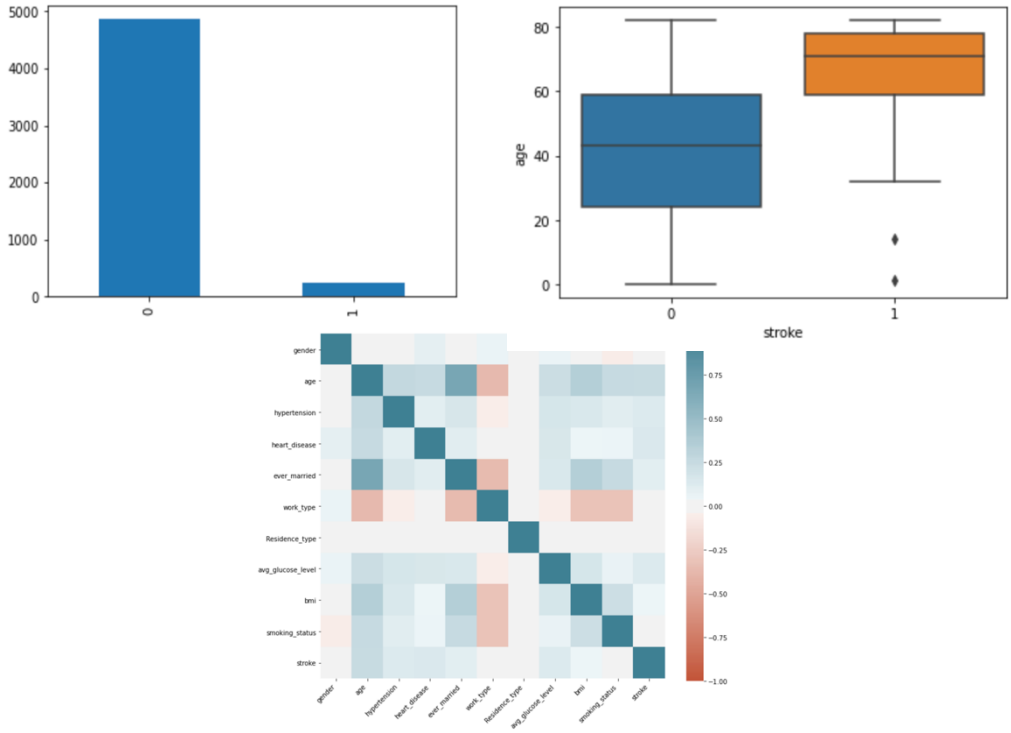
Stroke Prediction

ABSTRACT

Stroke is the sudden death of some brain cells due to lack of oxygen when the blood flow to the brain is lost by blockage or rupture of an artery to the brain, it is also a leading cause of dementia and depression. According to World Health Organization (WHO), stroke are the second leading cause of death and the third leading cause of disability globally. With the advancement of technology in the medical field, predicting the occurrence of a stroke can be made using Machine Learning.

DATASET

dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status



MODELING

The following models were trained on Jupyter Notebook after cleaning and Preprocessing.

The models used:
Logistic Regression.
Naive Bayes.
SVM.
KNN.

	Before Handling Imbalance		After Handling Imbalance using under sampling		After Handling Imbalance using over sampling	
ML Models	F1 score	Confusion matrix	Accuracy	Confusion matrix	Accuracy	Confusion matrix
Logistic Regression	0.0	[[1458 0] [75 0]]	0.76	[[56 19] [16 59]]	0.83	[[1385 74] [221 78]]
Naive Bayes	0.23	[[1319 139] [48 27]]	0.72	[[62 13] [29 46]]	0.80	[[1263 196] [150 149]]
Support Vector Machine (SVM)	0.0	[[1458 0] [75 0]]	0.75	[[53 22] [16 59]]	0.85	[[1442 17] [250 49]]
KNeighborsClassifier (KNN)	0.0	[[1451 7] [75 0]]	0.71	[[53 22] [22 53]]	0.86	[[1290 169] [84 215]]

RUSELT

Naive Bayes model before handling imbalance yields a good performance as indicated by the model F1 score which was found to be 0.23. After handling imbalance using under sampling Logistic Regression model yields a good performance as indicated by the model accuracy which was found to be 0.76. After handling imbalance using over sampling KNN model yields a good performance as indicated by the model accuracy which was found to be 0.86. It is clear from the above table that the accuracy is better after handling imbalance using over sampling.

CONCLUSION

Stroke is a critical medical condition that should be treated before it worsens. Building a machine learning model can help in the early prediction of stroke and reduce the severe impact of the future. This paper shows the performance of various machine learning algorithms in successfully predicting stroke.