

스마트폰 과의존 판별을 위한 기계 학습 기법의 응용*

김우성

건국대학교 경영학과 교수

한준희

부산대학교 산업공학과 교수

Application of Machine Learning Techniques for Problematic Smartphone Use

Woo-sung Kim^a, Jun-hee Han^b

^a School of Business, Konkuk University, South Korea

^b Department of Industrial Engineering, Pusan National University, South Korea

Received 26 June 2022, Revised 22 August 2022, Accepted 28 September 2022

Abstract

Purpose - The purpose of this study is to explore the possibility of predicting the degree of smartphone overdependence based on mobile phone usage patterns.

Design/methodology/approach - In this study, a survey conducted by Korea Internet and Security Agency(KISA) called “problematic smartphone use survey” was analyzed. The survey consists of 180 questions, and data were collected from 29,712 participants. Based on the data on the smartphone usage pattern obtained through the questionnaire, the smartphone addiction level was predicted using machine learning techniques. k-NN, gradient boosting, XGBoost, CatBoost, AdaBoost and random forest algorithms were employed.

Findings - First, while various factors together influence the smartphone overdependence level, the results show that all machine learning techniques perform well to predict the smartphone overdependence level. Especially, we focus on the features which can be obtained from the smartphone log data (without psychological factors). It means that our results can be a basis for diagnostic programs to detect problematic smartphone use. Second, the results show that information on users’ age, marriage and smartphone usage patterns can be used as predictors to determine whether users are addicted to smartphones. Other demographic characteristics such as sex or region did not appear to significantly affect smartphone overdependence levels.

Research implications or Originality - While there are some studies that predict smartphone overdependence level using machine learning techniques, but the studies only present algorithm performance based on survey data. In this study, based on the information gain measure, questions that have more influence on the smartphone overdependence level are presented, and the performance of algorithms according to the questions is compared. Through the results of this study, it is shown that smartphone overdependence level can be predicted with less information if questions about smartphone use are given appropriately.

Keywords: Smartphone Overdependence , Problematic smartphone use, Machine learning, Predictor.

JEL Classifications: C19, C49, I18

*이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음

^a First Author, E-mail: kimws@konkuk.ac.kr

^b Corresponding Author, E-mail: junhan@pusan.ac.kr

© 2022 The Institute of Management and Economy Research, All rights reserved.

I. 서론

정보통신 기술의 발달에 따라 스마트폰이 제공하는 편의가 다양해지면서, 스마트폰 이용자 수가 급격히 늘어나게 되었다. 이전의 휴대전화는 커뮤니케이션 기능만을 제공한 것에 비하여 스마트폰은 커뮤니케이션 뿐 아니라, 인터넷 접속을 통한 쇼핑, 게임, 음악 감상 등 다양한 여가 활동의 편의를 제공하고 있으며, 교육 콘텐츠를 활용한 학습 기능을 제공하기도 한다. 이러한 편의들로 인하여 스마트폰 사용률은 급격하게 증가한 것으로 나타나고 있는데, 한국 갤럽의 보고서에 따르면, 스마트폰 사용률은 2012년 1월 53%에서 2021년 95%로 증가하였으며, 저연령일수록 더 사용률이 빠르게 증가한 것으로 나타났다(한국갤럽, 2021). 실제로 스마트폰이 제공하는 다양한 기능들은 스마트폰 보급 이전에 휴대용 컴퓨터, PDA, PMP, MP3 플레이어, 전자 사전 등의 기기들이 제공하였는데, 스마트폰 보급 이후 이러한 시장 수요의 대부분을 스마트폰이 대체함에 따라 해당 기기들의 시장은 축소되는 등 현재 스마트폰은 하나의 기기로 이전에 다양한 기기들이 제공했던 기능과 편의들을 제공하고 있다.

한편 이러한 다양한 편의를 제공함에 따라 스마트폰에 과도하게 의존하는 사람들의 수도 증가하고 있다. 한국지능정보사회진흥원과 과학기술정보통신부가 2021년 발표한 스마트폰 과의존 실태조사에서는, 스마트폰 과의존을 “과도한 스마트폰 이용으로 스마트폰에 대한 현저성이 증가하고, 이용 조절력이 감소하여 문제적 결과를 경험하는 상태”로 정의하고 있다. 여기에서 현저성(salience)은 개인의 삶에서 스마트폰 이용이 다른 행태보다 두드러지고 가장 중요한 활동이 되는 것을 의미하며, 이용 조절력 감소(self-control failure)는 이용자의 스마트폰 이용에 관한 주관적인 목표와 비교했을 때 실제 스마트폰 이용에 대한 자기 조절 능력이 떨어지는 것을 의미한다. 문제적 결과(serious consequences)는 스마트폰 이용으로 신체적, 심리적, 사회적인 부정적인 결과를 경험함에도 불구하고 스마트폰을 계속 이용하는 것을 의미한다. 설문은 이러한 세 가지 요인에 기반한 척도 점수에 따라 대상을 고위험군, 잠재적위험군, 일반사용자군으로 분류하고 있으며, 각각 고위험군을 스마트폰 사용에 대한 통제력을 상실하여 신체적인 건강과 대인관계에 있어서 문제가 심각하게 발생한 상태로, 잠재적위험군을 스마트폰 사용에 대한 조절력이 약화되어 대인관계나 일상생활에서 문제가 발생하기 시작한 단계로 보고 있다. 다양한 선행연구에서도 스마트폰 과의존이 불안, 우울 등의 심리적인 증세를 발생시킴과 함께 대인관계나 자기 통제력에 부정적인 영향을 미친다고 밝히고 있으며, 또한 하나의 기기에서 다양한 편의를 제공하는 스마트폰의 특성상 SNS중독, 게임 중독, 쇼핑 중독 등 다른 중독으로 이어질 위험 또한 높아진다고 언급하고 있다(김병년, 2013; 김병년, 고은정, 최홍일, 2013; 황경혜, 유양숙, 조옥희, 2013). 2021년 기준으로 우리나라 만 3세~69세 스마트폰 이용자 중 24.2%는 스마트폰 과의존 위험군에 해당되는 것으로 나타났는데, 이는 2017년에 18.6%였던 것에 비해 증가한 수치이다(한국지능정보사회진흥원, 2021). 이 수치는 전 연령에서 증가하였는데 스마트폰 과의존 위험군 비율이 2019년 대비 60대 이상의 경우 2.6% 증가한 것에 비하여 성인은 4.5%, 청소년의 경우에는 6.8%, 유아동은 5.5% 증가한 것으로 나타나면서 특히 청소년과 유아동의 위험군 비율의 증가세가 두드러지는 것을 나타냈다.

이러한 스마트폰 과의존을 예방하기 위하여 과의존 정도를 진단하고, 측정하기 위한 다양한 척도들에 관한 연구들이 수행되었다(강희양, 박창호, 2012; 한국지능정보사회진흥원, 2021). 이러한 설문들의 경우 기존의 중독에 관련된 설문들을 기반으로 개발되거나, 스마트폰 과의존의 개념을 정의함으로써 개발되며, 이용자의 심리적인 문제(우울, 불안 등등)나 행동에 기반하고 있다. 예를 들어, 강희양, 박창호(2012)에서는 중독에 관한 기존의 설문들을 기반으로 스마트폰이라는 기기에 맞게 설문 문항을 개발하였으며 앞서 언급한 한국지능정보사회진흥원 설문의 경우에 스마트폰 과의존 요인들을 조절실패, 현저성, 문제적 결과로 정의하고 있다. 이러한 설문들은 스마트폰 이용에 따른 부작용이나 행동 양식에 관한 유용한 정보를 제공하지만, 어떤 이용자의 상태를 정확하게 진단하기 위해서는 사용자가 설문에 참여하고 진실되게 응답해야만 한다는 한계가 있다. 특히, 진단을 위한 설문들은 스마트폰 사용에 따른 심리적인 문제들이나 부작용들에 초점을 맞추고 있기 때문에, 스스로 민감하게 생각하여 진단을 거부하거나 솔직하게 응답하지 않는 경우에 통계적인 자료에 편향을 일으킬 수 있다. 선행 연구에서도, 조사 통계의 이론적인 측면은 무응답이 없다는 것을 전제로 하고 있기 때문에, 무응답이나 거부가 있는 경우에 응답한 사람들을 대상으로 측정된 조사 추정치는 편향된 표본에 의한 것일 수 있다고 언급하고 있다(김서영, 권순필, 2019).

이에 본 연구의 목적은 사용자의 심리에 기반한 설문이 아닌, 스마트폰 로그 데이터로 수집될 수 있는 객관적인 데이터들을 예측 변수로 사용하여 사용자의 과의존 여부를 예측하는 모형을 제시하는 것이다. 기존 스마트폰 과의존 관련 설문들은 사용자의 심리적인 요인들과 부작용에 초점을 두고 설계되었기 때문에 인구통계학적 특성이나 사용자의 스마트폰 사용 특성들에 대한 정보를 제공하기 어려운 측면이 있는데, 예를 들어, 어떠한 스마트폰 콘텐츠가 과의존과 더욱 밀접하게 연관되어 있는지, 또한 어떤 인구 특성적 요인이 과의존과 관련성이 있는지 등에 관한 정보를 도출하기는 어렵다. 또한, 위에서 언급했듯, 진단의 대상자가 설문에 답변하기를 거부하거나 솔직하게 답변하지 않는 경우, 과의존 정도에 관한 정보를 얻기 어렵다는 단점이 있다. 스마트폰 로그 데이터와 인구통계학적 특성들을 사용하여 스마트폰 과의존 진단 예측 모형들을 구축하는 경우, 스마트폰에 저장된 데이터를 기반으로 앱이나 프로그램들을 통하여 손쉽게 과의존 진단 예측을 할 수 있기 때문에, 다양한 진단이나 예방 프로그램 개발에 유용한 정보를 제공할 수 있을 것이다. 분석을 위하여 한국지능정보사회진흥원에서 조사한 스마트폰 과의존 실태조사 설문 데이터를 사용하였으며, k-NN (k-nearest neighbor), 그레디언트 부스팅(Gradient boosting), XGBoost, CatBoost, AdaBoost, 랜덤 포레스트와 같은 다양한 기계 학습 알고리즘을 사용하였다.

