BITCOIN

Presented by Team

สมาชิกกลุ่ม

นายนภัสธ	ชินอักษร	6434441923
นายเขมภัทร	ภูมา	6434409923
นายจิรพัฒน์	ธนสุทธิวัฒน์	6434412723

Dataset Sample

Attributes: 46

Instances: 8808

	address	label	n_transactions	block_height_max-m	fee_mean	fee_median	fee_std	fee_min	fee_max	fee_max-min	size_mean	size_median
0	1129yN1itF5a5Hqgw	coinjoin-like	100	31865	16304400	10000	161990110.1	0	1620000000	1620000000	6232.8	753
1	112b8ZjaJWGobxT7	gambling	72	8031	27788.59722	10000	66494.38973	0	450000	450000	2308.416667	782
2	112LhEwGZk5mqpL;	exchange	3	1919	167926.6667	67200	233815.6462	1356	435224	433868	3122.666667	223
3	1136GYGTdySKCoc	gambling	197	119799	249839.0457	40000	1402503.602	744	10000000	9999256	4060.101523	484
4	113Nu2g1dE4d3oiU	miner	6	4238	5135066.667	6120522	4338964.3	24034	9817000	9792966	49573.5	56165
5	113zXqPm8bUcqN3	miner	13	117940	1248242.077	318600	2748839.254	1084	10000000	9998916	9661.538462	1228
6	1142QZVPTNpWYh9	exchange	6	98612	32563.66667	27000	35662.15288	2738	100000	97262	1689.333333	364.5
7	114ADriuRWjkdmAD	miner	54	153368	1781464.481	37150	6219486.141	748	32509232	32508484	4813.425926	374
8	114dCn4TFweCWGc	services	8	31024	26025	6622.5	43737.4827	1130	121840	120710	5135.875	881
9	114i8FXzB8RLuKod	exchange	2	3	56450	56450	61589.00064	12900	100000	87100	2464.5	2464.5
10	115HCtNXsMHy751v	exchange	4	16946	1733867	513135	2674173.38	180800	5728398	5547598	8435.25	8464
11	115mjxm8A1GZ4rzu	exchange	2	2	2181132	2181132	2787601.608	210000	4152264	3942264	18282.5	18282.5
12	115NGnx1tJm2o2rPf	services	3	13	28983.66667	12628	28328.84566	12628	61695	49067	4300.666667	225
13	115Qn5pR6pAFUcK	services	5	773	789081	10405	1086889.73	10000	2253000	2243000	1641.2	521
14	115vqUzjcwsMAYtKu	exchange	8	85492	146105.125	65643.5	174690.9227	7106	517020	509914	3515.25	404.5
15	115xuNLCqjQ12UPY	exchange	62	188948	5095956.774	1775416	8701912.097	10000	45071068	45061068	14074.17742	15396.5
16	116AYmdheVE4o64	miner	8	51493	173944.875	106480.5	166289.921	5976	432460	426484	8044.75	3527
17	11d17SGG28V9K5z	coinjoin-like	4	10	108260	116480	35588.17032	61400	138680	77280	1655.75	1537.5
18	11NomAcyxruQ2t9is	exchange	8	2995	126176.875	100000	121516.2108	5621	380000	374379	8170.375	3711
19	121FDZK6HLCwBhC	coinjoin-like	114	23603	47631.57895	10000	75143.55149	10000	550000	540000	3864.640351	620.5
20	121p3NbdXCP9AKo	exchange	50	49804	3964751.4	858159	8724648.59	22763	47250164	47227401	16234.96	15426

DATA PREPROCESSING

SPLITTING DATA

85% data for modeling and 15% unseen data
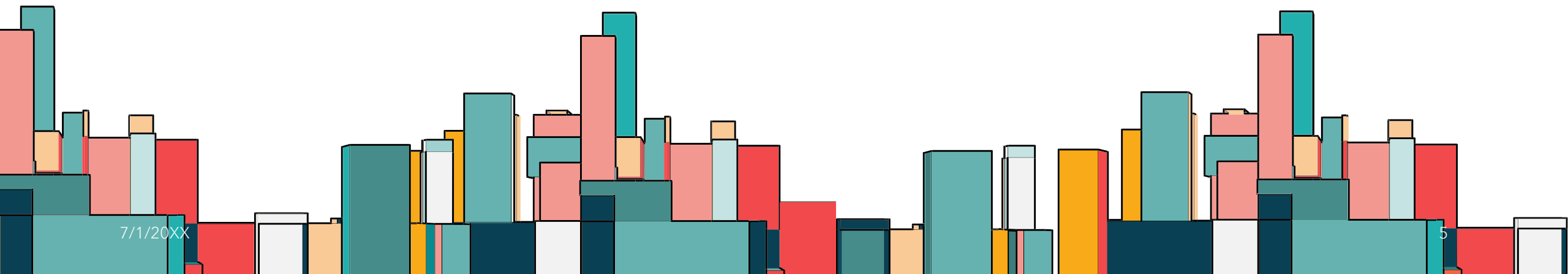
DATA REDUCTION

Drop the 'Unnamed: 0' column

DATA CLEANING

Remove rows with null values

DATA INTEGRATION

Calculate the correlation matrix and Iterate through the top correlated pairs and drop one feature from each pair
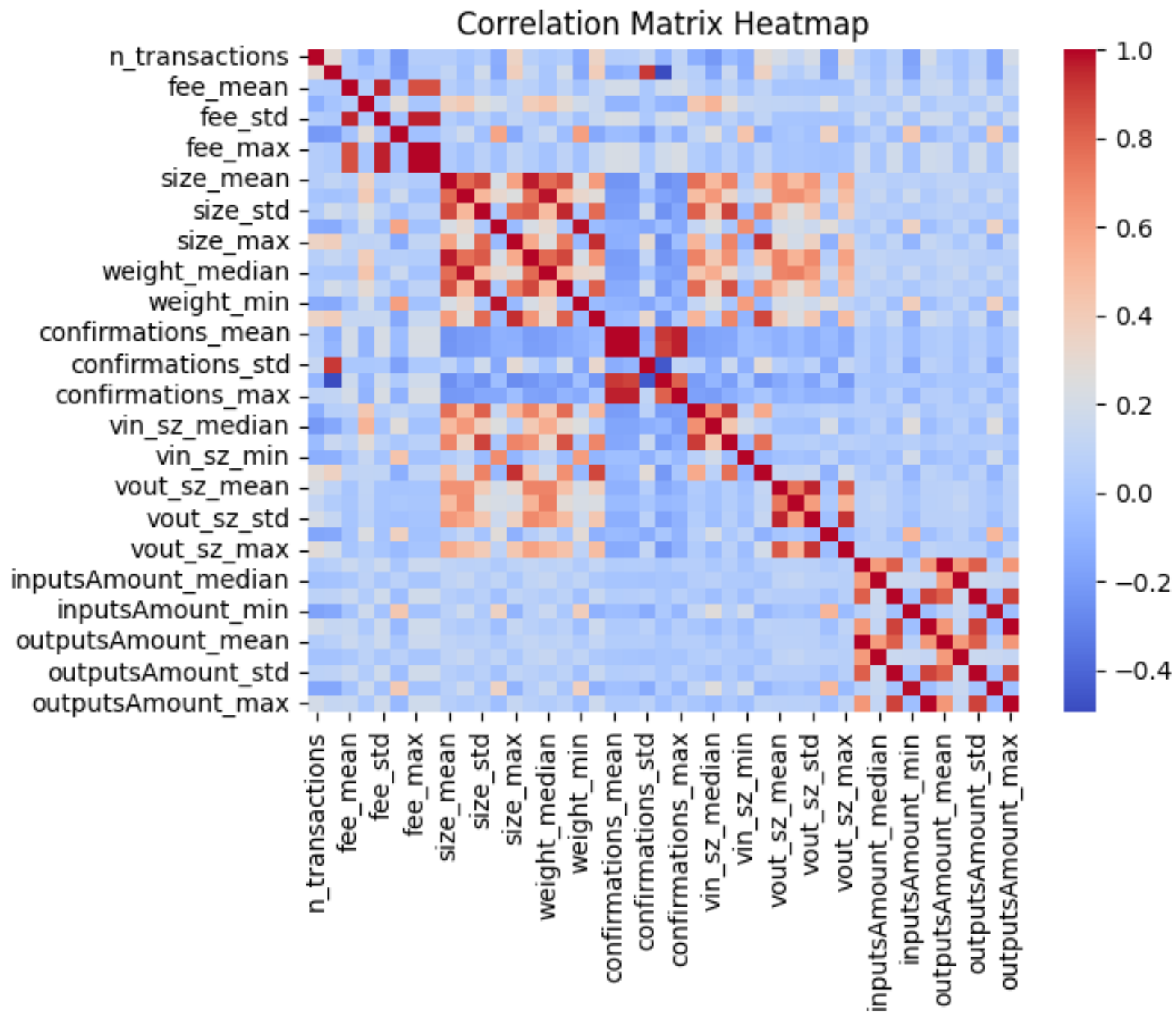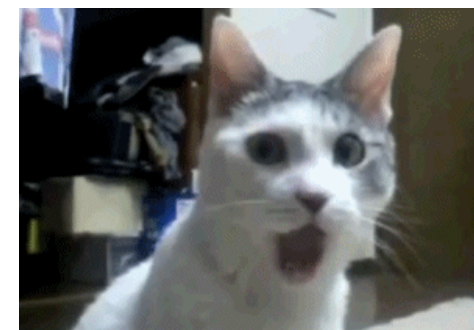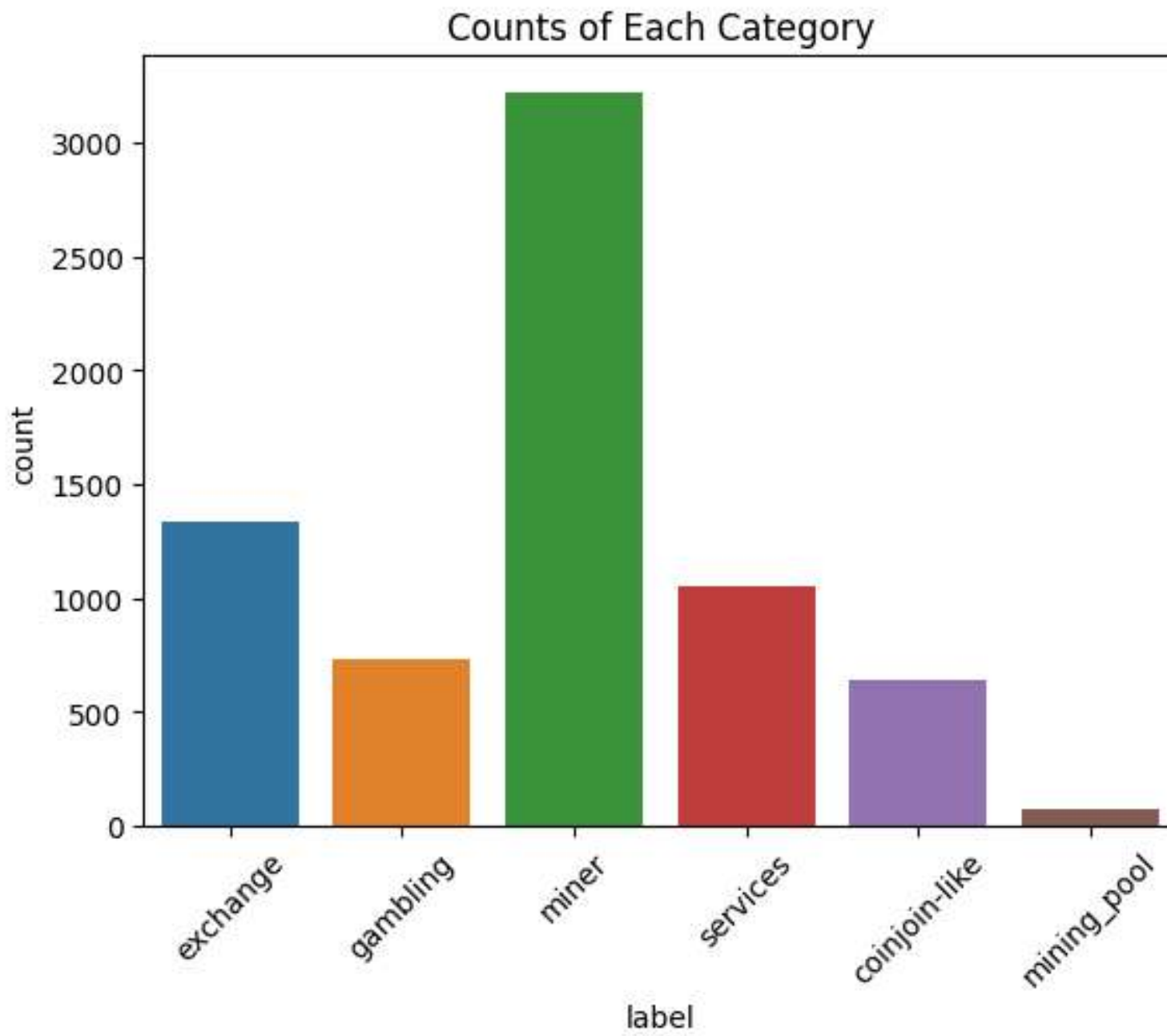
Correlation Matrix Heatmap

