Stroke predection

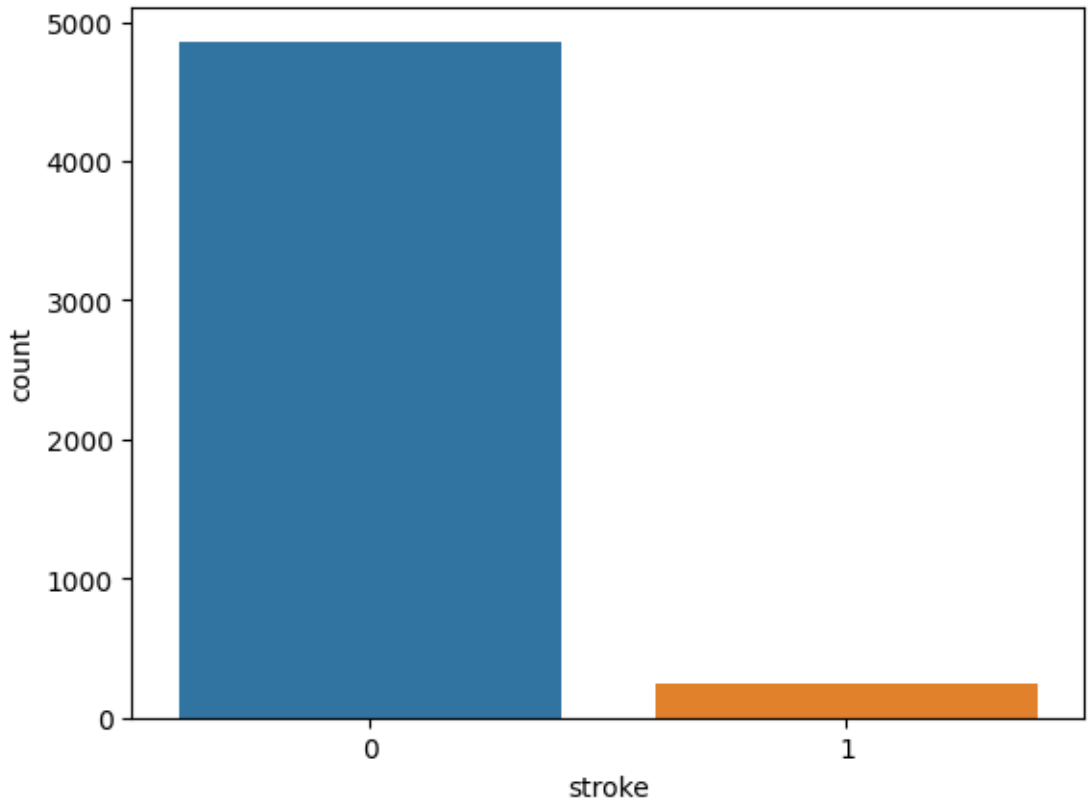
Dataset:

Link in Kaggle: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

- This data contains some health and social informations about humans[males and females] can be used to be a features and we target to predict if the person has stroke or no

-Data original shape:(5110,12)

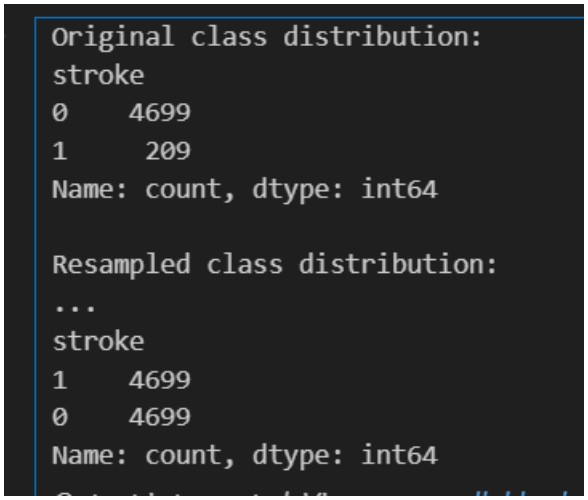
Data wrangling & Preprocessing



-According this plot we get there is imbalance between values of this column so , we use oversampling

SmoteAlgorithm tp balance this feature .

