

Evaluation of VI Index Forecasting Model by Machine Learning for Yahoo! Stock BBS using Volatility Trading Simulation

Kodai Sasaki
Nara Institute of
Science and Technology
sasaki.kodai.sb6@is.naist.jp

Hirohiko Suwa
Nara Institute of
Science and Technology
h-suwa@is.naist.jp

Yuki Ogawa
Ritsumeikan University
y-ogawa@fc.ritsumei.ac.jp

Eiichi Umehara
Tokyo City University
umehra@tcu.ac.jp

Tatsuo Yamashita
Yahoo Japan Corporation
tayamash@yahoo-crop.jp

Kota Tsubouchi
Yahoo Japan Corporation
ktsubouc@yahoo-crop.jp

Abstract

The risk avoidance is very crucial in investment and asset management. One commonly used index as a risk index is the VI index. Suwa et al.(2017) analyzed stock bulletin board messages and predicted it rise. In our study, we developed a simulation of trading Nikkei stock index options using intra-day data and verified the validity of the VI index prediction model proposed by Suwa et al. In a period from November 18, 2014, to June 29, 2016, we conducted a simulation using a long straddle strategy. The profit and loss from trading with the instructions of their model was +3,021 yen. The benchmark's profit and loss was -3,590 yen. The improvement with their model was +6,611 yen. Therefore, we confirmed that Suwa et al.'s VI index prediction model might be effective.

1. Introduction

Many investors and researchers investigated whether stock prices and stock indices were predictable. Financial instruments such as stock prices move randomly according to the efficient market hypothesis [1]. The price of financial instruments does not follow a Gaussian distribution, but rather stable distributions with infinite variance [2]. Thus, calculating the average and standard deviation is difficult and predicting the price of financial instruments is also complicated. However, various information technologies have been developed for applications in financial fields and many studies have identified the predictability of the price and movement of financial instruments.

For example, Miyazaki et al.[3] reported that Convolution Neural Networks (CNNs) predicted stock market crashes. Furthermore, with the development

of natural language technology, many studies showed that stock prices and stock indexes could be predicted through information, such as news articles and social media. News articles and press releases are important information sources in investor decisions. For analyzing news articles, Tetlock et al.[4] created pessimistic factors from Wall Street Journal columns using principal component analysis. Consequently, they reported the relationship between news articles and the Dow Jones Industrial Average. Atkins et al.[5] analyzed Reuter US News to predict both volatility and asset price movements using financial news as machine-learning input and concluded that volatility movements are more predictive than asset price movements. Feuerriegel et al.[6] analyzed the effects of topics found in such corporate press releases on stock market returns in the German market by using Latent Dirichlet Allocation. In an analysis of text information, an accuracy of the dictionary that polarizes texts is important. Loughran et al.[7] pointed out an importance of text evaluation in a finance context and created a finance dictionary.

Investor sentiments can also be expressed on social media. Sprenger et al.[8] and Bollen et al.[9] analyzed Twitter messages to predict stock market trends. In these studies, bull/bear classification or classification into six types of emotions analyzed the content of posted messages on social media.

The risk avoidance is very crucial in investment and asset management. Volatility, which is one of the risk indexes, represents a rate of a change in a product's price. A Nikkei VI index indicates an expected degree of fluctuation of the Nikkei stock index in the future and is used for risk indicators of the Nikkei stock index. Suwa et al.[10] analyzed stock BBS and predicted the VI index rises. If the VI index can be predicted, we can avoid a risk and profit from a volatility trading.

In our study, we verified the effectiveness of Suwa et al.'s model. Previous studies develop a trading simulation using past financial information to verify a effectiveness of a prediction model [11], [12], [13]. As with these previous studies, we developed a trading simulation of the Nikkei stock index option and conducted a trading simulation in accordance with Suwa et al.'s model. Our final goal is to develop a decision support system for investor risk management.

