

<http://dx.doi.org/10.17703/JCCT.2023.9.6.251>

JCCT 2023-11-31

세계 경제 지표를 활용한 머신러닝 기반 국제 경유 가격 예측 모델 개발

International Diesel Price Prediction Model based on Machine Learning with Global Economic Indicators

최아린*, 박민서**

A-Rin Choi*, Min Seo Park**

요약 국제 경유 가격은 산업, 교통 및 에너지 생산과 같은 여러 분야에서 중요한 역할을 수행하며, 세계 경제와 국제 무역에도 큰 영향을 미친다. 특히, 국제 경유 가격의 상승은 소비자에게 부담을 주고 인플레이션의 원인이 될 수 있다. 그러나 기존 연구들은 주로 휘발유에 초점을 맞추어 진행되었다. 따라서 본 연구는 국제 경유 가격 예측 모델을 제안하고자 한다. 이를 위해 다양한 세계 경제 지표들을 활용하여 머신러닝 방법론 중 하나인 선형 회귀 모델로 학습한다. 해당 모델은 세계 경제 지표들과 국제 경유 가격 간의 관계를 명확하게 파악함과 동시에 높은 정확도로 예측한다. 이는 시장 변화를 비롯한 전반적인 경제 흐름 파악에 도움이 될 것으로 기대된다.

주요어 : 세계 경제 지표, 선형 회귀 모델, 국제 경유, 머신러닝

Abstract International diesel prices play a crucial role in various sectors such as industry, transportation, and energy production, exerting a significant impact on the global economy and international trade. In particular, an increase in international diesel prices can burden consumers and potentially lead to inflation. However, previous studies have primarily focused on gasoline. Therefore, this study aims to propose an international diesel price prediction model. To achieve this goal, we utilize various global economic indicators and train a linear regression model, which is one of the machine learning methodologies. This model clearly identifies the relationship between global economic indicators and international diesel prices while providing highly accurate predictions. It is expected to aid in understanding overall economic trends including market changes.

Key words : Economic indicators, Linear Regression, International Diesel, Machine Learning

1. 서론

세계 경제의 흐름을 이해하는데 중요한 요소 중 하나가 에너지 자원의 생산과 소비다. 산업, 교통, 에너지 생산

등 다양한 분야에서 필수적으로 사용되는 에너지 자원의 가격 변동은 세계 경제 상황에 직접적인 영향을 미친다 [1][2][3]. 특히, 국제 무역 및 수출에 주로 사용되는 경유

*준회원, 서울여자대학교 데이터사이언스전공 석사과정

**정회원, 서울여자대학교 데이터사이언스학과 조교수

접수일: 2023년 10월 3일, 수정완료일: 2023년 10월 20일

게재확정일: 2023년 11월 5일

Received: October 3, 2023 / Revised: October 20, 2023

Accepted: November 5, 2023

**Corresponding Author: mpark@swu.ac.kr

Dept. of Data Science, Seoul Women's Univ, Korea

의 가격은 각 나라의 내부 시장뿐만 아니라 세계 시장에도 영향을 미친다. 실제 과거 몇 년간 발생한 여러 사회·경제적 위기 사례들은 국제 경유 가격과 밀접하게 관련되어 있다[4]. 그러나 기존 연구들은 대부분 휘발유에 초점을 맞추고 있다. 특히 휘발유 가격과 경제 지표 간의 상관관계를 중심으로 석유 소비 흐름을 파악하는 연구가 주를 이루고 있다. 기존 연구들은 다양한 요인들의 결합으로 형성되는 국제 경유 가격 변동을 충분히 고려하지 못하는 한계가 있다. 대부분 특정 기간 내에서의 시간 범위에 초점을 맞추었기 때문에 장기적인 트렌드나 영향력 분석에 있어 한계가 있다[5].

따라서 본 연구에서는 경유의 중요성을 인식하고, 그 가격 변동에 영향을 미치는 다양한 요인들을 고려하여 머신러닝 기반 국제 경유 가격 예측 모델을 제안한다.

