

머신러닝과 네트워크 분석을 활용한 연령집단별 후속출산의도 예측 모형 비교*

엄연용¹⁾

요약

본 연구의 목적은 후속출산의도를 예측하는 요인을 탐색하고 네트워크 분석을 통해 요인 간의 상호작용 및 예측요인의 중요도를 파악하는 것이다. 이를 위해 2022년 한국 영유아 교육·보육 패널 자료 중 어머니 1,922명의 자료를 분석하였다. 개인요인, 양육요인, 가구요인, 환경요인의 20개의 예측요인과 후속출산의도 간의 관계를 설정하고 랜덤포레스트, 그래디언트 부스트, 로지스틱, 스택킹 모델을 적용하였다. 연구 결과, 네 가지 모델의 성능은 F1 .60~.87으로 나타났다. 20대 집단에서 후속출산의도를 예측하는 주요 요인은 양육스트레스, 성격(그릿), 우울로 나타났고 30대 집단에서는 어머니연령, 부부갈등, 양육스트레스, 40대 집단에서는 사회적지지, 총자녀수, 학력으로 나타났다. 본 연구를 통해 후속출산의도를 예측할 수 있는 주요 요인을 탐색하였으며 후속출산을 지원하기 위한 실증적인 자료와 토대를 마련하였다는 점에서 의의가 있다.

주제어: 후속출산의도, 랜덤포레스트, 그래디언트부스팅, 스택킹, 네트워크분석

I. 서론

우리나라의 저출산 현상은 매우 심각한 사회 문제로 제기되고 있다. 인구동향조사에 따르면 2024년 출생아 수는 전년 대비 7.7% 감소하였으며 합계출산율을 살펴보면 2015년 1.24명, 2016년 1.17명, 2017년 1.05명, 2018년 0.98명, 2019년 0.92명, 2020년 0.84명, 2021년 0.81명, 2022년 0.78명, 2023년 0.72명으로 나타나 그 심각성을 확인할 수 있다(임영일·강현영·서주희 외, 2024). 우리나라는 매년 최저의 출산율을 기록하고 있으며

* 본 논문은 제2회 한국 영유아 교육·보육 패널 학술대회(2024. 8. 30)에서 발표한 원고를 수정·보완한 것임
1) 건양대학교 심리상담치료학과 겸임교수

저출산 문제는 인구 유지뿐만 아니라 인구 부족으로 인한 노동력 감소, 세수 감소 및 재정 부담의 증가, 지역사회 붕괴, 도시의 소멸, 교육 기관 및 필수 기관의 축소 등 다양한 사회적, 경제적, 문화적 문제를 초래할 수 있다.

후속출산이란 첫째 자녀를 출산한 후 둘째 이상의 자녀를 출산하는 것으로 개인의 선택으로 이어지는 출산을 말한다(이정원, 2009). 후속출산은 자녀를 양육했던 경험과 양육 환경, 경제적 여건, 양육 정책, 개인의 가치에 따라 선택의 폭이 더 크다고 볼 수 있다(김미나·이소은, 2018). 생애주기 모델에 따르면 첫째 자녀를 출산한 후 둘째 또는 그 이상의 후속 출산의 터울이 저출산 현상을 예측하는 데 중요하다고 보았다(Morgan & Taylor, 2006). Kohler(2001)는 저출산의 원인 중 하나로 첫째 자녀 출산 후 둘째 이상의 후속출산을 하지 않는 사회 현상을 주목하였다. 우리나라 경우도 저출산 특징 중 하나로 후속 출산이 이어지지 않는다는 점이 있다. 이러한 현상은 우리나라의 종단적 출생아 수에서 확인할 수 있다. 2021년 전국 출산력 및 가족보건·복지실태조사에서 기혼여성의 출생아 수 분포의 종단적 특성을 살펴보면 1명의 범주는 지속적으로 증가하였으나(2015년 24.4%, 2018년 26.4%, 2021년 30.0%) 2명~3명 이상은(2015년 55.9%, 2018년 53.3%, 2021년 46.0%) 감소하는 것으로 나타났다(박종서·임지영·김은정 외, 2021). 또한 인구 동향 조사에서도 첫째 출산(-4.6%)보다 둘째 출산(-11.4%)의 감소가 더 큰 것으로 나타났다(임영일·강현영·서주희 외, 2024). 이는 한국만의 문제는 아니며 저출산 국가에서 반복적으로 나타나고 있어 저출산 문제를 기혼 여성의 후속출산에 초점을 맞추어 다각적인 방면에서 살펴볼 필요가 있다(Gietel-Basten, Sobotka, Zeman et al., 2014).

한편, 후속출산의도는 어머니의 연령에 따라 다른 양상을 보임에도 불구하고 연령에 따른 집단을 구분하지 않고 가구소득, 자녀의 수, 어머니의 취업유무에 따른 집단 구분을 중점적으로 연구되어왔다. 선행연구에 따르면 소득구간과 취업유무에 따라 집단을 구분하고 후속출산계획에 영향을 미치는 요인을 탐색한 결과 모든 집단에서 어머니의 연령이 유의한 것으로 나타났다(강유선·김찬우, 2018). 출산한 자녀 수에 따라 출산의지를 살펴본 연구에서도 모든 집단에서 연령이 적을수록 출산의지가 강하였고 부모의 연령이 낮을수록 후속출산 계획과 자녀를 낳겠다는 반응도 높았다(김은경·박신애, 2019; 서미정, 2011; 서정연·김한곤, 2015). 또한 후속출산에 영향을 미치는 요인은 다양하고 복잡적이다. 후속 출산은 개인의 출산 의도가 중요하며 출산 의도에 영향을 미치는 개인적 요인, 자녀 양육과 관련된 요인, 경제적 요인, 환경요인, 가구특성 등 다양한 요인들이 서로 연관되어 영향을 미치는 것으로 보고되고 있다(김종훈·양소영·성지현, 2013; 백지희, 2021; 현정환, 2019a). 그러나 기존 선행연구들은 후속출산에 영향을 미치는 요인을 중점으로 변인의 인과성을 밝

히는 데 초점이 맞춰져왔다. 그리하여 어머니의 연령에 따라 어떤 요인들이 후속출산의도에 영향을 미치는지 비교하기 어렵고 요인 간의 상호작용을 알 수 없다는 한계점이 있다. 머신러닝 및 네트워크 분석은 복합적인 요인들을 토대로 후속출산의도를 예측할 수 있으며 주요 요인과 요인 간의 상호작용을 살펴볼 수 있다는 장점이 있어 많은 분야에서 사용되고 있다(Sebastian & Vahid, 2017).

따라서 본 연구는 어머니의 연령집단별 후속출산의도의 예측요인을 탐색하고 주요 요인을 확인하며 요인 간 상호관련성을 밝히고자 한다. 이를 위해 개인요인, 양육요인, 가구요인, 환경요인 등 다양한 요인들을 복합적으로 조사한 한국 영유아 교육·보육 패널 자료를 활용하고자 한다. 더불어 후속출산의도는 어머니의 연령에 따라 예측요인이 다르게 영향을 미칠 수 있음을 고려하여 20대, 30대, 40대 등 3개 집단으로 분리하여 살펴보고자 한다. 후속출산의도의 예측요인 탐색을 위한 머신러닝 모델 구축 및 네트워크 분석은 가족과 저출산의 근거 자료로써 활용할 수 있으며 후속출산을 장려하기 위한 실증적 자료가 될 것이다.

1. 이론적 배경

가. 후속출산

후속출산이란 첫째 자녀를 출산한 후 둘째 이상의 자녀를 출산하는 것으로 개인의 선택으로 이어지는 출산을 말한다(이정원, 2009). 후속출산의도는 자녀를 출산 및 양육했던 경험에 영향을 받으며 양육 환경, 경제적 여건, 양육 정책, 개인의 가치에 따라 선택의 폭이 크다고 볼 수 있다(김미나·이소은, 2018; Becker, 1992). 후속출산의도는 첫째 자녀 출산보다 사회적 및 문화적 압력이 적고 개인 또는 부부의 자발적 선택이라는 측면에서 저출산의 원인으로 주목하고 있다(Kohler, 2001). 후속출산의도는 개인요인, 양육요인, 가구요인, 환경요인 등 다양한 요인들이 서로 연관되어 영향을 미치는 것으로 보고되고 있다(김종훈·양소영·성지현, 2013; 백지희, 2021; 현정환, 2019a). 양육과정모델(process model)에 따르면 어머니의 심리적요인, 자녀특성, 양육요인, 상황 및 환경적 요인은 서로 영향을 주고받으며 어머니의 행동에 영향을 준다고 보았다(Belsky, 1984). 또한 최근에는 이러한 요인들 간의 상호연관성을 살펴보아야 한다는 주장이 제기되고 있다(강유선·김찬우, 2018).

