

Predictive Analysis of Problematic Smartphone Use by Machine Learning Technique

Yu Jeong Kim*, Dong Su Lee**

*Professor, Dept. of Nursing Science, Chosun Nursing College, Gwang-Ju, Korea

**Professor, Dept. of Computer Science and Statistics, Chosun University, Gwang-Ju, Korea

[Abstract]

In this paper, we propose a classification analysis method for diagnosing and predicting problematic smartphone use in order to provide policy data on problematic smartphone use, which is getting worse year after year. Attempts have been made to identify key variables that affect the study. For this purpose, the classification rates of Decision Tree, Random Forest, and Support Vector Machine among machine learning analysis methods, which are artificial intelligence methods, were compared. The data were from 25,465 people who responded to the '2018 Problematic Smartphone Use Survey' provided by the Korea Information Society Agency and analyzed using the R statistical package (ver. 3.6.2). As a result, the three classification techniques showed similar classification rates, and there was no problem of overfitting the model. The classification rate of the Support Vector Machine was the highest among the three classification methods, followed by Decision Tree and Random Forest. The top three variables affecting the classification rate among smartphone use types were Life Service type, Information Seeking type, and Leisure Activity Seeking type.

▶ **Key words:** Problematic Smartphone Use, Machine Learning, Smartphone Use Type, Decision Tree, Random Forest, and Support Vector Machine

[요 약]

본 연구는 스마트폰 과의존을 진단하고 예측하기 위하여 할 수 있는 분류분석 방법과 스마트폰 과의존 분류율에 영향을 미치는 중요변수를 규명하고자 시도되었다. 이를 위해 인공지능의 방법인 기계학습 분석 기법 중 의사결정트리, 랜덤포레스트, 서포트벡터머신의 분류율을 비교하였다. 자료는 한국정보화진흥원에서 제공한 '2018년 스마트폰 과의존 실태조사'에 응답한 25,465명의 데이터였고, R 통계패키지 (ver. 3.6.2)를 사용하여 분석하였다. 분석한 결과, 3가지 분류분석 기법은 정분류율이 유사하게 나타났으며, 모델에 대한 과적합 문제가 발생되지 않았다. 3가지 분류분석 방법 중 서포트벡터머신의 분류율이 가장 높게 나타났고, 다음으로 의사결정트리 기법, 랜덤포레스트 기법 순이었다. 스마트폰 이용 유형 중 분류율에 영향을 미치는 상위 3개 변수는 생활서비스형, 정보검색형, 여가추구형이었다.

▶ **주제어:** 스마트폰 과의존, 기계 학습, 스마트폰 이용 유형, 의사결정기법, 랜덤 포레스트, 서포트 벡터 머신

• First Author: Yu Jeong Kim, Corresponding Author: Dong Su Lee

*Yu Jeong Kim (fight1004@scnc.ac.kr), Dept. of Nursing Science, Chosun Nursing College

**Dong Su Lee (gs60311@hanmail.net), Dept. of Computer Science and Statistics, Chosun University

• Received: 2020. 01. 15, Revised: 2020. 02. 11, Accepted: 2020. 02. 14.

I. Introduction

한국은 모바일 인터넷 속도 세계 1위, 스마트폰 보유율 세계 1위, 지하철에서도 무료 Wi-Fi (wireless fidelity, 무선 데이터 전송 시스템)가 제공되는 스마트폰 최적 사용 국가이다[1]. 스마트폰은 휴대가 간편하고, 장소와 시간의 구애받지 않아서 기존의 커뮤니케이션 용도를 넘어 여가 추구, 정보 검색 및 생활 서비스로 활용도가 급증하고 있다. 스마트폰은 이용 형태에 따라 생활 만족도를 증가시키고, 대인관계, 학업 및 업무 역량을 강화시켜준다고 보고되었다[2]. 반면에 일부 스마트폰 이용자들은 다양한 역기능을 호소하고 있으며, 스마트폰 과의존은 대표적인 역기능으로서 7년 연속 증가하고 있다[3-7].

