|  |  |
| --- | --- |
| **Project Proposal** |  |

*LAMYA ALARWAN*

**Design**

|  |  |
| --- | --- |
| **Project Title** | Stroke Prediction in Patients. |
| **Project Goal** | Stroke is a blood clot or bleeding in the brain which can cause long-term neurological damage and complications. stroke doesn't only cost medical financial burden and permanent disability but can eventually lead to death.  This project predicts whether a person is at risk of having a stroke or not. |
| **Project Benefit** | Prevention and early detection of a stroke risk has many benefits on the patient and healthcare system. |
| **Who benefits from the project?** | Patients and healthcare workers. |

**Data**

|  |  |
| --- | --- |
| **Data Source** | The dataset source is Kaggle ([Stroke Prediction Dataset](https://www.kaggle.com/fedesoriano/stroke-prediction-dataset)).  It is containing 5110 records and12 features.  Record: Contains patient information and if patient had a stroke or not.  Target features: [stroke]=1 if the patient had a stroke or 0 if not. |
| **Describe Features** | 1- id: Unique id.  2- gender: "Male "or "Female".  3- age: age of the patient.  4- hypertension: binary feature (0 if the patient doesn't have hypertension,1 if the patient has hypertension).  5- heart\_disease: binary feature (0 if the patient doesn't have any heart diseases,1 if the patient has heart diseases.  6- ever\_married:"Yes" or "No".  7- work\_type:" children"," Private"," Self-employed'," Govt\_job" or " Never\_worked".  8- Residence\_type:" Rural" or " Urban".  9- avg\_glucose\_level: Average glucose level in blood.  10- bmi: Body Mass Index.  11-smoking\_status:" never smoked"," formerly smoked"," smokes"," Unknown".  12- stroke:1 if the patient had a stroke or 0 if not. |

**Algorithms**

|  |  |
| --- | --- |
| **Algorithm** | I will use classification algorithm. |
| **Evaluating Results** | * Confusion Matrix. * Recall. * Precision. * F1-Score. * Accuracy. * ROC/ AUC. * Log Loss. |

**Tools**

|  |  |
| --- | --- |
| **Data Processing** | Pandas, Numpy, Scipy. |
| **Modelling** | Scikit-Learn, XGBoost. |
| **Visualization** | Matplotlib,Seaborn,Bokeh. |
| **save Model** | Pickle. |