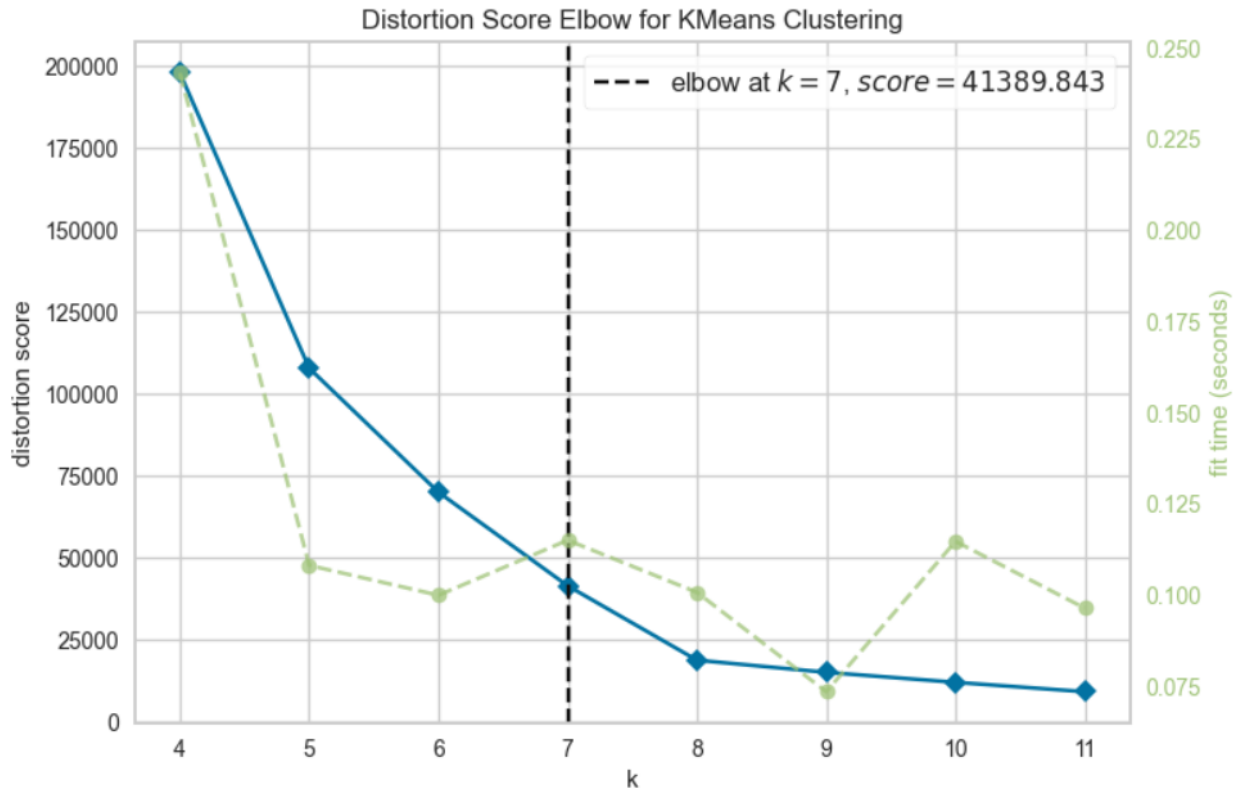
# 1. Kịch bản chia data



## 2. Ảnh training

```python
from statsmodels.tsa.arima.model import ARIMA

# Khởi tạo và phù hợp với mô hình ARIMA
model_arima = ARIMA(r_t, order=(2, 0, 2))
model_fit = model_arima.fit()

# In ra tóm tắt của mô hình
print(model_fit.summary())
```

✓ 1.0s

```python
BGM = BayesianGaussianMixture(n_components=7,covariance_type='full',random_state=1,n_init=15)
# fit model and predict clusters
preds = BGM.fit_predict(X)

#Adding the Clusters feature to the orignal dataframe.
df["clusters"]= preds
```

✓ 1.8s

```python
pp=BGM.predict_proba(X)# Calcualting the probabilities of each prediction
df_new=pd.DataFrame(X,columns=feats)
df_new[[f'predict_proba_{i}' for i in range(7)]]=pp # creating new dataframe columns of probabilites
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(7):
    n_inx=df_new[(df_new.preds==n) & (df_new.predict_proba > 0.68)].index
    train_index = np.concatenate((train_index, n_inx))
```

✓ 0.0s

```python
from sklearn.model_selection import StratifiedKFold
X_new=df_new.loc[train_index][feats]
y=df_new.loc[train_index]['preds']

params_lgb = {'learning_rate': 0.06,'objective': 'multiclass','boosting': 'gbdt','n_jobs': -1,'verbosity': -1, 'num_classes':7}

model_list=[]

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

    tr_dataset = lgb.Dataset(X_new.iloc[train_idx],y.iloc[train_idx],feature_name = feats)
    vl_dataset = lgb.Dataset(X_new.iloc[valid_idx],y.iloc[valid_idx],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)
```

15.5s

# 3. Ảnh kết quả

SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | y | No. Observations: | 365 |
| Model: | ARIMA(2, 0, 2) | Log Likelihood | 563.727 |
| Date: | Tue, 07 May 2024 | AIC | -1115.453 |
| Time: | 10:57:35 | BIC | -1092.054 |
| Sample: | 0 | HQIC | -1106.154 |
| | - 365 | | |
| Covariance Type: | opg | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -0.0002 | 0.001 | -0.211 | 0.833 | -0.002 | 0.001 |
| ar.L1 | -0.0635 | 0.066 | -0.956 | 0.339 | -0.194 | 0.067 |
| ar.L2 | 0.5935 | 0.061 | 9.713 | 0.000 | 0.474 | 0.713 |
| ma.L1 | -0.0425 | 0.051 | -0.838 | 0.402 | -0.142 | 0.057 |
| ma.L2 | -0.8414 | 0.054 | -15.610 | 0.000 | -0.947 | -0.736 |
| sigma2 | 0.0026 | 9.93e-05 | 26.270 | 0.000 | 0.002 | 0.003 |

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.04 | Jarque-Bera (JB): | 675.24 |
| Prob(Q): | 0.85 | Prob(JB): | 0.00 |
| Heteroskedasticity (H): | 1.74 | Skew: | 0.51 |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 9.58 |

## 4. Link git

https://github.com/ducjr/TH1_PhanTichChuoiThoiGian