본 연구의 기여도 및 기존 연구들과의 차별성은 다음과 같다. 첫 번째, 본 연구에서는 기계학습 기법들을 사용하여 인구통계학적 특성과 스마트폰 로그 데이터로 수집될 수 있는 데이터들로 스마트폰 과의존 정도를 어느 정도 예측할 수 있음을 보인다. 스마트폰 과의존과 이에 영향을 미치는 요인들 간에 내재된 복잡성으로 인하여 기존의 몇몇 연구들에서 기계 학습 기법을 사용하여 과의존 예측 모형을 개발하였다. 하지만, 이철현(2020)과 Kim and Lee (2020)에서 개발한 예측 모형에서는 대인 관계, 학업 열의, 주의 집중 등 심리적인 요인들에 대한 변수들이 예측 변수로 사용하였기 때문에 결국 예측을 위해서는 설문을 통한 데이터 수집이 필수적이다. Lee and Kim (2021)에서 스마트폰 로그 데이터만을 사용하여 과의존군 예측 모형을 개발하였으나, 나이대별로 모형을 따로 학습하여 적용하였으며, 본 연구에 비하여 낮은 분류 결과를 보인다는 한계점이 있다. 본 연구에서는 스마트폰 사용량이나 자주 사용하는 콘텐츠 등 비교적 객관적인 데이터들을 예측 변수로 사용하여 상대적으로 높은 분류 결과를 제공하는 모형을 제시하기 때문에 다양한 진단 프로그램 개발에 유용한 정보를 제공할 수 있을 것으로 생각된다. 둘째, 본 연구는 스마트폰 과의존을 예측하기 위한 적절한 예측 변수들에 관한 정보를 제공한다. 인구통계학적 특성 및 스마트폰 사용 특성들로 과의존 예측 모형을 구축하여 이를 앱이나 프로그램으로 구현하는 경우 가장 중요한 문제 중 하나는 로그 데이터에서 어떤 데이터들을 추출하여 예측 변수로 사용해야 하는지에 관한 것이다. 각 개인의 스마트폰에는 수많은 데이터가 존재하기 때문에 이 중 적절한 변수를 선별하여 사용하는 것이 중요하다. 예측 모형이 과도하게 많은 예측 변수들로 구성되는 경우, 이는 계산 복잡도를 증가시켜 비용을 증가시키며, 동시에 모형 학습에 더 많은 표본을 요구하게 된다. 반대로 적은 수의 예측 변수들로는 충분한 예측력을 보장하기 어렵다. Lee and Kim (2021)에 따르면 스마트폰 과의존 위험군과 그렇지 않은 집단들은 자주 사용하는 콘텐츠, 월 평균 지출 등 다양한 스마트폰 패턴에 있어서 서로 다른 행태를 보이지만, 단일한 변수로는 과의존 여부를 판단하기 어렵다고 언급하고 있다. 다시 말하면 스마트폰 과의존 척도는 심리적인 요소에 의한 진단이기 때문에 단순히 스마트폰 사용 횟수가 많거나 특정 콘텐츠를 자주 사용한다는 단일한 변수로는 사용자가 스마트폰 과의존 상태라고 판단하기는 어렵다는 것이다. 이는 스마트폰 과의존과 이에 영향을 미치는 요인들 간에 복잡성이 내재되어 있음을 암시한다. 따라서 실무적으로 과의존 예측을 위한 앱이나 프로그램 개발을 고려할 때, 적절한(최대한 적은) 변수들을 선별하여 높은 예측력을 보장하는 효과적인 예측 모형을 개발하는 것이 중요하며, 이에 본 연구에서는 효과적인 예측 모형 구축에 필요한 사용 패턴에 관한 변수들과 인구통계학적 요인들에 관한 정보를 제공한다. 몇몇 선행 연구에서 스마트폰 과의존 여부 예측을 위하여 기계 학습 기법들을 사용하였으나, 적절한 예측 변수 수에 따른 모형의 예측력을 분석한 연구는 진행된 바 없다. 따라서 본 연구에서는 주요 예측 변수들을 선정하기 위하여 정보획득량(information gain) 척도를 사용하여, 스마트폰 과의존 진단에 가장 사용하기 좋은 질문들을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련된 선행 연구들을 정리한다. 스마트폰 과의존과 관련된 선행 연구들과 함께 스마트폰 과의존 예측에 기계학습 기법이 사용된 연구들을 소개한다. 3장에서는 본 연구에서 사용되는 설문과 기계학습 기법에 관하여 설명한다. 설문 문항 구성과 함께 주요 변수들에 관한

기술통계가 제시된다. 4장에서는 분석 결과를 제시하며, 과의존 여부 판단에 주요한 요인들과 학습 모형의 예측력에 관한 결과가 제시된다. 5장에서는 본 연구의 결론과 의의를 정리한다.

II. 문헌 조사

본 장에서는 연구와 관련된 선행 논문들을 정리한다. 과도한 스마트폰 사용에 관한 문헌들과 함께 이를 예측하기 위한 기계 학습 기법의 응용 사례들을 소개한다.

과도한 스마트폰 사용의 부작용에 관한 다양한 선행 연구에도 불구하고, 아직까지 과도한 스마트폰 사용을 중독으로 볼 수 있는지에 관한 합의에는 이르지 못한 실정이다(이성수, 2022). 실제로 해외의 몇몇 문헌에서는 과도한 스마트폰 사용은 기존의 다른 중독과는 차이를 보이기 때문에 이를 중독이라고 정의할 수 없다고 언급하고 있으며, 이에 다른 중독과 구별하기 위하여 과의존(overdependence), 과다 사용(overuse) 혹은 문제를 일으키는 사용(problematic use)이라는 용어 등을 사용하고 있다(Panova and Carbonell, 2018). 스마트폰은 물리적인 매체이기 때문에 이에 중독되는 것은 마치 술이나 마약과 같은 다른 물리적인 매체에 중독되는 것과 유사한 특성이 나타나야 한다는 것이다. 문헌에서는 스마트폰 기기가 인터넷, 게임, 성인물, 쇼핑 등으로의 접근을 용이하게 하기 때문에, 게임이나 인터넷 중독과 같은 다른 중독을 일으키는데 영향을 미치지만, 이는 다른 중독으로 구분해야 하며 스마트폰 중독으로 보면 안된다고 언급하고 있다(Panova and Carbonell, 2018). 실제로 과학기술부와 한국지능정보사회진흥원이 매년 발간하는 스마트폰 과의존 실태조사 보고서에서도 중독이 아닌 과의존이라는 표현을 사용하고 있으며, 이에 본 연구에서도 과의존이라는 용어를 사용한다. 그럼에도 상당한 연구들에서 스마트폰의 과도한 사용을 중독이라고 정의하고 있는데, 이 연구들에서는 중독의 정의인 '특정한 기호, 습관 또는 행동에 빠지거나 자신을 내맡기는 상태'에 스마트폰 중독이 해당된다고 주장한다(강희양, 박창호, 2012). 그 밖에 박용민(2011)은 스마트폰 중독을 스마트폰에 지나치게 몰입한 나머지 스스로를 제어할 수 없는 상태에 이른 것이라고 정의하였으며, 강희양, 박창호(2012)에서는 스마트폰의 과도한 몰입으로 인하여 생기는 초조, 불안과 같은 일상생활에서의 장애로 정의하였다.

스마트폰의 과도한 사용을 중독으로 볼 수 있으나 여부와 관계없이, 많은 연구에서 스마트폰 과용의 부작용에 대하여 보고하고 있다. 스마트폰의 과다 사용은 수면장애를 유발하여 수면의 질을 저하시키며(Matar Boumosleh and Jaalouk, 2017), 시력 저하, 거북목증후군, 경직, 엄지손가락과 손목무력증 등 신체적인 부작용을 초래할 수 있다(Mohammadbeigi et al., 2016; AlAbdulwahab, Kachanathu and AlMotairi, 2017). 또한, 다른 중독과 마찬가지로 스마트폰 과의존 위험군의 경우 스마트폰을 사용하지 못하게 되는 상황에서 불안감을 느끼게 되는 것으로 보고되었으며, 게임, 음란물 등으로의 손쉬운 접근으로 인하여 충동성, 과잉 행동, 망상, 환각 등 심리적 질병이 관찰되기도 하였다(Peraman and Parasuraman, 2016). 강혜자(2016), 정구철(2016)은 스마트폰의 과용이 대인기피증, 불안, 우울, 분노 등을 일으킬 수 있다고 언급하였다. 그 밖에 Samaha and Hawi(2016)에서는 스마트폰 과용의 부작용으로 인지능력과 기억 능력의 저하가 나타났으며, 특히 학생들의 경우 이는 주의력 저하, 학업 지연을 발생시켜 학업성취에도 부정적인 영향을 끼친다고 보고하였다.

이러한 부작용들을 예방하기 위하여, 스마트폰 과의존 여부를 측정하기 위한 다양한 척도들이 개발되었다. Bianchi and Phillips(2005)는 약물 남용에 관한 설문에 기반하여 스마트폰 과의존을 측정 설문을 개발하였다. 설문은 내성, 통제 불가능, 금단, 갈망, 부정적 영향 등 다섯 요인에 관한 것으로 구성되어 있으며, 스마트폰 사용 문제를 파악하는 데 중점을 두고 있다. Leung(2008)은 갈망, 감정변화, 금단, 행위 상실 등의 네 가지 요인으로 구성된 척도를 개발하였으나, 각 요인 별로 문항 수가 불균형하다는 한계를 보였다. 앞선 연구들이 성인을 대상으로 한 설문에 초점을 둔 데 비하여, Choliz(2012)는 청소년을 대상으로 한 TMD척도를 개발하였으며, 이는 금단, 통제 불가능, 내성을 요인으로 하고 있다. 국내에서는 Kwon et al.(2013a)과 Kwon et al.(2013b)에서 인터넷 중독에 관한 설문에 기반을 두고 청소년들을 대상으로 한 척도를 개발하였으며, 요인은 일상생활 장애, 긍정적 기대, 금단, 가상 인간관계 지향성, 과도한 사용, 내성으로 정의하여 사용하였다. 그 밖에 강희양, 박창호(2012)는 기존 중독 척도와 휴대문

중독 척도 등을 조합하여 설문을 개발하였으며, 요인분석을 통하여 몰입, 생활문제, 일상성, 과용, 관계성 등의 요인을 도출하였다. 이정정 et al.(2021)은 기존 스마트폰 중독 척도와 인터넷 중독 척도를 참고하여 일상생활 장애, 학업 장애, 금단, 내성, 가상 인간관계 지향 등의 다섯 요인에 관한 척도를 개발하였다. 현재 매년 정부에서 발간하고 있는 스마트폰 과의존 실태조사 보고서(한국지능정보사회진흥원, 2021)에서는 조절실패, 현저성, 문제적 결과의 세 가지 요인으로 구성된 설문을 개발, 사용하고 있으며, 유아동의 경우에는 관찰자가 작성하는 설문으로 구분하여 사용하고 있다.

