Binary classification

Evgenii Karpuk

Data Scientist

+375 29 324 22 28

kek777@mail.ru

github.com/kek7777/Data_Analytics.git

Problem Definition

Predict the values of the target variable "Target"

Steps

01

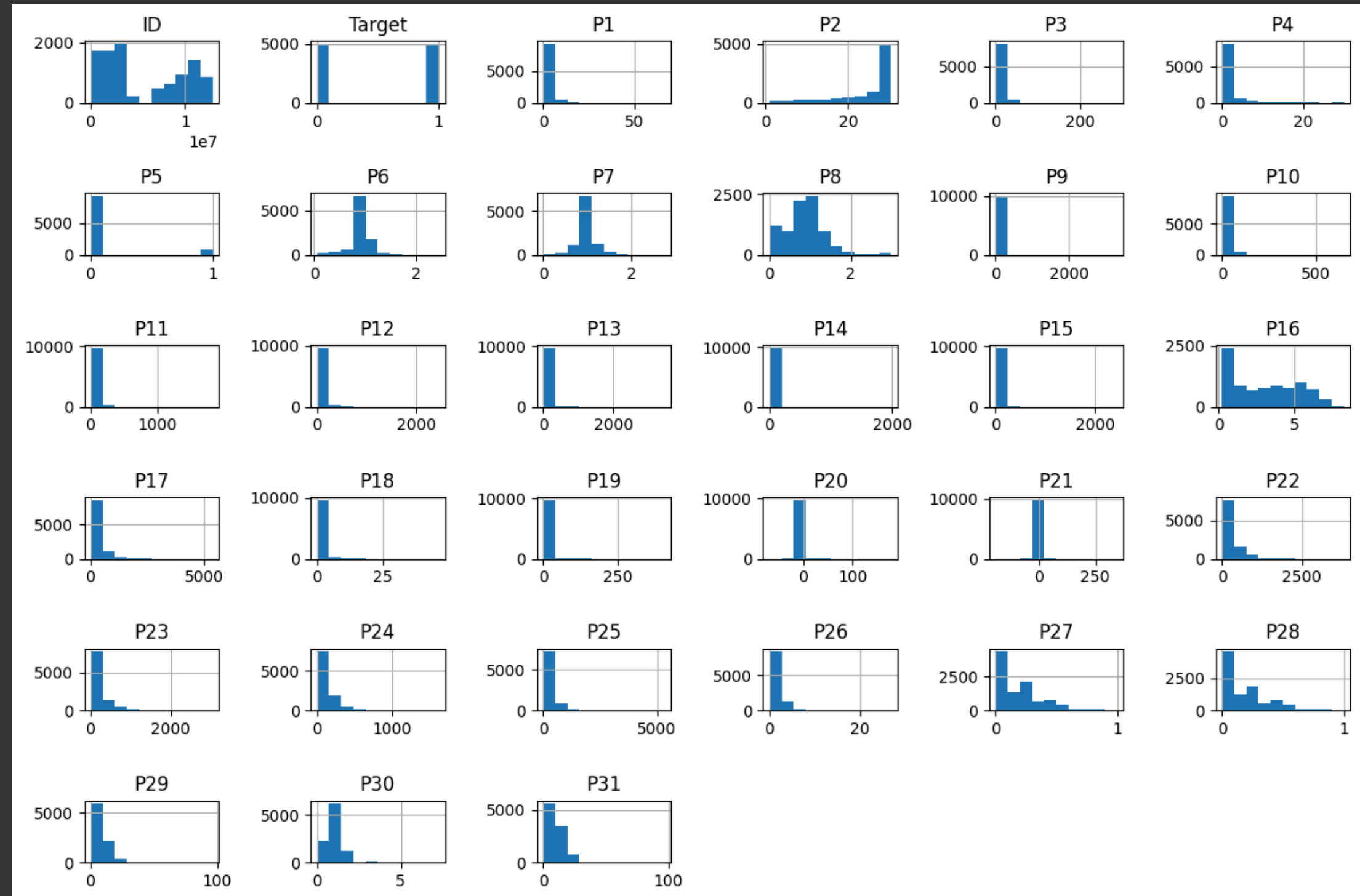