

**DECISION TREE & REGRESSION IN ACADEMIC INNOVATION AND
SCIENCE RESEARCH**

CHONG CAI NING

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS*

JUDUL: DECISION TREE AND REGRESSION IN ACADEMIC INNOVATION AND SCIENCE RESEARCH

SESI PENGAJIAN: 2014/2015

Saya CHONG CAI NING

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 tanadikan (/)

_____ SULIT (Mengandungi maklumat yang berdajah 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)

Alamat tetap: No.24, _____

Jalan Selesa 1, Taman Selesa, _____

44000, Kuala Kubu Bharu, Selangor.

Tarikh: _____

(TANDATANGAN PENYELIA)

DR. GOH ONG SING

Nama Penyelia

Tarikh: _____

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

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

DECLARATION

I hereby declare that this project report entitled
DECISION TREE & REGRESSION
IN ACADEMIC INNOVATION AND SCIENCE RESEARCH

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

STUDENT : _____ Date: _____

(CHONG CAI NING)

SUPERVISOR : _____ Date: _____

(DR. GOH ONG SING)

DEDICATION

To my beloved family, thanks for your support and caring when I am working hard to complete the project.

To my beloved supervisor, Dr. Goh Ong Sing, thanks for guiding me and also helping me again and again when I faced problems in the development of the project.

To my beloved friends who are willing to provide ideas and share their knowledge to me during the completion of this project.

ACKNOWLEDGEMENT

Firstly, I would like to thank my supervisor, Dr Goh Ong Sing for giving me a chance to develop this project. I am very appreciating with all the helps and guidance from him in my way in completing this system. Besides, I am also very glad that he gives me this opportunity to do this project as I had learnt something new too through this project.

Besides, I also would like to thank my family for giving me supports and caring while I am working hard in doing this project. In addition, I also would like to say thank you to my friends, Safar and Poh Ling for giving me ideas in doing this project.

ABSTRACT

In today's world, raising capital for early stage technologies has become a gruelling process, not only has government funding been cut, but companies and investors too. This project aims to promote academic research towards commercializing research projects through MeetInventor. MeetInventor is a portal that is being developed for the purpose to help university-based inventors bring their products to the marketplace quickly. MeetInventor portal comprises of three portals which are Education portal (MOOC), Funding portal (MeetInventor) and Marketplace portal (eStore). This study is to report a sales such as research products and training modules through Intelligent Analytic Dashboard (IAD). The data of sales will be retrieved from three databases. Furthermore, artificial intelligence techniques are being chosen to analyse the sales of products and those techniques are regression and decision tree. After the data are being analysed using these two techniques, results of the analysis are being reported in the IAD. This IAD are able to report sales, predict the future sales and also the performance of the sales in that particular day.

ABSTRAK

Pada era globalisasi, penyelidikan pembangunan teknologi bertambah sukar dan bukan sahaja pemotongan pembelanjaan telah dilakukan oleh kerajaan, malah banyak syarikat dan pelabur juga menghadapi banyak masalah yang sama. Projek ini telah dicadangkan dengan tujuan untuk mempromosikan penglibatan penyelidikan akademik kepada projek penyelidikan berbentuk komersial melalui MeetInventor. MeetInventor merupakan salah satu portal yang dibinakan bertujuan membantu semua pencipta produk di universiti yang dapat memasarkan produk mereka kepada orang ramai. Portal MeetInventor ini mempunyai tiga portal sokongan dan portal-portal tersebut adalah portal Pendidikan (MOOC), portal Dana (MeetInventor) dan portal Pemasaran (eStore). Projek ini adalah untuk membuat laporan tentang hasil jualan bagi produk penyelidikan dan modul latihan yang dipaparkan melalui 'Intelligent Analytic Dashboard' (IAD). Semua data jualan akan diambil daripada tiga pangkalan data ini. Dalam hal ini, teknik kepintaran buatan akan digunakan untuk menganalisis data jualan produk melalui teknik '*regression*' dan '*decision tree*'. Selepas semua data dianalisis melalui kedua-dua teknik ini, keputusan analisis akan dilaporkan melalui IAD. IAD ini merupakan sistem yang mampu melaporkan hasil jualan, membuat jangkaan tentang jualan pada masa depan dan juga prestasi jualan pada hari-hari tertentu.

