

INTELLIGENT HEART DISEASE PREDICTION SYSTEM

(IHDPS)

JENNY LIM YEN YEN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

**INTELLIGENT HEART DISEASE PREDICTION SYSTEM
(IHDPS)**

JENNY LIM YEN YEN

**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**

2011

BORANG PENGESAHAN STATUS TESIS

JUDUL: INTELLIGENT HEART DISEASE PREDICTION SYSTEM (IHDPS)

SESI PENGAJIAN: 2010/2011

Saya JENNY LIM YEN YEN
(HURUF BESAR)

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: J6596, Tmn Sri Merlimau
77300 Merlimau, Melaka, Malaysia

DR. CHOO YUN HUOY

Nama Penyelia

Tarikh : 13/7/2011

Tarikh : 13/7/2011

CATATAN: *Tesis dimaksudkan sebagai Laporan 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
INTELLIGENT HEART DISEASE PREDICTION SYSTEM
(IHDPS)

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

STUDENT :  Date: 13/7/2011
(JENNY LIM YEN YEN)

SUPERVISOR :  Date: 13/7/2011
(DR. CHOO YUN HUOY)

DEDICATION

This report is dedicated to my beloved parents; friends and supervisors who have provided encouragement and guidance throughout the process to complete the system and report.

ABSTRACT

The objective of this project is to develop a web-based Naïve Bayes decision support system namely Intelligent Heart Disease Prediction System (IHDPS) to predict heart disease risk level of the user. This three-tier application is developed based on object-oriented analysis and design (OOAD) methodology. WAMP Server Version 2.0 is used to develop IHDPS. IHDPS is tested via white-box and black-box strategy. There are five types of risk level can be provided for the heart disease prediction result which are Normal, Low, Medium, High and Very High. Besides, there are eight prediction parameters which can influence the heart disease risk level result which are Age, Gender, Chest Pain Type, Resting Blood Pressure, Serum Cholesterol, Fasting Blood Sugar, Resting Electrocardiography Results, and Thallium Scan. There are 303 data instances and 14 attributes exist in the system database to create a knowledge base for the Simple Naïve Bayes classifier to carry out heart disease risk level prediction. The strengths of IHDPS are this system applies Simple Naïve Bayes classifier for heart disease risk level prediction module, provides useful information about heart disease, can be accessed by users at anytime and also the user interfaces of this system are simple and user friendly. On the other hand, the weaknesses of IHDPS are browser compatibility, lack of the system security feature, and also this system has limited functionality and limited interaction with users. IHDPS can contribute on increasing awareness and vigilance about the heart disease to the users. There are some future improvements for IHDPS which are solving browser compatibility issues, this system can add in e-mail or SMS notification and printing function, and also can be upgraded to provide other kind of disease diagnosis such as diabetes to become a higher standard health care system.

ABSTRAK

Objektif projek ini ialah membina satu sistem penyokongan keputusan yang berasaskan laman untuk meramal tahap risiko penyakit jantung pengguna. Sistem ini dikenali sebagai *Intelligent Heart Disease Prediction System (IHDPS)*. IHDPS dibangunkan berasaskan metodologi *object-oriented analysis and design (OOAD)*. IHDPS merupakan satu laman web aplikasi yang berasaskan seni bina *three-tier* iaitu *Presentation Tier, Application Tier* dan *Data Tier*. *WAMP Server Version 2.0* dipasang untuk membangun sistem IHDPS kerana ia merupakan pakej penuh yang merangkumi pakej perisian pemasangan seperti *Apache, PHP* dan *MySQL*. IHDPS diuji melalui strategi kotak putih dan kotak hitam. Sistem ini mempunyai lima jenis tahap risiko untuk hasil ramalan penyakit jantung iaitu normal, rendah, sederhana, tinggi dan sangat tinggi. Selain itu, sistem ini juga mempunyai lapan ramalan parameter yang boleh mempengaruhi hasil ramalan tahap risiko penyakit jantung iaitu Umur, Jantina, Jenis Sakit Dada, Tekanan Darah Rehat, Serum Kolesterol, Gula Darah Puasa, Keputusan Elektrokardiografi Rehat, dan Talium Scan. Sistem ini mempunyai 303 data dan 14 atribut dalam pangkalan data untuk membina satu asas pengetahuan supaya pengelas *Simple Naive Bayes* dapat menjalani proses ramalan tahap risiko penyakit jantung. Kekuatan sistem ini adalah sistem ini menggunakan pengelas *Simple Naive Bayes* untuk modul ramalan tahap risiko penyakit jantung, menyediakan informasi yang berguna berkenaan penyakit jantung, boleh digunakan oleh pengguna pada bila-bila masa dan *user interfaces* sistem ini mudah dan mesra digunakan. Sebaliknya, kelemahan sistem ini ialah keserasian *browser*, kekurangan sistem keselamatan, dan juga kekurangan sistem fungsi. IHDPS memberi sumbangan dengan meningkatkan kesedaran berkenaan penyakit jantung kepada pengguna. Sistem ini mempunyai peningkatan untuk masa depan seperti penyelesaian isu keserasian *browser*, sistem ini boleh menambah *e-mail* atau *SMS* fungsi pemberitahuan dan fungsi percetakan, dan sistem ini boleh dinaikkan taraf untuk menyediakan ramalan penyakit lain seperti diabetes untuk menjadi sistem penjagaan kesihatan yang bertaraf lebih tinggi.

