**Application of Machine Learning based approach for Diabetes Prediction**

**A Mini Project Report**

***Submitted by***

**Amarjeet Kumar Singh 202002021073**

**Bhagyashree Nath 202002022099**

**Nerswn Baglary 202002022089**

***Under the supervision of***

**Dr Pankaj Pratap Singh**

Assistant Professor

**B. Tech. 6th Semester**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**



केन्द्रीय प्रौद्योगिकी संस्थान कोकराझार

**CENTRAL INSTITUTE OF TECHNOLOGY KOKRAJHAR**

**(Deemed to be University under MoE, Govt. of India)**

**May 2023**

**CERTIFICATE OF APPROVAL**

This is to certify that the mini project entitled “**Application of Machine Learning based approach for Diabetes Prediction**” submitted by **Amarjeet Kumar Singh (202002021073) , Bhagyashree Nath (202002022099) and Nerswn Baglary(202002022089)** of B.Tech. 6th Semester in the department of Computer Science & Engineering, Central Institute of Technology Kokrajhar is a record of bona fide work carried out by them under our supervision and guidance.

#### (Signature) (Signature)

#### 

#### Guide Head of the Department

#### Dr Pankaj Pratap Singh Dr. Amitava Nag

Assistant Professor Professor

Dept. of Computer Science & Engineering Dept. of Computer Science & Engineering

##### Certificate by the Board of Examiners

This is to certify that the project work entitled “Application of Machine Learning based approach for Diabetes Prediction” submitted by Amarjeet Kumar Singh (202002021073) , Bhagyashree Nath (202002022099) and Nerswn Baglary(202002022089) to the department of Computer Science & Engineering of Central Institute of Technology, Kokrajhar has been examined and evaluated.

This project work has been prepared as per the regulations of Central Institute of Technology and has been qualified for acceptance.

#### (Signature) (Signature)

#### Project Coordinator Board of Examiners

#### Tania Sarkar

Assistant Professor

Dept. of Computer Science & Engineering

## **ABSTRACT**

Diabetes is a chronic metabolic disorder affecting millions of people worldwide. Early detection and accurate prediction of diabetes can play a crucial role in effective management and prevention of complications. This project aims to develop a machine learning model for diabetes prediction that can accurately predict the likelihood of a person having diabetes based on various clinical features and give suggestion of relevant Insurance policy.

The project starts with data collection, where relevant data on patients with and without diabetes is gathered from healthcare databases, surveys, and medical records. Data cleaning techniques are then applied to clean and normalize the collected data, removing any inconsistencies or missing values.

Next, feature extraction methods are employed to identify the most significant predictors of diabetes from the available dataset.

Various machine learning algorithms are trained and evaluated to build the system.

## 

## i

## **ACKNOWLEDGEMENTS**

We would like to express our deepest gratitude to our guide, Dr. Pankaj Pratap Singh for his valuable guidance, consistent encouragement, personal care and timely help which provided us with an excellent atmosphere for doing our project. In spite of the busy schedule, he has extended his cheerful and cordial support to us, without which we could not have completed our project work.

We express our heartfelt thanks to our Head of the Department, Dr. Amitava Nag who has been actively involved and very influential from the start till the completion of our project.

We would also like to thank all teaching and non-teaching staffs of the Computer Science & Engineering Department for their constant support and encouragement.

Last but not the least it is our pleasure to acknowledge the support and wishes of our friends and well-wishers, both in academic and non-academic spheres.

ii

# CONTENTS

ABSTRACT

[i 5](#_Toc135519273)

[ACKNOWLEDGEMENTS 5](#_Toc135519274)

[CONTENTS 6](#_Toc135519275)

[Chapter 1 8](#_Toc135519276)

[Introduction 8](#_Toc135519277)

[1.1 Objectives 10](#_Toc135519278)

[1.2 Motivation 10](#_Toc135519279)

[Chapter 2 4](#_Toc135519280)

[Literature Survey 4](#_Toc135519281)

[Background details of different concepts used in your project along with related research work. Error! Bookmark not defined.](#_Toc135519282)

[2.1 Background Details Error! Bookmark not defined.](#_Toc135519283)

[Study of different concepts 4](#_Toc135519284)

[2.1.1….. 4](#_Toc135519285)

[2.1.1.1 …… 4](#_Toc135519286)

[2.2 Related Work 6](#_Toc135519287)

[Chapter 3 7](#_Toc135519288)

[Proposed System 7](#_Toc135519289)

[3.1 System Architecture 7](#_Toc135519290)

[3.2 Flow Chart/ Data Flow Diagram/E-R Diagram 7](#_Toc135519291)

[Give your relevant diagram 7](#_Toc135519292)

[3.3 Database Design 7](#_Toc135519293)

[Required in case your project contains a database 7](#_Toc135519294)

[3.4 Hardware Components 7](#_Toc135519295)

[3.5 Software tools 7](#_Toc135519296)

[Chapter 4 8](#_Toc135519297)

[4.1 Implementation Details 8](#_Toc135519298)

[4.2 Results 8](#_Toc135519299)

[4.3 Analysis 8](#_Toc135519300)

[Give necessary graphs 8](#_Toc135519301)

[4.4 Discussion 8](#_Toc135519302)

[6 8](#_Toc135519303)

[Chapter 5 9](#_Toc135519304)

[5.1 Conclusion 9](#_Toc135519305)

