Abstract

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. this project use to predict whether a patient is likely to get stroke or not based on the input parameters.

Data/ design

The data from kaggle.com(https://www.kaggle.com/fedesoriano/stroke-prediction-dataset) This dataset contains 5110 observations with 12 attributes and used to predict whether a people is likely to get stroke based on the input parameters :

1. id : unique identifier
2. gender : "Male", "Female" or "Other"
3. age : age of the patient
4. hypertension : 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
5. heart_disease : 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
6. ever_married : "No" or "Yes"
7. work_type : "children", "Govt_jov", "Never_worked", "Private" or "Self-employed"
8. Residence_type : "Rural" or "Urban"
9. avg_glucose_level : average glucose level in blood
10. bmi : body mass index
11. smoking_status : "formerly smoked", "never smoked", "smokes" or "Unknown"
12. stroke : 1 if the patient had a stroke, 0 the patient do not have a strok

In this project the I use Analysis ,Cleaning , Visualization model in the data

Algorithms

Logistic Regression

Tools

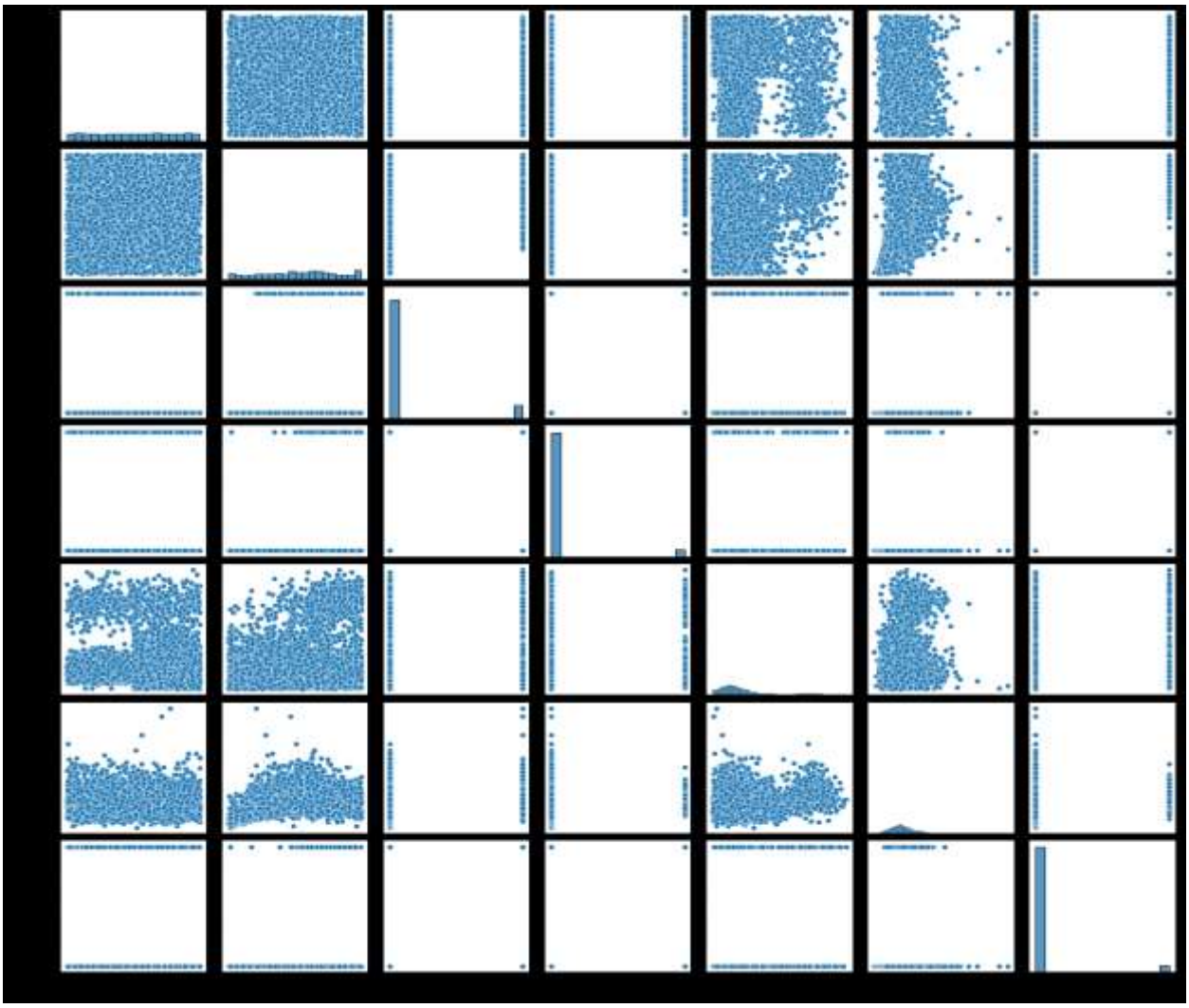
Pandas

Nubay

shuffle

seaborn and matplotlib for visualization data

Commniction



Percentage of pepole has a stroke: % 4.87 -->(249)

Percentage of pepole does not have a stroke: % 95.13 (4861)

Almost %95 of the instances of our target variable is 'No stroke'

4861 patient does not have a stroke

%5 of the instances of our target variable is 'Stroke'

249 patient have a stroke

Model accuracy

Logistic Regression with 95.46165884194053% accuracy