본 논문의 구성은 다음과 같다. 제2장에서는 국제 원유 가격과 관련된 선행 연구에 대해 살펴본다. 제3장에서는 월평균 세계 경제 지표를 활용한 머신러닝 기반 국제 경유 가격 예측 모델을 제안하고 그 성능을 평가한다. 마지막으로 제4장에서는 결론을 서술하고, 향후 연구 방향에 대해 논한다.

II. 관련 연구

본 장에서는 국제 원유 가격과 관련된 연구들을 살펴본다.

표수진[6]은 세계 경제 변화에 영향을 미치는 7가지 경제 지표를 이용하여 각각의 경제 지표와 국제 원유 가격의 일반화된 표준으로 공인받은 WTI(West Texas Intermediate)유, 두바이유, Brent유 가격과의 상관관계를 단순 회귀 모델을 통해 분석하였다. 그러나 이 연구는 7가지의 독립변수를 동시에 고려하지 않고, 각각의 독립 변수와 원유 가격과의 상관관계를 분석하였다. 송경재[7]는 시계열 분석을 통해 WTI(West Texas Intermediate)유 가격을 예측하는 모델을 제안하였다. 12분기 중 3분기를 제외한 영역의 변화 추이를 분석하여 세계 경제 지표와 WTI(West Texas Intermediate)유 가격의 변화를 예측하였다. 그러나 해당 연구 역시 한 가지 개별 지표와 원유 가격과의 관계를 분석하는 데 주력하여 국제 경유 가격과 여러 경제 지표와의 관계를 분석하고 예측하는 데는 한계가 있다.

따라서 본 논문은 기존 선행 연구에서 제시한 방법을

참고하고 다양한 세계 경제 지표를 활용하여 맞추어 국제 경유 가격에 영향을 미치는 요인을 찾고, 이를 기반으로 국제 경유 가격 예측 모델을 제안한다.

III. 연구 방법

본 연구는 머신러닝을 적용하여 국제 경유 가격 예측 모델을 제안한다. 이를 위해 데이터 수집, 데이터 전처리, 데이터 표준화, 모델링의 프로세스를 수행한다.

1. 데이터 수집

첫 번째 단계는 관련 데이터를 수집하고 탐색하는 것이다. 국제 경유 가격이 세계 경제 지표와 밀접한 연관이 있기 때문에, 다양한 경제 지표를 수집한다. 한국석유공사 유가정보 서비스(오피넷, Opinet)[8]에서 2012년 12월부터 2022년 12월까지의 월별 평균 국제 경유 가격을 수집하였다. 또한, 실시간 주식시장 데이터를 무료로 제공하는 인베스팅닷컴(Investing.com)[9]에서 동일한 기간의 S&P 500(Standard & Poor's 500 Stock), 실업률, WTI(West Texas Intermediate)유, 산업생산지수 (Industrial Product Index, IPI), Dollar Index 등 주요 경제 지표를 함께 수집하였다.

S&P 500은 미국 Standard & Poor's 사가 발표하는 주가지수로서 기업규모와 유동성 등을 고려하여 선정된 보통주 500종목을 대상으로 작성된다. 이 지수는 글로벌 시장에 대한 중요한 정보를 제공한다, S&P 500 내 포함된 회사들은 원유 시장에서 큰 비중을 차지하고 있기에 원유 가격과 밀접한 연관이 있다 [10][11].

실업률은 일할 능력과 취업 의사가 있는 사람 중 일자리가 없는 사람이 차지하는 비율로서 UN(United Nations)은 실업률 상승이 각국의 생산성과 구매력 저하로 이어져 세계 경제 성장에 부정적인 영향을 미친다고 보고하였다[12].

WTI(West Texas Intermediate)유는 북미에서 생산되며 세계적으로 널리 인정받는 원유의 표준 가격 지표 중 하나이다. 이 WTI유 가격은 세계 시장 동향에 민감하게 반응하는 경향이 있어, 세계 경제 및 에너지 시장의 변동을 예측하는 데 직접적인 요소로 고려된다[13].

