

MANAGEMENT&ECONOMICS

Income prediction of apple and pear farmers in Chungnam area by automatic machine learning with H2O.AI

Hyundong Jang¹, Sounghun Kim^{2*}

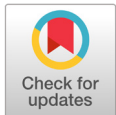
¹Division of Horticulture Research, Chungcheongnam-do Agricultural Research and Extension Service, Yesan 32418, Korea

²Department of Agricultural Economics, Chungnam National University, Daejeon 34134, Korea

*Corresponding author: sounghunkim@cnu.ac.kr

Abstract

In Korea, apples and pears are among the most important agricultural products to farmers who seek to earn money as income. Generally, farmers make decisions at various stages to maximize their income but they do not always know exactly which option will be the best one. Many previous studies were conducted to solve this problem by predicting farmers' income structure, but researchers are still exploring better approaches. Currently, machine learning technology is gaining attention as one of the new approaches for farmers' income prediction. The machine learning technique is a methodology using an algorithm that can learn independently through data. As the level of computer science develops, the performance of machine learning techniques is also improving. The purpose of this study is to predict the income structure of apples and pears using the automatic machine learning solution H2O.AI and to present some implications for apple and pear farmers. The automatic machine learning solution H2O.AI can save time and effort compared to the conventional machine learning techniques such as scikit-learn, because it works automatically to find the best solution. As a result of this research, the following findings are obtained. First, apple farmers should increase their gross income to maximize their income, instead of reducing the cost of growing apples. In particular, apple farmers mainly have to increase production in order to obtain more gross income. As a second-best option, apple farmers should decrease labor and other costs. Second, pear farmers also should increase their gross income to maximize their income but they have to increase the price of pears rather than increasing the production of pears. As a second-best option, pear farmers can decrease labor and other costs.



OPEN ACCESS

Citation: Jang H, Kim S. Income prediction of apple and pear farmers in Chungnam area by automatic machine learning with H2O.AI. Korean Journal of Agricultural Science 49:619-627. <https://doi.org/10.7744/kjoas.20220056>

Received: July 28, 2022

Revised: August 16, 2022

Accepted: August 19, 2022

Copyright: © 2022 Korean Journal of Agricultural Science



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Key words: apple, automatic machine learning, deep learning, GBM (gradient boosting machine), pear

Introduction

사과와 배는 우리 농가의 주요 과일로 농업 소득 창출에 상당한 기여를 하고 있다. 농가는 자신이 재배한 사과나 배를 판매하여 필요한 소득을 얻는데, 한정된 농업자원을 투입하여 가능한 높은 소득을 얻기 위한 의사결정이 단계마다 중요하게 작용한다. 환언하면, 농가는 자

신에게 주어진 생산 여건에서 비용을 들여 확보하는 농자재를 어떻게 투입하고, 시장에 판매하기 위한 가격과 물량을 어떻게 결정하는 지 등에 대해 매번 고민을 하고 있다.

이에 따라 연구자들은 농가가 최적화된 소득을 얻기 위한 의사결정을 하는 것에 도움을 주기 위한 연구를 다양하게 진행하고 있다. 먼저 농가 소득을 분석하여 소득 증대를 위한 시사점을 모색한 선행연구로 Jeon 등(2016)은 벼와 양념채소(고추, 마늘, 양파) 생산 농가의 소득 변동 요인을 분석하여 경영비 부담을 줄여 소득을 제고하기 위한 방안을 제시하였고, Hwang 등(2018)은 경상남도 청양군의 풋고추 농가의 실질 소득과 순수익의 불안정 요인을 경영비 분석을 통해 연구하여 소득 안정을 위한 방안을 제안하였으며, Jeon과 Jang (2020)은 복숭아 농가의 농업소득 결정 요인을 설문조사 등을 통해 분석하고 소득 개선을 위한 방안을 도출하였다. 한편 농가 소득에 영향을 주는 간접 요인들에 대한 연구도 진행되었는데, Jang과 Kim (2016)은 백합 농가의 주요 투입재인 종구의 국산화에 따른 경제적 가치를 추정하였고, Soon 등(2021)은 최근 농가의 재배 선호도가 급증한 샤인머스켓 포도의 생산 증가 및 품질 하락이 농가 소득에 미치는 영향을 경제모형을 적용하여 분석하였다.

