \in F(e_j)$, where $h_i \in U$ and $e_j \in E$, then $f(h_i, e_j) = 1$, otherwise $f(h_i, e_j) = 0$. For this, given $V(h_i, e_j) = \{0, 1\}$. Therefore, a soft set (F, E) can be considered as a Boolean-valued information system where $S = (U, A, V_{\{0,1\}}, f)$ and a soft set (F, E) can be represented in the form of Boolean table.

From Proposition 1, a soft set in (1) can be easily represented in Boolean table as follow.

Table 2.2. Tabular representation of soft set (F, E) in (1)

U / E	e_1	e_2	e_3	e_4	e_5	e_6	e_7
c_1	1	0	0	0	0	0	0
c_2	1	1	0	1	1	0	0
c_3	0	0	0	1	0	0	0
c_4	1	0	0	1	0	0	0
c_5	1	0	0	1	1	0	0
c_6	0	0	1	0	1	1	1
c_7	0	0	0	0	1	0	0
c_8	0	1	0	1	1	0	0
c_9	0	1	1	0	1	1	1
c_{10}	0	0	1	0	1	1	1

As can be seen in Table 2.2, '1' denotes the presence of the described parameters, while '0' means the parameter which is not part of the capabilities of the candidate.

Definition 2.3. See (Maji and Roy, 2002) The class of all value sets of a soft set (F, E) is called value-class of the soft set and is denoted by $C_{(F, E)}$.

2.7.3. Multi Soft Sets

Multi Soft Sets (Herawan, 2009) describes the idea of decomposition of the multi-value information system $S = (U, A, V, f)$ into $|A|$ numbers of Boolean-valued information system $S^i = (U, a_i, V_{\{0,1\}}, f)$, where $|A|$ is the cardinality of A . The decomposition of $S = (U, A, V, f)$ is based on decomposition of $A = \{a_1, a_2, \dots, a_{|A|}\}$ into the disjoint-singleton attribute $\{a_1\}, \{a_2\}, \dots, \{a_{|A|}\}$. At this stage, only complete information system is given the consideration. Let $S = (U, A, V, f)$ be an information system such that for every $a \in A$, $V_a = f(U, a)$ is a finite non-empty set and for every $u \in U$, $|f(u, a)| = 1$. For every a_i under i^{th} -attribute consideration, $a_i \in A$ and $v \in V_a$, Definition of map $a_v^i : U \rightarrow \{0,1\}$ such that $a_v^i(u) = 1$ if $f(u, a) = v$, otherwise $a_v^i(u) = 0$. The next result, A definition of binary-valued information systems as quadruple $S^i = (U, a_i, V_{\{0,1\}}, f)$. The information systems $S^i = (U, a_i, V_{\{0,1\}}, f)$, $i = 1, 2, \dots, |A|$ is referred to as a decomposition of a multi-valued information system $S = (U, A, V, f)$ into $|A|$ binary-valued information systems, as depicted in Figure 3.2. Every information system $S^i = (U, a_i, V_{a_i}, f)$, $i = 1, 2, \dots, |A|$ is a deterministic information system since for every $a \in A$ and for every $u \in U$, $|f(u, a)| = 1$ such that the structure of a multi-valued information system and $|A|$ number of binary-valued information systems give the same value of attribute related to objects.

Based on the notion of a decomposition of a multi-valued information system in the previous sub-section, in this sub-section, the presentation of notion multi-soft set representing multi-valued information systems is mentioned. Let $S = (U, A, V, f)$ be a multi-valued information system and $S^i = (U, a_i, V_{a_i}, f)$, $i = 1, 2, \dots, |A|$ be the $|A|$ binary-valued information systems. From Proposition 1, be given

$$\begin{aligned}
S = (U, A, V, f) &= \begin{cases} S^1 = (U, a_1, V_{\{0,1\}}, f) & \Leftrightarrow (F, a_1) \\ S^2 = (U, a_2, V_{\{0,1\}}, f) & \Leftrightarrow (F, a_2) \\ \vdots & \vdots \\ S^{|A|} = (U, a_{|A|}, V_{\{0,1\}}, f) & \Leftrightarrow (F, a_{|A|}) \end{cases} \\
&= ((F, a_1), (F, a_2), \dots, (F, a_{|A|}))
\end{aligned}$$

There are defined $(F, E) = ((F, a_1), (F, a_2), \dots, (F, a_{|A|}))$ as a multi-soft set over universe U representing a multi-valued information system $S = (U, A, V, f)$.

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

Binary-valued information system-1						
U	a_1					
	Va_{11}	Va_{12}	...	Va_{1k}	...	Va_{1n}
u_1	0	1	...	0	...	0
u_2	0	0	...	1	...	0
u_3	0	1	...	0	...	0
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
$u_{ U }$	1	0	...	0	...	0

Binary-valued information system- $ A $						
U	$a_{ A }$					
	$Va_{ A 1}$	$Va_{ A 2}$...	$Va_{ A k}$...	$Va_{ A n}$
u_1	0	0	...	1	...	0
u_2	0	0	...	0	...	1
u_3	0	1	...	0	...	0
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
$u_{ U }$	1	0	...	0	...	0

Figure 2.10. A decomposition of a multi-valued information system

2.7.4. Applications of Soft Set Theory

Soft set theory can be applied to various fields to the problem of uncertainty, and would be beneficial to extend the method proposed for further research. However, the approach should be more comprehensive in the future to solve related problems. Soft set theory is growing fast, some operations and application of soft sets are studied by many researchers to provide the first practical application of soft sets in decision-making problems. In addition soft set theory can also be applied to game theory and operations research. Soft set theory can improve the forecasting performance in terms of accuracy and stability (Xu et al., 2014). Attribute reduction in soft set is combinatorial problems. The use of soft set theory optimization algorithm can be done quickly (Kong *et al.*, 2015). Decision making on soft set theory can be applied to various fields that contain uncertainties such as computer science, game theory, and so on (Deli and Cagman, 2015).

2.8. Soft Set-based Clustering Attribute Selection Techniques

Here are two clustering techniques that use soft set theory, namely Novel soft sets and Maximum Relative Attribute.

2.8.1. A Novel Soft Set (NSS) Technique

A novel soft-set (NSS) approach has been proposed by (Hongwu *et al* in, 2010). In this approach, the re-definition of attribute dependency in the rough set theory is given in the context of soft set theory. In NSS, the degree of dependency is defined as:

$$k_{a_j}(a_i) = \frac{|\underline{a_j}(a_i)|}{|U|}. \quad (2.23)$$

The cardinality of $\underline{a_i}(a_j)$ can be calculated as

$$|\underline{a_i}(a_j)| = \sum_{k=D+1}^{D+|U/a_j|} |\underline{a_i}(e_k)|, \quad (2.24)$$

where lower approximation of attribute a_j with respect to attribute a_i is defined as

$$\underline{a}_i(a_j) = \{x | x \in \underline{a}_i(e_k), k = D+1, \dots, D+|U/a_j|\} \quad (2.25)$$

$$\text{where } D = \sum_{l=1}^{j-1} |U/a_l|.$$

As another alternative approach of the rough set theory, the NSS takes the highest total degree of dependency as the selected clustering attribute. The details of the NSS algorithm is shown in Figure 2.10.

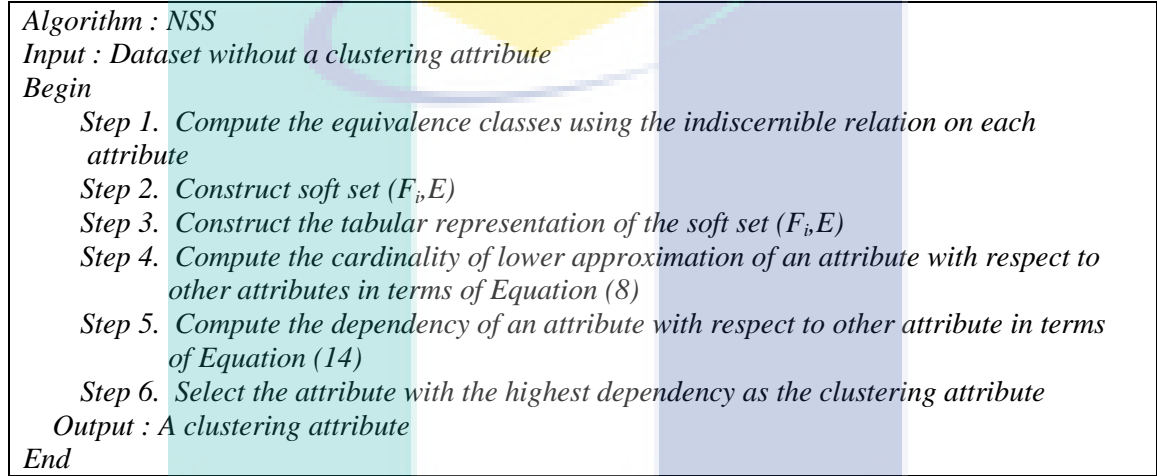


Figure 2.11. The NSS algorithm

2.8.2. Maximum Attribute Relative (MAR) Technique

A Maximum Attribute Relative (MAR) approach has been proposed by (Mamat *et al*, 2013). Throughout this section, a pair (F, A) , refers to multi-soft sets over the universe U representing a categorical valued information system $S = (U, A, V, f)$. Let (F, A) be a multi soft-sets over the universe U , where $(F, a_i), \dots, (F, a_{|A|}) \subseteq (F, A)$ and $(F, a_{i_j}), \dots, (F, a_{i_{|a_i|}}) \subseteq (F, a_i)$. Support of (F, a_{i_j}) by (F, a_{i_k}) denoted $\sup_{(F, a_{i_k})}(F, a_{i_j})$

is defined as $\sup_{(F, a_{i_k})}(F, a_{i_j}) = \frac{|(F, a_{i_j}) \cap (F, a_{i_k})|}{|(F, a_{i_k})|}$. Maximum support is a summation of

all support with value equals to 1. For each soft set (F, a_{i_j}) , the maximum support, $\max \sup_{(F, a_{i_j})}$, is defined as $\max \sup_{(F, a_{i_j})} = \sum (\sup_{(F, a_{i_k})}(F, a_{i_j}) = 1)$. Minimum support

is a summation of all support with value less than 1. For each soft set (F, a_{i_j}) , the minimum support is denoted by $\min \sup_{(F, a_{i_j})}$ is defined as $\min \sup_{(F, a_{i_j})} = \sum (\sup_{(F, a_{i_k})} (F, a_{i_j}) \neq 1)$. If

$\text{Mode} \left(\max \left(\max \sup_{(F, a_{i_j})}, \dots, \max \sup_{(F, a_{|m||n|})} \right) = 1 \right)$ then (F, a_{i_j}) is a clustering

attribute. If $\text{Mode} \left(\max \left(\max \sup_{(F, a_{i_j})}, \dots, \max \sup_{(F, a_{|m||n|})} \right) > 1 \right)$ then

$\max \left(\min \sup_{(F, a_{i_j})}, \dots, \min \sup_{(F, a_{|m||n|})} \right)$ is a clustering attribute, where max refers to the

value that is the highest in the probability distribution and mode refers to the value that is most frequently occurred in the probability distribution. The details of the MAR algorithm is shown in Figure 2.12.

```

Input: Categorical-valued data-set
Output: A Clustering attribute
Begin
1. Builds the multi-soft set approximation
2. Calculate Support, MaxSup and MinSuP
   for i = all categories
     for j = all categories
       intersection = Data(i) And Data(j)
       Sup(i,j) = Intersection/Data(j)
       if Sup (i,j) = 1 then
         MaxSup(i) = MaxSup(i) + Sup(i,j)
       else
         MinSup(i) = MinSup(i) + Sup(i,j)
       end
     end
   end
3. Finding Clustering Attribute
   if Mode(Max(MaxSup(data(1)..data(n)))) = 1 then
     Clustering Attribute = Max(MaxSup(Data(1)..Data(n)))
   else
     Clustering Attribute = Max(MinSup(Data(1)..Data(n)))
   end
End

```

Figure 2.12. The MAR algorithm

2.9. Educational Data Mining

The Educational Data Mining community has undergone tremendous growth in the past decade (Beck, 2014). Data Mining is very important in the field of education, especially when examining the behavior in the learning environment (Mohamad and Tasir, 2013). Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. Machine learning and data mining have developed a vast array of tools and techniques that have been well studied and examined in myriad context. Educational data mining can be applied to wide areas of research including elearning, intelligent tutoring systems, text mining, social network mining and etc. In education, EDM can function as a replacement for less accurate but more established psychometric techniques. Educational data mining is an interactive cycle of hypothesis formation, testing and refinements that alternates between two complementary types of activities. One type of activity is qualitative analysis, focuses on understanding individual tutorial events. The other type is the knowledge, the analysis of growth, and seize opportunities to improve the abilities and skills (Sachin and Vijay, 2012).

The emerging fields of academic analytics and educational data mining are rapidly producing new possibilities for gathering, analyzing, and presenting student data. University might soon be able to use these new data sources as guides for course redesign and as evidence for implementing new assessments and lines of communication between instructors and students. The EDM links the concepts of academic analytics, data mining in higher education, and course management system audits and suggest how these techniques and the data they produce might be useful to those who practice the scholarship of teaching and learning (Baepler and Murdoch, 2010).

The EDM process converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. This process does not differ much from other application areas of Data Mining (DM), like business, genetics, medicine and etc., because it follows the same

steps as the general DM process (Romero *et al*, 2004). Even so, there are some important issues that differentiate the application of DM, specifically to education, from how it is applied in other domains :

- a. Objective: The objective of DM in each application area is different. For example, in EDM, there are both applied research objectives, such as improving the learning process and guiding students' learning, as well as pure research objectives, such as achieving a deeper understanding of educational phenomena. These goals are sometimes difficult to quantify and require their own special set of measurement techniques.
- b. Data: In educational environments, there are many different types of data available for mining. These data are specific to the educational area, and therefore have intrinsic semantic information, relationships with other data, and multiple levels of meaningful hierarchy.
- c. Techniques: Educational data and problems have some special characteristics that require the issue of mining to be treated in a different way. Although most of the traditional DM techniques can be applied directly, others cannot and have to be adapted to the specific educational problem.

There are actually more groups involved with many more objectives, namely :

- a. Learners/students
- b. Educators/lecturers
- c. Course/educational/researchers
- d. Organizations/learning/providers/universities/privat training companies
- e. Administrators

Lecturers and academics section are in charge of planning, designing, building and maintaining the educational systems. Students use and interact with them. Starting from all the available information about courses, students, usage and interaction, different data mining techniques can be applied in order to discover useful knowledge that helps to improve the learning process. The discovered knowledge can be used not only by providers (lecturers) but also by own users (students). So, the application of data mining in educational systems can be oriented to different actors with each particular point of view (Figure 2.13). An existing learning management system is

improved by using data mining techniques and increasing the efficiency of the courses using custom modules (Blagojevic and Micic, 2012).

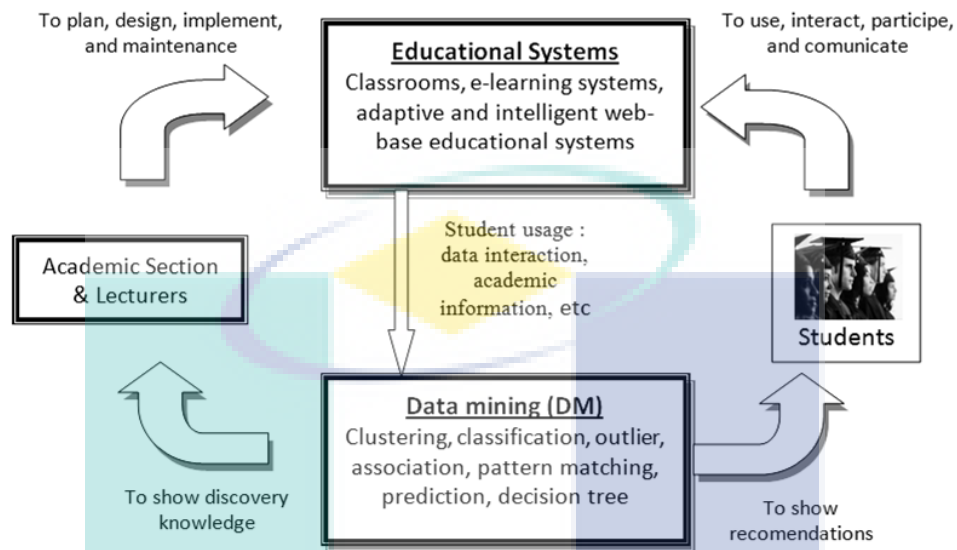


Figure 2.13. The cycle of applying data mining in educational systems

Model analysis of data mining and computer science heuristic assist in the analysis, data visualization, prediction, recommendation, provide feedback, identification, detection, and planning a variety of other educational activities (Wassan, 2015).

UMP

2.10. Discussion

Several clustering techniques have been presented in the previous section. In Table 2.3 discusses the advantages and disadvantages of each techniques.

Table 2.3. Summary of clustering techniques

No	Technique	Advantages	Disadvantages
1	K-means, k-mode, fuzzy k-modes, and fuzzy centroids	Linear and efficient for large data sets. Simple and fast	Require multiple runs with the initial values of the different modes to test the stability of clustering solutions
2	ROCK and QROCK	Strong clustering technique for categorical. Able to explore the concept of link to the data with the attribute category.	Sensitive to the threshold value. Produce a large cluster that includes the object of most of the class. Not guarantee the number of clusters generated.
3	COOLCAT, LIMBO	Outperform ROCK. ROCK has several limitations. The order processing point has a definite impact on the quality of the clustering.	Low accuracy and high computational complexity. The clustering results may be affected by the sample size and the distribution of the real
4	STIRR	An iterative algorithm based on nonlinear dynamic systems	Difficult to analyze the stability of the system for each combiner function when is useful
5	CACTUS	Finding clusters in a subset of all the attributes. Outperform STIRR	The algorithm unstable
6	Squeezer	Suitable for clustering data stream since it scans each tuple only once	Each dataset need a different threshold that makes the selection of threshold a difficult work for users
7	Total Roughness (TR)	Can be applied to more than one semantic domain.	The main difficulty is combinatorial explosion.
8	Min-Min Roughness (MMR)	The ability to handle uncertainty in the clustering process.	Low accuracy and high computational complexity.
9	Maximum Dependency Attributes (MDA)	Results of creating a better performance than MMR	Execution time is longer. Has good scalability poorly
10	A Novel Soft Set (NSS)	Giving a quick decision in choosing the clustering attribute then MDA	Execution time is longer. Has good scalability poorly
11	Maximum Relative Attributes (MAR)	Lower computing time Has good scalability	Low efficiency and accuracy

The last clustering technique for selecting attribute is Maximum Relative Attribute (MAR). It is based on a soft set theory by introducing the concept of the attribute relative in information systems. Based on the experiment on fourteen UCI datasets and a supplier dataset, the MAR technique achieved a lower execution time than the three rough set-based techniques, i.e. TR, MMR, MDA, and NSS. Furthermore, MAR has a good scalability, i.e. the executing time of the algorithm tends to increase linearly as the number of instances and attributes are increased respectively.

The MAR technique uses complex mathematical models, because this technique calculates the value of support, max_support and min_support. Support value = 1 is summed as max_support, while others as min_support. Clustering attribute is determined based on the largest amount of max_support. If there are greatest max_support of more than 1, then the clustering attribute is an attribute that has the largest value of min_support. From the analysis of the limitations of the MAR technique, there is a need to develop technique for clustering. This new technique is called the Maximum Degree of Domination in Soft set theory (MDDS).


 The logo of UMP (Universitas Muhammadiyah Palembang) is a large, stylized shield-like shape. It is composed of several geometric sections in shades of teal, light blue, and yellow. At the top center is a yellow diamond. Below it, the letters 'UMP' are written in a large, white, bold, sans-serif font. The background of the entire page features a faint, large-scale version of this UMP logo.

UMP

CHAPTER 3

PROPOSED METHOD

3.1.Soft Set Based Technique

The multi-valued information system, as described in (Pawlak, 1982), needs to be converted to multi-soft sets based on the notion of a decomposition of a multi valued information system. Let $S = (U, A, V, f)$ be a multi valued information system and $S^i = (U, a_i, V_{\{0,1\}}, f)$, where $i = 1, 2, \dots, |A|$ be a Boolean valued information system. As a result of decomposition, it is given as below :

$$S = (U, A, V, f) = \begin{cases} S^1 = (U, a_1, V_{\{0,1\}}, f) \Leftrightarrow (F, a_1) \\ \vdots \\ S^{|A|} = (U, a_{|A|}, V_{\{0,1\}}, f) \Leftrightarrow (F, a_{|A|}) \end{cases} \quad (3.1)$$

$$= ((F, a_1), (F, a_2), \dots, (F, a_{|A|}))$$

Defined $(F, E) = ((F, a_1), (F, a_2), \dots, (F, a_{|A|}))$ as a multi-soft sets over universe U representing a multi valued information system $S = (U, A, V, f)$. As an example, by using an information system in Table 3.1. which describes animals and their various attributes, the concept of multi soft-set is illustrated in Figure 3.1.

Table 3.1. Multi-valued information system

No	Animal	Hair	Teeth	Eye	Feather	Feet	Eat	Milk	Fly	Swim
1	Tiger	Y	Pointed	Forward	N	Claw	Meat	Y	N	Y
2	Cheetah	Y	Pointed	Forward	N	Claw	Meat	Y	N	Y
3	Giraffe	Y	Blunt	Side	N	Hoof	Grass	Y	N	N
4	Zebra	Y	Blunt	Side	N	Hoof	Grass	Y	N	N
5	Ostrich	N	N	Side	Y	Claw	Grain	N	N	N
6	Penguin	N	N	Side	Y	Web	Fish	N	N	Y
7	Albatross	N	N	Side	Y	Claw	Grain	N	Y	Y
8	Eagle	N	N	Forward	Y	Claw	Meat	N	Y	N
9	Viper	N	Pointed	Forward	N	N	Meat	N	N	N

From Table 3.1, the multi-soft sets is made as shown in Figure 3.1.

$$(F, E) = \left\{ \begin{array}{l} (F, \text{Hair}) = \begin{cases} (F, \text{Hair}_{Yes}) = \{1, 2, 3, 4\} \\ (F, \text{Hair}_{No}) = \{5, 6, 7, 8, 9\} \end{cases} \\ (F, \text{Teeth}) = \begin{cases} (F, \text{Teeth}_{Pointed}) = \{1, 2, 9\} \\ (F, \text{Teeth}_{Blunt}) = \{3, 4\} \\ (F, \text{Teeth}_{No}) = \{5, 6, 7, 8\} \end{cases} \\ (F, \text{Eye}) = \begin{cases} (F, \text{Eye}_{Forward}) = \{1, 2, 8, 9\} \\ (F, \text{Eye}_{Side}) = \{3, 4, 5, 6, 7\} \end{cases} \\ (F, \text{Feather}) = \begin{cases} (F, \text{Feather}_{No}) = \{1, 2, 3, 4, 9\} \\ (F, \text{Feather}_{Yes}) = \{5, 6, 7, 8\} \end{cases} \\ (F, \text{Feet}) = \begin{cases} (F, \text{Feet}_{Claw}) = \{1, 2, 5, 7, 8\} \\ (F, \text{Feet}_{Hoof}) = \{3, 4\} \\ (F, \text{Feet}_{Web}) = \{6\} \\ (F, \text{Feet}_{No}) = \{9\} \end{cases} \\ (F, \text{Eat}) = \begin{cases} (F, \text{Eat}_{Meat}) = \{1, 2, 8, 9\} \\ (F, \text{Eat}_{Grass}) = \{3, 4\} \\ (F, \text{Eat}_{Grain}) = \{5, 7\} \\ (F, \text{Eat}_{Fish}) = \{6\} \end{cases} \\ (F, \text{Milk}) = \begin{cases} (F, \text{Milk}_{Yes}) = \{1, 2, 3, 4\} \\ (F, \text{Milk}_{No}) = \{5, 6, 7, 8, 9\} \end{cases} \\ (F, \text{Fly}) = \begin{cases} (F, \text{Fly}_{No}) = \{1, 2, 3, 4, 5, 6, 9\} \\ (F, \text{Fly}_{Yes}) = \{7, 8\} \end{cases} \\ (F, \text{Swim}) = \begin{cases} (F, \text{Swim}_{Yes}) = \{1, 2, 6, 7\} \\ (F, \text{Swim}_{No}) = \{3, 4, 5, 8, 9\} \end{cases} \end{array} \right.$$

Figure 3.1 Multi-soft set composition from dataset

Definition 3.1. See (Herawan et al., 2011) *The class of all value sets of a soft set (F, E) is called value-class of the soft set and is denoted by $C_{(F, E)}$.*

3.2. Maximum Attribute Relative (MAR) Technique

The MAR technique (Mamat, 2013), a pair (F, A) , refers to multi-soft sets over the universe U representing a categorical valued information system $S = (U, A, V, f)$.

Definition 3.2: Let (F, A) be a multi soft-sets over the universe U , where $(F, a_i), \dots, (F, a_{|A|}) \subseteq (F, A)$ and $(F, a_{i_j}), \dots, (F, a_{i_{|a_i|}}) \subseteq (F, a_i)$. Support of (F, a_{i_j}) by (F, a_{i_k}) denoted $\sup_{(F, a_{i_k})}(F, a_{i_j})$ is defined as

$$\sup_{(F, a_{i_k})}(F, a_{i_j}) = \frac{|(F, a_{i_j}) \cap (F, a_{i_k})|}{|(F, a_{i_k})|} \quad (3.2)$$

Example :

$$Teeth_{\text{Pointed}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,9\}|} = \frac{2}{3} = 0.6667 \quad Teeth_{\text{Blunt}} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

Definition 3.3: Maximum support is a summation of all support with value equals to 1. For each soft set (F, a_{i_j}) , the maximum support, $\max \sup_{(F, a_{i_j})}$, is defined as

$$\max \sup_{(F, a_{i_j})} = \sum (\sup_{(F, a_{i_k})}(F, a_{i_j}) = 1) \quad (3.3)$$

Example :

$$\text{Max Sup } (F, \text{Hair}_{\text{Yes}}) = (1, 0, 0.6667, 1, 0, 0.5, 0.4, 0, 0.8, 0.4, 1, 0, 0, 0.5, 1, 0, 0, 1, 0, 0, 0.5714, 0.5, 0.4) = 5$$

Definition 3.4: Minimum support is a summation of all support with value less than 1. For each soft set (F, a_{i_j}) , the minimum support is denoted by $\min \sup_{(F, a_{i_j})}$ is defined as

$$\min \sup_{(F, a_{i_j})} = \sum (\sup_{(F, a_{i_k})}(F, a_{i_j}) \neq 1) \quad (3.4)$$

Example :

$$\text{Min Sup } (F, \text{Hair}_{\text{Yes}}) = (1, 0, 0.6667, 1, 0, 0.5, 0.4, 0, 0.8, 0.4, 1, 0, 0, 0.5, 1, 0, 0, 1, 0, 0, 0.5714, 0.5, 0.4) = 4.738$$

Definition 3.5: Mode refers to the value that most frequently occurs in the probability distribution.

