

# GROUP-E PROJECT

## PRESENTATION

---

Presented By: Nwe Ni, Phoe La Pyae, Sai Sai, Phone Pyae

# 01

## Physical Activity & Mental Health

# INTRODUCTION

Lack of physical activity and sleep deprivation can lead to mental issues like stress and depression.

These mental issues disrupt daily routines by affecting energy levels, leading to difficulty with basic tasks.

# DATASET DESCRIPTION

## VISION AND MISSION STATEMENT

---

The main goal of our project is to encourage the user to take some activities and rest by warning for daily happiness.

02

01

### Heart\_Rate\_BPM

The user's heart rate in beats per minute (BPM)

02

### Sleep\_Duration\_Hours

The number of hours the user slept per night.

03

### Physical\_Activity\_Steps

The total number of steps the user took in a day.

03

### Mood\_Rating

A self-reported mood rating on a scale of 1 to 10.

03

### Mental\_Health\_Condition

A binary target variable where 1 represents a high risk for mental health issues, and 0 indicates no risk.

# DATA PREPROCESSING

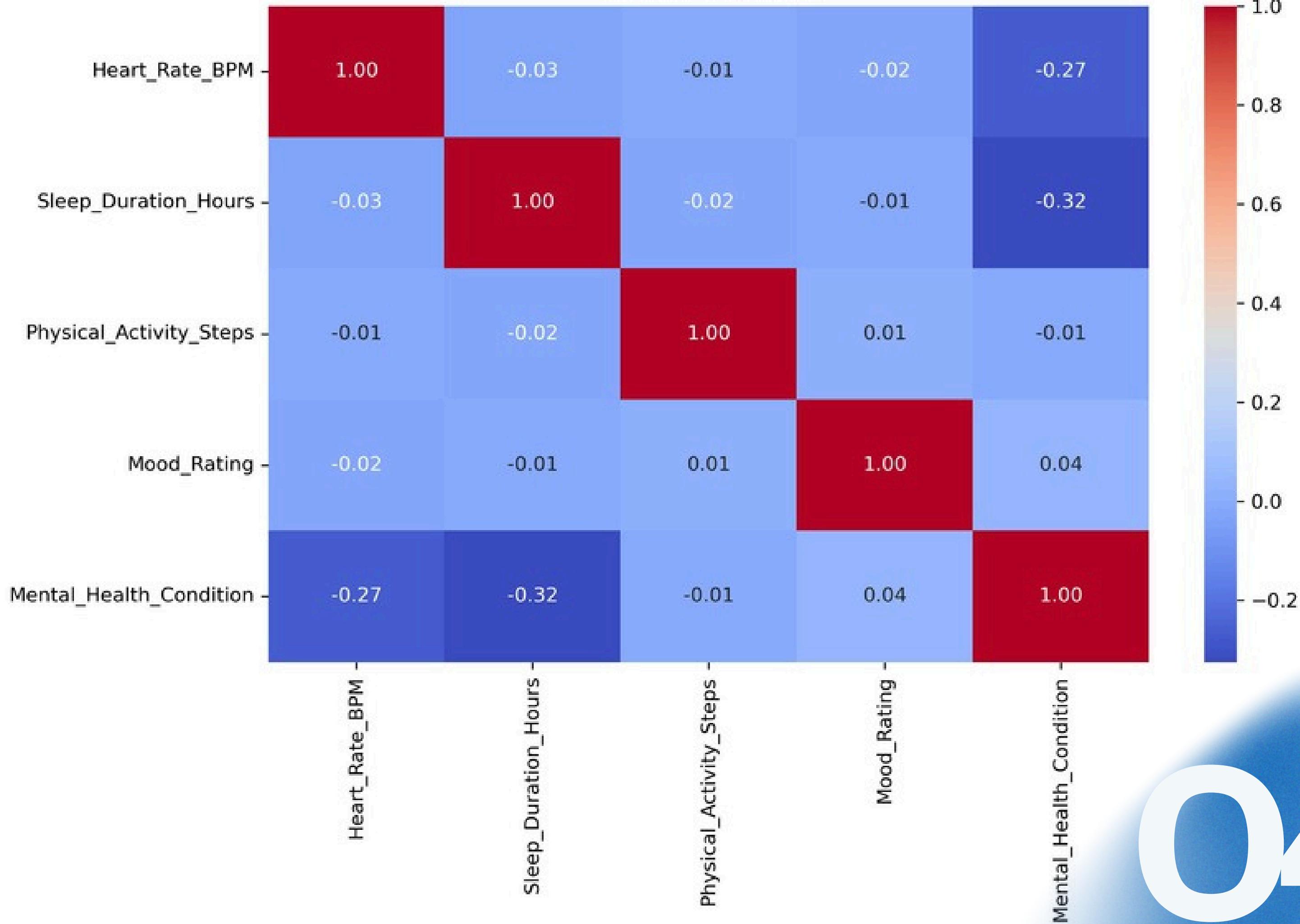
## SPLITTING & ANALYSING DATA

- SPLITTING DATA**  
Split data set into two file (60% train data & 40% test data)
- CORRELATIONS**  
Analysing the correlation between features.
- DISTRIBUTIONS**  
Statistical analysis for data distribution.
- DIMENTIONS**  
Data distribution among features.

