COMP 702 - MSc Project

Financial Market Prediction

Michael Tattersall - 201518853

Specification & Design

Contents

Introduction

An outline of the project is as follows:

1. To perform a literature review of financial market prediction using artificial intelligence techniques
2. To review the application of such methods by money managers
3. To develop a state of the art artificial intelligence based prediction model for a financial market

Objectives:

- To compare and contrast existing artificial intelligence techniques by data scientists
- To research the application of such methods by money managers
- To design an artificial intelligence based prediction model for a financial market
- To identify and collect appropriate data for use in the model
- To evaluate and assess the performance of this prediction model
- To complete the deliverables of the project
- To make a contribution to the study of the problem that might be useful to others

Approach:

- To deliver a project that is research-based and involves the development and evaluation of artificial intelligence software
- To demonstrate a deep understanding of the problem
- To implement an artificial intelligence solution to the problem

- To develop my own concepts and ideas by looking at the problem in a novel way
- To implement the research skills developed in COMP 516 - Research Methods in Computer Science
- To develop new skills and techniques as necessary, such as programming and algorithms
- To develop a piece of software in the context of the problem and to evaluate critically
- To report the results and to address them in a way which makes them potentially applicable in a broader context

Ultimately, it is intended that my research and development will fill a gap by blending the academic research and industry best practice.

Summary of the Proposal

The problem addressed by the project is how to predict the market price of publicly traded financial instruments. These include stocks (equity) of public companies, the bonds (debt) of sovereign governments and corporations, futures contracts (such as for commodities or market indices), foreign currencies (forex), and warrants/options (derivatives). The implication of having the knowledge and capability to predict such prices may empower a trader with the financial means to implement a profitable strategy in the markets, and to build capital. That said, it is empirically difficult to consistently and profitably predict financial markets.

This problem has been addressed in a number of ways by investors and speculators for as long as financial instruments have been traded. In recent decades increased resources have been devoted to the quantitative trading of financial markets, which uses quantitative analysis and mathematical models to analyse the change in price and volume of securities in the stock market.[1]

According to the Financial Stability Board (FSB)[2], the international body established to coordinate the work of national financial authorities, artificial intelligence (AI) is being rapidly adopted for a range of applications in the financial services industry. In particular, institutions are optimising capital with artificial intelligence, whilst hedge funds and other firms and individuals (money managers) are applying these techniques to find signals for higher and uncorrelated financial returns on the money they are responsible for managing for investors.

The increased availability of large datasets, digital data sources, computing power and artificial intelligence know-how, as well as improvements in data quality and granularity, has enabled market participants to collect and analyse information on a greater scale. This has enabled AI to be used to identify price signals and to make more effective analysis of the vast amount of data available. According to the FSB,, “The key task is to identify signals from data on which predictions relating to price level or volatility can be made, over various time horizons, to generate higher and uncorrelated returns.”

[1] https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/quantitative-trading/
[2] Financial Stability Board, Artificial intelligence and machine learning in financial services: market developments and financial stability implications, 1 November 2017, www.fsb.org

The basic problem of this project is how to predict the price of a financial market. This leads to the calculation of a return, which is the difference between the price of a financial instrument over a time interval. The problem may be further defined as forecasting a return or the actual asset price over a time interval, or will the price go up or down. This inevitably leads to the development of a trading strategy, such as to buy or to sell a financial instrument if a model predicts a particular price movement, and to wait for the next buy or sell signal from the model to close out the trade or add to the position.

Conduct of the Project

The project will be conducted in three phases:

1. Perform a literature review of financial market prediction using artificial intelligence techniques.
2. Review the application of such methods by money managers.
3. Develop a state of the art artificial intelligence based prediction model for a financial market.

The initial phase of the project is to perform a literature search and review. This proceeds firstly by identifying academic research related to the problem and focuses on studies of AI methods applied to major financial markets such as US stock markets, which are high liquid i.e. financial instruments can be bought and sold on a daily basis in relatively large quantities without individual trades having a significant impact on the price.

The literature review references include:

- Saad, Prokhorov & Wunsch (1998)[3] compared three types of neural network - time delay, recurrent and probabilistic - and tested predictability and forecasting on a variety of highly liquid US stocks. They concluded “predicting short term stock trends based on history of daily closing prices is possible using any of the three different networks.”
- Guresen, Kayakutlu & Daim (2011)[4] compared also compared three types of neural network - mutli-layer perceptron, dynamic artificial neural network and hybrid neural network - and used daily rate values of the NASDAQ Stock Exchange index. They concluded “the multilayer perceptron gives the most reliable best results in forecasting time series”.
- Kamalov (2020)[5] constructed and tested three neural network models - multilayer perceptron, convolutional neural network and long short-term memory net - with stock price data for four major US public companies. He concludes “the results show

[3] Saad, E. W., Prokhorov, D. V., & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. IEEE Transactions on neural networks, 9(6), 1456-1470.
[4] Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389-10397.
[5] Kamalov, F. (2020). Forecasting significant stock price changes using neural networks. Neural Computing and Applications, 32(23), 17655-17667.

that neural networks models are capable of forecasting significant changes in asset price with high degree of accuracy."