DATA PREPROCESSING

Correlation matrix to Heatmap

DATA PREPROCESSING

Counts of Each Category

Class label of the data set

CLASSIFICATION

PYCARET

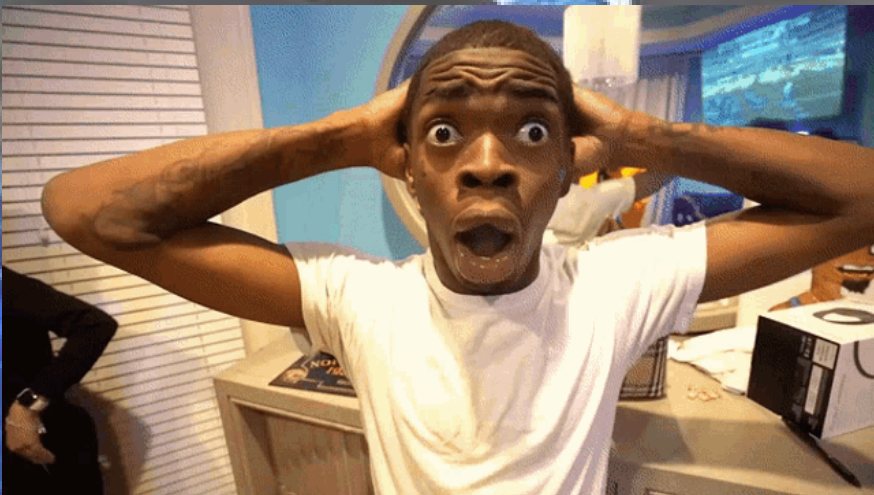Data Preparation | Model Training | Hyperparameter Tuning | Analysis & Interpretability | Model Selection | Experiment Logging

stonks

CLASSIFICATION

```
compare_models()
```

|	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
rf	Random Forest Classifier	0.6132	0.9530	0.6132	0.5431	0.5170	0.3718	0.4580	
knn	K Neighbors Classifier	0.6031	0.7932	0.6031	0.6008	0.5966	0.4336	0.4354	
et	Extra Trees Classifier	0.5818	0.9405	0.5818	0.7383	0.4570	0.3299	0.4015	
ridge	Ridge Classifier	0.5720	0.0000	0.5720	0.6152	0.4991	0.3211	0.3608	
gbc	Gradient Boosting Classifier	0.4631	0.6315	0.4631	0.2204	0.2953	0.0156	0.0920	
lightgbm	Light Gradient Boosting Machine	0.4612	0.8109	0.4612	0.2170	0.2927	0.0108	0.0639	
dt	Decision Tree Classifier	0.4572	0.5000	0.4572	0.2090	0.2869	0.0000	0.0000	
qda	Quadratic Discriminant Analysis	0.4572	0.5000	0.4572	0.2090	0.2869	0.0000	0.0000	
ada	Ada Boost Classifier	0.4572	0.5000	0.4572	0.2090	0.2869	0.0000	0.0000	
lda	Linear Discriminant Analysis	0.4572	0.5000	0.4572	0.2090	0.2869	0.0000	0.0000	
dummy	Dummy Classifier	0.4572	0.5000	0.4572	0.2090	0.2869	0.0000	0.0000	
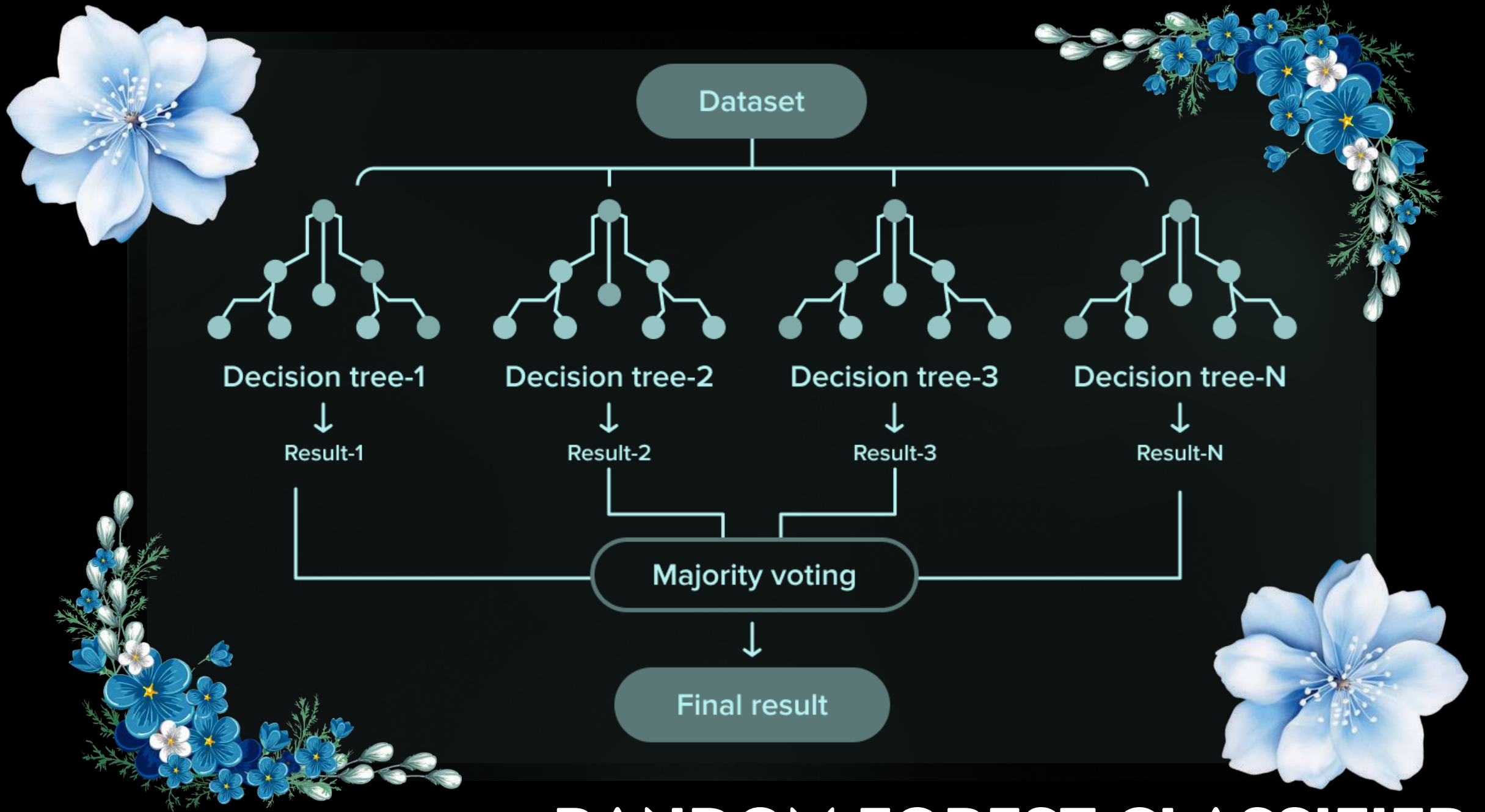
RANDOM FOREST CLASSIFIER

