

# Lasso 회귀분석을 활용한 농산물 가격예측 모델 변수 선정 연구

## A study on selection of variables in Forecasting Model of Agricultural Products Price using Lasso regression

이도영 (Doyoung Lee, 제1저자) | 전북대학교 농업경제학과 석사과정 | ldo608@hanmail.net  
양예원 (Yewon Yang, 공동저자) | 전북대학교 분자생물학과 학사과정 | yangyewon0579@gmail.com  
이주형 (Joohyung Lee, 공동저자) | 전북대학교 소프트웨어공학과 학사과정 | yamiblack95@gmail.com  
박지홍 (Jihonh Park, 공동저자) | 전북대학교 통계학과 학사과정 | jihong8090@gmail.com  
강민구 (Mingu Kang, 공동저자) | 전북대학교 소프트웨어공학 과 석사과정 | kang.mingu94@gmail.com

목 차	1. 서론
	2. 선행연구분석
	3. 연구설계
	4. 연구결과
	5. 결론

### 초 록

농산물 가격 파동 문제를 해결하기 위해 딥러닝 모델을 연구하였다. 선행연구를 기반으로 기본변수를 선정 후, lasso 회귀분석을 적용해 최적 변수를 선정하여 LSTM모델에 적용하여 RMSE로 성능을 비교하였다. 최적 변수를 이용했을 때, 양파와 대파의 1, 3, 5, 7일 후 가격 예측의 RMSE가 평균적으로 0.0229, 0.0019씩 감소하였다. lasso 회귀분석을 활용해 최적 변수 선정은 모델의 정확도를 향상시키며, lasso 회귀분석을 이용해 선정된 최적 변수들을 이용한 딥러닝 농산물 가격 예측 모델은 가격 파동을 완화에 활용될 수 있을 것으로 예상된다.

\* 키워드 : 농산물 가격예측, LSTM, Lasso회귀분석, 변수선정

### ABSTRACT

A deep learning model was studied to solve the problem of agricultural product price fluctuations. After selecting the basic variable based on previous studies, the optimal variable was selected by applying lasso regression and applied to the LSTM model to compare the performance with RMSE. When using the optimal variable, the RMSE of the price prediction decreased by 0.0229, 0.0019 on average after 1, 3, 5, and 7 days of onions and leek price. The selection of optimal variables using lasso regression was effective, and the LSTM agricultural product price prediction model using the optimal variables selected using lasso regression is expected to be used to alleviate price fluctuations.

\* **Keywords** : Selection of variables, Forecasting Model of Agricultural Products Price using, LSTM, Lasso regression

• 본 논문은 과학기술정보통신부가 주최하고 한국데이터산업진흥원이 주관하는 DB산업육성(데이터 청년인재 양성)사업의 일환으로 수행된 2021년 데이터 청년 캠퍼스 프로젝트 결과물임

• 논문접수 : 2020년 8월 30일 •최초심사일: 2020년 월 일 •심사완료일: 2020년 월 일

---

## 1. 서론

농산물은 생산의 비탄력적인 재화로 수요에 따라 생산을 조절하기가 어려운 특징을 가지고 있으며, 특히 농업 생산은 재배환경에 크게 영향을 받아 기상여건도 중요하다. 또한 농민들은 올해의 가격을 기준으로 내년의 재배면적을 결정하는 경향을 가지고 있어 농산물 가격의 농민들의 중요한 의사결정의 도움수단이 된다.

주요 채소류의 최근 20년 사이 가격변동률이 평균 15~40%나 될 만큼 농산물 가격은 불안정하며, 가격위험이 크다. 농산물 가격위험이 클수록 매년 작물 재배면적의 변화도 커지게 되어 이는 곧 매년 되풀이되는 농산물 가격 폭락 및 폭등으로 나타난다. 이러한 문제를 해결하기 위해 정부 부처와 기업에서는 농산물 가격을 예측하는 시스템을 도입하였는데, ‘경상남도 농산물 가격 예측 시스템’은 풋고추, 마늘, 양파 등 11개의 농산물 주간 가격을 예측을 진행하고 있으며 평균 80% 정도의 예측력을 가지고 있다. ‘서울시 농수산 식품공사’는 데이터 분석이 아닌 가락시장의 경매 사들의 의견을 정리하여 보고서로 예측하고 있으며, ‘팜에어 한경’은 변동 폭이 큰 주요 농산물의 정보를 제공하고 있다.