The literature review with positive findings concerning the use of AI-based methods for predicting financial markets needs to be balanced with studies that claim financial markets cannot be predicted. For example, the random walk hypothesis is a financial theory stating that stock market prices evolve according to a random walk (so price changes are random).

Additional literature references include:

- Kendall (1953)[6] found "in [a] series of prices which are observed at fairly close intervals the random changes from one term to the next are so randomised as to swamp any systematic effect which may be present. The data behave almost like [a] wandering series".
- Fama (1965)[7] explains that the random walk theory casts serious doubt on many other methods for describing and predicting stock price behavior, but also conceded, "It is unlikely that the random walk hypothesis provides an exact description of the behavior of stock market prices. For practical purposes, however, the model may be acceptable even though it does not fit the facts exactly. Thus although successive price changes may not be strictly independent, the actual amount of dependence may be so small as to be unimportant."
- Lo & MacKinley (2001)[8] collate and present evidence that reject the random walk theory, and give support to the view that trends in the stock market prices can be predicted and exploiting them can be profitable.

I have also researched data science articles written by programmers building models for quantitative trading. These references are as follows:

- How (not) to use Machine Learning for time series forecasting: Avoiding the pitfalls, Vergards Florvik, 7 June 2018.[9]
- S&P 500 Stock Price Prediction Using Machine Learning and Deep Learning, Shiyan Boxer, 16 August 2019.[10]
- Machine learning for stock prediction. A quantitative approach, Gianluca Malato, 27 February 2020.[11]
- A Technical Guide on RNN/LSTM/GRU for Stock Price Prediction, Dr. Dataman, 6 December 2020.[12]
- Predicting stock prices with AI: which method maximises your profit?, Justin Tang, 18 August 2021.[13]

---

[6] Kendall, M. G., & Hill, A. B. (1953). The analysis of economic time-series-part i: Prices. Journal of the Royal Statistical Society. Series A (General), 116(1), 11-34.
[7] Fama, E. F. (1995). Random walks in stock market prices. Financial analysts journal, 51(1), 75-80.
[8] Lo, A. W., & MacKinlay, A. C. (2001). A non-random walk down Wall St. Princeton University Press.
[9] https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424
[10] https://medium.com/shiyan-boxer/s-p-500-stock-price-prediction-using-machine-learning-and-deep-learning-328b1839d1b6
[11] https://towardsdatascience.com/machine-learning-for-stock-prediction-a-quantitative-approach-4ca98c0bfb8c
[12] https://medium.com/swlh/a-technical-guide-on-rnn-lstm-gru-for-stock-price-prediction-bce2f7f30346
[13] https://jztang.medium.com/predicting-stock-prices-with-ai-which-method-should-you-use-to-profit-5cbc4c9a2cc1

The literature review is augmented by researching information about how money managers use AI-based prediction models to gain trading/investment performance in financial markets. I will focus this research on the world's largest hedge fund management firms (ranked by assets under management)[14]:

1. Blackrock Advisors - based in New York, USA
2. AQR Capital Management - based in Greenwich, Connecticut, USA
3. Bridgewater Associates - based in Westport, Connecticut, USA
4. Renaissance Technologies - based in New York, USA
5. Man Group - based in London, UK
6. Elliott Management - based in New York, USA
7. Two Sigma Investments - based in New York, USA
8. Millennium Management - based in New York, USA
9. Davidson Kempner Capital Management - based in New York, USA
10. Citadel Advisors - based in Chicago, USA

Gaining this information may present some difficulties, and I will need to find public sources of information such as reports on their websites, statements made to financial authorities, and in some instances books written by or about their founders or by their associates[15].

I do not aim to perform a comprehensive study of how these firms employ AI-based prediction models. Instead I intend to highlight some examples of how AI techniques have been or are being applied by top professional investors managing real money.

According to the FSB[16], firms are hesitant to share information they view as proprietary and critical to their competitive advantage. Furthermore, there is not always a standard definition of what AI methods are when firms share information. Also, whilst AI may be used in the investment decision-making process, an investment decision may be mainly made on a discretionary basis, as opposed to a systematic AI-driven basis.

Please see the design section below for information on the data and skills required for the model development phase of the project.

