

## 머신러닝 알고리즘을 이용한 온실 딸기 생산량 예측

김나은<sup>1</sup> · 한희선<sup>2</sup> · 아를모지엘렌체잔<sup>1</sup> · 문병은<sup>3</sup> · 최영우<sup>1</sup> · 김현태<sup>4\*</sup>

<sup>1</sup>경상국립대학교 대학원 바이오시스템공학과 대학원생, <sup>2</sup>아이티아이즈 부장,

<sup>3</sup>경상국립대학교 스마트팜연구소 학술연구교수, <sup>4</sup>경상국립대학교 생물산업기계공학과(스마트팜연구소) 교수

## Prediction of Greenhouse Strawberry Production Using Machine Learning Algorithm

Na-eun Kim<sup>1</sup>, Hee-sun Han<sup>2</sup>, Elanchezhian Arulmozhi<sup>1</sup>, Byeong-eun Moon<sup>3</sup>,  
Yung-Woo Choi<sup>1</sup>, and Hyeon-tae Kim<sup>4\*</sup>

<sup>1</sup>Graduate Student, Department of Bio-Systems Engineering, Graduate School of Gyeongsang National University, Jinju 52828, Korea

<sup>2</sup>Department Head, Iteyes Inc., Seoul 07238, Korea

<sup>3</sup>Research Professor, Institute of Smart Farm, Gyeongsang National University, Jinju 52828, Korea

<sup>4</sup>Professor, Department of Bio-Industrial Machinery Engineering, Gyeongsang National University (Institute of Smart Farm), Jinju 52828, Korea

**Abstract.** Strawberry is a stand-out cultivating fruit in Korea. The optimum production of strawberry is highly dependent on growing environment. Smart farm technology, and automatic monitoring and control system maintain a favorable environment for strawberry growth in greenhouses, as well as play an important role to improve production. Moreover, physiological parameters of strawberry plant and it is surrounding environment may allow to give an idea on production of strawberry. Therefore, this study intends to build a machine learning model to predict strawberry's yield, cultivated in greenhouse. The environmental parameter like as temperature, humidity and CO<sub>2</sub> and physiological parameters such as length of leaves, number of flowers and fruits and chlorophyll content of 'Seolhyang' (widely growing strawberry cultivar in Korea) were collected from three strawberry greenhouses located in Sacheon of Gyeongsangnam-do during the period of 2019-2020. A predictive model, Lasso regression was designed and validated through 5-fold cross-validation. The current study found that performance of the Lasso regression model is good to predict the number of flowers and fruits, when the MAPE value are 0.511 and 0.488, respectively during the model validation. Overall, the present study demonstrates that using AI based regression model may be convenient for farms and agricultural companies to predict yield of crops with fewer input attributes.

**Additional key words :** big data, cloud, data mining, green revolution, smart farm

### 서 론

우리나라는 2018년에 채소 및 화훼류 온실의 재배면적이 각각 51,226ha 및 2,048ha 전제 면적은 53,314ha로써 2017년의 54,632ha에 비해 감소하였지만, 2013년 말 이후 꾸준히 50,000ha 이상을 유지하고 있다(Jeong 등, 2020b). 그중 딸기의 전체 생산면적은 약 6,000ha이다(Kang 등, 2018). 서부 경남 지역 중 딸기재배로 유명한 지역, 즉 진주시 수곡면, 대평면, 집현면, 금곡면과 산청군 및 의령군 지역의 40개 농가를 대상으로 한 조사에 따르면 일본 품종인 “장희”를 재배하는 농가

도 있지만, 국산품종이 약 97.5%를 차지하고 있다. 이 중 “설향”이 65.0%으로서 가장 선호하고 있는 것으로 나타났다(Jeong 등, 2018). 그리고 우리나라의 딸기 재배 면적은 감소하였지만, 농업 기술의 발달로 단위 면적당 생산량(kg·10/a)은 증가했다(KOSIS, 2019).

