

# Prediction of Strokes

T5 Data Science Bootcamp

BY

Text, letter

Description automatically generated

**Munira Alshahrani**

**Rahaf Almayouf**

[**Shatha Alghamdi**](https://github.com/ShathaAlghamdi)

# Introduction

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. 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 relavant information about the patient.

# Study [Methodolog](https://github.dihe.moe/sanjeevai/nyc_subway_data_analysis#p2)y

**The dataset consists of 12 features and 5110 observations.**

# Data Description

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Feature name** | **Description** | **Data type** | | --- | --- | --- | | Id | Identification number of the individual | int | | Age | age of the patient | float | | Gender | male or female | object | | Hypertension | person age | int | | heart\_disease | Health related parameter, does person have heart disease | int | | ever\_married | Personal information, is person married on not? | object | | work\_type | Nature of work place | object | | Residence\_type | Residence type of the individual | object | | avg\_glucose\_level | average glucose level in blood for the individual | float | | Bmi | body mass index of the individual | float | | smoking\_status | Habitual information | object | | Stroke | Our target, is person suffered heart attack | int | |
|  |

# Question:

1. Does age has impact on strokes?
2. Does body mass index and glucose levels in a person, propel a heart stroke?
3. Assumption: Smoking can induce Stroke, is it true?
4. Assumption: Heart with a Heart Disease is prone to Stroke, is it true?
5. Assumption: Workload results in high blood pressure and that could lead to Stroke, is it true?
6. Assumption: Males are most susceptible to strokes due to high work related stress, is it true?

# Tools and Libraries:

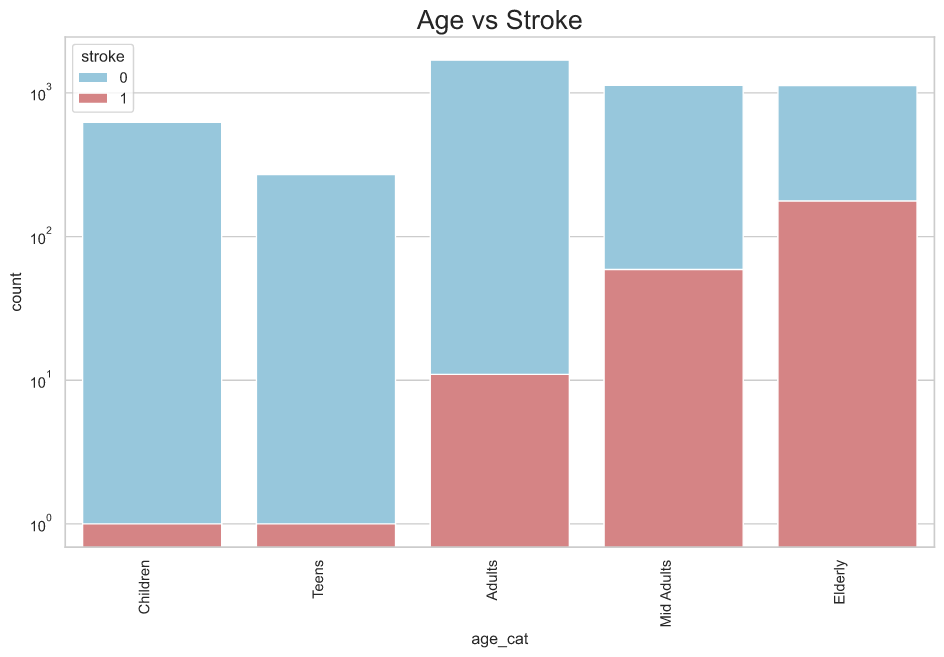
* **Libraries:** pandas, numpy, matplotlib, seaborn, plotly, sklearn.
* **Softwares:** Trello, GitHub, Jupyter, VSCode, Word & PowerPoint, Zoom.

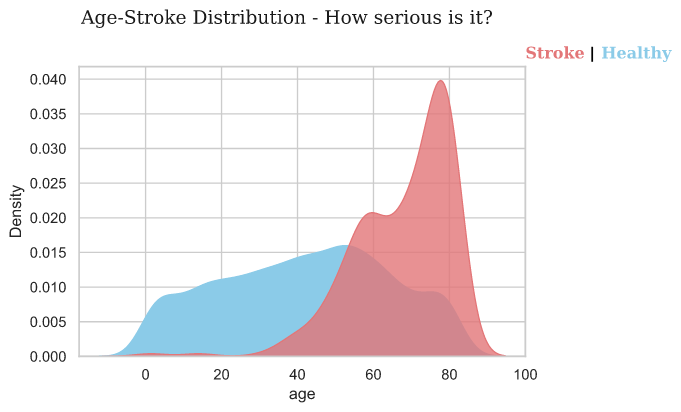
# Project Goals :

* Awaring of the most common causes of strokes.
* Concluding who are the most susceptible to strokes from people.
* Realizing the difference between smokers and non-smokers in having strokes.
* Knowing the ages most likely to have strokes.