Comparing before and after using smote:



-Encoding all categorical columns to 1,0 using map() function

-splitting data into train and test

With test size = 0.2
,random_state=42

X_train.shape(7518,7)

Y_train.shape(7518,)

X_test.shape(1880,7)

Y_test.shape(1880,)

Modeling

-we`ve used DecessionTree Algorithm with parameters:

[max_depth=9 ,
min_samples_leaf=1,

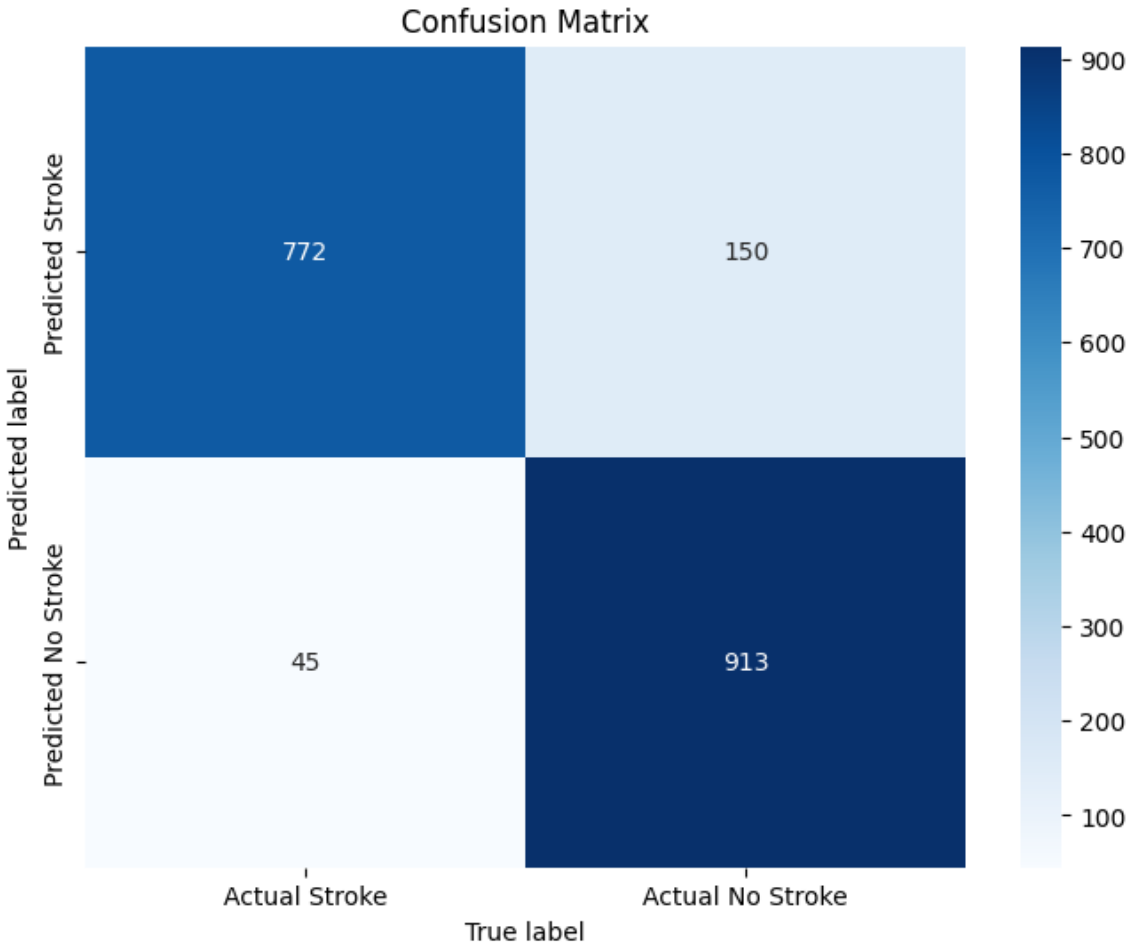
min_samples_split=3,

criterion=’gini’,

random_state=42]

