

A Multi-step Time Series Forecasting Model for Mid-to-Long Term Agricultural Price Prediction

Jonghyun Park*, Yeong-Woo Lim*, Do Hyun Lim*, Yunsung Choi**, Hyunchul Ahn*

*Master's Candidate, Graduate School of Business IT, Kookmin University, Seoul, Korea

*Ph.D. Candidate, Graduate School of Business IT, Kookmin University, Seoul, Korea

*Master's Candidate, Graduate School of Business IT, Kookmin University, Seoul, Korea

**Research director, ONTWOIN AILAB, ONTWOIN Co.,Ltd., Seoul, Korea

*Professor, Graduate School of Business IT, Kookmin University, Seoul, Korea

[Abstract]

In this paper, we propose an optimal model for mid to long-term price prediction of agricultural products using LGBM, MLP, LSTM, and GRU to compare and analyze the three strategies of the Multi-Step Time Series. The proposed model is designed to find the optimal combination between the models by selecting methods from various angles. Prior agricultural product price prediction studies have mainly adopted traditional econometric models such as ARIMA and LSTM-type models. In contrast, agricultural product price prediction studies related to Multi-Step Time Series were minimal. In this study, the experiment was conducted by dividing it into two periods according to the degree of volatility of agricultural product prices. As a result of the mid-to-long-term price prediction of three strategies, namely direct, hybrid, and multiple outputs, the hybrid approach showed relatively superior performance. This study academically and practically contributes to mid-to-long term daily price prediction by proposing an effective alternative.

▶ **Key words:** Multi-Step Time Series Forecasting, Agricultural Product Price, Mid-to-long-term Prediction, Daily Price Prediction, Hybrid Strategy

[요 약]

본 논문에서는 Multi-Step Time Series의 세 가지 전략을 비교 분석하기 위해 LGBM, MLP, LSTM, GRU를 사용하여 농산물 중장기 가격 예측에 대한 최적의 모형을 제안한다. 제안 모형은 다각도로 전략을 선택하여 모델과 전략간 최적의 조합을 찾도록 설계되었다. 기존 농산물 가격 예측 연구에서는 전통 계량경제 모델인 ARIMA를 비롯하여 LSTM 계열 모델이 주로 사용된 반면 Multi-Step Time Series 관련 농산물 가격 예측 연구는 매우 제한적이다. 본 연구에서는 농산물 가격의 변동성 정도에 따라 두 개의 기간으로 나누어 실험을 진행하였으며, Direct, Hybrid, Multiple Outputs 등 세 전략의 중장기 가격 예측 결과 Hybrid 접근법이 상대적으로 우수한 성능을 보였다. 본 연구 결과는 중장기 일별 가격 예측을 고도화할 수 있는 효과적인 대안을 제시한다는 측면에서 학술적, 실무적 의의를 갖는다.

▶ **주제어:** 다단계 시계열 예측, 농산물 가격, 중장기 예측, 일별 가격 예측, 혼합 전략

- First Author: Jonghyun Park, Corresponding Author: Hyunchul Ahn
- *Jonghyun Park (ppjjhh1027@kookmin.ac.kr), Graduate School of Business IT, Kookmin University
- *Yeong-Woo Lim (duddn7244@naver.com), Graduate School of Business IT, Kookmin University
- *Do Hyun Lim (ehgus7011@naver.com), Graduate School of Business IT, Kookmin University
- **Yunsung Choi (raphaelyun@ontwo.co.kr), ONTWOIN AILAB, ONTWOIN Co.,Ltd.
- *Hyunchul Ahn (hcahn@kookmin.ac.kr), Graduate School of Business IT, Kookmin University
- Received: 2022. 12. 30, Revised: 2023. 01. 30, Accepted: 2023. 01. 31.

