

STOCK PRICE PREDICTION USING SENTIMENT AND TECHNICAL
ANALYSIS

HUAQI GAO

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF ARTS

GRADUATE PROGRAM IN Information Systems and Technology
YORK UNIVERSITY
TORONTO, ONTARIO

August 2024

© Huaqi Gao, 2024

Abstract

With the rapid advancement of the economy, the stock market has garnered extensive attention in both business and academic fields. Due to the dynamic, unstable, information-sensitive nature of the stock market, obtaining an accurate stock price prediction is extremely challenging. This study explores the integration of sentiment data from financial news headlines with historical stock data to predict stock prices. This research gathered historical price and trading volume data for the S&P 500 index, sourced from Yahoo Finance, along with 106,494 financial news titles obtained from Reuter. This dataset encompasses the period around the 2008 financial crisis from Oct 20, 2006, to Nov 19, 2013. Empirical implementation of the proposed methodology revealed the substantial value of incorporating sentiment and historical information to enhance the accuracy of stock price prediction.

Acknowledgements

I would like to acknowledge and give my deep and sincere gratitude to my research supervisor, Prof. Jimmy Huang, for giving me the opportunity to do research and his ongoing mentorship and support through all the stage of my thesis. It was a great honor to study under his guidance. I am extremely grateful for what he has offered me.

I would also like to thank my graduating committee members: Prof. Gary Spraakman, Prof. Xiaohui Yu and Prof. George Georgopoulos, for their patience and insightful suggestions.

Last, but not least, I would like to thank my family for their unconditional love, continuous support, and understanding when undertaking my research and writing my thesis.

Preface

Before delving into the contents of this thesis, I would like to acknowledge a milestone in my academic journey. In "A Survey of Explainable E-Commerce Recommender Systems", co-authored with Shunke Zhou, which was published in Web Intelligence and Intelligent Agent Technology (WI-IAT) in 2022 [99], I had the privilege of contributing to the field of E-Commerce. This publication represents a valuable stepping stone in my research career. I am deeply grateful to Prof. Jimmy Huang and Sherry Zhu for their support and guidance throughout this publication journey. This work has not only deepened my understanding of the E-Commerce recommender system but has also paved the way for my research career.

TABLE OF CONTENTS

Abstract	ii
Acknowledgments	iii
Preface	iv
Table of Contents.....	v
List of Tables	viii
List of Figures	ix
List of Abbreviations	x
 Chapter 1: Introduction.....	 1
1.1 Background.....	1
1.2 The Importance of Study.....	3
1.3 Contribution.....	4
1.4 Thesis Outline	5
Chapter 2: Literature Review.....	7
2.1 Stock Price Prediction	7
2.1.1 Theories of Stock Price Prediction.....	7
2.1.2 Stock Price Prediction Approaches.....	8
2.1.3 Influence of News Articles on Stock Price Movement.....	10
2.2 Text Mining of Financial News in Stock Price Prediction.....	11
2.3 Review of Machine Learning Contribution in Stock Price Prediction.....	12
2.3.1 Technical Analysis.....	13

2.3.2	Sentiment Analysis of Financial News.....	15
2.3.3	Combination of Technical and Sentiment Analysis.....	17
Chapter 3: Methodology.....		20
3.1	Text data Preprocessing.....	20
3.2	Feature Selection.....	21
3.2.1	Technical Indicators.....	25
3.2.2	Sentiment Features.....	28
3.3	Data Splitting.....	30
3.4	Proposed Stock Price Prediction Models.....	30
3.4.1	Support Vector Machine.....	30
3.4.2	Multi-Layer Long Short-term Memory.....	33
Chapter 4: Stock Price Prediction Environment.....		38
4.1	Experiment Environment Setting.....	38
4.2	Data Collection.....	39
4.3	Evaluation Metrics.....	41
Chapter 5: Result and Analysis.....		43
5.1	Baseline Models.....	43
5.2	Proposed Models.....	44
5.2.1	SVR Models.....	46
5.2.2	Multi-Layer LSTM Models.....	50
5.3	Comparison of Baseline and Proposed Models.....	54
Chapter 6: Conclusion and Directions.....		56

6.1 Conclusion.....	56
6.2 Future Directions.....	57
References.....	60
Appendices.....	74
Appendix A: Technical Indicators Computing.....	74
Appendix B: Financial News Headline Extracting.....	76
Appendix C: News Headlines Preprocessing.....	79
Appendix D: Sentiment Analysis Using LMD.....	82
Appendix E: Data Splitting and Standardization.....	86
Appendix F: Effectiveness of Proposed SVR Model Experiment	88
Appendix G: Effectiveness of Proposed Multi-Layer LSTM Model Experiment.....	91
Appendix H: Predicted Results Using SVR with Technical Analysis Only.....	94
Appendix I: Predicted Results Using Multi-Layer LSTM with Technical Analysis Only.....	100
Appendix J: Predicted Results Using SVR with Technical and Sentiment Analysis (NLTK).....	106
Appendix K: Predicted Results Using SVR with Technical and Sentiment analysis (LMD).....	112
Appendix L: Predicted Results Using Multi-Layer LSTM with Technical and Sentiment Analysis (NLTK).....	118
Appendix M: Predicted Results Using Multi-Layer LSTM with Technical and Sentiment Analysis (LMD).....	124

LIST OF TABLES

Table 3.1.1:	Original Financial News Headlines for 2006-10-23.....	23
Table 3.1.2:	Preprocessed Financial News Headlines for 2006-10-23.....	24
Table 3.2.1.1:	Technical Indicators of Historical Data for First Two Months.....	28
Table 3.2.2:	Loughran-McDonald (LMD) Categories.....	29
Table 4.1:	Computing Environment Condition.....	39
Table 4.2.1:	Historical Stock Price and Trading Volume of First Two Months.....	40
Table 5.1.1:	Evaluation Results of Baseline Models.....	44
Table 5.2.1.1:	Best Combination of Hyperparameters for Two SVR Models.....	48
Table 5.2.1.2:	Evaluation Results of Two SVR Models.....	48
Table 5.2.2.1:	Best Combination of Hyperparameters for Two Multi-layer LSTM models.....	52
Table 5.2.2.2:	Evaluation Results of Two Multi-layer LSTM Models.....	53
Table 5.3:	Comparison of Evaluation Results for Prediction Models.....	55

LIST OF FIGURES

Figure 3.1: The Overall Research Process.....	22
Figure 3.4.1.1: Support Vector Classifier Architecture.....	31
Figure 3.4.1.2: Support Vector Regression Architecture.....	32
Figure 3.4.2.1: Multi-Layer LSTM.....	34
Figure 3.4.2.2: LSTM Cell Architecture.....	35
Figure 5.1.1: Prediction Result of SVR Baseline Model with Technical Information.....	45
Figure 5.1.2: Prediction Result of Multi-Layer LSTM Baseline Model with Technical Information....	46
Figure 5.2.1.1: Stock Price Prediction of SVR with NLTK.....	49
Figure 5.2.1.2: Stock Price Prediction of SVR with LMD.....	50
Figure 5.2.2.1: Stock Price Prediction of Muti-Layer LSTM with NLTK.....	53
Figure 5.2.2.2: Stock Price Prediction of Muti-layer LSTM with LMD.....	54

LIST OF ABBREVIATIONS

Here is a list of abbreviations used in this thesis:

EMH: The Efficient Market Hypothesis

SMA: Simple Moving Average

EMA: Exponential Moving Average

MACD: Moving Average Convergence/Divergence Rules

RSI: Relative-Strength Index

ADX: Average Directional Index

ROE: Return on Equity

D/E: Debt/Equity ratio

MC: Market Capitalization

P/S: Price/Sales ratio

P/B: Price/Book ratio

EPS: Earnings Per Share

P/E: Price/Earnings ratio

ROA: Return on Assets

DJIA: Dow Jones Industrial Average

NASDAQ: stock market index of Nasdaq stock exchange

S&P 500: stock market index of the 500 largest companies listed on stock exchanges in the United States

RNN: Recurrent Neural Network

MKL: Multiple Kernel Learning

AI: Artificial Intelligent

ANN: Artificial Neural Networks

SM: Support Vector Machines

PSO: Particle Swarm Optimization

LS-SVM: Least Square Support Vector Machine

RNN: Recurrent Neural Network

NLTK: Natural Language Toolkit

ARIMA: An Autoregressive Integrated Moving Average

MAPE: Mean Absolute Percentage Error

LSTM: Long Short-Term Memory

S&P 500: Standard and Poor's 500

LMD: Loughran-McDonald

NLTK: Natural language Toolkit

SVM: Support Vector Machine

SVC: Support Vector Classifier

SVR: Support Vector Regression

MSE: Mean Squared Error

R^2 : R-squared

MAE: Mean Absolute Error

RMSE: Root Mean Squared Error

RBF: Radial Basic Function

Chapter 1

Introduction

In the past century, the stock market has emerged as a major platform for financing. It enables companies to raise funds by issuing stock and enables investors to make profits by increasing the value of investments and distribution allocated by the company. The exponential growth of digital stock market has generated a wealth of valuable trading data. This data can be utilized for data mining and machine learning algorithms to analyze prospective trends, meet consumer needs, and facilitate informed investment decisions.

1.1 Background

The integration of the financial industry with advanced technology, as well as the index-level growth of information technology and artificial intelligence, has triggered the rapid growth of data [94]. The application and analysis capabilities of big data are becoming the core competitiveness of future development of financial investment.

In recent years, more than two exabytes of data have been generated daily from various sources. The rapidly growing data volume has created opportunities for financial institutions to radically revolutionize their service processes and enhance their competitive advantage [1] [2].

Easy and rapid availability to financial data provides many benefits in precision marketing, fraud prevention, consumer credit, risk assessment, supply chain finance, stock market forecasts, investment management, and other financial services.

As people's investment awareness increases, the stock market has become the primary financing platform to optimize resources allocation and effectively adjusting the risk and return. An accurate stock price prediction can provide investors with appropriate investment opportunities and promptly avoids investment risks. However, the stock market is relatively fragile and sensitive, and it is full of uncertainties as various factors influence the price fluctuation. The Efficient Market Hypothesis (EMH) states that the stock price movement is always driven by various types of new information, including technical and fundamental information [3], which is relevant to the stock market and can change the investors' perceptions and expectations.

With the rapid dissemination of web information, web media has been considered as a critical factor and a major threat to the stability of the stock market [4]. One of the representative media types in stock price prediction comes from financial news articles. To attract readers' attention, news expresses opinions while reporting incidents [5]. Financial news articles usually contain the operating conditions of companies, the activities involved in the company, and the opinions of market participants and expectations for future stock movements. The continuous availability of more financial news articles has inspired the application of text mining to discover the relationship between the sentiment of news articles and stock price movement.

Due to the unstable, information-sensitive, and noisy nature of the stock market, developing an accurate predictive model is complex and challenging. Most of the proposed methods rely heavily on the use of supervised learning. The core of these procedures involves the training

process with access to the output value. A supervised learning algorithm maps labeled input data to output data and produces an inferred function that can be applied to new examples.

According to the applications, information-driven stock movement prediction can be categorized into three types: (1) technical analysis, (2) fundamental analysis, and (3) sentiment analysis. Technical analysis utilizes historical stock prices and the trading volume for statistical analysis to draw trend conclusions. Fundamental analysis presents the intrinsic market value of stocks, which is heavily rely on the company's financial statements and annual reports. Sentiment analysis uses unstructured datasets such as news, messages, and data extracted from web media [6] [7] [8]. Many previous studies have used various machine learning models, based on technical analysis or fundamental analysis to study stock price predictions for a particular industry, and the reliability of the prediction results is disputed. These studies have employed structured and numerical databases in analyzing stock prices. Nevertheless, a recent proposed method to predict stock price prediction has proved that the stock price can be predicted more accurately by analyzing sentiment information from financial news.

1.2 The Importance of Study

The stock market has received extensive attention in both business and academic fields due to its commercial and educational value. Desirable stock price prediction can significantly provide investors with appropriate investment opportunities and promptly avoids investment risks in a timely manner.

Since stock price movements directly affect investors' profit, it has always been a challenging task to precisely predict stock prices. Past studies have proven the feasibility of stock

price prediction based solely on technical, fundamental, or sentiment analysis. Investors are increasingly susceptible to influence as the volume of daily financial news and readers increases. Investors learn the latest information about the stock market from financial news and immediately formulate corresponding investment strategies, which will have a direct impact on the volatility of the stock market. Hence, to improve the accuracy of stock price prediction, it is necessary to investigate the joint impact of technical information and sentiment information from financial news on the stock market.

1.3 Contribution

The major contribution of this research has been the development of procedures to highlight the significance of integrating sentiment and technical data in stock price prediction, increase prediction efficiency, and enable investors to use such predictions to assess past investment strategies, manage risk, make informed investment decision in varying market conditions, and help in preparing for future uncertainties.

This research is consistent with the literature on the topic of sentiment analysis in financial news. Although previous studies have provided a wealth of evidence, highlighting that sentiment analysis of news content improved prediction accuracy of stock price movement, this research suggests that extracting sentiment information from news headlines is more efficient and noise-reducing. With the development of the global economy and the rapid expansion of the internet, the volume of financial news has continued to grow. It is time-consuming and difficult to analyze stock price in combination with a large amount of news content. News headlines are designed to capture the most critical information concisely. Without the distraction of intricate information,

they highlight noteworthy announcements or events that are more likely to have a direct impact on stock prices. Full articles frequently provide context, stories, and other material that might not be immediately related to changes in stock prices. In the fast-paced world of stock trading, headlines help filter out this irrelevant information, enabling faster data processing and quicker response times for predictive models.

The majority of current research concentrates on using machine learning classification algorithms to predict stock price movements. Through investigation, this study has deepened the understanding of using machine learning regression models to predict real stock prices, allowing investors to make timely informed investment strategies on asset allocation and risk management under extreme conditions. The phrase “don’t put all your eggs in one basket” is commonly associated with the financial risk management and advocating for investment diversification. This concept suggests that it’s risky to put all your money in single asset, company, or industry. Instead, spreading your investments across a variety of assets, such as stocks, bonds, real estate, and different industries, can reduce risk [98]. If one investment performs poorly, the other investments may perform better, thereby balancing the overall portfolio and minimizing potential losses. This strategy is at the heart of modern portfolio theory, which seeks to optimize the balance between risk and return. Understanding market sentiment and investment behavior before, during, and after a crisis provided valuable insight for future crisis management and prevention.

1.4 Thesis Outline

This thesis consists of six chapters, a list of references and appendices. The organization of the thesis is as follows. Chapter 1 introduces an introduction of stock price prediction, including the background of stock market, the importance of study, contribution behind this thesis, and the

structure of this thesis. Chapter 2 discusses previous studies related to the stock price prediction, including the theories of stock price prediction, the existing approaches of stock price prediction, and the contribution machine learning in stock price prediction. Chapter 3 introduces the data used, extracted features and machine learning models adopted in this thesis. In Chapter 4, the experimental environment is described, consists of the computing environment, and evaluation metrics used in experiments. Chapter 5 presents detailed experimental results and analysis. Chapter 6 concludes the thesis and provides possible future directions.

Chapter 2

Literature Review

There are many studies on the topic of stock price prediction. This section primary reviews the studies using machine learning methods in stock price prediction based on technical and sentiment analysis, as these studies are the most related to this research.

2.1 Stock Price Prediction

Over the past two decades, stock price prediction has always been a popular topic. However, the volume of structured and unstructured data continues to grow unabated, making the stock price movements more uncertain and difficult to predict. Numerous studies have been devoted to discovering an accurate stock price prediction model using various machine-learning approaches. The following sections briefly explain two common theories and existing approaches in stock price prediction.

2.1.1 Theories of Stock Price Prediction

There are two common theories in predicting stock price, namely the Efficient Market Hypothesis (EMH) introduced by Fama [9] and the Random Walk Theory proposed by Malkiel [10].

The Efficient Market Hypothesis (EMH) states that current stocks trade at the fairest value, which means that as new information enters the market, and the stock price will immediately incorporate important information and changes quickly [63] [64]. The EMH can be divided into three versions: Weak, Semi-Strong, and Strong [65] [66]. The weak form suggests that the current stock prices reflect the assimilation of all historical information, and no prediction can be utilized based on technical analysis. The semi-strong form states that the current stock price incorporates all historical and public information, and that the prediction cannot use either technical or fundamental analysis. Current stock prices fully reflect the strong form, which contains historical, public, and private information. This means that no information can predict future stock prices. Recent studies have proven that the strong form of the Efficient Market Hypothesis is incorrect. Due to the limited accessibility of private information, it is impossible for investors to immediately adjust their investment decisions by analyzing all the information. Random Walk Theory is like the semi-strong form of EMH. It implies that stock prices are randomly volatile and cannot be predicted using historical and public information [67] [68] [69] [70].

2.1.2 Stock Price Prediction Approaches

These two conventional theories have been criticized by some researchers who believe that stock prices can be predicted using technical and fundamental analysis through various techniques and algorithms. A systematic review of 122 studies published between 2007 and 2018 on stock price predictions used machine learning algorithms, revealed that 66% of studies utilized technical analysis, 23% utilized fundamental analysis, and the rest of the studies were based on combination

analysis. Among these studies, the support vector machine and artificial neural network have proven to be the most popular algorithms for stock price prediction [11].

Technical analysis is a form of quantitative analysis. It aims to calculate several technical indicators based on historical price and volume data to identify the price movement trends and predict the future stock price. Some of the technical indicators developed in technical analysis discussed in [12][13][14][15] are Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence rules (MACD), Relative-Strength Index (RSI), William (%R), Average Directional Index (ADX), and Bollinger Bands. However, technical analysis will not work in predicting the stock price for a small-cap company because there is not enough trading activity to analyze the trend of the price movement.

The fundamental analysis estimates the intrinsic value of stocks by analyzing the financial and economic variables from three dimensions: the overall economy, the industry, and the company. The fundamental analyst uses published data to predict stock price movement, such as quarterly and annual reports, balance sheets, income statements, auditor's reports, economic activities, political changes, etc. Various financial ratios are applied to analyze the stock price movement including Return on Equity (ROE), Debt/Equity ratio (D/E), Market Capitalization (MC), Price/Sales Ratio (P/S), Price/Book ratio (P/B), Earnings Per Share (EPS), Price/Earnings ratio (P/E) and Return on Assets (ROA) [11]. Heeyoung et al. (2014) predicted the stock price movement by incorporating the textual analysis of financial reports. They collected 8-K financial reports for all S&P 500 companies between 2002 and 2012, whenever a major business event occurs, including bankruptcies, layoffs, the election of a director, a change in credit, etc [28]. According to regulatory requirements, publicly traded companies have to provide annual reports

that contain valuable information about the company's economic events, performance, strategy, and risks. Financial news articles often timely organize and analyze the annual reports in a timely manner to assess the growth of revenue, profitability, potential opportunities, and challenges [71]. As the contents of the financial news articles overlap with the companies' annual reports, readers can gain a deeper understanding into the company's financial situation, predict the stock price movements, and make more informed investment decisions simply by analyzing financial news articles. The following section will review the influence of financial news on stock price movements.

2.1.3 Influence of News Articles on Stock Price Movement

The daily financial news provides electronic versions of stock market. It typically includes global and regional economic conditions, political changes, companies' operations, activities companies are involved in, market participants' opinions, and expectations for future stock movements. For these reasons, researchers are inspired to investigate whether the news has an impact on the stock market.

Rohitha et al. (2007) collected financial news from CNN Headline News Channel, Fox News Channel, www.yahoo.com, and The New York Times for a ten-week period. They analyzed the impact of news article content on three prominent market indicators; DJIA (Dow Jones Industrial Average), NASDAQ (stock market index of Nasdaq stock exchange), and S&P 500 (stock market index of the 500 largest companies listed on stock exchanges in the United States). It confirmed that there is an association between news articles and stock price fluctuations [16].

Tetlock (2007) measured the interactions between the DJIA (Dow Jones Industrial Average) and daily news from the Wall Street Journal. The results suggested that highly pessimism predicted

a downward in market prices, while positive sentiment led to a higher market price. Furthermore, trading volume increases at extremely high or low levels of pessimism [17].

The flourish in electronic journalism has led to a massive increase in the number of daily news. It is both time-consuming and difficult to combine large amount of news content with stock price fluctuations. Therefore, many scholars have turned to studying the impact of news headlines on stock prices. Gaurav et. al (2020) and László et. al (2021) applied several sentiment analysis tools to classify the sentiment of news headlines as positive, negative, and neutral. The results demonstrated that financial news headlines have a significant impact on stock price movements in the absence of news content [18][19].

2.2 Text Mining of Financial News in Stock Price Prediction

Text mining, also known as text data mining, is the process of transforming unstructured text into structured data to derive meaningful patterns. Unstructured text usually includes resources from news, websites, social media, reviews, etc. The amount of textual data is constantly increasing, so text mining is a crucial process. Typically, text mining tasks involve text categorization, text clustering, concept extraction, production of granular taxonomies, document summarization, sentiment analysis, and entity relation modeling [20].

Text mining is used in stock price prediction to extract sentiment information from financial news. Existing text mining techniques have been employed to explore how news articles affect the stock market. [21][29][30][31][57] applied the Bag of Word technique to capture the statistical patterns of words. They stated that news articles related to listed companies had a

significant impact on their stock price movements. [22][32][33] took the dictionary-based approach to extract sentiment features and concluded that financial news articles help to predict stock price movements.

The Multiple Kernel Learning (MKL) techniques were employed in [23] to assign news articles into five categories (stock-specific, sub-industry-specific, industry-specific, group-industry-specific, and sector-specific), analyze them separately, and combine the prediction results. The prediction results showed that simultaneously analyzing five news categories outperforms approaches based on a lower number of news categories.

[19][34] utilized Python powerful natural language processing libraries (TextBlob and Natural Language Toolkit) to label news headlines with sentiment values (positive, negative, and neutral). Additionally, the Recurrent Neural Network (RNN) was used to categorize news headlines as either positive or negative. By comparing the results of stock price prediction, the RNN model outperformed than the other two text mining techniques.

2.3 Review of Machine Learning Contribution in Stock Price Prediction

Machine learning is a subfield of Artificial Intelligence (AI), and it devoted itself to letting systems understand and learn to solve a problem. Many previous studies have proven the application of machine-learning algorithms in stock market prediction. The following sections are going to explain the application of machine learning methods in stock price prediction based on technical and sentiment analysis.

2.3.1 Technical Analysis

Technical analysis often employs machine learning methods to identify stock price movement trends using historical price and volume.

Ince et al. (2007) utilized kernel principal component analysis and factor analysis to identify influential technical indicators for MLP and SVR forecasting models in stock price prediction [73].

Almeida et al. (2010) proposed an automatic stock trading method that combines technical analysis and the nearest neighbor classification. Several well know indicators were used including Moving Average Convergence/Divergence rules (MACD), Relative-Strength Index (RSI), and Simple Moving Average (SMA). The technique helps investors to buy(sell) the stock when predicting price movement is down(up) [24].

Kara et al. (2011) compared the stock price movement predicting performance of two classification models on Istanbul Stock Exchange (ISE) National 100 Index respectively, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Results revealed that the ANN model outperformed the SVM model [25]. Hegazy et al. (2013) optimized the least square support vector machine (LS-SVM) for predicting S&P 500 price movement by employed Particle swarm optimization (PSO). The proposed model achieved better accuracy than ANN (Artificial Neural Network) [26]. Oliveira et al. (2013) combined technical analysis, fundamental analysis, and time series analysis with artificial neural networks for stock prices prediction [75]. Ayşe et al. (2018) utilized ten integrating metaheuristics including to improve the accuracy of ANN (Artificial

Neural Network) model for predicting S&P 500 price movement [27]. Furthermore, Isacc et al. (2019) reviewed eighty technical-based studies published between 2007 and 2018 on stock price predictions with machine learning algorithms. It was observed that the most utilized technical indicators are Simple Moving Average (SMA), Moving Average Convergence/Divergence rules (MACD), and Relative-Strength Index (RSI). The survey also revealed that SVM (Support Vector Machine) and ANN (Artificial Neural Network) are the most popular machine learning algorithms in stock market predictions. A very high percentage of the reviewed studies used classification machine-learning algorithms to predict stock price movement, and few works predict the future stock price [11].

Ticknor (2013) developed a Bayesian regularized artificial neural network using technical indicators as input for forecasting financial market behavior, aiming to improve prediction quality and generalization [74].

Saahil (2015) improved the long-run prediction accuracies of stock price movement by employing SVM with the RBF kernel model based on four technical indicators including price volatility, price momentum, sector volatility, and sector momentum [13]. Kaur and Dharni (2019) proved the SVM with RBF model performs better in predicting the stock price movement across seven countries' stock markets [14].

Nelson et al. (2017) studied the usage of LSTM networks for predicting stock price trends based on historical price data and a set of technical indicators [76]. Dara and Bachala (2023) employed LSTM to predict the next day closing price of APPLE shares using 2 technical indicators to explore trends of movement and predict stock prices [55].

2.3.2 Sentiment Analysis of Financial News

As news about listed company continues to appear, both financial institutions and individual investors need to monitor the immediate impact on stock prices. This led researchers to investigate the impact of news articles on stock price prediction using machine learning algorithms.

Schumaker et al. (2009) examined the Support Vector Machine (SVM) approach to estimate a discrete stock price twenty minutes after a news article is released. They collected 9,211 financial news articles and 10,259,042 stock quotes covering the S&P 500 stocks for five weeks. Several different textual representations were applied such as bag of words, noun phrases, and named entities, results expressed that a proper Noun scheme performs better than the de facto standard of Bag of Words in all three metrics [57]. In a subsequent study, Schumaker et al. (2012) explored sentiment analysis of financial news articles using the Arizona Financial Text (AZFinText) system and found that subjective news articles were more easily to predict stock price direction [77]. Li et al. (2014) demonstrated that combining SVM with a dictionary-based sentiment analysis model outperforms the SVM with a bag-of-words model in predicting the stock price movement of the Hong Kong Stock Exchange [32]. Kirange et al. (2016) investigated the impact of news headlines on the stock price movement of two Indian companies (Infosys and Wipro) over 10 years. The emotion classification results present that SVM outperforms KNN and Naïve Bayes [33]. Chiong et al. (2018) introduced a support vector machine (SVM) with parameters optimized particle swarm optimization (PSO) approach for stock market prediction using news disclosures, incorporating historical stock market data for improved accuracy [80].

Ayman et al. (2017) analyzed the correlations between news sentiment and stock price movement of three companies traded in NASDAQ. News data is classified into either positive or negative sentiments using the Naïve Bayes algorithm, and then the K-NN classifier is applied to predict the stock trend fall or rise. This improved the accuracy of stock trend prediction [62].

Minh et al. (2018) proposed using deep learning approach to predict price trends for the S&P 500 index and the VN-index. Their method is based on a two-stream gated recurrent unit network and sentiment word embedding that was trained on financial news datasets [78].

Saloni et al. (2019) gathered historical stock prices and more than 265,000 financial news articles of S&P500. The NLTK (Natural Language Toolkit) libraries are used to give each document positive, negative, or neutral values. RNN (Recurrent Neural Network) performs better in predicting stock price than ARIMA (An Autoregressive Integrated Moving Average) [34].

Alzazah et al. (2021) conducted a survey comparing various machine learning and deep learning methods for sentiment analysis in stock market prediction, aiming to identify the most effective method for different types and amounts of data [81].

