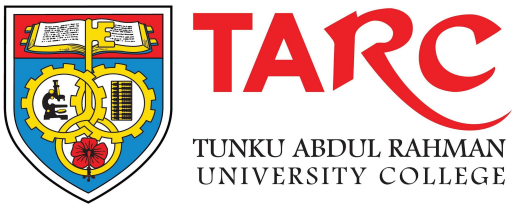
Stock Prediction Using Univariate Long Short-Term Memory Model

By

Pang Zheng Beng



FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

TUNKU ABDUL RAHMAN UNIVERSITY COLLEGE
KUALA LUMPUR

ACADEMIC YEAR
2021/2022

Stock Prediction Using Univariate Long Short-Term Memory Model

By

PANG ZHENG BENG

Supervisor : CHIN WAN YOKE

A project report submitted to the
Faculty of Computing and Information Technology
in partial fulfillment of the requirement for the
Bachelor of Science (Hons.)
Management Mathematics with Computing
Tunku Abdul Rahman University College

Department of Mathematical and Data Science
Faculty of Computing and Information Technology
Tunku Abdul Rahman University College
Kuala Lumpur

2021/2022

DECLARATION

The project report submitted herewith is a result of my own efforts in totality and in every aspect of the works. All information that has been obtained from other sources had been fully acknowledged. I understand that any plagiarism, cheating or collusion of any sorts constitutes a breach of Tunku Abdul Rahman University College rules and regulations and would be subjected to disciplinary actions.

Signature : Pang Zheng Beng

Name : Pang Zheng Beng

ID No. : 1901241

Date : 16-12-2021

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"STOCK PREDICTION USING UNIVARIATE LONG SHORT-TERM MEMORY MODEL"** was prepared by **Pang Zheng Beng** and has met the required standard for submission in partial fulfillment of the requirements for the award of Bachelor of Science (Hons.) Management Mathematics with Computing at Tunku Abdul Rahman University College.

Approved by,

Signature: _____ **Signature:** _____

Supervisor: _____ **Moderator:** _____

Date: _____ **Date:** _____

Specially dedicated to
my beloved grandmother, mother and father
(this dedication page is optional)

ACKNOWLEDGEMENTS

I would like to express my gratitude to my research supervisor, Dr Chin Wan Yoke for her invaluable advice, guidance and her enormous patience throughout the development of the research.

In addition, I would also like to express my gratitude to my loving parent and friends who had helped and given me encouragement.

ABSTRACT

The stock market is a chaotic place for prediction as there are plenty of factors that will affect the stock market simultaneously. Numerous studies have been conducted regarding to this field and the accuracy of the proposed methods have been improved. The stock price data have the characteristics of time series. To address this challenge, this project proposes long short-term memory (LSTM), which has the advantages of integrating relationships among time series data through its memory function, to predict the stock price. Adjusted closing price as the historical data from January 1, 2012, to November 3, 2021, including 2477 trading days is the dataset of this project. First, this project split 95% of the dataset into training set, then apply sliding window method on x train set before fitting into the model. And then, we adopt LSTM to predict the stock price with the extracted feature data. Root mean square error (RMSE) and coefficient of determination (R^2) are used to evaluate the model. According to the experimental results, the univariate LSTM can provide a reliable stock price forecasting with a high prediction accuracy of 97.27%. This forecasting method not only provides a new research idea for stock price forecasting but also provides practical experience for scholars to study financial time series data.

Keywords: LSTM, Machine Learning, Stock prediction

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction and definition

What is stock market? Stock market is a free market where shares of the public listed companies are traded. Two investors are involved in trading so stock market also known as Secondary Market (Sharma and Kaushik, 2017). A share is a unit of company's capital which is divided equally into a finite number and can be purchased by investors. At the most fundamental level, stock price in the market is determined by supply and demand. The market value also known as market capitalization of a publicly traded company is calculated by multiplying the number of its outstanding shares by the current share price.

1.2 Background

Stock market provides regulated, secure environment for companies to be traded publicly and raise capital. Companies raise capital from stock market by selling shares of ownership of the company. The addition of capital enables companies to repay its depositors and maintain financially stable in the short term. Capital also enables companies to develop their business, extend their operation. Stock market promotes investment for economic growth and for investors, a way to invest money. In fact, the stock market is considered as a primary indicator of a country's economic strength and potential. The performance of stock market shows the health of overall economy. Obviously, the performance of stock market and the health of economy is aligned. Besides, since stock market is a free market, the performance of stock market shows people financial behaviors.

1.3 Problem of statement

Stock market prediction's goal is to determine future movement of the stock value of a financial exchange. Prediction of stock market is one of the most challenging issues in the world. Due to many factors that involved in such as politics, interest rate, economic growth and current event affect the stock market, that makes the stock market unstable and hard to predict. Stock market is hard to predict by using the historical data. The past result cannot be used as an indicator of future success. Anything can happen in the world, the black swan events which is unpredictable, beyond most expectation like 2008financial crisis and covid19 pandemic, which makes the prediction of stock market become inaccurate. Besides, the prediction of stock market is mainly conducting by human brain in the past. Emotion, but not practical reasons may drive people choices when they are making financial decision. The main emotion that investors respond to is the fear of financial loss and excitement for potential profits. Thus, investors may vulnerable to errors in judgement.

Next, in order to conduct a research project on a stock by using fundamental and technical analysis, it is highly depending on the ability of the analyst where the analyst must possess proficient skill in finance, macroeconomics and etc. Furthermore, based on statistics and probability theory, some scholars use time series linear forecasting model to predict the short-term stock price with a large number of long-term data, such as vector autoregression (VAR) (Jung et al, 1996), Bayesian vector autoregression (BVAR) (Bessler et al, 2008) model, and autoregressive integrated moving average mode (ARIMA) (Ariyo et al, 2014). However, the accuracy of using time series model alone is questioned due to the uncertainty and high noise characteristics of financial time series and the relationship between the study variables are prone to dynamic changes over time, which limits its further application and expansion.

1.4 Aims and objectives

The objectives of this study are:

1. Using univariate long short-term memory model to predict future stock price accurately without finance knowledge.
2. To study the relationship between the independent variables (input data) and the dependent variables (output data).
3. To utilize the computational power of LSTM in forecasting stock price.

1.5 Scope of study

With the large number of the stock price data and the relationships among stock price data, there has been a corresponding increase in the difficulty of the prediction of future stock price. In view of this situation, this study analyses the characteristics of the stock price data. To this end, this project will use long short-term memory (LSTM) model to predict future stock price. This study is focused on the adjusted closing price of the companies such as Apple Inc, Tesla Inc, Alphabet Inc, and Netflix Inc. In the report, they are represented by an abbreviation

AAPL, TSLA, GOOG, and NFLX, respectively. Further, the study also involves an analysis of the volatility of the four stocks.

1.6 Significance of study

Stock market is a chaos and complex system. An accurate prediction of stock price movement is essential for investors to make more profits. Investors can avoid big loss in stock market. Accurate prediction of stock market is a meaningful and vital issue for both research purpose and government financial planning.

CHAPTER 2

LITERATURE REVIEW

2.1 Theories in finance

2.1.1 Efficient Market Hypothesis

There are various theories in financial economics. Efficient Market Hypothesis (EMH) is a good starting theory in the modern theory of finance. The theory states that the share prices reflect all available information. The term "efficiency" denotes that investor have no opportunity of gaining abnormal profits from the market. It is impossible to beat the market. A efficient market is defined if a market with great number of rational, profit-maximizers actively competing, with each trying to predict future market values of individual securities, and where current important market information is almost freely available to all investors (Fama, 1970).

According to Fama (1970), the weak EMH assumes the current prices of financial assets incorporate, regardless of hour, all the existing historical financial information. As a result, investors cannot gain unusual profits from investing in these financial assets, such as stock. The weak EMH form indicates that prices will show random walk.

The semi-strong EMH was introduced as the state of fact in which the price of financial assets reflects at any moment, all the information publicly existing on market, including historical price and other historical information. In addition, the price of financial asset changes rapidly and without biases to any new public information existing on the market. In other word, semi-strong incorporate the weak form of EMH. Neither technical or fundamental analysis can find out how an investor should manage his funds in order to obtain higher profit than investment in a random portfolio of financial assets, Ţiţan (2015).

Compared to weak EMH and semi-strong EMH, the strong form of EMH assumes that the share price mirrors all the public and private information on a market, which includes historical data, all public information, and all private information regarding the financial asset Fama (1969 and 1970).

Basic logic for modern risk-based theories of financial assets price, and frameworks such as consumption-based asset pricing and intermediary asset pricing can be considered as the combination of a model of risk with the EMH (Fama, 2014). Many decades of studies were elaborated to justify all the three types of EMH. Many researches invalidated the semi-strong EMH and the strong EMH, which are supported by financial data, while there is disagreement about weak form of EMH among researchers. EMH is a controversial and very controversial and of particular interest for financial economists, professors and researchers as confirmed by the large body of specialized literature, Ţiţan (2015).

2.1.2 Random Walk Hypothesis

The random walk hypothesis is a financial theory stating that price of securities in the stock market evolve according to a random walk. The hypothesis suggests that changes in stock price is independent of the movement in the price of another stock. It is still a reputable and challenging question if the financial data follows random walk. Some methods or approaches have been introduced to investigate whether the stock price follows a random walk. For

example, the variance ratio tests (Lo and MacKinlay, 1989), the Hurst exponent and surrogate data testing, (Nakamura and Small, 2007). As a result, any attempt tries to predict future price movement, either through fundamental or technical analysis is unsuccessful.

2.2 Different approaches to predict stock price

There are various studies in stock prediction. In the past, the techniques for stock price prediction include simple regression techniques, time series models and econometric tools. However, in the recent year, it has changed the trend modeling by using machine learning and deep learning.

2.2.1 Qualitative approaches

Despite Fama's hypothesis, the analysis of stock market has been studied repeatedly for the past. Forecasting the movement of stock market has always had a certain appeal for the analysts. No method has discovered to accurately forecast the stock price even there are numerous scientific approaches have been made. Several approaches have been introduced trying to forecast the stock price. Basically, the two most popular methods that investor use to analyze the stock price are fundamental analysis and technical analysis.

David Dodd and Benjamin Graham are considered as the creators of fundamental analysis, who first describe the rule of this method of forecasting in their Security Analysis, published in 1934 (Nazarowa 2014, p. 290). Fundamental analysis has developed numerous methods to determine intrinsic value of the share. Intrinsic value is the true value of the shares, the current price of the share may not equal to the intrinsic value but it's heading

towards over a longer period of time. The information related to macroeconomic time series, like Gross Domestics Product, interest rates, currency exchange rates, customer price index, are the most common used among others, (Boyacioglu & Avci, 2010).

Fundamental analysis is a persistently advancing science, as evidence by emergence of latest and latest valuation method (Borowski 2014, p. 10). As compared to technical analysis, fundamental analysis is fewer in number of literatures, because it is harder to build models that explain the fluctuation of stock price. Despite the numerical data, other information like financial news, brand power, management team and etc, are hard to quantify, because of its unstructured nature and non-continuous behaviour.

However, technical analysis has mostly relied when come to stock market analysis in short term period, especially among beginner investors, unlike fundamental analysis, technical analysis is better acknowledged, main reason is the ease of its use and high-level automatization, which does not require great learning in economic (Figurska and Wisniewski, 2016). Researchers of technical analysis claim that al new information, like financial news and macroeconomics variables, are already reflected in stock price (Bustos and Pomares-Quimbaya, 2020). Hafezi et al (2015) stated that technical analysis analyze data on market activity such as historical trends, price level and volumes to chart pattern in stock movement, while Krzywda (2010) claims fundamental analysis provides a mechanism for evaluating long-term forecasting of values of future circumstances based on historical data and a set of other factors which may involve. De Long et al. (1990) and Shleifer et al. (1997) have demonstrated that sentiment affects investment decisions. The stock returns were affected by the sentiments contained in financial reports or news articles (Schumaker, Zhang, Huang and Chen, 2012).

2.2.2 Traditional approaches

Autoregressive integrated moving average (ARIMA) models are known to be robust and efficient in financial time series forecasting especially short-term prediction. It has been extensively used in field of economics and finance. This model type is classified as ARIMA(p,d,q), where p denotes the autoregressive parts of the data set, d refers to integrated parts of the data set and q denotes moving average parts of the data set and p,d,q is all nonnegative integers. Ariyo et al (2014) presented the prediction of Nokia stock price by using ARIMA. In his research, ARIMA (2, 1, 0) is considered the best for Nokia stock pricedata.

Based on the paper from Mondal et al (2014), the accuracy of ARIMA model in predicting stock prices is above 85% for all the sectors, which indicates that ARIMA gives good accuracy of prediction. Mondal et al (2014) stated that from the result of hypothesis testing, the changes in the accuracy for different size of training datasets is not significant. Since the behaviour of stock prices cannot easily be captured, (Pai and Lin, 2005) proposed a hybrid model of ARIMA and the SVMs that has both linear and nonlinear modelling abilities for forecasting stock price.

2.2.3 Deep learning approaches

Definition of text mining is the computer automatically extracting information of new, previously unknown (Fan, Wallace, Rich and Zhang, 2006), a process to extract interesting and significant patterns to explore information from different written sources (Talib, Kashif, Ayesha and Fatima, 2016). The usage of text mining is to handle unstructured data or semi-structured data sets, such as email, full text documents, and HTML files (Fan, Wallace, Rich and Zhang, 2006). Event extraction is one of the subtasks of text mining, for the purpose of identification of associations among entities and other information in text.

The growth in web information has resulted in the recent works that has applied Natural Language Processing (NLP) techniques to explore financial news for market volatility prediction, (Ding et al, 2015). From the study of Ding et al (2014), the “actor” and “object” of an event can be captured, in order to structured event representations instead of words as

features, which made an improvement to stock market prediction. In the research of Ding et al (2015), they found that events are better for stock market prediction instead of words. Secondly, embedded event extraction models (EB-NN, EB-CNN) has consistently achieved better performance than the event extraction model (E-NN, E-CNN). Authors claimed that the problem of feature sparsity can be alleviated by low-dimensional dense vector.