나. 후속출산의도와 선행연구

어머니의 연령은 후속출산에 직접적인 영향을 미치는 요인으로 밝혀졌으며 연령이 낮을수록 후속출산의도가 높았다(배광일·김경신, 2011; 임현주, 2013). 어머니의 학력은 후속출산에 부정적 영향을 미쳤으며 어머니의 학력은 취업, 일가정 양립과도 관련이 있었다(류기철·박영화, 2009; 조병구·조윤영·김정호, 2007). 어머니의 우울 수준이 낮을수록 후속출산 의도가 높게 나타났으며 후속 출산 의도가 있는 어머니는 자기효능감이 높고 우울 수준은 낮았다(서미정, 2011; 송영주·이주옥·김춘경, 2011; 이영환, 2014). 자아존중감은 후속출산에 정적 영향을 미쳤으며 뿐만 아니라 우울, 자아존중감 등 어머니의 심리적 특성은 양육스트레스와 연관되어 있었고 이는 부부관계에도 서로 밀접한 관련이 있었다(임현주·이대균, 2013; 임현주·이대균·최항준, 2011; 이진화·민정원, 2015). 끈기(grit)는 목표를 달성할 수 있는 개인의 열정과 인내를 말한다(김희명·황매향, 2015; Duckworth, Peterson, Matthews et al, 2007). 끈기는 자신의 감정과 행동을 조절하여 어떠한 일을 해내는 본질적인 특징으로 자녀 양육이라는 어려운 과업에서 개인의 끈기는 부모역할 및 후속출산의도와 밀접한 관련이 있을 것으로 유추해 볼 수 있다.

양육관련 요인을 살펴보면 양육스트레스는 후속출산에 직접적인 영향을 미쳤으며 양육스트레스가 낮을수록 후속출산의도가 높게 나타났다(김은경·박신애, 2019; 임효진·박휴용·장경은 외, 2015). 취업모의 경우 자녀 가치는 후속출산의도에 유의미한 영향을 미쳤으며 후속 자녀를 출산하고자 하는 의도를 강하게 예측하였다(이정원, 2009; 최항준, 2015). 영아기 자녀를 둔 가구를 중심으로 후속출산계획에 영향을 미치는 요인을 탐색한 연구에서는 어머니의 자녀 가치관이 후속출산계획에 영향을 미쳤다(이형민, 2012). 유자녀 임신부의 후속출산 결정 경험 연구에서는 양육행동, 첫 자녀를 양육했던 경험이 후속출산과 관련이 있었다(최여진·조희선, 2018). 양육관 및 양육신념도 후속출산의도와 관련이 높았으며 아버지의 양육 참여는 후속출산의도에 직접적인 영향을 미쳤고 후속출산계획 자녀 수에도 영향을 미쳤다(김종훈·양소영·성지현, 2013; 이형민, 2012; 하민경, 2020; 현정환, 2019b).

가구 및 환경요인은 출산에 직접적인 영향을 미치는 것으로 알려져 왔다(Adsera, 2005; Lesthaeghe & Willems, 1999). 출산한 자녀 수에 따라 출산의지에 영향을 미치는 요인이 달랐으며(서정연·김한곤, 2015), 아동의 지출비는 미취업, 취업 어머니의 후속출산에 유의한 영향을 미치는 것으로 나타났다(강유선·곽승주, 2018). 소득에 따라 저소득 집단에서는 어머니의 연령이 유의하였으며 중소득 집단 및 고소득 집단에서는 어머니의 연령, 아동지출비가 후속출산계획에 영향을 미쳤다(강유선·김찬우, 2018). 다만, 경제적 측면에서 소득이 높은 경우 출산율이 증가한다는 연구 결과도 함께 제시되고 있어 복합적인 요인들

간의 관련성을 확인하고 요인들의 상호 연관성에 대한 고려가 함께 이루어져야 할 필요가 있다(김하나, 2008; 공선영, 2006; 이미란, 2010). 또한 최근 한국 사회의 다문화가구의 증가로 인한 가구의 다양성 및 그 영향도 살펴보아야 한다는 주장이 제기되고 있어 이 연구에서 함께 살펴보고자 한다(김지영·고선, 2021). 어머니들의 결혼만족도는 후속출산계획을 예측하는 것으로 밝혀졌으며 부모의 결혼만족도가 높을수록 후속 자녀 출산 의도가 높았다(김은경·박신애, 2019; 임양미, 2015; 정미라·홍소영·박수경, 2013). 한편 부부갈등이 높을수록 후속출산계획은 낮아졌으며 부부관계는 후속출산계획에 직접적인 영향을 미쳤다(이진화·김경은·임원신, 2015; 이진화·조하나·최혜윤, 2012). 가족 상호작용도 후속출산계획에 유의한 영향을 미쳤으며 사회적 지원은 후속출산계획에 정적 영향을 미치는 것으로 나타났다(강유선·곽승주, 2018; 이형민, 2012). 환경요인 중 지역사회 내 문화시설 접근 편리성도 후속출산계획 의지에 영향을 미쳤다(김은경·박신애, 2019).

II. 연구방법

1. 연구대상

2022년 한국 영유아 교육·보육 패널(Korean Early Childhood Education & Care Panel 이하 K-ECEC-P) 조사는 2022년에 출생한 아동을 모집단으로 기초조사 I, 기초조사 II, 생후 1개월(임부 2,786명), 1차 본조사(생후 4개월 아동 2,775명)가 진행되었다. 이 연구는 한국 영유아 교육·보육 패널 홈페이지에서 1차년도 본조사 자료를 받아 사용하였다. 배우자가 있는 어머니 2,442명 중 결혼값이 있는 데이터를 제외하고 20대 267명, 30대 1,498명, 40대 157명, 총 1,922명의 자료를 사용하였다. 연구대상의 일반적 특성은 <표 1>과 같다.

<표 1> 연구대상의 일반적 특성

		(N=1,922)
	구분	빈도(%)
연령	20대	267(13.9)
	30대	1,498(77.9)
	40대	157(8.2)
학력	중졸이하	18(9)
	고졸	262(13.6)

구분		빈도(%)
지역	전문대졸	399(20.8)
	대졸	1,046(54.4)
	대학원	175(9.1)
	서울	245(12.7)
	경기,인천	628(32.7)
	강원,충청권	204(10.6)
	경상권	626(32.6)
	전라,제주권	219(11.4)

2. 측정도구

가. 결과변수

후속출산의지를 결과변수로 설정하였다. 후속 자녀 출산에 대한 의지를 묻는 질문으로 단일 문항으로 후속출산 계획 여부에 낳지 않음 0점, 후속출산계획 여부에 낳겠음(또는 임신중) 1점으로 측정한다. 이 연구에서도 후속출산 계획 여부에 낳지 않음 0점, 후속출산계획 여부에 낳겠음(또는 임신중) 1점으로 사용하였다.

나. 예측요인

본 연구에서는 개인요인, 양육요인, 가구요인, 환경요인의 측면에서 총 20개의 요인을 예측요인으로 설정하였다.

〈표 2〉 후속출산의도 예측요인

연령: 1=10대, 2=20대, 3=30대, 4=40대
학력: 1=무학, 2=초졸, 3=중졸, 4=고졸, 5=전문대졸, 6=대졸, 7=대학원석사, 8=대학원박사
우울: K-EPDS, 총 10문항, '예전과 똑같았다(1점)'~'전혀 그렇지 않았다(4점)', 평균사용, Cronbach's α .87
성격(그릿): The Short Grit Scale, 총 8문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)', 평균사용, Cronbach's α .71
자아존중감: Rosenberg Self-Esteem Scale, 총 10문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(4점)', 평균사용, Cronbach's α .89
양육스트레스: Parenting Stress Scale, 총 10문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)', 평균사용, Cronbach's α .87

양육행동: The Parental Style Questionnaire, 총 9문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)', 평균사용, Cronbach's α .83
 자녀가치: 총 8문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)', 평균사용, Cronbach's α .78
 양육신념: 자녀에 대한 부모의 책임을 묻는 단일문항, 이항변수
 아버지양육참여: 총 4문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)', 평균사용, Cronbach's α .77
 어머니취업상태: 0=미취업, 1=취업
 월평균가구소득: 1=300만원 이하, 2=301~500만원, 3=501~700만원, 4=701만원 이상
 주관적 사회경제적 지위: 단일문항, '낮음(1점)'~'높음(10점)
 저소득가구 여부: 0=저소득가구 아님, 1=저소득가구
 다문화가구 여부: 0=다문화가구 아님, 1=다문화가구
 총자녀수(아동 출생순위): 1명(첫째), 2명(둘째), 3명(셋째), 4명(넷째), 5명(다섯째), 6명(여섯째)
 결혼만족도: 총 4문항, '전혀 만족하지 않는다(1점)'~'매우 만족한다(7점)', 평균사용, Cronbach's α .93
 부부갈등: 총 8문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)', 평균사용, Cronbach's α .92
 사회적지지: 단일문항, '전혀 그렇지 않다(1점)'~'매우 그렇다(5점)
 지역사회 양육환경 만족도: 단일문항, '전혀 좋지않음(1점)'~'매우 좋음(5점)'

