

데이터에서 정책으로: UN SDG 3.1 달성을 위한 기계학습과 설명 가능한 인공지능 기반 산모 사망률 연구

문영신¹, 유시연¹, 김나연¹, 문지훈²

¹ 덕성여자대학교 데이터사이언스학과 학부생

² 덕성여자대학교 데이터사이언스학과 교수

{munniie,dbtdbtdus,starship,jmoon25}@duksung.ac.kr

From Data to Policy: Machine Learning and XAI Approaches to Maternal Mortality in the Context of UN SDG 3.1

Yeongsin Mun¹, Siyeon Yu¹, Nayeon Kim¹, and Jihoon Moon²

^{1,2} Dept. of Data Science, Duksung Women's University

요 약

UN 지속가능발전목표(SDG 3.1)는 2030년까지 전 세계 산모 사망률(MMR)을 출산 10만 명당 70명 미만으로 낮추는 것을 목표로 한다. 본 연구는 2011~2016년 국가 단위 자료를 활용하여 여성 교육, 보건 지출, 영양 상태의 상호작용을 분석하였다. 랜덤 포레스트와 SHAP 결과, 여성 중등교육 등록률이 핵심 요인으로 나타났으며, 특히 등록률 75% 이상에서는 보건 지출 수준과 관계없이 MMR이 30~75 수준으로 안정화되었다. 반대로 40~50% 이하에서는 MMR이 높게 유지되며 보건 지출의 영향력이 크게 작용하였다. 이는 여성 교육과 보건 지출을 통합적으로 확대하는 전략이 SDG 3.1 달성에 이바지할 수 있음을 보여준다.

1. 서론

산모 사망률(Maternal Mortality Ratio, MMR)은 국가 보건 시스템의 효과성과 여성 건강권을 평가하는 핵심 지표로, UN 지속가능발전목표(SDG 3.1)는 2030년까지 전 세계 산모 사망률을 출산 10만 명당 70명 미만으로 줄이는 것을 목표로 하고 있다. 하지만 2011~2016년 개발도상국의 산모 사망률은 선진국 대비 약 13배 높게 나타났으며, 이는 단순한 의료 인프라의 차이를 넘어 교육, 영양 보건 재정수준 등 복합 요인의 상호작용에서 비롯된 것으로 분석된다.

기존 연구들은 주로 개별 요인의 단일 효과에 초점을 맞추어 왔다. 따라서 요인 간 상호작용과 비선형적 관계에 대한 체계적 분석이 부족하였으며, 효과적인 정책 조합을 제시하는 실증적 근거 또한 미흡하였다[1]. 이에 본 연구는 기계학습과 설명 가능한 인공지능을 활용하여 여성 교육, 영양 상태, 보건 지출 간 복합적 관계를 분석할 뿐만 아니라 산모 사망률 감소를 위한 정책적 시사점을 제시함으로써 UN SDG 3.1의 달성을 향한 논의와 실천을 뒷받침하고자 한다.

2. 연구 방법론

본 연구는 Kaggle 데이터 셋인 *Malnutrition-Underweight women, children and others*을 기반으로 모델링을 수행하였다[2]. 최종적으로 확보된 변수는 *Country* (국가명), *Country_Code* (국가 코드), *Year* (연도), *Maternal_Mortality_Ratio* (산모 사망률, 출생 10만 명당), *Health_Expenditure_GDP* (GDP 대비 정부 보건 지출 비율), *Female_Underweight_Prevalence* (성인 여성 저체중 비율 및 연령 표준화), *Female_Secondary_Enrollment*

(여성 중등교육 등록률)이다. 이 중 *Maternal_Mortality_Ratio*는 종속변수, *Health_Expenditure_GDP*, *Female_Underweight_Prevalence*, *Female_Secondary_Enrollment*는 입력변수로 설정하였다. 원본 1,398개 중 결측치가 포함된 행을 제거하여 최종 750개를 분석에 활용하였다.

