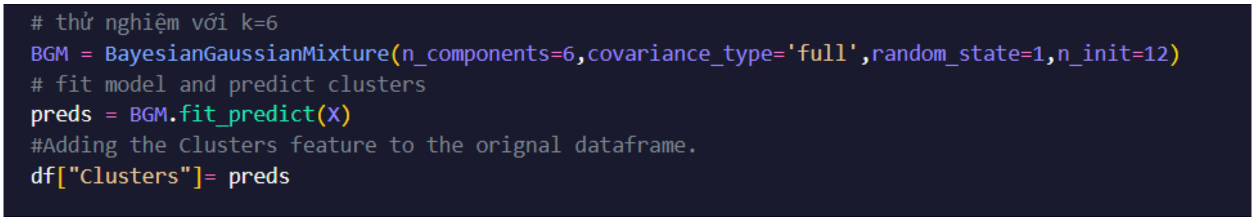
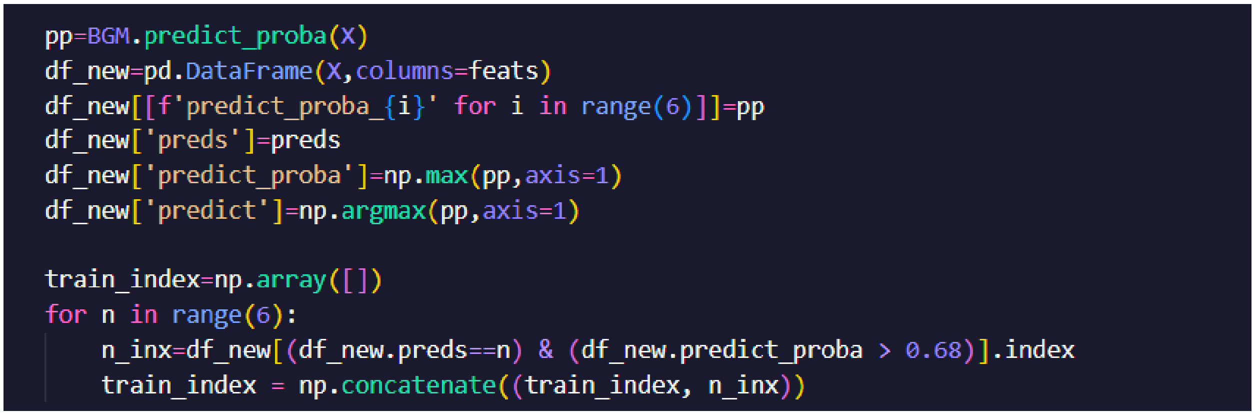
1. Gaussian

