



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

**Development of Machine Learning Methods to Improve
Decisions in Bulk Chartering Practice**

by

Sangseop Lim

Supervised by

Prof. Kihwan Lee

A dissertation submitted for the degree of

Doctor of Philosophy

Department of Shipping Management

Graduate School of Korea Maritime and Ocean University

February 2018

Approval Sheet

This dissertation, which is an original work undertaken by Sangseop Lim in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Business Administration, is in accordance with the regulations governing the preparation and presentation of dissertations at the Graduate School in Korea Maritime and Ocean University, Republic of Korea.

Approved by the Dissertation Committee:

Prof. Seong-Cheol Cho

Chairman

Prof. Jai-Min Lee

Member

Prof. Dong-Keun Ryoo

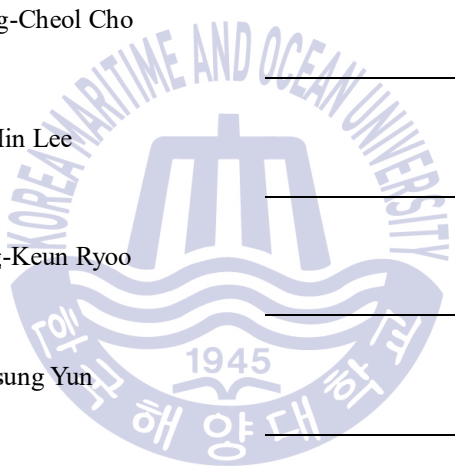
Member

Ph.D. Heesung Yun

Member

Prof. Ki-Hwan Lee

Member



Department of Shipping Management
Graduate School of Korea Maritime and Ocean University

December 2017

Contents

Contents	iii
List of Tables	v
List of Figures	vii
요 약	ix
Abstract	xi
Chapter 1 Introduction	1
1.1 Purposes and Contributions	3
1.2 Structure	4
Chapter 2 Chartering Practice	5
2.1 T/C Option to Extend.....	7
2.2 Chartering-out Strategy	10
Chapter 3 Literature Review	13
3.1 Regression-related Problem with Machine Learning	13
3.2 Classification-related Problem with Machine Learning	28
3.3 Maritime-related Problem with Machine Learning	31
Chapter 4 Methodology	37
4.1 Option Pricing.....	37
4.1.1 Benchmark: Black-Scholes-Merton model(BSM).....	37
4.1.2 Machine Learning Methods.....	38
4.1.2.1 Artificial Neural Networks(ANN)	39
4.1.2.2 Support Vector Regression(SVR).....	42

4.1.2.3	Random Forest(RF).....	48
4.2	Trading Strategy.....	51
4.2.1	Benchmark: Multinomial Logistic Regression(MLR).....	53
4.2.2	Artificial Neural Networks(ANN).....	56
4.2.3	Support Vector Machines(SVM).....	57
4.2.4	Random Forest(RF).....	63
4.3	Performance Measures	63
Chapter 5	Research Results	66
5.1	Valuation of T/C Extension Pption	66
5.1.1	Data and Frameworks.....	66
5.1.2	BSM Modeling	70
5.1.3	ANN Modeling.....	73
5.1.4	SVR Modeling.....	77
5.1.5	RF Modeling.....	79
5.1.6	Results and Discussion.....	81
5.2	Decision for Chartering-out Strategies.....	86
5.2.1	Data and Frameworks.....	86
5.2.2	MLR Modeling.....	91
5.2.3	ANN Modeling.....	91
5.2.4	SVM Modeling.....	93
5.2.5	RF Modeling.....	93
5.2.6	Results and Discussion.....	93
Chapter 6	Conclusion.....	98
	References.....	100

List of Tables

TABLE 1 THE DETAILS OF PANAMAX FIXTURES IN 2016	11
TABLE 2 APPLICATION OF ANN IN FINANCIAL MARKET	18
TABLE 3 APPLICATION OF SVM AND RF IN FINANCIAL MARKET	20
TABLE 4 APPLICATION OF MACHINE LEARNING METHODS IN PRICING OPTIONS	23
TABLE 5 APPLICATION OF MACHINE LEARNING METHODS IN OTHER AREA ...	26
TABLE 6 APPLICATION OF MACHINE LEARNING METHODS IN CLASSIFICATION-RELATED PROBLEMS	30
TABLE 7 APPLICATION OF MACHINE LEARNING METHODS IN MARITIME SECTOR	35
TABLE 8 TYPE OF KERNEL FUNCTION	47
TABLE 9 ALGORITHM OF THE BINARY RECURSIVE PARTITIONING TREES (CUTLER, CUTLER AND STEVENS, 2012)	48
TABLE 10 ALGORITHM OF RANDOM FORESTS (CUTLER, CUTLER AND STEVENS, 2012)	49
TABLE 11 ALGORITHM OF OUT-OF-BAG PREDICTIONS	50
TABLE 12 TYPE OF INDICATORS FROM TECHNICAL AND FUNDAMENTAL ANALYSIS	52
TABLE 13 TYPE OF PERFORMANCE MEASURES	64
TABLE 14 DESCRIPTIONS OF DATA	67
TABLE 15 DESCRIPTIVE ANALYSIS OF FREIGHT RATES AND THE US 3 MONTH T- BILL	68
TABLE 16 INPUT VARIABLES OF BSM	71
TABLE 17 OPTIMAL PARAMETERS OF ANN	75
TABLE 18 OPTIMAL PARAMETERS OF SVR	77
TABLE 19 OPTIMAL PARAMETERS OF RF	79
TABLE 20 SUMMARY OF OPTION PRICING(LEFT:CAPESIZE, RIGHT:PANAMAX)	

.....	81
TABLE 21 SUMMARY OF OPTION PRICING(LEFT:SUPRAMAX, RIGHT:HANDYMAX).....	81
TABLE 22 DESCRIPTIONS OF DATA.....	87
TABLE 23 ADDITIONAL VARIABLES FROM TECHNICAL AND FUNDAMENTAL ANALYSIS.....	88
TABLE 24 OPTIMAL PARAMETERS OF ANN.....	91
TABLE 25 OPTIMAL PARAMETERS OF SVM.....	93
TABLE 26 OPTIMAL PARAMETERS OF RF.....	93
TABLE 27 TEST RESULTS OF CAPESIZE.....	96
TABLE 28 TEST RESULTS OF PANAMAX.....	96
TABLE 29 TEST RESULTS OF SUPRAMAX.....	96
TABLE 30 SUMMARY OF CHARTERING-OUT DECISIONS.....	97



List of Figures

FIGURE 1 FREIGHT TRADING MIX	6
FIGURE 2 NO. OF PANAMAX FIXTURES.....	7
FIGURE 3 STRUCTURE OF T/C OPTION	7
FIGURE 4 SPECIFIED STRUCTURE OF T/C OPTION	9
FIGURE 5 COMPARISON OF CHARTERING-OUT VESSEL	10
FIGURE 6 STRUCTURE OF ANN.....	39
FIGURE 7 ϵ -INSENSITIVE SVR	43
FIGURE 8 ϵ -INSENSITIVE LOSS FUNCTION	44
FIGURE 9 HARD-M ARGINED SVM	57
FIGURE 10 SOFT-MARGINED SVM.....	59
FIGURE 11 NON-LINEAR DATA	62
FIGURE 12 TRANSFORMATION TO HIGHER DIMENSIONAL SPACE	62
FIGURE 13 FLOW CHART OF PRICING OPTIONS.....	66
FIGURE 14 BULK FREIGHT RATE.....	69
FIGURE 15 US 3-MONTH T-BILL RATE.....	70
FIGURE 16 VOLATILITY OF RETURN FOR THE UNDERLYING ASSET(TCT).....	72
FIGURE 17 ORIGINAL AND RANDOMIZED SERIES OF FREIGHT RATES.....	74
FIGURE 18 OPTIMAL PARAMETERS OF ANN.....	76
FIGURE 19 OPTIMAL PARAMETERS OF SVR.....	78
FIGURE 20 OPTIMAL M OF RF	80
FIGURE 21 COMPARISON OF T/C OPTION PRICES: CAPESIZE	82
FIGURE 22 COMPARISON OF T/C OPTION PRICES: PANAMAX.....	83
FIGURE 23 COMPARISON OF T/C OPTION PRICES: SUPRAMAX	84
FIGURE 24 COMPARISON OF T/C OPTION PRICES: HANDYMAX	85
FIGURE 25 FLOW CHART OF CLASSIFIED CHARTERING-OUT DECISION	86

FIGURE 26 NO. OF DESIRED CHARTERING-OUT DECISIONS.....	89
FIGURE 27 ORIGINAL AND RANDOMIZED SERIES OF FREIGHT RATES.....	90
FIGURE 28 OPTIMAL PARAMETERS OF ANN.....	92
FIGURE 29 OPTIMAL PARAMETERS OF SVM.....	94
FIGURE 30 OPTIMAL M OF RF	95



기계학습을 활용한 벌크선 용선 의사결정 지원에 관한 연구

임 상 섭

한국해양대학교 대학원, 해운경영학과

요 약

본 연구는 벌크선 용선에 수반되는 기간연장옵션의 가격 결정 문제와 일정기간 확보된 선박을 이용하여 다양한 대선 전략을 선택하는 문제를 다뤘다.

기간연장옵션은 관행상 계약당사자 사이에서 정확한 가치평가없이 사용되며 심지어는 신용이 좋은 용선주를 유인하기 위해 옵션프리미엄 없이 무상으로 주어진다. 기존 연구에서 제시된 인공신경망모델과 더불어 새로운 2 가지 기계학습방법을 제시하였으며 이를 이용하여 기간연장옵션의 가치를 실험적으로 평가하였다. 분석결과 기계학습방법이 기존 금융시장에서 사용되는 블랙숄즈모델과 인공신경망 모델보다 R^2 을 기준으로 98%에 육박한 아주 뛰어난 성능을 보였다.

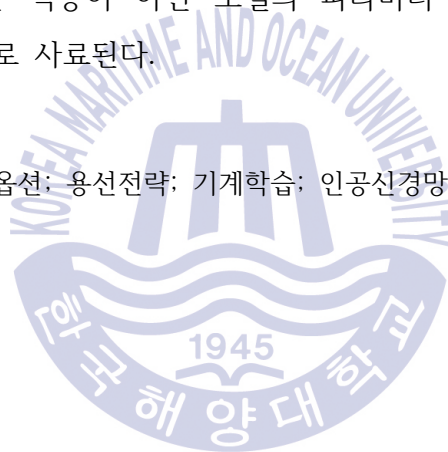
대선 전략을 선택하는 문제에 있어서는 위의 결과와 비교해볼 때 차이가 있었다. 기계학습모델들이 다항로지스틱회귀모델보다 성능이 뛰어났지만 제시된 모델 중 인공신경망 모델의 성능이 다른 기계학습모델보다 뛰어난 결과를 보였다.

결론적으로 학습시간, 모델의 복잡성, 그리고 해석의 용이성을 고려하면 용선의사결정에 가장 적합한 모델은 랜덤포레스트임을 알 수 있었다.

본 연구는 기계학습모델들을 이용하여 해운의사결정문제 적용가능성을 다뤘다는 측면에서 학문적 의의가 있고 새롭게 제시된 기계학습모델들의 가치평가 능력과 분류 능력을 고려하면 해운 실무에 시사하는 바가 크다고 할 수 있다.

연구의 한계로서 기계학습모델의 성능 차이는 연구자의 설계 능력과 배경지식에 더불어 사용하려는 데이터의 종류에 따라 결과에 많은 영향을 미치기 때문에 단순 적용이 아닌 모델의 파라미터 조정에 보다 정교한 연구가 필요할 것으로 사료된다.

핵심어: 용선기간연장옵션; 용선전략; 기계학습; 인공지능망; 서포트벡터머신; 랜덤 포레스트



Abstract

This study deals with the valuation of the T/C option corresponding to the Bulk chartering contract and the choice of the periods of chartering-out the vessel secured for a certain period.

The T/C option is granted without a precise valuation between the parties in practice and even free of charge with no option premium to attract a creditworthy charterer. In addition to the ANN presented in the previous research, two new machine learning methods are proposed and the value of the T/C option is evaluated empirically. As a result of the empirical analysis, it is shown that the machine learning methods are superior to the BSM and the ANN model used in the existing financial market, which is close to 98% based on R^2 .

In the problem of selecting the charter-out method, the results are slightly different from the above problem. Although the machine learning-based models perform better than the multinomial logistic regression models, the performance of the ANN among the proposed models is outstanding. In conclusion, considering the learning time, the model complexity, and the simple interpretability, it is found that the most promising model for decision-making in chartering practice is the random forest.

This study is of academic significance as it deals with the applicability of machine learning models in chartering problems. Furthermore, considering the evaluation ability and classification ability of the newly proposed machine learning models, it can be said that it is very important to the shipping industry.

As a limitation of the study, since it is considered that the difference between the performance of the machine learning models tends to depend on the type of data to be used as well as the design ability and background knowledge of the researcher, more elaborate research is necessary to carefully adjust the parameters of the models rather than the simple application of the models.

Keywords: Time charter option, Chartering-out strategy, Machine learning, Artificial Neural Networks, Support Vector Machines, Random Forest



Chapter 1 Introduction

We are now living in the era of the Fourth Industrial Revolution, which was conceptualized by Schwab at the World Economic Forum in 2016¹. At the heart of the Fourth Industrial Revolution is a technological integration based on artificial intelligence learning. Although it is needed to keep track of whether these revolutionary changes have the massive impact on the industries and societies, it is evident that the innovations through the inexorable integration of technology will significantly help decision-making.

In Schwab's presentation, he mentioned that compared with other revolutions, the Fourth is exponentially evolving. Since the applicability of Artificial Intelligence (AI) is inexhaustible and versatile, it applies to the tangible and visible technologies like a robot or autonomous car, as well as the intangible and invisible ones such as management, business, marketing, and so on. As a result, efforts in overall industries to introduce or develop these innovative technologies are underway to stay competitive in the market.

Nilsson (2010) defined AI as an activity to make machines intelligent, which are capable of functioning appropriately with the proper forecasts. Therefore, AI is not a machine but intelligence. However, in modern society, AI has been used confusingly with machine learning (ML). Bernard Marr demonstrated that ML is a machine

¹ <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond>

authorized for access to data and capable of self-learning². That is, ML is an application of AI.

In the shipping market, the Fourth Industrial Revolution is also a hot issue. Perhaps the world's first autonomous ship will commercialize in Norway, 2018³. Even the present papers only covered the operational efficiency of port and ship (Ahmed and Hasegawa, 2013; Bal Beşikçi *et al.*, 2016; Kourounioti, Polydoropoulou and Tsiklidis, 2016; De León *et al.*, 2017; Lazakis, Raptodimos and Varelas, 2017; Pagoropoulos, Møller and McAloone, 2017).

Unfortunately, these approaches are concerning the operation techniques only. Although these hardware innovations, of course, are also necessary and inevitable, it is not enough for supporting the decision-making in the shipping business. The more essential and fundamental points are relating to what kind of decisions in shipping can be supported by these tools, and the more improved determination that might be made through them.

However, the studies using ML for other shipping related problems have been carried out once in a blue moon, which mostly focused on forecasting Baltic Dry Index (BDI) or maritime traffic as mentioned in the literature.

In this context, this paper aims at applying the extent of AI technologies to the decision-making in shipping practice, especially the chartering-related tasks.

² <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#25f8c1212742>

³ <http://fortune.com/2017/07/22/first-autonomous-ship-yara-birkeland/>

1.1 Purposes and Contributions

The primary objective of this paper is to investigate the applicability of ML methods for the chartering practice and provide the improved framework for the decision-making.

For achieving these purposes, the problems met in chartering-desk was defined as follows,

- Evaluating the option to extend the period in time-charter contracts (T/C options)
- Making the decision to charter-out the vessel secured

Then, the state of the art ML methods will be applied to that questions. These approaches are as below

- Artificial Neural Networks (ANN)
- Support Vector Machines (SVM)
- Random Forest (RF).

This thesis can be differentiated from the other research for the following reasons. First, the paper shows that the ML tools can be used for chartering business. Notably, the SVM and RF can apply to pricing the T/C options and making chartering-decision. Second, with comparing the performance of the models, the paper will provide the suitable one among ML techniques. Above all, the article deals with the chartering and presents the new approach which is needed.

1.2 Structure

The remainder of the paper is organized as follows. The details of chartering practice are demonstrated in Chapter 2. The earlier research is scrutinized in Chapter 3. Chapter 4 introduces the methodologies of ML. In Chapter 5, it confirms the empirical results from each experiment. Finally, Chapter 6 summarizes the results and delivers the implicit interpretations of them.



Chapter 2 Chartering Practice

Chartering decisions in the freight market is fraught with uncertainties originated from a highly volatile condition that exists in the shipping market. As reported by Stopford (2009), one of the most active markets is the freight market, where the participates trade carrying capacities depending on the term of usage in return for freight rates. This market naturally exposes to the freight volatility.

The main decisions in shipping are divided into timing to buy or sell vessels and chartering in the spot or time charter contract (Karakitsos and Varnavides, 2014). Many pieces of research have been concerned with the timing of the investment for buying or selling carrying capacities as well as a ship itself. Most of them tried to find the right time to be a winner who beats the market (Alizadeh and Nomikos, 2006, 2007; Gouelmos and Gouelmos, 2009; Alizadeh and Talley, 2011; Chisté and van Vuuren, 2014).

A ship-owner or charterer having a vessel tries to sell its capacities with the specified period. Figure [1] represents the trading mix of carrying capabilities (Yun, Lim and Lee, 2017).

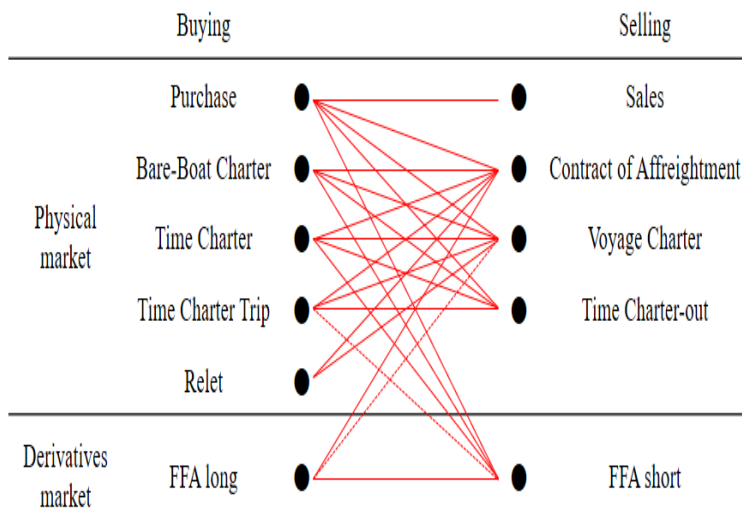


Figure 1 Freight Trading Mix

As shown, the structure of trading mix may seem too complicated. Briefly speaking, the one who intends to acquire the usage of the ship selects an instrument listed on the left-hand side of the figure. Then, he/she sells its capacities to the counterparty on the right-hand side of the picture who demands a specified period. This forms of trading is called 'Chartering' except for Sales and Purchase and Derivatives. The purpose of the chartering is to switch a freight rate between the parties. That is, the best of trading strategies is 'buying-low selling-high.'

According to Clarkson's source in Figure [2], the number of the time charter contract in Panamax sector was 711 fixtures for the last three years. Out of them, the 1year-term of the deal was 166 fixtures (23.3%). Various T/C contracts had buffering ranges from 0 to 6 months. The 3-month interval was about 40% among the fixtures.

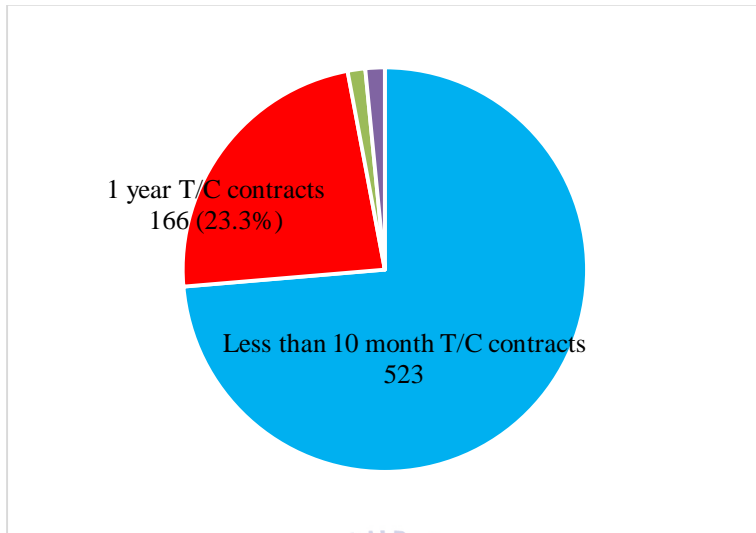


Figure 2 No. of Panamax Fixtures

2.1 T/C Option to Extend

Alizadeh and Nomikos (2009) introduced the implied options embedded in the time charter contract. These types of options usually consist of a mother contract plus an additional option (called T/C option) to extend the pre-specified period of the deal. Since the T/C option is given to the charterer for free without evaluating the economic

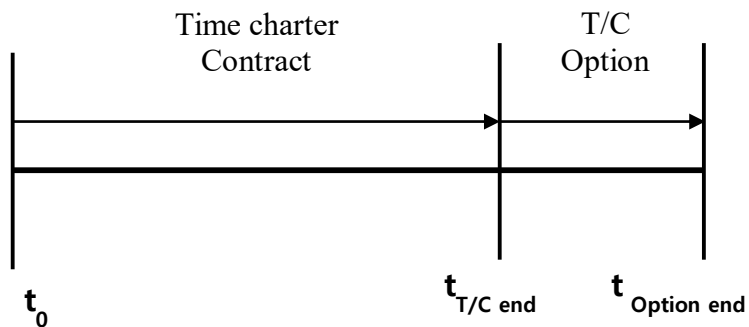


Figure 3 Structure of T/C Option

value, it might be quite profitable to the charterers. Figure [3] shows the structure of the T/C option.

The reasons why the ship-owners grant counterparties the T/C options at no cost are as follows:

1. To make the contract charming to the charterers.
2. To sustain the relationship with a credible charterer.

Starting off with the financial crisis, the world economic growth as the driver of the global trade has plunged. Unfortunately, consistently cumulated brand-new ships in shipping market have pushed the freight market down. These over-carrying capacities have strangled the players in the market. For ship-owner possessing the vessel built in the bull market, the expensive cost of the shipbuilding has been their constant curse. For the charterers borrowing the ship at a high charter rate, the crash of the freight rate has exhausted the chartering desk.

Under this state of shipping market, to attract the charterers, the T/C options might tend to be used when experiencing the bearish market like the era of post-global crisis.

This paper studies for pricing the 3-months T/C option embedded in their 1-year mother contract, especially in Panamax sector. These time conditions of the T/C contract and the option are reflected as shown in Figure [4].

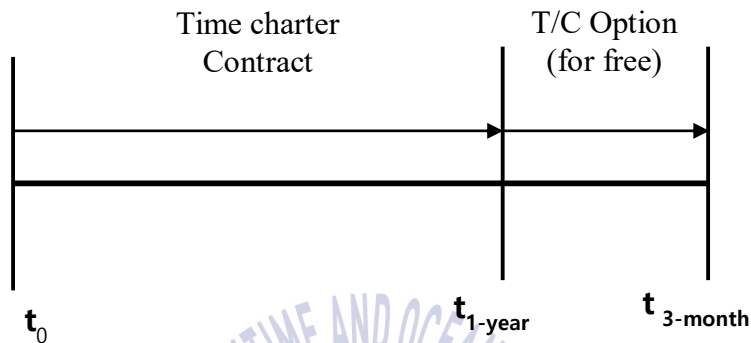


Figure 4 Specified Structure of T/C Option

The T/C options are different from the paper one usually traded in the derivative market. The distinction of the option between them is whether the mother contracts exist or not. This option cannot be individually traded in the market, as the optional period with physical operations of the vessel is not sufficient for time to generate the profit. However, this option might be so valuable that it should need to estimate the value of the options before chartering decision.

2.2 Chartering-out Strategy

The ship-owner or the charterer who secures carrying capacities has to find a trading counterparty willing to pay the freight rate in return for the term of usage. At this point, they may consider selling a part of the time of a ship. More specifically, the vessel obtained through the long-term can be chartered-out in forms of the short-term.

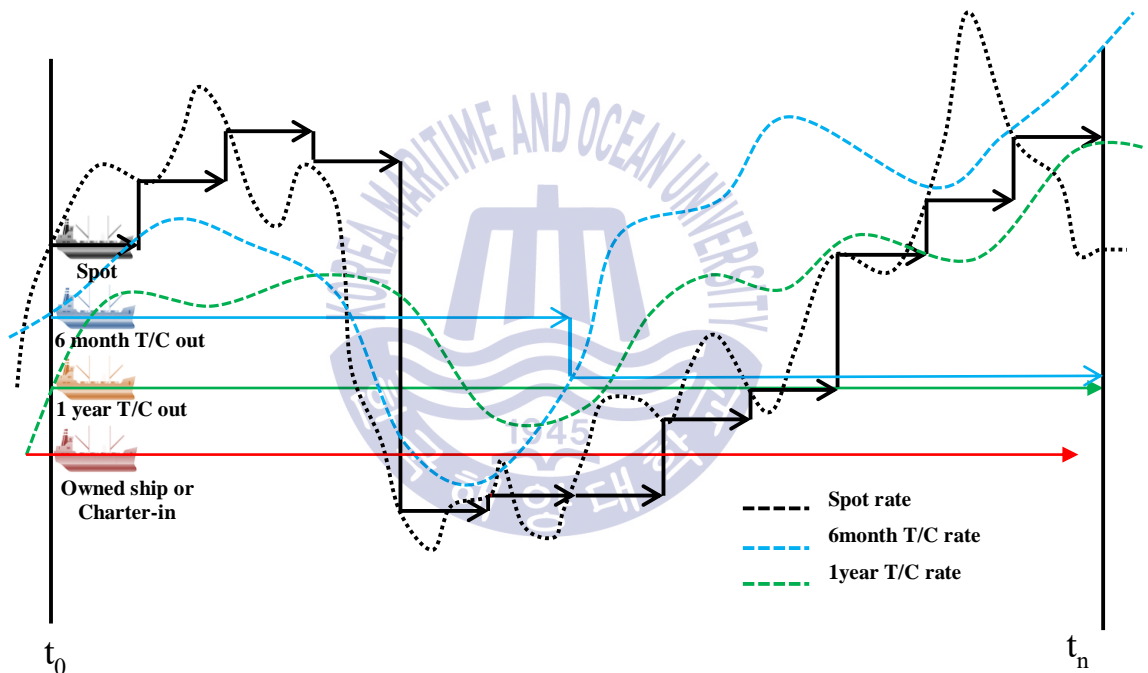


Figure 5 Comparison of Chartering-out Vessel

As confirmed in Table [1], the total number of Panamax fixtures was 2,400 contracts in 2016. To be specific, the 448 owners made contracts for 1,292 vessels. The owners had various chances to sell the obtained vessels as below tables. The vessels can be sold into time-charter contracts, spot contracts, and combined contracts.