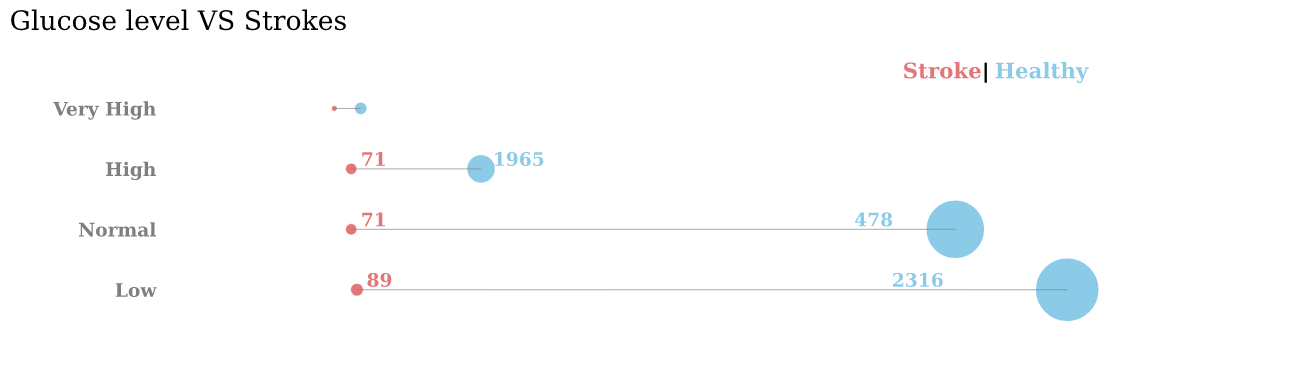
# EDA

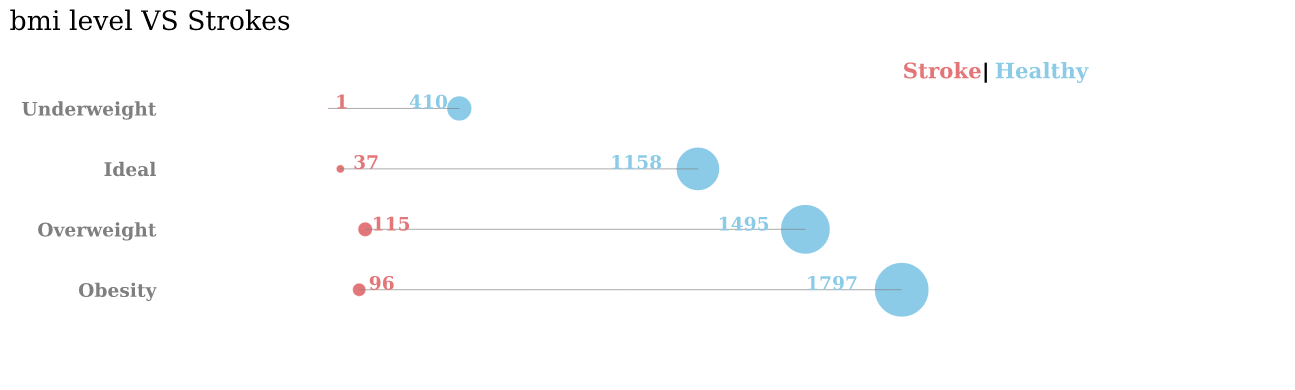
#### **Does age has impact on strokes?**



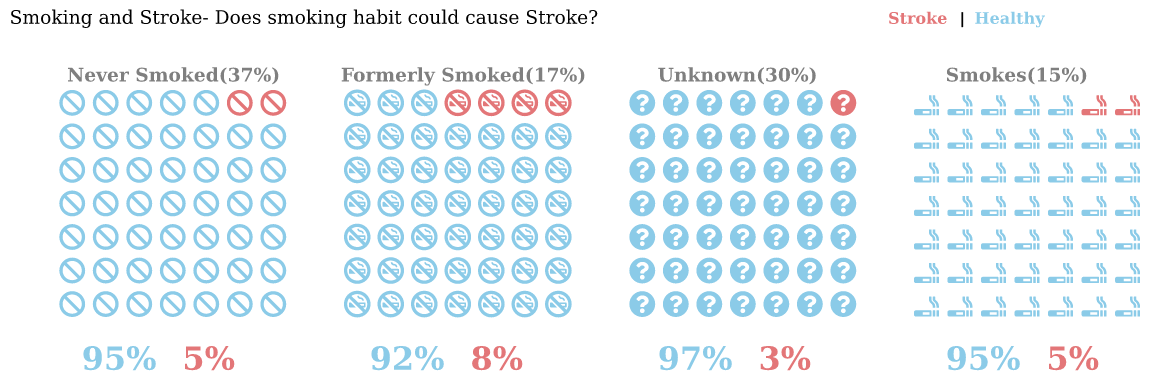


#### **2. Does body mass index and glucose levels in a person, propel a stroke?