

---

# Healthcare Stroke Prediction

---

Presented by :  
Tahani Alqahtani  
Amani Albalawi

---



# Presentation Outline



Today's Topics



Background



data Description



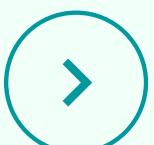
Modeling-classification



Experiment and modeling summary



Conclusion



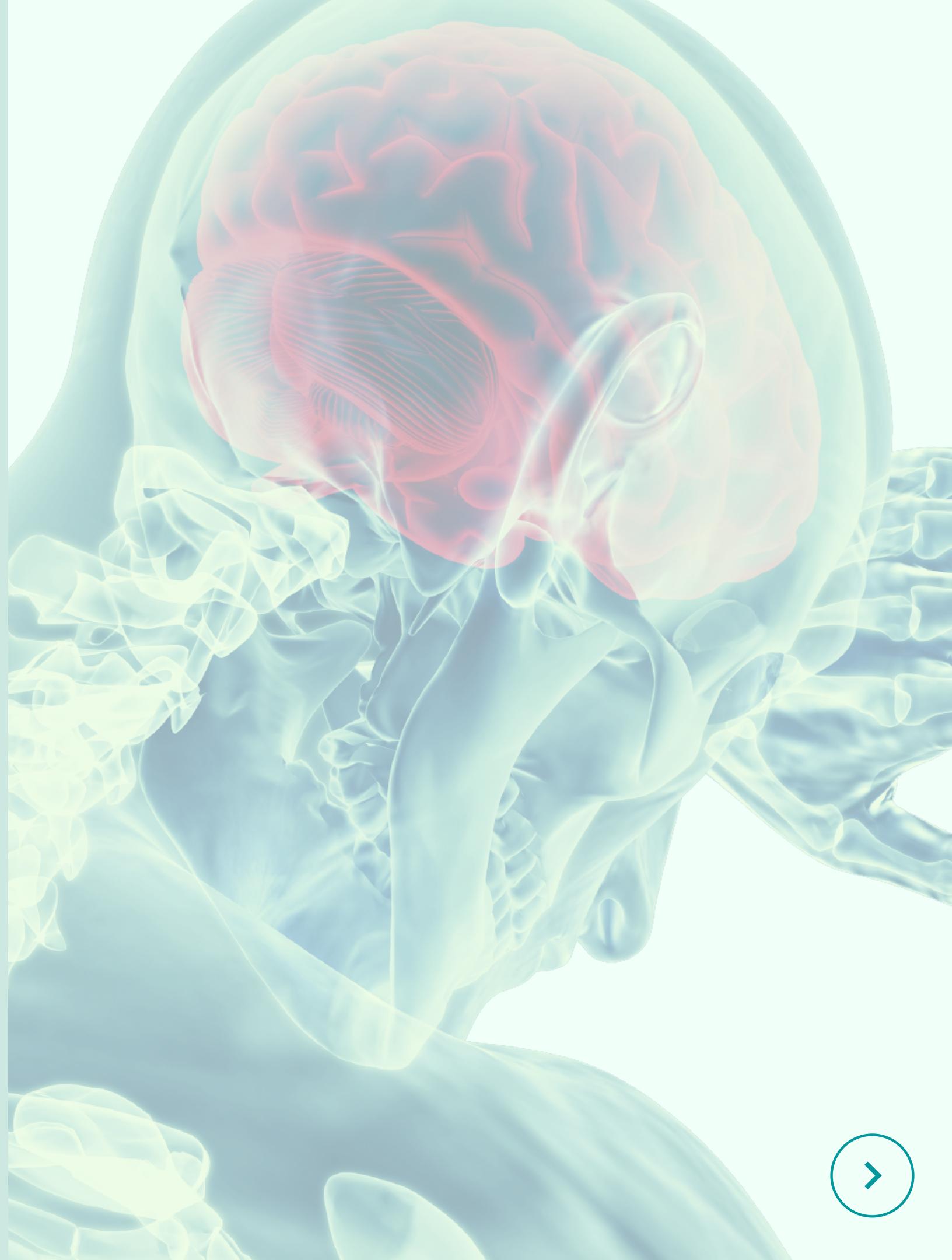
# Background

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally

**11%**

total deaths of stroke

This 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. Each row in the data provides relevant information about the patient.