TABLE OF CONTENTS

CHAPTER	SUBJECT	PAGE
	DECLARATION	i
	DEDICATION	ii
	ACKNOWLEDGEMENT	iii
	ABSTRACT	iv
	ABSTRAK	v
	TABLE OF CONTENTS	vi
	LIST OF TABLES	x
	LIST OF FIGURES	xii
CHAPTER I	INTRODUCTION	
	1.1 Introduction	1
	1.2 Problem Statements	3
	1.3 Objectives	4
	1.4 Scope	6
	1.5 Project Significance	6
	1.6 Excepted Output	7
	1.7 Conclusion	8

CHAPTER II	LITERATURE REVIEW AND PROJECT METHODOLOY	
	2.1 Introduction	9
	2.2 Facts and Findings	10
	2.2.1 Domain	11
	2.2.2 Existing System	11
	2.2.3 Technique	14
	2.3 Project Methodology	17
	2.4 Project Requirements	19
	2.4.1 Software Requirement	19
	2.4.2 Hardware Requirement	20
	2.5 Project Schedule and Milestones	20
	2.6 Conclusion	23
CHAPTER III	ANALYSIS	
	3.1 Introduction	24
	3.2 Problem Analysis	25
	3.3 Requirement Analysis	26
	3.3.1 Data Requirement	26
	3.3.2 Functional Requirement	27
	3.3.3 Non-Functional Requirement	31
	3.3.4 Others Requirement	33
	3.4 Conclusion	34
CHAPTER IV	DESIGN/THE PROPOSED TECHNIQUE	
	4.1 Introduction	35
	4.2 High-Level Design	36
	4.2.1 System Architecture	36
	4.2.2 User Interface	38
	4.2.3 Input Design	39
	4.2.4 Technical Design	39

4.2.5 Output Design	41
4.2.6 Database Design	42
4.2.6.1 Conceptual and Logical Database Design	42
4.3 Detailed Design	50
4.3.1 Software Design	50
4.3.2 Physical Database Design	51
4.4 Conclusion	59
CHAPTER V	IMPLEMENTATION
5.1 Introduction	60
5.2 Software Development Environment Setup	61
5.3 Software Configuration Management	70
5.3.1 Configuration Environment Setup	70
5.3.2 Version Control Procedure	70
5.4 Implementation Status	71
5.5 Conclusion	72
CHAPTER VI	TESTING AND ANALYSIS
6.1 Introduction	74
6.2 Test Plan	75
6.2.1 Test Organization	75
6.2.2 Test Environment	75
6.2.3 Test Schedule	76
6.3 Test Strategy	76
6.3.1 Classes of tests	77
6.4 Test Implementation	77
6.4.1 Experimental / Test Description	78
6.5 Test Results and Analysis	79
6.6 Conclusion	83

CHAPTER VII	PROJECT CONCLUSION	
	7.1 Introduction	84
	7.2 Observation on Weaknesses and Strengths	85
	7.2.1 Strengths of IAD	85
	7.2.2 Weakness of IAD	85
	7.3 Propositions for Improvement	86
	7.4 Project Contribution	86
	7.5 Conclusion	87
	REFERENCE	88
	APPENDIX	90

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Software Requirement	19
2.2	Hardware Requirement	20
3.1	Functional Requirement	27
3.2	Determine orders placed Use Case	28
3.3	Determine orders removed Use Case	29
3.4	View Report Use Case	29
3.5	Forecast Sales	30
3.6	View Sales performance	31
3.5	Non-functional Requirements	32
3.6	Table of Software Requirements	33
3.7	Hardware Requirements	34
4.1	mooc_commentmeta	42
4.2	mooc_comments	43
4.3	mooc_posts	44
4.4	mooc_postmeta	45
4.5	mooc_options	45
4.6	mooc_sr_abandoned_items	45
4.7	mooc_sr_order_items	46
4.8	mooc_terms	47
4.9	mooc_term_relationships	47