Yang and Yi (2021) proposed a blending ensemble learning method that combines LSTM and gated recurrent unit (GRU) to improve the accuracy of stock price movement for the S&P 500 Index [60]. Gite et al. (2021) focused on using machine learning techniques and deep learning techniques LSTM to predict BSE and NSE stock prices, along with sentiment derived from financial news headlines. The study revealed that the LSTM model with sentiment analysis creates a better predictive mode [82]. Omarkhan et al. (2021) addressed data imbalance in financial news sentiment analysis using CNN and LSTM, emphasizing the importance of machine learning and

deep learning algorithms in developing comprehensive studies on Financial News Sentiment Analysis [83]. Alexandre et al. (2021) adopt the VADER model to process the sentiments of news articles, then incorporate historical prices and sentiments as input features with the LSTM model to predict the price of the S&P 500 index. Experiments demonstrated that the predicted stock prices are more accurate when the sentiment analysis is considered [35]. Sidogi et al. (2021) collected 2621 Amazon-related news article headlines to investigate the influence of financial news headline sentiment on stock price prediction using LSTM networks with a natural language processing model, FinBERT. The result from two performance metrics, Root Mean Square Error (RMSE) and mean absolute error (MAE), suggest that the use of news headline sentiment features from FinBERT significantly improves the predictive performance of LSTM networks in intraday stock price prediction [61]. Gupta et al. (2022) incorporated historical and sentiment analysis with the LSTM model to predict stock prices efficiently, findings emphasized the strong correlation between the movement of stock prices and the publication of news articles [84].

Sharaf et al. (2022) proposed an efficient system that applies sentiment analysis of stock news headlines to predict the impact of the pandemic on stock price movement such as TSLA, AMZ, and GOOG stock [85].

2.3.3 Combination of Technical and Sentiment Analysis

To improve the accuracy of stock price prediction, researchers have entered a new area of studying the joint impact of technical and sentiment analysis.

Manuel et al. (2017) used Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to predict price movements of the S&P 500 index using financial news headlines and a set of technical indicators. Results showed that CNN is better at capturing sentiment information [38]. Pisut and Peerapon (2018) improved the accuracy of stock price movement prediction by applying Convolutional Neural Network (CNN) and Long Short-term Memory (LSTM) architectures with both event embedding vectors extracted from news headlines and a set of technical indicators. They used accuracy and annualized return as evaluation metrics and then examined performance in three datasets obtained from Reuters, Reddit, and Intrinio [58].

Jiawei et. al (2019) employed the LSTM (Long Short-term Memory) neural network to classify the sentiment of news as positive, negative, or even. They then combined sentiment features with various technical indicators to predict the Shanghai Composite Stock Index's price movement. It convinced that technical and sentiment analysis can improve the accuracy of stock trend prediction [39]. Xiaodong et. al (2020) set up an LSTM model to predict the stock price movement of Hang Seng Index (HSI) constituents. Four sentiment dictionaries are used to calculate the sentiment vectors for the financial news articles, which are SenticNet 5, SentiWordNet 3.0, Vader and Loughran-McDonald Financial Dictionary 2018. The results indicated that models incorporating technical and sentiment information outperform models that only use either technical indicators or news sentiments. Additionally, the Loughran-McDonald Financial Dictionary 2018 performs better than the other three dictionaries [59].

Li et al. (2020) proposed a layered deep learning model that incorporated the technical indicators of Hong Kong Stock Exchange stock prices and news sentiments using four different sentiment dictionaries for stock market prediction. The findings suggested that the finance domain-

specific sentiment dictionary (Loughran–McDonald Financial Dictionary) models the news sentiments better than the other three dictionaries [79].

Shengting et al. (2021) adopted the CNN (convolutional neural network) to calculate the sentiment index for financial news. Then they combined historical data, technical indicators, and sentiment index as the features, employing a long short-term memory network to predict the China Shanghai A-share market accurately. The experiments showed that the predicted stock price is closer to the true price using multiple data sources, and the MSE (mean absolute error) achieved 2.39, which is better than using a single data source [40].

Chapter 3

Methodology

The overall process is illustrated in Figure 3.1. The process includes data collection, data processing, feature extraction, machine learning model implementation, and model evaluation. The following sections will specify the details of each step. To accomplish the research, this study uses the Python programming language to implement the overall process.

3.1 Text Data Preprocessing

The text preprocessing is indeed a crucial step in building a machine learning model, especially when working with natural language data. It aids in comprehending the structure of text and preparing it for further analysis. All the processes in text preprocessing are done with the Python programming language. Taking the financial news headlines on October 23, 2006, as an example, Table 3.1.1 shows the original financial news headlines, and Table 3.1.2 shows the processed financial news headlines. Here are the steps involved in text preprocessing:

- Tokenization: Breaking down a sentence into words.
- Lowercasing: Converting all characters to lowercase ensures consistency and reduces the complexity of the text, as it treats words like “Economic” and “economic” as the same.

- Removing punctuation, special characters, numbers, and marks: Punctuation and special characters are often not necessary for understanding the context or sentiment of the text.
- Removing Stop Words: Stop words are common words like "and", "the", "is", etc., which are usually filtered out as they do not contribute much to the meaning of the text.
- Stemming and lemmatization: Reducing words to their root form. Stemming might chop off the ends of words, while lemmatization considers the context and converts the word to its meaningful base form.

3.2 Feature Selection

This study generates technical features and extracts sentiment features. The following sections will explain details of feature selections.

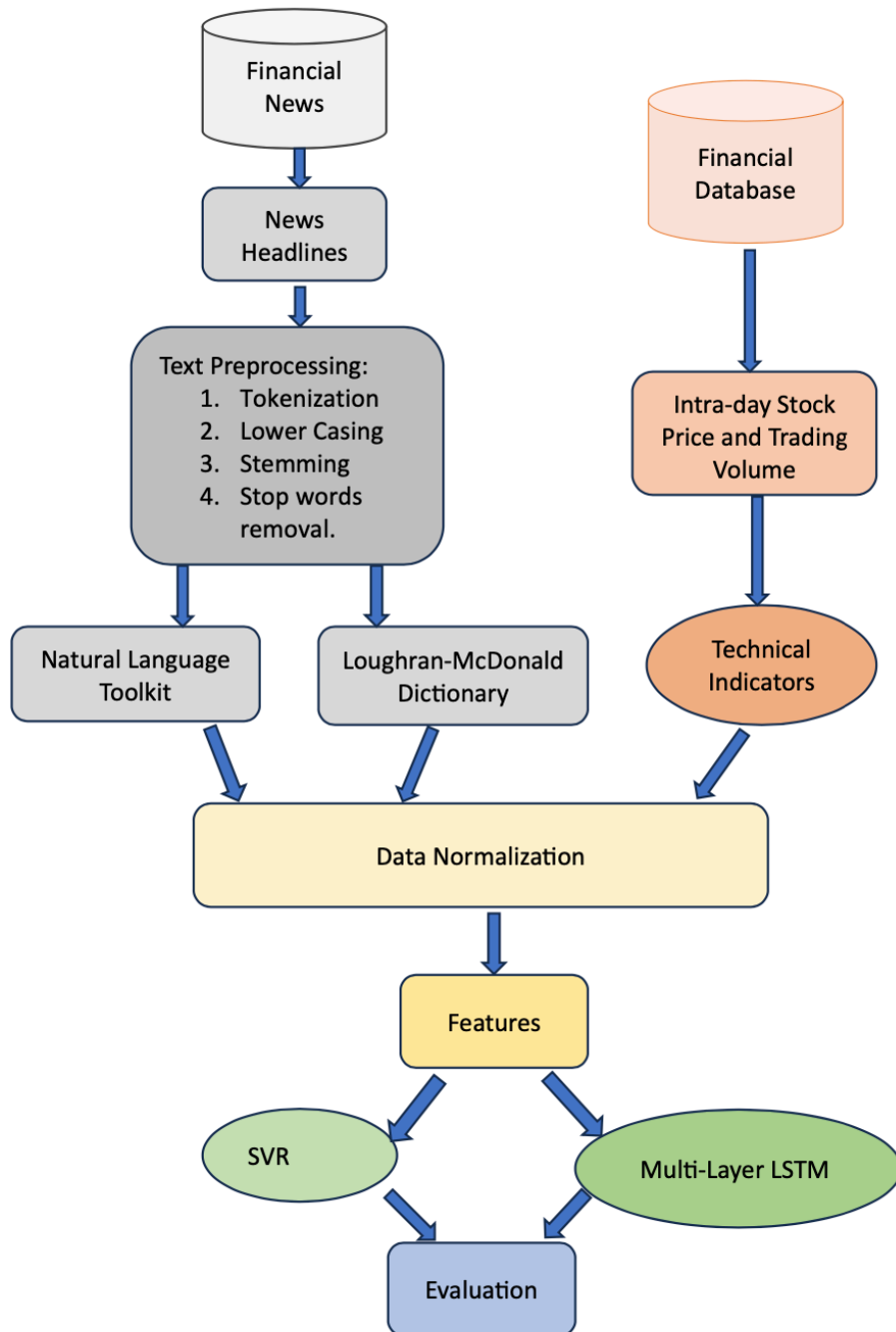


Fig.3.1: The Overall Research Process.

Financial News Headlines for 2006-10-23

2006-10-23 Australia's Foster's says confident on targets
2006-10-23 Macquarie bids \$1.4 bln for metering firm Techem
2006-10-23 DaimlerChrysler talks to 2 firms on small car: paper
2006-10-23 More solid earnings may keep rally going
2006-10-23 U.S. venture investors betting on energy, Web 2.0
2006-10-23 Shell Canada jumps 30 pct on Royal Dutch offer
2006-10-23 Kimberly-Clark sees higher costs this year
2006-10-23 Ford won't sell Ford Credit
2006-10-23 Ford sees fourth quarter weaker than third
2006-10-23 Hasbro profit beats expectations
2006-10-23 Merrill Lynch to buy Petrie Parkman
2006-10-23 Ford CEO: Jury out on European premium brands
2006-10-23 Kimberly-Clark profit higher, narrows view
2006-10-23 Wal-Mart October sales growth tracking below forecast
2006-10-23 Private equity groups form for Tribune bid: source
2006-10-23 AT&T earnings beat expectations
2006-10-23 Wrigley sees no change in communications policy
2006-10-23 Ford is reviewing all products, brands: CEO
2006-10-23 UPDATE 4-Wal-Mart to slow U.S. expansion, spending growth
2006-10-23 Royal Dutch bid for Shell Canada low: shareholder
2006-10-23 Shell looks to buy out Canada unit for C\$7.7 billion
2006-10-23 Dow ends at record high; Texas Instruments off late
2006-10-23 FDA rejects antibiotic from Replidyne, Forest
2006-10-23 Oil tumbles below \$59
2006-10-23 Wrigley goes beyond family for first time for CEO
2006-10-23 Ford posts biggest loss in 14 years
2006-10-23 FACTBOX-Key facts about Enron's Skilling
2006-10-23 Kraft 3rd-quarter profit up, but sales disappoint
2006-10-23 CORRECTED: American Express quarterly earnings rise
2006-10-23 American Express profit up as customers increase
2006-10-23 TI wireless chip sales, outlook disappoint
2006-10-23 Weatherford profit jumps 78 percent
2006-10-23 Saudi Arabia tells Japan to cut its Nov crude supply
2006-10-23 Bombardier could launch new jet this year: analyst

Table 3.1.1 Original Financial News Headlines for 2006-10-23

Preprocessed Financial News Headlines for 2006-10-23

2006-10-23 ['australia', 'foster', 'say', 'confid', 'target']
 2006-10-23 ['macquari', 'bid', '14', 'bln', 'meter', 'firm', 'techem']
 2006-10-23 ['daimlerchrysl', 'talk', '2', 'firm', 'small', 'car', 'paper']
 2006-10-23 ['solid', 'earn', 'may', 'keep', 'ralli', 'go']
 2006-10-23 ['u', 'ventur', 'investor', 'bet', 'energi', 'web', '20']
 2006-10-23 ['shell', 'canada', 'jump', '30', 'pct', 'royal', 'dutch', 'offer']
 2006-10-23 ['kimberlyclark', 'see', 'higher', 'cost', 'year']
 2006-10-23 ['ford', 'wont', 'sell', 'ford', 'credit']
 2006-10-23 ['ford', 'see', 'fourth', 'quarter', 'weaker', 'third']
 2006-10-23 ['hasbro', 'profit', 'beat', 'expect']
 2006-10-23 ['merril', 'lynch', 'buy', 'petri', 'parkman']
 2006-10-23 ['ford', 'ceo', 'juri', 'european', 'premium', 'brand']
 2006-10-23 ['kimberlyclark', 'profit', 'higher', 'narrow', 'view']
 2006-10-23 ['walmart', 'octob', 'sale', 'growth', 'track', 'forecast']
 2006-10-23 ['privat', 'equiti', 'group', 'form', 'tribun', 'bid', 'sourc']
 2006-10-23 ['att', 'earn', 'beat', 'expect']
 2006-10-23 ['wrigley', 'see', 'chang', 'commun', 'polici']
 2006-10-23 ['ford', 'review', 'product', 'brand', 'ceo']
 2006-10-23 ['updat', '4walmart', 'slow', 'u', 'expans', 'spend', 'growth']
 2006-10-23 ['royal', 'dutch', 'bid', 'shell', 'canada', 'low', 'sharehold']
 2006-10-23 ['shell', 'look', 'buy', 'canada', 'unit', 'c77', 'billion']
 2006-10-23 ['dow', 'end', 'record', 'high', 'texa', 'instrument', 'late']
 2006-10-23 ['fda', 'reject', 'antibiot', 'replidyn', 'forest']
 2006-10-23 ['oil', 'tumblr', '59']
 2006-10-23 ['wrigley', 'goe', 'beyond', 'famili', 'first', 'time', 'ceo']
 2006-10-23 ['ford', 'post', 'biggest', 'loss', '14', 'year']
 2006-10-23 ['factboxkey', 'fact', 'enron', 'skill']
 2006-10-23 ['kraft', '3rdquarter', 'profit', 'sale', 'disappoint']
 2006-10-23 ['correct', 'american', 'express', 'quarterli', 'earn', 'rise']
 2006-10-23 ['american', 'express', 'profit', 'custom', 'increas']
 2006-10-23 ['ti', 'wireless', 'chip', 'sale', 'outlook', 'disappoint']
 2006-10-23 ['weatherford', 'profit', 'jump', '78', 'percent']
 2006-10-23 ['saudi', 'arabia', 'tell', 'japan', 'cut', 'nov', 'crude', 'suppli']
 2006-10-23 ['bombardi', 'could', 'launch', 'new', 'jet', 'year', 'analyst']

Table 3.1.2 Preprocessed Financial News Headlines for 2006-10-23

3.2.1 Technical Indicators

In addition to historical stock price and trading volume, several technical indicators were calculated based on historical data to identify price movement trends using the Python package Talib. Previous studies highlighted the importance of technical analysis in stock price prediction and numerous indicators have demonstrated their ability to improve predictability. This study applies the most common used indicators discussed in Section 2.3.1 are Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence rules (MACD), Relative-Strength Index (RSI), William (%R), Average Directional Index (ADX), and Bollinger Bands (upper, middle, lower band). The technical indicators of historical data for first two months are shown in Table 3.2.1.1.

- **SMA (Simple Moving Average)** The SMA identifies current price trends. It calculates the average of a selected range of prices, usually closing prices, by the number of periods in that range [47].

$$SMA = \frac{P_1 + P_2 + \dots + P_n}{n} \quad (1)$$

Where P_n is the stock price at period n , n is the number of total periods.

- **EMA (Exponential Moving Average)** The EMA is similar to the Simple Moving Average (SMA), measuring trend direction over a period of time. EMA places more weight on the most recent data [47].

$$EMA = Price(t) * k + EMA(y) * (1 - k) \quad (2)$$

$$k = \frac{2}{number\ of\ days + 1} \quad (3)$$

Where t represents today, y represents yesterday, k is exponential smoothing constant.

- **MACD (Moving Average Convergence/Divergence rules)** The MACD is a trend-following momentum indicator that finds the difference between two moving averages of a stock price. The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA [48].
- **RSI (Relative-Strength Index)** The RSI gauges whether a stock is overbought (bearish) or oversold (bullish). The RSI scale is 0 to 100. Over 70 measurements are generally considered to be overbought, and measurements below 30 are considered oversold [48].

$$RSI = 100 - \left(\frac{100}{1 + RS} \right) \quad (4)$$

$$RS = \frac{\text{average gain}}{\text{average loss}} \quad (5)$$

- **William (%R)** The Williams %R measures overbought and oversold levels and generate entry signals when the price moves out of the overbought or oversold point. It compares a stock's closing price to the highest-lowest price range over a specific period, typically 14 days. Williams %R moves between 0 and -100, when it is above -20 is overbought, below -80, it is oversold [49].

$$Williams\%R = \frac{Highest - Close}{highest - Lowest} \quad (6)$$

Where *Highest* represents the highest price in 14 days, *Lowest* is the lowest price in 14 days, *Close* represents the most recent closing price.

- **ADX (Average Directional Index)** The ADX is used to indicate the strength of a trend. If positive directional indicator (+DI) is the higher number, market direction is up; if negative indicator (-DI) is the greater number, market direction is down. A value of

ADX below 20 indicates the trend is weak; a value over 40 indicates a strong trend [48].

$$+DI = \left(\frac{\text{smoothed} + DM}{ATR} \right) * 100 \quad (7)$$

$$-DI = \left(\frac{\text{smoothed} - DM}{ATR} \right) * 100 \quad (8)$$

$$DX = \left(\frac{|+DI - -DI|}{|+DI + -DI|} \right) * 100 \quad (9)$$

$$ADX = \frac{(\text{prior } ADX * 13) + \text{current } ADX}{14} \quad (10)$$

$$+DM = \text{current high} - \text{previous high} \quad (11)$$

$$-DM = \text{previous low} - \text{previous low} \quad (12)$$

Where $+DI$ represents the positive directional indicator, $+DM$ is positive directional movements, ATR is Average True range, $-DI$ is negative directional indicator, $-DM$ is negative directional movements, DX represents the directional index.

- **Bollinger Bands (upper, middle, lower band)** Bollinger Bands generate oversold or overbought signals. There are three lines that compose Bollinger Bands: A simple moving average (middle band) and an upper and lower band [50].

$$\text{upper band} = 20\text{day SMA} + (20\text{day SD} * 2) \quad (13)$$

$$\text{middle band} = 20\text{day SMA} \quad (14)$$

$$\text{lower band} = 20\text{day SMA} - (20\text{day SD} * 2) \quad (15)$$

Where SD represents the standard deviation.

Date	SMA_10	EMA_10	ADX	MACD	RSI	BBANDS_U	BBANDS_M	BBANDS_L	WILLR
2006-10-20									
2006-10-23									
2006-10-24									
2006-10-25									
2006-10-26									
2006-10-27									
2006-10-30									
2006-10-31									
2006-11-01									
2006-11-02	52.19307709	52.19307709							
2006-11-03	52.46461563	52.4446855							
2006-11-06	52.9707695	53.11907845							
2006-11-07	53.26923103	53.56595935							
2006-11-08	53.61230774	54.06305751							-2.316610449
2006-11-09	54.0907692	54.61103312			77.33421734				-5.390530892
2006-11-10	54.61000023	54.91951682			72.08207275				-16.76038205
2006-11-13	55.15384636	55.24603839			73.1238912				-10.79863354
2006-11-14	55.48307724	55.50200354			72.68298878				-11.69852432
2006-11-15	55.83846207	55.71422679			72.72726154				-11.47355162
2006-11-16	56.05153885	55.67247925			64.11115502	58.42715323	54.12230797	49.81746271	-29.73286488
2006-11-17	56.41230812	55.94741322			69.66521882	58.66704338	54.43846188	50.20988037	-5.747106781
2006-11-20	56.62923126	56.3793523			72.7109649	59.05872798	54.80000038	50.54127278	-3.940199631
2006-11-21	57.05538521	57.00828149			76.14442307	59.8221507	55.16230812	50.50246554	-1.20216358
2006-11-22	57.64769287	57.95642592			80.33693655	61.08512979	55.63000031	50.17487082	-1.810372395
2006-11-24	58.04384689	58.51679641			73.43268001	61.77832753	56.06730804	50.35628855	-18.91890443
2006-11-27	58.37769279	58.72213388			66.22839893	61.92940691	56.49384651	51.05828611	-38.90713249
2006-11-28	58.45615425	58.49992772			56.95310551	61.69552822	56.80500031	51.91447239	-57.20721063
2006-11-29	58.80538483	58.79924131	43.40461022		63.70294919	61.97311561	57.14423103	52.31534645	-26.2162429
2006-11-30	59.28076935	59.27630228	41.15402791		66.43782084	62.35996251	57.55961571	52.7592689	-11.26127917
2006-12-01	59.96307716	59.82746432	39.45536872		68.22405692	62.82527097	58.00730801	53.18934504	-2.654876578
2006-12-04	60.45384636	60.23925409	37.87804232		67.28508347	63.11276022	58.43307724	53.75339426	-5.132761383
2006-12-05	61.10615387	61.07687212	37.3903387		72.49706317	64.19032362	58.86769257	53.54506151	-2.787213528
2006-12-06	61.51769218	61.59995827	36.8258989		68.6790221	64.82157546	59.2865387	53.75150193	-10.67300237
2006-12-07	61.64384613	61.94262319	36.03695742	2.868001038	66.69000929	65.2903537	59.6457695	54.0011853	-14.81984047
2006-12-08	61.7384613	61.95025813	34.48789446	2.698582561	60.64372054	65.49425344	59.8911541	54.28805475	-28.07612954
2006-12-11	62.08153839	62.15510635	33.30084403	2.622229894	63.2561879	65.7430614	60.22961559	54.71616978	-18.42282875
2006-12-12	62.62000008	62.28774462	32.19858291	2.517185776	62.46997221	65.91927909	60.53807716	55.15687524	-20.12234261
2006-12-13	63.03538513	62.65360979	31.58020314	2.519108676	65.83568494	66.22940232	60.92038498	55.61136763	-7.61383021
2006-12-14	63.39461632	63.0830241	31.24688471	2.548975215	67.42577621	66.55577976	61.33769283	56.1196059	-1.29159278
2006-12-15	63.74538536	63.57981679	31.23452364	2.607144073	69.15467956	66.68325017	61.85423126	57.02521235	-0.765309027
2006-12-18	64.1284626	64.00586473	31.66787455	2.631598098	69.39021476	66.92859576	62.29115448	57.6537132	-15.5435234
2006-12-19	64.36846275	64.59500838	32.07027181	2.726311903	72.19885685	67.47765939	62.73730831	57.99695724	-0.11258513
2006-12-20	64.70923233	65.09801444	32.64131103	2.778653453	72.4363928	68.06331146	63.11346226	58.16361306	-7.110059134

Table 3.2.1.1 Technical Indicators of Historical Data for First Two Months

3.2.2 Sentiment Features

Two methods are utilized to extract sentiment features from financial news headlines. This study applies the powerful natural language preprocessing library NLTK (Natural language

Toolkit) in Python to compute the polarity score for each news headline, between $[-1.0, 1.0]$. Because the number of news published by the top 500 companies in the U.S. stock market varies each trading day. The average polarity score was calculated for each trading day and labelled negative sentiment (-1) if the score falls below zero, positive sentiment (1) if the score is above zero, and neutral sentiment (0) if the score is zero [111].

Since manually constructed dictionaries are carefully selected by linguistic experts, they are considered more accurate than automatically constructed dictionaries [32]. Therefore, this study also take the manually constructed sentiment dictionary Loughran–McDonald (LMD) financial sentiment dictionary to categorize the existing words of news headlines into seven groups. The LMD was constructed by Loughran and Bill McDonald [37] and used the 2022 version. It contains more than 4141 words, for detailed of which see Table 3.2.2. This study first matches the words in the news headlines with words under the seven sentiment categories in the LMD, and then counts the total occurrences for seven types of sentiment words in all news released on each trading day.

No.	Description	No. of words
1	Negative words	2355
2	Positive words	354
3	Uncertainty words	297
4	Litigious words	905
5	Modal words strong	19
6	Modal words weak	27
7	Constraining words	184

Table 3.2.2: Loughran-McDonald (LMD) Categories

3.3 Data Splitting

This study divides the data into training set (from 2006-10-20 to 2011-10-20), a test set (from 2011-10-21 to 2012-10-20), and a validation set (from 2012-10-21 to 2013-11-19). The separation was not random, which is a common strategy in artificial neural networks. As the data is a time series, the values are dependent on its predecessors, for example: the price of the day t depends directly on the price of the days $t-1$, therefore, so the separation of sets is carried out according to the date.

3.4 Proposed Stock Price Prediction Models

According to the literature review of previous studies in Chapter 2, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) have achieved great results in predicting stock price movement. SVM perform well in technical analysis-based studies, while LSTM is popular in sentiment-based analysis. This study evaluates and compares the performance of these two models in predicting stock prices by analyzing the joint impact of technical and sentiment information.

3.4.1 Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks, known as Support Vector Classifier (SVC) and Support Vector Regression (SVR), respectively [41]. The stock price movement prediction is to formulate the problem into a binary classification task, which is to predict whether the stock price will rise or down over a

certain time period. The SVC tries to find a hyperplane that best separates the data points of the two classes, as shown in Fig. 3.4.1.1.

This study uses a Support Vector Regression (SVR) to directly predict the future stock price. The aim of the SVR is to create a hyperplane that contains the maximum number of points for capturing the relationship between the features and the target stock price [41], as shown in Fig. 3.4.1.2

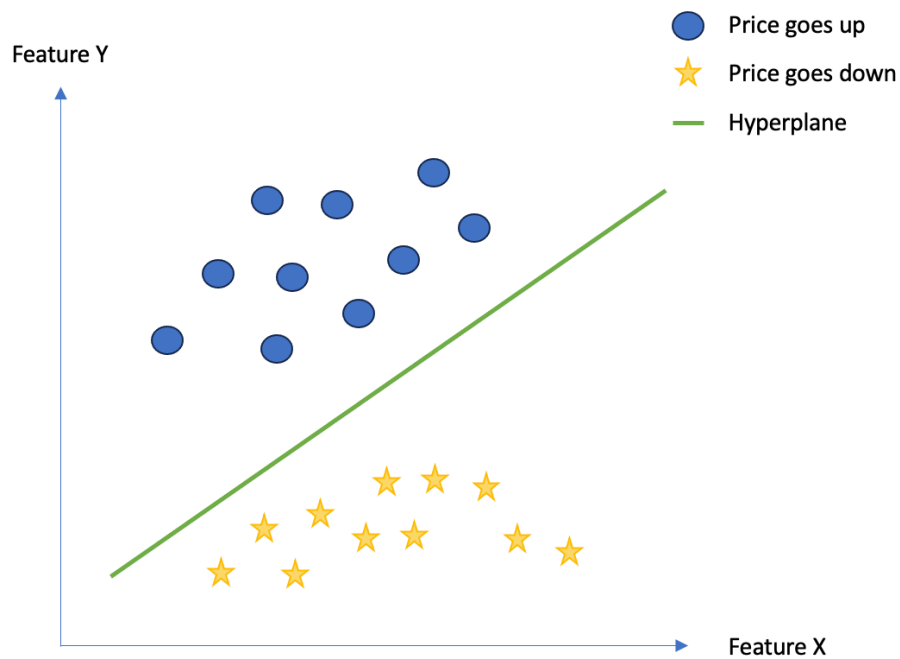


Fig.3.4.1.1 Support Vector Classifier Architecture

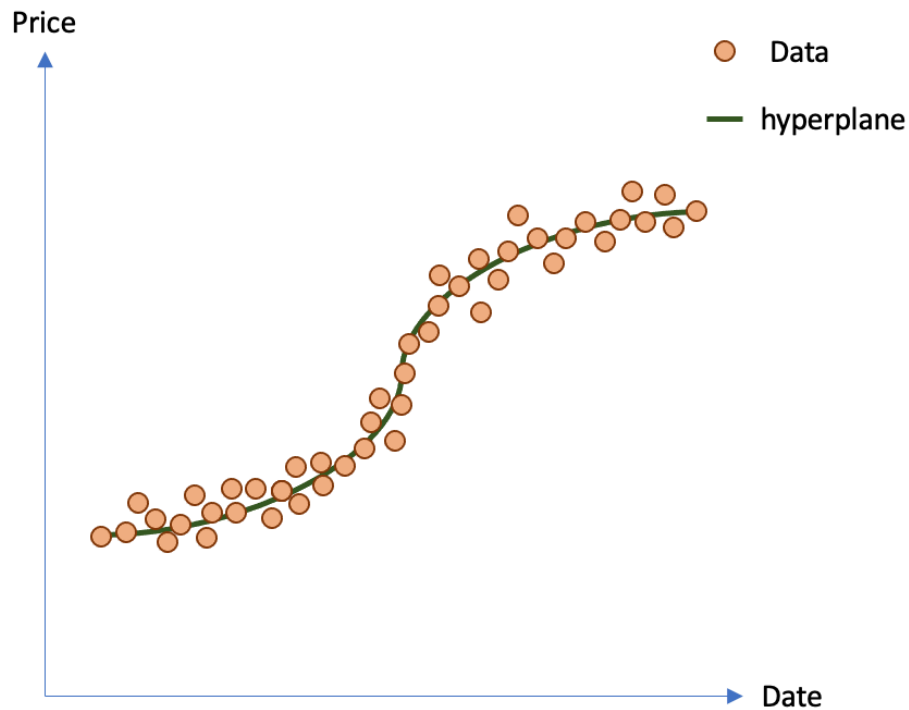


Fig.3.4.1.2 Support Vector Regression Architecture

Stock price data often exhibits non-linear patterns and trends. SVR can handle high-dimensional data effectively and can capture the non-linear relationships between a large number of input features (e.g., historical stock prices, trading volumes, technical indicators, news sentiment) and the target variable (future stock prices) by using kernel functions. This allows SVR to model complex relationships where linear regression models may not [42] [43] [44].