## Data description:

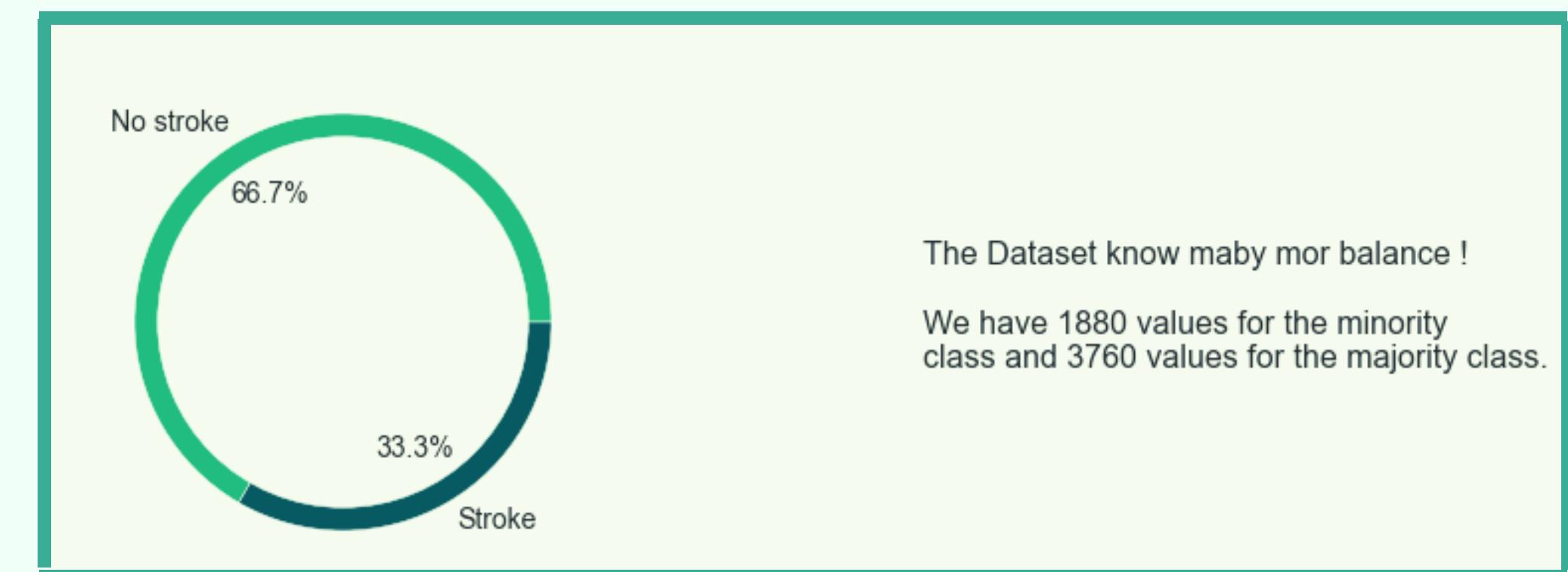
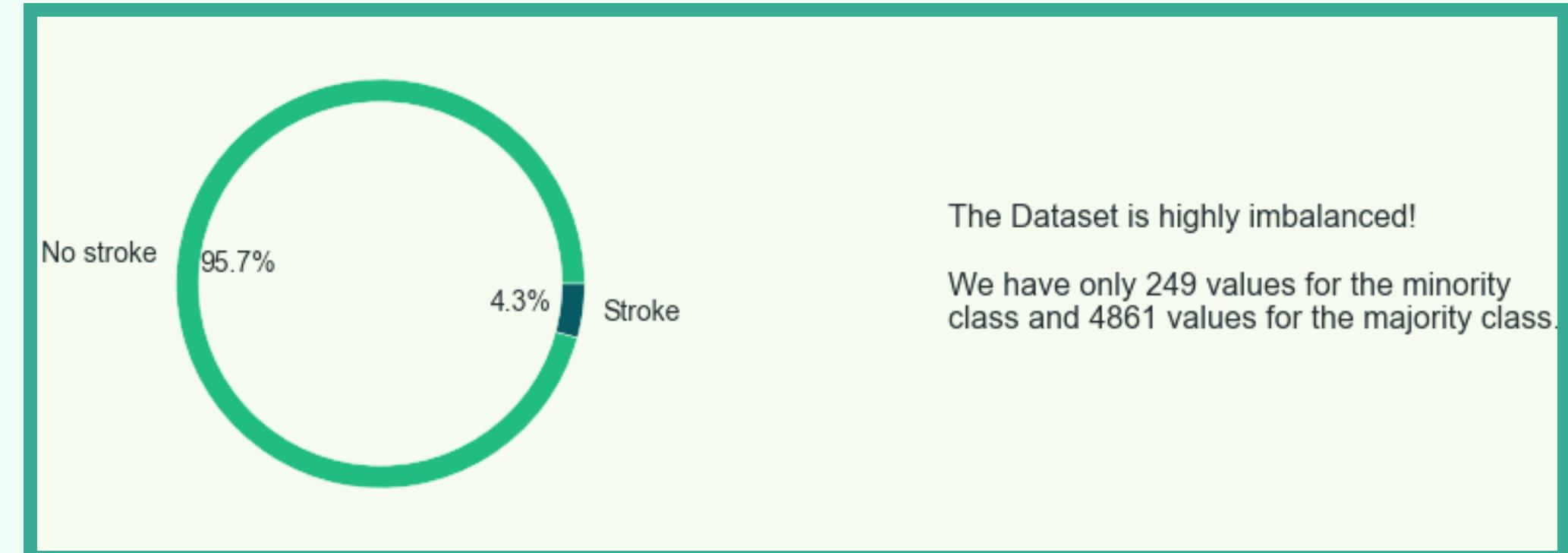
The original source for this data from kaggel . This data set is name as healthcare-dataset-stroke-data contain 12 column and 5111 rows.



