

AIRBNB CHICAGO PRICE PREDICTION



Group 18

Jai Agrawal, Yaohui Wu, Jennie Ryu, Xiaofan Zhu, Sahithi Muddana

CONTENTS OF THE REPORT

01

INTRODUCTION

- Market Analysis
- Data Overview

02

ANALYSIS

- Data Cleaning & Data Preprocessing
- Before Preprocessing
- After Preprocessing

03

CONCLUSION

- Key Takeaways
- Conclusion

MARKET ANALYSIS



5.6M+



4M+



\$110B+



220+

DATA OVERVIEW

LISTING

6,529

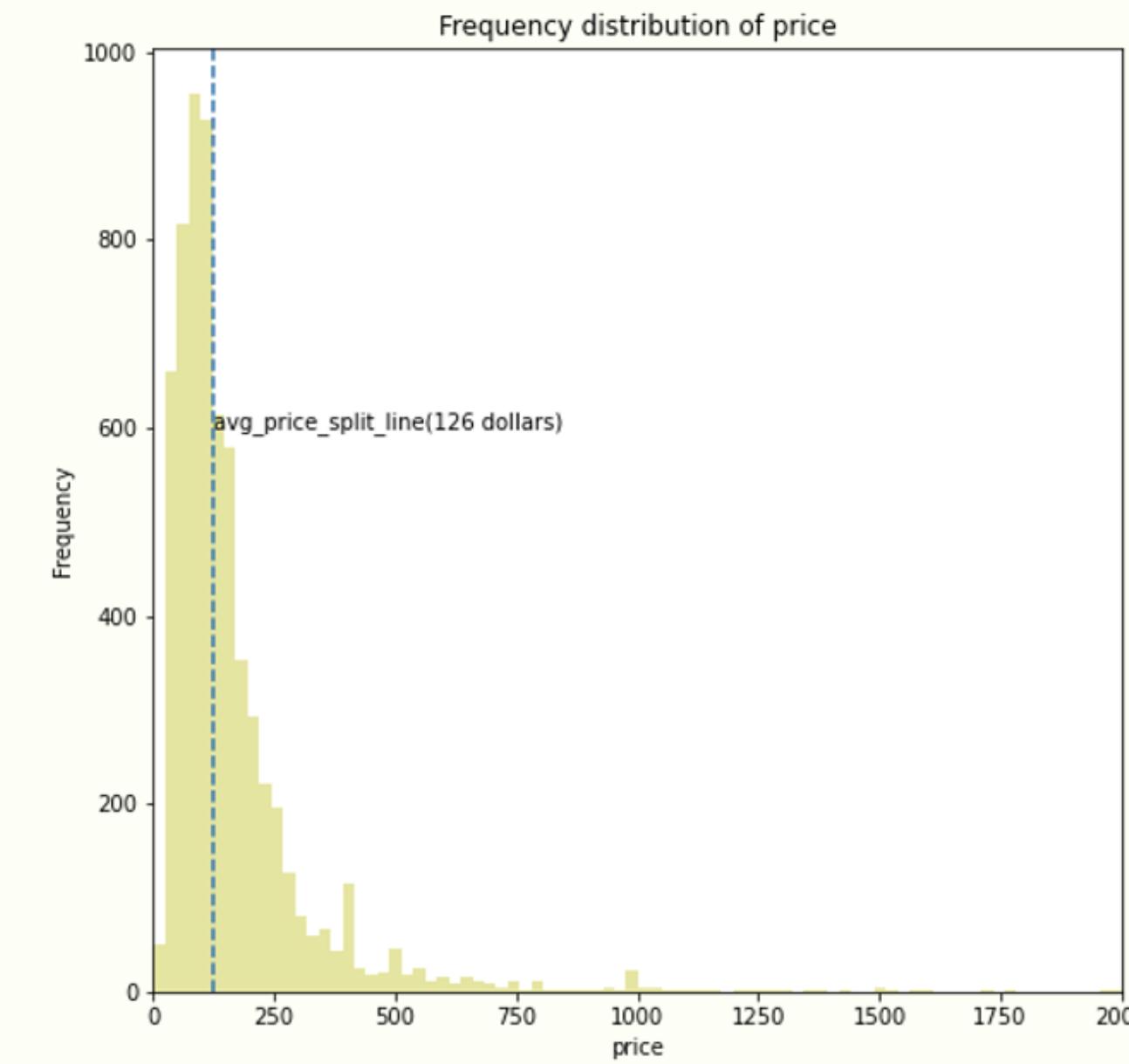
REVIEWS

302,750

36

FEATURES

DATA VISUALIZATIONS



DATA CLEANING & PREPROCESSING

CREATED classification variable using the "Price" feature

REPLACED NA values with averages grouped by "Neighborhood" features

TURNED categorical variables into numeric data

RAN sentiment analysis on text to convert it into numeric data

MODELS

**Logistic
Regression**

**Decision
Tree**

**Random
Forest**

**K-nearest
Neighbors**

LOGISTIC REGRESSION

Benchmark

Accuracy: 0.7570188871873405				
	precision	recall	f1-score	support
0	0.77	0.82	0.79	1106
1	0.74	0.68	0.71	853
accuracy			0.76	1959
macro avg	0.75	0.75	0.75	1959
weighted avg	0.76	0.76	0.76	1959

Accuracy : 0.76

F1-Score : 0.76

AUC : 0.85

After Pre-Processing

Accuracy: 0.7868453105968332				
	precision	recall	f1-score	support
0	0.79	0.85	0.82	926
1	0.79	0.70	0.74	716
accuracy			0.79	1642
macro avg	0.79	0.78	0.78	1642
weighted avg	0.79	0.79	0.78	1642

