

SYNTHETIC STROKE PREDICTION

Using Machine Learning

"Prevention is better than cure – with machine learning, prediction becomes prevention."

Computational Intelligence

SDG GOALS

SDG 3: Good Health and Well-Being

SDG 9 – Innovation in Healthcare Technology



Introduction

Stroke is one of the most critical medical emergencies, often leading to severe disability or death.

Predicting the chances of a stroke before it occurs can save lives through early diagnosis and preventive care.

Using machine learning, we can analyze health and lifestyle patterns to identify people at higher risk and support timely medical intervention.

Objective

- Develop a synthetic stroke prediction model using ML.
- Handle missing data and class imbalance effectively.
- Compare multiple algorithms for the most accurate prediction.
- Prepare the model for deployment in real-world systems.

Dataset

Synthetic Stroke Prediction (Source: Kaggle)

- Records: ~50,000
- Features: 10 + 1 target

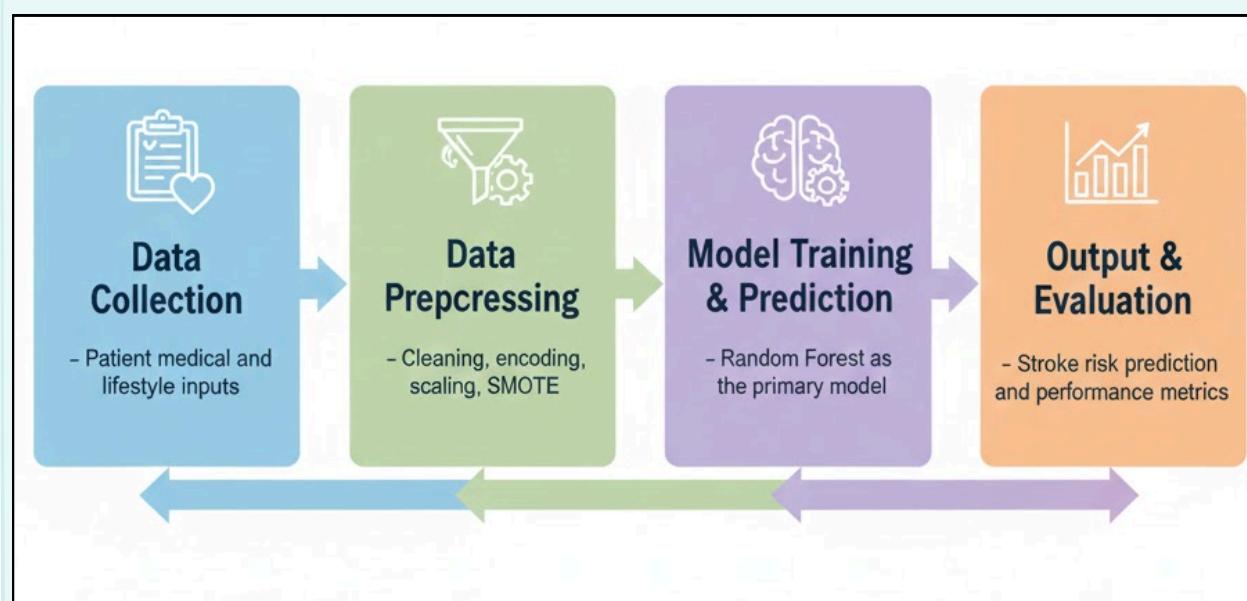
Machine Learning Model Used

A Random Forest Classifier model was used for stroke prediction. It works by building multiple decision trees and combining their outputs to improve accuracy and stability.

Why Random Forest?

- Handles both numerical and categorical data effectively
- Resistant to overfitting
- Provides feature importance insights
- High accuracy and interpretability

