

**Project Proposal**  
**Shinjon Ghosh**  
**Department of Mathematics, Illinois State University**

**Project Title:** Stroke Prediction with Regression based Analysis.

**Introduction:**

Stroke is one of the leading causes of morbidity and mortality worldwide. It has significant health and economic impacts on individuals and healthcare systems. Early identification of individuals at risk of stroke or predicting outcomes in patients who have experienced a stroke is crucial for implementing timely preventive measures and improving treatment strategies. This project aims to develop predictive models to expect stroke and identify factors influencing stroke risk by World Health Organization stroke data.

**Research Questions:**

1. What are the key factors influencing stroke?
2. Can we build a predictive model to forecast the stroke based on these factors?
3. How to measure performance of fitted model?

**Objectives:**

- Build a regression model to forecast stroke based on patient demographic and health data.
- Identify key factors influencing stroke risk.
- Evaluate model performance to identify early strokes and enhance patient treatment.

**Methodology:**

- **Data Collection:** Obtain data from Kaggle (Secondary data resource) of Stroke prediction Dataset, including patient demographics and health details.
- **Data Preprocessing:** Clean and prepare data for analysis, handling missing values, and outliers. Explore the dataset to understand distributions and correlations related to retention.
- **Model Development:** Train regression models, such as logistic regression, using preprocessed data to predict stroke probabilities.
- **Model Evaluation:** Evaluate the model's performance and analyze the impact of individual features on predicted stroke risk probabilities.

**Expected Outcomes:**

- A regression model to predict strokes with patient health details.
- Identification of significant predictors and their impact on stroke risk.

- Insights and recommendations for early stroke criteria to mitigate stroke risk.

**Significance:**

- Improve data driven decisions to proactively manage stroke risk and enhance treatment strategies.
- Contribute to a deeper understanding of patient details to predict stroke risk.
- Develop in stroke prevention and management.

**Resources:**

- Access to Stroke Prediction Dataset (collect from Kaggle).
- Statistical software (e.g., R) for data analysis and model development and visualization software (e.g., MS Power BI, Excel) for data visualization.

**Conclusion:**

This project aims to leverage predictive analytics to forecast stroke and provide actionable insights for optimizing stroke risk. By understanding patient demographic and health details, it can improve stroke prevention and management.