

기계학습 방법을 활용한 가격 예측 모형 개발:  
공군 항공유 구매 사례를 기반으로  
Developing a Machine Learning-Based Model  
for Price Forecasting:  
A Case Study on ROKAF Jet Fuel Procurement

임세환(주저자) · 민순홍(교신저자) · 최경환(공저자)  
Sehwan Lim(First Author) · Soonhong Min(Corresponding Author) · Kyunghwan Choi(Co-Author)

연세대학교 Yonsei University School of Business(limseh@yonsei.ac.kr)  
연세대학교 Yonsei University School of Business(smminscm@yonsei.ac.kr)  
방위사업청 Defense Acquisition Program Administration(borita@hanmail.net)

공군은 매년 수천억 원 규모의 항공유를 구매 중이며, 급격한 항공유 가격 변동은 작전 준비태세에 리스크 요인으로 작용하고 있다. 정확한 가격 예측은 유류 재고관리를 최적화하는 데 중요하지만, 기존 연구는 구매기관에서 바로 적용이 제한되는 거시적 수준의 경제지표에 초점을 맞추는 경우가 많았다. 본 연구에서는 구매기관 관점에서 항공유 가격 예측 정확도를 향상하기 위한 기계학습 방법의 효과를 평가한다. 이때, 예측 정확도 향상을 위해 공급망, Google 검색 트렌드(Google Trends), 지정학적 요인과 전통적인 경제 요인을 통합한 기계학습 예측모형을 제안한다. 본 연구의 결과는 XGBoost 모형이 RMSE(Root Mean Squared Error)를 67%까지 줄여 가장 정확한 성과를 나타내었다. 따라서 공군을 비롯한 구매기관이 기계학습 가격 예측 모형을 채택하면 공급 가격관리 능력이 크게 향상될 수 있다. 또한, 본 연구 결과는 구매가 반복되고 가격 변동성이 큰 다른 제품 관리에도 효과적으로 적용할 수 있다.

주제어: 석유제품, 항공유, 기계학습 방법, 가격 예측, 구매관리

The Republic of Korea Air Force (ROKAF) spends hundreds of billions of Korean won annually on jet fuel, with price fluctuations posing a significant logistical and operational readiness challenge. Accurate price forecasting is crucial for optimizing fuel inventory management, yet existing research often focuses on macro-level economic indicators with limited practical application. This study investigates the potential of machine learning (ML) for enhancing jet fuel price forecasting accuracy. We propose ML models incorporating supply chain data, Google Trends data, and geopolitical factors alongside traditional economic variables for more realistic predictions. Our results demonstrate that a XGBoost model achieves the best performance, reducing Root Mean Squared Error (RMSE) by 67%. Adoption of this model by the ROKAF could significantly improve supply price management capabilities. Furthermore, the study's findings have broader applicability, potentially benefiting the inventory management of other commodities with significant price volatility and recurring purchases.

Keyword: commodities, jet fuel, machine learning, price forecasting, purchasing

## 1. 서론

2022년 러시아-우크라이나 전쟁의 영향으로 국제 유가는 8년 만에 최고치를 달성하였는데, 이는 연초 예측치 대비 50% 이상 상승한 가격으로 석유제품을 주요 에너지원으로 활용하는 기업과 기관들에 큰 비용부담으로 작용하였다(이인애, 2022). 특히 항공유 구매비가 기업 운영비용의 20~30%를 차지하는 항공사들의 수익을 크게 악화시키는 물론, 고유가 상황에서 지상 서비스를 제공하는 공항들이 항공유를 충분히 확보하지 못하여 재급유를 받지 못한 항공사들이 운항 일정을 조정하여 서비스 차질이 빚어지기도 하였다(Financial News, 2021). 대한민국 공군(이하 공군)의 경우에도 고유가 상황에서는 일부 장비의 운영을 일정 기간 동안 필수 임무수행으로만 제한할 수 있으며, 연료 구매 예산을 추가 확보하기 위한 노력이 심화되는 등 작전 및 운영상 제약이 존재한다.

항공유를 포함한 석유제품은 국가 간 분쟁, 산유국의 감산 정책, 자연재해 등으로 인해 공급 및 가격

변동 등의 공급사슬 위험(supply chain risk) 발생 빈도가 높은 반면, 이들 품목은 기업(기관) 자산운용의 필수 자원이므로 전략적 구매관리가 반드시 필요하다(cf. Kraljic, 1983). 이에 따라 공군을 포함한 대다수 구매기업(기관)은 공급업체와 1년 이상의 장기계약을 기반으로 안정적 공급 관계를 유지하여 공급 리스크는 완화할 수 있으나, 구매대금을 시장 가격 변동과 연계하여 정산함으로써 가격 변동 리스크 해소는 어렵다. 특히, 항공유를 포함한 유류 제품의 가격 변동은 예측하기 어려워서 가격 리스크를 더욱 가중시킨다. 구매기업(기관)은 기존의 가격 변동 예측성을 높이기 위해 주요 시장조사 기관에서 발표하는 예측 보고서를 활용하고 있지만, <Table 1>과 같이 단기 예측값 조차도 실제 시장가격과의 오차가 커서 제품 구매 상황과 조건에 맞는 가격 예측 방식이 필요하다.

공군과 공급업체 간 항공유 구매계약은 월별 항공유 구매대금 정산을 위해 매월 1일에 지난달의 일별 가격의 산술평균치를 월별 단가로 책정한다. 이를 참고하여 공군은 가격 예측에 시계열 방법을 사용 중인데, 이는 항공유 가격이 안정적인 경우에는 유용하다.

<Table 1> 미국 에너지정보청(EIA) 단기 에너지 전망 (월간/분기)

단위: \$/배럴(BBL)

지표	2022년 1월			2022년 2월		
	예측(F)	실제(P)	오차 (E=P-F)	예측(F)	실제(P)	오차 (E=P-F)
Brent	83.2	85.5	2.3	87.0	94.0	7.0
WTI	86.5	82.9	-3.6	90.0	91.6	1.6
지표	2022년 3월			2022년 2분기		
	예측(F)	실제(P)	오차 (E=P-F)	예측(F)	실제(P)	오차 (E=P-F)
Brent	85.0	112.4	27.4	83.6	111.9	28.3
WTI	88.0	108.2	20.2	87.0	108.5	21.5

\* EIA(Energy Information Administration) Short-term Energy Outlook

그러나, 현실에서는 유가가 수요와 공급, 재고수준, 지정학적 요인, 거시경제 지표 등 다양한 요인으로 인해 변동성이 높아서(Miao et al., 2017), 가격 예측의 정확도를 높이는 데 한계가 있다. 공군이 항공유를 적절한 양을 적절한 시기에 적절한 가격에 확보하기 위해서는 반드시 항공유 가격 예측 정확도를 높여야 한다.

항공유 가격 예측 정확도를 높이기 위해서는 가격에 영향을 미치는 다양한 요인들 간 복잡한 상관관계를 고려할 수 있는 새로운 방법론을 모색하여야 한다. 최근 공급사슬관리(supply chain management) 분야에서는 기계학습(machine-learning) 방법을 적용하여 자재 소요량, 주문량 등에 관한 수요 예측 능력 향상, 공정 프로세스 오류 탐지 등의 문제 해결에 활용하여 긍정적 효과를 얻고 있다는 연구 결과가 존재한다(Bojer & Meldgaard, 2021; Nagar et al., 2021; Younis et al., 2022). 문제상황과 연관된 변수들의 관계에 대한 반복 학습을 수행하는 기계학습 알고리즘 특성과 범용성을 고려하면, 기계학습 방법은 제품가격 예측에도 우수한 성과를 낼 수 있을 것으로 기대할 수 있다(Nagar et al., 2021).

유가 예측에 기계학습 방식을 적용한 다수의 연구가 존재한다. 그러나 이들 연구가 제시한 예측 모형들은 다양한 석유제품의 원재료인 원유의 가격을 예측 대상으로 하고 있어, 석유 완제품 가격 예측이 필요한 구매기업이 활용하기에는 한계가 있다. 이에 더하여 이들 기계학습 기반 예측 모형들은 원유의 수요·공급 지표 등 한정적 범위의 변수 유형을 활용하기 때문에, 기계학습에 투입하는 데이터의 유형이 다양할수록 최적의 결과를 산출하는 기계학습 방법의 장점을 충분히 살리지 못하는 단점이 있다.

본 연구의 목적은 제품 가격 변동 리스크 축소에 기여하고자 하는 것이다. 이를 위해 본 연구는 공군 항공

유 구매사례를 활용하여 기존 산술평균법 등 전통적 가격 예측 모형은 물론 기존 기계학습 예측 모형의 단점을 보완하는 새로운 기계학습 모형을 제시하려고 한다. 구체적으로 본 연구가 제시하는 기계학습 기반 가격 예측 모형은 기계학습 예측 모형에 생산·수요, 공급사슬, 지정학 관련 데이터 등 다양한 유형의 변수를 활용한다. 본 연구의 결과물인 기계학습 기반 가격 예측 모형은 다양한 가격 변동 요인을 고려한 점, 예측 정확도를 높인 점을 고려할 때, 공군의 항공유뿐만 아니라 다른 유류 품목(경유 등)에도 적용할 수 있을 것으로 기대한다. 나아가, 본 연구에서 제안하는 예측 모형 일부를 수정하여 사용한다면, 공군 등 국가 기관은 물론 민간기업에서도 시기별 구매 물량 결정 등의 의사결정에 사용할 수 있을 것이다.

본 연구에서 제시하는 ‘기계학습 방법 기반 항공유 가격 예측 모형’의 개발 절차는 다음과 같다. 먼저 공군 항공유 구매 사례를 검토(Ⅱ장)하고 기계학습 기반 가격 예측의 특성을 제시한다(Ⅲ장). 그리고 항공유 구매와 가격 특성에 부합하는 예측 모형을 정립한 후(Ⅳ장), 제안하는 기계학습 기반 예측 모형을 실제 데이터를 활용하여 검증하고(Ⅴ장), 본 연구의 결론과 시사점을 제시한다(Ⅵ,Ⅶ장).

## II. 공군 항공유 구매 현황

### 2.1 항공유의 중요성: 작전과 운영 측면

항공유는 전투기, 수송기 등 항공 자산의 작전 가용시간에 직접적인 영향을 준다. 급작스러운 고유가 현상이 발생하면 필요한 물량을 확보하기 위한 예산

이 부족해져 공군의 작전지속능력을 제한할 수 있다(중앙일보, 2006). 특히 공군이 사용하는 항공유 중 JP-8<sup>1)</sup>은 경유, 휘발유는 물론 민수용 항공유인 Jet A-1과 다르게 상용 시장에서 쉽게 구매할 수 없으므로 항공작전능력 유지를 위해서는 안정적인 공급을 확보하는 것이 필수적이다. 또한 공군의 항공유 구매비용은 공군 군수 예산의 20% 수준으로 매우 높아서, 원재료인 등유 가격 상승으로 항공유 가격이 상승하면 다른 품목의 구매에 영향을 줄 수 있다. 이처럼 항공유 가격 변동은 군의 작전은 물론 군수 운영에 심각한 제약을 초래할 수 있다.

## 2.2 항공유 구매 프로세스

공군의 항공유(JP-8, Jet-A1) 구매비용은 현재 연간 3,000억원 이상으로써 군수 단일 품목 구매비 지출액으로는 가장 큰데, 향후 KF-21, F-35A 등 신규 항공 자산이 속속 도입될 예정이므로 항공유 구매비용은 점차 증가할 수밖에 없다. 따라서 공군은 안정적인 항공유 확보를 위해 조달물류(inbound logistics) 운송 수단별(송유관, 트럭, 기차 등)로 1년 이상 장기 구매계약을 체결하고 일정 기간 단위로 정기 발주하는 시스템을 갖추고 있다. 항공유 구매 가격은 제품의 시장가격 변동에 연계하여 매월 조정하는 '월별 변동단가' 방식으로 산정하는데, 이번 달의 구매 단가는 지난달 현물시장 일일 배럴(barrel) 당 가격의 산술평균값으로 매월 1일 확정하는 방식이다. 예를 들면, 2024년 3월의 배럴 당 구매 단가는 2024년 2월 한 달 동안 매일 배럴당 가격의 산술평균값으로 3월 1일 확정한다. 항공유 가격은 항

공유의 주원료인 등유 가격에 환율을 곱하고 부원료인 각종 첨가제 가격, 그리고 운송비 등 부대비용을 포함한 금액이다. 원가 구성 요소 중 주원료인 등유는 가격 변동성이 매우 커서 변동비로 정산하지만, 부원료 및 부대비용은 변동성이 낮아서 고정비로 계약한다. 즉, 항공유 배럴 당 가격은 변동비인 등유 1배럴당 가격에 환율을 곱하고 고정비인 등유 1배럴당 첨가해야 하는 부재료 가격 및 부대비용을 합산한 금액이다(〈Table 2〉). 따라서, 항공유 가격 변동성은 실질적으로 등유 가격과 환율의 변동성을 의미한다. 이때, 주원료인 등유 가격의 결정은 시장조사 기관인 Platts사가 중동 및 싱가포르 현물시장의 일일 거래가격을 바탕으로 작성하는 Means of Platt's Survey(MOPS)의 가격을 기반으로 한다.

