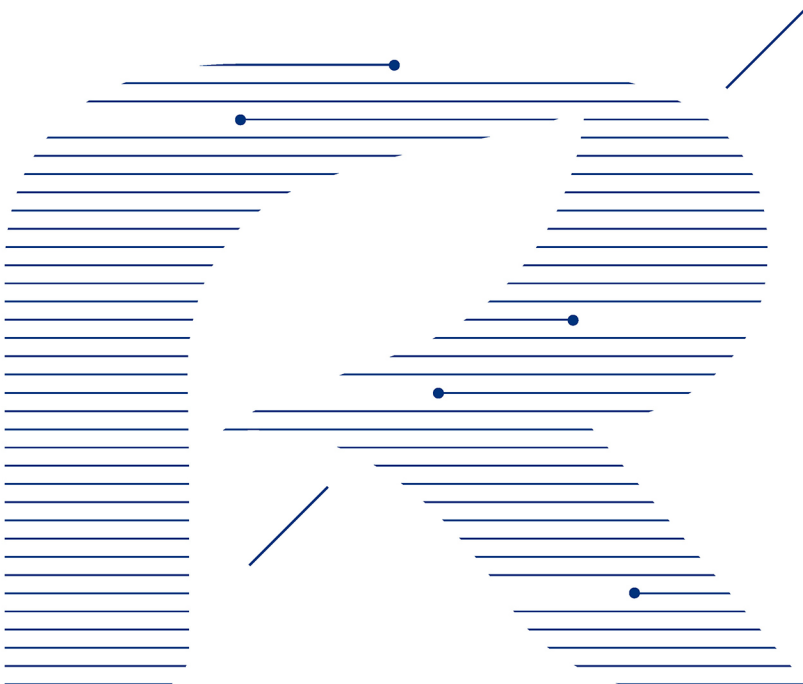


배추 가격 예측력 개선 연구

한은수 · 윤성욱 · 김지수



배추 가격 예측력 개선 연구

한은수 · 윤성욱 · 김지수



연구 담당

한은수 | 전문연구원 | 연구 총괄, 제1~6장 집필

윤성욱 | 전문연구원 | 제2장 집필

김지수 | 연구원 | 제1장 집필

특별연구 E2026-01

배추 가격 예측력 개선 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 | 더크리P&B (주)

I S B N | 979-11-6149-848-5 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.

배추는 국민 식생활과 김치 소비에 밀접하게 연관된 주요 채소로, 정부의 수급 관리와 한국농촌경제연구원(KREI) 농업관측센터의 전망 업무에 핵심적으로 다루어지는 품목이다. 그러나 최근 이상기상의 발생 빈도 증가, 재배 여건 변화, 출하 및 저장 여건의 불확실성 확대 등으로 배추 가격의 변동성이 높아지고 있으며, 이에 따라 수급 및 가격 전망의 안정성과 신뢰성을 높이기 위한 다양한 예측 전략의 도입 필요성이 커지고 있다.

이 연구는 배추 월별 도매가격을 대상으로 시계열 분해와 예측조합을 결합한 가격 예측 알고리즘을 구축하고, 그 예측 성과와 실무 적용 가능성을 검토하였다. 배추 가격 시계열의 계절성과 비계절적 변동을 구분하고 복수의 예측모형 결과를 조합함으로써, 단일 모형에 의존하는 방식의 한계를 보완하고자 하였다.

아울러 본 연구는 제안 알고리즘의 예측치와 KREI 농업관측센터 전망치를 비교하여, 정량적 예측 알고리즘이 기존 전망 체계에서 보완적으로 활용될 수 있는 방안을 검토하였다. 이는 현장 정보와 전문가 판단에 기반한 기존 전망 체계를 대체하기 위한 것이 아니라, 전망 과정의 객관성과 재현성을 높일 수 있는 정량적 예측 도구를 마련하는 데 목적이 있다.

이 연구가 배추 가격 전망의 안정성과 신뢰성을 높이고, 농업관측센터의 정량적 예측 도구 확대와 사후 평가체계 구축을 위한 기초자료로 활용되기를 기대한다. 연구 수행 과정에서 자료 제공과 자문에 도움을 주신 농업관측센터 관계자와 원내·외 전문가 여러분께 감사드린다.

2026. 6.

한국농촌경제연구원장 **한 두 봉**

연구 목적

○ 이 연구는 최근 배추 가격 변동성 확대에 대응하여 예측 불확실성을 완화하고, 안정적이고 객관적인 가격 전망을 위한 정량적 예측 체계를 마련하는 데 목적이 있다. 배추는 기상 여건, 재배면적 변화, 작황 및 출하 상황 등에 따라 가격이 크게 변동하는 대표적인 품목으로, 다양한 경제주체의 의사결정을 지원하기 위한 정확한 예측 정보의 제공이 중요하다.

○ 이를 위해 본 연구는 시계열 분해와 예측조합을 적용한 배추 가격 예측 알고리즘을 구축하고, 예측 성과를 평가하였다. 또한 제안 알고리즘의 예측치를 KREI 농업관측센터 전망치와 비교하여, 정량모형이 제공하는 정보가 기존 전망 체계를 어떻게 보완할 수 있는지 검토하였다. 이를 통해 농업관측 분야에서 객관적이고 재현성 있는 예측 도구의 적용 가능성을 제시하고자 한다.

연구 방법

○ 이 연구는 배추 월별 도매가격 자료를 대상으로 STL(Seasonal and Trend decomposition using Loess) 분해를 적용하여 시계열을 계절 성분과 계절조정 성분으로 구분하였다. 계절 성분에는 계절성 단순기법(seasonal naive)을 적용하고, 계절조정 성분에는 ETS 지수평활 모형, ARIMA 및 ARIMAX 모형, 신경망 자기회귀 모형을 적용하였다. 이후 계절 성분과 계절조정 성분의 예측치를 다시 결합하여 분해 기반 예측치를 도출하였다.

- 개별 예측모형은 시계열 분해를 적용하지 않은 원자료 기반 모형과 시계열 분해 기반 모형으로 구분하여 예측치를 도출하였다. 이 중 분해 기반 개별 예측치들은 단순 평균, 중앙값, MSE 기반 가중평균 방식으로 조합하여 최종적으로 시계열 분해-예측조합 모형의 예측치를 도출하였다. 예측 성과는 MAE(mean absolute error), RMSE(root mean squared error), sMASE(seasonal mean absolute scaled error)를 활용하여 평가하였다. 또한 농업관측센터 전망치와 제안 알고리즘 간의 정보 보완 가능성을 평가하기 위해 Encompassing 검정을 수행하였다.

연구 결과

- STL 시계열 분해와 예측조합을 결합한 모형은 개별 예측모형과 비교하여 전반적으로 낮은 예측오차를 나타내었다. 이는 계절성, 추세 변화, 일시적 변동이 복합적으로 나타나는 배추 가격 시계열에서 변동 구조를 분해하고 예측치를 조합하는 방식이 평균적인 예측 성과를 개선하는 데 기여할 수 있음을 보여준다.
- 다만 가격 급등·급락과 같이 비정상적 수급 충격이 발생하는 시기에는 정량적 예측 알고리즘만으로 실제 가격 변화를 충분히 반영하는 데 한계가 있는 것으로 나타났다. 이는 제안 알고리즘이 과거 가격 흐름과 정량 자료에 기반하여 예측치를 산출하는 구조상, 단기간에 발생하는 외생적 충격을 즉각적으로 포착하기 어렵기 때문이다. 따라서 제안 알고리즘과 함께 현장 기반 정보를 활용하는 농업관측센터 전망을 종합적으로 고려할 필요가 있다.

- 농업관측센터 전망치는 본 연구의 제안 알고리즘보다 전반적으로 우수한 예측 성과를 보였으나, 일부 작형에서는 제안 알고리즘이 상대적으로 낮은 예측오차를 나타내었다. Encompassing 검정 결과, 제안 알고리즘은 농업관측센터 전망치에 포함되지 않은 추가적인 정량 정보를 제공할 가능성이 있는 것으로 확인되었다. 이에 따라 제안 알고리즘은 농업관측센터 전망을 대체하기보다 전망 과정의 객관성과 일관성을 보완하는 정량적 참고자료로 활용될 수 있을 것으로 판단된다.

시사점

- 본 연구의 제안 알고리즘은 농업관측센터의 전망치에 포함되지 않은 추가 정보를 제공할 가능성이 확인된 만큼, 농업관측센터 품목 담당자의 판단을 보완하고 전망 결과의 객관성과 일관성을 높이는 데 기여할 수 있다.
- 배추 가격 급등·급락과 같은 수급 불안정 국면에서는 정량모형만으로 대응하는 데 한계가 있으므로, 현장 조사자료와 전문가 판단을 결합한 전망체계가 필요하다. 아울러 향후에는 확률예측이나 가격 급변에 대한 조기경보 기능을 보완하여 수급 충격에 대한 대응력을 높일 필요가 있다.
- 나아가 제안 알고리즘 예측치와 실제 가격 간 오차를 지속적으로 평가·관리하는 체계를 구축함으로써 모형 성능을 개선하고 적용 범위를 확대해 나갈 필요가 있다. 이를 통해 농업관측 분야의 데이터 기반 의사결정을 강화하고, 배추 가격 전망체계의 안정성을 높이는 데 기여할 수 있을 것으로 기대된다.

제1장 서론

1. 연구의 배경 및 목적	1
2. 연구 내용 및 방법	5
3. 선행연구 검토 및 차별성	8

제2장 배추 가격 전망 체계와 진단

1. 농업관측센터의 정보 수집 체계	15
2. 배추 가격 전망치 산출 구조	17
3. 배추 가격 전망 오차 진단 및 개선 방향	19

제3장 배추 가격 예측 알고리즘

1. 시계열 분해(time series decomposition)	25
2. 배추 가격 예측모형	28
3. 예측조합(forecast combination) 방법	36
4. 시계열 분해-예측조합 기반 배추 가격 예측 알고리즘	38

제4장 예측 성과 평가 방법

1. 예측오차 기반 평가지표	43
2. 예측력 차이 검정: Modified Diebold and Mariano(MDM) 검정	45
3. 예측 포괄성 검정: Encompassing 검정	47

제5장 배추 가격 예측모형 추정 및 예측력 평가 결과

1. 분석 자료	51
2. STL 시계열 분해 결과	59
3. 배추 가격 예측모형 추정 결과	62
4. 예측조합을 위한 가중치 적용 결과	64
5. 가격 예측 정확도 평가 결과	66

제6장 제안 알고리즘의 실무 적용 및 확대 방안

1. 제안 알고리즘과 농업관측센터 전망치 비교	75
2. 제안 알고리즘의 실무 적용 체계	81
3. 실무 적용 확대를 위한 향후 개선 과제	86

부록

1. 배추 도매가격 예측 실무용 R 코드	89
2. 실무용 R 코드 재현을 위한 통합 프롬프트	111

참고문헌	115
------------	-----

제2장

〈표 2-1〉 농업관측센터 주요 정보 수집 체계	16
〈표 2-2〉 농업관측센터 배추 가격 전망치의 월별 평균 오차	21

제3장

〈표 3-1〉 오차, 추세, 계절 성분에 따른 ETS 지수평활 모형의 주요 유형	31
--	----

제5장

〈표 5-1〉 배추 월별 도매가격의 기초 통계량	52
〈표 5-2〉 배추 가격 예측모형의 주요 변수 정의 및 산식	56
〈표 5-3〉 배추 가격 예측모형의 구조 및 주요 설정	64
〈표 5-4〉 분해 기반 개별 예측모형의 조합 방식별 가중치 및 선택 비율	65
〈표 5-5〉 배추 가격 예측 정확도 평가 결과	68
〈표 5-6〉 예측력 차이의 통계적 유의성 검정: MDM 검정 결과	72

제6장

〈표 6-1〉 제안 알고리즘과 농업관측센터 전망치의 예측 정확도 비교	76
〈표 6-2〉 제안 알고리즘과 농업관측센터 전망치의 작형별 예측 정확도 비교 ..	78
〈표 6-3〉 농업관측센터 전망치와 제안 알고리즘 간 Encompassing 검정 결과 ..	80

제1장

〈그림 1-1〉 배추 도매가격(원/10kg) 월별 변화 추이	2
---	---

제2장

〈그림 2-1〉 농업관측센터 배추 가격 전망치 산출 구조	19
〈그림 2-2〉 연도별·월별 농업관측센터 배추 가격 전망치의 오차	22

제3장

〈그림 3-1〉 단층다층 퍼셉트론 구조	34
〈그림 3-2〉 시계열 분해-예측조합 기반 배추 가격 예측 알고리즘 구조	41

제5장

〈그림 5-1〉 배추 도매가격의 분포	52
〈그림 5-2〉 배추의 월별 도매가격(원/10kg) 분포	53
〈그림 5-3〉 학습자료와 평가자료 구간의 구분	59
〈그림 5-4〉 배추 가격 시계열의 STL 분해 결과	61
〈그림 5-5〉 배추 가격 예측 정확도 평가 결과(RMSE 기준)	69
〈그림 5-6〉 본 연구 제안 알고리즘의 예측치와 실제값 비교	70

제6장

〈그림 6-1〉 제안 알고리즘 및 농업관측센터의 예측치와 실제값 비교	79
〈그림 6-2〉 본 연구 제안 알고리즘의 실무 적용 절차	85

1

서론

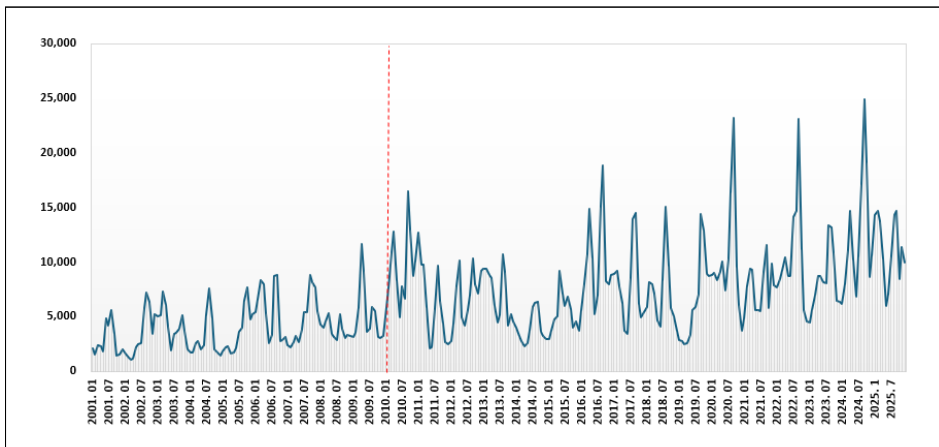
1. 연구의 배경 및 목적

- 한국농촌경제연구원(KREI) 농업관측센터는 36개 농축산물을 대상으로 재배 면적, 생산량, 가격 등의 관측 정보를 제공하고 있음. 이러한 수급 예측 정보는 농가, 유통·가공업체, 정부 등 다양한 경제주체의 의사결정과 계획 수립에 활용되는 중요한 자료 중 하나임.
- 생산 농가는 농업관측센터가 제공하는 품목별 재배면적, 가격 등의 수급 전망 정보를 바탕으로 재배 품목과 규모를 조정할 수 있음.
- 유통·가공 관련 업체는 수급 전망을 반영하여 원료 구매, 재고 운영, 판매 전략 등을 수립하고, 정부는 품목별 수급 불균형에 선제적으로 대응하기 위한 대책을 마련함.

○ 이처럼 농축산물 수급 예측 정보의 활용성과 중요성이 큼에도 불구하고, 최근 이상기상 발생 빈도의 증가, 생산비 상승, 재배 여건 변화 등으로 농산물 가격의 변동성과 예측의 불확실성은 확대되고 있음.

- <그림 1-1>은 2001년 1월~2025년 12월의 월별 배추 도매가격 변화 추이를 나타낸 것임. 배추 도매가격은 2010년 이후 여름철 고온 및 집중호우, 겨울철 한파 등 이상기상의 영향이 커지면서 가격 수준이 상승하고 변동성 또한 확대되는 경향을 보임.
- 따라서 배추 가격의 변동성 확대에 대응하여 예측의 안정성과 신뢰성을 높일 수 있는 전망 체계를 지속적으로 개선하고, 보완해 나갈 필요가 있음.

<그림 1-1> 배추 도매가격(원/10kg) 월별 변화 추이



자료: 서울특별시농수산물공사 유통정보 홈페이지(검색일: 2026. 2. 5.).

○ 농업관측센터는 농산물 수급 및 가격 예측 정보를 제공하기 위해 정성적(qualitative) 분석과 정량적(quantitative) 분석을 병행하고 있음.

- 정성적 분석은 현장 조사와 전문가 의견을 바탕으로 가격 흐름을 신속하게 파악할 수 있다는 장점이 있음. 다만 전문가의 주관적 판단에 의존한다는 점에서 객관적이고 일관된 예측치를 제시하는 데 일정한 제약이 있음.

- 이러한 제약을 보완하기 위해 농업관측센터는 전통적인 회귀 모형 및 시계열 모형과 더불어 최근에는 머신러닝(Machine Learning)·딥러닝(Deep Learning) 알고리즘의 다양한 정량적 기법을 품목별 가격 예측모형으로 활용하고 있음.

○ 이처럼 농업관측센터는 다양한 현장 조사자료와 정량적 예측모형의 결과를 통해 복수의 예측치를 생산하고 있음. 그러나 다양한 예측 정보를 평가하여 최종 전망치 산출 과정에 반영하기 위한 객관적인 기준과 절차가 체계적으로 마련되지 못한 한계는 존재함.

- 예컨대 후보 예측치의 선택 과정에서 농업관측센터 품목별 담당 연구자의 경험과 판단에 의존함으로써, 최종 전망치 선정의 근거를 명확하게 설명하기 어렵고 예측 결과의 일관성이 저하되는 사례가 발생함.
- 따라서 복수의 예측 정보를 체계적으로 평가한 후, 이를 최종 전망 과정에 반영할 수 있는 객관적인 정량적 기준을 마련해 나갈 필요가 있음.

○ 예측조합(forecast combination) 방식은 여러 예측모형의 결과를 조합하여 최종 예측치를 도출하는 방식으로, 농업관측센터의 이러한 한계를 보완할 수 있는 대안적 접근 중 하나로 적용해 볼 수 있음(Bates & Granger, 1969; Clemen, 1989; Wang et al., 2022).

- 예측조합은 단일 모형의 선택 과정에서 수반될 수 있는 예측 불확실성을 완화함으로써 예측의 정확성과 안정성을 동시에 제고하는 방법으로 평가됨.
- 단일 예측모형의 경우 특정 기간의 학습 및 테스트 자료에 대해 예측력이 우수하더라도, 시점 변화에 따라 시계열의 구조, 계절성, 외생변수의 영향력이 달라질 수 있어, 기존 모형의 예측 정확도는 일정하게 유지되기 어려울 수 있음.

- 특히 농산물 가격 시계열은 비선형성, 구조적 변화, 외부 충격의 영향이 빈번하게 나타나므로, 단일 모형만으로 복합적인 변동 특성을 충분히 반영하는 데 한계가 있을 수 있음.

○ 예측조합은 전통적으로 단순히 복수의 예측치를 평균하는 방식에서 개별 예측치에 각기 다른 가중치를 부여하여 조합하는 방법 등 다양한 실증 연구를 통해 적용된 바 있음(Wang et al., 2022). 최근에는 시계열의 복합적인 변동 구조를 반영하기 위한 분해 기반 예측조합(decomposition-based forecast combination) 접근으로 확장되고 있음(Sun et al., 2023).

- 분해 기반 예측조합은 원 시계열을 추세, 계절성, 잔차 등의 구성 성분으로 분해한 뒤, 각 성분의 특성에 적합한 모형을 적용하고, 이후 성분별 예측 결과를 다시 조합하는 방식임.
- 이러한 접근은 원 시계열에 포함된 증장기 추세, 계절적 반복성, 단기적 불규칙 변동을 구분하여 반영할 수 있다는 장점이 있음.
- 특히 시계열을 분해하지 않은 원자료를 직접 예측하는 경우보다 각 성분의 변동 구조를 단순화한 상태에서 예측모형을 적용할 수 있어, 시계열의 구조적 특성을 효과적으로 반영할 수 있음.

○ 배추는 정부의 주요 수급 관리 대상 품목으로, 최근 재배면적 감소와 기상 여건 변화 등의 영향으로 가격 변동성이 확대되고 있음. 이에 따라 단일 모형에 기반한 전망만으로는 시기별 예측 성과를 안정적으로 유지하기 어려울 수 있어 대안적 예측 전략의 도입이 필요함.

○ 본 연구는 배추 월별 도매가격을 대상으로 시계열 분해와 예측조합을 결합한 가격 예측 알고리즘을 구축하고, 그 예측 성과를 평가하는 데 목적이 있음.

- 아울러 본 연구는 제안 알고리즘의 실무 적용 가능성을 검토하고자 함.
 - 본 연구의 제안 알고리즘으로부터 도출된 예측치와 KREI 농업관측센터의 배추 가격 전망치의 예측 정확도 차이를 비교함.
 - 또한 두 예측치가 상호 보완적인 정보를 포함하는지 검토함으로써, 제안 알고리즘이 기존 농업관측센터 전망 체계에서 정량적 보조도로 활용될 수 있는 가능성을 평가함.
 - 최종적으로 실무자가 반복적으로 적용할 수 있는 예측 코드 운영 절차와 향후 개선 과제를 제시함으로써, 배추 가격 전망의 객관성, 재현성 및 안정성 제고에 기여하고자 함.

2. 연구 내용 및 방법

2.1. 연구 내용

○ 제1장. 서론

- 연구의 배경 및 필요성, 연구 목적, 주요 연구 내용 및 방법을 제시함.
- 국내외 선행연구를 검토하고, 본 연구의 차별성을 제시함.

○ 제2장. 배추 가격 전망 체계와 진단

- 농업관측센터의 배추 수급 및 가격 전망 절차와 정보 수집 체계를 정리함.
- 현행 배추 가격 전망 체계를 진단하고, 보완 필요성을 제시함.

○ 제3장. 배추 가격 예측 알고리즘

- 배추 가격 예측을 위한 시계열 분해 방법과 개별 예측모형의 구조를 제시함.
- 개별 예측치를 가중평균 방식으로 조합하는 방법을 제시함.

○ 제4장. 예측 성과 평가 방법

- 예측오차에 기초한 가격 예측 정확도 평가지표를 제시함.
- 예측력 차이의 통계적 유의성을 검토하는 방법과 농업관측센터 전망치와 제안 알고리즘 간 예측 정보의 보완성을 검토하기 위한 검정 방법을 제시함.

○ 제5장. 배추 가격 예측모형 추정 및 예측력 평가 결과

- 분석 대상 자료의 출처, 주요 변수의 정의, 학습 및 평가 구간을 제시함.
- STL 분해 결과와 개별 예측모형의 추정 결과를 정리함.
- 원자료 기반 예측, 분해 기반 예측, 원자료-예측조합 모형, 시계열 분해-예측조합 모형의 예측 정확도를 비교하고, 예측력 차이에 대한 통계적 검정을 수행함.

○ 제6장. 제안 알고리즘의 실무 적용 및 확대 방안

- 본 연구의 제안 알고리즘과 KREI 농업관측센터 전망치의 예측 정확도를 비교하고, 두 예측치의 정보 보완성과 조합 가능성을 검토함.
- 제안 알고리즘을 실제 전망 업무에 적용하기 위한 운영 절차와 향후 개선 과제를 제시함.

2.2. 연구 방법

- 본 연구는 선행연구 검토, 시계열 분해 및 예측모형 구축, 예측조합, 예측력 평가, 통계적 검정, 전문가 자문을 결합하여 수행함.
- 첫째, 선행연구 검토 및 관련 문헌 조사를 수행함.
 - 배추 가격 예측모형, 시계열 분해 방법 및 예측조합 방식을 활용한 국내외 선행연구를 검토함.
- 둘째, 배추 가격 예측을 위한 다양한 예측모형을 적용함.
 - 시계열을 계절(seasonal) 성분, 추세(trend) 및 잔차(remainder) 등의 성분으로 구분하기 위해 STL(Seasonal and Trend decomposition using Loess) 분해 기법을 활용함.
 - 가격 예측모형으로는 ETS(Error, Trend, Seasonality) 지수평활(exponential smoothing), ARIMA(autoregressive integrated moving average), ARIMAX(ARIMA with exogenous variables), 신경망 자기회귀(neural network autoregression: NNAR) 알고리즘을 적용함.
- 셋째, 개별 예측모형의 예측치를 조합하여 최종 예측치를 산출함.
 - 복수의 예측치를 비교적 쉽게 조합하는 방법으로 여러 예측치의 평균(average)과 중앙값(median)을 이용하는 방식을 활용함. 또한 평균제곱 오차(mean squared error: MSE)의 예측성과 기반 조합 방식을 적용함.

○ 넷째, 예측 정확도 평가 및 통계적 검정을 수행함.

- 예측 정확도를 평가하기 위해 MAE(mean absolute error), RMSE(root mean squared error), sMASE(seasonal mean absolute scaled error)를 활용함.
- 모형 간 예측력 차이가 통계적으로 유의한지 확인하기 위해 Modified Diebold and Mariano(MDM) 검정을 통해 검토함.
- 농업관측센터 전망치와 본 연구 제안 알고리즘 두 예측 정보가 상호 보완적으로 활용 가능한지 평가하기 위해 Encompassing 검정을 적용함.

○ 다섯째, 전문가 자문 및 의견 수렴을 통해 본 연구 제안 알고리즘의 성과를 평가하고, 실무적 적용 가능성을 점검함.

- 가격 예측모형 설계, 분석 과정, 결과 해석, 실무 적용 방안의 타당성을 검토하기 위해 전문가 자문회의를 개최하고 관련 의견을 반영함.

3. 선행연구 검토 및 차별성

3.1. 선행연구 검토

가) 배추 가격 예측 관련 국내 연구

○ 국내 배추 가격 예측과 관련하여 초기 수급 구조를 반영한 관측 모형 구축 연구에서 가격 시계열의 통계적 특성과 변동성 분석, 기상 및 생육 등 다양한 외생변수의 활용, 머신러닝 및 딥러닝 기반 예측모형의 적용으로 연구 범위가 점차 확대되었음.

- 먼저, 수급 구조 기반 연구로 한석호·김병률(2004)은 배추를 작형별로 구분하고, 재배면적, 단수, 생산, 소비량 및 가격 간의 연계 구조를 모형화함으로써 배추 수급 및 가격 변동을 설명하고자 하였음. 박지연·박영구(2013)는 배추·무의 예측모형 고도화를 목적으로 재배면적 반응 함수, 생육 단계별 기상정보를 반영한 단수 함수, 반입량과 전월 가격 등을 활용한 가격 함수를 추정하였음.
- 다음으로, 가격 시계열 자체의 통계적 특성과 변동성을 반영하여 예측력을 제고하려는 연구가 수행되었음.
 - 김배성(2005)은 마늘, 양파, 무, 배추의 일별 가격 자료를 대상으로 ARIMA, GARCH, 인공신경망(artificial neural networks) 모형의 예측력을 비교하였음. 분석 결과, 배추의 경우 ARMA(1,1)-GARCH(1,1) 조합모형이 상대적으로 우수한 예측 성과를 보이는 것으로 나타났음. 이는 농산물 가격 예측에서 평균적 가격 수준의 변화뿐만 아니라 가격 변동성 구조를 함께 고려하는 접근이 유용할 수 있음을 시사함.
- 이후에는 가격 자료에 기상, 생육, 재배면적, 수출입 등 다양한 외생 정보를 결합하여 배추 가격 변동의 예측력을 높이려는 연구가 이루어졌음.
 - 서교 외(2008)는 고랭지 배추를 대상으로 기상 요인과 도매시장 가격 간의 관계를 분석하고, 10일 전 가격, 누적 강우량, 최저기온을 활용한 가격 변동 방정식을 구성하여 가격 폭등 확률을 예측하였음.
 - 유도일(2016)은 기상 요인을 고려하여 채소가격 예측모형을 개발하였음.
 - 정호중·정원호(2016)는 배추 가격 변동 요인을 재배면적, 기온, 강수량, 태풍, 수출입량 등으로 구분하고, 작형별 재배단수 모형과 도매가격 결정모형을 추정하였음. 또한 시나리오 분석을 통해 기상 및 재배면적 변화가 배추 가격에 미치는 영향을 구조적으로 분석하였음.

