

TAIWANESE COMPANY

BANKRUPTCY PREDICTION

PROJECT

PRESENTER :
HU YVAN , NOUADJE ENGE

Project Objective

Develop a predictive model to identify the likelihood of bankruptcy for Taiwanese companies



TABLE OF CONTENTS

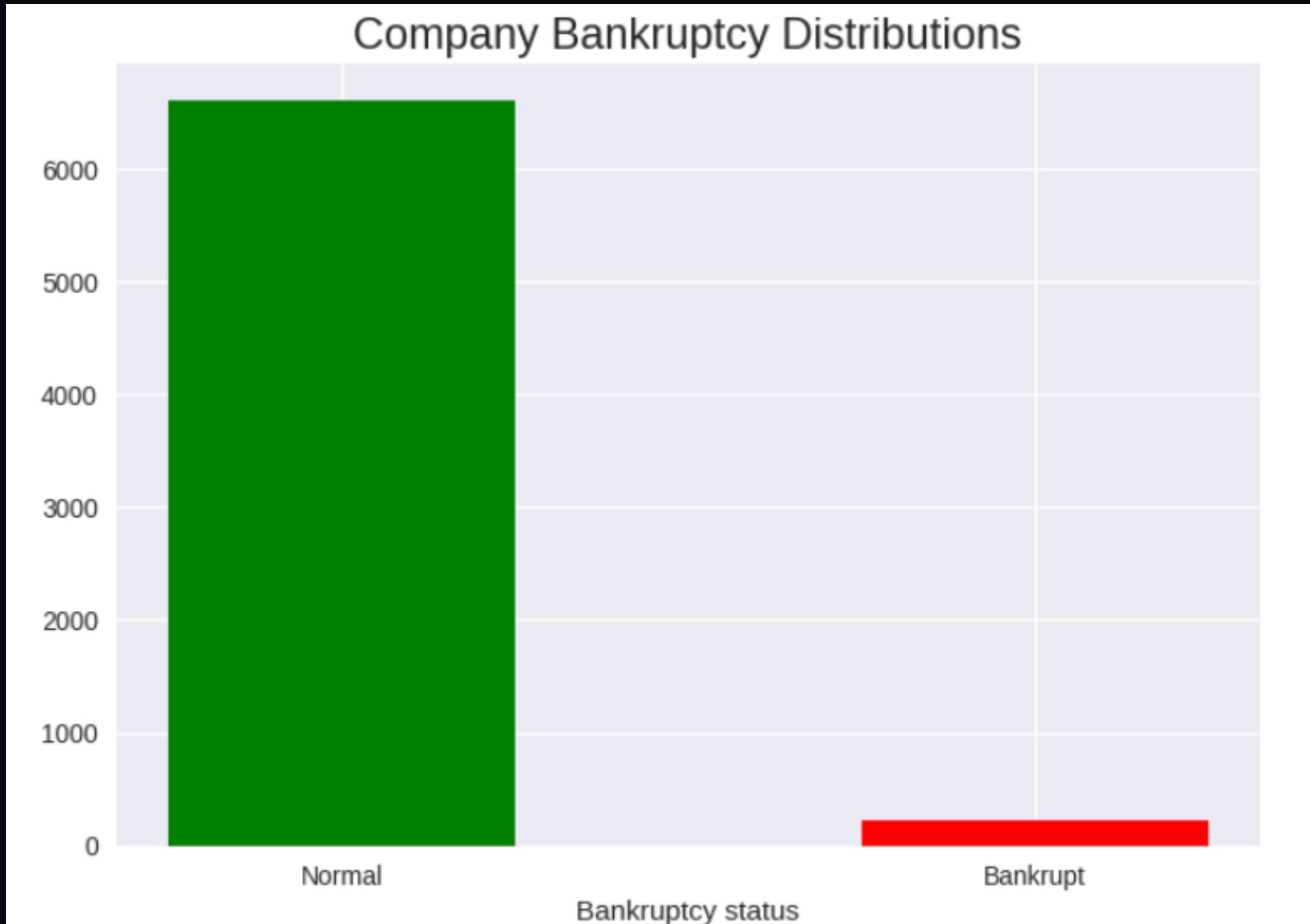
01 | Problem
Explanation

02 | Proposed
Solutions

03 | Conclusion



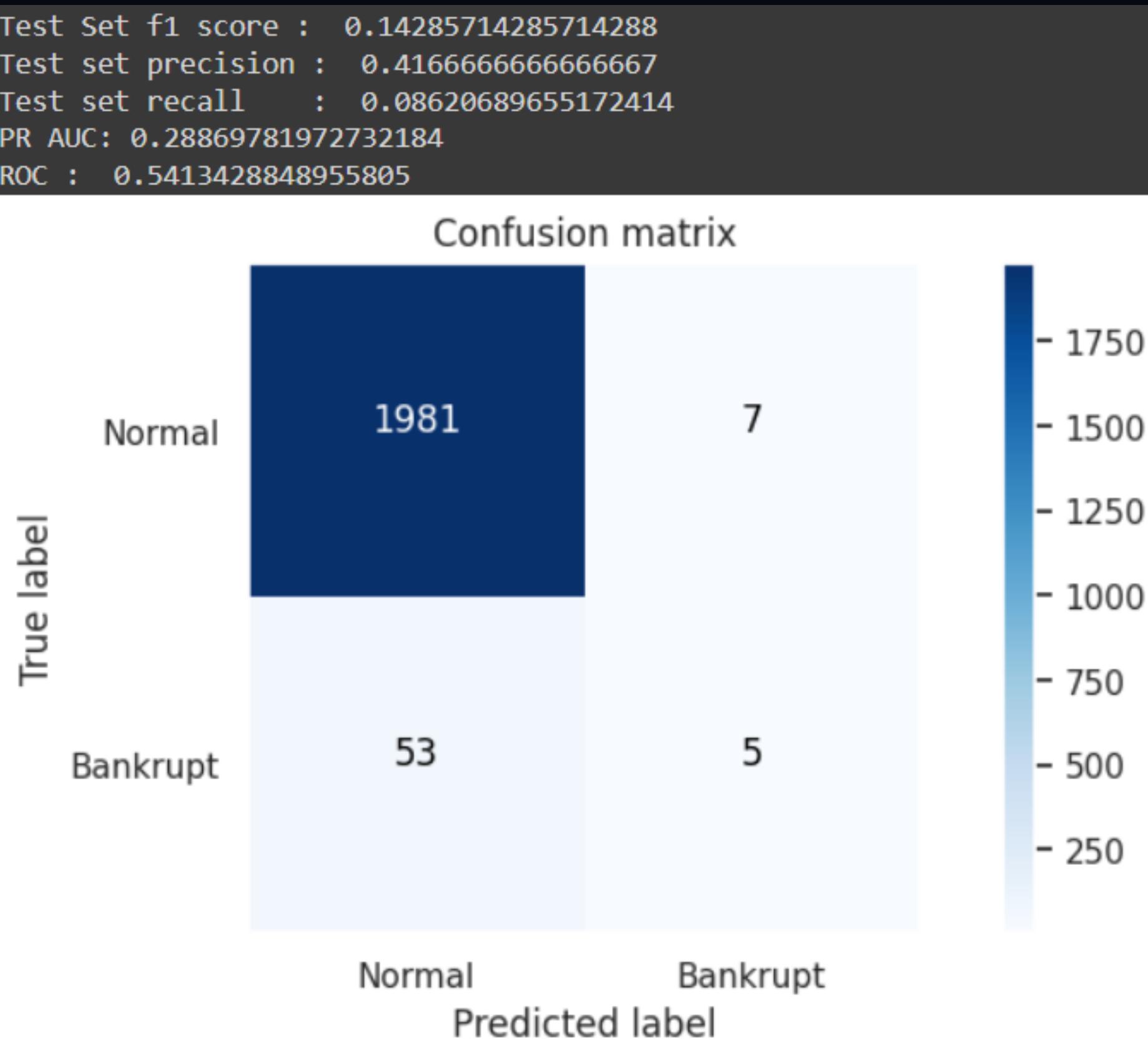
1. Problem Explanation



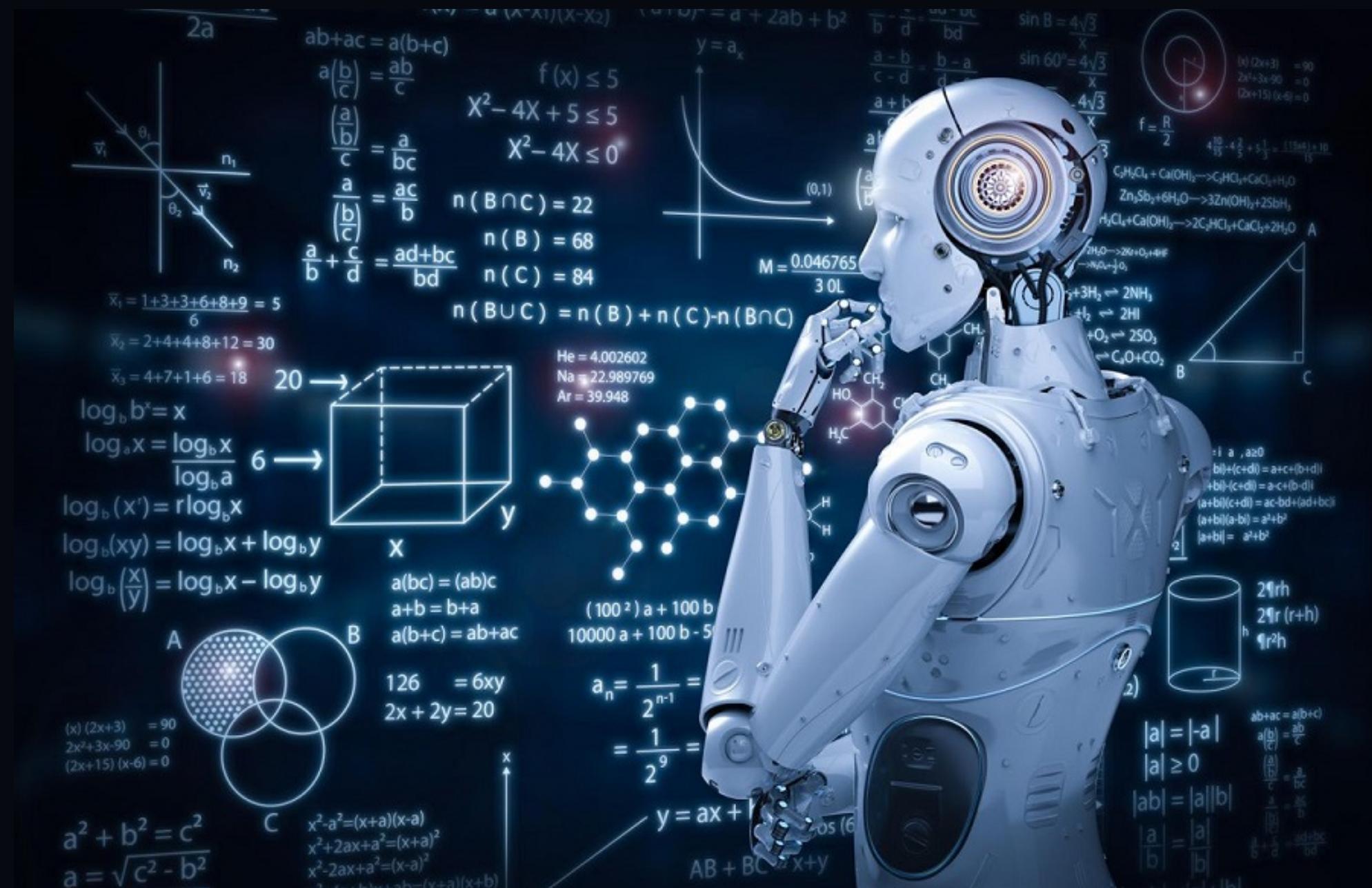
96.77% Normal
3,22% Bankrupt

Problem Explanation