There can be noise in stock price data, and outliers can have a significant impact on the regression models. Compared to traditional linear regression, the SVR is less sensitive to anomalies because it focuses on adjusting the "tube" around the data points rather than trying to minimize the sum of squared errors [109] [110].

Overfitting can be a significant concern in financial forecasting. SVR allows us to set a margin of error (epsilon) that controls the trade-off between fitting the training data closely and

having a wider margin to accommodate potential errors. This provides a degree of flexibility in the modeling the stock price prediction task and can help mitigate overfitting.

In the context of stock price prediction, which has a dataset with ‘n’ data points, a feature space m. Input features are $(x_1, x_2, \dots, x_m)$, the outputs are target stock prices $(y_1, y_2, \dots, y_n)$. SVR aims to find a regression hyperplane $f(x)$ that minimizes the objective function [45]:

$$\min \left(\frac{1}{2} * ||w||^2 + c * \sum \left(\varepsilon - (Y_i - f(x_i)) \right)^2 \right) \quad (16)$$

where $||w||^2$ represents the L2-norm of the weight vector w , which controls the complexity of the model. C is a regularization parameter that controls the trade-off between fitting the training data and minimizing the margin of error. $\sum (\varepsilon - (Y_i - f(x_i)))^2$ sums up the squared errors for all data points, where ε is the tube width, Y_i is the actual stock price for data point i , and $f(x_i)$ is the predicted stock price [46].

The constraints for the SVR ensure that the predicted values fall within the tube around the actual values, which is defined as follows [45]:

$$-\varepsilon \leq Y_i - f(x_i) \leq \varepsilon, \text{ for all data points } i \quad (17)$$

SVR typically uses a kernel trick to map the input features to higher-dimensional space, allowing the algorithm to capture complex relationships between the features and the target variables. This study considers common kernel functions from the literatures [41] [42] [43] [44] [45] include linear, polynomial, radial basis function (RBF), and sigmoid kernels.

3.4.2 Multi-Layer Long Short-term Memory

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that has been applied to stock prices prediction and other time-series prediction tasks [51]. The stock price data is inherently sequential, with today's stock price depending on the

previous day's price. LSTM networks are well-suited for modeling sequential data because they have a memory cell that captures and remembers information from previous time steps, which makes them very effective in predicting time-sequences [52] [108]. This is important for stock price prediction, as stock prices are influenced by a wide range of factors, some of which may have effects that persist for a period of time.

Stock price movements are often nonlinear and are influenced by complex interactions between various factors. LSTMs can automatically learn relevant features from the data, and model nonlinear relationships in data, which makes them more flexible than some traditional time-series forecasting models [52]. As new data becomes available, the LSTM network can update its internal state and adjust its predictions accordingly [51]. This adaptability is crucial in dynamic financial markets.

Fig. 3.4.2.1 illustrates the ordinary LSTM model at time t which is composed of three layers, which are the input layer, LSTM layer, and output layer. Each LSTM layers consists of multiple LSTM units (also called cells).

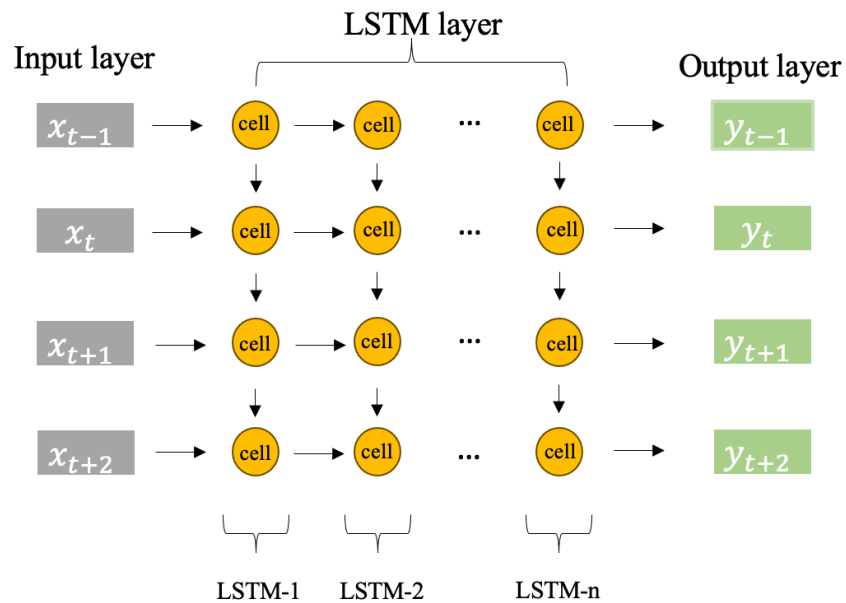


Fig. 3.4.2.1 Multi-Layer LSTM

LSTM units have three gates: the input gate i , the forget gate f , and the output gate o [53]. Fig. 3.4.2.2 shows the LSTM cell architecture.

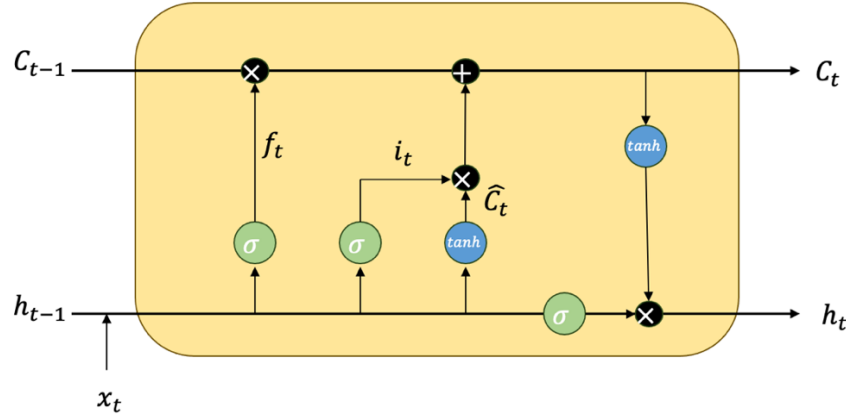


Fig. 3.4.2.2 LSTM Cell Architecture

The forget gate determines what information should be retained or forget. The forget gate takes the current input value x_t and the previous hidden state h_{t-1} to compute a value between 0 and 1 by a *sigmoid* function. The values near 0 are considered as information is not relevant and should be discard, the values near 1 are considered as valuable information. The equation for the forget gate f_t is given by:

$$f_t = \sigma(w_f * [h_{t-1}, x_t] + b_f) \quad (18)$$

Where σ is the *sigmoid* activation function. w_f is the weight matrix for the forget gate. $[h_{t-1}, x_t]$ represents the concatenation of the previous hidden state and the current input. b_f is the bias vector of the forget gate [72].

The input gate defines what new information should be kept in the cell gate. It uses the current input x_t and the previous hidden state h_{t-1} to compute a new candidate cell state. The two equations for the input gate i_t are given by:

$$i_t = \sigma (w_i * [h_{t-1}, x_t] + b_i) \quad (19)$$

$$\hat{C}_t = \tanh (w_c * [h_{t-1}, x_t] + b_c) \quad (20)$$

Where i_t is the input gate. w_i is the weight matrix for the input gate. b_i represents the bias vector for input gate. $\tanh$ is the hyperbolic *tangent* activation function. $\hat{C}_t$ is the candidate cell state. w_c is the weight matrix for the candidate cell state. b_c represents the bias vector for candidate cell state [72].

The cell state is the memory of the LSTM, which updated according to the forget gate and the input gate. The equation of cell state updated is shown as follow:

$$c_t = (f_t * C_{t-1}) + (i_t * \hat{C}_t) \quad (21)$$

Where c_t is the updated cell state. c_{t-1} represents the previous cell state [72].

The output gate determines the parts of the cell state should be exposed to the next hidden state and used as the output of the LSTM. The equations for the output gate are given by:

$$o_t = \sigma (w_o * [h_{t-1}, x_t] + b_o) \quad (22)$$

Where o_t is the output gate. w_o is the weight matrix for the output gate. b_o is the bias vector for the output gate [72].

The new hidden state is generated by the updated cell state and output gate:

$$h_t = o_t * \tanh (c_t) \quad (23)$$

Where h_t represents the new hidden state [72].

This study stacks multiple LSTM layers to create a deep LSTM network in order to capture more complex patterns and dependencies in the data, thereby improving prediction accuracy. The

number of LSTM layers is a hyperparameter that can be adjusted based on the complexity of the problem. Each LSTM layer consists of multiple LSTM units that capture patterns and dependencies in the input sequence. In the multi-layer LSTM network, the output of one LSTM layer serves as the input to the next LSTM layer, creating a hierarchical representation of the input sequence that enable the network to capture long- short term dependencies in the data [107]. The final LSTM layer is often followed by a dense layer that maps the LSTM's hidden state to a single output neuron. This output neuron is used to predict the future stock price [72].

Chapter 4

Stock Price Prediction Environment

4.1 Experiment Environment Setting

Table 4.1 provides an overview of the experimental environment for this study. The experiments use the Python programming environment along with the Scikit-learn and Keras APIs. All experiments are implemented on the machine configuration as shown in Table 4.1. Normalized data were constructed with the selected features as described in the previous section. Furthermore, the necessary steps are taken to reshape and split the data into training, test, and validation data sets. The purpose of this study is to predict the closing price of the S&P 500 index with high accuracy.

Machine Configuration	Intel Core i7 64bits with 16 GB memory laptop
Environment Architecture	Python 3.8, Scikit-learn, and Keras APIs, SVR model, Multi-layer LSTM model

Table 4.1 Computing Environment Condition

4.2 Data Collection

This study is dedicated to studying the price trends of the US stock market. The S&P 500 (Standard and Poor's 500) stock market index is considered one of the best measures of U.S. stocks, which is a broad stock market index that tracks the stock performance of the 500 largest companies listed on the U.S. stock exchanges [106]. It serves as a benchmark for measuring overall market performance and usually reflects the overall market sentiment and economic trends [97]. A rise in the S&P 500 generally indicates positive economic indicator, while a decline may indicate a slowdown or challenges. By analyzing the index, investors can determine broader market trends that may affect individual stock prices.

To better understand the impact of financial news on market sentiment and investor behavior, this study collects data from three periods: before, during, and after the financial crisis of 2008. These periods include significant market events, and analyzing predictions in this context can help determine how effectively models handle volatile markets and recoveries.

The financial news source for the 500 largest corporations listed on U.S. stock exchanges was obtained from [36] for this study. The data source includes 106,494 financial news from Reuters News corresponding the period from October 20, 2006, to November 20, 2013. Each article consists of its headline, content, and publishing date. This study uses only news headlines, which are concise and focused, often summarize the most crucial information of the news. This makes them easier to process and analyze for sentiment. In contrast, full news articles can contain a lot of additional, possibly irrelevant information that can dilute or obscure the key sentiment. Analyzing news headlines is computationally less intensive than processing the entire articles. This efficiency is crucial in the fast-paced world of stock trading where decisions must be made quickly based on the latest information. Moreover, previous studies have shown that news

headlines alone can be strong indicators of market sentiment and can have predictive power regarding stock price movements [18] [19] [33] [38]. Another data source required for this study is historical S&P 500 prices and trading volume. The dataset was downloaded from finance.yahoo.com¹ using the python package yfinance for the period of 2006.10.20 to 2013.11.20. The historical data of first two months are shown in Table 4.2.1.

Date	Open	High	Low	Close	Adj Close	Volume
2006-10-20	51	51.1230774	50.6846161	50.8615379	24.4390221	519740
2006-10-23	50.4615364	51.2076912	50.4615364	51.092308	24.5499058	374140
2006-10-24	52.0384636	52.6461525	51.7769241	52.592308	25.2706566	478530
2006-10-25	52.5999985	53.2461548	52.4923096	52.8692322	25.4037189	344760
2006-10-26	52.5384636	52.7538452	51.7000008	52.2923088	25.1265106	270010
2006-10-27	50.7692299	51.5076904	50.6153831	51.1153831	24.5609932	587990
2006-10-30	51.1230774	51.453846	50.8307686	51.2769241	24.6386147	543530
2006-10-31	53.2307701	53.5384636	52.7692299	53.3615379	25.6402702	882570
2006-11-01	53.3076935	53.4384613	52.9000015	53.1153831	25.5219917	398970
2006-11-02	52.8923073	53.592308	52.8846169	53.3538475	25.6365757	341120
2006-11-03	53.6538467	53.9846153	53.4615364	53.5769234	25.7437611	486980
2006-11-06	55.3846169	56.1538467	55.2692299	56.1538467	26.9819832	798330
2006-11-07	55.5999985	55.8384628	55.3923073	55.5769234	26.704771	284180
2006-11-08	55.453846	56.4384613	55.4307709	56.2999992	27.0522022	252200
2006-11-09	56.9230766	57.453846	56.7000008	57.0769234	27.4255123	309530
2006-11-10	56.5384636	56.8230782	56.1538467	56.3076935	27.0559006	349700
2006-11-13	56.3076935	56.8461533	56.1615372	56.7153854	27.2518005	271310
2006-11-14	56.5384636	56.7384605	56.3615379	56.6538467	27.2222271	393250
2006-11-15	56.5384636	56.9153862	56.4307709	56.6692314	27.2296295	201890
2006-11-16	56.3076935	56.6538467	55.3461533	55.4846153	26.6604099	281060
2006-11-17	56.9307709	57.2461548	56.7461548	57.1846161	27.4772663	587730
2006-11-20	58.2384605	58.546154	58.0769234	58.3230782	28.0242901	543920
2006-11-21	59.2692299	59.9230766	59.1230774	59.8384628	28.7524433	411580
2006-11-22	62.1538467	62.3846169	61.6153831	62.2230759	29.8982487	530140
2006-11-24	61.5	61.5	60.9230766	61.0384636	29.3290386	296140
2006-11-27	60.453846	60.7230759	59.4769249	59.6461525	28.6600399	471250
2006-11-28	57.6923065	57.6923065	53.8461533	57.5	27.6288071	867620
2006-11-29	58.6923065	60.1461525	58.6923065	60.1461525	28.9002914	593840
2006-11-30	61.1153831	61.4230766	60.4230766	61.4230766	29.5138474	517270
2006-12-01	61.8461533	62.5384636	61.1307678	62.3076935	29.9389095	463060
2006-12-04	61.8846169	62.092308	61.3461533	62.092308	29.835413	362050
2006-12-05	64.2692337	65.1615372	64.1538467	64.8461533	31.1586342	586430
2006-12-06	64.4538498	64.6461563	63.9461555	63.953846	30.7298851	336960
2006-12-07	63.7615395	64.1999969	63.4692307	63.4846153	30.504425	309790
2006-12-08	62.6153831	62.7307701	61.9230766	61.9846153	29.7836685	677820
2006-12-11	62.6769218	63.407692	62.3461533	63.0769234	30.3085251	331110
2006-12-12	63.0846138	63.3692322	62.7000008	62.8846169	30.216116	176540
2006-12-13	63.9615364	64.4230804	63.8692322	64.3000031	30.8962116	337220
2006-12-14	64.384613	65.0769196	64.3230743	65.0153885	31.2399559	206310
2006-12-15	65.5384598	65.907692	65.384613	65.8153839	31.6243572	203190
2006-12-18	66.407692	67.2538452	65.5307693	65.9230804	31.6761074	413920
2006-12-19	65.9153824	67.2461548	65.6923065	67.2461548	32.3118362	304460
2006-12-20	67.6923065	67.8384628	67.1846161	67.3615417	32.3672867	250250

Table 4.2.1 Historical Stock Price and Trading Volume of First Two Months

¹ <https://finance.yahoo.com/quote/^GSPC?p=^GSPC&.tsrc=fin-srch>

4.3 Evaluation Metrics

Assessing the predictive ability of machine learning models to predict stock price is essential to assess their performance accurately. Dattatray and Kumar (2019) concluded, after reviewing 50 research papers on stock price prediction area that the most commonly used performance metrics are Mean Average Percentage Error (MAPE), Root Mean Squared Error (RMSE), accuracy, Mean Squared Error (MSE), Mean Absolute Error (MAE), precision, and R-squared score (R2 score) [56]. This study uses four of the following indicators:

- **Mean Squared Error (MSE):**

The MSE measures the average squared difference between the predicted and actual stock price. It gives more weight to larger errors and can help identify outliers. The lower the MSE, the better the model performance [11].

$$MSE = \frac{1}{n} (y_i - \hat{y}_i)^2 \quad (24)$$

Where n is the number of data points. y_i is the actual stock price at time i . $\hat{y}_i$ is the predicted stock price at time i .

- **Mean Squared Error (RMSE):**

The RMSE is the square root of the MSE, which provides a measure of the average error in the same units as the target variable. It defines the gap between the predicted stock price and the actual price. The lower the RMSE, the better the model performance [54].

$$RMSE = \sqrt{MSE} \quad (25)$$

- **Mean Absolute Error (MAE):**

The MAE is a measure of average absolute difference between predicted and actual stock price. A lower MAE indicates better model performance [11].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (26)$$

Where n is the number of data points. y_i is the actual stock price at time i . $\hat{y}_i$ is the predicted stock price at time i .

- **R-Squared score (R^2 score):**

The R^2 score measures how close the data are to the fitted regression line. It ranges from 0 to 1, where a higher value indicates a better fit [11].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (27)$$

Where n is the number of data points. y_i is the actual stock price at time i . $\hat{y}_i$ is the predicted stock price at time i . $\bar{y}$ is the mean of the actual stock prices.

Chapter 5

Results and Analysis

5.1 Baseline Model

To provide a comprehensive comparison of the role of technical and sentiment information in predicting stock prices over 1 day ahead, this study uses SVR and multi-layer LSTM model that contain only technical information as baseline models. The features include historical prices (opening, closing, highest, lowest price, adjusted closing price), volume, and technical indicators mentioned in section 4.3.1. Normalized data were constructed with the selected features as described in the previous section. Necessary steps are also taken to split data into training, test, and validation datasets.

To evaluate the performance of the baseline models, this study uses standard evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared score to measure the prediction error between the baseline model's predictions and the actual stock prices in the validation set.

The prediction results of the baseline models over the validation datasets are shown in Appendix H and Appendix I, respectively, in Fig.5.1.1 and Fig.5.1.2, where the x-axis represents the date of trading, and the y-axis represents the adjusted closing price of the S&P 500 index. It

can be seen in Table 5.1.1 that the multi-layer LSTM model obtained a better prediction accuracy, RMSE:0.591, MSE: 0.349, MAE:0.448, and R-squared score: 0.651.

	RMSE	MSE	MAE	R-squared score
SVR	0.669	0.447	0.509	0.553
Multi-layer LSTM	0.591	0.349	0.448	0.651

Table 5.1.1 Evaluation Results of Baseline Models

The period from 2010 to 2013 was the economic recovery after the financial crisis of 2008. Driven by factors including the Federal Reserve's monetary stimulus program, the recovery of corporate earnings, and progressively bettering economic conditions, the S&P 500 Index was significantly affected and continued to rise. The index closed 2013 at about 1,848 points, having surpassed its pre-crisis peak by then [97]. Due to the complexity, noise, and volatility of stock price movement, SVR and Multi-layer LSTM models based on only historical and technical information are insufficient to accurately predict stock prices. The following sections will analyze and explain the performance of models that added sentiment information extracted from financial news headlines.

5.2 Proposed Models

To investigate the joint impact of technical and sentiment information on stock price prediction, this study adds sentiment information extracted from the financial news headlines as features into the baseline models. Four separate models were created for predicting stock price

over 1 day ahead. Model-1 is the SVR model with powerful natural language preprocessing library NLTK, Model-2 is the SVR model with the Loughran–McDonald (LMD) financial sentiment dictionary, Model-3 is the Multi-layer LSTM model with NLTK, and Model-4 is the Multi-layer LSTM model with Loughran-McDonald (LMD). The following section presents the results of respective models.

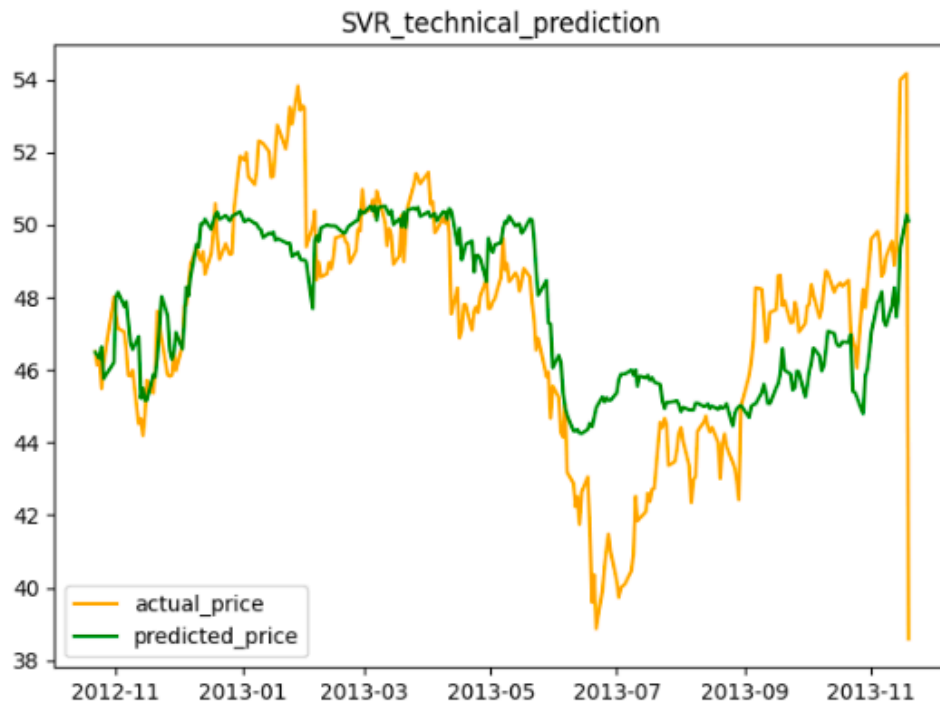


Fig. 5.1.1 Prediction Result of SVR Baseline Model with Technical Information

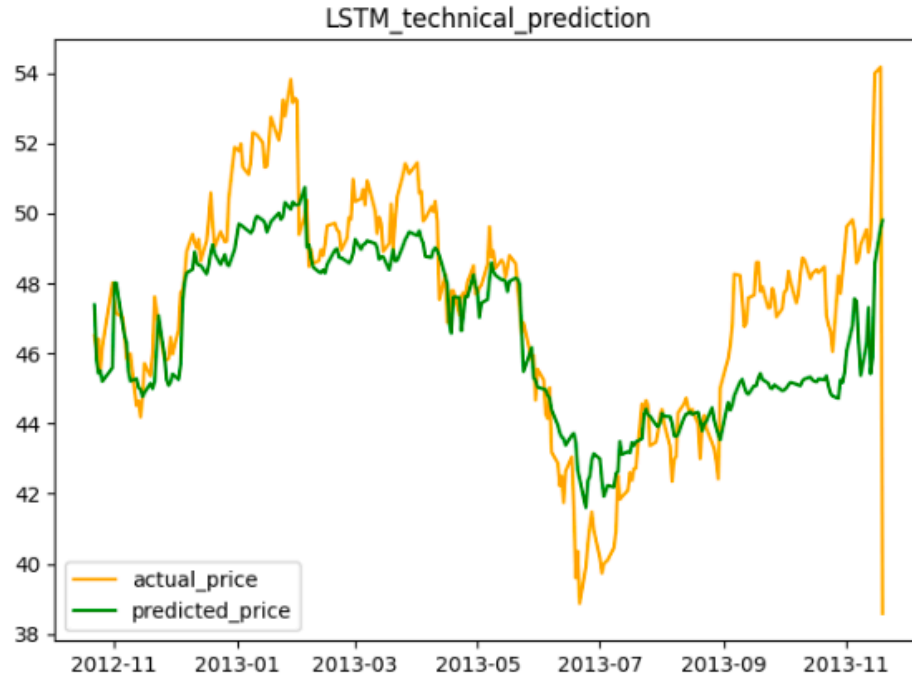


Fig. 5.1.2 Prediction Result of Multi-Layer LSTM Baseline Model with Technical Information

5.2.1 SVR Models

The best possible model architecture is selected by tuning a wide range of hyperparameters. The overall SVR model optimization procedure is divided into four categories — (a) Kernel type (linear, RBF, polynomial, sigmoid, etc.), (b) Regularization parameter (C), (c) Epsilon (ϵ) for ϵ -insensitive loss, and (d) Kernel-specific hyperparameters (e.g., gamma for RBF, degree for polynomial) [45] [103] [104] [105].

Since the kernel function determines how the Support Vector Regression model to capture the non-linear relationship in data, selecting the appropriate kernel is crucial as it can significantly impact the performance of a SVR model. Stock price prediction is considered as a non-linear problem due to it is influenced by various factors, including political events, financial news,

economic activities, etc. To address the non-linearity of stock price prediction, this study turns to non-linear kernel functions: Radial Basic Function (RBF), Polynomial Kernel, Sigmoid Kernel.

The regularization parameter (C) controls regularization strength of the SVR model. Regularization is used to prevent overfitting. Overfitting occurs when the value of C is too large, making the model fit training data too closely. Smaller values of C can prevent overfitting and result in stronger regularization, which encourages the model to have a simpler decision boundary with a greater margin. However, it can also cause underfitting if set too small. Thus, it is essential to determine the right balance between model complexity and performance of fit to the training data. This study set the value of C in [0.001, 0.01, 0.1, 0.5, 1, 10].

Gamma (γ) is a hyperparameter that affects the shape of the decision boundary in the RBF kernel. Smaller values of gamma will result in a wider and smoother shape of the decision boundary. This can help prevent overfitting but can lead to underfitting if set too small. Larger values of gamma result in a narrower and more complex decision boundary that may lead to overfitting if it set too large. In the scikit-learn SVR model, this study set the gamma to either 'scale' or 'auto.' The gamma value will be automatically determined based on the characteristics of our data. When the value of gamma is set to 'scale,' scikit-learn automatically calculates the value of gamma as

$$\gamma = \frac{1}{\text{the number of features}} \quad (28)$$

When the value of gamma is set to 'auto,' scikit-learn automatically calculates the value of gamma as

$$\gamma = \frac{1}{\text{number of samples}} \quad (29)$$

In SVR, epsilon (ϵ) represents the permissible deviation or error of predicted regression values from actual target value. The default value is often set to 0.1 in scikit-learn, this study set it to 0.1 and other similar small value [0.1, 0.2, 0.3].

This study applies the grid search for tuning hyperparameters in SVR models. It enables us to identify the best performing set from different combinations of hyperparameters.

The best hyperparameters and evaluation results for SVR with NLTK model and SVR with Loughran–McDonald (LMD) model are listed in Table 5.2.1.1 and Table 5.2.1.2. By comparing the results of the four evaluation metrics, it can be seen the SVR with Loughran model has a better performance.

