

PREDICTION OF STOCK PRICE DIRECTION

START →

6520422002 Apichai Fukaew
6520422022 Naris Chanpaiboonkit
6520422025 Pakwat Raksasin



PREDICTION OF STOCK PRICE DIRECTION

เป็นการสร้างโมเดลทำนายราคาหุ้นที่คาดว่าจะซื้อหรือขาย (ราคาปิดของ Period ข้างหน้ามากกว่า Period ปัจจุบัน โดย Period ปัจจุบันจะถูกบันทึกว่าต้องซื้อ (1) ในทุกๆ 15 นาที กับหุ้น 3 ตัว ได้แก่ ERW, TISCO, SPRC โดยดูจาก feature ข้อมูลการเปลี่ยนแปลงราคาในรอบ 15 นาที ที่สูงที่สุดที่โมเดลทำนายได้ของแต่ละหุ้น โดยข้อมูลที่ใช้ในการ train และ test โมเดลเป็นข้อมูลราคาหุ้น ในอดีตย้อนหลัง ประมาณ 60 วันโดยประมาณ ตั้งแต่ช่วง วันที่ทำการรับข้อมูลหักลบย้อนกลับไป 60 วัน

SECURITIES INFORMATION



- ประกอบธุรกิจเกี่ยวกับการลงทุนพัฒนาและดำเนินธุรกิจโรงแรม
- อุตสาหกรรมกลุ่มบริการ
- ธุรกิจหมวดการท่องเที่ยวและสันักงาน



- ประกอบธุรกิจการถือหุ้นในบริษัทอื่น ของกลุ่มธุรกิจทางการเงินทิสโก้ โดยมีธนาคารทิสโก้เป็นหลัก
- อุตสาหกรรมกลุ่มธุรกิจการเงิน
- ธุรกิจ報導การ



- โรงกลั่นน้ำมันของบริษัท พลิตพลิตภัณฑ์ปีโตรเลียม
- อุตสาหกรรมกลุ่มกรัพยากร
- ธุรกิจพลังงานและสารเคมี

