Stock Movement Prediction

Tae yoon Lim (tl2968)
Hojin Lee (hl3328)

Data Source and Scrape

Data Source:
https://www.dukascopy.com/swiss/english/marketwatch/historical/

Data Scraping with the existing github repo: https://github.com/giuse88/duka/

About Data

	Local time	Open_x	High_x	Low_x	Close_x	Volume_x	Open_y	High_y	Low_y	Close_y	Volume_y
0	02.01.2020 09:30:00.000 GMT-0500	296.248	296.268	295.367	295.678	0.7800	296.302	296.303	295.443	295.742	0.6639
1	02.01.2020 09:31:00.000 GMT-0500	295.768	295.818	295.657	295.748	0.0825	295.862	295.882	295.753	295.842	0.0808
2	02.01.2020 09:32:00.000 GMT-0500	295.748	295.778	295.678	295.758	0.0975	295.842	295.842	295.752	295.842	0.0806
3	02.01.2020 09:33:00.000 GMT-0500	295.758	295.838	295.637	295.838	0.4275	295.812	295.882	295.702	295.882	0.4231
4	02.01.2020 09:34:00.000 GMT-0500	295.838	296.897	295.838	296.868	1.8375	295.883	296.953	295.883	296.942	1.5123

- 1 min frequency Bid/Ask Candle AAPL data from Jan.2.2020 to Dec.11.2020

Preprocess

Open	Close	High	Low	Volume_x	Volume_y	MA	Returns	BBand_upper	BBand_middle	BBand_lower	Average Directional Index	Directional Index	MACD	MACD_signal	MACD hist	stochastic k	stochastic d	movement
-0.010889	122.185	-0.011785	-0.010825	0.292769	0.304425	-0.013394	0.971103	0.044383	0.070090	0.326082	0.138746	0.196830	0.943951	0.943662	0.769829	0.570910	0.378162	0
-0.010954	122.210	-0.011874	-0.010902	0.153018	0.147788	-0.013410	0.971103	0.044479	0.070199	0.326130	0.136073	0.150448	0.944056	0.943701	0.769962	0.832032	0.589005	0
-0.010889	122.235	-0.011810	-0.010890	0.324469	0.343363	-0.013393	0.971658	0.044522	0.070311	0.326245	0.135980	0.179121	0.944158	0.943755	0.770062	0.839690	0.754837	1
-0.010825	122.230	-0.011784	-0.010865	0.361204	0.387611	-0.013370	0.971733	0.044274	0.070409	0.326669	0.136874	0.190892	0.944233	0.943815	0.770104	0.874152	0.857092	1
-0.010851	122.455	-0.011219	-0.010683	0.712477	0.765929	-0.013285	0.972752	0.044603	0.070547	0.326496	0.154827	0.396480	0.944453	0.943914	0.770354	0.908502	0.882878	1

Preprocess with TA-LIB

- MA: Rolling Window of close price
- Returns: Log returns of MA
- BBand: Bollinger Band of Close Price
- MACD: Moving Average Convergence/Divergence
- Stochastic k,d: Stochastic Oscillator Slow in k and d period
- Label: Movement, Returns > 0 → 1, Returns <= 0 → 0

