

# 인공지능기반 해상운임예측 연구

---

A Study on the Prediction of Dry Bulk Freight Rate  
based on Long Term Short Term Memory

---

황수진 • 고병욱 • 최건우 • 안영균 • 김병주

2020. 12.



한국해양수산개발원  
KOREA MARITIME INSTITUTE

저자 |

황수진, 고병욱, 최건우, 안영균, 김병주

연구진 |

연구책임자 황수진 한국해양수산개발원 해운물류연구본부 전문연구원  
공동연구원 고병욱 한국해양수산개발원 해운물류연구본부 부연구위원  
공동연구원 최건우 한국해양수산개발원 해운물류연구본부 전문연구원  
공동연구원 안영균 한국해양수산개발원 해운물류연구본부 전문연구원  
공동연구원 김병주 한국해양수산개발원 해운물류연구본부 연구원

연구기간 |

2020. 04. 13. ~ 2020. 10. 12.

보고서 집필내역 |

연구책임자 |

황수진 연구총괄, 제1장 제1절~제3절 일부, 제2장 제1절~제3절 일부,  
제3장, 제4장, 제5장 일부

내부연구진 |

고병욱 제1장 제1절~제3절 일부  
최건우 제5장 일부  
안영균 제2장 제1절~제3절 일부  
김병주 제2장 제1절~제3절 일부

# 발간사

2008년 9월 15일 리먼사태로 촉발된 글로벌 금융위기에서부터 시작된 해운경기의 위축이 장기화되고 있습니다. 2017년에 국내 1위이자, 전 세계 컨테이너 운영 선사 중 TOP 10 안에 항상 이름을 올리던 컨테이너 국적 선사 한진해운의 파산으로 결국, 국가경제가 큰 타격을 입었습니다. 이로 인해 그 어느 때보다 국가 기간산업으로써의 해운산업에 대한 국민적 관심이 높아졌으며, 정부 지원의 필요성이 강조되고 있습니다.

당시 비전문 경영진의 영업전략 수립 실패가 한진해운의 파산 사태가 발생한 원인 중 하나로 지적되었으며, 영업전략은 기업 입장에서 사활을 건 중요한 사안임에도 불구하고, 경험과 직관에 의존해 전략을 수립했다는 평가였습니다. 영업전략은 해운시황을 전망한 결과를 토대로 수립되어야만, 기업 입장에서 시장수익가능성에 따라 수익 및 위험을 합리적으로 관리할 수 있기 때문에 해운시황 전망, 즉 해상운임예측은 기업 입장에서 중요한 의미를 가집니다. 그런 이유로 해상운임예측의 필요성이 더욱 강조되고 있습니다.

2019년 12월 말에 발생한 코로나19라는 예기치 못한 상황으로 전 세계 경기는 또다시 위기에 직면했으며, 해운경기도 예외는 아닙니다. 코로나19로 인해 산업의 디지털 혁신이 가속화되고 있으며, 4차 산업혁

명의 핵심기술인 인공지능, 클라우드, 빅데이터 등의 기술이 융합되어 일상의 변화를 가져오고 있습니다. 해상운임예측 연구도 기존의 방법에서 벗어나 새로운 방법론을 적용해 나아가야 할 것입니다.

이번 연구는 이러한 시대적 흐름에 대응하여 인공지능기반 해상운임예측 모델을 구축하기 위한 복수의 연구보고서(Series) 중 하나입니다. 이번 보고서는 운임의 데이터 특성을 활용할 수 있는 인공지능 알고리즘을 사용하여 모델을 구축함으로써, 앞서 수행된 연구와 차별화됩니다. 이번 연구가 가속화되고 있는 디지털 시대에 대응해 해운기업이 데이터기반의 영업전략을 수립하고, 나아가서는 정부가 데이터기반의 정책을 수립하는 데 도움이 될 수 있기를 기대합니다. 아울러 연구를 수행하는 데 도움을 주신 분들께 감사의 말씀을 전합니다.

2020년 12월  
한국해양수산개발원  
원장 장 영 태

# 목차

요 약	i
-----	---

Executive Summary	ix
-------------------	----

제 1장 서론	1
---------	---

제1절 연구의 배경 및 목적	1
-----------------	---

1. 연구의 배경	1
-----------	---

2. 연구의 목적	3
-----------	---

제2절 연구의 내용 및 방법	4
-----------------	---

1. 연구 범위 및 주요 내용	4
------------------	---

2. 연구 방법	5
----------	---

제3절 선행연구 및 연구의 차별성	6
--------------------	---

1. 선행연구	6
---------	---

2. 연구의 차별성	12
------------	----

제 2장 데이터 및 변수 선정	13
------------------	----

제1절 데이터	13
---------	----

제2절 데이터의 전처리	14
--------------	----

1. 기술적 분석	15
-----------	----

2. 시계열 분해	26
-----------	----

제3절 변수탐색 및 선정	27
---------------	----

1. 기술적 변수 선정	27
--------------	----

2. 기술적 통계 분석	28
--------------	----

<b>제 3장 순환신경망 모델</b>	<b>33</b>
제1절 순환신경망 모델의 고찰	33
1. RNN 개념 및 구조	34
2. LSTM 개념 및 구조	35
제2절 모델구조 및 학습방법	39
1. 모델 구축 과정	39
2. 학습데이터	40
3. 모델구조	40
4. 학습방법	42
<b>제 4장 분석결과</b>	<b>43</b>
제1절 순환신경망 모델 분석결과	43
1. 모델 평가 지표	43
2. 모델 분석결과	44
제2절 모델성능비교	50
1. 비교 모델	50
2. 모델성능비교 결과	51
<b>제 5장 결과</b>	<b>57</b>
제1절 연구결과 요약	57
1. 순환신경망 모델 분석결과	57
2. 비교 모델 분석결과	57
제2절 연구의 기여	59
제3절 추가연구	60
<b>참고문헌</b>	<b>63</b>

# 표목차

〈표 1-1〉 선행연구 요약 .....	10
〈표 2-1〉 데이터 요약 .....	14
〈표 2-2〉 기술적 지표 .....	15
〈표 2-3〉 추세 지표 .....	16
〈표 2-4〉 모멘텀 지표 .....	19
〈표 2-5〉 기타 지표 .....	24
〈표 2-6〉 기술적 변수 요약 .....	27
〈표 2-7〉 케이프 기술통계 .....	29
〈표 2-8〉 파나마스 기술통계 .....	30
〈표 2-9〉 수프라마스 기술통계 .....	31
〈표 4-1〉 데이터 구성에 따른 LSTM기반 케이프선 스팟운임 예측 모델의 성능 지표 비교 .....	47
〈표 4-2〉 데이터 구성에 따른 LSTM기반 파나마스선 스팟운임 예측 모델의 성능 지표 비교 .....	49
〈표 4-3〉 데이터 구성에 따른 LSTM기반 수프라마스선 스팟운임 예측 모델의 성능 지표 비교 .....	50
〈표 4-4〉 알고리즘 및 데이터 구성 조합 별 케이프선 스팟운임 예측 모델의 성능 지표 비교 .....	52
〈표 4-5〉 알고리즘 및 데이터 구성 조합 별파나마스선 스팟운임 예측 모델의 성능 지표 비교 .....	54
〈표 4-6〉 알고리즘 및 데이터 구성 조합 별 수프라마스선 스팟운임 예측 모델의 성능 지표 비교 .....	55
〈표 5-1〉 선형별 모델성능비교(RMSE값) .....	58
〈표 5-2〉 선형별 모델성능비교(MAPE값) .....	59

## 그림 목차

〈그림 2-1〉 데이터 전처리 과정 .....	14
〈그림 2-2〉 케이프션 운임과 장단기이동평균선 .....	17
〈그림 2-3〉 케이프션 운임과 MACD .....	18
〈그림 2-4〉 케이프션 운임과 모멘텀 지표 .....	21
〈그림 2-5〉 케이프션 운임과 변화율 지표 .....	21
〈그림 2-6〉 케이프션 운임과 상대강도지수 .....	22
〈그림 2-7〉 케이프션 운임과 윌리엄스 %R .....	23
〈그림 2-8〉 케이프션 운임과 스토캐스틱 .....	24
〈그림 3-1〉 인공신경망 모델 .....	34
〈그림 3-2〉 순환신경망 모델 .....	35
〈그림 3-3〉 LSTM 모델 .....	36
〈그림 3-4〉 LSTM 모델의 셀 구조 .....	38
〈그림 3-5〉 LSTM 모델의 구축과정 .....	39
〈그림 4-1〉 LSTM 모델의 학습 곡선 .....	45
〈그림 4-2〉 LSTM_ts 모델의 학습 곡선 .....	45
〈그림 4-3〉 LSTM 모델 예측 결과 (케이프션 스팟운임) .....	46
〈그림 4-4〉 LSTM 모델 성능 비교 (케이프션 스팟운임) .....	47
〈그림 4-5〉 LSTM 모델 예측 결과 (파나막스선 스팟운임) .....	48
〈그림 4-6〉 LSTM 모델 성능 비교 (파나막스선 스팟운임) .....	48
〈그림 4-7〉 LSTM 모델 예측 결과 (수프라막스선 스팟운임) .....	49
〈그림 4-8〉 LSTM 모델 성능 비교 (수프라막스선 스팟운임) .....	50
〈그림 4-9〉 DNN 모델과 예측 성능 비교(케이프션 스팟운임) .....	51
〈그림 4-10〉 지표 활용 모델의 예측성능 비교(케이프션 스팟운임) .....	52
〈그림 4-11〉 DNN 모델과 예측성능 비교 (파나막스선 스팟운임) .....	53
〈그림 4-12〉 지표 활용 모델의 예측성능 비교(파나막스선 스팟운임) .....	53
〈그림 4-13〉 DNN 모델과 예측성능 비교(수프라막스선 스팟운임) .....	54
〈그림 4-14〉 지표 활용 모델의 예측성능 비교 (수프라막스선 스팟운임) .....	55
〈그림 5-1〉 선형별 모델성능 비교(RMSE값) .....	58
〈그림 5-2〉 선형별 모델성능 비교(MAPE값) .....	59



## 요약

# 인공지능기반 해상운임예측 연구

## 1. 연구 목적

- ▶ 본 연구의 목적은 해운기업이 데이터 기반의 영업전략을 수립할 수 있도록 인공지능기반 해상운임예측 모델을 구축하는 것임
  - 기업입장에서는 반복되는 해운위기방지의 대안으로 시장수익가능성에 따라 수익 및 위험을 합리적으로 관리할 수 있어 해상운임예측은 중요한 의미를 가짐
  - 해상운임을 예측하는 연구는 과거에서 현재까지 계속 진행 중이나, 기존 모형은 대부분 회귀모형이나 계량 분석 방법에 의존하고 있음
  - 기존의 방법론은 예측 결과에 대해 설명 가능하다는 장점을 가지고 있는 반면, 정규성, 다중공선성과 같은 데이터의 특성을 가정해야 한다는 제약이 존재함
  - 새로운 방법론을 적용한 해상운임예측모델에 대한 요구가

- ▶ LSTM(Long Short term Memory)을 활용하여 예측 모델을 구축하고, 이전 연구에서 제시되었던 딥러닝 알고리즘과 비교한 결과를 제시하고자 함
  - 이번 연구에 앞서 딥러닝 기반 해상운임예측 모델을 구축하여, 기업에게 주간단위 운임예측 서비스를 제공 중에 있음
  - 새로운 방법론인 LSTM을 적용한 모델을 구축함으로써 제공되는 서비스의 운임예측 성능을 향상시키고자 함

## 2. 연구 특징 및 방법

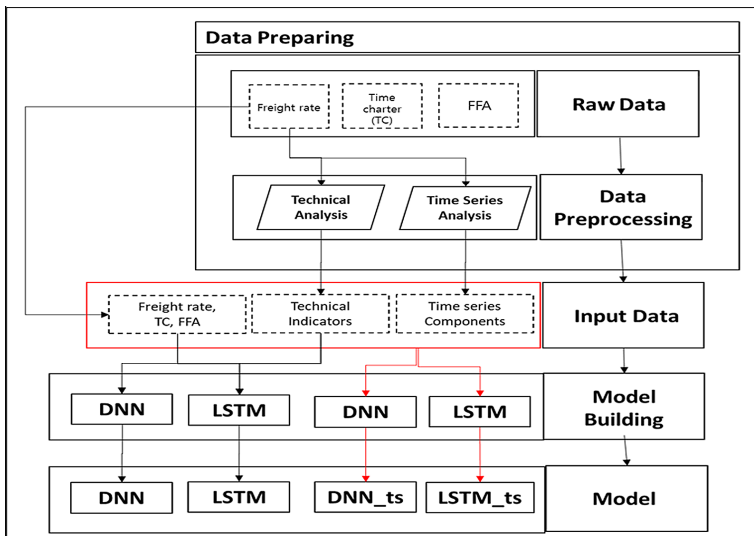
- ▶ 이번 연구에서는 LSTM 알고리즘을 이용함으로써 해상운임의 시계열적 특성을 활용하여 예측 모델을 구축함
  - LSTM은 인공지능망 알고리즘 중 하나로 데이터의 시계열적 특성을 활용할 수 있도록 고안됨
- ▶ 해상운임의 예측성능을 향상시키기 위해 모델 구축 과정에서 기술적 분석 및 시계열 성분 분해법으로 운임의 특성을 추출하여 입력변수로 활용함
  - 기술적 분석은 주로 주가예측에 활용되는 방법으로, 매도시점을 파악하기 위해 주가의 추세, 모멘텀, 변곡점을 확인할 수 있도록 고안되었음
  - 시계열 성분 분해는 정상성을 확보한 데이터의 추세, 순환, 계절성, 불규칙성을 추출하는 방법임

〈요약 표 1〉 연구 방법

연구흐름	세부내용	연구 방법
서론	<ul style="list-style-type: none"> <li>• 연구의 배경 및 목적</li> <li>• 선행연구 및 차별성</li> </ul>	선행연구 분석
데이터 및 변수 선정	<ul style="list-style-type: none"> <li>• 데이터 설명</li> <li>• 기술적 분석</li> <li>• 시계열 성분 분해</li> </ul>	선행연구 기술적 분석 시계열 성분 분해
순환신경망 모델	<ul style="list-style-type: none"> <li>• 순환신경망 모델 개요</li> <li>• 모델 구조 및 과정</li> </ul>	선행연구
분석결과	<ul style="list-style-type: none"> <li>• 순환신경망 모델의 분석결과 제시</li> <li>• DNN 모델과 성능비교</li> </ul>	분석
결론	<ul style="list-style-type: none"> <li>• 결론</li> </ul>	-

자료: 저자 작성

〈요약 그림 1〉 모델 구축 과정



자료: 저자 작성

### 3. 연구 결과

#### 1) 연구 결과 요약

- ▶ 건화물선의 선형별 운임, 용선료, FFA가격을 입력변수로 활용했으며, 특히 운임의 기술적 분석 및 시계열 성분 분석을 통해 추출한 특성을 입력변수로 활용해 LSTM기반 운임예측 모델을 구축함
  - 제시된 모델은 공통적으로 운임, 용선료, FFA가격을 변수로 활용하였으며, 또한 운임의 기술적 분석을 통해 추출된 지표도 입력변수로 활용함
  - 추가적으로 앞에서 언급한 변수 이외에 시계열 성분 분해를 통해 추출된 추세와 계절성, 불규칙성분을 입력변수로 학습한 모델도 구축함
- ▶ 비교모델로 동일한 입력변수를 활용한 DNN기반 운임예측 모델을 제시하고, 모델간의 성능은 평균제곱근편차, 평균절대백분율편차, 상관계수의 모형성능지표를 활용하여 비교하여 그 결과를 제시함

〈요약 표 2〉 모델

공통변수	추가 변수	인공지능 알고리즘	모델
운임, 용선료, FFA, 운임의 기술적 분석 지표	-	LSTM	LSTM
		DNN	DNN
	운임의 시계열 분해 성분 (추세, 계절, 불규칙성)	LSTM	LSTM_ts
		DNN	DNN_ts

자료: 저자 작성

- ▶ LSTM기반 해상운임예측 모형이 기존의 DNN기반 모형보다 성능이 향상된 것으로 나타났으며, 특히, 시계열 분해 성분을 입력변수로 활용하면 모형의 성능이 크게 개선되는 것을 확인함

〈요약 표 3〉 모형 성능 비교

모형	모형 성능 지표		
	RMSE	MAPE	CORR
LSTM_ts	687.23	4.96	0.9966
DNN_ts	753.61	4.70	0.9956
LSTM	1,201.52	7.39	0.9884
DNN	1,576.28	9.89	0.9843

자료: 저자 작성

- ▶ 이번 연구는 인공지능 알고리즘 중 고전적이라고 할 수 있는 딥러닝과 순환신경망 모형을 활용해 건화물선의 해상운임예측 모형을 구축함
- ▶ 구축된 모형의 예측 기간은 주간단위로 한정했음. 이는 건화물선 시장이 부정기선 시장이라는 특성을 감안했을 때, 기업의 단기 수익을 올리기 위한 기초 데이터로 활용 가능할 것으로 판단됨
- ▶ 하지만, 주간단위의 예측 기간은 기업의 의사결정을 내리기 위한 기초 데이터로의 활용하기에 그 범위가 제약적임
- ▶ 모형의 예측 성능 향상 및 예측 기간을 늘리기 위해 새롭게 제시되거나 개선된 인공지능 알고리즘의 활용과 수급 데이터의 사용을 검토하여 모형을 구축함으로써 활용범위를 확대할 필요가 있음

## 2) 정책대안 제시내용 및 정책화 활동

- ▶ 이번 연구에서 구축된 인공지능기반 단기운임예측 모델의 결과를 해운기업의 영업전략, 정부의 정책 수립에 활용될 수 있도록 제공함
  - 한진사태의 원인 중 하나가 주관적인 경험과 직관에 의존한 영업전략 수립 때문이며, 데이터에 근거한 전략 수립 환경이 구축되어야 한다는 필요성이 제기됨
  - 전망에 활용할 수 있는 기초 데이터는 해외 유수기관에서 제공하는 데이터를 제외하면 그 수가 부족한 상황임
  - 전망치는 적용된 분석방법에 따라 도출된 결과가 달라지며, 기업의 영업전략 및 정부의 정책수립을 위해서는 다양한 데이터가 필요함
- ▶ 인공지능기반 운임예측 모델 및 결과의 실용성과 활용성 제고를 위한 산·학·연 공동 연구 추진 및 협의회 구성 및 정례화
  - 대학과의 공동연구 추진을 통해 인공지능 알고리즘 고도화 및 해운업계의 의견수렴을 위한 협의체 구성·운영을 통한 도메인 지식 기반 모델 개발 추진
  - 해운특화 운임전망데이터의 공공성 및 활용성 제고 방안 수립
  - 해운기업 맞춤 정보제공이 지속적, 정기적으로 가능하도록 정보제공서비스 체계 마련
  - 정부 및 업계의 제공된 데이터 활용도 제고를 위한 모델 개념 및 제공데이터의 활용방안 설명회 개최

### 3) 정책적 기여 등 기대효과

- ▶ 이번 연구는 변동성이 큰 건화물선 운임시장의 단기 예측 모델을 구축함으로써 해운기업이 데이터기반의 영업전략 수립할 수 있도록 환경을 마련함
- ▶ 해운특화 데이터기반 영업전략 수립을 통해 세계 시장에서의 우리 기업의 경쟁력을 확보에 기여할 것으로 기대됨
- ▶ 정부의 4차 산업혁명에 대한 다양한 전략과 성과가 나오는 시점이며, 해운부분에서 단기 성과로써 의미가 있을 것으로 기대됨
  - 해운부분에서 4차 산업혁명의 핵심기술이 적용되고 있는 분야는 항만 등 인프라 시설의 스마트화, 플랫폼 구축 등과 같은 장기적인 관점의 투자에만 집중됨
  - 단기적-장기적 목표간의 균형이 필요하며, 이번 연구가 인공지능을 적용한 운임 예측모델을 구축하고 그 정보를 기업에게 제공한다는 점에서 4차 산업혁명에서의 해운부분의 단기적 성과로써 의미가 있을 것으로 기대됨
- ▶ 해운기업이 활용할 수 있는 인공지능기반 운임 예측모델 구축 및 데이터 수를 다양화함으로써, 해운기업의 연구비용 절감 및 정보 격차 해소에 기여할 것으로 기대됨
  - 도메인 부분이 아닌 새로운 분야의 방법론을 적용하여 모델을 구축하는 데는 시간, 예산이 필요함
  - 해운분야에서 매출수익이 높은 기업의 경우는 예산을 확보하여 데이터구매, 전문 인력을 확보한 모델 구축 및 분석이 가능하나, 대부분의 기업은 활용의 가능성이 높을 것이라는 이론적 가정에 투자한다는 것은 현실적으로 쉽지 않음





# A Study on the Prediction of Dry Bulk Freight Rate based on Long Short Term Memory

Su-Jin Hwang · Byoung-Wook Ko · Geon-U Choi ·  
Yeong-Gyun An · Byeong-Ju Kim

## 1. Purpose

- ▶ This study aims to establish a prediction model for ocean freight rates based on Long Short Term Memory (LSTM), which allows shipping companies to develop a data-based business strategy.
- Prediction of ocean freight rates bears significance from the perspective of companies. As an effective alternative for preventing recurring shipping crisis, this capacity allows for reasonable management of profits and risks in accordance with market profitability.
- Studies on predicting the freight rates have been continued from the past, however, most of the existing models relied on regression models and quantitative analysis methods.
- While having an advantage of the ability to explain on prediction results, existing methodologies possess limitations of making certain assumptions on the features of data including normality and multicollinearity.

