우수상

시설재배 오이 수확량 및 가격예측모형

이브람

시설재배 오이 수확량 및 가격예측모형

스마트팜 빅데이터를 활용한 딥러닝 모델링

이브람

Key words

수확량예측(Crop Yield Prediction), 가격예측(Price Prediction), LSTM순환신경망(LSTM)

ABSTRACT

Cucumbers are mainly cultivated in the greenhouse throughout the year and the production is relatively stable. However, prices fluctuate largely due to supply and demand in high or low temperature seasons. In the harvest soaring season, prices may drop by less than half, and in winter season, prices may soar more than tripled. This study predicts the crop yield of cucumbers using the Long Short-Term Memory (LSTM) recurrent neural network, and predicts the trading volume and price of the wholesale market using time series prediction machine learning models so that the farmer can predict the revenue of a crop period. Farmers may use this information to make agricultural decisions such as planting time, end of cropping season, and alternative crop cultivation. All data used for machine learning are collected from real farms in Gyeonggi-do, Korea from 2018 to 2020, and the price prediction model was trained with actual wholesale market transaction price. Prediction results are presented using comparative graphs and performance evaluation metrics, the mean absolute error and the coefficient of determination, R-Squared.

차례

1. 서론
2. 연구 방법
3. 연구 내용 및 결과
4. 결론

1. 서론

주로시설에서연중재배되는오이는생산량이안정적인편이나고온기와저온기의수급에따른가격변동이큰작물이다. 경기도의 경우 2~6월 봄작기와 9~11월 가을작기 재배가 일반적이며 시세가 좋은 1~2월에 출하하기 위해 겨울철 가온 재배를 하거나 봄작기와 가을작기 사이 7~8월에 짧은 여름작기 재배를 하기도 한다. 농가에 따라서는 토마토나 방울토마토 등의 대체 작물 재배로 수익증대를 꾀한다. 비슷한 시기에 작기가 진행되기 때문에 수확량이 집중되는 시기에 도매시장 가격이 연평균 거래가격의 50% 이하로 떨어져 농가 수익이 비용에 못 미치는 경우가 발생하기도 하여 농가에서는 정식을 앞당겨서 출하시기를 조정하기도 한다. 이 경우 난방비 등 추가 비용이 발생하게 되므로 비용상승에 대한 수익구조를 잘 살펴보아야 한다.

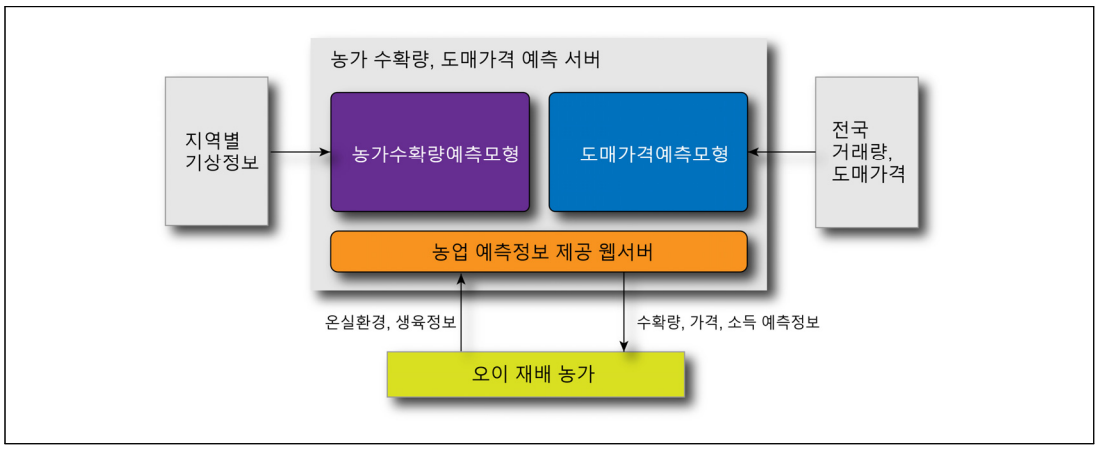
본 연구의 목적은 시설재배 오이의 수확량과 도매가격을 예측하여 농가의 정식시기나 작기 종료시기, 대체 작물 재배 등의 영농 의사결정에 활용하는 것이다.

2. 연구 방법

오이 스마트팜 시설재배 농가의 실측 생육조사 데이터와 온실 환경데이터를 학습데이터로 활용한 인공신경망 딥러닝 모델로 작기 주차별 수확량을 예측하고 농업관측통계시스템에서 제공하는 도매시장 가격정보를 활용한 가격예측 모델로 주차별 거래가격을 예측한다. 예측된 정보는 오이재배 농가에서 웹사이트를 통해 제공받을 수 있으며 예측을 위한 입력정보는 스마트팜 환경센서 수집정보 서버에서 전송되거나 직접입력 혹은 지역별 외부 기상정보를 활용할 수 있도록 한다. 가격예측모형은 전일까지의 도매시장 거래량과 가격으로 예측할 수 있도록 실시간 업데이트 될 수 있어야 한다.

위에서 제시한 서비스 모형 〈그림 1〉이 가능하도록 수확량예측모형과 가격예측모형을 머신러닝 기술을 통해 구현하였으며 먼저 연구에 사용된 자료와 모델링 방법을 살펴보고 예측결과를 실제 결과와 비교해 보도록 하겠다.

그림 1. 수급예측 서비스 모형



2.1. 연구에 사용된 자료

머신러닝 학습을 위해 스마트팜 시설재배 오이의 생육조사 데이터와 온실 환경데이터 그리고 전국 도매시장 경락가격 데이터를 전처리, 가공 및 표준화(평균이 0이고, 표준편차가 1이 되도록 변환)하여 사용하였다. 본 연구결과에서는 머신러닝 "학습"데이터와 "검증"데이터로 표현하기로 한다.

2.1.1. 수확량 예측모형에 사용된 자료

2018~2020년 경기도 오산, 평택, 안성 지역 9개 농가 33개 작기의 온실 환경데이터와 생육조사(경기도 농업기술원) 데이터를 사용하였다(표 1).

표 1. 연구에 사용된 작기 데이터

연도		2018			2019			2020		
작기		봄	여름	가을	봄	여름	가을	봄	여름	가을
봄합계	14	2			5			7		
여름합계	4		-			2			2	
가을합계	15			2			5			8

봄, 가을 각 3개 작기와 여름 1개 작기, 총 7개의 작기 데이터는 검증데이터로 활용하였고 나머지 26개 작기 데이터를 학습데이터로 사용하였다. 검증에 사용된 작기 데이터는 학습에 노출되지 않았다.

가) 온실 환경 데이터

온실 환경데이터는 시설 내·외부의 환경센서에 의해 분 단위로 측정된 데이터 중에서 광합성에 주로 관여하는 항목을 가공하여 활용하였으며 가공 항목과 기준을 〈표 2〉에 정리하였다. 상대습도는 온도에 따른 절대습도와 포화습도를 구한 후 수분부족분으로 변환하여 사용하였다. 1~5분 단위로 측정된 온실 환경데이터는 1주일 단위로 변환하였고 주 1회 진행한 생육조사일 기준으로 이전 7일간의 데이터를 평균 및 합산하여 주단위 환경데이터로 데이터세트를 구성하였다.

표 2. 온실환경 데이터 항목

	센서수집데이터	학습데이터가공	데이터 가공 기준
온실내부환경	내부온도	주간내부온도(℃)	주간 온실 내부온도 평균
	내부상대습도	야간내부온도(℃)	야간 온실 내부온도 평균
	CO2농도	주야간온도차(℃)	일출 전 후 4시간 평균 온도 차이
	내부일사	주간수분부족분(g/m^3)	주간 포화수분과 절대습도 차이
		야간수분부족분(g/m^3)	야간 포화수분과 절대습도 차이
		주간CO2농도(ppm)	일출 후 일몰 전 누적 CO2
		내부일사평균(W/m^2)	일출 후 일몰 전 일사량
		누적내부일사(W/m^2)	5분단위 일사량의 1주일간 누적
온실외부센서	외부온도	주간외부온도(℃)	주간 온실 외부온도 평균
	외부상대습도	야간외부온도(℃)	야간 온실 외부온도 평균
	외부일사	외부일사평균(W/m^2)	일출 후 일몰 전 일사량
		누적외부일사(W/m^2)	5분단위 일사량의 1주일간 누적

나) 생육조사 데이터

주 1회 농가별 3반복 각 3개체의 총 9개 개체에서 수집된 데이터를 가공하여 학습에 사용하였고 항목 및 가공 기준을 〈표 3〉에 정리하였다. 작기진행주차로 정식일 기준으로 주차를 표현하였으며 줄기직경과 엽장/엽폭은 생장점 아래 10번째 마디에서 측정된 것이다.

표 3. 생육조사 데이터 항목

	생육조사데이터	학습데이터가공	데이터 가공 기준
영양생장	정식일 마디수 줄기직경 엽장/엽폭 엽수	작기진행주차(주) 마디수(개) 줄기직경(cm) 엽면적(cm^2) 엽수(개)	정식일 기준으로 주차로 표현 생장점 아래 10번째 마디 경경 측정 마디의 잎으로 면적 근사[1)] 적엽 개수 반영
생식생장	착과수 수확량 적과수 열매길이	착과수(개) 수확량(개)	생육조사 당시의 착과 개수 전 주 열매길이 15cm 이상 수확으로 간주

주: 1) 엽면적 = 0.88×엽장×엽폭-4.27 Blanco, F. "Estimation Of Leaf Area for Greenhouse Cucumber"

2.1.2. 가격예측모형에 사용된 자료

농업경제연구원 농업관측 통계정보시스템에서 제공하는 전국도매시장 경락가격자료 raw 데이터를 사용하였으며 '06~'19년 자료를 학습데이터로, '20년 자료를 검증데이터로 활용하였다. 전체 거래량의 99%를 차지하는('20년 기준) 백다다기, 가시오이, 취청 3개 품종의 전국도매시장 거래량과 거래가격 데이터를 가공하여 사용하였다. 〈표 4〉는 가격예측모델 학습에 사용된 항목이다. 거래가격은 단위기간 거래금액의 합을 거래량의 합으로 나누어 산출하였으며 물가상승분은 별도로 학습에 반영하지 않았다.

