

Project Proposal

LAMYA ALARWAN

Design

Project Title	Stroke Prediction in Patients.
Project Goal	<p>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.</p> <p>This project predicts whether a person is at risk of having a stroke or not.</p>
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	<p>The dataset source is Kaggle (Stroke Prediction Dataset). It is containing 5110 records and 12 features.</p> <p>Record: Contains patient information and if patient had a stroke or not.</p> <p>Target features: [stroke]=1 if the patient had a stroke or 0 if not.</p>
Describe Features	<p>1- id: Unique id.</p> <p>2- gender: "Male "or "Female".</p> <p>3- age: age of the patient.</p> <p>4- hypertension: binary feature (0 if the patient doesn't have hypertension, 1 if the patient has hypertension).</p> <p>5- heart_disease: binary feature (0 if the patient doesn't have any heart diseases, 1 if the patient has heart diseases).</p> <p>6- ever_married: "Yes" or "No".</p> <p>7- work_type: " children", " Private", " Self-employed", " Govt_job" or " Never_worked".</p> <p>8- Residence_type: " Rural" or " Urban".</p> <p>9- avg_glucose_level: Average glucose level in blood.</p> <p>10- bmi: Body Mass Index.</p> <p>11- smoking_status: " never smoked", " formerly smoked", " smokes", " Unknown".</p> <p>12- stroke: 1 if the patient had a stroke or 0 if not.</p>

Algorithms

Algorithm	I will use classification algorithm.
Evaluating Results	<ul style="list-style-type: none">• 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.