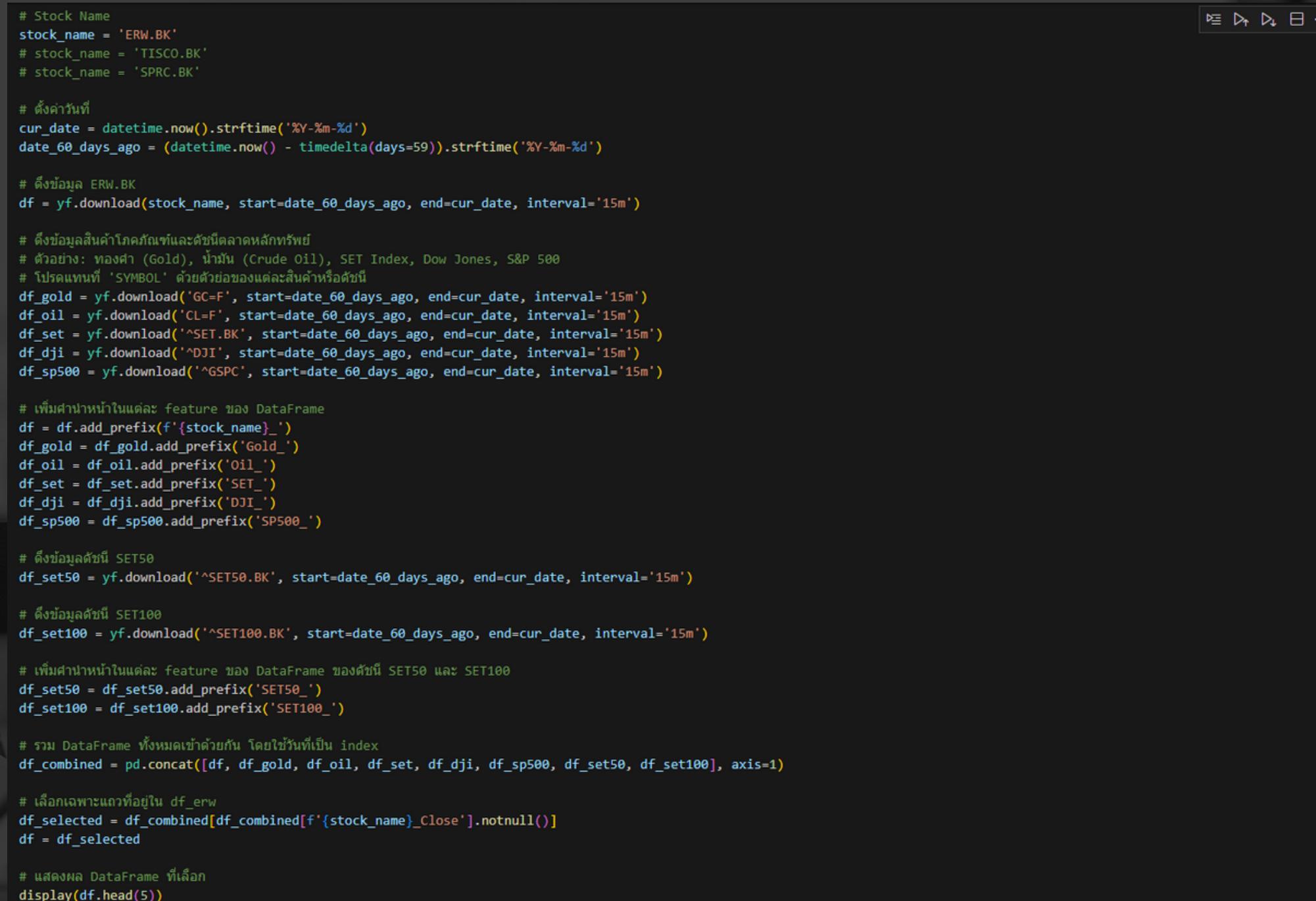
IMPORT ALL NECESSARY PACKAGES

```
import pandas as pd
import numpy as np
import warnings
import ta
from statsmodels.stats.outliers_influence import variance_inflation_factor
from sklearn.model_selection import (
    train_test_split,
    GridSearchCV,
    cross_val_score,
    TimeSeriesSplit,
    RandomizedSearchCV,
)
from sklearn.ensemble import (
    GradientBoostingClassifier,
    RandomForestClassifier,
    RandomForestClassifier,
)
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from lightgbm import LGBMClassifier
from xgboost import XGBClassifier
from mlxtend.feature_selection import SequentialFeatureSelector as SFS
from datetime import timedelta, datetime
import yfinance as yf
from sklearn.preprocessing import RobustScaler
from sklearn.feature_selection import RFE
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from bayes_opt import BayesianOptimization
from scipy.stats import chi2_contingency
from itertools import combinations
from sklearn.metrics import roc_auc_score
import seaborn as sns

# Set pandas options
pd.options.mode.chained_assignment = None
pd.set_option('display.max_columns', None)
warnings.filterwarnings('ignore', category=pd.errors.PerformanceWarning)
```

COLLECT DATA

- ดึงข้อมูลราคาหุ้น ERW.BK, TISCO.BK และ SPRC.BK
- ดึงข้อมูลสินค้าโภคภัณฑ์ ได้แก่ ทองคำ (Gold), น้ำมัน (Crude Oil), SET Index, Dow Jones, S&P 500 (โดยข้อมูลแต่ละค่าจะมีค่า Volume, Open, High, Low, Close และ Adj_Close)



```
# Stock Name
stock_name = 'ERW.BK'
# stock_name = 'TISCO.BK'
# stock_name = 'SPRC.BK'

# ดึงค่ารันที่
cur_date = datetime.now().strftime('%Y-%m-%d')
date_60_days_ago = (datetime.now() - timedelta(days=59)).strftime('%Y-%m-%d')