이러한 척도는 설문자의 스마트폰 과의존 여부를 파악하고 부작용을 측정하는데 유용한 정보를 제공하지만, 사용자의 심리적인 부분과 일상생활에서의 부작용을 진단하는 것에 초점을 두고 있기 때문에 스마트폰 사용 패턴과 같은 다른 요인들과의 연관성을 분석하는 것에는 한계가 있다. 예를 들어, 설문이 아닌 스마트폰 사용 패턴만으로도 사용자의 과의존 여부를 판별할 수 있는가? 성별이나 나이와 같은 인구통계학적 특성은 과의존 여부를 판별하는데 유용한 정보인가? 이와 같은 질문의 답을 얻기에는 기존의 설문들로는 한계가 있다. 또한 위에서 언급했듯 설문지 질문들은 심리적인 부분이나 일상생활에서의 부작용에 관한 질문들이 대부분이기 때문에, 응답자가 답변을 거부하거나, 솔직하게 답변하지 않는다면 부정확한 결과를 얻을 수도 있다. 이에 Lee and Kim (2021)에서는 스마트폰 사용 패턴에 관한 데이터를 기반으로 스마트폰 과의존 위험군 여부를 예측하였다. 문헌에 의하면, 위험군과 그렇지 않은 집단 간에 개별적인 스마트폰 사용 패턴에 관한 몇몇 변수에서 유의한 차이를 보였음에도 불구하고, 단일한 변수로 그 사용자가 과의존 위험군인지 아닌지를 판별하기는 어려운 것으로 나타났다. 예를 들어, 스마트폰 사용 시간과 같은 단일한 변수로 어떤 사용자가 과의존 위험군인지 아닌지 판별할 수 없다는 것이다. Lee and Kim (2021)에서는 단순히 몇몇 변수들로 판단하기 어려운 복잡성이 스마트폰 과의존 위험군 판별 문제에 내재되어 있다고 언급하고 있는데, 연구에서는 설문의 스마트폰 사용 패턴에 관한 변수들만을 선별하여, 기계 학습 기법으로 과의존 위험군 여부를 판별하였다. 그 밖에도 이철현(2020), Kim and Lee (2020) 등에서 기계 학습 기법을 사용하여 스마트폰 과의존 위험군 판별 문제를 분석하였다.

이러한 선행 연구들은 유용한 정보를 제공하고 있으나, 다음에 있어서 한계점을 보인다. 선행 연구들에서는 한국지능정보사회진흥원에서 진행하는 스마트폰 과의존 실태조사 설문 데이터를 분석에 이용하였다. 실제 다수의 스마트폰 사용 데이터들을 얻기는 각 사용자의 스마트폰에서 취득하기는 어렵기 때문에, 어찌 보면 이는 당연하다. 하지만 이철현(2020), Kim and Lee(2020)에서는 스마트폰 로그 데이터로부터 취득할 수 있는 데이터나 기초적인 인구통계학적 특성이 아닌, 사용자의 심리적인 요소들이나 대인 관계, 학업에 대한 요인까지 포함하여 예측을 수행하였다. 실제로 연구들에서 알고리즘들은 어느 정도 우수한 예측력을 보였으나, 실제로 과의존 예측 모형을 프로그램이나 어플리케이션으로 구현한다고 가정했을 때, 결국 사용자는 관련된 설문에 응답해야 한다는 한계를 지닌다. Lee and Kim (2021)에서 스마트폰 로그 데이터만을 사용하여 과의존군 예측 문제를 다루었으나, 어떤 질문(요인)들을 예측을 위하여 사용하는 것이 좋을지에 관한 결과를 제시하지 못하였다. 또한, 예측력을 보장하기 위하여 기존 실태조사에서는 세 그룹(일반사용자군, 잠재적위험군, 고위험군)으로 나누어져 있던 것을 두 그룹(고위험군과 이외)으로 합쳐서 분석한 것 또한 한계점이 될 것이다. 실태조사 보고서에서는 잠재적 위험군 또한, 과의존 요인 3가지 중 1~2 요인의 특성을 보이기 때문에 스마트폰 사용에 있어 주의를 요한다고 언급하고 있다. 이에, 본 연구에서는 스마트폰으로부터 수집할 수 있는 데이터들을 기계 학습 기법을 통하여 분석하여 스마트폰 과의존 수준을 판별하는 문제를 다룬다. 또한 과의존 수준 판별에 영향을 미치는 인구통계학적 요인과 스마트폰 사용과 관련된 변수들을 제시한다.

III. 데이터 및 연구방법론

1. 데이터

본 장에서는 연구에서 분석되는 데이터를 소개하고 기술통계를 제시한다. 본 연구에서 사용되는 데이터는 2017년에 과학기술정보통신부와 한국정보화진흥원에서 함께 조사한 스마트폰 과의존 실태조사 설문

자료이다. 설문 조사는 2017년 9월부터 11월까지 전국에서 시행되었으며, 가구 방문 면접 조사를 통하여 표본 10,000가구에 관한 설문 데이터를 수집하였다. 가구 내 만 3세에서 69세 사이의 스마트폰 이용자 29,712명이 설문에 응답하였으며, 응답자 중 유아동(만3~9세)은 2,651명, 청소년(만10~19세)은 5,144명, 성인(만20~59세)은 19,712명, 60대는 2,205명으로 나타났다. 각 시도별 표본은 각 시도의 가구 수 제곱근에 비례하여 할당, 조사하였으며, 각 시도의 조사구의 표본 또한 조사구의 연령, 성별 등 인구통계학적 특성을 비례하여 할당하여 조사하였다. 구체적인 표본 배분 및 자료 수집 방식, 그리고 데이터에 관한 다양한 기술통계에 관한 결과는 한국지능정보사회진흥원(2017)에 제시되어 있으며, 사용자의 유형별로 스마트폰 사용 관련 변수들의 값에 차이가 있는지에 대한 가설 검정 결과는 Lee and Kim(2021)에 제시되어 있다.

설문 자료는 스마트폰 이용현황, 스마트폰 과의존 척도, 게임 이용현황, 스마트폰 이용 관련 정책 인지도, 스마트폰 이용 관련 심리적 특성, 인구 통계 특성 등을 포함하여 약 180문항으로 이루어져 있다. 이 중 과의존 척도가 약 10문항(유아동의 경우 9문항)으로 구성되어 있는데, 설문에서는 스마트폰 과의존을 “과도한 스마트폰 이용으로 스마트폰에 대한 현저성이 증가하고 이용 조절력이 감소하여 문제적 결과를 경험하는 상태”로 정의하고 있으며, 조절 실패, 현저성, 문제적 결과를 스마트폰 과의존의 3요인으로 정의하고 있다. 척도는 각 요인에 관한 3~4문항으로 총 9~10문항으로 각 문항별 4점 만점 척도로 구성되어 있으며, 유아동 스마트폰 과의존 척도의 경우 관찰자(보호자)가 응답하게 되어 있다. 유아동 스마트폰 과의존 척도와 그 외(청소년, 성인, 고령층)에 관한 스마트폰 과의존 척도 설문이 각각 <Table 1>과 <Table 2>에 제시되어 있다.

Table 1. 유아동 스마트폰 과의존 척도(관찰자 응답)

요인	항목
조절실패 (역문항)	1) 스마트폰 이용에 대한 부모의 지도를 잘 따른다
	2) 정해진 이용 시간에 맞춰 스마트폰 이용을 잘 마무리한다.
	3) 이용중인 스마트폰을 빼앗지 않아도 스스로 그만둔다.
	4) 항상 스마트폰을 가지고 놓고 싶어한다.
현저성	5) 다른 어떤 것보다 스마트폰을 갖고 노는 것을 좋아한다.
	6) 하루에도 수시로 스마트폰을 이용하려 한다.
	7) 스마트폰 이용 때문에 아이와 자주 싸운다.
문제적 결과	8) 스마트폰을 하느라 다른 놀이나 학습에 지장이 있다.
	9) 스마트폰 이용으로 인해 시력이나 자세가 안 좋아진다.

Source: 한국지능정보사회진흥원(2017).

Table 2. 청소년, 성인, 고령층 스마트폰 과의존 척도

요인	항목
조절실패	1) 스마트폰 이용 시간을 줄이려 할 때마다 실패한다.
	2) 스마트폰 이용시간을 조절 하는 것이 어렵다.
	3) 적절한 스마트폰 이용 시간을 지키는 것이 어렵다.
	4) 스마트폰이 옆에 있으면 다른 일에 집중하기 어렵다.
현저성	5) 스마트폰 생각이 머리에서 떠나지 않는다.
	6) 스마트를 이용하고 싶은 충동을 강하게 느낀다.
	7) 스마트폰 이용 때문에 건강에 문제가 생긴 적이 있다.
문제적 결과	8) 스마트폰 이용 때문에 가족과 심하게 다툰 적이 있다.
	9) 스마트폰 이용 때문에 친구, 동료, 사회적 관계에서 심한 갈등을 경험한 적이 있다.
	10) 스마트폰 때문에 업무(학업 혹은 직업 등) 수행에 어려움이 있다.

Source: 한국지능정보사회진흥원(2017).