한편, 최근 새로운 분석법으로 각광받고 있는 머신러닝(machine learning)을 적용한 연구도 진행되고 있는데, Lee와 Lee (2021)는 우리나라 외식 소비자 중 경기도 양평군의 외식업체를 방문하는 고객들의 소비지출 패턴을 예측하기 위한 모델을 머신러닝 방법론을 적용하여 개발하였고, Choi 등(2022)은 농업 유전자원의 데이터 분석 플랫폼을 구축하기 위한 머신러닝 기법을 적용하였다. Oh 등(2019)은 과일을 최상의 상태에서 수확하여 최고가격을 받을 수 있는 수확시기 예측을 위해 머신러닝을 적용하여 분석하였고, Li 등(2021)은 딸기 수익 예측을 위해 머신러닝 기법을 적용하여 분석을 진행하였다.

이상의 연구들은 농가 소득에 영향을 주는 다양한 요인을 분석하여 시사점을 도출하거나 머신러닝 방법론을 식품산업 또는 농업부문에 적용하여 나름의 시사점을 도출하였다. 그러나 이러한 선행연구들이 농가 소득 분석에 머신러닝 방법론을 직접적으로 적용한 논의를 진행하지는 못하였다. 전술한 것과 같이, 데이터를 이해하는 알고리즘의 과학이자 애플리케이션인 머신러닝 기법이 다양한 부문에 적용되고 있는 상황에서 이를 농가 소득 증대를 위한 방안 모색에 적용하기 위한 연구의 필요성이 있다.

본 연구는 우리나라 대표 과일인 사과와 배 생산 농가의 소득을 머신러닝의 방법론을 적용하여 분석하고, 농가 소득 증대를 위한 시사점을 제시하는 것을 주요 목적으로 한다. 보다 구체적으로는 머신러닝 자동화 솔루션을 이용하여 지난 11년간 충남지역 사과, 배 소득관련 데이터를 분석하고 작목별 소득에 영향을 미치는 변수를 도출하여 해당 농가의 소득 증대를 위한 방안을 제안하도록 한다.

Materials and Methods

머신러닝 방법론의 개념과 개요

머신러닝은 스스로 데이터를 통해 학습할 수 있는 알고리즘으로 사람이 수동으로 대량의 데이터를 분석하여 규칙을 유도하고 모델을 만드는 과정을 대신한다. 머신러닝은 보다 빠른 시간에 데이터에서 지식을 추출하여 예측모델과 데이터 기반의 의사결정 성능을 점진적으로 향상시키는 강점을 가진다. 컴퓨터 과학에서 머신러닝은 더욱 중요해지고 있으며, 이미 우리 일상에서도 상당한 역할을 하고 있다. 예를 들면 이메일 스팸 필터, 편리한 텍스트와 음성인식 소프트웨어, 웹 검색 엔진, 체스 대결 프로그램, 자율주행 자동차 등이 해당된다.

머신러닝을 이용하여 필요로 하는 목적함수 값을 예측하기 위해서는 컴퓨터에 아나콘다(Anaconda) 프로그램을 설치하는 등 개발 환경을 마련하고, 피처 엔지니어링 등 데이터 전처리 작업 후에 사이킷런(scikit-learn) 등의 머신러닝 프로그램을 이용하여 예측 모델을 개발해야 한다. 사이킷런 프로그램을 이용한 머신러닝으로 종속변수 값에 예

측하는 방법은 Fig. 1과 같은데, 분석하고자 하는 빅데이터에서 독립변수(X)를 이용하여 종속변수(y)를 예측할 때 랜덤 포레스트(random forest) 알고리즘을 이용하고, 데이터 예측 정확도 검정을 위하여 훈련 데이터(training data)와 검정데이터(test data)를 일정비율(7:3)로 구분하여 모델을 만들어 예측하한 다음, 실측치와 예측치를 상호 비교하여 모델 적합성을 판단하게 된다.