## 2.3 항공유 구매 가격 예측 프로세스의 문제점

매월 단가를 변경하는 계약구조를 가지고 있는 공군 사례에서의 항공유 가격 예측은 다음 달 가격에 대한 예측을 의미하며, 매월 10일에서 15일 경에 예측을 시행한다. 가격 예측에 사용하는 데이터는 매월 1일부터 예측 실시일까지의 MOPS에 고시된 등유 가격에 고정비를 더한 값이다. 예를 들면, 2024년 3월 구매 가격 예측은 2월 15일에 실시하는데, MOPS에 고시된 2월 1일부터 2월 14일까지의 현물시장 등유 배럴 당 확정 가격을 활용하여 전통적 시계열 예측방법(산술평균법, 단순이동평균법, 가중평균법, 지수평활법 등)으로 등유 가격을 예측한 후, 등유 1배럴당 예상되는 가격에 항공유 제조를 위해 등유 1배럴당 첨가하는 부재료 및 부대비용의 고정값을 더

1) JP-8(Jet Propellant-8): 등유(kerosene)를 주원료로 제작하는 군사용 제트 연료로서, 민간 항공기 연료인 Jet A-1에 상용 부식 방지제, 윤활향상제, 결빙방지제 등을 첨가하여 제조한다(제품 적용 규격: MIL-DTL-83133).

\* JP-8의 가격은 Jet A-1 가격에 군이 요구하는 상용 첨가제 가격을 더하여 산출되므로 두 제품 가격변동은 동일한 양상을 보인다.

〈Table 2〉 정부기관의 대량유류 공급계약 특수조건

<p><b>〔국방부 항공유 공급계약 특수조건(조달청)〕</b>                  제56조(물가 변동으로 인한 계약금액 조정) ① 물가 변동으로 인한 계약금액 조정은 제37조에도 불구하고 다음 사항을 적용한다.                  1. 매월 계약 유종별로 변동된 단가를 적용한다.                  4. 월별 변동단가는 '제2호'의 기준에 따라 산정된 전월 월간가격(매월 1일 ~ 말일 평균)과 '제3호'에 따라 산정된 부대비를 합하여 산정한다.                  5. 항공유 Opinet가 적용은 Opinet(www.opinet.co.kr)의 '국제유가', '석유제품'의 '등유'를 기준으로 한다.</p> <p><b>〔해양경찰청 항공유 공급계약 특수조건(조달청)〕</b>                  제 10 조 (계약금액의 조정)                  ① 계약금액의 조정은 낙찰가 중 제품대를 제외한 금액은 고정하고, 전월 MOPS(원)의 평균가를 적용한 제품에 고정금액을 더한 가격으로 매월 조정 적용한다.</p>
--

한 가격을 3월의 항공유 가격 예측치로 확정할 수 있다. 매월 15일까지의 누적된 가격 데이터를 위의 시계열 예측방식에 활용할 경우에는 일반적으로 MAE가 2,000원/BBL, RMSE가 3,000원/BBL 수준으로 산술평균법(Table 3-1)이 가장 정확하였으며, 계절적인 추세는 없는 것으로 판단할 수 있다(Table 3-2). 다음 달의 항공유 가격 예측을 불과 15일 전에 실시해야 하는 이유는 다음 달 변동 가격이 이번 달 현물시장 거래가로 결정되므로, 예측모형에 정확값을 많이 포함할수록 다음 달 가격 예측의 정확성이 높아지기 때문이다(〈Table 4〉).

공군의 항공유 가격 예측결과는 이번 달 구매량을 결정 요인으로 작용할 수 있다. 예를 들면, 이번 달을 2023년 2월로 가정하면, 2월 15일 3월 가격 예측치를 확정하기 전까지 (2월 1일~15일)는 일별 평균 사용량만큼 구매하다가, 3월 가격 예측치가 2월 확정가격보다 높으면, 2월 남은 기간(2월 16일~28일) 평균 구매량보다 많은 양을 구매하고, 3월 가격 예측치가 2월 확정가격 대비 낮다면, 2월 남은 기간 구매량을 평균 구매량 보다 적게 구매함으로써 3월에 낮은 가격으로 구매하는 양을 늘린다. 이러한 가

격 예측 방식의 단점은 다음 달 가격이 이번 달 가격보다 하락할 것으로 예측하여 적은 양을 구매하였으나 예측 오류가 확인되면, 월별 가격 차이에 의한 추가 예산이 소요될 수 있다는 점이다. 반대의 경우에는 사전에 유류를 고액에 구매함으로써 고가 매입에 따른 손실은 물론 불필요한 재고유지 비용까지 발생할 수 있다. 이에 더하여, 항공유 예측 가격에 기반한 구매량 결정 메커니즘은 한정된 공급업체의 납품 능력과 납품한 항공유를 수령하는 공군의 작업 능력 제약으로 인해 원하는 구매량을 채우지 못하는 경우도 발생하게 만들 수 있다.

앞에서 살펴본 전통적 가격 예측 방식은 (1) 매월 초기에는 정확도가 72% 수준에 불과하여(〈Table 4〉), 공군은 예측 정확도를 높이기 위해 (2) 가격 예측 시기를 매월 중반으로 늦춰야 하며, 늦어진 가격 예측 시기로 인해 (3) 매월 구매량 확정도 늦어지게 되고, 이는 다시 (4) 공급업체 또는 공군의 한정된 운영 능력으로 인한 필요 재고량 확보 실패 또는 과잉 재고 보유를 유발하는 등의 악순환(vicious cycle)을 초래할 가능성이 커진다. 이러한 악순환을 끊기 위해서는 항공유 가격 예측을 월초에 실시하면서도 예측 정

〈Table 3〉 현재 예측방식에 따른 판단 오류 및 오차수준

예측 오차 판단 기준			
- MAE (Mean Absolute Error) / 평균절대오차			
= $\frac{1}{n} \sum_{i=1}^n  P_i - F_i $ : 실제가격(P)와 예측가격(F) 차이의 절대값 평균 (1)			
- RMSE (Root Mean Squared Error) / 평균제곱근오차			
= $\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - F_i)^2}$ : 실제가격(P)와 예측가격(F) 차이 제곱 평균의 제곱근 (2)			
시계열 예측방법	내용	② MAE (원/BBL)	③ RMSE (원/BBL)
① 산술평균법	• 예측시점까지 누적된 최근 일일데이터를 산술평균	2,205	3,190
② 산술평균법 (계절지수 포함)	• 산술평균법에 계절지수(seasonal index)를 곱하여 적용 * 월 단위 계절지수는 '08~'19년까지의 데이터를 활용하여, '1월~12월'의 '각 월의 평균가격/전체 월의 평균가격'을 산출 * 위의 산출된 계절지수를 '20~'23년 데이터에 곱하여 오차수준을 측정 (월별 계절지수 분포: 0.94 ~ 1.04)	4,518	5,997
③ 단순이동평균법	• 가장 최근의 3일 동안의 자료를 평균하여 다음일자의 예측값으로 활용	2,389	3,426
④ 가중이동평균법	• 가장 최근의 5일 동안의 자료에 가중치를 부여하여 다음일자의 예측값으로 활용 * 가중치: 최근일자 기준으로 40%/30%/20%/10%	2,380	3,414
⑤ 지수평활법	• 지수평활계수(0~1 범위), 이전의 실제 가격과 예측 가격을 활용하여 다음일자의 예측 가격 산출 * 지수평활계수는 0.1 적용(0.1, 0.3, 0.7 중 정확도 가장 우수)	2,442	3,498

〈Table 4〉 현재 예측방식에 따른 판단 오류 및 오차수준(산술평균법 기준)

예측시기	예측자료	구매판단 적중률 ('08~'22년/180개월)	MAE (원/BBL)	RMSE (원/BBL)
매월 1일	최근 5일치 가격 평균	72.1% (129/179)	5,369	7,665
매월 11일	당월 1~9일 가격 평균	86.1% (155/180)	3,128	4,540
매월 16일	당월 1~15일 가격 평균	91.1% (164/180)	2,205	3,190
매월 23일	당월 1~22일 가격 평균	95.0% (171/180)	1,185	1,765
매월 26일	당월 1~25일 가격 평균	96.1% (173/180)	758	1,096

\* 구매판단 적중률 산출  
- 월 단위 항공유 구매판단 일치 건수(다음달의 항공유 예측단가가 이번달 단가 대비 상승·하락 여부를 판단한 결과를 기준으로, 이번달 항공유를 최대 혹은 최소 구매 의사결정함) / 대상기간의 총 건수(180개월)

확도를 높여서 공급 및 재고 관리, 예산상 제약을 최소화할 수 있는 새로운 가격 예측 모형을 개발하는 것이 시급하다. 본 연구는 이를 위해 기계학습(machine learning, ML) 기반 예측 모형을 제안한다.

### III. 기계학습 기반 예측

#### 3.1 기계학습 예측 모형 일반

기계학습 모형은 조직 운영에서 발생하는 문제와 관련된 대용량의 내·외부의 데이터 학습하여 예측 정확도를 개선할 수 있다. 기계학습 운용자는 사전에 프로그래밍이 된 알고리즘에 문제에 적절한 데이터셋(dataset)을 구조화하여 투입하고 기계학습 모형이 자동으로 학습을 수행하도록 하는데, 일반적으로 데이터가 누적될수록 모형의 예측 정확도가 높아진다. 공급사슬관리 분야에서는 기계학습을 재고 관리, 수요 및 생산계획 관리, 신제품 개발 등에 활용하고 있는데, 데이터 유형들의 엄격한 구성은 예측 정확도와 결과에 대한 해석 수준을 높이는데 도움이 된다(Nagar et al., 2021).

#### 3.2 기계학습 예측 유형 및 절차

기계학습 기반 예측 방법은 일반적으로 크게 지도 학습(supervised learning), 비지도학습(unsupervised learning), 강화학습(reinforced learning) 3가지로 분류한다. Chou et al.(2023)에 따르면 지도 학습은 사용자의 이해에 기반한 인과관계가 구조화된, 즉 정답(label)이 있는, 데이터셋을 통해 학습 모형을 훈련하고, 테스트 데이터(test data)를 활용

하여 정확도를 높이는 학습 방식이다. 비지도학습은 정답이 없는 데이터셋을 통해 학습 모형을 훈련하는 방식으로서 주어진 데이터셋에서 일정한 구조(pattern)를 찾으려 학습시키는 방식이다. 강화학습은 학습 모형이 특정 환경(state)에서 가능한 의사결정 대안 중 하나를 선택하면 이에 대한 보상(reward)을 제공하고, 이에 따라 학습 모형은 보상을 최대화하는 방향으로 의사결정을 보정하는 과정을 통해 의사결정 정확도를 높이는 방식이다.

본 연구의 목적은 공군 항공유 구입 과정의 예산상, 운영상 한계점을 극복할 수 있도록 항공유 가격 예측 정확도를 높일 수 있는 정교한 항공유 가격 예측 모형을 개발하는 것이다. 선행연구에서는 유류 가격 예측 모형에 원재료인 원유의 가격, 수요, 공급 지표, 항공유 관련 공급망 요인 등 다양한 변수 유형 중 일부를 사용하였다. 그러나 항공유를 비롯한 석유제품 가격은 다양한 경제적, 정치적 상황에 영향을 받기 때문에 제한된 변수만을 활용하는 단순 시계열 모형만으로는 예측 정확도를 높이는 데 한계가 있다. 따라서 본 연구는 기계학습 기반 가격예측 모형을 구축하고, 이 예측모형에 항공유 가격 결정에 영향을 주는 다수의 변수를 포함한 데이터셋을 제공한 후, 이들 변수 간 상호관계와 패턴을 학습하여 예측 정확도를 높이는 지도학습 방식을 소개한다. 지도학습 방식의 기계학습에 사용하는 대표적인 알고리즘 방식에는 배깅(bagging)과 부스팅(boosting)이 있는데, 이들 유형에 대한 설명은 아래와 같다.

