ROAD ACCIDENTS PREDICTION BY USING ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM (ANFIS)

EFFA RIZAN BINTI SAMSUDDIN

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL: ROAD ACCIDENTS PREDICTION

SESI PENGAJIAN : 2016/2017

SAYA EFFA RIZAN BINTI SAMSUDDIN

mengaku membenarkan tesis (PSM/~~Sarjana/Doktor Falsafah~~) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. **Sila tandakan (/)

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

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

_____ TIDAK TERHAD

(TANDATANGAN PENULIS) (TANDATANGAN PENYELIA)

Alamat tetap: Lot 582, Kangkar Tebrau Baru, Ulu Pandan, 8100, Johor Bahru, Johor

PROF. MADYA DR. ABDUL SAMAD BIN SHIBGHATULLAH
Nama Penyelia

Tarikh: 18 August 2017 Tarikh: 21 August 2017

CATATAN:* Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)

**Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

ROAD ACCIDENTS PREDICTION BY USING ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM (ANFIS)

EFFA RIZAN BINTI SAMSUDDIN

This report is submitted in partial fulfillment of the requirement for the

Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2017

DECLARATION

I hereby declare that this project report entitled

ROAD ACCIDENTS PREDICTION

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT: _____ Date: 18 August 2017

(EFFA RIZAN BINTI SAMSUDDIN)

I hereby declare that I have read this project report and found this project report is sufficient in term of the scope and quality for the award of Bachelor of Computer Science (Artificial Intelligence) With Honours.

SUPERVISOR: _____ Date: 21 August 2017

(PROF. MADYA DR. ABDUL SAMAD BIN SHIBGHATULLAH)

DEDICATION

This project is a dedication for my beloved family, lecturers and friends whom have taught me that everything is achievable as long as there are desire that burn inside. Besides, knowledge is important that do not have an easy way in the process of gaining it and always work hard to achieve success. They are also available by my side and always cheering me up in my despair time and always helping me in giving idea during this project is being develop.

ACKNOWLEDGEMENTS

First I thank Him whom all flows.

I owe millions of thank to my beloved supervisor Prof. Madya Dr. Abdul Samad Bin Shibghatullah for his continuous supports and encouraging comments throughout this Final Year Project's creation. His advice and help have giving me strength in completing my project and report.

I also want to grab this opportunity to express my thanks to my beloved parents and sibling that always giving me their moral support and advice have help me in inmumerable ways. I do really appreciate that.

I want to express my gratitude toward all my friends that contribute to help me completing this Final Year Project. Your opinionns and all the time we had spent brainstorming together are brilliantly helpful to provide me with better ideas.

Last but not least, my thanks are dedicated to all parties or person that I have not mentioned. Thank you for your support.

ABSTRACT

One of the most essential aim of Jabatan Pengangkutan Jalan Raya is to accommodate safe environments for the road safety. Nowadays, the average statistic of road accidents is too high and amount people die in a day because of road accidents also increase. So that, the project entitled Road Accidents Prediction is developed as road accidents prediction in order to reduce the amount of severity of the future possible accidents. The Road Accidents Prediction is developed by using the Adaptive Network Based Fuzzy Inference System (ANFIS) technique. This project aims to give the awareness to the road user based on the risk of accident happen. From the road accidents data, this prediction will provide awareness to the the user be more careful. Besides that, this project also aims to give the details about the accidents to the user. In this project, a comparative study will be conducted by implement the prediction on dataset. This comparative study use real data and predicted data by make the comparison between both data to know the accuracy of the prediction. If the comparative is not the same then there is some changes on prediction model will be occur.

ABSTRAK

Salah satu matlamat yang paling penting dalam Jabatan Pengangkutan Jalan Raya adalah untuk menampung persekitaran yang selamat untuk keselamatan jalan raya. Pada masa kini, statistik purata kemalangan jalan raya adalah terlalu tinggi dan jumlah orang mati dalam satu hari kerana kemalangan jalan raya juga meningkat. Oleh yang demikian, projek bertajuk Ramalan Kemalangan Jalan Raya dibangunkan sebagai ramalan kemalangan jalan raya untuk mengurangkan jumlah kemalangan mungkin masa depan. Ramalan Kemalangan Jalan Raya juga dibangunkan dengan menggunakan teknik Sistem Kesesuaian Fuzzy Berdasarkan Rangkaian Penyesuaian. Projek ini bertujuan untuk memberi kesedaran kepada pengguna jalan raya berdasarkan risiko kemalangan berlaku. Dari data kemalangan jalan raya, ramalan ini akan memberi kesedaran kepada pengguna lebih berhati-hati. Selain itu, projek ini juga bertujuan untuk memberi butir-butir mengenai kemalangan kepada pengguna. Dalam projek ini, kajian perbandingan akan dijalankan dengan melaksanakan ramalan pada set data. Kajian perbandingan ini menggunakan data sebenar dan meramalkan data dengan membuat perbandingan antara kedua-dua data untuk mengetahui ketepatan ramalan. Jika perbandingan tidak sama maka ada beberapa perubahan pada model ramalan akan berlaku.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Introduction

One of the most essential aim of Jabatan Pengangkutan Jalan Raya is to accommodate safe environments for the road safety. The Data on road crash were analysed for a period of 14 years starting from 1985 to 1998 which was gathered from Statistical Report Road Crash. The number of accident for every 1,000,000,000 km registered motorcycle increased from 0.45 in year 1990 to 0.82 to in the year 1998 with an increase of 82.2%. Reducing the amount and of course the severity of the traffic accidents will require safer roads and provide major saving for the society.

The project entitled Road Accidents Prediction Analysis is developed to predict road accidents in order to reduce the amount of severity of the future possible traffic accidents. This system will provide number of accident happened previously and predicted number of accidents that will be happened. Therefore, road user will be aware and be careful if the predicted number of accidents is high. Based on the prediction, the user knows the accident zone by using the system. The system will give the details about the accidents to the user. It help the user to make decision to choose the road that have low risk of predicted accidents possibility the accidents will be occur.

1.2 Problem Statements

Nowadays, has been found that the total number of statistic road accidents in Malaysia are too high. Malaysia's accident rate is ranked as the one of the highest in the world. Every year, the statistic of the road accidents in Malaysia will increase drastically. This situation is very worrying and that the government is confident that the best way to handle the issue is to educate the public and create awareness of the importance of road safety.

The next problem is the road user are not aware with the accidents that have been occur in the specific state. Currently, most studies have proven that drivers are responsible for the main cause of accident. Most drivers tend to have the attitude of overestimating their own ability and the capability of their vehicles. Most of them are not aware with the accidents and may lead to the increases of accidents to be occur.

The last problem that has been identified is the road accidents will increase the number of people death. About 1.25 million people die each year because of the road accidents. Besides that, the road traffic injuries also can cause considerable economic that leads losses of individual, their family and to nations as a whole.

1.3 Objective

This project embarks on the following objectives:

1. To gather and analyse the raw data of road accidents dataset.
2. To predict number of road accidents will be occured each state in Malaysia.
3. To developed and approach to predict road accidents using fuzzy logic technique.

1.4 Scope

Road Accidents Prediction System will develop based on some scope as below:

1.4.1 Data

Data that will be used in this project is road accidents datasets from the official portal of Royal Malaysia Police, Bukit Aman. The data that will be used is the total road accidents by states in Malaysia from year 2003 until 2013.

1.4.2 Study Area

The project covered the specific road area in Malaysia in order to get the data to make the prediction.

1.4.3 Technique

In this project, researcher will predict the road accidents will be occur based on road accidents datasets by using fuzzy logic technique.

1.5 Project Significance

The main goal of this project is to develop a Road Accidents Prediction System which is basically this system can predict the road accidents that will be occur. Other than that, this prediction also will give awareness to the road user based on the risk of accident happen. The road user will be more aware and help the road user to make decission to choose the low risk possibility the accidents will be occur.

1.6 Expected Output

Based on the purpose project, the expected output would be produce the data analysis on statistic and road accidents dataset. Besides that, the Road Accidents Prediction will be able to predict the accidents that will be occur. Prediction by using fuzzy logic approach will be understand and improve by looking at the prediction performance.

1.7 Conclusion

This project report has been divided into 7 chapters and each chapters will be discussed on specific topic. As for chapter 1, it has discussed on the background of the study and in the next chapter will be discussed more about the literature review and project methodology.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

In this chapter tells the reader in details about Literature Review and Project Methodologies. Literature review describe the background of the project proposed which is Road Accidents Prediction. Besides that, this chapter provides summary and evaluation of the previous research or work that is the same or related to the project which can be developed. The purpose of the literature review is to justify the exact choices of research or a project. Furthermore, in order to understand the project, literature review can assist to gather the background information needed. It function as the key to show readers that the developer of the project is familiar with significant and the latest research that is relevant with the topic. Literature review provides chance to discover what has been investigated and what has not and also to discover how the project is related to the work of others.

Methodology is a method or process of activities that is use to developed the project. Methodology will be discussing on the project methodology and project milestones on the research and development of this project and on top of that, reporting project progress to keep up with the flow of research and development. In the methodology also be explaining on the lifecycle that is needed to be implemented for the project to be completed. Other than that, discussion on the project milestones that is designed to make sure all the task is completed by the specified time given is reviewed again.