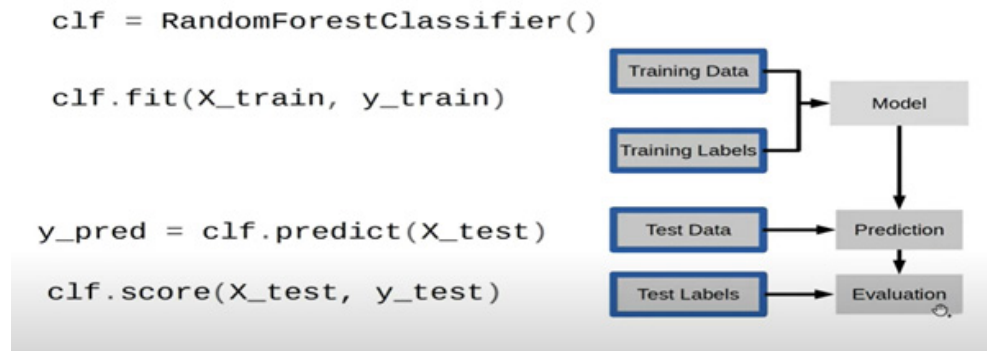


Fig. 1. Basic model of machine learning using scikit-learn.

머신러닝을 이용한 데이터 분석은 많은 시행착오를 통한 수작업으로 이루어지기 때문에 상당히 힘이 드는 과정이지만, 최근 이러한 분석과정을 자동으로 처리하는 머신러닝 자동화 솔루션이 개발되어 변수만 지정해주면 나머지 분석과정을 자동으로 처리하여 최적의 예측모델이 만들어지게 되었다.

본 연구에서는 바로 이러한 머신러닝의 자동화 솔루션 H2O.AI (H2O.AI, New York, USA)를 이용하여 충남 농산물 소득을 예측하였다. H2O.AI는 2012년에 미국 실리콘 벨리에서 Wells Fargo 등 여러 개 벤처회사가 공동 투자하여 개발되었다. H2O.AI의 장점으로는 시각화, 피쳐 엔지니어링, 모델 튜닝, 시계열 분석 등이 자동화 처리되고, 생성 모델의 활용 지원을 위한 자동화 파이프라인(automatic pipelines), 짧은 지연시간내 추론(low latency inferencing) 등이 가능하다. 그리고 기존의 인공지능과는 달리 인공지능의 결정에 대한 해석과 설명이 가능하며, 보안성, 확장성이 매우 뛰어난 장점이 있다.

H2O.AI의 머신러닝 자동화 분석 과정은 다음과 같다. 우선 데이터를 클라우드, 빅데이터, 데스크탑 등에서 끌어다 놓게 된다. 그 다음 설명변수와 종속변수를 정해 주면, 데이터 모양, 이상치, 결측치 등을 이해하고 자동적으로 시각화 분석을 실시한다. 그리고 최상의 연습 모델 레서피(best practice model recipes)와 CPU/GPU연산능력을 이용하여 진보된 피쳐 엔지니어링(feature engineering)과 변수 튜닝(parameter tuning)을 포함한 수 천개의 가능성 있는 모델을 탐색하여 반복 훈련하면서 자동화된 머신러닝 과정을 거친다. 마지막으로 자동화된 점수 파이프라인(scoring pipelines)을 통해 모델 예측 값을 판단하게 된다.

