

FLOOD FORECASTING APPLICATION SYSTEM BY USING
DEMPSTER SHAFER THEORY OF EVIDENCE

NUR IZZATI LIYANA BT MUHAMAD MAZLAN

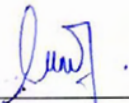
This report is submitted in partial fulfillment of the requirements for the
Bachelor of Computer Science (Artificial Intelligence)

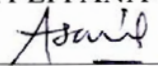
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2012

DECLARATION

I hereby declare that this project report entitled
**FLOOD FORECASTING APPLICATION SYSTEM BY USING
DEMPSTER SHAFER THEORY OF EVIDENCE**

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

STUDENT :  Date: 04/SEPTEMBER/2012
(NUR IZZATI LIYANA BT MUHAMAD MAZLAN)

SUPERVISOR :  Date: 04/SEPTEMBER/2012
(DR ABDUL SAMAD SHIBGHATULLAH)

DEDICATION

This report is dedicated to my beloved parents; friends,
supervisors and my dearest to me who have
provided encouragement and guidance all
the way during the completion
of the report.

ACKNOWLEDGEMENTS

Foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Abdul Samad Shibgatullah for his patience, motivation, enthusiasm, immense knowledge and continuous support through out my project. His valuable guidance and constructive evaluations have been of great value for me in all the time of research and writing of this thesis.

Besides that, I also want to thank Department of Irrigation and Drainage, Melaka with help of En Wan Idris on the data that I get and use in this project also the useful information about flood disaster in Melaka.

My sincere thanks also go towards my friends for their continuing supports and encouragement all the way to accomplish my project especially my coursemate and roommate in UTeM who always give moral support when I was discouraged. I am grateful to my beloved family for their patience, benevolence and supporting me spiritually throughout my life. Last but not the least, I offer my regards and blessings to all of those who supported me in any respect during the completion of my Final Year Project.

Thank you.

PENGHARGAAN

Pertama sekali, saya ingin merakamkan ucapan terima kasih yang amat mendalam dan ikhlas kepada penyelia saya Dr. Abdul Samad Shibghatullah atas kesabarannya, motivasi, semangat, pengetahuan yang besar dan sokongan yang berterusan sepanjang projek saya. Bimbingan beliau yang berharga serta membina amat bernilai bagi saya sepanjang penyelidikan dan penulisan tesis ini.

Selain itu, saya juga ingin mengucapkan terima kasih kepada Jabatan Pengairan dan Saliran Melaka dengan bantuan En. Wan Idris yang membantu memberikan data untuk digunakan di dalam projek ini serta maklumat yang berguna tentang bencana banjir di Melaka.

Terima kasih juga kepada kawan-kawan saya terhadap sokongan yang berterusan serta galakan dalam pelbagai aspek dalam menyempurnakan projek saya terutamanya rakan seperjuangan saya di UTeM yang sentiasa memberi sokongan moral sewaktu saya tidak bersemangat. Saya juga berterima kasih yang tidak terhingga kepada keluarga tercinta atas kesabaran mereka menyokong sepanjang hidup saya. Akhir sekali, saya ingin merakamkan rasa terima kasih kepada semua yang menyokong saya dalam menyempurnakan Projek Sarjana Muda saya.

Terima kasih.

ABSTRACT

The project is about an application on Flood Forecasting Application System using Dempster Shafer Theory. The scope for this project is around Malacca and this application is built using offline data that obtained from the Department of Irrigation and Drainage, Melaka. The purpose of this application is to predict the probability of getting involves in flood disaster around Melaka. This application also can perceive the area the most frequently label as the high potential for the flood disaster. Therefore, it is very important for residents to know the risky area as we can take the precautionary to face the problem. This application will compute all the possibility using the uncertainty theory which is Dempster Shafer Theory. It's came with the best probability calculation where it can give us the best detection area for the flood disaster. Based to this theory, it will calculate the probability based on several factors like proportion of rainfall, water level of nearest river and ratio for how frequent rainy day in a month. Here, the Dempster Shafer theory is used to calculate and predict the area of high potential to occur flood disaster based on the inputs or attributes from the datasets. This application will develops using the Matlab language and the algorithm of the Dempster Shafer theory will be applied into Matlab software.