각 문항 점수에 대한 점수 합을 산출 후에, 총점에 따라 스마트폰 과의존 수준이 고위험군, 잠재적 위험군, 일반사용자군의 3개의 유형으로 분류한다. 대상별로 과의존 유형을 구분하는 기준 점수는 상이한

데, 예를 들어 유아동의 경우 28점 이상이 고위험군으로 분류되지만, 청소년은 31점, 성인은 29점, 60대는 28점 이상일 때 고위험군으로 분류된다. 고위험군은 현저성, 조절실패, 문제적 결과 세 가지 요인의 특성을 모두 보이는 집단을 의미하며, 잠재적 위험군은 세 요인 중 1, 2가지 특성을 보이는 집단을 의미한다. 표본 전체 중 고위험군, 잠재적 위험군, 일반사용자군은 각각 2.7%, 15.9%, 81.4%로 나타났다. 본 연구의 목표가 인구통계학적 특성과 스마트폰 사용 패턴에 관한 정보를 사용하여 스마트폰 과의존 수준을 예측하는 것임을 참고하자. 따라서, 스마트폰 과의존 수준(고위험군, 잠재적 위험군, 일반사용자군)은 우리가 예측하고자 하는 종속변수(결과 변수)가 된다.

또한, 총 문항 180문항 가운데, 인구통계학적 특성과 스마트폰 사용에 관련된 변수들을 예측변수로 선정하였다. <Table 3>에는 설문 중 예측 변수로 선정할 변수들이 정리되어 있다. 어떠한 인구통계학적 특성들이 스마트폰 과의존 여부에 영향을 미치는지 알아보기 위하여 설문의 인구통계학적 특성에 해당하는 부분들을 사용하였으며, 설문에서 정서적이나 심리적인 부분에 관한 문항들이나 중독 상담 서비스에 대한 인지에 관한 문항들은 제외하고 스마트폰 사용에 대한 설문 데이터만을 사용하였다. Lee and Kim(2021)에서는 스마트폰 사용자 유형별로 자주 사용하는 스마트폰 사용 콘텐츠가 차이가 있는지에 관한 가설 검정을 수행하였다. 고위험군과 그렇지 않은 사용자 그룹 간에 게임, 영화, 소셜 및 만화 등 콘텐츠 관련 월 평균 요금과 각 콘텐츠의 사용 빈도 등 다양한 변수들에 있어서 유의한 차이를 보였으나, 연구에서는 단일한 문항들로는 스마트폰 과의존 수준을 예측하기 어렵다고 밝히고 있다. 스마트폰 과의존 수준의

Table 3. 예측 변수

	문항	설명
인구통계학적 특성	거주지역	응답자가 거주하는 시, 도
	주택형태	단독주택, 아파트, 연립주택/다세대 주택 거주 여부
	가구주 성별	가구주의 성별
	가구주 나이	가구주의 나이
	가구주와의 관계	배우자, 유아동자녀, 청소년 자녀, 미혼자녀, 기혼자녀, 부모, 조부모 등등의 가구주와의 관계
	본인 성별	성별
	본인 나이	나이
	총 가구원 수	총 가구원 수
	다문화 가족 여부	다문화 가족 여부 (이진 변수)
	한부모가족 여부	한부모 가족 여부 (이진 변수)
	월평균 가구 소득	소득 구간에 따라 6등급으로 나눔
	학력 및 재학 여부 (2문항)	최종 학력 및 재학 여부
	장애인 여부	이진 변수
	북한 이탈 주민 여부	이진 변수
스마트폰 사용 관련 변수	스마트폰 요금	월 평균 스마트폰 이용 요금
	콘텐츠별 이용 여부 (19문항)	이진 변수, 이용하지 않으면 0, 이용하면 1. 콘텐츠별로 각각 19문항으로, 콘텐츠는 뉴스, 학업(업무), 교육, 상품/서비스 정보, 교통 및 위치 정보, 일반적인 웹서핑, 게임, 성인용 콘텐츠, 영상(영화, TV, 동영상), 음악, 전자책(e-book, 웹툰, 웹소설), 스포츠 배팅, 이메일, 메신저, SNS, 상품 구매, 상품 판매, 금융(스마트폰 뱅킹, 주식 거래), 생활 관리(가계부, 일기 등)로 구성됨
	콘텐츠별 이용 정도 (19문항)	콘텐츠별로 각각 19문항으로, 콘텐츠는 위와 동일하지만 사용 정도를 측정하기 위하여 7점 척도를 사용함
	월 평균 결제 비용 (3문항)	게임, 영상(영화, TV, 동영상), 전자책(웹툰, 웹소설) 콘텐츠에 대하여 월 평균 과금 비용
	이용 횟수 및 시간 (4문항)	각각 주중, 주말의 이용 횟수(회) 및 시간(분)
	즐거 이용하는 게임 종류(2문항)	즐거 이용하는 게임 종류 1, 2순위 (예를 들어, 액션게임, 스포츠 게임 등등)

Source: 한국지능정보사회진흥원(2017).

판단 기준은 사용자의 심리적인 요인이나 일상생활에서 발생하는 문제들에 기반하기 때문에, 스마트폰 사용 패턴이나 인구통계학적 특성에 관한 하나의 변수로 과의존 여부를 예측하기 어려운 것이다. 예를 들어, 단순히 스마트폰 사용 시간이 길거나, 월 평균 요금이 높다는 이유 하나로 어떤 사용자가 스마트폰 중독인지 판단할 수 없다. 선행 연구에서도 과의존 위험군과 일반사용자 간에는 스마트폰 사용 패턴에 관한 몇몇 변수들이 유의한 차이를 보임에도 불구하고, 과의존 수준을 판별하기 위해서는 여러 변수들을 복합적으로 고려해야 한다고 언급한다. 이러한 복잡성을 고려하여 스마트폰 사용 패턴과 인구통계학적 특성으로 스마트폰 과의존 수준을 예측하기 위하여, 본 연구에서는 <Table 3>에서 정리한 예측 변수들로 기계 학습 기법을 사용하여 스마트폰 과의존 수준을 판별한다. 과의존 예측을 위하여 k-NN (k-nearest neighbor), 그레디언트 부스팅(Gradient boosting), XGBoost, CatBoost, AdaBoost, 랜덤 포레스트와 같은 다양한 알고리즘을 사용하였다.

2. 연구방법론

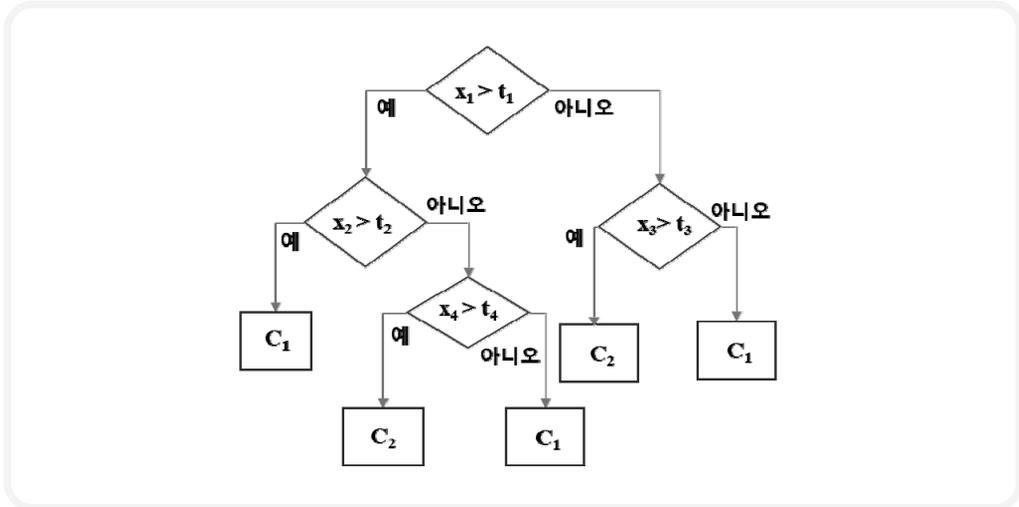
본 연구에서는 과의존 예측을 위하여 k-NN (k-nearest neighbor), 그레디언트 부스팅(Gradient boosting), XGBoost, CatBoost, AdaBoost, 랜덤 포레스트 알고리즘이 사용된다. 본 장에서는 각 알고리즘에 대한 간략한 설명을 제시한다.

k-NN, 또는 k-최근접 이웃 알고리즘은 어떤 데이터를 판별할 때 학습 데이터 중 그 데이터와 가장 가까운 k 개의 다른 데이터의 판별 결과를 참조하여 분류하는 알고리즘이다. 알고리즘에서는 어떤 사람이 과의존인지 아닌지를 판단하기 위하여 학습 데이터 중 가장 그 사람과 유사한 k 명의 사람들을 찾아낸다. 이때, 두 사람이 얼마나 유사한지를 측정하기 위하여 일반적으로 유클리디안 거리 척도가 사용된다. 본 연구에서는 임의의 두 사람 사이의 유사성을 측정하기 위하여 <Table 3>에 예측 변수들로 거리 척도를 계산한다. 이후 계산된 거리 척도들을 기반으로 가장 가까운, 다시 말하면 분류하고자 하는 사람과 가장 유클리디안 거리 척도 값이 작은 이웃 k명을 선정한다. 그리고 k명의 다수결 결과에 따라 해당 사람을 과의존 여부를 판별하게 된다. 결국, 어떤 응답자의 과의존 여부를 판단하기 위해서는, 첫 번째 단계로 그 응답자와 학습 데이터 집합에 속한 모든 사람들과의 거리 척도를 각각 계산한다. 두 번째로 거리 척도 값이 가장 작은 k명의 사람을 추출한다. 예를 들어서, k가 3인 경우, 응답자와 가장 유사한 3명을 학습 데이터에서 추출하게 된다. 세 번째로 k명의 분류 결과에 따라 다수결로 해당 응답자를 판별한다. 추출된 3명 중 2명이 일반 사용자군이고 한명이 고위험군이라면 응답자를 일반 사용자군으로 분류하는 것이다. k 값은 사용자가 결정하는 파라미터이며, 다수결을 위하여 홀수로 선정하는 것이 일반적이다. k-NN 알고리즘은 직관적인 거리 척도에 의하여 이웃들을 선정하고 다수결로 결정을 하기 때문에 특별히 모델들을 생성하지 않으며, 단순하면서도 효율이 높다는 장점이 있지만, 각 데이터를 분류할 때마다 응답자와 다른 모든 학습 데이터와의 거리 척도를 계산하기 때문에 데이터 크기가 커질수록 분류 단계가 느려진다는 단점을 갖는다.

