

# **Fraud Detection**

**Di Kang, Fang Li**

# Motivation

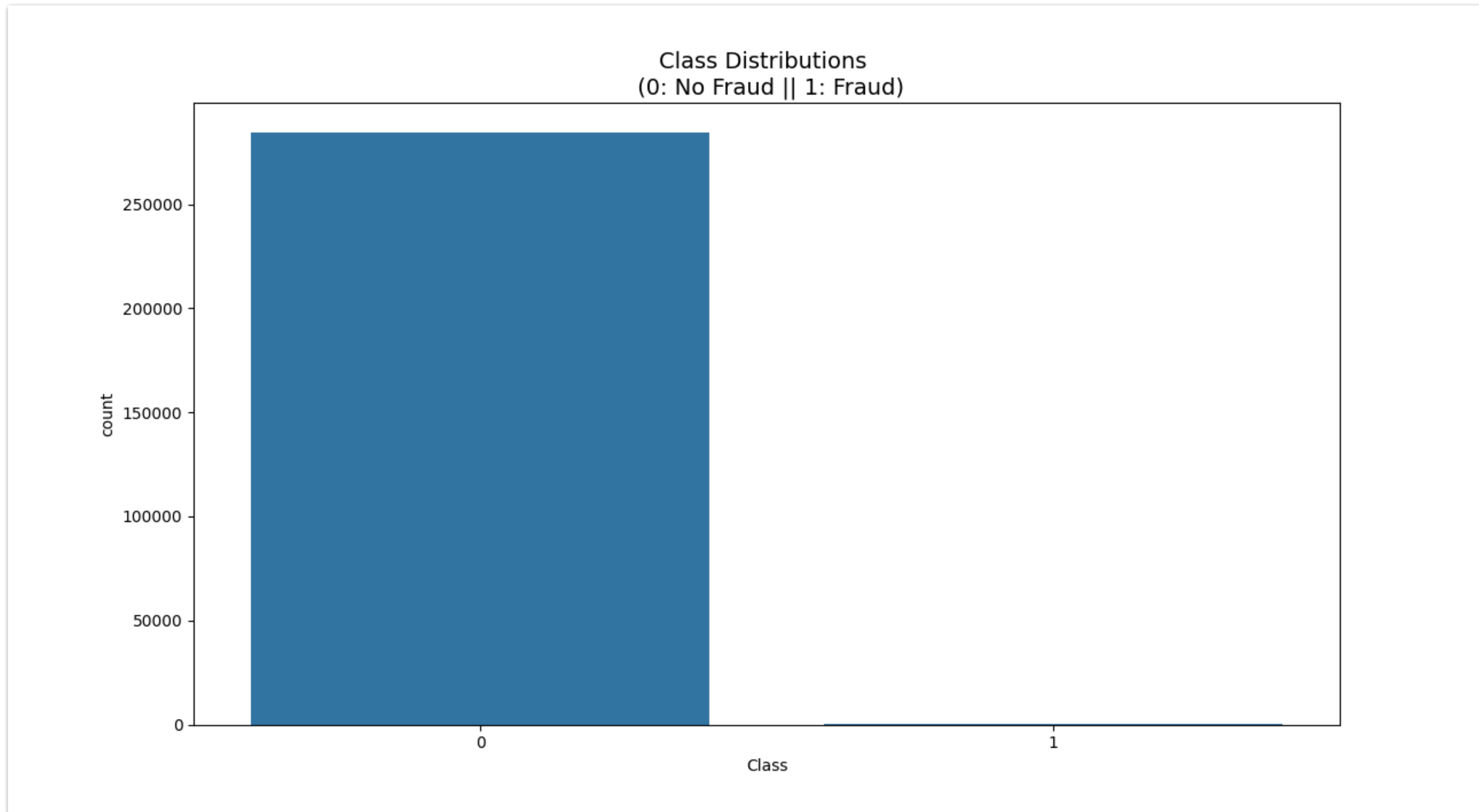
- Fraud Detection is important as customers should not be charged for items that they did not purchase.

