

머신러닝을 이용한 철광석 가격 예측에 대한 연구

(Forecasting of Iron Ore Prices using Machine Learning)

이 우 창¹⁾, 김 양 석²⁾, 김 정 민³⁾, 이 충 권^{4)*}

(Woo Chang Lee, Yang Sok Kim, Jung Min Kim, and Choong Kwon Lee)

요 약 철광석의 가격은 여러 국가와 기업들의 수요와 공급에 따라서 높은 변동성이 지속되고 있다. 이러한 비즈니스 환경에서 철광석의 가격을 예측하는 것은 중요해졌다. 본 연구는 머신러닝 기법을 이용하여 철광석이 거래되는 시점으로부터 한 달 전에 철광석 거래가격을 미리 예측하는 모형을 개발하고자 하였다. 예측 모형은 시계열 데이터를 활용한 예측 방법론으로 많이 활용되고 있는 시차분포 모형과 다층신경망 (Multi-layer perceptron), 순환신경망 (Recurrent neural network), 그리고 장단기 기억 네트워크 (Long short-term memory)와 같은 딥 러닝(Deep Learning) 모형을 사용하였다. 측정지표를 통해 개별 모형을 비교한 결과에 따르면, LSTM 모형이 예측 오차가 가장 낮은 것으로 나타났다. 또한, 앙상블 기법을 적용한 모형들을 비교한 결과, 시차분포와 LSTM의 앙상블 모형이 예측 오차가 가장 낮은 것으로 나타났다.

핵심주제어: 머신러닝, 철광석 가격, 그레인저 인과관계, 시계열 예측

Abstract The price of iron ore has continued to fluctuate with high demand and supply from many countries and companies. In this business environment, forecasting the price of iron ore has become important. This study developed the machine learning model forecasting the price of iron ore a one month after the trading events. The forecasting model used distributed lag model and deep learning models such as MLP (Multi-layer perceptron), RNN (Recurrent neural network) and LSTM (Long short-term memory). According to the results of comparing individual models through metrics, LSTM showed the lowest predictive error. Also, as a result of comparing the models using the ensemble technique, the distributed lag and LSTM ensemble model showed the lowest prediction error.

Keywords: Machine learning, Price of iron ore, Granger causality, Time-series forecasting

1. 서 론

* Corresponding Author: ckleee@kmu.ac.kr

Manuscript received February 10, 2020 / revised April 14, 20xx / accepted April 14, 2020

- 1) 계명대학교 경영정보학과, 제1저자
- 2) 계명대학교 경영정보학과, 제2저자
- 3) 계명대학교 경영정보학과, 공동저자
- 4) 계명대학교 경영정보학과, 교신저자

인류 역사에서는 불의 발견 및 이용에서부터 시작하여 종이와 화약, 그리고 전기의 발명과 같이 문명을 발전시키는 과학기술이 있었다. 이러한 과학기술 중에서 철의 사용은 기원전 1200년경부터 이루어졌다. 철기시대의 초기에는 인류가 철을 이용하여 이전보다 더욱 단단하고 정밀한 무기나 농기구를 제작할 수 있었고, 인간

이 삶을 영위하는 데 있어서 안전성과 편리성을 제공해 주었다. 현대 문명에서는 철강을 이용하여 총, 잠수함과 같은 군사 무기와 철제 농기구를 제작할 수 있었다. 철강은 교량이나 건물, 자동차와 같은 제조업에서 만들어내는 물품들의 주 소재로 사용되고 있다. 이러한 철강은 철광석의 채산, 제강, 압연공정의 과정을 통해 제조되는데 철강을 가공하는 획기적인 기술을 가지고 있지만 국내에서의 철광석 채광이 어려울 경우, 철광석을 수입하게 되는 경우가 발생한다. 반대로 원자재인 철광석을 필요 이상으로 많이 보유하고 있거나 철강을 가공하는 기술이 부족한 국가는 철광석을 수출하는 경우가 발생한다. 따라서 철광석의 가격은 많은 국가와 기업들의 수요와 공급에 따라서 가격이 결정짓게 된다.

주요 원자재 가격의 변동계수 분석에 의하면 2017년 철광석 가격의 변동계수가 2위를 기록하면서 철광석 가격의 변동성이 높은 수준을 유지하고 있는 것으로 나타났다 (Huh, 2018). 이는 철광석 공급 및 원료구매의 위험성이 확대되고 있음을 의미한다. 이러한 비즈니스 환경에서 철광석의 가격을 예측하는 문제가 중요하게 되었다. 따라서 본 연구는 철광석의 가격을 미리 예측하여 철강을 다루는 국가나 기업의 투자위험성을 줄일 수 있는 머신러닝 (Machine learning) 모형을 개발하고자 하였다.

본 연구에서 사용되는 머신러닝 모형은 시계열 데이터를 활용한 예측 방법론으로 많이 활용되고 있는 시차분포 (Distributed lag: DL) 모형과 MLP (Multi-layer perceptron), RNN (Recurrent neural network), LSTM (Long short-term memory)과 같은 신경망 모형이다. 해당 모형을 통해 철광석이 거래되는 시점으로부터 한 달 전에 철광석 거래가격을 미리 예측하는 모형을 구축하고자 한다. 시차분포 모형은 이러한 문제를 해결하기 위한 전통적인 통계 모형이다. MLP, RNN 그리고 LSTM은 시계열 데이터를 모델링 하는 딥 러닝 기법으로 알려져 있다. 해당 4개의 알고리즘은 기계학습을 통해 주어진 데이터를 입력받고 결과물을 제시한다는 점에서 모두 머신러닝 모형이라고 할 수 있다. 본 연구는 이러한 기법들을 예측에 적용하여 성과를 비교하고, 서로 다른 모델들을 앙상블 (Ensemble)하여 성과를 개선할 수 있는 방법을 제안하고자 한다.

2. 선행 연구

2.1 철광석 가격 예측과 관련된 선행연구

초기 연구에서는 철광석 가격의 예측보다 철강에 대한 가격 예측 연구가 먼저 이루어졌다. Malanichev and Vorobyev (2011)는 시장에 공급되는 압연강재의 공급량, 산업에서의 시장 소비량에 따른 철강의 생산량, 거시경제 변수, 세계 철강 보유량, 생산원가 변수를 사용하여 1년 단위의 장기적인 세계 철강 가격을 시계열 모형과 선형회귀 모형을 통해 예측하고자 하였다.

Wu et al. (2012)은 SVM (Support vector machine) 알고리즘을 이용하여 중국에서의 철광석 수입가격을 예측하는 모형을 개발하였고, 철광석 가격을 예측하는 것에 있어 머신러닝 모형에 대한 우수성을 검증하였다. Pustov et al. (2013)은 한계비용 접근법과 인센티브 가격 계산이라는 두 가지 접근법에 근거하여 철광석 시장의 구조적 변화를 관찰하였고, 몬테카를로 시뮬레이션 접근법을 사용하여 미래 철광석 가격에 대한 신뢰구간을 추정하였다.

Nam (2016)은 철광석 가격과 관련성이 높을 것으로 예상되는 거시경제, 원자재 경기, 철강 경기, 철강 수급의 4가지 부문에서 13개 지표를 수집하여 신호 접근법을 통해 중국 철광석 가격의 급등에 대한 조기경보 시스템을 연구하였다. 그 결과, 테스트 기간 동안 10번 중 7번의 확률로 1년 후의 철광석 가격의 위기 발생 유무를 비교적 정확하게 예측했다. 또한 해당 연구에서는 13개의 선행지표 이외에 다수의 선행지표 후보군을 대상으로 연구를 진행하였다면 더욱 높은 수준의 예측 정확도를 얻을 수 있었을 것이라는 시사점을 남겼다. Nam (2018)은 인공지능 기법인 SVM 모델과 랜덤 포레스트 (Random forest) 모델을 사용하여 기존 Nam (2016)의 연구보다 정확도를 10% 이상 향상시킬 수 있었다. Lee et al. (2019a)는 예측 시점으로부터 1주 전에 중국 철광석 가격에 영향을 미치는 요인을 탐색하고자 시도하였고 ARDL (Auto regressive distributed lag) 모형을 사용하여 예측 시점으로부터 과거의 중국, 일본, 한국 세 개국의 철강의 수출입과 관

런된 요인이 1주라는 짧은 기간 내에서 중국 철광석 가격에 영향을 미친다는 것을 모형을 통해 검증하였다.

