

기후가 화훼 농가 소득에 미치는 영향

신재은

소상공인시장진흥공단 부연구위원

The impact of climate on floriculture income

Shin, Jae Eun

Associate Research Fellow, Small Enterprise and Market Service, Daejeon, Korea

ABSTRACT

Climate change poses increasing threats to floriculture, a high-value agricultural sector that is particularly sensitive to environmental variability. This study examines climate variable impact on floriculture farm income in South Korea using household-level panel data from 2003~2023, employing traditional regression analysis and explainable machine learning models with SHAP (SHapley Additive exPlanations) analysis. Our analysis revealed two key findings: (1) temperature effects showed critical threshold patterns in which average temperatures between 12°C and 14°C positively affected income, while minimum temperatures less than 7°C reduced earnings by 16.3%; (2) adequate precipitation positively influenced income, but precipitation variability increased prediction uncertainty. Beyond climate factors, farm net income and agricultural gross revenue were the primary determinants of income variability, with minimum temperature emerging as the most influential climate factor. Farms with greater assets demonstrated higher climate resilience, which revealed adaptation inequalities. Machine learning models significantly outperformed linear regression, with the optimized Random Forest achieving the highest accuracy ($R^2=0.59$). Prediction accuracy showed a U-shaped pattern: high performance for short-term (2-year: $R^2=0.49$) and long-term (10-year: $R^2=0.53$) forecasts, but a decline for medium-term (6-year: $R^2=0.17$) predictions. These findings underscore the need for targeted climate interventions.

Key words : Climate Change, Floriculture Income, Explainable Machine Learning, SHAP and LIME, Regression and Prediction Modeling

1. 서론

최근 기후변화로 인한 지구 평균 기온 상승이 가속화되고 있다. IPCC (2022)에 따르면, 2011~2020년 전 세계 평균 지표 온도는 산업화 이전(1850~1900년) 대비 약 1.1°C 상승했으며, 이러한 추세가 지속될 경우 2021~2040년 사이 1.5°C 상승에 도달할 것으로 전망된다. 한국의 경우 지난 100년간 평균기온이 약 1.5°C 상승하여, 세계 평균보다 빠른 기후변화가 진행 중이다. 이러한 기후

시스템의 변화는 폭염, 한파, 집중호우 등 극단적 기상이변의 발생 빈도와 강도를 증가시켜 전 세계적으로 이상 기후 재해의 발생빈도가 증가하고 있으며, 농업 부문의 피해도 점차 심화되고 있다.

기후변화는 농업 생산성 및 소득의 불확실성을 높이는 주요 요인이다. 온도, 강수, 일조 등 기후요인의 변화는 작물 생육 및 수확량, 품질에 직접적인 영향을 미치며, 장기적으로는 재배 적지의 변화, 병해충 및 잡초 발생 양상 변화, 농업 기반시설의 파손 등 농업 전반에 걸쳐 복합적인

†Corresponding author : topaz2033@gmail.com (34097, 188 Bukyuseong-daero, Yuseong-gu, Daejeon, Korea. Tel. +82-42-363-3367) ORCID 신재은 0009-0006-6285-8901

리스크를 유발한다. 이러한 기후 리스크는 작물의 특성에 따라 그 영향 정도가 크게 달라지는데, 특히 화훼산업은 환경 조건에 대한 민감도가 높아 기후변화에 더욱 취약할 수 있다.

화훼산업은 단위면적당 부가가치가 높은 고소득 작물로서, 2023년 기준 화훼 재배농가의 평균 농업총수입은 약 8,400만원으로 일반 농가 대비 1.7배 높은 수준이다. 화훼산업은 단위면적당 부가가치가 높은 고소득 작물이다. 국내 화훼 재배농가는 일반 농가보다 높은 소득 수준을 유지하고 있으며, 화훼 산업은 전국 수만 개 농가에서 연간 수천억 원대의 생산액을 기록한다. 특히 경기·경남·전남 등 주요 주산지에서는 농업인의 상당 부분이 화훼업과 연관된 고용에 종사하며, 지역경제 활성화와 일자리 창출에 중요한 역할을 하고 있다. 화훼 유통, 가공, 장식 서비스업 등 연관 산업까지 포함할 경우, 그 고용 및 경제 파급 효과는 더욱 확대된다.

그러나 화훼산업은 기온, 강수, 일조량 등 환경요인에 매우 민감하여 기후변화에 특히 취약한 산업 중 하나이다. 국내외 연구에 따르면, 기후변화는 화훼 작물의 생육 기간 단축, 개화 불량, 품질 저하, 꽃 크기 감소, 색소 변이, 병해충 증가, 생산비용 상승 등 다양한 경로를 통해 화훼 산업의 생산성과 소득에 부정적 영향을 미친다고 일반적으로 보고되고 있다. 한 예로, 장미와 국화는 고온 스트레스에 취약해 꽃잎 수 감소, 색상 퇴색, 개화 실패율 증가 등의 현상이 관찰되며, 이는 시장성 하락과 농가 소득 감소로 직결된다(Liang, 2016; Zhou et al., 2025). 또한 잦은 집중 호우와 가뭄은 생육기 단계별 수분 스트레스를 유발해 뿌리 발달 억제, 영양분 흡수 장애 등의 문제를 초래한다(Shin et al., 2020). 특히 화훼농가는 시설투자비, 종묘비, 연료비 등 고정비용 비중이 높아 총수입 대비 경영비 비율이 높아서 소득 변동성이 큰 특징을 가지고 있어 기후 충격에 더욱 취약한 구조를 보여준다(Rural Development Administration, 2020).

이처럼 화훼산업은 지역 경제 활성화, 일자리 창출 등 사회·경제적으로 중요한 역할을 담당하지만, 기후변화 영향에 대한 정량적 분석이 부족하여 농가의 소득 안정화와 정책 설계의 근거 마련이 시급하다.

최근 정부와 학계에서는 기후변화에 대응한 농업 생산 기반 안정화, 스마트팜 등 첨단 기술 도입, 품종 개발, 경영 개선 지원 등 다양한 적응 전략의 필요성이 강조되고 있으나, 특히 화훼 부문에 대한 맞춤형 대응책은 충분히 마련되지 않은 상황이다. 이에 본 연구는 기온, 강수량, 일

조량 등 주요 기후 요인이 화훼 농가의 소득에 미치는 영향을 실증적으로 분석하고, 머신러닝 기반 예측모델을 통해 미래 소득의 변동성을 평가하고자 한다. 연구의 핵심 질문은 기후 요인이 화훼 소득에 어떠한 방식으로 영향을 미치는지에 대한 것으로, 이를 바탕으로 다음과 같은 두 가지 가설을 설정하였다.

첫째, 기온 상승은 특정 화훼 작물의 생육기간 단축과 품질 저하를 유발하여 농가 소득 감소로 이어질 것이다.

둘째, 강수 패턴 변화는 생장 및 개화 시기 변동성을 증가시켜 생산성과 소득의 불확실성을 높일 것이다.

본 연구는 농업 전체가 아닌 화훼라는 특정 고부가가치 작물에 초점을 맞춘 미시적 접근을 시도한다는 점, 그리고 기후 시나리오 기반의 장단기 소득 예측모델을 통해 화훼농가 경영 전략과 정책 지원의 실질적 근거를 제공한다는 점에서 차별성을 갖는다. 이에 본 연구는 계량경제학의 핵심 분석 도구인 OLS 회귀분석을 통해 변수 간 인과관계와 통계적 유의성을 확인하고, 추가적으로 설명 가능한 머신러닝(Explainable Machine Learning) 기법을 병행 적용함으로써 기후 변수와 경제적 요인 간의 복합적 상호작용과 비선형 관계를 종합적으로 규명한다는 점이 본 연구의 방법론적 강점이다. 화훼 농가 소득 결정 메커니즘은 기후 요인 간 복잡한 상호작용과 임계점 효과를 포함하고 있어 전통적인 선형 모델로는 포착하기 어려운 비선형적 특성을 갖는다. 이러한 한계를 극복하기 위해 본 연구는 Random Forest, XGBoost 등 앙상블 기법을 활용하여 변수 간 복합적 상호작용을 모델링하고, SHAP (Shapley Additive Explanations) 분석을 통해 블랙박스 모델의 해석 가능성을 확보하고자 한다. 궁극적으로, 기후변화에 취약한 화훼산업의 지속 가능성 확보와 농가 소득 안정화를 위한 실질적 대안을 제시하는 데 연구의 목적이 있다.