Table 1 The details of Panamax fixtures in 2016

Object		Fixtures	No. of T/C	No. of Spot	No of Combined (pure Double T/C)	
Owner	446	Spot contracts	2,317	55	1,088	149(4)
TTL Vessels	1,292	T/C contracts	223	(based on the number of vessels)		

As shown in Figure [5], there can exist many combinations to sell the secured vessel according to the charter periods. Yun et al (2016) revealed that the cause of the shipping companies' collapse might derive from the failure of the trading decision. These decisions happen very often in chartering practice. The wrong judgments have plagued the chartering desk and harassed the management of the shipping companies. In this regard, it raises another research question concerning the trading strategy in chartering. This question would intimately be linked to the well-known theory, Efficient Market Hypothesis.

Since Fama's studies (1965, 1970, 1991), whether the market is efficient has been controversial in academia even in the business fields. In the academic world, it accepted that there are three kinds of efficient market hypothesis (EMH): weak, semi-strong, and strong form depending on the level of information reflected. Shortly speaking for the EMH, it is impossible to beat the market persistently because the market is efficient and all information is already reflected. Although the fact might be true, there exist a winner and loser in the market. It easily overlooked that the timing and duration are crucial for investment.

The shipping market where its players participate might meet the conditions of perfect competition, which are a competitive market, identical products, free entry and exit, and perfect information. The academic research to test the efficiency of the shipping market can be found in the literature (Ådland and Koekebakker, 2004; Adland and Strandenes, 2006; Alizadeh and Nomikos, 2006, 2007). The mixed results implicate that there is the room for success to beat the market regardless of the efficiency.

Therefore, the ML tools are adapted to support chartering decisions. Adopting the ML tools are expected to improve the decision to be made in chartering practice.



Chapter 3 Literature Review

The research questions of this paper are related to regression and classification, which are the representatives of decision-making. The decision is subsequently made through the process of learning based on them. This chapter investigates the previous studies on that problems. Besides, this research concerning shipping or maritime is also reviewed.

3.1 Regression-related Problem with Machine Learning

In finance studies, there have been many efforts to apply artificial neural network model (ANN) to forecasting the price of the stock.(Roh, 2007; Chen, Härdle and Jeong, 2010; Guresen, Kayakutlu and Daim, 2011; Wang *et al.*, 2011; De Oliveira, Nobre and Zárate, 2013; Ticknor, 2013; Kumar and Thenmozhi, 2014), gold (Kocak and Un, 2014; Kristjanpoller and Minutolo, 2015), oil (Sompui and Wongsinlatam, 2014; Kristjanpoller and Minutolo, 2016) and other commodities (Zhang, 2003; Chen, Härdle and Jeong, 2010; Khashei and Bijari, 2010; Chaudhuri and Ghosh, 2016).

Zhang (2003) combined ANN and ARIMA for time series forecasting because the series can be composed of both linear and nonlinear parts. As their achievement of prediction is easily confirmed in their area, the former is better estimated by using the econometric model, and the latter by ANN.

Roh (2007) designed the framework to predict the volatility and the direction of the stock price. To overcome the lack of economic background, he carefully chose

the inputs from the parameters of financial time series models such as EWMA and ARCH-family. The finding was that the ANN-EGARCH model outweighed others.

Khashei and Bijari (2010) proposed the novel model equipped with ARIMA and ANN in predicting three types of data (Wolf's sunspot, Canadian lynx series, and UK/US exchange rate). They included the lags of errors from ARIMA forecasts as the variables as well as that of the original series. The difference with the Zang's hybrid model (2003) is whether the original series is included in what ANN learns.

Wang et al. (2011) proposed the neural network model assigning the pre-processed series to the input variable, which is called 'wavelet-decomposition' that can be used to divide the signal into high and low frequency. Although the model performance differed according to the level of decomposition of the series, one of the proposed model was more suitable for their data than ANN.

Guresen et al. (2011) used a dynamic architecture for ANN (DAN2) developed by Ghiassi and Saidane (2005) to compare the naïve model of ANN and Roh's hybrid model (Roh, 2007). Despite the possibility of DAN2, the simple ANN empirically proved to be the best one.

Wang et al. (2012) came up with the integrated model which linearly summates the results produced by each model like ANN, ARIMA, and ESM (exponential smoothing model). Their experiments revealed the outcome of which the hybrid one can be more fitted than others.

Ticknor (2013) tested ANN equipped with Bayesian regularization term to predict the stocks. To protect the likelihood of overfitting during the training process, the penalized term was added to the cost function.

De Oliveira et al. (2013) attempted to identify the variables according to technical, fundamental, and time series analysis of the stock, macroeconomic factors, and financial data. Even though they used the parsimonious form of ANN, the procedure of extracting the important ones from scrutinizing the relation among the variables was plausible and understandable.

Sometimes exceptional cases are found in literature, as it is with the findings of Kocak and Un (2014) which are contrary to others. They compared ANN with the conventional model, ARCH-family to predict the returns of gold. Their paper concluded that as unexpected, the extrapolation of GJR-GARCH (1,1) outweighed that of ANN.

Sompui and Wongsinlatam (2014) showed that for predicting the crude oil price, ANN having the appropriate number of hidden neuron had better performance than the least-square model.

Kristjanpoller and Minutolo (2015) devised the hybrid model including GARCH model and ANN model for forecasting the volatility of Gold price. They used the advantages of each model to overcome the limitation inherently included in them. For example, GARCH generalized by Bollerslev (1986) can capture the volatility clustering of time series. With ANN learning errors from GARCH forecasting, the hybrid one can improve the performance.

Kristjanpoller and Minutolo (2016) modified the ANN to learn the forecast from GARCH. At the same time, as the factors having an impact on oil price were orderly introduced in the model as the inputs to identify which one is the most important variable to predict the price.

Chaudhuri and Ghosh (2016) adopted NARX (nonlinear autoregressive models with exogenous input) neural network model (Lin *et al.*, 1996) as the modified form of recurrent neural network to gauge the future exchange rate of India/US. NARX is characterized by feeding the output of output layer as the input. Then, it can be shown in $y_{(t)} = f(y_{(t-1)}, y_{(t-2)}, \dots, y_{(t-n_y)}, u_{(t-1)}, u_{(t-2)}, \dots, u_{(t-n_u)})$ where $y_{(t)}$ is regressed on its previous values and other exogenous inputs, $u_{(t)}$.

In relatively recent time, new techniques of ML have been introduced in the literature. Support vector machine (SVM) and Ensemble learning model are representative ones. One of the main issues in the application of ANN is how well the model is designed to protect the occurrence of over-fitting or under-fitting outcomes.

The studies reviewed have shown that SVM is superior to ANN in forecasting the price of the stock (Trafalis and Ince, 2000; Pai and Lin, 2005), oil, and, other commodities rather than ANN.

Trafalis and Ince (2000) compared SVM with ANN and radial basis function neural network (RBFNN) concerning forecasting US stock prices. They searched the optimal parameters of SVM by changing the value of penalty term and kernel parameters in turn. However, the paper showed that the results varied case by case.

Tay and Cao (2001) used SVM and ANN to predict various financial time series such as government bonds and index futures. They selected the input variables extracting the relative change in the percentage of the price. In most experiments, SVM yielded the higher performance than ANN. They demonstrated the reason of the superiority of SVM over ANN. Specifically speaking, SVM has the structural risk minimization principle, fewer parameters to be tuned, and the convergence of global minimum in the cost function.

Pai and Lin (2005) followed the hybrid model that Zhang designed (Zhang, 2003). The concept is that the series of interest can be divided into linear and nonlinear components. ARIMA estimates the data, and then ANN learns the residual that ARIMA cannot estimate. Lastly, the hybrid model incorporates forecast of ARIMA and ANN.

Chen et al. (2010) devised the hybrid model organized by SVM and GARCH and compared it with other models such as moving-average model, GARCH, EGARCH, and ANN+GARCH. To estimate the parameters, they simulated the data under the assumption of the artificial condition, normal distribution, and student t distribution. Then, they found that the hybrid one performed the highest.

Kumar and Thenmozhi (2014) created the variety of the integrated models that combine ANN, SVM, RF, and ARIMA, which are the representatives of ML tools and time series analysis respectively. They compared the proposed models with other naïve models and found that the formers yielded the better performances of prediction than the latter.

It is difficult to find that the RF has been used for predicting in domestic literature. Suh (2016) confirmed a high prediction accuracy of RF regarding extrapolating the future path of foreign exchange. In particular, by using the feature selection of RF, the variables affecting on forecasting exchange rate were picked in order of the most importance.

Table 2 Application of ANN in Financial Market

Authors	Fields	Model and Benchmark Variables	Performance	Results
Zhang(2003)	Forecasting UK/US exchange rate	ANN+ARIMA vs. ARIMA, ANN Time-lagged variables	MSE, MAD	Hybrid model >>others
Roh(2007)	Forecasting volatility and direction of KRX stock index	ANN+EWMA and ANN+ARCH family vs. ANN Kospi 200, lags of Gov. bonds yield and price. Open interest, parameters of time series models	MAE, Hit ratio	ANN+EGARCH >>others
Khashei and Bijari(2010)	Forecasting UK/US exchange rate	ANN+ARIMA vs. ARIMA, ANN, Zhang model(2003) Time-lag	MAE, MSE	Hybrid model>>others
Guresen et al.(2011)	Forecasting NASDAQ stock exchange index	ANN, ANN+GARCH, Dynamic ANN, DANN+GARCH Time-lagged variables, parameters of GARCH	MSE, MAD	ANN >> others
Wang et al.(2011)	Forecasting Shanghai stock index	Wavelet DBPNN vs. ANN Decomposed series	MAE, RME, MAPE	WDBPNN >>ANN
Wang et al.(2012)	Forecasting Shenzhen and Dow Jones Indices	ANN, ESM(exponential smoothing), ARIMA, RWM(random walk) vs. hybrid model with their errors weighted errors from each model by Genetic Algorithm	MAE, RMSE, MAPE, ME, DA	Hybrid model>>other

		BRNN vs. ARIMA, ANN		
Ticknor(2013)	Forecasting US stocks	Daily high, low, and close price, 6 technical indicators	MAPE	BRNN>>others
De Oliveira et al.(2013)	Forecasting a Brazil stock	ANN vs. actual Variables from economic and financial theory, technical, fundamental, and time series analysis	MAPE, RMSE, THEIL, POCID(% of correct directional prediction)	ANN
Kocak and Un(2014)	Forecasting gold returns	ANN vs. ARCH-family Time-lagged variables	MSE, MAE	GJR-GARCH >>ANN
Sompui and Wongsinlatam(2014)	Forecasting oil price	ANN vs. LSM(least-square)	MSE	ANN>>LSM
Kristjanpoller and Minutolo(2015)	Forecasting gold price volatility	5 ANN+GARCH vs. GARCH Spot and futures price of gold, Daily variation of exchange rate and oil price, DJI and FTSE returns	MSE, RMSE, MAD, MAPE	Hybrid model incorporating all variables >> other models and GARCH model
Kristjanpoller and Minutolo(2016)	Forecasting oil price volatility	ANN+GARCH vs. ARFIMA, GARCH GARCH forecast, oil price return, Dow Jones, FTSE, and exchange rate	MSE, RMSE, MAD, HMAE(heteroscedasticity-adjusted)	ANN-GARCH>>
Chaudhuri and Ghosh(2016)	Forecasting Indian/US exchange rate	NARX vs. ANN, ARCH-family Real-time variables including returns of Dow Jones, HangSeng, and DAX, crude oil price, India and US VIX	MSE, R ² , TI(Theil Inequality)	NARX >> others

Table 3 Application of SVM and RF in Financial Market

Authors	Fields	Model and Benchmark Variables	Performance	Results
Trafalis and Ince (2000)	Forecasting stock prices	SVM vs. ANN, RBFNN(radial basis function)	MSE	Mixed results
Tay and Cao(2001)	Forecasting S&P stock index futures, US 30y bond and 10y bond, German 10y bond, French stock index futures.	SVM vs. ANN Time-lagged variables from 5days relative difference in percentage of price	NMSE(normalized),MAE,DS, WDS(weighted)	SVM>>others
Pai and Lin(2005)	Forecasting stock prices	Hybrid model vs. SVM, ARIMA	MAE, MAPE, MSE, RMSE	Hybrid model(SVM+ARIMA) >>SVM, ARIMA
Chen et al.(2010)	Forecasting NYSE index	SVM+GARCH vs. moving average, GARCH, EGARCH, ANN+GARCH Residuals of GARCH process	MAE, DA	Hybrid model>>others
Kumar and Thenmozhi(2014)	Forecasting S&P CNX Nifty index	SVM+ARIMA, ANN+ARIMA, RF+ARIMA, ARIMA, ANN, RF, SVM Time-lagged variables	NMSE, MAE, RMSE, DA	SVM+ARIMA >> others
Suh (2016)	Forecasting foreign exchange in Korea	Random Forest + GARCH vs. AR(autoregressive), GARCH	RMSE	Hybrid model >> others

Notice that the application of ML techniques in forecasting has been verified in the vast financial literature. Similarly, it can find that the body of studies on the pricing of the financial and commodity options by using ML methods.

Yao et al. (2000) carried out forecasting of Nikkei 225 stock index options with ANN. They carefully choose three parameters used in Black-Scholes Models (BSM) because of the model performance compared with previous studies. For the comparison between ANN and BSM, they separated the data according to moneyness of the option. BSM presented the better pricing at-the-money option than ANN, while it showed that the performance of the latter is useful for pricing in-the-money and out-of-money option.

Gencay and Qi (2001) adopted the regularization technique to reduce the possibility of overfitting in ANN. For aiming at pricing and hedging the option, they introduced three kinds of regularization such as Bayesian penalty term, early stopping criteria, and bagging. Also, the authors selected the normalized spot and option prices by strike price as variables. ANN with Bayesian term substantially reduced the error in pricing and hedging, whereas the bagging had the limitation of time-consuming, but the highest performance than others.

Morelli et al. (2004) used ANN and RBFNN to estimate the option price and Greeks. What is unique about this paper is that the experiments were partitioned into two types of the dimensionality of inputs. When all the variables of BSM were exploited, despite the fact that the time to calculate was consuming, the performance of ANN became equivalent to RBF.

Tseng et al. (2008) designed ANN with Grey model and EGARCH to capture the time-varying asymmetric volatility of return of the options in Taiwan. Grey model developed by Deng (1982) has been verified to extract the useful information from small samples and poor information and EGARCH pioneered by Nelson (1991) has been widely accepted to explain the asymmetry in the volatility of time series, which is more weighted on the negative information. Having combined ANN with Grey model and EGARCH, they tried to enhance pricing the option, but the results were not better than expected.

Liang et al. (2009) attempted to improve the valuation of the option, which follows the process that the forecasts from the binomial tree, finite difference, and Monte-Carlo simulation. The results from these experiments implicated that since the factors affecting the option markets are mixed with linear and nonlinear elements, instead of forecasting with a single model, they proposed the novel model combining the ML models and the above models. The proposed model produces the more precise forecasts of the option price as it can learn the predictions generated from each model.

Most results found in the literature commonly said that since the performance of the ANN is comparable to that of the BSM, the ANN is eligible for the alternative of the BSM.

Table 4 Application of Machine Learning Methods in Pricing Options

Authors	Fields	Model and Benchmark Variables	Performance	Results
Yao et al.(2000)	Forecasting option price in NIKKEI 225 index	ANN vs. BSM Spot price, strike price, time to maturity	NMSE(normalized)	Mixed results ATM: BSM better ITM,OTM: ANN better
Gencay and Qi(2001)	Pricing and hedging S&P500 index option	ANN with Bayesian regularization, early stopping, and bagging Dependent: C/X, independent: S/X, time to maturity	MSPE, AHE(average hedging error)	ANN+Bayesian regularization >> others
Morelli et al.(2004)	Pricing European and American option and Greek letters	ANN vs. RBFNN Spot price, time to maturity	Error unknown	Mixed results RBFNN had fast training time ANN is robust tool in Greek modeling.
Tseng et al.(2008)	Pricing Taiwan stock index option	ANN+Grey+EGARCH vs. ANN+EGARCH Spot price, strike price, risk-free rate, time to maturity, Grey+EGARCH volatility	RMSE, MAE, MAPE	Mixed results
Liang et al.(2009)	Forecasting option price in Hong Kong	ANN and SVM with BT(binomial tree), FD(finite difference), and MC(Monte Carlo) Forecasts from BT, FD, and MC	MAE, TAFE(total average absolute), ARFE(average relative), TARFE(total average relative)	Hybrid SVM>> ANN

In another field, the applicability of ML has been seen as well.

Diaz-Robles et al. (2008) exploited ANN to predict air quality in Chile. For the study, they built the hybrid model with ANN and ARIMA, Multi-linear regression (MLR) respectively. When it comes to the model performance, the one with ANN and ARIMA had the best.

Hung et al. (2009) proposed ANN to forecast the real-time rainfall in Bangkok. The best model chosen by comparing the performance of them had the time lag variable as well as the meteorological parameters.

Zhu and Wei (2013) made a hybrid model integrating ARIMA and the least square SVM (LSSVM) to approximate linear and non-linear patterns of the carbon price. They adopted a heuristic algorithm, particle swarm optimization (PSO) to optimize the parameters used in LSSVM.

Zhang et al. (2013) tried to seek the proper model revealing the high level of prediction in pork price in China. By the way, the demonstration about the structure of the model they used is insufficient in their paper. It noted that the prediction power of the model, however, is sufficient to forecast the pork price.

Lama et al. (2016) employed the time-delay neural network with GARCH for predicting the edible oil price in India and Global market. To find the optimal parameter of time-lags they used ARIMA and then, based on that, the forecasts of each model was combined. Unfortunately, it was surprisingly noticed that the performance of the combined model positioned between naïve GARCH and TDNN. The reason they demonstrated was that the performance largely depends on the size of bandwidth and kernel function used.

Mirakyan et al. (2017) applied SVM, ANN, and ridge regression to estimate the electricity price. Instead of the use of single model only, they used the weighted forecasts processed by giving equal-weight, inverse-RMSE weight, and OLS weight to the prediction of each model respectively. The most predictable model was the hybrid model with inverse-RMSE weight, while the simple ridge regression was comparable to the best.

Ahmad et al. (2017) utilized ANN and RF to forecast the energy consumption of a hotel in Spain. They included the social parameters like the number of guest and the number of room booked as well as the weather factors and time-lag of consumption. Through the feature selection of RF, the considered variables were ranked by the level of importance. Afterwards, they concluded that though the predictability of ANN surpassed RF, the latter may deserve to be comparable to the performance of the former.

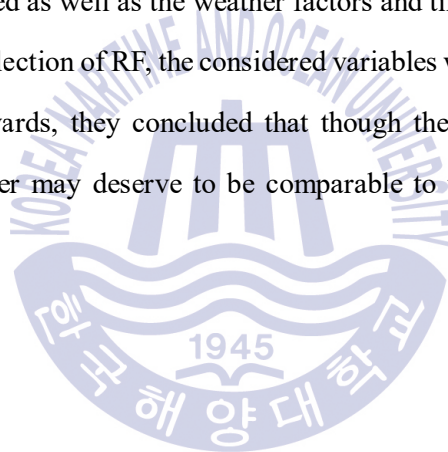


Table 5 Application of Machine Learning Methods in Other Area

Authors	Fields	Model and Benchmark Variables	Performance	Results
Diaz-Robles et al.(2008)	Forecasting air quality in Chile	ANN+ARIMA vs. MLR, ARIMA, ANN Time-lagged and Meteorological variables	E2(coefficient of efficiency), ARV(average relative variance), RMSE, R2, SEP(percent standard error of prediction), PI(persistence index), BIC	ANN+ARIMA >> others
Hung et al.(2009)	Forecasting real-time rainfall in Thailand	Seven ANN Rainfall as well as humidity, temperature pressure, cloudiness etc	EI(Efficiency index), RMSE, R ²	No comparison
Zhu and Wei (2013)	Forecasting carbon price	Hybrid model(ARIMA+LSSVM) >>ARIMA+ANN, LSSVM, ANN, ARIMA unknown	RMSE, Directional statistics	Hybrid model >> others
Zhang et al.(2013)	Forecasting pork price	SVM unknown	RMSE	No comparison

Lama et al.(2016)	Forecasting volatility of edible oil price	TDNN(Time-delay) +GARCH vs. TDNN, GARCH Times-lagged but variable specification is unknown	RMSE Theil-U DA	Mixed results The performance of Combined model positioned between other models
Mirakyan et al.(2017)	Forecast electricity price	SVM, ANN, RR(ridge regression), hybrid model Times-lagged but variable specification is unknown	RMSE, MAPE, MdAPE(median)	Mixed results but a hybrid model has consistently outperformed others in most cases
Ahmad et al.(2017)	Predicting energy consumption of building	ANN vs. RF Weather condition, time-span, number of guests, room booked and time-lagged variables	RMSE, CV(coefficient of variation), MAD, MAPE	ANN>>RF Though ANN outperforms RF, the latter is comparable to the former

3.2 Classification-related Problem with Machine Learning

The choice between some alternatives is intuitively no less sophisticated but no less straightforward than the regression-related problem.

Fernández-Rodríguez et al. (2000) employed ANN to test the trading strategy in Madrid stock index. According to the trading strategy from ANN, they traded the stock index and found the returns is more profitable than a prior model (Pesaran and Timmermann, 1992).

Kim (2003) suggested that the predictability of SVM outperformed ANN and case-based reasoning. He tuned the parameters with regard to penalty term C and kernel \mathcal{K} . He checked whether SVM consistently outweighs others by using McNemar tests.

Huang et al. (2005) used SVM for predicting the movement of NIKKEI 225 index. They selected S&P 500 index and US/JP exchange rate that significantly influence Japan's export as the inputs. Then they built a combining SVM to capture the relationship between a day before prior behavior of inputs and today's direction of NIKKEI index. Finally, they compared the forecast of the proposed model with SVM, the random walk model, Elman neural network (ENN), and other statistical methods. Moreover, they noticed that the reason why SVM outperforms others is based on 'the structural risk minimization principle.'

Kara et al. (2011) applied the proposed model to forecast the direction of the stock price in emerging market. They extensively experimented to find the suitable parameters of the model to be optimized. Afterwards, the designed SVM model outperformed the result of ANN and prior research.

Wang and Choi (2014) constructed SVM using the variables through the principal component analysis (PCA). Since the shortcoming of SVM is the calculation of transformed inputs to a high dimension, they reduced the number of stocks having an effect on the composed index under PCA and incorporated them with the macroeconomic factors as inputs. Also, the period of their experiments was designed to follow rolling windows of time in order that the model can be generalized any time.

Wang and Shang (2014) used the least square SVM (LSSVM) to forecast the direction of China Security index 300. The ten variables were chosen through technical analysis. Compared with the Bayesian neural network, QDA, and LDA, the SVM was accepted as the valuable model to predict. Eventually, they tested whether the LSSVM consistently surpasses others under McNemar test.

Tanaka et al. (2016) introduced RF technique for capturing the warning signal before bank failure. The distinction of the study is the choice of the variables. The former studies included the macroeconomic data to predict the financial crisis, while they made use of bank-level financial data only. They ended up with comparing the existing model such as logistic regression and decision trees with RF and then they achieved constructing the RF-based early warning system.

Table 6 Application of Machine Learning Methods in Classification-related Problems

Authors	Fields	Model and Benchmark Variables	Performance	Results
Fernández-Rodríguez et al.(2000)	Forecasting the optimal trading strategy	Trading with ANN vs. Buy & hold, prior model Time-lagged variables	Hit ratio Sharp ratio	ANN>>others
Kim(2003)	Forecasting the direction of KOSPI	SVM vs. ANN, CBR(case-based reasoning) 12 Variables from technical analysis	Hit ratio	SVM>>ANN, CBR
Huang et al.(2005)	Forecasting the movement of NIKKEI index	Hybrid model vs. SVM, ENN, LDA(linear discriminant), and QDA(quadratic discriminant) Binary variables of S&P500 and US/JP exchange rate	Hit ratio	Hybrid>SVM>>others
Kara et al.(2011)	Forecasting the direction of stock price in ISE	SVM vs. ANN, prior research models 10 indicator from technical analysis	Hit ratio	SVM >> others
Wang and Choi(2014)	Forecasting the direction of KOSPI and HIS	SVM+PCA vs. SVM, ANN, ANN+PCA, RW 2 stocks from PCA, 2 macroeconomic factors	Hit ratio	SVM+PCA >> others
Wang and Shang(2014)	Forecasting the direction of CSI 300	LSSVM vs. BNN, LDA, QDA	Hit ratio	LSSVM>>others
Tanaka et al.(2016)	Forecasting early warning signals in bank failure	RF vs. Decision tree, Logistic regression 48 indicator from groups like profitability ratio, capitalization, loan quality, and funding	Accuracy unknown	RF>>others

3.3 Maritime-related Problem with Machine Learning

In maritime sector, the studies on the application of the ML disciplines have occasionally been introduced. The problems discussed in this research are mainly related to forecasting.

Li and Parsons (1997) attempted to confirm the prediction power of ANN and suggest a useful framework with regard to ANN in forecasting of tanker freight rate. The variables they used are dirty spot rate, tanker demand, and tanker supply. This study had concentrated on tuning the appropriate parameters and if they are not correct, it might cause poor or good predictions. Also, they considered the time consuming and model complexity for constructing an ideal model. Compared with autoregressive moving averages model (ARMA), the rates ANN produced were more precisely and consistently matched to the actual rates.