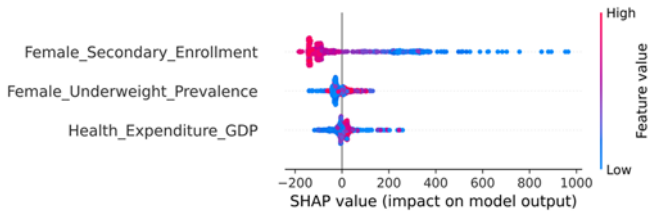
본 연구는 산모 사망률 예측에서 비선형성과 변수 간 상호작용을 효과적으로 포착하고 초매개변수 설정에 크게 의존하지 않고도 안정적 성능을 보장하는 랜덤 포레스트를 최종 모델로 선정하였으며 기존 문헌의 가이드라인에 따라 128개의 나무로 구성되어 학습을 진행하였다[3]. 다음으로 전체 데이터 셋을 대상으로 랜덤 포레스트 모델의 SHAP 분석을 통해 결과를 해석하였다. 분석 내용은 (1) 변수별 전역적 기여도, (2) 개별 관측치 수준에서의 영향, (3) 변수 간 상호작용 효과로 구분하였다. 또한, 보건 지출-여성 교육, 보건 지출-영양 상태 등 정책적으로 의미 있는 변수 조합에 대해 정책 표면을 도출하여, 최적 정책 조합의 시각화를 시도하였다.

3. 실험 및 결과

<표 1> 10겹 교차검증을 통한 모델별 성능 비교(± 표준편차)

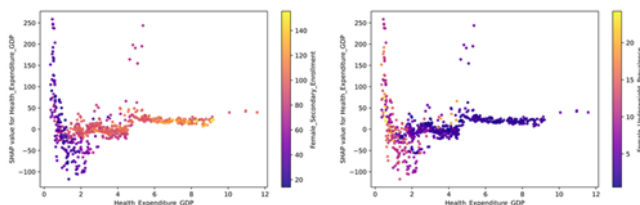
Models	MAE (MMR, 10만 출생아당)	RMSE (MMR, 10만 출생아당)	R ²
Linear Regression	96.95 (± 12.87)	148.78 (± 28.08)	0.596 (± 0.085)
Decision Tree	49.64 (± 12.94)	120.90 (± 31.82)	0.704 (± 0.177)
Random Forest	42.42 (± 8.28)	88.66 (± 23.06)	0.848 (± 0.081)

본 논문은 랜덤 포레스트 모델의 신뢰성을 검증하고자 다중선형 회귀, 의사결정 나무와 같이 모델 해석성이 우수한 기계학습 모델들을 10점 교차검증을 통해 성능을 비교하였으며, 표 1과 같이 랜덤 포레스트의 우수성을 확인할 수 있었다.



(그림 1) 랜덤 포레스트 모델의 SHAP Summary Plot

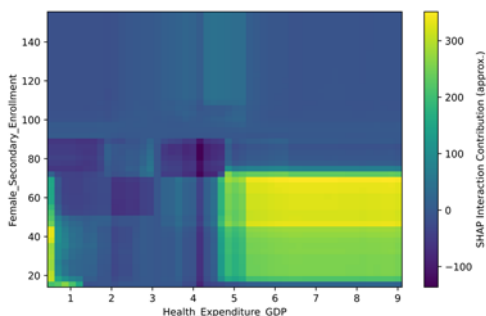
그림 1은 전체 데이터 셋을 학습한 랜덤 포레스트 모델의 SHAP summary plot 결과를 나타낸다. 여성 중등교육 등록률은 음의 효과를 보임으로써 교육 수준 향상이 산모 사망률 감소로 이어짐을 확인할 수 있었으며, 여성 저체중률은 양의 효과가 두드러져 영양 악화가 위험 증가에 이바지함을 알 수 있었다. 보건 지출 비중의 영향은 상대적으로 적지만, 낮은 수준에서 위험 증가와 연관됨이 드러났다.



