JPX Stock Market Prediction

DAN CHEN

2021.12

Problem Statement

HOW TO EVALUATE AROUND 2000 STOCKS RETURNS, AND RANK THE STOCKS FROM HIGHEST TO LOWEST RETURNS?

Background

When a stock or derivative is undervalued, it's time to buy, overvalued, sell. While most of finance decisions was made manually by professionals, technology was ushered in new opportunities for retail investors

For data scientist, make predictions based on trained models with programming skills on raw data.

In this project, I built a strategy to analyze the real JPX stock data with stock indicator analysis, bayes optimization, LGBT model tuning, etc. and use model to predict stock market and ranked the returns

Data: JPX Tokyo Stock dataset from (2017 – 2021)

Workflow

Data Collection

Downloaded dataset
using Kaggle API

Feature Engineer

Processed data and changed
and transformed features for
analysis

Stock Indicator

Built functions to executed the
evaluation with Bollinger
bands, MACD, RSI, Stochastic
Oscillator , and created
buy/sell signals

Model Tuning

Tuned hyperparameters with
bayes optimization and used
LGBT model to do prediction

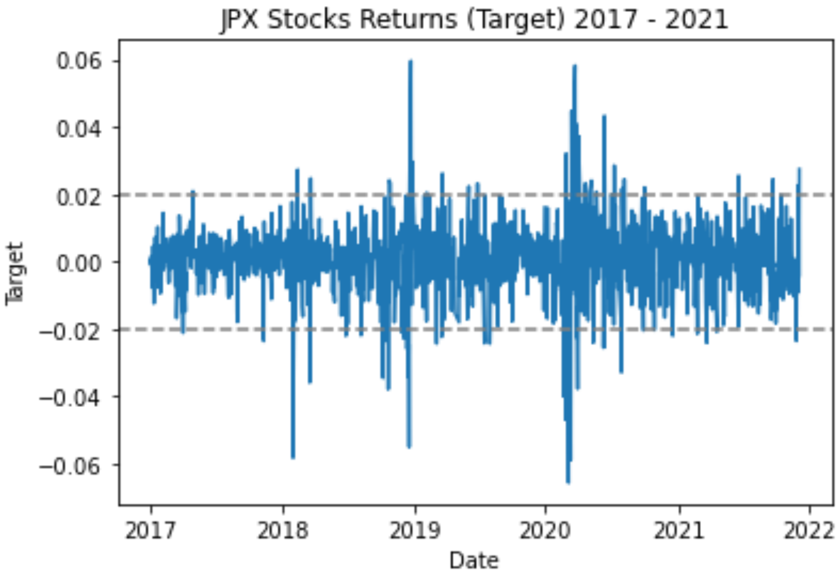
Rank Returns

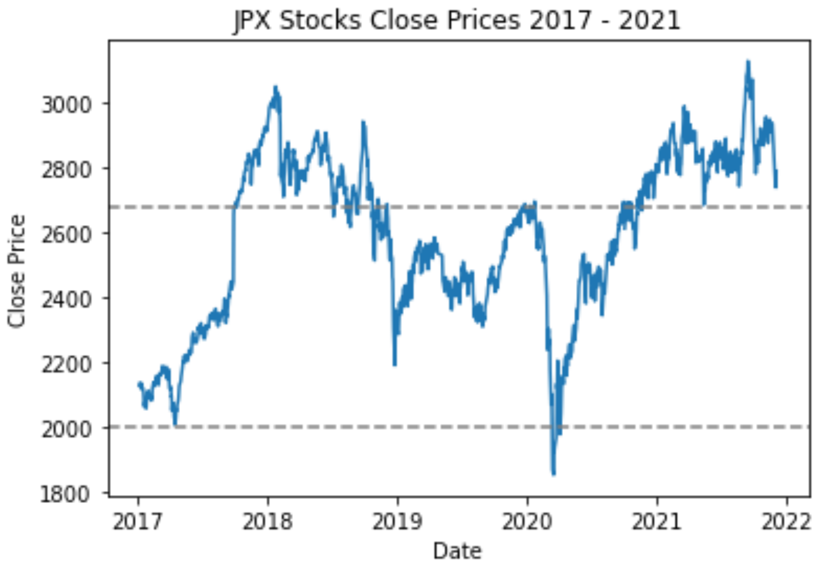
Calculated the returns and
ranked from high to low

