

양파 가격 예측 고도화를 위한 인공지능(AI) 활용방안 연구

김원태 · 문홍성 · 이미숙



양파 가격 예측 고도화를 위한 인공지능(AI) 활용방안 연구

김원태 · 문홍성 · 이미숙



연구 담당

김원태 | 전문위원 | 연구 총괄, 제1~6장 집필

문홍성 | 전문연구원 | 제3~5장 집필

이미숙 | 전문연구원 | 제2장 집필

특별연구 E2026-03

양파 가격 예측 고도화를 위한 인공지능(AI) 활용방안 연구

등 록 | 제6-0007호(1979. 5. 25.)

발 행 | 2026. 6.

발 행 인 | 한두봉

발 행 처 | 한국농촌경제연구원

우) 58321 전라남도 나주시 빛가람로 601

대표전화 1833-5500

인 쇄 처 | 크리커뮤니케이션

I S B N | 979-11-6149-850-8 95520

※ 이 책에 실린 내용은 한국농촌경제연구원의 공식 견해와 반드시 일치하는 것은 아닙니다.

※ 이 책에 실린 내용은 출처를 명시하면 자유롭게 인용할 수 있습니다.

무단 전재하거나 복사하면 법에 저촉됩니다.

최근 농산물 가격은 기후변화, 재배면적 변동, 생산량 변화, 저장 및 유통 여건, 소비패턴 변화 등 다양한 요인에 의해 변동성이 확대되고 있다. 특히 양파는 계절성과 저장성이 강한 농산물로 기상 여건과 수급 상황의 변화에 따라 가격 변동 폭이 크게 나타나는 품목이다. 따라서 시장 변화에 선제적으로 대응할 수 있는 정확하고 신뢰성 높은 가격 예측 체계의 구축이 그 어느 때보다 중요해지고 있다.

농업관측은 생산자와 소비자, 유통업체, 정책 담당자에게 시장 정보를 제공하고 합리적인 의사결정을 지원하는 중요한 역할을 수행하고 있다. 그러나 기존의 통계 기반 예측기법만으로는 가격 변동을 충분히 설명하거나 예측하는 데 한계가 있다. 따라서 다양한 데이터를 효과적으로 활용하고 복잡한 비선형 관계를 학습할 수 있는 인공지능(AI) 기술을 농업관측에 접목하려는 관심이 높아지고 있다.

본 연구는 양파 가격 예측의 정확도를 향상시키기 위해 시계열 모형과 인공지능 기법을 활용한 예측모형을 구축하고, 기존 예측방식과의 성능을 비교·분석하여 농업관측 업무에 적용 가능한 활용 방안을 제시하는 것을 목적으로 수행되었다.

본 연구의 결과는 양파 가격 예측의 정확도를 높이는 데 기여할 뿐만 아니라, 인공지능 기반 농업관측체계 구축을 위한 기초자료로 활용될 것으로 기대된다. 아울러 향후 마늘, 배추, 무 등 다른 주요 채소류 품목으로 인공지능(AI) 기반 예측 기법을 확대 적용하는 데에도 유용한 참고자료가 될 것으로 기대한다.

2026. 6.

한국농촌경제연구원장 한 두 봉

연구 목적

- 최근 기후변화의 심화, 생산 및 유통환경의 변화, 소비패턴의 다양화 등으로 양파 가격의 변동성이 확대되면서 정확한 가격 예측의 중요성이 더욱 커지고 있음. 기존 농업관측의 가격 전망은 기상변화와 시장환경의 복합적인 변동을 충분히 반영하는 데에는 한계를 보이고 있음. 특히 공급·수요뿐만 아니라 저장, 유통, 소비, 거시경제 요인 등이 복합적으로 작용하는 최근의 시장 환경에서는 보다 정교한 예측기법의 도입이 요구되고 있음.
- 이에 본 연구는 인공지능(AI) 기반 가격 예측기법을 활용하여 양파 가격 예측의 정확도를 향상시키고, 기존 가격 전망체계를 보완할 수 있는 예측모형을 개발하는 것을 목적으로 수행되었음. 또한 다양한 데이터의 활용 가능성을 검토하고, 실제 농업관측 업무에 적용 가능한 인공지능 기반 가격 예측 체계의 구축 방향과 정책적 활용 방안을 제시하고자 하였음.

연구 방법

- 본 연구에서는 2003년 1월부터 2026년 4월까지의 월별 양파 도매가격 자료를 구축하고, 가격에 영향을 미치는 다양한 설명변수를 함께 수집하여 분석하였음. 주요 설명변수는 국내 양파 반입량, 수입 양파 반입량, 소비자물가지수, 기상자료, 저장 관련 변수 등으로 구성하였으며, 양파의 생육 특성과 저장·유통 구조를 반영할 수 있도록 데이터를 전처리하였음.

- 가격 예측모형은 기존 시계열 모형과 인공지능 모형을 함께 구축하여 성능을 비교하였음. 비교 대상은 Seasonal Naive, SARIMA, SARIMAX, XGBoost, SARIMA-XGBoost 및 STL-ARIMA-XGBoost 등 총 6개 모형이며, RMSE, RMSPE, MAE, MAPE 및 MDA 등 다양한 예측 성능지표를 활용하여 모형의 예측력을 평가하였음. 또한 SHAP를 이용하여 예측 결과에 영향을 미치는 주요 변수의 중요도를 분석하고, 인공지능 모형의 설명가능성을 검토하였음.

연구 결과

- 기존 농업관측의 가격 전망은 전반적으로 양호한 수준의 예측력을 유지하고 있으나, 저장양파에서 햇양파로 전환되는 시기와 조생양파 출하 시기에는 예측오차가 상대적으로 크게 나타나는 것으로 분석되었음. 이는 기상요인과 저장물량, 출하 시기 조절 등 다양한 요인이 복합적으로 작용하기 때문으로 판단됨.
- 탐색적 데이터 분석 결과 양파 가격과 주요 설명변수 간에는 비선형적인 관계가 존재하는 것으로 나타났음. 특히 동일한 공급량 변화라도 저장 여건과 시장 상황에 따라 가격 반응이 달라지는 특성이 확인되어, 선형모형만으로는 가격 변동을 충분히 설명하기 어려운 것으로 분석되었음.
- 예측모형의 성능을 비교한 결과 머신러닝 기법의 XGBoost모형과 시계열 모형과 머신러닝을 결합한 하이브리드 모형이 우수한 예측력을 보였음. 점

추정에서는 XGBoost의 예측력이 가장 우수하였으나, 수확기인 3월과 5월의 예측력은 하이브리드 모형이 더 우수하였음. 따라서 양파 도매가격 예측에는 XGBoost모형과 하이브리드모형의 예측치를 다양하게 검토할 필요가 있음.

- SHAP 분석 결과 국산 양파 반입량, 저장 관련 변수, 소비자물가지수 및 기상요인이 가격 변동에 큰 영향을 미치는 것으로 나타났음. 특히 일부 기상 변수는 특정 시기와 조건에서 가격에 중요한 영향을 미치는 것으로 분석되어, 인공지능 모형이 복잡한 가격 결정요인을 효과적으로 반영하고 있음을 확인하였음.

시사점

- 기존 통계모형과 전문가 판단 중심의 가격 전망체계에 인공지능(AI) 기반 예측모형을 보완적으로 활용하는 하이브리드 농업관측체계를 구축할 필요가 있음. 인공지능은 전문가의 판단을 대체하기보다 객관적인 예측 정보를 제공하여 의사결정을 지원하는 도구로 활용되어야 함.
- 인공지능 예측모형의 성능 향상을 위해서는 학습데이터의 품질관리와 데이터 표준화를 강화해야 함. 조사자료의 정확성을 높이고, 기상자료, 저장 정보, 유통정보 및 경제지표 등을 통합적으로 관리할 수 있는 데이터 관리 체계를 구축할 필요가 있음.

- 예측 결과의 신뢰성과 활용성을 높이기 위해 SHAP와 같은 설명 가능한 인공지능 기법을 농업관측 업무에 적극 활용할 필요가 있음. 이를 통해 가격 예측 결과뿐만 아니라 가격 변동의 주요 원인을 함께 제시함으로써 정책 담당자의 이해도와 의사결정의 투명성을 높일 수 있음.
- 본 연구에서 개발된 인공지능 기반 가격 예측모형은 양파뿐 아니라 마늘, 배추, 무 등 주요 채소류 품목으로 확대 적용하여 농업관측 전반의 예측 정확도를 제고할 필요가 있음. 또한 위성영상, 드론영상, IoT 센서, 온라인 유통정보 등 다양한 디지털 데이터를 단계적으로 연계하여 미래형 인공지능 기반 농업관측체계를 구축하여야 함.
- 앞으로 Transformer 계열의 최신 시계열 딥러닝 모형과 실시간 데이터 연계 기술을 활용하여 예측모형을 지속적으로 고도화하고, 인공지능 기반 가격 예측 체계를 정부의 수급 안정 정책과 연계함으로써 보다 신속하고 과학적인 농산물 수급관리 기반을 구축할 필요가 있음

A Study on the Application of Artificial Intelligence (AI) to Improve Onion Price Forecasting

Purpose of Research

- In recent years, the importance of accurate onion price forecasting has increased as price volatility has intensified due to climate change, changes in production and distribution environments, and increasingly diverse consumption patterns. Conventional agricultural outlook systems have shown limitations in adequately reflecting the complex impacts of weather variability and rapidly changing market conditions. In particular, the current market environment is influenced not only by supply and demand but also by storage, distribution, consumption, and macroeconomic factors, highlighting the need for more sophisticated forecasting methodologies.
- Accordingly, this study aimed to improve the accuracy of onion price forecasting by applying artificial intelligence (AI)-based forecasting techniques and to develop a prediction model that complements the existing agricultural outlook system. In addition, the study examined the applicability of diverse data sources and proposed a framework for establishing an AI-based price forecasting system that can be practically implemented in agricultural outlook services, together with its policy implications.

Research Method

- This study constructed a monthly dataset of wholesale onion prices covering the period from January 2003 to April 2026 and collected a wide range of explanatory variables affecting price movements. The major explanatory variables included domestic market arrivals, import volumes, the Consumer Price Index (CPI), weather data, and storage-related variables. The data were preprocessed to reflect the biological characteristics of onion production as well as the storage and distribution structure of the onion market.
- To evaluate forecasting performance, both conventional time-series models and AI-based models were developed and compared. The models included Seasonal Naive, SARIMA, SARIMAX, XGBoost, SARIMA-XGBoost, and STL-ARIMA-XGBoost. Forecasting performance was assessed using multiple evaluation metrics, including RMSE, RMSPE, MAE, MAPE, and MDA. Furthermore, SHAP (SHapley Additive exPlanations) was employed to identify the relative importance of explanatory variables and to improve the interpretability of the AI-based forecasting models.

Research Results

- The existing agricultural outlook system generally demonstrated satisfactory forecasting performance. However, relatively larger

forecasting errors were observed during the early-season onion harvest period and the transition from stored onions to newly harvested onions. These forecasting difficulties appear to result from the combined effects of weather conditions, storage inventories, and shipment timing.

- Exploratory data analysis revealed nonlinear relationships between onion prices and the major explanatory variables. In particular, the market response to changes in supply varied depending on storage conditions and market circumstances, indicating that conventional linear models alone were insufficient to fully explain price fluctuations.
- The comparison of forecasting model performance showed that the XGBoost machine learning model and the hybrid model, which combines time series and machine learning approaches, exhibited superior predictive performance. In terms of point forecasting accuracy, the XGBoost model outperformed the other models. However, the hybrid model achieved higher forecasting accuracy during the harvest season, particularly in March and May. Therefore, for forecasting wholesale onion prices, it is necessary to consider the predictions generated by both the XGBoost and hybrid models in a complementary manner.

-
- The SHAP analysis indicated that domestic market arrivals, storage-related variables, the Consumer Price Index, and weather factors were among the most influential determinants of onion price fluctuations. In particular, several weather variables were found to exert significant impacts under specific seasonal and market conditions, demonstrating that the AI-based models effectively captured the complex mechanisms underlying price formation.

Implications

- A hybrid agricultural outlook system that integrates AI-based forecasting models with conventional statistical models and expert judgment should be established. Rather than replacing expert knowledge, AI should serve as a decision-support tool by providing objective and data-driven forecasting information.
- Improving the performance of AI forecasting models requires stronger quality management of training data and enhanced data standardization. It is essential to improve the accuracy of survey data and establish an integrated data management system that effectively combines weather data, storage information, distribution data, and macroeconomic indicators.
- Explainable AI techniques such as SHAP should be actively incorporated into agricultural outlook services to enhance the

reliability and usability of forecasting results. Presenting not only forecast values but also the major factors driving price changes will improve policymakers' understanding and promote greater transparency in decision-making.

- The AI-based forecasting model developed in this study should be extended beyond onions to other major vegetable crops, including garlic, Chinese cabbage, and radish, in order to improve the overall accuracy of agricultural outlook services. Furthermore, next-generation agricultural forecasting systems should progressively integrate various digital data sources, including satellite imagery, drone imagery, IoT sensor data, and online distribution information.
- Future research should continue to enhance forecasting performance by adopting state-of-the-art Transformer-based time-series deep learning models and real-time data integration technologies. Moreover, AI-based price forecasting systems should be closely linked with government supply-demand stabilization policies to establish a more timely, scientific, and evidence-based framework for agricultural market management.

Researchers: KIM Wontae, MUN Hongsung, LEE Misook

Research period: 2026. 2. - 2026. 6.

E-mail address: wtkim@krei.re.kr

제1장 서론

1. 연구의 필요성과 목적	1
2. 연구 내용 및 방법	4
3. 선행연구 검토 및 차별성	8

제2장 양파 가격 예측 체계 분석

1. 양파 가격 예측 체계	15
2. 양파 가격 전망치 산출 방식	17
3. 양파 가격 전망치 평가와 한계	19

제3장 양파 가격 예측 데이터 구축

1. 데이터 및 변수	25
2. 탐색적 데이터 분석	28

제4장 양파 가격 예측 모형 구축

1. 단일 예측 모형	36
2. 하이브리드 예측 모형	38
3. 예측 및 평가 설계	40

제5장 양파 가격 예측 모형 평가

1. 모형 적합 진단	45
2. 예측력 평가	49
3. 가격 예측 기여 효과	53

제6장 인공지능 기반 관측 고도화 방안

1. 가격 예측 모형 분석 요약	59
2. 인공지능 기반 관측체계 구축 방향	62
3. 정책적 시사점 및 향후 과제	66

부록

1. ARIMA 모형 차수 탐색 결과	73
2. 모형별 잔차 자기상관함수	75
3. 양파 예측 모형 운영 프로세스	77

참고문헌	83
------------	----

제2장

〈표 2-1〉 양파 공급량 전망을 위한 주요 조사 내용	16
〈표 2-2〉 연도별 양파 실제가격과 전망가격 간 오차율	19
〈표 2-3〉 최근 5개년 시기별 양파 가격 오차율	21
〈표 2-4〉 양파 가격 실제치와 전망치 간 오차 원인 분석	21

제3장

〈표 3-1〉 변수 정의 및 출처	27
〈표 3-2〉 기초통계량	28
〈표 3-3〉 양파 도매가격 단위근 검정	31
〈표 3-4〉 종속변수와외 상관계수	32
〈표 3-5〉 분산팽창계수(VIF)	33

제4장

〈표 4-1〉 평가지표 비교	43
-----------------------	----

제5장

〈표 5-1〉 ARIMA 차수 결과	46
〈표 5-2〉 SARIMA 추정 결과	46
〈표 5-3〉 STL-ARIMA 추정 결과	47
〈표 5-4〉 잔차 검정	47
〈표 5-5〉 근 안정성 결과	48
〈표 5-6〉 예측력 비교	50

〈표 5-7〉 월별 예측력 비교	52
〈표 5-8〉 Diebold-Mariano 검정 결과	53
〈표 5-9〉 SARIMAX 추정 결과	55

부록

〈부표 1-1〉 SARIMA 차수 탐색	73
〈부표 1-2〉 SARIMAX 차수 탐색	74
〈부표 1-3〉 STL-ARIMA 차수 탐색	75
〈부표 3-1〉 양파 예측 모형 모듈 구성	78
〈부표 3-2〉 모형 학습 명령어	80

제2장

〈그림 2-1〉 양파 가격 전망 체계	17
----------------------------	----

제3장

〈그림 3-1〉 양파 도매가격 추세	29
〈그림 3-2〉 양파 도매가격 STL 분해	30
〈그림 3-3〉 양파 도매가격 자기상관함수와 편자기상관함수	30
〈그림 3-4〉 변수 상관행렬	32

제5장

〈그림 5-1〉 모형별 예측력 지표 시각화	50
〈그림 5-2〉 실제치와 예측치 시각화	51
〈그림 5-3〉 모형별 SHAP 중요도	57

부록

〈부도 2-1〉 SARIMA 잔차의 ACF, PACF	75
〈부도 2-2〉 SARIMAX 잔차의 ACF, PACF	76
〈부도 2-3〉 STL-ARIMA 잔차의 ACF, PACF	76
〈부도 3-1〉 config 파일 설정(평가용과 예측용)	79
〈부도 3-2〉 원자료 입력	80

1

서론

1. 연구의 필요성과 목적

1.1. 연구 배경과 필요성

○ 양파는 국내 채소류 중에서도 대표적인 가격 변동 품목으로, 생산과 소비 구조의 특성상 수급 불균형이 빈번하게 발생하는 특징을 가짐. 특히 양파는 노지 재배 작물로서 기상 조건에 크게 의존하며, 재배면적과 생산량의 변동이 가격에 직접적인 영향을 미치고 있음. 또한 저장이 가능한 품목이라는 점에서 출하시기 조절에 따라 단기적인 가격 변동성이 더욱 확대되는 경향이 있음.

- 이러한 구조적 특성으로 인해 양파 가격은 계절적 요인뿐 아니라 기상, 재배 의사결정, 저장 및 유통 행태 등 다양한 요인의 복합적인 영향을 받아 불안정한 흐름을 보임.

○ 최근에는 이상기후 현상의 빈번한 발생으로 인해 생산 변동성이 더욱 확대되

고 있으며, 이에 따라 가격의 급등락 현상이 반복적으로 나타나고 있음. 특히 특정 시기에는 공급 과잉으로 가격이 급락하여 생산자 소득이 크게 감소하는 반면, 반대로 공급 부족 시에는 가격 급등으로 소비자 부담이 증가하는 등 시장 불균형 문제가 심화되고 있음.

- 이러한 가격 변동성은 단순한 시장 문제를 넘어 농가 경영 안정성과 소비자 후생, 그리고 정부의 수급 관리 정책 전반에 영향을 미치는 중요한 문제로 인식되고 있음.

○ 이와 같은 상황에서 양파 가격에 대한 정확한 예측은 생산자, 유통업자, 정책당국 모두에게 필수적인 의사결정 수단으로 작용함. 생산자는 정식 및 재배면적을 결정하는 데 있어 가격 전망에 의존하며, 유통업자는 저장 및 출하 전략을 수립하기 위해 가격 예측 정보를 필요로 함.

- 정책당국은 수급 조절, 비축 및 방출, 수입 관리 등의 정책을 적시에 수행하기 위해 신뢰성 있는 가격 예측 체계를 요구하고 있음. 따라서 가격 예측의 정확도는 시장 안정성과 정책 효과성을 좌우하는 핵심 요소라 할 수 있음.

○ 농업관측센터에서 수행하고 있는 양파 가격 예측은 가락도매시장 반입량과 도매가격을 변수로 추정하는 회귀모형, 신축성 계수를 이용한 추정, 시계열 모형 등 다양한 방법으로 예측하고 있음.

- 그러나 이러한 방식은 수급 상황이 안정적일 때는 높은 예측력을 보일 수 있으나 급격한 시장충격이 발생할 경우 예측력이 저하될 수 있다는 한계를 보이고 있음.

○ 또한 최근 농산물 시장은 과거와 달리 구조적 변화가 빈번하게 발생하고 있음. 소비 패턴의 변화, 유통 구조의 다양화, 수입 확대, 기후 변화 등 다양한 요인이 동시에 작용하면서 가격 형성 메커니즘이 더욱 복잡해지고 있음.

- 이러한 환경에서는 단일 시계열 모형만으로는 시장의 동태적 변화를 충분히 설명하기 어려움. 따라서 보다 유연하고 복합적인 분석이 가능한 새로운 접근방법이 요구됨.