현재의 농업은 제3의 녹색혁명이라고 일컬어질 만큼 스마트팜(Smart Farm)이 최근 4차 산업혁명 기술 적용으로 더욱 발전하고 있다(Kim 등, 2021). 스마트팜은 농업 현장에 사물인터넷(IoT), 빅데이터, 인공지능(AI) 등과 같은 제4차 산업혁명 기술이 접목된 농장을 말하며(Hwang과 Park, 2020), 이러한 스마트팜은 작물과 환경을 조사하여 데이터를 수집하고 데이터의 가공을 통해 농업의 자동화가 가능하다. 세계적으로 많은 민간 기업들과 단체들은 효율성이 높은 농업 환경을

\*Corresponding author: bioani@gnu.ac.kr

Received August 9, 2021; Revised November 10, 2021;

Accepted November 11, 2021

만들기 위해 새로운 기술들을 조사하고 개발하고 있다. 농업 분야는 표준을 개발할 때 서로 다른 이해관계와 시스템 및 기술 간의 상호 운용성을 향상시키는 것이 중요하다. 현재 시설 원예 농가에서는 스마트팜을 도입하여 생산량의 증대와 질적으로 향상을 기대하는 농가가 많아지고 있다. 특히 빅데이터를 활용하여 농가 의사결정을 하고자 하는 농가의 수요가 높다. 그러나 각 생육단계가 어떤 상황일 때 딸기의 생산량이 최적에 달하는지 대한 기준이 없으며, 이러한 판단기준은 아직까지 스마트팜에 경험이 있는 농업인의 의사에 달려있다는 문제점이 있다. 클라우드를 통한 빅데이터 플랫폼을 활용한 AI분석 방법을 접목하여 농가 내의 작물 생육상태를 관리하기 위한 빅데이터 분석이 필요하다고 판단된다. 그리고 더욱 정밀한 데이터 분석을 통해 문제점을 수정하고 보완하여 농업 환경을 개선하는데 의미가 있다(Choi 등, 2019).

먼저, 각종 센서를 통하여 수집된 대용량의 데이터의 분석 플랫폼을 생성할 때, 온프레미스(On-premise) 방식과 클라우드(Cloud) 방식이었다(Lee, 2018). 온프레미스 방식은 모든 것을 직접 설치하고 관리하는 전통적인 방식이며, 클라우드 방식은 인터넷상에 자료를 저장해두고, 필요한 정보를 컴퓨터에 설치하지 않아도 인터넷 접속만을 통해 이용할 수 있는 방식이다(Baek 등, 2018; Park 등, 2019). 온프레미스 방식은 데이터의 공유나 서비스 제공에 많은 이용자들이 발생한다면 자체적인 운영이기 때문에 문의 및 요구사항을 직접 응대할 인력이 필요할 수 있다. 그러므로 다양한 이해 관계자가 요구하는 스마트팜의 분석 모델을 개발하기 위해서는 클라우드 기반 IoT와 AI 기술을 활용한 빅데이터 분석 모델의 개발이 필요하고, 데이터의 활용도를 높이기 위해 농산업을 사물인터넷, 빅데이터, 클라우드 기술을 융합한 첨단 농업이 필요하다고 판단된다(Park과 Heo, 2020). IoT와 AI 기술의 활용은 농산물 전문가의 지식과 모델 규칙을 연계하여 생육속도 조절을 통해 출하량을 조절할 수 있고 전통적 농업으로부터 긍정적 변화를 가져올 수 있다.

따라서 본 연구는 기존의 농가 단위의 스마트팜의 작물 생장 상태를 모니터링하기 위하여 머신러닝을 이용한 딸기 생산량 예측모델을 도출함으로써, 선진화된 스마트팜 시스템을 구축하고자 한다.

## 재료 및 방법

### 1. 실험 장소

본 연구는 경상남도 사천시의 딸기 농가에서 수행하였으며, 총 3곳의 농가를 대상으로 데이터 수집을 진행하였다. 한 농가에서 랜덤으로 6개체를 조사하였으며, 세 곳의 농가에서 총 18

개체를 조사하였다. 농가의 세 곳 중 한 곳은 연동 형태의 온실 나머지 두 곳은 단독 형태의 온실이다. 실험 대상의 모든 온실 내에서 재배하는 딸기의 품종은 “설향”이다.