I. Introduction

최근 급변하는 기후 및 국제 정세로 농산물의 가격 변동성이 더욱 커지고 있으며, 자연재해 등이 발생한다면 견잡을 수 없이 폭등할 수 있다[1]. 가격이 급변할 시 생산자와 소비자 모두에게 큰 영향을 미치는 관계로 농산물의 가격 안정화는 국가 차원에서 다루고 있는 중요한 문제이다. 생산자는 매년 재배량과 재배면적을 신중하게 고려해야 하는데 이는 가격과 밀접한 관련이 있다. 하지만 수확 기간에 따른 생산량 조절이 어려워 가격의 급등과 급락을 반복하게 된다. 이 밖에도 기후나 병해충, 저장이 어려운 특성, 비탄력적 소비행태 등 수요와 공급의 균형이 맞지 않는 다양한 원인이 존재한다. 안정적이지 못한 농산물 수급은 가격에 영향을 미쳐 생산자, 소비자에게 영향을 주고 농산물 거래 활성화를 어렵게 한다. 한국농촌경제연구원 발표한 보고서에 의하면 소비자가 민감하게 여기는 물가 중 1순위로 농산물 가격이 집계되었다. 이는 농산물은 구입 빈도가 높을 뿐더러, 가계 지출에서 차지하는 비율 또한 크기 때문에 심리적 부담이 크게 작용하는 것으로 해석할 수 있다[2]. 또한 외식분야 등 실생활에 소비자의 체감 물가 수준이 반영되어 소비심리와 경제 활성화에 악영향을 미칠 수 있다.

이 문제에 대응하기 위해 정부는 주의·경계·심각 3단계로 구성된 농산물 수급 조절 매뉴얼을 배포하였으며, 농업인은 특히 수입산 부정 유통 문제에 대응하기 위해 ‘우리농업지키기운동본부’를 창설하였다. 또한 가격 관측의 일환으로 한국농촌경제연구원이 설립한 ‘농업관측본부’는 각종 통계자료를 수집, 분석하여 매일 관측일보를 발간하고 있다. 한국농수산식품유통공사 aT에서는 2020년부터 경진대회를 통해 농산물 가격을 예측할 수 있는 모델을 발굴하고 있으며, 경상남도에는 “주요 농산물 가격예측 시스템”을 통해 가격예측 정보를 제공하는 등 농산물 가격 예측에 대한 사회적 수요가 높아지는 추세이다. 하지만 이러한 서비스들은 길어야 7일 후 정도의 예측 가격만 제공한다. 단기 가격 예측 정보의 경우 농산물의 출하관리에는 도움이 될 수 있다. 하지만 어떤 작물을 파종할지 결정하고자 하는 농민이나 농산물 가격에 기반한 금융서비스를 제공하고자 하는 핀테크 사업자의 경우에는 보다 중장기적인 농산물 가격예측 정보가 필요하다. 학계에서도 농산물 가격을 예측하기 위한 연구를 오래 전부터 진행해왔다. 기상 및 토지 데이터 등 농산물 재배 관련 변수를 활용한 연구는 지속적으로 수행되었고[3], 농산물의 저장성이 출하량과 가격예측에 미치는 영향을 분석하기도 했다[4]. 또한 날씨 데이터의 2개월 후 가격 변화를 활용한 연구도 수행되

었다[5]. 그러나 30일 이상의 중장기 일별 가격 예측에 관한 연구는 매우 제한적이다.

다양한 범위의 미래 시점 농산물 가격을 예측하기 위해서는 다단계 시계열 예측 접근법(Multi-step Time Series Forecasting)을 적용해야 한다. 다단계 시계열 예측 시 일반적으로 사용되는 접근법은 Scikit-Learn에 구현되어 있는 Direct Multi-step Time Series Forecasting(이하 Direct), Recursive Multi-step Time Series Forecasting(이하 Recursive) 방법이며[6], 그 중 Recursive 방법이 일반적으로 가장 많이 사용되어 왔다. 하지만 Recursive는 익일 예측값을 활용하여 이틀 뒤의 값을 예측하고 이 값을 포함하여 그 다음 날 가격을 예측하는 원리이기 때문에 중장기 즉 예측해야 하는 시점이 멀어질수록 예측 오차가 점점 커질 수 밖에 없다.

이에 본 연구에서는 중장기 농산물 가격 예측 모형을 구축하고자 Brownlee[7]가 제안한 (1) Direct, (2) Direct-Recursive Hybrid Multi-step Time Series Forecasting(이하 Hybrid), (3) Multiple Outputs의 3가지 전략을 선정하였다. 특히 본 연구에서는 Direct와 Recursive의 장점을 결합한 Hybrid 기법에 주목하여 해당 기법이 중장기 가격예측에 효과적인지 확인해 보고자 한다.

