

# 랜덤포레스트 모델을 활용한 청년층 차입자의 채무 불이행 위험 연구\*

Predicting Default Risk among Young Adults with Random Forest Algorithm

인하대학교 소비자학과  
부 교 수 이 종 희\*\*

Dept. of Consumer Science, Inha University  
Associate Professor Lee, Jonghee

## 〈목 차〉

I. 서론	IV. 연구결과
II. 이론적 배경	V. 결론 및 함의
III. 연구방법	참고문헌

## 〈초록〉

청년층 및 저소득층을 포함한 취약계층과 제2금융권을 중심으로 한 부채 불이행에 대한 우려가 증가하고 있다. 청년층의 가계부채 건전성은 최근 고용 부진, 학자금대출 부담 증가, 제2금융권에서 고금리 대출 증가 등이 복합적으로 작용하여 더욱 취약해졌다. 본 연구의 목적은 한국의 청년층 차입자를 대상으로 채무 불이행 가능성을 진단하고, 그 가능성에 영향을 주는 요인을 예측하는 것이다. 이러한 목적을 달성하기 위하여 본 연구는 2021년 「가계금융·복지조사」를 활용하고, 청년층의 채무 불이행 가능성과 관련된 요인들을 포괄적으로 분석하기 위하여 머신러닝 알고리즘의 랜덤포레스트 방법을 적용하였다. 청년층 차입자의 채무 불이행 위험을 예측하는 모형을 탐색한 뒤 중요도 지수를 산출하고, 중요도가 높은 설명변수들을 선별한 뒤, 주요 결정요인들의 부분 의존성 도표를 제시하고자 하였다. 최종적으로 자산대비부채비율(DTA), 의료비 비중, 가계부실위험지수(HDRI), 통신비 비중, 주거비 비중이 주요한 변인으로 나타났다.

**주제어:** 청년층 차입자, 채무 불이행, 머신러닝, 랜덤포레스트

## 〈Abstract〉

There are growing concerns about debt insolvency among youth and low-income households. The deterioration in household debt quality among young people is due to a combination of sluggish employment, an increase in student loan burden and an increase in high-interest loans from the secondary financial sector. The purpose of this study was to explore the possibility of household debt default among young borrowers in Korea and to predict the factors affecting this possibility. This study utilized the 2021 Household Finance and Welfare Survey and used random forest algorithm to comprehensively analyze factors related to the possibility of default risk among young adults. This study presented the importance index and partial dependence charts of major determinants. This study found that the ratio of debt to assets(DTA), medical costs, household default risk index (HDRI), communication costs, and housing costs the focal independent variables.

**Key words:** Young Adults, Default Risk, Machine Learning, Random Forest

\*이 논문은 2019년 대한민국 교육부와 한국연구재단의 인문사회분야 신진연구자지원사업의 지원을 받아 수행된 연구임(NRF-2019S1A5A8032831).

\*\*주저자, 교신저자: 이종희(jay.jongheeleee@gmail.com), <https://orcid.org/0000-0002-0235-2689>

## I. 서론

한국 가계부채의 지속적인 총량 증가는 한국 경제의 중대한 위험 요인으로 지적되어 왔다. 부동산 가격 상승과 코로나 19로 경제 상황이 어려워졌고, 시중 금리가 상승하는 상황에서 사상 최대로 늘어난 가계부채가 한국 경제의 뇌관으로 작용할 수 있다는 우려가 나온다. 부채 총량 측면에서 2018년 1,493조 200억원이었던 가계부채(한국경제, 2018)는 2021년 4분기 말 1,862.1조원으로 약 25% 증가하였다(한국은행, 2021a). 가계의 순가처분 소득 대비 부채 비율은 2020년 사상 처음으로 200%를 넘어섰는데(OECD, 2021), 이는 한국은행이 관련 통계를 내기 시작한 2008년 이후 최대치에 달하는 수치이다.

한국 가계부채의 누증에 대한 해석은 다양하다. 금융위원회(2017)가 작성한 '가계부채 종합대책' 보고서에 의하면 한국의 가계부채는 주택담보대출 위주로 증가하여 실물자산도 동반하여 증가하였고, 소득 4-5분위 부채 점유율이 70% 수준에 이른다는 점에서 가계 부채 증가가 한국 금융시스템에 직접적인 위험 요인으로 작용하지 않을 것이라고 판단하였다. 반면 가계부채비율이 급격하게 증가한 것은 소득 증가에 비하여 부채의 증가가 더 큰 것이 원인이므로, 가계부채의 증가로 인한 가계의 상환 부담 증가는 한국경제의 전체적인 소비와 성장에 있어 제약요인으로 작용할 수 있다고 예측하기도 하였다(강중구, 2017).

한국의 가계대출 증가에 대한 해석은 일치되지 않았지만, 청년층 및 저소득층을 포함한 취약계층과 제2금융권을 중심으로 한 채무 불이행<sup>1)</sup>에 대한 우려는 존재한다. 청년층 차입자들이 보유한 가계부채는 꾸준히 증가하였는데, 가계대출 증감을 살펴보면 30대 미만의 청년층의 경우 2013년-18년 사이 전년 대비 가계대출이 연령대 중 가장 크게 증가하였다(연합뉴스, 2018). 또한, 가처분소득 대비 원리금 비율은 동기간 10.7%에서 14.0%까지 상승하였다. 최근 경제 악화와 취업난 속에서 청년 가계부채의 총량 증가는 이들 집단의 부채부실에도 심각한 영향을 미친 것으로 나타났다. 청년층 중 13%는 은행이 아닌 캐피탈, 카드사 등 고금리 금융기관을 이용한 경험이 있었다. 또한 연체경험자 중 약 32.3%는 금융 채무 불이행 등록 경험이 있었고(아시아경제, 2018), 3곳 이상 금융기관으로부터 돈을 빌린 20대 다중채무자가 2022년 36만 6,369만명에 달하였다. 이는 2019년 대비 21% 증가한 수치이었는데, 동기간 30대가 1.0%가 증가하였고, 40대가 0.2% 감소하였던 것과 비교하면, 청년층은 다른 연령대의 증가세를 압도하였다(매일경제, 2022).

청년층의 가계부채 건전성이 최근 더욱 취약해진 것은 몇 가지 이유로 설명할 수 있다. 첫째, 고용 부진으로 소득이 증

가하기 어려운 여건이 지속되고 있다(현대경제연구원, 2018). 2020년에 발생한 세계적 보건위기로 인해 경기침체와 함께 전반적인 고용위축이 예상되는데, 이는 신규구직 계층인 청년에게 특히 크게 작용할 수 있다(한요셉, 2020). 자산보다 부채가 많은 청년층의 경우 소득변동에 상대적으로 더 취약할 수 있다(파이낸셜뉴스, 2018). 구직에 걸리는 시간이 증가하면서 생활비를 위한 대출까지 이어지기도 한다. 개인회생을 신청한 20대 청년들이 처음 빚을 지게 된 이유로는 '생계비 마련 목적'(43%)이 가장 많았다(서울시복지재단, 2022). 둘째, 청년층의 학자금 대출 부담이 많이 증가한 점도 하나의 원인일 수 있다. 실제로 2021년 학자금대출은 1,841억원으로 전년 대비 16.2% 증가하였다(동아일보, 2021). 셋째, 신용이 악화된 청년층의 경우 제도권의 금융을 이용하는 것이 어렵다 보니, 제2금융권에서 고금리 대출을 받을 수밖에 없는 구조적인 환경에 놓이게 되고, 채무 악순환이 이루어질 수 있다. 개인회생을 신청한 20대 청년들의 실태를 분석한 결과 제2금융권 부채비율이 압도적으로(78%) 높았다(서울시복지재단, 2022).

가계부채 부실화 관련 선행 연구를 고찰한 결과 가계부채 채무 불이행 위험에 영향을 미치는 요인으로 대체로 채무자의 인구사회학적 특징, 경제적 특징, 부채 관련 특징들이 나타났다. 인구사회학적 특징으로는 20-40대의 비근로자(e.g. 박윤태·노정현, 2017), 50세 이상의 채무자(e.g. 김영일·진경희, 2018), 일용직 근로자 및 자영업자(e.g. 이종희, 2019)의 부실 위험이 높은 것으로 나타났다. 경제적 특징으로는 소득 수준이 낮을수록(e.g. 배영목, 2011), 자산 수준이 낮을수록(e.g. 이종희, 2019) 부실위험이 높았다. 부채 관련 특징으로는 자산 대비 채무액이 증가할수록(e.g. 신진욱·박정민·김영, 2019), 금리가 상승할수록(e.g. 주동현, 2019) 부실화 가능성이 상승하였다. 한편 가계부채 부실화를 분석한 선행 연구들의 연구 방법론을 살펴보면 토빗모형을 활용한 연구(e.g. 김현정·김우영, 2009; 박연우·허석균, 2018; 유경원, 2009), 회귀모형분석을 활용한 연구(e.g. 박윤태, 2018), 로지스틱 회귀분석을 활용한 연구(e.g. 이종희, 2019; 유재인·정호성, 2021), 기술적 통계 방법(e.g. 배영목, 2011)을 주로 활용하였다.

종합하자면 가계부채의 문제를 분석한 선행 연구 중 청년층 차입자들을 주요 대상으로 분석한 연구는 존재하였으나, 대부분 모수적 추정방법을 따르는 연구들이었다. 이러한 전통적인 통계 방법은 변수의 확률 분포에 대한 가정이나 조건이 엄격하게 유지되어야 하지만 가계 부채 문제를 결정하기 위한 차입자의 의사결정과정에서는 비선형성과 불확실성이 존재할 수 있다. 이러한 점에서 최근 보다 유연한 추정모형을 구현할 필요성이 제기되었고, 머신러닝 기법의 활용이 주목을 받고 있다. 따라서 머신러닝 기법을 활용하여 청년층 가계의 부채 문제를 다각도에서 분석할 필요가 있다.

본 연구의 목적은 한국의 청년층 차입자를 대상으로 채무 불이행 가능성을 진단하고, 그 가능성에 영향을 주는 요인을

1) 채무 불이행은 일반적으로 공채 및 사채에 대한 이자 지불 혹은 원금 상환이 불가능해진 상황을 의미한다.

예측하는 것이다. 이러한 목적을 달성하기 위하여 다음의 순서대로 연구를 진행하고자 한다. 통계청, 한국은행, 금융감독원이 공동으로 시행한 2021년 「가계금융·복지조사」를 활용하여 한국 청년층의 생활경제실태, 자산 및 부채실태, 채무 불이행 실태를 종합적으로 조망한다. 한편 청년층의 채무 불이행 가능성과 관련된 요인들을 포괄적으로 분석하기 위하여 머신러닝 방법을 활용한다. 본 연구에서 개발한 연구모형을 통하여 도출된 규칙을 근거로 청년층의 가계부채 부실위험과 관련된 함의를 제공한다. 또한 최근 사회과학 분야에서 적극적으로 도입되고 있는 머신러닝 기법이 가계부채 관련 연구에 어떻게 적용될 수 있는지 그 가능성도 검토해 볼 수 있을 것이다.

