

## 맞벌이 기혼여성의 출산의사 예측요인 탐색: 랜덤포레스트 기법의 적용

안리라<sup>1)</sup>

### 요약

이 논문은 맞벌이 기혼여성의 출산의사 예측요인 탐색을 위해 랜덤포레스트 기법을 적용하여 여성가족패널 4차(2012년)~9차(2022년) 자료를 분석했다. 분석에는 49세 이하 맞벌이 기혼여성 데이터 3,314건이 포함되었다. 첫째, 예측모델에 대한 변수별 중요도는 성평등 요인, 경제적 요인, 인식·문화적 요인 순이었다. 둘째, 성평등 요인인 남편의 가사노동 부담률과 아내의 경제적 협상력이 부부간 대등할 때 출산의사가 가장 높았다. 셋째, 가구소득의 증가는 출산·육아에 필요하다고 생각되는 적정 소득 구간까지 출산의사를 증가시켰으나 그 이후에는 출산의사가 감소했다. 이러한 결과는 저출생 위기 해소를 위해 성평등 제고가 중요한 과제이며, 변수 간 비선형 관계를 고려한 정책설계가 필요하다는 것을 시사한다.

주제어: 출산의사, 맞벌이 기혼여성, 성평등, 랜덤포레스트

## I. 서론

### 1. 문제제기

한국의 지속된 초저출생 현상은 국내문제를 넘어 국제적 관심을 받은 지 오래다. OECD에 따르면, 한국은 회원국 중 가장 낮은 합계출산율(0.78명, 2022년 기준)을 기록했는데 이는 OECD 회원국의 평균 출산율(1.51명)의 절반 수준에 불과하다(OECD, 2024). 한국의 합계출산율은 1983년 처음 대체출산율 이하로 떨어졌으며, 이후 감소 추세가 계속되었다. 2018년에는 합계출산율이 1명 이하로 떨어졌고, 2024년 합계출

1) 고려대학교 사회학 박사

산율이 0.75명까지 감소했다. 이에 저출생 위기를 해결하기 위해 여러 학술적, 정책적 노력이 기울여졌으나 출산율이 반등할 기미는 크게 보이지 않는 상황이다.

저출생에 대응하기 위해 2006년부터 5년마다 수립하고 있는 저출산고령사회기본 계획에 대한 평가에도 명암이 존재한다. 1차~4차까지의 기본계획을 통해 정부는 임신·출산·양육에 대한 사회적 책임 강화, 일·가정 양립 지원정책 확대, 일자리·주거지원 확대 등을 추진했다. 하지만 출산율 중심으로 목표를 설정하고, 여성을 재생산권 주체로서 인정하지 않는 등 저출생 문제를 여전히 인구통제 전략으로 접근하는 한계가 있었으며(배은경, 2021), 저출생 위기 해결의 기반이 되는 성평등 증진을 위한 사회구조적 변화보다는 현금성 급여 중심의 지원에 집중하는 등 경제적 요인이 과도하게 부각되는 경향을 보였다. 더불어 불충분한 공공지출 규모, 일·가정 양립 지원제도의 실효성 부재, 고용·주거 등 임신·출산과 관련된 종합적인 대응 미흡 등 여러 한계가 지적되었다(박선권, 2022).

학계에서도 출산에 영향을 미치는 요인을 다루는 여러 연구가 진행되었는데 주로 가구소득, 육아·교육비 부담, 주거형태 등과 같은 경제적 요인에 관심을 두고 있는 연구들이 주를 이루었다. 이는 양육비 등 경제적 부담이 청년세대가 출산을 기피하는 주된 이유로 지적되는 현실을 비롯해 청년세대 출산 지원정책이 고용·주거 등 경제적 양극화 해소에 중점을 두고 있다는 점과 무관하지 않다. 출산과 육아를 위해 부부 간 공평한 가사분담, 일·가정 양립 등 가정과 사회에서의 성평등이 필수적인 요소임에도 불구하고 성평등 요인에 대한 정책 및 학술영역에서의 관심이 상대적으로 부족했다. 그리고 기존 연구들은 출산에 영향을 미치는 요인을 선형관계 중심으로 분석해왔는데 향후 실효성 있는 출산·육아 지원정책 마련을 위해서는 변수 간 비선형 관계에 주목할 필요가 있다. 특정 조건에 따라 각 변수가 어떻게 다른 영향을 미치는지 분석함으로써 맞춤형 정책설계가 가능하다.

이에 본 연구는 랜덤포레스트 기법을 적용하여 출산의사에 영향을 미치는 요인을 탐색한다. 출산의사를 예측하는 주요 설명변수로 경제적 요인, 성평등 요인, 인식·문화적 요인, 제도적 요인 등을 함께 다룸으로써 어떤 요인이 출산의사를 예측하는데 중요도를 갖는지 분석한다. 또한 랜덤포레스트 기법은 비정형 데이터 분석 및 비선형 관계 분석에 적합하다. 이 연구에서는 랜덤포레스트 분석결과를 시각화하는 부분의 존성 도표 분석을 활용하여 각 설명변수와 출산의사 간 비선형 관계를 분석한다.

## 2. 이론적 배경

### 가. 합리적 선택이론

경제학적 관점에서 출산을 변화를 설명하는 대표적인 이론이다. 1960년대 베커(Gary Becker)에 의해 본격적으로 이론화되었다. 합리적 선택이론에서는 각 가정이 자녀를 하나의 소비재이자 투자재로 간주하며 비용과 편익, 수요와 공급의 관점에서 출산을 결정한다고 설명한다. 여기에는 부모의 소득, 자녀양육 및 교육 비용, 부모의 시간이 갖는 가치, 사회정책 등 다양한 요인이 작용한다(Becker, 1992). 여성의 노동 시장 참여가 증가하면서 여성이 육아에 투자하는 시간의 경제적 가치도 커지기 때문에 자녀를 양육하는 데 따르는 기회비용이 커져서 출산율이 감소한다고 본다. 그리고 자녀 양육의 질이 중요해지면서 인적자본에 대한 투자비용이 높아지는 것, 연금제도 등 사회보장제도가 발전하면서 노후 대비를 위한 자녀 양육의 필요성이 감소하는 것 등도 출산율 감소의 원인으로 지적된다.

후속연구에서는 출산과 관련한 비용 산출과 의사결정에 부부간 가사노동분담, 정책적 지원 등 다양한 요인이 고려될 수 있음이 밝혀지기도 했다. 뉘케와 킨더만(Doepke & Kindermann, 2019)은 베커의 경제적 관점을 발전시켜 경제적 비용 및 편익 계산에 따른 출산결정이 부부간 협상의 결과임을 강조했다. 이들은 남성의 출산의사보다 여성의 출산의사가 출산에 더 결정적인 영향을 미치며, 남성이 육아를 더 많이 담당하는 국가에서 출산율이 상대적으로 높고, 정부의 육아 지원정책이 어느 정도로 제공되는가에 따라 부부간 출산협상 결과가 달라질 수 있다고 보았다

합리적 선택이론에 대해서는 비판도 존재한다. 합리적 선택이론의 경제적 가정과 달리 실제 출산 결정은 사회적, 문화적 요인에 의해서 복합적으로 영향을 받으며, 현실에서는 자녀를 보는 관점 역시 경제적·도구적 관점을 넘어 심리적 만족, 가족 내 유대감 등이 작용한다는 것이다(Robinson, 1997).

### 나. 제2차 인구전환(the second demographic transition) 이론

1970년대 이후 서구 국가들에서 나타난 출산율 감소는 기존의 제1차 인구전환 모델로는 설명하기 어려웠다. 출산율과 사망률이 감소하여 인구가 대체출산율인 2.1명 수준에서 안정화될 것이라는 예측과 달리 1970년대부터 출산율이 계속해서 감소했고, 이를 설명할 새로운 이론적 모델로 제2차 인구전환이론이 등장하였으며, 1980년대 론 레스따게(Ron Lesthaeghe)에 의해 본격적으로 이론화되었다.

제2차 인구전환이론은 출산율의 감소가 경제적 요인보다는 가치관 변화와 같은 문화적 요인에서 기인한다고 강조한다. 제2차 인구전환이론은 매슬로우 욕구이론에 영향을 받았으며, 인간은 먹고사는 기본적인 생존 욕구가 충족되면 개인의 자아실현이나 성장과 같은 더 높은 차원의 욕구를 추구하기 때문에 가치관 변화로 인한 출산율 감소를 필연적 현상으로 본다(Zaidi & Morgan, 2017). 기존에는 결혼과 출산을 당연하게 여겨왔으나, 점차 자아실현과 삶의 질을 우선시하고, 개인주의적 가치, 자율성, 다양성이 확대되면서 결혼과 출산의 전통적인 연결이 약해지며, 동거, 비혼 출산, 이혼이 증가하는 등 가족구조가 변화하여 출산율도 지속적으로 감소하게 된다. 경제성장, 여성의 교육수준 향상, 가족구조 변화, 생애과정의 다양화, 유연화 등이 이러한 문화적 변동을 추동한다.