- Ảnh train

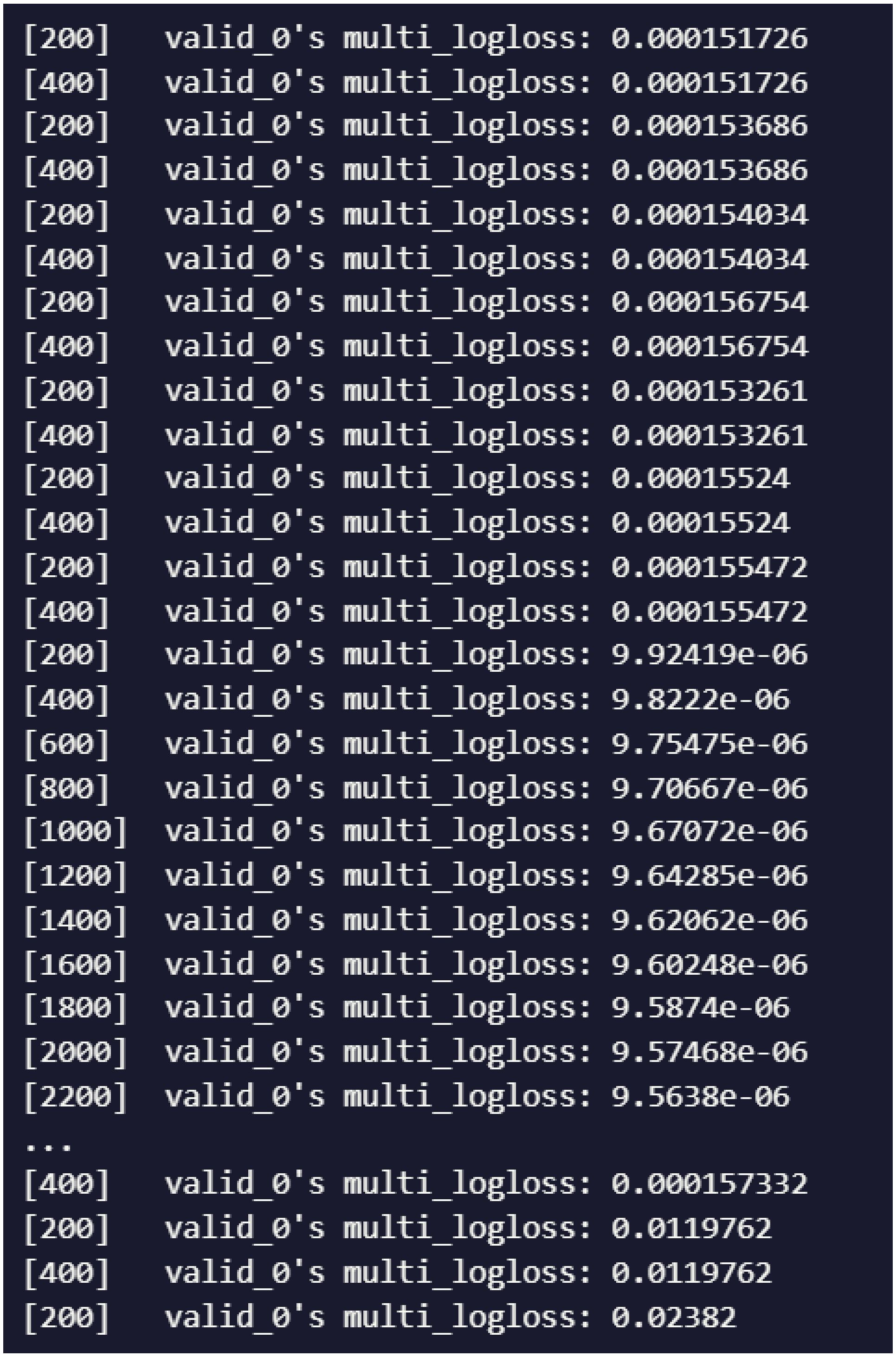


```
# thử nghiệm với k=6
BGM = BayesianGaussianMixture(n_components=6,covariance_type='full',random_state=1,n_init=12)
# fit model and predict clusters
preds = BGM.fit_predict(X)
#Adding the Clusters feature to the orignal dataframe.
df["Clusters"]= preds
```