2.2 Facts and Findings

More recently, neuro-fuzzy systems have gained increasing attention. They are a composition of artificial neural networks (ANN) and fuzzy logic (FL) approaches. Artificial neural networks reconstruct links between input–output pairs for the system being modelled. The ANNs have to be trained in order to generate the desired output. Artificial neural networks have been shown to give useful results in many fields of hydrology and water resources research (Chen *et al.*, 2006, Tingsanchali & Gautam, 2000). Fuzzy logic and fuzzy set theory, founded by Zadeh (1965), are used to identify the characteristics of decision making through a set of logical rules. Sugeno & Yasukawa (1993) have developed a fuzzy logic-based approach to qualitative modelling and have proposed the use of a fuzzy clustering method for the structure identification of models. The adaptive neural-based fuzzy inference system (ANFIS) model and its principles proposed by Jang (1992) have been applied to study many problems.

The model identifies a set of parameters through a hybrid learning rule combining the backpropagation gradient descent an a least squares method. It can be used as a basis for constructing a set of fuzzy IF–THEN rules with

appropriate membership functions in order to generate the previously stipulated input–output pairs. Some researchers have applied ANFIS in prediction modelling. Chang & Chang (2001) studied the intelligent control of a real-time reservoir operation model and found that, given sufficient information to construct the fuzzy rules, the ANFIS helps to ensure more efficient reservoir operation than the classical models based on rule curve. The main purpose of this paper is to develop an ANFIS methodology to estimate flows collectively for a longer period without any restrictive assumptions,by considering all factors affecting the flow simultaneously. Flow records cover the data set inputs and outputs that had been measured over a very long period of time.

2.2.1 Prediction and Classification

Data mining is the process of sorting large data sets then basically to identify the patterns and establish relationships to solve problems through the data analysis. Besides that, data mining also allow enterprises to predict the possible future trends. Data mining techniques are used in manyresearch areas, that including mathematics, cybernetics, genetic and also for marketing. Other than that, data mining techniques are a means to drive efficiencies and predict customer behavior. In that case if the technique used correctly, a business can set itself apart from its competition through the use of preditive analysis.

Data mining can be divided by two definition which is include the classification and prediction. These two forms of data analysis can be used for extracting models that describing the important classes or to predict future data trends. Classification models defined as prediction model for categorical class labels. While the prediction model are use for predict the continous valued function.

The major issue during data preparation for Classification and Prediction are involves the several activities. The activities that involves including data cleaning, relevance analysis, data transformation and reduction. Under data transformation and reduction can be classify two data transform method which is normalization and generalization.

The performance for both classification and prediction can be evalute by make the comparison based on the important criteria. The criteria are accuracy, speed, robustness, scalability and interpretability. Each of the criteria will be considered to make decision which of the data mining technique is the best.

2.2.2 Existing vs Proposed Technique Prediction

There are several road accidents data analysis on prediction that has been developed previously. There are certain features that the prediction do and do not provided which is basically the difference of the technique used between the road accidents prediction and the proposed technique prediction.

For example, the previous prediction that has been developed is traffic accidents predictions based on fuzzy logic approach for safer turban environments. In this study, the researcher use fuzzy logic technique that involves the fuzzification process, the production of the rule base and the defuzzification process in the fuzzy logic model.

Besides that, there is analysis study on time series analysis of road traffic accients in Zimbawe, Kudakwashe Mutangi. In the study, the reseacher focus on finding a suitable model for the annual Zimbawe Traffic Accidents statistic from 1997 to 2013 and to forecast the number of annual traffic accidents likely to occur in the future. The technique used for the prediction is ARIMA(0,1,0), ARIMA(1,1,0) and ARIMA(1,1,1)

Based on the summarization of the previous studies, it can conclude that the existing prediction have a good performances but there are some lacking features or weakness that can be found in that prediction technique. Therefore, it is hoped that this proposed prediction technique can fulfill the best requirement of the prediction technique.

2.2.3 Computional Technique

In other to obtain a better output of the project to be developed, there is a technique that is used to gathering all the data and develop the prediction model. Adaptive Neuro-based Fuzzy Inference System (ANFIS) technique was used as computional technique for prediction model development. The different between Road Accidents Prediction with others prediction is the technique that has been used to predict the data.

2.2.4 Data Analysis Mechanism

With the above discussion on Data Analysis, there must be a mechanism that is available out there to execute all this kind of operation. MATLAB programming language is the most used language for statistical computing and graphics where most professionals use it nowadays.

Figure 2.1: MATLAB programming software

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-environment where problems and solutions are expressed in familiar ma thematical notation. The MATLAB are mostly popular uses are for Math and computation, algorithm development, modeling, simulation, and prototyping and also use for data analysis, exploration and visualization

2.3 Project Methodology

There are lots of system development model that exist out there. And most of it are being widely used in the industry today. To name a few, we have waterfall model, agile model, scrum model and many more. However, with data analysis, we are bound to implement a lifecycle called data analytics lifecycle. Data analytics lifecycle is consist of 6 phases and those 6 phases are:

a. Discovery
b. Data Prep
c. Model planning
d. Model Building
e. Communicate Results
f. Operationalize

Data analytics is based on an iterative process because it closely portrays the real project scenario where project progress may rebound to the initial stage due to some restrict requirements where new information is learned and analyse. This iteration process also may occur due to a reason where individual explore more on the stages of data analytics lifecycle to drive efficiency and improve operational work as a team.

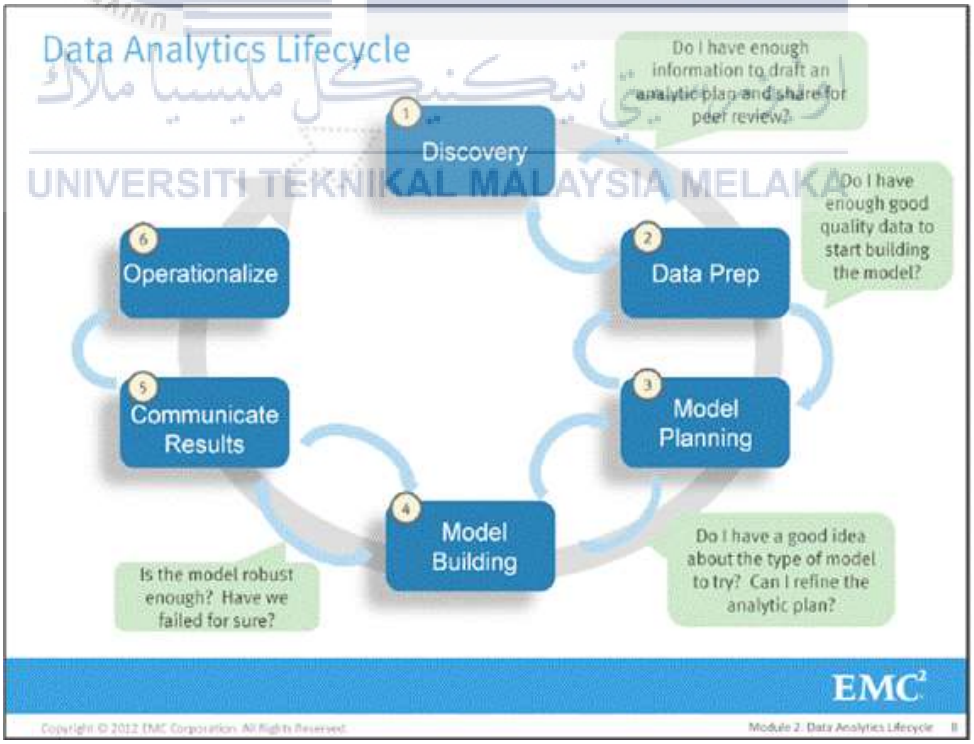


Figure 2.2: Data Analytics Lifecycle

2.3.1 Phase 1 – Discovery

Discovery phase is the kick starter of the data analytics lifecycle where in this phase, team of data analyst and data scientist learns about the business domain. This business domain is a significant information regarding to the history of the organization, whether the organization or business unit have pursue with the similar project from past or accessing available resources to support and improve the project performance. Available resources are in terms of people, technology, time, accessibility and data. In this phase also, the critical activities to be executed is framing the business problem where it would be the most challenging part in this phase where this activity can be affect in the following phases and formulating an initial hypotheses to test and learn the data.

2.3.2 Phase 2 – Data Prep

Data prepping is like preparing a meal for a dinner party, where we will cut all the ingredients and make sure that all ingredients are ready to be cooked into the meal. Same goes with data, preparing the data by filtering the required attributes or reducing data redundancy and etc. In this phase, the presence of an analytic sandbox is required where team can work together with data to perform analytics. This phase also includes the team to execute, load and transform (ELT) or extract, transform and load (ETL) to insert data into the sandbox. The ELT and ETL are sometimes abbreviated as ETLT. Data should be transformed in the ETLT process so the team can work with it and analyze it. In this phase, the team also needs to familiarize itself with the data thoroughly and take steps to condition the data.

2.3.3 Phase 3 – Model Planning

Phase 3 is the model planning phase. This is where the team of data scientist and data analyst determines the model which is the methods, techniques and workflow that they intend to build. Through this, the subsequent phase will be more objective oriented and ease the flow of the data mining. Data exploration also is initiated in this stage as well where relationships between variables and subsequently selects key variables and the most suitable models.

2.3.4 Phase 4 – Model Building

Model Building and Model Planning is a 2 separate activity, whereas model planning is just for the sake of making decision but as for model building is where the development of the process begins. In this phase, data sets for testing, training and production purposes is develop and all the models that have been planned during the model planning phase is build and executed at the same time. The team will also considers whether its existing tools will meet the requirements for running the models, or if it will need a more robust environment for executing models and work flows.

