



시계열 분석 모델을 이용한 조선 산업 주요물가의 예측에 관한 연구

함주혁[†]
한라대학교 스마트모빌리티공학부

A Study on the Prediction of Major Prices in the Shipbuilding Industry Using Time Series Analysis Model

Juh-Hyeok Ham[†]
School of Smart Mobility Engineering, Halla University

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oil and steel prices, which are major prices in the shipbuilding industry, were predicted. Firstly, the error of the moving average line (N=3-5) was examined, and in all three error analyses, the moving average line (N=3) was small. Secondly, in the linear prediction of data through existing theory, oil prices rise slightly, and steel prices rise sharply, but in reality, linear prediction using existing data was not satisfactory. Thirdly, we identified the limitations of linear prediction methods and confirmed that oil and steel price prediction was somewhat similar to actual moving average line prediction methods. Due to the high volatility of major price flows, large errors were inevitable in the forecast section. Through the time series analysis method at the end of this paper, we were able to achieve not bad results in all analysis items relative to artificial intelligence (Prophet). Predictive data through predictive analysis using eight predictive models are expected to serve as a good research foundation for developing unique tools or establishing evaluation systems in the future. This study compares the basic settings of artificial intelligence programs with the results of core price prediction in the shipbuilding industry through time series prediction theory, and further studies the various hyper-parameters and event effects of Prophet in the future, leaving room for improvement of predictability.

Keywords : Prediction of major prices in the shipbuilding industry(조선산업의 주요 물가 예측), Artificial intelligence(인공지능), Time series analysis(시계열 분석), Oil and steel price(유가와 철강가격)

1. 서론

인공지능과 관련하여 일반인들의 관심을 끌기 시작한 때는 오래되었으나 그 적용성과 효율성 등 한계점이 보이기 시작하여 관심에서 사라진듯하다가, 활성화 함수들의 개선으로 Vanishing 문제 등도 해결되고 컴퓨터의 성능과 GPU 활용 등에 힘입어 빠르게 계산 할 수 있는 하드웨어의 발달에 도움을 받아, 딥러닝을 선두주자로 수년전 제4차 산업혁명의 시대로 갑자기 바뀌는 분위기로 전환되는 듯한 과도기를 거쳐, 이제는 많은 분야에서 인공지능이 주도적 역할을 해내고 있다고 사료된다.

그러나 개인적으로 학습하고 그 이론을 적용하여 본인의 전공에 적용과 응용 (Ham, 2018)이 쉽지는 않다고 사료되지만, 우선 일반 인공지능과 요즘 각광을 받고 있는 딥러닝(deep learning)의 다양한 분야로의 응용 (Andreas & Seragado, 2017)과 활용 (Kujira, 2017; Min, 2017)에 관심을 가지게 되었다. 따라서 본 연구에서는 본인의 전공과의 융합적용 (Ham, 2021)에 앞서 그 적용의 첫 단계로 인공지능과 관련 data base를 상대적으로 취득하기 쉽고 인공지능 프로그램도 공개된 툴을 사용하여 조선 산업과 관련된 가장 주요한 물가들이라고 판단되는 유가와 철강가격을 예측하여 보고 실제 결과와 가능한 최소한의 유용한 기존 시계열 예측법과 비교하여 보았다. 본 연구자는 올해 65세로, 정년은 22학년도 한해 남았지만, 재직하

는 대학의 업적에 조금이라도 기여키 위해, 대학원생도 없는 사립지방대학교수 단독으로 60대에도 매년 한편씩은 게재해 왔으나 이번 논문은 시계열 전문가의 입장에서 보면 부족한 면도 많을 거라 사료된다. 그러나 기본적 공학 소양을 가진 비전문가들은 당연히 엄두도 내지 못했던 조선 산업의 주요 물가를, 가볍게 인공지능의 훌륭한 도구들이나 오픈소스를 이용하여, 조선 산업 전문가들도 예측해 볼 수 있는 응용 기회의 장이될 수 있다고 생각한다. 왜냐하면 저자는 국내 최대 조선사 조선사업부의 핵심 연구소에 대학원을 졸업하고 입사하여 햇수로 한 15년 근무한 경험이 있어, 계약한 선박가격이 약 1~2년 후 인도 시점에는 2배로 뛰거나 반 토막 나는 경우도 발생하여 조선관련 주요 물가인 철강가격이나 유가의 예측은 선박계약 시점에서 조선관련 주요물가가 어느 위치에 있는가가 조선소의 흥망성쇠를 가름하는 엄청나게 중요한 일이기 때문에 관련 분야 모든 이의 관심사이기 때문이다. 이러한 사유로 첫 시도한 바 있는 응용 논문으로, 작년에 게재불가 판정의 내용 (Ham, 2019a & 2019b)을 기존 시계열분석이론에 근거하여 예측결과 및 실제물가와 비교분석 단계를 삽입하고 심사자의 의견을 가능한 반영하였다 (Ham, 2021a & 2021b). 그리고 이를 통해 5년 이상 째째이 공부한 인공지능 응용과 관련하여 정년 한 해 앞선, 본인의 처음이자 마지막 게재논문이 되는 영광이 있었으면 한다. 서언이 길어져 다시 간단히 요약하면 연구의 이론적 배경은 조선 산업 관련 1~2년 후의 주요물가 예측이 매우 중요하나 제공되어 활용되는 프로그램의 확보가 쉽지 않고 관련된 경험 풍부한 예측 전문가도 찾기 힘들며 예측이 쉽지 않아 조선 산업의 탄력적 영업력을 효율성을 향상시키기 어려운 면이 많다고 사료된다. 따라서 본 연구의 방법으로 복잡한 파라미터의 처리 없이 손쉽게 예측할 수 있는 인공지능을 이용한 기존의 잘 알려진 툴의 활용성과 함께 다양한 예측 모델을 사용하여 조선 산업의 주요물가의 실제 데이터들에 적용하여 보고 그 결과들을 비교 분석하여 본다. 연구 목표는 다양한 예측모델을 통한 분석을 통하여, 기존 도구의 활용을 위한 개선의 여지나 다양한 예측모델을 통한 분석을 통해 앞으로의 조선 산업 주요물가 예측의 정확도 향상이나 다양한 예측모델 분석을 통한 효율성 있는 예측모델과 방법을 검토하여, 본 연구를 토대로 고유의 툴을 개발하거나 평가 시스템 구축하는 기초연구가 되었으면 한다.

2. Prophet 개요

인공지능 등의 프로그램들이 깃허브(github)에서 제공하는 오픈소스 개념으로 많이 제공되고 있어, 본 연구에서는 페이스북(facebook)의 연구원들이 예측에 초점을 두고 개발된 예측모델인 Prophet을 활용하여 예측을 수행하여 보았다. 시계열분석은 통계적으로 어렵고 복잡한 작업으로 통계학의 깊은 지식이 요구되나 간단히 데이터를 예측하는 가벼운 느낌의 유용한 도구로 잘 알려진 바도 있으나, Prophet은 페이스북의 Core Data Science팀에서 개발한 만큼 일반적인 경우 기본 값만 사용해도

높은 성능을 보여 주며, 현재는 구글에 적용 실적이 다양한 유명한 인공지능 프로그램(기여자 현재 약 124명)이다 (Prophet, 2019). 그리고 최근에는 2조원대의 유니콘 기업인 하이퍼커넥트의 데이터분석기들이 다양한 물가정보를 예측하고 변화에 대응하는 전략을 수립하는 업무영역에도 활용되고 있음을 확인한 바 있다 (Jang, 2020). 참고문헌을 인용할 때에는 peer-review가 된 출판물을 인용하는 것이 적절하다고 생각하나 본 프로그램은 내부 알고리즘을 공개하지 않아 간략한 모델개요관련 출판 전 문헌 (Taylor & Letham, 2017)을 근거로 소개함을 양해 바란다.

2.1 Prophet 예측 모델

Prophet 예측 모델은 비즈니스 시계열의 일반적인 특징을 처리하도록 설계된 시계열 예측 모델이다. 중요한 것은, 기본 모델의 세부 사항을 알지 않고도 조정할 수 있는 직관적인 파라미터를 갖도록 설계되어 있다는 점이다. 이는 분석가가 모델을 효과적으로 조정하기 위해 필요하다. 프로그램의 구현은 Prophet이라고 불리는 파이썬과 R에서 오픈 소스 소프트웨어로 이용 가능하다. Prophet은 세 가지 주요 기능을 가진 분해 가능한 시계열 모델 (Harvey & Peters 1990)을 사용한다. 모델 구성 요소로는 추세, 계절성 그리고 공휴일로 그것들은 다음 식 (1)과 같은 방정식으로 구성된다.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

여기서 $g(t)$ 는 시계열 값의 비주기적 변화를 모델링하는 추세 함수로서, $s(t)$ 는 주기적인 변화(예: 주간 및 연간 계절성)를 나타내며, $h(t)$ 는 1일 이상 동안 잠재적으로 불규칙한 일정에 발생하는 휴일의 영향을 나타낸다. 오차 항목인 ϵ_t 는 모델에 의해 수용되지 않는 모든 특이적 변화를 나타낸다.