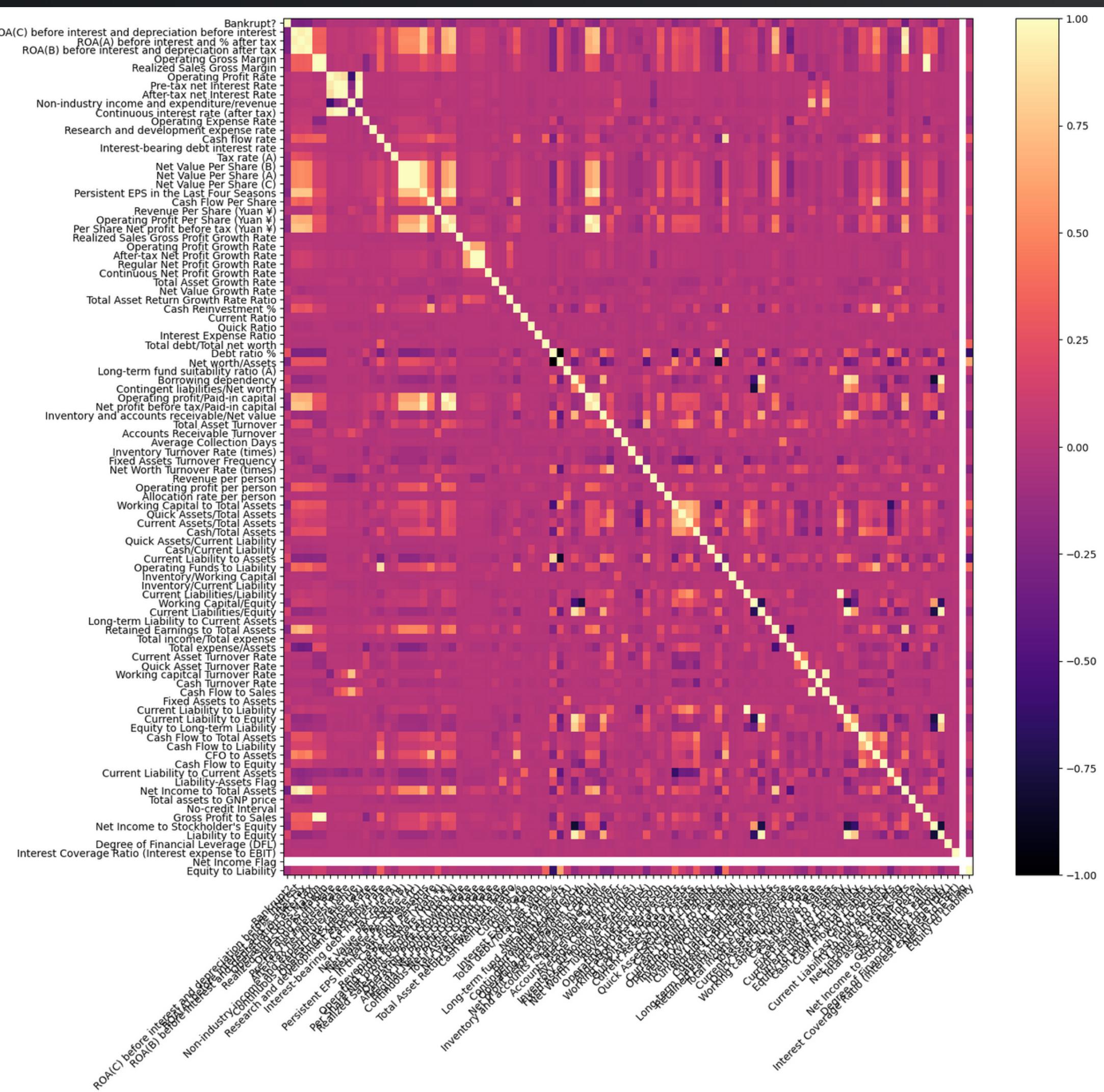
```
Test Set f1 score : 0.14285714285714288  
Test set precision : 0.4166666666666667  
Test set recall : 0.08620689655172414  
PR AUC: 0.28869781972732184  
ROC : 0.5413428848955805
```