Mostafa (2004) indicated that it is useful for ANN to be utilized in forecasting the traffic volumes of Suez canal. The distinct points that differentiated from other research are that he did not normalize the input data and did not adopt the biases terms in the hidden layer. Despite that, it noted that ANN achieved the higher performance than the traditional statistical model, ARIMA. This result shows again that since the performance of ANN strongly depends on the type of data to be used and the parameters to be tuned by an experienced expert, these limitations should be taken into account when using ANN.

Lyridis et al. (2004) built more than 100 ANN with the different number of inputs and hidden nodes. Then, they extracted the right models according to the out-of-sample performance corresponding with each forecast interval. Although the results

for a short-term interval was comparable to the unknown naïve model, the longer the forecast interval, the larger the difference between them.

Yang et al. (2008) revealed the forewarning system for freight rates in the shipping market. The variables they chose are CCFI, CCBFI, and BDI as they are the representatives reflecting information about shipping demand and supply. In particular, based on the opinions of an experienced expert, they extracted the warning signal from the freight rates. The forewarning power the model produced was accurate with 100% in the out-of-sample test, while it is disappointing that the reasons why they determined the level of warning signals are insufficient.

von Spreckelsen et al. (2012) investigated the potential of ANN in forecasting the spot rate and FFA for Tanker market and comparing the profit gains of the trading strategies stemmed from the results of univariate models and multivariate models. Perhaps, since the forecasting and trading interval was set as one-day ahead, the performance of other time series models might have been comparable to that of ANN. Furthermore, they pointed out that while it is helpful for FFA prices to predict the spot rate, the opposite does not hold.

Fan et al. (2013) constructed the neural networks replacing the standard transfer functions with the wavelet function that can extract the necessary information from noisy data. Though the forecasts of ARIMA was almost equal to WNN during the short-term period, the longer the prediction interval, the better the performance of WNN increases.

Lyridis et al. (2013) developed the ANN model with the proper number of hidden layers and input variables in order to forecast FFA prices. According to the correlation coefficient, the candidate variables were filtered out. Even though the results from

the final model satisfied for trading, the authors recommended to be careful about using it in real trade.

Santos et al. (2014) used the naïve ANN model and radial basis function (RBF) NN model that adopts RBF instead of the typical transfer function like a sigmoidal function. For predicting the future T/C rates of VLCC, they chose the lagged variables that represent the information about supply and demand and market condition. Although the RBFNN consistently outperformed others regardless of the forecasting time interval, it was done mainly for the short-term forecast.

Han et al. (2014) proposed SVM model incorporating the wavelet analysis that can denoise BDI. Prior to implementing SVM model, BDI was decomposed into high frequency and low frequency until attaining the satisfied level of denoising BDI. Compared with ANN and statistical models, the proposed model achieved the utmost results rather than others.

Daranda (2016) attempted to apply ANN to predict vessel routes at Baltic sea. He had not only analyzed the waypoints and routes the mass traffic had heavily used but also utilized ANN to train the pattern of the traffic routes. The model revealed quite accurate forecasts.

Bao et al. (2016) employed SVM with correlation-based feature selection that helps to reduce the redundant variables. They mentioned that they had approached the variable selection with the macroeconomic views. Due to the structural superiorities of SVM, the paper showed that its performance deeply outperformed the comparative model, ANN.

Eslami et al. (2017) came up with a hybrid model built with both ANN and an adaptive genetic algorithm (AGA) to foresee tanker freight rates. Three inputs were

chosen amid the suggested variables from the relevant literature through stepwise regression. The evolution mechanism of the AGA can parameterize the essential criteria. When it comes to the accuracy, the hybrid model improved the results compared with the previous research and the two traditional models.



Table 7 Application of Machine Learning Methods in Maritime Sector

Authors	Fields	Model and Benchmark Variables	Performance	Results
Li and Parsons(1997)	Forecasting Tanker freight rates	Univariate-ANN, Multivariate-ANN, ARMA Tanker spot rate, Tanker demand, and supply	Adjusted average of MSE(ADAMSE)	Multivariate-ANN>>others
Mostafa(2004)	Forecasting Suez canal traffic volume	ANN vs. ARIMA Time lagged variables	RMSE	ANN>>ARIMA
Lyridis et al.(2004)	Forecasting Tanker freight rates	Various ANN vs. Naïve model(Unknown) Oil demand, fleets, crude oil price and production, T/C rates, newbuilding and secondhand prices, bunker and scrap prices, oil stock, and 2 event dummies.	MSE	Mixed results
Yang et al.(2008)	Classifying the level of forewarning for freight rates	SVM CCFI, CCBFI, and BDI	Hit ratio	No comparison
von Spreckelsen et al.(2012)	Forecasting and trading Tanker spot rate and FFA	Univariate-ANN and ARIMA and Multivariate-ANN, VAR, and VECM and RW Spot rate, FFA 1M, and 2M	R^2 , RMSE, Theils-U, Trading profits	Mixed results but univariate-ANN are dominant overall.
Fan et al.(2013)	Forecasting Tanker freight rates	Wavelet-NN(WNN) vs. ARIMA Amex oil index, Brent oil price, S&P500 volatility, S&P Global1200, Dow Jones, and MSCIA world transportation	MAE, RMSE, MAPE	WNN >>ARIMA

Lyridis et al.(2013)	Forecasting dry FFA prices	4 ANN Spot rate, order-book, fleet development, scrap value, T/C rate, contracting, sales, deliveries, and newbuilding prices for Capesize	MSE	No comparison
Santos et al.(2014)	Forecasting T/C rate for VLCC	ANN, RBFNN vs. ARIMA 1-year and 3-year T/C rate, Spot rate, demolition price, world crude oil output, deliveries, demolitions	MAPE, RMSE, Theil-U	RBFNN>>others
Han et al.(2014)	Forecasting BDI	Wavelet-SVM vs. ARMA, ANN, VAR Time-lagged variables from decomposed wavelets	RMSE	WSVM>>others
Daranda(2016)	Forecasting vessel route	ANN Waypoints, speed, course, MMSI number, vessel dimension, type of vessel	-	No comparison
Bao et al.(2016)	Forecasting BDI	SVM vs. ANN World GDP growth, iron ore, coal, and grain demand, vessel supply, fuel price	RMSE	SVM>>ANN
Eslami et al.(2017)	Forecasting Tanker freight rate	ANN+AGA vs. regression, moving average Fleet productivity, crude oil price, and bunker price	RMSE	ANN+AGA>>others

Chapter 4 Methodology

4.1 Option Pricing

There are a variety of option pricing models invented. These models can roughly be classified into the parametric and non-parametric models. This chapter introduces some representative models among them in a nutshell.

4.1.1 Benchmark: Black-Scholes-Merton model (BSM)

The Black-Scholes-Merton option pricing model (BSM) was invented by three scholars named in the title of the model (Black and Scholes, 1973; Merton, 1973). Their contributions have been enormous and tremendous, and the model is still widely accepted as the benchmark model in the market trade and the academic research because of its analytical tractability (Lajbcygier and Conner, 1997; Andreou, Charalambous and Martzoukos, 2006; Shinde and Takale, 2012). Since the derivation of the model is beyond this paper, the specification of the BSM is briefly described as follows.

$$c = S_0 N(d_1) - K e^{-rT} N(d_2)$$
$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \quad , \quad d_2 = d_1 - \sigma\sqrt{T}$$

Where c is the European call option price, S_0 is the spot price at time 0, K is the strike price, $N(\cdot)$ is the cumulative probability distribution function, r is the

risk-free rate, σ is the spot price volatility, and T is the time to maturity of the option.

As reported in the literature, the main drawbacks of BSM are the unrealistic presumptions. Most of the parameters are assumed to be constant during the life of the option. In particular, the volatility of the spot price is unknown in the initial stage of option pricing. Nevertheless, because of the simplicity of the model to price the option, the model is still loved and attractive to the fields and academia.

If this formula can apply to pricing the TC options, the assumptions with regard to exercise the option must be required as well as the presumptions of BSM. The details of its premise will be discussed in Ch. 5.

4.1.2 Machine Learning Methods

Artificial intelligence was first programmed by Alan Turing and Arthur Samuel in the 1950s. Since then, various types of ML have been introduced, and although there have been considerable difficulties to solve, they have been improved and evolved through the efforts of scholars.

The concept of ML is to mimic the learning process in the human brain. The designed models go through the process that adjusts the weights of parameters in the learning algorithm. Then, the final one can be set by the generalization techniques like n-folds cross-validation, early stopping, and regularization.

The epitomes of ML are ANN, SVM, and Ensemble learning. The next sections demonstrate the conceptual and mathematical approaches to these methodologies.

4.1.2.1 Artificial Neural Networks(ANN)

As shown in the literature, the applicability of ANN has been quite diverse. In addition, Tkáč and Verner (2016) confirmed that ANN has been widely used in a variety of business areas including accounting, credit rating, decision support, derivatives pricing, bankruptcy, and so on. Based on the review papers summing up the applicability of ANN, the structure of ANN is illustrated in Figure [6].

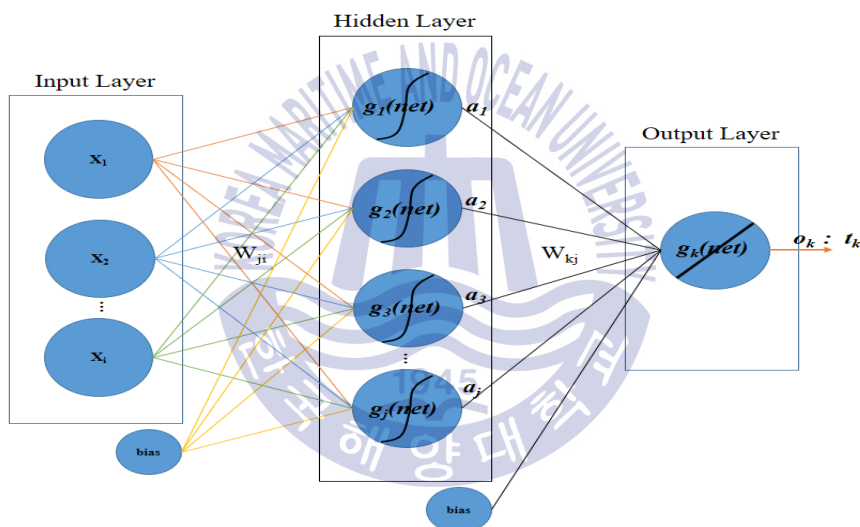


Figure 6 Structure of ANN

ANN is conceptually similar to the learning process of the neuron. More precisely, input layers accept the variable information, which is then passed throughout the networks between the layers. This mechanism can be expressed mathematically as follows.

For the input in the hidden nodes,

$$net_j = \sum_i x_i w_{ji}$$

where x_i is each input node and w_{ji} is the weights that reveal the connection strengths between the input node i and the hidden node j .

The outputs in the hidden nodes are the outcomes transformed through the transfer function, $g(z) = \frac{1}{1+e^{-z}}$.

$$a_j = g(net_j)$$

For the input in the output nodes,

$$net_k = \sum_j a_j w_{kj}$$

where w_{kj} is the weights between the hidden nodes j and the output nodes k .

The k th output of the out layers is $o_k = g(net_k)$.

For regression, the transfer function, $g(z)$, adopts the linear function to approximate numerical values.

The learning process of ANN is employing the error back-propagation that adjusts the synaptic weights. The following equations show how to process that algorithm in this discipline.

The E represents the loss function of the neural networks.

$$E = \frac{1}{2} \sum_j (t_k - o_k)^2$$

where t_k means a target value and o_k is the estimates of the model. To removing the square term when differentiating, the $1/2$ term is added to the equation.

Through the error back-propagation algorithm, the model adjusts the synaptic weights in the output nets until achieving the convergence to a certain threshold.

$$\Delta w_{kj} \propto -\frac{\partial E}{\partial w_{kj}}$$

where w_{kj} denotes a weight from the neuron j of the hidden layer to k neuron of the output layer.

Because the error is not directly a function of a weight, the partial derivative can be expanded by using the chain rule as follows.

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} = -\eta \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}}$$

where η is a learning rate, net_k denotes the k net input of output layer.

For convenient calculation, let's consider each partial derivative separately.

1. $\frac{\partial E}{\partial o_k} = \frac{\partial (\frac{1}{2} \sum_j (t_k - o_k)^2)}{\partial o_k} = -(t_k - o_k)$
2. $\frac{\partial o_k}{\partial net_k} = \frac{\partial (1 + e^{-net_k})^{-1}}{\partial net_k} = \frac{e^{-net_k}}{(1 + e^{-net_k})^2} = \frac{1}{1 + e^{-net_k}} \left(1 - \frac{1}{1 + e^{-net_k}}\right) = o_k(1 - o_k)$
3. $\frac{\partial net_k}{\partial w_{kj}} = \frac{\partial (w_{kj} a_j)}{\partial w_{kj}} = a_j$ where a_j is a activation value of hidden neuron j .

Substitute above equations into Δw_{kj}

$$\Delta w_{kj} = \eta \overbrace{(t_k - o_k) o_k (1 - o_k)}^{\delta_k} a_j$$

The final adjustment of error is derived from replacing the part of the local gradient with δ_k .

$$\Delta w_{kj} = \eta \delta_k a_j$$

In a similar way, the weights of the hidden nets can be updated.

$$\begin{aligned}
\Delta w_{ji} &\propto -\left[\sum_k \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial net_k} \frac{\partial net_k}{\partial a_j}\right] \frac{\partial a_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \\
\Delta w_{ji} &= -\eta \left[\sum_k \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial net_k} \frac{\partial net_k}{\partial a_j}\right] \frac{\partial a_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \\
&= \eta \left[\sum_k \overbrace{(t_k - o_k) o_k (1 - o_k)}^{\delta_k} w_{kj}\right] a_j (1 - a_j) x_i \\
&= \eta \left[\sum_k \delta_k w_{kj}\right] a_j (1 - a_j) x_i = \eta \left[\sum_k \overbrace{\delta_k w_{kj}}^{\delta_j}\right] a_j (1 - a_j) x_i \\
&= \eta \delta_j x_i \text{ where } x_i \text{ is the input variables.}
\end{aligned}$$

4.1.2.2 Support Vector Regression(SVR)

The proper use of ANN requires the appropriate estimates of the parameters and the sufficient number of sample. Therefore, ANN is always exposed to the empirical risk minimization.

For pursuing the principle of the structural risk minimization, the SVM was formulated by Vapnik (1995, 1997). This principle looks for the minimization of an upper bound of generalization error and simultaneously for the minimization of error from the training data.

Another core property of SVM is that the mechanism of the model is analogous to dealing with the quadratic programming (QP) problem so that SVM obtains the unique and globally optimal solution, while ANN is often getting stuck in local minima. This SVM can be expanded to regression problems, which is called 'SVR.'

Given a set of data (x_i, y_i) , where input vector $x_i \in \mathbb{R}^p$ and output scalar $y_i \in \mathbb{R}^1$. The latter, y , is treated as the target value in SVR. The aim of the model is to find a linear regression function $f(x)$, which estimates the real function, $g(x)$ as below.

$$f(x) = w^T x_i + b$$

The ε -insensitive SVR is harnessed to solve the linear regression.

$$\min \quad \frac{1}{2} w^T w$$

$$\text{s.t.} \quad \begin{aligned} w^T x_i + b - y_i &\leq \varepsilon \\ y_i - w^T x_i - b &\leq \varepsilon \end{aligned}$$

The ε -insensitive loss function, L_ε , is

$$L_\varepsilon(y, f(x)) = \begin{cases} |y - f(x)| - \varepsilon & \text{for } |y - f(x)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

The ε -insensitive SVR is ignoring the error less than ε , which forms the ε -tube as shown in Figure [7] and the ε -insensitive loss function is in Figure [8].

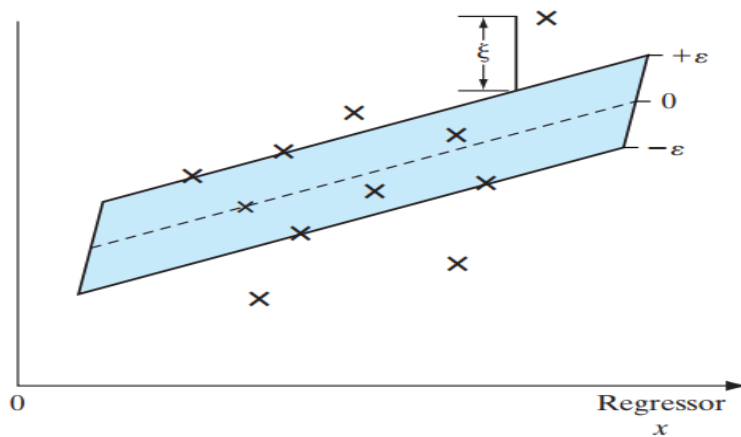


Figure 7 ε -insensitive SVR

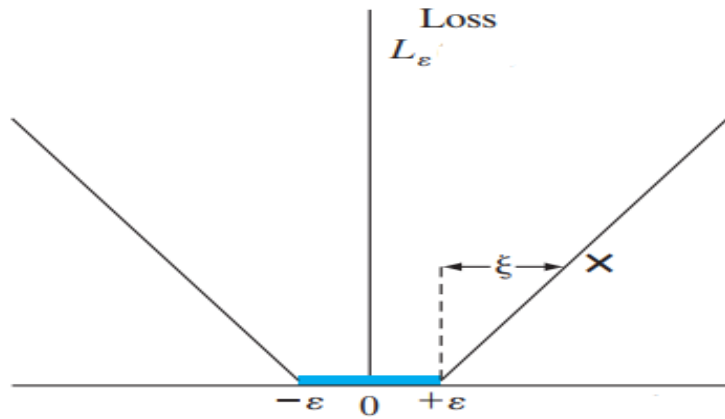


Figure 8 ε -insensitive Loss Function

Consider the infeasible case where the data cannot satisfy the constraints. The slack variables ξ_i, ξ'_i are introduced as illustrated in Figure [7] and [8].

$$\begin{aligned}
 \min \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n (\xi_i + \xi'_i) \\
 \text{s.t.} \quad & y_i - w^T x_i - b \leq \varepsilon + \xi_i \\
 & w^T x_i + b - y_i \leq \varepsilon + \xi'_i \\
 & \xi_i, \xi'_i \geq 0
 \end{aligned}$$

where penalty term, C , controls the trade-off relationship between the error and the flatness of the $f(x)$.

Then, the corresponding loss function is

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{if } |\xi| > \varepsilon \end{cases}$$

The Lagrange multiplier technique is required to solve the quadratic problem.

$$L = \frac{1}{2} w^T w + C \sum_{i=1}^n (\xi_i + \xi'_i) - \sum_{i=1}^n (\eta_i \xi_i + \eta'_i \xi'_i) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + w^T x_i + b) - \sum_{i=1}^n \alpha'_i (\varepsilon + \xi'_i + y_i - w^T x_i - b)$$

where η_i , η'_i , α_i , and α'_i denotes the Lagrange multipliers that satisfy the constraints of $\eta_i, \eta'_i, \alpha_i, \alpha'_i \geq 0$.

The minimization of the above function is conducted by partially differentiating with regard to w , b , and ξ .

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n (\alpha_i - \alpha'_i) x_i = 0 \rightarrow w = \sum_{i=1}^n (\alpha_i - \alpha'_i) x_i$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n (\alpha_i - \alpha'_i) = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0$$

$$\frac{\partial L}{\partial \xi'_i} = C - \alpha'_i - \eta'_i = 0$$

Substitute the results through partial derivative into the Lagrange function.

$$\begin{aligned} \max \quad & L(\alpha_i, \alpha'_i) \\ & = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha'_i)(\alpha_j - \alpha'_j) x_i^T x_j - \varepsilon \sum_{i=1}^n (\alpha_i - \alpha'_i) \\ & \quad + \sum_{i=1}^n y_i (\alpha_i - \alpha'_i) \\ \text{s. t.} \quad & \sum_{i=1}^n (\alpha_i - \alpha'_i) = 0 \\ & \alpha_i, \alpha'_i \in (0, C) \end{aligned}$$

Replace $w = \sum_{i=1}^n (\alpha_i - \alpha'_i) x_i$ with w in $f(x) = w^T x_i + b$.

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha'_i) x_i^T x + b$$

Finally, b can be calculated by using the Karush-Kuhn-Tucker (KKT) conditions. The key to them, that is, requires that the product between dual variables and constraints be zero at the point of the solution.

$$\alpha_i (\varepsilon + \xi_i - y_i + w^T x_i + b) = 0$$

$$\alpha'_i (\varepsilon + \xi'_i + y_i - w^T x_i - b) = 0$$

$$(C - \alpha_i) \xi_i = 0$$

$$(C - \alpha'_i) \xi'_i = 0$$

Because of the conditions above, the samples with corresponding $\alpha_i, \alpha'_i = C$ lie outside the ε -insensitive tube. Furthermore, $\alpha_i \alpha'_i = 0$ and the multipliers cannot be simultaneously nonzero. Hence, if $\alpha_i, \alpha'_i \in (0, C)$ has a certain value, the slack variables is zero.

$$b = y_i - w^T x_i - \varepsilon \quad \text{for } \alpha_i \in (0, C)$$

$$b = y_i - w^T x_i + \varepsilon \quad \text{for } \alpha'_i \in (0, C)$$

For the non-linear regression, the input space is transformed into higher feature space through the non-linear function, $\phi(x_i)$. The optimal regression function is $f(x) = w^T \phi(x_t) + b$.

By using the kernel function, $K(x_i, x) = \phi(x_i)^T \phi(x)$, the same procedures are conducted.

$$\begin{aligned}
\max \quad & L(\alpha_i, \alpha'_i) \\
& = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha'_i)(\alpha_j - \alpha'_j)K(x_i, x_j) - \varepsilon \sum_{i=1}^n (\alpha_i - \alpha'_i) \\
& \quad + \sum_{i=1}^n y_i(\alpha_i - \alpha'_i) \\
\text{s. t.} \quad & \sum_{i=1}^n (\alpha_i - \alpha'_i) = 0 \\
& \alpha_i, \alpha'_i \in (0, C)
\end{aligned}$$

Likewise, the w and the hyperplane are derived as

$$\begin{aligned}
w & = \sum_{i=1}^n (\alpha_i - \alpha'_i) \phi(x_i) \\
f(x) & = \sum_{i=1}^n (\alpha_i - \alpha'_i) K(x_i, x) + b.
\end{aligned}$$

The strongest advantage of SVM makes use of the kernel function, which can make the computational complexity of the high-dimensional space reduced. The exemplars of the kernel are as below Table [8].

Table 8 Type of Kernel Function

Kernel type	Equation
Linear	$K(x_t, x) = x_t^T x$
Polynomial	$K(x_t, x) = (x_t^T x + 1)^d$
Radial Basis Function	$K(x_t, x) = \exp\left(\frac{-\ x - x_t\ ^2}{2\sigma^2}\right)$

4.1.2.3 Random Forest (RF)

Having been inspired by the combination of decision trees and bagging ideas, Breiman (2001) proposed the random forest (RF).

This model is

- excellent versatile from classification to regression;
- relatively fast to learn;
- simple to tune few parameters;
- able to be applied to high-dimension problems;
- easy to be implemented in parallel. (Cutler, Cutler and Stevens, 2012)

A Random Forest is a tree-based ensemble method with each tree having a collection of random variables. The trees used in Random Forests are following the algorithm of the binary recursive partitioning trees in Table [9].

Table 9 Algorithm of the Binary Recursive Partitioning Trees (Cutler, Cutler and Stevens, 2012)

The training data $\mathcal{D} = \{(x_i, y_i): i = 1, 2, \dots, n\}$, $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})^T$

1. Begin with all observations $(x_1, y_1), \dots, (x_N, y_N)$ in a single node.
 2. Repeat the following steps recursively for each unsplit node until the stopping criterion is met:
 - a. Find the best binary split among all binary splits on all p predictors.
 - b. Split the node into two descendant nodes using the best split (Step 2a).
 3. For prediction at x , pass x down the tree until it lands in a terminal node. Let k denote the terminal node and let y_{k_1}, \dots, y_{k_n} denote the response values of the training data in node k . Predicted values of the response variable are given by:
 - $\hat{h}(x) = \bar{y}_k = \frac{1}{n} \sum_{i=1}^n y_{k_i}$ for regression
 - $\hat{h}(x) = \arg \max_y \sum_{i=1}^n I(y_{k_i} = y)$ for classification, where $I(y_{k_i} = y) = 1$ if $y_{k_i} = y$ and 0 otherwise.
-

A Random Forest uses trees $h_j(X, \theta_j)$ as base learners, where θ_j is a collection of random variables. For the training data $\mathcal{D} = \{(x_i, y_i): i = 1, 2, \dots, n\}$, where $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})^T$ denotes the p predictors and y_i denotes the response, and a particular realization θ_j of θ_j , the fitted tree is denoted $\hat{h}_j(x, \theta_j, \mathcal{D})$ (Breiman, 2001). Because of the random component θ_j , at least two randomnesses under above process are given to the model. First, as with bagging, each tree is fit to an independent bootstrap sample from the original data.. Second, when splitting a node, the best split is found over a randomly selected subset of m predictor variables instead of all p predictors, independently at each node. The detailed algorithm for Random Forest is as shown in Table [10]:

Table 10 Algorithm of Random Forests (Cutler, Cutler and Stevens, 2012)

The training data $\mathcal{D} = \{(x_i, y_i): i = 1, 2, \dots, n\}$, $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})^T$
 For $j = 1$ to J :

1. Take a bootstrap sample \mathcal{D}_j of size N from \mathcal{D} .
2. Using the bootstrap sample \mathcal{D}_j as the training data, fit a tree using binary recursive partitioning:
 - a. Begin with all observations in a single node.
 - b. Repeat the following steps recursively for each unsplit node until the stopping criterion is met:
 - (i) Choose m predictors at random from the p available predictors.
 - (ii) Find the best binary split among all binary splits on the m predictors from Step (i).
 - (iii) Split the node into two descendant nodes using the split from Step (ii).

To make a prediction at a new point x ,

- $\hat{f}(x) = \frac{1}{J} \sum_{j=1}^J \hat{h}_j(x)$ for regression
- $\hat{f}(x) = \arg \max_y \sum_{j=1}^J I(\hat{h}_j(x) = y)$ for classification

where $\hat{h}_j(x)$ is the prediction of the response variable at x using the j th tree

When a bootstrap sample is taken from the data, some observations do not make it into the bootstrap sample. These are called “out-of-bag data,” and are extremely useful for estimating generalization error and variable importance.