### 2. 데이터 수집

본 연구는 2019년 09월 23일부터 2020년 01월 07일까지 일주일에 한 번씩 16주 동안 한 농가당 6개체의 작물 생체 데이터를 수집하였으며, 매 30초 단위로 실시간으로 환경정보와 제어 정보를 수집하였다(Fig. 1). 작물 데이터의 수집 항목과 방법은 작물의 엽수, 꽃수, 과실수를 직접 측정하였으며 한 작물에서 가장 긴 초장과 짧은 초장, 한 작물에서 가장 긴 엽폭과 짧은 엽폭은 줄자를 이용하여 측정하였다. 그리고 엽록소 함량은 비파괴적 방법인 엽록소 측정기(SPAD-502 Plus, Minolta, Japan)를 사용하여 측정하였다. 환경 데이터의 수집 항목은 온도, 습도, 조도이며 온실의 중간지점에 위치하며 1m 10cm의 높이에서 무선센서(U-SN-W, UBN Co., Ltd., Korea)를 사용하여 환경정보를 수집하였다. 농장별 온실 환경 무선센서는 초 단위로 수집하였지만, 과실이나 꽃 등 작물 생육데이터는 일 단위로 수집하였다. 수집된 모든 데이터는 기존의 농가 PC에 저장하는 방식에서 벗어나 클라우드 기반의 실시간 서버에 저장하는 방식을 이용하여 여러 농가의 데이터를 실시간으로 한눈에 보기 용이하도록 하였다(Fig. 2).



Fig. 1. Data collection device (a) Chlorophyll meter; (b) Temperature and humidity wireless sensor.

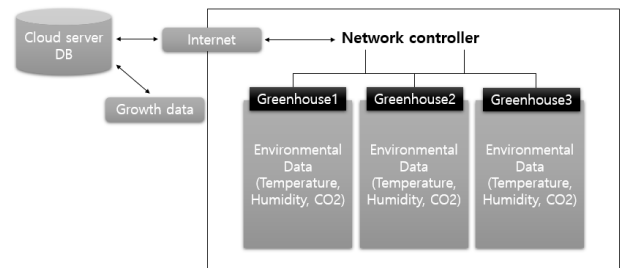


Fig. 2. Cloud-based real-time server.

### 3. 데이터 정리 및 분석

해당 농장 온실 환경 무선센서는 초 단위 정보를 생육데이터와 맞추기 위하여 일 단위 중앙치로 관리하였다. 그리고 분석 확정 데이터는 적재, 연관성 없는 데이터는 제외하였으며 결측이 많은 날짜의 데이터는 삭제하였다. 그리고 생체 데이터 측정 주기와 동일한 주기로 환경 데이터를 가공하였다. 데이터 분석을 위해 데이터 전처리를 진행하였으나, 모델 학습용 및 검정용을 사용해야 하지만 데이터의 개수가 120개로 부족하여 모델 생성용으로 사용하였다. 적은 데이터에서 머신러닝이 가능한 K-fold 교차검증으로 설계하였다. 1차 6월에서 8월, 2차는 12월로 시기상 과일이 열리지 않아 데이터가 존재하지 않으며, 세 능가 모두 12월 꽃 데이터만 존재하므로 목표 변수를 변경하여 데이터 분석을 수행하였다. 그 후 데이터 마닝 기초 자료로 활용하였다.

과실 데이터 중 분석이 용이한 딸기 생육데이터는 연동 온실 한곳과 단동 온실 한곳 2개 농가만 존재하였으며, 머신러닝 기법을 통해 샘플 데이터를 생성하고 데이터 분석을 수행하였다. 그리고 꽃 개수를 기준 변수로 데이터를 정제, 데이터 표본 생성, Lasso 회귀분석을 통한 생육 예측 모형을 생성하였다. 랜덤 포레스트 기법 중 K-fold 교차검증과 정확성 보완 등을 통해 분석 모델을 생성하였다. 분석은 Colaboratory(Google, USA.)를 사용하여 K-fold 교차검증, Lasso 회귀분석, MAPE 검증을 수행하였다(Kim, 2020).

#### 3.1 K-fold 교차검증

머신러닝에서 사용하는 K-fold 교차검증은 Fig. 3과 같이 수집한 데이터를 여러 개의 그룹으로 나누어 그 그룹 중에서 하나를 선택하여 테스트 셋으로 사용하는 것이다. 그리고 이 과정을 여러번 반복하여 나온 결과값을 평균으로 계산하여 검증 결과값으로 사용한다. 수집한 데이터의 크기가 작을 때, 테스트 셋의 성능 평가 신뢰성이 떨어질 때 사용하면 용이한 방법이다(Lee 등, 2021).