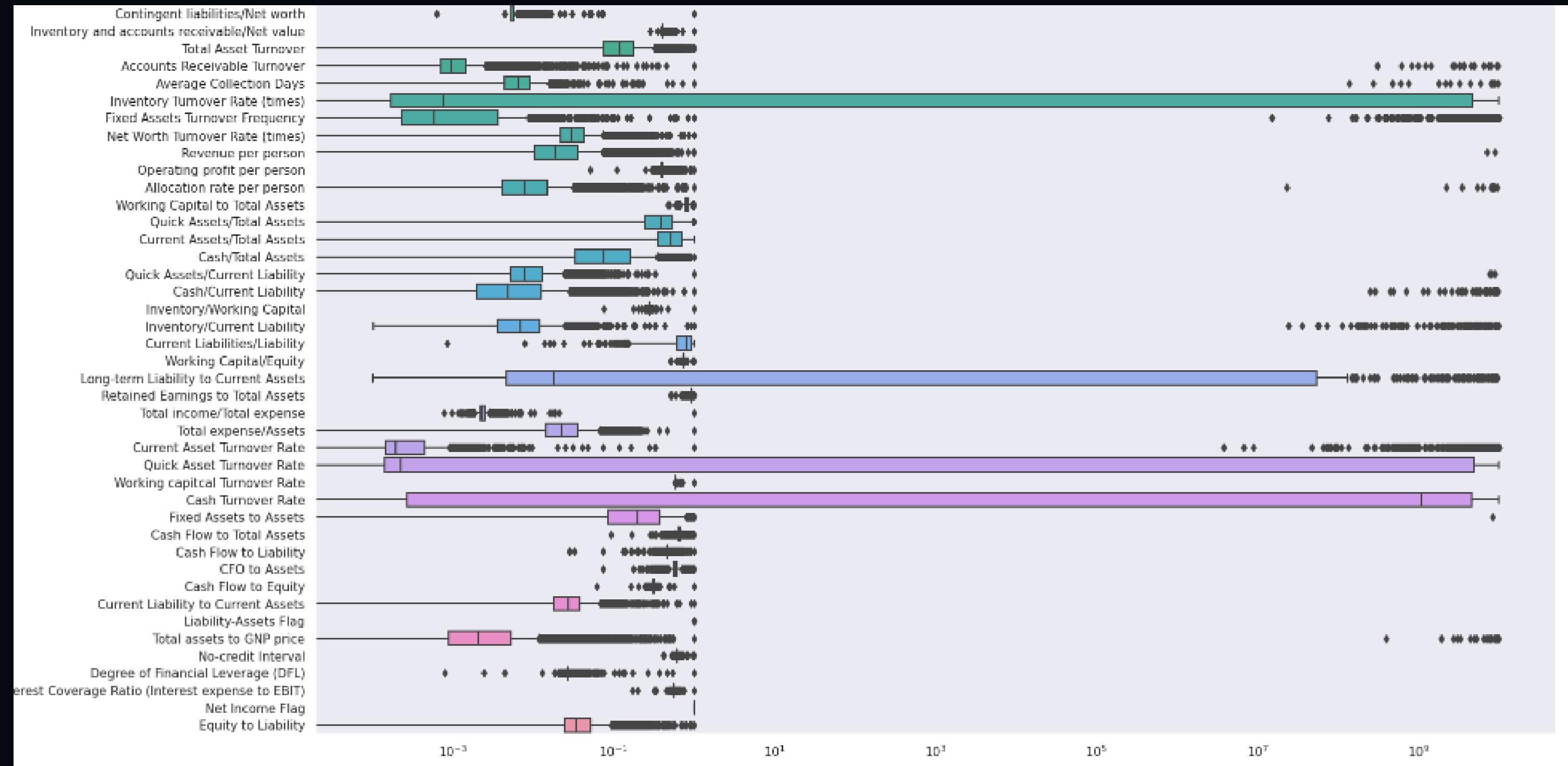
02. Proposed Solutions



Exploratory Data Analysis



Feature Boxplot



Data Preprocessing

DATA CLEANING

$$\text{Mean} = (10+19+12)/3 = 13.666$$