II. Theoretical Background

1. Agricultural Price Prediction

농산물 가격 예측에 대한 초기 선행 연구는 주로 전통계량경제학 모형인 ARMA(Autoregressive and Moving Average), ARIMA(Autoregressive Integrated Moving Average) 등을 활용하여 수행되었다[8-9]. 하지만 외[8]는 한국농식품유통공사의 농산물유통정보(KAMIS)에서 제공하는 데이터를 활용하여 시계열 모형의 예측값을 산출하였으며, 특히 한국농촌경제연구원이 공개한 예측치와 결합하여 보다 개선된 성능의 모델을 제시하였다. 단순히 다양한 품종을 대상으로 ARIMA 분석이 얼마나 유의성을 가지는지 비교분석한 연구[9] 외에도 전통적 통계기법을 보완하는 연구도 수행되었다. 먼저 정대호, 조영열[10]은 선형 모형인 ARIMA가 포착하지 못하는 비선형 패턴을 LSTM(Long Short-Term Memory)을 활용하여 보완하고자 하는 연구를 진행했다. 유동완, 박종범[11]은 LSTM에 Attention Mechanism을 적용하면 가격 변동성이 크지 않은 품종의 가격 예측이 효과적인 것을 확인하였다.

또한 농산물 거래 정보, 기상, 수출입, 물가지수 등을 예측 변수로 활용한 연구[10], 가격, 기상정보, 농가 소득 등 재배환경을 변수로 넣은 연구[12], 비정형 데이터인 온라인 키워드 분석을 반영한 연구[13] 등 가격 외에 다양한 변수를 입력 변인으로 사용한 연구들도 활발히 수행되었다.

이 외에도 ARIMA와 MLP(Multi-Layer Perceptron)를 비교하여 인공지능경망보다 전통적 통계기법의 예측력이 더 좋다는 결론을 보인 연구[9], 선형 회귀분석을 통해 독립변수를 선정 후 MLP를 활용하여 변수의 유효성을 검증한 연구[14], 강화학습을 활용한 연구[15] 등의 연구들이 이루어져 왔다.

2. Multi-step Time Series Forecasting

기존 시계열 예측은 보통 t시점까지의 관측치를 활용하여 t+1 시점의 예측을 수행한다. 그러나 본 연구에서는 국내 농산물 예측에 관한 연구에서 시도되지 않았던 Multi-step Time Series 예측을 시도해보고자 한다. Multi-step Time series Forecasting은 Brownlee[7]에 의해 제안된 모델로 총 4가지 전략이 존재하며, 여러 시점 후의 미래, 즉 t+n의 예측 결과를 얻을 수 있다.

2.1 Direct Multi-step Forecasting Strategy

Direct Multi-step Forecasting Strategy는 아래 제시된 수식과 같이 각 시점별 예측 모델을 독립적으로 구축하는 접근법으로 정의된다.

$$\begin{aligned} pred(t) &= model_1(obs(t-1), obs(t-2), \dots, obs(t-n)) \\ pred(t+1) &= model_2(obs(t-1), obs(t-2), \dots, obs(t-n)) \\ &\dots \\ pred(t+k) &= model_k(obs(t-1), obs(t-2), \dots, obs(t-n)) \end{aligned} \quad (1)$$

네트워크 대역폭 사용량의 예측이 연구된 바 있으며 [16], 준수한 성능을 보였다. 그러나 모델을 다수 개발해야 하므로 컴퓨팅 자원이 많이 소모되고, 각 시점별 예측 결과가 서로 독립적으로 생성되어 연계성이 떨어진다.

2.2 Recursive Multi-step Forecasting Strategy

Recursive Multi-step Forecasting Strategy는 아래 식에 표현된 것과 같이 재귀적으로 이전 단계의 예측 결과를 다음 단계 예측에 활용하는 방식으로 특정 시점의 값을 예측하는 전략을 말한다.

$$\begin{aligned} pred(t) &= model(obs(t-1), obs(t-2), \dots, obs(t-n)) \\ pred(t+1) &= model(pred(t), obs(t-1), \dots, obs(t-n)) \end{aligned} \quad (2)$$

$$\begin{aligned} &\dots \\ pred(t+k) &= model(pred(t+k-1), pred(t+k-2), \dots, obs(t-n)) \end{aligned}$$

이 접근법은 하나의 단일 모델로 여러 미래 시점의 예측 결과를 산출할 수 있다는 장점이 있다. 하지만 먼 미래를 예측, 즉 k가 커질수록 예측 오류차가 누적되므로 성과 저하가 뚜렷하게 나타난다. 본 연구에서도 파일럿 실험 결과 오류가 급증함을 확인하여 비교 전략에서 제외하였다.