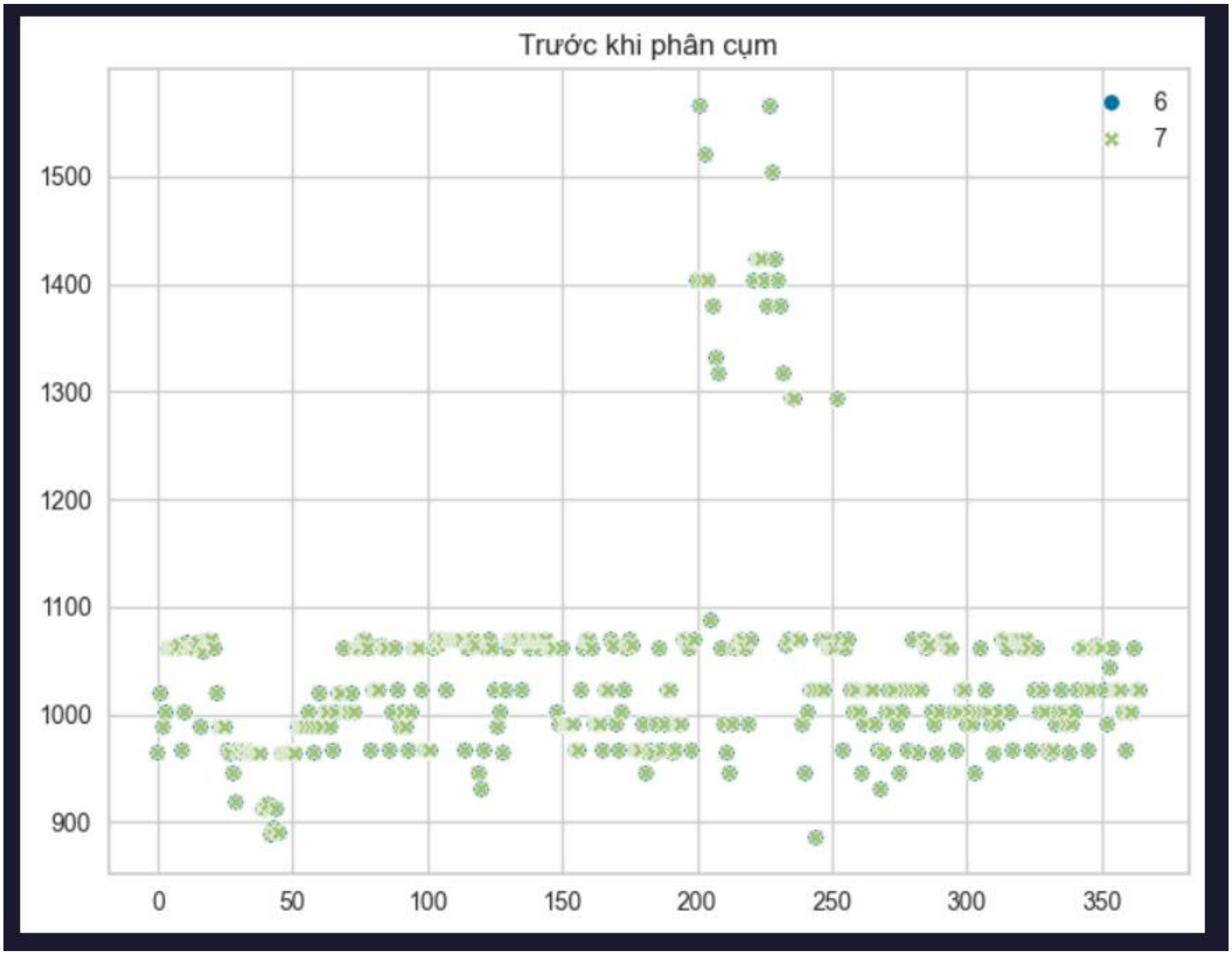
```
pp=BGM.predict_proba(X)
df_new=pd.DataFrame(X,columns=feats)
df_new[[f'predict_proba_{i}' for i in range(6)]]=pp
df_new['preds']=preds
df_new['predict_proba']=np.max(pp,axis=1)
df_new['predict']=np.argmax(pp,axis=1)

train_index=np.array([])
for n in range(6):
    n_inx=df_new[(df_new.preds==n) & (df_new.predict_proba > 0.68)].index
    train_index = np.concatenate((train_index, n_inx))
```