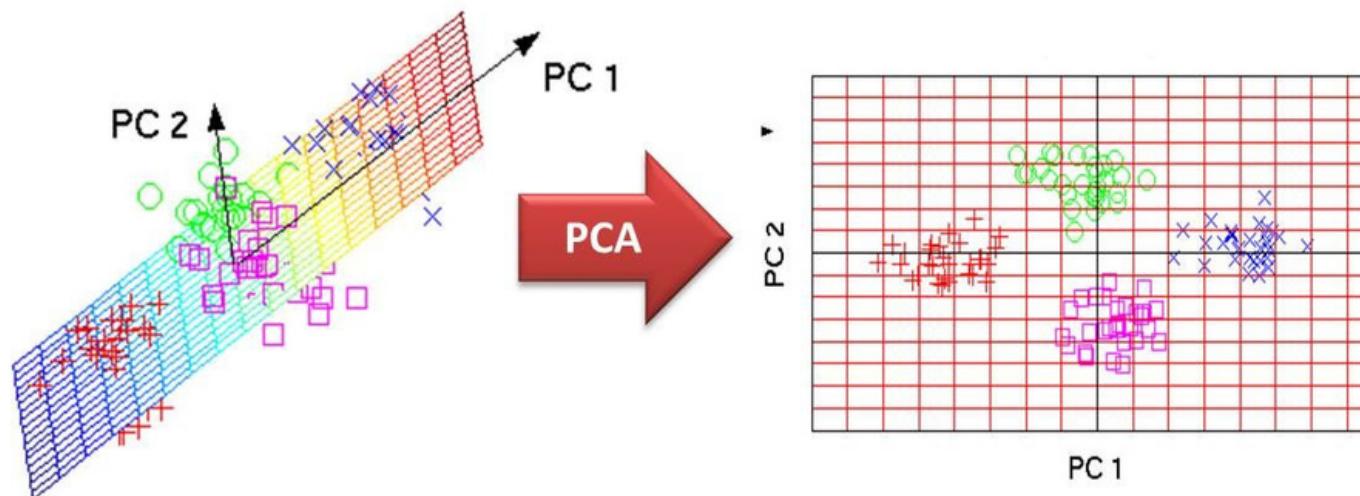
Column-1	Column-2
A	10
B	19
A	NaN
A	12



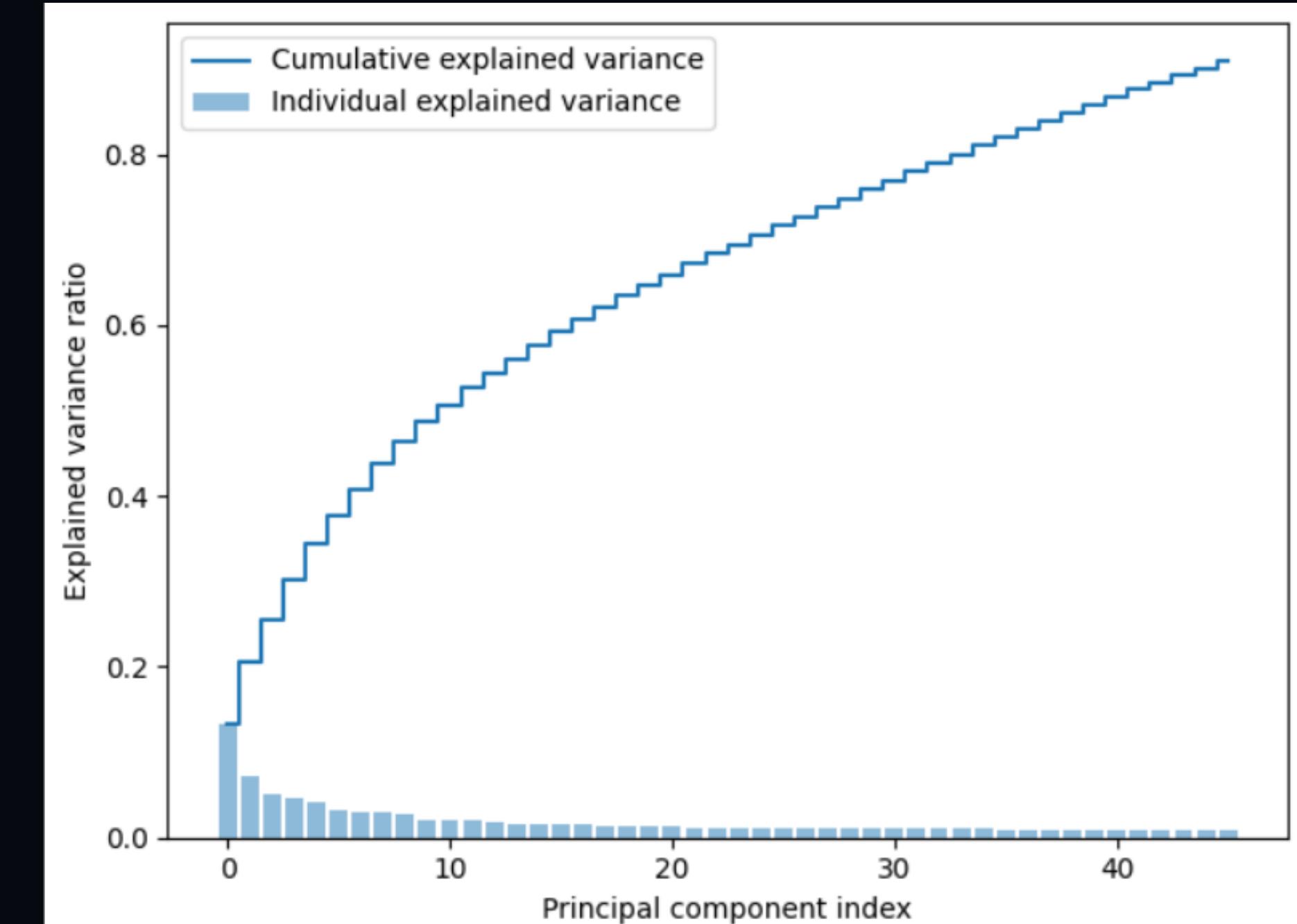
Column-1	Column-2
A	10
B	19
A	13.6666
A	12