- 장수희 외(2017)는 배추 도매가격만을 이용한 자기회귀(autoregressive; AR) 모형과 포털 문서 기반 비정형 농업기상 정보를 추가한 AR 모형을 구축하였음. 이후 작형별 주 출하 시기의 예측 성과를 비교함으로써 비정형 농업기상 정보의 활용 가능성을 검토하였음. 이는 정형화된 자료와 더불어 비정형 정보 역시 배추 가격 예측에 보조적으로 활용될 수 있음을 시사함.
- 허다솜·정승권(2024)은 가뭄 등 기상재해가 고랭지 배추 수급과 가격에 미치는 영향에 주목하여, 생육 시기별 주산지 기상정보를 구축하고 전년 동일 대비 도매가격 변동률을 예측하였음.

○ 최근에는 농산물 가격 변동의 비선형성과 변수 간 복합적 상호작용을 반영하기 위해 머신러닝 및 딥러닝 기반 예측 연구가 확대되고 있음.

- 박소연 외(2022)는 배추 도매가격에 생육 단계별 기상자료와 주산지 생산량 정보를 활용하고, LSTM(Long Short-Term Memory)과 GRU(Gated Recurrent Unit) 모형에 적용하여 가격 예측을 수행하였음.
- 이창민 외(2022)는 일별 도매가격에 생산량, 주산지 기상, 물가지수, 병해충, 호우 등 재해 정보를 활용하고, Linear, Ridge, Lasso 모형을 결합한 Voting 앙상블 기반 예측모형을 구축하였음.
- 박종현 외(2023)는 중장기 농산물 가격 예측을 위해 다단계 시계열 예측모형을 제안하였음. 해당 연구는 농산물 가격 예측에서 예측 기간과 예측 단계에 따라 모형 구조를 차별화할 필요성을 제시하였음.
- 박찬·이경순(2023)은 농산물 가격 자료의 이상치(outlier)로 인해 딥러닝 예측 성능이 저하될 수 있다는 점에 주목하였음. 해당 연구는 월별 자료를 기준으로 사분위 범위(Interquartile Range: IQR)를 이용하여 이상치를 탐지하고, 이동평균을 통해 시계열을 정제한 후 LSTM 모형을 적용하였음.
- 강동곤 외(2024)는 배추 가격, 기후 요인, 수요, 수입량 등 다양한 변수를

활용하여 LSTM 기반 농산물 가격 예측모형을 구축하였음.

- 이상의 국내 배추 가격 예측 연구는 수급 구조 기반 관측 모형에서 시계열 모형, 다양한 변수를 활용한 머신러닝·딥러닝 기반 예측 알고리즘까지 예측모형의 고도화가 이루어져 왔음. 다만 대부분의 연구는 주로 개별 예측모형의 예측력 비교나 특정 정보의 추가 활용 가능성 검토에 초점을 두고 있어, 복수의 예측치를 체계적으로 조합하여 예측의 정확성과 안정성을 제고하려는 논의는 상대적으로 제한적이었음.

나) 예측조합을 활용한 국내외 농업 분야 연구

- 국외 농업 분야에서는 축산물, 곡물, 농산물 가격 등을 대상으로 다양한 예측 조합 전략이 적용된 바 있음. 다수의 실증 연구는 예측조합을 통해 도출된 예측치가 개별모형의 예측치보다 안정적인 예측 성과를 나타냄을 제시함(Vere & Griffith, 1990; Gil & Albisu, 1993; Colino et al., 2012; Xu, 2017; Xu, 2020; Crespo Cuaresma et al., 2021; Heydari, 2023; Ramsey & Adjemian, 2024).
- 전통적인 예측조합 방법은 개별 예측치의 단순 평균 또는 중앙값 등을 이용하거나, 개별 예측치에 가중치를 달리 부여하여 가중 평균하는 방식이 있음.
 - 개별 예측치에 부여하는 가중치 추정 방법으로 우선, 실제값을 종속변수, 개별모형의 예측치를 설명변수로 설정한 뒤 회귀 분석을 통해 추정된 계수값을 가중치로 활용하는 방식이 있음.
 - 또한 MSE(mean squared error), RMSE(root mean squared error) 등 과거 예측 성과에 기반하거나, AIC 및 BIC 등의 정보 기준(information criterion) 기반 가중치 산정 방식이 다양한 실증 연구에서 적용되었음.

○ 최근에는 원 시계열을 추세, 계절성, 잔차 등의 성분으로 분해한 후 각 성분에 적합한 예측모형을 적용하고, 성분별 예측 결과를 다시 결합하는 분해 기반 예측조합(decomposition-based forecast combination) 전략도 활용되고 있음(Xiong et al., 2018; Jin et al., 2019; Fang et al., 2020; Fu et al., 2022; Jaiswal et al., 2022a; Jaiswal et al., 2022b; Zhu et al., 2022; Sun et al., 2023; Zhang & Tang, 2024).

- 이러한 접근은 원자료를 직접 예측하는 경우보다 시계열에 내재된 이질적 변동 구조를 구분하여 반영할 수 있다는 장점이 있음.
- 농업 분야에서도 STL(Seasonal and Trend decomposition using Loess), EEMD(ensemble empirical mode decomposition), VMD(variational mode decomposition) 등 다양한 시계열 분해 기법을 활용한 연구가 이루어지고 있음.

○ 한편, 예측조합 전략은 국외 농업 분야에서 다양한 품목과 자료를 대상으로 비교적 활발하게 적용된 바 있음. 다만 국내 농산물을 대상으로 한 실증 연구는 상대적으로 제한적이었으며, 최근 국내 일부 품목을 대상으로 Yin et al.(2020)과 한은수 외(2025)에 의해 예측조합 방법이 적용되었음.

- Yin et al.(2020)은 국내 주요 채소류를 대상으로 STL을 활용하여 원 시계열을 추세 및 계절 성분으로 분해한 뒤, LSTM(long short-term memory) 및 Attention 기반 LSTM을 결합한 STL-ATTTLSTM 알고리즘을 제안하였음. 분석 결과, STL 분해를 적용한 예측모형이 원자료를 직접 이용한 LSTM 모형보다 상대적으로 높은 예측 정확도를 보였음.
- 한은수 외(2025)는 양파, 마늘, 건고추를 대상으로 ETS 지수평활, ARIMA, 인공신경망 모형 등을 활용하여 개별 가격 예측치를 산출하고, 이를 조합하기 위해 단순 평균, 회귀 분석, 과거 예측 성과, 정보 기준 등 다양

한 방법을 적용하였음. 분석 결과, 예측조합 방식을 통해 도출된 예측치의 예측 오차가 개별 예측모형보다 낮게 나타났으며, 예측성과 기반의 조합 방식이 상대적으로 우수한 예측력을 보였음.

- Yin et al.(2020)과 한은수 외(2025)는 모두 예측조합을 적용한 예측치가 개별모형의 예측치보다 높은 예측 정확도를 보일 수 있음을 제시하였음. 이러한 결과는 예측조합 전략이 농산물 가격 전망 및 농업관측 업무에서 예측력 개선을 위한 실무적 대안으로 활용될 수 있음을 시사함.

3.2. 선행연구와의 차별성

- 앞서 검토한 국내 배추 가격 예측 연구에서는 다양한 자료와 방법론의 활용 가능성을 제시하였다는 점에서 의의가 있음. 그러나 대부분의 연구는 주로 다양한 예측모형의 성과를 비교한 후, 그중 예측력이 가장 우수한 단일 모형을 선택하는 데 초점을 두어 왔음.
- 특히 배추는 가격 변동성이 크게 나타나는 품목임에도 불구하고, 시계열 분해와 예측조합과 같은 다양한 예측 전략을 배추 가격 예측에 적용하고 이를 기존 농업관측센터 전망 체계와 연계하여 검토한 연구는 제한적임.
- 이에 본 연구는 배추의 월별 도매가격을 대상으로 시계열 분해와 예측조합을 결합한 예측 알고리즘을 적용하고, 개별 예측모형과의 비교를 통해 예측 성과를 체계적으로 평가한다는 점에서 기존 연구와 차별성을 가짐.

○ 본 연구의 구체적인 차별성은 다음과 같이 정리할 수 있음.

- 첫째, 본 연구는 Yin et al.(2020)의 시계열 분해 기반 예측 접근과 한은수 외(2025)의 가중평균 방식 예측조합 접근을 배추 가격 예측 문제에 결합하여 적용함.
- 둘째, ① 원자료 기반 예측, ② 시계열 분해 기반 예측, ③ 원자료-예측조합 기반 예측, ④ 시계열 분해-예측조합 기반 예측 간의 성과를 비교함으로써 제안된 예측 전략의 실질적 성과를 검토함.
- 셋째, 본 연구는 시계열 분해-예측조합 전략을 농업관측센터의 가격 전망 체계와 연계하여 검토함. 이는 본 연구가 단순한 예측모형 비교 연구에 머무르지 않고, 실제 농업관측센터의 실무 업무에서 활용 가능한지 평가함.

2

배추 가격 전망 체계와 진단

1. 농업관측센터의 정보 수집 체계

- 배추 가격은 재배면적, 단수, 생산량, 출하량, 저장량 등의 공급 요인뿐만 아니라 김치 수요, 소비 여건 변화 등의 영향을 받음. 또한 이상기상 발생에 따른 급격한 작황 변화로 단기간 가격 변동성이 확대될 수 있음. 따라서 배추 가격 전망의 정확성을 높이기 위해서는 다양한 현장 조사 자료와 통계 자료를 종합적으로 활용할 필요가 있음.
- 농업관측센터는 배추 수급 및 가격 변동 요인을 포괄적으로 파악하기 위해 표본농가 및 모니터 조사, 실측, 항공촬영, 소비 정보 분석 사업 등 복수의 정보 수집 체계를 운영하고 있음.
 - 표본농가 및 모니터 조사는 재배의향면적, 출하면적, 작황, 저장, 출하시기 등 산지 수급 여건을 파악하는 데 활용됨.
 - 실측 조사는 조사원이 재배 필지에 직접 방문하여 생육 상태, 병해충 발생

여부 등 생산량 추정에 필요한 기초 정보를 확보하는 방식임.

- 항공촬영 조사는 주산지의 재배면적과 수확 진행 상황을 공간 기반 자료로 파악하는 데 활용되며, 기존 현장 조사 자료를 보완하는 역할을 수행함.
- 소비 정보 조사는 기존 공급 중심의 관측 체계를 보완하기 위해 가구 부문과 급식·외식·식품 제조업 등 산업 부문의 소비량, 구매액, 원재료 구매 현황 등을 파악하는 데 활용됨.

○ 농업관측센터의 배추 가격 전망에 활용되는 주요 정보 수집 체계는 <표 2-1>과 같음.

〈표 2-1〉 농업관측센터 주요 정보 수집 체계

구분	대상	주요 조사 내용	가격 전망에서의 역할
표본농가	2,424호	- 재배의향면적, 출하면적 - 정식 및 출하시기 - 단수 및 작황 - 포전거래 가격 - 저장량 및 재고량 - 종자 판매 동향	- 재배면적, 단수, 출하 가능 물량 파악 - 산지 수급 상황 파악 - 단기 시장 변화 파악
모니터	90명 내외	- 산지 작황, 출하 동향, 포전 거래 - 시장 분위기 등 현장 정보	- 산지 수급 상황 파악 - 단기 시장 변화 파악
저장업체	137개소	- 저장량 및 저장 출하 시기 - 월별 출하 계획량, 저장 감모율	- 저장배추 공급 가능량 파악 - 출고 시점 파악
실측 조사	1,627호	- 농가필지 현황 - 품목별 생육 상태 - 병해충 발생 여부	- 생산량 추정 기초자료 제공
항공촬영	133개 읍면	- 주산지 품목별 재배면적 - 출하면적 및 시기 정보	- 공간 기반 재배·출하 정보 보완
수출입 정보	배추, 김치	- 수출입량, 수출입 금액	- 공급 여건 및 수입 대체 가능성 반영
소비 정보 조사	가구·소매유통·급식·외식업·식품제조업	- 소비량, 사용량, 구매액 - 원재료 구매 현황	- 수요 변화 및 소비 여건 파악

자료: KREI 농업관측센터 내부자료.

2. 배추 가격 전망치 산출 구조

○ 농업관측센터의 배추 가격 전망치는 크게 조사 기반 정보와 계량 모형 분석 결과를 종합하여 산출됨.

- 조사 기반 정보는 표본농가 및 모니터 조사, 실측 조사, 항공촬영, 소비 정보 분석 등을 통해 입수된 자료로서, 재배면적, 단수, 생산량, 출하량 등 배추 수급 여건을 파악하는 데 활용됨.
- 계량 모형 기반 정보는 재배면적 반응함수, 단수함수 및 가격함수 등을 추정하여 향후 재배면적, 단수, 생산량 및 가격을 전망한 결과임. 이는 조사자료만으로 파악하기 어려운 다양한 변수 간의 관계를 정량적으로 제시함으로써 조사 기반 정보를 보완하는 역할을 함.

○ 수급 변수별 전망은 재배면적과 단수 전망을 기초로 생산량을 산출한 후, 이를 출하량 및 공급량 전망으로 연계하는 방식으로 이루어짐.

- 재배면적 전망치는 표본농가 조사 결과를 바탕으로 전년 대비 면적 증감률을 산출한 후 지역별 재배면적 비중을 적용하여 추정함. 단수 전망치는 표본농가 및 모니터 조사, 생육 실측 조사, 기상자료 등을 종합하여 추정함.
- 이후 재배면적과 단수 전망치를 결합하여 생산량을 산출하고, 작형별 출하 비중, 월별 출하 패턴, 저장량 및 도매시장 반입 동향 등을 반영하여 월별 출하량 및 공급량을 전망함.

○ 가격 전망 단계에서는 생산량, 출하량, 도매시장 반입량 등 수급 변수별 전망 결과를 바탕으로 복수의 가격 전망치를 산출하고 이를 종합적으로 검토함.

- 신축성 계수 기반 전망은 과거 자료를 이용해 추정한 가격 신축성 계수를 활용하여 출하량 또는 반입량의 변동이 가격에 미치는 영향을 산출하는 방

식임. 예를 들어 가격 신축성 계수가 -0.5이고 반입량이 평년 대비 10% 증가할 것으로 전망될 경우, 가격은 약 5% 하락하는 것으로 추정할 수 있음 (가격 변동률=신축성 계수×반입량 변동률).

- 가격함수 기반 전망은 생산량, 반입량, 직전 작형 가격 등 주요 수급 요인을 설명변수로 설정하고, 이들 변수와 가격 간의 통계적 관계를 이용하여 도매가격을 추정하는 방식임.
- 이와 함께 유통인, 경매사, 농협 등 시장 참여자로부터 수집한 현장 정보를 바탕으로 전문가 전망을 병행함. 전문가 전망은 산지의 출하 동향과 시장 분위기, 품위 변화, 소비 변화 등 정량화하기 어려운 요인을 가격 전망에 반영하는 역할을 함.

○ 최종 가격 전망치는 신축성 계수, 가격 함수 및 전문가 전망 등을 통해 산출된 복수의 가격 전망 후보를 기초로, 관측센터 품목 담당자가 자료의 신뢰도, 최근 수급 여건 등을 종합적으로 검토하여 결정하는 구조임. 따라서 KREI의 배추 가격 전망 체계는 단일 정보 수집 체계에 의해 산출된 결과가 아니라 조사 자료와 계량 모형 분석 결과, 현장 정보 및 전문가 판단을 상호 보완적으로 활용하여 최종 전망치를 도출하는 체계로 이해할 수 있음.

○ KREI 농업관측센터의 배추 가격 전망치 산출 구조는 <그림 2-1>과 같음.

〈그림 2-1〉 농업관측센터 배추 가격 전망치 산출 구조



주: 본 그림의 도식화 과정에서 생성형 AI 도구인 ChatGPT(OpenAI)를 보조적으로 활용하였으며, 최종 내용은 저자가 검토·수정하였음.

자료: 저자 작성.

3. 배추 가격 전망 오차 진단 및 개선 방향

○ 농업관측센터의 배추 가격 전망 체계는 조사 자료, 계량 모형 분석 결과, 현장 정보 및 전문가 판단을 종합적으로 활용함으로써 배추 수급과 가격에 영향을 미치는 다양한 요인을 검토할 수 있다는 점에서 실무적 강점을 지님.

- 다만 배추는 기상 여건, 작황 변화, 출하 시기, 저장 배추의 품위 등 다양한 요인의 영향을 받아 단기간에도 가격이 큰 폭으로 변동할 수 있음. 특히 이상기상이나 예상하지 못한 작황 변화와 같이 전망 시점에 정확하게 파악하기 어려운 수급 충격이 발생할 경우, 전망 시점에 확보된 정보만으로 실제 가격 변화를 충분히 반영하는 데에는 한계가 있음.
- 이에 따라 다양한 정보를 종합적으로 활용하는 현행 전망 체계에서도 실제 도매가격과 전망치 간 오차가 발생할 수 있으며, 단기적인 수급 충격으로 가격 변동이 확대되는 시기에는 전망 오차도 상대적으로 커질 가능성이 있음.
- <표 2-2>는 2021~2025년 농업관측센터의 배추 가격 전망치와 실제 도매가격의 비교가 가능한 3~12월을 대상으로 월별 평균 전망 오차를 제시한 것임. 분석 기간의 평균 실제 가격은 10,406.0원/10kg, 평균절대오차(MAE)는 1,642.9원, 평균절대백분율오차(MAPE)는 14.9%로 나타남.
 - 월별 MAE는 9월이 3,303.8원/10kg으로 가장 컸으며, 다음으로 7월 2,304.5원, 10월 2,076.1원, 8월 1,941.3원 순으로 나타났음. 특히 7~10월의 MAE가 전반적으로 크게 나타나, 여름철 배추 주 출하 시기의 전망 오차가 상대적으로 확대되는 경향을 보였음.
 - MAPE는 7월이 25.8%로 가장 높았으며, 다음으로 11월 19.6%, 10월 19.3%, 9월 16.1% 순으로 나타남.
 - 월별 전망 오차에 대한 평가는 MAE와 MAPE 중 어느 지표를 기준으로 하는지에 따라 다르게 나타날 수 있음. 9월은 MAE가 가장 컸으나, 실제 가격 수준이 상대적으로 높아 MAPE는 7월보다 낮게 나타남. 반면 7월은 두 지표 모두 높은 수준을 보여, 가격 차이의 절대적 규모와 실제 가격 대비 상대적 오차가 동시에 큰 시기로 확인됨.

〈표 2-2〉 농업관측센터 배추 가격 전망치의 월별 평균 오차

구분	평균 실제 가격(원/10kg)	평균절대오차(원/10kg)	평균절대백분율오차(%)
3월	10,049.7	1,187.2	10.0
4월	10,687.1	1,541.8	14.3
5월	7,836.6	860.2	11.0
6월	7,266.9	568.4	7.6
7월	9,738.3	2,304.5	25.8
8월	13,699.4	1,941.3	13.2
9월	17,500.8	3,303.8	16.1
10월	11,066.6	2,076.1	19.3
11월	8,363.1	1,557.6	19.6
12월	7,851.5	1,096.5	12.7
평균	10,406.0	1,642.9	14.9

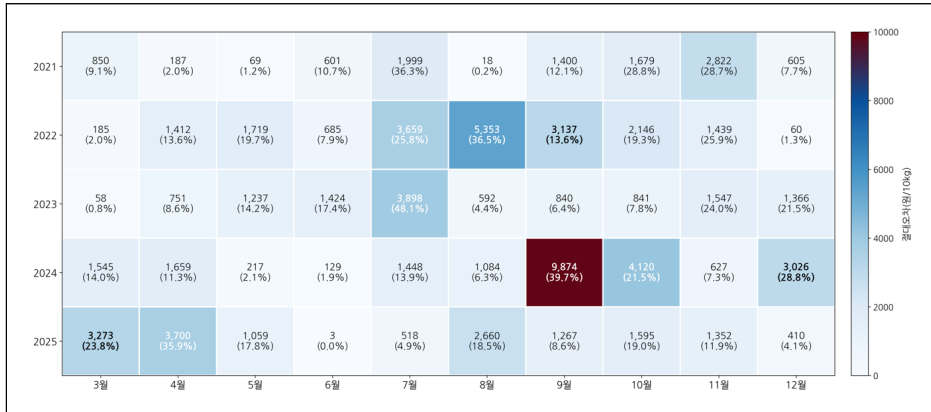
주: 평균절대오차(MAE)는 각 시점의 실제 도매가격(actual)과 농업관측센터 전망치(forecast) 간 차이의 절댓값(|actual-forecast|)을 월별로 평균한 값임. 평균절대백분율오차(MAPE)는 각 시점의 절대오차를 해당 시점의 실제 도매가격으로 나눈 값(|actual-forecast|/actual×100)을 월별로 평균한 값임.

자료: KREI 농업관측센터 자료를 이용하여 저자 산출.

○ 〈그림 2-2〉는 2021~2025년 농업관측센터의 배추 가격 전망치와 실제 도매 가격 간 절대오차 및 절대백분율오차를 연도별·월별로 나타낸 히트맵임. 〈표 2-2〉가 분석 기간의 월별 평균 오차를 요약한 것이라면, 〈그림 2-2〉는 전망 오차가 발생한 구체적인 연도와 월을 식별하고 시점별 오차 분포를 비교할 수 있도록 함.

- 분석 결과, 전망 오차는 모든 연도와 월에 걸쳐 균등하게 분포하기보다 일부 시점에 집중되는 경향을 보였음. 특히 2024년 9월, 2022년 8월, 2024년 10월, 2023년 7월 등에서 절대오차가 상대적으로 크게 나타났으며, 대부분 여름 배추 출하기인 7~10월에 분포하였음.

〈그림 2-2〉 연도별·월별 농업관측센터 배추 가격 전망치의 오차



주: 셀 안의 값은 절대오차(원/10kg)이며, 괄호 안은 실제 가격 대비 절대 오차율(%)인. 셀의 색이 진할수록 절대오차가 큰 것을 의미함. 본 그림의 도식화 과정에서는 생성형 AI 도구인 ChatGPT(OpenAI)를 보조적으로 활용하였으며, 최종 내용은 저자가 검토수정하였음.

자료: 저자 작성.

- 이상의 분석 결과는 배추 가격 전망 오차가 특정 시기와 수급 상황에서 확대되는 경향이 있음을 보여주며, 전망 정확도 제고를 위해 오차 발생 시점과 원인을 체계적으로 점검하고 그 결과를 이후 전망 과정에 반영할 필요가 있음.
- KREI의 배추 가격 전망 체계는 산지의 출하 동향, 시장 분위기, 소비 동향 등 정량화하기 어려운 정보를 전망 과정에 반영할 수 있다는 점에서 강점이 있음.
- 다만 복수의 정보와 품목 담당자의 판단을 종합하여 최종 전망치를 결정하는 구조에서는 전망 오차가 발생한 이후 각 정보의 상대적 기여도와 오차 발생 원인을 명확하게 식별하기 어려운 측면이 있음.
 - 예컨대 특정 시기의 전망 오차가 생산 및 출하 여건의 예상치 못한 변화에 기인한 것인지, 수요와 수입 여건의 변화를 충분히 반영하지 못한 결과인지, 또는 계량 모형의 추정 결과와 현장 정보를 조정하는 과정에서 발생한 것인지를 체계적으로 구분하기 쉽지 않음.

- 이는 다양한 정보를 탄력적으로 활용할 수 있다는 현행 체계의 장점인 동시에, 전망치 산출 과정과 결과에 대한 사후 평가의 제약 요인으로 작용할 수 있음.

○ 따라서 향후에는 현행 전망 체계의 종합적 판단 기능을 유지하되, 전망 과정과 결과를 사후적으로 점검할 수 있는 정량적 관리 체계를 보완할 필요가 있음.

- 우선 월별·작형별 전망치와 실제 가격 간 오차를 지속적으로 추적하고, 가격 급변이나 전망 오차 확대가 발생한 시점에 대해서는 수급 여건을 함께 검토하는 오차 진단 체계를 마련할 필요가 있음.
- 또한 신축성 계수 기반 전망, 가격 함수 기반 전망 및 전문가 전망 등 복수의 전망 후보와 최종 전망치의 조정 과정을 체계적으로 기록하고, 이를 실제 가격과 비교하여 각 전망 방식의 시기별·작형별 예측 성과를 평가할 필요가 있음.
- 나아가 추적된 평가 결과를 바탕으로 예측 성과가 상대적으로 우수한 전망 방식의 활용도를 높이거나 가중치를 차등적으로 부여하고, 반복적으로 나타나는 오차의 원인을 모형과 전망 과정의 개선에 반영하는 환류 체계를 구축할 필요가 있음. 이를 통해 현행 전망 체계가 지닌 실무적 유용성을 유지하면서도 전망 과정의 객관성과 투명성을 제고할 수 있을 것으로 판단됨.

3

배추 가격 예측 알고리즘

1. 시계열 분해(time series decomposition)

- 시계열 자료는 장기적인 방향을 나타내는 추세, 일정한 주기로 반복되는 계절성, 그리고 추세와 계절성으로 설명되지 않는 불규칙 변동이 결합된 형태로 나타날 수 있음. STL(Seasonal and Trend decomposition using Loess)은 국소 회귀 기법인 LOESS를 이용하여 시계열을 추세, 계절, 잔차 성분으로 분해하는 방법임. 이를 통해 시계열에 내재된 구조적 패턴을 구분하고, 각 성분의 특성에 적합한 예측모형을 적용할 수 있음(Yin et al., 2020; Hyndman & Athanasopoulos, 2021).
- 배추 가격 시계열을 P_t , 추세 성분을 T_t , 계절 성분을 S_t , 잔차 성분을 R_t 라고 하면, 덧셈(additive) 분해는 식 (3-1), 곱셈(multiplicative) 분해는 식 (3-2)와 같이 표현할 수 있음.

$$(3-1) P_t = T_t + S_t + R_t$$

$$(3-2) P_t = T_t \times S_t \times R_t$$

○ 덧셈형 분해는 시계열 수준이 변화하더라도 계절 변동과 불규칙 변동의 절대적인 크기가 비교적 일정하게 유지되는 경우에 적합함. 반대로 곱셈형 분해는 가격 수준이 높아질수록 계절적 변동이나 불규칙 변동의 폭도 함께 확대되는 경우에 적합함.

- 배추 가격은 가격 수준이 높은 시기에 변동 폭도 확대되는 특성을 보일 수 있으므로, 원자료에서는 곱셈적 변동 구조가 나타날 가능성이 있음. 이에 본 연구에서는 가격 수준에 따라 분산의 변화를 완화하고 곱셈적 관계를 덧셈적 관계로 전환하기 위해 배추 가격에 로그 변환을 적용하였음.
- 식 (3-2)의 양변에 로그를 취하면 곱셈형 구조는 식 (3-3)과 같이 덧셈형 구조로 변환됨.

$$(3-3) \log(P_t) = \log(T_t) + \log(S_t) + \log(R_t)$$

○ STL은 다양한 시계열 자료에 적용할 수 있는 유연한 분해 알고리즘으로 다음과 같은 특징을 가짐.

- 첫째, 분석자가 계절 주기를 지정할 수 있어 월별, 분기별 자료 등 서로 다른 주기의 시계열에 적용할 수 있음.
- 둘째, 계절 평활 모수의 설정을 통해 계절 성분이 시간에 따라 변화하는 정도를 조정할 수 있음.
- 셋째, 이상치의 영향을 줄이는 강건(robust) 추정 방식을 적용할 경우 일시적인 가격 급등락이나 이상치가 추세나 계절 성분 추정에 과도하게 반영되

는 것을 완화할 수 있음. 이 경우 추세와 계절 성분으로 설명되지 않는 일시적 변동은 주로 잔차 성분에 반영됨.