국가차원에서 스마트폰 과의존을 예방하기 위하여 국가정보화기본법을 개정하여 학교를 중심으로 스마트폰 과의존 예방교육을 의무실시하도록 하고[8], 관계부처 합동으로 대책마련에 고심을 하고 있음에도 불구하고, 스마트폰 과의존 위험군은 19.1%에 육박하고 있다. 이런 사회적 현상의 근본이유는 스마트폰 과의존 진단을 개인의 주관적 판단에 의존하여 실시하고 있기 때문이다. 스마트폰 과의존 진단이 미흡하기 때문에 개선노력이 따르지 않으며, 예방교육도 형식적일 수밖에 없다.

따라서 본 연구팀은 스마트폰 과의존 현상을 이용 유형과 개인적 특성에 따라 정확하게 진단 및 예측하기 위하여 인공지능(artificial intelligence : AI) 분석기법 중 기계학습(machine learning)을 이용하고자 한다. 대표적인 기계학습 분석 중에서 의사결정트리, 랜덤포레스트, 서포트벡터머신을 적용하여 가장 정확한 스마트폰 과의존 분석률을 나타내는 분석방법이 무엇인지 파악하여 스마트폰 과의존의 예측모형을 개발과 예방교육프로그램 개발을 위한 기초자료를 제공하고자 한다.

II. Preliminaries

1. Problematic Smartphone Use

스마트폰 과의존(problematic smartphone use)은 스마트폰에 대한 현저성이 증가하고, 이용 조절력이 감소하여, 문제적 결과를 경험하는 상태"를 의미한다[7]. 현저성(salience)은 스마트폰을 이용하는 행태가 다른 생활패턴보다 중요한 활동이 되는 것을 말한다. 조절 실패(self-control failure)는 스마트폰을 스스로 조절해서 사

용하는 능력이 떨어지는 것이며, 문제적 결과(serious consequences)는 스마트폰을 이용하면서 신체적·심리적·사회적 부정적인 결과를 경험하고 있음에도 불구하고 스마트폰을 지속적으로 이용하는 것을 말한다.

스마트폰 과의존시 발생할 수 있는 신체적·심리적·사회적 문제는 안구 건조증, 시력저하, 거북목 현상, 팝콘브레인, 손목터널증후군, 우울증, 디지털치매, 수면치매, 충동 조절장애, 대인 예민증 등이 있다. 또한 각종 유해정보 노출되면서 가상공간에서 정보침해, 인격침해, 일반사이버범죄, 사이버테러 등 사회적 문제로 확산될 수 있다[3-6].

매년 한국정보화진흥원에서 실시하는 스마트폰 과의존 실태조사에 따르면, 국내 스마트폰 이용자 5명 중 1명 정도가 과도한 스마트폰 이용으로 인한 역기능을 경험하고 있다[7]. 최근에는 스마트폰 과의존이 청소년층에 국한된 문제가 아니라 유아동 및 노년층에서도 급상승하고 있으며, 1인 가구의 스마트폰 이용률이 작년 대비 2.5%로 가장 높게 상승하였다[7]. 이러한 스마트폰 과의존 실태를 볼 때 범국민적 스마트폰 과의존 예방 노력을 이끌어 낼 수 있는 정책대안이 시급하며, 효과적이고 현실성 있는 대책마련을 위하여 조기에 스마트폰 과의존을 진단하고 예측할 수 있는 인공지능을 활용한 분석이 필요함을 알 수 있다.

