Journal of Health Informatics and Statistics

Original Article

J Health Info Stat 2026;51(2):102-117 • https://doi.org/10.21032/jhis.2026.51.2.102

pISSN 2465-8014 | eISSN 2465-8022

# 기계학습 기반 고립 · 은둔 청년의 사회적 고립 예측 및 취약요인 분석

**박현비[1], 최 별[2], 전보영[3]**

[1]명지전문대학 정보통신공학과 학부생, [2]세명대학교 간호학과 학부생, [3]명지전문대학 보건의료정보과 교수

# Prediction of Social Isolation and Vulnerability Factors among Isolated and Withdrawn Youth Using Machine Learning

**Hyunbi Park[1], Byeol Choi[2], Boyoung Jeon[3]**

*[1]Undergraduate Student, Department of Information and Telecommunications Engineering, Myongji College, Seoul; [2]Undergraduate Student, Department of Nursing, Semyung University, Jecheon; [3]Professor, Department of Health and Medical Information, Myongji College, Seoul, Korea*

**Objectives:** This study aimed to predict social isolation among youth using machine learning and to identify key vulnerability factors, thereby providing empirical evidence for early detection and preventive policy. **Methods:** Data were obtained from the 2022 Survey on Socially Isolated and Withdrawn Youth conducted by the Seoul Metropolitan Government. A total of 5,353 individuals aged 19–39 years were included in the analysis. Social isolation was defined as meeting both emotional and physical isolation criteria. Four machine learning classification models—k-nearest neighbors, Naive Bayes, decision tree, and random forest—were applied. Model performance was evaluated using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Feature importance was examined primarily using tree-based models. **Results:** Among the four models, the random forest model demonstrated the most stable and balanced predictive performance, showing high accuracy and AUC values. Key predictors of social isolation included frequency of going out, number of recent social interactions, household socioeconomic status, subjective health status, and employment-related factors. Social relationship and interaction variables consistently showed the strongest influence on isolation risk. **Conclusions:** Machine learning–based analysis is an effective approach for identifying youth at risk of social isolation. The findings suggest that behavioral and socioeconomic indicators can support early screening systems and contribute to data-driven policy strategies for preventing prolonged social isolation among youth.

**Key words:** Social isolation, Youth, Machine learning, Random forest, Prediction model

## 서 론

인간은 사회적 환경과의 상호작용 속에서 형성되는 존재로서, 타인과의 관계를 통해 정서적 안정과 일상생활을 유지한다[1]. 사회적 관계는 단순한 교류를 넘어 개인이 사회 구성원으로서의 소속감을 형성하고, 심리적·사회적 적응의 기반을 제공한다[2]. 그러나 이러한 관계망이 약화되거나 지속적으로 유지되지 못할 경우, 개인은 사회적 단절을 경험하게 되며 일상생활 전반에서 다양한 어려움에 마주치게 된다[3,4].

최근 사회적 단절의 양상은 특정 개인의 문제가 아니라 사회 구조적 변화와 맞물려[5] 청년층 전반에서 두드러지게 나타나고 있다. 특히 코로나19 팬데믹(pandemic)을 거치면서 국내 청년들의 사회적 고립 현상은 가파른 상승 곡선을 보인다. Jeong et al. [6]의 연구에 따르면, 2019년

**Corresponding author:** Boyoung Jeon
134 Gajwa-ro, Seodaemun-gu, Seoul 03656, Korea
Tel: +82-2-300-1168, E-mail: jeon.boyoung26@gmail.com

**Received:** January 14, 2026 **Accepted:** April 10, 2026 **Published:** May 31, 2026

*This work was carried out with the support of the Small Research, Good Seoul Program (2025-CR-01) funded by the Seoul Metropolitan Government through the Seoul Institute, and is based on a prior study. Parts of this study were also presented as a poster at the academic conference of the 2025 Korea Society of Artificial Intelligent in Medicine (KoSAIM).

No potential conflict of interest relevant to this article was reported.

How to cite this article:
Park H, Choi B, Jeon B. Prediction of social isolation and vulnerability factors among isolated and withdrawn youth using machine learning. J Health Info Stat 2026;51(2):102-117. Doi: https://doi.org/10.21032/jhis.2026.51.2.102

It is identical to the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2026 Journal of Health Informatics and Statistics

34만 명으로 집계되었던 고립 청년 사례는, 팬데믹 이후인 2021년 54만 명으로 급증하였고, 이들 중 은둔 상태에 놓인 청년은 24만 명으로 집계되었다. 또한 2023년 전국 실태조사에서도 응답자 21,360명 중 7,496명이 은둔을 경험한 것으로 나타났다. 조사에 참여하지 않은 잠재적 대상까지 고려할 때 실제 고립·은둔 청년은 훨씬 더 많을 것으로 예상된다[7,8].

이와 같은 고립의 심화는 학업, 취업, 주거, 경제적 불안정 등 복합적인 환경[9,10] 속에서 청년들은 사회적 관계를 지속적으로 형성·유지하는 데 어려움을 겪고 있으며, 이는 사회적 고립으로 이어질 가능성을 내포한다[11]. 이러한 고립 상태가 장기화한다면 사회참여의 축소와 생활 기능 저하로 이어질 수 있다는 점에서[12], 청년의 사회적 고립은 개인 차원의 문제가 아닌 사회적 대응이 요구되는 현상으로 인식되고 있다. 이러한 관점에서 사회적 고립은 단순한 사회적 관계의 부족을 넘어 심리적 불안과 우울을 유발하고, 자아개념 저하로 이어질 수 있는 심각한 사회문제로 지적되어왔다[13-15]. 최근 연구에서 고립·은둔 청년이 우울, 불안, 자살 충동 등 광범위한 정신건강 위험과 밀접하게 연관된 고위험군임을 보고하고 있다[16]. 특히 청소년기와 청년기의 사회적 고립은 은둔형 외톨이로 이행될 가능성이 높아, 사회복귀를 더욱 어렵게 만들 수 있다는 점에서 우려가 크다[17].

청년층의 사회적 고립은 특정 소수 집단의 문제가 아닌 전반적인 구조적 문제로 보고되고 있다. 실제 2019년 기준 고립 청년으로 인한 경제활동 포기 손실은 연간 약 6.7조 원에 달하는 것으로 분석되었다[18].

고립·은둔 청년은 일반 청년과 비교할 때 낮은 소득 수준, 불안정한 고용 상태, 교육 기회의 제약으로 인하여 사회경제적으로 불리하다. 또한 주관적으로 자신의 경제적 위치를 낮게 인식하는 청년은 전반적인 삶의 만족도가 낮은 경향을 보였다. 이러한 낮은 경제적 자기평가는 사회적 단절감을 심화시키고, 동시에 장기적인 은둔으로 이어질 가능성을 높이는 것으로 나타났다[19].

서울시 거주 청년 인구 중 약 2-4%가 고립 청년에 해당하는 것으로 보고되었고[20,21], 2023년 기준 19-29세 청년의 34.5%, 30-39세 청년의 27.5%가 사회적으로 고립된 상태에 놓여 있는 것으로 나타났다. 또한 물리적 고립 상태에 이르지 않아도, 상당수의 청년은 고립감과 외로움을 경험하고 있다[12]. 선행연구에 따르면 고립감을 느끼는 청년은 13.4%, 외로움을 경험하는 청년은 35.0%에 이르는 것으로 보고되었다[22]. 이는 사회적 고립이 은둔이라는 극단적 상태 이전부터, 일상적 관계 약화와 사회참여 감소의 형태로 이미 광범위하게 진행되고 있음을 보여준다.

사회적 고립은 명확히 합의된 개념 정의 없이 학문 분야나 연구 주제 및 대상에 따라 다차원적이고 다양한 방식으로 정의된다[15]. Park and Kim [14]은 사회적 고립을 사회적 관계의 양과 질에 대한 객관적 측면, 외로움은 사회적 관계에 대한 주관적 인식과 관련된 개념으로 구분하였다. Park [23]은 사회적 고립감을 사회적 관계망의 축소 및 상호작용 감소로 인해 발생하는 부정적 심리 상태로 설명하였다. Nicholson [24]은 사회적 고립을 타인과의 상호작용 빈도 감소뿐만 아니라 관계의 질적 결핍과 소속감 부재라는 측면에서 보다 포괄적으로 정의하였으며, Zavaleta et al. [25]은 사회적 관계의 양적·질적 결핍 모두를 포함하는 개념으로 사회적 고립을 설명하였다. Jung [26]은 사회적 고립을 특정 시기나 세대에 국한된 현상이 아니라, 생애과정 전반에 걸쳐 누적적으로 증가하는 경험으로 보았으며, 사회적 관계 및 참여 위축에 따른 위기까지를 포괄하는 개념으로 정의하였다.

생애과정 관점으로는, 은둔 상태에 있는 청년들은 성인기 이후 학업 중단, 반복적인 취업 실패, 직장 내 관계 갈등, 경제적 곤란 등 다양한 부정적 경험이 누적되면서 사회적 관계망이 점차 약화하고 심리적 안정감이 손상되는 양상을 보인다. 이러한 실패 경험이 단순한 사건을 넘어 사회적 역할 수행의 좌절로 인식될 경우, 점진적으로 사회적 관계망의 축소와 심리적 위축으로 이어질 수 있다[10,12].

이 과정에서 사회적 관계 자본이 극도로 결핍된 상태가 은둔으로 발현될 수 있다고 설명하였다. 은둔은 사회적 고립의 한 유형으로 6개월 이상 집에 머물며 사회적 상황과 관계를 회피하는 상태를 뜻한다. 이는 상당한 고통과 기능 장애를 수반하고[27], 더욱 폐쇄적이며 심화된 위기를 의미한다[21]. 반면 외로움은 단순한 사회적 접촉 부족이 아니라 주관적으로 경험하는 사회적 단절감[28]으로, 고립과 은둔 생활을 심화시키는 촉매 요인으로 작용할 수 있다[29]. 그러나 외로움은 주관적 인식의 차원에 해당하는 반면, 사회적 고립은 객관적 관계망의 단절을 중심으로 정의된다는 점에서 구분된다.

최근 연구에서는 청년의 사회적 고립을 단일 요인이 아닌 사회적 관계, 사회적 지지, 은둔 행동의 결합으로 나타나는 다차원적 현상으로 이해하고, 유형화 접근의 필요성을 제기하였다[30]. 이러한 연구는 고립의 구조적 특성을 설명하는 데 기여하였으나, 정책 현장에서 활용할 수 있는 조기 선별이나 예측 관점의 분석은 상대적으로 제한적인 측면이 있었다.

이에 서울특별시(이하 서울시) [21]는 정책적 개입을 위한 구체적인 기준을 제시하였다. 서울시는 사회적 고립 청년을 ① 비동거 가족 및 친척을 포함한 타인과의 교류가 부재하고, ② 생활·경제·정서적 곤란 상황에서 도움받을 수 있는 사람이 없는 청년으로 정의하였다[21]. 정서적 고립은 "중요하거나 어려운 일이 있을 때 조언을 구할 수 있는 사람", "급한 일이 있을 때 부탁할 수 있는 사람", "개인적으로 돈을 빌려야 할 때 부탁할 수 있는 사람", "낙심하거나 우울할 때 속마음을 털어놓을 수 있는 사람"이 전혀 없는 상태를 의미하고, 물리적 고립은 "친한 친구나 친한 사람과의 대면 교류", "이외 직장, 학교, 동네에서 알고 지내는 사람들과의 대면 교류(업무상 교류는 제외)"를 의미한다[21].