2.2 머신러닝 모형을 적용시킨 사례와 관련된 선행연구

본 연구에서는 시차분포, MLP, RNN, 그리고 LSTM을 사용하였다. 이러한 4 가지 머신러닝 모형들은 다양한 분야에서 사용되어왔다.

Rhee (2017)은 시차분포 모형을 이용하여 경제제재가 무역에 미치는 영향을 연구하였고, 1990년대 이후 경제제재가 부과된 6개국의 무역 반응을 추정된 결과, 4 또는 5분기 동안 무역 반응이 감소하였고 10 내지 12분기 후에는 영향이 사라진다는 것을 발견하였다. Kim and Kim (2018)은 시차분포모형을 사용하여 갈치 시장의 산지, 도매, 수입 가격 간의 인과관계를 분석하였고, 산지, 도매 그리고 수입 가격은 장기적으로 균형을 이루고 있으며 산지 가격과 수입가격은 단기적으로 서로 유의한 영향을 주고받고 있다는 것을 검증하였다. Chung and Lee (2018)는 시차분포 모형을 이용하여 주택 매매가격 대비 전세가격 비율을 예측하는 연구를 시도하였고, 주택 매매가격 대비 전세가격 비율을 결정하는 요인을 분석하였다.

Kim and Kim (2001)은 MLP 모형을 기반으로 한국어 숫자음의 인식 시스템 구현에 필요한 특징을 연구하였다. Hong (2009)은 MLP 모형을 이용하여 아파트 가격을 예측하는 모형을 개발하고자 하는 시도가 있었고 해당 연구에서 수치형 자료를 예측하는 데 있어 MLP와 같은 신경망 모델의 우수성을 검증하였다. Choi et al. (2019)은 환경요인을 이용하여 MLP 기반 온실내 기온 및 상대습도를 예측하는 연구를 진행하였다. Moon et al. (2019)은 MLP를 사용하여 온실가스 배출 및 에너지 부족 문제에 따라 에너지 효율성을 최적화시키는 효과적인 전력수요 예측 기법을 개발하였다.

Jung et al. (2017)은 대도시 도심부의 교통혼잡도를 예측하기 위해 RNN 모형을 사용하여 우수한 성능을 얻을 수 있었다. Song (2017)은

RNN 모형을 사용하여 고객들의 반복 구매제품의 재구매 시기를 예측하는 모형을 개발하였고 모형의 파라미터를 조정하여 최적 모형을 찾은 결과, 은닉층의 수와 훈련의 반복 횟수가 많아짐에 따라 과적합이 나타남을 발견하였다.

Kim and Hong (2017)은 LSTM 모형을 이용하여 코스피 200 지수에 대한 시계열 예측 모형을 개발하였고 RMSE (Root mean square error)를 통해 최적모형을 검증하였다. Kim et al. (2019)은 RNN 모형, LSTM모형 그리고 GRU 모형을 이용하여 태양광 발전량을 예측하는 모형을 설계하였고 LSTM 모형이 태양광 발전량을 예측하는 모델에 있어 가장 뛰어난 성능을 보임을 확인하였다. Lee et al. (2019b)는 MLP 모형, RNN 모형 그리고 LSTM 모형을 사용하여 가정에서 측정되는 전력 사용량을 예측 및 비교한 결과, LSTM 모형이 다른 모형과 비교하여 유의미한 예측 결과를 얻을 수 있음을 확인할 수 있었다.

Makridakis (1983)은 단일 예측 방법을 사용하는 대안으로 다수의 모델에서 얻은 예측값을 평균화하는 방법론을 제시하였다. 해당 방법은 다수의 모델을 앙상블하여 예측 정확도를 향상시킬 수 있는 방법론으로 제시되었다.

3. 변수의 설정

3.1 종속변수

철강은 국가 제조업 경쟁력의 핵심 요인으로 작용해왔으며 철강 소재의 원가 경쟁력과 품질은 국가 제조업 경쟁력의 발전과 매우 밀접하게 연관되어 있어 전 세계적으로 철강이 엄청난 규모로 생산 및 소비되고 있다 (Nam, 2018). 따라서 국가의 철강 생산량과 소비량의 글로벌 시장에서 차지하는 비중이 높을수록 국가 경쟁력에서 우위를 점할 가능성이 높다.

중국은 개혁개방 정책 이후로 철강시장에서 세계 1, 2위의 생산 및 소비 대국으로 등장하였다(Liu, 2012). 2014년 중국의 세계 철강 수출 비율은 37.2%로 1위를 차지하여 전 세계적으로 중국에서 생산된 철강을 많이 소비하는 것으로

나타나 중국 철강은 세계 각 국가의 산업과 경제에 지대한 영향을 미치고 있다 (Liu, 2012; Chu and Kong, 2015). 이에 본 연구에서는 중국 철강의 원재료인 중국 철광석의 가격을 미리 예측할 수 있다면 철강을 다루는 국가나 기업의 투자 위험성을 줄일 수 있을 것이라 판단하였고 중국 철광석 가격을 종속변수로 설정하였다.

3.2 설명변수

3.2.1 한·중·일 철강재 수출

2004년을 기점으로 중국은 철강재 순수출국으로 전환하여 철강 공급능력이 대폭 확대되었다 (Won and Yang, 2011). 중국 철강의 수출량은 앞서 언급한 것과 같이 세계 최대 규모를 가지고 있으며 자국에서의 철강 수출량은 철강의 원재료인 철광석 가격에 영향을 주는 중요한 요인이 될 수 있다(Chu and Kong, 2015; Nam, 2018).

세계 철강 수출에서 일본과 한국의 철강 수출 비중은 2014년을 기준으로 각각 9.1%, 7.1%로 중국에 뒤이어 2위, 3위를 차지하고 있어 세계 철강 시장에서 어느 정도 영향력을 미치고 있다 (Chu and Kong, 2015). 일본은 전반적으로 고도화된 기술로 만들어낸 수준이 높은 상품으로 경쟁력을 확보하고 있으며 중국은 이에 대비하여 빠른 경쟁력 강화를 추진 중에 있다 (Lee and Son, 2019). 한국은 H형강과 철근과 같은 철강에 대해 가파르다 중국 대비 경쟁력 우위를 유지한 것으로 나타난다 (Lee and Son, 2019). 이렇듯 한국과 일본의 철강 수출은 2000년 이후 중국의 수출입 증가만큼의 성장을 따라가기 어려우나 빠르게 성장해가고 있으며 철강시장에서 어느 정도 영향력을 행사하고 있다 (Liu and Han, 2012). 따라서 본 연구에서는 기존의 연구에 근거하여 한국의 철강재 수출량, 한국의 H형강 수출량, 철근 수출량 그리고 중국 및 일본의 철강재 수출량이 중국산 철광석 가격에 영향을 미칠 것이라 판단하였고 중국산 철광석 가격을 설명하는 변수로 설정하였다.

여기서 한국 및 일본의 수출량 변수는 중국으로 만의 수출량이 아닌 자국을 제외한 다른 국

가의 수출량을 의미한다.

3.2.2 철강 수급

20세기 중국은 전 세계의 철강 생산량에서 그 비중이 1위로 유지 중이며 세계 10대 철강 생산 기업 중 중국의 기업이 4곳이나 포함되어 있다 (Kim, 2010). 따라서 세계 최대 철강 생산국이자 최대 철광석 소비국인 중국의 철강 생산량은 철광석 수급에 가장 중요한 요인이다 (Nam, 2016). 또한 중국이 세계 철강 생산량에서 차지하는 비중은 2000년 15%에서 2017년 50%까지 증가하며 글로벌 철강 생산 대국으로 확고하게 자리매김하고 있다 (Nam, 2018). 따라서 철강을 생산하기 위해 철광석을 타국에서 수입하는 데 있어서도 큰 규모를 가지고 있다 (Kim, 2010). 따라서 본 연구는 기존의 연구에 근거하여 중국의 철강 생산량 그리고 중국의 철광석 수입량을 중국 철광석 가격을 설명하는 변수로 설정하였다.