Table 2.1 Comparison of models from Ding et al (2015)

Model	Accuracy	Matthews Correlation Coefficient
E-NN	58.94%	0.0711
WB-NN	60.25%	0.1649
WB-CNN	61.73%	0.2147
E-CNN	61.45%	0.2036
EB-NN	62.84%	0.3472
EB-CNN	65.08%	0.4357

- E-NN: structured events tuple (Ding et al, 2014) input and neural network model
- WB-NN: word embeddings input and neural network model (Ding et al, 2014)
- WB-CNN: word embeddings input and convolutional neural network model
- E-CNN: structured events tuple (Ding et al, 2014) input and convolutional neural network model
- EB-NN: event embeddings input and neural network prediction model (Ding et al., 2014)
- EB-CNN: event embeddings input and convolutional neural network prediction model

Public sentiment or mood states may play an equally important role though news mostly certainly influences stock market prices. It is believable to assume that the public sentiment can handle stock market values as much as news. OpinionFinder and GPOMS are used to measure variations in the public mood from tweets submitted. To avoid the ignorance of rich, multi-dimensional structure of human mood, unlike other researches, the tweets are labelled into 6 dimensional moods, namely Calm, Alert, Sure, Vital, Kind and Happy. Specific mood states of Twitter data able to forecast the Dow Jones Industrial Average (DJIA) value with 87.6% accuracy (Bollen, Mao and Zeng, 2011).

A study has been done by Deng et al (2018) using 18 million tweets relating to stocks, the objective of study is to determine whether the public sentiment has a relationship with stock market. The study shows that tweets do fluctuate with stock price. Authors noticed that the effect of negative sentiment is more crucial than positive sentiment. They concluded that 1% increase in negative tweets results in 0.03% drop in stock returns. Meanwhile, positive sentiment does not affect the stock return on daily basis as volatile as the negative sentiment does, however, in the long term, it does have repercussion over the stock.

Furthermore, there are numerous researchers regarding the impact of information from social media on stock price. Based on the research conducted by Yahyu Eru Cakra and Bayu Distiawan Trisedya, it stated that sentiment analysis model which using Random Forest Tree algorithm can classify Tweets into three classes (positive, negative, neutral) with 60.39% accuracy and the Naïve Bayes with the second highest accuracy which is 56.50%. Authors were using Twitter data in Indonesian to forecast stock price of companies from Indonesia, since the previous researches used the Tweets in English, the difference between authors' research and previous researches may shows that sentiment analysis can be performed in different language datasets. In particular, this paper exploits linear regression to perform price fluctuation prediction, margin percentage prediction and stock price prediction.

Convolutional neural network (CNN) is a class of artificial neural network, which were inspired by the mechanism of the biological vision system. Most commonly, it applied to analyze imagery. CNN can extract features from the input data. Referring to the study of Cao and Wang (2019), 5 stock indices (HIS, TSEC, DAX, NASDAQ and S&P500) are selected to train and test the models. Among the CNN-SVM, CNN, SVM, BP network, CNN-SVM and CNN can follow the trend best as one and the error series of the forecasting models fluctuate nears zero, follow by SVM and Back Propagation network. However, the same model has a far different fitting effect on different datasets, which is referred to the characteristics of the datasets itself. Datasets was provided by Yahoo Finance, not real time data, hence there are certain difficulty in practical application. Authors believed that the continuous addition of

data granularity and data size can improve the model, in order to better meet the actual application (Cao and Wang, 2019).

Table 2.2 Evaluation of models from Cao and Wang (2019)

Model	RMSE	Correlation coefficient	Determination coefficient
CNN	4.5459e − 04	0.9323	1.2009

Lu et al in 2020 stated that Long Short-Term Memory (LSTM) is a neural network model proposed by Schmidhuber et al in 1997. LSTM is used to solve the constant problems of gradient disappearance and explosion in Recurrent neural network (RNN) (Ta, Liu and Tadesse, 2020). Long Short-Term Memory (LSTM) consists of a cell state and three gates, therefore it has the ability of selectively learning information among units.

CNN-LSTM is a new deep learning approach to forecast the stock price. The time feature of data is extracted by CNN and LSTM is used to forecast data. The hybrid model can fully utilize the time sequence of stock price data to obtain more dependable forecasting. Authors compare the evaluation of CNN-LSTM with MLP, CNN, RNN, LSTM, and CNN-RNN. The obtained result show that the CNN-LSTM has the lowest mean absolute error (MAE) and root mean square error (RMSE) among all the methods. Also, the R^2 of CNN-LSTM is the highest among the six forecasting models. The paper concludes that CNN-LSTM is suitable for the prediction of stock price however it also stated that the ignorance of emotional factors into the prediction and suggest the increasing of sentiment analysis, in order to ensure the accuracy of stock prediction (Lu et al., 2020).

Table 2.3 Comparison of models from Lu et al (2020)

Model	MAE	RMSE	Determination coefficient
MLP	37.584	49.799	0.9442
CNN	30.138	42.967	0.9585
RNN	29.916	42.957	0.9593
LSTM	28.712	41.003	0.9622
CNN-RNN	28.285	40.538	0.9630

CNN-LSTM	27.564	39.688	0.9646

The ability of memorizing sequence of data makes the LSTM a special kind of RNNs. Moghar et al (2020) used different epochs (12 epochs, 25 epochs, 50 epochs and 100 epochs). for training data. The observation from the paper is that training with less data and more epochs can improve the testing result.

Chen et al 2015 studied a LSTM model to predict China's stock market in Shanghai and Shenzhen Exchanges (SSE). A single input layer, followed by multiple LSTM layers, a dense layer and a single output layer with several neurons are contained in his paper. Multiple stock features such as high price, low price, open price, close price, were experimented with six different methods to predict stock prices. This study indicated that the normalized features and SSE indices could increase the accuracy of forecasting.

Althelaya et al (2018) evaluated and compared the bidirectional LSTM (BLSTM) and stacked LSTM (SLSTM) models in stock price prediction. The paper used the closing price of Standard and Poor 500 index (S&P 500) data as the input dataset. The paper concluded that BLSTM model has good performance in both long-term and short-term prediction. However, SLSTM model performed well in short-term prediction only.

2.3 Summary

ARIMA able to predict the future stock price accurately in short-term prediction. However, the random walk behaviour of stock price in long-term will affect the accuracy of ARIMA in long-term. In this project, univariate long short-term memory is presented since LSTM has the ability to study the relationship among the time series data. Besides, high computational cost is the weakness of hybrid models. Market sentiment will highly affect the prediction of stock price. However, the lack of data source is one of the main barriers, hence only the stock price data will be considered.

CHAPTER 3

METHODOLOGY

3.1 Datasets

In this project, 4 stock prices will be studied. Besides, Standard and Poor's 500 Index (S&P 500) is the benchmark index of this study. Data from 01-01-2012 to 03-11-2021 will be selected. Using DataReader library to obtain the stock price data form Yahoo finance. Source of data is retrieved from https://finance.yahoo.com/. The data consists of 8 columns: Date, High, Low, Open, Close, Volume, Adj Close, and Company name.

Table 3.1 Description of the columns of the dataset

Column	Description
Date	a day that all order is executed in the market
High	maximum prices in a given day
Low	minimum prices in a given day
Open	prices at which a stock began trading in a day
Close	prices at which a stock ended trading in a day
Adj close	closing price reflects that stock's value after accounting for any corporate actions
Company name	a unique series of letters assigned to a security for trading purposes

This project only considers the ‘Adj Close’ column of the dataset. The closing price of a stock is much important than high or low-price levels. Many stock traders use close price in technical analysis. First, many analysts define the support and resistance level of a stock by using the closing price of a stock. Second, the closing price provides analysts a validation function. Closing price can validate the further price movement of the stock. Hence, the ‘adj close’ which indicates the real closing price of the stock is considered.

Table 3.2 Details of the dataset

Stock symbol	Description	Source
AAPL	Apple Inc	NASDAQ
TSLA	Tesla Inc	NASDAQ
NFLX	Netflix Inc	NASDAQ
GOOGL	Alphabet Inc	NASDAQ

3.2 Data Preprocessing

Time series is a series of data points indexed in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus, it is a sequence of discrete-time data. There is some potential for correlation between the sequences of data. In this report, the closing price of a stock is collected in each trading day, hence, the daily closing price of a stock is a time series data.

1. The first 95% of the dataset will be split into training dataset for fitting. The remaining 5% of the dataset will be used as the testing dataset for evaluating.
2. X_train - This includes the sliding window of data of closing price and these will be used to train the model.

3. X_test - This is the remaining portion of data of closing price which will not be used in the training phase and will be used to make predictions to test the accuracy of the model.
4. y_train - This is the data of closing price which needs to be predicted by this model.
5. y_test - This data has category labels for the test data, these labels will be used to test the accuracy between actual and predicted categories.

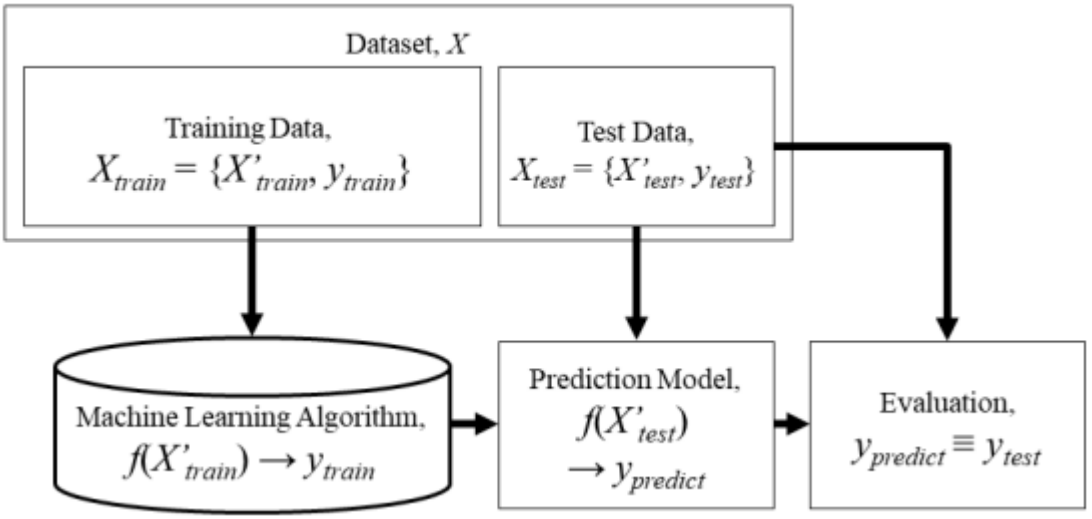


Figure 3.1 Flowchart of the model

Table 3.3 how the dataset is split into 4 portions.

Subset	Description
x_train	row 1-2353 will be selected and row 1-60, 2-61, 3-62 until row 2294-2353 will be grouped in 1
y_train	row 61 – 2354
x_test	row 2295 – 2476
y_test	row 2295 - 2477

The figure below shows that the data of closing price from row 1 to row 2354 will be rearranged into equal-size, small batches, which is the sliding window method. The window size is equal to 60, hence the size of batches is equal to 60.

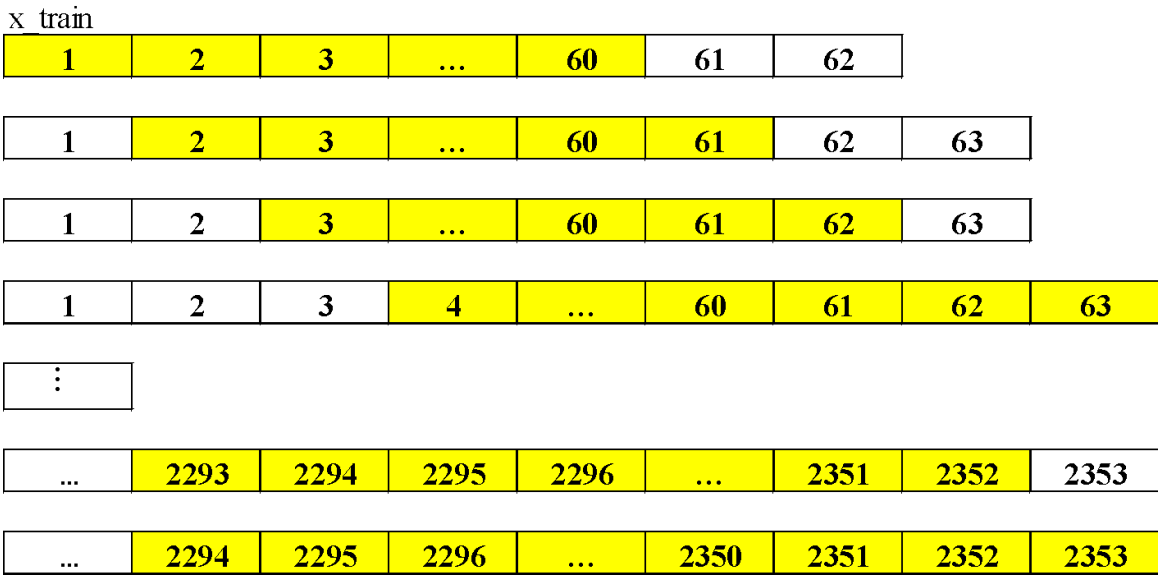


Figure 3.2 Sliding window framework of x_train

3.3 Daily return of a stock

The daily return of a stock is the percentage change of closing price between two consecutive days. A stock with lower positive or negative daily returns is commonly less risky than a stock with higher daily returns, which create larger swings in value.

$$r_i = \frac{p_i - p_{i-1}}{p_i} \tag{3.1}$$

where

p_i = closing price of the stock in current date

p_{i-1} = closing price of the stock of previous date

3.4 Mathematical framework of the model

The figure 3.3 below describes the structure of the univariate Long Short-Term Memory.

