**DISEASE PREDICTION USING MEDICAL DATA**

SUBMITTED IN PARTIAL FULFILMENT REQUIREMENT FOR THE

AWARD OF DEGREE OF

**Bachelor of Technology**

# (Computer Science & Engineering)



Submitted By : Submitted To :

Bhuvesh Aggarwal (1606662) Prof. Priyanka Arora

Prof. Kapil Sharma

Training Coordinators

CSE Department

Department of Computer Science & Engineering

# Guru Nanak Dev Engineering College

Ludhiana 141006

**

**Candidate’s Declaration**

I hereby certify that I have undergone six month industrial training at ANSH InfoTech and worked on project entitled, “**Project Name- DISEASE PREDICTION USING MEDICAL DATA**” , in partial fulfillment of requirements for the award of Degree of **Bachelor of Technology** in Department of Computer Science & Engineering at **GURU NANAK DEV ENGINEERING COLLEGE** under **PUNJAB TECHNICAL UNIVERSITY, JALANDHAR**, having University Roll No. 1606662, is an authentic record of my own work carried out during a period from June, 2019 to December, 2019 under the supervision ofProf. Priyanka Arora

(**BHUVESH AGGARWAL**)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(**Prof. Priyanka Arora**)

Professor

**Abstract**

Disease Detection is a methodology by which we determine that whether a person is having Disease or not. We use various features for determining the health condition of patient. The dataset taken from Kaggle is categorized into different sections based on various features. There are many tools related to disease prediction. But particularly heart related diseases, tuberculosis, malaria have been analyzed and risk level is generated. But generally there are no such tools that are used for prediction of general diseases. So Disease Predictor helps for the prediction of the general diseases. In this application, machine learning algorithms designed using enhanced libraries of python language for analyzing the patient, has been discussed using which we can determine whether a person is suffering from Disease or not. The outputs generated are highly accurate and hence can be of great help to the User. The main focus is on to use machine learning in healthcare to supplement patient care for better results. Machine learning has made easier to identify different diseases and diagnosis correctly. Predictive analysis with the help of efficient multiple machine learning algorithms helps to predict the disease more correctly and help treat patients. The healthcare industry produces large amounts of healthcare data daily that can be used to extract information for predicting disease that can happen to a patient in future while using the treatment history and health data.

**ACKNOWLEDGEMENT**

 I am highly grateful to the **Dr. Sehijpal Singh**, Principal, Guru Nanak Dev Engineering College (GNDEC), Ludhiana, for providing this opportunity to carry out the six month industrial training at ANSH Infotech. The constant guidance and encouragement received from **Prof. G.S. Sodhi** Dean T&P, GNDEC Ludhiana has been of great help in carrying out the project work and is acknowledged with reverential thanks.

I would like to express a deep sense of gratitude and thanks profusely to **Mr. Ansh Aneja** Director/CEO of Company. Without the wise counsel and able guidance, it would have been impossible to complete the report in this manner. The help rendered by **Mr. Akashdeep singh**, Data Scientist of CSE department for experimentation is greatly acknowledged. I express gratitude to other faculty members of Information Technology department of GNDEC for their intellectual support throughout the course of this work. Finally, I/WE are indebted to all whosoever have contributed in this report work and friendly stay at Ansh Infotech.

**Bhuvesh Aggarwal**

**LIST OF FIGURES**

**Fig. No. Figure Description Page No.**

Figure 2.1 Waterfall Model 10

Figure 3.1 Data flow Diagram 15

Figure 3.2 Prediction 16

Figure 3.3 E-R Diagram 18

Figure 3.4 Testing Sample Data 19

Figure 3.5 Training Sample Data 20

Figure 5.1 GUI of System 32

Figure 5.2 GUI of System 32

Figure 5.3 GUI of System 33

Figure 5.4 GUI of System 33

Figure 5.5 Tested Data 34

Figure 5.6 Trained Data 35

**LIST OF TABLES**

**Table No. Table Description Page No.**

Table 1 Testing Sample Data 19

Table 2 Training Sample Data 20

**table of CONTENTS**

* 1. **Contents Page No.**
  2. ***Certificate i***
  3. ***Table of Contents ii***
  4. ***Abstract iv***
  5. ***Acknowledgement v***
  6. ***Supervisor’s Certificate vi***
  7. ***Student’s Declaration vii***
  8. ***List of Figures viii***
  9. ***List of Tables ix***

**CHAPTER 1**

INTRODUCTION 1

1.1 INTRODUCTION TO ORGANIZATION 1

1.2 INTRODUCTION TO PROJECT 2

1.3 PROJECT CATEGORY 2

1.4 PROJECT OBJECTIVES 2

1.5 PROBLEM FORMULATION 3

1.6 IDENTIFICATION/REORGANIZATION OF NEED 3

1.7 EXISTING SYSTEM 3

1.8 PROPOSED SYSTEM 3

1.9 UNIQUE FEATURES OF THE SYSTEM 4

CHAPTER 2 5

REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION 5

2.1 FEASIBILITY STUDY 5

2.1.1 TECHNICAL FEASIBILITY 5

2.1.2 ECONOMICAL FEASIBILITY 6

2.1.3 OPERATIONAL FEASIBILITY 6

2.2 SOFTWARE REQUIREMENT SPECIFICATION 7

2.2.1 INTRODUCTION 7

2.2.2 SOFTWARE REQUIREMENTS 7

2.2.3 HARDWARE REQUIREMENTS 7

2.2.4 FUNCTIONAL REQUIREMENTS 8

2.2.5 NON FUNCTIONAL REQUIREMENTS 8

2.3 VALIDATION 9

2.4 EXPECTED HURDLES 9

2.5 SDLC MODEL TO BE USED 9

CHAPTER 3 12

SYSTEM DESIGN 12

3.1 DESIGN APPROCH 12

3.2 DETAIL DESIGN 12

3.3 SYSTEM DESIGN USING VARIOUS STRUCTURED ANALYSIS AND DESIGN TOOL 14

3.3.1 DATA FLOW DIAGRAMS 15

3.4 USER INTERFACE DESIGN 16

3.5 Database Design 17

3.5.1 ER Diagrams 17

3.5.2 Database manipulation 18

3.6 METHODOLOGY 20

CHAPTER 4 22

IMPLEMENTATION, TESTING AND MAINTENANCE 22

4.1 INTRODUCTION TO LANGUAGES, IDE’S AND TECHNOLOGY USED FOR IMPLEMENTATION 22

4.1.1 PYTHON 22

4.2. MACHINE LEARNING 24

4.3 TESTING TECHNIQUES AND TEST PLANS 27

CHAPTER 5 29

RESULTS AND DISCUSSIONS 29

5.1 BRIEF DISCRIPTION OF VARIOUS MODULES OF THE SYSTEM 29

5.1.1 LIBRARIES USED 29

5.2 SNAPSHOTS OF SYSTEM 32

5.3 BACK ENDS REPRESENTATION 34

5.3.1 SNAPSHOTS OF DATABASE TABLE 35

CHAPTER 6 36

CONCLUSION AND FUTURE SCOPE 36

6.1 CONCLUSION 36

6.1.1 ACCURACY 36

6.2 FUTURE SCOPE 37

REFERENCES 38