○ 이러한 한계를 보완할 수 있는 대안으로 인공지능(AI) 기반 분석기법이 주목받고 있음.

- 머신러닝 및 딥러닝 모형은 변수 간 비선형 관계를 학습하고, 다양한 외생변수를 동시에 고려할 수 있다는 점에서 기존 시계열 모형의 한계를 극복할 수 있는 잠재력을 지니고 있음. 특히 대규모 데이터 처리와 패턴 인식 능력을 바탕으로 복잡한 가격 변동 구조를 보다 정교하게 반영할 수 있음.

○ 그러나 인공지능 모형 역시 모든 문제를 해결하는 만능 도구는 아니며, 시계열 데이터가 가지는 추세 및 계절성과 같은 구조적 특성을 충분히 반영하지 못하는 경우도 존재함. 또한 해석 가능성 측면에서도 한계가 지적되고 있음.

- 따라서 최근에는 전통적인 시계열 모형과 인공지능 모형의 장점을 결합한 하이브리드 접근법이 효과적인 대안으로 제시되고 있음.

○ 이에 본 연구는 양파의 월별 도매가격 예측을 연구 대상으로 하되, 전통적인 시계열 모형 및 수급 모형과 인공지능 기법을 결합한 하이브리드 가격 예측 모형을 적용함으로써 데이터 제약하에서도 양파 도매가격 예측의 안정성을 제고하고 내부 수급 분석 및 정책 검토에 활용 가능한 예측 체계를 마련하고자 함.

1.2. 연구 범위 및 목적

○ 본 연구의 대상 품목은 양파이며, 가격 변수는 가락동 도매시장 양파 도매가격임. 연구의 시간적 범위는 최근 10년 이상 월별 자료이고, 예측 범위는 월별 가격임.

○ 본 연구의 목적은 양파 도매가격 변동의 구조적 특성과 비선형 요인을 동시에 반영하는 인공지능 기반 하이브리드 예측 모델을 개발하고, 이를 내부 수급 분석 및 정책 검토에 활용 가능한 형태로 제시하는 데 있음. 구체적인 목적은 다음과 같음.

- 첫째, 양파 도매가격의 월별 계절성, 추세, 자기상관 구조를 반영한 기본 예측 모형 구축
- 둘째, 기존 시계열 모형과 인공지능 보정 모형을 결합한 하이브리드 예측 체계 구축
- 셋째, 수급·기상·정책 요인을 반영한 월별 가격 변동 위험 예측 고도화
- 넷째, 품목별 수급 분석 및 정책 검토에 활용 가능한 월별 가격 전망 정보 제공

2. 연구 내용 및 방법

2.1. 연구 내용

○ 제2장에서는 기존 양파 가격 예측 체계를 분석하고 그 한계를 도출하고자 함. 현재 양파 가격 예측은 주로 단순 시계열 모형 또는 전문가 및 품목 담당자의

판단에 의존하는 경우가 많으며, 데이터 기반의 체계적인 예측 시스템은 제한적으로 운영되고 있음. 특히 예측 과정에서 수급 변수, 정책 변수 등이 충분히 반영되지 못하는 문제가 존재함.

- 또한 기존 예측 방식은 평균적인 가격 흐름을 설명하는 데에는 유용하나, 급격한 가격 변동이나 구조적 변화에 대응하는 데 한계가 있음. 이에 따라 보다 정교하고 유연한 예측 체계의 구축이 필요함.
- 농업관측센터에서 양파 가격을 예측하는 데 사용하고 있는 모형, 추정 방법 등을 정리하고 각 방법의 오차율을 정리한 후 개선방안을 도출함.

○ 제3장에서는 본 연구의 분석 데이터인 양파 가격을 예측하기 위한 데이터를 구축하고 인공지능 분석을 위한 데이터 전처리를 수행함.

- 먼저 공급측 요인으로 도매시장 국내 양파 반입량, 수입 양파 반입량을 사용하였음.
- 물가 요인으로는 소비자물가지수를 사용하였으며, 내생성을 피하고 전반적인 물가 통제와 실질 구매력을 대변하기 위해 전체 소비자물가지수를 선택하였음.
- 기상요인은 냉해, 봄철 생육 수분, 강수량 등을 사용하였으며, 양파 생육기와 월별 유통 실태를 반영하여 기상 시계열을 재가공하여 사용하였음.
- 양파의 저장이 진행될수록 한계저장비용이 체증하는 구조를 반영하기 위해 2차 함수 형태의 저장비용 구조 변수도 사용하였음.
- 또한 탐색적 데이터 분석을 통해 양파 도매가격 예측을 위한 비선형 모형 기법 도입의 필요성을 확인하고, 양파 가격 예측을 위한 6개 모형을 선정하였음.

○ 제4장에서는 탐색적 데이터 분석을 통해 벤치마크 모형으로 Seasonal Naive 를 두고, 단일 예측모형으로 시계열 모형인 SARIMA, SARIMAX, 머신러닝 기법의 XGBoost 등 3개의 모형을 선정하였음. 또한, 하이브리드 모형은 SARIMA-XGBoost, STL-ARIMA-XGBoost로 구성하여, 선형과 비선형 모형에 따른 예측력을 살펴보았음.

- 종속변수인 양파 도매가격은 로그변환하였으며, 설명변수는 국산 양파 반입량, 수입 양파 반입량, 전체 소비자물가지수, 냉해 관련 기상자료, 봄철 생육기 수분, 수확기 부패 요인, 저장비용 구조 등 7개로 구성하였음.
- XGBoost 모형은 설명변수 외에 1, 2, 3, 12개월 양파 가격, 3, 6, 12개월의 이동평균 및 이동표준편차를 생성하여 모형에 입력하였으며, 예측주기는 1개월, 계절주기는 12개월로 설정하였음.

○ 제5장에서는 양파 가격 예측 모형의 성능을 확인하기 위해 확장 윈도우 기반 롤링-오리진 평가 방법을 사용하였음. 비선형 모형의 변수 기여도를 살펴보기 위해 SHAP 변수 중요도를 추정하여 제시하였음.

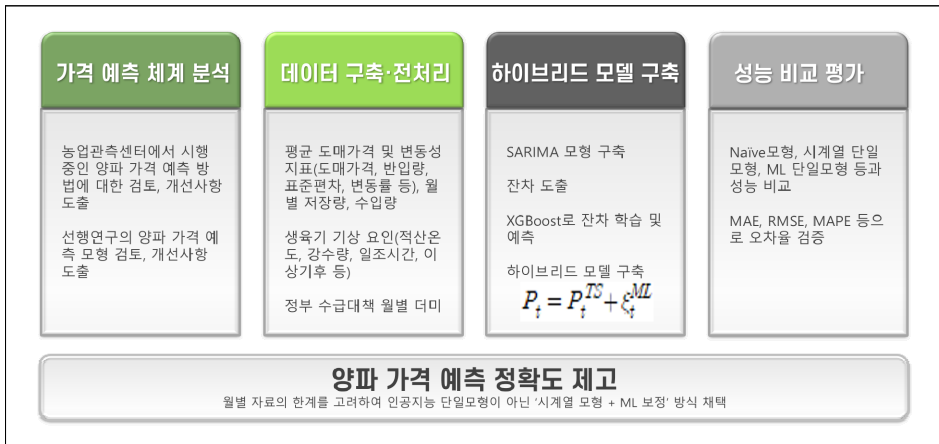
- 각 모형의 예측력은 RMSE, RMSPE, MAE, MAPE, MDA를 사용하여 종합적으로 평가하였음.

○ 제6장에서는 연구 결과를 요약하고 정책적 시사점을 제시하였으며, 향후 과제 및 인공지능 기반 관측체계 구축 방향을 단기, 중기, 장기로 구분하여 제시하였음.

2.2. 연구 방법

- 양파 재배면적, 생산량, 도매가격 관련 데이터 수집, 기상변수, 수입량, 저장량, 대체재 가격 등 양파 가격에 영향을 미치는 변수들의 데이터 수집, 양파 가격 예측 방법 등과 관련된 선행연구를 국내외 문헌조사를 통해 수행함.
- 양파 시계열 자료를 이용한 도매가격 예측 모델을 구축하고 양파의 도매가격을 예측한 뒤, 인공지능을 활용하여 오차항에 대한 학습을 시행함. 오차항을 구성하는 주요 변수를 XGBoost로 학습한 후 오차항을 추정하여 시계열 모형으로 예측한 값과 결합하는 하이브리드 모형을 구축함.
- 양파 하이브리드 모형 이외에도 다양한 비교모형을 추정한 후, 평균절대오차(MAE), 평균제곱근오차(RMSE), 평균절대백분율오차(MAPE), 대칭 평균절대백분율오차(sMAPE) 등을 이용하여 각 모형의 성능을 평가하고 검증하며, 최적의 양파 가격 예측모형을 선정한 후 품목 관측팀에서 사용 가능하도록 매뉴얼을 작성함.

2.3. 연구 추진 체계



자료: 저자 작성.

3. 선행연구 검토 및 차별성

3.1. 선행연구 검토

○ 박제인·정수진(2024)은 양파의 일별 도매가격 예측을 위해 ARIMA, SARIMA를 비롯하여 LSTM, LightGBM, XGboost 등 12개의 모형을 5대 도매시장별로 추정하여 각 모형의 성능을 비교 분석하였음. 분석 결과 대구와 부산에서 ARIMA와 SARIMA 모형이 우수한 성능을 보인 반면, 인공지능 기반 모델들은 상대적으로 저조한 성능을 나타냈음. 이는 각 지역의 데이터 특성과 변동성에 알맞은 최적의 예측 모형 선택의 중요성을 강조하고 있음.

○ 신성호 외(2018)는 대파, 양파, 애호박을 대상으로 다양한 기상변수와 경유, 물가상승률, 전년 수확량, 전년 수입량, 전년 재배면적 등 기타 변수를 대상으

로 LSTM을 이용하여 일별 소매가격을 예측하였음. 예측 결과 LSTM모형이 시계열 데이터 예측 모형이나 일반 회귀분석 모형보다 상대적으로 높은 예측력을 보였음.

- 홍성민·김종화(2023)는 VAR를 이용하여 양파의 주산지를 4개 권역으로 구분한 후, 각 권역별 산지가격 간 인과성을 분석하였음. 분석 결과 양파 생산량이 가장 많은 전라권이 아닌 저장업체가 가장 많은 경상권이 산지가격을 선도하고 있음을 보였음. 이를 통해 경상권 산지가격 정보를 토대로 수급정책을 수립하여야 하며, 전라권에는 저장량 확대를 위한 인프라 구축이 필요함을 주장하였음.
- 양진석 외(2020)는 벡터오차수정모형(Vector error correction model; VECM)을 이용하여 양파의 품종 간 가격 인과관계를 분석하였음. 양파는 극조생, 조생, 중생, 중만생, 만생 등의 작형 순으로 수확이 진행되는 바, 극조생 양파 가격은 조생과 이후 중만생 양파의 가격에 영향을 미치는 것으로 분석되었음. 또한 전년 중만생 양파 생산량은 극조생 양파의 가격에 영향을 미치는 것으로 나타나 양파 수급정책 수립 시 제주도 양파 가격에 대한 선제적 조치의 필요성을 주장하였음.
- 이도영 외(2021)는 농산물 가격 변동성을 완화하기 위해 양파와 대파를 대상으로 Lasso 회귀분석을 활용한 최적 변수 선정과 LSTM 기반 가격 예측모형을 개발하였음. 기존 선행연구에서 활용된 가격·거래 관련 변수를 기반으로 Lasso 회귀분석을 적용하여 예측에 유의한 변수를 선정하고, 선정 전·후 LSTM 모형의 예측 성능을 비교하였음. 분석 결과, 최적 변수를 적용한 모형에서 양파의 1일, 3일, 5일, 7일 후 가격 예측 RMSE가 평균 0.0229 감소하여 예측 정확도가 향상되는 것으로 나타났음.

- 김진·이정일(2019)은 농산물 가격 정보만을 활용한 강화학습 기반의 예측 모형을 제시하였음. 연구진이 제안한 하이브리드 가격 예측 시스템은 LSTM 기반 정책신경망과 강화학습을 결합한 형태로 기존의 시계열 모형이나 다변량 예측 기법에 비해 가격 변화 패턴을 적절하게 반영하는 예측 성능을 보였음.
- 박종현 외(2023)는 Direct, Hybrid, Multiple Outputs의 다단계 시계열 예측(Multi-step Time Series Forecasting)과 더불어 LightGBM, 다층 퍼셉트론(Multi-Layer Perceptron; MLP), LSTM, GRU(Gated Recurrent Unit) 등 다양한 머신러닝·딥러닝 모형형 적용하여 품종 및 예측 기간별 최적의 조합을 분석하였음. 분석 결과 Direct와 Recursive 모형의 장점을 결합한 Hybrid 모형이 MAPE 기준 가장 우수한 예측력을 보였으며, 특히 양파는 MLP와 Hybrid 모형의 결합이 중장기 예측에서 효과적인 것으로 나타났음.
- 윤민정 외(2024)는 농산물 가격 예측을 위해 경매가격, 기상, 생산 및 경제 변수를 활용한 딥러닝 기반 예측모형(LSTM, GRU)을 개발하였음. 분석 결과 GRU 모형이 LSTM보다 우수한 예측 성능을 보였으며, 특히 양파 가격 예측에서 높은 정확도를 나타냈음. 또한 기상 변수와 함께 국고채, M2 증가율, 경제불확실성지수(EPU), 생산자물가지수 등 경제 변수를 추가할 경우 예측력이 향상되는 것으로 확인되어 농산물 가격 예측 시 다양한 외부 요인을 반영한 인공지능 기반 모형이 수급 관리 및 가격 안정화에 활용될 수 있음을 제시하였음.
- Patil et al.(2023)은 농산물 가격 예측을 위해 ARIMA, SARIMA, SVR, XGBoost, LSTM 등 다양한 시계열 및 머신러닝 모형을 비교하고, 이 중 SARIMA와 LSTM의 장점을 결합한 하이브리드 모형(HySALS)을 제안하였음. SARIMA와 LSTM 모형은 MAPE 기준 4~8%로 다른 모형에 비해 우수한

예측력을 보였으며, 이를 결합한 HySALS 모형은 계절성, 추세, 비선형 패턴을 동시에 효과적으로 반영하는 것으로 나타났다.

○ Gu et al.(2022)은 농산물 가격 변동성을 예측하기 위해 Dual Input Attention-LSTM(DIA-LSTM) 모형을 제안하였음. 기존 LSTM 모형이 과거 가격 중심으로 학습하는 한계를 보완하기 위해 기상자료, 거래량 등 다양한 외생 변수를 함께 활용하였으며, Attention 메커니즘을 통해 가격 변화에 영향을 미치는 주요 변수를 선택하도록 설계하였음. 연구 결과 제안모형은 기존 시계열 모형 대비 우수한 예측 성능을 보였으며, 농산물 가격 예측에서 기상·시장 정보를 통합하는 것이 중요함을 제시하였음.

○ Kannan & Patel(2024)은 다양한 요인으로 인해 급격한 변동성을 보이고 있는 인도의 양파 시장 가격 예측을 위해 LSTM(Long Short-Term Memory) 모형을 사용하였음. 예측 결과 LSTM 모형은 88%의 예측 정확도를 기록하며 기존 모델보다 상대적으로 우수한 예측력을 보였음. 이 연구는 양파 가격 예측에 인공지능 기법을 본격적으로 적용한 사례이며, 향후 하이브리드 모형 개발과 장기 시계열 데이터 활용의 필요성을 제기하였다는 점에서 의의가 있음.

○ Tran et al.(2023)은 머신러닝과 딥러닝을 활용한 농산물 가격 예측 연구 27편을 분석한 결과 LSTM을 포함한 신경망 기반 모형이 가장 많이 활용되었으며, ARIMA, SARIMA와 같은 시계열 모델, SVR, 앙상블 학습 기법도 함께 사용되고 있음을 보였음. 특히 최근 연구들은 단일 모형보다는 전통적 시계열 모형과 LSTM, GRU, Attention 모형 등을 결합한 하이브리드 모형이 비선형성과 복잡한 시계열 패턴을 효과적으로 반영할 수 있다는 점을 주목하였음.

○ Zhang et al.(2024)은 미국 아보카도 가격 예측을 위해 TCN-MLP-Attention 기반의 융합 딥러닝 모형을 제안하였음. 2015~2018년 미국 아보카도 판매

자료 약 5만 건을 활용하여 판매량, 가격, 지역, 기상정보, 품종 등의 변수를 분석에 활용하였음. TCN은 가격 변화의 장·단기 시계열 패턴을 추출하고, MLP는 변수 간 비선형 관계를 학습하며, Attention은 주요 영향 요인에 가중치를 부여하는 구조로 설계하였음. 분석 결과 제안모형은 RMSE 1.23, MSE 1.51의 우수한 예측 성능을 보여 기존 통계 및 머신러닝 기반 모형보다 높은 정확도를 나타냈음.

○ Manogna et al.(2025)은 23개 농산물의 장기 일별 도매가격 자료를 활용하여 ARIMA와 같은 전통적 시계열 모형, SVR, XGBoost, MLP와 같은 기계학습 기법과 함께 RNN, LSTM, GRU, ESN 등 다양한 딥러닝 기법의 예측 성능을 비교한 결과, LSTM과 GRU 등 딥러닝 기반 모형이 양파를 포함한 변동성 높은 농산물 가격 예측에서 가장 우수한 성과를 보인다고 보고하였음.

○ Choudhary et al.(2025)은 농산물 가격의 비선형성과 비정상성을 고려하여 VMD(Variational Mode Decomposition)와 LSTM을 결합한 하이브리드 예측 모형을 개발하였음. 이 연구에서는 기존 단일 딥러닝 모형이 급격한 가격 변동을 충분히 반영하지 못하는 문제를 개선하기 위해 가격 시계열을 분해한 후 LSTM으로 예측하는 방법을 적용하였음. 결과적으로 제안모형은 기존 모형 대비 우수한 예측력을 보이며, 복잡한 농산물 가격 변동 분석에 적합함을 확인하였음.

3.2. 선행연구와의 차별성

○ 머신러닝 또는 딥러닝을 활용한 대부분의 농산물 가격 예측 선행연구는 부족한 학습용 데이터를 충족시키기 위해 일별 자료를 활용하여 진행되었음. 그러

나 농업관측센터에서 수행하는 월별 가격 예측에 인공지능을 활용하기 위해서는 월별 데이터를 사용할 수 밖에 없음. 따라서 본 연구는 계절성과 추세가 비교적 뚜렷한 양파의 월별 도매가격의 구조적 변동을 시계열 모형으로 설명하고, 기상 요인, 수급 불균형 등에 의해 발생하는 비선형 오차는 머신러닝의 일종인 XGBoost로 보완하는 하이브리드 모형을 적용한다는 점에서 선행연구와 차별성이 있음.

2

양파 가격 예측 체계 분석

1. 양파 가격 예측 체계

- 양파 가격 전망을 위해서는 월별 공급량 전망치가 필요하며, 해당 물량 산출을 위한 기초 자료는 다음의 수집 체계를 통해 생성됨.
- 농업관측사업 초기에는 표본농가와 모니터 조사 결과를 이용하여 수급전망을 시행하였음. 최근에는 조사방식 개선과 기술 발전 등에 따라 전문가를 활용한 산지동향 조사, 실측 조사, 항공촬영 등을 통한 자료 수집 방법이 추가되었음.
 - 표본농가 조사를 통해 품종별 재배의향면적, 정식시기, 재배면적, 생육상황, 단위면적당 수확량, 포전거래가격, 농가 저장량 등의 정보를 수집함.
 - 모니터 및 산지조사협회 조사에서는 시군별 재배면적, 육묘 상태, 정식시기, 생육상황, 병해충 발생 정도, 방제 동향, 기상재해, 단위면적당 수확량 등의 정보를 수집하고 있음.

- 저장업체 조사에서는 포전거래 및 계약재배 가격, 입고 예상량, 입고량, 입고가격, 출고량, 출고시기, 품위, 감모율 등의 정보를 수집하고 있음.
- 생육실측 및 항공촬영으로 재배면적, 생육상황, 생산량, 조생종 양파 수확면적 등의 정보를 수집하고 있음.
- 해외모니터와 수출입자료를 이용하여 중국의 양파 재배면적 및 생산량 전망, 산지 및 수출가격, 수출입량, 수출입액 등의 정보를 파악하고 있음.