본 연구에서 사용하는 나머지 알고리즘, 그레디언트 부스팅(Gradient boosting), XGBoost, CatBoost, AdaBoost, 랜덤 포레스트 알고리즘은 모두 의사결정 나무를 기반으로 하는 앙상블 학습 기법이다. 의사결정 나무는 의사결정 규칙들을 나무 구조로 나타내어 해당 데이터를 분류할 수 있도록 하는 기계 학습 기법이다. <Fig. 1>에 4개의 독립변수(x_1, x_2, x_3, x_4)를 사용하여 두 개의 클래스(C_1, C_2)로 분류하는 의사결정 나무 예시가 제시되어 있다. 의사결정나무에서는 데이터를 분류하기 위하여 상위노드부터 독립변수들의 값에 따라 가지를 따라 내려가게 된다. 예시에서 제시된 의사결정나무를 보면, 변수 x_1 의 값이 t_1 보다 크면, 왼쪽 가지를 따라 내려간다. 하위노드에서 변수 x_2 가 t_2 보다 크면, 해당 데이터를 클래스 C_1 으로 분류하게 된다. 만약 x_2 가 t_2 보다 작거나 같은 경우 그 하위노드에서 분류를 수행하게 되는데 x_1 값에 따라 클래스를 분류한다. 이와 같이, 의사결정나무에서는 나무 구조를 기반으로 상위노드부터 하위 노드로 매 단계 마다 의사결정 규칙들을 순차적으로 적용하며 분류 작업을 수행해 나간다. 나무 구조를 생성하게 될 때, 단계마다 분류 변수(분류하기 위해 사용되는 예측 변수)와 그 기준값을 정하는 것이 중요한데, 분류 변수와 기준값은 기준에 의하여 분할되는 두 학습 집합 간에는 이질성이

가장 커지도록, 집합 내에서는 동질성이 가장 커지도록 선택된다. 과적합을 피하기 위해, 나무 구조 크기는 여러 기준에 의하여 조절된다.

Fig. 1. 의사결정나무 예시



의사결정나무는 직관적으로 이해하기 쉬우며, 해석하기 용이한 방식으로 분류를 수행한다는 장점이 있지만, 학습 데이터 세트에 따라 생성되는 의사결정나무의 구조의 변동성이 크며 이에 학습 결과의 성능 또한 변동폭이 크다는 단점을 가지고 있다. 특히 의사 결정 나무는 계층적으로 분류 작업이 이루어지기 때문에 의사결정 나무 노드가 편향되게 형성되는 경우, 하위 노드에 계속 편향성이 전파되게 된다. 이에 실제로 분류 문제에는 여러 개의 의사결정나무를 생성하여 이를 결합하는 방식으로 분류를 수행하는 앙상블 기법들이 사용되고 있다. 다수의 의사결정나무를 결합하여 분류 작업을 수행하게 되면, 학습 데이터로 인하여 편향적인 의사결정나무가 생성되더라도, 이로 인하여 발생할 수 있는 오분류 가능성을 다른 의사결정나무 모형들과 통합하는 과정에서 줄일 수 있다.

다수의 의사결정 나무를 결합하여 앙상블 모형을 만드는 방식을 크게, 배깅(bagging)방식과 부스팅(boosting)방식으로 나눌 수 있는데, 랜덤 포레스트는 배깅 방식에 속하며, 나머지 그라디언트 부스팅(Gradient boosting), XGBoost, CatBoost, AdaBoost 기법은 부스팅 방식에 속한다. 배깅 방식은 bootstrap aggregation의 약자인데, 부스트래핑(bootstrapping)은 복원 추출을 사용한 표본 추출 방식을 의미한다. 배깅 방식에서는 복원 추출 방식을 통하여 다수의 학습 데이터 세트를 생성하고 각 학습 데이터들을 기반으로 의사결정 나무 모형을 학습시키게 된다. 이후 학습된 각각의 의사결정 나무의 결과들을 사용하여(다수결), 분류를 수행한다. 여러 개의 의사결정 나무를 얻은 후에, 분류해야 할 새로운 데이터를 각 트리에 동시에 통과시켜, 각 트리가 분류한 결과들을 놓고 다수결을 통하여 최종적으로 데이터를 분류하는 것이다. 특히, 랜덤 포레스트 기법에서는 다양한 의사결정나무를 생성시키기 위하여 학습 데이터 세트만 랜덤하게 추출하는 것이 아니라, 독립변수를 또한 일부만 선택하여 의사결정나무를 학습시키게 된다. 최종적으로 다수의 의사결정 나무들이 형성되게 되며 이는 다수의 나무로 이루어진 숲(forest)을 구성하게 된다.

부스팅(boosting) 방식은 다수의 분류기를 학습하여 생성하고 이를 결합하여 최종 분류를 한다는 점에서는 배깅 방식과 동일하지만, 동시에 학습을 진행하는 것이 아닌 순차적인 학습을 통하여 분류 모형을 개선해나간다는 점에서 차이가 있다. 배깅 방식에서는 각각 다른 분류 모형을 학습한 후 그 결과들을 취합하여 최종적인 분류를 하기 때문에 각각의 의사결정 나무들이 병렬적으로 학습과 분류 작업을 하게 되는데, 부스팅 방식에서는 한 분류 모형의 학습이 이루어지고 난 후, 이전의 분류 모형에서 발생한 오분류

들을 개선하는 방식으로 새로운 분류 모형을 얻게 된다. 학습된 의사 결정 나무의 분류 정확도를 관찰하여 오분류가 발생한 데이터에 가중치를 두어, 그 다음 학습되는 예측 모형의 성능이 개선되게끔 하는 것이다. 본 연구에서 사용하는 부스팅 방식의 알고리즘은 모두 순차적인 학습을 통하여 예측 모형을 개선해 나가지만, 개별 예측 모형의 약한 부분을 인식하는 방법과 그 가중치를 부여하는 방식에 있어서 약간 차이가 있다. Adaboost 기법이 잘못 분류된 데이터에 가중치를 부여하여 약한 학습기들을 순차적으로 학습시키고, 개별 학습기에 가중치를 부여하여 결합함으로써 개별 학습기보다 높은 정확도를 얻는데 비하여, 그레디언트 부스팅 알고리즘에서는 가중치를 업데이트할 때 경사하강법(gradient descent)를 이용한다. XGBoost 기법은 그레디언트 부스팅 방식에 기반하고 있지만, 알고리즘 효율성을 위한 최적화 방법을 통하여 기존의 그레디언트 부스팅 방식의 단점으로 꼽혔던 느린 학습 시간의 문제를 해결하였다. Catboost 방식에서는 비용함수(loss function)를 이용하여 잔차들을 계산하여 사용한다는 점과, 순열 기반의 대안인 순서형 부스팅(ordered boosting) 방법과 범주형 변수 처리를 위한 알고리즘에서 다른 알고리즘과 차이를 보인다. 각 알고리즘은 데이터 세트와 문제 상황에 따라 다른 성능을 보이게 되는데, 본 연구에서도 다양한 알고리즘들의 성능을 측정하고 비교한다.

IV. 분석 결과

다양한 알고리즘의 분류 성능을 비교하기 위하여 여러 성능척도에 관한 결과를 도출하였다. 과적합을 방지하고, 불균형한 종속변수 분포 문제를 해결하기 위하여 Stratified 5-fold 기법을 사용하였다. 교차 검증이란, 모델이 학습 데이터에만 과도하게 최적화되어 다른 데이터로 분류를 수행할 때 성능이 과도하게 떨어지는 과적합 문제(overfitting)를 해결하기 위하여 별도의 여러 세트로 구성된 학습, 검증 데이터 세트에서 학습과 평가를 수행하여 성능을 도출하는 것을 의미한다. 본 연구에서는 5-fold 교차 검증 기법을 사용하였는데, 이는 전체 데이터 집합을 5개의 세트로 분할하여, 각 데이터 세트를 검증 데이터로 사용하고 나머지 데이터 세트를 학습 데이터로 사용하여 5회 평가를 수행하는 것을 의미한다. Stratified 5-fold는 각 데이터 세트에 고위험군, 잠재적 위험군, 일반사용자군이 전체 데이터 세트의 분포와 유사하게 분배되도록 한다. 이는 전체 데이터 중 고위험군, 잠재적 위험군의 비율이 일반 사용자군에 비하여 상대적으로 적기 때문에, 편향된 표본을 얻지 않도록 하기 위함이다. <Table 3>의 변수들을 사용하여 과의존 수준(고위험군, 잠재적 위험군, 일반사용자군)을 분류한 결과가 <Table 4>에 제시되어 있다. 각 알고리즘의 개략적인 파라미터에 관한 정보는 <Table 5>에 제시되어 있다. 기계 학습 기법의 성능을 비교하기 위하여, 로지스틱 회귀 분석 또한 수행하여 결과를 측정하였다.

Table 4. 분류 성능 (전체 예측 변수 사용)

모형	정확도(accuracy)	F1 score	정밀도(precision)	재현율(recall)
Logistic Regression	0.735	0.758	0.763	0.818
k-NN	0.837	0.820	0.814	0.837
Random forest	0.872	0.856	0.861	0.872
Gradient boosting	0.853	0.825	0.838	0.853
XGBoost	0.885	0.875	0.876	0.885
CatBoost	0.868	0.853	0.855	0.868
AdaBoost	0.826	0.827	0.828	0.826

성능을 측정한 결과, 전체적으로 기계 학습 기법들이 로지스틱 회귀 기법보다 우수한 예측 성능을 발휘하고 있는 것으로 나타났다. 세부적으로 보면 XGBoost 알고리즘이 가장 뛰어난 성능을 보이는 것으로 나타났으며, 그 이후는 랜덤 포레스트, CatBoost 순으로 나타났다. 나머지 그레디언트 부스팅, AdaBoost, k-NN 알고리즘은 척도에 따라 다른 순위를 보였으나, 모든 알고리즘이 다양한 분류 척도에서 0.8 이상의 값으로 어느 정도 성공적인 분류 성능을 보여준다. 이는 선행연구인 Lee and Kim(2021)의 결과를 상회하

는 값이며, 인구통계학적인 특성들과 스마트폰 사용 패턴들이 과의존 수준을 예측하는 문제에 예측자로 사용될 수 있음을 의미한다.

