

# 케이프선 시장 운임의 결정요인 및 운임예측 모형 분석

임상섭\* · † 윤희성

\*한국해양대학교 해운경영학부, † 한국해양수산개발원 해운빅데이터연구센터

## An Analysis on Determinants of the Capesize Freight Rate and Forecasting Models

Sang-Seop Lim\* · † Hee-Sung Yun

\*Division of Shipping Management, Korea Maritime and Ocean University, Busan 49112, Korea

† Centre for Shipping Big Data Analytics, Korea Maritime Institute, Busan 49111, Korea

**요약** : 운임시장의 심한 변동성과 시계열 데이터의 불안정성으로 해운시황 예측에 대한 연구가 큰 성과를 내지 못하고 있지만 최근 대표적 비선형 모델인 기계학습모델을 적용한 연구들이 활발히 진행되고 있다. 대부분의 기존 연구가 계량모델의 설계단계에서 입력변수에 해당하는 요인들을 기존 문헌연구와 연구자의 직관에 의존하여 선정했기 때문에 요인선정에 대한 체계적인 연구가 필요하다. 본 연구에서는 케이프선 운임을 대상으로 단계적 회귀모형과 랜덤포레스트모형을 이용하여 중요 영향요인을 분석하였다. 해운시장에서 비교적 단순한 수급구조를 가져 요인파악이 용이한 케이프선 운임을 대상으로 하였으며 총 16개의 수급요인들을 사전 추출하였다. 요인간의 상호관련성을 파악하여 단계적 회귀는 8개 요인, 랜덤포레스트는 10개 요인을 분석대상으로 선정하였으며 선정된 변수를 입력변수로 하여 예측한 결과를 비교하였다. 랜덤포레스트의 예측성능이 아주 우수하였는데 주요요인이 주로 선정된 단계적 회귀분석과는 달리 공급요인이 비중 있게 선정되었기 때문인 것으로 판단된다. 본 연구는 운임예측 연구에 있어 운임결정요인에 대한 과학적인 근거를 마련하였으며 이를 위해 기계학습 기반의 모델을 활용하였다는데 연구적 의의가 있다. 또한 시장정보의 분석에 있어 실무자들이 어떤 변수에 중점을 두어야 하는지에 대해 합리적 근거를 제시한 측면에서 해운기업의 의사결정에 실질적 도움이 될 것으로 기대된다.

**핵심용어** : 케이프선 운임, 운임 결정요인, 단계적 회귀분석, 랜덤포레스트

**Abstract** : In recent years, research on shipping market forecasting with the employment of non-linear AI models has attracted significant interest. In previous studies, input variables were selected with reference to past papers or by relying on the intuitions of the researchers. This paper attempts to address this issue by applying the stepwise regression model and the random forest model to the Cape-size bulk carrier market. The Cape market was selected due to the simplicity of its supply and demand structure. The preliminary selection of the determinants resulted in 16 variables. In the next stage, 8 features from the stepwise regression model and 10 features from the random forest model were screened as important determinants. The chosen variables were used to test both models. Based on the analysis of the models, it was observed that the random forest model outperforms the stepwise regression model. This research is significant because it provides a scientific basis which can be used to find the determinants in shipping market forecasting, and utilize a machine-learning model in the process. The results of this research can be used to enhance the decisions of chartering desks by offering a guideline for market analysis.

**Key words** : Capesize Market, Freight Determinants, Stepwise Regression, Random Forest

### 1. 서 론

해운시장은 다른 상품시장보다 극심한 변동성에 노출되어 있다. 많은 해운시장 참여자들은 경기에 순행하여 호황기에 대량의 선박투자를 하고 뒤이어 도래하는 불황기에 상당한 재무적 부담을 지며 일부는 파산에 이르기도 하였다. 경기순행적인 투자에는 여러 원인들이 작용하겠지만 의사결정과정에서 경험에 의존하는 경향이 강하며 시황예측의 관점에서 과학적 방법의 적용이 이루어지지 않은 것이 중요한 요인으로 지목된다. 시황의 변동성이 시장참여자의 수익에 직접적인 영향

을 미치므로 학계나 산업계의 관심이 컸고, 시황예측 연구에서도 다양한 노력들이 있었다(Celik et al., 2009).

본 논문에서는 운임 결정에 영향을 미치는 요인을 분석하는데 기계학습 모델을 적용하고자 한다. 지금까지의 운임 예측연구에는 주로 다양한 계량경제학적 기법들이 적용되었다. Veenstra and Franses(1997)는 VECM(Vector Error Correction Model)을 사용하여 케이프선 3개 항로 운임과 파나마스 3개 항로 운임의 장단기 균형관계를 분석하였으며 Batchelor et al.(2007)는 스팟운임과 선도가격의 관계분석에 ARIMA(Autoregressive Integrated Moving Average), VAR(Vector Autoregression), 그리고 VECM 모

† Corresponding author : 정희원, heesung.yun@gmail.com

\* 정희원, limsangseop@kmou.ac.kr

형을 사용하였다. 이 논문에서는 수급요인보다는 선형에 따른 항로별 운임과 파생상품의 가격의 시차변수를 주요 요인으로 선정하고 이들 간의 관계를 분석하였는데, 계량모형의 가정을 충족시키는 과정에서 이론적 접근이 어려워질 뿐만 아니라 모델에 대한 이해와 결과의 해석이 까다로워져 현실 적용에 제약이 따른다.