○ 수집된 자료를 이용하여 재배면적과 단수 추정, 생산량을 전망함. 또한 표본 저장업체를 대상으로 입고량과 누적 출고량, 감모율 등을 조사하여 저장기간 동안의 출하량 추정에 활용하고 있음.

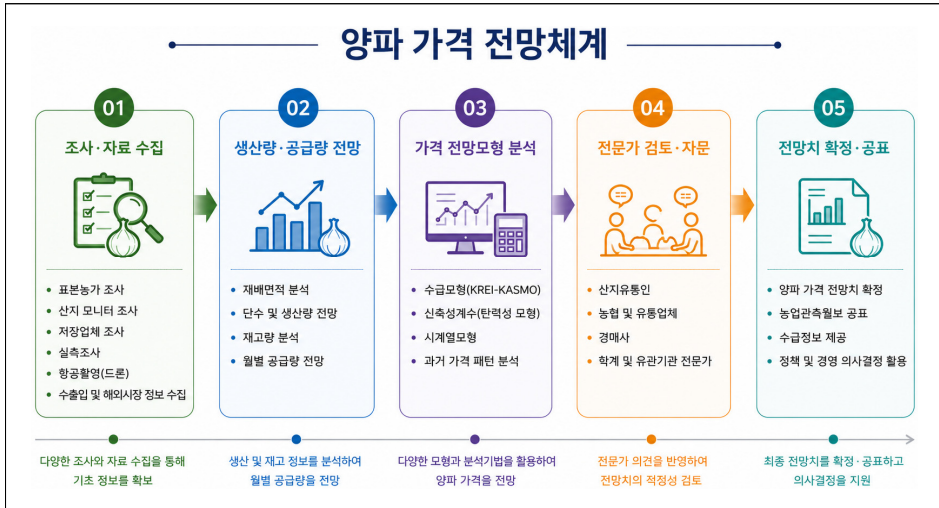
○ 당해 년의 생산량이 결정되면 품종별(조생종, 중만생종) 출하 비중, 월별 도매 시장 반입 비중, 재고량 등의 자료를 활용하여 월별 공급량 전망치를 산출하고 이를 토대로 가격 전망치를 산출함.

〈표 2-1〉 양파 공급량 전망을 위한 주요 조사 내용

구분	주요 조사 내용
표본농가	• 품종별 재배의향면적, 정식시기, 재배면적 • 생육상황, 단위면적당 수확량 • 포전거래가격, 농가저장량
모니터 산지조사협회	• 시군 재배면적, 정식시기, 육묘상태 • 생육상황, 병해충 발생 정도, 방제 동향, 기상재해, 단수
저장업체	• 포전거래 및 계약재배 가격, 입고예상량, 입고량, 입고가격 • 출고량, 출고시기, 품위, 감모율
실측조사	• 재배면적 : GPS 기반 표본농가 재배면적 실측 • 생육 : 초장, 엽초장, 엽수, 구직경, 1주 무게, 경엽무게, 병해충 등 • 생산량 : 조사구역(3.3㎡) 내 개체 수, 생구중(g), 구직경(mm)
항공촬영	• 드론을 활용한 주산지 재배면적 파악 • 조생종 양파 수확면적(제주 대정읍, 한경면, 전남 고흥 금산면)
해외모니터	• 중국 재배면적 및 생산량 전망, 저장량 및 입고가격, 산지거래동향, 수출가격
수출입자료	• 양파 형태별 수출입량, 수출입 금액

자료: 저자 작성.

〈그림 2-1〉 양파 가격 전망 체계



자료: 본 그림의 도식화 과정에서 생성형 AI 도구인 OpenAI(2025)를 보조적으로 활용함.

2. 양파 가격 전망치 산출 방식

- 농업관측센터는 수급모형, 신축성 계수(탄력성 모형), 시계열 모형, 과거 가격 패턴을 적용한 추정 등 다양한 방법으로 복수의 가격 전망치를 산출함.
- 공급과 수요의 균형에서 가격이 결정되는 수급모형은 가장 많이 사용되는 방법으로서 KREI-KASMO 모형이 대표적인 예임. 다만 월별 가격을 전망해야 하는 관측월보에 맞추어 모형을 구성하기 다소 어려운 측면이 있음.
- 양파 월별 가격 형성 구조를 설명하기 위해서는 품종별 출하시기, 수확입고 시기, 저장 출하기로 구분이 필요함. 공급 부분에서 출하 지연, 정부 정책 시행 등으로 공급량 변동이 예상될 경우 가격 영향을 신속하게 추정할 필요가

있음. 이러한 경우 과거 공급량 변화와 가격 변동 간의 관계를 활용하는 신축성 계수를 주로 활용하고 있음.

- (4~5월) 조생종 생산량, 저장양파 출하량, (6~7월) 중만생종 생산량, 초기 입고량, 출하가능물량, (8월~다음 해 3월) 입고량, 재고량, 재고율

○ 신축성 계수는 공급량이 변할 때 가격이 얼마나 변하는지 비율을 계산한 것으로 손쉽게 사용할 수 있음.

〈신축성 계수를 활용한 가격 예측 분석〉

① 신축성 계수 : 가격 전년 대비 등락률/공급량 전년 대비 증감률

전망 가격 등락률 = 신축성 계수 × 예상 공급량 증감률

② 가격 전망치 : $P_t = P_{t-1} + P_{t-1} \times (a_{t-1} \times Q^p)$

P_t =양파 가격 전망치, P_{t-1} =양파 전년 동월 가격,

a_{t-1} =전년 동월 신축성 계수, Q^p =공급량 전년 대비 증감률 전망치

○ 시계열 모형은 가격 데이터로 ARIMA, VAR, Prophet 같은 모형을 구축하여 미래 가격을 예측하는 방법임.

○ 가격 패턴 분석은 해당 시기와 생육 상황이나 생산량, 출하 동향, 가격 흐름이 유사한 시기의 가격 변화 패턴을 분석하여 적용하는 방법임. 당해 연도의 가격 형성이나 출하 경향이 특정 해와 유사한 패턴을 보일 것이라는 의견 등을 참고하여 적용해 볼 수 있음.

○ 위에서 열거한 분석 방법을 사용하여 양파 도매가격 전망 범위를 산출함. 이후 농업관측 프로세스에 따라 조사자료와 전문가 검토 결과를 종합하여 최종 전망가격을 공표함.

- 중앙자문회의는 농협, 산지유통인, 유통업체, 경매사 등 수급 여건을 실시간으로 파악하고 있는 자문위원 다수, 농촌진흥청, 학계 전문가 등도 포함되어 관측센터 조사 결과에 대한 적절성에 대해 판단할 수 있는 자리임.

3. 양파 가격 전망치 평가와 한계

3.1. 양파 가격 전망치 평가

○ 앞서 설명한 다양한 전망기법을 활용하여 산출된 양파 가격 전망치의 정확도를 평가하기 위해 최근 5개년(2021~2025년) 전망가격과 실제가격 간 오차를 분석하였음. 그 결과 평균 오차율(MAPE)은 약 9% 수준으로 나타났으나, 시기별로 편차가 있음.

- 평균 절대 백분율 오차(Mean Absolute Percentage Error: MAPE)는 실제값 대비 예측값의 평균 절대 백분율 오차를 나타내는 지표로 예측 정확도를 직관적으로 평가할 수 있음.
- 일반적으로 MAPE 값이 10% 미만일 경우 매우 높은 정확도를, 10~20%는 좋은 예측, 20~50%는 합리적 예측, 50% 이상은 부정확하여 낮은 예측력을 의미한다고 알려짐.

〈표 2-2〉 연도별 양파 실제가격과 전망가격 간 오차율

단위: %

구분	2021	2022	2023	2024	2025	5개년 평균
회계연도	8.3	9.6	10.3	6.9	7.5	8.5
연산연도	9.8	9.0	11.1	6.0	10.2	9.2

주: 회계연도는 양념채소 관측월보가 발행되는 3월~12월 평균, 연산연도는 4월~이듬해 3월 평균 기준임.
자료: 농업관측센터 내부자료.

- 월별 오차율을 살펴보면 특정 시기에 오차가 집중되는 특징이 나타남. 이는 양파 가격이 상대적으로 불안정한 단경기와 수확기에 가격 급등락이 빈번하게 발생하기 때문임.
- 단경기인 3월 오차율은 5년 평균 14.1%로 저장출하기(3~5%) 대비 높음.
 - 3월~4월 상순에는 전년산 저장양파 출하와 당해연도 햇양파(제주 조생종 중심) 출하가 동시에 이루어짐. 이에 따라 3~4월 가격에는 저장양파 재고량 뿐만 아니라 햇양파 작황과 출하시기 등이 동시에 영향을 미침.
- 저장성이 낮아 즉시 출하가 요구되는 조생종 성출하시기인 5월의 오차율은 21.9%로 상대적으로 크게 나타남.
- 6월에는 중만생종 양파 수확이 이루어지나 저장창고 입고작업에 따라 공급량이 조절되는 측면이 있어 평균 오차율은 13.7%로 5월 대비 낮은 편임.
 - 6월에는 수확 후 출하와 입고(저장)작업이 동시에 이루어짐. 향후 가격 전망의 방향성을 예측한 민간 저장업체의 입고의향 등에 따라 출하량 조절이 가능함.
 - 예컨대 많은 물량을 취급하는 민간 저장업체의 입고의향이 낮을 경우 수확 물량이 시장으로 일시에 출하되면서 가격 하락 폭이 클 수 있으며 반대로 입고의향이 높을 경우 예상보다 가격 상승 폭이 클 수 있음.

〈표 2-3〉 최근 5개년 시기별 양파 가격 오차율

단위: %

구분	저장, 햇양파	수확기				저장 출하기				
	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
2021	4.7	9.4	30.6	0.2	0.1	15.2	1.6	0.8	6.5	13.7
2022	19.6	0.0	16.1	36.7	0.4	4.1	9.8	4.4	3.1	2.0
2023	12.9	19.5	34.5	4.7	5.9	2.3	5.3	1.1	9.0	8.1
2024	20.8	11.8	5.1	3.3	9.5	2.6	5.6	7.4	0.9	1.6
2025	12.4	8.6	23.1	23.8	2.5	1.3	2.1	0.8	0.1	0.0
평균	14.1	9.9	21.9	13.7	3.7	5.1	4.9	2.9	3.9	5.1
평년	15.0	10.0	23.3	10.6	2.9	3.0	4.4	2.1	3.5	3.9
최소	4.7	0.0	5.1	0.2	0.1	1.3	1.6	0.8	0.1	0.0
최대	20.8	19.5	34.5	36.7	9.5	15.2	9.8	7.4	9.0	13.7

주: 오차율은 (전망가격-실제가격)/실제가격×100으로 도출하며, 절댓값 기준임.

자료: 저자 작성.

○ 〈표 2-4〉는 양파 가격 전망치와 실제치 간 오차가 10% 이상 발생한 특정 월을 대상으로 원인을 분석한 것임. 급격한 기상 변화로 조생종 양파 공급량(생산량, 출하량) 변동성이 커진 경우, 단경기 조생종 작황 변화에 따라 저장업체에서 출고 진행을 조절하여 공급량 변동성이 커지는 등 미리 예견할 수 없는 상황이 많았음.

- 조생종 양파는 월동 후 생육 재생기부터 수확기까지 생육 기간이 짧으므로 토양 내 수분과 기온 등에 따라 구 비대 차이가 큰 편임. 통상 조기 출하할수록 가격이 높으므로 잎이 쓰러지는 도복 시작 시점부터 시장 가격 추이를 고려하여 수확할 수 있음.

〈표 2-4〉 양파 가격 실제치와 전망치 간 오차 원인 분석

단위: 원/kg, %

연도	월	전망 가격	실제 가격	오차율	원인 분석(오차율이 10% 이상인 경우)
2021	5	750	574	30.6	조생종 생산량 전망치가 전월 대비 증가한데다, 수급대책 시행 지연됨에 따라 가격 하락
	8	750	884	15.2	저장업체 출하조절로 전월 대비 시장 출하량 감소
	12	850	747	13.7	재고량 증가 및 소비 부진

(계속)

연도	월	전망 가격	실제 가격	오차율	원인 분석(오차율이 10% 이상인 경우)
2022	3	500	418	19.6	조생종 생육이 회복되며 저장양파 출하 집중
	5	700	603	16.1	출하량 증가로 상중순 가격 큰 폭 하락 영향
	6	850	1,343	36.7	중만생종 생산량 감소와 저장업체의 높은 입고 의향
2023	3	1,400	1,608	12.9	재고량 적어 저장업체에서 출하 조절, 야간 기온 하락으로 구 비대 지연되어 제주 조생종 출하 지연
	4	1,000	1,242	19.5	재고량 감소, 4월 초 강우로 조생종 수확 지연
	5	750	1,146	34.5	4월 말 저온, 5월 초 집중호우로 조생종 출하 지연. 중순에는 병해 확산
2024	3	1,200	1,515	20.8	2~3월 저온과 일조량 부족으로 비대 부진하여 출하량 감소
	4	1,300	1,475	11.8	비대 부진하여 조생종 생산량 전망치 전월 대비 하향
2025	3	1,600	1,827	12.4	추가적인 가격 상승 기대로 저장업체 출고 조절, 저온과 강 우로 2025년산 조생종 양파 수확 지연
	5	1,000	812	23.1	전남지역에서 비로 수확 지연되며 상품성 낮은 극대구 생산 증가, 소비 부진으로 도매시장 미판매품 증가
	6	950	767	23.8	소비 부진, 품위 저하, 저장업체의 입고 지연 영향

주: 오차율은 (전망가격-실제가격)/실제가격×100으로 도출하며, 절댓값 기준임.

자료: 저자 작성.

3.2. 현행 가격 전망의 한계와 개선 필요성

○ 농산물은 일반 공산품과 달리 기상, 작황, 재고, 수입, 소비심리 등의 영향을 동시에 받기 때문에 현재의 전망 체계만으로는 일정 수준 이상의 오차 발생을 완전히 방지하기는 어려움. 주요 한계점은 아래와 같음.

○ 첫째, 앞서 언급한 공급량 중심 전망의 한계임. 월별 국산 양파 공급량 추정치는 일반적으로 평년 수준의 출하 행태(비중)를 가정하므로 기상에 따른 출하 시기 변동, 기민한 저장업체의 출하 조절 효과 반영이 어려움. 또한 소비 대체 관계에 있는 수입 양파의 유입 영향도 알 수 없음.

- 둘째, 분석 자료가 조사치 기반이므로 조사 시점과 월보 발행 시점 간의 시차로 인한 오차 발생 가능성. 또한 출하량 전망치와 실제 출하된 물량을 검증할 수 있는 통계가 없어 단경기와 같이 물량이 적은 시기에 공급량 전망치가 달라질 경우 가격 등락 폭이 더 커질 수 있음.
- 셋째, 수요 측면 반영이 부족함. 도매시장 반입량과 가격이 같은 방향으로 움직이는 예외적인 경우가 있음. 예컨대 품위나 품질 차이, 경기 침체, 재고 누적으로 중도매인 수요 감소 등 공급 외 요인의 영향을 받을 가능성이 있음. 또한 가구보다 산업 부문 소비 비중이 높은 특성상 외식 경기, 가공제조업체 수요, 소비 추세 변화 등을 고려해야 하나 최신의 정량적 자료 수집이 어려운 데다 해당 월에 미치는 영향을 명확히 구분하여 적용하기 어려움.
- 넷째, 기상 변수 활용이 부족함. 양파는 겨울 한파, 봄철 가뭄, 수확기 강우 등 기상 변화에 민감하게 반응함. 이는 생산량과 수확시기에 영향을 주어 공급가능물량을 결정하고, 상품성과 저장 중 감모율(부패 등)에도 영향을 미치는데 기존 모형에 반영이 쉽지 않았음.
- 다섯째, 시장심리 반영의 부재임. 실제 양파 시장은 생육 기간 중 과잉·과소 생산 예상, 재고 과잉·부족 우려, 정부 수매비축분 출하 우려, 수입 양파의 시장 대체 우려 등 기대효과에 의해 가격이 먼저 움직이는 경우가 많음.
- 향후 양파 가격 전망의 정확도 제고를 위해서는 공급량 중심 분석에서 벗어나 기상 변수, 저장업체 출하 행태, 수입 동향, 수요 관련 지표 등을 종합적으로 반영한 예측 체계 구축이 필요함. 또한 머신러닝 등 다양한 분석기법을 활용하여 비선형적 가격 변동 요인을 반영하고, 전망 결과에 대한 설명력을 강화하는 방향으로 지속적인 개선이 요구됨.

3

양파 가격 예측 데이터 구축

1. 데이터 및 변수

- 양파 가격 예측을 위해 공급 요인으로 도매시장 국산 양파 반입량, 수입 양파 반입량을 구성하였고, 물가 요인으로 소비자물가지수, 기상 요인으로 냉해, 생육 수분, 수확 부패, 재고 요인으로 저장비용 구조를 고려하였음.
- 공급요인으로는 국산 양파 반입량과 최근 양파 수입이 증가하면서 가격 하락 요인으로 나타나는 수입 양파 반입량으로 구성하였음.
 - 국산 양파 반입량은 국내 양파 생산량을 나타내는 것이 아닌 공급 대리 변수임. 이 국산 양파 반입량을 분해하면 양파 생산량과 저장 출고 의사결정, 출하 시기 선택, 도매시장 경유율 등으로 나타낼 수 있음.
 - 여기서 국산 양파 반입량의 분해요인 가운데 대부분이 가격 기대에 반응하는 형태로 국산 양파 반입량은 결국 시장 균형의 결과로 볼 수 있음.

- 따라서, 국산 양파 반입량은 인과 식별 목적이 아닌 예측 목적으로써 공급 측 여건을 반영하는 선행지표 또는 가격 변동에 대한 예측 정보를 제공하는 변수 등으로 한정함.

○ 소비자물가지수는 내생성을 피하기 위해 세부 품목이 아닌 전체 소비자물가지수를 고려하였음. 또한, 전반적인 물가 통제와 수요측 실질 구매력을 대변하기 위해 이용함.¹⁾

○ 기상요인은 기상 시계열을 그대로 활용하기보다는 양파 생육기와 각 월별로 유통되는 양파에 대응해 재가공하여 예측에 활용하였음.

- 양파가 출하가 시작되는 4월을 기준으로 이전에는 전년 생육기의 기상, 이후에는 당해 년 생육기의 기상을 활용하여 구성하였음. 또한, 주산지인 무안 지역을 대상으로 함.
- 냉해를 나타내는 기상 cold는 가을 정식 이후 12~2월의 저온의 영향을 설정하여 1월부터 3월은 재작년 12월부터 작년 2월의 평균 최저기온의 평균값으로 재가공하고, 4월부터는 작년 12월부터 올해 2월까지의 평균 최저기온의 평균값으로 산출하여 변수를 생성함.
- precip은 봄철 생육 수분을 나타내며, 1월부터 3월 가격 예측에는 작년 3월부터 5월의 강수량의 합을, 4월은 당해 년 3월의 강수량, 5월은 3~4월의 강수량 합계, 6월부터 12월은 당해 년 3~5월의 강수량 합계를 산출하여 변수를 생성하였음.
- 수확기 부패 관련 기상으로는 강수량 데이터를 이용하였고, 1~3월 가격 예

¹⁾ 소비자물가지수는 관측월보 예측치 산출 시점인 각 월 하순에 발표되지 않고 있음. 따라서, 실제 예측치 산출 환경에서는 관측이 가능하고, 소비자물가지수와 0.65의 높은 상관관계를 갖는 국내 유가(자동차용 경유)를 대리변수로 활용하는 것이 좋을 것으로 판단됨.

측에는 작년 5~6월의 강수량 합계를 이용하고, 7월부터는 저장양파이므로 당해 년 5~6월의 강수량 합을 설정하였음. 이때, 4~6월은 데이터 누수 방지를 위해 0 값으로 설정함.

○ decay 변수는 양파 저장비용 구조를 나타내는 변수로 저장이 진행될수록 한계저장비용이 체증하는 구조로 반영하기 위해 2차 함수 형태를 동일 t기로 구성하였음.²⁾ 즉, 시간 순서를 7월은 1, 8월은 2, 다음 해 3월은 9로 값을 두고 이를 제공하여 시계열을 구성하였음. 다른 월의 값은 0임.