**





#### **3. Assumption: Smoking can induce Stroke, is it true?**



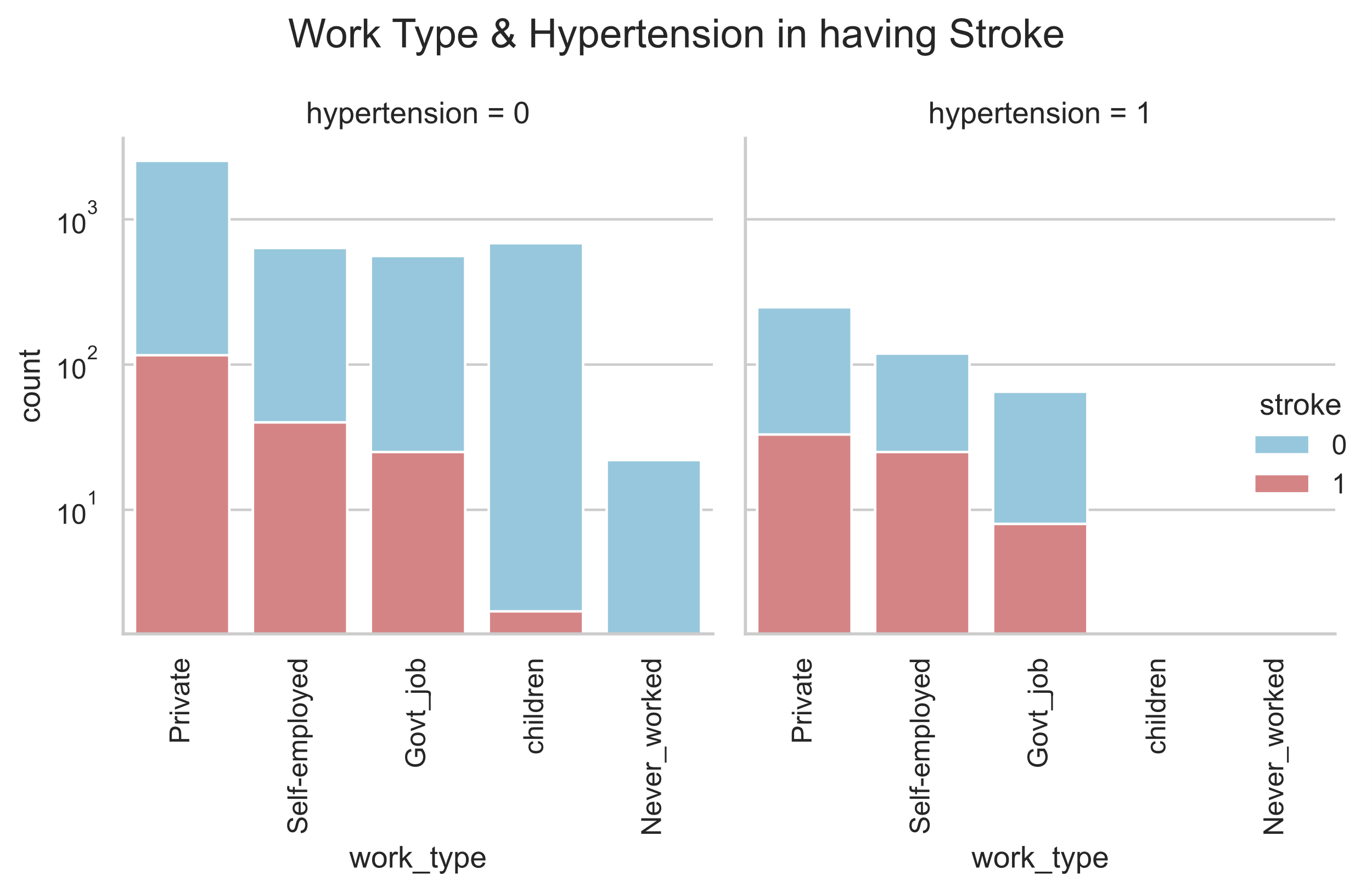
#### **4. Assumption: Heart with a Heart Disease is prone to Stroke, is it true?**