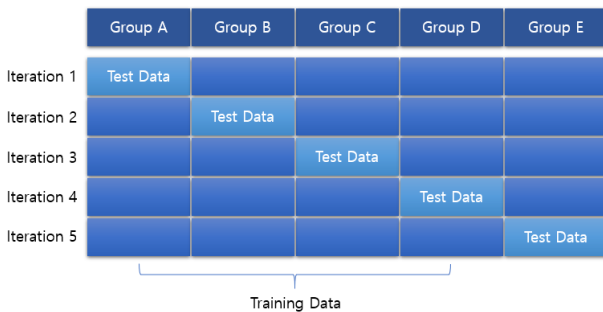


Fig. 3. K-fold cross validation.

#### 3.2 Lasso 회귀분석

Lasso(Least Absolute Shrinkage and Selection Operator)는 회귀함수의 문제점인 많은 설명변수로 인해 본래의 자료에 대한 설명력은 높지만 새로운 자료에 대한 해석력이 떨어지는 것을 해결하기 위해 Tibshirani(1996)에 의해 제시되었다. 이는 식(1)과 같이 L1 정규화(두 개의 벡터를 빼고 절대값을 취한 후 합하는 것)를 이용하여 중요하지 않은 변수들의 회귀계수에 벌점을 부여를 위해  $\lambda$ 를 사용하여 0으로 만든다. 이렇게 남은 주요 변수들로 회귀식을 만들어 모형 차원의 증가를 방지한다(Jeong 등, 2020a).

$$\hat{\beta}^{lasso} = \arg \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (1)$$

#### 3.3 MAPE 검증

MAPE(Mean Absolute Percentage Error)는 백분율 오차로 비올레러(Percentage Error)라고도 하며, 본 실험의 변수들은 상호연관성이 높아 MAPE 검증방법을 사용하였다(Myttenaere 등, 2016). 식(2)와 같이 MAPE 식에서  $A_t$ 는 실제 값이고  $F_t$ 는 예측값을 나타내며, 예측값과 실제값을 빼준 후 실제값으로 나눈다. 그리고 모두 더한 뒤  $n$ 으로 나눠 주고 백분율로 변환하기 위하여 100을 곱해주게 되는 것이다(Kim, 2018).

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2)$$

## 결과 및 고찰

### 1. 분석 모형화

꽃의 개수에 영향을 미치는 결정적 변수를 도출하여, 엽수, 토양의 온도, 긴 초장, 짧은 엽폭, 짧은 초장을 기준값으로 잡고 모형에서 꽃의 개수가 많은 규칙 집합 지식화를 진행하였다. 그리고 꽃의 개수가 많은 규칙 회귀 집합을 수행하였다. Fig. 4는 개화의 결정변수를 선택하는 모형이고, Fig. 5는 변수의 기준값을 선택하여 개화 속도에서 규칙을 발견한 결과 전체적으로 엽수는 19.5개 미만, 토양의 온도는 18.9°C 이하, 초장은 16.6cm 이하를 기준으로 규칙을 생성한 것임을 모형화한 그림이다. 그리고 각각 모형의 MAPE(Mean Absolute Percentage Error) 검증결과는 0.511, 0.488로 나타났으며, 이는 실측값과 예측값의 오차가 0.511%, 0.488%임을 의미한다.

생육상태별 수확량을 예측하기 위해 모델을 생성하였다. 위

의 Fig. 5의 기준값에 대한 규칙(Rule)을 기반으로 나타낸 결과를 정리한 Table 1을 보면, 엽수가 19.5개 이상, 토양 온도가 18.85 미만, 초장이 16.6cm 이상, 엽폭 7.4cm 미만, 초장이 18.5cm 이하일 때, 3.4개의 꽃이 필 것으로 예측할 수 있으며, 조금 다르게 엽수 19.5개 이상, 토양 온도 18.85°C 미만, 초장 16.6cm 이상, 잎 넓이 7.4cm 미만, 작물 길이 18.5cm 이상은 꽃의 개화가 1.66개가 일어날 것으로 예측된다. 여기서 작물 길이를 제외하고 4가지의 변수로 계산된 꽃의 개수는 4.14개이며, 잎 개수가 19.5개, 토양 온도가 18.9°C 미만일 때, 작물의 키가 16.6cm 이상과 미만일 때, 각각 3.11개 1.13개의 꽃이 개화한다는 것을 예측할 수 있다.