- 양파는 4~6월 수확 이후 다음 해 3월까지 저장된 양파를 소비하는 형태이므로 7월부터 다음 해 3월까지 양파 재고량에 따른 가격이 형성됨.
- 이 변수를 통해 저장 초기에는 감모와 비용 등이 적지만 후기로 갈수록 품질 하락과 비용 부담이 커지는 비선형적 구조를 포착하기 위해 설정함.

〈표 3-1〉 변수 정의 및 출처

변수	내용	단위	출처	비고
p	양파 도매가격	원/kg	서울특별시농수산식품공사	-
q	도매시장 국산 양파 반입량	톤	농업관측통계시스템	-
import_q	수입 양파 반입량	톤	농업관측통계시스템	-
cpi	소비자물가지수	-	한국은행 경제통계시스템	-
cold	냉해 관련 기상	℃	기상청 기상자료개방포털	평균 최저기온 파생
precip	생육 수분 관련 기상	mm	기상청 기상자료개방포털	월 합계 강수량 파생
rot	수확 부패 관련 기상	mm	기상청 기상자료개방포털	월 합계 강수량 파생
decay	양파 저장비용 구조	-	저자 계산	2차 함수 근사

자료: 저자 작성.

²⁾ 본 연구는 Working(1949)과 Brennan(1958)에서 시간이 지남에 따라 저장의 한계비용이 체증한다는 근거를 바탕으로 단순 볼록 형태인 2차 함수로 근사하여 변수를 설정하였음.

2. 탐색적 데이터 분석

○ 데이터는 취득 용이성과 파생변수 설정에 따라 2003년 1월부터 2026년 4월
까지의 280개의 시계열을 이용하였음. 추정과 학습에 224개의 훈련 데이터
를 활용하였고, 56개 테스트 데이터로 예측력을 평가하였음.

○ 예측 모형에 활용한 변수의 기초통계량은 <표 3-2>와 같음. 도매가격, 국산
양파 반입량, 수입 양파 반입량 등이 비대칭적, 두꺼운 꼬리 분포 등을 나타내
어 비선형 모형의 도입 필요성을 확인함.

- 양파 도매가격은 평균 923.9이고 표준편차는 359.7로 변동성이 큰 것을
확인하였음. 또한, 최댓값 2,342.4와 최솟값 274.6의 범위는 데이터 기간
내 가격 급등과 급락이 있었음을 시사함.
- 가격, 반입량, 재고비용 구조 등이 우측으로 치우쳐져 있는 편이고, 수입 반
입량은 강한 비대칭 분포를 보임. 그 외 변수들은 완만한 추세의 특성을 확
인하였음.

<표 3-2> 기초통계량

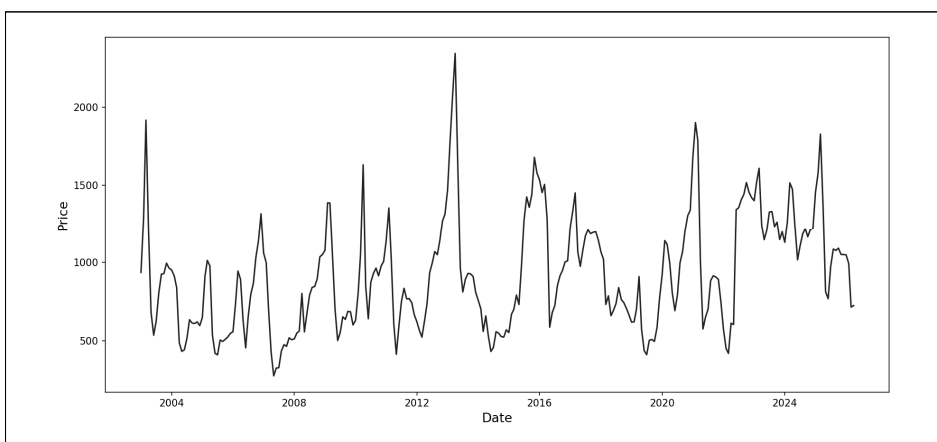
변수	평균	표준편차	최솟값	중앙값	최댓값	왜도	첨도
p	923.9	359.7	274.6	894.6	2,342.4	0.75	0.44
q	16,220.8	3,662.0	8,418.0	15,476.0	28,978.0	0.77	0.47
import_q	886.4	1,632.3	0.0	0.0	8,497.0	1.99	3.65
cpi	93.0	13.3	68.8	94.3	119.4	0.02	-0.82
cold	-2.1	1.1	-4.1	-2.0	0.3	0.28	-0.04
precip	219.7	96.1	19.0	215.5	492.5	0.52	0.88
rot	174.5	134.6	0.0	188.5	458.0	0.21	-0.91
decay	24.1	27.0	0.0	16.0	81.0	0.88	-0.56

자료: 저자 작성.

○ 양파 도매가격 추세는 <그림 3-1>에 나타냈음. 양파 도매가격 트렌드를 보면 급등락이 반복되어 선형 모형으로 포착하기 어려운 비선형적 구조가 존재할 수 있음을 확인함.

- 장기적인 추세는 보이지 않으나 평균 근처에서 상승과 하락이 주기적으로 나타나고, 2013년 4월이 가장 높은 가격을 기록하였음.

<그림 3-1> 양파 도매가격 추세

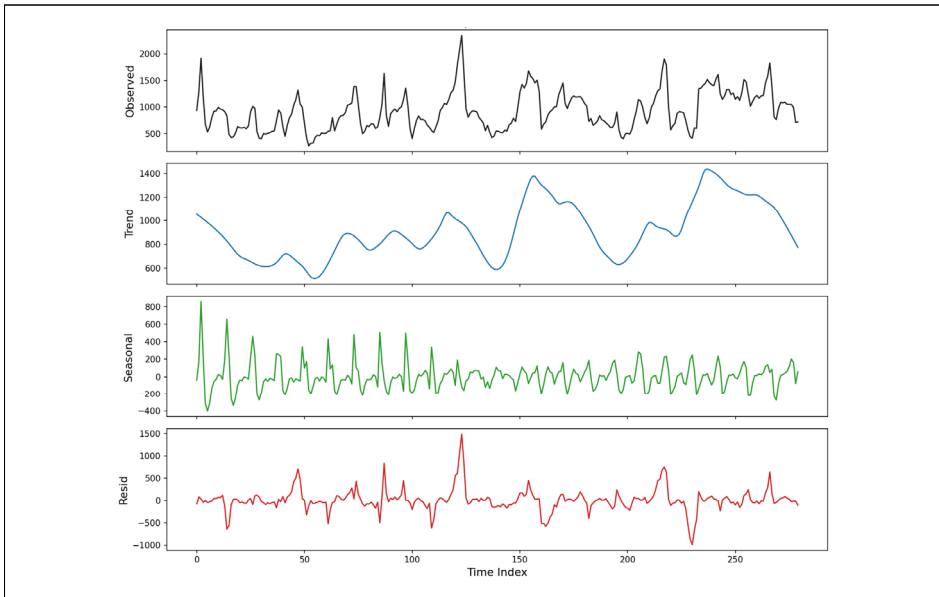


자료: 저자 작성.

○ 다음은 양파 도매가격의 STL 분해 결과를 <그림 3-2>에 제시하였음. 그 결과 선형적 분해 이후에서도 비선형성과 급등락 성분이 존재하는 것을 확인해 비선형적 모형 도입이 필요함을 시사함.

- 추세성분은 장기적 추세보다는 연 단위 주기로 순환적 변동을 보이는 것을 확인함.
- 계절성분은 우선 연 단위 주기로 반복적 패턴을 나타내나 그 진폭이 최근에 축소되고 있는 것으로 나타났음.
- 잔차성분의 경우 대체로 적게 나타났으나 특정 시기를 중심으로 큰 충격들이 존재하는 형태로 나타나 비선형적 구조를 보이는 것을 포착하였음.

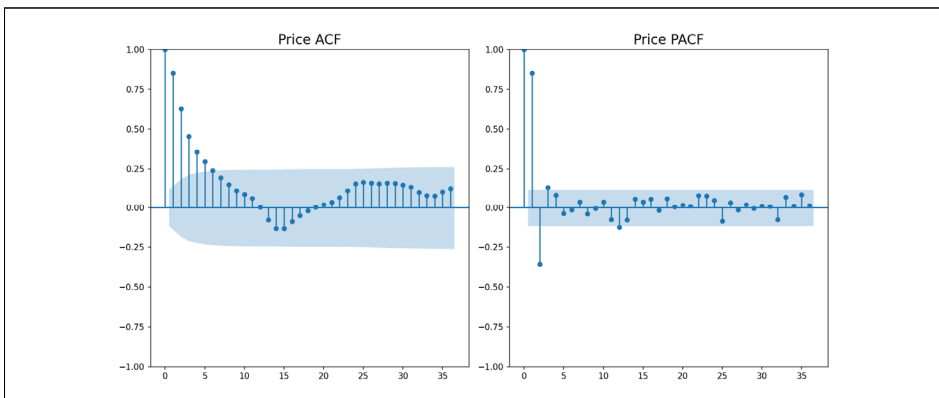
〈그림 3-2〉 양파 도매가격 STL 분해



자료: 저자 작성.

○ 양파 도매가격의 자기상관관계를 파악하기 위해 ACF와 PACF를 〈그림 3-3〉과 같이 나타냄. ACF와 PACF를 보면 AR(2) 구조에 가까우며, 완전한 12개월 주기의 변동이 존재하는 것을 확인함.

〈그림 3-3〉 양파 도매가격 자기상관함수와 편자기상관함수



자료: 저자 작성.

○ <표 3-3>은 종속변수인 양파 도매가격의 단위근 검정 결과임. 검정 결과 원계열을 차분하지 않고, 원자료 형태로 활용하고자 함.

- ADF, PP, KPSS 검정의 결과는 모두 귀무가설을 기각하여 결과가 상충하는 것으로 나타났음. 추가로, pmdarima의 일반 차분 차수와 계절 차분 차수 진단 결과 모두 0으로 나타나 차분이 불필요한 것으로 확인됨.
- KPSS를 제외한 ADF, PP, 차분 차수 진단이 일관되게 정상성을 확인함에 따라 본 연구에서는 원자료를 차분하지 않은 상태로 예측에 활용함.

<표 3-3> 양파 도매가격 단위근 검정

단위근 검정	statistic	p-value
ADF	-3.891	0.002
PP	-4.265	0.001
KPSS	0.759	0.012

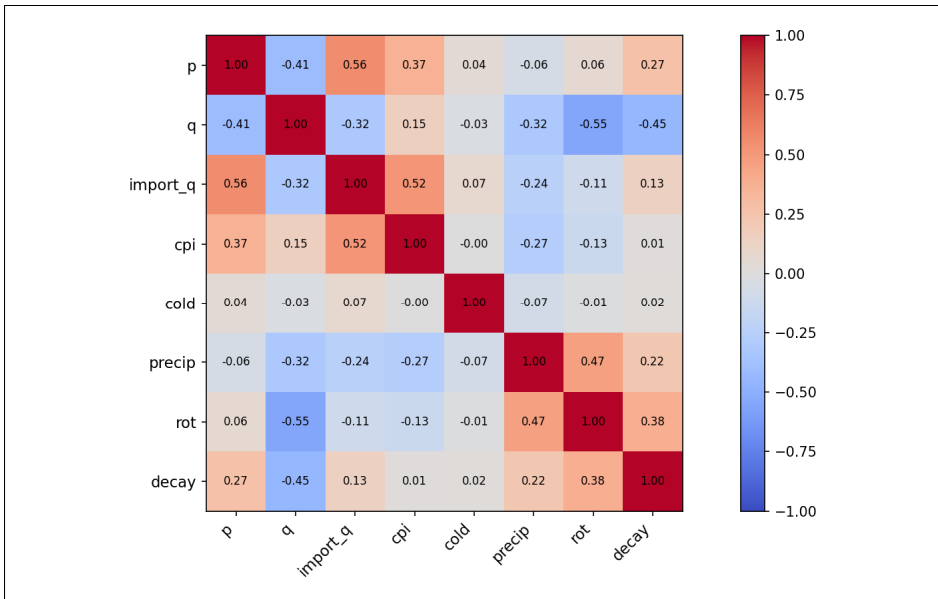
자료: 저자 작성.

○ <그림 3-4>는 변수 간 상관관계, <표 3-4>는 종속변수와 t 기, $t-1$ 기의 독립변수의 상관관계를 나타낸 것임. 대부분의 변수는 상관관계를 보였으나, 기상 변수는 상대적으로 낮은 상관계수를 나타냈음. 이는 기상요인이 양파 도매가격에 지속적으로 영향을 미치는 것이 아니라 생육 및 수확기 이상기온, 강수량 부족·과다 등 특정 조건에서 큰 영향을 미치는 특성이 있기 때문으로 판단됨. 따라서 단순 선형 상관분석에서는 관계가 약하게 나타나지만, 이러한 비선형 특성은 트리 기반 인공지능 모형에서 효과적으로 반영될 수 있음.

- 가격과 가장 높은 상관관계를 보이는 변수는 수입 양파 반입량으로 t 기와 $t-1$ 기에 각각 0.56, 0.47을 나타냈으며, 국산 양파 반입량은 -0.41과 -0.44로 나타남. 국산 양파 반입량은 경제학적 이론에 부합하나 수입 양파 반입량은 내생성을 내포할 수 있음.
- 소비자물가지수는 0.37~0.38로 물가의 추세와 연동효과를 보임.

- 저장비용 구조는 0.27로 감모에 상관관계를 보이는 것을 확인하였음.
- 기상 변수들은 약한 상관성을 보이는데, 연속적 시계열 형태가 아닌 수확기를 기준으로 문턱 패턴과 같은 형태로 작용하였기 때문으로 판단하였고, 이것은 비선형을 포착하는 트리 기반 모형에서 활용 가능함.

〈그림 3-4〉 변수 상관행렬



자료: 저자 작성.

〈표 3-4〉 종속변수와와의 상관계수

변수	t기	t-1기
q	-0.41	-0.44
import_q	0.56	0.47
cpi	0.37	0.38
cold	0.04	0.04
precip	-0.06	0.07
rot	0.06	0.14
decay	0.27	0.26

자료: 저자 작성.

○ 다중공선성을 파악하기 위해 <표 3-5>에 분산팽창계수(VIF)를 정리함. 실질적인 다중공선성은 없는 것으로 나타나 해당 변수들을 구성해 분석에 활용함.

- 절편 없이 추정된 VIF는 평균이 큰 변수 값이 과대 추정되는 경향이 있으며, 이러한 결과로 상수항을 미포함하는 VIF에서는 반입량과 CPI가 높게 나타났음.
- 그러나 상수항을 포함해 추정한 결과 VIF 값은 모두 5 미만의 값으로 낮게 나타나 해당 변수들 사이에서는 다중공선성 문제 가능성은 낮을 것으로 판단됨.

<표 3-5> 분산팽창계수(VIF)

변수	상수항 미포함	상수항 포함
q	38.08	2.22
import_q	2.83	2.07
cpi	62.15	1.78
cold	4.72	1.01
precip	7.61	1.45
rot	4.79	1.69
decay	2.31	1.56

주: 분석 활용에 맞게 q, import_q, cpi는 t-1기, 그 외는 t기로 활용함.

자료: 저자 작성.

○ 탐색적 데이터 분석(Exploratory Data Analysis) 결과를 종합하면 다음과 같으며, 해당 내용들을 바탕으로 예측 모형을 선택하고 설계하였음.

- 양파 도매가격은 장기 추세 없이 평균을 중심으로 변동하며 연 주기로 계절성이 존재하며, 강한 시간적 지속성과 자기회귀 구조를 보이고 있음.
- 원계열을 차분하지 않고, AR, STL 등 분리와 로그변환 설정이 필요함.
- 외생변수들이 가격과 시차 상관을 보여 외생변수를 결합한 예측 구조를 고려하는 것이 필요함.

- 데이터의 비대칭적 분포, STL 잔차에 존재하는 충격 잡음, 기상 변수의 비선형·문턱 패턴에 따른 비선형 모형 기법 도입이 필요함.
- STL 분해에서 시간에 따라 계절 진폭이 축소되는 시변 계절성이 확인되어 고정 시계열 형태와 다른 유연한 분해인 STL 기반 접근이 필요함.

○ 따라서, 베이스 모형으로 Seasonal Naive를 설정하고, 단일 모형으로 SARIMA와 SARIAMX 그리고 XGBoost를 구성함. 추가로 하이브리드 모형으로 SARIMA-XGBoost, STL-ARIMA-XGBoost를 구축해 총 6개 모형으로 양파 도매가격 예측 모형을 구축하고자 함.

4

양파 가격 예측 모형 구축

- 3장에서 수행한 탐색적 데이터 분석을 통해 Seasonal Naive를 벤치마크 모형으로 두고, 단일 예측 모형으로 시계열 모형인 SARIMA, SARIMAX, 머신러닝 모형인 XGBoost 등 3개의 모형을 선정하였음. 또한, 하이브리드 모형은 SARIMA-XGBoost, STL-ARIMA-XGBoost로 구성하여, 선형과 비선형에 따른 예측력을 살펴보고자 하였음.
- 예측에 이용하는 양파 도매가격은 p_t 로 표기하였고, 로그변환을 통해 예측 모델형구축에 활용하였음. 예측치는 지수함수를 이용해 원래 가격 단위로 환산하였음.
 - 모형 설명에서는 문자 혼동을 막고자 일반적인 종속변수를 나타내는 y 로 표기함.
- 예측 모형에 사용하는 설명변수는 7개이며, 미래 예측 시점에 알 수 있는 정보만 사용하기 위해 대부분 $t-1$ 기 값을 설정하였고, decay 변수만 t 기 값을 사용함.

- q 는 국산 양파 반입량, $import_q$ 는 수입 양파 반입량, cpi 는 전체 소비자물가지수, $cold$ 는 겨울철 냉해 관련 기상, $precip$ 은 봄철 생육기 수분, rot 는 수확기 부패 요인을 의미함. $decay$ 는 저장비용 구조 추세를 단순화하여 미래에도 결정적으로 계산되는 값임.

$$X_t = (q_{t-1}, import_{q_{t-1}}, cpi_{t-1}, cold_{t-1}, precip_{t-1}, rot_{t-1}, decay_t) \quad \text{식(1)}$$

- 머신러닝 모형인 XGBoost는 일반적으로 시간 종속성을 고려하는 시계열 모형이 아니기 때문에 설명변수 외에 자기가격 1, 2, 3, 12개월, 이동평균통계량으로 3, 6, 12개월의 이동평균과 이동표준편차를 생성해 모형에 활용하였음.
- 이동평균 데이터로는 최근 추세를 나타내고, 이동표준편차로는 최근 변동성을 반영하도록 하였음.
- 예측은 $h=1$ 인 1개월 미래를 예측하도록 모형을 구축하였고, 계절주기는 12개월로 설정하였음.

1. 단일 예측 모형

1.1. Seasonal Naive

- Seasonal Naive는 벤치마크로 다른 모형의 성능을 평가하는 기준 역할을 하고, 예측치는 전년 동월 값임. 계절성이 강한 품목은 이러한 규칙으로도 의미 있는 성능을 확인할 수 있음.

$$\hat{y}_t = y_{t-12}, \hat{p}_t = p_{t-12} \quad \text{식(2)}$$

1.2. SARIMA

○ SARIMA는 미래 예측에 자기가격과 과거 예측오차, 그리고 계절적 패턴으로 설명하는 선형 시계열 모형임. 외생변수는 활용하지 않고, 후방이동연산자 B 를 이용하면 식(3)과 같이 표현됨.

- $\phi_p(B)$ 는 비계절 자기회귀 다항식, $\theta_q(B)$ 는 비계절 이동평균 다항식, ϕ_p 와 θ_q 는 12개월 주기의 계절 자기회귀와 이동평균 다항식임. $(1-B)^d$ 는 d 차 차분, $(1-B^s)^D$ 는 계절성을 제거하는 D 차 계절 차분을 의미함.
- 본 연구에서 차수는 $(p, d, q)(P, D, Q, s)$ 로 표기하여 진행함.