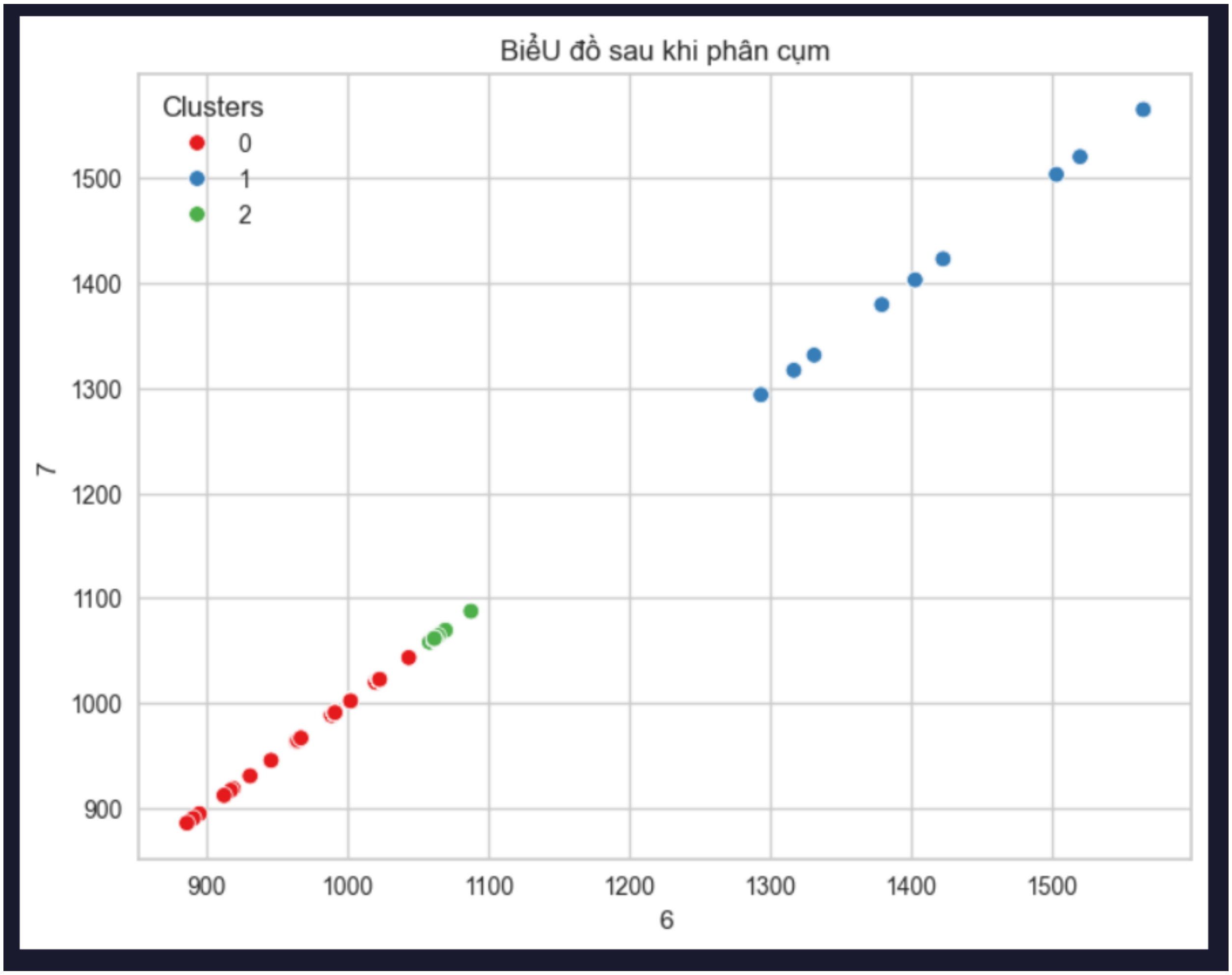
- Ảnh kết quả

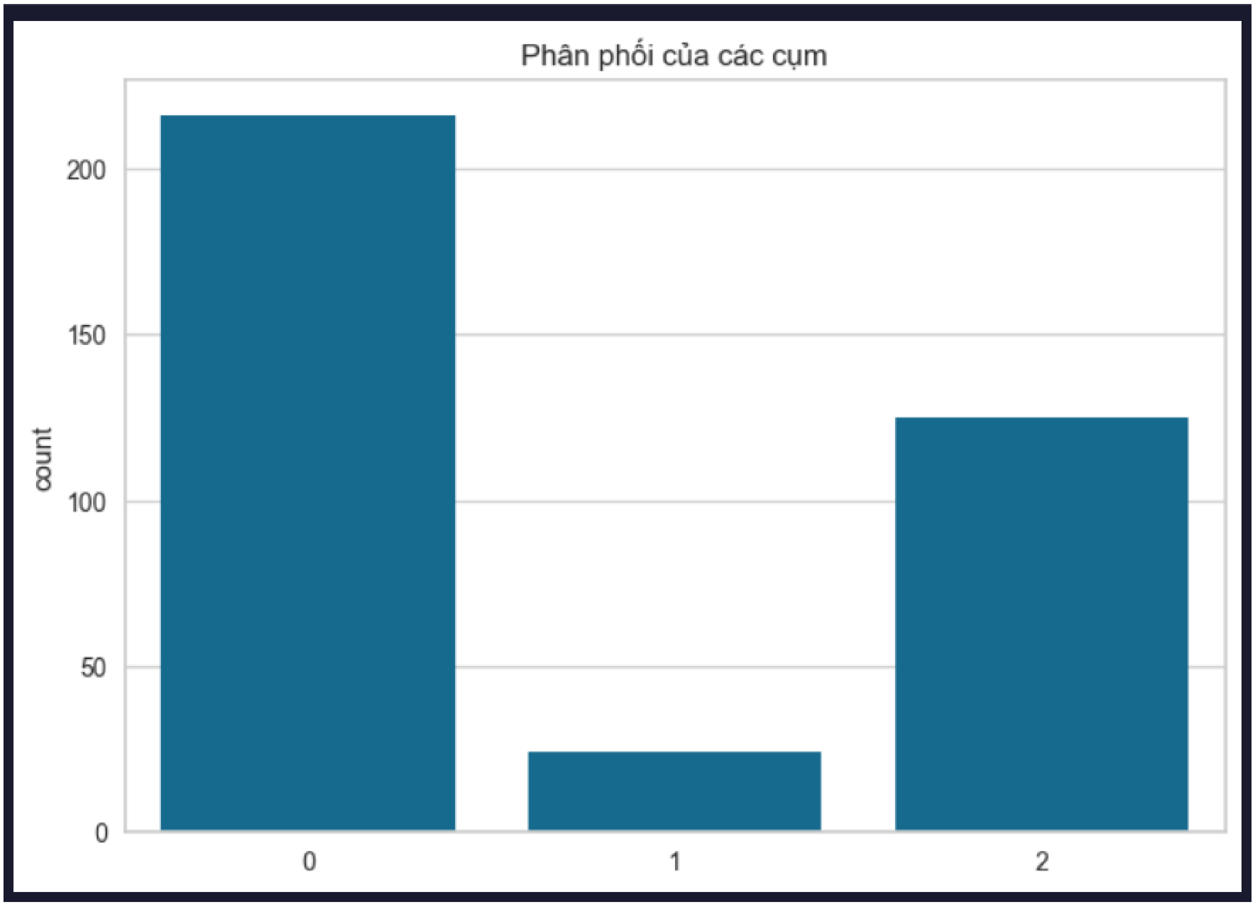


```
[200]   valid_0's multi_logloss: 0.000151726
[400]   valid_0's multi_logloss: 0.000151726
[200]   valid_0's multi_logloss: 0.000153686
[400]   valid_0's multi_logloss: 0.000153686
[200]   valid_0's multi_logloss: 0.000154034
[400]   valid_0's multi_logloss: 0.000154034
[200]   valid_0's multi_logloss: 0.000156754
[400]   valid_0's multi_logloss: 0.000156754
[200]   valid_0's multi_logloss: 0.000153261
[400]   valid_0's multi_logloss: 0.000153261
[200]   valid_0's multi_logloss: 0.00015524
[400]   valid_0's multi_logloss: 0.00015524
[200]   valid_0's multi_logloss: 0.000155472
[400]   valid_0's multi_logloss: 0.000155472
[200]   valid_0's multi_logloss: 9.92419e-06
[400]   valid_0's multi_logloss: 9.8222e-06
[600]   valid_0's multi_logloss: 9.75475e-06
[800]   valid_0's multi_logloss: 9.70667e-06
[1000]  valid_0's multi_logloss: 9.67072e-06
[1200]  valid_0's multi_logloss: 9.64285e-06
[1400]  valid_0's multi_logloss: 9.62062e-06
[1600]  valid_0's multi_logloss: 9.60248e-06
[1800]  valid_0's multi_logloss: 9.5874e-06
[2000]  valid_0's multi_logloss: 9.57468e-06
[2200]  valid_0's multi_logloss: 9.5638e-06
...
[400]   valid_0's multi_logloss: 0.000157332
[200]   valid_0's multi_logloss: 0.0119762
[400]   valid_0's multi_logloss: 0.0119762
[200]   valid_0's multi_logloss: 0.02382
```

