

머신러닝을 활용한 친환경에너지보급사업 만족도 영향요인 분석

양성준*, 김배성**

*제주대학교 일반대학원 농업경제학과, 한국농촌경제연구원

**제주대학교 산업응용경제학과

e-mail:jejuysj96@naver.com

An Analysis of Satisfaction Determinants in an Eco-Friendly Energy Support Program Using Machine Learning

Seong-Jun Yang*, Bae-Sung Kim**

*Jeju National University, Korea Rural Economic Institute

**Dept. of Applied Economics, Jeju National University

요약

본 연구는 해양수산부의 친환경에너지보급사업 참여 양식어가를 대상으로 사업 만족도에 영향을 미치는 요인을 분석하는 것을 목적으로 한다. 기존 성과평가는 주로 단순 만족도 수준을 파악하는 데 그쳐 영향요인에 대한 체계적인 분석이 미흡하였다. 이에 본 연구는 한국농어촌공사의 성과조사 자료를 활용하여 만족도 영향 요인을 보다 정밀하게 규명하고자 하였다. 에너지절감시설을 도입한 총 200개 어가를 대상으로 인구통계학적 요인, 경영 특성, 사업효과 및 운영관리 요인을 포함한 변수를 구성하였다. 머신러닝 알고리즘인 랜덤 포레스트(Random Forest)를 적용하여 변수 중요도를 도출하고, 계량경제기법인 순서형 프로빗 모형(OPM)을 적용하여 결과를 비교·검증하였다. 분석 결과, 연 매출액, 설비 고장 관련 요인, 품질 향상 효과 등이 주요 영향요인으로 나타났으며, 특히 에너지 비용 절감보다 품질 개선 및 거래처 반응과 같은 생산성 관련 요인이 만족도에 더 큰 영향을 미치는 것으로 확인되었다. 또한 설비 고장 빈도와 사후관리 수준은 만족도에 부정적인 영향을 미치는 핵심 요인으로 나타났다. 본 연구는 머신러닝을 활용한 정책 평가의 적용 가능성을 실증적으로 제시하고, 향후 보급사업의 효율적 운영을 위한 정책적 시사점을 제공한다.

1. 서론

해양수산부 친환경에너지보급사업은 양식어가를 대상으로 에너지 고효율 설비(히트펌프, 인버터) 설치를 지원함으로써 경영비 절감과 탄소 배출 저감을 동시에 달성하고자 추진된 정책사업이다. 해당 사업은 난방비 및 전력료 절감을 통해 양식어가의 경영 안정성을 제고하고, 일정 수온 유지에 따른 생산성 향상과 매출 증대 효과를 창출하는 것으로 평가되고 있다. 그러나 현재 사업 성과평가는 주로 설문조사를 기반으로 전반적 만족도 수준을 파악하는 데 그치고 있으며, 만족도 역시 높은 수준을 유지하고 있음에도 불구하고 이에 영향을 미치는 요인을 체계적으로 분석한 연구는 부족한 실정이다.

또한 만족도 영향요인 분석과 관련한 기존 연구는 주로 회귀모형이나 구조방정식 모형을 활용하여 분석이 이루어져 다양한 요인 간 복합적인 관계를 충분히 반영하지 못하는 한계가 있다. 한편, 최근 머신러닝은 데이터 기반 학습을 통해 변수 간 복잡한 비선형 관계를 효과적으로 분석할 수 있는 방법

으로 다양한 분야에서 활용되고 있다. 머신러닝은 기존 이론 중심 접근과 달리 데이터로부터 패턴을 도출하여 예측 및 설명력을 확보할 수 있다는 점에서 정책 효과 분석에도 유용한 도구로 주목받고 있다. 이에 본 연구는 머신러닝 알고리즘인 랜덤 포레스트(Random Forest)를 활용하여 친환경에너지보급사업 참여 어가의 만족도 영향요인을 분석하고, 계량경제기법인 순서형 프로빗 모형과의 비교를 통해 분석 결과의 타당성을 검증하고자 한다. 이를 통해 사업 만족도 결정요인을 규명하고, 향후 보급사업의 효율적 운영을 위한 정책적 시사점을 제시하고자 한다.