2.3.5 Phase 5 – Communicate Results

Phase 5 is where the communication between the team and stakeholders are initiated. the team, in collaboration with major stakeholders, determines if the results of the project are a success or a failure based on the criteria developed in Phase 1, discovery. The team should identify key findings, quantify the business value, and develop a report to summarize and convey findings to stakeholders.

2.3.6 Phase 6 – Operationalize

Phase 6, the team of data science and data analyst delivers final reports, briefings, code, and technical documents. In addition, the team may run a pilot project to implement the models in a production environment.

2.4 Project Requirements

2.4.1 Software Requirement

Table 2.1: Software Requirement

Software	Description
MATLAB R2015a	Design and coding tool
Microsoft Excel 2013	To store dataset
Microsoft Word 2013	Documentation
Microsoft Power Point 2013	Presentation
Microsoft Project 2013	Project planning and scheduling

2.4.2 Hardware Requirement

2.4.3 Table 2.2: Hardware Requirement

Hardware	Description
Personal Computer	Asus, Intel core i3, 64bit, 2GB (RAM)

2.5 Project Schedule

2.5.1 Milestone

Figure below shows the milestone during the project has been developed. The project milestone shows the project stage from chapter 1 until chapter 4 for the first stage of 'Projek Sarjana Muda 1' (PSM 1) and also chapter 5 until chapter 7 for second stage of 'Projek Sarjana Muda 2' (PSM 2).

Table 2.3: Milestone for PSM 1 and PSM 2

Week	Date	Activity	Note/Action
1	13-17 Feb	Proposal PSM: Discussion & Submission	Deliverable – Proposal Action – Student
		Proposal assessment & verification	Action – Supervisor, Evaluator
2	20-24 Feb	Proposal Correction / Improvement	Action - Student
		List of supervisor / title	Action – PSM/PD Committee
3	27 Feb-3 Mac	Proposal Presentation Chapter 1 (System Development Begins)	Deliverable – Proposal Presentation (PP) Action – Student
4	6-10 Mac	Chapter 1 Chapter 2	Deliverable – Chapter 1 Action – Student, Supervisor
5	13-17 Mac	Chapter 2	Action – Student
6	20-24 Mac	Chapter 2 Chapter 3	Deliverable – Chapter 2 Progress Presentation 1 Action – Student, Supervisor
7	27-31 Mac	Chapter 3 Chapter 4	Action – Student
8	3-7 April	MID SEMESTER BREAK	
9	10-14 April	Chapter 4 Project Demo	Deliverable – Chapter 3 Action – Student, Supervisor
10	17-21 April	Chapter 4 Project Demo	Deliverable – Progress Presentation 2 Action – Student, Supervisor
11	24-28 April	Project Demo	Action – Student
12	1-5 May	Project Demo PSM 1 Report	Action – Student, Supervisor
13	8-12 May	Project Demo PSM 1 Report	Action – Student, Supervisor
		Presentation schedule	Action – PSM/PD Committee
14	15-19 May	Project Demo PSM 1 Report	Deliverable – Complete PSM 1 Draft Report Action – Student, Supervisor
15	22-26 May	FINAL PRESENTATION & PROJECT DEMO	Action – Student, Supervisor, Evaluator
16	29 May-2 June	REVISION WEEK	Deliverable – Complete PSM 1 Logbooks Action – Student, Supervisor
17	5-18 June	FINAL EXAMINATION WEEKS	Action – Student, Supervisor, Evaluator
18	3-7 July	Chapter 4: Design Chapter 5: Implementation	Deliverable – Chapter 4 Action – Student
19	10-14 July	Chapter 5: Implementation	Deliverable – Progress Presentation 1 Action – Student, Supervisor, Evaluator
20	17-21 July	Chapter 5: Implementation Chapter 6: Testing	Deliverable – Chapter 5 Action – Student
21	24-28 July	Chapter 6: Testing	Deliverable – Progress Report Presentation 2 Action – Student, Supervisor
22	31 July-4 Aug	Chapter 6: Testing Chapter 7: Conclusion	Deliverable – Chapter 6 Action – Student, AJK PSM
23	7-11 Aug	Chapter 7: Conclusion Draft Report PSM	Deliverable – Chapter 7 Action – Student, Supervisor
24	14-18 Aug	FINAL PRESENTATION	Deliverable – Draft Report PSM

			Action – Student, Supervisor, AJK PSM
25	21-25 Aug	Draft Report PSM correction. Submit Marks	AJK PSM, Supervisor
26	28 Aug-1 Sept	Upload Report PSM and Logbooks already sign, and fully project material to the system.	Deliverable – Report PSM, Logbooks, Project Material Action – Student, Supervisor

2.5.2 Gantt Chart

Figure below shows that the Gantt chart or the due date before completed the task. The Gantt chart will illustrate the starting point of the task and the ending point of the task. The point of making a Gantt Chart is to ensure that the completion time will be met.

Table 2.4: Gantt Chart

Tasks \ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Discussion/Verification of project title and synopsis with supervisor	■															
Final proposal submission		■														
Preparation on Chapter 1: Introduction • Research on road accidents prediction			■													
Preparation on Chapter 2: Literature Review • Research on fuzzy logic implemented on data road accidents • Identify suitable method to be implemented on the project				■	■	■										
Preparation on Chapter 3: Methodology							■	■								
Analysis									■							

Design										■						
Implementation on identified methods											■	■	■	■		
Presentation of Final Year Project 1															■	
Presentation of Final Year Project 1 report																■

2.6 Conclusion

In this chapter, literature review and methodology is done. The summary of some of the research papers have been discussed at this chapter as well. On the other hand, methodology consist the method or technique that used throughout the whole process. Next chapter can explain about the analysis phases and how the project going to be develop can be discussed.

CHAPTER III

ANALYSIS

3.1 Introduction

In the analysis phase will be focusing more on the analysis from what information that have been gain from the previous chapter. In order to utilize the information, analysis that is being conducted will become the benchmark in getting a reliable information that could help and support in making strategic decision that could either open and lead towards success or failure of an organization. Usually, the problem in the analysis techniques used and requirement resources are required in the analysis. The problem was viewed and the process of identifying problem is the process for defining the differences to get the problem solving by finding the way to reduce the differences.

This chapter begin by presenting the problem analysis and showing how the current prediction run with the information background of the current prediction. During the analysis, the requirement for the new project was gained. The system requirements are including data requirement, functional requirement, non-functional requirement and other requirement. Data requirement show the type of data need for the data used to functioning properly. Functional requirement explained about the function that can be used during system implementation. It includes use-case diagram and description. While, non-functional requirement define the other basic function that a system should have.

3.2 Problem Analysis

Problem analysis is the way to understand and identify the problem in term of causes and effect. Furthermore, the analysis can show the constraint that can limit the solution of the problem. The analysis was made for the current system's to solve the problems from the lacking feature that the existing prediction not provided.

3.2.1 Background of Current Prediction

The current prediction that was develop is the Road Accidents Prediction which is as road accidents prediction in order to reduce the amount of severity of the future possible accidents. This project aims to give the awareness to the road user based on the risk of accident happen. From the road accidents data, this prediction will provide awareness to the the user be more careful. Besides that, this project also aims to give the details about the accidents to the user. In this project, a comparative study will be conducted by implement the

prediction on dataset. This comparative study use real data and predicted data by make the comparison between both data to know the accuracy of the prediction. If the comparative is not the same then there is some changes on prediction model will be occur.

3.2.2 Problem Statement

Nowadays, the total number of statistic road accidents in Malaysia are too high. Next, the problem is the road user are not aware with the accidents that have been occur in the specific state. Besides that, the problem has been faced is the road accidents will increase the number of death. The Road Accidents Prediction was develop to overcome the problem by identify the solution to achieve the objective of the project.

3.3 Requirement Analysis

The word of "Requirement" meaning that what some system must do or must have and what the characteristic it need to have. "Requirement can be either functional or non-functional in the nature. A functional requirement relates directly to the process the system for the next step of the analysis process. This is because it define the function that the system must have. Non-functional requirement refer to behavioral properties that the system need to have such as usability and perfomance." [Dennis A. And Wixom B. H., 2003]. The requirement analysis for this chapter for comprises of data requirement.

3.3.1 Data Requirement

Data requirement is important for identify the required data needed to be used. The table shows the total road accidents by states in Malaysia based on year from 2006 until 2015.

NEGERI *State*	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
PERLIS	1,160	1,364	1,417	1,633	1,548	1,791	1,881	1,895	1,888	1,861
KEDAH	15,505	16,172	16,520	17,701	17,966	19,699	19,935	20,228	20,159	22,016
PULAU PINANG	32,573	33,881	34,049	33,719	34,306	37,158	37,851	39,361	38,747	39,856
PERAK	27,432	29,203	30,539	32,327	32,072	33,506	34,714	35,408	35,131	36,736
SELANGOR	92,632	99,157	100,380	107,429	115,565	128,876	129,106	135,024	137,809	140,957
W.P. KUALA LUMPUR	46,254	49,454	48,671	51,942	53,493	58,795	61,872	64,527	63,535	64,664
NEGERI SEMBILAN	15,197	16,079	17,362	18,369	19,407	21,157	22,146	23,066	23,748	22,939
MELAKA	10,707	11,720	12,105	13,275	14,110	14,720	15,195	16,083	16,375	17,069
JOHOR	43,757	46,584	48,667	51,747	55,381	59,501	62,316	64,600	64,473	67,112
PAHANG	13,242	13,982	15,629	17,068	17,315	19,001	20,554	20,130	19,071	19,635
KELANTAN	7,337	8,116	8,842	9,549	9,707	9,603	9,968	9,748	10,326	9,960
TERENGGANU	7,098	8,155	8,814	10,118	10,106	10,684	10,861	10,996	9,383	10,381
SABAH	13,550	14,256	14,588	15,798	16,192	16,585	17,446	17,438	17,858	17,290
SARAWAK	14,808	15,196	15,488	16,655	17,253	17,964	18,578	18,700	17,693	19,130
JUMLAH *Total*	341,252	363,319	373,071	397,330	414,421	449,040	462,423	477,204	476,196	489,606

SUMBER: CAWANGAN TRAFIK BUKIT AMAN
Source: Traffic Branch Bukit Aman

Figure 3.1: Total Road Accidents by States in Malaysia

3.4 Conclusion

In this chapter, the discussion on the analysis have been discussed. Analysis phase focusses on the outcomes that is obtain through the raw data of road accidents dataset. In the next chapter, will be explained in details on the proposed technique that will be used.