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \quad \text{식(3)}$$

1.3. SARIMAX

○ SARIMAX는 SARIMA에 설명변수를 추가한 모형으로 가격을 외생변수로 설명하고 잔여 부분의 시계열을 SARIMA로 통제하는 구조임.

- β' 는 설명변수 벡터의 회귀계수이고, u_t 는 독립변수로 설명하고 남은 잔차로 이를 SARIMA 구조로 따르게 함.

$$y_t = \beta' X_t + u_t \quad \text{식(4)}$$

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D u_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t \quad \text{식(5)}$$

1.4. XGBoost

○ XGBoost(eXtreme Gradient Boosting)은 다수의 결정 트리를 순차적으로 더해 경사 부스팅 기반 비선형 머신러닝 기법임. 선형 모형과 다르게 변수 간 비선형 관계 등을 학습하는 형태임.

- 식(4)와 같은 데이터 벡터를 이용해 식(5)와 같이 K 개 트리 출력을 합산해 예측을 진행함.
- 이때, 학습은 식(6)과 같은 과정을 이용하며, Ω 는 과도한 복잡한 트리를 억제하는 규제항이고 T 는 트리 노드 수, w 는 노드 가중치, γ 와 λ 는 복잡도 패널티 강도를 조절하는 계수임.

$$z_t = (X_t, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-12}, m_t^{(w)}, s_t^{(w)}), \quad w \in 3, 6, 12 \quad \text{식(4)}$$

$$\hat{y}_t = \sum_{k=1}^K f_k(z_t) \quad \text{식(5)}$$

$$L = \sum_t (y_t - \hat{y}_t)^2 + \sum_K \Omega(f_k), \quad \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad \text{식(6)}$$

2. 하이브리드 예측 모형

○ 하이브리드 모형은 선형부의 선형 시계열 모형과 잔차부의 비선형 머신러닝을 결합한 2단계 모형임. 두 모형에서의 시계열은 선형 성분 L_t 와 비선형 성분 N_t 의 합으로 설명하는 식(7)과 같은 Zhang(2003)의 분해구조를 따름.

$$y_t = L_t + N_t \quad \text{식(7)}$$

- 선형 성분은 선형 시계열 모델을 바탕으로 설명될 수 있지만, 그 외 모형이 설명하지 못하는 잔차는 비선형 패턴으로 존재할 수 있어 이 잔차를 비선형 알고리즘인 XGBoost로 추가 학습 및 예측 보완에 활용하고자 함.

2.1. SARIMA-XGBoost

- 앞선 SARIMA 과정에서 선형 예측과 표본 내 잔차를 산출하고, XGBoost에서 이 잔차를 외생변수들과 잔차 과거 시차분포들로 학습하고, 최종 예측을 두 성분의 합으로 추정함.
 - G 는 XGBoost의 회귀함수이며, m 은 잔차 시차를 의미함. 이는 후술한 최적화 과정에서 결정됨.

$$e_t = y_t + \hat{L}_t \quad \text{식(8)}$$

$$\hat{N}_t = G(X_t, e_{t-1}, \dots, e_{t-m}) \quad \text{식(9)}$$

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad \text{식(10)}$$

2.2. STL-ARIMA-XGBoost

- 이 모형은 앞선 SARIMA와 다르게 STL(Seasonal-Trend decomposition using Loess) 분해로 처리하고, 계절적 패턴이 가변적인 것을 허용함.
 - 앞선 EDA에서 STL 분해 결과 계절성이 가변적인 것이 이 모형의 활용 근거임.

- 먼저, 식(7)과 같이 STL 분해를 진행해 계절성분인 S_t 를 분리하고 탈계절 d_t 를 산출함. 이때, T_t 는 추세 성분, R_t 는 잔여 성분임.
- 탈계절 d_t 에 계절성이 없는 ARIMA를 이용해 선형 예측과 잔차를 계산하고, 이 잔차를 XGBoost를 이용해 설명변수와 잔차 시차분포들을 학습하고 잔차를 예측함.
- 이후 최종 예측으로 식(13)처럼 계절 값 $\hat{S}_{m(t)}$ 와 앞서 추정된 탈계열과 잔차를 합해 진행함.

$$y_t = T_t + S_t + R_t, \quad d_t = y_t - S_t \quad \text{식(11)}$$

$$\hat{N}_t = G(X_t, e_{t-1}, \dots, e_{t-m}) \quad \text{식(12)}$$

$$\hat{y}_t = \hat{S}_{m(t)} + \hat{d}_t + \hat{N}_t \quad \text{식(13)}$$

3. 예측 및 평가 설계

3.1. 모형 차수 및 하이퍼파라미터 최적화

○ SARIMA 계열의 차수 $(p, d, q)(P, D, Q, s)$ 는 추세 제거를 위한 비계절 차분 d 는 KPSS, 계절 차분 D 는 OCSB 검정을 통해 결정하였음. 이 두 차분 차수가 결정되면 나머지 차수는 힌드만 칸다카르(Hyndman-Khandakar) 단계적 탐색을 이용해 AICc 값이 최소화되는 조합을 찾도록 함.

○ XGBoost의 예측력을 높이기 위한 하이퍼파라미터 최적화(Hyperparameter Optimization, HPO)는 Optuna 패키지를 이용해 TPE(Tree-structured

Parzen Estimator)를 표본추출기로 활용해 30회 시행으로 최적 조합을 탐색하였음.

- 이때, 각 성능은 시계열 교차검증을 이용해 5-Fold로 진행하고 목적함수는 RMSE를 최소화하도록 하였음.

$$\theta^* = \arg \min_{\theta} \frac{1}{K} \sum_{k=1}^K SE_k(\theta), \quad K = 5 \quad \text{식(14)}$$

- 탐색한 하이퍼파라미터는 트리 개수(n_estimators), 트리 최대 깊이(max depth), 학습률(learning rate), 트리당 사용 표본 비율(subsample), 트리당 사용 변수 비율(colsample bytree), 분기 최소 가중치(min child weight), 규제 강도(reg_alpha, reg_lambda)이고 탐색 범위를 설정하고 그 안에서 탐색하도록 함.

○ 양파 가격 예측 모형 구축을 위해 활용한 변수가 많고, 데이터 길이도 월별이라 짧은 편에 속하므로 과적합 가능성이 존재함. 이를 위해 시계열 교차검증을 진행하였음. 또한 하이퍼파라미터 탐색공간도 과적합을 억제하기 위한 범위로 설정하고 그 안에서 최적 조합을 찾도록 함.

- 트리당 사용 표본 비율(subsample), 트리당 사용 변수 비율(colsample bytree), 분기 최소 가중치(min child weight), 규제 강도(reg_alpha, reg_lambda) 등은 과적합 억제 파라미터이며, 이들의 범위를 강화시켜 과적합을 억제시킴.
- 모델 복잡도 상한도 트리 최대 깊이(max depth)를 2-8로 제한하였고, 학습률(learning rate)도 0.01-0.3 로그스케일로 적용하도록 함.

○ ARIMA도 단순 AIC를 기준으로 탐색하는 것이 아니라 소표본 보정을 설정한

AICc로 진행해 과대 차수 선택을 억제하고, 차수 상한도 비계절 차수는 6, 계절 차수는 2로 제한함.

3.2. 평가지표 및 방식

○ 본 연구는 예측력을 평가하기 위해 단일 분할 방식이 아닌 확장 윈도우 기반 롤링 오리진(rolling-origin) 방식으로 진행해 강건한 예측력을 확보하고자 하였음.

- 롤링 오리진 검증(Rolling Origin Validation) 과정에서 모형 간 예측 오차의 강한 선형종속성(Linear dependence)이나 표본 부족으로 인해 공분산 행렬의 특잇값 분해(SVD)가 수렴하지 않는 불안정성(Numerical Instability)이 감지될 경우, 과적합을 방지하기 위해 단순 평균 앙상블(Simple Average Ensemble)로 자동 전환되도록 강건한 제어 알고리즘을 구현함.

○ 기준점을 t+1씩 늘려 모형을 재추정 및 재학습하여 다음 월의 가격을 예측하도록 하여 우연에 의한 결과가 아닌 전반적인 평가 성능을 측정하여 실제 운영 절차와 동일한 실질적인 평가를 하고자 하였음.

○ 평가지표는 평균제곱근오차(RMSE), 평균제곱근비율오차(RMSPE), 평균절대오차(MAE), 평균절대백분율오차(MAPE), 평균 방향정확도(MDA) 등을 이용하여 종합적으로 예측력을 평가 비교하고자 함.

$$- RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

$$\begin{aligned}
- \text{ } RMSPE &= \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2} \times 100 \\
- \text{ } MAE &= \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \\
- \text{ } MAPE &= \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \\
- \text{ } MDA &= \frac{1}{n} \sum_{t=1}^n I[(y_t - y_{t-1})(\hat{y}_t - \hat{y}_{t-1}) > 0] \times 100
\end{aligned}$$

〈표 4-1〉 평가지표 비교

지표	측정대상	단위	특징	활용
RMSE	절대 오차	가격 단위	큰 오차 강조	가격 정확도 평가
RMSPE	상대 오차	%	큰 오차 강조+비율화	품목 간 비교
MAE	절대 오차	가격 단위	직관적 평균 오차	평균 예측 오차 평가
MAPE	상대 오차	%	해석 쉬움	예측 정확도 제시
MDA	방향성	%	등락 방향 평가	가격 추세 판단

자료: 저자 작성.

5

양파 가격 예측 모형 평가

1. 모형 적합 진단

1.1. ARIMA 차수 선정

○ ARIMA 모형의 최적 차수는 결과는 <표 5-1>에 제시하였음.

- 계절차분 차수는 OCSB 계절 단위근 검정, 비계절 차분 차수 d 는 KPSS, 자기회귀와 이동평균차수 (p , q , P , Q)는 힌드만-칸다카르(Hyndman-Khandakar) 알고리즘을 이용해 AICc가 최소화하는 조합을 탐색하였음.
- 탐색 범위는 p 와 q 는 6 이하, P 와 Q 는 2 이하, 계절 주기는 12로 설정하였고, 절편 포함 여부는 자동으로 선택하도록 하였음.
- SARIMA는 $(3, 1, 0)(1, 0, 2, 12)$ 가 최적 조합으로 나타났으며, 이 조합은 SARIMA-XGBoost에도 활용하였음.
- SARIMAX는 $(2, 0, 0)(1, 0, 1, 12)$ 로 선정됐고, STL-ARIMA는 계절성을 분리했으므로 계절항 없이 $(2, 1, 1)(0, 0, 0, 0)$ 으로 추정되었음.

〈표 5-1〉 ARIMA 차수 결과

모형	order	seasonal order	AIC	AICc	BIC
SARIMA	(3, 1, 0)	(1, 0, 2, 12)	-142.40	-141.72	-115.14
SARIMAX	(2, 0, 0)	(1, 0, 1, 12)	-65.16	-63.67	-24.27
SARIMA-XGBoost	(3, 1, 0)	(1, 0, 2, 12)	-142.40	-141.72	-115.14
STL-ARIMA-XGBoost	(2, 1, 1)	(0, 0, 0, 0)	-151.14	-150.86	-134.10

자료: 저자 작성.

1.2. 계수 추정 결과

○ SARIMA와 STL-ARIMA 추정 결과는 〈표 5-2〉와 〈표 5-3〉에 정리하였음.

- SARIMA 추정 결과를 보면 비계절 자기회귀항(ar.L1)과 계절항(ar.S.L12)이 1% 유의수준에서 유의성을 확보한 것으로 나타났음. 이는 강한 계절의 존성이 있는 것을 확인할 수 있음.

〈표 5-2〉 SARIMA 추정 결과

파라미터	계수	표준오차	p-value	CI lower	CI upper
intercept	0.0000	0.0005	0.9524	-0.0010	0.0011
ar.L1	0.2441***	0.0532	0.0000	0.1398	0.3484
ar.L2	-0.1456**	0.0641	0.0231	-0.2711	-0.0200
ar.L3	-0.1393	0.0712	0.0505	-0.2788	0.0003
ar.S.L12	0.9852***	0.0112	0.0000	0.9633	1.0070
ma.S.L12	-1.1404***	0.0659	0.0000	-1.2696	-1.0111
ma.S.L24	0.2481***	0.0667	0.0002	0.1173	0.3789
sigma2	0.0270	0.0021	0.0000	0.0228	0.0312

주: *** p<0.01, ** p<0.01, * p<0.01.

자료: 저자 작성.

- STL-ARIMA 결과에서는 자기회귀항들(ar.L1, ar.L2)과 이동평균항(ma.L1)이 1% 유의수준에서 유의한 것으로 나타났는데, 계절분해 이후에도 시간적 지속성과 이동평균 보정이 반영되고 있음을 알 수 있었음.

〈표 5-3〉 STL-ARIMA 추정 결과

파라미터	계수	표준오차	p-value	CI lower	CI upper
intercept	0.0001	0.0003	0.6729	-0.0004	0.0007
ar.L1	1.0905***	0.0462	0.0000	0.9999	1.1811
ar.L2	-0.2174***	0.0441	0.0000	-0.3038	-0.1311
ma.L1	-0.9813***	0.0561	0.0000	-1.0912	-0.8714
sigma2	0.0271	0.0017	0.0000	0.0239	0.0304

주: *** p<0.01, ** p<0.01, * p<0.01.

자료: 저자 작성.

1.3. 잔차 분석

○ 잔차 분석은 〈표 5-4〉와 같으며, 대체적으로 자기상관은 잘 제거되었으나 분포가 정규분포가 아닌 것으로 나타나 XGBoost와 같은 비선형 모형이 필요함을 확인함.

- Ljung-Box 결과 SARIMAX를 제외하고 모두 5% 유의수준에서 귀무가설을 기각할 수 없어 잔차가 백색잡음일 가능성이 높은 것을 확인함.
- 정규성과 이분산성 검정 결과를 보면 모든 모형에서 5% 유의수준에서 귀무가설을 기각하는 것으로 나타나 비선형성과 극단적 성분은 여전히 존재하는 것으로 나타나 비선형 모형으로 보완할 필요가 있는 것으로 나타났음.

〈표 5-4〉 잔차 검정

모형	Ljung-Box p		Jarque-Bera p		ARCH-LM p	
	statistic	p-value	statistic	p-value	statistic	p-value
SARIMA	2.734	0.997	269,304.664	0.000	20.534	0.025
SARIMAX	23.607	0.023	5,125.074	0.000	23.655	0.009
SARIMA-XGB	2.734	0.997	269,304.664	0.000	20.534	0.025
STL-ARIMA-XGB	0.209	1.000	272,348.951	0.000	31.118	0.001

자료: 저자 작성.

1.4. 근 안정성

○ 추정된 ARIMA가 통계적으로 의미가 있으려면 정상성(stationarity)과 가역성(invertibility)을 만족해야 함.

- 정상성은 자기회귀의 충격 효과가 시간이 지남에 따라 발산하지 않고 소멸하는 것을 의미함.
- 가역성은 이동평균이 모형에서 과거 관측치의 수렴하는 형태로 표현될 수 있어야 하는 것임.

○ <표 5-5>는 각 모형의 근 안정성(Root Stability) 결과로 모두 근 절댓값이 1을 초과해 정상성과 가역성을 모두 충족하는 것으로 나타남. 이는 각 ARIMA가 통계적으로 타당하며, 예측은 신뢰할 수 있는 모형을 바탕으로 산출됨을 의미함.

<표 5-5> 근 안정성 결과

모형	최소 AR 근 절댓값	최소 MA 근 절댓값
SARIMA	1.001	1.014
SARIMAX	1.131	1.188
SARIMA-XGB	1.001	1.014
STL-ARIMA-XGB	1.208	1.019

자료: 저자 작성.

2. 예측력 평가

2.1. 모형별 예측력 비교

- 모형별 예측력을 평가하기 위해 훈련 데이터와 테스트데이터를 8:2 비율로 설정하였고, 단일 분할 평가 방식이 아닌 확장 윈도우 기반 롤링-오리진(rolling-origin) 평가 방법을 이용하였음.
 - 단일 분할 평가 방식은 모형을 구축 후 데이터 입력만을 통해서 예측치를 산출하고 실제치와 비교하는 형태로 쉽고 빠르지만, 평가 시점에 따라 예측력이 크게 달라질 수 있고, 시간적 의존성을 반영하지 못할 수 있음.
 - 실제 가격 예측 모형을 활용하기 위해 매 시점의 데이터들로 학습하고, 그것을 바탕으로 $h=1$ 을 예측하는 형태로 평가를 진행해 평가의 안정성을 높이고자 하였음.
 - 이를 통해 실제 운영 환경에서 예측방식과 동일하게 진행해 실무 적용 성능을 현실적으로 평가하고자 하였음.
 - ARIMA 모형들은 매번 확장되는 데이터를 바탕으로 최적 시차를 선정하고, 그 바탕으로 예측치를 산출하고, XGBoost의 경우 학습 데이터로 튜닝된 하이퍼파라미터 값을 고정하고 매번 데이터를 이용해 최종 학습을 진행하고 예측치를 추정하는 형태로 진행함.
- 예측력은 RMSE, RMSPE, MAE, MAPE, MDA를 종합적으로 평가하고자 하였으며 <표 5-6>, <그림 5-1>과 같음.
- 예측 결과를 보면 단일모형인 XGBoost가 모든 지표에서 예측력 성능이 가장 우수한 것으로 나타났으며, 벤치마크 모형인 Seasonal Naive 보다 RMSE가 66.4% 개선된 것으로 확인됨.

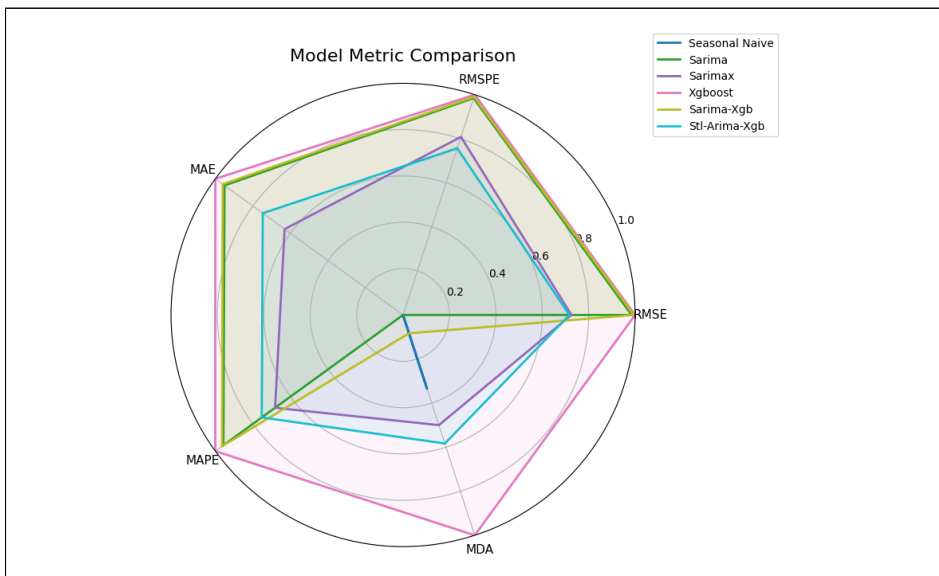
- XGBoost 뒤로 SARIMA-XGB와 SARIMA가 예측력 성능이 우수한 것으로 나타났다.
- 다만, 예측 방향 정확도를 나타내는 MDA는 XGBoost만 60%대를 상회한 64.3%를 보여 우수함을 보였지만, SARIMA-XGB와 SARIMA는 벤치마크 모형보다 낮게 나타나 방향 예측에서는 성능이 낮게 나타났다.

〈표 5-6〉 예측력 비교

모형	RMSE	RMSPE	MAE	MAPE	MDA
Seasonal Naive	534.6	78.3	402.3	44.7	50.0
SARIMA	184.9	16.6	140.2	12.9	42.9
SARIMAX	277.0	27.6	228.5	22.1	53.6
XGBoost	179.6	15.6	126.7	11.5	64.3
SARIMA-XGB	181.7	16.2	137.7	12.7	44.6
STL-ARIMA-XGB	280.1	30.9	196.4	19.7	55.4

자료: 저자 작성.