Table 11 Algorithm of Out-Of-Bag Predictions

\mathcal{D}_j denotes the j th bootstrap sample and $\hat{h}_j(x)$ denotes the prediction at x from the j th tree for $j = 1, \dots, J$. For $i = 1$ to N :

1. Let $\mathcal{T}_i = \{j : (x_i, y_i) \notin \mathcal{D}_j\}$ and let J_i be the cardinality of \mathcal{T}_i .
2. Define the out-of-bag prediction at x_i to be
 - $\hat{f}_{oob}(x_i) = \frac{1}{J_i} \sum_{j \in \mathcal{T}_i} \hat{h}_j(x_i)$ for regression
 - $\hat{f}_{oob}(x_i) = \arg \max_y \sum_{j \in \mathcal{T}_i} I(\hat{h}_j(x_i) = y)$ for classification

where $\hat{h}_j(x_i)$ is the prediction of the response variable at x_i using the j th tree.

For regression with squared error loss, generalization error is typically estimated using the out-of-bag mean squared error (MSE):

$$MSE_{oob} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_{oob}(x_i))^2$$

where $\hat{f}_{oob}(x_i)$ is the out-of-bag prediction for observation i .

For classification with the zero-one loss, generalization error rate is estimated using the out-of-bag error rate:

$$E_{oob} = \frac{1}{N} \sum_{i=1}^N I(y_i \neq \hat{f}_{oob}(x_i)).$$

4.2 Trading Strategy

The trading strategy is a pre-specified rule for obtaining the profitable return. This strategy is closely related to the efficient market hypothesis (EMH). The strategy aims to beat the market and hence, succeed the higher profit than the market.

The trading strategy is usually formed by exploiting the technical analysis and the fundamental analysis. Fama (1965, 1970, 1991) suggested that there may be three types of the market: a weak-form, a semi-strong form, and a strong form of efficiency. The first assumes that the pattern in past price will repeat in the future. Therefore, the technical analysis that can identify the pattern is used, which is like moving averages and filter rules. The second presumes that the price will reflect the firm value in the future. Hence, the fundamental analysis that can extract the latent value from financial data is usually exploited, which is known as the ratio analysis. Unfortunately, it is not possible to test the last form as it is unknown whether there exist the strategies using the internal information of the firm of interest.

Whether the freight market is efficient is controversial in academic research. Some (Hale and Vanags, 1989; Veenstra, 1999; Kavussanos and Alizadeh, 2002a, 2002b, Alizadeh and Nomikos, 2006, 2007) have rejected the efficiency, while others (Hale and Vanags, 1992; Glen, 1997; Ådland and Koekebakker, 2004; Adland and Cullinane, 2005; Adland and Strandenes, 2006) have concluded mixed results. They have used the technical or fundamental indicators like Table [12].

Table 12 Type of Indicators from Technical and Fundamental Analysis

Type	Description		
Technical Analysis	Filter Rule	Bollinger Bands	$CP - (MA_n \pm k * SD_n)$
		MA envelopes	$CP - (MA_n \pm k\%)$
	Moving-Average Rule	Moving Average Convergence Divergence(MACD)	$MA_5 - MA_{20}$
		Relative Strength Index(RSI)	$RS = \frac{\frac{1}{n} \sum_i^n UMP_i}{\frac{1}{n} \sum_i^n DMP_i}$
	Momentum and Oscillator		$RSI = 100 - \frac{100}{(1 + RS)}$
		Stochastic Oscillator	$k(\%) = \frac{CP_x - LP_n}{HP_n - LP_n}$
			$MAOS(\%) = \left(\frac{1}{n} \sum_i k_i \right) * 100$
		Profitability	$ROE = \frac{Net\ Income}{Avg\ Equity}$
			$ROA = \frac{Net\ Income}{TTL\ Assets}$
	Fundamental Analysis	Financial Ratio Analysis	Liquidity
Activity			$\frac{Asset\ turnover}{Net\ Sales} = \frac{Net\ Sales}{TTL\ Assets}$
		Leverage	$\frac{Debt\ ratio}{TTL\ Liabilities} = \frac{TTL\ Liabilities}{TTL\ Assets}$
Economic Analysis		Price-Earnings ratio	$PE = \frac{Price}{Earnings}$

Since analyzing the past behavior of the underlying assets is not enough for predicting the future price, Alizadeh and Nomikos (2007) used the price-earnings (P/E) ratio as the fundamental indicator for making up for the drawback of the technical indicators. The P/E ratio has proved to be the economic indicator for investment and divestment timing of stocks in financial markets (Fama and French, 1992; Campbell and Shiller, 1998).

This paper also uses the same indicators in order to overcome the lack of theoretical background in picking the variables. Additionally, this focuses on the point that the chartering-out strategies based on the ML methods can beat the returns of the shipping market.

4.2.1 Benchmark: Multinomial Logistic Regression (MLR)

The MLR is the expanded binary logistic regression (BLR), which can classify more than three multiple outcomes.

Before explaining the BLR, the definition of odds needs to be known. Since the dependent variables of the BLR are categorical, they should be dealt with the probability ratio and logarithm in order to make the discrete variables continuous.

This can be described as follows:

$$\ln\left(\frac{p_i(y_i = 1)}{1 - p_i(y_i = 1)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im}$$

where p_i is the probability that y_i has certain categorical value.

The final function can be derived,

$$\ln\left(\frac{p_i(y_i = 1)}{1 - p_i(y_i = 1)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} = z_i$$

$$\frac{p_i(y_i = 1)}{1 - p_i(y_i = 1)} = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im}} = e^{z_i}$$

$$p_i(y_i = 1) = (1 - p_i(y_i = 1))e^{z_i}$$

$$p_i(y_i = 1)(1 + e^{z_i}) = e^{z_i}$$

$$p_i(y_i = 1) = \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{1 + e^{-z_i}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im})}}$$

The coefficients of the BLR can be estimated by Maximum Likelihood Estimation (MLE) that maximize the likelihood of given observations.

$$prob_i(y_i | x_{i1}, x_{i2}, \dots, x_{im}) = \{p_i^{y_i} (1 - p_i)^{(1-y_i)}\}$$

Assume all the samples are independent, the likelihood function is taken.

$$L = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{(1-y_i)}$$

Adopt p_i into this function,

$$L = \prod_{i=1}^n \left(\frac{e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})}} \right)^{y_i} \left(\frac{1}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})}} \right)^{(1-y_i)}$$

Take the logarithm both sides of above function,

$$\begin{aligned} \ln L &= \sum_{i=1}^n [y_i \{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) - \ln(1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})})\} \\ &\quad + (1 - y_i) \{ \ln(1) - \ln(1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})}) \}] \\ &= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) - \ln(1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})})] \end{aligned}$$

In order to estimates the coefficients that maximize the probability, the function is partially differentiated with respect to each parameter β_k .

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{i=1}^n \left(y_i - \frac{e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im})}} \right) x_{ik} = 0$$

When there are several classes, more than three, the BLR is extended to MLR that can classify multiple categorical variables.

$$\ln\left(\frac{p_{ij}}{p_{i0}}\right) = \sum_{m=0}^M \beta_{jm} x_{im} \quad , (x_{i0} = 1)$$

$$\frac{p_{ij}}{p_{i0}} = \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)$$

where p_{ij} is the probability that i th sample is in j th dependent category.

By using $\sum_{j=1}^J p_{ij} = 1$ and $\sum_{j=1}^{J-1} p_{ij} = 1 - p_{iJ}$,

$$\frac{\sum_{j=1}^{J-1} p_{ij}}{p_{iJ}} = \frac{1 - p_{iJ}}{p_{iJ}} = \sum_{j=1}^{J-1} \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)$$

$$p_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)}$$

$$p_{ij} = \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right) p_{iJ} = \frac{\exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)}{1 + \sum_{j=1}^{J-1} \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)}$$

If among the elements of the denominators, 1 is manipulated to the ratio $\frac{p_{ij}}{p_{iJ}} =$

$$\frac{p_{i(y_i=J)}}{p_{i(y_i=J)}} = \exp\left(\sum_{m=0}^M \beta_{Jm} x_{im}\right),$$

$$p_{ij} = \frac{\exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)}{\exp\left(\sum_{m=0}^M \beta_{Jm} x_{im}\right) + \sum_{j=1}^{J-1} \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)} = \frac{\exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)}{\sum_{j=1}^J \exp\left(\sum_{m=0}^M \beta_{jm} x_{im}\right)}$$

As the similar methods to estimate the coefficients of BLR, MLE is utilized for the approximation of the parameters of MLR. To derivate the likelihood function, it is needed to incorporate the dummy variables, g_{ij} , where i th observation belongs to j th class and then, $\sum_{j=1}^J g_{ij} = g_{i1} + \dots + g_{iJ} = 1$.

The likelihood function is shown as

$$L = \prod_{i=1}^n [p_{i1}^{g_{i1}} \times p_{i2}^{g_{i2}} \times \dots \times p_{ij}^{g_{ij}}].$$

If taking natural logarithm,

$$\begin{aligned} \ln L &= \ln \left(\prod_{i=1}^n [p_{i1}^{g_{i1}} \times p_{i2}^{g_{i2}} \times \dots \times p_{ij}^{g_{ij}}] \right) = \sum_{i=1}^n [g_{i1} \ln(p_{i1}) + g_{i2} \ln(p_{i2}) + \\ &\dots + g_{ij} \ln(p_{ij})] = \sum_{i=1}^n \sum_{j=1}^J [g_{ij} \ln(p_{ij})]. \end{aligned}$$

To estimate the coefficients of MLR,

$$\frac{\partial \ln L}{\partial \beta_{mj}} = 0, \quad m = 0, 1, \dots, M, j = 1, 2, \dots, J - 1.$$

4.2.2 Artificial Neural Networks (ANN)

ANN for classification has a different feature from the one for regression. The major distinction is the transfer function in the nodes of output layers. For the classification, it is better for ANN to harness the sigmoidal function that it can be continuously and easily differentiable.

4.2.3 Support Vector Machines (SVM)

SVM was originally invented for the linear classification, then it was extended to the nonlinear case and regression-related problems.

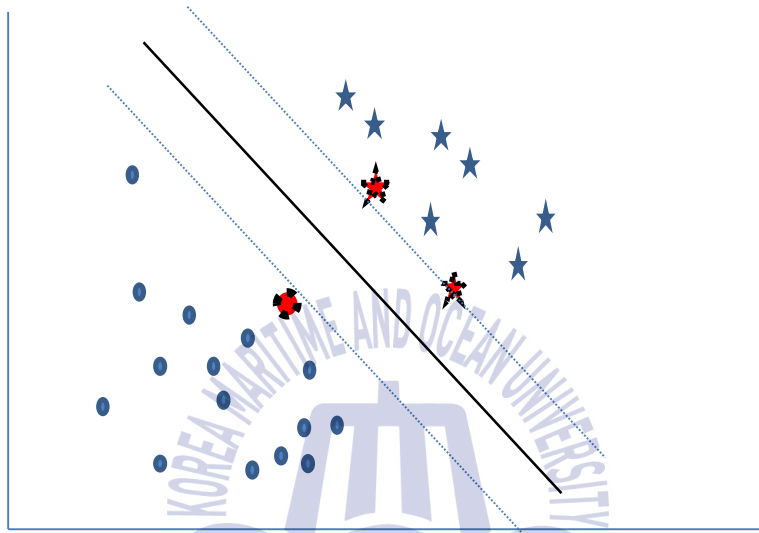


Figure 9 Hard-margin SVM

In the linearly separable case as Figure [9], the optimal boundary is the black line that has the equal distance between two classes. Since the red vectors are supporting to make an optimal boundary only, this method is called ‘support vector machines’.

The optimal boundary is $f(x) = w^T x + b = 0$ and a certain class is $f(x) = \text{sign}(w^T x + b)$. This optimal boundary is also called ‘optimal hyperplane’ in a case where the feature space is in high-dimension. The distance between both dot lines is called ‘margin,’ and this margin has $\frac{2}{\|w\|}$. Under the constraints $y_i(w^T x_i + b) \geq 1$, optimizing the hyperplane is to maximize the margin or to minimize the reverse of it, $\frac{1}{2} \|w\|$. The norm $\|w\|$ presents quadratic terms $w^T w$. This case can be expressed mathematically as follows,

$$\begin{aligned} \min \quad & \frac{1}{2} w^T w \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1. \end{aligned}$$

This constrained convex optimization can be solved by using Lagrange multipliers.

$$\begin{aligned} \min \quad L(w, b, \alpha) &= \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i (y_i(w^T x_i + b) - 1) \\ L(w, b, \alpha) &= \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i y_i w^T x_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \end{aligned}$$

This function must satisfy the Karush-Kuhn-Tucker (KKT) conditions as follows,

$$\begin{aligned} - \frac{\partial L(w, b, \alpha)}{\partial w} &= 0 \quad \text{and} \quad \frac{\partial L(w, b, \alpha)}{\partial b} = 0 \\ - \alpha_i (y_i(w^T x_i + b) - 1) &= 0 \\ - y_i(w^T x_i + b) - 1 &\geq 0 \\ - \alpha_i &\geq 0. \end{aligned}$$

For taking the derivative of $L(w, b, \alpha)$ with respect to w and b respectively

$$\begin{aligned} \frac{\partial L(w, b, \alpha)}{\partial w} &= w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i \\ \frac{\partial L(w, b, \alpha)}{\partial b} &= - \sum_{i=1}^n \alpha_i y_i = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned}$$

Substitute these conditions into $L(w, b, \alpha)$,

$$\begin{aligned} L(w, b, \alpha) &= \frac{1}{2} \sum_{i=1}^n \alpha_i y_i x_i \cdot \sum_{i=1}^n \alpha_i y_i x_i - \sum_{i=1}^n \alpha_i y_i x_i \cdot \sum_{i=1}^n \alpha_i y_i x_i + \sum_{i=1}^n \alpha_i \\ L(w, b, \alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j. \end{aligned}$$

This $L(w, b, \alpha)$ can be changed to $L(\alpha)$. Then, during the training phase, it is tried to find α that minimize $L(w, b, \alpha)$.

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \\ \text{s. t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \end{aligned}$$

So far, the SVM with a hard margin was explained. However, there can be many cases where the classes are non-separable linearly as Figure [10].

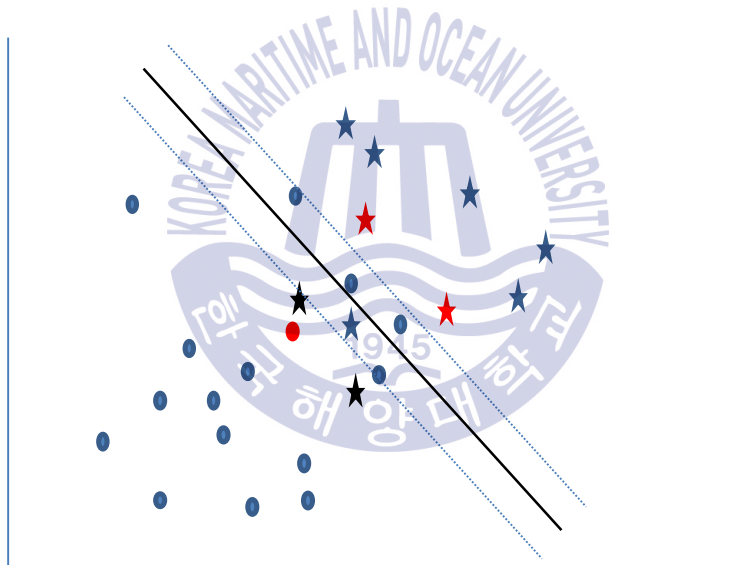


Figure 10 Soft-margined SVM

Given the dataset where the hard-margined SVM cannot be applied, introducing the slack variable is needed to relax the constraints. This is called ‘soft-margined SVM’ that can be derived from hard-margined SVM as below.

$$\begin{aligned} \min \quad & \frac{1}{2} w^T w \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1 \end{aligned}$$

When introducing the slack variable into the constraints, the objective function should tolerate the penalty term.

$$\begin{aligned} \min \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$

The relationship between the width of margin and the penalty C is trading-off. In a similar way of hard margined case,

$$\begin{aligned} \min \quad L(w, b, \alpha, \beta) = & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\ & - \sum_{i=1}^n \alpha_i (y_i(w^T x_i + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i \end{aligned}$$

$$\begin{aligned} L(w, b, \alpha, \beta) = & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i y_i w^T x_i - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i - \\ & \sum_{i=1}^n \alpha_i \xi_i - \sum_{i=1}^n \beta_i \xi_i. \end{aligned}$$

The KKT conditions are given as,

$$\begin{aligned} - \frac{\partial L(w, b, \alpha, \beta)}{\partial w} = 0 \quad , \quad \frac{\partial L(w, b, \alpha, \beta)}{\partial b} = 0 \quad , \quad \text{and} \quad \frac{\partial L(w, b, \alpha, \beta)}{\partial \xi} = 0 \\ - \alpha_i (y_i(w^T x_i + b) - 1 + \xi_i) = 0 \\ - y_i(w^T x_i + b) - 1 + \xi_i \geq 0 \end{aligned}$$

$$- \alpha_i \geq 0 \text{ and } \beta_i \geq 0.$$

For taking the derivative of $L(w, b, \alpha, \beta)$ with respect to w , b , and ξ respectively,

$$\frac{\partial L(w, b, \alpha, \beta)}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i$$

$$\frac{\partial L(w, b, \alpha, \beta)}{\partial b} = - \sum_{i=1}^n \alpha_i y_i = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0$$

$$\frac{\partial L(w, b, \alpha, \beta)}{\partial \xi} = C - \alpha_i - \beta_i = 0.$$

Substitute these conditions into $L(w, b, \alpha, \beta)$.

$$L(w, b, \alpha, \beta) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i y_i w^T x_i + \sum_{i=1}^n \alpha_i$$

$$L(w, b, \alpha, \beta) = \frac{1}{2} \sum_{i=1}^n \alpha_i y_i x_i \cdot \sum_{i=1}^n \alpha_i y_i x_i - \sum_{i=1}^n \alpha_i y_i x_i \cdot \sum_{i=1}^n \alpha_i y_i x_i + \sum_{i=1}^n \alpha_i$$

This $L(w, b, \alpha, \beta)$ can be changed to $L(\alpha)$. Then, during the training phase, it is tried to find α that minimize $L(w, b, \alpha, \beta)$.

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j$$

$$\text{s. t. } \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C$$

$0 \leq \alpha_i \leq C$ is due to $C - \alpha_i - \beta_i = 0$. Because $\alpha_i \geq 0$ and $\beta_i \geq 0$, C should be greater than or equal to α_i . The soft-margined SVM that can linearly classify was presented so far.

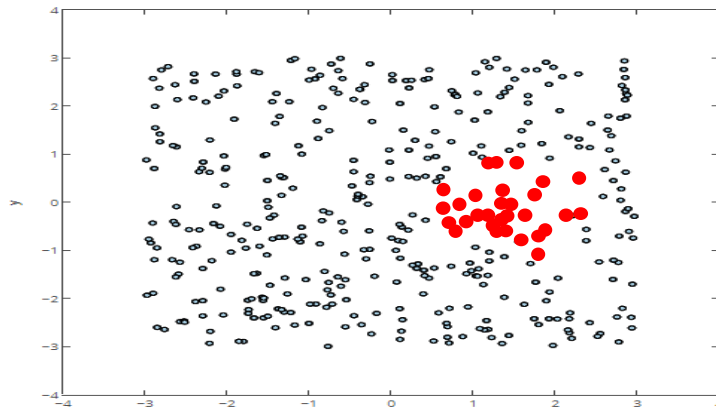


Figure 11 Non-linear data

Unfortunately, there exist non-linear cases in the real world as shown in Figure [11]. This data cannot be linearly separable through fore-mentioned methods. To solve the case SVM should be equipped with the kernel trick, $K(x, x') = \phi(x)^T \phi(x')$, that can transform the input features to higher dimensional feature space. Once the transformation is carried out, the hyperplane that classifies the non-linear dataset can be conveniently found as illustrated in Figure [12].

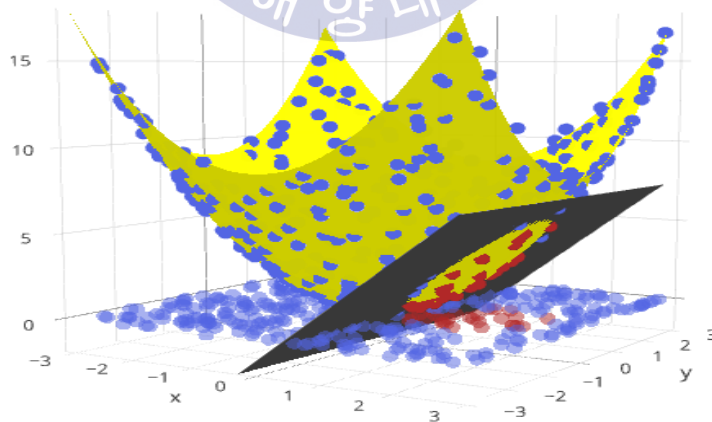


Figure 12 Transformation to Higher Dimensional Space⁴

⁴ <http://efavdb.com/svm-classification/>

The optimal hyperplane is $w^T \phi(x) + b = 0$. Furthermore, if $w = \sum_{i=1}^n \alpha_i y_i x_i$ is substituted into the hyperplane.

$$\sum_{i=1}^n \alpha_i y_i \phi(x_i)^T \phi(x) + b = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b = 0$$

Recall the optimal problem in soft margined SVM and then

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \rightarrow \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x) \\ \text{s. t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

4.2.4 Random Forest (RF)

Since the concept of RF is originated from the decision tree, the application of it is easily extended to the classification. After tuning the parameters of RF, dealing with the outcomes from trees are different from the regression. For classification, the called ‘voting’ is counting the number of the results from the constructed trees while RF for regression is to average the results from the trees. This procedure is confirmed in Part 4.1.2.3.

4.3 Performance Measures

This part presents the criterion of the model selection. To choose the best model with other candidates, or choose the parameters to achieve the optimal performance, the specific criteria for selection is needed. Although many performance measures have been presented in the literature, there has not been a consensus about which

measures are appropriate to evaluate the model performance. Hence, the most common methods are introduced as below Table [13].

Table 13 Type of Performance Measures

Measurements	Equations
MAE(Mean Absolute Error)	$\frac{1}{n} \sum_{i=1}^n x_i - \hat{x}_i $
RMSE(Root Mean Squared Error)	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}$
COR(Correlation Coefficients)	$r = \frac{Cov(x, \hat{x})}{\sigma_x \sigma_{\hat{x}}}$
R ²	r^2

· \hat{x} : Model estimation x : Actual observation, \bar{x} : Mean of actual observation. Cov() : Covariance, σ : Standard deviation.

MAE and RMSE are the scale-dependent measures. The former measures the magnitude of the deviation from the true value with disregard to its direction. The latter uses the square root of MSE to measure the forecast error. This measure can be more sensitive than other measures because RMSE imposes more weights on the larger error. Assume various candidate models produce equal MAE, the relationship between them can be expressed as follows.

$$MAE \leq RMSE \leq \sqrt{n}MAE$$

Moreover, if all individual errors in a model have an equal magnitude, two measures are the same.. Furthermore, the upper bound of RMSE becomes $\sqrt{n}MAE$

in an extreme case where only the one of the error set has $nMAE$ and the all of rest have zero. lthough RMSE is widely used in the literature, some misunderstanding or misinterpretation often had been found (Willmott and Matsuura, 2005).

RMSE can be affected by three factors, which are the variability within the distribution of error, \sqrt{n} , and MAE. Hence, Willmott and Matsuura (2005) suggested that it is more appropriate for MAE to be used for the performance measure.

The correlation coefficient is an alternative which shows a forecast verification. This measures can be interpreted as the strength of the linear relationship between the actual values and the estimates (Barnston, 1992). R^2 provides the model performance with the proportional variability. So far, it explained the types of the measures to be used in the regression problems.

For the classification, the mechanism of the measure is more straightforward. The most widely used score is the hit ratio. This counts the number of accurately matching the observations and the forecasts. So it is intuitively understandable and conceptually interpretable for a clumsy reader.

Chapter 5 Research Results

5.1 Valuation of T/C Extension Option

5.1.1 Data and Frameworks

Based on the backgrounds and the methodologies previously described, this chapter illustrates the detailed descriptions of the data and the blueprint of the model for two problems as mentioned before.

Figure [13] depicts the design of modeling the problems to be solved.

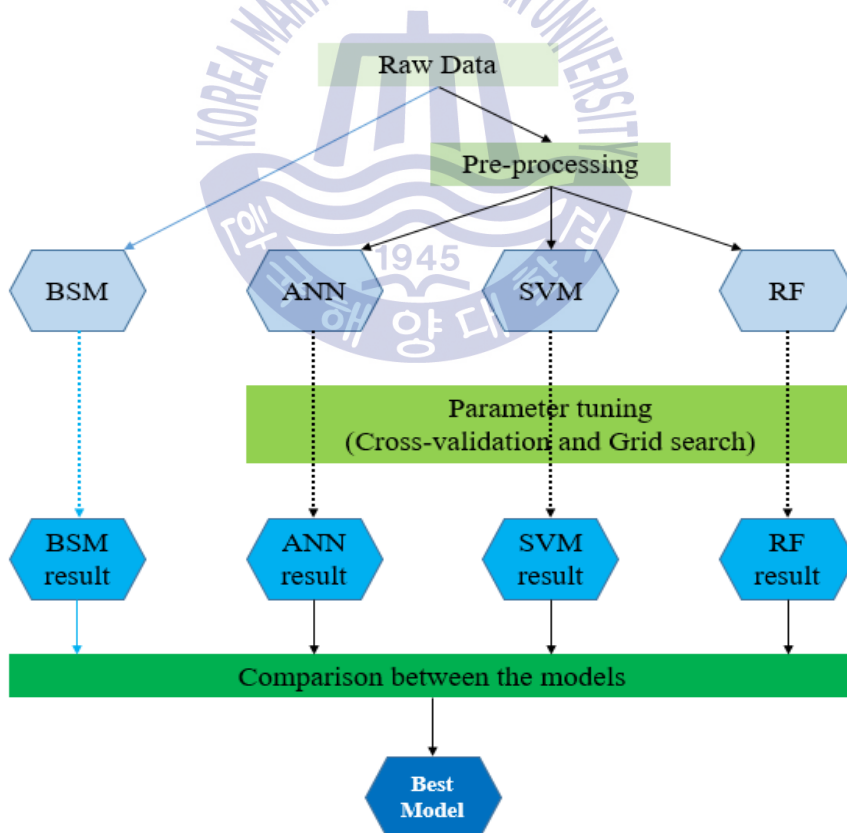


Figure 13 Flow chart of Pricing Options

The raw data used in this paper include four freight rates and US 90days treasury bond rate. The formers are downloaded from the Shipping Intelligence Network. Although the data having different periods as listed in Table[14] are taken, this fact does not impose a severe problem on this experiment because the primary purpose of the paper is to confirm the applicability of Machine Learning to the chartering practice. The first and last year of data is only used for producing the volatility of the freight rate and are then excluded from subsequent analysis.