그러나 제2차 인구전환이론에 대해서는 문화적 변화를 지나치게 강조한다는 비판, 결혼 감소, 동거 증가, 출산율 감소 등의 현상이 장기적인 문화적 변화가 아니라 경제적 불안정성으로 인한 일시적 현상이라는 비판 등이 제기되었다(Johnson, 2024). 또한, 제2차 인구전환이론은 단선적 발전 모델을 전제하고 있는데 이는 국가 간, 계층 간 차이 등을 간과하는 한계가 있으며, 경제적 요인, 성불평등, 세계 경제 환경 변화 등을 고려하지 못한다는 한계가 있다고 지적된다(Zaidi & Morgan, 2017).

이에 대해 제2차 인구전환이론을 지지하는 학자들은 동아시아에서도 결혼연령 상승, 비혼 증가 등의 현상이 유사하게 확인되고 있으며, 동거, 비혼 증가 및 출산 연기 등의 패턴이 경제적으로 불안정한 계층의 사람들만이 아니라 고학력자 등 경제적으로 안정된 계층에서도 광범위하게 일어나는 현상이라는 점에서 이것이 경제적 불안정성으로 인한 일시적인 현상보다는 가치관 변화 등으로 인한 장기적 변동으로 이해해야 한다는 등의 반론을 제기한다(Lesthaeghe, 2014).

#### 다. 다중균형(multiple equilibrium) 이론

다중균형이론은 가족구조의 변화가 하나의 단순한 방향이 아니라 여러 개의 균형 상태를 거치며 변화하는 과정이라 본다. 남성생계부양자 중심의 전통적인 가족구조에서 안정적인 균형상태에 있다가, 여성의 교육수준이 높아지고 노동시장 참여가 확대되면서 기존의 균형상태가 불안정해지면 결혼율과 출산율이 낮아지고 이혼이 증가하는 등 과도기적 불안이 나타나지만, 시간이 지나면서 가정과 사회의 성평등 수준이 높아지고, 여성의 노동시장 참여를 지원하는 각종 출산·육아지원, 일·가정 양립 지원 정책 등이 제공되면 결혼율과 출산율이 다시 증가하는 U자형 곡선 형태를 보인다(에

스핑-앤더슨, 2014; Esping-Andersen & Billari, 2015).

에스핑-앤더슨은 성평등의 계층화를 분석했다. 북유럽 국가에서 여성의 학력과 출산율 사이의 정적 관계가 발견되었으며, 특히 교육수준, 선호, 취향이 비슷한 고학력 동질혼 부부의 출산율이 증가했다(에스핑-앤더슨, 2014). 부부가 임금 등의 자원을 동등하게 보유하면 공평하게 가사·육아를 분담할 가능성이 높고, 이것이 출산 결정에 긍정적 영향을 미친다. 에스핑-앤더슨은 성평등이 계층화되어 있고, 모든 계층에 성평등이 확대되면 출산율이 상승할 것이며, 이를 위해 국가는 사회투자의 관점에서 여성의 노동시장 참여를 독려하고 아동교육과 복지를 확충하는 지원제도를 적극적으로 추진해야 한다고 강조했다.

다중균형모델에 대한 비판도 존재한다. 금융위기 등 경제적 불확실성이 큰 상황에서는 높은 수준의 성평등과 복지시스템 발전에도 불구하고 출산율이 감소하였으며(Vignoli, Guetto & Bazzani et al., 2020), COVID-19 팬데믹 시기 돌봄공백 해결을 위해 가사·육아 등 전통적 성역할이 다시 여성에게 집중적으로 요구되면서 에스핑-앤더슨 이론의 실효성이 비판 받았다(Morero-Minguez & Ortega-Gaspar, 2022). 더불어 국제경제의 불평등으로 인한 국가 간 사회투자 격차, 노동시장의 구조적 불평등에 대한 고려가 필요하다는 지적도 제기되었다(Ferragina, 2023).

이처럼 출산율 감소를 설명하는 다양한 이론이 변화하면서 발전되어 왔고, 기존 이론이 완전히 폐기된 것이 아니라 반론을 거듭하며 경합하고 있다. 이들 이론이 복합적으로 영향을 미친다고 보는 것이 적절해 보이며, 각 국가 또는 사회의 역사적, 문화적 배경에 따라, 그리고 시기적 상황에 따라 우세하게 영향을 미치는 이론적 관점과 요인이 있을 수 있다. 따라서 현재 한국의 출산율 감소를 설명하는 데 경제적, 문화적, 성평등 요인 중 어떤 요인이 더 영향력 있는지에 대한 분석이 중요하다.

### 3. 선행연구 고찰

출산의사에 영향을 미치는 요인에 대한 국내연구는 크게 개인 및 가정요인, 경제적 요인, 정책적 요인을 다룬 연구로 구분할 수 있다.

첫째, 개인 및 가정요인이다. 가족 및 성역할 가치관, 부부간 가사노동 분담, 결혼 만족도 등을 다룬다. 남편의 가사노동분담률, 결혼생활 만족도 등이 높을수록 후속 출산의사가 높아진다(남국현, 2022; 선민정·조상미, 2019). 자녀를 노후 부양 등 도구적 수단으로 인식하는 가치관을 갖고 있는 경우보다 부모가 된다는 것 자체를 가치 있는 일로 여기는 가치관이 후속 출산에 긍정적 영향을 미친다(이완·채재은, 2017).

자녀 유무에 따라 가치관이 출산에 미치는 영향이 상이할 수 있다. 자녀가 없는 기혼 여성은 자녀의 필요성, 성역할 규범에 동조할수록 출산의사가 높았고, 자녀가 있는 기혼여성은 혼외임신, 입양 등에 개방적 태도를 보일수록, 성역할 가치관은 부부간 가사노동 분담이 공평하다고 생각할수록 출산의사가 높았다(임병인·서혜림, 2021).

둘째, 경제적 요인이다. 고용형태에 따른 부부의 직업안정성, 가구소득, 주택 소유 여부 등에 따라 출산의사가 상이하게 나타났다. 남성은 정규직인 경우 자녀의 수가 많았으나(이진권, 2011; 정병삼, 2023), 여성은 비정규직인 경우 자녀 수가 더 많았다(이진권, 2011). 자녀의 출생순서에 따라 고용형태의 영향이 다르게 나타났는데, 첫째 자녀는 정규직, 비정규직, 자영업 모두 전업주부와 출산가능성에 큰 차이가 없었으나, 정규직, 비정규직 여성은 전업주부에 비해 둘째 자녀 출산 가능성이 낮았다(김현동, 2021). 16개 광역자치단체의 출산율을 분석한 연구에서도 여성의 경제활동 참여, 상용직 근무 등은 출산에 부정적 영향을 미쳤다(탁아림·이찬영, 2024). 경제적 비용과 관련해서 사교육비와 전세가격이 증가하면 출생률은 감소했다(박정희, 2024).

셋째, 정책적 요인이다. 주로 출산휴가, 육아휴직 등 가족친화제도에 주목한다. 자녀 출산 및 양육지원, 유연근무제, 가족친화적 조직문화 등 직장 내 가족친화적 문화와 제도는 여성의 출산의사에 긍정적인 영향을 미친다(송민영, 2018; 유계숙, 2010; 최지훈·안선희, 2017). 한편, 한국노동패널조사를 분석한 연구에서는 가족친화제도가 첫째자녀 출산에 긍정적 영향을 미쳤으나, 둘째자녀 출산부터는 그 효과가 약화된다는 것이 분석되었다(주익현, 2023).

출산의사에 영향을 미치는 요인을 다룬 기존 연구들은 주로 변수 간 선형관계에 주목했다. 유사한 변수에 대해 상반된 연구결과가 존재하는데 이는 변수들이 비선형 관계에 있기 때문일 수 있다. 이에 본 연구에서는 비정형 데이터 분석에 강점을 가진 랜덤포레스트 분석을 통해 출산의사 예측요인을 탐색하고 부분의존성 도표(PDP) 분석을 통해 설명변수와 반응변수 간의 관계를 구체적으로 밝히고자 한다.

## II. 연구방법

### 1. 분석자료 및 분석대상

본 연구는 한국여성정책연구원의 여성가족패널 4차(2012년)~9차(2022년) 자료를 활용하였다. 여성가족패널은 2007년 시작된 패널조사로 19세 이상 64세 이하 여성

을 표본으로 하여 여성의 경제활동, 가족실태, 가족가치관, 건강 등에 관한 사항을 다룬다. 본 논문은 가정에서의 성평등 수준을 측정할 수 있는 변수로서 부부의 가사 및 돌봄노동시간을 주요한 설명변수로 사용하고 있기 때문에 이들 변수가 조사되기 시작한 4차(2012년) 자료부터 분석에 포함하였다.

