CARDIOVASCULAR STROKE PREDICTION

Group Members

Phani Abhishek

Poojitha Chukkaluru

Rachan Reddy

BAN5573-01-F23

VISUAL ANALYTICS & BUS INTELL

Professor: Abdullah Asilkalkan

ABSTRACT

Heart disease is one of the major diseases in the world, and proper heart function is very important for physical health. There are various types of heart disease, including congenital heart disease, myocardial ischemia, cardiac arrest, myocardial infarction, coronary heart disease, peripheral heart disease. HCD. Also, men experience more HCD conditions than women, and heart attacks occur earlier in men than in women. According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.
This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.

Introduction:

Attribute information

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 or 0 if not
*Note: "Unknown" in smoking_status means that the information is unavailable for this patient

Problem statement

To analyse and compare the results between the various attributes of a person's who are more likely or prone to Heart stroke.

Scope of the project

This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.

Dataset – Healthcare-dataset-stroke-data.csv.

Data set is collected from a consequential set of patient's medical records. The term "heart disease" refers to multiple types of heart conditions which are hazardous to human heart health. The term" cardiovascular disease" is family of disease that neces- sitate the involvement of heart or blood vessel. Cardiac data is collected from Kaggle. In this Project we are using this database as it is mostly used by researchers for cardiac related research. The data set accommodates 5111 records, 12 attributes.