2.2 트렌드 모델

2.2.1 비선형, 포화 성장

데이터 생성 프로세스의 핵심 요소는 모집단이 어떻게 성장했으며 어떻게 지속적으로 성장할 것으로 예상하는지에 대한 모델이다. 페이스북에서 성장을 모델링하는 것은 비선형 성장이 있는 자연 생태계의 인구 증가와 유사하다 (Hutchinson, 1978). 예를 들어, 특정 영역의 페이스북 사용자 수는 휴대폰 등을 통한 인터넷에 접속할 수 있는 사람의 수일 것이다. 이러한 종류의 성장은 일반적으로 로지스틱 성장 모델을 사용하여 모델링되며, 가장 기본적인 형태는 다음 식 (2)와 같다.

$$g(t) = C / (1 + \exp(-k(t - m))) \quad (2)$$

여기서, C는 용량, k는 성장률 그리고 m은 오프셋 파라미터이다.

2.2.2 변경 점을 사용한 선형 추세

포화 성장을 나타내지 않는 문제를 예측하는 경우, 조각별 일정한 성장률은 점증적 형태이고 종종 유용한 모델을 제공한다. 트렌드 모델은 다음의 식 (3)과 같이 표시된다.

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (3)$$

여기서, k가 성장률이고 δ 는 속도 조절량, m은 오프셋 파라미터, γ 는 연속함수 관련 파라미터이다.

2.3 계절성

업무 시간 등은 대표하는 인간 행동의 결과로서 다주기적 계절성을 갖는 경우가 많다. 예를 들어, 주 5일 근무는 매주 반복되는 시계열에서 효과를 낼 수 있고, 방학 일정과 방학은 매년 반복되는 효과를 준다. 이러한 효과를 적절히 예측하기 위해 우리는 t의 주기적인 기능인 계절성 모델을 정해야 한다. 우리는 푸리에 급수(Fourier series)에 의존하여 주기적인 효과의 유연한 모델을 제공한다(Harvey & Shephard 1993). P는 시계열의 정규 기간(예: 일 단위로 시간 변수를 확장할 때 연간 데이터의 경우 P = 365.25 또는 주간 데이터의 경우 P = 7)이 되도록 한다. 식 (4)와 같은 표준 푸리에 급수로 임의의 매끄러운 계절적 효과를 대략적으로 알 수 있다.

$$s(t) = \sum_{n=1}^N (a_n \cos(\frac{2\pi n t}{P}) + b_n \sin(\frac{2\pi n t}{P})) \quad (4)$$

2.4 휴일 및 이벤트

휴일이나 이벤트는 크고 다소 예측 가능한 충격을 주며 종종

주기적인 패턴을 따르지 않기 때문에, 그 효과는 매끄러운 주기에 의해 잘 모델링되지 않는다. 음력을 따르는 주요 명절 등의 특정 공휴일이 시계열에 미치는 영향은 매년 비슷하기 때문에 이를 예측에 반영하는 것이 중요하다.

3. 데이터베이스 구축

주요물가 대상은 조선 산업의 주요 물가인 유가와 철강가격을 예측하기 위한 데이터베이스를 준비하여 보았다. 먼저 유가 데이터를 준비하였다. 데이터베이스관련 자료습득은 네이버 증권(선물)에서 취득한 두바이 유가의 평균가격 관련 데이터를 준비하였다. Fig.1에서 왼쪽은 2017년 7월 20일부터 2019년 7월 19일까지의 2년간의 두바이원유의 평균가격, 등락가격과 등락률 데이터 546개이며 오른쪽 자료는 2014년 7월 21일부터 2019년 7월 19일까지의 5년간의 두바이 원유의 평균가격, 등락가격과 등락률 데이터 1301개의 일부 데이터 화면이다.

두 번째는 철강가격 데이터를 준비하였다. 데이터베이스관련 자료습득은 Investing.com에서 취득한 다우존스 Iron&Steel의 평균가격 관련 데이터를 다음 Fig. 2와 같이 준비하였다.

본 그림에서 2014년 7월 23일부터 2019년 7월 19일까지의 5년간의 다우존스 Iron&Steel의 평균가격 등의 데이터 1257개의 데이터의 일부 화면이다. 오른쪽에 데이터는 날자가 한글이고 평균가격이 없어 프로그램이 돌아갈 수 있게 다운받은 파일을 오른쪽과 같이 데이터 처리를 하였다. 일반적으로 유가나 철강가격 등도 일반 주가와 마찬가지로 시시각각 변동성이 크다고 볼 수 있어 오차가 커서 예측이 불가하다고 해도 과언은 아니나 가능한 노이즈를 줄이기 위해 입력 자료를 일일 데이터를 최대 최소의 평균치인 일일 평균값으로 변환한 값만을 입력하여 노이즈를 크게 순화시켜 사용하였다.

	A	B	C	D	F	G	H	I
1	Date	AveragePrice			실제는 네이버증권(선물)에서 따온 증가임			
2	2019.07.19	62.59	0.28	0.44%				
3	2019.07.18	62.31	0.54	-0.85%	2019.07.19	2017.07.20	546	
4	2019.07.17	62.85	0.34	-0.53%				
5	2019.07.16	63.19	1.12	-1.73%				
6	2019.07.15	64.31	0.14	-0.20%				
7	2019.07.12	64.44	0.2	0.31%				
8	2019.07.11	64.24	0.38	-0.58%				
9	2019.07.10	64.62	1.85	2.94%				
10	2019.07.09	62.77	0.16	0.25%				
11	2019.07.08	62.61	0.08	0.12%				
12	2019.07.05	62.53	0.29	0.46%				
13	2019.07.03	62.25	1.22	2.00%				
14	2019.07.02	61.02	2.17	-3.42%				
15	2019.07.01	63.19	0.21	0.33%				
16	2019.06.28	61.76	0.03	0.04%				
17	2019.06.27	61.73	0.04	-0.06%				
18	2019.06.26	61.77	0.13	0.21%				
19	2019.06.25	61.63	0.01	-0.01%				
20	2019.06.24	61.64	0.17	-0.27%				
21	2019.06.21	61.81	0.23	0.36%				
22	2019.06.20	61.59	0.94	1.54%				
23	2019.06.19	60.65	0.01	0.01%				

	A	B	C	D	E	F	G	H
1	Date	AveragePrice						
2	2019.07.19	62.59	0.28	0.44%				
3	2019.07.18	62.31	0.54	-0.85%	2019.07.19	2014.07.21	1301	
4	2019.07.17	62.85	0.34	-0.53%				
5	2019.07.16	63.19	1.12	-1.73%				
6	2019.07.15	64.31	0.14	-0.20%				
7	2019.07.12	64.44	0.2	0.31%				
8	2019.07.11	64.24	0.38	-0.58%				
9	2019.07.10	64.62	1.85	2.94%				
10	2019.07.09	62.77	0.16	0.25%				
11	2019.07.08	62.61	0.08	0.12%				
12	2019.07.05	62.53	0.29	0.46%				
13	2019.07.03	62.25	1.22	2.00%				
14	2019.07.02	61.02	2.17	-3.42%				
15	2019.07.01	63.19	0.21	0.33%				
16	2019.06.28	61.76	0.03	0.04%				
17	2019.06.27	61.73	0.04	-0.06%				
18	2019.06.26	61.77	0.13	0.21%				
19	2019.06.25	61.63	0.01	-0.01%				
20	2019.06.24	61.64	0.17	-0.27%				
21	2019.06.21	61.81	0.23	0.36%				
22	2019.06.20	61.59	0.94	1.54%				
23	2019.06.19	60.65	0.01	0.01%				

Fig. 1 Average price of Dubai crude oil, fluctuating price and fluctuating rate data

