

Treatments Done on Parent Dataset							
Name of Derived Dataset	Gender (removed 'other' column)	Label Encoding of Categorical Variables	Handling Missing Values	Oversampling Using SMOTE	Removed 'id'	Removed 'Residence_type'	Outlier Treatment
Data1_treated.csv	Yes	Yes	Imputed Mean	Yes	No	No	No
Data2_treated.csv	Yes	Yes	Imputed Median	Yes	No	No	No
Data3_treated.csv	Yes	Yes	Imputed Mode	Yes	No	No	No
Data3-0_treated.csv	Yes	Yes	Imputed Mode	Yes	Yes	No	No
Data3-1_treated.csv	Yes	Yes	Imputed Mode	Yes	Yes	No	Dropped Outliers
Data3-2_treated.csv	Yes	Yes	Imputed Mode	Yes	Yes	No	Imputed Mode
Data3-3_treated.csv	Yes	Yes	Imputed Mode	Yes	Yes	Yes	Imputed Mode

Table of Treatments Done on Parent Dataset '5k_stroke-data.csv'

Evaluation Results of ML Classification Models Built on Derived Datasets			
Name of Derived Dataset	Confidence Threshold	F1 Score	Log Loss
Data1_treated.csv	0.50	0.9758794	0.067
Data2_treated.csv	0.50	0.9688150	0.087
Data3_treated.csv	0.50	0.9853095	0.050
Data3-0_treated.csv	0.50	0.9771372	0.065
Data3-1_treated.csv	0.50	0.9662699	0.093
Data3-2_treated.csv	0.50	0.9743853	0.064
Data3-3_treated.csv	0.50	0.9789579	0.067

Table of Evaluation Results of ML Models Built on Datasets Derived from '5k_stroke-data.csv'