

Strokes Classification

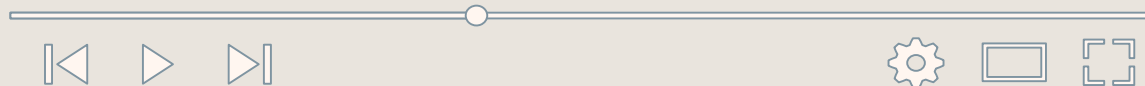
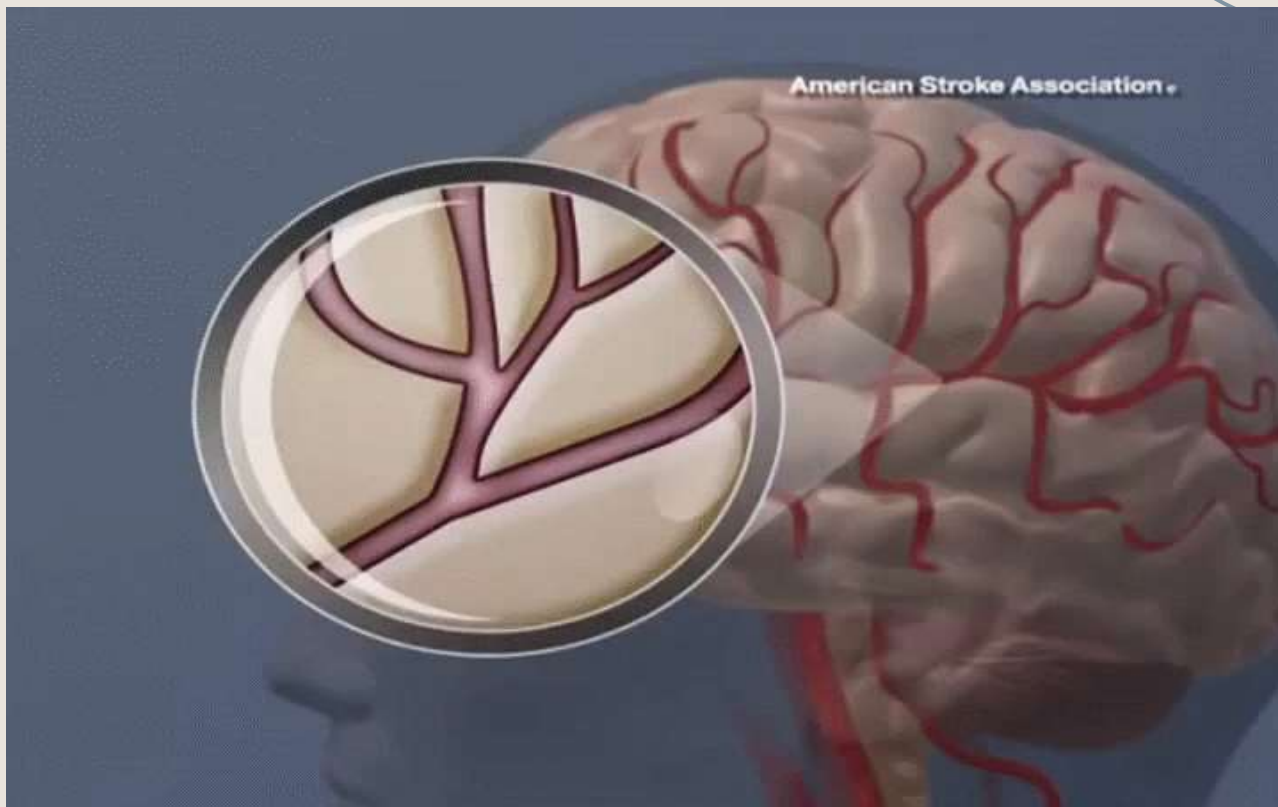


Introduction

A stroke occurs when the blood supply to part of the brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients. Brain cells begin to die in minutes.

