

# Cardiovascular Diseases Prediction

By: Ruqaiah Abdullah & Alanoud Alzahrany

# Presentation Outline

---

- **Introduction**
- **Mythology**
- **Conclusion**

# Introduction

- Diseases that affect the structures or function of someone heart.
- The leading cause of death globally.
- Early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes.



**What are cardiovascular diseases?**

# Mythology



## *Dataset*

from [kaggle.com](https://www.kaggle.com)

## *EDA*

imbalanced data and Oversampling

## *Machine Learning Models*

Experiments and results

# Dataset

(4238, 15)

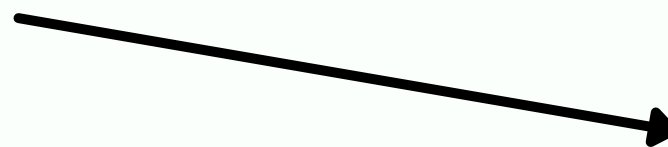
—

|                   |
|-------------------|
| Gender            |
| Age               |
| Current Smoker    |
| Cigs Per Day      |
| BP Medication     |
| Hypertensive      |
| Cholesterol level |
| Systolic BP       |
| Diastolic BP      |
| BML               |
| Heart Rate        |
| Glucose           |
| Ten years risk    |

# Dataset

(4238, 15)

Target: Predicting the number of people who will have cardiovascular diseases within ten years

