DIABETES PREDICTION

PROPOSED BY:

Rawan Alharbi
Sultanah Aldossari

PROBLEM STATEMENT

Diabetes is a common, chronic disease. Prediction of diabetes at an early stage can lead to improved treatment.

DATASET

The original dataset comes from the National Institute of Diabetes and Digestive and Kidney Diseases. The Pima Indians Diabetes Database provides this data, which can be found on Kaggle. There is only one target variable (Outcome). A patient's age, BMI, insulin level, number of pregnancies and diabetes are all predictor variables.

Pregnancies	Number of times pregnant.
Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test.
BLOODPRESSURE	Diastolic blood pressure (mm Hg)
SkinThickness	Triceps skinfold thickness (mm).
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m)^2).
DiabetesPedigreeFunction	Diabetes pedigree function
Age	Age (years)
Outcome	Class variable (0 or 1) 268 of 768 are 1, the others are 0

QUESTION

- What features are strongly related to predicting diabetes?
- Can these features predict diabetes?

TOOLS

- Python
- Jupyter Notebook

LIBRARIES

- Pandas
- Numpy
- Matplotlib
- Seaborn