PRINCIPAL COMPONENT ANALYSIS

Dimensionality Reduction & Principal Component Analysis

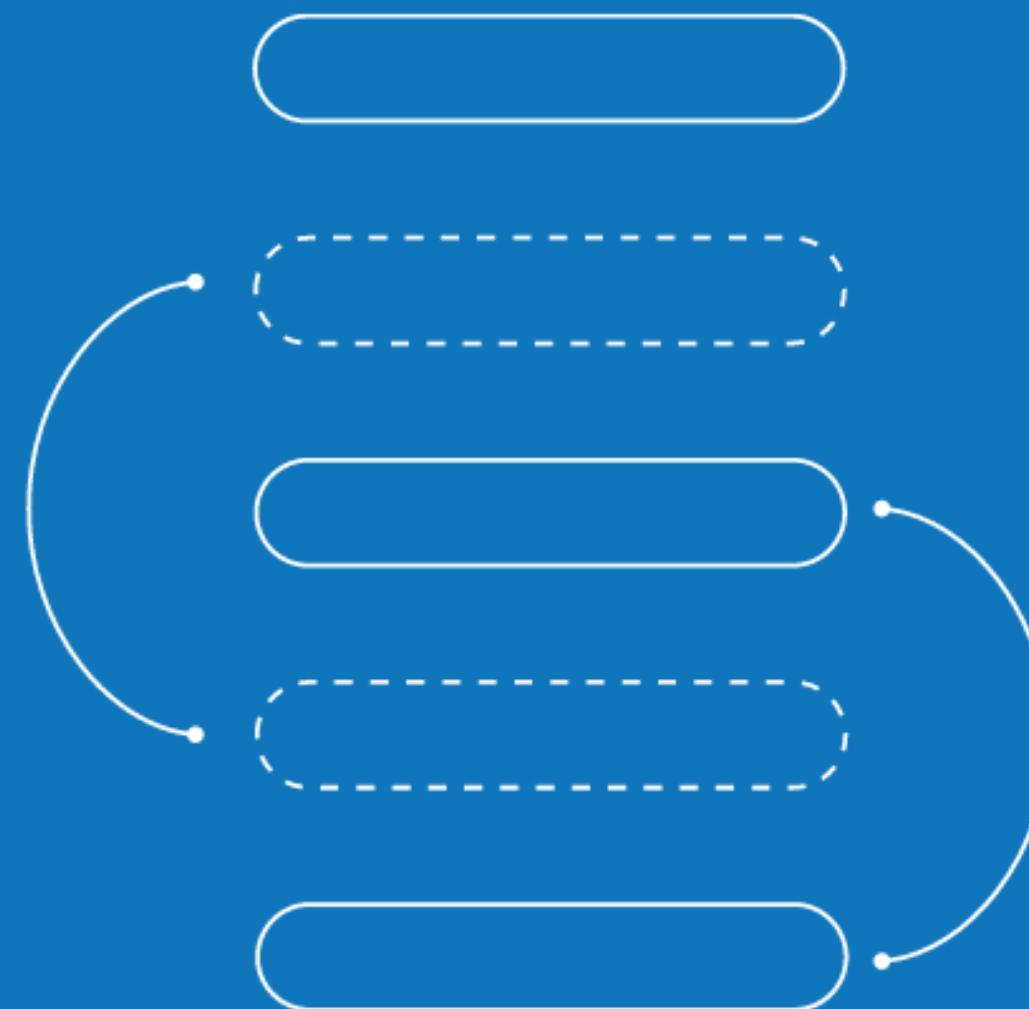


48% of features explain
>90% of the variance
in the dataset



RECURSIVE FEATURE ELIMINATION

RFE - Scikit-learn



Train a machine learning model

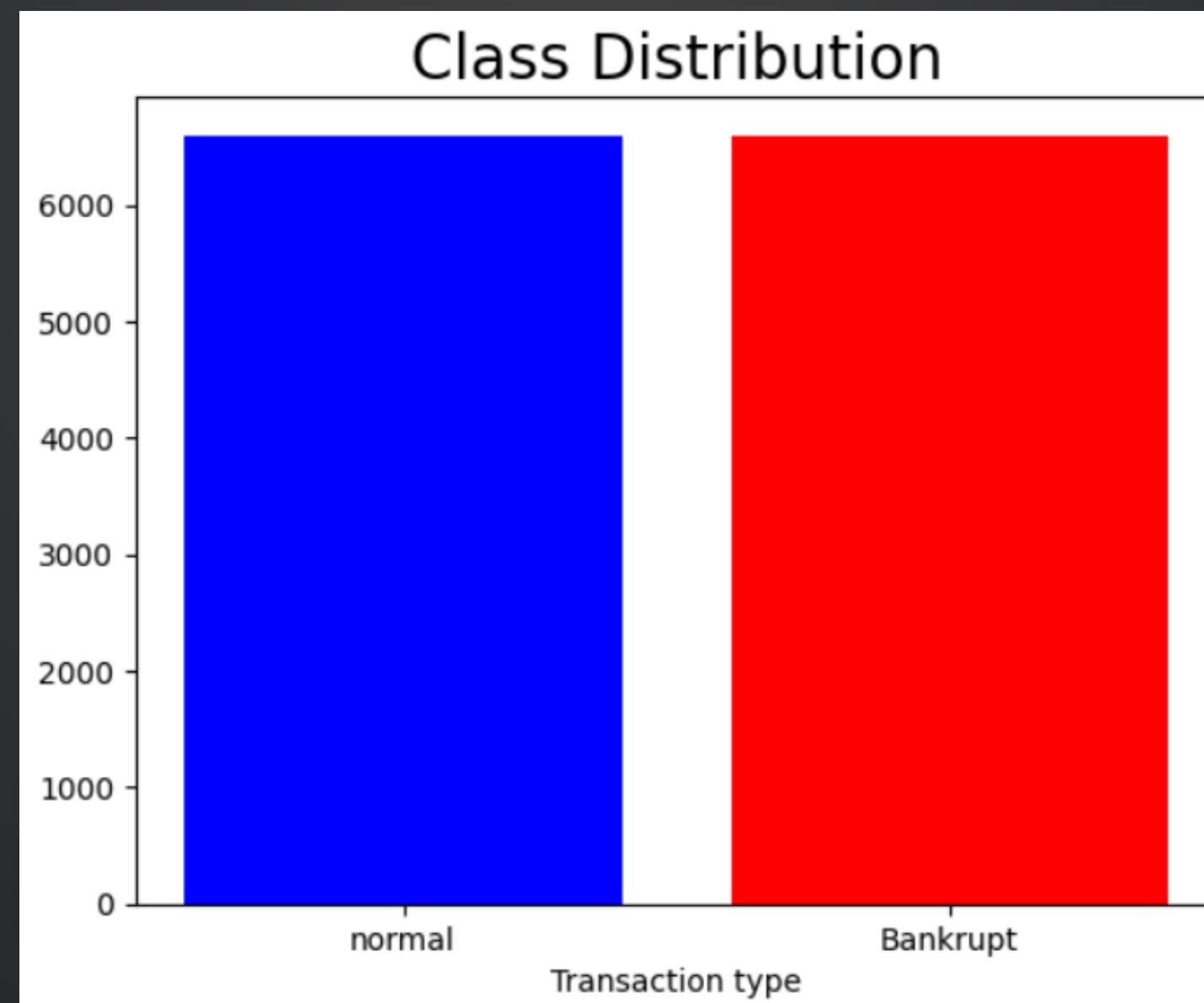
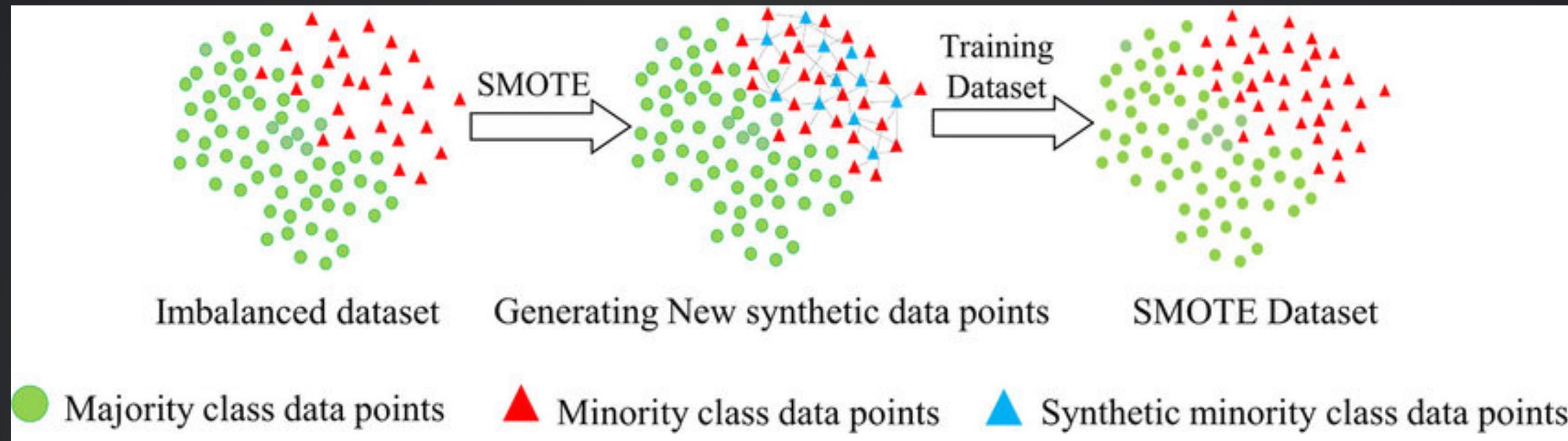
Obtain feature importance

Remove least important feature(s)

