Diabetes Detection

Prepared by:

Hala Aloqiel

**Introduction:**

Diabetes is a chronic health condition which affects how your body turns food into energy. The food that we consume is broken down into glucose(sugar) and is released into the bloodstream. Normally, when the sugar in the bloodstream increases, the pancreas releases insulin which acts as a key to let the blood in our body’s be uses as energy. However, a diabetic person either doesn’t make enough insulin or can’t use the insulin it makes as well as it should. Diabetes can cause health problems such as heart disease, vision loss, and kidney disease. Although there is no cure for diabetes, early diagnosis and a healthy lifestyle can mitigate the symptoms associated with this condition, which makes predictive models for diabetes risk important tools for public and public health officials.

**Scope:**

This project demonstrates the detection of diabetes using machine learning classification algorithms. Using the data found at Kaggle, which contains 22 columns and over 250 thousand rows. The features that I’m using, includes the Diabetes indicator, which has 3 classes. 0 is for no diabetes or only during pregnancy, 1 is for prediabetes, and 2 is for diabetes. BMI, HighBP, HighChol, age, stroke, HeartDiseaseorAttack, PhysActivity, GenHlth, PhysHlth, DiffWalk. The data is collected as part of a survey in 2015 by CDC . Th original survey contains responses from over 441,455 individuals and has 330 features. These features are either questions directly asked of participants, or calculated variables based on individual participant responses. This project applies logistic regression, decision Tree classification, gradient boost classification, and random forest classification model to predict whether a person is at a risk of getting diabetes.

**Tools:**

* Python
* Pandas
* NumPy
* scikit\_learn
* seaborn

**Sources:**

* [**https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html**](https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html)
* [**https://www.kaggle.com/alexteboul/diabetes-health-indicators-dataset?select=diabetes\_012\_health\_indicators\_BRFSS2015.csv**](https://www.kaggle.com/alexteboul/diabetes-health-indicators-dataset?select=diabetes_012_health_indicators_BRFSS2015.csv)