Diabetes Prediction Using Different Machine Learning Approaches

PRESENTED BY

R Ramya   1KS18CS077

Vedavedya B H   1KS18CS112

Veera Sreenidhi R   1KS18CS113

Venkatesh M N   1KS18CS114

UNDER THE GUIDANCE OF
Dr . Dayananda R B
Dept. Of CSE,KSIT

Overview

▶ Diabetes, also known as chronic illness, is a group of metabolic diseases due to a high level of sugar in the blood over a long period. In 2019, diabetes was the 9th leading cause of death with an estimated 1.5 million deaths directly caused by diabetes.

▶ Insulin is a hormone that regulates blood sugar. The insulin hormone in the body produced by the pancreas allows glucose to pass from the food into the bloodstream. The lack of that hormone due to malfunctioning of the pancreas forms diabetes.

▶ However, there is no long term cure for diabetes, but it can be controlled and prevented if an early prediction is accurately possible.

▶ The robust and accurate prediction of diabetes is highly challenging due to the limited number of labeled data and also the presence of outliers (or missing values) in the diabetes datasets.

➢ This project focuses on evaluating a dataset collected from 768 female diabetic patients from the Pima Indian population near Phoenix, Arizona [6].

➢ Pre-processing, in the proposed pipeline, is the heart of achieving the state-of-the-art result, which consists of outlier rejection, filling missing values, data standardization, feature selection, and K-fold cross-validation.

➢ We consider the mean value in the missing position of attribute rather than median value, as it has a more central tendency toward the mean of that attribute distribution.

Goal

Serious stage of diabetes is when, severely elevated blood sugar levels due to an actual lack of insulin or insufficient action of insulin. This leads to conditions called **diabetic ketoacidosis** or hyperosmolar coma, which can be life-threatening.

➢ The goal of the project is to prevent or delay the appearance of late disease complications

➢ The objective of the project is to employ machine learning algorithms to develop a prediction model for diabetes based on clinical data and deploy it using stream-lit library.

➢ Another objective of this project is to provide an effective model to predict the diabetes by least number of predictors.

DATASET

➢ The Diabetes dataset was collected from 768 female diabetic patients from the Pima Indian population near Phoenix, Arizona [6].

➢ This dataset consists of 268 diabetic patients (positive) and 500 non-diabetic patients (negative) with eight different attributes.

➢ Features include Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age.

➢ The Pedigree(Diabetes Pedigree Function) was calculated : $\text{Pedigree} = \dfrac{\sum_i K_i(88 - ADM_i) + 20}{\sum_j K_j(ALC_j - 14) + 50}$

➢ where i and j respectively denote the relatives who had developed and NOT developed diabetes. K is the percentage of shared genes by the relatives (K = 0.500 for the parent or full sibling, K = 0.250 for a half-sibling, grandparent, aunt or uncle and K = 0.125 for a half aunt, half-uncle or first cousin). ADMi and ACLj is the age of relatives, in years, at the time of diagnosing and at the last non-diabetic test respectively.

APPLICATION

➢ This project helps patients to tackle their condition and receive treatment in the early stages of the disease.

➢ The disease can be detected with the least possible tests and at low cost.

➢ To predict patients with diabetes using less number attributes while maintaining a higher accuracy.

➢ It will help the doctors to start the treatments early for the diabetes patients and also it will help to diagnose more patients within a less time period.

REQUIREMENT SPECIFICATION

Minimum Hardware Requirements:

Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher. An AMD equivalent processor will also be optimal.
RAM — 8 GB minimum.
ROM — 100 GB minimum
Operating System — Windows 10/ Linux .

Minimum Software Requirements:

Jupiter notebook / Anaconda
Python
Libraries: numpy, pandas, matplotlib, seaborn, streamlit, sklearn, imblearn, pickle, pyngrok, Collections

REFERENCE

➢ **Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers**

MD. KAMRUL HASAN 1 , MD. ASHRAFUL ALAM1 , DOLA DAS2 , EKLAS HOSSAIN 3 , (Senior Member, IEEE), AND MAHMUDUL HASAN 2 1Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Khulna 9203, Bangladesh 2Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna 9203, Bangladesh 3Oregon Renewable Energy Centre (OREC), Department of Electrical Engineering and Renewable Energy, Oregon Institute of Technology, Klamath Falls, OR 97601, USA

➢ **Predicting Diabetes in Healthy Population through Machine Learning**

Lejla Alic∗, Hasan T. Abbas† ‡, Marelyn Rios§, Muhammad AbdulGhani¶, and Khalid Qaraqe† ∗Magnetic Detection & Imaging Group, Faculty of Science & Technology, University of Twente, Enschede, The Netherlands †Dept. of Electrical & Computer Engineering, Texas A&M University at Qatar, Doha, Qatar 23874 §Dept. of Industrial & Systems Engineering, Texas A&M University, College Station, TX 77843-3128 ‡ Electronic & Nanoscale Engineering University of Glasgow, Glasgow, Scotland G12 8QQ ¶Division of Diabetes, University of Texas Health Science Center at San Antonio, Texas