- ▶ This study intends to develop a prediction model by utilizing LSTM (Long Short Term Memory), presenting the result after comparing it with deep-learning algorithm proposed in the previous study.
- Prior to this study, a deep-learning based freight rate prediction model was established, offering shipping companies with a weekly based service for forecasting freight rates.
- This study aims to improve the capacity of existing freight rate prediction by establishing a LSTM applied model, which is a new methodology.

## 2. Methodology and Feature

- ▶ Utilizing a LSTM-based algorithm, this study has developed a prediction model in which the time series features of freight rates have been applied.
- As one algorithm of artificial neural networks, LSTM is designed to utilize the time series features of data.
- ▶ During the process of establishing the model, features of freight rates were extracted through the methods of technical analysis and time series decomposition and used them as input variables in order to boost the predictability of freight rates.
- Generally used in forecasting stock price, technical analysis is designed to identify the general trend of stock price, momentum and tipping point in an effort to determine when to sell a stock.

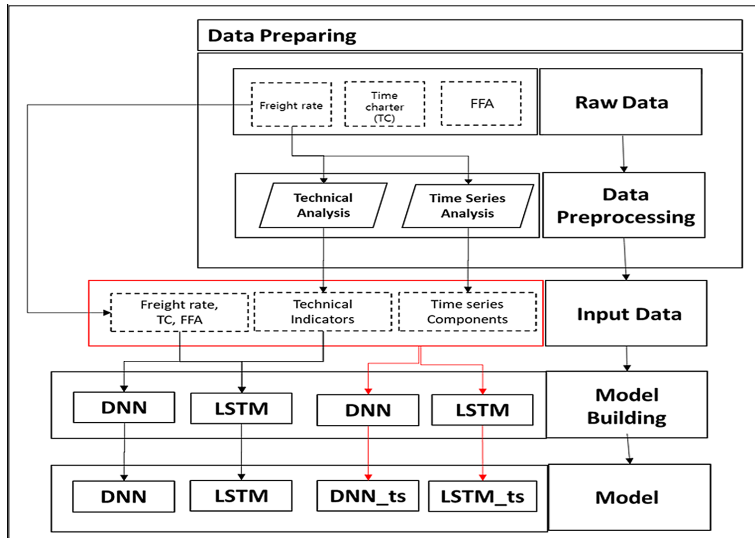
- Time series decomposition is intended to extract the trend, cycle, seasonality and irregularity of that data that possesses normality.

〈Summary Table 1〉 Methodology of the study

Research flow	Specific content	Method
↓	↓	↓
Introduction	<ul style="list-style-type: none"><li>• Background and purpose of the study</li><li>• Preceding studies and differentiation</li></ul>	Analysis of preceding studies
↓	↓	↓
Selecting data and variables	<ul style="list-style-type: none"><li>• Data explanation</li><li>• Technical analysis</li><li>• Time series decomposition</li></ul>	<ul style="list-style-type: none"><li>• Preceding studies</li><li>• Technical analysis</li><li>• Time series decomposition</li></ul>
↓	↓	↓
Model of Recurrent Neural Network (RNN)	<ul style="list-style-type: none"><li>• Overview of RNN models</li><li>• Model structure and process</li></ul>	Preceding studies
↓	↓	↓
Result of the analysis	<ul style="list-style-type: none"><li>• Presenting the analysis result of RNN models</li><li>• Performance comparison with DNN models</li></ul>	Analysis
↓	↓	↓
Conclusion	<ul style="list-style-type: none"><li>• Conclusion</li></ul>	-



〈Summary Figure 1〉 Flow diagram of model construction



### 3. Result

#### 1) Summary

- ▶ This study utilized the freight rate, charterage and FFA price of dry bulk ships by ship types, specifically the characteristics extracted by the technical analysis and time-series decomposition, as input variables to establish the freight rate prediction model based on LSTM.
- The presented models commonly used the freight rate, charterage and FFA price as variables as well as the indicators extracted through the technical analysis on the freight rate.
- In addition to variables discussed above, this study also developed a model which learnt the trend, seasonality and irregular

components extracted through the time-series decomposition as input variables.

- For a comparative model, the freight rate prediction model based on DDN with the same input variables was presented, and the model performance indicators of the root-mean-square deviation, mean absolute percentage deviation and correlation coefficient were utilized to compare performance between models.

〈Summary Table 2〉 Models

Common variables	Additional variables	AI algorithm	Model
Freight rates, charterage, FFA Technical analysis index of freight rates	—	LSTM	LSTM
		DNN	DNN
	Time series decomposition components of freight rates (Trend, seasonality, irregularity)	LSTM	LSTM_ts
		DNN	DNN_ts

- It is shown that the dry bulk freight rate prediction model based on LSTM has had better performance than the DNN-based model, and specifically, the performance of the model has been significantly improved when using the components from the time-series decomposition as input variables.

〈Summary Table 3〉 Performance comparison of models

Model	Performance index of models		
	RMSE	MAPE	CORR
LSTM_ts	687.23	4.96	0.9966
DNN_ts	753.61	4.70	0.9956
LSTM	1,201.52	7.39	0.9884
DNN	1,576.28	9.89	0.9843

- ▶ This study utilized the deep-learning and recurrent neural network models, which can be called classical among the artificial neural network algorithms, to establish the dry bulk freight rate prediction model.
- ▶ The prediction period for the model established in this study was limited to a weekly unit. When taking into consideration that the dry bulk ship market is the tramp liner market, this could be utilized as basic data to improve short-term earnings of companies.
- ▶ However, the weekly prediction period has a limited scope to be used as basic data for the decision-making of companies.
- ▶ It is required to expand its application scope by utilizing the LSTM algorithm which has been newly presented and enhanced in order to improve the prediction performance and period of the model and by considering the use of supply-demand data.

## 2) Policy suggestions and relevant activities

- ▶ This study provided the results of the short-term freight rate

prediction model based on LSTM so that they could be utilized for the sales strategies of shipping companies and the policy-making process of the government.

- Since it was considered that one of the causes of the Hanjin Shipping's bankruptcy was the establishment of sales strategies which relied only on subjective experiences and intuition, there raised the need of creating an environment where data-based strategies could be established.
- Currently, there are insufficient basic data which could be used for prediction, other than those provided by prominent overseas organizations.
- Estimates vary depending on the analysis methods applied, and there need various data in order to establish the sales strategies of companies and the policies of the government.
- ▶ It is required to promote the industry-academia-research joint research and form and hold a regular industry-academia-research meeting in order to improve the practicality and applicability of the LSTM-based freight rate prediction model and its results.
- It is required to create a more sophisticated LSTM algorithm through joint researches with universities and to work on the development of a domain knowledge-based model by forming and operating a consultative body to accommodate the opinions of the shipping industry.
- It is required to establish a plan to improve the public characteristics and applicability of the freight rate prediction

data specializing in the shipping industry.

- It is required to prepare the information-providing service so that information tailored to shipping companies could be provided in a continuous and regular way.
- It is required to hold a briefing session to explain the concept of the model and a way of utilizing the provided data in order to improve the applicability of data from the government and the industry.

### **3) Expected benefits including policy contribution**

- ▶ This study prepared an environment where shipping companies could establish data-based sales strategies by building the short-term freight rate prediction model for the dry bulk ship market with high variability.
- ▶ It is expected that the establishment of sales strategies based on the shipping market-specialized data could contribute to securing the competitiveness of Korean companies in the global market.
- ▶ Since now is the time when various strategies regarding the 4<sup>th</sup> industrial revolution by the government and their results are coming out, it is expected that they have a significance as a short-term performance in the shipping area.
- The areas with the core technologies of the 4<sup>th</sup> industrial revolution applied in the shipping field focus only on the long-term investments such as the smartification and platform-



building of infrastructure facilities, including ports

- It is required to strike a balance between short-term and long term objectives, and this study is expected to have significance as a short-term performance in the shipping area in terms of the 4<sup>th</sup> industrial revolution since it built the freight prediction model based on LSTM and provided its data for companies.
- ▶ By establishing a LSTM-based freight rate prediction model with which shipping companies are able to apply as well as diversifying the number of data, this study is expected to contribute to reducing research cost of shipping companies and addressing information gap.
- The application of a methodology in new areas other than domain requires time and budget for constructing a model.
- Companies with high sales profit in the shipping industry are able to secure the required budget to purchase the data, establish and analyze the model by professionals. However, it is not realistically feasible for most companies to put investment under the theoretical assumption of having high utilization.



## 제 1 장

## 서 론

## 제1절 연구의 배경 및 목적

## 1. 연구의 배경

반복되는 해운위기 방지의 대안으로 해상운임예측 모델의 필요성이 대두되고 있다. 해상운임예측은 기업 입장에서 시장수익가능성에 따라 수익 및 위험을 합리적으로 관리할 수 있다는 점에서 중요한 의미를 가진다. 해운기업은 해운시황의 향후 방향성을 예측하고, 시장 수익가능성에 따라 선박투자 여부, 용대선 비율 및 기간을 조정하는 등의 영업 전략을 수립할 수 있기 때문이다. 선박투자는 대규모의 자금이 필요하며, 일단 투자되면 그 선박이 시장에서 운영되는 한 경영 성과에 큰 영향을 미치기 때문에 향후 해운시황을 예측하는 일은 기업에게 중요한 과제가 된다. 기업의 미래수익과 직결된다는 측면에서 운임을 예측하는 능력은 해운기업이 반드시 확보해야 하는 중요한 역량이다.

우리나라 해운위기 반복의 원인 중 하나로 해운기업의 영업전략 수립 과정에서 시장의 높은 변동성을 고려하지 못한 것이 지목되었다. 해상운임예측이 기업에게 중요한 의미를 가지고 있음에도 불구하고, 대부분의 기업들이 경험과 직관에 의존하거나, 해외기관에서 발표한 시황전망 보고서 내용을 참고하고 있는 것으로 나타났다. 2000년대 중반 해운의 초호황기 시점에서 신조선박을 고가에 대량 발주하거나 장기용선 했던 해운기업이 세계 금융위기 발생으로 해운 경기가 불황으로 급전되자 심각한 경영위기에 봉착했다. 일부 기업은 파산 지경에 이르렀는데, 이는 기업이 영업전략 수립 시에 시황의 변화를 제대로 예측하지 못했기 때문이라는 의견이 제기되었다.

또한, 우리나라 해운분야에서 대기업인 한진해운이 경영난으로 결국 파산하게 되었는데, 파산의 원인 중 하나로 해상운임예측 능력부족이 거론되었다. 이러한 위기를 계기로 해상운임예측 모델 구축의 필요성이 더욱 부각되고 있는데, 이는 구축된 모델을 통해 해운시장의 향후 전개방향을 전망하고 상황에 따른 시나리오를 설계함으로써 대응방안을 마련하고자 하는 정부와 기업의 요구가 강해지고 있기 때문이다.

운임예측을 위해 시계열 모델로 ARIMA, VAR, VECM 모델 등이 사용되어 왔다. 시계열 모델은 장기운임예측보다는 단기운임을 예측하는 데 적합한 것으로 나타났다. 또한 동 모델들은 많은 양의 자료가 필요하며 신뢰구간에 대한 폭이 과대 추정될 가능성이 있다. 또한 불안정한 시계열을 활용하여 회귀분석을 실시할 경우 가성회귀의 결과를 초래할 수 있다는 점이 지적되고 있다. 또한 변수 간의 다중공선성(multicollinearity) 문제가 발생되면 차분을 통해 안정성을 확보할 수 있으나 이 과정에서 시계열의 고유 정보가 소실될 수 있다는 문제

점이 있다. 문제는 대부분 해상운임과 관련한 데이터는 불안정 시계열을 나타낸다는 점이다. 4차 산업혁명의 진전으로 데이터를 수집, 분석하기 위한 새로운 환경이 조성됨과 동시에 활용할 수 있는 데이터의 범위가 넓어졌다. 또한 기계학습(machine learning)기반의 데이터 분석 툴이 재조명되기 시작했다. 그중 가장 대표적 기술이 바로 인공지능(AI: Artificial Intelligence)이다. 인공지능이란 인간이 사고하는 사고과정을 기계가 재현해 낼 수 있도록 하는 개념이다. 이러한 과정을 재현해내기 위해 많은 기술적 방법론들이 등장하기 시작하였으며, 최근 들어 인공지능 툴에 대한 관심이 높아지면서 해운분야에도 이와 관련한 연구들이 꾸준히 진행 중이다. 하지만 기업과 정부 등의 요구에도 불구하고 현재까지 인공지능망을 활용해 해상운임을 예측한 선행연구의 수가 그리 많지 않은 실정이다.

## 2. 연구의 목적

본 연구는 해운기업이 데이터기반의 영업전략을 수립할 수 있도록 인공지능을 활용한 해상운임예측 모델을 개발하는 데 목적을 두고 있다. 해상운임예측 모델을 구축하는 과정에서 고려할 사항은 다음의 세 가지이다. 첫째, 인공지능망 모델 중 데이터의 시계열적 특성을 활용하는 알고리즘인 LSTM(Long Short term Memory)을 활용하여 예측 모델을 구축하고, 이전 연구에서 사용되었던 딥러닝 알고리즘과의 비교한 결과를 제시하고자 한다. 둘째, 모델의 입력변수를 기술적 분석(technical analysis) 방법을 통해 가공할 예정이다. 즉, 데이터 전처리 과정에서 입력변수의 특징을 기술적 분석을 활용해 추출한 후, 모델이 그 결과를 학습하게 함으로써 타당도를 높이하고자 한

다. 기술적 분석은 주로 주식 시장에서 주가를 예측할 때 활용되는 대표적인 방법으로, 시장움직임(market action)을 가격과 거래량만으로 예측하는 방법이다. 마지막으로 데이터의 시계열적 특성을 활용하기 위한 방법으로 데이터 전처리 과정에서 데이터를 분해한 값을 입력변수로 활용한다.

## 제2절 연구의 내용 및 방법

### 1. 연구 범위 및 주요 내용

본 연구는 건화물선 분야에 한정하여 선형별 운임을 예측하고자 하였다. 건화물선 시장은 부정기선 시장으로 해운시장에서 변동성이 커 운임예측이 쉽지 않은 것으로 알려진 분야이다. 부정기선 시장은 정해진 항로가 없는 시장으로, 선박의 입출항지가 정해져 있지 않다. 따라서 건화물선을 소유 또는 운항하는 기업은 시장의 상황에 따라 항로를 결정할 수 있으며, 선박의 용선과 대선을 통해 시황에서 노출을 조정함으로써 단기간에 기업의 영업 수익을 창출하기 위한 영업 전략을 수립할 수 있다. 따라서 다른 선종을 소유한 기업에 비해 운임의 예측에 대한 필요성이 크다. 반면 컨테이너선은 정기선으로 수요와 공급의 변화가 크지 않고, 운임이 수급보다는 외부 충격으로 인해 결정되는 경우가 많아 시황예측의 중요성이 건화물선에 비해 크지 않다. 유조선의 경우는 국내 해운기업 중 유조선을 소유하거나 운항관리 하여 수익을 얻는 기업이 많지 않으며, 이로 인해 노출 비중이 크지 않기 때문에 건화물선을 연구 범위로 선정하였다. 본 연구는

인공지능기반 선형별 운임예측 모형을 구축하기 위해 주식 시장에서 주가를 예측하기 위해 고안된 기술적 분석을 활용한다. 또한 운임의 시계열적 특성을 활용하기 위해 성분 분해를 실시하여, 운임의 추세, 계절, 순환, 불규칙 성분을 입력변수로 활용하였는데 이에 대한 설명은 2장에서 다루었다. 3장에서는 모형을 구축하기 위해 사용된 인공지능 모형인 LSTM에 대한 개념과 모델의 설계 및 구현에 대해 자세히 설명하였으며, 구축된 모델의 결과와 분석 내용을 4장에서 제시하고, 결론, 시사점 및 향후 연구 방향에 대해 5장에서 제시한다.

## 2. 연구 방법

인공지능기반 건화물선의 선형별 운임예측 모형을 구축하기 위해 모델의 학습과정에서 운임 데이터의 시계열적 특성을 활용하고자 하였다. 이를 위해 3가지 방법을 사용하였는데 첫째는 기술적 분석(technical analysis), 둘째는 시계열 분해, 셋째는 통계분석이다. 기술적 분석을 통해 건화물선 운임 자체의 움직임을 추출하고자 하였으며, 시계열 분해를 통해 시간의 경과에 따른 데이터의 특성인 추세, 계절, 순환성, 불규칙성을 성분으로 추출하였다. 마지막으로 통계 분석을 활용하여 데이터의 기초 특성인 표준편차, 백분위 등을 입력변수로 활용할 수 있도록 하였다. 모형을 구축을 위해 신경망 모델 중 하나로 순서가 있는 데이터를 기반으로 하여 분석을 진행했다. LSTM 알고리즘은 해상운임 데이터와 같이 시간에 따라 값의 변동이 있는 시계열 데이터를 활용해 미래 시점에 발생할 수 있는 값을 예측하는 데 활용된다. 비교되는 모델은 딥러닝 모델을 적용하여 성과를 비교하였다.

