



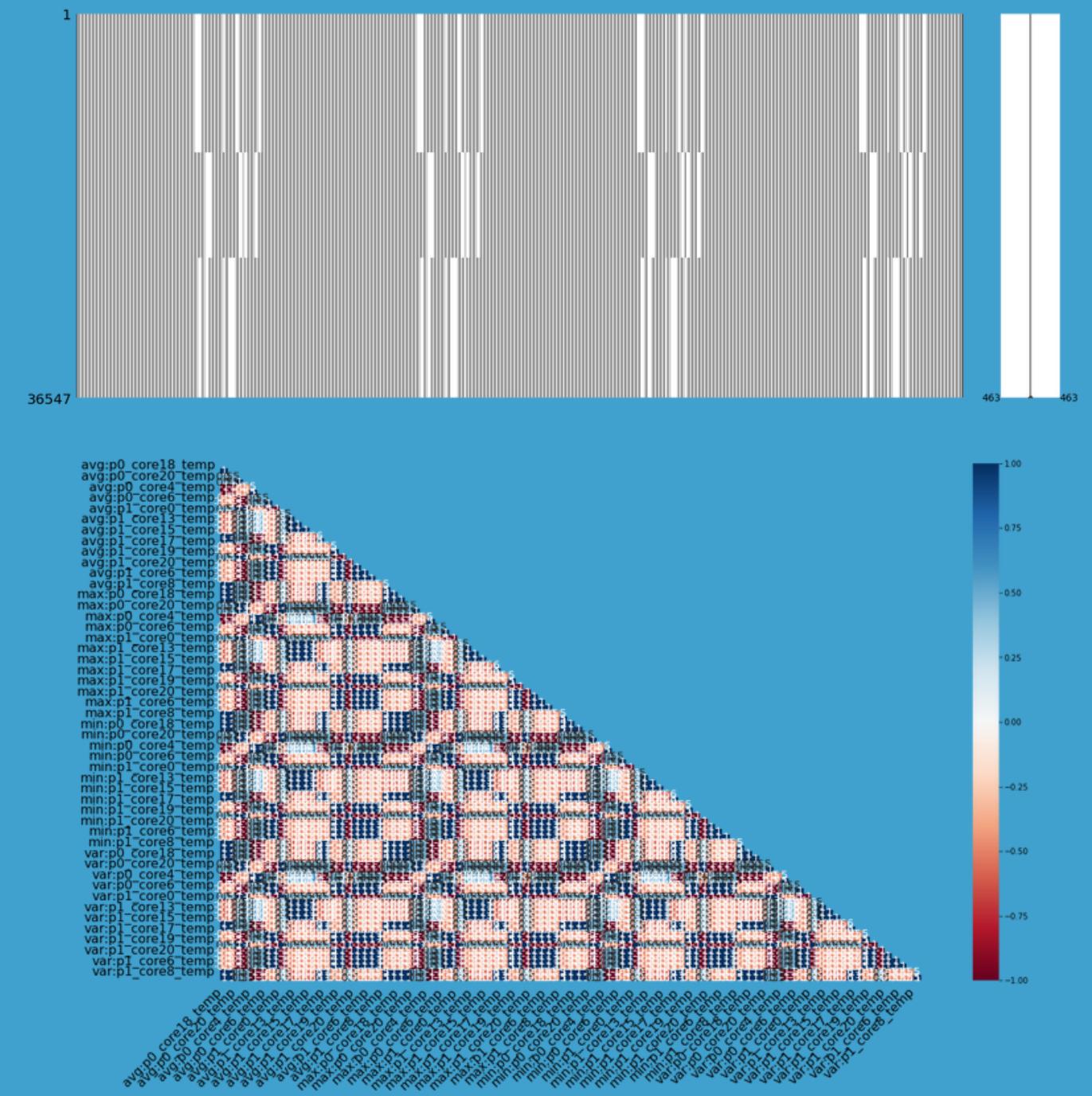
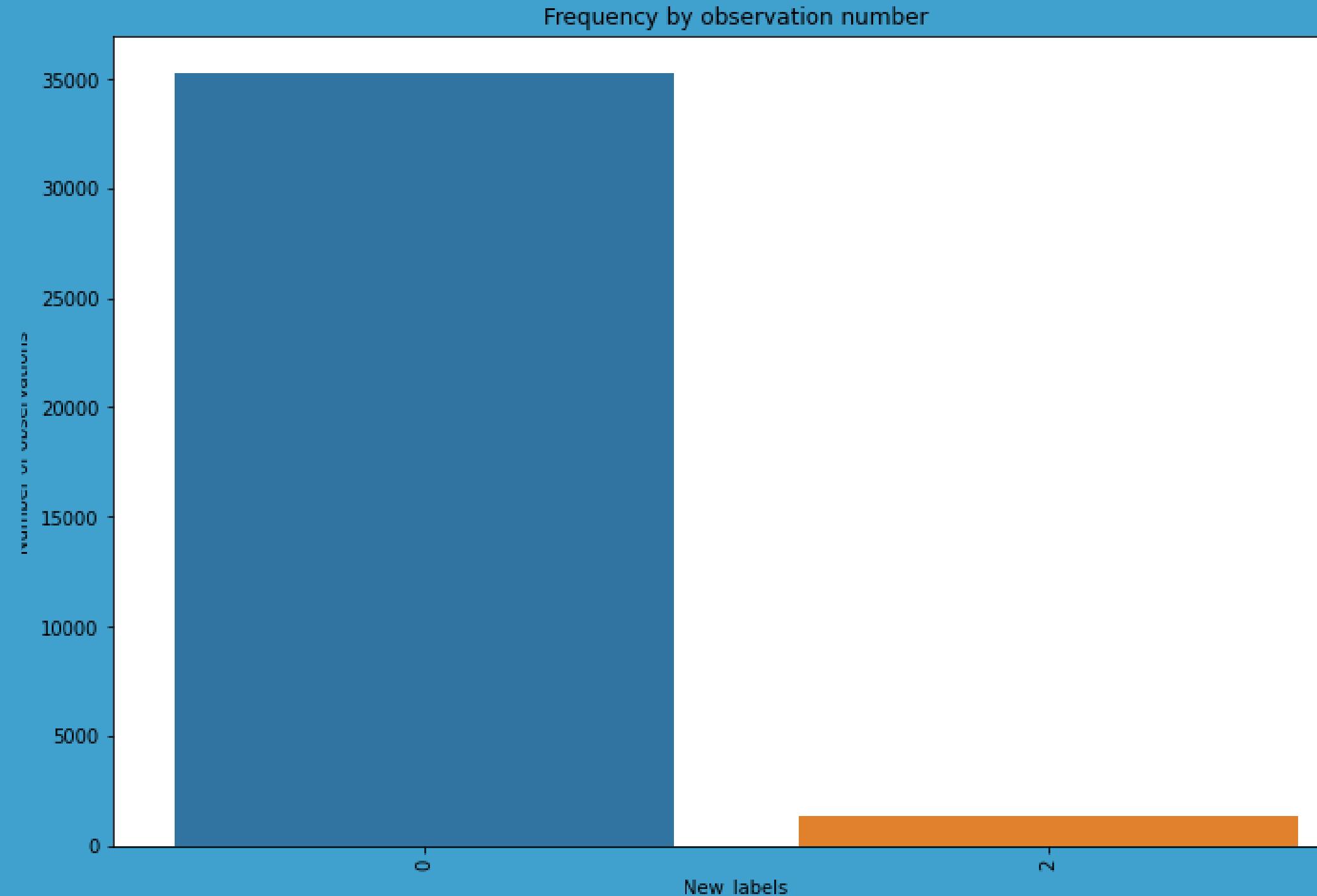
# Anomaly detection

Jyoti Yadav

# Contents

1. Data Exploration
2. Preprocessing
3. Categories of Machine learning approaches
  - a. Supervised
  - b. Unsupervised
  - c. Semi-supervised
  - d. Self-supervised
4. Models
5. Results
6. Conclusion

# Data Exploration



# Preprocessing

01

## Removal of Irrelevant columns

- Label
- Timestamp

02

## Removal of Nans

- Drop the columns which contain at least one NaN value
- Drop the rows which contain at least NaN values