(a) 의료 지출 × 여성 교육 (b) 의료 지출 × 여성 저체중
(그림 2) 의료 지출과 여성 요인 간 상호작용에 대한 SHAP 의존성 분석

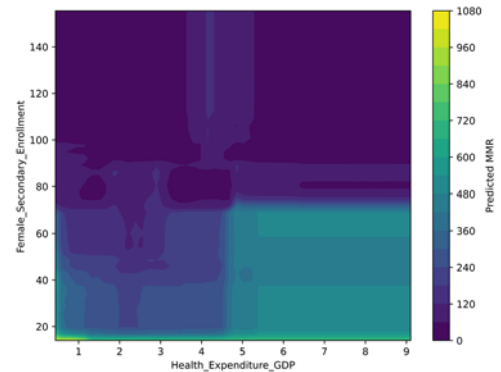
변수 간 상호작용 분석 결과(그림 2), 보건 지출과 여성 교육 간 상호작용 강도가 가장 높게 나타났다. 그림 2(a)의 의존성 플롯에서 확인할 수 있듯이, 교육 수준이 낮은 국가에서는 보건 지출이 증가할수록 산모 사망률 감소에 큰 기여를 보였으나, 교육 수준이 높은 국가에서는 보건 지출의 추가 효과가 점차 완화되는 경향을 확인할 수 있었다. 보건 지출이 GDP의 약 2~4% 수준에 도달하면, 그 이후로는 감소 효과가 상대적으로 안정화되는 양상이 나타났다.

그림 2(b)는 보건 지출과 여성 저체중 유병률 간의 상호작용을 보여준다. 저체중 유병률이 낮은 국가에서는 보건 지출 증가가 산모 사망률 감소에 뚜렷하게 이바지하는 반면, 저체중 유병률이 높은 국가에서는 동일한 수준의 지출 확대가 충분한 개선 효과로 이어지지 못했다. 이는 단순한 재정 투입만으로는 한계가 있으며, 영양 상태 개선과 의료 체계 강화가 병행되어야 함을 시사한다.



(그림 3) SHAP 상호작용: 보건 지출 × 여성 교육

그림 3은 보건 지출(Health_Expenditure_GDP)과 여성 중등교육 등록률(Female_Secondary_Enrollment) 간의 비선형적 상호작용을 보여준다. 전반적으로 여성 교육 수준이 높을수록 보건 지출의 효과가 완화되는 경향을 확인하였으며, 반대로 교육 수준이 낮을수록 보건 지출의 기여도가 상대적으로 커지는 패턴을 확인하였다. 이는 두 요인 간의 상호작용이 산모 사망률 변동을 설명하는 데 핵심적임을 시사한다.



(그림 4) 정책 표면 분석: 보건 지출 × 여성 교육

이를 바탕으로 그림 4의 정책 표면을 분석한 결과, 여성 저체중 유병률을 중간값(3.30%)으로 고정하였을 때, 여성 중등교육 등록률이 약 75% 이상으로 높아질 경우 보건 지출 수준과 관계없이 산모 사망률이 30~75 수준으로 안정화되는 양상이 관찰되었다. 반면 여성 중등교육 등록률이 40~50% 이하이면 산모 사망률이 평균 이상으로 높게 나타났으며, 이 구간에서는 보건 지출의 변화가 산모 사망률에 미치는 영향력이 두드러지게 나타났다.

4. 결론

본 연구는 여성 교육이 산모 사망률 감소의 가장 강력한 요인임을 실증적으로 확인하고, 교육과 보건 지출 간 시너지 효과를 제시하였다. 이는 단일 부문보다 통합적 접근이 효과적임을 시사하며, 개발도상국은 교육 확대를 우선하고 중간 소득국은 교육 및 보건 투자의 균형을 추구하는 전략의 필요성을 강조한다. 그럼에도 불구하고 국가 단위 지표만 활용한 한계가 있으므로, 향후에는 지역별 세분화, 시계열 분석, 정책 실험 검증 등을 수행하고자 한다. 더불어, 결측치로 인한 표본 손실을 보완하기 위해 다중대치 기법을 적용하여 분석의 신뢰성과 일반화를 높일 예정이다.

참고문헌

- [1] J. P. Souza, L. T. Day, A. C. Rezende-Gomes, J. Zhang, R. Mori, A. Baguiya, et al., "A global analysis of the determinants of maternal health and transitions in maternal mortality," *The Lancet Global Health*, Vol. 12, No. 2, pp. e306–e316, 2024.
- [2] S. Bose, "Malnutrition: Underweight Women, Children & Others" [Dataset], Kaggle, 2025.
- [3] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Berlin, Heidelberg: Springer, Jul. 2012, pp. 154–168.