**Capstone Project 1: Stroke Prediction Data**

**Source of the Data: - Kaggle.com**

**Domain: - Health Sector Domain**

**Nature Of Data: -In Categorical form**

**Machine Learning Algorithm: - Classification Algorithm will be use**

**About Dataset: -**

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths due to their unhealthy lifestyle. They want to use some of the most advanced machine learning techniques to study about the Main cause of this disease

There are 5110 Tuple along with 12 features.

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Type** | **Description** |
| **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 job", "Never worked", "Private" or "Self-employed" |
| **8** | Residence Type: | "Rural" or "Urban" |
| **9** | Avg glucose level | Average glucose level in blood of every Persons |
| **10** | BMI | Body Mass Index of each person |
| **11** | Smoking status | The person is "formerly smoked", "never smoked", "smokes" or "Unknown"\*  \*Note: "Unknown" in smoking status means that the information is unavailable for this patient |
| **12** | stroke: | 1 if the patient had a stroke or 0 if not |

**Problem Statement: -**

In this dataset my target variable will be Stroke and it’s in categorical form, so A patient will have stroke disease in ‘’future or not’’ based on the input parameters like gender, age, various diseases, and smoking status etc. which means I will be predicting the accuracy of stroke disease by using proper means of Exploratory Data Analysis and Classification Machine Learning Algorithms.