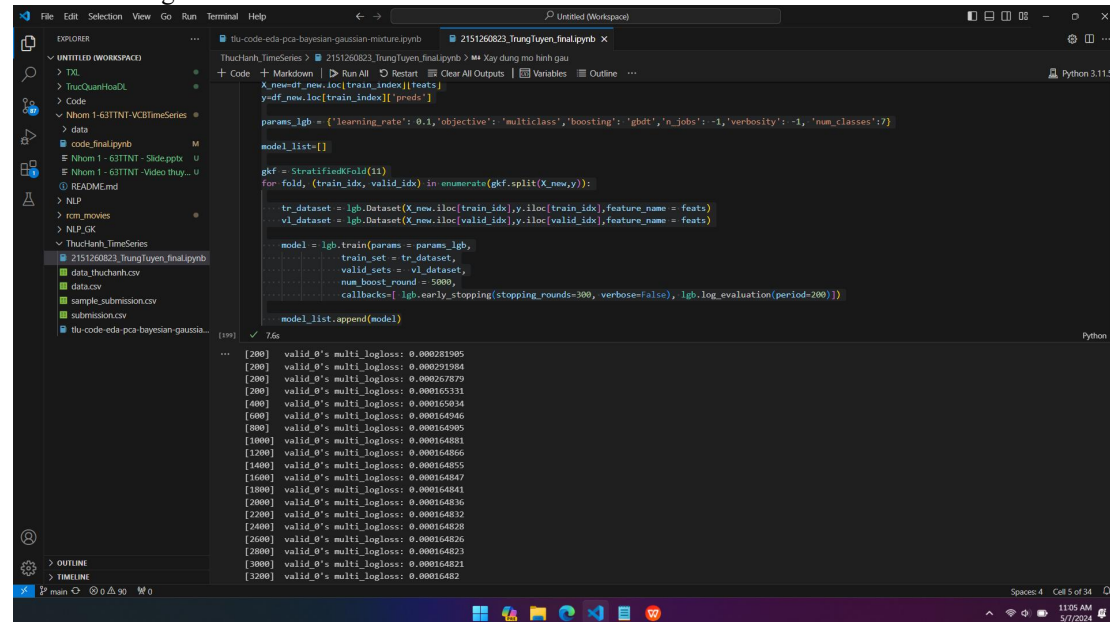


Nguyễn Trung Tuyển - 21512608 - Thực hành 01 - 07/05/2024

1. Kịch bản chia data

- Với mô hình LSTM sử dụng 30 ngày trước để dự báo ngày tiếp theo
- Với mô hình Sarima sử dụng các tham số tính chu kì là 30 ngày để dự đoán giá trị tiếp theo