## 제3절 선행연구 및 연구의 차별성

### 1. 선행연구

해상운임을 예측하기 위해 많은 연구가 수행되어 왔다. 해상운임은 변동성이 크다는 특징을 가지고 있다. 일반적으로 변동성이 크다고 알려진 주식 시장보다 해상운임의 변동성이 크다. 특히 부정기선 운임인 건화물선 운임지수(BDI: Baltic Exchange Dry Index)의 변동계수(CV: Coefficient of Variation)를 살펴보면, 코스피, 코스닥지수보다 각각 4.5배, 3.3배 큰 것으로 나타난다.<sup>1)</sup> 이러한 해상운임의 특성으로 운임을 예측하는 것은 쉽지 않다.<sup>2)</sup>

해상운임예측 모델을 구축하기 위해 활용되는 방법론을 크게 3가지로 정리된다.<sup>3)</sup> 첫째는 운임에 영향을 주는 것으로 알려진 수급요인을 선택하여 회귀 모델을 구축하는 방법, 둘째는 운임 데이터의 시계열적 특성을 활용하여 자기회귀변수를 독립변수로 계량분석 모델을 구축하는 것이며,<sup>4)</sup> 마지막으로 기계학습 모델을 구축하여 예측하는 방법이다. 대부분의 해상운임예측 모델은 회귀 또는 계량분석 방법론에 의존하고 있으며, 근래에 들어서야 인공지능에 대한 관심이 다시 높아지면서 연구가 점차 활기를 띠고 있다.

회귀분석, 시계열 분해 등과 같은 기존의 방법론은 예측 결과에 대해 어느 정도 설명이 가능하다는 장점을 가지고 있는 반면, 데이터의 특성(정규성, 다중공선성 등)을 가정해야 하거나, 정형적 데이터만으

1) 2015년 2월부터 2019년 12월까지 월단위의 데이터를 활용하여 변동계수를 구한 후, 비교하였다.

2) Lyridis *et al.*(2004); Eslami *et al.*(2017)

3) Eslami *et al.*, *op. cit.*, pp. 539-540.

4) Song and Chissom(1993); Chou(2008)



로 모델을 구축해야 한다는 한계가 있다. 기초통계량 분석방법 중 하나인 다변량 회귀분석의 경우에는 많은 양의 데이터를 한 번에 모델링하여 분석할 경우 다중공선성 문제가 발생할 확률이 높아 모델을 구축하기 어려워지게 된다.

또한 해운시황의 데이터는 대부분 불안정(non-stationary) 시계열이므로 원형의 데이터를 활용해 회귀분석을 실시할 경우 가성회귀(spurious regression)가 되어 원하는 결과를 얻을 수 없다. 따라서 회귀분석을 실시하기 전 차분하거나 로그변환하여 데이터의 안정성(stationarity)을 확보한 후에야 비로소 모델링 작업을 수행할 수 있다. 이러한 과정에서 원형 데이터가 가지고 있는 중요한 정보가 손실될 가능성이 커, 모델이 운임예측을 위해 계산 작업을 수행할 때 정보가 손실된 데이터로 인해 목표하는 종속변수의 값이 왜곡될 수 있다는 한계가 존재한다.

이러한 상황에서 컴퓨터 연산능력의 향상으로 기계학습기반의 컴퓨팅 알고리즘을 실행하는 것이 가능해짐으로써 데이터를 활용한 예측의 영역이 확장되었다. 대용량, 실시간의 데이터를 처리, 분석하는 기술이 빠른 속도로 진화하고 있으며 기존의 분석방법을 보완하는 새로운 분석방법으로 떠오르고 있다. 특히 자원과 재무 등의 분야에서 예측에 활용되어 성과가 나타나기 시작한 것으로 알려져 있으나, 해운분야에서의 활용은 비교적 최근 일이다. 인공지능에 의한 성과가 여러 분야에서 입증되기 시작하면서 해운업계에서도 이를 활용한 연구에 대한 요구가 점차 강해지고 있다.

Veenstra and Franses(1997)은 건화물선 해상운임 데이터가 시계열적 특성을 가지고 있다는 점에 착안하여 벡터자기회귀 모델(VAR: Vector Autoregressive model, 이하 “VAR”이라고 한다)을 활용하여

해상운임을 예측하였다.<sup>5)</sup> 건화물선 중 특정 선형인 케이프선과 파나막스선의 주요 항로별 운임을 VAR 모델로 예측하고자 시도하였으며, 그 결과 VAR 모델은 장기 운임예측보다는 단기 운임예측에 적합하다는 결론을 제시하였다.

Kavussanos and Alizadeh-M(2001)은 건화물선 선형별(케이프선, 파나막스선, 핸디선) 운임을 예측하고자 시도하였다.<sup>6)</sup> ARIMA (Auto-regressive Intergrated Moving Average) 모델과 VAR 모델이 운임의 계절을 분석하고 예측에 활용하기 적합한 모델이라는 결론을 제시하였다. Kavussanos and Alizadeh-M(2002)은 다른 연구에서 스팟운임과 용선료(1년, 3년) 간의 장기균형관계를 벡터오차수정 (Structural Vector Error Correction, 이하 “VECM”이라고 한다) 모델을 통해 분석하고자 하였다.<sup>7)</sup> 분석결과 변수 간 장기 균형 관계가 존재하지 않는다는 결론을 제시하였는데, 장기 균형관계가 존재하지 않는 이유에 대해 해운사가 시황을 고려하지 않고 선박 확보를 위한 무리한 투자를 하는 경우가 많기 때문이라고 해석했다.

Kavussanos and Visvikis(2004)는 운임선도거래(FFA: Forward Freight Agreement) 가격이 시장의 심리를 반영한다는 점에 착안하여 운임을 예측하려는 시도를 하였다.<sup>8)</sup> 파나막스선의 스팟운임과 FFA 가격 간의 관계를 분석하고, 두 변수 간에 공적분 관계가 있어 VECM 모델을 활용해 장기적 균형관계에 대한 정보와 단기적 움직임 을 파악하고자 시도하였다. 그 결과 FFA 가격이 운임을 예측하는 데 중요한 변수임을 입증하였으며, 두 변수를 활용한 VECM 모델의 유용성을 입증하였다. Alizadeh (2013)은 건화물선의 선형별(케이프,

---

5) Veenstra *et al.*(1997)

6) Kavussanos and Alizadeh-M (2001)

7) Kavussanos and Alizadeh-M (2002)

8) Kavussanos and Visvikis (2004)

파나막스, 수프라막스선) 운임을 예측하기 위해 항로별 FFA 가격을 변수로 한 VAR 모델을 구축하였다.<sup>9)</sup> 중대형선인 케이프와 파나막스선 FFA 가격이 거래량에 정의 영향을 미치는 반면, 소형선인 수프라막스선은 거래량이 FFA 가격에 정의 영향을 미쳤다는 결과를 제시하였다. FFA 가격과 거래량의 변동이 운임예측 모델을 구축하는 데 활용하기 적합한 변수임을 입증하였다.

모수원(2010)은 BDI의 움직임이 시차를 두고 서로 상호의존하고 있다는 점에 착안해 단일변량 모형을 구축하고자 하였다. 즉, BDI의 시계열적 특성을 기반으로 단변량 ARIMA 모형과 장기변동 추출에 적합한 HP(Hodrick-Prescott)필터 기법 기반 예측 모델을 구축하였으며, 모델의 결과를 활용해 BDI예측치의 상한선과 하한선을 결정하는데 활용하였다.<sup>10)</sup>

Lyridis(2004)는 인공신경망 모델을 활용해 대형원유선의 운임을 예측하는 연구를 수행하였다.<sup>11)</sup> 입력변수로는 수요측면에서 원유생산량, 공급측면에서는 선박량을, 기타 변수로 원유가 및 선가를 활용하여 운임을 예측하고자 시도하였다. 단기예측이 아닌 최소 1개월에서 1년까지 중장기적 관점에서 예측을 시도하였으며, 그 결과 운임의 예측 기간을 늘릴수록 성능이 떨어진다는 결론을 제시하였다.

이우식(2017)은 변동성이 높은 주가를 예측하기 위해 인공지능 모델을 구축하였는데, 이때 입력변수로 주가를 활용해 상대강도지수(RSI: Relative Strength Index), 모멘텀(Momentum), 변화율(ROC: Rate of Change) 등과 같은 기술적 지표를 활용하였다.<sup>12)</sup> 추출한 입력변수를 인공지능 모형인 서포트벡터머신(SVM: Support Vector

9) Alizadeh (2013)

10) 모수원(2010)

11) Lyridis *et al.*, *op. cit.*(2004)

12) 이우식(2017)

Machine), 인공지능망 알고리즘 등에 적용하였으며, 주가 예측에 앞서 언급한 방법이 적합함을 입증하였다.

Shynkevich *et al.*(2017)도 주가 예측을 위해 기술적 분석 지표인 이동평균(MA: Moving Average), 변화율, 상대강도지수, 윌리엄스%R 등을 입력변수로 활용하여 인공지능 모델을 구축하였다.<sup>13)</sup> 구축한 인공지능 모델은 인공지능망, 서포트벡터머신 등이며, 기술적 분석 지표가 주가 운임예측에 적절한 변수임을 보였다. 임상섭·윤희성은 건화물선 운임 중 수프라막스선에 한정하여 운임예측 모형을 구축하였다.<sup>14)</sup> 기술적 지표를 입력변수로 인공지능망 모형을 구축하였으며, 시황을 예측하는 데 기술적 지표가 유용하다는 결과를 제시하였다.

〈표 1-1〉 선행연구 요약

구분	선행연구와의 차별성		
	연구 목적	연구 방법	주요 연구내용
주요 선행 연구	1 <ul style="list-style-type: none"> <li>• 과제명: Forecasting Tanker Market Using Artificial Neural Networks</li> <li>• 연구자(연도): Lyridis <i>et al.</i> (2004)</li> <li>• 연구 목적: 인공지능망을 활용한 탱커 스팟운임 예측</li> </ul>	<ul style="list-style-type: none"> <li>• 인공지능망 모델 활용</li> <li>• 변수는 운임에 영향을 주는 수급 요인을 활용함</li> </ul>	<ul style="list-style-type: none"> <li>• 인공지능망기반 탱커 운임예측 모델을 구축함</li> <li>• 예측 기간을 중장기로 확장함 (1개월에서 1년)</li> </ul>
	2 <ul style="list-style-type: none"> <li>• 과제명: 2010년 BDI의 예측 - ARIMA 모델과 HP 기법을 이용하여</li> <li>• 연구자(연도): 모수원 (2010)</li> <li>• 연구 목적: 시계열 모델을 활용하여 BDI를 예측</li> </ul>	<ul style="list-style-type: none"> <li>• 시계열 분해 방법인 ARIMA 모형을 구축함</li> <li>• HP Filter(HP) 모델을 구축함</li> <li>• 입출력변수로 BDI를 활용함</li> </ul>	<ul style="list-style-type: none"> <li>• ARIMA와 HP기반 BDI 예측 모델을 구축함</li> </ul>
	3 <ul style="list-style-type: none"> <li>• 과제명: 딥러닝분석과 기술적 분석 지표를 이용한 한국 코</li> </ul>	<ul style="list-style-type: none"> <li>• 인공지능 알고리즘 중 의사결정나무, 서포트벡</li> </ul>	<ul style="list-style-type: none"> <li>• 다양한 인공지능 알고리즘을 활용하여 주가</li> </ul>

13) Shynkevich *et al.*(2017)

14) 임상섭·윤희성(2018)

구 분	선행연구와의 차별성		
	연구 목적	연구 방법	주요 연구내용
4	<ul style="list-style-type: none"> <li>스피주가 지수 방향성 예측</li> <li>연구자(연도): 이우식(2017)</li> <li>연구 목적: 기술적 지표와 인공지능 모형을 활용해 주가를 예측</li> </ul>	<ul style="list-style-type: none"> <li>터머신, 인공신경망을 활용함</li> <li>기술적 분석 방법을 이용해 입력변수 추출</li> </ul>	<ul style="list-style-type: none"> <li>예측 모형을 구축</li> <li>주가의 특성을 기술적 분석을 통해 추출해 입력변수로 활용함</li> </ul>
	<ul style="list-style-type: none"> <li>과제명: Forecasting price movements using technical indicators: Investigating the impact of varying input window length</li> <li>연구자(연도): Shynkevich <i>et al.</i> (2017)</li> <li>연구 목적: 기술적 지표와 인공지능 모형을 활용한 주가 예측</li> </ul>	<ul style="list-style-type: none"> <li>인공지능 알고리즘 중 서포트벡터머신, k-최근접 이웃을 활용함</li> <li>기술적 분석 방법을 이용해 입력변수를 추출</li> </ul>	<ul style="list-style-type: none"> <li>다양한 인공지능기반 주가 예측 모형을 구축함</li> <li>주가의 특성을 기술적 분석을 통해 추출하고 입력변수로 활용함</li> </ul>
	<ul style="list-style-type: none"> <li>과제명: Supramax Bulk Carrier Market Forecasting with Technical Indicators and Neural Networks</li> <li>연구자(연도): 임상섭·윤희성(2018)</li> <li>연구 목적: 기술적 지표와 인공지능 모형을 활용해 건화물선 운임예측</li> </ul>	<ul style="list-style-type: none"> <li>인공신경망 알고리즘 활용</li> <li>기술적 분석 방법을 이용해 입력변수를 추출</li> </ul>	<ul style="list-style-type: none"> <li>운임의 특성을 기술적 분석방법을 이용해 추출하고 이를 입력변수로 활용함</li> <li>인공신경망기반 모델을 구축함</li> </ul>
본 연구	<ul style="list-style-type: none"> <li>변수의 시계열적 특성을 활용하여 건화물선 시장을 예측하는 모형을 구축하고자 함</li> </ul>	<ul style="list-style-type: none"> <li>운임을 기술적 분석 방법을 통해 시간에 따른 운임의 변화를 관찰할 수 있는 입력변수를 추출함</li> <li>추가적으로 운임의 시계열적 특성을 활용하기 위해 추세, 계절, 불규칙성분을 추출해 입력변수로 활용함</li> <li>순환신경망기반 건화물선 운임예측 모델을 개발함</li> <li>다른 모델들과 예측성능 비교</li> </ul>	<ul style="list-style-type: none"> <li>기술적 지표, 시계열 분해 방법을 활용한 변수탐색</li> <li>순환신경망 모델을 활용한 모델 학습 및 분석</li> <li>기존의 인공신경망 모델과의 예측성능 비교</li> </ul>

## 2. 연구의 차별성

본 보고서의 차별성은 두 가지로 요약할 수 있다. 첫째는 데이터 전처리 과정에서 기술적 분석방법과 시계열 분해 방법을 활용했다는 점이다. 즉, 데이터 전처리 과정에서 입력변수의 특징을 기술적 분석을 활용해 추출한 후, 모형이 그 결과를 학습하게 함으로써 타당도를 높였다. 기술적 분석 방법으로 주식 추세의 방향성, 변곡점 등을 확인할 수 있다. 또한 시계열 요소 분해법을 활용하여 데이터의 추세, 순환, 계절, 불규칙성을 분해하여 데이터로 활용하였다. 가법모형(additive model)을 활용해 원시계열 자료를 분해했으며, 이를 변수로 활용하여 모형의 학습효과를 높였다. 둘째는 시계열 분해에 특화되어 있는 LSTM 신경망을 활용하여 해상운임예측 모델을 구축했다는 점이다. LSTM은 시계열 데이터를 활용하는 데 적합한 인공신경망 모델이다.

## 제 2 장

## 데이터 및 변수 선정

## 제1절 데이터

본 연구는 건화물선의 선형별 해상운임을 예측하는 모델을 제시하기 위해 수행되었으며, <표 2-1>에서 보는 것처럼, 모형을 구축하기 위해 선형별 운임과 용선료, FFA를 데이터로 사용하였다. 모형을 구축하는 데 사용한 데이터의 기간은 2014년 2월 24일부터 2020년 8월 28일까지이며, 스팟운임과 선도거래 데이터는 1,630개(일단위), 용선료는 340개(주단위)를 활용하였다.

〈표 2-1〉 데이터 요약

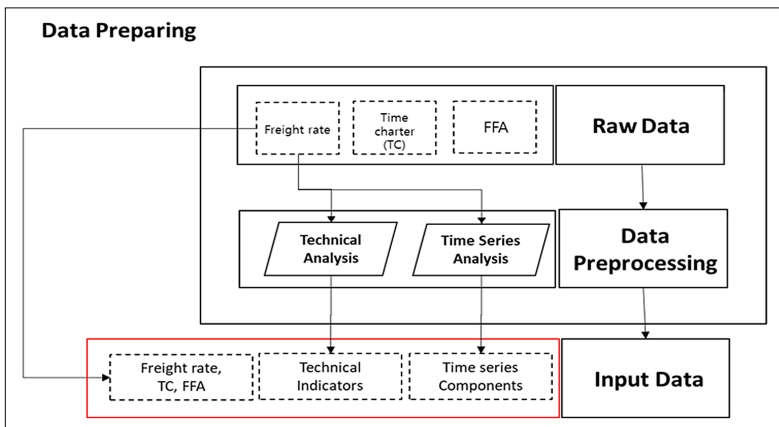
변수	설명	출처	단위
선형별 운임	케이프선 5TC 파나막스선 4TC 수프라막스선 6TC	클락슨 (Clarkson)	\$/day
용선료	단기용선료(6개월) 케이프선 18만 dwt 파나막스선 7.5K dwt 장기용선료(5년) 수프라막스선 5.8K dwt	클락슨	\$/day
FFA	케이프선 당월물, 익월물 파나막스선 당월물, 익월물 수프라막스선 당월물, 익월물	볼틱해운거래소 (Baltic Exchange)	\$/day

자료: 저자 작성

## 제2절 데이터의 전처리

본 연구에서는 모델을 구축하기 위해 운임을 기술적 분석과 시계열 분해라는 2가지 방법을 이용하여 변환하였으며, 다음의 그림과 같이 변수를 분석하여 모델 학습과정에서 입력변수로 활용하였다.

〈그림 2-1〉 데이터 전처리 과정



자료: 저자 작성



## 1. 기술적 분석

기술적 분석은 주식 시장을 분석할 때 주로 활용되며, 주가의 움직임에 주목하는 방법이다. 즉 분석 대상의 가치가 과대평가되어 있다면 매도될 것이고, 반대로 과소평가 되어 있다면 매수된다는 점에 착안하여 발전한 방법이다. 기술적 분석의 가정은 수요와 공급에 의해 주가가 결정되며, 그 결과를 그래프로 나타낸다는 점이다.<sup>15)</sup> 기술적 지표는 주가의 방향을 확인할 수 있는 추세 지표, 주가의 추세가 변하는 변곡점을 확인할 수 있는 모멘텀 지표가 있다. 동 연구에서는 두 가지 지표 이외에도 백분율, 회귀계수(Regression Beta) 등을 시장의 움직임을 관찰할 수 있는 지표로 고려하였다. 모델 구축을 위해 활용한 기술적 분석 지표에 대해 <표 2-2>와 같이 표로 정리하여 제시하였다.

<표 2-2> 기술적 지표

구분	사용목적	지표
추세 지표	주가의 방향 확인	이동평균, MACD
모멘텀 지표	주가의 추세가 변하는 변곡점 확인	K모멘텀, 변화율, 상대강도지수, 윌리엄스 %R, 스토캐스틱
기타	주가의 방향, 추세 등의 움직임 확인	백분위, 회귀계수, 로그수익률, 표준편차, 장단기 용선료 비율

자료: 저자 작성

15) 이우식, 앞의 책, p. 287.