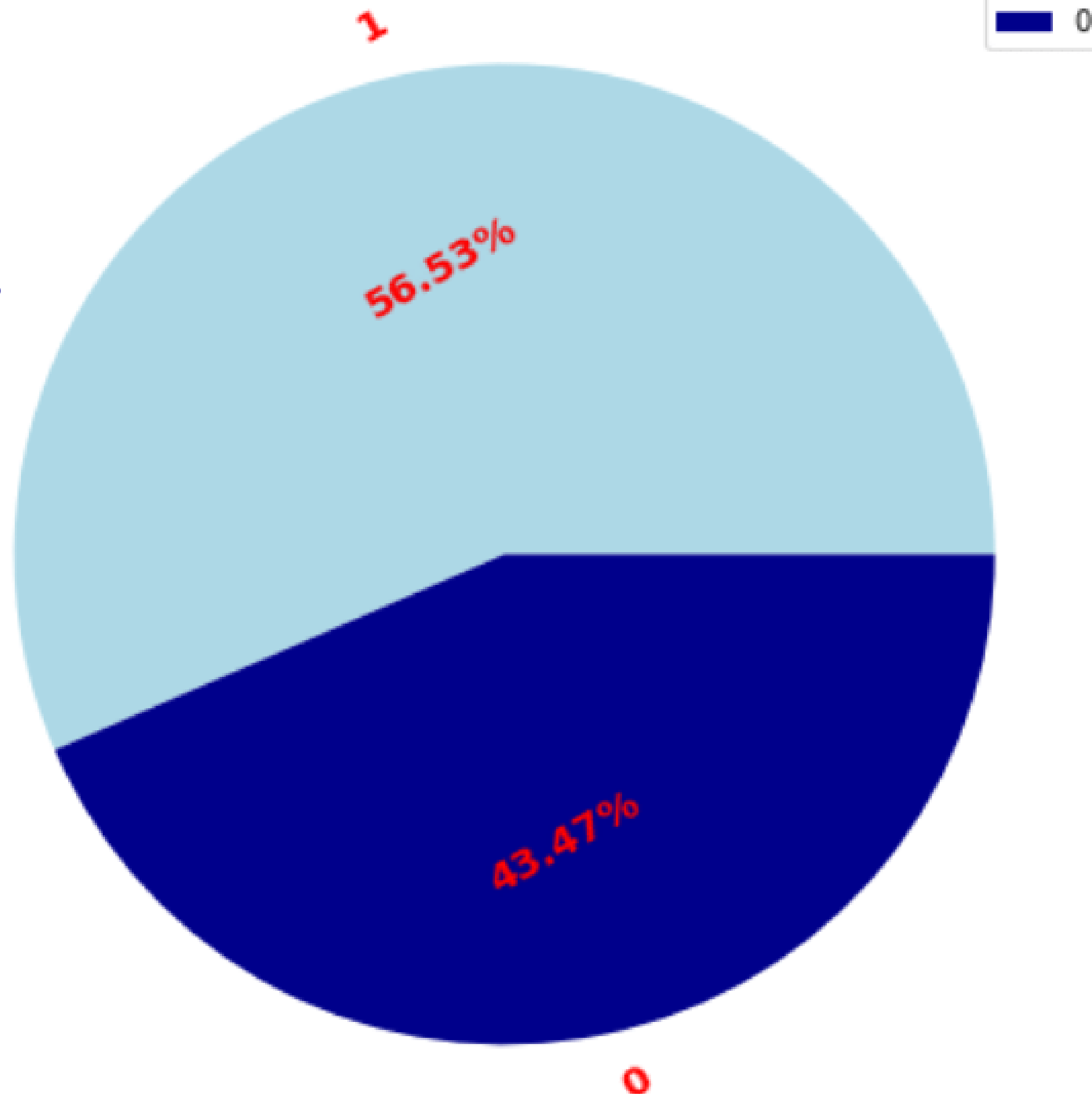


|                   |
|-------------------|
| Gender            |
| Age               |
| Current Smoker    |
| Cigs Per Day      |
| BP Medication     |
| Hypertensive      |
| Cholesterol level |
| Systolic BP       |
| Diastolic BP      |
| BML               |
| Heart Rate        |
| Glucose           |
| Ten years risk    |