CHAPTER IV

THE PROPOSED TECHNIQUE

4.1 Introduction

In this chapter, the outcomes of the high level design and detailed design of the Road Accidents Prediction was defined. In the high level design and detailed design will be discuss on the deatils of the proposed technique that will be used. The flowchart of the technique used has been shown to explain clearly how the process of the project can be done by using the technique. Besides that, in this chapter also fuzzy logic approach in this prediction model will be describe in details which is usef of the ANFIS modelling.

4.2 High-Level Design

In Road Accidents Prediction high-level design was focuses on the architecture that would be used for developing the prediction model. The archicture diagram provide an overview of an entire project to be developed. High-level design was important as to provide an overview of a solution, platform, product, services or process. This high-level design contains of input or output of the data for the Road Accidents Prediction which is include the flow diagram how the prediction will be done step by step from data gathering until get the final prediction.

4.2.1 ANFIS Model Design

Adaptive Network Based Fuzzy Inference System (ANFIS) is an intelligent neuro-fuzzy technique used for modeling and control of ill-defined and uncertain systems. ANFIS is based on the input-output data pairs of the system under consideration. The figure 4.1 shows the ANFIS modelling process that has been used for the Road Accidents Prediction.

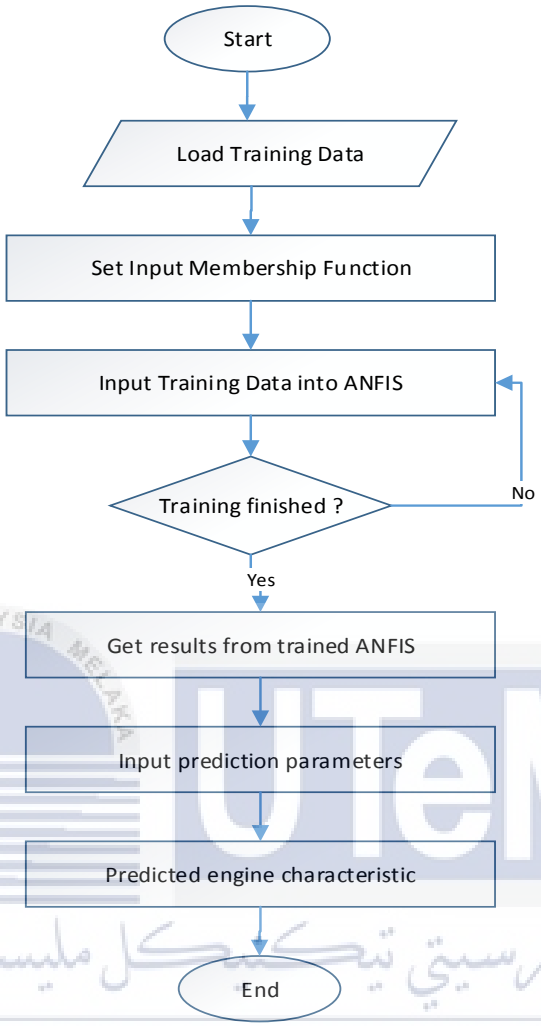


Figure 4.1: ANFIS Modelling Process

The ANFIS modelling process describe all the process of prediction that has been occur. The ANFIS modelling process describes the process from load training data until the prediction process.

4.3 Detailed Design

In this phase, the detailed design about the ANFIS work has been expained in more details for each step. ANFIS process are included input data, training and testing phase, measured actual data and ANFIS parameter adjustment. The figure 4.2 shows the overview on how the flowchart of the prediction by used of ANFIS has been explain in details.

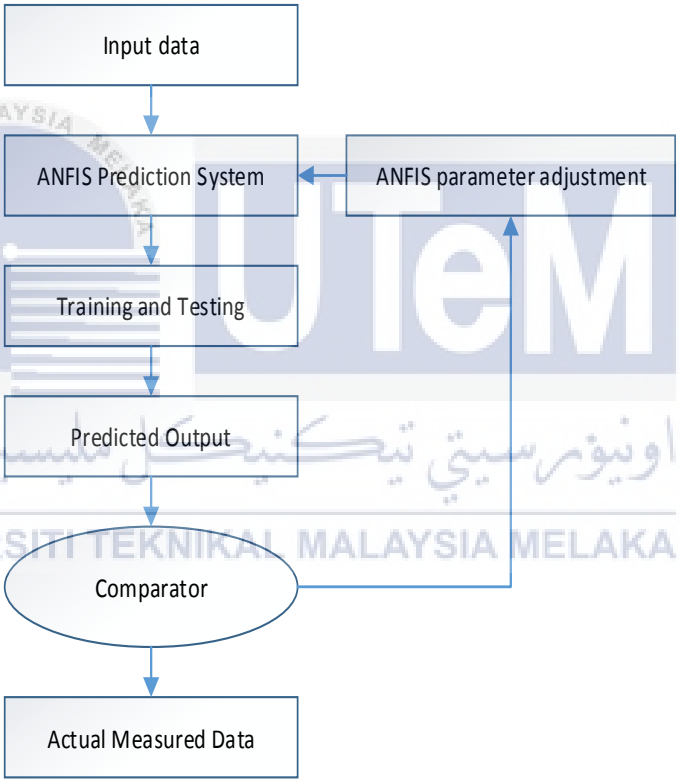


Figure 4.2: ANFIS prediction flowchart

Phase 1: Input data

Input data allows user to type some sort of information into the program and save that information into a variable that the program can

process the data. In this phase, start the process with gathering data which is collect all the data and store it the Microsoft Excel. Next, load the data into MATLAB and followed by the cleaning process of tha data. Data cleaning refer to methods for finding, removing and replacing bad or missing data. Data sets are require preprocessing data to ensure accurate, efficiency and meaningful analysis and performance.

Phase 2: Training and Testing

Testing data is the process by which the input vectors from input or output data sets on which was not trained and testing data will check the generalization capability of the resulting fuzzy inference system.. Next, for training data is the learning process of the model that has been develop. The model was trained until the results are obtained with minimum error. Based on the training and testing, if the dataests are not selected properly then the testing data cannot capture any of the features of the testing data.

Phase 3: Measured Actual Data

In this phase, after the predicted has been done and get the output the comparative analysis will be on work.This comparative study use real data and predicted data by make the comparison between both data to know the accuracy of the prediction. If the comparative is not the same then there is some changes on prediction model will be occur. To ensure that the prediction is accurate the ANFIS parameter was adjust to make the comparison and find the best solution.

Phase 4: ANFIS parameter adjustment

The ANFIS parameters associated with the membership functions changes through the learning process. The computation of these parameters which is refer to parameter adjustment is facilitated by a gradient vector. The gradient vector provides a measure of how well the fuzzy inference system is modeling the input or output data for a given set of parameters. Once the gradient vector is obtained, in order to adjust the parameters to reduce some error measure any of several optimization routines can be applied. This error measure is usually defined by the sum of the squared difference between actual and desired outputs

4.4 Conclusion

In this chapter, the proposed technique that will be used have been disussed in details. It focusses on the outcomes that is obtain through the high-level design and detailed design. In the next chapter will be talking on the project implementation.

CHAPTER V

IMPLEMENTATION

5.1 Introduction

This chapter will be discussing on the implementation of the project using MATLAB programming language in MATLAB to develop a prediction model. From the beginning of the implemenation which is the data resourcing (data discovery) until the prediction algorithm, everything will be implement in MATLAB. MATLAB has a high-performance language for technical computing which is easier for data scientist or data analyst to visualize and analyze their data in an interactive way. This Implementation Phase also put on place of the software, hardware and other important elements of this project.

5.2 Software Development Environment Setup

In the Road Accident Prediction, there are consist two parts that consist of user services and data services. The user services are for the graphical user interface that is displayed to the user and manipulated the system by this interface. Microsoft Excel acts as database to store the data in the system and this situation called data services. The Figure 5.1 below shows all the data will be store in Microsoft Excel before it will be process into the MATLAB.