본 연구는 '서울시 고립·은둔 청년 실태조사'에서 사용된 조작된 정의를 적용하여, 정서적 고립과 물리적 고립이 동시에 충족되는 경우를 사회적 고립으로 정의하였다. 따라서 정서적 고립과 물리적 고립이 동시에 나타나면서, 단순한 외로움이나 심리적 고립감과는 구별된다는 차이점을 보여준다. 또한 서울시 실태조사에서 고립 범주에 은둔을 포함하는 기준을 분석에 동일하게 적용하였다[21].

선행연구에 따르면 사회적 고립과 외로움은 우울과 같은 정신건강 문제와 상호 연관된 다요소적 문제이다[14]. 또한 사회적 고립은 삶의 만족도와 삶의 질을 저하시킬 뿐 아니라 신체적·정신적 건강에도 부정적인 영향을 미치는 것으로 보고되었다[23,31]. 그러나 대다수의 기존 연구에서는 고립·은둔 발생 요인을 단편적으로 분석하여, 다양한 특성이 강한 고립 청년의 집단을 자세하게 설명하는 데 한계를 보였다. 또한 개인의 생활여건, 생활방식, 사회참여도 등 복합적 요인을 동시에 고려하여 고립의 위험 수준을 체계적으로 분류하거나 예측하는 연구는 상대적으로 부족하였다[30]. 이에 따라 고립 청년의 욕구를 충분히 반영하고, 재고립을 예방하기 위한 정책적 개입에는 제한적이었다.

국내에서는 사회적 고립 청년을 대상으로 한 법률 및 정책적 대응이 일부 지방자치단체 조례와 개별 사업을 중심으로 이루어지고 있으나, 중앙정부 차원의 명확한 개념 정의와 법적 근거는 여전히 미흡하다는 평가를 받고 있다.

특히 사회적 고립 청년을 독립적인 정책 대상으로 규정하는 법률이 부재하고, 기존 정책 또한 선언적 수준에 머물거나 단기적·사후적 개입에 그치고 있어 조기 발견과 예방 차원의 체계적 대응의 한계점이 존재한다[22]. 반면 영국, 미국, 일본 등은 사회적 고립을 국가가 대응해야 할 사회적 위험으로 인식하고, 국가 차원의 전담부처 지정, 장기적 정책 프레임 구축, 지역사회 기반의 통합적 지원체계를 운영하고 있다는 점에서 우리나라와 차이를 보인다. 영국은 외로움과 사회적 고립을 공중보건 이슈로 설정하고 국가 차원의 전략과 측정 체계를 마련하였으며, 미국은 지역별 특성을 반영하여 단절 청년을 지원하고 있다. 일본은 장기적 고립 상태에 놓인 청년을 대상으로 조기 발견과 지속적 지원을 연계하는 체계를 구축하고 있어, 국내의 민간 중심, 분절적 구조에 있는 정책과 대비된다[22].

이러한 연구의 한계를 극복하기 위해 최근 청년의 사회적 고립과 은둔을 정책적 개입이 필요한 사회적 위험 상태로 인식하려는 제도적인 변화가 확대되고 있다. 서울시는「서울시 고립·은둔 청년 실태조사」를 바탕으로 고립과 은둔의 개념을 구분하고, 이를 기반으로 심리·정서적 지원, 사회적 관계 회복, 사회복귀 및 취업 연계를 포함하는 종합적인 지원 정책을 추진하고 있다. 서울시의 고립·은둔 청년 정책은 개인 단위의 심리지원뿐만 아니라, 대면 활동과 체험 중심의 관계 회복 프로그램, 생활 습관 형성 및 공동생활 지원, 진로 탐색과 직무 역량 강화를 포함한 사회복귀 지원을 포함하고 있으며, 고립·은둔 청년의 가족을 정책 대상에 포함하여 부모 교육 및 가족 관계 회복 프로그램을 병행하고 있다[21].

이에 본 연구는 「서울시 고립·은둔 청년 실태조사」 자료를 활용하여, 서울시가 제시한 사회적 고립 정의를 기준으로 기계학습 기반으로 청년의 사회적 고립 여부를 예측하고, 주요 취약 요인을 분석하고자 한다.

머신러닝은 대규모 데이터를 기반으로 변수 간의 비선형적 관계와 잠재적 상호작용을 효과적으로 반영할 수 있어, 복합적인 건강 문제의 예측에 유리하다[32]. 특히 정서적 고립과 물리적 고립 조건을 동시에 고려하여 고립군을 정의하고, 인구사회학적 특성, 생활 특성, 사회적 관계, 심리적 요인 등 다양한 변수를 종합적으로 분석하여 사회적 고립의 특성을 더욱 정밀하게 규명하고자 한다. 아울러 본 연구는 분석을 바탕으로 청년의 사회적 고립 여부를 예측하는 머신러닝 기반 모형을 구축하는 것을 목표로 한다. 이를 통해 청년 사회적 고립 예방 및 맞춤형 설계를 위한 실증적 근거를 제공하고자 한다.

## 연구 방법

### 연구자료 및 연구대상

본 연구는 서울시가 2022년에 실시한 「서울시 고립·은둔 청년 실태조사」 원자료를 활용하였다. 조사는 만 19세에서 39세 청년을 대상으로, 사회적 고립 및 은둔 상태의 규모와 특성을 파악하고, 정서적·물리적 고립 지표, 사회적 관계망, 생활 습관, 생애 경험, 심리·정서 상태 등을 다차원적으로 조사한 대규모 실태조사이다. 자료 수집은 온라인 구조화 방식으로 이루어졌고, 총 5,513명이 응답하였다.

조사 문항은 사회적 관계 및 교류 특성, 외출 및 일상생활, 심리사회적 지표, 생애 사건 경험, 건강 및 생활 습관, 인구사회학적 특성 등으로 구성되어 있다. 본 연구에서는 분석을 위해 전체 자료 중 결측치가 확인된 응답을 제외하였다. 결측치가 발생한 주요 문항은 직업, 혼인상태, 가정의 사회경제적 수준, 본인의 주관적 경제 수준, 지난 2주간 교류 상대, 지난 2주간 교류 횟수, 평균 수면시간이었다. 해당 문항들에서 결측값이 포함된 응답 사례 160명을 분석에서 제외하였고, 최종적으로 5,353명을 본 연구의 분석 대상자로 선정하였다.

### 변수정의

#### 종속변수

본 연구에서 종속변수는 사회적 고립 여부이다. 사회적 고립은 정서적 고립과 물리적 고립 기준을 모두 충족할 때 고립군으로 정의하였다. 정서적 고립은 "도움이 필요할 때 의지할 사람이 전혀 없는 상태", 물

리적 고립은 가족·친척 외 타인과의 대면 교류가 연 1-2회 이하로 거의 없으며, 이러한 상태가 6개월 이상 지속된 경우로 정의하였다. 두 기준 중 하나라도 충족하지 않는 경우 비고립군으로 분류하였다.

#### 독립변수

독립변수는 청년의 사회적 고립에 영향을 미칠 수 있는 요인을 인구사회학적 요인, 경제활동 및 사회경제적 요인, 사회적 관계 및 교류 요인, 심리적 요인, 생활 습관 및 건강 요인, 생애사적 경험 요인으로 구분하였다(Appendix Table 1). 독립변수의 이분화는 「서울시 고립·은둔 청년 실태조사」의 문항 응답 구조와 정책적 분류 기준을 반영하였다[21]. 원자료에서 범주형으로 수집된 변수는 해당 범주를 유지하였고, 필요 시 분포 및 해석의 일관성을 고려하여 범주를 통합하였다. 우울과 외로움의 경우, 점수의 연속적 특성을 유지하기 위해 연속형으로 분석하였다.

#### 인구사회학적 요인

인구사회학적 요인은 개인의 기본적 특성을 반영하는 변수로, 성별은 남자, 여자로 구분하였고 연령대는 만 19-24세, 만 25-29세, 만 30-34세, 만 35-39세의 네 구간으로 구분하였다. 동거인 수는 함께 거주하는 인원을 기준으로 혼자 거주(1명), 2인 이상 거주로 최종 학력은 고등학교 이하, 대학교 이상으로 구분하였다. 직업 상태는 정규직, 비정규직, 학생·무직 및 기타로 구분하였고, 혼인상태는 기혼, 미혼, 이혼·사별로 구분하였다. 주택 소유 형태는 자가 비보유자, 자가 보유자로 구분하였다.

#### 경제활동 및 사회경제적 요인

경제활동 및 사회경제적 요인은 개인의 노동 여부와 경제적 환경을 반영하는 변수로, 지난주 노동 여부는 최근 1주일의 노동 경험을 기준으로 일을 하지 않음, 일을 함으로 구분하였다. 가정의 사회경제적 수준은 중하 이하와 중상 또는 상으로 구분하였으며, 본인의 주관적 경제 수준은 부족함과 여유있음으로 구분하였다.

#### 사회적 관계 및 교류 요인

사회적 관계 요인은 청년의 대면 관계 및 비대면 교류 수준을 반영하는 변수로, 외출 빈도는 활동적 외출, 보통 외출, 비활동적 외출로 구분하였다. 온라인 대화 여부는 그렇다, 그렇지 않다로 구분하였다.

#### 심리적 요인

심리적 요인은 청년의 정서적·정신적 상태를 반영하는 변수로, 외로움과 우울 수준을 포함하였다. 외로움은 4개 문항의 평균값을 산출하여 연속형 변수로 구성하였다. 우울은 9개 문항의 평균값을 산출하여 연속형 변수로 구성하였다.

#### 생활습관 및 건강 요인

생활습관 및 건강 요인은 개인의 일상 습관과 건강 상태를 반영하는 변수로, 밤낮 역전 생활 여부, 수면 규칙성, 수면의 질은 각각 그렇지 않다, 그렇다로 구분하였다. 평균 수면 시간은 6시간 미만, 6시간 이상으로 구분하였다. 식사 횟수는 정상·규칙적 식사, 결식·저빈도 식사로 구분하였고, 식사 방식은 혼자 식사, 가족과 식사로 구분하였다. 자기관리 항목(옷 갈아입기, 샤워·목욕, 세수·양치, 방 정리)은 각각 자주 하지 않음과 매일로 구분하였다. 주관적 건강상태는 좋지 않음과 보통 이상으로 구분하였으며, 약물 복용 여부는 예와 아니오로 구분하였다.

#### 생애사적 경험 요인

생애사적 경험 요인은 성장기 및 성인기 동안 겪은 주요 경험을 반영하는 변수로, 성인기 전 경험은 가정의 경제적 어려움, 보호자로부터의 학대 또는 모욕 경험, 가족의 정서적 어려움, 가까운 사람의 상실, 전학·이사 경험, 따돌림 경험으로 구성되며, 각 문항은 없다, 있다로 구분하였다. 성인기 후 경험은 경제적 이유로 인한 학업 중단, 진로 및 취업에 대한 압박 경험, 원하는 시기의 대학 진학 실패, 희망 대학 진학 실패, 원하는 시기의 취업 실패, 희망 직장 진입 실패, 신뢰했던 사람으로부터의 배신 또는 사기 경험, 가까운 관계의 단절 또는 상실 경험으로 구성되며, 각 문항은 없다, 있다로 구분하였다.