03

# CORRELATIONS

Correlation Matrix



04

## Heart\_Rate\_BPM

Mean: 89.76

Median: 90.00

Mode: 80.00

Kurtosis: -1.20

## Sleep\_Duration\_Hours

Mean: 6.98

Median: 6.96

Mode: 4.00

Kurtosis: -1.20

## Physical\_Activity\_Steps

Mean: 8054.78

Median: 8121.00

Mode: 6566.00

Kurtosis: -1.21

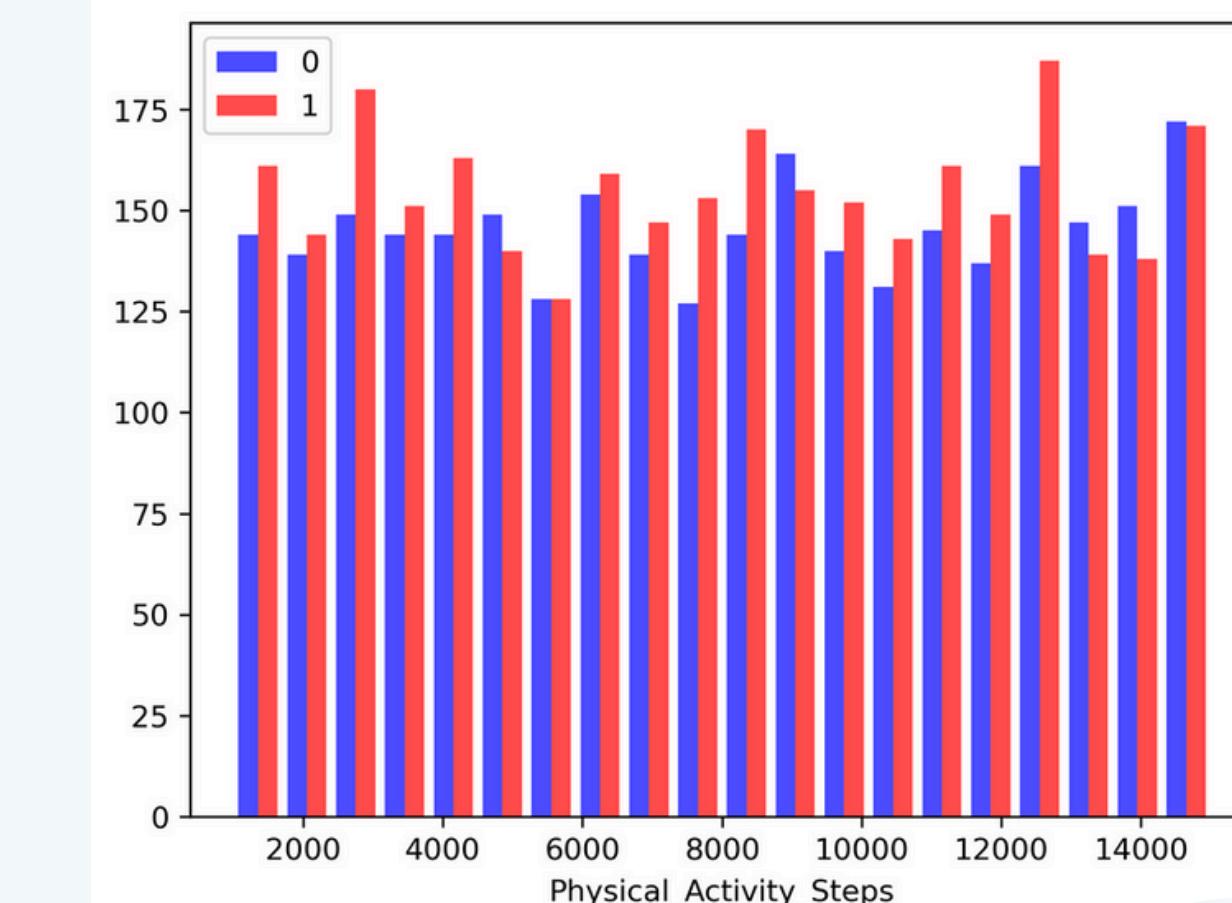
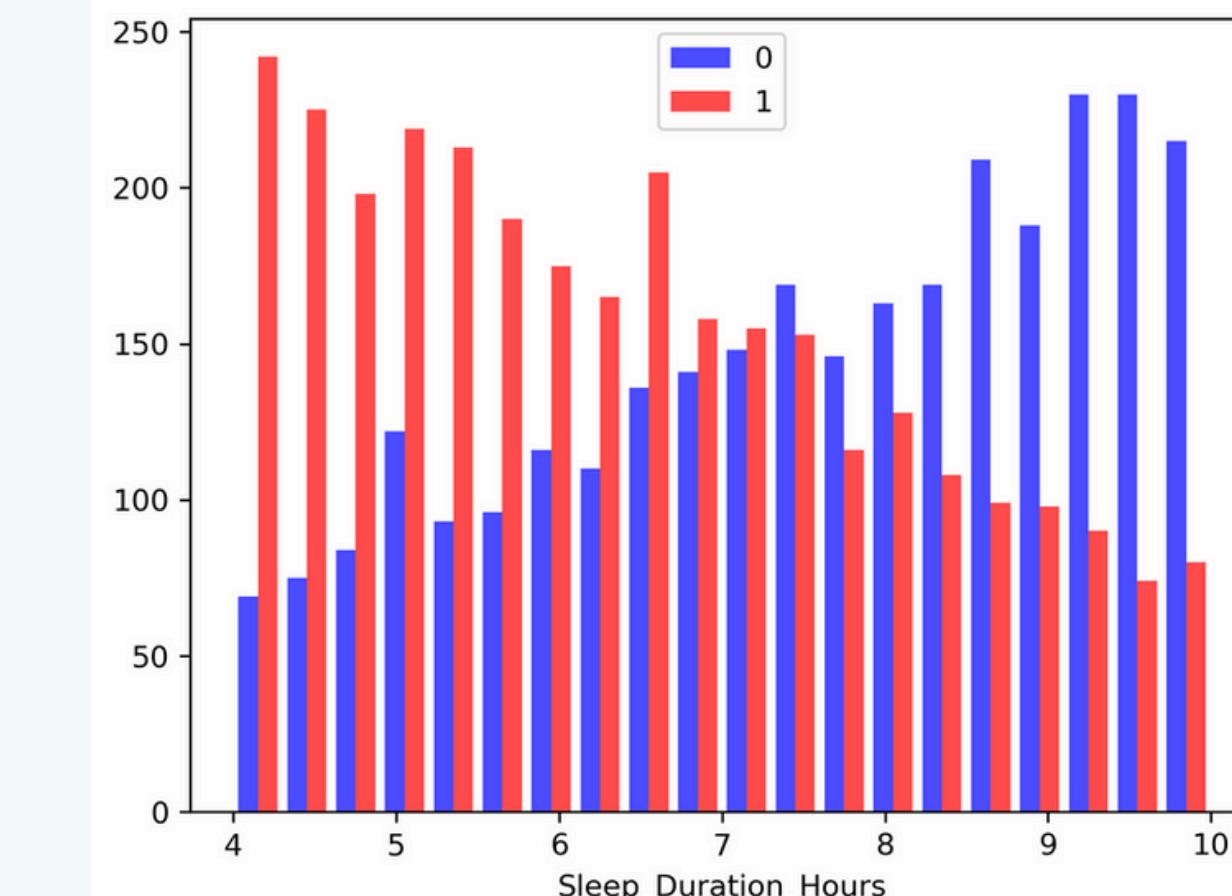
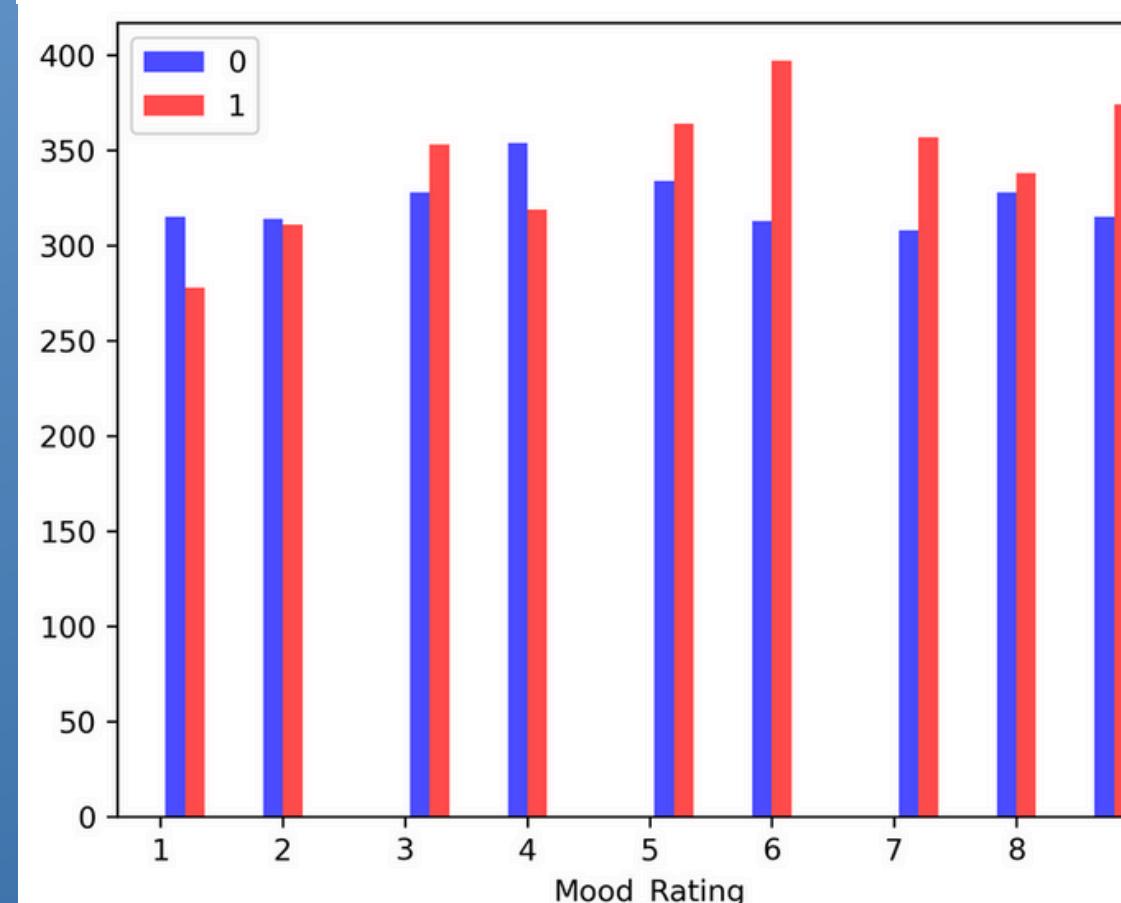
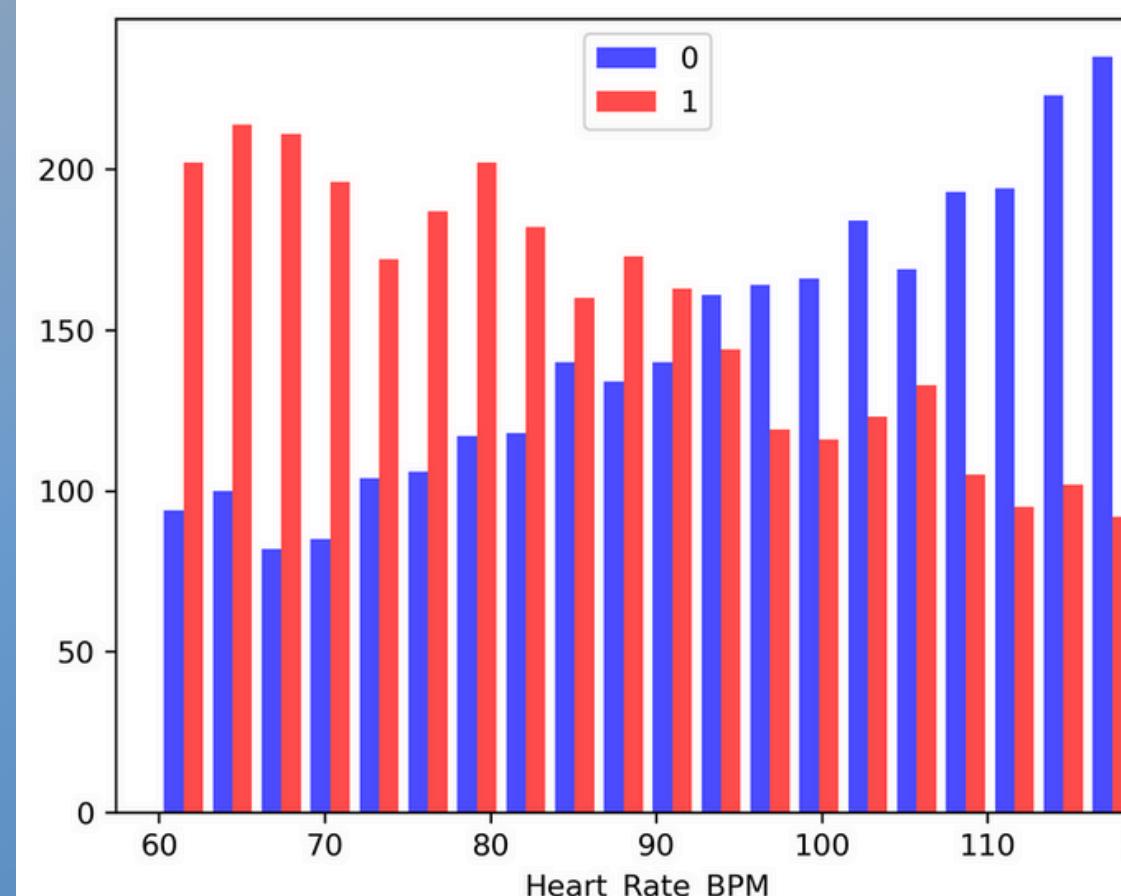
## Mood\_Rating