JUMLAH KEMALANGAN JALAN RAYA MENGIKUT NEGERI, MALAYSIA, 2003 – 2013
Total Road Accidents by States, Malaysia, 2003 - 2013

NEGERI State	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
JOHOR	36,445	41,854	42,606	43,757	46,584	48,667	51,747	55,381	59,501	62,316	64,600
KEDAH	12,456	14,196	14,484	15,505	16,172	16,520	17,701	17,966	19,699	19,935	20,228
KELANTAN	6,892	7,253	7,126	7,337	8,116	8,842	9,549	9,707	9,603	9,968	9,748
MELAKA	10,122	11,029	10,321	10,707	11,720	12,105	13,275	14,110	14,720	15,195	16,083
NEGERI SEMBILAN	13,868	14,505	14,461	15,197	16,079	17,362	18,368	19,407	21,157	22,146	23,066
PAHANG	12,303	13,349	13,381	13,242	13,982	15,548	17,060	17,315	19,001	20,554	20,130
PERAK	25,948	27,542	27,225	27,432	29,203	30,539	32,327	32,072	33,506	34,714	35,408
PERLIS	1,101	1,073	1,037	1,160	1,364	1,417	1,633	1,548	1,791	1,881	1,895
PULAU PINANG	27,817	31,008	30,934	32,573	33,881	34,049	33,719	34,306	37,158	37,851	39,361
SABAH	11,071	12,018	12,760	13,550	14,256	14,588	15,790	16,192	16,585	17,446	17,438
SARAWAK	12,483	13,865	14,209	14,808	15,195	15,488	16,655	17,253	17,964	18,578	18,700
SELANGOR	80,074	87,891	87,705	90,632	99,157	100,380	107,429	115,565	128,878	129,106	135,024
TERENGGANU	6,588	7,003	7,078	7,098	8,155	8,814	10,118	10,106	10,084	10,861	10,996
WILAYAH PERSEKUTUAN	41,492	44,240	45,000	46,254	49,454	48,671	51,942	53,493	58,795	61,872	64,527
JUMLAH Total	298,651	326,817	328,327	339,252	363,319	372,990	397,330	414,421	448,040	462,423	477,204

SUMBER : POLIS DIRAJA MALAYSIA

Figure 5.1: The data store in Microsoft Excel

The software configuration also involved in the environment setup. MATLAB2015a is needed to be installed before configure all the data. Other than that, the MATLAB software will be used as the platform to be used for implementation process. All the data will be implemented into the MATLAB and all the process will be occur in the software by their algorithm and source code.

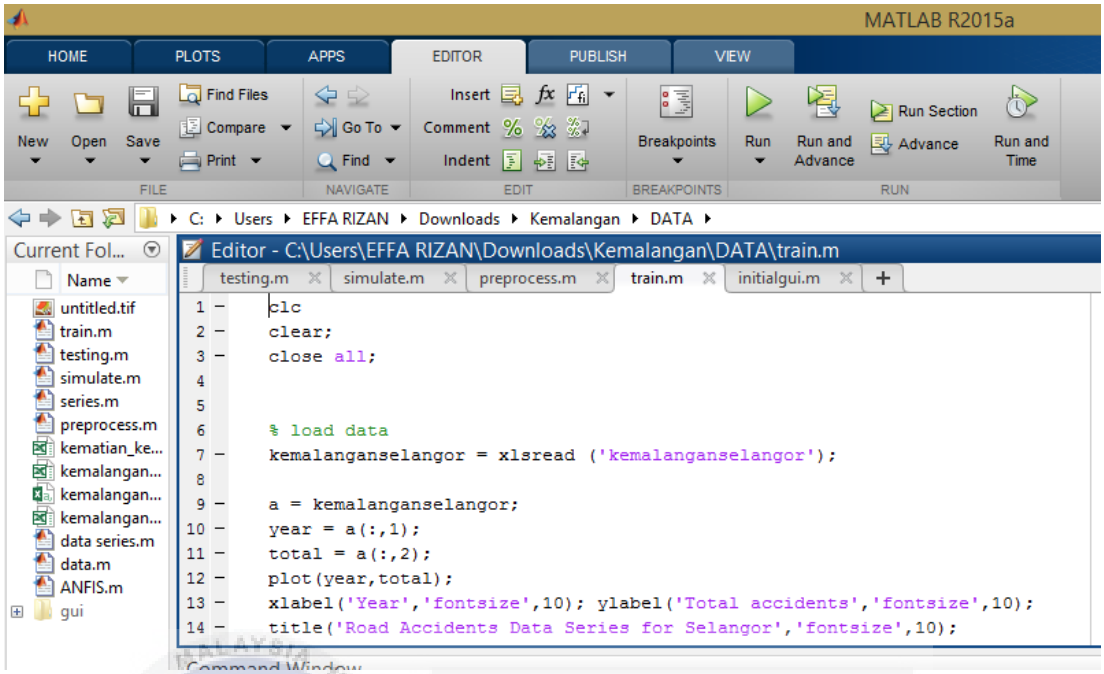


Figure 5.2: The MATLAB interface

The figure 5.2 above shows the interface of MATLAB which is consists of text editor, command window, workspace and others function that will be used for the implementation process. All the coding and the pre-process of the data will be occur in this software and give the performance result.

The software configuration of the MATLAB is to ensure that the function and the process can be work successfully. It is because the version of software use might be effect the process of the data. In this case MATLAB 2015 will be used since the latest version of the MATLAB is in the year of 2015.

5.3 The process of Implementation

In this phase, it is about the detailed process which occurs during the whole implementation process. The process include the data resource, data parsing, data testing and data visualization. All the process involved was implement into the MATLAB software.

5.3.1 Data Resource

Data is of course the heart and soul of this project, without data, this project would not be any benefit to be executed. But here lies the question, where do we get the data? There are many ways in getting the data, especially when dealing with social data one of the ways to collect social data is using data crawler. Data crawler works as a bridge that allows data to be extracted or transferred from any website to your local computer.

However , in this project data source is from the official portal of Royal Malaysia Police, Bukit Aman. In their portal they has been display the open source of data for the public in case to use for analysis or future studies. Figure 5.3 below shows the interface of the Royal Malaysia Police, Bukit Aman official Portal.

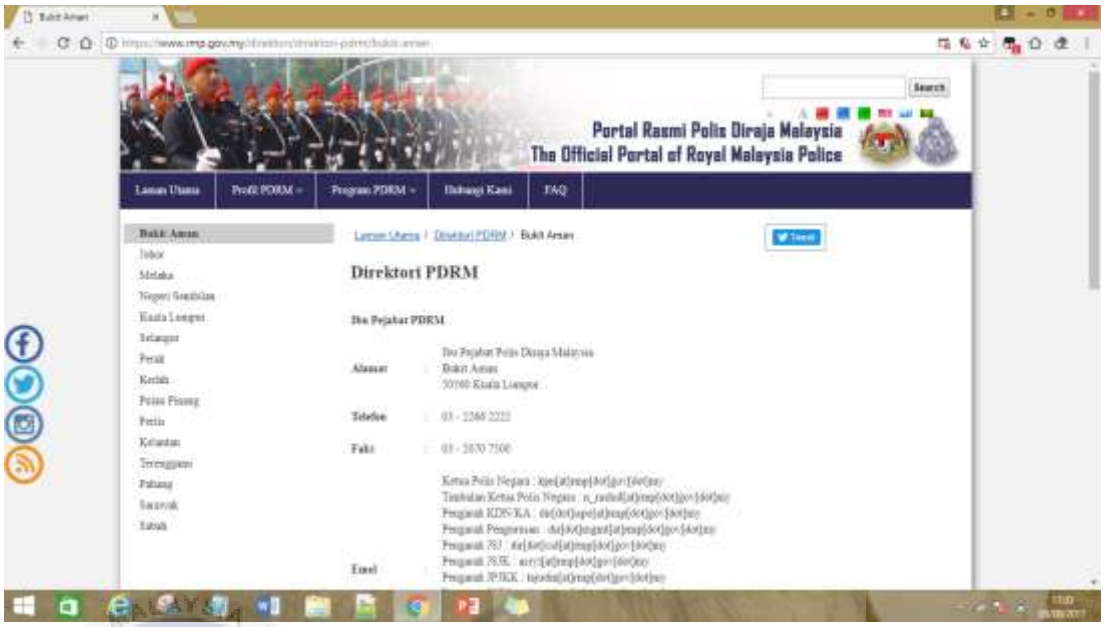
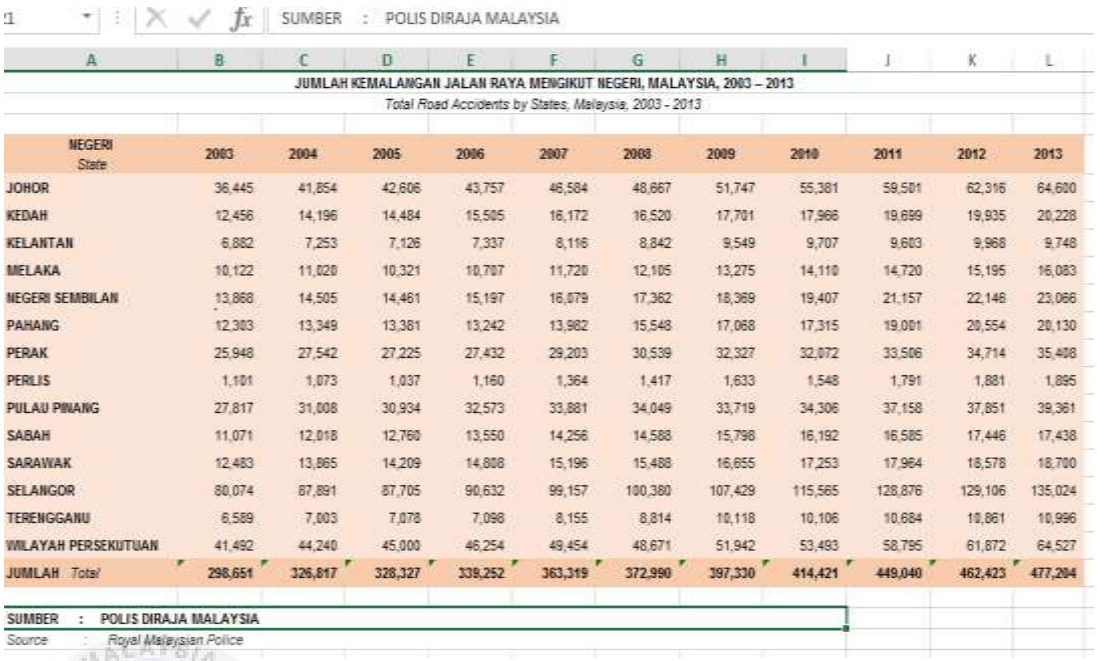


Figure 5.3: Official Portal of Royal Malaysia Police, Bukit Aman.