2. 선행연구

기후변화가 농업 생산성과 농가 소득에 미치는 영향은 국내외에서 지속적으로 연구되어 왔다. 특히, 온도 상승, 기후변동성, 극한기후의 증가 등이 작물 생산성은 물론 농가 소득 안정성에도 직접적이고 복합적인 영향을 미치는 것으로 나타났다. Ahmad et al. (2022)에 의하면, 최대 온도 2.8℃ 및 최저온도 2.2℃ 상승 시 벼 수확량은 15.2

~17.2%, 밀은 12~14.1% 감소할 것으로 예측하였으며, 이는 단순한 온도 변화가 아니라 개화기 고온 스트레스로 인한 수분 효율 저하와 같은 생리학적 메커니즘이 주요한 원인임을 밝혔다. EPA (2025)에 따르면 최신 보고서도 이러한 논의를 확장하면서, 기온이 1°C 상승할 때 주요 곡물의 생산성이 5~11% 감소한다는 정량적 예측을 제시하였다. 특히 주야간 온도차의 축소는 식물 호흡대사의 교란을 유발하여 생산성을 더욱 저하시킨다는 점에서, 단순 평균기온 상승이 아닌 온도 변화의 패턴 자체가 중요한 변수임을 알 수 있다. 또한, 작물 생육 단계별로 기후 민감도가 다르게 나타나며, 개화기 및 수분기에 가해지는 스트레스가 최종 수확량에 미치는 영향은 다른 시기보다 2~3배 더 큰 것으로 나타났다. 이에 더해, IPCC (2022)는 기후변화의 영향이 평균조건의 변화에 그치지 않고, 연속 고온일수와 같은 극한 기후 사건의 빈도 및 강도 증가가 농업에 비선형적 충격을 가하고 있다고 하였다. 예를 들어, 연속 3일 이상 고온현상이 발생할 경우 일반 고온에 비해 수확량 감소 효과가 평균 1.8배 더 크며, 이러한 현상은 최근 30년간 발생 빈도가 42% 증가한 것으로 나타났다.

기후변화의 영향은 단순히 생산성 저하에 그치지 않고, 농가 경제의 변동성과 불확실성을 심화시키는 방향으로 작용한다는 점도 점차 명확해지고 있다. Kim (2024)에 의하면 2003~2022년 9개 도의 농가 패널 데이터를 준모수적 방식으로 분석하여, 연평균기온 13°C를 기준으로 소득 효과가 양(+)에서 음(-)으로 전환되는 임계점을 확인하였다. 예컨대, 중부지역(평균 11~12°C)에서는 1°C 상승 시 농가소득이 4.2% 증가하였으나, 남부지역(14~15°C)에서는 오히려 3.8% 감소하는 양상이 나타나, 기후에 대한 지역 간 반응의 이질성이 분명하게 드러났다. 특히 화훼류는 생육에 적합한 온도 범위가 13.5±1.5°C로 비교적 좁아, 기후변화에 상대적으로 더 취약할 수 있다는 지적이 제기되었다. 또한, Schlenker and Roberts (2009)에 의하면 연평균 기온의 표준편차가 1°C 증가할 경우 농가소득의 변동계수가 평균 18.3% 상승하며, 이는 평균 기온 상승의 영향보다도 더 크게 작용한다는 메타분석 결과를 제시하였다. 이러한 결과는 기후변동성이 농가소득의 안정성에 더 심각한 위협이 될 수 있음을 시사한다. 아울러, Burke and Emerick (2021)에 따르면 고온과 강수량 간 상호작용의 중요성을 강조했는데, 동일한 고온 조건이라도 적절한 수분이 보장될 경우 고온 스트레스 효과가 76% 완화되는 반면, 건조 상태에서는 부정적 효과가 오히려 증폭된다는

복합적 메커니즘을 제시하였다. 이는 개별 기후변수의 독립적 영향만으로는 충분하지 않으며, 변수 간 상호작용에 대한 통합적 접근이 필요함을 의미한다.

이와 같은 기후위기에 대응하기 위해 다양한 농업 부문 적응 전략의 실효성에 관한 연구도 활발히 진행되고 있다. EPA (2025)의 비용-편익 분석결과에 따르면, 스마트 팜 기술의 도입은 물 사용 효율을 25% 향상시키고 병해충 피해를 18% 감소시키는 것으로 나타났으며, 이를 통해 농가 소득 변동성을 30% 이상 완화할 수 있었다. 이러한 기술적 개입은 기후변화의 부정적 영향을 상쇄하는 효과적인 수단으로 입증되고 있다. Ahmad et al. (2022)에 의하면 파키스탄 편자브 지역에서 파종시기 15일 조정과 정밀 관개 기술 도입이 수확량 감소를 42%까지 상쇄하며, 비용 대비 편익 비율(B/C ratio)은 2.7에 이르는 높은 경제적 타당성을 입증하였다. 이는 비교적 적은 비용의 기술적·행태적 조정만으로도 상당한 적응 효과를 달성할 수 있음을 보여준다. 한편, OECD (2023)에 따르면 기후 적응을 자율적 및 계획적 적응으로 구분하고, 자율적 적응은 농가 단위의 능동적 대응에 해당하는 반면, 계획적 적응은 정책과 제도적 지원에 기반한 체계적 개입으로 설명하였다. 두 적응 유형의 결합이 기후 충격 완화 효과를 약 60%까지 증가시킨다는 분석결과는 기후 대응에서 공공의 역할이 중요함을 재확인시켜 주었다.

그러나 이러한 선행연구들은 기후변화와 농업의 관계에 대한 유의미한 통찰을 제공했지만 몇 가지 한계가 있다. 첫째, 농업 연구개발 투자는 식량작물 중심으로 편중되어 있어, 화훼작물과 같은 고부가가치 원예작물에 대한 연구는 상대적으로 활발히 이루어지지 않고 있다. 실제로 미국 2014년 Farm Bill에서 과일, 채소, 화훼 등 특수작물에 대한 지출은 7억 달러에 그친 반면 곡물작물에는 50억 달러가 투입되어 약 7배의 투자 격차를 보이고 있다 (UC Davis Horticulture Innovation Lab et al., 2018). 이는 단위 면적당 경제적 가치가 높은 화훼 산업의 중요성을 고려할 때 심각한 연구 격차로 볼 수 있다. 특히 화훼 산업은 기상 조건에 대한 민감도가 높고 고부가가치 특성을 가지고 있어, 기후변화 대응 측면에서 차별화된 접근이 요구되에도 불구하고 연구적 관심이 부족한 것으로 나타났다. 둘째, 분석 방법론적 측면에서 대부분의 연구가 선형 회귀모델에 집중되어 있어, 기후와 농업 간의 비선형 관계나 임계점 효과를 충분히 포착하지 못하였다. 온도-생산성 관계에서 나타나는 비선형적 반응과 임계치 효과는 다수의 연구에서 관찰되에도 불구하고, 이를 정교하게 모

텔링하기 위한 비모수적 머신러닝 기법의 적용은 제한적이었다. 셋째, 시간적 범위의 제약도 뚜렷하게 나타났다. 대다수 연구가 5년 이내의 단기 분석에 머물러 있어, 기후 변화의 누적적·장기적 영향은 상대적으로 간과되었다. 특히 투자 회수 기간이 긴 화훼 작물의 경우, 중장기적 분석 없이는 정책적 개입의 시의성과 효과성을 판단하기 어렵다는 방법론적 한계가 있다.