Statement of Deliverables

The deliverables of the project in summary are:

- A final presentation with slides
- A dissertation
- A copy of the software code

The final presentation will include the following contents:

---

[14] https://www.investopedia.com/articles/personal-finance/011515/worlds-top-10-hedge-fund-firms
[15] For example, Lipton, A. (2021). The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution: by Gregory Zuckerman, Portfolio/Penguin (2019).
[16] Financial Stability Board, Artificial intelligence and machine learning in financial services: market developments and financial stability implications, 1 November 2017, www.fsb.org

- The aims and objectives of the project
- A summary of the design
- A description of what was produced in the project
- Interesting aspects of the implementation
- An evaluation of what has been achieved
- Any shortfall from the original proposal
- Suggestions for future development
- A demonstration of the software

The written dissertation will be presented using the LaTeX publisher and include the following contents:

- Abstract - a summary of the project
- Introduction - a brief overview of the project
- Background - the research done to acquire the necessary information and skills to carry out the project
- Design - a full description of the design
- Realisation - a detailed explanation of how the design was implemented, including changes made to the design in the course of implementation and testing of the implementation
- Evaluation - a critical assessment of the strengths and weaknesses of the project
- Learning points - a rundown of the skills and knowledge acquired and improved by the project, including what actions were crucial to the outcome of the project and any things that should be done differently
- Professional issues - how the project relates to the BCS Code of Conduct and Code of Good Practice
- Conclusion - including summary, main findings and suggestions for further research and development
- Bibliography - a cited list of research papers and other materials consulted during the project and referred to in the dissertation
- Appendices - as necessary for completeness

Experiments will be carried out as detailed in the design section below, including adjusting the hyperparameters of models, experimenting with time lags in time series data, creating multivariate time series, and adding additional features to the price data.

Design

The model will consist of components, principally an artificial intelligence mode fed by financial market data followed by an evaluation of the model performance.

It is anticipated the model will be an artificial neural network (ANN) designed to handle time-series datasets. ANNs are biologically inspired analogues of the brain. They have several advantages, including:

- They can be designed to detect linear or non-linear patterns.

- They can learn from their environment and can be adapted through training to improve their performance.
- They allow for the application of mathematical techniques to enable learning and analysis.

The exact choice of architecture for the ANN will be influenced after reflecting on the literature review and conducting research of how money managers use AI prediction models to gain trading/investment performance in financial markets. The final design decision will be made following the completion of a critical examination of the alternatives.

The design of the model is likely to be a deep learning ANN, whose variations include recurrent neural network, convolutional neural network and multi-layer perceptron. For each of these types, there are also a multitude of design options, including number of layers, the number of neurons in each layer, and the use of activation functions. These ANNs have been studied in the first two semesters of MSc Data Science & Artificial Intelligence.

The data used by the model will be structured historical time-series data from a financial market. In short, time series data is a flow of data, which is collected during a specific time period. United States financial market data is available from the kibot website.[17] The data I intend to use is packaged as standard comma-delimited text files, with a separate file for every financial instrument. Each file has fields as follows:

- Date - recorded in the “mm/dd/yyyy” format.
- Time - US Eastern Time (ET) time zone in the “hh:mm” format
- Price - the opening price, the highest price, the lowest price and the closing price of the financial instrument for that day on the exchange. The plan is to feed the model input of a single price per trading day, although it should be possible to also create a multivariate time series using two or more of the given daily prices, and to add additional features to the price data.
- Volume - the total number of instruments e.g. shares or contracts traded on that day on the exchange. This may be left out and not input to the model.

An alternative data source of financial market data is Yahoo Finance.[18]

The data may need to be scaled in order to handle prices in the different time intervals and for different financial instruments. This may involve some pre-processing of data and the use of a software library such as sklearn. However, if returns (or percentage changes) are used data scaling may not be necessary as the data will already be normalised.

The data will need to be organised into appropriate time intervals. For example, chunks of 30 or 60 days, which will require experimentation, and it may also be useful to analyse moving averages of daily prices, as is commonly performed by traders employing technical analysis in financial markets.

[17] www.kibot.com
[18] https://uk.finance.yahoo.com

The model will have a training mode, where synaptic weights are adjusted to learn a task, and an online mode, where input signals are converted to output ones. The dataset will be divided into a training dataset and a test dataset. The training dataset may be further divided to create a validation dataset to fine tune the hyperparameters of the model. The hyperparameters may include the learning rate, the number of epochs, the choice of loss function as well as the components of the ANN structure, such as the number of layers, the number of neurons in each layer and the use of activation functions.

