Determinants of Recurrent Stroke Incidence in a Clinical Sample

SUPPLEMENTAL TABLES AND FIGURES

Critical Thinking Group 3

Data 621

Dr. Jeffrey Edwards

CUNY School of Professional Studies

Coco Donovan

Jean Jimenez

Marjete Vucinaj

Matthew Roland

Table 1: Descriptive statistics of key numerical variables for patients in a recurrent stroke dataset, detailing means, standard deviations, medians, interquartile ranges, and value distributions along with counts of valid and missing observations.

Summary Statistics for Recurrent Stroke Numerical Variables

Stroke Dataset Analysis[1]

Variable	Mean	Standard Deviation	Median	Interquartile Range	Minimum	Maximum	Valid Observations	Missing Observations
age	71.36	14.45	73.00	21.00	18.00	121.00	23730	0
Length_of_stay_hours	167.63	214.77	118.00	127.00	-2,032.00	9,666.00	23728	2
MRS_discharge_score_cleaned	2.07	1.72	2.00	4.00	0.00	8.00	20345	3385
Arrival_NIHSS_score	6.55	10.68	3.00	8.00	-7.00	999.00	17788	5942
Arrival_NIHSS_score_cleaned	6.50	7.66	3.00	8.00	0.00	42.00	17786	5944
hasIVTPA	0.12	0.33	0.00	0.00	0.00	1.00	23730	0
BMI	27.79	7.67	26.78	7.40	2.03	259.18	19296	4434
TARGET	0.15	0.36	0.00	0.00	0.00	1.00	23730	0

[1] Analysis conducted on the stroke dataset.

Figure 1: Residuals plot comparing model fits for logistic regression, lasso regression, ridge regression, and decision tree models based on observed binary outcomes.

