**EFFICIENT HEART DISEASE PREDICTION SYSTEM USING DATA MINING ALGORITHMS AND DETERMINING THEIR ACCURACIES**

**A PROJECT REPORT**

***Submitted by***

**M. MANIKANDAN (810015104714)**

**R. SAMPATH RAM (810015104717)**

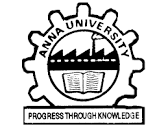
***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**



**UNIVERSITY COLLEGE OF ENGINEERING, BIT CAMPUS, TIRUCHIRAPPALLI**

**ANNA UNIVERSITY: CHENNAI 600 025**

**APRIL 2019**

**ANNA UNIVERSITY: CHENNAI 600 025**

**BONAFIDE CERTIFICATE**

Certified that this project report “**EFFICIENT HEART DISEASE PREDICTION SYSTEM USING DATA MINING ALGORITHMS AND DETERMINING THEIR ACCURACIES**” is the bonafide work of “**M. MANIKANDAN (810015104714) and** **R**. **SAMPATH RAM (810015104717)”** who carried out the project work under my supervision.

**Mr. D. VENKATESAN Mrs. C. USHA RANI**

Assistant Professor & Head Assistant Professor

Department of CSE Department of CSE

Anna University, BIT Campus Anna University, BIT Campus

University College of Engineering University College of Engineering

Tiruchirappalli - 620 024 Tiruchirappalli - 620 024

Submitted for the ANNA UNIVERSITY Viva-voce Examination held on

……………….

**Internal Examiner External Examiner**

**DECLARATION**

We hereby declare that the work entitled **“EFFICIENT HEART DISEASE PREDICTION SYSTEM USING DATA MINING ALGORITHMS AND DETERMINING THEIR ACCURACIES”** is submitted in partial fulfillment of the requirements for the award of the degree in B.E.-Computer Science and Engineering, University College of Engineering, BIT Campus, Tiruchirappalli, is a record of our own work carried out by us during the academic year 2018-2019 under the guidance of **Mrs. C. Usha Rani,** Assistant Professor, Department of Computer Science and Engineering, University College of Engineering, BIT Campus, Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places. The matter embodied in this work is original and has not been submitted for the award of any other Degree, either in this or any other University.

**M.MANIKANDAN**

**R. SAMPATH RAM**

I certify that the declaration made above by the candidate is true.

Signature of the Project Guide

**Mrs. C. Usha Rani**

Assistant Professor

Department of Computer Science and Engineering

University College of Engineering (BIT Campus)-Tiruchirappalli – 620 024

**ACKNOWLEDGEMENT**

First and foremost, of all, we would like to thank our beloved Parents and God Almighty for giving us the strength, knowledge, ability and opportunity to undertake this project study and to persevere and complete it satisfactorily.

We would like to convey our thanks to our honorable Dean **Dr. T. SENTHIL KUMAR**, Associate Professor, for having provided us with all required facilities not only to complete our project without hurdles but also for the entire course of study.

We extremely indebted to our Head of the Department **Mr. D. VENKATESAN**, Assistant Professor, Department of Computer Science and Engineering, for the devoted attention shown upon us and making the project a grant success.

We whole heartedly express our gratitude to our Internal Project Guide **Mrs. C. USHA RANI**, Assistant Professor, Department of Computer Science and Engineering for her valuable guidance and help for the successful completion of the project.

We also express our sincere thanks to our Project Coordinator **MR. C. RAMASAMY @ SANKARRAM** Assistant Professor**, Mr. P. KARTHIKEYAN** Assistant Professor**, Mr. C. SURESH KUMAR**, Teaching Fellow, Department of Computer Science and Engineering, for her constant inspiration and support.

We extend our thanks to all other Teaching and Non-Teaching Staff members for their constant encouragement and moral support.

And we would like to thank all our well-wishers who have encouraged us to do this project.

**ABSTRACT**

Largest-ever study of deaths shows heart diseases have emerged as the number one killer in world. About 25 per cent of deaths in the age group of 25- 69 years occur because of heart diseases. If all age groups are included, heart diseases account for about 19 per cent of all deaths. It is the leading cause of death among males as well as females. It is also the leading cause of death in all regions though the numbers vary. The proportion of deaths caused by heart disease is the highest in south India

(25 per cent) and lowest - 12 per cent - in the central region of India. The prediction of heart disease survivability has been a challenging research problem for many researchers. Since the early dates of the related research, much advancement has been recorded in several related fields. Therefore, the main objective of this manuscript is to report on a research project where we took advantage of those available technological advancements to develop prediction models for heart disease survivability. Data mining algorithms can provide great assistance in prediction of early stage heart disease. This dataset contained total 15 attributes in which we applied Naive Bayes, KNN, Random forest tree and Support Vector Machine (SVM) classification algorithms and calculated their prediction accuracy. An efficient feature selection algorithm helped us to improve the accuracy of each model by reducing some lower ranked attributes

**TABLE OF CONTENTS**

**CHAPTER NO TITLE PAGE NO**

**ABSTRACT** V

**LIST OF ABBREVIATIONS** IX

**LIST OF FIGURES** VIII

1. **INTRODUCTION** 
   1. Overview of Data Mining 1
   2. Data Mining Techniques 3
   3. Data Mining Processes 4
   4. Tasks of Data Mining 5

**2. LITERATURE SURVEY** 9

**3. SYSTEM ANALYSIS**

3.1 Existing System 13

3.2 Proposed System 14

3.3 Architecture of a proposed system 15

3.4 Analysis 16

3.5 Attributes of heart datasets 17

3.6 Algorithms in proposed system 20

**4. SYSTEM DESIGN**

4.1 System Architecture 25

4.2 Datasets Used 27

**5. SYSTEM REQUIREMENTS**

5.1 Hardware & Software details

**TABLE OF CONTENTS**

**CHAPTER NO TITLE PAGE NO**

5.2 About the Software 28

5.3 Python 30

**6. CONCLUSION AND FUTURE WORK** 31

**7. APPENDIX** 32

**8. REFERENCES** 41

**LIST OF FIGURES**

**FIGURE NO TITLE PAGE NO**

**1 SYSTEM ARCHITECTURE 24**

**2 IMPORTING DATA 34**

**3 READING THE CSV FILE 35**

**4 TRAINING DATASETS 36**

**5 TESTING DATASETS 37**

**6 TARGET DATA 38**

**7 ACCURACY DETECTION 38**

**8 DATA VISUALIZATION 39**

**9 SUPPORTING GRAPHS 39**

**LIST OF ABBREVIATIONS**

**DSS** Decision Support Systems

**HDPS** Heart Disease Prediction System

**KNN** K-Nearest Neighbors

**CART** Classification and Regression Trees

**CHAID** Chi Square Automatic Interaction Detection

**ADPS** Automated Disease Prediction System

**IHDPS** Intelligent Heart Disease Prediction System

**MAFIA** Maximal Frequent Itemset Algorithm

**ID3** Iterative Dichotomized 3

**EHR** Electronic Health Record

**PHR**  Patient Health Record

**BMI** Body Mass Index

**BP** Blood Pressure

**SVM**  Support Vector Machine