ABSTRAK

Projek ini adalah satu aplikasi tentang Sistem Ramalan Banjir menggunakan Teori Dempster Shafer. Projek ini telah dijalankan di sekitar Melaka dan dibina dengan menggunakan data yang sedia ada yang diperolehi daripada Jabatan Pengairan dan Saliran Melaka. Tujuan projek ini dikaji adalah untuk meramalkan kebarangkalian untuk bencana banjir ini berlaku di Melaka. Aplikasi ini juga boleh meramalkan kawasan yang berpotensi tinggi dalam bencana banjir. Ini kerana sangat penting bagi penduduk tempatan untuk mengetahui kawasan banjir agar mereka dapat mengambil langkah-langkah berjaga. Aplikasi ini akan mengira semua kebarangkalian menggunakan teori ketidakpastian Teori Dempster Shafer. Berikutan pengiraan tersebut, ia akan memberikan kebarangkalian yang terbaik untuk mengesan kawasan yang berpotensi untuk bencana banjir. Berdasarkan teori ini, kebarangkalian tersebut akan dikira melalui beberapa faktor seperti kadaran taburan hujan, takat air sungai dan ratio perkadaran air hujan dalam sebulan. Teori ini akan mengira serta memaal kawasan yang berpotensi untuk bencana banjir berdasarkan input ataupun atribut daripada dataset. Aplikasi ini dibina menggunakan bahasa Matlab serta algoritma Teori Dempster Shafer yang digunakan didalam perisian Matlab.

TABLE OF CONTENTS

CHAPTER	SUBJECT	PAGE
	DECLARATION	II
	DEDICATION	III
	ACKNOWLEDGEMENTS	IV
	PENGHARGAAN	IV
	ABSTRACT	VII
	ABSTRAK	VII
	LIST OF TABLES	IX
	LIST OF FIGURES	X
	LIST OF ABBREVIATIONS	IXII
CHAPTER 1	INTRODUCTION	1
1.0	INTRODUCTION	1
1.2	PROBLEM STATEMENT	4
1.3	OBJECTIVES	5
1.4	SCOPE	6
1.5	PROJECT SIGNIFICANT	6
1.6	EXPECTED OUTPUT	7
1.7	CONCLUSION	7

CHAPTER 2	LITERATURE REVIEW AND PROJECT	
METHODOLOGY		9
2.1	INTRODUCTION	9
2.2	FACTS AND FINDINGS	10
2.2.1	DOMAIN	10
2.2.2	EXISTING SYSTEM	13
2.2.3	TECHNIQUE	13
2.3	PROJECT METHODOLOGY	16
2.3.1	APPROACH	16
2.3.2	MODEL	17
2.4	PROJECT REQUIREMENT	19
2.4.1	SOFTWARE REQUIREMENT	19
2.4.2	HARDWARE REQUIREMENT	19
2.4.3	OTHER REQUIREMENT	20
2.5	PROJECT SCHEDULE AND MILESTONES	20
2.6	CONCLUSION	23
CHAPTER 3	ANALYSIS	24
3.1	INTRODUCTION	24
3.2	PROBLEM ANALYSIS	25
3.3	REQUIREMENT ANALYSIS	27
3.3.1	DATA REQUIREMENT	27
3.3.2	FUNCTIONAL REQUIREMENT	32
3.3.2.1	PRE-PROCESSING DATA	33
3.3.2.2	ANALYSING DATA	34

3.3.2.3	BASIC PROBABILITY ASSIGEMENT	34
3.3.2.4	COMBINING MASS VALUE	34
3.3.2.5	PERFORMING GRAPH PERCENTAGE	34
3.3.3	NON-FUNCTIONAL REQUIREMENT	38
3.3.3.1	PERFORMANCE	38
3.3.3. 2	SIZE OF DATA	38
3.4	CONCLUSION	39
CHAPTER 4	DESIGN	40
4.1	INTRODUCTION	40
4.2	HIGH LEVEL DESIGN	41
4.2.1	SYSTEM ARCHITECTURE	42
4.2.2	USER INTERFACE DESIGN	42
4.2.2.1	NAVIGATION DESIGN	43
4.2.2.2	INPUT DESIGN	47
4.2.2.3	TECHNICAL DESIGN	47
4.2.2.4	OUTPUT DESIGN	50
4.2.3	NON-DATABASE DESIGN	52
4.3	DETAILED DESIGN	54
4.3.1	SOFTWARE SPECIFICATION	54
4.3.1.1	POP-UP MENU CONTROL	54
4.3.1.2	GRAPH BUTTON	66
4.3.1.3	CLEAR BUTTON	66
4.3.1.4	EXIT BUTTON	67
4.3.2	PHYSICAL DATABASE	67