The data that will be used is the total road accidents by states in Malaysia from year 2003 until 2013 in .xls file format to utilised the information for public. In the Figure 5.4 shows the Road Accident Dataset in .xls format that will be used for the analysis project.



SUMBER : POLIS DIRAJA MALAYSIA

JUMLAH KEMALANGAN JALAN RAYA MENGIKUT NEGERI, MALAYSIA, 2003 – 2013

Total Road Accidents by States, Malaysia, 2003 - 2013

NEGERI *State*	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
JOHOR	36,445	41,854	42,606	43,757	46,584	48,667	51,747	55,381	59,501	62,316	64,600
KEDAH	12,456	14,196	14,484	15,505	16,172	16,520	17,701	17,966	19,699	19,935	20,228
KELANTAN	6,882	7,253	7,126	7,337	8,116	8,842	9,549	9,707	9,603	9,968	9,748
MELAKA	10,122	11,020	10,321	10,707	11,720	12,105	13,275	14,110	14,720	15,195	16,083
NEGERI SEMBILAN	13,868	14,505	14,461	15,197	16,079	17,362	18,369	19,407	21,157	22,146	23,066
PAHANG	12,303	13,349	13,381	13,242	13,982	15,548	17,068	17,315	19,001	20,554	20,130
PERAK	25,948	27,542	27,225	27,432	29,203	30,539	32,327	32,072	33,506	34,714	35,408
PERLIS	1,101	1,073	1,037	1,160	1,364	1,417	1,633	1,548	1,791	1,881	1,895
PULAU PINANG	27,817	31,008	30,934	32,573	33,881	34,049	33,719	34,306	37,158	37,851	39,361
SABAH	11,071	12,018	12,760	13,550	14,256	14,588	15,798	16,192	16,585	17,446	17,438
SARAWAK	12,483	13,865	14,209	14,808	15,196	15,488	16,655	17,253	17,964	18,578	18,700
SELANGOR	80,074	87,891	87,705	90,632	99,157	100,380	107,429	115,565	128,876	129,106	135,024
TERENGGANU	6,589	7,003	7,078	7,098	8,155	8,814	10,118	10,106	10,684	10,861	10,996
WILAYAH PERSEKUTUAN	41,492	44,240	45,000	46,254	49,454	48,671	51,942	53,493	58,795	61,872	64,527
JUMLAH *Total*	298,651	326,817	328,327	339,252	363,319	372,990	397,330	414,421	449,040	462,423	477,204

SUMBER : POLIS DIRAJA MALAYSIA

Source : Royal Malaysian Police

Figure 5.4: Road Accident Dataset in .xls format

In this dataset, it contains its own data attribute and the attributes in the Road Accident Dataset is state in Malaysia and year. Based on all the mention dataset in attribute, not all data will be applied in the analysis and will be use for prediction modelling. Only some of the data will be used in this case for training and testing purpose to make the prediction model of data.

5.3.2 Data Preparation

Data Preparation is the process of collecting, cleaning, and consolidating data into one file or data table for use in analysis. In Data preparation, the process of filtering attribute and fixing missing value is the most critical part of the prepping process. In this phase, the dataset that has been selected will undergo the cleaning process of the data.

The dataset shows the total road accidents by states in Malaysia from year 2003 until 2013. But in this project, the data that will be used is only the total road accidents from 2008 until 2013 only. The dataset in Microsoft Excel will be filter first before the data will be process into the MATLAB.

```
% CREATE TIME SERIES DATA

kemalanganselangor = xlsread ('kemalanganselangor');
x = (2003:1:2012)';

year = [2003 2004 2005 2006 2007 2008 2009 2010 2011 2012];
[Inputs, Targets] = CreateTimeSeriesData(x,year);
Inputs = Inputs';
Targets = Targets';
```

Figure 5.5: Code Read Data from .xls file

The figure 5.6 below shows the Total Road Accidents from 2008 until 2013 that has been used in this analysis after the cleaning process. The dataset is based on their attribute which is the states and the year for the accident.

```
>> kemalangan_negeri = xlsread ('kemalangan_negeri.xls')

kemalangan_negeri =

        2008        2009        2010        2011        2012        2013
       48667       51747       55381       59501       62316       64600
       16520       17701       17966       19699       19935       20228
        8842        9549        9707        9603        9968        9748
       12105       13275       14110       14720       15195       16083
       17362       18369       19407       21157       22146       23066
       15548       17068       17315       19001       20554       20130
       30539       32327       32072       33506       34714       35408
        1417        1633        1548        1791        1881        1895
       34049       33719       34306       37158       37851       39361
       14588       15798       16192       16585       17446       17438
       15488       16655       17253       17964       18578       18700
      100380      107429      115565      128876      129106      135024
        8814       10118       10106       10684       10861       10996
       48671       51942       53493       58795       61872       64527
      372990      397330      414421      449040      462423      477204
```

Figure 5.6: Dataset in .xls after cleaning process

In this analysis, not all year of accidents will be used. It is because if there are too many data will be used for data processing it will effect the process of training data in the MATLAB. It will make the training data will become lack and do not give the best performance of data processing. So that overcome the situation by select the certain data only for testing and training process.

5.3.3 Data Testing and Training

Data testing is the process by which the input vectors from input or output data sets on which was not trained and testing data will check the generalization capability of the resulting fuzzy inference system.

Next, for training data is the learning process of the model that has been develop. The model was trained until the results are obtained with minimum error. Based on the training and testing, if the dataests are not selected properly then the testing data cannot capture any of the features of the testing data.

In the Road Accident Prediction the data will be start with the testing data and followed by the training process. About 30% of the data will be used in testing process while 70% for training data. The data will undergo the testing process to check the performance capability of the data to be process in the ANFIS and produce the output.

```
% generate membership function
data = [rand(10,1) 10*rand(10,1)-5 rand(10,1)];
fis = genfis1(data,[3 7],char('pimf','trimf'));
[x,mf] = plotmf(fis,'input',1);
subplot(2,1,1), plot(x,mf);
xlabel('input 1 (pimf)');
[x,mf] = plotmf(fis,'input',2);
subplot(2,1,2), plot(x,mf);
xlabel('input 2 (trimf)');

x = (2003:1:2012)';
y = sin(5*x)./exp(x/5);
epoch_n = 20;
in_fis = genfis1([x y],5,'gbellmf');
out_fis = anfis([x y],in_fis,epoch_n);
plot(x,y,x,evalfis(x,out_fis));
legend('Training Data','ANFIS Output');
```

Figure 5.7: Code for Testing and Training data

ANFIS applied a learning process of the data by the training process which is by updating the FIS parameters. The training process will continue till the desired number of training steps (epochs) or the desired root mean square error (RMSE) between the desired and the generated output is achieved.

Figure 5.8: The training process

Figure 5.8 above shows how the training process has been occur. The training process is based on the input and output of the data used. The performance of the training data can be improve by looking the resulted output for every training process. To get the better performance, the training process of the data must be done with different value and make the comparison of it.

5.3.4 Data Visualization

Visualization plays a vital role in delivering information to the stakeholders. Through visualization a lot of information can be displayed and more information can be mined to improve day to day operation for business unit and etc. For this Road Accident Prediction, the visualization is based on the graphical user interface (GUI) that will display all the infomation about the data prediction.

In the Road Accident Prediction the output of the prediction process will be display by using the GUI. As can see, the figure 5.9 shows the interface for GUI for the Road Accident Data Prediction. The interface will display all the output for the prediction that has been occur in the ANFIS and make it easier by use the visualization of the process.

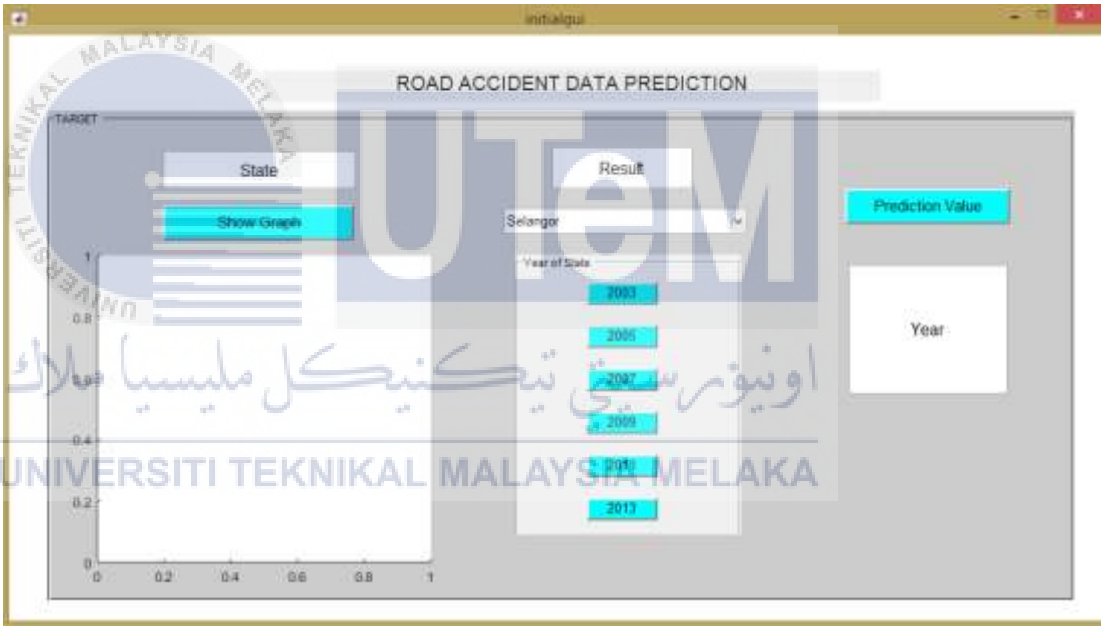


Figure 5.9: GUI of Road Accident Prediction

In the graphical user interface it have the function of load data, final output, the performance of the data and also the prediction value. All of the function well work to ensure that the representation of the analysis can be done by using data visualization.