머신러닝 분석을 위한 모형으로는 딥러닝(deep learning)모형, XGBoost (extreme gradient boosting) 모형, GBM (gradient boosting machine) 모형 등이 있는데, 딥러닝 모형은 머신러닝 기술 중 하나인 인공신경망을 수많은 계층 형태로 연결한 기법으로 입력과 출력 사이에 있는 인공 뉴런들을 층층이 쌓고 연결한 인공신경망 기법을 다루는 특성을 가진다. XGBoost 모형은 트리 기반의 앙상블 기법 중 하나로 그래디언트 부스팅(gradient boosting)을 분산환경에서 실행하여 다른 알고리즘보다 좋은 예측 성능을 가지는 강점이 있다. GBM 모형은 첫번째 모델 tree 1을 통해 y를 예측하고 잔차(실제값과 예측값의 차이)를 다시 두번째 단계 모델 tree 2를 통해 예측하고 여기에서 발생한 잔차를 모델 tree 3로 예측하여 점차 잔차를 작게 만들어 가는 기법이다. 본 연구에서는 전술한 3가지 모형을 모두 적용하여 사과와 배 농가의 소득 추정을 시도하되, 그중 예측치와 실측치와의 오차를 의미하는 평균절대오차(mean absolute error, MAE) 값이 가장 작게 나타나는 모형을 선정하여 분석에 활용하도록 한다.

분석자료

본 연구에서 사용한 자료는 충남지역에서 사과 또는 배를 재배하는 농가의 소득 및 경영비 자료를 2007년부터 2019년까지 수집하여 적용하였다. 사과와 배 농가의 기초자료에 대한 기술 통계량 분석 결과는 Table 1과 Table 2에 각각 제시되어 있는데, 사과의 경우, 평균 판매 가격이 2,455원/kg으로 배의 1,982원/kg보다 높게 나타났으나 10 a당 평균 생산량이 2.2톤으로 배의 2.8톤 보다는 낮아, 사과 농가의 10 a당 평균 소득(2,800 천원)이 배(2,974천원) 보다 낮게 나타났다.

사과와 배 생산에 소요되는 투입재 비용도 차이를 보였는데, 사과의 경우 감가상각비를 제외하고 농약비가 10 a당 평균 361천 원으로 가장 높은 반면 배는 재료비가 10 a당 평균 636 천원으로 가장 높은 모습을 보였다. 다만 사과와 배 모두 비용 중 광열동력비(사과: 95 천원/10a, 배: 87 천원/10a)가 가장 작은 비중을 차지하여 시설원에 작물과 달리 노지 과수원에서 재배하는 사과와 배의 특성을 반영하고 있다.

Table 1. Descriptive statistics of variables for apple farms (Unit: kg, Won/kg, Won/10 a).

Variable	Obs.	Min.	Max.	Mean	S.D.
Quantity	156	250	5,034	2,172	986
Price	156	1,111	9,360	2,455	908
Fertilizer	156	10,000	1,828,200	178,356	192,714
Pesticide	156	6,750	1,222,500	360,910	200,857
Energy	156	3,390	825,000	94,571	86,218
Material	156	4,991	1,323,180	357,200	239,277
Depreciation	156	2,000	1,376,000	425,507	254,975
Orchard building	156	3,300	1,775,148	245,327	285,297
Hired labor	156	39,486	2,732,880	611,883	487,124
Income	156	100,219	12,349,957	2,799,589	2,054,821

Obs., observation; Min., minimum; Max., maximum; S.D., standard deviation.

Table 2. Descriptive statistics of variables for pear farms (Unit: kg, Won/kg, Won /10 a).

Variable	Obs.	Min.	Max.	Mean	S.D.
Quantity	197	600	6,158	2,838	895
Price	197	830	7,500	1,982	669
Fertilizer	197	9,195	832,500	179,642	126,035
Pesticide	197	6,000	1,363,636	270,034	175,872
Energy	197	16,502	340,500	86,789	60,756
Material	197	33,341	3,195,600	635,762	441,889
Depreciation	197	12,500	3,195,186	460,544	397,294
Building	197	3,208	909,091	97,378	125,193
Hired labor	197	8,200	2,137,412	673,211	360,456
Income	197	187,244	8,178,504	2,973,551	1,304,489

S.D., standard deviation.

Table 3. Selection of prediction model for apple farms.

Model_ID	Mean_residual_deviance	RMSE	MSE	MAE	RMSLE
Deep learning	1.50233e + 12	1.2257e + 06	1.50233e + 12	757,003	nan
XGBoost	1.49404e + 12	1.22231e + 06	1.49404e + 12	766,578	0.632611

RMSE, root mean squared error; MSE, mean squared error; MAE, mean absolute error; RMSLE, root mean squared error; GBM, gradient boosting machine.