본 연구는 각 시점의 관측치를 횡단면 데이터로 모두 결합한 데이터를 사용했다. 랜덤포레스트는 비모수적 머신러닝 모델로서 전통적인 회귀분석에서 전제하는 엄격한 데이터 분포 가정에서 자유로우며, 개별 시점의 데이터를 독립된 관측값으로 인식하여 학습한다. 이 연구는 시간 흐름에 따른 변화 추적보다 구조적인 예측 패턴을 규명하는 것에 목적이 있기 때문에 예측 정확도 향상을 위해 4차~9차 자료를 결합하여 사용했다. 각 조사연도 기준으로 49세 이하의 맞벌이 기혼여성 데이터를 추출하여 분석하였다. 분석에 사용된 변수 중 결측치를 포함한 데이터는 제외하였으며, 최종적으로 분석에 사용된 유효표본은 총 3,134건이다.

## 2. 측정도구

### 가. 반응변수

본 연구의 반응변수는 출산의사가 있는지 여부이다. 출산의사는 “자녀를 가질 계획이 있으십니까?”라는 문항으로 조사되었으며, “1) 있다, 2) 없다, 3) 모르겠다”로 측정되었다. 본 연구에서는 기존 연구의 측정방법을 참고하여 ‘모르겠다’라고 응답한 경우는 출산계획이 없거나 출산에 소극적인 집단으로 보아 출산의사가 없는 집단에 포함하여 해당 변수를 ‘출산의사 있음/없음’의 이항변수로 재코딩하였다(강유진, 2020; 김정석, 2007). 전체 데이터 중 출산의사가 있는 경우는 약 4.2%에 해당했다.

### 나. 설명변수

- ① 성평등 요인: 부부간 성평등 수준을 파악할 수 있는 변수로 성역할에서의 평등과 경제적 평등 두 가지로 측정하였다. 첫째, 성역할에서의 평등은 남편의 가사노동 분담률로 측정하였다. 평일·주말을 포함한 일주일간 부부의 전체 가사·돌봄노동시간 중 남편의 가사·돌봄노동시간이 차지하는 비율이다. 둘째, 경제적 평등은 아내의 상대적인 경제적 협상력으로 측정하였다. 여성의 인적자본이 향상

되고 경제활동 참여가 증가하면서 부부의 생산성이 유사한 수준으로 수렴됨에 따라 각자가 임금 등과 같은 외부자원을 얼마나 확보할 수 있는지가 가사·육아 역할분담 협상에서 중요한 역할을 하게 되었다(에스핑-앤더슨, 2014). 부부간 경제적 협상력은 본 논문의 분석대상인 여성의 경제적 협상력을 기준으로 측정하였으며, 폴락(Pollak, 2005)의 변수측정을 참고하여 아내의 시간당 임금을 부부의 시간당 임금을 합산한 값으로 나눈 비율로 측정되었다.

- ② 경제적 요인: 가구소득은 지난 1년간 가구 총소득으로 측정하였다.
- ③ 인식·문화적 요인: 개인주의 가치관은 개인주의 및 자율성을 중시하는 가치관에 동의하는 정도를 나타내며, 성역할 가치관은 부부간 평등한 역할분담에 동의하는 정도를 의미한다. 가족 가치관은 개방적인 결혼·가족 가치관에 동의하는 정도를 측정했다. 자세한 측정문항은 <표 1>에 기술하였다. 각 가치관을 구성하는 문항들의 크론바흐 알파(cronbach's alpha) 값은 개인주의 가치관 0.522, 성역할 가치관 0.281, 가족 가치관 0.527로 낮은 편인데, 이는 각 문항들이 다차원적 하위요인을 포함하고 있어 문항 간 상관관계가 높지 않기 때문으로 보인다. 다만, 본 논문에서는 개인의 가치관이 다른 경제적 요인, 성평등 요인, 제도적 요인 등과 비교했을 때 출산 의사결정을 예측하는 데 설명력을 가지는지 밝히기 위한 탐색적 목적으로 활용하였기 때문에 넓은 범위에서 개념적 포괄성을 고려하여 설계하였다. 다만 결과 해석 시 이러한 한계를 감안해야 하며, 후속연구에서는 각 측정문항의 하위차원별 분석이 필요하다.
- ④ 제도적 요인: 직장에서의 출산·육아 관련 가족친화제도 제공 여부로 측정하였다. 근무하고 있는 직장에서 출산휴가, 육아휴직, 보육비 지원, 직장보육시설, 육아기 근로시간 단축제 등 가족친화제도를 최소 1가지 이상 제공하는지 여부이다. 이들 제도 중 1가지라도 제공하면 제공됨, 전혀 제공하지 않거나 제공하는지 모르는 경우는 제공되지 않음으로 재코딩하였다.
- ⑤ 개인적 요인: 연령, 학력, 자녀 수는 본 논문의 분석대상인 49세 이하 기혼여성의 연령, 학력, 자녀 수를 의미한다. 학력은 무학, 초졸, 중졸, 고졸, 전문대졸, 4년제 대학졸, 대학원졸 이상으로 측정하였다.

〈표 1〉 설명변수 문항 및 척도

변수	문항	
성평등 요인	남편의 가사노동 부담률	부부의 일주일간 전체 가사·돌봄노동시간 중 남편의 가사·돌봄노동시간이 차지하는 비율
	아내의 경제적 협상력	아내의 시간당 임금률을 부부의 시간당 임금률을 합산한 값으로 나눈 비율
경제적 요인	가구소득	지난 1년간 가구 총소득
인식· 문화적 요인	개인주의 가치관	“결혼보다 나 자신의 성취가 더 중요하다”, “결혼을 하면 나 자신의 생활이 구속받는다” 문항 평균점수 <역코딩> (1=전혀 그렇지 않다, 2=별로 그렇지 않다, 3=조금 그렇다, 4=매우 그렇다)
	성역할 가치관	“남성은 직장을 가지고 여성은 가정을 돌보는 것이 이상적이다”, “취학 전 자녀를 둔 주부가 일을 하면 자녀에게 부정적인 영향을 줄 것이다”, “맞벌이 부부는 집안일도 공평히 분담해야 한다 <역코딩>” 문항 평균점수 (1=매우 그렇다, 2=조금 그렇다, 3=별로 그렇지 않다, 4=매우 그렇다)
	가족 가치관	“결혼은 반드시 해야한다”, “자녀는 반드시 있어야 한다”, 자녀가 있어도 이혼할 수 있다<역코딩>”, “결혼을 전제로 하지 않아도 동거할 수 있다<역코딩>”, “결혼하지 않아도 아이를 낳아 기를 수 있다<역코딩>” 문항 평균점수 (1=매우 그렇다, 2=조금 그렇다, 3=별로 그렇지 않다, 4=매우 그렇다)
제도적 요인	가족친화제도	출산휴가, 육아휴직, 보육비지원, 직장보육시설, 육아기근로시간단축제 중 최소 1가지 이상 제공하는지 여부 (1=제공된다, 0=제공되지 않거나 모른다)
개인적 요인	연령	응답자 연령
	학력	응답자 학력 (1=무학, 2=초졸, 3=중졸, 4=고졸, 5=전문대졸, 6=대졸(4년제), 7=대학원 이상)
	자녀 수	응답자 자녀 수

### 3. 분석방법

랜덤포레스트는 머신러닝의 앙상블 기법 중 하나로 여러 개의 의사결정트리를 조합하여 예측을 수행한다. 각각의 트리는 배깅을 통해 무작위로 선택된 데이터를 사용하여 독립적으로 학습되기 때문에 트리 간 상관관계를 줄이고 일부 강한 변수가 모든 트리를 지배하는 문제를 방지한다(Strobl, Malley & Tutz, 2009).

랜덤포레스트는 분류와 회귀 예측이 가능하다. 본 논문에서처럼 출산을 할지 여부를 결정하는 경우는 분류에 해당하며, 회귀는 상품가격 등 연속된 숫자 값을 예측하는 경우이다. 분류는 다수의 트리 결과를 다수결 방식으로 종합하여 최종 예측하며, 회귀는 여러 트리의 평균값을 산출하여 최종 예측한다. 랜덤포레스트는 변수간 복잡한 상호작용을 자동으로 고려할 수 있어 비선형 데이터에도 강한 성능을 보이며 (Breiman, 2001; Strobl, Malley & Tutz, 2009), 이 때문에 변수 간의 외생성, 오차항의 분포, 등분산성, 다중공선성 등 전통적인 회귀분석에서 충족해야 하는 가정에서 자유롭다(이은정·송영수·김지하 외, 2020). 랜덤포레스트는 변수가 많고 관측치가 적은 데이터에서도 좋은 성능을 보이는데, 그 이유는 각 노드에서 일부 데이터를 복원 추출하여 학습시키며, 여러 트리의 결과를 평균 또는 다수결 방식으로 조합하여 예측의 변동성을 줄이기 때문이다(Biau & Scornet, 2016).