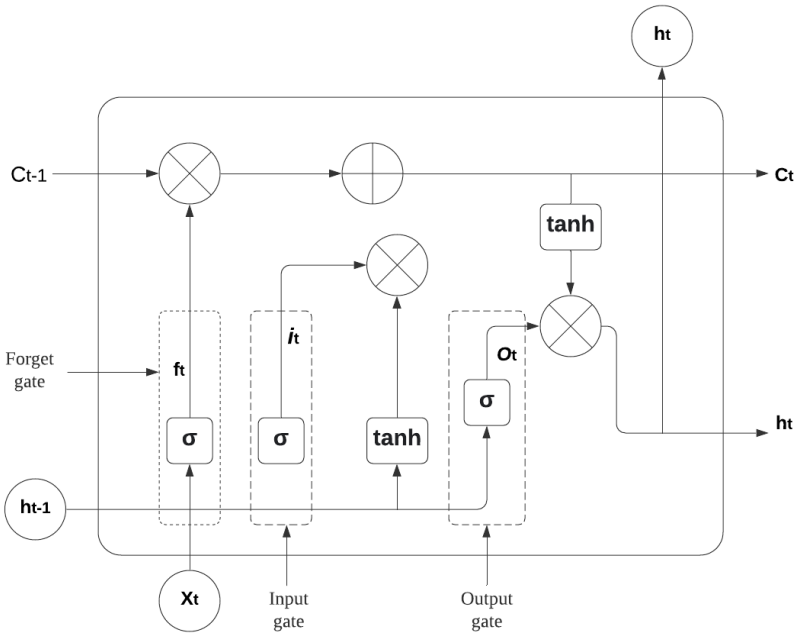


Figure 3.3 Structure of Long Short-Term Memory

When a new input come through, and the system wants to forget the property of the old subject:

$$f_{t=\sigma\left(W_f\cdot\left[h_{t-1},x_t\right]+b_f\right)}$$

(3.2)

The next step is to decide what new information we're going to store in the cell state. This has two parts. First, a sigmoid layer called the "input gate layer" decides which values we'll update. Next, a $tanh$ layer creates a vector of new candidate values, $\widetilde{C}_t$, that could be added to the state. In the next step, the system will combine these two to create an update to the state.

In this model, system wants to add the property of the new input to the cell state, to replace the old one system are forgetting.

$$i_{t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i)}$$

(3.3)

$$\widetilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

(3.4)

It's now time to update the old cell state, C_{t-1}, into the new cell state C_t. The previous steps already decided what to do.

Then, multiplying the old state by f_t, forgetting the things system decided to forget earlier. Then adding $i_t \times \widetilde{C}_t$. This is the new candidate values, scaled by how much system decided to update each state value.

In this model, this is where system actually drop the information about the old input's properties and add the new information, as system decided in the previous steps.

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t$$

(3.5)

Finally, system needs to decide what we're going to output. This output will be based on the cell state, but will be a filtered version. First, running a sigmoid layer which decides what parts of the cell state system are going to output. Then, putting the cell state through $tanh$ (to push the values to be between −1and 1) and multiply it by the output of the sigmoid gate, so that only the parts system decided to outputs.