##### 3.2.1 배깅(bagging) 방식 알고리즘

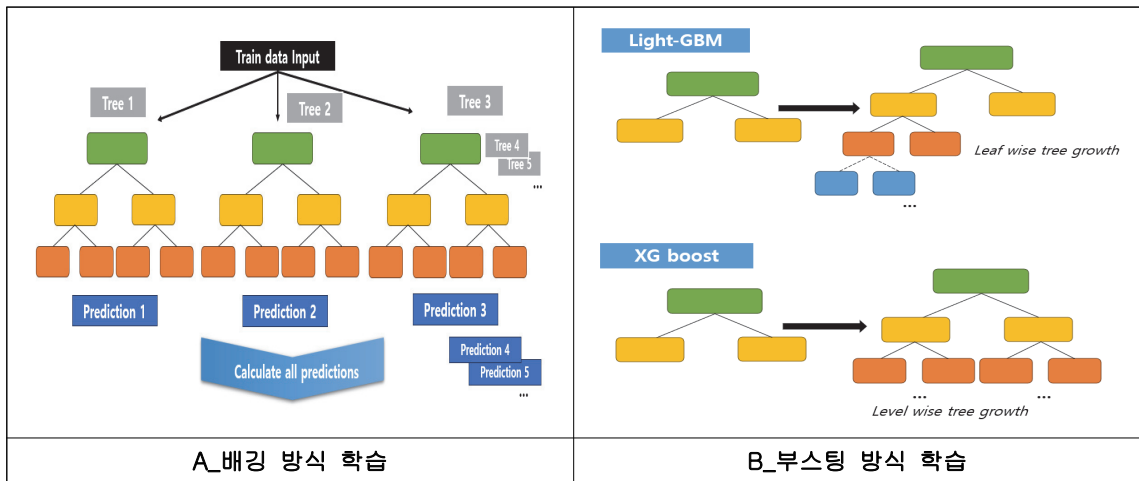
지도학습의 기본 알고리즘으로 여겨지는 의사결정 나무(decision tree)는 특정 변수를 기준으로 데이터를 분류하여 예측하는 형태이며, 예측구조의 레벨

(level)이 늘어날수록 데이터가 다수의 리프(leaf)로 분할되는 방식이다. 의사결정나무 알고리즘은 단순한 구조적 특성으로 해석이 용이한 장점이 있으나, 단일 학습모형에 예측을 의존하므로 학습에 사용한 데이터에 과대적합(overfitting)하여 미지의 데이터에 대해서는 예측력이 떨어지는 단점이 있다.

배깅 방식은 학습에 사용하는 데이터셋트를 시뮬레이션(simulation)을 통한 복원추출(bootstrapping)로 표본 수를 늘리고, 이를 표본추출(sampling)하여 다수의 부분 데이터셋트로 만들고, 이들 데이터셋트를 각각 다수의 학습 모형에 할당하여 각 모형이 독립적으로 학습하게 한 후, 각 독립 모형이 학습하여 얻은 예측 결과를 종합하는 방식으로 최종 예측 결과를 도출한다. 배깅 방식은 복원추출을 통해 데이터셋트 규모를 키우고, 다수의 학습 모형을 생성하여 예측값을 얻으므로, 단일 데이터셋트와 단일 학습 모형으로 인해 발생하는 과대적합 문제를 어느 정도 해소할 수 있다. 배깅의 대표적인 알고리즘은 랜덤 포레스트(random forest)로 다수의 독립적인 의사

결정나무들을 일부 변수만을 활용하여 무작위로 만들고 이들이 제시하는 예측값의 평균을 최종 예측값으로 활용한다. 랜덤 포레스트 예측 모형에 다양한 독립변수를 사용할 경우, 예측 정확도가 개선되는 것으로 알려져 있으며(Breiman, 2001), 기본적인 랜덤 포레스트 알고리즘의 학습 방식은 <Figure 1-A>과 같다.

본 연구에서 활용하는 데이터셋트는 시계열 항공유 가격 데이터를 기준으로 다수의 변수들을 구조화한 형태이다. 많은 변수들이 데이터셋트에 포함되어 있으므로 예측 과정에서 예측력이 뛰어난 변수들을 선별하고, 선별된 변수들을 바탕으로 예측 정확도를 향상하는 것이 중요하다. 랜덤 포레스트 알고리즘은 독립적인 예측 나무를 구성할 때마다 예측 변수들을 랜덤하게 선정하고 각 예측 결과를 평균하는 방식이므로 다수의 변수들을 적절하게 활용한 예측 방식에 유용할 수 있다. 실제로 원유 등 에너지 가격을 예측한 Herrera et al.(2019)의 연구 결과를 살펴보면 랜덤 포레스트 모형의 예측 정확도가 가장 높았다. 따



<Figure 1> 배깅 방식(random forest) / 부스팅 방식(Light-GBM, XGboost)



로서, Herrera et al.(2019)은 다양한 외부 변수를 활용하여 에너지 가격 예측 정확도를 향상할 수 있다고 주장하였다. 동일한 맥락에서 본 연구에서도 공군 항공유 가격에 영향을 미칠 수 있는 다수의 외부 요인들을 적절하게 선택하여 랜덤 포레스트 알고리즘으로 학습시킨다면 가격 예측 정확도를 향상시킬 수 있을 것으로 기대한다.

### 3.2.2 부스팅(boosting) 방식 알고리즘

부스팅 알고리즘도 여러 개의 학습 모형(예측 나무)을 생성하여 예측값의 오차를 줄인다는 점에서 배깅 알고리즘과 유사하다. 그러나 부스팅 알고리즘은 첫 번째 예측 나무의 학습 결과 도출한 예측값을 분석한 후, 예측값과 실제값의 차이가 큰 데이터에 가중치를 부여하여 다음 학습 모형을 만드는 방식으로 예측의 정확도를 향상시킨다. 이는 각 예측모형이 독립적으로 생성되는 배깅 방식(랜덤 포레스트)과 분명한 차이점이다. 부스팅 방식의 대표적인 알고리즘으로는 Light-GBM(Gradient Boosting Machine)과 XGboost(eXtream Gradient Boosting)이 있다. Light-GBM은 학습 모형 즉 예측 나무가 확장이 가능한 리프를 중심으로 레벨에 제약 없이 분할하는 방식이다. 반면에 XGboost 방식은 예측 나무의 각 레벨을 중심으로 모형을 확장하는 방식으로 두 방식 간 차이가 존재한다(Figure 1-B).

앞서 설명한 것과 같이 본 연구에서 활용하는 데이터세트는 항공유 가격에 영향을 주는 다수의 변수들로 구성되어 있다. 이때, 항공유 가격은 매일 최대 1회만 발표되어 1건의 데이터만 누적될 수 있어, 데이터세트 행(row)의 양이 부족한 상황이 예측 정확도에 부정적 영향을 줄 수 있다. 하지만, 본 연구모형의 데이터세트는 범위와 특성이 다른 다양한 변수들

을 포함하고 있는 특징이 있다. 이에 따라 개별 데이터가 지닌 의미를 최대한 활용하여 데이터가 적은 문제를 해결할 수 있는 예측모형이 필요하며, 부스팅 예측방식이 이러한 상황에 적절할 수 있다. 부스팅 예측모형은 학습을 반복하면서 실제값과 예측값의 차이가 큰 데이터에 집중하면서 각 학습 데이터에 포함된 변수의 특성을 활용하여, 이전 모델에서 발견하지 못한 패턴을 찾아 예측 정확도를 개선할 수 있기 때문이다. 항공유 데이터와 유사한 특성을 보이는 부동산, 주가가격, 국제 금 가격 시세 등 가격 예측 연구에서 부스팅 알고리즘의 정확도가 높았다는 점을 고려하면, 항공유 가격 예측에서도 우수한 예측 정확도를 기대할 수 있다(Avanijaa, 2021; Jabeur et al., 2021; Sharma et al., 2024).

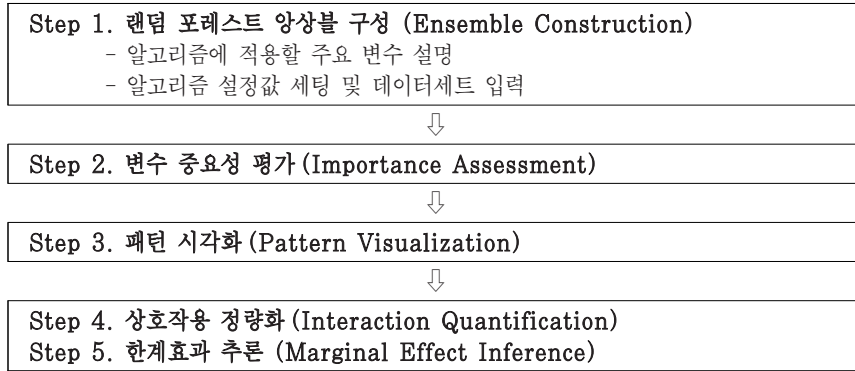
## IV. 항공유 가격 예측 모형/변수 검토

### 4.1 가격예측 절차

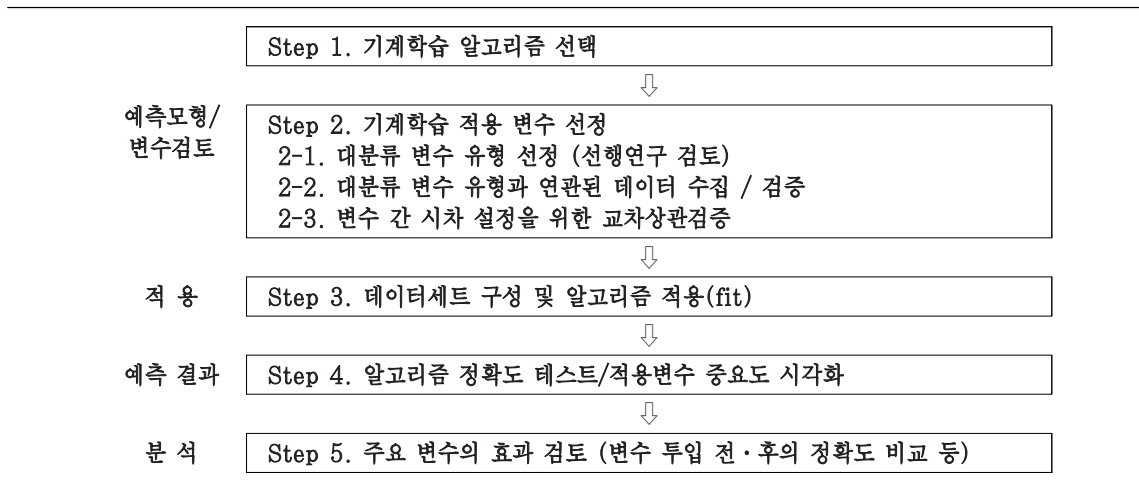
가격예측 연구 모형 개발을 위해 Chou et al.(2023)이 랜덤 포레스트 방식을 기준으로 제시한 기계학습 모형 활용 절차(〈Table 5〉)를 참고하였다. 랜덤 포레스트 방식만을 사용한 Chou et al.(2023)과는 다르게, 본 연구는 4가지 알고리즘(의사결정나무, 랜덤 포레스트, Light-GBM, XGboost) 기반 기계학습 예측 모형을 구축한 후, 데이터세트를 사용하여 각 모형의 예측 정확도를 비교하고, 이를 바탕으로 공군 항공유 구매사례에 적합한 예측모형을 찾는다. 본 연구의 기계학습 모형 활용 절차는 〈Table 6〉과 같다.

Step 1~2에서는 본 연구에서 사용할 기계학습 알고리즘을 선택한다. 다음으로 예측 모형의 학습에 활

〈Table 5〉 ‘랜덤 포레스트 알고리즘’ 적용 절차(Chou et al., 2023)



〈Table 6〉 항공유 가격 예측 기계학습 모형 적용 절차



용할 데이터세트의 구성을 위해 선행연구에서 확인한 석유제품에 영향을 주는 변수 선정의 기초 작업으로써 변수의 대분류를 도출한다. 그리고 변수 유형과 관련하여 데이터를 수집한 후 데이터의 적용 가능성을 검증한다. 이후, 선정된 변수들과 항공유 가격 간의 상관관계를 검증하고 상관관계가 확인된 예측 변수의 예측력을 높이기 위해 각 변수 데이터와

가격 데이터 간 매칭을 최적화할 수 있는 시차를 설정하기 위한 교차상관분석을 실시한다. Step 3~5에서는 항공유 가격 예측학습을 위해 구축한 데이터세트를 각 학습 모형에 투입하여 예측값을 구하고, 각 모형 간에 예측 정확도와 각 모형에 사용된 각 변수의 중요도를 상호 비교한다.

### Step 1 : 기계학습 알고리즘 선택

본 연구에서 적용한 지도학습 방식의 세부 알고리즘은 의사결정나무, 랜덤 포레스트, Light-GBM, XGboost 알고리즘이다. 해당 알고리즘들은 데이터 학습 방식에 차이가 있기에 예측 정확도, 중요 변수 비교가 용이하며, 이를 통해 항공유 가격 예측에 적절한 모형을 찾을 수 있을 것으로 기대한다.

### Step 2 : 기계학습 적용 변수 선정

기계학습에 적용할 변수의 선정은 3개의 세부 단계(Step 2-1, 2-2, 2-3)로 나누어 실시하였다. Step 2-1에서는 선행연구를 바탕으로 항공유 가격에 영향을 미칠 수 있는 요인의 유형(대분류)을 선정하였다. Step 2-2에서는 대분류 유형을 대표하는 세부 변수들을 선정하고 이들 변수의 데이터를 수집하고 수집한 데이터를 기반으로 변수 간 인과성(상관성 및 선행성)을 검증하였다. Step 2-3에서는 변수 간 시차 설정을 위한 교차상관검증을 실시하였다.