농산물 가격예측은 개인의 의견, 딥러닝, 통계기법 등 다양한 방법으로 예측하는데, 현재는 딥러닝 모델이 많이 활용된다. 과거의 딥러닝은 하드웨어의 기술적 한계에 막혀 딥러닝 모델 연구에 어려움이 컸지만 현재는 하드웨어의 환경이 크게 발전하여 빅데이터 학습을 가능하게 되었고 기존 모델의 개선 방법이 나타나 모델의 성능은 크게 향상되었다. 다양한 변수를 이용한 딥러닝 모델은 농산물 가격예측에 있어서 전통적인 통계기법인 시계열 분석방법보다 성능이 우수하여, 보다 정확한 농산물의 예측 가격을 제시해주고 있다.

하지만 딥러닝 모델을 적용하는 농산물 가격예측은 연구자 또는 시스템에 따라서 예측하는 방법, 변수는 모두 달라 어떤 변수가 농산물 가격을 정확히 설명하는지 구분할 수 없다. 농산물시장이 개방되고 유통이 복잡해지면서 농산물의 가격 결정요인은 다변화 되었으며, 농산물의 가격을 설명하는 변수를 선택하는 것에 큰 어려움을 가진다.

딥러닝 모델의 설명변수의 중요도를 규명하는 것은 선형회귀 모형과 같이 간단한 문제가 아니며 해석적인 관점에서 중요한 문제이다. 우리가 알지 못한 설명변수가 농산물 가격에 영향을 미칠 수 있으며 반대로 우리가 농산물 가격에 영향을 미칠 것이라 예상했던 변수가 실제 영향을 미치지 않을 수도 있다. 이처럼 농산물 가격 예측 모델을 설계할 때 적합한 설명변수를 선정하는 것은 모델 성능에 중요한 문제이며, 보다 많은 연구가 필요하다.

이에 본 연구는 딥러닝 기반의 농산물 가격 예측에 있어 최적 변수선정방법으로 lasso회귀분석을 제안하며, 기존의 선행연구를 기반으로 선정한 기본변수와 기본변수를 다시 lasso 회귀분석을 적용시

---

켜 최적변수를 선정한다. 구성된 두 개의 변수 그룹을 LSTM모델에 이용하여 농산물 가격 예측의 성능을 비교해 봄으로써 lasso회귀분석의 우수성을 알아보고자 한다. 2000년부터 2020년의 데이터를 MinMax정규화 과정을 거쳐 30일간 모델에 학습시킨 후 1, 3, 5, 7일 후 서울시의 양파와 대파 소매가격을 예측하며 성능지표로 RMSE를 이용한다. 다음 2장에서는 선행연구분석을, 3장에서는 연구설계, 4장에서는 연구결과, 5장에서는 결론을 다룬다.

## 2. 선행연구분석

### 2.1 농산물 가격예측

농산물 가격예측에 관한 연구는 전통적인 통계기법인 시계열 분석방법을 적용한 연구와 딥러닝을 적용한 연구로 구분할 수 있다. 시계열 분석방법을 적용하는 연구는 단일변수인 농산물 가격을 분석하여 미래가격을 예측하며, 딥러닝 모델을 적용하는 연구는 단일변수(가격) 또는 가격 이외의 다른 변수를 추가로 이용하여 미래가격을 예측하는 연구로 구분할 수 있다.