### 데이터 전처리

본 연구는 분석의 정확성과 일관성을 위해 단계적인 자료 전처리를 진행하였다. 독립변수의 결측치는 SimpleImputer (strategy = 'most_frequent')를 활용하여 보완하였고, 주요 변수에서 결측이 확인된 160명(전체 5,513명 중 2.9%)을 제외하고 분석을 진행하였다. 결측 비율이 5% 미만인 경우 단순 사례 제거를 적용하더라도 추정의 왜곡이 크지 않을 수 있다는 선행연구를 참고하여 종속변수에서 결측이 확인된 사례는 분석에서 제외하였다[33,34]. 이후 전체 자료는 기계학습 모형 간의 성능 비교를 위해 학습용(70%), 검증용(30%)으로 분할하였다. 분할 시 종속변수의 집단 비율을 유지하기 위해 stratified split을 적용하였고, 모든 모형 간 동일한 조건을 부여하기 위해 random state = 42를 사용하였다.

각 모형의 예측 성능을 안정적으로 평가하기 위해 학습 데이터 내부에서 5-fold 교차검증을 적용한 GridSearchCV (평가지표: ROC-AUC)를 활용하여 최적의 하이퍼파라미터를 탐색하였다. 또한 ROC 곡선을 기반으로 산출한 최적 임곗값을 분류 기준에 적용하였다. 이와 같은

전처리 절차를 통해 분석에 적합한 자료를 이용하여 네 가지 기계학습 알고리즘(KNN, Naive Bayes, Decision Tree, Random Forest)의 예측 성능을 비교 및 평가하였다.

### 분석모형

본 연구는 서울시 고립·은둔 청년의 사회적 고립 여부를 예측하고 각 기계학습 모형의 성능을 비교하기 위해 K-최근접 이웃(K-Nearest Neighbors, KNN), 나이브 베이즈(Naive Bayes), 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest) 등 네 가지 분류 알고리즘을 적용하였다. 모든 분석은 Python 3.12.12를 활용하였다.

KNN 모델은 관측치 간 거리를 활용하여 가장 가까운 K개 이웃을 기준으로 분류하는 알고리즘이다[35]. 본 연구에서는 n_neighbors, metric (euclidean, manhattan), weights (uniform, distance)를 탐색 범위로 설정하고, GridSearchCV (CV=5, ROC-AUC 기준)를 통해 최적 조합을 선정하였다.

나이브 베이즈 모델은 베이즈 이론에 기반하여 각 특성이 조건부 독립이라는 가정을 전제로 관측된 특징을 확률적으로 계산해 분류하는 분류 알고리즘이다[36]. GaussianNB와 BernoulliNB를 비교 분석한 후, 자료 특성에 적합한 나이브 베이즈 모형을 최종 적용하였고, var_smoothing을 탐색 범위로 설정하여 교차검증을 통해 최적 모형을 선정하였다.

의사결정나무 모델은 분류 규칙에 따라 데이터를 반복적으로 분할하여 반응 변수를 예측하는 알고리즘이다[37]. max_depth, min_samples_split, min_samples_leaf, class_weight를 탐색 범위로 설정하고, 교차검증을 통해 최적 파라미터를 도출하였다.

랜덤포레스트 모델은 의사결정나무 모형을 기반으로 하며 무작위성을 극대화하여 예측 오차를 줄이는 앙상블 알고리즘이다[38]. n_estimators, max_depth, min_samples_leaf를 탐색 범위로 설정하고, 교차검증 기반으로 최적 조합을 선정하였다. 랜덤포레스트와 의사결정나무는 변수 중요도를 제공하여, 사회적 고립에 영향을 미치는 주요 요인을 탐색하는 데 이용하였다. 모든 분류 모형의 기본 임계성능 평가는 Accuracy, Precision, Recall, F1-score, ROC-AUC를 기준으로 평가되었으며, 혼동행렬을 통해 고립군과 비고립군의 분류 정확도를 세부적으로 비교하였다. 분석을 통해 각 모형의 성능을 비교하고 사회적 고립 예측에 가장 적합한 모델을 도출하고자 하였다.

## 연구 결과

### 사회적 고립 여부에 따른 비교

총 5,513명의 응답자 중 결측값 160명을 제외한 최종 분석 대상자는 5,353명이었으며, 사회적 고립 여부에 따라 사회적 고립군 377명(7.04%)과 비고립군 4,976명(92.97%)으로 분류하였고, 사회적 고립 여부에 따른 각 요인의 특성과 주요 변수의 차이는 Table 1에 제시하였다.

#### 인구사회학적 요인

성별에 따른 분포를 살펴보면, 사회적 고립군 내에서는 남성 50.7%, 여성 49.3%, 사회적 비고립군에서는 남성 48.7%, 여성 51.3%로 나타나 사회적 고립 여부에 따른 성별 차이는 통계적으로 유의미한 차이는 없었다($p=0.503$). 반면 연령대에서는 사회적 고립군 내에서 만 35-39세가 35.8%로 가장 높았고, 사회적 비고립군에서는 만 25-29세가 31.0%로 가장 높게 나타나, 사회적 고립 여부에 따라 연령 분포에 통계적으로 유의미한 차이가 있었다($p=0.024$). 교육 수준에 따라서는 사회적 고립군 내에서 대학교 이상 96.8%, 고등학교 이하 3.2%, 사회적 비고립군에서는 각각 99.4%, 0.6%로 나타나 유의미한 차이가 있었다($p<0.001$). 고용형태에서는 사회적 고립군 내에서 학생·무직 및 기타가 41.4%로 가장 높았고, 정규직 34.0%, 비정규직 24.6% 순으로 나타났다. 반면 사회적 비고립군에서는 정규직이 62.3%로 가장 높았으며, 비정규직 20.2%, 학생·무직 및 기타 17.5%로 나타나 사회적 고립 여부에 따라 유의미한 차이가 있었다($p<0.001$). 혼인상태는 사회적 고립군 내에서 미혼 83.8%, 기혼 15.4%, 이혼·사별 0.8%, 사회적 비고립군에서는 각각 74.9%, 23.7%, 1.4%로 나타나, 두 집단 모두 미혼이 가장 높았으며 분포 차이는 통계적으로 유의미한 차이가 있었다($p<0.001$). 주택 소유 형태에서는 사회적 고립군 내에서 자가 비보유자 59.7%, 보유자 40.3%, 사회적 비고립군에서는 각각 51.9%, 48.1%로 유의미한 차이가 있었으나($p=0.004$), 동거인 수에 따른 차이는 통계적으로 유의미한 차이는 없었다($p=0.415$).

#### 경제활동 및 사회경제적 요인

지난주 노동 여부에 따라 사회적 고립군 내에서는 일을 한 경우 57.8%, 하지 않은 경우 42.2%, 사회적 비고립군에서는 각각 84.7%, 15.3%로 나타나 통계적으로 유의미한 차이가 있었다($p<0.001$). 가정의 주관적 사회경제적 수준에서는 사회적 고립군 내에서 '중하 이하' 63.1%, '중상 또는 상' 36.9%, 사회적 비고립군에서는 각각 28.7%, 71.3%로 나타나 유의미한 차이가 있었다($p<0.001$). 개인의 주관적 경제 상태에서도 사회적 고립군 내에서 '경제적으로 부족하다'고 인식한 경우 79.6%, 사회적 비고립군에서는 48.0%로 나타나 통계적으로 유의미한 차이가 있었다($p<0.001$).

#### 사회적 관계 및 교류 요인

외출 빈도에 따라 사회적 고립군 내에서는 활동적 외출 56.0%, 보통 31.6%, 비활동적 외출 12.5%, 사회적 비고립군에서는 각각 88.8%, 9.2%,

**Table 1.** Differences in general characteristics by social isolation status

| Variables | Categories | Non-isolated (n=4,976) | | Isolated (n=377) | | Chisq-test or t-test | *p* |
|---|---|---|---|---|---|---|---|
| | | n | (%) | n | % | | |
| Sociodemographic factors | | | | | | | |
| Sex | Male | 2,425 | 48.7 | 191 | 50.7 | 0.45 | 0.503 |
| | Female | 2,551 | 51.3 | 186 | 49.3 | | |
| Age group | 19-24 | 607 | 12.2 | 45 | 11.9 | 9.45 | 0.024 |
| | 25-29 | 1,543 | 31.0 | 94 | 24.9 | | |
| | 30-34 | 1,374 | 27.6 | 103 | 27.3 | | |
| | 35-39 | 1,452 | 29.2 | 135 | 35.8 | | |
| Education level | High school or lower | 32 | 0.6 | 12 | 3.2 | 24.71 | <0.001 |
| | College or higher | 4,944 | 99.4 | 365 | 96.8 | | |
| Employment status | Regular employment | 3,099 | 62.3 | 128 | 34.0 | 153.67 | <0.001 |
| | Non-regular employment | 1,004 | 20.2 | 93 | 24.7 | | |
| | Student, unemployed and others | 873 | 17.5 | 156 | 41.4 | | |
| Marital status | Married | 1,179 | 23.7 | 58 | 15.4 | 15.05 | <0.001 |
| | Un-married | 3,728 | 74.9 | 316 | 83.8 | | |
| | Divorced or widowed | 69 | 1.4 | 3 | 0.8 | | |
| House ownership | Non-homeowner | 2,583 | 51.9 | 225 | 59.7 | 8.18 | 0.004 |
| | Homeowner | 2,393 | 48.1 | 152 | 40.3 | | |
| Number of co-residents | 1 | 1,027 | 20.6 | 85 | 22.5 | 0.66 | 0.415 |
| | ≥2 | 3,949 | 79.4 | 292 | 77.5 | | |
| Economic activity and socioeconomic factors | | | | | | | |
| Worked during the past week | Did not work | 761 | 15.3 | 159 | 42.2 | 176.04 | <0.001 |
| | Worked | 4,215 | 84.7 | 218 | 57.8 | | |
| Household socioeconomic status | Lower-middle or below | 1,427 | 28.7 | 238 | 63.1 | 192.51 | <0.001 |
| | Upper-middle or above | 3,549 | 71.3 | 139 | 36.9 | | |
| Individual subjective economic status | Insufficient | 2,386 | 48.0 | 300 | 79.6 | 138.94 | <0.001 |
| | Sufficient | 2,590 | 52.0 | 77 | 20.4 | | |
| Social relationships and social interaction factors | | | | | | | |
| Frequency of going out | Active | 4,419 | 88.8 | 211 | 56.0 | 348.25 | <0.001 |
| | Moderate | 459 | 9.2 | 119 | 31.6 | | |
| | Inactive | 98 | 2.0 | 47 | 12.5 | | |
| Online communication | Yes | 2,846 | 57.2 | 142 | 37.7 | 53.41 | <0.001 |
| | No | 2,130 | 42.8 | 235 | 62.3 | | |
| Psychological factors | Loneliness (1-4, Mean±SD) | 2.20±0.75 | | 2.80±0.91 | | <0.001 | |
| | Depression (1-4, Mean±SD) | 1.60±0.68 | | 2.20±0.88 | | <0.001 | |
| Lifestyle and health factors | | | | | | | |
| Daily life patterns | | | | | | | |
| Reversed day–night routine | No | 2,344 | 47.1 | 162 | 43.0 | 2.24 | 0.134 |
| | Yes | 2,632 | 52.9 | 215 | 57.0 | | |
| Regular sleep schedule | No | 568 | 11.4 | 82 | 21.8 | 34.13 | <0.001 |
| | Yes | 4,408 | 88.6 | 295 | 78.2 | | |
| Sleep quality | No | 1,092 | 21.9 | 145 | 38.5 | 52.88 | <0.001 |
| | Yes | 3,884 | 78.1 | 232 | 61.5 | | |
| Self-care and cleanliness | | | | | | | |
| Changing clothes | Not frequent | 1,458 | 29.3 | 193 | 51.2 | 77.73 | <0.001 |
| | Daily | 3,518 | 70.7 | 184 | 48.8 | | |
| Bathing/Showering | Not frequent | 1,535 | 30.8 | 178 | 47.2 | 42.39 | <0.001 |
| | Daily | 3,441 | 69.2 | 199 | 52.8 | | |