	Kernel	C	γ	ϵ
SVR with NLTK	Sigmoid	0.01	auto	0.1
SVR with LMD	Sigmoid	0.01	auto	0.1

Table 5.2.1.1 Best Combination of Hyperparameters for Two SVR Models

	RMSE	MSE	MAE	R-Squared score
SVR with NLTK	0.634	0.402	0.482	0.598
SVR with LMD	0.518	0.268	0.388	0.732

Table 5.2.1.2 Evaluation Results of Two SVR Models

Additionally, the prediction results of the SVR models over the validation datasets are shown in Appendix J and Appendix K. Fig.5.2.1.1 and Fig.5.2.1.2 visualized the prediction performance of the two models respectively. The orange and green curves represent the actual target price and the predicted stock price. The x-axis represents the trading date, and the y-axis is stock price. It can be observed the stock price predicted by SVR with LMD model is significantly closer to the actual price than SVR with NLTK model.

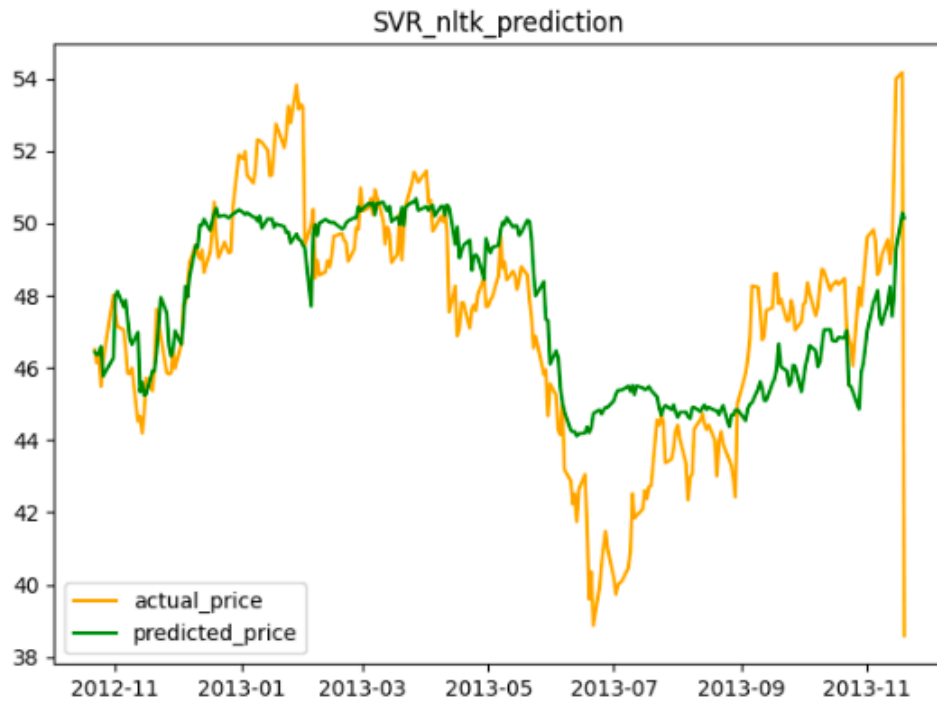


Fig. 5.2.1.1 Stock Price Prediction of SVR with NTLK

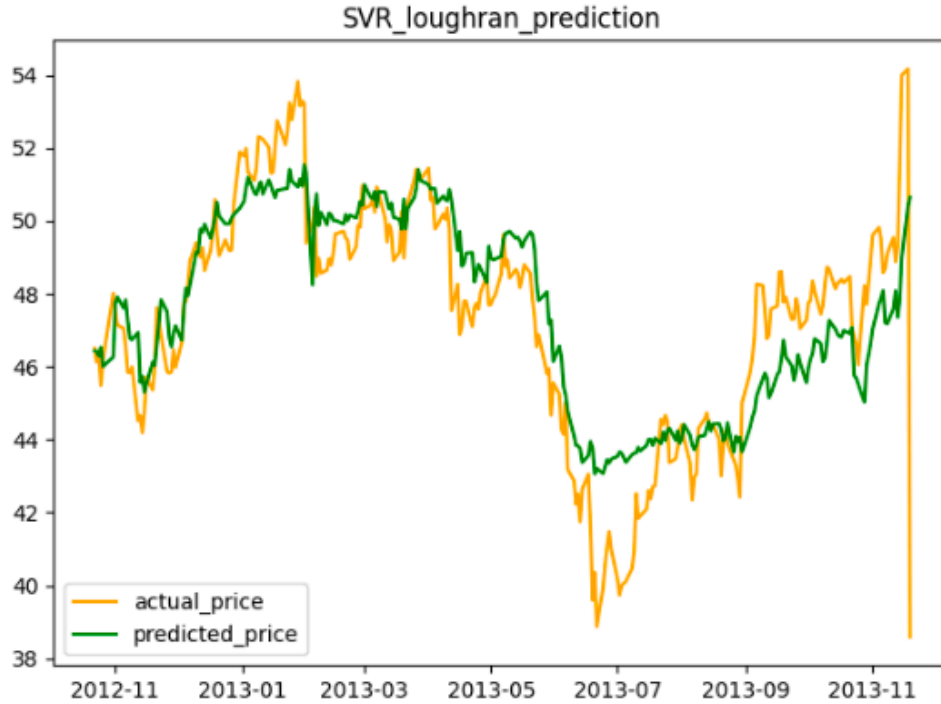


Fig. 5.2.1.2 Stock Price Prediction of SVR with LMD

5.2.2 Multi-Layer LSTM Models

Multi-layer LSTM models are trained on the training data set. The best model is selected based on the average evaluation score calculated on the test data set. Each experiment is repeated 30 times. Ultimately, the selected optimal model is assessed on the validation dataset to obtain the predictive result.

The hyperparameters of a multi-layer LSTM model exert a substantial influence on its performance. The hyperparameters considered in this study include: the number of LSTM layers, the number of LSTM neurons, dropout rate, batch size, learning rate, optimizer, epochs, and the number of dense layers [72] [100] [101] [102].

- The number of LSTM layers:

More LSTM layers help the model to learn more complex patterns but also increase computational complexity. This study set the number of LSTM layer in the range of 1 to 6.

- The number of LSTM neurons:

It decides the dimensionality of the hidden state. A large number of neurons have the capacity to capture more complex patterns but may result in issues of overfitting problem. [16, 32, 64, 128, 256] neurons are experimented in this study.

- Dropout rate:

The dropout rate pertains to the probability of a neuron being excluded during each iteration of training. Its purpose is to prevent overfitting by diminishing the interdependence between neurons. Sophisticated models necessitate a large dropout rate for regularization. This study used the common dropout rate in the range of 0.2 to 0.5.

- Batch size:

Batch size is how many samples are used for each training batch. Small batch sizes result in faster convergence but may produce noisier gradients. Large batch sizes can smooth gradient updates but require more memory. This study set it in the range of [4, 16, 32, 64]

- Learning rate:

The learning rate determines the update of the weight and bias when the gradient decreases during training. If the learning rate is set too high, it leads to exceeding the optimal weights; while the learning rate is set too low, and the convergence rate is low. This study set it as [0.01, 0.001]

- **Optimizer:**

The optimizer affects the training speed and convergence. This study employed the most common Adam optimizer in deep learning, because it can dynamically adjust the learning rates for each parameter during training, making it suitable for many tasks.

- **Epochs:**

An epoch represents a complete pass through the entire training dataset. The number of epochs refers to the number of times the model will iterate the entire training dataset. This study set it in a range of [11, 20, 30, 40, 50, 60, 70]

- **The Number of Dense layers:**

Adding one or more dense layers after the LSTM allows the network to learn complex relationships and patterns in the training data that may not be captured only by LSTM layers. This study set it in a range of 0 to 5.

The best hyperparameters and evaluation results for multi-layer LSTM with the NLTK model and multi-layer LSTM with the Loughran–McDonald (LMD) model are listed in Table 5.2.2.1 and Table 5.2.2.2. By comparing the results of four evaluation metrics, multi-layer LSTM with Loughran model performances better.

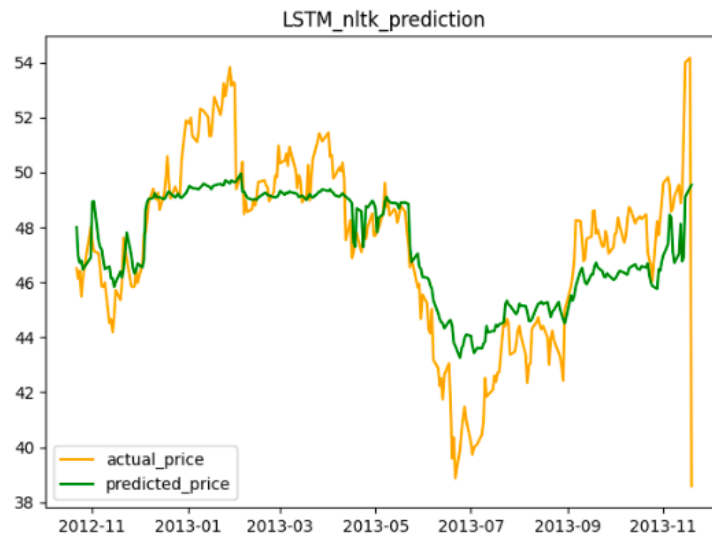
	Number of LSTM layers	Number of LSTM neurons	Dropout rate	Learning rate	Batch Size	Optimizer	Epochs	Number of Dense Layers
Multi-layer LSTM with NLTK	5	256	0.5	0.001	32	Adam	32	2
Multi-layer LSTM with LMD	5	256	0.5	0.001	32	Adam	15	2

Table 5.2.2.1 Best Hyperparameters of Multi-Layer LSTM

	RMSE	MSE	MAE	R-Squared score
Multi-layer LSTM with NLTK	0.573	0.328	0.439	0.672
Multi-layer LSTM with LMD	0.523	0.247	0.367	0.726

Table 5.2.2.2 Evaluation Results of Multi-Layer LSTM

For Multi-Layer LSTM models, see Appendices L and M for predictions on validation datasets. The comparison plot of the actual value versus the predicted value of closing price for validation dataset are plotted in Fig.5.2.2.1 and Fig.5.2.2.2 respectively. The orange and green curves represent the actual target price and the predicted stock price. The x-axis represents the trading date, and the y-axis is stock price. Comparing the line charts, it can be seen that the multi-layer LSTM using the LMD model is significantly more accurate in predicting stock prices than the multi-layer using the NLTK model.

**Fig. 5.2.2.1 Stock Price Prediction of Multi-Layer LSTM with NLTK Model**

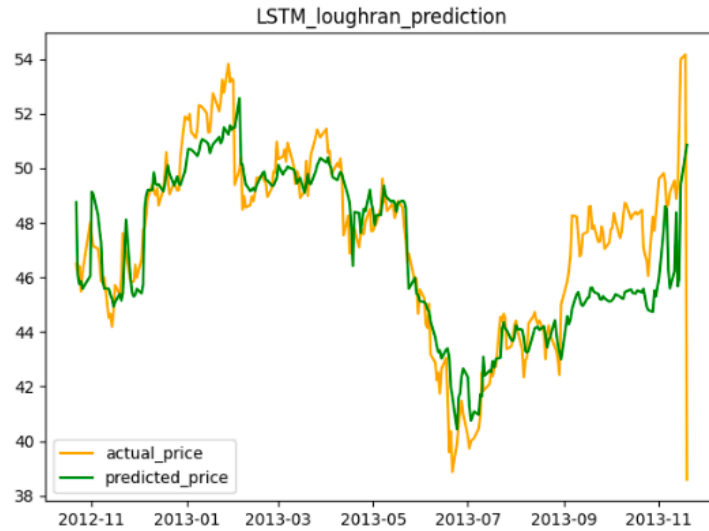


Fig. 5.2.2.2 Stock Price Prediction of Multi-Layer LSTM with LMD Model

5.3 Comparison of Baseline and Proposed Models

As mentioned earlier, this study has implemented models based solely on technical information and models that combine technical and sentiment information to predict stock price for the next day. This section will analyze the performance of these architectures by comparing the evaluation metrics obtained from experiments. The result of table 5.3 found that models using both technical and sentiment features are preferable to models based solely on technical information. This implies that sentiment information extracted from financial news headlines is beneficial for predicting stock price, while the combination of technical and sentiment features improves the accuracy of stock price prediction. Furthermore, the models using the Loughran–McDonald (LMD) financial sentiment dictionary have obtained better prediction accuracy than models with the automatic natural language preprocessing tool NLTK. SVR with the LMD model achieved better result in RMSE and R-squared score while the Multi-layer LSTM with the LMD model performs

better in terms of MSE and MAE. RMSE and R-squared indicate that SVR with the LMD model is better at capturing the overall variance and trends in the data but more sensitive to outliers. MAE and MSE suggest that Multi-layer LSTM with the LMD model produces more consistent, small errors, making it better at minimizing the average error but less focused on outlier errors.

	RMSE	MSE	MAE	R-Squared score
SVR with technical	0.669	0.447	0.509	0.553
Multi-layer LSTM with technical	0.591	0.349	0.448	0.651
SVR with NLTK	0.634	0.402	0.482	0.598
SVR with LMD*	0.518*	0.268	0.388	0.732*
Multi-layer LSTM with NLTK	0.573	0.328	0.439	0.672
Multi-layer LSTM with LMD	0.523	0.247*	0.367*	0.726

Table 5.3 Comparison of Evaluation Results for Prediction Models

Chapter 6

Conclusion and Future Directions

6.1 Conclusion

This study analyzes the joint impact of news headline sentiment and technical information on S&P 500 stock price around the financial crisis of 2008. Firstly, computing technical indicators. Secondly, constructing the sentiment dimensions using the Python natural language preprocessing library NLTK (Natural language Toolkit) and the Loughran–McDonald Financial Sentiment Dictionary. Then, implementing the SVR and LSTM models, assessing their prediction accuracy and comparing their performance on different commonly used evaluation metrics. This study experiments with the historical prices of the S&P 500 stock market index and news headlines corresponding before, during, and after the financial crisis of 2008, and the results showed:

1. When the economic situation changes unpredictably, the government releases major decisions, or under other extreme conditions, it is impossible to predict stock prices based solely on technical information.
2. Sentiment analysis of financial news headlines does help improve the prediction accuracy. The models with both sentiment and technical analysis outperform the model based on technical analysis only.

3. Simply focusing on positive and negative dimensions could not bring useful predictions. More precise sentiment classification can improve stock price prediction.
4. Manually constructed Loughran-McDonald financial sentiment dictionary performs better than automatically natural language preprocessing tool in stock price prediction.
5. Proposed models enable investors to use such predictions to identify potential market trend, adopt proactive actions for future uncertainties and make informed investment decision in varying market conditions.
6. Decision-makers can use such predictions to implement interventions to stabilize the financial market and assess their past crisis management strategies to make adjustment for the future condition.
7. Financial institutions can better manage their risk and avoid insolvency by predicting market movements.

6.2 Future Directions

Given the ever-increasing speed of trading, this study considered analyzing the joint impact of sentiment and technical information in intra-day stock price change. However, it is more technically difficult to capture intra-day trading information.

Using the S&P 500 as basis for predictions has certain limitations. The index is limited scope, and its prediction may be biased for smaller companies or specific sectors that are underrepresented. Moreover, the market performance of individual stocks can significantly outperform or underperform the overall market trend, depending on factors such as the company's

own earnings, management changes, or industry innovations, which are often not fully reflected in the S&P 500 index. Because the impact of any single company's performance is diluted by the other companies in the index, market measures such as the S&P 500 are typically less volatile than individual stocks. Therefore, improving the prediction accuracy of the stock prices of smaller company will be another future direction.

Since the Loughran–McDonald Financial Sentiment Dictionary is manually built, the expansion of dictionaries and other techniques for sentiment analysis will also improve the effectiveness of stock price prediction [112].

Although news headlines are effective materials for sentiment analysis, their information is limited for small companies. Complete news always provides the depth and context. Hence, to gain a more comprehensive understanding, combining headline analysis and content analysis can be beneficial when predict small companies' stock price.

Economic indicators offer valuable insights into the broader economic conditions that impact corporate profitability, investor sentiment, and, ultimately, stock prices. Analyzing these indicators can increase the effectiveness and accuracy of predictive models to help investors make informed predictions and decisions regarding stock investments.

Despite the advantage of LSTM networks and SVR algorithm, stock prices prediction is still an extremely challenging task, and its success cannot be guaranteed. Financial markets are influenced by a variety of factors, including economic, political, and psychological factors, which make them inherently noisy and unpredictable. Moreover, past performance is not necessarily indicative of future results, so the accuracy of stock price predictions can vary significantly. LSTM and SVR models are just two of many tools in the field of quantitative finance and algorithmic

trading, and the effectiveness of other machine learning models in stock price prediction is also worth further study.

In the future, the proposed ideas can be evaluated on a wider range of datasets and expanded the applications to multiple proposed frameworks and methods to diverse practical domains, such as biomedical and chemical IR research [86] [87] [88] [89] [90], recommendation systems [91] [92], and question answering systems [93], query expansion for IR [95] and Web usage mining [96].

References

- [1] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics.," *international Journal of Information Management*, vol. 35, no. 2, pp. 137-144, 2015.
- [2] D. Turner, M. Schroeck and R. Shockley, "Analytics: The real-world use of big data in financial services," IBM Global Business Services, New York, 2013.
- [3] E. F. Fama, "The Behavior of Stock-Market Prices," *Journal of Business*, vol. 38, no. 1, pp. 34-105, 1965.
- [4] Q. LI, Y. Chen, J. Wang, Y. Chen and H. Chen, "Web Media and Stock Markets: A Survey and Future Directions from a Big Data Perspective," in *IEEE Transactions on Knowledge and Data Engineering*, 2018.
- [5] N. Godbole, M. Srinivasaiah and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs," in *International Conference on Weblogs and Social Media*, Colorado, 2007.
- [6] A. V. Girija, M. P. M. M, R. M. Pai and A. Nayak, "Stock market prediction: A big data approach," in *IEEE*, Macao, 2015.
- [7] K. A. Althelaya, "Combining Deep Learning and Multiresolution Analysis for Stock Market Forecasting," in *IEEE*, 2021.
- [8] K. Chandar and M. K. Sharda, "A REVIEW ON STOCK MARKET PREDICTION TECHNIQUES," *INTERNATIONAL JOURNAL OF INFORMATION AND COMPUTING SCIENCE*, vol. 6, no. 2, pp. 74-79, 2019.

- [9] Fama, F. Eugene, "Efficient market hypothesis." *Diss. PhD Thesis, Ph. D. dissertation*, 1960.
- [10] B. G. Malkiel, *A Random Walk down Wall Street*. NEW YORK: NORTON, 1975.
- [11] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artificial Intelligence Review*, vol. 53, no. 4, pp. 3007–3057, 2019. doi:10.1007/s10462-019-09754-z
- [12] M. Agrawal, P. K. Shukla, and A. U. Khan, "Stock price prediction using technical indicators: A predictive model using optimal deep learning," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 2297–2305, 2019. doi:10.35940/ijrte.b3048.078219
- [13] S. Madge, "Predicting Stock Price Direction using Support Vector Machines," Princeton University, 2015
- [14] J. Kaur and K. Dharni, "Predicting Daily Returns of Global Stocks Indices: Neural Networks vs Support Vector Machines," *Journal of Economics, Management and Trade*, vol. 24, no. 6, pp. 1-13, 2019.
- [15] S. S. Hasan, R. Rahman, N. Mannan, H. Khan, J. N. Moni, and R. M. Rahman, "Improved stock price prediction by integrating data mining algorithms and technical indicators: A case study on Dhaka stock exchange," *Computational Collective Intelligence*, pp. 288–297, Sep. 2017.
- [16] R. Goonatilake and S. Herath, "The Volatility of the Stock Market and News," *International Research Journal of Finance and Economics*, no. 11, pp. 53-65, 2007.
- [17] P. C. TETLOCK, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007. doi:10.1111/j.1540-6261.2007.01232.x

- [18] G. Jariwala, H. Agarwal, and V. Jadhav, "Sentimental analysis of news headlines for stock market," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020. doi:10.1109/inocon50539.2020.9298333
- [19] L. Nemes and A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines," *Journal of Information and Telecommunication*, vol. 5, no. 3, pp. 375–394, Feb. 2021. doi:10.1080/24751839.2021.1874252
- [20] J. Han, M. Kamber, and J. Pei, "Data Mining Trends and Research Frontiers," *Data Mining*, pp. 585–631, 2012.
- [21] K. Joshi, B.H.N. and J. Rao, "STOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS," 7 July 2016. [Online]. Available: <https://arxiv.org/pdf/1607.01958.pdf>.
- [22] D. G. ía, "Sentiment during recessions," *The journal of Finance*, vol.68, no.3, pp.1267- 1300, 2013.
- [23] Y. Shynkevich, T. McGinnity, S. Coleman and A. Belatreche, "Predicting Stock Price Movements Based on Different Categories of News Articles," in *IEEE*, 2015.
- [24] L. A. Teixeira and A. L. de Oliveira, "A method for automatic stock trading combining technical analysis and nearest neighbor classification," *Expert Systems with Applications*, vol. 37, no. 10, pp. 6885–6890, 2010. doi:10.1016/j.eswa.2010.03.033
- [25] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5311–5319, 2011. doi:10.1016/j.eswa.2010.10.027
- [26] O. Hegazy, O. S. Soliman, and M. A. Salam, "A machine learning model for stock market prediction," *International Journal of Computer Science and Telecommunications*, vol. 4, no. 12, pp. 17–23, Dec. 2013.

- [27] DOSDOĞRU, Ayşe Tuğba, et al. "Assessment of hybrid artificial neural networks and metaheuristics for stock market forecasting." *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 27.1 (2018): 63-78.
- [28] H. Lee, M. Surdeanu, B. MacCartney, & D. Jurafsky, "On the Importance of Text Analysis for Stock Price Prediction," In *LREC*, 2014, pp. 1170-1175.
- [29] G. Fung, J. Yu, W. Lam, News sensitive stock trend prediction, *Adv. Knowl. Discov. Data Min.* (2002) 481–493.
- [30] D. Wu, G. Fung, J. Yu, Z. Liu, Integrating multiple data sources for stock prediction, *Web Inform. Syst. Eng. (WISE)* 2008 (2008) 77–89.
- [31] D. Wu, G. Fung, J. Yu, Q. Pan, Stock prediction: an event-driven approach based on bursty keywords, *Frontiers Comput. Sci. China* 3 (2009) 145–157.
- [32] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 14–23, 2014. doi:10.1016/j.knosys.2014.04.022
- [33] D. K. Kirange, and Ratnadeep R. Deshmukh. "Sentiment Analysis of news headlines for stock price prediction." *Composoft, An International Journal of Advanced Computer Technology*, 2016, pp, 2080-2084.
- [34] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis", *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications*, 2019, pp. 205–208
- [35] A. Heiden and R. S. Parpinelli, "Applying LSTM for stock price prediction with sentiment analysis," *Anais do 15. Congresso Brasileiro de Inteligência Computacional*, 2021. doi:10.21528/cbic2021-45

- [36] X. Ding, Y. Zhang, T. Liu, and J. Duan, “Using structured events to predict stock price movement: An empirical investigation,” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. doi:10.3115/v1/d14-1148
- [37] T. Loughran and B. McDonald, “Loughran-McDonald master Dictionary W/ sentiment word lists,” Software Repository for Accounting and Finance, <https://sraf.nd.edu/loughranmcdonald-master-dictionary/> (accessed May 27, 2023).
- [38]. M. R. Vargas, B. S. de Lima, and A. G. Evsukoff, “Deep learning for stock market prediction from financial news articles,” *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2017. doi:10.1109/civemsa.2017.7995302
- [39] J. Xu and T. Murata, “Stock Market Trend Prediction with Sentiment Analysis based on LSTM Neural Network”, *International Multiconference of Engineers and Computer Scientists*, 2019
- [40] S. Wu, Y. Liu, Z. Zou, and T.-H. Weng, “S_I_LSTM: Stock price prediction based on multiple data sources and sentiment analysis,” *Connection Science*, vol. 34, no. 1, pp. 44–62, 2021. doi:10.1080/09540091.2021.1940101
- [41] Chih-Wei. Hsu, Chih-Chung. Chang, Chih-Jen. Lin, “A Practical Guide to Support Vector Classification,” Initial version: 2003, Last updated version: 2010
- [42] U. Thissen, R. van. Brakela, A.P. de. Weijerb, W.J. Melssena, L.M.C. Buydensa, “Using support vector machines for time series prediction,” In: *Chemometrics and Intelligent Laboratory Systems* 69, pp.35– 49, 2003
- [43] Lijuan. Cao, “Support vector machines experts for time series forecasting,” In: *Neurocomputing* 51, pp. 321 – 339, 2003
- [44] J. Smola. Alex, SchoLkop. Bernhard, “A tutorial on support vector regression,” In: *Statistics and Computing* 14, pp.199–222, 2004

- [45] B. M. Henrique, V. A. Sobreiro, and H. Kimura, “Stock price prediction using support vector regression on daily and up to the minute prices,” *The Journal of Finance and Data Science*, vol. 4, no. 3, pp. 183–201, 2018. doi:10.1016/j.jfds.2018.04.003
- [46] C.-Y. Yeh, C.-W. Huang, and S.-J. Lee, “A multiple-kernel support vector regression approach for stock market price forecasting,” *Expert Systems with Applications*, vol. 38, no. 3, pp. 2177–2186, 2011. doi:10.1016/j.eswa.2010.08.004
- [47] R.K. Nayak, D. Mishra, A.K. Rath, "A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices," *Applied Soft Computing*, vol.35, no.1, pp. 670-680, 2015.
- [48] R. Handa, H. S. Hota, and S. R. Tandan, “Stock Market Prediction with Various Technical Indicators Using Neural Network Techniques,” *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 3, no. 6, pp. 604–608, Jun. 2015
- [49] K. C. D. II and J. R. Dahlquist, “Technical Analysis: The Complete Resource for Financial Market Technicians”, Upper Saddle River, N.J: *Financial Times*, 2016
- [50] T. T. L. Chong and W. K. Ng, “Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30,” *Applied Economics Letters*, vol. 15, no. 14, pp. 1111–1114, 2008.
- [51] S. Hochreiter, J. Schmidhuber, “Long short-term memory”, *Neural computation*, vol. 9, no.8, pp. 1735–1780, 1997
- [52] F. A. Gers, J. Schmidhuber, F. Cummins, “Learning to forget: Continual prediction with lstm”, *Neural computation*, vol. 12, no.10, pp. 2451–2471, 2000.
- [53] H. N. Bhandari *et al.*, “Predicting stock market index using LSTM,” *Machine Learning with Applications*, vol. 9, p. 100320, 2022. doi:10.1016/j.mlwa.2022.100320
- [54] W. Rodi, N. Fueyo, *Engineering turbulence modelling and experiments 5*, Elsevier, 2002.