## II. 이론적 배경

### 1. 청년층 가계부채

가계부채에 대한 연구는 경제학의 다양한 이론들을 적용하며 주로 진행되어 왔다. 경제학에서는 소비이론에 근거하여 가계부채를 설명하는 경향이 크며, 소비이론 중에서는 특히 생애주기이론에 그 기초를 두고 있다. Ando & Modigliani(1963)가 제안한 생애주기이론(life-cycle theory)에 의하면 가계는 대체로 전 생애에 걸친 소득을 고려하여 자원을 배분하는 식으로 현재의 소비를 결정하는데, 소득과 소비에 있어 생애주기에 따라 변화양상이 존재한다고 본다. 이 이론은 개인 또는 가계가 현재의 소득뿐 아니라 미래의 소득을 포함하는 일생 동안 벌게 될 전체 소득을 고려하여 소비수준을 결정한다고 가정하고(김현정·김우영, 2009), 과거로부터 축적해 온 자산까지 소비에 영향을 주는 명확한 요인으로 간주한다(김혜련·최현자, 1999). 개인의 부채는 개인이 효용을 극대화하는 과정에서 발생하게 되며 합리적 가계를 가정할 때 인적자본, 자산, 소득 등 항상소득에 영향을 미치는 요인과 시간선호, 교육수준, 생애주기 등 개인의 선호에 영향을 미치는 요인에 의해 결정된다(김현정·김우영, 2009).

생애주기 이론에 의하면 생애주기의 초반에는 당기 소득이 낮더라도 인적 자본을 토대로 한 미래 소득흐름을 예상하여 소비수준을 높게 유지할 수 있다. 이때 자본시장이 완전하다면 현재 소득과 소비의 괴리는 부(-)의 저축인 부채를 통해 해결될 수 있다. 청년층의 경우 생애주기상 다른 연령대보다 당기 평균소득 수준이 낮은 편이므로 전 생애주기에 걸쳐 소비평준화를 달성하려면 부채를 이용하는 것이 합리적 행동일 수 있다. 일반적으로 생애주기상 이 시기에 재무적으로 독립을 시작하며, 그 과정에서 부채를 활용하게 된다. 주택을 포함한 부동산 매입 등의 이유로 차입을 적극적으로 활용하기도 한다(노혜민·손상희, 2017). 합리적 의사결정을 거친 적정 수준의 차입

은 청년층의 경제적 선택의 폭을 확대하여 현재의 소비를 원활하게 하고, 가계부문의 후생을 증대할 수 있다. 그러나 청년층 가계의 입장에서 부채상환은 고정지출 항목으로 작용하여 이들 가계의 실질적인 소득 감소를 가져올 수 있다. 상환능력을 고려하지 않은 차입은 지불연체와 같이 가계의 재무건전성을 악화시킬 수 있으며, 개인파산과 같이 재정적 파탄을 초래할 수 있다(Dunkin, 2000). 실제로 20대의 파산 신청건수는 2016년은 전체에서 약 1.5%이었고, 2021년에는 약 1.8%를 차지하여 절대적인 수치로는 크지 않았으나, 이 기간에 732건에서 884건으로 약 21%가 증가한 것이었다. 이러한 수치는 전체 연령대의 개인파산 신청 건수가 동기간 5만 5,400건에서 4만 4,000건으로 약 20% 감소한 것과 비교된다(Legal Times, 2021).

한국 청년층의 부채 총량은 다른 연령대에 비하여 크지는 않다. 그러나 최근 들어 부채 증가폭이 크고, 학자금 등 소액으로 시작된 청년층의 신용사용이 고용부진과 대부업체 등의 공격적 영업, 신용불량자 급증으로 악순환으로 이어지는 추세이다(백종호, 2016). 2021년 들어 악화된 고용부진이 청년 체감경제고통지수에 부정적인 영향을 미쳤는데, 아르바이트를 하면서 재취업을 희망하는 자, 경제활동을 하지 않지만 취업의지가 있는 자 등 사실상 실업자를 포함한 청년 체감실업률은 2021년 현재 25.4%로 집계되었으며, 이 수치는 30대의 2.2배, 40대의 2.6배에 해당하였다(전국경제인연합회, 2021). 29세 이하 청년 가구주의 자산 대비 부채 비율은 2015년 16.8%로, '60세 이상'(13.4%)와 함께 가장 낮은 편이었으나 2017년(24.2%)에 전연령대 중 가장 높은 비율을 보였고, 2020년에는 32.5%로 최고치를 경신하였다. 이는 다른 연령대에 비하여 청년층의 부채 증가속도가 자산 증가속도를 압도하기 때문이다. 2015년 청년층 부채는 1,491만원에서 2020년 3,479만원으로 연평균 18.5% 오른 반면, 자산은 연평균 3.9% 증가하는데 그쳤다(한국경제연구원, 2021). 동기간 중 순자산은 전연령대 가운데 청년층만 유일하게 감소하였고, 다른 연령대에서는 순자산액이 최소 6,048만원에서 최대 10,892만원까지 증가한 것으로 나타났다(한국경제연구원, 2021).

### 2. 가계의 채무 불이행 결정요인

가계부채 부실 관련 선행 연구를 고찰한 결과 가계의 채무 불이행에 영향을 미치는 요인으로 대체로 채무자의 인구사회학적 특징, 경제적 특징, 부채 관련 특징들이 나타났다. 선행연구에 포함된 주요 설명변수를 정리한 결과를 <표 1>에 제시하였다.

가계부채 부실위험의 연관 요인을 살펴본 일부의 연구에서는 인구사회학적 특징들이 주요한 요인인 것으로 나타났다. 박윤태와 노정현(2017)은 가계금융복지조사를 이용하여 가구주의 연령별 가계부채 상환위험요인에 대해 분석하였다. 연구결과는 청장년층(20-40대)이 비근로자 일수록 상환위험이 높은

것으로 나타났다. 그 원인으로 청장년층은 소득 및 자산이 적어 경제적으로 불안정한 채무 상황을 들었다. 50대는 부동산 담보대출이 높을수록 상환위험이 높았는데, 이는 이들 연령대가 추가적으로 부동산을 구입하기 위한 자금 마련 방안으로 담보대출을 활용하기 때문인 것으로 보였다. 60대는 은퇴로 인하여 근로 소득이 감소하기 때문에 원리금 연체가 증가할수록 상환위험이 크다고 지적하였다. 김영일과 진경희(2018)은 한국신용정보원의 신용정보와 통계청의 미시자료를 연계한 자료를 활용하여 가계 채무 불이행 결정요인을 분석하였다. 50대 이상 채무자의 부실위험이 확대될 우려가 있다고 밝혔다. 따라서 해당 연령층에 대해서는 금융안전성 측면의 대응과 더불어 일자리대책의 병행이 필요하다고 제안하였다. 장기적으로는 부채의 구조조정 노력과 함께 단기적으로는 위험관리에 주력하는 한편 취약 차주와 금융기관에 대한 모니터링의 강화가 필요하며, 다수의 비은행권 대출을 보유한 다중채무자에 대한 모니터링 강화 및 감독 사각의 축소가 필요하다고 강조하였다. 이종희(2019)는 청년차입자의 부실화 가능성을 진단하고, 부실화 가능성에 영향을 미치는 변수들을 분석하기 위하여 다양한 지표들을 활용하였다. 2018년 가계금융·복지조사를 활용하여 분석한 결과 인구사회학적 변수 중 근무형태가 유의미한 변수이었다. 특히 일용직 근로자 및 자영업자와 같은 근무형태의 차입자의 경우 상용근로자보다 부채부실화 가능성이 높은 것으로 나타났다. 김태은과 양세정(2018)은 청년의 부채 상환에 영향을 미치는 요인을 분석하였다. 인구 통계적 요인, 경제적 요인, 금융역량 요인, 심리적 요인을 포함하여, 부채 경험이 있는 전국의 만 34세 이하 미혼 청년을 대상으로 다변량 분석을 실시하였다. 다른 독립변수들을 통제한 상태에서는 월평균 소득과 총자산이 많을수록, 전세나 월세를 사는 경우 부채상환 가능성이 높아졌다. 반면 금융역량요인이나 심리적 요인은 유의미한 영향요인이 아닌 것으로 나타났다.

가계부채 부실위험에 영향을 주는 요인으로 소득, 자산, 지출과 같은 경제적 특징을 살펴본 연구들도 존재한다. 배영목(2011)은 6년간의 노동패널자료를 활용하여 가계부채 상환 부담의 분포와 추이를 분석하였다. 가계부채의 상환부담은 대체로 소득분위가 낮을수록 증가하나, 최고소득층의 부담도 크다고 밝혔다. 한국의 부채규모의 현격한 증가세는 최고소득층의 부채상환 부담의 증가세와 연관이 큰 만큼 가계대출의 부실징후가 확산되거나 금리인상 등의 거시경제적 충격이 클 경우에는, 저소득층 뿐 아니라 고소득층의 부채상환도 문제가 될 수 있다고 지적하였다. 이동진과 한진현(2017)은 노동패널과 가계금융복지조사를 활용하여 과다부채 가계의 위험요인을 분석한 결과 과다부채 가계는 소득탄력성 및 자산가격탄력성이 높았다. 소득 및 부동산 가치 상승시 더 큰 폭으로 부채를 늘리며 반대로 부동산 가치가 하락시 상환부담이 더 크게 증가하는 것으로 나타났다. 연구결과를 바탕으로 과다부채 가계의 높은 부채비율 등을 감안할 때 개별 가계의 소득 및 자산대비 부채수

준에 대한 대출 심사를 더욱 강화할 것을 제안하였다. 이종희(2019)는 청년차입자의 부실화 가능성에 영향을 미치는 재무적 변수로 가처분 소득, 자산, 소비지출을 들었는데, 특히 가처분 소득과 자산은 차입자의 부채부실화 가능성을 감소시키는 변수이었으며, 주거비와 같은 소비지출은 차입자의 부채부실화 가능성에 증가시키는 변수라고 밝혔다.

일부의 선행연구들은 차입자의 상환여력에 영향을 미치는 요인으로 부채관련 변수들을 분석하였다. 신진욱 외(2019)는 소득감소, 원리금상환액 증가, 자산대비채무액 증가가 가구의 과중 채무상태로의 전이에 끼친 영향을 분석하였다. 2012~2017년 가계금융복지조사 패널자료를 활용한 결과 불운한 생애사건을 경험한 이후 총부채원리금상환비율(Debt Service Ratio, DSR)이 과중채무 위험요인으로 작용하였고, 자산대비부채비율(Debt to Asset Ratio, DTA)은 실업, 장애의 경우 사건 발생 시점에, 사별의 경우 사건발생 이후 과중채무 위험 상승과 연관되어 있음을 밝혔다. 주동현(2019)은 금리변화가 가계의 상환 능력에 미치는 영향을 분석하였다. 금리가 자산구조에 미치는 영향을 총자산 대비 금융자산 비중을 통해 분석하였다. 가계금융복지조사를 활용한 연구결과 정책금리의 변화는 가계 채무 구조에서 실물자산 구입을 위한 자금 조달 비용과 금융자산 운용 수익에 영향을 미치며, 금리 상승이 금융자산 비중 늘리고 실물자산 비중을 줄일 수 있다고 보였다. 이러한 경우 금리 상승은 재산 소득에 부정적 영향을 미치며 부채 측면의 원리금 상환 부담에 더해 채무 상환 능력을 낮추는 요인으로 작용할 가능성이 있다고 주장하였다.