A stroke is a medical emergency, and prompt treatment is crucial. Early action can reduce brain damage and other complications.





Objective

is to build a classification model to detect stroke and evaluate the model using some performance metrics

Methodology

**Problem
understanding**



**Gathering
data**



EDA



**Data
preparation**



**Classification
algorithms**



Data description

database that has been used is provided by [Kaggle](#). this data provided in .CSV format and contains 5110 rows and 12 columns

Label :

- | | |
|-----------------------|----------------------|
| 1- Gender | 6- Residence type |
| 2- Age of the patient | 7- avg glucose level |
| 3- Hypertension | 8- work type |
| 4- Heart disease | 9- body mass |
| 5- Ever married | 10- smoking status |



Target :

stroke

Feature Engineering

We added new column age group depending on column age contain 3 category:

Old

Adult

Child

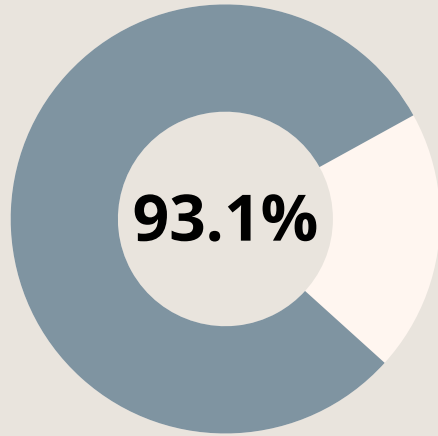




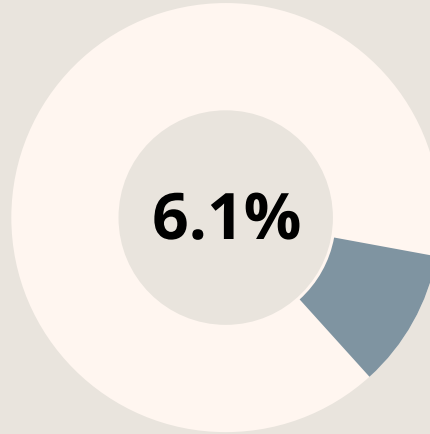
EDA



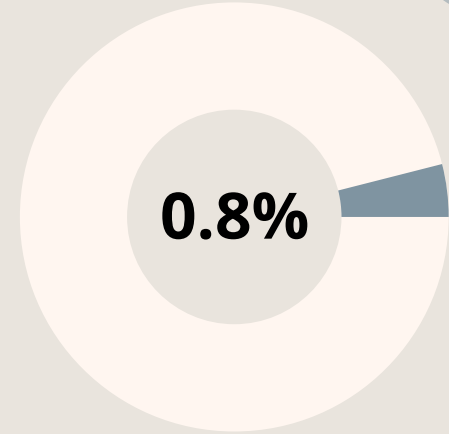
Which age group get stroke most ?



Old



Adult



Child

Which gender is got stroke most?



43.4%



56.6%

56.6%

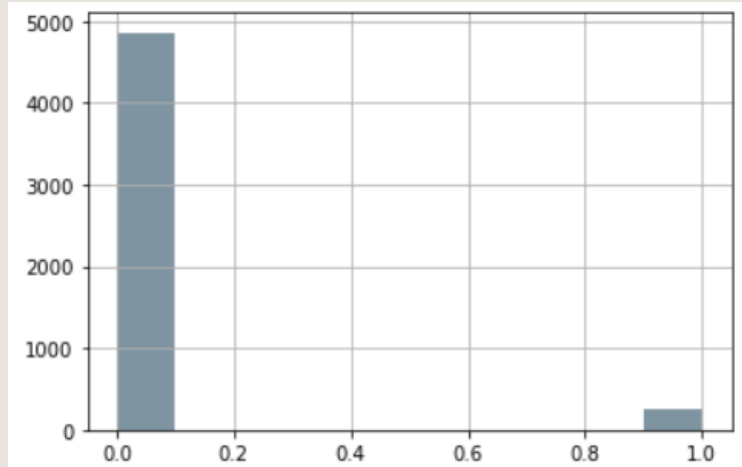
Female stroke

Depending on the data the female is get stroke more than male .

