credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

Share

+ Code + Text

Connect Gemini

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
```

```
train = pd.read_csv("fraudTrain.csv")
test = pd.read_csv("fraudTest.csv")

data = pd.concat([train, test])
data.describe()
```

	Unnamed: 0	cc_num	amt	zip	lat	long	city_pop	unix_time	merch_lat	merch_long	is_
count	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.852394e+06	1.85239
mean	5.371934e+05	4.173860e+17	7.006357e+01	4.881326e+04	3.853931e+01	-9.022783e+01	8.864367e+04	1.358674e+09	3.853898e+01	-9.022794e+01	5.21001
std	3.669110e+05	1.309115e+18	1.592540e+02	2.688185e+04	5.071470e+00	1.374789e+01	3.014876e+05	1.819508e+07	5.105604e+00	1.375969e+01	7.19921
min	0.000000e+00	6.041621e+10	1.000000e+00	1.257000e+03	2.002710e+01	-1.656723e+02	2.300000e+01	1.325376e+09	1.902742e+01	-1.666716e+02	0.00000
25%	2.315490e+05	1.800429e+14	9.640000e+00	2.623700e+04	3.466890e+01	-9.679800e+01	7.410000e+02	1.343017e+09	3.474012e+01	-9.689944e+01	0.00000
50%	4.630980e+05	3.521417e+15	4.745000e+01	4.817400e+04	3.935430e+01	-8.747690e+01	2.443000e+03	1.357089e+09	3.936890e+01	-8.744069e+01	0.00000
75%	8.335758e+05	4.642255e+15	8.310000e+01	7.204200e+04	4.194040e+01	-8.015800e+01	2.032800e+04	1.374581e+09	4.195626e+01	-8.024511e+01	0.00000

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text

```
print(train.shape)
print(test.shape)
```

```
(1296675, 23)
(555719, 23)
```

```
display(data.head())
print(data.describe())
print(data.isnull().sum())
```

	Unnamed: 0	trans_date_trans_time	cc_num	merchant	category	amt	first	last	gender	street	...	lat	long
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
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
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
[illegible]	[illegible]	[illegible]	[illegible]	fraud_Kutch, [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9443 Cynthia [illegible]	[illegible]	[illegible]	[illegible]

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text Connect Gemini

```
city_pop              0
job                   0
dob                   0
trans_num             0
unix_time             0
merch_lat             0
merch_long            0
is_fraud              0
dtype: int64
```

```
[ ] test.info
```

```
<bound method DataFrame.info of         Unnamed: 0 trans_date_trans_time            cc_num  \
0                0   2020-06-21 12:14:25  2291163933867244
1                1   2020-06-21 12:14:33  3573030041201292
2                2   2020-06-21 12:14:53  3598215285024754
3                3   2020-06-21 12:15:15  3591919803438423
4                4   2020-06-21 12:15:17  3526826139003047
...            ...                   ...               ...
555714      555714   2020-12-31 23:59:07    30560609640617
555715      555715   2020-12-31 23:59:09  3556613125071656
555716      555716   2020-12-31 23:59:15  6011724471098086
555717      555717   2020-12-31 23:59:24     4079773899158
555718      555718   2020-12-31 23:59:34  4170689372027579

                                   merchant        category    amt    first  \
0                     fraud_Kirlin and Sons   personal_care   2.86     Jeff
1                      fraud_Sporer-Keebler   personal_care  29.84   Joanne
2     fraud_Swaniawski, Nitzsche and Welch  health_fitness  41.28   Ashley
3                        fraud_Haley Group        misc_pos  60.05    Brian
4                    fraud_Johnston-Casper          travel   3.19   Nathan
...                                     ...             ...    ...      ...
```

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text

Connect Gemini

```
555714       519              Town planner  1966-02-13
555715     28739            Futures trader  1999-12-27
555716      3684                  Musician  1981-11-29
555717       129              Cartographer  1965-12-15
555718    116001               Media buyer  1993-05-10