- [55] D. R. Babu and B. Sathyanarayana, "Design and implementation of Technical Analysis based LSTM model for stock price prediction," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 4s, pp. 01–07, 2023. doi:10.17762/ijritcc.v11i4s.6301
- [56] D. P. Gandhmal and K. Kumar, "Systematic analysis and review of Stock Market Prediction Techniques," *Computer Science Review*, vol. 34, p. 100190, 2019. doi:10.1016/j.cosrev.2019.08.001
- [57] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using Breaking Financial News," *ACM Transactions on Information Systems*, vol. 27, no. 2, pp. 1–19, 2009. doi:10.1145/1462198.1462204
- [58] P. Oncharoen and P. Vateekul, "Deep learning for stock market prediction using event embedding and technical indicators," *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, 2018. doi:10.1109/icaicta.2018.8541310
- [59] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong," *Information Processing & Management*, vol. 57, no. 5, p. 102212, 2020. doi:10.1016/j.ipm.2020.102212
- [60] Y. Li and Y. Pan, "A novel ensemble deep learning model for stock prediction based on stock prices and news," *International Journal of Data Science and Analytics*, vol. 13, no. 2, pp. 139–149, 2021. doi:10.1007/s41060-021-00279-9
- [61] T. Sidogi, R. Mbuva, and T. Marwala, "Stock Price Prediction Using Sentiment Analysis," *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 46–51, 2021
- [62] A. E. Khedr, S.E.Salama, and N. Yaseen, "Predicting stock market behavior using data mining technique and news sentiment analysis," *International Journal of Intelligent Systems and Applications*, vol. 9, no. 7, pp. 22–30, 2017. doi:10.5815/ijisa.2017.07.03

- [63] A. Degutis and L. Novickytė, “The Efficient Market Hypothesis: A critical review of literature and methodology,” *Ekonomika*, vol. 93, no. 2, pp. 7–23, Jan. 2014.
doi:10.15388/ekon.2014.2.3549
- [64] A. G. Țițan, “The Efficient Market Hypothesis: Review of specialized literature and Empirical Research,” *Procedia Economics and Finance*, vol. 32, pp. 442–449, 2015.
doi:10.1016/s2212-5671(15)01416-1
- [65] M. R. Borges, “Efficient market hypothesis in European stock markets,” *The European Journal of Finance*, vol. 16, no. 7, pp. 711–726, Aug. 2010.
doi:10.1080/1351847x.2010.495477
- [66] G. Oricchio, “The efficient-market hypothesis (EMH) and financial crisis,” *Corporate Management in a Knowledge-Based Economy*, pp. 155–168, 2012.
doi:10.1057/9780230355453_7
- [67] T. T. Chitenderu, A. Maredza, and K. Sibanda, “The random walk theory and stock prices: Evidence from Johannesburg Stock Exchange,” *International Business & Economics Research Journal (IBER)*, vol. 13, no. 6, p. 1241, Oct. 2014. doi:10.19030/iber.v13i6.8918
- [68] J. J. Dsouza and T. Mallikarjunappa, “Do the stock market indices follow random walk?,” *Asia-Pacific Journal of Management Research and Innovation*, vol. 11, no. 4, pp. 251–273, Dec. 2015. doi:10.1177/2319510x15602969
- [69] “Weak form efficient or not? tests of random walk hypothesis,” *The Efficiency of China’s Stock Market*, pp. 107–144, Nov. 2017. doi:10.4324/9781351146920-14
- [70] S. Mathivannan and M. Selvakumar, “Test of random walk theory in the National Stock Exchange,” *Asian Journal of Managerial Science*, vol. 4, no. 2, pp. 21–25, Nov. 2015.
doi:10.51983/ajms-2015.4.2.1193
- [71] N. M. Guest, “Do journalists help investors analyze firms’ earnings news?,” *SSRN Electronic Journal*, 2017. doi:10.2139/ssrn.3067387

- [72] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Single Layer & Multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting," *Procedia Computer Science*, vol. 135, pp. 89–98, 2018. doi:10.1016/j.procs.2018.08.153
- [73] H. Ince and T. B. Trafalis, "Kernel principal component analysis and support vector machines for stock price prediction," *IIE Transactions*, vol. 39, no. 6, pp. 629–637, Mar. 2007. doi:10.1080/07408170600897486
- [74] J. L. Ticknor, "A Bayesian Regularized Artificial Neural Network for Stock Market Forecasting," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5501–5506, Oct. 2013. doi:10.1016/j.eswa.2013.04.013
- [75] F. A. de Oliveira, C. N. Nobre, and L. E. Zárate, "Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index – case study of PETR4, Petrobras, Brazil," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7596–7606, Dec. 2013. doi:10.1016/j.eswa.2013.06.071
- [76] D. M. Nelson, A. C. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017. doi:10.1109/ijcnn.2017.7966019
- [77] R. P. Schumaker, "Analyzing parts of speech and their impact on stock price," *Communications of the IIMA*, vol. 10, no. 3, Jun. 2014. doi:10.58729/1941-6687.1139
- [78] D. Lien Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon, "Deep Learning Approach for short-term stock trends prediction based on two-stream gated recurrent unit network," *IEEE Access*, vol. 6, pp. 55392–55404, 2018. doi:10.1109/access.2018.2868970
- [79] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong," *Information Processing & Management*, vol. 57, no. 5, p. 102212, Sep. 2020. doi:10.1016/j.ipm.2020.102212

- [80] R. Chiong *et al.*, “A sentiment analysis-based machine learning approach for financial market prediction via news disclosures,” *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, Jul. 2018. doi:10.1145/3205651.3205682
- [81] F. Subhi Alzazah and X. Cheng, “Recent advances in stock market prediction using text mining: A survey,” *E-Business - Higher Education and Intelligence Applications*, May 2021. doi:10.5772/intechopen.92253
- [82] S. Gite *et al.*, “Explainable stock prices prediction from financial news articles using sentiment analysis,” *PeerJ Computer Science*, vol. 7, Jan. 2021. doi:10.7717/peerj-cs.340
- [83] M. Omarkhan, G. Kissymova, and I. Akhmetov, “Handling data imbalance using CNN and LSTM in financial news sentiment analysis,” *2021 16th International Conference on Electronics Computer and Computation (ICECCO)*, Nov. 2021. doi:10.1109/icecco53203.2021.9663802
- [84] I. Gupta, T. K. Madan, S. Singh, and A. K. Singh, “Hisa-SMFM: Historical and sentiment analysis based stock market forecasting model,” arXiv.org, <https://arxiv.org/abs/2203.08143> (accessed Jun. 12, 2024).
- [85] M. Sharaf, E. E.-D. Hemdan, A. El-Sayed, and N. A. El-Bahnasawy, “An efficient hybrid stock trend prediction system during COVID-19 pandemic based on stacked-LSTM and news sentiment analysis,” *Multimedia Tools and Applications*, vol. 82, no. 16, pp. 23945–23977, Nov. 2022. doi:10.1007/s11042-022-14216-w
- [86] S. Shen, X. Liu, H. Sun, and D. Wang, “Biomedical knowledge discovery based on Sentence-BERT,” *Proceedings of the Association for Information Science and Technology*, 57(1), e362, 2020.
- [87] X. Huang, and Q. Hu, “A Bayesian learning approach to promoting diversity in ranking for biomedical information retrieval”, In J. Allan, J. A. Aslam, M. Sanderson, C. Zhai, & J. Zobel (Eds.), *Proceedings of the 32nd Annual International ACM SIGIR Conference on*

- Research and Development in Information Retrieval, SIGIR 2009*, pp. 307–314, Boston, MA, USA, July 19-23, 2009. <https://doi.org/10.1145/1571941.1571995>
- [88] X. Yin, J. X. Huang, Z. Li, and X. Zhou, "A Survival Modeling Approach to Biomedical Search Result Diversification Using Wikipedia," *IEEE Trans. Knowledge. Data Eng.*, 25(6), pp. 1201–1212, 2013. <https://doi.org/10.1109/TKDE.2012.24>
- [89] M. Lupu, F. Piroi, X. Huang, J. Zhu, and J. Tait, "Overview of the TREC 2009 Chemical IR Track," *The Proceedings of the 18th Text REtrieval Conference (TREC'09)*, pp. 1–26, 2009a.
- [90] M. Lupu, F. Piroi, X. Huang, J. Zhu, and J. Tait, "TREC-CHEM: large scale chemical information retrieval evaluation at TREC," *ACM SIGIR Forum*, 43(2), pp. 63–70, 2009b.
- [91] C. Huang, J. Chen, L. Xia, Y. Xu, P. Dai, Y. Chen, L. Bo, J. Zhao, and J. X. Huang, "Graph-Enhanced Multi-Task Learning of Multi-Level Transition Dynamics for Session-based Recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4123–4130, 2021. <https://doi.org/10.1609/AAAI.V35I5.16534>
- [92] L. Xia, C. Huang, Y. Xu, J. Zhao, D. Yin, and J. X. Huang, "Hypergraph Contrastive Collaborative Filtering," In E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, & G. Kazai (Eds.), *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 70–79, Madrid, Spain, July 11 - 15, 2022. <https://doi.org/10.1145/3477495.3532058>
- [93] Q. Chen, Q. Hu, J. X. Huang, L., and W. An, "Enhancing Recurrent Neural Networks with Positional Attention for Question Answering," In N. Kando, T. Sakai, H. Joho, H. Li, A. P. de Vries, & R. W. White (Eds.), *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 993–996, Shinjuku, Tokyo, Japan, August 7-11, 2017. <https://doi.org/10.1145/3077136.3080699>
- [94] K. Rabah, "Convergence of AI, IoT, Big Data and Blockchain: A Review," *The Lake Institute Journal*, vol. 1, no. 1, pp. 1–18, Mar. 2018.

- [95] J. X. Huang, J. Miao, and B. He, “High performance query expansion using adaptive co-training,” *Information Processing and Management*, vol. 49, no. 2, pp. 441–453, Mar. 2013. doi:10.1016/j.ipm.2012.08.002
- [96] X. Huang, A. An, and N. Cercone, “Comparison of interestingness functions for learning web usage patterns,” *Proceedings of the eleventh international conference on Information and knowledge management*, Nov. 2002. doi:10.1145/584792.584896
- [97] F. Balubaid, “S&P 500 Index Analysis: The effects of the 2008 financial crisis,” thesis, Halifax, 2015
- [98] J. Pawlowsky, “Treatment of hepatitis C: Don’t put all your eggs in one basket!,” *Gastroenterology*, vol. 132, no. 4, pp. 1611–1615, Apr. 2007. doi:10.1053/j.gastro.2007.03.014
- [99] H. Gao and S. Zhou, “A survey of explainable e-commerce Recommender Systems,” *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, Nov. 2022. doi:10.1109/wi-iat55865.2022.00115
- [100] K. Zhang, G. Zhong, J. Dong, S. Wang, and Y. Wang, “Stock market prediction based on generative Adversarial Network,” *Procedia Computer Science*, vol. 147, pp. 400–406, 2019. doi:10.1016/j.procs.2019.01.256
- [101] Y. Tsai and Q. Zhao, “An experimental study on the effectiveness of artificial neural network-based stock index prediction,” *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, Jul. 2019. doi:10.1109/icmlc48188.2019.8949282
- [102] M. Jiang, L. Jia, Z. Chen, and W. Chen, “The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm,” *Annals of Operations Research*, vol. 309, no. 2, pp. 553–585, Jun. 2020. doi:10.1007/s10479-020-03690-w
- [103] H. Ince and T. B. Trafalis, “Kernel principal component analysis and support vector machines for stock price prediction,” *2004 IEEE International Joint Conference on Neural*

Networks (IEEE Cat. No.04CH37541), vol. 3, pp. 2053–2058.
doi:10.1109/ijcnn.2004.1380933

- [104] C.-L. Huang and C.-Y. Tsai, “A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 1529–1539, Mar. 2009. doi:10.1016/j.eswa.2007.11.062
- [105] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, “Support vector regression with Chaos-based Firefly algorithm for stock market price forecasting,” *Applied Soft Computing*, vol. 13, no. 2, pp. 947–958, Feb. 2013. doi:10.1016/j.asoc.2012.09.024
- [106] Z. H. Przasnyski and L. S. Tai, “Stock performance of Malcolm Baldrige National Quality Award winning companies,” *Total Quality Management*, vol. 13, no. 4, pp. 475–488, Jul. 2002. doi:10.1080/09544120220149287
- [107] S. Banerjee, “Short term stock price prediction in Indian market: A neural network perspective,” *SSRN Electronic Journal*, 2020. doi:10.2139/ssrn.3701816
- [108] G. Ding and L. Qin, *Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM by Guangyu Ding, Liangxi Qin*. 2019.
- [109] Y. Guo *et al.*, “An adaptive SVR for high-frequency stock price forecasting,” *IEEE Access*, vol. 6, pp. 11397–11404, 2018. doi:10.1109/access.2018.2806180
- [110] J. Zheng, Y. Wang, S. Li, and H. Chen, “The stock index prediction based on SVR model with bat optimization algorithm,” *Algorithms*, vol. 14, no. 10, p. 299, Oct. 2021. doi:10.3390/a14100299
- [111] R. Batra and S. M. Daudpota, “Integrating stocktwits with sentiment analysis for better prediction of Stock Price Movement,” *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Mar. 2018. doi:10.1109/icomet.2018.8346382

- [112] J. Zhou, J. Zhao, J. X. Huang, Q. V. Hu, and L. He, “MASAD: A large-scale dataset for multimodal aspect-based sentiment analysis,” *Neurocomputing*, vol. 455, pp. 47–58, Sep. 2021. doi:10.1016/j.neucom.2021.05.040

Appendix A

Technical Indicators Computing

A.1 Instructions

In order to compute the technical indicators to identify trend of price movement in Chapter 3.2.1, you need to do the following:

- Downloading S&P 500 historical price data from yahoo finance.
- Implementing the Python library TA-Lib to calculate technical analysis indicators.

```
1. import talib
2. import yfinance as yf
3. # download S&P500 historical price data from yahoo finance
4. sp500_data = yf.download("SNP", start = "2006-10-20", end = "2013-11-20")
5. # Building Technical indicators
6. # Simple Moving Average (SMA) over a specific number of time periods (10
   days, 30 days, 60 days)
7. sp500_data['SMA_10'] = talib.SMA(sp500_data['Close'], timeperiod = 10)
8. sp500_data['SMA_30'] = talib.SMA(sp500_data['Close'], timeperiod = 30)
9. sp500_data['SMA_60'] = talib.SMA(sp500_data['Close'], timeperiod = 60)
10. # Exponential Moving Average (EMA) over a specific number of time period
    s (10 days, 30 days, 60 days)
11. sp500_data['EMA_10'] = talib.EMA(sp500_data['Close'], timeperiod = 10)
```

```

12. sp500_data['EMA_30'] = talib.EMA(sp500_data['Close'], timeperiod = 30)
13. sp500_data['EMA_60'] = talib.EMA(sp500_data['Close'], timeperiod = 60)
14. # Average Directional Movement (ADX)
15. sp500_data['ADX'] = talib.ADX(sp500_data['High'], sp500_data['Low'], sp500_data['Close'], timeperiod = 14)
16. # Moving Average Convergence Divergence (MACD)
17. sp500_data['MACD'] = talib.MACD(sp500_data['Close'], fastperiod=12, slowperiod=26, signalperiod=9) [0]
18. # Relative Strength Index (RSI)
19. sp500_data['RSI'] = talib.RSI(sp500_data['Close'], timeperiod=14)
20. # Bollinger bands (upper, middle, lower band)
21. # 20-day simple moving average (middle band)
22. # The upper and lower bands are typically 2 standard deviations +/- from a 20-day simple moving average
23. sp500_data['BBANDS_U'] = talib.BBANDS(sp500_data['Close'], nbdevup=2, nbdevdn=2, timeperiod = 20) [0]
24. sp500_data['BBANDS_M'] = talib.BBANDS(sp500_data['Close'], nbdevup=2, nbdevdn=2, timeperiod = 20) [1]
25. sp500_data['BBANDS_L'] = talib.BBANDS(sp500_data['Close'], nbdevup=2, nbdevdn=2, timeperiod = 20) [2]
26. # Williams % R moves between zero and -100.
27. # A reading above -20 is overbought.
28. # A reading below -80 is oversold.
29. sp500_data['WILLR'] = talib.WILLR(sp500_data['High'], sp500_data['Low'], sp500_data['Close'], timeperiod=14)

```

Appendix B

Financial News Headline Extracting

B.1 Instructions

In order to extract the headlines from financial news data source in Chapter 4.2, you need to do the following:

- Grouping the financial news that published on the same day.
- Extracting the news headlines and year, month, and date of publication.

```
1. import os
2. from typing import List, Tuple
3. import csv
4.
5.
6. class Time:
7.     def __init__(self, date: str, hour: int, minute: int):
8.         self.date = date
9.         self.hour = hour
10.        self.minute = minute
11.
12.
13. def pair_sort_key (pair: [str, Time]) -> int:
14.     news_time = pair [1]
15.     return int (news_time.date) * (24 * 60) + news_time.hour * 60 + news_time.
    minute
```

```

16.
17.
18. if __name__ == '__main__':
19.     directories = os. listdir("ReutersNews106521")
20.     pairs: List [Tuple [str, Time]] = []
21.     for sub_dir in directories:
22.         if sub_dir [0] == ".":
23.             continue
24.         file_names = os. listdir ("ReutersNews106521/" + sub_dir)
25.         for file_name in file_names:
26.             if file_name [0] == "1":
27.                 continue
28.             with open
29.                 ("ReutersNews106521/" + sub_dir + "/" + file_name) as file:
30.                     header = file. readline () [3:]
31.                     header = header [: len(header) - 1]
32.                     file. readline ()
33.                     time_str = file. readline ()
34.                     print(time_str)
35.                     parts = time_str. split (' ')
36.                     if len(parts) >= 6:
37.                         hour_str = parts [5]. split (':') [0]
38.                         last_part = parts [5]. split (':') [1]
39.                         minute_str = last_part [0: len(last_part) - 2]
40.                         flag_str = last_part[len(last_part) - 2:]
41.
42.                         minute = int(minute_str)
43.                         hour = int(hour_str)
44.                         if flag_str == "pm":
45.                             hour += 12
46.                         time = Time (sub_dir, hour, minute)
47.                         pairs. append ((header, time))
48.
49. #20000805
50. pairs = sorted (pairs, key=pair_sort_key)
51. with open ("result.csv", "w") as file:
52.     names = ["time", "header"]
53.     writer = csv. DictWriter (file, fieldnames=names)
54.     for pair in pairs:
55.         time_str = pair [1]. date
56.         year = time_str [0:4]

```

```
56.         month = time_str [4:6]
57.         date = time_str [6:8]
58.         writer.writerow ({
59.             "time": "{0}-{1}-{2}".format (year, month, date),
60.             "header": pair [0]
61.         })
```

Appendix C

News Headlines Preprocessing

C.1 Instructions

In order to preprocess news headlines for NLTK-based sentiment analysis in Chapter 3.1, you need to do the following:

- Implementing the Python library Natural Language Toolkit for preprocessing news headlines.
- Removing punctuation, converting uppercase letters to lowercase, filtering out all stopwords, removing the affix, and determining the lemma of words.

```
1. import string
2. import pandas as pd
3. from nltk.tokenize import word_tokenize
4. from nltk.corpus import stopwords
5. from nltk.stem.porter import PorterStemmer
6. from nltk.stem.wordnet import WordNetLemmatizer
7.
8. news_title = pd.read_excel('news_title.xlsx', index_col=0)
9.
10.
11. # remove punctuation
12. def remove_punctuation(text):
13.     no_punct = [words for words in text if words not in string.punctuation]
14.     words_no_punct = ' '.join(no_punct)
15.     return words_no_punct
16.
17.
```

```

18. news_title['title_no_punct'] = news_title['Headline']. apply
    (lambda x: remove_punctuation(x))
19. news_title. head ()
20.
21.
22. # Tokenization and convert to lower case
23. def tokenize(text):
24.     tokens = word_tokenize (text. replace ('-', ''))
25.     return tokens
26.
27.
28. news_title['title_no_punct_split'] = news_title['title_no_punct']. apply
    (lambda x: tokenize (x. lower ()))
29. news_title. head ()
30.
31.
32. # filter out stop words("of","a","the"..)
33. def remove_stopwords(text):
34.     stopword = set (stopwords. words('english'))
35.     text = [word for word in text if word not in stopword]
36.     return text
37.
38.
39. news_title['title_no_punct_split_stopwords'] = news_title['title_no_punct_split'
    ]. apply (lambda x: remove_stopwords(x))
40. news_title. head ()
41.
42. # Lexicon Normalization
43. # stemming
44.
45. stemmer = PorterStemmer ()
46. news_title['title_no_punct_split_stopwords_stemmer'] = news_title['title_no_punc
    t_split_stopwords']. apply (lambda x: [stemmer. stem(i) for i in x])
47. news_title. head ()
48.
49.
50. # Lemmatization
51. lemmatizer = WordNetLemmatizer ()
52. news_title['title_no_punct_split_stopwords_stemmer_Lemmatizer'] = news_title['ti
    tle_no_punct_split_stopwords_stemmer']. apply (lambda x: [lemmatizer.
    lemmatize(i) for i in x])

```

```
53. news_title. head ()  
54. print(news_title)  
55.  
56. news_title.to_excel("news_title_preprocessed.xlsx")
```

Appendix D

Sentiment Analysis Using LMD

D.1 Instructions

In order to preprocess news headlines for LMD-based sentiment analysis in Chapter 3.1, you need to do the following:

- Implementing the Python library Natural Language Toolkit for tokenizing words.
- Removing punctuation, converting uppercase letters to lowercase, filtering out all stopwords according the LoughranMcDonald stopwords list, removing the affix, and determining the lemma of words.
- Classifying the sentiment of words in news headlines according to the vocabulary list in the LoughranMcDonald financial sentiment dictionary.
- Counting the number of words belonging to 7 sentiment categories for each trading day.

```
1. import pandas as pd
2. import string
3. from nltk.tokenize import word_tokenize
4.
5. news_title = pd.read_excel('news_title.xlsx', index_col=0)
6. # remove punctuation from each word
7. def remove_punctuation(text):
8.     no_punct = [words for words in text if words not in string.punctuation]
```

```

9.     words_no_punct = ''. join (no_punct)
10.    return words_no_punct
11. news_title['Headline_no_punct'] = news_title['Headline']. apply
    (lambda x: remove_punctuation(x))
12.
13. # Tokenize the words and convert to lowercase
14. def tokenize(text):
15.     tokens = word_tokenize (text. replace ('-', ''))
16.     return tokens
17. news_title['Headline_tokenized'] = news_title['Headline_no_punct']. apply
    (lambda x: tokenize (x. lower ()))
18. news_title. head ()
19.
20. # load the LoughranMcDonald generic stopwords list
21. stopwords = pd. read_csv ("StopWords_GenericLong.txt", names = ['Word'])
22. stop_list = stopwords['Word']. tolist ()
23. print(stop_list)
24.
25. # Remove the stopwords
26. def remove_stopwords(text):
27.     text = [word for word in text if word not in stop_list]
28.     return text
29. news_title['Headline_no_stopwords'] = news_title['Headline_tokenized']. apply
    (lambda x: remove_stopwords(x))
30. news_title. head ()
31. print(news_title)
32. news_title.to_excel('news_title_loughran.xlsx')
33.
34. # Load master dictionary
35. master = pd. read_csv('Loughran-McDonald_MasterDictionary_1993-2021.csv')
36. positive = master[master["Positive"] > 0]
37. negative = master[master["Negative"] > 0]
38. uncertainty = master[master["Uncertainty"] > 0]
39. litigious = master[master["Litigious"] > 0]
40. strong_modal = master[master["Strong_Modal"] > 0]
41. weak_modal = master[master["Weak_Modal"] > 0]
42. constraining = master[master["Constraining"] > 0]
43.
44. pos_words = positive["Word"]. tolist ()
45. neg_words = negative["Word"]. tolist ()
46. uncer_words = uncertainty["Word"]. tolist ()

```

```

47. liti_words = litigious["Word"]. tolist ()
48. strong_words = strong_modal["Word"]. tolist ()
49. weak_words = weak_modal["Word"]. tolist ()
50. constr_words = constraining["Word"]. tolist ()
51. pos_words = [word. lower () for word in pos_words]
52. neg_words = [word. lower () for word in neg_words]
53. uncer_words = [word. lower () for word in uncer_words]
54. liti_words = [word. lower () for word in liti_words]
55. strong_words = [word. lower () for word in strong_words]
56. weak_words = [word. lower () for word in weak_words]
57. constr_words = [word. lower () for word in constr_words]
58.
59. def pos_word(text):
60.     text = [word for word in text if word in pos_words]
61.     return text
62. news_title['words_new_pos'] = news_title['Headline_no_stopwords']. apply
    (lambda x: pos_word(x))
63. news_title. head ()
64.
65. def neg_word(text):
66.     text = [word for word in text if word in neg_words]
67.     return text
68. news_title['words_new_neg'] = news_title['Headline_no_stopwords']. apply
    (lambda x: neg_word(x))
69. news_title. head ()
70.
71. def uncer_word(text):
72.     text = [word for word in text if word in uncer_words]
73.     return text
74. news_title['words_new_uncer'] = news_title['Headline_no_stopwords']. apply
    (lambda x: uncer_word(x))
75. news_title. head ()
76.
77. def liti_word(text):
78.     text = [word for word in text if word in liti_words]
79.     return text
80. news_title['words_new_liti'] = news_title['Headline_no_stopwords']. apply
    (lambda x: liti_word(x))
81. news_title. head ()
82.
83. def strong_word(text):

```

```

84.     text = [word for word in text if word in strong_words]
85.     return text
86. news_title['words_new_strong'] = news_title['Headline_no_stopwords']. apply
    (lambda x: strong_word(x))
87. news_title. head ()
88.
89. def weak_word(text):
90.     text = [word for word in text if word in weak_words]
91.     return text
92. news_title['words_new_weak'] = news_title['Headline_no_stopwords']. apply
    (lambda x: weak_word(x))
93. news_title. head ()
94.
95. def constr_word(text):
96.     text = [word for word in text if word in constr_words]
97.     return text
98. news_title['words_new_constr'] = news_title['Headline_no_stopwords']. apply
    (lambda x: constr_word(x))
99. news_title. head ()
100.
101.     news_title['pos_Score'] = news_title['words_new_pos']. str.len ()
102.     news_title['neg_Score'] = news_title['words_new_neg']. str.len ()
103.     news_title['uncer_Score'] = news_title['words_new_uncer']. str.len ()
104.     news_title['liti_Score'] = news_title['words_new_liti']. str.len ()
105.     news_title['strong_Score'] = news_title['words_new_strong']. str.len ()
106.     news_title['weak_Score'] = news_title['words_new_weak']. str.len ()
107.     news_title['constr_Score'] = news_title['words_new_constr']. str.len ()
108.
109.     news_title. head ()
110.     print(news_title)
111.     news_title.to_excel('news_title_loughran.xlsx')

```

Appendix E

Data Splitting and Standardization

E.1 Instructions

In order to split data into training, test, and validation set for proposed methods experiments in Chapter 3.4, you need to do the following:

- Filling blank data with median.
- Splitting 7 years historical data into training (5 years), test (1 year), and validation set (1 year).
- Implementing the Python Sklearn library to standardize the data values into a standard format.

```
1. # fill na
2. data.fillna(data.median(), inplace=True)
3. # print(data)
4.
5. # splitting data
6. x = data.drop(['Target'], axis=1)
7. y = data[['Target']]
8.
9. x_train = x ["2006-10-20":"2011-10-20"]
10. x_test = x ["2011-10-21":"2012-10-20"]
11. x_validation = x ["2012-10-21":"2013-11-19"]
```

```

12. y_train = y ["2006-10-20":"2011-10-20"]
13. y_test = y ["2011-10-21":"2012-10-20"]
14. y_validation = y ["2012-10-21":"2013-11-19"]
15. sc = StandardScaler ()
16. x_train= sc.fit_transform(x_train)
17. x_test = sc. transform(x_test)
18.
19. scy = StandardScaler ()
20. y_train = scy.fit_transform(y_train)
21. y_test = scy.fit_transform(y_test)
22.
23. scv = StandardScaler ()
24. x_validation = scv.fit_transform(x_validation)
25. y_validation = scv.fit_transform(y_validation)
26.
27.
28. x_train, y_train, = np. array(x_train), np. array(y_train)
29. x_train = np. reshape (x_train, (x_train. shape [0], x_train. shape [1], 1))
30. x_test, y_test = np. array(x_test), np. array(y_test)
31. x_test = np. reshape (x_test, (x_test. shape [0], x_test. shape [1], 1))
32. x_validation, y_validation = np. array(x_validation), np. array(y_validation)
33. x_validation = np. reshape (x_validation, (x_validation. shape [0],
    x_validation. shape [1],1))

```

Appendix F

Effectiveness of Proposed SVR Model Experiment

F.1 Instructions

In order to run the proposed SVR experiment in Chapter 3.4.1, and evaluating model in validation set, you need to do the following:

