


```
!pip install pandas numpy lightgbm seaborn matplotlib scikit-learn imblearn joblib streamlit geopy
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (2.0.2)
Requirement already satisfied: lightgbm in /usr/local/lib/python3.11/dist-packages (4.5.0)
Requirement already satisfied: seaborn in /usr/local/lib/python3.11/dist-packages (0.13.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (3.10.0)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: imblearn in /usr/local/lib/python3.11/dist-packages (0.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (1.5.1)
Requirement already satisfied: streamlit in /usr/local/lib/python3.11/dist-packages (1.46.1)
Requirement already satisfied: geopy in /usr/local/lib/python3.11/dist-packages (2.4.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: scipy in /usr/local/lib/python3.11/dist-packages (from lightgbm) (1.15.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (4.58.4)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.3)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: imbalanced-learn in /usr/local/lib/python3.11/dist-packages (from imblearn) (0.13.0)
Requirement already satisfied: altair<6,>=4.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (5.5.0)
Requirement already satisfied: blinker<2,>=1.5.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (1.9.0)
Requirement already satisfied: cachetools<7,>=4.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (5.5.2)
Requirement already satisfied: click<9,>=7.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (8.2.1)
Requirement already satisfied: protobuf<7,>=3.20 in /usr/local/lib/python3.11/dist-packages (from streamlit) (5.29.5)
Requirement already satisfied: pyarrow>=7.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (18.1.0)
Requirement already satisfied: requests<3,>=2.27 in /usr/local/lib/python3.11/dist-packages (from streamlit) (2.32.3)
Requirement already satisfied: tenacity<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (8.5.0)
Requirement already satisfied: toml<2,>=0.10.1 in /usr/local/lib/python3.11/dist-packages (from streamlit) (0.10.2)
Requirement already satisfied: typing-extensions<5,>=4.4.0 in /usr/local/lib/python3.11/dist-packages (from streamlit) (4.14.0)
Requirement already satisfied: watchdog<7,>=2.1.5 in /usr/local/lib/python3.11/dist-packages (from streamlit) (6.0.0)
Requirement already satisfied: gitpython!=3.1.19,<4,>=3.0.7 in /usr/local/lib/python3.11/dist-packages (from streamlit) (3.1.44)
Requirement already satisfied: pydeck<1,>=0.8.0b4 in /usr/local/lib/python3.11/dist-packages (from streamlit) (0.9.1)
Requirement already satisfied: tornado!=6.5.0,<7,>=6.0.3 in /usr/local/lib/python3.11/dist-packages (from streamlit) (6.4.2)
Requirement already satisfied: geographiclib<3,>=1.52 in /usr/local/lib/python3.11/dist-packages (from geopy) (2.0)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.11/dist-packages (from altair<6,>=4.0->streamlit) (3.1.6)
Requirement already satisfied: jsonschema>=3.0 in /usr/local/lib/python3.11/dist-packages (from altair<6,>=4.0->streamlit) (4.24.0)
Requirement already satisfied: narwhals>=1.14.2 in /usr/local/lib/python3.11/dist-packages (from altair<6,>=4.0->streamlit) (1.44.0)
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.11/dist-packages (from gitpython!=3.1.19,<4,>=3.0.7->streamlit) (4.0.12)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.27->streamlit) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.27->streamlit) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.27->streamlit) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.27->streamlit) (2025.6.15)
Requirement already satisfied: sklearn-compat<1,>=0.1 in /usr/local/lib/python3.11/dist-packages (from imbalanced-learn->imblearn) (0.1.3)
Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.11/dist-packages (from gitdb<5,>=4.0.1->gitpython!=3.1.19,<4,>=3.0.7->streamlit) (5.0.2)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2->altair<6,>=4.0->streamlit) (3.0.2)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->streamlit) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->streamlit) (2025.4.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->streamlit) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from jsonschema>=3.0->altair<6,>=4.0->streamlit) (0.25.1)
```

```
import pandas as pd
import numpy as np
import lightgbm as lgb
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification_report, roc_auc_score, confusion_matrix, roc_curve, auc
from imblearn.over_sampling import SMOTE
```

```
import joblib
import streamlit as st
from geopy.distance import geodesic
```

```
df = pd.read_csv('/content/dataset.csv')
df.head (10)
```