#### Step 2-1 : 대분류 유형 선정

본 연구에서는 (1) 유가 예측에 영향을 미치는 요인에 관한 선행연구 및 (2) 유가와 경제지표 간 관계를 분석한 선행연구를 검토하여 항공유 가격 변동의 원인 변수를 선정하기 위한 기초 작업으로서 가격 결정 요인의 유형 즉 대분류를 선정하였다. 연구모형에 적용한 대분류는 <Table 7>에 정리하였다. 첫 번째 대분류 유형은 원유와 석유제품의 가격지표이다. 항공유를 포함한 각종 석유제품은 원유의 정제공정에서 순차적으로 생산되므로 이들 제품의 가격은 원유 가격의 직접적 영향을 받으며, 각 제품의 수요와 공급은 다른 제품 가격에 직·간접으로 영향을 끼친다(Asche et al., 2010).

두 번째로 예측모형에 포함할 대분류 유형은 원유

와 석유제품 생산/수요 지표, 경제/무역 지표이다. Miao et al.(2017)은 원유 가격 예측에 원유의 공급, 수요, 금융시장 지표를 활용하였다. 노산하, 김남현(2019)도 유가 불확실성이 원유의 수요와 공급 변화에 기인한다고 하였다. Chai et al.(2011)도 석유 생산량 및 재고량, 주요 국가의 석유 수입량 등이 유가 변동에 영향을 미친다고 하였다. 이서진, 유종민(2022)은 유가 불확실성이 생산량은 물론, 글로벌 경제 현황, 금융지표 등 경제지표와도 밀접하게 관련되어 있다고 주장하였다. 따라서 본 연구의 기계학습 모형에는 (1) 원유와 석유제품의 생산/수요 지표 관련 세부 변수와 (2) 경제/무역 지표 관련 세부 변수를 예측 모형에 포함하여 예측 정확도를 향상시킬 수 있도록 두 대분류 유형을 선정하였다.

세 번째, 예측모형에 적용한 대분류 유형은 원유와 석유제품의 공급망/물류 관련 지표로 이다. 예를 들면, 조지성 외(2022)은 글로벌공급망압력지수(Global Supply Chain Pressure Index, GSCPI)가 유가에 유의미한 영향을 미침을 확인하였다.

네 번째, 대분류 변수 유형은 항공유 가격과 관련된 검색 실적이다. Boone et al.(2018)는 Google을 통한 검색 데이터를 활용하면 수요 예측 오류를 줄일 수 있다고 밝혔다. 김선미, 조두연(2022)도 'oil' 등 Google 검색 데이터를 활용하여 원유 가격 예측 모형의 정확도를 향상시켰다. 따라서, 본 연구는 항공유 가격과 관련된 Google 검색 기록에서 추출한 세부 변수를 예측 모형에 포함하여 예측 정확도 개선을 기대한다.

마지막으로 예측모형에 적용할 대분류 변수 유형으로는 지정학적/환경적 지표가 선정되었다. 예를 들면, Miao et al.(2017)는 지정학적 지표(중동과 북아프리카에서 발생한 테러 건수)가 유가 예측에 강력한 지표임을 확인하였다. 따라서, 본 연구는 주요 산유국 및

(Table 7) 항공유 가격 예측에 영향을 미치는 변수의 대분류 유형

구분	참고 문헌
① 원유/석유제품 가격	Asche et al.(2003), Li (2010)
② 생산/수요 지표	Miao et al.(2017), Chai et al.(2011), 노산하, 김남현(2019), 이서진, 유종민(2022)
③ 공급망/물류 지표	조지성 외(2022), 이서진, 유종민(2022)
④ 경제/무역 지표	Miao et al.(2017), Chai et al.(2011)
⑤ Google 검색 실적	Boone et al.(2018), 김선미, 조두연(2022)
⑥ 지정학·환경 지표	Miao et al.(2017)

소비국의 정치, 환경 과 관련한 세부 변수를 예측 모형에 포함하였다.

요약하면, 본 연구는 항공유 가격 예측 모형에 투입할 세부 변수 선정을 위한 기초 작업으로 원인 변수의 대분류 유형으로서 원유와 석유제품의 가격 지표, 생산/수요 지표, 경제/무역 지표, 공급망/물류 지표, Google 검색 실적 지표, 지정학·환경적 지표를 선정하였다.

**Step 2-2: 대분류 유형 관련 세부 변수 선정, 데이터 수집, 검증**

Step 2-2에서는 Step 2-1에서 선정한 변수의 대분류 유형(원유/석유제품 관련 가격지표, 생산/수요 지표, 경제/무역 지표, 공급망/물류 지표, Google 검색 실적 지표, 지정학적, 환경적 지표)을 대표하는 세부 변수 68개를 선정한 후, 이들 변수의 데이터를 수집 및 검증하였다. 첫째, 원유와 석유제품 가격지표는 국제유가 가격지표 3가지, 석유제품 가격 유형 6가지가 있으며, 한국석유공사에서 운영하는 Opinet을 통해 세부 데이터를 수집하였다.

둘째, 원유와 석유제품의 생산/수요 관련 세부 변수로는 미국 에너지정보청(U.S. Energy Information Administration, EIA)이 제공하는 19가지 원유/

석유제품의 생산, 공급, 재고, 소비 지표가 존재한다. 원유와 석유제품의 생산/수요 지표는 수요 공급 곡선에 따라 가격에 직접적인 영향을 미친다. 예를 들면, 석유제품의 단기적 소비량 증가는 가격 상승을 촉발하지만, 장기적 소비량 증가는 가격 안정화를 목적으로 하는 산유국의 증산, 주요 소비국의 비축유 방출 등으로 인해 가격이 하락할 수 있다(조지성 외, 2022).

셋째, 원유와 석유제품의 공급망/물류 대분류의 세부 변수로는 GSCPI, 항공운임지수(International Services: Air freight), 발틱석유제품선지수(Baltic Clean Tanker Index, BCTI), 발틱원유운반선지수(Baltic Dirty Tanker Index, BDTI)를 활용할 수 있다. GSCPI는 지표 발표 후 일정 시간 경과 후 유가에 영향을 미치는 것으로 나타났다(조지성 외, 2022). 대량 구매하는 원유와 석유제품의 운송은 주로 해상 운송에 의존하므로 원유/석유제품 운송과 직접적으로 관련된 BCTI와 BDTI를 예측 변수로 포함하였다.

넷째, 경제/무역지표의 세부 변수로는 원유와 석유제품의 수출, 수입 관련 지표를 선정하고 데이터를 수집하였다. 유가가 중·장기적으로 낮게 예측되는 경우, 각국 원유비축기관(예: 한국석유공사)과 정유업체는 원유 구매량을 실수로 대비 늘리게 된다. 대

표적인 원유 수입/석유제품 수출국인 한국의 석유산업 무역지표를 활용하면 유가 변동 예측력을 높일 수 있을 것이다. 따라서 본 연구는 관세청 통계포털을 통해 한국의 석유제품 수출, 수입 지표, 중동지역 수·출입 지표를 세부 변수로 선정하여 예측 모형 개발에 활용하였다.

다섯째, Google 검색 실적으로는 항공유 가격과 관련한 검색 데이터 실적을 세부 변수로 활용하기 위해 ‘항공유 가격(jet fuel price)’, ‘원유 가격(oil price)’, ‘OPEC 가격(OPEC price)’, ‘MOPS’ 월간 검색 실적을 수집하여 활용하였다.

여섯째, 지정학·환경적 지표의 세부 변수로서 방위산업체의 주가지수, 취약국가지수(Fragile States Index, FSI), 지정학적 위험 지수(Geopolitical Risk index, GPR Index), 석유 주요 소비국의 환경 관련 지수를 선정하여 데이터를 수집 및 활용하였다. 원유와 석유제품은 국가의 전략적 자원이므로 국제정치 이슈에 큰 영향을 받는다. 예를 들면, 러시아-우크라이나 전쟁 초기 유가가 50% 이상 상승하기도 하였다. Miao et al.(2017)은 원유 가격 예측

을 위한 지정학적 변수로서 글로벌테러리즘 데이터베이스(Global Terrorism Database) 만을 사용하였는데, <Figure 2>의 유가 전망 보고서에서 다루고 있는 이슈를 충분히 반영하였다고 볼 수 없다. 본 연구에서 활용한 지정학·환경적 지표의 세부 설명은 아래와 같다.

먼저, 방위산업체의 주가에 관해서는 세계 각지에서 발생하는 분쟁은 무기 수요를 증가시키고, 이는 다시 방위산업체의 중·장기적 수익 개선에 대한 기대감을 높여 글로벌 방위산업체 주가가 높인다(Zhang et al., 2022). 따라서, 세부 적용 변수로는 미국의 S&P Aerospace & Defense 지수,<sup>2)</sup> 2022년 기준 글로벌 1위 방위산업체인 Lockheed-Martin 주가가 활용되었다.

다음으로 활용한 지정학 요인의 세부지표는 비정부 기구인 평화기금(Fund for Peace)에서 산정하는 FSI와 GPR 지수이다. FSI는 정성적 미디어 분석, 관계지표 양적 분석, 전문가 그룹 대상 질적연구를 종합하여 세계 각국의 안보위협지수, 경제불평등지수 등 12개의 상대적 사회 취약지수를 산출하고 각국의



(Figure 2) 국내 유가전망자료(석유공사) 텍스트 검토결과

2) S&P Aerospace & Defense 지수는 글로벌산업분류기준에 따라 항공우주 및 방위산업으로 분류되는 종목 지수로 항공기 및 방위산업체, 군용 조선기업, 무기 완성품 및 부품 업체들이 포함되어 있다.

<Table 8> 항공유 가격 예측 기계학습 모형에 포함할 대분류 및 세부 변수 검토

구분	세부 변수	적용 판단	시차 적용	데이터 출처	
$Y$	항공유가격(1)	변동원가 요소 적용: 등유 가격(배럴당) × 환율(원/달러)	-	-	
$X_1$	원유/석유제품 가격(8/9)	X.1.1. Dubai	○	1개월	한국석유공사 (Opinet)
		X.1.2. WTI	○	1개월	
		X.1.3. Brant	○	1개월	
		X.1.4. 휘발유(보통)	○	1개월	
		X.1.5. 휘발유(고급)	○	1개월	
		X.1.6. 경유	○	1개월	
		X.1.7. 증유	○	1개월	
		X.1.8. 나프타	○	1개월	
		X.1.9. 등유(전세계 시장가격의 평균)	×(상관관계 낮음)		
		$X_2$	생산/수요 지표(8/19)	X.2.1. 미국 등유 생산량	
X.2.2. 미국 등유 공급량	×(상관관계 낮음)				
X.2.3. 미국 등유 재고	○			4개월	
X.2.4. OECD 석유제품 생산량	×(상관관계 낮음)				
X.2.5. US 석유제품 생산량	○			12개월	
X.2.6. Non-OECD 석유제품 생산량	×(상관관계 낮음)				
X.2.7. OPEC 석유제품 생산량	×(상관관계 낮음)				
X.2.8. OPEC 원유 생산량	×(상관관계 낮음)				
X.2.9. OPEC 기타제품 생산량	×(상관관계 낮음)				
X.2.10. 전세계 석유제품 생산량	×(상관관계 낮음)				
X.2.11. Non-OPEC 석유제품 생산량	○			12개월	
X.2.12. US 석유제품 소비량	○			12개월	
X.2.13. 유럽 석유제품 소비량	×(상관관계 낮음)				
X.2.14. OECD 석유제품 소비량	×(상관관계 낮음)				
X.2.15. non-OECD 석유제품 소비량	○			12개월	
X.2.16. 아시아 석유제품 소비량	○			12개월	
X.2.17. 전세계 석유제품 소비량	×(상관관계 낮음)				
X.2.18. US 석유제품 재고량	○			1개월	
X.2.19. OECD 석유제품 재고량	○			1개월	
$X_3$	공급망/물류 지표(4/5)	X.3.1. 공급망압력지수(GSCPI)	○	12개월	뉴욕연방은행
		X.3.2. 항공운임지수	○	1개월	미.노동통계국
		X.3.3. 해상운임지수(Dry)	×(석유제품과 관련성 낮음)		investing.com
		X.3.4. 해상운임지수(Clean)	○	2개월	
		X.3.5. 해상운임지수(Dirty)	○	2개월	

<Table 8> 항공유 가격 예측 기계학습 모형에 포함할 대분류 및 세부 변수 검토 (계속)