2. Previous research

2.1. Social media and decision support system

Hogenbooma et al.[14] reviewed event extraction techniques of decision support systems to be developing since the 1980s. These methods used big data, text mining and natural language processing. They gave a summarization of event extraction techniques for textual data, distinguishing between data-driven, knowledge-driven, and hybrid methods. Moreover, we discuss common decision support applications of event extraction from text corpora. As they mentioned, applying social media event extraction systems to decision support systems is one important stream of research issue. As a decision support system using social media text information, Jiang et al.[15] proposed a novel stakeholder-based event analysis framework that used online stylometric analysis to segment the forum participants by stakeholder groups to examine how important stakeholders evolved over time. Their results showed that some stakeholder groups identified by their system had stronger relationships with firms' market performance. Biswas et al.[16] examined hacker behavior in dark forums. They combined techniques from online forum features as well as text-mining and sentiment-analysis of messages. They mentioned such investigation of hacker behavior was particularly relevant for organizations in proactive prevention of cyber-attacks. As an example of a financial decision support system, Liu et al.[17] clarified an impact of social media on stock returns. As with these previous studies, our research objective is to develop a financial decision support system using messages posted on Yahoo! Stock BBS.

2.2. Research on stock market

Most of the distributions in finance theories were assumed a Gaussian distribution because calculating the average and standard deviation is easy. However, the financial instruments do not follow a Gaussian distribution, but rather stable distributions with infinite variance [2]. Calculating the average and standard

deviation is difficult. The discussion of predictability includes an efficient market hypothesis, which is a theory in financial economics that argues that asset prices completely reflect all the available information [18]. Since predicting stock prices is impossible, excess returns cannot be earned. An efficient market hypothesis consists of three hypotheses: weak, semi-strong, and strong. In weak-form efficiency, future prices cannot be predicted by analyzing past prices. Semi-strong-form efficiency implies that neither fundamental nor technical analysis techniques can reliably produce excess returns. In strong-form efficiency, stock prices reflect all of the information (public and private), and no one can earn excess returns. Not only past time series and public information but also many investor's opinions are posted on the Internet. No conclusions have been made whether stock prices and stock indices are predictable using many investor's opinions.

2.3. Social media and market index

There were many studies using investor's sentiments on social media.

Using natural language processing, Antweiler et al.[19] analyzed the number and content of messages posted to Yahoo! and Raging Bull for 45 companies, including the Dow Jones Industrial Index and the Dow Jones Internet Index. They examined the relationship between the stock market and social media and reached the following conclusions: 1. These stock bulletin boards do not predict stock returns. 2. When bullish and bearish opinions compete, trading becomes active. 3. A number of messages posted on a bulletin board predicts the volume of stock trading.

Many studies have focused on Twitter. Sprenger et al.[8] analyzed about 250,000 tweets about the stock market and found that their sentiments were related to abnormal changes in stock prices as well as changes in them the next day. They also investigated the diffusiveness of information on Twitter and found a correlation between the return on investment and a tweet's influence.

Bollen et al.[9] classified tweets into six different sentiment levels using natural language processing and discovered that "calm" feelings positively correlated with the Dow average stock price two and five days later.

Maruyama et al.[20] used natural language processing and machine learning to analyze the content of posted messages for each stock of the Yahoo! Japan stock bulletin boards and investigated the relationship between stock bulletin boards and stock prices in Japan. They classified posted messages as bullish, bearish, or neutral. They found that the number

of messages is a leading indicator of volatility and volume, and the difference between bullish and bearish messages (the bullish index) is related to returns, volatility, and the volume of individual stocks.

Recently, posted content was directly used both for forecasting a stock price and for creating new feature quantities. Xu et al.[21] predicted stock price movements from tweets and historical stock prices and presented a novel deep generative model that jointly exploited text and price signals for predicting stock price movements.

2.4. Trading strategy

Worasuchep et al.[11] conducted a stock trading simulation using the historical daily stock prices of twenty stocks from the NYSE and SET markets from 2015 to 2016 to verify a trading strategy based on their learning method that combined a set of technical trading signals that used a modified Particle Swarm Optimization. They verified the effectiveness of their learning method.