- Implementing the Python Sklearn library to build SVR model.
- Defining parameter range.
- Fitting the model for grid search to determine the best parameter.
- Evaluating the model in validation set and draw a plot.

```
1. import pandas as pd
2. import numpy as np
3. import matplotlib.pyplot as plt
4. from sklearn.preprocessing import StandardScaler
5. from sklearn.svm import SVR
6. from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
7. from sklearn.model_selection import GridSearchCV
8.
9. # Apply SVR model
10. # defining parameter range
11. param_grid = {'C': [0.001, 0.01, 0.1, 0.5, 1, 10],
```

```

12.         'gamma': ['auto', 'scale'],
13.         'tol': [0.001],
14.         'kernel': ['sigmoid', 'poly', 'rbf'],
15.         'epsilon': [0.1, 0.2, 0.3]}
16.grid = GridSearchCV (SVR (), param_grid, refit=True, verbose=3)
17.# fitting the model for grid search
18.grid.fit (x_train, y_train. ravel ())
19.
20.# print best parameter after tuning
21.print (grid. best_params_)
22.# print how our model looks after hyperparameter tuning
23.print (grid. best_estimator_)
24.# evaluating the model in the validation set
25.ypred2 = grid. predict(x_validation)
26.rmse2 = np. sqrt (mean_squared_error (y_validation, ypred2))
27.mse2 = mean_squared_error (y_validation, ypred2)
28.mae2 = mean_absolute_error (y_validation, ypred2)
29.r2_score1 = r2_score (y_validation, ypred2)
30.print ('Test RMSE: %.3f' %rmse2)
31.print ('Test MSE: %.3f' %mse2)
32.print ('Test MAE: %.3f' %mae2)
33.print ('R2 score= %.3f' %r2_score1)
34.print(ypred2)
35.
36.inverse_pre = scv. inverse_transform(ypred2)
37.y_validation = scv. inverse_transform(y_validation)
38.
39.print (inverse_pre. shape)
40.print (y_validation. shape)
41.inverse_pre = inverse_pre. flatten ()
42.y_validation = y_validation. flatten ()
43.print(inverse_pre)
44.print(y_validation)
45.
46.df_time = data. Index[-len(y_validation):]
47.df_time = np. Array (df_time.to_pydatetime (), dtype=np. datetime64)
48.print(df_time)
49.result = {'actual_price': y_validation, 'predicted_price': inverse_pre}
50.result = pd. DataFrame (result)
51.print(result)
52.result. index = df_time

```

```
53. print(result)
54.
55. result.to_excel('./SVM_nltk_predicted_result.xlsx')
56. result.head ()
57.
58.
59. plt. plot (df_time, y_validation, color='orange', label='actual_price')
60. plt. Plot (df_time, inverse_pre, color = 'green', label = 'predicted_price')
61. plt. title ('SVR_nltk_prediction')
62. plt. legend ()
63. plt. show ()
64. plt. Savefig ('/SVR_nltk_prediction.png', dpi=500)
```

Appendix G

Effectiveness of Proposed Multi-layer LSTM Model Experiment

G.1 Instructions

In order to run the proposed Multi-layer LSTM experiment in Chapter 3.4.2, and evaluating model in validation set, you need to do the following:

- Implementing the Python Sklearn and Keras library to build multi-layer LSTM model.
- Defining parameter range.
- Running the model with different set of parameters to determine the best parameter.
- Evaluating the model in validation set and draw a plot.

```
1. import pandas as pd
2. import numpy as np
3. import matplotlib.pyplot as plt
4. from sklearn.preprocessing import StandardScaler
5. import tensorflow as tf
6. from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
7. from keras.models import Sequential
8. from keras.layers import Dense, Dropout, LSTM
9. from keras.callbacks import ModelCheckpoint
```

```

10.
11. lstm_layers = [5]
12. dense_layers = [2]
13. units = [256]
14. adam = tf.keras.optimizers.legacy.Adam(
15.     learning_rate=0.001,
16.     beta_1=0.9,
17.     beta_2=0.999,
18.     epsilon=1e-8,
19.     decay=0.0
20. )
21.
22. for the_lstm_layers in lstm_layers:
23.     for the_dense_layers in dense_layers:
24.         for the_units in units:
25.             filepath = 'loughran_withoutpre'+'{val_mae:.2f}_{epoch:02d}_'+ f'ls
tm_{the_lstm_layers}_dense_{the_dense_layers}_units_{the_units}'
26.             checkpoint = ModelCheckpoint (
27.                 filepath=filepath,
28.                 save_weights_only=False,
29.                 monitor='val_mae',
30.                 mode='min',
31.                 save_best_only=True
32.             )
33.             # build LSTM model
34.             model = Sequential ()
35.             model.add (LSTM (the_units, input_shape = (x_train. Shape
[1], 1), activation='relu', return_sequences=True))
36.             model.add (Dropout (0.5))
37.             for i in range(the_lstm_layers):
38.                 model.add (LSTM (the_units, activation='relu',
return_sequences=True))
39.                 model.add (Dropout (0.5))
40.                 model.add (LSTM (the_units, activation='relu'))
41.                 model.add (Dropout (0.5))
42.             for i in range(the_dense_layers):
43.                 model.add (Dense (the_units, activation='relu'))
44.                 model.add (Dropout (0.5))
45.                 model.add (Dense (1))
46.
47.             model.compile (optimizer=adam, loss='mse', metrics=['mae'])

```

```

48.         model. fit
        (x_train, y_train, batch_size=32, epochs=70, validation_data = (x_test, y_test),
        callbacks=[checkpoint])
49.
50.
51. from keras. models import load_model
52. best_model = load_model('./loughran_withoutpre0.52_15_lstm_5_dense_2_units_256')
53.
54. pre = best_model. predict (x_validation)
55. rmse = np. sqrt (mean_squared_error (y_validation, pre))
56. mse = mean_squared_error (y_validation, pre)
57. mae = mean_absolute_error (y_validation, pre)
58. r2_score = r2_score (y_validation, pre)
59.
60. inverse_pre = scv. inverse_transform(pre)[: ,0]
61. y_validation = scv. inverse_transform(y_validation) [: ,0]
62. inverse_pre = inverse_pre. Flatten ()
63. y_validation = y_validation. Flatten ()
64. print ('Test RMSE: %.3f' %rmse)
65. print ('Test MSE: %.3f' %mse)
66. print ('Test MAE: %.3f' %mae)
67. print ('R2 score= %.3f' %r2_score)
68.
69. df_time = data. Index [-len(y_validation):]
70. df_time = np. Array (df_time.to_pydatetime (), dtype=np. datetime64)
71.
72. result = {'actual_price': y_validation,'predicted_price': inverse_pre}
73. result = pd. DataFrame (result)
74. result. index = df_time
75.
76. result.to_excel('\\LSTM_loughran_predicted_result_3.xlsx')
77. result. head()
78.
79. plt. plot (df_time, y_validation, color='orange', label='actual_price')
80. plt. plot (df_time, inverse_pre, color = 'green', label = 'predicted_price')
81. plt. title ('LSTM_loughran_prediction')
82. plt. legend ()
83. plt. show ()
84. plt. savefig ('/LSTM_loughran_prediction.png', dpi=500)

```

Appendix H

Predicted Results Using SVR with Technical Analysis Only

Predicted results using SVR with technical analysis only

	actual_price	predicted_price
2012-10-22	46.50396347	46.48928879
2012-10-23	46.12223816	46.40080817
2012-10-24	46.41620255	46.32114307
2012-10-25	45.48604584	46.63525687
2012-10-26	46.16173172	45.757926
2012-10-31	48.01328659	46.20694825
2012-11-01	47.52626419	48.00756497
2012-11-02	47.14455414	48.14205108
2012-11-05	47.04801559	47.74336942
2012-11-06	46.55662537	47.88282773
2012-11-07	45.85021591	47.30034352
2012-11-08	45.82827377	46.74754331
2012-11-09	45.99500656	46.56510038
2012-11-12	44.51200104	46.91838905
2012-11-13	44.65678024	45.22072733
2012-11-14	44.18732071	45.49335595
2012-11-15	44.76647568	45.1554999
2012-11-16	45.71858978	45.16707095
2012-11-19	45.3588028	45.87423491
2012-11-20	46.00377655	45.80545175
2012-11-21	47.61839676	46.33000182
2012-11-23	46.84617615	48.02363643
2012-11-26	45.8589859	47.49224648
2012-11-27	45.82387924	46.54545691
2012-11-28	45.854599	46.28970119
2012-11-29	46.46446991	46.49343847
2012-11-30	45.9774437	47.03724729
2012-12-03	46.66630554	46.57726131
2012-12-04	47.75442886	47.34840968
2012-12-05	47.79389954	48.28042737
2012-12-06	48.21511459	48.03278384
2012-12-07	48.90396881	48.57700942
2012-12-10	49.40415573	49.46307833
2012-12-11	49.15843201	49.39607511
2012-12-12	49.00926971	50.02685264
2012-12-13	49.25936508	50.02394763
2012-12-14	48.63632965	50.15493399
2012-12-17	49.206707	49.87647879
2012-12-18	49.81220245	50.1030956
2012-12-19	50.58441162	50.27533273
2012-12-20	49.6279068	50.34507866
2012-12-21	49.04875946	50.14939872
2012-12-24	49.47873688	50.25943832
2012-12-26	49.18476868	50.10645011
2012-12-27	49.206707	50.23122583

Predicted results using SVR with technical analysis only

2012-12-28	50.42207718	50.26911544
2012-12-31	51.88751221	50.35639092
2013-01-02	51.77344513	50.0766068
2013-01-03	51.98843765	50.11234958
2013-01-04	51.32152176	50.15283572
2013-01-07	51.10652542	50.03786255
2013-01-08	51.46631241	50.02428705
2013-01-09	52.30433273	49.90702112
2013-01-10	52.27801514	49.86962111
2013-01-11	52.24290848	49.64838236
2013-01-14	52.01476288	49.76849271
2013-01-15	51.30835724	49.74839048
2013-01-16	51.3434639	49.79726846
2013-01-17	52.07180405	49.5673918
2013-01-18	52.74748611	49.62327796
2013-01-22	52.08496094	49.50665413
2013-01-23	52.4491272	49.48419584
2013-01-24	53.2388916	49.48893651
2013-01-25	52.77380371	49.12975137
2013-01-28	53.8224411	49.23983541
2013-01-29	53.15554047	49.06211404
2013-01-30	53.28716278	49.02196104
2013-01-31	53.22134399	49.03558333
2013-02-01	49.39099884	48.96645944
2013-02-04	49.94382858	47.69230514
2013-02-05	50.3825798	49.56765975
2013-02-06	48.48275757	49.6951577
2013-02-07	48.97416306	49.53905492
2013-02-08	48.56612396	49.91445582
2013-02-11	48.64948654	49.99773356
2013-02-12	48.96101379	49.98019634
2013-02-13	48.7898941	49.97142525
2013-02-14	49.11457062	49.97114186
2013-02-15	49.64108276	49.96996258
2013-02-19	49.71566772	49.75931972
2013-02-20	49.52699661	49.79702739
2013-02-21	49.46995926	49.92491496
2013-02-22	48.94783783	49.98514795
2013-02-25	49.29007339	50.10037903
2013-02-26	49.8823967	50.12601291
2013-02-27	49.82536316	50.39358855
2013-02-28	50.9705162	50.31211589
2013-03-01	50.33431625	50.33228792
2013-03-04	50.42207718	50.5144646
2013-03-05	50.68532944	50.37851888
2013-03-06	50.23778534	50.52104932

Predicted results using SVR with technical analysis only

2013-03-07	50.93101883	50.11540263
2013-03-08	50.72481155	50.48280214
2013-03-11	50.06229782	50.52115539
2013-03-12	49.42170715	50.39293637
2013-03-13	49.89555359	50.28324889
2013-03-14	49.71128464	50.38883878
2013-03-15	48.9083519	50.00385196
2013-03-18	49.14527893	50.16432894
2013-03-19	50.26851273	49.94467614
2013-03-20	48.98292923	50.34074575
2013-03-21	49.82097244	49.92408477
2013-03-22	50.48349762	50.40706239
2013-03-25	51.13285446	50.46122266
2013-03-26	51.41804504	50.41113225
2013-03-27	51.29957199	50.48257047
2013-03-28	51.13285446	50.23367149
2013-04-01	51.44436646	50.35385071
2013-04-02	50.57123947	50.21859261
2013-04-03	50.63267517	50.26794208
2013-04-04	49.77709198	50.31449257
2013-04-05	49.85168457	50.1045393
2013-04-08	50.1895256	50.35586858
2013-04-09	50.04474258	50.34476016
2013-04-10	50.34747696	50.14288006
2013-04-11	49.77270126	50.44223966
2013-04-12	47.53942108	50.37490832
2013-04-15	48.26337051	49.45682768
2013-04-16	46.88568115	49.90243005
2013-04-17	47.10066223	49.03461258
2013-04-18	47.79389954	49.22988725
2013-04-19	47.79389954	49.46005869
2013-04-22	47.10066223	49.55003927
2013-04-23	47.63156128	48.69522677
2013-04-24	47.73686981	49.1502291
2013-04-25	47.57452774	49.16044496
2013-04-26	48.02645111	49.06932306
2013-04-29	48.50908661	48.424642
2013-04-30	47.69300079	49.63112223
2013-05-01	47.71054077	49.38432084
2013-05-02	47.87726974	49.23357537
2013-05-03	47.95624542	49.44599821
2013-05-06	48.54856873	49.50319068
2013-05-07	49.61915207	49.91150058
2013-05-08	48.75039673	50.09176977
2013-05-09	48.92591095	50.05215681
2013-05-10	48.4301033	50.23705791

Predicted results using SVR with technical analysis only

2013-05-13	48.67142868	49.95898943
2013-05-14	48.56612396	50.01067705
2013-05-15	48.16685104	50.01980107
2013-05-16	48.43888855	49.76585387
2013-05-17	48.79866791	49.84064564
2013-05-20	48.56612396	50.15430718
2013-05-21	47.86850357	50.13213373
2013-05-22	47.35075378	49.56610607
2013-05-23	46.54784012	48.74253374
2013-05-24	46.87690353	48.05622513
2013-05-28	45.80634308	48.46832215
2013-05-29	45.94673538	47.28825922
2013-05-30	44.66995621	47.26344862
2013-05-31	45.55186081	46.05915626
2013-06-03	45.24034119	46.39921565
2013-06-04	44.27506638	46.19775653
2013-06-05	44.14345169	45.35543336
2013-06-06	45.02861023	44.96065357
2013-06-07	43.18360138	44.75348354
2013-06-10	42.87532806	44.32271898
2013-06-11	42.2270813	44.31379204
2013-06-12	42.50814438	44.35937789
2013-06-13	41.74655533	44.26193618
2013-06-14	42.63960266	44.25049885
2013-06-17	43.05212784	44.36358548
2013-06-18	41.61961365	44.5306777
2013-06-19	39.59326553	44.43143522
2013-06-20	40.35485077	44.69028798
2013-06-21	38.86794662	44.95361624
2013-06-24	39.91059494	45.25155443
2013-06-25	40.58151627	45.11625295
2013-06-26	41.07110214	45.23671533
2013-06-27	41.47909546	45.16324077
2013-06-28	40.98044586	45.16472944
2013-07-01	40.16809464	45.3699814
2013-07-02	39.73198318	45.73233058
2013-07-03	40.00307465	45.87762419
2013-07-05	40.10915375	45.88708756
2013-07-08	40.45685959	45.99984571
2013-07-09	40.9459877	45.85267619
2013-07-10	42.51358414	46.00697886
2013-07-11	41.83585739	45.54661238
2013-07-12	41.90658569	45.8276165
2013-07-15	42.09516144	45.78189817
2013-07-16	42.60786819	45.8525435
2013-07-17	42.37214661	45.71767716

Predicted results using SVR with technical analysis only

2013-07-18	42.70804977	45.78754868
2013-07-19	42.72573853	45.66808151
2013-07-22	44.55850983	45.5950834
2013-07-23	44.41119766	45.14484552
2013-07-24	44.66460037	44.9347758
2013-07-25	44.51137543	45.10562603
2013-07-26	43.3680954	45.10513261
2013-07-29	43.47417068	45.13497043
2013-07-30	43.78060913	45.15985957
2013-07-31	44.2579689	45.04211735
2013-08-01	44.4052887	44.85296204
2013-08-02	44.02223206	44.96426941
2013-08-05	43.36220932	44.90214381
2013-08-06	42.34857941	44.9002983
2013-08-07	42.97325134	44.89751861
2013-08-08	43.06164551	45.08413909
2013-08-09	44.30511475	45.08561077
2013-08-12	44.56441879	44.98977873
2013-08-13	44.72943115	45.00845093
2013-08-14	44.44065475	45.0625169
2013-08-15	44.29333115	44.93515159
2013-08-16	44.41119766	44.99959927
2013-08-19	43.99277496	44.91157138
2013-08-20	43.00271225	44.9949295
2013-08-21	43.95152283	44.94290186
2013-08-22	44.24028015	44.97487701
2013-08-23	43.89258194	45.0697444
2013-08-26	43.45059586	44.45733802
2013-08-27	43.27969742	44.81182621
2013-08-28	42.92021561	44.92734559
2013-08-29	42.41930008	44.98494292
2013-08-30	45.0123024	45.01648703
2013-09-03	45.8538208	44.6884718
2013-09-04	46.24439621	44.99889207
2013-09-05	46.83326721	45.0447574
2013-09-06	48.26939392	45.05710151
2013-09-09	48.22733688	45.36770319
2013-09-10	47.54231644	45.60386639
2013-09-11	46.76717758	45.37413881
2013-09-12	46.86932373	45.07145209
2013-09-13	47.58439636	45.07261493
2013-09-16	47.66249466	45.51414901
2013-09-17	48.5938797	45.61989209
2013-09-18	48.605896	45.8908589
2013-09-19	47.7706604	46.59133079
2013-09-20	47.90886688	45.99842214

Predicted results using SVR with technical analysis only

2013-09-23	47.29595566	45.88151097
2013-09-24	47.31398392	45.44713798
2013-09-25	47.86079407	45.51326735
2013-09-26	47.64447784	45.97552422
2013-09-27	47.0435791	45.94051786
2013-09-30	47.27793503	45.25255367
2013-10-01	47.75263214	45.56176128
2013-10-02	47.80670929	45.94538931
2013-10-03	48.34750748	46.19480684
2013-10-04	48.05308151	46.60593124
2013-10-07	47.44016647	46.38613318
2013-10-08	47.68052292	45.97540062
2013-10-09	48.29343796	46.12617329
2013-10-10	48.73207474	46.56545945
2013-10-11	48.67199326	47.064662
2013-10-14	48.13719559	47.00299762
2013-10-15	48.28141403	46.74852513
2013-10-16	48.34149551	46.65855588
2013-10-17	48.39558792	46.66535899
2013-10-18	48.30544662	46.77450141
2013-10-21	48.47369385	46.7535244
2013-10-22	47.09166336	46.96797129
2013-10-23	46.8152504	45.41184797
2013-10-24	46.62896729	45.38074433
2013-10-25	46.05211639	45.24748708
2013-10-28	48.22132492	44.79239092
2013-10-29	47.7105751	45.83599463
2013-10-30	48.29343796	46.00886377
2013-10-31	48.91835785	46.44961613
2013-11-01	49.61539078	47.04826939
2013-11-04	49.8196907	47.86679878
2013-11-05	49.50722122	47.94672868
2013-11-06	48.58184814	48.15098301
2013-11-07	48.69602966	47.34610336
2013-11-08	49.10463715	47.21297981
2013-11-11	49.54327011	47.77445037
2013-11-12	48.88230896	48.26804191
2013-11-13	49.27889252	47.4566974
2013-11-14	51.49016571	47.92824562
2013-11-15	53.9958725	49.3714092
2013-11-18	54.17012405	50.28064275
2013-11-19	38.58275795	50.10121465

Appendix I

Predicted Results Using Multi-Layer LSTM with Technical Analysis Only

Predicted results using Multi-layer LSTM with technical analysis only

	actual price	predicted price
2012-10-22	46.50396347	47.38744354
2012-10-23	46.12223816	45.84207535
2012-10-24	46.41620255	45.43633652
2012-10-25	45.48604584	45.51340866
2012-10-26	46.16173172	45.19984055
2012-10-31	48.01328659	45.59539032
2012-11-01	47.52626419	48.01369095
2012-11-02	47.14455414	48.00914001
2012-11-05	47.04801559	46.88953781
2012-11-06	46.55662537	46.50809479
2012-11-07	45.85021591	46.31642914
2012-11-08	45.82827377	45.48814774
2012-11-09	45.99500656	45.208992
2012-11-12	44.51200104	45.27980042
2012-11-13	44.65678024	45.01916122
2012-11-14	44.18732071	44.98215866
2012-11-15	44.76647568	44.76108932
2012-11-16	45.71858978	44.882061
2012-11-19	45.3588028	45.14321899
2012-11-20	46.00377655	44.99658203
2012-11-21	47.61839676	45.19597626
2012-11-23	46.84617615	47.08544159
2012-11-26	45.8589859	45.93825531
2012-11-27	45.82387924	45.24469376
2012-11-28	45.854599	45.08943176
2012-11-29	46.46446991	45.18496704
2012-11-30	45.9774437	45.41641998
2012-12-03	46.66630554	45.2566452
2012-12-04	47.75442886	45.67935562
2012-12-05	47.79389954	47.51677704
2012-12-06	48.21511459	48.03634644
2012-12-07	48.90396881	48.30226135
2012-12-10	49.40415573	48.39974594
2012-12-11	49.15843201	48.89582443
2012-12-12	49.00926971	48.59082794
2012-12-13	49.25936508	48.5161705
2012-12-14	48.63632965	48.51764679
2012-12-17	49.206707	48.26768875
2012-12-18	49.81220245	48.51031876
2012-12-19	50.58441162	48.88219833
2012-12-20	49.6279068	49.10952377
2012-12-21	49.04875946	48.84472656
2012-12-24	49.47873688	48.53837967
2012-12-26	49.18476868	48.82235718
2012-12-27	49.206707	48.55841064

Predicted results using Multi-layer LSTM with technical analysis only

2012-12-28	50.42207718	48.4969368
2012-12-31	51.88751221	48.96929932
2013-01-02	51.77344513	49.69329453
2013-01-03	51.98843765	49.66464996
2013-01-04	51.32152176	49.621521
2013-01-07	51.10652542	49.49058914
2013-01-08	51.46631241	49.4361496
2013-01-09	52.30433273	49.52687073
2013-01-10	52.27801514	49.79399109
2013-01-11	52.24290848	49.92421341
2013-01-14	52.01476288	49.80375671
2013-01-15	51.30835724	49.73814774
2013-01-16	51.3434639	49.46967316
2013-01-17	52.07180405	49.60404968
2013-01-18	52.74748611	49.7606163
2013-01-22	52.08496094	50.00806046
2013-01-23	52.4491272	49.82704544
2013-01-24	53.2388916	49.91070938
2013-01-25	52.77380371	50.3115654
2013-01-28	53.8224411	50.11123276
2013-01-29	53.15554047	50.32538605
2013-01-30	53.28716278	50.24583054
2013-01-31	53.22134399	50.24104691
2013-02-01	49.39099884	50.24825668
2013-02-04	49.94382858	50.74630356
2013-02-05	50.3825798	49.03566742
2013-02-06	48.48275757	49.10303879
2013-02-07	48.97416306	48.71273041
2013-02-08	48.56612396	48.46077728
2013-02-11	48.64948654	48.32818604
2013-02-12	48.96101379	48.30348206
2013-02-13	48.7898941	48.38282776
2013-02-14	49.11457062	48.29654312
2013-02-15	49.64108276	48.54111862
2013-02-19	49.71566772	48.95072937
2013-02-20	49.52699661	48.98885727
2013-02-21	49.46995926	48.74487305
2013-02-22	48.94783783	48.73706818
2013-02-25	49.29007339	48.62488174
2013-02-26	49.8823967	48.5779686
2013-02-27	49.82536316	48.66769791
2013-02-28	50.9705162	48.86024094
2013-03-01	50.33431625	49.25738907
2013-03-04	50.42207718	48.96781158
2013-03-05	50.68532944	49.10985184
2013-03-06	50.23778534	49.12821579

Predicted results using Multi-layer LSTM with technical analysis only

2013-03-07	50.93101883	49.22030258
2013-03-08	50.72481155	49.194767
2013-03-11	50.06229782	49.14220428
2013-03-12	49.42170715	49.01989746
2013-03-13	49.89555359	48.68868637
2013-03-14	49.71128464	48.76150513
2013-03-15	48.9083519	48.75886917
2013-03-18	49.14527893	48.38440704
2013-03-19	50.26851273	48.64196777
2013-03-20	48.98292923	48.96562576
2013-03-21	49.82097244	48.64056778
2013-03-22	50.48349762	48.66193771
2013-03-25	51.13285446	49.09349442
2013-03-26	51.41804504	49.25201416
2013-03-27	51.29957199	49.34769821
2013-03-28	51.13285446	49.46827316
2013-04-01	51.44436646	49.36051559
2013-04-02	50.57123947	49.50531006
2013-04-03	50.63267517	49.28178787
2013-04-04	49.77709198	49.14225769
2013-04-05	49.85168457	48.77231979
2013-04-08	50.1895256	48.7401886
2013-04-09	50.04474258	48.9416008
2013-04-10	50.34747696	49.01551056
2013-04-11	49.77270126	48.94301987
2013-04-12	47.53942108	48.78990555
2013-04-15	48.26337051	48.18718719
2013-04-16	46.88568115	47.96802521
2013-04-17	47.10066223	46.86809921
2013-04-18	47.79389954	46.58068848
2013-04-19	47.79389954	47.60869217
2013-04-22	47.10066223	47.59626007
2013-04-23	47.63156128	46.65046692
2013-04-24	47.73686981	47.06664276
2013-04-25	47.57452774	47.62775803
2013-04-26	48.02645111	47.61281967
2013-04-29	48.50908661	48.25254822
2013-04-30	47.69300079	47.95119476
2013-05-01	47.71054077	47.77397919
2013-05-02	47.87726974	47.02767563
2013-05-03	47.95624542	47.44317627
2013-05-06	48.54856873	47.54392242
2013-05-07	49.61915207	48.11891937
2013-05-08	48.75039673	48.58937836
2013-05-09	48.92591095	48.34445572
2013-05-10	48.4301033	48.233181

Predicted results using Multi-layer LSTM with technical analysis only

2013-05-13	48.67142868	48.10125351
2013-05-14	48.56612396	48.10474777
2013-05-15	48.16685104	48.05184174
2013-05-16	48.43888855	47.76303864
2013-05-17	48.79866791	48.0575943
2013-05-20	48.56612396	48.15138245
2013-05-21	47.86850357	48.12637329
2013-05-22	47.35075378	47.98207474
2013-05-23	46.54784012	46.7226181
2013-05-24	46.87690353	45.48553848
2013-05-28	45.80634308	46.17178726
2013-05-29	45.94673538	45.30537796
2013-05-30	44.66995621	45.23396683
2013-05-31	45.55186081	45.04030609
2013-06-03	45.24034119	44.98909378
2013-06-04	44.27506638	44.95913696
2013-06-05	44.14345169	44.80475616
2013-06-06	45.02861023	44.72454834
2013-06-07	43.18360138	44.39337921
2013-06-10	42.87532806	43.95802689
2013-06-11	42.2270813	43.73806381
2013-06-12	42.50814438	43.61241913
2013-06-13	41.74655533	43.57090378
2013-06-14	42.63960266	43.38235092
2013-06-17	43.05212784	43.66577911
2013-06-18	41.61961365	43.72052002
2013-06-19	39.59326553	43.46405792
2013-06-20	40.35485077	42.67036819
2013-06-21	38.86794662	42.42351532
2013-06-24	39.91059494	41.60208511
2013-06-25	40.58151627	42.37314987
2013-06-26	41.07110214	42.49115753
2013-06-27	41.47909546	42.95802307
2013-06-28	40.98044586	43.13938141
2013-07-01	40.16809464	43.00016403
2013-07-02	39.73198318	42.36654663
2013-07-03	40.00307465	41.9186821
2013-07-05	40.10915375	42.23974609
2013-07-08	40.45685959	42.19024658
2013-07-09	40.9459877	42.5798645
2013-07-10	42.51358414	42.62949753
2013-07-11	41.83585739	43.4924469
2013-07-12	41.90658569	43.1186142
2013-07-15	42.09516144	43.196064
2013-07-16	42.60786819	43.16321564
2013-07-17	42.37214661	43.46054459

