

# Stock prediction using transformer

group 32

309552007 袁鈺勛

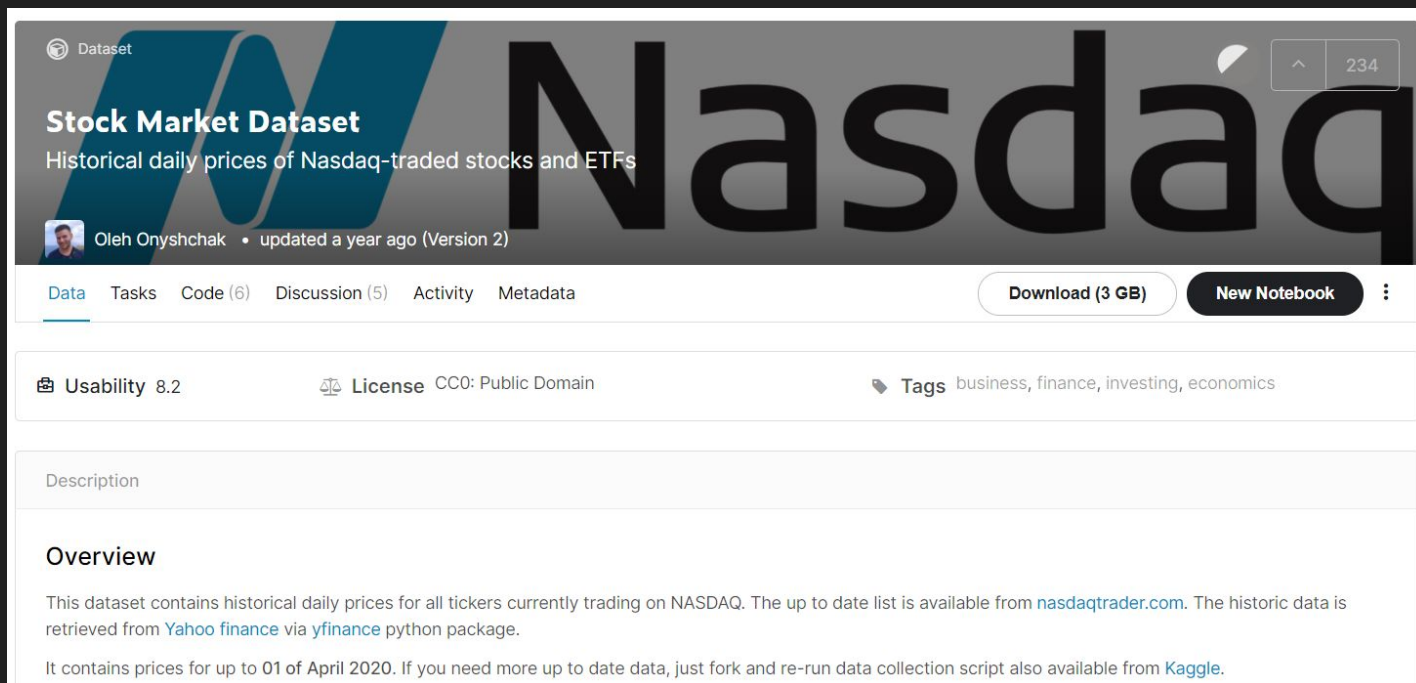
0716039 謝仁翔

# Introduction

- We want to predict companies' closing returns given stock features.
- There are 8 features used for the prediction:
  - Open
  - High
  - Low
  - Closing
  - Volume
  - Periodic time
  - Non-periodic time
  - Symbol

# Dataset

The dataset is from Kaggle.



Dataset

## Stock Market Dataset

Historical daily prices of Nasdaq-traded stocks and ETFs

Oleh Onyshchak • updated a year ago (Version 2)

Data Tasks Code (6) Discussion (5) Activity Metadata

Download (3 GB) New Notebook

Usability 8.2 License CC0: Public Domain Tags business, finance, investing, economics

Description

### Overview

This dataset contains historical daily prices for all tickers currently trading on NASDAQ. The up to date list is available from [nasdaqtrader.com](https://nasdaqtrader.com). The historic data is retrieved from [Yahoo finance](#) via [yfinance](#) python package.

It contains prices for up to 01 of April 2020. If you need more up to date data, just fork and re-run data collection script also available from [Kaggle](#).