○ STL 추정 과정은 외부 반복(outer loop)과 내부 반복(inner loop)의 두 단계 과정을 수행함. 외부 반복은 이상치로 판단되는 관측치에 상대적으로 작은 가중치를 부여하여 해당 값이 추세 및 계절 성분 추정에 과도한 영향을 미치지 않도록 조정하는 과정임. 내부 반복은 추세 성분과 계절 성분을 반복적으로 갱신하는 과정으로, 다음과 같이 요약할 수 있음.

- 1) 추세 제거(detrending): 배추 가격 시계열 P_t 에서 추정된 추세 성분 T_t 를 제거하여 계절성을 파악하기 위한 시계열을 구성함.
- 2) 부분 시계열 평활(subseries smoothing): 추세가 제거된 시계열을 1월부터 12월까지의 월별 부분 시계열로 구분하고, 각 부분 시계열에 LOESS 평활을 적용하여 초기 계절 성분을 도출함.
- 3) 저역통과 필터링(Low-pass filtering): 초기 계절 성분에 남아있는 장기적 변동을 제거하기 위해 저역통과 필터를 적용함.
- 4) 계절 성분 추정: 부분 시계열 평활 결과에서 저역통과 필터를 통해 추정된 장기 변동을 제거하여 최종 계절 성분 S_t 를 산출함.
- 5) 계절조정 시계열 구성(de-seasonalization): 배추 가격 시계열 P_t 에서 계절 성분 S_t 를 제거하여 계절조정 시계열을 구성함.
- 6) 추세 평활(trend smoothing): 계절조정 시계열에 LOESS 평활을 적용하여 최종 추세 성분 T_t 를 갱신함. 추세와 계절 성분의 추정이 완료되면 잔차 성분은 가격에서 추세와 계절 성분을 차감하여 산출함. 이러한 과정은 설정된 내부 반복 횟수만큼 반복함.

- 본 연구는 이러한 STL 시계열 분해 결과를 바탕으로 계절적 변동과 추세 및 잔차의 비계절적 변동을 구분하고, 이후 예측모형 구축 및 예측조합 과정에서 각 성분의 특성을 반영하고자 하였음.

2. 배추 가격 예측모형

- 본 연구는 배추 가격 예측을 위해 STL 시계열 분해를 통해 가격 시계열의 구조적 특성을 진단한 후, 전통적 시계열 모형, 인공신경망 계열의 예측모형을 적용함(Hyndman & Athanasopoulos, 2021).
 - 시계열 모형으로는 ETS 지수평활, ARIMA 및 계절 ARIMA, ARIMAX 모형을 활용하였으며, 인공신경망 계열 모형으로는 신경망 자기회귀(NNAR) 모형을 적용함.
 - 각 기법은 배추 가격 시계열의 서로 다른 특성을 반영하기 위해 선정하였음. ETS 지수평활 모형은 가격 시계열의 수준, 추세, 계절성 변동을 비교적 적은 수의 파라미터로 평활화하여 예측하는 데 유용하며, ARIMA 모형은 가격 시계열에 내재된 자기상관(autocorrelation) 구조를 통계적 검정 절차에 기초하여 반영할 수 있다는 장점이 있어 선정하였음. ARIMAX 모형은 ARIMA 구조에 반입량, 수입량 등 외생변수를 추가함으로써 가격 변동과 관련된 수급 정보를 함께 고려할 수 있다는 장점이 있음.
 - 한편, 최근 농산물 가격 예측 연구에서 기계학습 및 인공신경망 계열 모형의 적용이 확대되고 있음. 이에 본 연구에서는 시계열 자료에 적용 가능한 비교적 단순한 구조의 신경망 자기회귀(NNAR) 알고리즘을 후보 예측모형으로 포함함. NNAR 모형은 과거 가격 시차값과 계절 시차값을 입력변수로 사용하면서도, 입력값과 미래 가격 간의 관계를 선형적으로 제한하지

않는다는 점에서 전통적 자기회귀 모형과 차별성을 가진.

- 본 연구는 ETS, ARIMA, ARIMAX, NNAR 모형의 경우 원자료 기반 가격 시계열과 STL 분해 이후 계절조정 성분을 예측하는 데 활용됨. 다만 STL 분해를 통해 도출된 계절 성분은 별도의 복잡한 모형을 적용하기보다 월별 반복 패턴을 직접 반영할 수 있는 계절성 단순 모형(seasonal naive method: SNAIVE)을 적용함.
- 이와 같이 본 연구는 전통적 시계열 모형과 인공신경망 모형을 함께 적용함으로써 배추 가격 시계열에 내재한 수준 변화, 추세, 계절성, 자기상관, 외생변수 효과 및 비선형적 변동 가능성을 종합적으로 검토하고자 함.

2.1. ETS(Error, Trend, Seasonality) 지수평활 모형

- 지수평활은 과거 관측값의 가중평균(weighted average)을 이용하여 미래의 값을 예측하는 시계열 예측 기법임. 이때 최근 관측값에는 상대적으로 큰 가중치가 부여되고, 관측 시점이 오래될수록 가중치가 지수적으로 감소함. 따라서 지수평활은 최근의 가격 흐름을 보다 크게 반영하면서도 과거 정보 전체를 일정 수준 활용할 수 있다는 특징을 가짐.
- 지수평활 모형은 자료에 포함된 구성요소에 따라 단순 지수평활, 홀트(Holt) 모형, 홀트-윈터스(Holt-Winters) 모형으로 확장될 수 있음.
 - 단순 지수평활은 추세와 계절성이 뚜렷하지 않은 자료에 적용되며, 수준 성분만을 평활화하여 예측함.

- Holt 모형은 수준 성분과 함께 추세를 포함하며, 장기 예측에서 추세가 과도하게 증가하거나 감소하는 문제를 완화하기 위해 감쇠 추세(damped trend)를 적용할 수 있음.
 - Holt-윈터스 모형은 수준과 추세에 더해 계절성을 포함하는 모형으로, 계절 변동이 일정한 경우 덧셈 계절성, 시계열 수준에 비례하여 계절 변동 폭이 달라지는 경우 곱셈 계절성을 적용함.
- 최근에는 이러한 지수평활 계열 모형을 ETS(Error, Trend, Seasonality) 체계로 통합하여 표현하는 방식이 널리 사용됨. ETS 모형은 시계열을 오차(E), 추세(T), 계절성(S)의 조합으로 정의하며, 각 구성요소는 자료 특성에 따라 없음(N), 덧셈(A), 곱셈(M), 감쇠 추세(Ad) 등의 형태로 설정됨.
- ETS 모형은 일반적으로 ETS(E,T,S)의 형태로 표기됨. 여기서 E는 오차의 형태, T는 추세의 형태, S는 계절성의 형태를 의미함.
- 예를 들어 ETS(A,A,M)은 덧셈 오차(additive error), 덧셈 추세(additive trend), 곱셈 계절성(multiplicative seasonality)을 갖는 모형을 의미하며, ETS(M,Ad,N)은 곱셈 오차(multiplicative error), 덧셈 감쇠 추세(additive damped trend), 계절성 없음(no seasonality)을 가정한 모형을 의미함.
- <표 3-1>은 오차, 추세, 계절 성분의 조합에 따라 구분되는 ETS 모형의 주요 유형을 제시한 것임. ETS 모형은 각 구성요소의 형태에 따라 단순 지수평활, Holt 선형추세 모형, 감쇠 추세 모형, Holt-윈터스 계절모형 등으로 구분될 수 있으며, 이러한 구조는 자료의 추세 및 계절 변동 특성에 따라 선택됨.

〈표 3-1〉 오차, 추세, 계절 성분에 따른 ETS 지수평활 모형의 주요 유형

덧셈 오차 모형 (additive error models)		계절 성분		
		N (없음)	A (덧셈)	M (곱셈)
추세 성분	N (없음)	(A, N, N)	(A, N, A)	(A, N, M)
		단순 지수평활		
	A (덧셈)	(A, A, N)	(A, A, A)	(A, A, M)
		홀트 선형추세	덧셈 홀트-윈터스	곱셈 홀트-윈터스
	A_d (덧셈 감쇠)	(A, A_d, N)	(A, A_d, A)	(A, A_d, M)
		감쇠 추세		
곱셈 오차 모형 (multiplicative error models)		계절 성분		
		N (없음)	A (덧셈)	M (곱셈)
추세 성분	N (없음)	(M, N, N)	(M, N, A)	(M, N, M)
	A (덧셈)	(M, A, N)	(M, A, A)	(M, A, M)
	A_d (덧셈 감쇠)	(M, A_d, N)	(M, A_d, A)	(M, A_d, M)

자료: Hyndman & Athanasopoulos(2021)를 참고하여 저자 작성.

○ 본 연구는 통계 프로그램 R의 forecast 패키지에 포함된 ets() 함수를 이용하여 ETS 모형을 추정함. ets() 함수는 사용자가 오차, 추세, 계절성 구조를 사전에 지정하지 않을 경우 가능한 ETS 후보 모형을 학습자료에 적합한 뒤, 정보 기준(information criterion)을 이용하여 적합도가 우수한 모형을 자동적으로 선택함.

- 일반적으로 모형 선택에는 AIC, AICc, BIC 등의 정보 기준이 활용되며, forecast 패키지의 기본 선택 기준은 AICc임.

○ 본 연구에서 ETS 모형은 배추 가격 시계열의 수준, 추세 및 계절 변동을 반영하는 개별 예측모형으로 활용됨. 또한 원자료 기반 시계열과 STL 분해 이후 계절조정 성분에 ETS 모형을 적용함으로써, 반복적 계절 변동을 제외한 가격 흐름의 수준 및 추세 변화를 예측하고자 함.

2.2. ARIMA 및 ARIMAX 모형

○ ARIMA는 ETS 모형과 함께 시계열 예측에 널리 활용되는 기법임. ETS 모형이 시계열의 수준, 추세, 계절성에 대한 평활화에 기초한다면, ARIMA 모형은 시계열에 포함된 자기상관 구조를 모형화하는 데 목적이 있음. 따라서 본 연구는 ETS와 ARIMA의 서로 다른 모형화 관점을 고려하여 두 모형을 배추 가격 예측에 함께 적용하고, 예측 성과를 비교·평가함.¹⁾

○ ARIMA 모형은 자기회귀(autoregressive: AR) 과정, 차분(integration: I), 이동평균(moving average: MA) 과정을 결합한 모형임. 자기회귀 과정은 현재 값이 과거 값의 선형 결합으로 설명된다고 보는 방식이며, 이동평균 과정은 현재 값이 과거 예측오차의 선형 결합으로 설명된다고 보는 방식임. 비정상 시계열의 경우에는 차분을 통해 정상성을 확보한 뒤 ARMA 구조를 적용함.

- 일반적으로 ARIMA 모형은 $ARIMA(p, d, q)$ 로 표기함. 여기서 p 는 자기회귀 차수, d 는 차분 횟수, q 는 이동평균 차수를 의미함.

○ 월별 가격 자료와 같이 계절성이 존재하는 시계열의 경우 ARIMA 모형은 계절 ARIMA, 즉 SARIMA 모형으로 확장될 수 있음. SARIMA 모형은 비계절 부분인 (p, d, q) 와 계절 부분 $(P, D, Q)_m$ 로 구성됨.

- 여기서 P, D, Q 는 각각 계절 자기회귀 차수, 계절 차분 횟수, 계절 이동평균 차수를 의미하며, m 은 계절 주기를 나타냄. 월별 자료의 경우 일반적으로 $m=12$ 가 적용됨.

¹⁾ 선형 ETS 모형의 일부는 특정 ARIMA 모형과 이론적으로 대응될 수 있음. 예컨대, $ETS(A, N, N)$ 는 $ARIMA(0, 1, 1)$, $ETS(A, A, N)$ 는 $ARIMA(0, 2, 2)$, $ETS(A, Ad, N)$ 는 $ARIMA(1, 1, 2)$ 와 대응되는 것으로 알려져 있음(Hyndman & Athanasopoulos, 2021). 다만 이러한 대응관계는 일부 단순 구조에 해당하며, 실제 예측에서는 모형 선택 절차와 오차 구조에 따라 결과가 달라질 수 있음.

- 한편, ARIMA 모형에 배추 가격에 영향을 미치는 외생변수를 포함하면 ARIMAX로 확장될 수 있음. ARIMAX 모형은 가격 시계열의 자기상관 구조 뿐만 아니라 반입량, 수입량 등 수급 관련 정보를 함께 반영할 수 있다는 점에서 농산물 가격 예측에 유용하게 활용될 수 있음. 본 연구는 ARIMAX 모형을 함께 검토함으로써, 단변량 시계열 모형 대비 예측력 개선 가능성을 평가함.
- 본 연구는 R의 forecast 패키지에 포함된 `auto.arima()` 함수를 이용하여 ARIMA 및 ARIMAX 모형을 추정함. `auto.arima()` 함수는 단위근 검정, 정보 기준에 기초한 차수 선택, 최대우도추정(maximum likelihood estimation)을 결합하여 ARIMA 구조를 자동으로 선택함.
 - 이 함수는 Hyndman-Khandakar 알고리즘에 기초하여 후보 모형을 탐색하며, 구체적인 절차는 Hyndman & Athanasopoulos(2021)를 참고할 수 있음.
- 본 연구는 원자료 기반 가격 시계열과 STL 분해 이후 계절조정 성분에 각각 ARIMA 및 ARIMAX 모형을 적용함. 이를 통해 배추 가격의 자기상관 구조와 외생변수 효과가 원자료 기반 예측과 분해 기반 예측에서 어떠한 예측 성과를 보이는지 비교하고자 함.

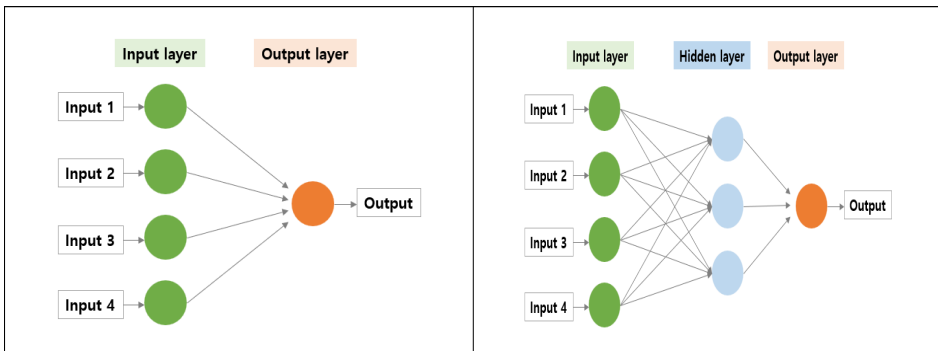
2.3. 신경망 자기회귀(Neural network autoregression: NNAR) 모형

- 인공신경망은 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성되며, 입력값과 출력값 간의 관계를 학습하는 기계학습 알고리즘임.
 - 입력값은 가중치를 통해 은닉층에 전달되고, 은닉층에서는 활성화 함수(activation function)를 통해 비선형 변환이 이루어짐. 이러한 구조를

통해 인공지능망은 선형 모형으로 포착하기 어려운 복잡한 비선형 관계를 학습할 수 있음.

○ 초기 형태의 퍼셉트론(perceptron)은 다수의 입력 신호를 가중합 한 뒤 임계값 또는 활성화 함수를 기준으로 출력을 결정하는 단순한 구조의 모형임. 단층(single-layer) 퍼셉트론은 입력층과 출력층으로 구성되며, 하나 이상의 은닉층을 추가하면 다층(multi-layer) 퍼셉트론으로 확장됨. 다층 퍼셉트론은 은닉층의 비선형 활성화 함수를 통해 입력변수와 출력변수 간의 복잡한 관계를 보다 유연하게 근사할 수 있음. 단층·다층 퍼셉트론의 기본 구조는 <그림 3-1>에 제시하였음.

<그림 3-1> 단층다층 퍼셉트론 구조



자료: 저자 작성.

○ 본 연구는 인공지능망 모형 중 시계열 예측에 적용 가능한 신경망 자기회귀 모형(NNAR)을 활용함. NNAR 모형은 과거 관측값의 시차항을 입력변수로 사용하고, 은닉층을 포함한 feed-forward neural network 구조를 통해 미래 값을 예측하는 방식임.

- 즉, 과거 가격 정보에 기초하여 미래 가격을 예측한다는 점에서는 자기회귀 모형의 성격을 가지지만, 입력값과 예측값 간의 관계를 선형으로 제한

하지 않는다는 점에서 전통적 자기회귀 모형과 차이가 있음.

- NNAR 모형은 일반적으로 $NNAR(p, P, k)_m$ 형태로 표현됨. 여기서 p 는 비계절 시차의 수, P 는 계절 시차의 수, k 는 은닉층 노드(node) 수, m 은 계절 주기를 의미함.
- 월별 배추 가격의 경우 $m = 12$ 가 적용되며, 최근 가격 시차값과 전년 동월 등 계절 시차값이 입력변수로 활용될 수 있음. 이 입력값들은 은닉층의 노드를 거쳐 비선형 방식으로 결합되며, 최종적으로 미래 배추 가격 예측값을 산출함.

○ 본 연구는 R의 forecast 패키지에 포함된 nnetar() 함수를 이용하여 NNAR 모형을 추정함. nnetar() 함수는 시계열 자료의 과거 시차값을 입력변수로 구성하고, 은닉층을 갖는 feed-forward neural network에 적합하여 예측값을 산출함.

- 비계절성 시계열의 경우 여러 후보값에 대해 $AR(p)$ 모형을 적합한 뒤 AIC 기준값이 가장 낮은 비계절 시차 p 를 선택함. 계절성 시계열의 경우 계절 시차 P 가 포함될 수 있음.
- 은닉층 노드 수 k 는 입력 노드 수를 바탕으로 설정되며, 일반적으로 $k = (p + P + 1)/2$ 의 정수값이 기본값으로 사용됨.

○ 본 연구에서 NNAR 모형은 배추 가격 시계열에 존재할 수 있는 비선형 자기회귀 패턴을 반영하기 위한 후보 예측모형으로 활용됨. 또한 원자료 기반 가격 시계열과 STL 분해 이후 계절조정 성분에 NNAR 모형을 적용함으로써, 비선형 구조가 분해 기반 예측에서 추가적인 예측력 개선에 기여하는지를 검토함.

3. 예측조합(forecast combination) 방법

- 예측조합은 서로 다른 예측모형에서 산출된 개별 예측치를 결합하여 최종 예측치를 도출하는 방법임. 개별 예측모형은 자료에 내재한 서로 다른 특징을 반영하므로, 특정 모형이 모든 시점에서 일관되게 가장 우수한 예측 성과를 보인다고 가정하기는 어려움. 따라서 복수의 예측치를 조합하면 개별모형의 일시적 예측 실패 가능성을 완화하고, 보다 안정적인 예측 결과를 얻을 수 있음(Bates & Granger, 1969; Clemen, 1989; Wang et al., 2022).
- 본 연구는 개별 예측치를 조합하는 방법으로 단순 평균, 중앙값, 평균제곱오차(mean squared error: MSE) 기반 가중평균(MSE-weighted average)을 적용함. 단순 평균과 중앙값 조합은 별도의 가중치 추정 과정이 필요하지 않은 비교적 단순한 방식이며, MSE 기반 가중평균은 과거 예측오차가 작은 모형에 상대적으로 큰 가중치를 부여하는 방식임.
- 단순 평균 조합은 각 개별 예측치에 동일한 가중치를 부여하여 조합 예측치를 산출하는 방법임. N 개의 개별 예측모형이 존재하고, 시점 t 에 대한 i 번째 모형의 예측치를 $\hat{y}_{i,t}$ 라고 하면, 단순 평균 조합 예측치는 식 (3-4)와 같이 정의할 수 있음.

$$(3-4) \hat{y}_t^{mean} = \frac{1}{N} \sum_{i=1}^N \hat{y}_{i,t}$$

- 단순 평균 조합은 여러 실증 연구에서 예측력 향상에 기여할 수 있는 예측 전략으로 제시되어 왔음. 정교한 가중치 추정 방식이 이론적으로 우월할 수 있음에도, 실제 자료에서는 단순 평균 조합이 더 안정적인 예측 성과를 보이는

사례가 다수 보고됨.

- 이는 단순 평균이 가중치 추정 과정에서 발생할 수 있는 추정오차와 불확실성을 피할 수 있고, 개별 예측치의 편향(bias)과 분산(variance)을 완화하는 데 기여할 수 있기 때문임.

○ 중앙값(median) 조합은 개별 예측치를 크기순으로 정렬한 뒤 중앙에 위치한 값을 조합 예측치로 사용하는 방식임. N 개의 개별 예측치가 주어졌을 때, 중앙값 조합 예측치는 식 (3-5)와 같이 표현할 수 있음.

$$(3-5) \hat{y}_t^{median} = median(\hat{y}_{1,t}, \hat{y}_{2,t}, \dots, \hat{y}_{N,t})$$

○ 중앙값 조합은 단순 평균 조합에 비해 극단적인 예측치(outlier)에 대한 민감도가 낮다는 장점이 있음. 따라서 가격 급등락 시기나 일부 모형의 예측 실패 가능성이 존재하는 경우, 중앙값 조합은 보다 강건한 조합 방식으로 활용될 수 있음.

○ MSE 기반 가중평균은 개별 예측모형의 과거 예측오차에 기초하여 가중치를 달리 부여하는 방식임. 이 방법은 과거 평균제곱오차가 작은 모형에 더 큰 가중치를 부여하고, 평균제곱오차가 큰 모형에는 상대적으로 작은 가중치를 부여함. 따라서 과거 평가기간에서 예측 성과가 우수했던 모형의 예측치를 최종 조합 예측치에 더 크게 반영할 수 있음.

○ Stock and Watson(1999)은 MSE에 기초한 조합 가중치 추정 방식을 제안하였음. i 번째 개별 예측모형의 과거 평가기간에 대한 평균제곱오차를 MSE_i 라고 하면, 해당 모형에 부여되는 가중치 w_i 는 식 (3-6)과 같이 정의할 수 있음.

- 식 (3-6)에서 분모는 예측조합에 포함되는 모든 모형의 MSE 역수를 합산한 값이며, 분자는 i 번째 모형의 MSE 역수를 의미함.
- 따라서 과거 예측오차가 클수록, 즉 MSE_i 가 클수록 해당 모형의 가중치는 작아지고, 반대로 MSE_i 작을수록 해당 모형의 가중치는 커짐.
- MSE 역수 기반 가중치를 이용한 조합 예측치는 식 (3-7)과 같이 산출됨.

$$(3-6) \quad w_i = \frac{MSE_i^{-1}}{\sum_{j=1}^N MSE_j^{-1}}$$

$$(3-7) \quad \hat{y}_t^{wmse} = \sum_{i=1}^N w_i \hat{y}_{i,t}$$

4. 시계열 분해-예측조합 기반 배추 가격 예측 알고리즘

- 본 연구는 앞서 살펴본 시계열 분해와 예측조합을 순차적으로 적용하여 배추 가격 예측 알고리즘을 구성하고, 이를 바탕으로 예측력을 평가함.
- 본 연구에서 제안한 알고리즘은 배추 가격 시계열을 계절 성분과 계절조정 성분으로 분리한 후, 각 성분의 특성에 적합한 예측모형을 적용하고, 최종적으로 복수의 개별 예측치를 조합하는 방식으로 구성됨. 즉, 반복적인 계절 변동은 별도로 분리하여 예측하고, 계절성이 제거된 가격 흐름에 대해서는 복수의 예측모형을 적용한 뒤, 이들 예측치를 조합함으로써 보다 안정적인 최종 예측치를 도출하고자 함.

- <그림 3-2>는 본 연구에서 제안하는 시계열 분해-예측조합 기반 배추 가격 예측 알고리즘의 전체 구조를 제시한 것으로, 세부 절차는 다음과 같음.

[Step 1] 시계열 분해

- 본 연구는 로그 변환된 배추 가격 시계열에 STL을 적용하여 추세, 계절, 잔차 성분으로 분해함.
- 분해 결과 중 월별 반복 패턴은 계절 성분으로 구분하고, 추세 성분과 잔차 성분은 합산하여 계절조정(seasonally adjusted) 성분으로 정의함. 이를 통해 배추 가격의 계절적 변동과 비계절적 변동을 구분한 후, 다음 단계에서 성분별 예측모형을 적용함.

[Step 2] 성분별 예측모형 적용

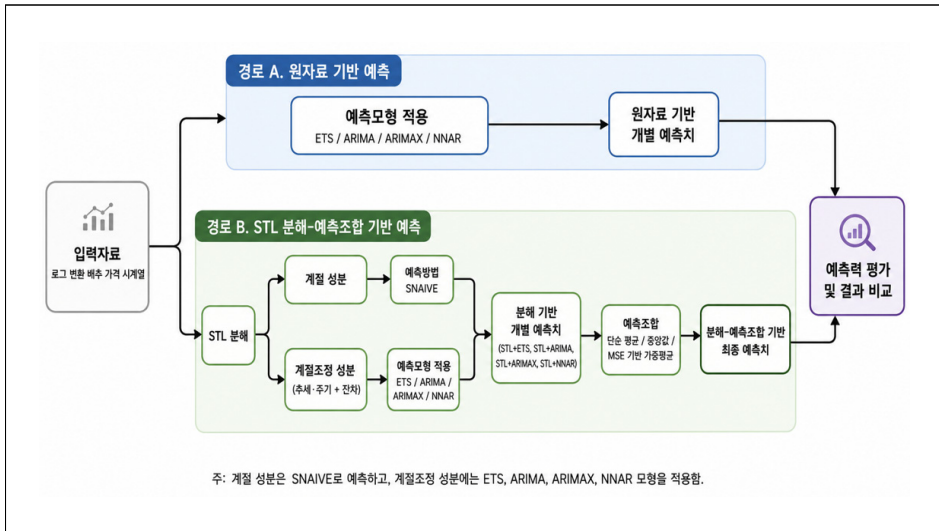
- 분해된 성분별 특성을 고려하여 각 성분에 적합한 예측 방법을 적용함. 먼저 계절 성분은 계절성 단순 모형(seasonal naive method: SNAIVE)을 이용하여 예측함.
 - SNAIVE는 예측 대상 시점과 동일한 계절에 해당하는 가장 최근의 관측값을 예측값으로 사용하는 방식임. 월별 자료의 경우 전년 동월의 값을 예측값으로 사용함. 따라서 STL 분해를 통해 도출된 계절 성분의 반복 패턴을 예측하는 데 적합함.
- 한편, 계절조정 성분에는 ETS 지수평활 모형, ARIMA 및 ARIMAX 모형, NNAR 모형을 적용하여 각각의 예측치를 산출함. 계절조정 성분은 수준 변화, 추세, 자기상관 구조, 외생변수 효과 및 비선형 변동을 보다 직접적으로 반영할 수 있음.

- 따라서 계절 성분의 예측치를 $\hat{S}_t$, i 번째 모형을 통해 산출한 계절조정 성분의 예측치를 $\hat{SA}_{i,t}$ 라고 하면, i 번째 분해 기반 개별 예측치는 $\hat{P}_{i,t} = \hat{S}_t + \hat{SA}_{i,t}$ 와 같이 정의할 수 있음.
- 즉, 계절 성분 예측치와 계절조정 성분 예측치를 결합함으로써 분해 기반 예측치를 산출함. 예를 들어 ETS 모형을 적용한 경우에는 STL+ETS, ARIMA 모형을 적용한 경우에는 STL+ARIMA의 형태로 복수의 분해 기반 개별 예측치를 구성할 수 있음.