# EDA

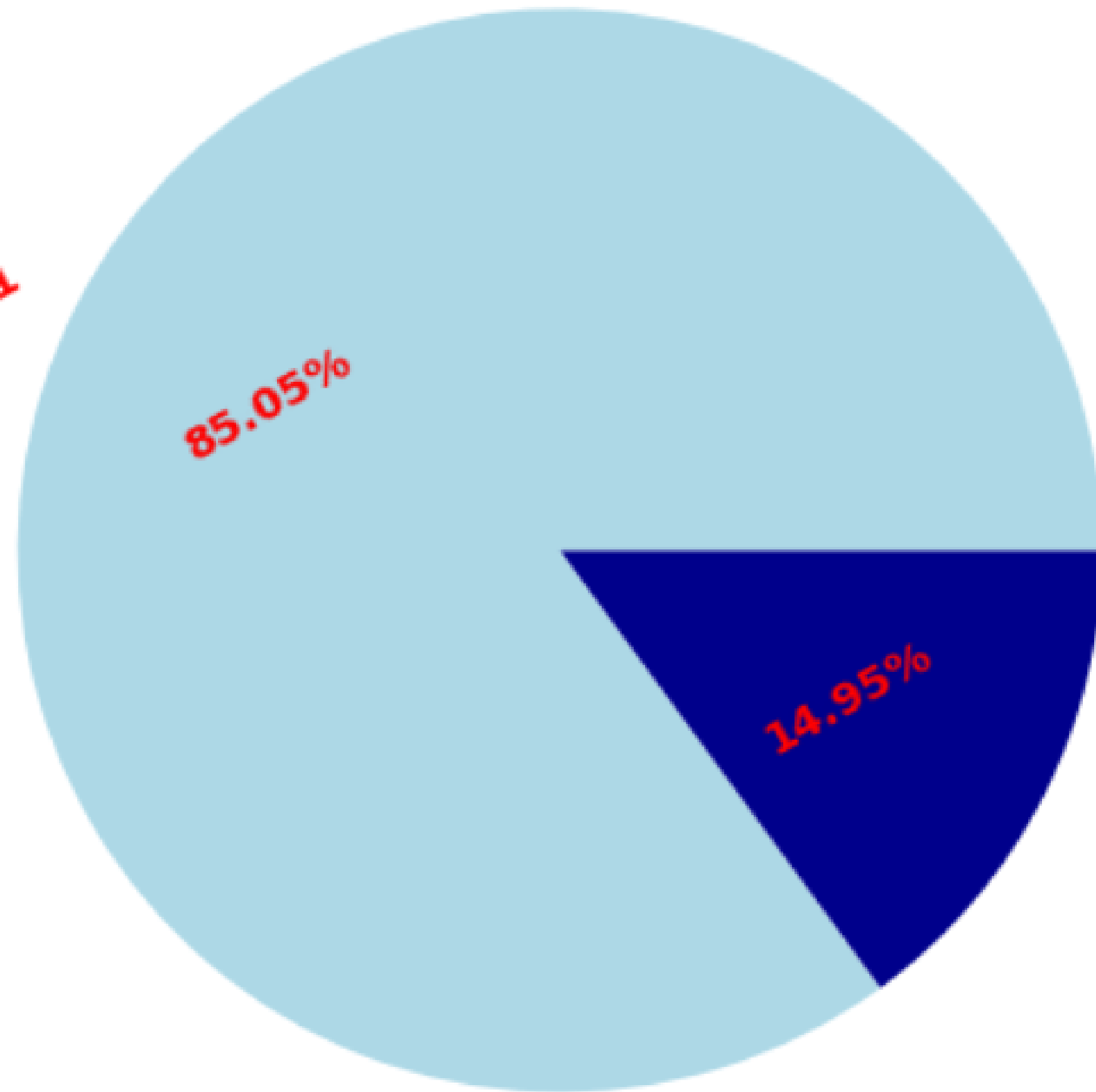
## Percentage of Gender Types



# EDA

Percentage of Ten years risk

1