3.5 Parameter setting of Univariate LSTM

The model of this project will use 2 LSTM layers and 2 dense layers. The batch size and the number of epochs are set to (8,10), (8, 50), (32, 10), (32,50) accordingly.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 60, 128)           66560
_________________________________________________________________
lstm_1 (LSTM)                (None, 64)                49408
_________________________________________________________________
dense (Dense)                (None, 25)                1625
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 26
=================================================================
Total params: 117,619
Trainable params: 117,619
Non-trainable params: 0
_________________________________________________________________
```

Figure 3.4 Summary of the univariate LSTM model

3.6 Evaluation of the model

In order to evaluate the forecasting result of the LSTM model, the mean square error (MSE), root mean square error (RMSE), and *R*-square (R^2) are used as the evaluation criteria of the methods. The MSE calculation formula is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \quad (3.6)$$

where

$\hat{y}_i$ = predictive value

y_i = the true value

n = total number of values

The mean square error (MSE) represents the difference between the original and predicted values extracted by averaged the absolute difference over the dataset. The smaller the value of MSE, the better the result of forecasting. The RMSE calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}$$

(3.7)

where

$\hat{y}_i$ = predictive value

y_i = the true value

n = total number of values

The smaller the value of RMSE, the better the forecasting.

The coefficient of determination, R^2, is the proportion of the variation in the dependent variable that is predictable from the independent variable(s). The value range of R^2 is (0,1).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2}$$

(3.8)

where

$\hat{y}_i$ = predictive value

y_i = the true value

$\bar{y}_i$ = average value

The closer the value of MAE and RMSE to 0, the smaller the error between the predicted value and the real value, the higher the forecasting accuracy. The closer R^2 is to 1, the better the fitting degree of the model is.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Data analysis

4.1.1 Trend of the stock

From Figure 4.1 below, all the stocks have an uptrend characteristic in common. A trend is a continually increase or decrease in the series over time. The importance of identifying the trend in time series data is to enhance the model performance.

Figure 4.1 Historical price of stock (a) AAPL, (b) GOOG, (c) NFLX, (d) TSLA

4.1.2 Sales volume of stocks

Sales volume of a stock is the total number of shares being traded during a given time period. Volume represents the overall activity of a stock or a whole market, it shows the demand and supply of a stock. From Figure 4.2, GOOG, NFLX, and TSLA have significant huge trading volume in certain time period. Normally, the unusual trading volume cause the price of security rise or fall sharply. Therefore, the result of prediction will be affected.

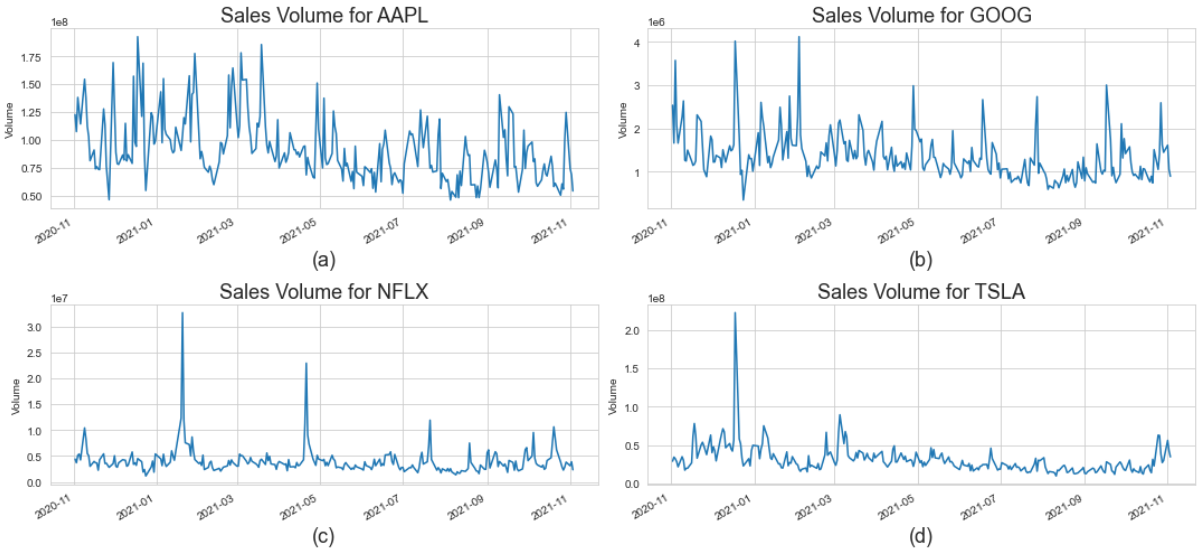


Figure 4.2 Sales volume of stock (a) AAPL, (b) GOOG, (c) NFLX, (d) TSLA

4.1.3 Daily return of stock

From Figure 4.3 below, the daily return of stocks was ranked in ascending order, AAPL< GOOG < NFLX < TSLA. The daily return calculation formula is as follows:

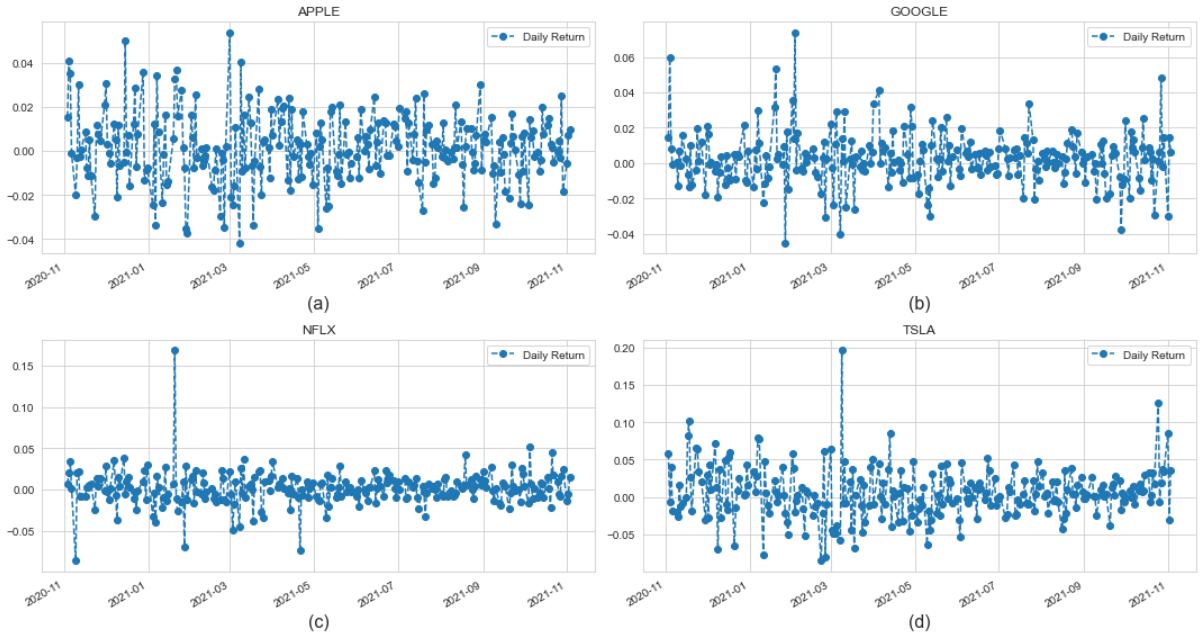


Figure 4.3 Daily return of stock (a) AAPL, (b) GOOG, (c) NFLX, (d) TSLA

4.1.4 Correlation of daily return of stock

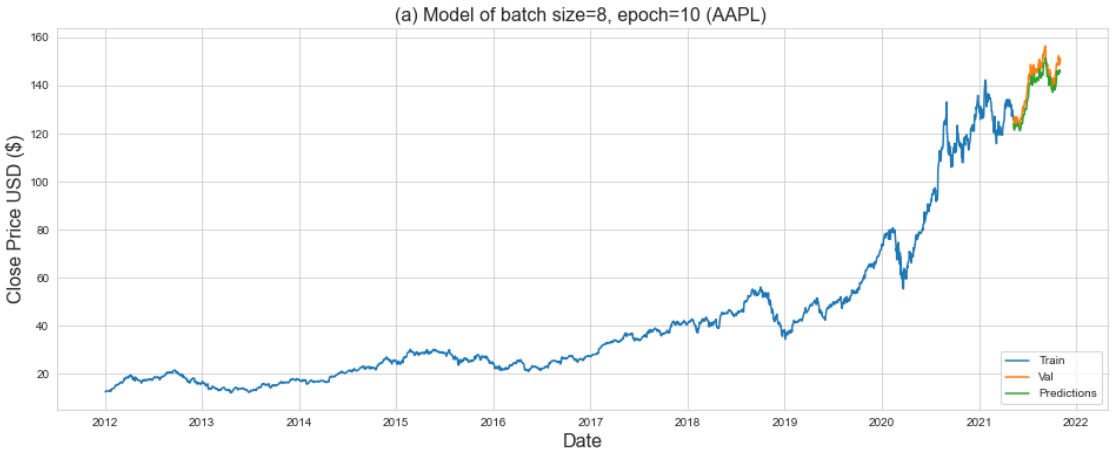
The correlation of daily return among stocks is the degree of level of co-movement of daily return between two stocks. The importance of interpreting correlation of daily return among stocks is to help investors build diversified portfolios, even correlation coefficients have no real predictive power beyond that.

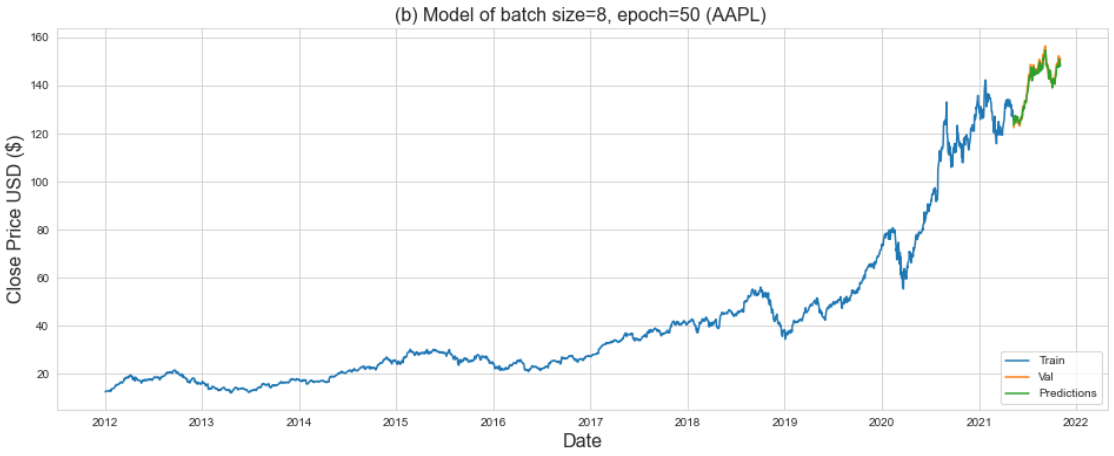
From Figure 4.4 below, all the stocks have positive correlation of daily return with each other, in other word, the daily rise or fall in price of a stock, other stocks will have the same direction of movement. It is interesting that all technology companies are positively correlated. Secondly, GOOG and AAPL have the strongest correlation of daily return. The correlation of daily return of TSLA between AAPL, GOOG, NFLX are the lowest. The correlation of daily return among AAPL, NFLX, and GOOG are the highest.

Figure 4.4 Correlation of daily return stock AAPL, GOOG, NFLX, and TSLA

4.2 Prediction of AAPL

The Figure 4.5 below show a whole picture of prediction and the real closing price. From the figure, it shows the impact of prediction in a long timeframe. The results from all 4 models with different hyperparameters are different. In a longer timeframe, all the 4 models can predict the trend of AAPL.





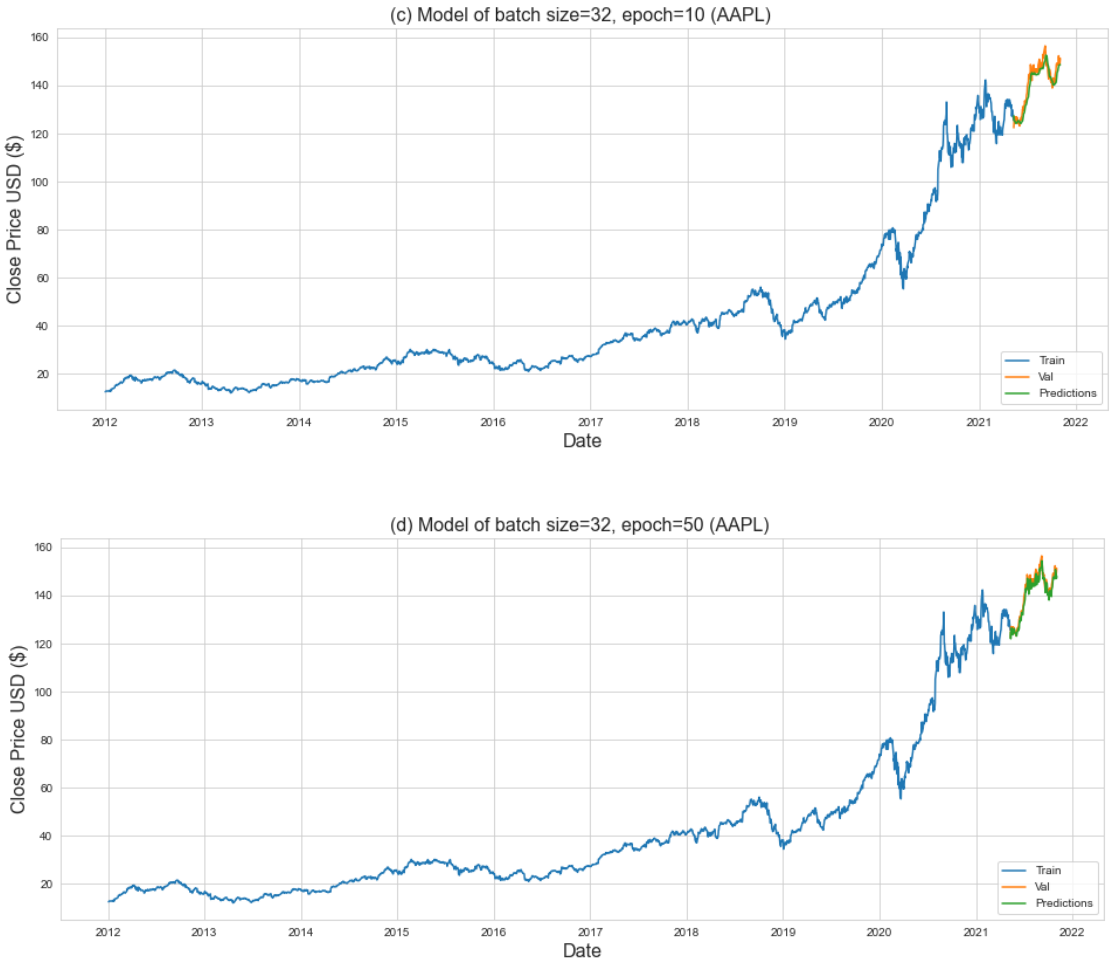


Figure 4.5 Overall of prediction of (AAPL) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