	Unnamed: 0	trans_date_trans_time	cc_num	merchant	category	amt	first	last	gender	street	...	lat	long	city_pop	job	dob	
0	0	2019-01-01 00:00:18	2703186189652095	fraud_Rippin, Kub and Mann	misc_net	4.97	Jennifer	Banks	F	561 Perry Cove	...	36.0788	-81.1781	3495	Psychologist, counselling	1988-03-09	0b242abb623afc5
1	1	2019-01-01 00:00:44	630423337322	fraud_Heller, Gutmann and Zieme	grocery_pos	107.23	Stephanie	Gill	F	43039 Riley Greens Suite 393	...	48.8878	-118.2105	149	Special educational needs teacher	1978-06-21	1f76529f8574734f
2	2	2019-01-01 00:00:51	38859492057661	fraud_Lind-Buckridge	entertainment	220.11	Edward	Sanchez	M	594 White Dale Suite 530	...	42.1808	-112.2620	4154	Nature conservation officer	1962-01-19	a1a22d70485983e
3	3	2019-01-01 00:01:16	3534093764340240	fraud_Kutch, Hermiston and Farrell	gas_transport	45.00	Jeremy	White	M	9443 Cynthia Court Apt. 038	...	46.2306	-112.1138	1939	Patent attorney	1967-01-12	6b849c168bdad6ff
4	4	2019-01-01 00:03:06	375534208663984	fraud_Keeling-Crist	misc_pos	41.96	Tyler	Garcia	M	408 Bradley Rest	...	38.4207	-79.4629	99	Dance movement psychotherapist	1986-03-28	a41d7549acf9078f
5	5	2019-01-01 00:04:08	4767265376804500	fraud_Stroman, Hudson and Erdman	gas_transport	94.63	Jennifer	Conner	F	4655 David Island	...	40.3750	-75.2045	2158	Transport planner	1961-06-19	189a841a0a8ba03
6	6	2019-01-01 00:04:42	30074693890476	fraud_Rowe-Vandervort	grocery_net	44.54	Kelsey	Richards	F	889 Sarah Station Suite 624	...	37.9931	-100.9893	2691	Arboriculturist	1993-08-16	83ec1cc84142af6
7	7	2019-01-01 00:05:08	6011360759745864	fraud_Corwin-Collins	gas_transport	71.65	Steven	Williams	M	231 Flores Pass Suite 720	...	38.8432	-78.6003	6018	Designer, multimedia	1947-08-21	6d294ed2cc447d2
8	8	2019-01-01 00:05:18	4922710831011201	fraud_Herzog Ltd	misc_pos	4.27	Heather	Chase	F	6888 Hicks Stream Suite 954	...	40.3359	-79.6607	1472	Public affairs consultant	1941-03-07	fc28024ce480f8e
9	9	2019-01-01 00:06:01	2720830304681674	fraud_Schoen, Kuphal and Nitzsche	grocery_pos	198.39	Melissa	Aguilar	F	21326 Taylor Squares Suite 708	...	36.5220	-87.3490	151785	Pathologist	1974-03-28	3b9014ea8fb80bd