**Re-train machine learning with
remaining features**

**Repeat until desired number of
features is reached**

SMOTE



Model Training and Evaluations

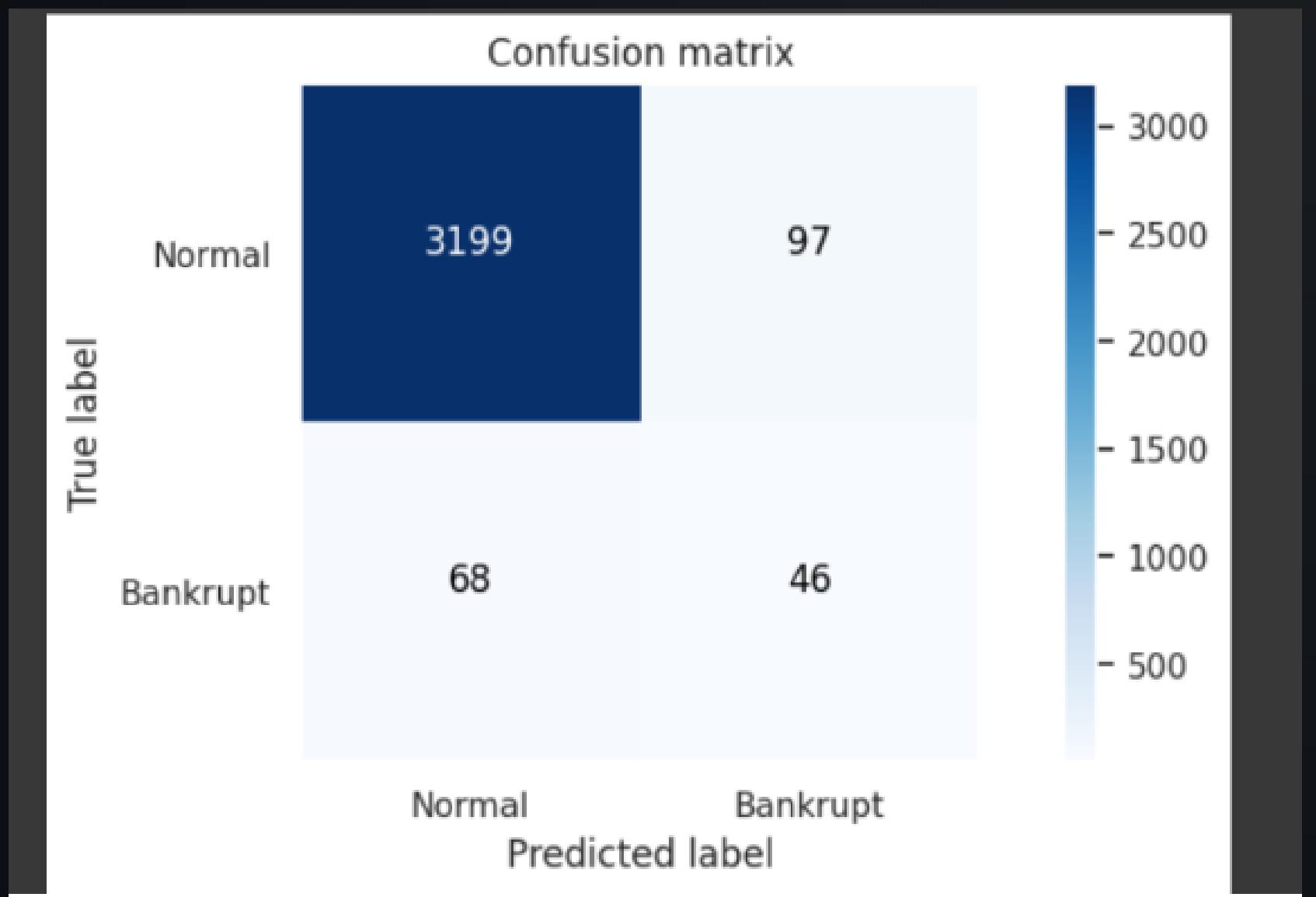
Metrics:

Precision
Recall
f1-score
ROC

Visualization:

Confusion Matrix

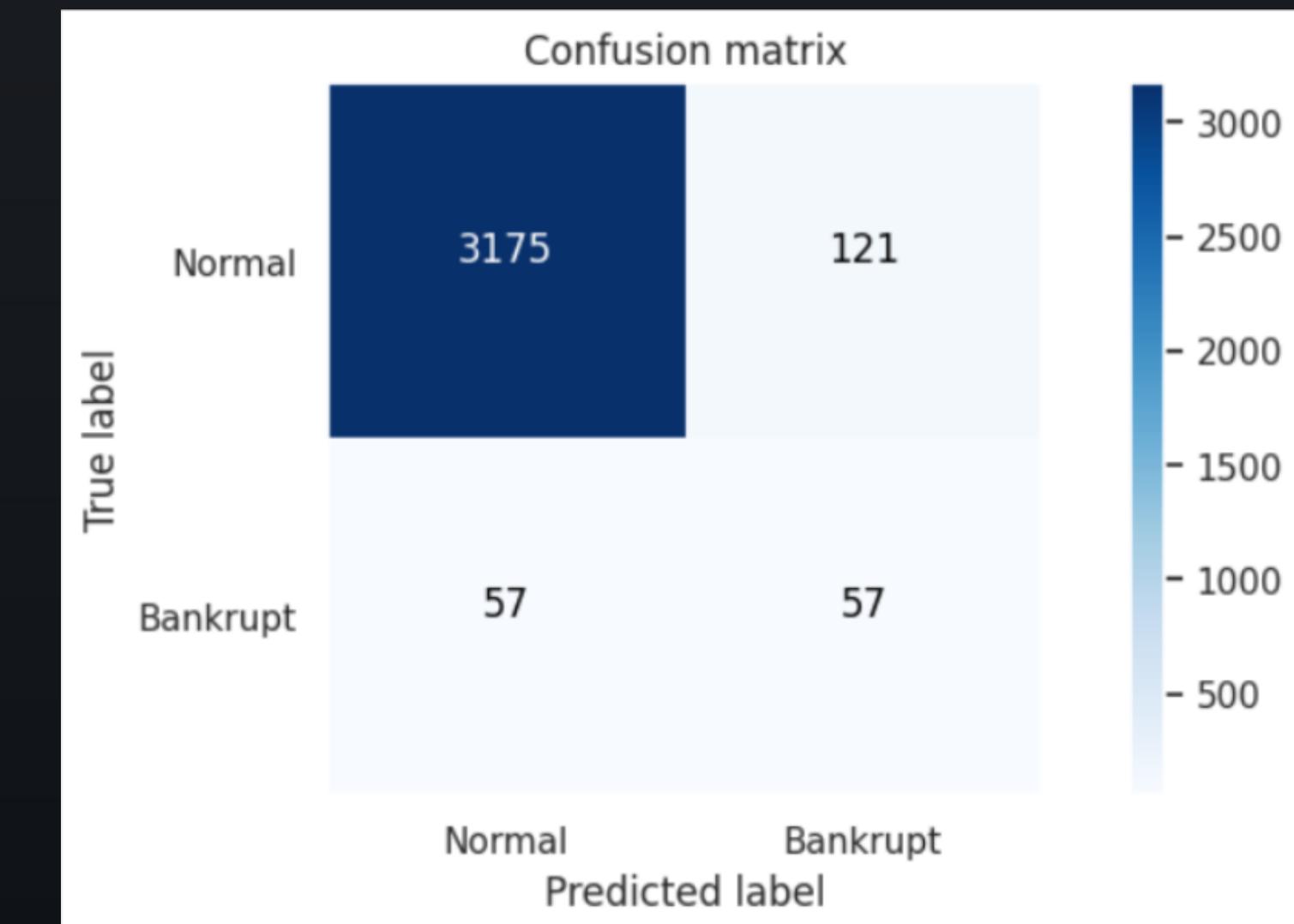
Ensemble methods



```
Confusion matrix, without normalization
Test Set f1 score : 0.35797665369649806
Test set precision : 0.32167832167832167
Test set recall : 0.40350877192982454
PR AUC: 0.2764661430293185
ROC : 0.6870395801396695
```

```
{'learning_rate': 0.2, 'max_depth': 5, 'min_samples_split': 2, 'n_estimators': 200}
```