[5.2 Future Scope 9](#_Toc135519306)

[References 10](#_Toc135519307)

iii

# Chapter 1

# Introduction

# Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Diabetes can lead to many serious long-term complicated disease like cardiovascular disease, stroke, kidney failure, heart attack, peripheral arterial disease, blood vessels, and nerves. The number of people with diabetes rose from 108 million in 1980 to 422 million in 2014. Diabetes can be divided into three types as (i) Type I diabetes (T1D), (ii) type II diabetes (T2D), and (iii) Gestational diabetes (GD). T1D, previously known as insulin-dependent, juvenile or childhood-onset is characterized by deficient insulin production and requires daily administration of insulin. The symptoms of T1D are polyuria, thirst, constant hunger, weight loss, vision changes and fatigue. In 2017 there were 9 million people with type 1 diabetes; the majority of them live in high-income countries. T2D occurs when our body doesn't make or use insulin properly. People who are middle-aged or older are most likely to get this kind of diabetes. T2D also affects kids and teens, mainly because of childhood obesity.T2D is the most common type of diabetes. There are about 29 million people in the U.S. with type 2. Another 84 million have prediabetes, meaning their blood sugar (or blood glucose) is high but not high enough to be diabetes yet. The symptoms of T2D are being very thirsty, blurry vision, being cranky, fatigue, tingling or numbness in hands or feet, weight loss without trying. The third type of diabetes is gestational diabetes. Actually, pregnant women are affected by gestational diabetes.

1

Chapter 1. *Introduction* 2

### Objectives

The objectives of our project are listed below:

* + - The aim of the project is to determine the appropriate machine learning based model that will be used for prediction of diabetes to figure out if a person is diabetic or non-diabetic so far. Based on the predicted result this model will suggest the appropriate life insurance policy.
    - Objective2

…….…..

### 1.2 Motivation

The number of diabetic patients is increased day by day as a result deaths are also increased day by day. Diabetes can damage every part of our body. it can give serious damage to the heart, blood vessels, nerve problems, foot problems, stroke and many more. Type 2 diabetes affects men and women proportionately, around 12 million men and 11.5 million women have diabetes.

**1.3 Problem Statement**

The number of people with diabetes rose from 108 million in 1980 to 422 million in 2014. The main goal of this project is to develop prediction modelling of the given medical data of patients with and without diabetes. We aim to predict and analyze the best algorithm that is suitable for Diabetes prediction and also find the efficiency. We have collected the dataset from Kaggle website which is used for this study and then analysis was done with data mining techniques.

# Chapter 2

# Literature Survey

# .

### 2.1 Classification and prediction of diabetes disease using machine learning paradigm :

### The paper was published in 2022 by the author Md. Maniruzzaman et al(3) . This paper focuses on classifying and predicting of diabetes patients using maching learning classifiers. They have used various machine learning classifier like naïve bayse (NB) , decision Tree (DT), adaboost(AB) , random forest(RF) to predict the diabetic patient. They used Logistic regression (LR) to identify the risk factors for diabetes disease based on p value and odds ratio (OR)..

### 2.1.1 Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications :

This paper was published in 2021 by the author Umair Muneer Butt et  al (3). This paper focuses on for predicting analysis of diabetes, long short-term memory (LSTM), moving averages (MA), and linear regression (LR) is used for predicting the diabetes patient.

### 2.1.1.1 Diabetes Prediction using Machine Learning Algorithms:

### This paper was published in 2019 by the author Aishwarya Mujumdara et  al (1), This paper aims to discover solutions to detect the diabetes by investigating and examining the patterns originate in the data via classification analysis by using Decision Tree and Naïve Bayes algorithms. The research hopes to propose a faster and more efficient method of identifying the disease that will help in well-timed cure of the patients.

Font style: Times new Roman

**Heading font size**

**Chapter title: 18, bold**

**1st level heading: 16, bold**

**2nd level heading:14, bold**

**3rd level heading:12, bold**

Report should contain no more than three levels of headings

Text font size:12

**Follow for every chapter**

3

Chapter 2. *Literature Survey*

6

### 

### Related Work

Liu Dan and Cao Xin in [1] carried out experiments on intelligent agriculture greenhouse monitoring system based on ZigBee technology……………………………………………

…………………………

…………………………

Discuss existing works related to your project.

# Chapter 3

# Proposed System

Overall working principle of your project

### System Architecture

Detailed system architecture

### Flow Chart/ Data Flow Diagram/E-R Diagram

### Give your relevant diagram

### Database Design

### Required in case your project contains a database

### Hardware Components

Different hardware components used in your project

### Software tools

Different software tools required for implementation of your project

5

# Chapter 4

**Implementation and Results**

Details of how you have implemented your project along with different results obtained

### Implementation Details

Give implementation details

### 

### Results

Give the results obtained

### Analysis

### Give necessary graphs

### Discussion

Any further discussion about your results

# 

# 6

# Chapter 5

**Conclusion and Scope**

In this section, you have to conclude your project and give future directions

### Conclusion

Conclude your work

……………….…………

### Future Scope

Write how your project can be improved in future to be used in different application areas.

7

# References

[1] Liu Dan and Cao Xin, ” Intelligent agent greenhouse environment monitoring system based on IoT technology”, *International Conference on Intelligent Transportation, Big Data & SmartCity,2015.*

Include other references……………….

8