3.2.3 원자재 경기

곡물, 비철금속 그리고 철강과 같은 세계의 주요 원자재 가격의 변동성은 나라의 경제성장, 국제수지, 물가, 고용과 같이 경제의 다양한 측면에서 적지 않은 영향을 준다 (Lee, 2006). 이러한 주요 원자재의 가격은 주요 국가들의 경제 성장에 따라 장기적으로 유사한 추세를 보이는 경향이 있다 (Nam, 2018). 본 연구는 기존의 연구에 근거하여 원자재 경기에 대해 대표성이 높은 구리 가격과 철광석의 대체품으로 활용되는 철스크랩의 가격을 중국 철광석 가격에 대한 설명변수로 설정하였다.

4. 본 론

4.1 데이터 수집

본 연구에서는 기존의 연구에 근거하여 중국 철광석 가격과 한·중·일 철강재 수출과 철강 수급 그리고 원자재 경기와 관련된 9개의 지표들

을 수집하였다. 데이터의 기간은 2010년 1월에서 2018년 3월까지의 월별 데이터 총 99개를 수집하였다. 총 10개의 지표들은 KOMIS (Korea mineral resource information service), KOSIS (Korean statistical information service), 중국 국가통계국, Trading Economics 그리고 세계철강협회에서 수집할 수 있었다.

4.2 중국 철광석 가격의 예측 및 평가 방법

본 연구는 머신러닝 모델을 이용하여 중국 철광석 가격이 결정짓게 되는 시점으로부터 한 달 전에 미리 철광석의 가격을 예측하는 머신러닝 모델을 구축하고자 한다. 머신러닝 모델을 훈련시키고 구축된 모형의 성능을 평가하기 위해서 원본 데이터를 훈련 데이터와 평가 데이터로 나누어 사용하였다. 훈련 데이터와 평가 데이터의 비율은 8 대 2의 비율로 설정하였다.

과거의 데이터를 머신러닝 모형이 학습을 하고 모형이 미래의 중국 철광석 가격을 예측하기 위해서는 연구에서 사용되는 설명변수들이 중국 철광석 가격에 시계열적으로 얼마만큼의 인과성을 가지는지 알아보고 사용 데이터를 인과관계 분석 결과에 따라 설명변수를 과거의 설명변수들로 재구성할 필요가 있다. 따라서 본 연구는 Granger (1969)가 제안한 인과모형을 사용하여 설명변수와 종속변수 사이의 시계열적 인과성을 알아보고 설명변수들을 유의한 그레인저 인과성이 있는 과거의 변수들로 재구성한다. 이렇게 재구성된 데이터를 사용하여 철광석이 거래되는 시점으로부터 한 달 전에 철광석 거래 가격을 미리 예측하는 모형을 구축한다. 그리고 머신러닝 모형이 학습하지 않은 평가 데이터를 사용하여 모형의 성능을 평가한다.

시계열 모형의 평가 지표는 MSE (Mean square error), RMSE (Root mean square error), MAE (Mean absolute error) 그리고 MAPE (Mean absolute percentage error)를 사용하였다. MSE는 오차의 제곱 값들의 평균을 나타낸다. RMSE는 MSE에서 제곱근 한 값이며, MAE는 오차의 절대값의 평균을 설명한다. 또한 MAPE는 실제값 대비 오차의 절대값들의

평균을 설명하고 있으며 오차의 정도를 %비율로 표현한다. 해당 4개의 측정지표들은 각 시점별 중국 철광석 가격의 실제값과 머신러닝 모형을 통해 미리 예측된 값에 대하여 오차를 측정하는데 적용된다. 4개의 측정지표들의 값이 낮은 수준으로 유지될수록 머신러닝 모형이 중국 철광석 가격을 정확하게 예측하였다고 볼 수 있다. 본 연구에서 사용한 측정지표들의 식은 (1), (2), (3) 그리고 (4)와 같다.

$$MSE = \frac{\sum_{t=1}^n (A_t - F_t)^2}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}} \quad (2)$$

$$MAE = \frac{\sum_{t=1}^n |A_t - F_t|}{n} \quad (3)$$

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|}{n} 100\% \quad (4)$$

여기서 n 은 데이터의 수, t 는 시간, A_t 는 실제값 그리고 F_t 는 예측값을 의미한다. Fig. 1은 본 연구의 머신러닝 모형 평가에 대한 흐름도를 설명해 준다.

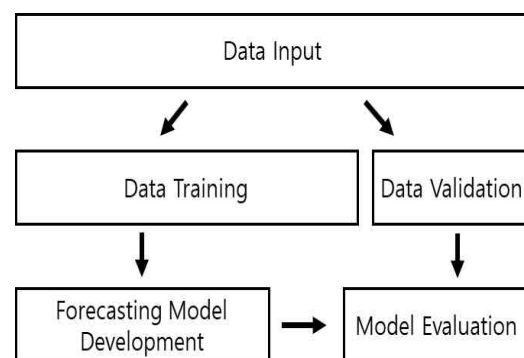


Fig. 1 Flowchart of Model Evaluation Process

4.3 그레인저 인과관계

본 연구에서는 Granger (1969)가 제안한 인과모형을 사용하여 설명변수와 종속변수 사이의 시계열적 인과성을 알아보려고 하였다. 그레인저 인과관계를 알아보기 위해서는 관측된 계열이 안정성을 가지는지 확인해야 한다. 안정되지 않은 시계열 자료를 분석하는 것은 분석 결과를 오도할 수 있기 때문이다 (Lee and Ryu, 2017). 본 연구에서는 시계열 자료의 안정성 확인을 위해 ADF (Augmented dickey-fuller) 단위근 검정을 사용하였다.

Table 1 Results of the ADF Test

Division	Variable	Stats	P-value
dependent variable	PIC(Price of the Iron Ore in China)	-7.751	0.000
Export volume of Steel products in China, Japan and South Korea	ESC(Export volume of Steel products in China)	-13.024	0.000
	ESJ(Export volume of Steel products in Japan)	-18.249	0.000
	ESK(Export volume of Steel products in South Korea)	-13.629	0.000
	EHK(Export volume of H-Steel products in South Korea)	-17.577	0.000
	ERK(Export volume of Rebar products in South Korea)	-15.253	0.000
Steel supply and demand	IIC(Import volume of Iron Ore in China)	-17.749	0.000
	PCSC(Production volume of Crude steel in China)	-14.736	0.000
raw material economy	PC(Price of the Copper)	-8.7399	0.000
	PIS(Price of the Iron Scrap)	-15.475	0.000

원계열의 자료들을 사용하여 1차적으로 단위근 검정을 시행한 결과, 원계열의 자료들은 모두 비안정성을 띠는 시계열임을 확인할 수 있었다. 따라서 원계열의 자료들을 1차 차분한 계열의 자료들을 사용하여 ADF 검정을 수행하였다. 그 결과, 본 연구에서 설정한 종속변수인 중국산 철광석 가격과 설명변수로 설정한 한·중·일 철강재 수출과 관련된 5개의 변수, 철강 수

급과 관련된 2개의 변수, 원자재 경기와 관련된 2개의 변수로 총 10개의 변수들이 유의수준 5% 하에 모두 안정성을 가지는 시계열로 나타났다. Table 1은 연구에서 사용되는 변수들의 ADF 검정 결과를 종합적으로 보여준다. 본 연구에서는 Table 1의 결과를 반영하여 모든 변수를 1차 차분한 계열의 자료들을 사용하여 분석을 진행하였다.