Results and Discussion

Table 1과 Table 2에서 제시된 사과와 배 소득 데이터를 시각적으로 분석하기 위해 히트맵(heatmap)을 이용한 시각화 분석을 진행하였는데, 분석 결과가 Fig. 2와 Fig. 3에서 제시된 바와 같이 사과 소득은 수량이 단가보다 상관성이 높았고, 배 소득은 단가가 수량보다 상관성이 높은 것으로 나타났다. 이는 사과농가는 소득을 극대화하기 위해 수량의 증대가 가장 중요하고, 배 농가는 배 단가를 높게 받는 것이 소득 극대화에 가장 중요함을 의미하는데, 이러한 결과는 후술할 머신러닝을 이용한 두 작목의 소득 예측 분석결과와 일치한다.

머신러닝의 자동화 솔루션 H2O.AI를 이용하여 작목별 소득을 추정된 결과, Table 3에서 제시된 것과 같이 사과 소득은 예측치와 실측치와의 오차를 의미하는 평균절대오차(MAE) 값이 가장 작은 딥러닝 모델이 가장 적합한 것으로 판단된다.

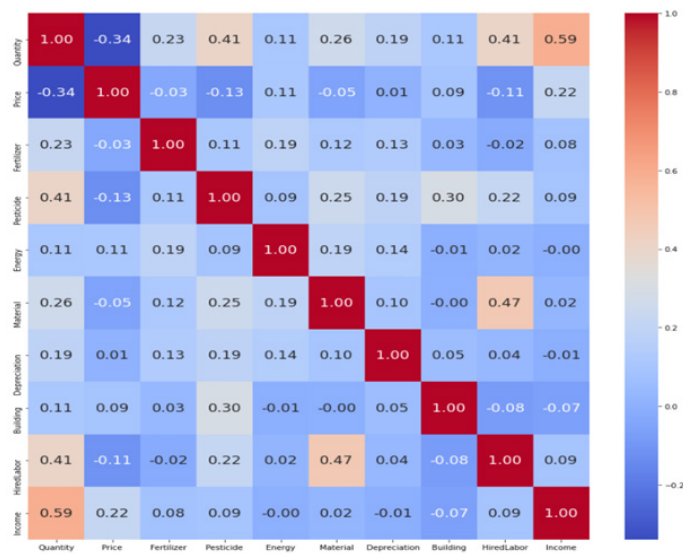


Fig. 2. Correlation among apple income variables using by heatmap.

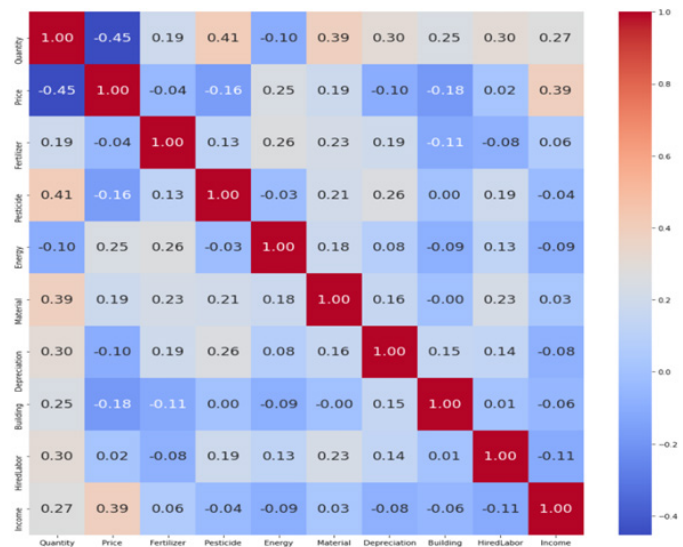


Fig. 3. Correlation among pear income variables using by heatmap.