Encoding
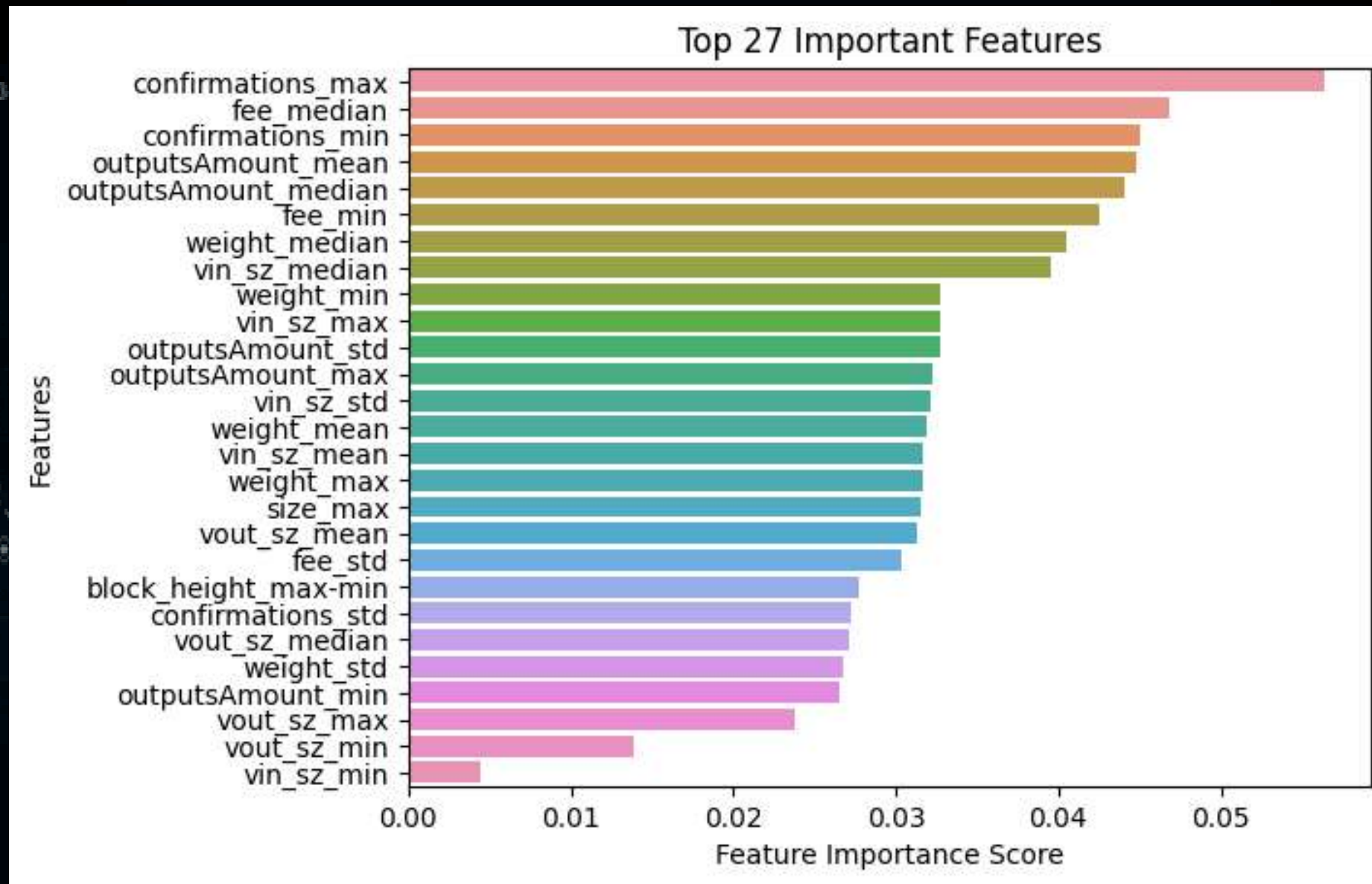
```
0        1
1        2
2        2
3        1
4        3
        ..
7041     5
7042     1
7043     2
7044     1
7045     3
Name: label_encoded, Length: 7041, dtype: int64
```