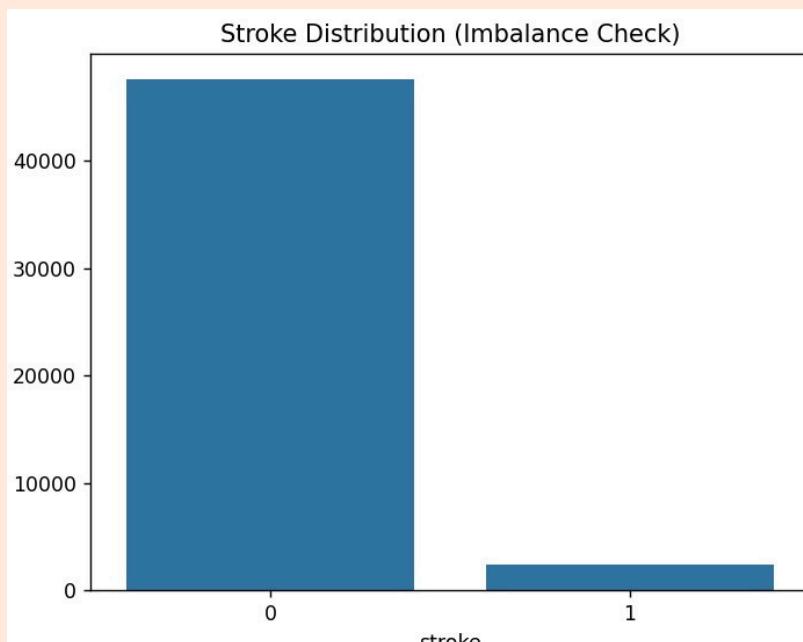
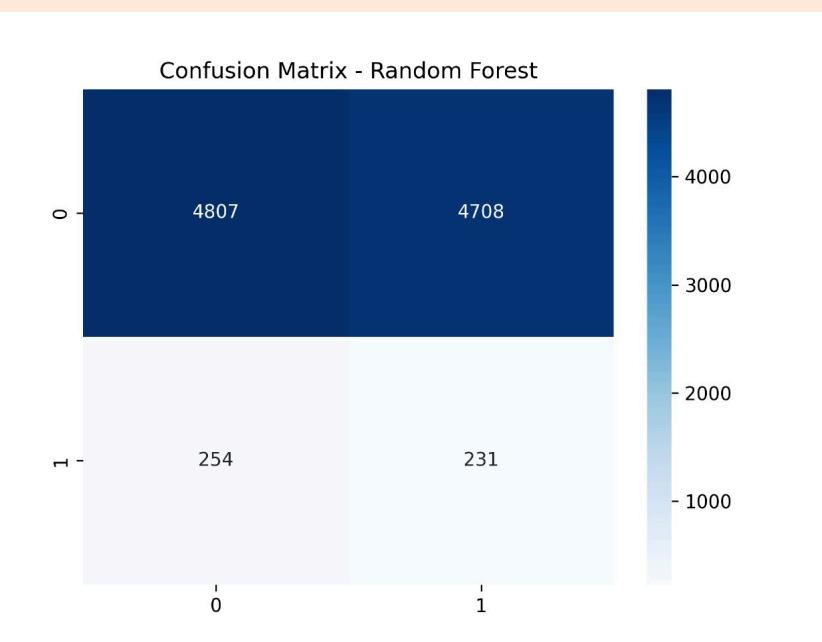
Machine Learning Pipeline



Technologies Used

- VS Code
- Python
- Pandas
- Scikit-learn
- Matplotlib
- Imbalanced-learn

Visualisations



Model Evaluation

==== MODEL RESULTS ====					
	precision	recall	f1-score	support	
0	1.00	0.96	0.98	7123	
1	0.92	1.00	0.96	2877	
accuracy				0.97	10000
macro avg	0.96	0.98	0.97	10000	
weighted avg	0.98	0.97	0.97	10000	
ROC-AUC: 0.9998296480761529					

Sample Output

Stroke Prediction System

Gender	Female
Age	25
Hypertension	No
Heart Disease	No
Ever Married	Yes
Work Type	Govt_job
Residence Type	Urban
Avg Glucose Level	146
BMI	23
Smoking Status	never smoked

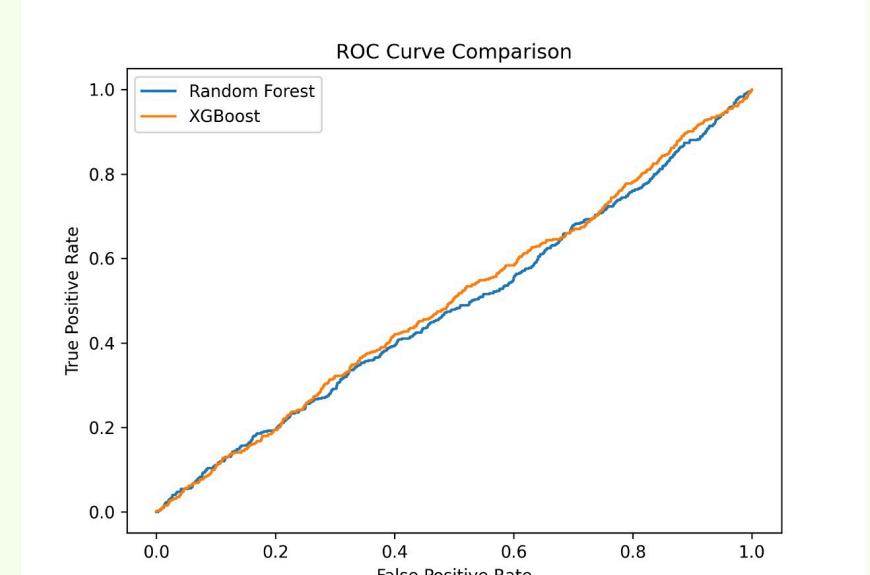
No Stroke Detected

You are currently at LOW risk of stroke.
Maintain a balanced diet and regular exercise.
Go for routine health check-ups annually.
Monitor glucose levels and blood pressure.
Avoid smoking and alcohol overuse.

Predict Stroke Risk

Performance Summary

- Random Forest performed significantly better than Logistic Regression on all major metrics
- Random Forest showed higher recall, important for detecting patients who may be at risk
- Logistic Regression provided good interpretability but lower predictive power



Guided By:
Dr. Abhilasha Joshi

Purva Nigade	202301070146
Aakanksha Sah	202301070148
Pratiksha Shinde	202301070155
Amruthavarshini Repalle	202301070158