Mean: 5.09

Median: 5.00

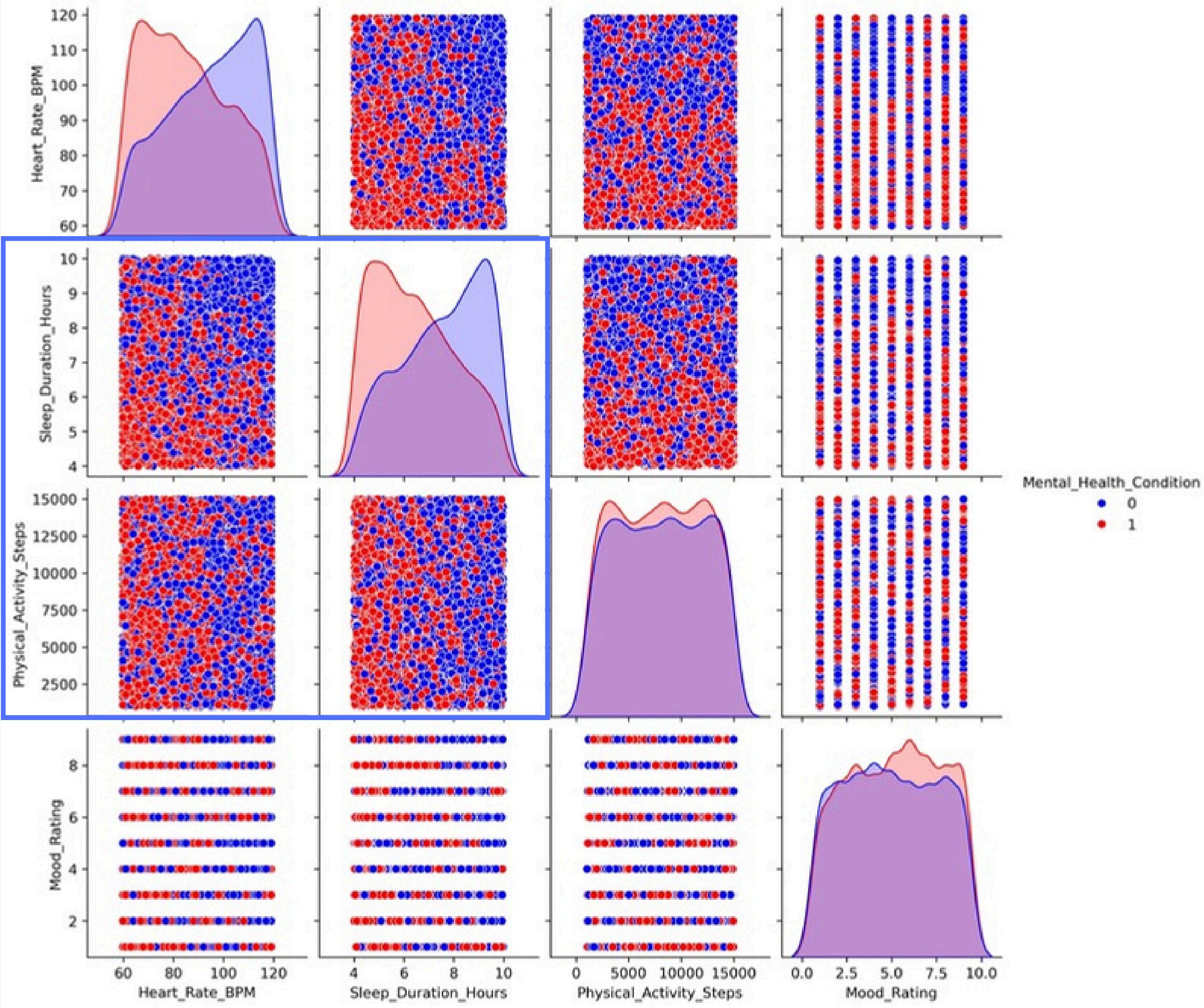
Mode: 6.00

Kurtosis: -1.19



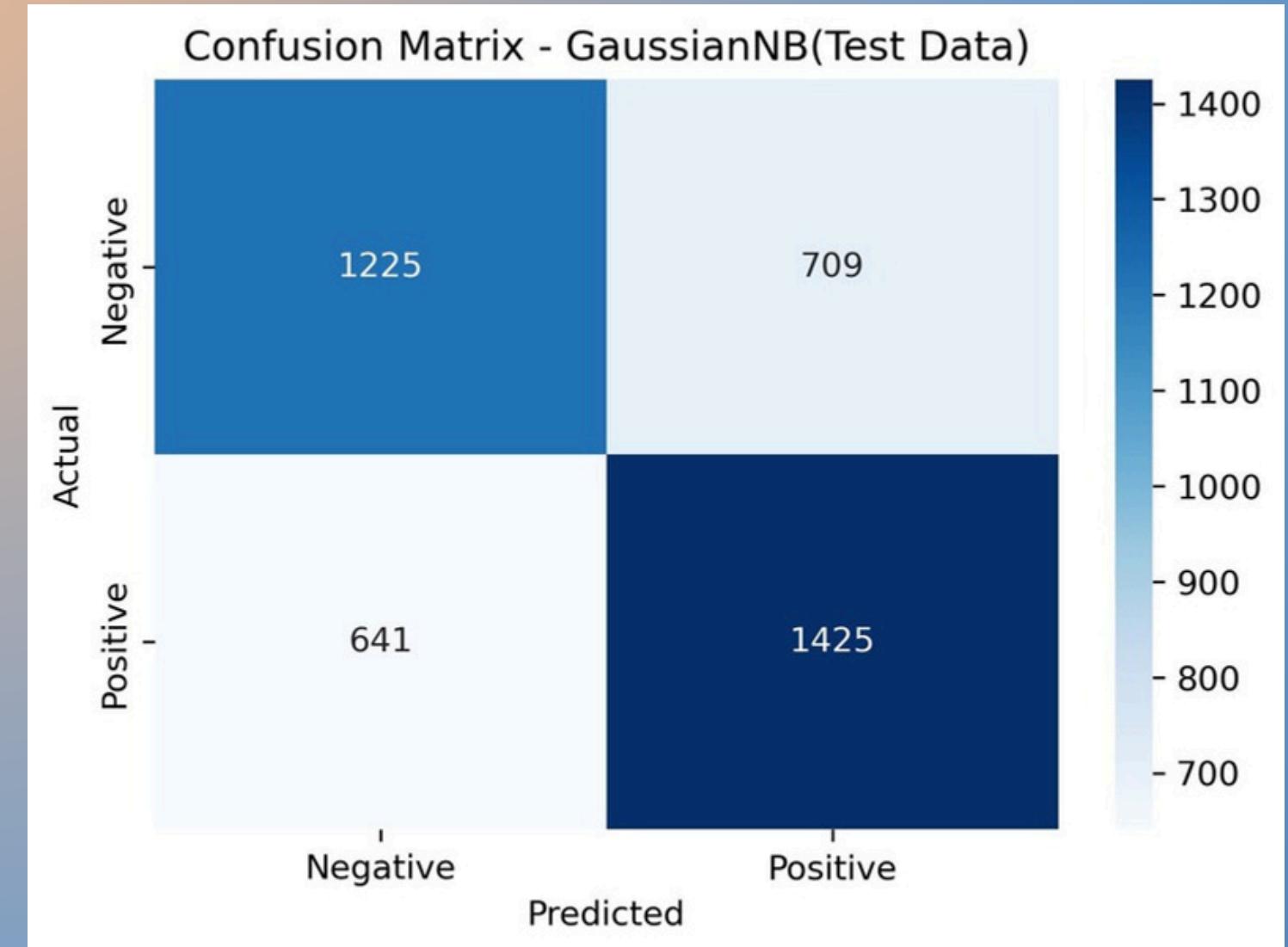
# 06

# TWO DIMENSIONAL



# Naive bayes

- Choosing model type according to data.
- Selecting dimension and features.
- Using StandardScaler Vs None.
- Final evaluation by using confusion matrix & accuracy score..



	precision	recall	f1-score	support
0	0.66	0.63	0.64	1934
1	0.67	0.69	0.68	2066
accuracy				0.66 4000