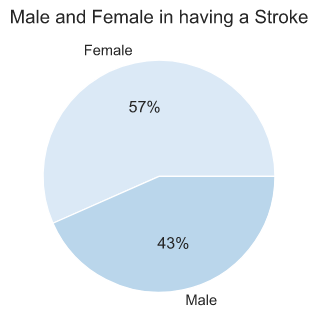
#### **5. Assumption: Workload results in high blood pressure and that could lead to Stroke, is it true?**

A picture containing calendar

Description automatically generated



#### **6. Assumption: Males are most susceptible to strokes due to high work related stress, is it true?**



# Algorithm

# SMOTE

# RF

# 

# DT

# 

# LR

# 

# 

# SVM

# 

# KNN

# 

# NB

# 

# LGBM

# 

# XGB

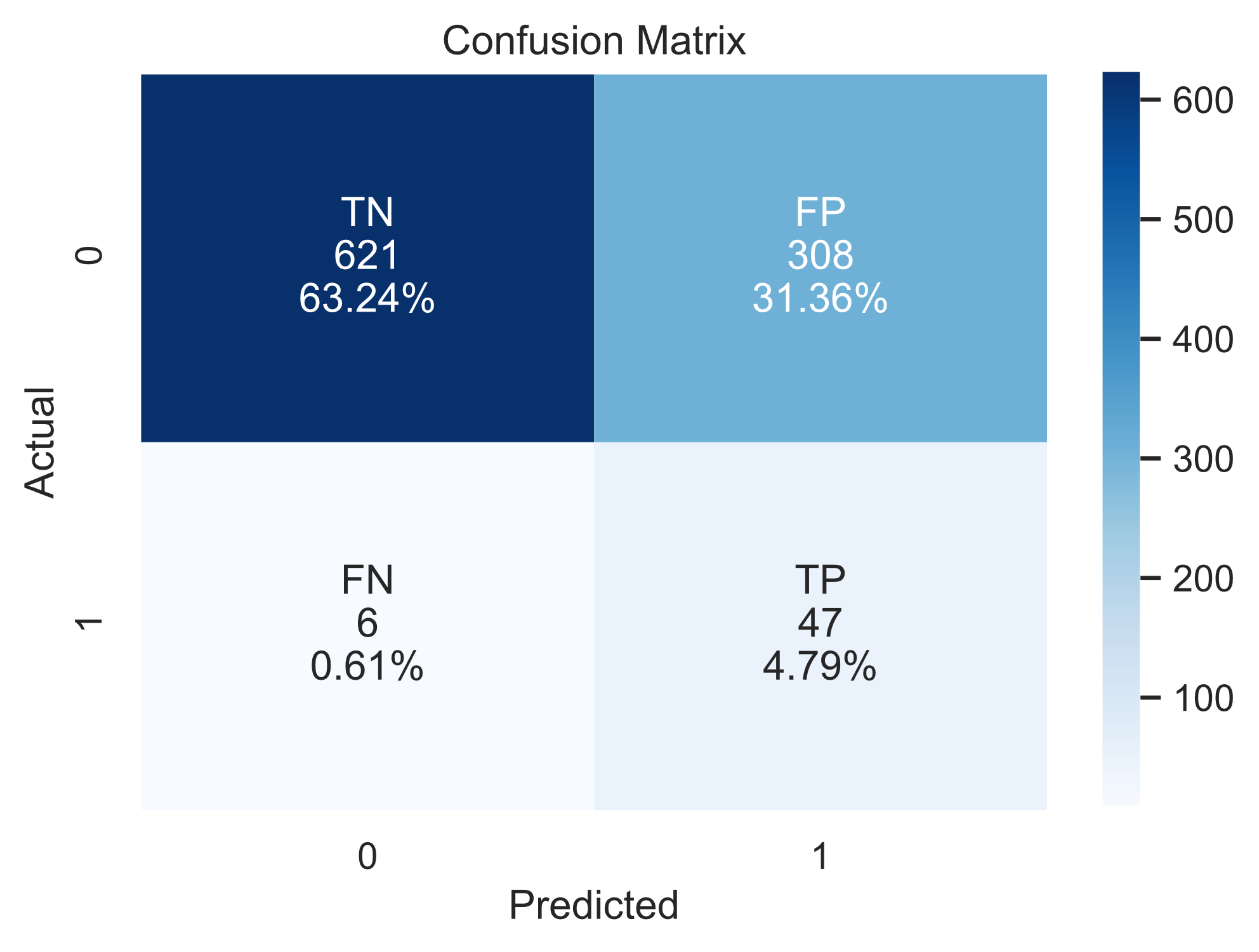
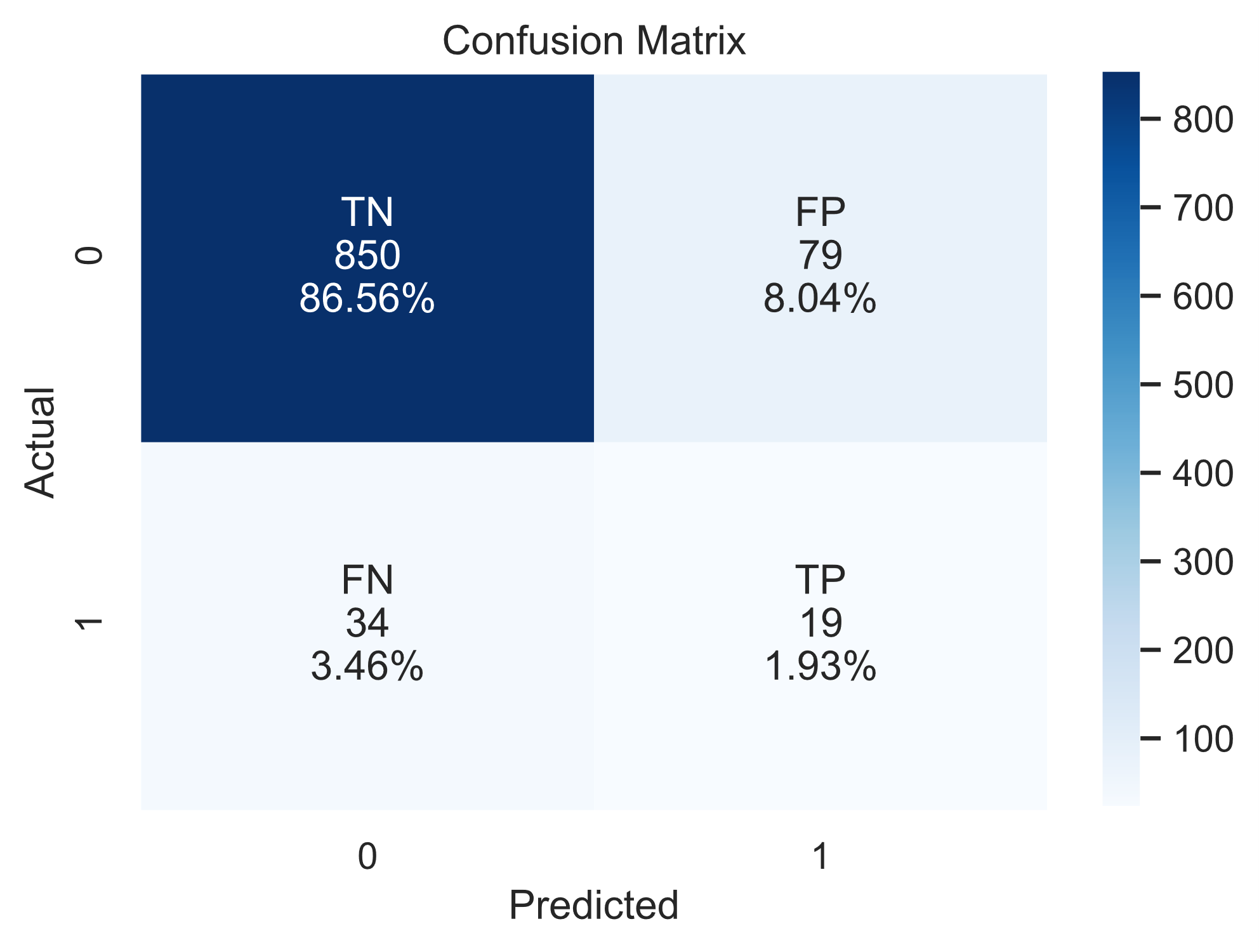
# 