# ดึงข้อมูล ERW.BK
df = yf.download(stock_name, start=date_60_days_ago, end=cur_date, interval='15m')

# ดึงข้อมูลสินค้าโภคภัณฑ์และดัชนีตลาดหลักทรัพย์
# ตัวอย่าง: ทองคำ (Gold), น้ำมัน (Crude Oil), SET Index, Dow Jones, S&P 500
# โปรดแทนที่ 'SYMBOL' ด้วยตัวชื่อของแต่ละสินค้าหรือดัชนี
df_gold = yf.download('GC=F', start=date_60_days_ago, end=cur_date, interval='15m')
df_oil = yf.download('CL=F', start=date_60_days_ago, end=cur_date, interval='15m')
df_set = yf.download('^SET.BK', start=date_60_days_ago, end=cur_date, interval='15m')
df_dji = yf.download('^DJI', start=date_60_days_ago, end=cur_date, interval='15m')
df_sp500 = yf.download('^GSPC', start=date_60_days_ago, end=cur_date, interval='15m')

# เพิ่มค่าไฟฟ้าในแต่ละ feature ของ DataFrame
df = df.add_prefix(f'{stock_name}_')
df_gold = df_gold.add_prefix('Gold_')
df_oil = df_oil.add_prefix('Oil_')
df_set = df_set.add_prefix('SET_')
df_dji = df_dji.add_prefix('DJI_')
df_sp500 = df_sp500.add_prefix('SP500_')

# ดึงข้อมูลดัชนี SET50
df_set50 = yf.download('^SET50.BK', start=date_60_days_ago, end=cur_date, interval='15m')

# ดึงข้อมูลดัชนี SET100
df_set100 = yf.download('^SET100.BK', start=date_60_days_ago, end=cur_date, interval='15m')

# เพิ่มค่าไฟฟ้าในแต่ละ feature ของ DataFrame ของดัชนี SET50 และ SET100
df_set50 = df_set50.add_prefix('SET50_')
df_set100 = df_set100.add_prefix('SET100_')

# รวม DataFrame ทั้งหมดเข้าด้วยกัน โดยใช้วันที่เป็น index
df_combined = pd.concat([df, df_gold, df_oil, df_set, df_dji, df_sp500, df_set50, df_set100], axis=1)

# เลือกเฉพาะแคร์ออยู่ใน df_ewr
df_selected = df_combined[df_combined[f'{stock_name}_Close'].notnull()]
df = df_selected

# แสดงผล DataFrame ที่เลือก
display(df.head(5))
```