RANDOM FOREST CLASSIFIER

OVER

SAMPLING
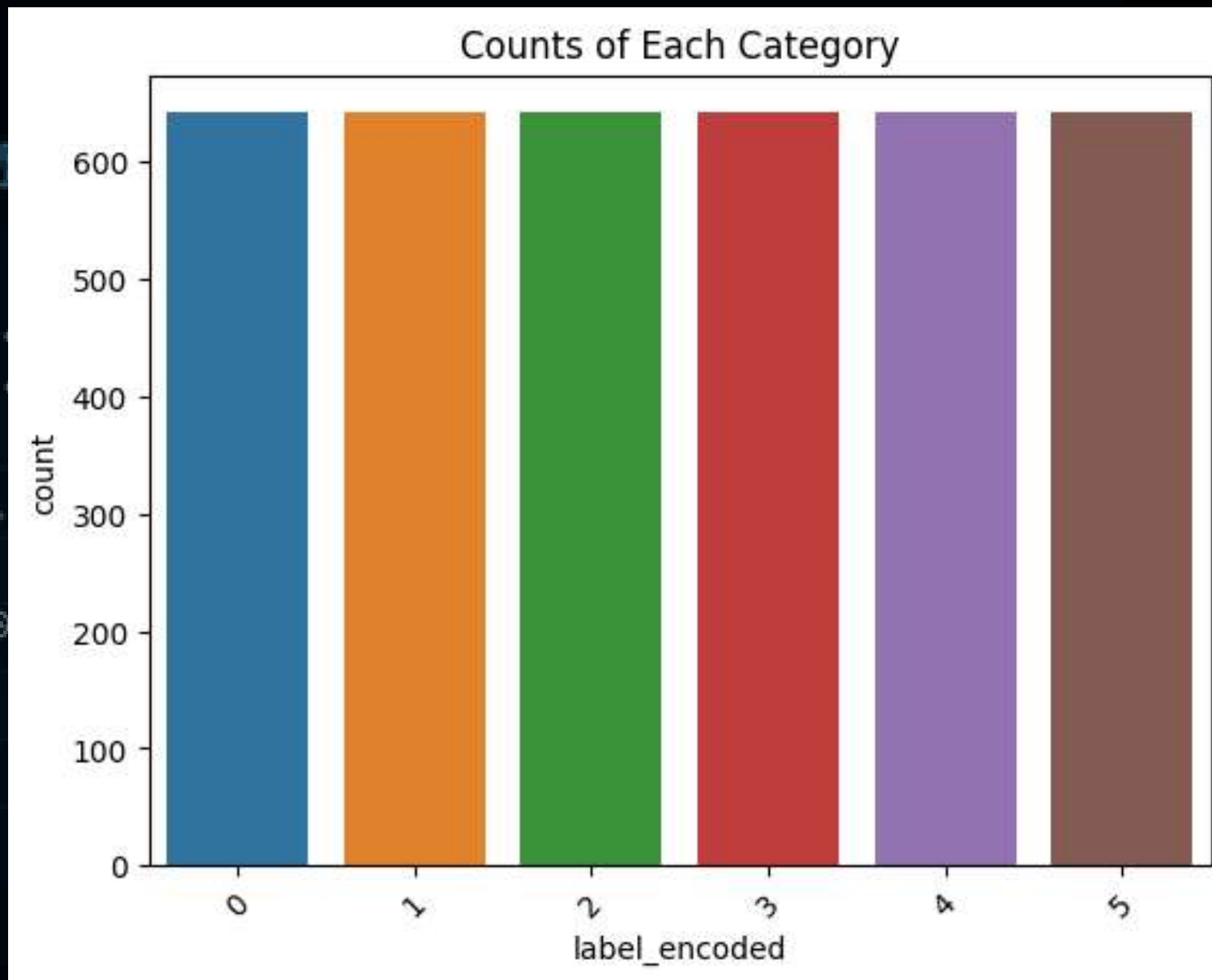
Counts of Each Category

RANDOM FOREST CLASSIFIER

Top 27 Important Features

RANDOM FOREST CLASSIFIER

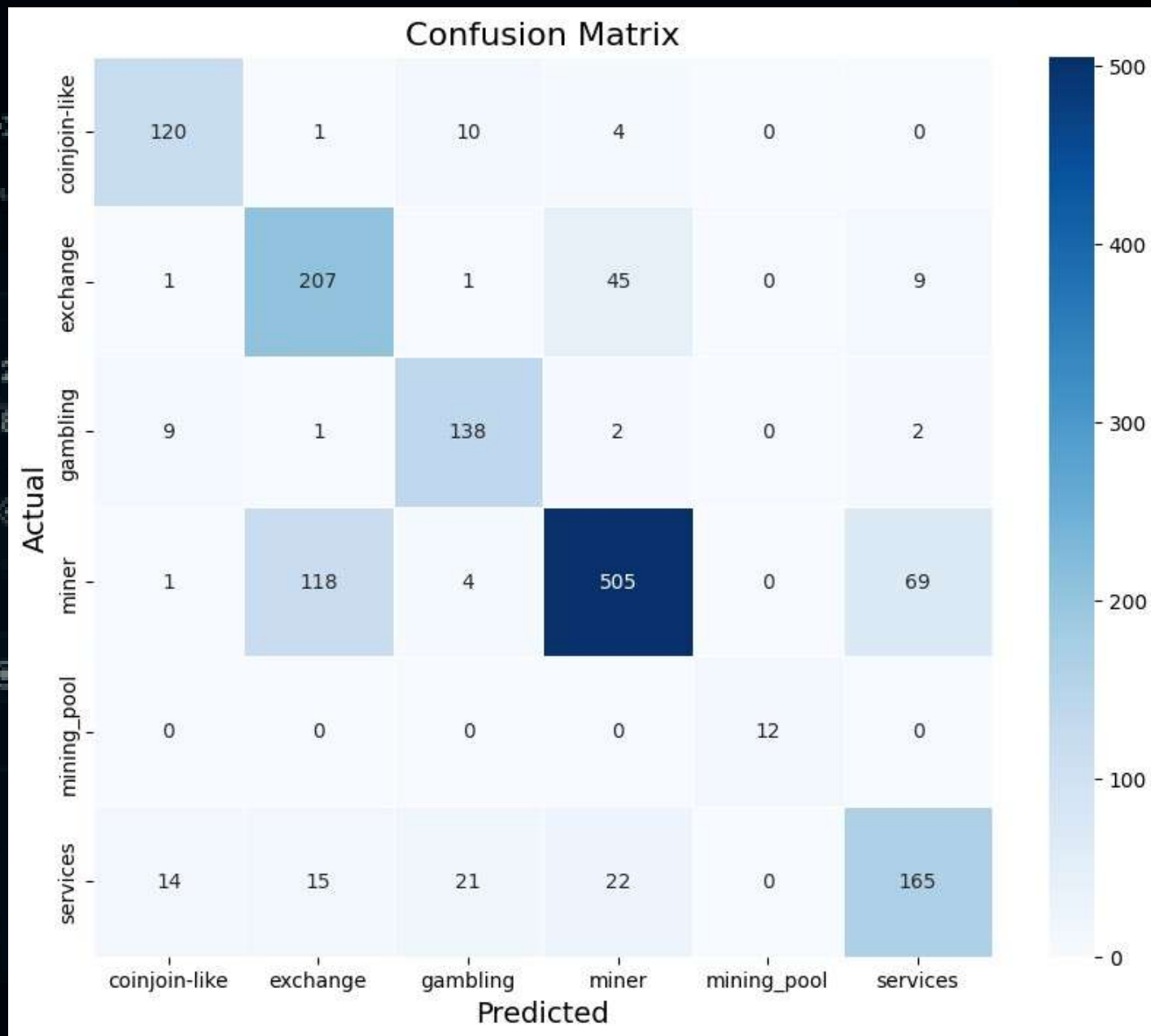# 2nd Encoding

```
0        1
1        2
2        2
3        1
4        3
        ..
7041     5
7042     1
7043     2
7044     1
7045     3
Name: label_encoded, Length: 7041, dtype: int64
```
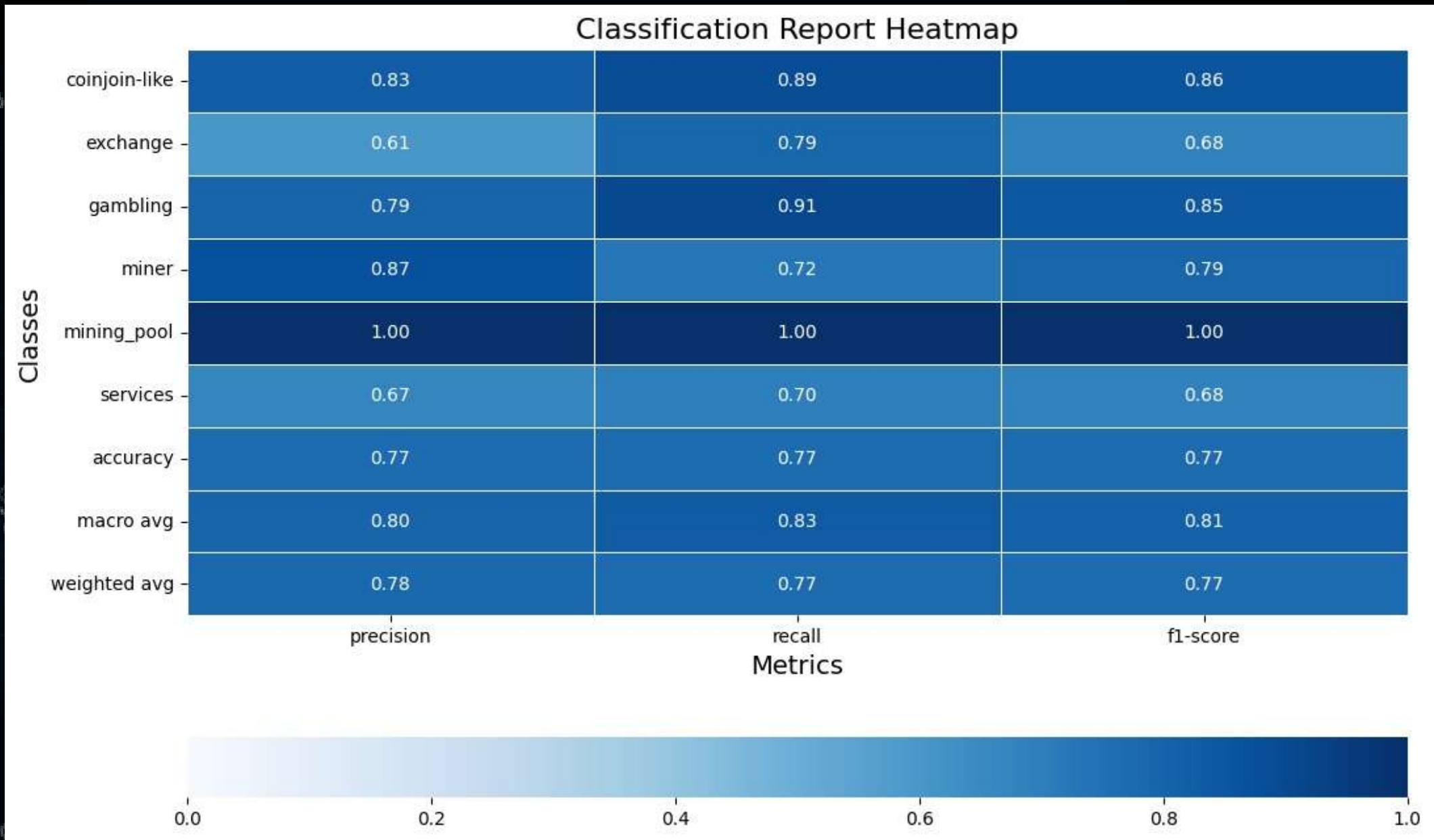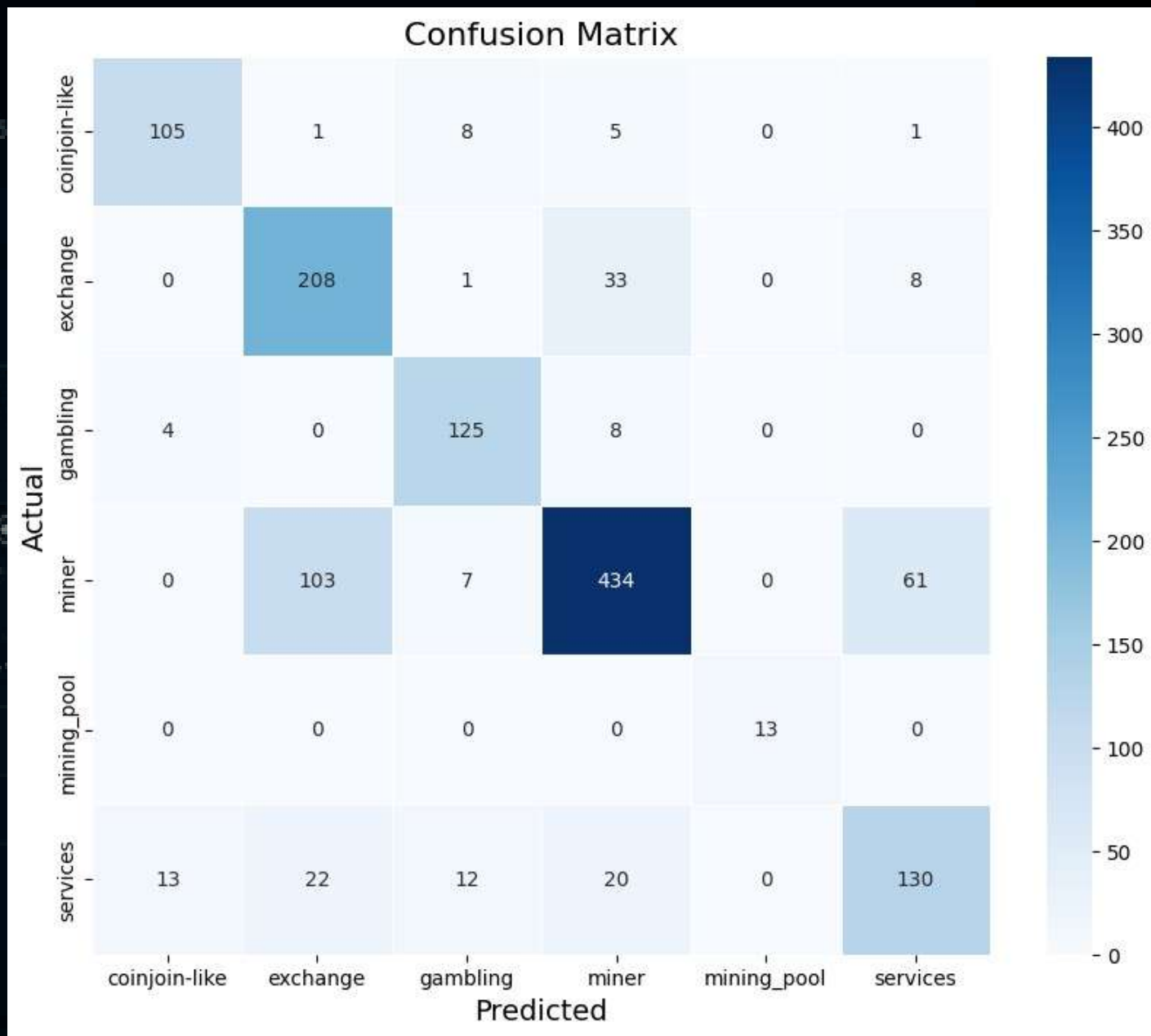
RANDOM FOREST CLASSIFIER