# Gradient Boosting

# 

# SMOTE + Tomek

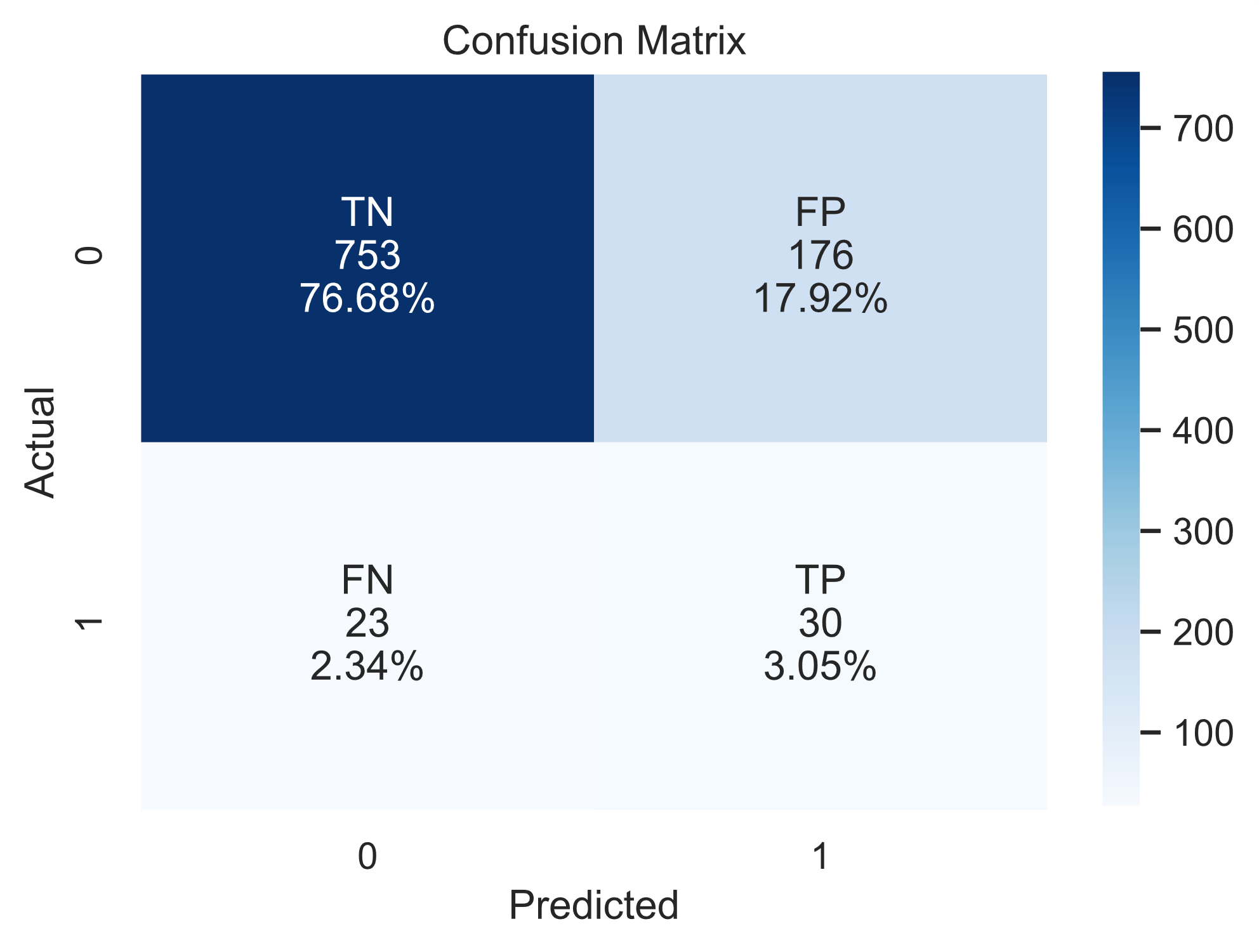
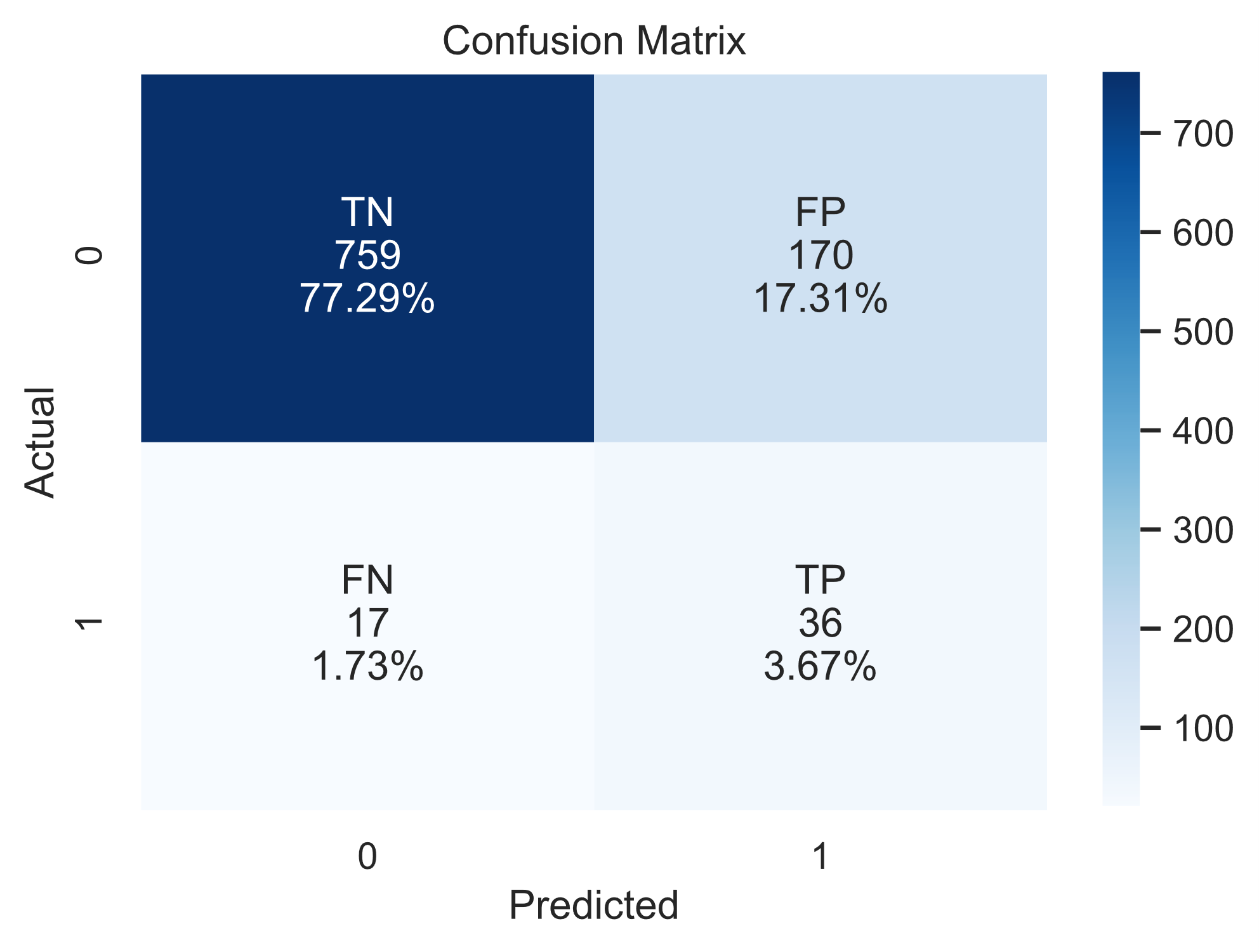
# 