본 연구는 이러한 한계점을 보완하고자, 2003년부터 2023년까지의 21년간 화훼 농가 패널 데이터를 구축하고, Random Forest, XGBoost 등 비선형 머신러닝 알고리즘을 활용하여 기후변수의 임계점 및 상호작용 효과를 정밀 분석하였다. 특히 SHAP (Shapley Additive Explanations) 분석을 통해 각 기후변수의 기여도를 가시화하고, 개화기 일사량과 강수량 간의 복합 작용을 고려한 생물학적 생육도일(BEGDD) 개념을 도입함으로써 미세기후 영향까지 포착하고자 했다. 이러한 접근은 기존 선형 모델에서 간과되었던 복합적 기후 상호작용을 식별하는 데 중요한 방법론적 진전을 제공한다. 또한, 경영주 특성별 이질성을 분석해 동일한 기후조건하에서도 농가의 적응 능력 차이를 밝혀냄으로써, 기후 변화 대응 정책의 정교화 및 계층화에 기여할 수 있는 실증적 기반을 마련했다. 이와 같은 다층적이고 통합적인 접근은 향후 농업 부문의 기후 복원력 제고와 지속 가능한 발전 전략 수립에 핵심적 토대를 제공할 것으로 기대된다.

3. 데이터 및 연구 방법론

3.1. 데이터 수집 및 구축

본 연구는 2003년부터 2023년까지의 농가경제조사 자료를 활용하여, 기후변화가 화훼 농가 소득에 미치는 영향을 분석하였다. 이를 통해 향후 화훼 농가의 소득 안정화 정책과 기후 적응 지원 방안 설계에 실질적인 시사점을 제공하는 것을 목표로 하였다. 연구에 사용된 주요 변수는 다음과 같다. 종속 변수는 화훼 농가의 연간 소득이며, 주요 독립 변수는 월평균 기온과 강수량 등 기후 요인이다. 추가로 경지 규모와 농업 자산 등의 경제적 변수도 설명 변수로 포함하였다. 분석 대상은 영농형태 코드가 5인 화훼 재배 농가로 한정하였으며, 시도별 기후 데이터를 매핑한 뒤 연도별 기온 평균값을 계산해 농가 데이터와 병합하였다. 기후 데이터는 기상청의 중관기상관측(ASOS) 자료를 활용하여 각 농가의 소재지와 가장 가까

운 관측소의 데이터를 할당하는 방식으로 구축하였다.

데이터 전처리 과정에서는 먼저 결측값을 중앙값으로 대체하여 데이터 왜곡을 최소화하였으며, 연속형 변수는 StandardScaler를 이용해 표준화하였다. 농가 소득은 로그 변환을 통해 정규성을 확보하였고, 범주형 변수는 성격에 따라 Label Encoding과 One-Hot Encoding으로 처리하였다. 특히 지역 변수는 온도와 강수량 패턴의 차이를 반영하기 위해 One-Hot Encoding을 적용하였다. 또한, 기후 변수 간 다중공선성 문제를 해결하기 위해 분산 팽창 계수(VIF)를 활용하여 높은 상관관계를 가진 변수를 제거하였다. 다중공선성 문제를 해결하기 위해 VIF 값 10을 기준으로 변수를 선별하였으며, 최종 모델에 포함된 모든 변수의 VIF 값은 허용 범위(VIF < 10) 내에 있는 것으로 나타났다.

3.2. 분석 방법론

분석 방법으로는 전통적인 OLS 회귀분석과 함께, 머신러닝 모델인 Random Forest, XGBoost, LightGBM, Linear Regression을 활용하였다. 먼저, OLS 회귀분석은 기후 변수(기온, 강수량)와 경제 변수(경지 규모, 자산 등)가 화훼 농가 소득에 미치는 영향을 통계적으로 평가하는 데 사용되었다(Greene, 2020). 해당 회귀식은 식 (1)과 같다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (1)$$

여기서 y 는 농가소득(로그변환 값), X_1, X_2, \dots, X_n 은 각각 기후 및 경제 변수, $\beta_1, \beta_2, \dots, \beta_n$ 은 해당 변수의 회귀계수, ϵ 는 오차항이며, 회귀 계수의 크기와 통계적 유의성을 통해 각 변수의 영향력을 판단하였다. 이와 함께, 비선형적 관계를 포착하고 예측 성능을 높이기 위해 머신러닝 기법을 적용하였다. Random Forest, XGBoost, LightGBM 모델은 Voting 및 Stacking 방식의 앙상블 기법을 통해 예측력을 향상시켰으며, GridSearchCV와 RandomizedSearchCV를 이용해 하이퍼파라미터를 최적화하였다. 모델의 예측 성능은 결정계수(R^2), 평균제곱근오차(RMSE), 평균절대오차(MAE)와 같은 지표로 평가하였다.

Random Forest는 다수의 의사결정나무를 구축하고 그 예측값의 평균을 사용하는 앙상블 방법으로, 과적합 위험을 줄이면서 예측 성능을 향상시키는 장점이 있다. XGBoost와 LightGBM은 그래디언트 부스팅 계열의 알고리즘으로, 이전 모델의 오차를 보완하는 방식으로 순차적

으로 모델을 구축함으로써 높은 예측 정확도를 달성한다. 이러한 머신러닝 모델은 변수 간 복잡한 상호작용과 비선형 관계를 효과적으로 포착할 수 있다는 점에서 기후 변수와 농가 소득 간의 관계 분석에 적합하다.

또한, 모델 해석력을 강화하기 위해 SHAP (Shapley Additive Explanations)와 LIME (Local Interpretable Model-agnostic Explanations) 기법을 사용하였다. SHAP 분석을 통해 각 특성이 예측값에 기여하는 정도를 정량화하였으며, LIME을 통해 개별 예측에 대한 로컬 설명을 제공함으로써 모델의 설명력을 보완하였다. 특히 SHAP 값은 협력게임이론에 기반한 방법으로, 각 변수가 예측에 미치는 한계 기여도를 일관성 있게 측정할 수 있어 변수 중요도 평가에 유용하다.

본 연구의 분석 절차는 다음과 같은 세 단계로 구성하였다.

첫째, 2003 ~ 2023년의 농가경제조사 및 기후 데이터를 기반으로 데이터의 분포와 정규성을 검토한 뒤, OLS 회귀분석을 통해 주요 변수들의 통계적 유의성을 확인하였다. 이 과정에서 Durbin-Watson 검정을 통해 자기상관 여부를 확인하고, Breusch-Pagan 검정을 통해 이분산성 존재 여부를 점검하였다.

둘째, 선형 회귀의 한계를 보완하고 변수 간 비선형 관계를 반영하기 위해 Random Forest, XGBoost, LightGBM 등 머신러닝 모델을 적용하여 화훼 농가의 소득을 2년, 4년, 6년, 10년 단위로 예측하였다. 이 과정에서 k-fold 교차검증(k=5)을 실시하여 모델의 일반화 성능을 평가하였으며, SHAP 분석을 통해 각 변수의 기여도를 시각화하고, 기후 변수의 비선형적 영향력을 해석하였다.

셋째, 예측 성능 향상을 위해 하이퍼파라미터 튜닝 및 Stacking 기법을 활용한 앙상블 모델을 구축함으로써, 단일 모델의 성능 한계를 극복하고 예측의 안정성과 정확도를 동시에 확보하고자 하였다. Stacking 기법은 Random Forest, XGBoost, LightGBM을 기반 모델로 하고, 메타 모델로는 Ridge Regression을 채택하여 개별 모델의 예측 강점을 결합하는 방식으로 구현하였다.

이에 본 연구는 기후 및 경제 변수의 복합적 영향을 정량적으로 평가하기 위해, 전통적인 통계 분석과 설명 가능한 머신러닝 기법을 결합하였다. 특히 SHAP 및 LIME과 같은 해석 기법을 활용하여 머신러닝 모델의 블랙박스 특성을 보완함으로써, 분석 결과의 실용성과 정책적 함의를 극대화하고자 하였다.