10 rows × 23 columns

```
df ['trans_date_trans_time'] = pd.to_datetime(df['trans_date_trans_time'])
```

```
df ['hour'] = df['trans_date_trans_time'].dt.hour
df ['day'] = df['trans_date_trans_time'].dt.day
df ['month'] = df['trans_date_trans_time'].dt.month
```

```
drop_columns = ['Unnamed: 0', 'trans_date_trans_time','first','last', 'street','city','state','zip','dob','job','trans_num']
df = df.drop(columns=drop_columns)
```

```
df.head (10)
```

	cc_num	merchant	category	amt	gender	lat	long	city_pop	unix_time	merch_lat	merch_long	is_fraud	hour	day	month	
0	2703186189652095	fraud_Rippin, Kub and Mann	misc_net	4.97	F	36.0788	-81.1781	3495	1325376018	36.011293	-82.048315	0	0	1	1	
1	630423337322	fraud_Heller, Gutmann and Zieme	grocery_pos	107.23	F	48.8878	-118.2105	149	1325376044	49.159047	-118.186462	0	0	1	1	
2	38859492057661	fraud_Lind-Buckridge	entertainment	220.11	M	42.1808	-112.2620	4154	1325376051	43.150704	-112.154481	0	0	1	1	
3	3534093764340240	fraud_Kutch, Hermiston and Farrell	gas_transport	45.00	M	46.2306	-112.1138	1939	1325376076	47.034331	-112.561071	0	0	1	1	
4	375534208663984	fraud_Keeling-Crist	misc_pos	41.96	M	38.4207	-79.4629	99	1325376186	38.674999	-78.632459	0	0	1	1	
5	4767265376804500	fraud_Stroman, Hudson and Erdman	gas_transport	94.63	F	40.3750	-75.2045	2158	1325376248	40.653382	-76.152667	0	0	1	1	
6	30074693890476	fraud_Rowe-Vandervort	grocery_net	44.54	F	37.9931	-100.9893	2691	1325376282	37.162705	-100.153370	0	0	1	1	
7	6011360759745864	fraud_Corwin-Collins	gas_transport	71.65	M	38.8432	-78.6003	6018	1325376308	38.948089	-78.540296	0	0	1	1	
8	4922710831011201	fraud_Herzog Ltd	misc_pos	4.27	F	40.3359	-79.6607	1472	1325376318	40.351813	-79.958146	0	0	1	1	
9	2720830304681674	fraud_Schoen, Kuphal and Nitzsche	grocery_pos	198.39	F	36.5220	-87.3490	151785	1325376361	37.179198	-87.485381	0	0	1	1	

```
cat_col = ['merchant', 'category','gender']
encoders = {}
for col in cat_col:
    encoders[col]=LabelEncoder()
    df[col] = encoders[col].fit_transform(df[col])
```

```
def haversine(lat1, lon1, lat2, lon2):
    return np.array([geodesic((a,b),(c,d)).km for a,b,c,d in zip (lat1,lon1,lat2,lon2)])
```

```
df ['distance'] = haversine(df['lat'],df['long'],df['merch_lat'],df['merch_long'])
```

```
df.head()
```

	cc_num	merchant	category	amt	gender	lat	long	city_pop	unix_time	merch_lat	merch_long	is_fraud	hour	day	month	distance	
0	2703186189652095	514	8	4.97	0	36.0788	-81.1781	3495	1325376018	36.011293	-82.048315	0	0	1	1	78.773821	
1	630423337322	241	4	107.23	0	48.8878	-118.2105	149	1325376044	49.159047	-118.186462	0	0	1	1	30.216618	
2	38859492057661	390	0	220.11	1	42.1808	-112.2620	4154	1325376051	43.150704	-112.154481	0	0	1	1	108.102912	
3	3534093764340240	360	2	45.00	1	46.2306	-112.1138	1939	1325376076	47.034331	-112.561071	0	0	1	1	95.685115	
4	375534208663984	297	9	41.96	1	38.4207	-79.4629	99	1325376186	38.674999	-78.632459	0	0	1	1	77.702395	

```
df.tail()
```

	cc_num	merchant	category	amt	gender	lat	long	city_pop	unix_time	merch_lat	merch_long	is_fraud	hour	day	month	distance
1296670	30263540414123	499	0	15.56	1	37.7175	-112.4777	258	1371816728	36.841266	-111.690765	0	12	21	6	119.696415
1296671	6011149206456997	2	1	51.70	1	39.2667	-77.5101	100	1371816739	38.906881	-78.246528	0	12	21	6	75.202184
1296672	3514865930894695	599	1	105.93	1	32.9396	-105.8189	899	1371816752	33.619513	-105.130529	0	12	21	6	98.987927
1296673	2720012583106919	509	1	74.90	1	43.3526	-102.5411	1126	1371816816	42.788940	-103.241160	0	12	21	6	84.688356
1296674	4292902571056973207	370	1	4.30	1	45.8433	-113.8748	218	1371816817	46.565983	-114.186110	0	12	21	6	83.845902