본 연구의 데이터는 출산의사가 있는 샘플이 전체 표본의 약 4.2%에 불과하여 불균형한 분포를 보였다. 데이터 불균형이 심하면 과적합 문제가 발생하거나 예측 결과가 많은 수의 샘플을 가진 클래스에 편향될 가능성이 크기 때문에 분석 전에 데이터 불균형을 해소하는 것이 중요하다. 이에 데이터 불균형을 해결하기 위해 소수 클래스를 증대시키는 기법인 SMOTE-ENN을 적용했다. SMOTE 기법은 기존 데이터를 단순 복제하는 것이 아니라 기존 데이터의 패턴을 유지하면서 새로운 샘플을 생성하는 방식으로 데이터 균형을 맞추기 때문에 예측모델이 특정 클래스에 치중하지 않고 학습할 수 있도록 데이터를 보완해준다(Alahmadi, 2025). ENN은 오버샘플링으로 인해 발생할 수 있는 불필요한 데이터 또는 노이즈를 제거하는 기법으로 클래스 간 경계를 더 명확하게 조정하여 머신러닝 예측모델의 학습 성능을 최적화한다(Siswara, Soleh & Wigena, 2024).

의사결정트리 분석은 단일 트리로 예측하기 때문에 분석결과가 직관적이고 해석하기 쉬운 반면, 랜덤포레스트는 많은 수의 트리를 결합하는 모델이기 때문에 직관적 해석이 어려워 다양한 시각화 기법이 필요하다(Siroky, 2009). 설명변수가 예측 성능에 영향을 미치는 정도를 직관적으로 보여주는 변수 중요도 그래프, 설명변수와 반응변수와의 관계를 시각화하는 부분의존성 도표 등이 있으며, 본 연구도 이들 시각화 도구를 통해 분석결과를 기술한다. 본 연구는 파이썬 인터프리터가 내장되어 있는 Google Colab을 활용하여 랜덤포레스트 분석을 진행하였다.

### Ⅲ. 연구결과

#### 1. 기술통계

아래 <표 2>는 분석모델에 포함된 설명변수의 기술통계이다. 전체 데이터의 평균 연령은 42세였다. 출산의사가 있는 집단의 평균 연령은 33세로 출산의사가 없는 집단(42세)에 비해 낮았다. 평균 자녀 수는 1.9명이었으며, 출산의사가 있는 집단의 평균 자녀 수(0.6명)가 출산의사가 없는 집단(2.0명)에 비해 적었다. 가구소득은 출산의사가 없는 집단(7,013만원)이 출산의사가 있는 집단(6,141만원)에 비해 높았는데, 이는 반응변수의 집단별 평균 연령을 고려할 때 응답자의 연령, 직장에서의 경력기간 등이 반영된 것일 수 있다. 아내의 경제적 협상력은 출산의사가 있는 집단(0.440)이 출산의사가 없는 집단(0.381)에 비해 높았고, 남편의 가사노동 분담률 역시 출산의사가 있는 집단(0.319)이 출산의사가 없는 집단(0.204)에 비해 높았다. 개인주의 가치관, 성역할 가치관, 가족 가치관은 모두 집단 간 큰 차이가 없었다. 가족친화제도의 경우 출산의사가 있는 집단은 회사로부터 가족친화제도를 제공받는 비율이 73.5%였으나, 출산의사가 없는 집단은 그 비율이 35.3%에 불과했다.

<표 2> 기술통계

설명변수	전체		출산의사 있음		출산의사 없음	
	평균	표준편차	평균	표준편차	평균	표준편차
남편의 가사노동 분담률	0.209	0.154	0.319	0.155	0.204	0.152
아내의 경제적 협상력	0.384	0.112	0.440	0.096	0.381	0.112
가구소득	6976.154	3133.989	6141.288	2603.406	7012.863	3150.561
개인주의 가치관	2.694	0.572	2.682	0.536	2.694	0.574
성역할 가치관	2.817	0.454	2.894	0.490	2.813	0.452
가족 가치관	2.277	0.444	2.327	0.496	2.275	0.442
가족친화제도	0.369	0.483	0.735	0.443	0.353	0.478
연령	42.406	5.196	33.394	3.852	42.802	4.880
학력	4.988	0.974	5.545	0.745	4.963	0.975
자녀수	1.906	0.780	0.621	0.546	1.963	0.739

## 2. 예측모델 평가

랜덤포레스트는 교차검증 없이 내부 데이터를 이용하여 모델 성능을 평가할 수 있으며, 주어진 데이터를 학습 데이터와 테스트 데이터로 분류하여 모델의 예측력을 검증한다. 여기에서는 전체 데이터 중 80%를 학습 데이터로 활용하여 예측모델을 구축하고, 전체 데이터의 20%는 예측모델 평가에 사용하였다. SMOTE-ENN을 적용하기 전 데이터는 총 3,134건이며, 출산의사가 없는 집단의 샘플 수는 3,002건(95.8%), 출산의사가 있는 집단은 132건(4.2%)으로 데이터 불균형이 심한 편이다. 학습데이터는 2,507건, 테스트 데이터는 627건으로 각각 80%, 20%의 비율로 분류하였고, 집단별 분포도 원본 데이터와 유사했다. 데이터 불균형을 해소하고 랜덤포레스트 모델의 예측성능을 최적화하기 위해 학습데이터에 SMOTE-ENN을 적용하였다. 테스트 데이터는 실제 현실 적용에 대한 일반화 능력을 높이고, 평가의 신뢰성을 확보하기 위해 데이터 증강 없이 원본 데이터를 유지했다.

SMOTE-ENN 적용 후 학습데이터는 총 2,597건으로 출산의사가 없는 집단의 샘플 수 1,882건, 출산의사가 있는 집단의 샘플 수 715건이다. 출산의사가 있는 집단의 비중이 27.5%로 증강되어 불균형이 개선되었고, SMOTE-ENN 적용 후 예측모델의 성능도 개선되었다. 아래 <표 3>은 예측모델의 평가결과이다.

본 연구에서는 기본 랜덤포레스트, 균형 랜덤포레스트(balanced random forest), 균형 랜덤포레스트에 SMOTE-ENN 적용, 균형 랜덤포레스트에 SMOTE-ENN 적용 및 임계값 조정 순으로 모델을 발전시키면서 성능이 어떻게 변하는지 살펴보았다. 기본 랜덤포레스트 모델은 정확도는 높으나 재현율이 낮았고, 균형 랜덤포레스트 모델은 재현율은 높지만 정밀도가 낮았다.

<표 3> 예측모델 평가결과

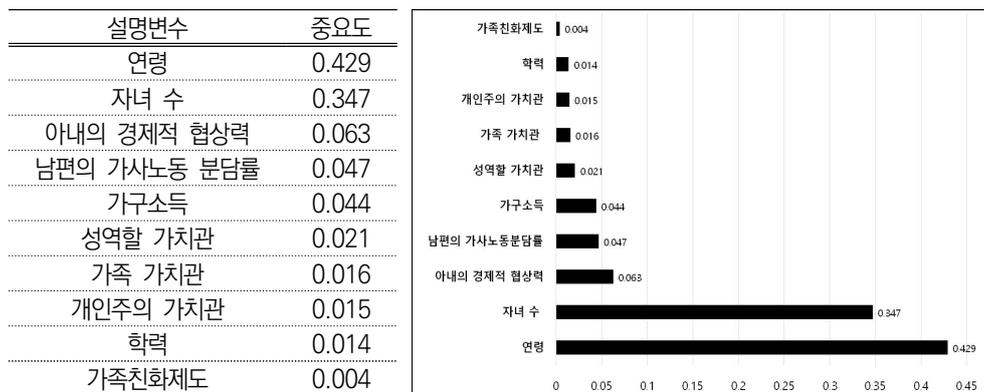
예측모델	정확도	정밀도	재현율	F-1 score
Random Forest(RF)	0.969697	0.818182	0.346154	0.486486
Balanced RF	0.918660	0.333333	0.961538	0.495050
Balanced RF+SMOTE-ENN	0.945774	0.425926	0.884615	0.575000
Balanced RF+SMOTE-ENN+임계값 조정	0.961722	0.523810	0.846154	0.647059

SMOTE-ENN을 적용한 균형 랜덤포레스트 모델은 재현율을 유지하면서 정확도가 소폭 향상되었다. 원본 데이터의 심한 불균형 때문에 재현율에 비해 정밀도가 충분한

수준은 아니었으나, 본 연구의 목적상 출산의사가 있는 응답자를 얼마나 잘 예측했는지가 중요하기 때문에 재현율을 우선적으로 고려하였다. 그중에서도 정밀도와 재현율 간의 균형을 극대화하도록 임계값을 조정한 모델이 가장 우수한 성능을 보여 해당 모델을 최종 선택했다(F-1 score=0.647059).

