

A MULTIDIMENSIONAL DATA
DESCRIPTOR TOOL BASED ON
FUZZY MIN MAX NEURAL NETWORK
ALGORITHM

FATIN NURJANNAH BINTI SALAUDDIN

Bachelor of Computer Science (Graphics and
Multimedia Technology)

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : FATIN NURJANNAH BINTI SALAUDDIN
Date of Birth : 24 JULY 1996
Title : A MULTIDIMENSIONAL DATA DESCRIPTOR TOOL BASED ON FUZZY MIN MAX NEURAL NETWORK ALGORITHM
Academic Session : 2018/2019

I declare that this thesis is classified as:

- CONFIDENTIAL (Contains confidential information under the Official Secret Act 1997)*
 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 reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis 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)


(Supervisor's Signature)

960724-14-5044

Date: 10 JANUARY 2019


Name of Supervisor

Date: 10 JANUARY 2019

NOTE : * If the thesis is CONFIDENTIAL, or RESTRICTED, please attach a thesis declaration letter.



SUPERVISOR'S DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the degree of ~~*Doctor of Philosophy/ Master of Engineering/ Master of Science~~ In *PSM*

A handwritten signature in black ink, appearing to read 'Mohammed Fatah'.

(Supervisor's Signature) Full

Name : *Mohammed Fatah*

Position :

Date : *10/1/19*



STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

A handwritten signature in black ink, appearing to be 'Fatin Nurjannah Binti Salauddin', written over a horizontal line.

(Student's Signature)

Full Name : FATIN NURJANNAH BINTI SALAUDDIN

ID Number : CD15052

Date : 09 January 2019

A MULTIDIMENSIONAL DATA DESCRIPTOR MODEL BASED ON FUZZY
MIN MAX NEURAL NETWORK ALGORITHM

FATIN NURJANNAH BINTI SALAUDDIN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
for the award of the degree of
Bachelor of Computer Science (Graphic and Multimedia Technology)

Faculty of Computer System & Software Engineering
UNIVERSITI MALAYSIA PAHANG

MAY 2018

ACKNOWLEDGEMENTS

All praises to Allah S.W.T because of the strength given and his bless as I able to complete this project. I would like to express my sincere appreciation to the individuals that have been involved in helping me to completed this project.

Also, my appreciation and thanks goes to my supervisor, Dr. Mohammed Falah Mohammed for all the guidance, assistance and all the comment that help me to complete this project according the schema and make me understand more about this project.

Last but not least, I would like to thanks my parents, En. Salauddin bin Mohd Hashim and Ainul Mardliah, and also to my siblings that always give me a moral support and motivation for me to finish this project.

ABSTRAK

Tujuan projek ini adalah untuk memperkenalkan pelbagai teknik dan kaedah yang berbeza yang pernah digunakan sebelum ini untuk menganalisis data. Objektif utama adalah untuk membina alat bagi data analisis untuk data yang terdiri dalam pelbagai dimensi. Dalam projek ini, Teknik yang akan digunakan adalah teknik Fuzzy Min Max. Dengan menggunakan method ini, visualisasi data akan memaparkan nilai minimum dan maksimum dalam nilai antara kosong ke satu. Dalam projek ini, pelbagai teknik akan dibandingkan dan menentukan teknik yang paling sesuai. Teknik - teknik yang dibandingkan adalah semua teknik yang ada dalam Neural Network seperti K-Nearest Neighbour Fuzzy Min Max, general Reflex Fuzzy Min Max, dimana untuk mendapatkan idea berkaitan method, algorithm, teknik dan konsep untuk projek dan kajian yang sedia ada. Selain itu, dalam projek ini, tiga sistem yang sedia ada akan dibandingkan dari segi kebaikan dan keburukan setiap sistem, seperti Tableau Public, Qlikview dan IBM InfoSphere Streams. Projek ini dibuat menggunakan Matlab Programming. Oleh itu, dalam projek ini, akan diterangkan lebih lanjut mengapa saya menggunakan Fuzzy Min Max berbanding dengan kaedah yang lain.

ABSTRACT

The purpose of this project is to introduce different techniques and methods that have been used before to analyze the data. The main objective is to build an data analytic tool for a multidimensional dataset. In this project, the technique that will be used is Fuzzy Min Max method. By using this method, the data visualization will displaying the minimum and maximum value in range of zero to one. In this project, it will be compare a few techniques and determine which techniques is the most suitable. The techniques is in the Neural Network which has a few popular techniques such as K-Nearest Neighbour Fuzzy Min Max, general Reflex Fuzzy Min Max to get some idea of methodologies, algorithm, techniques and concept of the whole existing project and research study. Besides that, in this paper, it will also compare three existing system such as Tableau Public, Qlikview and IBM InfoSphere Streams. They have been compare for their advantages and disadvantages. The implementation of Fuzzy Min Max Neural Network technique has been applied using Matlab Programming. Therefore, in this project, it will explain more about why I am using the Fuzzy Min Max method rather than other methods.

TABLE OF CONTENT

DECLARATION	
TITLE PAGE	
ACKNOWLEDGEMENTS	iiiiiii
ABSTRAK	iv
ABSTRACT	v
TABLE OF CONTENT	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xxi
CHAPTER 1 INTRODUCTION	12
1.1 Background	1-3
1.2 Problem Statement	3
1.3 Aim and Objectives	4
1.4 Scope	4
1.5 Significance of this project	4
1.6 Thesis Organization	4-5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Neural Network	6-7

2.2.1	Fuzzy Min Max Neural Network	7-8
2.2.2	K-Nearest Fuzzy Min Max	8-9
2.2.3	General Reflex Fuzz Min Max	9-10
2.3	Summary	11
2.4	Existing system of data analytical tools	11-12
2.4.1	Tableau public	12
2.4.2	Qlikview	12
2.4.3	IBM Infosphere Streams	12
2.5	Summary	13
2.6	Conclusion	14
	CHAPTER 3 METHODOLOGY	15
3.1	Introduction	15
3.2	Methodology	15-16
3.2.1	Analyse Background Study	16
3.2.2	Literature review	16-17
3.3	Fuzzy Min Max Neural Network	17-23
3.4	Context diagram and Use Case diagram	23
3.5	Flowchart and dialogue diagram	24
3.6	Proposed User interface	25

3.7	Hardware and Software	25
	3.7.1 Hardware requirement and specification	25
	3.7.2 Software requirement and specification	26-27
3.8	Gantt Chart	27
3.9	Conclusion	27
CHAPTER 4 RESULT AND DISCUSSION		
4.1	Introduction	28
4.2	Result Discussion	29
CHAPTER 5 CONCLUSION		
5.1	Introduction	30
5.2	Research Constraints	30-31
5.3	Future Work	31
REFERENCES		32-34
APPENDIX A SAMPLE APPENDIX 1		35

LIST OF TABLES

Table No	Table Title	Page Number
2.1	Advantages and Disadvantages	11
2.2	Advantages and Disadvantages	13
3.1	Hardware Requirements and Specifications	26
3.2	Software Requirements and Specifications	26