[Step 3] 예측조합

- Step 2에서 산출된 복수의 분해 기반 개별 예측치는 예측조합 방법을 적용하여 최종 예측치로 결합함. 본 연구에서는 앞 절에서 제시한 단순 평균, 중앙값, MSE 기반 가중평균을 적용하여 조합 예측치를 산출하고, 각 조합 방식의 예측 성과를 비교함.
- 이와 함께 원자료 기반 예측모형에서 산출된 개별 예측치도 비교 대상에 포함함. 즉, 본 연구의 예측력 평가는 원자료 가격 시계열에 개별 예측모형을 직접 적용한 결과와 STL 분해 이후 각각의 성분을 예측한 뒤 이를 조합한 결과를 함께 비교하는 방식으로 이루어짐.

〈그림 3-2〉 시계열 분해-예측조합 기반 배추 가격 예측 알고리즘 구조



주: 본 그림의 도식화 과정에서는 생성형 AI 도구인 ChatGPT(OpenAI)를 보조적으로 활용하였으며, 최종 내용은 저자가 검토·수정하였음.

자료: 저자 작성.

4

예측 성과 평가 방법

1. 예측오차 기반 평가지표

- 배추 가격 예측모형의 예측 정확도는 실제값과 예측치 간의 차이인 예측오차 (forecast error)를 통해 측정할 수 있음. 실제값을 P_t , 예측값을 $\hat{P}_t$ 라고 하면, 예측오차는 $e_t = P_t - \hat{P}_t$ 와 같이 정의됨.
- 본 연구는 가격 예측 성과를 평가하기 위한 기본 지표로 평균절대오차(mean absolute error: MAE)와 평균제곱근오차(root mean squared error: RMSE)를 활용함. 평가기간의 관측치 수를 N 이라고 하면, MAE와 RMSE는 각각 식 (4-1), 식 (4-2)와 같이 정의됨.

$$(4-1) \text{ MAE} = \frac{1}{N} \sum_{i=1}^N |P_{t+h,i} - \hat{P}_{t+h,i}|$$

$$(4-2) \text{ RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{t+h,i} - \hat{P}_{t+h,i})^2}$$

- MAE는 절대오차의 평균으로, 예측치가 실제값에서 평균적으로 얼마나 벗어나는지를 나타냄. 배추 가격과 같이 동일 단위(원/10kg)에서 여러 예측모형을 비교할 경우 해석이 비교적 직관적이며, RMSE에 비해 극단적인 오차의 영향을 상대적으로 덜 받는다는 특징이 있음. 따라서 MAE는 평균적인 오차 수준을 파악하는 데 유용함.
- RMSE는 제곱오차의 평균에 제곱근을 취한 지표로, 큰 오차에 더 큰 가중을 부여함. 따라서 실제 가격이 급등하거나 급락하는 시점에서 예측모형이 큰 오차를 보일 경우, RMSE는 이를 민감하게 반영함. 이러한 점에서 RMSE는 가격 변동성이 확대되는 구간에서의 예측 실패 정도를 확인하는 데 유용함.
- 본 연구는 평균적인 오차 규모를 나타내는 MAE와 큰 오차에 대한 민감도를 반영하는 RMSE를 함께 제시하여 배추 가격 예측모형의 성과를 평가함.
- 한편, MAE와 RMSE는 모두 원 단위 오차를 나타내므로 해석이 직관적이라는 장점이 있으나, 계절성이 강한 시계열에서 단순 기준 예측과 비교한 상대적인 예측 성과를 직접 보여주지는 못함. 배추 도매가격은 월별 계절성과 작형별 수급 여건에 따라 가격 수준과 변동 폭이 달라질 수 있으므로, 예측모형이 단순한 계절 반복 규칙보다 얼마나 개선된 성과를 보이는지를 함께 검토할 필요가 있음.
- 이에 본 연구는 MAE, RMSE와 함께 계절성 평균절대스케일오차(seasonal mean absolute scaled error: sMASE)를 추가로 산출함. sMASE는 예측모형의 평균절대오차를 계절성 단순 예측(SNAIVE)의 평균절대오차로 나누어 산출하는 지표임. 따라서 sMASE는 평가 대상 모형이 단순한 계절 반복 예측에 비해 예측 성과가 어느 정도 우수한지를 나타내는 상대적 평가지표로 활용될 수 있음.

○ 월별 자료의 계절 주기를 $m=12$, 학습자료의 관측치 수를 T , 평가기간의 관측치 수를 N 이라고 하면, sMASE는 식 (4-3)과 같이 정의됨.

- 식 (4-3)의 분자는 평가 대상 예측모형의 평균절대오차를 의미함. 즉, 본 연구에서 적용한 원자료 기반 개별모형, STL 분해 기반 개별모형 또는 예측조합 모형의 예측오차 평균임. 분모는 학습자료에서 계절성 단순예측을 적용했을 때의 평균절대오차를 의미함.
- 따라서 sMASE가 1보다 작으면 해당 모형의 평균절대오차가 계절성 단순예측보다 작음을 의미하므로, 기준 모형보다 예측 성과가 우수하다고 해석할 수 있음. 반대로 sMASE가 1보다 크면 해당 모형의 평균절대오차가 계절성 단순예측보다 크다는 의미임.

$$(4-3) \text{ sMASE} = \frac{\frac{1}{N} \sum_{t=1}^N |Y_t - \hat{Y}_t|}{\frac{1}{N-1} \sum_{t=m+1}^T |Y_t - Y_{t-m}|}$$

2. 예측력 차이 검정: Modified Diebold and Mariano(MDM) 검정

○ 본 연구는 MAE, RMSE, sMASE의 예측오차 기반 평가지표를 이용하여 배추 가격 예측모형의 예측 성과를 평가함. 다만 이러한 지표는 모형별 예측오차의 크기를 비교하는 데 유용하나, 모형 간 예측력 차이가 통계적으로 유의한지 여부를 판단하기에는 한계가 있음. 이에 본 연구에서는 Diebold & Mariano(1995)의 DM 검정을 소표본에 적합하도록 수정한 Modified Diebold and Mariano(MDM) 검정을 추가로 수행함(Harvey et al., 1997).

○ MDM 검정은 두 예측모형의 손실함수 차이가 통계적으로 유의한지 검정함으로써, 모형 간 예측력 차이의 유의성을 평가하는 방법임. 본 연구는 원자료 기반 개별모형, STL 분해 기반 개별모형, 원자료-예측조합 모형, 시계열 분해-예측조합 알고리즘 간의 예측력 차이를 비교하는 데 MDM 검정을 활용함.

○ 두 예측모형을 각각 모형 1과 모형 2라고 하고, 시점 t 에서 두 모형의 예측오차를 각각 e_{1t} , e_{2t} 라고 하면, 두 모형의 예측력 비교는 예측오차에 대한 손실함수(loss function)에 기초함. 본 연구에서는 제곱오차(squared error: SE) 손실 $L(e_{it}) = e_{it}^2$ 과 절대오차(absolute error: AE) 손실 $L(e_{it}) = |e_{it}|$ 을 사용하였으며, 각 시점의 손실 차이 d_t 는 식 (4-4)와 같이 정의됨.

$$(4-4) \quad d_t = L(e_{1t}) - L(e_{2t})$$

○ DM 검정의 귀무가설은 두 모형의 기대 손실이 동일하다는 것으로, 식 (4-5)와 같이 표현됨.

$$(4-5) \quad H_0 : E(d_t) = L(e_{1t}) - L(e_{2t}) = 0$$

○ 식 (4-4)의 정의에 따르면, $E(d_t) > 0$ 인 경우 모형 2의 평균 손실이 모형 1보다 작다는 의미이고, $E(d_t) < 0$ 인 경우 모형 1의 평균 손실이 모형 2보다 작다는 의미임. 따라서 MDM 검정 결과를 해석할 때에는 비교 기준 모형과 비교 대상 모형의 설정 순서를 명확히 할 필요가 있음.

○ DM 검정 통계량은 식 (4-6)과 같이 정의됨.

- 여기서 $\bar{d}$ 는 두 모형 간 손실 차이 d_t 의 표본평균이며, $\widehat{Var}(\bar{d})$ 는 $\bar{d}$ 의 분산 추정치임. DM 검정 통계량은 귀무가설 하에서 점근적으로 표준정규분포를 따름.

$$(4-6) DM = \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}}$$

○ 그러나 예측평가 표본 수가 제한적인 경우 DM 검정은 소표본 편향의 영향을 받을 수 있음. Harvey et al.(1997)은 이를 보완하기 위해 DM 검정 통계량에 조정계수를 적용한 MDM 검정 통계량을 제안하였으며, 표본 수가 제한적인 경우 자유도 $n-1$ 의 Student's t 분포를 이용하여 검정하도록 제시함.

○ 예측평가 표본 수를 n , 예측 범위를 h 라고 할 때, MDM 검정 통계량은 식 (4-7)과 같이 표현됨.

$$(4-7) MDM = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n} \right]^{1/2} DM$$

○ 본 연구는 1개월 앞 예측을 대상으로 하므로 예측 범위는 $h=1$ 임. 또한 제공오차 손실과 절대오차 손실을 각각 적용하여 MDM 검정을 수행하고, 손실함수 선택에 따라 모형 간 예측력 차이가 일관되게 나타나는지를 검토함.

3. 예측 포괄성 검정: Encompassing 검정

○ 본 연구는 KREI 농업관측센터 전망치와 제안 알고리즘 예측치가 서로 보완적인 예측 정보를 포함하는지 검토하기 위해 Encompassing 검정을 수행함. 이 검정은 한 예측치가 다른 예측치에 포함되지 않은 추가적인 정보를 제공하는지를 평가하는 방법임.

- KREI 전망치는 현장조사, 수급 동향, 산지 및 시장 정보, 담당자 판단 등을

종합적으로 반영하는 반면, 본 연구의 제안 알고리즘 예측치는 과거 가격 및 정량 자료에 기반하여 산출됨. 따라서 두 예측치가 서로 다른 정보를 반영할 경우, 이를 결합하여 활용함으로써 예측 성과가 개선될 가능성이 있음.

○ Harvey et al.(1998)은 두 예측치 간 정보 포괄 여부를 검정하기 위한 Encompassing 검정을 제시하였으며, 이석일 외(2015)는 이를 육계가격 예측모형의 예측력 비교 연구에 적용하여 KREI 전망치와 시계열 모형 예측치의 결합 가능성을 검토하였음. 본 연구에서도 이와 유사하게 KREI 배추 가격 전망치와 시계열 분해-예측조합 알고리즘 예측치 간의 보완 가능성을 검토함.

○ 본 연구는 KREI 농업관측센터 전망치를 기준 예측치로, 제안 알고리즘 예측치를 비교 예측치로 설정하여 Encompassing 검정을 수행함. 시점 t 의 KREI 전망치의 예측오차를 e_{1t}^{KREI} , 제안 알고리즘의 예측오차를 e_{2t}^{ALG} 라고 하면, Encompassing 검정식은 식 (4-8)과 같이 표현할 수 있음.

- Encompassing 검정은 KREI 전망치의 예측오차가 두 예측치의 오차 차이에 의해 설명되는지를 검토하는 방식으로 수행할 수 있음.
- 식 (4-8)에서 $\lambda = 0$ 이라는 귀무가설이 기각되지 않으면, KREI 전망치가 제안 알고리즘 예측치에 포함된 예측 정보를 포괄하는 것으로 해석할 수 있음.
- 반대로 λ 가 통계적으로 유의하게 나타나면, 제안 알고리즘 예측치가 KREI 전망치에 포함되지 않은 추가적인 예측 정보를 제공할 가능성이 있으며, 이는 두 예측치를 결합하여 활용할 수 있는 근거로 해석될 수 있음.

$$(4-8) \quad e_{1t} = \alpha + \lambda(e_{1t} - e_{2t}) + \epsilon_t$$

○ Encompassing 검정 결과 두 예측치의 조합 가능성이 확인되는 경우, KREI 전망치와 제안 알고리즘 예측치는 식 (4-9)와 같은 선형 결합 형태로 활용할 수 있음.

- 식 (4-9)에서 f_t^C 는 조합 예측치, f_t^{KREI} 는 KREI 전망치, f_t^{ALG} 는 본 연구의 제안 알고리즘 예측치를 의미함. $1 - \lambda$ 는 KREI 전망치에 부여되는 가중치이며, λ 는 제안 알고리즘 예측치에 부여되는 가중치임.

$$(4-9) \quad f_t^C = (1 - \lambda)f_t^{KREI} + \lambda f_t^{ALG}$$

- 본 연구는 Encompassing 검정 결과를 통해 KREI 전망치와 제안 알고리즘 예측 간의 정보 보완성을 평가하고, 두 예측치를 결합한 조합 전망의 실무적 활용 가능성을 검토함.

5

배추 가격 예측모형 추정 및 예측력 평가 결과

1. 분석 자료

1.1. 기초 통계량

○ <표 5-1>은 본 연구의 분석에 활용된 배추 월별 도매가격의 기초 통계량을 제시한 것임. 분석자료는 2001년 1월부터 2025년 12월까지의 월별 배추 도매가격이며, 전체 관측치 수는 300개임.

- 원자료 기준 배추 도매가격의 평균은 6,658원/10kg, 최솟값은 1,073원, 최댓값은 24,874원, 표준편차는 4,028원으로 나타났음. 또한, 왜도는 1.32, 첨도는 5.54로 분포가 우측으로 치우쳐 있으며, 정규분포보다 뾰족한 형태를 보임. Jarque-Bera 검정 결과에서도 정규성 가정이 기각되어($p=0.000$), 원자료의 분포는 정규분포와 차이가 있는 것으로 확인됨.
- 반면, 로그 변환 후에는 왜도가 -0.24로 나타나 분포의 비대칭성이 완화되었고, 첨도는 2.54로 낮아져 극단값의 영향도 감소한 것으로 나타남. 또한,

Jarque-Bera 검정 결과도 통계적으로 유의하지 않게 나타나($p=0.064$), 로그 변환 자료가 원자료에 비해 정규분포에 보다 가까운 특성을 보이는 것으로 해석할 수 있음.

- 이러한 결과는 배추 도매가격에 로그 변환을 적용하는 것이 가격 시계열의 분포 왜곡과 극단값(outlier)의 영향을 완화하는 데 유효함을 시사함.
- <그림 5-1>에서도 로그 변환 이후 분포가 원자료보다 더 대칭적인 분포를 보이는 점을 시각적으로 확인할 수 있음.

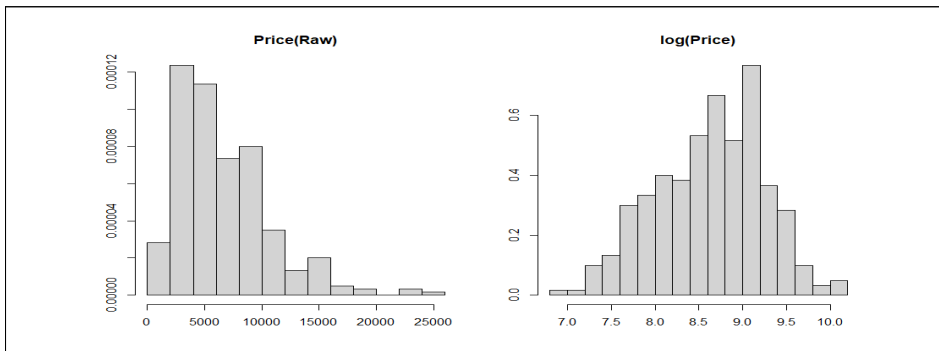
〈표 5-1〉 배추 월별 도매가격의 기초 통계량

구분	평균	중앙값	최대	최소	표준편차	왜도	첨도	Jarque-Bera 통계치
원자료	6,658.0	5,755.0	24,874.3	1,073.0	4,028.2	1.32	5.54	167.6 (0.000)
로그 변환	8.62	8.66	10.12	6.98	0.62	-0.24	2.54	5.5 (0.064)

주: ()는 Jarque-Bera 검정에 대한 p-value임.

자료: 서울특별시농수산물공사 유통정보 홈페이지(검색일: 2026. 2. 5.).

〈그림 5-1〉 배추 도매가격의 분포

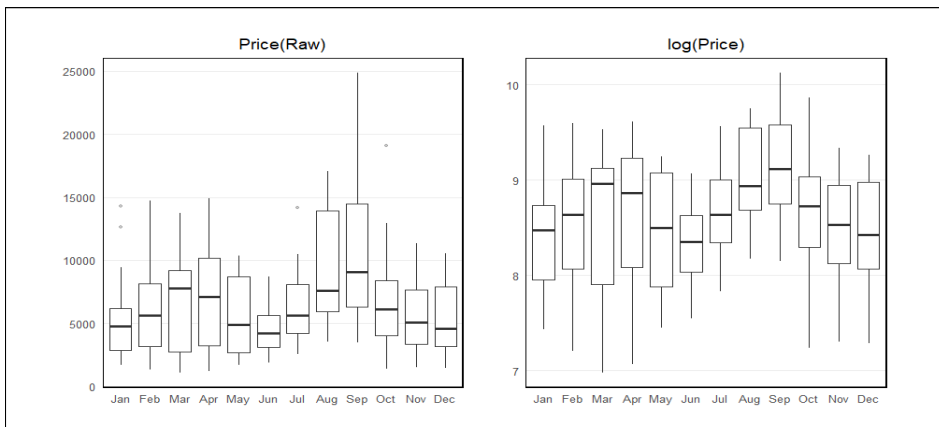


자료: 서울특별시농수산물공사 유통정보 홈페이지(검색일: 2026. 2. 5.).

○ <그림 5-2>는 배추 도매가격의 원자료와 로그 변환 자료의 월별 분포를 제시한 것임.

- 원자료 기준 월별 중앙값과 사분위 범위가 뚜렷하게 달라 배추 가격에 계절적 변동이 존재함을 확인할 수 있음.
- 특히 8~9월의 여름 배추 출하기에는 가격 수준이 상대적으로 높고 변동폭도 확대되는 경향을 보임. 반면 5~6월 봄배추 출하기에는 가격 수준이 상대적으로 낮고 변동폭도 비교적 안정적으로 나타남. 또한 일부 월에서는 극단값이 관측되어 특정 시기에 가격 급등이 발생할 수 있음을 보여줌.
- 로그 변환 이후에는 이상치의 영향이 완화되면서 월별 가격 분포가 원자료보다 안정적으로 나타남. 즉, 로그 변환 자료는 원자료에서 관측되는 큰 폭의 가격 변동을 완화하면서도 월별 계절적 패턴은 유지하는 것으로 볼 수 있음.

〈그림 5-2〉 배추의 월별 도매가격(원/10kg) 분포



자료: 서울특별시농수산물공사 유통정보 홈페이지(검색일: 2026. 2. 5.).

○ 따라서 본 연구는 가격 시계열의 분산 안정화와 분포 왜곡 완화를 위해 모든 예측모형에 로그 변환된 배추 도매가격을 공통적으로 적용함.²⁾

2) 인공신경망 모형에 포함된 연속형 외생변수는 학습 표본의 평균과 표준편차를 이용하여 표준화한 후 적용함. 다만 월별 더미변수와 같이 범주형 정보를 나타내는 변수는 표준화하지 않음.

- 또한 로그 변환을 통해 STL 분해가 전제하는 가법적 구조에 보다 부합하는 형태로 자료를 변환함.

1.2. 변수 정의 및 구성

○ 본 연구에서 사용한 주요 변수의 정의와 산식은 <표 5-2>에 제시하였음. 예측 대상 변수는 t 월의 배추 월별 도매가격 P_t 이며, 모형 추정 과정에서는 가격 수준에 따른 변동성 확대를 완화하기 위해 로그 변환 가격 $p_t = \log(P_t)$ 를 사용하였음.

- 원자료 기반 모형은 로그 가격 p_t 를 직접 예측대상 변수로 설정하였음. 반면 분해 기반 모형은 STL 분해를 통해 도출된 계절 성분과 계절조정 성분으로 분리한 후, 각 예측 대상을 성분으로 설정하였음. 계절조정 성분은 로그 가격에서 계절 성분을 제거한 값으로 정의되며, 이를 통해 월별 반복 패턴을 나타내는 계절 성분과 계절성을 제외한 가격 흐름을 구분하여 예측하고자 하였음.

○ 설명변수는 가격 기반 파생 변수와 수급 관련 변수로 구성하였음.

○ 가격 기반 파생 변수는 배추 가격의 계절성, 최근 가격 흐름, 작형별 가격 수준 차이를 반영하기 위해 설정하였음.

- 배추는 월별 계절성이 뚜렷하고, 겨울·봄·여름·가을 작형에 따라 주산지, 출하 시기, 저장 가능성, 기상 리스크가 다르게 나타남. 따라서 동일한 가격 수준이라도 해당 가격이 어느 시기 또는 어느 작형에서 관측되었는지에 따라 수급상 의미가 달라질 수 있음.

- 이에 본 연구는 단순한 전월 가격 수준만을 활용하기보다, 전년 동월 대비 가격 변화, 최근 3개월 평균 가격의 전년 동월 기간 대비 차이, 동일 작형의 과거 평균 가격과의 차이를 함께 고려하였음.
- 전년 동월 대비 로그 가격 변화는 동일 계절 조건에서 최근 가격 수준이 전년보다 높은지 또는 낮은지를 반영하기 위한 변수임.
- 최근 3개월 평균 가격의 전년 동월 기간 대비 차이는 특정 월의 일시적 가격 변동보다 최근 가격 흐름의 지속성을 반영하기 위해 사용하였음.
- 동일 작형 평균 대비 가격 수준은 예측 대상 월이 속한 작형의 통상적인 가격 수준과 비교하여 현재 가격이 상대적으로 높은지 또는 낮은지를 파악하기 위한 변수임.

○ 수급 관련 변수로는 배추 반입량과 김치 수입량의 시차변수를 사용하였음. 반입량은 도매시장에 실제로 공급되는 물량을 나타내는 단기 공급지표이며, 김치 수입량은 국내 배추 및 김치 수급 여건과 관련된 외부 수급지표로 볼 수 있음. 본 연구는 이들 변수를 1~3개월 시차변수로 구성하여 실제 예측 시점에서 활용 가능한 수급 정보를 반영하였음.

○ 본 연구는 설명변수 구성 과정에서 예측 대상 월의 실제 가격이나 사후적으로 확인되는 정보의 사용을 배제하고자 하였음. 즉, 예측 대상 월 직전까지 관측 가능한 가격, 반입량, 수입량 자료만을 이용하여 변수를 생성함으로써 정보 누출 가능성을 최소화하고, 예측모형의 실무 적용 가능성을 높이하고자 하였음.

〈표 5-2〉 배추 가격 예측모형의 주요 변수 정의 및 산식

구분		변수		산식
예측 대상 변수	원자료 기반	P_t	t 월의 배추 도매가격	P_t (t 는 예측 대상 월)
		p_t	로그 변환한 배추 도매가격	$p_t = \log(P_t)$
예측 대상 성분	분해 기반	S_t	STL 분해로 추출한 계절 성분	$p_t = T_t + S_t + R_t$ (T_t, S_t, R_t 는 각각 STL 분해의 추세 성분, 계절 성분, 잔차 성분)
		SA_t	STL 분해 후 계절 성분을 제거한 계절조정 성분	$SA_t = p_t - S_t$
설명 변수 (예측 변수)	가격 기반 변수	$\Delta^{(12)}p_{t-1}$	예측 대상 월의 전월 기준 전년 동월 대비 로그 가격 변화	$\Delta^{(12)}p_{t-1} = p_{t-1} - p_{t-13}$
		$G_{t-1}^{MA_3}$	예측 대상 월의 전월까지 최근 3개월 평균 로그 가격과 전년 동일 기간 평균 간 차이	$G_{t-1}^{MA_3} = MA_3(p)_{t-1} - MA_3(p)_{t-13}$ ($MA_3(p)_t$ 는 t 월까지의 최근 3개월 로그 가격 이동평균)
		C_{t-1}	예측 대상 월의 전월 로그 가격과 동일 작형 과거 평균 로그 가격 차이	$C_{t-1} = p_{t-1} - \bar{p}_{c,t-1}$ ($\bar{p}_{c,t-1}$ 는 예측 대상 월이 속한 작형 c 와 동일한 작형에 해당하는 과거 월별 로그 가격의 평균)
	반입량	q_{t-k}	배추 반입량의 로그 변환 시차 변수	$q_{t-k} = \log(Q_{t-k}), k = 1, 2, 3$
	수입량	m_{t-k}	김치 수입량의 로그 변환 시차 변수	$m_{t-k} = \log(M_{t-k}), k = 1, 2, 3$

자료: 저자 작성.

1.3. 학습자료와 평가자료 구간의 구분

○ 본 연구는 시계열 분해-예측조합 기반 배추 가격 예측 알고리즘의 예측 성과를 평가하기 위해 2001년 1월~2025년 12월의 월별 자료($n=300$)를 표본 내 모형 설정 및 추정(in-sample specification and estimation) 구간, 검증

(validation) 구간, 표본 외 평가(out-of-sample evaluation) 구간으로 구분하여 활용함.

○ 표본 내 모형 설정 및 추정 구간은 2001년 1월~2018년 12월(n=216)로 설정하였으며, 이 구간은 STL 분해 구조와 개별 예측모형의 기본 구조를 설정하는 데 활용함.

- 먼저 로그 변환한 가격 시계열에 STL 분해를 적용하여 계절 성분과 계절조정 성분을 산출하고, 이후 각 성분에 적용할 예측모형의 기본 구조를 설정함.
- STL 분해 파라미터는 계절 평활화 방식, 추세 평활화 구간, 저역통과 필터 구간 및 반복 횟수 등 주요 설정값을 검토한 뒤, 분해 잔차의 진단 결과, 계절 성분의 반복 안정성, 분해 결과의 해석 가능성 등을 종합적으로 고려하여 확정함. 개별 예측모형은 ETS 지수평활 모형, ARIMA 및 ARIMAX 모형, NNAR 모형을 대상으로 설정하였으며, 각 모형의 기본 구조는 표본 내 자료 특성과 모형 진단 결과를 바탕으로 결정함.
- ARIMA 계열 모형은 비계절 및 계절 차수의 후보 범위를 검토하고, ETS 지수평활 모형은 오차, 추세, 계절 성분의 조합 형태를 검토함. NNAR 모형은 비계절 및 계절 입력 시차, 은닉노드 수 등 기본 구조를 설정함. 외생변수 포함 모형에서는 외생변수의 종류, 시차 길이, 입력 조합을 사전에 정의함.

○ 검증 구간은 2019년 1월~2020년 12월(n=24)로 설정하였음. 이 구간은 표본 내 구간에서 확정한 STL 분해 방식과 개별 예측모형의 구조를 유지한 상태에서 예측 성과를 점검하는 데 활용함. 특히 본 연구에서는 검증 구간에서 산출된 모형별 예측오차를 이용하여 MSE 기반 가중평균 조합의 가중치를 추정함.

- 검증 구간의 예측치는 고정 길이 이동창(fixed-length rolling window) 방식으로 산출함. 즉, 각 예측 시점마다 동일한 길이의 학습자료를 사용하

되, 학습창을 1개월씩 이동하면서 모형을 재추정하고 1개월 앞 예측치 (1-step-ahead forecast)를 도출함.