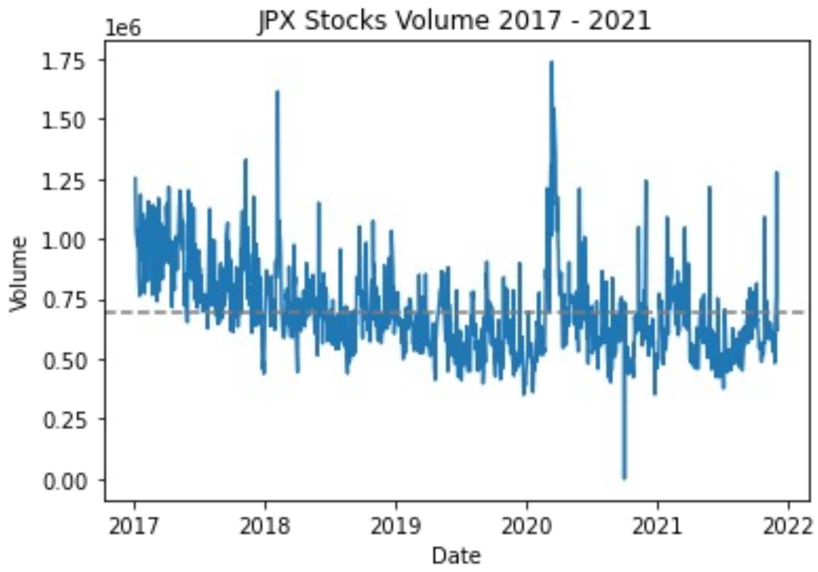
Data

- There are two datasets
  - Stock values dataset(train), data shape 2332531 X 12
  - Stock list datset, data shape 4417 X 16
- Merged dataset
  - Data shape 2332531 X 14
  - Columns: Date, SecuritiesCode, open, high, low, close, volume, AdjustmentFactor, ExpectedDividend, SupervisionFlag, Target, Name, SectorName
- EDA

