TU Delft Delft University of Technology

Date	28/04/2019
Faculty	EWI
Course	CS4180 - Deep Learning

Team members	Deep learning ID
Daan Cuppen	7482796
Benjamin De Bosscher	9166571
Kipras Paliušis	9971528
Isaan Seminck	4083087
Jeff Maes	13175487

Table of Contents

Problem

The purpose of this project is to predict stock market values based on a Long Short-Term Memory (LSTM) deep learning model. Only based on previous stock data (no news events will be considered), this time series model will predict whether a certain stock will rise or fall in the near future. This algorithm could help traders during the decision when to buy, keep or sell their stocks. As basis, an existing algorithm will be used. This model will be extended using other input data and alternating the time series window. This way, we hope to achieve better results which can be used for longer periods of time (as we do not want to retrain our model every trading day).

Data

The data for this project won't be a problem as stock data is widely available. It can be downloaded, imported using an API (eg. Quandl), etc. The data that will be used for this project are all the underlying stocks of the Amsterdam Exchange Index. For every trading day, the open/close/high/low will be of importance together with the traded volume. The data will be saved in .CSV files.

Method

The method for this project is to use a Long-Short Term Memory Recurrent Neural Network to predict future stock prices using stock data from the past. An implementation of this method already exists in the form of a tutorial on datacamp.com called "Stock Market Predictions with LSTM in Python". The existing implementation will be used as the base for our implementation.

While the existing model used the average price of the open/close stocks and the a fixed time series for the training dataset, the new model will be improved by first varying the time series interval of the training data. This will allow to analyse whether limiting the time series length can lead to better predictions. It will also be looked into whether not only using the mid price of the stocks but also including the volume can lead to a better prediction model (change in input data).

Relevant reading previous work

The following references list was identified:

- Eunsuk Chong, Chulwoo Han, Frank C. Park (2017) "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies" **DOI** https://doi.org/10.1016/j.eswa.2017.04.030

- Manuel R. Vargas ; Beatriz S. L. P. de Lima ; Alexandre G. Evsukoff (2017) “Deep learning for stock market prediction from financial news articles” **DOI:** 10.1109/PDP2018.2018.00060
- Sean McNally ; Jason Roche ; Simon Caton (2018) “Predicting the Price of Bitcoin Using Machine Learning” **DOI:** 10.1109/PDP2018.2018.00060
- Wen Long, Zhichen Lu, Lingxiao Cui, (2019) “Deep learning-based feature engineering for stock price movement prediction” **DOI:** https://doi.org/10.1016/j.knosys.2018.10.034

Research Question

The research question is: What is the effect of varying the time series length and the day traded volume of stocks to an LSTM model on the prediction performance of the stock price of companies in the Amsterdam Exchange Index (AEX).

This research question is labeled as a preliminary research question since it will most likely undergo adjustments once we actually start with the project and gain in depth knowledge on the stated problem.

Evaluation

We will evaluate our results by visual inspection of the predicted stock price and the actual stock price over time as well as the prediction error over time. Quantitatively, evaluation will be executed by the following performance statistics: Root Mean Square Error (RMS), linear correlation, average absolute error, maximum absolute error between the actual prices and the predicted price and lastly the 95% confidence interval of the predicted prices around the actual prices. Using these statistics the performance of the LSTM networks (different time-series length and stock volume as additional input variable) can be compared.

Timeline

This timeline shows an estimate on the work division in Q4. The distractions of the team members are not included in this overview since it is discussed among the team members that each member will be able to contribute the indicated time per week.

Week	**Benjamin**	**Daan**	**Isaac**	**Jeff**	**Kipras**
First phase: project proposal					
4.1	Problem statement, dataset **(2h)**	Timeline, research question and feasibility analysis **(2h)**	Evaluation of results and analysis method **(2h)**	Timeline, research question and feasibility analysis **(2h)**	Method specification and selecting relevant literature**(2h)**
Second phase: constructing model					
4.2	Literature study **(3h)**, understanding and implementing basis LSTM framework from literature **(4h)**	Literature study **(3h)**, understanding and implementing basis LSTM framework from literature **(4h)**	Literature study **(3h)**, understanding and implementing basis LSTM framework from literature **(4h)**	Literature study **(3h)**, understanding and implementing basis LSTM framework from literature **(4h)**	Literature study **(3h)**, understanding and implementing basis LSTM framework from literature **(4h)**
4.3	Validate LSTM framework **(3h)**, investigate addition volume input **(3h)**	Validate LSTM framework **(3h)**, investigate addition volume input **(3h)**	Investigate addition volume input **(6h)**	Investigate addition volume input **(3h)**, Construct varying time series input **(3h)**	Investigate addition volume input **(3h)**, Construct varying time series input **(3h)**
4.4	Implement volume input to the code and validation **(6h)**	Implement volume input to the code and validation **(6h)**	Implement volume input to the code and validation **(6h)**	Implement time series input to the code and validation **(6h)**	Implement time series input to the code and validation **(6h)**

Third phase: conducting experiments and gathering results					
4.5	Experiments with volume input **(4h)**, Combined experiments **(2h)**	Experiments with volume input **(4h),** Combined experiments **(2h)**	Experiments with volume input **(4h),** Combined experiments **(2h)**	Experiments with varying time series **(4h),** Combined experiments **(2h)**	Experiments with varying time series **(4h),** Combined experiments **(2h)**
4.6	Processing data experiments **(4h)**	Processing data experiments **(4h)**	Processing data experiments **(4h)**	Processing data experiments **(4h)**	Processing data experiments **(4h)**
Fourth phase: report and poster					
4.7	Report **(6h)**	Report **(6h)**	Report **(6h)**	Report **(6h)**	Report **(6h)**
4.8	Report **(6h)**	Report **(6h)**	Report **(6h)**	Report **(6h)**	Report **(6h)**
4.9	Poster **(2h)**	Poster **(2h)**	Poster **(2h)**	Poster **(2h)**	Poster **(2h)**
Total hours:	45 hours	45 hours	45 hours	45 hours	45 hours

Feasibility

Indicator	Specific	Explanation	Verdict
Resources			
	Time of members	Most of the team members have only a few courses of the Masters' left, meaning that sufficient time is available. However, at the start, a working schedule should be set-up in which everybody indicates when he/she can work on the project.	Very good
	Knowledge	The team consists of 4 aerospace engineering students and 1 mechanical engineering student. This does not have to be a problem as we are all eager to learn new things and have a sufficient background in mathematics and programming. Since we are also not developing a new model ourselves, but rather extend on existing code, we avoid difficult programming problems.	Good
	Available data	The data used is from the stock exchange which is widely accessible. For example through Quandl, and many other sources.	Very good
Technical Depth		Stock price prediction can be rather difficult so to make this project feasible we start with an LSTM framework provided by https://www.datacamp.com/community/tutorials/lstm-python-stock-market. The input to the framework is varied and the model is extended to comply with a daytrade volume input.	Good

Looking at the above indicators, it becomes clear that an LSTM framework to predict stock prices is an excellent candidate for the project of the CS4180 Deep learning course.