                               trans_num   unix_time  merch_lat  merch_long  \
0       2da90c7d74bd46a0caf3777415b3ebd3  1371816865  33.986391  -81.200714
1       324cc204407e99f51b0d6ca0055005e7  1371816873  39.450498 -109.960431
2       c81755dbbbea9d5c77f094348a7579be  1371816893  40.495810  -74.196111
3       2159175b9efe66dc301f149d3d5abf8c  1371816915  28.812398  -80.883061
4       57ff021bd3f328f8738bb535c302a31b  1371816917  44.959148  -85.884734
...                                  ...         ...        ...         ...
555714  9b1f753c79894c9f4b71f04581835ada  1388534347  39.946837  -91.333331
555715  2090647dac2c89a1d86c514c427f5b91  1388534349  29.661049  -96.186633
555716  6c5b7c8add471975aa0fec023b2e8408  1388534355  46.658340 -119.715054
555717  14392d723bb7737606b2700ac791b7aa  1388534364  44.470525 -117.080888
555718  1765bb45b3aa3224b4cdcb6e7a96cee3  1388534374  36.210097  -97.036372

        is_fraud
0              0
1              0
2              0
3              0
4              0
...          ...
555714         0
555715         0
555716         0
555717         0
555718         0

[555719 rows x 23 columns]>
```

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

Share

+ Code + Text

Connect Gemini

```
1296673  fraud_Reinger, Weissnat and Strosin    food_dining   74.90
1296674  fraud_Langosh, Wintheiser and Hyatt    food_dining    4.30

               first       last gender                       street  ...  \
0           Jennifer      Banks      F               561 Perry Cove  ...
1          Stephanie       Gill      F  43039 Riley Greens Suite 393  ...
2             Edward    Sanchez      M      594 White Dale Suite 530  ...
3             Jeremy      White      M    9443 Cynthia Court Apt. 038  ...
4              Tyler     Garcia      M              408 Bradley Rest  ...
...              ...        ...    ...                          ...  ...
1296670         Erik  Patterson      M      162 Jessica Row Apt. 072  ...
1296671      Jeffrey      White      M  8617 Holmes Terrace Suite 651  ...
1296672  Christopher  Castaneda      M     1632 Cohen Drive Suite 639  ...
1296673       Joseph     Murray      M          42933 Ryan Underpass  ...
1296674      Jeffrey      Smith      M          135 Joseph Mountains  ...

             lat      long  city_pop                                 job  \
0        36.0788  -81.1781      3495           Psychologist, counselling
1        48.8878 -118.2105       149  Special educational needs teacher
2        42.1808 -112.2620      4154         Nature conservation officer
3        46.2306 -112.1138      1939                     Patent attorney
4        38.4207  -79.4629        99      Dance movement psychotherapist
...          ...       ...       ...                                 ...
1296670  37.7175 -112.4777       258                        Geoscientist
1296671  39.2667  -77.5101       100   Production assistant, television
1296672  32.9396 -105.8189       899                     Naval architect
1296673  43.3526 -102.5411      1126               Volunteer coordinator
1296674  45.8433 -113.8748       218             Therapist, horticultural

                dob                         trans_num   unix_time  merch_lat  \
0        1988-03-09  0b242abb623afc578575680df30655b9  1325376018  36.011293
1        1978-06-21  1f76529f8574734946361c461b024d99  1325376044  49.159047
2        1962-01-19  a1a22d70485983eac12b5b88dad1cf95  1325376051  43.150704
3        1967-01-12  6b849c168bdad6f867558c3793159a81  1325376076  47.034331
```

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text Connect Gemini

```
data['transaction_month'] = data['trans_date_trans_time'].dt.month
data['transaction_day'] = data['trans_date_trans_time'].dt.day
data['transaction_hour'] = data['trans_date_trans_time'].dt.hour

data['birth_year'] = data['dob'].dt.year
data['birth_month'] = data['dob'].dt.month
data['birth_day'] = data['dob'].dt.day
data.drop(['trans_date_trans_time', 'dob'], axis=1, inplace=True)
train['trans_date_trans_time'] = pd.to_datetime(train['trans_date_trans_time'])
train['dob'] = pd.to_datetime(train['dob'])