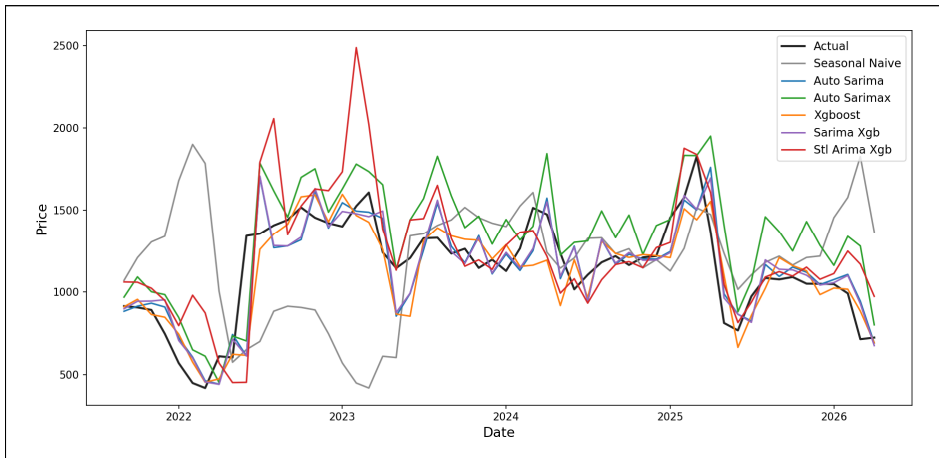
〈그림 5-1〉 모형별 예측력 지표 시각화



자료: 저자 작성.

○ 롤링 오리진으로 평가한 예측치를 시각화한 <그림 5-2>의 결과를 보면 벤치마크 모형은 실제치와 매우 다른 예측을 보이고 있으며, XGBoost, SARIMA, SARIMA-XGB가 실제치를 잘 따라가는 것을 확인함.

<그림 5-2> 실제치와 예측치 시각화



자료: 저자 작성.

○ 기존 모형과 예측력 비교는 <표 5-7>에 정리하였음. 예측이 중요한 3, 5, 6월 기준으로 비교한 결과 전체적인 예측력은 크게 좋아지지 않았지만, 세부적으로 기존 예측력이 떨어진 시점에서는 예측력이 더 개선된 것으로 확인됨.

- 기존 모형의 예측력인 <표 2-3>을 보면 저장 출하기에서는 비교적 안정적인 예측력을 보임.
- 3, 5, 6월은 각각 재고, 기상 충격 및 작황, 저장 출하의향 등의 영향으로 가격 구조가 바뀌는 시점으로 해당 시점의 예측력을 기존과 비교하였고, 그 기간은 예측력 산출이 가능한 2022~2025년으로 한정하였음.
- 구축한 5개 모형 예측력은 기존 예측력을 모두 개선한 경우는 없으나 모형 별·시점별로 보면 구축된 모형의 예측력이 대체로 우수하게 나타남.

- 모형에 따라 예측력이 우수하게 나타난다는 것은 각 모형별로 양과 시계열 패턴을 포착하는 방식이 다르기 때문인데, 향후 실무적으로는 이를 예측결합 형태로 예측치를 산출해야 할 방향성도 시사함.
- 다만, 기존 예측치 산출에 이용한 모형과 변수는 본 연구에서 활용한 데이터나 모형과 다르기 때문에 직접적인 수치 비교는 제한적으로 봐야 함.

〈표 5-7〉 월별 예측력 비교

시점	기존	SN	sarima	sarimax	xgb	sarima-xgb	stl-arima-xgb
'22년 3월	0.20	3.27	0.10	0.44	0.04	0.10	1.18
'22년 5월	0.16	0.05	0.23	0.16	0.11	0.23	0.26
'22년 6월	0.37	0.52	0.54	0.50	0.54	0.54	0.67
'23년 3월	0.13	0.74	0.10	0.07	0.08	0.11	0.24
'23년 5월	0.35	0.47	0.25	0.11	0.22	0.25	0.07
'23년 6월	0.05	0.11	0.24	0.17	0.31	0.24	0.21
'24년 3월	0.21	0.06	0.17	0.02	0.25	0.17	0.10
'24년 5월	0.05	0.07	0.13	0.10	0.07	0.12	0.19
'24년 6월	0.03	0.19	0.25	0.79	0.05	0.26	0.02
'25년 3월	0.12	0.17	0.17	0.01	0.23	0.17	0.01
'25년 5월	0.23	0.52	0.22	0.54	0.45	0.22	0.33
'25년 6월	0.24	0.32	0.13	0.23	0.13	0.13	0.03

주: 예측력은 MAPE 기준 값임.

자료: 저자 작성.

2.2. 예측력 검정

○ 모형 간 예측력 차이를 평가하기 위해 Diebold-Mariano 검정을 진행하였고, 소표본 보정을 위해 Harvey-Leybourne-Newbold 수정을 적용하였음. 이에 대한 결과는 〈표 5-8〉과 같음.

○ 벤치마크 모형 대비 1% 유의수준에서 유의하게 모든 모형이 우수한 것으로 나타났으며, XGBoost, SARIMA, SARIMA-XGB 간 예측력 차이는 유의하지 않아 통계적으로 예측력 차이가 없는 것으로 나타났음.

- STL-ARIMA-XGB는 XGBoost, SARIMA, SARIMA-XGB 대비 우수한 것으로 나타났음.

○ 우수한 세 모형의 예측력은 통계적으로 차이가 없지만, XGBoost가 점 추정 및 예측 방향에서 가장 우수함을 확인함.

〈표 5-8〉 Diebold-Mariano 검정 결과

모형	Seasonal Naive	SARIMA	SARIMAX	XGBoost	SARIMA-XGB
SARIMA	-3.99*** (0.0002)				
SARIMAX	-3.31** (0.0017)	3.96*** (0.0002)			
XGBoost	-4.00*** (0.0002)	-0.43 (0.6661)	-3.61*** (0.0007)		
SARIMA-XGB	-4.01*** (0.0002)	-1.08 (0.2831)	-4.01*** (0.0002)	0.17 (0.8647)	
STL-ARIMA-XGB	-3.75*** (0.0004)	2.23 (0.0296)	0.08 (0.9378)	2.29* (0.0262)	2.30* (0.0250)

주: *** p<0.01, ** p<0.01, * p<0.01.

자료: 저자 작성.

3. 가격 예측 기여 효과

○ 가격 예측 기여 효과를 보기 위해 선형 효과는 SARIMAX의 추정 계수로 확인하고, XGBoost는 SHAP의 피처중요도로 파악해 구조를 규명하고자 함.

3.1. 선형 모형 변수 효과

○ SARIMAX 추정 결과는 <표 5-9>와 같이 정리하였음.

- 가격을 유의하게 설명하는 변수는 반입량, 소비자물가지수, 냉해 관련 기상 요인 3가지로 나타났음.
- 먼저 국산 양파 반입량(q)은 1% 유의수준에서 통계적으로 유의하게 나타났으며, 경제이론에 부합하게 계수 부호가 (-)로 추정되었음. 그러나 음의 계수 크기는 매우 작게 나타남.
- 수입 양파 반입량(import_q) 또한 계수 부호가 (-)로 나타났지만, 통계적으로 유의성을 확보하지 못함.
- 소비자물가지수(cpi)의 계수는 0.078로 1% 유의수준에서 유의함을 보였음. 소비자물가지수가 한 단위(1포인트) 상승하면 양파 가격이 약 7.8% 상승하는 것으로 나타남.
- 냉해 관련 기상(cold)은 5% 유의수준에서 유의한 -0.0519로 추정되었고, 냉해가 심해지면 가격이 상승하는 영향을 주는 것으로 나타남.
- 그 외 다른 변수들은 선형 구조에서는 통계적으로 유의성을 보이지 않았음.

○ cpi를 제외하고 계수 절댓값이 매우 작게 나타났는데, 이는 log-linear 형태로 효과가 적은 것이라기보다는 변수의 단위 차이 때문임.

- log-linear로 진행한 이유는 설명변수에 0 값이나 음수가 존재하고, 종속 변수만을 로그변환하여 예측하고자 하는 가격 급등의 분산을 안정화하여 예측하고자 하기 때문임.

〈표 5-9〉 SARIMAX 추정 결과

파라미터	계수	표준오차	p-value	CI lower	CI upper
q_L1	-0.0000***	0.0000	0.0002	0.0000	0.0000
import_q_L1	-0.0000	0.0000	0.7935	0.0000	0.0000
cpi_L1	0.0782***	0.0028	0.0000	0.0727	0.0836
cold	-0.0519**	0.0260	0.0460	-0.1029	-0.0009
precip	0.0003	0.0002	0.2199	-0.0002	0.0007
rot	0.0002	0.0002	0.2959	-0.0002	0.0007
decay	0.0007	0.0006	0.2918	-0.0006	0.0019
ar.L1	0.9368***	0.0577	0.0000	0.8237	1.0500
ar.L2	-0.0463	0.0614	0.4505	-0.1666	0.0740
ar.S.L12	0.2059	0.2224	0.3545	-0.2300	0.6417
ma.S.L12	0.1261	0.1836	0.4920	-0.2336	0.4859
sigma2	0.0371	0.0036	0.0000	0.0300	0.0441

주: *** p<0.01, ** p<0.01, * p<0.01.

자료: 저자 작성.

3.2. 비선형 모형 변수 기여

○ 비선형 모형인 XGBoost, 잔차 예측에 이용한 SARIMA-XGB, STL-ARIMA-XGB별로 SHAP 변수 중요도를 추정하여 〈그림 5-3〉에 제시함.

- 변수 위치가 높을수록 가격 예측에 크게 기여하는 변수이며, 각 변수의 값이 붉은 색이면 그 값이 높은 수준, 파란색이면 그 값이 낮은 수준을 의미함. 또한, SHAP value 0 값 기준으로 오른쪽에 있으면 가격 상승쪽, 좌측에 있으면 가격 하락을 예측하는 형태임.

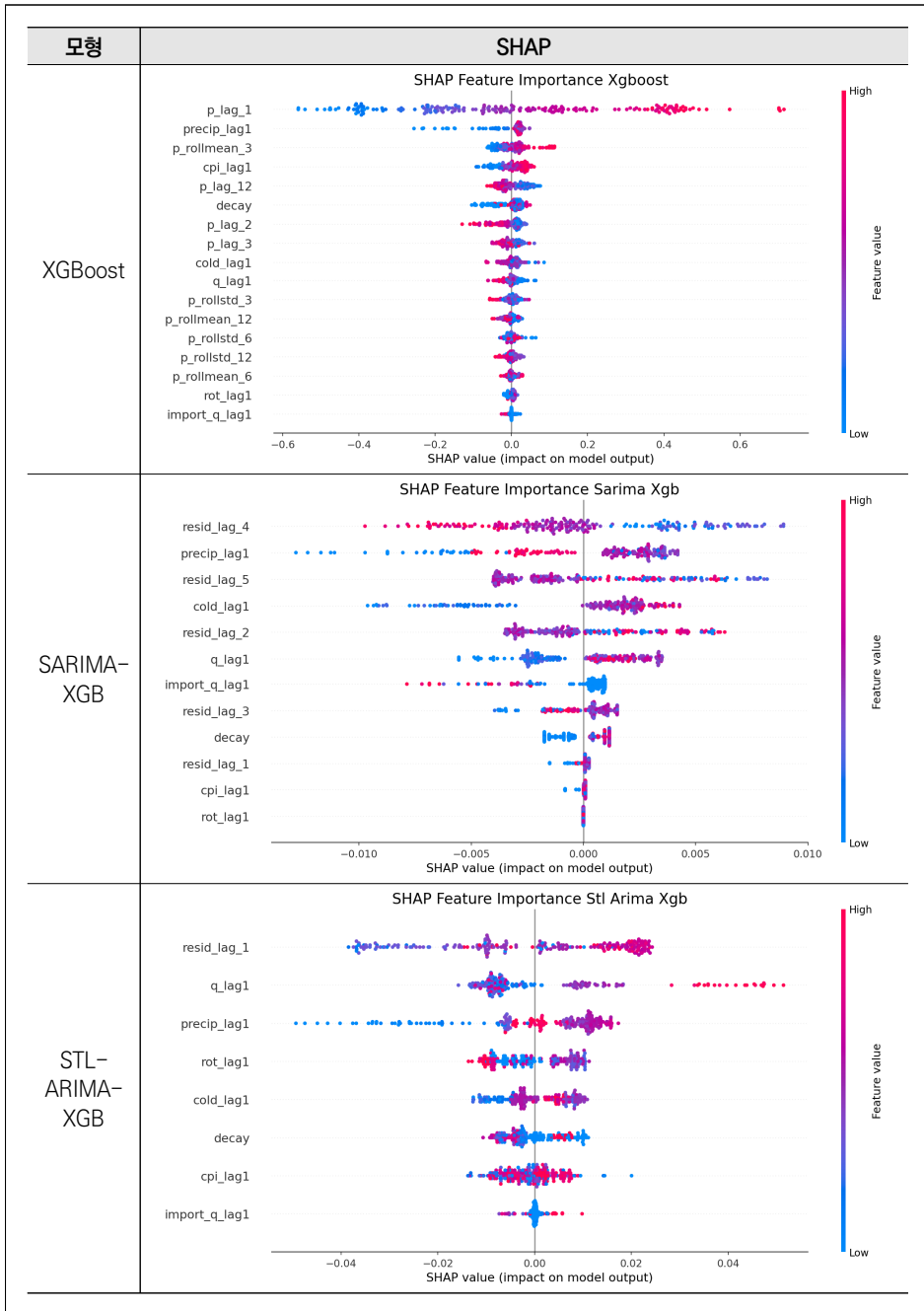
○ 먼저, XGBoost의 피쳐 중요도를 보면 t-1기 가격이 가격 예측에 가장 큰 기여를 한 것으로 나타났으며, 가격이 높을수록 예측 가격을 상승시키는 것으로 나타났음.

- 두 번째로는 생육 수분 관련 기상이 높을수록 가격 예측치를 상승시키는 주요 요인으로 나타났는데, 생육기에 강수량이 낮으면 예측 가격을 낮추는 방향으로 작용하는 비선형 패턴을 보임.
- 그 뒤로 3개월 이동평균, 소비자물가지수, 12개월 전 가격, 재고비용 구조 등으로 나타났음. 재고비용 구조 변수의 경우 값이 높거나 낮을 때 상관 없이 가격 예측을 상승시키거나 낮춰 비선형 영향을 주는 형태로 나타남.

○ SARIMA-XGB와 STL-ARIMA-XGB를 보면 선형 모형이 설명하고 남은 잔차에 기상과 공급, 재고비용 등 비선형 효과가 존재하는 것을 확인함. 또한 XGBoost가 이를 포착해 예측을 보완하고 있음을 시사함.

- 선형 모형에서 자기회귀 성분을 제외한 잔차에 잔차 시차와 함께 외생 변수들이 비선형적으로 가격 예측에 영향을 주는 것으로 나타남.

〈그림 5-3〉 모형별 SHAP 중요도



자료: 저자 작성.

3.3. 선형과 비선형 효과 비교

- 선형과 비선형에서 독립변수들이 다른 형태로 가격 예측에 활용되고 있는 것으로 나타났음.
- 선형 모형에서는 기상 요인들의 영향이 작게 나타났으나, XGBoost의 피쳐 중요도에서는 그 변수들이 가격 예측에 매우 중요한 기여를 하고 있는 것으로 나타났음.
- 이러한 두 접근에서의 차이는 독립변수들의 가격 예측 효과가 단조롭게 나타나는 것이 아니라 특정 국면이나 작기에 비선형적 또는 집중적으로 작동하고 있음을 시사함.
 - 이는 선형 모형이 포착하지 못하는 부분을 트리 기반 비선형 모형이 포착하였다는 증거이며 비선형 모형의 단독 활용 또는 하이브리드 모형 설계의 근거가 됨을 의미함.
- 양파 가격 예측에 있어서 전체 소비자물가지수는 두 접근에서 모두 중요한 변수로 나타났고, 거시 물가의 영향은 선형적 성격이 강한 것을 확인할 수 있었음.

6

인공지능 기반 관측 고도화 방안

1. 가격 예측 모형 분석 요약

○ 본 연구는 양파 가격 전망의 정확도 제고와 농업관측체계의 고도화를 목적으로 기존 양파 가격 전망 방식의 특징과 한계를 분석하고, 인공지능(AI)을 활용한 가격 예측 모형 구축 및 활용 가능성을 검토하였음.

- 최근 농산물 시장은 기상 변화, 생산 여건 변화, 수입 및 유통 구조 변화 등 다양한 요인의 영향으로 가격 변동성이 확대되고 있으며, 특히 양파는 생산량뿐만 아니라 저장 및 출하 조절, 기상 충격, 시장 참여자의 기대 변화 등이 가격 형성에 복합적으로 작용하는 품목임.
- 이에 따라 기존 수급 중심의 전망 방식과 함께 다양한 외부 요인을 반영할 수 있는 데이터 기반 예측 체계 구축의 필요성이 증가하고 있음.

○ 기존 양파 가격 전망 체계와 전망 오차를 분석한 결과, 농업관측센터의 양파 가격 전망은 최근 5개년(2021~2025년) 기준 실제 가격과 비교하여 평균 약

9% 수준의 오차율을 보여 전반적으로 안정적인 예측력을 확보하고 있는 것으로 나타났다.

- 그러나 월별로는 저장양파에서 햇양파로의 전환기, 조생종 출하시기 등 특정 시점에서 오차가 확대되는 현상이 확인되었음. 이는 생산량 변화뿐만 아니라 수확기 기상 조건, 저장업체의 입고 및 출하 전략, 시장 기대 변화 등 기존 수급자료만으로 설명하기 어려운 요인이 가격 변동에 영향을 미치기 때문으로 분석되었음.

○ 다음으로 인공지능 기반 양파 가격 예측모형 구축을 위해 2003년 1월부터 2026년 4월까지의 월별 양파 도매가격 자료를 기반으로 분석 자료를 구축하였음.

- 설명변수로는 국산 양파 반입량, 수입 양파 반입량, 소비자물가지수, 겨울철 냉해 관련 기상 변수, 봄철 생육기 수분 조건, 수확기 부패 관련 변수, 저장비용 구조 등을 활용하여 양파 가격에 영향을 미치는 수급·경제·기상 요인을 종합적으로 반영하였음.

○ 탐색적 데이터 분석 결과, 양파 도매가격은 장기적인 상승 또는 하락 추세보다는 일정한 계절성과 자기회귀 구조를 보이는 것으로 나타났다.

- 또한 연중 반복되는 계절성이 존재하는 동시에 특정 시기에 가격 급등락을 유발하는 충격 요인이 확인되어, 선형적인 시계열 모형만으로는 가격 변동 구조를 충분히 설명하기 어려운 것으로 분석되었음.
- 특히 기상 변수는 단순한 연속형 변수보다는 생육 단계별 임계 수준이나 특정 시기의 충격 요인으로 작용하는 특성이 있어 비선형 예측모형 활용의 필요성이 확인되었음.

○ 이러한 분석 결과를 바탕으로 본 연구에서는 기준모형을 계절성을 반영하는 Seasonal Naive 모형으로 설정하고, 전통적인 시계열 모형인 SARIMA와 외생변수를 포함한 SARIMAX, 머신러닝 기반의 XGBoost, 그리고 선형 시계열 모형과 머신러닝 모형을 결합한 하이브리드 모형인 SARIMA-XGBoost 및 STL-ARIMA-XGBoost 등 총 6개의 예측모형을 구축하였음.

- 이를 통해 기존 시계열 분석 방식과 인공지능 기반 비선형 예측 방식의 장점을 결합할 수 있는 하이브리드 예측 접근 방법을 검토하였음.

○ 연구 결과, 양파 가격 예측은 단일 시계열 모형보다는 머신러닝 기법의 XGBoost나 가격의 계절성과 자기상관 구조를 설명하는 시계열 모형과 다양한 외부 요인의 비선형 관계를 학습하는 머신러닝 모형을 결합하는 방식이 적합한 것으로 판단되었음.

- 점추정에서는 XGBoost의 예측력이 가장 우수했으나 하이브리드 모형과의 예측력에 통계적으로 차이가 없는 것으로 나타났음.
- 특히, <표 5-7>의 3월과 5월의 예측력을 보면 XGBoost보다는 하이브리드 모형의 예측력이 더 우수하게 나타나 단순히 XGBoost만을 이용하기 보다는 하이브리드 모형까지 고려해 모형의 예측치를 다양하게 검토할 필요가 있음.
- 이는 농산물 가격이 단순히 과거 가격 흐름만으로 결정되는 것이 아니라 생산, 유통, 기상, 저장 등 다양한 요인의 상호작용으로 결정된다는 특성을 반영한 결과임.
- 또한, 특정 시점에 따라 특정 모형의 예측력이 우수하게 나타나는데 이러한 점을 반영해 예측결합 또는 앙상블 기법을 도입해 실무적으로 예측력을 개선하는 방법을 시도해 볼 필요가 있음.