FEATURE ENGINEERING

สร้าง Indicator จากข้อมูล ปริมาณและราคากุ้น ที่มีโดยการอาศัย และมีการทำ Feature อื่นๆ เพิ่มเติมได้แก่ **Signal features** ที่บอกว่าราคา ตอนนั้นควรซื้อหรือขายและในบาง Feature จะคำนึงถึงความเป็น Time Series ของ Feature นั้นๆด้วยจะมี window หลายๆ แบบหรือมองย้อนไป k period โดย มีจำนวน Feature ที่สร้างขึ้นมา 185 ตัวซึ่งมีกั้ง **Numerical Features** และ **Categorical Features** นอกจากนี้ยังมี Feature ที่เกี่ยว กับเวลาอย่าง วันของสัปดาห์ และชั่วโมงของวัน observation นั้นๆ

Feature Name	Description	Feature Name	Description
SMA_{window}	Simple Moving Average ส่าหรับช่วงเวลา x, เป็นค่าเฉลี่ยของ ราคาปิดในช่วงเวลา n ต่อ	CMF_Buy_{window}	สัญญาณซื้อจาก CMF ส่าหรับช่วงเวลา x
EMA_{window}	Exponential Moving Average ส่าหรับช่วงเวลา x, คล้ายกับ SMA แต่ให้หนักมากขึ้นกับข้อมูลล่าสุด	CMF_Sell_{window}	สัญญาณขายจาก CMF ส่าหรับช่วงเวลา x
Stoch_%K_{window}	Stochastic Oscillator %K ส่าหรับช่วงเวลา x, ชี้วัดไม่เม่นตั้มของ ราคา	Parabolic_SAR_{window}	Parabolic SAR ส่าหรับช่วงเวลา x, ใช้เพื่อหาจุดกลับตัว ของแนวโน้ม
Stoch_%D_{window}	Stochastic Oscillator %D ส่าหรับช่วงเวลา x, เป็นค่าเฉลี่ย เคลื่อนที่ของ %K	Fibonacci_Retraction_23.6_{window}	Fibonacci Retracement 23.6% ส่าหรับช่วงเวลา x
Stoch_Buy_{window}	สัญญาณซื้อจาก Stochastic Oscillator ส่าหรับช่วงเวลา x	Fibonacci_Retraction_38.2_{window}	Fibonacci Retracement 38.2% ส่าหรับช่วงเวลา x
Stoch_Sell_{window}	สัญญาณขายจาก Stochastic Oscillator ส่าหรับช่วงเวลา x	ROC	Price Rate of Change, วัดการเปลี่ยนแปลงของราคาเทียบ กับช่วงเวลาก่อนหน้า
RSI_{window}	Relative Strength Index ส่าหรับช่วงเวลา x, วัดไม่เม่นตั้มของ การเปลี่ยนแปลงราคา	ROC_Buy	สัญญาณซื้อจาก ROC (ROC > 0 และก่อนหน้านั้น < 0)
RSI_Buy_{window}	สัญญาณซื้อจาก RSI ส่าหรับช่วงเวลา x (RSI < 30)	ROC_Sell	สัญญาณขายจาก ROC (ROC < 0 และก่อนหน้านั้น > 0)
RSI_Sell_{window}	สัญญาณขายจาก RSI ส่าหรับช่วงเวลา x (RSI > 70)	MACD	Moving Average Convergence Divergence, วัดแนวโน้ม และไม่เม่นตั้ม
WPR_{window}	Williams %R ส่าหรับช่วงเวลา x, วัดระดับราคาปัจจุบันเทียบกับ ช่วงราคาในอดีต	MACD_signal	สัญญาณของ MACD
ATR_{window}	Average True Range ส่าหรับช่วงเวลา x, วัดความผันผวนของ ราคา	MACD_diff	ความแตกต่างระหว่าง MACD กับสัญญาณของมัน
CCI_{window}	Commodity Channel Index ส่าหรับช่วงเวลา x, วัดการ เปลี่ยนแปลงของราคาเทียบกับค่าเฉลี่ย	MACD_Buy	สัญญาณซื้อจาก MACD (MACD > MACD_signal)
Bollinger_mavg_{window}	ค่าเฉลี่ยเคลื่อนที่ของ Bollinger Bands ส่าหรับช่วงเวลา x	MACD_Sell	สัญญาณขายจาก MACD (MACD < MACD_signal)
Bollinger_hband_{window}	แคบบนของ Bollinger Bands ส่าหรับช่วงเวลา x	Momentum	ไม่เม่นตั้มของราคา, วัดด้วยการเปลี่ยนแปลงของราคา
Bollinger_lband_{window}	แคบล่างของ Bollinger Bands ส่าหรับช่วงเวลา x	OBV	On-Balance Volume, วัดปริมาณการซื้อขายสะสม
VWAP_{window}	Volume-Weighted Average Price ส่าหรับช่วงเวลา x	Close_pct_change	เปลอร์เซ็นต์การเปลี่ยนแปลงของราคาปิด
VWAP_Buy_{window}	สัญญาณซื้อจาก VWAP ส่าหรับช่วงเวลา x	hour_x_cate	One-hot encoding ของชั่วโมงที่ x ของวัน
VWAP_Sell_{window}	สัญญาณขายจาก VWAP ส่าหรับช่วงเวลา x	day_x_cate	One-hot encoding ของวันที่ x ของสัปดาห์
CMF_{window}	Chaikin Money Flow ส่าหรับช่วงเวลา x, วัดการไหลเข้าออก ของเงินลงทุน		

หมายเหตุ: {window} หมายถึงระยะเวลาที่กำหนดสำหรับการคำนวณค่าเฉลี่ยเคลื่อนที่ โดยกำหนดให้มีค่า 5,10,15,20,25



FEATURE SELECTION

ทำ Feature selection ออกมา 5 วิธีเพื่อนำไปทำการทดลองกับทุกๆโมเดลที่นำมาฝึกฝน

1. Random Forest Feature Importance, Multi-Collinearity, Cramer's V
2. Recursive Feature Elimination (RFE)
3. Principal Component Analysis (PCA)
4. Domain Knowledge Selection
5. Stepwise selection

Domain Knowledge	Feature selection
Stoch_%K_5	SET50_Close
Stoch_Buy_5	VWAP_25
Stoch_Sell_5	hour_11_cate
RSI_5	hour_12_cate
RSI_Buy_5	hour_14_cate
RSI_Sell_5	hour_15_cate
SET_Close	hour_16_cate
Close_pct_change	day_1_cate
Oil_Close	day_2_cate
Gold_Close	day_3_cate
MACD_Buy	day_4_cate
MACD_Sell	close_col