Table 14 Descriptions of Data

	Size	Data Period	Observation
Capesize	172,000 dwt	2010-07-23 ~ 2016-11-18	331
Panamax	72,000 dwt	2002-03-01 ~ 2016-08-26	757
Supramax	52,000 dwt	2002-12-27 ~ 2016-04-01	693
Handymax	45,000 dwt	2002-03-01 ~ 2013-06-28	592

More precisely, each freight rate is classified into the time charter trip(tct), 3-month time charter(3m), and 1-year time charter(1yr). Among them, 3m is artificially generated by linearly interpolating tct and 6-month time charter rate in order to treat as the underlying assets. The rest is obtained from US Federal Reserve Bank of St. Louis (<https://fred.stlouisfed.org/series/TB3MS>). The frequency of all data is on weekly period. Table [14] shows the descriptive analysis of the data.

The statistic of the Jarque-Bera presents whether the distribution of the data satisfies the normality. The calculated statistic reveals that all of them are a non-satisfactory condition of the normal distribution. Additionally, the typical test for the existence of the unit root in particular series of interest is the augmented Dickey-Fuller (ADF) method that can analyze the stationarity.

Table 15 Descriptive Analysis of Freight Rates and the US 3 month T-bill

Statistics	Ctct	C3m	C1yr	Ptct	P3m	P1yr	Stct	S3m	S1yr	Htct	H3m	H1yr	rf
Mean	12301.0	14243.1	15499.6	21234.6	22602.1	21762.5	20503.4	21214.4	20308.6	20058.0	20460.5	19203.0	0.012
Median	8737.5	11187.5	13500.0	15250.0	16218.8	15250.0	15812.5	16093.8	15250.0	16375.0	16515.6	14750.0	0.003
Maximum	45125.0	42312.5	36500.0	93193.5	93846.8	82000.0	73125.0	73312.5	70000.0	65250.0	67937.5	60000.0	0.050
Minimum	1562.5	2668.8	4975.0	2350.0	3500.0	4750.0	2400.0	3075.0	4250.0	3437.5	4843.8	6750.0	0.000
Std. Dev.	9735.8	8704.4	7023.1	18308.5	18632.6	17434.3	14576.1	14790.0	14237.7	13034.4	13338.0	12454.7	0.016
Skewness	1.4	1.2	0.8	1.6	1.6	1.8	1.4	1.5	1.7	1.3	1.4	1.5	1.228
Kurtosis	4.1	3.8	3.0	5.4	5.4	5.8	4.7	4.8	5.2	4.3	4.5	4.4	3.133
Jarque-Bera ⁵	118.4	86.0	34.5	499.7	521.4	659.2	305.3	347.7	463.6	212.8	243.7	263.6	191.1
	0	0	0	0	0	0	0	0	0	0	0	0	0
Level	-4.69	-3.89	-2.07	-2.81	-2.98	-2.91	-2.86	-2.84	-2.71	-2.47	-2.24	-2.16	-1.92
ADF	0.00	0.01	0.56	-0.19	-0.14	-0.16	0.18	0.18	0.23	0.34	0.47	0.51	-0.64
test	-10.5***	-12.9***	-5.7***	-12.5***	-11.5***	-9.1***	-13.9***	-14.3***	-11.7***	-13.3***	-13.7***	-14.8***	-4.2***
1 st Diff	0	0	0	0	0	0	0	0	0	0	0	0	-0.004
Observations	331	331	331	757	757	757	693	693	693	592	592	592	757

⁵ $\frac{n}{6} \left[Sk^2 + \frac{(Ku-3)^2}{4} \right]$, where n is the number of observations, Sk is skewness, and Ku is kurtosis.

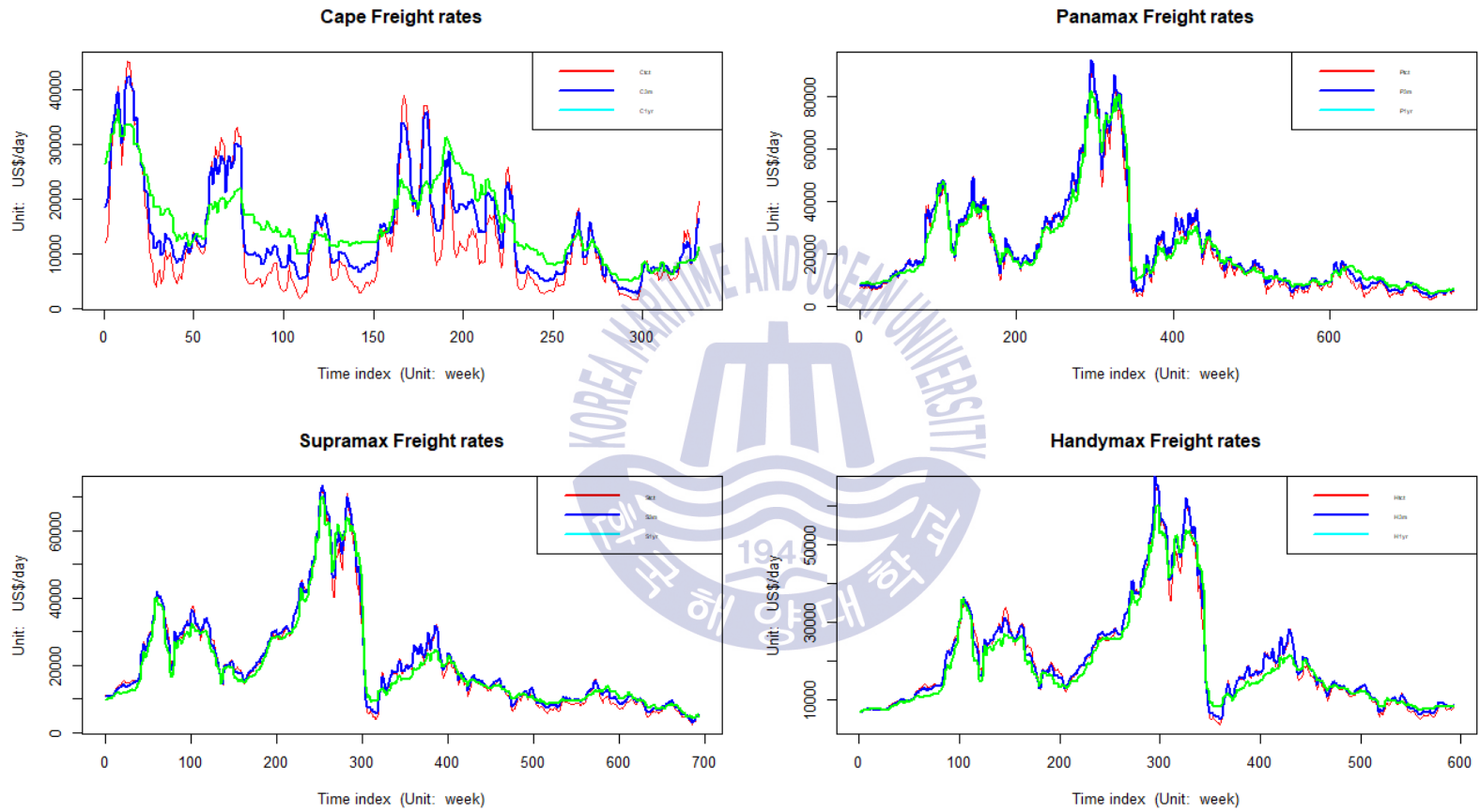


Figure 14 Bulk Freight Rate

Since the 1% critical value for the test statistic is -3.97, it then rejects the null hypothesis that the series is nonstationary in the logarithmic first difference. That is, the series does not have stationarity.

Furthermore, Figure [14] and [15] provides the exploration of the data with the charts.

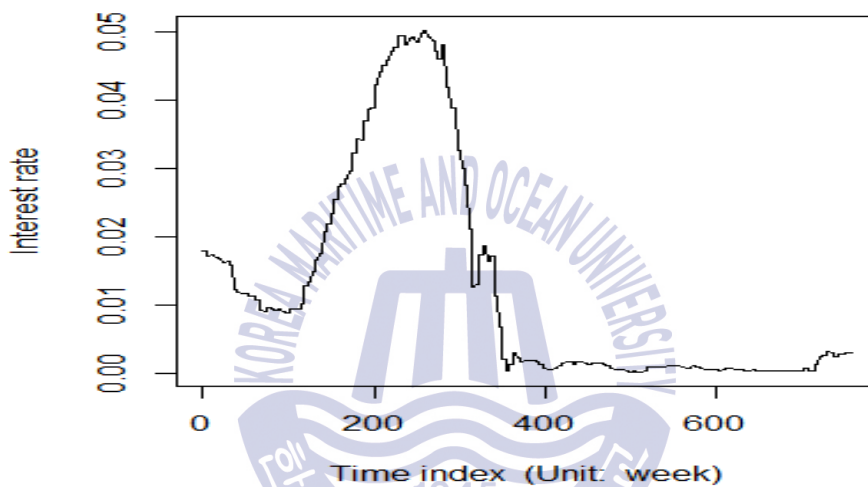


Figure 15 US 3-month T-bill rate

5.1.2 BSM Modeling

For applying BSM to pricing the T/C option, there should be pre-assumptions that bound the mechanism of the T/C option pricing. Yun et al. (2017) supposed; that redelivery flexibility is ignored; that the time lag between the contract and the delivery of the vessel does not occur; that the prices between the option period and the firm period are equal; that the exercise of the option is only limited at maturity; and that when exercising, the payoff of the option is based on a 3-month T/C rate at maturity.

The BSM needs five input variables such as spot price, strike price, time to maturity, risk-free rate, and spot return volatility to value the T/C option. The only unknown component is the volatility of the underlying asset. This paper uses the return of 3-month T/C rate for one year to yield the equally-weighted historical volatility as shown in Figure [16].

The Table [16] shows the input variables of BSM.

Table 16 Input Variables of BSM

	Variables	Data
S	Underlying price of the underlying asset	3m-T/C rate
X	Strike price	1yr-T/C rate
R_f	Risk-free rate	90d T-bill
σ	Volatility of return for S	1yr-SD of Spot rate
T	Maturity	1 year

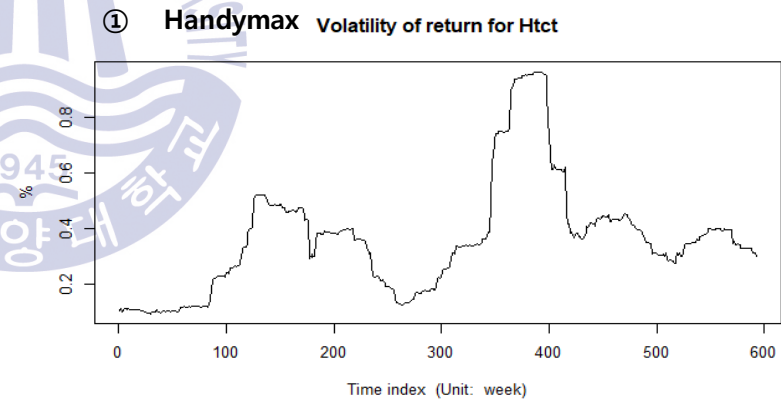
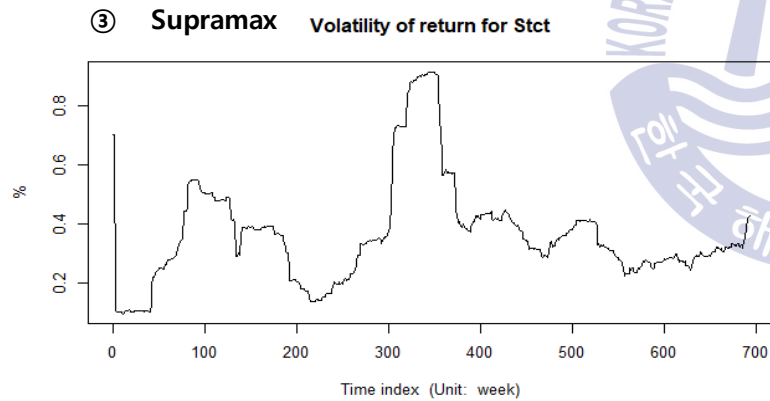
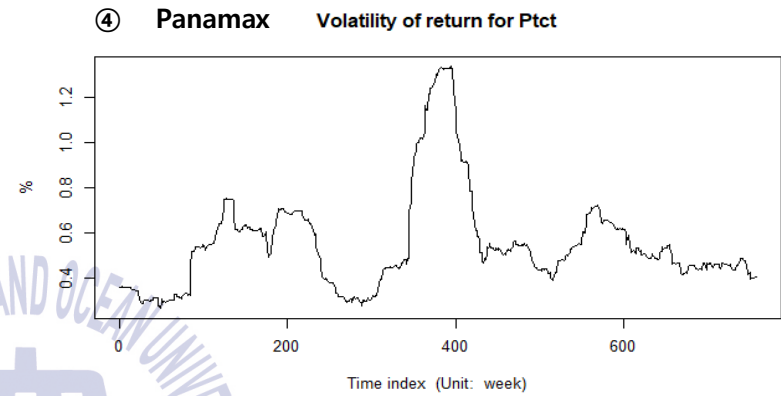
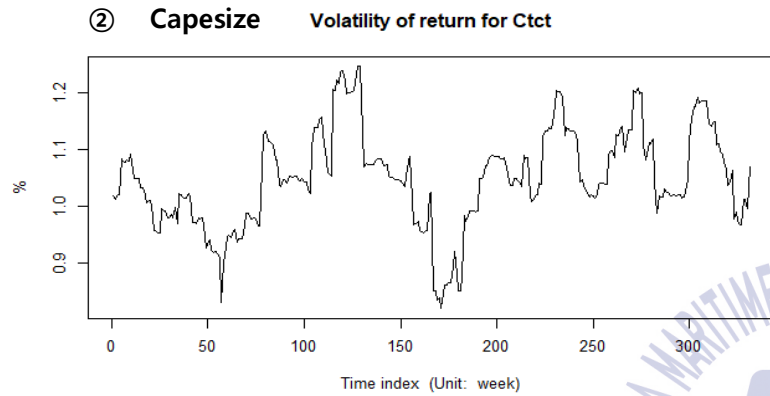


Figure 16 Volatility of Return for the Underlying Asset(tct)

5.1.3 ANN Modeling

Although the merit of the model is that the assumptions over the variables are not needed (Smith and Gupta, 2000; Zhang, 2003; Roh, 2007; Kristjanpoller and Minutolo, 2015), note that the results lack the economic background because of the learning process concealed under its ‘black box’ (Roh 2007). To circumvent the drawback of the ML techniques, the variables to be utilized are the same as the ones of BSM except for the time to maturity. Instead of the time to maturity, the spot rate is randomly added for capturing the market dynamics.

Despite the fact that the learning methods have an exceptional prediction power, they are not without drawbacks. For ANN, since it needs enough data size to fit the model, the training algorithm is relatively sluggish, and during the process of learning, the “over-fitting” often arise. Therefore, fitting the model has to be carefully carried out. In particular, the crucial parameters to be determined are involved with how many inputs, hidden layers, and hidden nodes are needed. This point is strongly associated with the model selection. To achieve successfully tuning the parameters the n -fold cross-validation technique and the grid search tools are applied. In addition to this, as shown in the body of the literature, one hidden layer is theoretically sufficient for approximating the non-linear functions (Kaastra and Boyd, 1996; Zhang, Patuwo and Hu, 1998; Basheer and Hajmeer, 2000; Fadlalla and Lin, 2001; Atsalakis and Valavanis, 2009).

For preprocessing, all the data at hand undergo the scaling called normalization, and they are randomly scattered with disregard to time sequence as it satisfies statistical assumption that the distribution of the return in any point of data is constant (Yun, Lim and Lee, 2017). This randomized distribution of data is depicted in Figure [17].

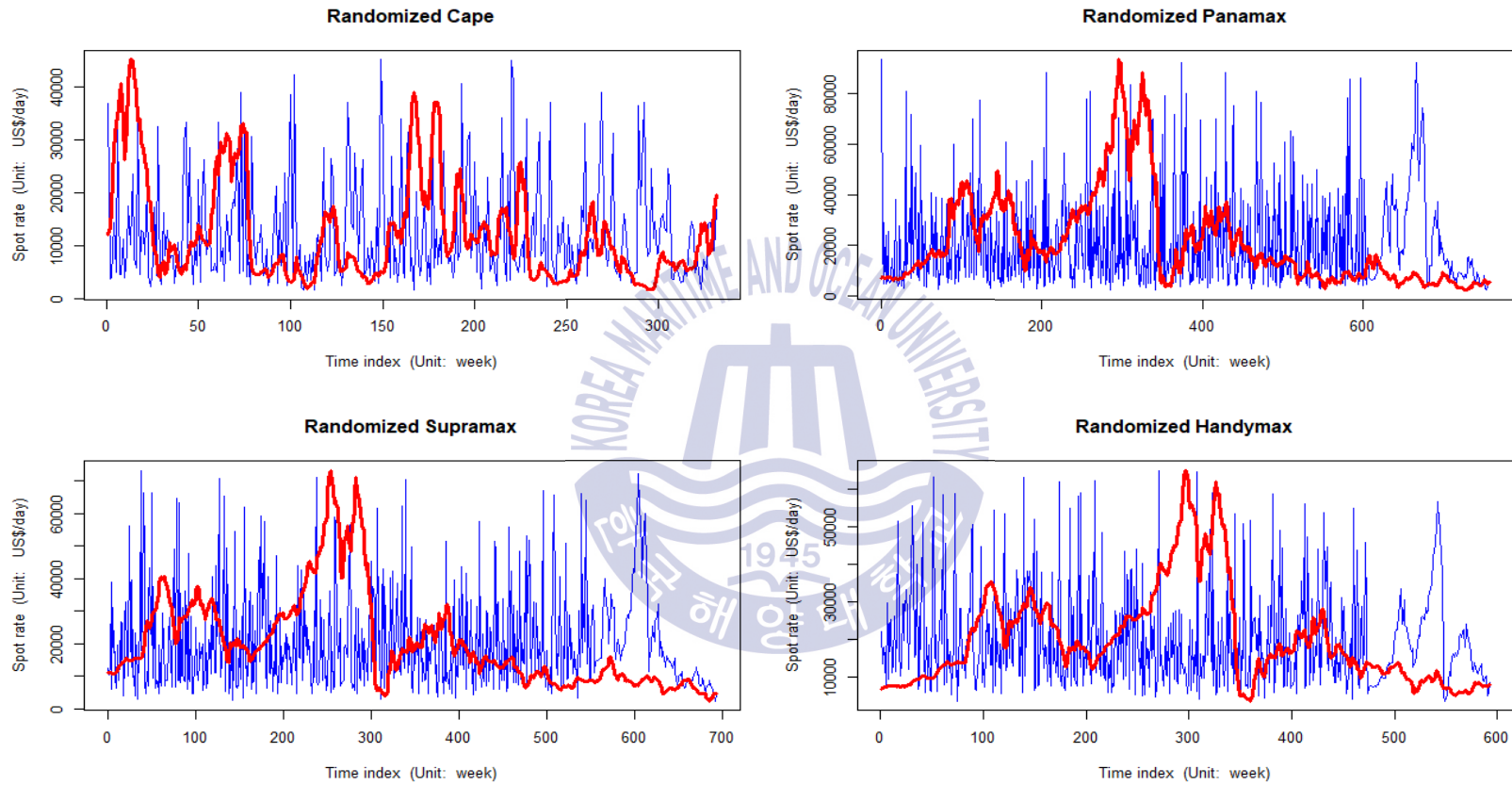


Figure 17 Original and Randomized Series of Freight Rates

Subsequently, the data is split into the training set and the test set according to the ratio of 80:20. Then, through the 10-folds cross-validation technique in the training set, the best parameters can be found that the weight decay and the number of the hidden neurons are shown in Table [17] and Figure [18]. According to the reviewed literature (Kaastra and Boyd, 1996; Zhang, Patuwo and Hu, 1998; Basheer and Hajmeer, 2000; Fadlalla and Lin, 2001; Atsalakis and Valavanis, 2009), they revealed that one hidden layer is enough for most problems.

Table 17 Optimal Parameters of ANN

Size	No. of Hidden nodes	Decay
Capesize	18	0.01
Panamax	14	0.01
Supramax	17	0.01
Handymax	13	0.01

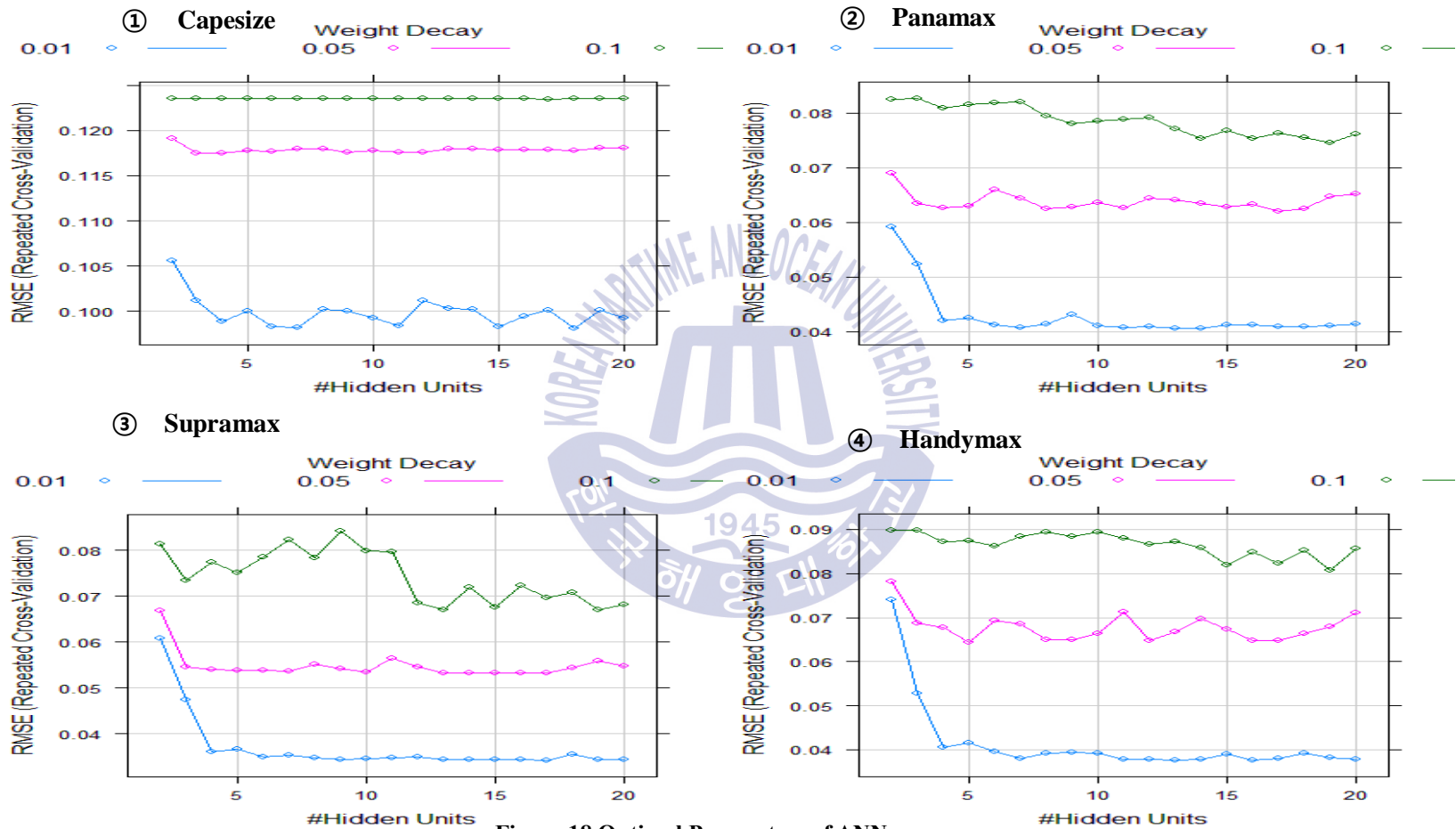


Figure 18 Optimal Parameters of ANN

5.1.4 SVR Modeling

For SVR, the usage of the kernel functions makes SVR a more powerful model than others. Furthermore, as mentioned before, the model pursues the structural risk principle different from others and many researchers pointed out that the advantages of SVR are the global optimality. However, the excellent performance of this model is highly dependent on choosing the cost and sigma parameters and the kernel function.

In order to discover the best fitting parameters, the cost, C , and sigma proved to be depicted in Table [18] and Figure [19] individually by using cross-validation. The kernel function to be used is the radial basis kernel (called Gaussian Kernel) that is mostly picked up in the literature.