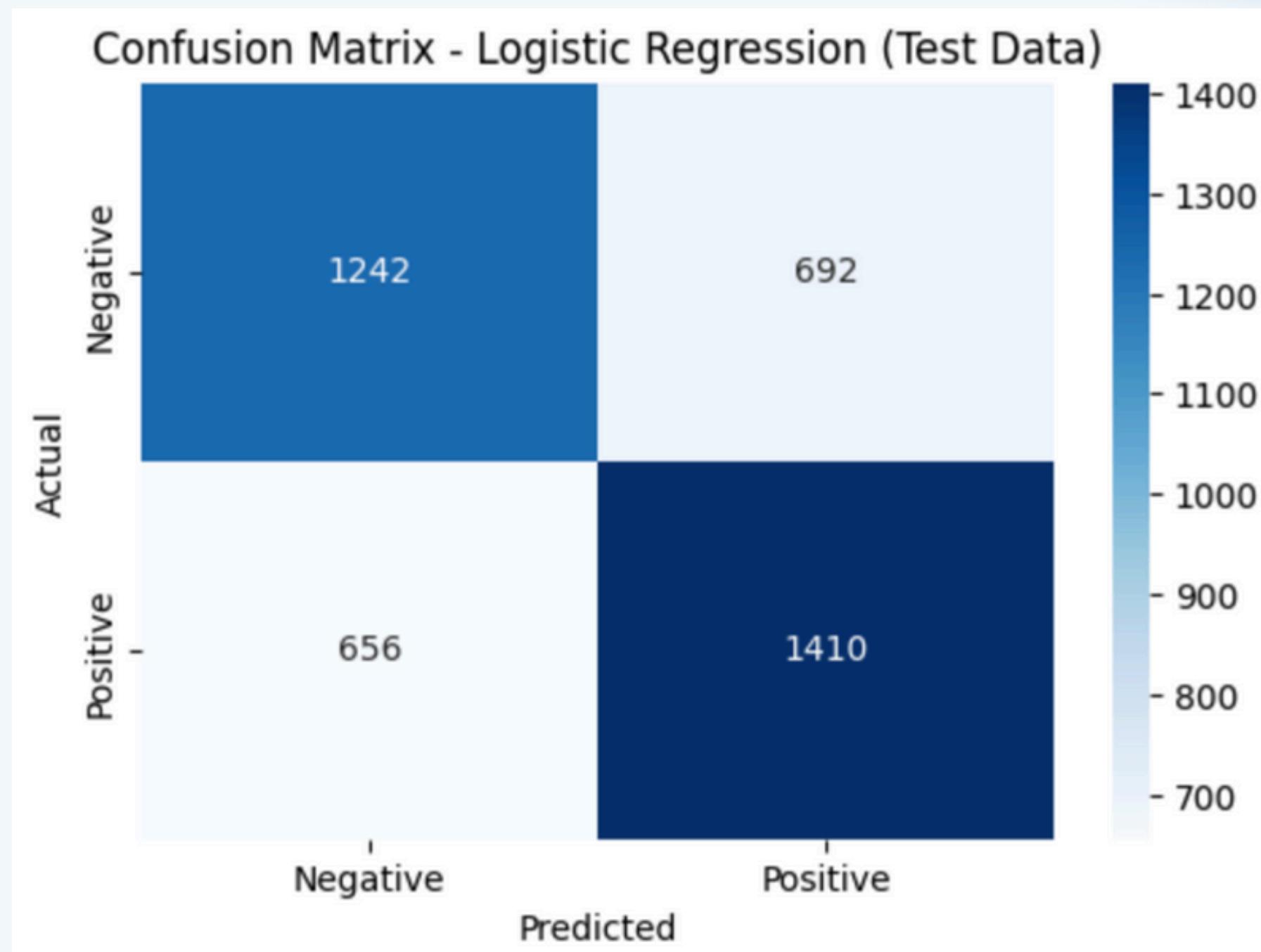
07

# 08

- 01** Split data into train data and test data
- 02** Standardize the train data using StandardScaler
- 03** Fit the train data into Logistic Regression model
- 04** Evaluate model using confusion matrix, accuracy score.