LIST OF FIGURES

Figure number	Figure title	Page number
Figure 2.1	General Reflex FMM	Page 10
Figure 3.1	Research Methodology	Page 14
Figure 3.2	FMM learning algorithm	Page 19
Figure 3.3	Example of expansion problem	Page 21
Figure 3.4	Hyperbox overlap test	Page 21
Figure 3.5	FMM pseudocode	Page 22
Figure 3.6	EFMM pseudocode	Page 23
Figure 3.7	Context Diagram for data analytic tool	Page 23
Figure 3.8	Use Case for data analytic tool	Page 24
Figure 3.9	Flowchart	Page 24
Figure 3.10	Dialogue Diagram	Page 25

LIST OF ABBREVIATIONS

Abbreviation	Title
ANN	ARTIFICIAL NEURAL NETWORK
FMM	FUZZY MIN MAX
EFMM	ENHANCED FUZZY MIN MAX
KNFMM	K-NEAREST FUZZY MIN MAX
NN	NEURAL NETWORK

CHAPTER 1

INTRODUCTION

1.1 Background

As the world are developing more and more, big companies are getting success and success, where the increasing amount of people means there are many data and information to be analyze. In fact, data analyses have been used in different areas of life like in finance, science and research, and business insight [1]. That analysis covering different kind of datasets including supervised, and unsupervised data sets [2]. Here and after, the focus will be on the supervised data set as one of the most important data in classification process.

In fact, analysing could help to diagnosis the limitation or the strength of the data and that could lead to improve the system performance. The data analysing are been use in many area such as medical, education, business and extra [3]. There are many data analytics tools that have been used in that area, for example, for the medical area, the most popular is the Hadoop (Apache platform) which is the open source distributed data processing. Initially, Hadoop was developed for the routine functions. Hadoop have been used widely for process extremely large amount of dataset and also can serve as two roles which are as analytical tools and data organizer. Even though Hadoop has many advantages, however person with Hadoop skill are rarely to be found due to the difficulty of installation and configuration processes [4].

Beside Hadoop, there are other tools that been used in others area such as Tableau and IBM. Tableau tool is useful for the data visualization process. It is very easy to be used even if the user doesn't have a strong background skill about it. Tableau using the simple drag and drop method which allow user to directly interact,

visualizing data, controls the analyses process, and generate the final report [5]. Although Tableau has many advantages, it is not completely stable due to its open source functionality. In case of IBM, it design to support IBM's global campaign for Business Analytics and Optimization (BAO). IBM has three popular platforms that supposed to support a real-time analytics processing (RTAP), such as, IBM's acquisition of Netezza which introduced in 2010, IBM InfoSphere BigInSights in 2011, and IBM InfoSphere Streams [6]. All these three IBM products can define the modern analytics, include built-in text analytics and provide the velocity for the big data on structured and unstructured data. But, not all tools can cooperate with the big data volumes at one time, then it will not support the specific analytic modelling if users need it, added to that, processing the data can be very slow [6]. Considering all these problems, there is a need to use a specific tool that could provide an easy and clear understanding for the nature of data in different domains. Hear, the focus will be on the classification domain. In order to keep focus, here and after, the focus will be on classification techniques.

In fact, analysing data could help to improve the classification performance for neural network, by avoiding some of the existing limitations and produce better classification accuracy [7]. Classification is one of the three major components for data mining. It is the technique that used to classify each item from a set of data into a predefined class. Classification are being used widely as it is a major technique in Data Mining, where it used in various fields such as historical credit rating, employment history, number and types of investment [8], and extra.

In fact, there are many techniques that have been designed to handle different pattern classification problems, such as, decision tree which is a simple technique comparing to other but it straightforward applied to solve different classification problems [9, 10]. Another techniques example are the rule-based classifier [11], support vector machines [12], naïve Bayes classifier [13], and the artificial neural networks which mainly been used in various field such as forecasting, data compression and extra for used in engineering purposes. [14].

Even though there were many attempts to enhance the classification performance by proposing different techniques, however, the door is still open to further enhance the prediction performance. One of the ways that could really help to improve classifiers performance could be achieved through data analysing process. This kind of process normally helps researchers to understand the nature of data, identify the limitations/drawbacks, then proposed a suitable solution in order to overcome the existing limitations. Based on that, many application and tools have been proposed for data analysis, for example, Artificial Neural Network (ANN) which has emerged as one of the popular methods in tackling pattern classification problems. ANNs are useful for handling noisy data collected from real environments. Even though the number of ANN model that have been developed from the first neural model are high, the focus of this thesis is on one of the latest proposed ANN for patterns classification, which is the Fuzzy Min-Max (FMM) neural network, in order to develop a useful pattern classification model. Based on that, we are going to propose a classification. Further discussion about the FMM neural network will be introduced in chapter 2.

1.2 Problem Statement

As highlighted in the introduction, there are many data analytic tool where a user can used freely according to their preference, however, all the data analytic have their own limitations or drawbacks that leads to the new tool. Therefore, in this research, it is focusing to develop a new data analytic tool that can improve the available tool for the visualization, which by using the Fuzzy Min Max technique. The tool will display the data in the range of 0 to 1. It will also display the overlap area between the classes of data.

1.3 Aim and Objectives

The aim of this project is to provide better understanding about the data relationship through proposing a new data analytical tool. Hence, the objectives of this research are:

- To study and analysis some of the current efficient data analyses techniques to identify the limitations/drawbacks.
- To propose a new tool that can visualize the relationship of a multidimensional supervised dataset.
- To test and validate the efficiency of the proposed tools using different benchmark/real-world dataset.

1.4 Scope

This study focus on supervised dataset and visualize the data relationship in a multidimensional way. That could help researchers to identify the problem/limitations and help them propose the suitable method/technique to overcome that limitations. In this study, we will use the Matlab to create a new data analytical tool.

1.5 Significance of this project

To make this project more impactful, there are some benefits on it. This study providing significances as following:

- i. This project focused on how to make a visualization for the user, so that it will make their work easier.
- ii. It is the new data analytical tools to help the big company to visualize their data in range of 0 to 1 and know which area is overlapped.

1.6 Thesis Organization

This research paper consists of six (5) chapters:

Chapter 1 discussed about the introduction of the research study, problem statement of data analysis for supervised dataset, objective, scope and significance for this project.

Chapter 2 will discuss on the literature review of research. The purpose of the review is to get the idea of the suitable method or technique that can be used to build the data analytics tool.

Chapter 3 will discuss the methodology. The purpose is to discuss what methodology will be used. This chapter will explain details of fuzzy min-max technique that will be used. Then will justify the hardware and software specification that will be used.

Chapter 4 is about result discussion. This chapter will discuss the result of the visualization for multidimensional dataset using the new tool.

Chapter 5 is about conclusion. This chapter concludes the project that need consists of future suggestion and enhancement of project to the next research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This literature review is for getting the information of the existing system for data analytical that have been used by the other researchers based on the same title. In order to learn about the Fuzzy Min Max neural network to develop a data analytic for multidimensional, which to get the idea, algorithm and technique used.