2. 분석 자료 및 방법

2.1 분석자료

본 연구는 한국농어촌공사의 친환경에너지보급사업 성과조사 자료를 활용하였다. 해당 조사는 2025년 9월부터 11월까지 전화조사 방식으로 수행되었으며, 최근 5년(2020~2024) 동안 사업에 참여한 양식어가를 대상으로 이루어졌다. 총

588개 어가 중 256개 어가가 응답하였으며, 결측치 및 응답 오류를 제외한 200개 유효표본을 최종 분석에 활용하였다. 종속변수는 전반적 만족도로, 1~5점 리커트 척도를 기반으로 측정된 값을 낮음(Low, 1~3), 보통(Mid, 4), 높음(High, 5)의 3개 범주로 재구성하여 순서형 변수로 설정하였다. 독립변수는 총 26개로 구성되며, 인구통계학적 요인(연령, 성별, 지역), 경영 특성 요인(연 매출액, 경영비, 양식 규모, 종사 기간), 사업효과 요인(에너지 비용 절감, 품질 향상, 거래처 반응 등), 운영 및 행정 요인(설비 고장 빈도, 사후조치 만족도, 운영관리 수월성 등)으로 구분하였다.

2.2 분석방법

친환경에너지보급사업 참여 어가의 전반적 만족도를 예측하고, 만족도에 영향을 미치는 주요 요인을 도출하기 위해 랜덤 포레스트(Random Forest) 알고리즘을 적용하였다. 랜덤 포레스트는 의사결정나무의 확장된 형태로, 여러 개의 의사결정나무를 생성한 뒤 예측 결과를 종합하는 앙상블 기법이다. 선형성, 정규성, 등분산성 등 전통적 모수모형에서 요구되는 통계적 가정이 상대적으로 적고, 변수 간 비선형 관계와 복잡한 패턴을 반영할 수 있다는 장점이 있다. 본 논문에서는 이를 바탕으로 만족도 범주를 예측하는 분류모형으로 랜덤 포레스트를 활용하였다.

모형 구축은 데이터 전처리, 학습자료와 검증자료 분할, 모형 학습, 예측 성능 평가의 순서로 수행하였다. 종속변수인 전반적 만족도는 보통 이하, 만족, 매우 만족의 3개 범주로 재구성되었으므로 본 연구의 분석은 회귀문제가 아니라 분류문제에 해당한다. 이에 따라 모형 성능 평가는 혼동행렬을 기반으로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 score 등을 활용하였다. 랜덤 포레스트 모형의 하이퍼파라미터는 R 4.3.0의 randomForest 패키지를 이용하여 조정하였다. 분류문제에서 일반적으로 사용되는 mtry와 nodesize는 기본값을 적용하였고, 의사결정나무의 개수인 ntree와 최대 노드 수를 의미하는 maxnodes를 중심으로 총 6개 시나리오를 구성하였다. 구체적으로 ntree는 500개, 1,000개, 2,000개 수준으로, maxnodes는 10개, 20개 수준으로 조합하여 시나리오 구성을 통해 모형 성능을 비교하였다. 최적 모형은 OOB(Out-of-Bag) 오차와 정확도, 정밀도, 재현율, F1 score를 종합적으로 고려하여 선정하였다.

분석 결과, OOB 오차만을 기준으로 보면 ntree=1,000, maxnodes=20인 시나리오의 오차가 가장 낮았으나, 정확도와 F1 score 등 예측 성능을 함께 고려할 경우 ntree=500, maxnodes=10인 시나리오가 가장 적합한 것으로 판단되었다. 따라서 본 연구에서는 해당 모형을 최종 모형으로 활용하

였다. 한편, 단일 학습자료와 검증자료 분할에 따른 결과의 불안정성을 보완하기 위해 5-fold 교차검증을 추가로 수행하였다. 교차검증 결과는 단일 분할 결과보다 다소 낮게 나타났으나, 이는 보다 엄격한 평가 기준을 적용한 결과로 해석할 수 있으며, 반복된 표본 분할에서도 주요 성능 지표가 비교적 안정적으로 유지되었다. 이를 통해 최종 모형의 일반화 가능성을 검토하였다.