2.3 (Direct-Recursive) Hybrid Strategy

2.1과 2.2 접근법의 장점을 결합한 Hybrid 전략으로 전문가들이 가장 추천하는 Multi-step 시계열 예측 방법이다. t+2 시점의 예측값을 얻고자 하면 먼저 t+1 시점의 예측값을 구하는 모델을 만든 후, 이 모델을 통해 예측된 값을 추가하여 t+2 시점의 예측값을 구하는 모델을 학습시킨 후 결과를 얻는 방식을 의미하며 식은 아래와 같다.

$$\begin{aligned} pred(t) &= model_1(obs(t-1), obs(t-2), \dots, obs(t-n)) \\ pred(t+1) &= model_2(pred(t), obs(t-1), \dots, obs(t-n)) \\ &\dots \\ pred(t+k) &= model_k(pred(t+k-1), pred(t+k-2), \dots, obs(t-n)) \end{aligned} \quad (3)$$

2.4 Multiple Outputs Strategy

Multiple Outputs Strategy는 한 번에 전체 미래 예측 값을 산출할 수 있는 전략을 의미한다. 이 전략에 적용할 수 있는 대표적인 분석기법으로 LSTM, GRU 등이 존재하며 식은 아래와 같다.

$$pred(t), \dots, pred(t+k) = model(obs(t-1), obs(t-2), \dots, obs(t-n)) \quad (4)$$

단일 모델로 원하는 시점까지의 예측값을 얻을 수 있는 반면, 모델이 과도하게 복잡해지는 단점이 있으며 학습이 오래 걸리고, 과적합을 피하기 위해서는 방대한 양의 학습 데이터가 요구된다.

이상 본 논문에서 소개한 4가지 Multi-step Forecasting 전략은 경제학 분야 연구에서 자주 활용되는 Static / Dynamic Time Series Forecasting의 관점에서 그 특징을 살펴볼 수 있다. 우선 Direct 전략의 경우 특정 미래시점의 값을 예측하게끔 설계되어 있다는 점에서 Static Forecasting과 같은 특성을 갖는다. 반면 Recursive, Hybrid 및 Multiple Outputs는 여러 미래 시점의 예측 값이 동시에 산출된다는 점에서 Dynamic Forecasting과 같은 특징을 갖는다고 할 수 있다. 하지만 본 연구에서 소개하는 Recursive, Hybrid 전략의 경우 입

력값으로 이전 시점의 예측값을 사용하도록 설계되어 있지만, Dynamic Forecasting 전략은 과거 관측치와 외생 변수를 사용하게끔 설계되어 있다는 점에서 차이가 있다.

III. Research Framework

2절에서 언급한 4가지 전략 중 파일럿 실험을 통해 오류가 증폭되는 것으로 확인된 Recursive 전략을 배제하고, 나머지 3가지 전략을 중심으로 실험을 진행하였다. Fig. 1과 같이 품종, 기간, 전략, 분석기법 등 다양한 조건을 설정/결합하여, 농산물 품종 및 시점별 최적의 모델을 살펴보고자 한다.

평가지표는 MAPE(Mean Absolute Percentage Error)를 사용한다. MAPE는 예측치와 실측치 간 오차를 비교한 것으로, 수식은 아래 (5)와 같다. 이 수치가 낮을수록 모형의 성능이 좋다고 할 수 있다. 농산품 품종별로 단가의 편차가 크므로 종합적인 비교를 위해 MAPE를 평가 지표로 선정하였다[12].

$$MAPE = 100\% \times \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| \quad (5)$$