선행연구에 따르면 스마트폰 과의존은 이용자, 이용 형태, 개인적 특성에 따라 다양하게 나타났다. 장노년층의 스마트폰 과의존 현상은 정보 추구형과 게임을 위한 스마트폰 사용이 증가할수록 높았고, 오락 추구형과 커뮤니케이션형에서는 나타나지 않았다[9]. 반면에 청소년층의 스마트폰 과의존 현상은 정보 추구형, SNS(social networking service) 및 게임을 위한 스마트폰 사용이 스마트폰 과의존에 영향을 미쳤고, 여가추구형 및 커뮤니케이션형에서는 영향을 미치지 못했다[11]. 개인적 특성 중 스마트폰 이용시간, 디지털 사용 역량이 높을수록, 대인관계와 생활 만족도가 불안하거나 낮을수록 스마트폰 과의존 경향이 나타난다고 보고되었다[7].

본 연구는 한국정보화진흥원에서 한국종단 데이터에서 조사되었던 스마트폰 이용 형태(커뮤니케이션형, 여가추구형, 정보검색형, 생활서비스형) 및 개인적 특성변수(스마트폰 이용시간, 디지털역량, 대인관계, 생활 만족도)를 스마트폰 과의존 위험군 선별을 위한 원인변수로 선정하여 분석하였다.

2. Machine Learning

기계학습은 AI방법으로 분류분석에 해당한다. 기계학습의 대표적인 분류기법으로 의사결정트리, 랜덤포레스트,

서포트벡터머신 기법이 있다[12]. 3가지 기계학습 기법을 살펴보고 가장 정확한 분류를 나타내는 분석방법이 무엇이며, 분류율에 영향을 미치는 중요 변수가 무엇인지를 규명하고자 한다.

2.1 Decision Tree

의사결정트리(decision tree, DT)는 지도학습에서 가장 유용한 분류와 예측 분석방법이다. 추론규칙이 나무처럼 펼쳐지므로 의사 결정 과정과 결정이 시각적으로 명료하게 표현된다. DT는 4단계를 걸쳐서 분류 예측이 된다.

1단계에서는 DT를 생성한다. 분석목적에 따라 적절한 분리기준과 정지규칙(stopping rule)을 정한다. 2단계에서는 가지를 친다. 분류 오류(classification error)를 크게 할 가능성이 있거나 부적절한 추론규칙(induction rule)을 갖는 가지(branch)를 제거한다. 3단계는 타당성 평가과정이다. 이익표(gains chart)나 위험도표(risk chart) 또는 검증용 자료를 토대로 교차타당성을 분석한다. 끝으로 4단계에서 해석 및 예측을 한다. DT를 전체적으로 해석하고 예측 모형을 설정한다. DT는 분리기준, 정지규칙, 평가기준 등을 어떻게 적용하느냐에 따라 다양한 형태의 모양으로 해석될 수 있다. DT는 과거의 학습한 데이터는 높은 예측을 나타내지만, 새로운 데이터에 대해서는 예측이 낮아지는 경우로 일반화에 문제가 발생할 수 있다.

2.2 Random Forest

랜덤포레스트(random forest, RT)는 의사결정트리의 기법의 모델을 반복하는 앙상블(ensemble) 학습기법의 일종이다. RT는 일반화에 문제가 발생할 수 있는 과적합(overfitting) 문제를 해결할 수 있다. Fig. 1과 같이 주어진 데이터로부터 여러 개의 모델을 학습한 다음, 여러 모델의 예측 결과들을 종합해서 예측하여 정확도를 높이는 기법이다.

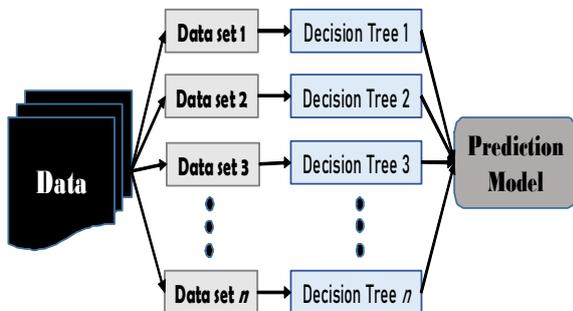


Fig. 1. Random Forest

RT는 대용량 데이터에 적합하며 많은 변수를 입력해도 변수 제거 없이 실행할 수 있어서 예측이 높은 편이다. 대표

적인 활용사례로는 태풍경로를 예측할 때나 선거결과에 가장 중요한 요인을 분류할 때 사용될 수 있다. RT는 변수들에 대한 정확도계수와 지니계수를 통해 중요도를 규명한다.