4. 연구결과

4.1. 데이터 분포 및 정규성 검토

본 연구에서는 농가 소득을 비롯한 주요 경제 변수 및 농업 관련 특성을 파악하기 위해 기초통계량 분석을 실시하였다. 이 분석은 변수들의 전반적인 분포, 평균, 중앙값, 최소-최대값 등을 확인함으로써 데이터의 이상치 여부 및 변수 간 차이를 탐색하는 데 목적이 있다. 분석 대상은 영농형태 코드가 5인 화훼 재배 농가 중 농가 소득이 0보다 큰 표본만을 포함하였으며, 주요 변수의 기초통계량은 Table 1과 같다.

분석 결과, 농가 소득의 평균은 약 4,739만 원, 중앙값은 약 3,216만 원으로 나타났다. 평균값이 중앙값보다 약 47% 높게 나타난 것은 소득 분포가 오른쪽 꼬리가 긴 비대칭적 분포(skewed distribution)를 보이며, 일부 대규모 농가의 높은 소득이 평균값을 끌어올린 것으로 해석된다. 소득의 최소값은 약 8천 원, 최대값은 약 9억 7천만 원으로 최대/최소 격차가 10만 배 이상으로 상당한 소득 불평등이 존재함을 확인할 수 있다. 이러한 극단적 격차는 화훼 농가 간 경영 규모와 효율성의 차이가 매우 크다는 점을 드러냈다.

기후 변수 분석 결과, 연평균 기온은 12.91°C, 평균 최저기온은 8.07°C, 평균 최고기온은 18.53°C로 나타났으며, 이는 연구 대상 지역이 비교적 온난한 기후 특성을 지닌다는 점을 시사한다. 최저기온과 최고기온의 격차(일교차)는 평균 10.46°C로, 작물의 생리적 스트레스에 영향을 미칠 수 있는 수준이다. 또한, 연간 강수량은 평균 1,291 mm로 최소 713 mm에서 최대 2,281 mm까지 지역 및 연도별 편차가 크게 나타났는데, 이는 강수 패턴의 변동성이 화훼 생산에 잠재적 리스크 요인으로 작용할 수 있음을 의미한다.

이외에도 평균 상대습도는 68.88%, 연간 일조시간은 2,308시간, 일사량은 4,735 MJ/m²로 집계되었다. 이들 기후 요인은 광합성 활동과 성장 조절에 직접적인 영향을 미치는 변수들로, 특히 화훼 작물의 개화 시기와 품질에 중요한 결정인자로 작용한다. 상대습도의 경우 최소 56.6%에서 최대 75.65%로 변동 폭이 크지 않았지만, 일조시간은 최소 1,449.7시간에서 최대 2,771시간으로 상당한 지역적 편차를 보였다.

이상의 분석 결과는 향후 회귀 및 예측 모델의 정확도 향상을 위해 데이터 변환과 이상치 처리가 필수적임을 보

Table 1. Descriptive statistics of key variables

Variable	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
Cultivated Area (ha)	2.51	1.83	1	1	2	3	9
Farm Income (1,000 KRW)	47,395	49,699	8	19,180	32,161	53,819	971,341
Farm Net Income (1,000 KRW)	33,419	43,684	-30,521	6,300	20,247	46,172	781,708
Agricultural Gross Revenue (1,000 KRW)	84,026	75,672	489,850	29,566	62,100	118,303	836,519
Farm Assets (1,000 KRW)	558,231	548,120	9,451	220,579	397,973	702,944	4,445,952
Land Area (m ²)	8,166	12,887	0	885	3,982	9,157	149,170
Average Temperature (°C)	12.91	1.18	10.59	11.9	12.67	13.58	16.67
Average Minimum Temperature (°C)	8.07	1.69	5.78	6.66	7.52	8.9	13.9
Average Maximum Temperature (°C)	18.53	0.84	15.91	17.96	18.45	18.89	20.67
Average Sea-level Pressure (hPa)	1,016.5	0.45	1,015.2	1,016.2	1,016.58	1,016.82	1,017.32
Total Precipitation (mm)	1,291.21	297.15	713	1,050.68	1,233.58	1,486.38	2,281.6
Average Relative Humidity (%)	68.88	4.29	56.6	65.08	69.32	72.4	75.65
Total Sunshine Duration (hr)	2,308.85	169.39	1,449.7	2,205.24	2,294.08	2,418.67	2,771
Total Solar Radiation (MJ/m ²)	4,735.11	689.96	2,133.62	4,437.68	4,873.16	5,180.99	5,962.13
Average Wind Speed (m/s)	1.85	0.58	1.28	1.53	1.58	1.84	4.18

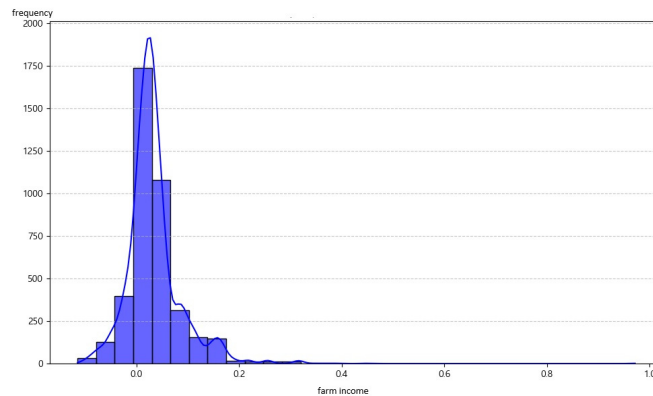


Fig. 1. Distribution of farm income and KDE visualization

여준다. 분석 결과를 종합하면, 향후 모델링에서는 데이터 변환과 이상치 처리가 핵심적 과제로 나타났다. 특히 소득 변수의 비대칭 분포는 로그 변환이나 비모수적 방법론의 도입을 필요하다. 기초 통계 분석 결과 농가 소득 분포의 왜도(skewness)가 크다는 점이 확인되어, 보다 정밀한 분포 분석을 위해 히스토그램 및 커널 밀도 추정(KDE) 시각화를 수행하였다(Fig. 1).

히스토그램과 KDE 곡선을 보면, 농가 소득은 대부분 0 ~ 2억 원 사이에 몰려있고, 소수 농가에서 높은 소득을 올리고 있는 것으로 나타났다. 분포는 전형적인 우측 편향 (Skewed Right) 형태를 띠며, 소득의 분산이 크고 고소득

농가가 전체 평균을 왜곡시키고 있음을 알 수 있다. 구체적으로 소득 상위 10%에 해당하는 농가들이 전체 소득의 35% 이상을 차지하는 높은 집중도를 보였다. 이러한 소득 분포의 비대칭성과 편차는 단순한 평균값만으로 농가의 실질적 수익성을 설명하기 어렵기 때문에 분포의 정규성을 보다 정확히 검토하기 위해 Q-Q 플롯 분석을 수행하였다(Fig. 2).