Table 5. 알고리즘 파라미터

모형	파라미터(parameters)
<i>k</i> -NN	Number of neighbors: 5 Metric: Euclidean distance Weight: Uniform
Random forest	Number of trees: 20 Number of attributes considered at each split: 5 Limit depth of individual trees: 3 Stopping rule: Do not split subsets smaller than 5 instances
Gradient boosting	Number of trees: 100 Learning rates: 0.1 Limit depth of individual trees: 3
XGBoost	Number of trees: 100 Learning rates: 0.3 Lambda : 1 Limit depth of individual trees: 6
CatBoost	Number of trees: 100 Learning rates: 0.3 Lambda : 3 Limit depth of individual trees: 6
AdaBoost	Number of trees: 50 Learning rate: 1 Classification algorithm: SAMME.R Regression loss function: Linear

Table 6. 정보 이득에 기반하여 도출한 문항 순위(1~20위)

순위	문항
1	스마트폰 콘텐츠별 이용 정도: 성인용 콘텐츠 (7점 척도)
2	스마트폰 콘텐츠별 이용 여부: 게임 (0: 이용안함, 1: 이용함)
3	가구주와의 관계 (가구주, 배우자, 유아동자녀, 청소년자녀, 미혼자녀, 기혼자녀 및 배우자 등등)
4	스마트폰 주중 1일 환산 평균 이용 횟수
5	스마트폰 주말 1일 환산 평균 이용 횟수
6	스마트폰 콘텐츠별 이용 정도: 전자책, 웹툰, 웹소설 (7점 척도)
7	스마트폰 콘텐츠별 이용 정도: 교육, 학습 (7점 척도)
8	스마트폰 콘텐츠별 이용 정도: 스포츠 배팅 (7점 척도)
9	스마트폰 콘텐츠별 이용 정도: 게임 (7점 척도)
10	스마트폰 콘텐츠별 이용 여부: 전자책, 웹툰, 웹소설 (0:이용안함, 1:이용함)
11	즐거 이용하는 게임 1순위 (롤플레이, 웹보드게임, 전략시뮬레이션 등등)
12	사용자의 나이
13	스마트폰 주중 1일 환산 평균 이용 시간(분)
14	스마트폰 주말 1일 환산 평균 이용 시간(분)
15	스마트폰 콘텐츠별 이용 여부: 교육, 학습 (0:이용안함, 1: 이용함)
16	스마트폰 콘텐츠별 이용 정도: 음악 (7점 척도)
17	학생 여부 (재학 중 여부, 휴학 포함)
18	스마트폰 콘텐츠별 이용 여부: 상품, 서비스 정보 검색 (0:이용안함, 1: 이용함)
19	스마트폰 콘텐츠별 이용 정도: 영화, TV, 동영상 (7점 척도)
20	스마트폰 콘텐츠별 이용 정도: 뉴스 검색 (7점 척도)

성능에 영향을 미치는 주요 요인들을 도출하기 위하여 분할 이전의 엔트로피에서 이후 엔트로피 값의 차이인 정보 이득(information gain)에 기반을 두어 주요 독립 변수들을 선정하였다. 과의존 수준을 예측하는데 필요한 변수들의 주요 순위는 <Table 6>과 같다.

주요 요인들에 따라 성능이 얼마나 차이가 나는지를 알아보기 위하여, 주요 변수 10, 15, 20가지만을 사용하여 같은 방식(Stratified 5-fold)으로 분류 성능을 측정하였다. 다시 말하자면, 주요 변수 10가지를 사용하여 성능을 측정하는 경우, 정보 이득에 기반을 둔 문항 순위 1~10위까지의 문항들만을 사용하여 과의존 수준을 분류하였고, 15가지를 사용하는 경우 1~15위까지의 문항들을, 20가지를 사용하는 경우 1~20위까지의 문항들을 사용하여 분류 성능을 측정하였다. 각각의 결과는 <Table 7>, <Table 8>, <Table 9>에 제시하였다.

Table 7. 분류 성능 (10가지 예측 변수 사용)

모형	정확도(accuracy)	F1 score	정밀도(precision)	재현율(recall)
Logistic Regression	0.684	0.738	0.764	0.816
k-NN	0.825	0.807	0.800	0.825
Random forest	0.838	0.818	0.815	0.838
Gradient boosting	0.835	0.791	0.812	0.835
XGBoost	0.842	0.818	0.819	0.841
CatBoost	0.838	0.806	0.812	0.838
AdaBoost	0.825	0.810	0.803	0.825

Table 8. 분류 성능 (15가지 예측 변수 사용)

모형	정확도(accuracy)	F1 score	정밀도(precision)	재현율(recall)
Logistic Regression	0.699	0.742	0.753	0.816
k-NN	0.829	0.811	0.805	0.829
Random forest	0.852	0.832	0.835	0.850
Gradient boosting	0.841	0.804	0.820	0.841
XGBoost	0.860	0.843	0.846	0.860
CatBoost	0.849	0.823	0.829	0.847
AdaBoost	0.842	0.828	0.824	0.842

Table 9. 분류 성능 (20가지 예측 변수 사용)

모형	정확도(accuracy)	F1 score	정밀도(precision)	재현율(recall)
Logistic Regression	0.710	0.744	0.751	0.816
k-NN	0.836	0.820	0.815	0.836
Random forest	0.868	0.852	0.855	0.869
Gradient boosting	0.846	0.816	0.827	0.846
XGBoost	0.877	0.867	0.867	0.876
CatBoost	0.863	0.846	0.848	0.861
AdaBoost	0.863	0.850	0.849	0.860

결과들을 살펴보면, 여기에서도 로지스틱 회귀 기법보다 기계 학습 기법들이 우수한 예측 성능을 발휘하고 있는 것으로 나타났다. <Table 7>을 살펴보면, 10가지의 주요 변수로도 다양한 척도에서 약 80%

정도의 분류율(과의존 고위험군, 잠재적 위험군, 일반 사용자군)을 보이며, 이는 10가지 주요 변수들만으로도 과의존 수준을 어느 정도 예측할 수 있음을 의미한다. 전체 변수를 사용했을 때와 10가지의 예측 변수만을 사용하였을 때의 정확도를 비교해 보면, k -NN 알고리즘의 경우 1.2% 정확도 차이를 보이며, 랜덤 포레스트는 3.4%, 그레디언트 부스팅, XGBoost, CatBoost, AdaBoost는 각각 1.8%, 4.3%, 3%, 0.1%의 정확도 차이를 보인다. 예측 변수의 수가 감소하더라도 모든 알고리즘에서 80% 이상의 정확도를 보이는 등 비교적 좋은 예측 성능을 보이고 있는데, 이는 스마트폰 로그 데이터를 이용하여 과의존 진단/예방 프로그램을 개발할 때 적은 데이터 수집 비용으로도 효과적으로 과의존 여부를 예측할 수 있다는 것을 의미한다. 마찬가지로 전체 변수를 사용했을 때의 경우와 20가지 변수를 사용할 때의 정확도를 비교해 보면 모든 알고리즘에서 정확도 차이가 1% 미만으로 나타난다. 이는 20개의 변수 안에 스마트폰 과의존 여부를 예측하는데 필요한 정보 대부분이 포함되어 있음을 의미한다. 다시 말하면, 설문에서 제시된 스마트폰 사용과 인구통계학적 특성에 관한 모든 변수를 사용하여 예측 모형을 개발하는 경우, 10개의 변수를 사용하는 것보다 평균적으로 2.3% 정도의 정확도 개선을 기대할 수 있으며 20개 변수를 사용하는 것 보다는 1% 미만의 정확도 개선을 기대할 수 있다. 예측 변수의 수가 증가함에 따라 데이터 수집 및 예측 모형 학습과 관련된 계산 복잡도와 비용들이 증가하기 때문에, 10개의 변수를 사용한 이후부터는 예측 변수 추가에 따른 정확도 개선을 통하여 비용을 상회하는 효과를 얻을 수 있는지를 고려하여 적절한 변수의 수를 선정할 필요가 있다. 제시한 알고리즘들이 비슷한 정확도를 보이지만, 앞선 결과와 같이 XGBoost 알고리즘이 가장 정확한 결과를 보이는 것으로 나타났으며, 그 다음 정확한 결과를 보이는 알고리즘은 랜덤 포레스트인 것으로 나타났다.

10가지 주요 변수를 살펴보면, 성인용 콘텐츠 사용 빈도가 과의존 예측에 가장 주요한 변수로 선정되었다. 같은 데이터를 다룬 Lee and Kim (2021)에서도 스마트폰 위험군과 일반 사용자군의 성인 콘텐츠 이용 빈도를 비교해 보면, 위험군이 일반 사용자 군보다 약 1,893배 더 자주 이용하는 것으로 나타나고 있다. 또한, 해당 요인이 주요한 원인으로는 요인이 콘텐츠에 대한 이용 빈도 뿐만 아니라 나이에 관한 정보를 포함하고 있기 때문으로 추측되는데 성인용 콘텐츠를 사용하는 응답자는 대부분 성인이기 때문에 이용 빈도가 잦은 경우 응답자가 성인일 가능성이 높다. 주요 변수 목록에서 볼 수 있듯이 나이는 과의존 예측에 필요한 주요 요인 중 하나이며, Lee and Kim (2021)에서도 주요 원인으로 보고 나이대 별로 모형을 따로 학습하고 있다. 그 다음 중요한 요인은 게임 이용 여부(이진 변수)로 나타났는데, 게임을 얼마나 자주 하는지 나타내는 빈도보다 게임을 하는지 안 하는지 여부가 더 주요한 요인으로 나타났다. 이는 오히려 고위험군이 아닌 사용자군을 관별하기 위한 것으로 추측되는데, 게임 콘텐츠 이용 빈도의 경우 고위험군의 응답 평균이 4.59인데 반하여, 다른 사용자군의 평균은 3.22로 나타난다. 표준 편차 또한 다른 사용자군의 경우 2.176으로 상대적으로 큰 값을 갖는 것으로 보아 아예 게임을 하지 않는 이용자들은 스마트폰 과의존군이 아닐 가능성이 높다. 실가구주와의 관계에 관한 문항이 3번째 주요 요인 이자, 인구통계학적 특성에 관한 문항 중 가장 주요한 문항으로 나타났는데, 이는 응답으로 나이대와 함께 기혼 여부를 동시에 알 수 있기 때문으로 추측된다. 4, 5번째 순위를 보면, 주중과 주말의 스마트폰 횟수가 주요 요인으로 나타났다. 스마트폰 사용 시간에 관한 변수 또한 주요 20가지 요인에 포함되었으나, 횟수가 더 관별에 주요한 영향을 미치는 것으로 나타나고 있다. 이는 향후 과의존 수준을 파악하기 위한 설문에서 응답자의 스마트폰 사용량을 측정하게 될 때, 시간보다 횟수에 기반을 두고 측정하는 것이 더 나을 수도 있다는 것을 시사한다. 그 밖에 다양한 스마트폰 사용 콘텐츠의 이용 빈도가 주요 요인으로 나타났다.