03

## Train Test split

- Splitting the dataset with 80% training and 20% testing

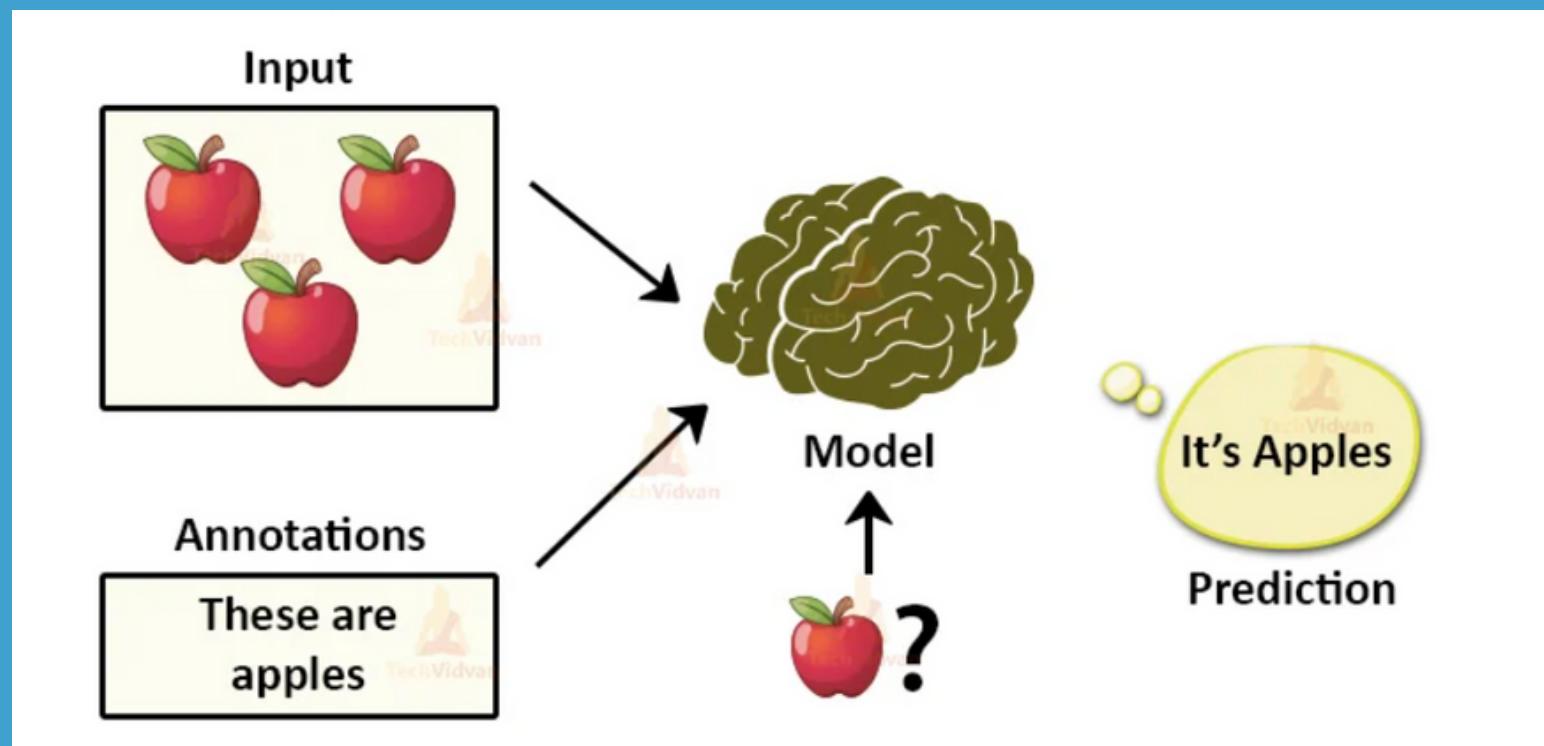
04

## Feature Scaling

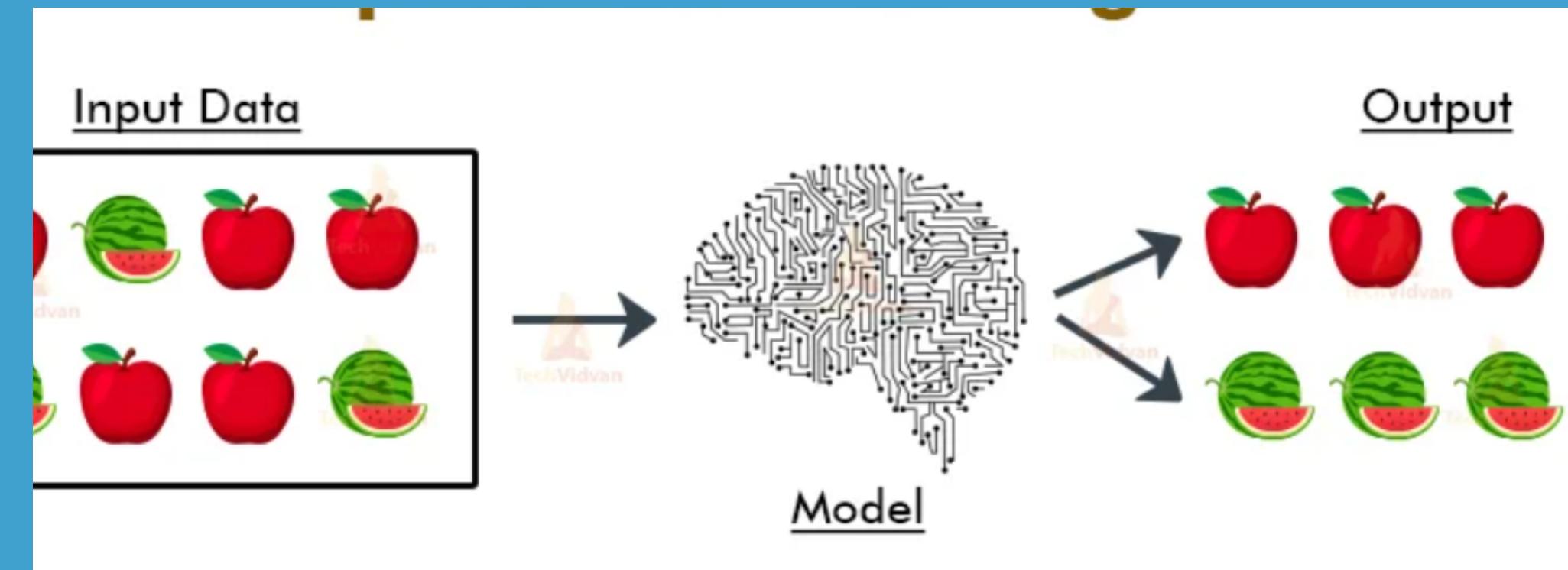
- Performing Minmax scaling on the dataframe