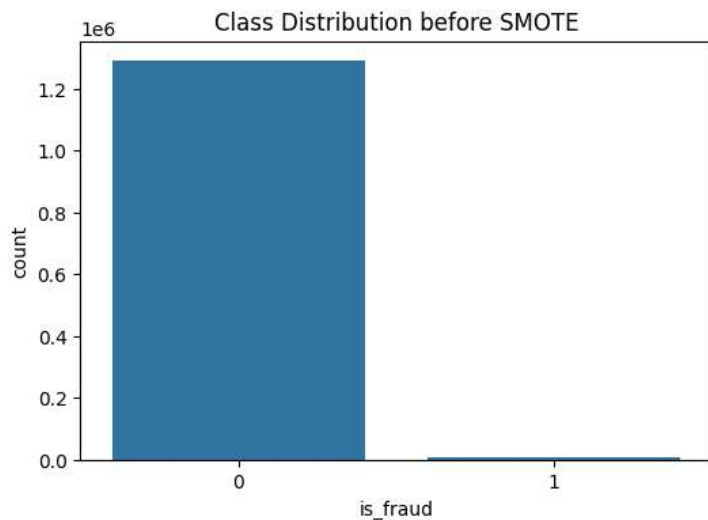
```
features = ['merchant', 'category', 'amt', 'cc_num', 'hour', 'day', 'month', 'gender', 'distance']
x = df [features]
y = df ['is_fraud']
```

```
plt.figure(figsize=(6,4))
sns.countplot(x='is_fraud',data=df)
plt.title('Class Distribution before SMOTE')
plt.show
```

```
matplotlib.pyplot.show
def show(*args, **kwargs) -> None

/usr/local/lib/python3.11/dist-packages/matplotlib/pyplot.py
Display all open figures.

Parameters
-----
block : bool, optional
```



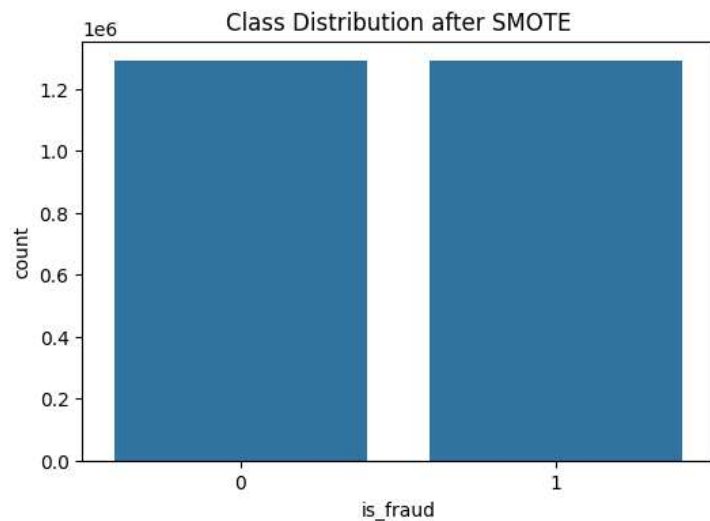
```
smote = SMOTE(random_state=42)
x_resample, y_resample = smote.fit_resample(x,y)
```

```
plt.figure(figsize=(6,4))
sns.countplot(x=y_resample)
plt.title('Class Distribution after SMOTE')
plt.show
```

```
matplotlib.pyplot.show
def show(*args, **kwargs) -> None

/usr/local/lib/python3.11/dist-packages/matplotlib/pyplot.py
Display all open figures.

Parameters
-----
block : bool, optional
```



```
x_train, x_test, y_train, y_test =train_test_split(x_resample,y_resample,test_size=0.2,random_state=42)
```

```
lgb_model = lgb.LGBMClassifier(
    boosting_type='gbdt',
    objective='binary',
    metric='auc',
    is_unbalance=True,
    num_leaves=31,
    max_depth=-1,
    learning_rate=0.05,
    n_estimators=200,
)
lgb_model.fit(x_train,y_train)
```