Table 18 Optimal Parameters of SVR

Size	Sigma	C
Capesize	1	10
Panamax	1	20
Supramax	1	100
Handymax	1	50

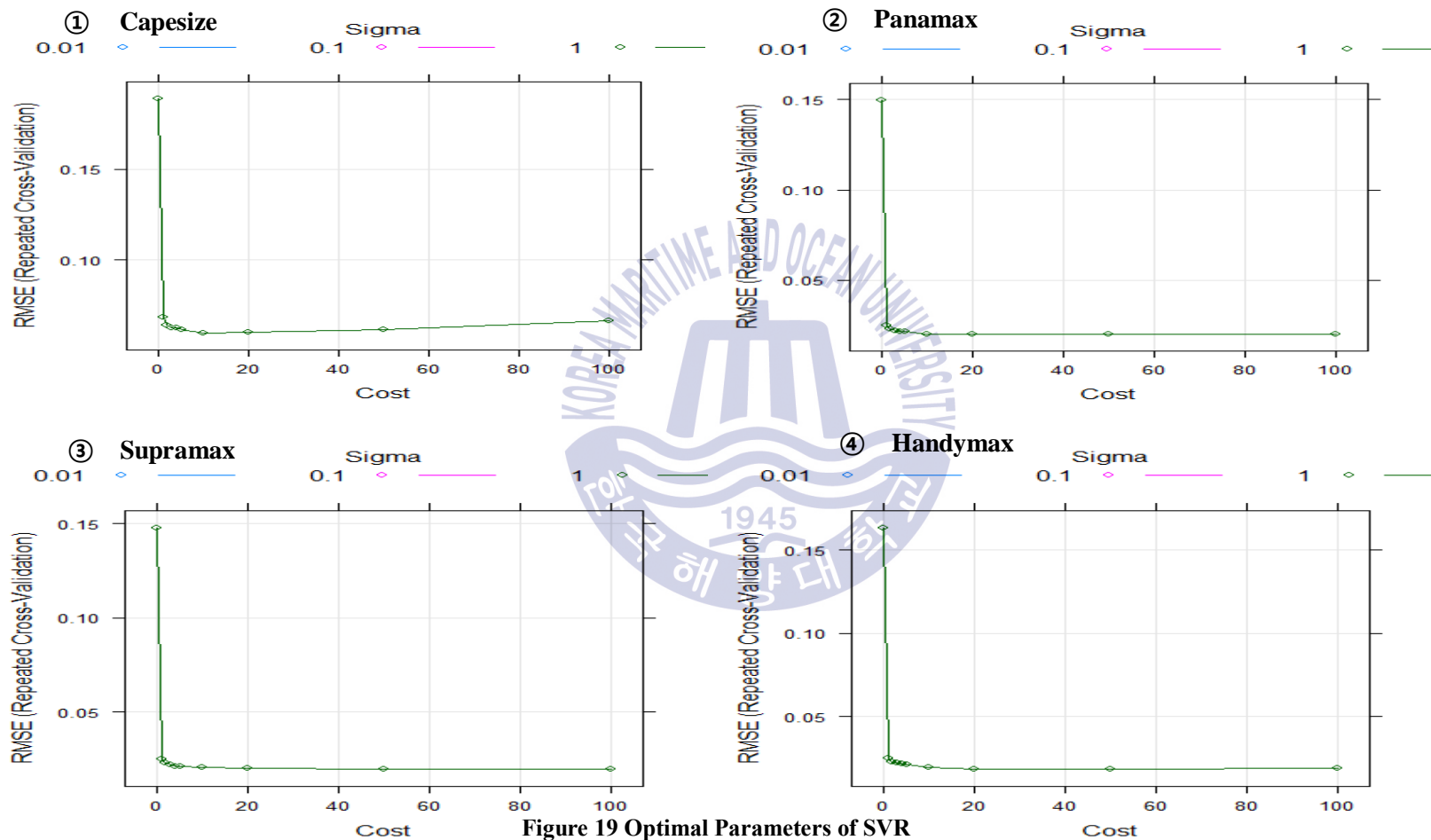


Figure 19 Optimal Parameters of SVR

5.1.5 RF Modeling

The RF model is the derivative of the decision tree, especially of the classification and the regression tree (CART). The main issue of the tree-based model is how to treat the overfitting problem as dissolved. It uses the stopping rule or the pruning for preventing overly fitted model to the in-sample data. Fortunately, Breiman (2001) devised the RF model with the similar mechanism of the bagging technique. The distinction between the RF and the bagging is that the bootstrapped trees generated in the learning process are de-correlated from each other as it randomly picks m predictors instead of the full number of p predictors. Typically, the parameter m is approximated to \sqrt{p} .

The appropriate number of m and trees are presented in Table [19] and Figure [20] respectively. The RF model can identify the essential variables that improve prediction accuracy. However, since the variable importance are not needed in this paper, this information is not given.

Table 19 Optimal Parameters of RF

Size	m	No. of trees
Capesize	5	500
Panamax	4	500
Supramax	3	500
Handymax	4	500

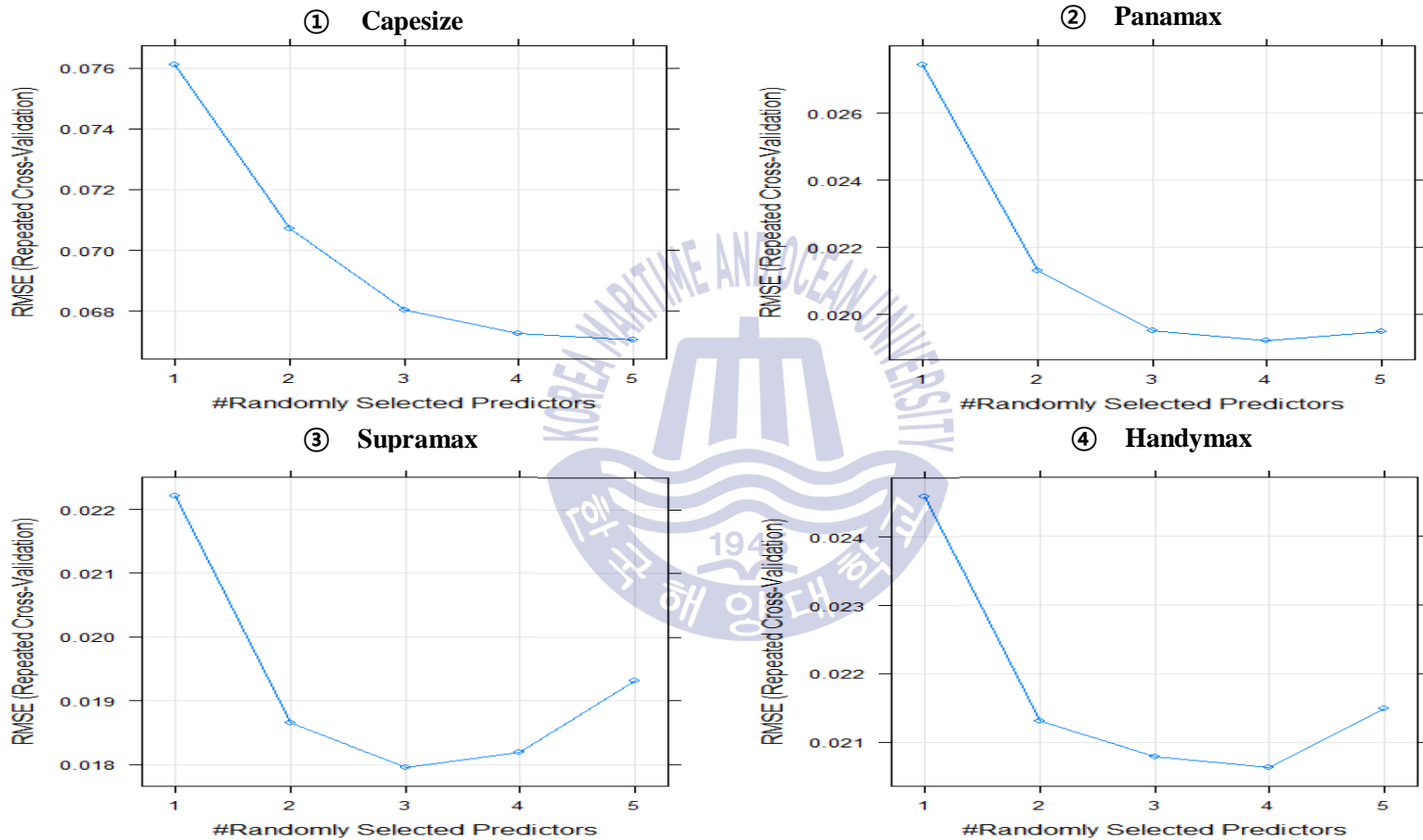


Figure 20 Optimal m of RF

5.1.6 Results and Discussion

Given the parameters tuned, each model is applied to the test sample respectively. The outcomes of the experiments are depicted in Figure [21] to Figure [24]. As illustrated in these figures, it is confirmed that the ML models far outweigh the conventional model. The following table [20] and [21] summarize the results based on the performance measures.

Table 20 Summary of Option pricing(Left:Capesize, Right:Panamax)

Measures	BSM	ANN	SVM	RF	Measures	BSM	ANN	SVM	RF
MAE	4958.2	222.4	142.6	145.2	MAE	2247.8	379.9	228.6	178.3
RMSE	6356.3	495.3	385.0	342.3	RMSE	2967.8	780.9	510.8	409.2
COR	-0.107	0.810	0.900	0.923	COR	-0.068	0.957	0.982	0.989
R ²	0.011	0.657	0.810	0.853	R ²	0.005	0.916	0.964	0.977

Table 21 Summary of Option pricing(Left:Supramax, Right:Handymax)

Measures	BSM	ANN	SVM	RF	Measures	BSM	ANN	SVM	RF
MAE	3091.1	277.3	130.1	108.7	MAE	3255.7	483.9	198.9	143.6
RMSE	4027.3	637.9	268.1	276.2	RMSE	3977.0	775.2	295.8	235.0
COR	0.05	0.96	0.99	0.99	COR	-0.26	0.94	0.99	1.00
R ²	0.002	0.920	0.986	0.985	R ²	0.066	0.888	0.986	0.991

On the basis of all criteria, the candidate models achieve the fair predictions rather than the benchmark model. In particular, SVM and RF as the newly introduced models in this paper appear to be promising. The best model in terms of the resultant performances is in the order of RF, SVM, ANN, and BSM.

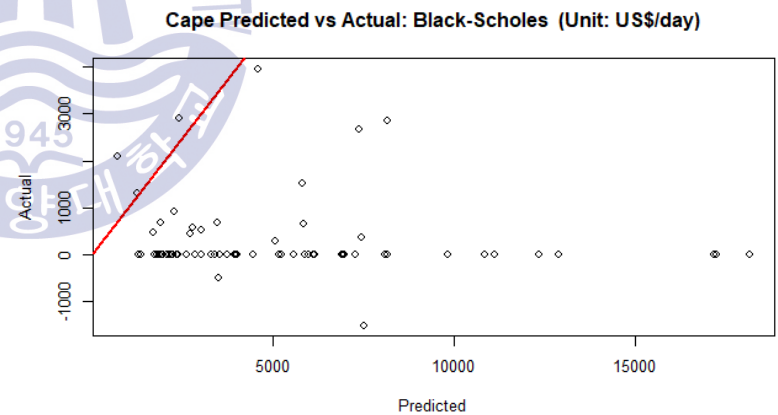
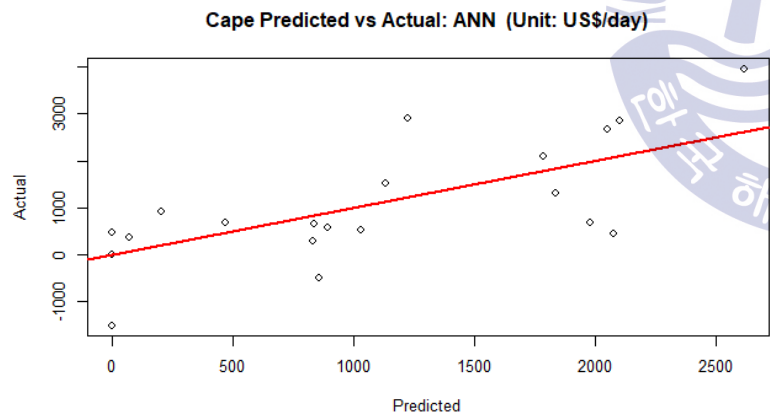
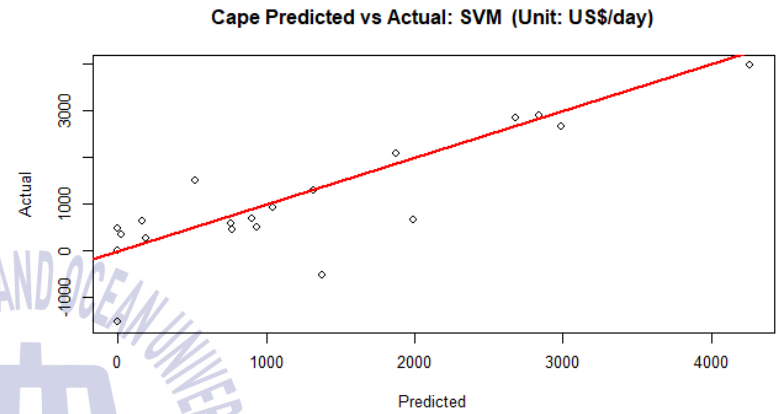
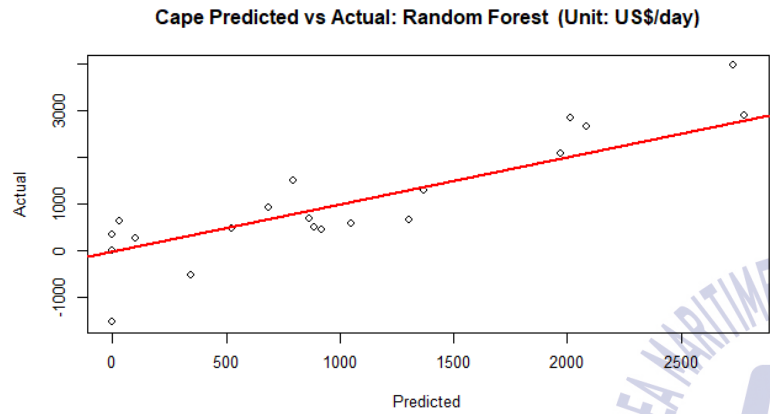


Figure 21 Comparison of T/C option prices: Capesize

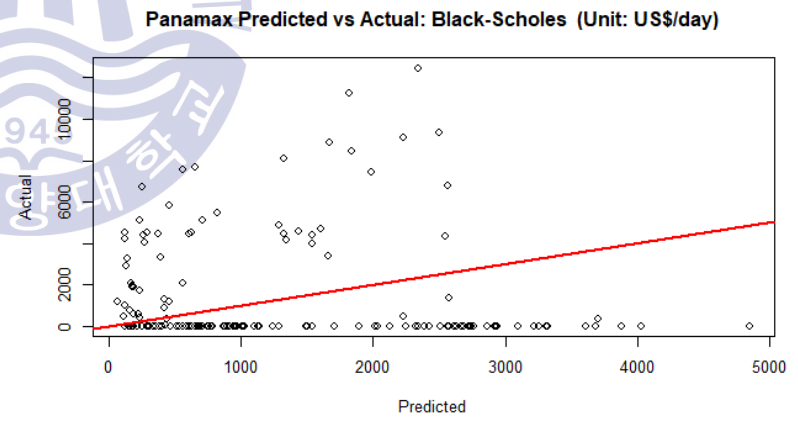
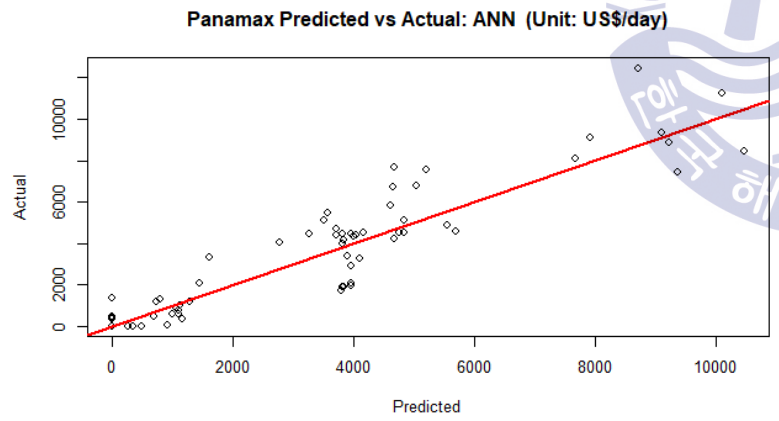
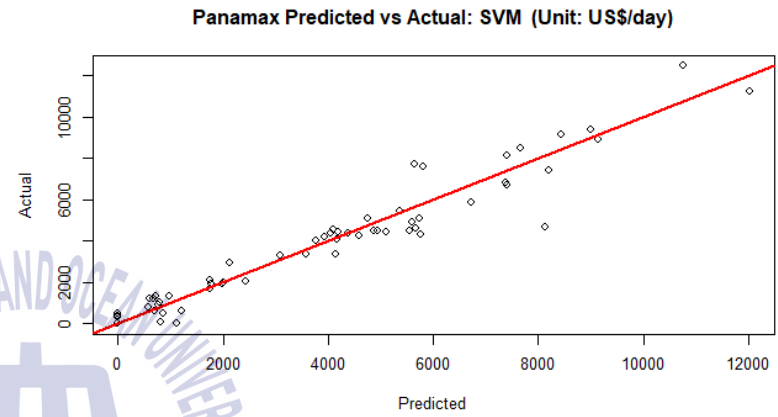
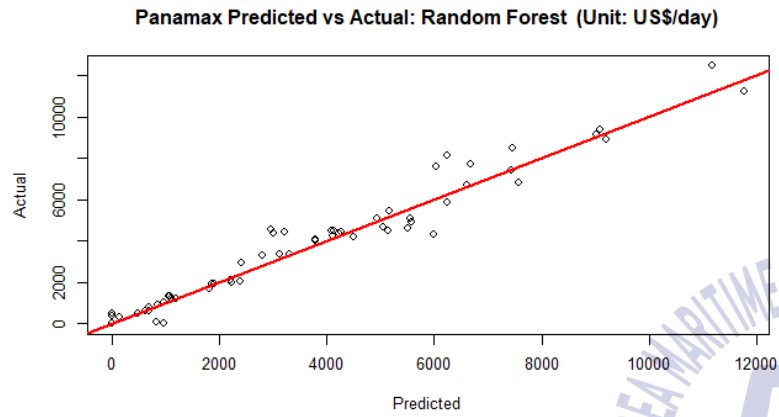


Figure 22 Comparison of T/C option prices: Panamax

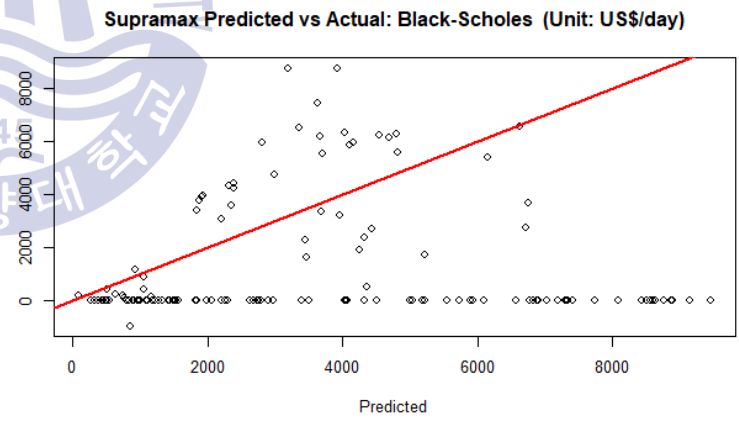
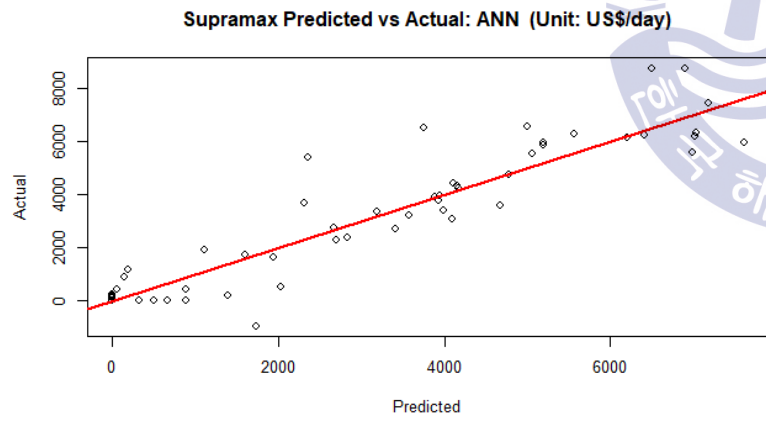
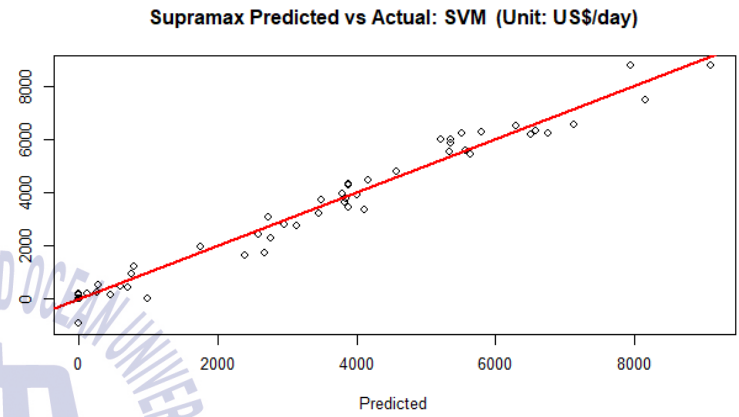
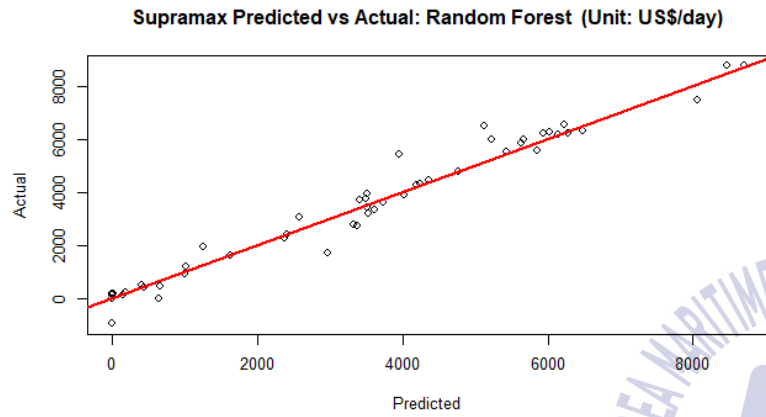


Figure 23 Comparison of T/C option prices: Supramax

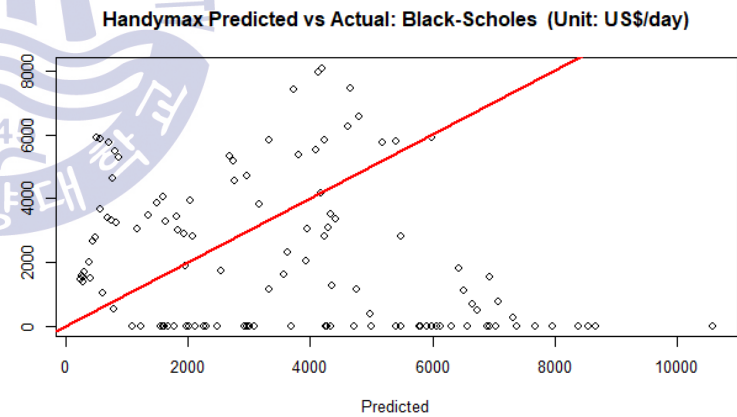
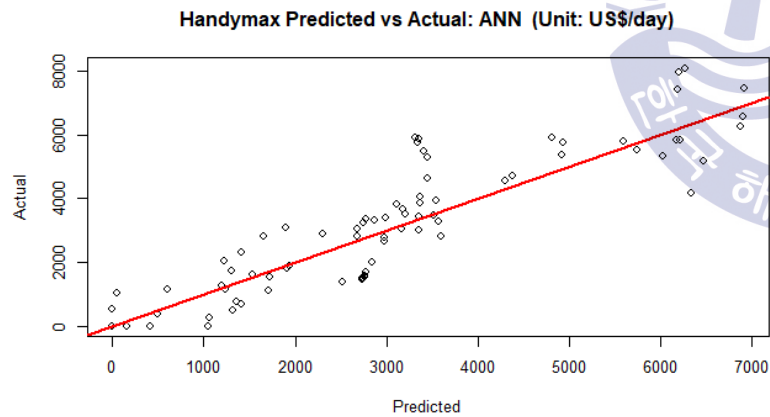
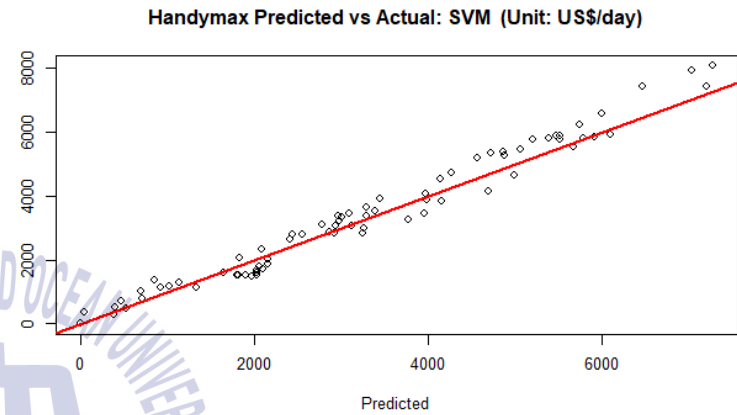
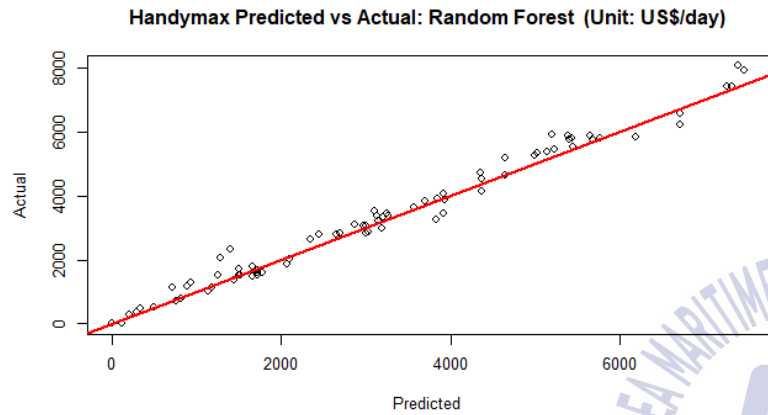


Figure 24 Comparison of T/C option prices: Handymax

5.2 Decision for Chartering-out Strategies

5.2.1 Data and Frameworks

This chapter deals with solving how long the contract in chartering is chosen. Since the type of this problem is the classification, the logistic regression is treated as the benchmark model, and the candidate models are used in the classifier model. Figure [25] exhibits the overall setting of the experiment.