# Machine Learning Approaches

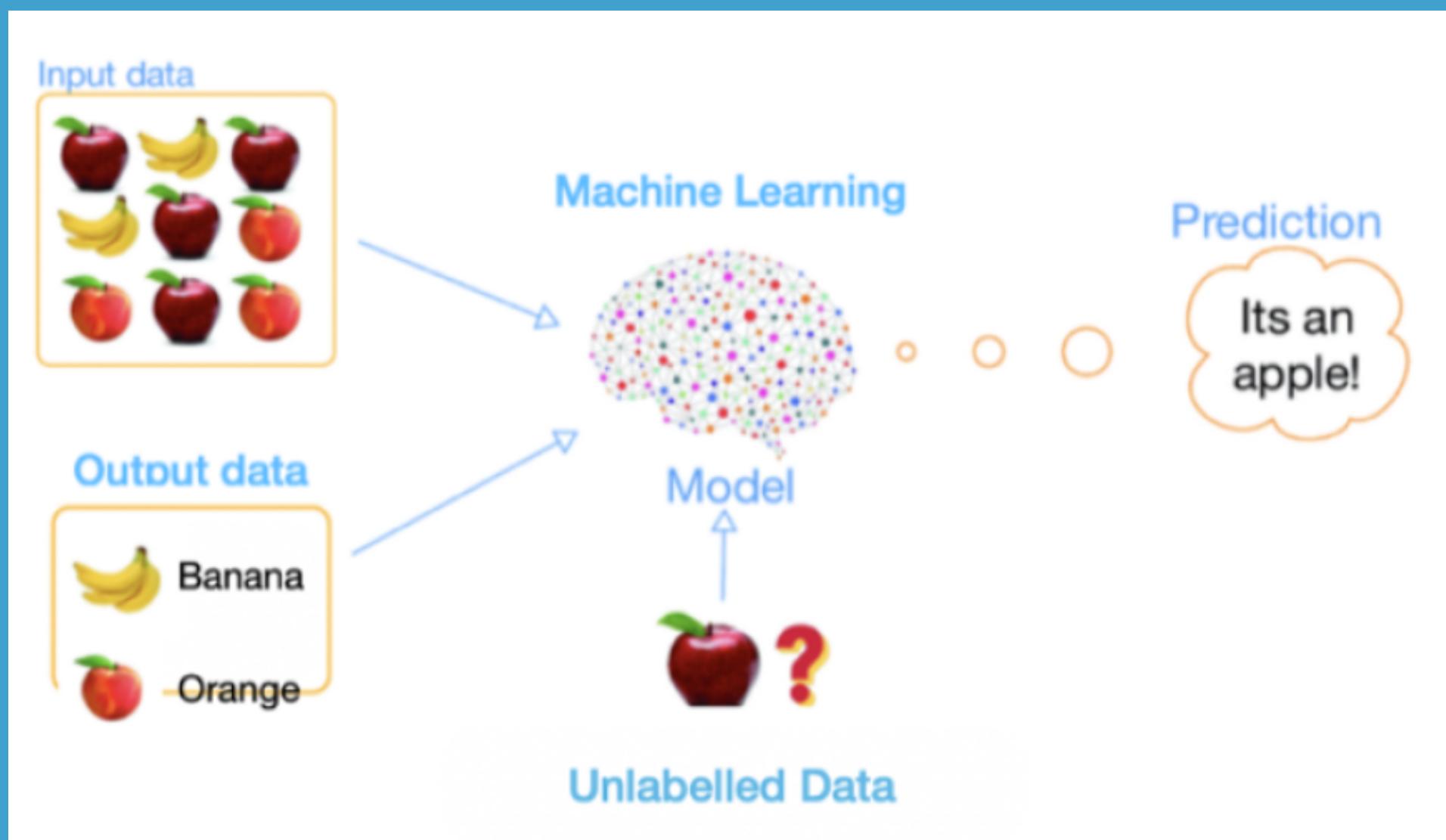
## Supervised Learning



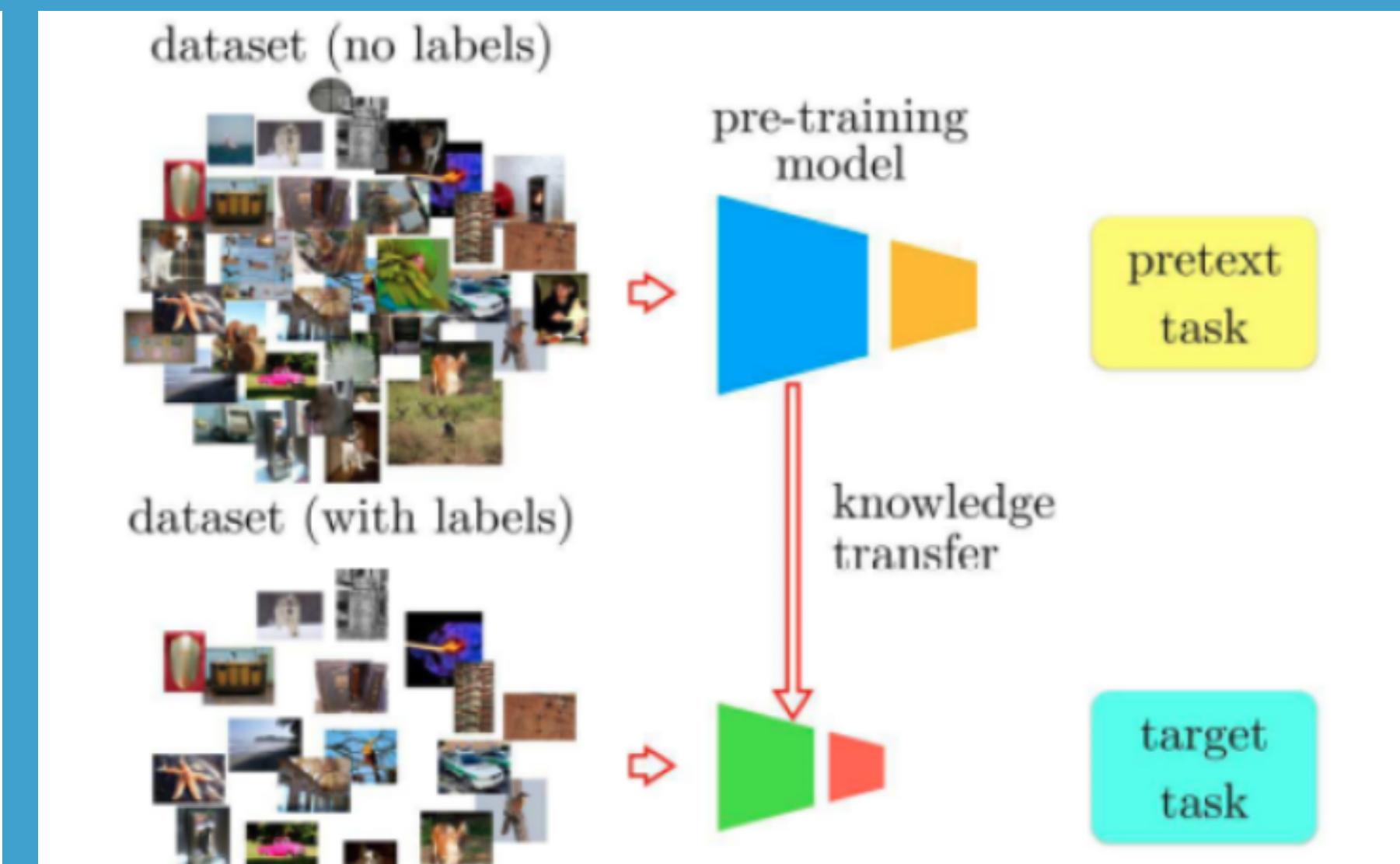
## Unsupervised Learning



# Semi supervised Learning



# Self supervised Learning



# Semi supervised Learning

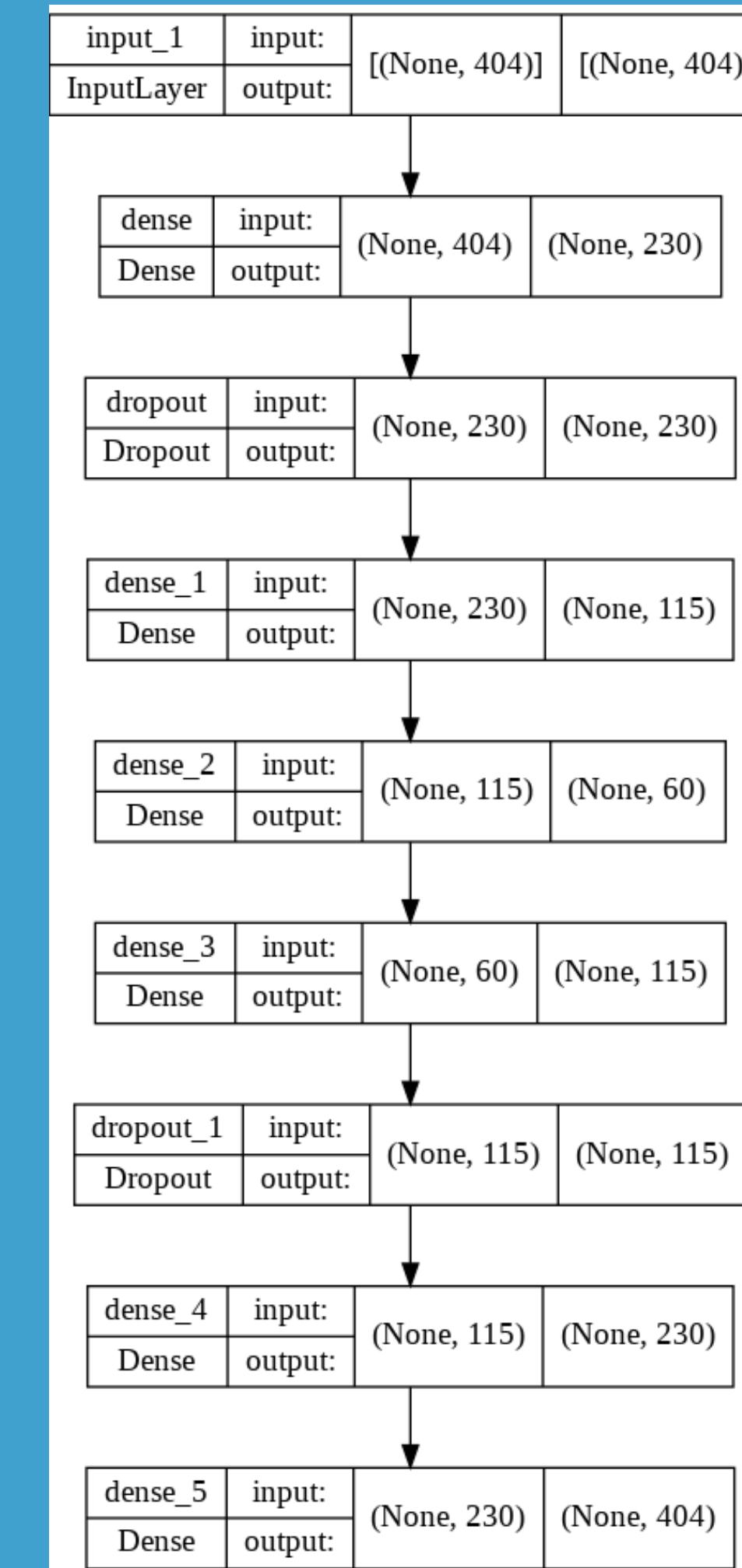
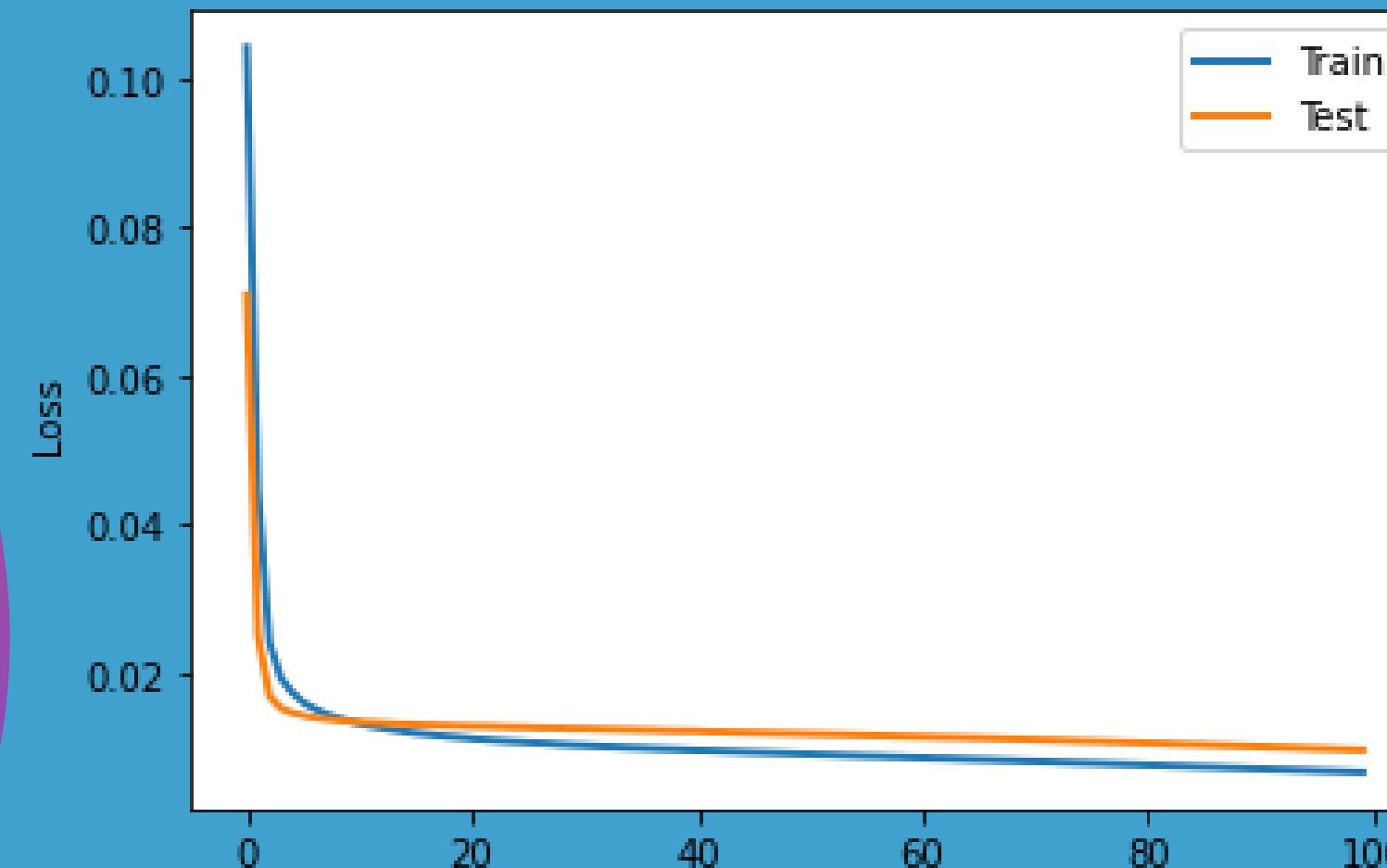
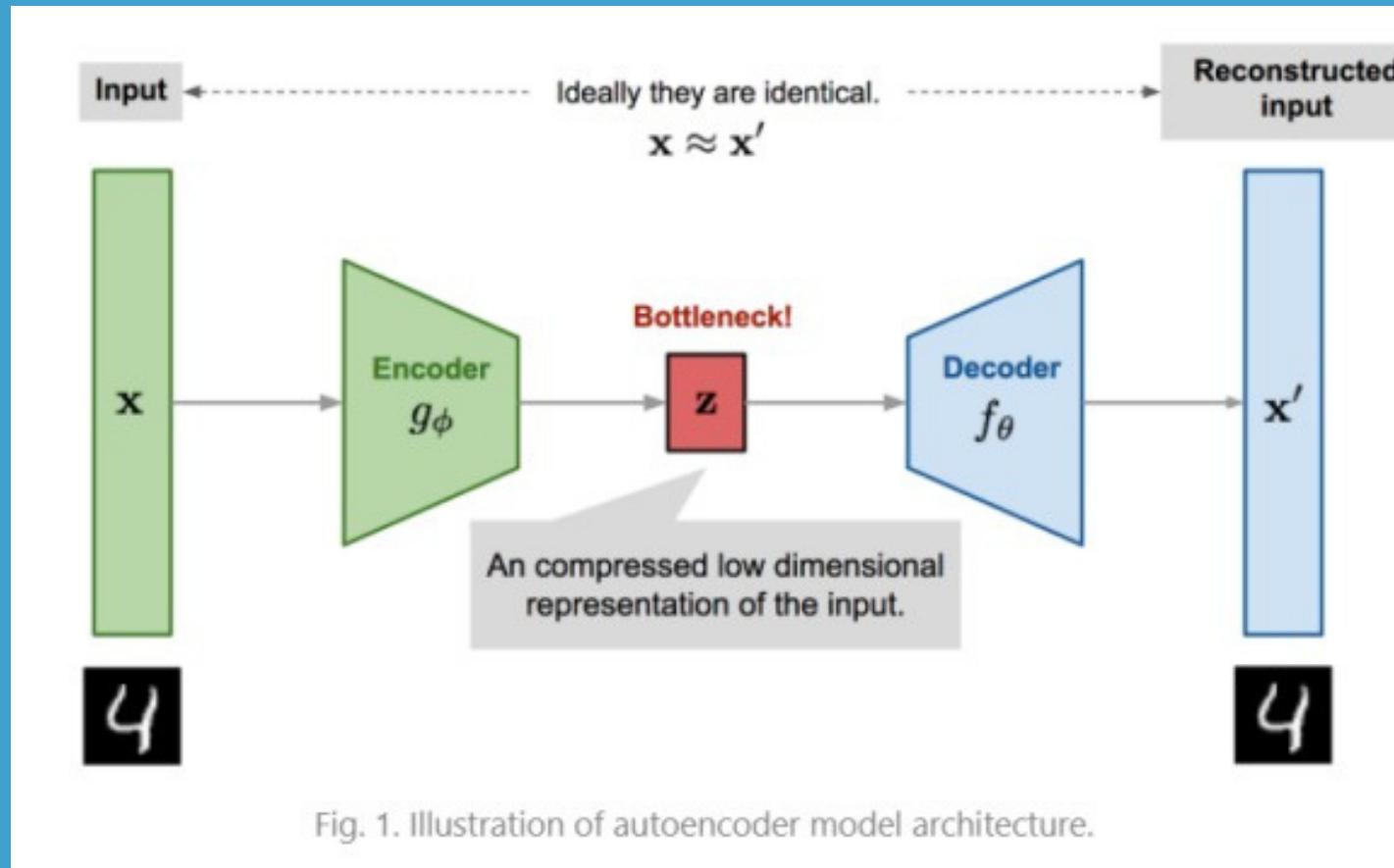
# Autoencoders