Confusion Matrix

TRAINING DATA

RANDOM FOREST CLASSIFIER

Classification Report Heatmap

Classes	precision	recall	f1-score
coinjoin-like	0.83	0.89	0.86
exchange	0.61	0.79	0.68
gambling	0.79	0.91	0.85
miner	0.87	0.72	0.79
mining_pool	1.00	1.00	1.00
services	0.67	0.70	0.68
accuracy	0.77	0.77	0.77
macro avg	0.80	0.83	0.81
weighted avg	0.78	0.77	0.77

Metrics

TRAINING DATA

RANDOM FOREST CLASSIFIER

TESTING DATA

Confusion Matrix

	coinjoin-like	exchange	gambling	miner	mining_pool	services
coinjoin-like	105	1	8	5	0	1
exchange	0	208	1	33	0	8
gambling	4	0	125	8	0	0
miner	0	103	7	434	0	61
mining_pool	0	0	0	0	13	0
services	13	22	12	20	0	130

Actual / Predicted

RANDOM FOREST CLASSIFIER

Classification Report Heatmap

Classes	precision	recall	f1-score
coinjoin-like	0.86	0.88	0.87
exchange	0.62	0.83	0.71
gambling	0.82	0.91	0.86
miner	0.87	0.72	0.79
mining_pool	1.00	1.00	1.00
services	0.65	0.66	0.65
accuracy	0.77	0.77	0.77
macro avg	0.80	0.83	0.81
weighted avg	0.78	0.77	0.77

Metrics

TESTING DATA

RANDOM FOREST CLASSIFIER

accuracy: 65.77% +/- 0.99% (micro average: 65.77%)

	true exchange	true gambling	true miner	true services	true coinjoin-like	true mining_pool	class precision
pred. exchange	16	0	0	0	0	1	94.12%
pred. gambling	2	580	2	11	100	0	83.45%
pred. miner	1307	128	3218	743	35	0	59.25%
pred. services	2	2	3	250	3	0	96.15%
pred. coinjoin-like	5	18	1	45	501	1	87.74%
pred. mining_pool	1	1	0	0	1	69	95.83%
class recall	1.20%	79.56%	99.81%	23.83%	78.28%	97.18%	

RAPID MINER
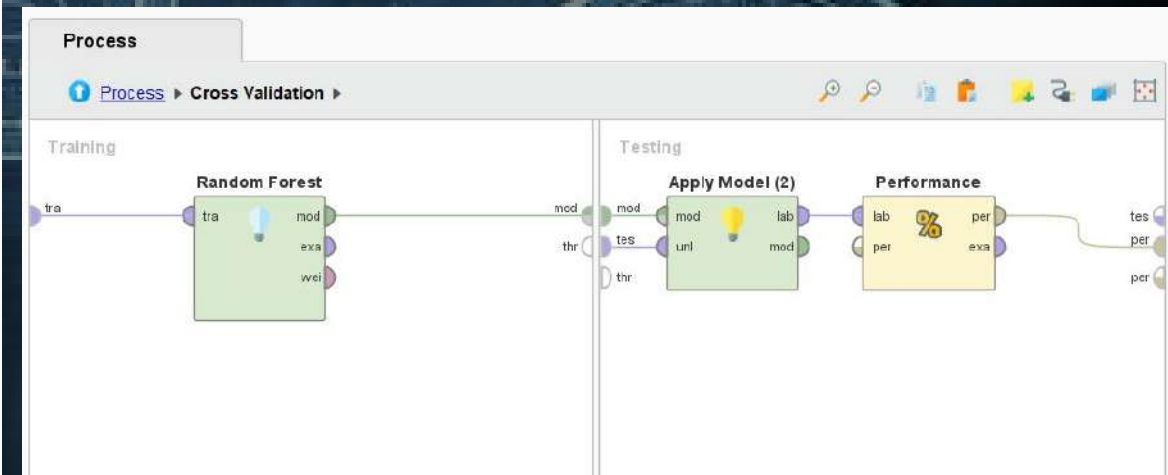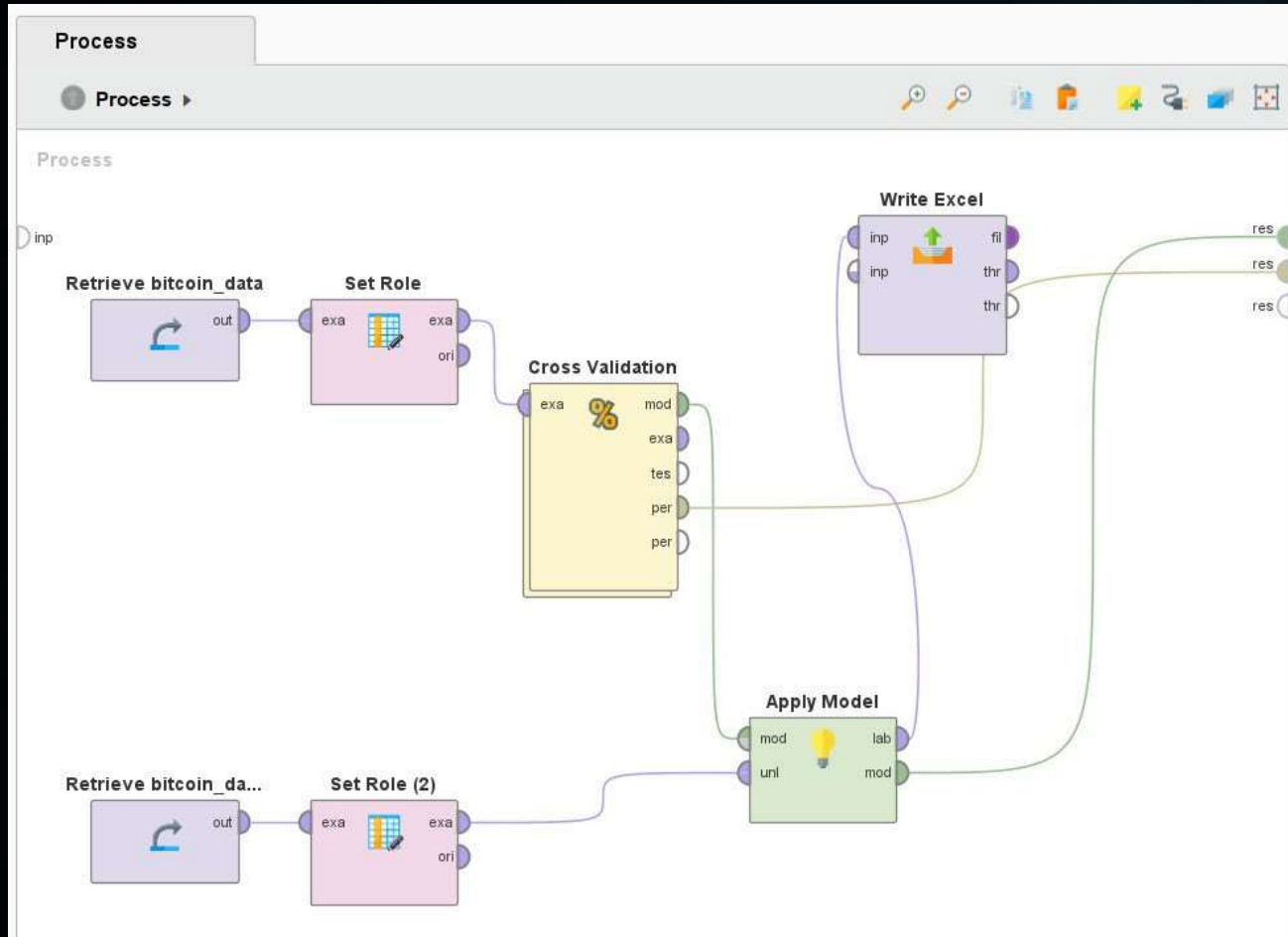
RANDOM FOREST CLASSIFIER