Definition 3.6: Max refers to the value that is the highest in the probability distribution.

Proposition 2: If $\text{Mode}\left(\max\left(\max\sup_{(F,a_{i_j})}, \dots, \max\sup_{(F,a_{|m||n|})}\right) = 1\right)$ then (F, a_{i_j}) is a clustering attribute.

Proof: Let (F, a_i) and (F, a_j) be two soft sets over the universe U , if $a_i = a_j$ then the support value of (F, a_i) is equal to 1, therefore it is said that a_i is relative to a_j . Therefore, a_i can be used to describe a_j and vice-versa. If $a_i \neq a_j$, then there exists (F, a_k) where $a_i = a_k$ and $a_j \neq a_k$ then, the support value of (F, a_i) is greater than 1. Based on Definition 3.2, it is clear that (F, a_i) is selected as a clustering attribute.

Corollary 1: If $\text{Mode}\left(\max\left(\max\sup_{(F,a_{i_j})}, \dots, \max\sup_{(F,a_{|m||n|})}\right) > 1\right)$ then $\max\left(\min\sup_{(F,a_{i_j})}, \dots, \min\sup_{(F,a_{|m||n|})}\right)$ is a clustering attribute.

Proof: The proof is clear from Definition 3.3 and Proposition 2.

3.2.1. The MAR Technique

The MAR technique was described in Figure 2.12. In this section, the calculation using the MAR is served. Dataset as in Table 3.1 are used as input data. Objects dataset contains 5 and 6 attributes. The technique begins by changing the input dataset into multiple sets of gentle, as shown in Figure 3.1. Second, the value of support for each other soft-set is obtained by using the formula as defined in Definition 2.3. MAR of all the attributes are shown in appendix C1.

All values support soft set are summarized in Table 3.2. After the value of the support has been raised, the calculation of the maximum support and minimum values for soft sets can now be done directly by using Definition 3.1 and 3.2 respectively. The first definition, the index of support for each set soft is counted. Finally, a clustering of attributes are selected using Definition 2.2. Table 3.3, the attribute as an attribute clustering. Fly is selected because it has the highest support index. This is followed by Milk and others.

Table 3.2 Support value of MAR technique of all soft set

Soft set	Support Value
$(F, Hair_{Yes})$	1,0,0.6667,1,0,0.5,0.4,0,0.8,0.4,1,0,0,0.5,1,0,0,1,0,0,0.5714,0.5,0.4
$(F, Hair_{No})$	0,1,0.3333,0,1,0.5,0.6,1,0.2,0.6,0,1,1,0.5,0,1,1,0,1,1,0.4286,0.5,0.6
$(F, Teeth_{Pointed})$	0.5,0.2,1,0,0,0.75,0,0,0.6,0.4,0,0,1,0.75,0,0,0.5,0.25,0,0.4286,0.5,0.2
$(F, Teeth_{Blunt})$	0.5,0,0,1,0,0,0.4,0,0.4,0,1,0,0,0,1,0,0,0.5000,0,0,0.2857,0,0.4
$(F, Teeth_{No})$	0,0.8,0,0,1,0.25,0.6,1,0,0.6,0,1,0,0.25,0,1,1,0,0.75,1,0.2857,0.5,0.4
$(F, Eye_{Forward})$	0.5,0.4,1,0,0.25,1,0,0.25,0.6,0.6,0,0,1,1,0,0,0.5,0.5,0.5,0.4286,0.5,0.4
(F, Eye_{Side})	0.5,0.6,0,1,0.75,0,1,0.75,0.4,0.4,1,1,0,0,1,1,1,0.5,0.5,0.5,0.5714,0.5,0.6
$(F, Feather_{No})$	0,0.8,0,0,1,0.25,0.6,1,0,0.6,0,1,0,0.25,0,1,1,0,0.75,1,0.2857,0.5,0.4
$(F, Feather_{Yes})$	1,0.2,1,1,0,0.75,0.4,0,1,0.4,1,0,1,0.75,1,0,0,1,0.25,0,0.7143,0.5,0.6
$(F, Feet_{Claw})$	0.5,0.6,0.6667,0,0.75,0.75,0.4,0.75,0.4,1,0,0,0,0.75,0,1,0,0.5,0.5,1,0.4286,0.75,0.4
$(F, Feet_{Hoof})$	0.5,0,0,1,0,0,0.4,0,0.4,0,1,0,0,0,1,0,0,0.5,0,0,0.2857,0,0.4
$(F, Feet_{Web})$	0,0.2,0,0,0.25,0,0.2,0.25,0,0,0,1,0,0,0,0,1,0,0.25,0,0.1429,0.25,0
$(F, Feet_{No})$	0,0.2,0.3333,0,0,0.25,0,0,0.2,0,0,0,1,0.25,0,0,0,0,0.25,0,0.1429,0,0.2
(F, Eat_{Meat})	0.5,0.4,1,0,0.25,1,0,0.25,0.6,0.6,0,0,1,1,0,0,0.5,0.5,0.5,0.4286,0.5,0.4
(F, Eat_{Grass})	0.5,0,0,1,0,0,0.4,0,0.4,0,1,0,0,0,1,0,0,0.5,0,0,0.2857,0,0.4
(F, Eat_{Grain})	0,0.4,0,0,0.5,0,0.4,0.5,0,0.4,0,0,0,0,0,1,0,0,0.25,0.5,0.1429,0.25,0.2
(F, Eat_{Fish})	0,0.2,0,0,0.25,0,0.2,0.25,0,0,0,1,0,0,0,0,1,0,0.25,0,0.1429,0.25,0
$(F, Milk_{Yes})$	1,0,0.6667,1,0,0.5,0.4,0,0.8,0.4,1,0,0,0.5,1,0,0,1,0,0,0.5714,0.5,0.4
$(F, Milk_{No})$	0,0.8,0.3333,0,0.75,0.5,0.4,0.75,0.2,0.4,0,1,1,0.5,0,0.5,1,0,1,0.5,0.4286,0.25,0.6
(F, Fly_{No})	0,0.4,0,0,0.5,0.25,0.2,0.5,0,0.4,0,0,0,0.25,0,0.5,0,0,0.25,1,0,0.25,0.2
(F, Fly_{Yes})	1, 0.6,1,1,0.5,0.75,0.8,0.5,1,0.6,1,1,1,0.75,1,0.5,1,1,0.75,0,1,0.75,0.8
$(F, Swim_{Yes})$	0.5,0.4,0.6667,0,0.5,0.5,0.4,0.5,0.4,0.6,0,1,0,0.5,0,0.5,1,0.5,0.25,,0.5,0.4286,1,0
$(F, Swim_{No})$	0.5,0.6,0.3333,1,0.5,0.5,0.6,0.5,0.6,0.4,1,0,1,0.5,1,0.5,0,0.5,0.75,0.5,0.5714,0,1

Table 3.3. Overall max and min support from the above example

Soft Set	MaxSup	MinSup	Categorical Rank	Attribute Rank
$(F, Hair_{Yes})$	5	4.738	6	2
$(F, Hair_{No})$	9	4.262	2	
$(F, Teeth_{Pointed})$	2	5.029	9	5
$(F, Teeth_{Blunt})$	3	2.486	8	
$(F, Teeth_{No})$	6	4.486	5	
$(F, Eye_{Forward})$	4	5.329	7	4
(F, Eye_{Side})	7	6.671	4	
$(F, Feather_{No})$	6	4.486	5	3
$(F, Feather_{Yes})$	8	4.514	3	
$(F, Feet_{Claw})$	3	8.245	8	8
$(F, Feet_{Hoof})$	3	2.486	8	
$(F, Feet_{Web})$	2	1.493	9	
$(F, Feet_{No})$	1	1.776	10	7
(F, Eat_{Meat})	4	5.329	7	
(F, Eat_{Grass})	3	2.486	8	
(F, Eat_{Grain})	1	3.693	10	
(F, Eat_{Fish})	2	1.493	9	2
$(F, Milk_{Yes})$	5	4.738	6	
$(F, Milk_{No})$	9	4.262	2	
(F, Fly_{No})	1	3.850	10	1
(F, Fly_{Yes})	11	7.150	1	
$(F, Swim_{Yes})$	3	7.295	8	6
$(F, Swim_{No})$	5	7.705	6	

3.2.2. Complexity of MAR

Suppose that in an information system, there are n objects, m attributes and l is the maximum distinct values of each attribute. Computational cost to determine the elementary set of all attributes is nm . This technique needs $ml(ml-1)$ times to determine the support for each category. Thus, the computational complexity for the MAR technique is $O(ml(ml-1) + nm + 1)$.

3.3. Maximum Degree of Domination in Soft Set Theory (MDDS) Technique

The proposed MDDS, for selecting a clustering attribute is as follows. First, the presentation of the idea of multi-soft sets is to deal with multi-valued information system. Second, the presentation of the notion domination in multi-soft sets. Finally, MDDS is presented to select the best clustering attribute.

Note that from Definition 2.1, the “standard” soft set deals with a binary-valued information system. For a multi-valued information system $S = (U, A, V, f)$, where $V = \bigcup_{a \in A} V_a$, V_a is the domain (value set) of attribute a which has multi values, a decomposition can be made from S into $|A|$ number of binary-valued information systems $S = (U, A, V_{\{0,1\}}, f)$. In this sub-section, presentation of the construction of multi-soft sets is represented as $S = (U, A, V, f)$ based on such decomposition.

3.3.1. Domination in Multi-soft sets

The notion of soft set-based domination is presented as follows.

Definition 3.7. Let (F, A) be multi-soft sets over U representing $S = (U, A, V, f)$ and $(F, a_i), (F, a_j) \in (F, A)$. Soft set (F, a_i) is said to be dominated by (F, a_j) , denoted as $(F, a_i) \leq (F, a_j)$ if for every $X \in C_{(F, a_i)}$, there exist $Y \in C_{(F, a_j)}$, such that $X \subseteq Y$.

The relation of soft set-based domination and functional dependency is presented as follows. The generalized soft set-based domination based on its degree is presented as follows.

Definition 3.8. Let (F, A) be multi-soft sets over U representing $S = (U, A, V, f)$ and $(F, a_i), (F, a_j) \in (F, A)$. (F, a_j) is said to be dominated in degree k by (F, a_i) , denoted as $(F, a_i) \leq_k (F, a_j)$, where

$$k = \left| \bigcup X : X \subseteq Y \right| / |U|, \quad (3.5)$$

Example :

$$Hair_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444 \quad Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

and, $X \in C_{(F,a_i)}$ and $Y \in C_{(F,a_j)}$.

Obviously $0 \leq k \leq 1$. If $k=1$, then (F, a_i) is dominated totally by (F, a_j) . Otherwise, (F, a_i) is dominated partially by (F, a_j) .

3.3.2. The MDDS Technique

Let (F, A) be multi-soft sets over U representing $S = (U, A, V, f)$, based on Definition 11, the soft set (F, a_i) with maximum degree of domination will be selected as a clustering attribute i.e.

$$\max\{k_1, k_2, \dots, k_n\}. \quad (3.6)$$

Example :

$$(F, Hair_{Yes}) = \{0.4444, 0, 0, 0, 0, 0, 0, 0, 0.4444, 0, 0, 0, 0, 0, 0, 0, 0.4444, 0, 0, 0.4444, 0, 0\} = 0.4444$$

Proposition 2. MAR and MDDS select the same clustering attribute.

Proof. Let (F, A) be a multi-soft sets over U representing an information system $S = (U, A, V, f)$. Given $(F, a_h), (F, a_i), (F, a_j) \subseteq (F, A)$, where $h \neq i \neq j$.

Fact: MAR and MDDS select the same clustering attribute, say a_h . Logically, U/a_h will be the coarsest and un-balanced partition. To prove the above proposition, indirect evidence is used as follows.

Suppose MDDS selects another attribute say a_j , where $j \neq h$. Thus, the value of k_j achieves the highest value, for $1 \leq j \leq |A|$. Let say the value achieved on (F, a_j) , where $j \neq l$. Thus from (19), the cardinality $|\bigcup X : X \subseteq Y|$ i.e. the domination degree of (F, a_j) by (F, a_l) must achieve maximum (the highest). To this, from (5.2) U/a_j

must be coarser and un-balanced partition. This contradicts with the fact that the coarsest un-balanced partition is on attribute a_h . The proof is complete. The proposed technique is described in Figure 3.2.

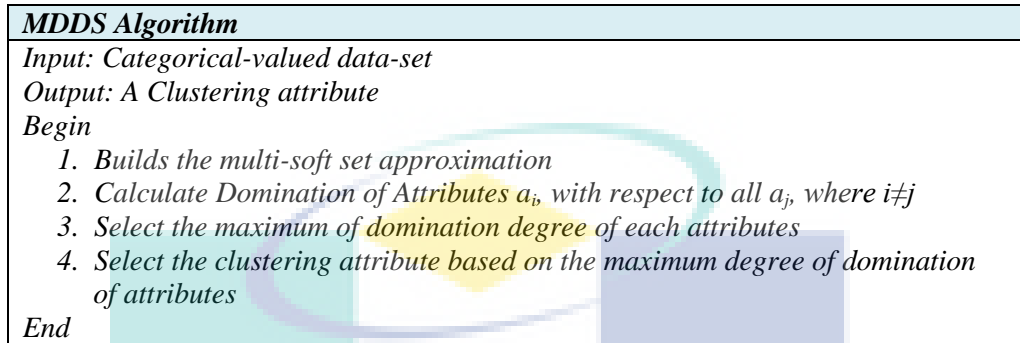


Figure 3.2. The MDDS algorithm

In this section, the calculation using MDDS is presented. Dataset as in Table 3.3 are used as input data. Dataset object contains 9 and 10 attributes. The technique begins with changing the input dataset into multiple soft sets, as shown in Figure 3.1. Second, the value of support for each other gently set by the soft-set is obtained by using the formula as defined in Definition 3.6 and 3.7. Calculation of the MDDS on each attribute from figure 3.1 are shown in appendix C2. The degree of dominance of each attribute calculations are shown in the Table 3.4.

Table 3.4. Domination degree of all softset

Soft Set	Domination Value	Max – Domination
$(F, Hair_{Yes})$	0.4444,0,0,0,0,0,0,0,0.4444,0,0,0,0,0,0,0.4444,0,0,0.4444,0,0	0.4444
$(F, Hair_{No})$	0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0.5556,0,0,0	0.5556
$(F, Teeth_{Pointed})$	0,0,0.3333,0,0,0.3333,0,0,0.3333,0,0,0,0,0.3333,0,0,0,0,0.3333,0,0	0.3333
$(F, Teeth_{Blunt})$	0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0.2222,0,0,0,0.2222,0,0,0.2222,0,0,0.2222,0,0	0.2222
$(F, Teeth_{No})$	0,0.4444,0,0,0.4444,0,0,0.4444,0,0,0,0,0,0,0,0,0.4444,0,0,0,0	0.4444
$(F, Eye_{Forward})$	0,0,0,0,0.4444,0,0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0,0,0	0.4444
(F, Eye_{Side})	0,0,0,0,0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0.5556
$(F, Feather_{No})$	0,0.4444,0,0,0.4444,0,0,0.4444,0,0,0,0,0,0,0,0,0,0.4444,0,0,0,0	0.4444
$(F, Feather_{Yes})$	0,0,0,0,0,0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0.5556,0,0	0.5556
$(F, Feet_{Claw})$	0,0,0,0,0,0,0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0.5556
$(F, Feet_{Hoof})$	0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0.2222,0,0,0,0.2222,0,0,0.2222,0,0,0.2222,0,0	0.2222
$(F, Feet_{Web})$	0,0.1111,0,0,0.1111,0,0.1111,0.1111,0,0,0,0.1111,0,0,0,0.1111,0,0,0,0.1111,0,0.1111,0	0.1111
$(F, Feet_{No})$	0,0.1111,0.1111,0,0,0.1111,0,0.1111,0,0,0.1111,0,0,0,0.1111,0.1111,0,0,0,0.1111,0,0.1111	0.1111
(F, Eat_{Meat})	0,0,0,0,0.4444,0,0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0,0,0	0.4444
(F, Eat_{Grass})	0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0.2222,0,0,0,0.2222,0,0,0.2222,0,0,0.2222,0,0	0.2222
(F, Eat_{Grain})	0,0.2222,0,0,0.2222,0,0.2222,0.2222,0,0.2222,0,0,0,0,0.2222,0,0,0.2222,0,0,0,0	0.2222
(F, Eat_{Fish})	0,0.1111,0,0,0.1111,0,0.1111,0.1111,0,0,0,0.1111,0,0,0,0.1111,0,0,0,0.1111,0	0.1111
$(F, Milk_{Yes})$	0.4444,0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0.4444,0,0,0.4444,0,0	0.4444
$(F, Milk_{No})$	0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5556,0,0,0,0	0.5556
(F, Fly_{No})	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.7778,0,0	0.7778
(F, Fly_{Yes})	0,0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0,0,0,0,0,0,0.2222,0,0,0,0	0.2222
$(F, Swim_{Yes})$	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.4444,0	0.4444
$(F, Swim_{No})$	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5556	0.5556

Table 3.3 and Table 3.4 show that the two techniques give the same result, namely Fly attribute was selected as the most dominant attribute. MDDS execution time is 0.165 seconds, while MAR is 0.350 seconds. The clustering result is based on the splitting attribute Fly :

--> { 1,2,3,4,5,6,9 }

--> { 7,8 }

In this example, the MAR and MDDS gives the same result, where there are more than two attributes, the splitting is on the attribute value which has the overall maximum domination compare to the other attributes. The partition at this stage can be represented as a tree and is shown in Figure 3.4.

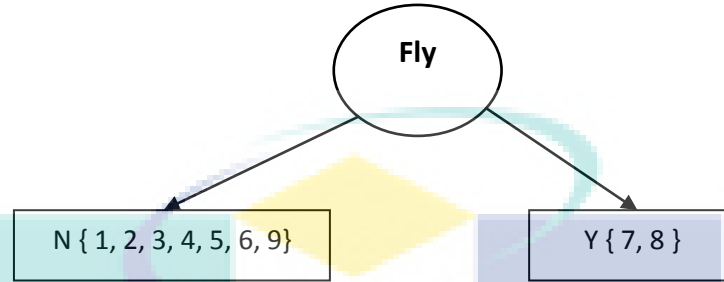


Figure 3.3. Result of clustering

For example from Table 3.3, there are two distribution classes, namely : N {1, 2, 3, 4, 9}, Y {5, 6, 7, 8}. Based on equation of 2.16 and 2.17, then the value of accuracy can be calculated as follows :

$$Fly_{No} = |\{1,2,3,4,5,6,9\} \cap \{1,2,3,4,9\}| = 5 \quad Fly_{No} = |\{1,2,3,4,5,6,9\} \cap \{5,6,7,8\}| = 2$$

$$Purity\ of\ Cluster_1 = \frac{5}{7} = 0.7143$$

$$Fly_{Yes} = |\{7,8\} \cap \{1,2,3,4,9\}| = 0 \quad Fly_{Yes} = |\{7,8\} \cap \{5,6,7,8\}| = 2$$

$$Purity\ of\ Cluster_2 = \frac{2}{2} = 1$$

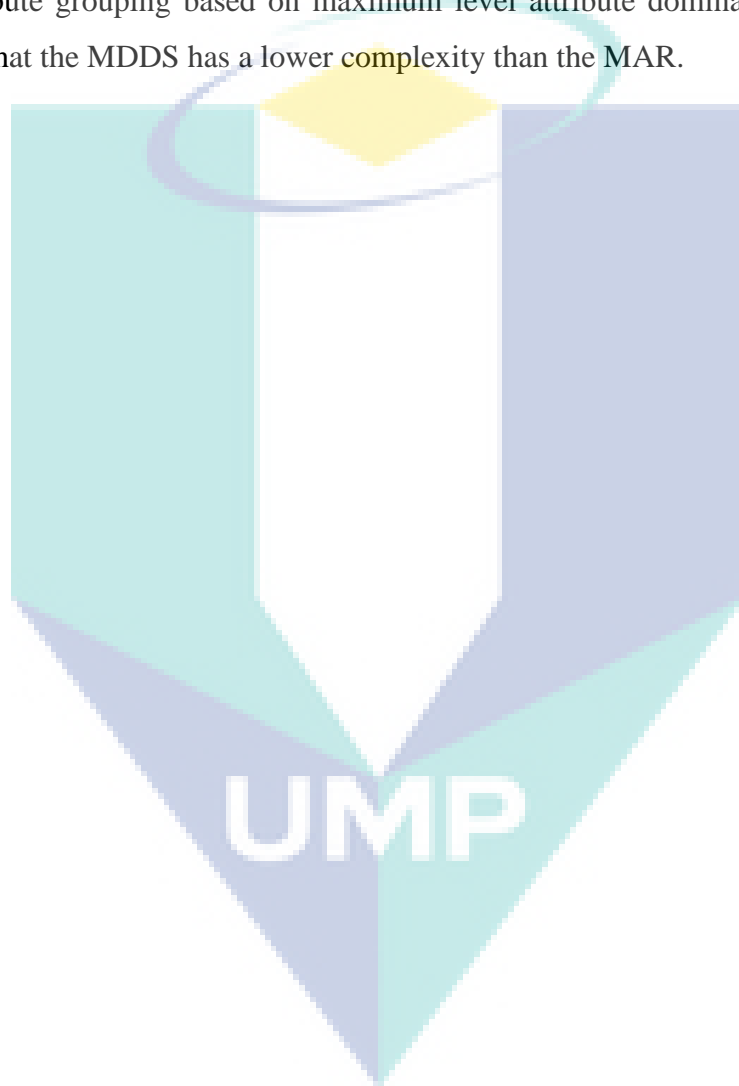
$$Overall\ Purity = \frac{0.7143 + 1}{2} = \frac{1.7143}{2} = 0.8571$$

3.3.3. Complexity of MDDS

Just as in the MAR technique, suppose that in an information system, there are n objects, m attributes and l is the maximum distinct values of each attribute. The computational cost to determine the elementary set of all attributes is nm . The proposed technique needs $m(m-1)$ times to determine the support for each category. The computational complexity for the proposed technique is $O(nm+m(m-1))$. After compared with MAR technique, it is clear that the proposed technique has a lower complexity.

3.4. Summary

This chapter analyzes the complexity of the MAR and MDDS techniques with categorical data as input. The first step is to develop multi soft approach set, then calculate the domination of many attributes a_i , with respect to all a_j , where $i \neq j$. After that the maximum level of dominance of each attribute is selected. The second step is to select attribute grouping based on maximum level attribute dominance. This example illustrates that the MDDS has a lower complexity than the MAR.



CHAPTER 4

EXPERIMENTAL RESULTS

4.1. Experimental Design

This section, The discussion of upon the comparison between MDDS and MAR. While the main focus of the experiments is on the performance measurement of the proposed technique in which execution time and accuracy are used as a parameter.

4.1.1. Data Sets

For comparisons, two techniques which have been discussed will be used with seventeen datasets obtained from the benchmark UCI machine learning repository and a supplier dataset.

a. Acute Inflammation

The acute inflammation data set consists of 120 instances, each of which is described by 6 attributes. The data was created by a medical expert, which will perform the presumptive diagnosis of two diseases of urinary system. The basis for rules detection was Rough Sets Theory. Each instance represents a potential patient. The data is in an ASCII file. Each line of the data file starts with a digit which tells the temperature of the patient. The data set consists of eight attributes, namely:

- a1. Temperature of patient { 35C-42C }
- a2. Occurrence of nausea { yes, no }
- a3. Lumbar pain { yes, no }
- a4. Urine pushing (continuous need for urination) { yes, no }
- a5. Micturition pains { yes, no }
- a6. Burning of urethra, itch, swelling of urethra outlet { yes, no }

d1. decision: Inflammation of urinary bladder { yes, no }

d2. decision: Nephritis of renal pelvis origin { yes, no }

b. Ballon

The ballon data set consists of 16 instances, each of which is described by 4 attributes. Attribute Information: (Classes Inflated T or F). Color: yellow, purple; size: large, small; act: stretch, dip; age: adult, child; inflated: T, F.

c. Bank Marketing

The bank marketing data set consists of 45,211 instances, each of which is described by 17 attributes. The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. Attributes of the data set, namely: age, job, marital, education, default: has credit in default?, housing: has housing loan?, loan, contact, month, day_of_week, duration, campaign, pays, previous, poutcomeemp.var.rate, cons.price.idx: consumer price index, cons.conf.idx: consumer confidence index, euribor3m: euribor 3 month rate-daily indicator, nr.employed: number of employees.

d. Car Evaluation

The car evaluation data set consists of 1,728 instances, each of which is described by 6 attributes. The data set consists of six attributes, namely:

Class Values: unacc, acc, good, vgood.