그레인저 인과관계의 검정단계에서 어떤 설명변수가 중국 철광석 가격에 시계열적으로 인과적인 영향을 주는지는 식 (5)와 (6)을 이용한다 (Lee and Ryu, 2017).

$$gY(i,t) = \sum_{j=1}^k \alpha(j)gY(i,t-j) \tag{5}$$

$$+ \sum_{j=1}^l \alpha(j)gX(i,t-j) + \epsilon(i,t)$$

$$gY(i,t) = \sum_{j=1}^k \alpha(j)gY(i,t-j) + \epsilon(i,t) \tag{6}$$

여기서 gY 는 종속변수, gX 는 설명변수, i 는 i 번째 설명변수, t 는 시간(월), k 와 l 은 시차, 그리고 $\epsilon(i,t)$ 는 오차항을 뜻한다. 식 (6)은 과거의 설명변수들에 의하여 현재의 중국 철광석 가격이 결정됨을 보인다. 이는 현재까지의 과거 자료를 이용하여 미래의 값을 예측할 때 사용되는 수식이다 (Lee and Ryu, 2017). 만약 식 (5)와 (6)을 비교하여 식 (5)가 (6)보다 통계적으로 우수하다면, i 번째 설명변수는 종속변수에 그레인저 인과적인 영향을 준다고 할 수 있다 (Granger, 1969). 두 식에 대한 통계적인 유의성을 확인하는 것은 식 (7)과 같은 검정통계량을 이용한다 (Granger, 1969).

$$F = \frac{(RSS_r - RSS_u/w)}{RSS_u/(T-s)} \tag{7}$$

해당 검정통계량은 분자 w 와 분모 $T-s$ 의 자유도를 가지는 F -분포를 따른다. w 는 식 (5)에서 사용된 회귀변수의 수에서 (6)에서 사용한 회귀변수의 수의 차를 나타내며, T 는 관측치의 수, s 는 식 (5)에서 사용된 회귀변수의 수를 나타낸다. 본 연구에서의 그레인저 인과관계의 시차 (Lag)는 한 달에서 다섯 달까지 설정하였다. 해당 시차를 고려하는 이유는 설명변수들의 어느

시점까지의 과거 변수가 종속변수의 현재 시점에 유의한 그레인저 인과관계가 있는지 알아보기 위함이며 해당 결과를 통해 머신러닝 모형을 학습시키는 설명변수를 과거의 설명변수들로 재구성하게 된다 (Granger, 1969). 여기서 그레인저 인과관계를 살펴보는 시차는 사용자의 판단에 따라 무한정 늘릴 수 있지만, 많은 시차를 사용하는 경우 다중공선성의 문제가 발생할 수 있어 적당한 시차를 고려해야 한다 (Silvey, 1969).

본 연구는 그레인저 인과관계를 이용하여 모형 개발에 적합한 속성을 선택하게 되므로 많은 시차를 사용하는 경우 과적합의 문제가 발생할 수 있어 연구에서 사용하는 시차는 다섯 시차까지 사용하였다. 주어진 시차 내에서 최적의 시차를 선택하는 기준은 AIC (Akaike information criterion)가 최소화되는 방향으로 결정하였으며 설명변수가 종속변수에 대해 유의한 그레인저 인과관계가 있는 시차를 선택하였다. Table 2는 그레인저 인과관계 검정을 시행한 결과를 보여준다. Table 2에서의 값들은 그레인저 인과관계 검정에 사용되는 F검정통계량과 유의확률이며 만약, 유의확률이 유의수준보다 작다면 설명변수가 종속변수의 l 시차만큼의 그레인저 인과를 가지지 않을 것이라는 귀무가설을 기각한다. 본 연구에서는 유의수준 10% 하에 각 설명변수마다 중국 철광석 가격과 그레인저 인과관계가 있는 설명변수들의 최적 시차를 탐색하였다. 설명변수들이 중국 철광석 가격에 그레인저 인과관

계가 있는지 탐색해본 결과, 한·중·일 철강재 수출과 관련된 변수에서 중국 철강재 수출량, 일본의 수출량, 한국 H형강 수출량, 한국 철근 수출량은 중국 철광석 가격에 유의한 그레인저 인과관계가 있는 것으로 나타났다. 이에 반해 한국 철강재 수출량은 중국 철광석 가격에 유의한 그레인저 인과관계가 없는 것으로 나타났다. 이러한 결과는 한국의 H형강과 철근과 같은 특정 철강금속은 중국 대비 경쟁력의 우위를 점하고 있어 중국 철광석 가격에 영향을 미칠 수 있다는 점에서 기존의 연구와 일치한다 (Lee and Son, 2019). 하지만 한국의 전체적인 철강재 수출량으로 본다면 경쟁력이 떨어지며 중국 철광석 가격에 미치는 영향력은 미미한 것으로 판단된다. 중국 철강 생산량과 중국 철광석 수입량은 중국 철광석 가격에 유의한 그레인저 인과관계가 있는 것으로 나타났다. 이러한 결과는 기존의 연구와 일치하는 결과를 보이고 있다.

원자재 경기 변수에서는 구리 가격이 중국 철광석 가격에 유의한 그레인저 인과관계가 있는 것으로 나타나 기존의 연구결과와 일치하는 것으로 보였으며, 철스크랩 가격은 중국 철광석 가격에 유의한 그레인저 인과관계가 없는 것으로 나타났다. 이러한 이유를 살펴보았을 때, 먼저 철스크랩은 철강재의 생산 및 가공 과정이나 철강제품의 사용 불능 상태에서 발생한 것을 재수집한 것으로 철광석의 소비 이후에 물량이 확보된다. 따라서 철스크랩 가격이 철광석의 가격에 인과적

Table 2 Results of the Granger Causality Test

Division	Variables	F-test statistics and significance probability by lag(l)					Optimal lag(l)	Granger causality
		1	2	3	4	5		
Export volume of Steel products in China, Japan and South Korea	ESC	0.727(0.369)	1.650(0.198)	2.179(0.096)	1.499(0.210)	1.244(0.296)	3	Y
	ESJ	2.512(0.116)	4.451(0.014)	1.799(0.153)	5.317(0.001)	3.920(0.003)	4	Y
	ESK	1.380(0.243)	0.679(0.509)	0.636(0.593)	0.642(0.634)	1.095(0.369)		
	EHK	6.893(0.010)	3.411(0.037)	2.957(0.037)	2.862(0.028)	3.154(0.012)	1	Y
	ERK	5.426(0.022)	4.680(0.012)	4.077(0.009)	4.291(0.003)	3.300(0.009)	4	Y
Steel supply and demand	IIC	0.234(0.630)	2.425(0.094)	1.272(0.289)	0.844(0.501)	1.702(0.143)	2	Y
	PCSC	0.049(0.825)	1.882(0.158)	2.256(0.087)	1.312(0.272)	0.945(0.457)	3	Y
raw material economy	PC	8.173(0.005)	5.017(0.008)	3.223(0.026)	2.748(0.033)	2.280(0.054)	1	Y
	PIS	0.187(0.666)	0.330(0.720)	0.280(0.847)	1.579(0.187)	1.127(0.353)		

인 영향을 미치기보다는 오히려 철광석 가격이 철스크랩 가격에 인과적인 영향을 미칠 수 있다. 기존의 연구에 따르면 철스크랩 가격의 하락에 대한 원인으로 낮은 철광석의 가격이 영향을 미친다는 연구결과가 있다 (Jang, 2018).

다음과 같은 이유로 본 연구에서는 한국 철강재 수출량 변수와 철스크랩 가격 변수는 중국 철광석 가격의 설명변수에서 제외하고 분석을 진행하였다. 또한 유의한 그레인저 인과성을 가지는 설명변수들은 각 변수별 최적의 시차에 맞게 과거의 변수들로 재구성하여 총 18개의 설명변수를 사용하였다.

4.4 머신러닝 모형

4.4.1 시차분포 모형

시차분포 모형은 종속변수의 현재 수준의 값을 설명변수들의 현재 수준 값 그리고 과거 시차 값의 선형함수로 설정된다 (Pearson, 2001). 시차분포 모형의 일반적인 형태는 식 (8)과 같다.

$$Y_t = u + \sum_{j=1}^k \sum_{i=0}^{q_j} \beta_{j,i} X_{j,t-i} + \epsilon_t \quad (8)$$

여기서 u 는 상수항, k 는 k 번째 설명변수, q_j 는 j 번째 설명변수의 시차, $\beta_{j,i}$ 는 X_j 의 i 번째 시차 변수에 대한 추정계수, ϵ_t 는 오차항을 나타낸다. 본 연구의 목적은 철광석 가격이 결정되기 한 달 전에 미리 예측하는 모형을 구축하는 것이 목적이므로 식 (8)에서 설명변수의 현재 수준 값은 제외하여 식 (9)과 같은 형태의 시차분포 모형을 사용하였다.

$$Y_t = u + \sum_{j=1}^k \sum_{i=1}^{q_j} \beta_{j,i} X_{j,t-i} + \epsilon_t \quad (9)$$

시차분포 모형에서는 설명변수들의 시차 q_j 를 결정해야 하는데 본 연구에서는 그레인저 인과성을 가지는 최적의 시차를 사용하였다.