The main title of this paper is Fuzzy Min-Max Neural Network based on analytical for supervised dataset. The main aim stated is to improve the accuracy of analytical for multidimension supervised dataset.

The main objective is to study the fuzzy min-max method for analytical for multidimension supervised dataset. Next is to implement fuzzy min-max methodology by using multidimension dataset and lastly is to test and validate the model.

2.2 Neural Network

Neural Network (NN) is the recent study for the development. An artificial Neural Network (NN) is an information processing standard. It is made by a large number of neuron which working in unison to solve the problems. Their speciality is to derive the complicated and imperfect database.[15].

2.2.1 Fuzzy Min Max Neural Network

The supervised FMM network [16] is made by a fuzzy set of hyperbox. This hyperbox can be defined as the area of n-dimensional which has pattern of full class membership.

There are the properties that motivated a portion of the development of the FMM classification as follows

- On-line Adaptation: It should be able to learn and refine existing and new classes quickly and without destroying old class information. This property is important to solve the stability plasticity dilemma.
- Off-line Adaptation: It can place a huge demand on the memory requirement and lead to longer training times.
- Nonlinear separability: The ability to build decision regions to separate classes of any size and shape.
- Overlapping classes: The ability to form the decision boundary to minimizes misclassification by eliminating the overlap region between different classes.
- Training time: The ability the decision boundaries of the classes within a short time training.
- Soft and hard decision: The ability to provide the soft and hard classification decision.
- Verification and Validation: It is important to have a mechanism to verifying and validate its performance.
- Turning Parameter: Should have a few parameters that help to tune in the system as possible.
- Nonparametric Classification: Assume the 'p priori' knowledge which about underlying probability density functions of each class.

In FMM, a data set are consist of input pattern and target class, Y_h , $h= 1, \dots, N$, where N is the total number of data samples. Then, depend on the data, FMM will

creates the hyperboxes, which represent the minimum (V_j) and maximum point (w_j). Its size is depend on the expansion coefficient (γ).

Each hyperbox fuzzy set is defined by the ordered set:

$$B_j = (\{Y_h, V_j, W_j, f((Y_h, V_j, W_j))\}) \quad \forall A_h \in I^n$$

(1)

Where:

- B_j is the hyperbox fuzzy set
- $A_h = (a_{h1}, a_{h2}, \dots, a_{hn}) \in I^n$ is the input pattern,
- $v_j = (v_{j1}, v_{j2}, \dots, v_{jn})$ is the minimum points of B_j .
- $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})$ is the maximum points of B_j .

Generally, the membership function is used which to find the closest hyperboxes to the data sample. It is calculated using the formula as follows:

$$B_j(A_h) = \frac{1}{2n} \sum_{i=1}^n [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - w_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, v_{ji} - a_{hi})))]$$

(2)(Simpson, P. K. (1992). Fuzzy Min-Max Neural Networks-Part, 3(5).)

2.2.2 K- Nearest Fuzzy Min Max (KNFMM)

A K-nearest hyperbox expansion rule Enhanced Fuzzy Min-Max (KnFMM) was developed by M.F. Mohammed and C.P. Lim [17]. When the size is set to 0, the Fuzzy Min Max will be related to the k-nearest expansion. This expansion of the point

will have a little loss in its performance but also have the computational advantages. [18].

The winning hyperbox is explained as the hyperbox with the highest membership function that will be used for the input sample to produce the expected target class. Besides, it is the exclusive hyperbox used for the expansion process. When the expansion criteria are not gratified, a new hyperbox will be formed. For the hyperbox expansion process, in the original Fuzzy Min-Max, it will only choose one winning hyperbox from the set of hyperbox, while in K-nearest hyperbox which has the similar classification will be choosing to define the final winning hyperbox. For example; initially, when the dimension of the hyperbox are correlated against the expansion, the winning hyperbox is chosen by using the inequality to determine that. If there is any problem with the winning hyperbox, the next closer one will go to the investigating step. The chosen one for the expansion is the first K-nearest hyperbox that meet all the inequality. The new hyperbox is created to conceal the input sample data if there is none of the K-nearest hyperboxes able to meet the inequality. K-nearest is able to avoid creation to many small hyperbox within the winning hyperbox during the learning stages.

2.2.3 General Reflex Fuzzy Min Max

A General Reflex Fuzzy Min Max Neural Network is proposed by Gabrys and Bargiela [19]. It is a set of fuzzy hyperbox based on supervised, unsupervised and semi-supervised learning algorithm. It is to learn the mixture of labelled and unlabelled data. General Reflex FMM is capable to extract the underlying structure of data and also a mix of FMM classification and clustering algorithm. It learns the data by creation and expansion of hyperbox and the contraction process avoids membership ambiguity n overlapping area. However, for the labelled data it avoids the use of contraction process which overcome the containment and overlapping issues [19].

If the membership of hyperbox is more than zero and average expansion is less than a user specified limit, it allow the expansion of a hyperbox to accommodate a training point. Hyperbox expansion is when the network tries to learn applied labelled in one of the existing hyperboxes of the same class or label [19].

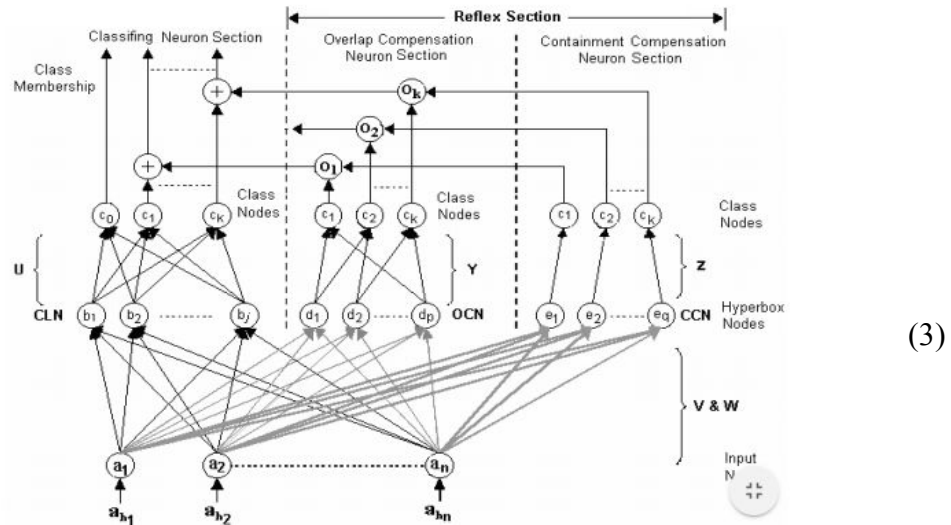


Figure 2.1 General Reflex FMM (Nandedkar, A. V, & Biswas, P. K. (2017). A general reflex fuzzy min-max neural network, *14*(2007), 2017)

