Diabetes Prediction – Preliminary Report

INTRODUCTION

Diabetes is a health condition that affects how the human body turns food into energy. A large amount of the food consumed is broken down into sugar (also known as glucose) and released into the bloodstream. In response to high blood sugar, the pancreas releases insulin. When diabetes is not managed properly, blood sugar levels can build up, increasing the risk of dangerous complications, including stroke and heart disease.

DATA DESCRIPTION

This project will use the Pima Indians Diabetes Database. This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether a patient has diabetes, based on certain diagnostic measurements included in the dataset. From a larger database, these instances were selected according to several constraints. All patients here are at least 21-year-old female Pima Indians.

PROPOSED ANALYSIS

We follow the quantitative methods. Our main emphasis is to predict whether a person has diabetes or not before they are diagnosed with it using Random Forest and AdaBoost algorithms. Random Forest is a powerful machine-learning algorithm that combines multiple decision trees together. And using Ada Boosting we can strengthen weak classifiers into one strong classifier. Using these major algorithms, we develop the predictive model.

ANALYSIS METHODS

A Diabetes prediction and recommendation system will be designed and implemented using machine learning classification techniques based on an analysis of the Diabetes dataset. A prediction system will be developed using Support Vector Machines and Random Forest, Gradient Boosting, and AdaBoost algorithms as well.

MILESTONES

To build a predictive model to predict the number of people diagnosed with diabetes, data has been cleansed and machine learning algorithms have been applied. The model is in its final stages of completion.

REFERENCES

Geek Culture, Level up coding, Kaggle