건화물 운임 예측에 기계학습 모델을 적용한 연구는 비교적 최근에 보고되고 있다. 대부분의 연구가 인공지능경망(Zeng and Qu, 2014; Zeng et al., 2016)과 서포트벡터머신(Yang, Jin and Wang, 2011; Han et al., 2014; Bao, Pan and Xie, 2016)을 이용한 분석이었다. Bao et al.(2016)의 연구를 제외하면 WD(Wavelet Decomposition) 또는 EMD(Empirical Mode Decomposition)을 이용하여 불안정한 운임 시계열을 고·저파동으로 분해한 후 이를 기계학습 모델의 입력변수로 사용하여 학습하는 방법을 채택하였다. Bao et al.(2016)의 경우에는 BDI에 영향을 미치는 요인으로 최대 자원수입국인 중국을 기준으로 최대 수출국의 환율과 장단기 이자율만을 채택하여 분석하였다. 지금까지의 연구들을 살펴보면 운임에 영향을 미치는 외생적 요인들을 선정시에는 연구자의 직관에 의해 선정되어 외생적 요인에 관한 일치된 견해가 부재하다. 따라서 본 연구는 수급요인이 비교적 단순하고 케이프션 운임시장을 대상으로 하여 운임에 영향을 미치는 요인들을 수급요인으로만 한정하였으며, 운임 결정에 가장 영향력이 큰 요인들을 ‘선정’하는 문제를 다뤘다.

GIGO(Garbage In, Garbage Out)는 해운 시황예측 모델에도 예외 없이 적용된다. 따라서 모델의 예측성과를 가장 크게 좌우하는 것은 “변수선택(feature selection)”이라고 할 수 있다. ‘빅데이터’ 연구에서도 유의미한 정보를 추출하기 위한 연구의 첫 번째 단계가 “변수선택”이다(Cai et al., 2018). 이제까지 운임 예측연구에 있어서는 통상적으로 기존 문헌과 연구자의 직관에 의해 변수를 선정하였으며 모델의 결과단계에서 변수에 대한 파라미터의 통계적 유의성에 대한 정보만을 제공하였다. 해운 시장과 밀접한 관계가 있는 거시경제변수를 선정한 경우도 있으나 수요·공급요인이 제외되어 거시적 관계에만 치우친 경향이 있다(Bao, Pan and Xie, 2016). 본 논문에서는 케이프션 운임의 결정에 영향을 미치는 수급요인들을 선정하고 변수의 중요도분석에 탁월한 성과를 보이는 단계적 회귀분석(Stepwise regression, LM)과 랜덤포레스트모형(Random forest, RF)을 채택하였다. 케이프션 운임을 선정한 이유는 케이프션 운임시장이 수요와 공급 요인의 변동에 상당히 민감하게 반응하기 때문이다. 선종 간 운임의 변동성 차이는 Fig. 1에 잘 나타나 있다.

케이프션 운송시장은 다른 해운시장에 비해 단순한 구조를 보인다. 운송화물은 대부분 철광석(iron ore)과 석탄(coal)이며 Fig. 2에서 보는 바와 같이 주요 수출지와 수입지가 명확하게 구분된다. 따라서 다른 선형의 운임에 비해 케이프션이 운임 결정에 영향을 미치는 수급요인의 식별이 용이하다. 석탄의 경우 원료탄(coking coal)은 제철산업과 밀접하게 연동

되어 있어 연료탄(steam coal)만을 분석에 포함하였다. 전 세계 철광석 무역량은 2017년 기준으로 14.7억 톤이며 호주가 56.1%, 브라질이 25.8%를 수출하고 주요 수입국인 중국이 전체의 약 73%를 수입한다.