Figure above is the General Reflex FMM Architecture where

- $a_{h1} - a_{hn}$: Input data samples
- $a_j - a_n$: Input nodes
- $b_1 - b_j$: Classification hyperbox nodes
- $C_0 - C_k$: Class nodes
- $d_1 - d_p$: Overlap compensation hyperbox nodes
- $e_1 - e_q$: Containment compensation hyperbox nodes
- $o_0 - o_k$: Overall compensation nodes

2.3 Summary of existing techniques for classification

	Fuzzy Min- Max (FMM) [16]	K-nearest Neighbor Fuzzy Min Max (KNFMM) [17][18]	General Reflex Fuzzy Min Max Neural Network (GRFMM) [19]
Advantages	<p>Cater for supervised learning combined ANN and fuzzy set theory into common framework for pattern classification problem.</p> <p>Avoid re-training</p>	<p>Simple to implement</p> <p>Naturally handles multi-class cases</p>	<p>Can handle the overlapping and containment problem of hyperboxes.</p> <p>Approximate the complex topology of dataset more better.</p>
Disadvantages		<p>Large search problem to find nearest neighbors.</p> <p>Storage of data</p>	

Table 2.1 Advantages and Disadvantages

2.4 Existing System of Data Analytics Tools

Data analytics is the process of observing the datasets to make a conclusion of the data with the help of the software and specialized system. Data analytics are being used

widely in all area in the industries to make more suitable decision. There are many examples of existing system that have been implemented to help the user make the visualization easier.

2.4.1 Tableau public

Tableau is the most popular and leading the other tools which focus on data visualization, data discovery and also dashboard. There are two types of tableau which are tableau public desktop and tableau public server. You do not need to have the skill to use Tableau as you can use it easily. It can help the user to make a decision at all levels [20].

2.4.2 Qlikview

Qlikview is an in-memory, for business discovery tools. It is very benefit for the business that help the organization whether it is big or small. Qlikview has three main component that make the tools function, which are as follow [21]:

- i. Qlikview desktop: it is a tool to create the graphical user interface or dashboard visualizations. It is also help for data extract and the transformation model.
- ii. Qlikview server: The application is stored on it special server which is Qlikview server. It is use to handle the communications that happens between the user and the application.
- iii. Qlikview publisher: It will loads the various data and send it to the user.

2.4.3 IBM InfoSphere Streams

IBM InfoSphere Streams is a software that enables the development execution of the tools which process the data. It provide the faster analysis of data and continuous that help to improve the speed for making decision. InfoSphere Streams contains many component such as resources, domains, and instances [22].

2.5 Summary of existing system

	Tableau Public	Qlikview	IBM InfoSphere Streams
Advantages	It can support the complex data, create the interactive visualization without need to wait longer.	It is a successful company as it is always released the new and up to date software, therefore, the user has no problem to use it.	It helps the user to analyse the big data at a faster speed and also faster in helping the user in making decision.
Disadvantages	It has no custom visual import. If there are a new visualization, it is needed to create again instead of importing.	The end user cannot create the dashboard by themselves without the trained developer and it is a very high cost software.	There is no automated error handling and recovery mechanism.

Table 2.2 Advantages and Disadvantages

2.6 Conclusion

In conclusion, there are many types of techniques and method has been proposed by previous researchers that can applied to produce a better classification accuracy performance of analytical for multidimension supervised dataset by using fuzzy min-max neural network. By using all these techniques, there are a lot of various tool that can be used for data analytical, but every of this tool, have their own limitation that leads to the development of new tools that are much better. Based on the literature review the most suitable model that can improve the data analytical tool is by using the FMM technique.

CHAPTER 3

METHODOLOGY

3.1 Introduction

In this chapter, Fuzzy Min-Max Neural Network method are being introduced to develop the pattern classification model. In this project is to propose a new data analytical tool with supporting ability for a multidimensional visualization. We will start with the methodology that have been chosen followed by the explanation for background of study, literature review, collection of data, hardware and software used and lastly the conclusion.

3.2 Methodology

Figure 3.1 shows the research methodology for a multidimensional supervised dataset. This research follows the following process and steps to obtained expected result.

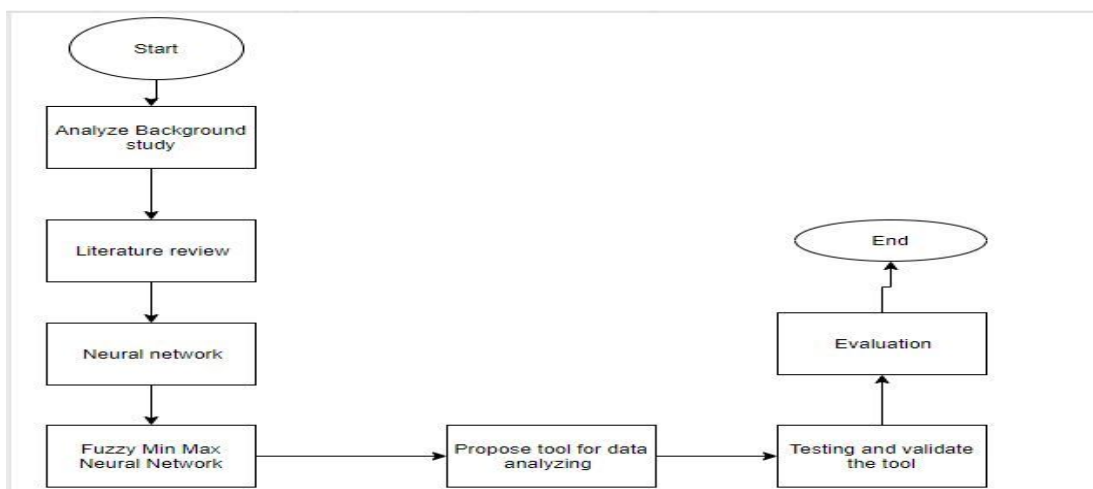


Figure 3.1: Research methodology for a multidimensional supervised dataset

3.2.1 Analyse Background Study

i. Research on problem background

It is a study about the tool that can help the user to get the visualization of their data to make their work easier. Then research on what method are more suitable to be propose for new data analytical tool for multidimensional dataset. In addition, focused problem study on the big company that will make a deep analyse and review of their data with many attributes from the data. So, without the suitable and easy visualization, it will take a longer time to analyse and get the wanted result.

ii. Determine problem study

This research paper is to develop a new data analytic application for multidimensional dataset based on FMM method.

iii. Determine objective and scope

After defining the problem on study, the objective for data analytical tool is defined. The objective for the application should be finish within a time frame and should be considered available resources. Based on the needs of the project as establishment of project timeline and project goals, the scope of this project id. It defined the goals to achieve the result.

3.2.2 Literature Review

Based on the review that has been conducted from the previous paper, the FMM model has proposed the technique on how to develop a tool. There are several other

technique that can be used to develop the tool such as data mining, fuzzy logic and others, however, in this project, Fuzzy Min Max are been chosen.