train['transaction_year'] = train['trans_date_trans_time'].dt.year
train['transaction_month'] = train['trans_date_trans_time'].dt.month
train['transaction_day'] = train['trans_date_trans_time'].dt.day
train['transaction_hour'] = train['trans_date_trans_time'].dt.hour

train['birth_year'] = train['dob'].dt.year
train['birth_month'] = train['dob'].dt.month
train['birth_day'] = train['dob'].dt.day

train.drop(['trans_date_trans_time', 'dob'], axis=1, inplace=True)

test['trans_date_trans_time'] = pd.to_datetime(test['trans_date_trans_time'])
test['dob'] = pd.to_datetime(test['dob'])

test['transaction_year'] = test['trans_date_trans_time'].dt.year
test['transaction_month'] = test['trans_date_trans_time'].dt.month
test['transaction_day'] = test['trans_date_trans_time'].dt.day
test['transaction_hour'] = test['trans_date_trans_time'].dt.hour
test['birth_year'] = test['dob'].dt.year
test['birth_month'] = test['dob'].dt.month
```

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text

```
print(train.shape)
print(test.shape)
print(data.shape)
```

```
(1296675, 23)
(555719, 23)
(1852394, 23)
```

```
print(data.head(0))
print(data.head())
print(data.describe())
print(data.isnull().sum())
```

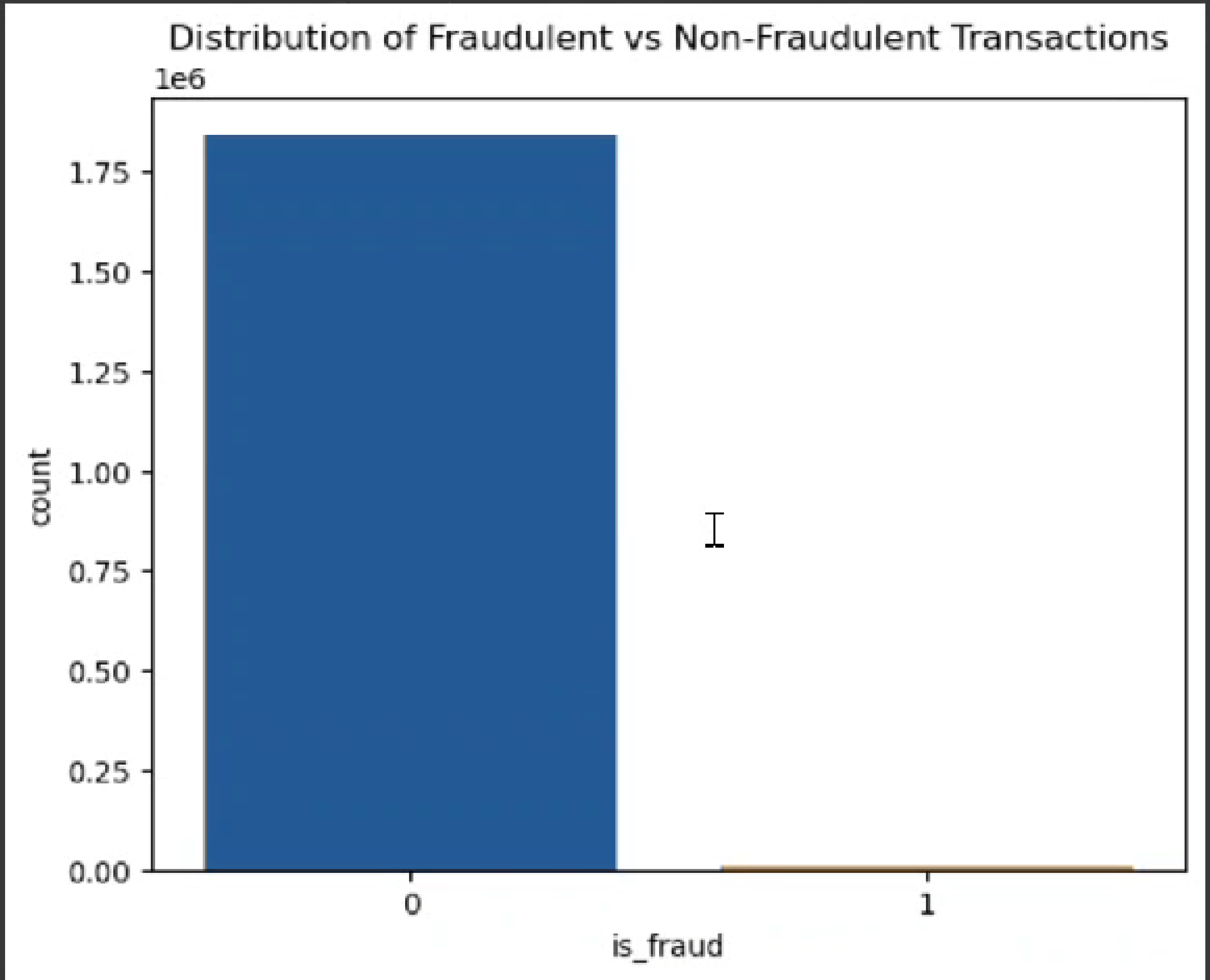
```
max    1.000000e+00  5.000000e+01  9.992100e+04  6.669330e+01 -6.795030e+01