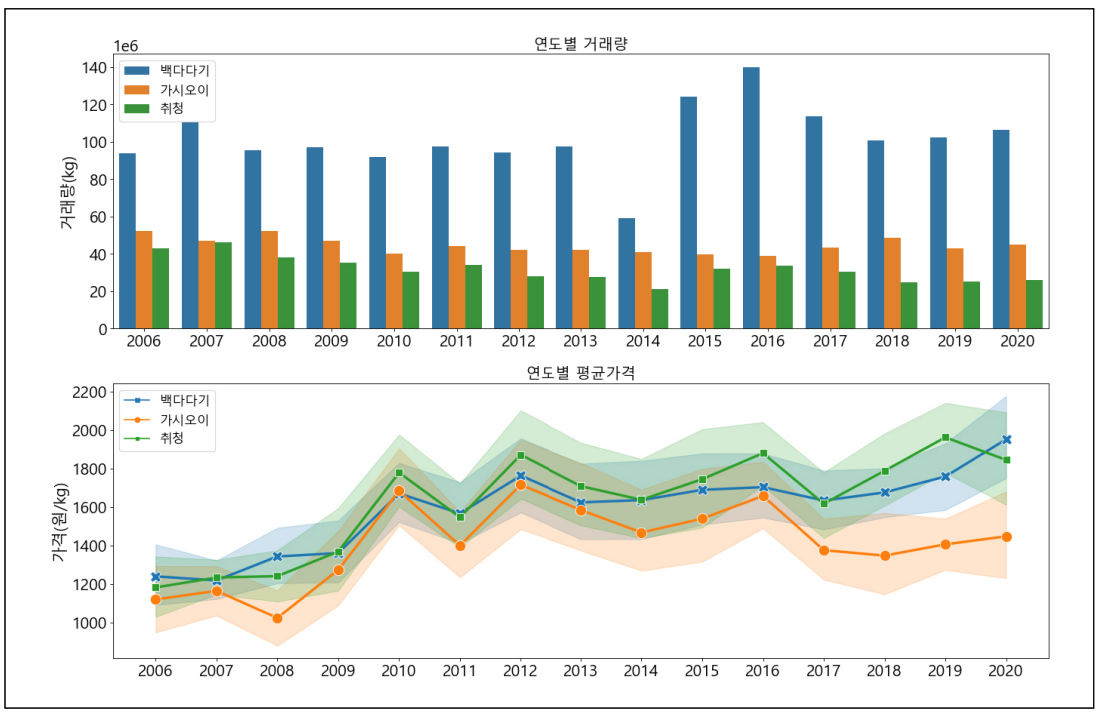
표 4. 가격예측모델 데이터 항목

	도매가격정보	학습데이터가공	데이터 가공 기준
전국도매시장 경락가격정보	거래년월일 품목명 출하지명 평균가격 거래량 단위코드	거래년월일 품목명 출하지역 거래가격(원/kg) 거래량(kg) 거래금액(원) 수요량지표	일일 거래량으로 모델 훈련 및 검증 백다다기, 가시오이, 취청 3개 품목 전국 13개 시도로 변환 일거래금액 ÷ 일거래량 일합계[거래량 × 평균가격] 일거래금액 ÷ 연평균거래가격

수요량지표는 수요가 예측에 반영되도록 하는 지표로써 거래금액을 연평균 거래가격으로 나누어 가격에 의한 기대수요량을 학습에 반영될 수 있도록 하였다(그림 8). 품종별 연간 거래

량은 백다다기는 증가하는 추세이고 취청은 감소하는 추세임을 〈그림 2〉에서 확인할 수 있으며 2012년 이후로 가격 변동은 크지 않음을 알 수 있다. 물가상승분은 거래량에 의한 가격예측에 왜곡이 있을 수 있어 배제하였다.

그림 2. 오이 품종별 연간 거래량, 가격 추이



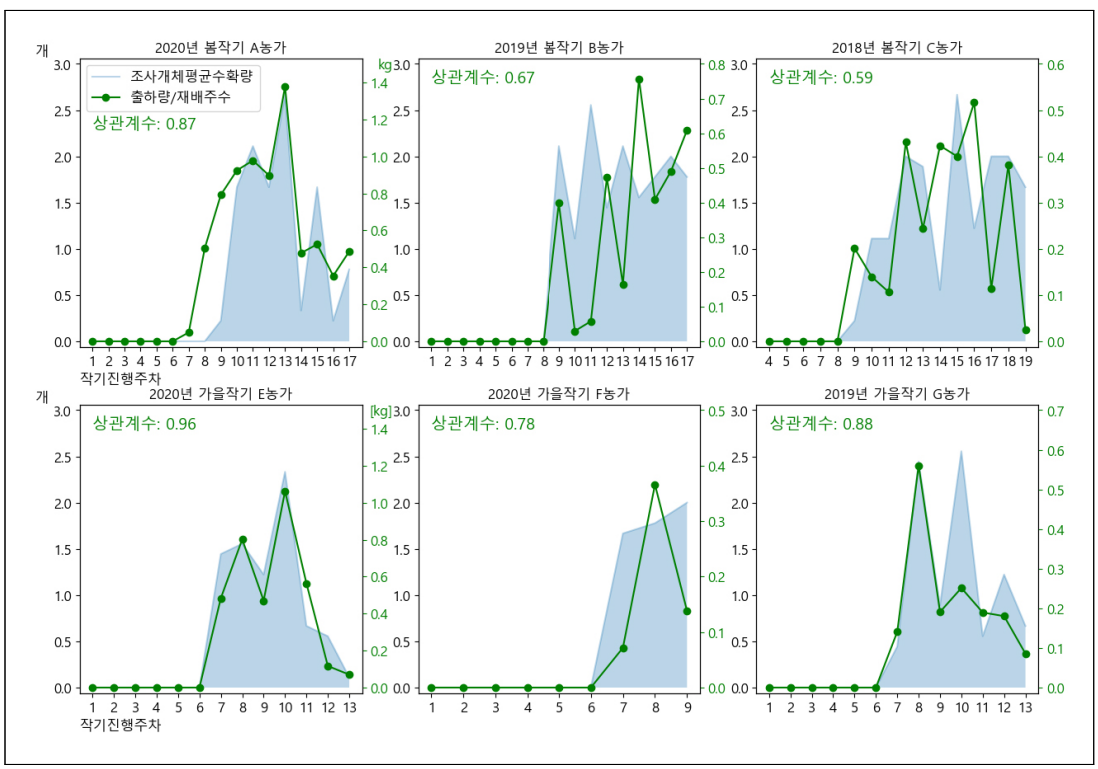
2.2. 모델링 방법

수확량과 가격예측 모델링은 파이썬 라이브러리를 활용한 머신러닝으로 진행하였고 딥러닝 신경망 모델과 Random Forest, SVR(Support Vector Regressor) 모델 결과를 비교하였다. 신경망 모델링 환경으로는 파이썬 케라스(Keras)-텐서플로(TensorFlow)엔진을 사용하였고, SVR(Support Vector Regressor)과 RF(Random Forest) 모델은 사이킷런(Scikit-learn) 라이브러리를 사용하였다.

2.2.1. 오이 수확량 예측 모형

학습에 사용된 수확량 데이터는 조사 개체로부터 수집한 샘플 데이터이며 실제 출하량과의 상관관계를 먼저 확인해보았다. 〈그림 3〉의 왼편 y축은 1개 조사 개체당 수확량을 나타내고, 오른편 y축은 1개체당 출하중량(kg)을 나타내며 농가의 주별 출하중량(kg)을 재배주수로 나눈 값이다. 출하내역은 관할 농협에 등록된 출하량만 보여주기 때문에 직거래 등의 미등록 내용이 있을 수 있으며 재배주수에 포함되지 않은 동일 농장주의 내역이 함께 포함되어 있을 수 있다. 따라서 그래프 스케일을 조사개체 수확량과 비슷하게 맞추어 주차별 추세를 확인할 수 있도록 하였다. 작기별 조사개체 샘플 수확량과 출하량의 상관계수는 평균 0.692이며 상관관계가 있다고 판단하고 학습을 진행하였다.

그림 3. 조사개체 수확량과 농가 출하량 비교



수확량 예측모형은 재배환경이 상이한 봄작기, 여름작기, 가을작기 모형을 각각 학습하였으며 작기주차에 따른 착과수와 수확량을 예측하도록 학습되었다. 〈표 5〉에서와 같이 입력변수를 구분하여 세 가지 상황을 가정하였다. 먼저 온실내부 환경정보와 생육정보를 모두 입력변수로 활용하는 경우로 작기 진행 중에 수확량을 예측하는 상황을 가정하였다. 두 번째는 온실내부 환경정보만을 입력변수로 활용하는 경우로 과거의 데이터나 온실의 평균적인 환경데이터를 활용하여 수확량을 예측하는 경우이다. 마지막으로 외부환경정보만을 입력변수로 사용하는 경우는 온실에 환경 센서가 없어 외부 기상조건만을 변수로 사용할 수 있는 경우이다. 두 번째와 세 번째의 경우는 예측서비스 제공 시 따로 입력변수를 입력하지 않고 서버로 기상정보나 환경데이터를 시스템을 연동하여 데이터를 전송하는 방식으로 구현 가능할 것이다.