**Batch size = 32**

**Epoch = 100**

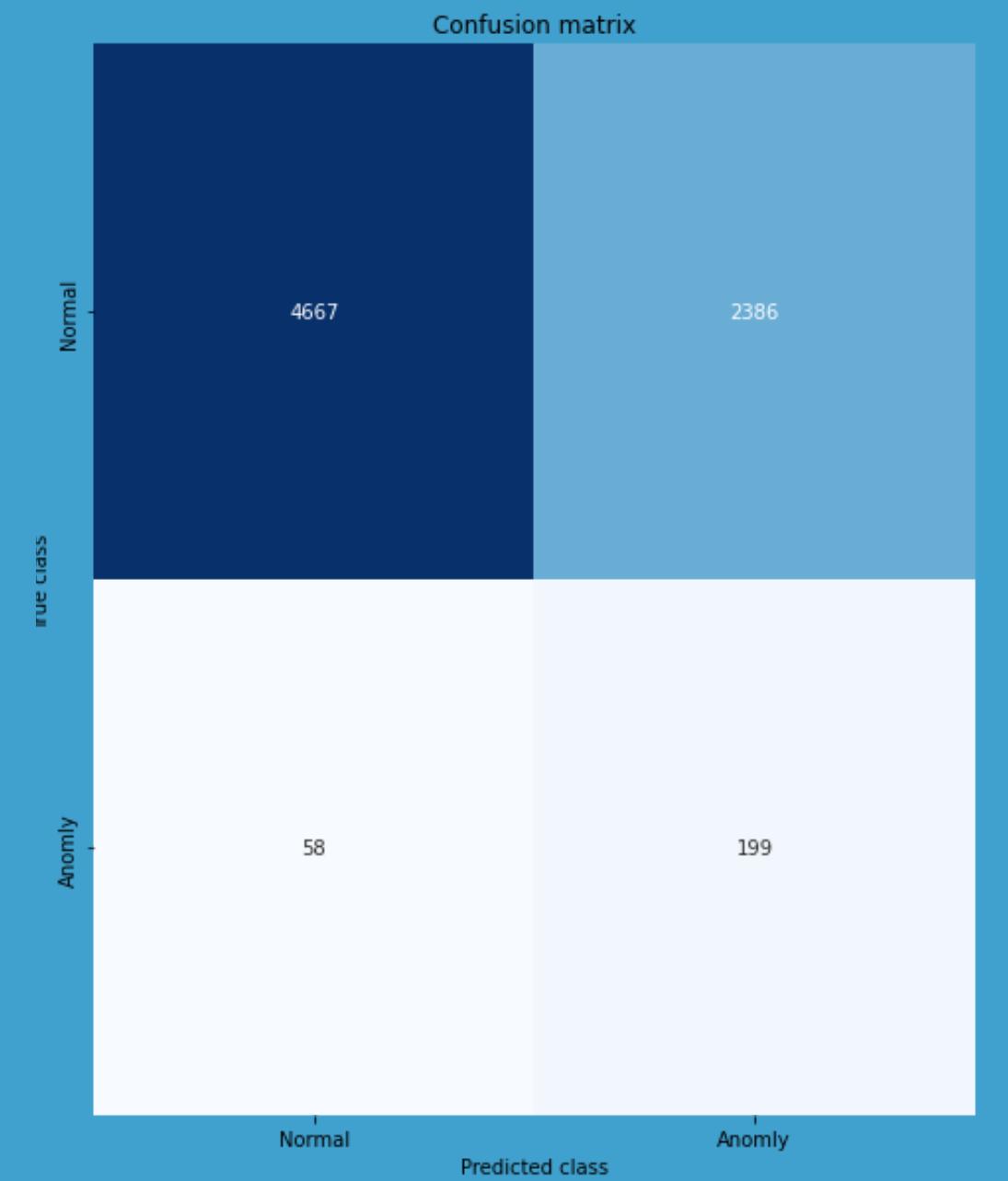
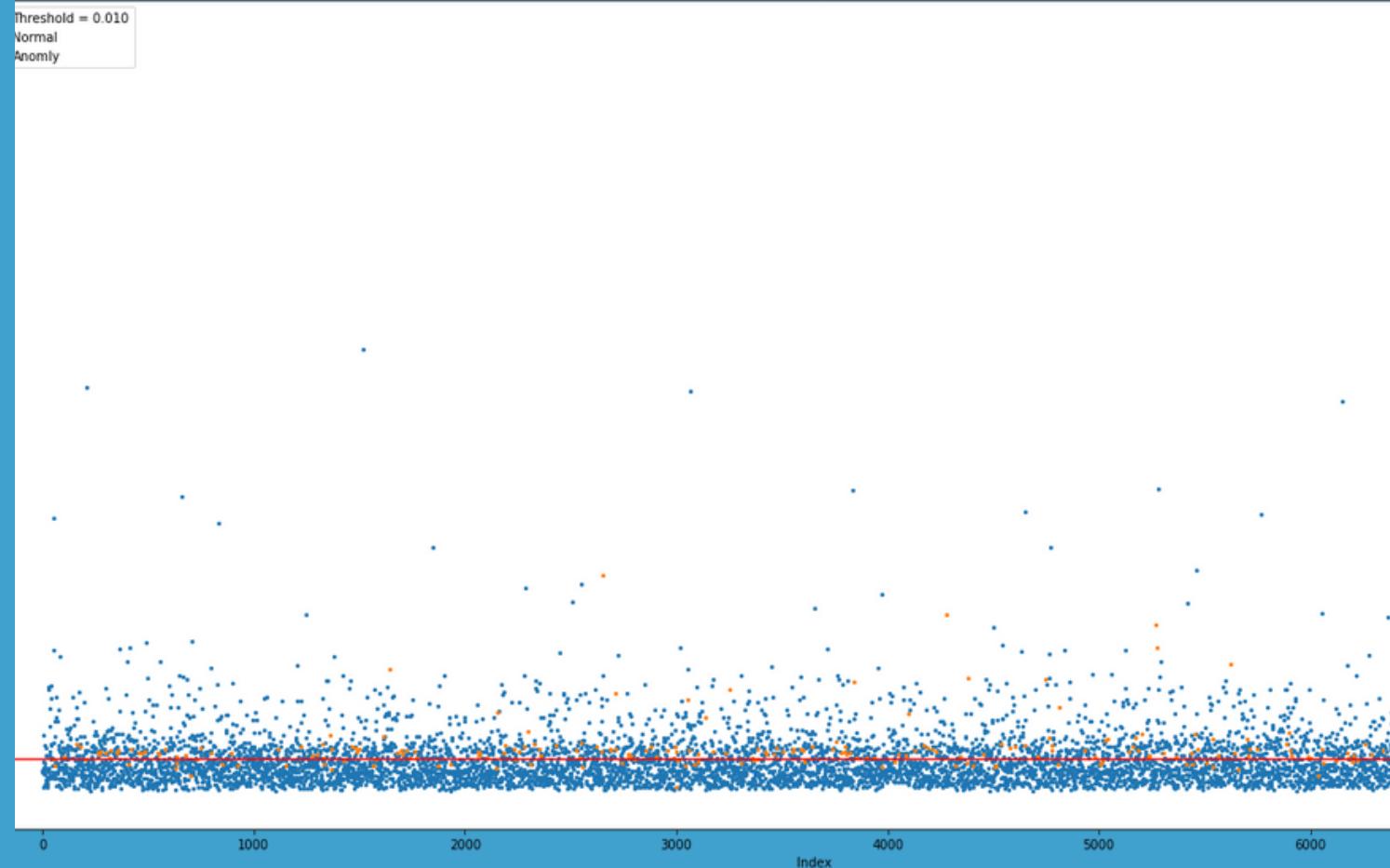
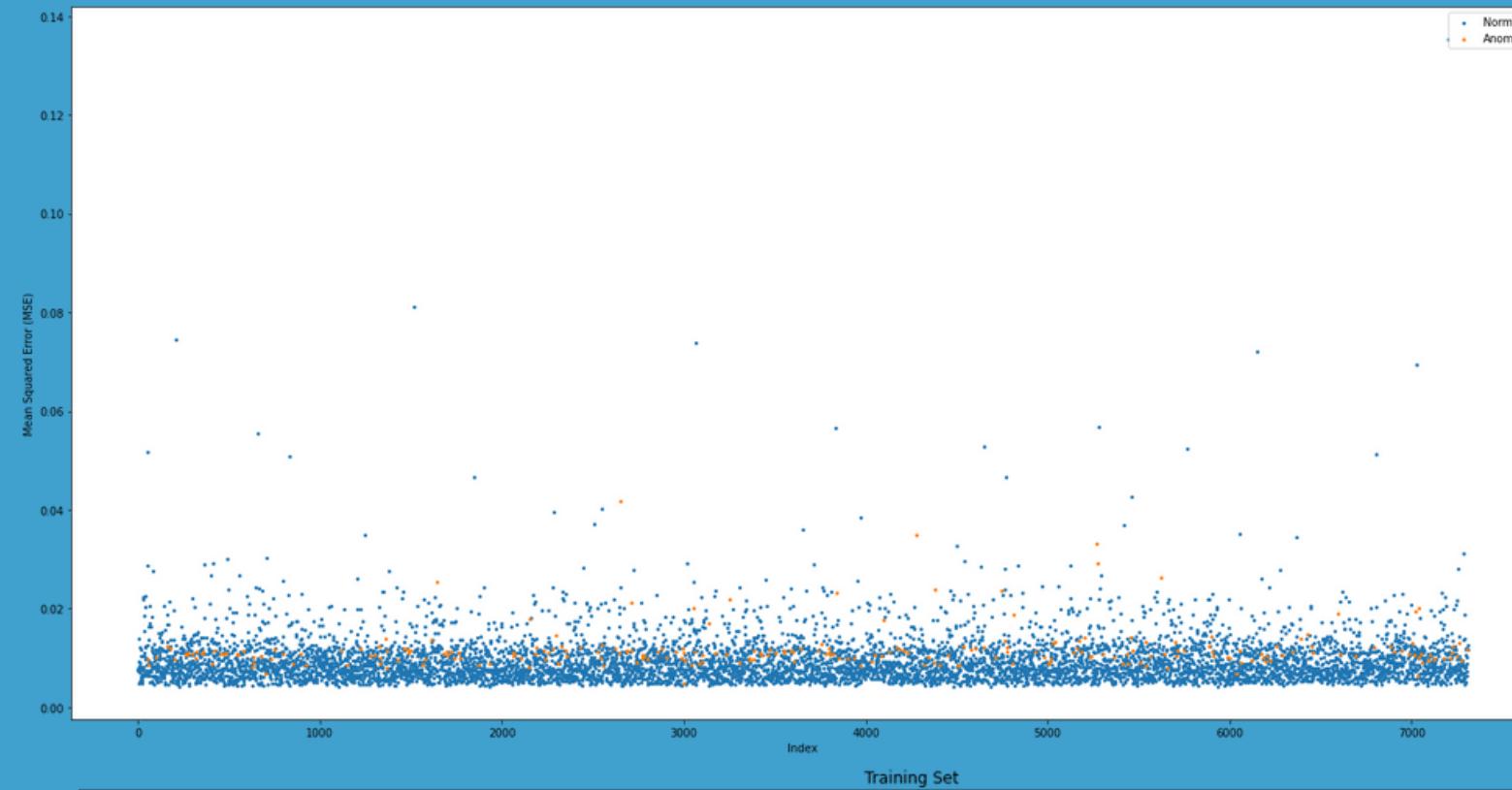
**Optimizer = SGD**