2.3 Support Vector Machine

서포트벡터머신(support vector machine, SVM)은 분류기법 중 다양하게 분포되어 있는 데이터를 정확하게 분류하는 경향이 있다. 다른 분류 기법들보다 분류율이 상대적으로 높아서 패턴인식, 자료 분류, 회귀분석에 적용이 가능한 분류분석 방법이다. SVM 모형은 3차원 공간에서 최적의 경계를 찾아 데이터를 분리하는 방식이다. 이 경계를 초평면(hyperplane)이라고 한다.

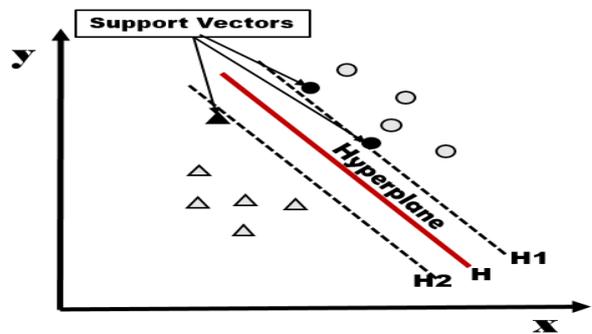


Fig. 2. support vector machines

위 그림에서와 같이 margin이 가장 큰 초평면을 분류기(classifier)로 사용할 때 새로운 데이터에 대한 오분류가 가장 낮아진다. SVM은 최대 마진 초평면(maximum margin hyperplane, MMH)을 찾는 분류방법이며 다음 식으로 표현된다.

$$f(x) = w^T \chi_i + b$$

위 식은 주어진 입력 벡터값에 대한 실제 목표 값의 편차내에서 가능한 좁은 범위의 가중치 벡터값 $w^T \chi_i$ 를 갖는 함수 $f(x)$ 를 찾는 것이며, 여기서 상첨자는 행렬의 전치(transpose)이며, b 는 바이어스(bias)이다.

III. The Proposed Scheme

1. Research Model

본 연구에서는 스마트폰 이용 형태 및 개인적 특성변수에 따라 스마트폰 과의존 여부가 분류되는지를 규명하기 위해 여러 변수를 설정하였다. 여러 원인변수들의 특성이 스마트폰 과의존 정도에 따라 다르게 분류되는지를 분석하였다. 분류분석 위해 설정한 8가지 원인변수는 스마트폰

이용 형태(커뮤니케이션형, 여가추구형, 정보검색형, 생활 서비스형)와 개인적 특성(스마트폰 이용시간, 디지털역량, 대인관계, 생활만족도) 이었다. 종속변수는 스마트폰 과의존 수준이다. 원인변수와 결과변수의 영향관계에서 기계 학습 분석 유형 따라 어떠한 분류율이 있는지 알아보고자 Fig. 3. 과 같이 연구모형을 설정하였다.

