# COMP5513

# Financial Computing

# Group Project Report

**Group 15**

**21000018G Li Yat Long**

**21009566G Mak Ka Wai**

**21009856G Chan Siu Fung**

**21014947G Yu Hin Chung Nikko**

## Table of Contents

*1. Paper Studies*......3

1.1 Li Yat Long's Paper – Stock Market's Price Movement Prediction with LSTM Neural Networks......3

1.2 Mak Ka Wai's Paper – Stock market analysis using candlestick regression and market trend prediction......5

1.3 Chan Siu Fung's Paper – Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM......7

1.4 Yu Hin Chung Nikko's Paper – Stock Closing Price Prediction Using Machine Learning Techniques..10

*2. Stock price prediction*......12

2.1 Feature selection......12

2.2 Machine Learning Model Training......14

2.3 Training Results......15

*3. Trading Strategies*......19

3.1 Trading Strategy 1......19

3.2 Trading Strategy 2......19

3.3 Trading Strategy 3......20

3.4 Trading Strategy 4......20

3.5 Trading Strategy 5......20

3.6 Results & Comparison......22

*References*......25

*Appendix*......26

Individual Report – 21000018G Li Yat Long......26

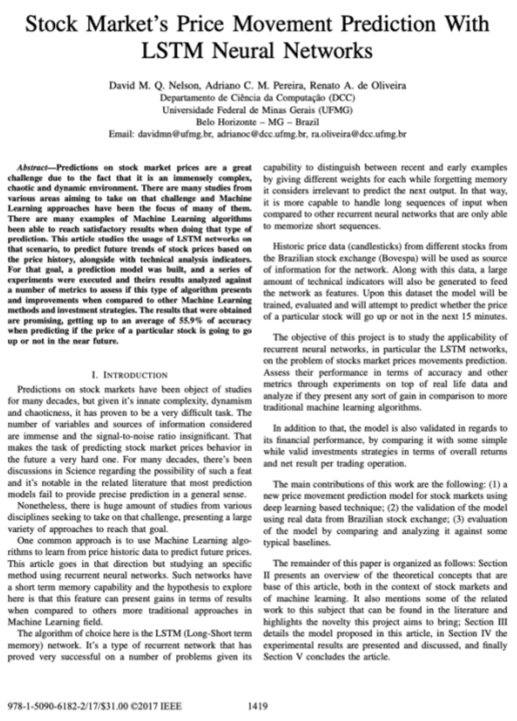
Individual Report – 21009566G Mak Ka Wai......29

Individual Report – 21009856G Chan Siu Fung......31

Individual Report – 21014947G Yu Hin Chung Nikko......34

# 1. Paper Studies

## 1.1 Li Yat Long's Paper – Stock Market's Price Movement Prediction with LSTM Neural Networks [1]



Stock Market's Price Movement Prediction With LSTM Neural Networks

David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira
Departamento de Ciência da Computação (DCC)
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brazil
Email: davidmn@ufmg.br, adrianoc@dcc.ufmg.br, ra.oliveira@dcc.ufmg.br

***Abstract*—Predictions on stock market prices are a great challenge due to the fact that it is an immensely complex, chaotic and dynamic environment. There are many studies from various areas aiming to take on that challenge and Machine Learning approaches have been the focus of many of them. There are many examples of Machine Learning algorithms been able to reach satisfactory results when doing that type of prediction. This article studies the usage of LSTM networks on that scenario, to predict future trends of stock prices based on the price history, alongside with technical analysis indicators. For that goal, a prediction model was built, and a series of experiments were executed and theirs results analyzed against a number of metrics to assess if this type of algorithm presents and improvements when compared to other Machine Learning methods and investment strategies. The results that were obtained are promising, getting up to an average of 55.9% of accuracy when predicting if the price of a particular stock is going to go up or not in the near future.**

I. INTRODUCTION

Predictions on stock markets have been object of studies for many decades, but given it's innate complexity, dynamism and chaoticness, it has proven to be a very difficult task. The number of variables and sources of information considered are immense and the signal-to-noise ratio insignificant. That makes the task of predicting stock market prices behavior in the future a very hard one. For many decades, there's been discussions in Science regarding the possibility of such a feat and it's notable in the related literature that most prediction models fail to provide precise prediction in a general sense.

Nonetheless, there is huge amount of studies from various disciplines seeking to take on that challenge, presenting a large variety of approaches to reach that goal.

One common approach is to use Machine Learning algorithms to learn from price historic data to predict future prices. This article goes in that direction but studying an specific method using recurrent neural networks. Such networks have a short term memory capability and the hypothesis to explore here is that this feature can present gains in terms of results when compared to others more traditional approaches in Machine Learning field.

The algorithm of choice here is the LSTM (Long-Short term memory) network. It's a type of recurrent network that has proved very successful on a number of problems given its capability to distinguish between recent and early examples by giving different weights for each while forgetting memory it considers irrelevant to predict the next output. In that way, it is more capable to handle long sequences of input when compared to other recurrent neural networks that are only able to memorize short sequences.

Historic price data (candlesticks) from different stocks from the Brazilian stock exchange (Bovespa) will be used as source of information for the network. Along with this data, a large amount of technical indicators will also be generated to feed the network as features. Upon this dataset the model will be trained, evaluated and will attempt to predict whether the price of a particular stock will go up or not in the next 15 minutes.

The objective of this project is to study the applicability of recurrent neural networks, in particular the LSTM networks, on the problem of stocks market prices movements prediction. Assess their performance in terms of accuracy and other metrics through experiments on top of real life data and analyze if they present any sort of gain in comparison to more traditional machine learning algorithms.

In addition to that, the model is also validated in regards to its financial performance, by comparing it with some simple while valid investments strategies in terms of overall returns and net result per trading operation.

The main contributions of this work are the following: (1) a new price movement prediction model for stock markets using deep learning based technique; (2) the validation of the model using real data from Brazilian stock exchange; (3) evaluation of the model by comparing and analyzing it against some typical baselines.

The remainder of this paper is organized as follows: Section II presents an overview of the theoretical concepts that are base of this article, both in the context of stock markets and of machine learning. It also mentions some of the related work to this subject that can be found in the literature and highlights the novelty this project aims to bring; Section III details the model proposed in this article, in Section IV the experimental results are presented and discussed, and finally Section V concludes the article.

978-1-5090-6182-2/17/$31.00 ©2017 IEEE 1419

Fig. 1 Screenshot of the paper

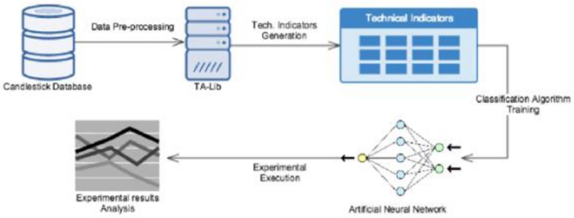

Data Pre-processing
Tech. Indicators Generation
Technical Indicators
Candlestick Database
TA-Lib
Classification Algorithm Training
Experimental Execution
Experimental results Analysis
Artificial Neural Network

Fig. 2 Methodology of the paper [1]

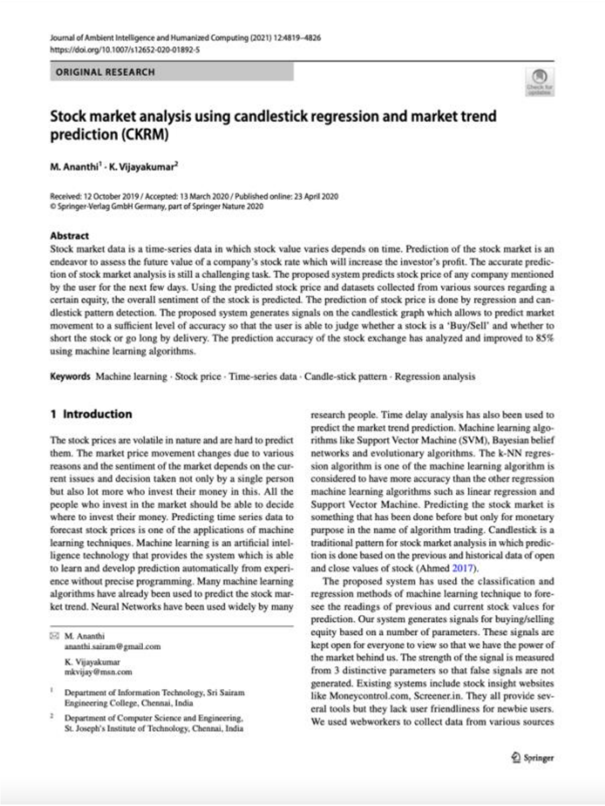
Studies on stock prices predictions methods have been investigated for decades due to its complexity and dynamism, where Machine Learning is one of the common approaches among them. The recurrent Long-Short term memory (LSTM) network which proves effective on Natural

Language Processing (NLP), was introduced to stock prices predictions in this paper. Compared to traditional feed-forwarded networks, LSTM allows neurons passing in multiple directions with weights given depending on the age of the data, forgetting memory when the data is evaluated as irrelevant and better performance for long data sequence input.

The historic price data from 2008 to 2015 were collected from the BM&F Bovespa stock exchange with the format of time series of candles including open, close, high, low and volume in a granularity of 15 minutes. After the data were collected, both log-return transformation and exponential smoothing were performed in order to stabilize the mean and variance of the time series, as well as reduce the random and noise on the price series. The data would be further transferred to the technical indicator using the TA-Lib library, generating 175 values which represent the future price, trading volume, movement tendency, visual graphical patterns of a stock in each period, by determining or predicting the characteristics of data. Finally, tuples of price data and technical indicators would be applied to train the neural network to predict if the closing price of the next period will be higher than the current closing price. The neural network would be trained at the end of each trading day to prepare for the prediction on the next day, by using the latest sets of data from the last 10 months till the current day with updated weights.

The LSTM model was compared to another three traditional methods including multi-Layer perception, Random Forest, and pseudo-random model with the same data input. The results showed that LSTM performed much better in terms of average accuracy, with 1 to 6 % better results. In addition, LSTM have achieved a high return ratio with significantly less risk compared to other strategies. However, the standard deviation was obviously larger in most cases compared to multi-Layer perception and Random Forest, which was the area the paper suggested to be improved.

## 1.2 Mak Ka Wai's Paper – Stock market analysis using candlestick regression and market trend prediction (CKRM) [2]



Journal of Ambient Intelligence and Humanized Computing (2021) 12:4819–4826
https://doi.org/10.1007/s12652-020-01892-5

ORIGINAL RESEARCH

Stock market analysis using candlestick regression and market trend prediction (CKRM)

M. Ananthi[1] · K. Vijayakumar[2]

Received: 12 October 2019 / Accepted: 13 March 2020 / Published online: 23 April 2020

Abstract

Stock market data is a time-series data in which stock value varies depends on time. Prediction of the stock market is an endeavor to assess the future value of a company's stock rate which will increase the investor's profit. The accurate prediction of stock market analysis is still a challenging task. The proposed system predicts stock price of any company mentioned by the user for the next few days. Using the predicted stock price and datasets collected from various sources regarding a certain equity, the overall sentiment of the stock is predicted. The prediction of stock price is done by regression and candlestick pattern detection. The proposed system generates signals on the candlestick graph which allows to predict market movement to a sufficient level of accuracy so that the user is able to judge whether a stock is a 'Buy/Sell' and whether to short the stock or go long by delivery. The prediction accuracy of the stock exchange has analyzed and improved to 85% using machine learning algorithms.

Keywords Machine learning · Stock price · Time-series data · Candle-stick pattern · Regression analysis

1 Introduction

The stock prices are volatile in nature and are hard to predict them. The market price movement changes due to various reasons and the sentiment of the market depends on the current issues and decision taken not only by a single person but also lot more who invest their money in this. All the people who invest in the market should be able to decide where to invest their money. Predicting time series data to forecast stock prices is one of the applications of machine learning techniques. Machine learning is an artificial intelligence technology that provides the system which is able to learn and develop prediction automatically from experience without precise programming. Many machine learning algorithms have already been used to predict the stock market trend. Neural Networks have been used widely by many research people. Time delay analysis has also been used to predict the market trend prediction. Machine learning algorithms like Support Vector Machine (SVM), Bayesian belief networks and evolutionary algorithms. The k-NN regression algorithm is one of the machine learning algorithm is considered to have more accuracy than the other regression machine learning algorithms such as linear regression and Support Vector Machine. Predicting the stock market is something that has been done before but only for monetary purpose in the name of algorithm trading. Candlestick is a traditional pattern for stock market analysis in which prediction is done based on the previous and historical data of open and close values of stock (Ahmed 2017).

The proposed system has used the classification and regression methods of machine learning technique to foresee the readings of previous and current stock values for prediction. Our system generates signals for buying/selling equity based on a number of parameters. These signals are kept open for everyone to view so that we have the power of the market behind us. The strength of the signal is measured from 3 distinctive parameters so that false signals are not generated. Existing systems include stock insight websites like Moneycontrol.com, Screener.in. They all provide several tools but they lack user friendliness for newbie users. We used webworkers to collect data from various sources

✉ M. Ananthi
ananthi.sairam@gmail.com

K. Vijayakumar
mkvijay@msn.com

[1] Department of Information Technology, Sri Sairam Engineering College, Chennai, India

[2] Department of Computer Science and Engineering, St. Joseph's Institute of Technology, Chennai, India

Fig. 3 Screenshot of the paper

In this paper, it proposed a stock prediction system which hast three main modules. The first module is trendline identification consisting of MACD and candlestick graph to indicate buy and sell signal. The second module is using K Nearest Neighbor algorithm to predict the stock prices. Features include MACD, RSI, Bollinger bands and candle stick pattern detection. The last module is pattern formation such as formed by Open, High, Low, Close (OHLC) and RSI over past few days to predict the trend of next day. Also, the proposed system is implemented Django web framework and dataset is using Flask REST API which is getting live data in an open-source platform called Quandl.

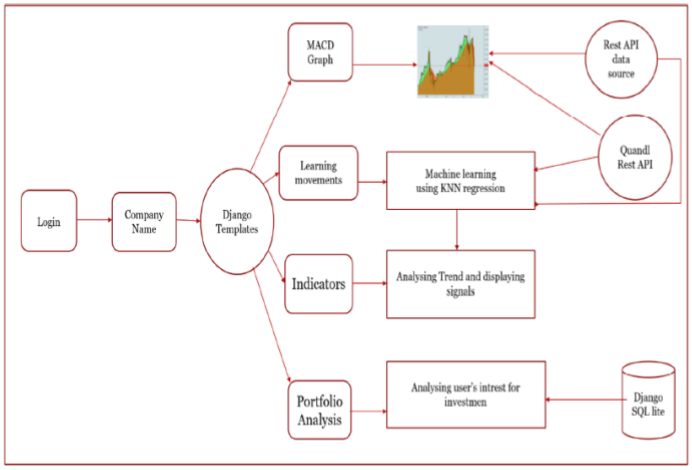
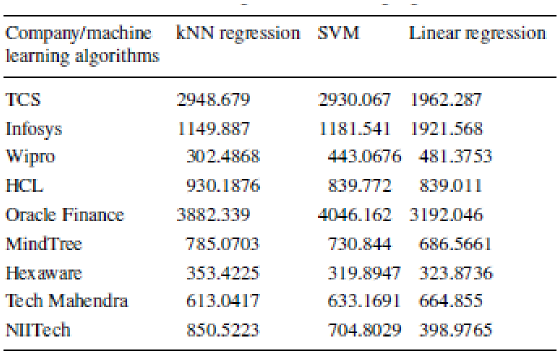


Fig. 4 Proposed system of the paper [2]

Moreover, this paper has compared the prediction accuracy from three common algorithms which are KNN regression, Support Vector Machine (SVM) and Linear regression to predict the stock prices of different companies (TCS, HCL, Oracle Finance, MindTree, NIITech, WiPro). Features include MACD, RSI, Bollinger bands, candle stick pattern detection. Moreover, this paper has used exponential moving average (EMA) and relative strength index for features and variation to have higher prediction for stocks.



| Company/machine learning algorithms | kNN regression | SVM | Linear regression |
|---|---|---|---|
| TCS | 2948.679 | 2930.067 | 1962.287 |
| Infosys | 1149.887 | 1181.541 | 1921.568 |
| Wipro | 302.4868 | 443.0676 | 481.3753 |
| HCL | 930.1876 | 839.772 | 839.011 |
| Oracle Finance | 3882.339 | 4046.162 | 3192.046 |
| MindTree | 785.0703 | 730.844 | 686.5661 |
| Hexaware | 353.4225 | 319.8947 | 323.8736 |
| Tech Mahendra | 613.0417 | 633.1691 | 664.855 |
| NIITech | 850.5223 | 704.8029 | 398.9765 |

Fig. 5 Prediction value of selected companies [2]

Below is the prediction accuracy from the paper. It points out that KNN regression has the best performance.  KNN has 90%- 96% accuracy compared to Linear regression (80% - 95%) and Support vector machine (60% - 80%).