### 3. 중요도 지수 분석

랜덤포레스트 분석에서는 설명변수 중요도 지수(Gini important index)와 부분의존성 도표(partial dependence plot)를 이용해 설명변수가 반응변수에 어떤 영향을 미치는지 파악할 수 있다. 중요도 지수를 통해서는 각 변수가 결과 예측에 어느 정도의 영향력을 갖는지 직관적으로 파악할 수 있으나, 각 변수가 미치는 영향의 방향 등을 파악하기 어렵다. 부분의존성 도표는 이러한 한계를 보완하는 것으로 각 변수가 반응변수에 어떤 방향으로 영향을 미치는지 파악할 수 있으며, 특정 구간에서 어떻게 다른 영향을 미치는지도 파악할 수 있다.



[그림 1] 설명변수 중요도 지수

본 연구의 예측모델에 투입된 변수의 중요도 분석결과는 [그림 1]과 같다. 랜덤포레스트 분석의 중요도 지수는 변수의 상대적 중요도 순위가 중요하다(Strobl, Malley & Tutz, 2009). 전체 데이터를 예측모델에 투입하였을 때, 설명변수의 중요도는 연령, 자녀 수, 아내의 경제적 협상력, 남편의 가사노동 분담률, 가구소득, 성역할 가치관, 가족 가치관, 개인주의 가치관, 학력, 가족친화제도 제공 순으로 나타났다. 연령, 자녀 수와 같은 개인 요인의 성능 기여도가 가장 높았고, 그다음으로는 성평등 요인,

경제적 요인 순으로 예측모델 성능에 기여도가 큰 것으로 확인되었다. 출산이 생물학적 과정이고, 부부가 생각하는 이상적인 자녀 수에 제한이 있기 때문에 연령, 자녀 수 등의 개인 요인의 기여도가 높게 나타난 것으로 보이며, 초혼 연령 및 초산 연령의 상승 등도 영향을 주었을 것으로 예상된다. 각 변수에 대한 해석은 부분의존성 도표 분석을 통해 좀 더 자세히 기술하기로 한다.

가족친화제도는 0에 가까운 값을 보여 해당 변수가 예측 성능에 기여하고 있는지 모델성능 변화에 대한 추가적인 분석을 진행했다. 가족친화제도 변수를 제외했을 때 F1-score가 0.575에서 0.480로 약 16.5% 이상 감소하였기 때문에 가족친화제도 변수가 중요도는 낮으나, 모델 예측 성능에 기여하는 것으로 확인되었다. 이는 다른 변수들과 상호작용하며 중요한 정보를 제공하고 있을 가능성을 시사한다.

#### 4. 부분의존성 도표 분석

부분의존성 도표 분석은 다른 변수들의 값을 평균화한 상태에서 특정 변수의 값이 변화할 때 나타나는 예측값의 평균적인 변화를 산출한다. 설명변수가 반응변수에 미치는 영향의 전반적인 증감 패턴과 변곡점을 파악하는 데 유용하다. 각 설명변수의 효과를 독립적으로 평가하고 싶은 경우, 비선형 관계나 상호작용 효과를 탐색하고 싶은 경우, 랜덤포레스트와 같은 복잡한 모델의 결과를 해석해야 하는 경우에 유용하게 활용된다(Siroky, 2009).

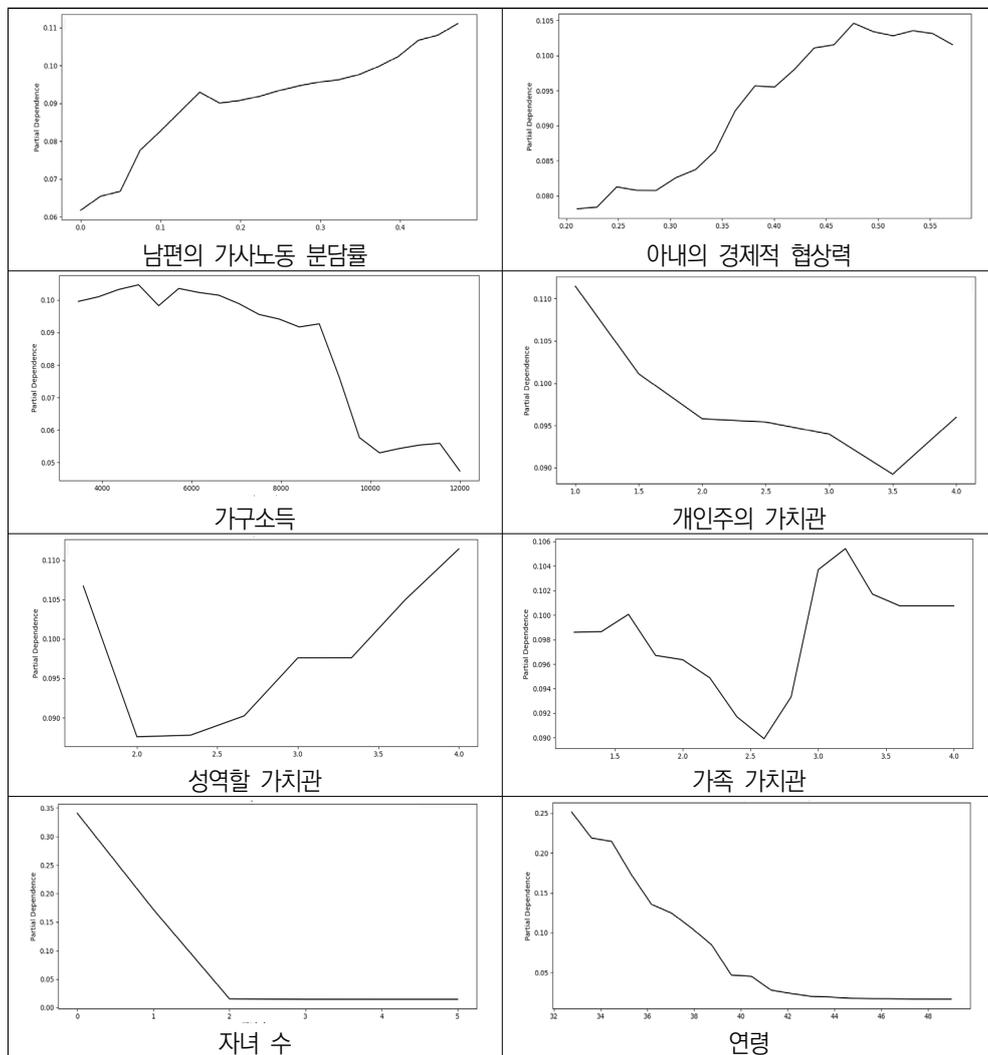
[그림 2]는 전체 데이터를 설명변수별로 분석한 부분의존성 도표 분석결과이다. 그래프의 x축은 각 설명변수의 범위를 의미하며, y축은 반응변수에 대한 평균 예측확률로 맞벌이 기혼여성이 출산의사가 있을 가능성을 의미한다.

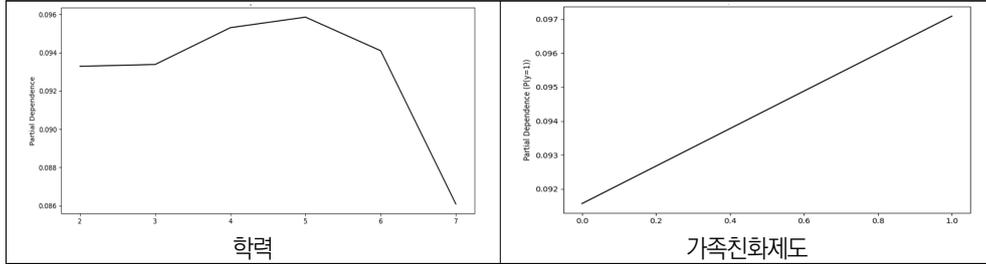
성평등 요인부터 살펴보면, 남편의 가사노동 부담률이 증가할수록 여성의 출산의사가 증가하는 경향을 보였다. 남편의 가사노동 부담률이 약 15% 지점을 넘어선 이후에는 출산의사가 소폭 감소하였으나, 전반적으로는 남편의 가사노동 부담률이 높아질수록 꾸준히 출산의사가 증가하는 패턴을 보였으며, 약 47% 지점에서 출산의사가 가장 높은 수준을 보였다. 이는 부부간 가사노동 부담이 공평해지는 정도까지 남편의 가사노동 참여가 출산의사를 증가시키는 데 기여한다는 것을 시사한다.

아내의 경제적 협상력 역시 부부간 대등한 수준으로 증가할 때 여성의 출산의사가 상승하는 패턴을 보였다. 특히 약 30~50% 구간에서 출산의사에 대한 평균 예측확률이 보다 뚜렷하게 높아졌다. 이는 여성의 소득 증가가 출산 결정에 긍정적 영향을 미친다는 기존 연구와 일치한다(Chung & Lee, 2021). 그러나 여성의 경제적 협상력이

약 50%를 넘어서는 구간부터는 출산의사에 대한 평균 예측확률이 감소하는 패턴을 보이는 비선형 관계가 확인되었다. 여성의 경제적 협상력이 남편을 역전하면 출산의사가 감소하는 경향을 보인다.