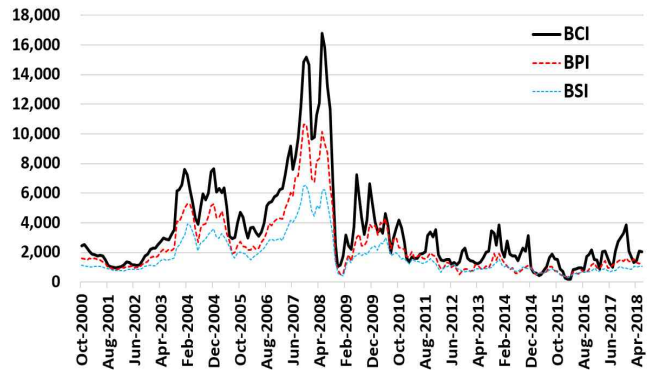


Fig. 1 Baltic indices for dry bulk market  
Source : Clarkson Research

석탄의 경우는 동년 기준으로 9.5억 톤이 거래되었으며, 인도네시아가 40.6%, 호주가 21.1%를 수출하며 수입측면에서 한·중·일 3국이 각 12.2%, 18.3%, 12.0%로 총 42.5%를 수입하고 있다. 이 데이터에서도 케이프션 시장의 대상화물인 석탄과 철광석의 주요 수출국과 수입국이 분명하게 드러남을 확인할 수 있다.

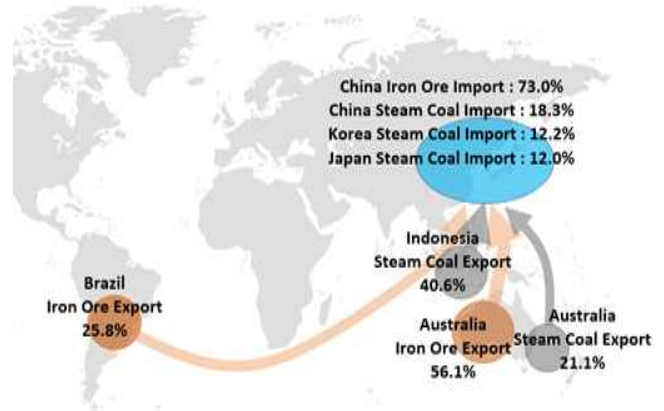


Fig. 2 World trade of iron ore and steam coal  
Source : Clarkson Research

Stopford(2009)는 운임의 결정요인으로 Table 1과 같은 수급요인을 제시하였다. 많은 선행연구들이 Stopford가 제시한 수요 요인에 근거하여 세계경제를 나타내는 거시경제변수를 채택하고 있다. 그리고 해운의 국제성, 높은 레버리지, 변동성에 기인한 시황위험과 관련된 거시지표를 중요요인으로 선정하기도 하였다. 하지만 운임에 영향을 미치는 수급요인과 거시지표의 복잡한 관계 때문에 중요요인을 식별하기 어렵다. 또한 일부 연구에서는 해운지수가 세계경제를 선행하는 변수로 분석되어 관계를 해석하기 복잡하다(Bakshi, Panayotov and

Skoulakis, 2011). 본 연구에서는 단순한 구조로 되어있는 케이프선 운임시장을 선택하였는데 수급요인을 다른 시장보다 명확히 파악함으로써 거시요인을 최대한 배제할 수 있도록 하였다.

Table 1 Freight determinants proposed by Stopford(2009)

Factor	1 <sup>st</sup> tier	2 <sup>nd</sup> tier	3 <sup>rd</sup> tier
Demand	World Economy	Seaborne Trades	Average Haul
Supply	World Fleet	Fleet productivity	Scrapping Delivery

본 연구는 케이프선 운임의 결정요인을 분석하는데 기계학습모델을 대표하는 랜덤포레스트 알고리즘을 적용하였다는 점에서 차별적 가치가 있다. 기존의 문헌연구에서 제시한 운임요인들은 기술적 지표(technical indicators)나 자기회귀변수를 포함하는 경우가 있는데, 심각한 변동성, 시계열의 불안정성과 수급요인 간 관계의 복잡성 때문에 분석에 어려움이 많았다(Lyridis et al., 2004; Gouelmos and Gouelmos, 2009). 본 논문에서는 케이프선 운임의 결정과정을 도식화하여 운송의 흐름에 따라 요인들을 도출하였으며, 계량모델과 기계학습모델을 이용하여 통계적 합리성과 과학적 근거를 제시하고 중요요인을 선정하였다. 뿐만 아니라 실무에서 시황정보를 분석하는데 초점을 설정하기 위한 기술적인 모델과 근거를 제시할 수 있으므로 실무적인 기여도 클 것으로 기대된다. 국내의 연구는 인공지능망을 활용한 항만물동량 예측(Shin et al., 2008, Shin and Jeong, 2011)이나 벌크운임의 거래전략(Yun et al., 2016)에 인공지능망을 활용한 연구로 한정되어 있는데 케이프선 운임의 결정요인 선정에 대하여 그동안 잘 활용되지 않은 랜덤포레스트를 적용함으로써 연구범위를 넓히고 새로운 모델을 제시하였다는 점이 연구의 성과라 할 수 있다.

