أكاديمية سدايا
SDAIA Academy

SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Predicting a stroke

Abstract

The goal of this project was to use a classification model to predict the factors and etiologies of stroke. In my project, we will analyze the most relevant risk factors for stroke as well as predict whether a patient is likely to have a stroke based on entry criteria such as gender, age, various diseases and smoking status. According to the World Health Organization (WHO), stroke is the second leading cause of death globally, and is responsible for approximately 11% of all deaths.

Design:

This project originates from Kaggle. Data provided by Fedsuriano. Patient data were collected from patients' age, gender, health status, smoking status, certain diseases, who had a stroke and based on patient data we see the likelihood of having a stroke, it can help make decisions about lifestyle changes to reduce complications and save high-risk patients.

Data:

Stroke data for the healthcare dataset from fedesoriano, 5110 Observations of 12 features.

Algorithms:

Resampling imbalance dataset

I take different machine learning algorithm as Logistic Regression, Random Forest Classifier.

Tools:

- Pandas: a library offers data structures and operations for manipulating numerical tables and time series.

- Numpy: a library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

- Matplotlib: a plotting library for the Python programming language and its numerical mathematics extension NumPy.

- Seaborn: a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.

- sklearn