또한 랜덤 포레스트 모형은 예측 성능 면에서는 우수하나, 내부 의사결정 과정을 직관적으로 파악하기 어렵다는 한계가 있다. 이에 본 연구는 설명가능한 AI(XAI)의 관점에서 변수 중요도 지수를 활용하였다. 먼저 MDA(Mean Decrease Accuracy) 기준 변수 중요도를 산출하여 전반적 만족도 예측에 기여하는 상위 변수를 도출하였다. 한편, 머신러닝 분석 결과의 적합성을 검토하기 위해 계량경제기법인 순서형 프로빗 모형(Ordered Probit Model, OPM)을 함께 추정하였다. 전반적 만족도는 보통 이하, 만족, 매우 만족의 순서를 갖는 범주형 변수이므로 일반 선형회귀모형보다 순서형 프로빗 모형이 적합하다. 순서형 프로빗 모형은 랜덤 포레스트와 동일한 변수 구성을 적용하였으며, 이를 통해 랜덤 포레스트에서 도출된 주요 영향요인이 전통적 계량모형에서도 일관되게 나타나는지를 비교하였다.

3. 분석 결과

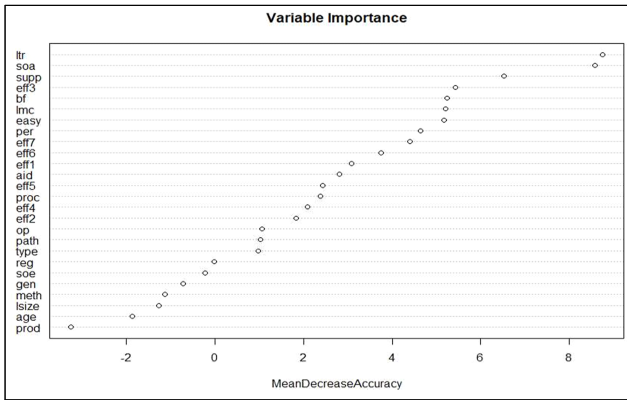
3.1 변수 중요도(MDA) 분석 결과

랜덤 포레스트 모형을 활용한 변수 중요도 분석 결과, 전반적 만족도에 영향을 미치는 주요 변수는 연 매출액, 고장 발생 후 조치 만족도, 정부 지원 필요성, 품질 향상 효과, 설비 고장 빈도, 경영비, 운영관리 수월성, 종사 기간 등으로 나타났다. 변수 중요도 상위에 속한 연 매출액은 본 사업의 효과가 어가의 경영 규모에 따라 차별적으로 체감될 수 있음을 시사한다. 즉, 일정 수준 이상의 생산 규모와 자본을 갖춘 어가일수록 에너지 절감 및 생산성 향상 효과를 보다 효율적으로 활용할 수 있으며, 이는 전반적 만족도로 이어지는 것으로 해석된다. 또한 설비 고장 빈도와 고장 발생 후 조치 만족도는 모두 높은 중요도를 보였으며, 이는 사업효과가 단순히 설비 도입 단계에서 결정되는 것이 아니라 운영 및 유지관리 과정 전반에 걸쳐 형성된다는 점을 의미한다. 특히 사후조치 만족도가 중요한 변수로 나타난 것은 A/S 체계의 질이 사업 만족도를 좌우하는 핵심 요인임을 보여준다. 한편 품질 향상 효과 역시 주요 변수로 나타났는데, 이는 사업의 성과가 단순한 비용 절감에 국한되지 않고 생산물의 품질 개선 및 부가가치 향상여부와도 관련이 있음을 시사한다. 이와 함께 운영관리 수

월성 및 종사 기간과 같은 변수도 일정 수준 이상의 중요도를 보이며, 경영 경험과 운영 효율성 역시 만족도 형성에 영향을 미치는 요인으로 확인되었다.

[표 1] MDA 기준 중요도 상위 10개 변수

순위	변수	중요도 지수 (MDA)
1	연 매출액	8.753
2	고장 발생 후 조치 만족도	8.579
3	정부 지속적 지원 필요	6.530
4	도입효과 3: 품질 향상 정도	5.427
5	설비 고장 빈도	5.240
6	연 경영비	5.206
7	설비 운영관리 수월 정도	5.172
8	현업 종사 기간	4.646
9	도입효과 7: 거래처 반응 향상 정도	4.403
10	도입효과 6: 어병 발생률 감소 정도	3.754



[그림 1] MDA 기준 중요도 상위 10개 변수

3.2 순서형 프로빗 모형 분석 결과

순서형 프로빗 모형 분석 결과, 품질 향상, 거래처 반응, 에너지 비용 절감, 설비 고장 빈도 등이 통계적으로 유의한 변수로 나타났다. 먼저 품질 향상과 거래처 반응 변수는 모두 양(+)의 영향을 미치는 것으로 나타났으며, 이는 사업 참여를 통해 생산물의 품질이 개선되거나 거래처의 반응이 긍정적으로 변화할 경우 전반적 만족도가 증가하는 경향이 있음을 의미한다. 특히 거래처 반응 변수의 유의성은 시장성과 직접적으로 연결된 성과가 어가의 체감 만족도에 중요한 영향을 미친다는 점을 보여준다.