4.10	mooc_term_taxonomy	48
4.11	mooc_usermeta	48
4.12	mooc_users	49
4.13	mooc_woocommerce_order_itemmeta	49
4.14	mooc_woocommerce_order_items	50
5.1	Environment Setup	70
5.2	Version Control	70
5.3	Configuration Access Control	71
5.4	Backup Management	71
5.2	Analysis Status	72
5.3	Implementation Status	72
6.1	Test Schedule	76
6.2	Test Cases	78
6.3	Result of sales forecasting	79
6.3	result of test cases	83

LIST OF FIGURES

DIAGRAM	TITLE	PAGE
2.1	Homepage of Smart Dashboard	13
2.2	Homepage of Influenza Surveillance Dashboard	14
2.3	Agiles Methodology Cycle	18
2.4	Flow Chart	21
2.5	Gantt Chart	22
4.1	Three-tier architecture	36
4.2	Flow chart of placing order in each portals	37
4.3	Flow chart of displaying of total sales report in IAD	38
4.4	user interface of IAD	39
4.5	Decision tree of decision rules used	40
4.5	Data Definition Language of mooc_commentmeta	51
4.6	Data Definition Language of mooc_comments	52
4.7	Data Definition Language of mooc_posts	53
4.8	Data Definition Language of mooc_postmeta	54
4.9	Data Definition Language of mooc_option	55
4.10	Data Definition Language of mooc_terms	55
4.11	Data Definition Language of mooc_term_relationships	56
4.12	Data Definition Language of mooc_term_taxonomy	56

4.13	Data Definition Language of mooc_usermeta	57
4.14	Data Definition Language of mooc_users	57
4.15	Data Definition Language of mooc_woocommerce_order_itemmeta	58
4.5	Data Definition Language of mooc_woocommerce_order_items	58
5.1	Setup of wampserver	62
5.2	Accept the licence agreement	62
5.3	Select Destination Location	63
5.4	Select Additional Tasks	63
5.5	Ready to Install	64
5.6	End of Setup	64
5.7	Select language and Continue	66
5.8	Configure database	66
5.9	Ready to Install	67
5.10	Form for inputting information needed	68
5.11	Wordpress is installed successfully	69
6.1	Interface of IAD	80
6.2	sales performance today	80
6.3	decision tree for proving the result in Figure 6.3	81
6.4	Interface of IAD with sales today:RM1473 and Orders:2	81
6.5	Condition medium for sales performance based on RM1473 sales today and 2 orders placed	82
6.6	Decision tree shows the result in Figure 6.6 is true based on the sales today: RM1473 and order:2	82

CHAPTER I

INTRODUCTION

1.1 Introduction

In today's world, raising capital for early stage technologies has become a grueling process, not only has government funding being cut, but companies and investors have stopped taking big risks on new projects though MeetInventor. MeetInventor aims to promote academic research towards commercializing research projects. MeetInventor enables this process by encouraging researchers to self-identify, articulate their projects, request funding, and then use that funding to bring their product closer to a marketplace. MeetInventor is a portal that is being developed for the purpose to help university-based inventors quickly bring their products to the marketplace.

Besides, this MeetInventor will be related to three portals which are Education portal (MOOC), Funding portal (MeetInventor) and Marketplace portal (eStore). All these three portals have their own missions, for examples, Education portal is a crowd that educate the researchers to think through a market lens, differently from those researchers that are looking for the funding; Funding portal is to facilitate innovators to promote their projects and ensure that necessary steps are followed when the faculty, students and researchers post projects; Marketplace portal is an online shopping portal where researchers or inventors will market their products directly to consumers.

For briefly explain the mission of the portals, the Education portal which is also known as MOOC is a portal that educate students which are also the member of the portals to become an innovative and creative inventors. They are being advised and encouraged to develop products based on the market niches nowadays but not focus on the funding that they will get. This is because they are different from those researchers that are looking for funding. Moreover, they will be educated and conducted under an instructor in the development of their products. The instructor will provide them some useful ideas so they can build a more reliable products.

