McKinsey Health Care

Problem Statement

Your Client, a chain of hospitals aiming to create the next generation of healthcare for its patients, has retained McKinsey to help achieve its vision. The company brings the best doctors and enables them to provide proactive health care for its patients. One such investment is a Centre of Data Science Excellence.

In this case, your client wants to have study around one of the critical disease "Stroke". Stroke is a disease that affects the arteries leading to and within the brain. A stroke occurs when a blood vessel that carries oxygen and nutrients to the brain is either blocked by a clot or bursts (or ruptures). When that happens, part of the brain cannot get the blood (and oxygen) it needs, so it and brain cells die.

Over the last few years, the Client has captured several health, demographic and lifestyle details about its patients. This includes details such as age and gender, along with several health parameters (e.g. hypertension, body mass index) and lifestyle related variables (e.g. smoking status, occupation type).

The Client wants you to predict the probability of stroke happening to their patients. This will help doctors take proactive health measures for these patients. Also find the predicted values.

Hint:- You can use Spark MLLib/Scikit-learn that helps you to do feature engineering and improve accuracy by using different models.

Note: Sample submission file is given in the folder.

**DATA DICTIONARY**

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| id | Patient ID |
| gender | Gender of Patient |
| age | Age of Patient |
| hypertension | 0 - no hypertension, 1 - suffering from hypertension |
| heart\_disease | 0 - no heart disease, 1 - suffering from heart disease |
| ever\_married | Yes/No |
| work\_type | Type of occupation |
| Residence\_type | Area type of residence (Urban/ Rural) |
| avg\_glucose\_level | Average Glucose level (measured after meal) |
| bmi | Body mass index |
| smoking\_status | patient's smoking status |
| stroke | 0 - no stroke, 1 - suffered stroke |