## 1) 추세 지표

추세 지표는 일반적으로 시간의 흐름에 따른 주가의 방향성(상승, 하락, 횡보)을 관찰하기 위해 고안된 지표로, 이동평균, MACD(Moving Average Convergence and Divergence)등이 추세 지표에 해당한다.

〈표 2-3〉 추세 지표

구분	지표	설명
추세 지표	이동평균	<ul style="list-style-type: none"> <li>• 주가 등의 변화를 살펴보기 위해 고안된 지표임</li> <li>• 일정 기간 동안 평균을 구한 뒤, 그 평균값의 흐름을 파악함</li> <li>• 이동평균선은 정해진 기간의 길이에 따라 단기와 장기로 구분함</li> <li>• 장기와 단기이동평균선 간의 관계에서 의미를 읽어냄. 골든크로스와 데드크로스가 대표적인 예임</li> </ul>
	MACD	<ul style="list-style-type: none"> <li>• 이동평균선 간의 관계를 살펴보기 위해 고안된 지표임</li> <li>• 기존 이동평균선의 시차(time lag)를 보완하기 위해 개발된 지표임</li> <li>• 이동평균선 간의 차가 확대되면 MACD가 상승, 반대로 차가 축소되면 MACD가 하락하도록 고안된 지표임</li> </ul>

자료: 저자 작성

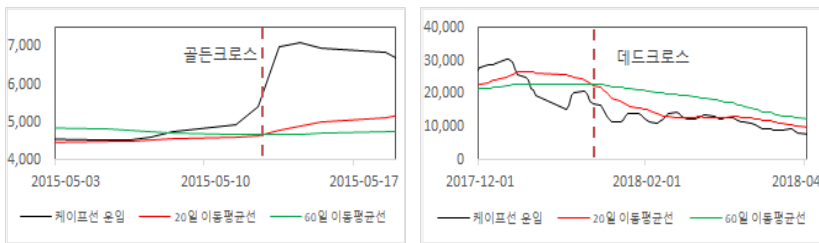
### (1) 이동평균

이동평균선은 시간에 따른 주가의 변화를 살펴보기 위해 고안되었으며, 일정 기간 동안의 주가의 평균을 구한 뒤, 그 평균값의 흐름을 통해 주가의 방향성을 파악할 수 있다. 이동평균선의 목적에 맞게 주가의 방향성을 관찰하기 원하는 시작 시점과 종료 시점을 설정할 수 있으며, 이동평균선의 관찰 기간은 증권거래소의 영업일을 기준(금·토요일 제외)으로 정하는 것이 일반적이다.

이동평균선은 설정한 기간에 따라 단기, 중기, 장기로 구분되며, 설정 기간에 따라 구분된 이동평균선 간의 조합을 통하여, 주가의 변

곡점을 판단할 수 있다. 이를 크로스 분석이라고 하는데, 대표적인 방법으로 골든크로스(Golden Cross)와 데드크로스(Dead Cross)가 있다. 골든크로스는 단기이동평균선이 장기이동평균선을 상향돌파하는 시점을 말하며, 주가의 상승세가 강할 때와 하락세를 보이던 주가가 반등했을 때 나타난다. 반대로 단기이동평균선이 장기이동평균선을 하향돌파하는 시점을 데드크로스라고 말하며, 주가의 하락세가 강할 때와 상승세를 보이던 주가의 흐름이 하락세로 전환될 때 나타난다.

〈그림 2-2〉 케이프선 운임과 장단기이동평균선



자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

## (2) MACD<sup>16)</sup>

MACD는 이동평균선이 주가의 추세를 뒤늦게 반영한다는 단점을 보완하기 위해 사용하는 기술적 지표 중 하나로, 기간이 다른 장기와 단기이동평균선 사이의 관계에서 추세 변화의 신호를 찾을 수 있도록 고안되었다. 동 지표는 MACD선과 MACD의 이동평균선으로 구성되며, 이를 식으로 나타내면 다음과 같다. 여기서 MACD선의 이동평균을 시그널(signal)이라 한다. 케이프선 스팟운임의 특정 기간 동안의 MACD와 시그널을 살펴보면 〈그림 2-3〉과 같다.

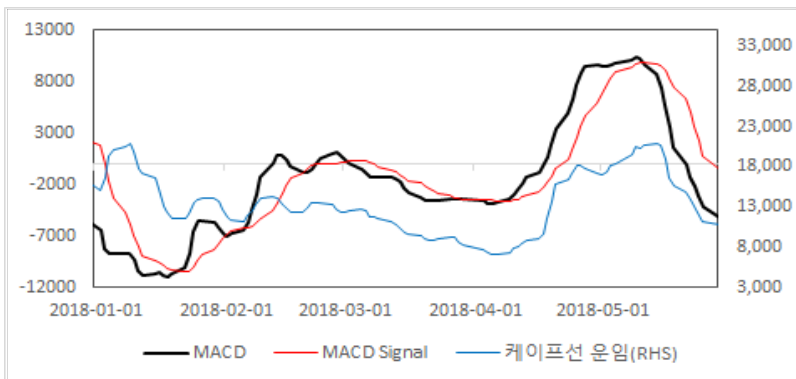
16) 윤희성 외(2019), pp.12-14, 재인용: John J.Murphy(1999), 최용석 번역(2000), pp.250-256.

$$MACD = MA_{short} - MA_{long}$$

$Signal = MACD$ 의 이동평균

$MA_{short}$  = 단기이동평균,  $MA_{long}$  = 장기이동평균

〈그림 2-3〉 케이프선 운임과 MACD



자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

## 2) 모멘텀 지표

모멘텀 지표는 시간에 따라 움직이는 주가의 변곡점을 판단하기 위해 고안된 지표로, 모멘텀, 변화율 등이 모멘텀 지표에 해당된다.

### (1) 모멘텀

모멘텀(Momentum)은 시간의 경과에 따른 주가의 를 정량화하기 위해 고안된 지표이며, 관찰하기 원하는 기간을 설정한 후, 그 기간의 시작 시점과 종료 시점 간의 주가 차이로 표현된다. 모멘텀 지표는 0을 기준으로 해석할 수 있으며, 모멘텀 지표가 0보다 크면 주가가 상승 중이며, 반대로 0보다 작으면 주가가 하락하고 있음을 의미

〈표 2-4〉 모멘텀 지표

구분	지표	설명
모멘텀 지표	모멘텀	<ul style="list-style-type: none"> <li>• 일정 기간 동안의 상승세와 하락세를 관찰하기 위해 고안된 지표임</li> <li>• 주가 간의 차이를 나타낸 지표임</li> </ul>
	변화율	<ul style="list-style-type: none"> <li>• 주가 간의 차이를 비율로 나타낸 지표</li> </ul>
	상대강도지수	<ul style="list-style-type: none"> <li>• 일정 기간 동안의 주가의 상승폭과 하락폭을 활용하여 범위를 설정하고 두 값을 활용함으로써 추세에 가지는 강도를 수치로 표현 가능하도록 고안된 지표임</li> </ul>
	윌리엄스 %R	<ul style="list-style-type: none"> <li>• 일정 기간 동안의 주가의 고점과 저점의 차이를 범위로 설정하여 현시점의 주가 수준을 평가하도록 고안된 지표임</li> <li>• 0에서 100 사이의 값으로 표현됨</li> </ul>
	스토캐스틱	<ul style="list-style-type: none"> <li>• 스토캐스틱은 일정 기간 동안의 최저가와 최고가 사이에서 현재시점의 주가와 최저가 간의 차이의 크기를 비교하여 0에서 100 사이로 나타낸 지표임</li> <li>• 기간 설정을 위해 3가지 값이 필요함</li> <li>• 스토캐스틱은 Slow(%K)와 Fast(%D)로 표현됨</li> <li>• %K와 %D의 신호선을 활용하여, %K가 %D를 상향돌파하면 골든크로스, 그 반대의 경우를 데드크로스로 해석함</li> </ul>

자료: 저자 작성

한다. 이해를 돕기 위해 케이프션 운임과 모멘텀 지표의 관계를 그림으로 나타내었다(〈그림 2-4〉 참조).

## (2) 변화율

변화율도 모멘텀 지표와 마찬가지로 일정 기간 동안의 주가의 추세를 정량화하기 위해 고안된 지표 중 하나로, 모멘텀 지표가 주가의 추세를 주가 간의 차이로 정량화하였다면, 변화율은 일정 기간 동안

의 주가의 비율로 정량화한다. 변화율은 1을 기준으로 정하여 해석하며, 1보다 지표가 클 경우를 주가가 상승 중에 있다고 해석하고, 반대로 지표가 1보다 작을 경우 주가가 하락 중에 있다고 해석한다. 케이프션 운임과 변화율에 대한 관계를 표현하면 아래의 〈그림 2-5〉와 같다.

### (3) 상대강도지수<sup>17)</sup>

상대강도지수는 추세의 변곡점을 알아보기 위해 사용하는 지표 중 하나로, 일정 기간 동안의 상승세와 하락세의 비율을 백분율로 나타내며, 동지수는 주식 시장의 과열 여부를 판단할 때 주로 활용된다. 상대강도지수를 해석할 때는 기준점을 30과 70으로 설정하는 경우가 일반적이다. 상대강도지수가 30선을 하향돌파하는 시점을 주식 시장이 과매도 상태인 것으로 간주하며, 70선을 상향돌파하는 시점은 과매수 상태로 해석한다. 상대강도 지수를 수식으로 표현하면 아래와 같으며, 케이프션 운임과 상대강도지수의 관계를 〈그림 2-6〉으로 나타내었다.

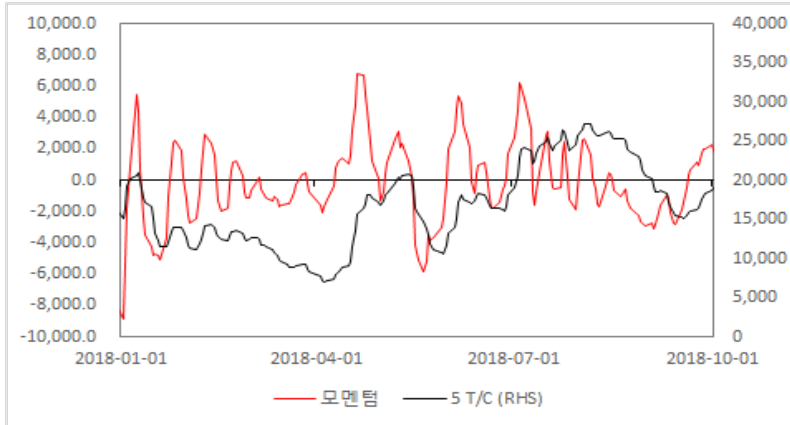
$$RSI = \frac{U}{U+D} \times 100$$

$U = \kappa$ 일간 주가의 상승폭 합계

$D = \kappa$ 일간 주가의 하락폭 합계

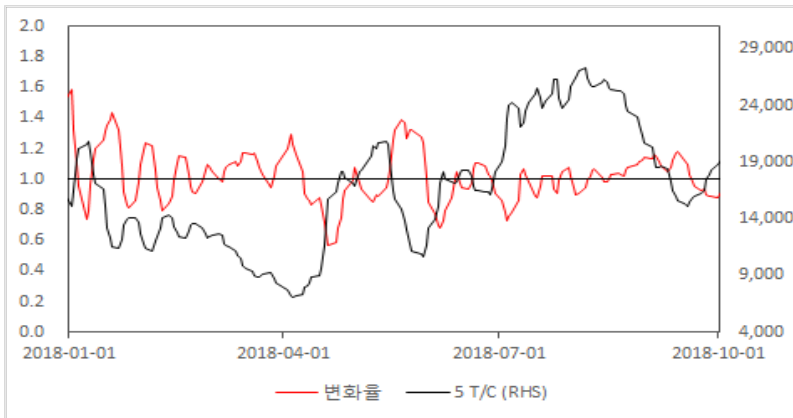
17) 윤희성 외(2019), p.16, 재인용: John J.Murphy(1999), 최용석 번역(2000), pp.260-267.

〈그림 2-4〉 케이프선 운임과 모멘텀 지표



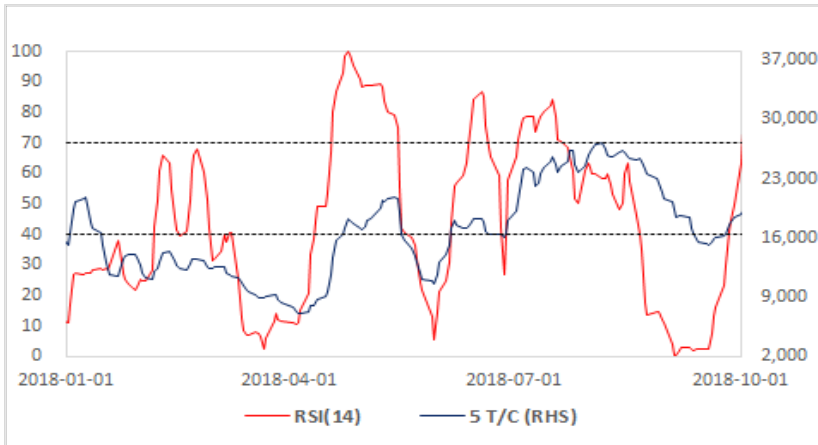
자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

〈그림 2-5〉 케이프선 운임과 변화율 지표



자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

〈그림 2-6〉 케이프선 운임과 상대강도지수



자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

#### (4) 윌리엄스 %R<sup>18)</sup>

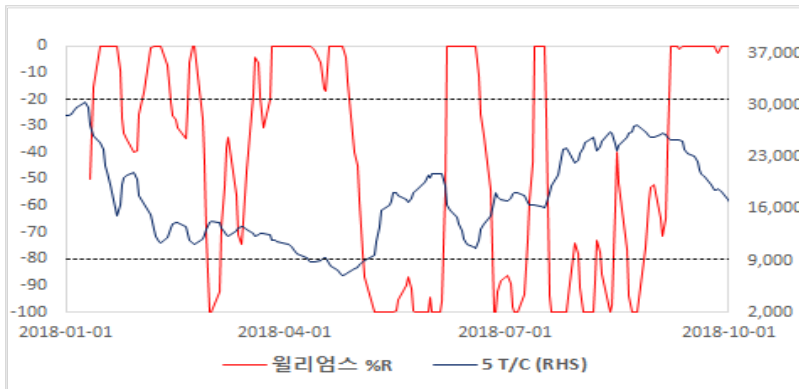
윌리엄스 %R은 일정 기간 동안의 주가의 변화 대비 고가와 현재 주가와의 차이를 이용하여 나타낸다. 동 지표의 특징 중 하나는 주가가 고점일 때를 기준으로 계산된다는 것과 주가의 변화를 -100에서 0 사이로 표현한다는 점이다. 이를 수식으로 표현하면 아래와 같다. 윌리엄스 %R은 주가가 고점에 가까워지면 0에 근접하게 되며, 저점에 가까워질수록 -100에 근접하도록 고안된 지표이다. 동 지표를 해석할 때는 -20과 -80을 기준으로 삼으며, 지표가 -20선을 하향돌파하는 형태를 보일 때를 주가의 추세가 하락으로 전환될 가능성이 있다고 보며, -80선을 상향돌파할 때를 주가의 추세가 상승으로 전환될 가능성이 높다고 해석한다. 케이프선 운임과 윌리엄스 %R의 관계를 〈그림 2-7〉로 제시하였다.

18) 윤희성 외(2019), pp.17-18, 재인용: John J.Murphy(1999), 최용석 번역(2000), pp.270-272.



$$Williams \%R = - \frac{H_{\kappa} - C_t}{H_{\kappa} - L_{\kappa}} \times 100$$

〈그림 2-7〉 케이프선 운임과 윌리엄스 %R



자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

### (5) 스토캐스틱(Stochastic)<sup>19)</sup>

스토캐스틱은 일정 기간 동안의 주가의 변화 대비 저가를 기준으로 현재 주가와 차이의 비율을 이용하여 나타내며, 주로 주가의 변곡점을 확인하기 위해 활용되는 지표이다. 스토캐스틱은 'Fast %K', 'Slow %K', 'Slow %D'로 구성되어 있으며, 각 지표는 0에서 100 사이로 표현된다. 주가의 'Fast %K'를 아래의 식을 활용하여 구한 뒤, 일정 기간 동안의 'Fast %K'의 이동평균 'Slow %K'를 계산한다. 이후 'Slow %K'값을 다시 일정 기간 동안의 이동평균으로 계산하게 되면 'Slow %D' 값을 얻게 되는데, 주가추세의 변화 시점을 'Slow %K'와 'Slow %D'의 관계로 포착한다. 스토캐스틱 지표의 해석은 이동평균선의 장단기 신호를 해석할 때와 방식이 유사하다.

19) 윤희성 외(2019), pp.18-19, 재인용: John J. Murphy(1999), 최용석 번역(2000), pp.268-270.

$$Fast \%K = \frac{C_t - L_{\kappa}}{H_{\kappa} - L_{\kappa}} \times 100$$

$C_t$  = t시점의 주가,  $H_{\kappa}$  =  $\kappa$ 일 중 최고가,  $L_{\kappa}$  =  $\kappa$ 일 중 최저가

〈그림 2-8〉 케이프선 운임과 스토캐스틱



자료: 클락슨 데이터를 활용하여 저자 작성(검색일: 2019. 8. 5)

### 3) 기타 지표

〈표 2-5〉 기타 지표

구분	지표	설명
기타	백분위(Percentile Rank)	• 일정 기간의 주가를 0에서 100으로 보고, 특정 시점의 주가가 어느 위치에 속해 있는지 나타내는 지표임
	회귀계수	• 추세를 파악하기 위해 활용함 • 회귀계수의 절댓값 크기로 추세의 강도를 파악 • 회귀계수의 부호로 추세를 파악
	로그수익률	• 금융 분야에서 주로 활용되며, 주가의 수익과 손실을 판단하기 위해 활용됨 • 부호에 따라 수익과 손실을 판단함
	표준편차	• 시장의 변동성을 파악하기 위해 활용됨 • 표준편차의 크기에 따라 변동성의 크고 작음을 판단함

자료: 저자 작성

### (1) 백분위

백분위는 일정 기간 동안의 관측치를 0에서 100으로 표현하기 때문에 특정 시점의 주가의 상대적 위치를 측정할 수 있다. 이러한 특성을 활용하여 주가의 추세가 전환되는 시점을 포착하기 위해 지표로 활용하였으며, 상대적 위치는 특정 시점 기준 1년 전까지를 기준으로 정하여 계산하였다. 백분위로 표현된 값은 주가가 강세를 보일수록 100에 가까워지며, 주가가 약세를 보일수록 0에 가까워지게 된다.

### (2) 회귀계수

회귀계수는 관측치가 일정한 값을 기준으로 회귀하려는 경향을 이용한 분석 방법인 회귀분석에서 입력변수와 출력변수 간의 관계를 정할 때 사용되는 계수(Coefficients)로 입력변수에 곱해지는 값이다. 본 연구에서는 회귀계수의 기준을 최근 20일(영업일 기준)로 정하고 산출하였다. 산출한 회귀계수의 부호는 추세의 판단근거로, 회귀계수의 절대값의 크기로 추세의 강도를 판단하였다. 회귀계수의 부호는 운임이 상승추세를 보이는 경우 양수이며, 하락 추세를 보이는 경우의 회귀계수 부호는 음수를 나타낸다.