DowJonesIronSteel(Investingdotcom)						dowJones5yprophet										
	A	B	C	D	E	F		A	B	C	D	E	F	G	H	
1225	2019년 06월 03일	198.31	201.38	194.92	201.8	194.82	25	1	2014-07-23	241.56	241.86	242.39	242.39	240.73	4.90M	-0.14%
1226	2019년 06월 04일	204.49	205.84	203.53	206.46	202.51	27	2	2014-07-24	241.19	239.35	242.33	243.7	238.68	4.58M	-1.04%
1227	2019년 06월 05일	204.67	205.81	206.26	206.26	203.07	28	3	2014-07-25	239.32	240.54	239.23	240.87	237.77	3.84M	0.50%
1228	2019년 06월 06일	205.23	205.65	205.79	207.11	203.34	29	4	2014-07-28	240.23	240.5	240.39	241.78	238.68	3.56M	-0.02%
1229	2019년 06월 07일	205.59	207.02	205.94	207.84	203.33	30	5	2014-07-29	240.69	238.97	240.99	242.4	238.97	6.36M	-0.64%
1230	2019년 06월 10일	208.98	207.59	208.78	210.68	207.27	31	6	2014-07-30	241.78	241.37	239.74	244.11	239.44	8.16M	1.00%
1231	2019년 06월 11일	211.5	211.35	209.76	213.29	209.71	32	7	2014-07-31	239.17	237.91	240.74	241.54	236.79	5.94M	-1.43%
1232	2019년 06월 12일	210	209.42	210.39	211.13	208.86	33	8	2014-08-01	237.55	237.33	237.12	240.08	235.02	5.62M	-0.24%
1233	2019년 06월 13일	211.81	213.13	210.51	213.77	209.85	34	9	2014-08-04	238.88	240.35	238.42	240.69	237.07	5.33M	1.27%
1234	2019년 06월 14일	211.14	210.51	212.42	212.44	209.84	35	10	2014-08-05	239.52	238.98	239.19	241.49	237.54	5.10M	-0.57%
1235	2019년 06월 17일	208.44	208.35	209.99	211.1	205.77	36	11	2014-08-06	240.2	241.34	238.23	242.17	238.23	4.01M	0.99%
1236	2019년 06월 18일	215.25	217.46	210.88	219.78	210.72	37	12	2014-08-07	240.38	239.44	241.74	242.46	238.3	3.76M	-0.79%
1237	2019년 06월 19일	222.38	222.83	220.5	224.26	220.5	38	13	2014-08-08	241.86	244.26	239.54	244.29	239.43	3.46M	2.01%
1238	2019년 06월 20일	224.59	222.82	225.84	227.83	221.34	39	14	2014-08-11	245.21	244.22	244.8	246.54	243.88	3.48M	-0.02%
1239	2019년 06월 21일	222.41	222.24	221.85	223.38	221.44	40	15	2014-08-12	243.51	243.68	243.95	244.79	242.22	2.78M	-0.22%
1240	2019년 06월 24일	220.8	220.65	221.8	222.53	219.06	41	16	2014-08-13	245.87	248.03	244	248.1	243.64	3.19M	1.79%
1241	2019년 06월 25일	223.51	222.87	221.53	226.16	220.85	42	17	2014-08-14	247.81	247.86	248.06	248.39	247.23	3.19M	-0.07%
1242	2019년 06월 26일	225.36	227.39	225.31	227.95	222.76	43	18	2014-08-15	247.3	246.71	248.14	248.92	245.68	4.14M	-0.46%
1243	2019년 06월 27일	226.74	227.14	227.44	227.98	225.5	44	19	2014-08-18	250.05	253.08	246.97	253.13	246.97	4.57M	2.58%
1244	2019년 06월 28일	228.37	229.28	228.14	230.06	226.68	45	20	2014-08-19	254.23	253.96	253.35	255.15	253.3	3.89M	0.35%
1245	2019년 07월 01일	230.95	231.36	231.3	232.4	229.49	46	21	2014-08-20	253.65	253.5	253.25	254.61	252.68	2.84M	-0.18%
1246	2019년 07월 02일	229.21	228.67	229.99	230.17	228.24	47	22	2014-08-21	251.94	251.86	253.48	253.48	250.4	3.38M	-0.65%
1247	2019년 07월 03일	228.65	228.54	229.18	229.63	227.66	48	23	2014-08-22	251.4	252.21	251.4	252.64	250.15	4.11M	0.14%
1248	2019년 07월 05일	227.36	229.25	226.25	229.72	225.14	49	24	2014-08-25	253.35	253.62	252.3	254.39	252.3	4.00M	0.56%
1249	2019년 07월 08일	226.28	223.88	228.06	229.72	222.83	50	25	2014-08-26	255.72	256.91	253.89	257.54	253.89	3.64M	1.30%
1250	2019년 07월 09일	221.65	224	222.08	224.13	219.17	51	26	2014-08-27	255.96	255.59	257.08	257.31	254.6	2.82M	-0.51%
1251	2019년 07월 10일	223.98	221.48	225.57	226.65	221.3	52	27	2014-08-28	252.94	251.93	254.39	254.39	251.48	4.27M	-1.43%
1252	2019년 07월 11일	220.94	220.98	220.8	222.06	219.81	53	28	2014-08-29	252.74	253.8	252.57	253.98	251.49	3.69M	0.74%
1253	2019년 07월 12일	224.2	225.93	222.16	226.27	222.12	54	29	2014-09-02	252.98	252.61	254.02	254.16	251.8	4.21M	-0.47%
1254	2019년 07월 15일	226.29	226.82	227.29	228.24	224.33	55	30	2014-09-03	255.68	256.45	253.75	257.6	253.75	6.42M	1.52%
1255	2019년 07월 16일	229.99	229.8	227.17	233.51	226.46	56	31	2014-09-04	258.95	258.93	256.47	261.42	256.47	5.23M	0.97%
1256	2019년 07월 17일	227.44	226.42	229.2	229.33	225.54	57	32	2014-09-05	257.98	259.18	258.58	259.19	256.76	3.13M	0.10%
1257	2019년 07월 18일	226.41	226.65	226.22	227.73	225.09	58	33	2014-09-08	257	256.01	258.87	259.02	254.98	4.81M	-1.22%
1258	2019년 07월 19일	230.97	232.69	229.14	234.24	227.7	59	34								

Fig. 2 Data on average prices for Dow Jones Iron & Steel

4. 인공지능 해석 결과

4.1 인공지능 유가 예측

상기 Fig. 1과 같은 2014년 7월 21일부터 2019년 7월 19일까지의 5년간의 두바이 원유의 평균 가격의 변동 그래프(Fig.3 참조)의 csv형 자료데이터를 Prophet의 입력데이터로 하여 분석과 예측을 수행하였다.

아래 Fig. 4와 같은, 2014년 7월 21일부터 2019년 7월 19일까지의 5년간의 두바이원유의 평균가격의 변동 그래프의 csv형 자료 데이터에서 두 가지의 주기데이터를 분류해 내었다. 상부의 그림은 주간 주기성이나 주말 토 & 일요일은 주식시장이 휴장으로 데이터가 없기 때문에 야기되는 현상으로 무의미하며 아래는 5년간 데이터에서 찾아낸 년 간 주기성으로 6월과 10월 경 고점을 보이며 1~2월이 저점을 보인다. 그리고 전체적인 패턴의 경향은 아래 Fig. 5과 같다.

즉 패턴에 주기성을 고려하여 다음 1년을 예측한 것으로 Fig. 6과 같이 지그재그 하향 선으로 예측하고 있다.



Fig. 3 Five years of average Dubai crude oil price (from Naver Securities Futures)

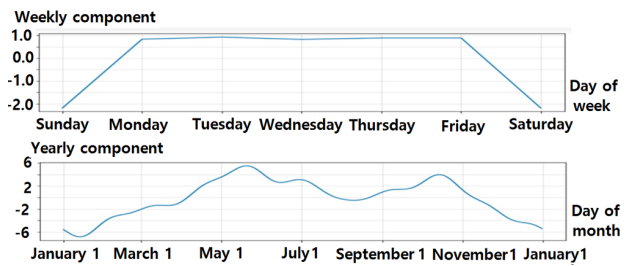


Fig. 4 Periodic component of the average Dubai crude oil price over five years (from Naver Securities Futures)

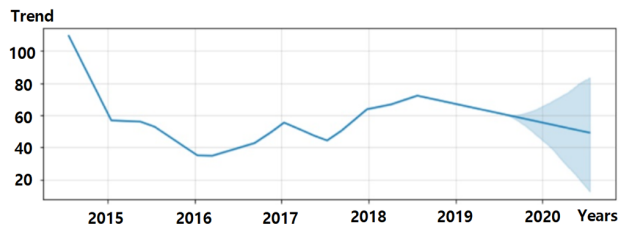


Fig. 5 A pattern of five years' average Dubai crude oil price (from Naver Securities Futures)

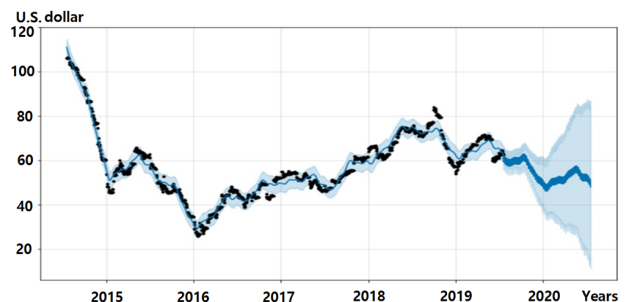


Fig. 6 Results of a one-year crude oil price forecast using data from Dubai for an average of five years (from Naver Securities)

4.2 인공지능 철강가격 예측

Fig. 7과 같이 2014년 7월 23일부터 2019년 7월 19일까지의 5년간의 다우존스 Iron & Steel의 평균가격의 변동 그래프의 csv형 자료 데이터를 Prophet의 입력데이터로 사용하여 분석과 예측을 수행하였다.