구분	세부 변수	적용 판단	시차 적용	데이터 출처	
X <sub>4</sub>	경제/무역 지표(4/11)	X.4.1. 원/달러 환율	×	(석유제품과 관련성, 상관관계 낮음)	-
		X.4.2. 중동지역 수출액	○	12개월	관세청
		X.4.3. 중동지역 수입액	○	1개월	
		X.4.4. 석유제품(HS code 27) 수출액	○	1개월	
		X.4.5. 석유제품(HS code 27) 수입액	○	1개월	
		X.4.6. 미국 비농업고용지수(NFP)	×	(석유제품과 관련성, 상관관계 낮음)	
		X.4.7. 미국 제조업지수(ISM)	×	(석유제품과 관련성, 상관관계 낮음)	미.공급자협회
		X.4.8. 미국 석유기업 증가1	×	석유화학 산업의 증가지표는 유가의 후행지표로서 주로 제시되어 제외함 (Faff, & Brailsford, 1999)	
		X.4.9. 미국 석유기업 증가2	×		
		X.4.10. 한국 석유기업 증가1	×		
		X.4.11. 한국 석유기업 증가2	×		
X <sub>5</sub>	Google 검색 실적 (1/4)	X.5.1. 'Jet fuel price' 검색 데이터	○	1개월	Google
		X.5.2. 'Oil price' 검색 데이터	×	(상관관계 낮음)	
		X.5.3. 'OPEC price' 검색 데이터	×	(상관관계 낮음)	
		X.5.4. 'MOPS' 검색 데이터	×	(상관관계 낮음)	
X <sub>6</sub>	지정학/환경 지표(10/20)	X.6.1. S&P Aerospace & Defense	○	2개월	-
		X.6.2. 미국 방산업체 증가	○	12개월	-
		X.6.3. 안보위협지수	×	(상관관계 낮음)	Fund for Peace
		X.6.4. 계층분열위협지수	○	12개월	
		X.6.5. 그룹분열위협지수	×	(상관관계 낮음)	
		X.6.6. 경제붕괴위협지수	○	12개월	
		X.6.7. 경제불평등지수	○	12개월	
		X.6.8. 지식계층유출지수	×	(상관관계 낮음)	
		X.6.9. 국가정당성지수	×	(상관관계 낮음)	
		X.6.10. 공공서비스지수	×	(상관관계 낮음)	Fund for Peace
		X.6.11. 인권지수	×	(상관관계 낮음)	
		X.6.12. 인구성장위협지수	○	12개월	
		X.6.13. 난민지수	×	(상관관계 낮음)	
		X.6.14. 외부침입지수	○	24개월	
		X.6.15. 전세계 지정학 위협지수	×	(상관관계 낮음)	American Economic Review
		X.6.16. 주요 산유국 지정학 위협지수	×	(상관관계 낮음)	
		X.6.17. 미국 10대 신문사 안보관련 기사수	×	(상관관계 낮음)	
		X.6.18. 미국 3대 신문사 안보관련 기사수	○	12개월	
		X.6.19. 각 데이터별 계절 반영	○	-	-
		X.6.20. 석유 주요 소비국 월별 기온 추세(미국 등 5개국)	○	1개월	기상청

취약성 점수와 순위를 도출한다. GPR 지수는 지정학적 위험을 테러, 군사개입, 전쟁 개시 등의 유형으로 구분하고 뉴욕타임즈(New York Times) 등 미국, 영국, 캐나다 주요 언론매체 기사에서 이들 위험의 언급 빈도를 국가별로 측정하여 산출한다(Caldara, 2022). 각 지정학적 위험 변수는 항공유 생산과의 연관성을 고려하여 석유 수출국 기구(Organization of the Petroleum Exporting Countries, OPEC) 회원국, 미국, 러시아 등 주요 산유국 관련 지표만을 연도별로 합산하여 세부 변수로 활용되었다.

마지막으로 환경 지표 관련 변수는 기온 변동으로 인한 난방용 석유제품의 수요량 영향을 고려하여 미국, 중국 등 10대 석유제품 소비국의 소비량을 가중 평균한 값을 세부 변수로 설정하고 데이터를 수집하였다.

**Step 2-3 : 변수 간 시차 설정을 위한 교차상관검증**

본 연구에서 정의한 교차상관계수는 시계열 특성을 가지는 변수  $X_t$ 와  $Y_t$ 가 있을 때,  $k$ 시점 전의  $X_{t-k}$ 와 현재의  $Y_t$ 와의 관계로 수식(3)과 같이 정의하였다.

$$\rho_{XY}(k) = \rho(X_{t-k}, Y_t) = \frac{Cov(X_{t-k}, Y_t)}{\sqrt{Var(X_{t-k})} \sqrt{Var(Y_t)}} = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y} \quad (3)$$

항공유 가격( $Y_t$ )과 초기 학습모형(Step 2-2)에 선정된 68개 변수( $X_t$ )와의 교차상관계수를 검토한 결과, 25개는 상관계수 절대값이 0.3 이하로 낮거나 변수가 항공유 가격 변동에 후행했을 때의 상관계수가 선행했을 때의 상관계수보다 높아, 즉 인과성이 부족하여, 모형에서 제외하였다. 본 연구가 기계학습 모형 개발을 위해 검토하고 최종 선정한 대분류 유형, 세부 변수, 그리고 데이터 출처를 <Table 8>에 정리하였다.

교차상관검증 결과에 따라 학습모형에 변수를 적용할 때에는 항공유 가격( $Y_t$ )에 대해 교차상관계수가 가장 높은 시점( $t-k$ )으로 변수( $X_{t-k}$ )를 선행시켜 적용하였다. 예를 들면, 교차상관분석 결과에 따라 해상운임지수(Baltic Clean Tanker Index, BCTI)와 항공유 가격의 상관계수가 가장 높은 시기는 항공유 가격 확정 전 2개월 전의 시점이며, 미국의

<Table 9> 변수별 시점 차이 적용 예시

구 분	교차상관계수	해상운임지수-항공유 가격 관계	구 분	교차상관계수	미국 석유재고-항공유 가격 관계
1개월	0.452		1개월	-0.679	
2개월	0.473		2개월	-0.669	
3개월	0.444		3개월	-0.656	
4개월	0.406		4개월	-0.640	
5개월	0.374		5개월	-0.624	
6개월	0.350		6개월	-0.605	



석유 재고량과 항공유 가격의 상관계수가 가장 높은 시기는 항공유 가격 확정 전 1개월 전의 시점이다. 이에 맞추어 항공유 가격( $Y_t$ )에  $BCTI(X_{t-2\text{개월}})$ , 미국 석유 재고량( $X_{t-1\text{개월}}$ ) 데이터를 연동하였다 (<Table 9>).

### Step 3 : 데이터셋 구성 및 알고리즘 적용

본 연구에서 활용한 기계학습 알고리즘은 의사결정 나무, 랜덤 포레스트, Light-GBM, XGBoost 이다. 학습모형에 적용하는 데이터셋은 2008년 1월 1일부터 2023년 1월 31일까지의 일일 기준 데이터로 총 3,810개이다. 각 학습모형 별로 훈련에 사용할 데이터(data for model training)로 90%를 할당하고 평가 데이터(data for model testing)로 10%를 할당하여 모형을 실행한 후 결과값을 비교하였다.

## V. Step 4, 5 : 항공유 가격 예측 결과

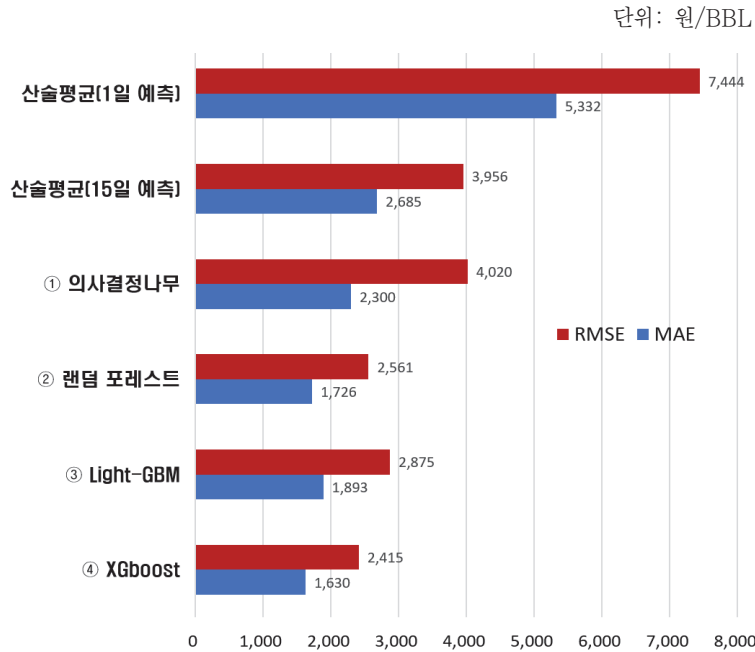
### 5.1 항공유 가격 예측 모형 종합 평가

본 연구에서 적용한 모형의 종합평가를 위해 기계 학습 회귀 알고리즘을 기반으로 한 4개의 가격 예측 모형과 기존의 예측방식의 정확도를 비교한 결과는 <Table 10>과 같다. 각 예측 모형 간 정확도 비교를 위해서 예측값과 실제값의 차이를 계산하는 평균 절대오차(Mean Absolute Error, MAE, 수식 1)와, 평균제곱근오차(Root Mean Squared Error, RMSE, 수식 2)를 활용하였다. 공군은 매달 항공유 구매규모 결정 시, 이번 달 가격과 다음 달 예측 가격을 비교하여 '최대 구매' 또는 '최소 구매' 판단을 할 수 있다. 이에 따라 아주 작은 가격 차이에도 의

사결정이 완전히 반대 방향으로 진행될 수 있다. 그러므로, 예측 모형의 정확도 평가에는 큰 오차값에 더 민감하게 반응하는 RMSE를 기반으로 예측 모형의 우수성을 최종 판단하였다. <Table 10>에서 보듯이 XGboost 알고리즘 기반 가격 예측 모형이 MAE, RMSE 값이 다른 모형 대비 모두 낮아 가장 우수한 예측 모형으로 판정되었다. XGboost 기반 예측 모형은 기존의 산술평균 예측 방식의 모형 대비 오차가 67% 감소한 결과를 나타내었다. 구체적으로 XGboost 알고리즘 기반 모형은 RMSE가 2,415원/BBL로 매월 15일까지의 시계열 데이터를 산술평균하여 예측하는 현재 예측 모형의 오차수준 보다도 뛰어나다. XGboost 기반 모형 다음으로는 랜덤 포레스트, Light-GBM 모형 순으로 현재 예측 모형 대비 예측 정확도가 높은 것으로 나타났는데, 이들 4개의 기계 학습 기반 모형들 모두 산술평균 방식의 모형 대비 예측 정확도가 최소 45% 이상 향상되었다. 정리하면, 기계학습 기반을 항공유 가격 예측 모형을 활용하면 지금까지 매월 15일까지 일일 가격 데이터가 누적되기를 기다리지 않고 매월 시작 시점 혹은 그 이전에 가격 예측을 실시하고도 높은 정확도로 항공유 가격을 예측할 수 있다.

본 연구에서 개발한 예측 모형이 공군의 항공유를 포함한 석유제품 구매절차 활용될 경우, 시기별 가격 차이에 따른 구매 규모를 현재 방식 대비 탄력적으로 조정할 수 있으므로 예산 절감과 운영적 성과가 기대된다. 또한, 항공유 구매 가격 예측 시점을 앞당김으로써 항공유 구매 의사결정의 유연성을 높이고, 항공유 운송수단 등 군수지원 자원의 활용도를 높일 수 있다.