MinMax Scaler for the Scaling

Embedding

Time2Vec: Learning a Vector Representation of Time

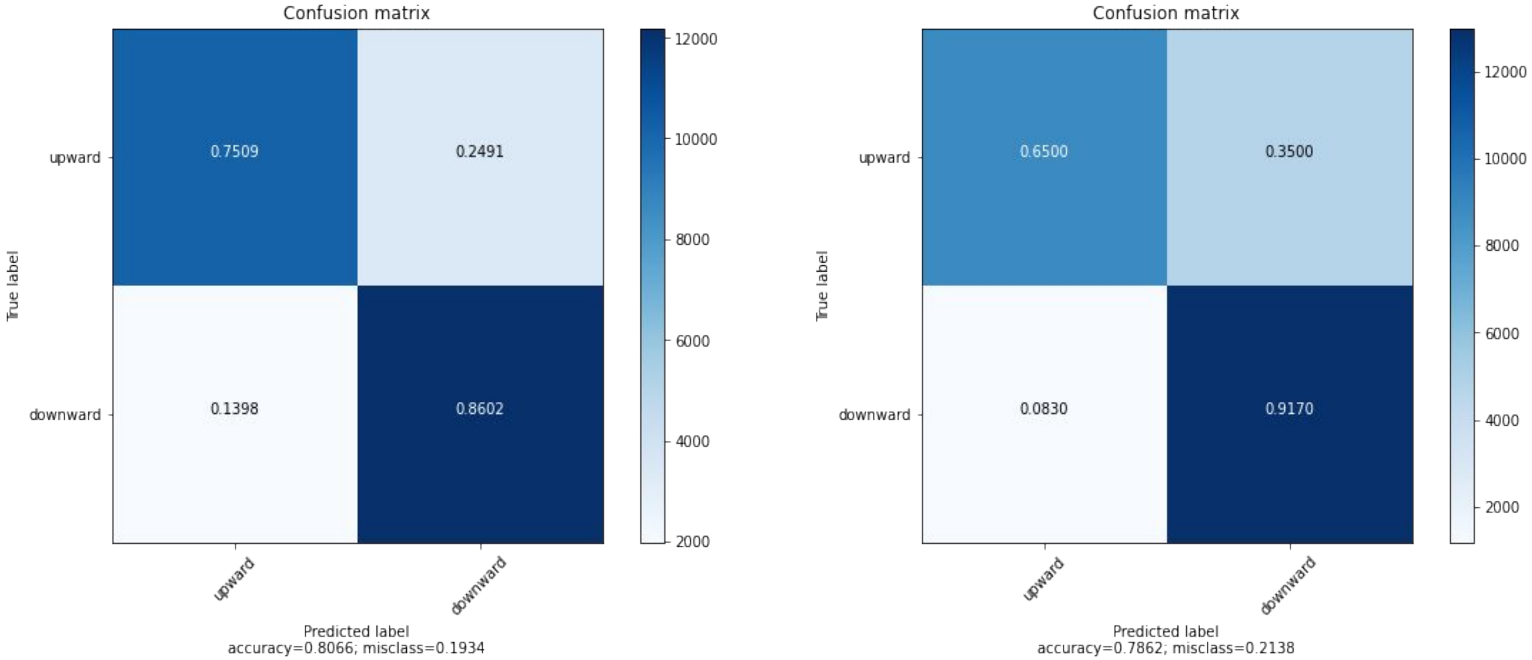
- Periodicity:
  - representation of time includes both periodic and non-periodic patterns
- Invariance to Time Rescaling
  - Time representation is not affected by different time increments
- Simplicity
  - A representation is easily consumable to other neural models

Model

We wanted to test whether different architecture of model yields better results

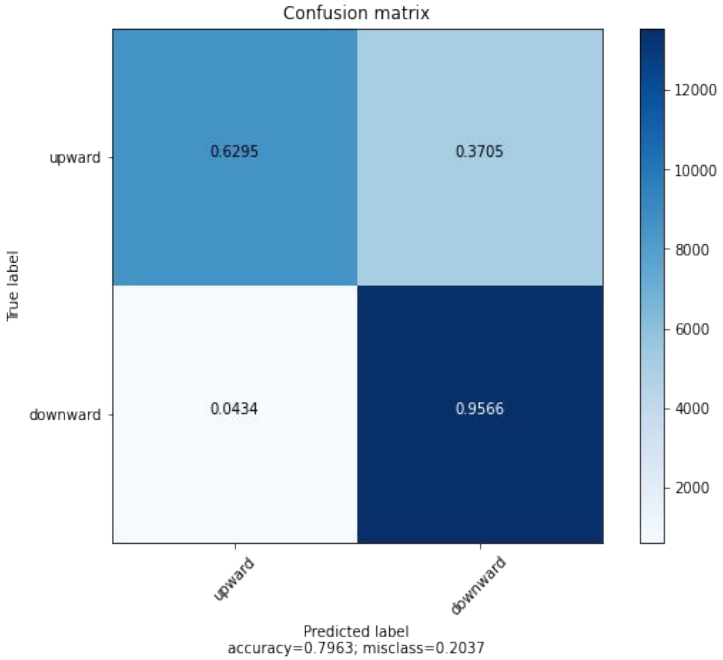
- LSTM
  - Well-known for processing the sequential data and yield a good results
- LSTM with Attention
  - As attention layers perform a great job in NLP task, we wanted to know whether it performs well on time series financial data
- Transformers
  - One of NLP's breakthrough, we wanted to know whether transformers perform a great job in time series financial data
- VLSTM (Very Long Short-Term Memory Networks)
  - By looking at the data in different period, the network can retrieve the long term signals and short term signals from the input sequence