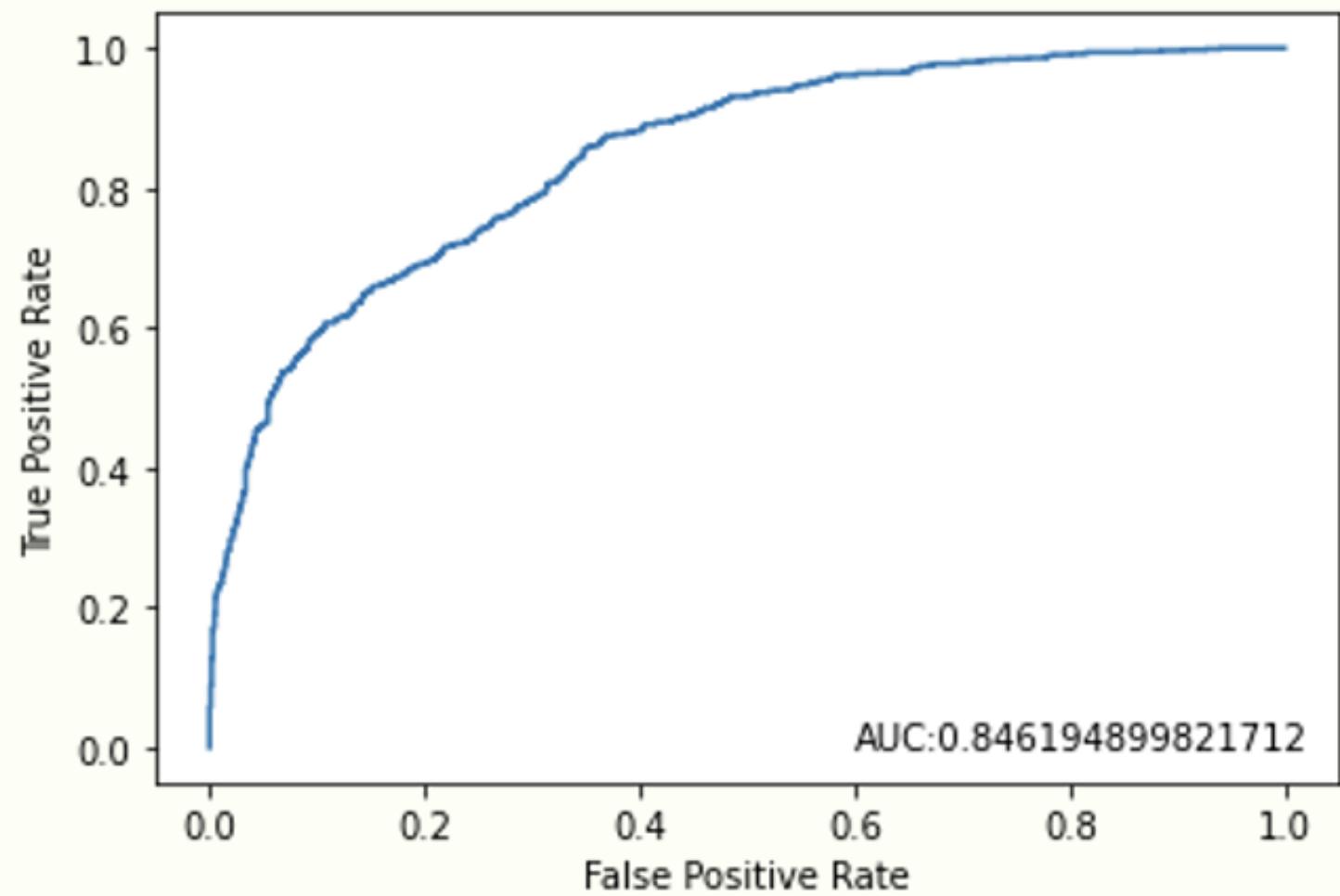
Accuracy : 0.79

F1-Score : 0.78

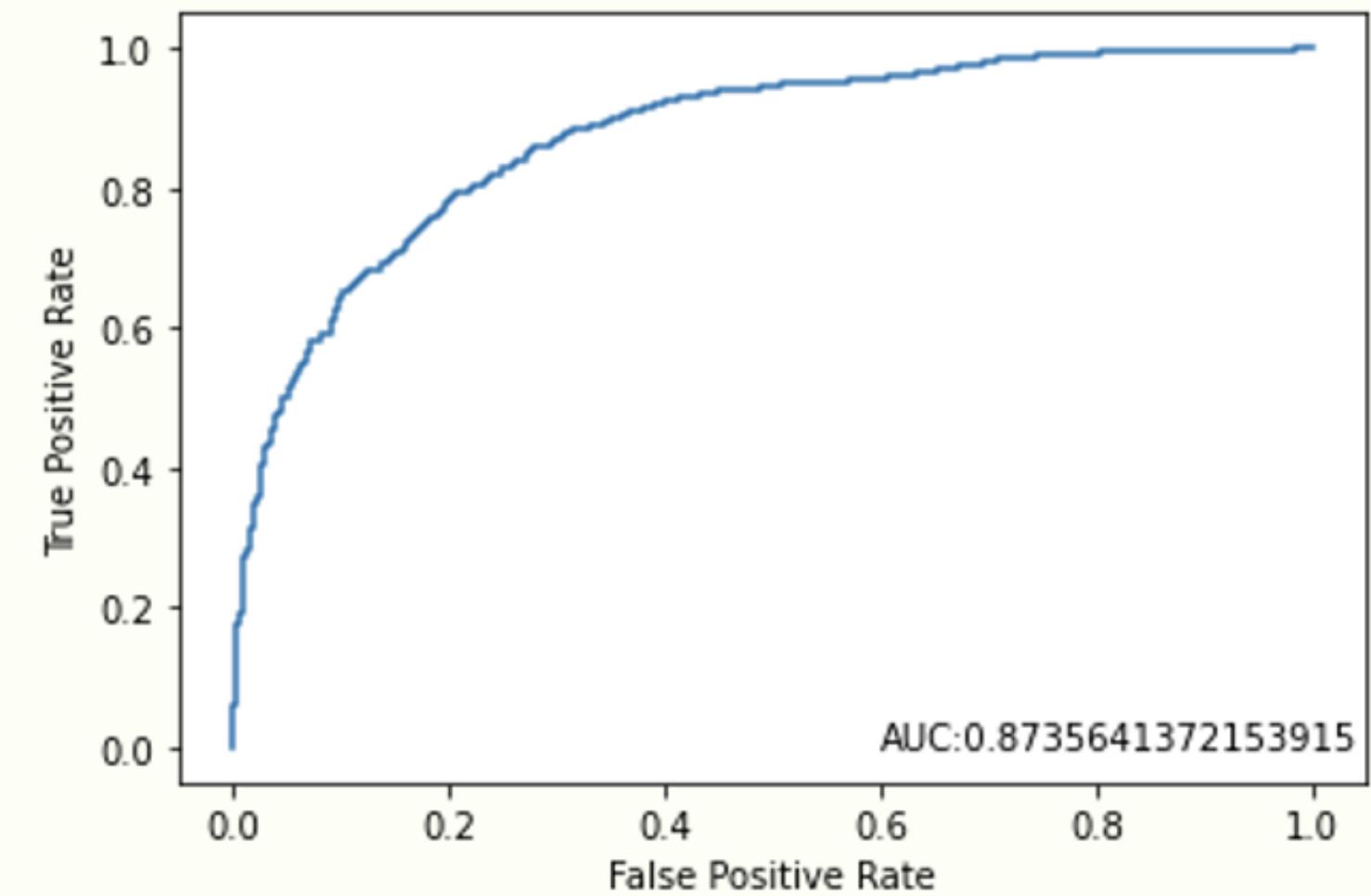
AUC : 0.87

LOGISTIC REGRESSION

Benchmark

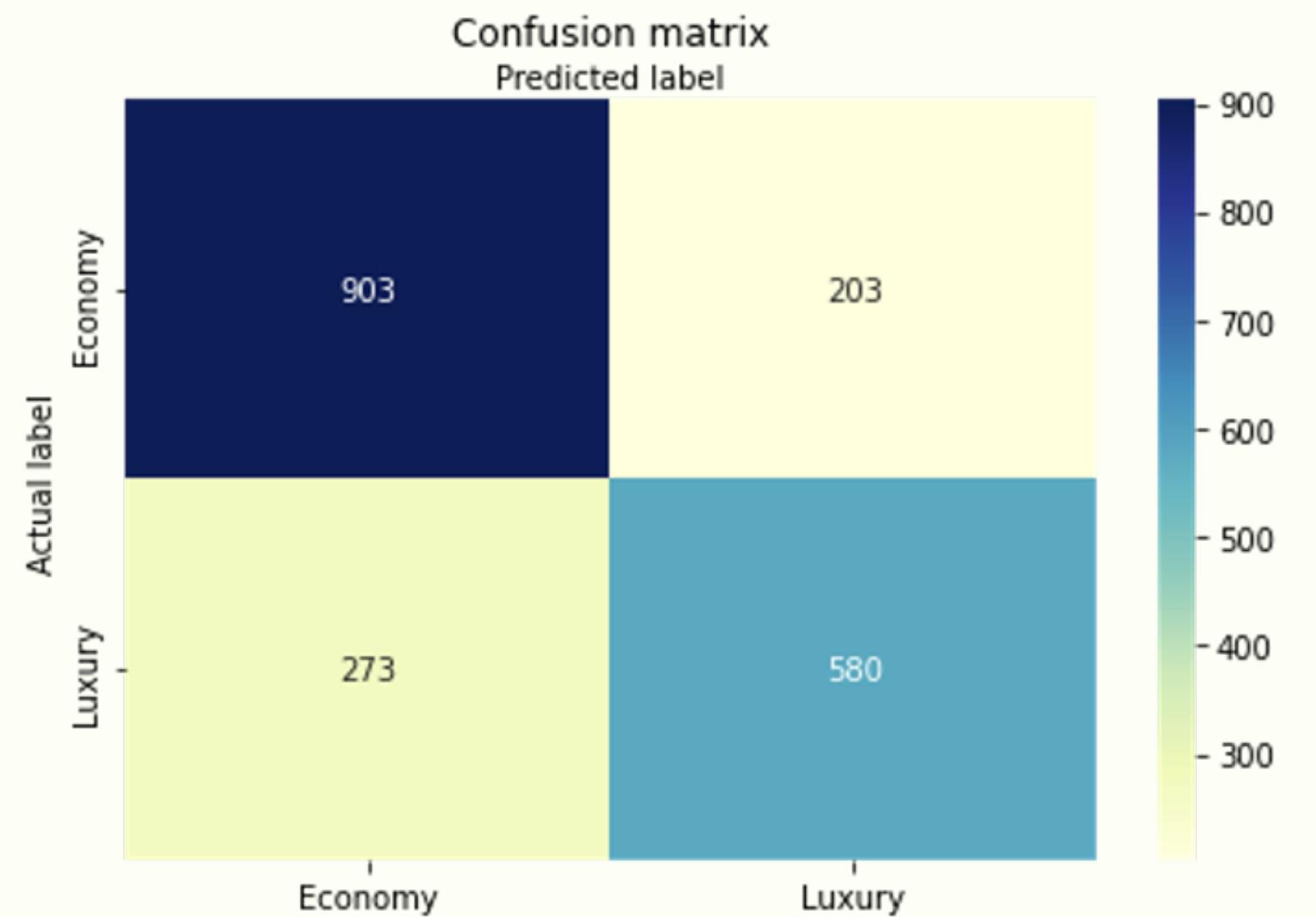


After
Pre-Processing

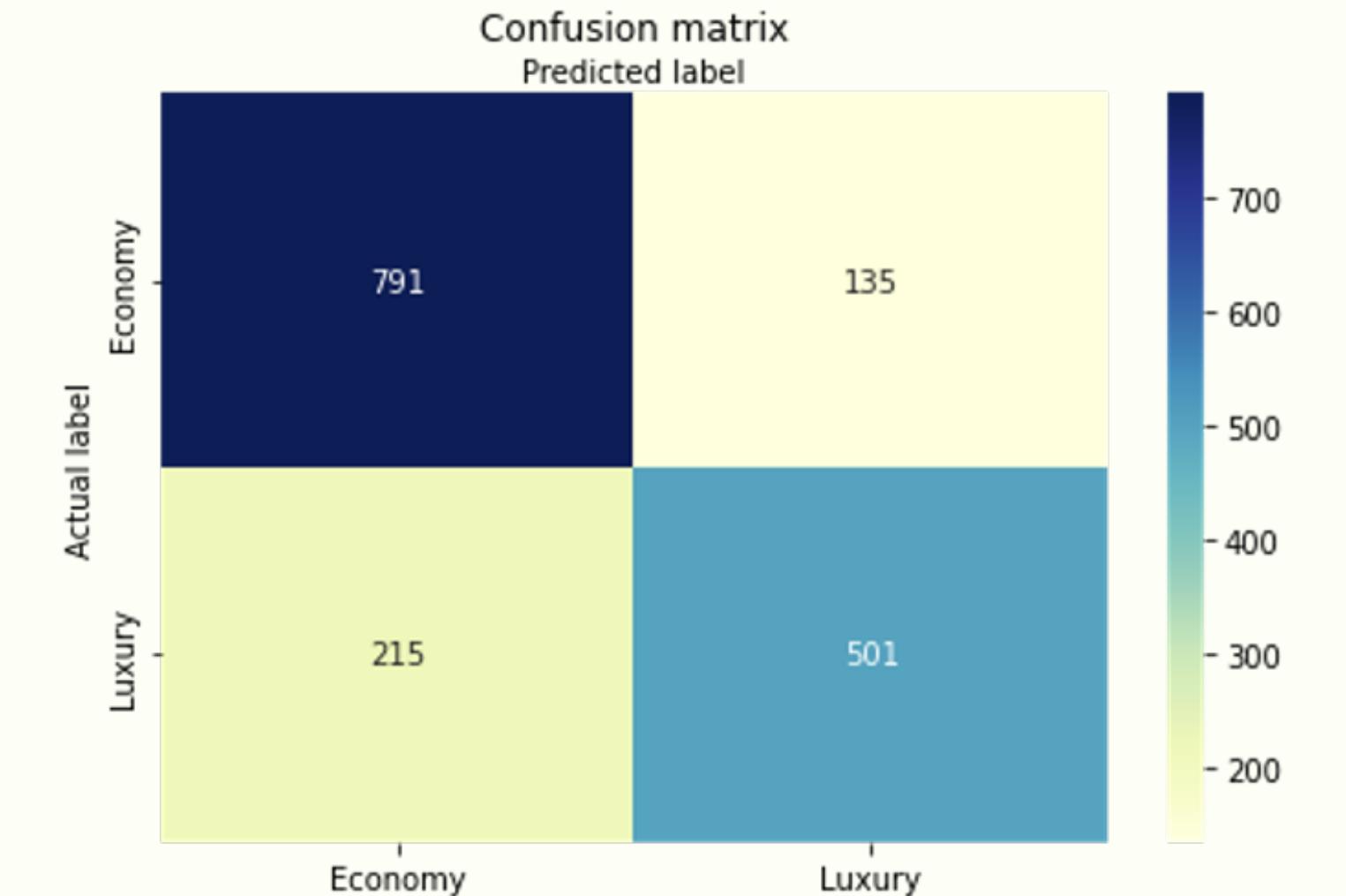


LOGISTIC REGRESSION

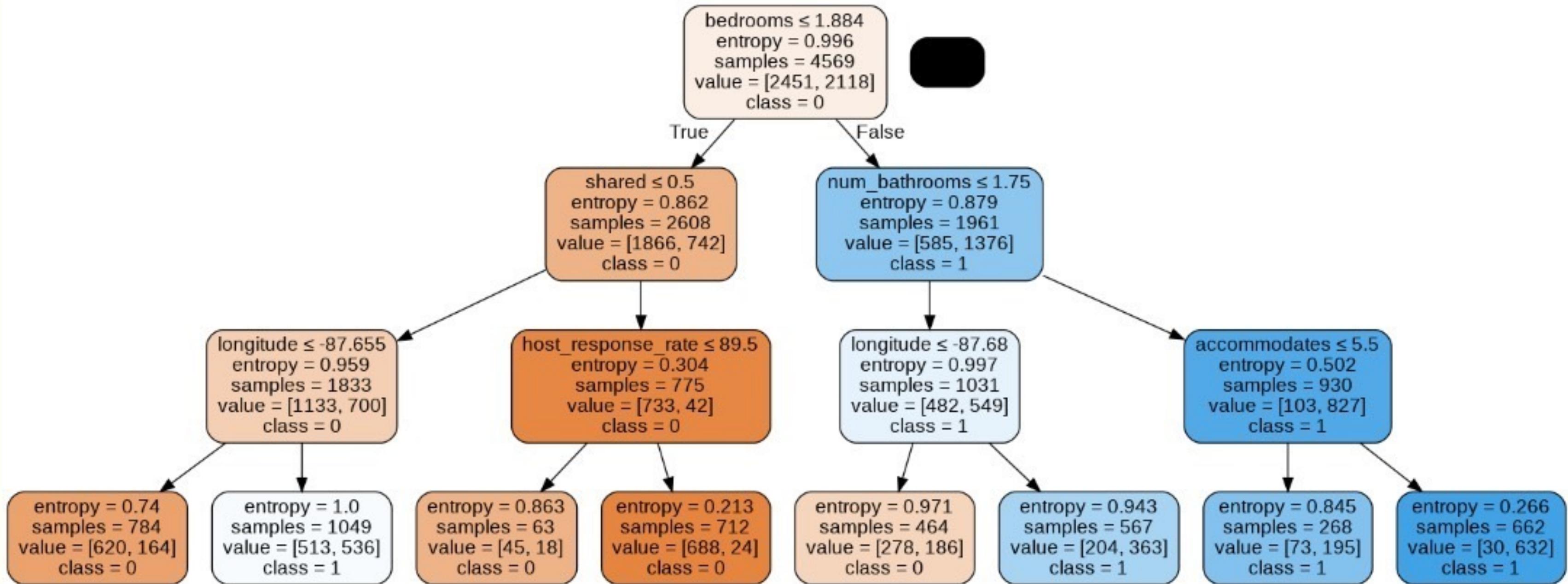
Benchmark



After
Pre-Processing

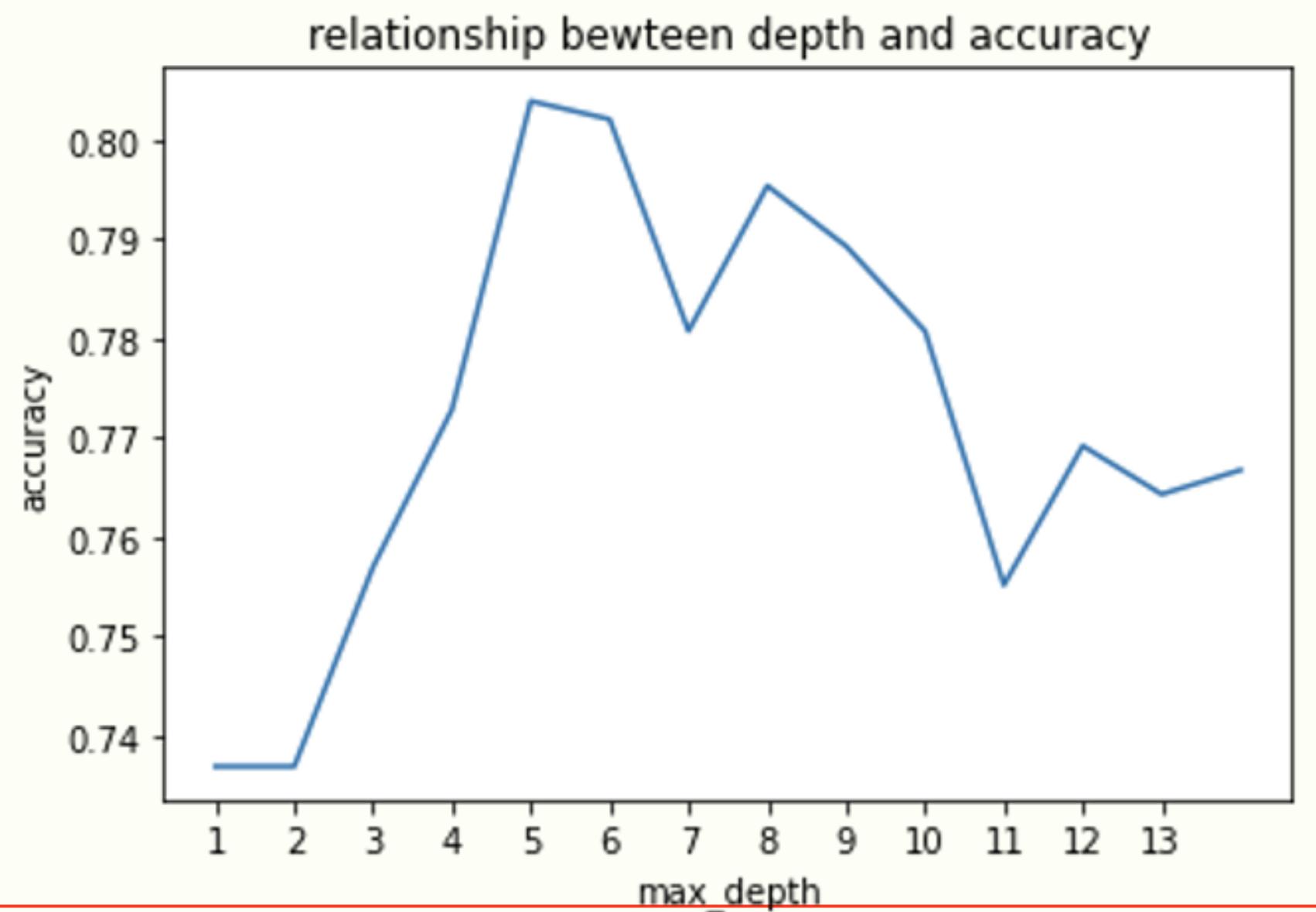


DECISION TREE



DECISION TREE

Depth vs. Accuracy



Optimal Depth : 5

DECISION TREE

Benchmark

Accuracy: 0.7815211842776927				
	precision	recall	f1-score	support
0	0.83	0.77	0.80	1106
1	0.73	0.79	0.76	853
accuracy			0.78	1959
macro avg	0.78	0.78	0.78	1959
weighted avg	0.79	0.78	0.78	1959

Accuracy : 0.78

F1-Score : 0.78

AUC : 0.86

After Pre-Processing

Accuracy: 0.8014616321559074				
	precision	recall	f1-score	support
0	0.80	0.86	0.83	926
1	0.80	0.72	0.76	716
accuracy			0.80	1642
macro avg	0.80	0.79	0.80	1642
weighted avg	0.80	0.80	0.80	1642