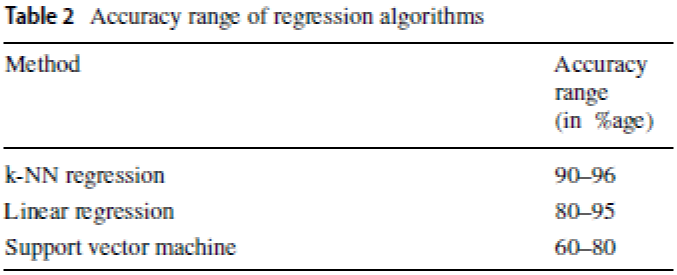
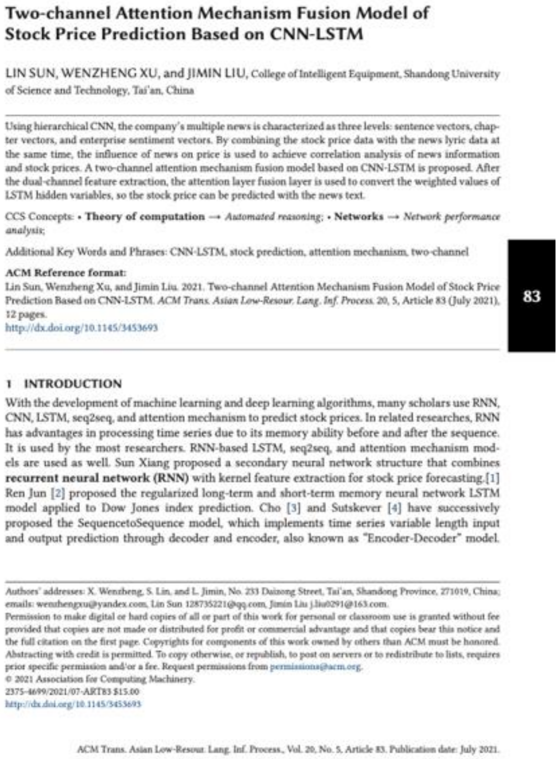


**Table 2** Accuracy range of regression algorithms

| Method | Accuracy range (in %age) |
|---|---|
| k-NN regression | 90–96 |
| Linear regression | 80–95 |
| Support vector machine | 60–80 |

Fig. 6 Finding of the paper

### 1.3 Chan Siu Fung’s Paper – Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM [3]



**Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM**

LIN SUN, WENZHENG XU, and JIMIN LIU, College of Intelligent Equipment, Shandong University of Science and Technology, Tai'an, China

Using hierarchical CNN, the company's multiple news is characterized as three levels: sentence vectors, chapter vectors, and enterprise sentiment vectors. By combining the stock price data with the news lyric data at the same time, the influence of news on price is used to achieve correlation analysis of news information and stock prices. A two-channel attention mechanism fusion model based on CNN-LSTM is proposed. After the dual-channel feature extraction, the attention layer fusion layer is used to convert the weighted values of LSTM hidden variables, so the stock price can be predicted with the news text.

CCS Concepts: • **Theory of computation** → *Automated reasoning*; • **Networks** → *Network performance analysis*;

Additional Key Words and Phrases: CNN-LSTM, stock prediction, attention mechanism, two-channel

**ACM Reference format:**
Lin Sun, Wenzheng Xu, and Jimin Liu. 2021. Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20, 5, Article 83 (July 2021), 12 pages.
http://dx.doi.org/10.1145/3453693

83

**1 INTRODUCTION**

With the development of machine learning and deep learning algorithms, many scholars use RNN, CNN, LSTM, seq2seq, and attention mechanism to predict stock prices. In related researches, RNN has advantages in processing time series due to its memory ability before and after the sequence. It is used by the most researchers. RNN-based LSTM, seq2seq, and attention mechanism models are used as well. Sun Xiang proposed a secondary neural network structure that combines **recurrent neural network (RNN)** with kernel feature extraction for stock price forecasting.[1] Ren Jun [2] proposed the regularized long-term and short-term memory neural network LSTM model applied to Dow Jones index prediction. Cho [3] and Sutskever [4] have successively proposed the SequencetoSequence model, which implements time series variable length input and output prediction through decoder and encoder, also known as "Encoder-Decoder" model.

Authors' addresses: X. Wenzheng, S. Lin, and L. Jimin, No. 233 Daizong Street, Tai'an, Shandong Province, 271019, China; emails: wenzhengxu@yandex.com, Lin Sun 128735221@qq.com, Jimin Liu j.liu0291@163.com.

2375-4699/2021/07-ART83 $15.00
http://dx.doi.org/10.1145/3453693

ACM Trans. Asian Low-Resour. Lang. Inf. Process., Vol. 20, No. 5, Article 83. Publication date: July 2021.

Fig. 7 Screenshot of the paper

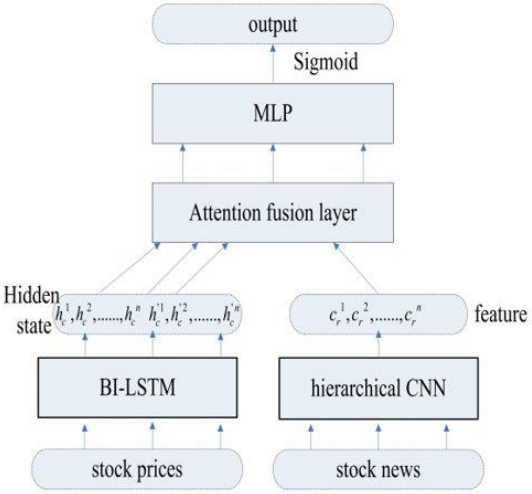


Fig. 8 Two-channel attention mechanism fusion CNN-LSTM Model [3]

Characteristics of proposed model

A dual channel attention mechanism fusion model combining both stock news and stock prices was introduced in this paper for stock price prediction.

In the model, the stock data and news corpus are processed separately in two different channels before the attention fusion layer that can avoid mutual interference between the stock price features and stock news features. A bidirectional LSTM model is used to process the stock prices while a CNN model is used to extract data from stock news.

To capture the semantic information related to the rise and fall of stocks from the long news items more accurately, the news is combined and analyzed together. News is segmented and words are stopped and embedded to vectors in embedding layer, the vectors are further compressed in three layers into different levels namely sentence, chapter, enterprise levels in the CNN model.

The fusion layer combines the features from both channels and then passes the output to MLP model to obtain final output. Through the fusion layer, a link between the corporate news and the stock price value can be established.

The feature and information fusion focusing on the derived attributes and data enhances the intrinsic perspective in stock market prediction. Meanwhile, the model fusion combines both the ability of CNN to express text space features and the ability of Bi-LSTM to express time dimension through attention mechanism, so this dual-channel fusion deep learning model has the ability to express time and text space of stock price and stock news data simultaneously.

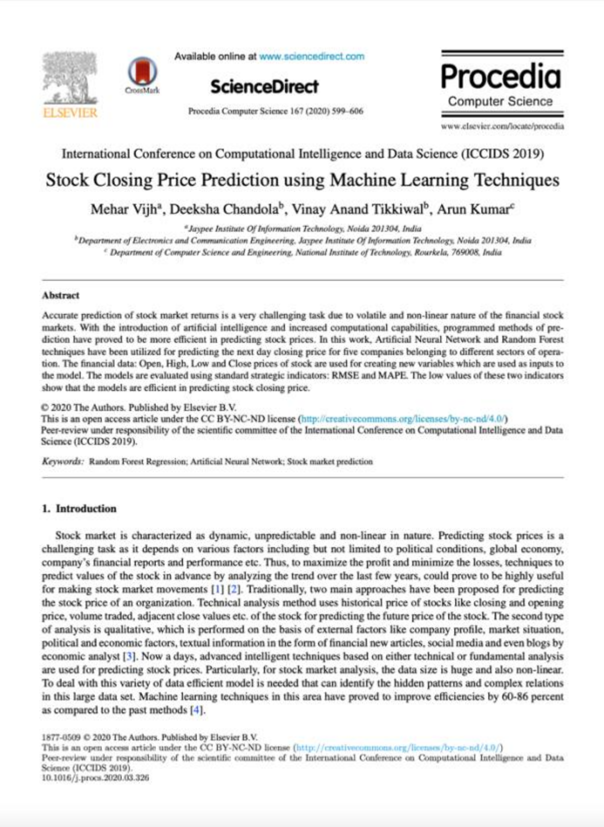
Comparison of models

The training set loss value and test set loss value of the model were compared with those of purely Bi-LSTM model and single-channel CNN-LSTM in terms of the average value of MSE, RSME and MAE of each stock. The stock data was derived from the closing price of 411 stocks in the period of January 2016 to December 2017 obtained by the tushare interface. The news data were extracted from over 23000 news items related to those stocks within that period from dozens of financial websites like Securities Times, China Securities Network, Sina Finance, and China Economic Net. The result showed that the proposed model has the least loss among three. This implies that the proposed model can improve the accuracy of stock price prediction.

Return from trading based on stock price prediction by the model

20 stocks were randomly selected and being forecasted with the model, trading was done by throwing at peak, purchasing at bottom, the average income of selected stocks was found. The total stock return is 290%, the annual compound yield is as high as 126.5%, the positive income ratio is 64% which far exceeds the benchmark CSI 300 index yield. This showed that the predictions based on this model can achieve sustained positive returns with good predictive results.

## 1.4 Yu Hin Chung Nikko’s Paper – Stock Closing Price Prediction Using Machine Learning Techniques [4]



Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 167 (2020) 599–606
Procedia Computer Science
www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

**Stock Closing Price Prediction using Machine Learning Techniques**

Mehar Vijh[a], Deeksha Chandola[b], Vinay Anand Tikkiwal[b], Arun Kumar[c]

[a] Jaypee Institute Of Information Technology, Noida 201304, India
[b] Department of Electronics and Communication Engineering, Jaypee Institute Of Information Technology, Noida 201304, India
[c] Department of Computer Science and Engineering, National Institute of Technology, Rourkela, 769008, India

**Abstract**

Accurate prediction of stock market returns is a very challenging task due to volatile and non-linear nature of the financial stock markets. With the introduction of artificial intelligence and increased computational capabilities, programmed methods of prediction have proved to be more efficient in predicting stock prices. In this work, Artificial Neural Network and Random Forest techniques have been utilized for predicting the next day closing price for five companies belonging to different sectors of operation. The financial data: Open, High, Low and Close prices of stock are used for creating new variables which are used as inputs to the model. The models are evaluated using standard strategic indicators: RMSE and MAPE. The low values of these two indicators show that the models are efficient in predicting stock closing price.

*Keywords:* Random Forest Regression; Artificial Neural Network; Stock market prediction

**1. Introduction**

Stock market is characterized as dynamic, unpredictable and non-linear in nature. Predicting stock prices is a challenging task as it depends on various factors including but not limited to political conditions, global economy, company’s financial reports and performance etc. Thus, to maximize the profit and minimize the losses, techniques to predict values of the stock in advance by analyzing the trend over the last few years, could prove to be highly useful for making stock market movements [1] [2]. Traditionally, two main approaches have been proposed for predicting the stock price of an organization. Technical analysis method uses historical price of stocks like closing and opening price, volume traded, adjacent close values etc. of the stock for predicting the future price of the stock. The second type of analysis is qualitative, which is performed on the basis of external factors like company profile, market situation, political and economic factors, textual information in the form of financial new articles, social media and even blogs by economic analyst [3]. Now a days, advanced intelligent techniques based on either technical or fundamental analysis are used for predicting stock prices. Particularly, for stock market analysis, the data size is huge and also non-linear. To deal with this variety of data efficient model is needed that can identify the hidden patterns and complex relations in this large data set. Machine learning techniques in this area have proved to improve efficiencies by 60-86 percent as compared to the past methods [4].

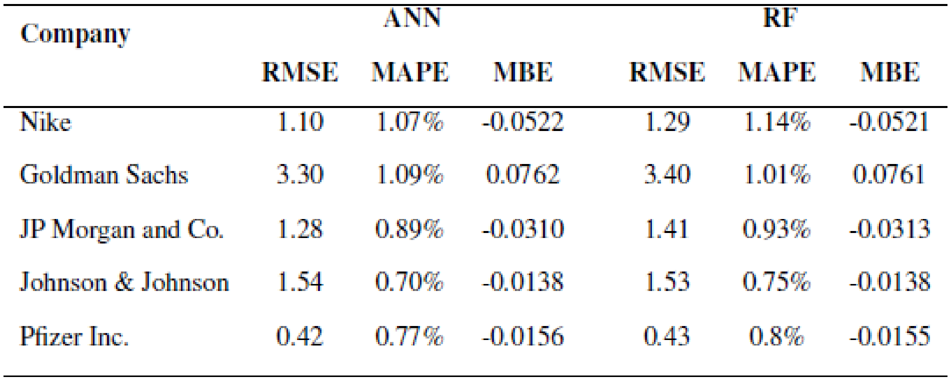
10.1016/j.procs.2020.03.326

Fig. 9 Screenshot of the paper

This paper adopts two machine learning methods to predict the stock price, namely Random Forest (RF) and artificial neural networks (ANN). There are several parameters they had made to make predictions, including:

- Stock High Price – Stock Low Price (H-L)
- Stock Close Price – Open Price (O-C)
- 7,14,21 Moving Average
- 7 days standard deviation

This paper uses stocks from companies like Nike, Goldman Sachs and JP Morgan for training and testing. As for evaluating the model performance, the paper uses Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Bias Error (MBE).



| Company | | ANN | | | RF | |
|---|---|---|---|---|---|---|
| | RMSE | MAPE | MBE | RMSE | MAPE | MBE |
| Nike | 1.10 | 1.07% | -0.0522 | 1.29 | 1.14% | -0.0521 |
| Goldman Sachs | 3.30 | 1.09% | 0.0762 | 3.40 | 1.01% | 0.0761 |
| JP Morgan and Co. | 1.28 | 0.89% | -0.0310 | 1.41 | 0.93% | -0.0313 |
| Johnson & Johnson | 1.54 | 0.70% | -0.0138 | 1.53 | 0.75% | -0.0138 |
| Pfizer Inc. | 0.42 | 0.77% | -0.0156 | 0.43 | 0.8% | -0.0155 |

Fig. 10 Paper results. [4]

The results shown that ANN generally behaves better than RF. For instance, using the Pfizer dataset, ANN has a score of 0.42 (RMSE) ,0.77 (MAPE) and –0.0156 (MBE), while RF has 0.43 (RMSE), 0.8 (MAPE) and –0.0155 (MBE).

This paper is part of our inspiration to develop this model as the orange software include ANN and RF models, as well as some of the parameters such as moving averages and the standard deviation.