The software used to design and implement the ANN will be python, with the use of open source software libraries such as pandas, numpy, sklearn and PyTorch. This programming language and tools have been studied in the first two semesters of MSc Data Science & Artificial Intelligence. The presentation slides will be prepared on Microsoft Powerpoint and the final dissertation will be published using LaTeX using the overleaf editor.

Evaluation

To evaluate the AI model I may seek to minimise the error of the model. For example, Mean Square Error (MSE) is a measure of the fitness of a model. It is calculated by the sum of squares of the prediction error (i.e. actual output minus predicted output), divided by the number of data points. It gives a measure of how much the predicted results deviate from the actual number.

For a classification problem, I may maximise the accuracy of the model, or alternatively, minimise the binary cross-entropy / log loss.

The evaluation of the model will also need to appreciate autocorrelation, which in the case of time series data, is the likeliness of the price at time ‘t+1’ to be close to the price at time ‘t’.

In addition to experimenting, testing and evaluating the software, it is intended to test the predictive capability of the model with a live trading account (e.g. Interactive Brokers[19]). The live trading account will be used to test whether the model generalises well to future prices and with real money. This will be used to include a trading account with a profit or loss in the final dissertation, which will also account for transaction costs, such as brokerage fees, taxes, forex costs, and the bid-offer spread.

Risks

The risks of the project include:

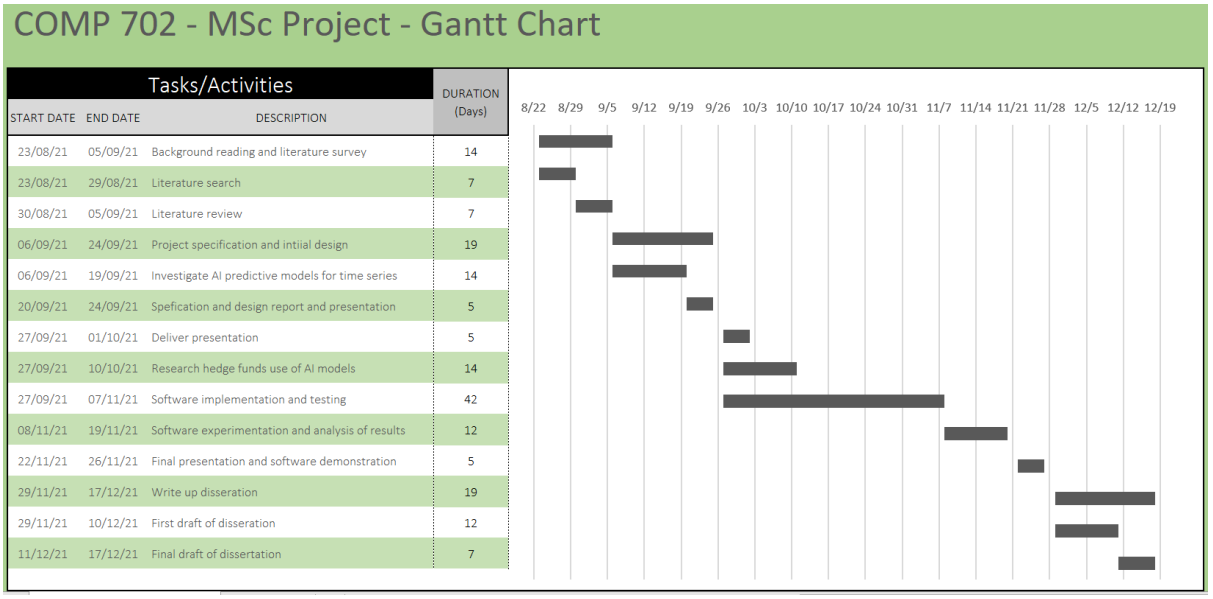
- Balancing the scope of the project with the resources (my learned and potential skills and abilities) and the time available. It is to be expected that forecasting or financial market data may become extremely complex, so finding a novel solution may not be a simple or straightforward task .

[19] https://www.interactivebrokers.co.uk

- A key issue anticipated is how to control the risk of a model overfitting. This may be addressed by applying methods to prevent overfitting, such as early stopping of training, dropout and regularisation. However, the real test will be whether the model generalises well to unseen data.
- Financial markets are very dynamic, and when market conditions change, the price signals generated by a model may decay in their usefulness. This will necessitate regular reviewing and testing of the model.
- An ANN model may risk creating a 'black box', whereby it is difficult to assess exactly why a decision has been made, and upon what factors a decision is based. This can be addressed in part by attempting credit assignment methods.

Plan

A Gannt chart for the project is shown above. This includes each of the specified deadlines and the scheduling of the main activities to be undertaken in the project.



References

Dawson, C. W. (2015). *Projects in computing and information systems: a student's guide*. Pearson Education.