**Learning rate = 1e-5**





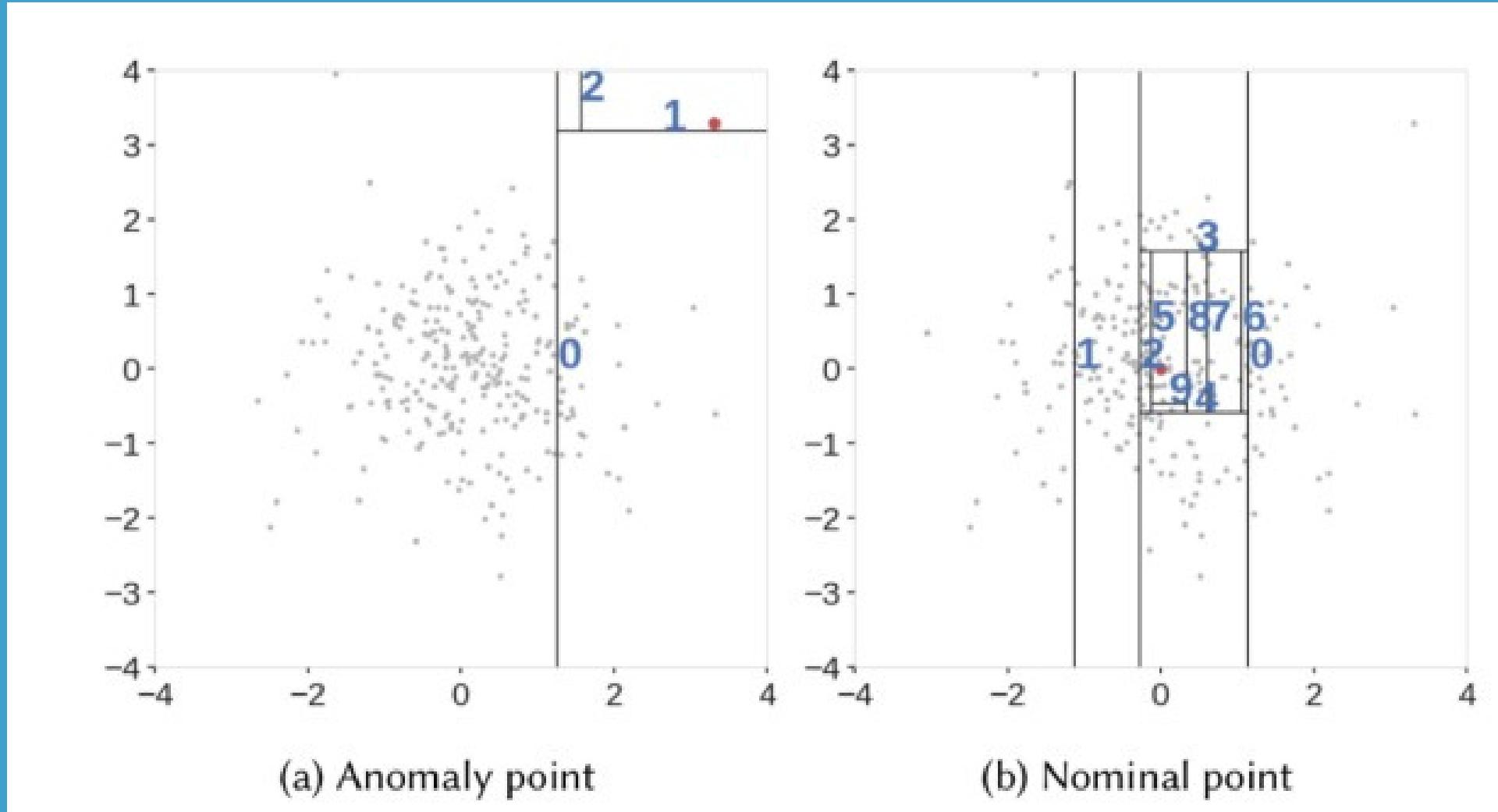
# Statistics



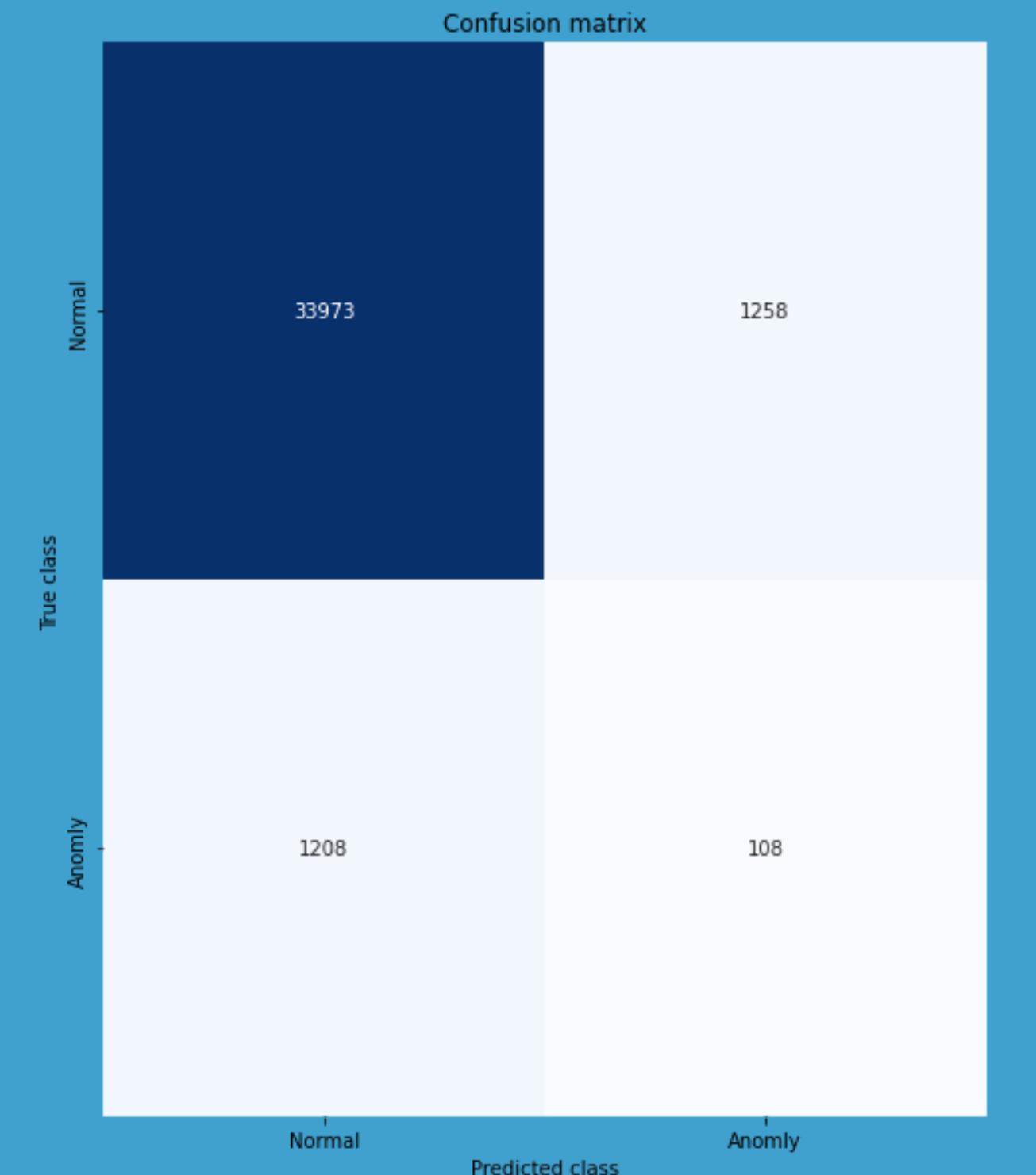
**Accuracy Score : 0.66**  
**ROC AUC score : 0.71**  
**recall : 0.77**  
**precision : 0.07**  
**f1\_score : 0.14**