농산물 가격 예측에 있어 딥러닝 모델이 시계열 분석보다 우수한 성능을 가지고 있다는 연구결과가 나타나고 있다. 특히, LSTM 모델을 활용한 농산물 가격 예측모델은 RMSE가 0.07로 다른 모델보다 예측성능이 월등히 높았다.

기존의 딥러닝 모델을 적용한 농산물 가격예측에 관한 연구는 선행연구를 기반으로 연구자가 연구하고자 하는 임의변수를 추가하여 모델화하였으며, 배경태 외(2016)는 21개의 기후데이터 중 7개의 기후데이터를 선별하여 변수로 이용하였다. 신성호 외(2018)는 시도별 날씨, 총 물가상승률, 총 품목별 물가상승률 등 108개의 변수를 이용하였고 YinHelin 외(2020)는 농산물 가격을 시계열 분해하여 계절성과 추세를 제외한 23주 전까지의 모든 잔차 값과 기후데이터, 농산물 반입량 등 많은 변수를 추가하여 연구하였다.

정부 연구기관인 농촌경제연구원은 KREI-KAMSO모형을 구축하여, 매년 농산물 수급과 균형가격을 전망한다. 농업 전망모형 KREI-KAMSO는 품목의 수요함수, 재배면적함수, 단수함수, 수급균형 항등식 등으로 구성된 계량경제학적 연립방정식체계로 위에서 언급한 선행연구와 변수에 있어 차이를 보인다. 예를 들어 수급모형의 재배면적은 생산사기기가 비슷한 작물의 재배면적까지 이용하여, 양파의 재배면적은 생산시기가 비슷한 양파, 대파, 보리의 재배면적이 변수로 이용된다. 이처럼 연구 모델마다 예측에 이용되는 변수에서 차이를 보이고 있다.

---

## 2.2 lasso를 활용한 변수 선정

연구자마다 모델을 설계할 경우 변수선택이 다르며, 이에 따른 성능 차이를 보이고 있다. 결국 변수 선정방식에 따른 딥러닝 예측모델의 성능은 달라지며 모델에 적합한 변수를 선정하는 것이 중요하다. 변수선정방법 중 lasso 회귀분석은 선형회귀분석과 비교하였을 때, 다중공선성을 감안한 계수 추정방법으로 에 따라서 독립변수 계수의 추정 값이 0이 된다. 일부 독립변수 계수를 0으로 추정되는 것은 독립변수 계수를 추정하는 동시에 유의미한 독립변수를 선택하는 것으로 머신러닝과 딥러닝에서 차원 축소 또는 변수선택 방법에 적용된다.

김나영 외(2020)는 중학생 학업열에 영향을 미치는 독립변수를 선정하기 위해 lasso회귀분석을 실시하였으며, 차성재 외(2018)는 변수선정방법과 딥러닝 모델의 접목하여 기업부도 예측모형을 비교 연구하여, 그 중 LSTM모델은 lasso회귀분석을 적용하는 변수선택방법이 유용함을 밝혔다.

농산물 가격예측모형의 변수는 연구자에 의해 다양한 변수가 이용되고 있으며, 반복적으로 이용되는 기후데이터도 조차 지역별, 온도, 습도, 지온 등 실제 이용되는 변수는 모두 다르다. 다양한 변수 중 농산물 가격 예측에 있어 적합한 변수를 선정하기 위한 방법이 필요하다.

본 연구는 LSTM모델을 적용하여 농산물 가격을 예측하며, 모델에 이용되는 변수는 선행연구와 KREI-KAMSO모형을 토대로 선정하며, 선정된 변수를 다시 lasso 회귀분석을 적용한 뒤, 모델에 적합한 변수로 간추린다.