Exploratory Data Analysis

02

Model Training

03

Forecasting

Exploratory Data Analysis
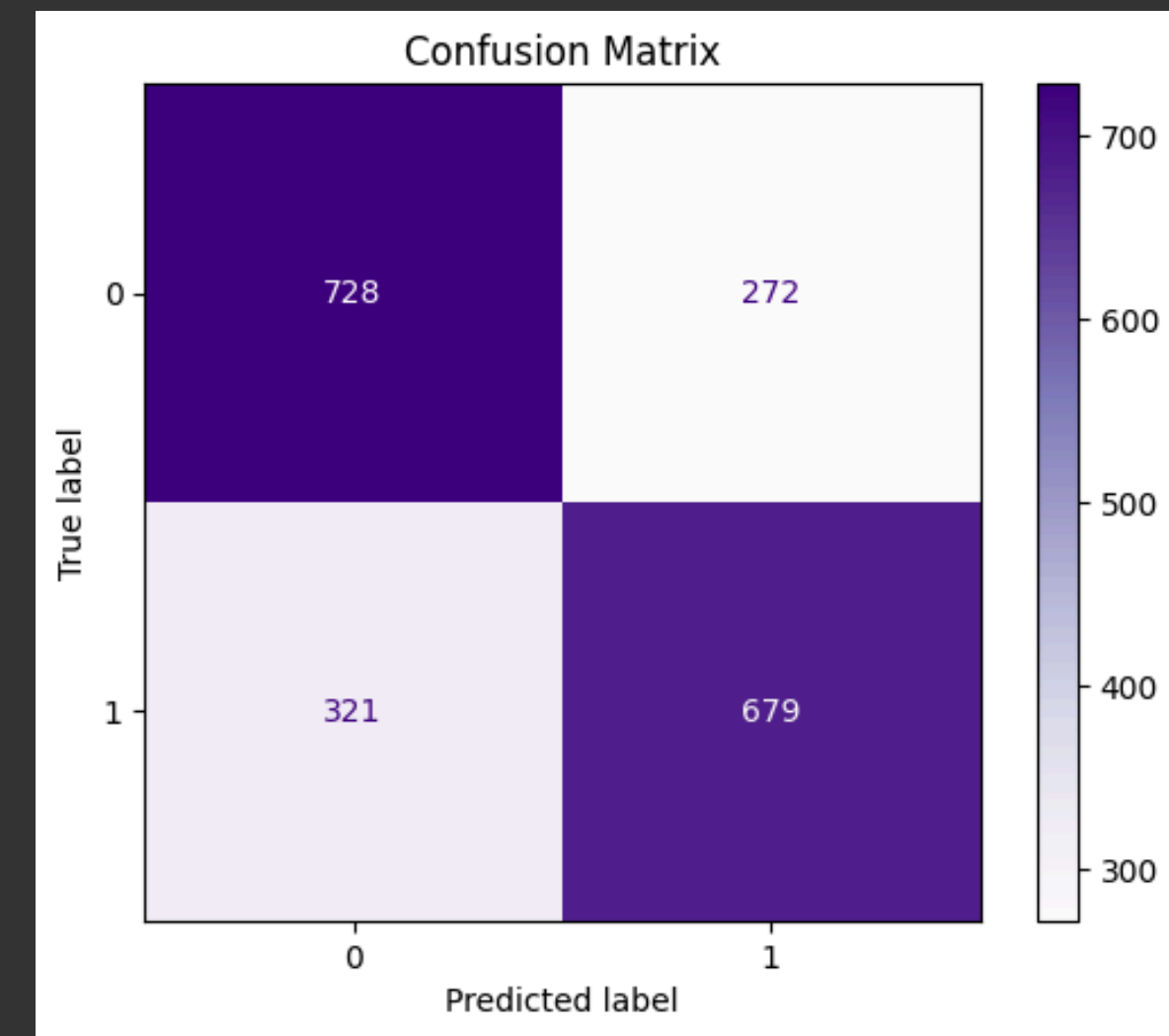