# Unsupervised learning

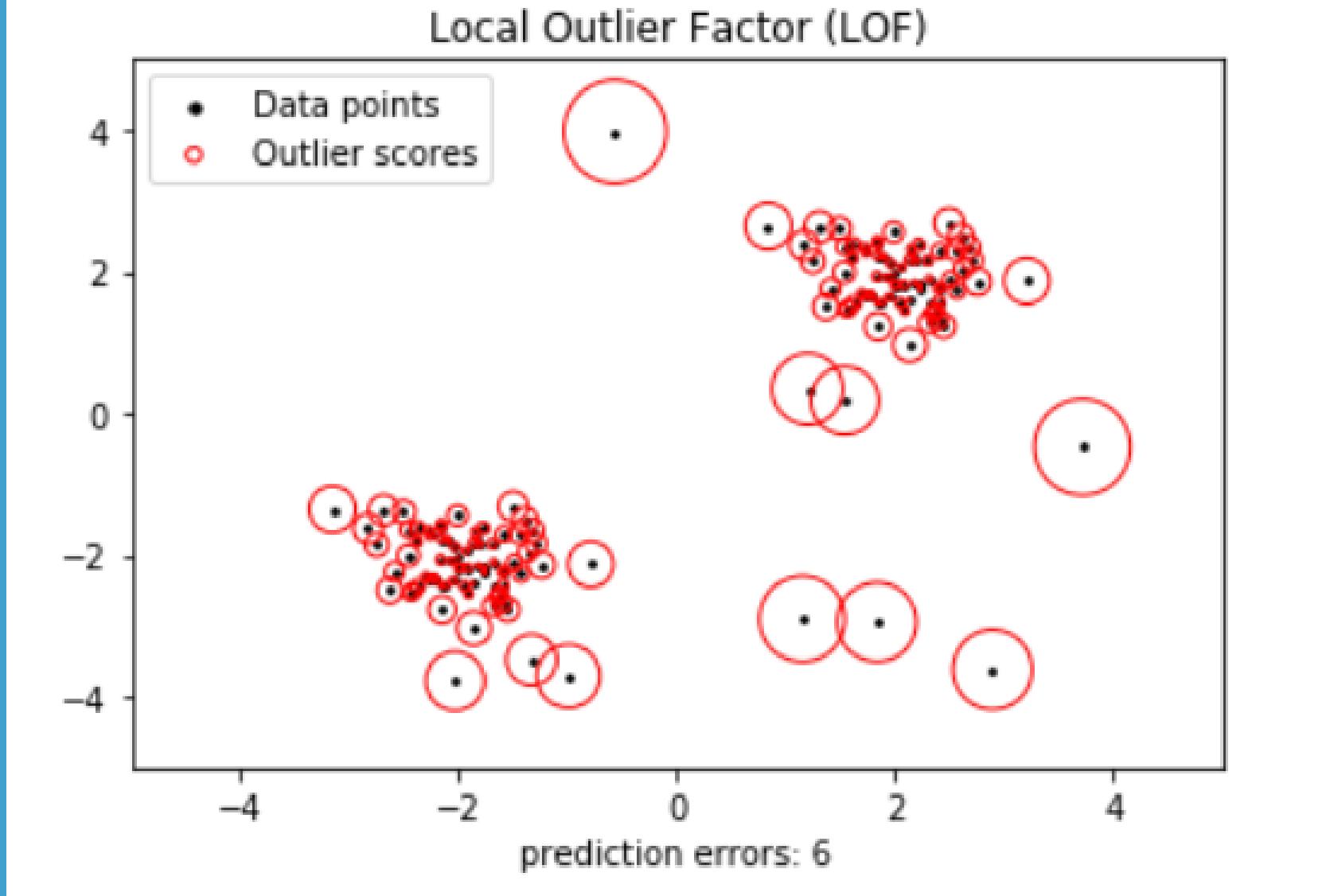
# Isolation Forest



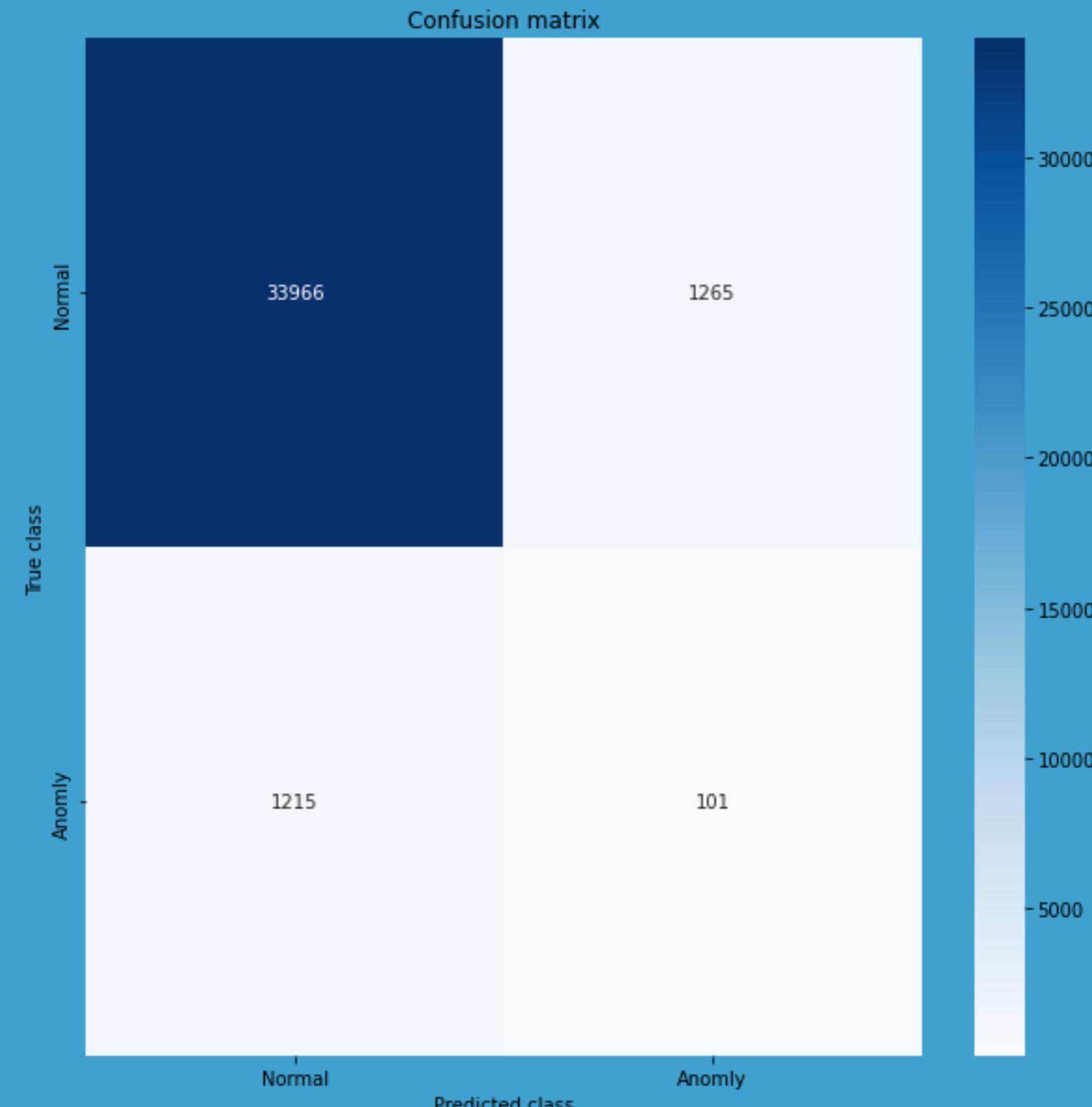
**Accuracy Score : 0.93**  
**ROC AUC score : 0.52**  
**recall : 0.08**  
**precision : 0.07**  
**f1\_score : 0.08**



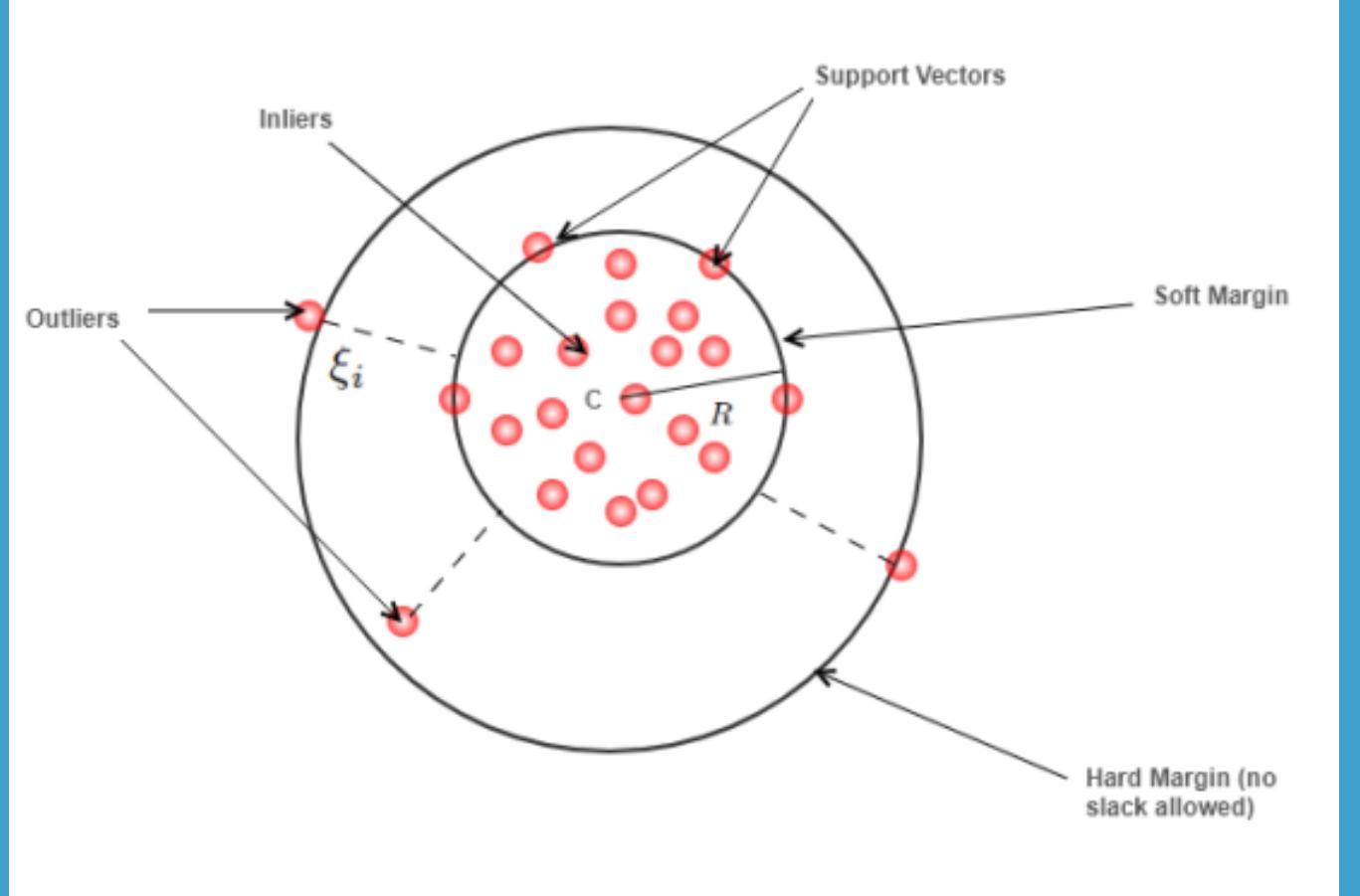
# Local Outlier Factor



**Accuracy Score : 0.93**  
**ROC AUC score : 0.52**  
**recall : 0.076**  
**precision : 0.073**  
**f1\_score : 0.075**