○ 또한 본 연구는 인공지능 기술이 기존 농업관측 체계를 대체하는 수단이 아니라, 전문가 중심 관측 방식을 보완하는 분석 지원 도구로 활용될 수 있음을 제시하였음.

- 인공지능 모형은 대규모 자료에서 관측업무 담당자가 발견하기 어려운 가격 변동 패턴을 탐색하고 위험 신호를 제공할 수 있으며, 전문가 판단과 결합할 경우 보다 신속하고 정밀한 가격 전망 체계를 구축하는 데 기여할 수 있음.

○ 본 연구는 양파 가격 전망 분야에서 기존 수급 중심 분석의 한계를 보완하고, 시계열 분석과 인공지능 기법을 결합한 데이터 기반 예측 체계의 방향을 제시하였다는 점에서 의의가 있음.

- 향후 지속적인 데이터 축적과 모형 고도화를 통해 인공지능 기반 가격 전망 시스템을 구축한다면, 농업관측의 정확성과 적시성을 높이고 생산자와 유통 주체의 합리적인 의사결정을 지원하는 기반으로 활용될 수 있을 것으로 기대됨.

2. 인공지능 기반 관측체계 구축 방향

○ 본 연구는 양파 가격 예측 분야에서 인공지능(AI) 활용 가능성을 검토하고, 기존 수급 중심 전망 체계를 보완할 수 있는 데이터 기반 예측모형의 방향을 제시하였음. 그러나 인공지능 기반 예측모형이 실제 농업관측 업무에 안정적으로 활용되기 위해서는 데이터 구축, 모형 고도화, 업무 적용 체계 마련 등 단계적인 접근이 필요함. 이에 향후 연구 및 시스템 구축 방향을 단기, 중기, 장기 단계로 구분하여 제시하면 다음과 같음.

2.1. 인공지능 기반 가격 예측모형 검증 및 데이터 기반 구축(단기)

- 단기적으로는 본 연구에서 구축한 양파 가격 예측모형의 지속적인 검증과 개선이 필요함.
 - 현재 모형은 양파 가격, 반입량, 수입량, 소비자물가지수, 기상 변수, 저장 관련 변수 등을 활용하여 설계되었으나, 실제 관측 업무에 적용하기 위해서는 매월 전망 시점에서 이용 가능한 자료만을 활용한 반복 검증 과정이 필요함.
- 첫째, 기존 농업관측 자료와 인공지능 학습 데이터를 체계적으로 관리할 수 있는 데이터 관리 체계를 구축해야 함.
 - 표본농가 조사, 산지 조사, 저장업체 조사 등 기존 조사자료뿐만 아니라 도매시장 거래자료, 기상자료, 수입 통계, 소비 관련 자료 등을 통합 관리하여 AI 학습에 적합한 데이터베이스를 구축할 필요가 있음.
- 둘째, 다양한 예측모형에 대한 지속적인 성능 평가가 필요함.
 - 본 연구에서는 SARIMA, SARIMAX, XGBoost 및 하이브리드 모형 등 다양한 접근 방법을 검토하였으나, 실제 운영 환경에서는 매월 신규 자료가 추가되는 상황을 고려한 재학습 및 모형 성능 비교 체계가 필요하며, 이를 통해 품목별·시기별로 가장 적합한 예측 모형을 선정하는 체계를 마련해야 함.
- 셋째, 인공지능 예측 결과를 기존 전문가 전망 과정에서 활용하기 위한 시범 운영이 필요함.
 - 인공지능 모형이 제시하는 전망값, 가격 변동 가능성, 주요 영향 변수 등을 전문가 검토 자료로 제공하여 기존 관측 프로세스와의 연계 가능성을 검증해야 함.

2.2. 인공지능 기반 농산물 가격 전망 지원 시스템 구축(중기)

- 중기적으로는 양파 품목을 중심으로 구축한 인공지능 예측 모형을 농업관측 업무 지원 시스템으로 발전시키는 단계가 필요함. 단순히 예측값을 산출하는 수준을 넘어, 데이터 수집부터 분석, 결과 제공까지 자동화된 플랫폼 구축이 목표가 되어야 함.
- 첫째, 실시간 또는 준실시간 데이터 연계 체계를 강화해야 함.
 - 현재 농산물 가격은 생산량뿐만 아니라 기상변화, 저장업체 출하 전략, 수입 동향, 소비 변화 등에 의해 영향을 받고 있으므로, 기존 조사자료와 함께 시장 상황을 빠르게 반영할 수 있는 데이터 확보가 중요함.
 - 특히 저장재고량, 도매시장 반입량, 산지 출하 동향 등은 가격 변동성을 설명하는 핵심 정보로 활용될 수 있음.
- 둘째, 설명 가능한 인공지능 기반 관측 지원 기능을 개발해야 함.
 - 농업관측 분야에서는 예측 정확도뿐만 아니라 “왜 이러한 가격 전망이 산출되었는가”에 대한 설명이 중요함.
 - 따라서 변수 중요도 분석, 가격 상승·하락 요인 분석, 과거 유사 사례 비교 기능 등을 포함하여 전문가와 정책 담당자가 쉽게 활용할 수 있는 형태로 발전시킬 필요가 있음.
- 셋째, 품목별 특성을 반영한 인공지능 모형 개발이 필요함.
 - 양파는 저장성이 높고 출하시기 조절이 가능한 품목이라는 특징이 있으므로, 배추·무·마늘 등 다른 품목에는 품목별 가격 형성 구조를 반영한 별도의 모형 설계가 필요함.
 - 이를 통해 농업관측 전 품목으로 인공지능 활용 범위를 확대할 수 있을 것임.

2.3. 인공지능 기반 농업관측 지능화 체계 구축(장기)

- 장기적으로는 인공지능을 활용한 농업관측 지능화 체계를 구축하여 기존 전문가 중심 관측 방식과 데이터 기반 분석 방식을 결합하는 방향으로 발전해야 함.
- 첫째, 농업관측 전 품목에 대한 통합 인공지능 예측 플랫폼 구축이 필요함.
 - 품목별 가격, 생산, 기상, 유통 데이터를 통합하고, 품목 특성에 맞는 다양한 인공지능 모형을 적용함으로써 농산물 수급 및 가격 전망의 정밀도를 높일 수 있음.
- 둘째, 예측 범위를 확대할 필요가 있음.
 - 현재 월별 가격 전망 중심에서 벗어나 단기 가격 변동 예측, 출하시기별 가격 전망, 위험 상황 조기 탐지 등 다양한 분석 기능으로 확대해야 함.
 - 특히 이상기상, 생산량 급감, 수입 증가 등 시장 충격 요인을 조기에 탐지할 수 있는 시스템 구축이 요구됨.
- 셋째, 인공지능과 전문가 판단이 결합된 협업형 관측 체계를 정착시켜야 함.
 - 농산물 시장은 과거 데이터만으로 설명하기 어려운 구조적 변화가 존재하므로, 인공지능 분석 결과와 현장 전문가의 경험을 결합하는 방식이 가장 효과적임.
 - 인공지능은 대규모 데이터를 기반으로 숨겨진 패턴과 위험 신호를 탐색하고, 전문가는 시장 상황과 정책 환경을 종합적으로 판단하는 역할 분담 체계를 구축할 필요가 있음.
- 인공지능 기반 농업관측 체계는 단순한 가격 예측 기술 도입을 넘어 데이터

수집, 분석, 전문가 판단, 정책 활용이 연계되는 종합적인 의사결정 지원 시스템으로 발전해야 함.

- 이를 통해 농업관측센터는 변화하는 농산물 시장 환경에 보다 능동적으로 대응하고, 생산자와 유통 주체에게 신뢰성 높은 정보를 제공하는 미래형 관측 체계를 구축할 수 있을 것으로 기대됨.

3. 정책적 시사점 및 향후 과제

3.1. 연구의 의의

○ 본 연구는 양파 가격 예측의 정확도를 향상시키기 위한 인공지능 기반 예측모형을 개발하고, 실제 농업관측 업무에 적용 가능한 활용 방안을 제시하였다는 점에서 학술적·정책적 의의를 가짐.

- 기존의 농산물 가격 예측 연구는 대부분 일별 가격자료를 활용하여 다양한 머신러닝 및 딥러닝 모형의 예측 성능을 비교하는 데 중점을 두었으나, 실제 농업관측 업무에서 활용되는 월별 가격 전망체계를 대상으로 인공지능 모형의 적용 가능성을 종합적으로 검토한 연구는 많지 않았음.

- 본 연구는 이러한 한계를 보완하기 위해 농업관측센터의 실제 업무환경을 고려하여 월별 가격자료를 활용하였으며, 기존 시계열 모형과 머신러닝을 결합한 하이브리드 예측 모형을 구축함으로써 현장 적용 가능성을 높였다는 점에서 기존 연구와 차별성을 가짐.

○ 또한 본 연구는 예측 정확도 향상뿐만 아니라 가격 변동의 원인을 함께 설명할 수 있는 분석체계를 제시하였음.

- SHAP 기반 변수 중요도 분석을 통해 국내 반입량, 저장 관련 변수, 소비자 물가지수, 기상요인 등이 가격 형성에 미치는 영향을 정량적으로 제시함으로써 예측 결과에 대한 해석 가능성을 높였음.
- 이는 단순히 가격을 예측하는 수준을 넘어 가격 변동의 원인을 객관적으로 설명할 수 있다는 점에서 정책 담당자의 의사결정을 지원하는 데 중요한 의미를 가짐.

○ 농업관측 업무 측면에서도 본 연구의 활용 가능성이 크다고 할 수 있음. 본 연구에서 개발된 예측모형은 기존 가격 전망 체계를 대체하기 위한 것이 아니라, 품목 담당자의 전문적인 판단을 지원하는 보조 의사결정 도구로 활용하는 것이 바람직함.

- 특히 인공지능 예측 결과와 전문가의 정성적 판단을 함께 활용할 경우 기상 재해, 저장업체의 출하 전략 변화, 수입 여건 변화 등 급격한 시장환경 변화에도 보다 유연하게 대응할 수 있을 것으로 기대됨.
- 또한 반복적인 가격 예측 업무의 자동화와 분석시간 단축을 통해 농업관측 업무의 효율성을 높이고, 담당자의 분석 역량을 보다 전략적인 업무에 집중할 수 있도록 지원할 수 있음.

3.2. 정책적 시사점

○ 본 연구는 향후 농업관측센터의 양파 가격 전망 체계를 고도화하고, 데이터 기반 농업관측으로 전환하기 위한 방향을 제시한다는 점에서 정책적 의미가 있음.

- 기존 양파 가격 전망은 생산량, 출하량, 재고량 등 주요 수급 정보를 기반으로 전문가 판단을 결합하는 방식으로 운영되어 왔으며, 최근 5개년 기준으로 비교적 안정적인 예측력을 확보하고 있음.
- 그러나 기상변화, 저장업체의 출하 전략, 시장 심리 변화 등 사전에 정량화하기 어려운 요인이 가격 변동에 미치는 영향이 확대되면서 기존 전망 방식만으로는 급격한 시장 변화를 충분히 설명하는 데 한계가 있음.

○ 첫째, 농업관측센터의 양파 가격 전망 체계는 기존 수급 분석 체계를 토대로 인공지능 기반 예측모형을 보완적으로 활용하는 방향으로 발전할 필요가 있음.

- 인공지능 모형은 과거 가격 흐름과 다양한 외생변수를 학습하여 사람이 사전에 규칙화하기 어려운 복잡한 가격 변동 패턴을 탐색할 수 있다는 장점이 있음.
- 특히 본 연구에서 확인된 바와 같이 양파 가격은 계절성과 자기회귀 구조뿐만 아니라 특정 시점에서 발생하는 비선형적 충격 요인이 존재하므로, 시계열 모형과 머신러닝 모형을 결합하는 방식이 기존 전망 체계를 보완하는 효과적인 방법이 될 수 있음.

○ 둘째, 농업관측 데이터의 수집 및 관리 체계를 고도화할 필요가 있음.

- 현재 양파 가격 전망은 표본농가 조사, 산지 조사, 저장업체 조사, 실측조사 등 다양한 자료를 기반으로 이루어지고 있으나, 조사자료 중심의 구조에서는 조사 시점과 실제 시장 상황 간 시차가 발생할 수 있음.
- 따라서 앞으로 기존 조사자료와 함께 도매시장 거래자료, 저장업체 입출고 정보, 기상·원격탐사 자료, 산지 생육 정보 등 다양한 데이터를 연계하여 인공지능 학습에 적합한 데이터 기반을 구축해야 함.

○ 셋째, 인공지능 예측 결과를 활용한 관측 업무 지원 체계를 확립할 필요가 있음.

- 인공지능 모형은 최종 전망가격을 자동으로 결정하는 도구라기보다, 기존 관측 과정에서 전문가가 판단할 수 있는 추가적인 정보와 위험 신호를 제공하는 보조 수단으로 활용하는 것이 바람직함.
- 예를 들어 인공지능 모형이 특정 월의 가격 급등 가능성을 사전에 제시할 경우, 해당 시기의 기상 상황, 산지 출하 동향, 저장업체 출하 의향 등을 집중적으로 점검하여 관측 판단의 근거 자료로 활용할 수 있으며, 이를 통해 기존 전문가 중심 관측체계를 데이터 기반 의사결정 체계로 발전시킬 수 있을 것임.

○ 넷째, 인공지능 기반 전망 시스템 구축을 위해서는 예측 정확도뿐만 아니라 결과에 대한 설명 가능성을 확보해야 함.

- 농업관측 결과는 생산자, 유통업체, 정책 담당자의 의사결정에 직접 활용되기 때문에 단순한 예측값 제공보다는 가격 변동의 주요 원인과 영향 요인을 함께 제시하는 것이 중요함. 따라서 향후에는 변수 중요도 분석, 영향 요인 시각화 등 설명 가능한 인공지능 기법을 적용하여 인공지능 전망 결과에 대한 신뢰성과 활용성을 높일 필요가 있음.

○ 다섯째, 본 연구에서 제시한 인공지능 기반 가격 예측 체계는 양파뿐만 아니라 가격 변동성이 높은 다른 주요 농산물에도 확대 적용할 수 있음.

- 마늘, 배추, 무 등 주요 관측 품목 역시 생산량, 기상, 저장, 소비, 수입 등 다양한 요인이 복합적으로 가격에 영향을 미치므로, 품목별 특성을 반영한 인공지능 예측 모형 개발을 통해 농업관측 전반의 정확성과 효율성을 높일 수 있을 것으로 기대됨.

○ 인공지능 기술의 도입은 기존 농업관측 체계를 대체하는 것이 아니라, 축적된 관측자료와 현장 전문성을 기반으로 예측 정확도를 향상시키는 방향으로 활용되어야 함.

- 향후 농업관측센터가 보유한 장기간의 조사자료와 시장 정보를 체계적으로 데이터화하고 인공지능 분석 기술과 결합한다면, 보다 신속하고 정밀한 농산물 가격 전망 체계를 구축하고 농업인의 합리적인 영농 의사결정을 지원하는 기반이 마련될 것으로 판단됨.

3.3. 향후 과제

○ 본 연구는 월별 자료를 기반으로 분석을 수행하였기 때문에 단기적인 시장 충격이나 일별 가격 변동을 충분히 반영하지 못한 한계가 있음. 또한 저장업체의 실제 출하 의사결정, 소비자의 구매행태, 시장참여자의 기대심리와 같은 비정형 정보를 정량화하는 데에도 제약이 있었음.

- 향후에는 일별 및 순별 자료를 활용한 고빈도 시계열 분석, 위성영상과 항공(드론)영상 등 공간정보의 활용, 온라인 거래정보와 소비정보를 포함한 다양한 디지털 데이터의 연계가 이루어진다면 예측 모형의 성능을 더욱 향상시킬 수 있을 것으로 기대됨.
- 또한 최근 시계열 예측 분야에서 우수한 성능을 보이고 있는 Transformer 기반 딥러닝 모형과 대규모 언어모델(LLM)을 활용한 설명형 인공지능 기술을 농업관측에 접목하는 연구도 지속적으로 추진할 필요가 있음.

○ 본 연구는 양파 가격 예측에 인공지능을 활용할 수 있는 실질적인 적용 가능성을 확인하고, 기존 시계열 모형과 머신러닝을 결합한 하이브리드 접근이 가

격 예측의 정확성과 안정성을 동시에 향상시킬 수 있음을 실증적으로 제시하였음.

- 데이터의 지속적인 축적과 품질관리, 알고리즘의 고도화, 다양한 디지털 데이터의 연계가 이루어진다면 인공지능 기반 가격 예측은 양파뿐만 아니라 마늘, 배추, 무 등 주요 채소류 전반으로 확대 적용될 수 있을 것임.
- 인공지능 기반 농업관측 체계는 우리나라 농산물 수급 안정 정책의 과학적 근거를 강화하고, 데이터 기반의 선제적 정책 결정을 지원하는 핵심 인프라로 자리매김할 것으로 기대됨.

1. ARIMA 모형 차수 탐색 결과

〈부표 1-1〉 SARIMA 차수 탐색

order	seasonal order	AIC	AICc	BIC
(3, 1, 0)	(1, 0, 2, 12)	-142.40	-141.72	-115.14
(4, 1, 0)	(1, 0, 2, 12)	-140.65	-139.80	-109.98
(2, 1, 0)	(1, 0, 2, 12)	-140.01	-139.48	-116.16
(2, 1, 1)	(1, 0, 2, 12)	-138.94	-138.26	-111.68
(4, 1, 1)	(1, 0, 2, 12)	-138.54	-137.51	-104.47
(2, 1, 2)	(1, 0, 2, 12)	-137.93	-137.08	-107.26
(3, 1, 1)	(1, 0, 2, 12)	-137.47	-136.63	-106.81
(2, 1, 1)	(2, 0, 1, 12)	-136.57	-135.89	-109.31
(1, 1, 2)	(1, 0, 2, 12)	-136.14	-135.47	-108.88
(2, 1, 2)	(1, 0, 1, 12)	-131.15	-130.48	-103.90
(1, 1, 1)	(1, 0, 2, 12)	-130.56	-130.04	-106.71
(2, 1, 1)	(2, 0, 2, 12)	-127.84	-126.99	-97.17
(2, 1, 2)	(1, 0, 0, 12)	-109.32	-108.80	-85.47
(2, 1, 1)	(0, 0, 2, 12)	-107.96	-107.44	-84.11
(2, 1, 2)	(0, 0, 2, 12)	-105.17	-104.50	-77.91
(3, 1, 0)	(0, 0, 2, 12)	-102.43	-101.91	-78.58
(3, 1, 0)	(0, 0, 1, 12)	-99.47	-99.08	-79.02
(2, 1, 0)	(0, 0, 2, 12)	-97.12	-96.73	-76.68
(2, 1, 0)	(0, 0, 1, 12)	-94.10	-93.82	-77.06
(0, 1, 1)	(0, 0, 1, 12)	-90.17	-89.99	-76.55
(1, 1, 0)	(1, 0, 0, 12)	-88.46	-88.27	-74.83
(0, 1, 0)	(0, 0, 0, 12)	-63.37	-63.35	-59.96
(0, 1, 0)	(0, 0, 0, 12)	-61.37	-61.32	-54.56

자료: 저자 작성.

