Chapter 3
Modeling Requirements With Diabetes Using Supervised Machine Learning Techniques

Priyal Dangi
CHRIST University (Deemed), India

ABSTRACT

Diabetes is characterized by either insufficient or inefficient insulin production by the body. High blood glucose levels result from this, which over time can harm a number of tissues and organs in the body. Diabetes can be brought on by a specific age, obesity, inactivity, insufficient physical activity, inherited diabetes, lifestyle, poor diet, hypertension, etc. This chapter explores modeling requirements with diabetes using supervised machine learning techniques.

INTRODUCTION

Diabetes is a widespread medical disorder that significantly actually harms wellbeing. Diabetes is characterized by levels of glucose in the blood which seem to be extraordinarily high and therefore are driven by inadequate secretion, inadequate physiologic consequences of glucose, maybe all in case. If diabetes is not treated, there are severe complications then. In the healthcare area, So, instead employing a single classifier, a classification or categorization techniques are usually used to segregate information into numerous groups in line with predefined criteria. Stated below are the different diabetes diagnoses:

Type 1 diabetes: An auto immune disease in which the immune system of the human body targets and damages insulin-producing pancreatic cells, type 1 diabetes is the most common form of the disease. Type 1 diabetics must inject themselves with insulin or use an insulin injection to stabilize their sugar levels in the blood.

DOI: 10.4018/978-1-6684-9809-5.ch003

Type 2 diabetes: The most typical diabetic condition, type 2, is brought on when the body develops a glucose intolerance or is unable to produce enough insulin to control the amount of sugar in the blood. Hypertension, obesity, a inactive lifestyle, and bad eating habits are frequently linked to type 2 diabetes. (Jain et al., 2023).

In order to gather information and forecast, prediction model utilises multiple approaches to machine learning, methods of statistics and data mining are two closely related disciplines. Significant conclusions and predictions can be made by using predictive analysis on healthcare data. Here approaches for learning algorithms such as SVM, Logistic regression, Naïve bayes, Decision tree, k-nearest neighbours (Knn), Random Forest, can be used for data modelling and prediction model. prediction model strives to improve healthcare outcomes by increasing patient care, optimizing resource use, and making the best possible disease diagnosis. Where Knn is giving the greater accuracy of i.e. (81%) among all the other models we have opted (Dangi et al., 2023).

Related Work

Nongyao et al. implemented a classification method for diabetic patients' risk. The author used the four well-known machine learning categorization methods Naive Bayes, Decision Tree (DT), Artificial Neural Networks (ANN), Naive Bayes, and Logistic Regression for the achievement of the specified objective. The proposed model's power is enhanced by using boosting and bagging strategies. Therefore, Random Forest method, out of all the algorithms used, produces the best outcomes, according to experiment results (Watkins et al., 2002).

Orabi et al. employed the system for predicting the diabetes disorder or disease, with the highly defined objective of predicting the presence of diabetes in an individual at a set age. By using decision trees, the proposed system can be developed using the principle of machine learning. Results were favourable since the system's architecture is good at forecasting diabetes prevalence rate at a specific age (Razavian et al., 2015).

A strategy for the analysis and the assessment of diabetic and insulin statistics was implemented using data and the Hadoop Map reduce technology and programming. The lifestyle factors and form of diabetes are anticipated by this strategy. The technology has been shown affordable for any healthcare business (Lipton et al., 2014).

Georga et al. centred on the glucose and, as a multivariate regression problem, used support vector regression (SVR) to forecast diabetes. Also, to increase the accuracy, more and more studies have been using ensemble approaches (Park et al., 2017).

K. Rajesh and V. Sangeetha (2012) a classification and categorization strategy was utilized. They have been using the C4.5 decision tree model to derive and detect the specific underlying patterns from the collection in order to perform effective and appropriate categorization (Svetnik et al., 2015).

Sajida et al. shows how Boost and Bagging cluster strategies are utilized for machine learning techinques through using the decision tree as the core effect in the detection of the people with diabetes and those without it based on factors that increase the likelihood of the condition (Li et al., 2017).

The Simple K-means algorithm called is included in the suggested combination prediction model, which then applies a classification algorithm to the outcomes of the classification method. Using the C4.5 decision tree approach, classifiers are created (Yue et al., 2008).