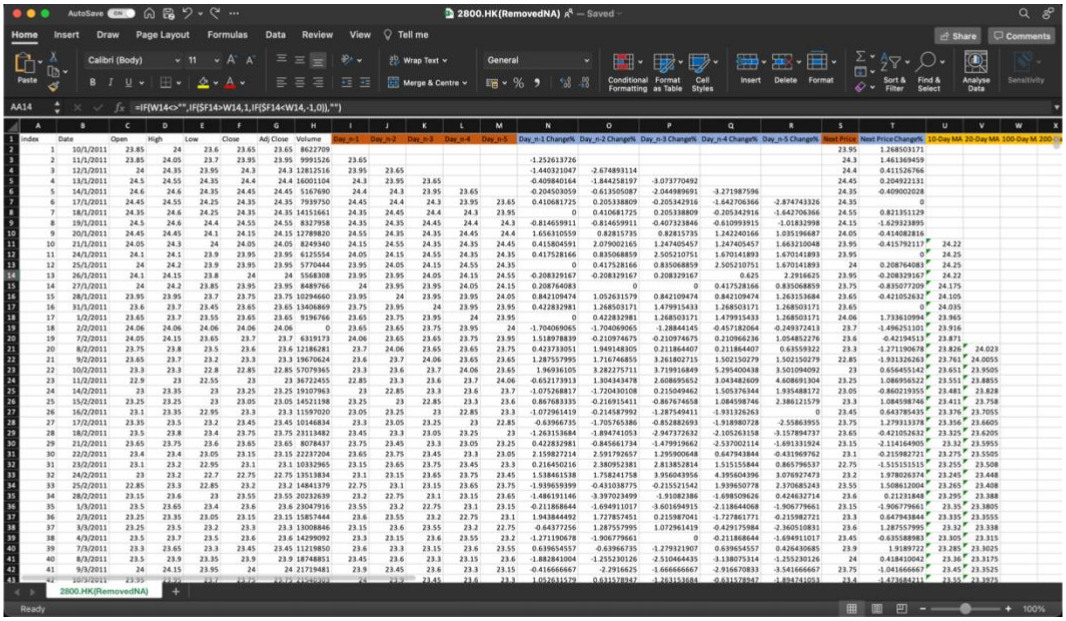
## 2. Stock price prediction

### 2.1 Feature selection

The 2800 stock price data from 2011 to 2021 were downloaded from the internet and features were selected and calculated from the raw data, these include:

1. Close price
2. Close price 1 to 5 days before the current date
3. Percentage change of price 1 to 5 days ago compared to today's price
4. 10-, 20-, 100- & 200-day moving average (MA) which calculates the average price of recent 10/ 20/ 100/ 200 days
5. Uptrend indicated when the closing price is higher than the 10-, 20-, 100- & 200-day moving averages or downtrend indicated when the closing price is lower than the 10-, 20-, 100- & 200-day moving averages
6. Buy signal indicated when the closing price moves over 10-, 20-, 100- & 200-day moving averages or sell signal when the closing price moves below 10-, 20-, 100- & 200-day moving averages
7. Buy signal indicated when the closing price above the moving averages rebounds upwards after touching the moving averages or sell signal indicated when the closing price below the moving averages rebounds downwards after touching the moving averages
8. Gain/ Loss in closing price compared to last trading day
9. Average Gain/ Loss of 14 days
10. Relative Strength – Average gain divided by average loss
11. Relative Strength Index
12. Buy signal indicated when the RSI leaves the oversold region or sell signal indicated when the RSI leaves the overbought region
13. 12- & 26-day exponential moving average (EMA)
14. MACD – Difference between 12-day EMA and 26-day EMA
15. Signal – 9-day EMA of MACD line
16. Histogram – MACD value minus Signal value
17. Simplified signal line crossover – Buy signal indicated when MACD line just goes over the signal line or sell signal indicated when MACD line just goes below the signal line
18. Lower/ Upper Bollinger band

19. Simplified mean reversion – Buy signal indicated when closing price below the lower Bollinger Band goes over it or sell signal indicated when closing price over the upper Bollinger Band goes below it

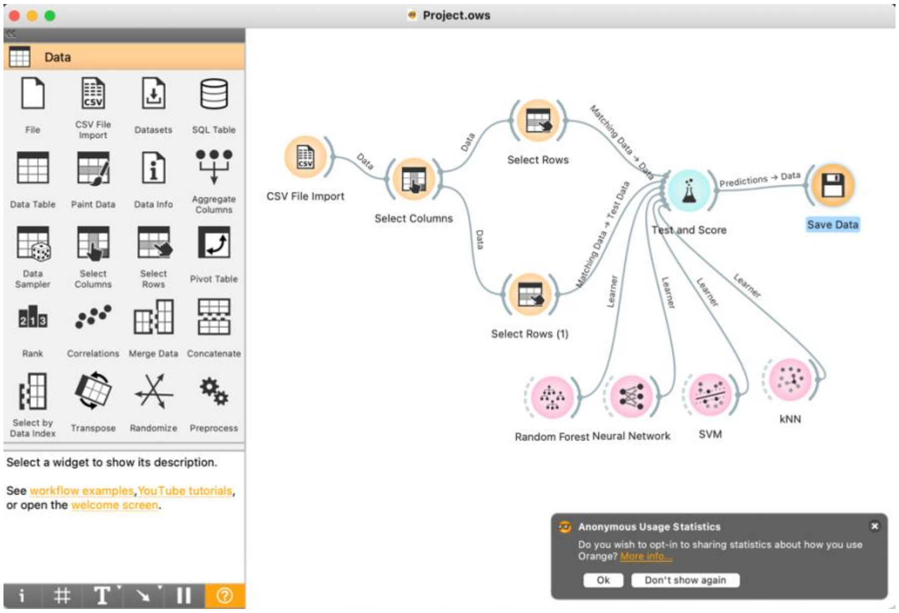


AA14 =IF(W14<>"",IF($F14>W14,1,IF($F14<W14,-1,0)),"")

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Index | Date | Open | High | Low | Close | Adj Close | Volume | Day_n-1 | Day_n-2 | Day_n-3 | Day_n-4 | Day_n-5 | Day_n-1 Change% | Day_n-2 Change% | Day_n-3 Change% | Day_n-4 Change% | Day_n-5 Change% | Next Price | Next Price Change% | 10-Day MA | 20-Day MA | 100-Day M | 200-Da |
| 2 | 1 | 10/1/2011 | 23.85 | 24 | 23.6 | 23.65 | 23.65 | 8622709 | | | | | | | | | | | 23.95 | 1.268503171 | | | | |
| 3 | 2 | 11/1/2011 | 23.85 | 24.05 | 23.7 | 23.95 | 23.95 | 9991526 | 23.65 | | | | | -1.252613726 | | | | | 24.3 | 1.461369459 | | | | |
| 4 | 3 | 12/1/2011 | 24 | 24.35 | 23.95 | 24.3 | 24.3 | 12812516 | 23.95 | 23.65 | | | | -1.440321047 | -2.674893114 | | | | 24.4 | 0.411526766 | | | | |
| 5 | 4 | 13/1/2011 | 24.5 | 24.55 | 24.35 | 24.4 | 24.4 | 16001104 | 24.3 | 23.95 | 23.65 | | | -0.409840164 | -1.844258197 | -3.073770492 | | | 24.45 | 0.204922131 | | | | |
| 6 | 5 | 14/1/2011 | 24.6 | 24.6 | 24.35 | 24.45 | 24.45 | 5167690 | 24.4 | 24.3 | 23.95 | 23.65 | | -0.204503059 | -0.613505087 | -2.044989691 | -3.271987596 | | 24.35 | -0.409002028 | | | | |
| 7 | 6 | 17/1/2011 | 24.45 | 24.55 | 24.25 | 24.35 | 24.35 | 7939750 | 24.45 | 24.4 | 24.3 | 23.95 | 23.65 | 0.410681725 | 0.205338809 | -0.205342916 | -1.642706366 | -2.874743326 | 24.35 | 0 | | | | |
| 8 | 7 | 18/1/2011 | 24.35 | 24.6 | 24.25 | 24.35 | 24.35 | 14151661 | 24.35 | 24.45 | 24.4 | 24.3 | 23.95 | 0 | 0.410681725 | 0.205338809 | -0.205342916 | -1.642706366 | 24.55 | 0.821351129 | | | | |
| 9 | 8 | 19/1/2011 | 24.5 | 24.6 | 24.4 | 24.55 | 24.55 | 8327958 | 24.35 | 24.35 | 24.45 | 24.4 | 24.3 | -0.814659911 | -0.814659911 | -0.407323846 | -0.610993915 | -1.01832998 | 24.15 | -1.629323895 | | | | |
| 10 | 9 | 20/1/2011 | 24.45 | 24.45 | 24.1 | 24.15 | 24.15 | 12789820 | 24.55 | 24.35 | 24.35 | 24.45 | 24.4 | 1.656310559 | 0.82815735 | 0.82815735 | 1.242240166 | 1.035196687 | 24.05 | -0.414082816 | | | | |
| 11 | 10 | 21/1/2011 | 24.05 | 24.3 | 24 | 24.05 | 24.05 | 8249340 | 24.15 | 24.55 | 24.35 | 24.35 | 24.45 | 0.415804591 | 2.079002165 | 1.247405457 | 1.247405457 | 1.663210048 | 23.95 | -0.415792117 | 24.22 | | | |
| 12 | 11 | 24/1/2011 | 24.1 | 24.1 | 23.9 | 23.95 | 23.95 | 6125554 | 24.05 | 24.15 | 24.55 | 24.35 | 24.35 | 0.417528166 | 0.835068859 | 2.505210751 | 1.670141893 | 1.670141893 | 23.95 | 0 | 24.25 | | | |
| 13 | 12 | 25/1/2011 | 24 | 24.2 | 23.9 | 23.95 | 23.95 | 5770444 | 23.95 | 24.05 | 24.15 | 24.55 | 24.35 | 0 | 0.417528166 | 0.835068859 | 2.505210751 | 1.670141893 | 24 | 0.208764083 | 24.25 | | | |
| 14 | 13 | 26/1/2011 | 24.1 | 24.15 | 23.8 | 24 | 24 | 5568308 | 23.95 | 23.95 | 24.05 | 24.15 | 24.55 | -0.208329167 | -0.208329167 | 0.208329167 | 0.625 | 2.2916625 | 23.95 | -0.208329167 | 24.22 | | | |
| 15 | 14 | 27/1/2011 | 24 | 24.2 | 23.85 | 23.95 | 23.95 | 8489766 | 24 | 23.95 | 23.95 | 24.05 | 24.15 | 0.208764083 | 0 | 0 | 0.417528166 | 0.835068859 | 23.75 | -0.835077209 | 24.175 | | | |
| 16 | 15 | 28/1/2011 | 23.95 | 23.95 | 23.7 | 23.75 | 23.75 | 10294660 | 23.95 | 24 | 23.95 | 23.95 | 24.05 | 0.842109474 | 1.052631579 | 0.842109474 | 0.842109474 | 1.263153684 | 23.65 | -0.421052632 | 24.105 | | | |
| 17 | 16 | 31/1/2011 | 23.6 | 23.7 | 23.45 | 23.65 | 23.65 | 13406869 | 23.75 | 23.95 | 24 | 23.95 | 23.95 | 0.422832981 | 1.268503171 | 1.479915433 | 1.268503171 | 1.268503171 | 23.65 | 0 | 24.035 | | | |
| 18 | 17 | 1/2/2011 | 23.65 | 23.7 | 23.55 | 23.65 | 23.65 | 9196766 | 23.65 | 23.75 | 23.95 | 24 | 23.95 | 0 | 0.422832981 | 1.268503171 | 1.479915433 | 1.268503171 | 24.06 | 1.733610994 | 23.965 | | | |
| 19 | 18 | 2/2/2011 | 24.06 | 24.06 | 24.06 | 24.06 | 24.06 | 0 | 23.65 | 23.65 | 23.75 | 23.95 | 24 | -1.704069065 | -1.704069065 | -1.28844145 | -0.457182064 | -0.249372413 | 23.7 | -1.496251101 | 23.916 | | | |
| 20 | 19 | 7/2/2011 | 24.05 | 24.15 | 23.65 | 23.7 | 23.7 | 6319173 | 24.06 | 23.65 | 23.65 | 23.75 | 23.95 | 1.518978839 | -0.210974675 | -0.210974675 | 0.210966236 | 1.054852276 | 23.6 | -0.42194513 | 23.871 | | | |
| 21 | 20 | 8/2/2011 | 23.75 | 23.8 | 23.5 | 23.6 | 23.6 | 12186281 | 23.7 | 24.06 | 23.65 | 23.65 | 23.75 | 0.423733051 | 1.949148305 | 0.211864407 | 0.211864407 | 0.63559322 | 23.3 | -1.271190678 | 23.826 | 24.023 | | |
| 22 | 21 | 9/2/2011 | 23.65 | 23.7 | 23.2 | 23.3 | 23.3 | 19670624 | 23.6 | 23.7 | 24.06 | 23.65 | 23.65 | 1.287557995 | 1.716746855 | 3.261802715 | 1.502150279 | 1.502150279 | 22.85 | -1.931326263 | 23.761 | 24.0055 | | |
| 23 | 22 | 10/2/2011 | 23.3 | 23.3 | 22.8 | 22.85 | 22.85 | 57079365 | 23.3 | 23.6 | 23.7 | 24.06 | 23.65 | 1.96936105 | 3.282275711 | 3.719916849 | 5.295400438 | 3.501094092 | 23 | 0.656455142 | 23.651 | 23.9505 | | |
| 24 | 23 | 11/2/2011 | 22.9 | 23 | 22.55 | 23 | 23 | 36722455 | 22.85 | 23.3 | 23.6 | 23.7 | 24.06 | -0.652173913 | 1.304343478 | 2.608695652 | 3.043482609 | 4.608691304 | 23.25 | 1.086956522 | 23.551 | 23.8855 | | |
| 25 | 24 | 14/2/2011 | 23 | 23.35 | 23 | 23.25 | 23.25 | 19107963 | 23 | 22.85 | 23.3 | 23.6 | 23.7 | -1.075268817 | -1.720430108 | 0.215049462 | 1.505376344 | 1.935488172 | 23.05 | -0.860219355 | 23.481 | 23.828 | | |
| 26 | 25 | 15/2/2011 | 23.25 | 23.25 | 23 | 23.05 | 23.05 | 14521198 | 23.25 | 23 | 22.85 | 23.3 | 23.6 | 0.867683335 | -0.216915411 | -0.867674658 | 1.084598746 | 2.386121579 | 23.3 | 1.084598746 | 23.411 | 23.758 | | |
| 27 | 26 | 16/2/2011 | 23.1 | 23.35 | 22.95 | 23.3 | 23.3 | 11597020 | 23.05 | 23.25 | 23 | 22.85 | 23.3 | -1.072961419 | -0.214587992 | -1.287549411 | -1.931326263 | 0 | 23.45 | 0.643785435 | 23.376 | 23.7055 | | |
| 28 | 27 | 17/2/2011 | 23.35 | 23.5 | 23.2 | 23.45 | 23.45 | 10146834 | 23.3 | 23.05 | 23.25 | 23 | 22.85 | -0.63966735 | -1.705765386 | -0.852882693 | -1.918980728 | -2.55863955 | 23.75 | 1.279313378 | 23.356 | 23.6605 | | |
| 29 | 28 | 18/2/2011 | 23.5 | 23.8 | 23.4 | 23.75 | 23.75 | 23113482 | 23.45 | 23.3 | 23.05 | 23.25 | 23 | -1.263153684 | -1.894741053 | -2.947372632 | -2.105263158 | -3.157894737 | 23.65 | -0.421052632 | 23.325 | 23.6205 | | |
| 30 | 29 | 21/2/2011 | 23.65 | 23.75 | 23.6 | 23.65 | 23.65 | 8078437 | 23.75 | 23.45 | 23.3 | 23.05 | 23.25 | 0.422832981 | -0.845661734 | -1.479919662 | -2.537002114 | -1.691331924 | 23.15 | -2.114164905 | 23.32 | 23.5955 | | |
| 31 | 30 | 22/2/2011 | 23.4 | 23.4 | 23.05 | 23.15 | 23.15 | 22237204 | 23.65 | 23.75 | 23.45 | 23.3 | 23.05 | 2.159827214 | 2.591792657 | 1.295900648 | 0.647943844 | -0.431969762 | 23.1 | -0.215982721 | 23.275 | 23.5505 | | |
| 32 | 31 | 23/2/2011 | 23.1 | 23.2 | 22.95 | 23.1 | 23.1 | 10332965 | 23.15 | 23.65 | 23.75 | 23.45 | 23.3 | 0.216450216 | 2.380952381 | 2.813852814 | 1.515155844 | 0.865796537 | 22.75 | -1.515151515 | 23.255 | 23.508 | | |
| 33 | 32 | 24/2/2011 | 23 | 23.2 | 22.7 | 22.75 | 22.75 | 13513834 | 23.1 | 23.15 | 23.65 | 23.75 | 23.45 | 1.538461538 | 1.758241758 | 3.956043956 | 4.395604396 | 3.076927473 | 23.2 | 1.978026374 | 23.245 | 23.448 | | |
| 34 | 33 | 25/2/2011 | 22.85 | 23.3 | 22.85 | 23.2 | 23.2 | 14841379 | 22.75 | 23.1 | 23.15 | 23.65 | 23.75 | -1.939659399 | -0.431038775 | -0.215521542 | 1.939650778 | 2.370685243 | 23.55 | 1.508612004 | 23.265 | 23.408 | | |
| 35 | 34 | 28/2/2011 | 23.15 | 23.6 | 23 | 23.55 | 23.55 | 20232639 | 23.2 | 22.75 | 23.1 | 23.15 | 23.65 | -1.486191146 | -3.397023499 | -1.91082386 | -1.698509626 | 0.424632714 | 23.6 | 0.21231848 | 23.295 | 23.388 | | |
| 36 | 35 | 1/3/2011 | 23.5 | 23.65 | 23.4 | 23.6 | 23.6 | 23047916 | 23.55 | 23.2 | 22.75 | 23.1 | 23.15 | -0.211868644 | -1.694911017 | -3.601694915 | -2.118644068 | -1.906779661 | 23.15 | -1.906779661 | 23.35 | 23.3805 | | |
| 37 | 36 | 2/3/2011 | 23.25 | 23.35 | 23.05 | 23.15 | 23.15 | 15857444 | 23.6 | 23.55 | 23.2 | 22.75 | 23.1 | 1.943844492 | 1.727857451 | 0.215987041 | -1.727861771 | -0.215982721 | 23.3 | 0.647943844 | 23.335 | 23.3555 | | |
| 38 | 37 | 3/3/2011 | 23.25 | 23.5 | 23.2 | 23.3 | 23.3 | 13008846 | 23.15 | 23.6 | 23.55 | 23.2 | 22.75 | -0.64377256 | 1.287557995 | 1.072961419 | -0.429175984 | -2.360510831 | 23.6 | 1.287557995 | 23.32 | 23.338 | | |
| 39 | 38 | 4/3/2011 | 23.5 | 23.7 | 23.5 | 23.6 | 23.6 | 14299092 | 23.3 | 23.15 | 23.6 | 23.55 | 23.2 | -1.271190678 | -1.906779661 | 0 | -0.211868644 | -1.694911017 | 23.45 | -0.635588983 | 23.305 | 23.315 | | |
| 40 | 39 | 7/3/2011 | 23.3 | 23.65 | 23.3 | 23.45 | 23.45 | 11219850 | 23.6 | 23.3 | 23.15 | 23.6 | 23.55 | 0.639654557 | -0.63966735 | -1.279321907 | 0.639654557 | 0.426430685 | 23.9 | 1.9189722 | 23.285 | 23.3025 | | |
| 41 | 40 | 8/3/2011 | 23.5 | 23.9 | 23.35 | 23.9 | 23.9 | 18748851 | 23.45 | 23.6 | 23.3 | 23.15 | 23.6 | -1.882841004 | -1.255230126 | -2.510464435 | -3.138075314 | -1.255230126 | 24 | 0.418410042 | 23.36 | 23.3175 | | |
| 42 | 41 | 9/3/2011 | 24 | 24.15 | 23.95 | 24 | 24 | 21719481 | 23.9 | 23.45 | 23.6 | 23.3 | 23.15 | -0.416666667 | -2.2916625 | -1.666666667 | -2.916670833 | -3.541666667 | 23.75 | -1.041666667 | 23.45 | 23.3525 | | |
| 43 | 42 | 10/3/2011 | 23.95 | 23.95 | 23.7 | 23.75 | 23.75 | 21540303 | 24 | 23.9 | 23.45 | 23.6 | 23.3 | 1.052631579 | 0.631578947 | -1.263153684 | -0.631578947 | -1.894741053 | 23.4 | -1.473684211 | 23.55 | 23.3975 | | |

Fig. 11 Excel files for calculating the features

## 2.2 Machine Learning Model Training



Select a widget to show its description.