Goshima et al.[12] estimated the polarity of news articles and conducted a stock trading simulation. They got Reuters News's articles about the Japanese stock market with polarity values based on stock data and vectorized them using a bag-of-words. With these data, they used machine learning to assign polarity values to the news articles of the test data. Finally, they constructed a long-short strategy based on the polarity values given by machine learning and did a trade simulation, which showed that they got active returns by analyzing news articles by machine learning.

Sasaki et al.[13] developed a simulation program for trading Nikkei stock index options and verified a validity of a volatility index prediction model proposed by Suwa et al.[10]. They used an opening price as a trading price. An exercise price for selecting an option was determined based on the opening price of the Nikkei stock index. However, since some of the options had very low liquidity, the timestamp of their opening prices might not have been 9 a.m..

Therefore, we use intra-day data to simulate trading with a trading time that is set to the same time for reliable results.

3. Prediction model of VI index

We reproduced the prediction model of Suwa et al.[10] and conducted our trading simulation. Thus in this section, we scrutinized a model that predicts the VI index using the stock BBS topics proposed by Suwa et al. [10].

3.1. Overview

First, Suwa et al.[10] extracted Japanese messages from Yahoo! Stock BBS and decomposed them into words using morphological analysis. They used Latent Dirichlet Allocation (LDA) topic model [22] to extract topics from these words. Finally, they created a model that predicts the increase in the VI index by Logistic Regression as classifier.

3.2. VI index

They focused on the VI index, which was gaining attention as a risk index.

The VI index, which expresses the fear of investors, is closely related to the risk of a depression. It shows the expected rate one month into the future of the Nikkei stock index in the current market. Using 14 years of market analysis, Whaley [23] explained why the VI index was called the fear index and found that it was closely related to the risk of crisis in the stock market.

The VI index is defined as follows:

$$VIX_T = \sqrt{E^* \left[\frac{1}{T} \right] \int_T^0 v^2 dt}. \quad (1)$$

A histogram of the VI index from November 19, 2014 to July 4, 2016 is shown in Fig. 1.

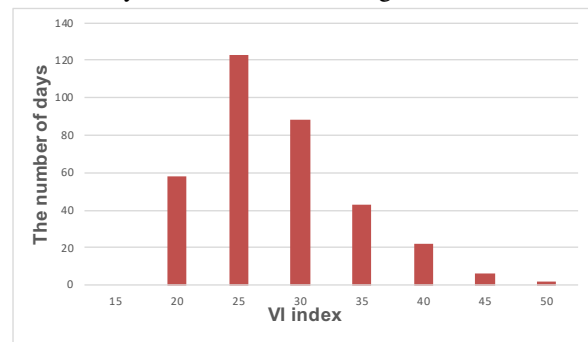


Figure 1. Histogram of VI index

3.3. Topic classification

Suwa et al.[10] assumed that market trends changed if topics that people discussed changed. Therefore they used LDA to measure what topics people discussed.

To determine the topics of each message, They used the LDA topic model of Blei et al.[22]. This probabilistic model assumes K potential topics Z_i that occur in document d :

$$\sum_{i=1}^K p(Z_i)p(d|Z_i). \quad (2)$$

With LDA, $p(w|Zi)$ is obtained for word w . In other words, by assuming that topic Zi is one cluster, words can be clustered. Therefore, topic generation probability can be obtained.

They used morphological analysis (MeCab) and classified messages into 100 topics using LDA. Next, they calculated 100 topics generation probability of each message. They calculated standard deviations of topic generation probability. They defined the threshold as the 2 standard deviation. If this generation probability was less than the threshold, they set the generation probability of message was 0. Finally, they summed up the generation probabilities for every 100 topics on a daily base and divided by the number of messages per day. They defined it as a posting rate of each topic. As a result, they calculated 100 features. They called them as posting rates of each topic. A posting frequency of each topic was defined as multiplied a posting rate of each topic by the total number of messages per day.

