**Overview**

**Stroke Prediction**

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.

In this hackathon, we challenge the data science enthusiasts to predict whether a patient is likely to get a stroke based on the parameters like gender, age, various diseases, and smoking status.

**Data Dictionary**

The dataset contains relevant information about the patient which can be useful for stroke prediction.

**Train File**

CSV containing the patient parameters for whom **‘stroke’** is known.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| id | Unique id |
| gender | Gender |
| age | Age |
| hypertension | Hypertension binary feature |
| heart\_disease | Heart disease binary feature |
| ever\_married | Has the patient ever been married? |
| work\_type | Work type of the patient |
| residence\_type | Residence type of the patient |
| average\_glucose\_level | Average glucose level in blood |
| bmi | Body Mass Index |
| smoking\_status | Smoking status of the patient |
| stroke | Stroke event |

**Test File**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| id | Unique id |
| gender | Gender |
| age | Age |
| hypertension | Hypertension binary feature |
| heart\_disease | Heart disease binary feature |
| ever\_married | Has the patient ever been married? |
| work\_type | Work type of the patient |
| residence\_type | Residence type of the patient |
| average\_glucose\_level | Average glucose level in blood |
| bmi | Body Mass Index |
| smoking\_status | Smoking status of the patient |

**Submission File Format**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| id | Unique id |
| stroke | Predicted stroke event |

**Public and Private LeaderBoard**

Test file is further divided into Public (25%) and Private (75%).

● Your initial responses will be checked and scored on the Public data.

● The final rankings would be based on your private score which will be published once the competition is over.

**Evaluation Criteria**

**Rubrics**

|  |  |
| --- | --- |
| **Component** | **Weightage** |
| Data Cleaning and Data Visualization | 25% |
| Model Building and Evaluation | 60% |
| Pipeline and Deployment  (Dashboard/Webapp) | 15% |