## 2. 데이터 및 모델링

### 2.1 데이터

케이프선 운임의 결정요인을 선정하기 위하여 운임에 직접적인 영향이 있는 수급요인을 식별하였다. Fig. 2와 같이 케이프선의 수요는 철광석과 석탄으로 대별된다. 철광석의 경우 주요 공급지인 브라질과 호주의 운송거리에 차이가 있어 지역별 수출량의 변화에 따라 톤·마일 수요가 달라진다. 석탄의 경우, 인도네시아는 대부분 파나마스 이하의 선박으로 운송하기 때문에 호주와의 수출량 비율이 중요한 영향을 미칠 수 있다. 공급측면에서 케이프선 선박량에 영향을 주는 요인으로 발주량, 인도량, 해체량, 해체선가 등이 있다. 이중 발주량은 발주와 인도 사이의 시차를 고려해야하기 때문에 입력변수로 투입하기 전에 시점을 조정할 필요가 있다. 이 연구에서는 발주량과 다른 입력변수의 시점을 일치시키기 위하여 통상적인

건조기간인 2년을 반영하였다. 공급적인 측면에서 선박의 운임에 따른 수송서비스 생산효율이 운임에 중요한 영향을 미친다. 이는 선속(speed)과 관련이 있으며 보통 연료유가가 높을수록 선속이 낮아진다.

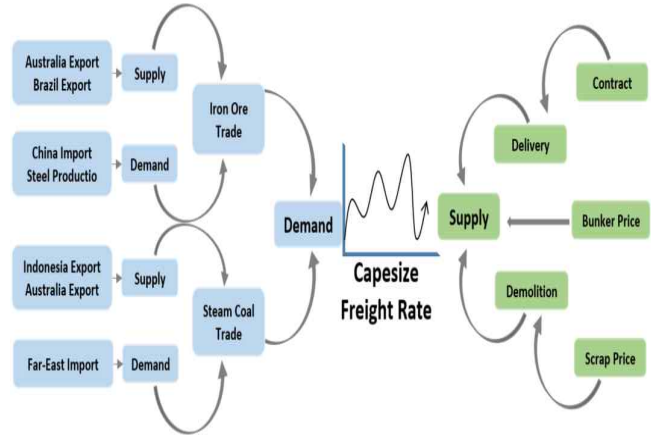


Fig. 3 Causality map of capesize freight determinants

따라서 선박의 가동률을 나타내는 대리변수로 연료유가를 선택하였다. 이를 종합하면 Fig. 3과 같이 인과관계를 나타낼 수 있다. 선정된 데이터의 통계적 특성은 Table 2와 같으며 데이터의 시점을 일치시키기 위해 2002년 1월부터 2015년 12월까지로 하여 각 변수 당 168개의 관측치를 사용하였다.

본 연구에서 식별한 16개 변수 중 케이프선 운임에 상당히 영향을 미치는 요인이 어떤 변수인지 알아보기 위해 단계적 회귀모형을 벤치마크모델로 하고 대응모델로 기계학습모델의 일종인 랜덤포레스트 모델을 적용하여 결과를 비교하였다. 많은 변수를 모델에 입력하면 모델의 복잡성으로 인해 계산에 상당한 시간이 필요하며 해석도 용이하지 않을 뿐만 아니라 과적합(overfitting) 문제를 노출하는 경우도 발생한다. 모델링에서 가장 중요한 작업은 ‘변수선택(feature selection)’에 관한 문제이다(Cai et al., 2018).

이 두 모델을 적용함에 있어 변수간의 통계적 상관성을 측정하는 필터방법(filter methods)이 아니라 여러 변수의 조합을 실제모델에 투입하여 모델의 성능에 기여도가 높은 변수를 식별하는 래퍼방법(wrapper methods)을 따랐다(Talavera, 2005; Jovic, Brkic and Bogunovic, 2015). 또한 기계학습 모델의 성능개선을 위해 식(1)을 적용하여 데이터를 정규화(normalization)하였다.(Li and Parsons, 1997)

$$\hat{y}_t = \frac{(y_t - y_{\min})}{(y_{\max} - y_{\min})} \quad (1)$$

또한 모델의 성능 검증을 위해 주어진 데이터를 무작위 추출을 하여 8:2로 나누어 훈련샘플과 검증샘플로 구분하였으며,

훈련단계에서 10-분할 교차검증(10-folds cross-validation)방법을 적용하여 모델의 파라미터를 조정하였다.