4.4.2 MLP

MLP는 인공신경망의 한 종류로, 하나의 입력층 (Input layer), 하나 이상의 은닉층 (Hidden layer) 그리고 하나의 출력층 (Output layer)으로 구성된다. MLP의 입력 신호는 층과 층을 지나는 순방향 (Forward direction)으로 전파하여 출력값을 얻게 된다 (Kang and Kim, 1996). 입력층에는 학습에 들어갈 입력 변수의 수만큼의 노드가 들어가며, 은닉층과 출력층에 있는 각 노드는 활성화함수를 통해 이전 층에서 입력값을 받아 가중치를 곱하여 출력값을 계산하게 된다. Fig. 2는 본 연구에서 사용하는 다층 퍼셉트론 모형 중 두 개의 은닉층을 가진 퍼셉트론의 구조를 보여준다.

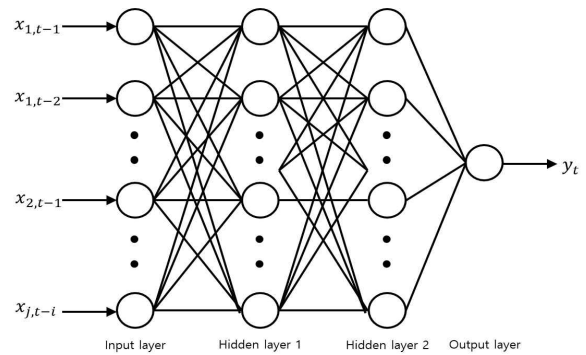


Fig. 2 Structure of MLP

본 연구에서는 입력층에 들어갈 입력 변수의 수를 Table 2의 결과에서 나타난 유의한 그레인저 인과관계가 있는 최적의 시차만큼의 변수의 개수인 18로 사용하였다. Table 3은 MLP에서 설정한 세부 옵션을 보여준다. 이러한 세부 옵션은 MLP뿐만 아니라 RNN, LSTM에도 동일하게 적용된다.

Table 3 Options of MLP, RNN, and LSTM

Options	Setting
Dropout	0.3
batch size	1
epochs	100
learning rate	0.01
activation function	Hyperbolic Tangent
loss function	MSE
optimizer	adam

4.4.3 RNN

RNN은 은닉층으로 순환하는 층을 이용하여 이전 상태의 은닉층 출력이 현재 상태의 은닉층 입력으로 들어가 피드백 작용을 하는 신경망 모델을 의미한다 (Lee et al., 2019b; Jung et al., 2017). RNN은 주로 시계열 데이터를 모델링 하는데 주로 사용되며, 앞서 언급한 MLP와 같이 입력층, 은닉층, 출력층을 가지고 있다. 다만 본 연구에서 사용한 RNN 모델의 구조인 Fig. 3과 같이 시간에 흐름에 따라 데이터를 처리한다는 점에서 MLP와 차이가 있다.

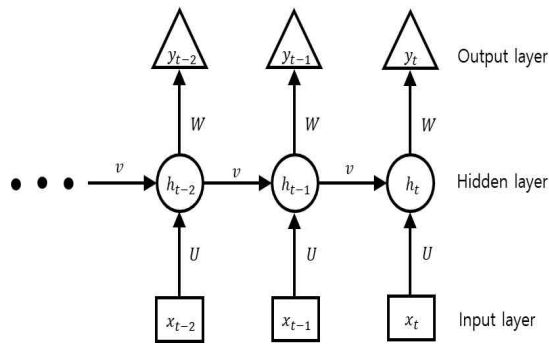


Fig. 3 Structure of RNN

RNN의 출력을 수식으로 표현하면 식 (10)과 같다 (Lee et al., 2019b). 수식에서 y_t 는 현재 입력에 대한 출력을 나타낸다. f 는 활성화함수, U 는 입력에 대한 연결계수, x_t 는 현재 입력을 말하며, v 는 이전 은닉층에 대한 연결계수, h_{t-1} 은 이전 은닉층의 출력 그리고 b_k 는 바이어스(Bias) 값을 의미한다. 여기서 입력에서는 유의한 그래인저 인과관계가 있는 최적의 시차만큼의 설명변수들이 입력으로 들어가게 된다.

$$y_t = f([Ux_t, v h_{t-1}] + b_k) \quad (10)$$

4.4.4 LSTM

LSTM의 구조는 RNN과 유사하며 RNN에서 망각 게이트, 입력 게이트 그리고 출력 게이트를 통해 장기 의존성 문제를 해결시킨 모델이다 (Shin et al., 2018). Fig. 4는 LSTM의 전체적인

구조를 보여주며 Fig. 5는 LSTM cell의 세부구조를 보여준다.

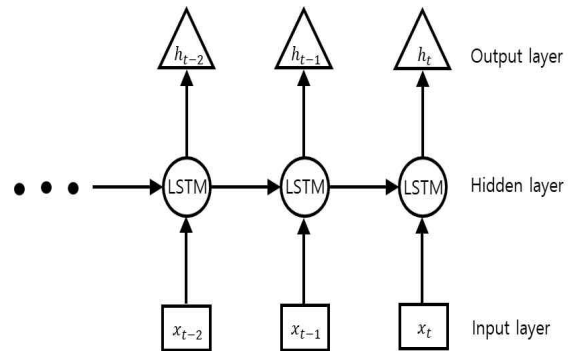


Fig. 4 Structure of LSTM

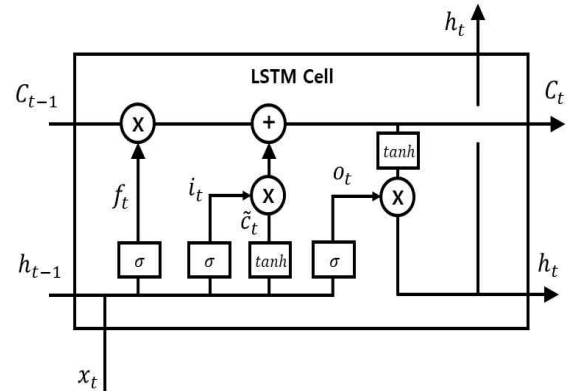


Fig. 5 Structure of LSTM cell

망각 게이트에서는 유지할 정보를 결정하게 된다. 식 (11)은 망각 게이트에서 이전 셀 상태와 현재 셀 입력 값의 조합을 보고 어떤 정보를 버릴지 결정하는 수식이다. 여기서 σ 는 활성화 함수, W 는 가중치, b 는 바이어스 값을 나타낸다.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (11)$$

입력 게이트에서는 새로운 정보가 계속해서 저장될지를 결정한다. 식 (12)와 (13)은 입력 게이트에서 새로운 입력 값을 만드는 수식이다. 여기서 C_{t-1} 을 새로운 셀 상태인 C_t 로 수정하는 식은 (14)과 같다 (Lee et al. 2019b).

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (12)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (13)$$

$$C_t = (f_t \times C_{t-1}) + (i_t \times \tilde{C}_t) \quad (14)$$

마지막으로 출력 게이트에서는 망각 게이트와 입력 게이트를 통해 업데이트된 정보를 반영하여 출력값을 결정하게 된다. 식 (15)과 (16)은 해당 과정을 수식으로 설명해 준다.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (15)$$

$$h_t = o_t \times \tanh(C_t) \quad (16)$$

4.5 실험 결과

본 연구는 4개의 머신러닝 모델을 이용하여 4개의 측정지표를 기준으로 평가 데이터에서 중국 철광석 가격을 얼마나 정확하게 예측하는지 실험하였다. MLP, RNN 그리고 LSTM의 경우는 노드의 수와 은닉층의 수에 따라 최적의 파라미터를 탐색해 보았다. 또한 중국 철광석 가격에 대한 예측 정확도를 높이기 위해 각 모형에서의 최적의 파라미터를 기준으로 4개의 모형을 앙상블 하는 연구를 시도하였다. 본 연구의

앙상블 모형은 각 모형으로부터 얻어진 출력값을 평균하여 도출하는 식으로 생성된다. 해당 방법론은 다수의 모델에서 얻은 예측값을 평균하여 도출함으로써 기존에 하나의 모델에서 얻어지는 예측 정확도보다 예측 정확도를 더욱 향상시키는 효과가 있다 (Makridakis, 1983).