2. Ảnh training:



```
ThucHanh_TimeSeries > 2151260823_TrungTuyen_final.ipynb > Xay dựng mô hình gao
+ Code + Markdown | ▶ Run All | ⌂ Restart | 🗑 Clear All Outputs | 📄 Variables | 📖 Outline ...
Python 3.11.5

X_new=df_new.loc[train_index|:feats]
y=df_new.loc[train_index|:'preds']

params_lgb = {'learning_rate': 0.1,'objective': 'multiclass','boosting': 'gbdt','n_jobs': -1,'verbosity': -1, 'num_classes':7}

model_list=[]

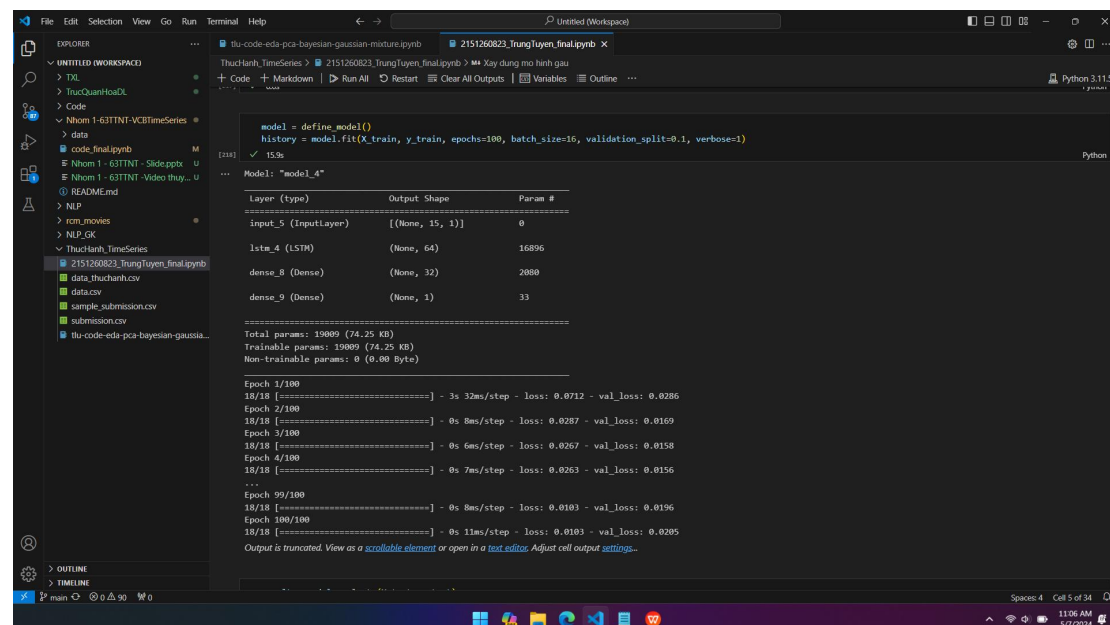
gkf = StratifiedKFold(11)
for fold, (train_id, valid_id) in enumerate(gkf.split(X_new,y)):

    tr_dataset = lgb.Dataset(X_new.iloc[train_id],y.iloc[train_id],feature_name = feats)
    vl_dataset = lgb.Dataset(X_new.iloc[valid_id],y.iloc[valid_id],feature_name = feats)

    model = lgb.train(params = params_lgb,
                      train_set = tr_dataset,
                      valid_sets = vl_dataset,
                      num_boost_round = 5000,
                      callbacks=[ lgb.early_stopping(stopping_rounds=300, verbose=False), lgb.log_evaluation(period=200)])

    model_list.append(model)

[200] valid_0's multi_logloss: 0.000231505
[200] valid_0's multi_logloss: 0.000231504
[200] valid_0's multi_logloss: 0.000267879
[200] valid_0's multi_logloss: 0.000165331
[400] valid_0's multi_logloss: 0.000165034
[600] valid_0's multi_logloss: 0.000164936
[800] valid_0's multi_logloss: 0.000164905
[1000] valid_0's multi_logloss: 0.000164881
[1200] valid_0's multi_logloss: 0.000164866
[1400] valid_0's multi_logloss: 0.000164855
[1600] valid_0's multi_logloss: 0.000164847
[1800] valid_0's multi_logloss: 0.000164841
[2000] valid_0's multi_logloss: 0.000164836
[2200] valid_0's multi_logloss: 0.000164832
[2400] valid_0's multi_logloss: 0.000164828
[2600] valid_0's multi_logloss: 0.000164826
[2800] valid_0's multi_logloss: 0.000164823
[3000] valid_0's multi_logloss: 0.000164821
[3200] valid_0's multi_logloss: 0.00016482
```



```
ThucHanh_TimeSeries > 2151260823_TrungTuyen_final.ipynb > Xay dựng mô hình gao
+ Code + Markdown | ▶ Run All | ⌂ Restart | 🗑 Clear All Outputs | 📄 Variables | 📖 Outline ...
Python 3.11.5

model = define_model()
history = model.fit(X_train, y_train, epochs=100, batch_size=16, validation_split=0.1, verbose=1)

Model: "model_4"
=====
Layer (type) Output Shape Param #
-----
input_5 (InputLayer) [(None, 15, 1)] 0
lstm_4 (LSTM) (None, 64) 16896
dense_8 (Dense) (None, 32) 2080
dense_9 (Dense) (None, 1) 33
=====
Total params: 19009 (74.25 KB)
Trainable params: 19009 (74.25 KB)
Non-trainable params: 0 (0.00 Byte)

Epoch 1/100
18/18 [=====] - 3s 32ms/step - loss: 0.0712 - val_loss: 0.0286
Epoch 2/100
18/18 [=====] - 0s 8ms/step - loss: 0.0287 - val_loss: 0.0169
Epoch 3/100
18/18 [=====] - 0s 6ms/step - loss: 0.0267 - val_loss: 0.0158
Epoch 4/100
18/18 [=====] - 0s 7ms/step - loss: 0.0263 - val_loss: 0.0156
...
Epoch 99/100
18/18 [=====] - 0s 8ms/step - loss: 0.0103 - val_loss: 0.0196
Epoch 100/100
18/18 [=====] - 0s 11ms/step - loss: 0.0103 - val_loss: 0.0205
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings.
```

```
File Edit Selection View Go Run Terminal Help
2151260823_TrungTuyen_final.ipynb x
Thuchanh_TimeSeries > 2151260823_TrungTuyen_final.ipynb > ...
+ Code + Markdown | Run All | Restart | Clear All Outputs | Variables | Outline ...
Python 3.11.5

p, d, q = 0, 1, 1
P, D, Q, s = 0, 1, 1, 30
[119] ✓ 0.0s Python

model_sarima = SARIMAX(df_train, order=(p, d, q), seasonal_order=(P, D, Q, s))
model_fit_sarima = model_sarima.fit()
[120] ✓ 3.5s Python

forecast_diff = model_fit_sarima.forecast(steps=len(df_test))
print(forecast_diff)
[121] ✓ 0.0s Python

... 328 998.085858
329 950.742045
330 961.778368
331 968.375809
332 962.620394
333 960.852329
334 967.410280
335 958.843832
336 966.506996
337 976.900969
338 1000.929808
339 981.891761
340 1013.225806
341 1042.078888
342 991.934165
343 984.766590
344 1036.961011
345 1030.219899
346 1036.746924
347 1041.829119
348 1042.265608
349 1035.300803
350 1063.858510

Python main 0 0 112 W 0 Spaces: 4 Cell 44 of 44 11:20 AM 5/7/2024
```

3. Ảnh kết quả chạy