5.4 Implementation Status

Implementation Status is the milestone for the whole project. It is used to see the status and also the progress of the project in a specific time. The Table below shows the modules for the project that need to be complete:

Table 5.1: Implement Status of Each Module

Module	Description	Duration to Complete
Collecting Data	Collect the data that will be used in the project.	2 week
Cleaning Data	The data that has been collected will be through the cleaning process of the data.	1 week
Train and Test Data	The process of train and test data is the most important part of pre-processing data.	3 week
Technique Used	Implement the fuzzy logic approach on the data to develop the prediction model.	4 week
Prediction Model	The process of data prediction that will give the result of prediction value.	2 week

5.5 Conclusion

In this chapter, the main explanation regarding the implementation phase for prediction modelling process have been described. This chapter also explain the entire required setup environment to show the development of the project during the implementation process. Testing and integration of the project will be carried out later to make an improvement. In the next chapter, will be more focusing on the testing of the project.

CHAPTER VI

RESULTS

6.1 Introduction

In this chapter comprises of the Road Accident Prediction testing. The crucial in this project testing phase is the correct expected outcome can be obtained by testing the algorithm. For this chapter, the testing phase of the project can be discussed. Software testing is an important part that must be carried out to evaluate the capability of the project and to determine the requirement it need to fulfill. Test plan and the results of the of testing can be discussed in this chapter to find out whether it fulfils the requirement for the project. The test can make certain of the system quality, verify and validate the project functionality. The testing process can be discussed first by test plan then follow by test implementation and test results.

6.2 Test Implementation

Test implementation is the process of implement the testing process into the project to ensure that each of the process can be done successfully. In this phase, it is consists of detailed experimental setting and the procedures involve. Besides that, the process of data collection and testing data also will be explain further.

6.2.1 Experimental / Test Description

In Road Accident Prediction, the time series prediction by using ANFIS is used as a model for developed the prediction model. ANFIS consists of a set of fuzzy 'if-then' rules with appropriate membership function that is acts as to generate the input-output pairs. Here the membership function are tuned to the input-output data and excellent result are possible.

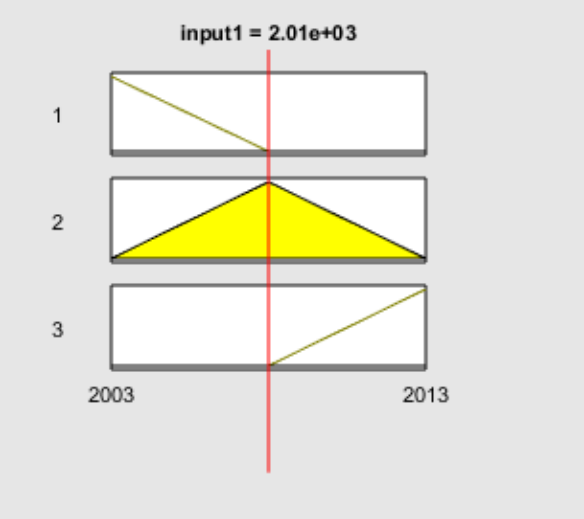


Figure 6.1: The rules in membership function

Figure 6.1 shows the rules that has been used to create the membership function. Each of rules have different condition that refer to the input in this project. For road accident the input that will be used is the year of the accident occur, while for the output is the number of accident for each year.

6.2.2 Test Data

The training data for ANFIS consists of the Z-score calculated in the previous section. From the whole set of road accident dataset, only one sets of data will be used for training ANFIS and the rest are used for testing the performance results of the model. With a proper training scheme and fine filtered data-sets, ANFIS capable of predicting the number of accident quite accurately since it learns from the training data.

In developing any prediction system, the known values are used of the time series up to the point in that time, to predict the value at some point in the future. The method used for this type of prediction is to create a mapping from a previous sample data and compared the value with the predicted data.

For Road Accident Prediction, the value of the number of accident that has been occur in each year will be compared the value of the number of accident after the process of prediction. If the value shows the best accurate then the prediction model will be used to predict the value for the next year.

6.3 Test Results

The ANFIS model can predict from the past values of the number road accident, the training data format is by calculated from the previous value. For example, if the data of current year relates to year 2008 and prediction is to be done for year 2013. The process of prediction is based on the previous values and will be undergo the training process until get the best performance results.

As Figure 6.2 shows the comparison between the expected actual output and the output produce by ANFIS. Based on the graph the comparison will be conducted and if the comparative is not the same then there is some changes to the prediction model.

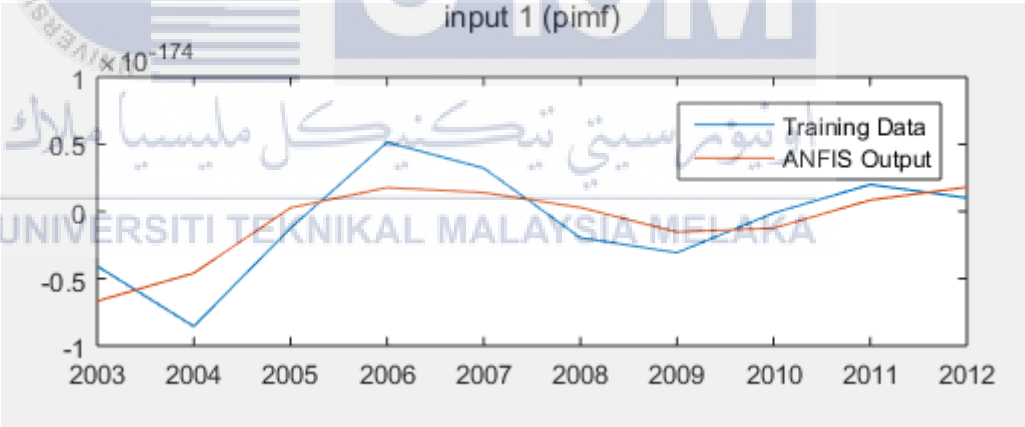


Figure 6.2: The actual output vs ANFIS Output

Figure 6.3 shows the value of road accident prediction for each year in Malaysia. From the resulted value, the prediction value of number of accident will be compare to the actual value. The prediction might be not accurate because of the error that data has been undergo the preprocess frequently.

Variables - kemalanganselangor

+1 | kemalanganselangor | out_fis

11x2 double

	1	2	3
1	2003	80074	
2	2004	NaN	
3	2005	87705	
4	2006	90632	
5	2007	99157	
6	2008	100380	
7	2009	107429	
8	2010	115565	
9	2011	128876	
10	2012	129106	
11	2013	135024	
12			

Figure 6.3: The road accident prediction value

As has been know that the value of prediction cannot be 100% accurate. There will be some of the value will give more or less to the actual value. If the difference of the predicted value with the actual value is not to big then it show the prediction is good. The implementation of ANFIS in this prediction model shows very good learning and prediction capabilities which is an efficient method.

6.4 Accuracy Output

After doing all the prediction process, the important thing that need to be consider is the accuracy of the result. Based on the result, the comparative study has been done to ensure that the final output can be used for the prediction model. In the Road Accident Prediction, the accuracy will be check based on the performance with different number of state.

In this case, the comparison is based on the performance of the prediction by using three different number of state. For the first, the training data will do for 1 state, 3 state and also for 5 state. Based on the result for each condition it will give the different value of the epoch error which is the less value is the best result.

Table 6.1: The Epoch error based on Number of State

Epoch	Number of State	Epoch error	Average error on FIS output	Accuracy	FIS data structure
5	1 state	0.76150	0.71890	94.41%	Same
5	3 state	0.55325	0.62551	88.45%	Same
5	5 state	1.14600	0.91280	79.65%	Same

Based on the Table 6.1 above show the epoch error based on the different number of state used to test the performance of the data. From that, the best performance is state with one state only which is shows the lowest error compare to the others. When the value of error is smaller that it can say that the accuracy of the result is the best to be used.

Besides that, the value of accuracy also can be consider as the factor to know how close the prediction value with the actual value. As can shows in the Table 6.1 the value of accuracy was given in the percentage value. If the value is near to 100% percent, so that the model is the best and give the most accurate value. In the Road Accident Prediction the most accurate value is by one number of state with the highest value of percentage.

6.5 Conclusion

In this chapter, the entire test implementation and test results has been discussed. The test result is shown in the last part and most of the result is OK and have some NOT OK for the test phase. This proves that the testing phase is a useful phase in order to determine the error or mistake in the project. The result can be used in the future to make corrections. It also can give contribution in further system development.

For the next chapter, it can be the conclusion for whole project which include the project strength and weakness, plans for future development, and the contribution of the project.