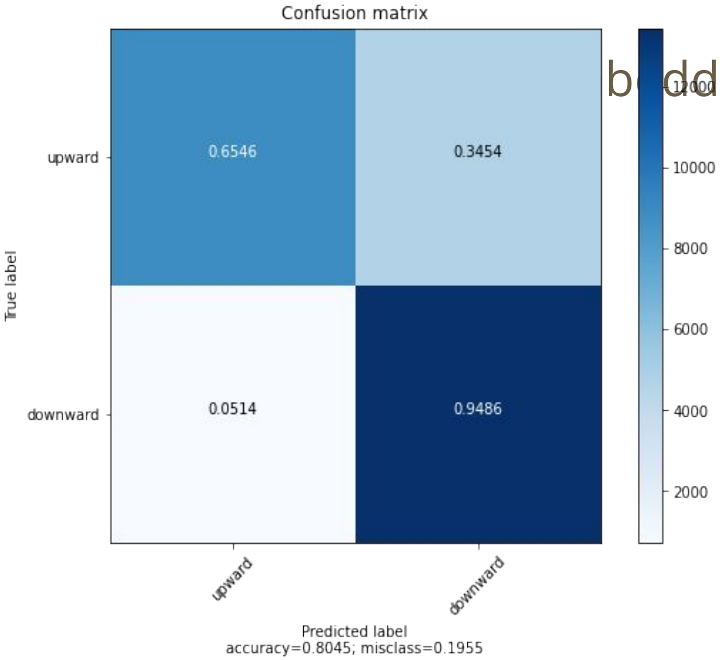
Result: LSTM



Left: w/ Time2Vec, Right: w/o Time2Vec

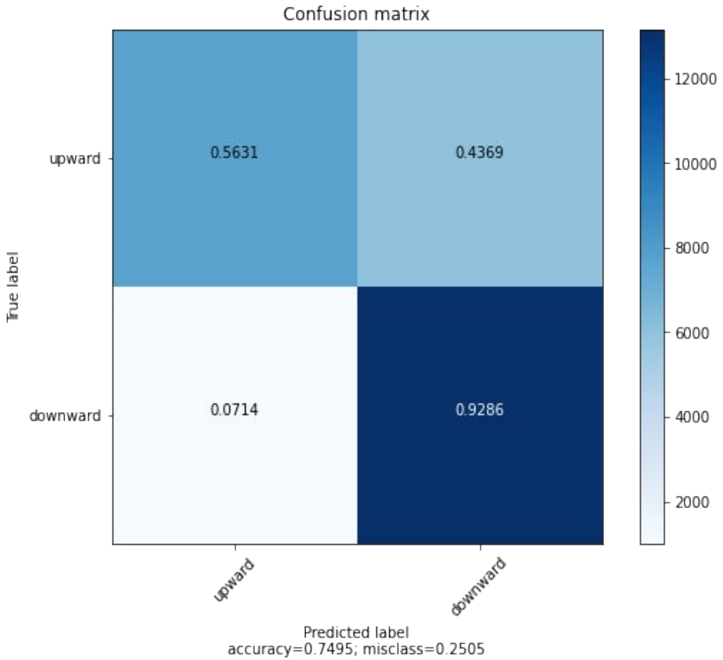
Result: LSTM with Attention

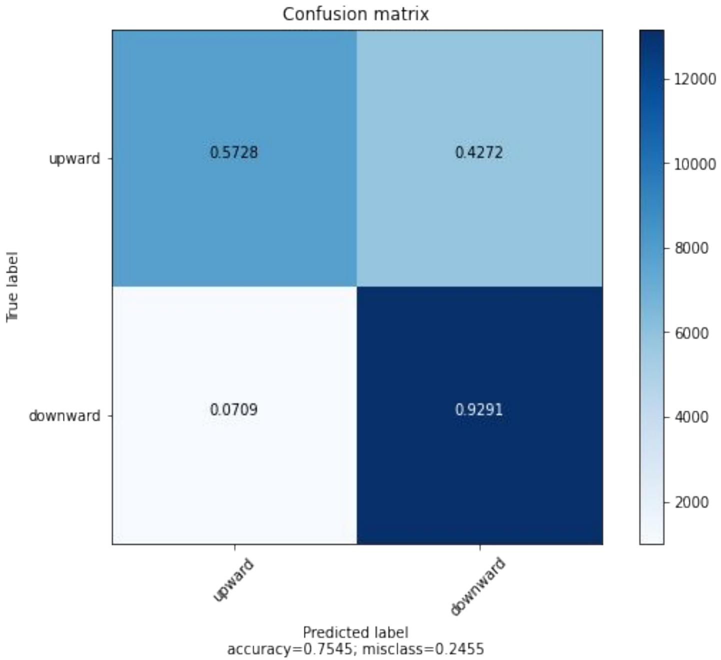




Left: w/ Time2Vec, Right: w/o Time2Vec

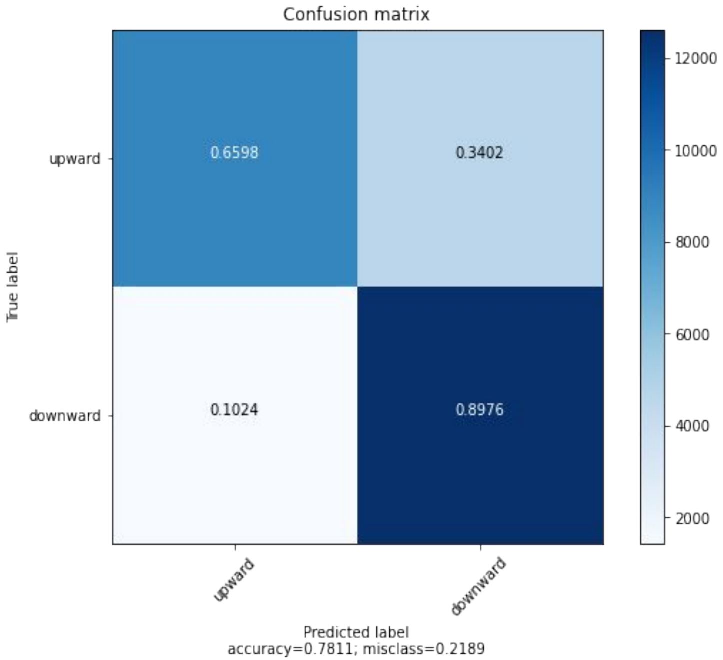
Result: Transformers (Attention is all you need)