## 3. 연구설계

### 3.1 변수선정 및 데이터처리

양파와 대파 소매가격 예측에 활용한 변수는 일간 데이터인 전국 날씨, 월간 데이터인 총 지수, 수입량, 수출량, 수입금액, 수출금액 등이며 연간 데이터인 재배면적, 생산량, 생산시기가 유사한 작물의 재배면적, 생산량이며 <표1>로 정리하였다. <그림1>, <그림2>는 양파와 대파의 소매가격을 시계열 분해한 결과이며, 변수는 1일전부터 20일까지 모든 잔차와 40일전 잔차, 60일전 잔차로 구성하였다. 또한 양파 가격을 예측하는 변수에는 양파 가격의 잔차만 포함되며, 대파 가격의 잔차는 포함하지 않았다. 반대로 대파 가격을 예측하는 변수에 양파의 잔차는 포함하지 않았다. 이외 다른 변수는 모두 동일하게 포함된다.

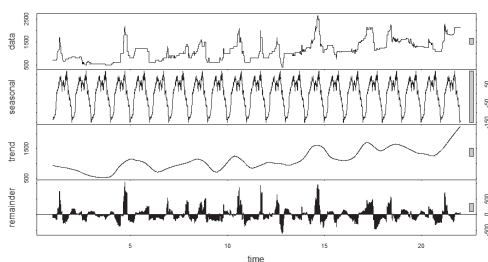
예측시점에서 수집 가능한 데이터를 선정하기 위해 월간 데이터는 전월 기준으로 연간 데이터는 전년도 기준으로 활용하였으며, 예를 들어 2019년 2월 4일의 가격을 예측하기 위해서는 1월의 월간 데이터들과 2018년의 연간 데이터가 활용하였다. 또한 데이터 크기를 맞추기 위해 월간과 연간데이터는

일일데이터로 적용하였으며, 결측치는 제거하였다.

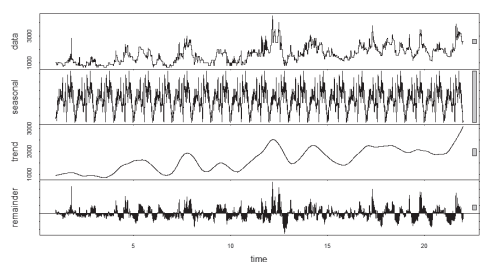
연구에 활용된 데이터는 2000년부터 2020년의 가격데이터가 기준이 되었으며, 월간데이터는 1999년 12월부터 2020년 11월, 연간데이터는 1999년부터 2019년까지 데이터를 적용하였다. 2001년부터 2020년까지 총 7,670일 중 농산물 소매가격은 공휴일을 제외하고 존재하였으며, 결측치가 발생한 일 을 제외한 5,091일의 데이터를 이용하여 데이터셋을 구성하였고 학습 데이터는 60%, 검증 데이터는 20%, 테스트 데이터는 20%로 구분하여 MinMax정규화 과정을 진행하였다.

<표 1> 변수선정

일간	변수	월간	변수	연간	변수
소매가격	양파가격	물가지수	총 지수	생산면적	양파
	대파가격		곡물 및 식품		대파
전국날씨	평균 기온		야채지수	생산량	보리
	평균 최저기온		대파지수		양파
	평균 최고기온		양파지수		대파
	평균 강수량		경유지수		보리
	최고 강수량	수출량	양파		
	평균 풍속	수입량	대파		
	평균 습도		양파		
	일사합		대파		
시계열 분해	일조합	수출금액	양파		
	잔차		대파		
		수입금액	양파		
			대파		



<그림 1> 양파 가격 시계열분해



<그림2> 대파 가격 시계열분해

소매가격은 농산물 유통정보에서, 전국 기상자료는 기상 자료 개방 포털, 수입과 수출에 관련 데이터는 무역 통계 진흥원, 무역협회, 물가지수와 생산면적, 생산량은 통계청에서 수집하였으며, 시계열 분석은 통계프로그램 R의 STL함수로 계산하였다.