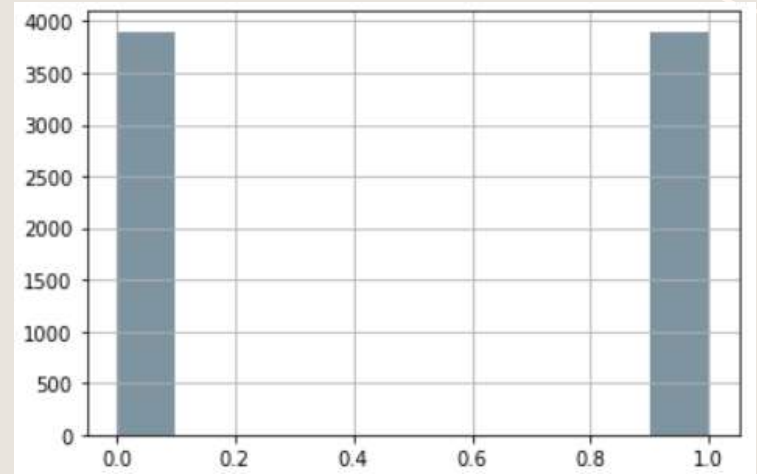


● **Classificaton**

Handling with Imbalance data



Imbalanced of target
variable



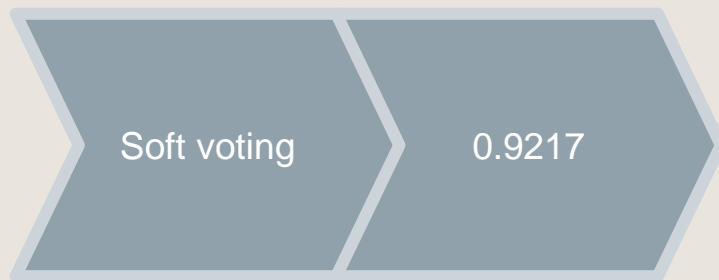
Balanced of target
variable

Model Evaluation

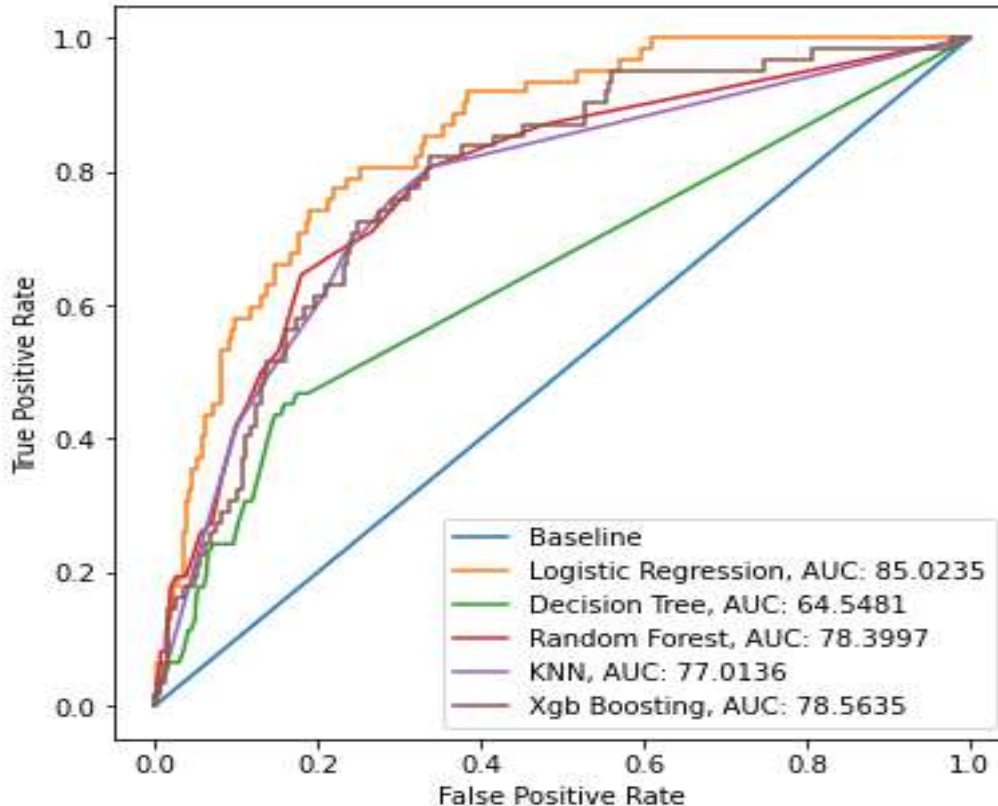
	Accurcy	F-1	precision	Recall
Logistic regression	0.7485	0.2801	0.1694	0.8064
Knn	0.8072	0.2676	0.1736	0.5806
Decision tree	0.9021	0.1666	0.1724	0.1612
Random forest	0.9344	0.0821	0.2727	0.0483
XGboost	0.9354	0.1538	0.375	0.0967