# One class SVM



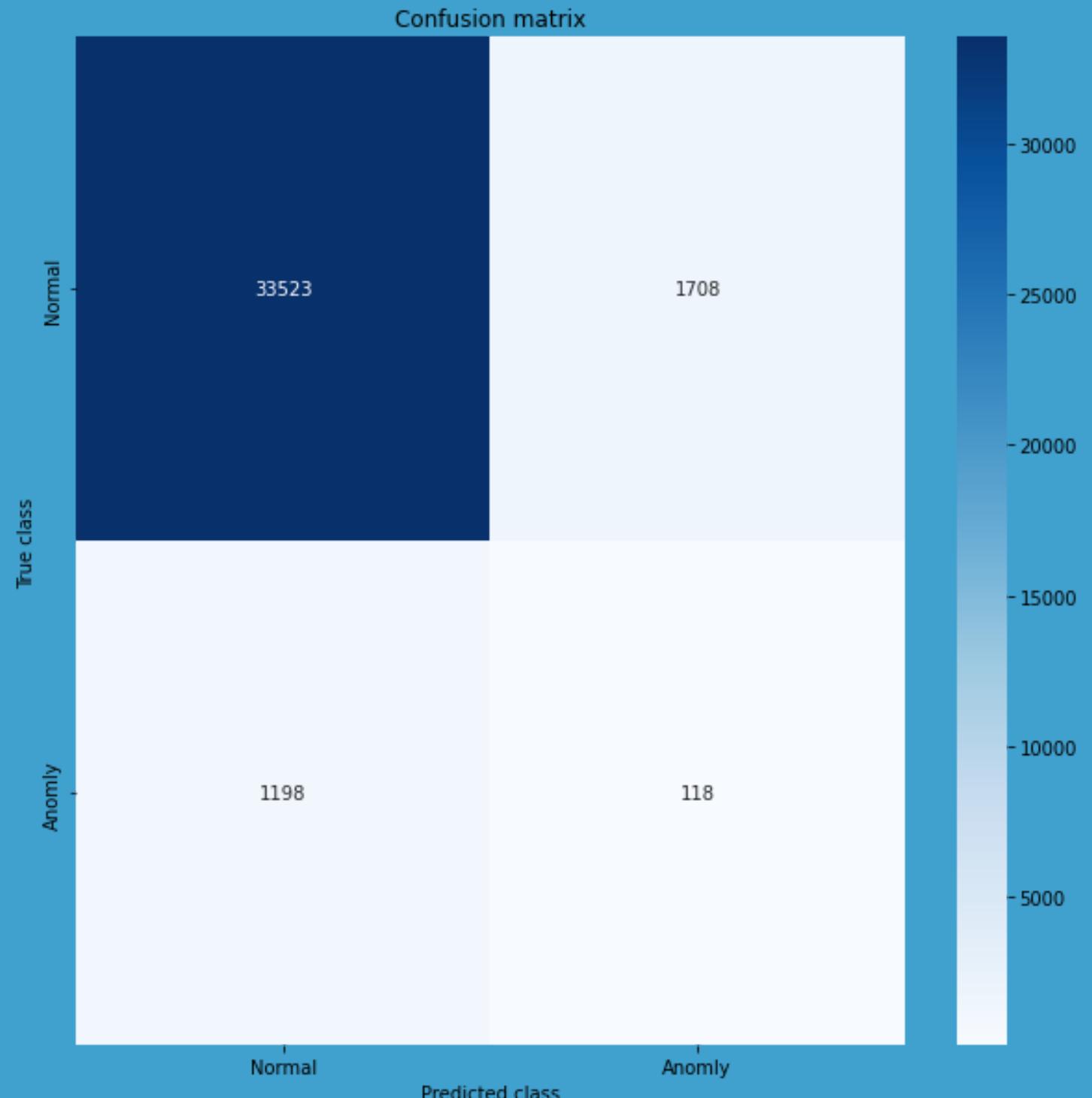
**Accuracy Score : 0.92**

**ROC AUC score : 0.52**

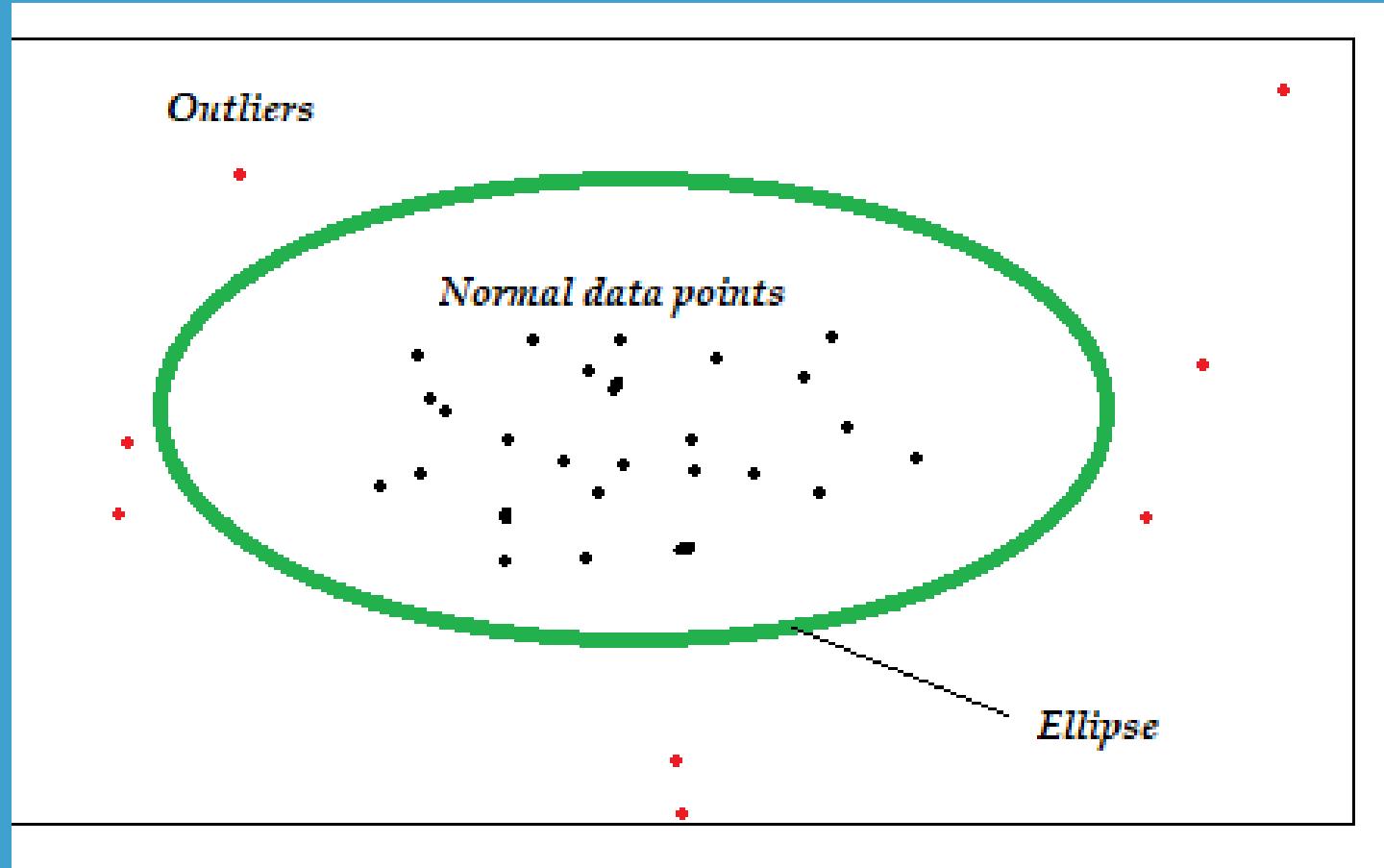
**recall : 0.08**

**precision : 0.064**

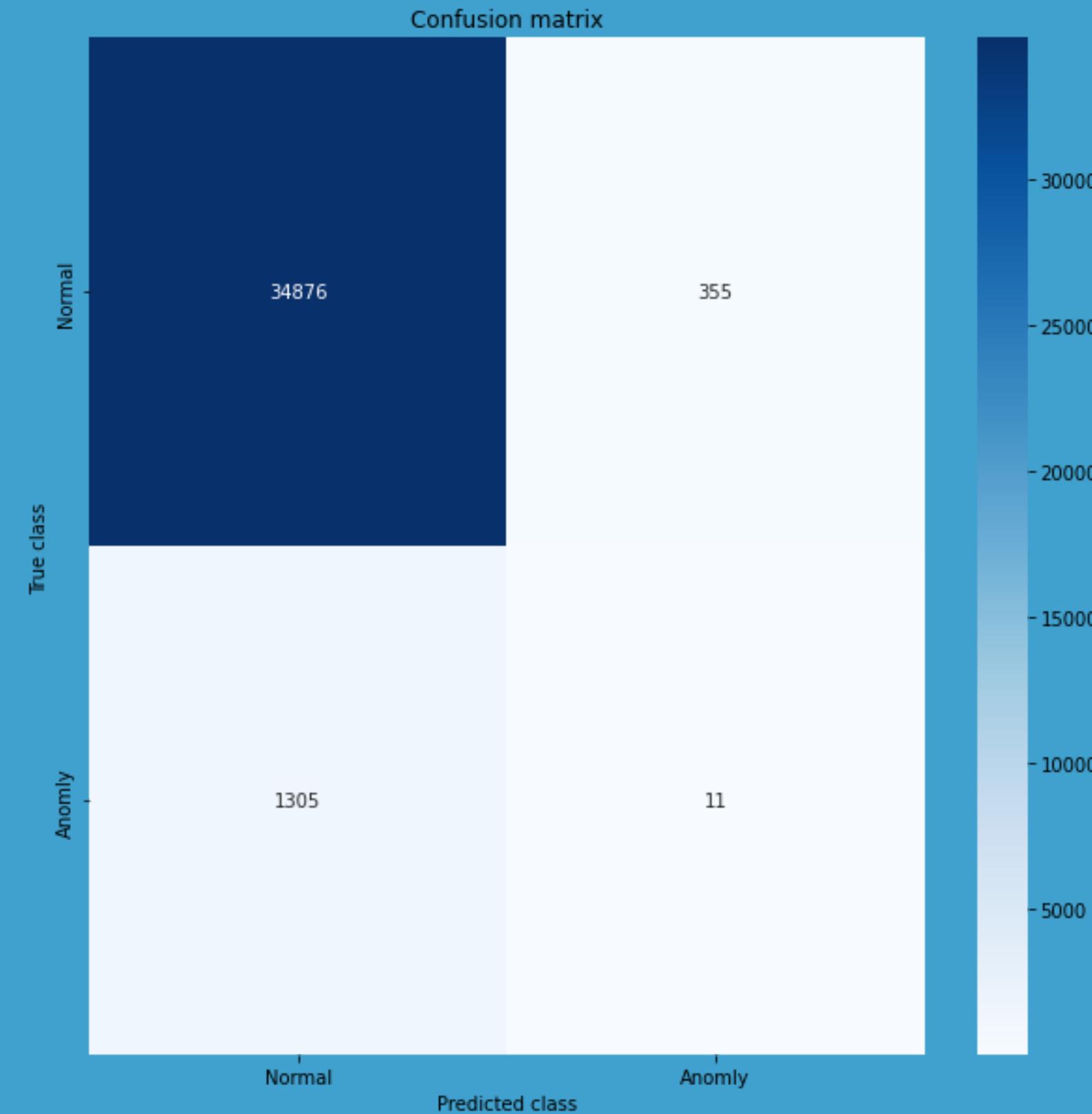
**f1\_score : 0.075**



# Elliptic Envelope

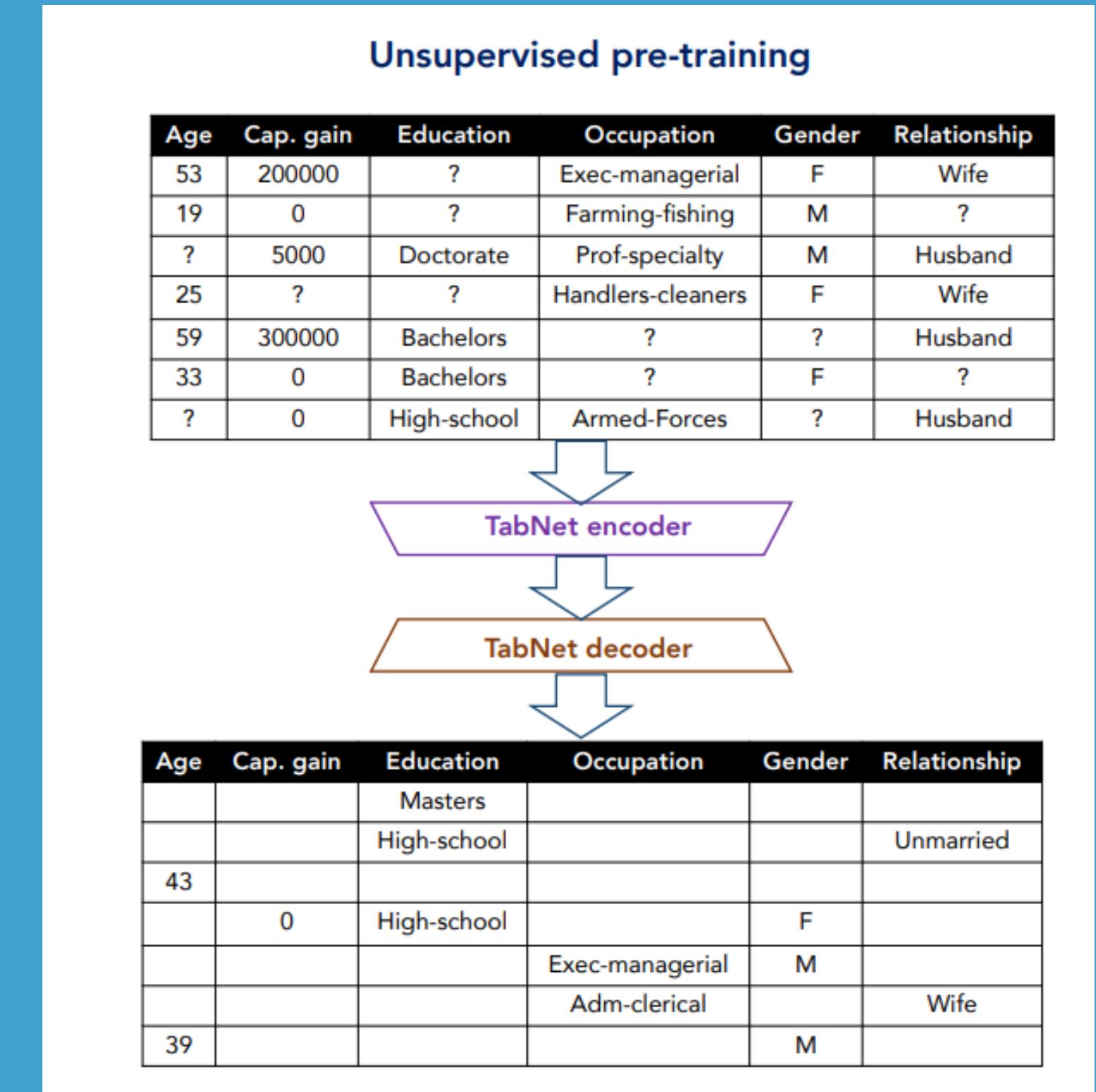
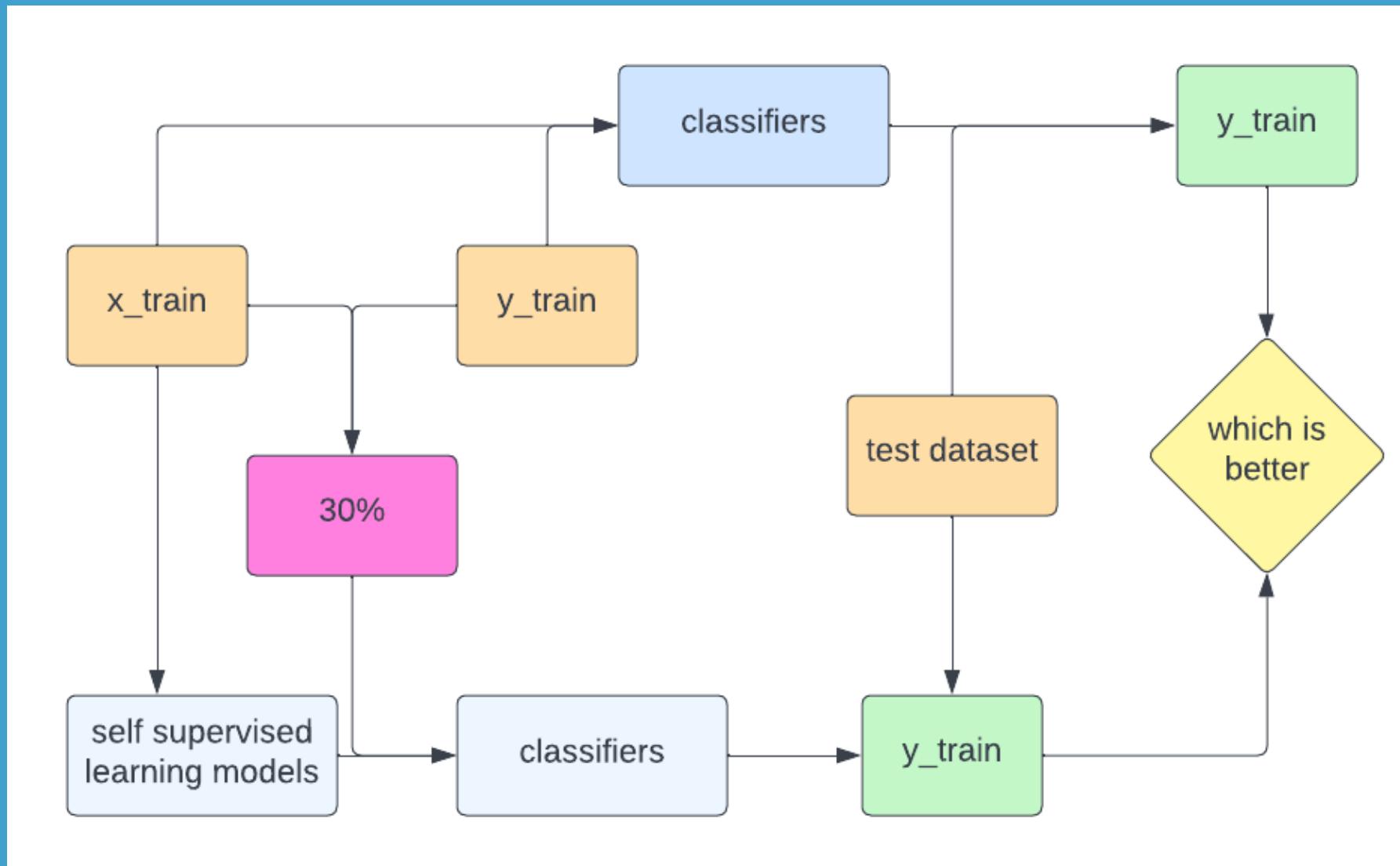