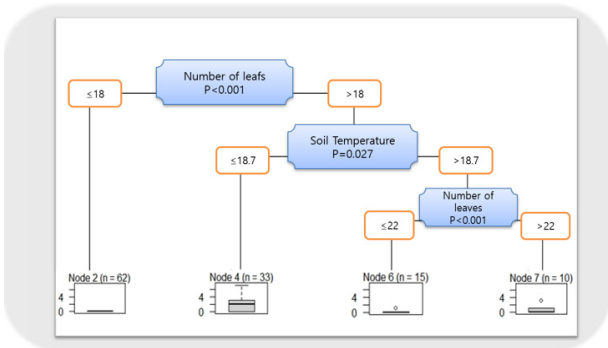


Fig. 4. Flowering determinant selection.

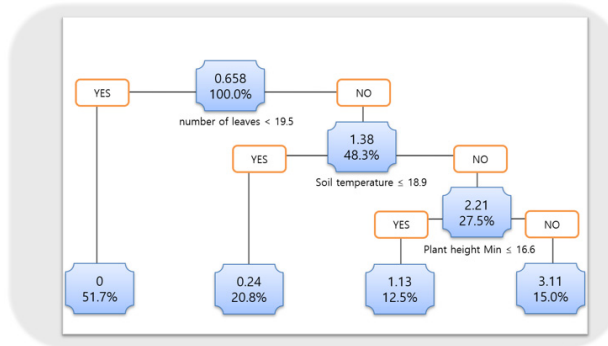


Fig. 5. Variable reference value selection.

Table 1. Yield prediction by growth status.

No.	Number of leaves	Soil temperature	Shortest plant height	Shortest leaf width	Longest plant height	Number of Flowers
1	Over 19.5	Less than 18.85°C	Over 16.6 cm	Less than 7.4 cm	Less than 18.5 cm	3.4
2	Over 19.5	Less than 18.85°C	Over 16.6 cm	Less than 7.4 cm	Over 18.5 cm	1.66
3	Over 19.5	Less than 18.85°C	Over 16.6 cm	Over 7.4 cm	-	4.14
4	Over 19.5	Less than 18.9°C	Over 16.6 cm	-	-	3.11
5	Over 19.5	Less than 18.9°C	Less than 16.6 cm	-	-	1.13

Table 1에서 결과적으로, 잎 개수 19.5개 이상, 토양 온도 18.85°C 미만, 작물 길이 16.6cm 이상, 잎 넓이 7.4cm 이상의 생육상태일 때에 딸기의 개화가 다른 상태의 딸기보다 높게 나타났다.

## 2. 딸기 수확량 예측모델

딸기 수확량 예측을 한 두 모델은 꽃 수와 과실 수 데이터를 활용하여 생성하였고, Lasso 회귀분석을 통하여 꽃수에 따른 수확량 예측모델은 Fig. 6과 같이 도출하였다.

Fig. 6의 위의 x축은 계수 값이 0이 아닌 변수의 수이고, 아래 x축은 람다( $\lambda$ )를 뜻하며 이때 가장 좋은 람다 값을 선택하기 위해 평균의 제곱 오차 추정을 할 때 가장 널리 쓰이는 k-fold 교차검증 방법을 적용한다(Lee 등, 2020a).

위의 모델에서의 꽃을 기준으로 한 수확량에 대한 예측값과 실측값은 Table 2와 같다. 꽃은 만개하기 시작한 11월 둘째 주부터 데이터를 수집하기 시작하여 총 실험 기간 중 8주 동안의 데이터를 사용하여 모델을 생성하였다. Table 2의 딸기 데이터 첫 번째 실측값이 1일 때 모델의 예측값은 1.52209485이며, 두 번째 실측값이 2일 때 1.86954707로 나타났다. 세 번째 실측값은 3으로 측정되었으며, 모델 예측값은 2.14584930이다. 네 번째 실측값은 2로 측정되었으며, 모델 예측값은 1.46552462이다. 다섯 번째와 여덟 번째 실측값은 모두 0으로 측정되었으며, 모델 예측값은 각각 0.63602713, 0.16543208, 0.24785719, -0.05233223으로 나타났다.