<Table 10> 항공유 가격 예측모형 종합 평가(매월 1일 예측 모형 적용 기준)



\* 기계학습 방식(①~④): 매월 1일, 다음월 가격에 대한 예측 시행기준

\* 가격 예측모형 종합 평가 활용 데이터 기준

- 기계학습법: '08.1월~'23.1월 중 기계학습 모형 평가(test, 10%)에 활용된 379개 데이터

- 산술평균(1일/15일 예측): [X월 일일가격 예측값] X-1월 25일~31일/X월 1~14일 일일가격 산술평균값

구 분		오차수준 (원/BBL)	
		MAE	RMSE
현재 방식	매월 1일 예측	5,332	7,444
	매월 15일 예측	2,685	3,956
기계학습 방식	① 의사결정나무	2,300	4,020
	② 랜덤 포레스트	1,726	2,561
	③ Light-GBM	1,893	2,875
	④ XGboost	1,630 (↓69%)	2,415 (↓67%)

## 5.2 주요 예측 모형 별 예측 결과

### 5.2.1 랜덤 포레스트 알고리즘 기반 가격 예측 모형

랜덤 포레스트 알고리즘 기반 예측 모형은 다수의

독립적인 의사결정나무를 생성하고 각 의사결정나무가 도출한 예측값을 평균하여 예측 정확도를 향상할 수 있다. 항공유 가격 예측을 위한 랜덤 포레스트 알고리즘 기반 모형은 원유/석유제품 가격에 절대적으로 의존하고 있으며, 변수 중요도의 상위 6개 변수

가 모두 원유/석유제품 가격 변수로 이들의 중요도 합계는 67.4%이다(〈Table 11〉). 이중 경유 가격은 변수 중요도가 19%로 가장 큰데, 이는 경유와 '등유(항공유의 주원료)'가 난방과 장비 가동용 연료로서 유사한 수요 패턴을 보이기 때문으로 판단한다. 〈Figure 3〉과 같이 석유제품의 시기별 가격 변동 추세를 살펴보면, 경유는 등유와 거의 유사한 가격 변동 추세를 나타낸다. 반면에 등유와 휘발유는 전체적으로 유사한 가격 변동 추세를 보이나 특정 기간(2022년 1~4월 등)에는 정반대의 가격 추세를 보

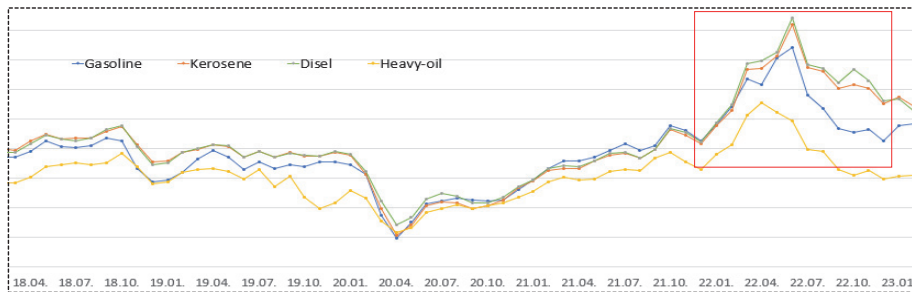
이기도 하며, 가격 급등과 급락 폭도 차이를 보였다. 랜덤 포레스트 알고리즘 기반 가격 예측 모형은 경유 가격을 비롯한 원유/석유제품 가격에 대한 예측 의존도가 높으므로, 원유/석유제품 가격 변동 원인을 고려하여 신중한 예측 결과 판단이 필요하다.

### 5.2.2 Light-GBM 알고리즘 기반 예측 모형

Light-GBM 알고리즘 기반 예측 모형은 각 예측 나무에서 발생하는 오차에 가중치를 부여하며, 예측

〈Table 11〉 랜덤 포레스트 기반 모형 분석 결과 및 변수 중요도 산출 결과

데이터 훈련/평가 비율	90 : 10	예측오차 (원/BBL)	MAE	1,726
			RMSE	2,561
<b>중요 활용 변수</b> : 경유 가격(19.4%), 국제유가 (Brent, Dubai, WTI), 휘발유 가격, 무역지표 등				



〈Figure 3〉 석유제품 유형별(휘발유, 등유, 경유 등) 가격 흐름

〈Table 12〉 Light-GBM 기반 모형 분석결과 및 변수 중요도 산출결과

데이터 훈련/평가 비율	90 : 10	예측오차 (원/BBL)	MAE	1,893
			RMSE	2,875
<b>중요 활용 변수</b> : 경유 가격(7.3%), S&P Aerospace & Defense(6.8%), 중유 가격, 국제유가(WTI), 해상운임지수				
	⋮			

모형을 리프(leaf)를 중심으로 비대칭적으로 확장하여 예측 정확성을 향상하는 방식이다. 각 변수에 대한 중요도는 〈Table 12〉에 정리한 것과 같이 경유 가격, S&P Aerospace & Defense 지수, 중유 가격, 국제유가(WTI) 순으로 나타났다. Light-GBM 기반 예측 모형에서는 가장 중요도, 즉 설명력이 높은 변수인 경유 가격이 예측에 활용된 비중이 7%에 불과하며, 원유/석유제품 가격, 지정학 지표, 공급망/물류 지표, Google 항공유 가격 검색 실적 등 다양한 변수들이 예측에 활용된 것을 확인할 수 있다. 이러한 결과는 경유 가격이 절대적인 예측 변수로 나타난 랜덤 포레스트 예측 모형과는 다른 결과이다. 따라서, 다양한 변수들 간의 관계를 고려한 Light-GBM 예측 모형을 활용하면 안정적 가격 예측이 가능할 것으로 보인다.

### 5.2.3 XGboost 알고리즘 기반 가격 예측 모형

XGBoost 알고리즘 기반 예측 모형은 각 예측 나무에서 발생한 오차를 기준으로 오차가 가장 큰 데이터에 가중치를 부여하여 레벨별로 일정하게 확장

하는 방식이므로 그 변수와 종속변수 간의 관계를 세밀하게 찾아내는 장점이 있다. 또한, XGboost 알고리즘은 일반적인 Gradient Boosting Machine (GBM) 대비 병렬 및 분산처리를 통하여 신속하게 데이터를 학습하는 장점을 가지고 있다. XGboost 기반 예측 모형이 가장 높은 정확도를 보여준 것은 학습 능력의 확장성을 기반으로 희소한 데이터의 특성을 놓치지 않고 종속변수와의 연관성을 찾아내는 알고리즘 특성 때문인 것으로 추정할 수 있다(Chen & Guestrin, 2016). XGboost 알고리즘 기반 항공유 가격 예측 모형에서는 항공유 가격 예측에 경유 가격, 지정학적 위험지수, Google 검색실적, 항공운임지수 등의 변수가 영향을 미쳤음을 확인할 수 있었다(〈Table 13〉). 다만, 항공유 가격 예측 결정력의 약 50%가 4개 변수(경유 가격, 경기침체지수, 외부개입지수, Google 검색실적)에 집중되어 있어 이들 변수에서 예상치 못한 변동이 생기면 예측 값이 크게 변동하므로 정확도가 낮아진다. 따라서 XGboost 기반 모형의 예측 결과 해석 시, 주요 활용 변수들의 급변성 추세 등을 면밀히 검토할 필요가 있다.

〈Table 13〉 XGboost 분석결과 및 변수 중요도 산출결과

데이터 훈련/평가 비율	90 : 10	예측오차 (원/BBL)	MAE	1,630
			RMSE	2,415
<b>중요 활용 변수</b> : 경유 가격(16.1%), 지정학 지표, Google 항공유 가격 검색실적, 국제유가(Brent), 항공운임지수				

〈Table 14〉 기계학습 예측모형별 대분류 유형의 변수 중요도(feature importance)

단위: %, ( ): 순위

대분류 유형	세부 변수(개)	랜덤 포레스트	Light-GBM	XGboost		
				방법①	방법②	방법③
원유/석유제품가격( $X_1$ )	8	78.3 (1)	54.2 (1)	36.9 (1)	59.0 (1)	82.6 (1)
생산/수요( $X_2$ )	8	5.3 (4)	7.1 (4)	10.9 (3)	4.8 (4)	3.8 (4)
공급망/물류( $X_3$ )	4	3.4 (5)	16.4 (3)	7.3 (5)	13.3 (3)	4.9 (3)
경제/무역( $X_4$ )	4	5.8 (2)	4.2 (5)	1.7 (6)	2.8 (5)	1.4 (6)
Google 검색 실적( $X_5$ )	1	1.6 (6)	1.2 (6)	10.2 (4)	0.6 (6)	1.8 (5)
지정학·환경( $X_6$ )	10	5.6 (3)	16.9 (2)	33.0 (2)	19.5 (2)	5.4 (2)

\*방법1(gain): 각 변수가 예측모형에 기여한 평균적인 성과 수준을 기준으로 중요도 측정

\*방법2(weight): 각 변수가 예측모형 트리 분할에 얼마나 자주 사용되었는지 빈도를 기준으로 중요도 측정

\*방법3(permutation): 각 변수별로 수치를 무작위 혼합하여, 모형에 얼마나 중요한 영향을 미치는 지를 측정

### 5.3 추가분석: '지정학·환경/Google 검색 실적' 대변수 유형의 영향성

본 연구의 예측모형은 유가 예측에 원유 생산/수요, 경제지표를 주요 변수로 활용한 기존 연구와는 달리 지정학·환경적 지표와 Google 검색 실적도 예측요인으로 포함하였다. 비록 이들 두 가지 대분류 변수 유형이 항공유 가격 예측에 미치는 영향 정도가

기계학습 모형과 변수의 중요도를 산출하는 방식에 따라 차이가 있지만, 두 변수 유형 모두 예측모형이 데이터를 학습하고 정확도를 개선하는데 기여한다는 것을 확인하였다. 〈Table 14〉을 살펴보면 부스팅 예측 모형은 지정학·환경적 지표를 원유와 석유제품 가격 다음으로 비중 있게 예측에 사용하였다. 특히, 가장 높은 예측 정확도를 보인 XGboost 예측 모형은 변수 중요도 측정 방식과 관계없이 지정학·환경

변수들을 의미있게 활용함을 확인할 수 있다. 결과적으로 본 연구에서는 주요 산유국들의 지정학적 상황, 석유 소비국의 환경적 요인 등이 석유제품 가격 예측 정확도 향상에 효과를 준다는 것을 알 수 있었다. Google 항공유 가격 검색실적의 경우에도 'jet fuel price'의 검색 실적만을 연구모형에 활용했지만 XGboost 모형에서 중요도가 10% 수준(방법1 기준)으로 예측 모형 정확도를 결정하는 주요 요인으로 드러났다.

## VI. 결론

국방 분야에서 운영하는 대다수의 장비는 석유제품을 연료로 사용하는 만큼, 석유제품 운영의 작전적 중요성은 높다. 경제적으로도 연료 구매비용은 무기체계 전체 유지비의 10% 수준(선미선 외, 2016)으로 효율적인 국방 예산관리를 위해서는 연료구매 가격의 정확한 예측을 기반으로 효과적인 재고관리가 필수적이다. 그러나 유류 가격의 높은 변동성으로 인해 군은 유류 및 석유제품의 운영관리에서 심각한 리스크에 노출되어 있다. 군은 유류 공급업체와의 장기 계약을 통해 공급의 안정화를 꾀하고 있지만, 여전히 예측이 어려운 구매 가격은 작전능력 유지에 필수적인 재고량의 안정적 확보의 저해 요소이다. 다음 달 가격 예측을 위해 군은 누적된 항공유 일일가격 실적 데이터를 시계열 분석할 수 있지만 다양한 요인에 의해 영향을 받는 유류 가격 특성으로 정확도가 높지 않은 문제점이 있다. 이러한 상황에서 공군자산의 안정적 운영을 위해서는 높은 가격으로라도 소요량을 확보해야 하므로 공군에 재무적 부담 요인으로도 작용하고 있다.

이에 따라 본 연구에서는 항공유 가격 예측의 정확도를 높이기 위해 예측모형에 기계학습 알고리즘을 적용하였다. 기계학습 가격 예측 모형에는 기존 연구에서 이미 검증된 변수에 더하여 본 연구진이 새롭게 개발한 변수를 포함하였다. 본 연구가 제시한 4개의 예측 모형은 현재 전통적 시계열 기반 예측 모형 대비 최소 45%에서 최대 67%까지 오차를 줄여서 모형 활용의 적절성을 보여주었다.

본 연구에서 제안한 기계학습 기반 항공유 가격 예측 모형에 사용한 데이터는 한국석유공사(Opinet)에서 제공하는 2008년 1월부터 2023년 1월까지 등유(항공유 원료) 가격을 기준으로 다른 변수들을 연결한 데이터이다. 향후 시간이 지남에 따라 누적되는 추가 데이터를 본 연구 모형에 투입하여 모형을 학습시킬 경우, 예측 정확도를 점진적으로 개선할 수 있을 것이다.

