# Milestone 3

## Predictive Model Performance Report

### Objective

To develop and optimize predictive models to classify whether a patient is at risk of stroke, based on healthcare attributes. The best-performing model will be selected based on evaluation metrics and suitability for healthcare applications

### Dataset

* Source File: preprocessed\_stroke\_data.csv
* Target Variable: stroke (binary classification)
* Preprocessing: StandardScaler, Train-test split (80/20)

### Models Developed

* Logistic Regression:
  + Accuracy: ~0.93
  + ROC AUC: ~0.84
* Random Forest:
  + Accuracy: ~0.95
  + ROC AUC: ~0.92
* Gradient Boosting:
  + Accuracy: ~0.96
  + ROC AUC: ~0.94
* Neural Network:
  + Accuracy: ~0.94
  + ROC AUC: ~0.90

### Model Optimization

* **Techniques Used:**
  + RandomizedSearchCV for faster hyperparameter sampling
  + GridSearchCV for fine-tuning best parameters
* **Applied to:** Random Forest and Gradient Boosting models

### Final Model Selection

Selected Model: Gradient Boosting Classifier

**Justification:**

* Highest **ROC AUC** and **Recall** among all models
* Robust performance in detecting positive stroke cases
* Lower false negatives, critical for healthcare scenarios
* Tuned hyperparameters showed stable and consistent results

### Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **ROC AUC** |
| Logistic Regression | ~0.93 | Moderate | Lower | ~0.75 | ~0.84 |
| Random Forest | ~0.95 | High | High | ~0.89 | ~0.92 |
| Gradient Boosting | ~0.96 | High | High | ~0.91 | ~0.94 |
| Neural Network | ~0.94 | Moderate | Moderate | ~0.86 | ~0.90 |