Gradient Boosting



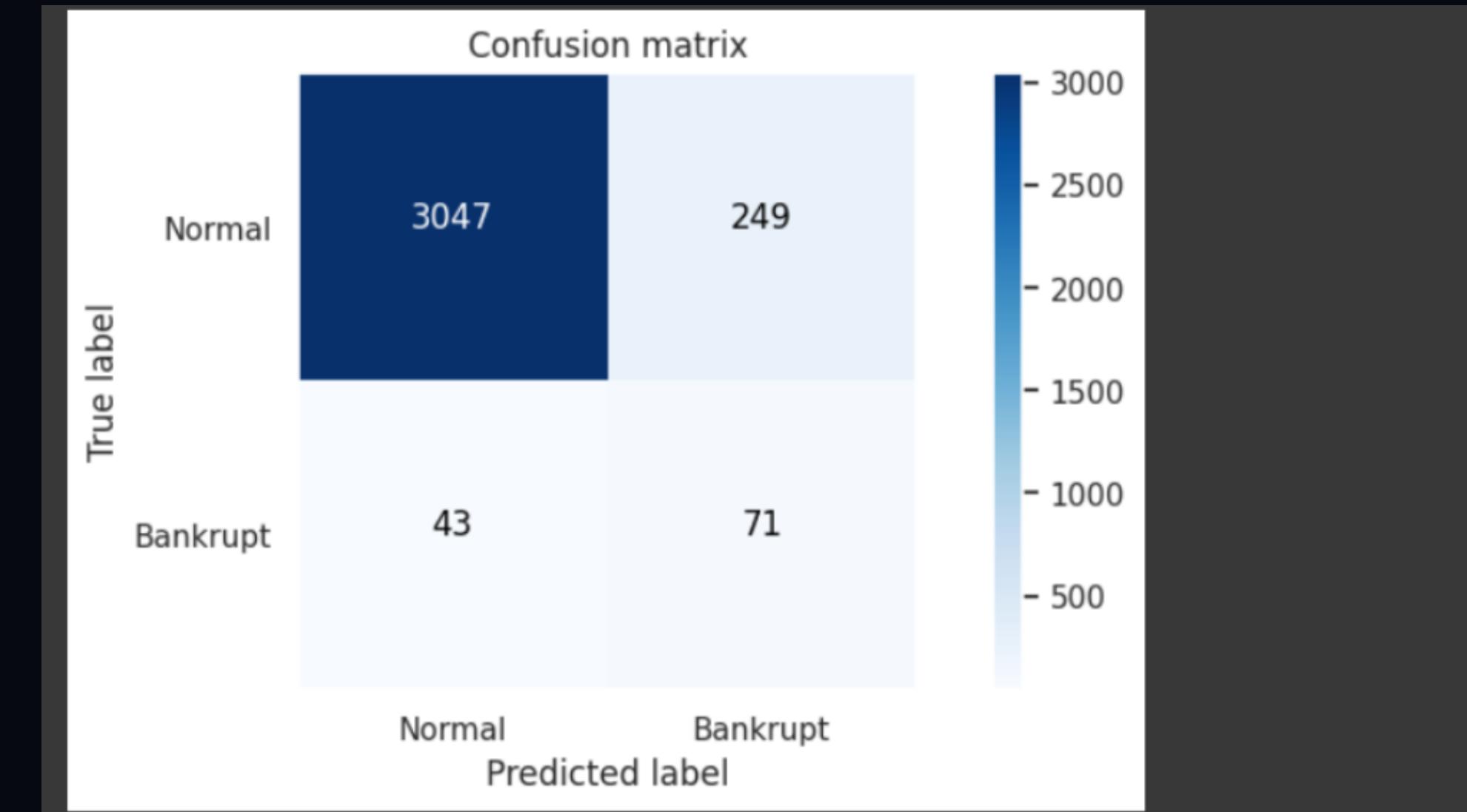
```
Confusion matrix, without normalization
Test Set f1 score : 0.3904109589041096
Test set precision : 0.3202247191011236
Test set recall : 0.5
PR AUC: 0.3066365766641073
ROC : 0.7316444174757282
```

```
{'max_depth': 25, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 100}
```

Random Forest

Best Model & Preprocessing

**MLPClassifier with RFE
and SMOTE**



```
Best Parameters: {'alpha': 0.0001, 'hidden_layer_sizes': (50,), 'learning_rate': 'constant'}
Confusion matrix, without normalization
Test Set f1 score : 0.3271889400921659
Test set precision : 0.221875
Test set recall    : 0.6228070175438597
PR AUC: 0.28428001451628193
ROC : 0.7736304505195027
```

03. Conclusion

**Bankruptcy prediction isn't totally accurate
(we might need non-financial data, such as the industry, age, localisation of the company etc.)**

**Computational costs were heavily reduced
(40% faster on average)**

High "Interest-bearing debt interest rate", "Persistent EPS in the Last Four Seasons", "Total debt/Total net worth", "Equity to Liability" and "Debt Ratio %" tended to lead to bankruptcy more often

THANK YOU

FOR LISTENING