Predicted results using Multi-layer LSTM with technical analysis only

2013-07-18	42.70804977	43.36778259
2013-07-19	42.72573853	43.47563553
2013-07-22	44.55850983	43.56938934
2013-07-23	44.41119766	44.26687241
2013-07-24	44.66460037	44.41761017
2013-07-25	44.51137543	44.25021362
2013-07-26	43.3680954	44.21871567
2013-07-29	43.47417068	43.96598434
2013-07-30	43.78060913	43.90541458
2013-07-31	44.2579689	44.02175903
2013-08-01	44.4052887	44.29741669
2013-08-02	44.02223206	44.2093277
2013-08-05	43.36220932	44.20304489
2013-08-06	42.34857941	44.00279999
2013-08-07	42.97325134	43.65412903
2013-08-08	43.06164551	43.62632751
2013-08-09	44.30511475	43.69895554
2013-08-12	44.56441879	44.16439819
2013-08-13	44.72943115	44.27425003
2013-08-14	44.44065475	44.288414
2013-08-15	44.29333115	44.34077072
2013-08-16	44.41119766	44.26707458
2013-08-19	43.99277496	44.32548141
2013-08-20	43.00271225	44.0687561
2013-08-21	43.95152283	43.78738785
2013-08-22	44.24028015	43.99574661
2013-08-23	43.89258194	44.04968643
2013-08-26	43.45059586	44.45105743
2013-08-27	43.27969742	44.10222626
2013-08-28	42.92021561	43.88196182
2013-08-29	42.41930008	43.69125366
2013-08-30	45.0123024	43.53410339
2013-09-03	45.8538208	44.60774612
2013-09-04	46.24439621	44.38042831
2013-09-05	46.83326721	44.52983856
2013-09-06	48.26939392	44.79326248
2013-09-09	48.22733688	45.19031906
2013-09-10	47.54231644	45.26056671
2013-09-11	46.76717758	45.04468918
2013-09-12	46.86932373	44.88473892
2013-09-13	47.58439636	44.83481216
2013-09-16	47.66249466	45.08370209
2013-09-17	48.5938797	45.08222961
2013-09-18	48.605896	45.30683136
2013-09-19	47.7706604	45.43247604
2013-09-20	47.90886688	45.26048279

Predicted results using Multi-layer LSTM with technical analysis only

2013-09-23	47.29595566	45.1486702
2013-09-24	47.31398392	45.01883316
2013-09-25	47.86079407	45.0055542
2013-09-26	47.64447784	45.10764694
2013-09-27	47.0435791	45.06307983
2013-09-30	47.27793503	44.93878555
2013-10-01	47.75263214	44.96155548
2013-10-02	47.80670929	45.01795578
2013-10-03	48.34750748	45.1067009
2013-10-04	48.05308151	45.17307281
2013-10-07	47.44016647	45.12551498
2013-10-08	47.68052292	45.09762573
2013-10-09	48.29343796	45.08081055
2013-10-10	48.73207474	45.22125626
2013-10-11	48.67199326	45.26958466
2013-10-14	48.13719559	45.33006668
2013-10-15	48.28141403	45.24481583
2013-10-16	48.34149551	45.20446396
2013-10-17	48.39558792	45.18737411
2013-10-18	48.30544662	45.27148056
2013-10-21	48.47369385	45.24692154
2013-10-22	47.09166336	45.36139679
2013-10-23	46.8152504	45.07495499
2013-10-24	46.62896729	44.86063385
2013-10-25	46.05211639	44.78477097
2013-10-28	48.22132492	44.71849442
2013-10-29	47.7105751	45.24687195
2013-10-30	48.29343796	45.1678009
2013-10-31	48.91835785	45.46051788
2013-11-01	49.61539078	46.07417679
2013-11-04	49.8196907	46.80044174
2013-11-05	49.50722122	47.56609344
2013-11-06	48.58184814	47.48892593
2013-11-07	48.69602966	46.30974197
2013-11-08	49.10463715	45.37539673
2013-11-11	49.54327011	46.26986313
2013-11-12	48.88230896	47.31125259
2013-11-13	49.27889252	45.42591095
2013-11-14	51.49016571	45.88447571
2013-11-15	53.9958725	48.56448746
2013-11-18	54.17012405	49.57865906
2013-11-19	38.58275795	49.79971313

Appendix J

Predicted Results Using SVR with Technical and Sentiment Analysis (NLTK)

Predicted results using SVR with technical and sentiment analysis (NLTK)

	actual_price	predicted_price
2012-10-22	46.50396347	46.43909685
2012-10-23	46.12223816	46.35624554
2012-10-24	46.41620255	46.37479683
2012-10-25	45.48604584	46.58808299
2012-10-26	46.16173172	45.76496024
2012-10-31	48.01328659	46.2707241
2012-11-01	47.52626419	47.98110631
2012-11-02	47.14455414	48.11257042
2012-11-05	47.04801559	47.68232105
2012-11-06	46.55662537	47.87169094
2012-11-07	45.85021591	47.32342995
2012-11-08	45.82827377	46.80849402
2012-11-09	45.99500656	46.64270516
2012-11-12	44.51200104	46.97919497
2012-11-13	44.65678024	45.3316147
2012-11-14	44.18732071	45.6018155
2012-11-15	44.76647568	45.23168263
2012-11-16	45.71858978	45.26803865
2012-11-19	45.3588028	45.92139644
2012-11-20	46.00377655	45.90221686
2012-11-21	47.61839676	46.41405623
2012-11-23	46.84617615	47.94600125
2012-11-26	45.8589859	47.51378655
2012-11-27	45.82387924	46.62004892
2012-11-28	45.854599	46.31815257
2012-11-29	46.46446991	46.51241186
2012-11-30	45.9774437	47.02208941
2012-12-03	46.66630554	46.64916216
2012-12-04	47.75442886	47.31613345
2012-12-05	47.79389954	48.24997506
2012-12-06	48.21511459	47.95348641
2012-12-07	48.90396881	48.53084845
2012-12-10	49.40415573	49.38708551
2012-12-11	49.15843201	49.33264485
2012-12-12	49.00926971	49.94856995
2012-12-13	49.25936508	49.94639552
2012-12-14	48.63632965	50.11123871
2012-12-17	49.206707	49.79914304
2012-12-18	49.81220245	50.06551646
2012-12-19	50.58441162	50.29254194
2012-12-20	49.6279068	50.41818236
2012-12-21	49.04875946	50.18210624
2012-12-24	49.47873688	50.22181253
2012-12-26	49.18476868	50.14214322
2012-12-27	49.206707	50.20521267

Predicted results using SVR with technical and sentiment analysis (NLTK)

2012-12-28	50.42207718	50.24472762
2012-12-31	51.88751221	50.36839918
2013-01-02	51.77344513	50.2927338
2013-01-03	51.98843765	50.23866011
2013-01-04	51.32152176	50.28734969
2013-01-07	51.10652542	50.1684675
2013-01-08	51.46631241	50.15705971
2013-01-09	52.30433273	50.05455487
2013-01-10	52.27801514	50.19547242
2013-01-11	52.24290848	49.85760143
2013-01-14	52.01476288	50.1209952
2013-01-15	51.30835724	50.10459297
2013-01-16	51.3434639	50.120655
2013-01-17	52.07180405	49.88018186
2013-01-18	52.74748611	49.98149376
2013-01-22	52.08496094	49.9075537
2013-01-23	52.4491272	49.73793213
2013-01-24	53.2388916	49.75609276
2013-01-25	52.77380371	49.45165257
2013-01-28	53.8224411	49.70353912
2013-01-29	53.15554047	49.54940549
2013-01-30	53.28716278	49.51709448
2013-01-31	53.22134399	49.38469685
2013-02-01	49.39099884	49.31926736
2013-02-04	49.94382858	47.70394463
2013-02-05	50.3825798	49.72603281
2013-02-06	48.48275757	49.97101488
2013-02-07	48.97416306	49.65213683
2013-02-08	48.56612396	49.99989164
2013-02-11	48.64948654	50.10034054
2013-02-12	48.96101379	50.06508531
2013-02-13	48.7898941	50.02853798
2013-02-14	49.11457062	50.01679521
2013-02-15	49.64108276	50.02417572
2013-02-19	49.71566772	49.83891989
2013-02-20	49.52699661	49.87626507
2013-02-21	49.46995926	50.00245548
2013-02-22	48.94783783	50.06022051
2013-02-25	49.29007339	50.15085896
2013-02-26	49.8823967	50.14525934
2013-02-27	49.82536316	50.45405458
2013-02-28	50.9705162	50.33851356
2013-03-01	50.33431625	50.39562798
2013-03-04	50.42207718	50.54744732
2013-03-05	50.68532944	50.49758132
2013-03-06	50.23778534	50.57213753

Predicted results using SVR with technical and sentiment analysis (NLTK)

2013-03-07	50.93101883	50.22904889
2013-03-08	50.72481155	50.54392431
2013-03-11	50.06229782	50.58305407
2013-03-12	49.42170715	50.44855739
2013-03-13	49.89555359	50.34116001
2013-03-14	49.71128464	50.46911321
2013-03-15	48.9083519	50.04235867
2013-03-18	49.14527893	50.17646525
2013-03-19	50.26851273	49.96421918
2013-03-20	48.98292923	50.43994321
2013-03-21	49.82097244	49.93711512
2013-03-22	50.48349762	50.45721175
2013-03-25	51.13285446	50.58358081
2013-03-26	51.41804504	50.57023817
2013-03-27	51.29957199	50.68134646
2013-03-28	51.13285446	50.34920024
2013-04-01	51.44436646	50.46077646
2013-04-02	50.57123947	50.35082869
2013-04-03	50.63267517	50.4696774
2013-04-04	49.77709198	50.49704283
2013-04-05	49.85168457	50.19755768
2013-04-08	50.1895256	50.46149812
2013-04-09	50.04474258	50.47697055
2013-04-10	50.34747696	50.21615102
2013-04-11	49.77270126	50.50796707
2013-04-12	47.53942108	50.42502459
2013-04-15	48.26337051	49.41670419
2013-04-16	46.88568115	49.89714511
2013-04-17	47.10066223	49.03898076
2013-04-18	47.79389954	49.22396349
2013-04-19	47.79389954	49.39226704
2013-04-22	47.10066223	49.52867556
2013-04-23	47.63156128	48.6921967
2013-04-24	47.73686981	49.06438845
2013-04-25	47.57452774	49.13270893
2013-04-26	48.02645111	49.04327934
2013-04-29	48.50908661	48.43098189
2013-04-30	47.69300079	49.57576108
2013-05-01	47.71054077	49.34042364
2013-05-02	47.87726974	49.17854759
2013-05-03	47.95624542	49.33255521
2013-05-06	48.54856873	49.38783177
2013-05-07	49.61915207	49.82414543
2013-05-08	48.75039673	50.02786403
2013-05-09	48.92591095	49.96605949
2013-05-10	48.4301033	50.15151427

Predicted results using SVR with technical and sentiment analysis (NLTK)

2013-05-13	48.67142868	49.88628416
2013-05-14	48.56612396	49.93579535
2013-05-15	48.16685104	49.92415644
2013-05-16	48.43888855	49.6631095
2013-05-17	48.79866791	49.74083969
2013-05-20	48.56612396	50.07872508
2013-05-21	47.86850357	50.04230683
2013-05-22	47.35075378	49.5267848
2013-05-23	46.54784012	48.64337641
2013-05-24	46.87690353	47.99134943
2013-05-28	45.80634308	48.38430767
2013-05-29	45.94673538	47.33635715
2013-05-30	44.66995621	47.31042393
2013-05-31	45.55186081	46.10072666
2013-06-03	45.24034119	46.46779739
2013-06-04	44.27506638	46.27150185
2013-06-05	44.14345169	45.42344109
2013-06-06	45.02861023	45.02528939
2013-06-07	43.18360138	44.7887908
2013-06-10	42.87532806	44.27804825
2013-06-11	42.2270813	44.22600796
2013-06-12	42.50814438	44.23730253
2013-06-13	41.74655533	44.11018436
2013-06-14	42.63960266	44.17127055
2013-06-17	43.05212784	44.19051804
2013-06-18	41.61961365	44.36132397
2013-06-19	39.59326553	44.2027632
2013-06-20	40.35485077	44.36062183
2013-06-21	38.86794662	44.73230228
2013-06-24	39.91059494	44.83499512
2013-06-25	40.58151627	44.71497148
2013-06-26	41.07110214	44.82897091
2013-06-27	41.47909546	44.90630886
2013-06-28	40.98044586	44.90872022
2013-07-01	40.16809464	45.08757423
2013-07-02	39.73198318	45.2650514
2013-07-03	40.00307465	45.37501801
2013-07-05	40.10915375	45.38805688
2013-07-08	40.45685959	45.49086278
2013-07-09	40.9459877	45.34884355
2013-07-10	42.51358414	45.50955639
2013-07-11	41.83585739	45.24910518
2013-07-12	41.90658569	45.49302987
2013-07-15	42.09516144	45.44829622
2013-07-16	42.60786819	45.39259906
2013-07-17	42.37214661	45.41752528

Predicted results using SVR with technical and sentiment analysis (NLTK)

2013-07-18	42.70804977	45.47928618
2013-07-19	42.72573853	45.37589705
2013-07-22	44.55850983	45.19178818
2013-07-23	44.41119766	44.87784735
2013-07-24	44.66460037	44.68009535
2013-07-25	44.51137543	44.87971859
2013-07-26	43.3680954	44.95296736
2013-07-29	43.47417068	44.84570806
2013-07-30	43.78060913	44.97086252
2013-07-31	44.2579689	44.77318046
2013-08-01	44.4052887	44.63219316
2013-08-02	44.02223206	44.7656958
2013-08-05	43.36220932	44.77464316
2013-08-06	42.34857941	44.65144942
2013-08-07	42.97325134	44.59574247
2013-08-08	43.06164551	44.89686634
2013-08-09	44.30511475	44.90608338
2013-08-12	44.56441879	44.78908624
2013-08-13	44.72943115	44.90220395
2013-08-14	44.44065475	44.96671729
2013-08-15	44.29333115	44.84379922
2013-08-16	44.41119766	44.90485753
2013-08-19	43.99277496	44.82330297
2013-08-20	43.00271225	44.8122936
2013-08-21	43.95152283	44.69771368
2013-08-22	44.24028015	44.8481118
2013-08-23	43.89258194	44.95489789
2013-08-26	43.45059586	44.37041201
2013-08-27	43.27969742	44.68882476
2013-08-28	42.92021561	44.68962158
2013-08-29	42.41930008	44.83274677
2013-08-30	45.0123024	44.83524853
2013-09-03	45.8538208	44.53644592
2013-09-04	46.24439621	44.86618806
2013-09-05	46.83326721	44.99139609
2013-09-06	48.26939392	45.02736136
2013-09-09	48.22733688	45.4084592
2013-09-10	47.54231644	45.61541382
2013-09-11	46.76717758	45.42390832
2013-09-12	46.86932373	45.09332505
2013-09-13	47.58439636	45.10113806
2013-09-16	47.66249466	45.57899409
2013-09-17	48.5938797	45.68894889
2013-09-18	48.605896	45.95946134
2013-09-19	47.7706604	46.65619702
2013-09-20	47.90886688	46.02856602

Predicted results using SVR with technical and sentiment analysis (NLTK)

2013-09-23	47.29595566	45.91358488
2013-09-24	47.31398392	45.50610227
2013-09-25	47.86079407	45.61273942
2013-09-26	47.64447784	46.07050494
2013-09-27	47.0435791	46.03768542
2013-09-30	47.27793503	45.33335769
2013-10-01	47.75263214	45.66078999
2013-10-02	47.80670929	46.03913395
2013-10-03	48.34750748	46.28579779
2013-10-04	48.05308151	46.61716742
2013-10-07	47.44016647	46.47068606
2013-10-08	47.68052292	46.03019785
2013-10-09	48.29343796	46.1679357
2013-10-10	48.73207474	46.63332048
2013-10-11	48.67199326	47.04685949
2013-10-14	48.13719559	47.05617944
2013-10-15	48.28141403	46.75781965
2013-10-16	48.34149551	46.73154311
2013-10-17	48.39558792	46.73747048
2013-10-18	48.30544662	46.84149206
2013-10-21	48.47369385	46.82435979
2013-10-22	47.09166336	47.026689
2013-10-23	46.8152504	45.51930973
2013-10-24	46.62896729	45.4766845
2013-10-25	46.05211639	45.33455375
2013-10-28	48.22132492	44.85702666
2013-10-29	47.7105751	45.88408522
2013-10-30	48.29343796	46.09727317
2013-10-31	48.91835785	46.52030253
2013-11-01	49.61539078	47.03157378
2013-11-04	49.8196907	47.80656785
2013-11-05	49.50722122	47.93902962
2013-11-06	48.58184814	48.1367496
2013-11-07	48.69602966	47.38101849
2013-11-08	49.10463715	47.19611404
2013-11-11	49.54327011	47.78656215
2013-11-12	48.88230896	48.24646541
2013-11-13	49.27889252	47.42347211
2013-11-14	51.49016571	47.93248018
2013-11-15	53.9958725	49.28860141
2013-11-18	54.17012405	50.28512294
2013-11-19	38.58275795	50.14118729

Appendix K

Predicted Results Using SVR with Technical and Sentiment Analysis (LMD)

Predicted results using SVR with technical and sentiment analysis (LMD)

	actual_price	predicted_price
2012-10-22	46.50396347	46.43554356
2012-10-23	46.12223816	46.41019042
2012-10-24	46.41620255	46.30295157
2012-10-25	45.48604584	46.53220406
2012-10-26	46.16173172	46.00058129
2012-10-31	48.01328659	46.26899383
2012-11-01	47.52626419	47.74203743
2012-11-02	47.14455414	47.90845526
2012-11-05	47.04801559	47.60431507
2012-11-06	46.55662537	47.83292278
2012-11-07	45.85021591	47.25864254
2012-11-08	45.82827377	46.78077691
2012-11-09	45.99500656	46.73913008
2012-11-12	44.51200104	46.93593171
2012-11-13	44.65678024	45.56582545
2012-11-14	44.18732071	45.73399382
2012-11-15	44.76647568	45.2938612
2012-11-16	45.71858978	45.53096669
2012-11-19	45.3588028	46.12908617
2012-11-20	46.00377655	46.02885739
2012-11-21	47.61839676	46.59549077
2012-11-23	46.84617615	47.83338379
2012-11-26	45.8589859	47.50394355
2012-11-27	45.82387924	46.80083549
2012-11-28	45.854599	46.54236126
2012-11-29	46.46446991	46.92214031
2012-11-30	45.9774437	47.12495286
2012-12-03	46.66630554	46.72239415
2012-12-04	47.75442886	47.59364842
2012-12-05	47.79389954	48.14878141
2012-12-06	48.21511459	47.94588888
2012-12-07	48.90396881	48.267558
2012-12-10	49.40415573	49.14172268
2012-12-11	49.15843201	49.13625287
2012-12-12	49.00926971	49.7731959
2012-12-13	49.25936508	49.69257528
2012-12-14	48.63632965	49.90886773
2012-12-17	49.206707	49.52172434
2012-12-18	49.81220245	49.79543538
2012-12-19	50.58441162	50.10835291
2012-12-20	49.6279068	50.51211311
2012-12-21	49.04875946	50.12915502
2012-12-24	49.47873688	49.91810372
2012-12-26	49.18476868	49.92418532
2012-12-27	49.206707	50.07983623

Predicted results using SVR with technical and sentiment analysis (LMD)

2012-12-28	50.42207718	50.17321921
2012-12-31	51.88751221	50.34761323
2013-01-02	51.77344513	50.53761293
2013-01-03	51.98843765	50.90462564
2013-01-04	51.32152176	51.18943576
2013-01-07	51.10652542	50.7697553
2013-01-08	51.46631241	50.71846499
2013-01-09	52.30433273	50.91925085
2013-01-10	52.27801514	51.06671635
2013-01-11	52.24290848	50.73513103
2013-01-14	52.01476288	51.12262667
2013-01-15	51.30835724	50.96473982
2013-01-16	51.3434639	50.78010379
2013-01-17	52.07180405	50.63322266
2013-01-18	52.74748611	50.8375881
2013-01-22	52.08496094	50.88329958
2013-01-23	52.4491272	50.8945034
2013-01-24	53.2388916	51.41170769
2013-01-25	52.77380371	51.09093499
2013-01-28	53.8224411	50.92966866
2013-01-29	53.15554047	51.15232577
2013-01-30	53.28716278	50.95150573
2013-01-31	53.22134399	51.53834368
2013-02-01	49.39099884	51.3228526
2013-02-04	49.94382858	48.24574233
2013-02-05	50.3825798	50.24291479
2013-02-06	48.48275757	50.74070718
2013-02-07	48.97416306	49.87067536
2013-02-08	48.56612396	50.23373826
2013-02-11	48.64948654	49.8932814
2013-02-12	48.96101379	50.21282384
2013-02-13	48.7898941	50.16652752
2013-02-14	49.11457062	50.00706925
2013-02-15	49.64108276	50.01575184
2013-02-19	49.71566772	49.91611933
2013-02-20	49.52699661	50.17482197
2013-02-21	49.46995926	50.00421266
2013-02-22	48.94783783	50.15480042
2013-02-25	49.29007339	50.07633683
2013-02-26	49.8823967	50.25588636
2013-02-27	49.82536316	50.53176937
2013-02-28	50.9705162	50.42573152
2013-03-01	50.33431625	50.99056735
2013-03-04	50.42207718	50.65128757
2013-03-05	50.68532944	50.60533009
2013-03-06	50.23778534	50.81009468

Predicted results using SVR with technical and sentiment analysis (LMD)

2013-03-07	50.93101883	50.38861725
2013-03-08	50.72481155	50.8087889
2013-03-11	50.06229782	50.78933462
2013-03-12	49.42170715	50.53684992
2013-03-13	49.89555359	50.3259692
2013-03-14	49.71128464	50.48776276
2013-03-15	48.9083519	50.09740019
2013-03-18	49.14527893	50.03830614
2013-03-19	50.26851273	49.77195871
2013-03-20	48.98292923	50.59279827
2013-03-21	49.82097244	49.78635336
2013-03-22	50.48349762	50.34673404
2013-03-25	51.13285446	50.64098034
2013-03-26	51.41804504	50.8916475
2013-03-27	51.29957199	51.4099706
2013-03-28	51.13285446	51.1132778
2013-04-01	51.44436646	50.99242053
2013-04-02	50.57123947	50.86724183
2013-04-03	50.63267517	50.88160749
2013-04-04	49.77709198	50.88371487
2013-04-05	49.85168457	50.48329083
2013-04-08	50.1895256	50.67805903
2013-04-09	50.04474258	50.58128693
2013-04-10	50.34747696	50.55277374
2013-04-11	49.77270126	50.8649573
2013-04-12	47.53942108	50.55000961
2013-04-15	48.26337051	49.19039297
2013-04-16	46.88568115	49.70656255
2013-04-17	47.10066223	48.75577017
2013-04-18	47.79389954	48.83252177
2013-04-19	47.79389954	49.11434656
2013-04-22	47.10066223	49.12928162
2013-04-23	47.63156128	48.32590318
2013-04-24	47.73686981	48.55200803
2013-04-25	47.57452774	48.80555775
2013-04-26	48.02645111	48.69452317
2013-04-29	48.50908661	48.30725453
2013-04-30	47.69300079	49.30259212
2013-05-01	47.71054077	48.98541576
2013-05-02	47.87726974	48.9280522
2013-05-03	47.95624542	48.94996929
2013-05-06	48.54856873	49.02609944
2013-05-07	49.61915207	49.48911732
2013-05-08	48.75039673	49.66556986
2013-05-09	48.92591095	49.67128879
2013-05-10	48.4301033	49.7140708

Predicted results using SVR with technical and sentiment analysis (LMD)

2013-05-13	48.67142868	49.51645255
2013-05-14	48.56612396	49.54823682
2013-05-15	48.16685104	49.45192653
2013-05-16	48.43888855	49.27320172
2013-05-17	48.79866791	49.4487866
2013-05-20	48.56612396	49.69704942
2013-05-21	47.86850357	49.62950958
2013-05-22	47.35075378	49.20563908
2013-05-23	46.54784012	48.31479672
2013-05-24	46.87690353	47.81003305
2013-05-28	45.80634308	48.05205535
2013-05-29	45.94673538	47.17455975
2013-05-30	44.66995621	47.2790355
2013-05-31	45.55186081	46.14303103
2013-06-03	45.24034119	46.5659681
2013-06-04	44.27506638	46.23871572
2013-06-05	44.14345169	45.44384658
2013-06-06	45.02861023	45.21708819
2013-06-07	43.18360138	44.78686731
2013-06-10	42.87532806	44.10849272
2013-06-11	42.2270813	43.83800631
2013-06-12	42.50814438	43.82690693
2013-06-13	41.74655533	43.76113167
2013-06-14	42.63960266	43.37283668
2013-06-17	43.05212784	43.56147721
2013-06-18	41.61961365	43.94802666
2013-06-19	39.59326553	43.74978279
2013-06-20	40.35485077	43.05738856
2013-06-21	38.86794662	43.20880948
2013-06-24	39.91059494	43.06824678
2013-06-25	40.58151627	43.25554718
2013-06-26	41.07110214	43.45224463
2013-06-27	41.47909546	43.35418419
2013-06-28	40.98044586	43.46711713
2013-07-01	40.16809464	43.53899422
2013-07-02	39.73198318	43.65961494
2013-07-03	40.00307465	43.62339263
2013-07-05	40.10915375	43.37519057
2013-07-08	40.45685959	43.58383631
2013-07-09	40.9459877	43.62211412
2013-07-10	42.51358414	43.63845351
2013-07-11	41.83585739	43.78866949
2013-07-12	41.90658569	43.68862619
2013-07-15	42.09516144	43.839028
2013-07-16	42.60786819	44.00703612
2013-07-17	42.37214661	43.99994962

Predicted results using SVR with technical and sentiment analysis (LMD)

2013-07-18	42.70804977	43.83848916
2013-07-19	42.72573853	44.03987579
2013-07-22	44.55850983	43.89001765
2013-07-23	44.41119766	44.20276724
2013-07-24	44.66460037	43.97619865
2013-07-25	44.51137543	44.24102037
2013-07-26	43.3680954	44.31511522
2013-07-29	43.47417068	43.98383168
2013-07-30	43.78060913	44.24001184
2013-07-31	44.2579689	43.89635662
2013-08-01	44.4052887	44.02432354
2013-08-02	44.02223206	44.41156782
2013-08-05	43.36220932	44.12143313
2013-08-06	42.34857941	43.85638964
2013-08-07	42.97325134	43.7236758
2013-08-08	43.06164551	43.85747482
2013-08-09	44.30511475	44.07380363
2013-08-12	44.56441879	44.11615816
2013-08-13	44.72943115	44.2993499
2013-08-14	44.44065475	44.49737509
2013-08-15	44.29333115	44.23773806
2013-08-16	44.41119766	44.42312048
2013-08-19	43.99277496	44.43580542
2013-08-20	43.00271225	44.28486546
2013-08-21	43.95152283	43.96833763
2013-08-22	44.24028015	44.03065462
2013-08-23	43.89258194	44.45871276
2013-08-26	43.45059586	43.65593801
2013-08-27	43.27969742	44.09110927
2013-08-28	42.92021561	43.94280026
2013-08-29	42.41930008	44.0589694
2013-08-30	45.0123024	43.66307995
2013-09-03	45.8538208	44.34991366
2013-09-04	46.24439621	44.61321105
2013-09-05	46.83326721	44.7817799
2013-09-06	48.26939392	45.18959529
2013-09-09	48.22733688	45.66273901
2013-09-10	47.54231644	45.82114624
2013-09-11	46.76717758	45.68724322
2013-09-12	46.86932373	45.14909133
2013-09-13	47.58439636	45.26310381
2013-09-16	47.66249466	45.78987003
2013-09-17	48.5938797	45.8534171
2013-09-18	48.605896	46.21082674
2013-09-19	47.7706604	46.72371924
2013-09-20	47.90886688	46.27621762