# Data Description

- The Data set includes transactions made by credit cards in September 2013 by European cardholders.
- There are 492 frauds out of 284,807 transactions. The positive class (frauds) account for 0.172% of all transactions.
- There are 30 features including time and transaction amount and 28 features derived from the original data masked by PCA.

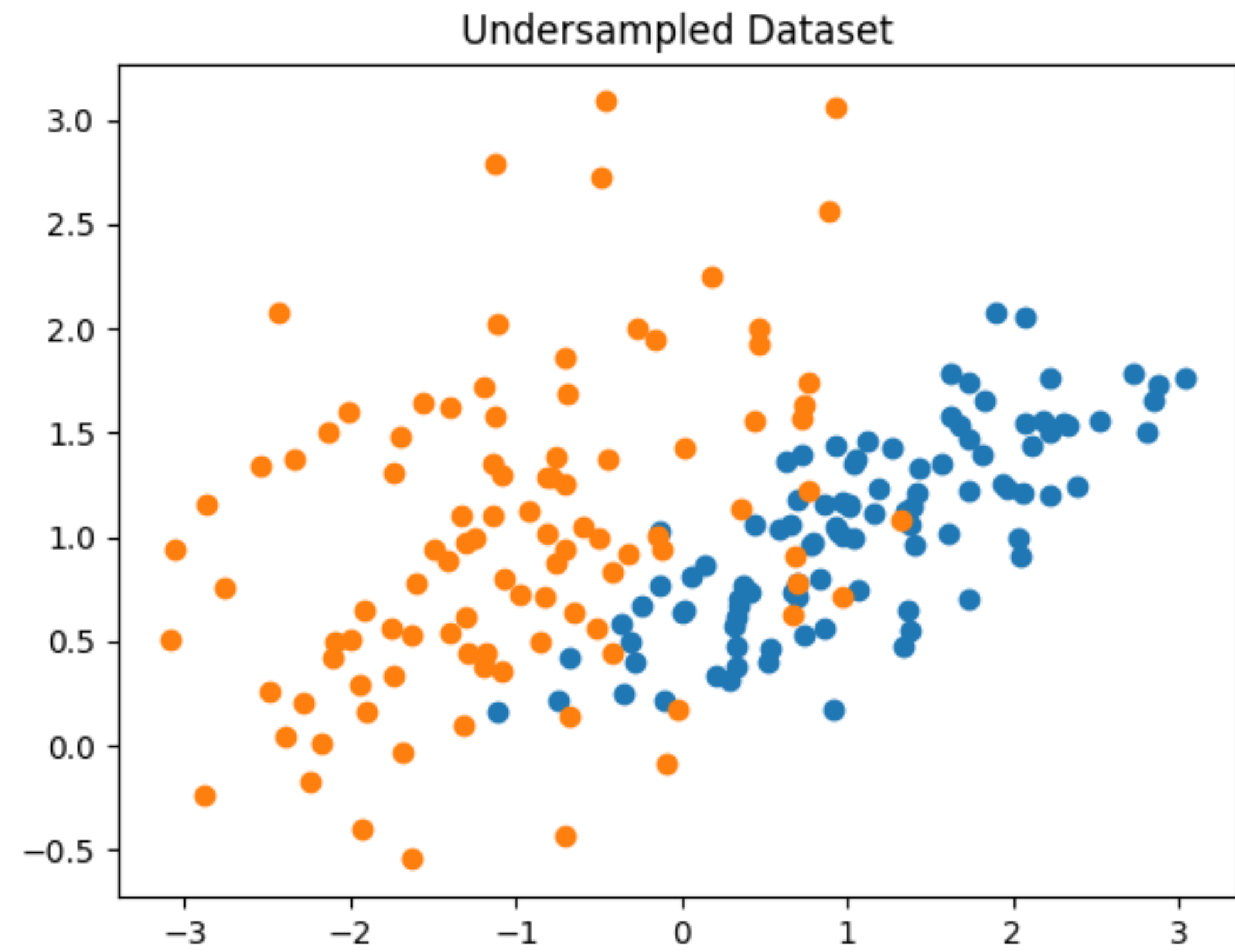
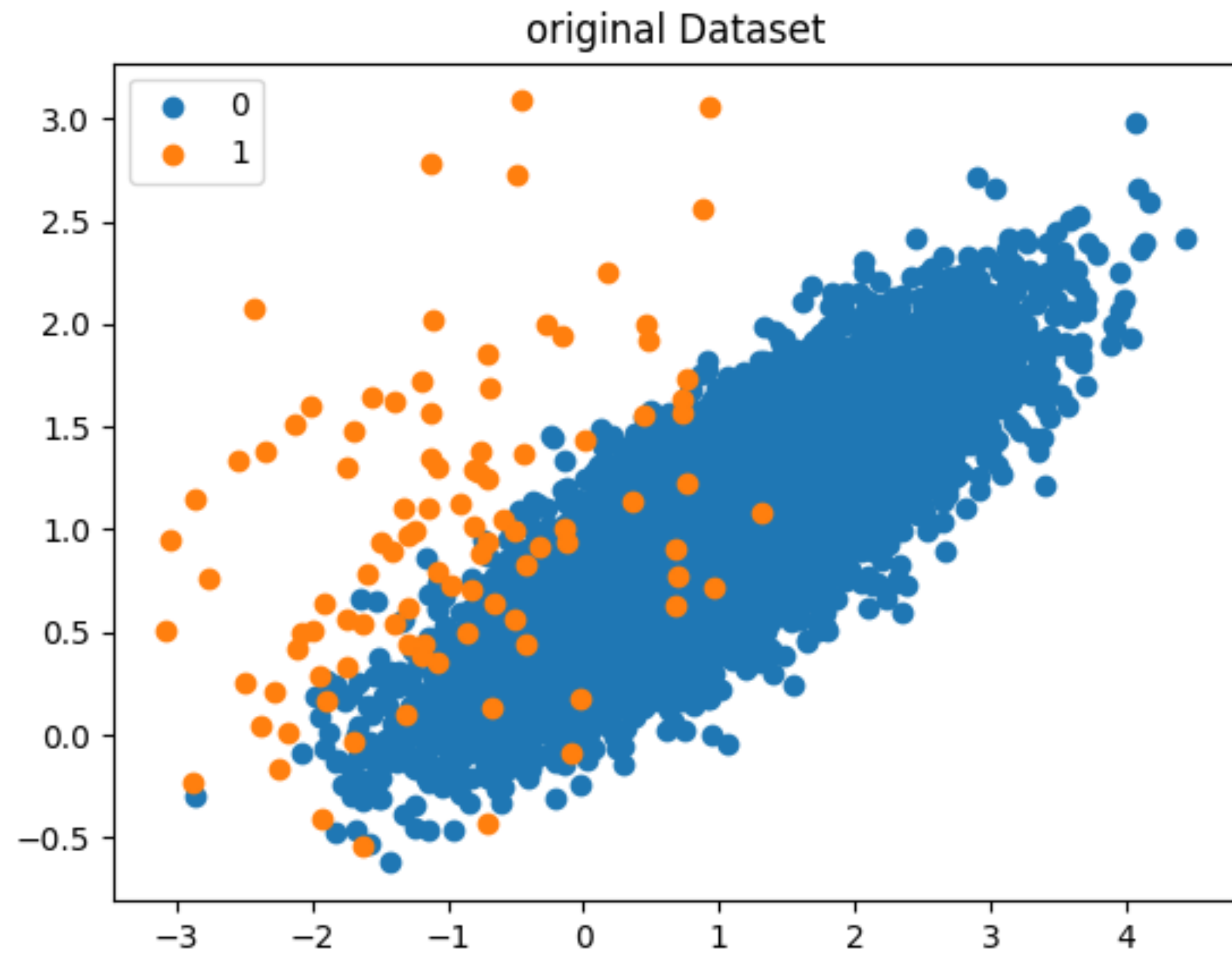
# Data Preprocessing