TRANSFORMATION

Categorical Data

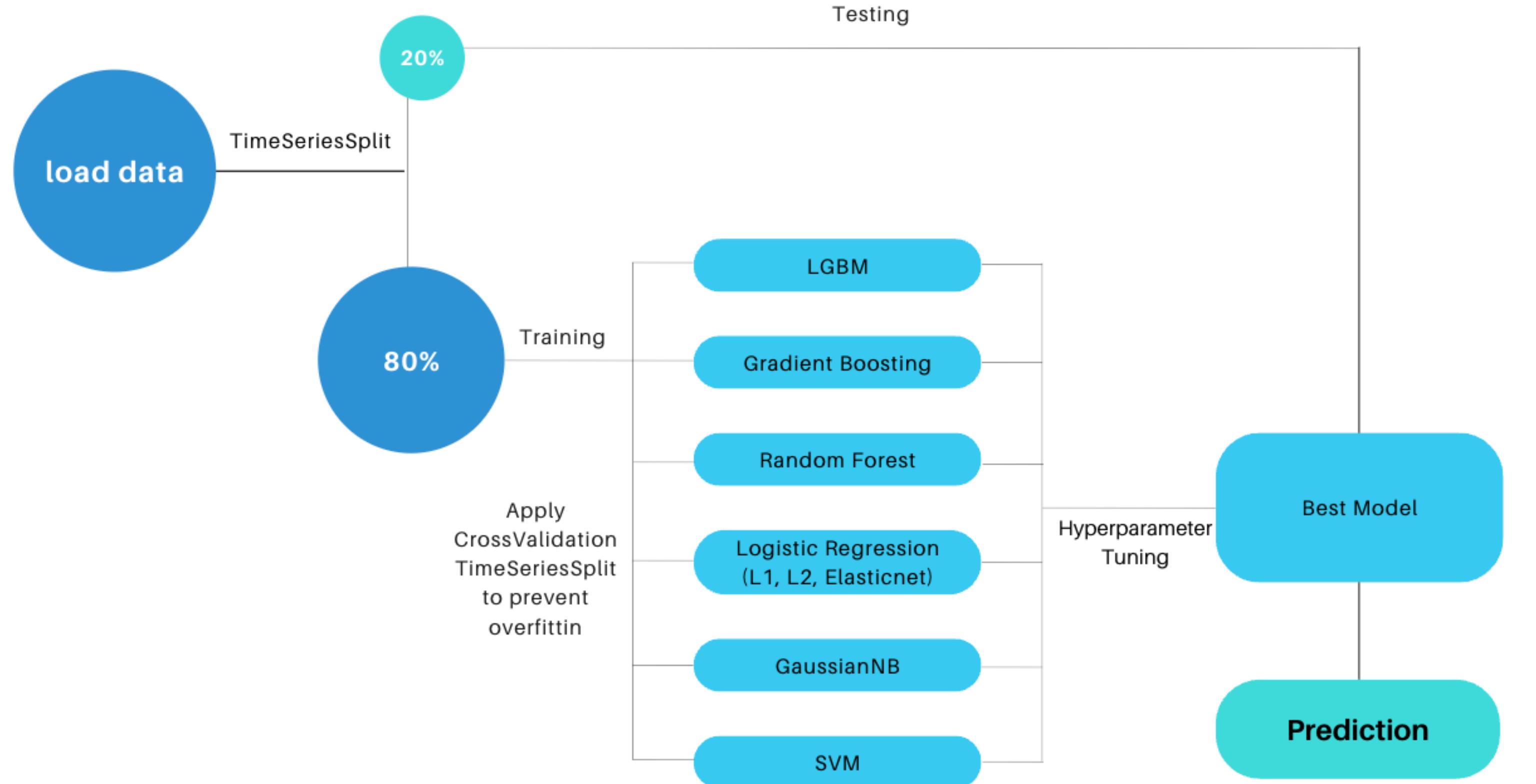
- Signal: เพื่อดูสำหรับขายหรือซื้อ ซึ่งได้แก่ RSI, MACD, STO และ CMF
- Time : วันที่เกิดการซื้อขาย, เวลาชั่วโมงของวันของข้อมูลหุ้นนั้นๆ
- ใช้ One hot encoding เพื่อให้ข้อมูล Categorical สามารถนำเข้าไปฝึกฝนในโมเดลได้

Numerical Data

- Robust Scaler : ให้ข้อมูลมีการกระจายที่น่าเสียดายแม้ว่าจะมีค่าผิดปกติ (outliers) โดยมีการใช้ median และ interquartile range (IQR) เป็นหลักในการปรับปรุง และง่ายต่อการนำไปใช้ในกระบวนการต่อไป



MODELING



SUMMARY AND EVALUATION



- Model : Gaussian naïve bayes
- Feature: Stepwise_Gaussian
- Train_Score: 0.739
- Test_Score: 0.838
- Avg. CV score: 0.703
- Overfitted_ratio: 0.882



- Model: xgb bayesian
- Feature:
- Stepwise_gradientboosting
- Train_Score: 0.785
- Test_Score: 0.784
- Overfitted_ratio: 1.002



- Model: Gaussian naïve bayes
- Feature: Stepwise_Gaussian
- Train_Score: 0.734
- Test_Score: 0.722
- Overfitted_ratio: 1.017

A black and white photograph showing a close-up of two people's hands shaking over a table. On the table, there are some papers and a pen. The person on the left is wearing a striped shirt, and the person on the right is wearing a dark suit jacket. The background is slightly blurred.

**THANK
YOU**