

HEART DISEASES PREDICTION USING DATA MINING TECHNIQUES

**VI SEMESTER
IT8611 MINI PROJECT REPORT**

Submitted by
YUVARAJ R (312417205058)

**BACHELOR OF TECHNOLOGY
in
INFORMATION TECHNOLOGY**



**St. JOSEPH'S INSTITUTE OF TECHNOLOGY
CHENNAI 600 119**



**ANNA UNIVERSITY, CHENNAI 600 025
APRIL 2020**

ANNA UNIVERSITY: CHENNAI 600 025



BONAFIDE CERTIFICATE

Certified that this project report “**HEART DISEASES PREDICT DATA MINING TECHNIQUES**” is the bonafide work of **YUVARAJ R (312417205058)** who carried out the project work under my supervision, for the partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Information Technology.

SIGNATURE

Dr. R. PRISCILLA, M.E., Ph.D

Professor

HOD (Lab Affairs)

Department of Information Technology

St.Joseph's Institute of Technology

Old Mamallapuram Road

Chennai-600119

SIGNATURE

Dr.A. MENAKA PUSHPA,B.E,M.E.,Ph.D.

Associate Professor

SUPERVISOR

Department of Information Technology

St.Joseph's Institute of Technology

Old Mamallapuram Road

Chennai-600119

Submitted for the Viva-Voce held on _____

(INTERNAL EXAMINER)

(EXTERNAL EXAMINER)

CERTIFICATE OF EVALUATION

College Name : St. Joseph's Institute of Technology

Branch & Semester : Information Technology (VI)

S.No.	NAMES OF STUDENTS	TITLE OF THE PROJECT	NAME OF THE SUPERVISOR WITH DESIGNATION
1.	MANOJ M (312417205026)	HEART	Dr.A.MENAKA PUSHPA,M.E.,Ph.D. Associate Professor
2.		DISEASES	
		PREDICTION	
	YUVARAJ R (312417205058)	USING DATA	
		MINING	
3.	ZAKIR	TECHNIQUES	
	HUSSAIN B (312417205059)		

The report of the project work submitted by the above students in partial fulfillment for the award of Bachelor of Technology degree in Information Technology of Anna University were evaluated and confirmed to be reports of the work done by the above students and then evaluated.

(INTERNAL EXAMINER)

(EXTERNAL EXAMINER)

ABSTRACT

Data Mining is the most popular knowledge extraction method for knowledge discovery (KDD). The healthcare industry contains a huge amount of data. But most of it is not effectively used. Heart disease is one of the main reasons for the death of people in the world. Nearly 47% of all deaths are caused by heart diseases. We use four algorithms including Decision Tree, Hoeffding Tree, Naïve Bayes and Sequential Minimal optimization to predict heart diseases. Accuracy of the prediction level is high when using more number of attributes. Using ROC curve, the prediction technique is identified effectively. Our aim is to perform predictive analysis using these data mining techniques on heart diseases and conclude which techniques are effective and efficient.

LIST OF FIGURES

Figure No.	Figure Name	Page No
Fig 1.1	Types Of Diseases chart	4
Fig 1.2	Naive Bayes	6
Fig 1.3	Hoeffding Tree Algorithm	7
Fig 1.4	Hoeffding Tree	9
Fig 1.5	Decision Table	12
Fig 1.6	SMO Equation	12
Fig 1.7	SMO Algorithm	13
Fig 1.8	SMO Margin Support	13
Fig 1.9	SMO Linear Kernel	14
Fig 3.1	Weka GUI	17
Fig 4.1	Training Phase	18
Fig 4.2	Testing Phase	19
Fig 6.1	Visualizing Attributes	25
Fig 6.2	Class Attribute	26
Fig 6.3	Class Visualization	26
Fig 6.4	Kappa Statistics	27
Fig 6.5	Execution Time	27
Fig 6.6	Root Mean Squared Error	28
Fig 6.7	Mean Absolute Error	28
Fig 6.8	Root Relative Squared Error	28
Fig 6.9	Correctly Classified Instances	28
Fig 6.10	Incorrectly Classified Instances	29
Fig 6.11	ROC Curve	30
Fig 6.12	Knowledge Flow	31

LIST OF TABLES

Table No	Table Name	Page No
Table 6.1	Confusion Matrix For Decision Table	22
Table 6.2	Confusion Matrix For Hoeffding Tree	22
Table 6.3	Confusion Matrix For Naive Bayes	22
Table 6.4	Confusion Matrix For SMO	23
Table 6.5	Performance of Attributes	23
Table 6.6	Attributes Information	25
Table 6.7	Attributes Types	25

LIST OF ABBREVIATIONS

Abbreviation	Expansion	Page No
ARFF	Attribute-Relation File Format	6
WEKA	Waikato Environment for Knowledge Analysis	17
SMO	Sequential Minimal Optimization	20
ROC	Receiver Operating Characteristic Curve	30

TABLE OF CONTENT

CHAPTER	TITLE	PAGE NO
	ABSTRACT	iv
	LIST OF FIGURES	v
	LIST OF TABLES	vi
	LIST OF ABBREVIATIONS	vii
1	INTRODUCTION	1
	1.1 DATA MINING	1
	1.2 PROBLEM STATEMENT	1
	1.3 HEART DISEASE	2
	1.3.1 Symptoms	2
	1.3.2 Types of Heart Disease	3
	1.4 OBJECTIVE	4
	1.4.1 General Objective	4
	1.4.2 Specific Objective	4
	1.5 SCOPE OF THE PROJECT	5
	1.6 DATA MINING TECHNIQUES	5
2	LITERATURE SURVEY	15
3	SYSTEM ANALYSIS	16
	3.1 EXISTING SYSTEM	16
	3.2 PROPOSED SYSTEM	16
	3.3 SYSTEM REQUIREMENTS	16
	3.3.1 Hardware Requirements	16
	3.3.2 Software Requirements	16

4	SYSTEM DESIGN	18
	4.1 ARCHITECTURE DESIGN	18
5	SYSTEM IMPLEMENTATION	20
	5.1 LIST OF MODULES	20
	5.2 MODULE DESCRIPTION	20
	5.3 SUPERVISED LEARNING	20
6	RESULTS AND DISCUSSION	22
7	CONCLUSION	33
	7.1 FUTURE SCOPE	32