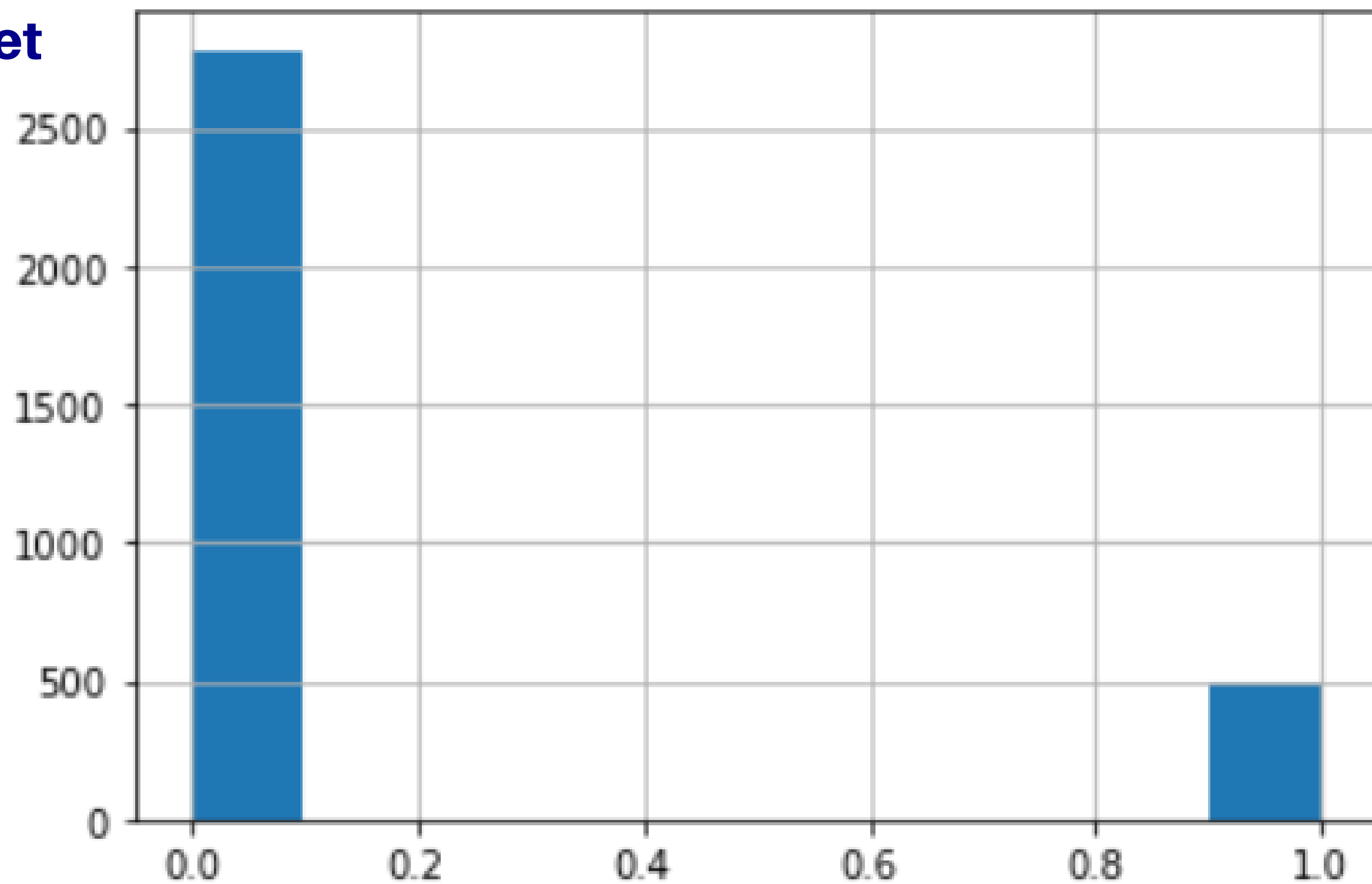
0





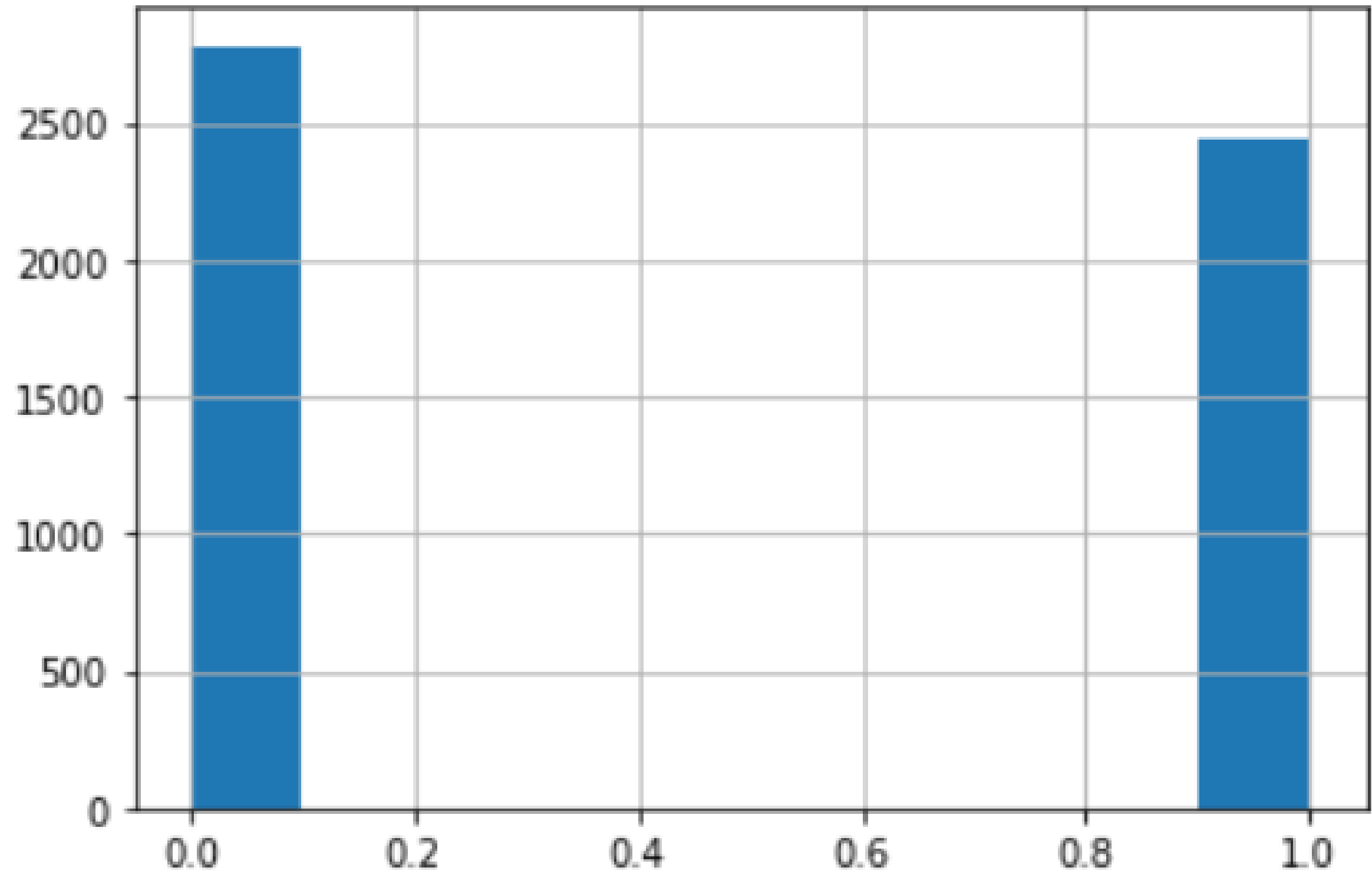
# EDA

## Imbalanced Dataset



# EDA

## Oversampling