Attributes: buying: vhigh, high, med, low; maint: vhigh, high, med, low; doors: 2, 3, 4, 5 more; persons: 2, 4, more; lug_boot: small, med, big; safety: low, med, high.

e. Chees

The chees data set consists of 3,196 instances, each of which is described by 36 attributes. There are two classes, namely: white-can-win ("won") and white-cannot-win ("nowin"). Class Distribution : 1,669 objects of the positions (52%), White can win and 1,527 objects of the positions (48%), White cannot win. The

format for instances in this database is a sequence of 37 attribute values. Each instance is a board-description for this chess end game. The first 36 attributes describe the board. The last (37th) attribute is the classification: "win" or "nowin".

f. Cylinder Band

The car evaluation data set consists of 512 instances, each of which is described by 40 attributes including the class attribute, 20 attributes are numeric, 20 are nominal.

g. Flag

The car evaluation data set consists of 194 instances, each of which is described by 30 attributes. This data file contains details of various nations and their flags. From this data, it can be used to predict the religion of a country from its size and the color in its flag. There are 10 attributes containing numeric data and others boolean or nominal value.

h. Lenses

The car evaluation data set consists of 24 instances, each of which is described by 5 attributes. Attribute information, there are three classes, namely: the patient must be equipped with hard contact lenses, soft, or the patient may not be fitted with contact lenses. As for the other attributes, namely : age of the patient : (1) young, (2) pre-presbyopic, (3) resbyopic; spectacle prescription : (1) myope, (2) hypermetrope; astigmatic : (1) no, (2) yes; tear production rate : (1) reduced, (2) normal.

i. Lung Cancer

The lung cancer data set consists of 32 instances, each of which is described by 56 predictive attributes and 1 class attribute. All predictive attributes are nominal, taking on integer values 0-3.

j. Monk

The monk data set consists of 432 instances, each of which is described by 8 attributes and 1 class attribute including class attribute. The MONK's problem

were the basis of a first international comparison of learning algorithms. The result of this comparison is summarized in "The MONK's Problems - A Performance Comparison of Different Learning algorithms". Attribute information: class: 0, 1; a1: 1, 2, 3; a2: 1, 2, 3; a3: 1, 2; a4: 1, 2, 3; a5: 1, 2, 3, 4; a6: 1, 2; Id: (A unique symbol for each instance).

k. Mushroom

The mushroom data set contains 8,124 instances where each instance contains information of a single mushroom. There are 22 categorical attributes to represent the physical characteristic of each mushroom. Each mushroom is classified into two classes: poisonous (3,916 instances) and edible (4,208 instances).

l. Nursery

The nursery data set consists of 12,960 instances, each of which is described by 8 attributes. The Nursery Database contains examples with the structural information removed, i.e., directly relates NURSERY to the eight input attributes: parents, has_nurs, form, hildren, housing, finance, social, health.

m. Solar Flare

The solar flare data set consists of 1,389 instances, each of which is described by 10 attributes. The database contains 3 potential classes, one for the number of times a certain type of solar flare occurred in a 24 hour period. Each instance represents captured features for 1 active region on the sun.

n. Soybean

The soybean small data set consists of 47 instances, each of which is described by 35 attributes. While the soybean large data set consist of 307 instances, each of which is described by 35 attributes. There are 19 classes, only the first 15 of which have been used in prior work. The folklore seems to be that the last four classes are unjustified by the data since they have so few examples. There are 35 categorical attributes, some nominal and some ordered. The value "dna" means does not apply. The values for attributes are encoded numerically, with the first value encoded as "0," the second as "1," and so forth. The classes are: diaporthes-stem-canker,

charcoal-rot, rhizoctonia-root-rot, phytophthora-rot, brown-stem-rot, powdery-mildew, downy-mildew, brown-spot, bacterial-blight, bacterial-pustule, purple-seed-stain, anthracnose, phyllosticta-leaf-spot, alternarialeaf-spot, frog-eye-leaf-spot, diaporthe-pod-&-stem-blight, cyst-nematode, 2-4-d-injury, herbicide-injury.

o. Supplier

The supplier data set consists of 27 instances, each of which is described by 7 attributes.

p. Zoo

The zoo data set consists of 101 instances of animal, each of which is described by 17 attributes. Among these 17 attributes, the first attribute which denotes the name of animals does not participate in the clustering, 15 boolean attributes correspond to the presence of hair, feathers, eggs, milk, backbone, fins, tail; and whether airborne, aquatic, predator, toothed, breather, venomous, domestic, catsize, 1 non-boolean attribute corresponds to the number of legs in the set {0,2,4,6,8}. Each animal is classified into seven classes, namely, mammal, fish, bird, invertebrate, insect, amphibian, and reptile.

Various kinds of data are in the form of numerical data, characters, and more. These data were converted into data categories. Data categories include: 1, 2, 3, 4, and so on depending on the range of data. All the selected data sets are different from one another in terms of size, either horizontally or vertically aimed to analyze the performance of the proposed technique when involving a high number of records as well as the high number of attributes. Some datasets have been modified by removing instances that have incomplete data and removing an attribute that only have one categorical value. Summary information about seventeen sets of data, are shown in Table 4.1.

Table 4.1. The summary information about the seventeen data sets

No	Data Sets	Number of Objects	Number of Attributers	Size of Data
1	Acute Imflammation	120	6	720
2	Ballon	16	5	80
3	Bank Full	45,211	17	768,587
4	Car	1,728	7	12,096
5	Chees	3,196	36	115,056
6	Cilinder Band	520	14	7,280
7	Flag	64	3	192
8	Lenses	24	5	120
9	Lung Cancer	32	57	1,824
10	Monk	432	7	3,024
11	Mushroom	8,124	22	178,728
12	Nursery	12,960	9	116,640
13	Solar Flare	1,066	13	13,858
14	Soybean	47	35	1,645
15	Soybean Large	265	36	9,540
16	Supplier	27	7	189
17	Zoo	101	16	1,616

4.1.2. Language and Platform for Implementation

The two techniques MAR and MDDS are implemented using Matlab programming language version R2009a under Windows 7 Home Edition operating system powered by Intel i5 processor with 4 GB memory.

4.2. Performance Analysis

Validating the clustering results is heavy work. It needs to be measured against the accuracy results clustering in a certain way. The methods are used to evaluate the measures of the accuracy of clustering (Parmar, 2007), in addition to the accuracy, the

computing time is also very important as it relates to efficiency. The faster the results obtained, decision making can be done faster.

4.2.1. Execution Time

In this sub-section, the experimental results of the two techniques will be presented. The execution time in selecting a clustering attribute is presented in Table 4.2. The results of the comparison are made between the MDDS and MAR in terms of execution time. To calculate the increase in the relative improvement MAR and MDDS the following formula is used :

$$Impr(\%) = \frac{|MAR - MDDS|}{MAR} \times 100\% . \quad (4.1)$$

Table 4.2. Execution time of comparison results

No	Data Sets	MAR	MDDS	Improvement (%)
1	Acute Imflammation	0.958	0.394	58.87
2	Ballon	0.051	0.023	54.90
3	Bank Full	31.824	17.472	45.10
4	Car	0.327	0.246	24.77
5	Chees	4.763	2.274	52.26
6	Cilinder Band	1.560	0.780	50.00
7	Flag	0.020	0.015	25.00
8	Lenses	0.038	0.027	28.95
9	Lung Cancer	6.158	2.597	57.83
10	Monk	0.146	0.101	30.82
11	Mushroom	9.610	4.824	49.80
12	Nursery	1.552	0.877	43.49
13	Solar Flare	0.931	0.453	51.34
14	Soybean	1.477	0.596	59.65
15	Soybean Large	3.885	1.670	57.01
16	Supplier	0.173	0.056	67.63
17	Zoo	0.399	0.262	34.34
Average of overall improvement				43.99

Based on experiments on UCI datasets and other datasets, the proposed MDDS technique achieves better execution time which is lower than the MAR technique. Finally, the average improvement time of MDDS has increased to 43.99%. From Table 4.2, the fastest execution time on the flag dataset is 0.020 seconds at MAR and 0.015 seconds at MDDS with improvement of 25.00%. The Flag dataset has three attributes and 64 objects. The longest execution time on the bank full dataset is 31.824 seconds on the MAR and 17.472 seconds on the MDDS with improvement 45.10%. The bank full dataset has 17 attributes and 45,211 objects. The lowest Improvement is 24.77% on the car dataset with the seven attributes and 1,728 objects, where time of MAR is 0.327 seconds and time of MDDS 0.246 seconds. While the highest Improvement was 67.63% on the Supplier dataset with seven attributes and 27 objects, where the time of MAR is 0.173 seconds and time of MDDS is 0.056 seconds. The difference between the highest and lowest of improvement is 42.86%. While the average improvement is of 43.99% of 17 datasets.

4.2.2. Scalability

In this subsection, scalability is compared between MAR and MDDS techniques between seventeen datasets. The first is scalability on the number of instance, the second is on the scalability the number of attributes, while the third is scalability on the number of data size.

a. Scalability of MDDS to the number of instances

Figure 4.1 shows the scalability to the number of instants of using MAR and MDDS in seventeen datasets. It can be observed that the execution time of these techniques tends to increase. Hence, in general both techniques increases non-linearly as the number of instances in the seventeen datasets are increased.

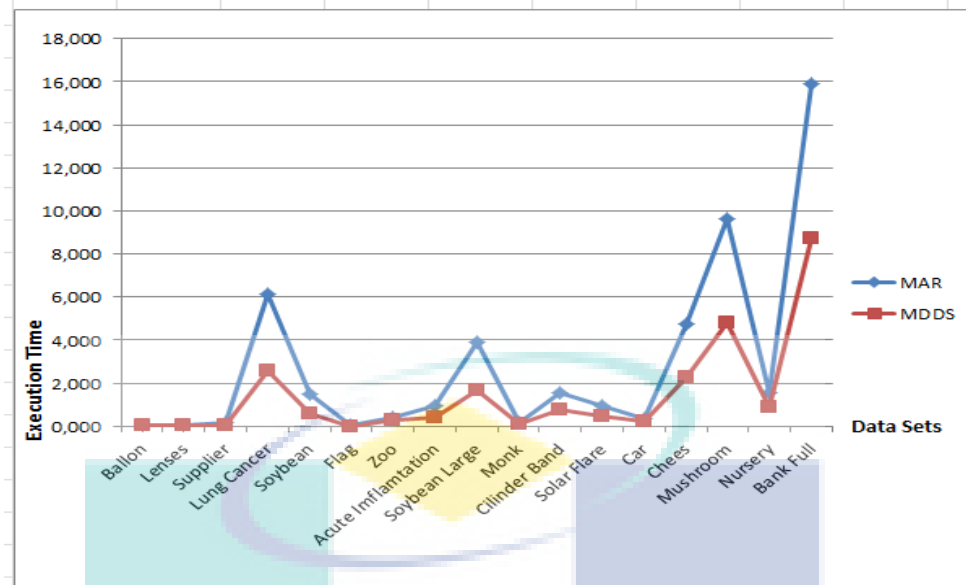


Figure 4.1. The scalability of MAR and MDDs to the number of instances

b. Scalability of MDDs to the number of attributes

Figure 4.2. shows scalability to the number of attributes of using MAR and MDDs in seventeen datasets. From Figure 4.2, it can be observed that the execution time of these techniques tends to increase the datasets with attribute data size.

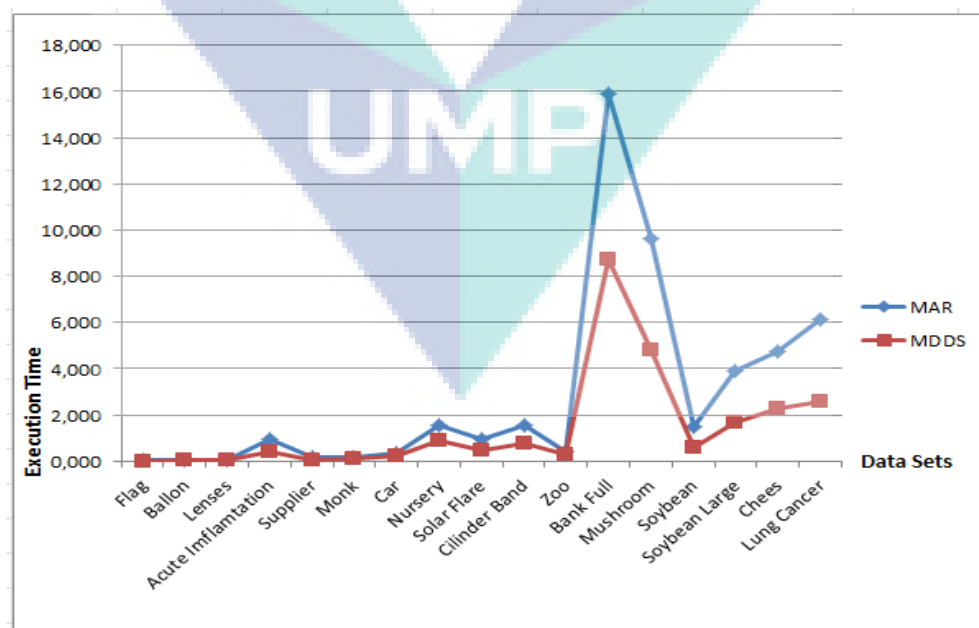


Figure 4.2. The scalability of MAR and MDDs to the number of attributes

c. Scalability of MDDS to the size of data

In reviewing the scalability of MAR and MDDS techniques on the seventeen data sets, the value varies in several numbers of instances and attributes. However, on the whole, these techniques have a good scalability to data size. The data size here is based-on the complexity of computing entries in data table. From the seventeen data sets in Table 4.1, Figures 4.3 described the scalability of these techniques to the data size. It is clearly depicted their good scalability to the data size.

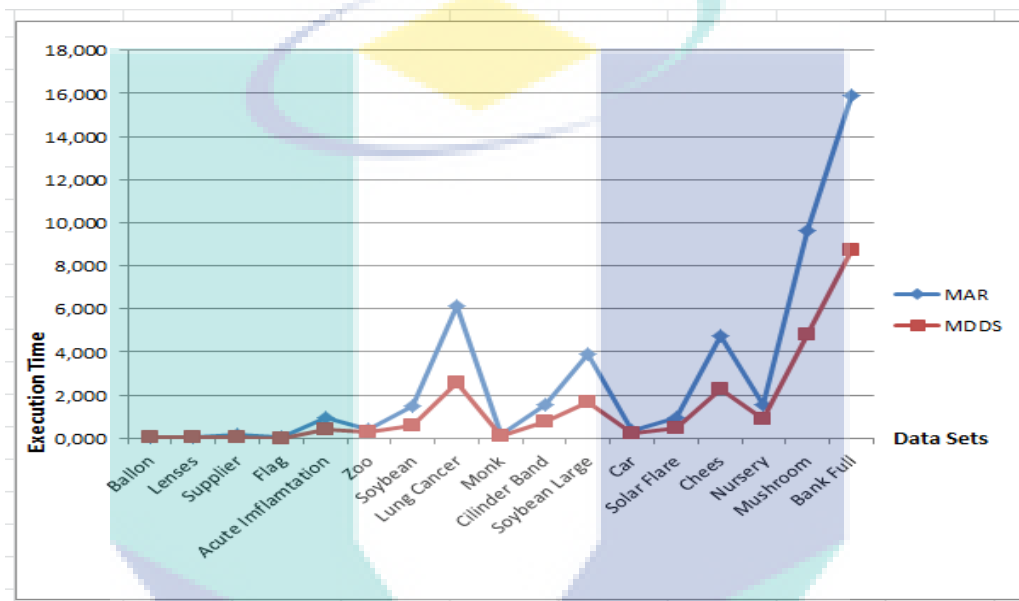


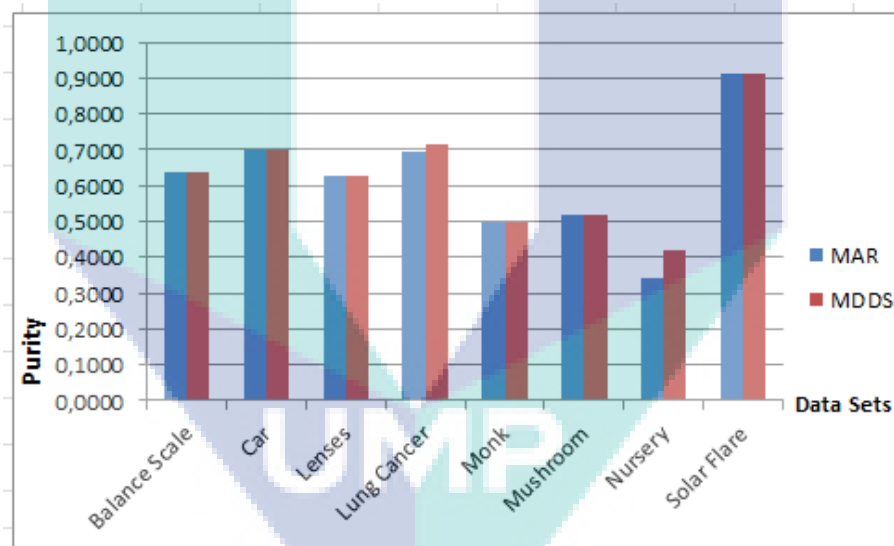
Figure 4.3. The scalability of MAR and MDDS to the number of data size

4.2.3. Accuracy

In addition to the execution time, to evaluate the results of clustering, is to look for its accuracy levels. Getting closer to the value 1, then it will be close to the actual conditions. So the result will be the better, (Parmar, 2007) to calculate the value of accuracy by following: For each(i), the number of data occurring in both the (i)th cluster and its corresponding class, then divided with the number of data in the data set. The overall amount is calculated from the accuracy of all accuracy and then it is divided by the number of classes. Calculation based on accuracy can be compared to the level of accuracy MAR and MDDS techniques. Results calculation of the accuracy of some data that has a class attribute is shown in table 4.3 and Figure 4.4

Table 4.3. Accuracy comparison results

No	Dataset	Attributes	Instances	Classes	Decision Attribute		Accuracy		Improve ment(%)
					MAR	MDDS	MAR	MDDS	
1	Balance Scale	4	625	3	1	1	0.64	0.64	0.00
2	Car	6	1,728	4	4	1	0.70	0.70	0.00
3	Lenses	4	24	3	2	1	0.63	0.63	0.00
4	Lung Cancer	56	32	3	1	47	0.69	0.72	3.35
5	Monk	6	432	2	3	1	0.50	0.50	0.00
6	Mushroom	21	8,124	2	15	15	0.52	0.52	0.00
7	Nursery	8	12,960	5	6	1	0.34	0.42	22.51
8	Solar Flare	10	1,066	3	10	10	0.91	0.91	0.00
Average of overall improvement									3.23

**Figure 4.4.** The accuracy of MAR and MDDS

4.2.4. Iterations

The technique of MDDS has proposed number of iterations that is lower than the technique of MAR. The average number of iterations in the MDDS is reduced to 15.26%. From Table 4.4, the fewest number of iterations on the balloon dataset is 72 in MAR and is 56 in MDDS, so reduced to 22.22%. The number of iterations at most on a

data set that is band Cilinder 55,389 in MAR and 42,222 in MDDS is, so reduced to 23.77%. A reduction in the number of iterations most 65.86% in Acute Inflammation dataset, whereas there is a reduction in the lowest amount of 1.80% in lung cancer data set. Figure 4.5 shows a comparison of the two methods the number of iterations.

Table 4.4. Number of Iterations comparison results

No	Data Sets	Number of Iterations		Reduction (%)
		MAR	MDDS	
1	Ballon	72	56	22.22
2	Lenses	90	69	23.33
3	Supplier	342	288	15.79
4	Monk	380	325	14.47
5	Car	462	387	16.23
6	Nursery	992	877	11.59
7	Acute Imflamtation	2,970	1,014	65.86
8	Zoo	1,332	1,236	7.21
9	Solar Flare	2,450	2,215	9.59
10	Soybean	5,256	5,050	3.92
11	Chees	5,402	5,253	2.76
12	Bank Full	11,990	11,037	7.95
13	Soybean Large	12,656	12,126	4.19
14	Mushroom	13,110	12,298	6.19
15	Lung Cancer	26,082	25,613	1.80
16	Flag	32,137	24,901	22.52
17	Cilinder Band	55,389	42,222	23.77
Average of overall improvement				15.26

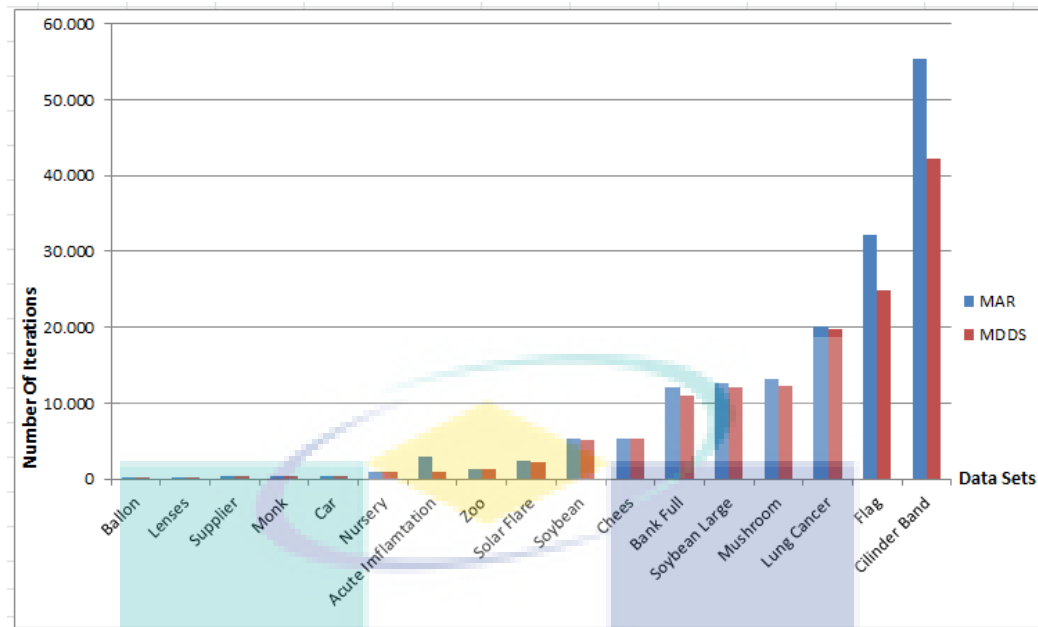


Figure 4.5. Comparison of iterations number MAR and MDDS

4.3. Comparison between MAR and MDDS Techniques

MDDS technique has successfully improved the MAR technique in terms of efficiency and accuracy. In this experiment, the time and the number of iterations is used as a criterion to measure efficiency. The lower the number of iterations, the less execution time is required, so that efficiency will be increased. Table 4.2-4.4 shows the execution time comparisons, clustering, and accuracy results of the number of iterations of the algorithm. While the Figure 4.1-4.5 show performance comparison between the MAR and MDDS techniques.

Data set are used in this research, it has shown that MDDS technique takes time and the number of iterations less. While the accuracy of the grouping of eight datasets that have class attribute has also increased. This can happen because, the MAR technique uses complex mathematical models, because this technique should calculate the value of support, max_support and min_support. Support value = 1 is summed as max_support, while others as min_support. Clustering attribute is determined based on the largest amount of max_support. If there are greater max_support of more than 1,

then the clustering attribute is an attribute that has the largest value of $\min_support$. The MDDS technique is simpler and just need three steps, namely :

- a. Calculate Domination of Attributes a_i , with respect to all a_j , where $i \neq j$
- b. Select the maximum of domination degree of each attributes
- c. Select the clustering attribute based on the maximum degree of domination of attributes.

Data clustering under soft set theory can be considered as a technique for data mining. Maximum Degree of Domination in Soft set theory is applied to select the clustering attributes. To find meaningful clusters from a dataset, a clustering attribute is conducted so that attributes within the clusters made will have a high correlation or highly interdependent to each other while the attributes in other clusters are less correlated or more independent. In the experimental results, it is seen how to determine the dominant attributes of a set of attributes of an assessment list by using the soft set theory.

4.4. Summary

A series of experiments were conducted to evaluate the clustering performance, clustering efficiency and scalability of MAR and MDDS algorithms. The experimental result show that MDDS achieves better clustering accuracy and stability than MAR algorithm, at the same time increases the efficiency. MDDS has obvious advantage against MAR on large data sets in terms of clustering efficiency as well as clustering accuracy. In addition, The MDDS technique has better scalability. It can be applied on small categorical data sets as well as large categorical data sets.

CHAPTER 5

APPLICATIONS OF MDDS TECHNIQUE IN EDUCATIONAL DATA MINING

5.1. Data Sets of Assessment

This section explains and discusses the experimental results of the proposed technique. The main focus of the experiments is on the performance measurement of the proposed technique in which execution time is used as a parameter. The Data is taken from the evaluation of Information Engineering and Architecture Department. The data is shown in Table 5.1.

Table 5.1. Data sets of Assessment

No	Courses	Object	Attributes
1	Algorithm	99	5
2	Software Engineering	260	6
3	System Security	269	6
4	File System	190	5
5	Artificial Intelligence	173	5
6	Database	88	5
7	Architecture Design	94	10
8	Infrastructure Design Studio	34	29
9	Architecture Studio	181	7
10	Contextual Studio	34	30

All the data we taken from University Technology of Yogyakarta (UTY) Indonesia for four years with periods from 2011 to 2014. The assessment consists of several attributes which were different. Each of the majors and courses does not have the same assessment criteria, all of it are in the form of assignments, midterm and final

exams. Midterms be done in the middle of the semester and is done in writing. The final exams are given at the end of the semester. Both are done on a scheduled basis. Students' name, age, race, and finance were not necessary in this assessment.

Evaluation criteria used ranges from 0 to 100. Nevertheless the data transformation is in the data category as shown in the table 5.2. From these data later in the process, it is to give weight or a certain percentage to get the final mark in the form of grade A for the highest value of the grade of E to the lowest mark. The proposed MDDS technique is used to determine the most dominant attribute and further to clustering the students.

Table 5.2. Transformation data assessment into categorical data

No	Data Assessment	Category
1	0 – 20	1
2	21 – 40	2
3	41 – 60	3
4	61 – 80	4
5	81 - 100	5

5.2. Data Descriptions

a. Algorithm Course

Assessment of Algorithm course has six attributes, namely task1 to task4, mid-term, and final exam as shown in Table 5.3.

Table 5.3. Assessment of Algorithm course

Attr	Description	2011					2012					2013				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	5	0	7	8	10	0	0	1	21	6	1	0	5	23	12
T2	Task 2	0	0	5	25	0	0	1	3	18	6	5	0	9	24	3
T3	Task 3	7	0	0	18	5	0	0	6	20	2	9	0	1	17	14
T4	Task 4	6	0	0	17	7	0	0	4	22	2	12	0	1	24	4
MT	Midterm	1	0	1	22	6	0	0	4	17	7	10	0	2	20	9
FE	Final Exam	2	2	3	7	16	0	0	3	18	7	1	0	1	12	27

Table 5.3 is a recapitulation of the Algorithm course assessment. The number of students who achieved mark #1 for task 1 (T1) in 2011 is 5 students. The number of students who achieved mark #3 for task 3 (T3) in 2012 is 6 students. While the number of students who achieved mark #5 for final exam (FE) in 2013 is 27 students.

b. Software Engineering Course

Assessment of Software Engineering course has six attributes, namely task1 to task4, mid-term, and final exams shown in Table 5.4.