전술한 바와 같이 차입자들의 부채부실 가능성에 영향을 미칠 것으로 예상되는 차입자들의 특징은 인구사회학적 특징, 경제적 여력, 부채 관련 특성으로 나눌 수 있는데, 청년층의 가계 부채 부실화 가능성을 분석하는데 이러한 변수들을 활용할 수 있을 것이다. 연구는 미시자료를 활용하여 취약 청년가계를 중심으로 부채의 부실화 가능성 파악하고 이에 기초한 대응방안을 모색하고자 한다.

종합적으로 선행연구들은 인구사회학적 변수, 경제적 변수, 부채관련 변수 등으로 나누어 가계 부채 상황을 설명하고자 하였다. 선행연구들은 주로 인구사회학적 변수 중에서는 연령, 성별, 종사상 지위를 주요한 변수로 포함하였고, 경제적 변수로는 소득, 자산, 입주형태를 주요한 변수로 포함하였다. 또한 부채 관련 변수 중에서는 대출형태 및 부채비율을 중요한 변수로 선택하였다. 연구방법에서 있어서는 로짓모형, 토빗모형과 같은 다변량 모형을 사용하거나, 기술통계를 사용하는 등 전통적인 선형모형을 사용한 선행연구가 대부분이었다. 이러한 전통적인 통계 방법은 변수의 확률 분포에 대한 가정이나 조건을 엄격하게 유지하여야 하지만, 머신러닝은 보다 유연한 추정모형을 구현할 수 있다. 이러한 면에서 청년 차입 가계의 채무 불이행을 연구하는데 머신러닝 기법을 활용한 시도가 필요한 상황이다(표 1) 참조.

〈표 1〉 선행연구의 주요 설명변수

연구자	인구사회학적 변수	경제적 변수	부채관련 변수	연구자료	연구대상	분석방법
김영일·진경희 (2018)	연령, 가족원수, 성별, 혼인상태, 종사상지위	소득, 입주형태	대출형태, 대출건수, DTI	통계청 가계부채 통합DB	구체적 설명없음	로짓모형
박연우·허석균 (2018)	연령, 종사상 지위	소득, 자산, 입주형태		가계금융 복지조사 (2014)	부채보유가계	토빗모형
박운태·노정현 (2017)	연령, 거주지역, 종사상지위	소득, 자산, 입주형태	대출형태, 부채규모	가계금융 복지조사 (2014-2016)	부채보유가계	기술통계
배영목(2011)		소득, 자산, 입주형태	대출형태, DTI, DTA	한국노동패널자료 (2004-2009)	부채보유가계	기술통계
신진욱·박정민·김영 (2019)	성별, 연령, 혼인상태, 종사상지위	소득, 자산	부채규모	가계금융 복지조사 (2012-2017)	구체적 설명없음	로짓모형
이동진·한진현 (2017)		소득, 자산	대출형태	노동패널 (2013) 가계금융복지 (2014)	구체적 설명없음	조건부 중절분위회 귀모형
이종희 (2019)	성별, 연령, 교육정도, 혼인상태	소득, 자산, 입주형태, 소비지출	대출기관, DTA, DSR	가계금융 복지조사 (2018)	부채보유가계	로짓모형
이종희 (2020)	종사상 지위	입주형태	차입목적, DTA, DSR	가계금융 복지조사 (2019)	차입불 이행응답자	의사결정나무 모형

### 3. 가계부채 연구방법

가계부채 관련 선행연구들은 고찰한 결과 다수의 선행연구들이 통계적 가정을 기본으로 하는 모수적 추정방법을 활용하여 가계부채 보유여부 및 가계부채 규모의 결정요인들을 분석한 것으로 나타났다. 가령 박연우와 허석균(2018)은 통계청의 가계금융복지조사를 활용하여 가계부채의 결정요인을 추정하고 취약계층의 특성을 파악하고자 하였다. 구체적으로 토빗모형을 활용하여 가계부채 결정모형을 추정하고, 주요 취약계층이 가계부채로 인해 겪고 있는 재무적 어려움을 평가하였다. 김현정과 김우영(2009)은 2000~2007년 한국노동패널 자료를 활용하여 가계채무부담이 소비에 미치는 영향을 분석하였다. 토빗 모형을 이용하여 소득 분위, 가구주 연령, 자가보유, 순자산 분위 및 변수들의 교호항의 포함하여 재무곤경을 파악하고자 하였다. 추정 결과 소비에 대한 부채의 전반적인 효과는 정적인 것으로 나타나 가계신용 확대와 같은 유동성 제약 완화가 소비 증가에 영향을 미친 것으로 해석하였다. 유경원(2009)은 자산과 부채에 대한 가계의 의사결정에서 나타나는 상호의존성을 토빗모형을 이용하여 분석하였고, 그 결과 부채규모가 큰 가구의 경우 금융자산 규모가 작아 유동성 측면에서 재무건전성이 취약할 수 있음을 보여주었다. 또한 연구 결과를 바탕으로 부채의 크기가 부동산 구입 등과 유의미하게 관련되어 있다고 밝혔다. 박운태(2018)는 회귀모형을 사용하여 가계부채 상환불능위험도 추정모형을 제시하였다. 연구 결과 20-30대인 2인 가구와 60대 이상인 1인 가구는 자산 5분위에서 가계부채 상환불능 위험도가 높은 것으로 나타났다. 자산 1·2분위의 경우 거주주택 마련 대출비율, 전월세 보증금마련 대출비율 및

생활비 마련 대출비율이 높을수록 가계부채 상환불능 위험도가 높았다.

최근 기술의 발전에 따라 정보처리를 위한 비용이 감소되고 시간이 단축되면서 사회과학 분야에서도 빅데이터를 활용한 연구가 활발히 행해지고 있는 상황이다. 이러한 빅데이터의 등장으로 좀 더 강력한 분석도구가 필요하게 되었는데 (Varian, 2014), 빅데이터의 장점을 활용하기 위해서는 전통적인 선형모형보다 좀 더 유연한 추정모형을 구현할 필요성이 제기되면서 머신러닝 기법의 활용이 주목을 받고 있다. 최근 머신러닝 기법을 활용하여 가계부채 문제를 분석한 연구들이 등장하기 시작하였다. 가령 이종희(2020)는 머신러닝의 분류모형을 통해 가계부채 상환연체 가능성이 높은 가계를 탐지하고자 하였다. 특히 머신러닝의 예측 모형 중에서 분류모형으로 많이 사용되는 모형으로 이항 로지스틱 회귀분석모형과 의사결정나무모형을 이용하였다. 의사결정나무 모델은 비교적 분석이 간단하고, 나무구조를 통해 시각적으로 표현할 수 있다는 장점을 갖는다 (김태진 외, 2018). 하지만 의사결정나무는 몇가지 한계점을 지닌다. 첫째, 의사결정나무는 높은 설명력에 비하여 예측력이 떨어지고, 교차타당화를 실시하는 데도 불구하고 모형이 불안정하다 (유진은, 2015). 둘째, 예측에 사용되는 변수가 늘어날수록 과적합(overfitting) 문제가 발생한다. 반면 랜덤포레스트는 의사결정나무 모델이 갖는 모형의 불안정성과 과적합 문제를 극복할 수 있다는 장점이 있다. 가령 랜덤포레스트는 의사결정 나무들 간 상관을 낮추어 예측오차를 줄이는 방식을 취하며(Breiman, 2001), 잡음(noise)이나 이상치(outlier)로부터 영향을 크게 받지 않는 강건성(robustness)을 가진다(Hamza & Larocque, 2005). 이러한 장점 때문에 머신러닝 중 랜덤포레스트

트 모델을 활용한 사례들이 등장하였다. 가령 박정민과 송태민(2021)은 소셜 빅데이터와 머신러닝을 이용하여 가계부채 부실 위험을 예측하는 모형을 개발하였다. 자료로 2014년-2018년 5년 동안 등장한 총 280개의 온라인 채널에 등장한 채무 관련 문서를 사용하였고, 랜덤포레스트와 의사결정나무 모형을 적용하였다. 연구 결과 머신러닝 알고리즘이 소득, 부채액, 원리금 상환액, 신용거래와 같은 특성 요인들이 가계부채의 부실위험을 예측할 수 있음을 보여주었다.

종합하자면 가계부채를 결정하는 데에는 차입자의 의사결정 과정 속에 포함된 비선형성과 불확실성이 존재할 수 있으므로 기존의 모수적 추정 방법만으로는 현실에 대한 설명력이 떨어질 수 있으므로, 청년층 가계부채의 문제를 다각도에서 분석하기 위하여 다양한 머신러닝 기법을 활용하는 것이 필요하다. 따라서 본 연구는 머신러닝의 한 종류인 랜덤포레스트를 활용하여 최적의 연구모형을 구축하고자 한다.

### III. 연구방법

#### 1. 연구자료

본 연구의 수행을 위하여 「가계금융·복지조사」를 활용한다. 이 자료는 통계청, 금융감독원, 한국은행이 1년마다 전국의 2만여 가구를 대상으로 하며, 가계의 자산, 부채, 소득 등의 규모, 구성, 분포와 미시적 재무건전성을 파악하기 위하여 조사한 자료이다(한국은행, 2021b). 이 자료는 차입자의 원리금 상환여부, 원리금 상환이 생계에 주는 부담과 더불어 차입자의 사회경제학적 특징을 포함하고 있으므로, 청년층의 가계부채 부실위험에 영향을 미치는 요인을 도출하고 규칙을 발견하기 위한 최적의 연구 자료이다. 가장 최신의 연구 자료인 2021년 「가계금융·복지조사」 데이터 사용하여 연구를 수행하였다.

#### 2. 연구문제

연구문제 1. 청년층의 전반적인 생활 경제 실태, 자산 및 부채 실태, 채무 불이행 실태는 어떠한가?

연구문제 2. 가계 채무 불이행 위험이 높은 취약 청년 가계를 분류하고 예측할 수 있는 연구 모형은 무엇인가?

#### 3. 연구방법

본 연구는 청년층 차입자들의 가계부채 부실위험을 예측하기 위하여 랜덤포레스트(Random Forest) 예측 모형을 구축한다. 전통적인 회귀분석에서는 변수들간의 외생성, 오차항의

분포, 등분산성 등의 가정이 전제되어야 하는데 다양한 독립변수를 연구모형에 포함시키는 경우 이를 만족시키는 것은 어렵다. 연구모형에 포함된 설명변수들이 높은 상관관계를 보이는 경우 그 외 설명변수들이 갖는 고유한 설명력이 매우 낮아져서 설명변수가 종속 변수에 미치는 독자적인 영향력이 회귀계수에 제대로 반영되지 않아 분석 결과의 해석에 한계가 발생한다(오주한·정석원, 2013). 그러나 랜덤포레스트는 예측인자와 종속변수가 서로 관련되어 있다는 가정하에 설명변수가 종속변수 예측에 어떤 역할을 하는지 분석하기에 까다로운 회귀분석의 가정에서 자유롭다는 장점이 있다(최필선·민인식, 2018).

랜덤포레스트는 Breiman(2001)에 의해 제시된 앙상블 학습(ensemble learning)으로 부트스트랩(bootstrap) 방식을 이용하여 다수의 표본을 생성하고 의사결정나무 모형에 적용하여 그 결과를 종합하는 방법으로 형태이다(서종덕, 2016). 이 모형은 의사결정나무 모형의 단점을 보완한 기법으로 예측력이 뛰어나며 과적합 문제도 개선된 결과를 제시한다(Breiman, 2001). 또한 이 모형은 실무적인 유용성이 뛰어나는데, 다양한 예측 인자들의 중요도 지수(importance index)를 활용하여 종속변수를 설명하는 변수들의 상대적인 예측력을 비교하는 것이 가능하며, 부분 의존성 도표는 설명변수 값이 종속변수 예측에 어떻게 영향을 보여주는지 나타낸 그래프로 개별의 독립변수와 종속변수 간의 영향력의 패턴을 시각적으로 보여줄 수 있다(Hastie, Tibshirani & Friedman, 2009).