(*Continued to the next page*)

**Table 1.** Continued

| Variables | Categories | Non-isolated (n=4,976) | | Isolated (n=377) | | Chisq-test or t-test | *p* |
|---|---|---|---|---|---|---|---|
| | | n | (%) | n | % | | |
| Washing face/Brushing teeth | Not frequent | 846 | 17.0 | 57 | 15.1 | 0.76 | 0.385 |
| | Daily | 4,130 | 83.0 | 320 | 84.9 | | |
| Room organization | Not frequent | 3,212 | 64.5 | 268 | 71.1 | 6.3 | 0.012 |
| | Daily | 1,764 | 35.5 | 109 | 28.9 | | |
| Self-rated health status | Poor | 569 | 11.4 | 142 | 37.7 | 207.08 | <0.001 |
| | Fair or better | 4,407 | 88.6 | 235 | 62.3 | | |
| Use of medication | Yes | 388 | 7.8 | 40 | 10.6 | 3.4 | 0.065 |
| | No | 4,588 | 92.2 | 337 | 89.4 | | |
| Average sleep duration | <6 hours | 2,168 | 43.6 | 182 | 48.3 | 2.96 | 0.085 |
| | ≥6 hours | 2,808 | 56.4 | 195 | 51.7 | | |
| Average meal frequency | Normal, regular meals | 4,682 | 94.1 | 313 | 83.0 | 67.03 | <0.001 |
| | Skipped/low-frequency meals | 294 | 5.9 | 64 | 17.0 | | |
| Usual eating pattern | Eat alone | 2,597 | 52.2 | 241 | 63.9 | 18.91 | <0.001 |
| | Eat with family | 2,379 | 47.8 | 136 | 36.1 | | |
| Life-course experience factors | | | | | | | |
| Pre-adulthood experiences | | | | | | | |
| Financial hardship before adulthood | No | 3,114 | 62.6 | 173 | 45.9 | 40.5 | <0.001 |
| | Yes | 1,862 | 37.4 | 204 | 54.1 | | |
| Abuse or humiliation by caregivers | No | 3,374 | 67.8 | 205 | 54.4 | 27.92 | <0.001 |
| | Yes | 1,602 | 32.2 | 172 | 45.6 | | |
| Family member's emotional distress | No | 2,952 | 59.3 | 166 | 44.0 | 33.08 | <0.001 |
| | Yes | 2,024 | 40.7 | 211 | 56.0 | | |
| Loss of a close person | No | 3,231 | 64.9 | 223 | 59.2 | 4.87 | 0.027 |
| | Yes | 1,745 | 35.1 | 154 | 40.8 | | |
| Frequent relocation or school transfer | No | 3,671 | 73.8 | 270 | 71.6 | 0.73 | 0.392 |
| | Yes | 1,305 | 26.2 | 107 | 28.4 | | |
| Bullying or harassment | No | 3,382 | 68.0 | 201 | 53.3 | 33.33 | <0.001 |
| | Yes | 1,594 | 32.0 | 176 | 46.7 | | |
| Adulthood experiences | | | | | | | |
| Dropping out due to financial reasons | No | 4,483 | 90.1 | 324 | 85.9 | 6.15 | 0.013 |
| | Yes | 493 | 9.9 | 53 | 14.1 | | |
| Career or employment pressure | No | 3,677 | 73.9 | 231 | 61.3 | 27.69 | <0.001 |
| | Yes | 1,299 | 26.1 | 146 | 38.7 | | |
| Failure to enter college at desired time | No | 3,522 | 70.8 | 246 | 65.3 | 4.88 | 0.027 |
| | Yes | 1,454 | 29.2 | 131 | 34.7 | | |
| Failure to enter desired college | No | 2,781 | 55.9 | 194 | 51.5 | 2.61 | 0.106 |
| | Yes | 2,195 | 44.1 | 183 | 48.5 | | |
| Failure to obtain employment at desired time | No | 2,703 | 54.3 | 144 | 38.2 | 35.95 | <0.001 |
| | Yes | 2,273 | 45.7 | 233 | 61.8 | | |
| Failure to enter desired workplace | No | 2,516 | 50.6 | 153 | 40.6 | 13.56 | <0.001 |
| | Yes | 2,460 | 49.4 | 224 | 59.4 | | |
| Betrayal or fraud by a trusted person | No | 3,395 | 68.2 | 208 | 55.2 | 26.55 | <0.001 |
| | Yes | 1,581 | 31.8 | 169 | 44.8 | | |
| Separation or loss of close relationships | No | 2,907 | 58.4 | 182 | 48.3 | 14.36 | <0.001 |
| | Yes | 2,069 | 41.6 | 195 | 51.7 | | |

SD, standard deviation.

2.0%로 나타나 유의미한 차이가 있었다($p<0.001$). 온라인 소통 여부 또한 사회적 고립군 내에서 온라인 대화를 하지 않는 경우 62.3%, 사회적 비고립군에서는 42.8%로 나타나 통계적으로 유의미한 차이가 있었다($p<0.001$).

#### 심리적 요인

외로움 점수는 사회적 고립군이 평균 2.8점(±0.914), 사회적 비고립군이 평균 2.2점(±0.748)으로 나타나, 사회적 고립 여부에 따라 통계적으로 유의한 차이가 있었다($p<0.001$). 우울 점수는 사회적 고립군이 평균 2.2점(±0.884), 사회적 비고립군이 평균 1.6점(±0.682)으로 나타나, 사회적 고립 여부에 따라 유의한 차이가 있었다($p<0.001$).

#### 생활 습관 및 건강 요인

수면의 질에 따라 사회적 고립군 내에서는 수면의 질이 낮은 경우 38.5%, 좋은 경우 61.5%, 사회적 비고립군에서는 각각 21.9%, 78.1%로 나타나 유의미한 차이가 있었으나($p<0.001$), 평균 수면시간에 따른 분포는 통계적으로 유의미한 차이는 없었다($p=0.085$). 자기관리 및 청결 항목 중 옷 갈아입기, 목욕·샤워, 방 정리 여부는 사회적 고립 여부에 따라 유의미한 차이가 있었으나($p<0.001$), 세수·양치 여부는 유의미한 차이가 없었다($p=0.385$). 주관적 건강 상태는 사회적 고립군 내에서 건강 상태를 좋지 않음으로 인식한 경우 37.7%, 사회적 비고립군에서는 11.4%로 나타나 유의미한 차이가 있었으며($p<0.001$), 약물 복용 여부는 유의미한 차이가 없었다($p=0.065$). 식사 습관에서는 사회적 고립군 내에서 결식·저빈도 식사 17.0%, 정상·규칙적 식사 83.0%, 사회적 비고립군에서는 각각 5.9%, 94.1%로 나타나 유의미한 차이가 있었고($p<0.001$), 식사 방식 또한 사회적 고립군 내에서는 혼자 식사 63.9%, 가족과 식사 36.1%, 사회적 비고립군에서는 각각 52.2%, 47.8%로 나타나 통계적으로 유의미한 차이가 있었다($p<0.001$).

#### 생애사적 경험 요인

성인기 이전 경험을 살펴보면, 사회적 고립군 내에서는 가정의 경제적 어려움 경험 54.1%, 보호자로부터의 학대 또는 모욕 경험 45.6%, 가족의 정서적 어려움 경험 56.0%, 따돌림 경험 46.7%로 나타났으며, 사회적 비고립군 내에서는 각 경험 항목에서 각각 37.4%, 32.2%, 40.7%, 35.1%로 나타나, 사회적 고립 여부에 따라 통계적으로 유의미한 차이가 있었다($p<0.001$). 반면 전학·이사 경험은 두 집단 간 유의미한 차이가 없었다($p=0.392$).

성인기 이후 경험의 경우, 사회적 고립군 내에서는 경제적 이유로 인한 학업 중단 14.1%, 진로 및 취업에 대한 압박 경험 38.7%, 원하는 시기의 취업 실패 61.8%, 희망 직장 진입 실패 59.4%, 신뢰했던 사람으로부터의 배신 또는 사기 경험 44.8%, 가까운 관계의 단절 또는 상실 경험 51.7%로 나타났으며, 사회적 비고립군 내에서는 각 경험 항목에서 각각 9.9%, 29.2%, 45.7%, 49.4%, 31.8%, 41.6%로 나타나, 사회적 고립 여부에 따라 유의미한 차이가 있었다($p<0.001$). 반면 희망 대학 진학 실패 경험은 통계적으로 유의미하지 않았다($p=0.106$).

### 기계학습 기반 모형별 사회적 고립 예측 성능 비교

#### K-최근접 이웃(KNN) 분석 결과

Figure 1A는 K-최근접 이웃 모형의 사회적 고립 여부 분류 결과를 혼동행렬 형태로 제시한 것이다. ROC 기반 최적 임곗값(threshold = 0.073)을 적용하여 KNN 모형의 예측 성능을 분석한 결과, 전체 정확도는 73.0%로 나타났다. 전체 1,606명 중 1,168명을 올바르게 분류하였다. 실제 비고립군 1,493명 중 1,082명을 비고립군으로 올바르게 예측하였다. 반면, 고립군은 전체 113명 중 86명(76.1%)을 고립군으로 올바르게 분류하였고, 27명(23.9%)은 비고립군으로 분류되었다. 고립군의 재현율(recall)은 0.76, 정밀도는 0.17, F1-score는 0.28로 나타났다. ROC-AUC 값은 0.7918로 나타났다. KNN은 재현율(0.76)은 높았으나 정밀도(0.17)와 정확도(0.73)가 낮아, 고립군 탐지 과정에서 오탐이 증가하는 분류 성능을 보였다.

#### 나이브 베이즈(Naive Bayes) 분석 결과

Figure 1B는 나이브 베이즈 모형의 사회적 고립 여부 분류 결과를 혼동행렬 형태로 제시한 것이다. ROC 곡선 기반 최적 임곗값(threshold = 0.064)을 적용하여 나이브 베이즈 모형의 예측 성능을 분석한 결과, 전체 정확도는 79.0%로 나타났다. 전체 1,606명 중 1,272명을 올바르게 분류하였으며, 비고립군에서 실제 1,493명 중 1,187명을 정확히 예측하였다. 반면, 고립군은 전체 113명 중 85명(75.2%)을 정확히 분류하여 재현율은 0.75로 나타났다. 정밀도는 0.22로 나타났으며, F1-score는 0.34이었다. ROC-AUC 값은 0.8345로 나타났다. 나이브 베이즈는 재현율(0.75)이 높았으나 정밀도(0.22)와 정확도(0.79)가 낮아, 고립군 판별의 정밀성이 충분하지 않은 분류 성능을 보였다.