Table 5.4. Assessment of Software Engineering course

Attr	Description	2011					2012					2013				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	6	1	20	48	12	51	0	4	14	18	25	3	18	25	15
T2	Task 2	3	0	8	58	18	1	0	9	59	18	3	0	5	58	20
T3	Task 3	6	0	37	44	0	5	0	31	51	0	13	0	36	37	0
T4	Task 4	0	4	3	22	58	0	1	9	27	50	1	2	5	25	53
MT	Midterm	5	17	49	14	2	0	16	55	13	3	0	24	48	10	4
FE	Final Exam	5	18	30	25	9	4	5	43	22	13	4	11	37	29	5

Table 5.4 is a recapitulation of the Software Engineering course assessment. The number of students who achieved mark #1 for task 1 (T1) in 2011 is 6 students. The number of students who achieved mark #3 for task 3 (T3) in 2012 is 31 students. While the number of students who achieved mark #5 for final exam (FE) in 2013 is 5 students.

c. System Security Course

Assessment of System Security course has six attributes, namely task1 to task4, mid-term, and final exam as shown in Table 5.5

Table 5.5. Assessment of System Security course

Attr	Description	2011					2012					2013				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	1	1	24	64	0	0	0	8	82	0	0	0	15	74	0
T2	Task 2	0	2	10	77	1	0	0	2	88	0	0	0	9	80	0
T3	Task 3	1	2	23	64	0	0	1	10	79	0	0	1	13	75	0
T4	Task 4	1	5	12	45	27	0	0	2	8	80	0	1	5	20	63
MT	Midterm	10	24	32	21	3	4	11	32	34	9	6	22	31	26	4
FE	Final Exam	5	3	26	46	10	0	0	11	68	11	2	0	22	56	9

Table 5.5 is a recapitulation of the System Security course assessment. The number of students who achieved mark #1 for task 1 (T1) in 2011 is 1 students. The number of students who achieved mark #3 for task 3 (T3) in 2012 is 10 students. While the number of students who achieved mark #5 for final exam (FE) in 2013 is 9 students.

d. File System Course

Assessment of File System course has five attributes, namely task1 to task 3, mid-term, and final exam as shown in Table 5.6

Table 5.6.Assessment of File System course

Attr	Description	2011					2012					2013				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	3	4	15	12	28	5	2	15	8	37	0	2	2	2	55
T2	Task 2	4	0	4	54	0	3	0	7	57	0	0	0	0	61	0
T3	Task 3	1	0	2	49	10	2	0	1	62	2	0	0	0	52	9
MT	Midterm	2	21	22	13	4	5	13	29	15	5	4	7	21	24	5
FE	Final Exam	3	7	19	21	12	2	6	32	18	9	3	4	12	20	22

Table 5.6 is a recapitulation of the File System course assessment. The number of students who achieved mark #1 for task 1 (T1) in 2011 is 3 students. The number of students who achieved mark #3 for task 3 (T3) in 2012 is 1 student. While the number of students who achieved mark #5 for final exam (FE) in 2013 is 22 students.

e. Artificial Intelligence Course

Assessment of Artificial Intelligence course taken in 2014 has five attributes, namely task1 to task 3, mid-term, and final exam as shown in Table 5.7.

Table 5.7. Assessment of Artificial Intelligence course

Attr	Description	#1	#2	#3	#4	#5
T1	Task 1	10	45	98	19	1
T2	Task 2	45	34	63	9	22
T3	Task 3	26	1	2	0	144
MT	Midterm	13	1	0	0	159
FE	Final Exam	14	0	7	135	17

Table 5.7 is a recapitulation of the Artificial Intelligence course assessment. The number of students who achieved mark #1 for task 1 (T1) is 10 students. The number of students who achieved mark #3 for task 3 (T3) is 2 students. While the number of students who achieved mark #5 for final exam (FE) in 2013 is 17 students.

f. Databases Course

Assessment of Database course taken in 2014 has five attributes, namely task1 to task 3, mid-term, and final exam as shown in Table 5.8.

Table 5.8. Assessment of Databases course

Attr	Description	#1	#2	#3	#4	#5
T1	Task 1	8	42	30	4	4
T2	Task 2	5	25	27	19	12
T3	Task 3	6	1	10	65	6
MT	Midterm	2	2	21	57	6
FE	Final Exam	7	11	29	39	2

Table 5.8 is recapitulation of the Databases course assessment. The number of students who achieved mark #1 for task 1 (T1) is 8 students. The number of students who achieved mark #3 for task 3 (T3) is 10 students. While the number of students who achieved mark #5 for final exam (FE) is 2 students.

g. Architectural Design Course

Assessment of Architectural Design course has ten attributes, namely task1 to task8, mid-term, and final exam as shown in Table 5.9.

Table 5.9. Assessment of Architectural Design course

Attr	Description	2011					2012					2013				
		#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	9	0	2	23	0	2	9	10	12	0	5	0	22	0	0
T2	Task 2	8	0	7	19	0	3	2	16	12	0	13	0	14	0	0
T3	Task 3	9	11	9	5	0	5	0	5	23	0	9	0	18	0	0
T4	Task 4	13	3	15	3	0	6	7	7	13	0	9	0	18	0	0
T5	Task 5	9	4	5	14	2	4	2	6	21	0	14	0	13	0	0
T6	Task 6	19	4	6	5	0	4	9	11	9	0	15	0	12	0	0
T7	Task 7	15	0	11	8	0	9	0	24	0	0	15	0	12	0	0
T8	Task 8	16	0	3	15	0	14	0	19	0	0	3	10	7	6	1
MT	Midterm	3	2	20	9	0	14	0	19	0	0	6	1	4	11	5
FE	Final Exam	3	4	22	5	0	5	0	28	0	0	9	0	0	18	0

Table 5.9 is a recapitulation of the Architectural Design course assessment. The number of students who achieved mark #1 for task 1 (T1) in 2011 is 9 students. The number of students who achieved mark #3 for task 5 (T5) in 2012 is 6 students. While the number of students who achieved mark #5 for final exam (FE) in 2013 is not student.

h. Infrastructure Design Studio Course

Assessment of Architectural Design course has ten attributes, namely task1 to task27, mid-term, and final exam as shown in Table 5.10.

Table 5.10. Assessment of Infrastructure Design Studio course

Attr	Description	#1	#2	#3	#4	#5
T1	Task 1	5	2	7	16	4
T2	Task 2	6	3	7	16	2
T3	Task 3	5	0	7	19	3
T4	Task 4	7	1	5	17	4
T5	Task 5	6	2	7	18	1
T6	Task 6	9	4	5	13	3
T7	Task 7	5	1	2	19	7
T8	Task 8	6	5	2	17	4
T9	Task 9	6	1	7	14	6
T10	Task 10	7	1	5	11	10
T11	Task 11	10	2	1	16	5
T12	Task 12	9	4	5	12	4
T13	Task 13	10	3	2	12	7
T14	Task 14	8	2	6	13	5
T15	Task 15	8	5	5	14	2
T16	Task 16	13	4	1	10	6
T17	Task 17	14	3	4	9	4
T18	Task 18	8	2	3	11	10
T19	Task 19	9	1	4	11	9
T20	Task 20	9	2	2	12	9
T21	Task 21	11	3	8	6	6
T22	Task 22	8	0	6	15	5
T23	Task 23	8	0	5	15	6
T24	Task 24	19	1	4	7	3
T25	Task 25	12	0	8	10	4
T26	Task 26	8	3	8	9	6
T27	Task 27	11	1	5	10	7
MT	Midterm	8	2	5	14	5
FE	Final Exam	11	2	4	4	13

Table 5.10 is a recapitulation of the Infrastructure Design Studio course assessment. The number of students who achieved mark #1 for task 1 (T1) is 5 students. The number of students who achieved mark #3 for task 10 (T10) is 5 students. While the number of students who achieved mark #5 for final exam (FE) is 13 students.

i. Architecture Studio Course

Assessment of Architectural Studio course has ten attributes, namely task1 to task5, mid-term, and final exam as shown in Table 5.11.

Table 5.11. Assessment of Architecture Studio course

Attr	Description	#1	#2	#3	#4	#5
T1	Task 1	8	67	68	35	3
T2	Task 2	20	62	72	23	4
T3	Task 3	4	63	93	18	3
T4	Task 4	2	63	88	25	3
T5	Task 5	6	64	68	38	5
MT	Midterm	24	61	61	25	10
FE	Final Exam	16	66	65	20	14

Table 5.11 is a recapitulation of the Architecture Studio course assessment. The number of students who achieved mark #1 for task 1 (T1) is 8 students. The number of students who achieved mark #3 for task 4 (T4) is 88 students. While the number of students who achieved mark #5 for final exam (FE) is 14 students.

j. Contextual Studio Course

Assessment of Contextual Studio course has ten attributes, namely task1 to task28, mid-term, and final exam as shown in Table 5.12.

Table 5.12. Assessment of Contextual Studio course

Attr	Description	#1	#2	#3	#4	#5
T1	Task 1	4	2	1	23	4
T2	Task 2	5	0	7	20	2
T3	Task 3	7	3	9	13	2
T4	Task 4	5	0	6	20	3
T5	Task 5	7	2	5	17	3
T6	Task 6	5	1	6	20	2
T7	Task 7	9	5	5	10	5
T8	Task 8	6	4	2	19	3
T9	Task 9	6	1	7	15	5
T10	Task 10	6	2	4	14	8
T11	Task 11	10	1	2	12	9
T12	Task 12	9	4	3	13	5
T13	Task 13	7	1	5	12	9
T14	Task 14	9	4	3	14	4
T15	Task 15	8	3	6	16	1
T16	Task 16	10	3	3	11	7
T17	Task 17	16	5	4	7	2
T18	Task 18	9	3	1	11	10
T19	Task 19	9	1	5	10	9
T20	Task 20	8	2	1	15	8
T21	Task 21	11	4	5	5	9
T22	Task 22	11	0	8	12	3
T23	Task 23	9	0	5	15	5
T24	Task 24	10	0	6	12	6
T25	Task 25	12	0	8	10	4
T26	Task 26	8	3	8	9	6
T27	Task 27	11	1	5	10	7
T28	Task 28	22	1	2	4	5
MT	Midterm	11	3	6	11	3
FE	Final Exam	7	1	6	6	34

Table 5.12 is recapitulation of the Contextual Studio course assessment. The number of students who achieved mark #1 for task 1 (T1) is 4 students. The number of students who achieved mark #3 for task 15 (T15) is 6 students. While the number of students who achieved mark #5 for final exam (FE) is 34 students.

5.3. Result and Discussion

The user interface of the MDDS software is shown in Figure 5.1. The data used can be benchmark data categorical valued. In addition to the benchmark, the data is available on this software, the data can also be used with other categories as input. Software will provide the results of the calculation of the domination degree in the form of a matrix and clustering visualization. Another output is selected for attributes and execution time.

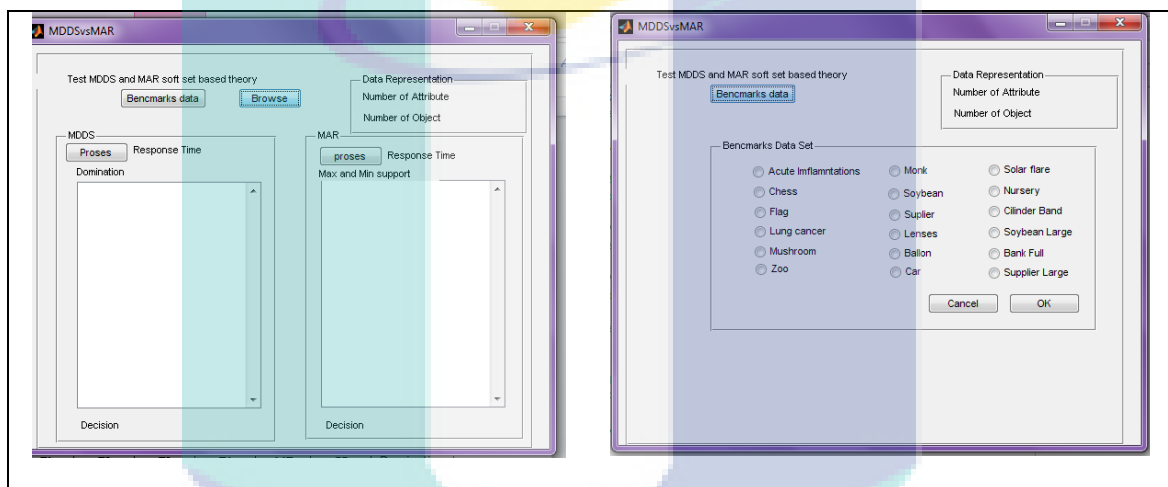


Figure 5.1. User interface of the developed MDDS software

Table 5.13. Matrix results from Algorithm course 2011

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.00	0.00	0.00	0.07	0.07	0.07
T2	0.60	0.00	0.17	0.23	0.27	0.70	0.70
T3	0.00	0.00	0.00	0.00	0.07	0.13	0.13
T4	0.00	0.00	0.00	0.00	0.07	0.13	0.13
MT	0.17	0.00	0.00	0.00	0.00	0.17	0.17
FE	0.00	0.00	0.00	0.00	0.07	0.00	0.07

Table 5.13 is a matrix that indicates the degree of dominance attribute value of Algorithm course in 2011. MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 0.7. Hence T2

is selected as a clustering attribute, whereas visualization is divided in two clusters as shown in Figure 5.2.

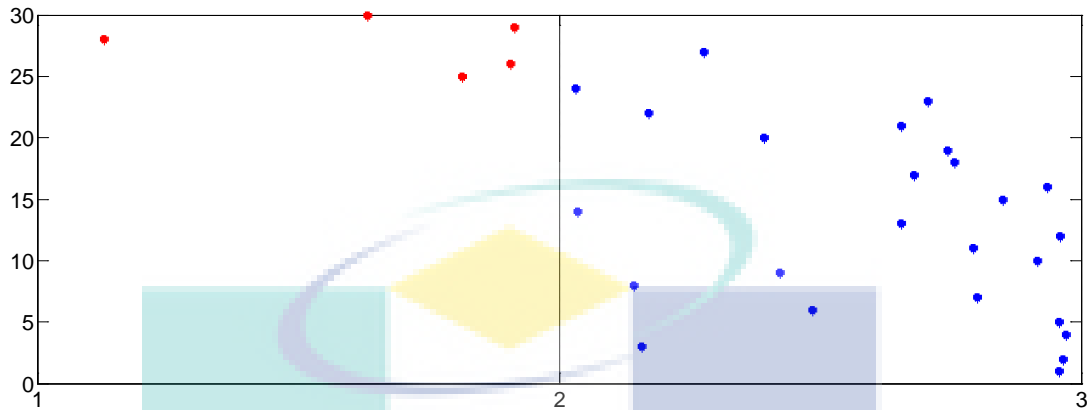


Figure 5.2. Clustering visualization of students on Algorithm course 2011

Table 5.14. Matrix results from Algorithm course 2012

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.04	0.00	0.07	0.00	0.11	0.11
T2	0.04	0.00	0.07	0.07	0.00	0.11	0.11
T3	0.04	0.14	0.00	0.07	0.00	0.00	0.14
T4	0.04	0.25	0.07	0.00	0.00	0.11	0.25
MT	0.04	0.14	0.00	0.00	0.00	0.11	0.14
FE	0.04	0.14	0.00	0.14	0.00	0.00	0.14

Table 5.14 is a matrix that indicates the degree of dominance attribute value of Algorithm course in 2012. MDDS technique provides a T4 as the most dominant attribute compared to other attributes, where the maximum domination is 0.25. Hence T4 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.3.

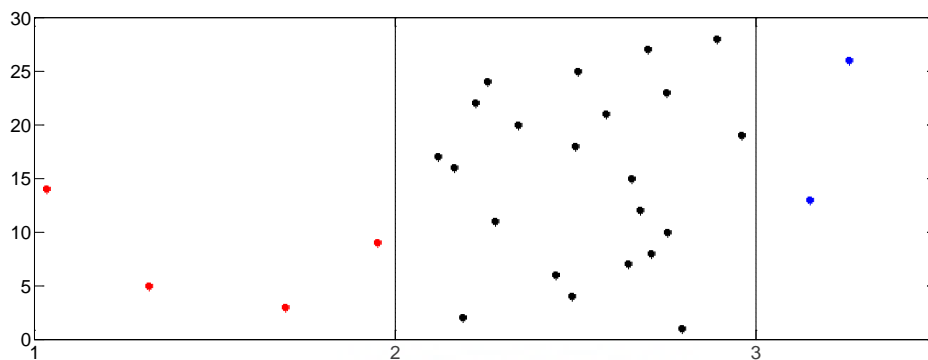


Figure 5.3. Clustering visualization of students on Algorithm course 2012

5.15. Matrix results from Algorithm course 2013

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.12	0.02	0.02	0.05	0.05	0.12
T2	0.02	0.00	0.02	0.02	0.00	0.05	0.05
T3	0.02	0.00	0.00	0.02	0.05	0.05	0.05
T4	0.02	0.00	0.02	0.00	0.24	0.05	0.24
MT	0.02	0.00	0.02	0.12	0.00	0.05	0.12
FE	0.02	0.07	0.02	0.12	0.00	0.00	0.12

Table 5.15 is a matrix that indicates the degree of dominance attribute value of Algorithm course in 2013. MDDS technique provides a T4 as the most dominant attribute compared to other attributes, where the maximum domination is 0.24. Hence T4 is selected as a clustering attribute, whereas visualization is divided in four clusters as shown in Figure 5.4.

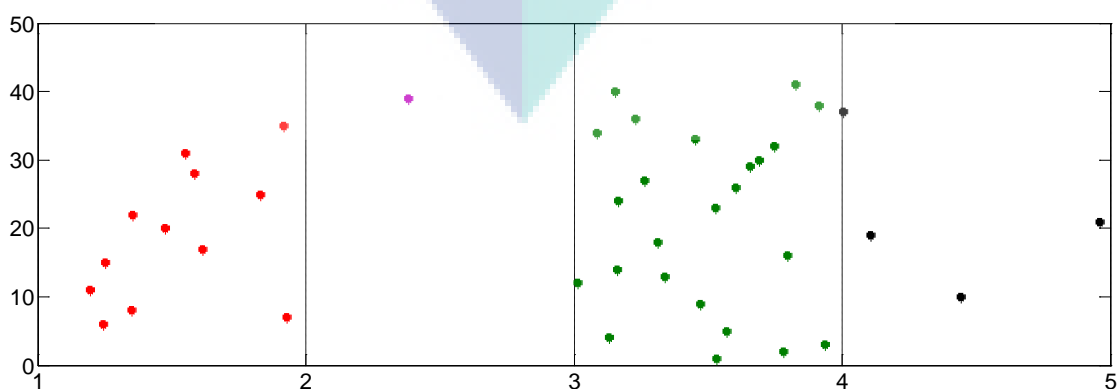
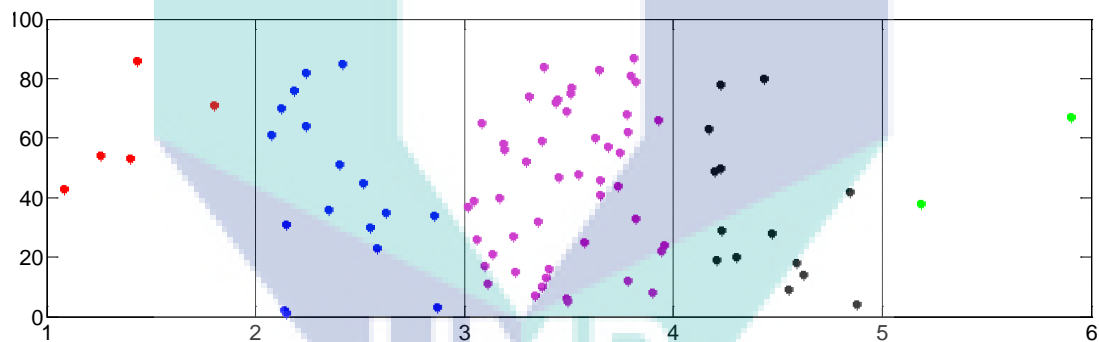


Figure 5.4. Clustering visualization of students on Algorithm course 2013

Table 5.16. Matrix results from Software Engineering course 2011

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T2	0.01	0.00	0.00	0.00	0.00	0.00	0.01
T3	0.01	0.00	0.00	0.00	0.00	0.00	0.01
T4	0.01	0.00	0.00	0.00	0.02	0.00	0.02
MT	0.01	0.03	0.00	0.03	0.00	0.00	0.03
FE	0.01	0.00	0.00	0.00	0.02	0.00	0.02

Table 5.16 is a matrix that indicates the degree of dominance attribute value of Software Engineering course in 2011. MDDS technique provides a MT as the most dominant attribute compared to other attributes, where the maximum domination is 0.03. Hence MT is selected as a clustering attribute, whereas visualization is divided in five clusters as shown in Figure 5.5.

**Figure 5.5.** Clustering visualization of students on Software Engineering course 2011**Table 5.17.** Matrix results from Software Engineering course 2012

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.01	0.06	0.01	0.00	0.00	0.06
T2	0.00	0.00	0.00	0.01	0.00	0.06	0.06
T3	0.00	0.01	0.00	0.01	0.03	0.00	0.03
T4	0.00	0.01	0.00	0.00	0.00	0.00	0.01
MT	0.00	0.01	0.06	0.01	0.00	0.00	0.06
FE	0.00	0.01	0.00	0.01	0.00	0.00	0.01

Table 5.17 is a matrix that indicates the degree of dominance attribute value of Software Engineering course in 2012. MDDS technique provides a T1 as the most dominant attribute compared to other attributes, where the maximum domination is 0.06. Hence T1 is selected as a clustering attribute, whereas visualization is divided in four clusters as shown in Figure 5.6.

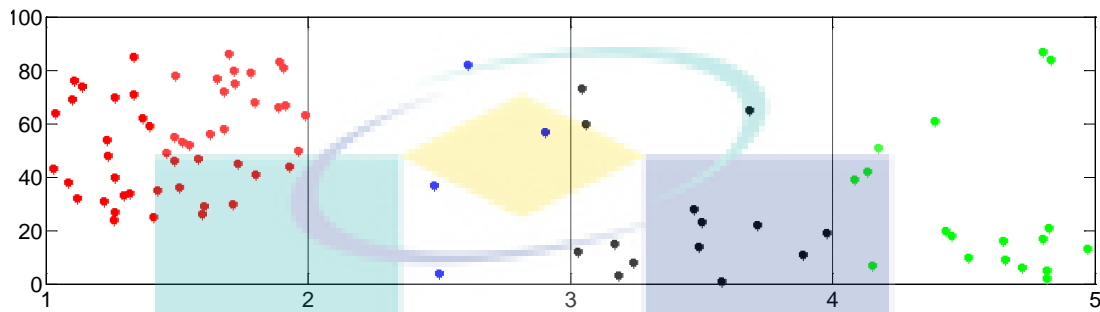


Figure 5.6. Clustering visualization of students on Software Engineering course 2012

Table 5.18. Matrix results from Software Engineering course 2013

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.00	0.00	0.01	0.00	0.00	0.01
T2	0.00	0.00	0.00	0.07	0.00	0.00	0.07
T3	0.00	0.06	0.00	0.03	0.00	0.00	0.06
T4	0.00	0.00	0.00	0.00	0.16	0.00	0.16
MT	0.00	0.00	0.00	0.09	0.00	0.00	0.09
FE	0.00	0.00	0.00	0.03	0.00	0.00	0.03

Table 5.18 is a matrix that indicates the degree of dominance attribute value of Software Engineering course in 2013. MDDS technique provides a T1 as the most dominant attribute compared to other attributes, where the maximum domination is 0.16. Hence T4 is selected as a clustering attribute, whereas visualization is divided in five clusters as shown in Figure 5.7.

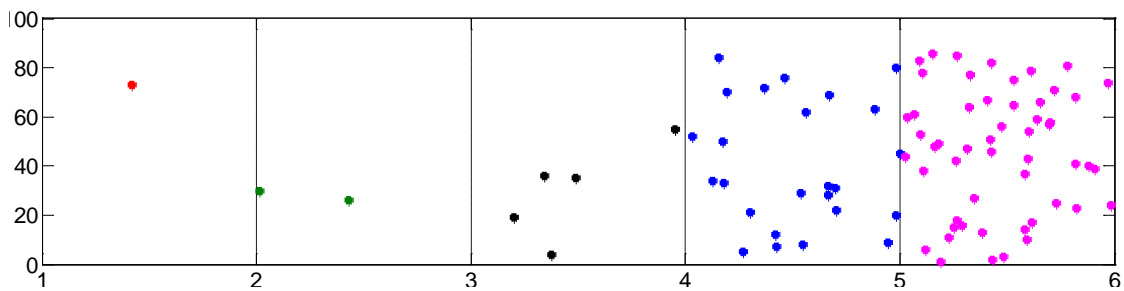


Figure 5.7. Clustering visualization of students on Software Engineering course 2013

Table 5.19. Matrix results from System Security course 2011

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.12	0.01	0.44	0.03	0.14	0.44
T2	0.02	0.00	0.01	0.51	0.03	0.00	0.51
T3	0.02	0.01	0.00	0.31	0.03	0.11	0.31
T4	0.02	0.01	0.03	0.00	0.00	0.00	0.03
MT	0.02	0.01	0.03	0.01	0.00	0.00	0.03
FE	0.02	0.01	0.01	0.01	0.03	0.00	0.03

Table 5.19 is a matrix that indicates the degree of dominance attribute value of System Security course in 2011. The MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 0.51. Hence T2 is selected as a clustering attribute, whereas visualization is divided in four clusters as shown in Figure 5.8.

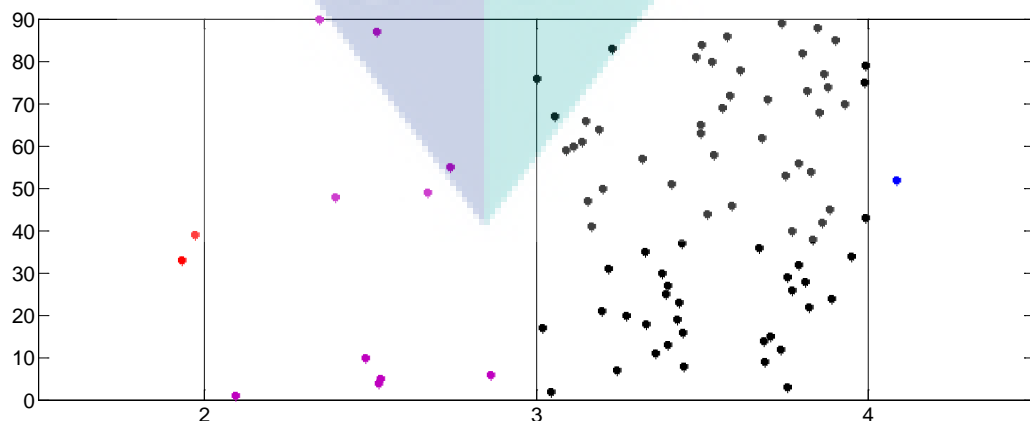
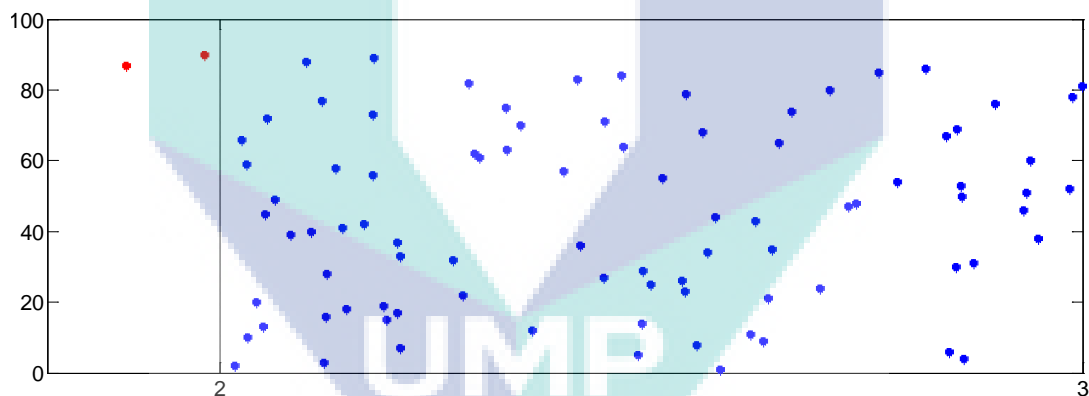


Figure 5.8. Clustering visualization of students on System Security course 2011

Table 5.20. Matrix results from System Security course 2012

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.02	0.89	0.91	0.10	0.12	0.91
T2	0.91	0.00	0.89	1.00	0.27	0.24	1.00
T3	0.00	0.00	0.00	0.09	0.10	0.12	0.12
T4	0.00	0.02	0.89	0.00	0.10	0.12	0.89
MT	0.00	0.00	0.01	0.00	0.00	0.00	0.01
FE	0.00	0.02	0.01	0.02	0.10	0.00	0.10

Table 5.20 is a matrix that indicates the degree of dominance attribute value of System Security course in 2012. The MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 1.00. Hence T2 is selected as a clustering attribute, whereas visualization is divided in two clusters as shown in Figure 5.9.