The Figure 4.6 below illustrates the real vs predicted value of models with 4 different hyperparameters. The coefficient of determination of model (a), (b), (c), and (d) are 0.80145651, 0.960402897, 0.887283835, and 0.940150202. The performance of model, in ascending order, are (a) < (c) < (d) < (b). The closer the real value curve and the predicted value curve, the higher the coefficient of determination, the better the performance of the model. The coefficient of determination of model (a) is the lowest, since the gap between real and predicted value is largest. The two curves of real and predicted value almost overlap, therefore model (b) has the highest coefficient of determination.

Figure 4.6 Real vs prediction of (AAPL) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

From Figure 4.7 below, there is some similarity in model (a) and model (c), the rate of decline of training loss are relatively lower than model (b) and model (d). The model (b) and (d) have higher number of epochs as compared to model (a) and (c). Therefore, model (b) and (d) can obtain low training loss quickly.

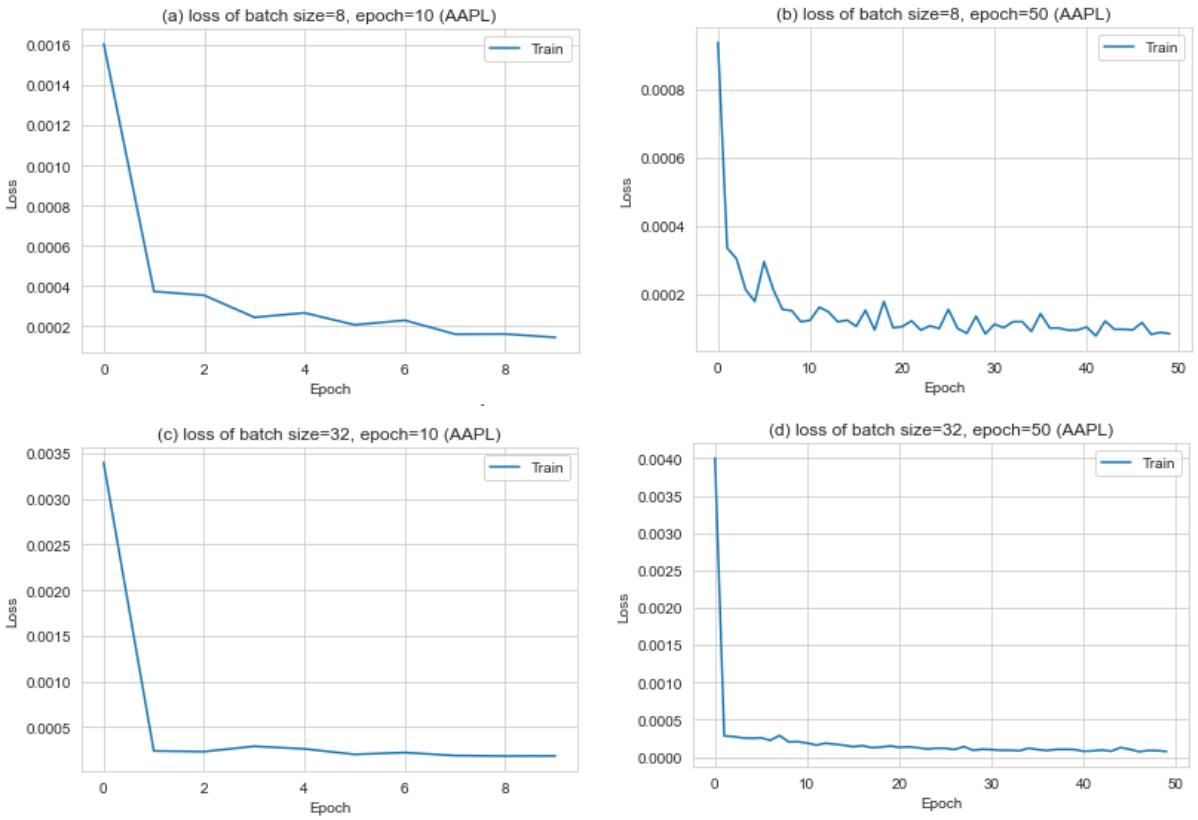
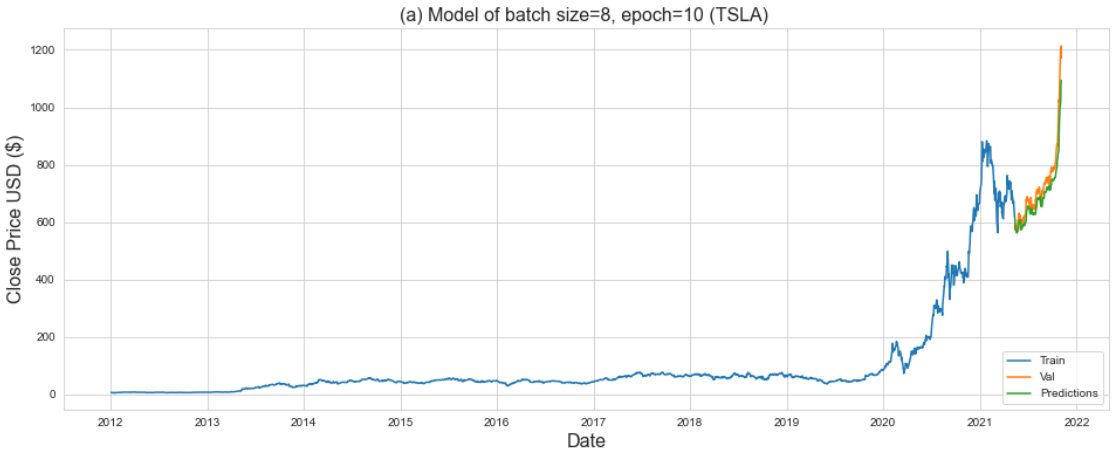
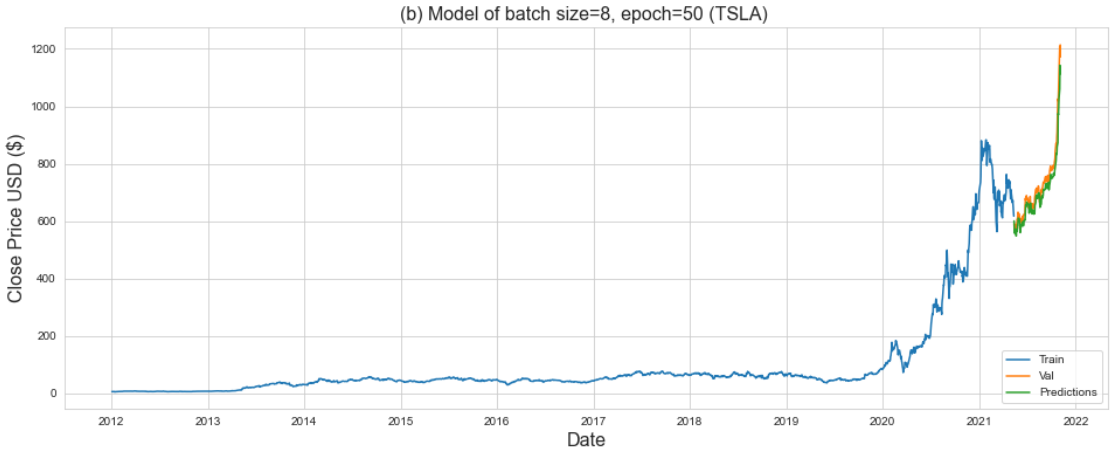


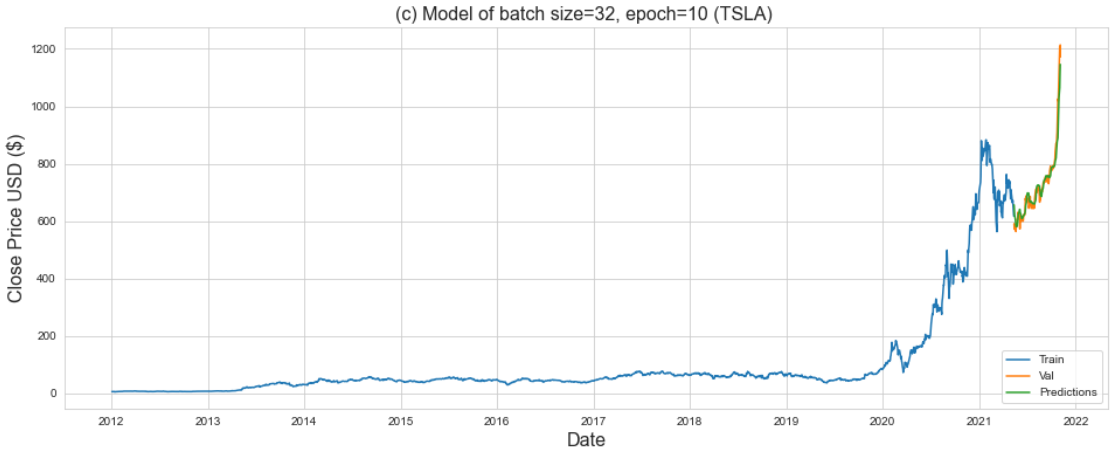
Figure 4.7 Training loss of (AAPL) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

4.3 Prediction of TSLA

The Figure 4.8 below shows the overall of prediction in a long-term period. In a long-term period, the movement of share price are far more important than the actual value. In short, from the figure, the predicted value is not too far apart from the actual value.


(a) Model of batch size=8, epoch=10 (TSLA)
Close Price USD ($)
Date
Train
Val
Predictions


(b) Model of batch size=8, epoch=50 (TSLA)
Close Price USD ($)
Date
Train
Val
Predictions


(c) Model of batch size=32, epoch=10 (TSLA)
Close Price USD ($)
Date
Train
Val
Predictions

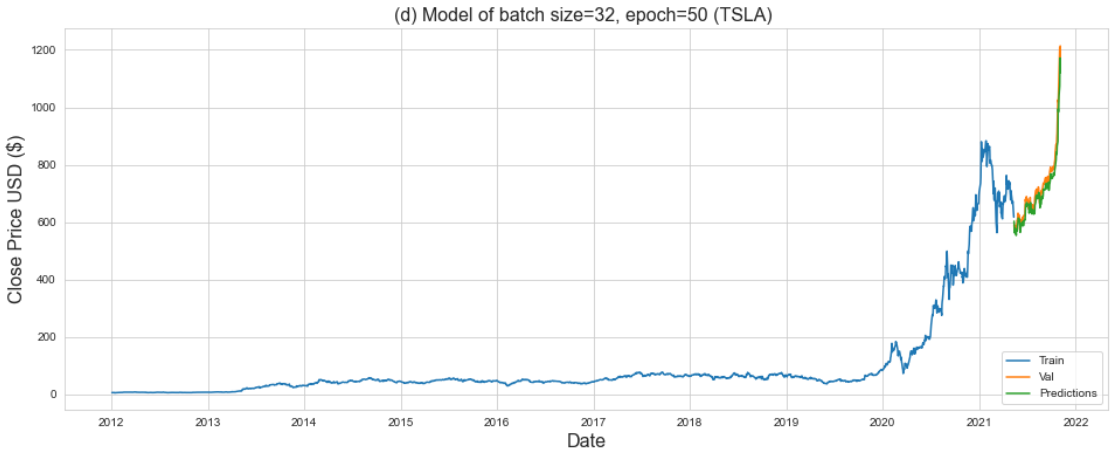


Figure 4.8 Overall of prediction of (TSLA) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

The Figure 4.9 shows the real vs predicted value of model (a), (b), (c), and (d). The coefficient of determination of model (a), (b), (c), and (d) is 0.864074544, 0.913996903, 0.934175634, and 0.931187336. The performance of the models, in ascending order, is model (a) < model (b) < model (d) < model (c). In model (a) 86.4% of the variation is explained by the model.

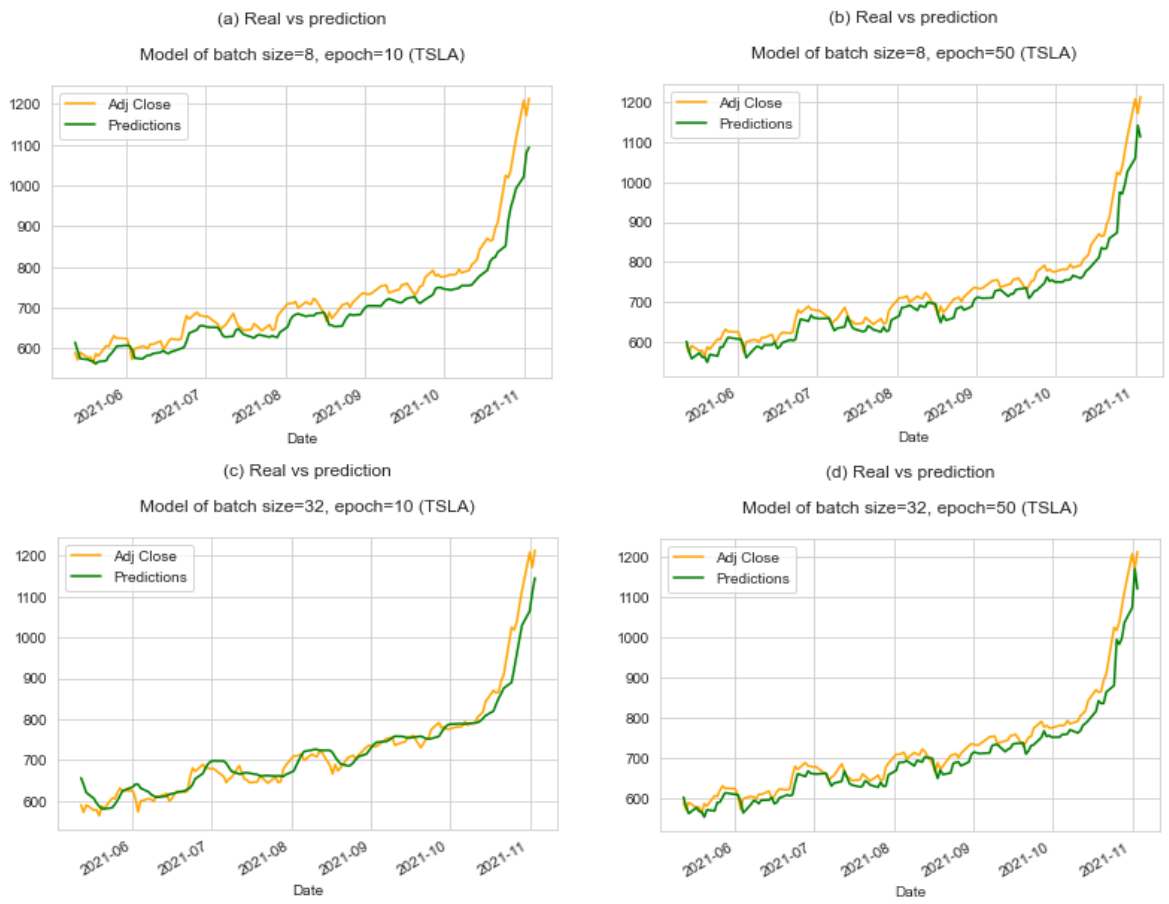


Figure 4.9 Real vs predicted of (TSLA) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

The Figure 4.10 below shows that the training loss of model (b) almost becomes static, barely improve when the training loss is around 0.0001. With the same number of epochs, which is 50, the training loss of model (d) decreases slowly when the training loss is around 0.00025. The larger batch size of model (d) means the model makes larger gradient updates and leads to a relatively low accuracy.

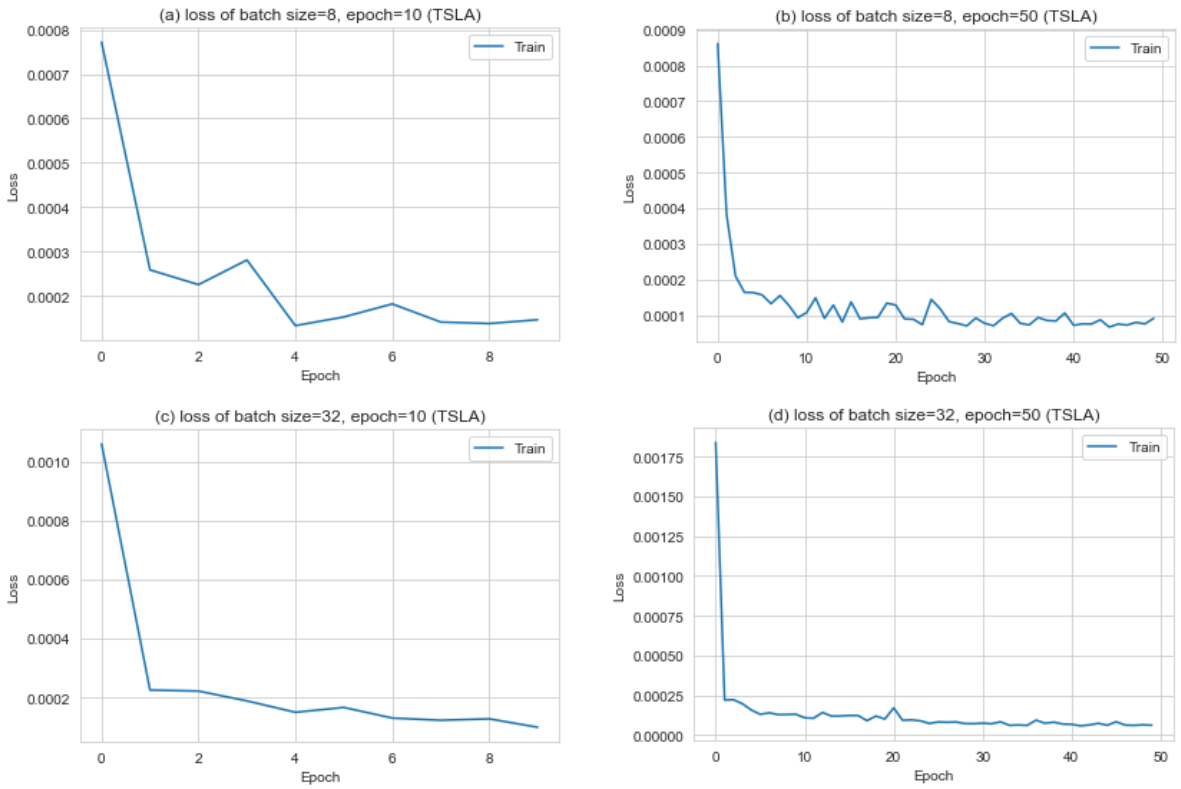


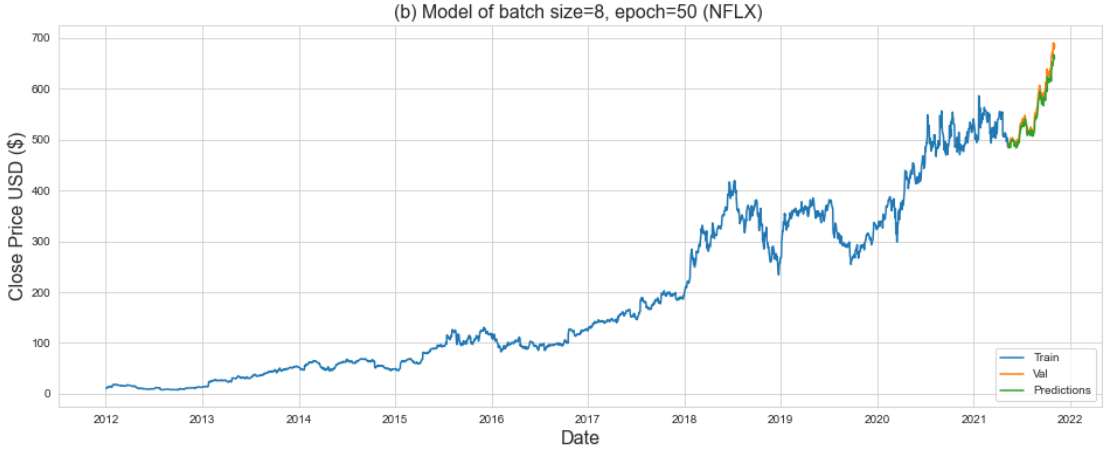
Figure 4.10 Training loss per epoch of (TSLA) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

The rate of decline of training loss is slower, when the number of epochs is 10, as compared to the number of epochs is 50. The training loss will be decreased if the number of epochs increases.

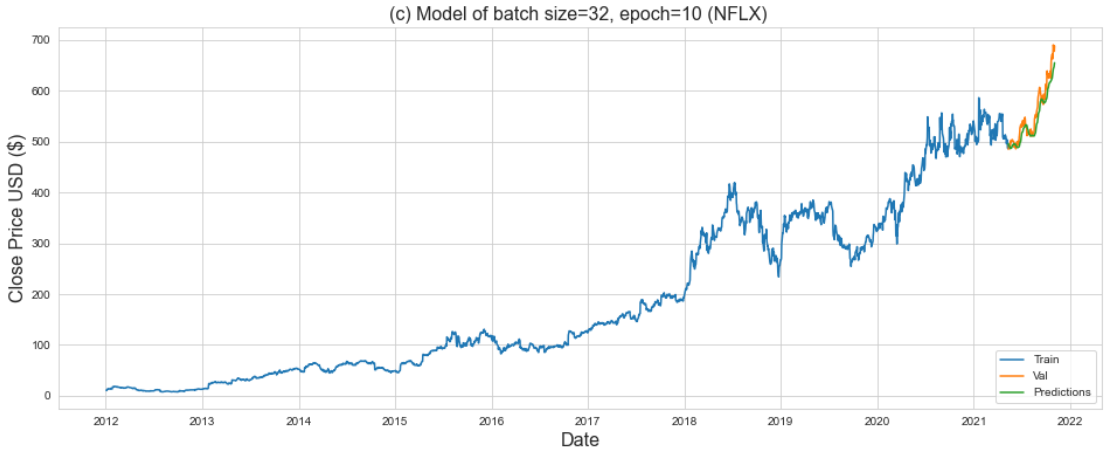
4.4 Prediction of NFLX

The Figure 4.11 below represents the overall of historical price and predicted price of 4 models with different hyperparameters. The predicted value curve, from all the 4 models, are not far away apart from the actual value curve. As long as the models can predict the future movement and trend of the share price, the results can be an extra reference for the investors.

(a) Model of batch size=8, epoch=10 (NFLX)



(b) Model of batch size=8, epoch=50 (NFLX)



(c) Model of batch size=32, epoch=10 (NFLX)

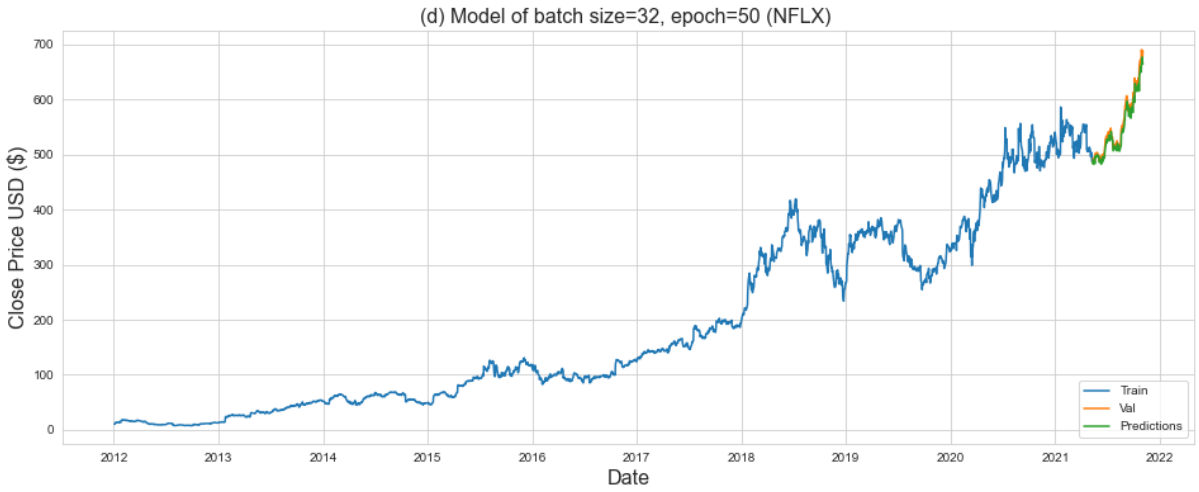


Figure 4.11 Overall of prediction of (NFLX) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

The Figure 4.12 below shows the real vs predicted value of NFLX model (a), (b), (c), and (d). The coefficient of determination of model (a), (b), (c), and (d) is 0.972696689, 0.947534362, 0.866418883, and 0.956960429. The model (a) has the highest coefficient of determination, 97.26% of the variation is explained by the predicted value curve, follow by the model (d), 95.69% of the variation is explained by the predicted value curve. The model (c) has the lowest coefficient of determination, which is 0.866418883, 86.64% of the variation is explained by the predicted value curve.

Figure 4.12 Real vs predicted of (NFLX) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

From the Figure 4.12, all the prediction of 4 models can follow the actual movement of real closing price of NFLX. The curve of predicted and real value of model (a) almost overlap together. The gap between predicted and actual value curve, of model (d) is slightly smaller than model (b), since the batch size of model (d) is higher than model (b).