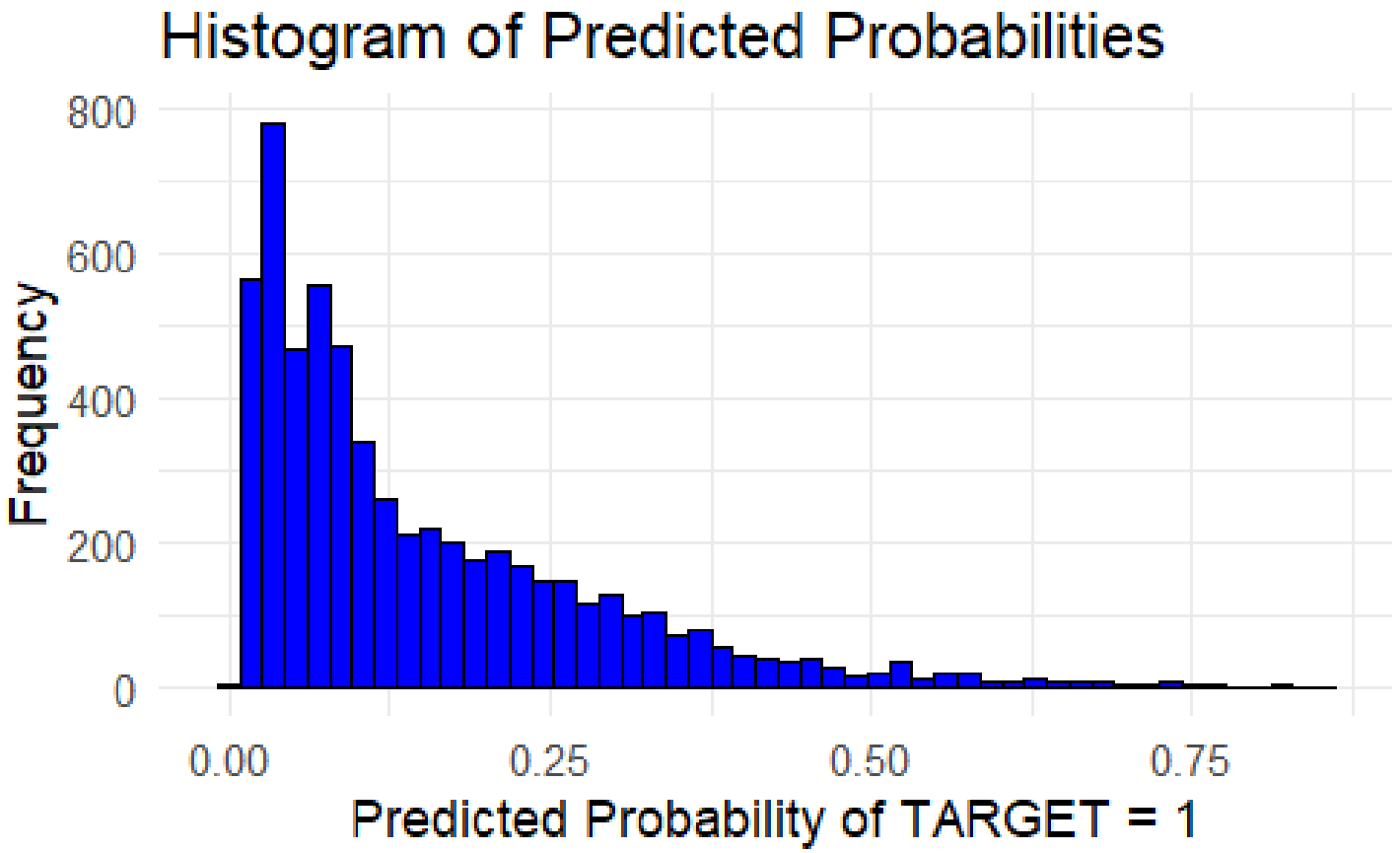

Figure 2: Histogram displaying the distribution of predicted probabilities for the binary outcome TARGET = 1 (stroke is recurrent) in the predictive model

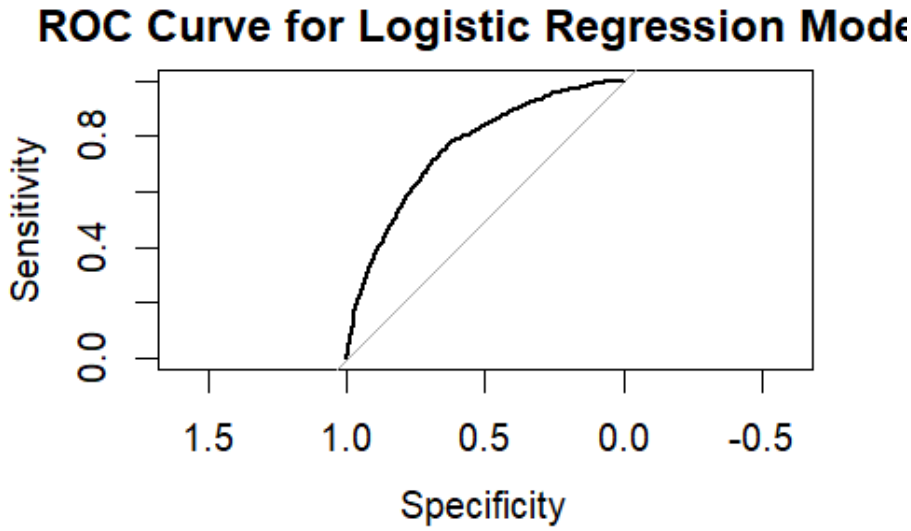

Figure 3: Receiver Operating Characteristic (ROC) curve for evaluating the performance of a logistic regression model, plotting sensitivity against 1-specificity