표 5. 입력변수 항목 구분

구분	입력변수
1. 온실내부환경과 생육정보를 입력변수로 사용	주간내부온도, 야간내부온도, 주야간온도차, 주간수분부족분, 야간수분부족분, 주간CO2농도, 내부일사평균, 누적내부일사, 작기진행주차, 마디수, 줄기직경, 엽면적, 엽수
2. 온실내부환경만 입력변수로 사용	주간내부온도, 야간내부온도, 주야간온도차, 주간수분부족분, 야간수분부족분, 주간CO2농도, 내부일사평균, 누적내부일사, 작기진행주차
3. 외부환경만 입력변수로 사용	주간외부온도, 야간외부온도, 주간/야간 외부상대습도 외부일사평균, 누적외부일사, 작기진행주차

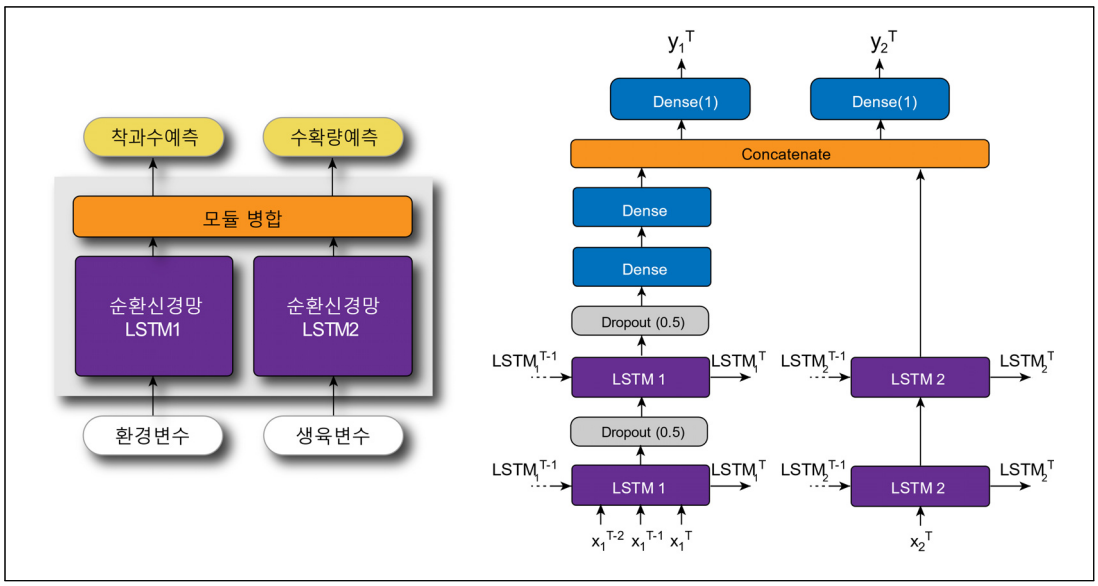
모델링은 시계열 분석을 위해 순환신경망 LSTM(Long Short-Term Memory)을 사용하였으며 〈그림 4(a)〉에서와 같이 환경변수를 처리하는 순환신경망 모듈과 생육변수를 처리하는 순환신경망 모듈을 따로 구성하여 출력 전 단계에서 병합하였다. 순환신경망의 layer 구성은 〈그림 4(b)〉에 표현하였다. 환경변수 순환신경망 모듈은 2개의 LSTM층과 2개의 Dense층으로 구성되어 있으며 각 LSTM층에 Dropout(=0.5)층을 추가하여 과대적합 감소를 꾀하였다. 생육변수 순환신경망 모듈은 입력변수로 생육변수와 작기진행주차 정보를 받는다. LSTM 2개층으로 구성되며 Dropout층은 추가하지 않았다.

두 순환신경망 모듈의 출력은 Concatenate층에서 병합되어 2개의 Dense출력층 -착과수와 수확량- 을 거치게 된다. 모델의 모든 Dense층은 활성화 함수로 'Relu' activation을 사용하였고 출력층 Dense도 'Relu' activation을 통해 음수값은 0이 되도록 하였다. 비용함수

(Cost Function)를 최소화하기 위한 옵티마이저로 Adam optimizer를, loss는 MSE(Mean Squared Error, 평균제곱오차)를 사용하였다.

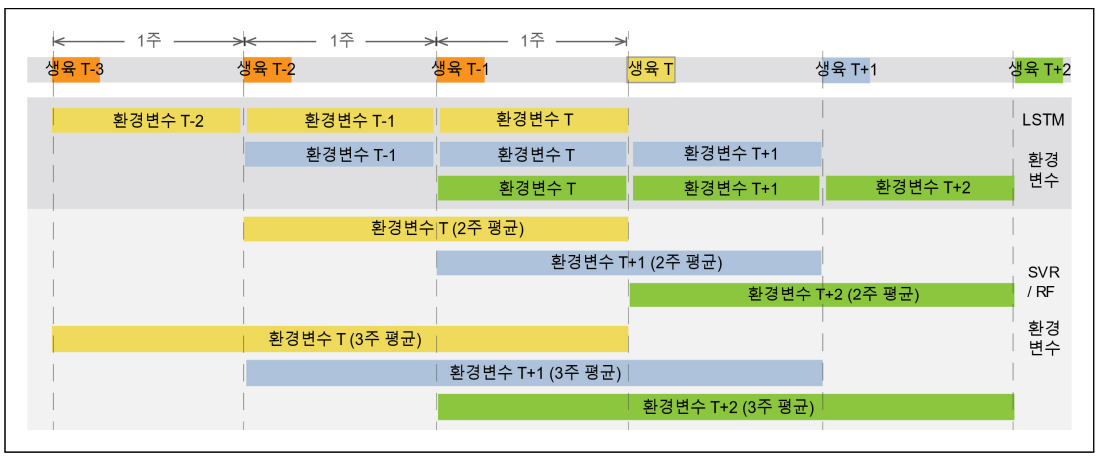
〈그림 4(b)〉에서 입력 x_1^{T-2},x_1^{T-1},x_1^{T}는 환경입력변수이며 각각 관측 시점의 3주전부터의 1주평균 데이터이고, x_2^{T}는 생육입력변수, y_1^{T}과 y_2^{T}는 출력인 착과수와 수확수를 의미한다.

그림 4. 수확량 예측 순환신경망 모델



〈그림 5〉는 환경입력변수와 생육입력변수의 시계열 데이터 시퀀스를 표현한 것이다. 환경변수 순환신경망 모델은 시계열 데이터 시퀀스로 3개의 타임스텝 x_1^{T-2},x_1^{T-1},x_1^{T}(하나의 타임스템은 1주)를 하나의 학습데이터로 입력하게 되며 다음 학습데이터는 한 스텝씩 Shfit한 x_1^{T-1},x_1^{T},x_1^{T+1}를 입력하게 된다. 비교모델로 사용한 SVR과 RF모델의 입력변수는 시계열 시퀀스가 아니므로 환경변수를 1주, 2주평균, 3주평균으로 가공하여 학습한 후 예측 결과를 비교하였다.

그림 5. 입력변수의 시계열 시퀀스



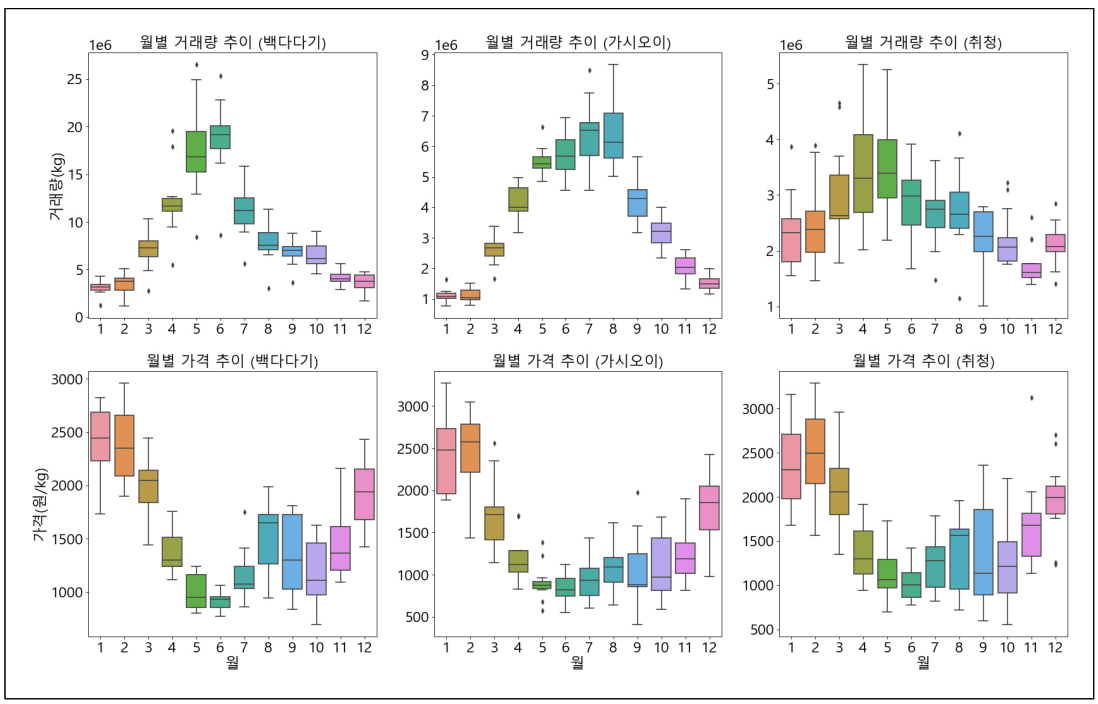
2.2.2. 가격예측 모형

농산물 가격예측은 ARIMA모형 등의 시계열예측으로 주로 진행되어 왔고 최근에는 순환신경망(Recurrent Neural Network)을 이용한 시계열예측 연구가 활발히 이루어지고 있다. 기상변수와, 유가, 전년 수확량 등의 변수를 활용한 LSTM 순환신경망 모델로 향후 7일 가격을 예측한 연구가 있으며, 기상변수와 도매시장 가격을 변수로 RNN 순환신경망을 적용하여 배추 가격을 예측한 연구가 있다.