EDA





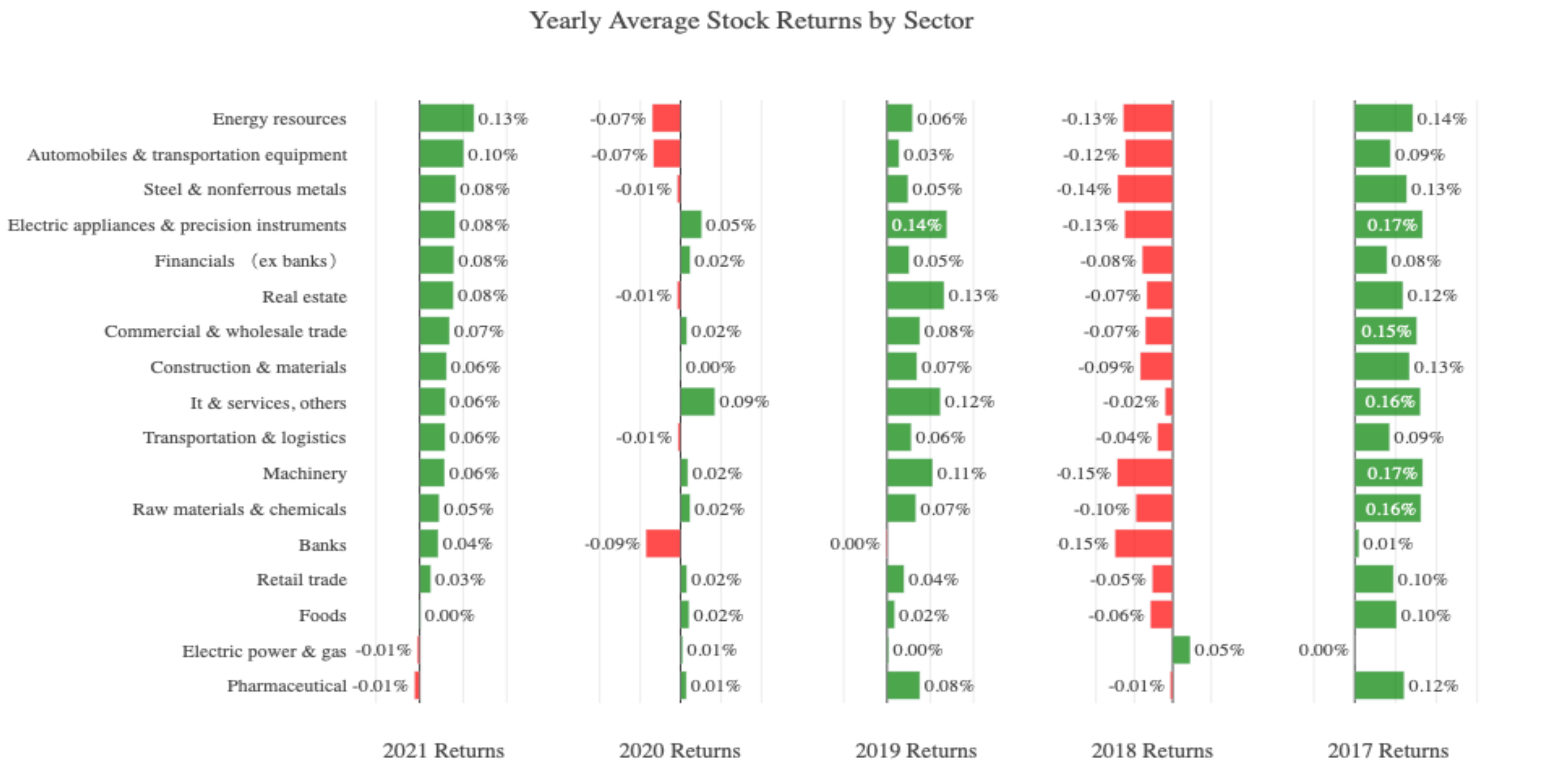


Returns range between -2% and 2%
With 5 years

Close prices show increased trend,
But also show deep dive in 2020,
Probability due to COVID19