**Accuracy Score : 0.95**  
**ROC AUC score : 0.49**  
**recall : 0.008**  
**precision : 0.03**  
**f1\_score : 0.013**



# **self supervised learning**

# TABNET



# Cases

01

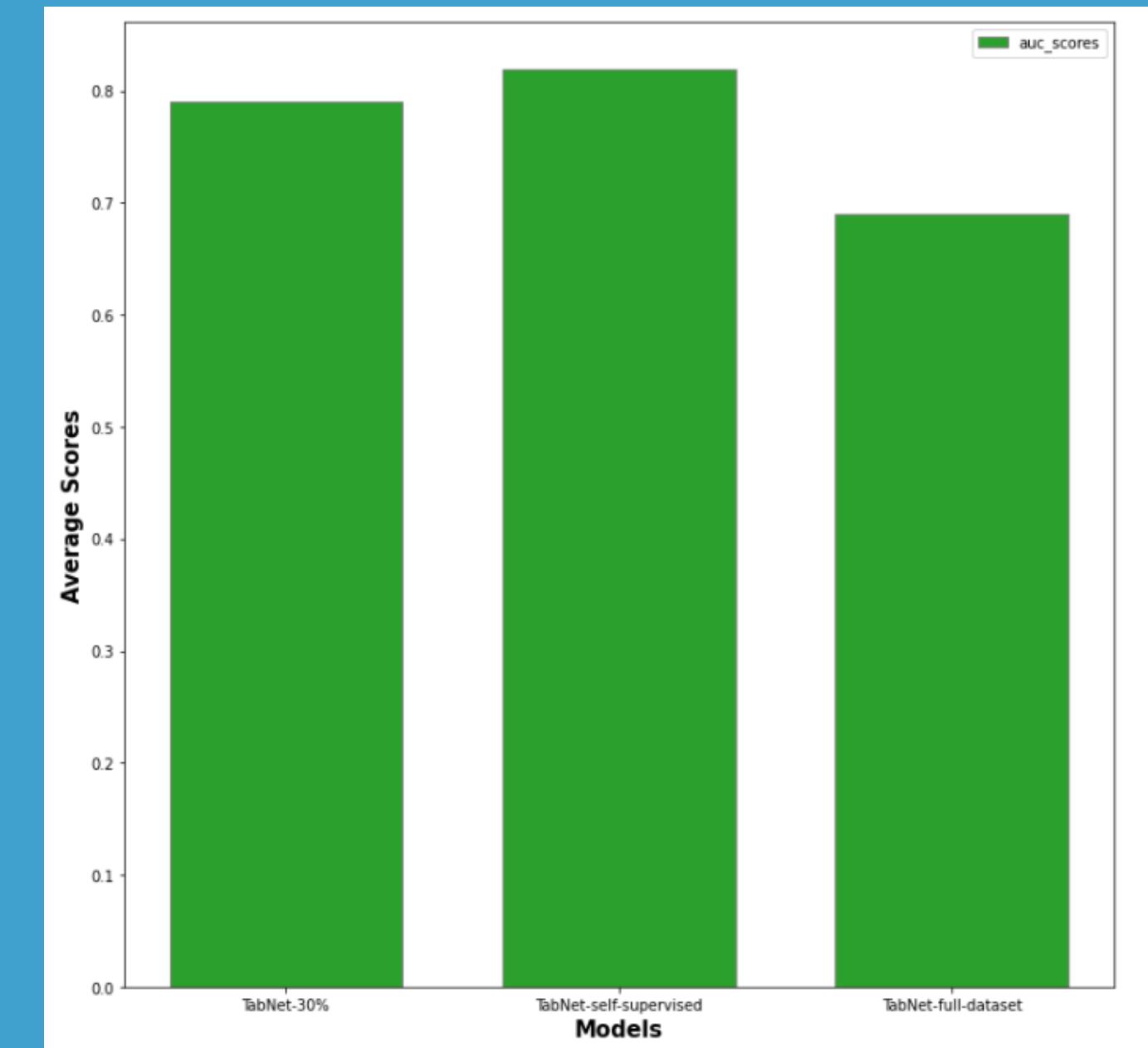
Train the TABNET classifier with  
30% of the dataset

02

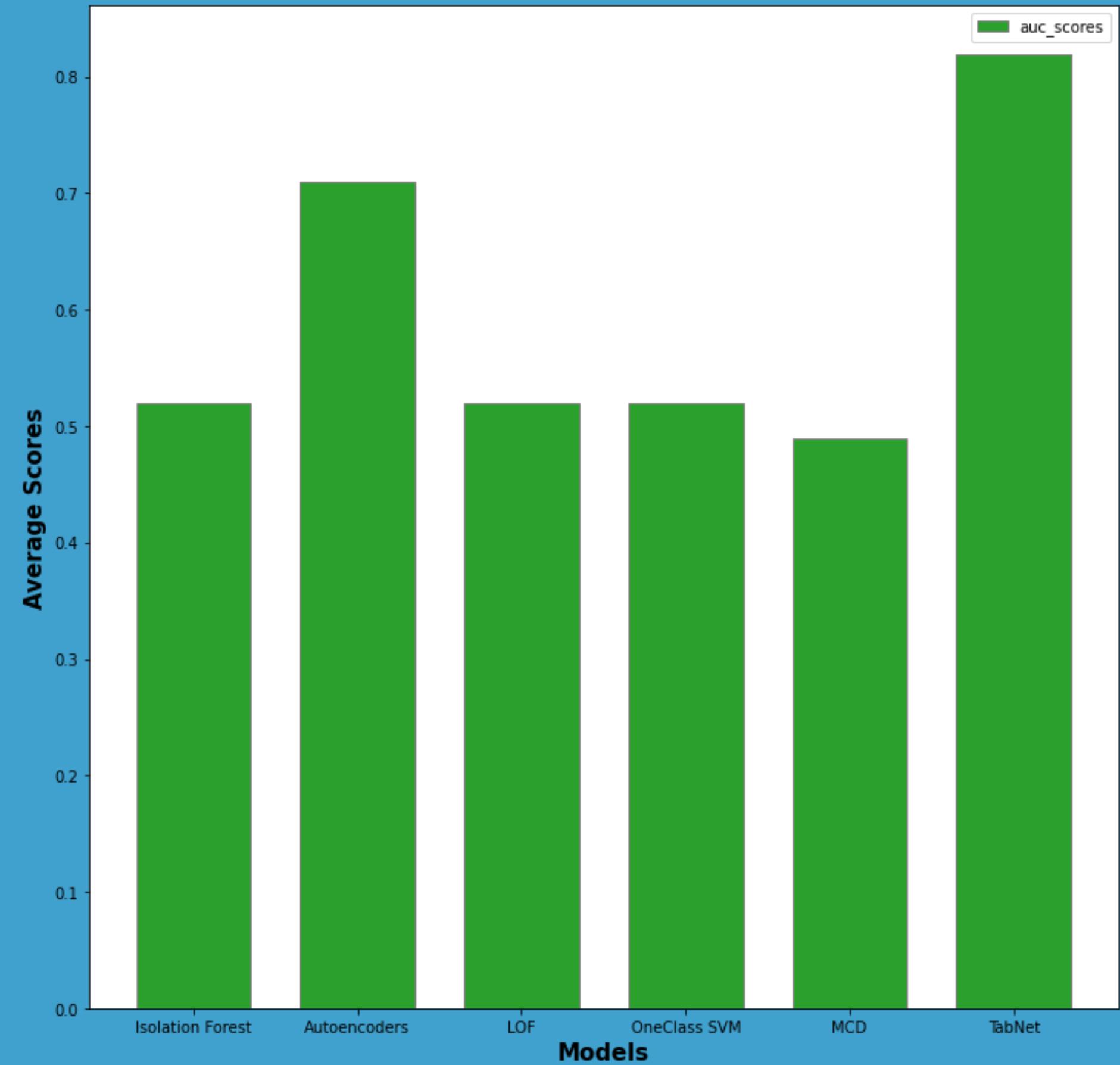
Train a TABNET with self supervised  
setting

03

Train a TABNET with complete  
dataset



# conclusion





# Thank You