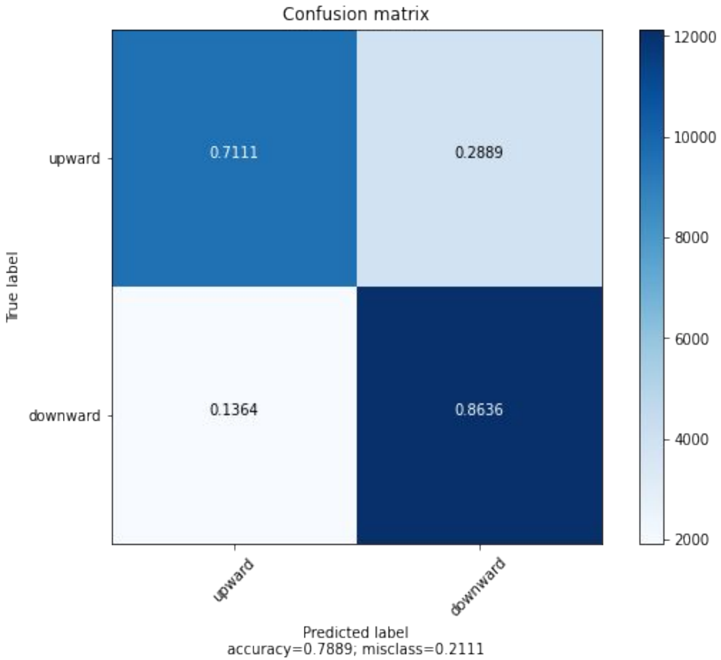




Left: w/ Time2Vec, Right: w/o Time2Vec

Result: VLSTM

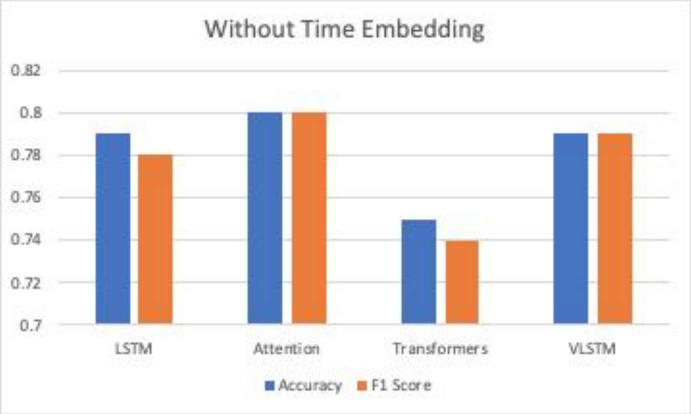


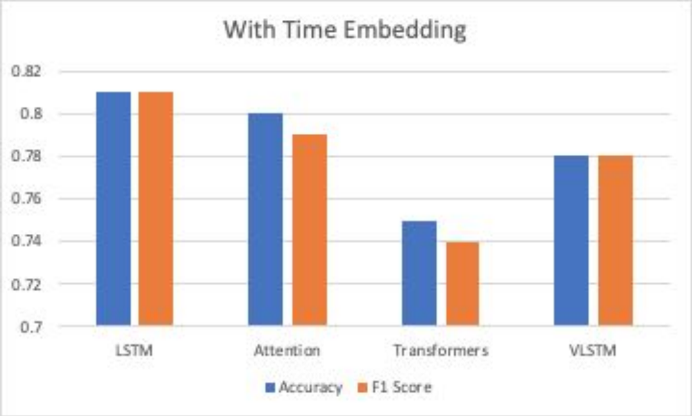


Left: w/ Time2Vec, Right: w/o Time2Vec

Result Table

	Without Time Embedding				With Time Embedding			