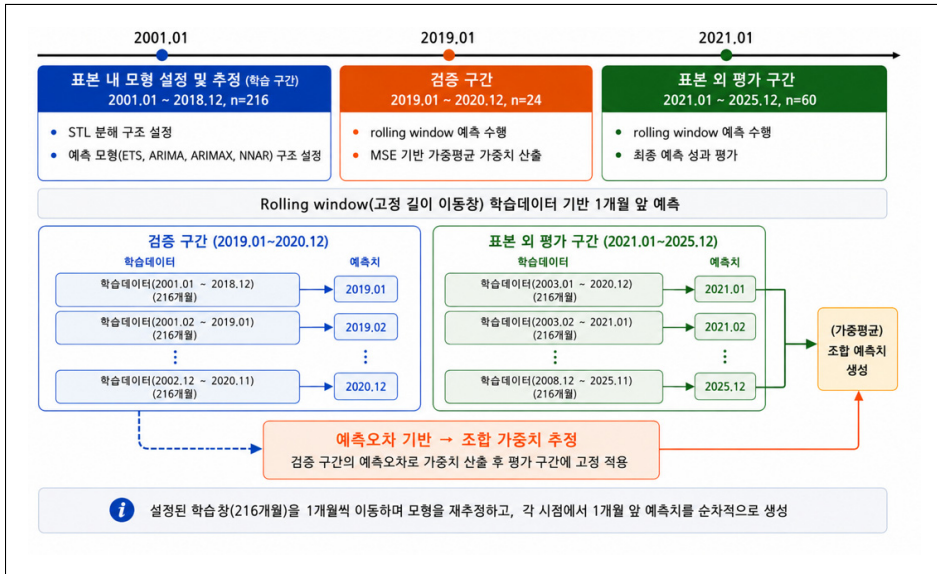
- 예를 들어 2019년 1월 예측에는 2001년 1월~2018년 12월의 216개월 자료를 사용하고, 2019년 2월 예측에는 학습창을 1개월 이동하여 2001년 2월~2019년 1월의 자료를 사용함. 이와 같은 방식으로 2020년 12월까지 총 24개의 1개월 앞 예측치를 산출함.
- 단순 평균과 중앙값 조합은 사전적인 가중치 추정 과정이 필요하지 않으므로, 검증 구간의 예측오차는 주로 MSE 기반 가중평균 조합의 모형별 가중치를 산출하는 데 활용함. 이를 통해 표본 외 평가 구간에 적용할 조합 가중치를 사전에 확정함으로써 평가기간의 실제값이 가중치 추정에 사용되는 것을 방지함.

○ 표본 외 평가 구간은 2021년 1월~2025년 12월($n=60$)로 설정하였음. 이 구간은 앞선 두 구간에서 확정한 분해 구조, 개별 예측모형의 구조 및 예측조합 가중치를 바탕으로 최종 예측 성과를 평가하는 데 활용함.

- 표본 외 평가 구간에서도 검증 구간과 동일하게 고정 길이 이동창 방식을 적용하여 각 예측 시점의 학습자료에 개별 예측모형을 재적합하고, 1개월 앞 예측치를 순차적으로 산출함.
- 이후 각 시점에서 산출된 개별 예측치에 단순 평균, 중앙값, MSE 기반 가중평균 조합 방식을 적용하여 조합 예측치를 생성함.
- 마지막으로 표본 외 평가 구간의 실제값과 개별모형 예측치 및 조합 예측치 간의 예측오차를 바탕으로 모형별 예측 성과를 비교·평가함.
- 이와 같은 구간 구분은 모형 구조 설정, 조합 가중치 산출, 최종 예측력 평가의 역할을 명확히 분리함으로써 예측 성과 평가의 객관성과 실무 적용 가능성을 높이기 위한 것임.

○ <그림 5-3>은 본 연구에서 설정한 학습자료와 평가자료 구간의 구분을 제시한 것임.

<그림 5-3> 학습자료와 평가자료 구간의 구분



주: 본 그림의 도식화 과정에서 생성형 AI 도구인 ChatGPT(OpenAI)를 보조적으로 활용하였으며, 최종 내용은 저자가 검토수정하였음.
자료: 저자 작성.

2. STL 시계열 분해 결과

○ 본 연구는 배추 도매가격 시계열의 계절성과 비계절적 변동을 구분하기 위해 STL 분해를 적용함. STL 분해에는 계절 주기, 추세 성분의 평활화 구간, 계절 성분의 평활화 방식, 저역통과 필터의 평활화 구간, 내부 반복 횟수, 외부 반복 횟수 등이 주요 설정값으로 활용됨.

- STL 분해 파라미터는 표본 내 모형 설정 및 추정 구간(2001년 1월~2018년

12월)의 로그 가격 시계열을 대상으로 여러 후보 조합을 적용한 후 최종 설정값을 확정하였음.

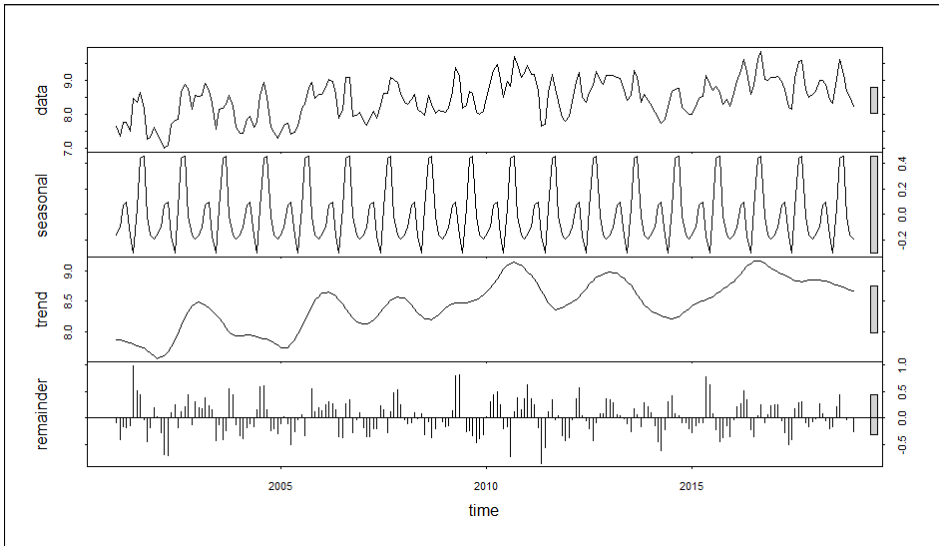
- 본 연구는 월별 가격 자료의 특성을 고려하여 계절 주기는 12개월로 설정하였음. 또한 계절 성분의 평활화 방식은 고정된 월별 계절 패턴을 반복적으로 적용하는 “periodic”으로 설정하였으며, 추세 성분의 평활화 구간은 19, 저역통과 필터의 평활화 구간은 13으로 설정함. 내부 반복 횟수는 2, 외부 반복 횟수는 0으로 설정함.

○ <그림 5-4>는 로그 변환한 배추 도매가격 시계열에 STL 분해를 적용한 결과를 제시한 것임. 분해 결과, 원 시계열에서는 시기별 가격 수준 변화와 반복적인 계절 패턴이 함께 관측되었음. 이는 배추 도매가격이 단순한 장기 흐름뿐만 아니라 작형 및 출하 시기와 연계된 계절적 변동을 동시에 포함하고 있음을 보여줌.

- 계절 성분(seasonal)은 월별로 유사한 형태가 반복되는 패턴을 보였으며, 이를 통해 배추 도매가격 시계열에 내재된 월별 계절성이 비교적 뚜렷함을 확인할 수 있음. 특히 본 연구는 계절 성분의 평활화 방식을 “periodic”으로 설정하였기 때문에, 계절 효과가 시간에 따라 크게 변화하기보다는 일정한 월별 패턴이 반복되는 형태로 추정됨. 이는 배추 가격의 구조적 월별 계절성을 안정적으로 반영하기 위한 설정임.
- 추세 성분(trend)은 표본기간 동안 완만한 상승과 하락을 반복하는 형태를 보였으며, 일부 시점에서는 상대적으로 높은 수준이 관측됨. 이는 배추 도매가격에 단기적인 계절 변동뿐만 아니라 중장기적 가격 수준 변화도 존재함을 의미함.
- 잔차 성분(remainder)은 전반적으로 0을 중심으로 분포하였으나, 일부 시점에서는 비교적 큰 양(+) 또는 음(-)의 변동이 나타남. 이는 추세 및 계

질 성분으로 설명되지 않는 일시적 가격 충격이나 불규칙 변동이 존재함을 시사함. 이러한 잔차 변동은 기상 여건, 작황 변화 등 단기 수급 충격 요인과 관련될 가능성이 있음.

〈그림 5-4〉 배추 가격 시계열의 STL 분해 결과



주: 로그 변환한 월별 배추 도매가격 시계열에 STL 분해를 적용한 결과임. 계절 주기는 12개월, 계절 성분은 고정된 월별 패턴이 반복되는 방식 (s.window=periodic)으로 설정하였으며, 추세 평활화 구간 (t.window)은 19, 저역통과 필터 평활화 구간(l.window)은 13, 내부 반복(inner)은 2, 외부 반복(outer)은 0을 적용함.

자료: 저자 작성.

○ 이러한 분해 결과를 바탕으로 본 연구는 STL을 통해 추출된 계절 성분과 계절 조정 성분을 구분하여 예측모형을 적용함.

- 계절 성분에는 월별 반복 패턴을 반영할 수 있는 계절성 단순 기법 (SNAIVE)을 적용함. 반면 계절조정 성분에는 ETS 지수평활 모형, ARIMA 및 ARIMAX 모형, NNAR 모형을 적용하여 예측치를 산출함.
- 이후 계절 성분 예측치와 계절조정 성분 예측치를 결합하여 분해 기반 개별 예측치를 산출함.

3. 배추 가격 예측모형 추정 결과

○ <표 5-3>은 배추 가격 예측에 적용한 개별모형의 구조와 주요 설정을 제시한 것임. 각 모형의 기본 구조는 표본 내 모형 설정 및 추정 구간인 2001년 1월부터 2018년 12월 자료를 기준으로 선정하였음. 이후 rolling window 예측에서는 각 예측 시점마다 학습자료를 갱신하여 모형을 재추정하되, <표 5-3>에서 제시한 ETS 구성 형태, ARIMA 차수, NNAR 입력 시차 및 은닉노드 수 등 기본 구조는 동일하게 유지함.

- 이러한 방식은 예측 시점마다 모형 구조를 새롭게 선택할 경우 발생할 수 있는 구조 변동을 줄이고, 모형별 예측 성과를 동일한 기준에서 비교하기 위한 것임.
- 즉, 본 연구에서는 표본 내 구간에서 확정한 모형 구조를 기준으로 하되, rolling window별로 새롭게 추가되는 자료를 반영하여 모형의 계수, 상태값 및 신경망 가중치를 재추정하고 1개월 앞 예측치를 산출함.

○ 원자료 기반 모형은 로그 변환한 가격 시계열에 예측모형을 직접 적용하는 방식으로 구성하였으며, ETS 지수평활 모형, ARIMAX 모형, 외생변수를 포함하는 NNARX 모형을 포함함.

- ETS(A,N,A)는 덧셈 오차(additive error), 추세 없음(no trend), 덧셈 계절성(additive seasonality)을 가정한 지수평활 모형임. 이는 로그 가격 시계열에 추세를 포함하지 않고, 월별 계절성을 덧셈 형태로 반영하는 구조임.
- ARIMAX(1,1,2)(2,0,0)[12]는 비계절 차분을 적용한 계절 ARIMA 구조에 외생변수를 추가한 모형임. 외생변수로는 배추 반입량과 김치 수입량의 시차 변수를 포함하였음.

- NNARX(3,1,2)[12]는 가격 시계열의 비계절 시차와 계절 시차를 입력으로 사용하고, 은닉층 노드 2개를 통해 비선형 관계를 학습하는 신경망 자기회귀 모형임. 여기에 가격 기반 특성 변수를 추가 입력변수로 포함함.

○ 시계열 분해 기반 모형은 STL 분해를 통해 로그 가격 시계열을 계절 성분과 계절조정 성분으로 분리한 뒤, 각 성분의 특성에 맞게 예측모형을 적용하는 방식임. 계절 성분은 계절성 단순 기법(SNAIVE)을 이용하여 예측하였고, 계절조정 성분에는 ETS, ARIMA, ARIMAX, NNAR 및 NNARX 모형을 각각 적용함.

- STL-ETS(A,N,N)는 계절조정 성분에 ETS 모형을 적용한 결과, 덧셈 오차를 가정하되 추세 및 계절성을 별도로 포함하지 않는 구조가 선택된 것임. 이는 STL 분해 과정에서 계절 성분이 이미 제거되었기 때문에 계절조정 성분에 대해서는 별도의 계절항이 필요하지 않았음을 의미함.
- STL-ARIMA 및 STL-ARIMAX는 계절조정 성분의 자기상관 구조를 반영하도록 설정된 모형임. 특히 STL-ARIMAX는 ARIMA 구조에 배추 반입량과 김치 수입량의 시차 변수를 외생변수로 포함한 형태임.
- STL-NNAR 및 STL-NNARX는 계절조정 성분에 내재될 수 있는 비선형 자기회귀 패턴을 반영하기 위한 모형이며, STL-NNARX는 가격 기반 특성 변수를 추가 입력변수로 활용한 구조임.

〈표 5-3〉 배추 가격 예측모형의 구조 및 주요 설정

구분	모형	구조(structure) 및 주요 설정(specifications)	
원자료 기반	ETS 지수평활	① ETS(A,N,A)	• 덧셈 오차(A), 추세 없음(N), 덧셈 계절성(A)
	ARIMA	② ARIMAX(1,1,2)(2,0,0)[12]	• 계절성 ARIMA(p,d,q)(P,D,Q)[m] • 외생변수: 반입량·김치 수입량 시차 변수
	인공 신경망	③ NNARX(3,1,2)[12]	• 계절성 NNAR(p,P,k)[m] • 비 계절 시차(p): AR(p)로 최적 시차 결정(AIC 기준) • 계절 시차(P): 기본값 1 • 은닉층의 노드 수(size): 2 • 외생변수: 가격 기반 특성 변수
분해 기반	ETS 지수평활	④ STL-ETS(A,N,N)	• 덧셈 오차(A), 추세·계절성 없음(N)
	ARIMA	⑤ STL+ARIMA(1,1,2)(2,0,1)[12]	• 계절성 ARIMA(p,d,q)(P,D,Q)[m]
		⑥ STL+ARIMAX(0,1,5)(2,0,0)[12]	• 계절성 ARIMA(p,d,q)(P,D,Q)[m] • 외생변수: 반입량·김치 수입량 시차 변수
	인공 신경망	⑦ STL+NNAR(3,1,2)[12]	• 계절성 NNAR(p,P,k)[m]
		⑧ STL+NNARX(2,1,2)[12]	• 계절성 NNAR(p,P,k)[m] • 외생변수: 가격 기반 특성 변수

자료: 저자 작성.

4. 예측조합을 위한 가중치 적용 결과

○ 〈표 5-4〉는 〈표 5-3〉에 제시한 시계열 분해 기반 개별 예측모형의 예측치를 조합하기 위해 적용한 조합 가중치 및 선택 비율을 제시한 것임. 본 연구는 분해 기반 개별 예측치에 대해 단순 평균, 중앙값, MSE 기반 가중평균 조합을 적용하였음.

○ MSE 기반 조합 가중치는 검증 구간인 2019년 1월~2020년 12월(n=24)의 예측오차를 이용하여 산출하였음. 구체적으로 분해 기반 개별 예측모형의 예측치와 실제값을 비교하여 평균제곱오차(MSE)를 계산하고, MSE가 작은 모형일수록 상대적으로 큰 가중치가 부여되도록 설정함. 이후 검증 구간에서 산

출된 MSE 기반 가중치는 표본 외 평가 구간인 2021년 1월~2025년 12월의 개별 예측치에 고정 적용하여 조합 예측치를 산출함.

○ MSE 역수 기반 가중치 산출 결과, STL+ARIMAX 모형의 가중치가 0.212로 가장 높게 나타났음. 이는 검증 구간에서 STL+ARIMAX 모형의 예측오차가 상대적으로 작았음을 의미함.

- 다만 모형 간 가중치 차이는 전반적으로 크지 않았음. 이는 검증 구간에서 분해 기반 개별 예측모형 간 예측오차 차이가 크지 않았으며, 조합 대상 모형들의 예측 성과가 비교적 유사하였음을 의미함.

○ 단순 평균 조합은 별도의 성과 기반 가중치 추정 과정 없이 5개의 개별 예측모형에 동일한 가중치를 부여하는 방식임. 따라서 단순 평균 조합에서는 모든 개별 예측모형에 각각 1/5, 즉 0.200의 가중치가 부여됨.

○ 중앙값 조합은 각 예측 시점에서 5개의 개별 예측치를 크기순으로 정렬한 뒤, 중앙에 위치한 예측치를 조합 예측치로 사용하는 방식임. 따라서 중앙값 조합은 단순 평균이나 MSE 기반 가중평균처럼 사전에 고정된 가중치를 부여하는 방식이 아님. <표 5-4>의 중앙값 열은 표본 외 평가 구간에서 각 모형의 예측치가 중앙값으로 선택된 비율을 나타낸 것임.

〈표 5-4〉 분해 기반 개별 예측모형의 조합 방식별 가중치 및 선택 비율

모형		구조 및 주요 설정	단순 평균 가중치	중앙값 선택 비율	MSE 기반 가중치
분해 기반	ETS 지수평활	④ STL-ETS(A,N,N)	0.200	0.125	0.209
	ARIMA	⑤ STL+ARIMA(1,1,2)(2,0,1)[12]	0.200	0.292	0.165
		⑥ STL+ARIMAX(0,1,5)(2,0,0)[12]	0.200	0.208	0.212
	인공 신경망	⑦ STL+NNAR(3,1,2)[12]	0.200	0.167	0.206
		⑧ STL+NNARX(2,1,2)[12]	0.200	0.208	0.208

자료: 저자 작성.

5. 가격 예측 정확도 평가 결과

○ <표 5-5>는 표본 외 평가 구간(2021년 1월~2025년 12월, n=60)을 대상으로 원자료 기반 개별 예측모형, STL 분해 기반 개별 예측모형, 원자료-예측조합 모형, 분해-예측조합 모형의 예측 정확도를 비교한 결과를 제시한 것임.

- 예측 정확도는 RMSE, MAE, sMASE를 기준으로 평가하였으며, 각 지표는 값이 작을수록 예측오차가 작음을 의미함.
- 본 연구의 예측모형은 로그 변환 가격을 대상으로 추정되었으나, 예측 정확도 평가는 원 단위 가격 기준으로 수행함.³⁾

○ 분석 결과, 표본 외 평가 구간에서 본 연구의 제안 알고리즘인 분해-예측조합 모형은 원자료 기반 개별모형 및 분해 기반 개별모형에 비해 전반적으로 낮은 예측오차를 보였음.

- RMSE 기준으로는 분해-예측조합 평균값 모형의 RMSE가 2,457.1로 가장 낮게 나타났으며, 분해-예측조합 예측성과(MSE) 모형의 RMSE도 2,458.9로 유사한 수준을 보였음.
- MAE 기준으로는 분해-예측조합 예측성과(MSE) 모형이 1,856.0으로 낮은 오차를 보였으며, 분해-예측조합 평균값과 중앙값 모형도 각각 1,858.7, 1,858.9로 유사한 수준의 예측 성과를 나타냄.
- sMASE 기준에서도 분해-예측조합 평균값 및 예측성과(MSE) 모형이 각각 0.626으로 낮게 나타나, 계절성 단순 예측을 기준으로 하였을 때도 상대적

³⁾ 로그 예측치 $\hat{p}_t$ 를 원 단위 가격으로 환산할 때 단순 역변환 $\exp(\hat{p}_t)$ 을 적용하면 과소 추정될 가능성이 있음. 이에 본 연구에서는 로그 예측오차의 분산 $\hat{\sigma}^2$ 을 반영한 보정 역변환을 적용하였으며, 원 단위 예측치는 $\hat{p}_t = \exp(\hat{p}_t + \hat{\sigma}^2/2)$ 로 산출하였음.

으로 우수한 예측 성과를 보였음.

○ 개별 예측모형 간 비교에서는 STL 분해 기반 모형이 원자료 기반 모형보다 대체로 낮은 예측오차를 보였음.

- ARIMA 계열의 경우 원자료 기반 ARIMAX 모형의 RMSE는 2,882.4인 반면, STL-ARIMA, STL-ARIMAX 모형은 각각 2,548.6, 2,730.8로 나타나 STL 분해 적용에 따른 예측오차 감소가 확인됨.
- 인공신경망 계열에서도 원자료 기반 NNARX 모형의 RMSE는 2,989.7로 높게 나타난 반면, STL+NNARX 모형은 2,504.7로 크게 낮아졌음.
- 다만, 모든 모형군에서 STL 분해 적용에 따른 개선 효과가 일관되게 나타나지는 않음. ETS 계열에서는 원자료 기반 ETS(A,N,A)의 RMSE가 2,606.2, STL-ETS(A,N,N)는 2,635.8로 나타나, STL-ETS 모형의 RMSE가 원자료 기반 ETS 모형보다 다소 높았음.
- 이는 STL 분해가 예측 성과 개선에 기여할 수 있으나, 모형 구조와 자료 특성에 따라 그 효과가 달라질 수 있음을 시사함.

○ 예측조합 방식 간 비교에서는 분해-예측조합 모형의 성과가 전반적으로 우수하게 나타났음. 동일한 조합 방식의 원자료 기반-예측조합 모형과 비교하면, 분해-예측조합 모형은 RMSE 기준으로 3.9~6.8%, MAE 기준으로 4.8~5.4%, sMASE 기준으로 4.9~5.3% 낮은 예측오차를 보였음.

- 이는 시계열 분해를 통해 계절 성분과 계절조정 성분을 구분한 뒤 개별 예측모형을 적용하고, 이를 조합하는 방식이 원자료에 직접 예측조합을 적용하는 방식보다 예측 안정성 측면에서 유리할 수 있음을 시사함.

〈표 5-5〉 배추 가격 예측 정확도 평가 결과

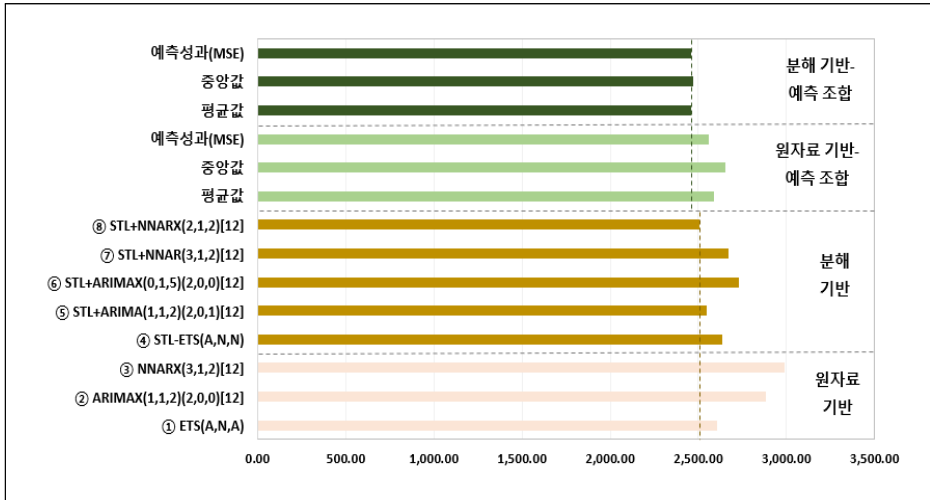
모형			RMSE	MAE	sMASE
원자료 기반	ETS 지수평활	① ETS(A,N,A)	2,606.2	2,017.2	0.680
	ARIMA	② ARIMAX(1,1,2)(2,0,0)[12]	2,882.4	2,187.7	0.737
	인공 신경망	③ NNARX(3,1,2)[12]	2,989.7	1,985.4	0.669
분해 기반	ETS 지수평활	④ STL-ETS(A,N,N)	2,635.8	2,010.5	0.678
	ARIMA	⑤ STL+ARIMA(1,1,2)(2,0,1)[12]	2,548.6	1,991.7	0.671
		⑥ STL+ARIMAX(0,1,5)(2,0,0)[12]	2,730.8	2,051.5	0.691
	인공 신경망	⑦ STL+NNAR(3,1,2)[12]	2,673.6	1,957.2	0.660
		⑧ STL+NNARX(2,1,2)[12]	2,504.7	1,807.4	0.609
비교 모형	원자료 기반- 예측조합	평균값	2,590.1	1,951.8	0.658
		중앙값	2,651.8	1,964.3	0.662
		예측성과(MSE)	2,557.4	1,959.0	0.660
본 연구 제안 알고리즘	분해 기반- 예측조합	평균값	2,457.1 (5.1)	1,858.7 (4.8)	0.626 (4.9)
		중앙값	2,472.2 (6.8)	1,858.9 (5.4)	0.627 (5.3)
		예측성과(MSE)	2,458.9 (3.9)	1,856.0 (5.3)	0.626 (5.2)

주: () 안의 값은 동일한 조합 방식의 원자료-예측조합 모형 대비 분해-예측조합 모형의 평가지표가 감소한 비율을 의미함. 개선율은(비교 모형의 평가 지뒃값- 본 연구 제안 알고리즘의 평가 지뒃값)/비교 모형의 평가 지뒃값×100으로 산출함. 굵은 글씨는 각 평가지표별로 최우수 모형을 의미함.

자료: 저자 작성.

○ 〈그림 5-5〉는 주요 예측모형별 RMSE를 비교한 결과임. 원자료 기반 개별모형은 모형 간 RMSE 편차가 상대적으로 크게 나타난 반면, 분해-예측조합 모형은 낮고 안정적인 RMSE 수준을 보였음. 이는 표본 외 평가 구간에서 시계열 분해와 예측조합을 결합한 모형의 예측 성과가 상대적으로 안정적이었음을 보여줌.

〈그림 5-5〉 배추 가격 예측 정확도 평가 결과(RMSE 기준)



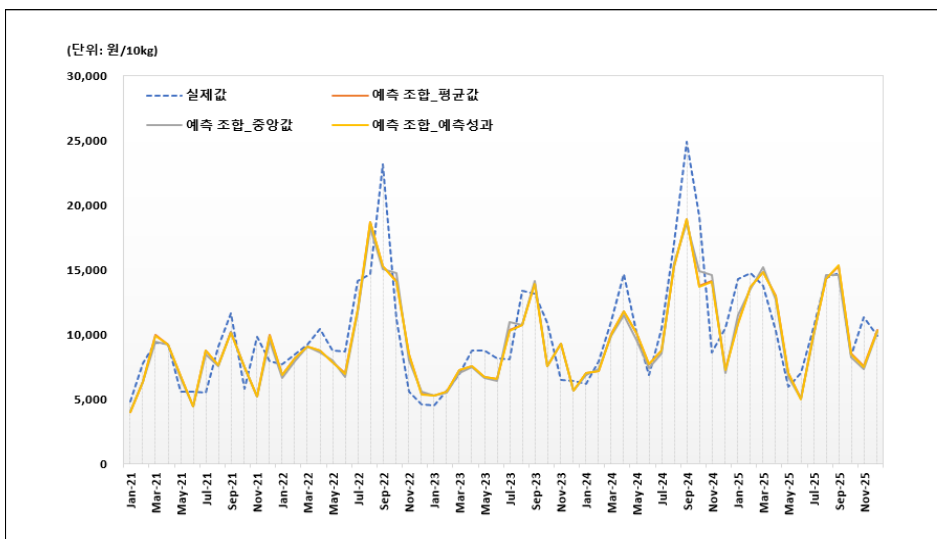
자료: 저자 작성.

○ 〈그림 5-6〉은 본 연구에서 제안한 분해 기반 예측조합 알고리즘의 배추 가격 예측치와 실제값을 비교한 결과임.

- 제안 알고리즘의 예측치는 실제 배추 가격의 전반적인 상승-하락 흐름과 계절적 변동 양상을 비교적 잘 반영하는 것으로 나타남.
- 다만 실제 가격이 급격하게 상승하는 일부 시점에서는 예측치와 실제값 간 차이가 확대되는 양상이 확인됨. 이는 배추 가격이 기상 여건, 작황 변화, 출하량 변동 등 단기 수급 충격에 민감하게 반응하는 특성을 가지기 때문으로 해석할 수 있음.
- 따라서 본 연구의 제안 알고리즘은 배추 가격의 전반적인 변동 흐름과 계절적 패턴을 파악하는 정량적 예측 도구로 활용될 수 있으나, 단기 수급 충격에 따른 급격한 가격 변동을 보완적으로 점검하기 위해서는 현장 정보 기반의 예측 전략이 병행될 필요가 있음.