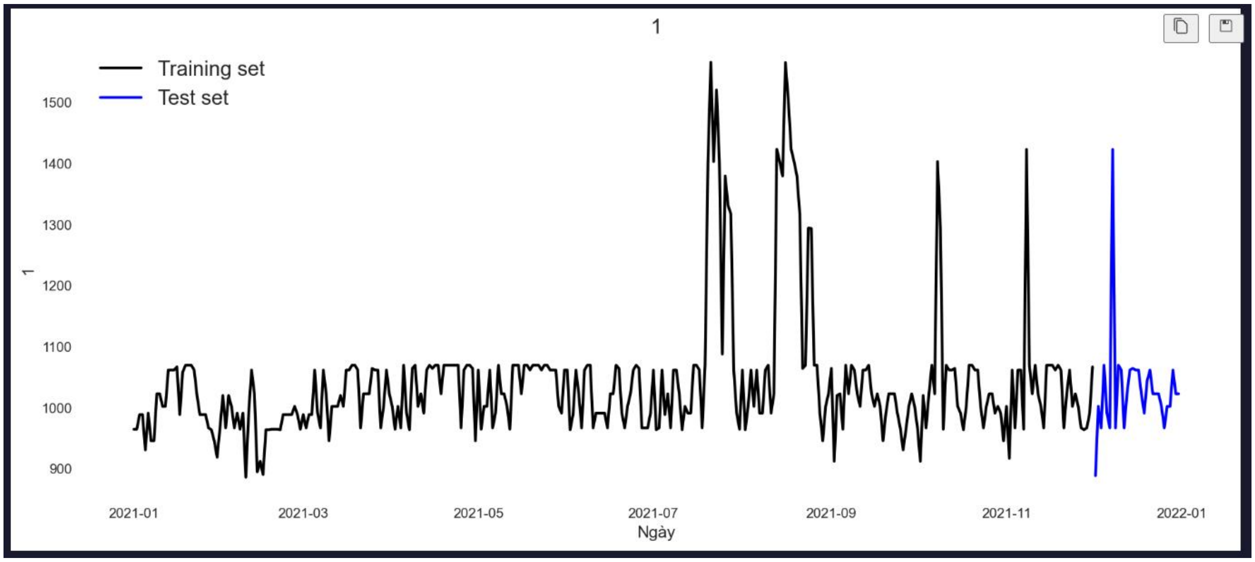
2. EDA

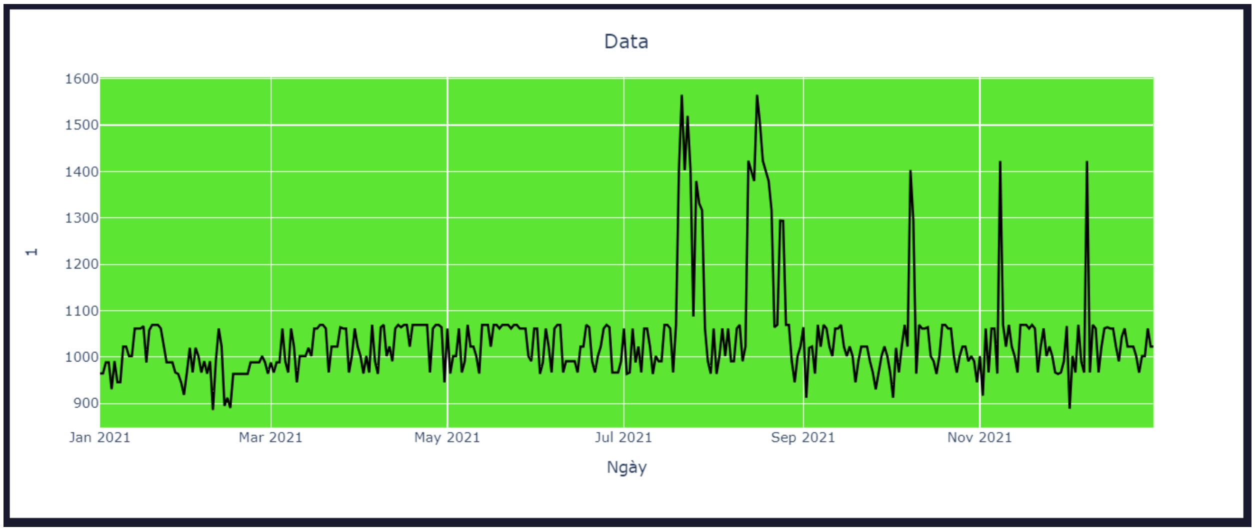

Trước khi phân cụm
● 6
× 7
1500
1400
1300
1200
1100
1000
900
0
50
100
150
200
250
300
350


BiểU đồ sau khi phân cụm
Clusters
0
1
2
1500
1400
1300
1200
1100
1000
900
7
900
1000
1100
1200
1300
1400
1500
6