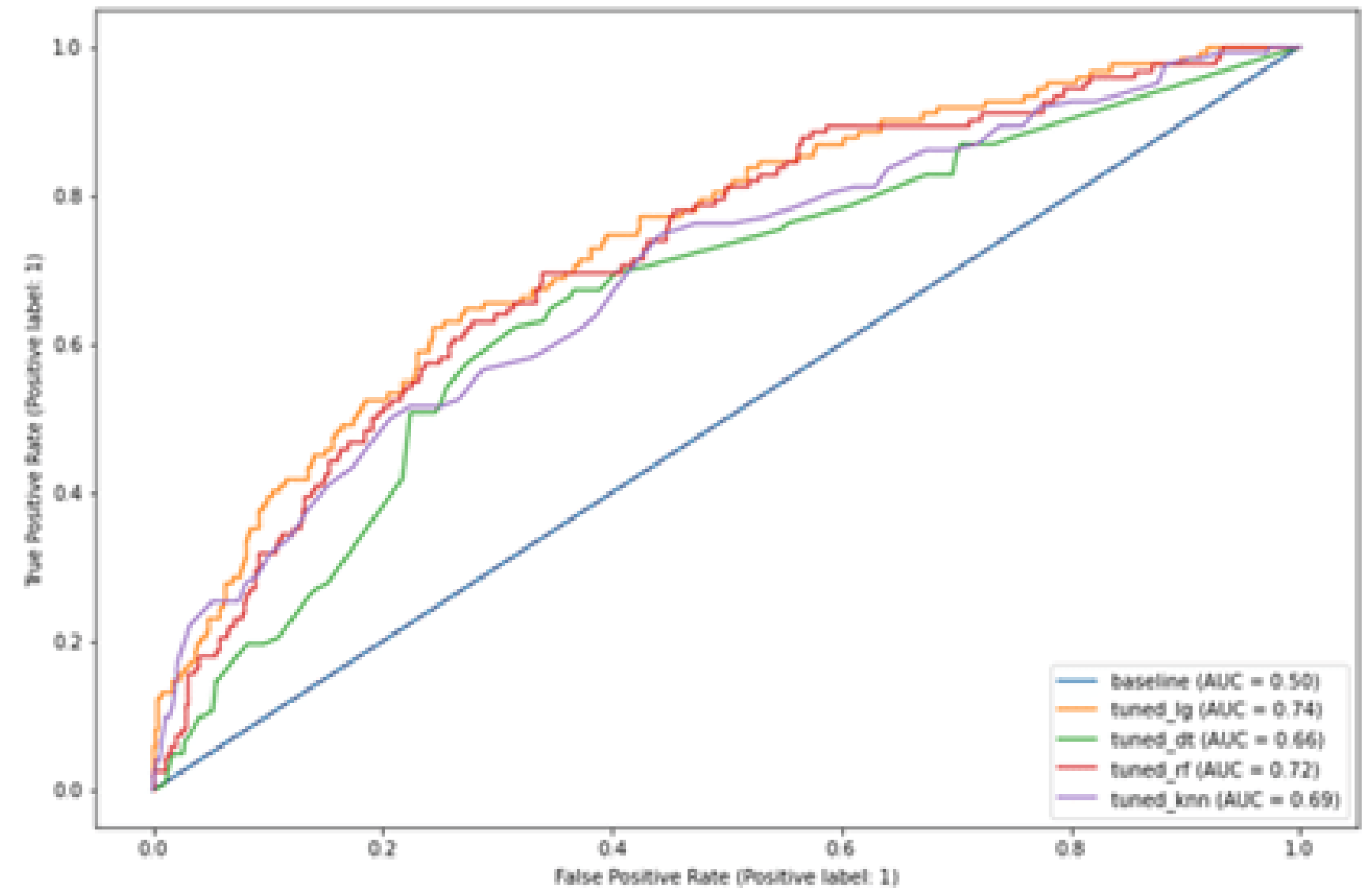
# Machine Learning Models

- **Baseline Model**
- **logistic regression**
- **Decision Tree**
- **Random Forest**