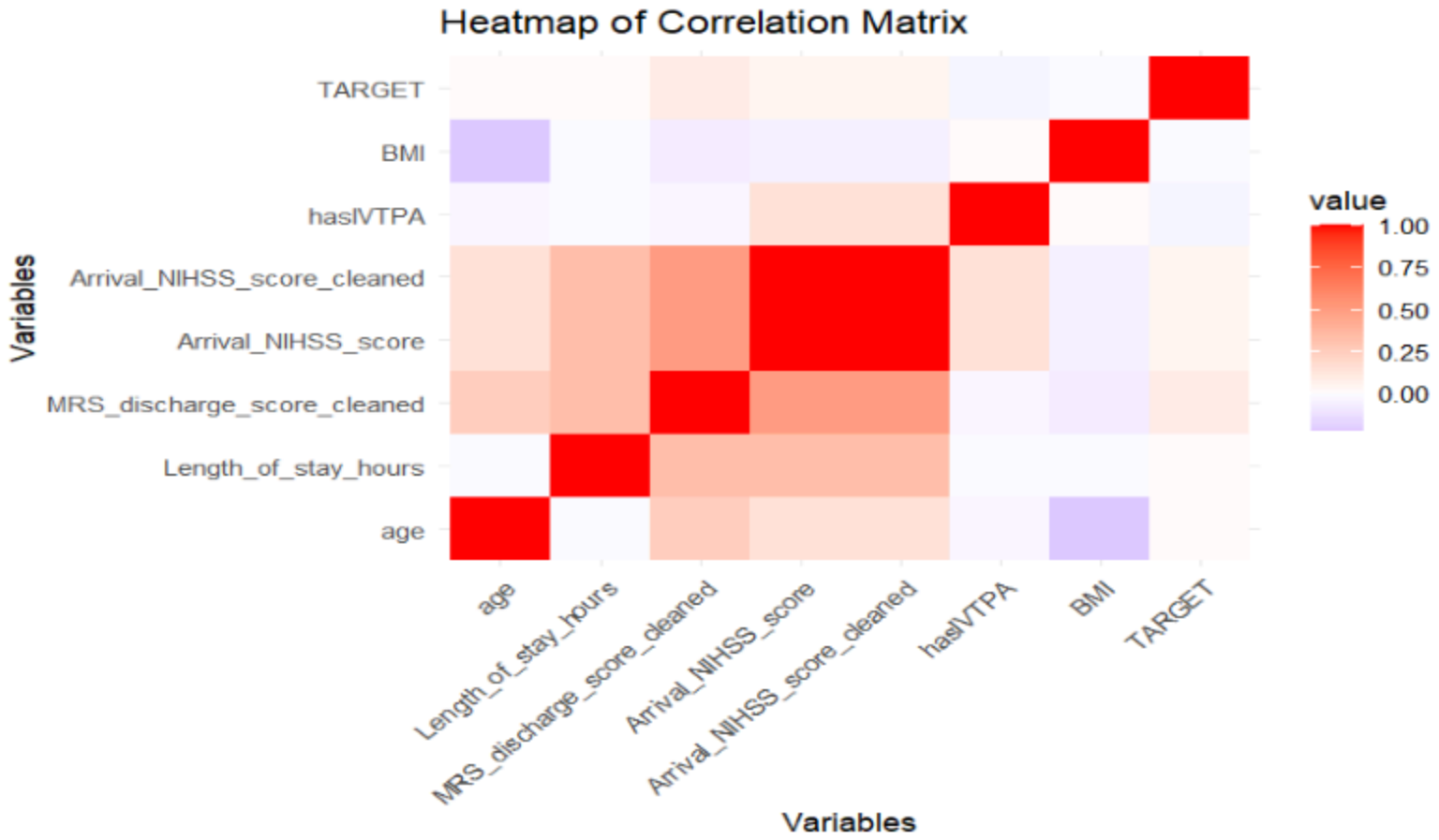

Figure 4: Heatmap displaying the correlation matrix of selected variables in a stroke patient dataset, with color intensities representing the strength of correlation between each pair of variables.

Table 2: Comparison of mean squared errors (MSE) across different predictive models including logistic regression, lasso regression, ridge regression, and decision tree, evaluating their performance in a predictive analytics context.

Model Comparison

Mean Squared Errors of Predictive Models

Model	Mean Squared Error
Logistic Regression	0.1094
Lasso Regression	0.1094
Ridge Regression	0.1095
Decision Tree	0.1277