Models	LSTM	Attention	Transformers	VLSTM	LSTM	Attention	Transformers	VLSTM
Accuracy	0.79	0.8	0.75	0.79	0.81	0.8	0.75	0.78
F1 Score	0.78	0.8	0.74	0.79	0.81	0.79	0.74	0.78





Possible Reason of Failure

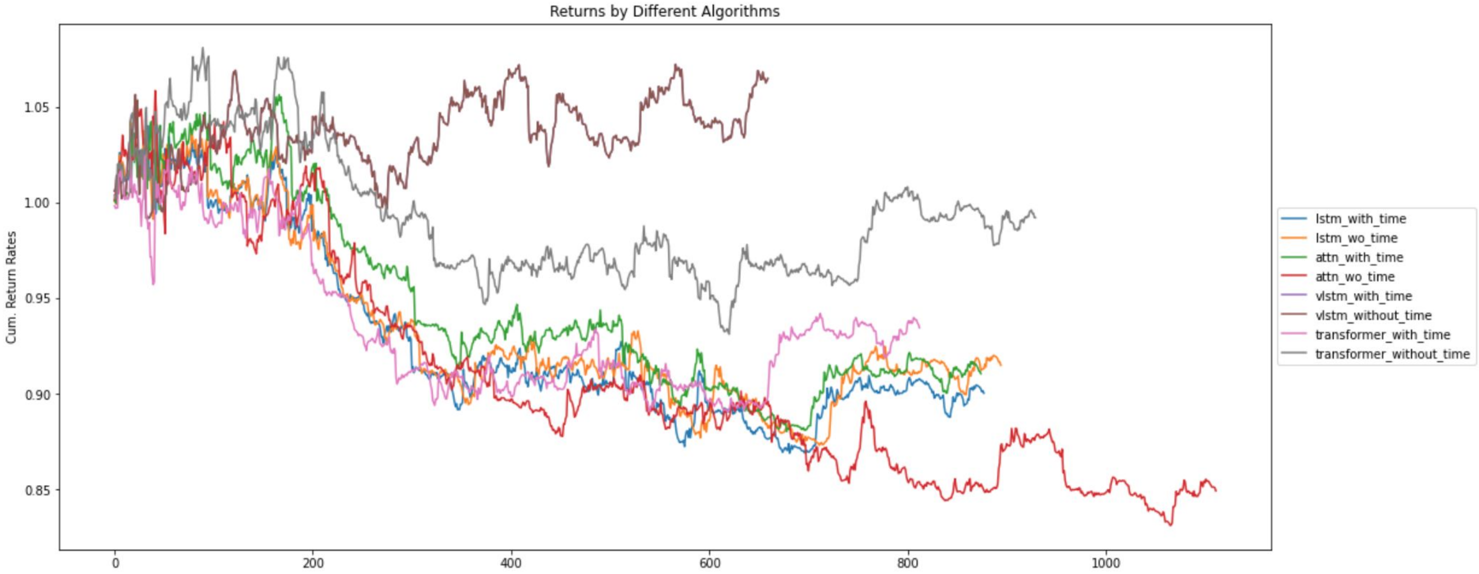
Despite more complicated architecture, VLSTM or Transformers do not yield better result than the baseline model (LSTM)