Table 2 Description of determinants

	Factor	Description
	Cape	Capesize 4TC rate(\$/day)
Supply	Cont	Capesize Contract(DWT)
	Deliv	Capesize Delivery(DWT)
	Demol	Capesize Demolition(DWT)
	Scrap	Capesize Scrap Value(m\$)
	Bunker	Singapore 380cst bunker price(\$/ton)
Demand	IronChnImp	China Iron Import(kton)
	ChnSteelProd	China Steel Production(kton)
	IronAusExp	Australia Iron Export(kton)
	IronBrzExp	Brazil Iron Export(kton)
	AusBrzRatio	Australia-Brazil Iron Export Ratio(%)
	CoalChnImp	China Steam Coal Import(kton)
	CoalIndExp	Indonesia Steam Coal Export(kton)
	CoalAusExp	Australia Steam Coal Export(kton)
	IndAusRatio	Indonesia-Australia Steam Coal Export Ratio(%)
	CoalJpnImp	Japan Steam Coal Import(kton)
	CoalKorImp	Korea Steam Coal Import(kton)

2.2 단계적 회귀분석(Stepwise Regression)

다중회귀모형에 여러 변수를 투입하는 경우에는 단계적 회귀분석이 적절한 접근법이다(Shepperd & MacDonell, 2012; Silhavy, Silhavy, & Prokopova, 2017). 수식으로는 식(2)와 같이 벡터로 다중회귀모형을 나타낼 수 있다.

$$y = X\beta + \varepsilon \tag{2}$$

파라미터는 OLS(ordinary least squares)추정으로 식(3)과 같이 구할 수 있다.

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{3}$$

단계적 회귀분석은 여러 설명변수(predictor)의 추가와 제거를 반복적으로 시행하여 예측오차가 가장 작은 최소변수집합(subset of variables)을 찾는 방법이다. 즉, 최소의 변수를 투입하여 최고의 성능을 보이는 효율적인 모델을 도출하는 것이다. 단계적 회귀분석에는 변수를 순차적으로 늘려가면서 성능을 비교하는 전진선택법(forward selection), 변수를 순차적으로 줄여가며 성능을 비교하는 후진선택법(backward selection) 그리고 전후진 선택법이 결합된 단계적 선택법(stepwise selection)이 있으며 본 연구에서는 단계적 선택법을 적용한 회귀 알고리즘을 사용하여 최적의 변수 조합을 찾았다.

2.3 랜덤포레스트(Random Forest)

Breiman (2001)이 결정나무(decision tree)모델과 배깅(bagging)을 결합한 RF모델을 제안하였다. 이 모델의 특징은 분류(classification)문제부터 회귀(regression)문제까지 광범위하게 적용될 수 있으며, 학습하는 속도가 다른 기계학습 모델에 비해 빠를 뿐만 아니라 조정할 파라미터도 적으며 특히 본 연구의 특성처럼 많은 차원을 계산해야 하는데 특히 뛰어난 성능을 보인다(Cutler, Cutler and Stevens, 2012). Fig. 4는 랜덤포레스트의 알고리즘을 도식화한 것이다.

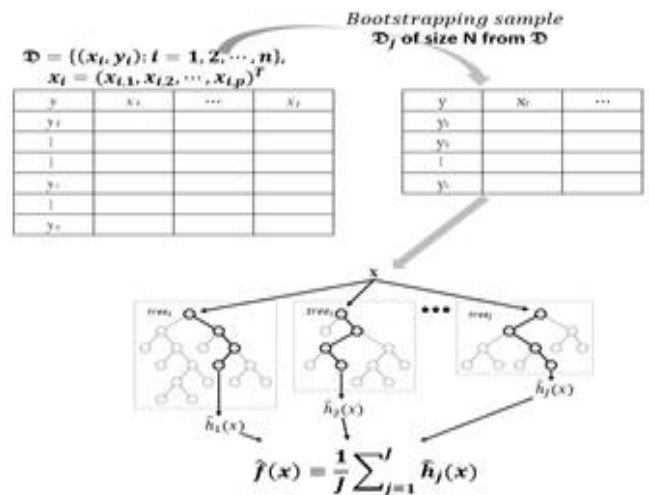


Fig. 4 Procedure of random forest

2.4 모델의 예측성능 평가기준

각 모델에서 선정된 변수들을 이용하여 예측한 결과와 실측치를 비교하여 선정된 요인들의 반영이 모델성능에 얼마나 영향을 미치는지 평가하고 한다. 성능검증기준은 아주 다양하지만 어느 기준이 성능평가지표로 채택되어야 하는지에 대한 일치된 의견은 없다. 하지만 Table 3에 제시된 기준들은 많은 연구들에서 일반적으로 받아들여지는 지표들이다(Zhang, Patuwo and Hu, 1998; Paliwal and Kumar, 2009). MAE와 RMSE는 평균오차를 나타내는 공통점이 있다. MAE는 큰 오차에 덜 민감하지만 RMSE는 큰 오차에 대해 패널티가 큰 특징이 있다. MAPE는 오차와 실측치의 비율로 나타낸 특징이 있지만 실측치가 0이면 값을 가질 수 없는 단점이 존재한다. COR는 예측치와 실측치의 상관계수를 나타낸 것이다. MAE, RMSE, MAPE는 수치가 낮을수록 우수한 성능이라고 해석하며 COR는 1에 가까울수록 예측력이 높다고 할 수 있다.