3. 분석방법

머신러닝 알고리즘은 데이터를 토대로 모델을 훈련시켜 요인을 분류하고 예측할 수 있으며, 새로운 데이터에 적용하여 일반화할 수 있는 장점이 있다. 머신러닝 분석은 개인요인, 양육요인, 가구요인, 환경요인 등 다양한 요인을 동시에 검증할 수 있어 후속 출산의 다양한 예측 요인을 탐색하는 데 유용하다. 다만, 분석 모델에 따라 결과에 편향이 있을 수 있어 몇 가지 모델을 비교하여 최적의 모델을 선택하는 것이 권장된다(Sebastian & Vahid, 2017). 랜덤 포레스트(Random forest) 모델은 결정 트리의 앙상블을 사용하여 예측하는 모델로, 일반화 성능이 높고 과적합의 위험이 적어 안정적인 모델을 만들 수 있다(Breiman, 2001). 그래디언트 부스팅(Gradient Boosting) 모델은 이전 모델의 오차에 가중치를 부여하여 보완하는 부스팅 방식의 앙상블 기법으로, 뛰어난 추론 성능을 얻을 수 있어 많이 활용되고 있다(Natekin & Knoll, 2013). 로지스틱 회귀(Logistic Regression) 모델은 선형 방정식을 사용하여 입력 변수와 결과 변수 간의 관계를 모델링하고, 로지스틱 함수를 통해 이를 확률로 변환한다. 이 모델은 간단하면서도 효과적이며, 데이터가 선형의 경우에 성능이 우수하다(Cox, 1958). 스택킹(Stacking) 모델은 메타 모델의 앙상블 기법으로, 다양한 예측 요인에서 최적의 성능으로 모델을 예측할 수 있다(Syarif, Zaluska, Prugel-Bennett et al., 2012). 네트워크 분석은 밀도 분석, 중심성 분석 등 변인들의 관계를 구조적으로 표현하고 계량화하는 분석 방법이다. 밀도 분석은 변인들 간의 관계를 노

드(Node)와 링크(Link)로 연결하여 네트워크 행렬을 시각화하며, 상호 연계성을 토대로 구조적 관계를 파악할 수 있다. 연결 중심성(degree centrality)은 특정 변인의 중심 위치를 보여주어 중심성에 따라 변인의 영향이 크다는 것을 의미한다. 근접 중심성(closeness centrality)은 변인과 변인이 얼마나 가까이 있는지를 살펴봄으로써 관련성을 추측할 수 있다. 마지막으로 매개 중심성(betweenness centrality)은 특정 변인이 다른 변인들을 연결해 주는 매개 역할과 흐름을 살펴봄으로써 중재적 역할을 확인해 볼 수 있다(Sebastian & Vahid, 2017).

이 연구에서는 후속출산의도의 예측요인 탐색 및 네트워크 분석을 위해 Google의 Colaboratory(Colab)를 사용하였다. 후속출산의도를 예측하는 요인 탐색 및 네트워크 분석을 하기 위해 데이터의 결측치, 극단값을 처리하였으며 데이터를 정규화, 표준화, 데이터 전 처리를 하였다. 모델을 pipeline으로 구축하였고 GridSearch로 최적의 성능의 하이퍼파라미터를 적용하였다(Sebastian & Vahid, 2017). 일반화하기 위해 K-fold 교차 검증을 사용하였고 모델의 성능을 평가하기 위해 metrics 라이브러리로 Confusion Matrix를 구하고 정확도, 정밀도, 재현율, F1값, AUC을 산출하였다. feature_importances를 활용하여 중요도를 산출하였다. networkx, matplotlib를 활용해 네트워크를 산출하고 시각화하였다.

〈표 3〉 하이퍼파라미터

Model	Random Forest	gradientboosting	Logistic Regression	Stacking
Hyperparameter	criterion=gini n_estimators=20 max_features=4 max_depth=5	n_estimators=35 max_features=2 max_depth=5	max_iter=1000	Random Forest gradientboosting Logistic Regression
Grid Search	n_estimators=range(5, 51, 5) max_features=range(1, 5)		max_depth=range(1, 6) cv=10, n_jobs=-1	

정확도(CA: Classification Accuracy)=(TP+TN)/(TP+TN+FP+FN) 정확하게 예측 비율
 정밀도(Precision) =TP/(TP+FP) 양성을 예측한 데이터 중 실제 양성인 값을 계산한 값
 재현율(Recall) =TP/(TP+FN) 전체 양성 중에서 예측으로 양성을 찾아낸 비율
 오류율 =(FN+FP)/(TN+FP+FN+TP)
 가중평균(F1)=2×(Precision×Recall)/(Precision+Recall) 모델 성능 평가에 유용한 지표
 ROC 곡선은 FPR(False Positive Rate)이 변할 때 TPR(True Positive Rate)이 어떻게 변하는지 나타내는 곡선
 FPR(False Positive Rate): FP/(FP+TN), 1-TNR, TPR(True Positive Rate): TP/(FN+TP), 재현율
 AUC(Area Under Curve) 값은 ROC 곡선 밑에 면적을 구한 값 (10이 가까울수록 좋은 값)

III. 연구결과

1. 집단별 후속출산 여부

후속출산의도 연령별로 살펴보면 <표 4>와 같다. 20대 집단의 경우 다른 연령의 집단보다 후속출산의도가 상대적으로 높았다. 30대 집단은 20대 집단보다 감소하였으나 40대보다 높은 것으로 나타났다. 마지막으로 40대 집단의 경우 다른 연령 집단보다 후속출산의도가 현저히 낮았다. 전반적으로 후속출산의도는 20대를 제외하고 낮은 편이었다.

<표 4> 집단별 출산 의도

(N=1,922)

후속출산의도	날지않음	날것음(또는 임신 중)
20대(%)	121(45.3)	146(54.7)
30대(%)	1,009(67.4)	489(32.6)
40대(%)	141(89.8)	16(10.2)
Total(%)	1,271(66.1)	651(33.9)

2. 후속출산의도 모델의 성능 평가

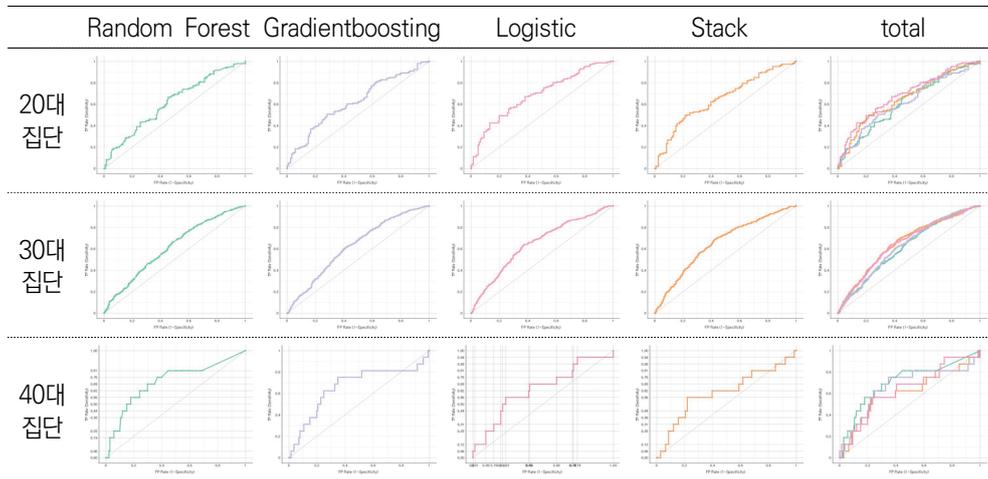
<표 5>에 제시한 바와 같이 CA는 후속출산의도를 정확하게 예측한 비율이며 20대 집단은 .60~.65으로 나타났다. 30대 집단에서는 .64~.68이었고 40대 집단은 .88~.89이었다. Precision를 살펴보면 20대 집단에서는 .61~.69이었고 30대 집단에서는 .61~.64이었으며 40대 집단에서는 .80~.86으로 나타났다. Recall의 경우 20대 집단에서는 .68~.75으로 나타났다고 30대 집단에서는 .64~.68이었으며 40대 집단에서는 .88~.89으로 나타났다. 모델의 성능 평가 F1을 살펴보면 20대 집단에서는 .66~.68이었고 30대 집단에서는 .60~.62이었으며 40대 집단에서는 .85~.87이었다. 20대 집단과 30대 집단에서 모든 모델이 보통 수준의 성능을 보였으며 40대 집단에서는 모든 모델이 우수한 성능을 보였다.