3.4. Creation of model that predicts increase in VI index using machine learning

To predict increases in the VI index, they created a model using machine learning. They used the value of the posting rate and the posting frequency of each topic introduced in Section 3.3 as feature quantities. They also used 12 metrics for topic as follows:

1. Posting frequency
2. Posting rate
3. Difference between yesterday and today's posting frequency
4. Difference between yesterday and today's posting rate
5. Ratio of yesterday and today's posting frequency
6. Ratio of yesterday and today's posting rate
7. Average posting frequency of 7 days
8. Average posting rate of 7 days
9. Difference between average posting frequency of previous 7 days and today's posting frequency
10. Difference between average posting rate of previous 7 days and today's posting rate
11. Ratio of average posting frequency of previous 7 days and today's posting frequency
12. Average posting rate of previous 7 days and today's posting rate

Therefore, they used 1200 feature quantities of the topic. They also used the VI index and the total number of messages as feature quantities. These were 12 metrics of stock market as follows:

1. VI index
2. Total number of messages
3. Difference between yesterday and today's VI index
4. Difference between yesterday and today's total number of messages
5. Ratio of yesterday and today's VI index
6. Ratio of yesterday and today's total number of messages
7. VI index average over 7 days
8. Average of total number of messages over 7 days
9. Difference between average VI index over previous 7 days and today's VI index
10. Difference between average of total number of messages over previous 7 days and today's total number of messages
11. Ratio of VI index average of previous 7 days and today's VI index
12. Ratio of average of total number of messages over previous 7 days and today's total number of messages

Therefore, the total number of their feature quantities was 1212.

Regarding the increase in the VI index, Suwa et al. [10] defined the days when it rose more than 2 as increasing. They defined the days with "increasing" over seven subsequent days as "increasing days". They used it as objective variables for Machine learning. They predicted "increasing days" using Logistic Regression. The precision of "increasing day" was 0.45 and the recall of "increasing day" was 0.45.

Since they do not study the effectiveness of actual transactions, we conduct volatility trading using the Nikkei stock index option. Suwa et al.'s output is whether "increasing days" or not for each day. We define it as the trading instructions for trading simulations.

4. Simulation of volatility trading using Nikkei stock index option

4.1. Purpose

Suwa et al.[10] predicted increases in the VI index. We call it their prediction model. If their prediction model is correct, trading should be profitable. We do trading simulations by following their trading instructions based on their prediction results and verify their effectiveness. In the following, we explain the market mechanism handled in our trading simulation.

4.2. Nikkei stock index option

Our simulations use 225 Nikkei stock index options listed on the Osaka Stock Exchange. An option of Nikkei stock index is called a "right to buy" of the "right to sell" at a predetermined price (exercise price) at a predetermined date (exercise date) [24]. The right to buy is a call option, and the right to sell is a put option. The options are issued by the maturity date of each month and the exercise price (bid or offer prices). On the Osaka Stock Exchange, the maturity date is the second Friday of every month, and the exercise price is listed in increments of 125 yen. Options trading can be bought and sold with only commissions and margins. Therefore, the initial investment amount can be quite small. Options are classified into three types based on the relationship between the Nikkei stock index price and the exercise price (Table 1).

Table 1. Option classification

	CALL	PUT
Nikkei > Exercise	InTheMoney	OutofTheMoney
Nikkei = Exercise	AtTheMoney	AtTheMoney
Nikkei < Exercise	OutofTheMoney	InTheMoney

An option whose exercise price equals the Nikkei stock index price is described as at-the-money. In call option, a option whose exercise price lower than the Nikkei stock index price is described as in-the-money. An option whose exercise price higher than the Nikkei stock index price is described as out-of-the-money. On the contrary, in put option, a option whose exercise price lower than the Nikkei stock index price is described as out-of-the-money. An option whose exercise price higher than the Nikkei stock index price is described as in-the-money. At the exchange, the volume of options trade near at-the-money is high, but the volume of other options trade is low.

The option theoretical price is calculated using the Black-Scholes differential equation. In it, a volatility is one of the parameters. Therefore, predicting the

volatility of the Nikkei stock index is critical for options trading. One indicator that predicts this volatility is a VIX. We define a VIX of Nikkei stock index as the VI index. If we can predict the VI index, it can be used for the investment strategy of a volatility trading using options. Accordingly, we verify the validity of their prediction model.