**Figure 5.9.** Clustering visualization of students on System Security course 2012**Table 5.21.** Matrix results from System Security course 2013

Attribute (with respect to)	Domination Degree						Maximum Domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.10	0.01	0.78	0.04	0.12	0.78
T2	0.83	0.00	0.85	0.78	0.04	0.12	0.85
T3	0.00	0.00	0.00	0.72	0.04	0.10	0.72
T4	0.00	0.00	0.01	0.00	0.04	0.10	0.10
MT	0.00	0.00	0.01	0.01	0.00	0.00	0.01
FE	0.00	0.00	0.01	0.01	0.04	0.00	0.04

Table 5.21 is a matrix that indicates the degree of dominance attribute value of System Security course in 2013. The MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 0.85. Hence T2 is selected as a clustering attribute, whereas visualization is divided in two clusters as shown in Figure 5.10.

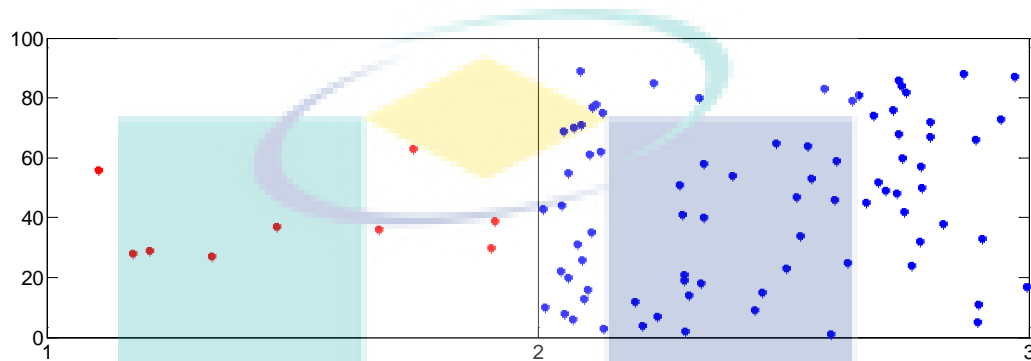


Figure 5.10. Clustering visualization of students on System Security course 2013

Table 5.22. Matrix results from File System course 2011

Attribute (with respect to)	Domination Degree					Maximum Domination
	T1	T2	T3	MT	FE	
T1	0.00	0.06	0.02	0.06	0.00	0.06
T2	0.76	0.00	0.05	0.10	0.19	0.76
T3	0.06	0.00	0.00	0.03	0.00	0.06
MT	0.00	0.00	0.05	0.00	0.00	0.05
FE	0.00	0.00	0.05	0.00	0.00	0.05

Table 5.22 is a matrix that indicates the degree of dominance attribute value of File System course in 2011. The MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 0.76. Hence T2 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.11.

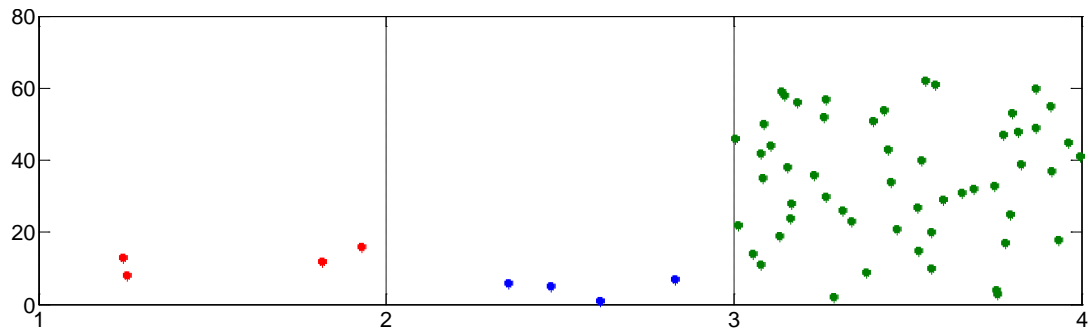


Figure 5.11. Clustering visualization of students on File System course 2011

Table 5.23. Matrix results from File System course 2012

Attribute (with respect to)	Domination Degree					Maximum Domination
	T1	T2	T3	MT	FE	
T1	0.00	0.10	0.04	0.07	0.00	0.10
T2	0.15	0.00	0.04	0.07	0.36	0.36
T3	0.22	0.00	0.00	0.27	0.22	0.27
MT	0.00	0.00	0.04	0.00	0.00	0.04
FE	0.03	0.10	0.01	0.00	0.00	0.10

Table 5.23 is a matrix that indicates the degree of dominance attribute value of File System course in 2012. The MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 0.36. Hence T2 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.12.

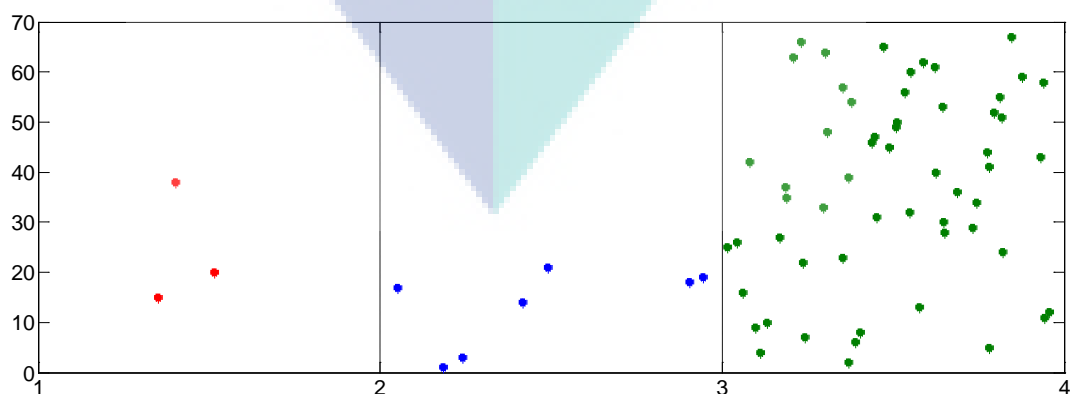
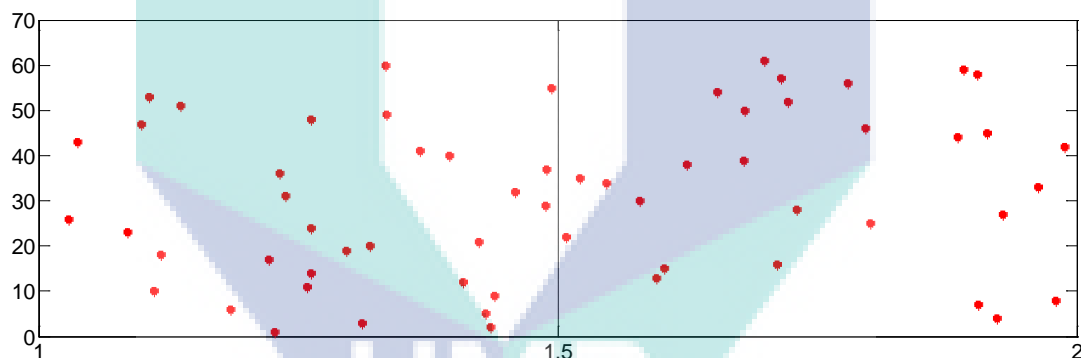


Figure 5.12. Clustering visualization of students on File System course 2012

Table 5.24. Matrix results from File System course 2013

Attribute (with respect to)	Domination Degree					Maximum Domination
	T1	T2	T3	MT	FE	
T1	0.00	0.00	0.00	0.26	0.48	0.48
T2	1.00	0.00	1.00	1.00	1.00	1.00
T3	0.07	0.00	0.00	0.18	0.31	0.31
MT	0.03	0.00	0.00	0.00	0.00	0.03
FE	0.07	0.00	0.00	0.00	0.00	0.07

Table 5.24 is a matrix that indicates the degree of dominance attribute value of File System course in 2013. The MDDS technique provides a T2 as the most dominant attribute compared to other attributes, where the maximum domination is 1.00. Hence T2 is selected as a clustering attribute, whereas visualization is divided in a clusters as shown in Figure 5.13.

**Figure 5.13.** Clustering visualization of students on File System course 2013**Table 5.25.** Matrix results from Artificial Intelligence course 2014

Attribute (with respect to)	Domination Degree					Maximum Domination
	T1	T2	T3	MT	FE	
T1	0.00	0.00	0.01	0.01	0.00	0.01
T2	0.01	0.00	0.02	0.01	0.00	0.02
T3	0.01	0.05	0.00	0.00	0.00	0.05
MT	0.01	0.38	0.02	0.00	0.00	0.38
FE	0.01	0.00	0.01	0.01	0.00	0.01

Table 5.25 is a matrix that indicates the degree of dominance attribute value of Artificial Intelligence course in 2014. The MDDS technique provides a MT as the most dominant attribute compared to other attributes, where the maximum domination is 0.38. Hence MT is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.14.

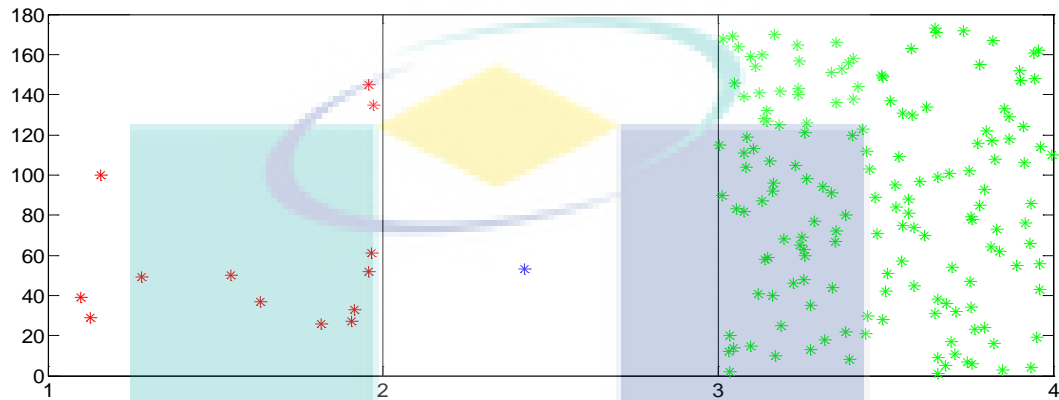


Figure 5.14. Clustering visualization of students on Artificial Intelligence course 2014

Table 5.26. Matrix results from Databases course 2014

Attribute (with respect to)	Domination Degree					Maximum Domination
	T1	T2	T3	MT	FE	
T1	0.00	0.00	0.08	0.02	0.00	0.08
T2	0.00	0.00	0.01	0.00	0.00	0.01
T3	0.10	0.00	0.00	0.02	0.02	0.10
MT	0.10	0.00	0.01	0.00	0.02	0.10
FE	0.00	0.00	0.01	0.00	0.00	0.01

Table 5.26 is a matrix that indicates the degree of dominance attribute value of Databases course in 2014. The MDDS technique provides a T3 as the most dominant attribute compared to other attributes, where the maximum domination is 0.1. Hence T3 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.15.

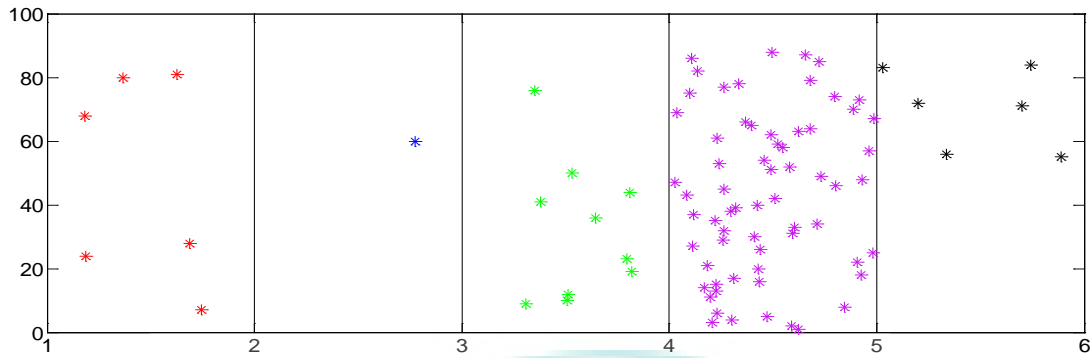


Figure 5.15. Clustering visualization of students on Databases course 2014

Table 5.27. Matrix results from Architectural Design course 2011

Attribute (with respect to)	Domination Degree										Max. Domi- nation
	T1	T2	T3	T4	T5	T6	T7	T8	MT	FE	
T1	0.00	0.00	0.00	0.09	0.06	0.15	0.00	0.00	0.26	0.12	0.26
T2	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.12
T3	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
T4	0.06	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.15	0.00	0.15
T5	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.09
T6	0.00	0.00	0.00	0.18	0.26	0.00	0.00	0.00	0.00	0.12	0.26
T7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.12
T8	0.00	0.00	0.00	0.18	0.44	0.26	0.00	0.00	0.00	0.00	0.44
MT	0.06	0.00	0.00	0.18	0.06	0.00	0.00	0.09	0.00	0.12	0.18
FE	0.00	0.00	0.15	0.09	0.00	0.15	0.24	0.00	0.09	0.00	0.24

Table 5.27 is a matrix that indicates the degree of dominance attribute value of Architectural Design course in 2011. The MDDS technique provides a T8 as the most dominant attribute compared to other attributes, where the maximum domination is 0.44. Hence T8 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.16.

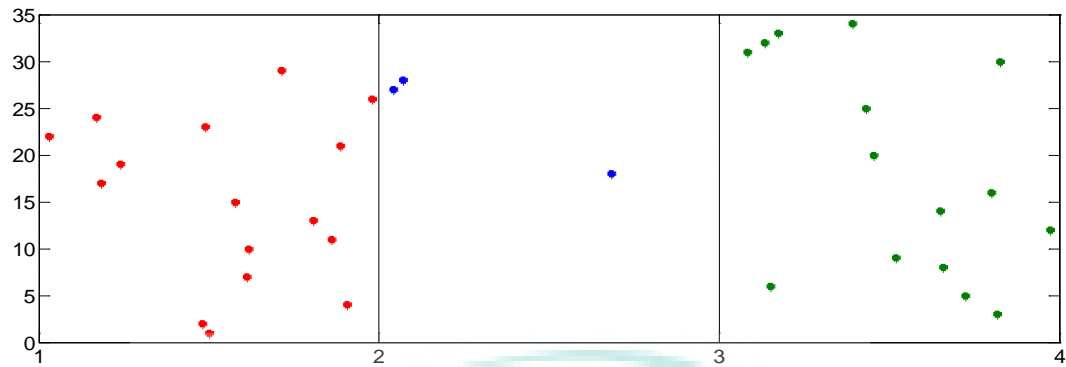


Figure 5.16. Clustering visualization of students on Architectural Design course 2011

Table 5.28. Matrix results from Architectural Design course 2012

Attribute (with respect to)				Domination Degree							Max. Domi- nation
	T1	T2	T3	T4	T5	T6	T7	T8	MT	FE	
T1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T2	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
T3	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.39
T4	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15
T5	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
T6	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
T7	0.06	0.06	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
T8	0.00	0.09	0.00	0.00	0.06	0.12	0.00	0.00	0.00	0.15	0.15
MT	0.00	0.09	0.00	0.00	0.06	0.12	0.00	0.00	0.00	0.15	0.15
FE	0.00	0.06	0.00	0.00	0.06	0.27	0.00	0.58	0.58	0.00	0.58

Table 5.28 is a matrix that indicates the degree of dominance attribute value of Architectural Design course in 2012. The MDDS technique provides a FE as the most dominant attribute compared to other attributes, where the maximum domination is 0.58. Hence FE is selected as a clustering attribute, whereas visualization is divided in two clusters as shown in Figure 5.17.

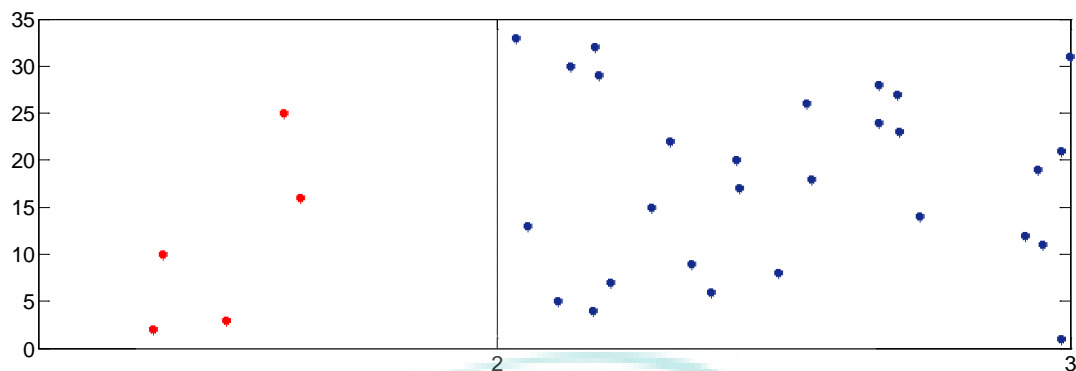


Figure 5.17. Clustering visualization of students on Architectural Design course 2012

Table 5.29. Matrix results from Architectural Design course 2013

Attribute (with respect to)				Domination Degree							Max. Domi- nation
	T1	T2	T3	T4	T5	T6	T7	T8	MT	FE	
T1	0.00	0.00	0.00	0.00	0.48	0.44	0.44	0.26	0.63	0.00	0.63
T2	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.04	0.04	0.00	0.33
T3	0.00	0.52	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.52
T4	0.00	0.52	0.00	0.00	0.48	0.00	0.44	0.04	0.04	0.00	0.52
T5	0.19	0.00	0.00	0.33	0.00	0.00	0.00	0.15	0.41	0.00	0.41
T6	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.37	0.00	0.37
T7	0.19	0.00	0.00	0.33	0.00	0.00	0.00	0.15	0.26	0.00	0.33
T8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.04
MT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.15
FE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.44	0.00	0.44

Table 5.29 is a matrix that indicates the degree of dominance attribute value of Architectural Design course in 2013. The MDDS technique provides a T1 as the most dominant attribute compared to other attributes, where the maximum domination is 0.63. Hence T1 is selected as a clustering attribute, whereas visualization is divided in two clusters as shown in Figure 5.18.

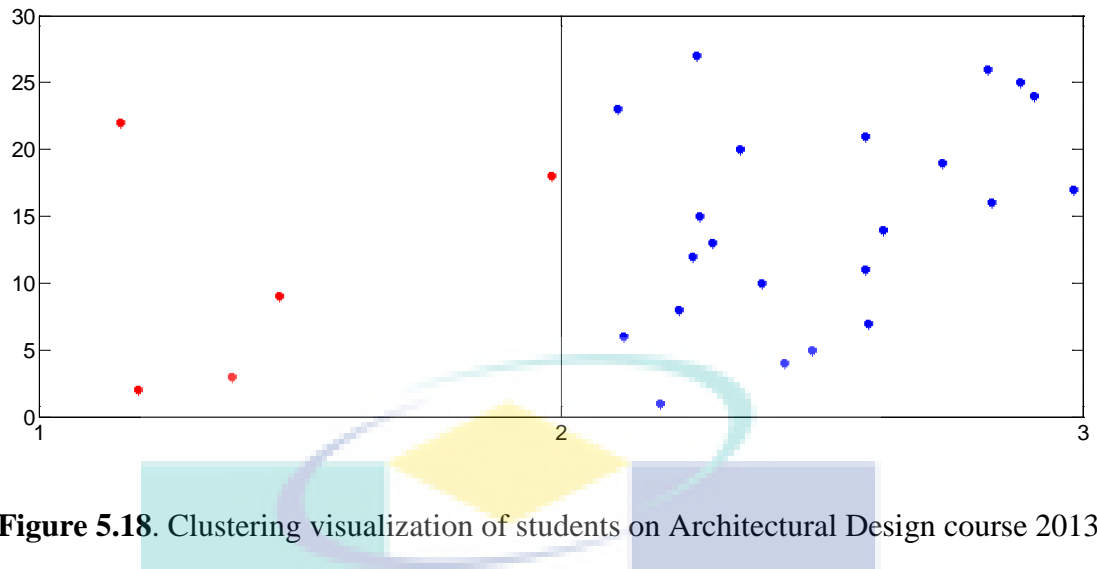


Figure 5.18. Clustering visualization of students on Architectural Design course 2013

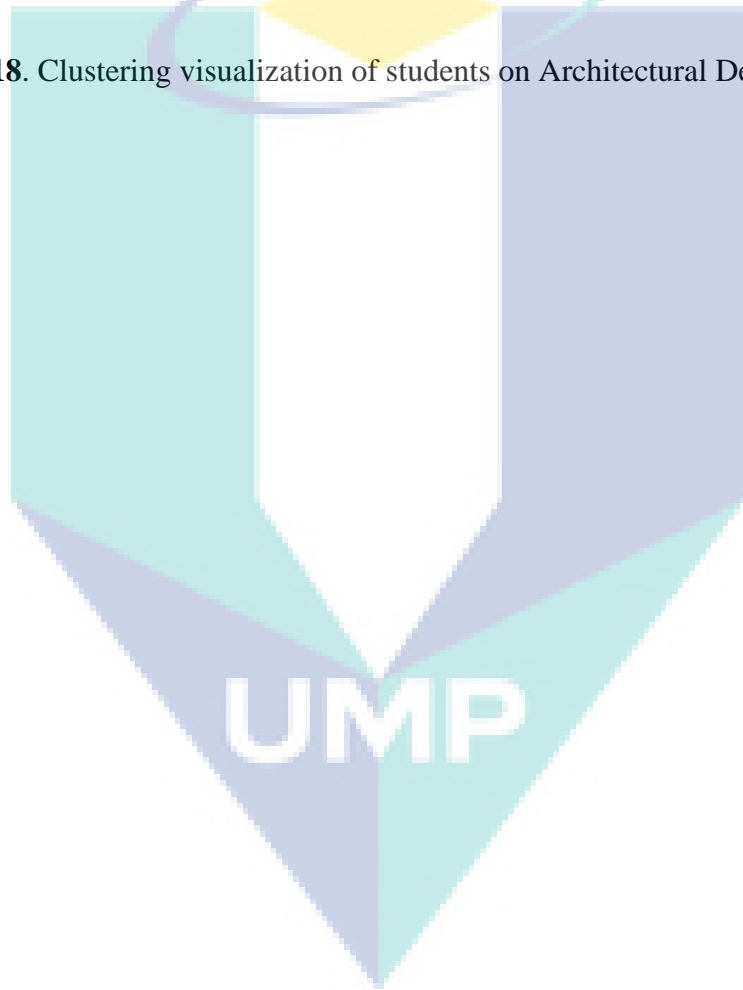


Table 5.30. Matrix results from Infrastructure Design Studio course 2014
(Columns 1 through 15)

Attribute (with respect to)	Domination Degree														
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1	0,00	0,06	0,00	0,03	0,03	0,12	0,18	0,12	0,03	0,03	0,03	0,00	0,09	0,00	0,06
T2	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,00	0,03	0,03	0,03	0,00	0,00	0,00	0,06
T3	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,00	0,03	0,03	0,03	0,15	0,00	0,00	0,06
T4	0,15	0,06	0,00	0,00	0,09	0,09	0,18	0,06	0,03	0,03	0,09	0,12	0,06	0,06	0,06
T5	0,15	0,06	0,09	0,15	0,00	0,00	0,18	0,18	0,21	0,03	0,03	0,00	0,00	0,00	0,00
T6	0,21	0,15	0,00	0,03	0,03	0,00	0,18	0,00	0,03	0,03	0,03	0,00	0,06	0,00	0,00
T7	0,15	0,06	0,00	0,03	0,03	0,15	0,00	0,06	0,03	0,03	0,09	0,00	0,06	0,06	0,15
T8	0,15	0,00	0,00	0,03	0,21	0,12	0,18	0,00	0,03	0,03	0,09	0,00	0,00	0,06	0,06
T9	0,15	0,00	0,00	0,03	0,03	0,00	0,18	0,00	0,00	0,03	0,09	0,00	0,00	0,06	0,00
T10	0,15	0,06	0,00	0,03	0,21	0,00	0,18	0,18	0,21	0,00	0,18	0,00	0,00	0,00	0,00
T11	0,15	0,00	0,00	0,03	0,21	0,00	0,18	0,24	0,21	0,03	0,00	0,00	0,00	0,00	0,00
T12	0,15	0,06	0,00	0,03	0,09	0,00	0,18	0,00	0,03	0,03	0,03	0,00	0,06	0,00	0,00
T13	0,15	0,24	0,00	0,03	0,03	0,00	0,18	0,06	0,03	0,03	0,03	0,00	0,00	0,00	0,00
T14	0,21	0,00	0,09	0,03	0,03	0,00	0,18	0,00	0,03	0,03	0,03	0,00	0,00	0,00	0,06
T15	0,15	0,06	0,09	0,15	0,03	0,00	0,18	0,00	0,03	0,03	0,03	0,00	0,00	0,21	0,00
T16	0,15	0,32	0,00	0,24	0,03	0,00	0,18	0,06	0,03	0,03	0,03	0,00	0,00	0,00	0,00
T17	0,21	0,00	0,15	0,03	0,26	0,00	0,18	0,18	0,21	0,24	0,09	0,00	0,00	0,00	0,00
T18	0,15	0,24	0,00	0,03	0,03	0,00	0,18	0,06	0,03	0,03	0,09	0,00	0,06	0,00	0,00
T19	0,15	0,06	0,09	0,03	0,21	0,00	0,18	0,24	0,03	0,24	0,03	0,00	0,00	0,00	0,24
T20	0,21	0,06	0,00	0,03	0,03	0,00	0,18	0,00	0,03	0,03	0,09	0,00	0,00	0,00	0,00
T21	0,15	0,06	0,00	0,03	0,21	0,00	0,18	0,18	0,03	0,38	0,09	0,00	0,00	0,06	0,06
T22	0,15	0,00	0,09	0,03	0,26	0,00	0,18	0,18	0,03	0,03	0,03	0,00	0,06	0,00	0,00
T23	0,15	0,00	0,00	0,03	0,41	0,00	0,24	0,18	0,03	0,03	0,09	0,00	0,00	0,06	0,15
T24	0,15	0,24	0,00	0,03	0,26	0,09	0,18	0,18	0,21	0,24	0,09	0,00	0,00	0,00	0,00
T25	0,15	0,00	0,00	0,03	0,21	0,00	0,18	0,29	0,21	0,03	0,03	0,00	0,09	0,00	0,06
T26	0,21	0,00	0,00	0,03	0,03	0,00	0,18	0,00	0,03	0,03	0,09	0,00	0,00	0,06	0,06
T27	0,15	0,18	0,00	0,03	0,03	0,00	0,18	0,18	0,21	0,03	0,03	0,00	0,00	0,00	0,06
MT	0,32	0,24	0,00	0,03	0,03	0,00	0,18	0,12	0,03	0,03	0,09	0,00	0,06	0,06	0,06
FE	0,32	0,24	0,00	0,03	0,03	0,00	0,18	0,00	0,03	0,03	0,03	0,00	0,06	0,00	0,06