위의 Table 2의 결과에 5번째에서 8번째는 대부분의 꽃이 과실로 변하여 꽃의 데이터가 과실 데이터로 넘어갔을 시기에 측정된 값으로 판단된다.

Lasso 회귀분석을 통하여 과실 수에 따른 수확량 예측모델은 Fig. 7과 같이 도출하였다.

위의 모델에서의 과실을 기준으로 한 수확량에 대한 예측값과 실측값은 Table 3과 같다. 딸기 데이터의 첫 번째 실측값이 7일 때 모델의 예측값은 7.621239이며, 두 번째 실측값이 8일 때 7.600018로 나타났다. 세 번째 실측값은 9로 측정되었으며, 모델 예측값은 8.258211이다. 네 번째 실측값은 7로 측정

머신러닝 알고리즘을 이용한 온실 딸기 생산량 예측

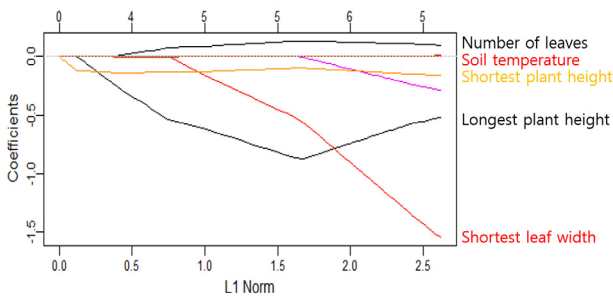


Fig. 6. Yield prediction model according to the number of flowers.

Table 2. Comparison of Model Predicted Values and Actual Values (Flower).

Data	Model Predicted Values	Actual Values
1	1.52209485	1
2	1.86954707	2
3	2.14584930	3
4	1.46552462	2
5	0.63602713	0
6	0.16543208	0
7	0.24785719	0
8	-0.05233223	0

되었으며, 모델 예측값은 6.413387이다. 다섯 번째와 여섯 번째 값 3에서 모델 예측값은 각각 3.966320, 3.007086으로 나타났다, 일곱 번째 값은 2이며 모델 예측값은 2.254328로 나타났다. 여덟 번째 실측값 3에서 모델 예측값은 2.879411로 나타났다.

꽃과 과일을 기준으로 예측모델 생성하였을 때, Table 2의 5, 6, 7, 8번째의 실측값이 같은 0이지만 모델의 예측값은 각각 0.63602713, 0.16543208, 0.24785719, -0.05233223과 Table 3의 5, 6, 8번째의 실측값이 같은 3이지만 모델의 예측값은 각각 3.966320, 3.007086, 2.879411과 같이 일정한 값을 갖지 못한 것을 확인할 수 있다. 이는 데이터의 개수가 적은 문제점으로 K-fold 교차검증을 통해 편향 현상을 줄이는 작업을 수행하였지만(Lee 등, 2021), 데이터의 개수가 적은 것에 기본적으로 생산량 예측에 영향을 주며(Na 등, 2017), 농작물은 기본적으로 생산량에 영향을 주는 요인이 매우 다양하기 때문에 광합성량, 생체중 등 본 연구에서는 수집하지 못한 더 많은 종류의 데이터가 필요할 것으로 사료된다(Lee 등, 2020b). 더욱이 수확량에 대해서는 과실의 개수보다는 과실의 무게를 중점으로 농가의 수확량이 결정되고 있으며 딸기의 연구 또한 수확량을 예측할 때, 과실의 무게로 측정하고 있다(Choi 등, 2018).

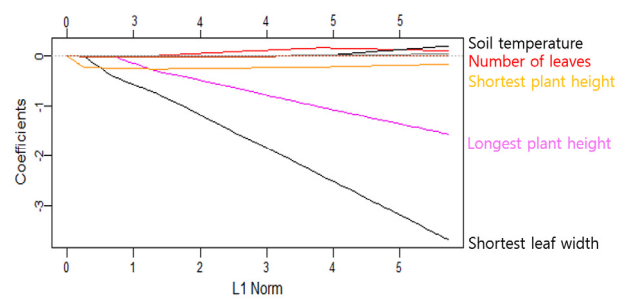


Fig. 7. Yield prediction model according to the number of fruits.