## VII. 연구의 의의, 한계점 및 향후 연구 방향

### 7.1 연구의 의의

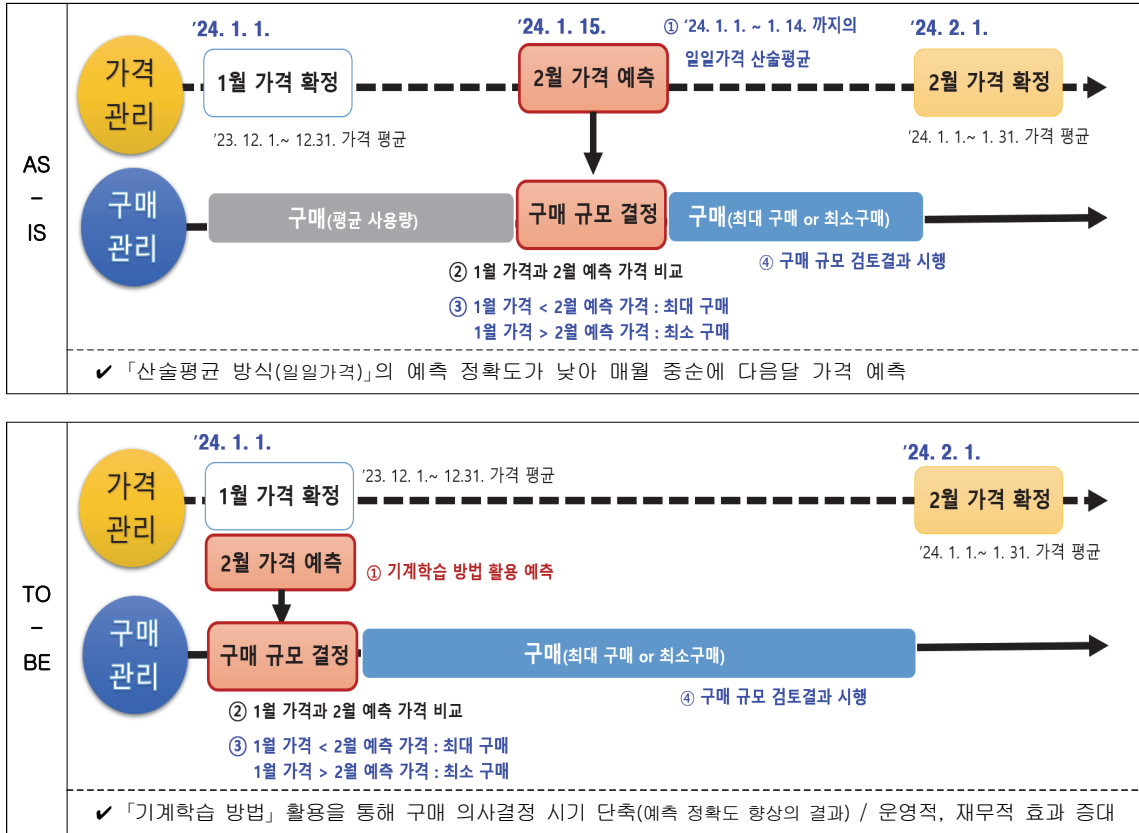
본 연구의 의의는 매우 높은 가격 변동성을 보이는 군용 항공유 가격 예측에 기계학습 알고리즘을 활용한 모형을 개발하여 가격 예측의 정확도가 향상될 수 있다는 것이다. 본 연구에서 테스트한 항공유 가격 예측 모형을 공군에서 활용한다면, 월별 변동 가격으로 계약하는 항공유 구매 상황에서 항공유 가격 예측 시기를 앞당겨 항공유 구매 규모를 조기에 조정함으로써 작전수행능력을 유지하면서도 구매비용 절감 효과를 기대할 수 있다. 예를 들면 기존에 산술 평균법을 사용하면 예측 정확도 향상을 위해 매일 약

15일까지 가격 데이터를 수집하기 위해 기다려야 했다. 이러한 현행 프로세스는 월별 가격 차이에 기반한 구매량 결정이 늦어지게 되어 군 운영에 지장을 초래할 수 있음은 물론, 제한적인 예산 절감 효과만을 기대할 수 있었다. 향후, 기계학습 방법을 항공유 가격 예측에 적용한다면 우수한 예측 정확도로 인해 매월 초에 구매 의사결정이 가능하여 항공유 수송, 하역, 검사 계획 등을 효율적으로 조정할 수 있는 등의 항공기 운용에 대한 성과 개선을 예상할 수 있다 (Figure 4). 재무적으로도 기계학습 예측모형을 통해 구매량 조정 규모를 확대할 수 있고, 의사결정의 정확도가 향상되기에 성과 수준이 산술평균 예측 방식을 운영하는 것 보다 약 10억 원 이상 구매비용을 절감할 수 있을 것으로 기대한다(부록 참조). 예상되는 절감액 10억원 정도는 공군의 총 구매비용(연간 약 3,000억원) 대비 예산 절감 규모가 크지는 않으며, 더욱이 민간기업 경영자의 시각에서 보면 비용 절감 효과가 적다고 할 수 있다. 그러나, 군에서는 연중 긴급을 요하는 소요 제기가 많은 반면, 정부예산의 조정과 추가 확보는 오랜시간이 소요되어 갑작스러운 항공유 등의 가격 변동에 따른 긴급 예산을 확보하기 어려운 것이 현실이다. 따라서 기계학습 기반 예측에 기반한 정확도 향상으로 예상되는 예산 절감액은 군수 운영에 큰 유연성을 제공할 수 있다. 다만, 본 연구가 제시한 예산 절감액은 연간 예산 수준, 구매 규모, 일정한 수요량과 물류능력 등의 다수 요소를 일정하다고 가정하고 얻은 결과이다. 따라서, 향후 연구에서는 정확한 예산감축 효과 판단을 위해서 월간 항공유 수요량 추세, 물류능력의 변동을 고려하여 예산 절감액을 세부적으로 분석할 필요성이 있다. 이에 더하여, 항공유 이외 물품에 본 연구가 제시한 예측 방법론을 적용하기 위해서는 각 물품의 운영 상황에 맞는 재무적 성과측정 모형을 향후 연

구할 필요가 있다.

또한, 항공유 가격 예측 능력 향상은 가격 불확실성에 대비하기 위해 재고를 추가 보유하려는 공급사슬 주체들의 행동을 억제하여 효율적 재고관리에 기여할 수 있을 것으로 기대한다. 본 연구에서는 항공유의 구매사례에서 기계학습 방법에 기반한 가격 예측 모형의 정확도 개선 효과를 확인하였다. 향후 연구에서는 본 연구가 제안한 가격 예측 모형을 수정하여 다른 석유제품 가격 예측에 사용할 수 있을 것이다. 그리고, 기관(기업) 운영에서 반복적으로 구매하고 가격 변동성이 심한 품목에 대해 계약조건 등 상황에 맞게 모형을 수정하여 적용할 수 있을 것으로 판단한다.

다음으로, 본 연구모형에는 기계학습 기반 예측을 위해 기존 유가 예측 관련 연구에서 활용한 다양한 원인 변수들이 데이터셋으로 활용되었다. 기존 연구에서 널리 활용하지 않은 GSCPI, 항공운임지수 등 공급망 관련 지표는 물론, 지정학적, 환경적 이슈 관련 변수, Google 검색 실적 등도 연구모형에 포함하였고 그 유효성을 기계학습 방법을 통해서 검증하였다. 따라서, 다양한 요인에 의해 영향을 받는 원유 및 석유제품 가격 예측 연구에 주요 영향 요인 분류, 세부 변수 선정 등 연구 범위를 넓히는데 기여할 수 있을 것으로 기대한다. 다만, 기계학습 예측 모형을 운영하기 위해서는 기존 시계열 예측 모형 대비 예측을 위한 다양한 데이터의 추가 확보, 알고리즘 실행 시간이 필요할 수 있다. 따라서, 향후 연구에서는 각 알고리즘에서 공통으로 도출된 주요 변수를 적극 활용하는 방법으로 데이터 수집에 필요한 노력을 줄이거나, 최적 변수 선정 알고리즘(Dynamic Feature Importance based Feature Selection, DFIFS) 사용하는 방식 등을 채택한다면 가격 예측 정확도를 확보하면서도 예측에 투입되는 자원을 최소화할 수 있을 것이다(Wei et al., 2020).



〈Figure 4〉 항공유 가격/구매관리 프로세스 변화 (매월)

## 7.2 연구의 한계점 및 향후 연구 방향

본 연구에서 제시한 예측모형의 예측 정확도를 향상하고 안정적으로 모형을 활용하기 위해서는 추가적인 데이터 유형의 탐색과 연구에서 적용한 알고리즘 이외에 다른 알고리즘 적용방안 검토가 필요하다. 그리고 기계학습 알고리즘의 일반적인 제한점으로 논의되는 예측 과정에서 상세한 해석이 제한되는 점이 연구 과정에 존재하며, 연구에 적용된 변수들의 데이터 구조화 과정에서도 일간 데이터와 월간 단위 이상의 데이터를 연결하는 다양한 방법이 존재하여 이를

합리적으로 구조화할 수 있는 연구가 필요하다.

또한, 예측 정확도를 높인 기계학습 기반 가격 예측 모형을 기초로 기타 다양한 요인들을 고려하여, 더욱 정교한 의사결정을 실행할 수 있는 의사결정 종합모형을 향후 공군 구매관리에 도입하는 것도 고려할 수도 있다. 예를 들면, 시나리오 플래닝(scenario-planning)과 같은 방법을 활용한다면 복잡한 의사결정 상황에서 직관적인 구매 판단이 가능할 것으로 보인다. 시나리오 플래닝은 주요 불확실성과 성과에 영향력이 높은 요인에 집중함으로써 핵심 상황에만 집중할 수 있도록 도와줄 수 있는 기법이다(Amer



**구매 의사결정 주요 영향요인 선정**

- [주요] 가격 예측 모형의 정확도, [기타] 물류지원능력, 재고수준

**(1) 가격 예측 모형 정확도를 고려한 예측시기 설정**

: ex) 1회차(매월 1일) & 2회차(매월 16일)

**(2) 핵심 의사결정 상황 검토: 예측모형 정확도, 물류지원능력**

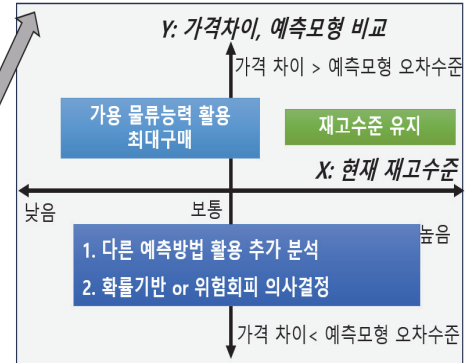
기준(1회차)	변경(2회차)	다음달 가격 하락 (최소 구매)	다음달 가격 상승 (최대 구매)
다음달 가격 하락 (최소 구매)	리스크 없음	리스크 없음	리스크 높음 (물류능력 제한)
다음달 가격 상승 (최대 구매)	리스크 보통 (잔여기간 구매 축소)	리스크 보통	리스크 없음

→ 다음달 가격이 하락에서 상승으로 예측되는 경우에 대한 상황에 집중

**(3) 핵심 상황에 대한 세부 계획 수립**

\* 가격 예측 모형 정확도, 재고수준 고려

- 1회차: 최소구매 / 2회차: 최대구매 예측 상황



- \* 확률 기반 의사결정: 과거 실적을 기준으로 각 의사결정에 기대 이익, 손실에 확률을 곱한 뒤 합산
- \* 위험회피 의사결정: 손실이 최소화 되는 의사결정 시행(재고를 중간 수준으로 유지 등)

〈Figure 5〉 가격 예측 모형과 기타 요인을 고려한 구매 시나리오 예시

et al., 2013). 구체적으로 〈Figure 5〉와 같이 가격 예측 모형을 기반으로 하고 물류지원 능력, 재고수준 등의 기타 핵심 의사결정 요인을 활용하여 의사결정 시나리오를 구성한 후, 다양한 구매 상황에 대한 대책을 제시할 수도 있을 것이다. 따라서, 향후 연구에서는 기계학습 기반 가격 예측 모형과 시나리오 플래닝 의사결정 방법을 통합하는 연구를 시도할 필요가 있다.

**참고문헌**

김선미, 조두연(2022), "Machine Learning과 Google Trends Data를 이용한 유가 예측 및 분석," *The Korean Journal of Economics*, 제29권 2호, pp. 175-193.