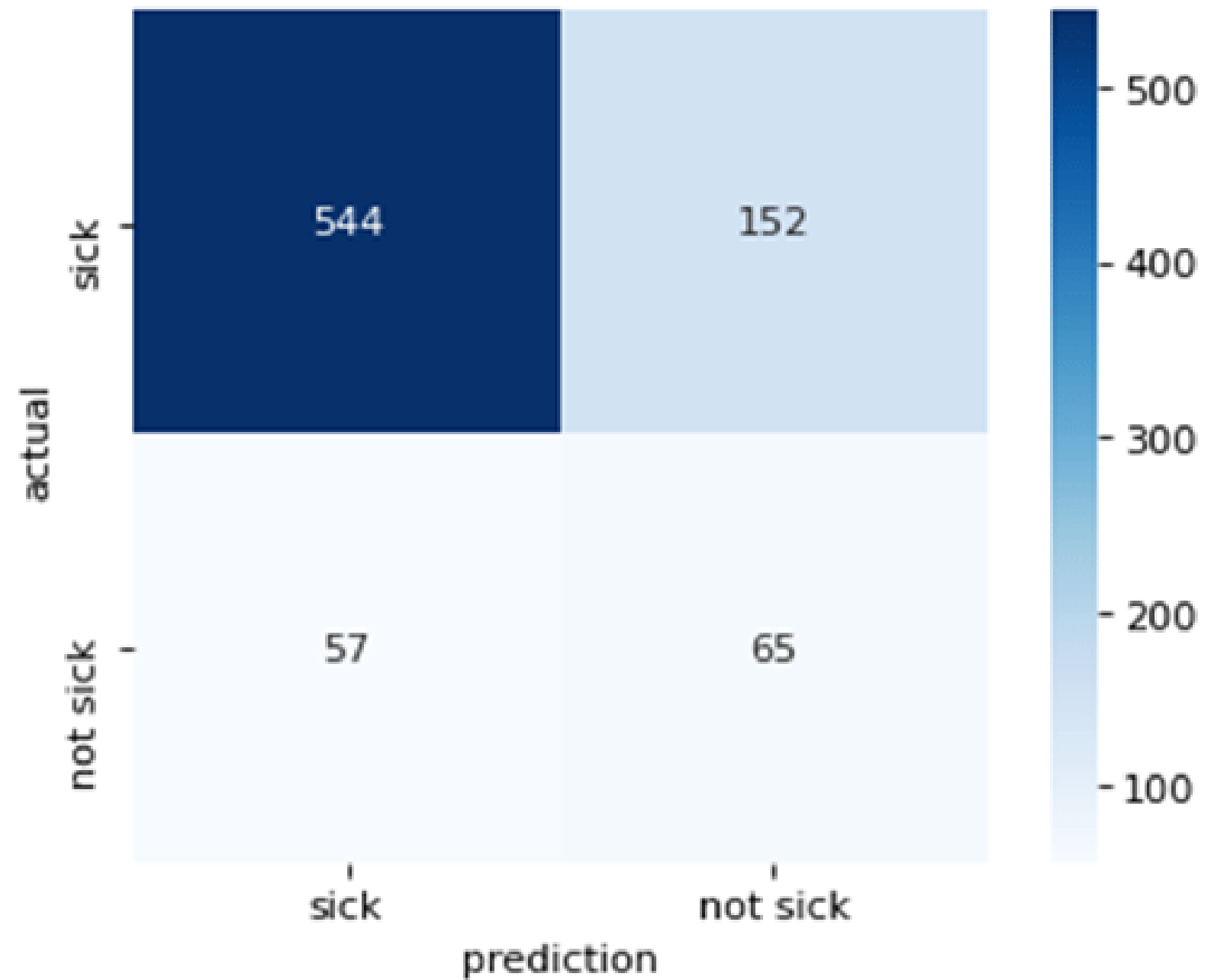
| Model                                   | Precision | Recall | <u>F(1)</u> | Accuracy |
|---|-----------|--------|-------------|----------|
| Baseline                                | 0.29      | 0.64   | 0.40        | 0.70     |
| logistic regression<br>(StandardScaler) | 0.29      | 0.64   | 0.40        | 0.71     |
| Decision Tree                           | 0.25      | 0.62   | 0.36        | 0.67     |
| Random Forest                           | 0.28      | 0.60   | 0.39        | 0.71     |



# Best model



# Confuion Matrixe



# CONCLUSION

We tried several Experiment models to achieve our goal is to predict whether the patient has 10-year risk of future coronary heart disease .

And we found the best model -  
logistic regression



—  
**The End**

**THANKS  
FOR  
LISTENING**

*Any Question?*