경제적 요인인 가구소득은 약 4,800~4,900만원 구간에서 출산의사에 대한 평균 예측확률이 증가하는 경향을 보였으나, 이후부터는 감소하는 패턴으로 나타나 비선형 관계가 확인되었다. 특히 약 9,000만원~1억원 구간에서는 출산의사 예측확률이 급격하게 감소하는 경향을 보였다.





[그림 2] 부분의존성 도표 분석

인식·문화적 요인 중 개인주의 가치관에 동의하는 정도가 강할수록 출산의사에 대한 평균예측확률이 전반적으로 감소하는 경향을 보였다. 3.5점~4점 구간에서 예측확률이 다시 증가하는 비선형적 패턴이 나타났으나, 데이터 분포를 고려하여 해석에 유의할 필요가 있다. 약 96.5%의 데이터가 1점~3.5점 구간에 집중되어 있고, 특히 2점~3점 구간에 약 82.5%의 데이터가 밀집되어 있어 해당 구간의 예측 신뢰도가 높다. 반면, 3.5점~4점 구간에서는 데이터가 적어 예측의 신뢰성이 상대적으로 낮을 수 있다. 따라서 데이터 밀집 구간(2점~3점)에서 개인주의 가치관에 동의하는 정도가 강할수록 출산의사 예측확률이 감소하는 패턴이 지속적으로 나타난다는 점에서 개인주의 가치관이 출산의사에 부정적 영향을 미칠 가능성이 있다고 해석하는 것이 적절할 것으로 보인다.

성역할 가치관은 2점~3.5점 구간에 데이터의 약 92.7%가 집중되어 있는데, 해당 구간에서는 평등한 성역할 가치관에 동의할수록 출산의사에 대한 평균 예측확률이 증가하는 경향을 보인다. 이는 공평한 가사분담의 필요성에 대해 동의할수록 출산의사가 높아진다는 기존 연구결과와 일치한다(임병인·서혜림, 2021; 정아름, 2018).

가족 가치관은 개방적인 가족가치관에 대한 동의수준이 높아질수록 출산의사에 대한 평균 예측확률이 감소했다 증가하는 패턴의 비선형 관계를 보인다. 비선형 관계는 데이터 분포가 약 93.6% 집중되어 있는 1.5~3점 구간에서 확인된다. 약한 수준의 동의 구간에서 출산의사에 대한 평균 예측확률이 증가했다가 동의수준이 높아지면 다시 출산의사에 대한 평균 예측확률이 감소하는 추세로 전환된다.

제도적 요인인 가족친화제도가 제공되었을 때 출산의사의 예측값이 높아지는 것으로 나타났다. 다만, 중요도 분석 결과에서 설명한 바와 같이 가족친화제도는 예측모델에서 중요도가 낮았음에도 불구하고 해당 변수를 제외했을 때 모델 예측성능이 저하되어 중요도는 낮지만, 예측에 기여하는 것으로 확인되었다. 이는 가족친화제도가

다른 변수들과 상호작용하며 중요한 정보를 제공하고 있을 가능성을 시사한다. 예를 들어, 남편의 가사노동 분담률, 아내의 경제적 협상력 등 부부의 성평등 수준이 어느 정도인지, 가구소득이 부부의 육아휴직 이용으로 인한 소득의 감소를 감당할 수 있는지 등에 따라 가족친화제도의 영향이 상이할 수 있다. 이에 대해서는 후속연구를 통해 심도 있는 분석이 필요하다.

개인요인 중 자녀 수의 증가는 출산의사에 부정적인 영향을 주었다. 자녀가 2명일 때까지 급격하게 감소하다가 2명 이후에는 낮은 수준에서 안정화되는 패턴을 보였다. 여성의 학력은 비선형 관계를 보였다. 전문대 졸업의 학력까지는 출산의사가 증가하였으나, 4년제 대학교 졸업에 진입하는 구간부터 출산의사가 감소하여 대학원 졸업 이상의 학력을 가진 여성의 경우 가장 큰 폭으로 출산의사가 감소한다. 학력이 높을수록 결혼연령이 늦어지거나, 노동시장에서의 임금소득 또는 사회경제적 지위가 높을 가능성이 높으며, 이 때문에 출산, 육아로 인한 기회비용이 높아져 출산의사가 감소할 수 있을 것으로 해석된다.

#### IV. 논의 및 결론

본 논문은 한국여성정책연구원의 여성가족패널 4차~9차 자료를 이용한 랜덤포레스트 분석을 통해 맞벌이 기혼여성의 출산의사에 영향을 미치는 요인을 분석했다. 주요한 분석결과 및 그에 따른 해석과 논의는 다음과 같다.

첫째, 설명변수의 중요도 분석결과 개인 요인, 성평등 요인, 경제적 요인, 인식·문화적 요인 순으로 출산의사 예측모델에 대한 영향력이 높았다. 출산이 생물학적 과정이기 때문에 개인의 건강상태 등과 관련되는 연령의 중요도가 높게 나타난 것으로 보인다. 그리고 경제적, 인식·문화적 요인에 비해 성평등 요인의 영향력이 더 높았는데, 이는 한국의 저출생 위기 해결을 위해 가정 및 사회에서의 성평등 수준 향상이 여전히 중요한 과제임을 시사한다.

둘째, 성평등 요인인 남편의 가사노동 분담률과 아내의 경제적 협상력이 부부간 대등한 수준일 때 출산의사에 대한 평균 예측확률이 가장 높았다. 본 논문의 분석대상이 맞벌이 기혼여성임을 고려하면 남편의 가사노동 분담률이나 아내의 경제적 협상력이 평등한 수준일 때 부부간 합의를 통해 출산을 결정할 수 있기 때문으로 보인다. 특히 남편의 가사노동 참여는 여성의 가사·양육 부담을 완화하고, 여성의 행복감과

가사노동 만족도 등에 긍정적 영향을 미쳐 출산의사를 높인다(최지훈·안선희, 2017; Cheung, 2023). 그리고 남성이 가사노동에 적극적으로 참여한다는 것은 여성에게 출산 이후의 자녀돌봄 상황에서 남편의 협력을 기대할 수 있다는 신호로 작용한 것이라 볼 수 있다(Fanelli & Profeta, 2021).

셋째, 여성의 경제적 협상력은 출산의사에 대해 비선형 관계를 보였다. 약 50% 구간까지는 여성의 소득 증가가 출산의사에 대한 평균 예측 확률에 긍정적 영향을 주었으나, 50% 이상의 구간에서는 오히려 여성의 소득 증가가 출산의사에 대한 평균 예측 확률에 부정적 영향을 주었다. 이에 대해서는 두 가지 해석이 가능하다. 첫째, 출산·육아로 인해 포기해야 하는 기회비용이 남편에 비해 더 높아지기 때문에 여성의 소득기여도가 높을수록 출산가능성이 낮아질 수 있다(주익현, 2023). 둘째, 여성의 소득이 남편의 소득을 역전하게 되면 전통적인 남성적 권위와 역할을 유지하기 위해 오히려 남성이 가사노동을 회피하는 경향이 나타날 수 있다(Syrda, 2023).

넷째, 경제적 요인인 가구소득은 일정 지점까지는 소득이 증가하면 출산의사에 대한 평균 예측확률이 증가하는 패턴을 보였으나, 해당 지점 이후부터는 가구소득이 증가하면 출산의사에 대한 평균 예측확률은 오히려 감소했다. 패턴이 변화되는 변곡점은 출산과 육아에 필요하다고 생각되는 적정 소득구간으로 예상된다. 가구소득에 대해서는 베커의 합리적 선택이론의 시각과 제2차 인구전이론에서 강조하는 인식·문화적 요인이 함께 작동한다고 해석할 수 있다. 자녀 양육에 필요한 적정 가구소득에 도달할 때까지는 가구소득의 증가가 출산의사 증가를 견인하지만, 그 이상의 소득 구간부터는 여성의 경력단절로 인한 기회비용 증가를 비롯해 이상적인 선호 자녀 수 고려, 삶의 질, 육아·교육의 질에 대한 투자 등이 복합적으로 작용하여 출산의사를 감소시키는 것이라고 예상해 볼 수 있다. 후속연구를 통해 다른 변수와의 상호작용 등을 살펴볼 필요가 있다.