Logistic

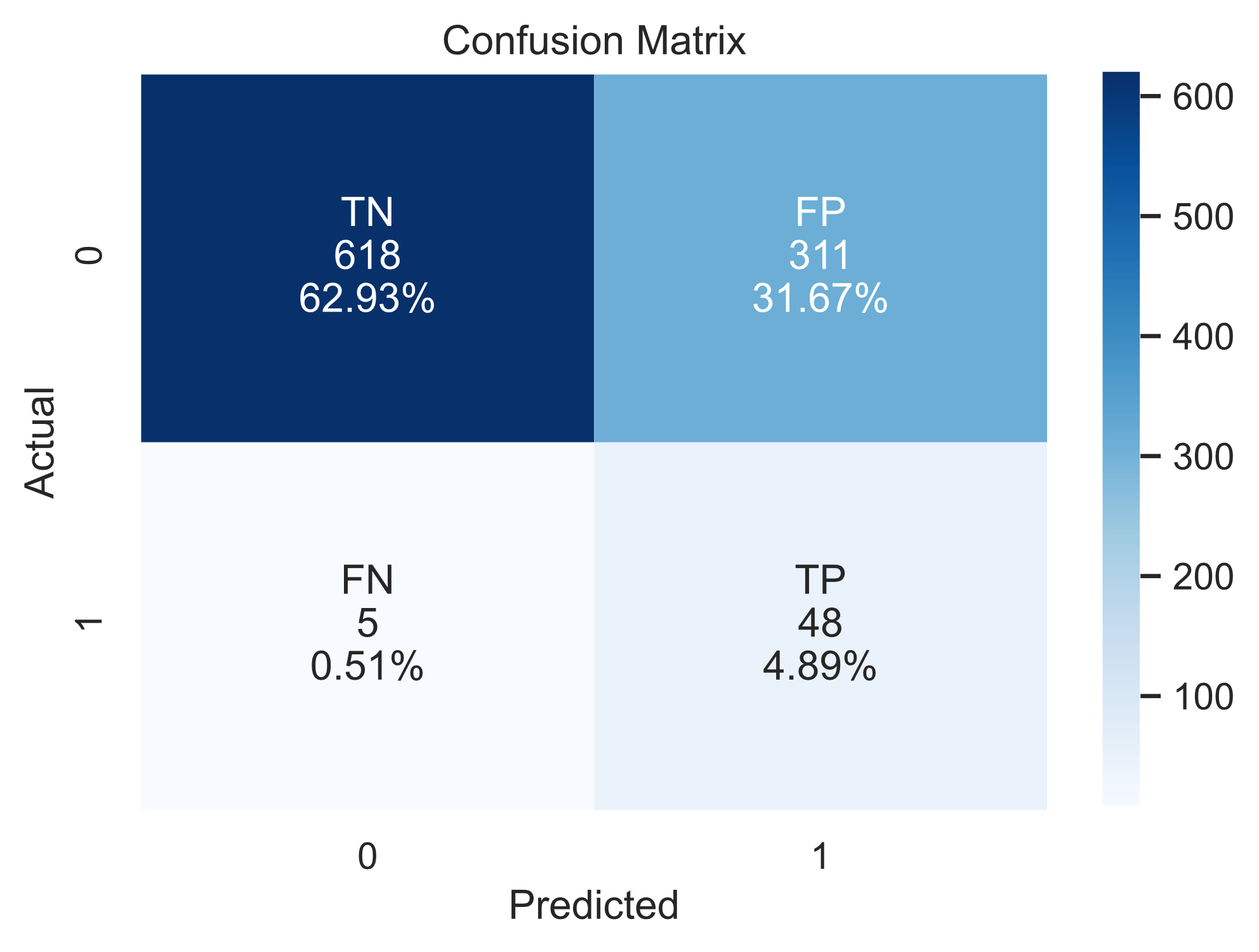
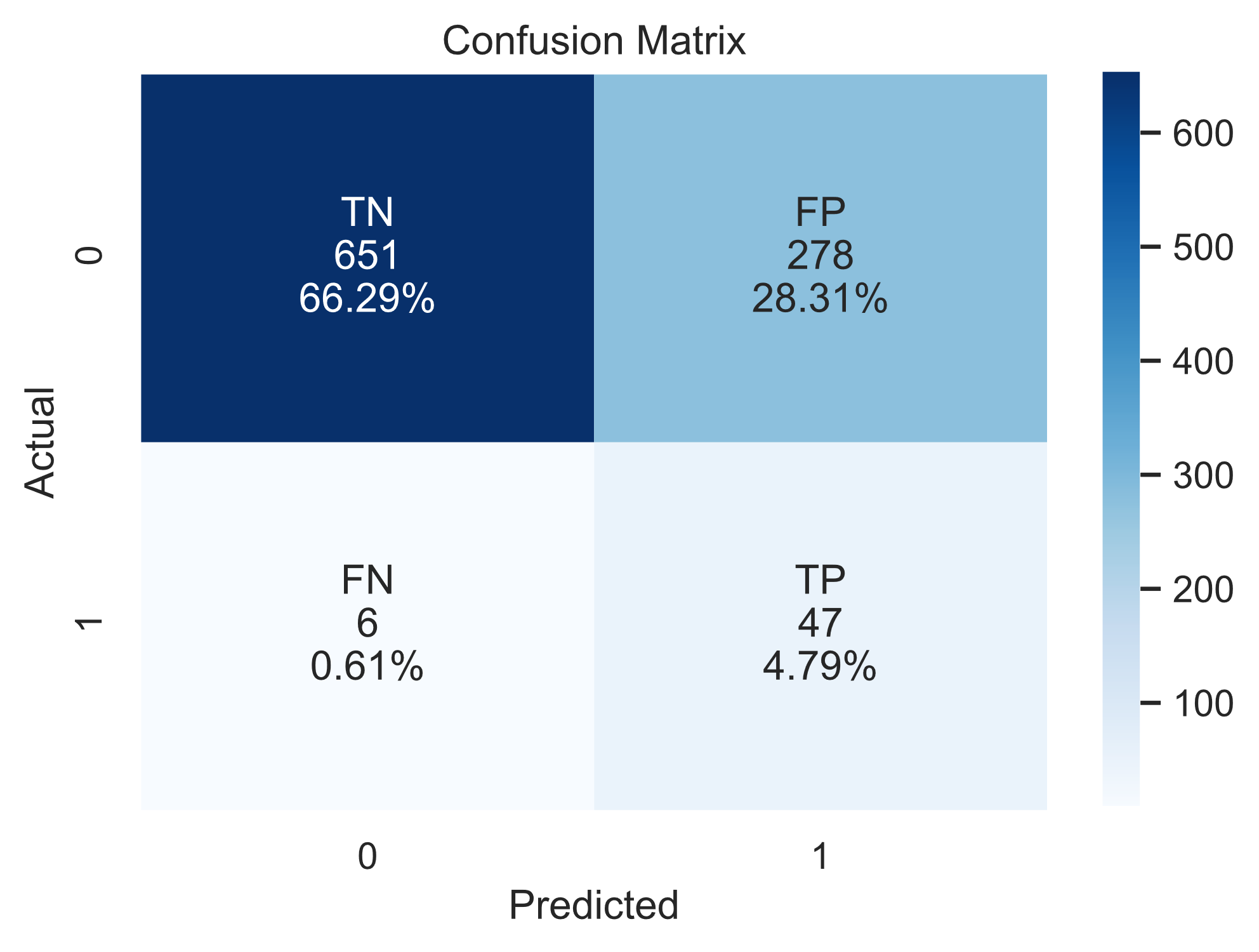
Regression