### (3) 로그수익률(Log-Return)

로그수익률은 0을 기준으로 부호에 따라 수익과 손실을 판단할 수 있으며, 계산이 간편해 금융 분야에서 주로 사용한다. 특정 시점의 운임 대비 그 이전 시점의 운임의 비율에 자연로그를 취해 계산되며, 본 연구에서는 5일(영업일 기준)을 기준으로 로그수익률을 계산하였다.

$$\text{로그수익률} = \ln\left(\frac{y_t}{y_{t-k}}\right) = \ln(y_t) - \ln(y_{t-k}), \quad y_t = t\text{시점 시장 가격}$$

#### (4) 표준편차(Standard Deviation)

표준편차는 관측치가 평균을 중심으로 떨어진 거리의 합을 제곱근으로 표현한 것을 말한다. 본 연구에서는 20일 동안(영업일 기준)의 표준편차를 계산하여 시장의 변동성을 파악하고자 하였다. 표준편차의 크기가 작을수록 시장의 변동성이 크지 않다는 뜻이며, 크기가 클수록 시장의 변동성이 크다는 것을 의미한다.

## 2. 시계열 분해

시계열 분해란 관측치가 시간의 흐름에 따라 나열되어 있는 시계열 데이터(time series data)의 특성을 활용하여 추세를 파악하거나, 패턴을 파악하여 원인 분석 및 예측에 사용하는 분석방법이다. 시계열 데이터는 일반적으로 추세(trend), 순환(cycle), 계절(seasonality), 불규칙(irregular) 요인으로 구성된다. 추세는 시간의 경과에 따른 데이터의 방향성을 말하며, 순환과 계절은 데이터가 반복적으로 보이는 패턴을 의미하나 계절은 1년을 기준으로 주, 월, 분기단위로 반복되는 패턴을, 순환은 1년 이상의 장기간 동안 간격을 두고 보이는 데이터의 변화 패턴을 말한다. 마지막으로 불규칙이란 추세, 순환, 계절로는 설명되지 않는, 규칙성이 발견되지 않는 요인이다. 본 연구에서는 시계열 데이터인 운임의 구성요인을 분해(decomposition)하여 인공지능 모델의 학습과정에서 입력변수로 활용하였다.

### 제3절 변수탐색 및 선정

#### 1. 기술적 변수 선정

본 연구에서 인공지능기반 모델을 구축하기 위해 학습과정에서 사용한 입력변수와 관찰 기간을 정리하여 <표 2-6>에서 제시하였다.

<표 2-6> 기술적 변수 요약

변수	수식	파라미터
이동평균비율	$\frac{\frac{1}{s} \sum_{j=0}^{s-1} TC_{t-j}}{\frac{1}{l} \sum_{i=0}^{l-1} TC_{t-i}}$	$(s, l) = (5, 20)$
MACD	$\frac{1}{s} \sum_{j=0}^{s-1} TC_{t-j} - \frac{1}{l} \sum_{i=0}^{l-1} TC_{t-i}$	$(s, l, m) = (5, 20, 5)$
시그널	$\frac{1}{m} \sum_{i=0}^{m-1} MACD_{t-i}$	
오실레이터	$MACD - Signal$	
모멘텀	$TC_t - TC_{t-k}$	$k = 5$
변화율	$\frac{TC_t}{TC_{t-k}}$	
상대강도지수	$U = \sum_{i=0}^{k-1} \max(TC_{t-i} - TC_{t-i-1}, 0)$ $D = \sum_{i=0}^{k-1} \max(TC_{t-i-1} - TC_{t-i}, 0)$ $\frac{U}{U+D}$	$k = 14$

변수	수식	파라미터
Slow %K	$H_t(k) = \max(TC_{t-k+1}, TC_{t-k+2}, \dots, TC_t)$ $L_t(k) = \min(TC_{t-k+1}, TC_{t-k+2}, \dots, TC_t)$ $Fast \% K_t = \frac{TC_t - L_t(k)}{H_t(k) - L_t(k)}$	$(k, m, n) = (10, 6, 6)$
Slow %D	$Slow \% K_t = \frac{1}{m} \sum_{i=0}^{m-1} Fast \% K_{t-i}$ $Slow \% D = \frac{1}{n} \sum_{i=0}^{n-1} Slow \% K_{t-i}$	
윌리엄스 %R	$-\frac{H_t(k) - TC_t}{H_t(k) - L_t(k)}$	$k = 14$
백분위	$\frac{1}{k} \sum_{i=0}^{k-1} I(TC_t \geq TC_{t-k})$	$k = 250$
회귀계수	$\frac{S_{xy_t}(k)}{S_{xx(k)}}$ $S_{xy_t}(k) = \sum_{i=0}^{k-1} (x_i - \bar{x})(TC_{t-i} - \overline{TC_t}(k))$ $x_i = i + 1$ $\overline{TC_t}(k) = \frac{1}{k} \sum_{i=0}^{k-1} TC_{t-i}$	$k = 20$
로그수익률	$\log(T_t) - \log(T_{t-k})$	$k = 5$
표준편차	$\frac{1}{k} \sqrt{\sum_{i=0}^{k-1} (TC_{t-i} - \overline{TC_t}(k))^2}$	$k = 20$

자료: 윤희성 외(2019), pp. 27-28, 재인용: John J. Murphy(1999), 최용석 번역(2000).

## 2. 기술적 통계 분석

건화물선의 선형별 독립변수들의 기초통계량은 다음의 표와 같이 나타났다.

〈표 2-7〉 케이프 기술통계

변수	최솟값	1분위수	중앙값	평균	3분위수	최댓값
5TC	1.885	6.960	11.325	13.559	17.574	47.398
6개월 용선료	3.650	10.650	15.275	16.150	19.900	40.900
5년 용선료	8.500	13.450	15.400	15.997	18.650	27.900
FFA 당월물	1.648	7.355	11.726	13.617	16.972	43.341
FFA 익월물	2.282	8.310	12.732	13.886	17.516	40.705
FFA-스팟 비율	0.182	0.989	1.125	1.193	1.340	3.027
MACD	-14.710	-1.440	-0.139	-0.100	1.403	11.012
시그널	-13.432	-1.383	-0.103	-0.101	1.371	10.701
오실레이터	-4.826	-0.515	-0.073	-0.099	0.353	3.333
모멘텀	-17.823	-1.443	-0.173	-0.104	1.308	11.720
변화율	0.353	0.785	0.891	0.921	1.017	1.918
상대강도지수	-0.100	0.115	0.389	0.382	0.637	0.900
윌리엄스 %R	-1.100	-1.100	-0.657	-0.626	-0.150	-0.100
Slow %K	-0.100	-0.029	0.357	0.372	0.760	0.900
Slow %D	-0.100	0.014	0.361	0.373	0.721	0.900
백분위	-0.096	0.132	0.432	0.422	0.696	0.900
회귀계수	-1.730	-0.263	-0.102	-0.100	0.097	1.406
로그수익률	-0.783	-0.204	-0.104	-0.099	-0.004	0.495
표준편차	-0.005	0.815	1.439	1.975	2.684	10.221
2일 이동평균	1.900	6.963	11.319	13.561	17.534	47.323
3일 이동평균	1.910	6.979	11.315	13.562	17.497	47.208
4일 이동평균	1.912	6.989	11.332	13.563	17.503	47.061
5일 이동평균	1.910	6.986	11.360	13.563	17.474	46.941
6일 이동평균	1.914	6.968	11.325	13.563	17.477	46.881
7일 이동평균	1.924	6.992	11.329	13.563	17.491	46.845
8일 이동평균	1.933	6.963	11.298	13.563	17.489	46.806
9일 이동평균	1.946	6.962	11.289	13.563	17.446	46.743
10일 이동평균	1.963	6.977	11.353	13.563	17.424	46.639

자료: 저자 작성

〈표 2-8〉 파나막스 기술통계

변수	최솟값	1분위수	중앙값	평균	3분위수	최댓값
4TC	2.160	6.260	8.638	9.204	11.894	27.229
6개월 용선료	4.525	8.900	10.900	11.628	13.650	31.400
5년 용선료	7.450	9.750	11.650	11.506	13.125	19.275
FFA 당월물	2.415	6.315	8.830	9.312	11.740	26.839
FFA 익월물	2.495	6.861	9.275	9.700	12.109	27.786
FFA-스팟 비율	0.336	0.977	1.105	1.167	1.287	2.657
MACD	-3.655	-0.698	-0.107	-0.118	0.524	3.524
시그널	-3.525	-0.683	-0.109	-0.119	0.499	3.459
오실레이터	-1.423	-0.278	-0.091	-0.098	0.079	1.138
모멘텀	-3.887	-0.703	-0.139	-0.116	0.451	3.380
변화율	0.577	0.832	0.895	0.905	0.968	1.501
상대강도지수	-0.100	0.041	0.404	0.391	0.728	0.900
윌리엄스 %R	-1.100	-1.100	-0.643	-0.610	-0.100	-0.100
Slow %K	-0.100	-0.079	0.377	0.387	0.853	0.900
Slow %D	-0.100	-0.029	0.377	0.386	0.794	0.900
백분위	-0.096	0.156	0.484	0.445	0.736	0.900
회귀계수	-0.584	-0.179	-0.100	-0.103	-0.021	0.379
로그수익률	-0.461	-0.160	-0.104	-0.099	-0.036	0.317
표준편차	-0.019	0.293	0.544	0.659	0.863	2.800
2일 이동평균	2.160	6.264	8.632	9.206	11.876	27.221
3일 이동평균	2.165	6.273	8.652	9.209	11.878	27.132
4일 이동평균	2.174	6.265	8.640	9.211	11.865	27.081
5일 이동평균	2.182	6.259	8.647	9.212	11.846	26.966
6일 이동평균	2.193	6.267	8.656	9.214	11.856	26.870
7일 이동평균	2.202	6.277	8.657	9.215	11.820	26.720
8일 이동평균	2.213	6.297	8.667	9.216	11.832	26.536
9일 이동평균	2.221	6.292	8.675	9.217	11.811	26.365
10일 이동평균	2.231	6.306	8.666	9.217	11.813	26.157

자료: 저자 작성



〈표 2-9〉 수프라막스 기술통계

변수	최솟값	1분위수	중앙값	평균	3분위수	최댓값
6TC	2.444	7.454	9.407	9.778	11.364	21.295
6개월 용선료	4.150	9.400	11.400	11.639	13.650	26.400
5년 용선료	7.150	9.494	10.900	11.052	12.400	17.400
FFA 당월물	2.520	7.640	9.448	9.810	11.323	21.794
FFA 익월물	2.915	7.950	9.680	10.023	11.690	23.891
FFA-스팟 비율	0.714	0.948	1.067	1.146	1.260	2.341
MACD	-2.409	-0.466	-0.098	-0.131	0.286	1.655
시그널	-2.330	-0.470	-0.092	-0.133	0.271	1.617
오실레이터	-0.804	-0.188	-0.099	-0.097	-0.016	0.705
모멘텀	-2.646	-0.395	-0.090	-0.114	0.218	1.724
변화율	0.682	0.868	0.901	0.901	0.932	1.284
상대강도지수	-0.100	-0.012	0.408	0.406	0.825	0.900
Slow %K	-0.100	-0.096	0.452	0.412	0.896	0.900
Slow %D	-0.100	-0.063	0.438	0.410	0.869	0.900
윌리엄스 %R	-1.100	-1.100	-0.566	-0.593	-0.100	-0.100
백분위	-0.096	0.128	0.440	0.436	0.744	0.900
회귀계수	-0.402	-0.150	-0.099	-0.104	-0.051	0.134
로그수익률	-0.335	-0.132	-0.098	-0.101	-0.070	0.197
표준편차	-0.062	0.102	0.269	0.353	0.511	1.690
2일 이동평균	2.449	7.446	9.411	9.780	11.363	21.279
3일 이동평균	2.459	7.454	9.413	9.782	11.362	21.268
4일 이동평균	2.465	7.444	9.415	9.783	11.371	21.260
5일 이동평균	2.475	7.451	9.418	9.784	11.366	21.246
6일 이동평균	2.486	7.441	9.415	9.786	11.381	21.225
7일 이동평균	2.496	7.446	9.420	9.787	11.369	21.200
8일 이동평균	2.511	7.450	9.419	9.789	11.379	21.167
9일 이동평균	2.525	7.465	9.424	9.791	11.378	21.139
10일 이동평균	2.545	7.470	9.422	9.793	11.373	21.210

자료: 저자 작성



## 제 3 장

## 순환신경망 모델

## 제1절 순환신경망 모델의 고찰

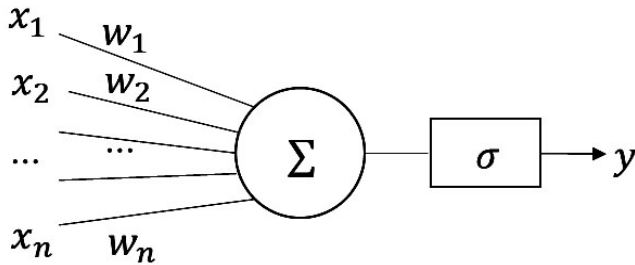
순환신경망(Recurrent neural networks) 모델은 인공신경망 모델의 한 종류로, 신경망 모형에 시간의 개념을 도입한 모델이다. 기존의 인공신경망 모델(NN, Neural network model)이 입력 데이터를 독립적으로 학습하는 구조로 되어 있어 시간에 따른 정보를 처리하지 못한다는 한계를 보완한 모델이 순환신경망 모델이다.

일반적인 인공신경망 모델은 다음의 식과 같이 나타내며, 입력값에 가중치를 곱한 후, 편향값을 더하여 나온 값을 활성화 함수를 활용해 출력값을 계산한다. 이와 같이 인공신경망 모델은 모든 입력이 서로 독립적이라고 가정하며, 시간의 정보를 고려하지 않는 모형이다.

$$y = XW + b$$

여기서,  $y$ 는 출력값,  $X$ 는 입력값,  $w$ 는 가중치,  $b$ 는 편향을 뜻한다. 각각은 모두 행렬이다.

〈그림 3-1〉 인공신경망 모델



자료: 저자 작성

## 1. RNN 개념 및 구조

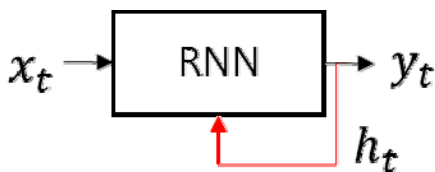
RNN(Recurrent Neural Network)은 인공신경망의 한 종류이며, 순차적으로 등장하는 데이터인 시계열 데이터를 처리하는 데에 적합한 모델이다. 순환신경망이 기존 신경망과 다른 점은 순환하는 형태라는 점이다. 2000년대 후반에 접어들면서 구글, 애플 등에서 시계열적 특성을 활용한 순환신경망을 자연어 처리 문제에 적용해 놀라운 성과를 보여주면서 알려지기 시작한 방법론으로, 주식 거래에도 많이 적용되는 방법론 중 하나이다.<sup>20)</sup>

RNN의 알고리즘은 순환(loop)하는 구조로 되어있다. 〈그림 3-2〉의 붉은색으로 표시된 것처럼, 기존의 인공신경망이 데이터 학습 과정에서 데이터가 순환할 수 있도록 순환경로를 만들어 주면서, RNN 모델은 학습하는 과정에서 과거의 정보를 기억하는 동시에 최신의

20) Ng(2020), pp.1-7의 내용을 요약·정리하였음

데이터로 정보를 갱신할 수 있게 되었다. 즉, 은닉층(hidden layer)이 일종의 메모리 역할을 하게 되는 것이다. 이를 식으로 나타내면 다음과 같다.

〈그림 3-3〉 순환신경망 모델



자료: 저자 작성

$$y_t = \tanh(h_{t-1}W_h + X_tW_i + b)$$

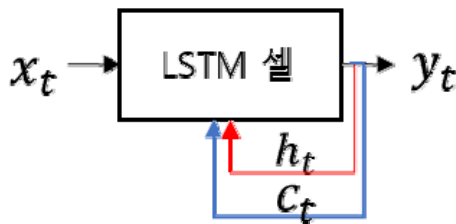
이때  $t$ 는 시각을 나타낸다.  $x_t$ 는 시계열 데이터가 RNN 모델에 입력된 것을 표현한 것이다. 방향을 가진 엣지(Edge)를 통해 이전 정보가 포함된  $h_{t-1}$ 가  $t$ 시점에 영향을 주고,  $h_t$ 를 다시 입력할 수 있는 것이다. 즉, 시각 정보  $t$ 와 가중치  $w$ 를 추가하여 이전 데이터가 다음에 영향력을 미치거나 받을 수 있는 재귀구조를 가진다.

## 2. LSTM 개념 및 구조

LSTM(Long short term memory)는 체인구조를 가지고 있으며, 각 모듈은 다른 구조를 가진 순환신경망이다. RNN은 설계 구조상 이전 정보에 대한 기억만을 활용해,  $h_{t-1}$  이전의 정보가 현재에 영향을 미치는 중요한 정보를 포함하였다 하더라도 활용할 수 없는 문제가 발생한다. 이를 장기 의존성(long-term dependency)라고 하며, 이러한

구조적 결함을 보완하기 위해 LSTM은 셀 스테이트(Cell State)라는 개념을 도입해 직전 정보뿐만 아니라  $h_{t-1}$  이전의 장기간 정보를 활용하여 예측에 활용할 수 있도록 설계해 RNN의 장기 의존성 문제를 해결하였다.

〈그림 3-4〉 LSTM 모델



자료: 저자 작성

LSTM은 셀이라는 구조를 가지고 있으며, 셀은 망각게이트(forget gate), 입력게이트(input gate)와 출력게이트(output)로 구성되는데, 게이트를 통해 이전 단계의 정보를 저장하여 cell의 형태로 흘려보낸다는 특징을 가지고 있다. 즉, 이 과정을 통해 불필요한 정보에 부여하는 가중치의 크기를 조정하는 것이다. RNN이 이전 시점  $t-1$ 의 은닉층값과 현재  $t$ 의 입력층값에 가중치를 곱한 결과를 합하여 활성화 함수 중 하나인 하이퍼볼릭 탄젠트(hyperbolic tangent)함수에 적용하여 출력값을 계산하는 과정을 반복했다면, LSTM은 셀이라는 개념을 도입해 단계별로 연산과정을 수행함으로써 RNN의 장기 의존성 문제를 해결하였다.

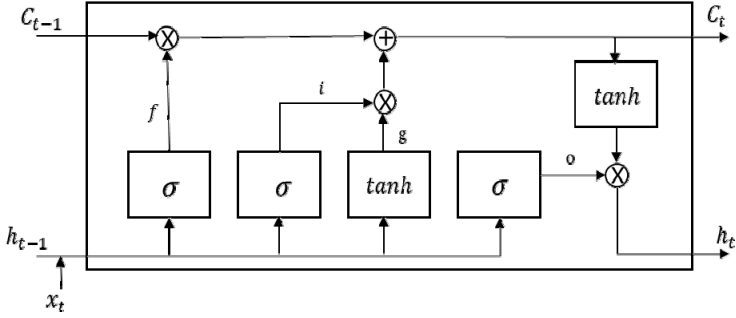
망각게이트는 과거 정보의 중요도를 결정하는 단계로 연산과정은 다음과 같다. 현시점의 정보와 과거의 은닉층 값의 값에 각각 가중치를 곱하여 더한 후, 활성화 함수 중 시그모이드(sigmoid) 함수를 활용

하여 이전 시점의 셀  $C_{t-1}$ 에 곱해준다. 이때, 시그모이드 함수를 활용하였기 때문에 그 값이 0에서 1 사이를 갖게 되는데, 1에 가까울수록 과거의 정보를 많이 활용한다는 의미이며, 0에 가까울수록 과거의 정보를 거의 활용하지 않겠다는 의미가 된다.