3.3 Fuzzy Min Max Neural Network

The supervised FMM network [23] has explained the dynamic network structure with the online learning capability. The number of hyperbox can be increased when it is necessary.

There are several properties that associated with FMM that motivated a portion for the development of FMM neural network. The properties as follow:

- i. Online Adaptation: This property are also being called as online learning. It able to learn the new classes of data and refine it quickly without losing the old information of classes.
- ii. Nonlinear separability: Able to build the decision regions for classes that separate any shape or size.
- iii. Overlapping classes: able to build a decision boundary to minimize the misclassification for all classes that overlapping.
- iv. Training time: Able to learn the decision boundaries of nonlinear in the short time.
- v. Soft and hard decision: Able to prepare hard and soft classification decision. Hard decision is either 0 or 1, while the soft decision gives a value to describe the degree that a pattern fits for the class.
- vi. Verification and validation: It's a important mechanism for verifying and validating the performance. To been able for performing this function, closed form solution, Contour plots and scatter plots are being used.

The three steps in FMM are known as hyperbox overlap test, hyperbox contraction and the hyperbox expansion [24]. It will start with a data that have input and target classes. Then, depend on the data, FMM will create the hyperboxes, which represent the minimum (v_j) and maximum point (w_j). Its size is depend on the expansion coefficient (γ).

Each hyperbox fuzzy set is defined by the ordered set:

$$B_j = \{(A_h, v_j, w_j, f((A_h, v_j, w_j)))\} \quad \forall A_h \in I^n \quad (8)$$

Where :

- B_j is the hyperbox fuzzy set
- $A_h = (a_{h_1}, a_{h_2}, \dots, a_{h_n}) \in I^n$ is the input pattern,
- $v_j = (v_{j_1}, v_{j_2}, \dots, v_{j_n})$ is the minimum points of B_j .
- $w_j = (w_{j_1}, w_{j_2}, \dots, w_{j_n})$ is the maximum points of B_j .

Generally, the membership function is used which to find the closest hyperboxes to the data sample. It is calculated using the formula as follows:

$$B_j(A_h) = \frac{1}{2n} \sum_{i=1}^n [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - w_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, v_{ji} - a_{hi})))] \quad (9)$$

a) Hyperbox Expansion, Overlap Test and Contraction

When the hyperbox expansion is performed, the input pattern will be included where the class of hyperbox must not exceed the expansion coefficient. To include the new pattern, the following constraint must be follow [26].

$$B_j(A_h) = \frac{1}{2n} \sum_{i=1}^n [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - w_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, v_{ji} - a_{hi})))] \quad (10)$$

The new hyperbox will be created if the constraint is not satisfied to encode the pattern. However, when the new hyperbox is created, it can lead to overlapping between hyperboxes. Therefore, the overlap test will take over to check the area that overlapping is occurred after the expansion process between expanded hyperbox and the existing hyperbox are belong to other classes. Four cases are being test for the overlapping test. It is considered that there exist overlap between two hyperbox if one of the cases are satisfied [25].

<p><i>Case 1:</i></p> $V_{ji} < V_{ki} < W_{ji} < W_{ki}, \quad \delta^{\text{new}} = \min(W_{ji} - V_{ki}, \delta^{\text{old}}).$ <p><i>Case 2:</i></p> $V_{ki} < V_{ji} < W_{ki} < W_{ji}, \quad \delta^{\text{new}} = \min(W_{ki} - V_{ji}, \delta^{\text{old}}).$ <p><i>Case 3:</i></p> $V_{ji} < V_{ki} < W_{ki} < W_{ji}$ $\delta^{\text{new}} = \min(\min(W_{ki} - V_{ji}, W_{ji} - V_{ki}), \delta^{\text{old}}).$ <p><i>Case 4:</i></p> $V_{ki} < V_{ji} < W_{ji} < W_{ki}$ $\delta^{\text{new}} = \min(\min(W_{ji} - V_{ki}, W_{ki} - V_{ji}), \delta^{\text{old}}).$

(11)(Simpson, P. K. (1992). Fuzzy Min-Max Neural Networks-Part, 3(5).)

Next, the contraction of hyperbox will begin to eliminate the area of overlapping when the hyperbox from the different classes are overlapping. However, if the overlapping are is from the other classes, it is allowed. To identify whether the contraction step is necessary, it will continue to check the next step. In contraction process, the hyperbox size is kept large as possible by adjusting one of the dimensions in overlapped hyperbox

Hence, the overlapping regions are eliminated by adjusting each hyperbox. For the proper adjustment, the same four cases are examined.

Case 1:
 $V_{j\Delta} < V_{k\Delta} < W_{j\Delta} < W_{k\Delta}, W_{j\Delta}^{new} = V_{k\Delta}^{new} = \frac{W_{j\Delta}^{old} + V_{k\Delta}^{old}}{2}$

Case 2:
 $V_{k\Delta} < V_{j\Delta} < W_{k\Delta} < W_{j\Delta}, W_{k\Delta}^{new} = V_{j\Delta}^{new} = \frac{W_{k\Delta}^{old} + V_{j\Delta}^{old}}{2}$

Case 3a:
 $V_{j\Delta} < V_{k\Delta} < W_{k\Delta} < W_{j\Delta}$ and $(W_{k\Delta} - V_{j\Delta}) < (W_{j\Delta} - V_{k\Delta}), V_{j\Delta}^{new} = W_{k\Delta}^{old}$

Case 3b:
 $V_{j\Delta} < V_{k\Delta} < W_{j\Delta} < W_{k\Delta}$ and $(W_{k\Delta} - V_{j\Delta}) > (W_{j\Delta} - V_{k\Delta}), W_{j\Delta}^{new} = V_{k\Delta}^{old}$

Case 4a:
 $V_{k\Delta} < V_{j\Delta} < W_{j\Delta} < W_{k\Delta}$ and $(W_{k\Delta} - V_{j\Delta}) < (W_{j\Delta} - V_{k\Delta}), W_{k\Delta}^{new} = V_{j\Delta}^{old}$

Case 4b:
 $V_{k\Delta} < V_{j\Delta} < W_{j\Delta} < W_{k\Delta}$ and $(W_{k\Delta} - V_{j\Delta}) > (W_{j\Delta} - V_{k\Delta}), W_{k\Delta}^{new} = W_{j\Delta}^{old}$

(12) (Simpson, P. K. (1992). Fuzzy Min-Max Neural Networks-Part, 3(5).)