Table 4, Table 5, Table 6은 MLP, RNN 그리고 LSTM의 노드의 수와 은닉층의 수에 따른 4개의 측정지표가 나타난 결과를 보여준다. 그리고 Table 7은 본 연구에서 시도한 앙상블 모형들의 예측 오차의 측정 결과를 보여준다.

4개의 개별 모형을 측정된 결과, 시차분포 모형은 MSE의 값이 42.087, RMSE는 6.487, MAE의 값이 4.803 그리고 MAPE의 값이 6.928%로 나타났다. MLP는 노드의 수가 16 그리고 은닉층의 수가 4일 때 MSE의 값이 50.581, RMSE는 7.112, MAE는 5.582 그리고 MAPE의 값이 7.778%로 나타났다. RNN은 노드의 수가 16, 은닉층의 수가 4일 때 MSE의 값이 43.485, RMSE는 6.594, MAE는 5.347 그리고 MAPE가 7.697%로 나타났다. LSTM은 노드

Table 4 Results of Measurement of the MLP at Different Number of Hidden Layers and Neurons

Number of hidden layer	Indicators of Measurement	Number of neurons at each layer							
		8	16	32	64	128	256	512	1024
1	MSE	54.071	55.846	56.439	53.874	61.399	58.661	66.86	71.771
	RMSE	7.353	7.473	7.513	7.340	7.836	7.659	8.177	8.472
	MAE	5.789	6.065	5.909	5.948	6.355	6.538	6.132	6.376
	MAPE	8.154	8.656	8.393	8.423	8.984	9.405	8.701	9.308
2	MSE	60.749	61.807	51.683	72.352	73.917	58.909	53.764	63.925
	RMSE	7.794	7.862	7.189	8.506	8.598	7.675	7.332	7.995
	MAE	6.348	6.321	5.676	7.066	6.746	5.995	5.855	6.420
	MAPE	9.072	8.952	8.117	10.138	9.758	8.700	8.432	9.162
3	MSE	59.999	59.257	57.436	63.392	60.769	65.709	60.488	54.806
	RMSE	7.746	7.698	7.579	7.962	7.795	8.106	7.777	7.403
	MAE	6.281	5.986	5.842	6.016	6.012	6.809	5.995	5.771
	MAPE	8.892	8.415	8.126	8.467	8.647	9.946	8.691	8.419
4	MSE	62.300	50.581	59.476	56.919	59.467	65.125	72.643	65.027
	RMSE	7.893	7.112	7.712	7.544	7.711	8.070	8.523	8.064
	MAE	6.329	5.582	6.111	6.144	6.434	6.500	6.690	6.853
	MAPE	8.938	7.778	8.484	8.475	9.301	9.273	9.580	9.990

Table 5 Results of Measurement of the RNN at Different Number of Hidden Layers and Neurons

Number of hidden layer	Indicators of Measurement	Number of neurons at each layer							
		8	16	32	64	128	256	512	1024
1	MSE	59.251	59.767	60.296	61.613	64.616	63.312	67.401	67.433
	RMSE	7.697	7.731	7.765	7.849	8.038	7.957	8.210	8.212
	MAE	6.276	6.303	6.360	6.455	6.561	6.516	6.706	6.730
	MAPE	9.239	9.265	9.325	9.412	9.601	9.532	9.893	10.011
2	MSE	59.484	57.830	72.011	71.963	50.966	74.044	63.099	75.594
	RMSE	7.713	7.605	8.486	8.483	7.139	8.605	7.944	8.694
	MAE	6.314	6.145	6.500	6.213	6.077	7.239	6.494	7.081
	MAPE	8.968	8.732	9.100	8.573	8.521	10.259	9.070	10.142
3	MSE	62.173	74.270	81.243	66.503	78.576	64.058	78.370	77.355
	RMSE	7.885	8.618	9.013	8.155	8.864	8.004	8.853	8.795
	MAE	6.451	7.425	6.833	6.714	7.707	6.467	7.009	7.410
	MAPE	9.141	10.413	9.721	9.370	10.817	9.268	9.816	10.578
4	MSE	69.959	43.485	86.272	74.668	72.195	61.991	68.363	71.881
	RMSE	8.364	6.594	9.288	8.641	8.497	7.873	8.268	8.478
	MAE	6.430	5.347	7.610	6.512	6.413	6.440	6.898	6.748
	MAPE	9.032	7.697	10.846	9.292	8.974	9.071	9.899	9.555

Table 6 Results of Measurement of the LSTM at Different Number of Hidden Layers and Neurons

Number of hidden layer	Indicators of Measurement	Number of neurons at each layer							
		8	16	32	64	128	256	512	1024
1	MSE	61.867	59.985	62.184	66.083	64.516	66.907	65.898	68.066
	RMSE	7.866	7.745	7.886	8.129	8.032	8.180	8.118	8.250
	MAE	6.434	6.305	6.412	6.676	6.588	6.662	6.596	6.736
	MAPE	9.408	9.342	9.449	9.971	9.788	9.814	9.700	10.138
2	MSE	61.465	27.196	48.327	66.453	62.788	56.384	63.352	71.916
	RMSE	7.840	5.215	6.952	8.152	7.924	7.509	7.959	8.480
	MAE	6.175	4.227	5.813	6.842	6.480	6.198	6.080	6.651
	MAPE	8.694	6.242	8.164	9.665	9.360	8.811	8.792	9.695
3	MSE	64.215	60.489	67.142	73.770	60.647	68.985	66.115	71.272
	RMSE	8.013	7.777	8.194	8.589	7.788	8.306	8.131	8.442
	MAE	6.602	6.046	6.762	7.106	6.480	6.311	6.647	6.206
	MAPE	9.229	8.408	9.456	9.987	9.255	9.145	9.699	8.891
4	MSE	56.349	65.712	52.656	64.771	69.350	63.689	86.833	77.734
	RMSE	7.507	8.106	7.256	8.048	8.328	7.981	9.318	8.817
	MAE	6.044	6.105	5.826	6.509	6.967	6.205	7.310	6.716
	MAPE	8.516	8.497	8.085	9.442	9.792	9.040	10.560	9.757

의 수가 16, 은닉층의 수가 2일 때 MSE의 값이 27.196, RMSE는 5.215, MAE는 4.227 그리고 MAPE의 값이 6.242%로 나타나 4개의 개별적

Table 7 Results of Measurement of the Ensemble

Model	Indicators of Measurement	Values
DL+MLP	MSE	39.504
	RMSE	6.285
	MAE	5.057
	MAPE	7.169
DL+RNN	MSE	38.134
	RMSE	6.175
	MAE	5.005
	MAPE	7.214
DL+LSTM	MSE	22.546
	RMSE	4.748
	MAE	4.013
	MAPE	5.870
MLP+RNN	MSE	39.778
	RMSE	6.307
	MAE	5.398
	MAPE	7.628
MLP+LSTM	MSE	28.489
	RMSE	5.338
	MAE	4.659
	MAPE	6.622
RNN+LSTM	MSE	27.879
	RMSE	5.280
	MAE	4.461
	MAPE	6.494
DL+MLP+RNN	MSE	37.057
	RMSE	6.087
	MAE	5.132
	MAPE	7.310
DL+MLP+LSTM	MSE	26.922
	RMSE	5.189
	MAE	4.458
	MAPE	6.382
DL+RNN+LSTM	MSE	26.830
	RMSE	5.180
	MAE	4.418
	MAPE	6.423
MLP+RNN+LSTM	MSE	29.258
	RMSE	5.4009
	MAE	4.757
	MAPE	6.798
DL+MLP+RNN+LSTM	MSE	28.664
	RMSE	5.354
	MAE	4.660
	MAPE	6.683

인 머신러닝 모형 중에서는 예측 오차가 가장 낮은 것으로 나타났다.