본 연구에서는 기상변수를 사용하지 않고 도매시장 거래량과 가격 자료만으로 중장기 가격을 예측할 수 있는 모델을 적용하여 검증해보았다. 입력 변수를 최소화하고 학습 비용이 낮은 모델을 사용하여 요청이 들어왔을 때 실시간으로 최신정보를 학습하여 웹서비스를 통해 가격 예측결과를 제공할 수 있는 모델을 연구 방향으로 설정하였다.

〈그림 6〉의 '06~'20년 월별 거래량과 가격 추이를 보면 뚜렷한 계절성과 품종별로 거래량 추이가 상이한 것을 확인할 수 있다. 본 연구에서는 이러한 품종별 거래량과 가격의 계절성을 예측에 적용하였다.

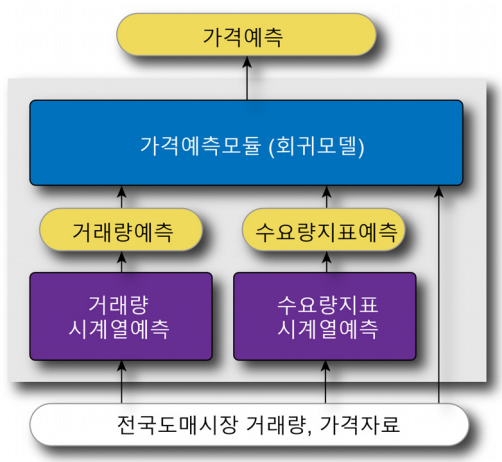
그림 6. 오이 품종에 따른 월별 거래량 및 가격 추이('06~'20년)



가격예측모형 〈그림 7〉은 거래량과 가격, 그리고 계절성 반영을 위한 거래일자를 입력변수로 받아 가격을 예측하며 거래량 시계열예측, 수요량지표 시계열예측, 가격예측의 3개 모듈로 구성되어 있다. 시계열예측 모듈은 파이썬 Prophet 라이브러리를 활용하였다. Prophet은 페이스북에서 개발한 시계열예측 라이브러리로 학습속도가 빠르고 계절성이 있는 모델 예측에 높은 성능을 보이며 계절성, 트랜드 등의 파라미터로 모델 최적화가 가능하다. 가격예측 모듈은 회귀모델이며 시계열예측 모듈에서 예측된 거래량과 수요량지표, 그리고 거래일자를 입력변수로 최종 가격을 예측한다. 본 연구에서는 SVR, RF, Lasso회귀모델의 결과를 비교하였다. 각각의 회귀모델은 사이킷런의 GridSearchCV를 활용하여 최적의 하이퍼파라미터[1]를 적용하였다.

1) SVR: kernel="rbf", C=1000, gamma=1
RF: n_estimator=100
Lasso Regressor: Polynomial degree=5, alpha=0.01(백다다기), alpha=1(가시오이, 취청)

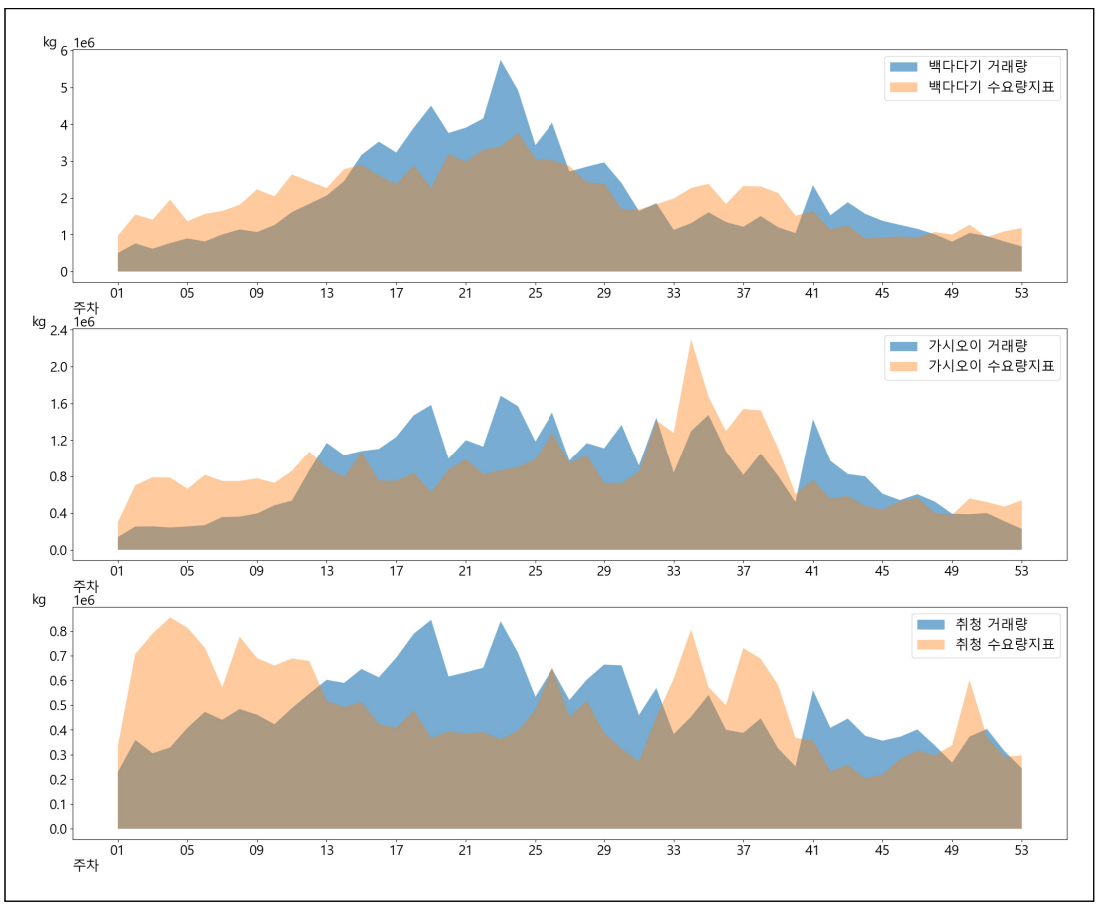
그림 7. 가격예측 모형



가격예측에 수요를 반영하기 위해 '2.1.2. 가격예측모형에 사용된 자료'에서 언급된 수요량지표를 활용하였다. 수요량지표 (1)은 기준가격(연평균거래가격)에서의 기대거래량을 예상해보는 방법을 사용하였다. 〈그림 8〉에서 파란색 영역이 실제 거래량(공급물량)이며 수요량지표를 나타내는 주황색 영역이 더 높은 경우는 수요에 비해 공급물량이 부족하다고 볼 수 있고, 반대로 파란색 영역이 더 높은 경우는 수요보다 공급이 많아 가격이 기준가격 보다 낮아진 것으로 추정할 수 있다.

$$(1)\ \text{수요량지표} = \text{거래량} \times \frac{\text{거래가격}}{\text{연평균거래가격}}$$

그림 8. 주차별 거래량과 수요량지표 비교(2020년)



연구 결과에서는 거래량과 수요량지표의 시계열예측결과를 오이 품종별로 확인해보고 예측된 거래량과 수요량지표를 통해 가격예측모듈에서 최종 예측된 결과를 회귀모델 별로 비교해보았다.

3. 연구 내용 및 결과

수확량예측모형과 가격예측모형은 시계열예측 모형이므로 머신러닝을 위한 학습세트와 검증세트를 무작위 추출하지 않고 작기와 기간을 선택하여 데이터의 시퀀스를 유지하였다. 수확량 예측을 위한 검증세트는 작기별로 연도와 농가가 골고루 검증에 포함될 수 있도록 구성하였고, 가격예측을 위한 데이터세트는 '06~'19년 데이터를 학습세트로, '20년 데이터를 검증세트로 구성하여 결과를 확인하였다.

검증결과의 성능 지표는 (2) MAE(절대오차평균, Mean Absolute Error)와 (3) 결정계수 R^2(결정계수, R-Squared)을 사용하였다.

(2) $MAE = \frac{\sum|y-\hat{y}|}{n}$

(3) $R^2 = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}$

y는 관측값, $\bar{y}$는 관측값의 평균, $\hat{y}$는 예측값을 나타낸다. 성능 지표는 검증세트의 지표만 표기하였다.

3.1. 오이 수확량 예측 결과

수확량 예측모형으로 작기진행주차별 착과수와 수확량을 예측하였으며, 결과의 값은 오이 1개체에서 관측되는 열매의 개수이다. 예측 모델은 봄작기, 여름작기, 가을작기로 나누어 학습하였으며 학습에 노출되지 않은 작기 데이터로 검증을 진행하였다. 검증은 봄 3작기, 여름 1작기, 가을 3작기 데이터를 사용하였다.

33개 작기의 총 4,488개의 데이터세트를 사용하였으며 학습세트 80%, 검증세트 20%로 구성되어있다(표 6).

표 6. 수확량 예측에 사용된 학습세트와 검증세트 개수

구분	봄작기	여름작기	가을작기
학습세트	2,021	240	1,379
검증세트	462	65	321
합계	2,483	305	1,700

수확량 예측 테스트는 아래의 항목으로 진행하였고 결과를 그래프와 예측성능으로 정리하였다.

가) 계절별 착과수와 수확량 예측

환경조건이 상이한 봄, 여름, 가을 작기를 구분하여 예측 결과를 확인하였다.

- 봄 3작기(A, B, C농가): 착과수 〈그림 9〉, 〈표 7〉, 수확량 〈그림 10〉, 〈표 8〉
- 여름 1작기(D 농가): 착과수, 수확량 〈그림 11〉, 〈표 9〉
- 가을 3작기(E, F, G농가): 착과수 〈그림 12〉, 〈표 10〉, 수확량 〈그림 13〉, 〈표 11〉

LSTM 순환신경망 모델이 작기별로 안정적인 예측결과를 보여주는 반면 SVR, RF 모델은 검증세트에 따라 편차가 큰 것을 확인할 수 있다.