b) FMM learning algorithm

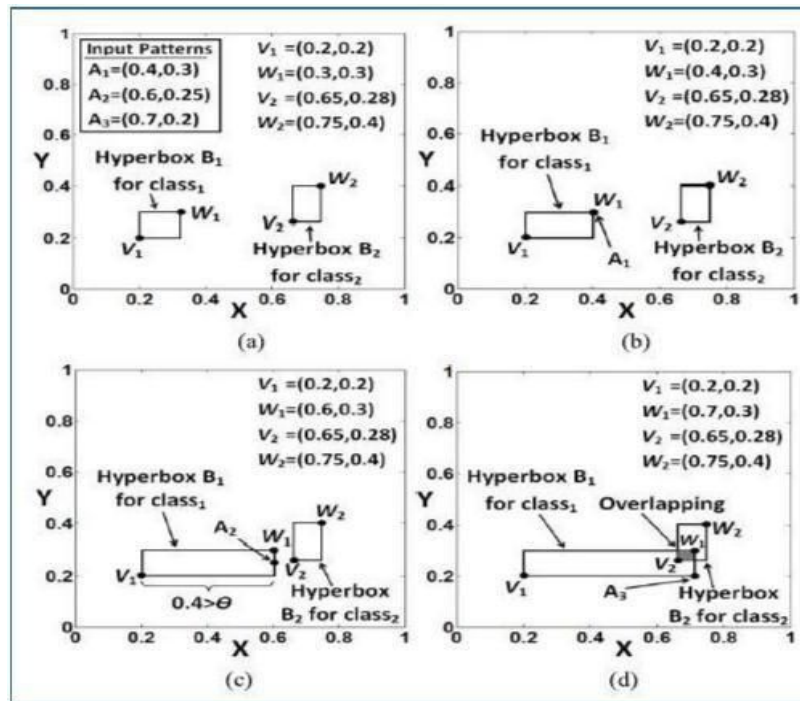


Figure 3.2: FMM learning algorithms (Mohammed, M. F., & Chee Peng Lim. (2015). An Enhanced Fuzzy Min–Max Neural Network for Pattern

Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 26(3), 417–429. <https://doi.org/10.1109/TNNLS.2014.2315214>)

The learning process is online which refer to the ability of learning model to create a new class and refined the existing classes without affecting any information in the model. This allow FMM to add a new class and refine the existing class to avoid retraining. This is to tackle the stability plasticity dilemma [25]. Stability plasticity dilemma is a process where the learning model is able to able new information.

```
1  Training
2  For the first to the last input vectors ( $A_i$ )
3      If there is any hyperbox
4          Determine the membership values (Eq. 1), checking all dimensions (M) for all hyperboxes (N)
5              Winner= the hyperbox with highest membership value
6          If the input vector does not fit into the winning hyperbox
7              Determine the expansion condition of the winning hyperbox (Eq. 2)
8                  If the winning hyperbox is able to expand
9                      Do the expansion process to include the input vector
10                     Then perform the hyperbox overlap test
11                     If all dimensions (M) are overlapped
12                         Do the contraction process to contract the dimension with the minimum
13                         overlapped space
14                     End If
15                 Else create a new hyperbox
16                     New hyperbox= the input vector (A)
17                 End If
18             Else winner= the current hyperbox with a full membership value
19             End If
20         Else create a new hyperbox
21             New hyperbox= the input vector (A)
22         End If
23     End For
```

Figure 3.5: FMM pseudocode

```

1  Training using k-nearest neighbors selection rule
2  For the first to the last input vectors ( $A_i$ )
3      If there is any hyperbox
4          Determine the membership values (Eq. 1), checking all dimensions (M) for all hyperboxes (N)
6          For the selected k number of hyperboxes that have highest membership values from the same
7              target class as the training sample
8                  If the input vector does not fit into the selected hyperbox
9                      Determine the expansion condition of the selected hyperbox (Eq. 2)
10                     If the selected hyperbox is able to expand
11                         Winner= the current hyperbox
12                         Do the expansion process to include the input vector
13                         Then perform the hyperbox overlap test
14                         If any dimension (M) is overlapped
15                             Do the contraction process to contract all overlapped dimensions
16                         End If
17                     Break For
18                 End If
19                 Else winner= the current hyperbox with a full membership value
20                 Break For
21             End If
22         End For
23         If none of the selected hyperbox is able to expand
24             Create a new hyperbox where new hyperbox= the input vector (A)
25         End If
26     Else create a new hyperbox
27         New hyperbox= the input vector (A)
28     End If
29 End For
30
31 Pruning using prediction set
32 For the first to the last hyperbox (N)
33     Determine the confidence factor of each hyperbox, i.e,  $CF_{kj}$  (Eq. 4)
34     Remove all hyperboxes with  $CF_{kj}$  lower than a pruning threshold ( $CF_{kj} < \delta$ )
35 End For

```

Figure 3.6: EFMM pseudocode

In figure 3.5 and figure 3.6 are the pseudocode of algorithms for FMM and EFMM. Both algorithms are being used for the implementation of FMM and EFMM from previous research paper.

3.4 Context Diagram and Use Case Diagram

Data analytical tool are developed in order to enhance the existing data analytical tool. This system can be accessed by any user through online.

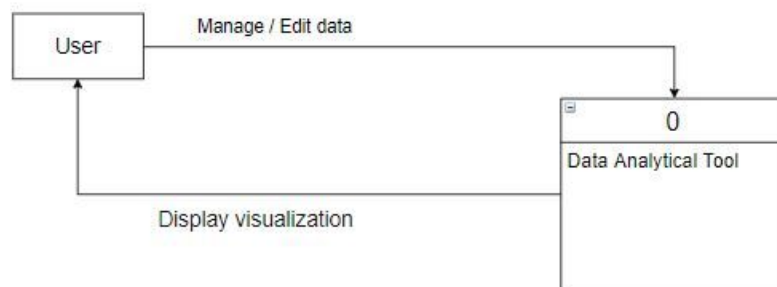


Figure 3.7: Context diagram for data analytical tool

The data analytical tool will display the visualization of the data when the user run the data analytic tool.

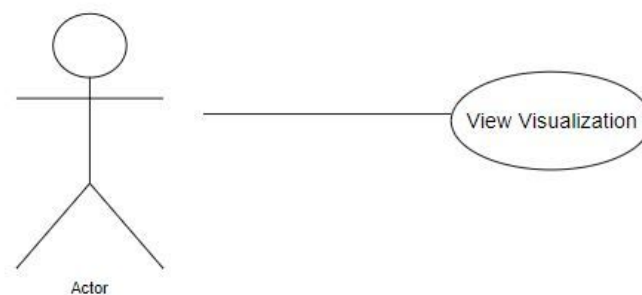


Figure 3.8: Use case Diagram for data analytical tool

When the user run the tool, the tool will display the data in the graph with the range of 0 to 1. The data will display the minimum and maximum value of each dataset.