Table 5.30. (Continued) Matrix results from Infrastructure Design Studio course 2014
(Columns 16 through 30)

Domination Degree														Maximum Domination
T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	MT	FE	
0,03	0,00	0,00	0,03	0,06	0,18	0,00	0,00	0,03	0,12	0,00	0,03	0,21	0,12	0,21
0,03	0,12	0,00	0,03	0,00	0,09	0,00	0,00	0,12	0,00	0,00	0,03	0,00	0,00	0,12
0,03	0,00	0,06	0,03	0,00	0,09	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,12	0,15
0,03	0,00	0,00	0,03	0,06	0,18	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,12	0,18
0,03	0,00	0,00	0,29	0,00	0,18	0,15	0,18	0,12	0,00	0,00	0,03	0,06	0,06	0,29
0,03	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,12	0,00	0,00	0,03	0,00	0,00	0,21
0,03	0,09	0,00	0,03	0,12	0,00	0,00	0,00	0,15	0,00	0,00	0,03	0,21	0,18	0,21
0,03	0,09	0,15	0,15	0,12	0,00	0,00	0,00	0,03	0,12	0,00	0,03	0,00	0,06	0,21
0,03	0,09	0,06	0,03	0,12	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,06	0,18
0,03	0,12	0,00	0,03	0,00	0,00	0,00	0,18	0,03	0,00	0,00	0,03	0,00	0,06	0,21
0,03	0,12	0,15	0,03	0,12	0,00	0,00	0,24	0,03	0,00	0,00	0,32	0,21	0,12	0,32
0,03	0,00	0,09	0,03	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,06	0,18
0,03	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,06	0,00	0,24
0,03	0,00	0,00	0,03	0,06	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,00	0,21
0,03	0,00	0,00	0,03	0,00	0,18	0,00	0,00	0,03	0,00	0,09	0,03	0,00	0,12	0,21
0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,12	0,32
0,03	0,00	0,00	0,03	0,06	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,00	0,26
0,03	0,00	0,00	0,03	0,06	0,18	0,00	0,00	0,03	0,00	0,00	0,03	0,24	0,00	0,24
0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,09	0,03	0,00	0,00	0,24
0,03	0,00	0,00	0,03	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,00	0,21
0,03	0,09	0,00	0,03	0,12	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,06	0,06	0,38
0,03	0,12	0,06	0,03	0,06	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,06	0,18	0,26
0,03	0,09	0,00	0,03	0,06	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,06	0,06	0,41
0,03	0,00	0,24	0,03	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,24	0,32	0,32
0,03	0,12	0,00	0,03	0,00	0,00	0,00	0,24	0,03	0,00	0,09	0,03	0,00	0,06	0,29
0,03	0,00	0,00	0,03	0,06	0,00	0,00	0,00	0,03	0,00	0,00	0,03	0,00	0,00	0,21
0,03	0,00	0,00	0,03	0,06	0,00	0,00	0,00	0,03	0,12	0,00	0,03	0,00	0,00	0,21
0,03	0,00	0,24	0,03	0,06	0,00	0,00	0,00	0,03	0,12	0,00	0,03	0,00	0,00	0,32
0,21	0,00	0,24	0,03	0,00	0,00	0,00	0,00	0,12	0,00	0,00	0,03	0,24	0,00	0,32

Table 5.30 is a matrix that indicates the degree of dominance attribute value of Infrastructure Design Studio course in 2014. The MDDS technique provides a T23 as the most dominant attribute compared to other attributes, where the maximum domination is 0.41. Hence T23 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.19.

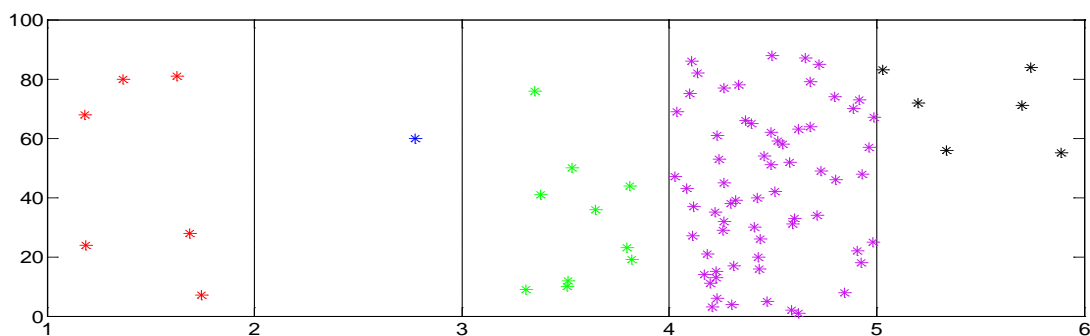


Figure 5.19. Clustering visualization of students on Infrastructure Design Studio course 2014

Table 5.31. Matrix results from Architecture Studio course 2014

Attribute (with respect to)	Domination Degree							Maximum Domination
	T1	T2	T3	T4	T5	MT	FE	
T1	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.02
T2	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
T3	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01
T4	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
T5	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01
MT	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01
FE	0.00	0.01	0.00	0.00	0.00	0.05	0.00	0.05

Table 5.31 is a matrix that indicates the degree of dominance attribute value of Architecture Studio course in 2014. The MDDS technique provides a FE as the most dominant attribute compared to other attributes, where the maximum domination is 0.05. Hence FE is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.20.

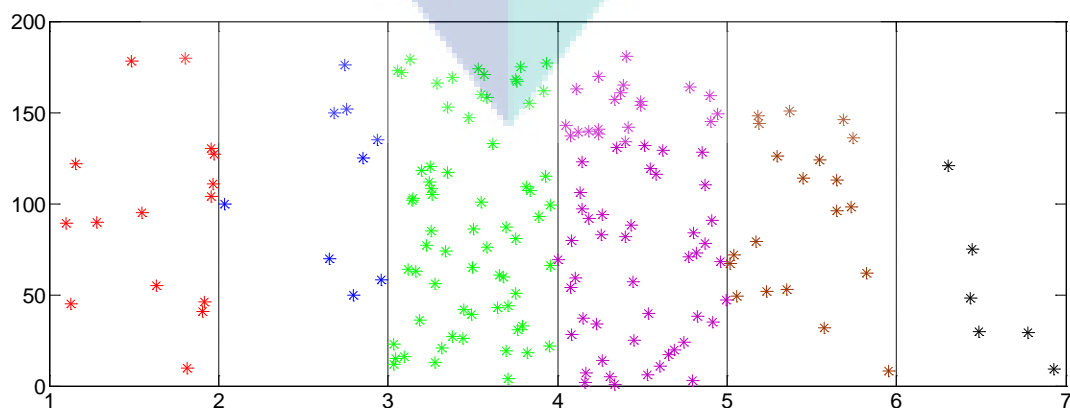


Figure 5.20. Clustering visualization of students on Architecture Studio course 2014

Table 5.32. Matrix results from Contextual Studio course 2014
(Columns 1 through 15)

Attribute (with respect to)	Domination Degree														
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1	0,00	0,06	0,09	0,00	0,06	0,21	0,15	0,15	0,03	0,06	0,09	0,00	0,38	0,00	0,03
T2	0,15	0,00	0,00	0,00	0,15	0,24	0,15	0,09	0,03	0,12	0,03	0,15	0,03	0,00	0,12
T3	0,15	0,00	0,00	0,00	0,09	0,09	0,00	0,00	0,03	0,00	0,03	0,00	0,03	0,00	0,03
T4	0,15	0,06	0,00	0,00	0,15	0,09	0,00	0,00	0,03	0,06	0,09	0,00	0,18	0,00	0,03
T5	0,15	0,21	0,06	0,09	0,00	0,18	0,15	0,32	0,03	0,18	0,03	0,00	0,03	0,00	0,03
T6	0,21	0,21	0,06	0,09	0,06	0,00	0,00	0,00	0,03	0,24	0,09	0,00	0,03	0,09	0,03
T7	0,15	0,21	0,09	0,09	0,00	0,18	0,00	0,00	0,03	0,00	0,03	0,00	0,03	0,00	0,03
T8	0,15	0,15	0,06	0,00	0,15	0,18	0,15	0,00	0,03	0,18	0,09	0,00	0,03	0,12	0,03
T9	0,15	0,15	0,00	0,00	0,00	0,18	0,00	0,12	0,00	0,00	0,09	0,00	0,03	0,00	0,03
T10	0,15	0,21	0,00	0,09	0,00	0,24	0,00	0,18	0,03	0,00	0,09	0,00	0,03	0,00	0,03
T11	0,15	0,15	0,00	0,00	0,00	0,18	0,00	0,26	0,21	0,24	0,00	0,00	0,03	0,00	0,03
T12	0,15	0,21	0,00	0,15	0,00	0,24	0,00	0,09	0,03	0,00	0,09	0,00	0,03	0,00	0,03
T13	0,15	0,21	0,00	0,09	0,06	0,18	0,00	0,00	0,03	0,00	0,03	0,00	0,00	0,00	0,03
T14	0,21	0,15	0,06	0,00	0,06	0,18	0,00	0,12	0,03	0,00	0,03	0,00	0,03	0,00	0,03
T15	0,15	0,21	0,06	0,00	0,00	0,18	0,00	0,00	0,03	0,00	0,03	0,00	0,03	0,12	0,00
T16	0,15	0,21	0,00	0,15	0,00	0,18	0,00	0,06	0,03	0,00	0,03	0,00	0,03	0,00	0,03
T17	0,21	0,15	0,29	0,00	0,21	0,18	0,00	0,18	0,21	0,18	0,03	0,00	0,24	0,00	0,03
T18	0,15	0,21	0,06	0,00	0,06	0,18	0,00	0,12	0,03	0,00	0,03	0,00	0,03	0,00	0,12
T19	0,15	0,21	0,06	0,00	0,06	0,18	0,00	0,18	0,03	0,18	0,03	0,00	0,03	0,00	0,03
T20	0,21	0,21	0,00	0,09	0,00	0,18	0,00	0,00	0,03	0,00	0,03	0,00	0,03	0,00	0,03
T21	0,15	0,21	0,06	0,09	0,00	0,18	0,00	0,18	0,03	0,18	0,03	0,00	0,24	0,00	0,03
T22	0,15	0,15	0,06	0,00	0,26	0,18	0,00	0,26	0,03	0,18	0,09	0,00	0,03	0,00	0,03
T23	0,15	0,15	0,00	0,00	0,06	0,18	0,00	0,18	0,03	0,18	0,09	0,00	0,03	0,00	0,12
T24	0,15	0,21	0,06	0,00	0,00	0,24	0,00	0,24	0,21	0,18	0,03	0,00	0,03	0,00	0,03
T25	0,15	0,15	0,06	0,00	0,00	0,18	0,00	0,18	0,21	0,18	0,03	0,00	0,03	0,12	0,03
T26	0,21	0,15	0,00	0,00	0,00	0,18	0,00	0,00	0,03	0,00	0,03	0,00	0,03	0,09	0,03
T27	0,15	0,15	0,00	0,00	0,00	0,18	0,00	0,00	0,35	0,00	0,03	0,00	0,03	0,00	0,03
T28	0,32	0,21	0,26	0,15	0,00	0,24	0,00	0,35	0,21	0,35	0,32	0,35	0,24	0,21	0,03
MT	0,32	0,21	0,00	0,09	0,00	0,24	0,15	0,00	0,03	0,00	0,03	0,00	0,03	0,00	0,03
FE	0,32	0,21	0,00	0,09	0,00	0,24	0,15	0,06	0,03	0,00	0,03	0,00	0,03	0,00	0,03

Table 5.32. (Continued). Matrix results from Contextual Studio course 2014
(Columns 16 through 30)

Domination Degree															Maximum Domination
T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	MT	FE	
0,09	0,21	0,03	0,03	0,03	0,15	0,00	0,00	0,00	0,41	0,09	0,03	0,35	0,18	0,21	0,41
0,41	0,12	0,03	0,18	0,03	0,15	0,00	0,00	0,18	0,12	0,00	0,03	0,24	0,26	0,21	0,41
0,09	0,00	0,03	0,03	0,09	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,09	0,03	0,15
0,00	0,00	0,12	0,03	0,09	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,26	0,21	0,26
0,00	0,06	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,32
0,00	0,06	0,03	0,03	0,09	0,00	0,09	0,15	0,18	0,00	0,00	0,03	0,09	0,00	0,03	0,24
0,00	0,00	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,21
0,09	0,00	0,12	0,18	0,09	0,00	0,09	0,00	0,00	0,00	0,00	0,03	0,21	0,00	0,03	0,21
0,09	0,00	0,12	0,03	0,09	0,00	0,00	0,00	0,00	0,12	0,00	0,03	0,15	0,09	0,03	0,18
0,09	0,06	0,12	0,03	0,03	0,00	0,00	0,15	0,00	0,00	0,00	0,03	0,15	0,00	0,03	0,24
0,00	0,12	0,03	0,03	0,03	0,00	0,00	0,15	0,00	0,00	0,00	0,03	0,03	0,09	0,03	0,26
0,09	0,00	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,09	0,00	0,03	0,24
0,00	0,06	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,09	0,09	0,03	0,21
0,00	0,06	0,03	0,03	0,09	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,21
0,09	0,06	0,03	0,03	0,03	0,00	0,09	0,00	0,00	0,00	0,00	0,03	0,03	0,09	0,03	0,21
0,00	0,06	0,03	0,03	0,26	0,00	0,00	0,00	0,00	0,00	0,09	0,03	0,03	0,00	0,03	0,26
0,09	0,00	0,03	0,03	0,09	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,29
0,00	0,06	0,00	0,03	0,03	0,00	0,09	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,21
0,00	0,06	0,03	0,00	0,03	0,00	0,09	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,21
0,09	0,06	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,09	0,03	0,03	0,00	0,03	0,21
0,00	0,06	0,03	0,03	0,03	0,00	0,09	0,15	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,24
0,00	0,18	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,26
0,09	0,00	0,03	0,03	0,09	0,15	0,00	0,00	0,00	0,00	0,00	0,03	0,15	0,26	0,21	0,26
0,09	0,00	0,12	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,24
0,09	0,12	0,03	0,03	0,03	0,00	0,00	0,26	0,00	0,00	0,09	0,03	0,09	0,09	0,03	0,26
0,00	0,00	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,00	0,03	0,21
0,00	0,00	0,03	0,03	0,03	0,00	0,00	0,15	0,00	0,12	0,00	0,00	0,03	0,09	0,24	0,35
0,00	0,12	0,03	0,29	0,03	0,00	0,00	0,26	0,29	0,35	0,24	0,03	0,00	0,09	0,03	0,35
0,00	0,00	0,03	0,03	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,09	0,00	0,24	0,32

Table 5.32 is a matrix that indicates the degree of dominance attribute value of Contextual Studio course in 2014. The MDDS technique provides a T28 as the most dominant attribute compared to other attributes, where the maximum domination is 0.35. Hence T28 is selected as a clustering attribute, whereas visualization is divided in three clusters as shown in Figure 5.21.

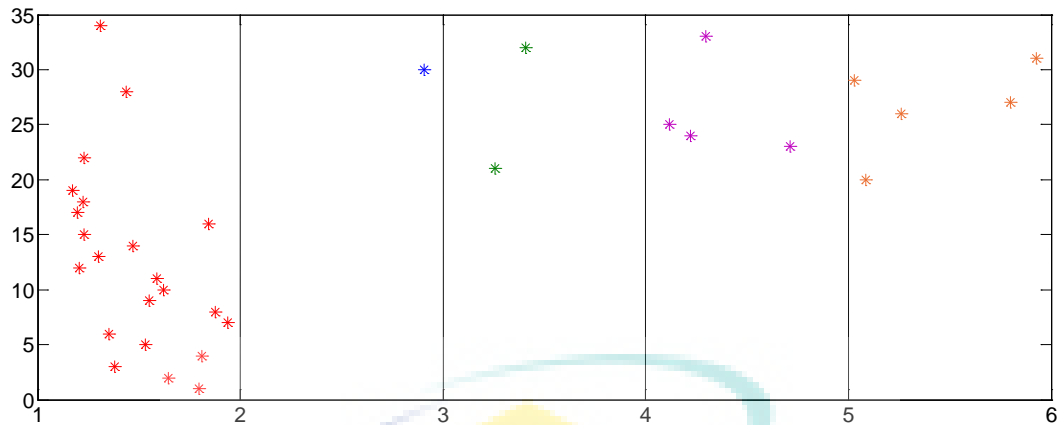


Figure 5.21. Clustering visualization of students on Contextual Studio course 2014

5.4. Execution Time on Data Sets Assessment

In this sub-section, experimental results of the two techniques are presented for the assessment of data sets. Execution time in selecting a clustering of attributes is presented in Table 5.33. Based on the experimental data sets assessment, MDDS technique achieves faster computing time than MAR technique. The average repair time of MDDS has increased to 50.49%. Comparison of computing time between MAR and MDDS techniques in the assessment of data sets are shown in Figure 5.22.

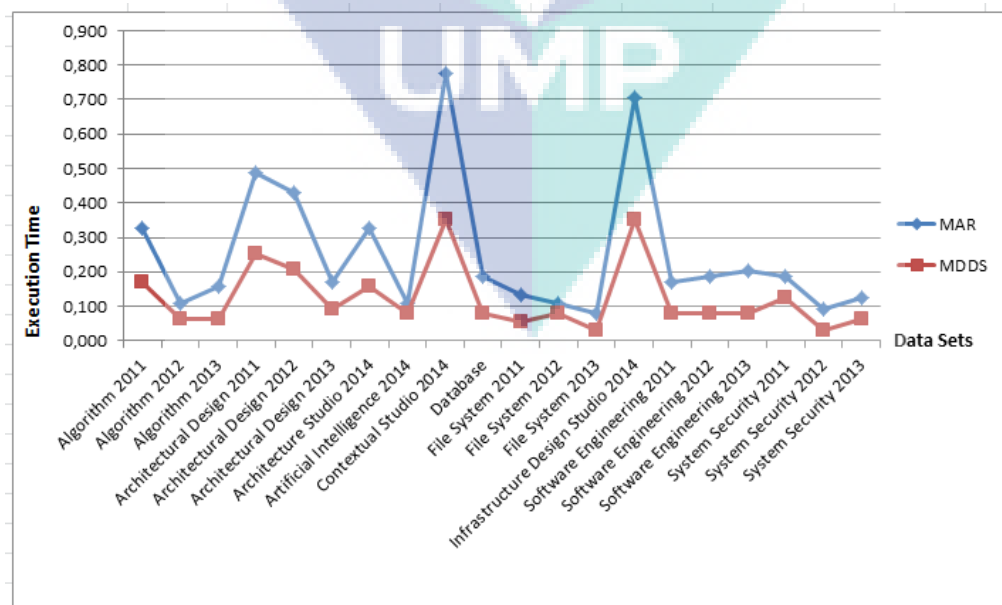


Figure 5.22. Comparison of execution time MAR and MDDS on datasets assessment

Table 5.33. Execution time of comparison results on datasets assessment

No	Course Datasets	MAR	MDDS	Improvement(%)
1	Algorithm 2011	0.327	0.172	47.40
2	Algorithm 2012	0.109	0.063	42.20
3	Algorithm 2013	0.156	0.062	60.26
4	Architectural Design 2011	0.489	0.255	47.92
5	Architectural Design 2012	0.432	0.208	51.81
6	Architectural Design 2013	0.172	0.093	45.93
7	Architecture Studio 2014	0.328	0.158	51.83
8	Artificial Intelligence 2014	0.109	0.078	28.44
9	Contextual Studio 2014	0.777	0.354	54.52
10	Database	0.187	0.078	58.29
11	File System 2011	0.133	0.055	58.65
12	File System 2012	0.109	0.078	28.44
13	File System 2013	0.078	0.031	60.26
14	Infrastructure Design Studio 2014	0.705	0.351	50.18
15	Software Engineering 2011	0.171	0.078	54.39
16	Software Engineering 2012	0.187	0.078	58.29
17	Software Engineering 2013	0.203	0.078	61.58
18	System Security 2011	0.187	0.125	33.16
19	System Security 2012	0.093	0.031	66.67
20	System Security 2013	0.125	0.063	49.60
Average of overall improvement				50.49

5.5. Summary

The MDDS is applied to select attributes of clustering on the data sets assessment. Results of this experiment showed that the selection of a dominant attribute of the data sets assessment can be performed faster than the MAR technique. The speed increases up to 50.49%. This speed is obtained because of the simplification process, so that the number of iterations is reduced. As the selection process attributes can be done faster, so the clustering of students will also be faster.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1. Contributions and Limitations

In chapter 1 have highlighted that exiting clustering techniques are based on rough set theory and soft set theory, but there techniques have limitations, namely a long execution time and low accuracy. At present, the most widely used criteria for evaluating clustering technique include accuracy and efficiency of clustering. Clustering accuracy measures the quality of the clustering technique. A higher value indicates a better clustering result. Clustering efficiency is measured by the execution time. The longer the time of execution, the lower the efficiency. Whereas the accuracy demands a high value. On the contrary, direct optimization-based techniques have high accuracy, while the clustering has a low efficiency compared with a heuristic technique. Therefore, a new clustering technique for categorical data with high efficiency and high classification accuracy is very much required.

Chapter 2 has discussed some issues relating to clustering techniques and the majority of the problems are related to efficiency, low accuracy, and scalability poorly. The clustering technique for selecting attribute is Maximum Relative Attribute (MAR). It is based on a soft set theory by introducing the concept of the attribute relative to the information systems. Based on the experiment on fourteen UCI datasets and a supplier dataset, the MAR technique achieved a lower execution time than the three rough set-based techniques, i.e. TR, MMR, MDA, and NSS. Furthermore, MAR has a good scalability, i.e. technique execution, technique execution time tends to increase linearly as the number of instances and attributes is increased respectively. However, the MAR technique uses complex mathematical models, because this technique calculates the

value of support, max_support and min_support. Support value = 1 is summed as max_support, while others as min_support. Clustering attribute is determined based on the largest amount of max_support. If there is greatest max_support of more than 1, then the clustering attribute is an attribute that has the largest value of min_support.

From the analysis of the limitations of the MAR technique, there is a need to develop clustering algorithm for data categories. The proposed technique is the Maximum Degree of Domination in Soft Set theory (MDDS). The steps of MDDS technique are as follows:

- a. Build the multi-soft set approximation.
- b. Calculate Domination of Attributes.
- c. Select the maximum of domination degree of each attributes.
- d. Select the clustering attribute based on the maximum degree of domination of attributes.

As the input is categorical data and the output is the clustering attribute, MDDS can overcome the limitations of MAR. The MDDS technique has succeeded in improving performance. The execution time and the number of iterations is lower than the MAR technique. Average execution time of 17 data sets is 43.99% faster, at the same time the average number of iterations of the 17 sets of data is reduced to 15.26%. While the accuracy of eight data sets which have a class attributes has increased 3.23%. The number of clusters was not determined from the start, so it will be more for user convenience.

Suppose that in an information system, there are n objects, m attributes and l is the maximum distinct values of each attribute. Computational cost to determine the elementary set of all attributes is nm . The MAR technique needs $ml(ml-1)$ times to determine the support for each category. Thus, the computational complexity for the proposed technique is $O(ml(ml-1)+nm+1)$. The MDDS technique needs $m(m-1)$ times to determine the support for each category. The computational complexity for the proposed technique is $O(nm+m(m-1))$. After comparison with MAR technique, it is clear that the MDDS technique has a lower complexity.

The MAR and MDDS techniques were applied to select attributes of clustering on the data sets assessment. Results of this experiment showed that the selection of a dominant attribute of the data sets assessment can be performed faster than the MAR technique. The speed increases by up to 50.49%. This speed is obtained because of the simplification process, so that the number of iterations is reduced. As the selection process attributes can be done faster, so the clustering of students will also be faster. But accuracy can not be determined because the dataset there has no decision attribute.

The MDDS gives better results than the previous techniques, however it also has some limitations, which includes: The MDDS is more focused on categorical data, whereas in a real database, variety and range of data is enormous. Data is transformed into a form category. Not all data can be processed well by this technique. Data must be transformed into data category. In this technique, this is still done separately. Data transformation has not been conducted properly. The amount of cluster was not specified by the user, and the number of clusters generated may not match expectations. Likewise, big data variations and prevalent in every attribute will generate a lot of clusters, so it is difficult to distinguish from each other. This happens because the distance between the clusters have become very close. MDDS technique was tested on assessment data at the University, while experiments on elementary and secondary education data have not been conducted.