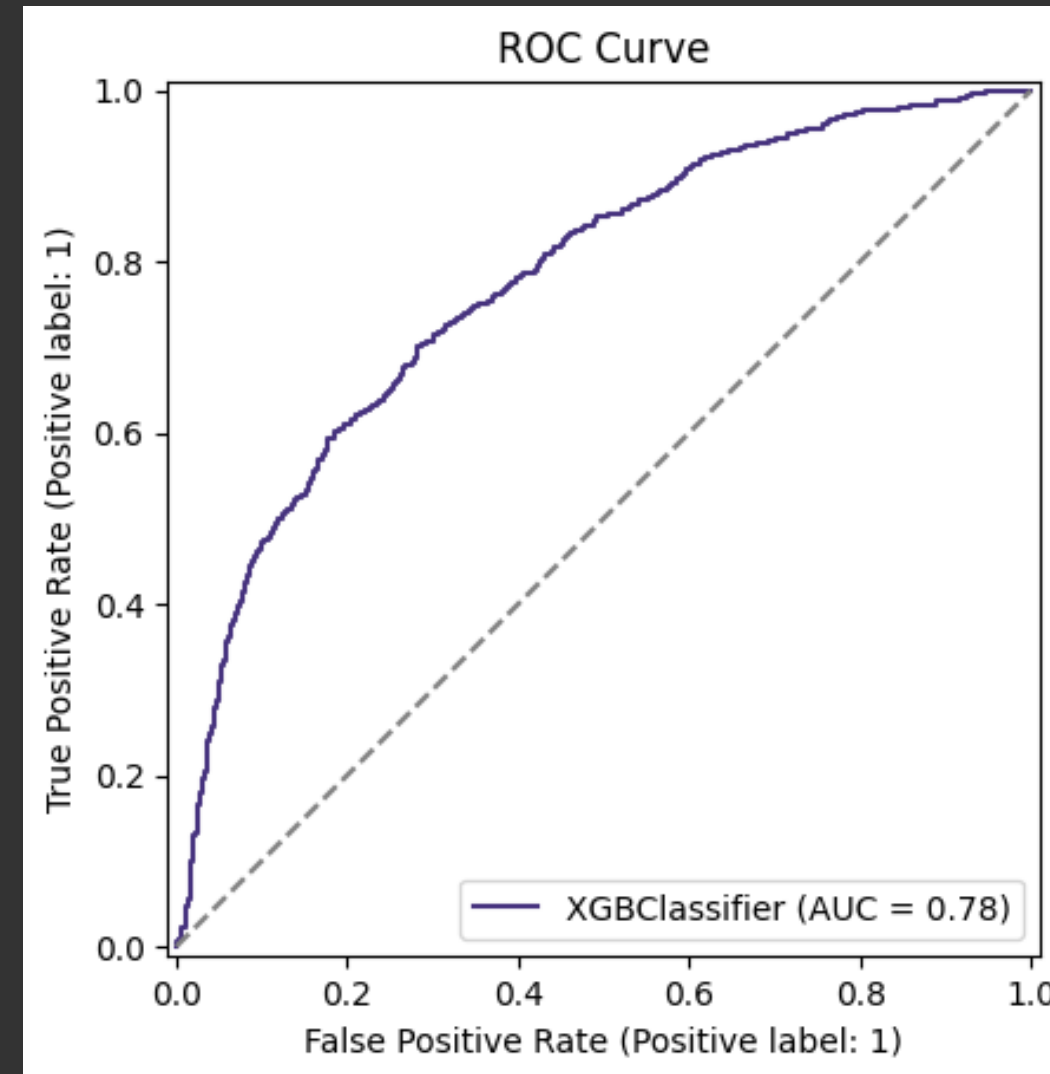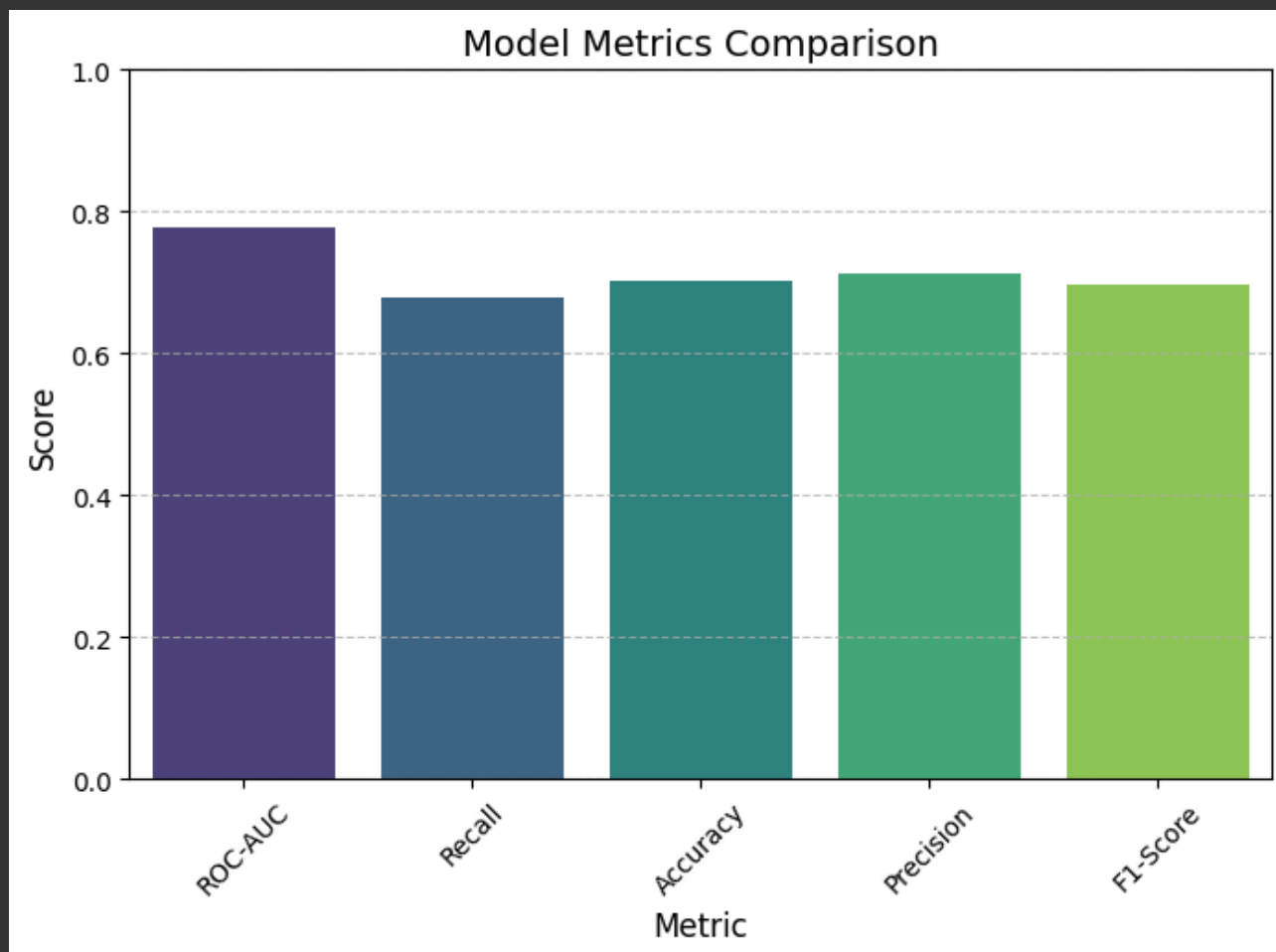
- **Dataset :**   10000