본 연구에서는 랜덤포레스트 기법을 활용하여 청년층 차입자의 채무 불이행 위험을 예측하는 모형을 탐색한 뒤 중요도 지수를 산출하고, 중요도가 높은 설명변수들을 선별한 뒤, 주요 결정요인들의 부분 의존성 도표를 제시하고자 한다. 이를 통해 청년층 차입자의 채무 불이행 개선을 위해 우선적으로 집중해야 하는 지표들을 확인하고, 시사점을 제공하고자 한다.

#### 4. 주요변수

##### 1) 청년층 차입자

청년을 정의할 때 사회 경제적 의미를 담기도 하지만 일차적 의미로 특정 연령대에 포함된 인구 집단 전체를 의미한다. 통계청은 15~29세의 남녀를 청년으로 간주하나, 법령과 조례에 따라 청년의 기준이 되는 연령을 다르게 적용하는 경우도 있다. 일반적으로 20대와 30대를 청년으로 정의하는데, 대통령 직속 청년위원회는 19~39세까지의 남녀를 청년으로 규정하였고(최효미 외, 2016), 한국청소년 정책연구원(2020)은 청년의 사회·경제 실태를 분석하기 위하여 청년층을 15세에서 39세로 한정하였다. 일반적인 청년의 연령대보다 최대한 포괄적인 연령(15~39세)으로 조사하였는데, 법률적으로 취업이 가능한 최저연령인 15세를 청년의 시작으로 보았고, 취업, 결혼이나

출산 등과 관련해 최고연령으로 39세로 삼았다. 이러한 경향에 따라 본 연구는 다음과 같이 청년층 차입자를 정의하였다. 본 연구의 자료인 가계금융복지조사가 조사한 19세 이상의 응답자부터 이루어지므로, 19세 이상 39세 이하의 조사대상자를 ‘청년’으로 정의한다. 또한 채무 불이행 관련 문항에 응답한 조사대상자를 ‘차입자’라고 정의하였다.

가계금융복지조사의 채무 불이행 관련 문항에 응답한 자들 중 19세 이상 39세 이하의 응답자들을 청년층 차입자로 정의하였고, 1,596가계가 이에 해당하였다. 랜덤포레스트를 시행하기 위한 데이터 전처리 과정에서 독립변수들 중 이상치가 존재함을 발견하였다. 주요변수가 이상치를 갖는 경우 왜곡된 검정 결과가 도출될 수 있으므로, 이상치를 제거하는 것이 바람직하다고 보고, 이러한 가계를 데이터에서 제거하였다. 구체적으로 가처분 소득이 음수가 나온 6가계를 제거하였으며, 최종적으로 1,590가계를 분석에 이용하였다.

## 2) 종속 변수

가계부채 부실위험의 측정은 대개 가계의 부채상환 능력의 분석에 기초한다(박정민·송태민, 2021). 청년층 가계의 부채 부실화를 평가하기 위하여 응답자의 가계부채 상환연체 여부를 활용하였다. 구체적으로는 다음의 문항을 사용하였다. “지난 1년간 귀책에서는 원금을 상환하거나 이자를 낼 날짜를 지나친 적이 있습니까?” 이 항목에서 “있음”이라고 응답한 경우 부채부실 위험이 있는 것으로 간주하였고, 이 항목에서 “없음(1)”이라고 응답한 경우 부채부실 위험이 없는 것(0)으로 간주하였다.

## 3) 독립 변수

가계부채 부실화를 예측하는 설명변수는 크게 (1)인구사회학적 변수, (2)경제적 변수, (3)부채변수로 나누어 연구모형에 포함하였다. 설명변수의 구성을 살펴보면, 인구사회학적 요인은 성별, 거주지역, 종사상 지위, 입주형태, 혼인상태, 교육정도를 포함한다. 경제적 요인으로는 소득, 자산, 목적별 소비지출비율을 포함하였다. 부채관련 변수로 자산부채 부채비율(Debt to Asset Ratio, DTA)과 원리금상환비율(Debt Service Ratio, DSR)을 포함하였다. 원리금상환비율(DSR)은 처분가능 소득 중 원리금상환액이 차지하는 비율을 의미한다. DSR은 소득에 대비하여 원리금의 상환부담이 높은 상황에서 채무상환의 요구가 집중될 경우 실물자산의 처분 없이는 단기간에 부채를 상환할 수 없는 가구를 식별할 수 있다는 점에서 단기부채 변수로 활용할 수 있다. DTA는 가계가 위기 상황에 직면하였을 때 보유한 자산의 유동화를 통하여 소비평활화를 가능하게 하는지(남상호, 2015) 파악할 수 있으므로, 이 비율을 통하여 장기적인 가계부채의 위험성을 평가할 수 있다. 또한 DSR와 DTA 두 기준지표의 합성합수형식을 사용하여, 가계의 소득,

금융, 실물자산을 종합적으로 고려한 가계부실위험지수(Household Default Risk Index; HDRI)도 포함한다. 이러한 합성합수를 독립변수로 사용할 경우 독립변수가 강한 상관관계가 발생하여 다중공선성(multicollinearity)이 나타날 수 있다. 단, VIF를 확인해 본 결과 10을 넘는 변수가 없는 것을 확인하여, 합성합수 변수를 연구모형에 포함하였다. 한편 청년 차입가계가 어떠한 용도로 차입을 하였는지를 파악하기 위한 변수들도 포함하였다. 가령 부채지표 중 부실위험지수(HDRI)는 가구의 소득 흐름은 물론 금융 및 실물 자산까지 종합적으로 고려한 지표로 정부는 가계가 100을 초과하는 가구일 경우 다른 가구보다 상대적으로 취약한 “부실위험”가구로 판단하였다. 한편 정부가 가계대출 규제를 본격화하면서 적용하기 시작하며 고 DSR 가구(DSR > 40%), 고 DTA 가구(DTA > 100%)으로 구분하였다. 최종적으로 본 연구의 연구모형에 예측 모형에 투입된 변수는 <표 2>와 같다.

## IV. 연구결과

### 1. 연구대상자의 전반적 특징

#### 1) 생활경제실태

연구대상자의 인구사회학적 특징을 정리한 결과를 <표 3>에 제시하였다. 본 연구의 관심대상인 19세 이상 39세 이하와 40세 이상의 응답자의 결과를 구분하여 제시하였다. 전체 가계는 9,362가계이고, 그 중 청년 응답자는 1,590가계이었다. 변수의 유무에 따라 통계적으로 유의미한 차이가 있는지 평가하기 위해 카이제곱분석 및 T-test를 진행하였다. 분석결과 수도권 거주여부, 교육수준, 종사상 지위, 혼인상태, 입주형태, 처분가능소득, 자산의 경우 연령 간 유의미한 차이가 존재하였다. 청년 응답자의 약 41%는 수도권에 거주하고 있으며, 약 79%는 대학을 졸업하였으며, 약 72%는 상용근로자이었고, 약 61%는 기혼이었고, 약 41%는 자가를 소유하였다. 소득 및 자산 평균은 각각 5,020만원과 3억 5,095만원이었다. 40세 이상 응답자의 경우 약 34%는 수도권에 거주하고 있으며, 약 43%는 대학을 졸업하였으며, 약 44%는 상용근로자이었고, 약 74%는 기혼이었고, 약 68%는 자가를 소유하였다. 소득 및 자산 평균은 각각 5,405만원과 5억 4,334만원이었다. 마지막으로 청년층의 부채 평균은 1억 1,304만원이었으며, 40세 이상의 부채 평균은 1억 2,726만원이었다. 청년층과 40세 이상의 연구대상자를 비교하면 청년층은 대출 이상, 상용근로자, 수도권 거주 비율이 통계적으로 더 높았고, 기혼비율, 자가 보유 비율은 청년층이 더 낮은 편이었다. 처분가능소득, 자산, 부채 모두 40세 이상의 연구대상자가 청년층보다 높았다.

〈표 2〉 설명변수 설명

변수	구분	측정	
인구사회학적 변수	성별	남자, 여자	
	교육정도	대졸이상, 그 외	
	혼인상태	기혼, 그 외	
	수도권여부	수도권, 그 외	
경제적 변수	종사상지위	상용근로자, 그 외	
	입주형태	자기집, 그 외	
	목적별 소비지출 비중	식료품지출목적	식료품/가처분소득
		주거비지출목적	주거비/가처분소득
		교육비지출목적	교육비/가처분소득
		의료비지출목적	의료비/가처분소득
		교통비지출목적	교통비/가처분소득
통신비지출목적	통신비/가처분소득		
부채지표	DTA	DTA=부채/자산	
	DSR	DSR=부채원리금상환액/처분가능소득	
	HDRI	HDRI=100·[1+(DSR-0.4)]{1+(DTA-1)}	
부채 관련 변수	목적별 부채 비중	거주주택구입 목적	거주주택구입 목적 부채/금융부채
		거주이외 부동산 구입 목적	거주이외 부동산 구입 목적 부채/금융부채
		전월세보증금 목적	전월세보증금 목적 부채/금융부채
		부채상환 목적	부채상환 목적 부채/금융부채
		사업자금 목적	사업자금 목적 부채/금융부채
		생활비 목적	생활비 목적 부채/금융부채

〈표 3〉 연구대상자의 일반적 특징

변수	항목	19-39세 (n= 1,590)		40세이상 (n=7,772)		× 2	T
		빈도	%	빈도	%		
수도권	수도권	651	40.94	2,640	33.97	28.17	***
	그외	939	59.06	5,132	66.03		
성별	남성	1,261	79.31	6,207	79.86	0.25	
	여성	329	20.69	1,565	20.14		
교육정도	대졸이상	1,259	79.18	3,378	43.46	673.67	***
	그외	331	20.82	4,394	56.54		
종사상지위	상용근로자	1,156	72.70	3,381	43.50	450.67	***
	그외	434	27.30	4,391	56.50		
혼인상태	기혼	972	61.13	5,718	73.57	100.15	***
	그외	618	38.87	2,054	26.43		
입주형태	자가	663	41.70	5,310	68.32	405.11	***
	그외	927	58.30	2,462	31.68		
처분가능연소득		5,020.81만원		5,405.02만원			-3.23**
자산		35,095.89만원		54,334.83만원			-9.80***
부채		11,304.81만원		12,726.51만원			-2.43*

\*p<.05, \*\*p<.01, \*\*\*p<.001



청년층 가계의 채무 불이행 가능성을 조망하기 위하여 청년층 가계가 전체 가계에서 차지하는 소득 및 자산의 비중을 조망한다. <표 4>는 부채보유 여부와 상관없이 청년층의 가계소득, 자산, 부채 점유율을 나타낸 표이다. 연구 결과 청년층의 인구 규모는 전체 조사대상자의 16.9%를 차지하고 있었는데, 소득(경상소득, 근로소득, 사업소득, 재산소득, 이전소득을 포함)은 전체의 약 16%를 차지하였으며, 자산은 전체 자산의 약 12%, 부채는 15%를 차지하였다. 청년층이 전체 연령대에서 차지하는 인구 비중을 고려하면, 청년층의 소득 비중, 자산 비중, 부채 비중은 그보다 낮은 편이었다.