Fig. 8와 같은, 2014년 7월 23일부터 2019년 7월 19일까지의 5년간의 평균 다우존스 Iron & Steel 평균가의 변동 그래프의 csv형 자료데이터로부터 두 가지의 주기데이터를 분류해내었다.

Fig. 8의 상부 그림은 주간 주기성으로 주말 토와 일요일은 주식시장이 휴장으로 데이터가 없기 때문에 야기되는 현상으로 그림 자체에는 큰 의미는 없으며 하부 그림은 데이터에서

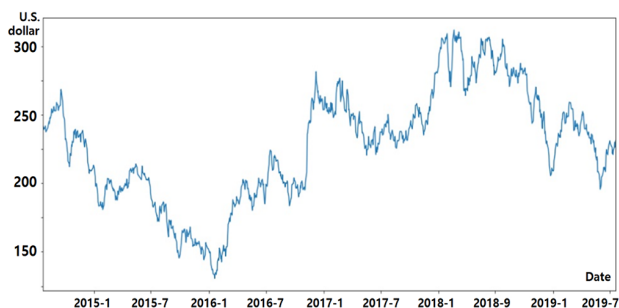


Fig. 7 Five-year Dow Jones iron & steel average charts (from Investing.com)

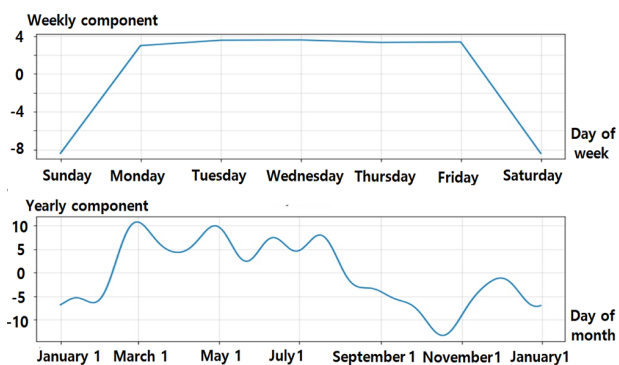


Fig. 8 Two periodic data classified from a five-year average of Dow Jones Iron & Steel's fluctuating graph.

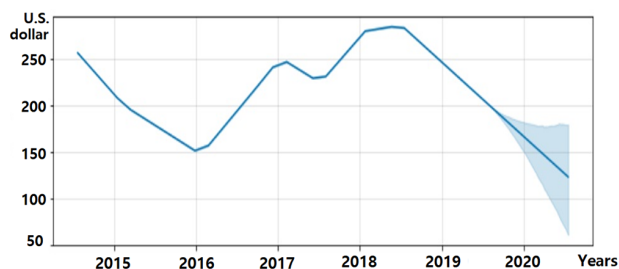


Fig. 9 Five-year Dow Jones iron & steel average price patterns

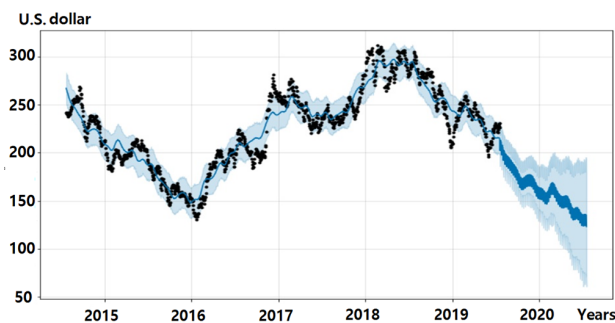


Fig. 10 1 Year's forecast result using five-year Dow Jones Iron & Steel average

찾아낸 년 간 주기성으로 3-7월의 고점 대 영역을 보이며 11월초가 저점을 보인다. 그리고 전체적인 패턴의 경향은 아래 Fig. 9와 같다.

그리고 마지막으로 패턴에 주기성을 고려하여 다음 1년을 예측한 것으로 거의 지속적인 하 방향으로 예측되고 있다.

5. 기존 이론을 통한 분석

5.1 유가 예측

상기 Fig. 3과 같은 2014년 7월 21일부터 2019년 7월 19일까지의 그래프에 추가적으로 약 1년 치의 실제 데이터 값을 예측과 비교하기 위해 2020년 7월 17일까지의 6년간의 두바이 원유의 평균 가격의 변동 그래프의 엑셀 자료데이터 중 약 5년을 입력데이터로 하여 1년 후를 기존의 기본 예측이론에 근거하여 예측하여 보고 1년 후의 결과자료와 비교 분석을 다양한 방법으로 수행하였다.

5.1.1 유가 예측을 위한 평균이동선의 오차분석

이동평균선의 n 일 이동평균선 중 n=3~5 영역의 오차를 분석하여 보았다. 전체 범위의 그래프에서는 각 이동평균선 그래프가 구분이 가지 않아 약 3개월간의 이동평균선을 위주로 실제 평균 주가와 함께 Fig. 11에 나타내었다.

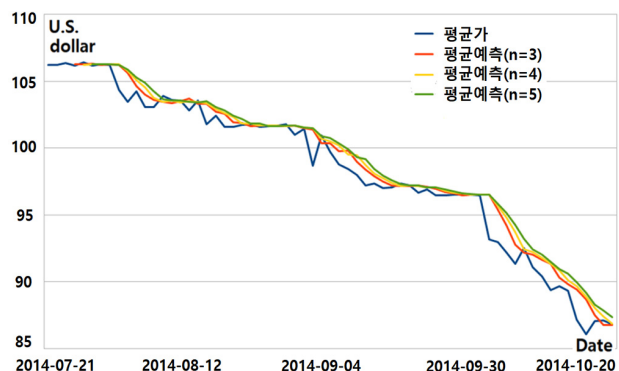


Fig. 11 Average moving line for predicting oil prices(n=3~5)

Table 1 Error estimation of average moving lines for estimating oil prices

Error analysis method	n=3	n=4	n=5
MAE	0.79	0.90	1.00
MSE	1.92	2.35	2.80
MAPE(%)	1.58	1.80	2.01

n=3이 n=4와 n=5에 비해 3 가지 오차 분석법인 MAE(Mean absolute error), MSE(Mean squared error) 그리고 MAPE (Mean absolute percentage error)에서 오차가 상기 Table 1과 같아 N=3 이 공통되게 모두 작아, 예측 작업에 n=3 평균 선을 적극 활용하였다.

5.1.2 유가 예측을 위한 데이터 분석 예

시계열 해석의 유가 예측 데이터의 분석의 한 예를 간략히 제시하여 보았다. 이전의 값을 기초로 한, 선형의 AR(Autoregressive) Model을 구하였고 t-1부터 t-3까지의 기울기와 절편은 각각 1.165, -0.336 0.017 그리고 8.479이다. t-1과 t-3은 양의 값으로 +영향을 보이고, t-2는 음의 영향을 미친다. P값은 단지 t-1만이 1.94E-11의 값을 보여 significant 하다고 사료된다.. 즉 t-1만이 심대한 영향을 미친다고 볼 수 있다. 결정계수인 R square 값은 거의 85%로 전날 값으로 약 85%로 설명될 수 있으며 최종 AR모델의 상관관계식은 $y(t) = 1.165y(t-1) - 0.336y(t-2) + 0.017y(t-3) + 8.479$ 로 예측될 수 있다(Table 2 참조).

Table 2 Example of data analysis for predicting oil prices(AR model)

Regression analysis statistics	
Multiple correlations	0.923
Determination factor	0.851
Adjusted Det. Factor	0.843
Standard error	5.225
No. of observation	58

ANOVA					
	DOF	Square sum	Squared average	F ratio	Significant F
Regression	3	8450.496	2816.832	103.172	2.40E-22
Residual	54	1474.327	27.302		
Total	57	9924.824			

	Coefficient	Standard error	t statistics	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Y intercept	8.479	2.893	2.931	0.005	2.68	14.279	2.68	14.279
t-1	1.165	0.138	8.439	1.94E-11	0.888	1.442	0.888	1.442
t-2	-0.336	0.212	-1.589	0.118	-0.761	0.088	-0.761	0.088
t-3	0.017	0.131	0.128	0.899	-0.246	0.279	-0.246	0.279

5.1.3 유가 선형 예측

예측기간 약 1년을 포함한 총 6년간의 월평균데이터와 앞선 5년 데이터를 이용한 약 1년간의 선형 예측 그래프를 Fig. 12 제시하였고 Table 2에는 예측부의 값을 표시하였다.

선형 예측은 Table 3과 같이 미세한 증가를 보이거나 실제 월 평균치는 급락현상을 보여 본 내용에서 선형 예측은 바람직하지 못한 것으로 사료된다.