From the Figure 4.13 below, the training loss of model (c) becomes static after the training loss decreases below 0.001. However, the training loss of model (d) continue decreasing when the training loss is around 0.001 and becomes static when the training loss is below 0.0005. Therefore, from the comparison of training loss per epoch of model (c) and model (d), both the models have the same batch size, which is 32, the higher number of epochs can decrease the training loss to lowest.

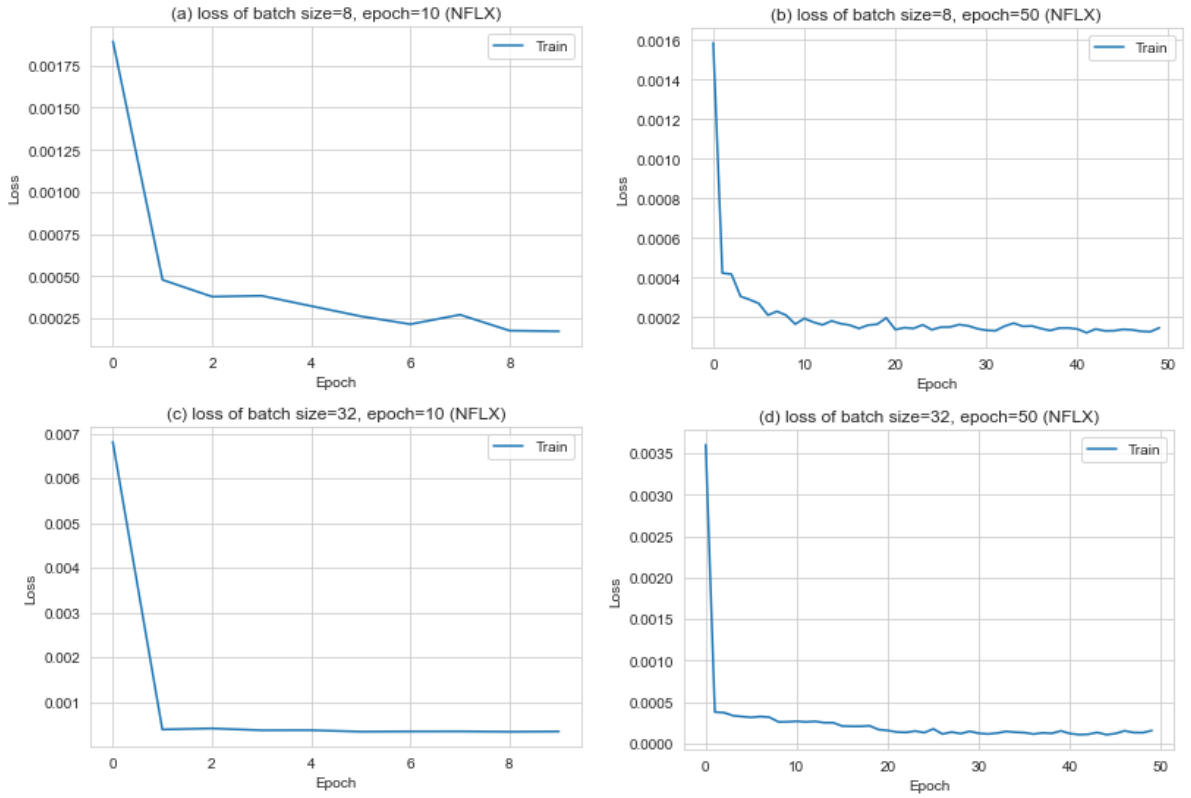
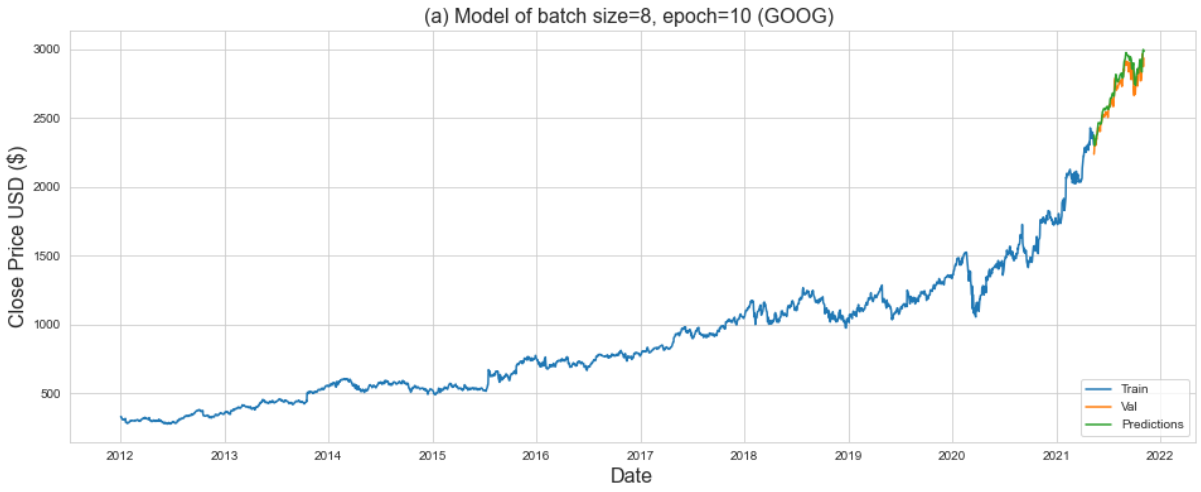
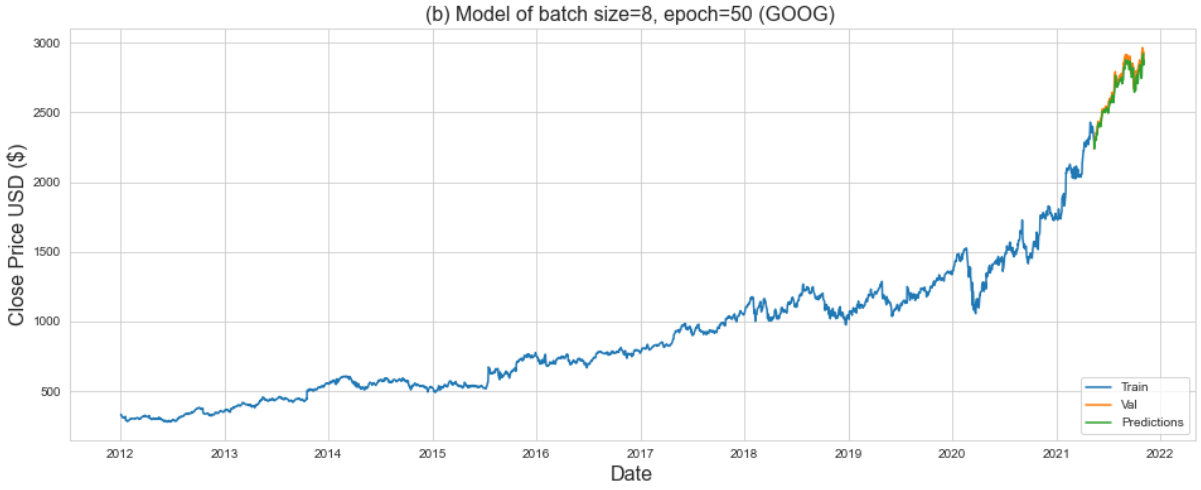


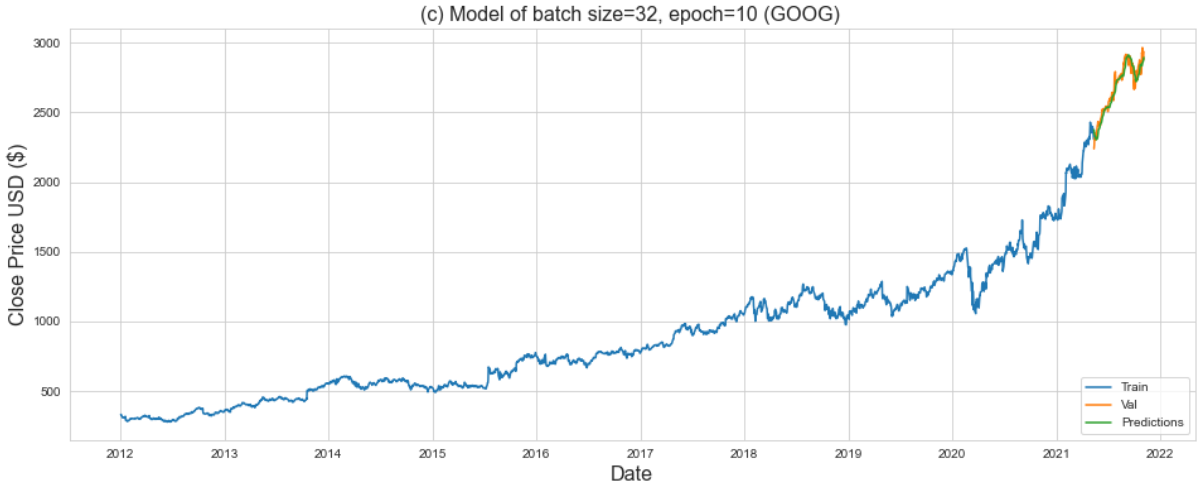
Figure 4.13 Training loss per epoch of (NFLX) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

4.5 Prediction of GOOG

The Figure 4.14 below shows a whole picture of historical and predicted closing price in a long-term period. The 4 models, with different hyperparameters, can predict the future movement of closing price in common.


(a) Model of batch size=8, epoch=10 (GOOG)
Close Price USD ($)
Date
Train
Val
Predictions


(b) Model of batch size=8, epoch=50 (GOOG)
Close Price USD ($)
Date
Train
Val
Predictions


(c) Model of batch size=32, epoch=10 (GOOG)
Close Price USD ($)
Date
Train
Val
Predictions

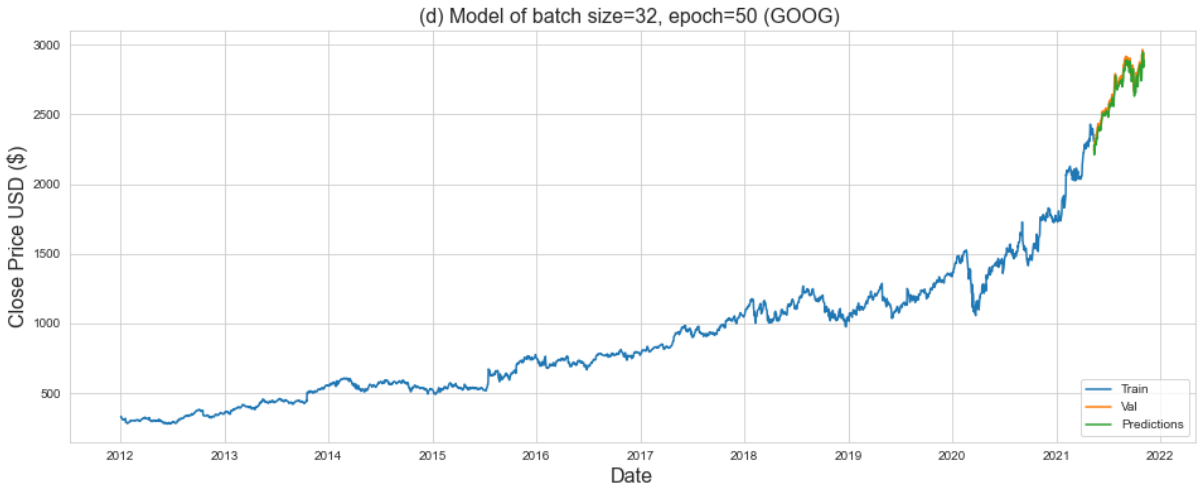


Figure 4.14 Overall of prediction of (GOOG) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

The Figure 4.15 shows that the real vs predicted value of GOOG with 4 hyperparameter. All the 4 models can predict the trend of GOOG. Model (a), the coefficient of determination is 0.879433305, 87.94% of the variation is explained by model (a). Model (b), the coefficient of determination is 0.939955899, 93.99% of the variation is explained by model (b). Model (c), the coefficient of determination is 0.910006292, 91% of the variation is explained by model (c). Model (c), the coefficient of determination is 0.937022724, 93.7% of the variation is explained by model (d). In this dataset, the models with high number of epochs have greater coefficient of determination as compared to the models with low epochs. The models with large batch size have greater coefficient of determination as compared to the models with small batch size.

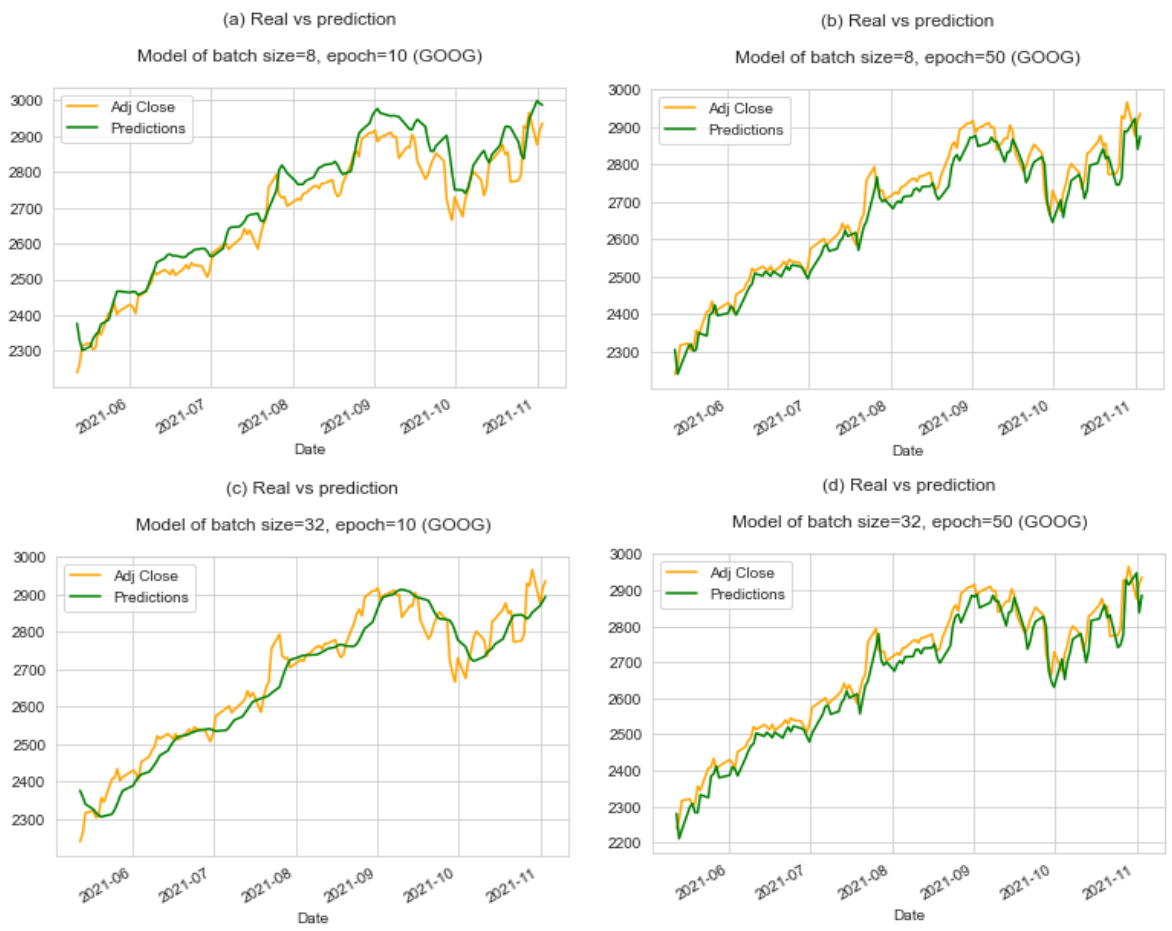


Figure 4.15 Real vs predicted of (GOOG) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

From the Figure 4.16 below, the rate of decline in training loss of the 4 models becomes slower in common, after the training loss reduced to certain level. The rate of decline in training loss of model (a) and (b) becomes static, after a sharp drop to 0.0002. After the training loss of model (c) and (d) drop sharply to 0.0005, the rate of decline in training loss becomes static.

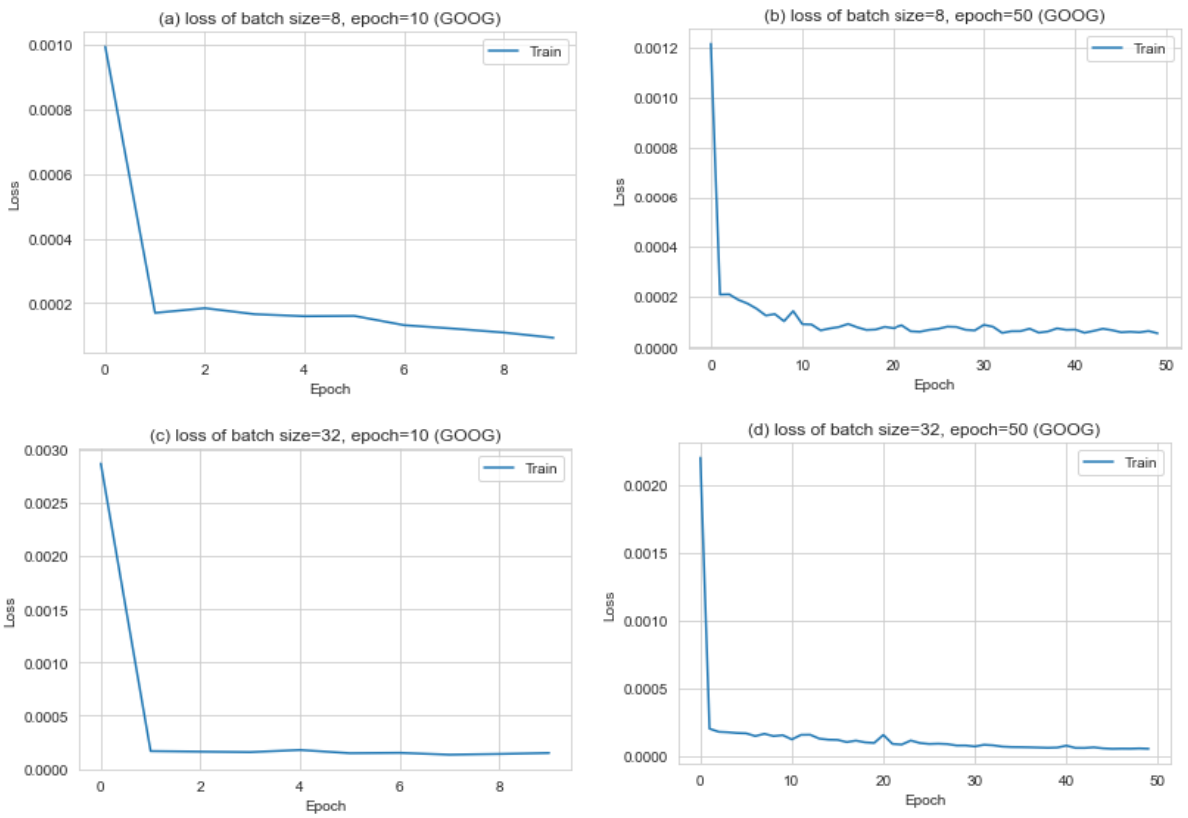


Figure 4.16 Training loss per epoch of (GOOG) of model (a) batch size=8, epoch=10, (b) batch size=8, epoch=50, (c) batch size=32, epoch=10, and (d) batch size=32, epoch=50

4.6 Conclusion

From the Figure 4.17 below shows the coefficient of determination of 4 models in chart. The model with batch size equals to 32 and number of epochs equals to 50, the coefficient of determination of prediction of AAPL, TSLA, NFLX, GOOG are not much different. However, the model with batch size equals to 8 and number of epochs equals to 10, the coefficient of determination of AAPL, TSLA, NFLX, TSLA are quite different. The reason of the difference between two models is the large batch size makes large gradient step and large gradient updates. The number of epochs defines the number times that the learning algorithm will work through the entire training dataset. The model with epochs equals to 50 will work through the entire training dataset 5 times than the model with epochs equals to 10.

In the prediction of NFLX, the performance of batch size=32, epochs=50 is slightly worse than batch size=8, epochs=50. However, the performance of batch size=8, epochs=10 is not better than batch size=32, epochs=10. In conclusion, the accuracy of prediction also needs to depend on the size of the dataset, the characteristics of the dataset, the number of hidden layers of the model, other than the batch size and the number of epochs.

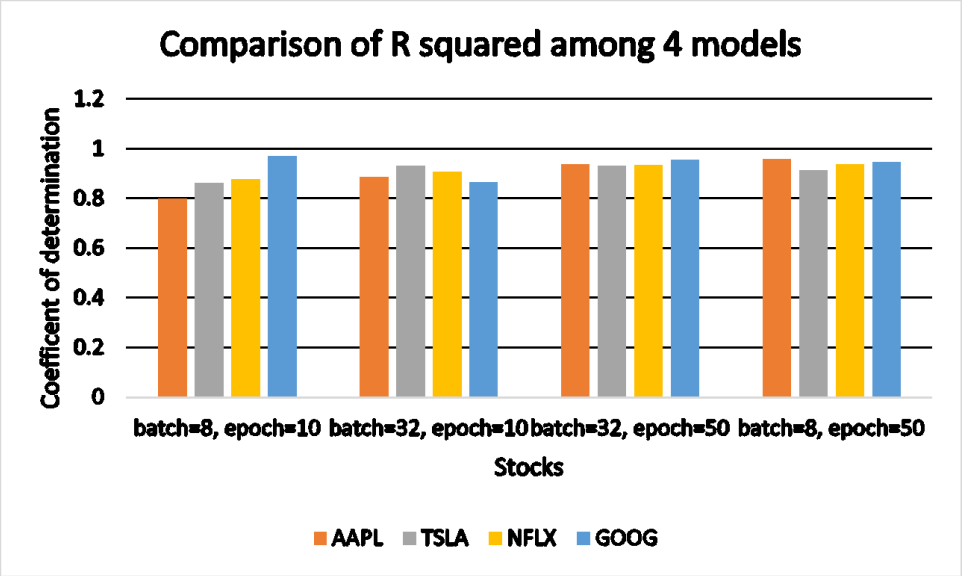


Figure 4.17 Comparison of coefficient of determination

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Volatility of stocks

Beta is an indicator of the volatility of a stock in comparison with the market as a whole. A benchmark index is used as the measurement for the market. In this project, the benchmark index will be SPY. SPY is an exchange-traded fund which is designed to track the S&P 500 stock market index, and also the largest ETF in the world. Knowing how volatile a stock's price is can help an investor decide whether it is worth the risk.

The baseline number for beta is 1, which indicates that the stock price moves exactly as the market moves. A stock price will move more 50% than the benchmark index if the beta of the stock is 1.5, and vice versa if the beta of a stock is 0.5.