## Imbalanced Data



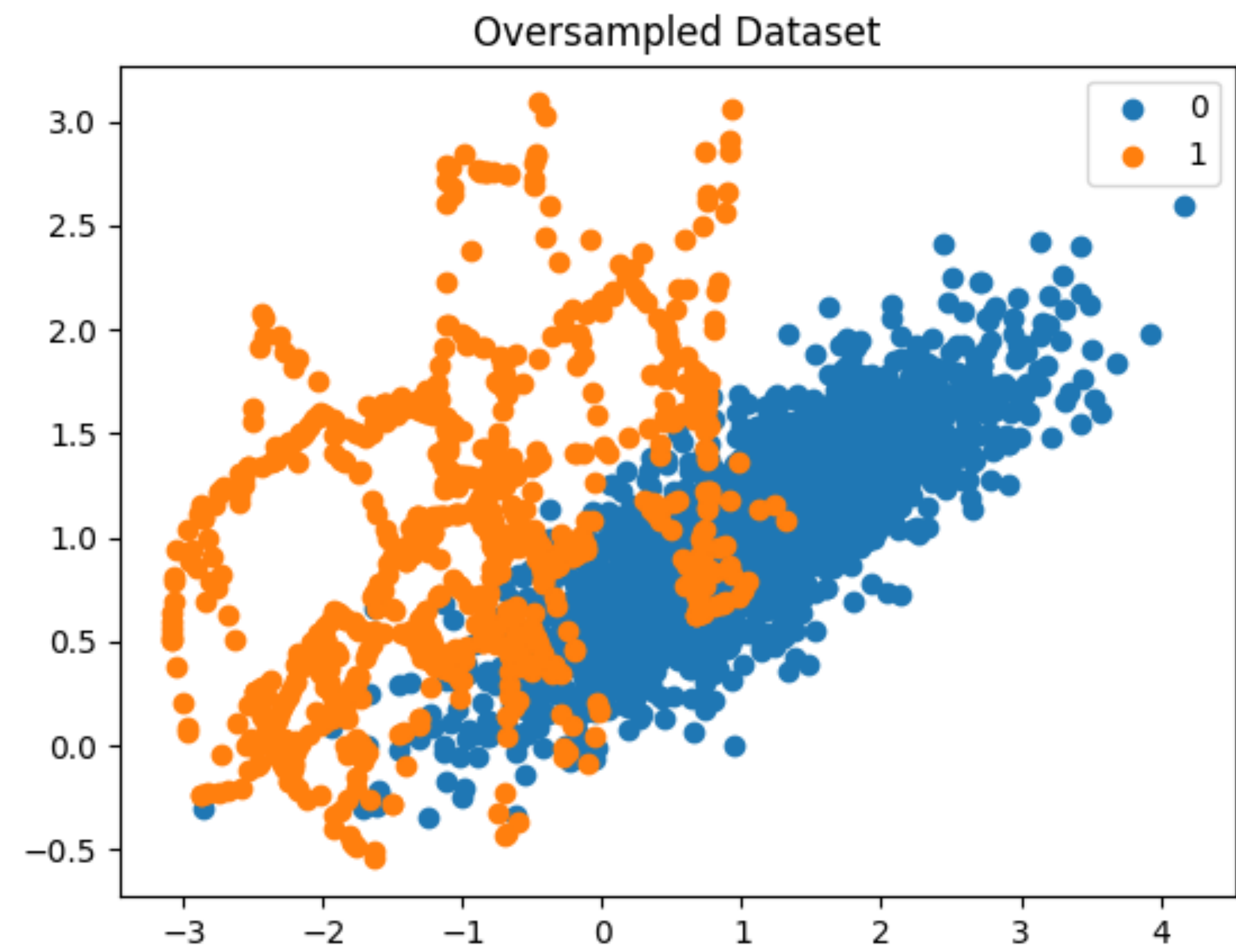