앙상블 모형에서는 시차분포와 LSTM의 앙상블 모형이 MSE의 값이 22.546, RMSE는 4.748, MAE는 4.013, MAPE는 5.870%로 나타나 연구에서 시도된 앙상블 모형 중 가장 낮은 예측 오차를 보이는 것으로 나타났다.

4개의 개별 모형과 11개의 앙상블 모형들의 측정지표들을 종합적으로 비교해보았을 때 시차분포와 LSTM의 앙상블 모형이 본 연구에서 시도한 15개의 모형 중 가장 높은 성능을 보였으며 개별 모형만을 사용한 경우보다 앙상블 모형을 사용하였을 때 더 좋은 예측 성능을 가짐을 검증할 수 있었다. Fig. 6, Fig. 7, Fig. 8, Fig. 9, 그리고 Fig. 10은 본 연구에서 사용한 4개의 개별 머신러닝 모형과 앙상블 모형 중 가장 좋은 성능을 가진 시차분포와 LSTM의 앙상블 모형을 통해 평가 데이터를 예측 결과를 보여준다.

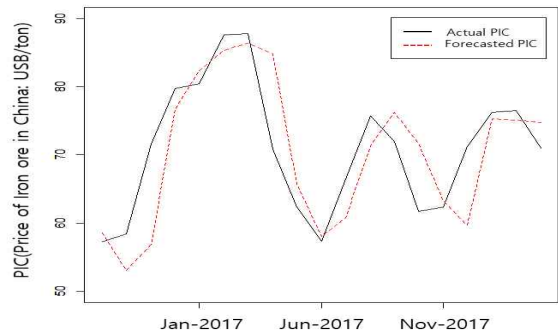


Fig. 6 Results of Forecasting PIC using DL Model

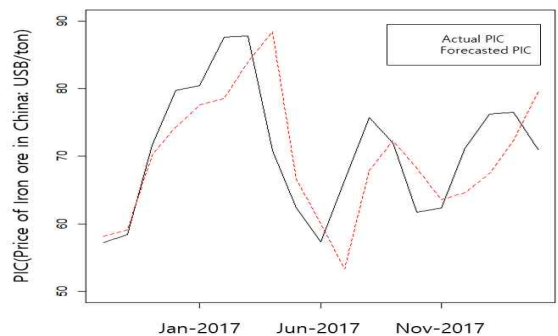


Fig. 7 Results of Forecasting PIC using MLP Model

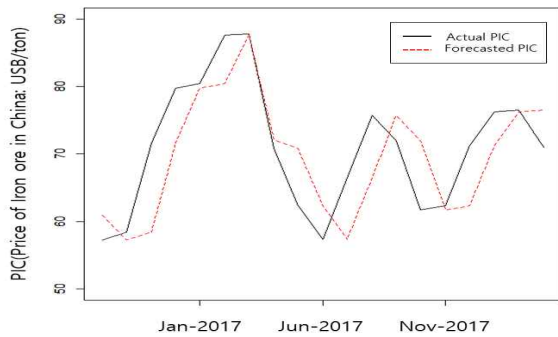


Fig. 8 Results of Forecasting PIC using RNN Model

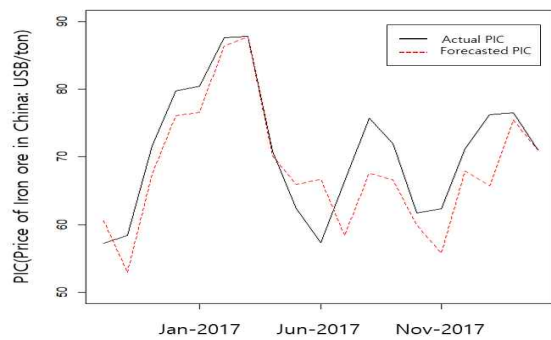


Fig. 9 Results of Forecasting PIC using LSTM Model

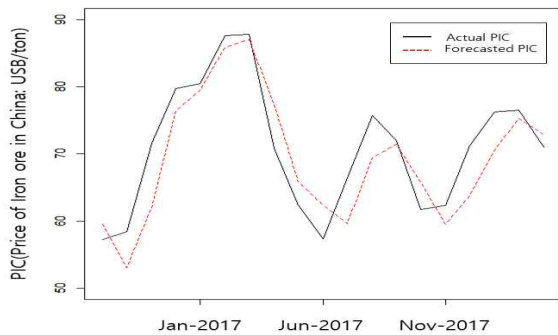


Fig. 10 Results of Forecasting PIC using DL+LSTM Model

5. 결론

본 연구는 중국 철광석이 거래되는 시점으로부터 한 달 전에 가격을 미리 예측하는 머신러닝 모델을 구축하였고, 학습하지 않은 평가 데이터를 예측하여 각 모델에서 4개의 측정지표를

통하여 실제값과 예측값에 대한 오차를 비교하였다. 본 연구의 분석 결과를 통하여 다음과 같은 시사점을 제시하고자 한다.

첫째, 그레인저 인과관계 분석을 통해 중국 철광석 가격에 시계열적 인과관계를 가지는 변수를 탐색할 수 있었다. 해당 분석을 통해 일본과 같은 이웃 국가의 철강재 수출량 변화에 따라 중국 철광석 가격에 변동성이 생긴다는 사실을 알 수 있었다. 또한 일본과 달리 한국은 H형강과 철근과 같이 특정한 제품만이 중국 철광석 가격에 영향을 미치는 것으로 나타났다. 중국의 철강 수급과 관련된 변수들은 중국 철광석 가격에 시계열적 인과관계를 가진다는 사실을 알 수 있었고, 원자재 가격의 경우에는 구리 가격이 중국 철광석 가격에 시계열적 인과관계를 가지지만, 철스크랩과 같은 경우에는 시계열적 인과관계를 가지지 않는 것으로 나타났다.

둘째, 본 연구는 중국 철광석이 거래되는 시점으로부터 한 달 전에 가격을 미리 예측하는 모델을 개발하였고 모델을 통해 나타난 예측값과 실제값에 대한 오차가 낮은 수준인 것으로 나타났다. 시차분포 모형과 LSTM 모형의 앙상블 모형은 MSE의 값이 22.546, RMSE는 4.748으로 나타나서 중국 철광석의 실제 가격과 예측 가격 간의 오차 제곱에 대한 측정지표 수준이 낮은 것으로 확인되었다. 또한 MAE의 값은 4.013로 나타났는데 이러한 결과는 머신러닝 모형이 거래 시점으로부터 한 달 전에 현재 시점의 중국 철광석 가격을 미리 예측하였을 때 실제값과 예측값이 평균적으로 4.013USD/ton의 오차가 나타남을 시사한다. 또한 모형들 중에서 시차분포 모형과 LSTM 모형의 앙상블 모형의 MAPE는 5.870%로 나타났는데 이는 평가 데이터에서 머신러닝 모형이 거래 시점으로부터 한 달 전에 중국 철광석 가격을 미리 예측하였을 때 실제값 대비 예측 오차의 비율이 5.860%라는 것을 알 수 있다. 이러한 결과를 종합적으로 보았을 때, 시차분포 모형과 LSTM 모형의 앙상블 모형은 거래 시점으로부터 한 달 전에 중국산 철광석 가격을 미리 예측하였을 때 실제값과 예측값에 대한 오차의 정도가 굉장히 낮은 수준임을 알 수 있다.

본 연구의 결과는 중국 철광석 가격과 관련된 변수들에 대한 과거 데이터를 이용하여 머신러닝 모델을 통해 현재 시점의 중국 철광석 가격을 미리 예측할 수 있음을 시사한다. 이러한 시사점은 철강을 취급하는 국가나 기업의 중국 철광석에 대한 투자 위험성을 줄일 수 있을 것으로 기대된다. 또한 본 연구에서 시도한 머신러닝 모델들뿐만 아니라 더 다양하고 성능이 뛰어난 모델들을 사용한다면 중국 철광석 가격을 예측하는 것에 있어서 보다 정확한 예측 결과를 얻을 수 있을 것이라 기대된다.