Table 3 Model performances

Criteria	Equation
MAE	$\frac{1}{n} \sum_{i=1}^n  (Y_i - \hat{Y}_i) $
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$
MAPE	$\frac{100}{n} \sum_{i=1}^n \frac{Y_i - \hat{Y}_i}{Y_i}$
COR	$\frac{Cov(Y, \hat{Y})}{\sigma_Y \sigma_{\hat{Y}}}$

### 3. 분석 결과

#### 3.1 단계적 회귀분석 결과

훈련샘플을 이용하여 10분할 교차검증과 단계적 변수조합을 시행한 결과 Fig. 5와 같이 최적변수가 8개일 때 모델의 성능이 가장 뛰어났다.

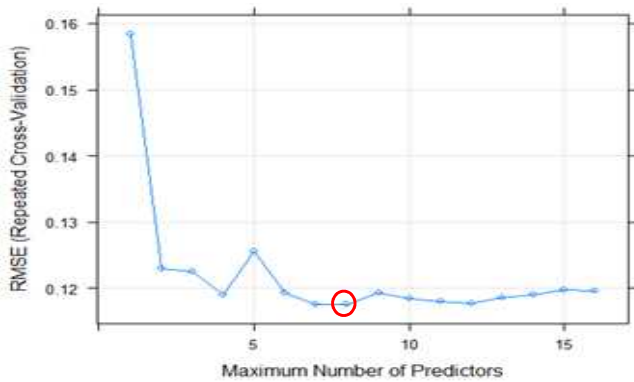


Fig. 5 The optimal number of predictors(LM)

Fig. 6은 모델의 성능개선에 가장 큰 영향을 미치는 요인을 보여주고 있는데 중요도순으로 인도네시아 석탄 수출량(CoalIndExp), 중국 철광석 수입량(ChnSteelProd), 중국 철강 생산량(ChnSteelProd), 한국 석탄 수입량(CoalKorImp), 케이프선 해체량(Demol), 호주 철광석 수출량(ChnSteelProd), 중국 석탄 수입량(CoalChnImp), 호주 석탄 수출량(CoalAusExp)이 차지하였다. 케이프선 운송서비스의 수요측면에서 상당히 많은 변수들이 중요도 순에서 상위권을 차지하였는데 단계적 회귀분석에서 중요하게 보는 요소는 공급보다는 수요요인에 비중을 두고 있음을 알 수 있다.

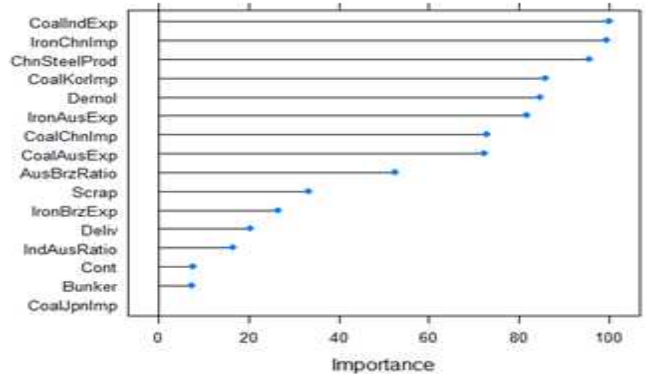


Fig. 6 Variable importance of LM

#### 3.2 랜덤포레스트 분석 결과

단계적 회귀분석의 훈련단계와 동일하게 10분할 교차검증을 시행한 결과 Fig. 7과 같이 최적변수가 10개일 때 모델의 성능이 가장 뛰어났다.

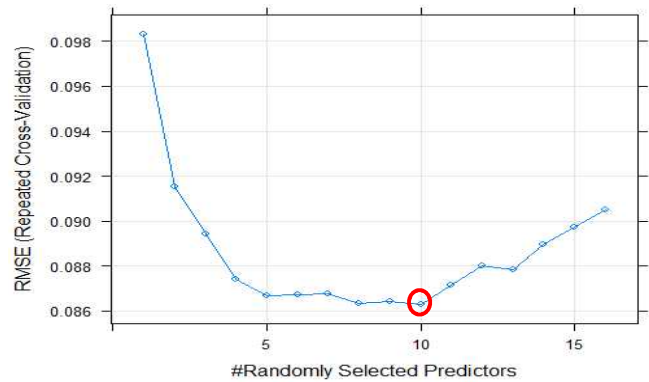


Fig. 7 The optimal number of predictors(RF)

Fig. 8은 모델의 성능개선에 가장 큰 영향을 미치는 요인을 보여주고 있는데 중요도순으로 케이프선 해체가(Scrap), 케이프선 해체량(Demol), 중국 철강 생산량(ChnSteelProd), 호주 철광석 수출량(ChnSteelProd), 한국 석탄 수입량(CoalKorImp), 중국 석탄 수입량(CoalChnImp), 인도네시아 석탄 수출량(CoalIndExp), 연료유가(Bunker), 일본 석탄 수입량(CoalJpnImp), 중국 철광석 수입량(ChnSteelProd)이 차지하였다. 단계적 회귀분석 결과와 달리 랜덤포레스트는 공급요인을 가장 중요한 요인으로 선정하였다.