〈부표 1-2〉 SARIMAX 차수 탐색

order	seasonal order	AIC	AICc	BIC
(2, 0, 0)	(1, 0, 1, 12)	-65.16	-63.67	-24.27
(2, 0, 0)	(0, 0, 0, 12)	-61.69	-60.65	-27.62
(1, 0, 0)	(0, 0, 0, 12)	-61.27	-60.42	-30.60
(1, 0, 0)	(0, 0, 0, 12)	-59.37	-58.33	-25.29
(1, 0, 1)	(0, 0, 0, 12)	-56.71	-55.46	-19.23
(1, 0, 0)	(1, 0, 1, 12)	-41.79	-40.30	-0.90
(1, 0, 1)	(1, 0, 1, 12)	-41.70	-40.21	-0.81
(1, 0, 0)	(0, 0, 1, 12)	-41.09	-40.05	-7.01
(1, 0, 0)	(1, 0, 0, 12)	-39.44	-38.41	-5.37
(2, 0, 0)	(0, 0, 0, 12)	-34.38	-33.13	3.10
(2, 0, 0)	(1, 0, 1, 12)	-34.58	-32.84	9.71
(2, 0, 0)	(2, 0, 0, 12)	-24.07	-22.59	16.82
(1, 0, 0)	(1, 0, 1, 12)	-16.82	-15.57	20.66
(2, 0, 0)	(2, 0, 1, 12)	-17.06	-15.32	27.23
(1, 0, 0)	(1, 0, 0, 12)	-14.74	-13.49	22.74
(2, 0, 0)	(1, 0, 0, 12)	-14.60	-13.34	22.88
(2, 0, 2)	(1, 0, 1, 12)	-9.11	-6.79	42.00
(2, 0, 0)	(1, 0, 2, 12)	-6.95	-5.20	37.35
(1, 0, 0)	(0, 0, 1, 12)	-4.39	-3.14	33.09
(3, 0, 0)	(1, 0, 1, 12)	-4.72	-2.98	39.57
(3, 0, 1)	(1, 0, 1, 12)	-1.83	0.19	45.87
(2, 0, 0)	(0, 0, 1, 12)	-1.01	0.24	36.47
(2, 0, 0)	(0, 0, 2, 12)	26.90	28.38	67.78
(2, 0, 0)	(2, 0, 2, 12)	32.31	34.33	80.01
(0, 0, 1)	(0, 0, 1, 12)	267.84	269.09	305.32
(0, 0, 1)	(0, 0, 0, 12)	275.36	276.40	309.43
(0, 0, 0)	(0, 0, 0, 12)	448.00	448.85	478.67
(2, 0, 1)	(1, 0, 1, 12)	1,085.78	1,087.52	1,130.07
(2, 0, 1)	(0, 0, 0, 12)	1,275.12	1,276.61	1,316.01
(0, 0, 0)	(0, 0, 0, 12)	1,432.75	1,433.42	1,460.00

자료: 저자 작성.

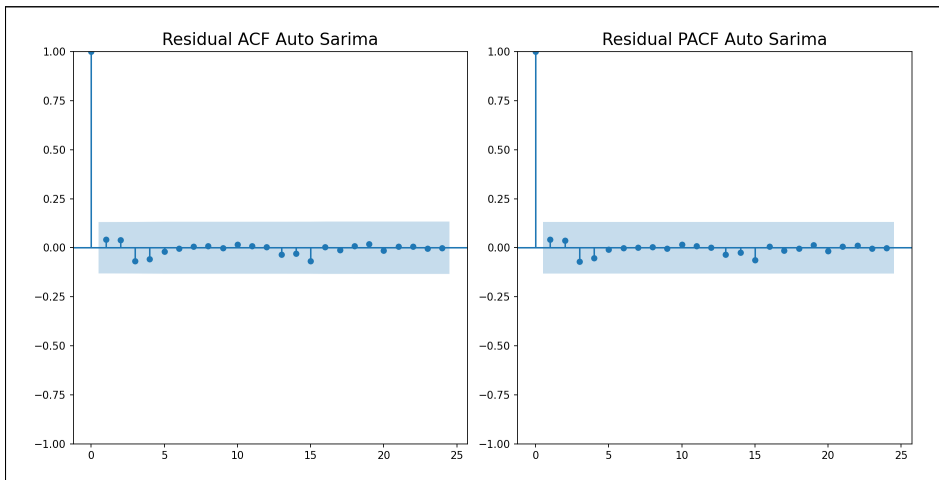
〈부표 1-3〉 STL-ARIMA 차수 탐색

order	seasonal order	AIC	AICc	BIC
(2, 1, 1)	(0, 0, 0, 0)	-151.14	-150.86	-134.10
(2, 1, 2)	(0, 0, 0, 0)	-149.32	-148.93	-128.87
(0, 1, 0)	(0, 0, 0, 0)	-141.62	-141.61	-138.22
(2, 1, 0)	(0, 0, 0, 0)	-141.65	-141.46	-128.02
(0, 1, 1)	(0, 0, 0, 0)	-141.25	-141.14	-131.03
(1, 1, 0)	(0, 0, 0, 0)	-140.51	-140.40	-130.29
(3, 1, 0)	(0, 0, 0, 0)	-140.55	-140.27	-123.51
(1, 1, 1)	(0, 0, 0, 0)	-140.35	-140.17	-126.73
(0, 1, 0)	(0, 0, 0, 0)	-139.63	-139.57	-132.81

자료: 저자 작성.

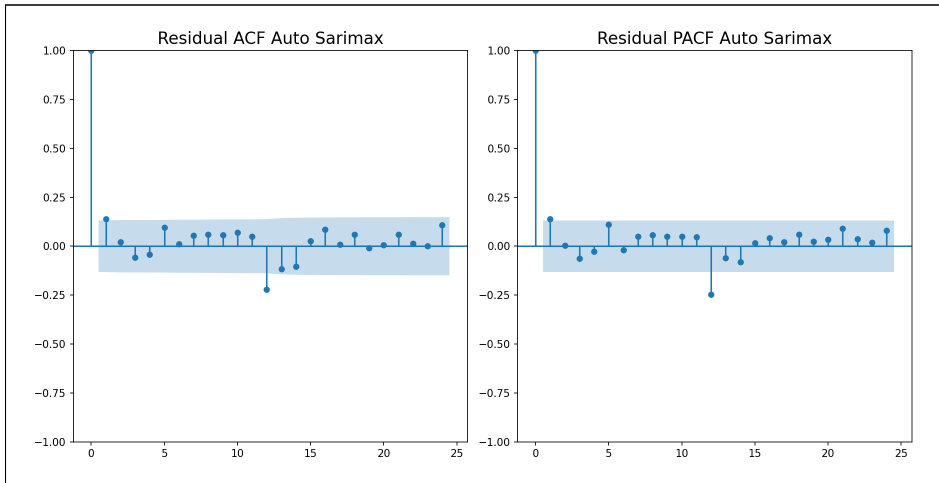
2. 모형별 잔차 자기상관함수

〈부도 2-1〉 SARIMA 잔차의 ACF, PACF



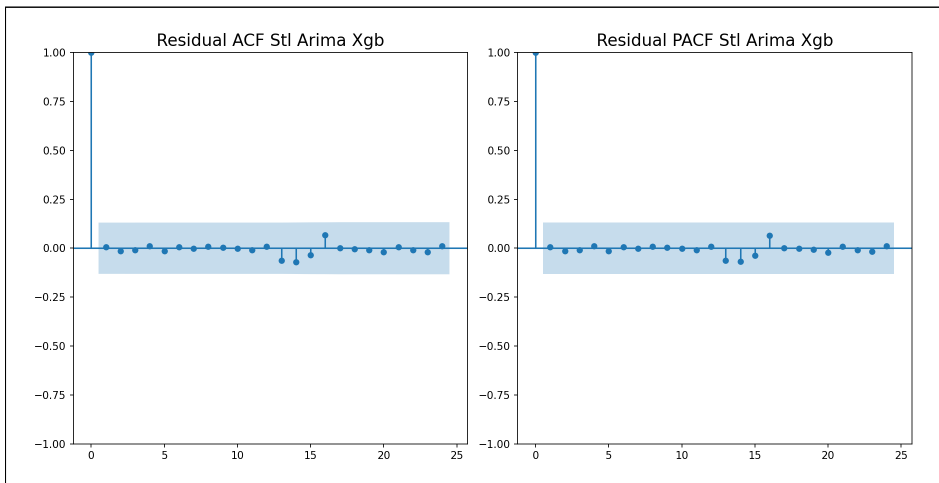
자료: 저자 작성.

〈부도 2-2〉 SARIMAX 잔차의 ACF, PACF



자료: 저자 작성.

〈부도 2-3〉 STL-ARIMA 잔차의 ACF, PACF



자료: 저자 작성.

3. 양파 예측 모형 운영 프로세스

○ 양파 예측 모형 모듈은 원자료, 모델 구성, 예측치 산출 등으로 구성되고 자세한 구성은 <부표 3-1>과 같음.

- 엑셀로 구성된 data 파일은 양파 예측을 위한 원자료와 변수 정의, 출처 등의 내용으로 구성 되어있고, 미래 예측을 위해서 매월 관측된 신규 데이터를 입력해야 함. 향후 api 호출을 통해 자동으로 데이터를 불러오도록 하여 자동화할 수 있음.
- config는 데이터 경로, 변수 설정·변환, 학습데이터 범위, 평가 방식, 튜닝 기준·범위, 앙상블 방식 등 예측 모형의 설정과 기준 등의 값들을 입력하는 모듈임.
- data_processor는 원자료를 바탕으로 탐색적 데이터 분석(EDA)을 진행하고, 모형의 입력데이터로 변환하는 역할을 함. 이용자의 편의를 위해 최근 시계열까지 구축된 데이터를 모형의 입력데이터로 자동으로 변환해 주는 모듈임.
- models은 양파 예측 모형으로 구축된 sarima, sarimax, xgboost, sarima-xgboost, stl-arima-xgboost에 대한 구조를 정의하였고, 모형 최적화를 위한 탐색 범위도 구성 되어있는 공간임.
- trainer는 config, data_processor, models의 모듈 내용을 읽어와 모형의 최적화 및 학습을 진행하고, 훈련데이터 비율에 따라 예측력을 평가하도록 구축 되어있음.
- predictor는 미래 예측 운영에 있어서 예측하는 모듈로, 학습이 끝나고 저장된 모형 파일을 읽어와 새로운 데이터를 바탕으로 미래 예측치를 산출하는 모듈임.

- diagnostics는 학습이 완료된 모형에 대한 강건성 등을 검정하는 모듈로 학습된 모형이 문제가 없는지 확인하는 부분임.
- utils는 data_processor, trainer, predictor 등에서 필요한 모형 정보, 로그, 재현성 설정, 예측력 지표 등이 작성된 모듈로 각 모듈 실행해 결과의 일관성을 유지하게 하는 역할을 함.

〈부표 3-1〉 양파 예측 모형 모듈 구성

모듈	파일 형식	내용
data	xlsx	데이터 파일, 매월 갱신, 예측 월 추가
config	yaml	데이터 경로/처리, 튜닝, 피처, 검증 설정
data_processor	py	전처리, 시차변환, EDA 산출
models	py	모형 정의, 탐색공간 설정
trainer	py	모형 튜닝, 학습, 평가
predictor	py	미래 예측 산출
diagnostics	py	잔차, 계수, SHAP 진단 등
utils	py	공통 유틸
future predictions	xlsx	미래 예측치 배포

주: 각 모듈의 이름은 구축자가 편의상 설정한 것임. 향후 모형 구성이나 기능 확장 등에 따라 모듈명이 달라질 수 있음.

자료: 저자 작성.

○ 월별 미래 예측은 ① config 파일에서 train_size=1 설정, ② data 파일에 실측치 입력 및 예측 월 행 추가, ③ 파이프라인 실행, ④ 예측치 산출로 구성됨.

- config는 데이터 경로, 처리, 튜닝 등의 방식 및 기준 등을 설정하는 부분으로 모형 평가 및 앙상블 가중치 설정을 위해서는 〈부도 3-1〉의 빨간 점선 표시 부분의 train_size를 1 미만으로 설정하면 됨. 보통 8:2나 9:1로 해서 구축을 진행함.
- 실제 미래 예측을 위해서는 〈부도 3-1〉의 config 파일 구성 중에 train_size를 1로 설정하면 됨.

〈부도 3-1〉 config 파일 설정(평가용과 예측용)

```
data:
  sheet_name: "data"
  date_col: "date"
  target_col: "p"
  freq: "MS"
  # observed variables
  lagged_exog_cols: ["q", "import_q", "oil"]
  # variables known at the forecast time t
  known_exog_cols: ["cold", "precip", "rot", "decay"]

transform:
  log_transform: true
  winsor_limits: null

pipeline:
  # set below 1.0 for backtesting and evaluation, set to 1.0 for operational deployment
  # predictor only runs when this equals 1.0 so that inference uses full sample models
  train_size: 1
  forecast_horizon: 1
  ml_multistep_strategy: "recursive"
```

자료: 저자 작성.

- data 파일에서 관측 가능한 양파 도매가격, 국산 양파 반입량, 수입 양파 반입량, 소비자물가지수(또는 유가), 기상 등 최신까지의 과거 데이터를 입력 후 〈부도 3-1〉과 같이 1번 부분에 예측하고자 하는 월을 추가함.
- 그 이후 2번 칸과 같이 예측하고자 하는 종속변수인 p는 공백으로 두고, 과거 시차 변수(lagged)도 빈칸으로 유지함. 미래의 알려진 변수(known)은 3번과 같이 채워서 예측을 위한 데이터 구성을 완료함.
- 위와 같이 구성하고 “python data_processor.py”를 실행하면 예측하고자 하는 월의 행에 과거 시차변수와 미래에 알려진 변수가 같은 행으로 정렬되고 p를 예측하도록 자동으로 구성됨.

〈부도 3-2〉 원자료 입력

date	p	q	import_q	oil	cold	precip	rot	decay
2025-05-01	812.4815	24301	1560	1502.15	-2.33333	75.5	0	0
2025-06-01	767.44	18784	747	1505.43	-2.33333	175.5	0	0
2025-07-01	974.2222	14523	1459	1531.55	-2.33333	175.5	341	1
2025-08-01	1085.64	12804	2440	1535.74	-2.33333	175.5	341	4
2025-09-01	1076.923	14514	2901	1530.67	-2.33333	175.5	341	9
2025-10-01	1091.25	13193	2736	1536.82	-2.33333	175.5	341	16
2025-11-01	1051.042	13254	2364	1619.1	-2.33333	175.5	341	25
2025-12-01	1050.231	13663	2746	1650.49	-2.33333	175.5	341	36
2026-01-01	1048.4	12737	2580	1600.08	-2.33333	175.5	341	49
2026-02-01	992.1	11925	2283	1587.26	-2.33333	175.5	341	64
2026-03-01	714.24	15637	2266	1828.92	-2.33333	175.5	341	81
2026-04-01	723.96	19277	2001	1979.31	-2.53333	97.5	0	0
2026-05-01	570.038	25770.05	94.56	2,005.71	-2.53333	97.5	0	0
2026-06-01	705	19567.78	80.35	1,998.65	-2.53333	97.5	0	0
2026-07-01	1,014.68	25693.19	70.03704	1870.89	-2.53333	97.5	125	1
2026-08-01					-2.53333	97.5	125	4

①
②
③

자료: 저자 작성.

- 데이터까지 구축이 되면 그 다음 과정으로는 모형 학습 과정임. 최근 데이터까지 학습된 모형까지 이용이 필요하면 “python trainer.py --tune”을 진행하고, 기존 모형 정보를 바탕으로 예측하려면 “python trainer.py --refit”을 진행하면 됨.
- 일반적으로 기존 학습된 모형 정보를 이용해 예측치를 산출하는 것을 추천하며, 예측력이 저하 또는 최근 데이터의 급등락이 심화될 경우 또는 정기적 학습일 때만 --tune을 진행함.

〈부표 3-2〉 모형 학습 명령어

구분	명령어
최신 데이터까지 학습	python trainer.py --tune
기존 모형 정보 이용	python trainer.py --refit

자료: 저자 작성.

- 모형 학습이 완료되면 “python predictor.py”를 입력해 미래 예측치인 $t+1$ 의 값을 산출할 수 있음. 이때, 각 모형별 예측치와 함께 가중치로 계산된 앙상블 값이 함께 출력·저장됨.

○ 모형 재학습을 위해서는 config 모듈에서 훈련데이터 비율인 train_size를 1 미만으로 설정하고, predictor 모듈을 제외한 나머지 모듈들을 실행하여 진행함.

- models에서 모형의 튜닝 범위를 확장하거나 축소할 수 있음.
- diagnostics에서 모형의 학습 과정과 잔차 분석, 계수 값 등을 검토하여 모형 방향을 수정할 수 있음.

- 김진·이정일(2019), “딥러닝 기반의 농산물 가격 예측 시스템에 관한 연구”, 한국정보기술학회논문지, 17(6): 27-34.
- 박제인·정수진(2024), “양파 도매 가격 예측을 위한 12가지 모델 성능 및 지역별 결과 비교 분석”, 한국정보처리학회 학술발표대회 논문집, 31(2): 908-909.
- 박종현·임영우·임도현·최연승·안현철(2023), “A Multi-step Time Series Forecasting Model for Mid-to-Long Term Agricultural Price Prediction”, 한국컴퓨터정보학회논문지, 28(2): 201-207.
- 신성호·이미경·송사광(2018), “LSTM 네트워크를 활용한 농산물 가격 예측 모델”, 한국콘텐츠학회논문지, 18(11): 416-429.
- 양진석·김배성·김화년(2020), “주요산지 양파 작형별 가격간 인과관계 분석”, 한국산학기술학회논문지, 21(2): 440-447.
- 윤민정·최수현·이소현·김희웅(2024), “농산물 가격 예측을 위한 딥러닝 모델 개발: 날씨와 경제 변수를 중심으로”, 경영정보학연구, 26(3): 223-239.
- 이도영·양예원·이주형·박지홍·강민구(2021), “Lasso 회귀분석을 활용한 농산물 가격예측 모델 변수 선정 연구”, 디지털문화아카이브지, 4(2): 123-136.
- 홍성민·김종화(2023), “VAR 모형을 이용한 산지별 양파의 선도가격 분석”, 한국산학기술학회논문지, 24(9): 719-726.
- Brennan, M. J.(1958), “The Supply of Storage”, The American Economic Review, 48: 50-72.
- Choudhary, K., G. K. Jha, R. Jaiswal & R. R. Kumar(2025), “A genetic algorithm optimized hybrid model for agricultural price forecasting based on VMD and LSTM network”, Scientific Reports, 15, 9932.
- Gu, Y. H., D.Jin, H. Yin, R. Zheng, X. Piao & S. J. Yoo(2022), “Forecasting agricultural commodity prices using dual input attention LSTM”, Agriculture, 12(2): 256.
- Kannan, M.K. J. & S. Patel(2024), “Sustainable Information Retrieval Techniques for Onion Market Instability Prediction using Machine Learning and Deep Learning Approaches”, International Journal of Advance Research, Ideas and Innovations in Technology, 10(6): 357-362.

- Manogna, R. L., V. Dharmaji & S. Saran(2025), “Enhancing agricultural commodity price forecasting with deep learning”, Scientific reports, 15: 1-24.
- Patil, A., D. Shah, A. Shah & R. Kotecha(2023), “Forecasting Prices of Agricultural Commodities using Machine Learning for Global Food Security: Towards Sustainable Development Goal 2”, International Journal of Engineering Trends and Technology, 71(12): 277-291.
- Tran, N. Q., T. N. Ngoc, Q. Tran, A. Felipe, T. Huynh, A. Tang & T. Nguyen (2023), Predicting Agricultural Commodities Prices with Machine Learning: A Review of Current Research, Proceedings of SOICT 2023, Hochiminh, Vietnam.
- Working, H.(1949), “The theory of price of storage”, The American economic review, 39(6): 1254-1262.
- Zhang, L., L. Feng & R. Liang(2024), Avocado Price Prediction Using a Hybrid Deep Learning Model: TCN-MLP-Attention Architecture.

〈AI 자료〉

OpenAI(2025), ChatGPT(GPT-5, 2026. 6. 12.), 대형언어모델(LLM), <https://chatgpt.com>
본 보고서 〈그림 2-1〉 도식화를 위해 AI 도구를 보조적으로 활용하였음.

〈관련 자료〉

- Bates, J. M. & C. W. Granger(1969), “The combination of forecasts”, Journal of the operational research society, 20(4): 451-468.
- Wang, Y. et al.(2019), “A hybrid ensemble method for pulsar candidate classification”, Astrophysics and Space Science, 364(139): 1-13, Springer.

KREI

www.krei.re.kr



**양파 가격 예측 고도화를 위한
인공지능(AI) 활용방안 연구**

한국농촌경제연구원

전라남도 나주시 빛가람로 601 T.1833-5500 F.061) 820-2211