# Handling The Class Imbalance

Before Modeling Training -  
Resampling The Data

With oversampling



# Modeling-classification

01

LOGISTIC REGRESSION

02

KNEIGHBORSCLASSIFIER

03

DECISIONTREECLASSIFIER

04

RANDOM FOREST

05

XGB-CLASSIFIER



## Evaluations

### logistic regression

fbeta\_score

0.4593

Precision

0.1382

Recall

0.6190

# Baseline Model

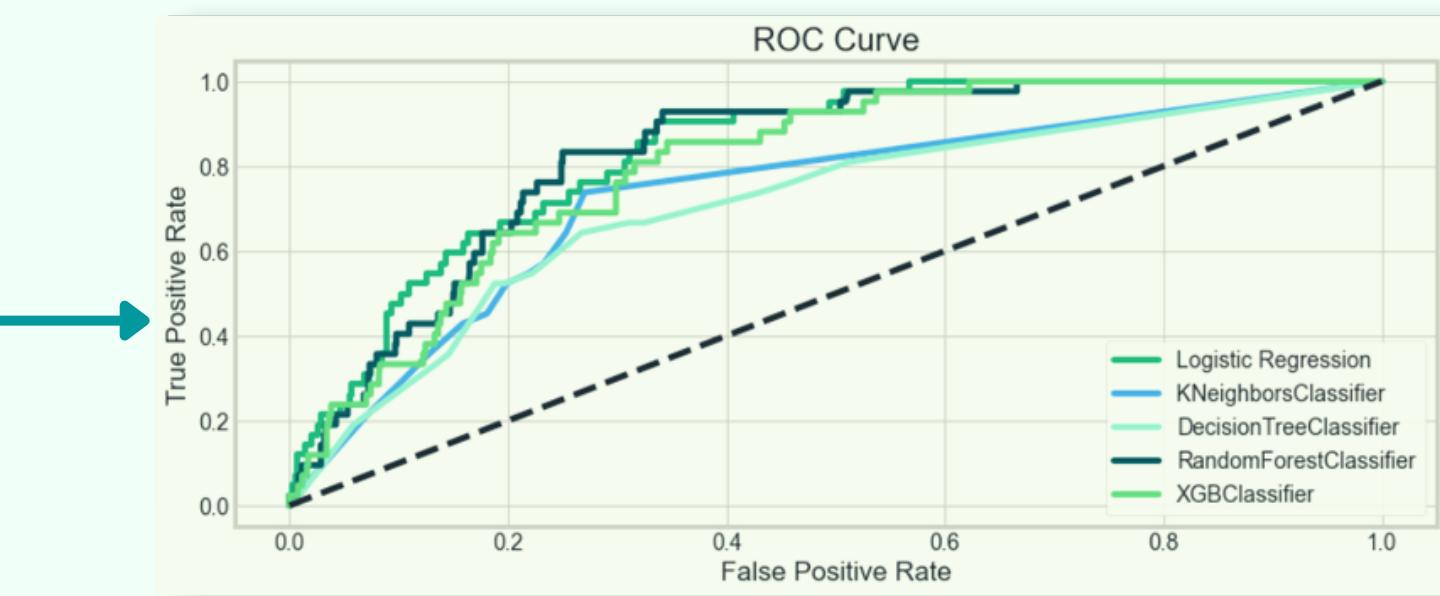
# Experiment and modelling summary

Evaluations	logistic regression	KNeighbors Classifier	DecisionTree Classifier	Random forest	XGB-Classifier
fbeta_score	0.4728	0.3350	0.3760	0.3591	0.3112
Precision	0.1398	0.1005	0.1062	0.1258	0.1176
Recall	0.6428	0.4523	0.5238	0.4523	0.3809