다섯째, 인식·문화적 요인 중 개인주의 가치관은 출산의사에 부정적 영향을 미쳤다. 이는 제2차 인구전이론에서 설명하는 개인주의 및 자율성 가치가 확산됨에 따라 개인의 삶의 질에 대한 투자, 자아실현의 욕구 증가 등으로 인해 출산율이 감소하게 된다는 내용과도 연결된다. 개인주의적, 자유주의적 가치관에 동의하면 혼인과 출산을 당연한 것으로 여기는 전통적 형태의 가족제도와 가부장적 가족문화를 수용하지 않는 경향을 보일 수 있다(임재연, 2021). 본 연구에서 평등한 성역할 가치관은 출산의사에 긍정적 영향을 미쳤으며, 이는 남편의 가사노동 분담률, 여성의 경제적 협상력이 대등할 때 출산의사에 대한 평균 예측확률이 가장 높다는 분석결과와도 일

치한다. 가족가치관은 개방적인 가족관에 대해 약한 수준의 동의를 보일 때 출산의사에 대한 평균 예측확률이 증가했다가 동의수준이 높아지면 출산의사에 대한 평균 예측확률이 감소하는 비선형 관계를 보였다. 이는 기혼여성에게서 결혼관과 자녀관이 혼재된 모습이 나타난다고 분석한 기존 연구결과를 참조할만하다. 기혼여성의 경우 개방적 결혼관에 동의할수록, 전통적 자녀관에 동의할수록 출산의사가 높아졌다(임병인·서혜림, 2021). 본 연구에서는 결혼·가족가치관을 함께 통합하여 분석하였는데 후속연구에서는 분리해서 분석하는 것도 고려할 필요가 있을 것으로 보인다.

여섯째, 가족친화제도는 중요도는 낮았지만, 다른 변수들과의 상호작용하며 중요한 정보를 제공할 가능성이 있음이 확인되었다. 가족친화제도가 첫째자녀 출산에는 긍정적 효과가 있으나 둘째자녀부터는 그 효과가 약해진다는 연구(주익현, 2023), 연령이 낮을수록, 도시에 거주할수록, 자녀가 꼭 필요하다고 생각하는 여성일수록 육아휴직 이용이 출산의사에 긍정적 영향을 미친다는 연구(Jung, Jang & Park, 2024) 등 기존 연구를 참고하여 후속연구에서 다른 변수와의 상호작용을 분석할 필요가 있다.

일곱째, 개인요인 중 자녀 수 증가는 출산의사에 부정적 영향을 주었다. 특히 2명 구간에서 가장 낮은 값을 보인 후 안정화되는 패턴을 보였는데, 이는 이상적인 자녀수로 2명을 선호한다는 경향과도 관련된다. 한국보건사회연구원의 조사에 따르면 이상적인 자녀수가 2명이라고 응답한 비중이 가장 높았다(박종서·임지영·김은정 외, 2023). 한편, 연령이 증가할수록 출산의사는 감소했다. 약 41세 구간까지 지속적으로 감소하였으며, 이후 구간에서는 매우 낮은 수준에서 안정화되는 패턴을 보였다. 이는 생물학적 이유와 더불어 결혼연령이 늦어지면서 소수의 자녀에게 자원을 집중하는 전략이 동원되고 있기 때문일 수 있다(주익현, 2023).

## V. 정책적 제언 및 시사점

본 논문의 이론적, 정책적 함의는 다음과 같다.

첫째, 한국사회의 저출생 문제 해결을 위해 여전히 성평등 관점의 접근이 중요하다는 것을 확인했다. 예측결과 성평등 요인의 설명력이 가장 높았으며, 이는 다중균형 이론이 설명하는 것처럼 부부간 성평등의 증대가 출산율에 긍정적 영향을 미친다는 점을 뒷받침한다. 그리고 그 균형이 깨져 역전되면 성평등에 역행하는 방향으로 남편의 가사노동 참여가 감소하는 경향을 보이는 비선형 관계도 확인할 수 있었다. 향후

저출산고령사회기본계획 등 정책설계에 있어서 부부간 공평한 가사노동 분담을 위한 일·가정 양립 지원정책이나 여성이 가정 내에서 경제적 협상력을 가질 수 있도록 노동시장 구조개선에 대한 정책적 노력이 확대되어야 한다.

둘째, 성평등의 양극화 해소가 필요하다. 대기업·정규직 중심으로 육아휴직 등 가족친화제도 이용이 편중되는 등 사각지대가 여전히 존재하고, 노동시장 양극화로 인한 임금 격차가 지속되는 상황에서 가정 내 성평등 역시 양극화되고 있는 것이 현실이다. 에스핑-앤더슨의 지적처럼 성평등의 양극화가 해소되어야 가정과 사회 전반에서의 성평등 수준이 향상되고, 이것이 출산율 제고로 이어지는 사회구조 개선이 가능하다. 이에 가족친화제도 지원 및 이용의 사각지대를 해소하고 성평등의 양극화를 개선할 수 있는 정책 마련이 필요하다.

둘째, 가구별 상황에 맞는 촘촘한 정책이 필요하다. 개별 가구의 객관적 조건을 결정하는 가구소득 등 경제적 요인, 연령, 자녀 수, 학력 등 개인요인 관련 변수들은 여성의 출산의사에 대해 비선형 관계를 보였다. 이는 여성의 연령, 자녀 수 및 가구소득 구간에 따라 출산을 결정하게 하는 핵심 요인이 상이할 수 있으며, 출산·육아를 위해 필요로 하는 정책적 지원이 다를 수 있음을 의미한다. 변수의 비선형 관계 및 변수 간 상호작용을 고려한 맞춤형 정책설계가 필요하다.

셋째, 실제 자녀 수와 이상적으로 생각하는 자녀 수 간의 격차를 줄일 수 있는 정책 설계가 필요하다. 본 연구결과에 따르면 소득, 주거 등의 경제적 요인, 성평등 등 기타 사회적 요인이 충족되더라도 개인이 선호하는 이상적인 자녀의 수를 넘어서 출산을 계획하지 않는 경향을 보였다. 첫 자녀 출산 이후 후속 출산을 고려하는 가구에 대한 분석을 통해 적절한 정책적 개입이 필요하다.

넷째, 정책대상자의 연령에 따라 출산지원 정책을 고도화할 필요가 있다. 출산의사가 급감하는 임계 연령에 도달하기 전에 정책적 개입이 필요하며, 특히 혼인과 출산 연령이 높아지고 있다는 점을 고려하여 늦은 출산을 선택한 이들에 대한 사회적 안전망과 유연한 제도설계가 필요하다. 결혼과 출산에 대한 의사가 형성되는 청년 초기와 실제 결혼과 출산이 이루어지는 시기에 필요한 지원이 상이할 수 있기 때문에 이에 대한 생애주기 맞춤형 지원이 고려되어야 한다.

이 논문은 몇 가지 한계를 가진다. 첫째, 데이터 불균형이 심했다. 한국의 낮은 출산율을 고려할 때 출산의사가 있다고 응답하는 비중이 현저히 낮아 데이터 불균형이 불가피했고, 이를 보완하기 위해 데이터 불균형을 개선하는 기법을 활용했다. 둘째, 조사문항의 특성상 기혼여성을 대상으로만 분석했다. 미혼여성의 출산의사, 비혼출산

을 고려하고 있는 여성의 출산의사 등을 고려하지 못했다. 결혼, 가족, 자녀 등에 대한 인식이 변화하고 있음을 고려할 때 다양한 형태의 출산계획에 대해서도 후속연구가 필요하다. 셋째, 변수 간 상호작용을 분석하지 못했다. 남편의 가사노동 참여율, 아내의 경제적 협상력, 가구소득, 자녀 수 등 출산의사와 비선형 관계를 보인 변수들과 다른 변수들 간의 상호작용을 살펴볼 수 있는 후속연구가 필요하다. 넷째, 랜덤포레스트는 비선형 관계 분석이나 변수의 예측 기여도를 파악하는 데는 강점이 있으나 변수 간 인과관계나 통계적 유의성 검증에는 한계가 있다. 이를 보완하기 위해 부분 의존성 도표 분석을 진행하였으나, 이 역시 예측모델을 해석하기 위한 보조적 수단에 불과하다. 따라서 이 논문의 분석결과 및 정책적 시사점은 변수 간 인과관계나 정책 결정요인을 단정하기 위한 것이 아니며, 예측결과에 기반한 경향성 중심으로 해석한 것임을 밝힌다.

출산은 생물학적 과정이자 개인의 선택이지만, 동시에 사회적 맥락에 의해 영향을 받는 결정이다. 사회와 가정에서의 성평등 수준, 직장에서의 경력전망, 일·가정 양립, 출산·육아와 관련한 정책적 지원, 출산·육아에 대한 사회적 인식 등이 복합적으로 영향을 미친다. 따라서 여성의 출산 결정은 다양한 모습으로 나타날 수 있으며, 이 점에서 다양한 변수 간의 상호작용 등을 고려한 후속연구가 이어지길 기대한다.