Table 3. Comparison of Model Predicted Values and Actual Values (Fruit).

Data	Model Predicted Values	Actual Values
1	7.621239	7
2	7.600018	8
3	8.258211	9
4	6.413387	7
5	3.966320	3
6	3.007086	3
7	2.254328	2
8	2.879411	3

본 실험의 결과는 작물의 성장 속도와 작물의 상태에 따라 수확량을 예측하는 모델을 머신러닝을 통해 도출하였다. 추후 연구에서 위의 문제점을 보완한다면, 더욱 정밀한 스마트팜 시스템 구축에 기초자료로 활용되고 스마트팜 데이터의 정립을 위해 AI를 통하여 생육상태별 수확량을 예측하여 농가 및 농업 관련 기업에 활용함으로써 농업서비스가 편리해질 것으로 판단된다.

적 요

서부 경남 지역 중 딸기재배로 유명한 지역 40개 농가를 대상으로 한 조사에 따르면 국산품종 중에서 “설향”이 65.0%로 가장 선호하고 있는 것으로 나타났다. 그리고 현재의 농업은 4차 산업혁명으로 스마트팜(Smart Farm)의 기술이 더욱 발전하고 있는 실정이다. 그러나 각 생육단계가 어떤 상황일 때 딸기의 생산량이 최적에 달하는지 대한 기준이 없으며, 이러한 판단기준은 아직까지 스마트팜에 경험이 있는 농업인의 의사에 달려있다는 문제점이 있다. 따라서 본 연구에서는 딸기의 생육상황에 대한 생산량 예측을 통해 선진화된 스마트팜 시스템을 구축하고자 한다. 실험 장소는 경상남도 사천시의 딸기 농가에서 수행하였으며, 총 3곳을 대상으로 데이터 수

집을 진행하였다. 실험 대상의 모든 온실 내에서 재배하는 딸기의 품종은 ‘설향’이다. 작물 데이터의 수집 항목은 작물의 엽수, 꽃수, 과실수, 초장, 잎의 길이, 엽록소 함량이며, 환경 데이터의 수집 항목은 온도, 습도, 조도이다. 기존의 농가 단위의 스마트팜의 문제점 보완 및 개선을 통하여 고품질의 작물 생산 상태를 유지하기 위해 K-fold 교차검증, Lasso 회귀분석, MAPE 검증을 통해 예측모델을 도출하였으며, MAPE 검증 결과 값으로 0.511(꽃 예측)과 0.488(과일 예측)의 값이 나타났다. 본 연구는 스마트팜 데이터 구축을 위해서는 AI를 통해 성장상태별 수확량을 예측하였으며, 이를 농가 및 농업 관련 기업에 활용해 농업 서비스가 편리할 것으로 판단된다.

**추가 주제어:** 녹색혁명, 데이터 마이닝, 빅데이터, 스마트팜, 클라우드

## 사 사

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 1세대 스마트플랜트팜 산업화 기술개발 사업의 지원을 받아 연구되었음(320112-1).