Q-Q 플롯을 통한 정규성 진단 결과, 원데이터는 양쪽 극단에서 직선 이탈이 심각하여 정규분포 가정을 만족하지 못했다. 우측 이상치가 특히 분포 왜곡의 주요 원인으로 확인되어 모수적 분석에는 한계가 있었으나, 이상치

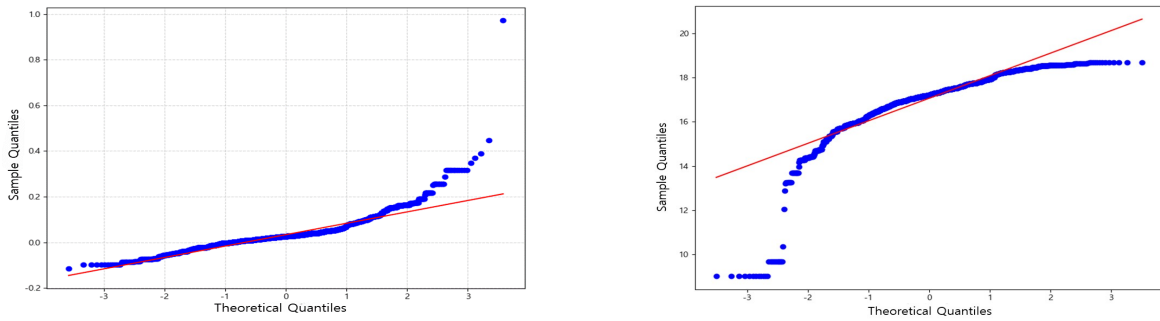


Fig. 2. Results of Q-Q plot analysis

제거와 로그 변환을 순차적으로 적용한 결과 점들이 45도선에 근접하며 정규성이 크게 개선되었다. 그러나 여전히 일부 극단값은 대각선에서 벗어나 있는 것으로 보아, 농가 소득은 완벽한 정규분포를 따르지 않으며, 로그 변환과 이상치 처리 이후에도 비정규성은 일정 부분 존재함을 알 수 있다. 이는 화훼 농가 소득이 개별 요인들의 독립적인 영향을 단순 합산한 결과가 아니라, 기후·경제·경영 요인들 간의 복합적 상호작용에 의해 형성됨을 보여준다. 따라서 이후 분석에서는 이러한 비선형적 특성을 반영할 수 있는 모델 선택이 필요하며, 예측 모델링과 변수 해석에서도 이를 고려해야 한다.

4.2. 다중 선형 회귀 분석결과

4.2.1. OLS 회귀 분석 결과

독립 변수 간 다중공선성을 점검하기 위해 분산 팽창 계수(Variance Inflation Factor, VIF)를 계산하였다. VIF는 특정 독립 변수가 다른 독립 변수들과 얼마나 밀접한 상관관계를 갖는지를 판단하는 데 활용되는 지표로, 일반적으로 값이 10을 초과할 경우 다중공선성이 심각한 것으로 간주한다. Table 2는 최종적으로 선정된 변수들의 VIF 분석 결과를 나타낸 것이다.

분석 결과, 대부분의 변수에서 VIF 값이 5 미만으로 나타나 다중공선성 문제는 크지 않은 것으로 확인되었다. 다만, 평균최저기온(6.37)은 상대적으로 높은 VIF 값을 보여주었는데, 이는 온도 변수 간 상관성과 소득 관련 변수 간 유사성에 기인한 것으로 해석된다.

하지만 모든 변수의 VIF가 10 미만이므로 회귀분석을 진행하는 데 통계적으로 유의한 제약은 없는 것으로 판단된다. 경제적 해석의 일관성을 고려하여 본 연구에서는 모든 변수들을 유지한 상태에서 분석을 진행하였다. 다중 선형회귀(OLS, Ordinary Least Squares) 분석을 통해 농가

소득에 영향을 미치는 주요 요인들을 파악하였다. 종속 변수는 농가소득이며, 독립 변수로는 경지규모, 농업총수입, 농가자산, 토지 관련 변수, 주요 기후 변수(평균기온, 강수량 등), 그리고 지역 더미 변수(시도), 인구통계학적 특성(성별, 교육 수준 등)을 포함하였으며, 분석 결과는 Table 3과 같다.

분석 결과, 회귀모델의 설명력(R²)은 0.439이고 Adj. R² 또한 0.437로 유사한 값을 보여 모델의 적합성이 적절한 것으로 판단된다. F-통계량은 375.0으로 나타났으며 모델 전체가 통계적으로 유의함을 확인하였다.

주요 변수들의 영향을 살펴보면, 농가순소득은 계수값 1.0714로 가장 강한 설명변수로 나타났으며, 이는 농가순소득이 1단위 증가할 때 농가소득이 약 107% 증가하는 관계를 의미한다. 이러한 강한 양의 관계는 농가의 실질적인 수익성이 전체 소득 형성에 핵심적 역할을 함을 보

Table 2. Results of multicollinearity VIF analysis

Variable	VIF
Average Minimum Temperature (°C)	6.37
Average Maximum Temperature (°C)	4.92
Average Sea-level Pressure (hPa)	2.1
Total Precipitation (mm)	1.28
Average Relative Humidity (%)	1.99
Total Sunshine Duration (hr)	1.36
Total Solar Radiation (MJ/m ²)	1.17
Cultivated Area	2.43
Agricultural Gross Revenue	1.58
Land Area	2.44
Land Value	1.27
Major Plant Assets	1.16

저온 스트레스가 국화의 발아 수 감소, 광합성 능력 저하, 세포막 손상, 로제트 현상 유발 등을 통해 생산성과 관상적 품질에 부정적 영향을 미치며, 이로 인해 겨울철 가열 시설의 필요성과 경제적 손실이 증가한다고 보고하였다. 이러한 결과들은 화훼 작물이 특정 온도 범위에서 최적 생육 조건을 가지며, 이 범위를 벗어날 경우 생산성과 품질에 모두 부정적인 영향을 받을 수 있음을 시사한다.

합계 강수량(0.0581)과 합계 일조시간(0.0395)도 통계적으로 유의미한 양의 영향을 미치는 것으로 나타났다. 이는 적정 수준의 수분 공급과 충분한 광합성 활동이 화훼 생산성을 높인다는 것을 보여준다. 특히 강수량의 경우, 화훼 작물의 생육기 수분 스트레스를 완화시켜 품질과 생산량 향상에 도움을 준다고 해석할 수 있다.

지역 변수들의 경우, 대부분의 지역 더미 변수가 음의 계수로 나타나며, 이는 기준 지역(인천광역시) 대비 다른 지역의 화훼 농가 소득이 전반적으로 낮은 경향을 보임을 의미한다. 특히 경상남도(-0.3651)와 경상북도(-0.3046)의 계수가 가장 큰 음의 값을 나타내어 지역 간 소득 격차 상당히 존재함을 드러낸다. 이러한 지역 간 차이는 기후적 요인 외에도 시장 접근성, 농업 인프라, 지역별 정책 지원의 차이 등 여러 가지 요인에 기인한다고 볼 수 있다.

인구통계학적 변수 중 교육 정도는 -0.0132로 나타나, 교육 수준이 높을수록 농가 소득이 낮아지는 경향이 확인되었다. 이러한 결과는 직관적으로 예상되는 방향성과는 다소 상반되는데, 이는 고학력자의 경우 영농 경험이 부족하거나, 농업 이외의 경제활동과 병행하는 비율이 높아 농업 소득에 대한 집중도가 상대적으로 낮기 때문으로 해석할 수 있다. 한편, 성별 변수의 경우 10% 유의수준($p < 0.1$)에서 통계적 유의성이 나타났으며, 남성 농가주거 여성 농가주거에 비해 다소 낮은 소득을 기록하는 경향이 나타났다.

선형 회귀모델의 진단 결과를 살펴보면, Durbin-Watson 통계량이 0.573으로 나타나 양의 자기상관 문제가 존재함을 시사한다. 또한 Jarque-Bera 테스트 결과 통계량이 2.570이고 p-value가 0.00으로 나타나 잔차의 정규성 가정이 위배되는 것으로 확인되었다. 이러한 진단 결과는 선형 회귀모델의 기본 가정들이 일부 충족되지 않음을 의미하며, 이로 인해 모델의 예측 정확성과 신뢰구간 추정에 한계가 있을 수 있다. 따라서 후속 분석에서는 이러한 한계를 보완할 수 있는 비선형 모델링 접근이 필요하다고 판단된다. 반면, 머신러닝 모델은 이러한 통계적 가정을 완화하고, 비선형 관계와 복잡한 상호작용을 포착할 수 있어

차별화된 강점을 보인다. 본 연구에서는 각 분석 방법의 고유한 목적에 따라 활용하였다. OLS 회귀분석은 기후 변수가 화훼 농가 소득에 왜 그리고 어느 정도 영향을 미치는지를 규명하고, 정책적 개입의 근거를 마련하는 데 목적이 있었다. 반면 머신러닝 기법은 복잡한 기후와 경제의 상호작용 속에서 농가 소득이 언제, 어떻게 변동할지를 예측하는 모델을 구축하는 데 초점을 두었다. 이처럼 분석 목적의 차별화를 통해 화훼 농가 소득 결정 메커니즘에 대한 설명과 예측을 상호 보완적인 관점에서 종합적으로 접근하였다.