DECISION TREE CLASSIFIER

DECISION TREE CLASSIFIER

Encoding
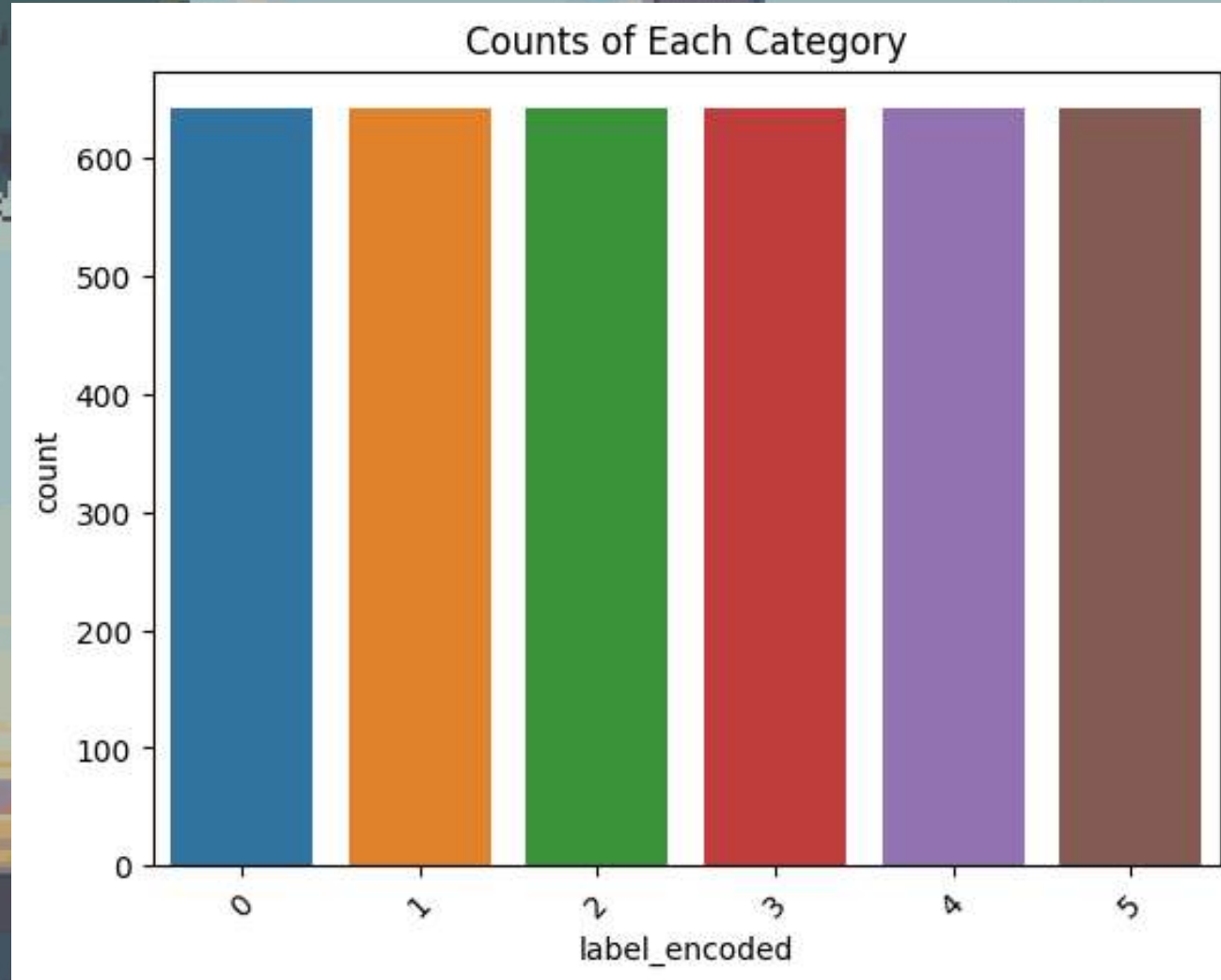
```
0       1
1       2
2       2
3       1
4       3
       ..
7041    5
7042    1
7043    2
7044    1
7045    3
Name: label_encoded, Length: 7041, dtype: int64
```
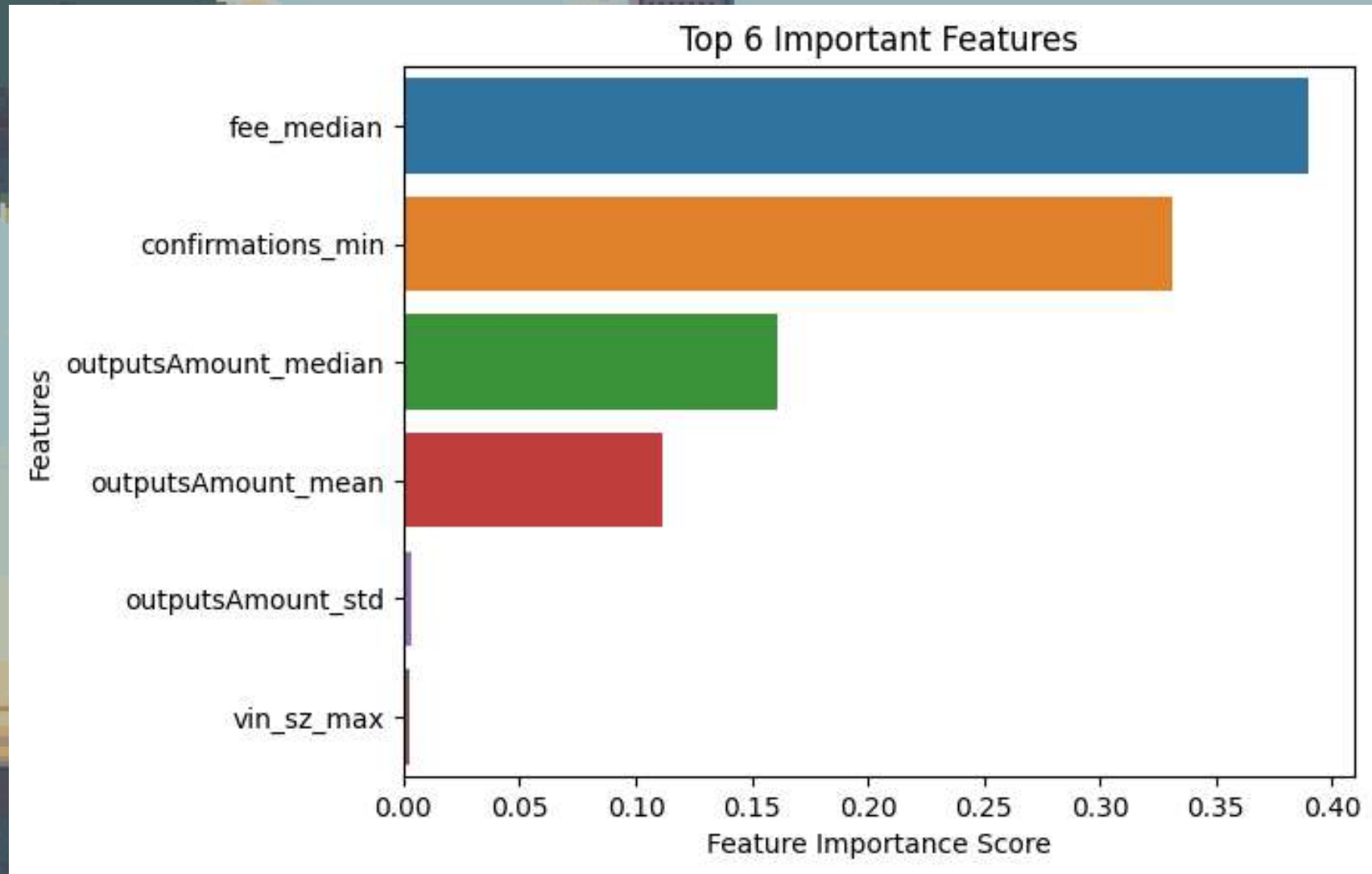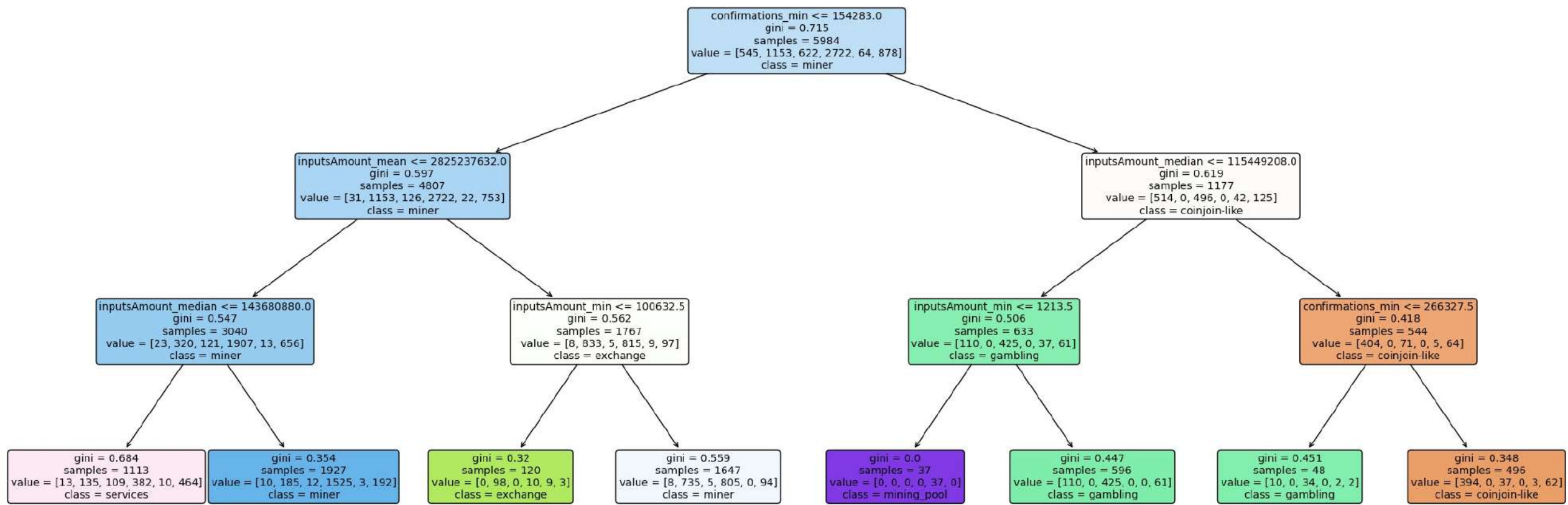
DECISION TREE CLASSIFIER