나) 예측모형에 따른 결과 비교

LSTM 순환신경망(그래프의 1행), SVR(2행), RF(3행) 모델의 예측결과를 비교하였다. SVR과 RF 모델은 싸이킷런의 GridSearchCV를 활용하여 최적의 하이퍼파라미터를 적용하였다.

LSTM 순환신경망 모델은 1주일 간격의 3개의 환경데이터를 하나의 시퀀스 데이터로. SVR, RF 모델은 입력 환경변수를 1주일, 2주평균, 3주평균으로 각각 학습 및 검증하였고, 그 중 결과가 가장 좋은 3주평균의 예측 결과를 비교하였다.

SVR, RF 모델의 경우 전제 데이터세트에서 검증세트를 무작위 추출하여 검증하는 교차검증(Cross Validation) 결과의 성능 수치는 LSTM 순환신경망 모델과 비슷한 수준으로 나왔으나 시퀀셜 검증 데이터의 경우 낮은 예측 성능을 보였다.

다) 입력변수에 따른 결과 비교

결과 그래프의 파란색 영역은 실제 값이며 주황색은 온실 내부환경변수와 생육변수를 모두 입력변수로 예측한 결과, 초록색은 내부환경변수로만 예측한 결과이며, 빨간색은 외부환경변수로만 예측한 결과이다. 내부환경변수와 생육변수를 모두 사용한 결과의 예측 성능이 높게 나타났으며 SVR, RF 모델은 입력변수에 따른 편차가 매우 큰 것을 확인할 수 있다.

그림 9. 봄작기 착과수 예측 결과

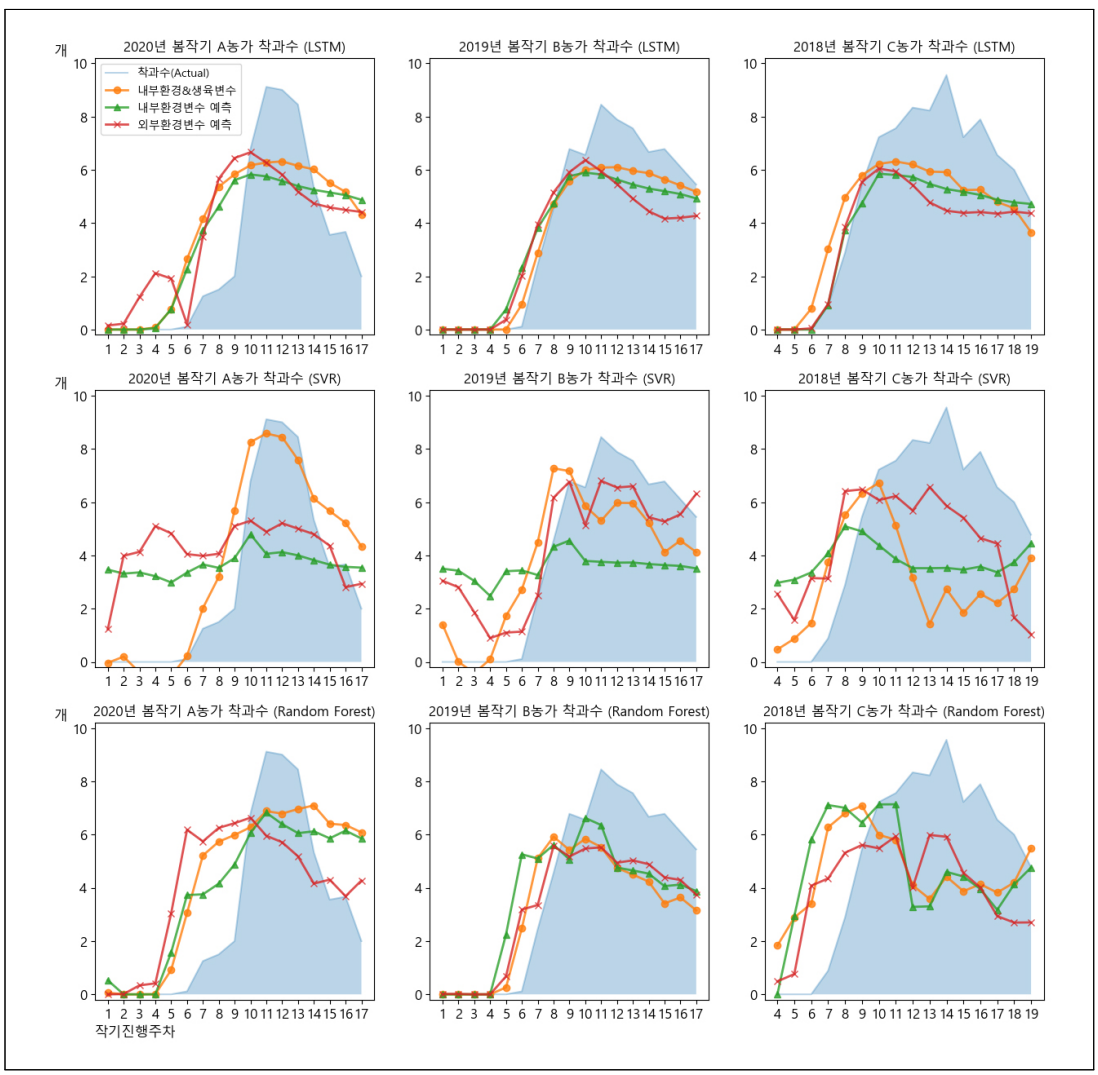


표 7. 봄작기 착과수 예측성능 비교

봄작기 착과수 예측	LSTM		SVR		Random Forest	
	MAE	R^2	MAE	R^2	MAE	R^2
내부환경&생육변수	1.39	0.69	0.58	0.45	0.61	0.38
내부환경변수	1.52	0.64	0.85	0.11	0.71	0.14
외부환경변수	1.82	0.53	0.64	0.39	0.68	0.29

그림 10. 봄작기 수확량 예측 결과

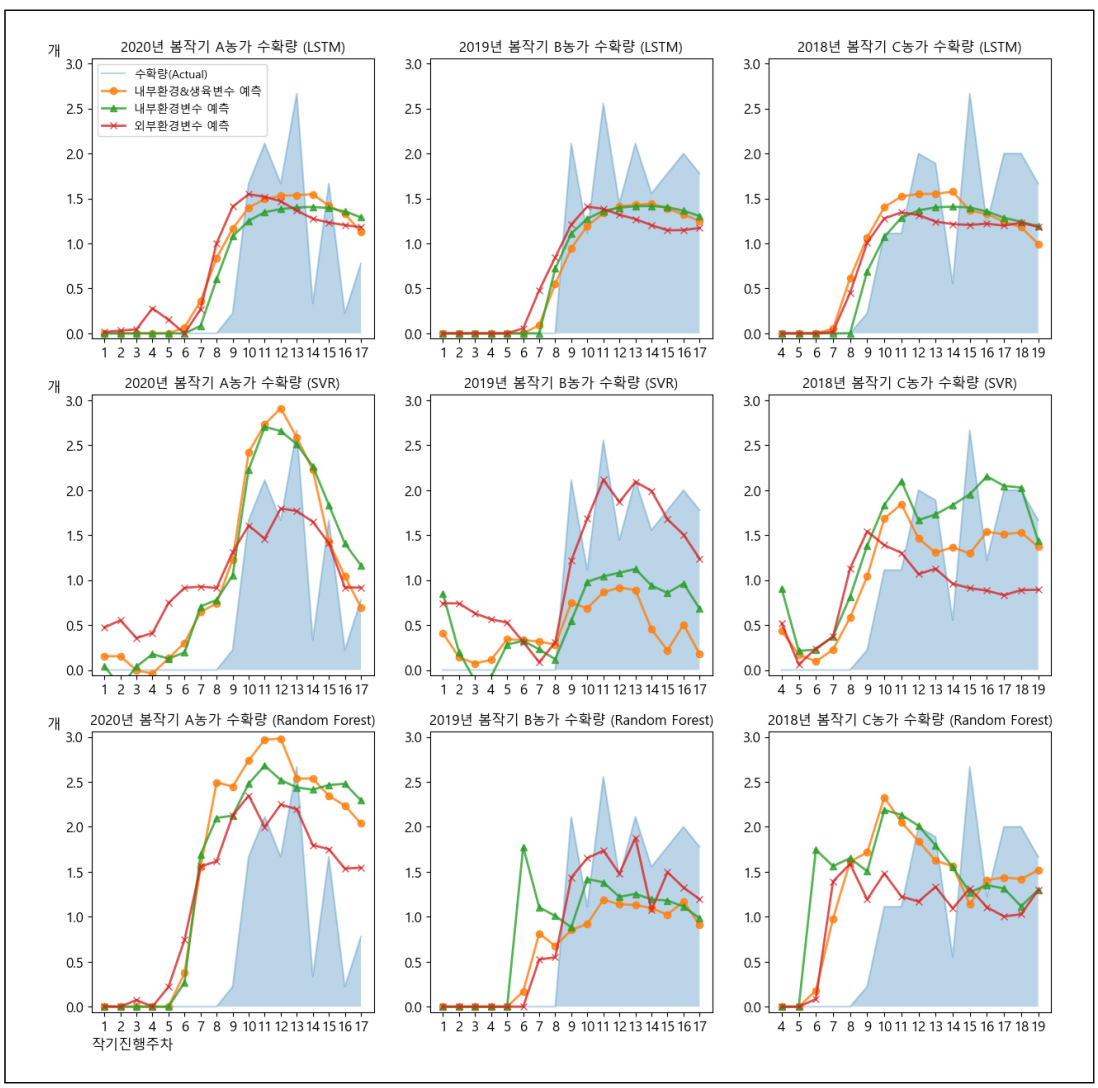


표 8. 봄작기 수확량 예측성능 비교

봄작기 수확량 예측	LSTM		SVR		Random Forest	
	MAE	R^2	MAE	R^2	MAE	R^2
내부환경&생육변수	0.58	0.43	0.64	0.21	0.77	−0.09
내부환경변수	0.57	0.43	0.65	0.27	0.88	−0.17
외부환경변수	0.64	0.37	0.70	0.30	0.67	0.20