$$MAE = \sum \frac{\widehat{y}_i - y_i}{n} \quad (1)$$

Table 4는 해당 딥러닝 모델의 추정 결과를 평가한 결과 값을 보여주고 있는데, 모델의 재현성을 위하여 데이터를 4개로 나누어 평가했기 때문에(cv_1 ~ cv_4) 해당 결과들의 평균값(mean)과 표준편차(standard deviation, s.d.)가 각각 표시되어 있다. 결과적으로 평균값을 기준으로 모델의 추정 결과를 평가하게 되는데, 표준편차(s.d.) 값이 너무 크면 모형의 안정성이 떨어짐을 의미한다. 표에서 모델의 설명력에 해당하는 r^2 값은 0부터 1의 값을 가지며 1에 가까울수록 성능이 우수함을 뜻하는데, 본 모델의 값은 0.68 정도로 나타났다.

Table 5는 해당 모형에 사용한 변수들이 농가 소득(income)에서 차지하는 중요도(importance)를 나타낸다. 표에서 제시된 값인 중요도는 0과 1 사이의 값을 가지는데, 1 (100.0%)에 가까울수록 중요한 변수로 평가된다. 표에 제시된 사과 소득에 영향을 미치는 변수 중요도를 보면 수량(25.6%), 단가(15.4%), 고용노동비(12.8%), 비료비(11.2%), 농약비(9.4%), 재료비(8.1%), 광열동력비(6.4%), 감가상각비(6.3%), 과원조성비(4.8%) 순으로 나타났다. 이를 통해 사과 농가의 소득을 높이기 위해서는 경영비 절감보다는 총수입 증대가 상대적으로 더 유리한 것으로 분석되는데, 총수입 증대를 위해서는 수량 증대가 단가 상승보다 더 중요한 것임이 확인된다. 보다 구체적으로는 사과를 재배하는 농가는 개별 사과의 크기나 당도 등의 품질을 높여서 가격을 올리기보다는 사과 생산량을 높이는 방식으로 재배관련 의사결정을 하는 것이 보다 타당한데, 이는 사과의 단수를 높이기 위한 노력이 강화되어야 함을 의미한다. 한편, 경영비에서는 고용노동비, 비료비, 농약비를 가장 중점적으로 절감하는 것이 유리한 것으로 도출되었다.

다음으로 배를 대상으로 머신러닝을 적용한 분석을 진행하였는데, Table 6에 제시된 것과 같이 평균절대오차(MAE) 값이 가장 작은 GBM 모델이 가장 적합한 것으로 나타나 해당 모델을 분석에 적용하였다. 추가로 GBM 모델의 모델 설명력(r^2)은 Table 7에 제시된 바와 같이 69.1%로 사과의 경우보다는 다소 높은 수치를 보였다.

배 소득에 영향을 미치는 변수 중요도는 Table 8에 제시되었는데, 단가(36.3%), 수량(26.65%), 고용노동비(8.55%), 광열동력비(6.5%), 농약비(6.0%), 재료비(5.4%), 비료비(4.4%), 과원조성비(3.5%), 감가상각비(3.1%) 순으로 나타났다. 이를 통해 배 또한 소득 증대를 위해서는 총수입 증대가 경영비 절감보다 중요한 것으로 나타났지만, 배 농가의 총수입은 수량보다는 단가 상승이 더 중요한 것으로 나타나 사과의 경우와 차이를 보인다. 즉, 생산량 증대에 초점을 두어야 하는 사과와 달리 배는 수확되는 배의 당도나 크기 등의 품질을 보다 높이기 위해 노력하는 것이 전체적인 소득 증대에 유리한 것으로 판단된다. 한편, 경영비 절감 부분에서는 고용노동비, 광열동력비, 농약비를 가장 중점적으로 절감하는 것이 효과적인 것으로 도출되었다.

Table 4. Model accuracy by deep learning.