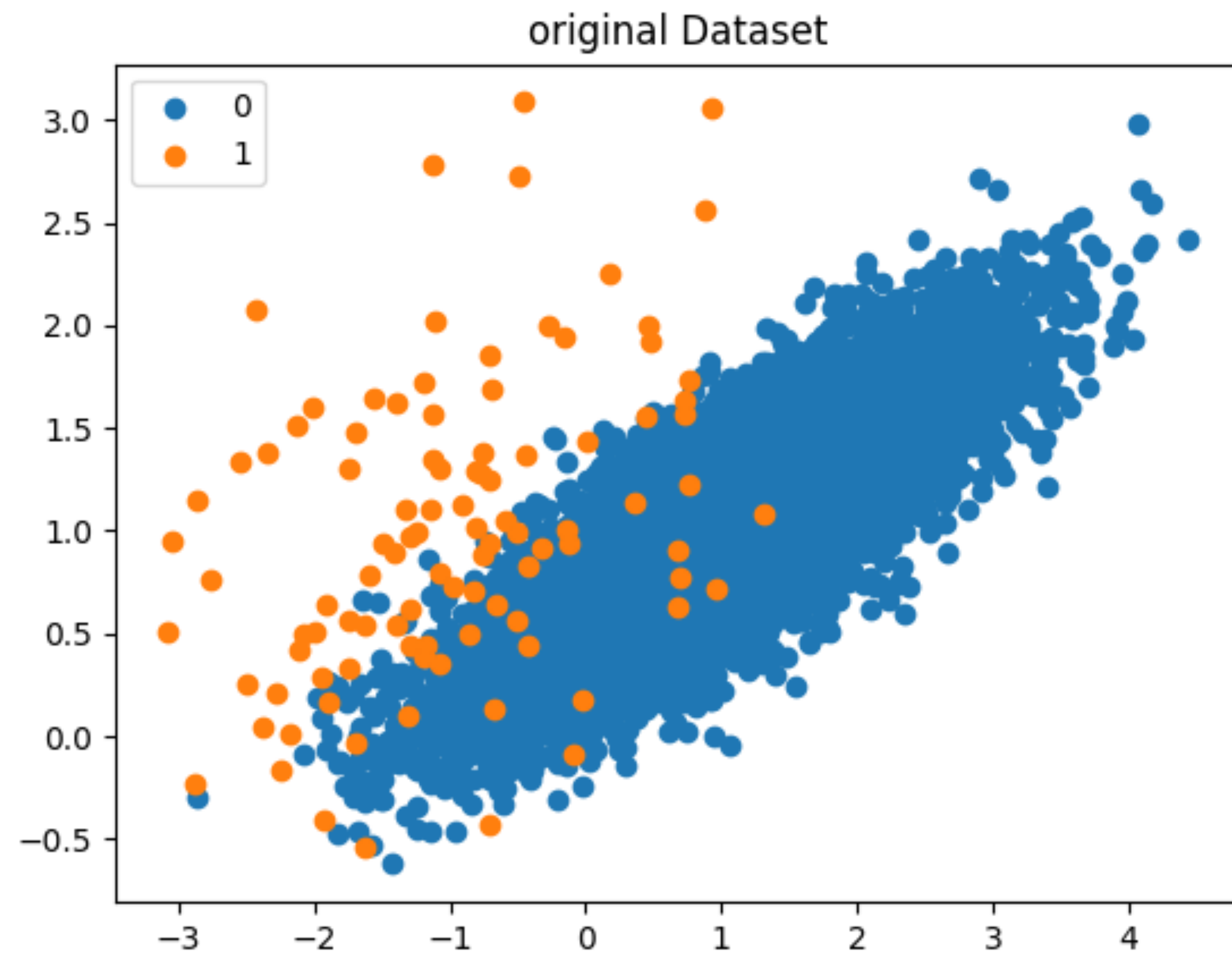
# Data Preprocessing