입력게이트는 현시점  $t$ 의 정보를 기억하기 위한 연산과정을 수행하는 단계이다. 입력게이트에서는 두 개의 값을 출력하는데, 기억셀  $C_{t-1}$ 에 새로운 정보를 추가하는  $g$ 와  $g$ 의 각 원소가 새롭게 입력된 값의 가치를 판단하여 크기로 반환하는 역할을 하는 게이트의 출력값  $i$ 를 도출한다. 입력게이트에서 출력값  $i$ 는 현시점  $t$ 에서 입력한 정보  $x_t$ 가 얼마나 중요한지를 정해주는 단계이다. 현재 시점의  $x_t$ 와 입력게이트로 이어지는 가중치  $w_x^i$ 를 곱한 값과 이전 시점  $t-1$ 의 은닉상태값  $h_{t-1}$ 에 입력게이트로 이어지는 가중치  $w_h^i$ 를 곱한 값을 더하여 시그모이드 함수를 통해 출력된 값을 말하며, 0에서 1 사이의 값을 도출한다.

현재 시점  $t$ 의  $x_t$ 값과 입력게이트와 이어지는 가중치  $w_x^i$ 를 곱한 후 이전 시점  $t-1$ 의 은닉상태값  $h_{t-1}$ 에, 역시 입력게이트로 이어지는 가중치  $w_h^i$ 를 곱한 값을 더하여 하이퍼볼릭 탄젠트 함수를 통해 출력된 값이  $g$ 이며, -1에서 1 사이의 값을 갖게 된다. 즉, LSTM은 우선 시그모이드 함수를 이용해서 정보의 중요도를 결정하고, 시그모이드 함수와 하이퍼볼릭 탄젠트 함수를 이용해 셀 스테이트에 저장될지 여부를 결정한다. 셀 스테이트는 새로운 정보를 업데이트한 후, 하이퍼볼릭 탄젠트 함수를 통해 최종적인 값을 출력한다.

〈그림 3-5〉 LSTM 모델의 셀 구조



자료: 저자 작성

$$f = \sigma(X_t W_x^f + h_{t-1} W_h^f + b^f)$$

$$g = \tanh(X_t W_x^g + h_{t-1} W_h^g + b^g)$$

$$i = \sigma(X_t W_x^i + h_{t-1} W_h^i + b^i)$$

$$o = \sigma(X_t W_x^o + h_{t-1} W_h^o + b^o)$$

$$c_t = f \odot c_{t-1} + g \odot i$$

$$h_t = o \odot \tanh(c_t)$$

이때  $f$ 는 망각게이트를  $g$ 와  $i$ 는 입력게이트의 출력값을,  $o$ 는 출력게이트의 출력값을 말하며,  $C_t$ 는 저장 셀 스테이트를,  $h_t$ 는 은닉상태값을 말한다.

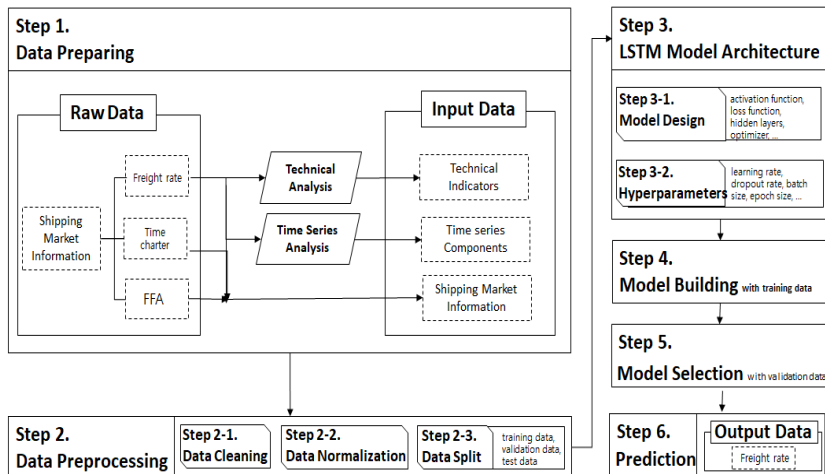


## 제2절 모델구조 및 학습방법

### 1. 모델 구축 과정

모델의 구축과정은 <그림 3-5>와 같다. 모델을 구축하기에 앞서 데이터를 준비한다. 데이터는 입력변수로 활용되기 전에 전처리 과정을 거친 뒤, 기계의 학습을 위한 훈련데이터, 학습된 모델의 검증을 위한 검증데이터, 모델의 예측성능을 최종적으로 판단하기 위한 테스트 데이터로 구분하여 준비한다. 이후, 모델의 구조를 설계(Model Design)하는 과정에서 활성화 함수, 손실함수, 은닉층의 개수 등을 결정한 뒤, 하이퍼파라미터(Hyperparameters)를 선정한다. 이후 설계된 모델에 따라 기계가 데이터를 학습할 수 있도록 학습시킨 뒤, 형성된 모델들 중 검증 과정을 거쳐 최종 모델을 선정한다.

<그림 3-6> LSTM 모델의 구축과정



자료: 저자 작성

## 2. 학습데이터

학습에 사용된 데이터는 2014년 2월 24일부터 2020년 8월 28일까지인 1,630개를 사용하였다. 구성되었다. 동 학습데이터는 모델 평가를 위해 훈련 세트와 시험 세트로 나누어 구성하였다. 훈련 세트은 2014년 2월 24일부터 2018년 9월 14일까지 1,141개, 시험 세트은 2018년 9월 17일부터 2020년 8월 28일까지 489개로 구성하였다.

## 3. 모델구조

본 연구에서는 3개의 은닉층을 가지고 있는 LSTM 모델을 활용했다. 각 은닉층 노드 개수는 32개로 동일하게 구성하였으며 활성화 함수로는 시그모이드 함수와 하이퍼볼릭 탄젠트 함수를 사용하였다. 학습에 사용하는 손실함수는 평균제곱오차(MSE: mean square error)를 최소화하는 방향으로 모델 학습을 진행하였다.

### 1) 손실함수(Loss functions)

손실함수는 순환신경망 모형 구축 시에 함수의 가중치와 편차를 정하기 위해 사용되는 함수로써, 실제값과 예측값과의 차이로 정의된다. 손실함수가 크면 클수록 예측값과 실제값과의 차이가 크다는 의미로, 기계는 손실함숫값이 최소가 되도록 가중치를 조정한다. 주로 손실함수로 MSE가 사용된다.

$$L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

## 2) 활성화 함수(Activation functions)

활성화 함수는 초기 신경망 모형이 일반화되지 못했던 이유인 기울기 소실(gradient vanishing) 문제를 해결하기 위해 고안된 것이다. 기울기 소실 문제란 기계가 신경망 모형 구조를 학습을 진행하면서 가중치와 편향의 크기를 정하게 되는데, 이때 층의 개수가 늘어나게 되면 역전파를 통해 전달되는 기울기의 값이 0에 가까워지거나 0으로 수렴하면서 발생하는 문제를 말한다. 이를 해결하기 위해 활성화 함수를 활용하여 선형적인 데이터의 결합을 비선형적으로 변환시켜 기울기 소실 문제를 해결한다. 활성화 함수에는 시그모이드 함수, 하이퍼볼릭 탄젠트 함수 등이 있다. 시그모이드 함수는 0에서 1 사이의 값을 출력하도록 고안된 활성화 함수이다. 동 함수를 사용하면 입력 신호가 클 경우 1을 출력하며, 반대로 입력신호가 작을 경우 0에 근접한 값을 출력해 다음 노드에 전달한다. 이를 식으로 표현하면 [식 3-1]과 같다.

$$f(z) = \frac{1}{1 + e^{-u}} \dots\dots\dots \text{[식 3-1]}$$

하이퍼볼릭 탄젠트 함수는 입력된 신호를 -1에서 1로 출력하도록 고안된 활성화 함수이다. 동 함수를 활용하면 신호가 클 경우에는 1을 출력하고, 반대로 입력신호가 작을 경우 -1에 가까운 값을 출력하여 다음 노드에 전달한다. 이를 식으로 표현하면 [식 3-2]와 같다.

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \dots\dots\dots \text{[식 3-2]}$$

## 4. 학습방법

### 1) 데이터 정규화(normalization)

데이터의 정규화는 데이터 전처리 단계에서 기계의 학습이 잘되도록 단위가 다른 값들을 일정한 범위로 변환하는 것을 말한다. 이와 같이 범위가 다른 데이터를 정규화하는 이유는 데이터의 범위 차이가 클 경우, 기계가 데이터의 패턴을 학습하며 가중치와 편향의 크기를 정할 때 왜곡된 결정을 내릴 수 있기 때문이다. 즉, 학습이 제대로 이루어지지 않게 되어, 원하는 예측값을 얻지 못하게 된다. 본 연구에서는 아래의 식을 활용하여 입력변수의 범위를 -1에서 1 사이로 정규화하였다. 이미 범위가 고정된 상대강도지수, 스톡캐스틱, 백분위 등은 정규화 과정에서 제외하였다.

$$\frac{2x - (\max - \min)}{\max - \min}$$

### 2) 모델학습

본 연구에서는 데이터를 임의 추출하여 훈련 세트 1,600개와 검정 세트(validation set) 457개를 20번 반복하여 생산하였다. 에포크(epoch)는 10,000번으로 설정하였으며, 손실함수는 앞서 언급한 MSE를 활용하였다. 손실함수의 값이 더 이상 개선되지 않으면 조기 종료한 후, 생성된 모델들 중 MSE값이 가장 작은 모델을 최종 모델로 선정하였다.

## 제 4 장

## 분석결과

## 제1절 순환신경망 모델 분석결과

## 1. 모델 평가 지표

## 1) 평균제곱근편차

예측 모델 간의 기본적인 성능평가를 위하여 다음과 같이 정의되는 RMSE(Root Mean Square Error)를 사용하였다. 여기서  $y_i$ ,  $\hat{y}_i$ 는 각각 실제값과 예측값을 나타내며,  $n$ 은 데이터의 수를 나타낸다. RMSE가 작을수록 예측 모델의 성능이 우수하다고 할 수 있으며, 지표 자체가 직관적이라는 장점을 가지고 있다.

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

## 2) 평균절대백분율편차

심층신경망(DNN: Deep Neural Networks)과 LSTM의 모델의 성능을 비교하기 위해서 MAPE(Mean Absolute Percentage Error)를 활용하였다. MAPE는 백분율로 표현되기 때문에 RMSE가 같거나 비슷한 경우에 모델의 예측값의 변동성을 파악하는 데 유용하다.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{(y_i - \hat{y}_i)}{y_i} \right|$$

## 3) 상관계수

마지막으로, 모델의 선형적인 일치성과 방향성을 측정하기 위한 수단으로 상관계수(Correlation, 이하 “CORR”이라고 한다)를 이용하였다.

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

## 2. 모델 분석결과

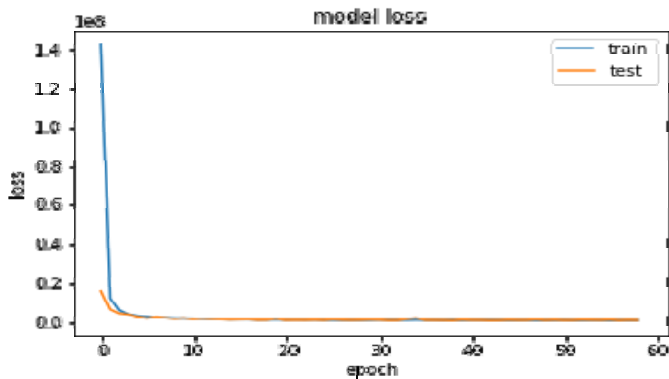
### (1) 학습 과정 및 추정 결과

LSTM 모델의 학습 진행 과정을 <그림 4-1>과 <그림 4-2>에서 제시하였다. 여기서 에포크란 모델 구축과정에서 기계가 학습하는 훈련데이터 셋의 노출 횟수를 말하며, 손실(loss)은 학습을 통해 구축된 모델이 검증데이터의 입력값을 활용해 예측한 값과 실제값 간의 차이를 말한다. 모델의 과적합을 방지하기 위해 손실값이 더 이상 떨어

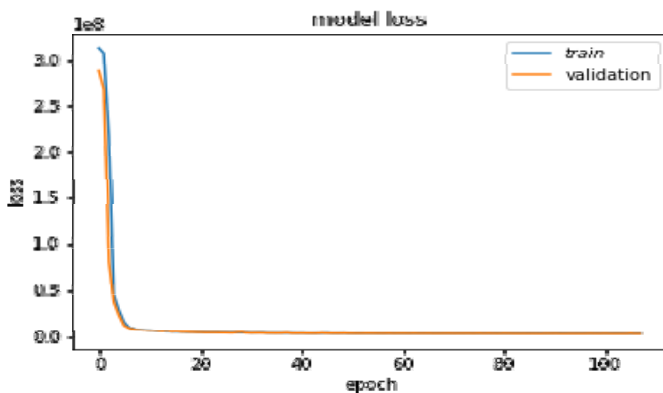
지지 않는 시점에서 기계의 학습을 멈추었는데 이를 조기 종료(early stopping)이라 한다.

LSTM의 모델은 2가지를 제시하였는데, 하나는 데이터가 스팟운임의 기술변수와 FFA, 용선료를 입력변수로 하는 모델과 동 데이터에 스팟운임을 시계열 분해 방법 중 하나인 가법모형으로 분해한 데이터인 추세, 시계열, 랜덤이 추가된 데이터 셋을 입력변수로 하는 모델 LSTM\_ts을 제시하였다.

〈그림 4-1〉 LSTM 모델의 학습 곡선



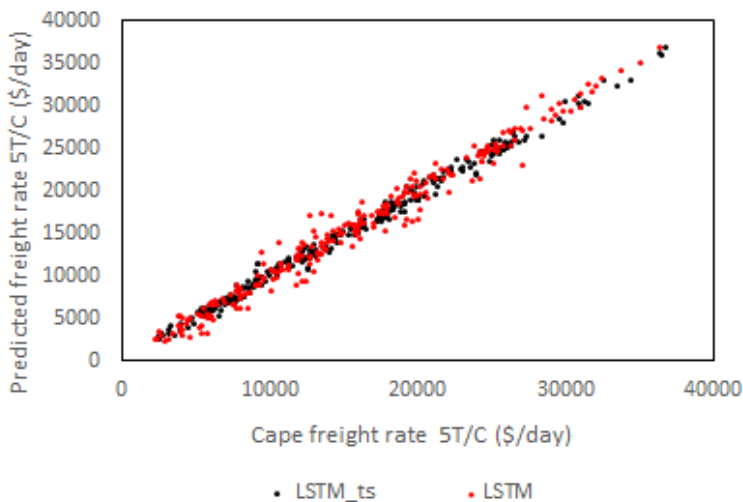
〈그림 4-2〉 LSTM\_ts 모델의 학습 곡선



## 1) 케이프선

LSTM를 이용하여 케이프선 스팟운임을 예측하는 모델을 구축했다. 주간단위 예측을 수행하였으며, LSTM의 모델은 2가지를 제시하였다. 각각의 모델의 예측값과 실제값을 비교한 결과를 살펴보면 <그림 4-3>과 같다. 제시한 두 모델 모두 편향은 낮아, 실제값에 거의 근사하게 예측하는 것을 볼 수 있다. 반면, LSTM 모델의 결과가 LSTM\_ts 모델보다 상대적으로 분산이 크다.

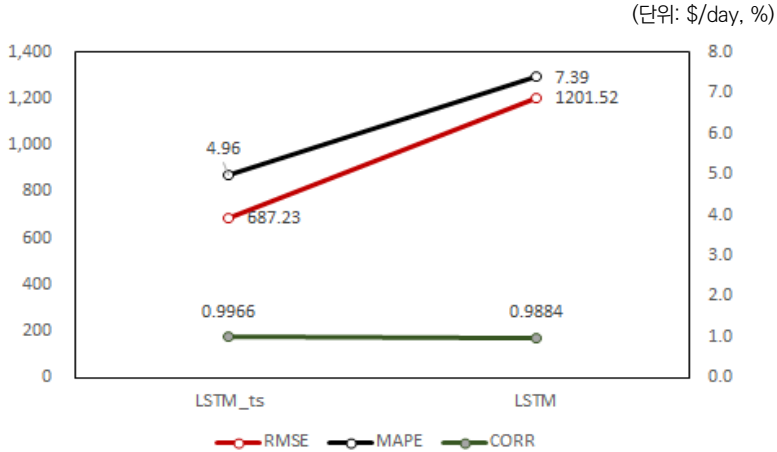
<그림 4-3> LSTM 모델 예측 결과 (케이프선 스팟운임)



모델의 성능을 지표로 비교해보면, CORR은 0.9 이상으로 실제값과 예측값 간의 방향성이 유사함을 알 수 있다. LSTM 모델의 RMSE는 1,201.52이며, LSTM\_ts 모델의 RMSE는 687.23으로 성능 면에서 큰 차이를 보이고 있다. 이는 MAPE에서도 나타나는데 LSTM은 7.39%로 LSTM\_ts(4.96%)보다 오차치가 큰 것을 알 수 있다(<그림 4-4>와 <표 4-1>).



〈그림 4-4〉 LSTM 모델 성능 비교 (케이프션 스팟운임)



〈표 4-1〉 데이터 구성에 따른 LSTM기반 케이프션 스팟운임 예측 모델의 성능 지표 비교

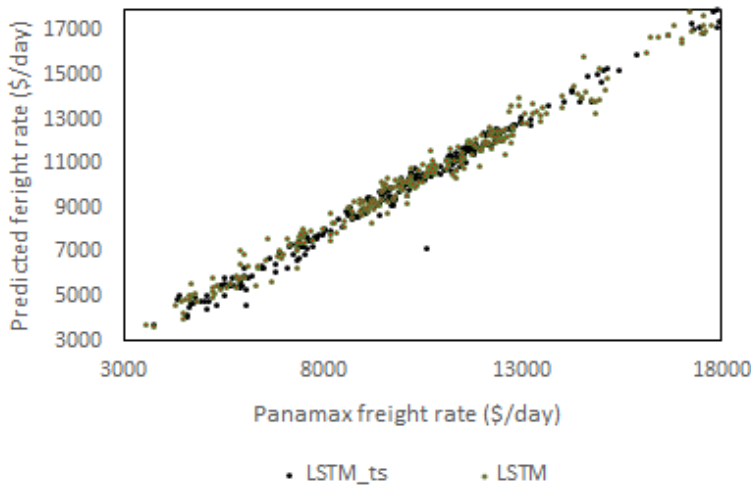
구분	LSTM	LSTM_ts
MAPE (%)	7.39	4.96
RMSE (\$/day)	1201.52	687.23
CORR	0.9884	0.9966

## 2) 파나막스선

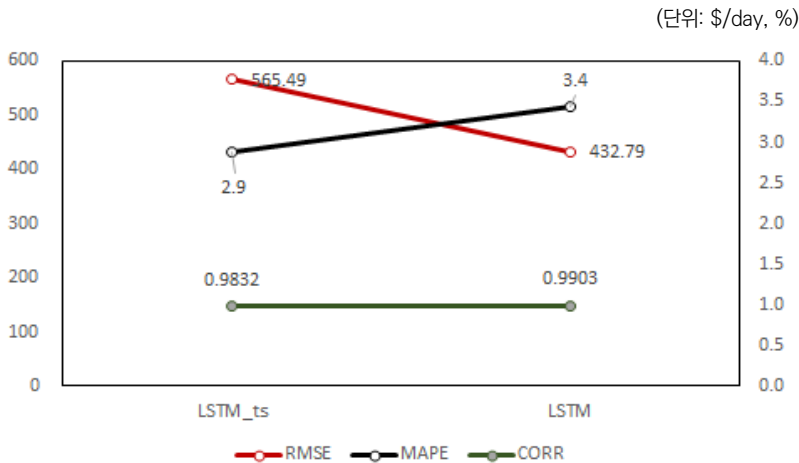
파나막스선 스팟운임을 예측하기 위해 LSTM을 활용하였다. 데이터 셋의 구성에 따라 LSTM과 LSTM\_ts 모델을 제시하였는데, 그 결과는 다음과 같다. 두 모델의 실제값과 예측값의 산점도 그래프를 그려보면, 두 모델 모두 분산과 편향이 낮은 것으로 나타난다. 모델 성능 지표를 비교해보면, LSTM 모델의 RMSE는 432.79로 LSTM\_ts(565.49)보다 예측 성능이 좋은 것으로 나타났다. 다만, MAPE를 비교하면 LSTM은 3.4%, LSTM\_ts는 2.9%로 실제 값에 가깝게 예측하는 데는 LSTM\_ts가 성능이 더 좋은 것으로 나타났다. 산점도 그래프상에

LSTM\_ts의 이상치 값이 발견됨에 따라 RMSE가 LSTM보다 크게 나타난 것으로 보여, LSTM의 분산이 LSTM\_ts의 분산보다 낮은 것이 원인으로 분석된다.