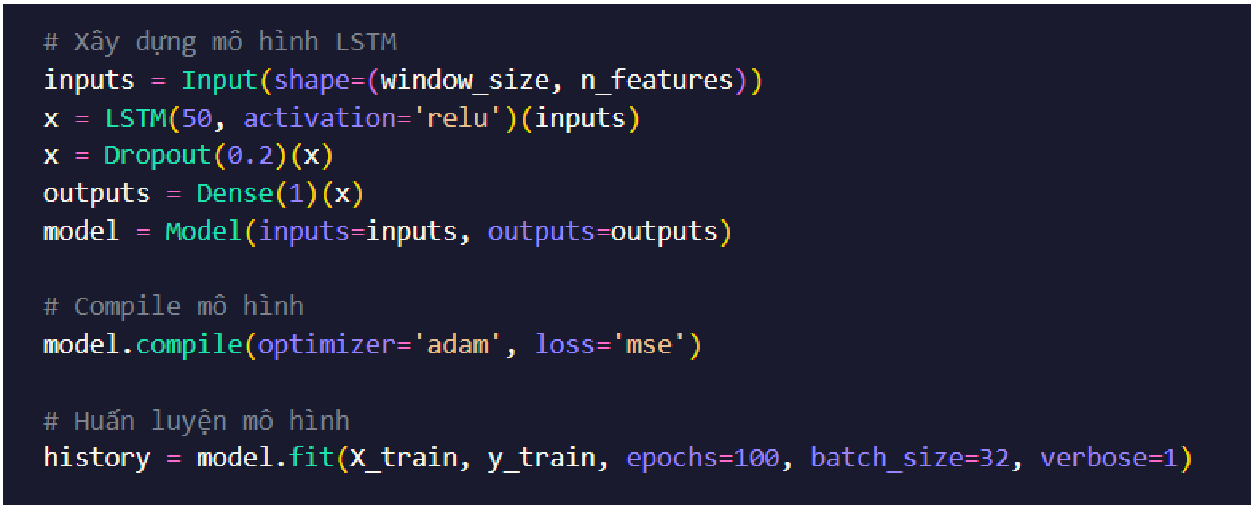


- Biểu đồ thể hiện train và test trước khi dự đoán



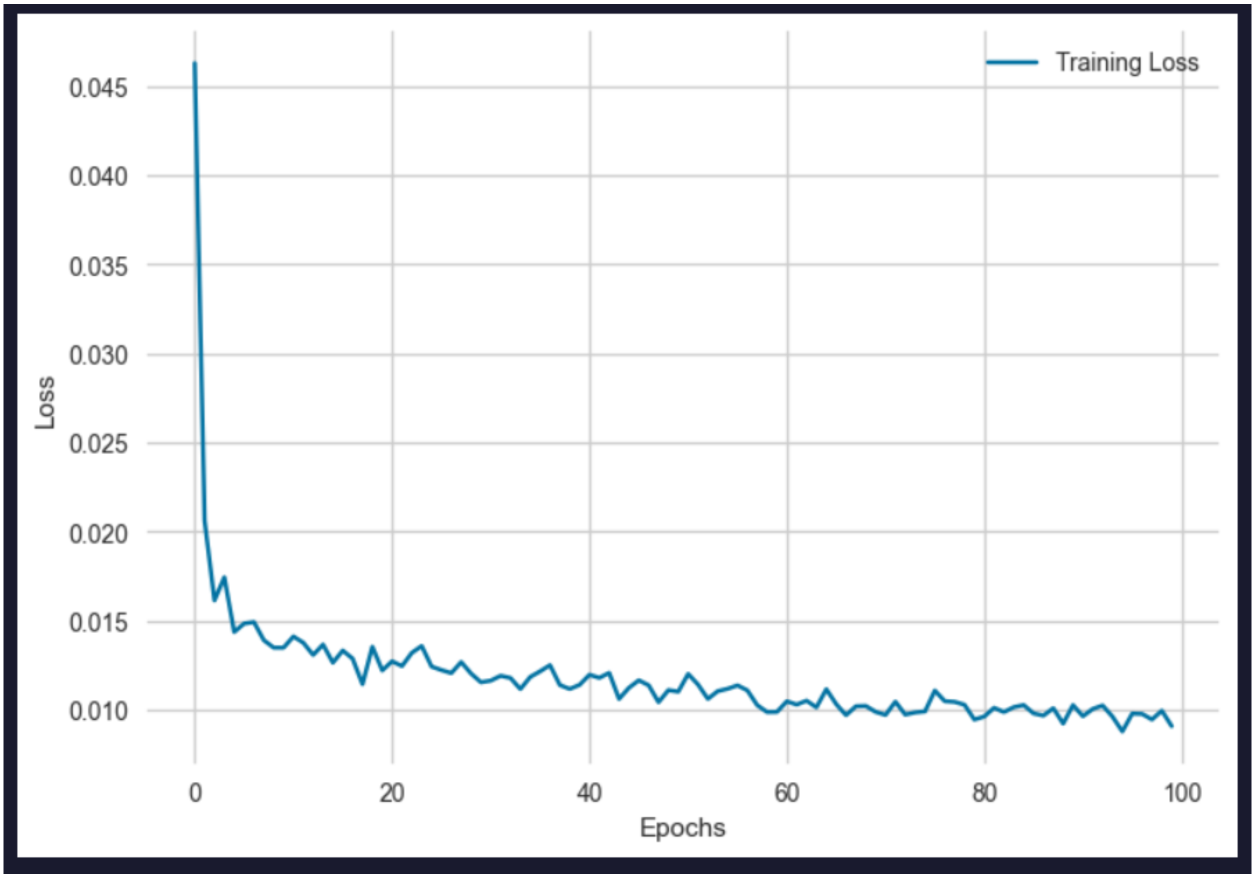


3. LSTM



```
# Xây dựng mô hình LSTM
inputs = Input(shape=(window_size, n_features))
x = LSTM(50, activation='relu')(inputs)
x = Dropout(0.2)(x)
outputs = Dense(1)(x)
model = Model(inputs=inputs, outputs=outputs)

# Compile mô hình
model.compile(optimizer='adam', loss='mse')

# Huấn luyện mô hình
history = model.fit(X_train, y_train, epochs=100, batch_size=32, verbose=1)
```

- Kết quả



Link github

https://github.com/ducjr/TH2_PhanTichChuoiThoiGian