(Kim, S., and Cho, D.(2022), "Forecasting Crude Oil Prices with Google Trends Data Based on Machine Learning Methods," *The Korean Journal of Economics*, 29(2). pp.175-193.)  
 김정욱(2006), "기름값 너무 올라 공군 훈련 차질," *중앙일보*, 2006.03.28.  
 (Kim, J.(2006). "Oil prices rise too high, disrupting air force training," *The Joongang*, 2006.03.28., <https://www.joongang.co.kr/article/1658535#home> ('24.1.30.))  
 노산하, 김남현(2019), "유가 불확실성이 국내 경제변수에 미치는 영향," *국제경제연구*, 제25권 1호, pp.1-38.  
 (Noh, S., and Kim, N.(2019), "The Effects of Oil Price Uncertainty on Korean Economic Variables," *International Business Journal*, 25(1), pp.1-38.)  
 선미선, 이성문, 이혁수, 송병규, 오병훈, 권남연, 이미지, 정재원(2016), "무기체계 획득 시 운영유지비 분석

- 체계 발전방안,” 한국국방연구원 연구보고서, 운 2016-3882.
- (Sun, M., Lee, S., Lee, H., Song, B., Oh, B., Kwon, N., Lee, M., and Jung, J.(2016), “A Study on the Development of Operation Maintenance Cost Analysis System in the Acquisition of Weapon Systems,” *KIDA Research Report*, 3882.)
- 송경재(2021), “미 항공유 부족에 항공사들 몸살,” *파이낸셜뉴스*, 2021.07.28.
- (Song, K.(2021), “Airlines are struggling due to US aviation fuel shortage,” *Financial News*, 2021.07.28., <https://www.fnnews.com/news/202107280452179467> (24.1.30.))
- 이서진, 유종민(2022), “예측오차의 확률분포에 기반한 국제유가 불확실성 분석,” *에너지경제연구*, 제21권 1호, pp. 87-113.
- (Lee, S., and Yu, J.(2022), “Causes and Consequences of Oil Price Uncertainty,” *Korean Energy Economic Review*, 21(1), pp.87-113.)
- 이인애(2022), “VLSFO 가격 750-770 달러 사상 최고: 러-우크라이나 정치외교상황 영향, 정치가 석유 펀더멘털 앞서 원유가 8 년만의 최고수준, HSFO 가격도 7 년전 수준으로 올라,” *해양한국*, 제2022권 3호, pp.33-33.
- (Lee, I.(2022), “VLSFO price \$750-770, all-time high: Russia-Ukraine political and diplomatic situation influenced, politics ahead of oil fundamentals Crude oil is at the highest level in 8 years, HSFO price also rose to the level of 7 years ago,” *Maritime Korea*, 2022 (3), pp.33-33.)
- 조지성, 이연경, 권보배, 김아름, 김엄지(2022), “글로벌 공급망 리스크별 영향분석 및 대응방안 연구 - 에너지 · 곡물을 중심으로,” [KMI] 연구보고서.
- (Jo, J., Lee, E., Kwon, B., Kim, A., and Kim, U. (2022), “A Study on Impact Analysis and Response Measures of Global Supply Chain Risks - Focusing on Energy and Crops,” [KMI] *Research Report*.)
- 한국석유공사(2023), “2023년 국제유가 전망, 2023년 1월 주간 유가전망,” *Petronet*.
- (Korea National Oil Corporation(2023), “international oil price forecast for 2023, weekly oil price forecast for January 2023,” *Petronet*.)
- Amer, M., Daim, T. U., and Jetter, A.(2013), “A review of scenario planning,” *Futures*, 46, pp.23-40.
- Asche, F., Gjøølberg, O., and Völker, T.(2003), “Price relationships in the petroleum market: an analysis of crude oil and refined product prices,” *Energy Economics*, 25(3), pp.289-301.
- Avanijaa, J.(2021), “Prediction of house price using xgboost regression algorithm,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(2), pp.2151-2155.
- Bojer, C. S., and Meldgaard, J. P.(2021), “Kaggle forecasting competitions: An overlooked learning opportunity,” *International Journal of Forecasting*, 37(2), 587-603.
- Boone, T., Ganeshan, R., Hicks, R. L., and Sanders, N. R.(2018), “Can Google trends improve your sales forecast?,” *Production and Operations Management*, 27(10), pp.1770-1774.
- Breiman, L.(2001), “Random forests,” *Machine Learning*, 45, pp.5-32.
- Caldara, D., and Iacoviello, M.(2022), “Measuring geopolitical risk,” *American Economic Review*, 112(4), pp.1194-1225.
- Chen, T., and Guestrin, C.(2016), “Xgboost: A scalable tree boosting system,” In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*:

- August 13, 2016: San Francisco, California, USA, pp.785-794.
- Chou, Y. C., Chuang, H. H. C., Chou, P., and Oliva, R.(2023), "Supervised machine learning for theory building and testing: Opportunities in operations management," *Journal of Operations Management*, 69(4), pp.643-675.
- Chai, J., Guo, J. E., Meng, L., and Wang, S. Y. (2011), "Exploring the core factors and its dynamic effects on oil price: an application on path analysis and BVAR-TVP model," *Energy Policy*, 39(12), pp.8022-8036.
- EIA(Energy Information Administration, US) (2022), "SHORT-TERM ENERGY OUTLOOK".
- Faff, R. W., and Brailsford, T. J.(1999), "Oil price risk and the Australian stock market," *Journal of Energy Finance & Development*, 4(1), pp.69-87.
- Gao, S., and Lei, Y.(2017), "A new approach for crude oil price prediction based on stream learning," *Geoscience Frontiers*, 8(1), pp.183-187.
- Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J. J., and Naranpanawa, A. (2019), "Long-term forecast of energy commodities price using machine learning," *Energy*, 179, pp.214-221.
- Jabeur, S. B., Mefteh-Wali, S., and Viviani, J. L. (2021), "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Annals of Operations Research*, pp. 1-21.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., and Liu, T. Y.(2017), "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing systems*, 30.
- Kraljic, P.(1983), "Purchasing Must Become Supply Management," *Harvard Business Review*, September-October, pp.109-117.
- Nagar, D., Raghav, S., Bhardwaj, A., Kumar, R., Singh, P. L., and Sindhwani, R.(2021), "Machine learning: Best way to sustain the supply chain in the era of industry 4.0," *Materials Today: Proceedings*, 47, pp.3676-3682.
- Jurado, K., Ludvigson, S. C., and Ng, S.(2015), "Measuring uncertainty," *American Economic Review*, 105(3), pp.1177-1216.
- Li, R.(2010), "The Relationship among petroleum prices," *International Conference on Applied Economics-ICOAE*, pp.459-463.
- Miao, H., Ramchander, S., Wang, T., and Yang, D. (2017), "Influential factors in crude oil price forecasting," *Energy Economics*, 68, pp.77-88.
- Simon, A., Deo, M. S., Venkatesan, S., and Babu, D. R.(2016), "An overview of machine learning and its applications," *International Journal of Electrical Sciences & Engineering*, 1(1), pp.22-24.
- Sharma, H., Harsora, H., and Ogunleye, B.(2024), "An Optimal House Price Prediction Algorithm: XGBoost," *Analytics*, 3(1), pp.30-45.
- Wei, G., Zhao, J., Feng, Y., He, A., and Yu, J.(2020), "A novel hybrid feature selection method based on dynamic feature importance," *Applied Soft Computing*, 93, Article 106337.
- Younis, H., Sundarakani, B., and Alsharairi, M. (2022), "Applications of artificial intelligence and machine learning within supply chains: systematic review and future research directions," *Journal of Modelling in Management*,

17(3), pp.916-940.  
Zhang, Z., Bouri, E., Klein, T., and Jalkh, N.(2022),  
“Geopolitical risk and the returns and vola-

tility of global defense companies: A new  
race to arms?,” *International Review of  
Financial Analysis*, 83, Article 102327.

- 저자 임세환은 현재 연세대학교 경영대학 Operations Management 전공 석사과정에 재학 중이다. 공군사관학교 국방경영학과를 졸업하였으며, 공군 군수분야에 복무 중이다. 주요 연구 관심분야는 물류서비스품질, 공급사슬 리스크 관리, 운영관리에서의 예측이다.
- 저자 민순홍(Ph.D., Haslam College of Business, the University of Tennessee)은 연세대학교 경영대학 오퍼레이션/공급사슬 담당 교수이자 ESG/기업윤리연구센터 부센터장으로 일하고 있다. 주요 연구/강의 분야는 지속가능경영, 공급사슬관리 (SCM), B2B 관계 전략 등이다. *International Journal of Logistics Management*에 발표한 논문으로 'Emerald 최우수논문상', 생산관리학회지에 발표한 논문으로 '현우 광수일 생산관리 학술상', 국제적으로 인용지수 높은 학술지 논문으로 '한국로지스틱스대상 학술상' 및 연세대학교 우수업적교수상 연구부문 최우수상을 수상하였다.
- 저자 최경환은 방위사업청에서 근무 중이다. 국방대학교를 졸업하였으며, 국방관리 석사와 군사 운영분석 박사를 취득하였다. 그후에는 방위사업청에서 F-35A, F-15K 등과 같은 전투기 사업을 담당해 왔고, 2020년부터 2년간 국방대학교에서 교수를 역임하기도 했다. 주요연구분야는 공급체인관리, 군수, 방위산업 등이다.

### 〈부록〉 산술평균 예측방법과 기계학습 방식의 예산감축 효과 비교(예상)

단위: (단가) 천원/BBL, (재무성과) 천원

구분	① 적용 가격	② 최적 구매규모	③ 산술평균 예측방법 (매월 15일, 산술평균값 활용)			④ ML활용 (매월 1일)		⑤ 산술평균 방법에서 시기 조정 (매월 1일, 산술평균값 활용)		
			예측 단가	구매 의사결정 최대/불일치	재무 성과(예상)	재무 성과(예상)	예측 단가	구매 의사결정 최대/불일치	재무 성과(예상)	
○년 1월	91	최소	90	최대/불일치	-91,221	-182,441	89	최대/불일치	-182,441	
○년 2월	88	최소	93	최소	315,110	630,220	94	최소	630,220	
○년 3월	75	최소	76	최소						
○년 4월	48	최소	60	최소						
○년 5월	26	최대	30	최대	238,284	476,567	35	최소/불일치	-476,567	
○년 6월	36	최대	27	최대						
○년 7월	50	최대	49	최대						
○년 8월	53	최소	52	최소	33,240	66,479	50	최대/불일치	-66,479	
○년 9월	51	최소	52	최소						
○년 10월	46	최대	46	최대						
○년 11월	48	최대	49	최대	36,544	73,089	51	최대	73,089	
○년 12월	51	최대	49	최대						
○+1년 1월	59	최대	57	최대						
○+1년 2월	64	최대	62	최대	20,183	40,366	61	최대	40,366	
○+1년 3월	72	최대	70	최대						
○+1년 4월	75	최소	77	최소						
○+1년 5월	75	최대	72	최대	149,096	298,191	77	최소	298,191	
○+1년 6월	81	최대	80	최대						
○+1년 7월	85	최대	83	최대						
○+1년 8월	88	최소	88	최소	61,050	122,099	87	최대/불일치	-122,099	
○+1년 9월	86	최대	86	최대	187,958	375,916	89	최대	375,916	
○+1년 10월	93	최대	89	최대						
○+1년 11월	110	최소	108	최소						
○+1년 12월	105	최소	109	최소	116,837	233,673	101	최대/불일치	-233,673	
○+2년 1월	99	최대	97	최대	388,057	776,115	110	최소	330,814	
○+2년 2월	114	최대	109	최대						
○+2년 3월	127	최대	125	최대						
○+2년 4월	163	최대	163	최소/불일치	-72,131	-144,263	133	최대	776,115	
○+2년 5월	166	최대	159	최대						
○+2년 6월	181	최대	184	최대						
○+2년 7월	209	최소	204	최소	395,332	790,663	170	최대	1,665,515	
○+2년 8월	176	최소	182	최소						
○+2년 9월	174	최소	162	최대/불일치	-163,049	-326,099	171	최대		1,665,515
○+2년 10월	168	최대	180	최대						
○+2년 11월	176	최소	174	최소	832,757	1,665,515	186	최소	1,665,515	
○+2년 12월	165	최소	175	최소						
합 계 (3년)					3,468,283 (34.7억원)	6,936,565 (69.4억원)				4,044,535 (40.4억원)
연 평균					1,156,094 (11.6억원)	2,312,188 (23.1억원)			1,348,178 (13.5억원)	

- 기본조건(가정사항)

- 매월 구매량을 동일(가정: 연간 예산 3,000억원, 평균가격: 3개년 평균 월간 적용가격) 하게 운영했을 경우와 대비하여, 다음달의 예측가격과 현재가격을 비교하여 구매량을 차이를 두었을 때를 비교  
\* 재무성과:  $(P_{t+1월} - P_{t월}) \times (Q_{t월구매량} - Q_{월평균소요})$ , P: 실제가격, Q: 구매량  
\* ex) ○년 1월에 현재가격(91) 대비, 2월에는 상승할 것으로 예측(93)하여 최대 구매 의사결정하였다 가정할 수 있으나, 실제로는 하락(88)하여 최적 의사결정은 최소라고 볼 수 있음.
- 항공유 가격은 실제 가격 데이터를 기준(등유×환율)이며, ○년 이후, 3년 간의 재무적 성과 파악(예상)
- 구매 의사결정이 이전달과 동일하면 재무적 손익은 첫 달 한번만 계산
- 항공유의 월간 수요량은 일정하며, 저장용량의 증·감 등 물류요소의 변동은 없다고 가정
- 구매 의사결정은 가격 예측 결과에만 따름(기타 정성적 요소는 고려하지 않음)

- 각 열별 세부내용

- 계약조건에 따라, 전월 일일가격의 평균치
- 다음월의 적용 가격 대비, 현재 가격이 낮으면 '최대', 높으면 '최소' 구매가 최적 구매 방안임.
- ③-①: 산술평균 방법과 기계학습 방법은 예측 정확도 수준은 유사하다고 가정(RMSE 수준이 3,000원/BBL 내외)하나, 예측시기에 기계학습 방법이 매월 1일, 산술평균 방법은 매월 15일로 기계학습 방법은 구매량 조정 규모를 산술평균 방법 대비, 2배로 운영할 수 있다고 가정함.  
\* 가정사항: (산술평균 방법) 매월 구매량 동일 대비 10% 조정 가능 / (기계학습 방법) 20%까지 조정 가능
- ⑤ 산술평균 방법 적용시기로 매월 1일 앞당기는 경우의 성과 추가 비교