## Literature Cited

- Baek J.H., J.W. Heo, H.H. Kim, Y. Hong, and J.S. Lee 2018, Research-platform design for the korean smart greenhouse based on cloud computing. *Protected Hort Plant Fac* 27:27-33. (in Korean) doi:10.12791/KSBEC.2018.27.1.27
- Choi H., H. Ahn, Y. Jeong, and B. Lee 2019, A smart farm environment optimization and yield prediction platform based on IoT and deep learning. *The Journal of Korea Institute of Information, Electronics, and Communication Technology* 12:672-680. (in Korean) doi:10.17661/jkiiect.2019.12.6.672
- Choi H.G., H.J. Jeong, G.L. Choi, S.H. Choi, S.C. Chae, S.W. Ann, H.K. Kang, and N.J. Kang 2018, Effects of supplemental LED lighting on productivity and fruit quality of strawberry (*fragaria ananassa* duch.) grown on the bottom bed of the two-bed bench system. *Protected Hort Plant Fac* 27:199-205. (in Korean) doi:10.12791/KSBEC.2018.27.3.199
- Hwang D.H., and G.J. Park 2020, An analysis on the educational needs for the smart farm: focusing on SMEs in jeon-nam area. *Journal of the Korea Academia-Industrial cooperation Society* 21:649-655. (in Korean) doi:10.5762/KAIS.2020.21.1.649
- Jeong K.H., and H.J. Lim 2020a, Analyzing bankruptcy prediction in energy and environment industries using logistic lasso. *Korean Energy Economic Review* 19:117-152. (in Korean)
- Jeong Y.K., J.G. Lee, E.K. Ahn, J.S. Seo, and Y.C. Yoon 2020 b, Performance improvement of a temperature and humidity measuring system for strawberry cultivation greenhouse. *Protected Hort Plant Fac* 29:110-119. (in Korean) doi:10.12791/KSBEC.2020.29.2.110
- Jeong Y.K., J.G. Lee, S.W. Yun, H.T. Kim, and Y.C. Yoon 2018, Field survey of greenhouse for strawberry culture - case study based on western gyeongnam area -. *Protected Hort Plant Fac* 27:253-259. (in Korean) doi:10.12791/KSBEC.2018.27.3.253
- Kang J.H., H.M. Kim, H.M. Kim, H.W. Jung, H.R. Lee, H.S. Hwan, B.R. Jeong, N.J. Kang, and S.J. Hwang 2018, Gibberellin application method and concentration affect to growth, runner, and daughter plant production in maehyang strawberry during nursery period. *Protected Hort Plant Fac* 27:407-414. (in Korean) doi:10.12791/KSBEC.2018.27.4.407
- Kim D.K. 2020, Google colab. *Magazine of the SAREK* 49:110-111. (in Korean)
- Kim J. 2018, Franchise business sales forecasting by comparison of neural network models. *Korean Journal of Marketing* 33:73-90. (in Korean) doi:10.15830/kmr.2018.33.3.73
- Kim J.C., S.H. Kwon, I.D. Ha, and M.H. Na 2021, Survival analysis for tomato big data in smart farming. *Journal of the Data and Information Science Society* 32:361-374. (in Korean) doi:10.7465/jkdi.2021.32.2.361
- KOSIS 2019, 2019 Strawberry production survey, <https://kosis.kr/>. Accessed 01 July 2021
- Lee D.K., Y.W. Kim, and H.J. Park 2020a, Exploring variables affecting college graduates' job satisfaction : an application of sparse group lasso model. *Asian Journal of Education* 21: 1069-1097. (in Korean) doi:10.15753/aje.2020.12.21.4.1069
- Lee J., C. Moon, and B. Kim 2020b, Big data model for analyzing plant growth environment informations and biometric informations. *Journal of the Korea Industrial Information Systems Research* 25:15-23. (in Korean) doi:10.9723/jkisi.2020.25.6.015
- Lee K.P. 2018, Where would you build your big data analytics platform? - comparative analysis between cloud and on-premise. *Journal of Research Methodology* 3:101-117. (in Korean) doi:10.21487/jrm.2018.11.3.2.101
- Lee Y.E., N.J. Choi, Y.H. Byun, D.W. Kim, and K.C. Kim 2021, Rubber O-ring defect detection system using K-fold cross validation and support vector machine. *Journal of The Korean Society of Visualization* 19:68-73. (in Korean) doi:10.5407/jksv.2021.19.1.068
- Myttenaere A.D., B. Golden, B.L. Grand, and F. Rossi 2016, Mean absolute percentage error for regression models. *Neurocomputing* 192:38-48. doi:10.1016/j.neucom.2015.12.114
- Na M.H., Y. Park, and W.H. Cho 2017, A study on optimal environmental factors of tomato using smart farm data. *Journal of the Korean Data And Information Science Society* 28:1427-1435. (in Korean) doi:10.7465/jkdi.2017.28.6.1427
- Park H.D., S.Y. Lee, and K.I. Ko 2019, A study on development plan by comparing korea-us japan agricultural cloud service. *Korea Rural Economic Research Institute Basic Research*

- Report pp 112-120. (in Korean)
- Park S.G., and C.M. Heo 2020, The effect of the perception of ICT technical characteristics in agricultural industry on the intention to start smart farm: focusing on the mediating effects of effort expectation and acceptance intention of smart farm. *Asia-Pacific Journal of Business Venturing and Entrepreneurship* 15:19-32. (in Korean)
- Tibshirani R. 1996, Regression shrinkage and selection via the lasso. *J R Statist Soc B* 58:267-288.