Top 6 Important Features

DECISION TREE CLASSIFIER

2nd Encoding

```
0        1
1        2
2        2
3        1
4        3
        ..
7041     5
7042     1
7043     2
7044     1
7045     3
Name: label_encoded, Length: 7041, dtype: int64
```
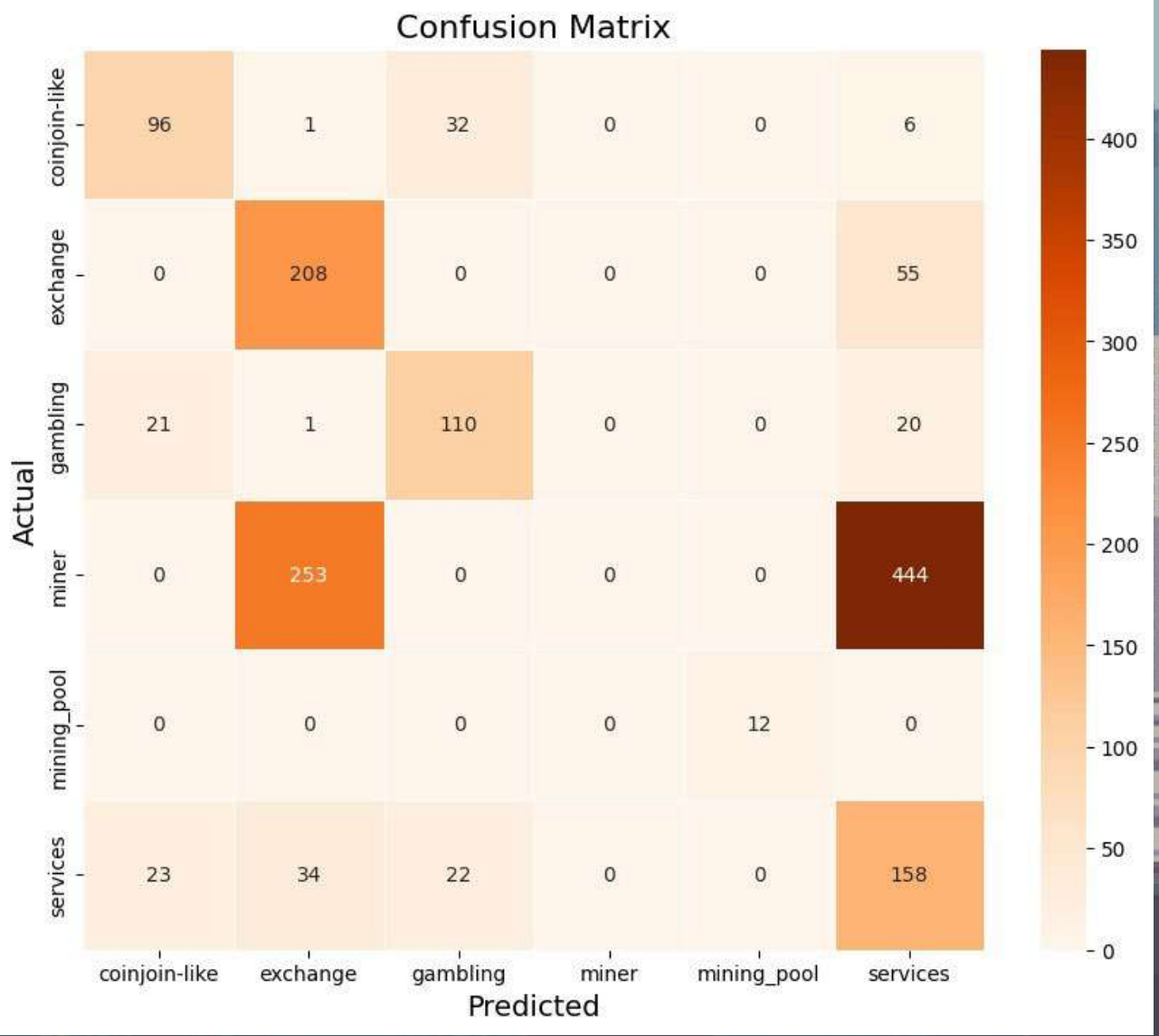
DECISION TREE CLASSIFIER
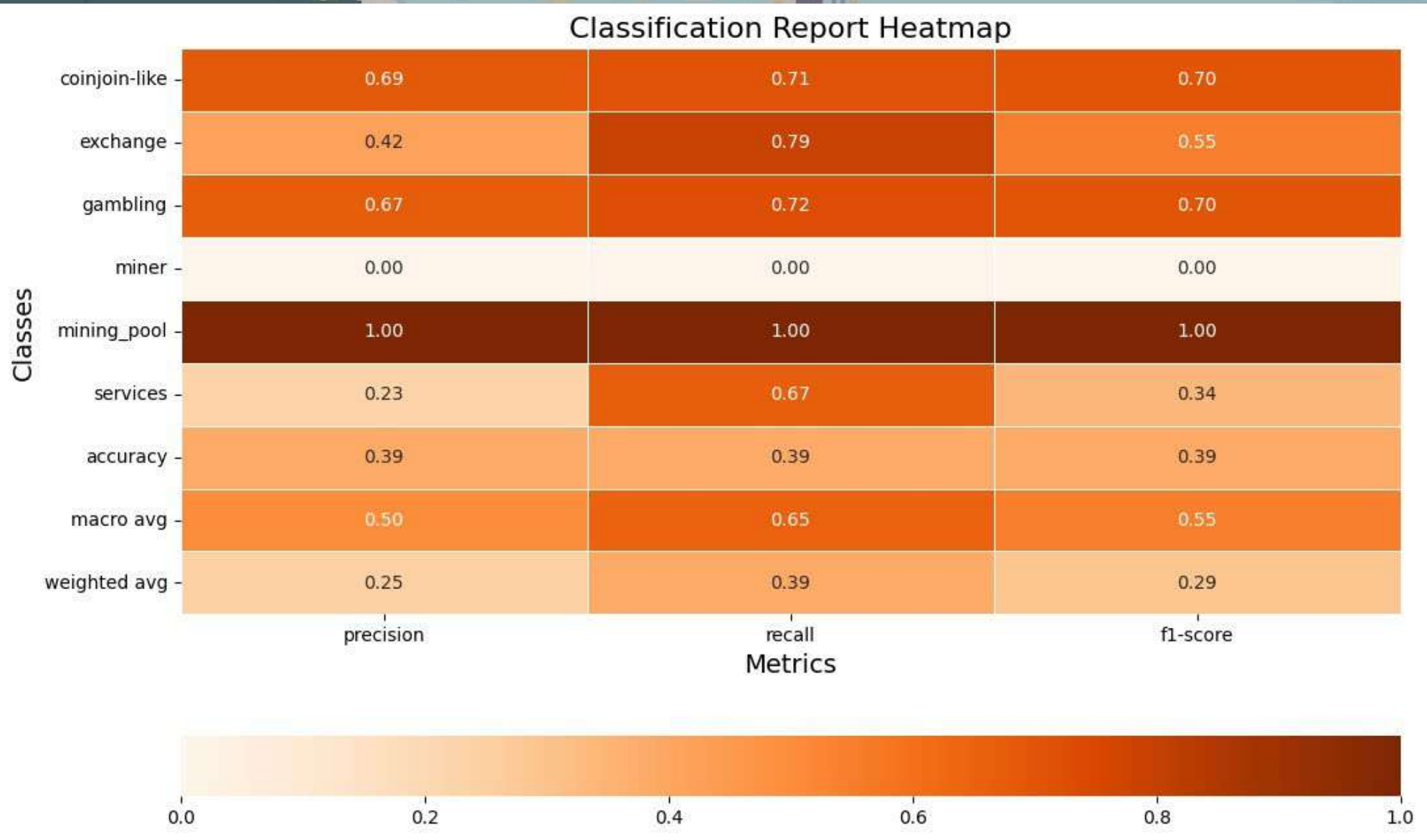

Decision Tree - Gini Criterion
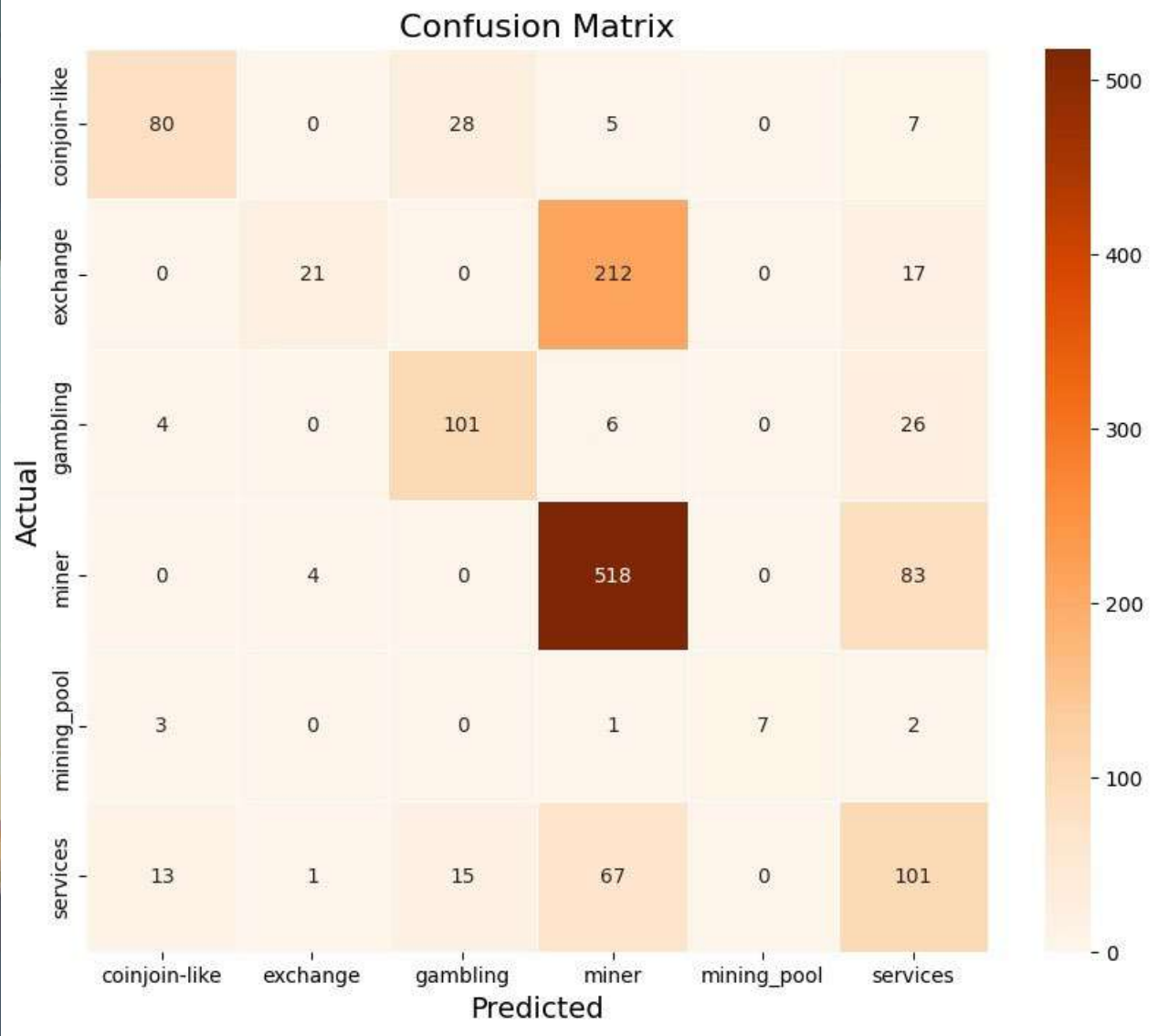
DECISION TREE CLASSIFIER