그림 11. 여름작기 착과수, 수확량 예측 결과

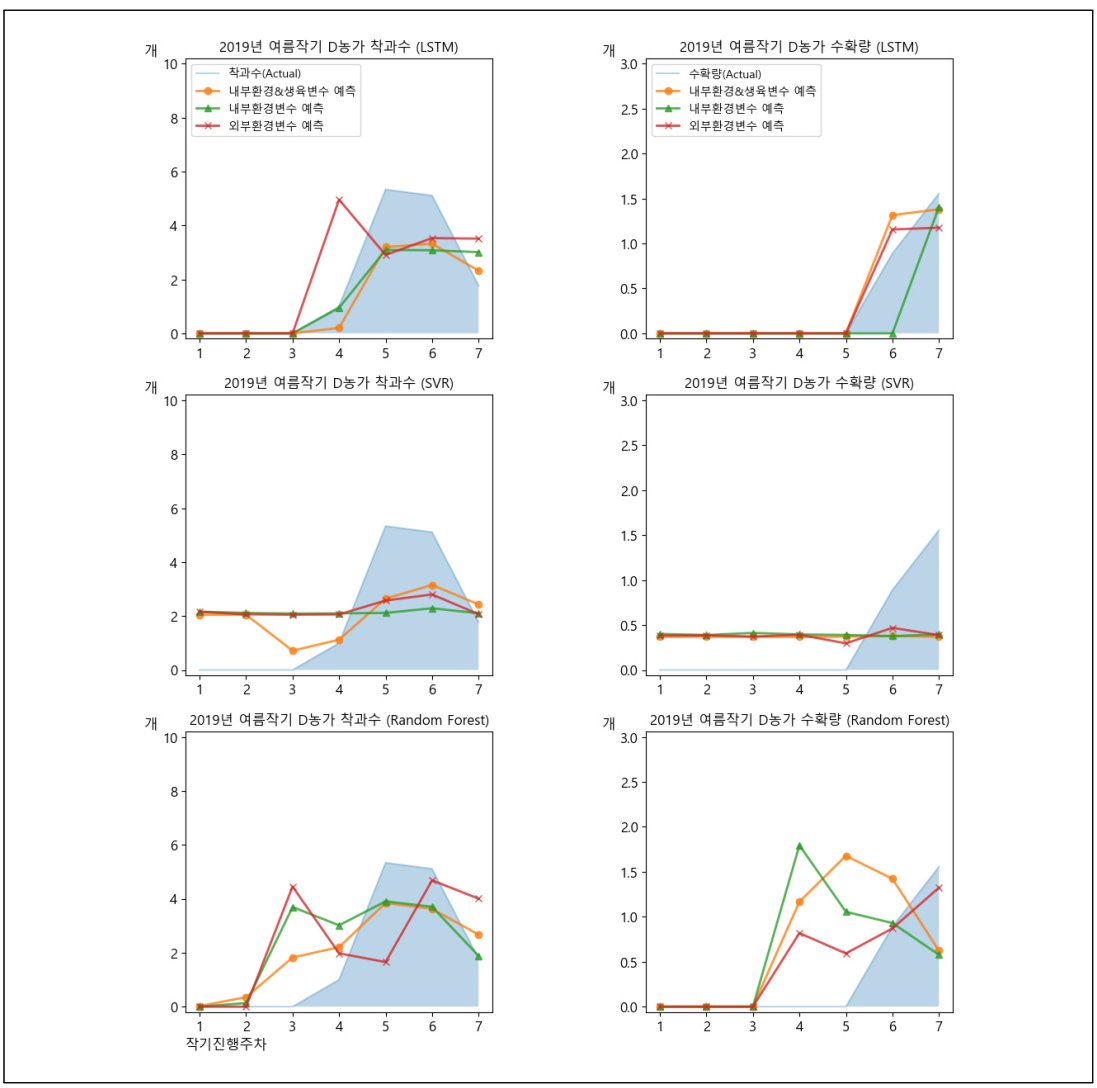


표 9. 여름작기 착과수, 수확량 예측성능 비교

여름작기 예측		LSTM		SVR		Random Forest	
		MAE	R^2	MAE	R^2	MAE	R^2
착과수	내부환경&생육변수	0.93	0.61	0.63	0.45	0.52	0.54
	내부환경변수	0.99	0.55	0.96	0.23	0.90	0.33
	외부환경변수	1.54	0.16	1.86	0.09	0.57	0.55
수확량	내부환경&생육변수	0.20	0.56	0.73	−0.07	0.73	−0.01
	내부환경변수	0.23	0.39	0.81	0.28	0.78	0.06
	외부환경변수	0.23	0.52	0.72	−0.03	0.65	0.22

그림 12. 가을작기 착과수 예측 결과

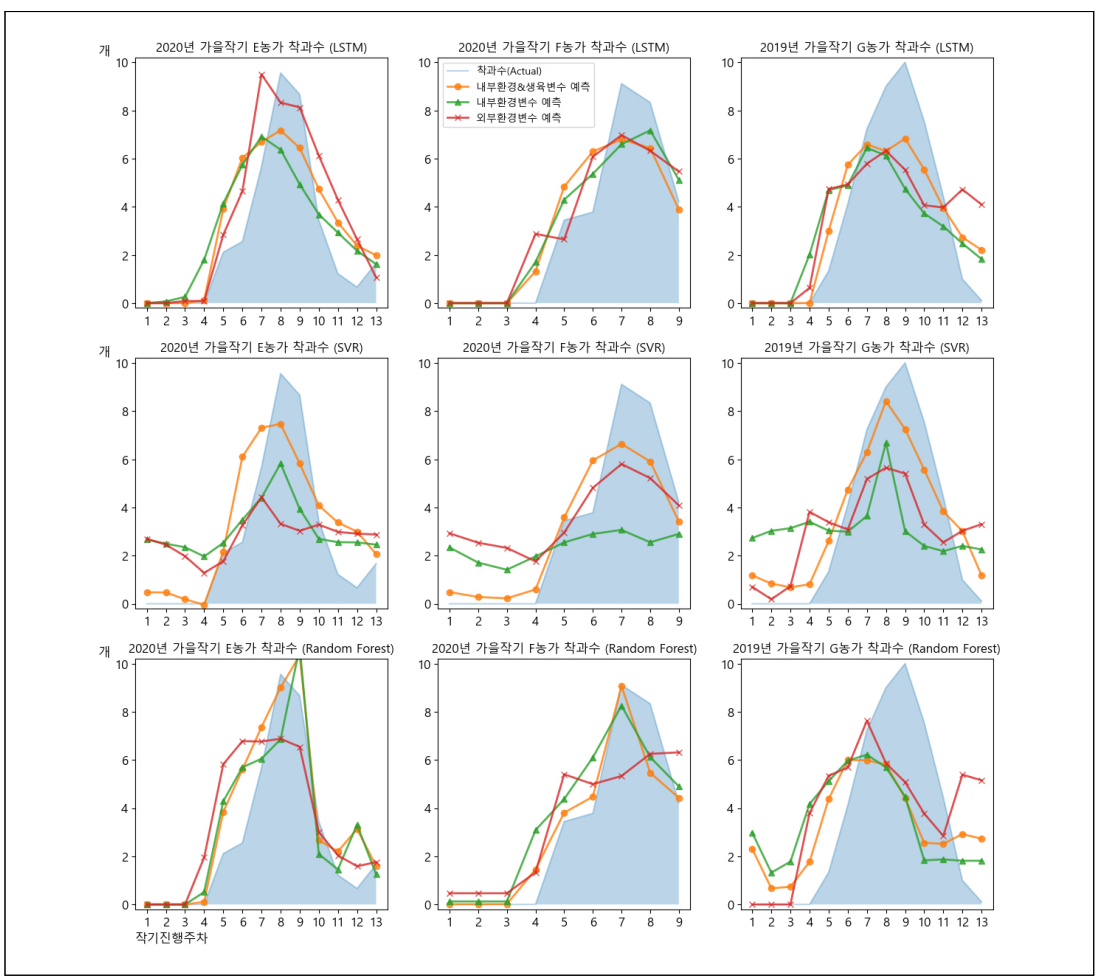


표 10. 가을작기 착과수 예측성능 비교

가을작기 착과수 예측	LSTM		SVR		Random Forest	
	MAE	R^2	MAE	R^2	MAE	R^2
내부환경&생육변수	1.50	0.64	0.41	0.62	0.49	0.53
내부환경변수	1.70	0.54	0.69	0.19	0.55	0.42
외부환경변수	1.79	0.51	0.64	0.27	0.55	0.40