○ 종합하면, 배추 가격 예측에서 원자료 기반 단일 모형보다 STL 분해 기반 개별모형의 예측 성과가 대체로 양호하였으며, 특히 본 연구에서 제안한 분해 기반 예측조합 모형이 가장 안정적인 예측 성과를 보였음. 이러한 결과는 배추 가격과 같이 계절성, 추세 변화 및 일시적 충격이 복합적으로 나타나는 농산물 가격 시계열의 경우, 시계열 분해와 예측조합을 결합한 접근이 예측오차를 완화하는 데 효과적일 수 있음을 보여줌.

〈그림 5-6〉 본 연구 제안 알고리즘의 예측치와 실제값 비교



자료: 저자 작성.

○ 앞서 제시한 RMSE, MAE, sMASE 기준의 예측오차 평가 결과, 본 연구의 분해 기반 예측조합 모형은 원자료 기반 및 분해 기반 개별 예측모형에 비해 전반적으로 낮은 예측오차를 보였음. 그러나 예측오차 지표는 모형 간 예측오차의 크기를 비교하는 데 유용한 반면, 이러한 차이가 통계적으로 유의한 예측력 차이인지 여부를 직접적으로 판단하기에는 한계가 있음. 이에 본 연구는 예측오차 기반 평가 결과에 대한 통계적 근거를 제시하기 위해 Modified Diebold and Mariano(MDM) 검정을 추가 수행함.

- <표 5-6>은 본 연구의 최종 제안 알고리즘인 분해-예측조합 모형 중 예측성과(MSE) 조합 방식과 원자료 기반 및 분해 기반 비교 모형 간의 MDM 검정 결과를 제시한 것임. 검정은 표본 외 평가 구간인 2021년 1월~2025년 12월의 60개 시점에 대한 예측오차를 대상으로 수행하였으며, 손실함수는 제곱오차와 절대오차를 각각 적용함.
- 본 연구에서는 MDM 통계량을 “제안 알고리즘의 손실값-비교 모형의 손실값” 기준으로 산출함. 따라서 MDM 통계량이 음(-)의 값을 가지면 제안 알고리즘의 손실값이 비교 모형보다 작아 예측오차가 상대적으로 낮았음을 의미함. 반대로 양(+)의 값은 제안 알고리즘의 손실값이 비교 모형보다 크게 나타났음을 의미함.
- 분석 결과, 본 연구의 제안 알고리즘은 원자료 기반 ARIMAX 모형에 비해 제곱오차 및 절대오차 기준 모두에서 음(-)의 MDM 통계량을 보였으며, 두 손실 기준 모두 5% 유의수준에서 통계적으로 유의한 예측오차 감소가 확인됨. 또한 원자료 기반 NNARX 모형과의 비교에서는 제곱오차 기준에서 예측오차 감소가 10% 유의수준에서 통계적으로 유의하게 나타남.
- 분해 기반 개별모형과의 비교에서도 일부 모형에 대해 통계적으로 유의한 예측력 차이가 나타남. 제안 알고리즘은 STL+ARIMAX 모형에 비해 제곱오차 및 절대오차 기준 모두에서 음(-)의 MDM 통계량을 보였으며, 두 기준 모두 5% 유의수준에서 유의한 차이가 확인됨. STL+ARIMA 모형과의 비교에서는 절대오차 기준에서 유의한 차이가 나타났고, STL+NNAR 모형과의 비교에서는 제곱오차 기준에서 유의한 차이가 확인됨.
- 반면 ETS 계열 모형, STL+ETS, STL+NNARX 및 원자료-예측조합 모형과

의 비교에서는 통계적으로 유의한 차이가 확인되지 않음. 이는 본 연구의 제안 알고리즘이 예측오차 지표상으로는 전반적으로 낮은 오차를 보였더라도, 모든 비교 모형에 대해 통계적으로 유의한 우위를 보였다고 해석하기는 어렵다는 점을 의미함.

- 특히 분해-예측조합과 원자료-예측조합 모형과의 예측력은 통계적으로 유의한 차이가 나타나지 않음. 이는 시계열 분해를 적용하지 않은 원자료를 이용하더라도 예측조합 방식을 활용할 경우 예측 성과가 안정될 수 있음을 의미함.

○ 종합하면, MDM 검정 결과 본 연구의 제안 알고리즘은 일부 원자료 기반 및 분해 기반 개별모형에 비해 예측 손실이 낮게 나타났으며, 그 차이가 통계적으로 유의한 것으로 확인됨. 다만 유의성은 비교 모형과 손실함수에 따라 다르게 나타났으므로, 본 연구의 제안 알고리즘이 모든 개별모형에 대해 일관되게 통계적으로 우월하다고 해석하기보다는, 예측오차 지표와 MDM 검정 결과를 함께 고려하여 상대적 예측 성과를 판단하는 것이 적절함.

〈표 5-6〉 예측력 차이의 통계적 유의성 검정: MDM 검정 결과

기준 모형: 본 연구 제안 알고리즘(“예측조합_예측성과(MSE)”)			제곱 오차	절대 오차
비교 대상 모형				
원자료 기반	ETS 지수평활	① ETS(A,N,A)	-0.661	-1.186
			(0.256)	(0.120)
	ARIMA	② ARIMAX(1,1,2)(2,0,0)[12]	-1.689	-1.709
			(0.048)**	(0.046)**
	인공 신경망	③ NNARX(3,1,2)[12]	-1.482	-0.658
			(0.072)*	(0.257)

(계속)

기준 모형: 본 연구 제안 알고리즘(“예측조합_예측성과(MSE)”)			제곱 오차	절대 오차
비교 대상 모형				
분해 기반	ETS 지수평활	④ STL-ETS(A,N,N)	-0.679	-1.042
			(0.250)	(0.151)
	ARIMA	⑤ STL+ARIMA(1,1,2)(2,0,1)[12]	-0.882	-1.605
			(0.191)	(0.057)*
		⑥ STL+ARIMAX(0,1,5)(2,0,0)[12]	-1.792	-2.071
	(0.039)**		(0.021)**	
인공 신경망	⑦ STL+NNAR(3,1,2)[12]	-1.406	-0.931	
		(0.083)*	(0.178)	
	⑧ STL+NNARX(2,1,2)[12]	-0.283	0.443	
(0.389)		(0.670)		
비교 모형	원자료 기반- 예측조합	평균값	-0.882	-0.804
			(0.191)	(0.212)
		중앙값	-0.991	-0.810
			(0.183)	(0.211)
		예측성과(MSE)	-0.800	-0.949
			(0.213)	(0.173)

주: **, *는 각각 5%, 10% 유의수준을 의미함. () 안의 값은 p-value임. MDM 통계량의 음(-)의 값은 기준 모형인 본 연구의 제안 알고리즘의 손실값이 비교 모형보다 작음을 의미함.
 자료: 저자 작성.

6

제안 알고리즘의 실무 적용 및 확대 방안

1. 제안 알고리즘과 농업관측센터 전망치 비교

- 본 연구의 제안 알고리즘은 과거 가격 및 관련 정량 자료에 기반하여 사전에 설정된 분석 절차에 따라 예측치를 산출하는 방식임. 반면 KREI 농업관측센터 전망치는 예측 정보 제공 시점까지 수집된 산지 작황, 출하 동향, 시장 상황 및 전문가 판단을 종합적으로 반영하여 산출됨. 따라서 두 전망치는 배추 가격을 예측한다는 점에서 동일한 목적을 가지지만, 활용 정보와 산출 방식에는 차이가 있음.
- 특히 단기 가격 변동성이 크게 나타날 수 있는 배추 품목의 특성을 고려할 때, 현장정보와 담당자 판단을 반영하는 농업관측센터 전망치는 단기 수급 변화와 가격 급등락 요인을 비교적 직접적으로 반영할 가능성이 있음. 반면 제안 알고리즘은 동일한 절차에 따라 반복적으로 예측치를 산출할 수 있다는 점에서 전망 과정의 객관성, 재현성 및 일관성을 보완하는 데 의의가 있음.

○ 이에 본 절에서는 제안 알고리즘과 KREI 농업관측센터 전망치의 예측 정확도를 비교하고, 두 전망치 간 예측력 차이의 통계적 유의성을 검토함. 또한 Encompassing 검정을 통해 두 예측치가 상호 보완적인 예측정보를 포함하는지 분석하고, 향후 실무 전망 과정에서 두 전망치를 함께 활용할 가능성을 검토함.

○ <표 6-1>은 제안 알고리즘과 농업관측센터 전망치의 예측 정확도를 비교한 결과임. 농업관측센터의 배추 가격 전망치는 매년 3~12월에 공식적으로 제공되므로, 비교는 2021~2025년 중 전망치가 존재하는 공통 평가 표본을 기준으로 수행함.

- 분석 결과, 공통 평가 표본에서 농업관측센터 전망치가 제안 알고리즘보다 전반적으로 낮은 예측오차를 보였음.
- RMSE 기준으로 농업관측센터 전망치는 2,362.6인 반면, 제안 알고리즘은 조합 방식에 따라 2,623.2~2,654.0 수준으로 나타남. MAE와 sMASE 기준에서도 농업관측센터 전망치가 더 낮은 예측오차를 보여, 공통 평가 표본에서는 농업관측센터 전망치의 예측 성과가 상대적으로 양호한 것으로 평가됨.

<표 6-1> 제안 알고리즘과 농업관측센터 전망치의 예측 정확도 비교

구분			RMSE	MAE	sMASE
본 연구 제안 알고리즘	예측조합	평균값	2,623.2	2,025.6	0.683
		중앙값	2,654.0	2,032.6	0.685
		예측성과(MSE)	2,623.2	2,020.4	0.681
KREI 농업관측센터 전망치			2,362.6	1,642.9	0.554

자료: 저자 작성.

○ <표 6-2>는 제안 알고리즘과 농업관측센터 전망치의 예측 정확도를 배추 작형별로 비교한 결과임. 본 연구는 농업관측센터의 작형 구분을 준용하여 겨울

배추는 1~4월, 봄배추는 5~6월, 여름배추는 7~10월, 가을배추는 11~12월로 설정함. 다만 농업관측센터 전망치가 3~12월에 제공되므로, 공통 평가 표본에서 겨울배추 비교는 실질적으로 3~4월 관측치를 대상으로 함.

- 작형별 결과를 살펴보면, 겨울 작형에서는 제안 알고리즘이 농업관측센터 전망치보다 낮은 RMSE를 보였음. 제안 알고리즘의 RMSE는 조합 방식에 따라 1,394.6~1,429.0으로 나타난 반면, 농업관측센터 전망치는 1,813.7로 나타남. 이는 특정 작형에서는 정량모형 기반 예측치가 상대적으로 양호한 성과를 보일 수 있음을 시사함.
- 여름 작형에서는 농업관측센터 전망치와 제안 알고리즘 간 RMSE 차이가 크지 않았음. 농업관측센터 전망치의 RMSE는 3,251.6이고, 제안 알고리즘은 3,187.9~3,196.0 수준으로 나타나 두 전망치가 유사한 수준의 예측 성과를 보였음.
- 반면 봄과 가을 작형에서는 농업관측센터 전망치가 제안 알고리즘보다 낮은 예측오차를 보였음. 특히 가을 작형에서는 농업관측센터 전망치의 RMSE가 1,619.1인 데 반해 제안 알고리즘은 3,131.8~3,238.7 수준으로 나타나 두 전망치 간 격차가 상대적으로 가장 크게 확인됨.
- 가을배추는 김장철 수요가 집중되는 시기이며, 정부의 수급안정대책, 비축방출, 출하 조절 등 정책적·시장적 대응이 가격 형성에 영향을 미칠 수 있음. 따라서 이 시기에는 과거 가격 및 수급 변수 중심의 정량모형보다 현장 정보와 정책 여건을 함께 반영하는 농업관측센터 전망체계의 실무적 강점이 상대적으로 크게 나타날 수 있음.

○ 이상의 결과는 공통 평가 표본 전체에서는 농업관측센터 전망치의 예측 성과가 상대적으로 양호하였으나, 작형 및 시기에 따라 본 연구 알고리즘도 일정 수준의 예측력을 보일 수 있음을 시사함. 따라서 두 전망치가 단순한 대체 관

계로 보기보다, 품목·시기별 정보 반영 방식의 차이를 고려하여 함께 검토할 필요가 있음.

〈표 6-2〉 제안 알고리즘과 농업관측센터 전망치의 작형별 예측 정확도 비교

구분				RMSE	MAE	sMASE
본 연구 제안 알고리즘	예측조합	평균값	겨울(1~4월)	1,422.6	1,095.3	0.369
			봄(5~6월)	1,390.6	1,272.0	0.429
			여름(7~10월)	3,191.7	2,497.3	0.842
			가을(11~12월)	3,141.6	2,694.8	0.908
		중앙값	겨울(1~4월)	1,394.6	1,077.8	0.363
			봄(5~6월)	1,401.7	1,273.6	0.429
			여름(7~10월)	3,187.9	2,505.0	0.844
			가을(11~12월)	3,238.7	2,714.3	0.915
		예측성과(MSE)	겨울(1~4월)	1,429.0	1,096.4	0.370
			봄(5~6월)	1,398.9	1,280.1	0.431
			여름(7~10월)	3,196.0	2,491.1	0.840
			가을(11~12월)	3,131.8	2,680.6	0.904
KREI 농업관측센터 전망치			겨울(1~4월)	1,813.7	1,362.0	0.459
			봄(5~6월)	923.8	714.3	0.241
			여름(7~10월)	3,251.6	2,406.4	0.811
			가을(11~12월)	1,619.1	1,325.5	0.447

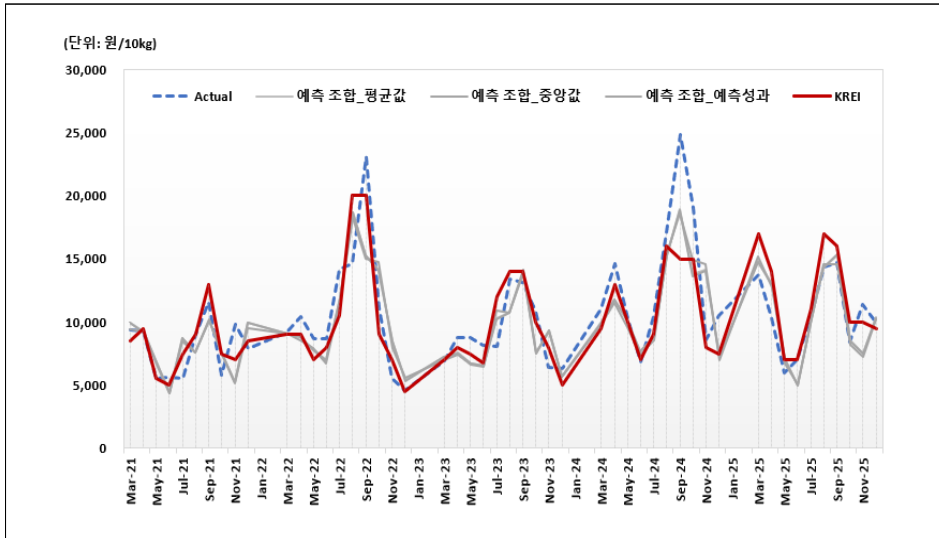
자료: 저자 작성.

○ 〈그림 6-1〉은 제안 알고리즘과 농업관측센터 전망치를 실제값과 비교한 결과를 제시한 것임.

- 농업관측센터 전망치는 일부 가격 급등 시점에서 실제 가격 상승을 비교적 잘 반영하는 것으로 나타남. 다만 모든 시점에서 농업관측센터 전망치가 실제값에 더 근접한 것은 아니며, 일부 시점에서는 제안 알고리즘의 예측치가 실제 가격 수준과 유사하게 나타나는 경우도 확인됨.
- 이는 두 전망치가 서로 다른 정보 기반에 의해 산출되기 때문에 가격 변동을 포착하는 방식에도 차이가 있음을 보여줌. 따라서 예측 성과의 단순 비

교와 함께 두 전망치가 서로 다른 정보를 제공하는지 검토할 필요가 있음.

〈그림 6-1〉 제안 알고리즘 및 농업관측센터의 예측치와 실제값 비교



자료: 저자 작성.

○ Encompassing 검정은 두 예측치가 서로 다른 정보를 포함하고 있는지, 그리고 두 예측치를 조합할 경우 추가적인 예측력 개선이 있는지를 검토하는 방법임.

○ 본 연구는 KREI 농업관측센터 전망치를 기준 예측치로, 제안 알고리즘을 비교 예측치로 설정하여 Encompassing 검정을 수행함. 구체적으로 농업관측센터 전망치의 예측오차를 e_t^{KREI} , 제안 알고리즘의 예측오차를 e_t^{ALG} 라고 할 때, $e_t^{KREI} = \alpha + \lambda(e_t^{KREI} - e_t^{ALG}) + \varepsilon_t$ 의 회귀식으로 추정함.

- 회귀식에서 $\lambda = 0$ 이라는 귀무가설이 기각되지 않으면, 농업관측센터 전망치가 제안 알고리즘의 예측치에 포함된 정보를 포괄하는 것으로 해석할 수 있음. 반대로 귀무가설이 기각되면($\lambda \neq 0$), 제안 알고리즘의 예측치가

KREI 전망치에 포함되지 않은 추가적인 예측 정보를 제공할 가능성이 있음을 의미함.

○ <표 6-3>은 농업관측센터 전망치와 제안 알고리즘 간 Encompassing 검정 결과를 제시한 것임. 분석 결과, λ 의 추정치는 0.347로 나타났으며 5% 유의수준에서 통계적으로 유의하였음. 이는 농업관측센터 전망치가 제안 알고리즘의 예측 정보를 완전히 포괄하기 어렵고, 제안 알고리즘이 농업관측센터 전망치에 포함되지 않은 추가적인 정량 정보를 제공할 가능성이 있음을 시사함.

○ 조합 가중치 측면에서 제안 알고리즘에 부여되는 가중치(λ)는 0.347, 농업관측센터 전망치에 부여되는 가중치($1 - \lambda$)는 0.653으로 해석할 수 있음.

- 다만 이 가중치는 2021~2025년 공통 평가기간 대상 사후적으로 추정된 결과이므로, 향후 전망 업무에 그대로 적용할 고정 가중치로 해석하기보다는 두 전망치의 선형 조합 가능성을 보여주는 시범적 기준으로 이해하는 것이 적절함. 실제 운영 단계에서는 새로운 실제값과 예측값이 축적됨에 따라 조합 가중치의 안정성을 주기적으로 점검하고, 필요할 경우 재추정하는 방식으로 운영하는 것이 바람직함.

<표 6-3> 농업관측센터 전망치와 제안 알고리즘 간 Encompassing 검정 결과

구분	상수항	λ
회귀계수	371.9	0.347**
	(0.261)	(0.047)

주: () 안의 값은 p-value이고, **, *는 각각 5%, 10% 유의수준을 의미함.
자료: 저자 작성.

○ 이상의 분석 결과를 종합하면, 농업관측센터 전망치와 제안 알고리즘은 상호 보완적으로 활용될 가능성이 있는 것으로 판단됨. 농업관측센터 전망치는 산지 작황, 출하 동향, 시장 상황 및 전문가 판단을 반영한다는 점에서 전반적으

로 낮은 예측오차를 보였음. 반면 제안 알고리즘은 정량 자료와 사전에 설정된 분석 절차에 기반하여 예측치를 산출한다는 점에서 전망 과정의 객관성, 재현성 및 일관성을 보완할 수 있음.

- 특히 Encompassing 검정 결과는 제안 알고리즘이 농업관측센터 전망치에 포함되지 않은 추가적인 예측 정보를 제공할 가능성을 보여줌. 따라서 향후 농업관측센터 전망 체계에서는 현장 정보와 전문가 판단에 기반한 기존 전망 절차를 중심으로 하되, 제안 알고리즘의 예측치를 정량적 참고지표 또는 조합 예측치로 병행 활용하는 방안을 검토할 수 있음.

2. 제안 알고리즘의 실무 적용 체계

- 본 연구에서 제안한 배추 도매가격 예측 알고리즘을 실제 전망 업무에 적용하기 위한 실무 적용 절차는 ① 원자료 업데이트 및 R 코드 실행 준비, ② 실무자 입력부 수정, ③ 예측 가능 여부 점검 및 코드 실행, ④ 예측 결과 출력 확인, ⑤ 결과의 종합 검토 및 활용의 단계로 구성됨. 이하에서는 2026년 5월 배추 도매가격 예측을 예시로 각 단계의 내용을 제시함.

① 원자료 업데이트 및 R 코드 실행 준비

- 실무자는 먼저 원자료 엑셀 파일의 raw_data 시트를 예측 대상 월의 직전 월까지 갱신해야 함. 예를 들어 2026년 5월 가격을 예측하는 경우, 2026년 4월까지의 배추 도매가격, 반입량, 김치 수입량의 관측값이 모두 입력되어 있어야 함. 특히 반입량과 김치 수입량은 예측모형에서 1~3개월 시차변수 생성에 활용되므로, 최근 월 자료의 누락 여부를 함께 확인할 필요가 있음.

○ 원자료 갱신 후에는 R 스크립트 파일과 원자료 엑셀 파일이 동일한 폴더에 저장되어 있는지 확인함. 본 연구의 실무용 코드는 동일 폴더 내 원자료 파일을 불러오는 방식으로 구성되어 있으므로, 파일 저장 위치가 다르거나 파일명이 변경될 경우 코드가 정상적으로 실행되지 않을 수 있음. 따라서 예측 작업에 앞서 R 스크립트와 원자료 엑셀 파일의 저장 위치 및 파일명을 확인하는 절차가 필요함.

② 실무자 입력부 수정

○ 다음으로 실무자는 R 코드 내 ‘실무자 입력부’에서 예측 대상 월, KREI 전망치, 최종 권장 방식의 세 가지 항목을 수정함. 2026년 5월 가격을 예측하고 KREI 전망치가 6,000원/10kg인 경우, 입력부는 <글상자 6-1>과 같이 설정할 수 있음.

- target_ym은 예측 대상 월을 지정하는 항목임.
- krei_forecast는 KREI 자체 전망 업무를 통해 산출된 배추 도매가격 전망치를 의미함.
- final_recommendation_type은 최종 권장 예측치의 산출 방식을 지정하는 항목임. 예를 들어 “krei_algorithm”으로 설정할 경우, 알고리즘 조합 예측치와 KREI 전망치를 사전에 설정된 가중치로 결합한 값이 최종 권장 예측치로 제시됨.
- 실무 적용 단계에서 매월 수정이 필요한 항목을 위 세 가지로 제한한 것은 반복적인 예측 업무에서 입력 오류를 줄이고, 예측 작업의 일관성을 확보하기 위한 것임.

〈글상자 6-1〉 실무자 입력부 설정 예시

```
target_ym      <- "2026-05"  
krei_forecast  <- 6000  
final_recommendation_type <- "krei_algorithm"
```

자료: 저자 작성.

③ 예측 가능 여부 점검 및 코드 실행

- 입력값 수정 후에는 예측 가능 여부를 점검함. 원자료에 누락된 관측값이 있거나 날짜 형식이 올바르게 입력되지 않은 경우, 코드 실행 과정에서 오류 메시지가 출력됨. 이 경우 실무자는 오류 내용을 확인한 뒤 원자료를 보완하고, 예측 가능 여부를 재점검해야 함.
- 자료 이상이 확인되지 않으면 RStudio에서 전체 R 코드를 실행함. 코드 실행은 RStudio의 Source 기능 또는 단축키(Ctrl+Shift+Enter)를 이용하여 수행할 수 있음. 코드가 정상적으로 실행되면 자료 불러오기, 변수 구성, 결측 및 중복 월 점검, 개별모형 예측, 조합 예측치 산출, KREI 전망치 결합, 결과 출력이 순차적으로 수행됨.

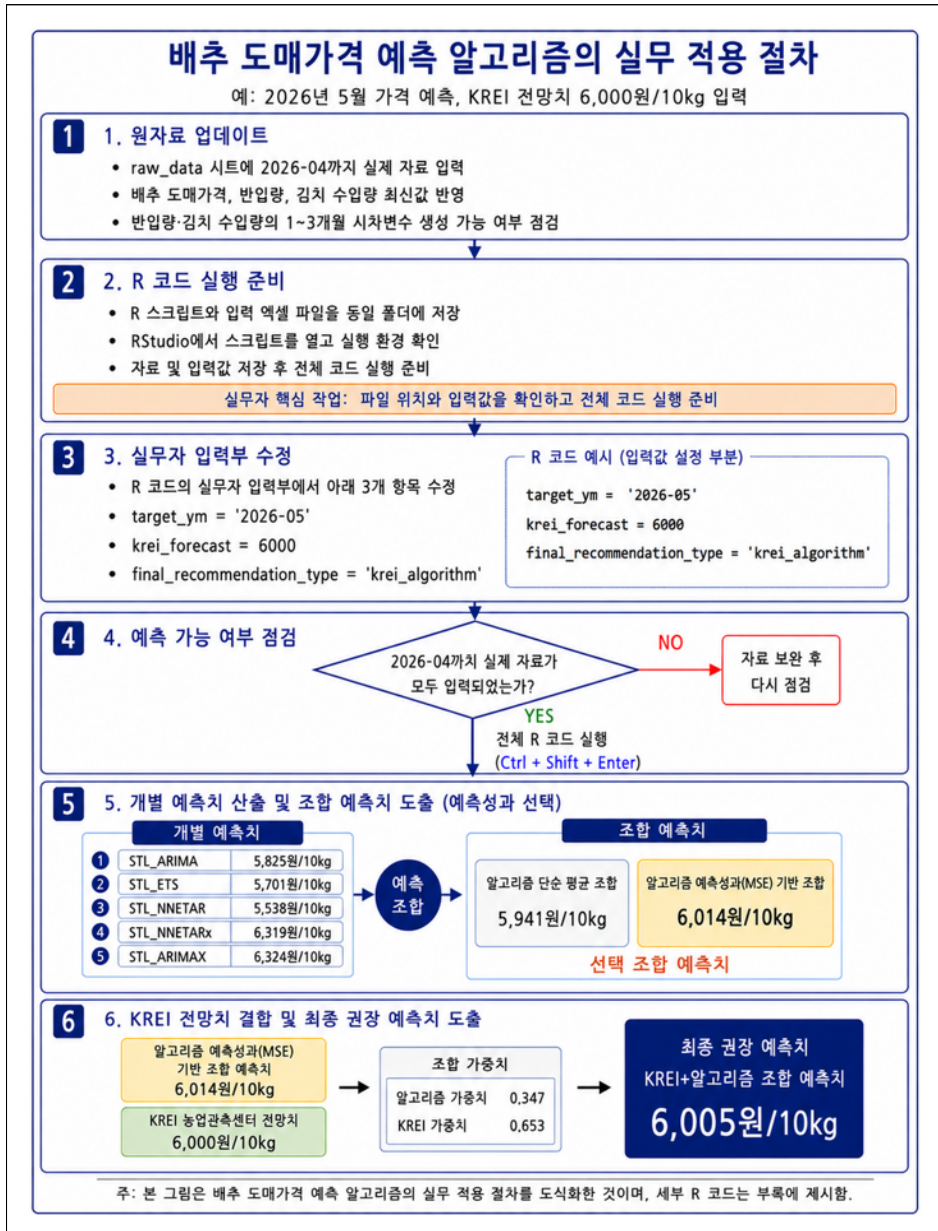
④ 예측 결과 출력 확인

- 코드 실행이 완료되면 콘솔(Console) 화면에 간략 예측 보고서가 출력됨. 출력 결과에는 예측 대상 월, 사용 자료의 관측 월, 개별모형 예측치, 알고리즘 조합 예측치, KREI 전망치, 최종 권장 예측치 등이 포함됨. 이를 통해 실무자는 예측 결과의 전반적인 수준과 개별모형 간 차이를 일차적으로 확인할 수 있음.