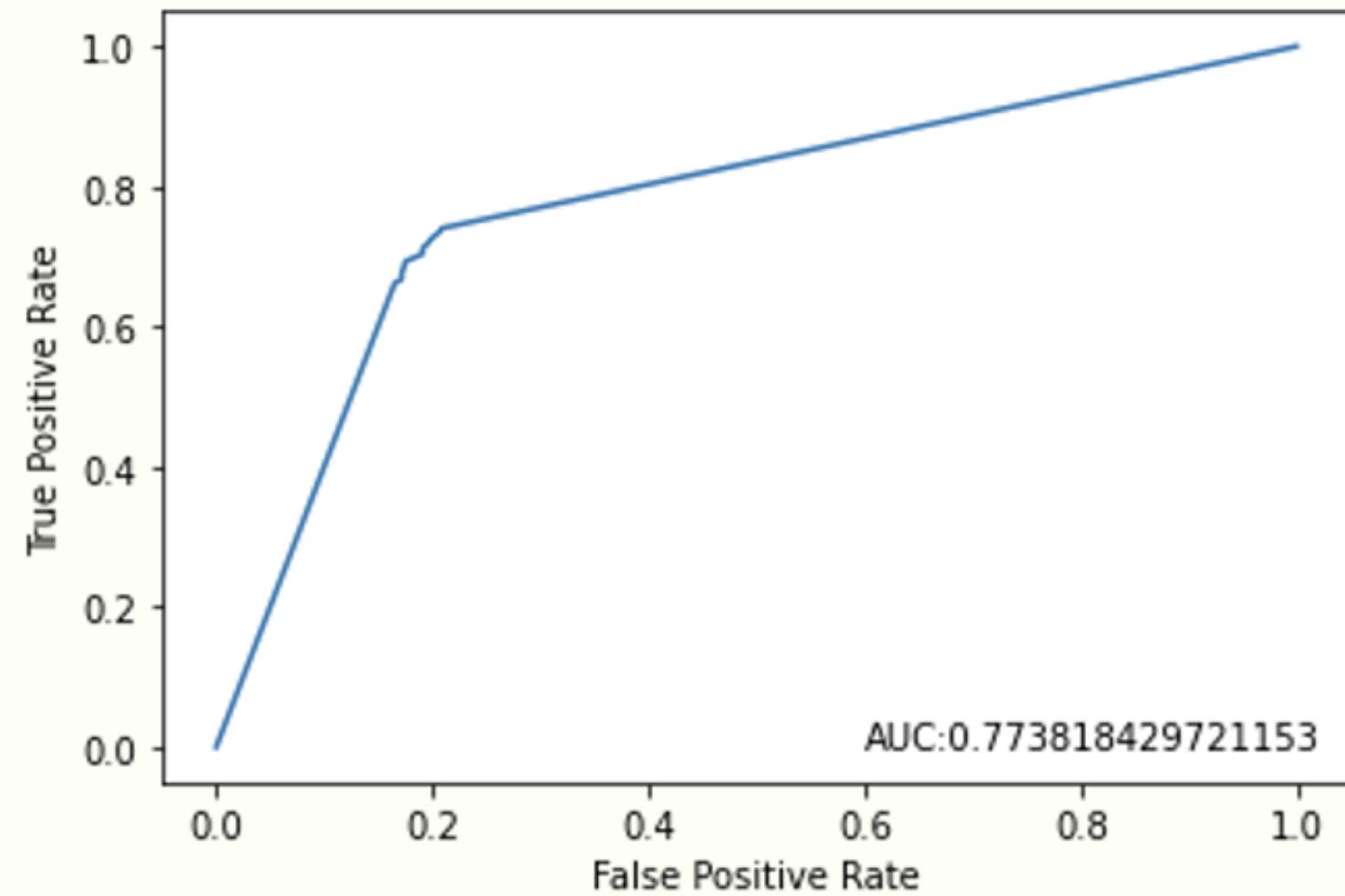
Accuracy : 0.79

F1-Score : 0.80

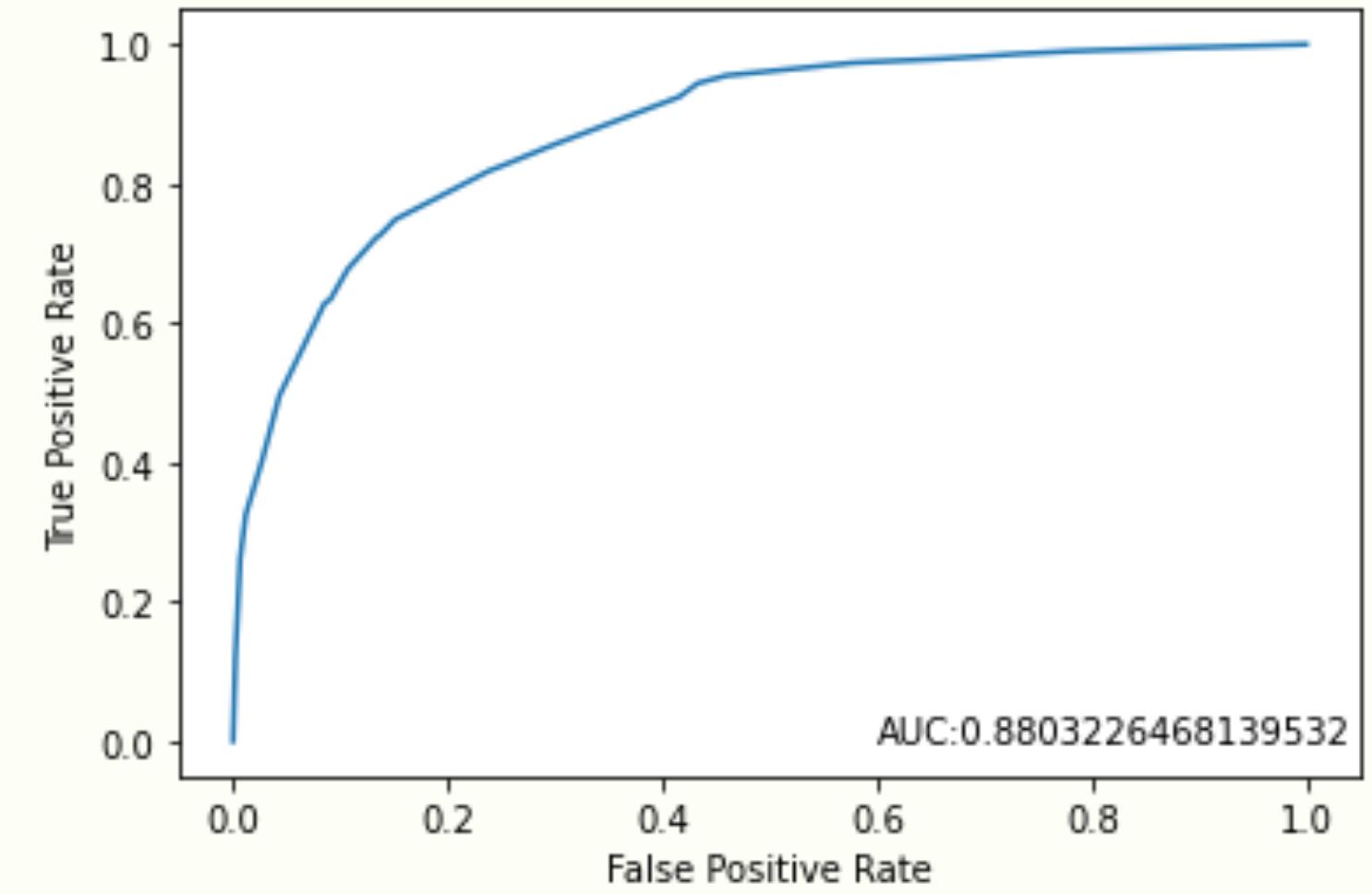
AUC : 0.87

DECISION TREE

Benchmark

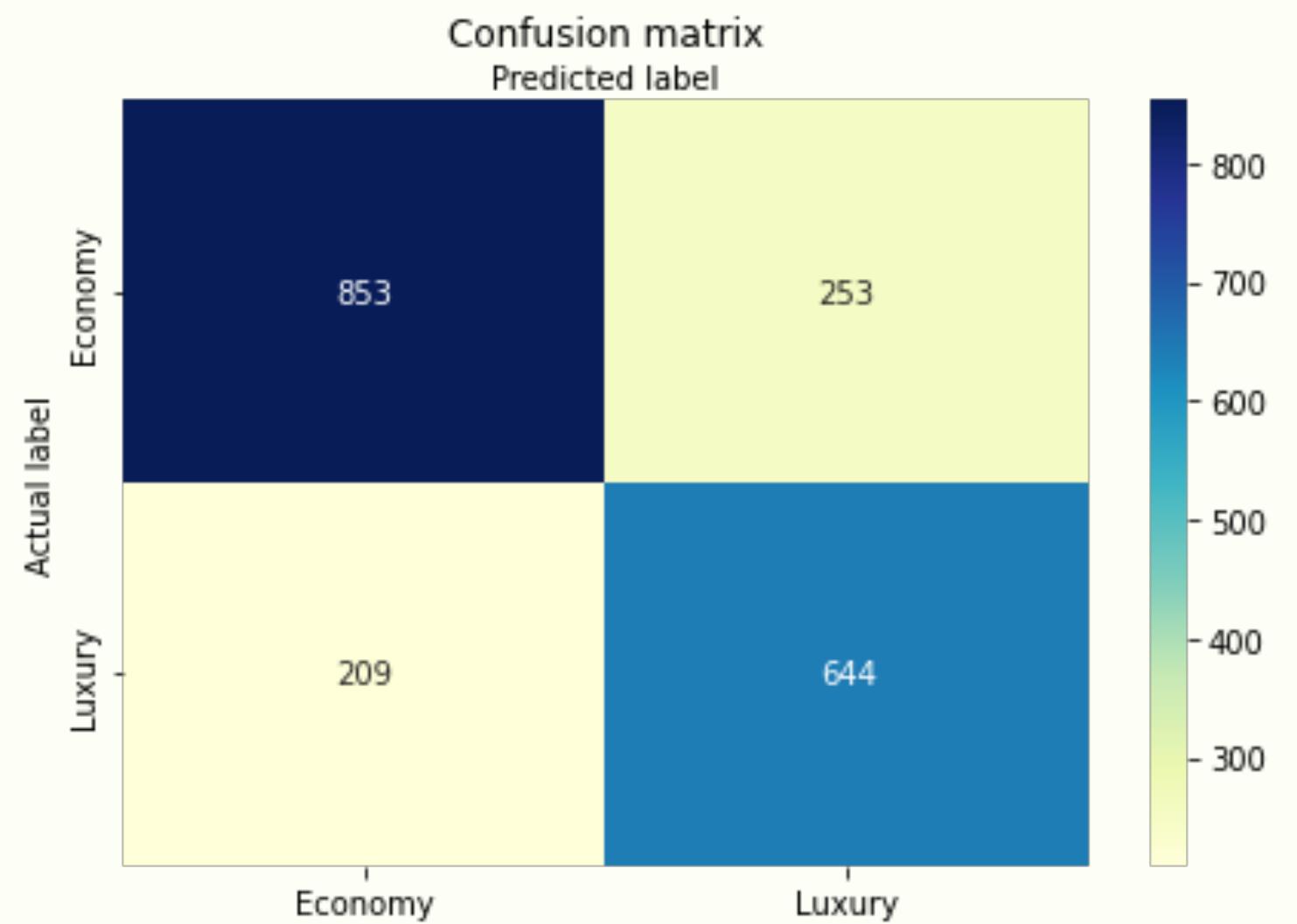


After
Pre-Processing

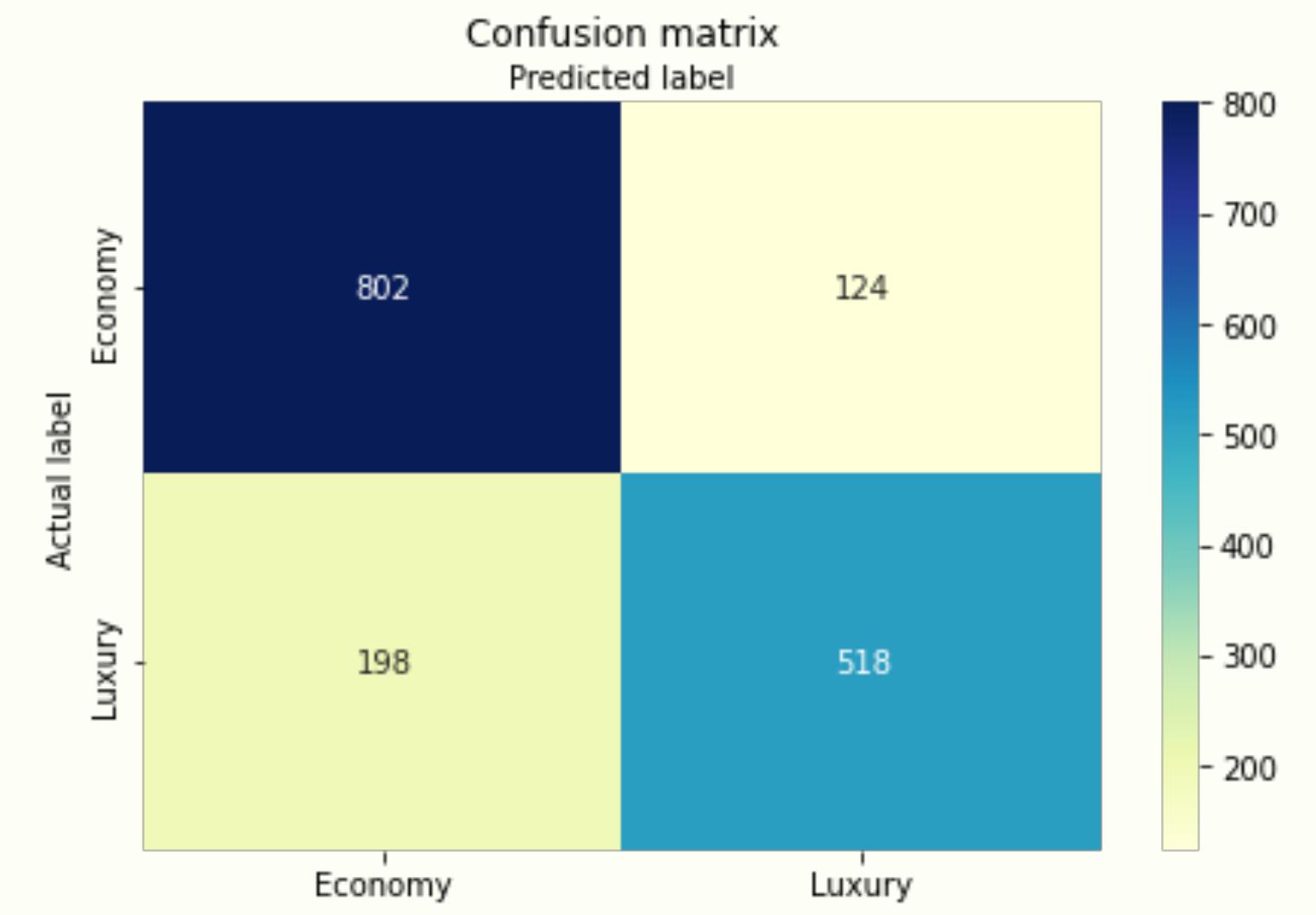


DECISION TREE

Benchmark

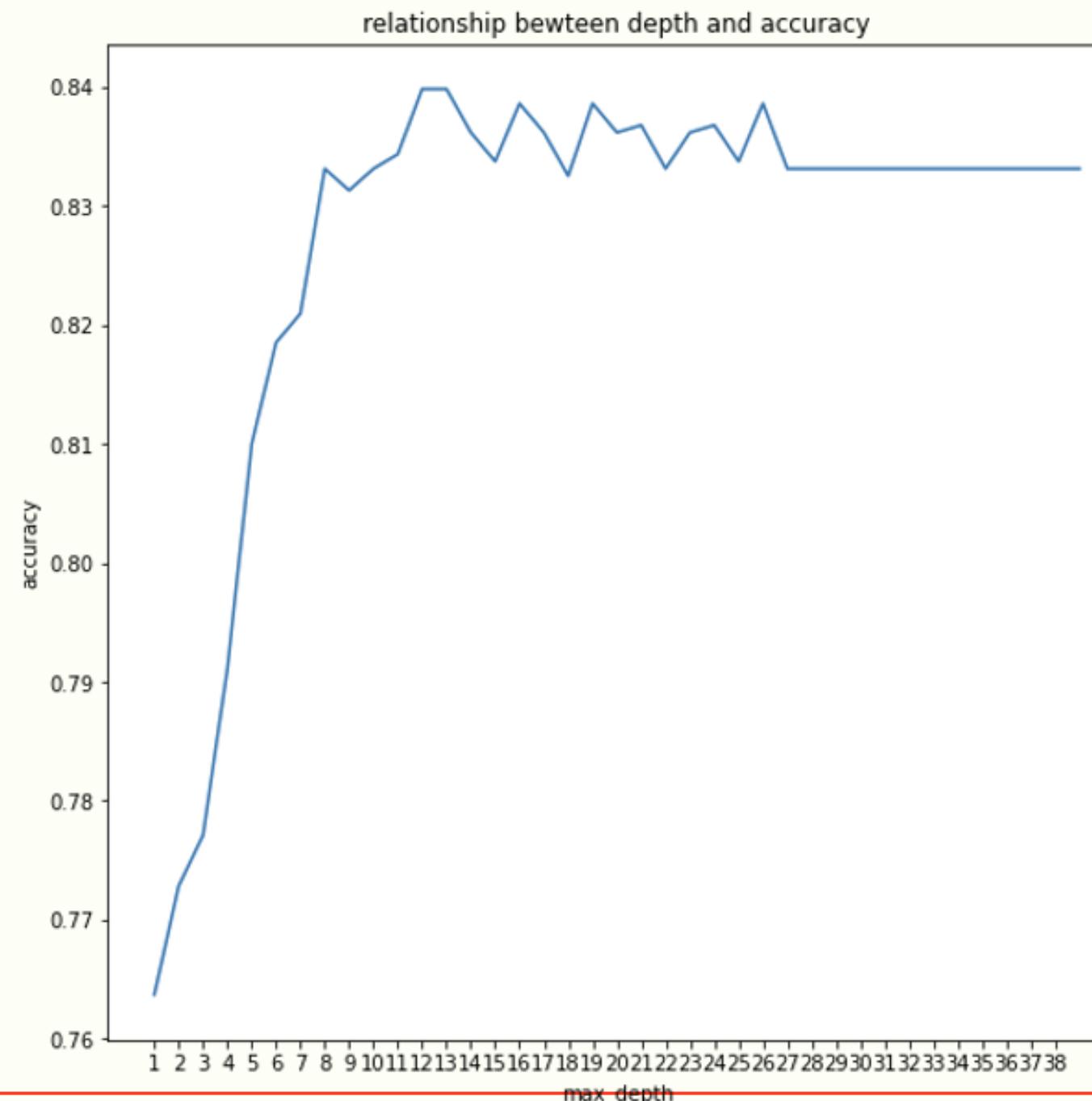


After
Pre-Processing



RANDOM FOREST

Depth vs. Accuracy



Optimal Depth : 12

RANDOM FOREST

Benchmark

Accuracy: 0.8289943848902501			
[[940 166]			
[169 684]]			
precision	recall	f1-score	support
0	0.85	0.85	0.85
1	0.80	0.80	0.80
accuracy		0.83	1959
macro avg	0.83	0.83	0.83
weighted avg	0.83	0.83	0.83

Accuracy : 0.79

F1-Score : 0.79

AUC : 0.88

After Pre-Processing

Accuracy: 0.8398294762484775			
[[817 109]			
[154 562]]			
precision	recall	f1-score	support
0	0.84	0.88	0.86
1	0.84	0.78	0.81
accuracy		0.84	1642
macro avg	0.84	0.83	0.84
weighted avg	0.84	0.84	0.84