4.4	CONCLUSION	68
CHAPTER 5	IMPLEMENTATION	80
5.1	INTRODUCTION	80
5.2	SOFTWARE AND HARDWARE DEVELOPMENT ENVIRONMENT SETUP	81
5.3	SOFTWARE OR HARDWARE CONFIGURATION MANAGEMENT	82
5.3.1	CONFIGURATION ENVIRONMENT SETUP	83
5.3.2	VERSION CONTROL PROCEDURE	84
5.4	IMPLEMENTATION STATUS	86
5.5	CONCLUSION	88
CHAPTER 6	TESTING	90
6.1	INTRODUCTION	90
6.2	TEST PLAN	91
6.2.1	TEST ORGANIZATION	91
6.2.2	TEST ENVIRONMENT	92
6.2.3	TEST SCHEDULE	93
6.3	TEST STRATEGY	94
6.3.1	CLASSES OF TESTS	95
6.3.1.1	UNIT TESTING	95
6.3.1.2	SYSTEM TESTING	96

6.4	TEST IMPLEMENTATION	97
6.4.1	TEST DESCRIPTION	98
6.4.2	TEST DATA	100
6.5	TEST RESULT AND ANALYSIS	102
6.6	CONCLUSION	104
CHAPTER 7	PROJECT CONCLUSION	105
7.1	OBSERVATION ON WEAKNESS AND STRENGTHS	105
7.1.1	WEAKNESS	106
7.1.2	STRENGTHS	107
7.2	PROPOSITIONS FOR IMPROVEMENT	107
7.3	CONTRIBUTION	108
7.4	CONCLUSION	109
REFERENCES		110
BIBLIOGRAPHY		111
APPENDICES		112

LIST OF TABLES

TABLE	TITLE	PAGE
Table 2.1 :	Software Requirement	19
Table 2.2 :	Project Schedule and Milestones PSM I	21
Table 2.3 :	Project Schedule and Milestones PSM II	24
Table 3.1 :	Data Requirement January	31
Table 3.2 :	Data Requirement February	32
Table 3.3 :	Data Requirement March	32
Table 3.4 :	Data Requirement April	32
Table 3.5 :	Data Requirement May	32
Table 3.6 :	Data Requirement June	33
Table 3.7 :	Data Requirement July	33
Table 3.8 :	Data Requirement August	33
Table 3.9 :	Data Requirement September	33
Table 3.10 :	Data Requirement October	34
Table 3.11:	Data Requirement November	34
Table 3.12 :	Data Requirement December	34
Table 3.13 :	Rules of evidences	35
Table 4.1 :	Training Input	51
Table 4.2 :	Navigation Input	51
Table 5.1 :	Environment Setup	83
Table 5.2 :	Version Control Procedures for FFAS	84
Table 5.3 :	Implementation Status for FFAS	86
Table 6.1 :	Test Organization for FFAS	92
Table 6.2 :	Hardware and Firmware Configuration FFAS	93
Table 6.3 :	Test Schedule for FFAS	93
Table 6.4 :	Software Testing based to Black Box and White Box Strategy	95
Table 6.5 :	System Testing for FFAS	96
Table 6.6 :	System Test Description for FFAS	98

LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 2.1 :	Basic Probability Assignment formula	15
Figure 2.2 :	Combination of Rules formula	15
Figure 2.3 :	Overall Process	15
Figure 2.4 :	Waterfall SDLC Design	18
Figure 3.1 :	Examples of Water level Graph	31
Figure 3.2 :	Processing Data	36
Figure 3.3 :	Flow Chart – (First)	39
Figure 3.4 :	Flow Chart-(Second)	40
Figure 3.5 :	Data Flow Diagram – Context Diagram	41
Figure 3.6 :	Data Flow Diagram-First Level Diagram	41
Figure 4.1 :	Interface Design	46
Figure 4.2 :	Pop-up Menu Design	47
Figure 4.3 :	Data Analysis Design	48
Figure 4.4 :	Percentage Flood Interface	49
Figure 4.5 :	Graph Performing Design	49
Figure 4.6 :	Clear Button Design	50
Figure 4.7 :	Data Analysis Design	54
Figure 4.8 :	Output of Predicted Result and Graph	55
Figure 4.9 :	Database Data	57
Figure 5.1 :	Environment Setup	82
Figure 6.1 :	First Dataset Used	101
Figure 6.2 :	Real Dataset Used	101
Figure 6.3 :	System Result I	102
Figure 6.4 :	System Result II	103

LIST OF ABBREVIATIONS

PSM	-	Projek Sarjana Muda
FFAS	-	Flood Forecasting Application System
SDLC	-	System Development Life Cycle
BPA	-	Basic Probability Assignment
DFD	-	Data Flow Diagram