#### 의사결정나무(Decision Tree) 분석 결과

Figure 1C는 의사결정나무 모형의 사회적 고립 여부 분류 결과를 혼동행렬 형태로 제시한 것이다. ROC 곡선 기반 최적 임곗값(threshold = 0.749)을 적용하여 의사결정나무 모형의 예측 성능을 평가한 결과, 전체 정확도는 83.0%로 나타났다. 전체 1,606명 중 1,332명을 올바르게 분류하였다. 실제 비고립군 1,493명 중 1,260명(84.4%)을 비고립군으로 정확히 예측하였다. 반면, 고립군은 전체 113명 중 72명(63.7%)을 정확히 분류하여 재현율은 0.64로 나타났다. 정밀도는 0.24로 나타났

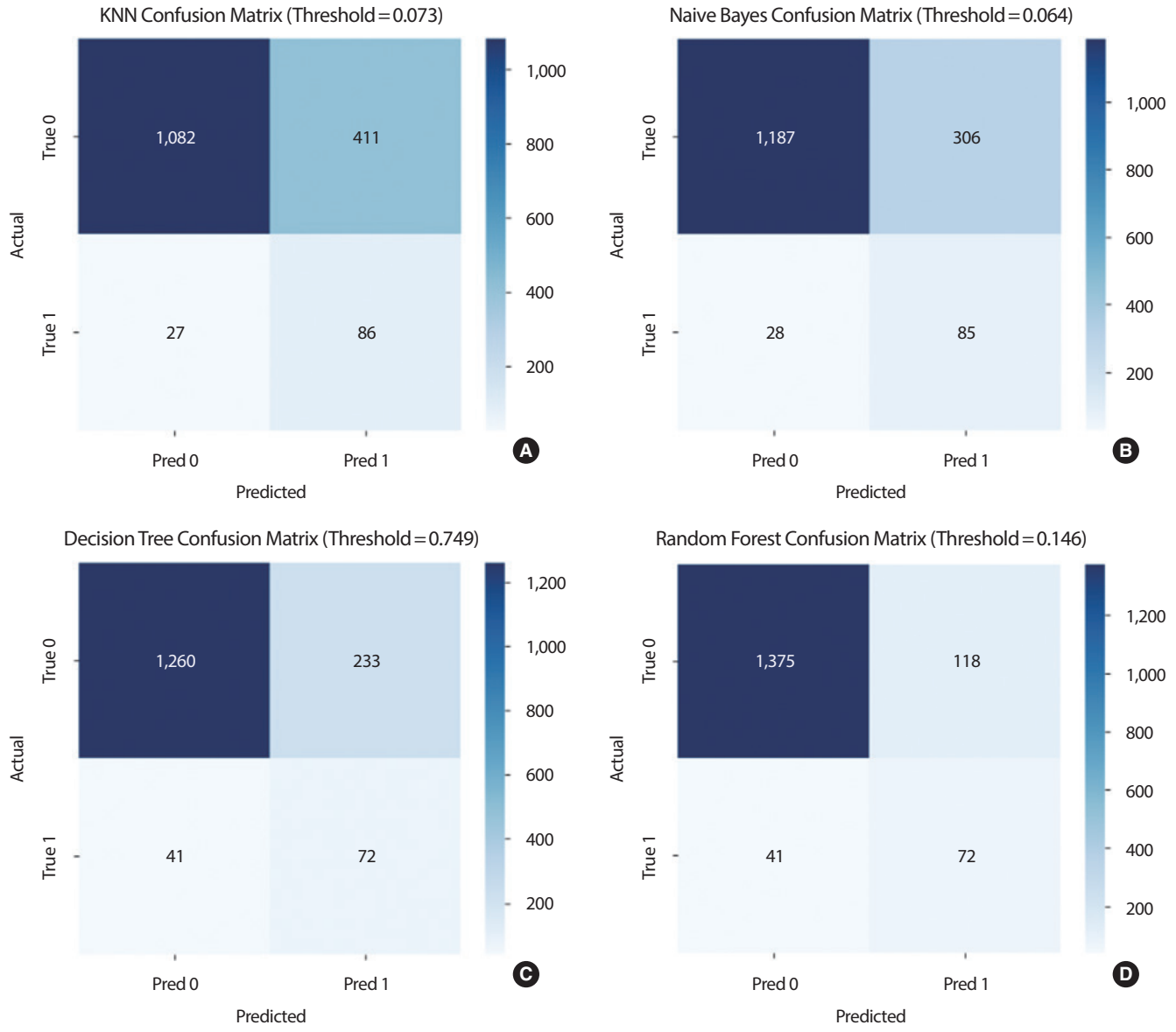


**Figure 1.** Comparison of confusion matrices for four machine learning models. (A) KNN, (B) Naive Bayes, (C) Decision Tree, (D) Random Forest. KNN, k-nearest neighbors.

으며, F1-score는 0.34로 나타났다. ROC-AUC 값은 0.7802로 나타났다. 의사결정나무는 정확도(0.83)와 ROC-AUC (0.7802)가 비교적 양호했으나 정밀도(0.24)가 낮아, 정확성에 한계가 있는 분류 성능을 보였다.

#### 랜덤포레스트(Random Forest) 분석 결과

Figure 1D는 랜덤포레스트 모형의 사회적 고립 여부 분류 결과를 혼동행렬 형태로 제시한 것이다. ROC 기반 최적 임곗값(threshold = 0.146)을 적용하여 랜덤포레스트 모형의 예측 성능을 평가한 결과, 전체 정확도는 90.0%로 나타났다. 전체 1,606명 중 1,447명을 올바르게 분류하였다.

비고립군의 경우 실제 1,493명 중 1,375명을 정확히 예측하였다. 반면 고립군은 전체 113명 중 73명(63.7%)을 정확히 분류하여 재현율은 0.64로 나타났다. 정밀도는 0.38로 나타났으며, F1-score는 0.48로 나타났다. ROC-AUC 값은 0.8538로 나타났다. 랜덤포레스트는 정확도(0.90)와 ROC-AUC (0.8538)가 비교적 높고 정밀도(0.38)도 상대적으로 양호하여, 전반적으로 균형적인 분류 성능을 보였다.

**Table 2.** Comparison of classification performance metrics across machine learning models

| Model | Accuracy | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|
| KNN | 0.7273 | 0.1730 | **0.7611** | 0.2820 | 0.7918 |
| Naive Bayes | 0.7920 | 0.2174 | 0.7522 | 0.3373 | 0.8345 |
| Decision tree | 0.8294 | 0.2361 | 0.6372 | 0.3445 | 0.7803 |
| Random forest | **0.9010** | **0.3789** | 0.6372 | **0.4752** | **0.8538** |

KNN, k-nearest neighbors; AUC, area under the curve.
Boldface indicates the highest value for each evaluation metric.

### 기계학습 모형 간 성능 지표 비교

Table 2는 사회적 고립 여부 예측을 위해 적용한 네 가지 기계학습 모형의 분류 성능 지표를 비교한 것이다. 성능 지표(정확도, 정밀도, 재현율, F1-score, ROC-AUC)를 비교한 결과, 각 모형은 사회적 고립 여부를 예측하는 데 서로 다른 성능 특성을 보였다. 정확도 기준에서는 랜덤포레스트(0.9010)가 가장 우수한 성능을 보였고, 의사결정나무(0.8294), 나이브 베이즈(0.7920), KNN (0.7273) 순으로 나타났다. 정밀도는 네 모형 모두에서 낮은 수준으로 나타났다. 이는 고립군의 실제 표본 비율이 113명(7%)으로 적어, 고립군과 비고립군 간 변별력이 제한되었다. 네 모형 중에서는 랜덤포레스트가 가장 높은 정밀도(0.3789)를 보였으며, 의사결정나무(0.2361), 나이브 베이즈(0.2174), KNN (0.1730) 순이었다.

재현율은 모형 간 차이가 가장 뚜렷하였다. KNN (0.7611)이 가장 높은 재현율을 보여 고립군 탐지 능력이 상대적으로 우수했으며, 나이브 베이즈(0.7522)가 유사한 수준으로 뒤를 이었다. 반면 의사결정나무(0.6372)와 랜덤포레스트(0.6372)는 상대적으로 낮은 재현율을 보여 고립군 탐지 측면에서 제한이 있었다.

F1-score는 정밀도와 재현율 간 균형을 반영하는 지표로, 랜덤포레스트(0.4752)가 가장 높아 전반적으로 균형적인 분류 성능을 보였다. 다음으로 의사결정나무(0.3445), 나이브 베이즈(0.3373), KNN (0.2820) 순으로 나타났다. 그중 KNN은 재현율이 높지만, 정밀도가 낮아 F1-score가 상대적으로 낮게 나타났다.

Figure 2는 네 가지 기계학습 모형의 분류 성능을 ROC 곡선을 통해 비교한 것이다. ROC-AUC를 기준으로 비교한 결과, 랜덤포레스트(0.8538)가 가장 우수한 판별력을 보였고, 나이브 베이즈(0.8345)가 뒤를 이었다. KNN (0.7918)과 의사결정나무(0.7803)는 상대적으로 낮아 ROC 기준의 분류 성능은 제한적이었다.

종합적으로, 랜덤포레스트는 높은 정확도와 우수한 ROC-AUC, 높은 F1-score로 네 모형 중 안정적인 예측 성능과 가장 균형 잡힌 성능을 보였다. 나이브 베이즈는 고립군 탐지(재현율)가 높았지만, 정밀도가 낮아 오탐 가능성이 있었고, KNN은 높은 재현율을 보였지만 정밀도와 F1-score가 낮아 고립군 판별에 한계가 있었다. 의사결정나무는 정확도

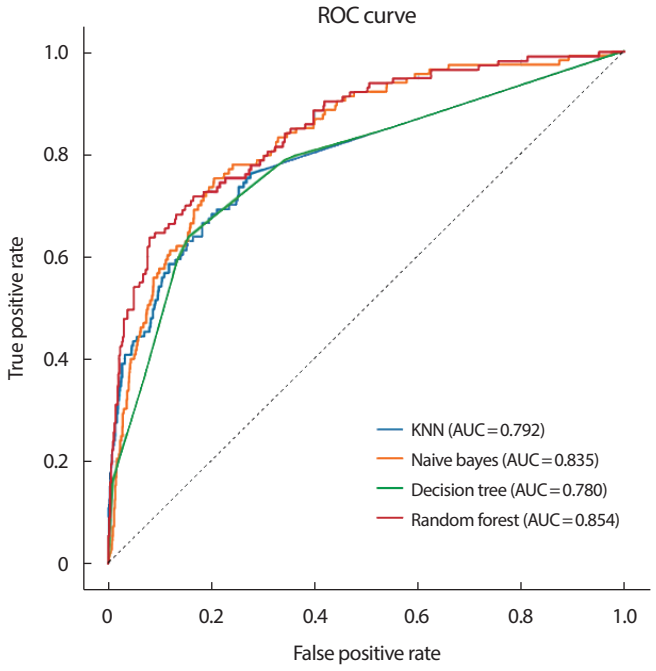


**Figure 2.** Comparison of ROC curves across machine learning models.
ROC, receiver operating characteristic.

는 중간 수준이었지만, AUC와 F1-score가 상대적으로 낮아 전반적인 성능이 낮았다. ROC-AUC는 이진 분류기의 전반적인 성능을 평가하는 데 널리 사용되는 지표이나[39], 정책적 선별 관점에서는 재현율 중심 평가가 중요할 수 있다. 따라서 향후 정책적 적용 시 cut-off 조정을 통해 재현율과 정밀도의 균형을 상황에 맞게 설정할 필요가 있다.