산업생산지수(IPI)는 모든 생산 활동의 기초를 나타내는 지표로서 산업별 생산 활동 수준의 변화를 파악할 수 있다[14]. 제조업체, 광산 및 가스 산업 등을 대표하는 약

300개 업체를 선택하여 이들이 생산한 인덱스 값을 총생산량으로 나눈 후 해당 비중으로 인덱스 값을 조정하여 사용한다.

마지막으로 Dollar Index는 달러 외환 시장에서 거래되는 달러 강도를 나타내주는 지수이다. 원유는 국제 시장에서 주로 미국 달러로 거래되기 때문에, 달러 강도가 원유 가격에 큰 영향을 미친다[15].

2. 데이터 전처리

수집된 데이터를 그대로 모델링에 사용하는 것은 정확도에 문제를 야기시킬 수 있다. 원시 데이터에는 결측치나 이상치가 포함되어 있을 수 있고, 각 데이터의 척도가 다를 수 있다. 이러한 데이터는 모델링 결과를 왜곡시킬 수 있다. 이러한 문제들을 방지하기 위해 수집된 데이터를 전처리한다. 전처리 방법으로 본 연구에서는 S&P 500 (Standard & Poor's 500 Stock), 실업률, WTI (West Texas Intermediate) 선물 가격, 산업생산지수 (Industrial Product Index, IPI), Dollar Index 데이터를 독립변수로 사용하고 해당 변수의 특징에 따라 시차를 반영하고, 각 변수의 분산 팽창지수 (Variance Inflation Factor, VIF)를 적용하여 전처리하였다.

표 1은 독립변수의 시차 반영 결과이다. t는 국제 경유 가격을 예측하고자 하는 월을 의미한다. S&P 500과 WTI 선물 가격은 값의 변화가 실시간으로 반영되므로 t값 조정 없이 그대로 사용하였다. 반면, 실업률, IPI, Dollar Index의 경우, 1개월의 시차가 존재하여 t-1로 조정하여 반영하였다.

표 1. 시점을 반영한 독립변수

Table 1. Independent variables reflecting the point in time
 * t: 2012년 12월 ~ 2022년 12월의 월

독립변수	S&P 500	실업률	WTI 선물가격	IPI	Dollar Index
시점	t	t-1	t	t-1	t-1

표 2는 각 변수의 분산팽창지수(VIF)를 보여준다. 분산팽창지수(VIF)는 독립변수들 간의 독립성을 측정하는 지표로서, 다중공선성 문제를 진단하는 데에 주로 사용된다. 다중공선성이란 회귀 분석에서 사용되는 독립변수들 사이에 높은 상관관계가 존재할 때 발생하는 현상으로, 다중공선성이 높은 경우 회귀 분석 결과의 정확도를 떨어뜨릴 수 있다.

서로 높은 상관관계를 보이는 독립변수들은 회귀분석 결과에 오류를 일으킬 수 있기에[16]. 분산팽창지수

(VIF) 10 이하를 기준으로 데이터를 전처리하였다.

3. 데이터 표준화

각 변수는 서로 다른 척도와 범위를 가질 수 있기 때문에 데이터 분석 및 모델링 시 결과가 편향될 수 있다. 데이터 편향 문제를 해결하기 위해 종속변수를 제외한 독립변수에 대해 표준화 과정을 진행하였다. 표준화란

표 2. 각 변수의 분산팽창지수

Table 2. VIF(Variance inflation factor) for each variables

독립변수	VIF Factor
S&P 500	2.5
전월 실업률	1.3
WTI 선물 가격	3.5
전월 IPI	5.2
전월 Dollar Index	3.2

각 변수의 평균을 0, 분산과 표준편차를 1로 만드는 변환 과정을 의미한다[17]. 이 과정을 통해 서로 다른 척도와 범위를 가진 변수들이 동일한 척도로 이상치의 영향력을 줄일 수 있어, 모델의 안정성과 예측 정확도가 개선될 수 있다. 본 연구에서는 python 프로그래밍 언어의 Scikit-learn 라이브러리에 내장된 StandardScaler 함수를 사용하여 데이터를 표준화한다.