<표 4> 청년층의 소득, 자산, 부채 점유율

	소득	자산	부채
비중	15.96%	11.67%	15.37%

청년층 차입자들의 소비지출 비중에 대한 특징은 <표 5>에 제시하였다. 소득대비 소비지출 비중을 살펴보면 연령대를 막론하고 가장 많은 비중을 차지하는 소비지출은 식료품비였으며, 이후 기타지출, 주거비 순으로 나타났다. 통계적으로 차이

가 있는 소비지출 비중은 주거비지출, 의료비지출, 기타지출이었다. 청년층의 경우 40세 이상 응답자들보다 주거비 지출과 기타지출 목적의 소비지출 비중이 높았으며, 40세 이상 응답자들은 의료비 지출 목적의 비중이 높았다.

## 2) 가계부채 관련 실태

청년층의 DTA, DSR, HDRI평균은 0.48, 0.30, 45.99 이었고, 40세 이상의 DTA, DSR, HDRI평균은 1.65, 0.27, 126.07 이었는데, DSR의 경우 청년층과 40세 이상 조사대상자 간 통계적으로 유의미한 차이는 없는 것으로 나타났다. 추가적으로 <표 6>는 연구모형에 포함된 DSR, DTA 및 HDRI로 활용된 차입자들의 가계부채비율을 계산한 결과이다. 청년층 가계의 경우 모든 자산을 매각하여도 부채 상황이 어려운 청년층(1(DTA)은 5%, DSR이 0.4를 초과하는 가계는 약 16%, 가계의 채무상환능력을 소득과 자산의 측면에서 동시에 측정할 경우에는 청년층의 약 6%가 위험도가 높은 것으로 나타났다. 40세 이상 가계의 경우 자산매각을 통해서도 부채상황이 어려운 응답자는 약 5%이었다. 최근 정부가 가계대출 본격화를 하면서 DSR 40%를 적용하고자 하는데, 이 수치를 초과하는 가계는 약 17%, 가계부실위험지수(HDRI) 100의 기준을 초과

<표 5> 연구대상자의 경제적 특징 (소비지출)

변수	19-39세		40세 이상		T
	빈도	%	빈도	%	
소득대비 소비지출 비중	식료품지출 목적	0.33	0.26	2.05 <sup>*</sup>	
	주거비지출 목적	0.18	0.11	3.03 <sup>**</sup>	
	교육비지출 목적	0.04	0.08	-1.36	
비중	의료비지출 목적	0.03	0.06	-3.82 <sup>***</sup>	
	교통비지출 목적	0.09	0.08	0.45	
	통신비지출 목적	0.07	0.05	2.02 <sup>*</sup>	
	기타지출 목적	0.26	0.18	3.13 <sup>**</sup>	

<sup>\*</sup>p<.05, <sup>\*\*</sup>p<.01, <sup>\*\*\*</sup>p<.001

<표 6> 연구대상자의 가계부채 특징 (부채지표)

변수	19-39세		40세 이상		sig.
	빈도	%	빈도	%	
DTA	DTA<0.4	909	57.17	5,667	72.92
	0.4≤DTA≤1	600	37.74	1,651	21.24
	1(DTA)	81	5.09	454	5.84
DSR	DSR<0.2	1,037	65.22	4,951	63.70
	0.2≤DSR≤0.4	292	18.36	1,485	19.11
	0.4(DSR)	261	16.42	1,336	17.19
HDRI	HDRI<32	902	56.73	5,495	70.70
	32≤HDRI≤100	589	37.04	1,766	22.72
	100(HDRI)	99	6.23	511	6.58

<sup>\*</sup>p<.05, <sup>\*\*</sup>p<.01, <sup>\*\*\*</sup>p<.001

한 청년층 위험가구는 6%이었다. 연구대상자의 연령에 따른 부채 비율의 차이를 알아보기 위해 카이제곱 검증을 실시하였다.

〈표 7〉은 연구대상자들의 목적별 부채 비중을 정리한 표이다. 청년층 차입자의 경우 거주 주택마련(35%), 전월세 보증금 목적(32%) 및 생활비 마련을 위한 부채(12%) 비중이 높은 편이었다. 목적별 부채 비중 중 전월세 보증금 목적으로 차입한 부채의 비율이 상당히 높았다. 40세 이상 차입자의 경우 거주 주택마련(36%) 비중이 가장 높았으며, 그 다음으로 생활비 마련을 위한 부채 비중(17%)과 사업자금 목적으로 이용한 부채 비중(17%)도 높은 편이었다. 청년층 차입자들이 높았던 부채 비중은 전월세보증금 목적 부채비중인 것으로 나타났고, 40세 이상 차입자들이 높았던 부채 비중은 거주이외 부동산 목적 부채, 부채상환목적 부채, 사업자금 목적 부채, 생활비 목적부채인 것으로 나타났다.

〈표 8〉은 가계부채 상환 여부를 묻는 문항에 대한 응답을 정리한 표이다. 청년층 응답자중 7.9%와 40세 이상의 차입자의 8.1%는 원리금 연체가 있다고 응답하였다. 연령대 별로 유의미한 차이가 존재하지 않았다( $p=.8353$ ).

## 2. 랜덤포레스트 분석

### 1) 의사결정트리 생성 과정 반복

분류 문제를 해결하기에 앞서 데이터의 분포를 확인한 결과 본 연구에서 사용한 데이터의 경우 예측해야 하는 결과값의 분포가 8:2 정도로 불균형한 분포를 띄고 있는 것으로 나타났다. 불균형 문제를 해결하지 않으면 연구모델은 가중치가 높은 집

단을 더 예측하려고 하기 때문에 정확도(accuracy)는 높아질 수 있지만, 분포가 작은 값에 대한 정밀도(Precision)는 낮을 수 있으며, 분포가 작은 집단의 재현율(recall)이 낮아지는 문제가 발생할 수 있다. 따라서 본 연구에서는 이러한 클래스 불균형 문제를 해결하기 위하여 몇 가지 샘플링을 진행하였다. 구체적으로 Oversampling, Downsampling, SMOTE(Synthetic Minority Over-sampling Technique), ROSE(Random Over Sampling Examples)sampling을 시도하였는데, 분류정확도를 개선하는데 있어서 ROSE sampling이 가장 우수한 것으로 나타나 이를 사용하였다.

랜덤포레스트는 표본과 변수를 무작위로 선택해 의사결정 트리를 생성하는 과정을 반복한다. 청년층 차입자의 부채부실 위험을 예측하는 랜덤포레스트 모형의 구축에는 나무 개수의 기본값인 500개( $n_{tree}=500$ )을 생성하도록 하였다. 500개의 나무 개수를 통하여 청년차입자의 상환연체를 예측하는데 충분한지를 확인하기 위해 나무의 개수에 따른 예측 오차 비율을 점검해 보았다. 또한 채무부실 위험 모형의 변수 선별을 위하여 각 의사결정나무 모형의 마디에서 사용하는 변수 개수의 최적화 값은 Brieman(2001)이 제안한 예측변수 채택 방식을 사용하여, 4개( $\sqrt{21} = 4.58$ )로 설정하였다. 4개의 마디를 사용하여 각 분할에서 4개의 변수가 무작위로 자동생성하도록 하였다. 〈그림 1〉에서 보는 바와 같이 94에서 OOB error(Out of Bag error)의 값이 가장 낮게 확인되어 예측 오차를 수렴하는데 의사결정나무를 500개로 설정하는 것이 충분하다고 볼 수 있다.

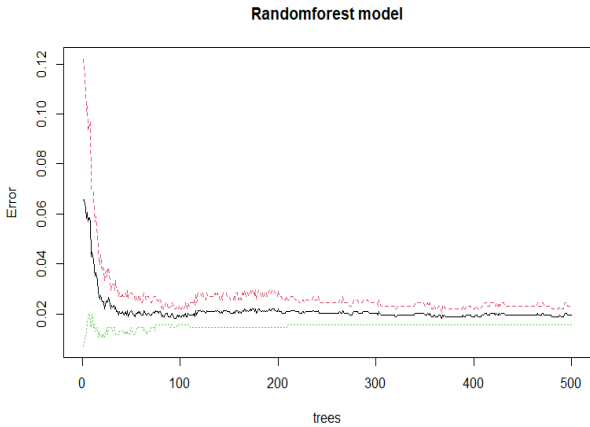
〈표 7〉 연구대상자의 차입 목적

변수		19-39세	40세이상	T
목적별 부채 비중	거주주택구입 목적 부채	0.35	0.36	0.93
	거주이외부동산 목적 부채	0.05	0.09	-4.64***
	전월세보증금 목적 부채	0.32	0.09	25.12***
	부채상환목적 부채	0.02	0.04	-3.06**
	사업자금 목적부채	0.06	0.17	-9.13***
	생활비 목적 부채	0.12	0.17	-4.55***
	기타목적 부채	0.08	0.08	1.44

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

〈표 8〉 연구대상자의 채무 불이행

변수	19-39세		40세 이상		x <sup>2</sup>	
	빈도	%	빈도	%		
원리금 연체	있음	126	7.92	628	8.08	0.0432
	없음	1,464	92.08	7,144	91.92	



〈그림1〉 채무 불이행 위험

전술한 바와 같이 Oversampling, Downsampling, SMOTE, ROSE를 사용하였는데, 분류정확도가 가장 많이 개선된 ROSE sampling 결과는 다음과 같았다. 민감도는 65%로 나타났으며, 특이도는 96%로 나타났다. 종합적인 모형의 정분류율은 78%였다. 순차적으로 모형에 시험자료를 투입하여 결과를 재확인하였다. ROC는 False Positive Rate에 따른 True Positive Rate를 의미하는데 즉, 민감도에 따른 1 - 특이도를 의미한다. 일반적으로 ROC가 0.9~1이면 ‘excellent’, 0.8~0.9이면 ‘good’, 0.7~0.8이면 ‘fair’, 0.6~0.7이면 ‘poor’로 해석할 수 있는데, 채무부실 위험의 ROC는 0.91로 나타나, 모형의 성능이 훌륭하다(excellent)고 볼 수 있다.