Volume slightly decreased
With 5 years

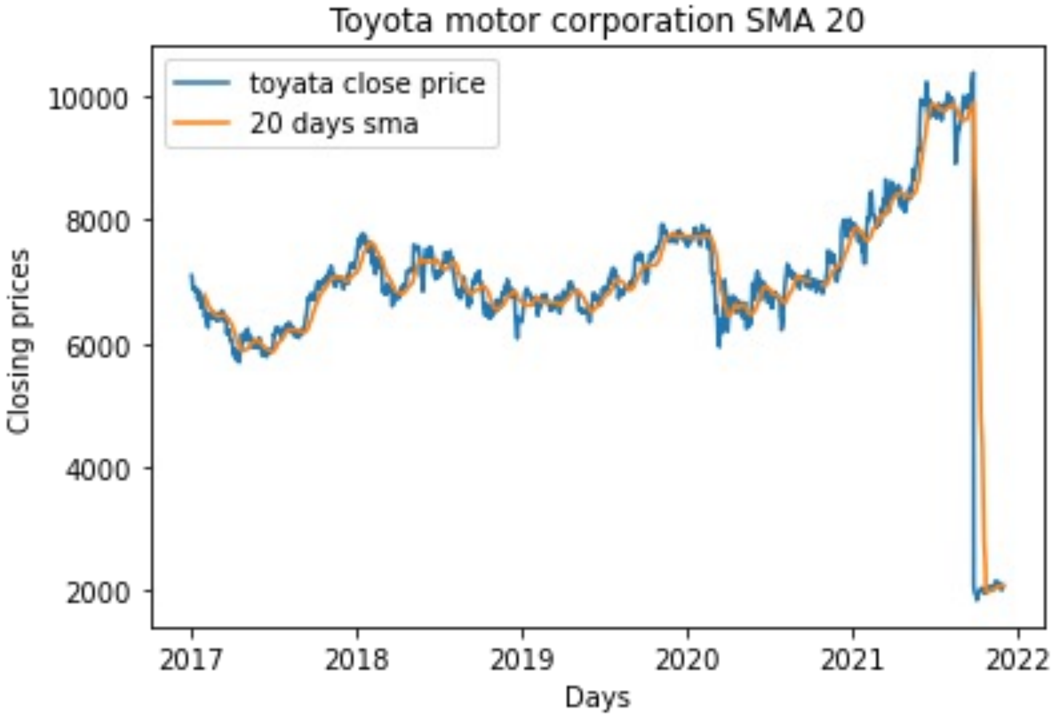
Annual Returns



Stock Prices Movements

Automobiles and transportation equipment stocks showed unclear trend. There was a significant dive early 2020 due to COVID19. After that, there was an increased trend

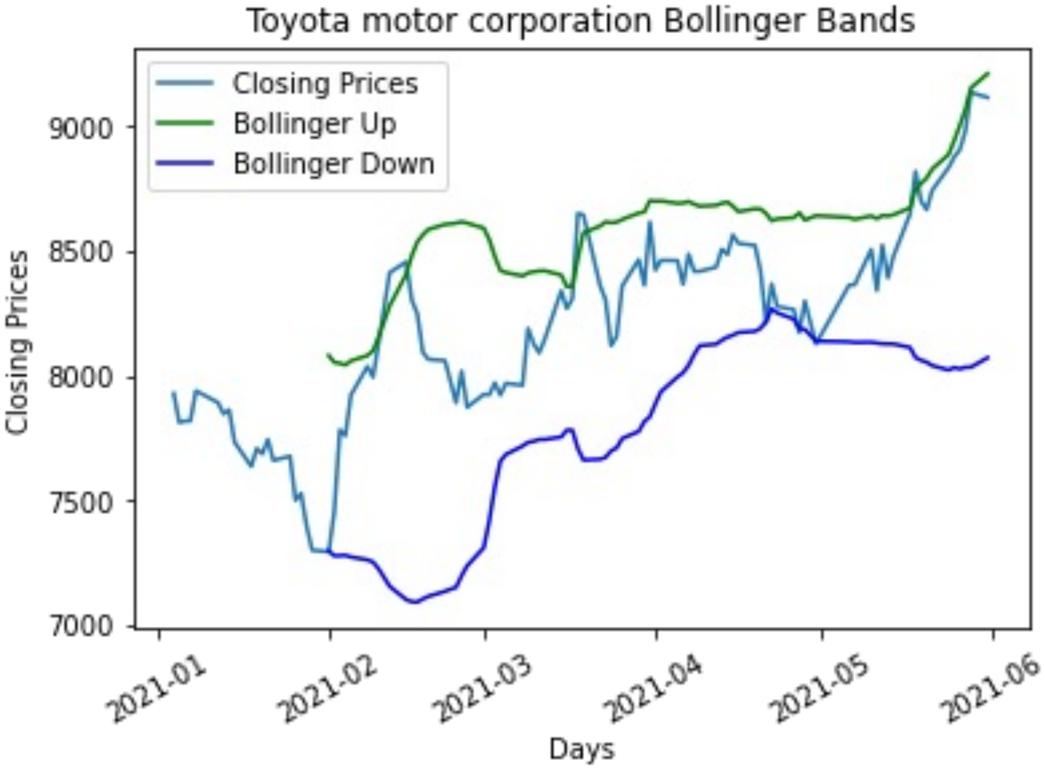
Stock Evaluation (TOYOTA)