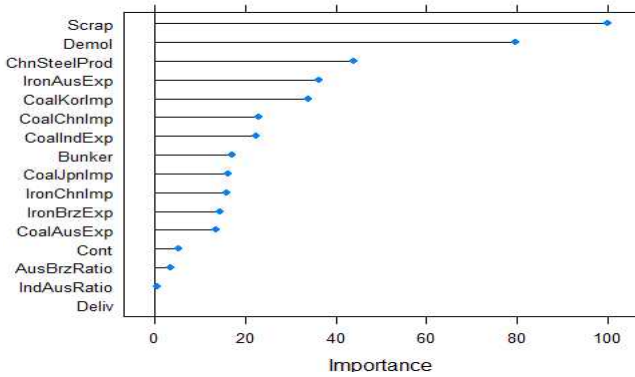


Fig. 8 Variable importance of RF

### 3.3 변수선택에 의한 예측결과 비교

각 모델에서 선정된 최적변수는 단계적 회귀분석 모델이 8개, 랜덤포레스트가 10개로, 이들을 입력변수로 하여 예측한 결과와 실측치를 비교하여 모델의 성능을 검증하였다. Table 4의 결과를 보면 모든 기준에서 랜덤포레스트 모형이 단계적 회귀모형보다 예측성능이 우수함을 알 수 있다.

Table 4 Model performances

Criterion	Random Forest	Stepwise Regression
MAE	<b>12,301.53</b>	16,606.95
RMSE	<b>22,290.20</b>	23,824.91
MAPE	<b>51.27</b>	71.54
COR	<b>0.92</b>	0.84

모델의 예측결과는 Fig. 9와 같이 그림으로 나타낼 수 있으며 변동성이 아주 큰 시점을 제외하면 랜덤포레스트 모델이 단계적 회귀분석보다 실측치에 더 유사함을 알 수 있다.

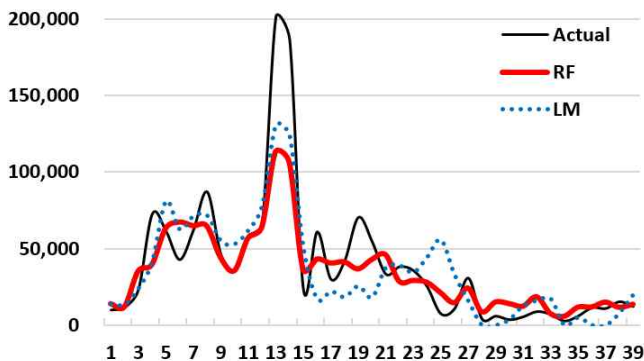


Fig. 9 Prediction results based on the optimal number of features

## 4. 결 론

해운시장의 심한 변동성과 수급요인들의 복잡성은 시황예측에 상당한 어려움을 주지만 해운업에 있어 예측의 중요성은 간과될 수 없기 때문에 학계나 산업계가 많은 관심을 가진다. 해운시황 예측에 있어서 세련된 모델을 이용하여 도출된 근거를 바탕으로 합리적이고 과학적인 의사결정을 할 수 있도록 기여하고자 본 연구를 진행하였다.

본 연구에서는 케이프션 운임예측에 랜덤포레스트를 적용하였으며 전통적인 계량모형인 단계적 회귀분석과 비교하였다. 케이프션 운임시장의 단순한 시장구조로 인해 수급요인들을 비교적 명확히 파악할 수 있었다. 사전 선정된 16개의 변수들에 대해 모델들의 알고리즘을 이용하여 최적변수를 도출하였다. 단계적 회귀모형은 케이프션 운임예측에 수요요인을 위주로 최적변수를 선정하였지만 랜덤포레스트모형은 공급요인을 선정하여 상이한 결과를 보였다. 선택된 변수들을 입력변수로 하여 추정된 예측치를 비교한 결과, 랜덤포레스트가 모든 성능지표를 기준으로 우수하였다. 케이프션 운임이 랜덤포레스트모형이 제시한 공급요인들에 민감하게 반응하기 때문인 것으로 해석될 수 있으며 이는 향후 실무적으로 시황을 예측할 때 어떤 요인을 비중 있게 모니터링할 것인가에 대한 가이드를 제공할 수 있을 것으로 기대된다.

이 연구에서는 비교적 단순한 케이프션 운임시장만을 대상으로 하였으나 향후 연구과제로 파나마크스, 수프라마크, 그리고 탱커시장의 VLCC, 수에즈막스 등 하위시장으로 확대 적용하여 모델의 일반화가 시도될 수 있을 것이다.