⑤ 결과의 종합 검토 및 활용

- 실무자는 최종 권장 예측치만을 단독으로 참고하기보다, 개별모형 예측치, 알고리즘 조합 예측치, KREI 전망치를 함께 검토할 필요가 있음. 특히 개별모형 간 예측치 편차가 과도하게 크지 않은지, 알고리즘 조합 예측치와 KREI 전망치 간 괴리가 큰지, 최근 산지 작황이나 시장 상황에 비추어 최종 권장 예측치가 과도하게 높거나 낮지 않은지를 종합적으로 점검해야 함.
- 이러한 검토 과정은 정량모형의 산출 결과와 실무적 판단 간의 정합성을 확인하는 절차로 이해할 수 있음. 즉, 본 연구의 실무용 예측 코드는 최종 전망치를 자동으로 확정하기 위한 도구라기보다는, 농업관측센터의 중앙자문회의 및 보고서 작성 과정에서 활용 가능한 정량적 참고자료를 제공하는 역할을 수행함.
- 이상의 실무 적용 절차는 <그림 6-2>에 제시하였음. 이러한 절차를 통해 실무자는 매월 동일한 방식으로 배추 도매가격 예측 작업을 수행할 수 있음.
 - 실무용 R 코드의 세부 내용은 부록에 수록하였으며, 원자료 업데이트에서 간략 예측 출력까지의 전체 과정을 포함함.

〈그림 6-2〉 본 연구 제안 알고리즘의 실무 적용 절차



주: 본 그림의 도식화 과정에서 생성형 AI 도구인 ChatGPT(OpenAI)를 보조적으로 활용하였으며, 최종 내용은 저자가 검토·수정하였음.
자료: 저자 작성.

3. 실무 적용 확대를 위한 향후 개선 과제

- 앞서 본 연구의 제안 알고리즘과 KREI 농업관측센터 전망치의 예측 성과를 비교하고, 이를 실제 전망 업무에 적용하기 위한 실무 절차를 제시하였음. 본 절에서는 제안 알고리즘의 실무 활용도를 높이고, 향후 농업관측센터의 가격 전망체계와 보다 유기적으로 연계하기 위한 개선 과제를 제시함.
- 첫째, 제안 알고리즘의 예측 결과는 현장정보와 병행하여 활용될 필요가 있음. 본 연구의 알고리즘은 과거 가격 및 관련 정량 자료에 기반하여 예측치를 산출하므로, 단기간에 발생하는 수급 충격은 현장 조사자료와의 종합적 검토가 필요함.
- 둘째, 가격 수준에 대한 점 예측(point forecast)을 보완하기 위해 가격 급등 가능성을 사전에 점검할 수 있는 확률예측모형의 도입을 검토할 필요가 있음. 제안 알고리즘은 특정 시점의 배추 도매가격 수준을 예측하는 데 초점을 두고 있으나, 실무적으로는 가격이 어느 수준으로 형성될 것인가와 함께 가격 급등 위험이 높아지고 있는지를 파악하는 것도 중요함.
 - 배추 가격 전망체계에서 점 예측 알고리즘과 함께 가격 급등 확률을 제시하는 보조적 모형을 병행한다면, 중앙자문회의 및 수급 대응 과정에서 가격 급등 위험을 사전에 점검하는 데 활용될 수 있을 것으로 판단됨.
- 셋째, 농업관측센터 전망치와 제안 알고리즘 예측치의 조합 가중치는 실제 운영 과정에서 주기적으로 재점검할 필요가 있음. 본 연구에서 제시한 조합 가중치는 특정 평가기간의 예측 성과에 기초하여 산출된 값이므로, 향후 실제 전망 업무에서 예측치와 실제값이 축적됨에 따라 일정 주기로 재추정하는 것이 바람직함.

○ 넷째, 예측모형에 활용되는 설명변수는 실무 활용성과 예측력 간의 균형을 고려하여 단계적으로 확장할 필요가 있음. 향후 기상, 생육, 저장량, 산지 출하의향, 품위 변화 등 다양한 조사자료와 현장정보를 추가로 활용할 수 있으나, 이들 변수가 예측 시점에 안정적으로 확보 가능한지, 미래값을 별도로 예측해야 하는지, 시차변수만으로 활용 가능한지에 대한 검토가 선행되어야 함.

- 설명변수 확대는 더 많은 정보를 반영할 수 있다는 장점이 있으나, 자료 수집·갱신 비용, 미래값에 대한 예측 오차, 모형 복잡성, 과적합 가능성도 함께 증가시킬 수 있음.

- 따라서 단순한 변수 확대나 머신러닝 모형 도입이 예측력 개선을 보장한다고 보기보다는, 추가 변수의 예측 기여도와 실무 운영 비용을 함께 고려하여 선별적으로 반영하는 것이 바람직함.

○ 다섯째, 예측 결과에 대한 사후 평가체계를 지속적으로 운영할 필요가 있음. 실무 적용 과정에서는 매월 예측치와 실제값을 축적하고, 월별·작형별·가격 급등락 시기별 예측 오차를 점검할 필요가 있음. 이러한 사후 평가는 특정 시기나 작형에서 예측 오차가 반복적으로 확대되는지를 확인하고, 모형 구조, 설명변수, 조합 가중치 등을 보완하는 기초자료로 활용될 수 있음.

○ 여섯째, 제안 알고리즘의 품목 확대 적용 가능성을 검토할 필요가 있음. 본 연구의 결과는 배추 가격 예측이라는 개별 품목 차원의 성과 검토에 그치지 않고, 작형별 계절성과 가격 변동성이 뚜렷한 노지 엽근채소 품목으로의 확대 적용 가능성을 검토하는 기초자료로 활용될 수 있음. 특히 배추는 작형별 출하시기가 구분되고 단기 수급 충격에 따른 가격 변동성이 큰 품목이라는 점에서, 본 연구의 시계열 분해-예측조합 접근은 유사한 수급 및 가격 구조를 갖는 품목에도 적용 가능할 것으로 판단됨. 향후에는 농업관측센터의 주요 전망 품목 중 노지 엽근채소류를 중심으로 적용 범위를 확대하여, 제안 알고리즘의

일반화 가능성과 실무 활용성을 추가적으로 검토할 필요가 있음.

- 일곱째, 실무 활용도를 높이기 위해 예측 코드의 운영 편의성과 관리 체계를 고도화할 필요가 있음. 본 연구에서는 실무자가 예측 대상 월, KREI 전망치, 최종 권장 방식 등 일부 입력값만 수정하면 예측 결과를 확인할 수 있도록 R 코드를 구성하였음. 향후에는 원자료 업데이트, 예측 실행, 결과 저장, 시각화, 사후 평가표 작성 등이 자동으로 연계되도록 시스템화함으로써 전망 업무의 효율성과 재현성을 높일 수 있음.
- 종합하면, 본 연구의 제안 알고리즘은 농업관측센터의 기존 배추 가격 전망 체계를 대체하기 위한 수단이라기보다, 전망 과정에서 활용 가능한 정량적 보조도로 이해하는 것이 적절함. 향후 실제 전망 업무에서 알고리즘 예측치와 KREI 전망치를 함께 추적하고, 예측 오차와 조합 가중치를 주기적으로 점검한다면 배추 가격 전망의 객관성, 재현성 및 안정성을 제고하는 데 기여할 수 있을 것으로 판단됨. 나아가 가격 급등 가능성을 별도로 제시하는 보조적 확률예측모형과 연계한다면, 가격 수준 전망뿐만 아니라 수급 위험에 대한 조기 경보 기능도 강화할 수 있을 것으로 기대됨.

배추 도매가격 예측 실무용 R 코드

```
# -----
# 실무자는 [1. 실무자 입력부]의 세 항목만 수정하여 사용함
# - target_ym          : 예측 대상월
# - krei_forecast      : KREI 농업관측센터 전망치
# - final_recommendation_type : 최종 권장 예측치 선택 방식
# -----

# =====
# 0. 패키지 설치 및 로드
# =====

pkgs <- c("readxl", "dplyr", "zoo", "forecast")

need <- pkgs[!pkgs %in% installed.packages()[, "Package"]]
if (length(need) > 0) install.packages(need)

suppressPackageStartupMessages({
  library(readxl)
  library(dplyr)
  library(zoo)
  library(forecast)
})

# =====
# 1. 실무자 입력부: 매월 수정하는 부분
# =====

# -----
# 1-1. 예측 대상월
# -----
# 예: 2026년 5월 가격 예측 → "2026-05"
target_ym <- "2026-05"

# -----
# 1-2. KREI 농업관측센터 전망치 입력
# -----
# KREI 전망치가 있으면 숫자로 입력
```

```

# 예: krej_forecast <- 6000
#
# KREI 전망치가 없거나 알고리즘 예측치만 확인하려면 NA_real_ 입력
# 예: krej_forecast <- NA_real_
krej_forecast <- 6000

# -----
# 1-3. 최종 권장 예측치 선택 방식
# -----
# "krej_algorithm" : KREI 전망치와 알고리즘 예측성과 기반 조합 예측치를 결합
# "performance"   : 알고리즘 예측성과 기반 조합을 최종 권장 예측치로 사용
# "mean"          : 알고리즘 단순 평균 조합을 최종 권장 예측치로 사용
# "krei"          : KREI 전망치를 최종 권장 예측치로 사용
final_recommendation_type <- "krej_algorithm"

# =====
# 2. 분석 파일 및 모형 설정
# =====

# 원자료 엑셀 파일명
input_excel_name <- "cabbage_raw_data_template.xlsx"

# 원자료가 입력된 시트명
input_sheet <- "raw_data"

# 가격 출력 단위
price_unit <- "원/10kg"

# Sliding window 길이: 최근 216개월 사용
L <- 216L

# -----
# 2-1. 알고리즘 내부 예측성과 기반 조합가중치
# -----
# 주: 아래 값은 예시값이며, 연구용 코드에서 산출한 검증구간 MSE 기반
#      가중치로 교체하여 사용할 수 있음. 가중치 합계는 1이어야 함.
performance_weights <- c(
  STL_ARIMA = 0.15,
  STL_ETS   = 0.25,
  STL_NNETAR = 0.10,
  STL_NNETARx = 0.20,
  STL_ARIMAX = 0.30
)

# -----
# 2-2. KREI 전망치와 알고리즘 예측치 간 결합가중치
# -----
# 주: 아래 값은 Encompassing 검정 결과에 기초한 예시값임.

```

```

# 운영 단계에서는 실제값과 예측값이 축적됨에 따라 재점검할 필요가 있음.
krei_algorithm_weights <- c(
  ALGORITHM = 0.347,
  KREI      = 0.653
)

# -----
# 2-3. 연구 결과 기반 모형 설정값
# -----

# STL 분해 설정
stl_fixed <- list(
  s.window = "periodic",
  t.window = 19,
  l.window = 13,
  inner    = 2,
  outer    = 0
)

# 계절조정 성분에 대한 고정 모형 구조
sa_arima_order <- c(1, 1, 2)
sa_arima_seasonal <- c(2, 0, 1)

sa_arimax_order <- c(0, 1, 5)
sa_arimax_seasonal <- c(2, 0, 0)

# ETS(A,N,N)
sa_ets_model <- "ANN"

# NNETAR 설정
nn_p <- 3L
nn_P <- 1L
nn_size <- 2L
nn_repeats <- 20L
set_seed <- 123

# =====
# 3. 보조 함수
# =====

get_script_dir <- function() {
  if (requireNamespace("rstudioapi", quietly = TRUE)) {
    path <- tryCatch(
      rstudioapi::getSourceEditorContext()$path,
      error = function(e) ""
    )
  }
}

```

```

    if (!is.null(path) && nzchar(path)) {
      return(dirname(normalizePath(path)))
    }
  }

  cmd_args <- commandArgs(trailingOnly = FALSE)
  file_arg <- grep("^--file=", cmd_args, value = TRUE)

  if (length(file_arg) > 0) {
    path <- sub("^--file=", "", file_arg[1])
    return(dirname(normalizePath(path)))
  }

  getwd()
}

parse_ym <- function(x) {

  if (inherits(x, "Date")) {
    return(as.Date(format(x, "%Y-%m-01")))
  }

  if (inherits(x, c("POSIXct", "POSIXlt"))) {
    return(as.Date(format(as.Date(x), "%Y-%m-01")))
  }

  if (is.numeric(x)) {
    return(as.Date(format(as.Date(x, origin = "1899-12-30"), "%Y-%m-01")))
  }

  x0 <- trimws(as.character(x))
  x0[x0 %in% c("", "...", "NA", "NaN", "NULL")] <- NA_character_

  out <- rep(as.Date(NA), length(x0))

  for (i in seq_along(x0)) {

    if (is.na(x0[i])) next

    xi <- x0[i]
    xi <- gsub("년", "-", xi)
    xi <- gsub("월", "", xi)
    xi <- gsub("/", "-", xi)
    xi <- gsub("\\.", "-", xi)
    xi <- gsub("\\s+", "", xi)

    if (grepl("^[0-9]{4}-[0-9]{1,2}$", xi)) {
      parts <- strsplit(xi, "-")[[1]]
    }
  }
}

```

```

yy <- parts[1]
mm <- sprintf("%02d", as.integer(parts[2]))
out[i] <- as.Date(paste0(yy, "-", mm, "-01"))
next
}

if (grepl("^[0-9]{4}-[0-9]{1,2}-[0-9]{1,2}$", xi)) {
  parts <- strsplit(xi, "-")[1]
  yy <- parts[1]
  mm <- sprintf("%02d", as.integer(parts[2]))
  out[i] <- as.Date(paste0(yy, "-", mm, "-01"))
  next
}

if (grepl("^[0-9]{6}$", xi)) {
  yy <- substr(xi, 1, 4)
  mm <- substr(xi, 5, 6)
  out[i] <- as.Date(paste0(yy, "-", mm, "-01"))
  next
}
}

out
}

as_num <- function(x) {
  as.numeric(gsub(",", "", as.character(x)))
}

ym_to_c <- function(ym) {
  ym <- as.yearmon(ym)
  c(as.integer(format(ym, "%Y")), as.integer(format(ym, "%m")))
}

ym_str <- function(ym) {
  format(as.yearmon(ym), "%Y-%m")
}

ts_to_yearmon <- function(x) {
  tt <- time(x)
  yr <- floor(tt)
  mo <- round((tt - yr) * frequency(x) + 1e-8) + 1
  as.yearmon(paste(yr, mo), "%Y %m")
}

lag_vec <- function(x, k) {
  n <- length(x)
  if (k <= 0) return(x)

```

```

c(rep(NA_real_, k), x[1:(n - k)])
}

format_price <- function(x) {
  ifelse(
    is.finite(x),
    paste0(format(round(x, 0), big.mark = ","), price_unit),
    "미산출"
  )
}

get_stl_components <- function(log_y_ts, stl_params) {

  fit <- stl(
    log_y_ts,
    s.window = stl_params$s.window,
    t.window = stl_params$t.window,
    l.window = stl_params$l.window,
    inner = stl_params$inner,
    outer = stl_params$outer
  )

  seasonal_ts <- ts(
    as.numeric(fit$time.series[, "seasonal"]),
    start = start(log_y_ts),
    frequency = frequency(log_y_ts)
  )

  sa_ts <- ts(
    as.numeric(log_y_ts) - as.numeric(seasonal_ts),
    start = start(log_y_ts),
    frequency = frequency(log_y_ts)
  )

  list(sa_ts = sa_ts, seasonal_ts = seasonal_ts)
}

season_fc1_snaive <- function(seasonal_ts) {
  as.numeric(snaive(seasonal_ts, h = 1)$mean[1])
}

get_resid_var <- function(fit) {

  res <- tryCatch(as.numeric(residuals(fit)), error = function(e) NA_real_)
  res <- res[is.finite(res)]

  if (length(res) < 2) return(0)

```

```

sigma2 <- var(res, na.rm = TRUE)

if (!is.finite(sigma2) || sigma2 < 0) return(0)

sigma2
}

inv_log_biasadj <- function(mu_hat, sigma2_hat = 0) {

  if (!is.finite(mu_hat)) return(NA_real_)

  if (!is.finite(sigma2_hat) || sigma2_hat < 0) {
    sigma2_hat <- 0
  }

  exp(mu_hat + 0.5 * sigma2_hat)
}

align_y_xreg_lastok <- function(y_ts, xmat) {

  y_ts <- as.ts(y_ts)
  xm <- as.matrix(xmat)
  storage.mode(xm) <- "double"

  ok <- is.finite(as.numeric(y_ts)) &
    apply(xm, 1, function(r) all(is.finite(r)))

  if (!isTRUE(ok[length(ok)])) {
    return(list(y = NULL, x = NULL, last_ok = FALSE))
  }

  y_num <- as.numeric(y_ts)
  y_keep <- y_num[ok]
  first_idx <- which(ok)[1]
  start_ym <- ts_to_yearmon(y_ts)[first_idx]

  y_keep_ts <- ts(
    y_keep,
    start = ym_to_c(start_ym),
    frequency = frequency(y_ts)
  )

  list(y = y_keep_ts, x = xm[ok, , drop = FALSE], last_ok = TRUE)
}

zscore_cont_only <- function(x_train, x_future, cont_cols, eps = 1e-8) {

  x_train <- as.matrix(x_train)

```

```

x_future <- as.matrix(x_future)

mu <- colMeans(x_train[, cont_cols, drop = FALSE], na.rm = TRUE)
sdv <- apply(x_train[, cont_cols, drop = FALSE], 2, sd, na.rm = TRUE)
sdv[!is.finite(sdv) | sdv < eps] <- 1

x_train[, cont_cols] <- sweep(
  sweep(x_train[, cont_cols, drop = FALSE], 2, mu, "-"),
  2, sdv, "/"
)

x_future[, cont_cols] <- sweep(
  sweep(x_future[, cont_cols, drop = FALSE], 2, mu, "-"),
  2, sdv, "/"
)

list(x_train = x_train, x_future = x_future)
}

make_price_features <- function(price.ts) {

  idx_ym <- ts_to_yearmon(price.ts)
  date_key <- ym_str(idx_ym)

  lp <- log(as.numeric(price.ts))
  dp12 <- lp - lag_vec(lp, 12)

  ma3 <- zoo::rollmean(lp, k = 3, align = "right", fill = NA_real_)
  ma3_lag1 <- lag_vec(ma3, 1)
  ma3_lag12 <- lag_vec(ma3, 12)

  level_rel_12_ma3 <- ma3_lag1 - ma3_lag12
  level_rel_12_ma3_lag1 <- lag_vec(level_rel_12_ma3, 1)

  month_num <- as.integer(format(as.Date(idx_ym), "%m"))

  crop_id <- rep(NA_integer_, length(month_num))
  crop_id[month_num %in% 1:4] <- 1L
  crop_id[month_num %in% 5:6] <- 2L
  crop_id[month_num %in% 7:10] <- 3L
  crop_id[month_num %in% 11:12] <- 4L

  crop_rel_level <- rep(NA_real_, length(lp))

  for (i in seq_along(lp)) {
    if (i <= 1) next

```

```

current_crop <- crop_id[i]
hist_idx <- which(crop_id == current_crop & seq_along(lp) < i & is.finite(lp))

if (length(hist_idx) >= 3 && is.finite(lp[i - 1])) {
  crop_mean_hist <- mean(lp[hist_idx], na.rm = TRUE)
  crop_rel_level[i] <- lp[i - 1] - crop_mean_hist
}
}

data.frame(
  Date = date_key,
  dp12_lag1 = lag_vec(dp12, 1),
  level_rel_12_ma3 = level_rel_12_ma3_lag1,
  crop_rel_level = lag_vec(crop_rel_level, 1),
  stringsAsFactors = FALSE
)
}

make_qm_features <- function(in_qty.ts, kimchi_imp.ts, max_k = 3) {

  idx_ym <- ts_to_yearmon(in_qty.ts)
  date_key <- ym_str(idx_ym)

  qv <- as.numeric(log1p(in_qty.ts))
  mv <- as.numeric(log1p(kimchi_imp.ts))

  out <- data.frame(Date = date_key, stringsAsFactors = FALSE)

  for (k in 1:max_k) {
    out[[paste0("q_lag", k)]] <- lag_vec(qv, k)
  }

  for (k in 1:max_k) {
    out[[paste0("m_lag", k)]] <- lag_vec(mv, k)
  }

  out
}

get_xreg_pair <- function(df, cols, y_dates_ym, target_date_ym) {

  idx_train <- match(ym_str(y_dates_ym), df$Date)
  idx_future <- match(ym_str(target_date_ym), df$Date)

  if (any(is.na(idx_train)) || is.na(idx_future)) {
    stop("설명변수 날짜 정렬을 확인하십시오.")
  }
}

```

```

x_train <- as.matrix(df[idx_train, cols, drop = FALSE])
x_future <- matrix(as.numeric(df[idx_future, cols, drop = FALSE]), nrow = 1)

colnames(x_train) <- cols
colnames(x_future) <- cols

list(x_train = x_train, x_future = x_future)
}

# =====
# 4. 원자료 파일 인식 및 불러오기
# =====

script_dir <- get_script_dir()
input_file <- file.path(script_dir, input_excel_name)

if (!file.exists(input_file)) {

  xlsx_files <- list.files(
    path = script_dir,
    pattern = "\\.*xlsx$",
    full.names = TRUE
  )

  xlsx_files <- xlsx_files[!grep("^~\\$", basename(xlsx_files))]

  if (length(xlsx_files) == 1) {
    input_file <- xlsx_files[1]
    message(
      "\n[알림] 지정한 파일명은 찾지 못했지만, 같은 폴더의 엑셀 파일을 사용합니다.\n",
      basename(input_file), "\n"
    )
  } else {
    stop(
      paste0(
        "\n엑셀 파일을 찾을 수 없습니다.\n\n",
        "방법 1) 본 R 코드 파일과 같은 폴더에 다음 엑셀 파일을 넣어주세요:\n",
        input_excel_name, "\n\n",
        "방법 2) 같은 폴더에 원자료 엑셀 파일을 하나만 남겨주세요.\n\n",
        "현재 R 코드가 인식한 폴더:\n",
        script_dir, "\n\n",
        "현재 같은 폴더에서 발견된 xlsx 파일 수: ",
        length(xlsx_files), "\n"
      )
    )
  }
}
}

```

```

sheet_names <- readxl::excel_sheets(input_file)

if (!input_sheet %in% sheet_names) {
  if ("raw_data" %in% sheet_names) {
    input_sheet <- "raw_data"
    message("\n[알림] raw_data 시트를 사용합니다.\n")
  } else if ("Raw_Data" %in% sheet_names) {
    input_sheet <- "Raw_Data"
    message("\n[알림] Raw_Data 시트를 사용합니다.\n")
  } else if ("RAW_DATA" %in% sheet_names) {
    input_sheet <- "RAW_DATA"
    message("\n[알림] RAW_DATA 시트를 사용합니다.\n")
  } else if (length(sheet_names) == 1) {
    input_sheet <- sheet_names[1]
    message("\n[알림] 첫 번째 시트를 사용합니다: ", input_sheet, "\n")
  } else {
    stop(
      paste0(
        "\n원자료 시트를 찾을 수 없습니다.\n",
        "지정한 시트명: ", input_sheet, "\n",
        "엑셀 파일 내 시트명: ", paste(sheet_names, collapse = ", "), "\n"
      )
    )
  }
}

raw0 <- readxl::read_excel(input_file, sheet = input_sheet)
names(raw0) <- trimws(names(raw0))

if ("time" %in% names(raw0) && !"date" %in% names(raw0)) {
  names(raw0)[names(raw0) == "time"] <- "date"
}

if ("quantity" %in% names(raw0) && !"in_qty" %in% names(raw0)) {
  names(raw0)[names(raw0) == "quantity"] <- "in_qty"
}

if ("import" %in% names(raw0) && !"kimchi_imp" %in% names(raw0)) {
  names(raw0)[names(raw0) == "import"] <- "kimchi_imp"
}

required_cols <- c("date", "price", "in_qty", "kimchi_imp")
missing_cols <- setdiff(required_cols, names(raw0))

if (length(missing_cols) > 0) {
  stop(
    paste0(
      "엑셀 파일에 다음 열이 없습니다: ",

```

```

paste(missing_cols, collapse = ", "),
"\n\n현재 엑셀 열 이름: ",
paste(names(raw0), collapse = ", "),
"\n\n필요 열 이름: date, price, in_qty, kimchi_imp",
"\n또는 대체 열 이름: time, price, quantity, import"
)
)
}

raw <- raw0 %>%
  mutate(
    date = parse_ym(date),
    price = as_num(price),
    in_qty = as_num(in_qty),
    kimchi_imp = as_num(kimchi_imp)
  ) %>%
  filter(!(is.na(price) & is.na(in_qty) & is.na(kimchi_imp))) %>%
  arrange(date)

if (nrow(raw) == 0) {
  stop("원자료 시트에 실제 원자료가 입력되어 있지 않습니다.")
}

if (any(is.na(raw$date))) {
  bad_rows <- which(is.na(raw$date))
  stop(
    paste0(
      "date 열에 인식할 수 없는 값이 있습니다.\n",
      "확인 필요 행 번호: ",
      paste(bad_rows + 1, collapse = ", "),
      "\n권장 날짜 형식: 2001-01 또는 2001-01-01"
    )
  )
}

if (any(duplicated(raw$date))) {
  dup_dates <- raw$date[duplicated(raw$date)]
  stop(
    paste0(
      "date 열에 중복된 기준월이 있습니다: ",
      paste(format(dup_dates, "%Y-%m"), collapse = ", "),
      "\n월별 자료는 한 행만 입력하십시오."
    )
  )
}

```

```

# =====
# 5. 예측 대상월 및 자료 범위 점검
# =====

target_date <- as.Date(paste0(target_ym, "-01"))
target_y <- as.yearmon(target_date)

last_required_y <- target_y - 1 / 12
last_required_date <- as.Date(last_required_y)

last_data_date <- max(raw$date[
  !is.na(raw$price) &
  !is.na(raw$in_qty) &
  !is.na(raw$kimchi_imp)
])

last_data_y <- as.yearmon(last_data_date)

if (last_data_y < last_required_y) {
  stop(paste0(
    "자료가 부족합니다. ", target_ym, " 가격을 예측하려면 최소 ",
    ym_str(last_required_y), "까지의 실제 price, in_qty, kimchi_imp 자료가 필요합니다."
  ))
}

full_dates <- seq.Date(from = min(raw$date), to = target_date, by = "month")

full_df <- data.frame(date = full_dates) %>%
  left_join(raw %>% select(date, price, in_qty, kimchi_imp), by = "date") %>%
  arrange(date)

check_df <- full_df %>%
  filter(date <= last_required_date)

if (any(is.na(check_df$price) | is.na(check_df$in_qty) | is.na(check_df$kimchi_imp))) {

  bad_months <- check_df$date[
    is.na(check_df$price) |
    is.na(check_df$in_qty) |
    is.na(check_df$kimchi_imp)
  ]

  stop(paste0(
    "예측 대상월 직전 월까지 결측이 있습니다. 확인 필요 월: ",
    paste(format(bad_months, "%Y-%m"), collapse = ", ")
  ))
}

```

```

start_c <- c(
  as.integer(format(min(full_df$date), "%Y")),
  as.integer(format(min(full_df$date), "%m"))
)

price_all.ts <- ts(full_df$price, start = start_c, frequency = 12)
in_qty_all.ts <- ts(full_df$in_qty, start = start_c, frequency = 12)
kimchi_imp_all.ts <- ts(full_df$kimchi_imp, start = start_c, frequency = 12)

# =====
# 6. 1개월 앞 예측 실행
# =====

end_ym <- target_y - 1 / 12
start_ym <- target_y - L / 12

y_hist <- window(price_all.ts, start = ym_to_c(start_ym), end = ym_to_c(end_ym))
log_hist <- log(y_hist)

if (length(y_hist) < L) {
  stop(paste0("분석에 필요한 과거 관측치가 부족합니다. 최소 ", L, "개월 자료가 필요합니다."))
}

price_feat_df <- make_price_features(price_all.ts)
qm_df <- make_qm_features(in_qty_all.ts, kimchi_imp_all.ts, max_k = 3)

comp <- get_stl_components(log_hist, stl_fixed)

sa_hist <- comp$sa_ts
seasonal_hist <- comp$seasonal_ts
s_future <- season_fc1_snaive(seasonal_hist)
sa_dates <- ts_to_yearmon(sa_hist)

# -----
# 6-1. STL + ARIMA
# -----

fit_arima <- Arima(
  sa_hist,
  order = sa_arima_order,
  seasonal = list(order = sa_arima_seasonal, period = 12),
  include.mean = FALSE,
  include.constant = FALSE
)

sa_fc_arima <- as.numeric(forecast(fit_arima, h = 1)$mean[1])
fc_stl_arima <- inv_log_biasadj(
  sa_fc_arima + s_future,