Fig. 12 Oil price linear prediction and actual monthly average

Table 3 Oil price linear prediction and actual monthly average

Date	Linear prediction	Actual monthly average
2019-08-01	60.23	58.8
2019-09-01	60.29	60.08
2019-10-01	60.34	58.75
2019-11-01	60.40	61.45
2019-12-01	60.45	63.74
2020-01-01	60.51	65.63
2020-02-01	60.56	54.45
2020-03-01	60.61	37.99
2020-04-01	60.67	22.9
2020-05-01	60.72	28.89
2020-06-01	60.78	40.34
2020-07-01	60.83	43.11

5.1.4 계절성을 고려한 유가 예측

앞의 선형예측의 부정확성을 확인하고 보다 바람직한 예측을 위해 예측하려는 기간의 최근 3년 전까지의 평균 계절지수를 계산하여 보았고 아래 Fig. 13 및 Table 4과 같이 계산되었다.

월평균치(n=3)곡선에 예측 당월의 계절지수를 곱하여 월평균 예측치를 구하는 반복법을 사용하여 두바이유가의 약 1년간의 월 평균치 주가를 예측하여 보았고 다음 Fig. 14 및 Table 5와 같다.

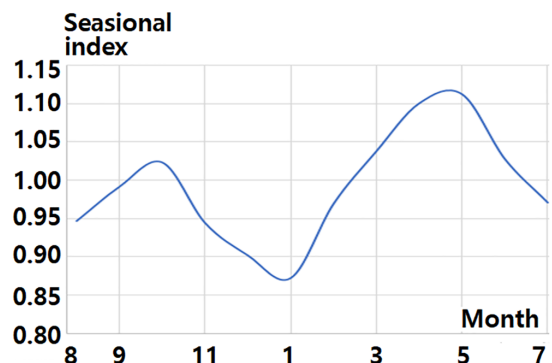


Fig. 13 Dubai oil seasonal index

Table 4 Seasonal index of Dubai oil price

Month	Seasonal index
August	0.946
September	0.991
October	1.023
November	0.944
December	0.901
January	0.872
February	0.969
March	1.038
April	1.100
May	1.112
June	1.028
July	0.971

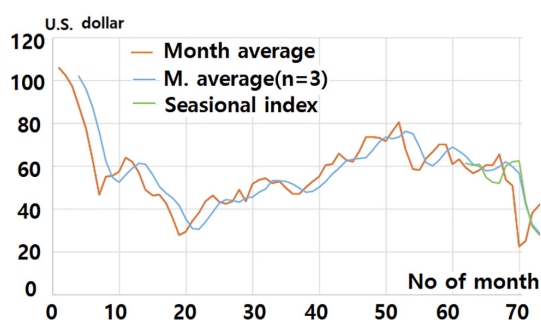


Fig. 14 Price forecasts per barrel of Dubai oil considering seasonal index

Table 5 Price forecasts per barrel of Dubai oil considering seasonal index

Date	Month average	N=3	Prediction
2019.08.01	59.40	64.71	61.24
2019.09.01	56.69	61.11	60.58
2019.10.01	58.11	59.73	61.11
2019.11.01	60.54	58.07	54.81
2019.12.01	60.41	58.45	52.69
2020.01.01	65.44	59.69	52.04
2020.02.01	53.50	62.13	60.21
2020.03.01	50.95	59.78	62.04
2020.04.01	22.46	56.63	62.31
2020.05.01	25.30	42.30	43.47
2020.06.01	38.32	32.90	31.94
2020.07.01	42.26	28.69	27.85

5.1.5 유가의 신뢰구간 설정

약 1년간의 두바이유가 예측치의 98% 신뢰 구간을 설정하여 Fig. 15 및 Table 6에 제시하였다.

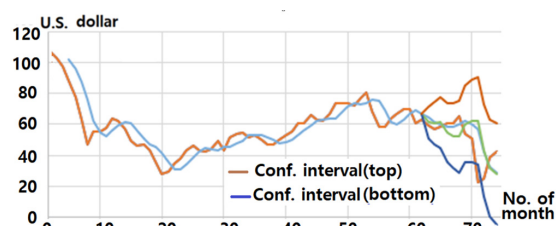


Fig. 15 Set confidence intervals for oil prices

Table 6 Set confidence intervals data for oil prices

Date	Prediction	confidence interval(top)	confidence interval (bottom)
2019.08.01	61.24	50.98	71.50
2019.09.01	60.58	46.78	74.39
2019.10.01	61.11	44.49	77.73
2019.11.01	54.81	35.78	73.83
2019.12.01	52.69	31.52	73.85
2020.01.01	52.04	28.93	75.15
2020.02.01	60.21	35.30	85.12
2020.03.01	62.04	35.45	88.63
2020.04.01	62.31	34.14	90.49
2020.05.01	43.47	13.79	73.14
2020.06.01	31.94	0.83	63.04
2020.07.01	27.85	-4.63	60.33

5.1.6 유가의 예측비교

1년간 유가의 Prophet을 통한 인공지능 예측법과 기존 예측법과 실제 값을 비교하여 보았고 상세 내역은 다음의 Table 7에 제시하였다.

Table 7 Artificial intelligence method of oil price for one year prediction, comparing actual value with existing forecast method

Date	Month average	Prophet	Time series analysis
2019.08.01	59.40	59.27	59.48
2019.09.01	56.69	59.13	60.09
2019.10.01	58.11	61.17	60.93
2019.11.01	60.54	56.72	55.94
2019.12.01	60.41	54.62	51.36
2020.01.01	65.44	52.34	48.56
2020.02.01	53.50	58.21	50.53
2020.03.01	50.95	59.75	51.35
2020.04.01	22.46	59.14	53.58
2020.05.01	25.30	44.91	55.57
2020.06.01	38.32	33.48	52.86
2020.07.01	42.26	29.02	51.14

본 연구는 1차 단계로 하이퍼파라메타를 Prophet의 디폴트 로 설정하여 평가하였다. 2020년 4월은 전 달보다 반값이하의 폭락으로 오일가격의 오차는 크나 세 가지 평균선의 전반적인 추세는 우 하향으로 유사하다. 그리고 예측 결과의 에러 평가는 다음 Table 8 및 Table 9와 같다.

Prophet의 하이퍼파라메타의 디폴트 설정을 이용하여 경직 성이 상대적으로 우세하여, 실제 가격의 변화에 대응이 빠르지 않아 오차가 시계열 결과 보다 미세하게 크나 최대에러는 시계열 분석이 크다. 추후 여러 하이퍼파라메타나 이벤트성 변동의 효과를 반영하는 조절을 통해 시계열 결과 이상의 좋은 예측치 획득을 위해 노력하겠다.

Table 8 Error analysis results of oil price for ten month prediction, comparing actual value with Time series analysis

Date	MAE	MSE	MAPE
2019-08-01	0.1	0.0	0.2
2019-09-01	2.4	6.0	4.3
2019-10-01	3.1	9.4	5.3
2019-11-01	3.8	14.6	6.3
2019-12-01	5.8	33.5	9.6
2020-01-01	13.1	171.6	20.0
2020-02-01	4.7	22.2	8.8
2020-03-01	8.8	77.4	17.3
2020-04-01	36.7	1345.4	163.3
2020-05-01	19.6	384.6	77.5
2020-06-01	4.8	23.4	12.6
2020-07-01	13.2	175.3	31.3
Average	9.7	188.6	29.7

Table 9 Error analysis results of oil price for ten month prediction, comparing actual value with Prophet

Date	MAE	MSE	MAPE
2019-08-01	0.1	0.0	0.1
2019-09-01	3.4	11.6	6.0
2019-10-01	2.8	8.0	4.9
2019-11-01	4.6	21.2	7.6
2019-12-01	9.1	81.9	15.0
2020-01-01	16.9	284.9	25.8
2020-02-01	3.0	8.8	5.6
2020-03-01	0.4	0.2	0.8
2020-04-01	31.1	968.5	138.6
2020-05-01	30.3	916.3	119.6
2020-06-01	14.5	211.4	37.9
2020-07-01	8.9	78.8	21.0
Average	10.4	216.0	31.9

5.2 철강가격 예측

4장에서 소개한 바 있는 Fig. 7과 같이 2014년 7월 23일부터 2020년 7월 17일까지의 6년간의 다우존스 Iron & Steel의 평균가격의 변동 그래프의 엑셀 자료데이터 중 약 5년을 입력 데이터로 하여 1년 후를 예측하여 보고 1년 후의 결과자료와 비교 분석을 다양한 방법으로 수행하였다.

5.2.1 철강가격 예측을 위한 평균이동선의 오차분석

이동평균선의 N 일 이동평균선 중 N=3~5 영역의 오차를 분석하여 보았다. 전체 범위의 그래프에서는 각 이동평균선 그래프가 구분이 가지 않아 약 3개월간의 이동평균선만 실제 평균 주가와 함께 그려보았다(Fig. 16 참조).

Table 10에서와 같이 n=3이 n=4와 n=5에 비해 3 가지 오차 분석 (MAE, MSE, MAPE)에서 오차가 n=3에서 공통되게 모두 작아 예측에 n=3을 적극 활용하였다.