## 3.2 연구방법

### 3.2.1 Lasso 회귀분석

선형회귀모델의 추정방법 중 최소제곱법(OLS)은 최적 선형 비편향 추정량임이지만, 모델의 독립변수의 선정은 완벽해야하며, 독립변수 간 또는 독립변수와 잔차는 연관은 없어야하며, 잔차는 등분산을 가져야 한다는 강력한 가정의 조건이 필요하다. 하지만 현실적으로 완벽한 독립변수 선정은 불가하며, 잘못 설정된 독립변수로 인해 모델의 추정 정확도가 낮아지며 동시에 모델 해석을 어렵게한다.

이러한 두가지 문제점을 해결하기 위한 방법으로 벌점화 추정방법이 개발되었으며, 이 방법을 통해 모델의 계수는 벌점함수(penalty function, )를 사용하여 제한한 상태에서 잔차 제곱합을 최소화하는 추정량을 얻는 방법이다. 벌점화 추정 방식 중 하나인 lasso 회귀분석은 Tibshirani(1996)에 의해 제안된 방법으로 기존의 OLS추정방식에  $\lambda$ 의 식을 추가하여 다중공선성을 고려한 계수를 추정하는 방법이다.

lasso 회귀분석의 추정식은 (1)이며  $\lambda$ 의 식에 의해서 다중공선성이 높은 독립변수의 계수가 0으로 추정되어 머신러닝에서 자동적으로 차원축소 또는 변수 선택 방법으로 활용된다. Lasso 회귀분석은 최소제곱추정법과 비교하였을 때 약간의 편향(bias)을 허용하기 때문에 분산을 감소시킴으로 예측의 효율성은 하락하지만, 예측의 정확도를 높일 수 있다는 장점이 있다.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \cdots (1)$$

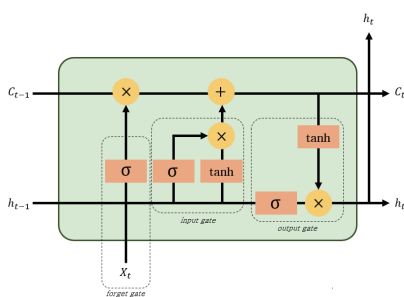
식(1)에서  $\lambda$ 에 따라 회귀계수가 결정되므로, 본 연구에는 교차타당화를 실시하여 최소값으로 선정하였다.

### 3.2.2 LSTM 모델

<그림3>은 LSTM 모델구조이며, LSTM 모델은 RNN구조의 장기 의존성 문제를 극복하려고 제안된 구조이며, 과거의 정보를 기억할 수 있는 장점을 가지고 있으며 Forget gate, Input gate, Output gate로 구성되어있다.

Forget gate는 이전의 단계와 현단계의 입력으로부터 sigmoid 활성화 함수를 활용하여 0과 1 사이의 값을 출력하는데, 0은 정보를 완전히 제거하는 의미이며 1은 정보를 유지한다. Input gate는 tanh 함수를 활용하여 입력된 새로운 저장을 결정하며 Output gate는 이전의 입력값과 현재의 입력값을 이용하여 값을 출력한다.

<표2> hyper-parameters 값을 정리한 것이며, Loss함수는 MSE로 사용하며 평가지표는 RMSE로 사용한다. 본 연구는 구글 Colab 가상환경에서 연구를 수행하였으며, PyTorch를 이용하여 LSTM모델을 구축하였다.



<그림 3> LSTM 구조

<표2> hyper-parameters 값

Hyper-parameters	값
batch size	64
epochs	100
learning rate	0.001
dropout	0.01
layer	1
hidden_size	100

## 4. 연구결과

### 4.1 Lasso 회귀분석 결과

<표 3>, <표 4>는 lasso 회귀분석 계수이며, 양파의 는 0.1877, 대파의 는 0.1266로 추정하였다. 양파의 변수 중 전국평균온도, 양파 수출량, 6일전 잔차, 7일전 잔차, 8일전 잔차, 9일전 잔차, 10일전 잔차, 11일전 잔차, 13일전 잔차, 18일전 잔차가 제외되었으며, 대파의 변수 경우 대파 수출량, 1일전 잔차, 2일전 잔차, 5일전 잔차가 제외되었다. 즉 양파 가격을 예측하기 위해서 전국평균온도, 양파 수출량, 6일전 잔차, 7일전 잔차, 8일전 잔차, 9일전 잔차, 10일전 잔차, 11일전 잔차, 13일전 잔차, 18일전 잔차는 불필요한 변수이며, 대파 가격을 예측하는데 대파 수출량, 1일전 잔차, 2일전 잔차, 5일전 잔차는 불필요한 변수이다.