## References

- [1] Bakshi, G., Panayotov, G. and Skoulakis, G.(2011), "The Baltic Dry Index as a Predictor of Global Stock Returns, Commodity Returns, and Global Economic Activity," SSRN. <http://dx.doi.org/10.2139/ssrn.1787757>.
- [2] 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.
- [3] Batchelor, R., Alizadeh, A. and Visvikis, I.(2007), "Forecasting spot and forward prices in the international freight market," International Journal of Forecasting, Vol. 23, No. 1, pp. 101-114.
- [4] Breiman, L.(2001), "Random Forests," Machine Learning, Vol. 45, No. 1, pp. 5-32.
- [5] Cai, J. et al.(2018), "Feature selection in machine learning: A new perspective," Neurocomputing, 300, pp. 70-79.
- [6] Celik, M. et al.(2009), "An integrated fuzzy QFD model

- proposal on routing of shipping investment decisions in crude oil tanker market,” *Expert Systems with Applications*, Vol. 36, No. 2, pp. 6227 - 6235.
- [7] 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.
- [8] 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*, Vol. 36, No. 2, pp. 261-286.
- [9] Han, Q. et al.(2014), “Forecasting dry bulk freight index with improved SVM,” *Mathematical Problems in Engineering*, Vol. 2014(ID: 460684), pp. 1-12.
- [10] Jovic, A., Brkic, K. and Bogunovic, N.(2015), “A review of feature selection methods with applications,” in 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1200-1205.
- [11] Li, J. and Parsons, M. G.(1997), “Forecasting tanker freight rate using neural networks,” *Maritime Policy and Management*, Vol. 24, No. 1, pp. 9-30.
- [12] Lyridis, D. et al.(2004), “Forecasting Tanker Market Using Artificial Neural Networks,” *Maritime Economics & Logistics*, Vol. 6, pp. 93-108.
- [13] Paliwal, M. and Kumar, U. A.(2009), “Neural networks and statistical techniques: A review of applications,” *Expert Systems with Applications*, Vol. 36, No. 1, pp. 2-17.
- [14] Shepperd, M. and MacDonell, S.(2012), “Evaluating prediction systems in software project estimation,” *Information and Software Technology*, Vol. 54, No. 8, pp. 820-827.
- [15] Shin, C. H. and Jeong, S. H.(2011), “A Study on Application of ARIMA and Neural Networks for Time Series Forecasting of Port Traffic,” *Journal of Korean Navigation and Port Research*, Vol. 35, No. 1, pp. 83-91.
- [16] Shin, C. H., Kang, J. S., Park, S. N. and Lee, J. H.(2008), “A Study on the Forecast of Port Traffic using Hybrid ARIMA-Neural Network Model,” *Journal of Korean Navigation and Port Research*, Vol. 32, No. 1, pp. 81-88.
- [17] Silhavy, R., Silhavy, P. and Prokopova, Z.(2017), “Analysis and selection of a regression model for the Use Case Points method using a stepwise approach,” *Journal of Systems and Software*, Vol. 125, pp. 1-14.
- [18] Stopford, M.(2009), “*Maritime Economics*,” Routledge.
- [19] Talavera, L.(2005), “An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering,” in Famili A.F., Kok J.N., Peña J.M., Siebes A., F. A. (ed.) *Advances in Intelligent Data Analysis VI. IDA 2005*. Springer Berlin Heidelberg.
- [20] Veenstra, A. W. and Franses, P. H.(1997), “A co-integration approach to forecasting freight rates in the dry bulk shipping sector,” *Transportation Research Part A: Policy and Practice*, Vol. 31, No. 6, pp. 447-458.
- [21] Yang, Z., Jin, L. and Wang, M.(2011), “Forecasting Baltic Panamax Index with Support Vector Machine,” *Journal of Transportation Systems Engineering and Information Technology*, Vol. 11, No. 3, pp. 50-57.
- [22] Yun, H., Lim, S. and Lee, K.(2016), “Trading Strategies in Bulk Shipping: the Application of Artificial Neural Networks,” *Journal of Korean Navigation and Port Research*, Vol. 40, No. 5, pp. 337-343.
- [23] Zeng, Q. et al.(2016), “A new approach for Baltic Dry Index forecasting based on empirical mode decomposition and neural networks,” *Maritime Economics & Logistics*, Vol. 18, No. 2, pp. 192-210.
- [24] Zeng, Q. and Qu, C.(2014), “An approach for Baltic Dry Index analysis based on empirical mode decomposition,” *Maritime Policy & Management*. Routledge, Vol. 41, No. 3, pp. 224-240.
- [25] Zhang, G., Patuwo, B. E. and Hu, M. Y.(1998), “Forecasting with artificial neural networks: The state of the art,” *International Journal of Forecasting*, Vol. 14, No. 1, pp. 35-62.

---

Received 27 August 2018

Revised 4 October 2018

Accepted 4 October 2018