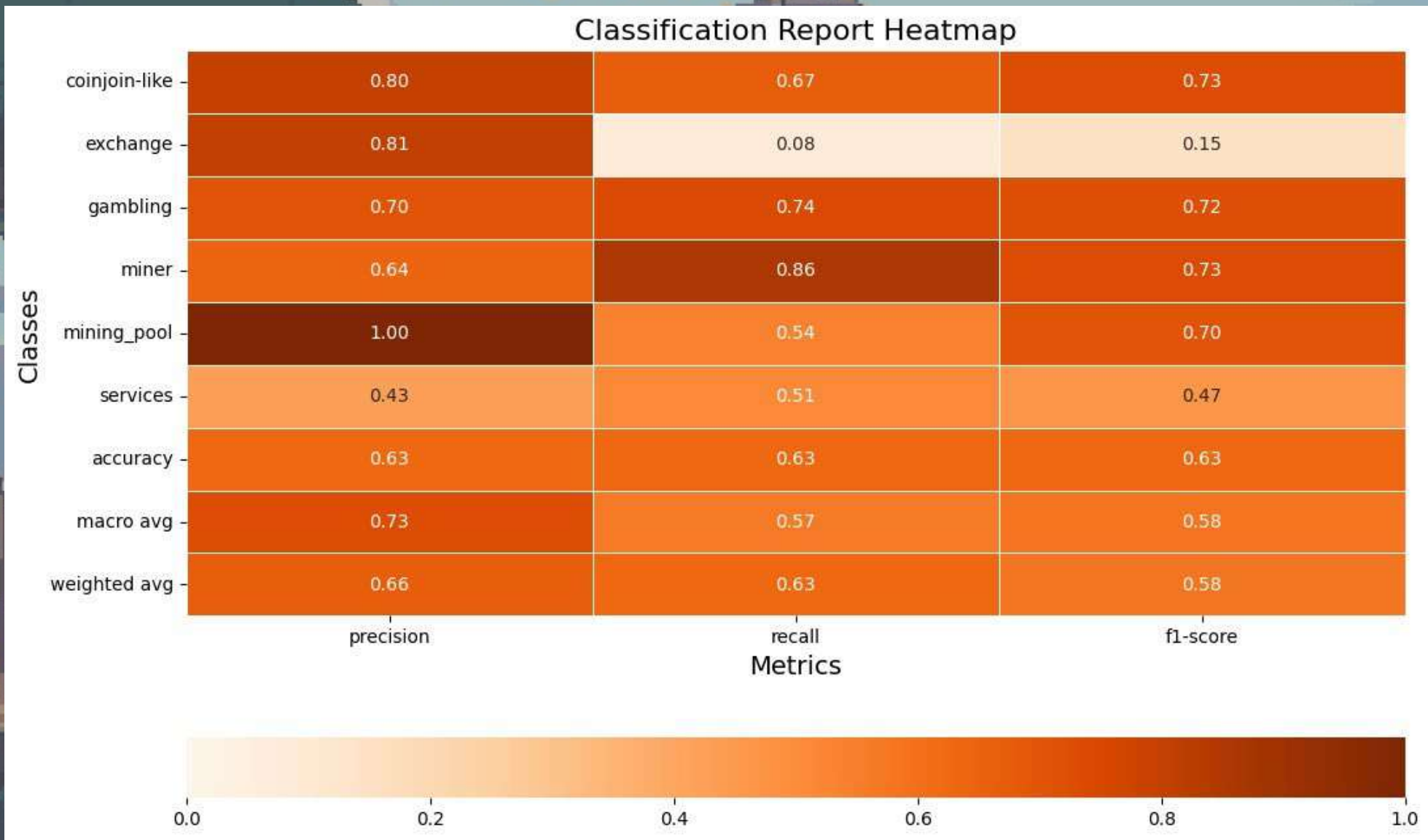Confusion Matrix

TRAINING DATA

DECISION TREE CLASSIFIER

Classification Report Heatmap

Classes	precision	recall	f1-score
coinjoin-like	0.69	0.71	0.70
exchange	0.42	0.79	0.55
gambling	0.67	0.72	0.70
miner	0.00	0.00	0.00
mining_pool	1.00	1.00	1.00
services	0.23	0.67	0.34
accuracy	0.39	0.39	0.39
macro avg	0.50	0.65	0.55
weighted avg	0.25	0.39	0.29

Metrics

DECISION TREE CLASSIFIER

Confusion Matrix

DECISION TREE CLASSIFIER

Classification Report Heatmap

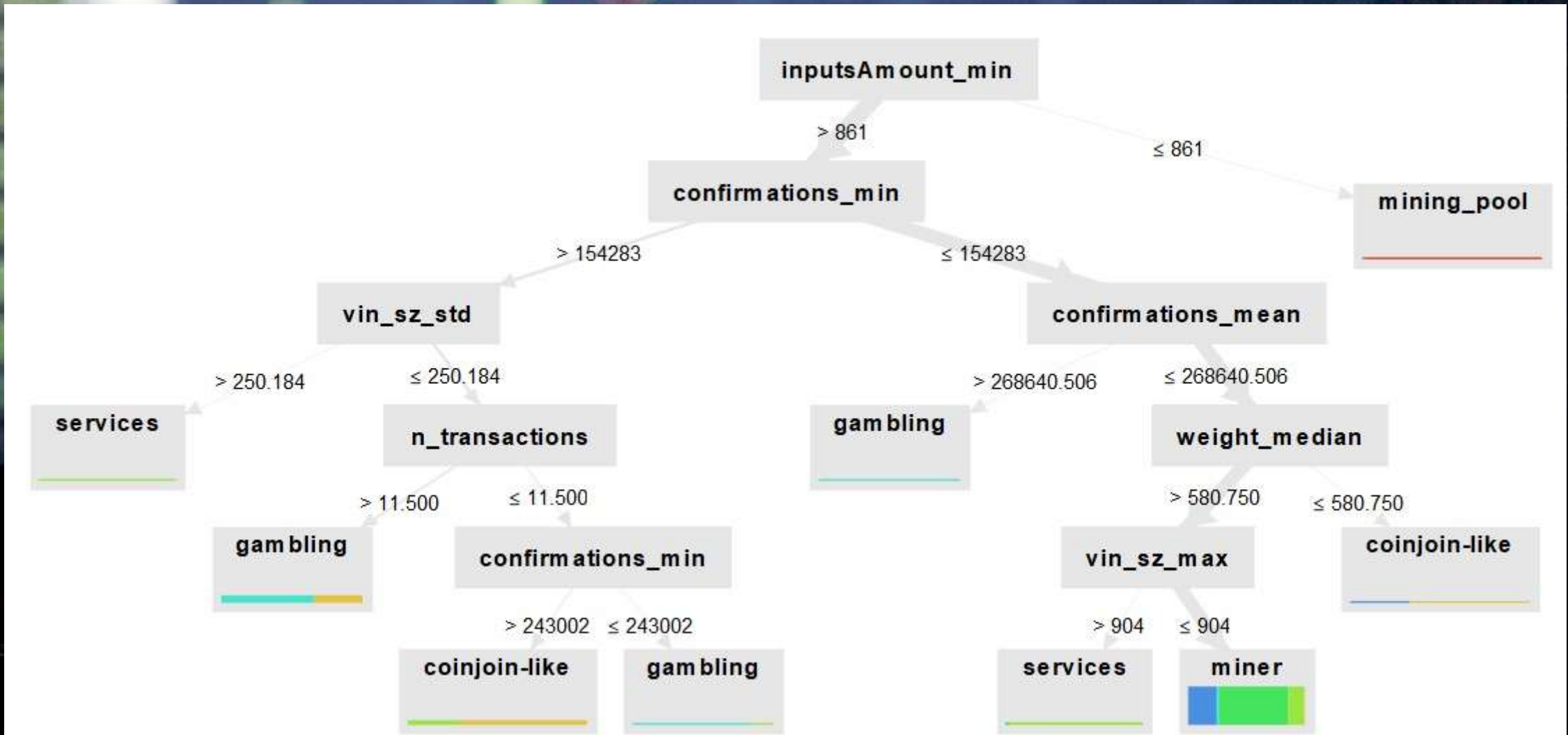| Classes	precision	recall	f1-score
coinjoin-like	0.80	0.67	0.73
exchange	0.81	0.08	0.15
gambling	0.70	0.74	0.72
miner	0.64	0.86	0.73
mining_pool	1.00	0.54	0.70
services	0.43	0.51	0.47
accuracy	0.63	0.63	0.63
macro avg	0.73	0.57	0.58
weighted avg	0.66	0.63	0.58

Metrics

DECISION TREE CLASSIFIER

accuracy: 61.24% +/- 1.88% (micro average: 61.24%)

	true exchange	true gambling	true miner	true services	true coinjoin-like	true mining_pool	class precision
pred. exchange	0	0	0	1	0	0	0.00%
pred. gambling	1	586	1	7	293	1	65.92%
pred. miner	1327	132	3218	785	37	0	58.52%
pred. services	0	2	5	132	1	0	94.29%
pred. coinjoin-like	4	9	0	124	309	0	69.28%
pred. mining_pool	1	0	0	0	0	70	98.59%
class recall	0.00%	80.38%	99.81%	12.58%	48.28%	98.59%	

DECISION TREE CLASSIFIER

DECISION TREE CLASSIFIER

RAPID MINER