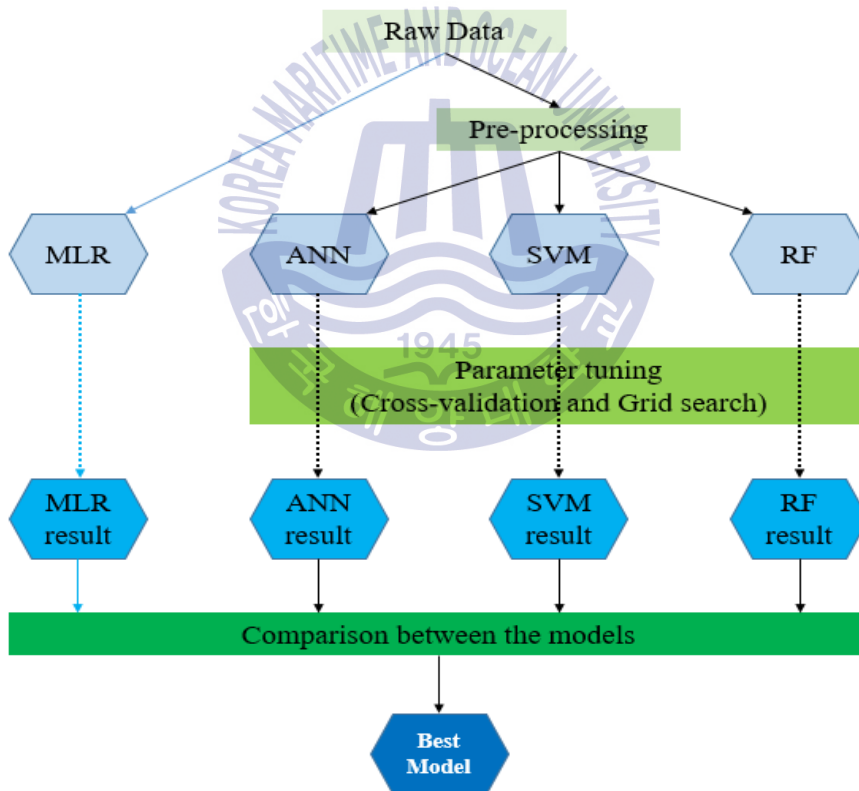


Figure 25 Flow Chart of Classified Chartering-out Decision

The data used are spot rate(tct), 6-month time charter(6m), 1-year time charter(1yr), earnings(earning), and newbuilding price(nb) in Capesize, Panamax, and Supramax sector respectively. All of them are downloaded from the Shipping Intelligence Network. The period of the data is revealed in Table [22]. The reason why Handymax is excluded is the availability of the data. However, the data for the first 11 weeks and the last year are excluded because the former period is calculated for the moving average and the latter period for the profit of the spot play.

Table 22 Descriptions of Data

Size	Data Period	Hire rate	Observation
Capesize	2009-07-17 ~ 2016-10-21	\$17,000/day	380
Panamax	2002-03-01 ~ 2016-08-26	\$12,000/day	757
Supramax	2002-12-27 ~ 2016-04-01	\$10,000/day	693

In addition to the freight rates, further variables are derived from the technical analysis and the fundamental analysis. From the point of view of technical analysis, investigating the past price behaviors itself has been known for helping to forecast the future dynamics of its prices. From another point of view, analyzing the public information about the financial data of individual firms can be conducive to forecasting the prices. The first point has usually been utilized for testing the weak-form of the efficient market and the second one for verifying the semi-strong form of it. Table [23] shows the indicators that this paper adopts.

Table 23 Additional Variables from Technical and Fundamental Analysis

Type	Description		
Technical Analysis (Alizadeh and Nomikos, 2009)	Term ratio	$\frac{FR_{tct}}{FR_{1yr}}$	This ratio reveals the relativeness between two prices.
	Size Difference	$2FR_{1yr}^{pmx} - FR_{1yr}^{cape}$ $3FR_{1yr}^{smx} - FR_{1yr}^{cape}$	This is a kind of 'spread trading'. It is the difference between twice of Panamax and one Cape and the difference between triple Supramax and one Cape.
	MA ratio	$\frac{4MA}{12MA}$	The moving average is most commonly used in the market and academia. It is the ratio between fast and slow moving averages.
Fundamental analysis	P/E ratio	$\ln\left(\frac{FR_{1yr}}{FR_{earn}}\right)$	The price-earnings ratio as the valuable indicator can be used to identify the proper timing of investment or divestment. (Campbell and Shiller, 1987, 1998, Alizadeh and Nomikos, 2006, 2007)

The ML models used in this study are kind of supervised learning where the target values are previously known in training phase. There are three kinds of decisions, which are the spot, 6-month, and 1-year charter-out contracts, and these desired decisions can be estimated through the real operations of each charter-out decision at the specific point of time on the basis of the charter-in rate as shown in Table [22]. The fact that the voyage charter rate is equal to the time-charter equivalent of spot rate (TCE) should be assumed for this experiment. Furthermore, it notes that the assumption of random hire rate does not affect the profit and loss of the real operation.

Figure [27] provides how many the number of the target decision are. The 'A' type decision denotes spot-play, 'B' 6-month, and 'C' 1-year.

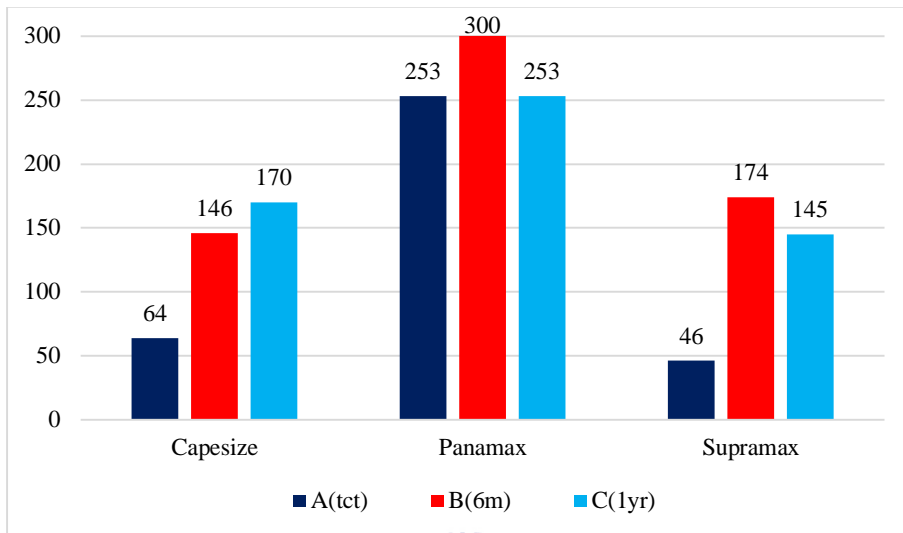


Figure 26 No. of Desired Chartering-out Decisions

In a similar way to preprocess the data in the previous problem, all the data is normalized and randomly scattered with disregard to time sequence because it satisfies statistical assumption that the distribution of the return in any point of data is constant (Yun, Lim and Lee, 2017). The modified data is charted in Figure [27] The samples are split into the training set and the test set according to the ratio of 80:20.

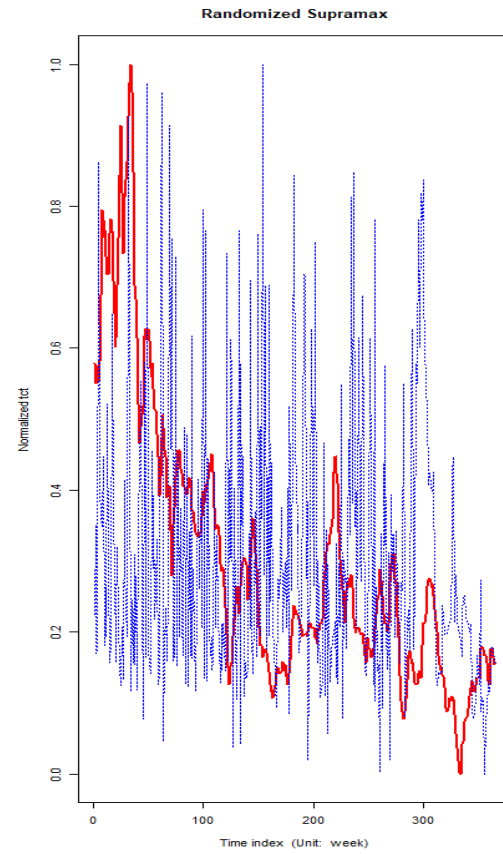
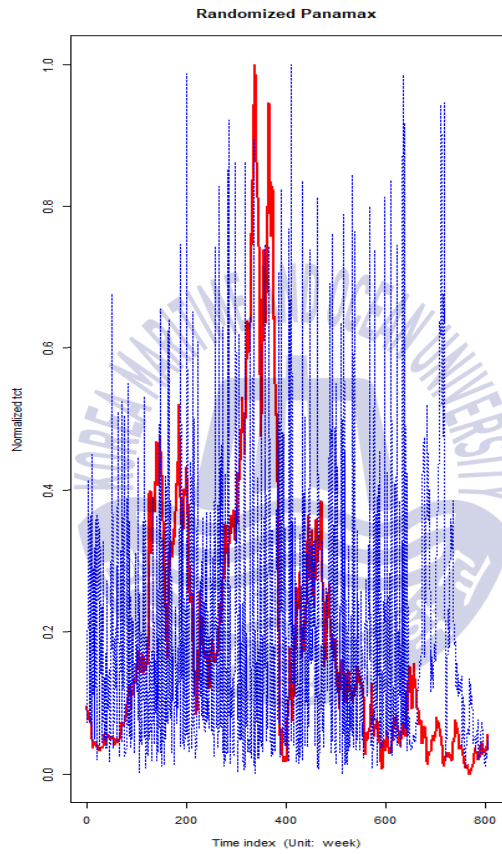
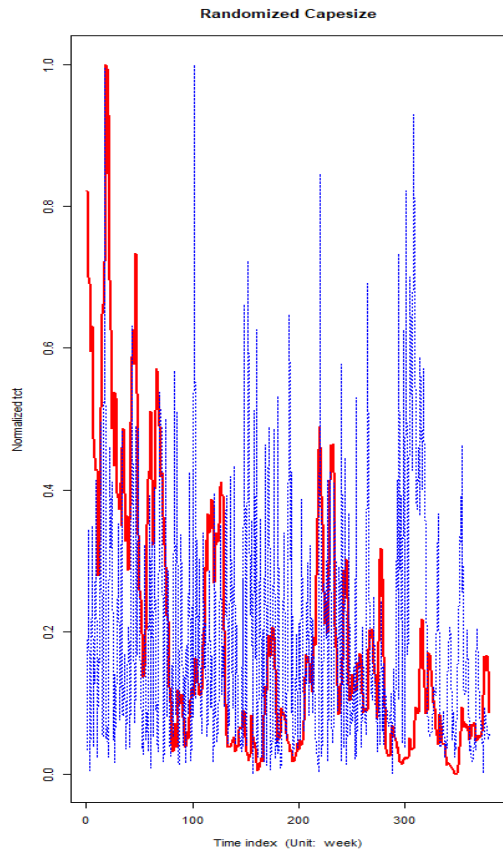


Figure 27 Original and Randomized Series of Freight Rates

5.2.2 MLR Modeling

The conventional MLR does not need to tune the parameters. Therefore, the preprocessing phase is excluded.

5.2.3 ANN Modeling

In order to find the optimal parameters such as the weight decay and the number of hidden nodes, the 10-folds cross-validation and the grid search are carried out. The weight decay and the number of the hidden nodes are presented in Table [24] and Figure [28].

Table 24 Optimal Parameters of ANN

Size	No. of Hidden node	Decay
Capesize	24	0.01
Panamax	21	0.01
Supramax	26	0.01

The number of the hidden layer is also important from the overfitting point of view. Since as reported in previous literature, the one hidden layer is sufficient for learning, the one hidden layer is the setting in this experiment.

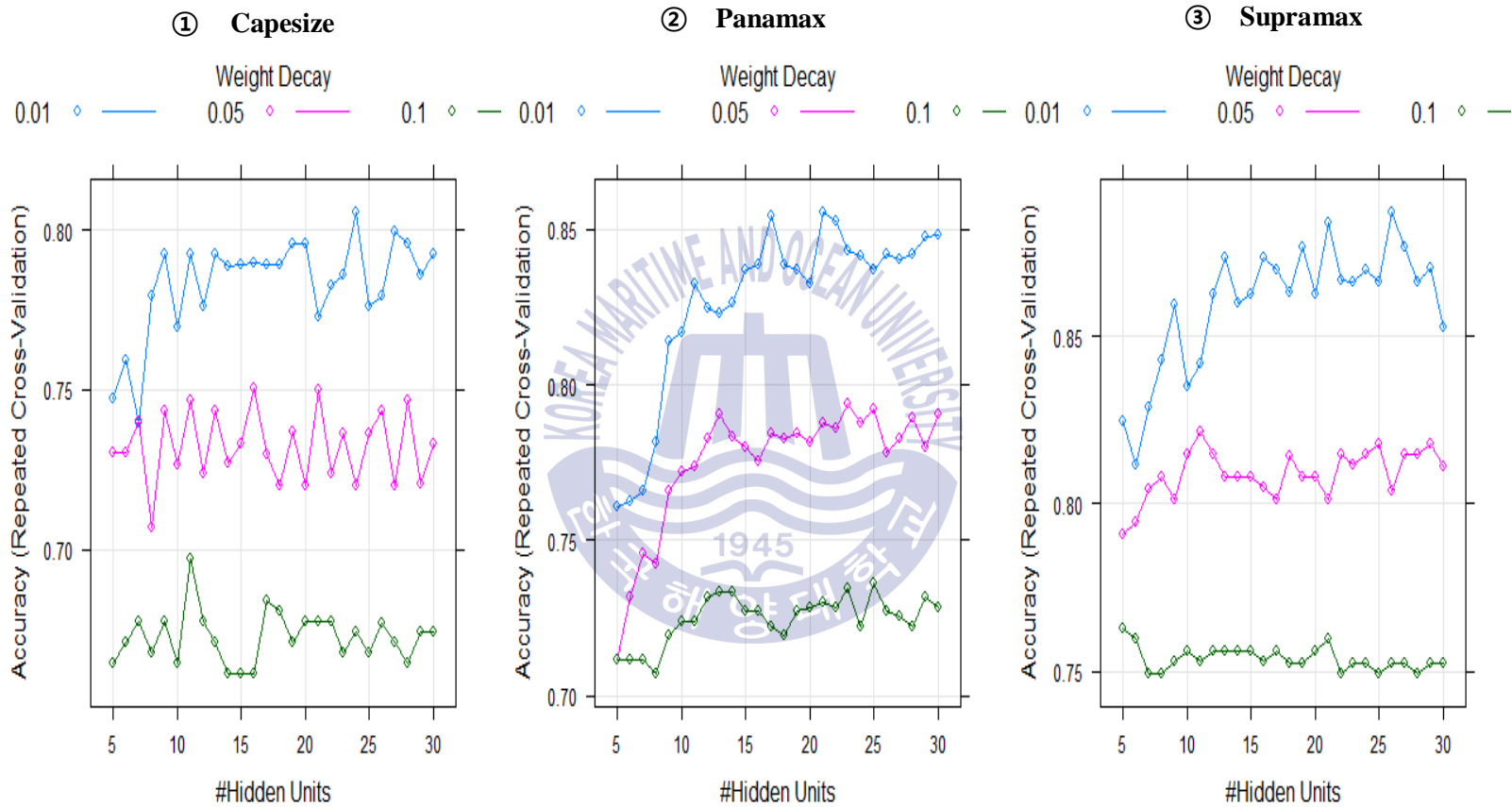


Figure 28 Optimal Parameters of ANN

5.2.4 SVM Modeling

Similar to the regression case, the best parameters can be obtained through 10-folds cross-validation and the grid search. The value of the sigma in the Gaussian kernel function is 1 and the cost, C, is in Table [25].

Table 25 Optimal Parameters of SVM

Size	Sigma	C
Capesize	1	8
Panamax	1	10
Supramax	1	2

The results of the parameter tuning are depicted in Figure[29] as well.

5.2.5 RF Modeling

The Table [26] and Figure [30] present that how many the optimal number is out of 9 predictors. Furthermore, the number of the constructed trees is confirmed in Table [26]. These parameters have undergone 10-folds cross-validation in order for tuning.

Table 26 Optimal Parameters of RF

Size	m	No. of trees
Capesize	5	500
Panamax	8	500
Supramax	6	600

5.2.6 Results and Discussion

Table [27] to [29] show the results of the models respectively.

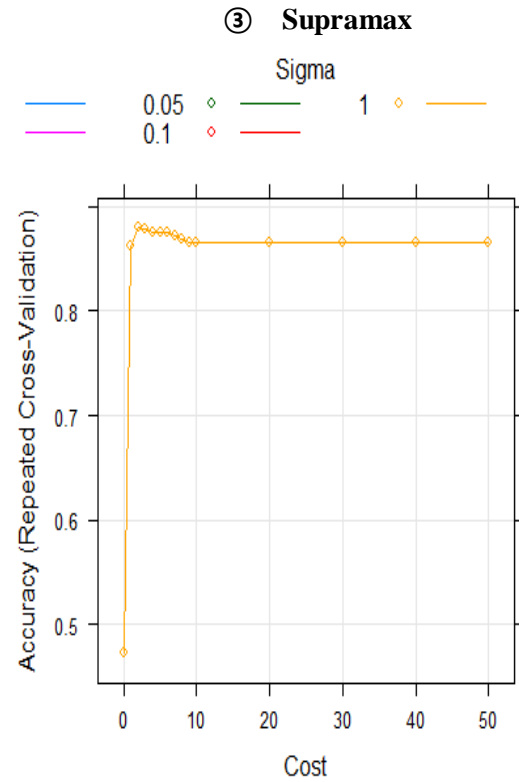
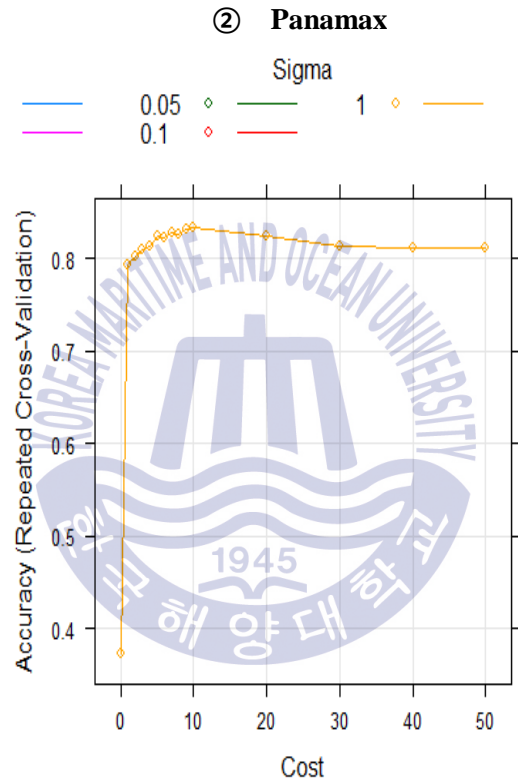
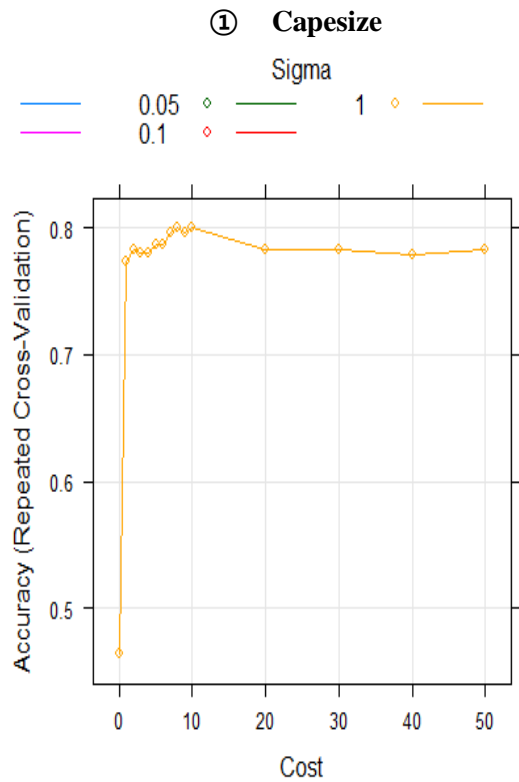


Figure 29 Optimal Parameters of SVM

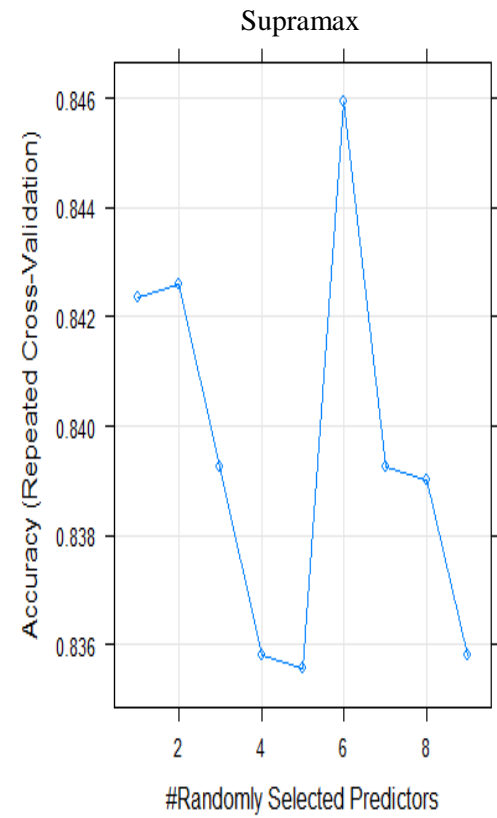
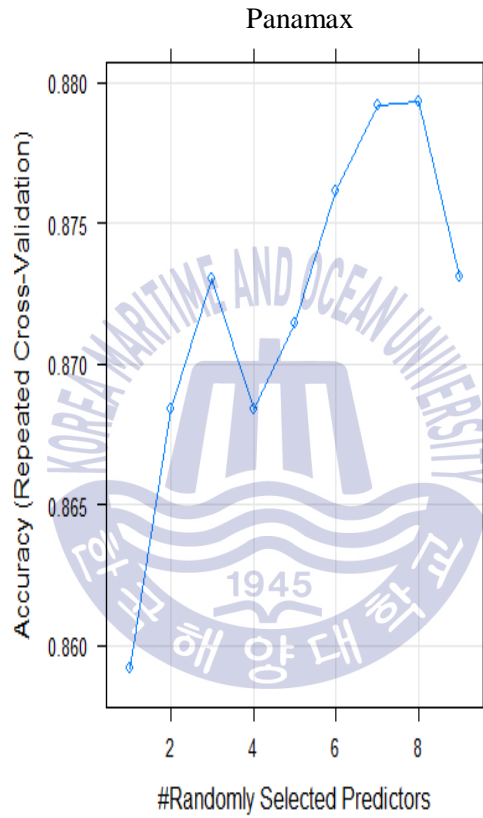
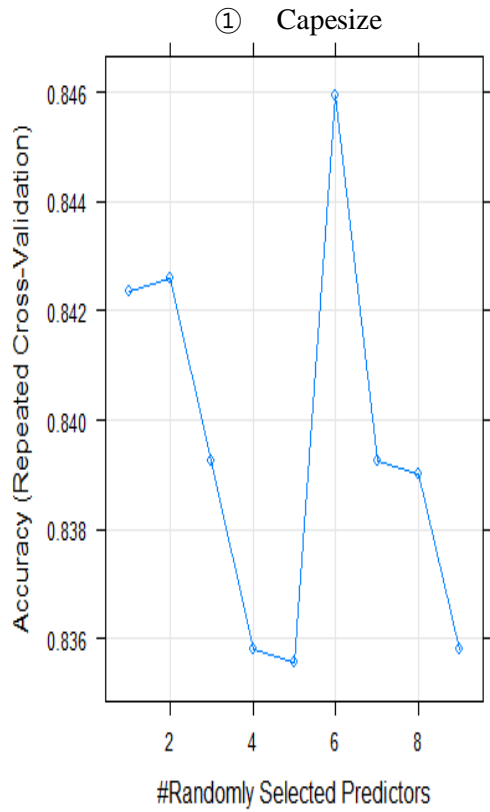


Figure 30 Optimal m of RF

Table 27 Test Results of Capesize

Capesize	Predicted default												Row TTL
	A(tct)				B(6m)				C(1yr)				
	MLR	ANN	SVM	RF	MLR	ANN	SVM	RF	MLR	ANN	SVM	RF	
A	4	9	11	11	5	4	3	2	6	2	1	2	15
(tct)	0.05	0.12	0.15	0.15	0.07	0.05	0.04	0.03	0.08	0.03	0.01	0.03	
B	1	2	2	2	16	26	26	25	15	4	4	5	32
(6m)	0.01	0.03	0.03	0.03	0.21	0.34	0.34	0.33	0.20	0.05	0.05	0.07	
C	0	0	0	0	6	4	4	3	23	25	25	26	29
(1yr)	0.00	0.00	0.00	0.00	0.08	0.05	0.05	0.04	0.30	0.33	0.33	0.34	
Column TTL	5	11	13	13	27	34	33	30	44	31	30	33	76

Table 28 Test Results of Panamax

Panamax	Predicted default												Row TTL
	A(tct)				B(6m)				C(1yr)				
	MLR	ANN	SVM	RF	MLR	ANN	SVM	RF	MLR	ANN	SVM	RF	
A	38	37	40	39	3	4	4	5	3	3	0	0	44
(tct)	0.24	0.23	0.25	0.24	0.02	0.03	0.03	0.03	0.02	0.02	0.00	0.00	
B	16	10	7	5	30	39	41	49	14	11	12	6	60
(6m)	0.10	0.06	0.04	0.03	0.19	0.24	0.26	0.30	0.09	0.07	0.08	0.04	
C	14	4	5	2	11	9	9	5	32	44	43	50	57
(1yr)	0.09	0.03	0.03	0.01	0.07	0.06	0.06	0.03	0.20	0.27	0.27	0.31	
Column TTL	68	51	52	46	44	52	54	59	49	58	55	56	161

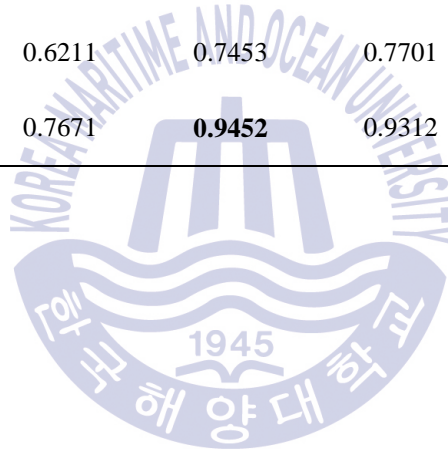
Table 29 Test Results of Supramax

Supramax	Predicted default												Row TTL
	A(tct)				B(6m)				C(1yr)				
	MLR	ANN	SVM	RF	MLR	ANN	SVM	RF	MLR	ANN	SVM	RF	
A	13	13	11	13	0	0	1	0	0	0	1	0	13
(tct)	0.18	0.18	0.15	0.18	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	
B	1	0	0	1	21	33	34	32	14	3	2	3	36
(6m)	0.01	0.00	0.00	0.01	0.29	0.45	0.47	0.44	0.19	0.04	0.03	0.04	
C	0	0	0	0	2	1	1	1	22	23	23	23	24
(1yr)	0.00	0.00	0.00	0.00	0.03	0.01	0.01	0.01	0.30	0.32	0.32	0.32	
Column TTL	14	13	11	14	23	34	36	33	36	26	26	26	73

The final results are summed up in Table [31]. As shown in the table, the performance of the RF is slightly preferred to the ANN. If considering the complexity of the calculation of ANN, the RF model is more competitive.