Statistics	Mean	S.D.	cv_1_valid ^w	cv_2_valid ^x	cv_3_valid ^y	cv_4_valid ^z
MAE	757,468.5	247,475.72	691,805.6	995,824.4	1036,851.2	488,939.94
Mean_residual_deviance	1.50558894E12	1.13098713E12	1.04649569E12	2.30357257E12	3.05629259E12	3.62023682E11
MSE	1.50558894E12	1.13098713E12	1.04649569E12	2.30357257E12	3.05629259E12	3.62023682E11
r^2	0.683959	0.15958728	0.74150735	0.48053882	0.5677315	0.8832872
Residual_deviance	1.50558894E12	1.13098713E12	1.04649569E12	2.30357257E12	3.05629259E12	3.62023682E11
RMSE	1,152,434.6	471,013.97	1,022,983.75	1,517,752.5	1,748,225.5	601,684.06
RMSLE	0.4943031	0.13350378	NaN	0.6226205	NaN	0.50413245

S.D., standard deviation; MAE, mean absolute error; MSE, mean squared error; RMSE, root mean squared error; RMSLE, root mean squared error.

^w Variance factor of the first data after dividing the data into four for model validation.

^x Variance factor of the second data after dividing the data into four for model validation.

^y Variance factor of the third data after dividing the data into four for model validation.

^z Variance factor of the fourth data after dividing the data into four for model validation.

Table 5. Variable importance on apple income.

Variable	Scaled_importance ^y	Percentage ^z (%)
Quantity	1.000000	25.6
Price	0.602951	15.4
Hired labor	0.498674	12.8
Fertilizer	0.439432	11.2
Pesticide	0.367785	9.4
Material	0.314785	8.1
Energy	0.251341	6.4
Depreciation	0.245831	6.3
Building	0.189135	4.8

^y Importance value calculated by converting the largest value to 1.

^z Importance value converted based on 100% overall.

Table 6. Selection of prediction model for pear farms.

Model_id	Mean_residual_deviance	RMSE	MSE	MAE	RMSLE
GBM	5.24367e + 11	724,132	5.24367e + 11	526,306	0.389953
XGBoost	5.27488e + 11	726,283	5.27488e + 11	526,505	0.404025
Deep learning	6.12187e + 11	782,424	6.12187e + 11	544,032	0.357269

RMSE, root mean squared error; MSE, mean squared error; MAE, mean absolute error; RMSLE, root mean squared error; GBM, gradient boosting machine.

Table 7. Model accuracy by GBM (gradient boosting machine).

Statistic	Mean	S.D.	cv_1_valid ^v	cv_2_valid ^w	cv_3_valid ^x	cv_4_valid ^y	cv_5_valid ^z
MAE	526,489.4	47,438.77	452,590.06	567,002.2	515,749.1	568,682.44	528,423.1
Mean_residual_deviance	5.25085999E11	1.37926214E11	3.38336678E11	5.8094525E11	5.7077144E11	6.9493201E11	4.40444649E11
MSE	5.25085999E11	1.37926214E11	3.38336678E11	5.8094525E11	5.7077144E11	6.9493201E11	4.40444649E11
r ²	0.69149387	0.050927166	0.76439166	0.70089525	0.6234243	0.67349046	0.6952678
Residual deviance	5.25085999E11	1.37926214E11	3.38336678E11	5.8094525E11	5.7077144E11	6.9493201E11	4.40444649E11
RMSE	719,328.94	97,799.83	581,667.1	762,197.6	755,494.2	833,625.8	663,660.06
RMSLE	0.37896115	0.104741655	0.35614514	0.34479228	0.559897	0.2871909	0.3467804

S.D., standard deviation; MAE, mean absolute error; MSE, mean squared error; RMSE, root mean squared error; RMSLE, root mean squared error.

^v Variance factor of the first data after dividing the data into four for model validation.

^w Variance factor of the second data after dividing the data into four for model validation.

^x Variance factor of the third data after dividing the data into four for model validation.

^y Variance factor of the fourth data after dividing the data into four for model validation.

^z Variance factor of the fifth data after dividing the data into four for model validation.

Table 8. Variable importance on pear income.

Variable	Scaled_importance ^y	Percentage ^z (%)
Price	1.000000	36.3
Quantity	0.734705	26.6
Hired labor	0.235771	8.5
Energy	0.178553	6.5
Pesticide	0.157002	5.7
Material	0.149279	5.4
Fertilizer	0.121498	4.4
Building	0.096477	3.5
Depreciation	0.084953	3.1

^y Importance value calculated by converting the largest value to 1.