ACKNOWLEDGEMENTS

This project is completed with the encouragement and help of many people. The author would like to express sincere gratitude to all those who have contributed directly and indirectly in completing this project.

Firstly, the author would like to thank and express her deep gratitude to her project supervisor, Dr. Choo Yun Huoy for giving her assistance, patience, guidance and continuous encouragement to complete this project successfully. Her valuable guidance has been of great value for me to accomplish this project.

Besides that, the author would also like to thank Dr. Azah Kamilah for offering her the brilliant suggestions and comments to make her project more perfect.

Last but not least, she would also like to thank her classmates and roommates for their continuing supports and encouragement along the way to accomplish and complete her project who always give moral support when she was frustrated and discouraged.

Finally, the author would like to express her sincere gratitude to her beloved parents for their inspiration, benevolence and supporting her spiritually throughout her studies.

Thank you very much to all from the bottom of my heart!

TABLE OF CONTENTS

CHAPTER	SUBJECT	PAGE
	TITLE	i
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENTS	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xii
	LIST OF FIGURES	xiv
	LIST OF ABBREVIATIONS	xix
CHAPTER I	INTRODUCTION	
	1.1 Project Background	1
	1.2 Problem Statements	2
	1.3 Objective	2
	1.4 Scope	2
	1.4.1 Target User	3
	1.4.2 Modules	3
	1.4.3 Platform	4
	1.5 Project Significance	4
	1.6 Expected Output	4
	1.7 Conclusion	5
CHAPTER II	LITERATURE REVIEW AND PROJECT METHODOLOGY	
	2.1 Introduction	7
	2.2 Facts and Findings (based on topic)	8

2.2.1	Domain	8
2.2.1.1	Heart Disease Diagnosis System	8
2.2.1.2	Naïve Bayes Classifier	9
2.2.1.3	Heart Disease Risk Factor	9
2.2.2	Existing System	11
2.2.2.1	Heart Attack Prediction System Using Neural Network	11
2.2.2.2	A Fuzzy Expert System for Heart Disease Diagnosis	14
2.2.3	Technique	19
2.2.3.1	Naïve Bayes	20
2.2.3.2	Rules Set Classifier	21
2.2.3.3	Artificial Neural Network	21
2.2.3.4	Fuzzy Logic	23
2.3	Project Methodology	24
2.4	Project Requirements	25
2.4.1	Software Requirement	25
2.4.2	Hardware Requirement	27
2.4.3	Other Requirement	28
2.5	Project Schedule and Milestones	28
2.6	Conclusion	28

CHAPTER III ANALYSIS

3.1	Introduction	30
3.2	Problem Analysis	30
3.3	Requirement Analysis	31
3.3.1	Data Requirement	32
3.3.2	Functional Requirement	35
3.3.3	Non-functional Requirement	37
3.3.4	Others Requirement	38
3.3.4.1	Software Requirement	38
3.3.4.2	Hardware Requirement	41

