Diabetes

Abstract:

Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high. Blood glucose is your main source of energy and comes from the food you eat. Insulin, a hormone made by the pancreas, helps glucose from food get into your cells to be used for energy.

The chronic metabolic disorder diabetes mellitus is a fast-growing global problem with huge social, health, and economic consequences. It is estimated that in 2010 there were globally 285 million people (approximately 6.4% of the adult population) suffering from this disease. This number is estimated to increase to 430 million in the absence of better control or cure. An ageing population and obesity are two main reasons for the increase. Furthermore, it has been shown that almost 50% of the putative diabetics are not diagnosed until 10 years after onset of the disease, hence the real prevalence of global diabetes must be astronomically high.

The goal of this project the goal of this project is to classify the patient as having diabetes or not, and to show most of the factors that have causes associated with diabetes.

Also, make the people knowledge about diabetes and how it is Serious illness.

Design:

Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high. Blood glucose is your main source of energy and comes from the food you eat. Insulin, a hormone made by the pancreas, helps glucose from food get into your cells to be used for energy.

Data:

This dataset was on Kaggle and were exploring by jupyter notebook to predict and select the best model to predict if they have a diabetes or not.

Data size: (253K Rows x 22 Columns).

Rows: 253K.

Columns: 22.

which are:

Diabetes binary–High BP–High Chol–Chol Check–BMI–Smoker– Stroke–Heart Disease or Attack Phys Activity– Fruits– Veggies– Hvy Alcohol Consump– Any Healthcare– No Docbc Cost– GenHlth MentHlth– PhysHlth– DiffWalk– Sex–Age–Edcation–Income.

Algorithms:

The dataset was clean.

Calculate the factor how have the height correlation by heatmap.

Calculate the recall for the models and see what is the good one.

Calculate the accuracy and see which models is have high an accuracy.

Calculate the Mean Squared Error.

Calculate the Root Mean Squared Error.

Tools:

The technologist that I used are Python, Jupyter Notebook.

The libraries are Pandas, NumPy, Matplotlib, seaborn, Scikit Learn, scipy, ensemble RandomForestClassifier, DecisionTreeClassifier, LogisticRegression, SVM, KNN, math, metrics,

Communication:

there is no specific reason can lead to diabetes but there are some risk factors that can make the probability of been diabetic are high.

BMI has the highest impact on Diabetes then the Age.

We use some models to find the high accuracy and the SVM turned out to be the best model.