## 참고문헌

- 강유진(2020). 무자녀와 한자녀 여성의 출산계획유형에 영향을 미치는 요인: 자녀가치관과 부부평등성 요인을 중심으로. 육아정책연구, 14(3), 3-26.
- 김정석(2007). 기혼여성의 출산아수별 추가출산계획. 한국인구학, 30(2), 97-116.
- 김현동(2021). 여성배우자 전업주부여부와 종사상 지위가 출산에 미치는 영향. 사회과학연구, 28(1), 48-74.
- 남국현(2022). 기혼여성의 결혼만족도와 출산계획, 직무만족도에 관한 연구. 보건사회연구, 42(1), 166-182.
- 박선권(2022). 제4차 「저출산고령사회 기본계획」의 문제점과 개선방향: 저출산 대응을 중심으로. 국회입법조사처 연구보고서.
- 박정희(2024). 국내 출산율의 장기적 결정요인에 대한 연구. 산업경제연구, 37(3), 379-403.
- 박중서·임지영·김은정·변수정·이소영·장인수·조성호·최선영·이혜정·송지은(2023). 2021년도 가족과 출산조사(수정본). 한국보건사회연구원.

- 배은경(2021). '저출생'의 문제제기를 통해 본 한국 인구정책의 패러다임 전환 모색: 재생산 주체로서 여성의 행위성과 저출산·고령사회정책의 검토. *페미니즘연구*, 21(2), 137-186.
- 선민정·조상미(2019). 기혼 취업여성의 일과 가정 관련 요인이 둘째 자녀 출산의지에 미치는 영향: 직장 내 성차별 문화에 대한 일·가정 양립 갈등의 조절효과를 중심으로. *사회과학연구논총*, 35(2), 5-35.
- 송민영(2018). 일가정양립정책, 직무만족도가 부부관계를 매개하여 취업모의 후속출산의도와 자녀가치에 미치는 영향. *육아지원연구*, 13(3), 5-32.
- 에스핑-앤더슨, 요스타(2014). 끝나지 않은 혁명: 성역할의 혁명, 고령화에 대응하는 복지 국가의 도전. 서울: 나눔의집. 주은선·김영미 역. Esping-Andersen, Gøsta(2009). *Incomplete Revolution: Adapting to Women's New Roles*.
- 유계숙(2010). 기업의 가족친화제도가 근로자의 자녀출산에 미치는 영향. *한국가정관리학회지*, 28(2), 89-96.
- 이완·채재은(2017). 기혼여성의 자녀가치관과 추가출산의향 간의 관계에서 출산장려정책의 조절효과 분석. *디지털융복합연구*, 15(9), 65-78.
- 이은정·송영수·김지하·오수현(2020). 랜덤 포레스트를 활용한 4년제 대학 중도탈락률 예측요인 탐색: 대학 수준 결정요인을 중심으로. *교육공학연구*, 36(1), 191-219.
- 이진권(2011). 경제적 불확실성이 출산 의사결정에 미치는 효과: 중도절단된 포아송 회귀 분석 및 중도절단된 음이항 회귀분석을 이용한 자녀 수 결정모형 분석. *Journal of The Korean Data Analysis Society*, 13(3), 1521-1531.
- 임병인·서혜림(2021). 여성의 가족가치관과 결혼 및 출산의향. *보건사회연구*, 41(2), 123-140.
- 임재연(2021). 한국 청년세대의 출산 의향에 영향을 미치는 사회적 요인: 가족가치관 유형의 효과를 중심으로. *한국인구학*, 44(4), 47-74.
- 정병삼(2023). 청년층의 결혼 및 출산의사에 영향을 미치는 요인 분석. *차세대융합기술학회논문지*, 7(12), 2244-2254.
- 정아름(2018). 기혼여성의 청소년기 가정환경요인, 가족가치관이 출산의지에 미치는 영향. *한국사회과학연구*, 37(3), 147-174.
- 주익현(2023). 모성보호제도 사용의 기혼여성 임금근로자 출산효과: 출생순서와의 상호작용 효과를 중심으로. *공공사회연구*, 13(3), 202-226.
- 최지훈·안선희(2017). 가족친화제도와 남편의 자녀돌봄 참여가 기혼여성의 추가출산의향에 미치는 영향. *한국가정관리학회지*, 36(1), 75-85.
- 탁아림·이찬영(2024). 여성의 경제활동참가와 고용지위가 출산율에 미친 영향. *노동경제논집*, 47(3), 33-62.

- Alahmadi, M. (2025). A risk-optimized framework for data-driven IPO under-performance prediction in complex financial systems. *Systems*, *13*(3), 1-38.
- Becker, G. S. (1992). Fertility and the economy. *Journal of Population Economics*, *5*(3), 185-201.
- Biau, G. & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*, 197-227.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32.
- Cheung, A. K. L. (2023). Couples' housework participation, housework satisfaction and fertility intentions among married couples in Hong Kong. *Asian Population Studies*, *20*(3), 289-307.
- Chung, M. S. & Lee, K. (2021). A recent change in the relation between women's income and childbirth: Heterogeneous effects of work-family balance policy. *Journal of Demographic Economics*, *88*, 419-445.
- Doepke, M. & Kindermann, F. (2019). Bargaining over babies: Theory, evidence, and policy implications. *American Economic Review*, *109*(9), 3264-3306.
- Esping-Andersen, G. & Billari, F. C. (2015). Re-theorizing family demographics. *Population and Development Review*, *41*(1), 1-31.
- Fanelli, E. & Profeta, P. (2021). Fathers' involvement in the family, fertility, and maternal employment: Evidence from central and eastern Europe. *Demography*, *58*(5), 1931-1954.
- Ferragina, E. (2023). The 'two lives' of Esping-Andersen and the revival of a research programme: Gender equality, employment and redistribution in contemporary social policy. *Social Policy & Administration*, *59*(1), 1-19.
- Johnson, S. (2024). From economic turbulence to demographic change: Tracing the pathways of the second demographic transition in post-socialist contexts. *AUC Geographica*, *60*(1), 1-22.
- Jung, Y. H., Jang, Y. S. & Park, E. C. (2024). Impact of parental leave system on the childbirth plan among working married women: A three-year follow-up study of the Korean longitudinal survey of women and families. *BMC Pregnancy and Childbirth*, *24*(1), 1-9.
- Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development. *Proceedings of the National Academy of Sciences*, *111*(51), 18112-18115
- Morero-Mínguez, A. & Ortega-Gaspar, M. (2022). A change in work-family/life or a return to traditional normative patterns in Spain? *Frontiers in Sociology*,

7, 1-13.

OECD(2024). Society at a Glance 2024-Country Notes: Korea. <https://www.oecd.org>

Pollak, R. A. (2005). Bargaining power in marriage: Earnings, wage rates and household production. *National Bureau of Economic Research Working Paper*, No. 11239.

Robinson, W. C. (1997). The economic theory of fertility over three decades. *Population Studies*, 51(1), 63-74.

Siroky, D. S. (2009). Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys*, 3, 147-163.

Siswara, D., Soleh, A. M., Wigena, A. H. (2024). Classification modeling with RNN-based, random forest, and XGBoost for imbalanced data: A case of early crash detection in ASEAN-5 stock markets. *Scientific Journal of Informatics*, 11(3), 569-582.

Strobl, C., Malley, J. & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, 14(4), 323-348.

Syrda, J. (2023). Gendered housework: Spousal relative income, parenthood and traditional gender identity norms. *Work, Employment and Society*, 37(3), 794-813.

Vignoli, D., Guetto, R., Bazzani, G., Pirani, E. & Minello, A. (2020). A reflection on economic uncertainty and fertility in Europe: The narrative framework. *Genus*, 76(28), 1-27.

Zaidi, B. & Morgan, S. P. (2017). The second demographic transition theory: A review and appraisal. *Annual Review of Sociology*, 43, 473-492.

- 논문접수: 6월 11일 / 수정본 접수 7월 21일 / 게재 승인 8월 12일
- 교신저자: 안리라, 고려대학교 사회학 박사, all7137@naver.com

## Abstract

### Exploring Predictors of Fertility Intention Among Dual-Earner Married Women: Using the Random Forest Method

Lira Ahn

This study applied the random forest method to explore the factors influencing the fertility intentions of married, dual-income women and to determine how each factor impacts their decision-making. The analysis utilized data from the 4<sup>th</sup>(2012) to the 9<sup>th</sup>(2022) waves of the Korean Longitudinal Survey of Women and Families(KLoWF), and included 3,314 cases of married, dual-earner women under the age of 49. The main findings are as follows. First, the variable importance rankings in the predictive model revealed that gender equality factors were the most influential, followed by economic and cultural-attitudinal factors. Second, fertility intention was highest when both the husband's housework contribution and the wife's economic bargaining power were at equal levels between spouses. Third, household income exhibited a non-linear relationship with fertility intention: intention increased with income up to a perceived adequate level for childbearing, but decreased beyond that threshold. This indicates that promoting gender equality in the household and society is crucial for addressing Korea's low-birthrate crisis. Additionally, enhancing policy effectiveness will require the design of tailored policies that take into account the nonlinear relationships between economic factors and fertility intentions.

Keyword: fertility intention, dual-earner married women, gender equality, random forest