<표3>양파변수의 lasso 회귀분석 결과

변수	계수	변수	계수
대파가격	0.047531	양파 수입량	0.002359
평균 최저기온	-2.28183	양파 수출금액	-0.00202
평균 최고기온	2.472404	양파 수입금액	-0.00253
평균 강수량	-1.22644	대파 수출량	-76.4978
최고 강수량	0.087815	대파 수입량	-0.83698
평균 풍속	-4.45629	대파 수출금액	64.7589
평균 습도	1.177428	대파 수입 금액	0.709865
일사합	12.23304	잔차	0.98558
일조합	-6.83324	1일전 잔차	0.003323
대파 재배면적	0.008471	2일전 잔차	0.010649
대파 생산량	-0.00065	3일전 잔차	0.000301
양파 재배면적	-0.00451	4일전 잔차	0.002594
양파 생산량	0.000433	5일전 잔차	0.000657
보리 재배면적	-0.00411	12일전 잔차	-0.00015
보리 생산량	0.001089	14일전 잔차	-0.02308
총 지수	9.4926	15일전 잔차	-0.025
곡물 및 식품 지수	3.955681	16일전 잔차	-0.04225
야채지수	-0.04149	17일전 잔차	-0.03413
대파지수	-0.05389	19일전 잔차	-0.0048
양파지수	8.48491	20일전 잔차	-0.15855
경유지수	-2.33111	40일전 잔차	-0.11831
		60일전 잔차	0.060479



<표4> 대파변수의 lasso 회귀분석 결과

변수	계수	변수	계수
양파가격	0.104496	양파 수출금액	0.192576
평균기온	1.454041	양파 수입금액	0.032551
평균 최저기온	-6.31366	대파 수입량	-1.00561
평균 최고기온	6.997741	대파 수출금액	11.18701
평균 강수량	-0.26499	대파 수입금액	0.886643
최고 강수량	0.071601	잔차	1.054061
평균 풍속	-10.1603	3일전 잔차	-0.00888
평균 습도	-0.14611	4일전 잔차	-0.01572
일사합	-0.61664	6일전 잔차	-0.00161
일조합	2.76624	7일전 잔차	-0.02392
대파 재배면적	0.152697	8일전 잔차	-0.00886
대파 생산량	-0.0047	9일전 잔차	-0.00772
양파 재배면적	-0.03269	10일전 잔차	-0.00967
양파 생산량	0.00066	11일전 잔차	-0.0191
보리 재배면적	-0.00706	12일전 잔차	-0.01087
보리 생산량	0.001259	13일전 잔차	-0.01317
총 지수	34.42209	14일전 잔차	-0.00487
곡물 및 식품 지수	-16.1798	15일전 잔차	-0.01608
야채지수	-1.77887	16일전 잔차	-0.00403
대파지수	10.29787	17일전 잔차	-0.00836
양파지수	-0.2999	18일전 잔차	-0.02151
경유지수	-6.57533	19일전 잔차	-0.00064
양파수출량	-0.06003	20일전 잔차	-0.15597
양파 수입량	-0.01168	40일전 잔차	-0.11527
		60일전 잔차	0.043611

<표 5>는 연구 결과이다. lasso 회귀분석을 적용하지 않고 선행연구 분석으로만 선정한 변수로 양파의 미래 가격을 예측하였을 경우(기본변수), RMSE는 1일 후는 0.2457, 3일 후는 0.2277, 5일 후는 0.2326, 7일 후는 0.2775이다. 기본변수에 lasso 회귀분석을 적용한 변수로 양파의 미래가격을 예측하였을 경우(lasso변수), RMSE는 1일 후는 0.2270, 3일 후는 0.2004, 5일 후는 0.2239, 7일 후는 0.2406으로 평균적으로 RMSE는 0.0229가 감소하여 예측의 정확도가 상승하여 성능은 상승하였다.