Simple Moving Average (SMA)

Toyota stock for example

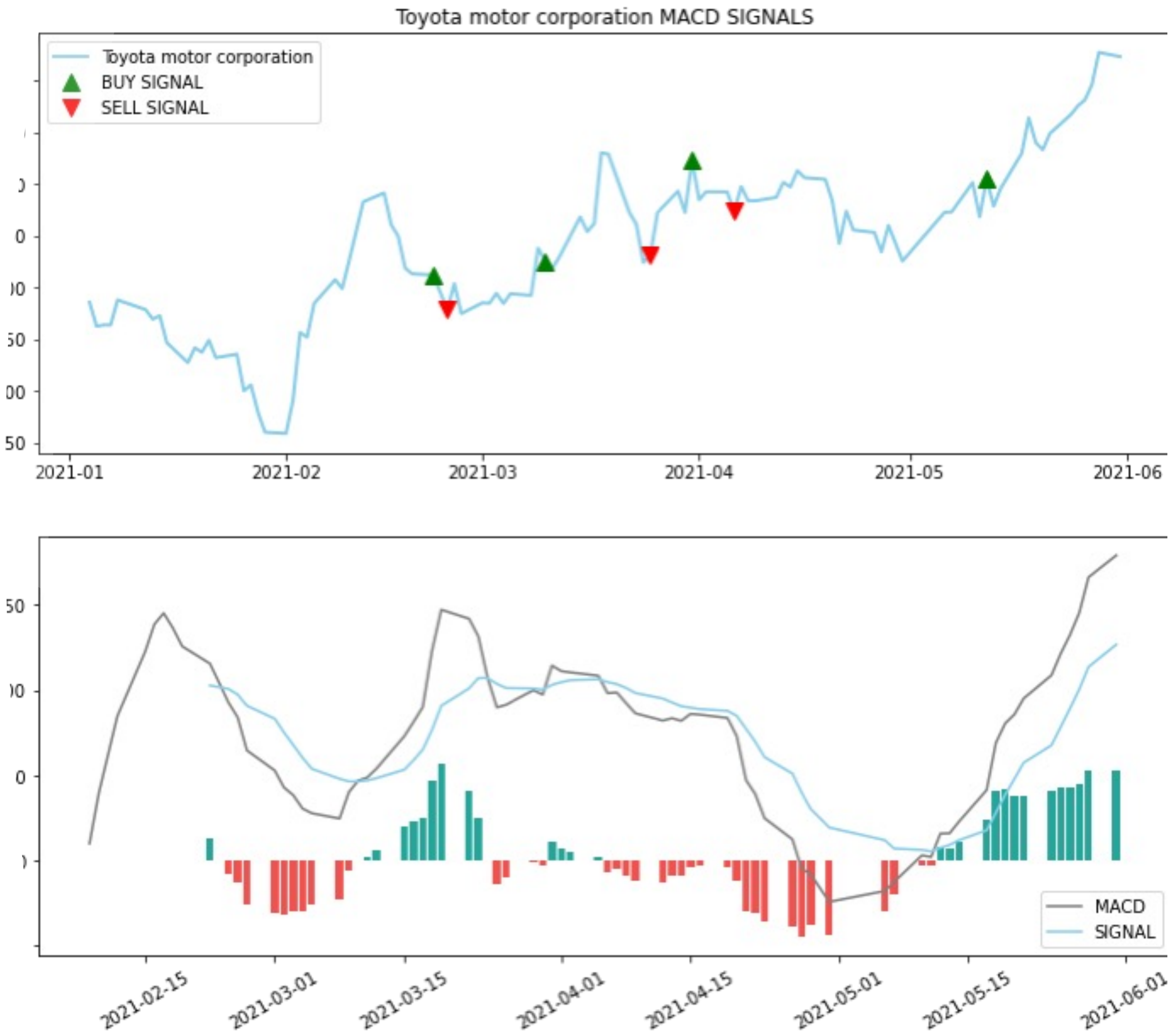
1. SMA was calculated by averaging past 20 days' close prices, which provided many more reversal signals than 100 days moving average.It provided more accurate signals on historical data.
2. Increased SMA20 trends collapsed in 2020