       ...     merch_lat    merch_long      is_fraud  transaction_year  \
count  ...  1.852394e+06  1.852394e+06  1.852394e+06      1.852394e+06
mean   ...  3.853898e+01 -9.022794e+01  5.210015e-03      2.019501e+03
std    ...  5.105604e+00  1.375969e+01  7.199217e-02      4.999996e-01
min    ...  1.902742e+01 -1.666716e+02  0.000000e+00      2.019000e+03
25%    ...  3.474012e+01 -9.689944e+01  0.000000e+00      2.019000e+03
50%    ...  3.936890e+01 -8.744069e+01  0.000000e+00      2.020000e+03
75%    ...  4.195626e+01 -8.024511e+01  0.000000e+00      2.020000e+03
max    ...  6.751027e+01 -6.695090e+01  1.000000e+00      2.020000e+03

       transaction_month  transaction_day  transaction_hour    birth_year  \
count       1.852394e+06     1.852394e+06      1.852394e+06  1.852394e+06
mean        7.152067e+00     1.585076e+01      1.280612e+01  1.973289e+03
std         3.424954e+00     8.876245e+00      6.815753e+00  1.739057e+01
min         1.000000e+00     1.000000e+00      0.000000e+00  1.924000e+03
25%         4.000000e+00     8.000000e+00      7.000000e+00  1.962000e+03
```

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text

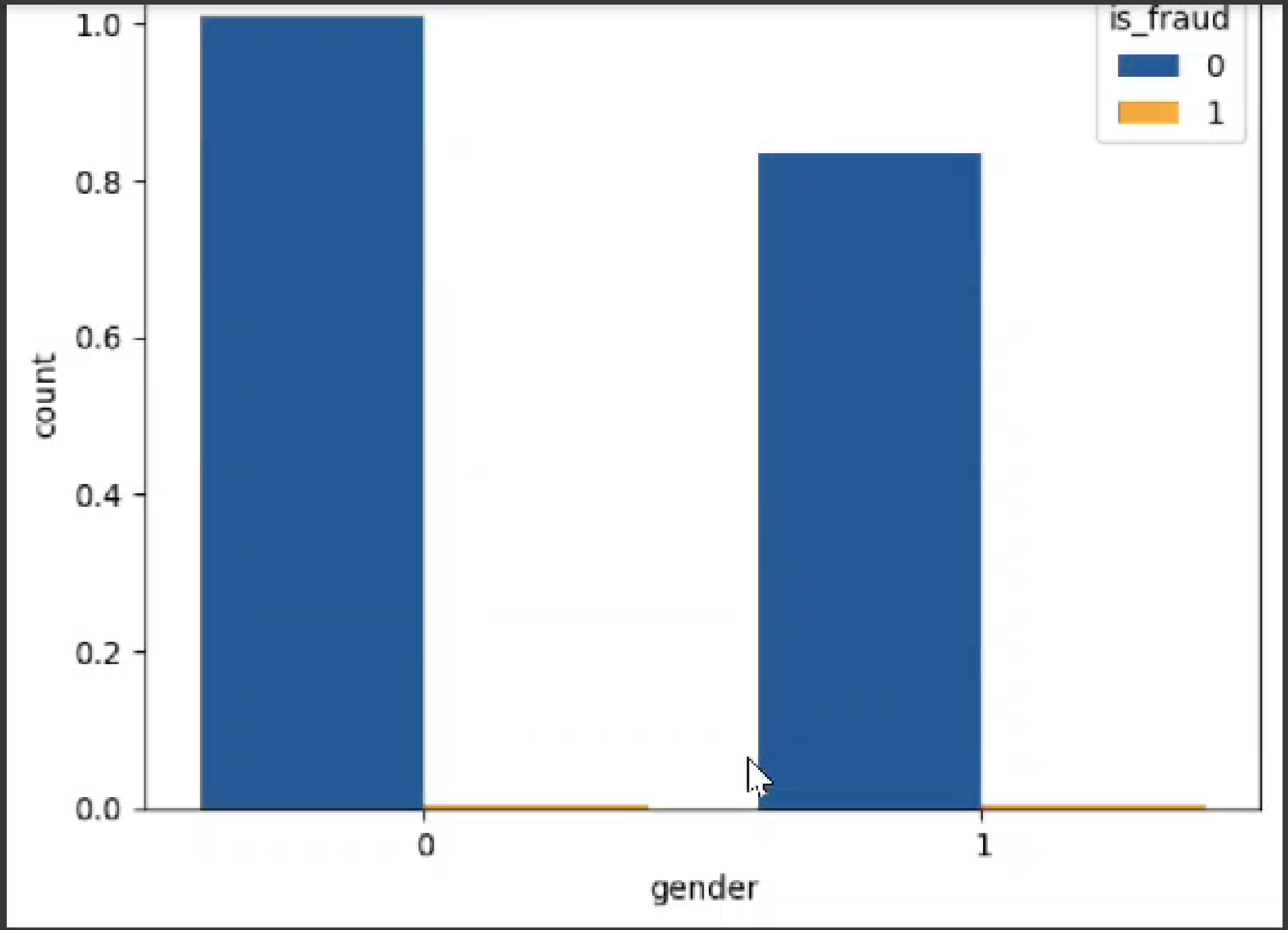
```
sns.countplot(data=data, x='is_fraud')
plt.title('Distribution of Fraudulent vs Non-Fraudulent Transactions')
plt.show()
```



credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text




```
plt.figure(figsize=(20,10))
sns.heatmap(data.corr(), annot=True, cmap='Blues')