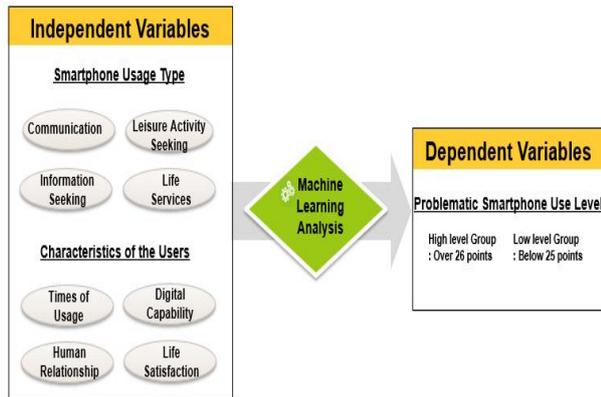


Fig. 3. Research Model

2. Data Analysis

본 연구는 한국정보화진흥원에서 제공한 '2018 스마트폰 과의존 실태조사' 패널 데이터를 활용하였다. 조사당시 자료수집 기간은 2018년 8월부터 10월까지 이었다. 조사대상은 전국 10,000개 가구 내 만 3~69세 스마트폰(인터넷) 이용자(최근 1개월 이내 1회 이상 이용)이었다. 조사 당시 표본수는 28,575명이었으나 9세 이하의 주돌봄자의 관찰에 근거한 자료이므로 데이터의 신뢰성 확보에 어려움이 있어 본 연구에서는 분석에 활용하지 않았다. 본 연구의 활용된 응답자는 청소년 4,889명과 성인 18,571명, 고령자 2005명을 합하여 총 25,465명이다. 수집된 자료는 빅데이터 분석 도구인 R 통계패키지(ver. 3.6.2)를 사용하여 분석하였다.

3. Variable Measurement Level

첫째, 여가추구형은 총 7문항으로 구성되며 7점 척도를 사용하였다. 점수가 높을수록 스마트폰을 여가 추구에 많이 이용한다는 것을 의미한다.

둘째, 정보검색형은 총 5문항으로 구성되며 7점 척도를 사용하였다. 점수가 높을수록 스마트폰을 정보 검색에 많이 이용한다는 것을 의미한다.

셋째, 커뮤니케이션형은 총 4문항으로 구성되며 7점 척도를 사용하였다. 점수가 높을수록 스마트폰을 커뮤니케이션에 많이 이용한다는 것을 의미한다.

넷째, 생활서비스형은 총 4문항으로 구성되며 7점 척도를 사용하였다. 점수가 높을수록 스마트폰을 생활 서비스에 많이 이용한다는 것을 의미한다.

다섯째, 스마트폰 이용시간은 순위척도를 등간척도로 환산하여 5점 척도로 사용하였다. 1시간 미만은 1점, 1시간~2시간 미만은 2점, 2시간~3시간 미만은 3점, 3시간~4시간 미만은 4점, 4시간 이상은 5점으로 하였다.

여섯째, 스마트폰 과의존은 총 10문항의 4점 척도를 사용하였다. 문항은 조절실패 요인 3문항, 현저성 요인 3문항, 문제적 결과 요인 4문항으로 구성되어 있으며, 점수가 높을수록 스마트폰 과의존이 높다는 것을 의미한다.

일곱째, 생활 만족도는 총 7문항으로 구성되며, 4점 척도를 사용하였다. 점수가 높을수록 생활 만족도가 높다는 것을 의미한다.

여덟째, 디지털 사용역량은 총 6문항으로 구성되며 4점 척도를 사용하였다. 점수가 높을수록 디지털 사용 역량이 높다는 것을 의미한다.

아홉째, 대인관계는 총 3문항으로 구성되며 4점 척도를 사용하였다. 점수가 높을수록 대인관계가 폭넓다는 것을 의미한다.

IV. Results

1. Problematic Smartphone Use level of Decision Tree technique

DT기법으로 스마트폰 과의존을 분류하기 위해 분지규칙과 정지규칙을 적용하였다. 각 변수 값들에 대한 표현을 나뉘어가지 모양으로 학습용 데이터(train data)와 검증용 데이터(test data)로 구분하였다. 이중 학습용 데이터를 스마트폰 과의존에 대한 의사결정트리 기법의 모델에 적용하였다. 스마트폰 과의존을 적용한 모델의 적합성 여부는 검증용 데이터를 통해 확인했다. 스마트폰 과의존에 대한 의사결정트리 분석결과는 Table 1. 과 같다. 전체 데이터 25,465개를 학습용 데이터는 70%(17,847개), 검증용 데이터 30%(7,618개)로 구분하였다.