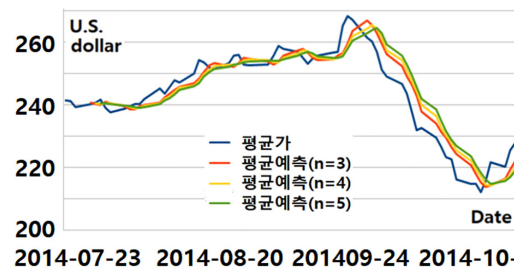


Fig. 16 Average moving line for predicting steel prices

Table 10 Error estimation of average moving line for predicting steel prices

Error analysis method	n=3	n=4	n=5
MAE	3.72	4.17	4.59
MSE	23.51	29.72	36.02
MAPE	1.73	1.94	2.14

5.2.2 철강가격 예측을 위한 예측데이터 분석 예

Time series 해석의 철강가격 예측 데이터의 분석의 한 예를 간략히 제시하여 보았다. 이전의 값을 기초로 한, 선형의 AR(Autoregressive) Model을 구하였고 t-1부터 t-3까지의 기울기와 절편은 각각 1.118, -0.245 0.050 그리고 17.055이다. t-1과 t-3은 양의 값으로 +영향을 미치고 t-2는 음의 영향을 미친다. P값은 단지 t-1만이 5.84E-11의 값을 보여 significant 하다고 사료된다. 즉 t-1만이 심대한 영향을 미친다고 볼 수 있다. 결정계수인 R square 값은 거의 88%로 전날 값으로 약 88%만큼 설명될 수 있으며 최종 AR모델의 상관관계식은 $y(t) = 1.118y(t-1) - 0.245y(t-2) + 0.050y(t-3) + 17.055$ 로 평가될 수 있다(Table 11 참조).

Table 11 Example of data analysis for steel price prediction(AR model)

Regression analysis statistics	
Multiple correlations	0.937
Determination factor	0.877
Adjusted Det. Factor	0.87
Standard error	15.078
No. of observation	57

ANOVA					
	DOF	Square sum	Squared average	F ratio	Significant F
Regression	3	86056.3	28685.4	126.182	4.00E-24
Residual	53	12048.7	227.334		
Total	56	98105			

	Coefficient	Standard error	t statistics	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Y intercept	17.055	11.539	1.478	0.145	-6.089	40.199	-6.089	40.199
t-1	1.118	0.137	8.177	5.84E-11	0.844	1.392	0.844	1.392
t-2	-0.245	0.204	-1.204	0.234	-0.654	0.164	-0.654	0.164
t-3	0.05	0.138	0.36	0.72	-0.228	0.328	-0.228	0.328

5.2.3 철강가격 선형 예측

5년간의 월평균데이터를 이용한 약 1년간의 선형 예측이다. 아래에 관련 그래프(Fig. 17)를 제시하였고 표 12에는 예측부의 값만 표시하였다.

선형 예측은 뚜렷한 상승세를 보이거나 실제로 급락현상 보여 본 내용에서도 선형 예측은 바람직하지 못한 것으로 사료된다.

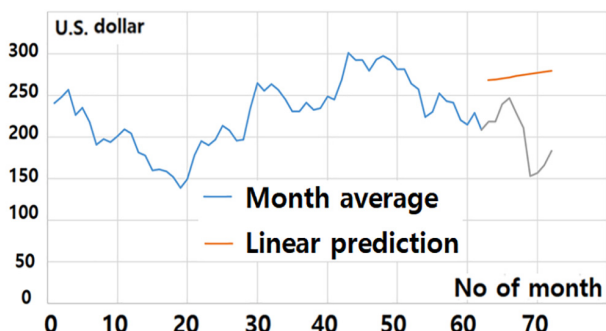


Fig. 17 Linear prediction of steel prices

Table 12 Linear prediction of steel price and actual monthly average

Date	Linear prediction	Actual month average
2019-09-01	267.91	218.32
2019-10-01	269.14	218.51
2019-11-01	270.42	239.11
2019-12-01	271.65	246.48
2020-01-01	272.92	228.79
2020-02-01	274.19	211.20
2020-03-01	275.38	152.71
2020-04-01	276.65	156.95
2020-05-01	277.88	166.07
2020-06-01	279.15	183.08

5.2.4 계절성을 고려한 철강가격 예측

앞의 선형예측의 부정확성을 확인하고 보다 바람직한 예측을 위해 예측하려는 기간의 최근 3년 전까지의 계절지수를 계산하여 보았고 아래 Fig. 18 및 Table 13과 같이 계산되었다.

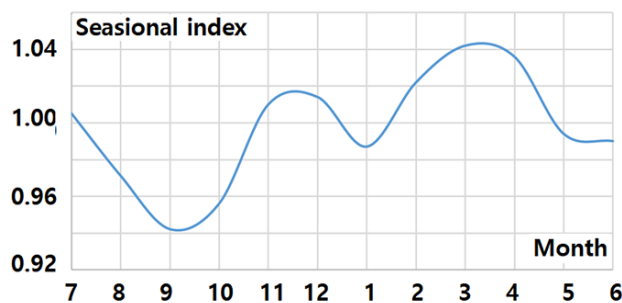


Fig. 18 Seasonal index for predicting steel prices

Table 13 Seasonal index of steel prices

Month	Seasonal index
July	1.005
August	0.971
September	0.942
October	0.956
November	1.01
December	1.014
January	0.987
February	1.022
March	1.042
April	1.036
May	0.994
June	0.99

월평균치(n=3)곡선에 예측 당월의 계절지수를 곱하여 월평균 예측치를 구하는 반복법을 사용하여 철강가격의 약 1년간의 월평균치 주가를 예측하여 보았고 Fig. 19 및 Table 14와 같다.

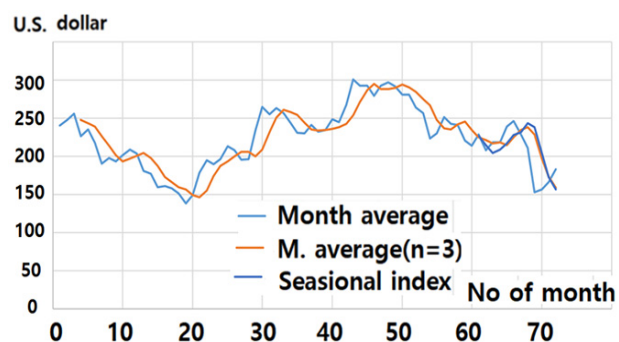


Fig. 19 Prediction of steel price considering seasonal index and actual price

Table 14 Prediction of steel price considering seasonal Index and actual price

Date	Month average	Ignore seasonality	Prediction
2019-07-01	229.0	225.40	226.53
2019-08-01	208.2	221.31	214.90
2019-09-01	218.3	217.22	204.62
2019-10-01	218.5	218.50	208.89
2019-11-01	239.1	215.00	217.15
2019-12-01	246.5	225.31	228.47
2020-01-01	228.8	234.70	231.65
2020-02-01	211.2	238.13	243.37
2020-03-01	152.7	228.82	238.43
2020-04-01	157.0	197.57	204.68
2020-05-01	166.1	173.62	172.58
2020-06-01	183.1	158.58	156.99

5.2.5 철강가격의 신뢰구간 설정

10개월간의 철강가격 예측치의 98% 신뢰 구간으로 표현하여 보았고 상세 내역은 다음 Fig. 20 및 Table 15에 제시하였다.

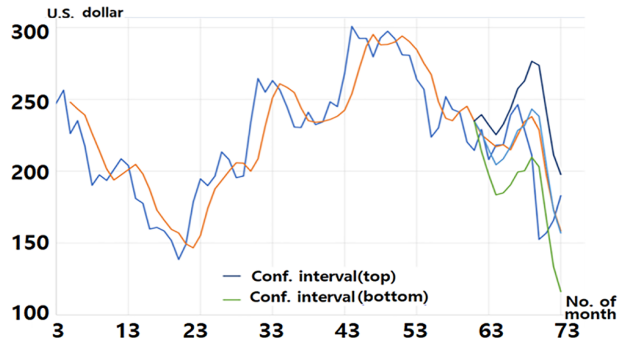


Fig. 20 Establishment of confidence interval for steel price

Table 15 Establishment of confidence interval for steel price

Date	Month average	N=3	Prediction	Confidence (-)	Confidence (+)
01-Jul	229	225	227	214	239
01-Aug	208	221	215	198	232
01-Sep	218	217	205	184	226
01-Oct	219	219	209	185	233
01-Nov	239	215	217	191	244
01-Dec	246	225	229	199	258
01-Jan	229	235	231	200	263
01-Feb	211	238	243	210	277
01-Mar	152	229	238	204	274
01-Apr	157	198	205	167	242
01-May	166	174	173	134	212
01-Jun	183	159	157	116	198

5.2.6 철강가격의 예측

약 1년간의 철강가격의 인공지능 방법, 기존 예측법과 실제 값을 비교하여 보았고 상세 내역은 Fig. 21 및 Table 16에 제시하였다.