4. 머신러닝 기반 국제 경유 가격 예측 모델 개발

머신러닝 중 선형 회귀는 종속변수와 독립변수 간의 인과 관계를 모델링하는 방법론으로 독립변수의 변화에 따른 종속변수의 변화를 추정하고 미래 값을 예측하는데 효과적이다.[18]. 본 연구 역시 다양한 세계 경제 지표들을 사용하여 국제 경유 가격을 관계를 살펴보고 원인을 밝히는데 목적이 있기에 선형 회귀 모델로 연구를 진행하였다. 전처리 된 세계 경제 지표들을 선형 회귀모델의 독립변수로 사용하여 국제 경유 가격 변동에 미치는 요인을 분석하고, 이를 바탕으로 예측 모델을 제안한다. 해당 결과는 표 3과 같다.

표 3. 국제 경유 가격 예측을 위한 제안된 회귀식

Table 3. Proposed multiple regression model

$$y = 207.1819(X_1) + 92.1081(X_2) + 62.5033(X_3) + 24.0398(X_4) - 15.6400(X_5)$$

주) Y=국제 경유(0.001%)

X₁=WTI선물 가격(t)

X₂=Dollar Index(t-1)

X₃=IPI(t)

X₄=S&P 500

X₅=실업률(t-1)

* t: 2012년 12월 ~ 2022년 12월의 월

표 3는 국제 경유 가격 예측을 위한 도출된 선형 회귀 식이다. WTI 선물 가격, 전월 Dollar Index, S&P 500, 전월 산업생산지수(IPI)는 국제 경유 가격과 양의 상관관계를 보이는 반면, 전월 실업률은 국제 경유 가격과 음의 상관관계를 보인다. 실업률은 다른 지표들과 달리 국제 경유 가격 상승을 억제하는 역할을 한다.

제안한 모델은 무작위로 나는 80%의 훈련 데이터와 20%의 테스트 데이터를 활용하여 학습되었다. 모델의 성능 평가를 위해, 결정계수(R-squared), 조정 결정 계수(Adjusted R-squared), 정확도(Accuracy), 평균 제곱근 오차 (Root Mean Squared Errors, RMSE)등을 사용하였다. 표 4는 해당 결과를 보여준다.

결정계수(R-squared)는 회귀 분석에서 모델의 적합도를 평가하는 척도로 모델의 설명력을 의미한다. 값이 1에 가까울수록 모델이 데이터를 잘 설명하는 것으로 평가될 수 있다. 훈련 데이터셋(training set)의 결정계수는 0.9226, 테스트 데이터셋(test set)에 대한 결정계수는 0.9110으로 국제 경유 가격의 예측에 있어서 제안된 모델이 상당히 높은 설명력을 가지고 있다는 것을 의미한다.

조정 결정 계수(Adjusted R-squared)는 독립변수의 수가 증가하면서 표본 크기에 따라 발생할 수 있는 결정계수(R-squared)의 과적합 현상을 보정할 수 있는 방법으로 이 값 역시 높을 수록 모델이 데이터를 잘 설명한다는 것을 의미한다. 본 연구의 훈련 데이터셋(training set)에서의 조정 결정계수는 0.9180, 테스트 데이터셋(test set)에서는 0.8876으로 모델이 국제 경유 가격을 잘 설명하고 있음을 보여준다.

정확도(Accuracy) 테스트에서도, 테스트 데이터셋(test set)이 90.57%로 상당히 높은 정확도로 보였다.

평균 제곱근 오차(Root Mean Squared Errors, RMSE)는 회귀 모델의 예측 성능을 오차 기반으로 평가하는 지표이다. 이는 실제 값과 예측한 값 간의 차이를 수치화하여 나타낸다. 평균 제곱근 오차는 작을수록 실제 값과 예측 값 사이의 차이가 작다는 것으로 해석되는 것은 오차가 작다는 것을 의미하여 모델의 예측 정확도가 높다고 판단할 수 있다. 본 연구에서는 훈련 데이터셋(training set)에 대한 평균 제곱근 오차(RMSE)는 53.2219, 테스트 데이터셋(test set)은 87.2675로 나타났다.