For the Funding portal which is also known as MeetInventor is a portal that “investor” post projects and waits for the Inventor proposals. Once projects start getting proposals, “Investor” are notified through email and within their dashboard. “Investor” can browse all proposals from their dashboard at any time, and from there, select the best candidate for each project. Inventors apply to projects will then considering the project skills and delivering time. In addition, the dashboard of this portal is the main hub for all the main actions in MeetInventor™ as “Investor” and inventors can monitor their activity, view notifications, view purchases, edit projects, edit proposals, cancel projects, cancel proposals and others. In this portal, there are sharing role capabilities. “investor” are only allowed to post projects and inventor can only apply to projects, but site owners have the option to allow sharing role capabilities. This allows the “investor” to apply projects and inventors to post projects. This capability will only be available when the user choose as investor/inventor while registration.

For the Marketplace portal which is named as eStore is a portal that all the research products will be marketed. The products that are developed from MOOC and MeetInventor are being manufactured by industry and are being sold through the Marketplace portal. Free standard delivery service is also provided for all the orders that are being placed.

As mentioned before, these three portals are somehow related to each others as the research products from MOOC and MeetInventor will be manufactured and marketed through eStore once an order is placed. Therefore, the users will need to

view the total sales report of their products from these portals. Therefore, a reporting should be done through multiple databases of these three portals.

In this project, we are going to develop an Intelligent Analytic Dashboard (IAD). Intelligent Analytic Dashboard is a dashboard that is developed for the purpose of reporting the sales reports. This dashboard is unique from other reporting dashboard before as this dashboard is developed to report the total sales reports of MOOC, MeetInventor and eStore. So, the data of sales will be retrieved from three databases especially which are MOOC, MeetInventor and eStore respectively. Because of each databases consists of large number of data, optimization is required to be done on those databases in order to retrieve and analyse the data from those three databases for the use of reporting. In this project, too, decision tree, a tree-like analyzing model artificial intelligent technique, is decided to be used in analyzing the data model. Besides, regression method is also being used to predict the future sales of those three portals. After optimization is successfully carried out, results will be reported through Intelligent Analytic Dashboard.

In conclusion, the data of sales reports are retrieved mainly from the completed orders placed in Woocommerce tables and some other tables such as posts, postmeta and others. The data that are being retrieved are large in number and is then being optimized and analyzed through the decision tree method and also regression method. The optimized data are then being displayed as sales reports for the reference of the users.

1.2 Problem Statements

Here we discuss the problems that will be faced during the development of Intelligent Analytic Dashboard and the needs to develop this dashboard:

- i. Inventors or university need a single dashboard to view their sales information of their products and reporting tool.

- Since this project is developed for the university, they will need to view all the sales reports of these three portals. As for demand, a single dashboard is needed to be developed for the purpose to display the sales report instead of developing dashboard for each portals. Although this is convenient for the inventors or university in viewing the sales information, the developing of single dashboard to display the sales information of those three portals maybe difficult as important data should be known and accurate when the data is retrieved from those three databases. This step is very important in order to give the correct sales information to the inventors or university.
- ii. Inventors or investors need to know the future sales information of each portals
- Besides the updated sales information from each portals, the inventors and investors may want to view the future sales information. By knowing the forecasting demand, they can estimate how the sales of their will be. Besides, this forecasting sales function plays an important role to ensure that the sales of products to reach their expected goals.

1.3 Objective

The problem stated from previous subtitle probably can be solved by developing the Intelligent Analytic Dashboard. This is because the development of Intelligent Analytic Dashboard is done with the following objectives:

- i. To build an Intelligent Analytic Dashboard (IAD)
 - Intelligent Analytic Dashboard is being developed to report the sales report of the users of MOOC, MeetInventor and EStore portals. Data will be extracted from each databases such as mooc, meetinventor and eStore and are being analysed using the chosen artificial intelligence technique which are regression method and decision tree method. After the data are being analyzed, the analyzed data will then display on the IAD. Therefore, as conclusion, the IAD is a system that report the total sales reports of multiple databases.

- ii. To predict the future sales through regression technique
 - Regression is one of the techniques that is being used in this project. By using this technique, we are going to predict the future sales for those Eduction portal, Funding Portal and MarketPlace portal.

- iii. To classify the sales performance by using decision tree.
 - Decision tree is a tree-like analytic technique that is easy to understand. It can help to classify today sales performance in a more effective way.