〈그림 4-5〉 LSTM 모델 예측 결과 (파나막스선 스팟운임)



〈그림 4-6〉 LSTM 모델 성능 비교 (파나막스선 스팟운임)



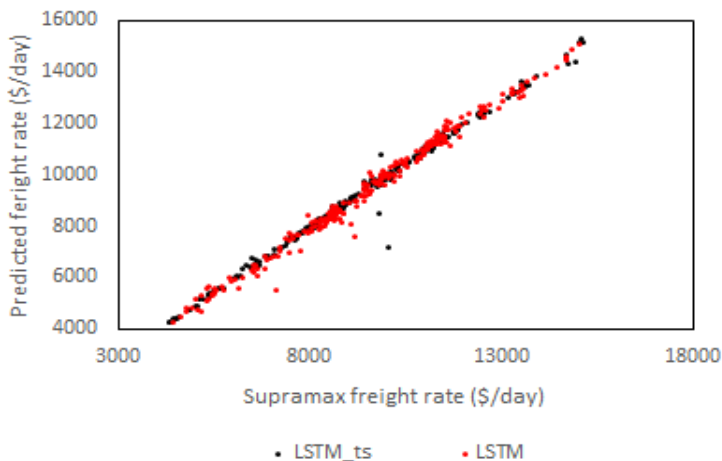
〈표 4-2〉 데이터 구성에 따른 LSTM기반 파나마maks선 스팟운임 예측 모델의 성능 지표 비교

구분	LSTM	LSTM_ts
MAPE (%)	3.4	2.9
RMSE (\$/day)	432.79	565.49
CORR	0.9903	0.9832

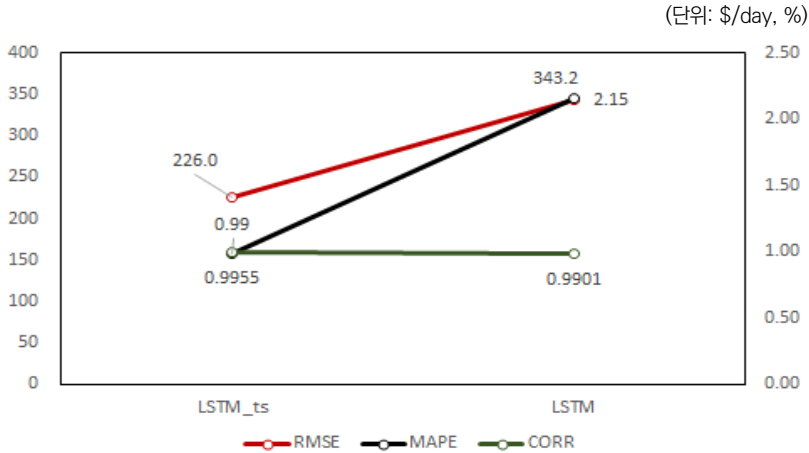
### 3) 수프라maks선

LSTM기반 수프라maks선 예측 모델도 데이터 셋의 구성에 따라 두 가지 모델이 제시되었다. 산점도 그래프인 〈그림 4-7〉에서 보는 바와 같이 두 모델 모두 편향이 낮은 것으로 나타난다. 상대적으로 LSTM 모델의 분산이 LSTM\_ts보다 높은 것으로 관찰된다. 성능 지표를 비교해보면, 시계열 분해 데이터가 추가된 LSTM\_ts의 성능이 LSTM보다 좋은 것을 알 수 있다. LSTM의 RMSE는 343.1로 LSTM\_ts보다 크기가 크며, MAPE도 LSTM이 0.99%인데 반해 LSTM\_ts는 2.15%로 계산되었다.

〈그림 4-7〉 LSTM 모델 예측 결과 (수프라maks선 스팟운임)



〈그림 4-8〉 LSTM 모델 성능 비교 (수프라막스선 스팟운임)



〈표 4-3〉 데이터 구성에 따른 LSTM기반수프라막스선 스팟운임 예측 모델의 성능 지표 비교

구분	LSTM	LSTM_ts
MAPE (%)	2.15	0.99
RMSE (\$/day)	343.2	225.983
CORR	0.9901	0.9955

## 제2절 모델성능비교

### 1. 비교 모델

비교 모델은 딥러닝에서 가장 기본으로 불리는 모델인 심층신경망 (DNN: Deep Neural Networks)을 선정했다. 심층신경망은 은닉층이 여러 개로 구성되어 있다.

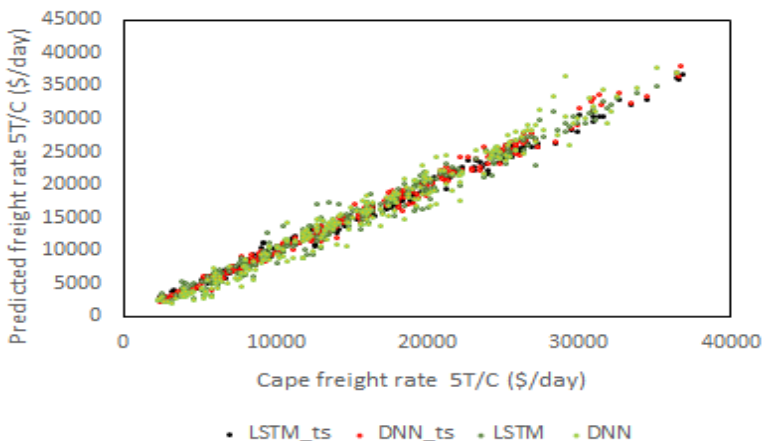
## 2. 모델성능비교 결과

### 1) 케이프선

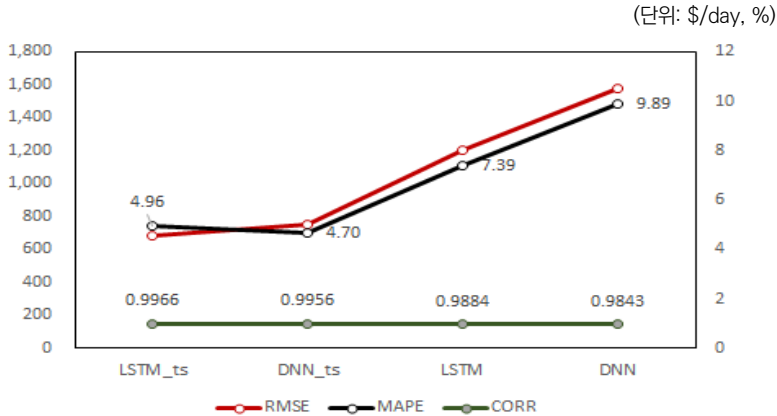
보고서에서 데이터 구성과 인공지능 알고리즘 조합에 따라 총 4개의 모델을 제시하였다. 데이터의 구성에 따라 스팟운임과 기술분석, FFA, 용선료를 데이터로 활용한 모델과 여기에 시계열 분해 방법을 이용해 스팟운임의 추세, 순환, 계절, 불규칙을 포함한 데이터를 입력변수로 하는 모델이다. 그다음으로 기존 예측에 주로 활용되는 DNN와 시계열을 예측하는 데 특화된 인공지능 모델인 LSTM을 활용한 모델을 제시하였다.

RMSE를 비교해보면 LSTM\_ts 모델이 가장 예측 성능이 좋으며 그다음으로 DNN\_ts, LSTM, DNN 순으로 나타난다. 반면 MAPE는 DNN\_ts, LSTM\_ts, LSTM, DNN 순으로 시계열의 특성을 분해한 데이터를 활용해 구축한 모델의 성능이 더 좋은 것으로 나타났다.

〈그림 4-9〉 DNN 모델과 예측 성능 비교(케이프선 스팟운임)



〈그림 4-10〉 지표 활용 모델의 예측성능 비교(케이프션 스팟운임)



〈표 4-4〉 알고리즘 및 데이터 구성 조합 별 케이프션 스팟운임 예측 모델의 성능 지표 비교

구분	LSTM_ts	DNN_ts	LSTM	DNN
MAPE (%)	4.96	4.70	7.39	9.89
RMSE (\$/day)	687.23	753.61	1201.52	1576.28
CORR	0.9966	0.9956	0.9884	0.9843

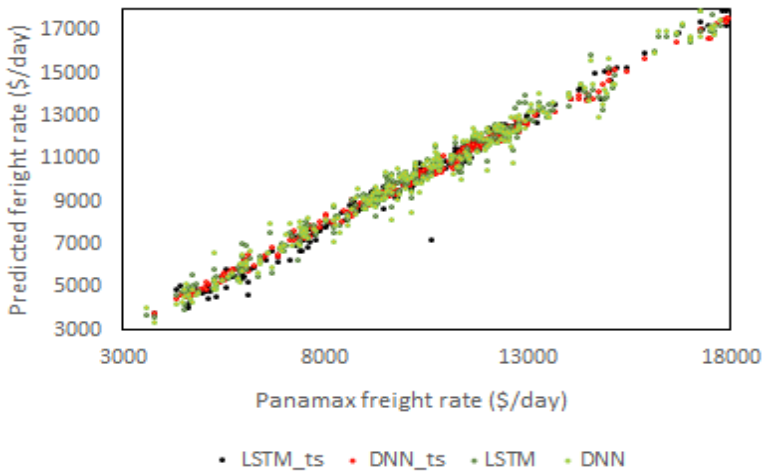
## 2) 파나마스선

파나마스선의 예측 성능을 살펴보면 LSTM을 비롯하여 제시한 모든 모델의 분산과 편차가 크지 않은 것으로 나타났다. 산점도 그래프 상에서 가장 분포가 높은 것은 DNN으로 나타났으나, RMSE는 LSTM\_ts이 566.49로 DNN(493.85)보다 큰 값이 나왔는데, 이는 LSTM\_ts에서 나타나는 이상치 때문인 것으로 분석된다. 〈그림 4-11〉에서도 LSTM\_ts의 산점도가 다른 모델에 비해 분산과 편차가 낮은 것으로 나타나, 이상치가 나타남을 알 수 있다. 이러한 이상치가 RMSE에 부정적인 영향을 미친 것으로 보인다.

MAPE를 비교하면 2.9%로 LSTM\_ts가 가장 실제값에 근사한 값을

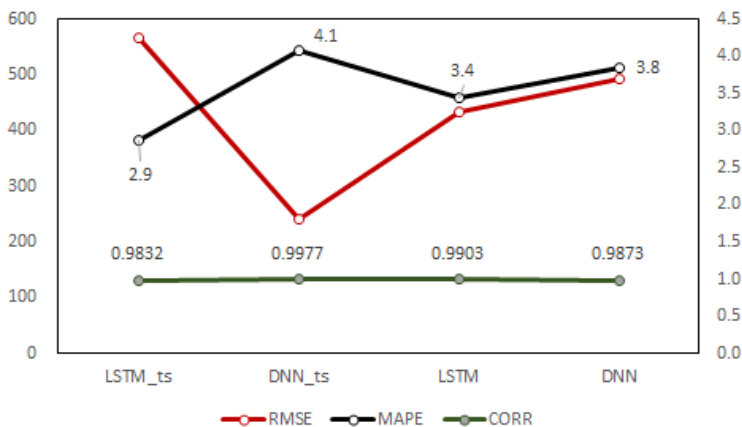
예측하는 것으로 나타나며, 그 뒤를 이어 LSTM이 3.4%, DNN이 3.8%, DNN\_ts가 4.1%이다. RMSE에서 가장 작은 값을 보인 DNN\_ts가 MAPE로 비교했을 때, 가장 예측 성능이 떨어진 것으로 나타났다.

〈그림 4-11〉 DNN 모델과 예측성능 비교 (파나마스선 스팟운임)



〈그림 4-12〉 지표 활용 모델의 예측성능 비교(파나마스선 스팟운임)

(단위: \$/day, %)



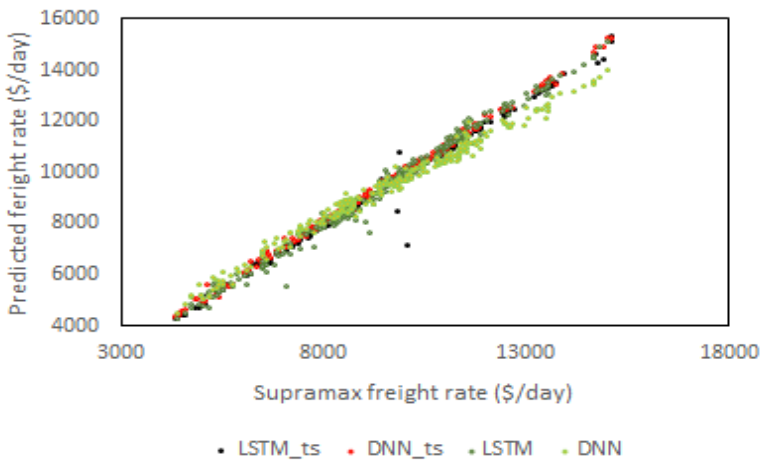
〈표 4-5〉 알고리즘 및 데이터 구성 조합 별파나막스선 스팟운임 예측 모델의 성능 지표 비교

구분	LSTM_ts	DNN_ts	LSTM	DNN
MAPE (%)	2.9	4.1	3.4	3.8
RMSE (\$/day)	565.49	241.10	432.79	493.85
CORR	0.9832	0.9977	0.9903	0.9873

### 3) 수프라막스선

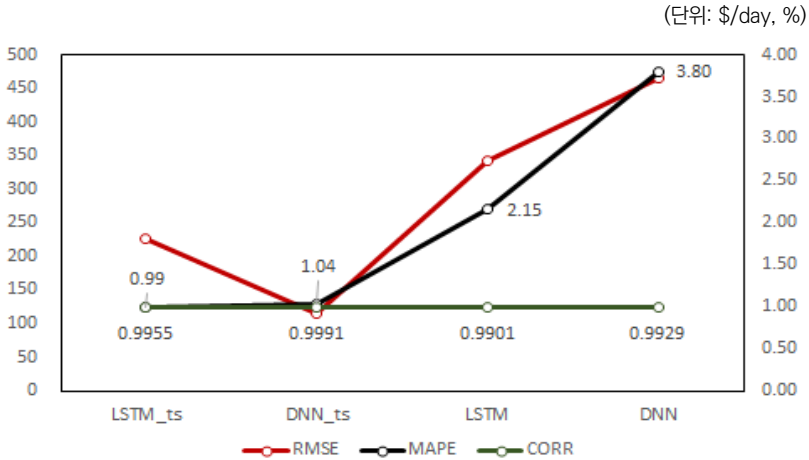
수프라막스선의 예측 성능을 비교해보면, RMSE는 DNN\_ts가 가장 작은 값을 보이는 것으로 나타났으며, 그 뒤를 이어 LSTM\_ts, LSTM, DNN 순이다. 반면 MAPE는 LSTM\_ts가 실제값과 가장 근사한 값을 예측하는 것으로 나타났으며(0.99%), DNN\_ts도 1.04%로 예측력이 높은 것으로 나타났다. 그 뒤를 이어 LSTM(2.15%), DNN(3.80%) 순이다. 수프라막스선에 있어서는 LSTM보다 DNN의 예측 성능이 좋은 것으로 나타났다.

〈그림 4-13〉 DNN 모델과 예측성능 비교(수프라막스선 스팟운임)





〈그림 4-14〉 지표 활용 모델의 예측성능 비교 (수프라막스선 스팟운임)



〈표 4-6〉 알고리즘 및 데이터구성 조합별 수프라막스선 스팟운임 예측 모델의 성능 지표 비교

구분	LSTM_ts	DNN_ts	LSTM	DNN
MAPE (%)	0.99	1.04	2.15	3.80
RMSE (\$/day)	225.983	115.379	343.2	465.26
CORR	0.9955	0.9991	0.9901	0.9929



## 제 5 장

## 결 론

## 제1절 연구결과 요약

## 1. 순환신경망 모델 분석결과

순환신경망 모델을 활용하여 건화물선 운임을 예측하였다. 입력변수로는 각 선형별 운임의 기술적 분석 지표와 시계열 분해 요소를 활용하였고, LSTM을 적용하여 주간단위의 운임을 예측하였다. 최종적으로 성능 지표가 가장 좋은 모델을 채택하였다. 그 결과 케이프선의 MAPE는 4.96%, 파나마스에서는 2.9%, 수프라막스에서는 0.99%로 나타났다.

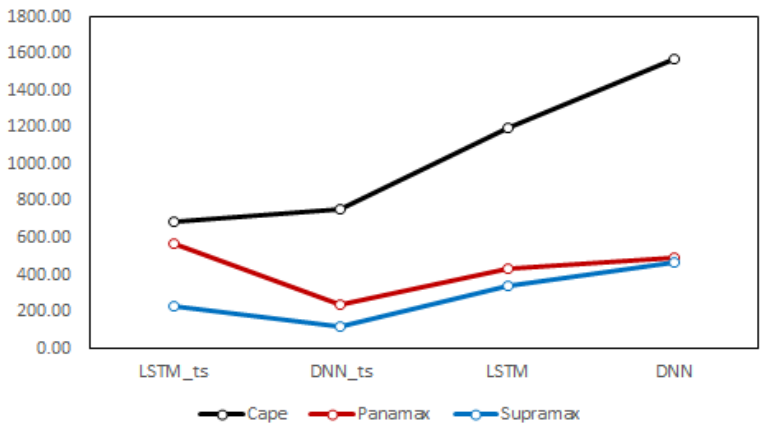
## 2. 비교 모델 분석결과

선형별 모델의 운임예측 성능을 RMSE를 기준으로 비교해보면, DNN\_ts의 성능이 가장 좋은 것으로 나타났으며, 그 뒤를 이어

LSTM\_ts로 시계열 분해 결과를 데이터로 포함했을 때, 예측 성능이 좋아지는 것을 알 수 있다.