스마트폰 과의존에 대한 정분류율은 학습용 데이터에서 87.63%(0.8763378), 검증용 데이터에서 88.00%(0.880021)로 나타났다. 학습용 데이터와 검증용 데이터의 분류율 차이는 유사하여 과적합(overfitting) 문제가 없는 것으로 판단되었다.

Table 1. Analysis result of Decision Tree

| Train data | | | | |
|--|----------------------|-------|------------|----------------------------|
| Response: train \$ Problematic Smartphone Use1 | | | | |
| Inputs: train\$InS , train\$LAS, train\$ Com, train\$Lse, train\$Tus, train\$Dca, train\$Hre, train\$Lsa | | | | |
| Number of observations: 17,847 | | | | |
| tr.train | low | high | Sum | |
| | low | 2207 | 17847 | |
| | high | 0 | 0 | |
| | Sum | 15640 | 2207 17847 | |
| > sum(tree2) | | | | |
| [1] 17847 | | | | |
| > diag(tree2) | | | | |
| | low | high | | |
| | 15640 | 2207 | | |
| > sum(diag(tree2))/sum(tree2) | | | | |
| # Classification rate | | | | |
| [1] 0.8763378 | | | | |
| Test data | | | | |
| Response: test \$ Problematic Smartphone Use1 | | | | |
| Inputs: test\$InS , test\$LAS, test\$Com, test\$Lse, test\$Tus, test\$Dca, test\$Hre, test\$Lsa | | | | |
| Number of observations: 7,618 | | | | |
| te.test | low | high | Sum | |
| | low | 914 | 7618 | |
| | high | 0 | 0 | |
| | Sum | 6704 | 914 7618 | |
| > sum(tree5) | | | | |
| [1] 7618 | | | | |
| > diag(tree5) | | | | |
| | low | high | | |
| | 6704 | 914 | | |
| > sum(diag(tree5))/sum(tree5) | | | | |
| # Classification rate | | | | |
| [1] 0.880021 | | | | |
| Dca | : Digital Capacity | | InS | : Information seeking |
| Hre | : Human Relationship | | LAS | : Leisure Activity Seeking |
| Lsa | : Life Satisfaction | | Com | : Communication |
| tr | : train | | | |
| te | : TEST | | | |
| | | | Tus | : Times of Usage |

2. Problematic Smartphone Use level of Random Forest Technique

스마트폰 과의존에 대한 RF기법 분석결과 Table 2. 와 같다. 랜덤포레스트(RF)로 스마트폰 과의존 분류를 하기 위하여 전체 데이터를 학습용 데이터와 검증용 데이터로 구분하였다. 결측값을 제외한 전체 데이터 개수는 21,050개였으며, 이중 학습용 데이터는 전체 데이터의 2/3 (14,033개), 검증용 데이터는 1/3(7,017개)로 구분하였다.