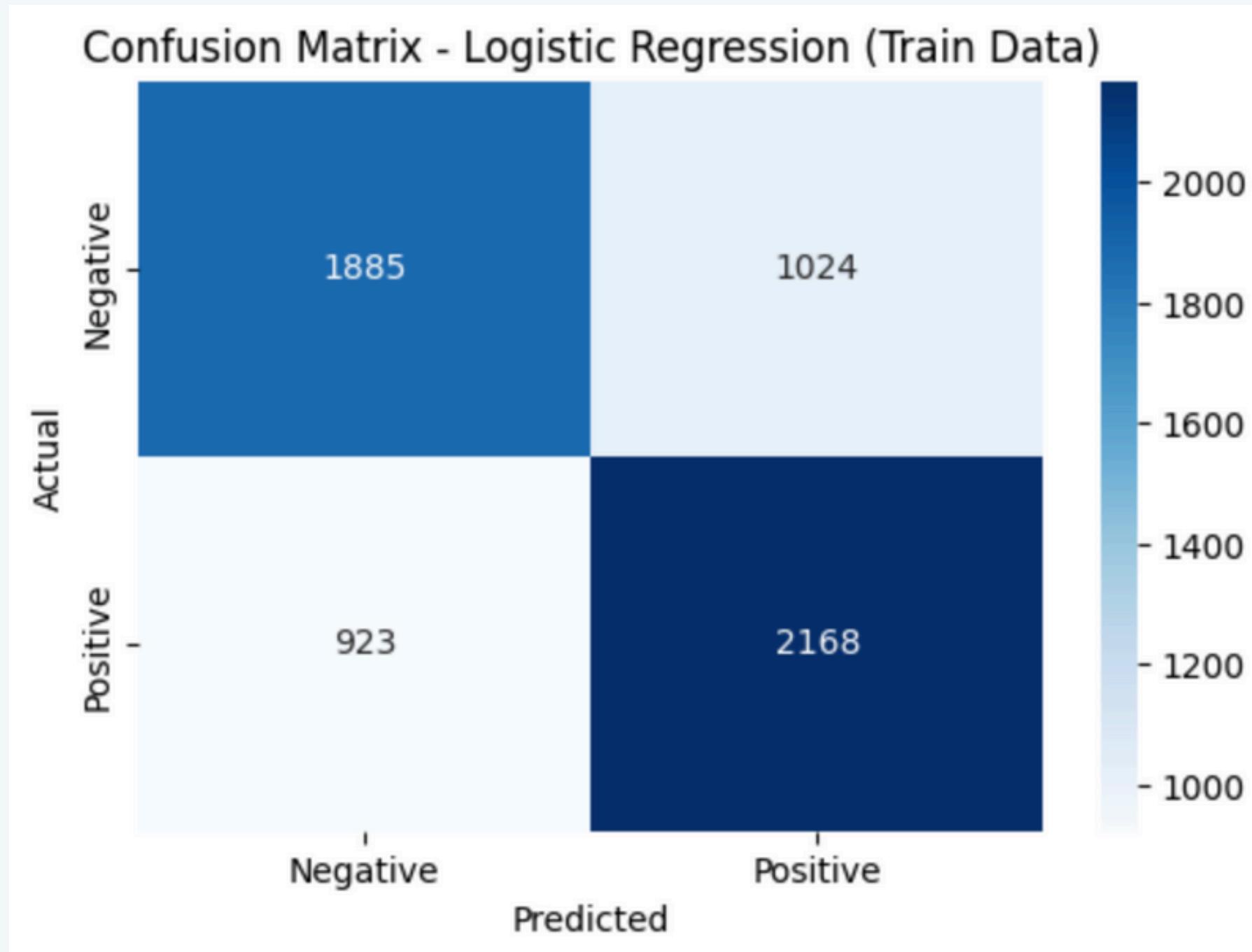
## MODEL TESTING – LOGISTIC REGRESSION

# CONFUSION MATRIX\_TEST DATA



09

# CONFUSION MATRIX\_TRAIN DATA

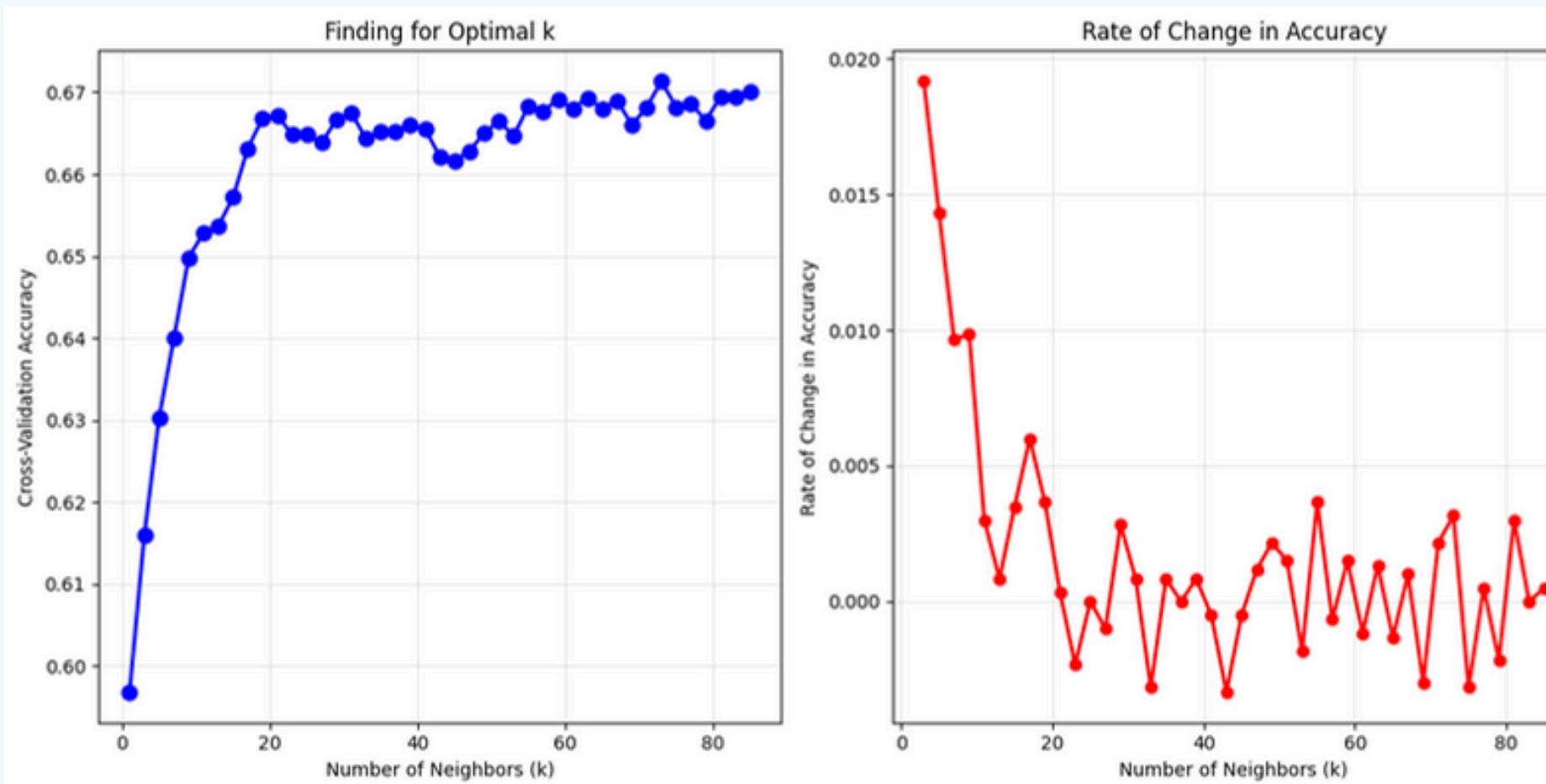


no

# IMPLEMENTATION

## USING KNN CLASSIFIER

- Firstly standardize the features using StandardScaler
- Find the optimal value of 'k=73' based on odd values of 'k' from 1 to 85 (where increasing k no longer improves cv=5 accuracy score is chosen as the best compromise)
- Train the model with 'k=73'



# EVALUATIONS AND RESULTS

## USING KNN CLASSIFIER

### ☰ Training Set Evaluation ☰

Confusion Matrix:

`[[1909 1000]`

`[ 914 2177]]`

Accuracy : 0.6810

Precision: 0.6852

Recall : 0.7043

F1 Score : 0.6946

ROC AUC : 0.7538

### ☰ Testing Set Evaluation ☰

Confusion Matrix:

`[[1210 724]`

`[ 625 1441]]`

Accuracy : 0.6627

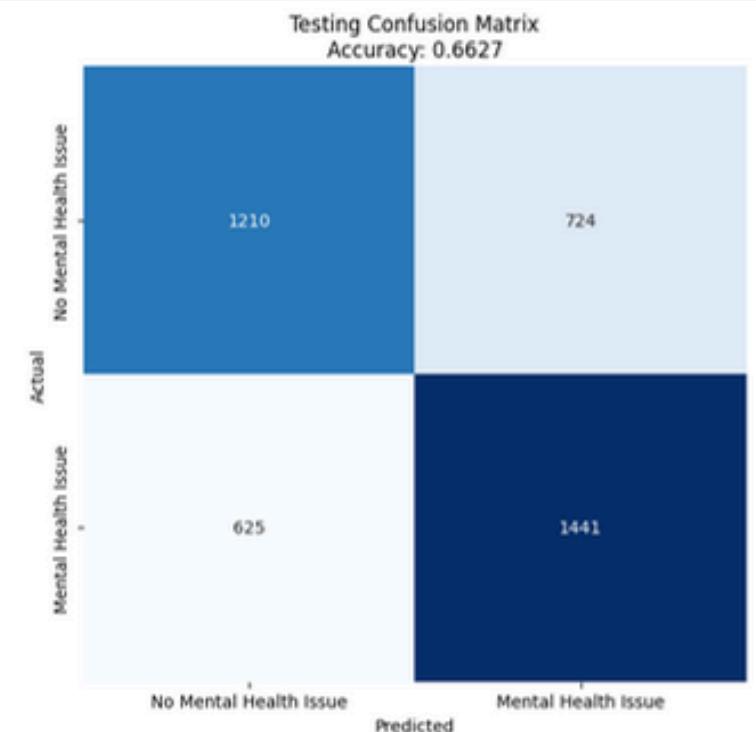
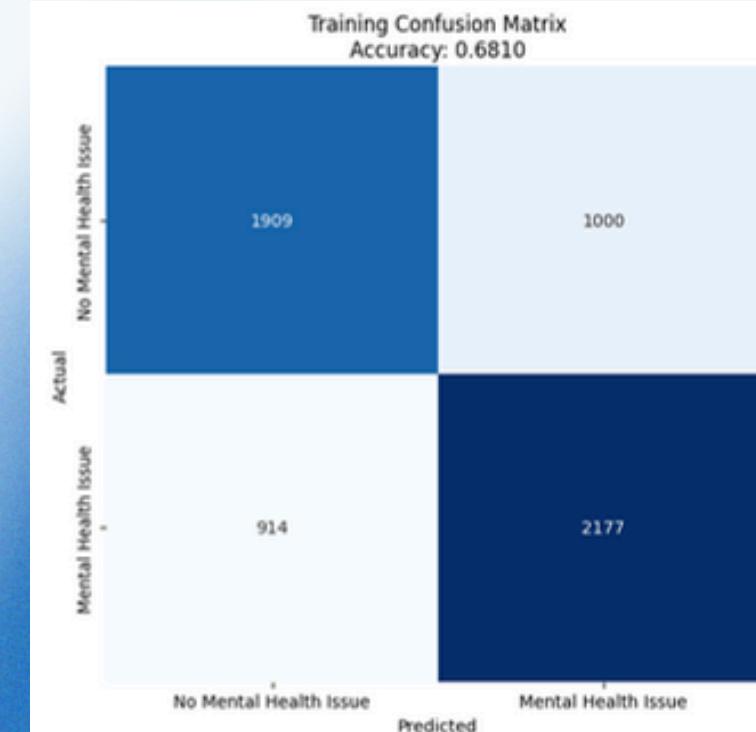
Precision: 0.6656

Recall : 0.6975

F1 Score : 0.6812

ROC AUC : 0.7249

- The model missed 625 individuals who were at high risk on the unseen data.
- The model correctly identified over 1,400 high-risk individuals and 1,200 no-risk individuals.
- While the rate of detection (70% recall) is stable, this absolute number highlights the real-world consequence of the model's errors and an area for future improvement.



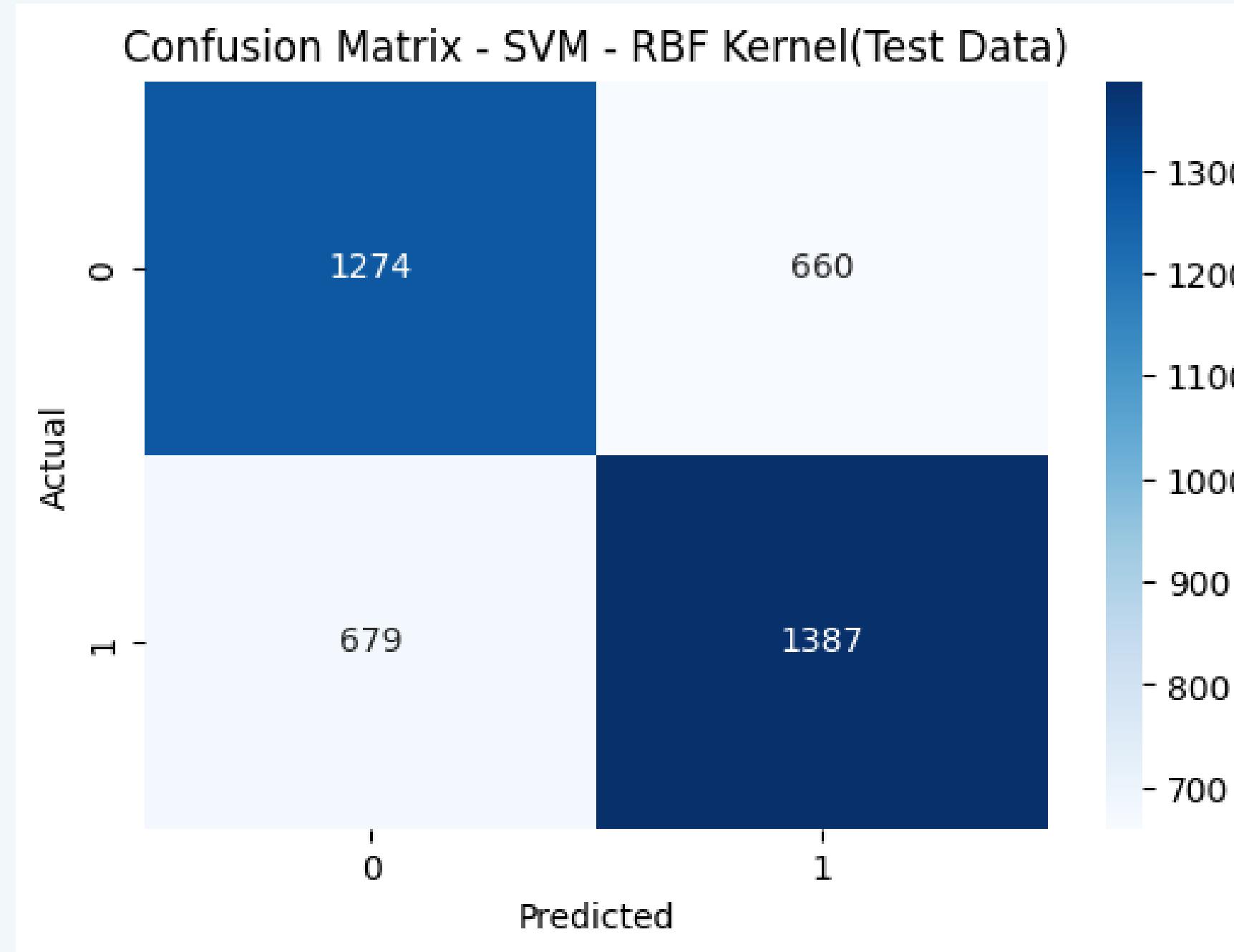
-  **01** Split data into train data and test data
-  **02** Standardize the train data using StandardScale
-  **03** Fit the train data into RBF model
-  **04** Evaluate model using confusion matrix, accuracy score and cross validation