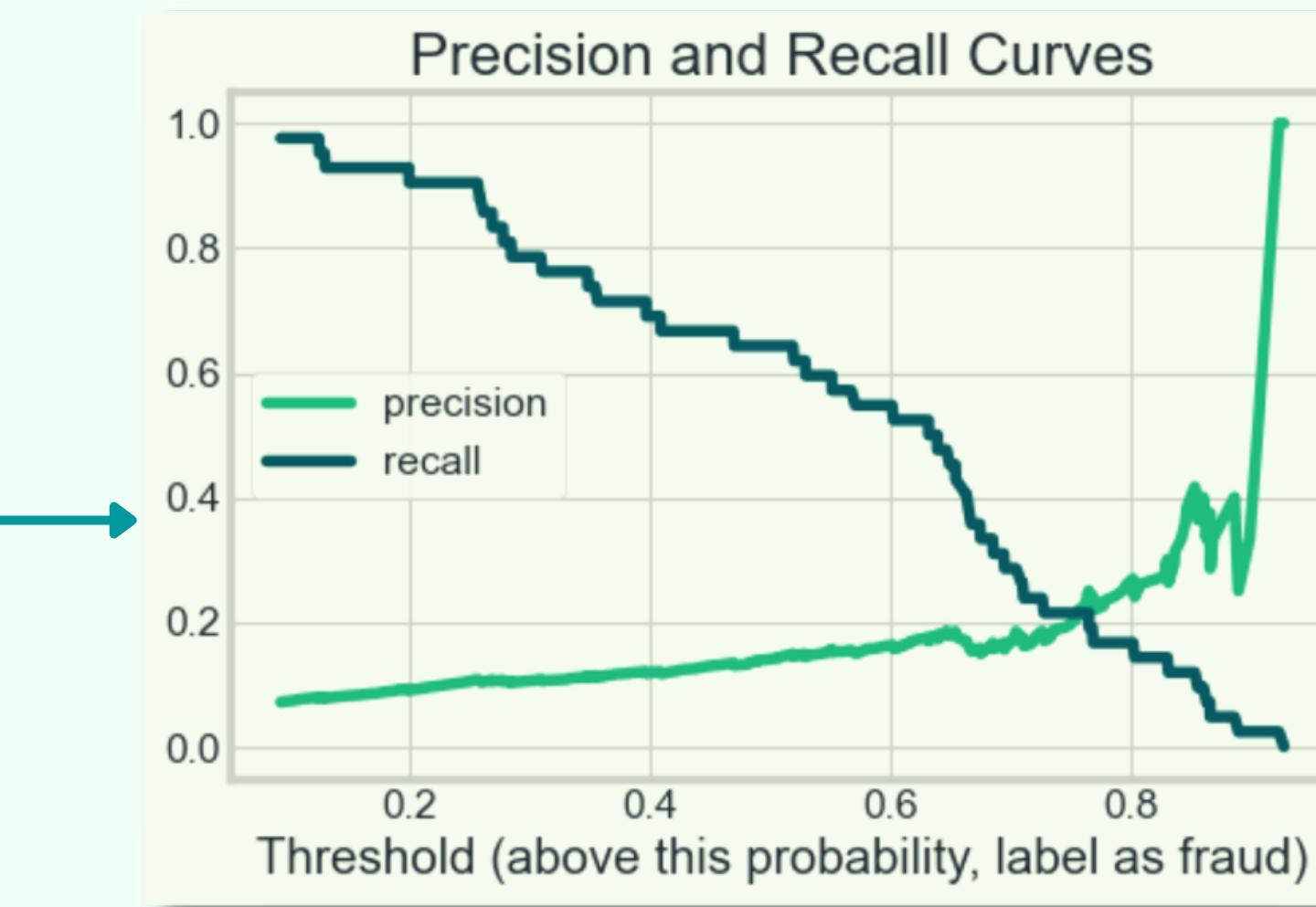


# Conclusion

The Best Model is Logistic Regression based on fbeta\_score and ROC curves



Logistic Regression precision and recall curves



THANK  
YOU

Do you have  
any  
Questions?