4.3. Volatility trading

Volatility trading is feasible using options. In volatility trading, investors predict the volatility of the Nikkei Stock index and decide on a trading strategy that adapts to changes in the market. As a volatility strategy using options, straddle strategies combine two options or butterfly strategies that combine three or four options.

4.4. Price of call option

If an investor buys a call option, he can limit the maximum loss to a premium and earn a profit if the stock price increases. Fig. 2 shows the profit and loss (P/L) diagram for a maturity date of a call option.

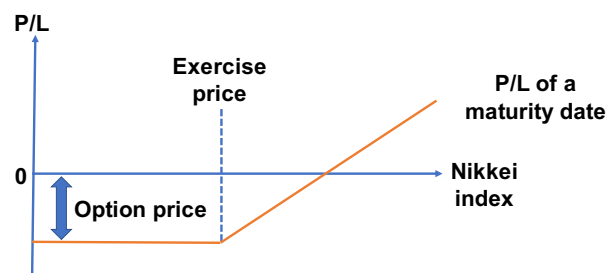


Figure 2. Profit and loss diagram of call option

4.5. Price of put option

If an investor buys a put option, he can limit the maximum loss to a premium and get a profit if the stock price decreases. Fig. 3 shows the P/L diagram for a maturity date of a put option.

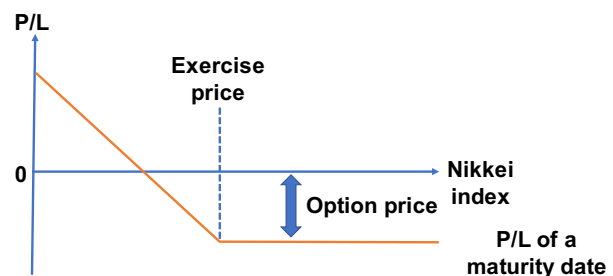


Figure 3. Profit and loss diagram of put option

4.6. Long straddle strategy

Since their prediction model predicts that the VI index will move significantly, we use a long straddle strategy. Fig. 4 shows the P/L diagram for a maturity date of an option. A loss occurs if the market moves slightly but a profit occurs if the market moves significantly.

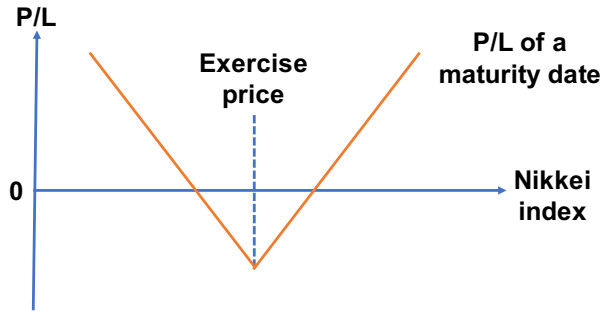


Figure 4. Profit and loss diagram of long straddle

4.7. Intra-day data

We conduct simulations using intra-day data, which are also called tick data. They are a record of intra-day trade history in a time series. With intra-day data, put and call options can be simultaneously bought and sold for accurate trading simulations. We perform simulations using the trading day, the brand code, the time, the trading price, and the price type (open, medium, or close) from the intra-day data. We get the intra-day data from the JPX Data Cloud[25].

5. Trading simulation

The simulation period is from November 18, 2014 (the day after Suwa et al.'s learning period ended) to June 29, 2016. Fig. 5 shows the Nikkei stock index and the VI index during the simulation period. The dotted line is the transition of the Nikkei stock index. The solid line is the VI index transition. The vertical line on the left shows the Nikkei stock index price, the vertical line on the right shows the VI index, and the horizontal line shows the days.

In our simulation, we set new position using opening price on the next day after the instructions of their prediction model. We settle this position up to six trading days from the day after the position was set because to consistent whit their prediction model and our simulation. when set new position and settle this position we need trade tow options(a call and a put) The difference of the trading time of tow options is within ten minutes.