## 2) 중요도 지수를 활용한 변수선별

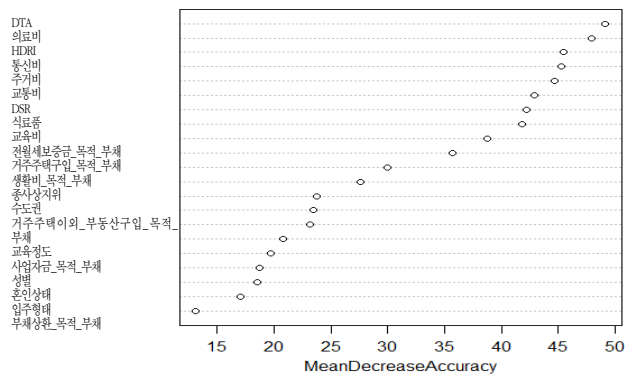
분류에 영향을 미치는 변인 선별을 위하여 중요도 지수(variable of importance index)를 활용하였다. 청년층의 채무 불이행 가능성을 예측하는 설명변수의 중요도를 확인하기 위하여 모형 정확도 개선지수(Mean Decrease Accuracy: MDA)와 지니개선지수(Mean Decrease Gini: MDG)를 활용하여 예측력이 높은 상위 변수들을 도출하였다. MDG는 MDA에 비해 상대적으로 편의(bias)가 크며, 추정 결과가 불안정하다는 한계를 갖는 것으로 알려져 있다(Sandri & Zuccolotto, 2010). 이에 본 연구는 MDA와 MDG에 따른 결과를 모두 제시하나, 분석 결과 해석은 MDA 값을 활용하였다. 중요도 지수 값은 해당 변수의 예측 영향력이 클수록 그 값이 증가하는데, 일반적으로 최소 기준값으로 3으로 한다(Strobl, Malley & Tutz, 2009). 본 연구에서 활용한 데이터를 예측모형에 투입한 결과 청년층 가계부채 위험의 모든 설명변수의 중요도값이 3이상으로 나타나 예측력을 해석하기 위한 최소 기준값을 초과하는 것으로 나타났다. 청년층 가계부채의 부실을 분류하는데 있어서 중요한 지수가 높은 변수들을 정리한 결과는 〈표 9〉과 같다. 분석 결과를 살펴보면 부채 관련 변수 중 DTA(MDA=49.13), HDRI(MDA

=45.51), DSR(MDA=42.17)이 순서대로 중요한 변수로 관찰되었고, 차입한 목적에 따른 부채 비중은 중요도가 높은 변수는 아니었다. 채무 관련 변수 중 의료비 비중(MDA=47.92), 통신비 비중(MDA=45.23), 주거비 비중(MDA=44.71), 교통비 비중(MDA=42.90) 순으로 주요한 변수임이 확인되었다. 그 외 성별(MDA=18.72), 혼인상태(MDA=18.56), 입주형태(MDA=17.07)와 같은 인구 사회학적 변수들은 주요한 변수가 아닌 것으로 나타났다.

〈표 9〉 설명변수의 중요도

영역	변수	MDA	MDG
인구사 회학적 변수	성별	18.72	6.05
	교육정도	20.76	13.70
	혼인상태	18.56	6.97
	수도권여부	23.50	7.82
경제적 변수	총자산지위	23.72	16.40
	입주형태	17.07	6.78
	식료품	41.86	56.26
	주거비	44.71	80.24
	교육비	38.79	35.30
	의료비	47.92	50.93
	교통비	42.90	71.82
	통신비	45.23	75.46
	DTA	49.13	94.15
부채 관련 변수	DSR	42.17	59.86
	HDRI	45.51	83.88
	거주주택구입 목적 부채 비중	30.00	22.20
	거주이외 부동산 구입 목적 부채 비중	23.18	10.43
	전월세보증금 목적 부채 비중	35.67	30.12
	부채상환 목적 부채 비중	13.13	3.34
	사업자금 목적 부채 비중	19.69	10.09
생활비 목적 부채 비중	27.64	25.57	

본 연구에서 구축한 청년차입자의 채무 불이행 위험을 예측하는 랜덤포레스트 모형을 설명하는 변수들의 중요도 지수 결과는 〈그림 2〉와 같다.



〈그림 2〉 청년층 채무 불이행 중요도 지수

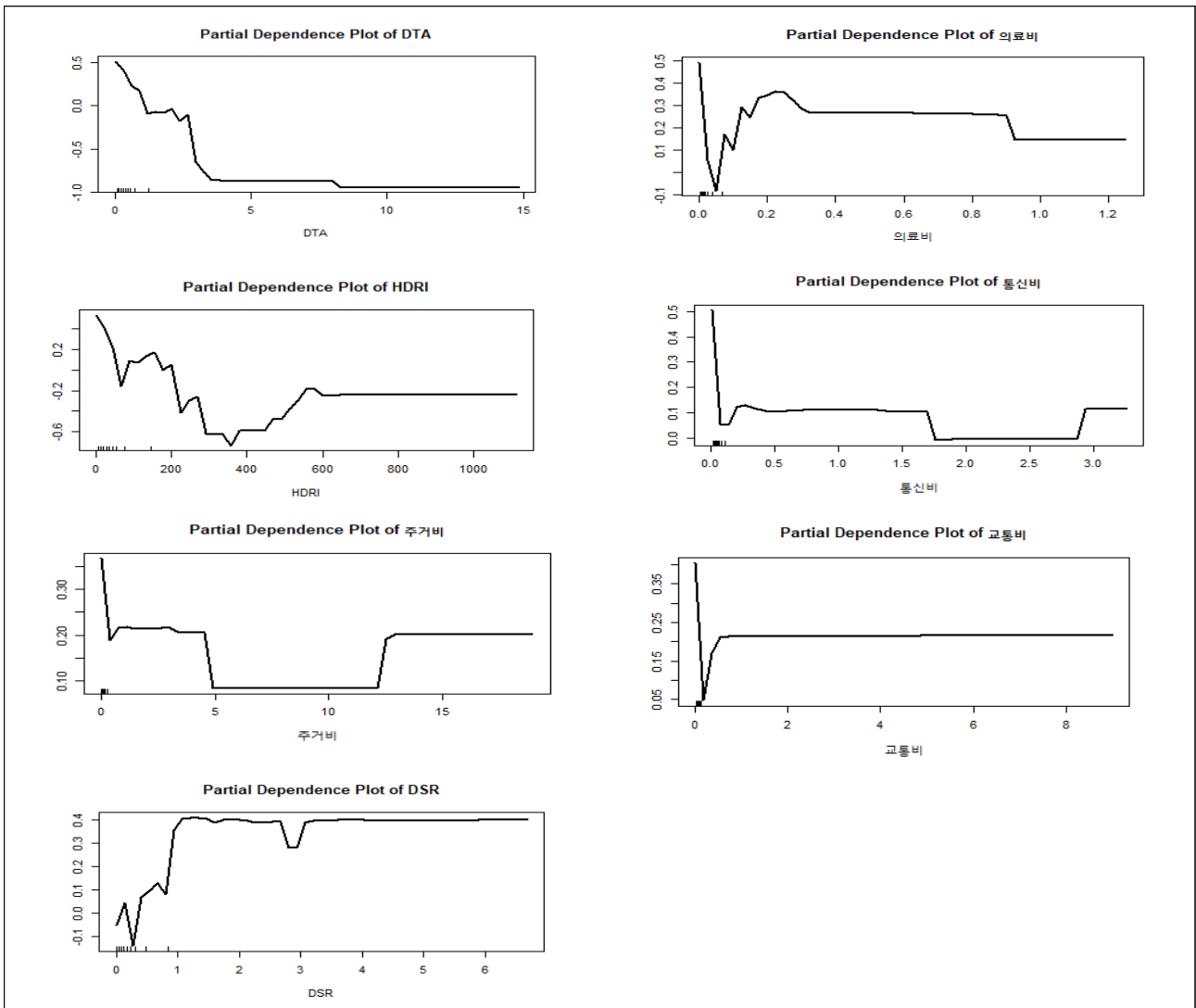
### 3) 부분 의존성 도표

각각의 독립변수가 종속변수에 미치는 영향력을 보여주기 위하여 부분 의존성 도표(partial dependence plot)를 작성하였다. 중요도 지수가 높았던 상위 30%의 변수들은 DTA, 의료비, HDRI, 통신비, 주거비, 교통비, DSR 이었으며, 이들 변수의 부분 의존성 도표를 순서대로 제시하였다(〈그림 3〉참조).

부채관련 변수 중에서는 DTA, HDRI, DSR이 주요한 독립변수이었는데, HDRI는 75지점부터 채무 불이행 가능성이 급상승하였다. DSR은 30% 이후 채무 불이행이 상승하다가, 1이 되는 지점(즉, 소득을 모두 부채 상황에 사용하는 수준)이 되면 채무 불이행 가능성이 급상승하였다. 그러나 자산 대비 부채 비율(DTA)은 주요한 독립변수이면서도 예상과는 다르게 자산에 비하여 부채 비율이 증가한다고 채무 불이행 가능성이 증가하는 것은 아니었다.

〈표 4〉에서도 나타나듯 청년층 차입자들은 생애주기상 자산의 축적이 많이 이루어진 시기가 아니었다. 40세 미만의 청년 응답자의 자산은 40세 이상의 응답자들보다 낮은 수준이었다. 이러한 점에서 청년층 차입자들은 자신의 현재 보유한 자산보다는 현재 벌어들이고 있는 당기 소득이 채무 불이행에 중대한 영향을 미친다고 볼 수 있다.

채무관련 변수 중 의료비, 통신비, 주거비, 교통비의 비중이 특히 중요한 독립변수로 나타났다. 단, 해석을 위해서는 가처분 소득 대비 해당 소비지출이 2배를 넘는 극단적 사례들의 한계효과는 고려하지 않았다. 부분 의존성 도표를 살펴보면 의료비의 경우 약 10% 수준까지 채무 불이행 가능성이 급격하게 증가하였고, 통신비의 경우 약 20% 수준까지 채무 불이행 가능성이 급격하게 증가하였다. 주거비의 경우 약 60% 수준까지, 교통비의 경우 50% 수준까지 채무 불이행 가능성이 크게 상승하였으며, 이후 지속적인 상승세를 유지하였다.



〈그림 3〉 부채위험에 영향을 미치는 상위설명변수의 부분 의존성 도표

## V. 결론 및 함의

한국의 가계부채가 지속적으로 증가하는 가운데 가계부채 문제가 한국경제의 중대한 위험 요인으로 대두되고 있다. 부동산 가격 상승과 2020년에 발생한 전 세계적 보건위기와 같은 악재로 경제 상황이 어려워졌고, 시중 금리가 상승하는 상황에서 사상 최대로 늘어난 가계부채가 한국 경제의 뇌관으로 작용할 수 있다는 우려가 커지고 있다. 한국 가계대출의 누증과 함께 청년층 및 저소득층을 포함한 취약계층을 중심으로 한 채무 불이행에 대한 우려도 크다. 최근 경제 악화와 취업난 속에서 청년층의 부채 증가는 이들의 채무 불이행에도 심각한 영향을 미친 것으로 나타났다. 청년층의 가계부채 건전성이 최근 더욱 취약해진 것은 몇 가지 이유로 설명할 수 있다. 첫째, 최근의 경기침체와 함께 전반적인 고용위축은 신규구직 계층인 청년에게 크게 작용할 수 있다(한요셉, 2020). 구직에 걸리는 시간이 증가하면서 생활비 충당을 위하여 대출을 받는 비율도 크게 증가하였다(서울시복지재단, 2022). 둘째, 신용이 악화된 청년층의 경우 금리가 높은 제2금융권에서 대출을 받을 수밖에 없는 구조적인 환경에 놓이게 되는 것도 한 원인이다. 셋째, 저금리 환경에서 투자를 목적으로 돈을 차입하여 주식 혹은 가상 자산 등에 투자한 청년들의 경우 최근의 자산가격 조정에 따라 경제적, 심리적으로 어려움을 겪기도 한다(금융위원회, 2022).