### 변수 중요도 분석

본 연구에서 사회적 고립 여부에 영향을 미치는 핵심 요인을 탐색하기 위해 의사결정나무와 랜덤포레스트 모형의 변수 중요도를 분석하였다. 두 모형 모두 사회적 고립을 분류하는 데 기여도가 높은 변수를 상위 10개의 중요 변수로 산출하였으며, 비교 결과 일부 변수들이 공통적으로 높은 영향력을 보이는 것으로 확인되었다.

Figure 3A는 의사결정나무 모형에서 산출된 변수 중요도 상위 10개

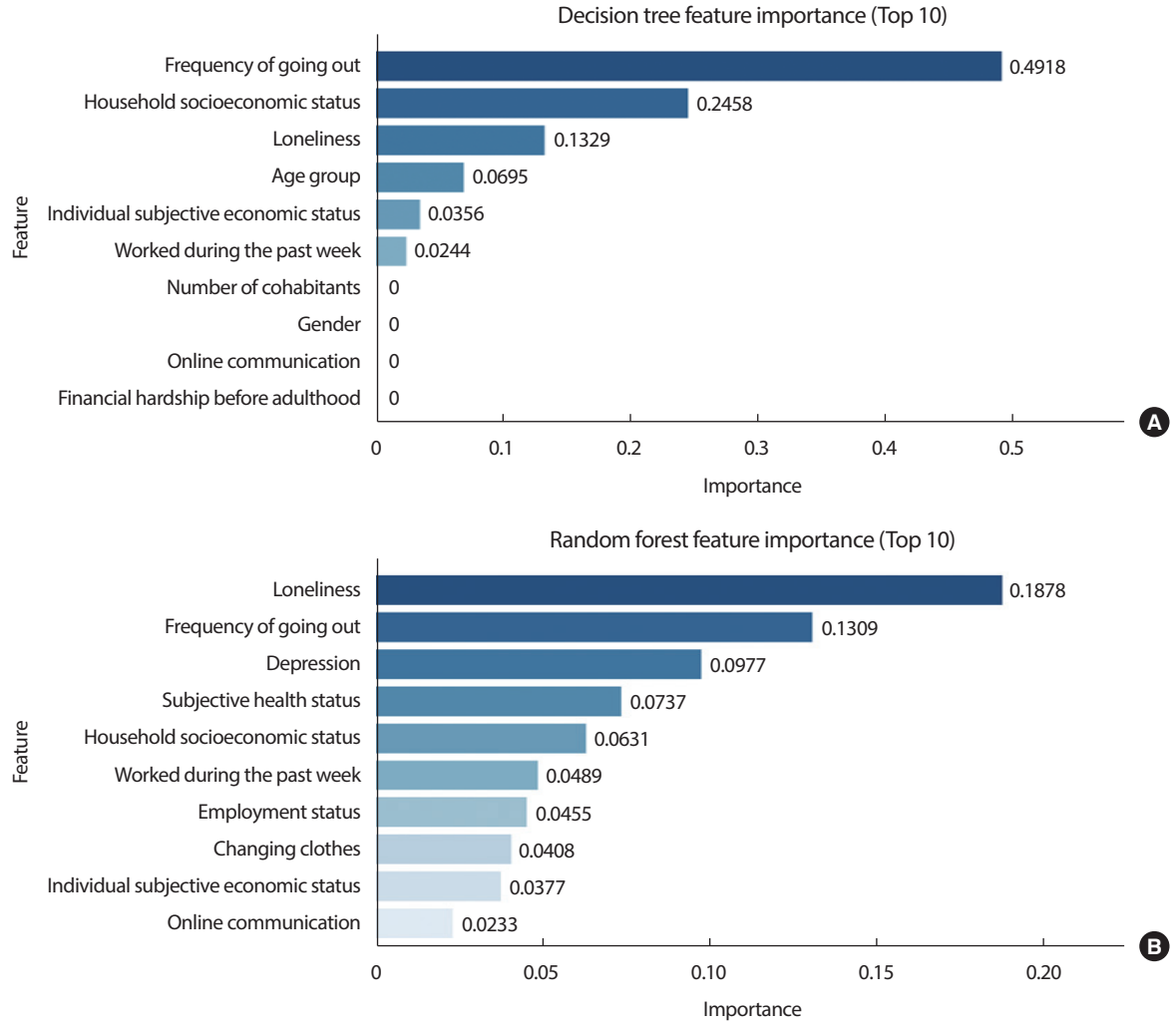


**Figure 3.** Top 10 feature importances of the prediction models. (A) Top 10 feature importances of the Decision Tree model. (B) Top 10 feature importances of the Random Forest model.

를 제시하였다. 분석 결과 외출 빈도가 중요도 0.4918로 가장 높게 나타나 사회적 고립을 예측하는 핵심 분기로 활용되었다. 다음으로는 가정의 사회경제적 수준(0.2458), 외로움(0.1329), 연령대(0.0695), 본인의 주관적 경제 수준(0.0356), 지난주 노동 여부(0.0244)가 뒤를 이었다. 그러나, 동거인 수, 성별, 온라인 대화 여부, 성인기 이전 경제적 곤란 경험은 중요도가 매우 낮거나(0.0000) 분류에 기여가 제한적인 것으로 나타났다.

Figure 3B는 랜덤포레스트 모형을 기준으로 산출한 변수 중요도 상위 10개를 시각화한 것이다. 분석 결과 외로움이 중요도 0.1878로 가장 높은 중요도를 보였으며, 외출 빈도(0.1309), 우울(0.0977)이 이어 주요 예측 요인으로 도출되었다. 또한 주관적 건강 상태(0.0737), 가정의 사회경제적 수준(0.0631), 지난주 노동 여부(0.0489), 직업 상태(0.0455), (위생/생활 관리 관련) 옷 갈아입기(0.0408), 본인의 주관적 경제 수준(0.0377), 온라인 대화 여부(0.0233) 등이 주요 변수로 나타났다.

의사결정나무에서는 외출 빈도와 가정의 사회경제적 수준이 높은 영향력을 보였고, 랜덤포레스트에서는 외로움·우울 등 심리적 요인과 외출 행동, 건강 및 사회경제적 요인이 함께 상위 변수로 나타났다. 이는 사회적 고립 예측에서 행동 지표(외출)와 사회경제적 조건, 심리·건강 요인이 사회적 고립을 예측하는 데 핵심적인 역할을 하는 것을 확인하였다.

## 고찰 및 결론

본 연구는 2022년 서울시 실태조사 자료를 활용하여 기계학습 알고리즘을 통해 청년의 사회적 고립 여부를 예측하였다. 각 모형의 성능 비교 및 주요 예측 요인을 분석한 결과, 이에 따른 고찰 및 결론은 다음과

같다.

기계학습 모형의 예측 성능 비교 결과, 랜덤포레스트 모형이 가장 안정적인 성능을 보였다. 연구 결과, 전체 대상자 중 사회적 고립군은 7.04% (113명)로 나타나 자료의 불균형이 발생하였다. 이를 고려하여 KNN, 나이브 베이즈, 의사결정나무, 랜덤포레스트 모형을 적용하고 정확도, 정밀도, 재현율, F1-score, ROC-AUC를 중심으로 성능을 평가하였다. 모형 비교 결과, 랜덤포레스트 모형이 정확도(0.9010)와 ROC-AUC (0.8538), F1-score (0.4752)에서 가장 우수한 성능을 보였으며, 정밀도(0.3789)와 재현율(0.6372)의 균형이 비교적 안정적으로 유지되었다. 이는 불균형 환경에서 단일 모델보다 앙상블 기법이 과적합 방지와 안정적인 예측 성능을 나타냄을 보여준다.

반면 KNN과 나이브 베이즈 모형은 재현율이 각각 0.7611, 0.7522로 높게 나타나 고립군 탐지에 민감한 특성을 보였으나, 정밀도(0.1730, 0.2174)가 낮아 오탐 가능성이 높은 한계를 보였다. 의사결정나무는 정확도가 0.8294로 비교적 높았으나 재현율(0.6372)과 F1-score (0.3445)가 중간 수준에 머물렀다. 최종적으로 랜덤포레스트는 정확도(0.9010), ROC-AUC (0.8538), F1-score (0.4752) 전반적으로 가장 안정적인 성능을 보이며, 고립군 탐지의 효율성과 예측의 신뢰성을 동시에 고려할 때 가장 적합한 모형으로 판단된다.

변수 중요도 분석 결과, 사회적 고립은 행동적·심리적·사회경제적 요인이 복합적으로 작용하는 양상으로 나타났다.

의사결정나무 모형에서는 '외출 빈도'가 가장 높은 중요도(0.4918)를 차지하였다. 그다음으로 '가정의 사회경제적 수준'과 '외로움'이 주요 변수로 나타났다. 랜덤포레스트 모형에서는 '외로움'이 가장 높은 중요도(0.1878)를 보였으며, 이어 '외출 빈도', '우울', '주관적 건강 상태', '사회경제적 수준' 등이 상위 변수에 포함되었다.

이는 사회적 고립이 단일 요인에 의해 설명되기보다는, 사회적 활동 감소와 같은 행동 변화와 심리적 취약성, 그리고 경제적·환경적 조건이 상호작용하며 나타나는 복합적 현상임을 시사한다. 그러나 랜덤포레스트 기반 변수 중요도는 변수 간 상관관계 및 측정 수준의 영향을 받을 수 있으므로, 본 연구의 결과는 인과적 해석보다는 예측 기여도의 관점에서 해석될 필요가 있다.

이러한 결과는 사회적 고립을 사회적 관계의 양과 질이라는 객관적 차원에서 이해해야 한다는 기존 연구의 개념적 구분과 일치한다. Park and Kim [14]은 사회적 고립을 사회적 관계의 객관적 특성으로, 외로움은 해당 관계 상태에 대한 개인의 주관적 인식으로 구분하였다. 본 연구에서 외출 및 교류 변수가 핵심 예측 요인으로 작용한 점은, 개념적으로 구분되어 실제 자료에서도 일정 부분 반영되고 있음을 의미한다. 또한 청년기 사회적 고립이 은둔으로 이행될 가능성이 높다는 점을 지적한 Noh et al. [17]의 연구를 고려할 때, 본 연구에서 확인된 외출 및 교류 감소는 은둔 이전 단계에서 발견할 수 있는 행동 수준의 위험 지표로 해석될 수 있다. 이는 사회적 고립이 특정 시점에 발생하기보다는, 일상적 활동 감소의 누적 과정에서 심화할 가능성을 보여준다.

변수 중요도 분석에서 경제적 요인으로 가정의 사회경제적 수준과 개인의 주관적 경제 수준이 상위 변수로 반복적 도출되었다. 이는 사회적 고립이 개인의 문제가 아닌 구조적·환경적 조건과 결합한 현상으로 이해해야 한다는 기존 연구의 문제의식과 부합한다. Kim et al. [30]은 청년의 사회적 고립을 사회적 관계, 사회적 지지, 은둔 행동이 결합한 다차원적 현상으로 유형화하고, 그 과정에서 인구사회학적·경제적 변수가 유형 구분에 중요한 역할을 한다고 보고하였다. 본 연구는 이러한 논의를 예측 모형의 관점에서 확장하여, 경제적 여건이 관계 단절 및 심리적 취약성과 결합되면서 고립 위험을 구분하는 데 중요한 변수로 작용하는 것으로 나타났다.