```

```

    get_resid_var(fit_arima)
  )

# -----
# 6-2. STL + ETS
# -----

fit_ets <- ets(sa_hist, model = sa_ets_model)

sa_fc_ets <- as.numeric(forecast(fit_ets, h = 1)$mean[1])
fc_stl_ets <- inv_log_biasadj(
  sa_fc_ets + s_future,
  get_resid_var(fit_ets)
)

# -----
# 6-3. STL + NNETAR
# -----

set.seed(set_seed)

fit_nn <- nnetar(
  sa_hist,
  p = nn_p,
  P = nn_P,
  size = nn_size,
  repeats = nn_repeats
)

sa_fc_nn <- as.numeric(forecast(fit_nn, h = 1)$mean[1])
fc_stl_nn <- inv_log_biasadj(
  sa_fc_nn + s_future,
  get_resid_var(fit_nn)
)

# -----
# 6-4. STL + NNETARx
# -----

pf_cols <- c("dp12_lag1", "level_rel_12_ma3", "crop_rel_level")

xr_pf <- get_xreg_pair(
  df = price_feat_df,
  cols = pf_cols,
  y_dates_ym = sa_dates,
  target_date_ym = target_y
)

```

```

aligned_pf <- align_y_xreg_lastok(sa_hist, xr_pf$x_train)

fc_stl_nnetarx <- NA_real_

if (isTRUE(aligned_pf$last_ok) && all(is.finite(xr_pf$x_future))) {

  zs_pf <- zscore_cont_only(
    x_train = aligned_pf$x,
    x_future = xr_pf$x_future,
    cont_cols = pf_cols
  )

  set.seed(set_seed)

  fit_nnx <- nnetar(
    aligned_pf$y,
    xreg = zs_pf$x_train,
    p = nn_p,
    P = nn_P,
    size = nn_size,
    repeats = nn_repeats
  )

  sa_fc_nnx <- as.numeric(
    forecast(
      fit_nnx,
      h = 1,
      xreg = zs_pf$x_future
    )$mean[1]
  )

  fc_stl_nnetarx <- inv_log_biasadj(
    sa_fc_nnx + s_future,
    get_resid_var(fit_nnx)
  )
}

# -----
# 6-5. STL + ARIMAX
# -----

qm_cols <- c(
  "q_lag1", "q_lag2", "q_lag3",
  "m_lag1", "m_lag2", "m_lag3"
)

xr_qm <- get_xreg_pair(
  df = qm_df,

```

```

cols = qm_cols,
y_dates_ym = sa_dates,
target_date_ym = target_y
)

aligned_qm <- align_y_xreg_lastok(sa_hist, xr_qm$x_train)

fc_stl_arimax <- NA_real_

if (isTRUE(aligned_qm$last_ok) && all(is.finite(xr_qm$x_future))) {

  fit_arimax <- Arima(
    aligned_qm$y,
    order = sa_arimax_order,
    seasonal = list(order = sa_arimax_seasonal, period = 12),
    xreg = aligned_qm$x,
    include.mean = FALSE,
    include.constant = FALSE
  )

  sa_fc_arimax <- as.numeric(
    forecast(
      fit_arimax,
      h = 1,
      xreg = xr_qm$x_future
    )$mean[1]
  )

  fc_stl_arimax <- inv_log_biasadj(
    sa_fc_arimax + s_future,
    get_resid_var(fit_arimax)
  )
}

# =====
# 7. 예측조합 및 결과 정리
# =====

model_forecasts <- c(
  STL_ARIMA = fc_stl_arima,
  STL_ETS = fc_stl_ets,
  STL_NNETAR = fc_stl_nn,
  STL_NNETARx = fc_stl_nnetarx,
  STL_ARIMAX = fc_stl_arimax
)

# 7-1. 알고리즘 단순 평균 조합
ensemble_mean <- mean(model_forecasts, na.rm = TRUE)

```

```

# 7-2. 알고리즘 예측성과 기반 조합
if (!all(names(performance_weights) %in% names(model_forecasts))) {
  stop("performance_weights의 이름이 model_forecasts와 일치하지 않습니다.")
}

if (abs(sum(performance_weights) - 1) > 1e-6) {
  stop("performance_weights의 합계가 1이 아닙니다.")
}

ensemble_performance <- sum(
  model_forecasts[names(performance_weights)] * performance_weights,
  na.rm = TRUE
)

# 7-3. KREI 전망치와 알고리즘 조합 예측치 결합
has_krei <- is.finite(krei_forecast)

krei_algorithm_combination <- NA_real_

if (has_krei) {

  if (abs(sum(krei_algorithm_weights) - 1) > 1e-6) {
    stop("krei_algorithm_weights의 합계가 1이 아닙니다.")
  }

  krei_algorithm_combination <-
    krei_algorithm_weights["ALGORITHM"] * ensemble_performance +
    krei_algorithm_weights["KREI"] * krei_forecast
}

# 7-4. 최종 권장 예측치 선택
if (final_recommendation_type == "krei_algorithm") {

  if (has_krei) {
    final_value <- krei_algorithm_combination
    final_name <- "KREI_ALGORITHM_COMBINATION"
    final_desc <- "KREI 전망치 + 알고리즘 예측성과 기반 조합"
  } else {
    final_value <- ensemble_performance
    final_name <- "ENSEMBLE_PERFORMANCE"
    final_desc <- "예측성과 기반 조합(KREI 전망치 미입력)"
  }

} else if (final_recommendation_type == "performance") {

  final_value <- ensemble_performance
  final_name <- "ENSEMBLE_PERFORMANCE"
  final_desc <- "알고리즘 예측성과 기반 조합"
}

```

```

} else if (final_recommendation_type == "mean") {

  final_value <- ensemble_mean
  final_name <- "ENSEMBLE_MEAN"
  final_desc <- "알고리즘 단순 평균 조합"

} else if (final_recommendation_type == "krei") {

  if (!has_krei) {
    stop("final_recommendation_type이 'krei'이지만 kre_forecast가 입력되지 않았습니다.")
  }

  final_value <- kre_forecast
  final_name <- "KREI_FORECAST"
  final_desc <- "KREI 농업관측센터 전망치"

} else {
  stop("final_recommendation_type은 'krei_algorithm', 'performance', 'mean', 'krei' 중 하나여야 합니다.")
}

forecast_result <- data.frame(
  target_month = target_ym,
  forecast_type = "one-step-ahead",
  model = c(
    names(model_forecasts),
    "ENSEMBLE_MEAN",
    "ENSEMBLE_PERFORMANCE",
    "KREI_FORECAST",
    "KREI_ALGORITHM_COMBINATION",
    "FINAL_RECOMMENDED"
  ),
  forecast_price = c(
    as.numeric(model_forecasts),
    ensemble_mean,
    ensemble_performance,
    ifelse(has_krei, kre_forecast, NA_real_),
    kre_algorithm_combination,
    final_value
  ),
  note = c(
    rep("개별모형 예측치", length(model_forecasts)),
    "개별모형 단순 평균 조합",
    "검증구간 MSE 기반 예측성과 조합",
    ifelse(has_krei, "실무자 입력 KREI 전망치", "KREI 전망치 미입력"),
    ifelse(has_krei, "KREI 전망치와 알고리즘 예측치의 선형 조합", "KREI 전망치 미입력으로 미산출"),
    "최종 권장 예측치"
  ),

```

```

row.names = NULL
)

# =====
# 8. 실무자용 간략 예측 보고서 출력
# =====

model_structure_text <- c(
  paste0("① STL_ARIMA : STL 계절성 + ARIMA(",
    paste(sa_arima_order, collapse = ","),
    ")",
    paste(sa_arima_seasonal, collapse = ","),
    ")[12]"),
  "② STL_ETS : STL 계절성 + ETS(A,N,N)",
  paste0("③ STL_NNETAR : STL 계절성 + NNETAR(p=", nn_p,
    ", P=", nn_P,
    ", size=", nn_size,
    ", repeats=", nn_repeats, ")"),
  paste0("④ STL_NNETARx : STL 계절성 + NNETARx(p=", nn_p,
    ", P=", nn_P,
    ", size=", nn_size,
    ", repeats=", nn_repeats,
    ", xreg=가격 파생변수)"),
  paste0("⑤ STL_ARIMAX : STL 계절성 + ARIMAX(",
    paste(sa_arimax_order, collapse = ","),
    ")",
    paste(sa_arimax_seasonal, collapse = ","),
    ")[12], xreg=반입량 김치수입량 시차변수")
)

forecast_lines <- paste0(
  sprintf("%-12s", names(model_forecasts)),
  " : ",
  format_price(as.numeric(model_forecasts))
)

ensemble_lines <- c(
  paste0("알고리즘 단순 평균 조합 : ", format_price(ensemble_mean)),
  paste0("알고리즘 예측성과 기반 조합 : ", format_price(ensemble_performance))
)

if (has_krei) {
  krei_lines <- c(
    paste0("KREI 농업관측센터 전망치 : ", format_price(krei_forecast)),
    paste0("KREI+알고리즘 조합 예측치 : ", format_price(krei_algorithm_combination)),
    paste0(
      "조합 가중치 : 알고리즘 ",
      round(krei_algorithm_weights["ALGORITHM"], 3),
    )
  )
}

```

```

    ", KREI ",
    round(krei_algorithm_weights["KREI"], 3)
  )
} else {
  krei_lines <- c(
    "KREI 농업관측센터 전망치   : 미입력",
    "KREI+알고리즘 조합 예측치   : 미산출",
    "참고                        : KREI 전망치를 입력하면 KREI+알고리즘 조합 예측치가 산출됨"
  )
}

cat("\014")

cat(
  "\n",
  "===== \n",
  "      배추 도매가격 1개월 앞 예측 보고서\n",
  "===== \n\n",

  "■ 1. 실무자 입력 정보\n",
  "----- \n",
  "예측 대상월       : ", target_ym, "\n",
  "KREI 전망치       : ",
  ifelse(has_krei, format_price(krei_forecast), "미입력"), "\n",
  "최종 권장 방식    : ", final_desc, "\n\n",

  "■ 2. 예측 개요\n",
  "----- \n",
  "예측 방식         : 1개월 앞 예측(one-step-ahead forecast)\n",
  "모형 추정 사용기간 : ", ym_str(start_ym), " ~ ", ym_str(end_ym), "\n",
  "사용 관측치 수     : ", length(y_hist), "개월\n",
  "입력 엑셀 파일     : ", basename(input_file), "\n\n",

  "■ 3. 개별 예측모형 구조\n",
  "----- \n",
  paste(model_structure_text, collapse = "\n"), "\n\n",

  "■ 4. 개별 예측모형별 예측치\n",
  "----- \n",
  paste(forecast_lines, collapse = "\n"), "\n\n",

  "■ 5. 알고리즘 앙상블 예측치\n",
  "----- \n",
  paste(ensemble_lines, collapse = "\n"), "\n\n",

  "■ 6. KREI 전망치 및 KREI+알고리즘 조합\n",
  "----- \n",

```

```

paste(krei_lines, collapse = "\n"), "\n\n",

"■ 7. 최종 권장 예측치\n",
"-----\n",
"최종 예측모형      : ", final_name, "\n",
"최종 예측가격      : ", format_price(final_value), "\n\n",

"■ 8. 해석\n",
"-----\n",
target_ym, " 배추 도매가격의 최종 권장 예측치는 ",
format_price(final_value), "입니다. ",
"본 예측치는 ", ym_str(start_ym), "부터 ", ym_str(end_ym),
"까지의 최근 ", length(y_hist),
"개월 자료를 이용하여 산출되었습니다. ",
ifelse(
  has_krei && final_recommendation_type == "krei_algorithm",
  "최종 권장값은 KREI 전망치와 알고리즘 예측성과 기반 조합 예측치를 선형 결합한 결과입니다.",
  "최종 권장값은 선택된 알고리즘 또는 KREI 기준에 따라 산출되었습니다."
),
"\n",
"-----\n",
sep = ""
)

```

실무용 R 코드 재현을 위한 통합 프롬프트

- 본 연구에서는 배추 도매가격 예측 알고리즘의 실무 적용 가능성을 높이기 위해, 연구 결과를 기반으로 한 실무용 R 코드를 부록에 제시하였음.
- 최근 생성형 AI 도구를 활용한 코드 작성 및 업무 자동화가 확대되고 있는 점을 고려하여, 본 부록에서는 실무용 R 코드 작성 과정에서 활용할 수 있는 프롬프트를 함께 제시함. 이는 코드 자체의 재현성뿐만 아니라, 코드 작성 요구 조건을 명확히 구조화하는 방법을 제시한다는 점에서 의의가 있음.
- 다만 본 부록의 프롬프트는 코드 작성을 보조하기 위한 예시이며, 생성형 AI 가 산출한 코드를 그대로 사용하는 것을 의미하지 않음. 최종 코드는 연구자가 분석 목적, 자료 구조, 모형 설정, 예측 절차 및 출력 결과를 검토하여 수정·보완한 후 사용해야 함. 특히 예측모형의 구조, 조합 가중치, KREI 전망치와의 결합 방식 등은 본 연구의 분석 결과와 실무 적용 목적에 맞게 연구자가 최종적으로 확정해야 함.

〈실무용 R 코드 재현을 위한 통합 프롬프트〉

다음 조건을 만족하는 배추 도매가격 1개월 앞 예측용 R 코드를 작성하라.

1. 코드의 목적

- 월별 배추 도매가격을 대상으로 1개월 앞 예측치를 산출한다.
- 실무자가 매월 반복적으로 사용할 수 있도록 코드 구조를 단순하고 명확하게 구성한다.

- 최종 출력은 콘솔에 간략 예측 보고서 형태로 제시한다.

2. 입력자료 구조

- 원자료는 엑셀 파일로 제공되며, R 코드 파일과 같은 폴더에 저장되어 있다고 가정한다.
- 엑셀 파일명은 "cabbage_raw_data_template.xlsx"로 설정한다.
- 원자료 시트명은 "raw_data"로 설정한다.
- 원자료에는 다음 열이 포함되어야 한다.
 - ① date: 기준월
 - ② price: 배추 도매가격
 - ③ in_qty: 배추 반입량
 - ④ kimchi_imp: 김치 수입량
- 날짜 형식은 "YYYY-MM" 또는 "YYYY-MM-DD"를 인식할 수 있도록 처리한다.
- 숫자 변수는 쉼표가 포함되어 있어도 숫자형으로 변환할 수 있도록 처리한다.

3. 실무자 입력부

- 코드 상단에 실무자가 매월 수정해야 하는 입력부를 별도로 둔다.
- 실무자 입력부에는 다음 세 가지 항목을 포함한다.
 - ① target_ym: 예측 대상 월, 예: "2026-05"
 - ② krei_forecast: KREI 농업관측센터 전망치, 예: 6000
 - ③ final_recommendation_type: 최종 권장 예측치 선택 방식
- final_recommendation_type은 다음 선택지를 허용한다.
 - ① "krei_algorithm": KREI 전망치와 알고리즘 예측치를 결합
 - ② "performance": 알고리즘 예측성과 기반 조합 예측치 사용
 - ③ "mean": 알고리즘 단순 평균 조합 예측치 사용
 - ④ "krei": KREI 전망치 사용

4. 자료 점검

- 예측 대상 월의 직전 월까지 price, in_qty, kimchi_imp 자료가 모두 존재하는지 확인한다.
- date 열에 중복 월이 있으면 오류 메시지를 출력한다.
- 필수 열이 없거나 결측이 있으면 실무자가 이해할 수 있는 한글 오류 메시지를 출력한다.
- 예측에 필요한 최소 학습자료 길이는 216개월로 설정한다.

5. 모형 설정

- 예측에는 최근 216개월의 고정 길이 rolling window 학습자료를 사용한다.
- 가격 변수는 로그 변환하여 모형을 추정한다.
- STL 분해는 로그 가격 시계열에 적용한다.
- STL 설정은 다음과 같이 고정한다.

① s.window = "periodic"

② t.window = 19

③ l.window = 13

④ inner = 2

⑤ outer = 0

- STL 분해 후 계절 성분은 SNAIVE 방식으로 1개월 앞 예측한다.
- 계절조정 성분에는 다음 모형을 적용한다.

① STL+ARIMA

② STL+ETS

③ STL+NNETAR

④ STL+NNETARx

⑤ STL+ARIMAX

6. 개별모형 구조

- STL+ARIMA는 ARIMA(1,1,2)(2,0,1)[12] 구조를 사용한다.
- STL+ETS는 ETS(A,N,N) 구조를 사용한다.
- STL+NNETAR는 $p=3$, $P=1$, size=2, repeats=20으로 설정한다.
- STL+NNETARx는 가격 기반 특성 변수를 외생변수로 포함한다.
- 가격 기반 특성 변수는 전년 동월 대비 변화, 최근 3개월 평균의 전년 대비 차이, 작형별 과거 평균 대비 차이 등을 활용한다.
- STL+ARIMAX는 ARIMAX(0,1,5)(2,0,0)[12] 구조를 사용한다.
- ARIMAX의 외생변수는 반입량과 김치 수입량의 시차변수로 구성한다.

7. 예측값 환산

- 모든 모형은 로그 가격 기준으로 예측한다.
- 원 단위 가격으로 환산할 때 단순 역변환 $\exp(\hat{y})$ 을 사용하지 말고, 로그 예측오차 분산을 반영한 보정 역변환을 적용한다.
- 보정 역변환은 $\exp(\hat{y} + \sigma^2/2)$ 형태로 계산한다.

8. 예측조합

- 개별모형 예측치는 다음 방식으로 조합한다.

① 단순 평균 조합

② 검증구간 MSE 기반 예측성과 조합

- MSE 기반 조합 가중치는 코드 상단에서 벡터 형태로 수정할 수 있도록 작성한다.
- 가중치의 합이 1이 아니면 오류 메시지를 출력한다.

9. KREI 전망치와 알고리즘 예측치 결합

- KREI 전망치가 입력된 경우, KREI 전망치와 알고리즘 예측성과 기반 조합 예측치를 선형 결합한다.
- 결합 가중치는 코드 상단에서 수정할 수 있도록 작성한다.
- 기본 예시는 알고리즘 0.347, KREI 0.653으로 설정한다.
- KREI 전망치가 입력되지 않은 경우에는 알고리즘 예측치만 산출되도록 한다.

10. 출력

- 코드 실행이 완료되면 콘솔에 간략 예측 보고서를 출력한다.
- 보고서에는 다음 항목을 포함한다.

① 예측 대상 월

② KREI 전망치 입력 여부

③ 모형 추정 사용기간

④ 사용 관측치 수

⑤ 개별모형별 예측치

⑥ 알고리즘 단순 평균 조합 예측치

⑦ 알고리즘 예측성과 기반 조합 예측치

⑧ KREI+알고리즘 조합 예측치

⑨ 최종 권장 예측치

- 가격은 원/10kg 단위로 표시하고, 천 단위 구분 기호를 적용한다.

11. 코드 작성 방식

- 실무자가 이해할 수 있도록 코드에 한글 주석을 충분히 포함한다.
- 주요 단계별로 구분선을 넣어 코드 구조를 명확히 한다.
- 패키지 설치 및 로드, 자료 불러오기, 자료 점검, 변수 생성, 모형 추정, 예측조합, 결과 출력 순서로 작성한다.
- 오류 메시지는 가능한 한 실무자가 원인을 파악할 수 있도록 구체적으로 작성한다.

- 강동곤·장영민·이주석·이성수(2024), “장단기 기억 신경망을 사용한 다변수 데이터 농산물 가격 예측 모델”, 전기전자학회논문지, 28(3): 451-457.
- 김배성(2005), “채소가격 예측을 위한 응용기법별 예측력 비교”, 농업경제연구, 46(4): 89-113.
- 박소연·김동연·김민수(2022), “생육시기 기상을 활용한 배추 가격 예측”, Journal of The Korean Data Analysis Society, 24(5): 1667-1679.
- 박종현·임영우·임도현·최윤성·안현철(2023), “증장기 농산물 가격 예측을 위한 다단계 시계열 예측 모델”, 한국컴퓨터정보학회논문지, 28(2): 201-207.
- 박지연·박영구(2013), 배추·무 예측모형 고도화 방안, 한국농촌경제연구원 기타연구보고: 1-80.
- 박찬·이경순(2023), “농산물 가격의 특이치 탐지 및 처리를 통한 딥러닝 기반 가격 예측”, 디지털콘텐츠학회논문지, 24(8): 1899-1906.
- 서교·이정재·김태곤(2008), “신뢰성 해석기법을 이용한 배추 가격폭등 예측 모형의 개발”, 한국농공학회논문집, 50(3): 91-104.
- 유도일(2016), “기상변인을 고려한 채소가격 예측모형 개발”, 농업경제연구, 57(1): 1-24.
- 이석일·김선웅·윤병삼(2015), “육계가격 예측모형의 예측력 비교 평가”, 농업경영·정책연구, 42(2): 232-248.
- 이창민·송성광·정성욱(2022), “양상불 Voting 기법을 활용한 배추 가격 예측에 관한 연구”, 융합정보논문지, 12(3): 1-10.
- 장수희·전희주·조인호·김동환(2017), “비정형 농업기상자료를 활용한 배추 도매가격 예측모형 연구”, 한국데이터정보과학회지, 28(3): 617-624.
- 정호중·정원호(2016), “배추수급모형 추정을 통한 가격변동 시나리오 분석”, 농업경영·정책연구, 43(4): 897-920.
- 한석호·김병률(2004), “시장개방하의 배추 수급 모형과 전망”, 농촌경제, 27(3): 35-54.
- 한은수·김선웅·윤병삼(2025), “예측조합을 활용한 양념채소 가격 예측 및 예측성과 분석”, 농촌경제, 48(3): 1-31.
- 허다솜·정승권(2024), “고랭지배추의 생육 시기별 주산지 기상정보를 활용한 출하시기 도매가격 예측모형”, 한국정보기술학회논문지, 22(9): 11-22.

- Bates, J. M. & C. W. J. Granger(1969), "The Combination of Forecasts", *Journal of the Operational Research Society*, 20(4): 451-468.
- Clemen, R. T.(1989), "Combining Forecasts: A Review and Annotated Bibliography", *International Journal of Forecasting*, 5(4): 559-583.
- Colino, E. V., S. H. Irwin & P. Garcia(2012), "Composite and Outlook Forecast Accuracy", *American Journal of Agricultural Economics*, 94(2): 514-520.
- Crespo Cuaresma, J., J. Hlouskova & M. Obersteiner(2021), "Agricultural Commodity Price Dynamics and Their Determinants: An Econometric Approach", *Journal of Forecasting*, 40(6): 1041-1057.
- Diebold, F. X. & R. S. Mariano(1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13(3): 253-263.
- Fang, Y., B. Guan, S. Wu & S. Heravi(2020), "Optimal Forecast Combination Based on Ensemble Empirical Mode Decomposition for Agricultural Commodity Futures Prices", *Journal of Forecasting*, 39(6): 877-886.
- Fu, L., X. Ding & Y. Ding(2022), "Ensemble Empirical Mode Decomposition-Based Preprocessing Method with Multi-LSTM for Time Series Forecasting: A Case Study for Hog Prices", *Connection Science*, 34(1): 2177-2200.
- Gil, J. M. & L. Albu(1993), "Composite Forecasting Methods: An Application to Spanish Maize Prices", *Journal of Agricultural Economics*, 44(2): 264-271.
- Harvey, D., S. Leybourne & P. Newbold(1997), "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, 13(2): 281-291.
- Harvey, D., S. Leybourne & P. Newbold(1998), "Tests for Forecast Encompassing", *Journal of Business & Economic Statistics*, 16(2): 254-259.
- Heydari, R.(2023), "A Flexible Combination Forecasting Method for Modeling Agricultural Commodity Prices: A Case Study Iran's Livestock and Poultry Meat Market", *Journal of Agricultural Economics and Development*, 37(2): 177-202.
- Hyndman, R. J. & G. Athanasopoulos(2021), *Forecasting: Principles and Practice*, 3rd ed., OTexts.

- Jaiswal, R., G. K. Jha, K. Choudhary & R. R. Kumar(2022a), “STL Decomposition Based LSTM Model for Seasonal Agricultural Price Forecasting”.
- Jaiswal, R., K. Choudhary & R. R. Kumar(2022b), “STL-ELM: A Decomposition-Based Hybrid Model for Price Forecasting of Agricultural Commodities”, *National Academy Science Letters*, 45(6): 477-480.
- Jin, D., H. Yin, Y. Gu & S. J. Yoo(2019), “Forecasting of Vegetable Prices Using STL-LSTM Method”, in 2019 6th International Conference on Systems and Informatics, 866-871.
- Ramsey, A. F. & M. K. Adjemian(2024), “Forecast Combination in Agricultural Economics: Past, Present, and the Future”, *Applied Economic Perspectives and Policy*, 46(4): 1450-1478.
- Stock, J. H. & M. W. Watson(1999), “A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series”, in R. F. Engle and H. White eds., *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, Oxford University Press.
- Sun, F., X. Meng, Y. Zhang, Y. Wang, H. Jiang & P. Liu(2023), “Agricultural Product Price Forecasting Methods: A Review”, *Agriculture*, 13(9): 1671.
- Vere, D. T. & G. R. Griffith(1990), “Comparative Forecast Accuracy in the New South Wales Prime Lamb Market”, *Australian Journal of Agricultural Economics*, 34(2): 103-117.
- Wang, X., R. J. Hyndman, F. Li & Y. Kang(2022), “Forecast Combinations: An Over 50-Year Review”, *International Journal of Forecasting*, 39(4): 1518-1547.
- Xiong, T., C. Li & Y. Bao(2018), “Seasonal Forecasting of Agricultural Commodity Price Using a Hybrid STL and ELM Method: Evidence from the Vegetable Market in China”, *Neurocomputing*, 275: 2831-2844.
- Xu, X.(2017), “Short-Run Price Forecast Performance of Individual and Composite Models for 496 Corn Cash Markets”, *Journal of Applied Statistics*, 44(14): 2593-2620.
- _____(2020), “Corn Cash Price Forecasting”, *American Journal of Agricultural Economics*, 102(4): 1297-1320.

- Yin, H. L., D. Jin, Y. H. Gu, C. J. Park, S. K. Han & S. J. Yoo(2020), “STL-ATTLLSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM”, *Agriculture*, 10(12): 612.
- Zhang, T. & Z. Tang(2024), “Agricultural Commodity Futures Price Prediction Based on a New Hybrid Forecasting Model Combining Quadratic Decomposition Technology and LSTM Model”, *Frontiers in Sustainable Food Systems*, 8: 1334098.
- Zhu, H., R. Xu & H. Deng(2022), “A Novel STL-Based Hybrid Model for Forecasting Hog Price in China”, *Computers and Electronics in Agriculture*, 198: 107068.

〈온라인 자료〉

서울특별시농수산물공사 유통정보 홈페이지(<https://www.garak.co.kr/youtong/index.do>), 검색일: 2026. 2. 5.

〈AI 자료〉

OpenAI(2025), ChatGPT(GPT-5, 2026. 5. 2.), 대형언어모델(LLM), <https://chatgpt.com>
본 보고서의 도식화를 위해 AI 도구를 보조적으로 활용하였으며, 연구자의 검토를 통해 일부 수정하여 본 보고서에 수록함.

KREI

www.krei.re.kr



배추 가격 예측력 개선 연구

한국농촌경제연구원

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



9 791161 498485
ISBN 979-11-6149-848-5 (PDF)