인구 통계학적 특성을 살펴보면 주요 요인 20가지 중 인구통계학적 특성과 관련된 문항은 3가지 포함되었는데 주요 변수들을 보면 성별, 거주지역, 가구원 수, 수입, 다문화/한부모 가정 여부나 장애인 여부, 가구 소득 등과 같은 인구통계학 특성들은 과의존 수준을 관별하는데 주요 영향을 미치지 않는 것으로 나타났으며, 나이, 기혼 여부, 학생(재학) 여부가 주요 변수로 작용하는 것으로 나타났다. 아마도 나이대와 기/미혼 여부, 학생인지 아닌지에 따라서 집단의 스마트폰 과의존 패턴이 다르기 때문으로 추측되며, 이는 부분적으로 Lee and Kim (2021)의 결과를 뒷받침한다. 연구에서는 정확도를 높이기 위하여 나이대 별(10대, 20대, 30대)로 예측 모형을 따로 학습하였는데, 이는 나이가 과의존 관별에 주요 요인으로 사용됨을 의미한다. 또한 연구에서는 스마트폰 사용 패턴으로는 성별을 예측할 수 없다고 밝히며, 성별에 따른

스마트폰 사용 패턴의 차이가 없음을 언급하고 있는데, 본 연구에서도 성별은 과의존 판별을 위한 주요 요인으로 밝혀지지 않았다. 하지만 선행 연구에서는 인구통계학적 특성 중 미/기혼 여부나 재학 여부와 같은 정보들은 포함되지 않았는데, 본 연구에서는 해당 정보들이 주요 요인들로 밝혀졌다. 예측 모형에 사용된 다른 요인들이 차이가 없는 것을 고려해 보면, 해당 인구통계학적 특성들이 두 연구에서의 예측력 차이에 어느 정도 영향을 미치는 것으로 추정된다.

대부분의 주요 변수들은 스마트폰 사용과 관련된 문항들로 나타났다. 다양한 스마트폰 콘텐츠 사용에 관한 문항들이 요인으로 선정되었으나, 교통 및 위치 정보, 일반적인 웹서핑, 이메일, 메신저, SNS, 상품 구매, 상품 판매, 금융(스마트폰 뱅킹, 주식 거래), 생활 관리(가계부, 일기 등) 등과 같은 콘텐츠 이용 빈도 및 이용 여부는 판별을 위한 주요 요인으로 선정되지 않았다. 흥미로운 것은 SNS나 메신저와 같이 대인 관계와 관련된 요인들 또한 주요 요인으로 선정되지 않았다는 점이다. 이는 선행 연구에서 스마트폰이 다양한 편의를 제공하기 때문에 SNS 중독으로 이어질 수도 있다는 주장과는 다소 배치되는 결과이다. 본 연구의 결과로서는 스마트폰 과의존은 SNS 중독과는 다른 패턴을 보이며, 이에 관하여는 더 심도 있는 연구가 필요할 것으로 보인다. 또한 스마트폰 요금이나 게임 과금과 같이 스마트폰 사용이나 콘텐츠 관련된 비용은 주요 요인으로 선정되지 않았다. 아마도 이는 스마트폰 요금의 경우, 현재 한국의 많은 장소에서 무료 와이파이를 제공하기 때문에 한 장소에서 사용할 때는 요금을 지불하지 않는 경우가 많으며, 또한 사용량에 맞는 다양한 요금제가 존재하기 때문으로 추측된다.

분류 성능에 관한 척도들의 경우, 대부분의 알고리즘에서 더 많은 정보가 주어질수록 성능이 높아지는 것으로 나타나지만 그 개선의 정도는 미약한 것으로 판단된다. k -NN 알고리즘의 경우에는 20가지 변수를 사용하는 경우와 전체 변수를 모두 사용하는 경우의 분류 성능이 유사한 것으로 나타났다. Adaboost 알고리즘의 경우 20가지 변수를 사용했을 때의 예측 성능이 전체 변수를 사용했을 때 예측 성능보다 높게 나타났는데, 알고리즘에 따라 변수들의 개수에 따라 성능이 상이한 것을 볼 수 있다. k -NN, Adaboost 알고리즘을 사용하는 경우, 20가지 요인들을 사용했을 때도 알고리즘이 충분한 분류 성능을 발휘하게 되는 것으로 보인다.

V. 결론

스마트폰 보급률 증가에 따라 이에 과도하게 의존하는 사람들의 수도 증가하고 있으며, 이를 예방하기 위하여 과의존 수준을 진단하고, 측정하기 위한 다양한 척도들이 개발되었다. 이러한 과의존 수준 측정 척도들은 스마트폰 사용의 심리적 문제, 일상생활에서의 부작용 등에 관한 유용한 정보를 제공하며, 이에 관련된 다양한 연구들이 수행되었지만, 인구통계학적 특성과 스마트폰 사용 패턴을 기반으로 과의존 수준을 판별하는 연구는 드문 편이다. 스마트폰 과의존 위험군 예측의 경우, 스마트폰 사용에 관한 단일한 예측 변수로는 과의존 여부를 예측할 수 없기 때문에 다수의 변수들을 복합적으로 고려해야 하며, 전통적인 통계 기법을 판별 문제에 사용하기에 어려운 점이 있다(Lee and Kim, 2021). 이에, 본 연구에서는 인구통계학적 특성들과 스마트폰 사용 패턴에 관한 데이터들을 기계 학습 기법을 통하여 분석하여 스마트폰 과의존 수준을 판별하는 문제를 다루었다. 2017년 한국진흥정보사회진흥원에서 시행한 스마트폰 과의존 실태조사 설문에서 관련 변수들을 선정하여, 분석에 사용하였으며, k -NN (k -nearest neighbor), 그레디언트 부스팅(Gradient boosting), XGBoost, CatBoost, AdaBoost, 랜덤 포레스트 알고리즘을 적용하였다. 또한 정보 이득(information gain)에 기반하여 과의존 수준 판별에 영향을 미치는 인구통계학적 요인과 스마트폰 사용과 관련된 주요 변수들을 제시하고, 예측 변수들의 수에 따른 알고리즘의 성능 또한 제시하였다. 모든 알고리즘의 정확도(accuracy), 정밀도(precision), 재현율(recall), F1 점수가 0.8이상으로 비교적 정확한 분류 성능을 보이는 것으로 나타났다.

본 연구의 시사점은 다음과 같다. 학문적으로 볼 때, 이는 기계 학습 기법이 스마트폰 과의존 판별에 있어서 유용하게 사용될 수 있다는 가능성을 시사한다. 정확도의 경우 XGBoost 알고리즘이 가장 높은 분류율을 보였으며, 그 다음으로 랜덤 포레스트 알고리즘의 성능이 우수한 것으로 나타났으나, 대부분의 알고리즘이 80%를 상회하는 분류 정확도를 보였으며, 이는 선행 연구보다 개선된 것이다. 또한, 본 연구의