4.3. 예측 모델 비교 및 SHAP 분석 결과

4.3.1. 예측 모델 비교

본 연구는 화훼 농가의 미래 소득을 예측하기 위해 5가지 머신러닝 모델을 적용하였다. 적용된 모델은 Random Forest, Gradient Boosting, XGBoost, LightGBM, Linear Regression 이며, 각각 2년, 4년, 6년, 10년 후 농가소득을 예측하였다.

Table 4의 결과를 보면, 단기 예측(2년 후)에서는 Random Forest 모델이 가장 우수한 성능을 보인다(MSE=0.1165, $R^2=0.49$). 이는 2년 후 농가 소득 변동의 49%를 설명할 수 있음을 의미한다. XGBoost($R^2=0.49$)와 LightGBM($R^2=0.49$)도 유사한 예측력을 나타낸다. 이러한 결과는 트리 기반 앙상블 모델들이 화훼 농가 소득의 복잡한 결정 구조를 효과적으로 포착할 수 있음을 보여준다.

반면, Linear Regression은 MSE가 0.2599로 가장 높고 R^2 이 -0.14로 음수 값이 나타난다. 음수 R^2 는 해당 모델의 예측력이 단순히 종속변수의 평균값을 예측하는 것보다도 못하다는 의미로, 모델이 데이터의 패턴을 전혀 포착하지 못했음을 나타낸다. Linear Regression의 일관된 음수 R^2 값은 화훼 농가 소득 예측에 있어 머신러닝 기법의 필요성을 실증적으로 뒷받침하며, 기후-농업 경제 관계가 단순한 선형 관계를 넘어서는 복잡한 시스템임을 의미한다.

4년 후에는 Random Forest가 $R^2=0.36$ 으로 여전히 최고 성능을 보이지만, 설명력이 감소한다. 6년 후에는 모든 모델의 성능이 $R^2=0.17$ 수준으로 급격히 저하된다. 이는 중기적 소득 변동이 현재 변수들로 예측하기 어렵다는 것을 보여주며, 화훼 산업의 구조적 변화와 기후 변동성이 복합적으로 작용하는 기간이기 때문으로 해석된다.

특히 장기 예측(10년 후)에서는 중기 예측과 달리 성능이 개선되는 경향이 관찰된다. Random Forest가

Table 4. Short- and long-term income forecasting results for floriculture farms

Prediction Time	Model	Mean Squared Error (MSE)	R-squared	Model Evaluation
2 years later	Random Forest	0.1165	0.49	Most suitable for short-term prediction
	Gradient Boosting	0.1405	0.38	Medium performance
	XGBoost	0.1166	0.49	Similar to Random Forest
	LightGBM	0.1168	0.49	Similar to Random Forest
	Linear Regression	0.2599	-0.14	Poor performance
4 years later	Random Forest	0.1584	0.36	Most suitable for medium-term prediction
	Gradient Boosting	0.1990	0.19	Low performance
	XGBoost	0.1656	0.33	Close to Random Forest
	LightGBM	0.1669	0.32	Medium performance
	Linear Regression	0.2472	0	Poor performance
6 years later	Random Forest	0.2940	0.17	Medium performance
	Gradient Boosting	0.3350	0.06	Low performance
	XGBoost	0.2944	0.17	Similar to Random Forest
	LightGBM	0.2963	0.17	Similar to Random Forest
	Linear Regression	0.3696	-0.04	Poor performance
10 years later	Random Forest	0.1097	0.53	Most suitable for long-term prediction
	Gradient Boosting	0.1913	0.17	Low performance
	XGBoost	0.1109	0.52	Close to Random Forest
	LightGBM	0.1126	0.51	Medium performance
	Linear Regression	0.2486	-0.07	Poor performance

MSE=0.1097, $R^2=0.53$ 으로 가장 우수한 성능을 기록했으며, XGBoost (0.1109, $R^2=0.52$)와 LightGBM (0.1126, $R^2=0.51$)도 근접한 예측력을 나타낸다. 이러한 장기 예측의 성능 향상은 단기적 변동성이 상쇄되고, 기후 및 경제적 요인의 장기적 추세가 더욱 명확하게 포착되기 때문으로 이해된다. 특히 Random Forest는 각 의사결정나무의 독립적 학습과 배깅(bagging)을 통한 편향-분산 트레이드 오프(bias-variance trade-off) 최적화가 장기 예측에 유리하게 작용한 것으로 판단된다.

분석 결과를 종합하면, Random Forest, XGBoost, LightGBM은 대부분의 예측 시점에서 안정적인 예측력을 보인 반면, 선형 회귀 모델은 모든 시점에서 예측 성능이 상대적으로 낮게 나타났다. 이러한 모델별 성능 차이는 화훼 농가 소득 결정 메커니즘의 비선형성과 복잡성을 뒷받침한다.

OLS 회귀분석의 설명력($R^2=0.439$)과 Random Forest 모델의 10년 후 예측 성능($R^2=0.53$)을 비교하면, 장기 예측에서 머신러닝 기법이 전통적인 회귀분석보다 더 나은 설명력을 가질 가능성을 시사한다. 이는 Random Forest가

비선형 관계와 복합적 상호작용을 효과적으로 모델링할 수 있기 때문이며, 예측 시점에 따라 최적의 모델을 선택하는 차별화된 전략이 필요함을 보여준다.

Table 5는 4개 주요 모델의 예측 성능을 시점별로 비교한 결과이다. MSE(평균제곱오차)는 예측값과 실제값의 차이를 제공한 값의 평균으로, 값이 작을수록 예측 정확도가 높다. R^2 (결정계수)는 모델이 데이터 변동의 얼마나 많은 부분을 설명하는지를 나타내며, 1에 가까울수록 설명력이 높다.

첫째, 머신러닝 모델과 선형회귀 간의 명확한 성능 차이가 확인되었다. Random Forest, XGBoost, LightGBM은 모든 예측 시점에서 안정적인 성능을 보인 반면, Linear Regression은 일관되게 음수 R^2 값(-0.14 ~ -0.07)을 기록했다. 이는 선형 회귀 모델의 예측력이 단순히 종속변수의 평균값을 예측하는 것보다 못함을 의미하며, 화훼 농가 소득 결정 메커니즘이 단순한 선형 관계를 넘어서 복잡한 비선형적 특성을 지닌다는 것을 입증한다.

둘째, 예측 성능에서 독특한 U자형 패턴이 관찰되었다. 단기(2년) 예측에서는 $R^2=0.49$ 로 양호한 성능을 보였으

Table 5. Performance comparison of machine learning models for floriculture farm income prediction (2, 4, 6, 10 years)

Forecast Period	Evaluation Metric	Random Forest	XGBoost	LightGBM	Linear Regression
2-year	MSE	1.17×10^{15}	1.17×10^{15}	1.17×10^{15}	2.60×10^{15}
	R-squared	0.49	0.49	0.49	-0.14
4-year	MSE	1.58×10^{15}	1.66×10^{15}	1.67×10^{15}	2.47×10^{15}
	R-squared	0.36	0.33	0.32	-0.00
6-year	MSE	2.94×10^{15}	2.94×10^{15}	2.96×10^{15}	3.70×10^{15}
	R-squared	0.17	0.17	0.17	-0.04
10-year	MSE	1.10×10^{15}	1.11×10^{15}	1.13×10^{15}	2.49×10^{15}
	R-squared	0.53	0.52	0.51	-0.07

나, 중기(4년: $R^2=0.33$, 6년: $R^2=0.17$)에서는 급격히 저하되었고, 장기(10년) 예측에서는 $R^2=0.51 \sim 0.53$ 으로 다시 향상되는 패턴을 보였다. 이는 화훼 소득 예측에서 중기적 변동성이 가장 예측하기 어려운 구간임을 시사하며, 중기 예측의 어려움이 정책 설계 시 특별한 주의를 요한다는 것을 보여준다.