CHAPTER VII

CONCLUSION

7.1 Introduction

This would be the last chapter that concludes the findings and output of the project where in this chapter will be discussing more on the weakness and strengths of the project and of course to seek for room of improvement for the project to be running in a much more fluent and efficient way.

7.2 Observation on Weakness and Strengths

The observation of the weaknesses and strengths of the project that have been developed, its contribution to the user and proposition for improvements in the future. In this chapter, it can review the project whether the project developed is successfully achieved its scope and objectives that was discussed in the previous chapter. The weaknesses and the strengths are normal to have in any project development. The

strengths in this project are good thing that can be the project be successful while the weaknesses in this project are the thing that important to address the future development.

7.2.1 Strengths

After doing a full analysis on the whole project, there are strengths that could be the point of value for this project. Point of value that would be the benchmark that is applicable in the near future for the same or different purposes.

The first thing that would highlight the strength of this project would be the extraction of information through the raw data. The prediction part would be something that is time consuming due to the fact that we need a prior data understanding in order to extract the information but, the process is worth to be explore as it strengthen the prediction modelling process.

Next would be the visualization of the data. As we all are aware, to deliver information effectively especially towards an individual with no background in analytics or related subject, visualization is all that matters. With the right visualization model, we can improve end user understanding and at the same time creates a data interaction between the the user and information.

7.2.2 Weakness

No systems are created in a perfect manner as perfection takes time to be developed. The same concept applies to this project as well because this would be the first publish project and of course it will require some improvement in the near future.

Truth be spoken, the system is lacking in so many ways and have no doubt regarding that. This is because the system itself is relying on road accident dataset which mainly focusing on states and year only. So, the prediction for each day for road accident that has been occur in Malaysia cannot be shown.

Other than that, the total data that is available is not being utilized completely. As we can observe in the data overview tab, there are a lack of data source that be used in this project. The information gain may not be fully extracted from the raw data which means there are still lots of information that is obtainable.

7.3 Propositions for Improvement

Improvement is definitely needed in every project that has been developed. This is to ensure a gradual increase in user satisfaction and also to increase reliability of the project. As for this project, there are several improvement that can be maded and this improvement is to recover all the weakness and find the suitable solution to make this project more useful.

The first would be applying this analysis project into the system. This is to ensure that an individual or goverment can use this system effectively and efficiently. But here lies the problem, to develop a system and implement the prediction model into the system is not easy and might be take a long time to complete it.

Next would be the exploration of data. Providing better insight would be one of the major improvement that could increase user understanding. In order to provide better insight, exploration of all available data would be something that should be done to understand more about the data prediction modelling.

7.4 Project Contribution

- Contribution to university: This analysis is the results of the project of a degree student in Universiti Teknikal Malaysia Melaka (UTeM). Thus it can be as reference for other student who doing the similar project and domain.

- Contribution for an individual: For the individual who do this project, it let them to gain more knowledge on procedure to make the project analysis.

- Contribution for Jabatan Pengangkutan Jalan Raya: The road accident prediction that has been developed can be used as a platform to reduce the amount of severity of the future possible accidents.

7.5 Conclusion

The project documentation for the project is come to the end for this part. The introduction, literature review, project methodology, analysis and high-level design have been successfully accomplished in the provided time period. Most of the requirements and specification of the project has been fulfilled with achieved the objectives of this project.

REFERENCES

[1] Balogan .O.S.*, O.O.Awoeyo and Dawudo .O.O. "Use of Time Series Analysis of Road Accident Data in Lagos State" *International Journal of Advanced Research on 2014.* Volume 2, pp 1045-1059.

[2] Gajendran C, Serin VK, Seenu SG, Swati P. "Different Methods of Accident Forecast Based on Real Data" *Civil & Environmental Engineering Eng 2015.* Volume 5, pp 1-4.

[3] Talib Ria Jaafar , Mohd Faizar Mustafa, Sutiman Kemin, Ramlan Kasiran. "Kemalangan Jalan Raya : Analisis Data Membabitkan Pengguna Motosikal" *Jurnal Teknologi, Universiti Teknologi Malaysia, Jun, 2003*. Pp 1-14.

[4] D. Karlis, E. Hermans. "Time Series Models for Road Safety Accident Prediction" *Onderzoekslijn Risicobepaling on Feb, 2012*. RA-MOW-2011-020.

[5] Kudakwashe Mutangi. "Time Series Analysis of Road Traffic Accidents in Zimbabwe" *International Journal of Statistics and Applications 2015, Department of Mathematics.* 5(4) pp. 141-149.

[6] Chungwon Lee, Bong Gyou Lee, Kisu Kim, Hye Sun Lee. "A VDS Based Traffic Accident Prediction Analysis and Future Application" *University of Seoul, Dept of Transportation Engineering.* pp. 901-902.

[7] Ridwan A. Sanusi, F.B Adebola, N.A Adegoke. "Cases of Road Traffic Accident in Nigeria: A Time Series Approach" *Mediterranean Journal of Social Sciences on March 2016.* Volume 7, pp 542-552.

[8] M. Erol Keskin, Dilek Taylan, Ozlem Terzi. "Adaptive Neural-based Fuzzy Inference System (ANFIS) approach for Modelling Hydrological Time Series" *Hydrological Sciences Journal,* 51:4, pp 588-598.

[9] Jin Xue Bo, Wang Jiang Feng, Zhang Hui Yan, Cao Li Hong. "ANFIS Model for Time Series Prediction" *Beijing Technology and Business University, 2013.*

APPENDICES

APPENDIX A: GANTT CHART FOR PSM 1

Week / Tasks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Discussion/Verification of project title with supervisor															
Final proposal submission															
Preparation on Chapter 1: Introduction • Research on road accidents prediction															
Preparation on Chapter 2: Literature Review • Research on fuzzy logic implemented on data road accidents • Identify suitable method to be implemented on the project															
Preparation on Chapter 3: Methodology															
Analysis															
Design															
Implementation on identified methods															
Presentation of Final Year Project 1															
Presentation of Final Year Project 1 report															

APPENDIX B: GANTT CHART FOR PSM 2

Week / Tasks	1	2	3	4	5	6	7	8	9
Completion of Chapter 5: Implementation	■	■							
Completion of Chapter 6: Testing/Result			■	■	■				
Completion of Chapter 7: Conclusion						■			
Complete and Submit the PSM 2 report							■		
Presentation PSM 2								■	
Refine PSM 2 report								■	■
Submission updated report to the system									■

APPENDIX C: MILESTONE FOR PSM 1 AND PSM 2

Week	Date	Activity	Note/Action
1	13-17 Feb	Proposal PSM: Discussion & Submission	Deliverable – Proposal Action – Student
		Proposal assessment & verification	Action – Supervisor, Evaluator
2	20-24 Feb	Proposal Correction / Improvement	Action - Student
		List of supervisor / title	Action – PSM/PD Committee
3	27 Feb-3 Mac	Proposal Presentation Chapter 1 (System Development Begins)	Deliverable – Proposal Presentation (PP) Action – Student
4	6-10 Mac	Chapter 1 Chapter 2	Deliverable – Chapter 1 Action – Student, Supervisor
5	13-17 Mac	Chapter 2	Action – Student
6	20-24 Mac	Chapter 2 Chapter 3	Deliverable – Chapter 2 Progress Presentation 1 Action – Student, Supervisor
7	27-31 Mac	Chapter 3 Chapter 4	Action – Student
8	3-7 April	MID SEMESTER BREAK	
9	10-14 April	Chapter 4 Project Demo	Deliverable – Chapter 3 Action – Student, Supervisor
10	17-21 April	Chapter 4 Project Demo	Deliverable – Progress Presentation 2 Action – Student, Supervisor
11	24-28 April	Project Demo	Action – Student
12	1-5 May	Project Demo PSM 1 Report	Action – Student, Supervisor
13	8-12 May	Project Demo PSM 1 Report	Action – Student, Supervisor
		Presentation schedule	Action – PSM/PD Committee

14	15-19 May	Project Demo PSM 1 Report	Deliverable – Complete PSM 1 Draft Report Action – Student, Supervisor
15	22-26 May	FINAL PRESENTATION & PROJECT DEMO	Action – Student, Supervisor, Evaluator
16	29 May-2 June	REVISION WEEK	Deliverable – Complete PSM 1 Logbooks Action – Student, Supervisor
17	5-18 June	FINAL EXAMINATION WEEKS	Action – Student, Supervisor, Evaluator
18	3-7 July	Chapter 4: Design Chapter 5: Implementation	Deliverable – Chapter 4 Action – Student
19	10-14 July	Chapter 5: Implementation	Deliverable – Progress Presentation 1 Action – Student, Supervisor, Evaluator
20	17-21 July	Chapter 5: Implementation Chapter 6: Testing	Deliverable – Chapter 5 Action – Student
21	24-28 July	Chapter 6: Testing	Deliverable – Progress Report Presentation 2 Action – Student, Supervisor
22	31 July-4 Aug	Chapter 6: Testing Chapter 7: Conclusion	Deliverable – Chapter 6 Action – Student, AJK PSM
23	7-11 Aug	Chapter 7: Conclusion Draft Report PSM	Deliverable – Chapter 7 Action – Student, Supervisor
24	14-18 Aug	FINAL PRESENTATION	Deliverable – Draft Report PSM Action – Student, Supervisor, AJK PSM
25	21-25 Aug	Draft Report PSM correction. Submit Marks	AJK PSM, Supervisor
26	28 Aug-1 Sept	Upload Report PSM and Logbooks already sign, and fully project material to the system.	Deliverable – Report PSM, Logbooks, Project Material Action – Student, Supervisor