3.5 Flowchart and Dialogue diagram

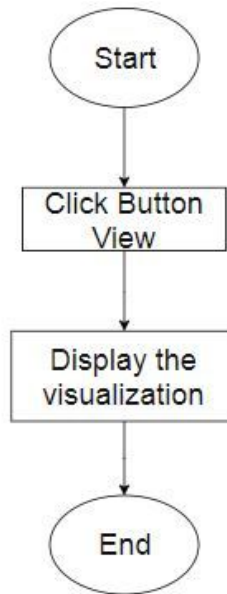


Figure 3.9: Flowchart of the data analytical tool

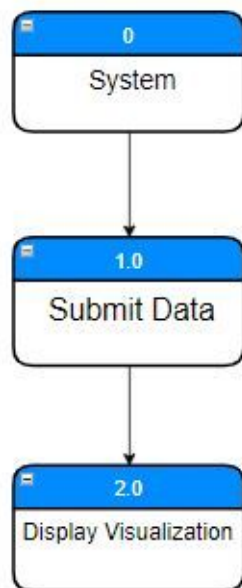
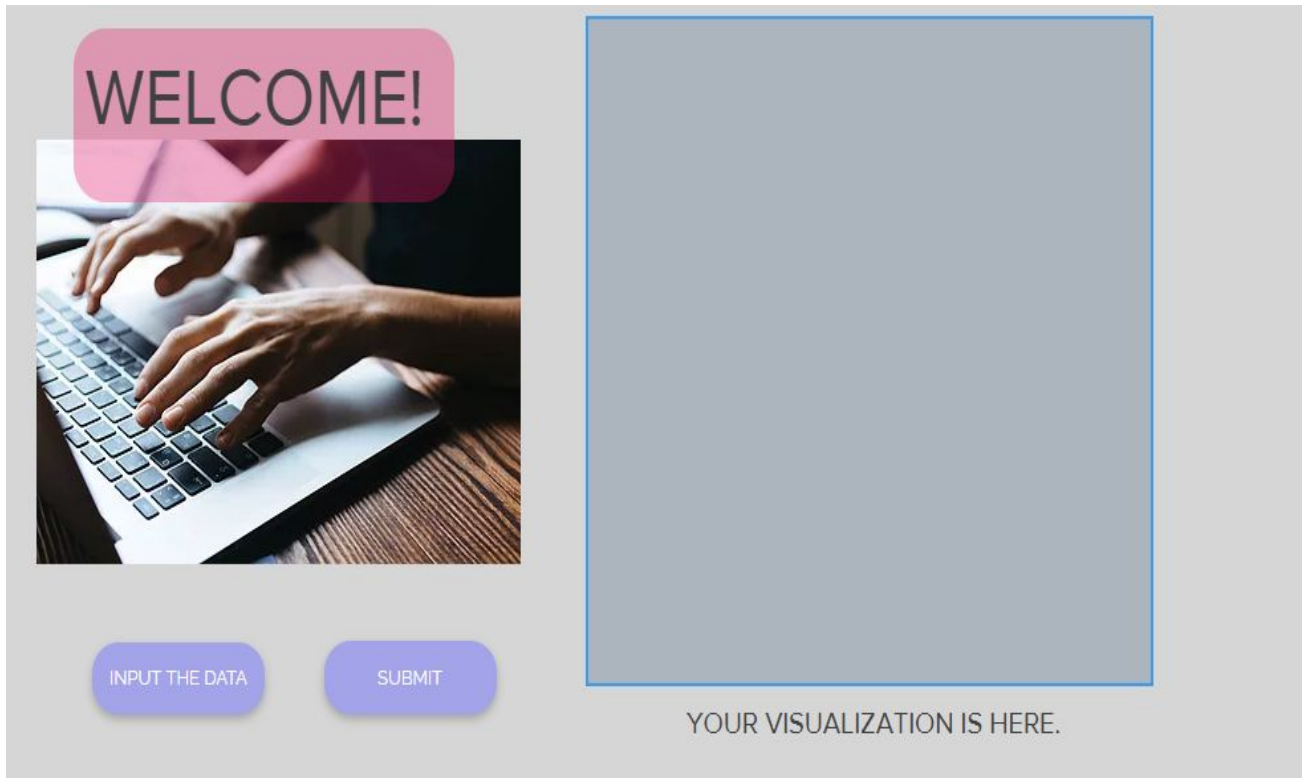


Figure 3.10: Dialogue Diagram of data analytical tool

3.6 Proposed User Interface



3.7 Hardware and Software

Hardware and software requirement are defined based on the needed during the developing the data analytical tool. The software and hardware specification are shown below:

3.7.1 Hardware Requirements and Specification

Hardware	Purpose
Personal Computer	Device to develop and create the system

Table 3.1: Hardware Requirements and Specifications

Table 3.1 shows the hardware and specifications used in the training of classification stage. Personal computer being used is ASUS.

3.7.2 Software Requirements and Specification

Software	Purpose
Windows 8	Platform of operating system
Microsoft Word 2016	Used for documentation
Microsoft Project 2010	Create Gantt Chart
Draw.io	To create the context diagram, flow chart, use case diagram and dialogue diagram
Wix	To design the proposed user interface
Matlab	Build the application using the Fuzzy Min Max coding

Table 3.2: Software Requirements and Specifications

Table 3.2 shows the software used during training phase for developing the tool. Each of the software has important roles from the beginning of developing process.

Microsoft Office Word 2016 is used for writing project documents from chapter 1 until chapter 5. This software is used because it contains more function that can be edited such as its margin, page number, type and size font and also line spacing.

Microsoft Project 2010 is used because it provides function to create a timeline to produce a Gantt chart.

Draw.io is used because it can easily create the context, use case diagram, flowchart and dialogue diagram without any problem.

Wix is used to create the interface as it is an application that allow the user to create the interfaces based on their needs.

MATLAB is used to train the dataset in order to give result of the prediction of the neural network.

3.8 Gantt Chart

Gantt Chart is a type of chart that illustrates a project schedule. Gantt chart illustrated the start and finish dates of each task and summary of element of project. Gantt chart of this research is in appendix A.

3.9 Conclusion

In conclusion, this chapter discussed the methodology of research in details. A method is chosen to be implemented in the project. Flow chart is included to show the flow of the project step by step.

CHAPTER 4

RESULT AND DISCUSSION

4.1 INTRODUCTION

For chapter 4, the implementation of Fuzzy Min Max Neural Network technique has been applied by using Matlab Programming is discussed. The technique is use to develop the data analytic tool which been built using Matlab Programming to produce the accurate result by displaying the data in the range of 0 to 1. The main purpose of this chapter is to discussing about the tools and result by that tool that been developed using the Fuzzy Min Max Neural Network technique.