셋째, 머신러닝 모델 간에는 상대적으로 작은 성능 차이가 나타났다. 10년 예측에서 Random Forest가 $R^2=0.53$ 으로 가장 높은 성능을 보였고, XGBoost (0.52), LightGBM (0.51)이 근소한 차이로 뒤를 이었다. 이는 세 모델 모두 화훼 소득의 복합적 결정 구조를 효과적으로 포착할 수 있음을 의미한다.

넷째, MSE 값의 변화 패턴은 R^2 와 일치하는 경향을 보였다. 6년 예측에서 MSE가 가장 높게 나타났다가(2.94×10^{15}), 10년 예측에서는 다시 감소(1.10×10^{15})하는 패턴을 보여 예측 정확도의 U자형 변화를 뒷받침했다.

이러한 결과는 화훼 농가 소득 예측에서 예측 시점별 차별화된 접근 전략의 필요성을 제시한다. 단기와 장기 예측에서는 머신러닝 모델들이 유용한 성능을 보이는 반면, 중기 예측의 어려움은 정책 설계 시 이 구간에 대한 특별한 주의가 필요함을 시사한다.

4.2.2. 변수 중요도 해석(Shap/LIME)

머신러닝 기반 예측 모델의 해석을 위해, SHAP (Shapley Additive Explanations) 기법을 적용하였다. 이 분석 기법은 단순한 예측 정확도 비교를 넘어, 각 변수가 모델 예측에 어떻게 기여했는지를 정량적으로 분해해 보여준다(Fig. 5).

Fig. 5는 SHAP 분석 결과를 시각화한 것이다. x축은 각 변수가 예측에 미치는 영향의 크기와 방향을 나타내

며, 점의 색상은 해당 변수 값의 크기를 의미한다. 빨간색은 높은 값을, 파란색은 낮은 값을 나타낸다.

분석 결과, 농가 순소득(Farm Net Income)이 예측에 가장 큰 영향을 미쳤으며, 총 농업수익(Total Agricultural Income), 작물가치(Crop Value), 농가 자산(Farm Assets) 등도 주요 영향 변수로 확인되었다. 이들 변수는 SHAP 값의 분포가 넓고, 변수 값이 높을수록 소득 예측치 증가에 기여하는 경향을 보였다.

기후 변수 중에서는 평균 최저기온(Average Minimum Temperature)의 중요도가 상대적으로 높았으며, 값이 낮을수록 농가 소득에 부정적인 영향을 미치는 것으로 나타났다. 이는 저온 스트레스가 화훼의 생산성과 품질에 악영향을 줄 수 있음을 시사한다. 평균기온(Average Temperature), 평균 최고기온(Average Maximum Temperature), 평균 풍속(Average Wind Speed)도 일정 수준의 영향을 미쳤으나, 경제 변수에 비해서는 중요도가 낮았다.

구조적 특성 변수인 토지 규모(Land Area), 교육 수준(Education Level), 토지 감정가(Land Appraisal Value) 등도 일정한 예측 기여를 보였지만, 경제 및 기후 변수에 비해서는 영향력이 제한적이었다.

이러한 SHAP 분석 결과는 농가의 경제적 기반과 기후 조건이 소득 안정성에 미치는 복합적인 영향을 시각적으로 제시하며, 향후 정책 설계 시 중점적으로 고려해야 할 개입 지점을 실증적으로 보여준다.

4.4. 모델 적합성 개선 및 하이퍼파라미터 튜닝

모델 예측 성능 향상을 위해 앙상블 모델링과 하이퍼파라미터 튜닝 기법을 적용하였다. 먼저 Random Forest 모델에 대해 주요 하이퍼파라미터($n_estimators$,

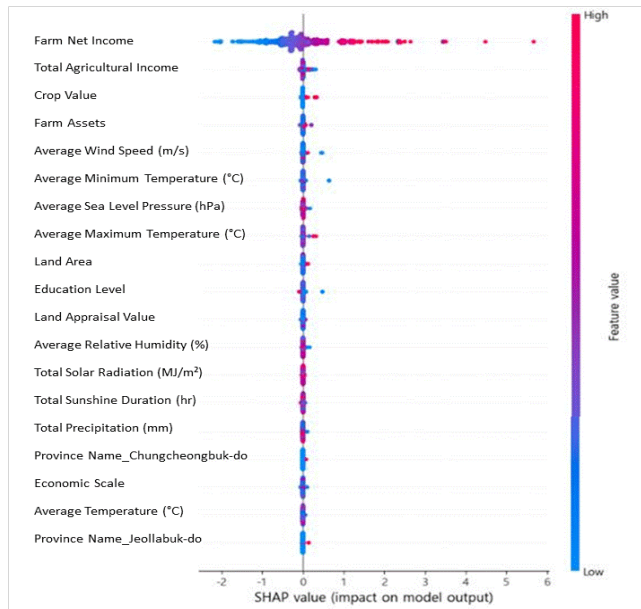


Fig. 5. SHAP analysis results

Table 6. Comparison of prediction performance: Stacking model vs. random forest with hyperparameter optimization

Model	MSE	R-squared
Stacking Model	0.50	0.58
Random Forest (with optimized hyperparameters)	0.50	0.59

max_depth, min_samples_split, min_samples_leaf)를 대상으로 GridSearchCV를 활용하여 최적 조합을 탐색하였다. 5-Fold 교차검증을 통해 단일 데이터 분할의 한계를 보완하고 모델의 일반화 성능을 확보하였다.

다음으로, Random Forest, XGBoost, LightGBM을 기본 학습기로 하고 Linear Regression을 메타 예측기로 사용하는 Stacking Regressor를 구성하였다. 모든 기본 학습기는 n_estimators=100으로 설정하였으며, 개별 모델들의 예측값을 종합하여 향상된 예측 성능을 목표로 하였다.

Table 6을 보면, 하이퍼파라미터 최적화를 통해 Random Forest 모델의 성능이 크게 향상되었다. 2년 후 예측 기준으로 최적화 이전 R²=0.49에서 최적화 후 R²=0.59로 약 20.4%의 성능 개선을 달성하였다. 최적화된 Random Forest(R²=0.59)는 Stacking 모델(R²=0.58)과 유사하거나 약간 높은 성능을 보였다. 이는 복잡한 앙상블 기법이 항상 단일 모델보다 우수한 성능을 보장하지 않음을 알 수 있다. 본 연구의 화훼 농가 소득 데이터에서는

체계적인 하이퍼파라미터 최적화가 효과적인 접근법인 것으로 나타났다.

이러한 결과는 농업 경제 분야의 머신러닝 적용에서 몇 가지 시사점을 제공한다. 첫째, 모델 복잡성과 성능 향상이 항상 비례하지 않는다는 점이다. 제한된 데이터셋에서는 단순하고 잘 조정된 모델이 복잡한 앙상블만큼 효과적일 수 있다. 둘째, 실무적 관점에서 계산 효율성을 고려할 때, 단일 모델의 체계적 최적화가 보다 실용적인 접근법일 수 있다는 점이다. 본 연구의 화훼 농가 소득 데이터에서는 체계적인 하이퍼파라미터 최적화가 효과적인 접근법을 확인할 수 있었다. 특히 Random Forest 모델의 경우, 이러한 최적화를 통해 실용적이면서도 높은 예측 성능을 보였다.

5. 요약 및 결론

본 연구는 2003년부터 2023년까지 21년간의 농가 단위

데이터를 활용하여 기후변화가 화훼 농가 소득에 미치는 영향을 실증 분석하고, 머신러닝 기법을 통해 미래 소득 변동성을 예측하였다. 서론에서 설정한 두 가지 가설에 대한 검증 결과는 다음과 같다.