- **Features :**   31 (numerical)

- **Label :**   2 (1 or 0)

- **Characteristic:**
    structured
    balance class
    missing values
    outliers

Model training

Model : XGBClassifier

Train/Test dataset:
    80/20(%)

Confusion Matrix

ROC Curve

Model Metrics Comparison

Metrics :   AUC - 0.78
Precision - 0.71
F1 - 0.69

Forecasting

Dataset: 20000
Class 1: 4492
Class 0: 15508
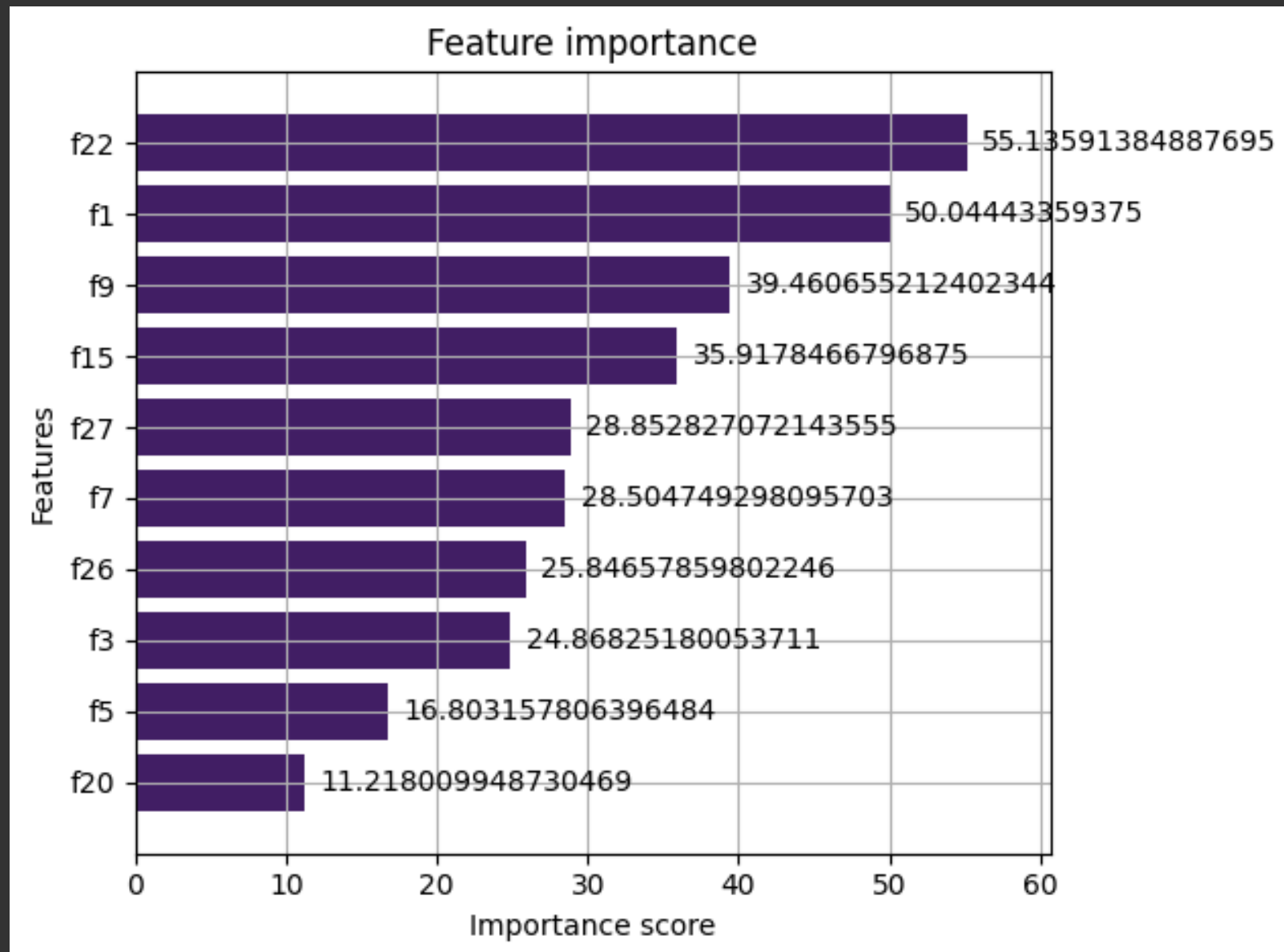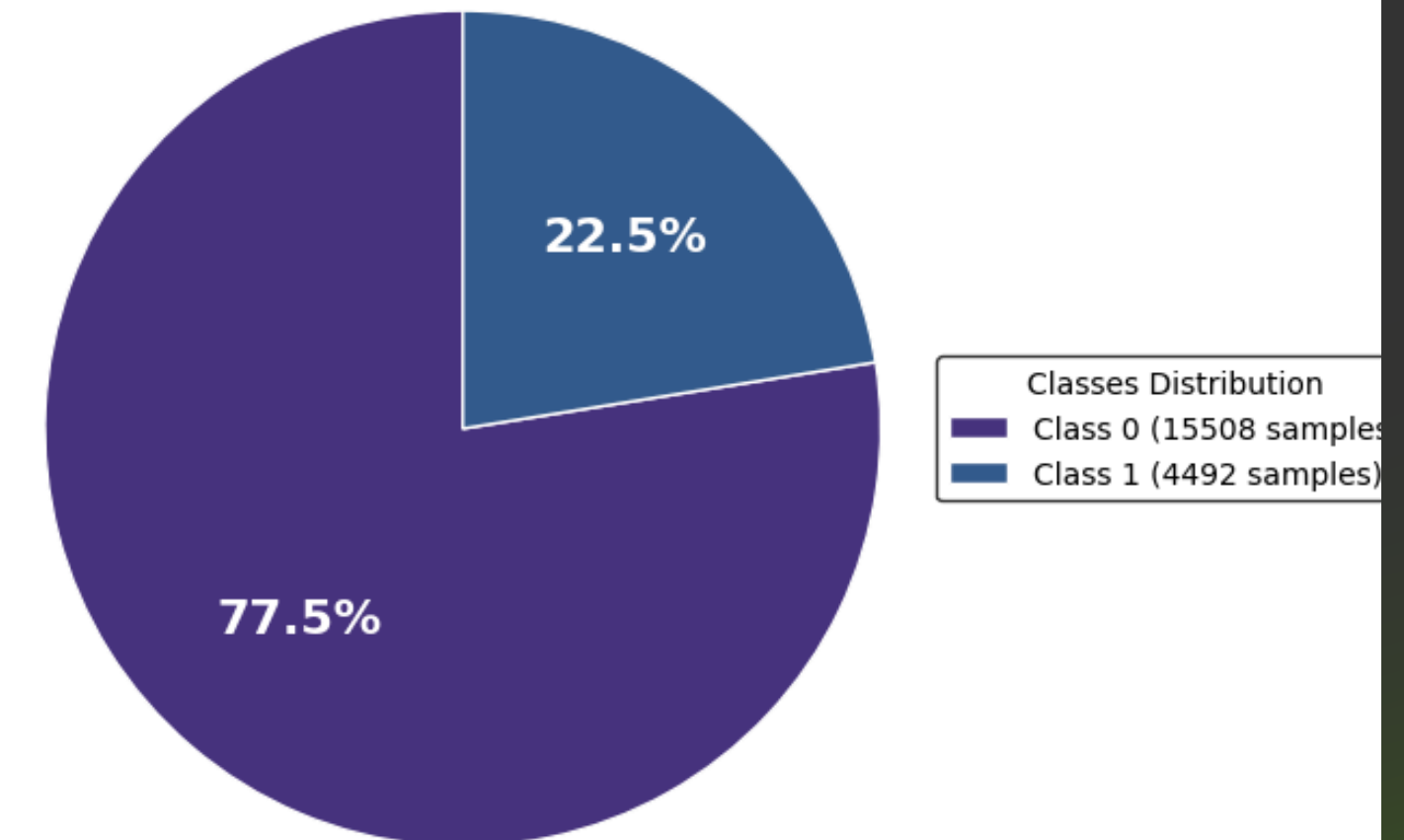
Importance features： P22

P1

P9