3.3.4.3 Network Requirement	41
3.4 Conclusion	41
CHAPTER IV DESIGN	
4.1 Introduction	43
4.2 High-Level Design	44
4.2.1 System Architecture	44
4.2.2 User Interface Design	45
4.2.2.1 Navigation Design	58
4.2.2.2 Input Design for IHDPS	59
4.2.2.3 Technical Design	61
4.2.2.4 Output Design	62
4.2.3 Database Design	63
4.3 Detail Design	64
4.3.1 Software Hardware Design	65
4.3.2 Physical Database Design	68
4.4 Conclusion	71
CHAPTER V IMPLEMENTATION	
5.1 Introduction	72
5.2 Software Development Environment Setup	72
5.3 Software Configuration Management	73
5.3.1 Software Configuration Environment Setup	73
5.3.2 Version Control Procedure	81
5.4 Implementation Status	82
5.5 Conclusion	84
CHAPTER VI TESTING	
6.1 Introduction	85
6.2 Test Plan	85
6.2.1 Test Organization	86
6.2.2 Test Environment	86

6.2.3	Test Schedule	87
6.3	Test Strategy	88
6.3.1	Classes of Test	90
6.3.1.1	Unit Testing	90
6.3.1.2	Integration Testing	91
6.3.1.3	System Testing	91
6.3.1.4	User Acceptance Testing	92
6.4	Test Implementation	92
6.4.1	Test Description	92
6.4.2	Test Data	100
6.5	Test Results and Analysis	102
6.6	Conclusion	107
 CHAPTER VII PROJECT CONCLUSION		
7.1	Observation on Weaknesses and Strengths	108
7.1.1	Weaknesses	109
7.1.2	Strengths	109
7.2	Proposition for Improvement	110
7.3	Contributions	111
7.4	Conclusion	111

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Classification of Systolic Blood Pressure	15
2.2	Classification of Heart Disease Risk Level	16
3.1	member	33
3.2	administrator	33
3.3	prediction	33
3.4	IHDPS Functional Requirements	35
3.5	IHDPS Non-Functional Requirements	38
4.1	Input Design for IHDPS	59
4.2	Output Design for IHDPS	62
4.3	member	68
4.4	administrator	69
4.5	prediction	69
5.1	Procedure and Control Source Code Version of IHDPS	81
5.2	Implementation Status of IHDPS	82
6.1	Hardware and Software Configurations of IHDPS	87
6.2	Test Schedule of IHDPS	88
6.3	Types of Testing for White-Box and Black-Box Testing Strategy	89
6.4	Unit Testing for User Login Module	93

6.5	Unit Testing for User Sign-Up Module	94
6.6	Unit Testing for Forgot Password Module	95
6.7	Unit Testing for Personal Profile Management Module	96
6.8	Unit Testing for Member Account Management Module	97
6.9	Unit Testing for Heart Disease Risk Level Prediction Module	98
6.10	System Testing for IHDPS	99
6.11	Test Result for Each Test Case	99
6.12	Test Data for IHDPS	100
6.13	Test Results for User Login Module	102
6.14	Test Results for User Sign-Up Module	103
6.15	Test Results for Forgot Password Module	104
6.16	Test Results for Personal Profile Management Module	104
6.17	Test Results for Member Account Management Module	106
6.18	Test Results for Heart Disease Risk Level Prediction Module	106

LIST OF FIGURES

DIAGRAM	TITLE	PAGE
2.1	Heart Attack Prediction with No Risk Level	12
2.2	Heart Attack Prediction with High Risk Level	13
2.3	Heart Attack Prediction with Medium Risk Level	13
2.4	Membership Functions of Systolic Blood Pressure	16
2.5	Membership Function of Result	17
2.6	Rule Base of the System	18
2.7	Membership function of input variable “Age”	23
2.8	Membership function of output variable “Result”	24
2.9	Rule Base of the System	24
3.1	Conceptual Database Design for IHDPS	32
3.2	Use Case Diagram of the System (IHDPS)	34
3.3	Use Case Diagram of IHDPS for User	36
3.4	Use Case Diagram of IHDPS for Administrator	37
4.1	System Architecture for IHDPS	45
4.2	Login Page of IHDPS	46
4.3	Navigation Flow of Login Module	47
4.4	Sign Up Page of IHDPS	47
4.5	Navigation Flow of Sign-Up Module	48
4.6	Secure Answer Confirmation Page	49
4.7	Retrieve Forgot Password Page	49