2000년대 후반 세계 금융 위기 이후부터 이루어진 가계부채 관련 선행 연구를 고찰한 결과 채무 불이행 위험에 영향을 주는 변수들은 크게 세 가지로 나눌 수 있었다. 첫째는 인구사회학적 변인으로 종사상 지위, 연령과 같은 변수가 중요하였다. 둘째는 경제적 변인으로 소득, 자산, 및 소비지출 등이 중요한 변수이었다. 셋째는 부채 관련 변수로 원리금 상환액, 부채 총량, 부채 비율, 금리 등이 채무 불이행 가능성에 영향을 미치는 것으로 나타났다. 한편 채무 불이행을 분석한 선행 연구들의 연구 방법론을 살펴보면 토빗모형, 선형회귀모형, 로지스틱 회귀모형, 기술적 통계방법 등 모수적 추정방법을 따른 연구들이 대부분이었다. 가계 부채의 부실화를 분석한 선행 연구들 중에서 청년층 차입자들을 주요 대상으로 분석한 연구는 존재하나, 머신러닝 기법을 활용하여 청년층 가계부채의 문제를 분석한 연구는 드문 편이었다.

본 연구의 목적은 한국의 청년 차입자를 대상으로 채무 불이행 가능성을 진단하고, 그 가능성에 영향을 주는 요인을 예측하는 것이다. 이러한 목적을 달성하기 위하여 통계청, 한국은행, 금융감독원이 공동으로 시행한 2021년 「가계금융·복지조사」를 활용하여 한국 청년층의 일반적 생활경제실태, 자산 및 부채실태, 채무 불이행 실태를 종합적으로 조망하였다. 또한 선행연구의 연구 방법론적 단점을 보완하고, 청년부채 문제에 대한 의미 있는 해석을 위한 선례를 제공하고자 하였다. 이를 위하여 머신러닝 알고리즘의 하나인 랜덤포레스트 방법을

활용하여 청년층의 채무 불이행 가능성과 관련된 요인들을 분석하였다. 본 연구에서 개발된 연구모형을 통하여 도출된 규칙을 근거로 청년층의 가계부채 부실위험과 관련된 함의를 제공하고자 하였다. 또한 최근 사회과학 분야에서 적극적으로 도입되고 있는 머신러닝 기법이 가계부채 관련 연구에 어떻게 적용될 수 있는지 그 가능성도 살펴보려고 하였다.

청년층 차입자의 자산, 부채 및 채무 불이행 실태를 살펴보면 다음과 같았다. 청년층 가계의 경우 모든 자산을 매각하여도 부채 상황이 어려운 청년층(1(DTA)은 5%, DSR 이 0.4를 초과하는 가계는 약 16%, 가계의 채무상환능력을 소득과 자산의 측면에서 동시에 측정할 경우에는 위험 정도가 높은 청년층은 약 6%이었다. 청년층 차입자의 목적별 부채 비중을 살펴보면 거주 주택마련(35%), 전월세보증금 목적(32%) 및 생활비 마련을 위한 부채(12%) 비중이 높은 편이었다. 청년층 차입자의 채무 불이행을 살펴보면, 응답자 중 7.9%가 이에 해당하는 것으로 나타났다. 연령별로 유의미한 차이가 존재하지는 않았다.

랜덤포레스트를 활용하여 청년층 차입자의 가계부채 부실을 분석한 결과는 다음과 같았다. 우선 본 연구에서 사용한 데이터의 경우 예측해야 하는 결과 값의 분포가 8:2 정도로 불균형한 분포를 띄고 있는 데이터이었다. 이러한 불균형 데이터 문제를 해결하기 위하여 다양한 샘플링을 진행하였고, 분류정확도를 개선하였다. 청년층 차입자들의 채무 불이행 분류정확도를 나타낸 결과 ROC가 0.91로 나타나, 모형의 성능이 훌륭한 편이었다.

가계부채 부실 분류에 영향을 미치는 변인을 선별하기 위하여 중요도 지수를 활용하였다. 모형 정확도 개선지수(MDA)를 활용하여 예측력이 높은 상위 변수들을 도출한 결과 부채관련 변수 중 자산부채비율(DTA), 가계부실위험지수(HDRI), 가계의 원리금상환비율(DSR)이 순서대로 중요한 변수로 관찰되었다. 자산부채비율(DTA), 가계의 원리금상환비율(DSR)은 배영목(2011), 신진욱 외(2019), 이종희(2019)의 연구결과와 마찬가지로 가계의 과도한 부채부담은 채무 불이행과 연관이 된다고 볼 수 있었다. 또한 가계부실위험지수(HDRI)은 가계부채의 부실을 예측할 수 있는 새로운 지표로서의 역할을 실증적으로 보여주었다. 일반적 예상과 달리 자산부채비율(DTA)의 경우 이 비율이 증가할수록 채무 불이행 가능성이 증가하지 않았다. 자산 대비 과도한 부채가 채무 불이행으로 이어진다는 선행연구와는 일치하지 않았다. 즉, 가계부채 채무 불이행을 설명하기 위해서는 가계가 축적한 자산보다는 부채 상황을 감당할 수 있는 당기 소득의 역할이 더 중요하다고 해석할 수 있다.

채무관련 변수 중 의료비 비중, 통신비 비중, 주거비 비중, 교통비 비중, 식료품 비중, 교육비 비중 순으로 주요한 변수임이 확인되었다. 선행 연구에서 유의미한 변수이었던 가계 소득은 DSR, HDRI와 같은 합성변수를 만드는데 사용하여, 그 독립적인 역할을 파악하기는 어려웠다. 선행 연구(e.g. 김영일·진

경희, 2018; 박연우·허석균, 2018; 박윤태·노정현, 2017; 배영목, 2011; 신진욱 외, 2019; 이동진·한진현, 2017; 이종희, 2019)의 연구결과와 어느 정도 일치한다고 볼 수 있으나, 단순 가계부채 총량 자체보다는 가계부채 총량 대비 가계 소득이 증가할수록 가계의 채무 불이행 가능성은 감소할 것이라고 예상할 수 있었다. 그 외 차입자 가구의 성별, 혼인상태, 입주형태와 같은 인구 사회학적 변수들과 부채 목적에 따른 비중은 주요한 설명 변수가 아닌 것으로 나타났다.

청년층 차입자 중 가계부채 부실 위험을 안고 있는 고위험 계층이 존재하지만 이들이 보유하고 있는 가계 부채량이 전체 금융시장에서 차지하는 비중이 크지 않으므로 금융시장 전체를 위협하는 수준에는 이르지 못할 수 있다. 그럼에도 불구하고 식별된 고위험군 청년층 차입자들이 40세 이상 차입자들도 소득 및 자산의 규모가 작으며, 경기침체와 고용위축이 예상되는 상황이기에 사회안전망 차원에서 이들의 채무 불이행을 방지하기 위한 정책적 노력이 필요하다고 할 수 있다. 본 연구의 결과를 바탕으로 청년층 차입자를 위한 제언, 후속 연구자를 위한 제언 및 정책입안자를 위한 제언을 제시하면 다음과 같다.

첫째, 청년층 차입자들을 위한 제언은 다음과 같다. 청년층 차입자의 채무 불이행 가능성은 차입자들의 부채 관련 비율과도 연관이 되어 있었다. 정부는 부채지표 부실위험지수(HDRI)가 100을 초과하는 가구를 취약한 “부실위험”가구로 판단하였으나, 실제로 HDRI가 75% 수준이 되면 가계부채 부실 가능성이 급상승하는 것으로 나타났다. 따라서 가계의 채무 불이행 가능성을 판단하기 위해서는 자산부채비율과 원리금 상환비율을 모두 고려하여야 할 것이다. 또한 2022년 초부터 시행된 차주별 DSR 2단계 규제에 따라 총대출액이 2억원이 넘는 대출자는 연간 원리금 상환액이 연소득의 40%(은행권 기준)를 넘으면 추가 대출을 받을 수 없게 된다. 2022년 7월부터 개인별 총부채원리금상환비율(DSR) 규제 대상도 7월부터 1억원 초과 대출자로 확대된다(금융위원회, 2021). 그러나 연구결과 DSR의 경우 약 30%를 기준으로 가계의 채무 불이행 가능성이 높아졌으므로, HDRI 및 DSR 비율이 높은 청년층 차입자들은 자신의 상황을 보수적으로 평가하여 차입 및 상환 계획을 수립하여야 할 것이다.

둘째, 후속연구자를 위한 제언은 다음과 같다. 본 연구에서 관심의 대상이 된 채무 불이행 가구는 청년층 가계 중 약 7.9%이었다. 이처럼 데이터의 불균형을 처리하지 않을 경우 정확도는 높아질 수 있지만, 정밀도 및 재현율이 낮아지는 문제가 발생할 수 있다. 따라서 데이터 불균형 해소를 위한 노력이 필요하다고 보고, 본 연구에서는 이러한 문제를 해결하기 위하여 Oversampling, Downsampling, SMOTE, ROSE 샘플링과 같은 다양한 샘플링을 시도하였다. 후속 연구자들은 불균형 데이터에서의 과적합 문제를 해결하고, 민감도를 보다 향상시킬 수 있는 다양한 샘플링 기법들의 적용하는 것이 필요하겠다.

셋째, 정책입안자를 위한 제언은 다음과 같다. 본 연구결과에서 발견된 채무 불이행 우려가 큰 청년층 가계를 중심으로 지속적인 모니터링을 하며, 이들을 대상으로 한 위기관리에 활용할 수 있을 것이다. 현재 청년층 중에서 부실 위험 우려가 큰 취약 청년 가계들은 금리상승과 같은 외부충격이 발생 시 더욱 취약해질 것으로 예상된다. 가령 일반적으로 청년층의 소비지출 비중은 식료품, 주거비, 기타 소비지출이 높은 편이지만 의료비의 지출이 많은 일부 청년층의 경우 가계부채 부담의 증가로 귀결될 가능성이 상당히 크고, 가계부채 부실에 심각한 영향을 줄 것으로 예상된다. 저신용의 청년층 채무자의 경우 의료비 충당을 위하여 제2금융권의 대출을 받을 가능성이 있으므로, 정부는 청년들을 위한 공적채무조정 상담 지원 등을 제공하여야 할 것이다. 가령 채무조정 신청 전 1개월 이내에 3개월 이상 입원 치료가 필요한 질병을 진단받은 채무자의 경우 신속채무조정 지원이 가능한데 이러한 연체 전 채무조정에 대한 확대 뿐 아니라 교육 및 안내 등이 필요할 것이다.

본 연구는 연구의 내용과 방법에 있어서 다음의 한계점이 있음을 밝힌다. 본 연구는 청년층의 채무 불이행 위험을 측정하기 위하여 부채상환을 연체한 경험의 여부를 묻는 문항을 사용하였다. 그러나 연체의 기간 측면에서 30일 이내의 실수나 단순 착오로 인한 상환 연체와 장기간 연체는 가계경제에 미치는 영향이 다를 것이며, 각각의 연체의 채무조정 방법도 다르다. 따라서 상환 여력이 부족하여 발생하는 장기간 연체인지, 실수나 단순 착오로 인한 단기간의 연체인지 구분하는 것이 바람직할 것이다. 그러나 가계복지금융조사의 경우 1,590 청년 가계 중 약 92%에 해당하는 1,464가계가 이 문항에 응답하지 않아서 30일 이상 연체인지 여부를 확인할 수 없었기에 장기 및 단기 연체를 구분할 수 없었다는 한계를 지닌다. 또한 저금리 환경에서 돈을 차입하여 주식 혹은 가상자산 등에 투자한 청년들의 경우 자산가격 조정에 따라 경제적으로, 심리적으로 어려움을 경험하고 (금융위원회, 2022), 투자 손실을 입은 청년층 차입자들은 채무 불이행을 할 가능성도 존재한다. 그러나 연구 자료의 한계로 인하여 청년층 차입자가 주식 투자를 위한 목적이나 가상자산 투자를 위한 목적으로 차입을 하였는지 여부를 판한할 수 없었다.