또한 주관적 건강 상태, 생활 패턴, 우울 관련 변수 등이 상위 예측 요인으로 포함된 점은 사회적 고립이 사회적 관계 차원에 국한되지 않음을 시사한다. 이는 사회적 고립이 일상 기능과 생활 패턴, 정신건강 상태와 밀접하게 연결된 과정임을 보여준다. 이러한 결과는 사회적 고립이 특정 시점에서 갑작스럽게 발생하기보다 생애과정 전반에 걸쳐 점진적으로 심화한다는 Jung [26]의 논의와 연결된다. 또한 사회적 고립과 외로움이 우울과 같은 정신건강 문제와 상호 연관된 다요소적 현상이라는 기존 연구 결과[14,31]를 뒷받침한다. 따라서, 본 연구에서 건강 및 생활 습관 요인이 사회적 관계 및 경제적 요인과 함께 중요 변수로 도출된 결과는, 고립 위험이 단일 영역에서 발생하는 문제가 아님을 보여준다. 즉, 관계·환경·건강 요인이 상호작용하는 구조 속에서 축적되는 현상으로 이해할 수 있다. 이는 사회적 고립을 단편적 지표로 파악하기보다는, 다차원적이며 통합적인 관점에서 이해하고 접근해야 할 필요성을 의미한다.

본 연구는 기존 연구에서 제시된 사회적 고립의 다차원적 특성과 전반적으로 일치하며, 예측 모형을 통해 각 요인의 상대적 기여도를 제시하였다는 점에서 차별성을 가진다. 선행 연구에서 사회적 고립을 외로움·우울 등 심리적 요인[14,31]과 사회적 관계의 축소, 생애과정에서 누적 및 심화하는 과정으로 설명하고[26], 청년기의 사회적 활동 감소는 고립과 은둔으로 이어질 수 있다고 하였다[17]. 이는 본 연구에서 행동적 요인(외출 빈도), 심리적 요인(외로움·우울), 사회경제적 요인이 중요 변수로 도출되었다는 점과 유사하다.

한편, 의사결정나무 모형에서 외출 빈도가 가장 높은 중요도를 보였다. 이는 심리 변수보다 강한 기준으로 나타났다는 점에서, 행동의 사회적 활동 감소가 고립 예측에서 핵심적인 신호로 작용할 수 있음을 실증적으로 보여준다. 선행 연구에서 심리적 취약성 및 관계 인식에 초점을 두었던 것과는 달리[14,31], 일상적 활동 감소라는 관찰 가능한 지

표의 변화를 통해 예측적 의미를 구체적으로 확인할 수 있었다. 또한 사회적 고립의 개념 및 이론적 설명이었던 것과는 달리, 예측 모형을 통해 실제 분류 과정에서 어떤 변수가 주요 요인으로 작용하는지 정량적으로 제시했다는 점에서 차별성을 가진다.

이와 같은 결과를 통해 현재 국내의 사회적 고립 청년 지원 정책의 한계를 보완할 수 있는 시사점을 제공한다. 국내 정책은 일부 지자체 조례와 개별 사업을 중심으로 추진되고 있다. 그러나 법적 제도의 미비와 단기적·사후적 개입 중심의 구조로 인하여 조기 발견과 예방적 대응에 한계점이 있다고 지적됐다[22]. 이에 따라 본 연구에서 확인된 외출 및 교류 감소와 같은 행동 지표는 고립이 장기화하기 이전 단계에서 위험 신호를 포착할 수 있는 실증적 근거로 해석될 수 있다.

특히 영국, 미국, 일본 등 해외 사례에서는 사회적 고립을 국가가 대응해야 할 사회적 위험으로 인식하고 있다[22]. 국가 차원의 측정 체계와 장기적 정책 프레임을 구축하고 있다는 점에서 국내와 차이를 보인다[22]. 이러한 점을 고려하면, 본 연구의 예측 결과는 국내 정책에도 적용 가능한 함의를 제시한다. 즉, 사회적 고립을 단일 사건이 아닌 과정적 위험으로 인식하고, 행동·환경·건강 요인을 종합적으로 고려한 조기 선별 관점의 접근이 필요함을 보여준다.

본 연구는 의사결정나무와 랜덤포레스트 모형에서 공통으로 도출된 변수 중요도를 통해, 청년의 사회적 고립이 단면적으로 적용되는 것이 아니라 사회적 관계·교류 요인, 사회경제적 여건, 건강·생활 습관 요인이 복합적으로 작용함을 확인하였다.

이를 통해 고립·은둔 청년 발굴 및 지원 체계에서 단일 지표에 의존하기보다, 핵심 행동 지표(외출·교류), 환경적 조건(경제 수준), 생활·건강 요인을 함께 고려한 통합적 스크리닝 도구로 활용될 수 있음을 시사한다. 또한 랜덤포레스트의 ROC-AUC (0.8538)는 핵심 문항 기반 자료만으로도 일정 수준의 조기 선별 가능성을 보여준다. 본 연구에서 도출된 예측 결과를 통해 실제 정책 현장에서 고립 위험 청년을 조기 식별하고 위험군을 분류하며 개입의 우선순위를 설정하는 데 활용될 수 있다. 예를 들어 외출 빈도 감소와 같은 행동 지표는 1차 스크리닝 기준으로 적용할 수 있으며, 심리적·경제적 취약성이 함께 나타나는 경우 우선 지원 대상자로 분류하는 등 선제적 개입의 근거로 쓰일 수 있을 것이다.

본 연구는 모델의 정밀도가 0.17-0.38 수준으로, 상대적으로 낮게 나타났다는 한계를 지닌다. 이는 실제 고립이 아닌 청년을 고립군으로 오분류할 가능성, 즉 위양성(false positive)이 증가할 수 있음을 의미한다. 이러한 결과는 고립군 비율이 낮은 클래스 불균형이 모델 학습과 예측에 영향을 미친 결과로 해석된다. 그러나 본 연구의 예측 모형은 명확한 진단 도구라기보다 고위험군을 조기에 선별하기 위한 스크리닝 도구의 가능성을 탐색한 것으로, 위양성 사례는 후속 심층 상담이나 추가 평가 절차를 통해 재확인·조정되는 2단계 선별 체계에서 보완될 수 있다. 향후 연구에서는 텍스트 마이닝 기반 정보 확장 또는 딥러닝 기반 불균형 자료 처리 기법 등을 적용하여 모델의 정밀도를 개선하고 오탐률을 낮추는 방안을 모색할 필요가 있다.

또한 2022년 서울시 고립·은둔 청년 실태조사 자료를 활용하였다는 점에서 지역적 특수성이 존재하고, 연구 결과에 따른 일반화 가능성에 한계점이 있다. 따라서 연구의 결과를 전국 청년을 대상으로 일반화하는 데에는 신중한 해석이 필요하다. 실제 수도권과 비수도권 청년의 고립 예측을 비교한 선행 연구 결과 연령, 상대적 박탈감, 우울과 같은 공통 요인 외에도, 수도권에서는 주거 인식과 물질적 박탈, 비수도권에서는 미래 전망 및 경제위기 통제 가능성이 유의한 예측 요인으로 나타나 지역 간 차이가 존재함을 확인할 수 있다[40]. 이에 따라 후속 연구에서는 타 광역시 또는 전국 단위의 자료를 활용한 비교 분석을 통하여 지역 간 차이를 검증하고, 예측 모형의 타당도를 확장할 필요가 있다.

아울러 서울시 정책 기준에 따라 사회적 고립 여부를 이분형 변수로 구성하였으나, 고립의 정도나 연속적 변화 양상을 충분히 반영하지 못하는 한계가 있다. 향후 연구에서는 다범주 분류 또는 점수 기반 고립 지수 구성을 통해 고립의 단계성과 심각도를 더욱 구체적으로 반영할 필요가 있다. 결측치 처리는 완전 제거법을 적용하였다. 그러나 단순 사례 제거는 가장 전통적인 결측치 처리 방법이지만, 결측 발생 메커니즘에 대한 검토 없이 적용할 경우 표본 편향 가능성이 존재한다. 따라서 향후 연구에서는 대체 값에 임의 오차를 포함하여 불확실성을 반영하는 다중 대치를 적용하여 안정성과 일반화 가능성을 정교하게 검증할 필요가 있다[41].

본 연구는 2022년 서울시 실태조사 자료를 분석하여 기계학습 기반의 사회적 고립 예측 모델을 구축하고 주요 취약 요인을 확인하였다. 랜덤포레스트 모형은 불균형 데이터 환경에서도 가장 안정적인 예측 성능을 보였으며, 외로움·우울과 같은 심리적 요인과 외출 행동, 건강 및 사회경제적 요인 등의 변수들이 사회적 고립 예측에 가장 큰 영향을 미치는 것으로 확인되었다. 이는 사회적 고립이 개인의 주관적 정서에 국한된 현상이 아닌, 외출·교류와 같은 사회활동과 사회경제적·건강 요인이 함께 반영되는 복합적 현상임을 시사한다. 따라서 핵심 예측 요인들은 현재 국내 사회적 고립 청년 지원 정책에서 상대적으로 미흡한 조기 발견 및 예방적 접근의 필요성을 뒷받침하는 실증적 근거로 해석될 수 있다.

향후 연구에서는 SHapley Additive exPlanations (SHAP) 등 설명 가능한 인공지능(XAI, eXplainable AI) 기법을 활용하여 각 변수의 기여 크기와 방향성을 세밀하게 해석할 수 있다[42,43]. 또한, SHAP 요약 플롯(summary plot)을 통한 시각화로 요인 간 상호작용 효과까지 설명함

으로써[44], 고립 위험 청년의 조기 선별 및 개입 전략 수립에 실질적인 근거를 제공할 수 있을 것으로 기대된다.

본 연구는 사회적 고립 청년 발굴 과정에서 단일 지표에 의존하기보다는 행동·환경·생활 요인을 종합적으로 고려하여 데이터 기반의 선별 접근 가능성을 제시한다는 점에서 의의를 가진다.