⌵

LightGBM

[Info] Number of positive: 1030687, number of negative: 1031983

LightGBM

[Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.066247 seconds. You can set `force_row_wise=true` to remove the overhead. And if memory is not enough, you can set `force_col_wise=true`.

LightGBM

[Info] Total Bins 1105

LightGBM

[Info] Number of data points in the train set: 2062670, number of used features: 9

LightGBM

[Info] [binary:BoostFromScore]: pavg=0.499686 -> initscore=-0.001257

LightGBM

[Info] Start training from score -0.001257

⌵

LGBMClassifier

ⓘ

LGBMClassifier(is_unbalance=True, learning_rate=0.05, metric='auc', n_estimators=200, objective='binary')

```
y_pred = lgb_model.predict(x_test)
```

```
print("classification_report: \n",classification_report(y_test,y_pred))
print("ROC AUC Score:", roc_auc_score(y_test,y_pred))
```

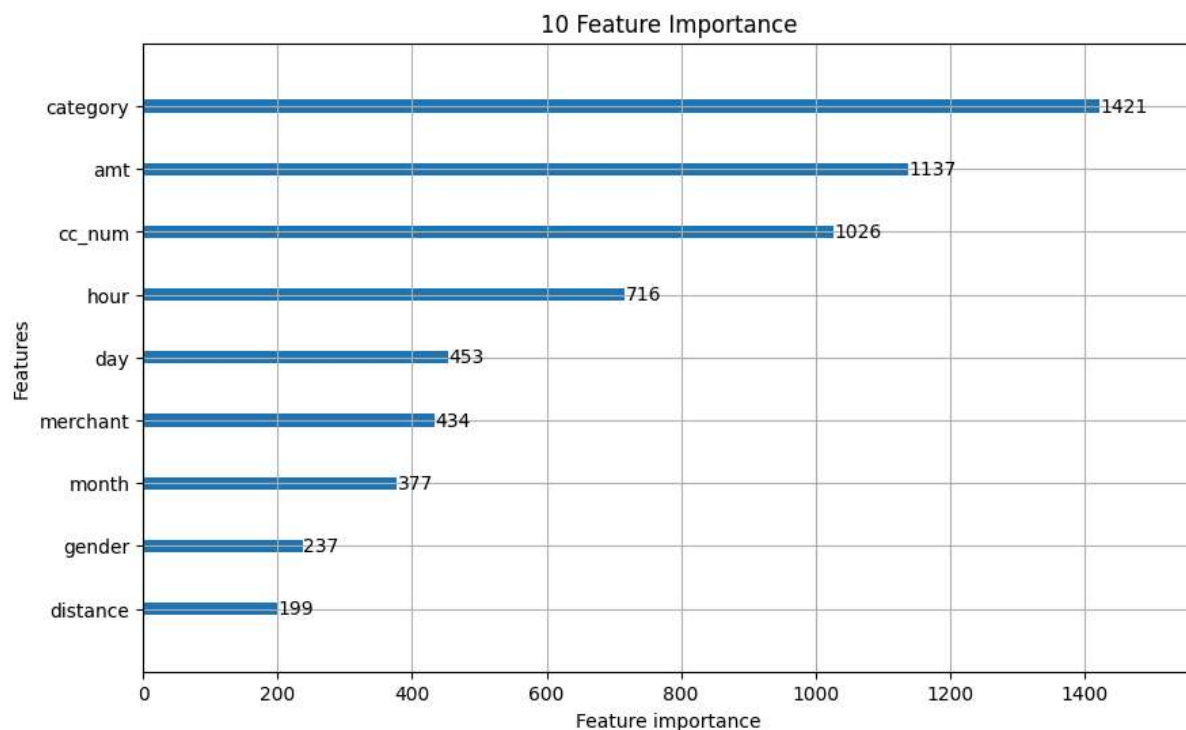
```
↗ classification_report:
      precision    recall  f1-score   support

     0       0.92      0.97      0.95    257186
     1       0.97      0.92      0.94    258482

 accuracy          0.94    515668
 macro avg          0.95    515668
 weighted avg       0.95    515668

ROC AUC Score: 0.9447563205132311
```

```
lgb.plot_importance(lgb_model,figsize=(10,6),max_num_features=10,importance_type='split')
plt.title('10 Feature Importance')
plt.show()
```