# MODEL TESTING-SUPPORT VECTOR MACHINE

RBF KERNEL

13

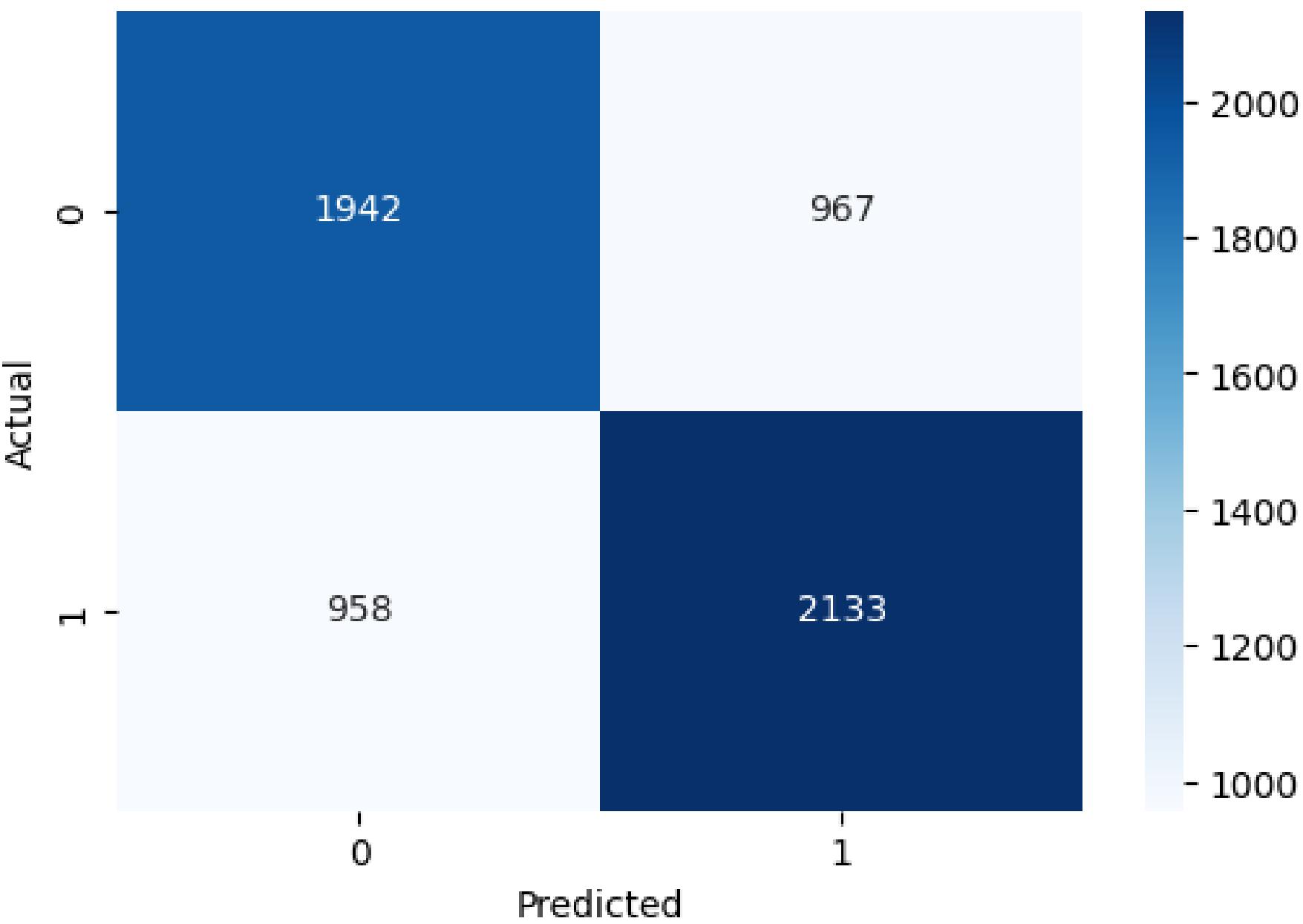
# CONFUSION MATRIX\_TEST DATA



14

# CONFUSION MATRIX\_TRAIN DATA

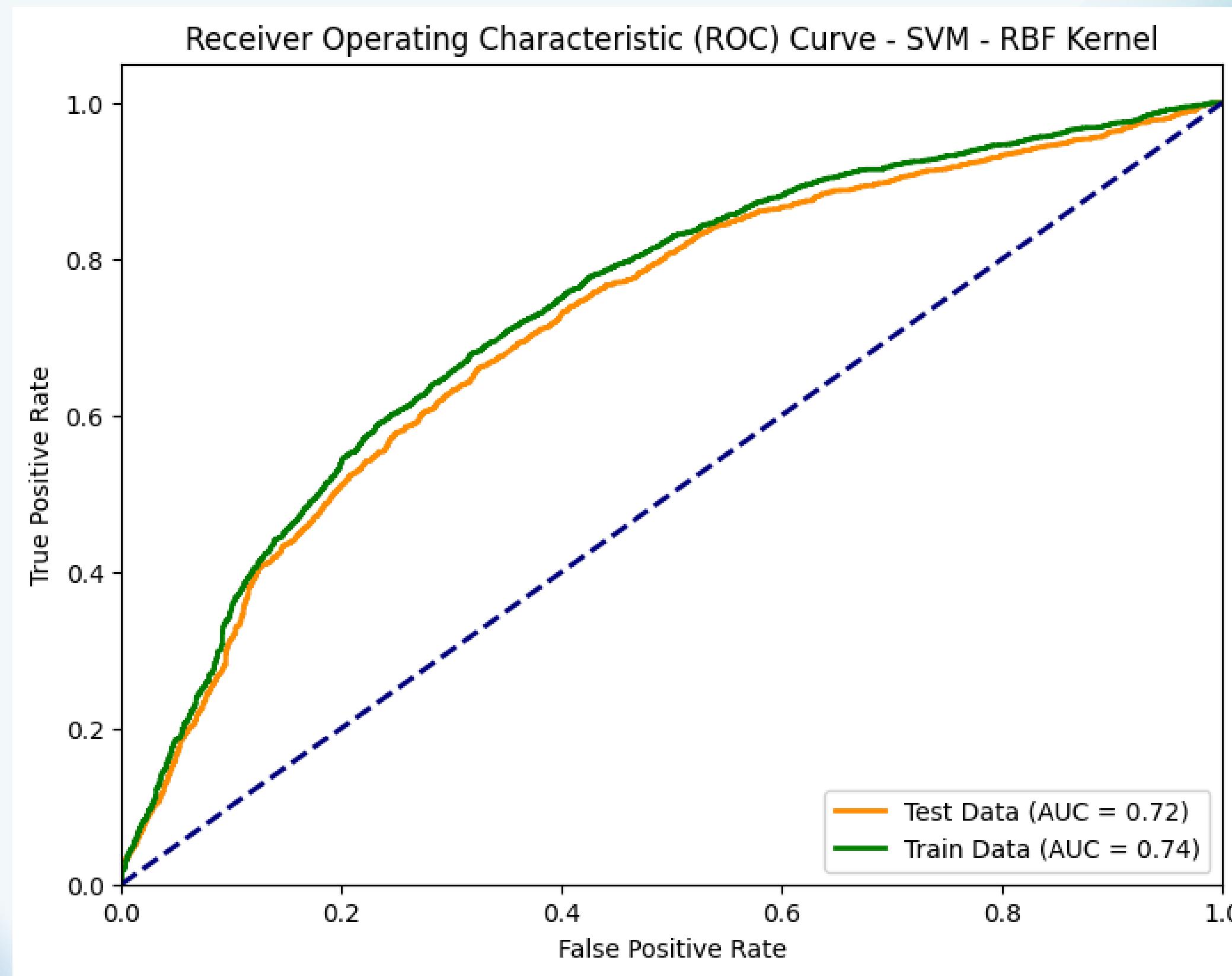
Confusion Matrix - SVM - RBF Kernel (Train Data)