References

- Chung, E. C., and Lee, C. M. (2018). Determinants of Chonseil Deposit to Housing Sales Price Ratio Based on Rental Housing Market Equilibrium Condition: Focused on Housing Price Volatility, *Journal of the Korea Real Estate Analysts Association*, 24(4) 5-20.
- Choi, H., Moon T., Jung, D. H., and Son, J. E. (2019). Prediction of Air Temperature and Relative Humidity in Greenhouse via a Multilayer Perceptron Using Environmental Factors, *Protected Horticulture and Plant Factory*, 28(2), 95-103.
- Chu, J. M., and Kong, M. K. (2015). Changes and Implications of Steel Export Structure in Korea, China and Japan, *POSRI Issue Report*, 2015(9), 1-10.
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica*, 37(3), 424-438.
- Huh, J. S. (2018). Mid-term Global Iron Ore Market and Purchasing Issue Analysis and Implications, *POSRI Issue Report*, 2018(8), 1-10.
- Hong, H. K. (2009). A Study on the Apartment Price Model using Neural Network Model, *Proceedings of the Spring Conference on Korea Society of Industrial Information Systems Research*, 220-226, Korea.
- Jung, H. J., Yoon, J. S., and Bae S. H. (2017). Traffic Congestion Estimation by Adopting Recurrent Neural Network, *Journal of Korea Intelligent Transport Systems*, 16(6), 67-78.
- Jang, W., (2018). Factors Determining China's Iron Scrap Export Volume, *POSRI Issue Report*, 2018(12), 1-9
- Kim, K. M., and Kim, D. H. (2018). A Causality Test on Hairtail Prices among Import and Domestic Markets Using a Vector Error Correction Model (VECM), *Journal of the Ocean and Polar Research*, 40(1), 49-58.
- Kim, I. C., and Kim D. Y. (2001). A Study on the Features for Building Korea Digit Recognition System Based on Multilayer Perceptron, *Journal of the Korea Industrial Information Systems Research*, 6(4), 81-88.
- Kim, S. S., and Hong, K. J. (2017). Development and Performance Analysis of Predictive Model for KOSPI 200 Index using Recurrent Neural Networks, *Journal of the Korea Industrial Information Systems Research*, 22(6), 23-29.
- Kim, H., Tak, H., and Cho, H. G. (2019). Design of Photovoltaic Power Generation Prediction Model with Recurrent Neural Network, *Journal of the Korean Institute of Information Scientists and Engineers*, 46(6), 506-514.
- Kim, M. S. (2010). *The Research on Liquidity and Issues of Steel Market in the Construction Industry*, M. E. Thesis, Graduate School of Seoul Industry University, Seoul, Korea.
- Kang, K., and Kim, C. (1996). ATM Connection Admission Control Scheme Using

- Multilayer Perceptron, *Journal of KISS(A): Computer Systems and Theory*, 23(12), 1225-1237.
- Lee, W. C., Kim, J. M., and Lee, C. K. (2019a). A Study on the Factors Affecting the Price of Iron Ore, *Proceedings of the Spring Conference on The Korea Society Management Information Systems*, 102-106, Korea.
- Lee, D. G., Sun, Y. G., Kim, S. H., Sim, I., Hwang, Y. M., and Kim, J. Y. (2019b). Comparison of Power Consumption Prediction Scheme Based on Artificial Intelligence, *Journal of The Institute of Internet, Broadcasting and Communication*, 19(4), 161-167.
- Liu, J. (2012). *A Comparative Study on the Competitiveness of Chinese Steel Industry against Korea and Japan*, Ph. D. Thesis, Graduate School of Dong-Eui University, Busan, Korea.
- Lee, B., and Son, W. (2019). A Comparative Analysis of Structural Changes of Korea, China, and Japan in Product Space, *Policy References in Korea Institute for International Economic Policy*, 19(4), 1-99.
- Liu, J., and Han, K. (2012). A Study on the Trade Structure of Steel Industry between China and Japan, *Journal of the Korean-Japanese Economic & Management Association*, 56(0), 111-127.
- Lee, I. N. (2006). The Impact of Raw Material Price Increase on Domestic Prices: An Input-Output Analysis of Price Pass-through, *Korea Association Of Public Policy*, 21, 61-89.
- Lee, S., and Ryu, Y. U. (2017). Industrial Economic Growth and IT Investment: Is Economic Growth an Effect of IT Investment, or a Determinant of Decision-making for IT Investment, *Journal of The Information Systems Review*, 19(1), 185-202.
- Lee, S. M., Sun, Y. G., Lee, J., Lee D., Cho, E. I., Park D. H., Kim, Y. B., Sim, I., and Kim, J. Y. (2019). Short-term Power Consumption Forecasting Based on IoT Power Meter with LSTM and GRU Deep Learning, *Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, 19(5), 79-85.
- Malanichev, A. G., and Vorobyev P. V. (2011). Forecast of Global Steel Prices, *Studies on Russian Economic Development*, 22(3), 304-311.
- Moon, J., Park, S., and Hwang, E. (2019). A Multilayer Perceptron-Based Electric Load Forecasting Scheme via Effective Recovering Missing Data, *Journal of the Korea Information Processing Society*, 8(2), 67-78.
- Makridakis, S., and Winkler R. L. (1983). Averages of Forecasts: Some Empirical Results, *Journal of the Management Science*, 29(9), 987-996.
- Nam, D. Y. (2016). A Study on Early Warning System of Iron Ore Price based on Signal Approach Model, *Research in Chinese Studies*, 76, 131-149.
- Nam, D. Y. (2018). A Study on Early Warning System of China Iron Ore Price Based on AI Models, *China Academic Conference*, 61, 215-232.
- Pustov, A., Malanichev, A., and Khobotilov I. (2013). Long-term Iron Ore Price Modeling: Marginal Costs vs. Incentive Price, *Journal of the Resources Policy*, 38(4), 558-567
- Pearson, M. H., Shin, Y., and Smith, R. J. (2001). Bounds Testing Approaches to the Analysis of Level Relationship, *Journal of Applied Econometrics*, 16, 289-326.
- Rhee, B. K. (2017). Effect of Economic Sanctions on Trade: Distributed Lag Analysis, *Journal of the Korea International Commerce Review*, 32(4) 399-413.

Song, H. S. (2017). Comparison of Performance between MLP and RNN Model to Predict Purchase Timing for Repurchase Product, *Journal of Information Technology Applications & Management*, 24(1), 111-128.

Silvey, S. D. (1969). Multicollinearity and Imprecise Estimation, *Journal of the Royal Statistical Society*, 31(3) 539-552.

Shin, S., Lee, M., and Song, S. K. (2018). A Prediction Model for Agricultural Products Price with LSTM Network, *Journal of the Korea Contents Association*, 18(11), 416-429.

Wu, J. Q., Wu, J. Q., and Chen, X. B. (2012). Import Iron Ore Price Forecasting based on PSO-SVMs Model, *The 7th International Computer Science & Education Conference*, Jul, 14-17, Melbourne, VIC, Australia, pp. 32-35.

Won, D. H., and Yang, C. H. (2011). A Study on the Validity of Logistics Collaboration for Steel Exports and Imports - Focused on Trade with China and Japan, *Korea Trade Research Association*, 36(3), 321-342.



이 우 창 (Woo Chang Lee)

- 정회원
- 계명대학교 통계학과 학사
- (현) 계명대학교 경영정보학과 석사과정
- 관심분야: 빅데이터, 머신러닝, 인공지능



김 양 석 (Yang Sok Kim)

- 서울시립대학교 경제학과 학사
- University of Tasmania 컴퓨터공학 석사
- University of Tasmania 컴퓨터공학 박사
- (현) 계명대학교 경영대학 경영정보학과 교수
- 관심분야: 머신러닝, 추천시스템



김 정 민 (Jung Min Kim)

- 계명대학교 통계학과 학사
- (현) 계명대학교 경영정보학과 석사과정
- 관심분야: Text Mining, Big data, Machine Learning



이 충 권 (Choong Kwon Lee)

- 종신회원
- 계명대학교 경영정보학과 학사
- Southeast Missouri State Univ. MBA
- Univ. of Nebraska-Lincoln 경영정보시스템 박사
- (현) 계명대학교 경영대학 경영정보학과 교수
- 관심분야: 빅데이터, IT인력