반면 설비 고장 빈도는 통계적으로 유의한 음(-)의 영향을 미치는 것으로 나타났다. 이는 설비 운영 과정에서 발생하는 고장 및 유지관리 부담이 어가의 만족도를 저해하는 주요 요인으로 작용함을 의미하며, 설비 도입 이후의 관리 체계가 정책 성과에 중요한 역할을 한다는 점을 시사한다.

에너지 비용 절감 효과 또한 유의한 변수로 나타났으나, 그 영향력은 품질 향상이나 거래처 반응과 같은 생산성 관련 변수에 비해 상대적으로 제한적인 것으로 해석된다. 이는 최근

농어업용 전기료(농사용(을))가 급등함에 따라 어가가 에너지 비용 절감 효과를 크게 체감하지 못한 것으로 사료된다.

[표 2] OPM 분석 결과

구분	계수	표준오차	p-value
양식품목(참조집단: 기타)			
넙치	-0.3828	0.463	0.409
전복	-0.2561	0.494	0.604
민물장어	-0.0643	0.601	0.915
새우	-0.8406*	0.466	0.071
치어/종묘	-0.2573	0.458	0.575
현업 종사 기간	0.0296***	0.010	0.005
사업효과 1 (에너지비용 절감)	0.2284*	0.133	0.086
사업효과 3 (품질 향상)	0.3390**	0.157	0.032
사업효과 7 (거래처 반응)	0.4060**	0.158	0.010
고장 빈도	-0.2575***	0.094	0.006
사업 절차 만족	0.2847**	0.144	0.049
사업 인지경로(참조집단: 홍보물)			
지인 소개	0.8498***	0.299	0.004
공무원	1.2524***	0.293	0.000
공사 직원	0.8870***	0.320	0.006
Log pseudolikelihood = -156.42123			
Number of obs = 200			
Wald chi(36) = 103.78			
Prob > chi2 = 0.0000			
Pseudo R ² = 0.2442			

주: 지면 제약으로 인하여 통계적으로 유의한 변수들 위주로 추정 결과를 정리함.

3.3 종합 비교

랜덤 포레스트와 순서형 프로빗 모형을 비교한 결과, 주요 영향요인은 전반적으로 유사한 방향성을 보였으나, 두 모형 간에는 분석 결과의 범위와 해석 방식에서 차이가 확인되었다. 순서형 프로빗 모형은 통계적 유의성을 기준으로 주요 변수를 도출함에 따라 비교적 제한된 변수만이 유의하게 나타난 반면, 랜덤 포레스트 모형은 변수 간 비선형 관계와 상호작용을 고려하여 보다 다양한 영향요인을 도출하였다. 특히 랜덤포레스트에서는 경영 규모, 운영관리, 사후조치와 같은 변수들이 주요 요인으로 나타나, 정책 효과가 복합적인 요인에 의해 결정된다는 점을 보다 폭넓게 반영하였다.

이러한 결과는 머신러닝 기반 분석이 전통적 계량모형을 보완하여 정책 효과를 보다 현실적으로 설명할 수 있는 유용한 도구임을 시사하며, 향후 정책 평가 연구에서 두 접근법을 병행 활용할 필요가 있음을 보여준다.

4. 요약 및 결론

본 연구는 친환경에너지보급사업 참여 양식어가를 대상으로 사업 전반적 만족도에 영향을 미치는 요인을 보다 정밀하게 분석하기 위해 머신러닝 기법인 랜덤 포레스트 모형을 활용하였다. 기존의 성과평가 연구는 주로 회귀모형과 같은 계량경제기법에 의존하여 변수 간 선형 관계를 중심으로 분석이 이루어졌으나, 본 연구는 변수 간 비선형 관계와 상호작용을 반영할 수 있는 머신러닝 기법을 적용함으로써 보다 현실적인 영향요인 분석을 시도하였다. 또한 전통적 계량모형인 순서형 프로빗 모형과의 비교를 통해 머신러닝 기반 분석 결과의 타당성과 해석 가능성을 검토하였다.