<표 5> 후속출산의도 예측 성능

	CA	Precision	Recall	F1	AUC	
20대	Random Forest	.65	.69	.68	.68	.61
집단	Gradientboosting	.60	.64	.69	.67	.62
	Logistic	.61	.62	.70	.66	.68
	Stacking	.60	.61	.75	.67	.66

		CA	Precision	Recall	F1	AUC
30대 집단	Random Forest	.64	.61	.64	.61	.61
	Gradientboosting	.65	.61	.66	.61	.62
	Logistic	.68	.64	.68	.62	.65
	Stacking	.67	.62	.67	.60	.65
40대 집단	Random Forest	.89	.80	.89	.85	.71
	Gradientboosting	.88	.83	.88	.85	.66
	Logistic	.89	.86	.89	.87	.64
	Stacking	.89	.80	.89	.85	.61

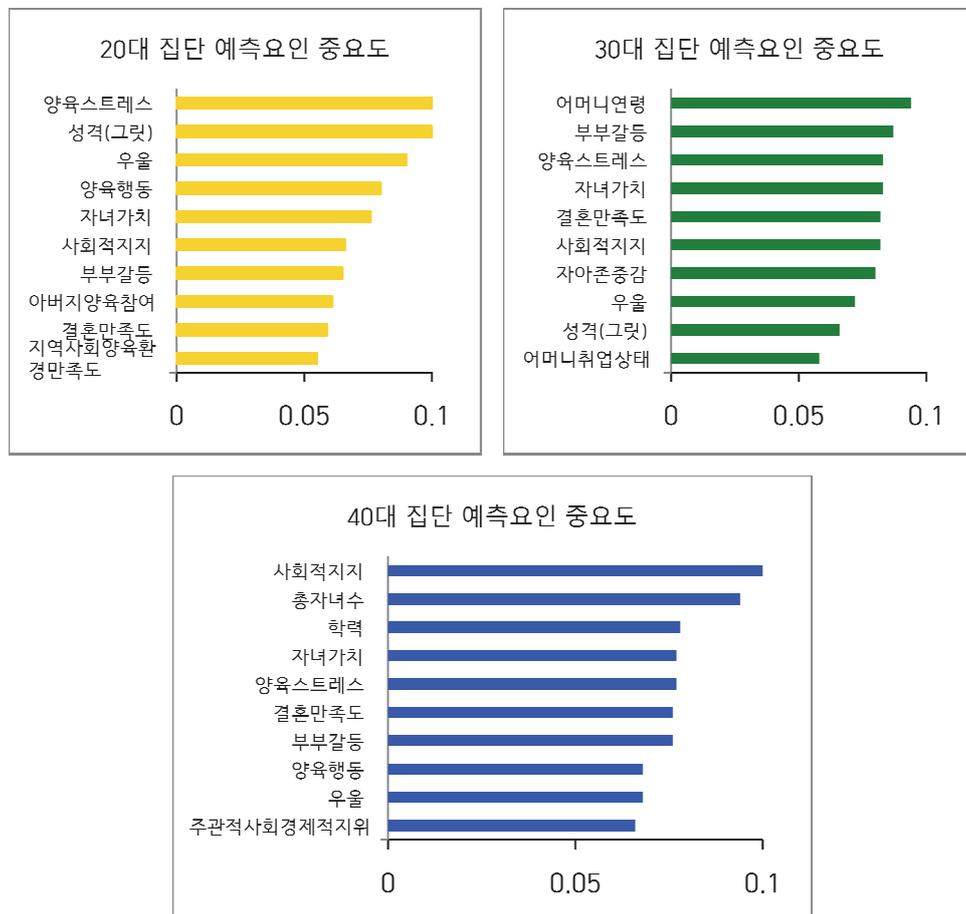
ROC(Receiver Operating Characteristic) 곡선은 분류 모델의 성능을 평가하고 비교하는 데 유용한 시각적 도구로 모델이 얼마나 잘 예측하는지 볼 수 있다. [그림 1]에 제시한 바와 같이 ROC 결정곡선을 살펴보면 20대 집단의 경우 Random Forest .61, Gradientboosting .62, Logistic Regression .68, Stack .66으로 .61~.68 범위로 나타났다. 30대 집단의 경우 Random Forest .61, Gradientboosting .62, Logistic Regression .65, Stack .65으로 .61~.65 범위로 나타났다. 40대 집단의 경우 Random Forest .71, Gradientboosting .66, Logistic Regression .64, Stack .61으로 .61~.71 범위로 나타났다.



[그림 1] ROC 결정곡선

3. 예측 요인의 중요도

20대 집단의 후속출산의도를 예측하는 상위 10개 주요 요인은 [그림 2]와 같이 양육스트레스, 성격(그릿), 우울, 양육행동, 자녀가치, 사회적지지, 부부갈등, 아버지양육참여, 결혼만족도, 지역사회양육환경만족도로 나타났다. 30대 집단은 어머니연령, 부부갈등, 양육스트레스, 자녀가치, 결혼만족도, 사회적지지, 자아존중감, 우울, 성격(그릿), 어머니취업상태로 나타났다. 40대 집단의 경우 사회적지지, 총자녀수, 학력, 자녀가치, 양육스트레스, 결혼만족도, 부부갈등, 양육행동, 우울, 주관적사회경제적지위로 나타났다.



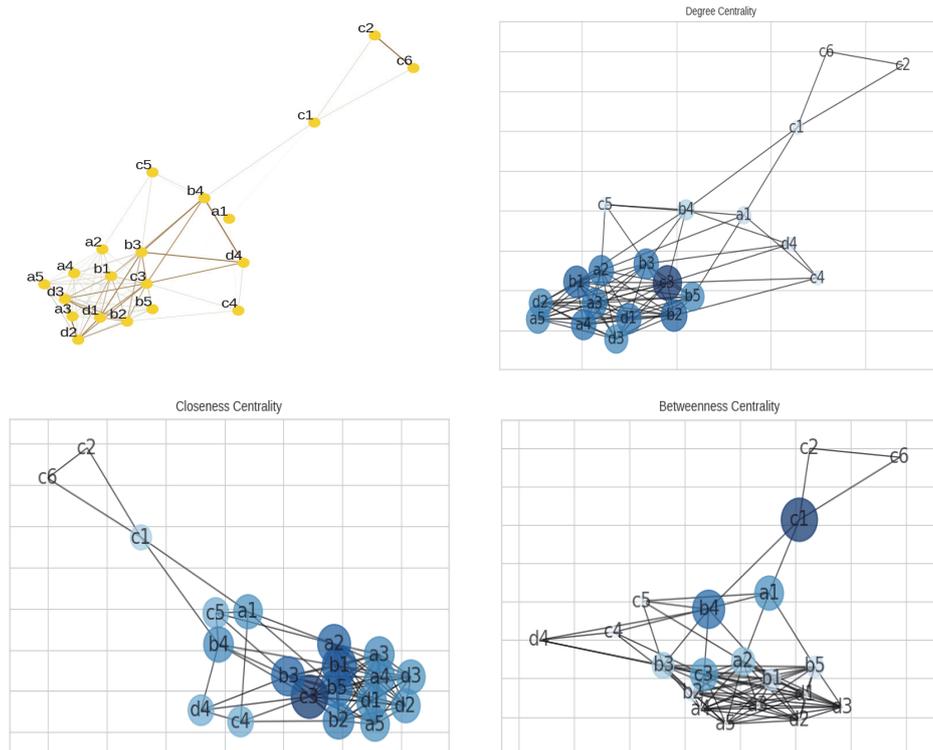
[그림 2] 후속출산의도 예측요인 중요도

이러한 결과는 모든 집단에서 양육스트레스, 우울, 자녀가치, 사회적지지, 부부갈등, 결혼만족도가 공통적으로 주요 요인임을 알 수 있다. 그리고 공통적으로 나타난 주요 요인들은 어머니들이 자녀를 양육하는 과정에서 경험하는 심리적 어려움 및 부부관계에서 비롯된 요인임을 알 수 있다. 연령에 따라 살펴보면 20대 집단의 경우 아버지양육참여, 양육행동, 성격(그릿), 지역사회양육환경만족도가 주요 요인으로 나타난 반면, 30대 집단은 어머니연령, 자아존중감, 어머니취업상태, 성격(그릿)으로 나타났으며 40대 집단에서 학력, 총자녀수, 주관적사회경제적지위, 양육행동으로 나타나 후속출산의도를 예측하는 주요 요인이 연령에 따라 차이가 있다는 것을 알 수 있다. 요인을 중점으로 살펴보면 성격(그릿)은 20대 집단과 30대 집단에서 공통적으로 나타났으나 40대 집단에서는 주요 요인으로 나타나지 않았다. 아버지양육참여와 지역사회양육환경만족도는 20대 집단에서 후속출산의도를 예측하는 주요 요인으로 나타났다. 어머니연령, 자아존중감, 어머니취업상태는 30대 집단에서 후속출산의도를 예측하는 주요 요인으로 나타났으며 총자녀수, 학력, 주관적사회경제적지위는 40대 집단에서 주요 요인으로 나타났다.