결과는 효과적인 과의존 관별을 위하여 얼마나 많은 변수들이 필요하며 어떤 변수들을 사용해야 하는지에 관한 유용한 정보를 제공한다. 현재 진행하는 과의존 실태조사 설문은 약 180문항으로 구성되어 있는데, 다양한 요인 중 스마트폰 과의존을 관별하기 위한 주요 요인들을 정보 이득(information gain) 척도를 사용하여 선별하였다. 인구통계학적 특성 중에는 나이, 혼인 여부, 학생 여부 등과 같은 요인은 과의존 관별에 주요 변수로 선별된 데 반하여, 성별이나 지역 주거 형태, 가구원 수 등등의 다른 요인들은 주요 요인으로 선정되지 않았다. 특히 성별의 경우, 전호선, 장승욱(2014), 이하나, 양승목(2018)과 같은 몇몇 선행 연구에서 스마트폰 이용 용도가 스마트폰 과의존에 미치는 영향이 성별로 다르다고 언급하고 있다. 하지만, 본 연구에서는 주요 요인으로 선정되지 않았는데, 이는 성별에 따라 다르게 나타나는 스마트폰 이용 패턴이 이미 스마트폰 콘텐츠 사용 빈도에 관한 다른 변수들에 포함되어 있기 때문으로 해석된다. 예를 들어, 이하나, 양승목(2018)에서는 게임 콘텐츠의 이용이 중, 고등학교 남학생의 스마트폰 의존도를 높이는 반면에 여학생과는 관계가 없었다고 언급하고 있는데, 예측 모형에서는 이미 게임 콘텐츠 이용에 관한 변수들이 있기 때문에 성별에 관한 정보 없이도 과의존을 관별할 수 있는 것이다. 스마트폰 사용 콘텐츠들에 관련된 변수들은 대부분 주요 요인에 들어가게 되었으나, 교통 및 위치 정보, 일반적인 웹서핑, 이메일, 메신저, SNS, 상품 구매, 상품 판매, 금융(스마트폰 बैं킹, 주식 거래), 생활 관리(가계부, 일기 등) 등과 같은 콘텐츠 이용 빈도 및 이용 여부는 관별을 위한 주요 요인으로 선정되지 않았으며, 스마트폰 요금이나 게임이나 영화에 사용하는 비용 또한 주요 변수로 선정되지 않았다. 과의존 예측에 주요 변수로 선정된 20가지 변수들로 학습된 예측 모형이 전체 변수를 사용한 예측 모형과 거의 동일한 (1% 이내) 분류 정확도를 갖는 것을 미루어 보았을 때, 이는 나이, 혼인 여부, 학생 여부와 같은 몇몇 인구통계학적 특성과 자주 사용하는 콘텐츠들에 대한 정보만으로도 충분히 과의존 관별이 가능하다는 것을 의미한다. 실무적으로 볼 때는 이는 스마트폰 로그 데이터에 기반한 효과적인 과의존 진단 프로그램이나 어플리케이션 개발 등에 사용될 수 있을 것이다. 앞서 언급했듯, 기존의 과의존 관별을 위한 설문들은 응답자의 심리적인 부분이나 일상생활에서의 부작용에 관한 질문들이 대부분이기 때문에, 응답자가 답변을 거부하거나, 솔직하게 답변하지 않는다면 부정확한 결과를 얻을 수도 있다. 반면, 스마트폰 로그 데이터를 사용하여 과의존을 진단한다면 보다 객관적인 정보를 기반으로 과의존을 예측할 수 있으며, 스마트폰 이용 패턴과 과의존과의 관계에 관한 유용한 정보를 제공할 수 있을 것으로 생각된다. 스마트폰 로그 데이터를 사용하게 되는 경우 중요한 이슈 중의 하나는 스마트폰 기기에서 정보를 효과적으로 수집하여, 좋은 예측 성능을 발휘하는 것이다. 스마트폰 기기에서 더 많은 종류의 데이터를 수집하여 예측 모형을 학습하는 경우, 높은 예측 성능을 낼 가능성이 높지만 데이터를 수집하고, 예측 모형을 학습하는데 드는 비용 또한 증가하게 된다. 이는 결국 데이터 수집과 예측 모형 학습을 위한 비용과 예측 성능 사이의 균형을 추구하는 문제가 된다. 본 연구의 결과에 따르면, 설문에서 제시된 스마트폰 사용과 인구통계학적 특성에 관한 모든 변수를 사용하는 경우, 10개의 변수를 사용하는 것에 비하여 평균적으로 2.3% 정도의 정확도 개선을, 20개 변수를 사용하는 것보다는 1% 미만의 정확도 개선을 기대할 수 있는 것으로 나타났다. 실제 프로그램이나 어플리케이션을 개발하는 경우, 더 나은 예측 성능을 얻기 위하여 얼마나 많은 비용을 들이느냐를 고려하게 될 때, 본 연구 결과는 이에 관련된 유용한 결과를 제시한다.

본 연구는 다양한 방향으로 확장될 수 있다. 먼저, 본 연구에서는 스마트폰 사용에 관한 설문 데이터를 사용하고 있다. 이는 다수의 스마트폰 사용에 관한 실태데이터를 얻기가 힘들기 때문에 어쩔 수 없는 일이다. 하지만, 대량의 스마트폰 사용 로그 데이터들에 관한 데이터를 확보하여 분석한다면, 더 정확하고 심도 있는 분석을 할 수 있으며, 본 연구에서 제시하는 결과와 일치하는지 확인할 수 있을 것이다. 또한, 현재는 지도 학습 기법들을 적용하였지만, 다양한 비지도 학습 기법 또한 적용할 수 있다. 연관 규칙이나 클러스터링 기법을 사용하여 사용 패턴이 유사한 집단을 선정하거나, 각 독립 변수들 간의 관계를 분석하는 것 또한 유용한 정보를 제공할 수 있을 것이다. 마지막으로, 본 연구 결과를 활용하여 더욱 효과적으로 과의존을 관별할 수 있는 설문 척도를 개발하는 것 또한 의미가 있을 것이다.

References

- 김희양, 박창호 (2012), “스마트폰 중독 척도의 개발”. *한국심리학회지: 일반*, 31(2), 563-580.
- 강혜자 (2016), “스마트폰 사용과 정신건강 및 신체건강 간의 관계”. *Journal of Digital Convergence*, 14(3), 483-488.
- 김병년 (2013), “청소년의 스마트폰 중독이 사회성 발달에 미치는 영향”. *한국콘텐츠학회논문지*, 13(4), 208-217.
- 김병년, 고은정, 최홍일 (2013), “대학생의 스마트폰 중독에 영향을 미치는 요인에 관한 연구: 중독 위험군 분류에 따른 차이를 중심으로”. *한국청소년연구*, 24(3), 67-98.
- 김서영, 권순필 (2009), “응답 거부와 부재율이 무응답 오차에 미치는 영향: 경제활동인구조사를 중심으로”. *응용통계연구*, 22(3), 667-676.
- 박용민 (2011), *성인들의 스마트폰 중독과 정신건강에 관한 연구 (석사학위논문)*. 상지대학교.
- 이성수 (2022), “스마트폰 중독 경향성에 대한 소외감의 영향에서 스트레스를 통한 마음챙김의 조절된 매개효과”. *한국융합학회논문지*, 13(4), 185-194.
- 이정정, 이정화, 김은미, 박병기 (2021), “대학생의 스마트폰 중독 척도 개발 및 타당화”. *아동교육*, 30(1), 157-184.
- 이철현 (2020), “초등학생의 스마트폰 중독 관련 데이터를 활용한 머신러닝 알고리즘의 성능 분석”. *한국실과교육학회지*, 33, 103-119.
- 이하나, 양승목 (2018), “청소년의 스마트폰 의존도와 영향 변인의 관계에 대한 학령 및 성별 비교 분석: 부모의 양육태도, 청소년의 심리적 특성, 스마트폰 이용 용도를 중심으로”. *한국언론학보*, 62(5), 175-214.
- 전호선, 장승욱 (2014), “스트레스와 우울이 대학생 스마트폰 중독에 미치는 영향: 성별 조절효과를 중심으로”. *청소년학연구*, 21(8), 103-129.
- 정구철 (2016). “대학생들의 정신건강과 인터넷 중독 및 스마트폰 중독 간의 관계”. *한국콘텐츠학회논문지*, 16(4), 655-665.
- 한국갤럽 (2021), 2012-2021 “스마트폰 사용률, 브랜드, 스마트워치, 무선이어폰에 대한 조사”. Available from <https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=1217> (accessed May 16, 2022).
- 황경혜, 유양숙, 조옥희 (2012), “대학생의 스마트폰 중독사용 정도에 따른 상지통증, 불안, 우울 및 대인관계”. *한국콘텐츠학회논문지*, 12(10), 365-375.
- 한국지능정보사회진흥원 (2017), “2017년 스마트폰 과의존 실태조사”. Available from https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=19592&parentSeq=19592 (accessed May 16, 2022).
- 한국지능정보사회진흥원 (2021), “2021년 스마트폰 과의존 실태조사”. Available from https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=24288&parentSeq=24288 (accessed May 16, 2022).
- AlAbdulwahab, S. S., S. J. Kachanathu, and M. S. AlMotairi (2017), “Smartphone Use Addiction can Cause Neck Disability”. *Musculoskeletal care*, 15(1), 10-12. <http://dx.doi.org/10.1002/msc.1170>
- Bianchi, A., and J. G. Phillips (2005), “Psychological Predictors of Problem Mobile Phone Use”. *CyberPsychology & Behavior*, 8(1), 39-51. <http://doi.org/10.1089/cpb.2005.8.39>
- Choliz, M. (2012), “Mobile-Phone Addiction in Adolescence: the Test of Mobile Phone Dependence (TMD)”. *Progress in health sciences*, 2(1), 33-44.
- Kim, Yu Jeong, and Dong Su Lee (2020). “Predictive Analysis of Problematic Smartphone Use by Machine Learning Technique”. *Journal of The Korea Society of Computer and Information*, 25(2), 213-219. <https://doi.org/10.9708/jksci.2020.25.02.213>
- Kwon, Min., Dai-Jin Kim., Hyun Cho, and Soo Yang. (2013a), “The Smartphone Addiction Scale: Development and Validation of a Short Version for Adolescents”. *PLoS one*, 8(12), e83558. <https://doi.org/10.1371/journal.pone.0083558>
- Kwon, Min, Joon-Yeop Lee, Wang-Yoon Won, Jae-Woo Park, Jung-Ah Min, Changtae Hahn, Xinyu Gu et al. (2013b). Development and Validation of a Smartphone Addiction Scale (SAS). *PLoS one*, 8(2), e56936. <https://doi.org/10.1371/journal.pone.0056936>

- Lee, Jooyeong. and Woosung Kim (2021), "Prediction of Problematic Smartphone Use: A Machine Learning Approach". *International Journal of Environmental Research and Public Health*, 18(12), 6458. <https://doi.org/10.3390/ijerph18126458>
- Leung, L. (2008), "Linking Psychological Attributes to Addiction and Improper Use of the Mobile Phone among Adolescents in Hong Kong". *Journal of children and media*, 2(2), 93-113. <https://doi.org/10.1080/17482790802078565>
- Matar Boumosleh, J. and D. Jaalouk (2017), "Depression, Anxiety, and Smartphone Addiction in University Students-A Cross Sectional Study". *PLoS one*, 12(8), e0182239. <https://doi.org/10.1371/journal.pone.0182239>
- Mohammadbeigi, A., R. Absari, F. Valizadeh, M. Saadati, S. Sharifimoghadam, A. Ahmadi et al. (2016), "Sleep Quality in Medical Students; the Impact of Over-Use of Mobile Cellphone and Social Networks". *Journal of research in health sciences*, 16(1), 46.
- Panova, T. and X. Carbonell. (2018), "Is Smartphone Addiction Really an Addiction?". *Journal of behavioral addictions*, 7(2), 252-259. <https://doi.org/10.1556/2006.7.2018.49>
- Peraman, R., and S. Parasuraman (2016), "Mobile Phone Mania: Arising Global Threat in Public Health". *Journal of natural science, biology, and medicine*, 7(2), 198.
- Samaha, M., and N. S. Hawi (2016). "Relationships among Smartphone Addiction, Stress, Academic Performance, and Satisfaction with Life". *Computers in human behavior*, 57, 321-325. <https://doi.org/10.1016/j.chb.2015.12.045>