Evaluating Model

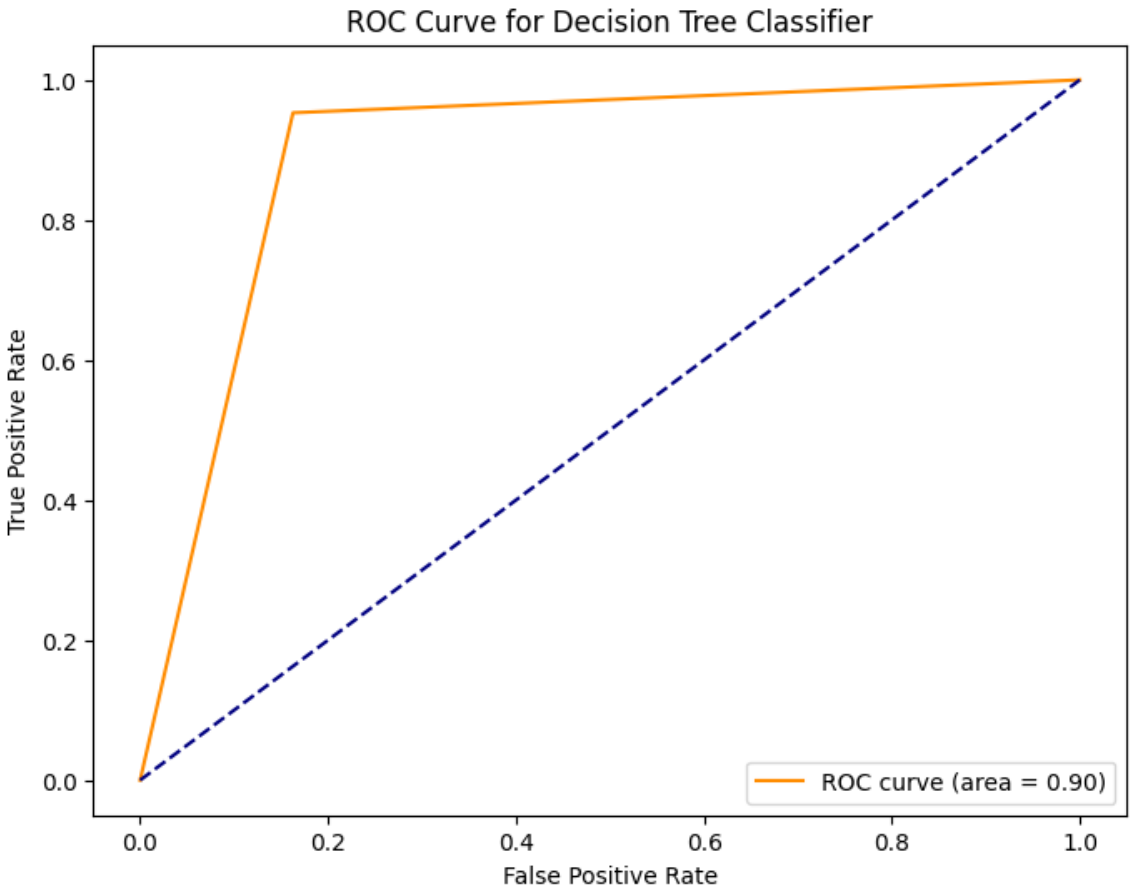
Confussion matrix:



Classification_report

	precision	recall	f1-score	support
0	0.94	0.84	0.89	922
1	0.86	0.95	0.90	958
...				
accuracy			0.90	1880
macro avg	0.90	0.90	0.90	1880
weighted avg	0.90	0.90	0.90	1880

ROC curve


ROC Curve for Decision Tree Classifier
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.90)