KNN



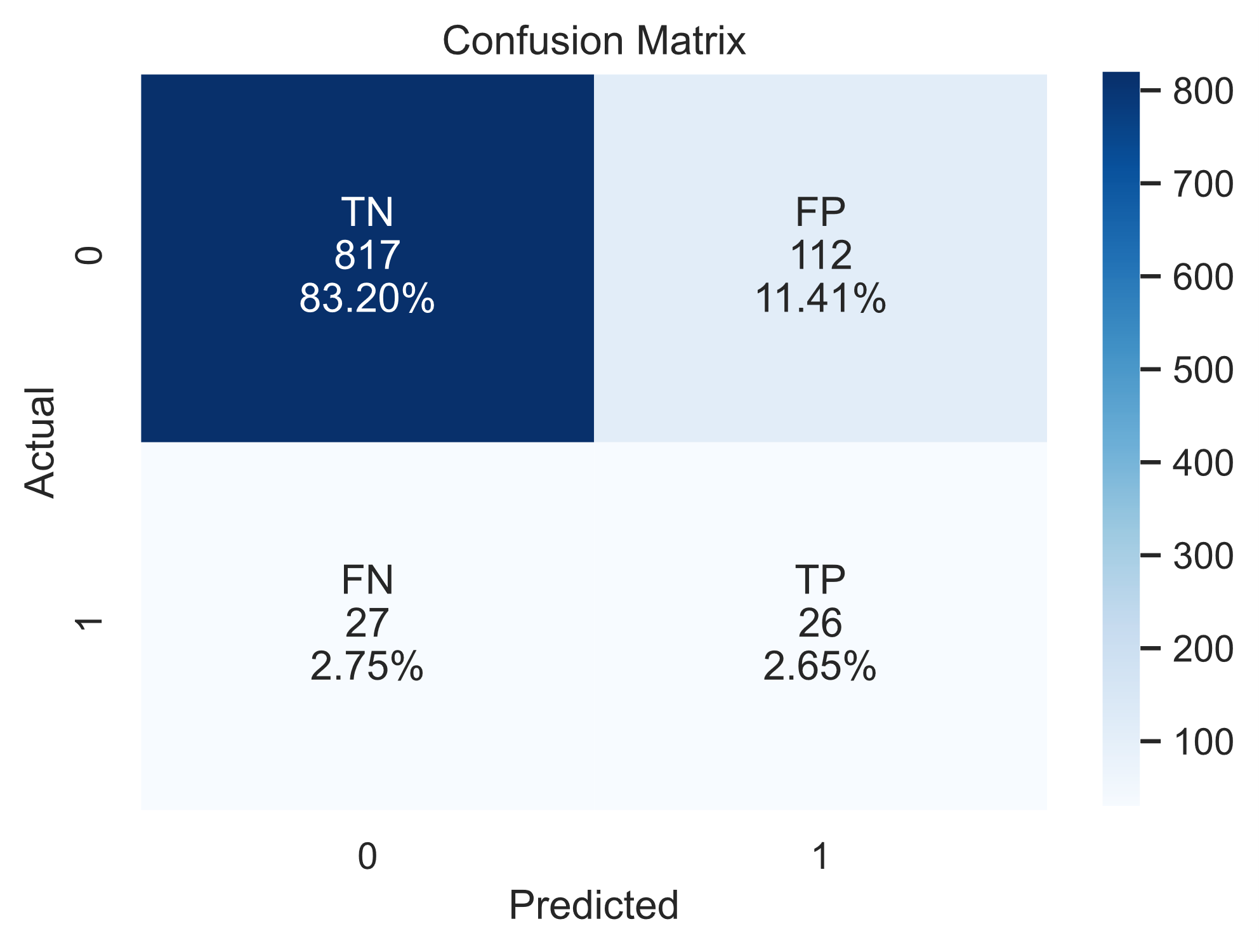
Random Forest

Decision Tree



Naive Bayes

SVM



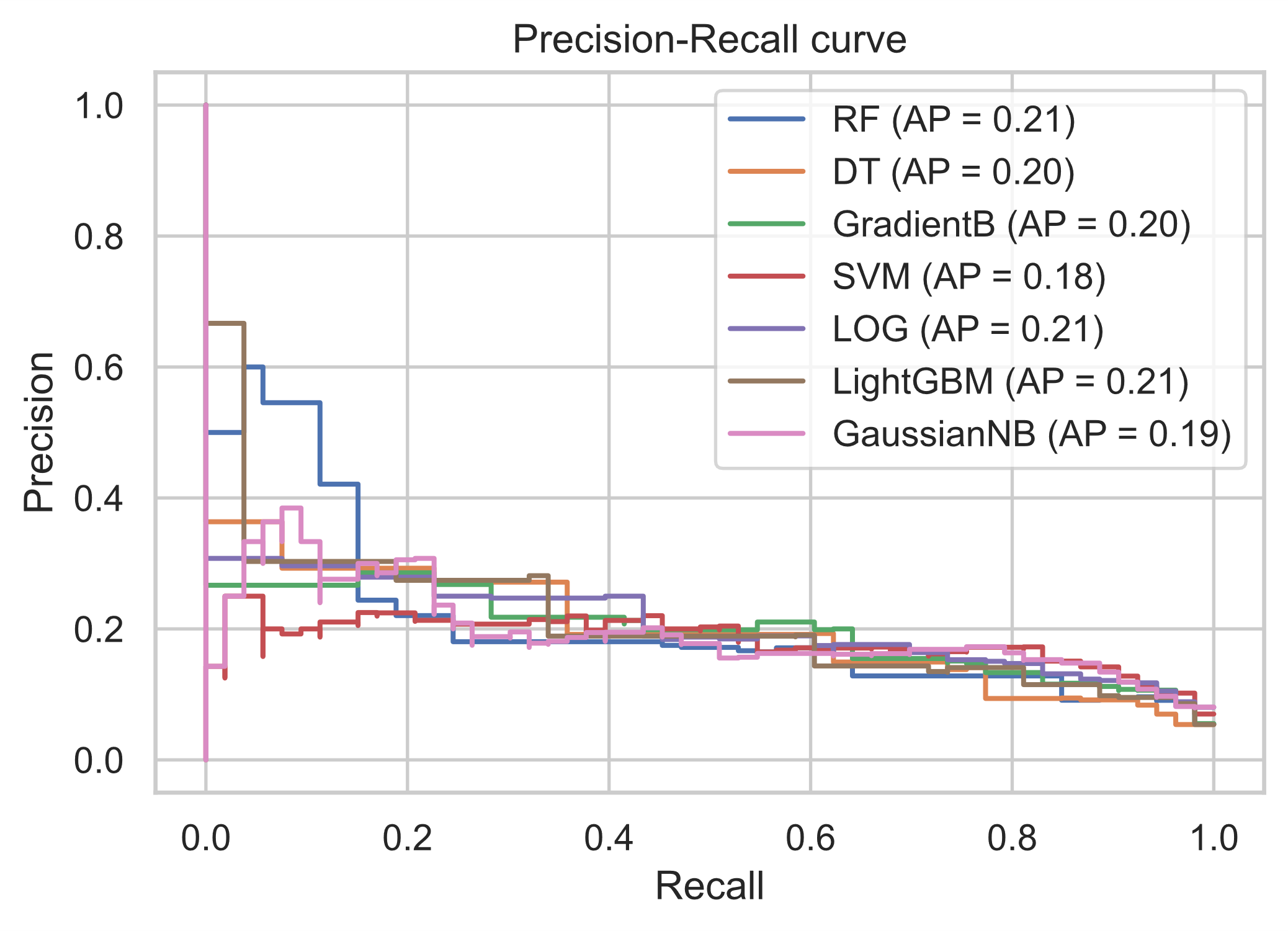
XGBoost

Gradient Boosting

# 

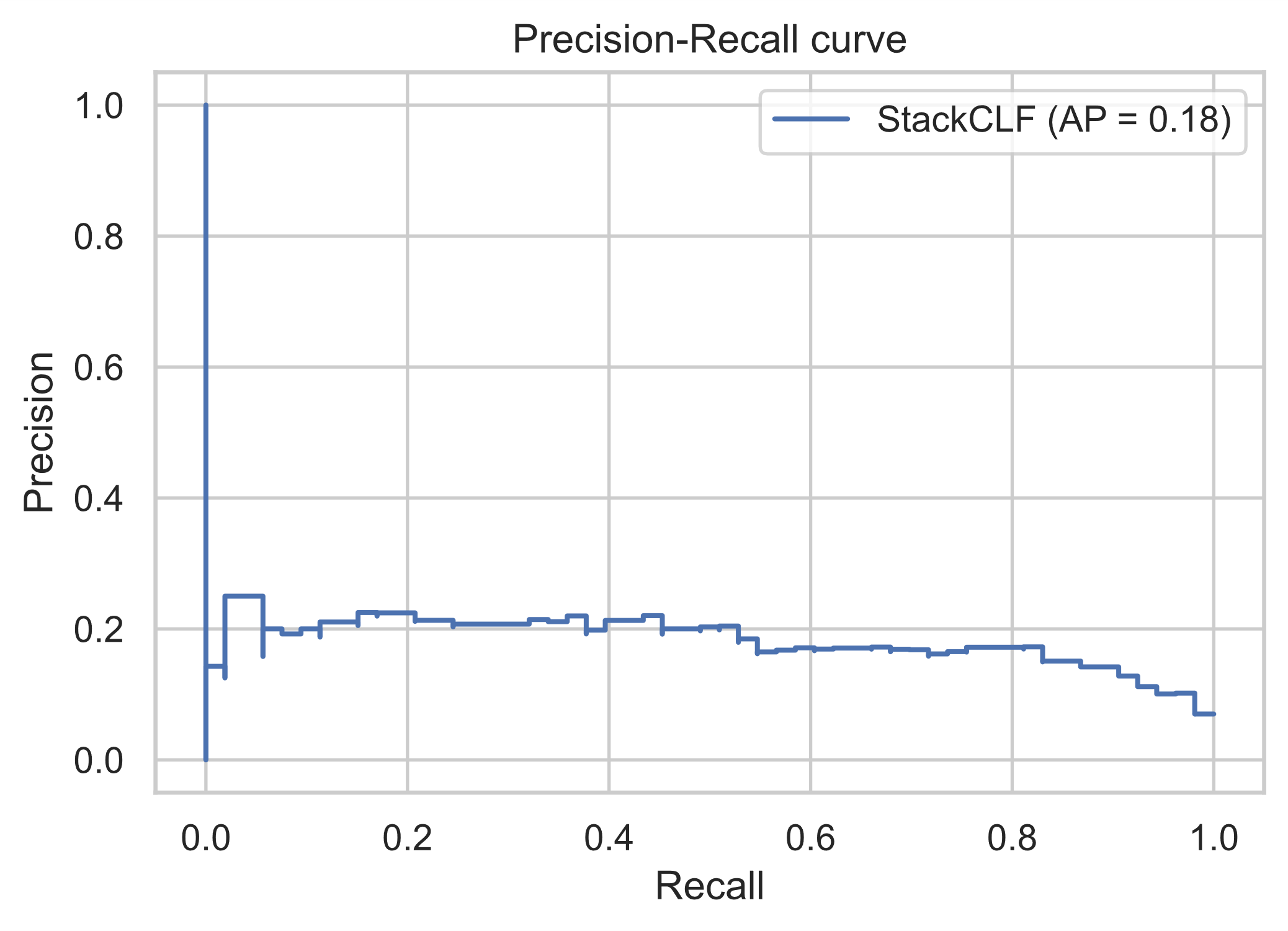
LightGBM

# Stacking



PRC for Models

Stacking Model



PRC for Stacking Model

# Conclusion

# Do an awareness program that explains all the causes for being exposed to get a stroke, teach people how to do the first medical help until the Ambulance.

# Build a platform to predict if any patient is exposed to get a stroke at any time, to call the patient and take the first step to prevent any bad situation.

# Reference

<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>