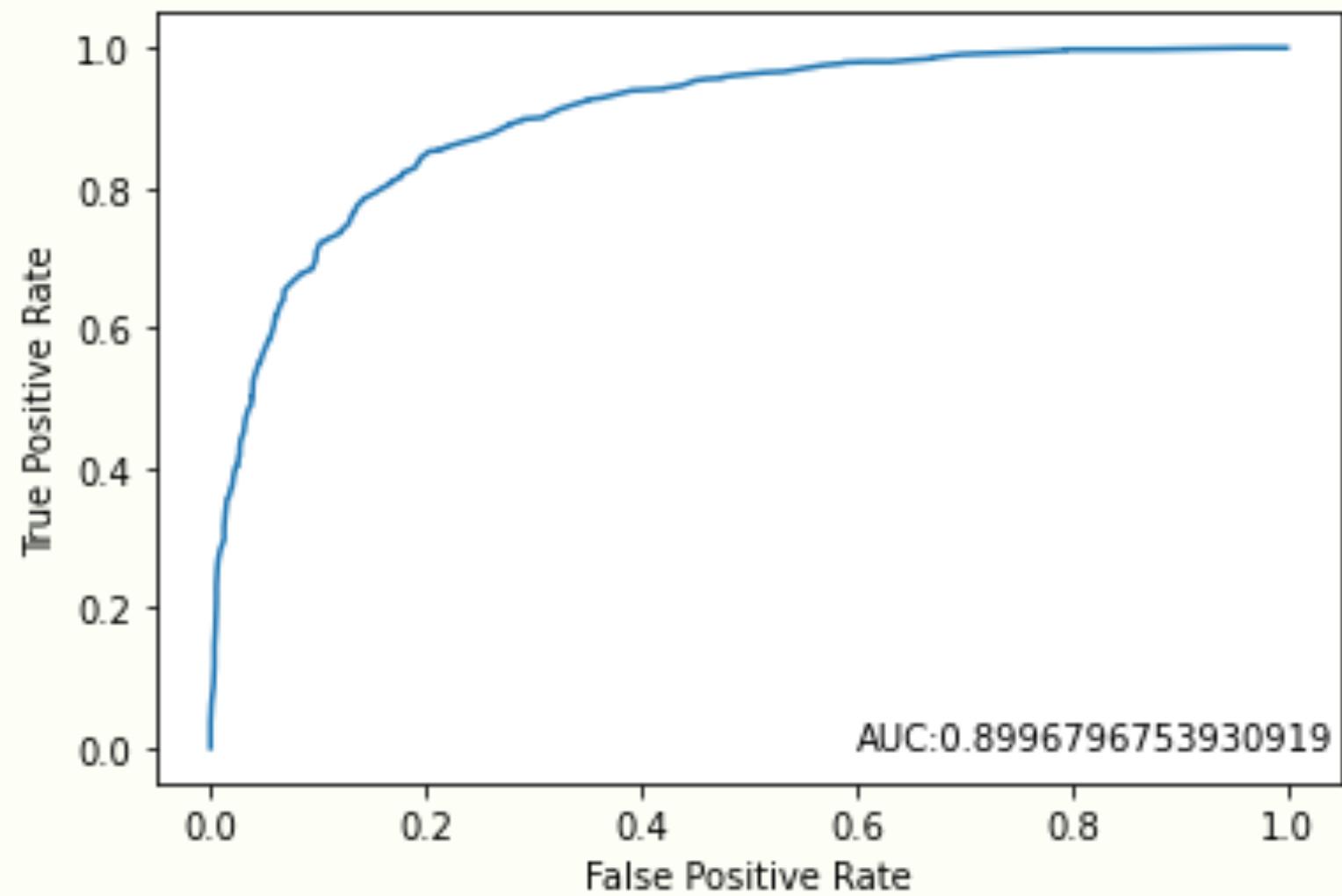
Accuracy : 0.82

F1-Score : 0.82

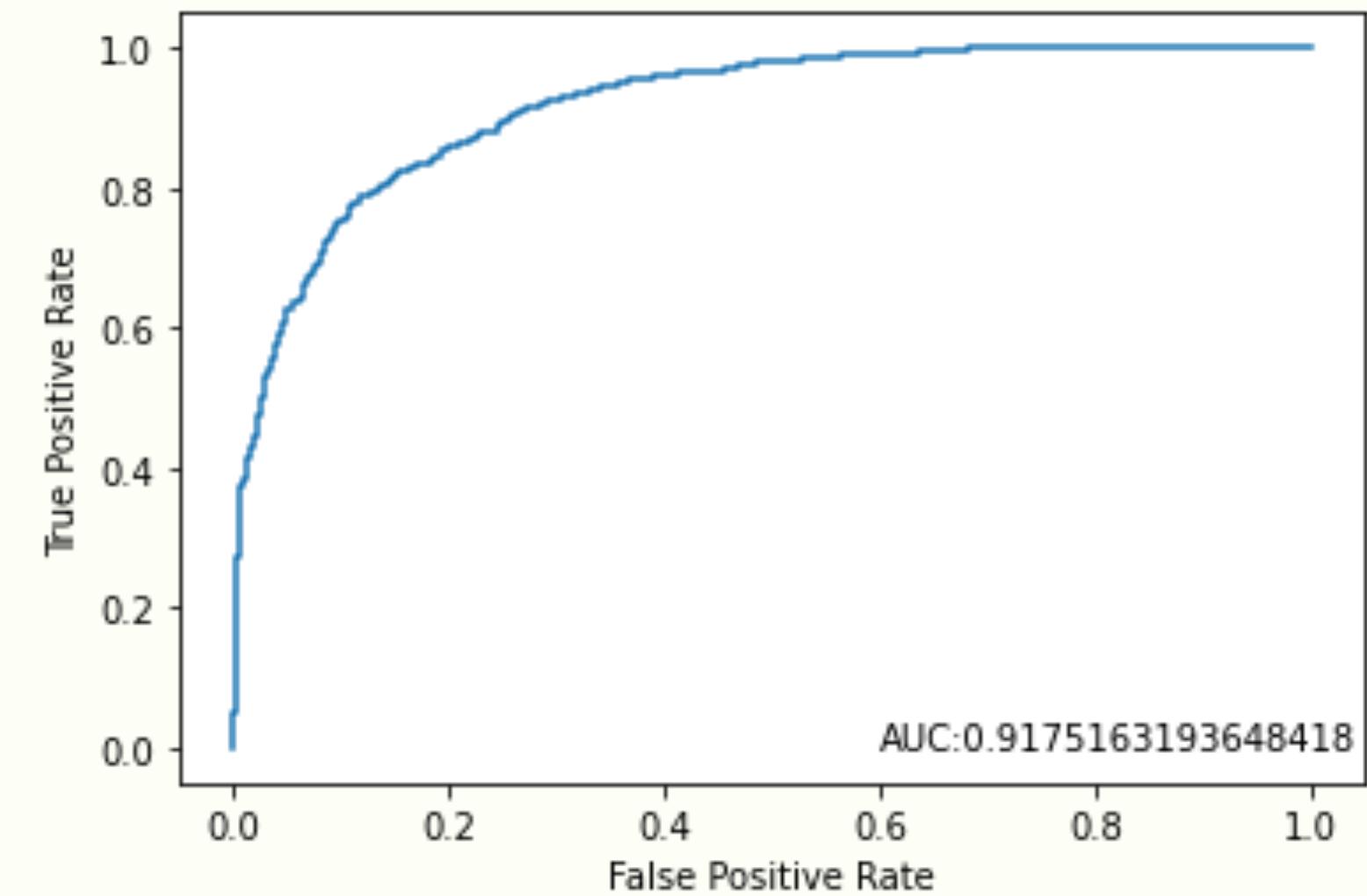
AUC : 0.90

RANDOM FOREST

Benchmark

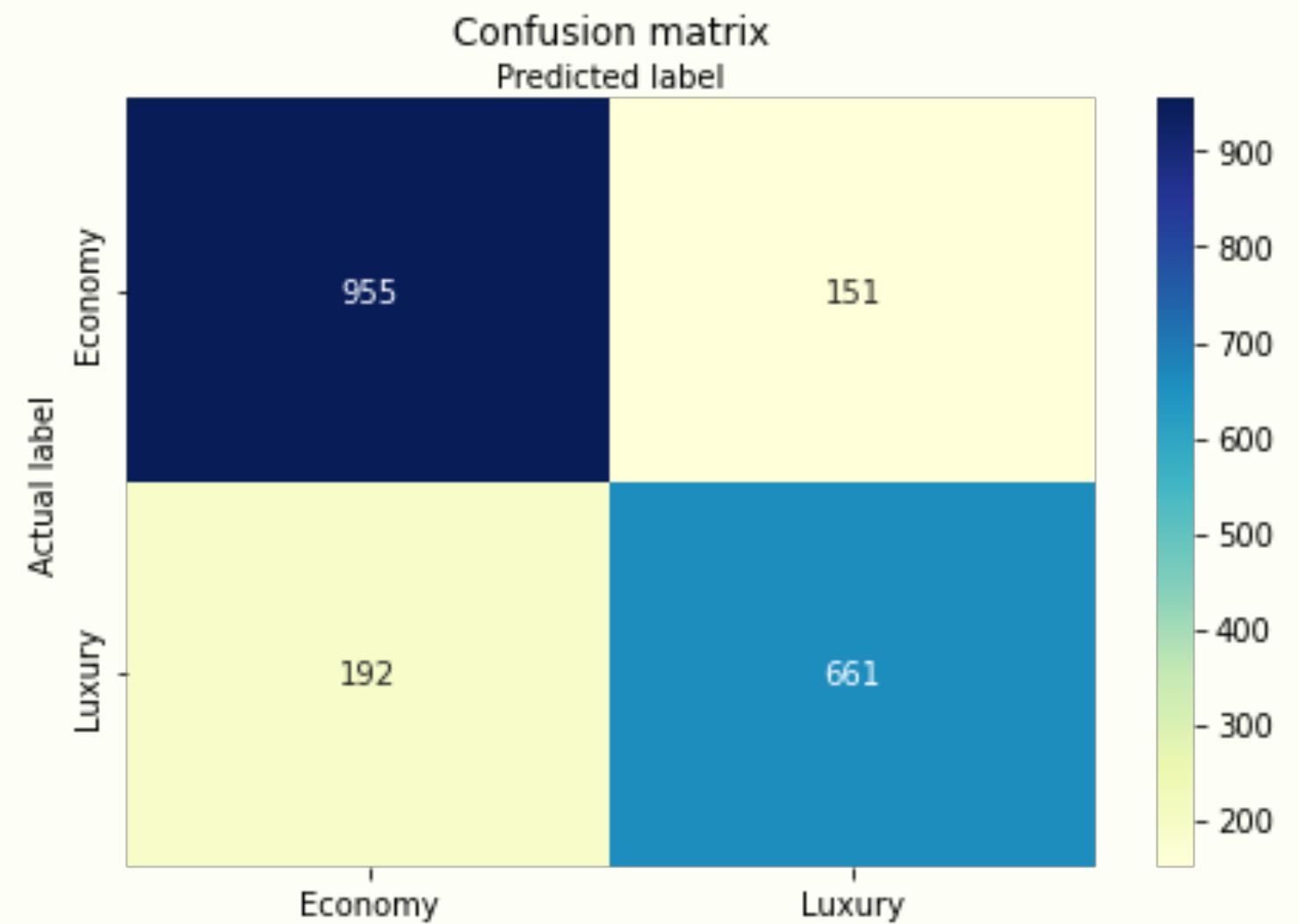


After
Pre-Processing

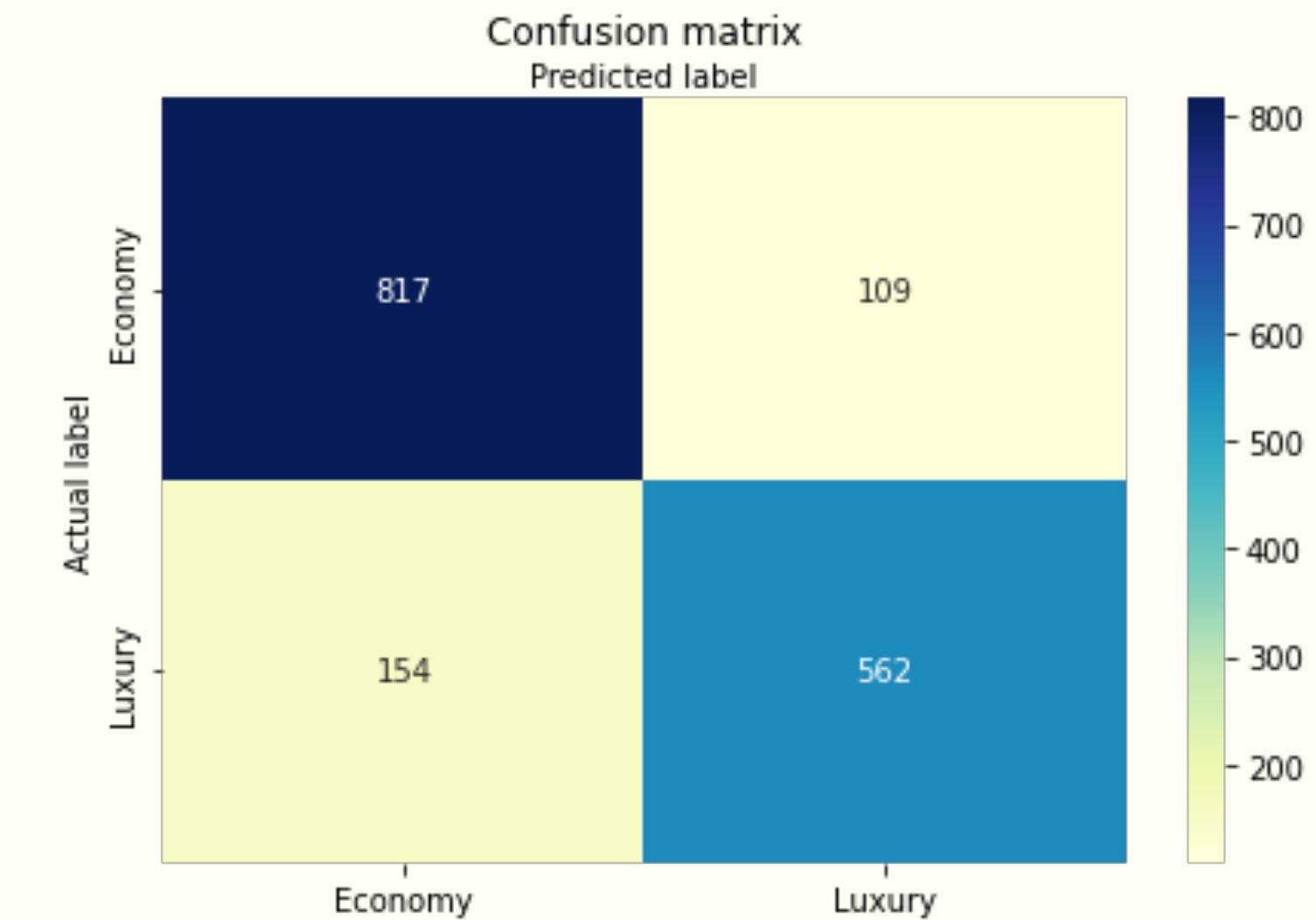


RANDOM FOREST

Benchmark

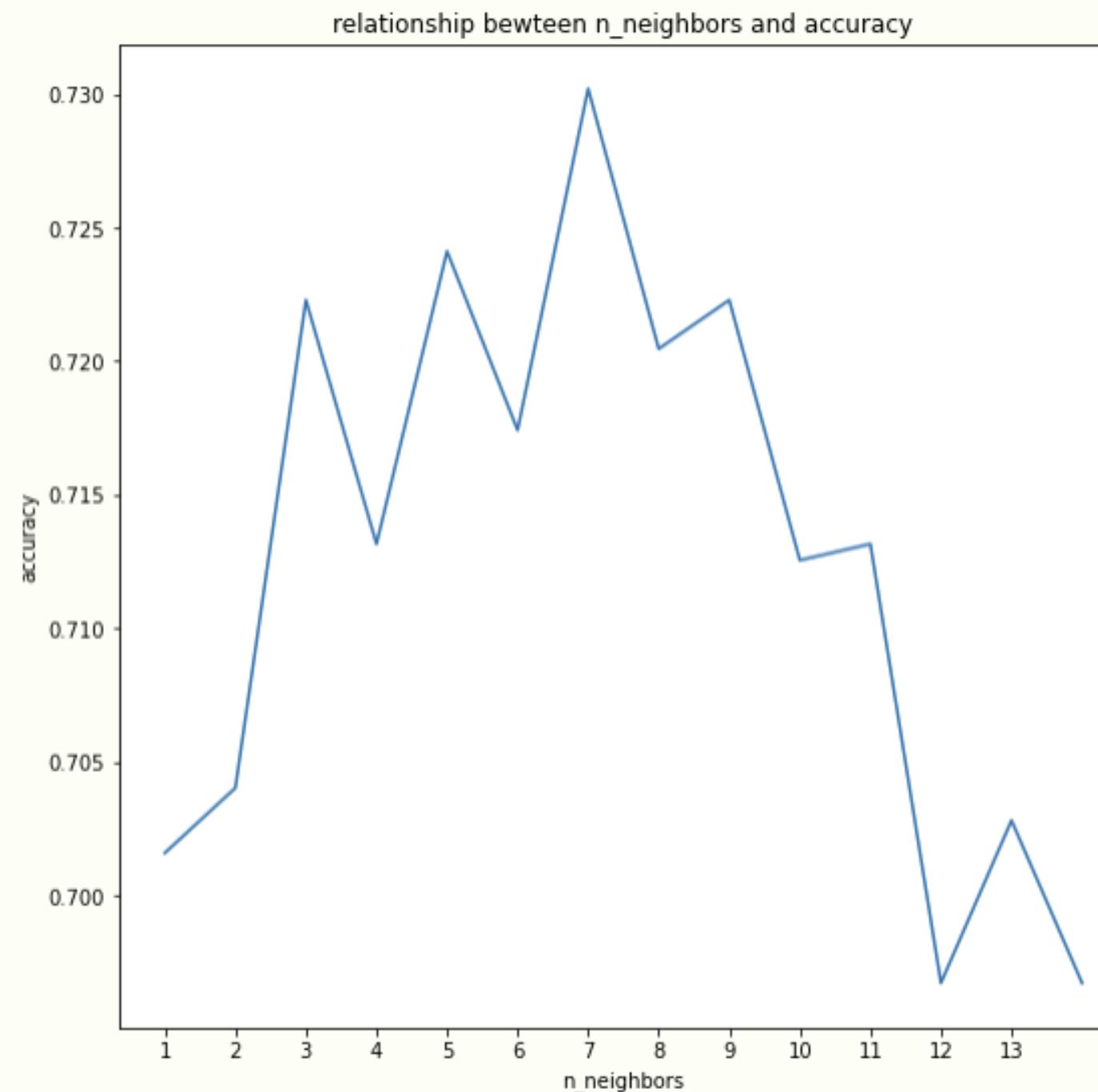


After Pre-Processing



K-NEAREST NEIGHBORS

Depth vs. Accuracy



Optimal K : 7

K-NEAREST NEIGHBORS

Benchmark

Accuracy: 0.8289943848902501
[[940 166]
[169 684]]
precision recall f1-score support
0 0.85 0.85 0.85 1106
1 0.80 0.80 0.80 853
accuracy
macro avg 0.83 0.83 0.83 1959
weighted avg 0.83 0.83 0.83 1959

Accuracy : 0.68

F1-Score : 0.69

AUC : 0.73

After Pre-Processing

0.7302070645554202
[[761 165]
[278 438]]
precision recall f1-score support
0 0.84 0.88 0.86 926
1 0.84 0.78 0.81 716
accuracy
macro avg 0.84 0.83 0.84 1642
weighted avg 0.84 0.84 0.84 1642