# Features

- Open, high, low, closing, and volume
  - These 5 features are converted to percentage changes.
- Periodic time and non-periodic time
  - Periodic time is the sine function of the mean value of the 5 features above.
  - Non-periodic time is the linear function of the mean value of the 5 features above.
- Symbol
  - The sum of the embeddings of each character in the symbol.

# Model Structure

- We use a multihead transformer encoder and three linear layers to predict the percentage change during training process.
- The predicted percentage changes will be used to produce the closing returns in inference.
- Input of the model
  - (batch size, sequence length, 5)
- Input of the transformer encoder
  - (batch size, sequence length, 8)
- Output
  - Percentage changes

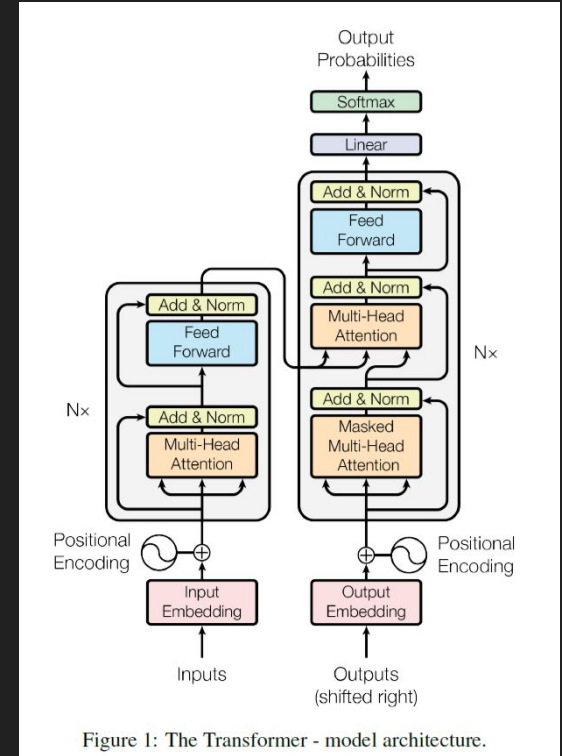


Figure 1: The Transformer - model architecture.