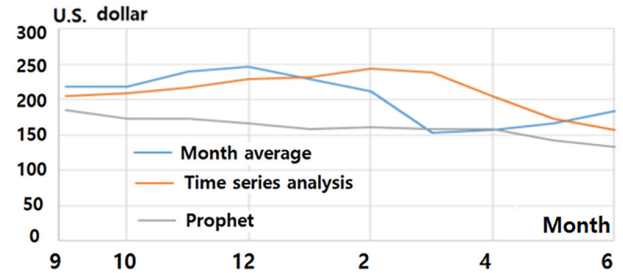


Fig. 21 Artificial intelligence method of steel price for ten month prediction, comparing actual value with existing forecast method

Table 16 Artificial intelligence method of steel price for ten month prediction, comparing actual value with existing forecast method

Date	Month average	Time series analysis	Prophet
2019-09-01	218.32	204.62	184.83
2019-10-01	218.51	208.89	172.91
2019-11-01	239.11	217.15	172.93
2019-12-01	246.48	228.47	166.15
2020-01-01	228.79	231.65	160.05
2020-02-01	211.20	243.37	243.37
2020-03-01	152.71	238.43	157.60
2020-04-01	156.95	204.68	157.57
2020-05-01	166.07	172.58	142.17
2020-06-01	183.08	158.99	131.54

본 연구는 첫 단계로 Prophet의 하이퍼파라미터를 디폴트로 설정하여 평가하였다. 2020년 3월의 폭락으로 철강 가격 오차는 크나 세 가지 선의 전반적인 추세는 우하향의 패턴이 공히 유지된다. 예측 결과의 오차평가는 Table 17와 Table 18와 같다. 즉, 시계열분석 예측 결과가 실제가격의 급격한 등락변동에 잘 대응하여 미세하게 모든 오차(MAE, MSE & MAPE)에서 Prophet 예측결과 보다 작으나 최대오차는 시계열분석이 크게 나타났다.

Prophet의 하이퍼파라미터의 기본값 설정으로 경직성이 상대적으로 우세하여 실제 가격의 변화에 대응이 크지 않아 오차가 시계열 결과 보다 크다고 사료된다. 추후 여러 하이퍼파라미터나 이벤트성 변동의 효과를 반영하는 추가적인 다양한 경험적 조절을 통해 시계열 결과 이상의 좋은 예측치 획득이 가능할 것으로 사료된다.

Table 17 Error analysis results of steel price for ten month prediction, comparing actual value with Time series analysis

Date	MAE	MSE	MAPE
2019-09-01	13.7	187.69	6.28
2019-10-01	9.62	92.54	4.40
2019-11-01	21.96	482.24	9.18
2019-12-01	18.01	324.36	7.31
2020-01-01	2.86	8.18	1.25
2020-02-01	32.17	1034.91	15.23
2020-03-01	85.72	7347.92	56.13
2020-04-01	47.73	2278.15	30.41
2020-05-01	6.51	42.38	3.92
2020-06-01	26.09	680.69	14.25
Average	26.44	1247.91	14.84

Table 18 Error analysis results of steel price for ten month prediction, comparing actual value with Prophet

Date	MAE	MSE	MAPE
2019-09-01	33.49	1121.58	15.34
2019-10-01	45.60	2079.36	20.87
2019-11-01	66.18	4379.79	27.68
2019-12-01	80.33	6452.91	32.59
2020-01-01	70.71	4999.90	30.91
2020-02-01	50.70	2570.49	24.01
2020-03-01	4.89	23.91	3.20
2020-04-01	0.62	0.38	0.40
2020-05-01	23.9	571.21	14.39
2020-06-01	50.54	2554.29	27.61
Average	42.70	2475.38	19.70

Table 19 Overall oil price and steel price forecasting with each predictive model

Oil price prediction														Seasonality		Seasonality	
														12		12	
														7. Prophet		8. MA	
														avg.	MAE	avg.	MAE
														avg.	MAE	avg.	MAE
2019-08-01	58.8	62.62	-6.50%	60.23	-2.43%	65.08	-10.68%	64.22	-9.22%	60.96	-3.67%	63.19	-7.47%	59.27	-0.80%	61.24	-4.15%
2019-09-01	60.08	63.34	-5.43%	60.29	-0.35%	62.89	-4.68%	63	-4.86%	61.78	-2.83%	65.08	-8.32%	59.13	1.58%	60.58	-0.83%
2019-10-01	58.75	64.85	-10.38%	60.34	-2.71%	63.72	-8.46%	63.44	-7.98%	62	-5.53%	62.89	-7.05%	61.17	-4.12%	61.11	-4.02%
2019-11-01	61.45	60.51	1.53%	60.4	1.71%	63.9	-3.99%	63.47	-3.29%	62.35	-1.46%	63.72	-3.69%	58.72	7.70%	54.81	10.81%
2019-12-01	63.74	56.63	11.15%	60.45	5.16%	63.6	0.22%	63.36	0.60%	62.66	1.69%	63.90	-0.25%	54.82	14.31%	52.89	17.34%
2020-01-01	65.83	51.69	21.24%	60.51	7.80%	63.71	2.93%	63.42	3.37%	62.83	4.27%	63.50	3.24%	52.34	20.25%	52.04	20.71%
2020-02-01	54.45	52.88	2.88%	60.56	-11.22%	63.7	-16.99%	63.41	-16.46%	63.00	-15.70%	63.71	-17.00%	58.21	-6.91%	60.21	-10.58%
2020-03-01	37.89	53.91	-41.91%	60.61	-59.54%	63.64	-67.52%	63.4	-66.89%	63.14	-66.20%	63.70	-67.68%	59.75	-57.28%	62.04	-63.31%
2020-04-01	22.9	54.93	-139.87%	60.67	-164.93%	63.68	-178.08%	63.41	-176.90%	63.24	-176.16%	63.64	-177.89%	59.14	-158.25%	62.31	-172.10%
2020-05-01	28.89	59.92	-107.41%	60.72	-110.18%	63.67	-120.39%	63.4	-119.45%	63.33	-119.21%	63.68	-120.43%	44.91	-55.45%	43.47	-50.47%
2020-06-01	40.34	58.68	-45.46%	60.78	-50.67%	63.66	-57.81%	63.4	-57.16%	63.4	-57.16%	63.67	-57.84%	33.48	17.01%	31.94	20.82%
2020-07-01	43.11	58.61	-35.95%	60.83	-41.10%	63.67	-47.69%	63.4	-47.07%	63.45	-47.18%	63.66	-47.68%	29.02	32.68%	27.85	35.40%

Steel price prediction														Seasonality		Seasonality	
														12		12	
														7. Prophet		8. MA	
														avg.	MAE	avg.	MAE
														avg.	MAE	avg.	MAE
2018-09-01	218.32	225.42	-3.25%	207.91	-22.71%	220.00	-0.77%	219.34	-0.47%	220.41	-0.96%	220.11	-0.82%	184.83	15.34%	204.82	6.28%
2018-10-01	218.51	219.26	-0.34%	209.14	-23.17%	221.84	-1.52%	220.12	-0.74%	220.33	-0.83%	220.00	-0.68%	172.91	20.87%	208.89	4.40%
2018-11-01	239.11	225.96	5.50%	270.42	-13.09%	220.65	7.72%	219.71	8.11%	220.63	7.73%	221.84	7.22%	172.93	27.68%	217.15	9.18%
2018-12-01	248.48	225.22	8.63%	271.85	-10.21%	220.83	10.41%	219.76	10.84%	220.63	10.49%	220.65	10.48%	168.15	32.59%	228.47	7.31%
2020-01-01	228.79	223.04	2.51%	272.92	-19.29%	221.10	3.36%	219.83	3.92%	220.67	3.55%	220.83	3.48%	180.05	30.05%	231.85	-1.25%
2020-02-01	211.2	230.93	-9.34%	274.19	-29.82%	220.86	-4.57%	219.78	-4.06%	220.76	-4.53%	221.10	-4.69%	243.37	-15.23%	243.37	-15.23%
2020-03-01	152.71	232.91	-52.52%	275.38	-80.33%	220.93	-44.67%	219.79	-43.93%	220.78	-44.57%	220.86	-44.63%	157.8	-3.20%	238.43	-56.13%
2020-04-01	158.95	232.3	-48.01%	278.85	-76.27%	220.97	-40.79%	219.79	-40.04%	220.81	-40.69%	220.93	-40.77%	157.57	-0.40%	204.88	-30.41%
2020-05-01	188.07	228.58	-37.64%	277.88	-67.33%	220.92	-33.03%	219.79	-32.35%	220.84	-32.98%	220.97	-33.06%	142.17	14.39%	172.58	-3.92%
2020-06-01	183.08	228.67	-24.90%	279.15	-52.47%	220.94	-20.68%	219.79	-20.05%	220.86	-20.63%	220.92	-20.67%	131.54	28.15%	168.99	14.25%

6. 각 예측 모델을 통한 유가와 철강 가격 예측 종합

Table 19에 각 예측 모델을 통한 유가와 철강 가격 예측 종합하여 나타내었다. 예측모델 총 7개 중 1, 7과 8의 예측모델을 앞 선 내용에서 기 설명한 자료이며 2~6번 예측모델은 모델의 다양성을 위해 추가한 내용이다.