CHAPTER I

INTRODUCTION

1.1 Project Background

Malaysia is one country that is free from devastating natural disasters. Serious disaster like hurricanes, tornadoes, volcanic eruptions and earthquake are not directly affected Malaysia. However, natural disaster such as floods is unavoidable because of natural causes such as high intensity leads to flash flood, heavy widespread rain leads to land inundation and another causes of flood was, human induced floods, because of disposal of solid wastes into rivers, sediments from land clearance and construction areas, increase in impervious areas and Obstruction and constriction in the rivers. Currently, Malaysia seems like been attack with many unexpected flash flood. Besides that, Malaysia is one country that is located on the equator, and has summer and rainy all the time, so there are high potential for the unexpected flood disaster to occur.

Flood is an overflow or inundation that comes from a river or other body of water and causes or threatens damage (*Kansas Water Science Center, 2011*). The disasters was a serious disruption of the functioning of a community or a society causing widespread human, material, economic or environmental losses which exceed the ability of the affected community or society to cope using its own resources (*WHO, 201*).

Another fact that can show us about the unexpected flood disaster is an unusual amount of rain on Friday, low bridges and an incomplete flood wall along Sungai Jelok in Kajang, had all combined to cause one of the country's worst flash floods in decades. Drainage and Irrigation Department (DID) water resources management and hydrology division director Hanipa Mohd Noor said a 200m section of the flood wall could not be completed as traders along there were still in the midst of negotiations to move out (*The SunDaily, 2011*) The waist high water trapped nine people, three of them senior citizens, who were rescued by personnel from the Selangor Fire and Rescue Department. (*News Straits Times, 2011*).

So from that, we can conclude that this flood disaster become large one problem that difficult in solve further. So because of that, we should take precautionary measure to prevent any problems such as unexpected flood disaster. Hence, the flood disaster area forecasting can predict the place or area that frequently occur the flood disaster especially around the research area.

Furthermore, the Dempster-Shafer theory which also known as the theory of belief functions is a generalization of the Bayesian theory of subjective probability. Whereas the Bayesian theory requires probabilities for each question of interest, belief functions allow us to base degrees of belief for one question on probabilities for a related question. These degrees of belief may or may not have the mathematical properties of probabilities; how much they differ from probabilities will depend on how closely the two questions are related.

The Dempster-Shafer theory is based on two ideas which are the idea of obtaining degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence.

Dempster-Shafer evidence theory offers an alternative to traditional probabilistic theory for the mathematical representation of uncertainty. It has been widely applied in various applications. The significant innovation of Dempster-Shafer theory is that it deals with measures of "belief", and is based on the non-classical idea of "mass" as opposed to probability.

In addition, Dempster-Shafer theory does not require an assumption regarding the probability of the individual constituents of the set or interval. It has a unique advantage of making inferences from incomplete and uncertain knowledge. This is a potentially valuable tool for the evaluation of risk and reliability in engineering applications when it is not possible to obtain a precise measurement from experiments, or when knowledge is obtained from expert elicitation. An important aspect of this theory is the combination of evidence obtained from multiple sources and the modeling of conflict between them. It allows other alternative scenarios for the system, such as "unknown".

So, the implementation of this application is in order to predict the area that most frequently label as the high potential for the flood disaster. This is very important for us to know the area as we can take the precautionary in order to face this problem. So this system will be build and calculate all the possibility using the uncertainty theory which is Dempster Shafer Theory. This theory came with the best probability calculation where it can give us the best prediction area for the flood disaster. This application will calculate the probability based on several factors or attributes such as proportion of rain fall, ratio for how frequent rainy day in one month and water level of the river.

In this case, the Dempster Shafer theory is used to calculate and predict the probability of the certain area in Melaka to get involves with the flood disaster based on the inputs or attributes from the datasets which get from the Department of Irrigation and Drainage Melaka.

The main focus area for this application is around Melaka especially in the data station of telemetry online system in Durian Tunggal U/S in order to define whether that data station in high risk of flood disaster.

By using the Dempster Shafer theory this application will calculate the probability for that area to get in disaster flood. It will be develop using the Matlab software R2009b and the form will show the graph and percentage about the probability for these uncertainty problems in order to detect the high-risk area to floods.

1.2 Problems Statement(s)

This project needs to be carried out due to problems that exist in the current systems. The current system that uses the Dempster Shafer theory as their algorithms is about prediction of the cloud formation and detects the area that could be rain. So, this application is using the same theory but in different prediction.