첫째, 기온 상승에 대한 가설은 부분적으로 검증되었다. 평균 최저기온이 7℃ 이하로 떨어질 경우 농가 소득에 유의미한 부정적 영향을 미쳤으나, 12~14℃ 범위의 적정 온도에서는 오히려 긍정적인 효과가 관찰되었고, 이는 온도의 임계점 효과를 확인하는 결과를 가져왔다.

둘째, 강수 패턴 변화에 대한 가설은 부분적으로 지지되었다. 적정 수준의 강수량은 농가 소득에 긍정적인 영향을 미쳤으나, SHAP 분석 결과, 강수량의 변동성이 소득 예측의 불확실성을 증가시키는 것으로 나타났다.

주요 연구 결과는 다음과 같다. 화훼 농가 소득은 농가 순소득과 농업총수입 등 경제적 요인에 의해 주로 설명되었다. 기후 요인 중에서는 평균 최저기온이 가장 중요한 영향을 미쳤으며, 저온 스트레스가 농가 소득 감소의 주요 원인으로 확인되었다. 평균 최고기온과 강수량은 일정 범위 내에서 긍정적인 영향을 미쳤으나, 비선형적이며 임계값을 중심으로 반전되는 특성을 보였다.

머신러닝 기반 예측 모델 비교 결과, Random Forest와 XGBoost는 선형 회귀에 비해 일관되게 우수한 예측 성능을 보였다. 예측 시점별로 U자형 패턴이 관찰되었고, 단기(2년: $R^2=0.49$)와 장기(10년: $R^2=0.53$) 예측에서는 높은 설명력을 보인 반면, 중기(6년: $R^2=0.17$) 예측에서는 성능이 저하되었다. SHAP 분석을 통해 농가 자산 규모가 클수록 기후변화 적응력이 높다는 점이 확인되어, 농가 간 기후 리스크 대응 능력의 불균형을 실증적으로 제시하였다. 하이퍼파라미터 최적화를 통한 Random Forest 모델 ($R^2=0.59$)이 복합 앙상블 기법인 Stacking 모델 ($R^2=0.58$)과 유사한 성능을 보여, 모델 복잡성보다는 체계적 최적화가 더 효과적임을 확인하였다.

본 연구의 정책적 시사점은 기후 리스크에 대한 정교한 대응, 농가의 구조적 역량 강화, 그리고 기후 변수의 복잡한 반응을 반영한 정책 설계의 필요성으로 요약될 수 있다.

첫째, 저온 리스크에 대한 선제적 대응이 필요하다. 연구 결과, 평균 최저기온이 7℃ 이하로 떨어질 때 농가 소득에 유의미한 부정적 영향을 미치는 것으로 나타났다. 이에 따라 에너지 효율적 온실 시설 보급, 지역별 맞춤형 스마트팜 확산, 저온 저항성 품종 개발 등을 통한 기반 강화가 요구된다.

둘째, 농가 경영 역량 강화와 자산 기반 격차 완화를 위

한 정책적 노력이 시급하다. SHAP 분석을 통해 농가 자산 규모가 클수록 기후변화 적응력이 높다는 점이 확인되었으며, 이에 따라 특히 소규모 농가를 대상으로 한 기후 적응 기술 보급과 금융 지원 확대가 필수적이다.

셋째, 정책 설계 시 기후 변수의 비선형 반응과 임계점 효과를 반영한 차별화된 접근이 필요하다. 온도와 강수량이 일정 범위에서는 긍정적 영향을 미치지만, 임계값을 초과할 경우 부정적으로 전환되는 특성이 관찰되었기 때문에, 이러한 비선형성을 고려한 정밀한 정책 설계가 요구된다.

본 연구는 화훼 산업을 대상으로 기후변화의 경제적 영향을 실증 분석하고 예측 모델을 개발함으로써, 농업 부문 기후 적응 정책 설계에 실질적인 기초자료를 제공하였다. 특히 선형·비선형 모델의 비교를 통해 기후 영향의 복합성을 규명하고, 농가 자산 격차에 따른 기후 적응력의 차이를 실증적으로 제시하였다는 점에서 학술적 기여가 있다.

그러나 품목별 특성이나 재배 방식에 따른 이질성을 충분히 반영하지 못한 한계가 있으며, 이를 보완한 후속 연구가 필요하다.

첫째, 화훼 품목별 생육 특성과 기후 민감도 차이를 충분히 고려하지 못하였다. 장미, 국화, 백합 등 주요 품목마다 최적 생육 조건이 상이함에도 불구하고, 본 연구에서는 전체 화훼농가를 단일 집단으로 분석하였다. 향후 연구에서는 품목별 세분화된 분석을 통해 품목 특성에 부합하는 기후 대응 전략을 제시할 필요가 있다.

둘째, 재배 방식에 따른 이질성도 충분히 반영되지 못하였다. 노지재배와 시설재배는 기후 변화에 대한 민감도가 다르며, 특히 시설의 구조와 자동화 수준에 따라 소득 구조 역시 달라질 수 있다. 향후에는 재배 형태별 구분 분석이나 군집 분석을 통해 이러한 이질성을 체계적으로 반영할 필요가 있다.

이러한 한계를 보완한 보다 정교한 분석을 통해, 실제 정책 설계에 활용 가능한 실행력 있는 기후 적응 전략이 도출되기를 기대한다.

Reference

- Ahmad M, Afzal M, Wang H, McClean CJ. 2022. Climate change and staple crops production: A systematic review of the impact of climate change on

- wheat and rice. *Front Plant Sci* 13: 835604. doi: 10.3389/fpls.2022.835604
- Burke M, Emerick K. 2021. Adaptation to climate change: Evidence from US agriculture. *Am Econ J-Econ Polic* 13: 346-379. doi: 10.1257/pol.20200051
- Chen Q, Gao K, Xu YR, Sun YH, Pan B, Chen D, Luo C, Cheng X, Liu H, Huang C. 2023. Research advance on cold tolerance in chrysanthemum. *Front Plant Sci* 14: 1259229. doi: 10.3389/fpls.2023.1259229
- EPA (Environmental Protection Agency). 2025. Climate change impacts on agricultural productivity in the United States. Washington, DC: United States Environmental Protection Agency.
- Greene WH. 2020. *Econometric analysis*. 8th ed. Boston: Pearson.
- IPCC (Intergovernmental Panel on Climate Change). 2022. Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
- Kim DP. 2024. Analysis of climate change impact on agricultural income using semi-parametric approach (in Korean with English abstract). Seoul, Korea: Korea Rural Economic Institute. doi: 10.36464/jrd.2024.47.1.007
- Liang S. 2016. Changes in flower size and number under heat stress in rose. *J Korean Soc Hortic Sci* 34: 215-221.
- Mortensen LM, Gislérød HR. 2005. Effect of air humidity and K/Ca ratio on growth, morphology and keeping quality of roses. *Sci Hortic* 104: 407-420. doi: 10.1016/j.scienta.2004.09.008
- OECD (Organisation for Economic Co-operation and Development). 2023. Agricultural policy and climate change: An integrated approach to adaptation and mitigation. Paris: OECD Publishing. doi: 10.1787/9789264279391-en
- Rural Development Administration. 2020. Current status and challenges of the floriculture industry. *Green Mag* 175: 34-42.
- Schlenker W, Roberts MJ. 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proc Natl Acad Sci USA* 106: 15594-15598. doi: 10.1073/pnas.0906865106
- Shin JH, Kim DH, Lee SH. 2020. Utilization of eco-friendly biopolymer to reduce environmental stress in plants. *Ecol Resilient Infrastruct* 7: 34-46. doi: 10.17820/eri.2020.7.2.034
- UC Davis Horticulture Innovation Lab, UC Davis World Food Center, UC Davis Program in International and Community Nutrition. 2018. Aligning the food system to meet dietary needs: Fruits and vegetables. Davis, CA: University of California, Davis.
- Zhou Y, Zhang L, Wang H. 2025. Anthocyanin degradation drives heat-induced petal fading in *Chrysanthemum morifolium* at full bloom. *Front Plant Sci* 16: 1198. doi: 10.3389/fpls.2025.1198