Table 5.1 Beta value of the stocks

Stock	Beta
AAPL	1.3487191169135797
TSLA	1.6190338150861687
NFLX	0.8863744452626994
GOOG	1.2519941805817065

Figures 5.1, 5.2, 5.3 and 5.4 below describe the daily return of stock AAPL, TSLA, NFLX, and GOOG compare with the benchmark index, SPY in linear regression.

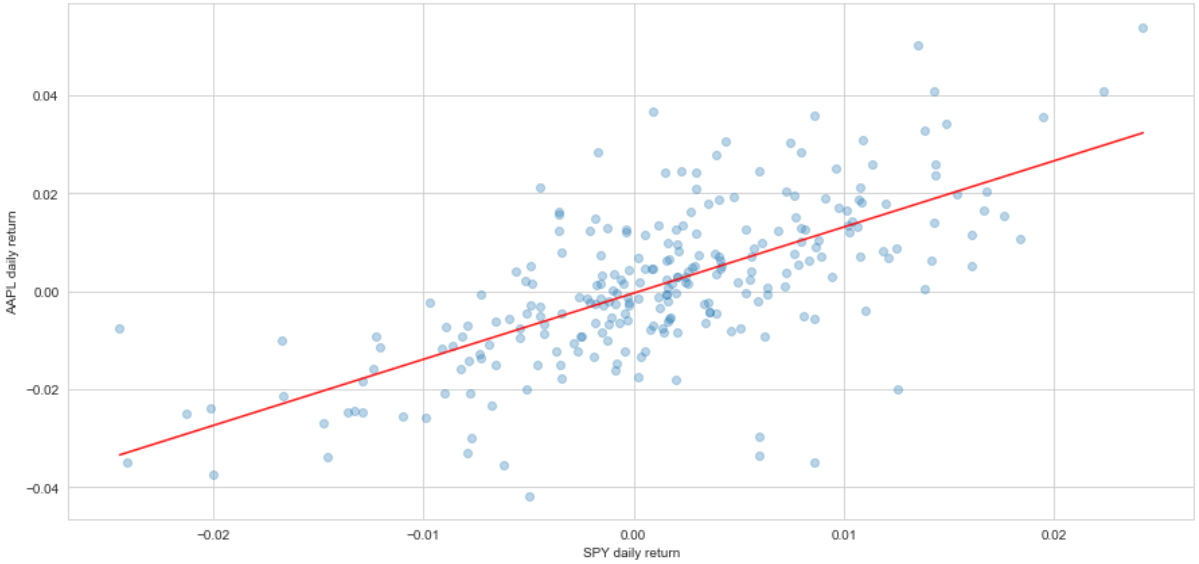


Figure 5.1 AAPL daily return vs SPY daily return

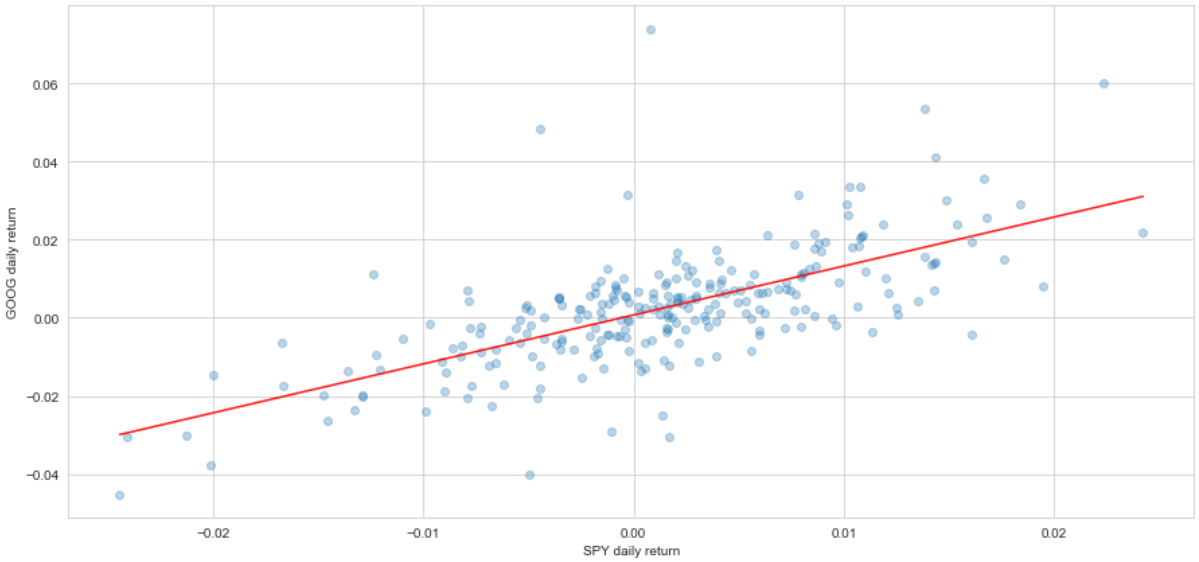


Figure 5.2 GOOG daily return vs SPY daily return

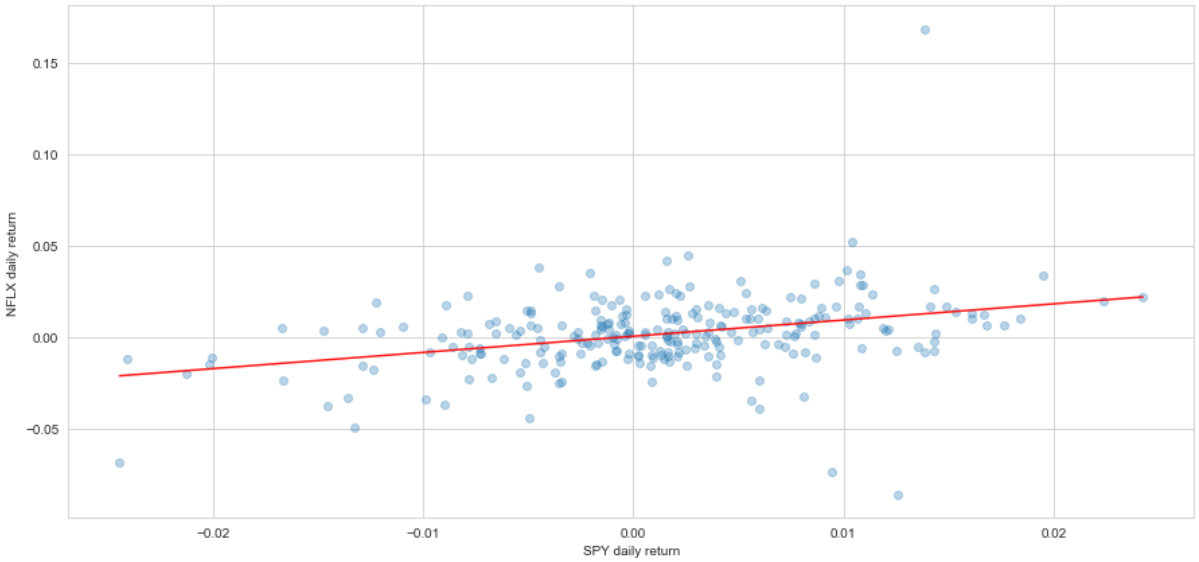


Figure 5.3 NFLX daily return vs SPY daily return

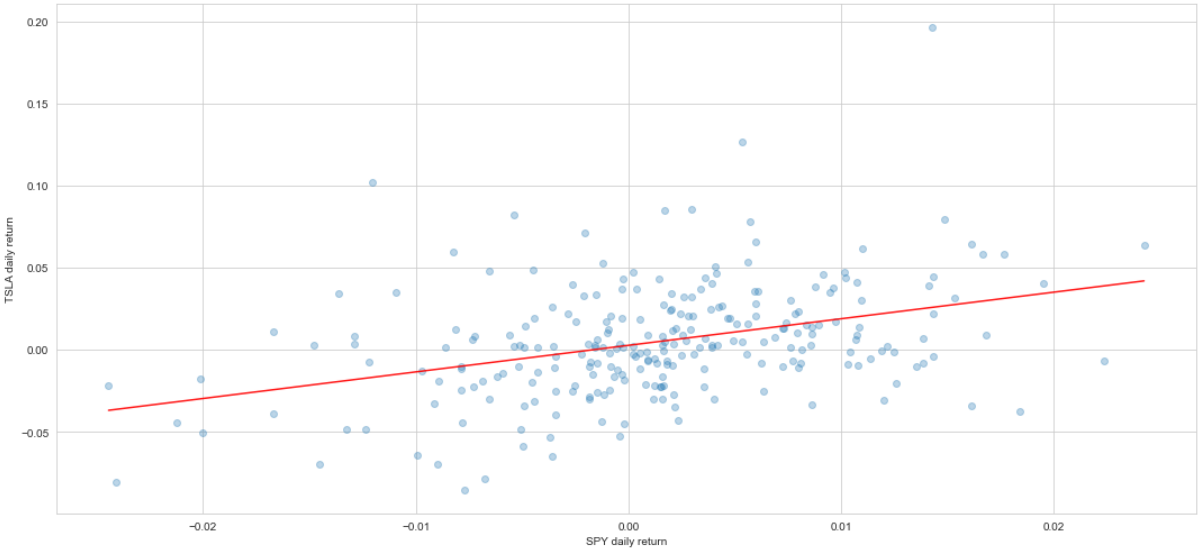


Figure 5.4 TSLA daily return vs SPY daily return

From the Figure 5.5, TSLA has the beta value among all the stocks and NFLX has the lowest beta value among all the stocks. According to the chart, the difficulty to forecast an accurate, stable result of stock price of TSLA will be the highest. At the same time, TSLA has the highest alpha value among all the stocks. This indicates that the successful of forecasting of TSLA will give the highest return theoretically. GOOG has the second lowest of beta value among all the stocks, but it has the second highest of alpha value among all the stocks. In

other word, the forecasting of GOOG is easier as compared to TSLA but the return is the second highest if the forecasting is successful.

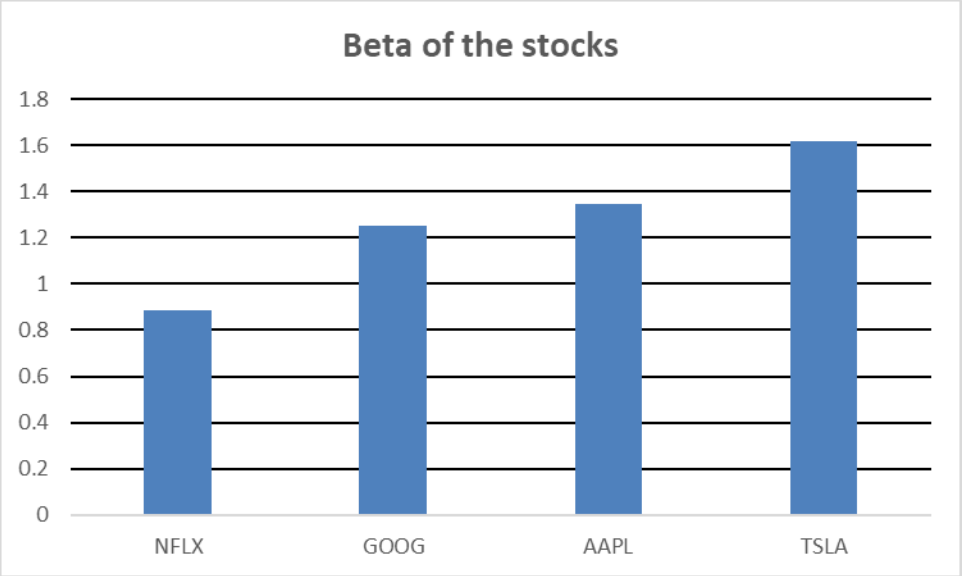


Figure 5.5 Comparison of beta among the stock AAPL, GOOG, NFLX, TSLA

5.2 Factors that affect the accuracy of model

5.2.1 Volatility of stock

The higher the beta value of a stock, the more volatile of the stock as compared to the benchmark market index. The more volatile of a stock will affect the result of performance.

Hyperparameters of model

First, increasing of hidden layer may improve the performance of the model. Ian et al (2016) stated that machine learning algorithms will generally perform best when their capacity is appropriate for the true complexity of the task they need to perform and the amount of training data they are provided with. Models with insufficient capacity are unable to solve complex tasks. Models with high capacity can solve complex tasks, but when their capacity is higher than needed to solve the present task, they may overfit.

Secondly, Keskar et al (2017) stated that stochastic gradient descent is sequential and uses small batches, so it cannot be easily parallelized. Using larger batch sizes would allow the model to parallelize computations to a greater degree. Hence, this could significantly speed up model training.

5.2.2 Other factors that affect stock price

The only feature that input into the models only the close price of the stocks. However, there are many factors can cause the stock price rise or fall. Some specific news may affect the share price, for example: news about a company's earning and profits. Company news and performance will highly change how a investor feel about the stock.

Investor sentiment is one of the vital factors of the fluctuation of stock price. If prices in a particular market are expected to keep rising, investors are said to be bullish, as well as investor optimism, and vice versa.

5.2.3 Difficulty of neural network in prediction

An important issue of neural network in stock prediction is the unstable of result in each prediction. The result is different in every prediction. Thus, investors cannot make short-term or mid-term trading depends on the predicted value

Furthermore, multivariate timeseries which using more than more than 1 series to predict the next value or next sequence of output. Univariate time series comprised of a single series of observations and a model is required to learn from the series of past observations to predict the next value in the sequence, which has limitation on prediction.

Conclusions

LSTM models are demonstrated on small contrived time series problems intended to give the flavour of the type of time series problem being addressed. Since the chosen models is arbitrary and not optimized for each problem, therefore a customized LSTM model which can produce highest accuracy is necessary.

According to the chronological characteristics of stock price data, this paper proposes a LSTM to predict the stock closing price of the next day. The method uses adjusted close price of the stock as the input, making full use of the time sequence characteristics of the stock data. Function of LSTM is to learn the feature data and predict the closing price of the stock the next day. This paper takes the relevant data of the AAPL, TSLA, NFLX, GOOG as an example to verify the model result. The experimental results show that the LSTM with the batch_size=32 and epoch=100 has a higher forecasting accuracy and a better performance compared to the LSTM with the batch_size=1 and epoch=1, and R^2 is close to 1. LSTM is potential for the forecasting of stock prices and considerable of referential importance for investment decision making. LSTM also provides the proposal of practical experience for people's research on financial time series data. However, the model still has some weakness. For example, it only considers the impact of stock price data on closing prices and fails to quantify emotional factors such as news and national policy into the forecast.