4. 네트워크 분석

네트워크 분석은 후속 출산에 영향을 미치는 요인들이 서로 어떻게 관련되어 있는지, 그리고 어떤 요인이 가장 중요한지 요인들 간의 관련성을 탐색할 수 있다. 집단에 따라 네트워크, 연결중심성, 근접중심성, 매개중심성을 산출하였으며 시각화하였다. 네트워크 중심성 분석 결과는 <표 6>, 시각화는 [그림 3]~[그림 5]와 같다. 먼저 20대 집단의 예측 요인들 간 상호관련성을 살펴보면 [그림 3]과 같이 양육신념, 자녀가치, 양육스트레스, 아버지양육참여, 양육행동 등 양육관련 요인들은 결혼만족도, 부부갈등, 사회적지지, 지역사회양육환경만족도와 밀접한 연결고리를 가지고 있었고 우울, 성격(그릿)과도 연결되어 있었다. 월평균가구소득은 총자녀수와 직접적으로 연결되어 있었으며 어머니 취업상태와 연결되어 있었다. 20대 집단의 주요 요인들 간 연결 중심성 분석을 살펴보면 주관적사회경제적지위가 가장 중심적 역할을 하는 요인으로 나타났다. 주관적사회경제적지위는 양육스트레스, 양육행동과 가장 밀접한 연결을 보였으며 자녀가치, 우울과도 연결되어 있었다. 두 번째로 근접 중심성은 특정 요인이 다른 요인과의 경로가 얼마나 짧은지를 나타내며 연결 중심성과 다르게 직간접적으로 연결된 요인들을 추정한다. 주관적사회경제적지위는 양육스트레스와 가까운 연결을 보였으며 자녀가치, 아버지양육참여, 사회적지지와도 근접한 연결을 보였다. 마지막으로 매개 중심성은 요인 간의 관계를 매개하는지 추정한다. 학력은 어머니연령을 매개로 어머니취업상태와 연결되어 있었으며 자녀가치는 양육신념을 매개로 양육

스트레스와 연결되어 있었다. 종합하면 20대 집단의 후속출산의도를 예측하는 주요 요인들은 주관적사회경제적지위를 중심으로 자녀 양육과 관련된 요인들과 밀접하게 연결되어 있었으며 자녀가치, 사회적지지와 상호 관련이 있는 것으로 판단된다.

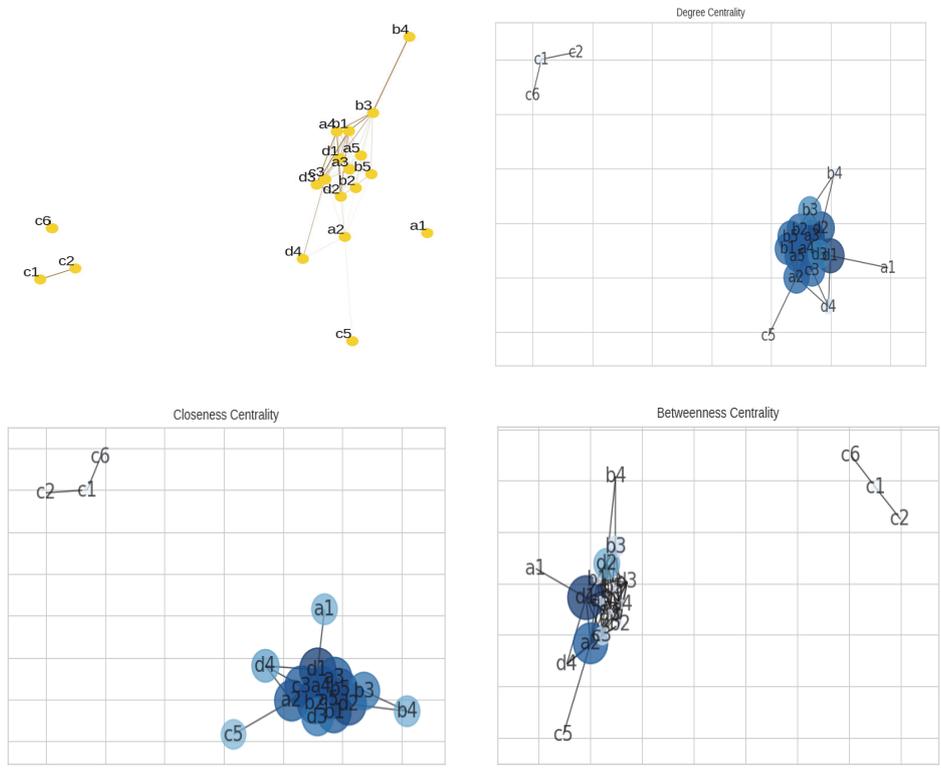


a1=어머니연령, a2=학력, a3=우울, a4=성격(그릿), a5=자아존중감, b1=양육스트레스, b2=양육행동, b3=자녀가치, b4=양육신념, b5=아버지의양육참여, c1=어머니취업상태, c2=월평균가구소득, c3=주관적사회경제적지위, c4=저소득가구, c5=다문화가구, c6=총자녀수, d1=결혼만족도, d2=부부갈등, d3=사회적지지, d4=지역사회양육환경만족도

[그림 3] 20대 집단의 중심성 네트워크

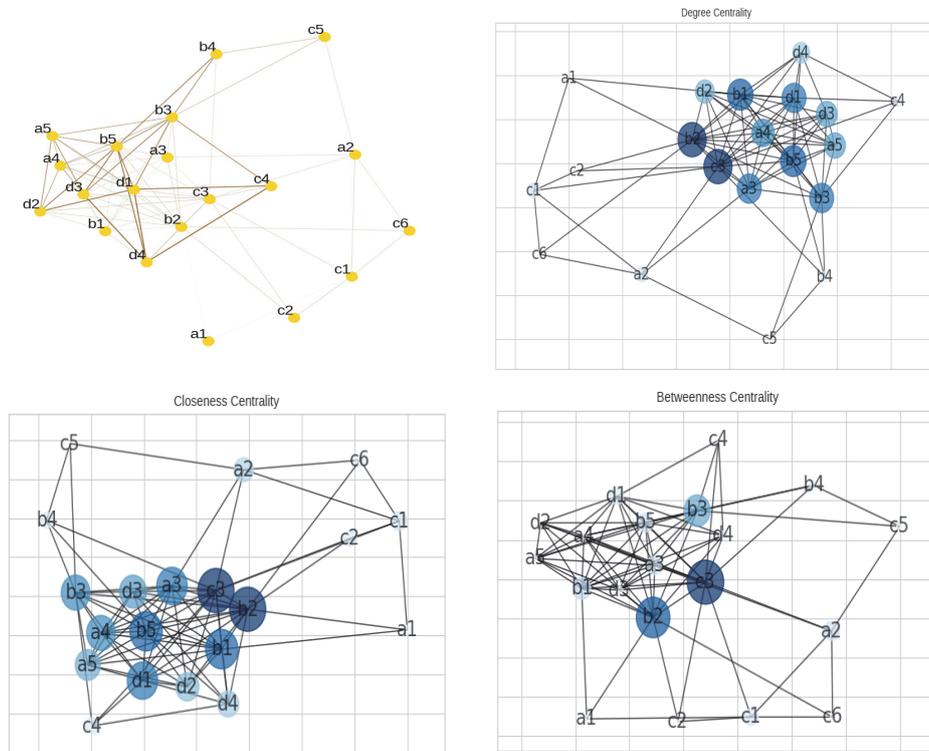
30대 집단의 예측 요인들 간 상호관련성을 살펴보면 [그림 4]와 같이 우울을 중심으로 결혼만족도, 자아존중감, 아버지양육참여, 양육행동, 양육스트레스, 주관적사회경제적지위와 밀접한 연결을 보였으며 자녀가치는 양육신념과 연결고리를 가지고 있었다. 어머니취업상태는 월평균가구소득과 직접적으로 연결되어 있었으나 총자녀수와는 연결되어 있지 않았다. 30대 어머니 집단의 주요 요인들 간 연결 중심성 분석을 살펴보면 결혼만족도가 중

심적 역할을 하는 요인으로 나타났다. 결혼만족도를 중심으로 양육스트레스, 양육행동, 아버지양육참여가 연결되어 있었으며 우울, 성격(그릿), 자아존중감도 연결되어 있었다. 30대 어머니 집단의 근접 중심성 분석을 살펴보면 결혼만족도는 부부갈등, 아버지양육참여와 근접한 연결을 보였으며 우울, 성격(그릿), 자아존중감과 연결되어 있었다. 30대 집단의 주요 요인들 간 매개 중심성 분석을 살펴보면 결혼만족도는 양육스트레스를 매개로 부부갈등과 연결되어 있었다. 종합하면 30대 집단의 후속출산의도를 예측하는 주요 요인들은 결혼만족도를 중심으로 자녀 양육 요인, 우울, 자아존중감과 밀접한 상호 관련이 있는 것으로 판단된다.



a1=어머니연령, a2=학력, a3=우울, a4=성격(그릿), a5=자아존중감, b1=양육스트레스, b2=양육행동, b3=자녀가치, b4=양육신념, b5=아버지의양육참여, c1=어머니취업상태, c2=월평균가구소득, c3=주관적사회경제적지위, c4=저소득가구, c5=다문화가구, c6=총자녀수, d1=결혼만족도, d2=부부갈등, d3=사회적지지, d4=지역사회양육환경만족도