표 4. 제안된 국제 경유 가격 예측 모델의 평가지표
Table 4. The evaluation value for the proposed model

Dataset	R-Square	Adjusted R-squared	Accuracy	RMSE
Training	0.9226	0.9180	92.37%	59.2219
Test	0.9110	0.8876	90.57%	87.2675

제안된 예측 모델의 독립변수들의 유의성을 검증하기 위해 p-value를 활용하여 한번 더 검증하였다. 일반적으로 p-value가 0.05 미만일 경우 해당 변수가 유의미하다고 판단한다. 표 5는 제안한 선형 회귀 모델의 각 독립변수에 대한 p-value를 보여준다. 모든 변수들이 유의함을 알 수 있다.

표 5. P>|t| 값
Table 5. P>|t| Parameters

독립변수	p-value
S&P 500	0.000
전월 실업률	0.022
WTI 선물 가격	0.021
전월 IPI	0.000
전월 Dallar Index	0.000

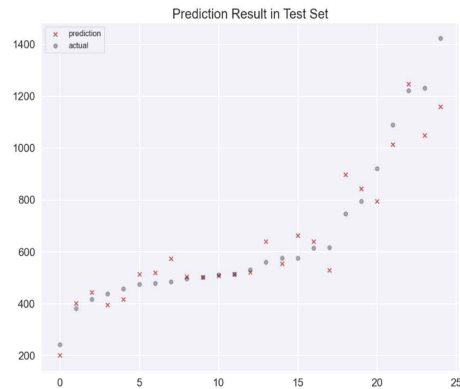


그림 1. 선형 회귀 모델 예측 결과
Figure 1. Result of Linear regression prediction

그림 1은 실제 값과 예측 값을 그래프로 나타낸 것이다. x축은 테스트 데이터 셋의 실제 값 기준으로 정렬한 index값이며, y축은 해당 index에 대한 국제 경유 가격을 의미한다. 빨간색 'X'기호는 예측 값을 나타내며, 회색 'O'기호는 실제 값으로 예측 값과 실제 값 사이에 큰 차이가 없음을 알 수 있다. 그러나, 일부 구간에서는 차이가 발생하였다. 상세 분석 결과, 해당 구간들은 분쟁 및 비정상적인 이벤트가 발생한 구간으로 경기가 악화되며 S&P 500지수가 급격하게 하락된 구간으로 판단된다.

IV. 결 론

본 연구는 WTI 선물 가격, Dollar Index, 산업생산지수(IPI), S&P 500, 실업률 등의 세계 경제 지표를 활용하여 국제 경유 가격을 예측하는 선형 회귀 모델을 제안하였다.

성능 평가 결과, 결정계수와 조정 결정 계수는 1에 가까운 값을 나타내며, 테스트 데이터셋(test dataset)에 대한 정확도 역시 1과 가까운 값을 보였다. 실제 값과 예측 값 사이의 차이를 나타내는 평균 제곱근 오차 역시 실제 값과 예측 값 사이의 차이가 적으며 유의미한 결과값을 보였다. 마지막으로 각 독립 변수들의 유의성을 검증하기 위해 p-value 검증을 하였으며 이 역시 각 변수들이 국제 경유 가격 예측에 유의미한 영향력을 갖고 있음을 확인할 수 있었다.

그러나 본 연구에는 한계점이 있다. 사용된 데이터가 국제 경유 가격이 공개되기 시작한 2012년 12월부터의 데이터로 구성되어 있어, 이전의 과거 데이터와 관련된 분석에는 한계가 있을 수 있다. 또한 실제 현실에서는 지정학적 위험, 정책 변화, 기술 발전 등 다양한 외부 요인들이 국제 경유 가격에 영향을 줄 수 있으나 본 연구에서는 세계 경제 지표만을 고려하였다. 향후, 세계 경제 지표만이 아닌 다른 유의미한 사회적 요인들도 함께 고려하여 모델을 개발하면 보다 정확한 국제 경유 가격 예측 모델을 개발할 수 있을 것으로 기대한다.