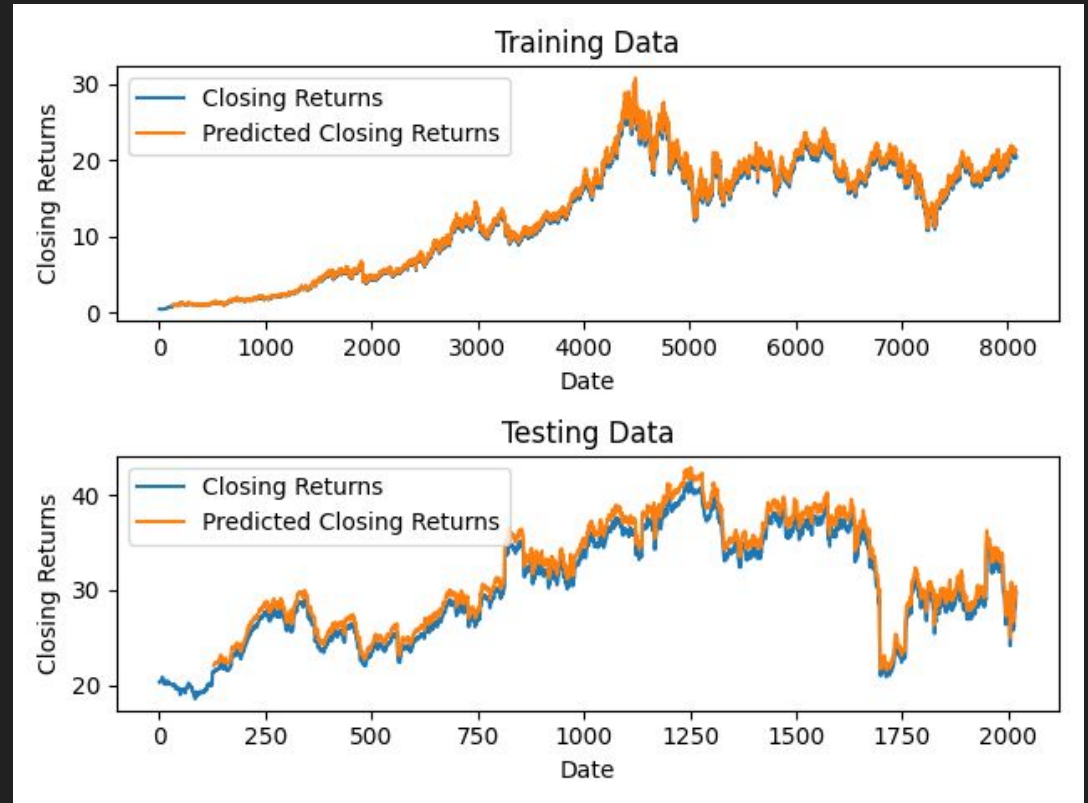
# Loss Function

- We use L2 loss as loss function.
- $E_{x,y \sim p_{\text{data}}}[(y - y'(x))^2]$

# Hyperparameters

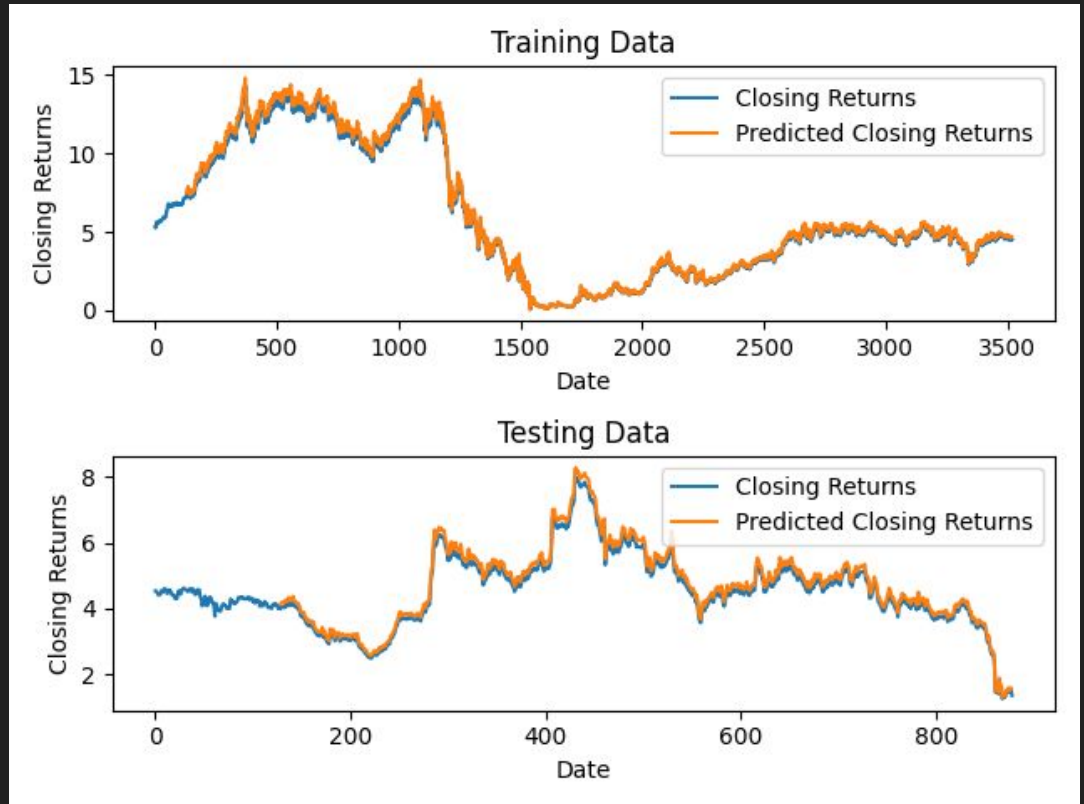
- Batch size: 256
- Learning rate: 0.001
- Sequence length : 128
- Number of encoders: 6
- Attention dimension: 128
- Hidden size: 512
- Dropout: 0.3
- Number of heads: 8

# ConAgra Brands, Inc. (CAG)





# Drive Shack Inc. (DS)



# Magellan Health Inc. (MGLN)



# Conclusion

- Predicted closing return is not accurate enough.
- Future works:
  - Use the date as time feature directly
  - Longer sequence length and deeper model
  - Use another way to embed the symbol

Q&A