plt.title('Correlation Heatmap')
plt.show()
```

credit card fraud detection

File Edit View Insert Runtime Tools Help Last saved at 7:04 PM

+ Code + Text Connect Gemini

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Training the model for logisticRegression
log_model = LogisticRegression(max_iter=1000)
log_model.fit(X_train, y_train)
y_pred = log_model.predict(X_test)
print(classification_report(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
C:\Users\chatt\anaconda3\Lib\site-packages\sklearn\linear_model\_logistic.py:460: ConvergenceWarning: lbfgs failed to converge (status=2):
ABNORMAL_TERMINATION_IN_LNSRCH.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
C:\Users\chatt\anaconda3\Lib\site-packages\sklearn\metrics\_classification.py:1469: UndefinedMetricWarning: Precision and F-score are ill-defined and
  _warn_prf(average, modifier, msg_start, len(result))
C:\Users\chatt\anaconda3\Lib\site-packages\sklearn\metrics\_classification.py:1469: UndefinedMetricWarning: Precision and F-score are ill-defined and
  _warn_prf(average, modifier, msg_start, len(result))
              precision    recall  f1-score   support

           0       0.99      1.00      1.00    368526
           1       0.00      0.00      0.00      1953

    accuracy                           0.99    370479
   macro avg       0.50      0.50      0.50    370479
weighted avg       0.99      0.99      0.99    370479

[[368526      0]
```