## Undersampling



# Data Preprocessing

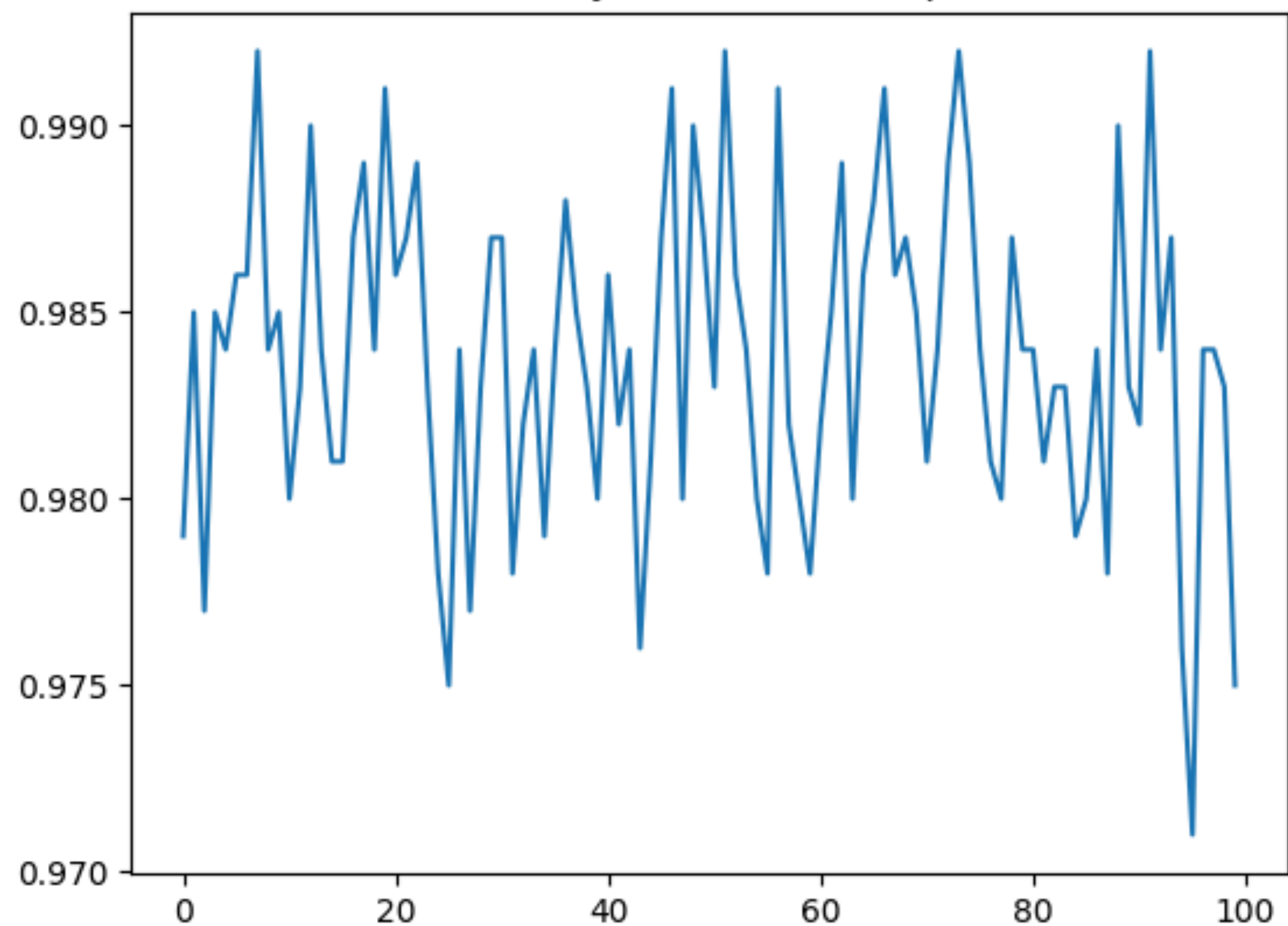
## Oversampling - SMOTE



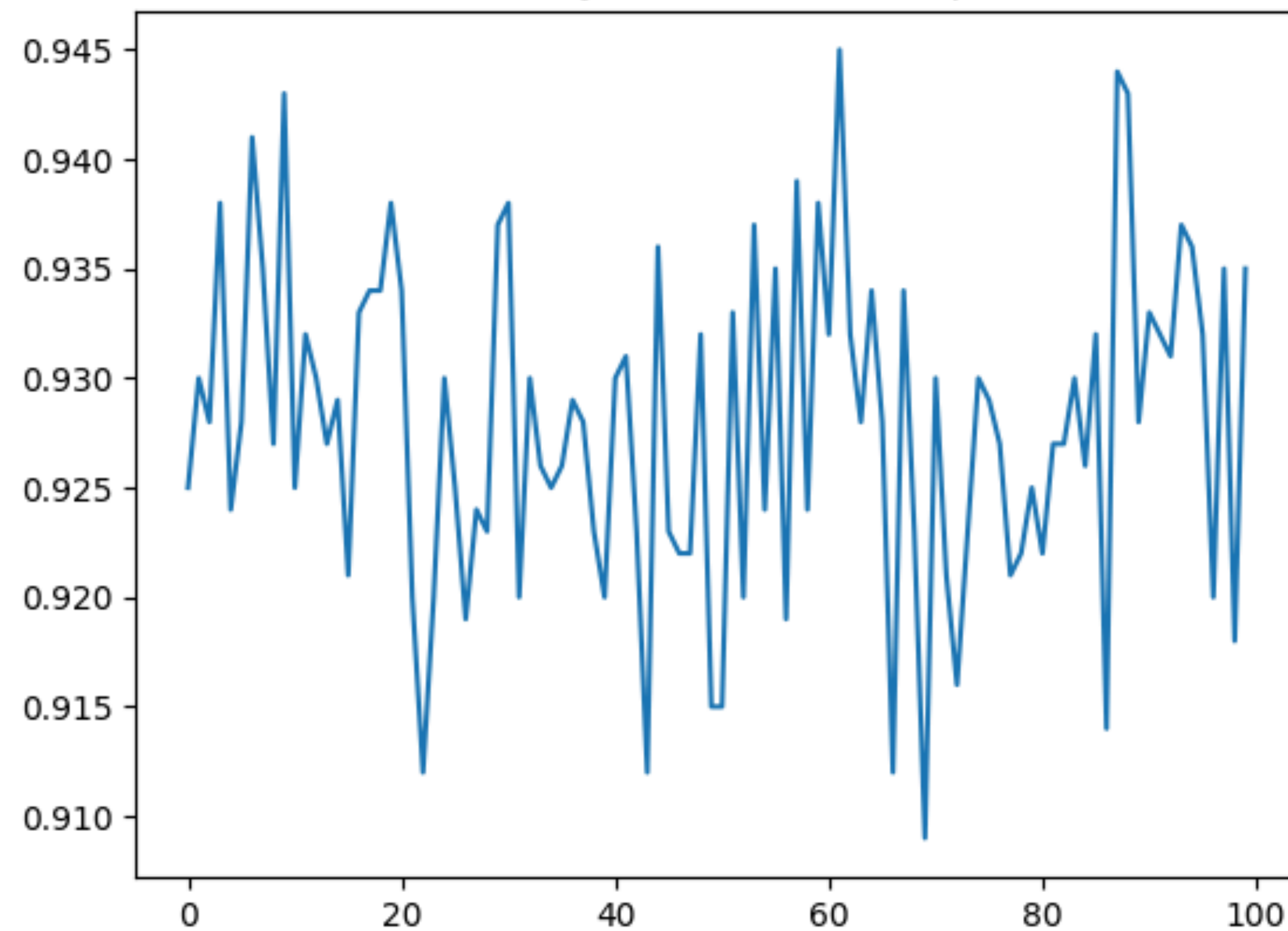


# Performance

Accuracy: the whole Sample



Accuracy: the Positive Sample





# Further work

- **Optimize the network structure**
- **Tune the hyper-params**
- **Validate the model**