\hat{y}_t : t시점에서의 예측가격
 y_t : t시점에서의 실제가격
 n : 전체 예측된 표본의 수

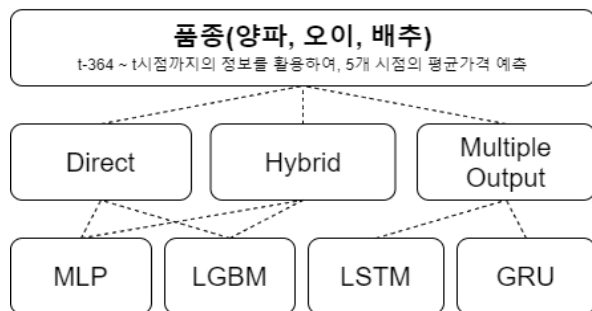


Fig. 1. Research Framework

IV. Empirical Test

1. Experimental Design

실험은 Python 3.6 버전에서 진행하였으며 Google Tensorflow, matplotlib, statsmodels 등의 라이브러리를 활용하였다. 실험은 총 3개의 농산물 품종(양파, 무, 배추)을 대상으로 진행하였다. 모델 학습 및 예측에 필요한 농산물 가격 데이터는 농넷에서 수집하였다. 2014년 1월 3일부터 2021년 12월 31일까지의 데이터를 활용하여 30

일(월), 90일(분기), 180일(반기), 270일, 365일(1년) 총 5개 시점의 미래 가격 예측을 수행하였다. 수집한 일별 데이터 중 결측치는 python에서 제공하는 pandas 모듈과 asfreq 함수를 활용하여 보완하였다.

농산물의 가격은 각 품종별로 가격대가 상이하여 비교 분석에 주의가 필요하다. 구체적으로 각 품종별 가격의 기술통계 분석결과는 다음의 Table 1과 같다.

Table 1. Descriptive Statistics

| | Onion | Cucumber | Cabbage |
|-------|----------|----------|----------|
| Count | 3,078 | 3,132 | 2,895 |
| Mean | 755.75 | 1,893.70 | 531.62 |
| STD | 326.63 | 756.34 | 270.04 |
| Min | 92.07 | 527.81 | 100.00 |
| 25% | 493.64 | 1,289.58 | 331.60 |
| 50% | 683.34 | 1,785.76 | 447.92 |
| 75% | 983.16 | 2,390.92 | 654.04 |
| Max | 1,927.61 | 5,300.32 | 2,310.00 |

Train Set은 2014년 1월 3일부터 2020년 12월 31일까지의 품종별 가격 데이터이다. 농산물 수확의 주기는 1년 이므로, 봄부터 겨울까지의 1년 사이에 가격 변화의 패턴을 확인하기 위해 Test Set은 2021년 1월 1일부터 2021년 12월 31일까지의 가격 데이터로 설정하였다.

마지막으로, 본 연구에서는 Direct 및 Hybrid 전략이 Multiple Outputs 전략에서 사용한 일반적인 시계열 분석 기법인 LSTM과 GRU(Gated Recurrent Unit) 대비 얼마나 효과적인지 살펴보고자 하였다. 이 때 Direct, Hybrid에서는 MLP와 LGBM(Light GBM)을 사용하였다.

2. Experimental Results

먼저 LGBM의 예측 결과를 Table 2에 제시하였다. 3개 품종의 5개 시점별 예측가격을 살펴본 결과, 양파는 Direct, 오이와 배추는 Hybrid가 더 적합한 전략임을 알 수 있었다. 이어 MLP의 예측 결과를 Table 3에 제시하였다. MLP 모델에서 양파는 Hybrid 오이와 배추는 Direct 전략이 평균적으로 더 낮은 MAPE를 산출하고 있는 것으로 나타났다. 특히 양파의 경우, Hybrid와 결합되어 사용할 때 우수한 성능을 보이는 것을 확인할 수 있었다.

다음으로 LSTM과 GRU의 예측 결과를 Table 4에 제시하였다. Direct 및 Hybrid 전략의 비교군으로 실험된 Multiple Outputs 전략의 예측값은 LGBM, MLP의 최적 모델 대비 MAPE가 전반적으로 높은 것을 확인하였다. 특히 학습 데이터의 양이 제한적인 상황에서 실험을 진행해 서인지 5개 시점의 세부 예측 결과 중 90일 이후에 대한 예측부터는 변별력 있는 예측을 보여주지 못하였다.