On the hands, another the current system that already build in Malaysia only detect the flood disaster area by guessing the possibility by looking forward on the attributes like water level of the nearest river and proposition of the rain drop for that day. But that current system doesn't have right percentage that been calculated in order to know the prediction on flood forecasting. By that problem, the application that builds is to improve the current systems to predict the flood disaster area that high possibility to get involve with the flood using the Dempster Shafer theory.

Furthermore, this application will enhanced the current system by given the percentage of the probability of the research area by calculating the probability using the theory chosen for doing this project and forecast the flood disaster for the next months based to the offline data of the three months before that next month. This is important to know the next month probability especially during monsoon season that occur every years. So the resident of that research areas can prepare more if their place get involve to the flood disaster based to the percentage that have been calculated in this application.

1.3 Objectives

In certain seasons, like monsoon season, it might be necessary to do this application so that the prediction will be the precaution step for the residents to prepare to face out this serious problem. If they only know a day before that disaster occurs, so they might not have enough time do a preparation. So, this application is develop to overcome these problems. Therefore, the objectives in building this such application are:

1. To improve the current system.

The current system is about detection the cloud formation and detects the area that could be rain by using the same theory as this developing application. So the application that builds is improvement from the current system to detect the flood disaster area that high possibility to get involve with the flood using the Dempster Shafer theory. It also improves the current system for prediction of flood but only predict by randomly by looking forward to the attribute like water level and rainfall status. So, this application will enhance this current system.

2. To provide convenience to residents in the flood problem.

When the system is able to detect high-risk areas to flood, residents will be ready to face any eventualities that will occur. So this will give convenience to resident to be prepared with any possibility. The prediction from this application also can help the resident to have much time to prepare as this application will prediction the flood disaster for the whole next month using the three next month before.

1.4 Scope

The scopes for this project are divided based on target user of this application and place of research area. Firstly, the target user for this application is government in in order to predict which month that high-risk to flood and they can give information to residents to be careful and prepare in order to cope with flood. For examples, it can be used by Department of Irrigation and Drainage of Melaka State or Malaysian Meteorological Department.

Secondly, the research area for this project is data station of telemetry online system which is Durian Tunggal U/S and this data station is use as the catchment area to be the research area in order to know the water level of that area and rainfall status. For this project, the data use is twelve month from the past three years in this data station. For develop this project, this data station is using to get the water level and rainfall status that be uses as attribute for this application.

1.5 Project Significance

This project will bring benefits to the university especially UTeM as this application will be one of the university product. By this intelligent system, it can help our government in order to give fully prediction on flood forecasting specially for the Department of Irrigation and Drainage of Melaka State, where this application will give the prediction for next whole months of the next year by using data of three years before.

In a close future, some improvement can be done to make this such application more intelligent, effective and can be distributed to community like the resident itself can use this application directly. Besides that, this application can be enhance more by using the online data for doing the prediction of flood forecasting and improve more where this application can be use around Malaysia.

1.6 Expected Output

From this application, all the calculation which uses the Dempster Shafer algorithm will do on the back side of the GUI interface of this application. From the GUI interface, the target user can choose which months that they need to know about the flood forecasting and this application will give the graph and right percentage whether that area have possibility to get involve in flood disaster on that chosen month like how much percentage of that place getting flood disaster and vice versa.

Besides, only the administrator of this application can know about all the calculation that have been done in the back side and this application will have monthly maintenance as this application only use the offline data from the three years data.

1.7 Conclusion

Since Malaysia is one of the tropical countries that only have two kind of weather which are whether rain or hot, there are very important to have an application for prediction about weather forecasting. Despite our weather always changes by months to months, it quite difficult to predict on the natural disaster like flood, as it is unavoidable disaster to our country. So this project has come out with one application that can predict the flood forecasting especially around the research area.

The aim of this project is to design and develop an application that can give briefly prediction on flood forecasting in the research area for this project. Furthermore, this application also is the first application on prediction of flood forecasting by using the Dempster Shafer theory that actually is the better theory to predict on natural disaster.

This simple intelligent application will give the percentage for the flood forecasting on the chosen month of the Durian Tunggal data station and the prediction is done for a month. So the residents can have enough time to get prepare if their place have high percentage for get involve in flood disaster. Despite this application have limitation on its ability, but for sure this application can be very useful to the community around the Melaka state.

In future significance, some improvement can be added to make this such application to be more intelligently, effectively functional and can be implemented in more wide area. Some improvement likes this application can be enhance to work with the online data and the research area can be wider as this application can be use around Malaysia.