AN APPROACH FOR THE PREDICTION OF ACCURACY IN TYPE 2 DIABETES MELLITUS

A PROJECT REPORT

Submitted by

RAJAMANI.P (810015205062)

VETRI PRIYANGA.N (810015205096)

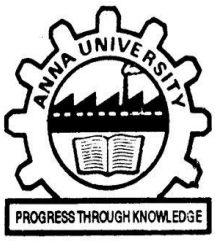
in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY



DEPARTMENT OF INFORMATION TECHNOLOGY

UNIVERSITY COLLEGE OF ENGINEERING -BIT CAMPUS

TIRUCHIRAPPALLI-620 024

ANNA UNIVERSITY: CHENNAI 600 025

APRIL 2019

UNIVERSITY COLLEGE OF ENGINEERING

BIT CAMPUS

TIRUCHIRAPPALLI-620 024

BONAFIDE CERTIFICATE

Certified that this project report titled **"AN APPROACH FOR THE PREDICTION OF ACCURACY IN TYPE 2 DIABETES MELLITUS"** is a bonafide work of MS.**P.RAJAMANI (810015205062)** and MS.**N.VETRI PRIYANGA (810015205096)** who carried out the project work under my supervision.Certified further that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr.D.VENKATESAN
HEAD OF THE DEPARTMENT
Department of Information Technology,
University College of Engineering,
Anna University-BIT Campus,
Tiruchirappalli-620 024.

Mr.M.KATHIRVEL
TEACHING FACULTY
Department of Information Technology,
University College of Engineering,
Anna University-BIT Campus,
Tiruchirappalli-620 024.

Submitted for the VIVA-VOCE to be held on _____

Internal Examiner **External Examiner**

DECLARATION

We hereby declare that the work entitled **"AN APPROACH FOR THE PREDICTION OF ACCURACY IN TYPE 2 DIABETES MELLITUS"** is submitted in partial fulfilment of the requirement for the award of the degree in B.Tech, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli, is record of our own work carried out by us during the academic year 2018– 2019 under the supervision and guidance of **Mr.M.KATHIRVEL,** Teaching Faculty, Department of Information Technology, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli. The extent and source of information are derived from the existing literature 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 degree, either in this or any other University.

P.RAJAMANI (81001205062)

N.VETRI PRIYANGA (810015205019)

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

Signature of the Guide,
Mr.M.KATHIRVEL
TEACHING FACULTY
Department of Information Technology,
University College of Engineering,
Anna University-BIT Campus,
Tiruchirappalli-620 024.

ACKNOWLEDGEMENT

It is a great opportunity to express our sincere thanks to all the people who have contributed to the successful completion of our project work through their support encouragement and guidance.

Our first and foremost thanks goes to **Dr.T.SENTHIL KUMAR,** Dean, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli for providing us the necessary facilities and supportive atmosphere for the completion of this project work.

It is our privilege to render our sincere thanks to **Dr.D.VENKATESAN,** Head of the Department of Information Technology, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli for providing us with excellent lab facilities and constantly encouraging us to pursue new goals and ideas.

We wish to record our heartfelt gratitude to our esteemed guide **Mr.M.KATHIRVEL,** Teaching Faculty, Department of Information Technology, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli for his excellent guidance, enterprising and valuable suggestions, encouragement and inspiration offered throughout the project.

It is our responsibility to thank our project co-ordinator **Mr.M.PRASANNA KUMAR,** Teaching Faculty, Department of Computer Science and Engineering, deserves a special vote of thanks for his constant inspiration that she has been all through the project period. We also thank the faculty members of the Department of Information Technology, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli for their remarkable help in completing the project.

We thank all our friends who have very understood, co-operative and appreciative and also understood with us as pillar of support during our good and bad times. On the whole, we express our heartfelt greatfulness to our parents without whom we cannot be shaped up this in our carrier. Last but never the least once again we thank one and all who have helped us either directly or indirectly in completing this dissertation.

ABSTRACT

In data mining technique, data pre-processing plays an important role in this model. Data mining technology is an appropriate study field, many algorithms are used to attain accuracy. In this model Pima Indian Diabetes Dataset is used. The Pima Indian Diabetes Dataset consists of information on patients. It is used to compare the result to attain higher accuracy. Anaconda Navigator provides a various editor such as Jupyter notebook is used. Jupyter notebook is a web applications used to visualize. The main problems is to solve are to improve the higher accuracy of prediction model for Type 2 Diabetes Mellitus. To further evaluate the performance of the model, Random forest and logistics regression algorithm techniques are used. Both the algorithms shows the good performance that the data set quality is sufficient. Naive bayes algorithm is used to train the dataset. The model is shown to be useful for realistic health management of diabetes.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	V
	LIST OF FIGURES	Viii
	LIST OF ABBREVIATIONS	ix
1	**INTRODUCTION**	1
	1.1 MACHINE LEARNING 1.1.1 MACHINE LEARNING TYPES 1.1.2 APPLICATIONS OF MACHINE LEARNING	1 1 2
	1.2 DIABETES DATA ANALYSIS 1.3OBJECTIVE	3 5
	1.4CLASSIFICATION 1.4.1 LOGISTIC REGRESSION 1.4.2 NAIVE BAYES CLASSIFIER 1.4.3 RANDOM FOREST ALGORITHM 1.5 REGRESSION 1.6 PREDICTION	5 5 6 6 7 7
2	**LITERATURE SURVEY**	8
3	**SYSTEM ANALYSIS**	11
	3.1 EXISITING SYSTEM	11
	3.1.1 DISADVANTAGES	12
	3.2 PROPOSED SYSTEM	12
	3.2.1ADVANTAGE	13
4	**SYSTEM DESIGN** 4.1 SYSTEM ARCHITECTURE	14 14

	4.2 MODULE DESCRIPTION	15
	4.2.1 DATA SET COLLECTION	15
	4.2.2 DATA PRE-PROCESSING	15
	4.2.3 DATA CLEANING	16
	4.2.4 STEPS FOR TEST AND TRAIN DATASET	17
	4.2.5 TEST DATASET	18
	4.2.6 TRAINING DATASET	19
	4.3 MODEL BUILDING	19
	4.3.1 TRAINING DATA	19
	4.4 APPLY ML ALGORITHMS	20
	4.4.1 NAÏVE BAYES CLASSIFIER	20
	4.4.2 LOGISTIC REGRESSION	20
	4.4.3 RANDOM FOREST CLASSIFIER	21
	4.4.4 EXPERIMENTAL RESULTS	21
5	**SYSTEM SPECIFICATION**	22
	5.1 SOFTWARE REQUIREMENTS	22
	5.2 HARDWARE REQUIREMENTS	22
	5.3 ABOUT THE SOFTWARE	22
6	**RESULT**	26
7	**CONCLUSION AND FUTURE WORKS**	28
APPENDIX 1 SOURCECODE		29
APPENDIX 2 SAMPLE SCREENSHOTS		33
REFERENCES		40

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
4.1	ARCHITECTURE OF THE PROPOSED SYSTEM	14
6.1	LOGISTICS REGRESSION	26
6.2	RANDOM FOREST	27

LIST OF ABBREVIATIONS

LOG	LOGISTIC REGRESSION
RF	RANDOM FOREST
ML	MACHINE LEARNING
T2DM	TYPE 2 DIABETES MELLITUS
GUI	GRAPHICAL USER INTERFACE