



أكاديمية سدايا
SDAIA Academy

Data Science Bootcamp

Project Proposal 3

**Apply Different classification
Algorithms to predict the
Diabetes using Python**

Presented by

Samah Alrefaei

30 Oct. 21



Introduction

In this project, I plan to predict is to diagnostically whether or not a patient has diabetes, from dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases., based on certain diagnostic measurements included in the dataset.

The datasets consist of several medical predictor (independent) variables and one target (dependent) variable, Outcome. Independent variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on

Objective

I will try to build a machine learning classification model to accurately predict whether or not the patients in the dataset have diabetes or not?

Data Description

The dataset we will use contains the Number of times pregnant, Plasma glucose, Diastolic blood pressure (mm Hg),Triceps skin fold thickness (mm),Insulin (mu U/ml),BMI: Body mass index (weight in kg/(height in m)²), Diabetes pedigree function , Age ,Outcome: Class variable (0 or 1)
The data are taken from the kaggle website. The following table explains the

dataset in detail:

Variables	Data type
Pregnancies	Integer
Glucose	Integer
BloodPressure	Integer
SkinThickness	Integer
Insulin	Integer
BMI	Float
DiabetesPedigreeFunction	Float
AgeOutcome	Integer
Outcome	Integer

There are 768 observations with 8 medical predictor features (input) and 1 target variable (output 0 for "no" or 1 for "yes")

Tools

I will analyze and explore the data in python by using jupyter ,and I will use more package such as : SXLearn for modeling , pandas , Matplotlib, numpy,seaborn

Conclusion

I expect the classification model will predict the property (outcome) , this research helps people with diabetes to the early disease prediction.