See workflow examples, YouTube tutorials, or open the welcome screen.

Do you wish to opt-in to sharing statistics about how you use Orange? More info...

Fig. 12 Orange training Interface

The software *Orange* was selected to train and build the machine learning models for predicting the stock price, which includes:

- Random Forest – it predicts the results by comparing the input with the root of the decision tree. The predicted price is the mean of the past prices which lie in the region that are larger / smaller than the root node of the branch.
- Neural Network – past prices are applied through interconnected hidden layer network with nodes activated by the value of threshold which is calculated by the input of features and the weight to train a model. By inputting the new values to the trained neural network in input layer, the values will be classified in output layer through activating the nodes in hidden layer.
- Support Vector Machine (SVM) – price is classified by finding a line separating the data into two classes which the margin by comparing with the support vector in both classes should be the largest.
- k Nearest Neighbor (kNN) – features are inputted, the square root of the sum of the squares of the difference between the feature of a test datapoint and that of a point in the model is

calculated as the distance and the mean or weighted mean of the top k-closest datapoint will be the predicted value.

On the other hand, the features selected on section 2.1 were spilt into 5 sets for training, producing 5 sets of stock price prediction model, which includes:

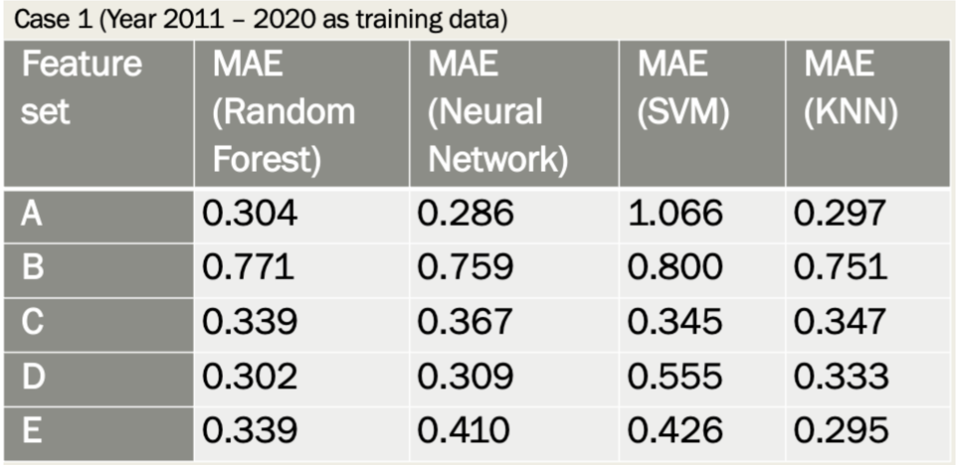
A. Close Price, Close price 1 to 5 days before
B. Close Price, Percentage change of price 1 to 5 days ago compared to today's price
C. Close Price, 10 day moving average, RSI, MACD, Signal, Histogram
D. Close Price, RSI
E. Close Price, MACD, Signal, Histogram

The above feature sets A-E were applied to 3 test cases in Orange, generating 60 price prediction model for comparison.

- Case 1: use price from year 2011 – 2020 as training data
- Case 2: use price from year 2014 – 2020 as training data
- Case 3: use price from year 2016 – 2020 as training data

### 2.3 Training Results

Below is the training result of case 1 which used Year 2011 – 2020 historical data to predict the stock price of 2021.



Case 1 (Year 2011 – 2020 as training data)

| Feature set | MAE (Random Forest) | MAE (Neural Network) | MAE (SVM) | MAE (KNN) |
|---|---|---|---|---|
| A | 0.304 | 0.286 | 1.066 | 0.297 |
| B | 0.771 | 0.759 | 0.800 | 0.751 |
| C | 0.339 | 0.367 | 0.345 | 0.347 |
| D | 0.302 | 0.309 | 0.555 | 0.333 |
| E | 0.339 | 0.410 | 0.426 | 0.295 |

Fig. 13 Training results of Case 1

Below is the bar chart of training results of Case 1. We found that feature A has the lowest MAE. For instance, Neural Networks of Feature A got 0.286 MAE which is the lowest result of training results of Case 1.

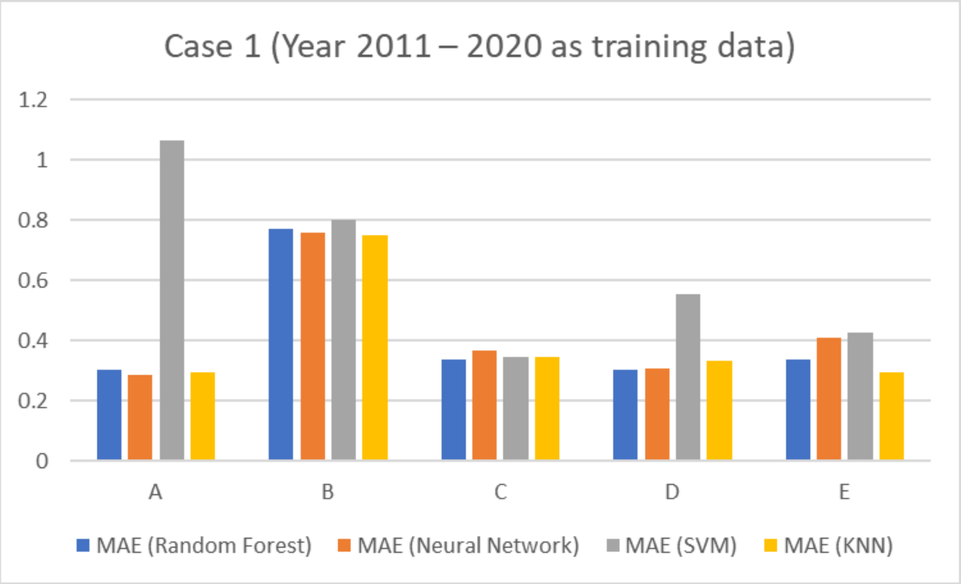
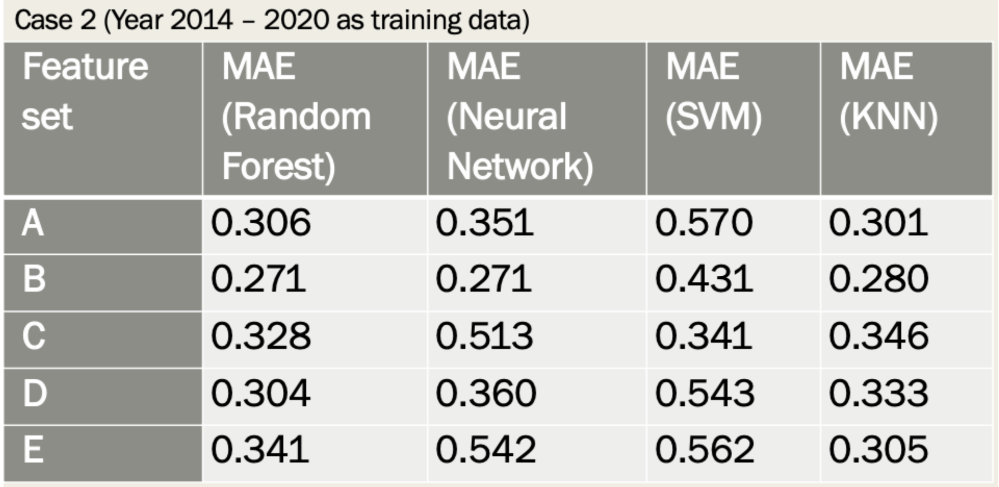


Fig. 14 Training results of Case 1 (bar chart)

Below is the training result of case 2 which used 2014 – 2020 historical data for training.



Case 2 (Year 2014 – 2020 as training data)

| Feature set | MAE (Random Forest) | MAE (Neural Network) | MAE (SVM) | MAE (KNN) |
|---|---|---|---|---|
| A | 0.306 | 0.351 | 0.570 | 0.301 |
| B | 0.271 | 0.271 | 0.431 | 0.280 |
| C | 0.328 | 0.513 | 0.341 | 0.346 |
| D | 0.304 | 0.360 | 0.543 | 0.333 |
| E | 0.341 | 0.542 | 0.562 | 0.305 |

Fig. 15 Training results of Case 2

Below is the bar chart of training results of Case 2. We found that feature B has the lowest MAE. For instance, Neural Networks and Random Forest of Feature B got 0.271 MAE which is the lowest result of training results of Case 2.  It showed that Case 1 and Case 2 have different outcomes in different features.

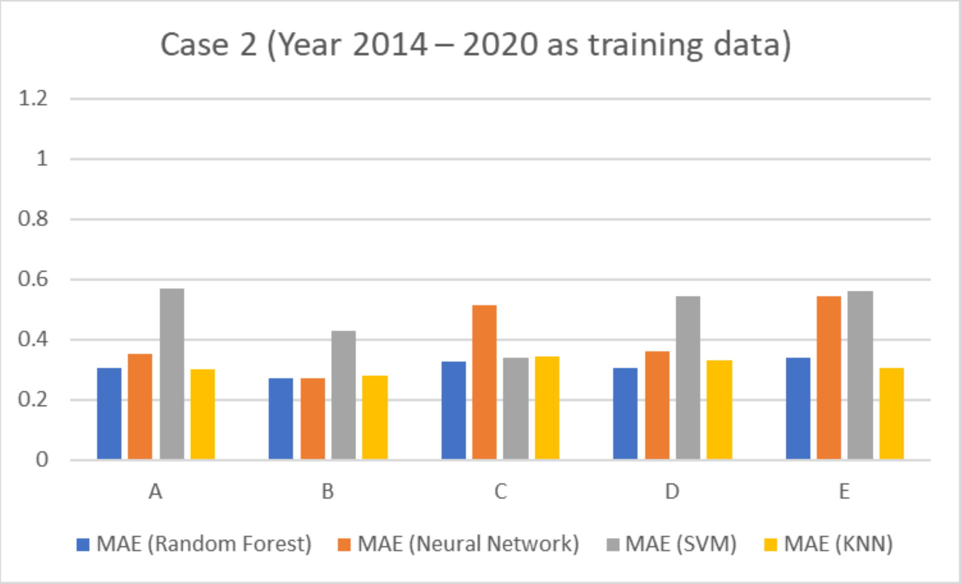
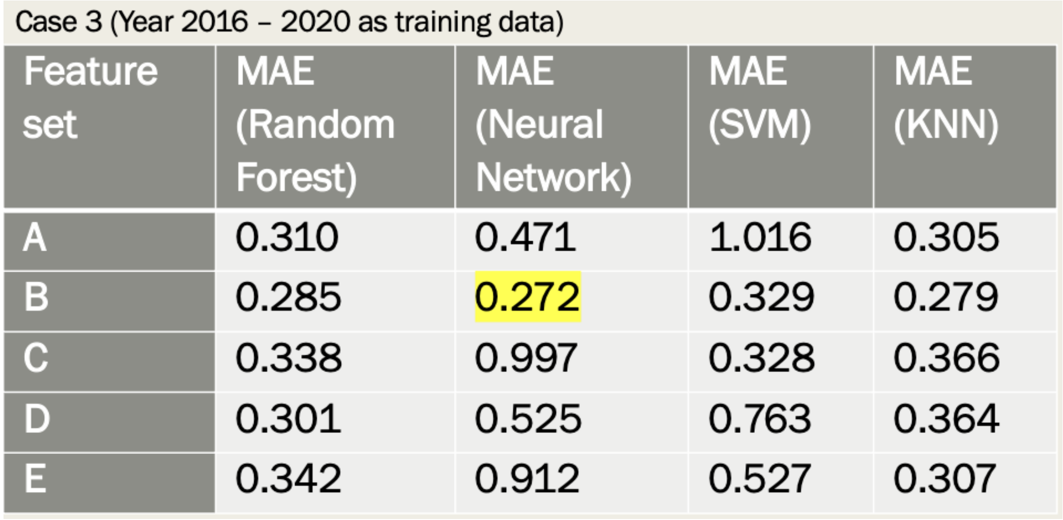


Fig. 16 Training results of Case 2 (bar chart)

Below is the training result of case 3 which used historical data within 2016 – 2020 for training our model.



Case 3 (Year 2016 – 2020 as training data)

| Feature set | MAE (Random Forest) | MAE (Neural Network) | MAE (SVM) | MAE (KNN) |
|---|---|---|---|---|
| A | 0.310 | 0.471 | 1.016 | 0.305 |
| B | 0.285 | 0.272 | 0.329 | 0.279 |
| C | 0.338 | 0.997 | 0.328 | 0.366 |
| D | 0.301 | 0.525 | 0.763 | 0.364 |
| E | 0.342 | 0.912 | 0.527 | 0.307 |

Fig. 17 Training results of Case 3

Below is the bar chart of training results of Case 3. We found that feature B has the lowest MAE. For instance, Neural Networks of Feature B got 0.272 MAE which is the lowest result of training results of Case 3.  It showed that Case 2 and Case 3 have same training result for predicting 2800 stock price in 2021.

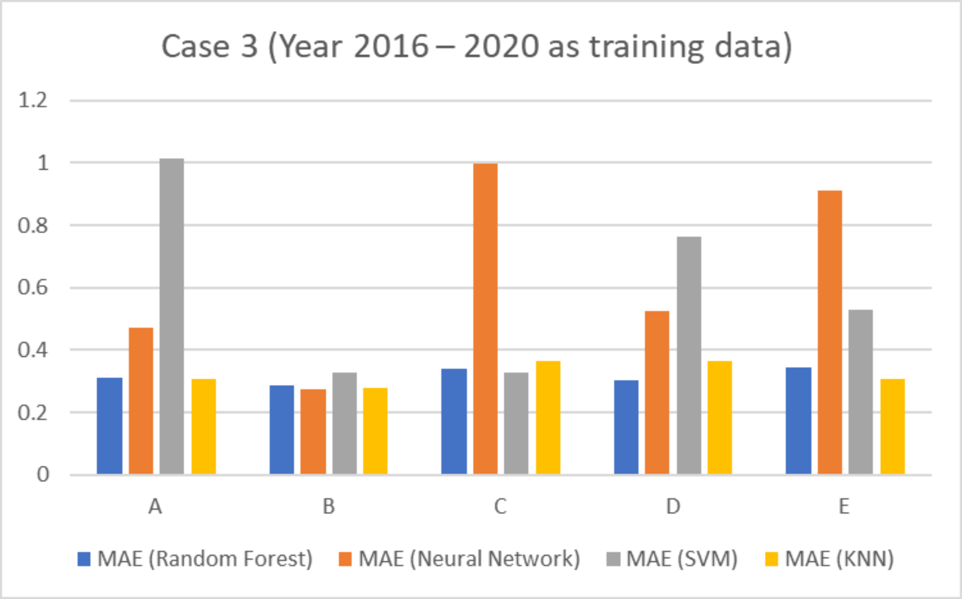


Fig. 18 Training results of Case 3 (bar chart)

In sum up of training phase, we found out that feature B (Close Price, Percentage change of price 1 to 5 days ago compared to today’s price) has highest accuracy in Case 2 and Case 3. For instance, Neural Network algorithm has 0.271 MAE in Case 2 and 0.272 MAE in Case 3. We believe Feature B is a better approach for predicting stock based on our studies.