Table 2. LGBM Results from Direct/Hybrid Strategies

| | Strategy | MAPE(%) | | | |
|------|----------|---------|----------|---------|-------|
| | | Onion | Cucumber | Cabbage | Avg. |
| 30 | Direct | 24.60 | 24.51 | 33.02 | 27.38 |
| | Hybrid | 29.51 | 24.21 | 30.04 | 27.92 |
| 90 | Direct | 38.40 | 26.61 | 37.95 | 34.32 |
| | Hybrid | 44.50 | 25.37 | 39.63 | 36.50 |
| 180 | Direct | 37.50 | 28.13 | 34.21 | 33.28 |
| | Hybrid | 42.05 | 26.76 | 31.19 | 33.33 |
| 270 | Direct | 40.03 | 27.70 | 29.61 | 32.44 |
| | Hybrid | 42.60 | 25.89 | 28.50 | 32.33 |
| 365 | Direct | 28.87 | 27.47 | 38.16 | 31.50 |
| | Hybrid | 43.89 | 24.61 | 25.18 | 31.23 |
| Avg. | Direct | 33.88 | 26.88 | 34.59 | 31.78 |
| | Hybrid | 40.51 | 25.37 | 30.91 | 32.26 |

Table 3. MLP Results from Direct/Hybrid Strategies

| | Strategy | MAPE(%) | | | |
|------|----------|---------|----------|---------|-------|
| | | Onion | Cucumber | Cabbage | Avg. |
| 30 | Direct | 25.09 | 33.23 | 36.63 | 31.65 |
| | Hybrid | 15.94 | 26.93 | 37.87 | 26.91 |
| 90 | Direct | 50.38 | 30.81 | 62.03 | 47.74 |
| | Hybrid | 16.06 | 29.03 | 46.95 | 30.68 |
| 180 | Direct | 43.92 | 30.46 | 45.34 | 39.90 |
| | Hybrid | 21.05 | 31.14 | 30.27 | 27.49 |
| 270 | Direct | 29.75 | 32.05 | 39.43 | 33.74 |
| | Hybrid | 23.93 | 38.26 | 30.47 | 30.89 |
| 365 | Direct | 28.11 | 28.46 | 34.34 | 30.30 |
| | Hybrid | 23.52 | 37.35 | 120.01 | 60.29 |
| Avg. | Direct | 35.45 | 31.00 | 43.55 | 36.67 |
| | Hybrid | 20.10 | 32.54 | 53.11 | 35.25 |

Table 4. Results from Multiple Outputs Strategy

| | Model | MAPE(%) | | | |
|------|-------|---------|----------|---------|-------|
| | | Onion | Cucumber | Cabbage | Avg. |
| 30 | LSTM | 44.61 | 28.25 | 29.85 | 34.23 |
| | GRU | 44.61 | 28.25 | 29.85 | 34.23 |
| 90 | LSTM | 38.24 | 27.67 | 41.36 | 35.75 |
| | GRU | 40.23 | 30.02 | 42.28 | 37.51 |
| 180 | LSTM | 28.81 | 29.17 | 34.47 | 30.81 |
| | GRU | 28.30 | 28.29 | 37.83 | 31.47 |
| 270 | LSTM | 28.18 | 29.77 | 42.69 | 33.54 |
| | GRU | 27.80 | 31.07 | 60.49 | 39.79 |
| 365 | LSTM | 28.33 | 26.88 | 41.08 | 32.10 |
| | GRU | 27.90 | 21.58 | 53.13 | 34.20 |
| Avg. | LSTM | 33.63 | 28.35 | 37.89 | 33.29 |
| | GRU | 33.77 | 27.84 | 44.72 | 35.44 |

마지막으로 각 품종의 기간별 전략 및 분석기법의 최우수 조합을 Table 5에 제시하였다. 양파 품종의 경우 MLP와 Hybrid 전략의 조합이 가장 우수한 예측력을 보였고, 오이와 배추 품종의 경우 LGBM과 Hybrid 전략이 가장 높은 성능을 나타냈다. 모델의 관점에서 볼 때, 최우수 조합 15개 중 품종에 따라 MLP와 LGBM을 Hybrid와 결합한 모델이 12회 관찰되어, Hybrid 전략이 보편적으로 우수함을 검증하였다.

Table 5. Best Models and Strategies

| Dataset | Onion | Cucumber | Cabbage |
|---------|------------|-------------|-------------|
| 30 | MLP+Hybrid | LGBM+Hybrid | LSTM,GRU |
| 90 | MLP+Hybrid | LGBM+Hybrid | LGBM+Direct |
| 180 | MLP+Hybrid | LGBM+Hybrid | MLP+Hybrid |
| 270 | MLP+Hybrid | LGBM+Hybrid | LGBM+Hybrid |
| 365 | MLP+Hybrid | GRU | LGBM+Hybrid |