[그림 4] 30대 집단의 중심성 네트워크



a1=어머니연령, a2=학력, a3=우울, a4=성격(그릿), a5=자아존중감, b1=양육스트레스, b2=양육행동, b3=자녀가치, b4=양육신념, b5=아버지의양육참여, c1=어머니취업상태, c2=월평균가구소득, c3=주관적사회경제적지위, c4=저소득가구, c5=다문화가구, c6=총자녀수, d1=결혼만족도, d2=부부갈등, d3=사회적지지, d4=지역사회양육환경만족도

[그림 5] 40대 집단의 중심성 네트워크

40대 어머니집단의 예측 요인들 간 상호관련성을 살펴보면 [그림 5]와 같이 사회적지지는 아버지양육참여, 부부갈등, 지역사회양육환경만족도와 직접적인 연결고리를 가지고 있었고 성격(그릿), 자아존중감과도 연결되어 있었다. 학력은 어머니취업상태와 연결되어 있었고 어머니취업상태는 월평균가구소득, 주관적사회경제적지위, 총자녀수와 연결고리를 가지고 있었다. 40대 집단의 주요 요인들 간 연결 중심성 분석을 살펴보면 양육행동과 주관적사회경제적지위가 가장 중심적 역할을 하는 요인으로 나타났다. 양육행동은 양육스트레스, 아버지양육참여와 밀접한 연결을 보였으며 부부갈등과도 연결되어 있었다. 주관적사회경제적지위의 경우 우울, 성격(그릿)과 밀접한 연결을 보였다. 40대 집단의 근접 중심성 분석을 살펴보면 양육행동을 중심으로 주관적사회경제적지위가 가장 근접한 연결을 보였으며 양육스트레스, 우울과 연결되어 있었다. 40대 집단의 주요 요인들 간 매개 중심성 분석

을 살펴보면 양육행동은 주관적사회경제적지위를 매개로 우울과 연결되어 있었다. 종합하면 40대 집단의 후속출산의도를 예측하는 주요 요인들은 주관적사회경제적지위를 중심으로 경제 관련 요인들과 밀접한 상호 관련이 있으며 이는 양육행동, 양육스트레스와도 상호 관련이 있는 것으로 판단된다.

〈표 6〉 요인들 간 중심성 분석 결과

요인	20대 집단			30대 집단			40대 집단		
	연결	근접	매개	연결	근접	매개	연결	근접	매개
어머니연령	.36	.61	.13	.05	.40	.00	.15	.50	.00
학력	.57	.65	.05	.61	.62	.10	.26	.57	.03
우울	.63	.61	.01	.61	.65	.00	.57	.70	.03
성격(그릿)	.57	.59	.00	.61	.65	.00	.52	.67	.00
자아존중감	.57	.59	.01	.61	.65	.00	.47	.63	.00
양육스트레스	.68	.70	.03	.61	.65	.00	.63	.73	.04
양육행동	.68	.63	.01	.61	.65	.00	.73	.79	.15
자녀가치	.63	.67	.06	.50	.56	.09	.57	.67	.09
양육신념	.31	.59	.09	.05	.34	.00	.21	.52	.01
아버지양육참여	.57	.65	.02	.61	.65	.00	.63	.73	.03
어머니취업상태	.26	.48	.20	.11	.11	.00	.26	.54	.02
월평균가구소득	.10	.33	.00	.05	.07	.00	.15	.52	.00
주관적사회경제적지위	.78	.76	.07	.61	.62	.01	.73	.79	.18
저소득가구여부	.31	.55	.01	.01	.02	.00	.21	.52	.00
다문화가구여부	.26	.57	.04	.05	.36	.00	.15	.48	.00
총자녀수	.10	.33	.00	.05	.07	.00	.15	.51	.00
결혼만족도	.63	.61	.00	.72	.73	.13	.57	.70	.02
부부갈등	.63	.67	.02	.61	.65	.00	.42	.61	.00
사회적지지	.52	.57	.00	.50	.54	.00	.47	.65	.00
지역사회양육환경만족도	.42	.59	.01	.16	.44	.00	.36	.59	.00

IV. 논의 및 결론

본 연구는 한국 영유아 교육·보육 패널 자료를 활용하여 랜덤포레스트, 그래디언부스팅, 로지스틱 회귀 모델, 스택킹 모델을 통해 후속출산의도의 예측 요인을 탐색하고 네트워크

분석을 통해 예측 요인 간의 상호작용을 확인하였다. 연령에 따라 20대 집단, 30대 집단, 40대 집단으로 구분하여 각 집단별 모델 예측 성능을 비교하였으며 후속출산의도에 미치는 주요 요인의 중요도를 확인하였다. 그 결과 첫째, 후속출산의도를 예측하는 랜덤포레스트 모델, 그래디언트부스팅 모델, 로지스틱 회귀 모델, 스택킹 모델은 F1 값을 기준으로, 60~.87 예측력을 보였고 보통 수준 이상의 성능을 보였다.

둘째, 후속출산의도를 예측하는 주요 요인은 집단에 따라 차이가 있었다. 20대 집단의 후속출산의도를 예측하는 상위 10개 주요 요인은 양육스트레스, 성격(그릿), 우울, 양육행동, 자녀가치, 사회적지지, 부부갈등, 아버지 양육참여, 결혼만족도, 지역사회양육환경만족도로 나타났으며 30대 집단의 후속출산의도를 예측하는 상위 10개 주요 요인은 어머니연령, 부부갈등, 양육스트레스, 자녀가치, 결혼만족도, 사회적지지, 자아존중감, 우울, 성격(그릿), 어머니취업상태로 나타났다. 마지막으로 40대 집단의 경우 후속출산의도를 예측하는 상위 10개 주요 요인은 사회적지지, 총자녀수, 학력, 자녀가치, 양육스트레스, 결혼만족도, 부부갈등, 양육행동, 우울, 주관적사회경제적지위로 나타났다. 이러한 결과는 20대 집단, 30대 집단, 40대 집단에서 후속출산의도를 예측하는 주요 요인은 양육스트레스, 우울, 자녀가치, 사회적지지, 부부갈등, 결혼만족도가 공통 요인이며 이는 개인의 심리적 요인, 양육요인, 환경요인 등 복합적인 요인이 후속출산의도와 관련이 있다는 선행연구와도 같은 맥락이다(김종훈·양소영·성지현, 2013; 백지희, 2021). 한편, 20대 집단의 경우 아버지양육참여, 양육행동, 성격(그릿), 지역사회양육환경만족도가 후속출산의도를 예측하는 주요 요인으로 나타났으며 30대 집단에서는 어머니연령, 자아존중감, 어머니 취업상태, 성격(그릿)이 후속출산의도를 예측하는 주요 요인으로 나타났다. 마지막으로 40대 집단에서는 학력, 총자녀수, 주관적사회경제적지위, 양육행동이 후속출산의도를 예측하는 주요 요인으로 나타났다. 이는 생애주기에 따라 후속출산의도에 영향을 미치는 요인이 다르며 관련 요인의 영향력이 변화할 수 있다는 선행연구와 동일한 결과이다(강현구·심도현·박윤현 외, 2014; 임현주, 2013). 또한 어머니의 연령에 따라 후속출산의도를 예측하는 주요 요인이 다를 수 있음을 확인할 수 있었다.

셋째, 네트워크 분석을 통해 후속 출산에 영향을 미치는 요인들의 상호 관련성은 집단에 따라 다르게 나타났다. 20대 어머니 집단의 후속출산의도를 예측하는 주요 요인들의 상호 관련성을 살펴보면 주관적사회경제적지위를 중심으로 자녀 양육과 관련된 요인들과 밀접하게 연결되어 있었으며 자녀가치, 사회적지지와 상호 관련이 있었다. 30대 어머니 집단의 후속출산의도를 예측하는 주요 요인들의 경우 결혼만족도를 중심으로 자녀 양육 요인, 우울, 자아존중감이 밀접한 상호 관련이 있었다. 40대 어머니 집단의 경우 주관적사회경제적

지위를 중심으로 경제 관련 요인들과 밀접한 상호 관련이 있으며 이는 양육행동, 양육스트레스와도 상호 관련이 있었다.

이 연구는 다음과 같은 몇 가지 제한점이 있다. 첫째, 어머니만을 대상으로 연구를 진행하여 아버지와의 차이점을 살펴보지 못하였다. 추후 연구에서는 어머니와 아버지 간의 공통점과 차이점은 무엇인지 확인해 볼 필요가 있다. 둘째, 선행연구를 토대로 맞벌이 여부와 가구 지출을 반영하고자 하였으나, 결측값이 많아 이 연구에 포함하지 못했다는 제한점이 있다. 또한, 결측치 처리 과정에서 데이터 수가 감소하여 일반화 가능성에 한계가 있을 수 있다. 셋째, 머신러닝에서 많이 활용되는 모델을 적용하여 후속출산의도를 예측하는 모델 구축 및 주요 요인을 살펴보았으나 주요 요인이 어떤 영향을 주는지(부적, 정적), 인과관계를 살펴보지 못하였다는 한계점이 있다. 후속 연구에서 SHAP 라이브러리를 활용하거나 구조방정식 모델을 활용하여 인과관계를 살펴볼 필요가 있다. 그럼에도 이 연구에서는 기존의 연구들이 살펴보지 못했던 복합적 요인들을 토대로 후속출산의도를 예측하는 모델을 구축하고 주요 요인 및 상호작용을 확인했다는 점에서 중요한 기여라 볼 수 있다. 또한 어머니를 단일한 집단으로 보지 않고 연령에 따라 집단을 구분했다는 점, 어머니의 연령에 따라 후속출산의도를 예측하는 주요 요인이 다르다는 점을 규명했고, 네트워크 분석을 활용하여 주요 요인 간의 상호 연관성을 탐색하였다는 점에서 저출산 정책 방향성의 근거 자료로써 활용할 수 있으며 후속 연구의 방향을 제시하였다는 점에서 의의가 있다.