## 3. Trading Strategies

The predicted stock price obtained from Case 3 feature set B were used for evaluating the effectiveness of the trading strategies. Total 9 strategies were developed (strategy 1A, 2A, 2B, 3A, 3B, 4A, 4B, 5A & 5B) and $1000000 initial cash were applied to all strategies to simulate stock bought/ sold in year 2016 and 2011.

An assumption had been made to all strategies with A suffix, which the fixed daily transaction amount approach was applied:

- Max transaction amount = Total cash/Number of trade days in the year
- Max number of shares to buy (Nmax) = Max transaction amount/Close price of the day

On the other hand, for all strategies with B suffix, fixed percentage approach was applied which bought with fixed percentage of cash and sold fixed percentage of shares.

### 3.1 Trading Strategy 1 (1A)

For strategy 1A, Nmax of shares are purchased each day regardless the stock is rise or fall.

### 3.2 Trading Strategy 2 – Purely based on price prediction

For strategy 2A, Nmax of shares are purchased when predicted price rise and Nmax shares are sold when predicted price fall. No action when there no price change in prediction.

For strategy 2B, 50% of the remaining capital are used to purchase shares when rise is predicted and 50% of the shares are sold when fall is predicted. No action when there no price change in prediction.

### 3.3 Trading Strategy 3 – Partially based on price prediction, partially based on technical indicator signals

For strategy 3A, half of Nmax shares is purchased when rise predicted and half of Nmax shares is sold when fall predicted. In addition, another half of Nmax shares is purchased when there is buy signal which can be indicated by moving average cross-over, RSI overbought/oversold, MACD cross-over or mean reversion, and half of Nmax shares is sold when there is sell signal.

For strategy 3B, 25% of capital is used to purchased stocks when rise predicted and 25% of stocks is sold when fall predicted. In addition, another 25% of capital is used to purchased stocks when there is buy signal and 25% of stocks is sold when there is sell signal.

### 3.4 Trading Strategy 4 – Purely based on technical indicator signals

For strategy 4A, Nmax of shares is purchased when there are buy signal and Nmax of shares is sold when there are sell signal.

For strategy 4B, 50% of capital is used to purchase shares when there are buy signal and 50% of shares is sold when there are sell signal.

### 3.5 Trading Strategy 5 – Purely based on 4 moving averages (10-Day MA, 20-Day MA, 100-Day MA, 200-Day MA)

For strategy 5A, quarter of Nmax shares are purchased when the close price is higher than the 10-day MA and quarter of Nmax shares are sold when the close price is lower than the 10-day MA.

The same practice also applied to the remained 3 quarters of Nmax shares with 20-Day MA, 100-Day MA, 200-Day MA and the buy or sell signal.

For strategy 5B, 1/10 of capital are used to purchase shares when the close price is higher than the 10-day MA and 1/10 of shares are sold when the close price is lower than the 10-day MA. In addition, 1/20 of capital are used to purchase shares when the close price is higher than the 20-day MA and 1/20 of shares are sold when the close price is lower than the 20-day MA. Furthermore, 1/100 of capital are used to purchase shares when the close price is higher than the 100-day MA and 1/100 of shares are sold when the close price is lower than the 100-day MA. Finally, 1/200 of capital are used to purchase shares when the close price is higher than the 200-day MA and 1/200 of shares are sold when the close price is lower than the 200-day MA.

### 3.6 Results & Comparison

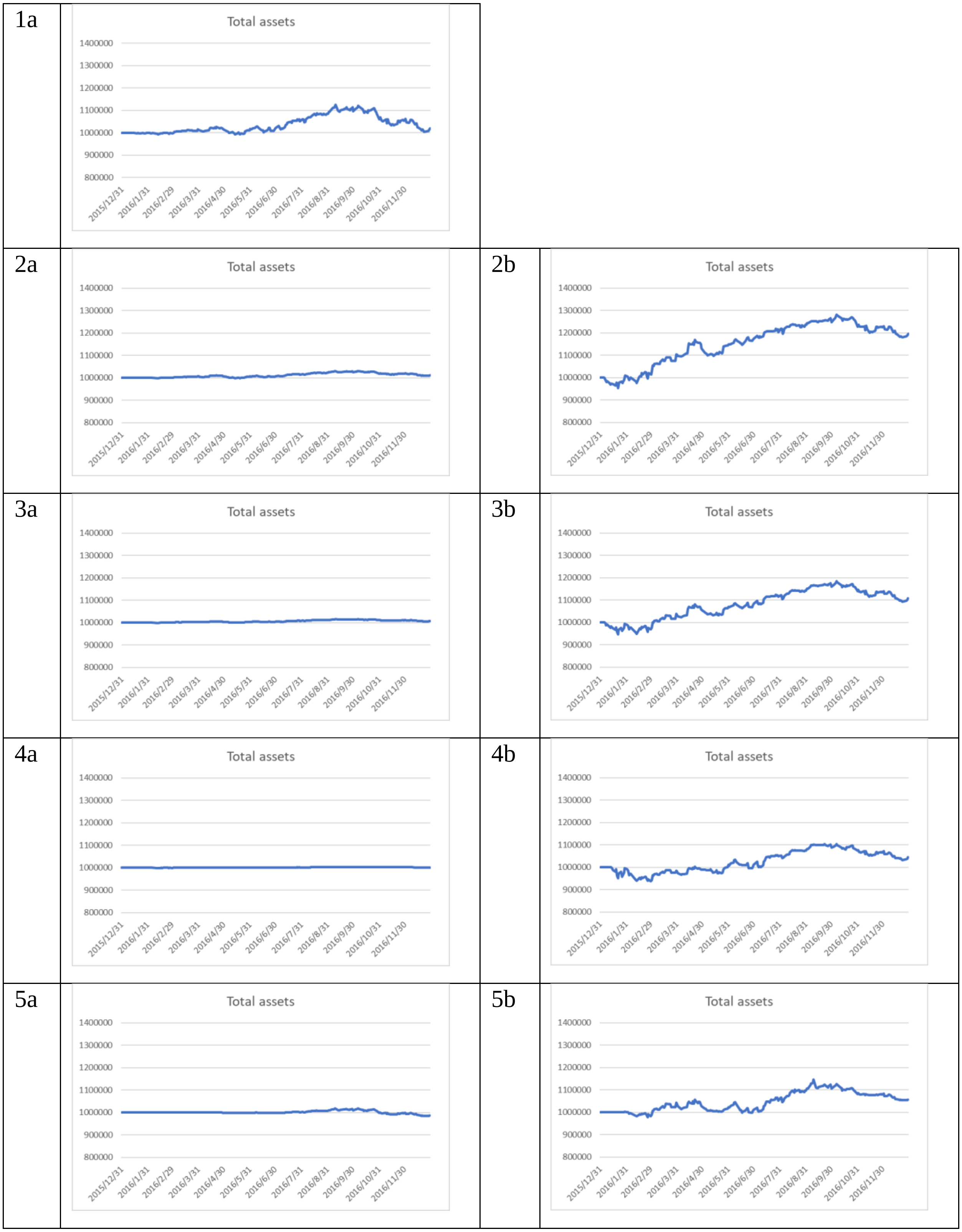


Fig. 19 Variation of total assets (2016)

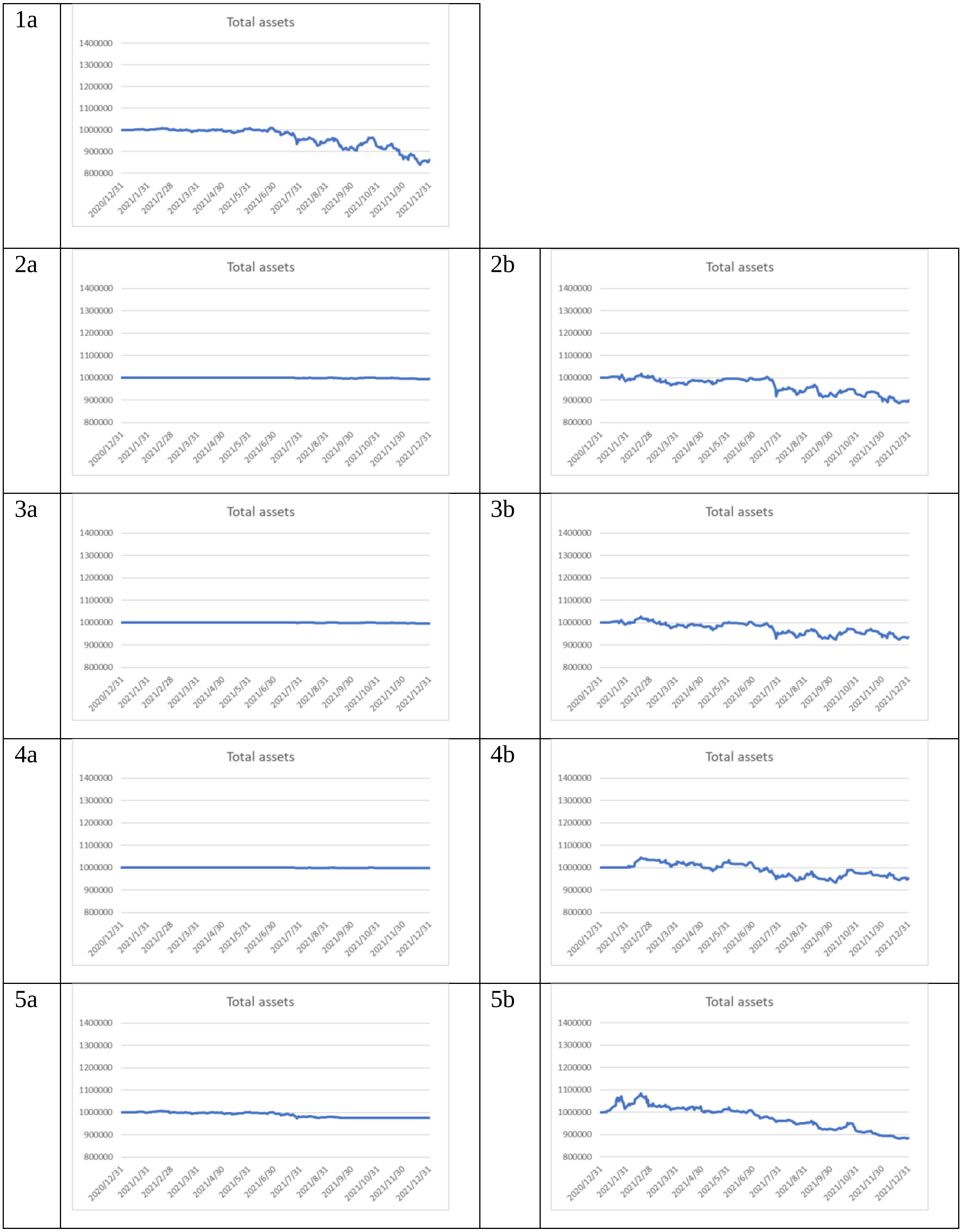
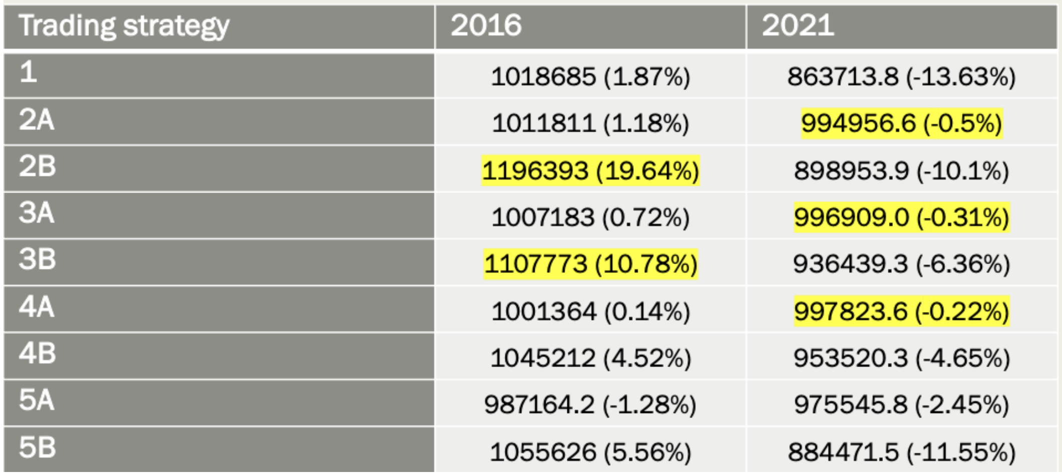


Fig. 20 Variation of total assets (2021)



| Trading strategy | 2016 | 2021 |
|---|---|---|
| 1 | 1018685 (1.87%) | 863713.8 (-13.63%) |
| 2A | 1011811 (1.18%) | 994956.6 (-0.5%) |
| 2B | 1196393 (19.64%) | 898953.9 (-10.1%) |
| 3A | 1007183 (0.72%) | 996909.0 (-0.31%) |
| 3B | 1107773 (10.78%) | 936439.3 (-6.36%) |
| 4A | 1001364 (0.14%) | 997823.6 (-0.22%) |
| 4B | 1045212 (4.52%) | 953520.3 (-4.65%) |
| 5A | 987164.2 (-1.28%) | 975545.8 (-2.45%) |
| 5B | 1055626 (5.56%) | 884471.5 (-11.55%) |

Fig. 21 Results of all trading strategy

Among all trading strategies, strategy 2B which based on purely price prediction, achieved better profits than other strategies in bull market – 2016, with 19.64% increase in capital. Strategy 3B also did well in 2016 with 10.78% gains in capital. In contrast, strategy 5A did the worst with -1.28% loss in capital.

However, when it was bear market (year 2021), strategy 4B achieved the least loss with only 0.22% loss in capital. Strategy 3B, and 2B also did well with 0.31% and 0.5% loss. In contrast, strategy 5B and 1 did the worst with 11.55% and 13.63% loss.

For buyers who consider profit is the most important and have the ability accept larger loss, strategy 2B are suggested which performs the best in bull market while having larger loss than strategy 3B/4B in bear market. On the other hand, for the buyers who consider defence the most, strategy 4B is suggested which performs the best in bear market while having 0.14% gain in capital in bull market.

Meanwhile, strategy 1 and 5 are not suggested for any buyers:

- For strategy 1, it achieved the worst in the bull market and only gain 1.87% increase in bull market.
- For strategy 5, 5A is the only strategy obtain loss in bull market and 5B is obtain a huge loss in bear market.

**References**

[1] D. M. Q. Nelson, A. C. M. Pereira, and R. A. d. Oliveira, “Stock Market’s Price Movement Prediction With LSTM Neural Networks,” presented at 2017 Stock Market’s Price Movement Prediction With LSTM Neural Networks. [Online]. doi: 10.1109/IJCNN.2017.7966019 [Accessed Feb 14, 2022].

[2] M. Ananthi, and K. Vijayakumar, “Stock market analysis using candlestick regression and market trend prediction (CKRM),” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 4819-4826, 2021. [Online]. doi: 10.1007/s12652-020-01892-5 [Accessed Feb 16, 2022].

[3] L. Sun, W. Xu, and J. Liu, “Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process. 20*, vol. 5, 2021. [Online]. doi: 10.1145/3453693 [Accessed Feb 18, 2022].