^z Importance value converted based on 100% overall.

Conclusion

사과와 배는 우리나라 주요 과일 중의 하나로 생산 농가의 소득 창출의 주요 수단이 되고 있는데, 농가는 한정된 자원을 효율적으로 투입하고 시장에 효과적으로 판매하기 위한 의사결정을 매년 진행한다. 그동안 농가 소득을 분석하여 시사점을 제시한 선행연구들이 다수 진행되었으나, 머신러닝 방법론을 적용한 연구는 거의 없었다. 본 연구에서 과거 11년간 충청남지역의 사과, 배 소득조사 결과 데이터를 이용하여 머신러닝 자동화 솔루션 H2O.AI로 두 작목의 소득에 영향을 미치는 변수들을 예측한 결과, 사과 농가는 수량의 증대에 초점을 두고, 배 농가는 단가 상승을 위해 노력하여야 함을 발견할 수 있었다. 또한 농가 소득을 높이기 위해서는 경영비 절감보다는 총수입 증대가 중요하다는 것도 확인되었다.

본 연구에 적용된 머신러닝의 방법론은 기존의 경제모형분석 방법론들에 비해 이론적으로 단순하지만, 새로운 방법론적 접근을 통해 보다 풍부한 학술적 논의를 촉진시킬 수 있을 것으로 기대된다. 본 연구의 분석 결과는 소득 추정을 위한 축적된 데이터 양이 적어 추정모델에 대한 적합도가 70%에 못 미치는 수준인 한계점을 가지지만, 향후 보다 많은 빅데이터 수집 분석을 통해 현실 설명력이 높은 소득 추정모델 개발이 가능할 것으로 생각된다.

Conflict of Interests

No potential conflict of interest relevant to this article was reported.

Acknowledgements

이 연구는 2021년도 충청남도농업기술원 기본연구과제로 수행되었다.

Authors Information

Hyundong Jang, <https://orcid.org/0000-0001-6869-3016>

Soungun Kim, <https://orcid.org/0000-0001-7175-8239>

References

- Choi JM, Kim JH, Choo HK, Park CL, Chae HJ. 2022. Machine learning-based agricultural genetic resources analysis platform. *KIISE Transactions on Computing Practices* 28:57-62. [in Korean]
- Hwang MJ, Jeong HJ, Cho JH. 2018. The plan of stabilizing Gyeongnam Chung-yang green pepper farm income. *Journal of the Korea Academia-Industrial Cooperation Society* 19:387-396. [in Korean]
- Jang HD, Kim SH. 2016. Analysis of the economic value of the production of lily bulbs in Korea. *Korean Journal of Agricultural Science* 43:481-495. [in Korean]
- Jeon JM, Jang DH. 2020. A study on factors influencing agricultural income of peach farmers. *Journal of Industrial Economics and Business* 33:1201-1223. [in Korean]
- Jeon JY, Yoo CH, Park JK. 2016. Analysis on the agricultural income variability of rice and spicy vegetables. *Journal of Rural Development* 39:23-48. [in Korean]
- Lee GT, Lee GO. 2021. A study on the estimation of consumption expenditure of foodservice customers by machine learning algorithm. *International Journal of Tourism and Hospitality Research* 35:161-173. [in Korean]

- Li S, Wu P, Guan Z. 2021.\ Machine learning techniques for strawberry yield forecasting. AAEA Conference Paper, Annual Meeting, July 26-27, Kansas City, Missouri, USA.
- Oh JW, Kim HK, Kim IT. 2019. Design and implementation of fruit harvest time predicting system based on machine learning. Korean Journal of Smart Media Science 8-1:74-81. [in Korean]
- Soon BM, Cho SM, Kim SH. 2021. Impact of a reduction in the quality of shine muscat on the grape variety market using the Armington model. Korean Journal of Agricultural Science 48:911-926. [in Korean]