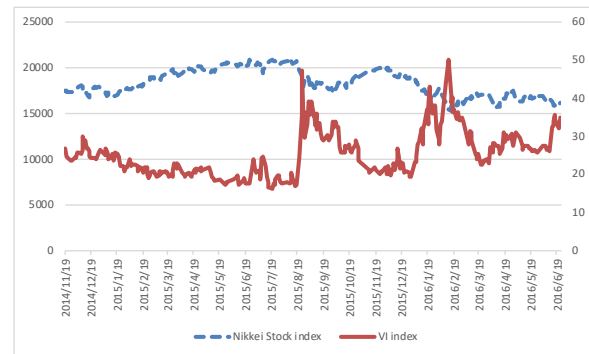


Figure 5. Nikkei stock index and VI index

5.1. Simulation method

We conduct the simulation as follows:

1. Based on the opening price of the Nikkei stock index on the day after the trading instructions are given, we obtain the Call and Put options from the brand code of intra-day data.
2. We obtain the times and the trading prices where the price type is "open", based on the brand code of the Call and Put options and the trading day.
3. If the time, at which each option's opening price is attached, is within ten minutes of the difference, we buy each option one by one and set a position.
4. If the time difference exceeds ten minutes, we search for the middle price of the other brand within ten minutes based on the late time with the opening price; if it is found, we set a position.
5. If the Nikkei Stock index is more than 500 yen away from the day when the price of the day the position was created or six trading days have passed since the position was created, we settle the position.

5.2. Exception handling

If there is no price when setting a new position, no trading is done. When we settle a position, if there is no price for a call option but a put option exists, we use a synthetic call option with a put option and a Nikkei stock index future. Similarly, if there is no price for a put option, but a call option exists, we use the synthetic put option with a call option and a Nikkei stock index future.

Furthermore, there were no days without a put option and a call option.

6. Trading simulation results

We summarize our simulation results, the total number of trading days, and the total P/L in Table 2.

Table 2. Trading simulation results

	Model	Benchmark
Trading days	81	335
Hit days	57	180
Miss days	24	155
Hit rate	70%	54%
Final P/L	3,021	-3,590
Final profit	6,526	19,955
Final loss	-3,505	-23,545

We call the day when the profit occurred as "hit day" and the day when the loss occurred as "miss day".

The total number of trading days during the simulation period was 394. The total number of trade instructions in accordance with Suwa et al.'s prediction model was 92. However, the number of trading days of the benchmark case was 335, which was less than 394. Likewise, in their prediction model, the number of trading days was 81, which was less than 92, because there were trading days when no option tradings were executed at the exchange. The total P/L of the instructions to trade in accordance with their prediction model was +3,021 yen. The total P/L of the benchmark case was -3,590 yen.

The details about simulation contents are omitted. Simulation contents in specific cases are shown in Table3.

Table 3. Simulation contents in specific cases

	Setting				Settlement				
	Date	Nikkei index	Call	Put	Date	Nikkei index	Call	Put	P/L
case1	2014/12/17	16689	345	450	2014/12/19	17511	920	110	235
case2	2015/5/25	20332	270	290	2015/6/2	20605	360	125	-75
case3	2016/2/25	15983	405	445	2016/3/3	16696	815	90	55
case4	2015/3/16	19245	335	440	2015/3/24	19719	535	175	-65

Case 1 shows a case where the VI index rose 2 points and the Nikkei index moved more than 500 yen. While Case 2 shows one where the VI index rose 2 points and the Nikkei index did not move more than 500 yen. Case 3 shows a case where the VI index did not change and the Nikkei index moved more than 500 yen. While Case 4 shows one where the VI index did not change and the Nikkei index did not move more than 500 yen.