대파의 미래 가격을 기본변수로 예측하였을 경우, RMSE는 1일 후는 0.0712, 3일후는 0.0792, 5일 후는 0.0747, 7일 후는 0.0664로 좋은 성능을 나타냈다. lasso 변수로 예측하였을 경우, RMSE는 1일 후는 0.0706, 3일 후는 0.0676, 5일 후는 0.0765, 7일 후는 0.0691으로 1,3일후 예측가격의 RMSE는 감소하여 예측의 정확도는 상승하였지만, 5,7일후 RMSE는 오히려 상승하여 예측의 정확도는 감소하였다. 하지만 평균적으로 RMSE는 0.0019 감소하였다.

대파는 선행연구를 통해서 선정된 기본변수로 예측하였을 경우 RMSE가 약 0.07로 정확한 예측을 하였다. 이는 선행연구를 통한 변수선정에 유의미하였으며, 유의미한 기본변수를 선정하였다면, Lasso회귀분석의 적용한 변수선정은 무의미한 것으로 판단된다.

<표 5> 연구 결과

RMSE	예측기간	1일 후	3일 후	5일 후	7일 후
양파	기본 변수	0.2457	0.2277	0.2326	0.2775
	Lasso 변수	0.2270	0.2004	0.2239	0.2406
대파	기본 변수	0.0712	0.0792	0.0747	0.0664
	Lasso 변수	0.0706	0.0676	0.0765	0.0691

## 5. 결론

농작물의 가격 파동은 매년 반복되고 있으며, 가격 파동은 소비자 물가와 농가 수입 등에도 영향을 끼치고 있다. 농작물 가격 파동을 해결하기 위한 정부부처와 기업에서 다양한 가격 예측 시스템을 도입하였고, 선행연구에 의하면 딥러닝 모델 중 LSTM 모델을 적용한 농산물 가격 예측이 우수한 성능을 가졌다. 딥러닝 모델을 적합한 설명변수를 선정하는 것은 매우 중요한 문제이며, 우리의 연구에서 적절한 변수 설정을 위해서 lasso 회귀분석을 활용하였다. 선행연구를 기반으로 기본변수를 선정한 후, lasso 회귀분석을 적용해 최적 변수를 선정하고 기본변수와 Lasso 변수의 모델 성능을 비교하였다.

---

선행연구를 기반으로 동일한 데이터에서 동일한 방법으로 양파와 대파의 기본변수를 선정하였고, 가격을 예측하였을 경우, 양파의 RMSE는 약 0.22이상으로 성능이 낮았으나, 대파의 RMSE는 약 0.07로 성능이 우수했다. 즉, 농산물 품목마다 설명하는 변수가 다른 것으로 해석되며, 대파의 경우는 선행연구로만 선정한 변수로도 충분히 소매가격을 잘 설명하는 변수구성이었다. 반대로 양파의 경우 선행연구로 선정한 변수로는 소매가격을 잘 설명하지 못하는 변수구성이었다.

Lasso 회귀분석을 통해 최적변수를 선정하여 농산물 가격을 예측하였을 경우 RMSE는 양파와 대파 모두 평균적으로 0.0229, 0.0019 감소하였고 이는 lasso 회귀분석을 적용해 최적 변수 선정이 유효했다는 것으로 해석된다.