Predicted results using SVR with technical and sentiment analysis (LMD)

2013-09-23	47.29595566	46.02527096
2013-09-24	47.31398392	45.62659964
2013-09-25	47.86079407	45.89011267
2013-09-26	47.64447784	46.33097185
2013-09-27	47.0435791	46.1143529
2013-09-30	47.27793503	45.55773726
2013-10-01	47.75263214	45.96464325
2013-10-02	47.80670929	46.20033746
2013-10-03	48.34750748	46.35996105
2013-10-04	48.05308151	46.76369079
2013-10-07	47.44016647	46.63184105
2013-10-08	47.68052292	46.13537488
2013-10-09	48.29343796	46.25106681
2013-10-10	48.73207474	46.72423548
2013-10-11	48.67199326	47.26876402
2013-10-14	48.13719559	47.02896304
2013-10-15	48.28141403	46.89482094
2013-10-16	48.34149551	46.82863961
2013-10-17	48.39558792	46.81362345
2013-10-18	48.30544662	46.9915199
2013-10-21	48.47369385	46.9065363
2013-10-22	47.09166336	47.06901992
2013-10-23	46.8152504	45.76162017
2013-10-24	46.62896729	45.68729852
2013-10-25	46.05211639	45.52479907
2013-10-28	48.22132492	45.02356744
2013-10-29	47.7105751	46.0547309
2013-10-30	48.29343796	46.32522503
2013-10-31	48.91835785	46.63542516
2013-11-01	49.61539078	47.04508194
2013-11-04	49.8196907	47.7240207
2013-11-05	49.50722122	47.8795134
2013-11-06	48.58184814	48.08373866
2013-11-07	48.69602966	47.20485253
2013-11-08	49.10463715	47.17879798
2013-11-11	49.54327011	47.5891121
2013-11-12	48.88230896	48.08249365
2013-11-13	49.27889252	47.35189661
2013-11-14	51.49016571	47.82096712
2013-11-15	53.9958725	49.00651672
2013-11-18	54.17012405	50.41108837
2013-11-19	38.58275795	50.6449757

Appendix L

Predicted Results Using Multi-Layer LSTM with Technical and Sentiment Analysis (NLTK)

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (NLTK)

	actual_price	predicted_price
2012-10-22	46.50396347	48.00773239
2012-10-23	46.12223816	46.96313477
2012-10-24	46.41620255	46.70833969
2012-10-25	45.48604584	46.79268646
2012-10-26	46.16173172	46.4621582
2012-10-31	48.01328659	46.90415573
2012-11-01	47.52626419	48.92370224
2012-11-02	47.14455414	48.94688416
2012-11-05	47.04801559	47.52309799
2012-11-06	46.55662537	47.27742386
2012-11-07	45.85021591	47.20538712
2012-11-08	45.82827377	46.81116104
2012-11-09	45.99500656	46.48485565
2012-11-12	44.51200104	46.57540131
2012-11-13	44.65678024	46.13352585
2012-11-14	44.18732071	46.14038086
2012-11-15	44.76647568	45.82577896
2012-11-16	45.71858978	45.99903488
2012-11-19	45.3588028	46.3944931
2012-11-20	46.00377655	46.17167282
2012-11-21	47.61839676	46.46908569
2012-11-23	46.84617615	47.80269623
2012-11-26	45.8589859	46.99155045
2012-11-27	45.82387924	46.51931381
2012-11-28	45.854599	46.31840515
2012-11-29	46.46446991	46.44528961
2012-11-30	45.9774437	46.68378067
2012-12-03	46.66630554	46.53924179
2012-12-04	47.75442886	46.8315506
2012-12-05	47.79389954	48.57231903
2012-12-06	48.21511459	48.89801788
2012-12-07	48.90396881	49.03130341
2012-12-10	49.40415573	49.07063675
2012-12-11	49.15843201	49.23912811
2012-12-12	49.00926971	49.11153793
2012-12-13	49.25936508	49.09246826
2012-12-14	48.63632965	49.09165955
2012-12-17	49.206707	49.02776718
2012-12-18	49.81220245	49.11006927
2012-12-19	50.58441162	49.23347473
2012-12-20	49.6279068	49.30570984
2012-12-21	49.04875946	49.22705078
2012-12-24	49.47873688	49.0871048
2012-12-26	49.18476868	49.21894073
2012-12-27	49.206707	49.10084915

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (NLTK)

2012-12-28	50.42207718	49.07118607
2012-12-31	51.88751221	49.25031662
2013-01-02	51.77344513	49.50291824
2013-01-03	51.98843765	49.46907043
2013-01-04	51.32152176	49.42717743
2013-01-07	51.10652542	49.41053009
2013-01-08	51.46631241	49.3854866
2013-01-09	52.30433273	49.44965363
2013-01-10	52.27801514	49.49826813
2013-01-11	52.24290848	49.58937454
2013-01-14	52.01476288	49.49292374
2013-01-15	51.30835724	49.46772003
2013-01-16	51.3434639	49.38981247
2013-01-17	52.07180405	49.48614883
2013-01-18	52.74748611	49.52088547
2013-01-22	52.08496094	49.58704758
2013-01-23	52.4491272	49.52414703
2013-01-24	53.2388916	49.53104019
2013-01-25	52.77380371	49.7137413
2013-01-28	53.8224411	49.59374619
2013-01-29	53.15554047	49.71016312
2013-01-30	53.28716278	49.66698837
2013-01-31	53.22134399	49.64276123
2013-02-01	49.39099884	49.67224884
2013-02-04	49.94382858	49.95545959
2013-02-05	50.3825798	49.29263687
2013-02-06	48.48275757	49.30234528
2013-02-07	48.97416306	49.18302917
2013-02-08	48.56612396	49.07183838
2013-02-11	48.64948654	49.00733566
2013-02-12	48.96101379	49.00030518
2013-02-13	48.7898941	49.03151321
2013-02-14	49.11457062	48.99634933
2013-02-15	49.64108276	49.08234406
2013-02-19	49.71566772	49.23996735
2013-02-20	49.52699661	49.2506218
2013-02-21	49.46995926	49.1711235
2013-02-22	48.94783783	49.16487885
2013-02-25	49.29007339	49.0957489
2013-02-26	49.8823967	49.09644318
2013-02-27	49.82536316	49.11016846
2013-02-28	50.9705162	49.17814636
2013-03-01	50.33431625	49.32488251
2013-03-04	50.42207718	49.18064499
2013-03-05	50.68532944	49.25965881
2013-03-06	50.23778534	49.23496246

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (NLTK)

2013-03-07	50.93101883	49.30799866
2013-03-08	50.72481155	49.2716713
2013-03-11	50.06229782	49.2346077
2013-03-12	49.42170715	49.20491028
2013-03-13	49.89555359	49.10255051
2013-03-14	49.71128464	49.12447357
2013-03-15	48.9083519	49.15729523
2013-03-18	49.14527893	48.99516678
2013-03-19	50.26851273	49.11473083
2013-03-20	48.98292923	49.21527481
2013-03-21	49.82097244	49.11188507
2013-03-22	50.48349762	49.0847702
2013-03-25	51.13285446	49.23340225
2013-03-26	51.41804504	49.30735779
2013-03-27	51.29957199	49.29875565
2013-03-28	51.13285446	49.38541412
2013-04-01	51.44436646	49.3174324
2013-04-02	50.57123947	49.38734818
2013-04-03	50.63267517	49.29327393
2013-04-04	49.77709198	49.25391006
2013-04-05	49.85168457	49.1665535
2013-04-08	50.1895256	49.11314392
2013-04-09	50.04474258	49.18313217
2013-04-10	50.34747696	49.23613358
2013-04-11	49.77270126	49.16466141
2013-04-12	47.53942108	49.11575317
2013-04-15	48.26337051	48.96356201
2013-04-16	46.88568115	48.86991882
2013-04-17	47.10066223	47.43831635
2013-04-18	47.79389954	47.28847504
2013-04-19	47.79389954	48.6908226
2013-04-22	47.10066223	48.57817459
2013-04-23	47.63156128	47.27593994
2013-04-24	47.73686981	47.7529335
2013-04-25	47.57452774	48.77243423
2013-04-26	48.02645111	48.727314
2013-04-29	48.50908661	48.96489716
2013-04-30	47.69300079	48.87996292
2013-05-01	47.71054077	48.69249344
2013-05-02	47.87726974	47.81581116
2013-05-03	47.95624542	48.34154129
2013-05-06	48.54856873	48.48167419
2013-05-07	49.61915207	48.92734146
2013-05-08	48.75039673	49.11251831
2013-05-09	48.92591095	49.00216675
2013-05-10	48.4301033	48.91702271

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (NLTK)

2013-05-13	48.67142868	48.8766098
2013-05-14	48.56612396	48.89027023
2013-05-15	48.16685104	48.84685516
2013-05-16	48.43888855	48.68030548
2013-05-17	48.79866791	48.89723969
2013-05-20	48.56612396	48.90557861
2013-05-21	47.86850357	48.87439728
2013-05-22	47.35075378	48.85396194
2013-05-23	46.54784012	47.34741211
2013-05-24	46.87690353	46.73417664
2013-05-28	45.80634308	47.04914093
2013-05-29	45.94673538	46.56983566
2013-05-30	44.66995621	46.49721146
2013-05-31	45.55186081	46.1991539
2013-06-03	45.24034119	46.16073227
2013-06-04	44.27506638	46.10798264
2013-06-05	44.14345169	45.8860321
2013-06-06	45.02861023	45.76055908
2013-06-07	43.18360138	45.3537941
2013-06-10	42.87532806	44.87782669
2013-06-11	42.2270813	44.66638565
2013-06-12	42.50814438	44.5520134
2013-06-13	41.74655533	44.50286102
2013-06-14	42.63960266	44.32903671
2013-06-17	43.05212784	44.58145142
2013-06-18	41.61961365	44.63085938
2013-06-19	39.59326553	44.40597534
2013-06-20	40.35485077	43.79695129
2013-06-21	38.86794662	43.68865585
2013-06-24	39.91059494	43.25592041
2013-06-25	40.58151627	43.6026001
2013-06-26	41.07110214	43.70567703
2013-06-27	41.47909546	43.97925568
2013-06-28	40.98044586	44.09755325
2013-07-01	40.16809464	44.04980469
2013-07-02	39.73198318	43.68906021
2013-07-03	40.00307465	43.43250656
2013-07-05	40.10915375	43.60938644
2013-07-08	40.45685959	43.5940361
2013-07-09	40.9459877	43.76289749
2013-07-10	42.51358414	43.85000992
2013-07-11	41.83585739	44.40824509
2013-07-12	41.90658569	44.17758942
2013-07-15	42.09516144	44.22686005
2013-07-16	42.60786819	44.22134399
2013-07-17	42.37214661	44.4497261

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (NLTK)

2013-07-18	42.70804977	44.38312531
2013-07-19	42.72573853	44.46697617
2013-07-22	44.55850983	44.54352951
2013-07-23	44.41119766	45.20412445
2013-07-24	44.66460037	45.33011246
2013-07-25	44.51137543	45.18886948
2013-07-26	43.3680954	45.15497971
2013-07-29	43.47417068	44.90172577
2013-07-30	43.78060913	44.84368515
2013-07-31	44.2579689	44.95558548
2013-08-01	44.4052887	45.21950531
2013-08-02	44.02223206	45.14818954
2013-08-05	43.36220932	45.14273453
2013-08-06	42.34857941	44.93598938
2013-08-07	42.97325134	44.60093307
2013-08-08	43.06164551	44.59399796
2013-08-09	44.30511475	44.65655136
2013-08-12	44.56441879	45.10116959
2013-08-13	44.72943115	45.21782684
2013-08-14	44.44065475	45.23223877
2013-08-15	44.29333115	45.29261017
2013-08-16	44.41119766	45.2102623
2013-08-19	43.99277496	45.27614212
2013-08-20	43.00271225	45.00360107
2013-08-21	43.95152283	44.73498917
2013-08-22	44.24028015	44.93086243
2013-08-23	43.89258194	44.9839325
2013-08-26	43.45059586	45.28220749
2013-08-27	43.27969742	45.00670242
2013-08-28	42.92021561	44.82008743
2013-08-29	42.41930008	44.64899063
2013-08-30	45.0123024	44.50677872
2013-09-03	45.8538208	45.52047348
2013-09-04	46.24439621	45.33853912
2013-09-05	46.83326721	45.51640701
2013-09-06	48.26939392	45.86260986
2013-09-09	48.22733688	46.40736008
2013-09-10	47.54231644	46.5241127
2013-09-11	46.76717758	46.22557068
2013-09-12	46.86932373	45.99642181
2013-09-13	47.58439636	45.93054199
2013-09-16	47.66249466	46.2879715
2013-09-17	48.5938797	46.28726196
2013-09-18	48.605896	46.58594513
2013-09-19	47.7706604	46.70841217
2013-09-20	47.90886688	46.54039764

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (NLTK)

2013-09-23	47.29595566	46.40382004
2013-09-24	47.31398392	46.20827866
2013-09-25	47.86079407	46.18314743
2013-09-26	47.64447784	46.32965088
2013-09-27	47.0435791	46.26285172
2013-09-30	47.27793503	46.09105682
2013-10-01	47.75263214	46.12116623
2013-10-02	47.80670929	46.19645309
2013-10-03	48.34750748	46.31647491
2013-10-04	48.05308151	46.41439056
2013-10-07	47.44016647	46.34966666
2013-10-08	47.68052292	46.31021881
2013-10-09	48.29343796	46.28253937
2013-10-10	48.73207474	46.5074234
2013-10-11	48.67199326	46.55499649
2013-10-14	48.13719559	46.65275574
2013-10-15	48.28141403	46.5209198
2013-10-16	48.34149551	46.47859573
2013-10-17	48.39558792	46.45599365
2013-10-18	48.30544662	46.58545685
2013-10-21	48.47369385	46.53959656
2013-10-22	47.09166336	46.68898392
2013-10-23	46.8152504	46.26026154
2013-10-24	46.62896729	45.973629
2013-10-25	46.05211639	45.86355209
2013-10-28	48.22132492	45.76405716
2013-10-29	47.7105751	46.4723053
2013-10-30	48.29343796	46.41963959
2013-10-31	48.91835785	46.75753403
2013-11-01	49.61539078	46.98188019
2013-11-04	49.8196907	47.39373779
2013-11-05	49.50722122	48.45526505
2013-11-06	48.58184814	48.35184479
2013-11-07	48.69602966	47.1007309
2013-11-08	49.10463715	46.70428848
2013-11-11	49.54327011	47.03546906
2013-11-12	48.88230896	48.1248436
2013-11-13	49.27889252	46.76127625
2013-11-14	51.49016571	46.92199707
2013-11-15	53.9958725	49.11238861
2013-11-18	54.17012405	49.45035553
2013-11-19	38.58275795	49.54118347

Appendix M

Predicted Results Using Multi-Layer LSTM with Technical and Sentiment Analysis (LMD)

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (LMD)

	actual_price	predicted_price
2012-10-22	46.50396347	48.75430298
2012-10-23	46.12223816	46.09690094
2012-10-24	46.41620255	45.75801086
2012-10-25	45.48604584	45.8595047
2012-10-26	46.16173172	45.57371902
2012-10-31	48.01328659	46.04946899
2012-11-01	47.52626419	49.13850784
2012-11-02	47.14455414	49.03869629
2012-11-05	47.04801559	48.26817322
2012-11-06	46.55662537	47.75989151
2012-11-07	45.85021591	47.29558945
2012-11-08	45.82827377	45.88478088
2012-11-09	45.99500656	45.58639526
2012-11-12	44.51200104	45.57919693
2012-11-13	44.65678024	45.35549927
2012-11-14	44.18732071	45.21029663
2012-11-15	44.76647568	44.92036057
2012-11-16	45.71858978	45.09085464
2012-11-19	45.3588028	45.40036392
2012-11-20	46.00377655	45.14846039
2012-11-21	47.61839676	45.39884186
2012-11-23	46.84617615	48.10956955
2012-11-26	45.8589859	45.93333054
2012-11-27	45.82387924	45.43011475
2012-11-28	45.854599	45.29700851
2012-11-29	46.46446991	45.36545181
2012-11-30	45.9774437	45.57841492
2012-12-03	46.66630554	45.41548157
2012-12-04	47.75442886	45.74802017
2012-12-05	47.79389954	48.48340607
2012-12-06	48.21511459	48.98538589
2012-12-07	48.90396881	49.19097137
2012-12-10	49.40415573	49.20902252
2012-12-11	49.15843201	49.84492493
2012-12-12	49.00926971	49.43004227
2012-12-13	49.25936508	49.38536835
2012-12-14	48.63632965	49.41707993
2012-12-17	49.206707	49.15146255
2012-12-18	49.81220245	49.41234207
2012-12-19	50.58441162	49.81988525
2012-12-20	49.6279068	50.10544205
2012-12-21	49.04875946	49.79768753
2012-12-24	49.47873688	49.36157608
2012-12-26	49.18476868	49.70004654
2012-12-27	49.206707	49.4383049

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (LMD)

2012-12-28	50.42207718	49.38056183
2012-12-31	51.88751221	49.89195633
2013-01-02	51.77344513	50.70038605
2013-01-03	51.98843765	50.68901443
2013-01-04	51.32152176	50.68237686
2013-01-07	51.10652542	50.51674652
2013-01-08	51.46631241	50.44425583
2013-01-09	52.30433273	50.61130905
2013-01-10	52.27801514	50.90743637
2013-01-11	52.24290848	51.06785583
2013-01-14	52.01476288	50.9534874
2013-01-15	51.30835724	50.8743515
2013-01-16	51.3434639	50.5501442
2013-01-17	52.07180405	50.72257233
2013-01-18	52.74748611	50.87498474
2013-01-22	52.08496094	51.141922
2013-01-23	52.4491272	50.90688324
2013-01-24	53.2388916	51.03695679
2013-01-25	52.77380371	51.49370575
2013-01-28	53.8224411	51.23284912
2013-01-29	53.15554047	51.56437302
2013-01-30	53.28716278	51.41537094
2013-01-31	53.22134399	51.48176575
2013-02-01	49.39099884	51.47568512
2013-02-04	49.94382858	52.56591034
2013-02-05	50.3825798	50.13330078
2013-02-06	48.48275757	50.13951492
2013-02-07	48.97416306	49.75386047
2013-02-08	48.56612396	49.4184227
2013-02-11	48.64948654	49.16016006
2013-02-12	48.96101379	49.20569992
2013-02-13	48.7898941	49.27598572
2013-02-14	49.11457062	49.16293716
2013-02-15	49.64108276	49.36450958
2013-02-19	49.71566772	49.82755661
2013-02-20	49.52699661	49.87297058
2013-02-21	49.46995926	49.57365799
2013-02-22	48.94783783	49.55097198
2013-02-25	49.29007339	49.35303879
2013-02-26	49.8823967	49.36600876
2013-02-27	49.82536316	49.49716949
2013-02-28	50.9705162	49.63646317
2013-03-01	50.33431625	50.12158966
2013-03-04	50.42207718	49.76730347
2013-03-05	50.68532944	49.90037918
2013-03-06	50.23778534	49.93928909

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (LMD)

2013-03-07	50.93101883	50.0591507
2013-03-08	50.72481155	50.02112198
2013-03-11	50.06229782	49.95344925
2013-03-12	49.42170715	49.77569199
2013-03-13	49.89555359	49.44625473
2013-03-14	49.71128464	49.54702759
2013-03-15	48.9083519	49.64700317
2013-03-18	49.14527893	49.10425949
2013-03-19	50.26851273	49.4149437
2013-03-20	48.98292923	49.78647995
2013-03-21	49.82097244	49.43235397
2013-03-22	50.48349762	49.42129135
2013-03-25	51.13285446	49.87853622
2013-03-26	51.41804504	50.10909653
2013-03-27	51.29957199	50.22020721
2013-03-28	51.13285446	50.36901474
2013-04-01	51.44436646	50.21031189
2013-04-02	50.57123947	50.39258194
2013-04-03	50.63267517	50.11046219
2013-04-04	49.77709198	50.00414658
2013-04-05	49.85168457	49.67279053
2013-04-08	50.1895256	49.56034851
2013-04-09	50.04474258	49.7431221
2013-04-10	50.34747696	49.86123276
2013-04-11	49.77270126	49.77150726
2013-04-12	47.53942108	49.54401398
2013-04-15	48.26337051	48.97956085
2013-04-16	46.88568115	48.71284103
2013-04-17	47.10066223	47.8310318
2013-04-18	47.79389954	46.41567993
2013-04-19	47.79389954	48.3897934
2013-04-22	47.10066223	48.35398102
2013-04-23	47.63156128	47.61992264
2013-04-24	47.73686981	48.07775497
2013-04-25	47.57452774	48.50907135
2013-04-26	48.02645111	48.4155426
2013-04-29	48.50908661	49.2070694
2013-04-30	47.69300079	48.71870804
2013-05-01	47.71054077	48.44656754
2013-05-02	47.87726974	47.90525055
2013-05-03	47.95624542	48.27436447
2013-05-06	48.54856873	48.29036713
2013-05-07	49.61915207	48.83552933
2013-05-08	48.75039673	49.35793304
2013-05-09	48.92591095	49.02880859
2013-05-10	48.4301033	48.91675568

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (LMD)

2013-05-13	48.67142868	48.76002121
2013-05-14	48.56612396	48.81357956
2013-05-15	48.16685104	48.72627258
2013-05-16	48.43888855	48.39899826
2013-05-17	48.79866791	48.73464584
2013-05-20	48.56612396	48.81044006
2013-05-21	47.86850357	48.76852798
2013-05-22	47.35075378	48.54581833
2013-05-23	46.54784012	46.83091354
2013-05-24	46.87690353	45.57880402
2013-05-28	45.80634308	45.99689484
2013-05-29	45.94673538	45.40427399
2013-05-30	44.66995621	45.34372711
2013-05-31	45.55186081	45.12900543
2013-06-03	45.24034119	45.10214996
2013-06-04	44.27506638	45.01800537
2013-06-05	44.14345169	44.82397079
2013-06-06	45.02861023	44.76325226
2013-06-07	43.18360138	44.39180374
2013-06-10	42.87532806	43.79706192
2013-06-11	42.2270813	43.47300339
2013-06-12	42.50814438	43.25920868
2013-06-13	41.74655533	43.32170105
2013-06-14	42.63960266	43.0267601
2013-06-17	43.05212784	43.34809494
2013-06-18	41.61961365	43.39489746
2013-06-19	39.59326553	43.02565002
2013-06-20	40.35485077	41.96396255
2013-06-21	38.86794662	41.64093018
2013-06-24	39.91059494	40.43122101
2013-06-25	40.58151627	41.60319138
2013-06-26	41.07110214	41.74835587
2013-06-27	41.47909546	42.35576248
2013-06-28	40.98044586	42.65589142
2013-07-01	40.16809464	42.33872604
2013-07-02	39.73198318	41.28920746
2013-07-03	40.00307465	40.74486542
2013-07-05	40.10915375	41.08382416
2013-07-08	40.45685959	40.96446609
2013-07-09	40.9459877	41.7179451
2013-07-10	42.51358414	41.63519669
2013-07-11	41.83585739	43.08613968
2013-07-12	41.90658569	42.40346146
2013-07-15	42.09516144	42.59510803
2013-07-16	42.60786819	42.52500153
2013-07-17	42.37214661	42.93085861

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (LMD)

2013-07-18	42.70804977	42.70433426
2013-07-19	42.72573853	43.00118256
2013-07-22	44.55850983	43.05675125
2013-07-23	44.41119766	44.11102676
2013-07-24	44.66460037	44.35650635
2013-07-25	44.51137543	44.12447739
2013-07-26	43.3680954	44.06699371
2013-07-29	43.47417068	43.69638062
2013-07-30	43.78060913	43.66432953
2013-07-31	44.2579689	43.81595612
2013-08-01	44.4052887	44.24109268
2013-08-02	44.02223206	44.10581207
2013-08-05	43.36220932	44.07762527
2013-08-06	42.34857941	43.7788353
2013-08-07	42.97325134	43.29495621
2013-08-08	43.06164551	43.24754715
2013-08-09	44.30511475	43.29404449
2013-08-12	44.56441879	43.92898178
2013-08-13	44.72943115	44.12854385
2013-08-14	44.44065475	44.1559639
2013-08-15	44.29333115	44.19696426
2013-08-16	44.41119766	44.07195282
2013-08-19	43.99277496	44.20748901
2013-08-20	43.00271225	43.86203766
2013-08-21	43.95152283	43.43502808
2013-08-22	44.24028015	43.69913864
2013-08-23	43.89258194	43.82080078
2013-08-26	43.45059586	44.42064285
2013-08-27	43.27969742	43.87639999
2013-08-28	42.92021561	43.50600433
2013-08-29	42.41930008	43.23168182
2013-08-30	45.0123024	42.99782944
2013-09-03	45.8538208	44.57865906
2013-09-04	46.24439621	44.28636169
2013-09-05	46.83326721	44.46660995
2013-09-06	48.26939392	44.84493256
2013-09-09	48.22733688	45.37134552
2013-09-10	47.54231644	45.45729828
2013-09-11	46.76717758	45.21790314
2013-09-12	46.86932373	45.00167847
2013-09-13	47.58439636	44.95542908
2013-09-16	47.66249466	45.24398422
2013-09-17	48.5938797	45.26396179
2013-09-18	48.605896	45.61360168
2013-09-19	47.7706604	45.6190567
2013-09-20	47.90886688	45.47000504

Predicted results using Multi-Layer LSTM with technical and sentiment analysis (LMD)

2013-09-23	47.29595566	45.38904572
2013-09-24	47.31398392	45.205616
2013-09-25	47.86079407	45.18336487
2013-09-26	47.64447784	45.30932236
2013-09-27	47.0435791	45.22882462
2013-09-30	47.27793503	45.11134338
2013-10-01	47.75263214	45.12913895
2013-10-02	47.80670929	45.17220306
2013-10-03	48.34750748	45.25597382
2013-10-04	48.05308151	45.38001633
2013-10-07	47.44016647	45.28847504
2013-10-08	47.68052292	45.23594284
2013-10-09	48.29343796	45.21519852
2013-10-10	48.73207474	45.45933914
2013-10-11	48.67199326	45.48722076
2013-10-14	48.13719559	45.55267715
2013-10-15	48.28141403	45.43061447
2013-10-16	48.34149551	45.41749191
2013-10-17	48.39558792	45.40790939
2013-10-18	48.30544662	45.51377106
2013-10-21	48.47369385	45.4522171
2013-10-22	47.09166336	45.57784271
2013-10-23	46.8152504	45.24975967
2013-10-24	46.62896729	44.93090057
2013-10-25	46.05211639	44.80547333
2013-10-28	48.22132492	44.72853851
2013-10-29	47.7105751	45.5131073
2013-10-30	48.29343796	45.29366302
2013-10-31	48.91835785	45.63166046
2013-11-01	49.61539078	45.9751358
2013-11-04	49.8196907	47.75212097
2013-11-05	49.50722122	48.60503006
2013-11-06	48.58184814	48.49137115
2013-11-07	48.69602966	46.2508316
2013-11-08	49.10463715	45.58957291
2013-11-11	49.54327011	46.22819901
2013-11-12	48.88230896	48.37440491
2013-11-13	49.27889252	45.67212296
2013-11-14	51.49016571	45.93632889
2013-11-15	53.9958725	49.40774536
2013-11-18	54.17012405	50.5168457
2013-11-19	38.58275795	50.85276031