Bollinger Bands

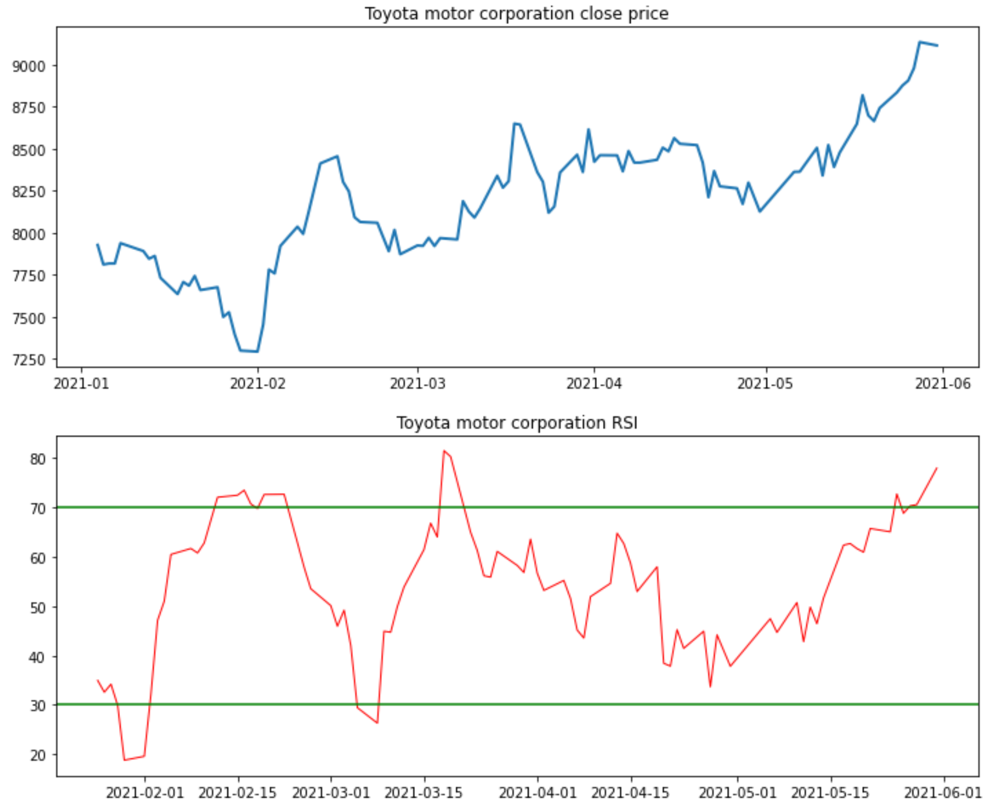
Bollinger bands help to identify sharp, short-term price movements, and potential entry and exit points. Above showed increased Bollinger bands