## 참고문헌

- 1) 강종구(2017). 가계부채가 소비와 경제성장에 미치는 영향: 유량효과와 저장효과 분석. 한국은행 경제연구원, 23(2), 28-57.
- 2) 금융위원회(2017.10.24). 가계부채 종합대책. [https://www.moef.go.kr/nw/nes/detailNesDtaView.do?searchBbsId1=MOSFBBS\\_00000000028&searchNttId1=MOSF\\_00000000](https://www.moef.go.kr/nw/nes/detailNesDtaView.do?searchBbsId1=MOSFBBS_00000000028&searchNttId1=MOSF_00000000)

- 0011337&menuNo=4010100.
- 3) 금융위원회(2021.10.26). [보도자료] 「가계부채 관리 강화방안」 발표. <https://www.fsc.go.kr/no010101/76740>.
  - 4) 금융위원회(2022.7.14). 금융부문 민생안정 과제 추진현황 및 계획 금융위원장 브리핑. <https://www.fsc.go.kr/no010103/78089>.
  - 5) 김영일·진경희(2018). 가계 채무불이행 위험의 결정요인에 대한 분석과 시사점. 금융연구, 32(2), 63-92.
  - 6) 김태진·홍정식·전윤수·박종률·안태욱(2018). 랜덤포레스트를 이용한 모기업의 하향 거래처 기업의 분류: 자동차 부품산업의 가치사슬을 중심으로. 한국전자거래학회지, 23(1), 1-22.
  - 7) 김태은·양세정(2018). 청년층의 부채상환에 대한 영향요인 연구. 소비자학연구, 29(6), 105-133.
  - 8) 김혜련·최현자(1999). 유형별 가계소득과 자산이 소비지출에 미치는 영향: 주거보유형태별 분석. 대한가정학회지, 37(10), 91-106.
  - 9) 김현정·김우영(2009). 가계부채가 소비에 미치는 영향: 미시자료를 중심으로. 경제분석, 15(3), 1-36. UCI : G704-000447.2009.15.3.001.
  - 10) 남상호(2015). 우리나라 가계 소득 및 자산 분포의 특징. 보건 복지 Issues & Focus, 277, 1-8. DOI: 10.23064/2015.04.277.
  - 11) 노혜민·손상희(2017). 청년층 소비자의 신용등급관리역량에 한 연구. 소비자정책교육연구, 13(2), 1-27. DOI : 10.15790/cope.2017.13.2.001.
  - 12) 동아일보(2021.9.23). 빚으로 사는 20대, 가계대출 증가폭 전세대의 2배. <https://www.donga.com/news/article/all/20210923/109354480/1>.
  - 13) 매일경제(2022.4.30). “20대 ‘카드빚’ 무섭게 늘어나고 있다”...코로나19 때 뭐했나 보니. <https://www.mk.co.kr/news/economy/view/2022/04/384107/>.
  - 14) 박윤태·노정현(2017). 가구 연령별 가계부채 상환위험요인에 관한 연구. 대한부동산학회지, 35(2), 223-242.
  - 15) 박윤태(2018). 자산분위별 가계부채 상환불능위험도 추정 연구. 대한부동산학회지, 36(1), 193-210.
  - 16) 박연우·허석균(2018). 가계부채 결정요인과 부채부담 취약계층의 재무공경 및 신용위험 분석. 금융정보연구, 7(1), 1-31. DOI : 10.35214/rfis.7.1.201802.001.
  - 17) 박정민·송태민(2021). 소셜 빅데이터와 머신러닝을 활용한 가계부채 부실위험의 예측. 사회보장연구, 37(3), 71-90.
  - 18) 배영목(2011). 가계부채 상환부담의 분포와 추이. 경제발전연구, 17(1), 87-120. UCI : G704-000850.2011.17.1.001
  - 19) 백종호(2016). 국내 청년층 금융 현황 및 발전 방향. 하나금융경영연구소 연구보고서.
  - 20) 신진옥·박정민·김영(2019). 불운한 생애사에 따른 과중 채무 경험의 확률 추정 및 원인 분석. 한국사회복지학, 71(2), 279-305. DOI : 10.20970/kasw.2019.71.2.011.
  - 21) 서울시복지재단(2022.5.4). 서울시복지재단, ‘개인회생’ 신청 청년 실태조사 발표. <https://www.welfare.seoul.kr/web/contents/archive1-6.do?schM=view&id=20321&schBcid=press>.
  - 22) 서종덕(2016). 데이터 마이닝 기법을 이용한 환율예측: GARCH와 결합된 랜덤포레스트모형. 산업경제연구, 29(5), 1607-1628. UCI : G704-001438.2016.29.5.011
  - 23) 아시아경제(2018.6.22). [대한민국 부채보고서②] '빚' 썸(SOME)의 청춘. <https://www.asiae.co.kr/article/2018062212095556888>에서 2022년 1월 3일 인출.
  - 24) 오주한·정석원(2013). 임상의를 위한 다변량 분석의 실제. 대한건강관절학회지, 16(1), 63-72. UCI : G704-SER000010153.2013.16.1.005
  - 25) 이동진·한진현(2017). 종합적 상환여건을 반영한 과다부채 가계의 리스크 요인 분석. 한국응용경제학회, 19(3), 5-37.
  - 26) 이종희(2019). 한국 청년가계의 부실화 가능성 연구. 한국가족자원경영학회지, 23 (2), 99-115. DOI : 10.22626/jkfma.2019.23.2.006.
  - 27) 이종희(2020). 데이터마이닝 기법을 이용한 중년층의 가계부채 연체 가능성 분류 연구. 가정과학의질연구, 38(3), 1-16. DOI : 10.7466/JKHMA.2020.38.3.1.
  - 28) 유경원(2009). 가계부채에 관한 문체 분석: 미시자료를 중심으로. 경제분석, 15(4), 1-32. UCI : G704-000447.2009.15.4.004.
  - 29) 유진은(2015). 랜덤 포레스트. 교육평가연구, 28, 427-448. UCI : G704-000051.2015.28.2.00.
  - 30) 유재인·정호성(2021). 개인 소비 지출 및 소득 충격에 의한 채무 불이행 행태 분석. 금융정보연구, 10(2), 1-38. DOI : 10.35214/rfis.10.2.202106.001.
  - 31) 연합뉴스(2018.10.7). 1천 50조 가계부채 2금융권, 취약계층부터 부실화 조짐. <https://www.yna.co.kr/view/AKR20181005179900002>.
  - 32) 주동현(2019). 금리 상승이 수도권 및 비수도권 가계의 채무상환능력에 미치는 영향: 자산측면을 중심으로. 한국자료분석학회, 21(3), 1329-1340. DOI : 10.37727/jkdas.2019.21.3.1329.
  - 33) 전국경제인연합회(2021.11.14). 청년 체감경제고통지수 분석 보도자료. [https://www.fki.or.kr/FkiAct/Promotion/Report/View.aspx?content\\_id=5356f082-b2c9-4350-8ae2-230975dbcaa1&cPage=&search\\_type=0&search\\_keyword=](https://www.fki.or.kr/FkiAct/Promotion/Report/View.aspx?content_id=5356f082-b2c9-4350-8ae2-230975dbcaa1&cPage=&search_type=0&search_keyword=).
  - 34) 최필선·민인식(2018). 재정패널조사를 이용한 우리나라 복권지출의 역진성 분석. 재정학연구, 11(1), 49-76.

- 35) 최효미 · 유해미 · 김지현 · 김태우(2016). 청년층의 비혼에 대한 인식과 저출산 대응 방안. 연구보고 2016-19.
- 36) 파이낸셜뉴스(2018.1.24). 한국은행 가계부채 건진성. 청년층-노년층이 증장년층보다 취약. <https://www.fnnews.com/news/201801241130213610>.
- 37) 한국경제(2018.10.1). '1500조 가계부채' 금리인상 타격... 금융당국 일제히 현황 점검. <https://www.hankyung.com/economy/article/201810079257Y>.
- 38) 한국경제연구원(2021.11.15). 보도자료. '21년상 청년 채감경제고통지수 27.2', 15년 집계 이후 최고 수준. [http://www.keri.org/web/www/news\\_02?p\\_p\\_id=EXT\\_BBS&p\\_p\\_lifecycle=0&p\\_p\\_state=normal&p\\_p\\_mode=view&\\_EXT\\_BBS\\_struts\\_action=%2Fext%2Fbbs%2Fview\\_message&\\_EXT\\_BBS\\_messageId=356277](http://www.keri.org/web/www/news_02?p_p_id=EXT_BBS&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&_EXT_BBS_struts_action=%2Fext%2Fbbs%2Fview_message&_EXT_BBS_messageId=356277).
- 39) 한국은행a(2021.2.22). 2021년 4/4분기 가계신용. 보도자료. <https://eiec.kdi.re.kr/policy/materialView.do?num=223814&topic=>.
- 40) 한국은행b(2021.12.16). 2021년 가계금융복지조사 결과 보도자료. <https://www.bok.or.kr/portal/bbs/P0000559/view.do?nttId=10067944&menuNo=200690&pageIndex=1>.
- 41) 한국청소년정책연구원(2020). 청년 사회 · 경제 실태 및 정책방안 연구. 연구보고 16-R08.
- 42) 한요셉(2020). 청년 고용의 현황 및 정책제언. KDI 경제전망. 보고서.
- 43) 현대경제연구원(2018). 청년층 경제 활동 제약의 5대 특성과 시사점. 현대경제연구원. 18-07(통권 782호).
- 44) Legal Times(2021.10.16). 개인 · 법인파산 신청 전년 대비 10.5% 증가. <https://www.legaltimes.co.kr/news/articleView.html?idxno=63131>.
- 45) Ando, A. & Modigliani, F.(1963). The life-cycle hypothesis of saving: Aggregate implications and tests. *American Economic Review*, 53, 55-84.
- 46) Breiman, L.(2001). Random forests. *Machine learning*, 45(1), 5-32.
- 47) Dunkin, T. A.(2000). Credit cards: Use and consumer attitudes, 1970-2000. *Federal Reserve Bulletin*, 623-634.
- 48) Hamza, M. & Larocque, D.(2005). An empirical comparison of ensemble methods based on classification trees. *Journal of Statistical Computation and Simulation*, 75, 629-643.
- 49) Hastie, T., Tibshirani, R. & Friedman, J. H.(2009). *The elements of statistical learning: data mining, inference, and prediction*(2nd ed.). New York: Springer.
- 50) OECD(2021). Household debt. <https://data.oecd.org/hha/household-debt.htm>.
- 51) Sandri, M. & Zoccolotto, P.(2010). Analysis and correction of bias in total decrease in node impurity measures for tree-based algorithms. *Statistics and Computing*, 20(4), 393-407.
- 52) 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.
- 53) Varian, H. R.(2014). Big data: New tricks for econometrics. *Journal of Economic Perspective*, 28, 3-28.

■ 투 고 일 : 2022년 06월 20일  
 ■ 심 사 일 : 2022년 07월 13일  
 ■ 게 재 확 정 일 : 2022년 08월 10일