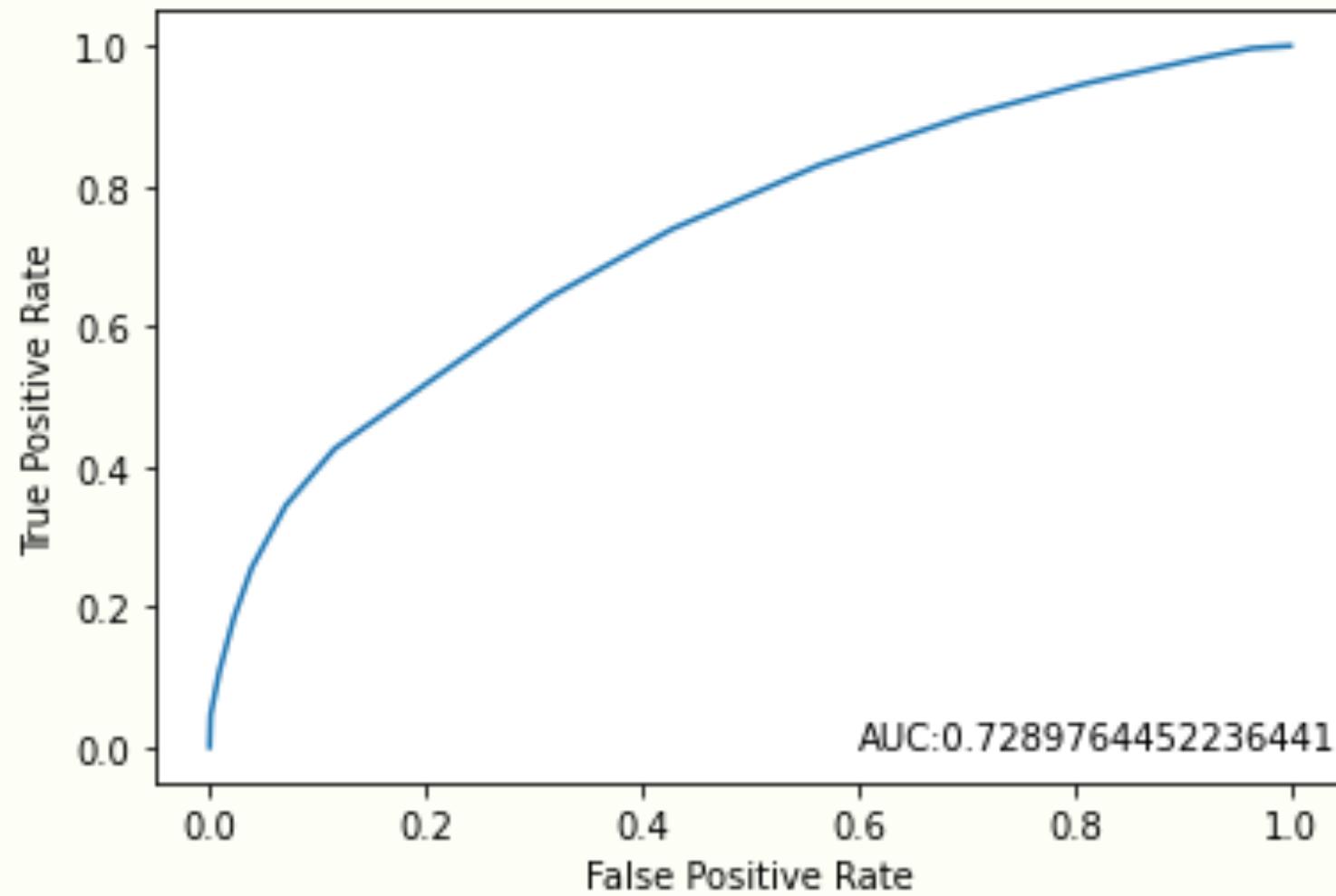
Accuracy : 0.72

F1-Score : 0.82

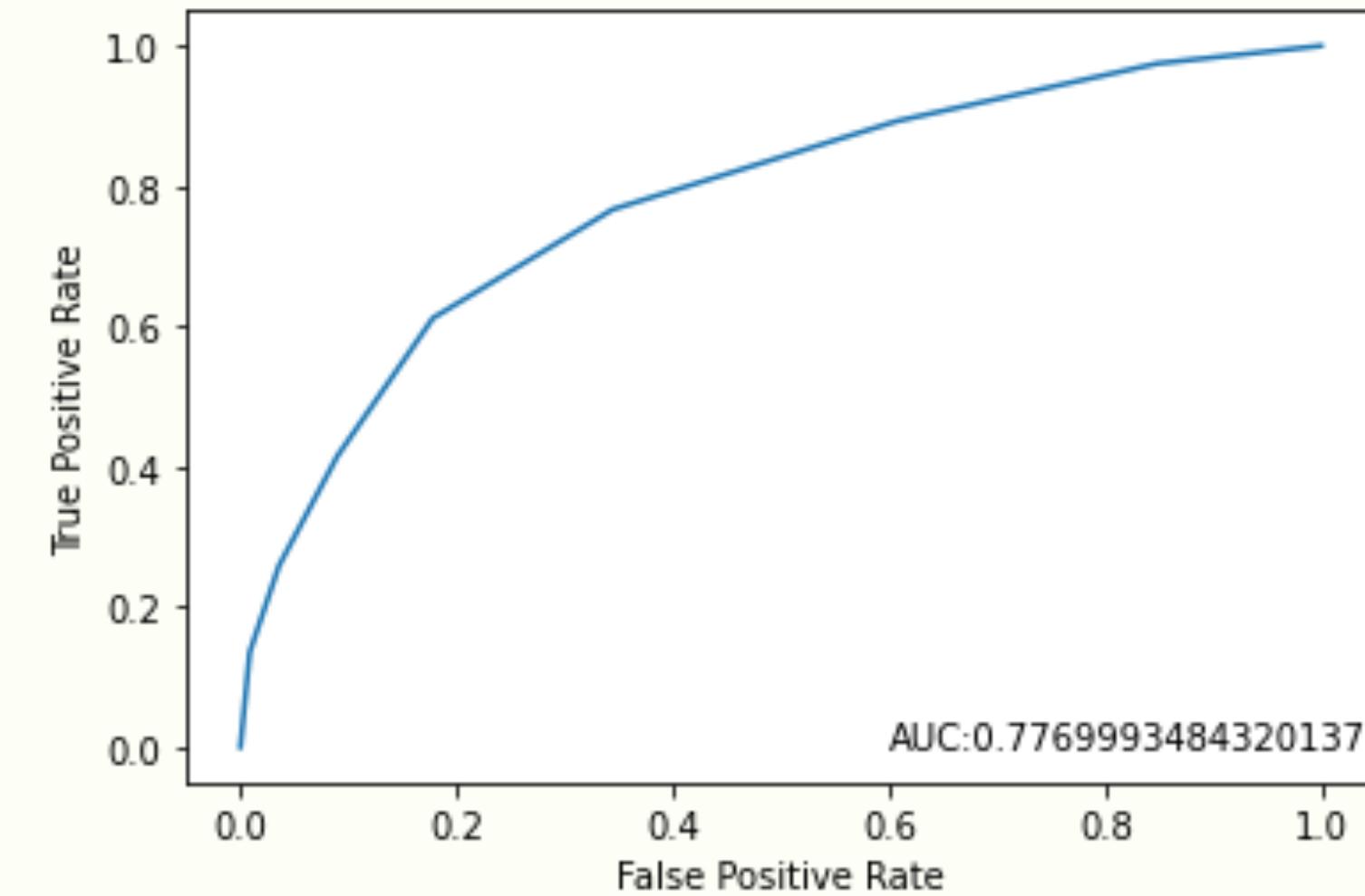
AUC : 0.78

K-NEAREST NEIGHBORS

Benchmark

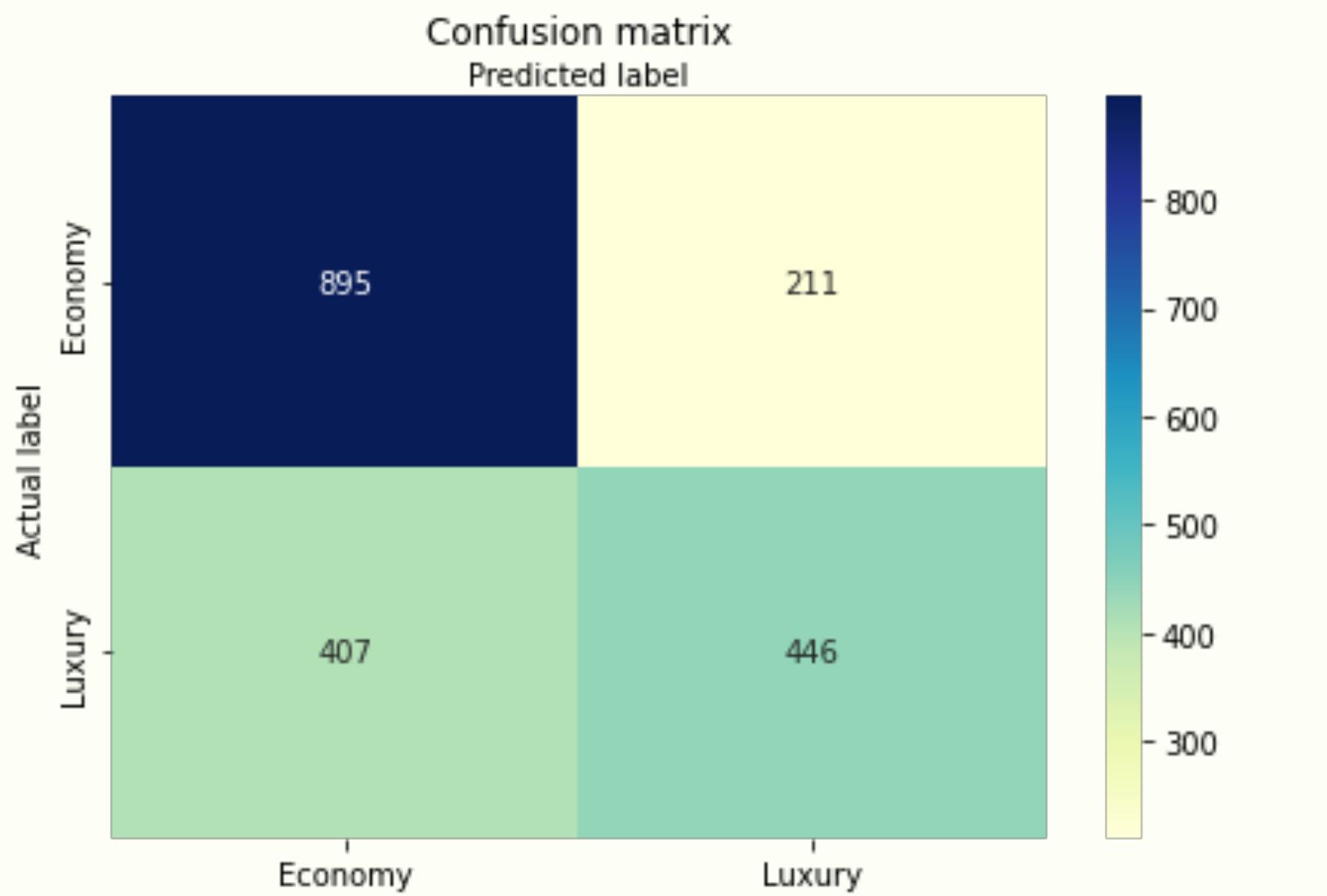


After
Pre-Processing

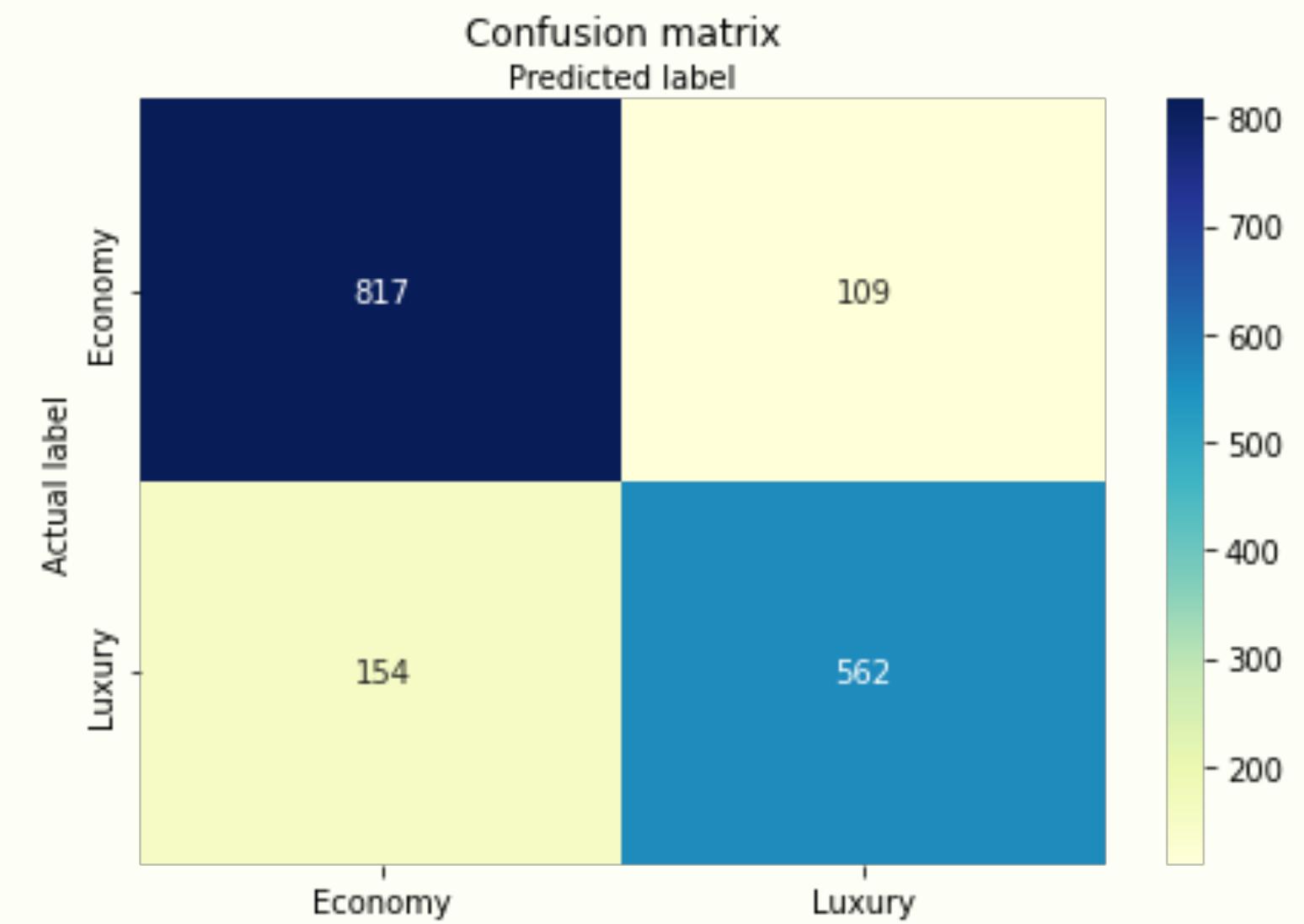


K-NEAREST NEIGHBORS

Benchmark



After
Pre-Processing



KEY TAKEAWAYS

1

Random
Forest

2

Decision
Tree

3

Logistic
Regression

4

K-nearest
Neighbors

Method	Accuracy	Precision	Recall	AUC	F1-score
LogisticRegression(benchmark)	0.76	0.76	0.76	0.85	0.76
LogisticRegression(pre-processing)	0.79	0.79	0.79	0.87	0.78
DecisionTree(benchmark)	0.78	0.79	0.78	0.86	0.78
DecisionTree(pre-processing)	0.79	0.8	0.8	0.87	0.8
RandomForest(benchmark)	0.79	0.79	0.79	0.88	0.79
RandomForest(pre-processing)	0.82	0.82	0.82	0.9	0.82
KNN(benchmark)	0.68	0.71	0.68	0.73	0.69
KNN(pre-processing)	0.72	0.82	0.82	0.78	0.82