Predicting Diabetes Using Machine Learning Models

Problem statement:

Can we predict diabetes knowing patient's clinic and demographic information?

Diabetes background:

Diabetes is a chronic (long-lasting) health condition that affects how your body turns food into energy. Most of the food you eat is broken down into sugar (also called glucose) and released into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin. Insulin acts like a key to let the blood sugar into your body's cells for use as energy.

If you have diabetes, your body either doesn't make enough insulin or can't use the insulin it makes as well as it should. When there isn't enough insulin or cells stop responding to insulin, too much blood sugar stays in your bloodstream. Over time, that can cause serious health problems, such as heart disease, vision loss, and kidney disease.

There isn't a cure yet for diabetes, but losing weight, eating healthy food, and being active can really help. Taking medicine as needed, getting diabetes self-management education and support, and keeping health care appointments can also reduce the impact of diabetes on your life. Therefore, there is still a great need to have access to machine learning models that can improve diabetes diagnosis.

Criteria for Success:

Build a machine learning model that predict diabetes with high accuracy.

Data Source:

The Pima Indian diabetes dataset is originally from the National Institute of

Diabetes and Digestive and Kidney Diseases and was obtained from the Kaggle website.

Data Description:

The Pima Indians diabetes dataset contains records for 768 subjects along with nine columns of predictive and target variables. The eight predictive variables are pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, and age. The outcome variable is a class variable (0 if non-diabetic, 1 if diabetic).

Scope of solution space:

Since diabetes is a worldwide disease affecting millions of people, it is essential that each affected country effectively allocate resources to control this deadly disease. Millions of people are not even aware that they have diabetes. Having access to robust machine learning models that predict diabetes with high accuracy will help countries' regions to receive the help they need to control and reduce diabetes.

Constraints within solution space:

This study will benefit by replicating the analysis using more recent diabetes data. Then, as future research we can include more predictive variables.