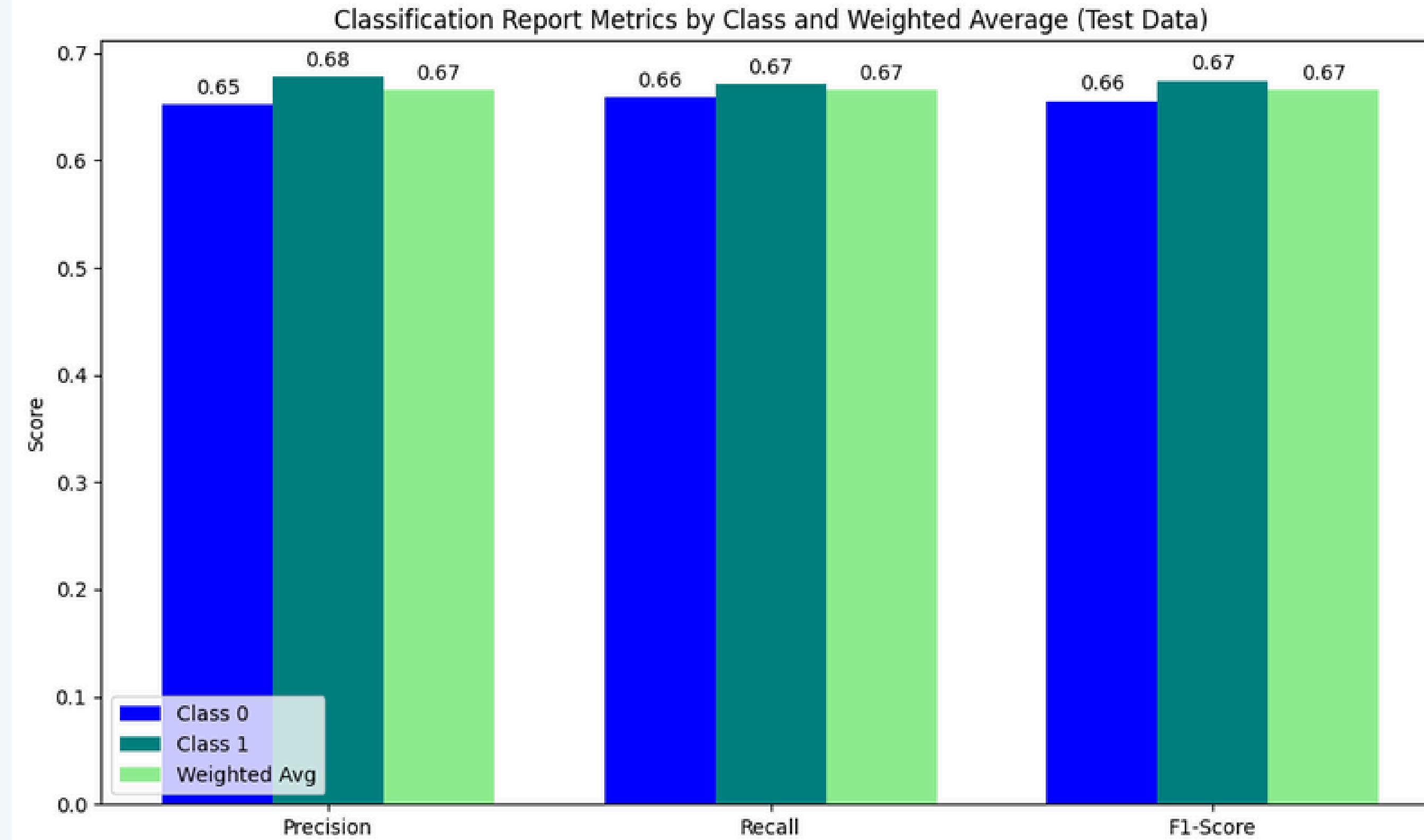
15

# ROC AUC

16

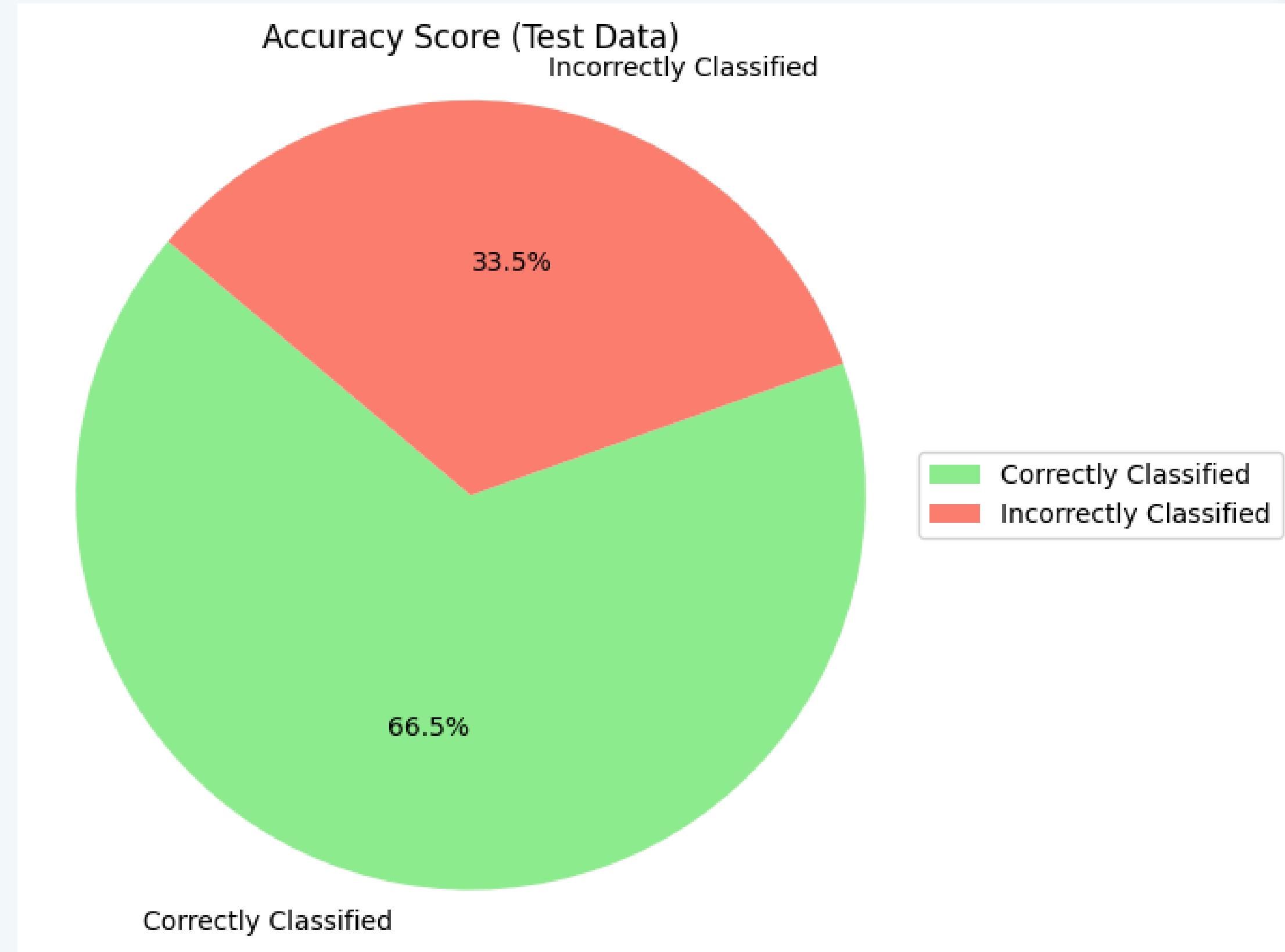


# EVALUATING MODEL\_TEST DATA



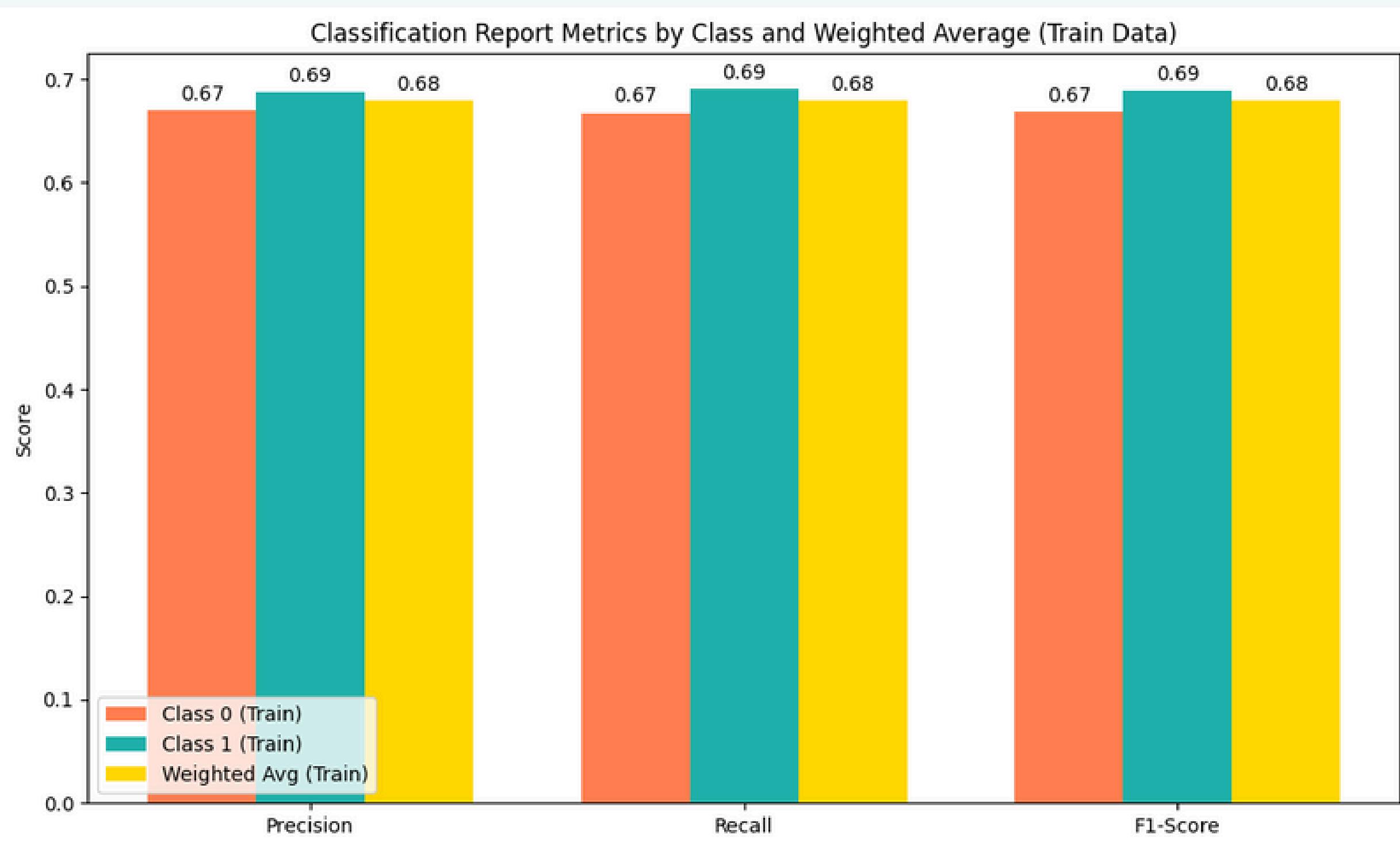
17

# 18

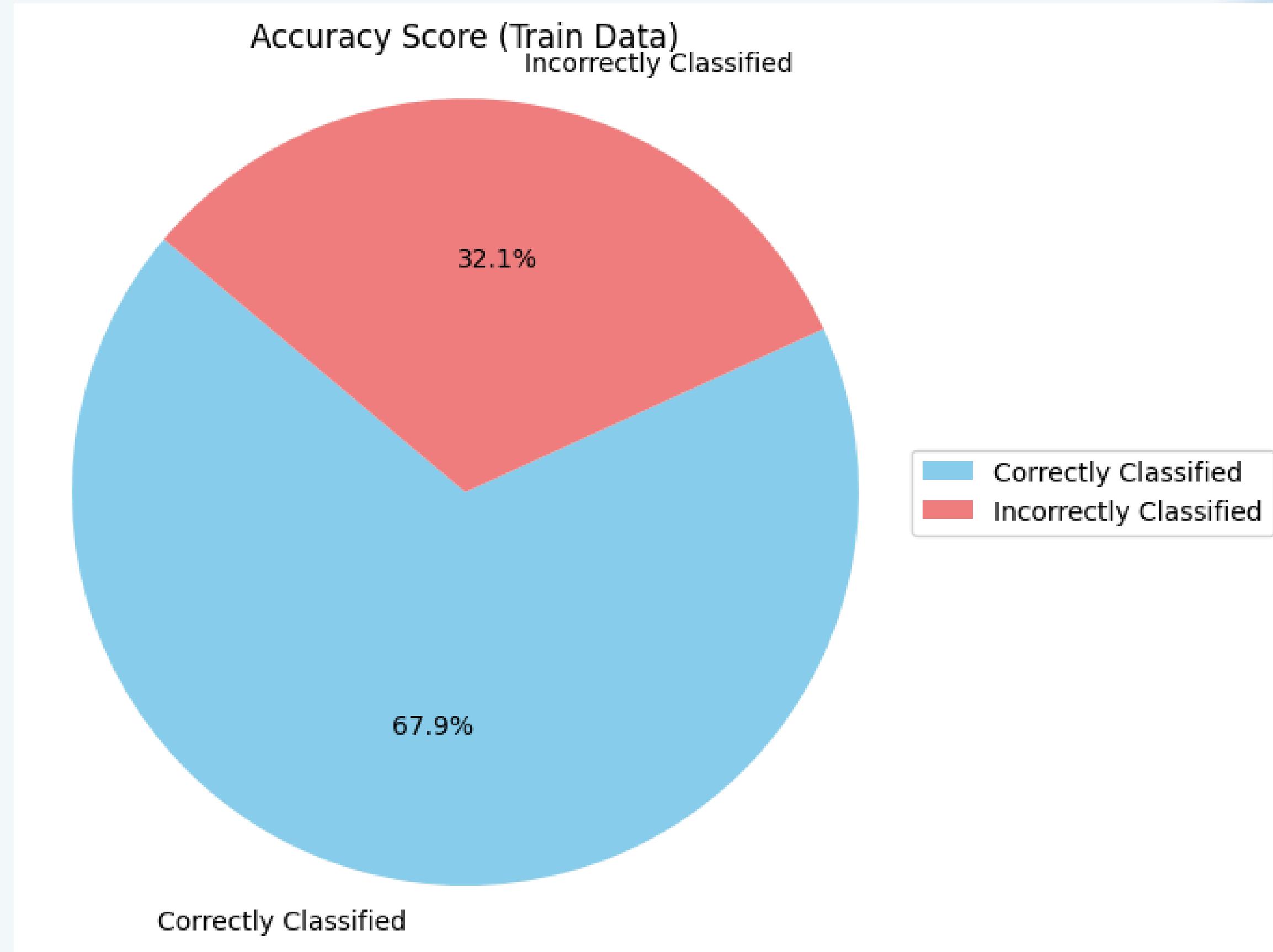


# EVALUATING MODEL\_TRAIN DATA

19

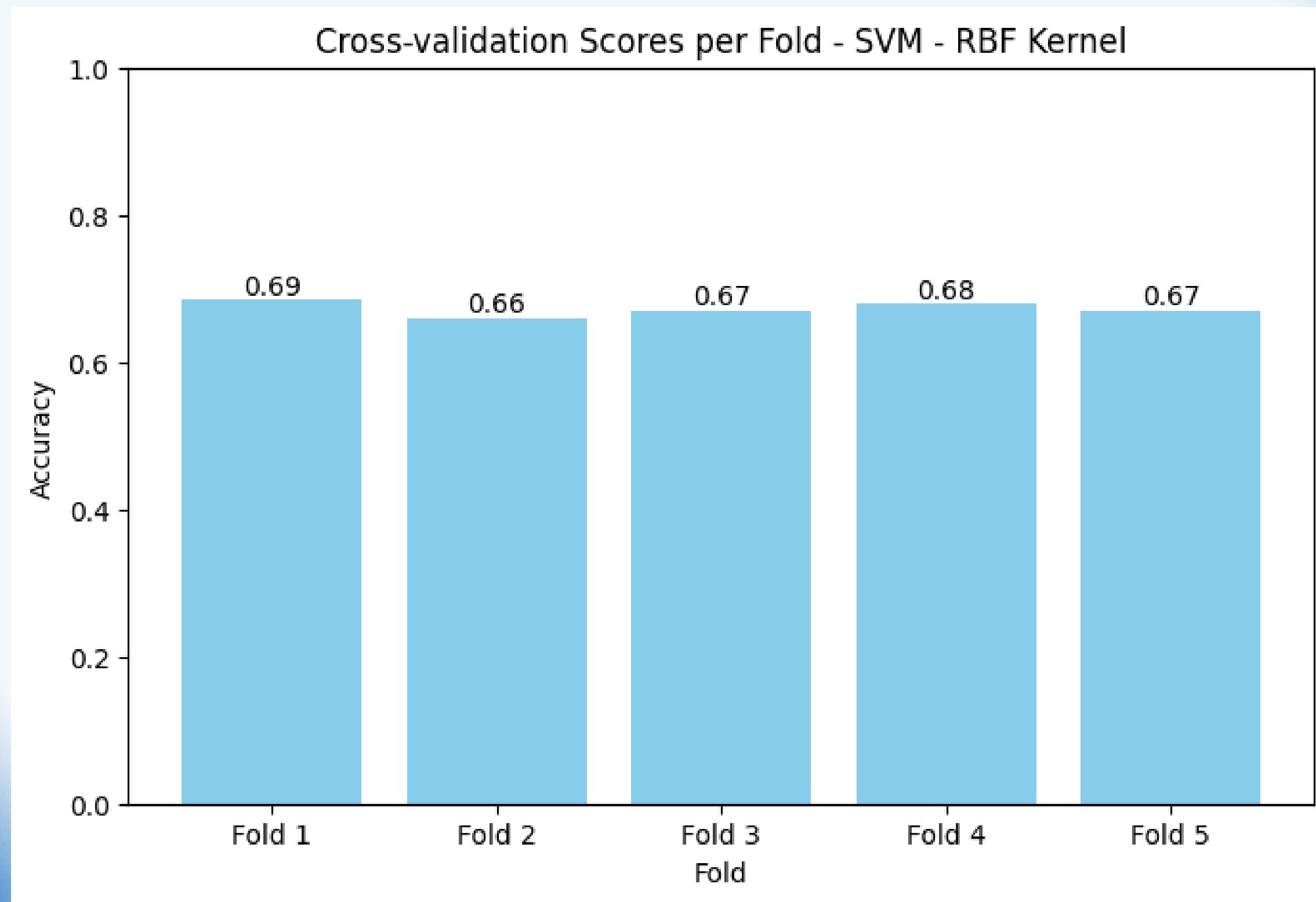


# 20



# CROSS-VALIDATION(5 FOLDS)

21



# 22 FINAL EVALUATION

Group\_E

Mental health risk detection

2025

## SVM-RBF

	precision	recall	f1-score
0	0.65	0.66	0.66
1	0.68	0.67	0.67
accuracy	0.67		

## KNN

	precision	recall	f1-score
0	0.66	0.63	0.64
1	0.67	0.70	0.68
accuracy	0.66		

## LR

	precision	recall	f1-score
0	0.65	0.64	0.65
1	0.67	0.68	0.68
accuracy	0.66		

## Gaussian NB

	precision	recall	f1-score
0	0.66	0.64	0.65
1	0.67	0.69	0.68
accuracy	0.66		

# 23

## CONCLUSION AND FUTURE WORKS

- There are no notable differences among models' performance
- Models seems to be underfitting
- We choose KNN model for its recall score being 0.70

In the future, we would find

- good and sufficient real-world dataset by
  - self-collect
  - relevant to Myanmar
  - consult with sectors professional for features

# THANK YOU

&  
ANY QUESTIOTN?