Table 30 Summary of Chartering-out Decisions

Size	MLR	ANN	SVM	RF
Capesize	0.5657	0.7894	0.8153	0.8157
Panamax	0.6211	0.7453	0.7701	0.8571
Supramax	0.7671	0.9452	0.9312	0.9315



Chapter 6 Conclusion

This paper proposes the prominent models to support the decision in chartering practice based on the forecasting accuracy. The problems defined in the previous chapter are the valuation of the T/C options and the decision to choose the period of charter-out. They can be closely linked to the regression-related problem and the classification-related problem. This point precisely matches to the applicability of the ML disciplines.

First, the commonly used instruments to attract the charterers is the T/C options in time charter contracts. Although options have the considerable economic values, two parties have been using them without evaluating. The BSM eminent in the financial market is considered as the benchmark model and three ML models as the candidates. It is worth noting that the key of modeling the ML techniques is how well the parameters in them are tuned to prevent “overfitting.” The cross-validation technique and the grid search algorithm are simultaneously adapted to tune them. The results from the fitted models greatly surpass the benchmark one. The relatively recently developed model, RF closely approximates the real value of the T/C option. Considering the time consumption, the model complexity, and easily interpretable results, the RF is more competitive than others.

Second, the chartering desks often face whether they sell a part of the period of the secured carrying capacities or a whole set. The service life of the charter-in vessel is usually determined as one year. Using the obtained vessels, the person in charge of the chartering decision has to determine to make them exposed to the spot, 6-month time charter, or 1-year time charter. The target selections of chartering decision are

discovered on the basis of the real profits or losses made by the real operation of the vessel. The multinomial logistic regression as the statistical model is treated as the benchmark model and the rests as candidates. In similar ways above, the parameters in the models have to be determined well in order to resolve the “overfitting.” Consequently, the RF model stands out well against ANN and SVM except for Supramax where the forecast accuracy of ANN is relatively higher than SVM and RF despite the results from much literature have told the SVM and RF are superior to the ANN. Likewise, the application of the ML methods is carefully carried out as the performance of them highly depends on the type of data to be used, and the structure of the models. In that sense, it is noted that the experience and the background knowledge of the researchers are quite important in applying the models to the problems. Oddly, it is ironic that the deeper the technology evolves, the more the researcher’s ability will be important. This might be that the machine models had been originated from the learning process of the human brain.

So far, this paper explores the applicability of the ML methodologies in chartering practice. The implication of the outstanding results of them is significant in the shipping industry. The questions relating to these problems are so pervasive in the maritime sector that can be solved by applying these models. In this sense, this paper will significantly contribute to triggering the post studies in the pricing of charter and decision-making of maritime business. However, if one is not a guru in the shipping industry or does not understand the ML methods, it is quite challenging to find the practical problems to be dealt with. Consequently, the collaboration of experts in the fields and academia is crucial.

References

- Ådland, A. O. and Koekebakker, S. (2004) 'Market Efficiency in the Second-hand Market for Bulk Ships', *Maritime Economics and Logistics*, 6(1), pp. 1–15.
- Adland, R. and Cullinane, K. (2005) 'A Time-Varying Risk Premium in the Term Structure of Bulk Shipping Freight Rates', *Journal of Transport Economics and Policy*, 39(2), pp. 191–208.
- Adland, R. and Strandenes, S. (2006) 'Market efficiency in the bulk freight market revisited', *Maritime Policy & Management*, 33(2), pp. 107–117.
- Ahmad, M. W., Mourshed, M. and Rezgui, Y. (2017) 'Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption', *Energy and Buildings*, 147, pp. 77–89. doi: 10.1016/j.enbuild.2017.04.038.
- Ahmed, Y. A. and Hasegawa, K. (2013) 'Automatic ship berthing using artificial neural network trained by consistent teaching data using nonlinear programming method', *Engineering Applications of Artificial Intelligence*, 26(10), pp. 2287–2304. doi: 10.1016/j.engappai.2013.08.009.
- Alizadeh, A. H. and Nomikos, N. K. (2006) 'Trading strategies in the market for tankers', *Maritime Policy & Management*, 33(2), pp. 119–140.
- Alizadeh, A. H. and Nomikos, N. K. (2007) 'Investment timing and trading strategies in the sale and purchase market for ships', *Transportation Research Part B: Methodological*, 41(1), pp. 126–143.
- Alizadeh, A. H. and Nomikos, N. K. (2009) *Shipping Derivatives and Risk Management*. London: Palgrave Macmillan UK.
- Alizadeh, A. H. and Talley, W. K. (2011) 'Vessel and voyage determinants of tanker freight rates and contract times', *Transport Policy*, 18(5), pp. 665–675. doi: 10.1016/j.tranpol.2011.01.001.
- Andreou, P. C., Charalambous, C. and Martzoukos, S. H. (2006) 'Robust Artificial

Neural Networks for Pricing of European Options', *Computational Economics*, 27, pp. 329–351. doi: 10.1007/s10614-006-9030-x.

Atsalakis, G. S. and Valavanis, K. P. (2009) 'Surveying stock market forecasting techniques - Part II: Soft computing methods', *Expert Systems with Applications*, 36, pp. 5932–5941. doi: 10.1016/j.eswa.2008.07.006.

Bal Beşikçi, E., Arslan, O., Turan, O. and Ölçer, A. I. (2016) 'An artificial neural network based decision support system for energy efficient ship operations', *Computers & Operations Research*, 66, pp. 393–401. doi: 10.1016/j.cor.2015.04.004.

Bao, J., Pan, L. and Xie, Y. (2016) 'A new BDI forecasting model based on support vector machine', in *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, pp. 65–69. doi: 10.1109/ITNEC.2016.7560320.

Barnston, A. G. (1992) 'Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score', *Weather and Forecasting*, 7(4), pp. 699–709. doi: 10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2.

Basheer, I. A. and Hajmeer, M. (2000) 'Artificial neural networks: Fundamentals, computing, design, and application', *Journal of Microbiological Methods*, 43(1), pp. 3–31.

Black, F. and Scholes, M. (1973) 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy*, 81(3), pp. 637–654.

Bollerslev, T. (1986) 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, 31(3), pp. 307–327. doi: 10.1016/0304-4076(86)90063-1.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. doi: 10.1017/CBO9781107415324.004.

Campbell, J. Y. and Shiller, R. J. (1987) 'Cointegration and Tests of Present Value Models', *Journal of Political Economy*, 95(5), pp. 1062–1088.

- Campbell, J. Y. and Shiller, R. J. (1998) 'Valuation ratios and the long run stock market outlook', *Journal of Portfolio Management*, 24(2), pp. 11–26. doi: 10.3905/jpm.24.2.11.
- Chaudhuri, T. D. and Ghosh, I. (2016) 'Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework', *Journal of Insurance and Financial Management*, 1(5), pp. 92–123.
- Chen, S., Härdle, W. K. and Jeong, K. (2010) 'Forecasting volatility with support vector machine-based GARCH model', *Journal of Forecasting*, 29(4), pp. 406–433. doi: 10.1002/for.1134.
- Chisté, C. and van Vuuren, G. (2014) 'Investigating the cyclical behaviour of the dry bulk shipping market', *Maritime Policy and Management*, 41(1), pp. 1–19.
- Cutler, A., Cutler, D. R. and Stevens, J. R. (2012) 'Random Forest', in Zhang, C. and Ma, Y. (eds) *Ensemble Machine Learning-Methods and Applications*. New York: Springer. doi: 10.1007/978-1-4419-9326-7.
- Daranda, A. (2016) 'Neural Network Approach to Predict Marine Traffic', *Baltic journal of Modern Computing*, 4(3), pp. 483–495.
- Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G. and Moncada-Herrera, J. A. (2008) 'A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile', *Atmospheric Environment*, 42(35), pp. 8331–8340. doi: 10.1016/j.atmosenv.2008.07.020.
- Eslami, P., Jung, K., Lee, D. and Tjolleng, A. (2017) 'Predicting tanker freight rates using parsimonious variables and a hybrid artificial neural network with an adaptive genetic algorithm', *Maritime Economics & Logistics*, 19(3), pp. 538–550. doi: 10.1057/mel.2016.1.
- Fadlalla, A. and Lin, C. H. (2001) 'An Analysis of the Applications of Neural Networks in Finance', *Interfaces*, 31(4), pp. 112–122. doi: 10.2307/25062724.
- Fama, E. F. (1965) 'Random Walks in Stock Market Prices', *Financial Analysts*

Journal, 21(5), pp. 55–59.

Fama, E. F. (1970) 'Efficient Capital Markets-A Review of Theory and Empirical Work', *The Journal of Finance*, pp. 383–417. doi: 10.2307/2329297.

Fama, E. F. (1991) 'Efficient Capital Markets II', *Journal of Finance*, 46(5), pp. 1575–1617. doi: 10.2307/2328565.

Fama, E. and French, K. (1992) 'The Cross-section of Expected Stock Returns', *The Journal of Finance*, 47(2), pp. 427–465. doi: 10.1111/j.1540-6261.1992.tb04398.x.

Fan, S., Ji, T., Gordon, W. and Rickard, B. (2013) 'Forecasting Baltic Dirty Tanker Index by Applying Wavelet Neural Networks', *Journal of Transportation Technologies*, 3(January), pp. 68–87.

Fernández-Rodríguez, F., González-Martel, C. and Sosvilla-Rivero, S. (2000) 'On the profitability of technical trading rules based on artificial neural networks', *Economics Letters*, 69(1), pp. 89–94. doi: 10.1016/S0165-1765(00)00270-6.

Gençay, R. and Qi, M. (2001) 'Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging', *IEEE Transactions on Neural Networks*, 12(4), pp. 726–734. doi: 10.1109/72.935086.

Ghiassi, M. and Saidane, H. (2005) 'A dynamic architecture for artificial neural networks', *Neurocomputing*, 63, pp. 397–413. doi: 10.1016/j.neucom.2004.03.014.

Glen, D. R. (1997) 'The market for second-hand ships: Further results on efficiency using cointegration analysis', *Maritime Policy & Management*, 24(3), pp. 245–260.

Gouelmos, A. M. and Gouelmos, M. A. (2009) 'The problem of timing in decisions to buy or to charter a vessel', *International Journal of Transport Economics*, 36(2), pp. 261–286.

Guresen, E., Kayakutlu, G. and Daim, T. U. (2011) 'Using artificial neural network models in stock market index prediction', *Expert Systems with Applications*, 38(8), pp. 10389–10397. doi: 10.1016/j.eswa.2011.02.068.

- Hale, C. and Vanags, A. (1989) 'Spot and period rates in the dry bulk market', *Journal of Transport Economics and Policy*, 23, pp. 281–291.
- Hale, C. and Vanags, A. (1992) 'The market for second-hand ships: Some results on efficiency using cointegration', *Maritime Policy & Management*, 19(1), pp. 31–39.
- Han, Q., Yan, B., Ning, G. and Yu, B. (2014) 'Forecasting dry bulk freight index with improved SVM', *Mathematical Problems in Engineering*, 2014(ID: 460684), p. 12 pages. doi: 10.1155/2014/460684.
- Huang, W., Nakamori, Y. and Wang, S.-Y. (2005) 'Forecasting stock market movement direction with support vector machine', *Computers & Operations Research*, 32(10), pp. 2513–2522. doi: 10.1016/j.cor.2004.03.016.
- Hung, N. Q., Babel, M. S., Weesakul, S. and Tripathi, N. K. (2009) 'An artificial neural network model for rainfall forecasting in Bangkok, Thailand', *Hydrology and Earth System Sciences*, 13, pp. 1413–1425. doi: 10.5194/hess-13-1413-2009.
- Ju-Long, D. (1982) 'Control problems of grey systems', *Systems & Control Letters*, 1(5), pp. 288–294. doi: 10.1016/S0167-6911(82)80025-X.
- Kaasra, I. and Boyd, M. (1996) 'Designing a neural network for forecasting financial and economic time series', *Neurocomputing*, 10(3), pp. 215–236.
- Kara, Y., Acar Boyacioglu, M. and Baykan, Ö. K. (2011) '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*, 38(5), pp. 5311–5319. doi: 10.1016/j.eswa.2010.10.027.
- Karakitsos, E. and Varnavides, L. (2014) *Maritime Economics: A Macroeconomic Approach*. London: Palgrave Macmillan UK.
- Kavussanos, M. G. and Alizadeh, A. H. (2002a) 'Efficient pricing of ships in the dry bulk sector of the shipping industry', *Maritime Economics & Logistics*, 29(3), pp. 303–330.

Kavussanos, M. G. and Alizadeh, A. H. (2002b) 'The Expectations Hypothesis of the Term Structure and Risk Premia in Dry Bulk Shipping Freight Markets', *Journal of Transport Economics and Policy*, 36(2), pp. 267–304.

Khashei, M. and Bijari, M. (2010) 'An artificial neural network (p, d, q) model for time series forecasting', *Expert Systems with Applications*, 37(1), pp. 479–489. doi: 10.1016/j.eswa.2009.05.044.

Kim, K. (2003) 'Financial time series forecasting using support vector machines', *Neurocomputing*, 55(1–2), pp. 307–319. doi: 10.1016/S0925-2312(03)00372-2.

Kocak, H. and Un, T. (2014) 'Forecasting the Gold Returns with Artificial Neural Network and Time Series', *International Business Research*, 7(11), pp. 139–152. doi: 10.5539/ibr.v7n11p139.

Kourouniotti, I., Polydoropoulou, A. and Tsiklidis, C. (2016) 'Development of Models Predicting Dwell Time of Import Containers in Port Container Terminals - An Artificial Neural Networks Application', in *Transportation Research Procedia*, pp. 243–252. doi: 10.1016/j.trpro.2016.05.061.

Kristjanpoller, W. and Minutolo, M. C. (2015) 'Gold price volatility: A forecasting approach using the Artificial Neural Network-GARCH model', *Expert Systems with Applications*, 42(20), pp. 7245–7251. doi: 10.1016/j.eswa.2015.04.058.

Kristjanpoller, W. and Minutolo, M. C. (2016) 'Forecasting volatility of oil price using an artificial neural network-GARCH model', *Expert Systems with Applications*, 65, pp. 233–241. doi: 10.1016/j.eswa.2016.08.045.

Kumar, M. and Thenmozhi, M. (2014) 'Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models', *International Journal of Banking, Accounting and Finance*, 5(3), pp. 284–308. doi: 10.1504/IJBAAF.2014.064307.

Lajbcygier, P. R. and Conner, J. T. (1997) 'Improved option pricing using artificial neural networks and bootstrap methods', *International Journal of Neural Systems*, 8(4), pp. 457–471.

Lama, A., Jha, G. K., Gurung, B., Paul, R. K., Bharadwaj, A. and Parsad, R. (2016) 'A Comparative Study on Time-delay Neural Network and GARCH Models for Forecasting Agricultural Commodity Price Volatility', *Journal of the Indian Society of Agricultural Statistics*, 70(1), pp. 7–18.

Lazakis, I., Raptodimos, Y. and Varelas, T. (2017) 'Predicting ship machinery system condition through analytical reliability tools and artificial neural networks', *Ocean Engineering*. Elsevier Ltd, pp. 1–12. doi: 10.1016/j.oceaneng.2017.11.017.

De León, A. D., Lalla-Ruiz, E., Melián-Batista, B. and Marcos Moreno-Vega, J. (2017) 'A Machine Learning-based system for berth scheduling at bulk terminals', *Expert Systems with Applications*, 87, pp. 170–182. doi: 10.1016/j.eswa.2017.06.010.

Li, J. and Parsons, M. G. (1997) 'Forecasting tanker freight rate using neural networks', *Maritime Policy and Management*, 24(1), pp. 9–30. doi: 10.1080/03088839700000053.

Liang, X., Zhang, H., Xiao, J. and Chen, Y. (2009) 'Improving option price forecasts with neural networks and support vector regressions', *Neurocomputing*. Elsevier, 72, pp. 3055–3065. doi: 10.1016/j.neucom.2009.03.015.

Lin, T., Horne, B. G., Tiiio, P. and Giles, C. L. (1996) 'Learning Long-Term Dependencies in NARX Recurrent Neural Networks', *IEEE Transactions on Neural Networks*, 7(6), pp. 1329–1338.

Lyridis, D., Zacharioudakis, P., Iordanis, S. and Daleziou, S. (2013) 'Freight-Forward Agreement Time series Modelling Based on Artificial Neural Network Models', *Strojniški vestnik – Journal of Mechanical Engineering*, 59(9), pp. 511–516. doi: 10.5545/sv-jme.2013.947.

Lyridis, D., Zacharioudakis, P., Mitrou, P. and Mylonas, A. (2004) 'Forecasting Tanker Market Using Artificial Neural Networks', *Maritime Economics & Logistics*, 6, pp. 93–108.

Merton, R. C. (1973) 'Theory of rational option pricing', *The Bell Journal of*

Economics and Management Science, 4(1), pp. 141--183.

Mirakyan, A., Meyer-Renschhausen, M. and Koch, A. (2017) 'Composite forecasting approach, application for next-day electricity price forecasting', *Energy Economics*, 66, pp. 228–237. doi: 10.1016/j.eneco.2017.06.020.

Morelli, M. J., Montagna, G., Nicosini, O., Treccani, M., Farina, M. and Amato, P. (2004) 'Pricing financial derivatives with neural networks', *Physica A: Statistical Mechanics and its Applications*, 338(1), pp. 160–165.

Mostafa, M. M. (2004) 'Forecasting the Suez Canal Traffic: A Neural Network Analysis', *Maritime Policy & Management*, 31(2), pp. 139–156.

Nelson, D. B. (1991) 'Conditional Heteroskedasticity in Asset Returns: A New Approach', *Econometrica*, 59(2), pp. 347–370. doi: 10.2307/2938260.

Nilsson, N. J. (2010) *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. 1st edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511819346.

De Oliveira, F. A., Nobre, C. N. and Zárata, L. E. (2013) '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*, 40(18), pp. 7596–7606. doi: 10.1016/j.eswa.2013.06.071.

Pagoropoulos, A., Møller, A. H. and McAlloone, T. C. (2017) 'Applying Multi-Class Support Vector Machines for performance assessment of shipping operations: The case of tanker vessels', *Ocean Engineering*, 140, pp. 1–6. doi: 10.1016/j.oceaneng.2017.05.001.

Pai, P.-F. and Lin, C.-S. (2005) 'A hybrid ARIMA and support vector machines model in stock price forecasting', *Omega*, 33(6), pp. 497–505. doi: 10.1016/j.omega.2004.07.024.

Pesaran, M. H. and Timmermann, A. (1992) 'A simple nonparametric test of predictive performance', *Journal of Business and Economic Statistics*, 10(4), pp. 461–465. doi: 10.2307/1391822.

Roh, T. H. (2007) 'Forecasting the volatility of stock price index', *Expert Systems with Applications*, 33(4), pp. 916–922. doi: 10.1016/j.eswa.2006.08.001.

Santos, A. A. P., Junkes, L. N. and Pires Jr, F. C. M. (2014) 'Forecasting period charter rates of VLCC tankers through neural networks: A comparison of alternative approaches', *Maritime Economics & Logistics*, 16(1), pp. 72–91.

Shinde, A. S. and Takale, K. C. (2012) 'Study of Black-Scholes model and its applications', in *International Conference on Modelling, Optimisation and Computing*, pp. 270–279. doi: 10.1016/j.proeng.2012.06.035.

Smith, K. A. and Gupta, J. N. D. (2000) 'Neural networks in business: techniques and applications for the operations researcher', *Computers & Operations Research*, 27(11), pp. 1023–1044.

Sompui, M. and Wongsinlatam, W. (2014) 'Prediction Model for Crude Oil Price Using Artificial Neural Networks', *Applied Mathematical Sciences*, 8(80), pp. 3953–3965.

von Spreckelsen, C., von Mettenheim, H.-J. and Breitmer, M. H. (2012) 'Short-Term Trading Performance of Spot Freight Rates and Derivatives in the Tanker Shipping Market: Do Neural Networks Provide Suitable Results?', in Jayne, C., Yue, S., and Iliadis, L. (eds) *Engineering Applications of Neural Networks (EANN 2012)*, *Communications in Computer and Information Science*. Berlin, Heidelberg: Springer, pp. 443–452.

Stopford, M. (2009) *Maritime Economics*. 3rd edn. Routledge.

Suh, J.-D. (2016) 'Foreign Exchange Rate Forecasting Using the GARCH extended Random Forest Model (Korean)', *Journal of Industrial Economics and Business*, 29(5), pp. 1607–1628.

Tanaka, K., Kinkyō, T. and Hamori, S. (2016) 'Random forests-based early warning system for bank failures', *Economics Letters*, 148, pp. 118–121. doi: 10.1016/j.econlet.2016.09.024.

Tay, F. E. H. and Cao, L. (2001) 'Application of support vector machines in

financial time series forecasting', *Omega*, 29(4), pp. 309–317. doi: 10.1016/S0305-0483(01)00026-3.

Ticknor, J. L. (2013) 'A Bayesian regularized artificial neural network for stock market forecasting', *Expert Systems with Applications*, 40(14), pp. 5501–5506. doi: 10.1016/j.eswa.2013.04.013.

Tkáč, M. and Verner, R. (2016) 'Artificial neural networks in business : Two decades of research', *Applied Soft Computing*, 38, pp. 788–804.

Trafalis, T. B. and Ince, H. (2000) 'Support vector machine for regression and applications to financial forecasting', *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, 6, pp. 348–353. doi: 10.1109/IJCNN.2000.859420.

Tseng, C.-H., Cheng, S.-T., Wang, Y.-H. and Peng, J.-T. (2008) 'Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices', *Physica A: Statistical Mechanics and its Applications*, 387(13), pp. 3192–3200.

Vapnik, V. N. (1995) *The Nature of statistical Learning Theory*. New York: Springer. doi: 10.1007/978-1-4757-2440-0.

Vapnik, V. N. (1997) *Statistical Learning Theory*. New York: Wiley. doi: 10.2307/1271368.

Veenstra, A. W. (1999) 'The term structure of ocean freight rates', *Maritime Policy & Management*, 26(3), pp. 279–293.

Wang, J.-J., Wang, J.-Z., Zhang, Z.-G. and Guo, S.-P. (2012) 'Stock index forecasting based on a hybrid model', *Omega*, 40(6), pp. 758–766. doi: 10.1016/j.omega.2011.07.008.

Wang, J. Z., Wang, J. J., Zhang, Z. G. and Guo, S. P. (2011) 'Forecasting stock indices with back propagation neural network', *Expert Systems with Applications*, 38(11), pp. 14346–14355. doi: 10.1016/j.eswa.2011.04.222.

Wang, S. and Shang, W. (2014) 'Forecasting direction of China security index 300 movement with least squares support vector machine', in *Procedia Computer Science*, pp. 869–874. doi: 10.1016/j.procs.2014.05.338.

Wang, Y. (2014) 'Market Index and Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HSI', *International Journal of Business Intelligence and Data Mining*, (2), p. 145. doi: 10.1504/IJBIDM.2014.065091.

Willmott, C. J. and Matsuura, K. (2005) 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', *Climate Research*, 30(1), pp. 79–82. doi: 10.3354/cr030079.

Yang, H., Dong, F. and Ogandaga, M. (2008) 'Forewarning of Freight Rate in Shipping Market Based on Support Vector Machine', in Mao, B., Tian, Z., Huang, H., and Gao, Z. (eds) *Traffic and Transportation Studies*. ASCE, pp. 295–303. doi: 10.13334/j.0258-8013.pcsee.2005.07.020.

Yao, J., Li, Y. and Tan, C. L. (2000) 'Option price forecasting using neural networks', *Omega*, 28(4), pp. 455–466.

Yun, H., Lim, S. and Lee, K. (2017) 'The value of options for time charterparty extension: an artificial neural networks (ANN) approach', *Maritime Policy and Management*, pp. 1–14. doi: 10.1080/03088839.2017.1392630.

Zhang, G. P. (2003) 'Time series forecasting using a hybrid ARIMA and neural network model', *Neurocomputing*, 50, pp. 159–175. doi: 10.1016/S0925-2312(01)00702-0.

Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998) 'Forecasting with artificial neural networks: The state of the art', *International Journal of Forecasting*, 14(1), pp. 35–62.

Zhang, J.-H., Li, Z.-M., Kong, F.-T., Dong, X.-X., Chen, W. and Wang, S.-W. (2013) 'Prediction of Pork Prices based on SVM', in Xu, S. (ed.) *Proceeding of 2013 World Agricultural Outlook Conference*. Springer, pp. 173–178. doi: 10.1007/978-3-642-54389-0_15.

Zhu, B. and Wei, Y. (2013) 'Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology', *Omega (United Kingdom)*, 41(3), pp. 517–524. doi: 10.1016/j.omega.2012.06.005.