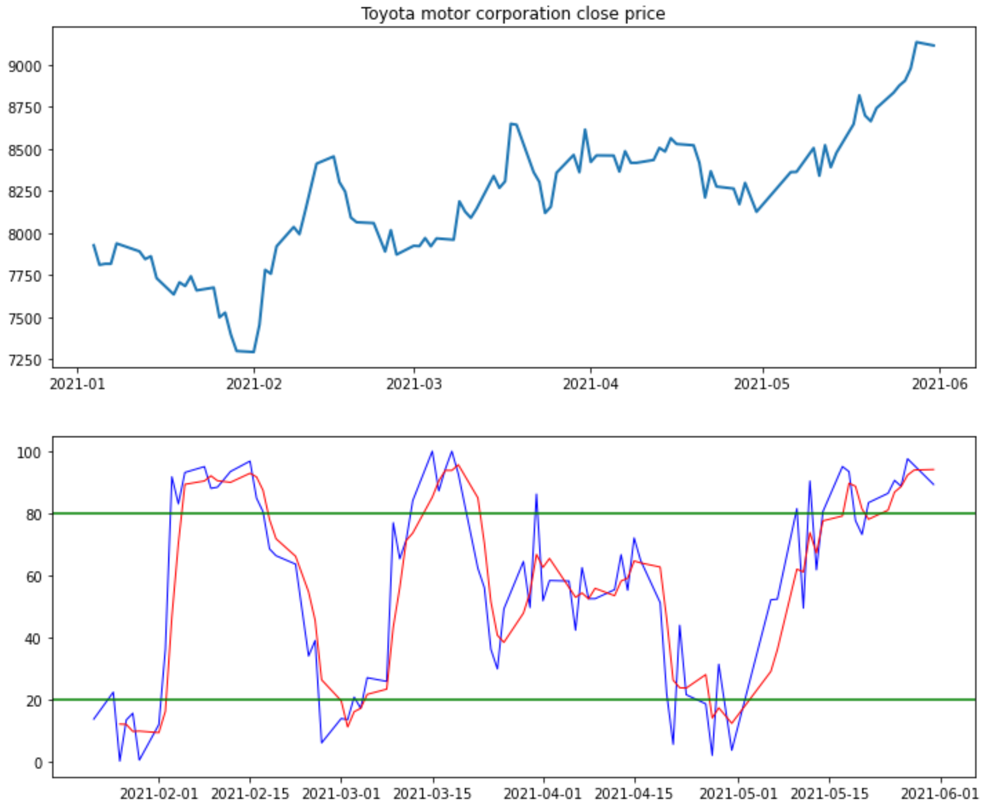
Buy / Sell Signal & MACD



Above: there are several buy and sell signals within 5 years. Entry at green signal, exit at red.
Below: MACD. Green shows bullish trends, red shows bearish trends.

RSI & Stochastic Oscillator





Relative Strength Index (RSI)

RSI measures speed and magnitude of price changes, which range between 0 and 100. RSI is considered overbought when above 70, and oversold when below 30. RSI shows prices vibrate between green lines

stochastic oscillator

The stochastic indicator establishes a range with values indexed between 0 and 100. A reading of 80+ points to a security being overbought, and is a sell signal. Readings 20 or lower are considered oversold and indicate a buy. Different from RSI, stochastic oscillator is more useful in sideways or choppy markets

Feature Engineering

- Adjusted the close prices
  - With cumulative adjustment close price
- Created new columns
  - return, moving average, exponential moving average, volatility columns
  - 9, 12, 26, 50 key days
  - Stock sharpe ratio, spread return
- Split features, label