* All models after tuned

Voting & Stacking



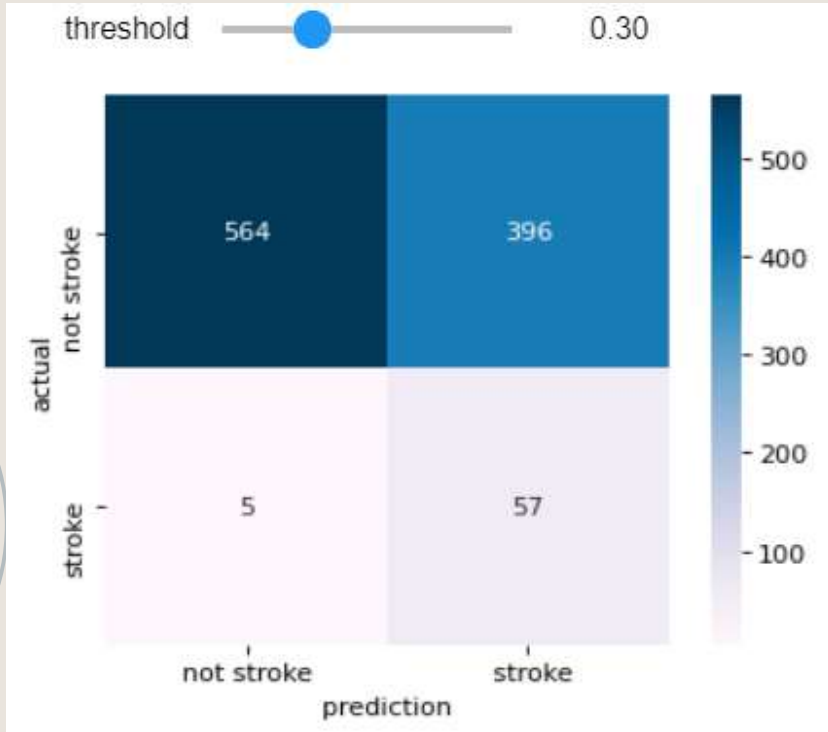
ROC CURVE



■ **Logistic Regression**

Is the beast model.

Best Model & Result



Logistic Regression

Accuracy: 74%
Recall: 91%
Precision: 12%
AUC: 85%



Conclusion:

- At the end from Roc curve, we can assume that the best model is Logistic Regression.
- In the feature work we want to improve accuracy for the best model.





Thanks

Do you have any questions?

Presented by Modhi and Razan