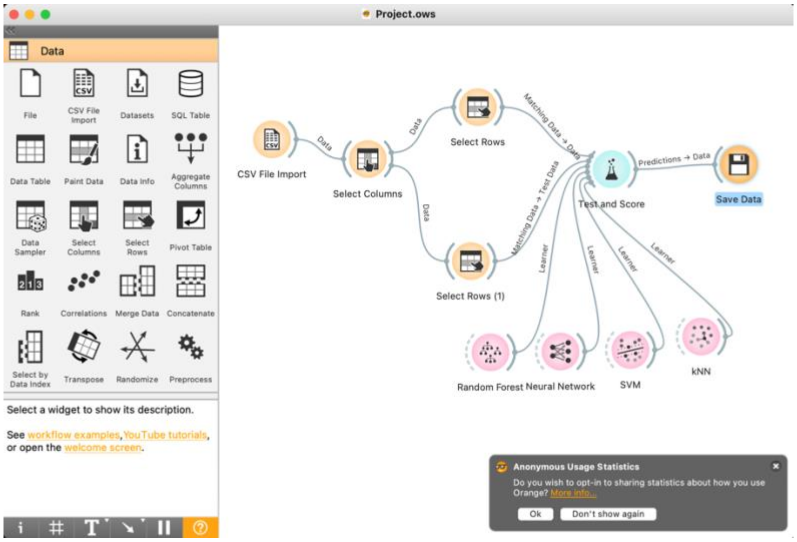
[4] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, “Stock Closing Price Prediction Using Machine Learning Techniques,” *Procedia Computer Science*, vol. 167, pp. 599-606, 2020. [Online]. doi: 10.1016/j.procs.2020.03.326 [Accessed Feb 18, 2022].

## Appendix

### Individual Report – 21000018G Li Yat Long

In this project, I oversee feature selection, model training, calculating MAE on case 3 feature set B, verifying results of trading strategies, and presentation slides & report writing.

During feature selection, different combinations of feature sets were discussed, and feature set A and B were selected to follow the method in the lab to ensure its performance while offering an opportunity to compare the results with other feature sets. Feature sets C tried included all indicators learned during lecture and feature sets D and E only included one type of indicators to compare the performance with feature set C and the methods learned from lab (feature sets A and B).



Select a widget to show its description.

See workflow examples, YouTube tutorials, or open the welcome screen.

Anonymous Usage Statistics

Do you wish to opt-in to sharing statistics about how you use Orange? More info...

Fig. 22 Model training by Orange

In the training process, I have assisted on inserting the CSV file to Orange, selected the appropriate training and testing data, and obtained the results from Test and Score module. All feature sets except feature B can be directly concluded from Orange. The MAE results obtained from Orange in feature set B calculated by the percentage change in price, which was different from the MAE of actual price our team would like to obtain. Therefore, I assisted in calculating the MAE in feature sets B following the methods learned from lab.

On the other hand, while implementing trading strategies, I assisted in verifying the excel table made by my groupmates. This was an important step not only to verify the correctness of the results but also to improve my understanding of the implementation process and the difference between the variety of strategies made by my groupmates.

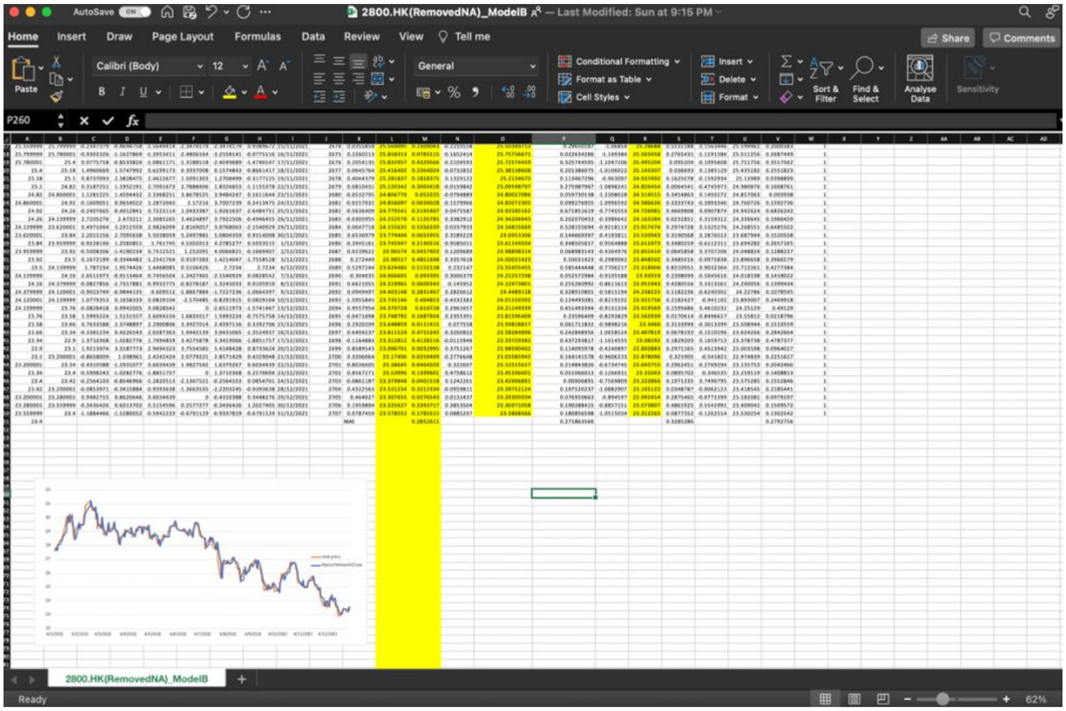
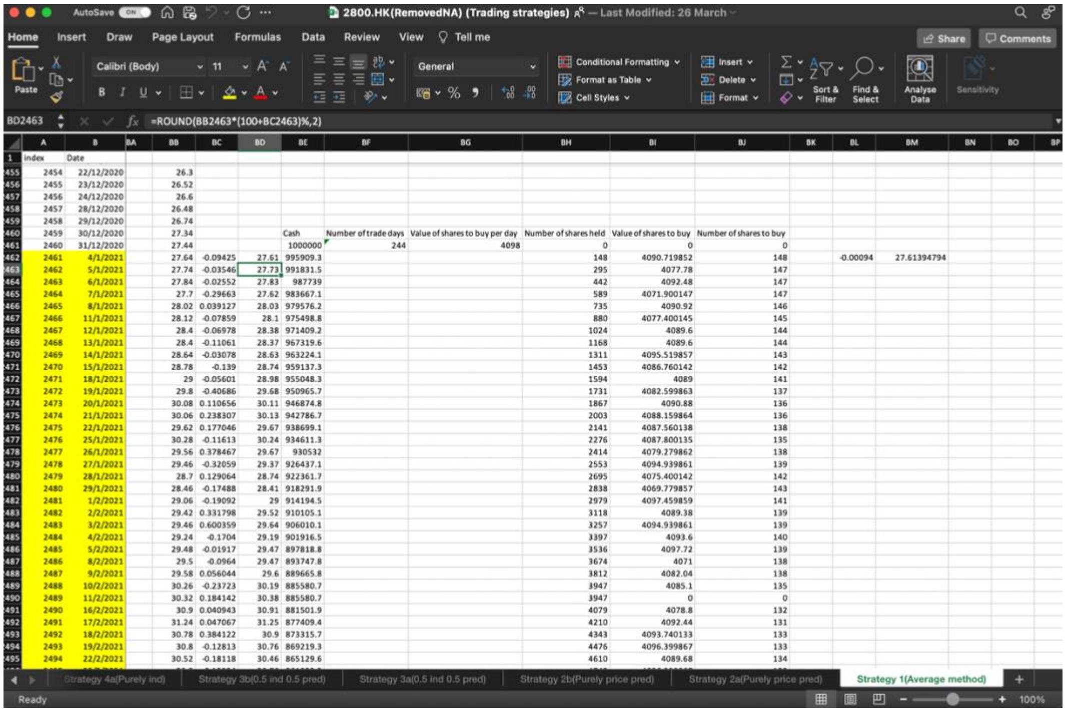


Fig. 23 Calculate the actual MAE of feature sets B



BD2463 =ROUND(BB2463*(100+BC2463)%,2)

| | A | B | BA | BB | BC | BD | BE | BF | BG | BH | BI | BJ | BK | BL | BM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | index | Date | | | | | | | | | | | | | |
| 455 | 2454 | 22/12/2020 | | 26.3 | | | | | | | | | | | |
| 456 | 2455 | 23/12/2020 | | 26.52 | | | | | | | | | | | |
| 457 | 2456 | 24/12/2020 | | 26.6 | | | | | | | | | | | |
| 458 | 2457 | 28/12/2020 | | 26.48 | | | | | | | | | | | |
| 459 | 2458 | 29/12/2020 | | 26.74 | | | | | | | | | | | |
| 460 | 2459 | 30/12/2020 | | 27.34 | | | Cash | Number of trade days | Value of shares to buy per day | Number of shares held | Value of shares to buy | Number of shares to buy | | | |
| 461 | 2460 | 31/12/2020 | | 27.44 | | | 1000000 | 244 | 4098 | 0 | 0 | 0 | | | |
| 462 | 2461 | 4/1/2021 | | 27.64 | -0.09425 | 27.61 | 995909.3 | | | 148 | 4090.719852 | 148 | | -0.00094 | 27.61394794 |
| 463 | 2462 | 5/1/2021 | | 27.74 | -0.03546 | 27.73 | 991831.5 | | | 295 | 4077.78 | 147 | | | |
| 464 | 2463 | 6/1/2021 | | 27.84 | -0.02552 | 27.83 | 987739 | | | 442 | 4092.48 | 147 | | | |
| 465 | 2464 | 7/1/2021 | | 27.7 | -0.29663 | 27.62 | 983667.1 | | | 589 | 4071.900147 | 147 | | | |
| 466 | 2465 | 8/1/2021 | | 28.02 | 0.039127 | 28.03 | 979576.2 | | | 735 | 4090.92 | 146 | | | |
| 467 | 2466 | 11/1/2021 | | 28.12 | -0.07859 | 28.1 | 975498.8 | | | 880 | 4077.400145 | 145 | | | |
| 468 | 2467 | 12/1/2021 | | 28.4 | -0.06978 | 28.38 | 971409.2 | | | 1024 | 4089.6 | 144 | | | |
| 469 | 2468 | 13/1/2021 | | 28.4 | -0.11061 | 28.37 | 967319.6 | | | 1168 | 4089.6 | 144 | | | |
| 470 | 2469 | 14/1/2021 | | 28.64 | -0.03078 | 28.63 | 963224.1 | | | 1311 | 4095.519857 | 143 | | | |
| 471 | 2470 | 15/1/2021 | | 28.78 | -0.139 | 28.74 | 959137.3 | | | 1453 | 4086.760142 | 142 | | | |
| 472 | 2471 | 18/1/2021 | | 29 | -0.05601 | 28.98 | 955048.3 | | | 1594 | 4089 | 141 | | | |
| 473 | 2472 | 19/1/2021 | | 29.8 | -0.40686 | 29.68 | 950965.7 | | | 1731 | 4082.599863 | 137 | | | |
| 474 | 2473 | 20/1/2021 | | 30.08 | 0.110656 | 30.11 | 946874.8 | | | 1867 | 4090.88 | 136 | | | |
| 475 | 2474 | 21/1/2021 | | 30.06 | 0.238307 | 30.13 | 942786.7 | | | 2003 | 4088.159864 | 136 | | | |
| 476 | 2475 | 22/1/2021 | | 29.62 | 0.177046 | 29.67 | 938699.1 | | | 2141 | 4087.560138 | 138 | | | |
| 477 | 2476 | 25/1/2021 | | 30.28 | -0.11613 | 30.24 | 934611.3 | | | 2276 | 4087.800135 | 135 | | | |
| 478 | 2477 | 26/1/2021 | | 29.56 | 0.378467 | 29.67 | 930532 | | | 2414 | 4079.279862 | 138 | | | |
| 479 | 2478 | 27/1/2021 | | 29.46 | -0.32059 | 29.37 | 926437.1 | | | 2553 | 4094.939861 | 139 | | | |
| 480 | 2479 | 28/1/2021 | | 28.7 | 0.129064 | 28.74 | 922361.7 | | | 2695 | 4075.400142 | 142 | | | |
| 481 | 2480 | 29/1/2021 | | 28.46 | -0.17488 | 28.41 | 918291.9 | | | 2838 | 4069.779857 | 143 | | | |
| 482 | 2481 | 1/2/2021 | | 29.06 | -0.19092 | 29 | 914194.5 | | | 2979 | 4097.459859 | 141 | | | |
| 483 | 2482 | 2/2/2021 | | 29.42 | 0.331798 | 29.52 | 910105.1 | | | 3118 | 4089.38 | 139 | | | |
| 484 | 2483 | 3/2/2021 | | 29.46 | 0.600359 | 29.64 | 906010.1 | | | 3257 | 4094.939861 | 139 | | | |
| 485 | 2484 | 4/2/2021 | | 29.24 | -0.1704 | 29.19 | 901916.5 | | | 3397 | 4093.6 | 140 | | | |
| 486 | 2485 | 5/2/2021 | | 29.48 | -0.01917 | 29.47 | 897818.8 | | | 3536 | 4097.72 | 139 | | | |
| 487 | 2486 | 8/2/2021 | | 29.5 | -0.0964 | 29.47 | 893747.8 | | | 3674 | 4071 | 138 | | | |
| 488 | 2487 | 9/2/2021 | | 29.58 | 0.056044 | 29.6 | 889665.8 | | | 3812 | 4082.04 | 138 | | | |
| 489 | 2488 | 10/2/2021 | | 30.26 | -0.23723 | 30.19 | 885580.7 | | | 3947 | 4085.1 | 135 | | | |
| 490 | 2489 | 11/2/2021 | | 30.32 | 0.184142 | 30.38 | 885580.7 | | | 3947 | 0 | 0 | | | |
| 491 | 2490 | 16/2/2021 | | 30.9 | 0.040943 | 30.91 | 881501.9 | | | 4079 | 4078.8 | 132 | | | |
| 492 | 2491 | 17/2/2021 | | 31.24 | 0.047067 | 31.25 | 877409.4 | | | 4210 | 4092.44 | 131 | | | |
| 493 | 2492 | 18/2/2021 | | 30.78 | 0.384122 | 30.9 | 873315.7 | | | 4343 | 4093.740133 | 133 | | | |
| 494 | 2493 | 19/2/2021 | | 30.8 | -0.12813 | 30.76 | 869219.3 | | | 4476 | 4096.399867 | 133 | | | |
| 495 | 2494 | 22/2/2021 | | 30.52 | -0.18118 | 30.46 | 865129.6 | | | 4610 | 4089.68 | 134 | | | |

Fig. 24 Verifying the results of trading strategies

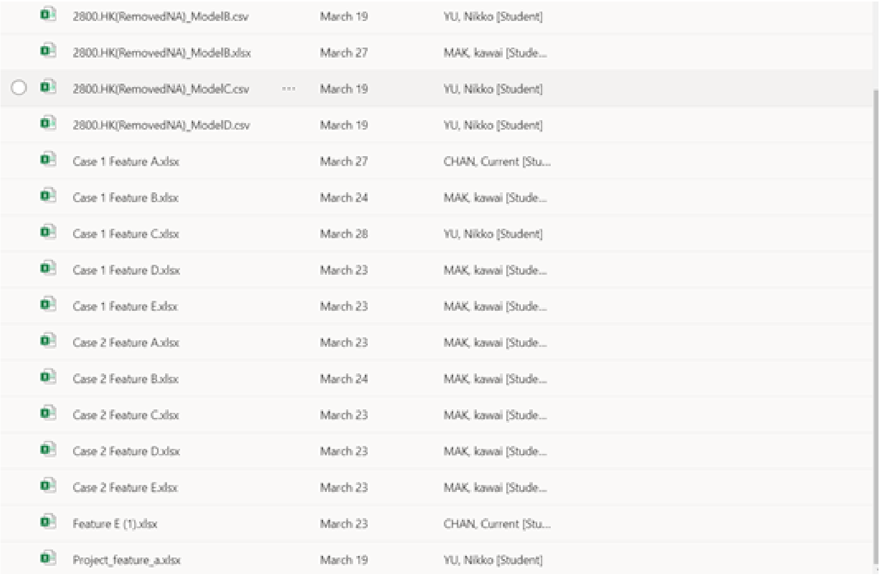
Finally, during the report and presentation slides writing, I discussed with my teammates to include which kind of graphs or table to represent our results clearly to our audience. One of the results was the chart shown in Fig. 23. In addition, I have also implemented both section 2 and 3 of report and presentation slides partly. Through this process, I have understood more of the parts not implemented by me like trading strategies and have better understanding of the results on both section 2 and 3. For section 1, I have understood more on the stock price prediction using LSTM which is not mentioned in the lectures, with is methodology and mechanism.

## Individual Report – 21009566G Mak Ka Wai

In this project, I am responsible for feature selection discussion, prediction tools discussion and trading strategy discussion, model training, calculating MAE of feature set B in case 1 and case 2, writing presentation slides & report and presentation.