V. 정책적 제언 및 시사점

본 연구는 한국 영유아 교육·보육 패널 자료를 활용하여 랜덤포레스트, 그래디언트부스팅, 로지스틱 회귀 모델, 스택킹 모델을 통해 후속출산의도의 예측 요인을 탐색하고 네트워크 분석을 통해 예측 요인 간의 상호작용을 확인하였다. 이를 통해 후속출산의도 예측 모델의 성능을 비교하고, 연령별로 중요한 예측 요인을 도출하였다. 이러한 결과는 후속출산의도를 예측하고 지원하는 데 있어 중요한 시사점을 제공한다. 우선 모든 연령대에서 양육스트레스와 우울이 주요 예측 요인으로 나타났다. 이는 어머니들의 정신건강과 양육 과정에서 스트레스 관리가 후속출산의도에 중요한 영향을 미친다는 것을 의미하며, 이는 양육자의 정서 지원을 위한 개입 및 시스템이 마련되어야 할 필요가 있음을 시사한다(임효진·박휴용·장경은 외, 2015; 하민경, 2020). 기존의 부모교육 프로그램은 주로 양육 기술, 양육 효능감 증진, 부모-자녀 관계 향상 등에 중점을 두고 있어 어머니들의 정신건강과 양육의

어려움을 해결하는 데 한계가 있다(송유미·이선자, 2016). 다음으로 결혼만족도와 부부갈등, 자녀가치도 모든 연령대에서 공통된 주요 예측 요인으로 나타났으며, 이 요인들은 자녀양육 요인들과 밀접한 연결을 보인다는 점을 주목할 필요가 있다. 이는 자녀의 양육 책임이 개인, 특히 여성에서 집중 되어있는 한국 사회의 구조적 측면과 연결되며(엄연용·송원영, 2022), 공동양육과 부부관계의 질이 후속출산의도를 예측할 수 있다는 것을 의미한다. 저출산을 극복하고 있는 프랑스는 출산과 자녀 양육을 국가의 책임으로 인식하고 가족 중심의 문화 형성을 지원하고 있다. 이러한 인식을 반영하여 국내에서도 가족 친화적인 문화 조성을 토대로 건강하고 균형 있는 공동양육 및 부부관계 질을 향상 시킬 수 있는 정책이 필요할 것이다(임현주·이대균, 2013; 정은희·최유석, 2013). 사회적지지도 후속출산의도를 예측하는 주요 요인으로 나타났다. 이는 어머니들이 사회적 지원을 받을 수 있는 네트워크를 구축하는 것이 중요함을 보여준다. 현재 우리나라에서 시행되고 있는 출산 장려 프로그램과 정책은 경제적 지원에 초점을 맞추고 있지만, 이러한 지원은 출생아 수 증가에 크게 기여하지 못하는 것으로 나타났다(최분희·이창원, 2017). 해외의 경우 핀란드에서는 사회적 지원이 첫째 자녀부터 이루어지고 있으며 프랑스의 경우 일시적인 현금지원이 아닌 임신부터 휴가 및 휴직 보장, 출산 및 종합 육아 수당, 무상보육 및 무상교육, 모자보건센터를 통한 무료 건강지원, 심리지원 등 포괄적이며 장기적인 지원을 하고 있다(정은희·최유석, 2013). 따라서 첫째 자녀 양육 단계에서부터 장기적이고 포괄적인 지원을 제공함으로써 부모들이 양육의 어려움, 양육불안 없이 후속 자녀를 출산할 수 있는 사회구조적 지원체계가 필요하다.

후속출산의도를 예측하는 주요 요인이 연령대별로 다르게 나타났다는 점에서, 연령별 맞춤 접근의 중요성이 강조된다. 20대 집단에서는 아버지양육참여, 양육행동, 성격(그릿), 지역사회양육환경만족도가 주요 요인으로 나타났고, 30대 집단에서는 어머니연령, 자아존중감, 어머니취업상태, 성격(그릿)이 나타났으며 40대 집단에서는 학력, 총자녀수, 주관적사회경제적지위가 주요 요인으로 나타났다. 이는 출산과 양육을 고려한 정책이 일률적일 수 없으며, 후속출산의도를 장려하기 위한 정책과 프로그램 설계 시 연령별로 특화된 접근이 필요함을 시사한다. 예를 들어, 20대 어머니들은 경제적 어려움과 함께 자녀 양육에 대한 경험 부족, 혼자 자녀를 양육할 가능성이 높다. 따라서 이들을 대상으로 한 개입은 공동양육, 양육스트레스 감소, 경제적 부담 및 보육 부담을 경감할 수 있는 정보와 지원을 포함해야 하고 아버지의 양육참여를 위한 일-가정 양립지원 정책이 뒷받침되어야 한다. 30대 어머니들은 연령, 어머니취업상태, 자아존중감, 성격(그릿)이 주요 요인으로 나타났으며 어머니의 연령을 고려할 때 자녀 출산 후 경력 단절을 경험할 수 있다. 따라서 이들을 대상으로

한 정책은 육아휴직제도의 활성화와 함께 신뢰할 수 있는 무상 보육 및 무상 교육 서비스 제공, 취업과 양육 지원 정책이 고려되어야 할 것이다. 마지막으로 40대 어머니들은 학력, 총자녀수, 주관적사회경제적지위가 주요 요인으로 나타났으며 주관적사회경제적지위를 중심으로 양육 요인들이 연결되어 있다는 점을 고려할 때 자녀 양육의 경제적 부담을 낮출 수 있는 자녀돌봄 및 교육지원, 의료지원, 자녀의 성장까지 지속적으로 도움을 받을 수 있는 사회적 지지체계 지원이 필요할 것으로 사료된다. 종합하면 후속출산의도에 영향을 미치는 요인은 개인요인, 양육요인, 가구요인, 환경요인 등 다양한 측면에서 나타난다. 따라서 종합적인 지원 시스템을 구축하여 어머니들에게 다양한 측면에서 지원을 제공해야 할 필요가 있다. 본 연구 결과를 토대로 예측요인을 예측하고 지원하는 데 있어 중요한 자료와 방향성을 제시하였으며 각 연령대별로 맞춤형 정책과 프로그램을 개발하고, 어머니들의 정신건강과 양육지원을 강화함으로써 후속출산을 촉진하는 데 실증적인 자료로 활용될 수 있을 것이다.

참고문헌

- 강유선·곽승주(2018). 어머니의 가족 상호작용과 육아지원기관 충분성 인식이 후속출산계획에 미치는 영향. 유아교육학논집, 22(3), 299-320.
- 강유선·김찬우(2018). 후속출산에 영향을 미치는 요인 탐색. 한국영유아보육학, 110, 59-84.
- 강현구·심도현·박윤현·이순형(2014). 지역경제적 특성 및 생애주기적 특성이 어머니의 후속출산의도에 미치는 영향. 육아정책연구, 8(2), 175-208.
- 공선영(2006). 경제위기 이후 인구학적 특성 및 사회경제적 특성이 기혼여성의 출산에 미친 영향. 보건과 사회과학, 19(1), 119-149.
- 김미나·이소은(2018). 보육서비스 만족도와 양육효능감이 후속출산의도에 미치는 영향 36개월 이하 한 자녀 어머니를 대상으로. 유아교육·보육복지연구, 22(3), 123-144.
- 김은경·박신애(2019). 후속 자녀 출산 계획에 영향을 미치는 부모 특성 및 지역사회 환경요인에 관한 연구. 예술인문사회 융합 멀티미디어 논문지, 9(7), 443-454.
- 김종훈·양소영·성지현(2013). 아버지의 자녀양육참여와 사회적 지원이 후속출산계획에 미치는 영향: 어머니의 양육스트레스 매개를 중심으로. 아동학회지, 34(1), 87-102.
- 김지영·고선(2021). 출신국 출산율과 이주여성의 자녀 출산. 다문화사회연구, 14(3), 109-143.
- 김하나(2008). 여성의 결혼, 출산자녀 수 결정에 영향을 주는 사회경제적 요인분석. 2008년 제2회 여성가족패널 학술대회 자료집, 391-416.