1.4 Scope

i. User:

This system is developed mainly for users as below:

- Inventors or investors
- Since those three portals are developed under a project in the university, therefore those inventors or university will prefer to have a review of the sales information of those three portals.

ii. Platform

- Web-based
- This system is a web-based system as user can surf for the website by internet browser. They can look through the reporting of their sales information of their products through the website.

1.5 Project Significance

In this project, Intelligent Analytic Dashboard is built in order to report the sales information of three different portals which are MOOC, MeetInventor and EStore. Basically, this three portals have different databases and each of them have very large number of data. Therefore, decision tree which is useful tree-like analytical data mining model is being used to optimise the queries while extracting the data from each of the databases. The results after the optimisation is then display through the Intelligent Analytic Dashboard.

In addition, this project is mainly developed to the users such as inventors who are the members of the Mooc, MeetInventor and EStore portals. As the member of MOOC portal, he or she will be trained to be an inventor and teach them the way to market their products. For the member of the MeetInventor portal, they are encouraged to propose their projects and then investors who are interested with their

products will help them to market through the EStore portal. Therefore, they will need to know the sales information of their products. For their ease to view their sales information, Intelligent Analytic Dashboard is required to be developed.

For the business value, this project is unique and quite significant for those businessman or dashboard users who decided to do reporting through multiple databases. As we mentioned just now, multiple databases will have large number of databases. Besides, suitable data mining methods should be determined wisely and suitably. This is because data will be reported wrongly if we choose the wrong techniques. Therefore, as an example or the beginning, decision tree is chosen to analyse the sales information of each portals and the Intelligent Analytic Dashboard will prove that the reporting can be done through multiple databases with large number of data.

To increase the significance of the development of Intelligent Analytic Dashboard, this features of this systems are as below:

- Able to analyse the large number of data from multiple databases
- Able to obtain the up-to-date sales information when there are changes in the sales of each portals
- Able to report the sales information in each of the portals
- Able to report the forecasted sales information in each of the portals

As conclusion, the development of Intelligent Analytic Dashboard is a very significant project for the purpose to report any information from multiple databases with large number of data.

1.6 Expected Output

A automatic up-to-date sales information of portals such as mooc, meetinventor and eStore can be reported by Intelligent Analytic Dashboard. When

the customer or member of one of the portals places orders or the orders are being cancelled, the sales of the portals will be changed. Therefore, the Intelligent Analytic Dashboard will be able to detect the changes in sales and then display the updated sales information.

1.7 Conclusion

In this chapter, we have described the problems that are faced by the investors and inventors nowadays besides the purpose of developing this project. The problem statements describe the problems that are faced in this project. For the objective and scope, they are being determined to emphasize the needs in the development of reporting for those separated databases which are mooc, meetinventor and eStore.

The Intelligent Analytic Dashboard will gain the data from the woocommerce order tables from those three different databases and display the up-to-date information. Once there are orders placed in any portals, the new information will be added to the existing data and also is displayed.

For the next chapter, we will describe the literature review and project methodology. In that chapter, the comparison between existing dashboard and smart-reporting dashboard will be discussed. In addition, software and hardware requirements that are required in the development of Intelligent Analytic Dashboard are also being discussed in the next chapter.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

In this chapter, we will discuss about the literature review and methodology that is used in the development of Intelligent Analytic Dashboard.

According to University of North Carolina at Chapel Hill, literature review is simple summary for information that had been published in a particular area. This literature review is also a combination of summary and synthesis in which the summary is a recap for important information while the synthesis is the re-organization of information. By doing the literature review, the writer may gain new ideas from the old materials or improve the old materials with their new ideas. Moreover, it may trace the intellectual progression of the field, including major debates and may evaluate the sources besides advise the reader on the most pertinent depending on the situation.

In this chapter, a few sections will be discussed too, Those sections are such as facts and findings, project requirements and project schedule and milestones. In the other hand, methodology also being discussed in this chapter. There are many types of methodologies such as Waterfall model, agile management, Spiral model and Rational Unified Process. In this project, selecting a suitable methodology is very important as a suitable methodology can reduce the risk level more effectively.