발표 1 토론

「머신러닝을 활용한 후속출산의도를 예측하기 위한 모델 구축 및 네트워크 분석」 토론문

김나영(육아정책연구소 연구위원)

본 연구는 「한국 영유아 교육·보육 패널」 자료를 활용하여 1) 랜덤포레스트, 그래디언부스팅, 로지스틱 회귀 모델, 스태킹 모델을 통해 후속출산의도를 예측하는 요인을 살펴보고, 2) 네트워크 분석을 통해 요인 간의 상호작용 및 예측요인의 중요도를 파악하고자 하였다. 기존 관련 연구들과는 또 다른 연구방법을 통하여 후속출산을 지원하기 위한 실증적인 자료를 제공하였다는 점에서 큰 의의가 있는 연구이다.

후속출산 관련 통계자료를 살펴보면, 올해 초인 2월 28일에 통계청이 발표한 「2023년 인구동향조사 출생·사망통계(잠정)」에 따르면, 전년대비 첫째아는 6천 7백 명(-4.6%), 둘째아는 9천 5백 명(-11.4%) 감소한 것으로 나타났다. 출생아 중 첫째아 비중은 60.1%, 둘째아 32.3%로, 전년대비 첫째아 비중은 1.9%p 증가한 것으로 나타났다. 이와 같이 우리나라의 출산은 후속출산으로 이어지지 못하는 상황인 것으로 나타나 이를 개선하기 위한 정책적 지원이 보다 적극적으로 요구되는 시점이다.

본 연구 도입에서도 지적하였듯이 후속출산에 영향을 미치는 요인은 매우 다양하여 그 영향이 복합적이다. 무엇보다도 후속 출산은 **개인의 출산에 대한 의향**이 절대적으로 중요하며, 이러한 의향에 영향을 미치는 요인이 개인적 요인, 자녀 양육과 관련된 요인, 경제적 요인, 양육 환경요인, 정책지원과 관련된 요인 등 매우 다양한 요인들이 서로 연관되어 영향을 미치게 된다.

이러한 복잡한 관계를 본 연구에서는 적절한 분석 틀을 통해 전반적으로 의미 있는 결과를 보여주었고, 그 중에서 특히 몇 가지 흥미로운 내용들을 살펴보면, 우선 분석 집단을 20대, 30대, 40대로 구분하여 분석함으로써 20대 집단의 경우 다른 연령 집단 보다 후속출산의도 비중이 높다는 것을 보여 주었다. 이는 후속출산의도에 있어 연령별 차이가 분명하게 드러난다는 점을 확인해 주었고, 나아가 후속출산의도를 예측하는 중요도는 집단에 따라 상이한 결과를 나타낸다는 것을 보여주었다. 20대 집단의 경우 다른 집단에서는 유의미하지 않았던 **아버지의 양육참여**, 30대 집단에서는 다른 집단에서 유의미하지 않았던 **어머니 연령**, 40대 집단에서는 다른 집단에서는 유의미하지 않았던

학력과 주관적 사회 경제적 지위가 후속출산의도에 영향을 미치는 중요한 요인으로 나타났다. 이는 다른 연구들에서도 논의된, 생애주기에 따라 후속출산의도에 영향을 미치는 요인이 다르며 관련 요인의 영향력이 변화할 수 있다는 것을 지지하는 연구결과를 다시 한 번 제공하면서 추가적으로 어떤 요인들이 주요한가를 파악할 수 있게 하였다.

최근과 같은 초저출생 시기에는 본 연구에서와 같이, 집단별 구분을 통한 분석 결과를 바탕으로 집단 간 공통된 이슈는 무엇이고, 분명하게 차이를 보이는 이슈는 무엇인가를 시의적절하게 파악하여 보다 세심한 정책을 수립하여 시행해야 할 것이다. 또한 본 연구의 연구자께서도 말미에 언급한 '정서적 지원', '문화조성' 등과 관련하여 초저출생이 심화되어 갈수록 유형의 지원과 더불어 무형의, 가치 및 정서와 관련된 부분을 개선해 나갈 수 있는 정책마련에 더 무게를 두는 것이 무엇보다 중요하다.