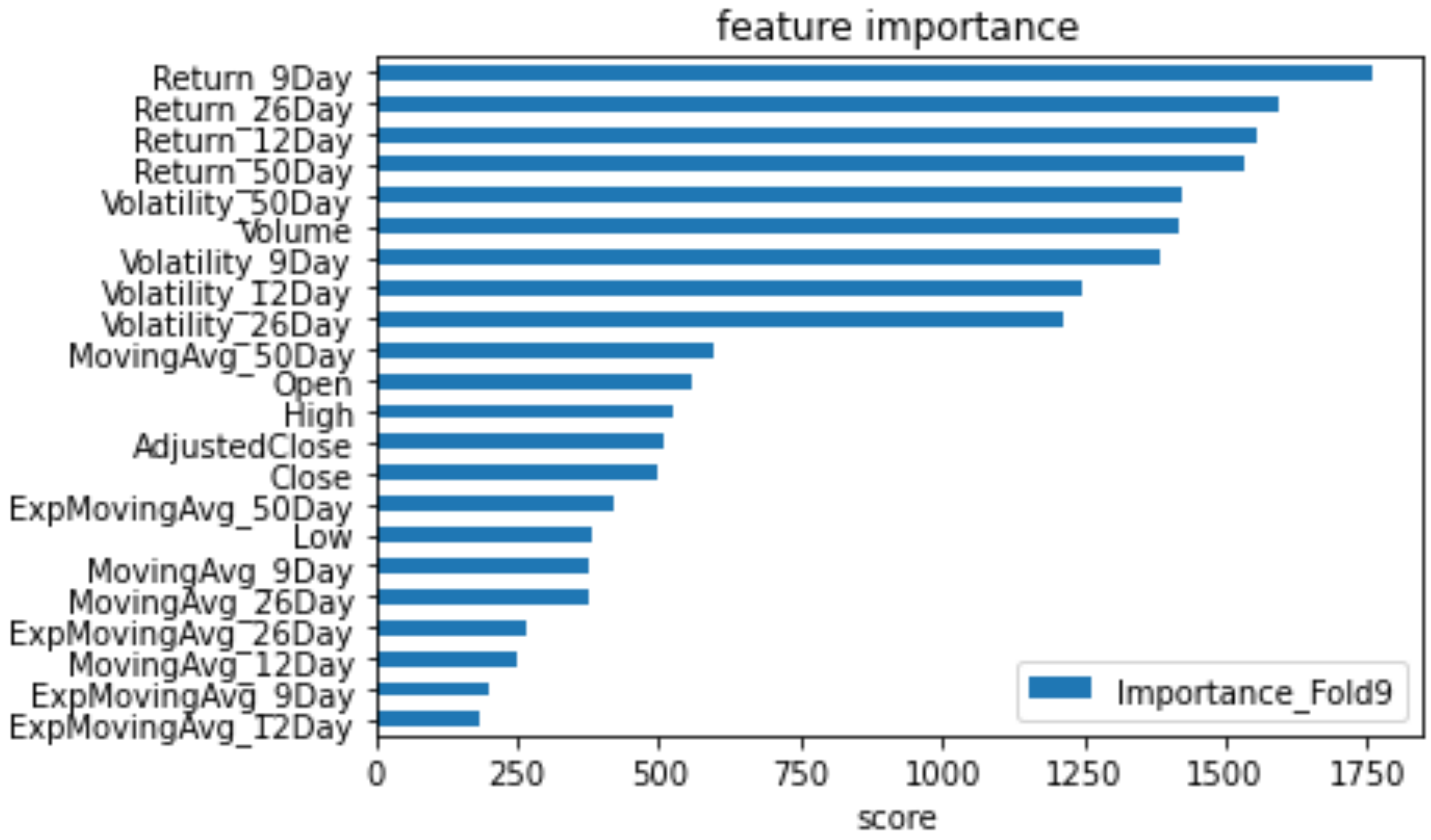
Bayes Optimization

- LGBT parameters tuning
  - LGBT parameters optimization
  - With cumulative adjustment close price
- Bayesian optimization
  - Compared the rmse to selection hyperparameters

```
params = {
          'num_leaves': 187,
          'learning_rate': 0.073,
          'max_depth': 12,
          'min_data_in_leaf': 167,
          'min_sum_hessian_in_leaf': 0.003,
          'bagging_fraction': 0.855,
          'feature_fraction': 0.837,
          'drop_rate': 0.261,
          'random_state': 42
}
```

- Updated LGBT model with selected parameters to predict stock
- Calculated the sharpe ratio

Feature Importance



9, 12, 26, 50 days returns are more important features, than exponential moving average

Plot Tree