〈그림 5-1〉 선형별 모델성능 비교(RMSE값)

(단위: \$/day)



〈표 5-1〉 선형별 모델성능비교(RMSE값)

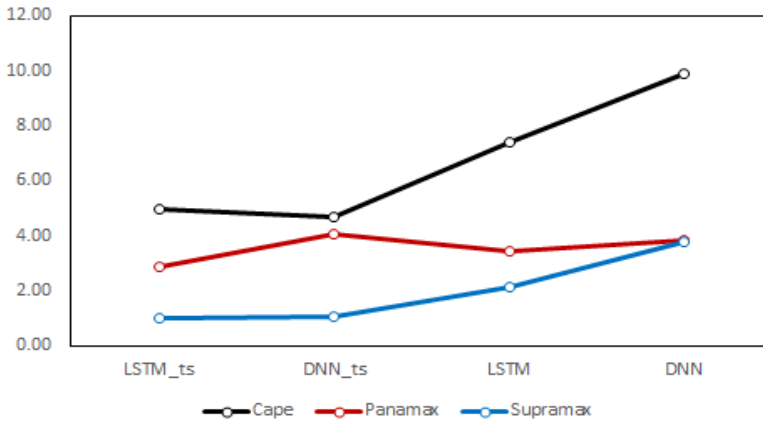
(단위: \$/day)

구분	LSTM_ts	DNN_ts	LSTM	DNN
케이프선	687.23	753.61	1201.52	1576.28
파나막스선	565.49	241.10	432.79	493.85
수프라막스선	225.98	115.38	343.20	465.26

MAPE를 기준으로 모델의 성능을 살펴보면 RMSE와는 다른 결과가 나온다. 스왈운임을 시계열 특성인 추세, 순환, 계절, 불규칙을 따로 분해해 입력변수로 활용한 모델이 특성을 활용하지 않은 모델에 비해 성능이 좋은 것으로 나타난 것은 같으나, MAPE만을 놓고 보면 LSTM을 활용해 모델을 구축했을 때, 실제값에 더 가깝게 예측할 수 있는 것으로 나타났다.

〈그림5-2〉 선형별 모델성능 비교(MAPE값)

(단위: %)



〈표 5-2〉 선형별 모델성능비교(MAPE값)

(단위: %)

구분	LSTM_ts	DNN_ts	LSTM	DNN
케이프선	4.96	4.70	7.39	9.89
파나막스선	2.9	4.1	3.4	3.8
수프라막스선	0.99	1.04	2.15	3.80

## 제2절 연구의 기여

본 연구는 운임의 시계열적 특성을 활용하여 주간단위 예측에 대한 인공지능 모델의 유효성을 검증했다. 주간단위의 예측은 해운기업이 영업전략 수립 중 용대선 시점을 결정하는 데 중요하게 활용되는 정보이다. 주간단위 예측 정보는 특히 항차단위의 용대선을 할 경우 특히 유용한 정보가 된다. 동 연구에서는 주간단위의 단기 운임을 예측하는 데 운임의 시계열적 특성을 활용할 수 있도록 고안된 순환

신경망 모델을 활용하고 이에 대한 유효성을 검증하였다. 일반적으로 해상운임은 시계열적 특성을 가지고 있다고 알려진 데이터로, 이러한 데이터 특성을 활용하기 위해 ARIMA, VAR, VECM를 활용하여 운임예측 모형을 구축한다. 하지만 기존의 DNN 모델은 모델 구조의 특성상 해상운임의 시계열적 특성을 활용하지 못한다는 한계가 있었다. DNN 모델은 주어진 입력값을 독립적으로 학습하도록 만들어진 구조이기 때문에 추세 등과 같은 입력값의 시계열적 특성을 활용하지 못한다. 시간에 따른 정보를 처리하지 못하는 DNN 모델의 구조적 한계를 보완한 모델이 순환신경망 모델이다. 동 연구에서는 운임 데이터의 시계열적 특성을 시계열 분해법을 활용하여 1차적으로 활용했으며, 2차적으로는 시계열적 특성을 활용하도록 고안된 LSTM과 같은 순환신경망 모델을 활용하여 모델을 구축하였다. 본 연구를 통해 제시된 LSTM기반 모델이 DNN 모델보다 향상된 운임예측 성능을 보였다.

### 제3절 추가연구

본 연구는 인공지능 알고리즘 중 고전적이라고 할 수 있는 순환 신경망 모델을 활용해 건화물선의 해상운임을 예측하였다. 건화물선 해상운임예측 모형의 예측 성능 향상 및 예측 기간을 늘리기 위해 새롭게 제시되거나 개선된 인공지능 모델의 활용을 검토할 필요가 있다. 최근 인공지능 모델도 대표적인 지도학습<sup>21)</sup> 알고리즘인 DNN, RNN, LSTM 이외에도 기존의 알고리즘의 한계를 개선하거나 조합한

21) 지도학습이란 모델의 학습과정에서 예측하려는 값을 알려주고 학습시키는 방법을 말한다.

하이브리드 형태의 모델들이 등장하고 있다. 동 연구에서 LSTM과 같은 순환신경망 모델을 활용하여 운임 데이터에서 얻을 수 있는 시계열적 특성을 활용하고자 하였다. LSTM과 같은 순환신경망 모델은 시계열적 특성을 활용할 수 있다는 강점을 가지고 있는 반면, 예측값이 기계가 학습한 바로 직전 값을 기반으로 수렴하려는 경향이 있다. 즉, 예측의 패턴이 예측 직전의 패턴과 유사하다는 것이다. 이 한계를 해결하기 위해 최근 새롭게 제안되고 있는 모델이 양방향 순환신경망(Bi-RNN), 하이브리드 모델이다. 인공지능망 모델 알고리즘의 개발 흐름이 빠르다는 점을 감안하여, 해상운임예측 모형 구축에도 다양한 알고리즘을 적용하여야 할 것이다.

또한 해운기업의 중장기적인 영업전략 수립 시에 활용될 수 있도록 예측 기간을 확장하는 연구가 추가되어야 할 것이다. 주간단위의 예측은 변동성이 큰 건화물선 시장, 특히 스팟으로 용대선 하는 경우 해운기업에게 유용한 정보로 활용될 수 있으나, 신조선 및 중고선 확보 등의 큰 규모의 금액을 투자하는 경우에는 단기적 예측보다는 중장기적 예측 정보가 필요하다. 예측 기간을 늘리기 위해 다양한 인공지능 모델 적용과 입력변수 선정에 대한 연구가 수행되어야 한다. 해상운임은 수급균형에 따라 결정되며, 예측 기간이 길어질수록 운임 이외에 운임을 결정하는 공급과 수요 및 외생변수를 확보해야 할 것이다.





## 참고문헌

## 〈국내 문헌〉

- 나이젤 루이스 지음, 이원상 옮김, 『R로 하는 쉬운 딥러닝』, 에이콘, 2017.
- 마틴 스토포드 지음, 양창호·이충배·이동현·신승식 옮김, 『해운경제학』, 박영사, 2015.
- 모수원, 「2010년 BDI의 예측 - ARIMA 모델과 HP 기법을 이용하여」, 『한국항만경제학회지』, 제26집 제1호, 2010.
- 사이토 고키 지음, 개앞맵시 옮김, 『밑바닥부터 시작하는 딥러닝: 파이썬으로 익히는 딥러닝 이론과 구현』, 한빛미디어, 2017.
- 야마시타 타카요시 지음, 심효섭 옮김, 『그림과 수식으로 배우는 통통 딥러닝』, 제이펍, 2017.
- 이우식, 「딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측」, 『한국데이터정보과학회지』, 제28권 제2호, 2017.
- 임상섭·윤희성, 「Supramax Bulk Carrier Market Forecasting with Technical Indicators and Neural Networks」, 『한국항해항만학회지』, 제42권 제5호, 2018.
- 윤희성·최건우·황수진·박동원, 「딥러닝 기반의 건화물선 시황예측 연구」, 한국해양수산개발원, 2019.
- 정봉민, 『해운경제학』, 블루&노트, 2015.
- 존 J. 머피 지음, 최용석 옮김, 『금융시장의 기술적 분석』, 국일증권경제연구소, 2000.

## 〈국외 문헌〉

- Alizadeh, A.H. "Trading Volume and Volatility in the Shipping Forward Freight Market", *Transportation Research Part E: Logistics and Transportation Review*, 49(1), 2013. 250-265.
- Chou, M.T., "A Fuzzy Time Series Model to Forecast the BDI", *Forth International Conference(IEEE)*, 2, 2008. 50-53.

- Eslami, P., Jung, K., Lee, D., and Tjolleng, A., "Predicting tanker freight rates using parsimonious variables and a hybrid artificial neural network with an adaptive genetic algorithm", *Maritime Economics & Logistics*, 19(3), 2017. 538-550.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.
- Kavussanos, M.G., and Alizadeh-M, A.H., "Seasonality Patterns in Dry Bulk Shipping Spot and Time Charter Freight Rates", *Transportation Research Part E: Logistics and Transportation Review*, 37(6), 2001. 443-467.
- Kavussanos, M.G., and Alizadeh-M, A.H., "Seasonality Patterns in Tanker Spot Freight Rate Markets", *Economic Modelling*, 19(5), 2002. 747-782.
- Kavussanos, M.G., and Visvikis, I.D., "Market Interactions in Returns and Volatilities between Spot and Forward Shipping Freight Markets", *Journal of Banking & Finance*, 28(8), 2004. 2015-2049.
- Lyridis, D. V, Zachraoudakis, P., Mitrou, P., and Mulonas A., "Forecasting tanker market using artificial neural networks," *Maritime Economics & Logistics*, 6(2), 2004., 96-108.
- Shynkevich, Y., McGinnity, T.M., S.A., Belatreche, A., and Li, Y.. "Forecasting price movements using technical indicators: Investigating the impact of varying input window length," *Neurocomputing* 264, 2017. 71-88.
- Song, Q., Chissom, Brad S., "Fuzzy time series and its models", *Fuzzy Sets and Systems*, 54(3), 1993. 269-277.
- Ng, Andrew, "Reinforcement Learning and Control," *Stanford CS229 Lecture Note Part XIII*, Stanford, 2020. 1-7.
- Veenstra, A., and Franses, P.H., "A Co-Integration Approach to Forecasting Freight Rates in the Bulk Shipping Sector", *Transportation Research. Part A: Policy & Practice*, 1997. 447-458.

## 〈인터넷 자료〉

Dive into Deep Learning 홈페이지, [https://ko.d2l.ai/chapter\\_deep-learning-basics/underfit-overfit.html](https://ko.d2l.ai/chapter_deep-learning-basics/underfit-overfit.html) (검색일: 2020. 10. 3)

Medium 블로그, <https://medium.com/@ankitnitjsr13/math-behind-svm-support-vector-machine-864e58977fdb> (검색일: 2020. 10. 6)

Medium 블로그, <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d> (검색일: 2020. 10. 6)

OpenEye Python Cookbook 홈페이지, <https://docs.eyesopen.com/toolkits/cookbook/python/plotting/roc.html> (검색일: 2020. 9. 30)

Tistory 블로그, <https://sumniya.tistory.com/26?category=818582> (검색일: 2019. 9. 30)

## 〈데이터〉

볼틱해운거래소 홈페이지, <https://www.balticexchange.com> (검색일: 2020. 8. 28)

클락슨 홈페이지, <https://sin.clarksons.net> (검색일: 2020. 8. 28)

# 기본연구보고서 발간목록

2020년

01	해양수산업의 지역 간 연관구조 분석	장정인
02	해양수산 분야 기술 대외의존도 분석연구 - 스마트항만을 중심으로 -	전형모
03	지역 해양수산 오픈 플랫폼 구축방안 연구	최지연
04	갯벌 거버넌스 개선방안에 관한 연구	육근형
05	해양환경정책의 능동적 추진을 위한 자원체계 개선 방안 연구	박수진
06	항만 대기환경 관리 표준 및 평가모형 연구	안용성
07	해양수산분야 사회문제해결형 R&D 기반 구축 연구	좌미라
08	해양 유입 하천쓰레기 관리체계 개선방안	이윤정
09	수산물 품질·안전관리 제도 개선방안 연구	이현동
10	국제법 변화에 대응한 어선원 안전 및 권리 제고방안 연구	한덕훈
11	스마트 양식 클러스터 추진 방안 수립 연구	이상철
12	해양포유류 보호에 관한 수산업 대응 방안 연구	정명화
13	수산물 수급통계 개선 방안 연구	김수현
14	IMO 온실가스 규제 대응 정책방향 연구	박한선
15	퇴직전문가 활용을 통한 해운업 경쟁력 강화방안 연구	안영균
16	글로벌 선사들의 물류통합화 전략에 대한 국적선사의 대응방안	전형진
17	내항여객운송항로 정책 발전방안 연구	김태일
18	블록체인 기술기반 식품콜드체인 체계 구축 연구	조지성
19	항만자동화 도입 관련 노무 갈등 해소 방안 연구	김찬호
20	스마트항만과 스마트도시 연계 발전 방안 연구	이연경
21	항만의 회복탄력성 측정 모형 구축에 관한 연구	김성기
22	IMO 규제기반 해사산업의 글로벌 지속발전방안 연구 -新해사산업의경제적파급효과분석연구(4차년도)-	박한선
23	국내 항만연계 산업의 가치사슬 및 공급사슬 연계성 강화방안 - 자동차 산업을 중심으로	신수용

## 2019년

01	한반도 평화 체제 수립 대비 접경수역 연구	최지현
02	수산부문 전망모형 「KMI-FOSiM」 구축 연구(2차년도)	이현동
03	4차 산업혁명 시대의 스마트 어촌 구축방안 연구	박상우
04	해양수산분야 예견적 위험분석 기반구축 연구	최석우
05	재해대응 및 환경보전을 위한 연안토지 매수 이행방안 연구	윤성순
06	선박 대기오염물질 배출저감 기술의 평가·인증체계 구축 및 활용방안 연구	안용성
07	해양레저관광서비스 산업 육성방안 연구	최일선
08	주요국의 나고야의정서 이행체계 강화에 대응한 해양유전자원정책 개선 연구	박수진
09	하천·해양 수질의 연계·통합 관리 체계 연구	장원근
10	총허용어획량(TAC) 기반 수산자원관리 강화 방안 연구	이정삼
11	우리나라 수산업·어촌의 공익적 기능에 관한 연구	류정곤
12	우리나라 원양산업의 사회적 책임 실천 강화를 위한 정책연구	정명화
13	수산업 노동시장 구조 분석과 노동정책 변화에 따른 대응 연구	한광석
14	우리나라 정기선 해운정책 개선 방안 연구	김태일
15	국내외 해운금융 비교를 통한 국내 해운금융 역량강화 방안 연구	전형진
16	해상 사이버 보안체계 강화방안 연구	박한선
17	해운 기업 비즈니스 모델과 경쟁우위 분석 연구	윤희성
18	국내 항만산업이 지역경제에 미치는 영향 분석	하태영
19	동북아 공동배출규제해역 지정 필요성 및 추진방안 연구	이기열
20	수출기업의 글로벌 공급사슬협력 수준 분석과 물류정책 방안 연구 - 화주~물류기업 협력을 중심으로 -	김은수
21	신선식품 수출입 증대를 위한 우리나라 항만물류 개선 방안	신수용
22	환동해권 물류 및 지역개발 비즈니스 모델 제안과 협력방안	박한선
23	IMO 규제기반 해사산업의 글로벌 지속발전방안 연구 - 우리나라 新해사산업의 경쟁력 분석에 관한 연구 -	최나영환

## 수시연구보고서 발간목록

2020년

01	포스트 코로나 19 해양수산 분야 정책방안	박광서
02	생분해성 어구 사용 활성화 방안 연구	심성현
03	해양법 전문인력 양성 방안 연구	박영길
04	무인도서 해양주권 강화와 이용 활성화를 위한 제도 개선방안	정지호
05	지역 해양수산 재정분권 대응방향	황재희
06	데이터 3법 개정에 따른 항만·물류 데이터 활용도 제고 방안	이기열
07	양식 활어 유통 효율화 방안 연구	마창모
08	해양레저관광사업 추진을 위한 제도정비 방안	홍장원
09	해양바이오기업의 규제 정비 방안 연구	최석문
10	비상체제 시 선원의 안전을 위한 선박-항만-항공 이동경로 구축방안	이혜진

2019년

01	일본의 수산물 수입 구조 분석과 對일본 수출 전략 연구	임경희
02	선박부착생물에 대한 선제적 대응을 위한 정책방향 연구	박수진
03	수산물 위생안전을 위한 저온유통체계 구축방안 연구	장홍석
04	전국 사업체조사 마이크로데이터를 활용한 어촌지역 고용통계 (2016, 2017)	한광석

## 일반연구보고서 발간목록

2020년

01	지역 해양관광 경쟁력 지수 체계화 연구	최일선
02	AIS 기반 글로벌 선박 배기가스 배출량 분석 연구	강무홍
03	김 중기 수급전망모형 「Gim-MFoS」 구축 연구	허수진

04	중앙 북극 공해 비규제어업방지협정 이행방안 연구	김민수
05	동해 평화관광구역 조성 방안 연구	윤인주
06	디지털 공급사슬 물류정보통합 구축전략 연구(II) - 일반 수출입 컨테이너 정보교환방식 중심 -	이연경
07	스트레스 테스트를 통한 우리나라 해운·조선 기업의 안정성 분석	박성화
08	인공지능기반 해상운임예측 연구	황수진
09	국제물류주산업 실태분석 및 경쟁력 제고 방향 연구	최나영환
10	디지털화에 따른 개별 직업의 대체 가능성 추정 연구	박희대
11	시계열 분석을 통한 해운시장 분석 및 예측 연구	고병욱
12	해양수산분야 글로벌 경제효과 분석모형(KMI-GEM) 시범 구축 연구	임병호
13	해운항만물류 인력양성사업 관리운영 제도 개선 방안 연구	이자연

## 2019년

01	양식수산물의 경영비 조사체계 구축에 관한 연구	백은영
02	딥러닝 기반의 건화물선 시황예측 연구	윤희성
03	해운-조선산업 관계분석 연구 - 컨테이너 신조발주량 예측 -	최건우
04	디지털 공급사슬 물류정보통합 구축전략 연구( I ) - 디지털 공급사슬 생태계 정보통합지도 구축 중심 -	최상희
05	시민 건강증진을 위한 해안관리 방향	정지호
06	우리나라 주요 국제여객항 운영체계 개선방안 연구 : 운영, 시설, 제도를 중심으로	이경한
07	북한 서해 해양관광 활성화 방안	윤인주
08	EU의 수산물 소비 특성 및 수출 전략(프랑스, 스페인을 중심으로)	김지연
09	해양수산 위성계정 기초연구( I )	장정인
10	우리나라 해운·물류기업의 중국 자유무역시범구 활용방안 연구	김형근
11	지역 해양수산 혁신성장체계 구축 연구	최지연
12	도미니카공화국 뱀장어 양식사업 타당성 조사	마창모
13	강원도 어촌 신활력 제고방안(2차년도) - 어촌 활성화 거버넌스 구축과 신활력 사업구상 -	박상우



일반연구 2020-08

## 인공지능기반 해상운임예측 연구

---

2020년 12월 29일 인쇄

2020년 12월 31일 발행

---

발 행 인 | 장 영 태

발 행 처 | 한국해양수산개발원

49111 부산시 영도구 해양로 301번길 26(동삼동)

연 락 처 | 051-797-4800 (FAX 051-797-4810)

등 록 | 1984년 8월 6일 제313-1984-1호

조판·인쇄 | 한성애드컴 (02-2266-6559)

---

판매 및 보급: 정부간행물판매센터 Tel : 02-394-0337

정가 6,000원