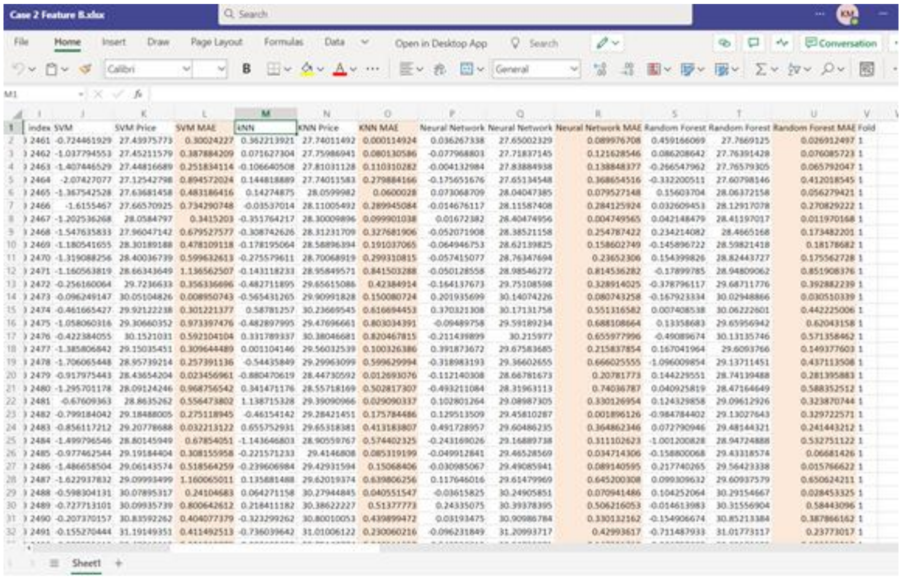
Firstly, I have discussed with team members how to predict 2800 stock price. For instance, what tools we are planning to use. Which years are we predicting? Which features will we use to train our algorithm? After discussion, we decided to use Orange to train our dataset. Also, we will have 5 set features shown in reports to see which feature will get the highest accuracy. Moreover, we set 3 different cases to predict 2800 stock price in 2021.

Apart from discussion, I have helped to use Orange to train the model for case 1 and case 2 to find out the finding of our research. After finding out the result of each feature, I have exported the finding as excel and marked down in our report and presentation slide. Also, I have calculated MAE of feature B in excel based on the instruction of the workshop.



| | | |
|---|---|---|
| 2800.HK(RemovedNA)_ModelB.csv | March 19 | YU, Nikko [Student] |
| 2800.HK(RemovedNA)_ModelB.xlsx | March 27 | MAK, kawai [Stude... |
| 2800.HK(RemovedNA)_ModelC.csv | March 19 | YU, Nikko [Student] |
| 2800.HK(RemovedNA)_ModelD.csv | March 19 | YU, Nikko [Student] |
| Case 1 Feature A.xlsx | March 27 | CHAN, Current [Stu... |
| Case 1 Feature B.xlsx | March 24 | MAK, kawai [Stude... |
| Case 1 Feature C.xlsx | March 28 | YU, Nikko [Student] |
| Case 1 Feature D.xlsx | March 23 | MAK, kawai [Stude... |
| Case 1 Feature E.xlsx | March 23 | MAK, kawai [Stude... |
| Case 2 Feature A.xlsx | March 23 | MAK, kawai [Stude... |
| Case 2 Feature B.xlsx | March 24 | MAK, kawai [Stude... |
| Case 2 Feature C.xlsx | March 23 | MAK, kawai [Stude... |
| Case 2 Feature D.xlsx | March 23 | MAK, kawai [Stude... |
| Case 2 Feature E.xlsx | March 23 | MAK, kawai [Stude... |
| Feature E (1).xlsx | March 23 | CHAN, Current [Stu... |
| Project_feature_a.xlsx | March 19 | YU, Nikko [Student] |

Fig. 25 Training result of case 1 and case 2



| index | SVM | SVM Price | SVM MAE | KNN | KNN Price | KNN MAE | Neural Network | Neural Network | Neural Network MAE | Random Forest | Random Forest | Random Forest MAE | Fold |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2461 | -0.724461929 | 27.43975773 | 0.30024227 | 0.362213921 | 27.74011492 | 0.000114924 | 0.036267338 | 27.65002329 | 0.089976708 | 0.459166069 | 27.7669125 | 0.026912497 | 1 |
| 2462 | -1.037794553 | 27.45211579 | 0.387884209 | 0.071627304 | 27.75986941 | 0.080130586 | -0.077968803 | 27.71837145 | 0.121628546 | 0.086208642 | 27.76391428 | 0.076085723 | 1 |
| 2463 | -1.407446529 | 27.44816689 | 0.251834114 | -0.106640508 | 27.81031128 | 0.110310282 | -0.004132984 | 27.83884938 | 0.138848377 | -0.266547962 | 27.76579305 | 0.065792047 | 1 |
| 2464 | -2.07427077 | 27.12542798 | 0.894572024 | 0.144818889 | 27.74011583 | 0.279884166 | -0.175651676 | 27.65134548 | 0.368654516 | -0.332200511 | 27.60798146 | 0.412018545 | 1 |
| 2465 | -1.367542528 | 27.63681458 | 0.483186416 | 0.14274875 | 28.0599982 | 0.0600028 | 0.073068709 | 28.04047385 | 0.079527148 | 0.15603704 | 28.06372158 | 0.056279421 | 1 |
| 2466 | -1.6155467 | 27.66570925 | 0.734290748 | -0.03537014 | 28.11005492 | 0.289945084 | -0.014676117 | 28.11587408 | 0.284125924 | 0.032609453 | 28.12917078 | 0.270829222 | 1 |
| 2467 | -1.202536268 | 28.0584797 | 0.3415203 | -0.351764217 | 28.30009896 | 0.099901038 | 0.01672382 | 28.40474956 | 0.004749565 | 0.042148479 | 28.41197017 | 0.011970168 | 1 |
| 2468 | -1.547635833 | 27.96047142 | 0.679527577 | -0.308742626 | 28.31231709 | 0.327681906 | -0.052071908 | 28.38521158 | 0.254787422 | 0.234214082 | 28.4665168 | 0.173482201 | 1 |
| 2469 | -1.180541655 | 28.30189188 | 0.478109118 | -0.178195064 | 28.58896394 | 0.191037065 | -0.064946753 | 28.62139825 | 0.158602749 | -0.145896722 | 28.59821418 | 0.18178682 | 1 |
| 2470 | -1.319088256 | 28.40036739 | 0.599632613 | -0.275579611 | 28.70068919 | 0.299310815 | -0.057415077 | 28.76347694 | 0.23652306 | 0.154399826 | 28.82443727 | 0.175562728 | 1 |
| 2471 | -1.160563819 | 28.66343649 | 1.136562507 | -0.143118233 | 28.95849571 | 0.841503288 | -0.050128558 | 28.98546272 | 0.814536282 | -0.17899785 | 28.94809062 | 0.851908376 | 1 |
| 2472 | -0.256160064 | 29.7236633 | 0.356336696 | -0.482711895 | 29.65615086 | 0.42384914 | -0.164137673 | 29.75108598 | 0.328914025 | -0.378796117 | 29.68711776 | 0.392882239 | 1 |
| 2473 | -0.096249147 | 30.05104826 | 0.008950743 | -0.565431265 | 29.90991828 | 0.150080724 | 0.201935699 | 30.14074226 | 0.080743258 | -0.167923334 | 30.02948866 | 0.030510339 | 1 |
| 2474 | -0.461665427 | 29.92122238 | 0.301221377 | 0.58781257 | 30.23669545 | 0.616694453 | 0.370321308 | 30.17131758 | 0.551316582 | 0.007408538 | 30.06222601 | 0.442225006 | 1 |
| 2475 | -1.058060316 | 29.30660352 | 0.973397476 | -0.482897995 | 29.47696661 | 0.803034391 | -0.09489758 | 29.59189234 | 0.688108664 | 0.13358683 | 29.65956942 | 0.62043158 | 1 |
| 2476 | -0.422384055 | 30.1521031 | 0.592104104 | 0.331789337 | 30.38046681 | 0.820467815 | -0.211439899 | 30.215977 | 0.655977996 | -0.49089674 | 30.13135746 | 0.571358462 | 1 |
| 2477 | -1.385806842 | 29.15035451 | 0.309644489 | 0.001104146 | 29.56032539 | 0.100326386 | 0.391873672 | 29.67583685 | 0.215837854 | 0.167041964 | 29.6093766 | 0.149377603 | 1 |
| 2478 | -1.706065448 | 28.95739214 | 0.257391136 | -0.54435849 | 29.29963099 | 0.599629994 | -0.318983193 | 29.36602655 | 0.666025555 | -1.096009854 | 29.13711451 | 0.437113508 | 1 |
| 2479 | -0.917975443 | 28.43654204 | 0.023456961 | -0.880470619 | 28.44730592 | 0.012693076 | -0.112140308 | 28.66781673 | 0.20781773 | 0.144229551 | 28.74139488 | 0.281395883 | 1 |
| 2480 | -1.295701178 | 28.09124246 | 0.968756542 | 0.341471176 | 28.55718169 | 0.502817307 | -0.493211084 | 28.31963113 | 0.74036787 | 0.040925819 | 28.47164649 | 0.588352512 | 1 |
| 2481 | -0.67609363 | 28.8635262 | 0.556473802 | 1.138715328 | 29.39090966 | 0.029090337 | 0.102801264 | 29.08987305 | 0.330126954 | 0.124329858 | 29.09612926 | 0.323870744 | 1 |
| 2482 | -0.799184042 | 29.18488005 | 0.275118945 | -0.46154142 | 29.28421451 | 0.175784486 | 0.129513509 | 29.45810287 | 0.001896126 | -0.984784402 | 29.13027643 | 0.329722571 | 1 |
| 2483 | -0.856117212 | 29.20778688 | 0.032213122 | 0.655752931 | 29.65318381 | 0.413183807 | 0.491728957 | 29.60486235 | 0.364862346 | 0.072790946 | 29.48144321 | 0.241443212 | 1 |
| 2484 | -1.499796546 | 28.80145949 | 0.67854051 | -1.143646803 | 28.90559767 | 0.574402325 | -0.243169026 | 29.16889738 | 0.311102623 | -1.001200828 | 28.94724888 | 0.532751122 | 1 |
| 2485 | -0.977462544 | 29.19184404 | 0.308155958 | -0.221571233 | 29.4146808 | 0.085319199 | -0.049912841 | 29.46528569 | 0.034714306 | -0.158800068 | 29.43318574 | 0.06681426 | 1 |
| 2486 | -1.486658504 | 29.06143574 | 0.518564259 | -0.239606984 | 29.42931594 | 0.15068406 | -0.030985067 | 29.49085941 | 0.089140595 | 0.217740265 | 29.56423338 | 0.015766622 | 1 |
| 2487 | -1.622937832 | 29.09993499 | 1.160065011 | 0.135881488 | 29.62019374 | 0.639806256 | 0.117646016 | 29.61479969 | 0.645200308 | 0.099309632 | 29.60937579 | 0.650624211 | 1 |
| 2488 | -0.598304131 | 30.07895317 | 0.24104683 | 0.064271158 | 30.27944845 | 0.040551547 | -0.03615825 | 30.24905851 | 0.070941486 | 0.104252064 | 30.29154667 | 0.028453325 | 1 |
| 2489 | -0.727713101 | 30.09935739 | 0.800642612 | 0.218411182 | 30.38622227 | 0.51377773 | 0.24335075 | 30.39378395 | 0.506216053 | -0.014613983 | 30.31556904 | 0.58443096 | 1 |
| 2490 | -0.207370157 | 30.83592262 | 0.404077379 | -0.323299262 | 30.80010053 | 0.439899472 | 0.03193475 | 30.90986784 | 0.330132162 | -0.154906674 | 30.85213384 | 0.387866162 | 1 |
| 2491 | -0.155270444 | 31.19149351 | 0.411492513 | -0.736039642 | 31.01006122 | 0.230060216 | -0.096231849 | 31.20993717 | 0.42993617 | -0.711487933 | 31.01773117 | 0.23773017 | 1 |

Fig. 26 Calculating MAE of feature B

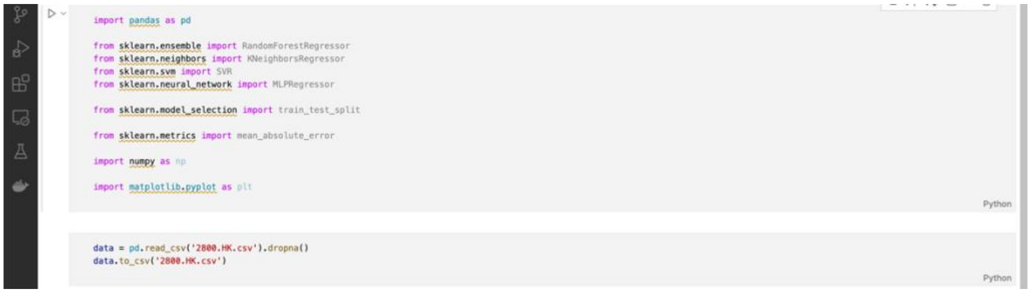
Apart from training models, I have discussed trading strategy and graphs of our findings in the presentation slide with my team members. For instance, I am responsible for presenting the training results during the presentation.

After working on this project, I have learnt how to analysis and predict a stock using machine learning algorithms in real world stock market. This project enhanced my confidence in investing in stocks or building up my financial portfolio. Moreover, I think Orange is easy to implement different machine learning algorithms to predict stocks in the market.

## Individual Report – 21009856G Chan Siu Fung

Contribution

During the project, I have mainly contributed for the calculation-required tasks and graph-plotting. I have removed the rows with empty values from the raw historical data that my groupmates have downloaded from the Internet by using the skill learnt in the extra Python tutorial.

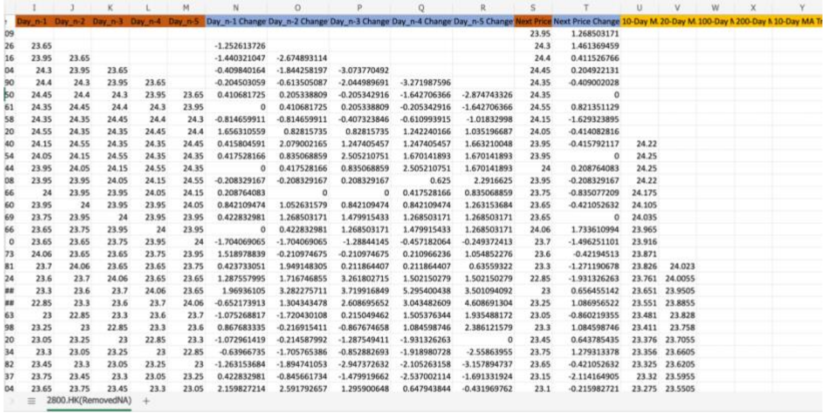


```
import pandas as pd

from sklearn.ensemble import RandomForestRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.svm import SVR
from sklearn.neural_network import MLPRegressor

from sklearn.model_selection import train_test_split

from sklearn.metrics import mean_absolute_error

import numpy as np

import matplotlib.pyplot as plt
```

```
data = pd.read_csv('2800.HK.csv').dropna()
data.to_csv('2800.HK.csv')
```

Fig. 27 Python codes for removing empty entries in the csv files downloaded from the Internet

I have calculated the technical indicators and prepared the derived features by using Excel.



| | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Day_n-1 | Day_n-2 | Day_n-3 | Day_n-4 | Day_n-5 | Day_n-1 Change | Day_n-2 Change | Day_n-3 Change | Day_n-4 Change | Day_n-5 Change | Next Price | Next Price Change | 10-Day M. | 20-Day M. | 100-Day N | 200-Day N | 10-Day MA Tr |
| 09 | | | | | | | | | | | 23.95 | 1.268503171 | | | | | |
| 26 | 23.65 | | | | | -1.252613726 | | | | | 24.3 | 1.461369459 | | | | | |
| 16 | 23.95 | 23.65 | | | | -1.440321047 | -2.674893114 | | | | 24.4 | 0.411526766 | | | | | |
| 04 | 24.3 | 23.95 | 23.65 | | | -0.409840164 | -1.844258197 | -3.073770492 | | | 24.45 | 0.204922131 | | | | | |
| 90 | 24.4 | 24.3 | 23.95 | 23.65 | | -0.204503059 | -0.613505087 | -2.044989691 | -3.271987596 | | 24.35 | -0.409002028 | | | | | |
| 50 | 24.45 | 24.4 | 24.3 | 23.95 | 23.65 | 0.410681725 | 0.205338809 | -0.205342916 | -1.642706366 | -2.874743326 | 24.35 | 0 | | | | | |
| 61 | 24.35 | 24.45 | 24.4 | 24.3 | 23.95 | 0 | 0.410681725 | 0.205338809 | -0.205342916 | -1.642706366 | 24.55 | 0.821351129 | | | | | |
| 58 | 24.35 | 24.35 | 24.45 | 24.4 | 24.3 | -0.814659911 | -0.814659911 | -0.407323846 | -0.610993915 | -1.01832998 | 24.15 | -1.629323895 | | | | | |
| 20 | 24.55 | 24.35 | 24.35 | 24.45 | 24.4 | 1.656310559 | 0.82815735 | 0.82815735 | 1.242240166 | 1.035196687 | 24.05 | -0.414082816 | | | | | |
| 40 | 24.15 | 24.55 | 24.35 | 24.35 | 24.45 | 0.415804591 | 2.079002165 | 1.247405457 | 1.247405457 | 1.663210048 | 23.95 | -0.415792117 | 24.22 | | | | |
| 54 | 24.05 | 24.15 | 24.55 | 24.35 | 24.35 | 0.417528166 | 0.835068859 | 2.505210751 | 1.670141893 | 1.670141893 | 23.95 | 0 | 24.25 | | | | |
| 44 | 23.95 | 24.05 | 24.15 | 24.55 | 24.35 | 0 | 0.417528166 | 0.835068859 | 2.505210751 | 1.670141893 | 24 | 0.208764083 | 24.25 | | | | |
| 08 | 23.95 | 23.95 | 24.05 | 24.15 | 24.55 | -0.208329167 | -0.208329167 | 0.208329167 | 0.625 | 2.2916625 | 23.95 | -0.208329167 | 24.22 | | | | |
| 66 | 24 | 23.95 | 23.95 | 24.05 | 24.15 | 0.208764083 | 0 | 0 | 0.417528166 | 0.835068859 | 23.75 | -0.835077209 | 24.175 | | | | |
| 60 | 23.95 | 24 | 23.95 | 23.95 | 24.05 | 0.842109474 | 1.052631579 | 0.842109474 | 0.842109474 | 1.263153684 | 23.65 | -0.421052632 | 24.105 | | | | |
| 69 | 23.75 | 23.95 | 24 | 23.95 | 23.95 | 0.422832981 | 1.268503171 | 1.479915433 | 1.268503171 | 1.268503171 | 23.65 | 0 | 24.035 | | | | |
| 66 | 23.65 | 23.75 | 23.95 | 24 | 23.95 | 0 | 0.422832981 | 1.268503171 | 1.479915433 | 1.268503171 | 24.06 | 1.733610994 | 23.965 | | | | |
| 0 | 23.65 | 23.65 | 23.75 | 23.95 | 24 | -1.704069065 | -1.704069065 | -1.28844145 | -0.457182064 | -0.249372413 | 23.7 | -1.496251101 | 23.916 | | | | |
| 73 | 24.06 | 23.65 | 23.65 | 23.75 | 23.95 | 1.518978839 | -0.210974675 | -0.210974675 | 0.210966236 | 1.054852276 | 23.6 | -0.42194513 | 23.871 | | | | |
| 81 | 23.7 | 24.06 | 23.65 | 23.65 | 23.75 | 0.423733051 | 1.949148305 | 0.211864407 | 0.211864407 | 0.63559322 | 23.3 | -1.271190678 | 23.826 | 24.023 | | | |
| 24 | 23.6 | 23.7 | 24.06 | 23.65 | 23.65 | 1.287557995 | 1.716746855 | 3.261802715 | 1.502150279 | 1.502150279 | 22.85 | -1.931326263 | 23.761 | 24.0055 | | | |
| ## | 23.3 | 23.6 | 23.7 | 24.06 | 23.65 | 1.96936105 | 3.282275711 | 3.719916849 | 5.295400438 | 3.501094092 | 23 | 0.656455142 | 23.651 | 23.9505 | | | |
| ## | 22.85 | 23.3 | 23.6 | 23.7 | 24.06 | -0.652173913 | 1.304343478 | 2.608695652 | 3.043482609 | 4.608691304 | 23.25 | 1.086956522 | 23.551 | 23.8855 | | | |
| 63 | 23 | 22.85 | 23.3 | 23.6 | 23.7 | -1.075268817 | -1.720430108 | 0.215049462 | 1.505376344 | 1.935488172 | 23.05 | -0.860219355 | 23.481 | 23.828 | | | |
| 98 | 23.25 | 23 | 22.85 | 23.3 | 23.6 | 0.867683335 | -0.216915411 | -0.867674658 | 1.084598746 | 2.386121579 | 23.3 | 1.084598746 | 23.411 | 23.758 | | | |
| 20 | 23.05 | 23.25 | 23 | 22.85 | 23.3 | -1.072961419 | -0.214587992 | -1.287549411 | -1.931326263 | 0 | 23.45 | 0.643785435 | 23.376 | 23.7055 | | | |
| 34 | 23.3 | 23.05 | 23.25 | 23 | 22.85 | -0.63966735 | -1.705765386 | -0.852882693 | -1.918980728 | -2.55863955 | 23.75 | 1.279313378 | 23.356 | 23.6605 | | | |
| 82 | 23.45 | 23.3 | 23.05 | 23.25 | 23 | -1.263153684 | -1.894741053 | -2.947372632 | -2.105263158 | -3.157894737 | 23.65 | -0.421052632 | 23.325 | 23.6205 | | | |
| 37 | 23.75 | 23.45 | 23.3 | 23.05 | 23.25 | 0.422832981 | -0.845661734 | -1.479919662 | -2.537002114 | -1.691331924 | 23.15 | -2.114164905 | 23.32 | 23.5955 | | | |
| 04 | 23.65 | 23.75 | 23.45 | 23.3 | 23.05 | 2.159827214 | 2.591792657 | 1.295900648 | 0.647943844 | -0.431969762 | 23.1 | -0.215982721 | 23.275 | 23.5505 | | | |

Fig. 28 Screenshot of derived features of technical indicators

I have calculated the next price change percentage, the moving averages (10-day, 20-day, 100-day, 200-day), found if it is an upward trend or downward trend by comparing the moving averages with the close price, the buy/sell signals detected by looking for the cross-over between close price line and moving average lines or the pull-back from the moving average lines.

I have also calculated the daily gain, daily loss, average gain, average loss, relative strength, and relative strength index, looked for buy/sell signals due to bullish or bearish divergence.

I have then calculated the 12-day EMA and 26-day EMA for calculating the MACD and signal, and thus the histogram and the cross-over between MACD line and signal line.

I have also calculated the lower and upper Bollinger bands and looked for mean reversion. In the trading strategies part, I have proposed different strategies, actualized the trading and calculated the final total assets for each strategy.

I have plotted clustered column bar charts for illustrating the training error of different models trained with different sets of features in different cases and the plots for the daily variation of total assets with different trading strategies.

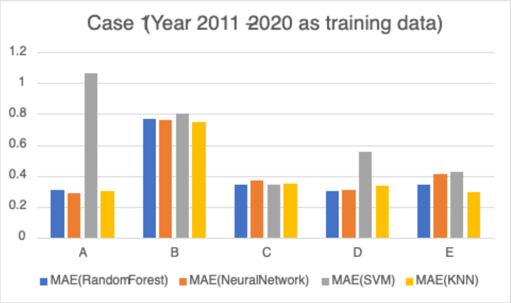


Fig. 29 Clustered column bar chart for comparing training errors of different models trained with different sets of features in Case 1

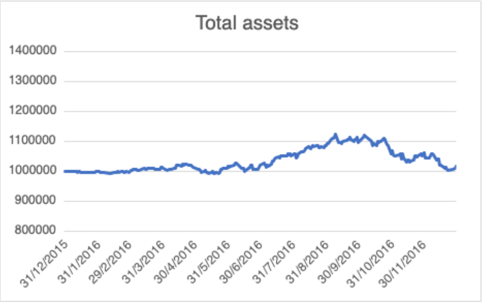


Fig. 30 Daily variation of total assets with different trading strategies

## Learning reflection

Through the study of paper, I think I can understand the paper with what I have learnt in the lectures in both this course and the course of Machine Learning and Data Analytics, as there are so many technical terms in the Machine Learning aspect (terms like features, MAE, MSE, RMSE, MLP, CNN, RNN, LSTM, sigmoid function, tanh function, hidden states, softmax function, etc.) and the Finance aspect (terms like return rate, Sharpe ratio). For understanding better, I have also checked the difference in meaning between some similar terms MLP and CNN, the former one is fully connected while the latter one is sparsely or partially connected.

More specific to the paper that I have selected, I have learnt that there is existing model which can involve both stock prices and stock news for stock price prediction. I have learnt that the use of dual-channel fusion can avoid interference between two types of features and at the same time gain the benefit from the fusion of two different models.

I have also learnt the text data extraction by using CNN which involves segmentation of text, stopping of words, embedding of words, compression of word vectors into sentence vectors, that of sentence vectors into chapter vectors, and that of chapter vectors into enterprise vector.

Through the stock prediction and the trading strategies parts, I have applied what I have learnt in the lectures about how to calculate various types of technical indicators (e.g. moving averages, RSI, MACD, Bollinger bands, etc.) and how to preliminarily look for buy/sell signals regardless of my limited skill for further checking for uptrend or downtrend from a more macroscopic perspective.

I have also learnt how to carry out mock trading by using Excel.

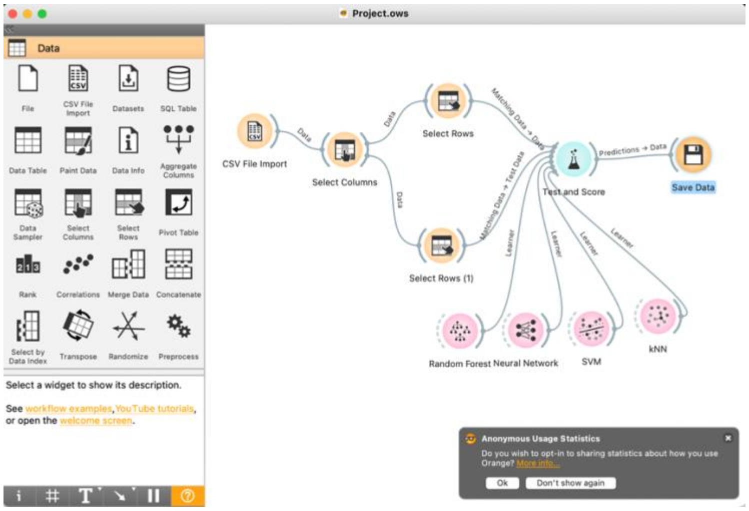
I have learnt that when copying the data from the table in Powerpoint slides to Excel, I need to need to remove last character in order to have numbers being recognizable by the Excel in order to plot graphs like clustered column bar charts.

I have also learnt that the axes of different charts should be scaled to be the same for fair comparison, and the minimum value of y-axis should be tuned for obvious contrast.

Through the preparation of all reports and materials for this project, I have learnt to cooperate, and share works with my groupmates. Without my groupmates, I think I cannot finish this project by myself. We shared the work well as we did what we are good at. They have trained and tested the models with different features prepared by using the Orange, prepared the Powerpoint slides for presentation and the written report. I have learnt from how they worked on their works as well.

## Individual Report – 21014947G Yu Hin Chung Nikko

I have done the design of the model. The model is a direct inspiration from the laboratory exercises. The figure is as shown. Some parameters in the CSV file report are inspired by the research paper using ANN and random forest such as moving average. The actual implementation is done by my teammates.



Select a widget to show its description.

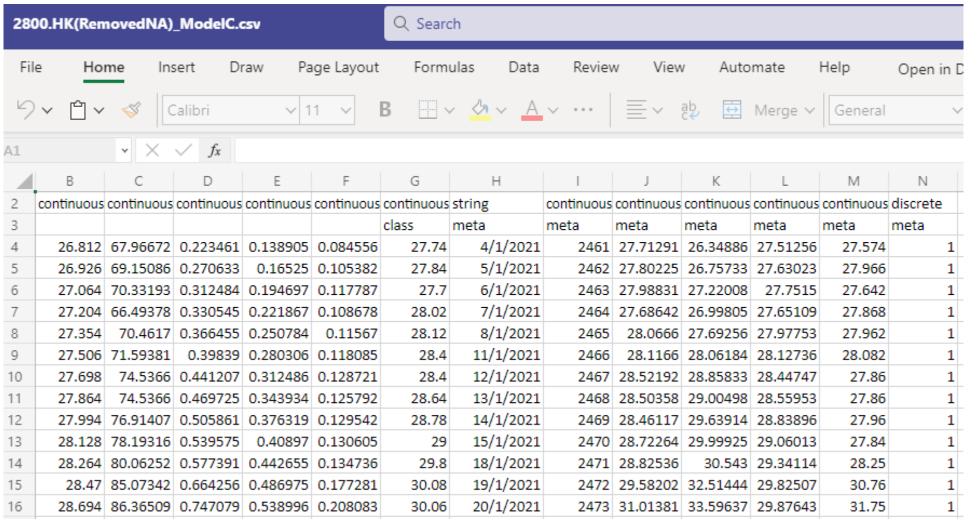
See workflow examples, YouTube tutorials, or open the welcome screen.

Anonymous Usage Statistics

Do you wish to opt-in to sharing statistics about how you use Orange? More info...

Fig. 31 The model

Meanwhile, I have also assisted the design of the features set, such as suggesting feature set C. Putting the csv files and generating the errors and metrics for feature sets A-D.



| | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | continuous | continuous | continuous | continuous | continuous | continuous | string | continuous | continuous | continuous | continuous | continuous | discrete |
| 3 | | | | | | class | meta | meta | meta | meta | meta | meta | meta |
| 4 | 26.812 | 67.96672 | 0.223461 | 0.138905 | 0.084556 | 27.74 | 4/1/2021 | 2461 | 27.71291 | 26.34886 | 27.51256 | 27.574 | 1 |
| 5 | 26.926 | 69.15086 | 0.270633 | 0.16525 | 0.105382 | 27.84 | 5/1/2021 | 2462 | 27.80225 | 26.75733 | 27.63023 | 27.966 | 1 |
| 6 | 27.064 | 70.33193 | 0.312484 | 0.194697 | 0.117787 | 27.7 | 6/1/2021 | 2463 | 27.98831 | 27.22008 | 27.7515 | 27.642 | 1 |
| 7 | 27.204 | 66.49378 | 0.330545 | 0.221867 | 0.108678 | 28.02 | 7/1/2021 | 2464 | 27.68642 | 26.99805 | 27.65109 | 27.868 | 1 |
| 8 | 27.354 | 70.4617 | 0.366455 | 0.250784 | 0.11567 | 28.12 | 8/1/2021 | 2465 | 28.0666 | 27.69256 | 27.97753 | 27.962 | 1 |
| 9 | 27.506 | 71.59381 | 0.39839 | 0.280306 | 0.118085 | 28.4 | 11/1/2021 | 2466 | 28.1166 | 28.06184 | 28.12736 | 28.082 | 1 |
| 10 | 27.698 | 74.5366 | 0.441207 | 0.312486 | 0.128721 | 28.4 | 12/1/2021 | 2467 | 28.52192 | 28.85833 | 28.44747 | 27.86 | 1 |
| 11 | 27.864 | 74.5366 | 0.469725 | 0.343934 | 0.125792 | 28.64 | 13/1/2021 | 2468 | 28.50358 | 29.00498 | 28.55953 | 27.86 | 1 |
| 12 | 27.994 | 76.91407 | 0.505861 | 0.376319 | 0.129542 | 28.78 | 14/1/2021 | 2469 | 28.46117 | 29.63914 | 28.83896 | 27.96 | 1 |
| 13 | 28.128 | 78.19316 | 0.539575 | 0.40897 | 0.130605 | 29 | 15/1/2021 | 2470 | 28.72264 | 29.99925 | 29.06013 | 27.84 | 1 |
| 14 | 28.264 | 80.06252 | 0.577391 | 0.442655 | 0.134736 | 29.8 | 18/1/2021 | 2471 | 28.82536 | 30.543 | 29.34114 | 28.25 | 1 |
| 15 | 28.47 | 85.07342 | 0.664256 | 0.486975 | 0.177281 | 30.08 | 19/1/2021 | 2472 | 29.58202 | 32.51444 | 29.82507 | 30.76 | 1 |
| 16 | 28.694 | 86.36509 | 0.747079 | 0.538996 | 0.208083 | 30.06 | 20/1/2021 | 2473 | 31.01381 | 33.59637 | 29.87643 | 31.75 | 1 |

Fig. 32 One of my work in generating the result csv and performing MAE metrics for model C.

Apart from my individual work, I have also helped verifying other teammate’s model (csvs) such as the curves, the formulas and more. Teammates and I had suggested different strategies during the brainstorming, and we agreed upon making strategies 1-5B.