자동회귀(AR)모델도 5장에서 기 분석한 바 있으나 물가나 주가 등에서는 불규칙하게 발생하는 이벤트 등과 같은, 에러항목

과 유사한 영향을 미치는 많은 고려해야 할 인자들은, AR모델의 특징인 전일 가격과의 선형회귀와의 상관성과 관련짓기 난해하여 이동평균(MA)유형의 모델이 상대적으로 간편하고 물가나 주가 등에서 많이 사용되고 있어 관련 유형의 모델을 추가하였다.

총 7개의 예측 모델로 1번은 단순평균모델, 2번은 선형회귀 모델, 3번은 이동평균모델(MA), 4번은 가중이동평균모델(WMA), 5번은 지수평활모델(ESM), 6번은 치수조정 지수평활 모델(TAESM), 7번은 인공지능모델(Prophet) 그리고 8번은 계절성을 고려한 이동평균모델이다. 4번의 모델에서 사용한 가중

치는 t-1에서 t-3까지 각각 0.43, 0.33, 0.23을 사용하였고 6 번의 치수조정 지수평활모델에서는 지수평활계수와 감쇠인수를 각각 0.2, 0.8을 사용하였다.

각 모델의 예측을 통한 평균절대예러(MAE)는 유가예측에서 예측 모델 순으로 각각 35.81%, 38.15%, 43.29%, 42.77%, 41.76%, 43.21%, 31.36% 그리고 34.21%로 3~6번 모델 군들의 오차가 40%대로 제일 크고 단순평균과 1차 선형모델과 같은 단순한 모델 측의 1~2번 모델이 오차가 30% 후반대로 줄어들었고 7~8번 모델 30% 전반대로 가장 우수하였다.

각 모델의 예측을 통한 평균절대예러(MAE)는 철강가격 예측에서는 예측 모델 순으로 각각 19.26%, 39.47%, 16.75%, 16.45%, 16.70%, 16.65%, 18.79% 그리고 14.84%로 단순평균과 1차 선형모델과 같은 단순한 모델 측의 1~2번 모델이 오차가 20% 이상으로 가장 높고 3~6번 모델 군들의 오차가 16~17% 범위이며 7~8번 모델은 14~18% 범위를 보여주고 있다.

각 예측모델의 예측분석을 통한 자료는 앞으로 고유의 툴을 개발하거나 평가 시스템 구축하는 좋은 연구기초자료가 될 것으로 사료된다.

7. 결론

페이스북에서 다양한 데이터를 예측하는, 많은 반복을 통해 개발한 모델예측 인공지능 시스템인 Prophet을 개인용 컴퓨터와 구글 클라우드에 설치하여 보았고 조선 산업과 관련된 가장 주요한 물가들이라고 판단되는 유가와 철강가격을 예측하였다.

주요물가 대상은 조선 산업의 주요 물가인 유가와 철강가격을 예측하기 위한 D/B를 준비하여 데이터의 오차를 가능한 순화키 위해 일일평균가를 입력 자료로 사용하였다.

Prophet 예측결과와 비교 분석을 위해 기존 예측 법을 이용하여 동일한 데이터로 예측을 수행하였다.

먼저 입력 데이터의 이동평균선 N=3~5의 오차를 검토하여 보았고 유가와 철강가격 공히 N=3이 N=4와 N=5에 비해 3 가지 오차 분석(MAE(Mean absolute error), MSE(Mean squared Error), MAPE(Mean absolute percentage error))에서 오차가 이동평균선 N=3 이 공통되게 모두 작아, 예측 작업에 이동평균선 N=3을 적극 활용하였다.

두 번째로 기존 이론을 통한 데이터의 선형 예측만을 고려한 경우, 유가에서는 미세한 증가를 보이나 실제 값은 급락현상 보여주고, 철강가격에서는 예측치가 뚜렷한 상승세를 보이나 실제는 급락현상 보이고 있어 기존 데이터를 이용한 선형 예측만으로는 부족한 면이 크다고 사료된다.

세 번째로, 선형예측의 한계를 파악하고 계절성을 추가한 예측을 통해 유가 및 철강가의 예측은 어느 정도 예측 추세와 실제 변동가격과 공히 유사한 추세를 보여주고 있음을 확인하였다. 또한 마지막으로 예측 값에 0.98의 신뢰구간을 설정하여 Prophet 결과와 같은 신뢰구간 형식의 유가와 철강가격의 상하

한 예측범위를 추가 설정하여 보았다.

간단한 사용결과였지만 Prophet은 내부적으로 전반적인 추세패턴, 주기적 성분의 분류를 제시하고 이를 통한 예측으로 예측의 이해도를 높일 수 있다고 사료된다.

주요 물가들의 흐름이 특별히 변동성이 매우 큰 구간이 발생하여, 예측구간의 큰 오차 발생이 불가피하였다. 비록 가장 큰 오차항목이 시계열분석에서 발생하였지만, 간편하게 정도 있는 예측 값을 준다고 알려진 prophet 과 비교하여 본 논문의 마지막 시계열분석법도 전 오차분석 항목에서 나쁘지 않은 결과를 얻을 수 있었다.

다양한 시계열분석과 Prophet 예측결과와 실제 월평균 데이터의 비교결과 전반적으로 기본 패턴은 유지되므로, 추가 등에도 전문가적 시계열의 깊은 지식적 접근을 통한 수익 창출의 접근 보다는 가벼운 느낌의 손실 가능성을 줄여주는 예측 도구로 나름 적합성은 보이고 있다고 사료된다.

8가지의 다양한 예측모델을 사용한 예측분석을 통한 예측자료는 앞으로 고유의 툴을 개발하거나 평가 시스템 구축하는 좋은 연구기초자료가 될 것으로 사료된다.

본 연구는 초기단계 연구로 인공지능프로그램의 기본세팅과 기존 시계열 예측이론을 통한 조선 산업의 주요물가의 예측 결과를 비교하는 차원이며 차후 Prophet의 다양한 하이퍼파라미터와 이벤트성 효과를, 비록 내부 알고리즘이 공개되진 않았지만 추가 연구하여 많은 경험이 축적되면 예측 성능을 더 향상시킬 수 있는 여지는 크다고 사료된다.

후기

이 연구는 논문발행일 기준으로 2021학년도 한라대학교 자율형 학술연구비의 지원에 의해 연구되었기에 이에 관련한 관계자 여러분께 감사드립니다.

References

- Andreas, M., & Seragado, M., 2017. *Introductory to Machine Learning with Python*, Hanbit Media.
- Ham, J.H., 2018. Design system of doubler plate of ship plate members under various in-plane and out-of-plane loads. *Journal of the Society of Naval Architects of Korea*, 55(6), pp.521-526.
- Ham, J.H. 2019a. A study on the prediction of major prices in the shipbuilding industry using artificial intelligence. *Autumn Conference of the Society of Naval Architects of Korea*, Kyungju, pp.167-171.
- Ham, J.H. 2019b. A study on the prediction of stock price of large shipyards using deep learning. *Autumn Conference of the Society of Naval Architects of Korea*, Kyungju, pp.172-176.
- Ham, J.H., 2021. Development of hybrid prototype dual load

- cell structure. *Journal of the Society of Naval Architects of Korea*, 57(6), pp.373–380.
- Ham, J.H. 2021a. A study on the prediction of major prices in the shipbuilding industry using artificial intelligence – Comparison analysis with existing prediction methods. *Spring Conference of the Society of Naval Architects of Korea*, Incheon.
- Ham, J.H. 2021b. A study on the prediction of stock price of large shipyards using deep learning– Comparison analysis with existing prediction methods. *Spring Conference of the Society of Naval Architects of Korea*, Incheon.
- Harvey, A. & Peters, S., 1990. Estimation procedures for structural time series models. *Journal of Forecasting*, 9, pp.89–108.
- Harvey, A. C. & Shephard, N., 1993. Structural time series models, in G. Maddala, C. Rao & H. Vinod, eds, ‘Handbook of Statistics’, Vol. 11, Elsevier, chapter 10, pp.261–302.
- Hutchinson, G. E. 1978. *An introduction to population ecology*, Yale university press, Publication date: September 10.
- Jang, E. A., 2020. *Introducing the time series prediction package prophet*, Hyper connect Tech Blog, <https://hyperconnect.github.io/2020/03/09/prophet-package.html> [Accessed 02 June 2021].
- Kujira, H., 2017. *Introduction to machine running, deep learning practical development using python*, Wikibooks.
- Prophet, 2019. *Forecasting at Scale*, <https://facebook.github.io/prophet> [Accessed 01 June 2021].



함주혁