V. Conclusion

본 연구는 재귀적 형태의 일반적인 시계열 예측 전략 대비 새로운 전략들이 중장기 농산물 가격 예측에 얼마나 효과적인지 검증하고자 양파, 오이, 배추 등 3가지 품종을 대상으로 다양한 전략 및 모델별 실험을 진행하였다. 특히 전문가들이 중장기 가격 예측 전략으로 추천한 Hybrid 전략을 실제 분석에 적용해 봄으로써 해당 접근법의 유효성을 살펴본 것이다. 분석 결과를 종합한 Table 5를 참고하면 품종 및 기간별 차이는 존재하지만 전반적으로 Multiple Outputs 전략 대비 Multi-step Forecasting, 그 중에서도 특히 Hybrid 전략이 상대적으로 더 나은 성능 조합을 보였다.

본 연구의 학술적 의의는 다음과 같다. 첫째, 기존 연구에서 거의 다루어지지 않은 30일 이상의 중장기 일별 가격 예측을 시도하였다. 둘째, Direct를 비롯하여 전문가들이 가장 추천한 Hybrid 전략의 효과를 Multiple Outputs 전략과 함께 실증적으로 분석하였다. 분석 결과 Hybrid 전략이 단순히 이론적인 탁월함에 그치지 않고 경우에 따라 타 접근법보다 효과적인 것을 알 수 있었다는 점에서 학술적 의의가 있다.

본 연구의 실무적 의의는 다음과 같다. 첫째, 생산자와 소비자 모두의 삶에 큰 영향을 미치는 농산물 가격의 중장기 일별 예측을 고도화할 수 있는 접근법을 제시하였다. 품종마다 최적의 전략이 다르게 나타나고 있음이 확인된 바, 품종별로 차별화된 전략을 통해 접근하면 보다 정확한 가격 예측이 가능할 것이다. 둘째, 본 연구에서 적용된 접근법들은 기본적인 원리가 동일하기 때문에 농산물 외에도 다른 도메인에서의 중장기 일별 가격 예측에도 유용하게 적용될 수 있을 것이다.

본 연구의 한계 및 향후 연구 방향은 다음과 같다. 첫째, 본 연구에서는 여러 가지 농산물 중 세 가지 품종에 한정하여 실험을 진행하였다. 향후 다양한 품종을 대상으로 실험을 진행한다면 보다 풍부한 결과를 얻을 수 있을 것이다. 둘째, 본 연구에서는 다양한 접근법의 효과를 비교 분석하는 과정에서 가격 데이터만을 활용하여 실험을 진행하였다. 하지만 농산물 가격에는 수입농산물 가격, 수출입

량, 재해발생여부, 기상정보 등 수많은 요인들이 영향을 미칠 수 있다. 따라서 향후 연구에서는 가격을 포함한 다양한 변인을 고려한 연구를 시도해 볼 필요가 있다.