References

[1] D. I. Stern, "The role of energy in economic growth," *Annals of the New York Academy of Sciences*, Vol. 1219, pp. 26-51, 2011. DOI:https://doi.org/10.1111/j.1749-6632.2010.05921.x

[2] R. H., Rasche, J. A, Tatom, "Energy resources and potential GNP," *Review*, 1977.

[3] A. K. Singh, N. Kumar, Amardeep and P. Kumar, "A Comparative Study on the Performance and Emission Analysis of a Dual Fuelled Diesel Engine with Karanja Biodiesel and Natural Gas," *IJACT*, Vol. 4, No. 1, pp. 10-18, 2016. DOI:https://doi.org/10.17703/IJACT.2016.4.1.10

[4] Z. Wadud, "Diesel demand in the road freight sector in the UK: Estimates for different vehicle types," in *Applied Energy*, vol. 165, pp. 849-857,

2016. DOI: https://doi.org/10.1016/j.apenergy.2015.12.102

[5] R. A. Ratti and J. L. Vespignani, "Oil prices and global factor macroeconomic variables," in *Energy Economics*, vol. 59, pp. 198-212, 2016. DOI:https://doi.org/10.1016/j.eneco.2016.06.002

[6] S. J. Pyo, M. Y. Jung, M. S. Cho and K. Y. Lee, "Interrelationship analysis between crude oil price and the world economic indices," *Korean Institute Of Industrial Engineers The Fall Conference*, pp. 412-416, 2012.

[7] K. J. Song and H. M. Yang, "A Study on the Nymex WTI Prices Forecasting Using Time Series Analysis," *Journal of the Korean Official Statistics*, Vol. 10, No. 1, pp. 4-4, 2005.

[8] Korean National Oil Corporation, "Opinet," Korean National Oil Corporation. [Online]. Available: <http://www.opinet.co.kr>. [Accessed: 14- 09- 2023].

[9] Investing.com, "Investing.com," Fusion Media Ltd. [Online]. Available: <https://www.investing.com>. [Accessed: 14- 09- 2023].

[10] Y. I. So, J. M. Ko, and W. K. Choi, "How Does Economic News Affect S&P 500 Index Futures?," *The Korean journal of Financial management*, Vol. 13, No. 1, pp. 341-357, 1996.

[11] O. Ratanapakorn, S. C. Shama, "Dynamic analysis between the US stock returns and the macroeconomic variables," *Applied Financial Economics*, Vol. 17, No. 5, pp. 369-377. 2007. DOI:https://doi.org/10.1080/09603100600638944

[12] International Labour Office, "ILO Skills for Employment Initiative: A Synthesis of Lessons Learned and Ways Forward," *International Labour Organization*, 2010.

[13] W. C. Yun, "Comparison of Price Predictive Ability between Futures Market and Expert System for WTI Crude Oil Price," *Environmental and Resource Economics Review*, Vol. 14, No. 1, pp. 201-222, 2005.

[14] J. H. Stock and M. W. Watson, "Forecasting inflation," *Journal of Monetary Economics*, vol. 44, no. 2, pp. 293-335, 1999. DOI:https://doi.org/10.1016/S0304-3932(99)00027-6

[15] M. B. Lee, "Energy Focus," *Korea Energy Economics Institute*, No. 1, Vol. 9, pp. 52-55, October 2004.

[16] R. M. O'brien, "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality & Quantity*, vol. 41, no. 5, pp. 673-690, October 2007.

[17] M. S. Gal and D. L. Rubinfeld, "Data

- standardization," *NYUL Rev.*, vol. 94, pp. 737, 2019. DOI:<http://dx.doi.org/10.2139/ssrn.3326377>
- [18]S. J. Kim, "Analysis of Apple Colors and Sugar Contents Using Linear Regression," *JCCT*, Vol. 8, No. 1, pp. 201 - 207, 2022. DOI:<https://doi.org/10.17703/JCCT.2022.8.1.201>