본 연구는 공공데이터를 기반으로 양파와 대파의 소매가격을 예측하였으므로 다른 농산물의 소매가격 예측으로 확대시킬 수 있을 것으로 기대한다. 또한 lasso 회귀분석을 적용한 최적 변수 선정하고, 선정된 변수를 이용한 LSTM 농산물 가격 예측 모델은 가격 파동문제를 완화하는데 활용될 수 있을 것으로 예상된다.

---

## 참고문헌

- 김나영, 황혜영 (2020) Lasso 벌점회귀모형을 적용한 중학생의 학업열의 관련 변수 탐색, 교육문화연구 vol.26, no.6, 통권 72호 pp. 337-355
- 배경태, 김창재 (2016) 인공지능망의 은닉층 최적화를 통한 농산물 가격예측 모델, 한국정보기술학회 논문지 14(12), 161-169
- 신성호, 이미경, 송사광 (2018). LSTM 네트워크를 활용한 농산물 가격 예측 모델, 한국콘텐츠학회 논문지, 18(11), pp.416-429
- 서홍석, 김충현, 김준호 (2020) 농업부문 전망모형 KREI-KASMO 2020 운용·개발 연구, 한국농촌경제연구원, 연구보고서
- 이정환, 김한호, 위태석, 하석건, 류상모 (2021). 농산물가격과 농업경영 안정을 위한 새로운 길. 시선집중 GSnJ, (286), 1-22.
- 차성재, 강정석(2018). 딥러닝 시계열 알고리즘 적용한 기업부도예측모형 유용성 검증, 한국지능정보시스템학회, 지능정보연구 24(4), pp.1-32
- 홍정명. 경남도, 농산물 가격 예측 정보 실시간 제공, 뉴시스, 2020년 4월 19일, [https://newsis.com / view/?id=NISX20200419\\_0000998481](https://newsis.com/view/?id=NISX20200419_0000998481)
- Yin, Helin; Jin, Dong; Gu, Yeong Hyeon; Park, Chang Jin; Han, Sang Keun. Yoo, Seong Joon (2020). STL-ATTN LSTM:Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM. Agriculture, Vol 10, Iss 612, p 612

## 국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Bae Kyung-Tae, Kim Chang-Jae. An Agricultural Estimate Price Model of Artificial Neural Network by Optimizing Hidden Layer. The Journal of Korean Institute of Information Technology 14(12), 2016.12, 161-169(9 pages)
- Cha Sung-jae, Kang Jung-seok. Corporate Default Prediction Model Using Deep Learning Time Series Algorithm, RNN and LSTM. Journal of Intelligence and Information Systems 24(4), 2018.12, 1-32(32 pages)
- Hong Jeong-myeong. Gyeongnam, Real-time information on agricultural product prices. Newsis. April 7, 2021. [https://newsis.com/view/?id=NISX20200419\\_0000998481](https://newsis.com/view/?id=NISX20200419_0000998481)
- Kim Na-young, Hwang Hye-young. Exploring Variables on Middle School Students' Academic

---

Engagement Using Lasso Regression. Journal of Education & Culture. Dec 30, 2020  
26(6):337

Lee Jung-hwan, Kim Han-ho, Wi Tae-seok, Ha Seok-gun, Ryu Sang-mo. Agricultural prices and new ways to stabilize agricultural management. GS&J Institute (286), 1-22.

Seo Hong-seok, Kim Chung-hyeon, Kim Jun-ho. KREI-KASMO 2020 Operation and Development Research. Korea Rural Economic Institute

Shin Sung-ho, Lee Mi-kyoung, Song Sa-kwang. A Prediction Model for Agricultural Products Price with LSTM Network. JOURNAL OF THE KOREA CONTENTS ASSOCIATION 18(11), 2018.11, 416-429(14 pages)

STL-ATTTLSTM:Vegetable Price Forecasting Using STL and Attention Mechanism Based LSTM. Agriculture, Vol 10, Iss 612, p 612

Yin, Helin; Jin, Dong; Gu, Yeong Hyeon; Park, Chang Jin; Han, Sang Keun. Yoo, Seong Joon (2020).