REFERENCES

- [1] J. H. Ha, S. Seo, and S. W. Kim, "Improving Forecasting Performance for Onion and Garlic Prices," *Journal of Korean Society of Rural Planning*, Vol. 25, No. 4, pp. 109-117, Nov. 2019. <https://doi.org/10.7851/Ksrp.2019.25.4.109>
- [2] Y. S. Lee, J. J. Kim, K. S. Kim, D. H. Ahn, and J. S. Cho, Influence on aggregate general price level of agricultural pricing and policy implications. Policy Research Report, Korea Rural Economic Institute (KREI), Jun. 2013. <http://repository.krei.re.kr/handle/2018.oak/20472>
- [3] S. Shin, M. Lee, and S. Song, "A Prediction Model for Agricultural Products Price with LSTM Network," *Journal of Korean Contents Society*, Vol. 18, No. 11, pp. 416-429, Nov. 2018. <https://doi.org/10.5392/JKCA.2018.18.11.416>
- [4] K. Suh and J. J. Lee, "Analysis of the Effect of Shipping Control depending on the Limited Storage Life of Agricultural Products," *Journal of Korean Society of Rural Planning*, Vol. 10, No. 3, pp. 53-58, Aug. 2004.
- [5] J. K. Seo, D. J. Choi, K. H. Ko, J. Pak, "Analyzing Significant Variables from a Linear Regression-Based Prediction Model for Rice Prices", *Proceedings of Korea Society of Computer Information*, Vol. 30, No. 2, pp. 39-42, July. 2022.
- [6] J. A. Rodrigo, J. E. Ortiz, "Skforecast: time series forecasting with Python and Scikit-learn" Attribution 4.0 International (CC BY 4.0), Feb. 2021. <https://www.cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html>
- [7] J. Brownlee, *Introduction to Time Series Forecasting With Python - Discover How to Prepare Data and Develop Models to Predict the Future*, Machine Learning Mastery, Feb. 2017. <https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python/>
- [8] J. H. Ha, S. T. Seo, and S. W. Kim, "Evaluation on the Performance of Onion and Garlic Forecasts," *Proceedings of the Summer Academic Conference of the Korea Food Trade Association*, pp. 559-572, Jul. 2019.
- [9] B. Choi and I. C. Choi, "Monthly price Forecasting of Fruit-type Vegetables Using Time-Series Analyses," *Journal of Rural Development*, Vol. 30, No. 1, pp. 129-148, Apr. 2007. <http://doi.org/10.36464/jrd.2007.30.1.012>
- [10] D. H. Jung and Y. Y. Cho, "A Prediction Model for Hallabong Tangor Product Prices using LSTM (Long Short-term Memory) Network," *Horticultural Science and Technology Journal*, Vol. 40, No. 5, pp. 571-577, Jun. 2022. <http://doi.org/10.7235/HORT.20220051>
- [11] D. W. Yoo and J. B. Park, "Prediction of Agricultural Prices Using LSTM" *Proceedings of the Korean Institute of Information and Commucation Sciences Conference*, pp. 710-712, May 2022.
- [12] K. H. Nam and Y. C. Choe, "A Study on Onion Wholesale Price Forecasting Model," *Journal of Agricultural Extension & Community Development*, Vol. 22, No. 4, pp. 423-434, Jan. 2015. <https://doi.org/10.12653/jecd.2015.22.4.0423>
- [13] S. Jang, H. Chun, I. Cho, and D. Kim, "A study on cabbage wholesale price forecasting model using unstructured agricultural meteorological data," *Journal of the Korean Data and Information Science Society*, Vol. 28, No. 3, pp. 617-624, May 2017. <https://doi.org/10.7465/jkdi.2017.28.3.617>
- [14] K. T. Bae and C. J. Kim, "An Agricultural Estimate Price Model of Artificial Neural Network by Optimizing Hidden Layer," *Journal of the Korean Society of Information Technology*, Vol. 14, No. 12, pp. 161-169, Dec. 2016. <http://doi.org/10.14801/jk iit.2016.14.12.161>
- [15] J. Kim and J. I. Lee, "A Study on Agricultural Price Prediction System based on Deep Learning," *Journal of the Korean Society of Information Technology*, Vol. 17, No. 6, pp. 27-34, Jun. 2019. <http://doi.org/10.14801/jkiit.2019.17.6.27>
- [16] D. Sahoo, N. Sood, U. Rani, G. Abraham, V. Dutt, and A. D. Dileep, "Comparative Analysis of Multi-Step Time-Series Forecasting for Network Load Dataset," *Proceedings of 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020. <http://doi.org/10.1109/ICCCNT49239.2020.9225449>

Authors



Jonghyun Park received a B.A. degree from Kookmin University, Korea, in 2021. He is currently pursuing the M.S. degree in Business IT Graduate School, Kookmin University, Korea.

Jonghyun Park is interested in business analytics, deep learning and text mining.



Yeong-Woo Lim obtained his bachelor's and master's degrees from Seoul National University in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree in Business IT Graduate School, Kookmin

University, Korea. Yeong-Woo Lim is interested in data analytics, data mining, machine learning, social network analysis.



Do Hyun Lim obtained a BS in Industrial Psychology from the Department of Industrial Psychology at Kwangwoon University. He is currently enrolled in the master's program of the Graduate School of Business IT at

Kookmin University. Do Hyun Lim is interested in machine learning, fraud detection, and deep learning.



Yunsung Choi received a bachelor's and a master's degree in computer science from Sogang University. He is currently working as a research director at the Corporate Research Institute in ONTWOIN Co., Ltd.

Yunsung Choi is interested in machine learning, big data and AI platform.



Hyunchul Ahn received a BS in Industrial Management from KAIST, and a ME and PhD from KAIST Graduate School of Management. He is currently working as a professor of the Graduate School of Business

IT at Kookmin University. His main research areas include AI applications in finance and customer relationship management as well as behavioral models related to information system acceptance.