7. Discussion

7.1. Evaluation of prediction model

The number of trading days in accordance with Suwa et al.'s prediction model was 81 and the total P/L was

+3,021 yen. The average P/L per trade was 37.2 yen. The number of trading days of the benchmark case was 335, and the total P/L was -3,590 yen. The average P/L per trade was -10.7 yen. The improvement with their prediction model was +6,611 yen. The number of days that the Nikkei stock index moved more than 500 yen was 180. On the other hand, the Nikkei stock index stayed within 500 yen on 155 days. In 155 days, their prediction model gave trading instructions on 24 days. It was 15%(24/155). Therefore, we can avoid 85% of the possibility of loss.

In 54%(180/335) of the days, the Nikkei stock index moved more than 500 yen during the simulation period. In 70%(57/81) of the days that their prediction model gave the trading instructions, the Nikkei stock index moved more than 500 yen. The improvement was 16 points. We confirmed Suwa et al.'s VI index prediction model might be effective and earn a profit.

7.2. Discussion of objective variables

Our trading instructions were based on the predictions of Suwa et al.'s model. They defined that a day when the VI index increased more than 2 over seven subsequent days was "increasing days". They used "increasing days" as objective variables of machine learning. We verified the validity of this definition in the simulation period. The number of days labeled their objective variables and the P/L are shown in Table 4. We called the number of days labeled "increasing days" as "correct answer days". The correct answer days were 136. The profit days were 79 days and the loss days were 57 days in the correct answer days. If their prediction model could perfectly predict for 136 days, the total P/L was +8,645 yen. However, losses occurred in the correct answers days. Thus, we wanted to avoid the 57 loss days was "correct answer days".

Table 4. Objective variable of correct answers

	Days	P/L
Correct answers	136	8,645
Profit	79	17,098
Loss	57	-8,453

Focusing on the 81 days of instructions of the Suwa et al.'s prediction model, we discussed the forecast model's accuracy. Table 5 shows this result. The objective variables were applied on 39 days. They were not applied on 42 days. The total P/L was positive for the incorrect answers, but the P/L was also negative. Thus, the accuracy of Suwa et al.'s VI index prediction model might be improved.

Table 5. Results with instructions

	Days	P/L
Instructions	81	3,021
Correct answers	39	2,796
Incorrect answers	42	255

We discussed the relationship with the objective variables focusing on the 254 days of trades that did not give trading instructions. Table 6 shows this result. Their prediction model avoided losses 157 times. Furthermore, the amount of loss that could be avoided by the instructions of their prediction model was 12,459 yen. From this result, their prediction model might achieve the goal of risk avoidance. Although we applied them to the definition of the objective variables, no trading instructions were given on 97 days and their P/Ls were +5,849 yen. Therefore, the extra verification will be necessary.

Table 6. Results without instructions

	Days	P/L
No instructions	254	-6,610
Correct answers	97	5,849
Incorrect answers	157	-12,459

To improve a machine learning accuracy, it might need to improve objective variables of machine learning. This point is future work.

8. Conclusion and future research

We developed a simulation for trading Nikkei stock index options using intra-day data and verified the validity of the VI index prediction model proposed by Suwa et al. [10]. With intra-day data, our simulations were closer to actual deals. Our period was from November 18, 2014 to June 29, 2016. We conducted two simulations using a long straddle strategy. One was a benchmark case to trade on every day during the period, and the other was in accordance with the instructions of Suwa et al.'s VI index prediction model. The P/L of trading with instructions was +3,021 yen. The P/L of the benchmark was -3,590 yen. The improvement using their model was +6,611 yen. We confirmed that Suwa et al.'s VI index prediction model might be effective. In one case, the model did not give instructions although trading made a profit every day. In another case, the trading suffered a loss, although the prediction model gave instructions. Thus, the accuracy of Suwa et al.'s VI index prediction model might be improved.

In the future, we improve the VI index prediction model and analyze which feature quantities of topics

is important. Based on the results of the trading simulation in our study, the VI index prediction model might be applied to a risk management. Therefore, one of our future research goals is to develop a decision support system of a risk management using the VI index prediction model.