References

1. Ariyo, A., Adewumi, A. and Ayo, C., 2014. Stock Price Prediction Using the ARIMA Model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.
2. Althelaya, K.A., El-Alfy, E.S.M. and Mohammed, S., 2018, April. Evaluation of bidirectional LSTM for short-and long-term stock market prediction. In 2018 9th international conference on information and communication systems (ICICS) (pp. 151-156). IEEE.
3. Bollen, J., Mao, H. and Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), pp.1-8.
4. Borowski, K., 2014. *Analiza fundamentalna. Metody wyceny przedsiębiorstwa*. Difin SA.
5. Boyacioglu, M. and Avci, D., 2010. An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange. *Expert Systems with Applications*, 37(12), pp.7908-7912.
6. Bustos, O. and Pomares-Quimbaya, A., 2020. Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, 156, p.113464.
7. Cakra, Y. and Distiawan Trisedya, B., 2015. Stock price prediction using linear regression based on sentiment analysis. *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*.
8. Cao, J. and Wang, J., 2019. Stock price forecasting model based on modified convolution neural network and financial time series analysis. *International Journal of Communication Systems*, 32(12), p.e3987.

9. Chen, K., Zhou, Y. and Dai, F., 2015, October. A LSTM-based method for stock returns prediction: A case study of China stock market. In 2015 IEEE international conference on big data (big data) (pp. 2823-2824). IEEE.
10. De Long, J., Shleifer, A., Summers, L. and Waldmann, R., 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), pp.703-738.
11. Deng, S., Huang, Z.J., Sinha, A.P. and Zhao, H., 2018. The interaction between microblog sentiment and stock return: An empirical examination. *MIS quarterly*, *42*(3), pp.895-918.
12. Ding, X., Y. Zhang, T. Liu and J. Duan, 2015. Deep learning for event-driven stock prediction. Proceedings of the 24th International Joint Conference on Artificial Intelligence, July 25-31, 2015, AAAI Press, pp: 2327-2333.
13. Jung, C. and Boyd, R., 1996. Forecasting UK stock prices. Applied Financial Economics, 6(3), pp.279-286.
14. De Long, J., Shleifer, A., Summers, L. and Waldmann, R., 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), pp.703-738.
15. Ding, X., Y. Zhang, T. Liu and J. Duan, 2015. Deep learning for event-driven stock prediction. Proceedings of the 24th International Joint Conference on Artificial Intelligence, July 25-31, 2015, AAAI Press, pp: 2327-2333.
16. Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), p.383.
17. Fama, E., 2014. Two Pillars of Asset Pricing. American Economic Review, 104(6), pp.1467-1485.
18. Figurska, M. and Wisniewski, R., 2016. Fundamental Analysis–Possiblity of Application on the Real Estate Market. *Real Estate Management and Valuation*, *24*(4), pp.35-46.
19. Hafezi, R., Shahrabi, J. and Hadavandi, E., 2015. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Applied Soft Computing*, 29, pp.196-210.
20. Keskar, N. S. et al. (2017) “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima”, arXiv [cs.LG]. Available at: http://arxiv.org/abs/1609.04836.

21. Krzywda M., 2010, GPW II. Akcje i analiza fundamentalna w praktyce (Shares and Fundamental Analysis in Practice), Wydawnictwo Złote Myśli, Gliwice.
22. Lo, A. and MacKinlay, A., 1989. The size and power of the variance ratio test in finite samples. *Journal of Econometrics*, 40(2), pp.203-238.
23. Lu, W., Li, J., Li, Y., Sun, A. and Wang, J., 2020. A CNN-LSTM-Based Model to Forecast Stock Prices. *Complexity*, 2020, pp.1-10.
24. Moghar, A. and Hamiche, M., 2020. Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science*, 170, pp.1168-1173.
25. Mondal, P., Shit, L. and Goswami, S., 2014. Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications, 4(2), pp.13-29.
26. Nakamura, T. and Small, M., 2007. Tests of the random walk hypothesis for financial data. *Physica A: Statistical Mechanics and its Applications*, 377(2), pp.599-615.
27. Nazarowa J., 2014, Portfolio Structure Planning and Its Future Price Forecasting Model, 8th International Scientific Conference "Business and Management", pp. 290-291.
28. Pai, P. and Lin, C., 2005. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), pp.497-505.
29. Shleifer, A. and Vishny, R., 1997. The Limits of Arbitrage. *The Journal of Finance*, 52(1), pp.35-55.
30. Schumaker, R., Zhang, Y., Huang, C. and Chen, H., 2012. Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3), pp.458-464.
31. Ta, V., Liu, C. and Tadesse, D., 2020. Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. Applied Sciences, 10(2), pp.437-457.
32. Ţiţan, A., 2015. The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research. *Procedia Economics and Finance*, 32, pp.442-44.
33. Bessler W., Lückoff P. (2008) Predicting Stock Returns with Bayesian Vector Autoregressive Models. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_59