## ORCID

Hyunbi Park *https://orcid.org/0009-0009-3640-0843*
Byeol Choi *https://orcid.org/0009-0000-7733-7603*
Boyoung Jeon *https://orcid.org/0000-0002-2261-253X*

## REFERENCES

1. Yun S, Lee GM, Kim J. Association between the level of social relationships and depression: Focusing on employment status. Korea J Hos Manag 2025;30(1):29-44 (Korean).
2. Shin H. Social networks of friends and helpers: Structural features, social interaction, and social influence. Korean J Psychol Gen 2022;41(1):39-61 (Korean). DOI: 10.22257/kjp.2022.3.41.1.39
3. Jung HS. The impact of negative experiences in early life on social isolation among young adults [dissertation]. Daegu University; 2024, Korea.
4. Jang SG, Lee JS, Lee JY. The need for comprehensive measures to combat loneliness: Fostering social relationships and a connected society in Jeonbuk (Issue Brief No. 286). Jeonbuk Research Institute; 2023.
5. Park CU, Kim N, Liu F, Yoon M. A study of social isolation and its socio-demographic characteristics in Korea. 2020;21(2):41-73 (Korean). DOI: 10.22862/kjsr.2020.21.2.002
6. Jeong S, Ryu J, Kang Y, Kim S, Ham S, Kim D, et al. Survey on the lives of youth in 2022. Sejong: Korea Institute for Health and Social Affairs; 2022 (Korean).
7. Kim SA, Kim MG, Lim DY, Noh HJ, Noh HJ. Gorip·eundun cheongnyeon siltae josa yeongu [Survey on socially isolated and reclusive youth]. Sejong: Korea Institute for Health and Social Affairs; 2023 (Korean).
8. Seo YK, Ahn SK, Chung IJ. Exploring the feasibility of developing a reward application program to alleviate physical isolation of reclusive youth. J Korea Game Soc 2024;24(1):27-44 (Korean). DOI: 10.7583/JKGS.2024.24.1.27
9. Jeong H, Kim J. How young adults in Seoul experience social isolation: Focusing on the individual characteristics and neighborhood environmental factors. J Korean Reg Sci Assoc 2025;41(3):39-58 (Korean). DOI: 10.22669/krsa.2025.41.3.039
10. Kim J. A study on social isolation in youth: Focusing on social capital, economic stability, and relative deprivation. Korean Public Adm Q 2023;35(3):315-337 (Korean). DOI: 10.21888/KPAQ.2023.9.35.3.315
11. Lee JH. Youth loneliness from a first-person perspective. Seoul: Thaumazein; 2024, p. 140-147 (Korean).
12. Ryu J, Seong K, Kim S. How do young adults become socially isolated? J Soc Sci 2023;34(3):47-71 (Korean). DOI: 10.16881/jss.2023.07.34.3.47
13. Kim J, Park E. The effects of achievement type of developmental tasks of early adulthood on the perceived social isolation in youth. Stud Korean Youth 2016;27(3):257-284 (Korean). DOI: 10.14816/sky.2016.27.3.257
14. Park M, Kim S. Characteristics and types of loneliness, social isolation, and mental health problems in single-person households: Focusing on single-person households in Seoul. Health Soc Welf Rev 2022;42(4):127-141 (Korean). DOI: 10.15709/hswr.2022.42.4.127
15. Choi J, Cho M, Lee S. The development of a scale of social isolation among young adults. Korean J Soc Welf 2022;74(2):53-78 (Korean). DOI: 10.20970/kasw.2022.74.2.003
16. Park SM, Joo MJ, Lim JH, Jang SY, Park EC, Ha MJ. Association between hikikomori (social withdrawal) and depression in Korean young adults. J Affect Disord 2025;380:647-654. DOI: 10.1016/j.jad.2025.03.125
17. Noh GB, Lee SM, Kim JH. Analysis of the experiences and causes of the hikikomori in youth. Korean J Soc Welf 2021;73(2):57-81 (Korean). DOI: 10.20970/kasw.2021.73.2.003
18. Lee H, Nah G. Characteristics and mental health of isolated and reclusive young adults. J Korean Soc Wellness 2025;20(2):255-261 (Korean). DOI: 10.21097/ksw.2025.5.20.2.255
19. Lee SB, Shin YR, Yoon MS. The effects of relative deprivation of youth on suicide: The serial mediation effect of prospects and social isolation. Health Soc Welf Rev 2022;42(2):369-389 (Korean). DOI: 10.15709/hswr.2022.42.2.369
20. Byun G, Park M, Ryu A. The lives of youth based on the 2022 Seoul Youth Panel. Seoul: The Seoul Institute; 2024 (Korean).
21. Seoul Metropolitan Government. Survey report on socially isolated and withdrawn youth. Seoul: Seoul Metropolitan Government; 2022 (Korean).

22. Yoo M, Shin D, Lee M. A study on the status of social isolation among youth and support measures. Sejong: National Youth Policy Institute; 2021 (Korean).
23. Park D. The effects of social isolation and depression on life satisfaction among young adults: Moderated mediation effect of material deprivation. J Korea Acad Indstr Coop Soc 2025;26(5):187-196 (Korean). DOI: 10.5762/KAIS.2025.26.5.187
24. Nicholson Jr NR. Social isolation in older adults: An evolutionary concept analysis. J Adv Nurs 2009;65(6):1342-1352. DOI: 10.1111/j.1365-2648.2008.04959.x
25. Zavaleta D, Samuel K, Mills CT. Measures of social isolation. Soc Indic Res 2017;131(1):367-391. DOI: 10.1007/s11205-016-1252-2
26. Jung S. A study on support measures for preventing social isolation among people in their 20s and 30s in Gyeongsangbuk-do. Andong: Gyeongbuk Women's Policy Development Institute; 2024 (Korean).
27. Teo AR. A new form of social withdrawal in Japan: A review of hikikomori. Int J Soc Psychiatry 2010;56(2):178-185. DOI: 10.1177/0020764008100629
28. Hyland P, Shevlin M, Cloitre M, Karatzias T, Vallières F, McGinty G, et al. Quality not quantity: Loneliness subtypes, psychological trauma, and mental health in the US adult population. Soc Psychiatry Psychiatr Epidemiol 2019;54(9):1089-1099. DOI: 10.1007/s00127-018-1597-8
29. Byun GS, Kim JS. The loneliness, types of social isolation, and mental health among young adults: Focusing on differences by experience of moving to Seoul. Soc Welf Policy 2024;51(3):67-108 (Korean). DOI: 10.15855/swp.2024.51.3.67
30. Kim A, Lee H, Noh H. Identifying and predicting of patterns of social isolation among young adults. Soc Welf Policy 2025;52(1):187-218 (Korean). DOI: 10.15855/swp.2025.52.1.187
31. Yu J, Yang J, Kim J. The impact of social isolation on depression among young single-person households: Focusing on the moderating effects of the frequency of contact with family, friends, and neighbors. J Korean Soc Wellness 2024;19(3):279-285 (Korean). DOI: 10.21097/ksw.2024.8.19.3.279
32. Son JH, Ryu W. Prediction of recurrent falls in community-dwelling older adults using machine learning techniques. J Health Info Stat 2025;50(4):305-313 (Korean). DOI: 10.21032/jhis.2025.50.4.305
33. Ko K, Tak H, Lee B. Impact of missing values on survey research and relevancy of multiple imputation techniques. Korean J Policy Anal Eval 2014;24(3):49-75 (Korean). DOI: 10.23036/kapae.2014.24.3.003
34. Graham JW. Missing data analysis: Making it work in the real world. Annu Rev Psychol 2009;60:549-576. DOI: 10.1146/annurev.psych.58.110405.085530
35. Kwon S. Practical machine learning for real-world applications. Seoul: Hanbit Media; 2022 (Korean).
36. Bagayatkar A, Ivin B. Survey paper on machine learning and deep learning driven applications using bayesian techniques. 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). Pune, India, 2024, p. 1-7. DOI: 10.1109/I2CT61223.2024.10543602
37. Cho W. Decision tree analysis for discrete choice data [dissertation]. Korea University; Korea, 2020.
38. Yoo JE. Random forests, an alternative data mining technique to decision tree. J Educ Eval 2015;28(2):427-448 (Korean).
39. Lee SB, Jeon Y. A study on factors affecting social isolation among young adults-Focusing on regional comparison-. Crisisonomy 2023;19(6):125-141 (Korean). DOI: 10.14251/crisisonomy.2023.19.6.125
40. Tafvizi A, Avci B, Sundararajan M. Attributing auc-roc to analyze binary classifier performance. arXiv:2205.11781 (2022). DOI: 10.48550/arXiv.2205.11781
41. Schafer JL. Multiple imputation: A primer. Stat Methods Med Res 1999;8(1):3-15. DOI: 10.1177/096228029900800102
42. Xu S, Cüppers J, Vreeken J. Succinct interaction-aware explanations. arXiv:2402.05566 (2024). DOI: 10.48550/arXiv.2402.05566
43. Park S, Moon J, Jung SW, Jung SM, Hwang E. SHAP-based explainable influenza occurrence forecasting using LightGBM. Proceedings of the Korean Institute of Information Scientists and Engineers. 2020, Seoul, Korea, p. 666-668.
44. Xu C, Yu X, Ding Z, Fang C, Gao M, Liu W, et al. Artificial intelligence-assisted metastasis and prognosis model for patients with nodular melanoma. PLoS One 2024;19(8):e0305468. DOI: 10.1371/journal.pone.0305468

**Appendix Table 1.** Definitions and measurement of study variables

| Category | Predictor | Description and measurement |
|---|---|---|
| Sociodemographic factors | Gender | 1 = Male, 2 = Female |
| | Age group (y) | 1 = 19-24, 2 = 25-29, 3 = 30-34, 4 = 35-39 |
| | Number of cohabitants | 1 = Living alone, 2 = Two or more persons |
| | Education level | 1 = High school or below, 2 = College or above |
| | Employment status | 1 = Regular employment, 2 = Non-regular employment, 3 = Student, Unemployed, or Others |
| | Marital status | 1 = Married, 2 = Unmarried, 3 = Divorced or widowed |
| | House ownership | 0 = Non-owner, 1 = Homeowner |
| Economic and socioeconomic factors | Worked during the past week | 0 = Did not work, 1 = Worked |
| | Household socioeconomic status | 0 = Lower-middle or below, 1 = Upper-middle or above |
| | Individual subjective economic status | 0 = Insufficient, 1 = Sufficient |
| Social relationship and interaction factors | Frequency of going out | 1 = Active outings, 2 = Moderate outings, 3 = Inactive outings |
| | Online communication | 1 = Yes, 2 = No |
| Psychological factors | Loneliness scale | Mean score of four loneliness items (e.g., "I feel that there are not enough people who are willing to stay with me"); dichotomized as 0 = Not lonely (at or below the mean), 1 = Lonely (above the mean) |
| | Depression scale | Mean score of nine depression items; dichotomized as 0 = Not depressed (at or below the mean), 1 = At risk of depression (above the mean) |
| Lifestyle and health factors | Daily life patterns | |
| | Reversed day–night routine | 0 = No, 1 = Yes |
| | Regular sleep schedule | 0 = No, 1 = Yes |
| | Sleep quality | 0 = No, 1 = Yes |
| | Self-care and cleanliness | |
| | Changing clothes | 0 = Not frequent, 1 = Daily |
| | Bathing/Showering | 0 = Not frequent, 1 = Daily |
| | Washing face/Brushing teeth | 0 = Not frequent, 1 = Daily |
| | Room organization | 0 = Not frequent, 1 = Daily |
| | Subjective health status | 0 = Poor, 1 = Fair or good |
| | Medication use | 1 = Yes, 2 = No |
| | Average sleep duration | 1 = Less than 6 hours, 2 = 6 hours or more |
| | Meal frequency | 1 = Normal/Regular meals, 2 = Skipped/Low-frequency meals |
| | Meal method | 1 = Eating alone, 2 = Eating with family |
| Life-course experience factors | Pre-adulthood experiences | |
| | Financial hardship before adulthood | 1 = No, 2 = Yes |
| | Abuse or humiliation by caregivers | 1 = No, 2 = Yes |
| | Family member's emotional distress | 1 = No, 2 = Yes |
| | Loss of a close person | 1 = No, 2 = Yes |
| | Frequent relocation or school transfer | 1 = No, 2 = Yes |
| | Bullying or harassment | 1 = No, 2 = Yes |
| | Adulthood experiences | |
| | Dropping out due to financial reasons | 1 = No, 2 = Yes |
| | Career or employment pressure | 1 = No, 2 = Yes |
| | Failure to enter college at desired time | 1 = No, 2 = Yes |
| | Failure to enter desired college | 1 = No, 2 = Yes |
| | Failure to obtain employment at desired time | 1 = No, 2 = Yes |
| | Failure to enter desired workplace | 1 = No, 2 = Yes |
| | Betrayal or fraud by a trusted person | 1 = No, 2 = Yes |
| | Separation or loss of close relationships (e.g., estrangement, death) | 1 = No, 2 = Yes |