**Predicting Probability of Stroke**

**Abstract** :

A stroke is a sudden interruption in the blood supply of the brain. Most strokes are caused by an abrupt blockage of arteries leading to the brain **(**[**ischemic stroke**](http://www.strokecenter.org/patients/about-stroke/ischemic-stroke/)**).** Other strokes are caused by bleeding into brain tissue when a blood vessel bursts **(**[**hemorrhagic stroke**](http://www.strokecenter.org/patients/about-stroke/intracerebral-hemorrhage/)**).** Because stroke occurs rapidly and requires immediate treatment, stroke is also called a brain attack. According to the World Health Organization, **ischemic** heart disease and stroke are the world’s biggest killers.

In this **project** we will try to predict the probability of an observation belonging to a category (in our case probability of having a stroke i.e. **1** and not having a stroke i.e. **0**).

**Goal of The Project** :

Predict whether a person is prone to have a stroke or not by considering few individual health factors, so that the person can take proper measures to avoid future probability of having a stroke.

**Data Set Information** :

We have used a historic dataset of various factors in patients, including their demographic, lifestyle and medical conditions along with their earlier symptoms :-

|  |
| --- |
| 1. Age |
| 1. Gender |
| 1. Hypertension |
| 1. Heart Disease |
| 1. Marital Status |
| 1. Work Type |
| 1. Residence |
| 1. Average Glucose Level |
| 1. BMI |
| 1. Smoking Status |

Depending upon the above factors we will try to predict whether a person is prone to have a stroke or not.

**Process Overview** :

In the dataset for simplicity we have marked the stroke -patients as ‘1’ and non-stroke patients as ‘0’,this will be our *target variable* or *dependent variable* because this is the one we want to predict .Now for each case we have corresponding individual demographic, lifestyle and health factors, we will call these factors as *independent variable* or *features* . Using machine learning techniques we have tried to find out the relations between these independent and dependent variables and we have discovered that our target variable is influenced by independent variables so there are underlying relationships .

**Technique Used** :

Our aim is to convert this problem into a Data Science problem and there by implementing a CRISP-DM method (Cross Industrial Standard Process). We will sequentially understand the Business problem, Exploring the data, Cleaning and Preprocessing the data, Building Machine Learning Models to classify the patients and finally selecting the better model with higher performance efficiency for deployment.

**Outcome** :

ML-Model with better performance in prediction is chosen as the final model that will be deployed.

**Conclusion** :

The model performance indicates that we have successfully achieved our project goal. All the technical details have been described in the following Project PPT.