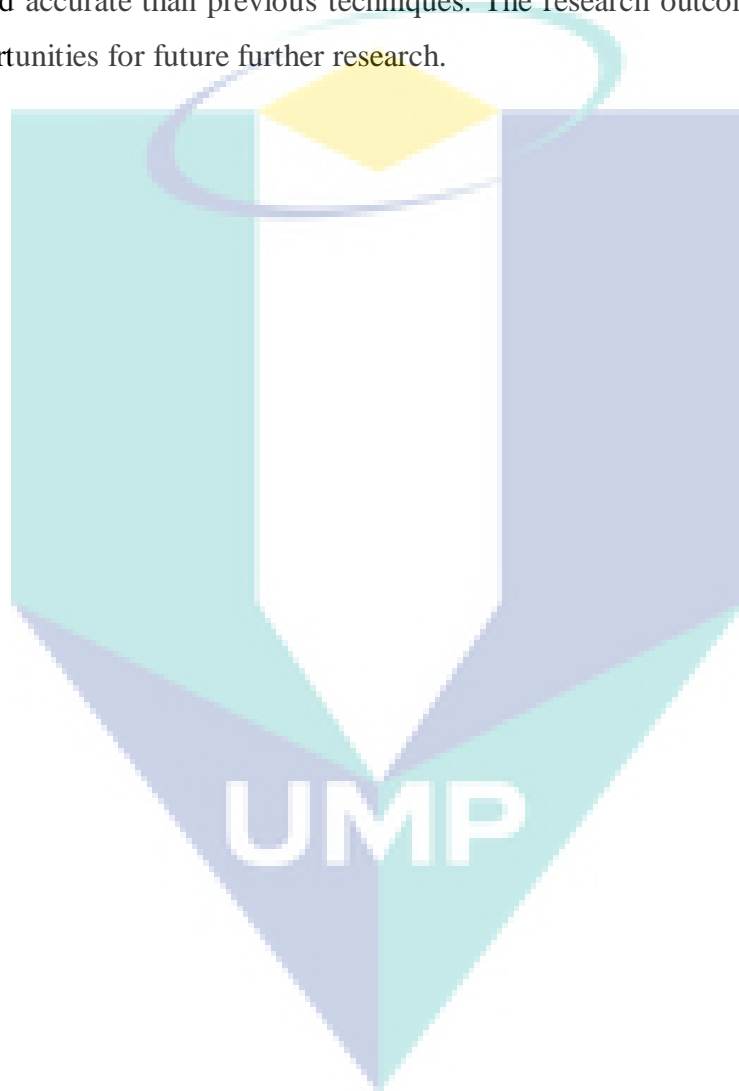
6.2. Recommendations

Based on the results of this thesis and its limitations, the following area of research and development can be conducted in the future:

- a. MDDS technique is more focused on categorical data and therefore future research should focus on technique can handle various types of data.
- b. There is a need for an integrated system to transform the raw data into a categorical data.
- c. As the number of clusters cannot be predicted, it is desirable that in in future research the cluster number should be determined by the user.
- d. The efficiency and accuracy is an issue that must be continuously improved.

6.3. Summary

This chapter presented the research summary namely conclusions and recommendations. Contributions and limitations of the study are fully described. Based on the results of the research, discussion, and evaluation, research objectives in chapter 1 have been achieved. This new clustering technique of categorical data that is more efficient and accurate than previous techniques. The research outcome have opened up many opportunities for future further research.



REFERENCES

- Abdullah, Z., Herawan, T., Ahmad, N., and Deris, M.M., 2011, Extracting highly positive association rules from students enrollment data." *Procedia - Social and Behavioral Sciences*, 28, pp. 107-111.
- Ahmed, S.R., 2004, Applications of data mining in retail business. *Information Technology: Coding and Computing*, 2, pp. 455-459.
- Andritsos, P., 2004, Scalable Clustering of Categorical Data and Applications, Ph.D Thesis, University of Toronto.
- Atkinson-Abutridy, J., Mellish, C., and Aitken, S., 2004, Combining Information extraction with genetic algorithms for text mining, *IEEE Intelligent System*, 19(3), pp. 22-30.
- Baepler, P. and Murdoch, C. J., 2010, Academic Analitics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), pp. 1-9.
- Bagga, S., Singh, G.N., 2011, Three Phase Iterative Model of KDD, *International Journal of Information Technology and Knowledge Management*, 4(2), pp. 695-697.
- Banfield, J.D. and Raftery, A.E., 1993, Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49, pp. 803-821.
- Barbara, D., Li, Y., and Couto, J., 2002, COOLCAT: an entropy-based algorithm for categorical clustering. In *Proceedings of ACM International Conference CIKM'02*, pp. 582-589.
- Beck, J., 2014, The field of EDM: where we came from and where we're going, *Proceedings of the 7th International Conference on Educational Data Mining*, London, United Kingdom.
- Berry, M.J.A., and Linoff, G.S., 2004, *Data mining techniques second edition for marketing, sales, and customer relationship management*. Wiley.
- Bhandari, D., Murthy, C.A., and Pal, S.K., 1996, Genetic Algorithm with elitist model and its convergence. *Internat. J. Patt. Recog. And Art. Intell.*, 10, pp. 731-747.
- Blagojevic, M. and Micic, Z., 2012, A web-based intelligent report e-learning system using data mining techniques. *Computers and Electrical Engineering*, pp. 1-10.
- Caruana, R., Elhawary, M.F., Nguyen, N., and Smith, C., 2006, Meta clustering. In *ICDM*, pp. 107-118.

- Chen, M.S., Han, J., and Yu, P.S., 1996, Data mining: an overview from a database perspective. *IEEE Trans. On Knowledge And Data Engineering* 8, pp. 866-883.
- Chen, Y.L., Hsu, C.L., and Chou, S.C., 2003, Constructing a multi-valued and multilabeled decision tree, *Expert Systems with Applications*, 25, pp. 199–209.
- Cristofor, D. and Simovici, D., 2002, Finding median partitions using information-theoretical-based genetic algorithms, *Journal of Universal Computer Science*, 8(2), pp.153–172.
- Davidson, I. and Qi, Z., 2007, Finding alternative clusterings using constraints. In *ICDM*, pp. 240–249.
- Deli, I. and Cagman, N., 2015, Intuitionistic fuzzy parameterized soft set theory and its decision making, *Applied Soft Computing*, 28, pp. 109-113.
- Deng, S., He, Z., and Xu, X., 2010, G-ANMI : A mutual information based genetic clustering algorithm for categorical data. *Knowledge-Based Systems*, 23, pp. 144–149.
- DuMouchel, W. and Schonlau, M., 1998, A fast computer intrusion detection algorithm based on hypothesis testing of command transition probabilities, In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD98)*, pp. 189–193.
- Dutta, M., Mahanta, A.K., and Pujari, A.K., 2005, QROCK: A quick version of the ROCK : algorithm for clustering of categorical data, *Pattern Recognition Letters*, 26, pp. 2364-2373.
- Fawcett, T. and Provost, F., 1997, Adaptive fraud detection. *Data Mining and Knowledge Discovery*, 1(3), pp. 291–316.
- Fayyad, U., Shapiro, G.P., and Smyth, P., 1996, From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence Magazine*, pp. 37-54.
- Fraley C. and Raftery A.E., 1998, “How Many Clusters? Which Clustering Method? Answers Via Model - Based Cluster Analysis”, Technical Report No. 329. Department of Statistics University of Washington.
- Filho, J.L.R., Treleaven, P.C., and Alippi, C., 1994, Genetic Algorithm programming environments. In *IEEE Computer*, pp.28-43.
- Ganti, V., Gehrke, J., and Ramakrishnan, R., 1999, CACTUS—clustering categorical data using summaries. In *Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 73–83.

- Giannotti, F., Gozzi, G., and Manco, G., 2002, Clustering transactional data. In Proceeding of PKDD'02, pp. 175–187.
- Gibson, D., Kleinberg, J., and Raghavan, P., 2000, Clustering categorical data: an approach based on dynamical systems. *The Very Large Data Bases Journal*, 8 (3-4), pp. 222–236.
- Gondek, D. and Hofmann, T., 2004, Non-redundant data clustering. In *ICDM*, pp. 75–82.
- Guha, S., Rastogi, R., and Shim, K., 2000, ROCK : a robust clustering algorithm for categorical attributes. *Information Systems*, 25(5), pp. 345–366.
- Hadi, A.S., 1994, A modification of a method for the detection of outliers in multivariate samples, *Journal of the Royal Statistical Society, B*, 56(2).
- Haimov, S., Michalev, M., Savchenko, A., and Yordanov, O., 1989, Classification of radar signatures by autoregressive model fitting and cluster analysis. *IEEE Transactions on Geo Science and Remote Sensing*, 8(1), pp. 606–610.
- Han, J., Kamber, M., and Pei, J., 2011, *Data Mining Concept and Techniques*. 3rd, Morgan kaufmann.
- He, Z., Xu, X., and Deng, S., 2002, Squeezer : an efficient algorithm for clustering categorical data. *Journal of Computer Science & Technology*, 17(5), pp. 611–624.
- He, Z., Xu, X., and Deng, S., 2004, A link clustering based approach for clustering categorical data. In *Proceedings of the WAIM Conference*.
- He, Z., Xu, X., and Deng, S., 2008, k-ANMI : a mutual information based clustering algorithm for categorical data, *Information Fusion*, 9(2), pp. 223–233.
- Herawan, T., Deris, M.M., 2009, On multi-soft sets construction in information systems, in: *ICIC (Lecture Notes in Computer Science)*, Springer-Verlag, pp. 101–110.
- Herawan, T., Deris, M.M., and Abawajy, J.H., 2010, A Rough Set Approach for Selecting Clustering Attribute. *Knowledge-Based Systems*, 23(3): pp 220–231.
- Herawan, T., and Deris, M.M., 2011, A Soft Set Approach for Association Rules Mining. *Knowledge Based Systems*, 24(1), pp.186–195.
- Herawan, T., 2012, Rough Clustering for Cancer Datasets, *International Journal of Modern Physics: Conference Series*, 9(1), pp. 240–258.
- Hongwu, Q., Xiuqin, M., Zain, J.M., and Herawan, T., 2012, T.A Novel Soft Set Approach for Selecting Clustering Attribute. *Knowledge-Based Systems*, 36, pp. 139–145.

- Hoppner, F. and Klawonn, F., 2004, "Learning Fuzzy Systems - An Objective Function-Approach", *Mathware & Soft Computing*, 11, pp.143-162.
- Huang, Z., 1997, A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining, *Proceedings of 1997 SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, pp. 1-8.
- Huang, Z., 1998, Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3), pp. 283–304.
- Huang, Z., and Michael, K.N., 1999, A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transaction on Fuzzy Systems*, 7(4), pp. 446–452.
- Jain, A.K., Murty, M.N., and Flynn, P.J., 1999, Data Clustering : A Review, *ACM Computing Surveys*, 31(3), pp. 264-323.
- Jiang, D., Tang, C., and Zhang, A., 2004, Cluster analysis for gene expression data: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 16(11), pp. 1370–1386.
- Kaufman, L. and Rousseeuw, P., 1990, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons.
- Kong Z., Zhang G., Wang L., Wu Z., Qi S., and Wang, H., 2014, An efficient decision making approach in incomplete soft set, *Applied Mathematical Modelling*, 38(7–8), pp. 2141-2150.
- Kong, Z., Jia, W., Zhang, G., and Wang, L., 2015, Normal parameter reduction in soft set based on particle swarm optimization algorithm, *Applied Mathematical Modelling*, 39, pp. 4808-4820.
- Lin, T.Y and Cercone, N., 1997, *Rough Set and Data Mining Analysis of Imprecise Data*, Kluwer Academic Publisher, Massachussets, USA.
- Luxburg, U.V., Bubeck, S., Jegelka, S., and Kaufmann, M., 2007, Consistent Minimization of Clustering Objective Functions, published at NIPS, pp. 1-8.
- Maimon, O. and Rokach, L., 2010, *Data Mining and Knowledge Discovery Handbook* 2nd ed., Springer Science+business media.
- Maji, P.K., and Roy, A.R., 2002, An application of soft sets in a decision making problem. *Computers and Mathematics with Applications*, 44, pp 1077–1083.
- Mamat, R., Herawan, T., and Deris, M.M., 2013, MAR Maximum Attribute Relative of Soft Set in selecting Clustering Attribute. To appear in *Knowledge-Based Systems*, pp. 1-10.

- Mathieu, R., and Gibson, J., 2004, A Methodology for large scale R&D planning based on cluster analysis, *IEEE Transactions on Engineering Management*, 40(3), pp.283–292.
- Mazlack, L.J., He, A., and Zhu, Y., 2000, A Rough Sets Approach in Choosing Partitioning Attributes. I Proceeding of ICSCA 13th. International Conference (CAINE-2000), pp. 1-6.
- Mesakar, S.S., and Chaudhari, M.S., 2012, Review Paper On Data Clustering Of Categorical Data, *International Journal of Engineering Research & Technology (IJERT)*, 1(10), pp. 1-3.
- Mohamad, S.K., and Tasir, Z., 2013, Educational data mining: A review, *The 9th International Conference on Cognitive Science, Procedia - Social and Behavioral Sciences* 97, pp. 320 – 324
- Molodtsov, D., 1999, Soft set theory - First results. *Computer and Mathematics with Applications*, 37(4-5): pp. 19-31.
- Parmar, D., Wu, T., and Blackhurst, J., 2007, MMR: An Algorithm for Clustering Categorical Data using Rough Set Theory. *Data and Knowledge Engineering*, 63(3), pp. 879–893.
- Pawlak, Z., 1982, Rough sets. *International Journal of Computer and Information Science*, 11, pp. 341-356.
- Pawlak, Z. and Skowron, A., 2007, Rudiments of rough sets. *International Journal of Computer and Information Science*, 177(1), pp. 3-27.
- Rezaee, M.R, Lelieveldt, B.P.F., and Reiber, J.H.C., 1998, A new cluster validity index for the fuzzyc-mean. *Pattern Recognition Letters*, 19, pp. 237–246.
- Romero, C., Ventura, S., and Bra, P. D., 2004, Knowledge discovery with genetic programming for providing feedback to courseware author. Kluwer Academic Publisher.
- Saagari, S., Anusha, P.D., Priyanka, C.L., and Sailaja, V.S.S.N., 2013, Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Essential Elements to Support Decision-Making Process in Industries, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2(6), pp. 88-93.
- Sachin, B.R. and Vijay, S.M., 2012, A Survey and Future Vision of Data mining in Educational Field. In *Proceedings of the Second International Conference on Advanced Computing & Communication Technologies*, IEEE Computer Society.
- Tan, P.N., Steinbach, M., and Vipin, K., 2006, Introduction to Data Mining. Addison-Wesley.

- Wassan, J.T, 2015, Discovering Big Data Modelling for Educational World, Procedia - Social and Behavioral Sciences, 176 pp. 642 – 649.
- Wong, P., and Bergeron. R.D., 1997, 30 Years of Multidimensional Multivariate Visualization. In G. M. Nielson, H. Hagan, and H. Muller (Eds), Scientific Visualization -Overviews, Methodologies and Techniques, IEEE Computer Society Press,pp. 3-33.
- Xu, W., Xiao, Z., Dang, X., Yang, D., and Yang, X., 2014, Financial ratio selection for business failure prediction using soft set theory, Knowledge-Based Systems, 63, pp. 59-67
- Yang, Y., Guan, X., and You, J., 2002, Clope: A fast and effective clustering algorithm for transactional data. Proc. of ACM SIGKDD Conference, pp. 682–687.
- Yanto, I.T.R., Herawan, T., and Deris, M.M., 2011, Data clustering using Variable Precision Rough Set. Intelligent Data Analysis, 15(4), pp. 465–482.
- Yanto, I.T.R., Vitasari, P., Herawan, T., and Deris, M.M., 2012, Applying Variable Precision Rough Set Model for Clustering Student Suffering Study's Anxiety. Expert System with Applications, 39(1), pp. 452–459.
- Zaki, M.J., and Peters, M., 2005, "Clicks: Mining Subspace Clusters in Categorical Data via Kpartite Maximal Cliques," Proc. Int'l Conf.Data Eng. (ICDE), pp. 355-356.
- UCI Repository of Machine Learning Database, retrieved from <http://www.ics.uci.edu/~mlern/MLRRepository.html>


 UMP

APPENDIX A

LIST OF PUBLICATIONS

The following publications has been made out of thesis.

Journals :

Suhirman, Tutut Herawan, Haruna Chiroma, and Jasni Mohamad Zain, 2014, Data Mining for Education Decision support : A review, *International Journal : Emerging Technologies in Learning (iJET)*, (6), pp. 4-19.

Suhirman, T. Herawan, I.T.R. Yanto, J.M. Zain, H. Qin and Z., Abdullah. Soft Set Approach for Clustering Student Assessment Data, *Journal of Computational and Theoretical Nanoscience*, Accepted 2015. SCI (IF: 1.343)

Suhirman, Tutut Herawan, Iwan Tri Riyadi Yanto, and Jasni Mohamad Zain, Qin Hongwu, Zailani Abdullah, and Jemal Abawajy, An Alternative Soft Set-based Technique for Selecting a Clustering Attribute, *Journal of Computational and Theoretical Nanoscience*, Accepted 2015. SCI (IF: 1.343)

Conferences :

Suhirman and Jasni Mohamad Zain, Applying Rough Set Theory for Student Clustering on Assessment Datasets, 2013, Proceedings of The 2013 International Conference on Computer Science and Information Technology, Yogyakarta, Indonesia, pp. 63-68.

Suhirman and Jasni Mohamad Zain, On Maximal Degree of Domination for Decision Making, Accepted in The second edition DaEng-2015 is going to be held in Bali, Indonesia on April 25th-26th, 2015.

APPENDIX B

PROGRAMMING CODE

function varargout = MDDSVsMAR(varargin)

```
% MDDSVsMAR M-file for MDDSVsMAR.fig
% MDDSVsMAR, by itself, creates a new MDDSVsMAR or raises the existing
% singleton*.
%
% H = MDDSVsMAR returns the handle to a new MDDSVsMAR or the handle to
% the existing singleton*.
%
% MDDSVsMAR('CALLBACK', hObject, eventData, handles,...) calls the local
% function named CALLBACK in MDDSVsMAR.M with the given input arguments.
%
% MDDSVsMAR('Property','Value',...) creates a new MDDSVsMAR or raises the
% existing singleton*. Starting from the left, property value pairs are
% applied to the GUI before MDDSVsMAR_OpeningFcn gets called. An
% unrecognized property name or invalid value makes property application
% stop. All inputs are passed to MDDSVsMAR_OpeningFcn via varargin.
%
% *See GUI Options on GUIDE's Tools menu. Choose "GUI allows only one
% instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help MDDSVsMAR

% Last Modified by GUIDE v2.5 01-Mar-2014 10:15:42

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name', mfilename, ...
    'gui_Singleton', gui_Singleton, ...
    'gui_OpeningFcn', @MDDSVsMAR_OpeningFcn, ...
    'gui_OutputFcn', @MDDSVsMAR_OutputFcn, ...
    'gui_LayoutFcn', [] , ...
    'gui_Callback', []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargin
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT
```

```

% --- Executes just before MDDSvsMAR is made visible.
function MDDSvsMAR_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)
% varargin   command line arguments to MDDSvsMAR (see VARARGIN)

% Choose default command line output for MDDSvsMAR
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes MDDSvsMAR wait for user response (see UIRESUME)
% uiwait(handles.figure1);

% --- Outputs from this function are returned to the command line.
function varargout = MDDSvsMAR_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;

% --- Executes on button press in browse.
function browse_Callback(hObject, eventdata, handles)
% hObject    handle to browse (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)
set(handles.Out_MAR, 'String', []);
set(handles.out_MDDS, 'String', []);

set(handles.RT_MDDS, 'String', []);
set(handles.RT_MAR, 'String', []);
[filename,pathname]=uigetfile({'*.txt'; '*.xls'; '*.xlsx'});
f = fullfile(pathname,filename);
if isequal(filename,0);
    %initialize_gui(gcbbf, handles, true);
else

    test=find(filename=='.');
    str=filename(test+1:end);
    if strcmp(str, 'txt')==0&strcmp(str, 'xlsx')==0 & strcmp(str, 'xls')==0
        initialize_gui(gcbbf, handles, true);
    else
        if strcmp(str, 'xlsx')==1
            data = xlsread(f);
        else

```

```

        data = textread(f);
    end

    [m,n]=size(data);
    set(handles.attribute, 'String', n);
    set(handles.Object, 'String', m);
    handles.metricdata.dataA = data;

%   handles.metricdata.browsw = data;
%   handles.metricdata.dataset=0;
%   handles.metricdata.nama_data=filename;
    set(handles.nama_data, 'String', filename);
%   set(handles.MDA_list, 'String', []);
%   set(handles.order, 'String', []);
%   set(handles.Spliting, 'String', []);
    guidata(hObject,handles);
    end
end

% --- Executes on button press in proses_MAR.
function proses_MAR_Callback(hObject, eventdata, handles)
% hObject    handle to proses_MAR (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
data=handles.metricdata.dataA;
[maxsp minsp sup indat time]=MAR(data);
set(handles.RT_MAR, 'String', time);
kls=mod(indat,1000);
indat2=(indat-kls)/1000;
ddd=[ maxsp' minsp' indat2' kls'];
outplh=num2str(ddd,'%8.3f    %4.3f  attribute %4.0f class %4.0f\n');
set(handles.Out_MAR, 'String', outplh);

%=====
attDec=find(maxsp==max(maxsp));
if length(attDec)>1
    attDec2=find(minsp(attDec)==max(minsp(attDec)));
    attDec=attDec(attDec2(1));
end
Ind_att_dec=mod(indat(attDec),1000);
attMAR=(indat(attDec)-Ind_att_dec)/1000;
set(handles.dec_mar, 'String', attMAR);
function [new_dom,MIndat_att]=cek_attribute(MIndat_att,Indat_att, dom,domi)
dd=dom([Indat_att,:]);
idd=find(dd~=max(domi));
dd(idd)=0;
next_att_cand=find(sum(dd)'==max(sum(dd)'));
newdom=dom([Indat_att,:]);
new_dom=newdom([next_att_cand,:]);
MIndat_att=MIndat_att(next_att_cand);

%=====

```

```

% --- Executes on selection change in Out_MAR.
function Out_MAR_Callback(hObject, eventdata, handles)
% hObject    handle to Out_MAR (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: contents = cellstr(get(hObject,'String')) returns Out_MAR contents as cell array
%         contents{get(hObject,'Value')} returns selected item from Out_MAR

% --- Executes during object creation, after setting all properties.
function Out_MAR_CreateFcn(hObject, eventdata, handles)
% hObject    handle to Out_MAR (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: listbox controls usually have a white background on Windows.
%         See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'), get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on button press in proses_MDDS.
function proses_MDDS_Callback(hObject, eventdata, handles)
% hObject    handle to proses_MDDS (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
data=handles.metricdata.dataA;
[dom, time, indat]=MDD_S(data);
set(handles.RT_MDDS, 'String', time);
dd=[ max(dom)' indat'];
outpl=num2str(dd,'%8.3f attribute %4.0f \n');
set(handles.out_MDDS, 'String', outpl);
%=====
domi=max(dom)';
Indat_att=find(domi==max(domi));
MIndat_att=Indat_att;
MIndat_att2=0;
% jika lebih dari satu cari
while length(MIndat_att)>1 & length(MIndat_att)~=length(MIndat_att2)
[dom,MIndat_att]=cek_attribute(MIndat_att,Indat_att, dom,domi);
domi=max(dom)';
Indat_att=find(domi==max(domi));
MIndat_att2=MIndat_att;
MIndat_att=MIndat_att;

% attribute yang memuat maksimum domination
end
atMDDS=MIndat_att(1);
set(handles.dec_mdds, 'String', atMDDS);
%=====
% --- Executes on selection change in out_MDDS.

```

```

function out_MDDS_Callback(hObject, eventdata, handles)
% hObject    handle to out_MDDS (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: contents = cellstr(get(hObject,'String')) returns out_MDDS contents as cell array
%         contents{get(hObject,'Value')} returns selected item from out_MDDS

% --- Executes during object creation, after setting all properties.
function out_MDDS_CreateFcn(hObject, eventdata, handles)
% hObject    handle to out_MDDS (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: listbox controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'), get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on button press in B_data.
function B_data_Callback(hObject, eventdata, handles)
% hObject    handle to B_data (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
set(handles.benmarks_panel, 'visible', 'on');
set(handles.uipanel2, 'visible', 'off');
set(handles.uipanel3, 'visible', 'off');
set(handles.browse, 'visible', 'off');
set(handles.radiobutton1, 'value', 0);
set(handles.radiobutton2, 'value', 0);
set(handles.radiobutton3, 'value', 0);
set(handles.radiobutton4, 'value', 0);
set(handles.radiobutton5, 'value', 0);
set(handles.radiobutton6, 'value', 0);
set(handles.radiobutton7, 'value', 0);
set(handles.radiobutton8, 'value', 0);
set(handles.radiobutton9, 'value', 0);
set(handles.attribute, 'String', []);
    set(handles.Object, 'String', []);
    set(handles.nama_data, 'String', []);
    set(handles.RT_MDDS, 'String', []);
set(handles.RT_MAR, 'String', []);
set(handles.Out_MAR, 'String', []);
set(handles.out_MDDS, 'String', []);

% --- Executes on button press in radiobutton14.
function radiobutton14_Callback(hObject, eventdata, handles)
% hObject    handle to radiobutton14 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

```

% Hint: get(hObject,'Value') returns toggle state of radiobutton14

% --- Executes on button press in radiobutton15.
function radiobutton15_Callback(hObject, eventdata, handles)
% hObject handle to radiobutton15 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton15

% --- Executes on button press in radiobutton16.
function radiobutton16_Callback(hObject, eventdata, handles)
% hObject handle to radiobutton16 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton16

% --- Executes on button press in radiobutton17.
function radiobutton17_Callback(hObject, eventdata, handles)
% hObject handle to radiobutton17 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton17

% --- Executes on button press in radiobutton18.
function radiobutton18_Callback(hObject, eventdata, handles)
% hObject handle to radiobutton18 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton18

% --- Executes on button press in radiobutton19.
function radiobutton19_Callback(hObject, eventdata, handles)
% hObject handle to radiobutton19 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton19

% --- Executes on button press in radiobutton20.
function radiobutton20_Callback(hObject, eventdata, handles)
% hObject handle to radiobutton20 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton20

% --- Executes on button press in radiobutton21.
function radiobutton21_Callback(hObject, eventdata, handles)

```

% hObject    handle to radiobutton21 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton21

% --- Executes on button press in radiobutton22.
function radiobutton22_Callback(hObject, eventdata, handles)
% hObject    handle to radiobutton22 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of radiobutton22

% --- Executes on button press in OK.
function OK_Callback(hObject, eventdata, handles)
% hObject    handle to OK (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in cancel.
function cancel_Callback(hObject, eventdata, handles)
% hObject    handle to cancel (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes during object creation, after setting all properties.
function benchmarks_panel_CreateFcn(hObject, eventdata, handles)
% hObject    handle to benchmarks_panel (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% --- Executes on button press in pushbutton8.
function pushbutton8_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton8 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

set(handles.benchmarks_panel, 'visible', 'off');
set(handles.uipanel2, 'visible', 'on');
set(handles.uipanel3, 'visible', 'on');
set(handles.browse, 'visible', 'on');

% --- Executes on button press in pushbutton9.
function pushbutton9_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton9 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
set(handles.benchmarks_panel, 'visible', 'off');
set(handles.uipanel2, 'visible', 'on');
set(handles.uipanel3, 'visible', 'on');
set(handles.attribute, 'String', []);