- Input data is more important than the model architecture
  - Previously, we have used only mid price and bid ask spread for the prediction, but the model does not learn anything from the data (only producing ‘upward’ label)
  - With additional financial indices and indicators, we have better result
- VLSTM
  - Used different input data
  - Difference in preprocessing and feature engineering
- Attention is not relevant for the time series data
  - NLP contains the same event (same token) whereas time series do not have such. Transformers are good at working with repeated tokens

Trading using predictions

Simple strategy: Buy at upward, sell at downward with single stock unit

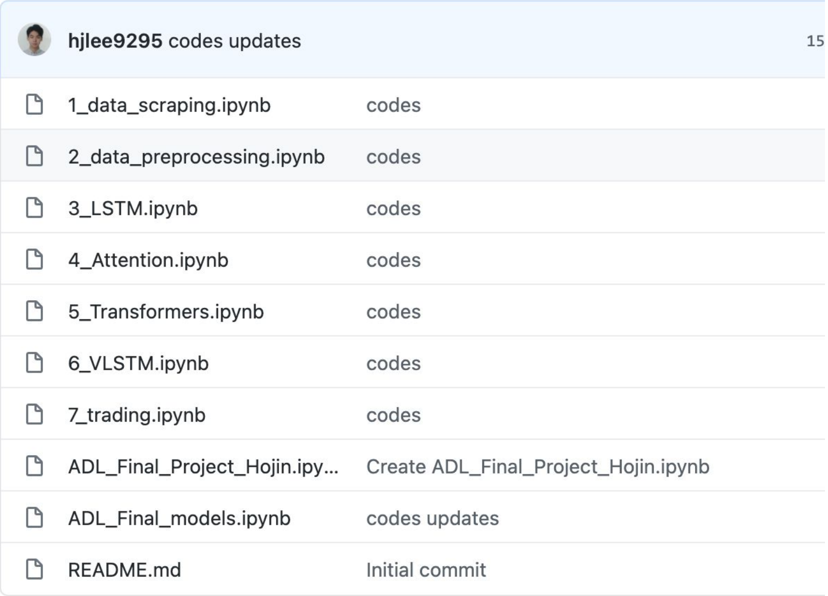
<matplotlib.legend.Legend at 0x7fdc338ad160>



Future Work

- Hyperparameter Tuning for model training
- Create better features and preprocessing with the data
- Build the trading strategy with the reinforcement learning
- Calculate loss from trading strategy for prediction
- Use different time frequency

Code Structure



- Total of 7 notebooks
- Run them in order
- https://github.com/hjlee9295/Stock_prediction_through_DeepLearning

Reference

- Kazemi, S., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., . . . Brubaker, M. (2019, July 11). Time2Vec: Learning a Vector Representation of Time. Retrieved December 22, 2020, from https://arxiv.org/abs/1907.05321
- Ganesh, P., & Rakheja, P. (2020, October 22). VLSTM: Very Long Short-Term Memory Networks for High-Frequency Trading. Retrieved December 22, 2020, from https://arxiv.org/abs/1809.01506
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., . . . Polosukhin, I. (2017, December 06). Attention Is All You Need. Retrieved December 22, 2020, from https://arxiv.org/abs/1706.03762
- Ntakaris, A., Magris, M., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2020, March 11). Benchmark Dataset for Mid-Price Forecasting of Limit Order Book Data with Machine Learning Methods. Retrieved December 22, 2020, from https://arxiv.org/abs/1705.03233
- Jan Schmitz, Cap Market, (2020), GitHub repository, https://github.com/JanSchm/CapMarket