4.2 RESULT DISCUSSION

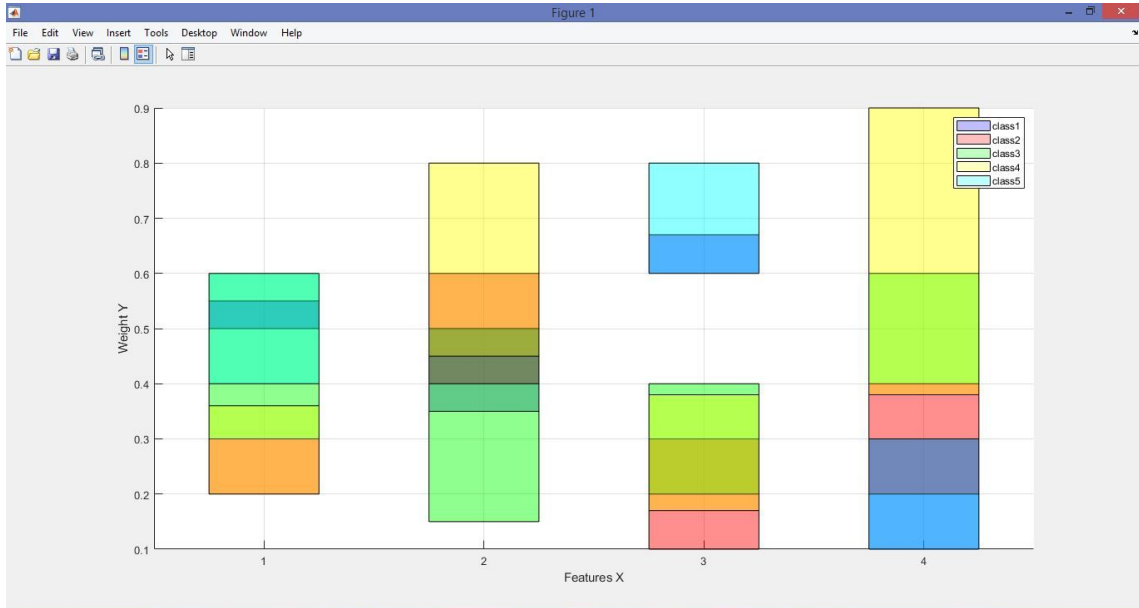


Figure 4.1 Result of data visualization

From the figure above, we can see that there are 5 hyperboxes being visualized which each hyperbox are being represent by different colour. From the visualization, we also can see the overlapping area clearly that show the overlap between several hyperboxes. Each hyperbox are consist of four features / dimensions, each minimum and maximum value are being plotted.

CHAPTER 5

CONCLUSION

5.1 INTRODUCTION

For this chapter, which is the last chapter, it summarizes all the content of this project including the summary of problem statement, literature review, methodology and also implementation phase and result obtained.

In order to make the accurate visualization, a suitable method or system must be evaluated. In literature review, there are many free data analyze tool that the user can use anytime and also many technique can be used for the classification, however all this system and techniques has their constraint. Therefore, in this paper, we used Fuzzy Min Max Neural Network technique as the neural networks is popular for their accuracy in result.

Based on the problem statement stated above, this project paper achieved the objectives to develop the tool. The objectives consist of to study and analysis some of the current efficient data analyses techniques to identify the limitations/drawbacks, to propose a new tool that can visualize the relationship of a multidimensional supervised dataset and to test and validate the efficiency of the proposed tools using different benchmark/real-world dataset.

5.2 RESEARCH CONSTRAINTS

i. Limited time

Student always has a limited time to completed their project and report. This is because they spend a lot of time on creating the coding for the system and test it until they satisfied the objectives of project. It takes time to understand on how the system

work and how the coding can be created so that the system can successfully be run. This problem can be overcome if the student considered the time limit is their first priority and follow the deadline for each phase.

5.3 FUTURE WORK

Neural Network has proven to be chosen as a method to make a visualization data tool based on Fuzzy Min Max technique. The future work that can be implemented is to using the Fuzzy Min Max to make data analyze tool that the classes are not limited and can be implemented in any platform.

REFERENCES

- [1] <https://www.digitalvidya.com/blog/data-analytics-applications/> (online)
- [2] Koturwar, P., Girase, S. and, & Debajyoti, M. (2014). A Survey of Classification Techniques in the Area of Big Data. *International Journal of Advance Foundation and Research in Compute*, 1(11), 1–7.
- [3] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.
- [4] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
<https://doi.org/10.1186/2047-2501-2-3>
- [5] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.
- [6] Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter, 19(4)*, 1–34.
- [7] Mohammed, M. F., & Chee Peng Lim. (2015). An Enhanced Fuzzy Min–Max Neural Network for Pattern Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 26(3), 417–429.
<https://doi.org/10.1109/TNNLS.2014.2315214>
- [8] <https://www.digitalvidya.com/blog/data-analytics-applications/> (online)
- [9] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.
- [11] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts , Decision Trees , and.

[11] Data Mining – Rule Based Classification (online)

https://www.tutorialspoint.com/data_mining/dm_rbc.htm

[12] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts ,
Decision Trees , and.

[13] Concepts, B., Trees, D., & Evaluation, M. (n.d.). Classification : Basic Concepts ,
Decision Trees , and.

[14] Kesavaraj, G., & Sukumaran, S. (2013). A Study On Classification Techniques in
Data Mining.

[15] Christos Stergiou and Dimitrios Siganos (2015). Neural Networks (Online).

https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction
to neural networks

[16] [25] Simpson, P. K. (1992). Fuzzy Min-Max Neural Networks-Part, 3(5).

[17] [23] [26] Mohammed, Mohammed Falah, and Chee Peng Lim. "An Enhanced
Fuzzy Min & Max Neural Network for Pattern Classification". 2013 IEEE Transactions
on Neural Networks and Learning Systems 26.3 (2015): 417-429. Web

[18] [24] Mohammed, Mohammed Falah, and Chee Peng Lim. "Improving the Fuzzy
Min-Max Neural Network with A K-Nearest Hyperbox Expansion Rule for Pattern
Classification". Applied Soft Computing 52 (2017): 135-145. Web.

















[19] Nandedkar, A. V, & Biswas, P. K. (2017). A general reflex fuzzy min-max neural network, *14*(2007), 2017.

[20] <https://intellipaat.com/blog/what-is-tableau/> (online)

[21] <http://www.learnallbi.com/what-is-qlikview/> (online)

[22] https://www.ibm.com/support/knowledgecenter/en/SSCRJU_4.0.1/com.ibm.streams.welcome.doc/doc/ibminfospherestreams-introduction.html (online)

APPENDIX A

	PSM title review	1 day	Thu 2/22/18	Thu 2/22/18	
	Chapter 1 review	2 days	Fri 2/23/18	Mon 2/26/18	
	Research problem background	2 days	Tue 2/27/18	Wed 2/28/18	
	Determine problem statement	2 days	Thu 3/1/18	Fri 3/2/18	
	Determine project and scope	3 days	Sat 3/3/18	Tue 3/6/18	
	Submit chapter 1	3 days	Wed 3/7/18	Fri 3/9/18	
	Literature review	15 days	Sun 3/4/18	Thu 3/22/18	
	Study research article	5 days	Sun 3/4/18	Thu 3/8/18	
	Study related neural network	4 days	Fri 3/9/18	Wed 3/14/18	
	Study related FMM	3 days	Tue 3/13/18	Thu 3/15/18	
	Submit chapter 2	1 day	Fri 3/16/18	Fri 3/16/18	
	Chapter 3 methodology	30 days	Sun 3/18/18	Thu 4/26/18	
	Neural network design	3 days	Mon 3/19/18	Wed 3/21/18	
	Data source	3 days	Fri 3/23/18	Tue 3/27/18	
	Submit chapter 3	1 day	Wed 5/2/18	Wed 5/2/18	
	Submit final report	1 day	Fri 5/4/18	Fri 5/4/18	