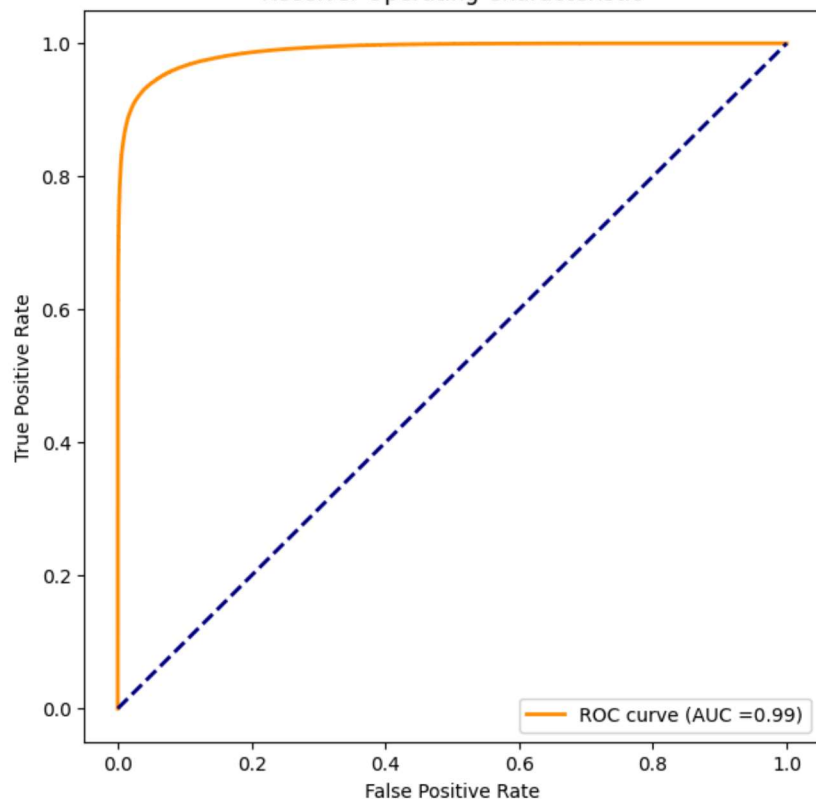


```
fpr, tpr, thresholds = roc_curve(y_test, lgb_model.predict_proba(x_test)[:,:1])
roc_auc = auc(fpr,tpr)
```

```
plt.figure(figsize=(7,7))
plt.plot(fpr,tpr,color='darkorange',lw=2,label=f'ROC curve (AUC ={roc_auc:.2f})')
plt.plot([0,1],[0,1],color='navy',lw=2,linestyle='--')
plt.title('Receiver Operating Characteristic')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.legend(loc="lower right")
plt.show()
```



Receiver Operating Characteristic



```
joblib.dump(lgb_model, 'Creditcard_fraud_detection_model.job')  
joblib.dump(encoders, 'lable_encoder.job')
```



['lable_encoder.job']