그림 13. 가을작기 수확량 예측 결과

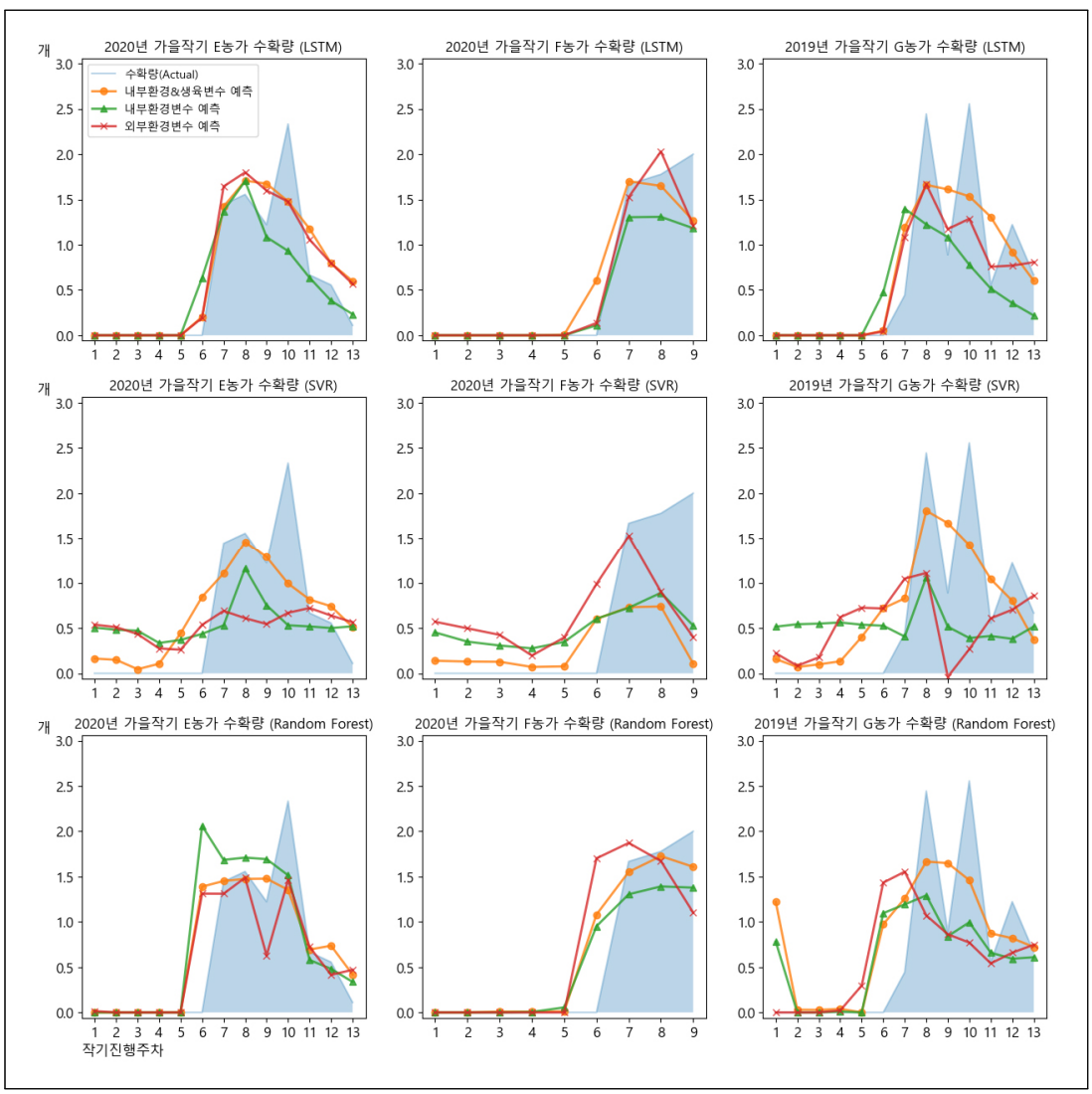


표 11. 가을작기 수확량 예측성능 비교

가을작기 수확량 예측	LSTM		SVR		Random Forest	
	MAE	R^2	MAE	R^2	MAE	R^2
내부환경&생육변수	0.47	0.46	0.56	0.28	0.57	0.29
내부환경변수	0.51	0.34	0.70	0.10	0.56	0.24
외부환경변수	0.45	0.45	0.71	0.06	0.58	0.20

3.2. 오이 가격예측 결과

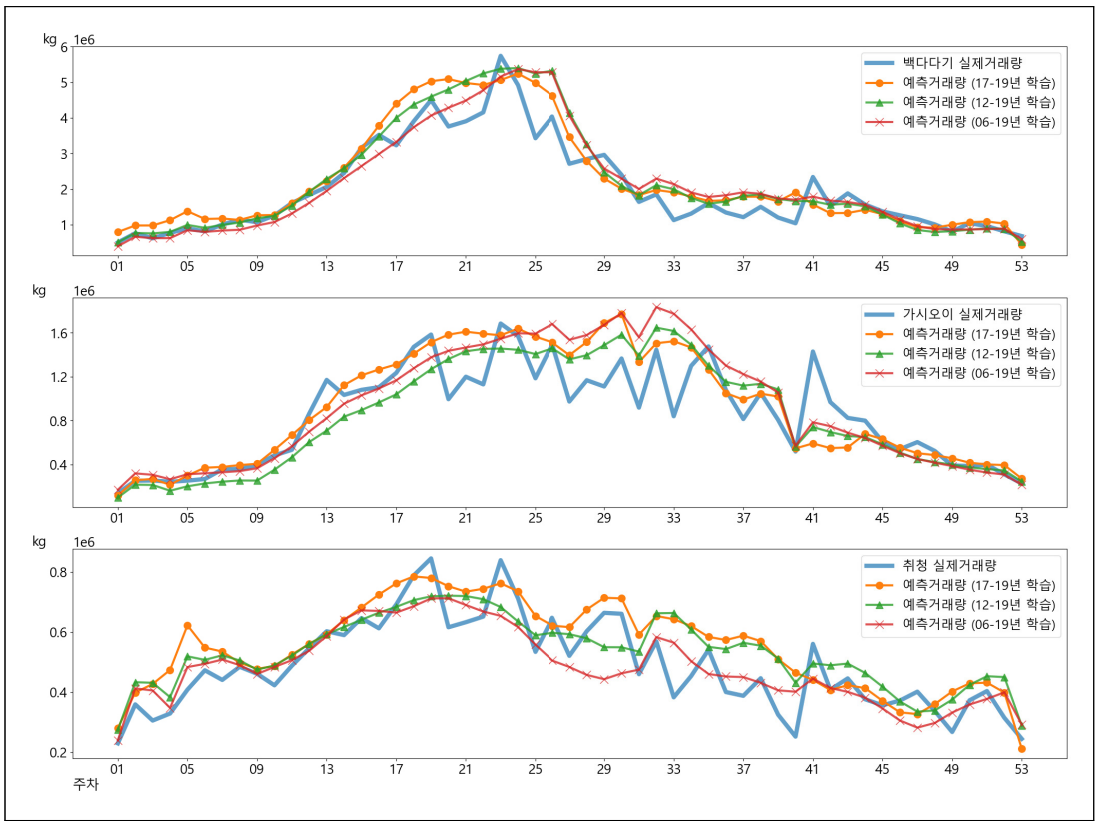
가격예측은 품종별로 결과를 확인하였다. 〈표 12〉는 가격예측에 사용된 '06~'20년 데이터 세트의 개수이다.

표 12. 가격예측에 사용된 학습세트와 검증세트 개수

구분	백다다기	가시오이	취청
학습세트	4,594	4,601	4,599
검증세트	366	366	366
합계	4,960	4,967	4,965

거래량과 수요량지표의 시계열예측 결과를 그래프로 확인하고 최종 가격예측 결과를 그래프와 예측 성능수치로 확인하였다.

그림 14. '20년 거래량 시계열예측 결과



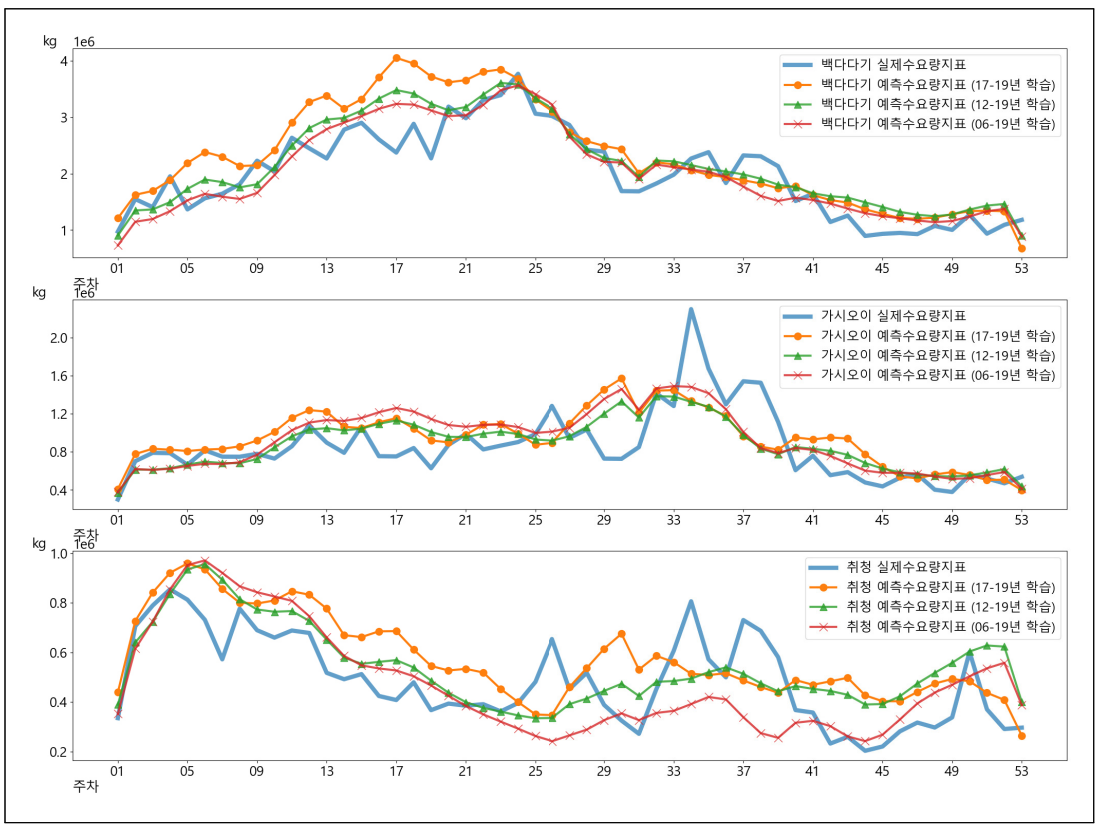
가) 거래량과 수요량지표 시계열예측

거래량과 수요량지표는 30일 단위로 학습데이터를 업데이트하여 예측하였다. 시계열예측은 과거 데이터의 계절성과 트렌드를 반영하므로 학습데이터의 기간을 '06~'19년, '12~'19년, '17~'19년으로 변경하여 예측 결과를 비교하였다. 〈그림 14〉는 거래량 시계열예측 결과이며 취청오이의 경우 학습데이터의 기간에 따라 예측 결과의 차이가 크게 나타나는 것을 확인할 수 있다. 예를 들어 '17~'19년 여름철(27~30주차) 취청오이의 생산량이 이전보다 많았다고 해석이 가능하다.