- 김희명·황매향(2015). 한국판 아동용 끈기(Grit) 척도의 타당화. *교육논총*, 35(3), 63-74.
- 류기철·박영화(2009). 한국여성의 출산을 변화와 출산간격 영향요인. *한국인구학*, 32(1), 1-23.
- 박종서·임지영·김은정·변수정·이소영·장인수·조성호·최선영·이혜정·송지은(2021). 2021년도 가족과 출산 조사: (구)전국 출산력 및 가족보건복지 실태조사. 한국보건사회연구원.
- 배광일·김경신(2011). 생태학적 변인이 출산의도에 미치는 영향 연구. *한국가족관계학회지*, 16(2), 123-149.
- 백지희(2021). 한 자녀 부모의 후속출산 관련 요인에 관한 체계적 문헌고찰. *학습자중심교과교육연구*, 21(6), 65-81.
- 서미정(2011). 부모의 심리사회적 특성이 후속 출산계획에 미치는 영향. *육아정책연구*, 5(1), 127-148.
- 서정연·김한곤(2015). 출산자녀수별 기혼여성의 출산의지에 영향을 미치는 요인. *한국인구학*, 38(3), 1-24.
- 송영주·이주옥·김춘경(2011). 어머니의 취업여부에 따른 후속출산계획 관련 심리사회적 요인 탐색. *열린유아교육연구*, 16(3), 185-206.
- 송유미·이선자(2016). Q방법을 활용한 취업모의 둘째 자녀 출산기피 요인의 유형화. *한국콘텐츠학회논문지*, 16(2), 714-728.
- 엄연용·송원영(2022). 아동기 자녀를 둔 부모의 양육불안 경험에 대한 질적 연구. *한국심리학회지 상담 및 심리치료*, 34(4), 1229-1251.
- 이미란(2010). 기혼 여성이 처한 환경적 요인이 자녀에 대한 가치관과 출산 행태에 미치는 영향. *한국가족관계학회지*, 15(1), 99-121.
- 이영환(2014). 어머니 결혼만족도와 우울의 종단관계 연구. *한국영유아보육학*, 87, 163-185.
- 이정원(2009). 유자녀 취업여성의 후속출산의도 결정 경로 분석: 가치관적 요인을 중심으로. *사회복지연구*, 40(1), 323-351.
- 이진화·김경은·임원신(2015). 한자녀 가정의 후속출산계획에 영향을 미치는 변인에 관한 연구: 가족 환경적 특성과 유아변인 간의 구조적 관계를 중심으로. *육아정책연구*, 9(2), 1-22.
- 이진화·민정원(2015). 어머니의 개인 심리적 특성과 가정환경자극 수준이 후속출산계획에 미치는 영향. *생태유아교육연구*, 14(1), 79-100.
- 이진화·조하나·최혜윤(2012). 모의 후속출산계획에서 한자녀 취업모의 일반적 변인에 따른 차이와 부모됨이 미치는 영향. *육아정책연구*, 6(1), 125-141.
- 이형민(2012). 후속출산계획의 영향요인에 관한 연구 한국의 영아기 자녀를 둔 가구를 중심으로 한국의 영아기 자녀를 둔 가구를 중심으로. *한국영유아보육학*, 70, 1-31.
- 임양미(2015). 의사결정나무분석을 적용한 비취업모와 취업모의 후속출산계획 예측요인 탐

- 색. 한국가정과교육학회지, 27(4), 155-172.
- 임영일·강현영·서주희·김경미(2024). 인구동향조사 출생·사망통계. 통계청.
- 임현주(2013). 인구, 사회학적 변인, 경제변인, 양육, 출산변인에 따른 후속출산계획 연구. 유아교육연구, 33(3), 35-58.
- 임현주·이대균(2013). 부부특성 및 사회적 지원과 어머니의 심리적 특성이 후속출산에 미치는 효과 모의 취업과 후속출산계획 여부에 따른 다중집단분석. 유아교육연구, 33(2), 121-144.
- 임현주·이대균·최향준(2011). 부부특성, 출산관련 정서, 어머니의 정서특성이 후속출산계획에 미치는 변인탐색: 구조방정식을 중심으로. 유아교육연구, 31(5), 183-202.
- 임효진·박휴용·장경은·이지은(2015). 어머니의 심리적, 관계적 변인과 양육스트레스, 후속출산의도의 구조적 관계. 열린유아교육연구, 20(2), 297-318.
- 정미라·홍소영·박수경(2013). 첫 자녀를 둔 가구의 후속출산에 영향을 미치는 변인 탐색. 유아교육연구, 33(1), 259-277.
- 정은희·최유석(2013). 기혼여성의 둘째자녀 출산계획 및 출산과 관련된 요인. 보건사회연구, 33(1), 5-34.
- 조병구·조운영·김정호(2007). 출산지원정책의 타당성 및 지원효과 분석. 서울: 한국개발연구원.
- 최분희·이창원(2017). 우리나라 광역지방자치단체 출산장려 프로그램의 효과성에 관한 실증적 연구. 사회적경제와 정책연구, 7(1), 187-213.
- 최여진·조희선(2018). 유자녀 임신부의 후속출산결정경험. 가정과삶의질연구, 36(1), 25-43.
- 최향준(2015). 취업모의 심리변인, 사회적 지원변인, 경제변인 및 자녀(영아)의 기질변인이 후속출산계획에 미치는 영향. 한국영유아보육학, 93, 89-114.
- 하민경(2020). 어머니가 인식하는 양육스트레스와 아버지 양육참여가 후속출산계획 자녀수에 미치는 영향. 학습자중심교과교육연구, 20(3), 1139-1155.
- 현정환(2019a). 부모의 후속출산에 영향을 미치는 요인에 대한 탐색. 한국영유아보육학, 114, 25-43.
- 현정환(2019b). 영유아 자녀를 둔 부모의 비모성적 양육관이 후속출산의지에 미치는 영향에 대한 연구. 한국영유아보육학, 118, 101-122.
- Adsera, A. (2005). Vanishing children from high unemployment to low fertility in developed countries. *American Economic Review*, 95(2), 189-193.
- Becker, G. S. (1992). Fertility and the economy. *Journal of Population Economics*, 5(3), 185-201.
- Belsky, J. (1984). The determinants of parenting: A process model. *Child*

Development, 55(1), 83-96.

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B(Methodological)*, 20(2), 215-232.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087-1101.
- Gietel-Basten, S., Sobotka, T., Zeman, K., Lutz, W., Butz, W. P., & KC, S. (2014). *Future fertility in low fertility countries. World Population and Human Capital in the Twenty-First Century: An Overview*. Oxford.
- Kohler, H. P. (2001). *Parity Progression Ratios with Continued Fertility Postponement: Assessing the Implications of Delayed Childbearing for Fertility in Sweden, the Netherlands and Spain. International Perspectives on Low Fertility: Trends, Theories and Policies*. Tokyo: Mauldin.
- Lesthaeghe, R., & Willems, P. (1999). Is low fertility a temporary phenomenon in the European Union. *Population and Development Review*, 25(2), 211-228.
- Morgan, S. P., & Taylor, M. G. (2006). Low fertility at the turn of the twenty-first century. *Annual Review of Sociology*, 32(1), 375-399.
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines: A tutorial. *Frontiers in Neurorobotics*, 7, 1-21.
- Sebastian, R., & Vahid, M. (2017). *Python Machine Learning - Second Edition*. 박해선(역) (2019). 머신러닝 교과서 with 파이썬, 사이킷런, 텐서플로. 서울:길벗.
- Syarif, I., Zaluska, E., Prugel-Bennett, A., & Wills, G. (2012). Application of bagging, boosting and stacking to intrusion detection. *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 7376(8), 593-602.

- 논문접수: 11월 5일 / 수정본 접수 12월 4일 / 게재 승인 12월 23일
- 교신저자: 엄연용, 건양대학교 심리상담치료학과 겸임교수, eyy83@hanmail.net

Abstract

Comparison of Predictive Models for Subsequent Childbirth Intentions by Age Group Using Machine Learning and Network Analysis

Yun-Yong Eom

This study aimed to explore the factors predicting the intention of subsequent childbirth and to identify the interactions and importance of these predictive factors through network analysis. For this purpose, data from 1,922 mothers in 2022 Korean Early Childhood education and care panels were analyzed. Relationships between 20 predictive factors, including individual, parenting, household, and environmental factors, and intention for subsequent childbirth were established, and Random Forest, Gradient Boosting, Logistic Regression, and Stacking models were applied. The results showed that the performance of the four models ranged from F1 .60 to .87. In the 20s age group, parenting stress, personality (grit), and depression were the main factors predicting subsequent childbirth intentions. The main factors in the 30s group, were mother's age, marital conflict, and parenting stress, whereas in the 40s group, the main factors were social support, total number of children, and education level. This study is significant in that it explored the main factors that can predict the intention of a subsequent childbirth and provided empirical data and a foundation to support subsequent childbirth through the interrelationships among these factors.

Keyword: subsequent childbirth, RandomForest, GradientBoosting, Stacking, NetworkAnalysis