4.8	Password Recovery Page	50
4.9	Navigation Flow of Forgot Password Module	51
4.10	User Home Page	51
4.11	Navigation Flow of Edit User Personal Profile Module	52
4.12	Edit User Personal Profile Page	53
4.13	Heart Disease Risk Level Prediction Page	54
4.14	Navigation Flow of Heart Disease Risk Level Prediction Module	55
4.15	Administrator Home Page	55
4.16	Edit Administrator Personal Profile Page	56
4.17	Delete User Account Page	56
4.18	Navigation Flow of Delete User Account Module	57
4.19	The Navigation Design for the IHDPS	58
4.20	Conceptual Database Design for IHDPS	64

LIST OF ABBREVIATIONS

AI	-	Artificial Intelligence
ERD	-	Entity Relational Diagram
IHDPS	-	Intelligent Heart Disease System
LAN	-	Local Area Network
MySQL	-	My Structure Query Language
NBC	-	Naïve Bayes Classifier
OOAD	-	Object-oriented Analysis and Design
OOD	-	Object-oriented Design
OOP	-	Object-oriented Programming
PC	-	Personal Computer
PHP	-	PHP: Hypertext Preprocessor
RDMS	-	Relational Database Management System
UML	-	Unified Modeling Language

CHAPTER I

INTRODUCTION

1.1 Project Background

Heart disease is an umbrella term used to refer and describe variety types of cardiovascular diseases that affect one or more of the components of the heart. The term heart disease is often used interchangeably with cardiovascular disease. Diseases associated with the term of heart disease can be possibly acute or fatal, but also can be diagnosed, predicted or prevented. The various diseases that fall under the umbrella term of heart disease include coronary artery disease (CAD); heart rhythm problems (arrhythmias); peripheral artery disease, heart attack; heart infections; and congenital heart defects. (K.Srinivas, B.Kahvita Rani and Dr.A.Govrdhan, “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attack” K.Srinivas et al. / (IJCSSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255.)

Most countries face high and increasing rates of cardiovascular disease. Nowadays, heart disease is on the rise in Malaysia in spite of improvement of health facilities and services. According to National Heart Association of Malaysia, heart disease is the second leading killer in 2006, accounting for 15.5 percent of those who died in government hospitals. By 2010, heart diseases are projected to be the leading cause of death in Malaysia and other developing countries. (“Heart Diseases on the Rise, Second Leading Killer”, 7 April 2008, <http://www.malaysianheart.org/article.php?aid=35>)

1.2 Problem Statements

Based on the project background above, the main problem of the project is most of the medical diagnosis web applications currently undergo medical diagnosis by providing professional online medical consultation. Professional online medical consultation is unable to produce the medical diagnosis result immediately due to the online medical practitioners not always online 24 hours. Hence, this problem can be solved by developing an online medical diagnosis system that is able to provide diagnosis result instantly.

1.3 Objective

As a guideline to develop this system, the objectives of this project are mentioned as below:

- To propose a web-based medical system which is known as Intelligent Heart Disease Prediction System (IHDPS) that applies the Artificial Intelligent (AI) technique to predict heart disease risk level.
- To apply Simple Naïve Bayes classifier technique in the heart disease risk level prediction module of IHDPS.
- To design and implement all the design elements such as the system architecture, user interfaces, database, input and output of the system in order to explain the process flow of the system.
- To perform testing to ensure that the system function properly and free from error.

1.4 Scope

This part explains about the scope involved in the project Intelligent Heart Disease Prediction System (IHDPS). The scope of IHDPS will be mentioned in three subparts which are scope for target user, modules and platform.

1.4.1 Target User

IHDPS is available for the system administrator and anyone who access to the internet. The administrator of the IHDPS is only responsible to manage his/her personal account and delete the users account. Anyone who wants to get access IHDPS need to register as member of IHDPS for heart disease risk level prediction.

1.4.2 Modules

Intelligent Heart Disease Prediction System (IHDPS) will contain modules as mentioned in the following:

- a) User Login Module: This module manages the member authentication through the user login. This module can be used by administrator and users.
- b) User Registration Module: This module will be used to manage new user registration.
- c) Forgot Password Module: This module enables members of this system to recover and retrieve their forgotten password.
- d) Personal Profile Management Module: This module will be used to manage user's personal profile. This module is available for administrators and members.

- e) Heart Disease Risk Level Prediction Module: This module will predict and analysis the risk level of having heart disease by using Naïve Bayes classification method based on historical heart disease database. This module is only available for members of this system.