분석 결과, 연 매출액과 같은 경영 규모 변수, 설비 고장 빈도 및 사후조치 만족도와 같은 운영관리 요인, 그리고 품질 향상 및 거래처 반응과 같은 생산성 관련 요인이 주요 영향요인으로 확인되었다. 특히 단순한 에너지 비용 절감 효과보다 품질 개선 및 시장 반응과 같은 생산성 중심 요인이 만족도에 더 큰 영향을 미치는 것으로 나타났다. 또한 설비 고장 및 유지관리 과정에서의 문제는 만족도를 저해하는 핵심 요인으로 확인되었다.

랜덤 포레스트와 순서형 프로빗 모형을 비교한 결과, 두 모형은 주요 영향요인에서 전반적으로 유사한 결과를 보였으나, 랜덤포레스트는 경영 규모, 운영관리, 사후조치와 같은 복합적 요인을 보다 폭넓게 반영하는 것으로 나타났다. 이는 머신러닝 기반 분석이 전통적 계량모형을 보완하여 정책 효과를 보다 입체적으로 설명할 수 있는 유용한 접근법임을 시사한다.

이러한 분석 결과를 바탕으로 다음과 같은 정책적 시사점을 도출할 수 있다. 첫째, 설비 도입 이후의 운영 단계에서 발생하는 고장 및 유지관리 부담이 만족도에 큰 영향을 미치는 만큼, A/S 대응 체계의 신속성과 전문성을 강화하고 사후관리 지원을 제도적으로 보완할 필요가 있다. 둘째, 정책의 성과를 단순한 에너지 비용 절감 중심으로 홍보하기보다 품질 향상, 생산성 개선, 거래처 반응 개선 등의 효과를 강조하는 방향으로 정책 설계 및 홍보 전략을 개선할 필요가 있다. 셋째, 경영 규모가 작은 어가의 경우 사업효과를 충분히 체감하지 못할 가능성이 있으므로, 소규모 어가를 대상으로 한 맞춤형 지원 및 운영관리 교육을 강화하여 사업효과의 체감도를 제고할 필요가 있다. 넷째, 설비의 안정적 운영을 위해 표준화된 유지관리 매뉴얼을 보급하고, 현장 중심의 기술지원 체계를 구축함으로써 설비 활용 효율성을 높일 필요가 있다.

본 연구는 머신러닝 기법을 활용하여 정책사업 만족도 결정요인을 실증적으로 분석하고, 계량경제모형과의 비교를 통해 분석 결과의 신뢰성을 검토하였다는 점에서 의의를 가진다.

다만, 단면자료를 활용한 분석이라는 한계가 존재하며, 향후 패널데이터 구축 및 다양한 분석기법을 적용한 추가 연구가 필요하다.

참고문헌

- [1] 김지영 “머신러닝 기법을 활용한 청소년의 온·오프라인 학교생활 만족도 영향요인 탐색”, 교육혁신연구, 제 32권 3호, pp. 21-48, 2022년.
- [2] 박소영·정혜원 “머신러닝 기법을 적용한 기업 구성원의 직무만족 및 조직몰입에 영향을 미치는 요인 탐색”, 교육학연구, 제 57권 4호, pp.183-213, 2019년.
- [3] 이나라·조영일 “랜덤 포레스트를 활용한 다문화청소년의 삶의 만족도에 미치는 영향요인 탐색”, 사회과학연구, 제 49권 1호, pp.95-118, 2023년.
- [4] 이유나·주하연 “서열프로빗 모형을 적용한 항공사 선택 속성이 만족도와 행동의도에 미치는 영향”, 호텔관광연구, 제 26권 4호, pp.253-268, 2024년.
- [5] 한은진·이승길 “서열프로빗모형을 적용한 한식만족도 영향요인에 관한 연구”, 관광연구저널, 제 26권 6호, pp.175-190, 2012년.
- [6] 김길재·김병주 “머신러닝을 활용한 유치원 정보공시 만족도 영향 요인 분석”, 교육재정경제연구, 제 32권 3호, pp.187-214, 2023년.
- [7] 김민선 “랜덤포레스트 머신러닝 기법을 활용한 장애인의 삶의 만족도 영향요인 탐색”, 사례관리연구, 제 15권 2호, pp.103-123, 2024년.
- [8] 임다혜·권영상 “랜덤 포레스트 모형을 활용한 청년들의 주택 유형별 주거환경 만족도 영향요인 중요도 분석”, 한국도시설계학회지, 제 23권 6호, pp.103-122, 2022년.