# Stroke Prediction Model Summary

## Overview

This project involved developing a machine learning model using R to predict stroke risk. The dataset includes patient characteristics such as age, BMI, glucose level, and medical history. The goal is to identify high-risk individuals and enable early intervention.

## Methods Used

1. Data Preprocessing: Handled missing values, encoded categorical variables.  
2. Exploratory Data Analysis (EDA): Identified risk factors using visualizations and correlation analysis.  
3. Model Training: Implemented Logistic Regression, Decision Tree, Random Forest, and XGBoost.  
4. Model Evaluation: Compared models using Accuracy, Precision, Recall, and F1-score.  
5. Deployment Considerations: Future steps include deploying via RShiny API for real-time prediction.

## Key Findings

• The XGBoost model provided the highest accuracy at 94.6%.  
• Age, BMI, and average glucose level were the most significant features in predicting stroke risk.  
• The dataset was highly imbalanced; future work should include techniques like SMOTE to balance classes.

## Conclusion

The project successfully built a stroke prediction model with promising accuracy. Further improvements could include hyperparameter tuning and integration into a web-based interface for broader accessibility.