References

- [1] S. Basu, "Investment performance of common stocks in relation to their price-earnings ratios : A test of the efficient market hypothesis," *The Journal of Finance*, vol. 32, no. 3, 1977.
- [2] B. Mandelbrot, "The variation of certain speculative prices," *The Journal of Business*, vol. 36, no. 4, pp. 394–419, 1963.
- [3] K. Miyazaki and Y. Matsuo, "Stock prediction analysis using deep learning technique," in *Proceedings of The 31st Annual Conference of Japanese Society for Artificial Intelligence*, 2017.
- [4] 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.
- [5] A. Atkins, M. Niranjan, and E. Gerding, "Financial news predicts stock market volatility better than close price," *The Journal of Finance and Data Science*, vol. 4, pp. 120–137, 2018.
- [6] S. Feuerriegel, A. Ratku, and D. Neurnann, "Analysis of how underlying topics in financial news affect stock price using latent dirichlet allocation," in *Proceedings of the 49th Hawaii International Conference on System Sciences*, 2016.
- [7] T. Loughran and B. McDonald, "When is a liability not a liability ? textual analysis, dictionaries, and 10-ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [8] T. O. Sprenger, A. Tumasjan, P. G. Sandner, and I. M. Welpe, "Tweets and trades: The information content of stock microblogs," *European Financial Management*, vol. 20, no. 5, pp. 926–957, 2014.
- [9] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, pp. 1–8, 2011.
- [10] H. Suwa, Y. Ogawa, E. Umehara, K. Kakigi, T. Yamashita, and K. Tsubouchi, "Develop method to predict the increase in the nikkei vi index," in *Proceedings of The 2nd International Workshop on Application of BigData for Computational Social Science in IEEE Bigdata 2017*, 2017.
- [11] C. Worasuchee, S. Nuannimnoi, R. Khamvichit, and P. Attagonwantana, "An automatic stock trading system using particle swarm optimization," in *Proceedings of The 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunication and Infomatin Technology*, 2017.
- [12] K. Goshima, T. Takahashi, and T. Terano, "Estimating news articles' negative-positive by deep learning," in *Proceedings of The 29th Annual Conference of Japanese Society for Artificial Intelligence*, 2016.
- [13] K. Sasaki, Y. Hirose, E. Umehara, H. Suwa, Y. Ogawa, T. Yamashita, and K. Tsubouchi, "Simulation of volatility trading using nikkei stock index option based on stock bulletin board," in *Proceedings of The Third*

International Workshop on Application of BigData for Computational Social Science in IEEE Bigdata 2018, 2018.

- [14] F. Hogenboom, F. Frasinca, U. Kaymak, F. De Jong, and E. Caron, "A survey of event extraction methods from text for decision support systems," *Decision Support Systems*, vol. 85, pp. 12–22, 2016.
- [15] S. Jiang, H. Chen, J. F. Nunamaker, and D. Zimbra, "Analyzing firm-specific social media and market: A stakeholder-based event analysis framework," *Decision Support Systems*, no. 67, pp. 30–39, 2014.
- [16] B. Biswas, A. Mukhopadhyay, and G. Gupta, "' leadership in action: How top hackers behave" a big-data approach with text-mining and sentiment analysis," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
- [17] L. Liu, J. Wu, P. Li, and Q. Li, "A social-media-based approach to predicting stock comovement," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3893–3901, 2015.
- [18] E. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
- [19] W. Antweiler and M. Z. Frank, "Is all that talk just noise? the information content of internet stock message boards," *The Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.
- [20] K. Maruyama, H. Suwa, E. Umehara, and T. Ohta, "Is stock bbs content correlated with the stock market?: A japanese case," in *2009 IEEE International Conference on Systems, Man and Cybernetics*, pp. 4601–4606, 2009.
- [21] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018.
- [22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [23] R. E. Whaley, "The investor fear gauge," *The Journal of Portfolio Management*, vol. 26, no. 3, pp. 12–17, 2000.
- [24] "Nikkei255 options." <https://www.jpx.co.jp/derivatives/products/domestic/225options/index.html>. Tokyo Stock Exchange, Inc.(2019/4/28).
- [25] "Jpx data cloud." <http://db-ec.jpx.co.jp/item/C430509.html>. Tokyo Stock Exchange, Inc.(2019/4/28).