```



```

set(handles.Object, 'String', []);
set(handles.nama_data, 'String', []);
set(handles.browse, 'visible', 'on');

% --- Executes when selected object is changed in benchmarks_panel.
function benchmarks_panel_SelectionChangeFcn(hObject, eventdata, handles)
% hObject    handle to the selected object in benchmarks_panel
% eventdata  structure with the following fields (see UIBUTTONGROUP)
%   EventName: string 'SelectionChanged' (read only)
%   OldValue: handle of the previously selected object or empty if none was selected
%   NewValue: handle of the currently selected object
% handles    structure with handles and user data (see GUIDATA)
if (hObject == handles.radiobutton1)
    pilih=2;
elseif (hObject == handles.radiobutton2)
    pilih=3;
elseif (hObject == handles.radiobutton3)
    pilih=4;
elseif (hObject == handles.radiobutton4)
    pilih=5;
elseif (hObject == handles.radiobutton5)
    pilih=6;
elseif (hObject == handles.radiobutton6)
    pilih=7;
elseif (hObject == handles.radiobutton7)
    pilih=8;
elseif (hObject == handles.radiobutton8)
    pilih=9;
%=====
elseif (hObject == handles.radiobuttontb1)
    pilih=10;
elseif (hObject == handles.radiobuttontb2)
    pilih=11;
elseif (hObject == handles.radiobuttontb3)
    pilih=12;
elseif (hObject == handles.radiobuttontb4)
    pilih=13;
elseif (hObject == handles.radiobuttontb5)
    pilih=14;
elseif (hObject == handles.radiobuttontb6)
    pilih=15;
elseif (hObject == handles.radiobuttontb7)
    pilih=16;
elseif (hObject == handles.radiobuttontb8)
    pilih=17;
elseif (hObject == handles.radiobuttontb9)
    pilih=18;
%=====
elseif (hObject == handles.radiobutton9)
    pilih=1;
end

```



```

%=====
load data.mat;
switch pilih
case 1
    dataA=zoo;
    nama=sprintf('zoo');
case 2
    dataA =AcuteImflamations;
    nama=sprintf('AcuteImflamations');
case 3
    dataA = chess;nama=sprintf('chess');
case 4
    dataA = flag;nama=sprintf('flag');
case 5
    dataA = lungCancer;nama=sprintf('lungCancer');
case 6
    dataA=mushroom;nama=sprintf('mushroom');
case 7
    dataA=monk;nama=sprintf('monk');
case 8
    dataA=soybean;nama=sprintf('soybean');
case 9
    dataA=suplier;nama=sprintf('suplier');
%=====
case 10
    dataA=lenses;nama=sprintf('Lenses');
case 11
    dataA=ballon;nama=sprintf('Ballon');
case 12
    dataA=car;nama=sprintf('Car');
case 13
    dataA=solar_flare;nama=sprintf('Solar Flare');
case 14
    dataA=Nursery;nama=sprintf('Nursery');
case 15
    dataA=cilinder_band;nama=sprintf('Cilinder Band');
case 16
    dataA=soybean_large;nama=sprintf('Soybean Large');
case 17
    dataA=bank_ok;nama=sprintf('Bank Full');
case 18
    dataA=suplier_large;nama=sprintf('Supplier Large');
%=====

end
set(handles.nama_data, 'String', nama);
[m,n]=size(dataA);
if isempty(dataA)
    set(handles.attribute, 'String', []);
    set(handles.Object, 'String', []);
else

```

```

    set(handles.attribute, 'String', n);
    set(handles.Object, 'String', m);
end
handles.metricdata.dataA = dataA;
guidata(hObject,handles);

```

function [A ovP]=purity(data,ad,kl)

```

CC=data(:,ad);
m=unique(CC);
att=unique(kl);
for i=1:length(m)
    ind=find(CC==m(i));
    for j=1:length(att) ;
        cc=find(kl==att(j));
        dd=intersect(ind,cc);
        A(i,j)=length(dd);
    end
end
B = max(A')./sum(A');
ovP=mean(B);
A=[A B'];

```

function [x,y]=gambar(data,kl)

```

klas_att=data(:,kl);
a=unique(klas_att);
x=[];y=[];
for i=1:length(a)
    yy=find(klas_att==a(i));
    xx=i*ones(1,length(yy))+rand(1,length(yy));
    set(0,'DefaultAxesColorOrder',[rand(1,3)]);
    plot(xx,yy,'*');hold on
    x=[x; xx];
    y=[y; yy];
end
hold off;

```

APPENDIX C1

MAR TECHNIQUE CALCULATION EXAMPLE

Attribute Hair^(Yes)

$$Teeth_{\text{Pointed}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,9\}|} = \frac{2}{3} = 0.6667$$

$$Teeth_{\text{Blunt}} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Teeth_{\text{No}} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Eye_{\text{Forward}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eye_{\text{Side}} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{\text{No}} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4,9\}|} = \frac{4}{5} = 0.8$$

$$Feather_{\text{Yes}} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Feet_{\text{Claw}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{\text{Hoof}} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{\text{Web}} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Feet_{\text{No}} \Rightarrow \frac{|\{0\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{\text{Meat}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eat_{\text{Grass}} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{\text{Grain}} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Eat_{\text{Fish}} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{\text{Yes}} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4\}|} = \frac{4}{4} = 1$$

$$Milk_{\text{No}} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Fly_{\text{No}} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{4}{7} = 0.5714$$

$$Fly_{\text{Yes}} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{\text{Yes}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Swim_{\text{No}} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Hair^(No)

$$Teeth_{\text{Pointed}} \Rightarrow \frac{|\{9\}|}{|\{1,2,9\}|} = \frac{1}{3} = 0.3333$$

$$Teeth_{\text{Blunt}} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Teeth_{\text{No}} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{9\}|}{|\{1,2,3,4,9\}|} = \frac{1}{5} = 0.2$$

$$Feet_{Claw} \Rightarrow \frac{|\{5,7,8\}|}{|\{1,2,5,7,8\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Eat_{Meat} \Rightarrow \frac{|\{8,9\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eat_{Grain} \Rightarrow \frac{|\{5,7\}|}{|\{5,7\}|} = \frac{2}{2} = 1$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{5,6,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Swim_{Yes} \Rightarrow \frac{|\{6,7\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Side} \Rightarrow \frac{|\{5,6,7\}|}{|\{3,4,5,6,7\}|} = \frac{3}{5} = 0.6$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8,9\}|}{|\{5,6,7,8,9\}|} = \frac{5}{5} = 1$$

$$Fly_{Yes} \Rightarrow \frac{|\{7,8\}|}{|\{7,8\}|} = \frac{2}{2} = 1$$

$$Swim_{No} \Rightarrow \frac{|\{5,8,9\}|}{|\{3,4,5,8,9\}|} = \frac{3}{5} = 0.6$$

Attribute **Teeth**_(Pointed)

$$Hair_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,3,4,9\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Hair_{No} \Rightarrow \frac{|\{9\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{|\{3,4,5,6,7\}|} = \frac{0}{5} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{9\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{9\}|}{|\{3,4,5,8,9\}|} = \frac{1}{5} = 0.2$$

Attribute Teeth_(Blunt)

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,9\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,5,7,8\}|} = \frac{0}{5} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{2}{7} = 0.2857$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,6,7\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Teeth_(No)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4,9\}|} = \frac{0}{5} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{7,8\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8,9\}|} = \frac{4}{5} = 0.8$$

$$Eye_{Side} \Rightarrow \frac{|\{5,6,7\}|}{|\{3,4,5,6,7\}|} = \frac{3}{5} = 0.6$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Eat_{Meat} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Eat_{Grain} \Rightarrow \frac{|\{5,7\}|}{|\{5,7\}|} = \frac{2}{2} = 1$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{5,6\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{2}{7} = 0.2857$$

$$Swim_{Yes} \Rightarrow \frac{|\{6,7\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8,9\}|} = \frac{4}{5} = 0.8$$

$$Fly_{Yes} \Rightarrow \frac{|\{7,8\}|}{|\{7,8\}|} = \frac{2}{2} = 1$$

$$Swim_{No} \Rightarrow \frac{|\{5,8\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Eye(Forward)

$$Hair_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,9\}|} = \frac{3}{3} = 1$$

$$Teeth_{No} \Rightarrow \frac{|\{8\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,3,4,9\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2,8\}|}{|\{1,2,5,7,8\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,8,9\}|}{|\{1,2,8,9\}|} = \frac{4}{4} = 1$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Hair_{No} \Rightarrow \frac{|\{8,9\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{8\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{8,9\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Fly_{Yes} \Rightarrow \frac{|\{8\}|}{|\{7,8\}|} = \frac{1}{2} = 0.5$$

$$Swim_{No} \Rightarrow \frac{|\{8,9\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Eye_(Side)

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6,7\}|}{|\{5,6,7,8\}|} = \frac{3}{4} = 0.75$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,9\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Claw} \Rightarrow \frac{|\{5,7\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{5,7\}|}{|\{5,7\}|} = \frac{2}{2} = 1$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{3,4,5,6\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{4}{7} = 0.5714$$

$$Swim_{Yes} \Rightarrow \frac{|\{6,7\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7\}|}{|\{5,6,7,8\}|} = \frac{3}{4} = 0.75$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Fly_{Yes} \Rightarrow \frac{|\{7\}|}{|\{7,8\}|} = \frac{1}{2} = 0.5$$

$$Swim_{No} \Rightarrow \frac{|\{3,4,5\}|}{|\{3,4,5,8,9\}|} = \frac{3}{5} = 0.6$$

Attribute Feather_(No)

$$Hair_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4\}|} = \frac{4}{4} = 1$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,9\}|} = \frac{3}{3} = 1$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Hair_{No} \Rightarrow \frac{|\{9\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4\}|} = \frac{4}{4} = 1$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,3,4,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{5}{7} = 0.7143$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{9\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4,9\}|}{|\{3,4,5,8,9\}|} = \frac{3}{5} = 0.6$$

Attribute Feather_(Yes)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Eye_{Forward} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Feet_{Claw} \Rightarrow \frac{|\{7,8\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Eat_{Meat} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Eat_{Grain} \Rightarrow \frac{|\{5,7\}|}{|\{5,7\}|} = \frac{2}{2} = 1$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{5,6\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{2}{7} = 0.2857$$

$$Swim_{Yes} \Rightarrow \frac{|\{6,7\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,9\}|}{|\{5,6,7,8,9\}|} = \frac{4}{5} = 0.4$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{5,6,7\}|}{|\{3,4,5,6,7\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8,9\}|} = \frac{4}{5} = 0.8$$

$$Fly_{Yes} \Rightarrow \frac{|\{7,8\}|}{|\{7,8\}|} = \frac{2}{2} = 1$$

$$Swim_{No} \Rightarrow \frac{|\{5,8\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Feet_(Claw)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Eye_{Forward} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,9\}|} = \frac{2}{5} = 0.4$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,8\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Eat_{Grain} \Rightarrow \frac{|\{5,7\}|}{|\{5,7\}|} = \frac{2}{2} = 1$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,5\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2,7\}|}{|\{1,2,6,7\}|} = \frac{3}{4} = 0.75$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,9\}|}{|\{5,6,7,8,9\}|} = \frac{4}{5} = 0.8$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{5,6,7\}|}{|\{3,4,5,6,7\}|} = \frac{3}{5} = 0.6$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7\}|}{|\{5,6,7,8\}|} = \frac{3}{4} = 0.75$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,7,8\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Fly_{Yes} \Rightarrow \frac{|\{7,8\}|}{|\{7,8\}|} = \frac{2}{2} = 1$$

$$Swim_{No} \Rightarrow \frac{|\{5,8\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Feet_(Hoof)

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,9\}|} = \frac{2}{5} = 0.4$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{2}{7} = 0.2857$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,6,7\}|} = \frac{0}{4} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Feet_(Web)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{6\}|}{|\{3,4,5,6,7\}|} = \frac{1}{5} = 0.2$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4,9\}|} = \frac{0}{5} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Fly_{No} \Rightarrow \frac{|\{6\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{1}{7} = 0.1429$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{6\}|}{|\{1,2,6,7\}|} = \frac{1}{4} = 0.25$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{|\{3,4,5,8,9\}|} = \frac{0}{5} = 0$$

Attribute Feet_(No)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{9\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{9\}|}{|\{1,2,9\}|} = \frac{1}{3} = 0.3333$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Eye_{Forward} \Rightarrow \frac{|\{9\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Feather_{No} \Rightarrow \frac{|\{9\}|}{|\{1,2,3,4,9\}|} = \frac{1}{5} = 0.2$$

$$Eat_{Meat} \Rightarrow \frac{|\{9\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{1}{7} = 0.1429$$

$$Swim_{Yes} \Rightarrow \frac{|\{6\}|}{|\{1,2,6,7\}|} = \frac{1}{4} = 0.25$$

$$Eye_{Side} \Rightarrow \frac{|\{6\}|}{|\{3,4,5,6,7\}|} = \frac{1}{5} = 0.2$$

$$Feather_{Yes} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{9\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{9\}|}{|\{3,4,5,8,9\}|} = \frac{1}{5} = 0.2$$

Attribute Eat_(Meat)

$$Hair_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,9\}|} = \frac{3}{3} = 1$$

$$Teeth_{No} \Rightarrow \frac{|\{8\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,8,9\}|}{|\{1,2,8,9\}|} = \frac{4}{4} = 1$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,3,4,9\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2,8\}|}{|\{1,2,5,7,8\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Hair_{No} \Rightarrow \frac{|\{8,9\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{|\{3,4,5,6,7\}|} = \frac{0}{5} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{8\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{8,9\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Fly_{Yes} \Rightarrow \frac{|\{8\}|}{|\{7,8\}|} = \frac{1}{2} = 0.5$$

$$Swim_{No} \Rightarrow \frac{|\{8,9\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Eat_(Grass)

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,9\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,5,7,8\}|} = \frac{0}{5} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{2}{7} = 0.2857$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,6,7\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{1}{2} = 1$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Eat_(Grain)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,7\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4,9\}|} = \frac{0}{5} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{5,7\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Hair_{No} \Rightarrow \frac{|\{5,7\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{5,7\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,7\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,7\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Fly_{No} \Rightarrow \frac{|\{5\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{1}{7} = 0.1429$$

$$Fly_{Yes} \Rightarrow \frac{|\{7\}|}{|\{7,8\}|} = \frac{1}{2} = 0.5$$

$$Swim_{Yes} \Rightarrow \frac{|\{7\}|}{|\{1,2,6,7\}|} = \frac{1}{4} = 0.25$$

$$Swim_{No} \Rightarrow \frac{|\{5\}|}{|\{3,4,5,8,9\}|} = \frac{1}{5} = 0.2$$

Attribute Eat_(Fish)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,8,9\}|} = \frac{0}{4} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{6\}|}{|\{3,4,5,6,7\}|} = \frac{1}{5} = 0.2$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4,9\}|} = \frac{0}{5} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8\}|} = \frac{1}{4} = 0.25$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,5,7,8\}|} = \frac{0}{5} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{6\}|}{|\{5,6,7,8,9\}|} = \frac{1}{5} = 0.2$$

$$Fly_{No} \Rightarrow \frac{|\{6\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{1}{7} = 0.1429$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{6\}|}{|\{1,2,6,7\}|} = \frac{1}{4} = 0.25$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{|\{3,4,5,8,9\}|} = \frac{0}{5} = 0$$

Attribute Milk_(Yes)

$$Hair_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4\}|} = \frac{4}{4} = 1$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8,9\}|} = \frac{0}{5} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,9\}|} = \frac{2}{3} = 0.6667$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4,9\}|} = \frac{4}{5} = 0.8$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{5,6,7,8\}|} = \frac{0}{4} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{|\{5,7\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{4}{7} = 0.5714$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{7,8\}|} = \frac{0}{2} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{|\{3,4,5,8,9\}|} = \frac{2}{5} = 0.4$$

Attribute Milk_(No)

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,8,9\}|}{|\{5,6,7,8,9\}|} = \frac{5}{5} = 1$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{9\}|}{|\{1,2,9\}|} = \frac{1}{3} = 0.3333$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Eye_{Forward} \Rightarrow \frac{|\{8,9\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Side} \Rightarrow \frac{|\{5,6,7\}|}{|\{3,4,5,6,7\}|} = \frac{3}{5} = 0.6$$

$$Feather_{No} \Rightarrow \frac{|\{9\}|}{|\{1,2,3,4,9\}|} = \frac{1}{5} = 0.2$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7,8\}|}{|\{5,6,7,8\}|} = \frac{4}{4} = 1$$

$$Feet_{Claw} \Rightarrow \frac{|\{5,7,8\}|}{|\{1,2,5,7,8\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Eat_{Meat} \Rightarrow \frac{|\{8,9\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eat_{Grain} \Rightarrow \frac{|\{7\}|}{|\{5,7\}|} = \frac{1}{2} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{5,6,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Swim_{Yes} \Rightarrow \frac{|\{6,7\}|}{|\{1,2,6,7\}|} = \frac{2}{4} = 0.5$$

$$Eat_{Grass} \Rightarrow \frac{|\{\emptyset\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Fly_{Yes} \Rightarrow \frac{|\{7,8\}|}{|\{7,8\}|} = \frac{2}{2} = 1$$

$$Swim_{No} \Rightarrow \frac{|\{5,8,9\}|}{|\{3,4,5,8,9\}|} = \frac{3}{5} = 0.6$$

Attribute Fly_(No)

$$Hair_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4\}|} = \frac{4}{4} = 1$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,9\}|} = \frac{3}{3} = 1$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,3,4,9\}|}{|\{1,2,3,4,9\}|} = \frac{5}{5} = 1$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2,5\}|}{|\{1,2,5,7,8\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,9\}|}{|\{1,2,8,9\}|} = \frac{3}{4} = 0.75$$

$$Eat_{Grain} \Rightarrow \frac{|\{5\}|}{|\{5,7\}|} = \frac{1}{2} = 0.5$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{|\{1,2,3,4\}|} = \frac{4}{4} = 1$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2,6\}|}{|\{1,2,6,7\}|} = \frac{3}{4} = 0.75$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,9\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4,5,6\}|}{|\{3,4,5,6,7\}|} = \frac{4}{5} = 0.8$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,9\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Swim_{No} \Rightarrow \frac{|\{3,4,5,9\}|}{|\{3,4,5,8,9\}|} = \frac{4}{5} = 0.8$$

Attribute Fly_(Yes)

$$Hair_{Yes} \Rightarrow \frac{|\{\emptyset\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{7,8\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Teeth_{\text{Pointed}} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,9\}|} = \frac{0}{3} = 0$$

$$Teeth_{\text{No}} \Rightarrow \frac{|\{7,8\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Eye_{\text{Forward}} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Feather_{\text{No}} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4,9\}|} = \frac{0}{5} = 0$$

$$Feet_{\text{Claw}} \Rightarrow \frac{|\{7,8\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{\text{Web}} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Eat_{\text{Meat}} \Rightarrow \frac{|\{8\}|}{|\{1,2,8,9\}|} = \frac{1}{4} = 0.25$$

$$Eat_{\text{Grain}} \Rightarrow \frac{|\{7\}|}{|\{5,7\}|} = \frac{1}{2} = 0.5$$

$$Milk_{\text{Yes}} \Rightarrow \frac{|\{\phi\}|}{|\{1,2,3,4\}|} = \frac{0}{4} = 0$$

$$Swim_{\text{Yes}} \Rightarrow \frac{|\{7\}|}{|\{1,2,6,7\}|} = \frac{1}{4} = 0.25$$

$$Teeth_{\text{Blunt}} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eye_{\text{Side}} \Rightarrow \frac{|\{7\}|}{|\{3,4,5,6,7\}|} = \frac{1}{5} = 0.2$$

$$Feather_{\text{Yes}} \Rightarrow \frac{|\{7,8\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Feet_{\text{Hoof}} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{\text{No}} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{\text{Grass}} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{\text{Fish}} \Rightarrow \frac{|\{\phi\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{\text{No}} \Rightarrow \frac{|\{7,8\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Swim_{\text{No}} \Rightarrow \frac{|\{8\}|}{|\{3,4,5,8,9\}|} = \frac{1}{5} = 0.2$$

Attribute Swim_(Yes)

$$Hair_{\text{Yes}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{\text{Pointed}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,9\}|} = \frac{2}{3} = 0.6667$$

$$Teeth_{\text{No}} \Rightarrow \frac{|\{6,7\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Eye_{\text{Forward}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Feather_{\text{No}} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4,9\}|} = \frac{2}{5} = 0.4$$

$$Feet_{\text{Claw}} \Rightarrow \frac{|\{1,2,7\}|}{|\{1,2,5,7,8\}|} = \frac{3}{5} = 0.6$$

$$Feet_{\text{Web}} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Hair_{\text{No}} \Rightarrow \frac{|\{6,7\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Teeth_{\text{Blunt}} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eye_{\text{Side}} \Rightarrow \frac{|\{6,7\}|}{|\{3,4,5,6,7\}|} = \frac{2}{5} = 0.4$$

$$Feather_{\text{Yes}} \Rightarrow \frac{|\{6,7\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Feet_{\text{Hoof}} \Rightarrow \frac{|\{\phi\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Feet_{\text{No}} \Rightarrow \frac{|\{\phi\}|}{|\{9\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eat_{Grain} \Rightarrow \frac{|\{7\}|}{|\{5,7\}|} = \frac{1}{2} = 0.5$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,6\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{3}{7} = 0.4286$$

$$Eat_{Grass} \Rightarrow \frac{|\{\emptyset\}|}{|\{3,4\}|} = \frac{0}{2} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{|\{6\}|} = \frac{1}{1} = 1$$

$$Milk_{No} \Rightarrow \frac{|\{6,7\}|}{|\{5,6,7,8,9\}|} = \frac{2}{5} = 0.4$$

$$Fly_{Yes} \Rightarrow \frac{|\{7\}|}{|\{7,8\}|} = \frac{1}{2} = 0.5$$

Attribute Swim_(No)

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{9\}|}{|\{1,2,9\}|} = \frac{1}{3} = 0.3333$$

$$Teeth_{No} \Rightarrow \frac{|\{5,8\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Eye_{Forward} \Rightarrow \frac{|\{8,9\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Feather_{No} \Rightarrow \frac{|\{3,4,9\}|}{|\{1,2,3,4,9\}|} = \frac{3}{5} = 0.6$$

$$Feet_{Claw} \Rightarrow \frac{|\{5,8\}|}{|\{1,2,5,7,8\}|} = \frac{2}{5} = 0.4$$

$$Feet_{Web} \Rightarrow \frac{|\{\emptyset\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{8,9\}|}{|\{1,2,8,9\}|} = \frac{2}{4} = 0.5$$

$$Eat_{Grain} \Rightarrow \frac{|\{5\}|}{|\{5,7\}|} = \frac{1}{2} = 0.5$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{|\{1,2,3,4\}|} = \frac{2}{4} = 0.5$$

$$Fly_{No} \Rightarrow \frac{|\{3,4,5,9\}|}{|\{1,2,3,4,5,6,9\}|} = \frac{4}{7} = 0.5714$$

$$Hair_{No} \Rightarrow \frac{|\{5,8,9\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4,5\}|}{|\{3,4,5,6,7\}|} = \frac{3}{5} = 0.6$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,8\}|}{|\{5,6,7,8\}|} = \frac{2}{4} = 0.5$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{|\{9\}|} = \frac{1}{1} = 1$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{|\{3,4\}|} = \frac{2}{2} = 1$$

$$Eat_{Fish} \Rightarrow \frac{|\{\emptyset\}|}{|\{6\}|} = \frac{0}{1} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,8,9\}|}{|\{5,6,7,8,9\}|} = \frac{3}{5} = 0.6$$

$$Fly_{Yes} \Rightarrow \frac{|\{8\}|}{|\{7,8\}|} = \frac{1}{2} = 0.5$$

APPENDIX C2

MDDS TECHNIQUE CALCULATION EXAMPLE

Domination of attribute $Hair_{Yes}$

$$Hair_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Hair_{No}$

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,8,9\}|}{9} = \frac{5}{9} = 0.5556$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8,9\}|}{9} = \frac{5}{9} = 0.5556$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute **Teeth_{Pointed}**

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{1,2,9\}|}{9} = \frac{3}{9} = 0.3333$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,9\}|}{9} = \frac{3}{9} = 0.3333$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,9\}|}{9} = \frac{3}{9} = 0.3333$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,9\}|}{9} = \frac{3}{9} = 0.3333$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,9\}|}{9} = \frac{3}{9} = 0.3333$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Teeth_{Blunt}$

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Fly_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

Domination of attribute $Teeth_{No}$

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Eye_{Forward}$

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,8,9\}|}{9} = \frac{4}{9} = 0.4444$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,8,9\}|}{9} = \frac{4}{9} = 0.4444$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Eye_{Side}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4,5,6,7\}|}{9} = \frac{5}{9} = 0.5556$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Feather_{No}$

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,3,4,9\}|}{9} = \frac{5}{9} = 0.5556$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,3,4,9\}|}{9} = \frac{5}{9} = 0.5556$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Feather_{Yes}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9} = 0.4444$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Feet_{Claw}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{1,2,5,7,8\}|}{9} = \frac{5}{9} = 0.5556$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Feet_{Hoof}

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{0}{9} = 0.2222$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Fly_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

Domination of attribute Feet_{Web}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Feather_{Yes} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Milk_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Fly_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Swim_{Yes} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Feet_{No}$

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eat_{Meat} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Fly_{No} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{9\}|}{9} = \frac{1}{9} = 0.1111$$

Domination of attribute Eat_{Meat}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{1,2,8,9\}|}{9} = \frac{4}{9} = 0.4444$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{1,2,8,9\}|}{9} = \frac{4}{9} = 0.4444$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Eat_{Grass}

$$Hair_{Yes} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Eye_{Side} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4\}|}{9} = \frac{2}{9} = 0.2222$$

Domination of attribute Eat_{Fish}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Swim_{Yes} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Hair_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Feather_{Yes} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Milk_{No} \Rightarrow \frac{|\{6\}|}{9} = \frac{1}{9} = 0.1111$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Milk_{Yes}$

$$Hair_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,3,4\}|}{9} = \frac{4}{9} = 0.4444$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Milk_{No}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{5,6,7,8,9\}|}{9} = \frac{5}{9} = 0.5556$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{5,6,7,8,9\}|}{9} = \frac{5}{9} = 0.5556$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Fly_{No}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{1,2,3,4,5,6,9\}|}{9} = \frac{7}{9} = 0.7778$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Fly_{Yes}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{7,8\}|}{9} = \frac{2}{9} = 0.2222$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{7,8\}|}{9} = \frac{2}{9} = 0.2222$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{7,8\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{7,8\}|}{9} = \frac{2}{9} = 0.2222$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{7,8\}|}{9} = \frac{2}{9} = 0.2222$$

$$Fly_{Yes} \Rightarrow \frac{|\{7,8\}|}{9} = \frac{2}{9} = 0.2222$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute Swim_{Yes}

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{1,2,6,7\}|}{9} = \frac{4}{9} = 0.4444$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

Domination of attribute $Swim_{No}$

$$Hair_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Pointed} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Forward} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Feather_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Claw} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Web} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Meat} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0.$$

$$Eat_{Grain} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Hair_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Teeth_{Blunt} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eye_{Side} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feather_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{Hoof} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Feet_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Grass} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Eat_{Fish} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Milk_{No} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Fly_{Yes} \Rightarrow \frac{|\{\phi\}|}{9} = \frac{0}{9} = 0$$

$$Swim_{No} \Rightarrow \frac{|\{3,4,5,8,9\}|}{9} = \frac{5}{9} = 0.5556$$