- f) Member Account Management Module: This module is only available for the system administrators to delete member account of IHDPS.

1.4.3 Platform

To access Intelligent Heart Disease Prediction System (IHDPS), user must have network connection and internet browsers such as Mozilla Firefox, Internet Explorer, and etc. Furthermore, WampServer is used as a web server and the operating system that is used to develop this system is Windows Vista. Moreover, this system is developed by using PHP language and MySQL as database system.

1.5 Project Significance

Intelligent Heart Disease Prediction System (IHDPS) is a web-based medical prediction and decision support system in prediction of heart disease risk. This system only lists an estimated order of likelihood and probabilities of heart disease risk levels based on historical heart disease database. Therefore, this system is not really replacing doctors but is being used to help users in making deduction of probable diagnosis about the heart disease.

This system is useful and important for user because the prediction of heart disease or probable heart disease diagnosis assumption can create awareness for the users and encourage them to seek the medical practitioners. This system is available for all of the users who can access to the internet. With this online Intelligent Heart Disease Prediction System, users can carry out their self-prediction process via internet at anytime. Hence it can help users to save the time.

1.6 Expected Output

The expected outputs for the project of Intelligent Heart Disease Prediction System are as below:

- System Administrator is responsible to delete users account and update his/her own personal details.
- Users can undergo the heart disease risk prediction so that they can seek the medical professional or doctor for further heart disease diagnosis information.
- Output or result of the heart disease risk level is normal, low, medium, high and very high.

1.7 Conclusion

As a conclusion, Intelligent Heart Disease Prediction System (IHDPS) is a web-based medical prediction and decision support system in prediction of heart disease risk. This system only lists an estimated order of likelihood and probabilities of heart disease risk levels based on historical heart disease database. Intelligent Heart Disease Prediction System (IHDPS) is implemented with the data mining classification technique which is known as Naïve Bayes for the prediction of heart disease events.

This system is useful and important for user because the prediction of heart disease or probable heart disease diagnosis assumption can create awareness for the users and encourage them to seek the medical practitioners. With this online Intelligent Heart Disease Prediction System, users can carry out their self-prediction process via internet at anytime. Hence it can help users to save the time.

Next, chapter II is going to list about the literature review and project methodology.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

In chapter 2, literature review and project methodology will be discussed. The purpose of writing literature review is to convey what knowledge and ideas have been established on my system project, and what of its strengths and weaknesses are. Literature review can be done by searching and studying the relevant sources such as references books, web journals, web documents and others. The outcome of the literature review is taken into consideration in developing this project after the project methodology is chosen to accomplish this project.

The fact and findings part is divided into three sections which are Domain, Existing System and Technique. The domain will describe what type or the criteria of the system to be built and the field of that system. On the other hand, Existing System will discuss and state about my approaches and related research, references, case study and other findings that relate to Intelligent Heart Disease Prediction System (IHDPS). Technique is the section which states other approaches than I use that are also applicable and related.

Project methodology section will describe the methodology that will be implemented in Intelligent Heart Disease Prediction System (IHDPS). Besides, it will also describe the activities in every stage of Intelligent Heart Disease Prediction System (IHDPS). Next, project requirements section will discuss about software requirement, hardware requirement and other requirements that will be used to

accomplish this project. Lastly, project schedule and milestones section will explain my actions plan prior to the end of the project. Hence, this part will list and describe stage by stage of the activities of this project.

2.2 Facts and Findings

Facts and Findings part will discuss about the references and case study of Intelligent Heart Disease Prediction System (IHDPS) in three viewpoints which are domain of the system, existing system and technique used in the system.

2.2.1 Domain

Intelligent Heart Disease Prediction System (IHDPS) is mainly developed or designed to predict and diagnose heart disease risk level of people by implementing with the data mining classification technique which is known as Naïve Bayes. Therefore, the domains for IHDPS can be discussed in three sub parts which are heart disease diagnosis system, Naïve Bayes classifier, and heart disease risk factor.

2.2.1.1 Heart Disease Diagnosis System

In most of countries, heart disease is a major cause of morbidity and mortality. Heart disease diagnosis is extremely important but is a complicated task that should be performed accurately and efficiently.

2.2.1.2 Naïve Bayes Classifier

Naïve Bayes classifier is the simplest form of Bayesian network. The Naïve Bayes classifier use the Naïve Bayes formula to calculate the probability of each