〈그림 15〉의 수요량지표도 마찬가지로 학습데이터의 기간을 변경하여 확인하였으며 백다다기의 경우 '17~'19년 학습데이터의 경우 상반기 지표가 실제보다 높게 나오는 것을 확인할

수 있으며 이는 '17~'19년 상반기에 백다다기의 가격이 다소 높았음을 예측할 수 있다. 자기상관을 이용하는 시계열예측의 경우 결과의 정확도를 높이기 위해 학습데이터의 기간 설정 시 트렌드를 감안한 최적화가 필요하다는 것을 알 수 있다.

그림 15. '20년 수요량지표 시계열예측 결과



나) 가격예측 결과

앞서 시계열예측 모듈에서 예측된 거래량과 수요량지표를 입력변수로 사용하여 가격예측 모듈에서 예측한 결과를 회귀모델별(SVR, RF, Lasso회귀)로 비교하였다. 〈그림 16〉은 '17~'19년 데이터로 학습하여 '20년 데이터로 검증한 결과이며 전체적인 가격추세를 제대로 예측하고 있으나 가격 피크구간의 예측은 미흡한 것으로 나타났다.

〈표 13〉은 모델별 예측 성능 비교이며 SVR모델의 성능이 비교적 높게 나왔다. 품종별 수요와 공급의 시계열예측 최적화에 대한 추가적인 연구가 필요하다.

그림 16. '20년 수요량지표 시계열예측 결과

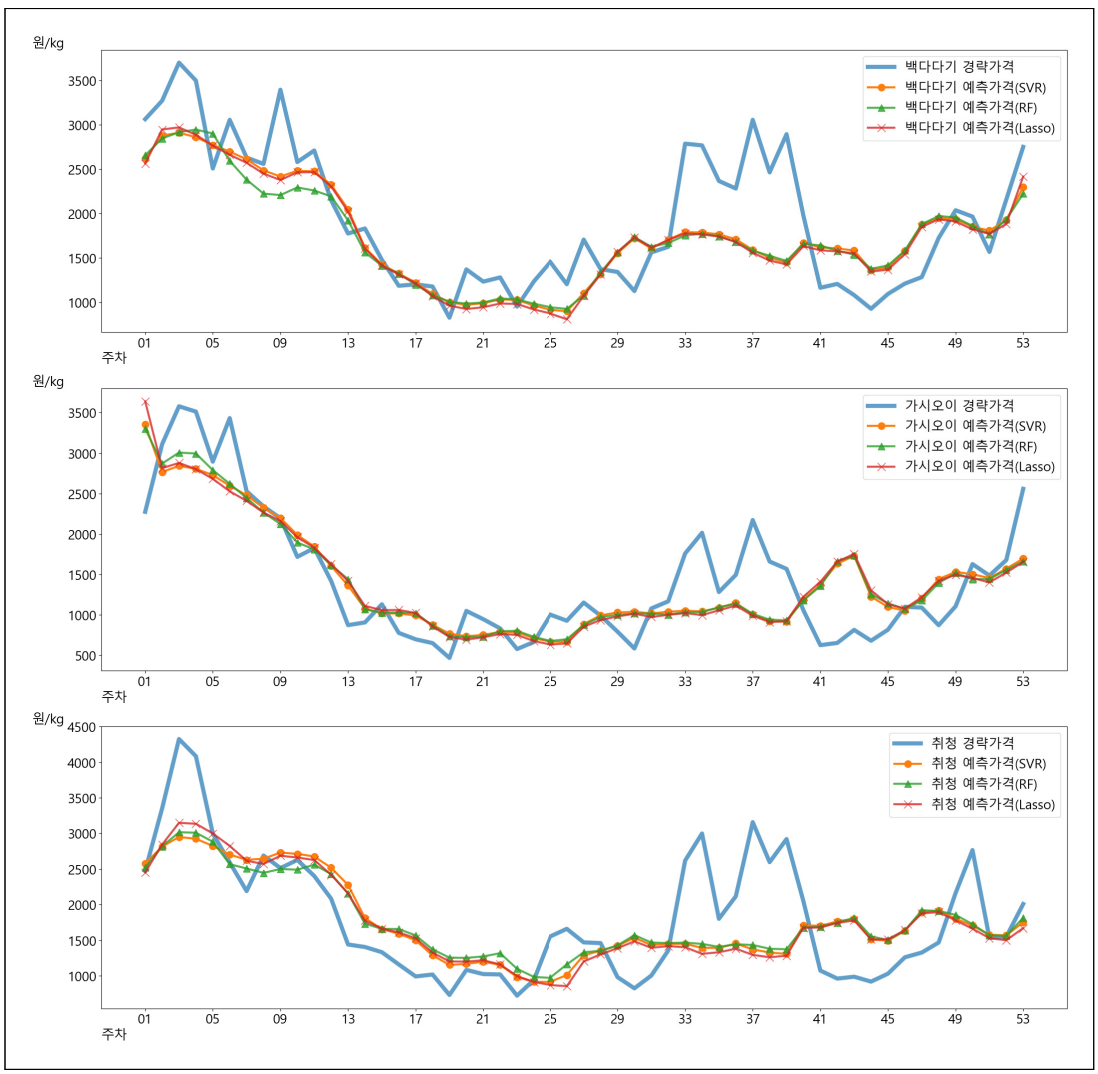


표 13. 가격예측 성능 비교

품종	SVR		Random Forest		Lasso Regressor	
	MAE	R^2	MAE	R^2	MAE	R^2
백다다기	457.08	0.44	511.47	0.26	518.85	0.28
가시오이	568.23	0.30	594.29	0.23	639.82	0.07
취청	455.16	0.50	426.17	0.56	487.85	0.42

4. 결론

머신러닝을 통해 시험 재배가 아닌 실제 재배 환경에서의 데이터 학습으로 비교적 높은 예측 성능을 얻을 수 있는 것을 확인하였다. 연구에 사용된 생육데이터는 병충해, 기상 악조건 등 현실의 모든 환경 변수가 녹아 있는 결과이며 관측데이터가 축적된다면 더 정확한 예측이 가능할 것이다. 다만 일관된 기준으로 정확하게 수집된 데이터가 머신러닝의 성능을 효과적으로 끌어올릴 수 있으므로 실측데이터 수집 계획 단계에서 장기적인 활용 방안이 고려되어야 하겠다.

LSTM 순환신경망이 SVR, RF 모델보다 모든 검증세트에서 안정적인 예측 결과를 보여주었으며 앞으로 순환신경망 기술을 활용한 농업 분야의 시계열예측 연구가 활발히 진행될 것이라 생각한다.

본 연구의 수확량 예측모형은 경기도 시설재배 농가의 데이터를 사용하였으며 향후 전국에서 수집된 데이터를 적용한다면 더 정확한 결과와 다양한 인사이트를 얻을 수 있을 것이다. 센서의 유지보수 등을 통해 정확한 환경데이터가 수집되도록 하는 것도 농업 수급예측 기술 발전의 중요한 요인이 될 것이다.

가격예측모형의 경우 수요와 공급의 시계열 예측 부분을 최적화한다면 실시간 업데이트가 가능한 점과 최소한의 정보로 중장기 예측이 가능하다는 장점으로 예측 정보 제공 서비스에 활용도가 높을 것으로 생각되며 오이 외 다른 농작물에 적용하는 것도 어렵지 않을 것이다.

스마트팜 농가가 확대되고 데이터가 축적되어 연동된 시스템에서 빅데이터를 스스로 학습시키고 활용하는 농업 수급예측 기술의 비약적인 발전이 기대된다.

참고문헌

신성호, 이미경, 2018. “LSTM 네트워크를 활용한 농산물 가격예측 모델.”「한국콘텐츠학회논문지」. 한국과학기술정보연구원.

김진, 이정일, 2019. “딥러닝 기반의 농산물 가격예측 시스템에 대한 연구.”「한국정보기술학회논문지」 제17권 제6호(JKIIT, Vol.17, No.6).

임진모, 김월용, 2018. “인공지능을 이용한 과일 가격예측 모델 연구.” The Journal of the Convergence on Culture Technology(JCCT, 2018.5.26). 한세대학교.

채명수, 정성관. 2017. “딥러닝 기반의 기상정보를 반영한 배추 가격예측.” 한국정보통신학회 2017 추계 종합학술대회. 한국과학기술원(KAIST) IT융합연구소.

Johnathon, S and Tryambak, G. 2020. “Crop Yield Prediction Integrating Genotype and Weather Variables Using Deep Learning.” arXiv:2006.13847.

Gangopadhyay, T and Shook, J. 2020. “Interpreting the Impact of Weather on Crop Yield Using Attention.” *AI for Earth Sciences Workshop at NeurIPS 2020*.

Khaki, S and Wang, L. 2020. “A CNN-RNN Framework for Crop Yield Prediction.” *Front. Plant Sci., 24 January 2020*.

Yin, H. and Jin, D. 2020. “STL-ATT LSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM.” *Agriculture 2020, 10(12), 612*.

Alhnaity, B and Pearson, S. 2019. “Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments.” *International Symposium on Advanced Technologies and Management for Innovative Greenhouses* 10.17660/ActaHortic.2020.1296.55.

Blanco, F and Folegatti, M. 2005. “Estimation of Leaf Area For Greenhouse Cucumber by Linear Measurements Under Salinity and Grafting.” Sci. agric.(Piracicaba, Braz.).