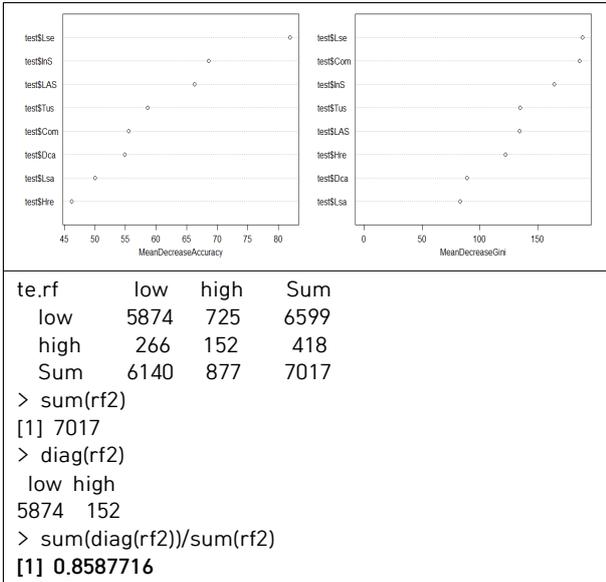
스마트폰 과의존에 대한 학습용 데이터의 정분류율은 86.62%(0.8662439)로 나타났다. 변수들의 분류율에 영향을 미치는 중요도는 정확도계수와 지니계수를 살펴보면 알 수 있다. 학습용 데이터의 정확도계수에서 중요도는 생활서비스형(110.84), 정보검색형(98.19), 여가추구형(95.17), 커뮤니케이션형(93.71) 순으로 나타났다. 학습용 데이터의 지니계수에서 중요도는 생활서비스형(332.49), 정보검색형(305.13), 커뮤니케이션형(290.70), 스마트폰 이용시간(232.68) 순으로

나타났다. 스마트폰 과의존에 대한 분류율에 영향을 미치는 중요 변수는 생활서비스형, 정보검색형, 커뮤니케이션형, 여가추구형, 스마트폰 이용시간 등으로 분석되었다.

스마트폰 과의존에 대한 검증용 데이터의 정분류율은 85.87%(0.8587716)로 나타났다. 검증용 데이터의 정확도계수에서 중요도는 생활서비스형(81.92), 정보검색형(68.55), 여가추구형(66.30), 스마트폰 이용시간(58.66) 순으로 나타났다. 검증용 데이터의 지니계수에서 중요도는 생활서비스형(189.19), 커뮤니케이션형(186.55), 정보검색형(164.67), 스마트폰 이용시간(135.20) 순으로 나타났다. 스마트폰 과의존에 대한 분류율에 영향을 미치는 변수로 생활서비스형, 정보검색형, 커뮤니케이션형, 여가추구형, 스마트폰 이용시간 등이 중요함을 알 수 있었다.

Table 2. Analysis Result of Random Forest

| Train data | | | | | |
|-------------------------------------|----------|-----------|------------|-----------|--|
| > round(importance(rf_out2), 2) | | | | | |
| | Low | High | MDA | MDG | |
| train\$InS | 83.44793 | 35.76456 | 98.19206 | 305.1251 | |
| train\$LAS | 73.30983 | 64.62912 | 95.17049 | 179.4788 | |
| train\$Com | 86.44342 | 20.2293 | 93.70996 | 290.6993 | |
| train\$Lse | 98.74434 | 31.57962 | 110.84351 | 332.4883 | |
| train\$Tus | 47.30518 | 67.78466 | 71.55834 | 232.6756 | |
| train\$Dca | 65.37091 | 48.04987 | 78.87015 | 161.9762 | |
| train\$Hre | 71.40857 | 28.70044 | 81.65945 | 185.6069 | |
| train\$Lsa | 79.33765 | 25.24759 | 85.88680 | 159.0187 | |
| | | | | | |
| tr.rf | low | high | Sum | | |
| | low | 1436 | 13336 | | |
| | high | 441 | 256 697 | | |
| | Sum | 12341 | 1692 14033 | | |
| > sum(rf1) | | | | | |
| [1] 14033 | | | | | |
| > diag(rf1) | | | | | |
| | low | high | | | |
| | 11900 | 256 | | | |
| > sum(diag(rf1))/sum(rf1) | | | | | |
| [1] 0.8662439 | | | | | |
| Test data | | | | | |
| > randomForest::importance(rf_out3) | | | | | |
| | Low | High | MDA | MDG | |
| test\$InS | 61.69320 | 20.908419 | 68.55639 | 164.67900 | |
| test\$LAS | 54.97530 | 39.401529 | 66.30111 | 134.77853 | |
| test\$Com | 51.31996 | 13.723844 | 55.50956 | 186.55679 | |
| test\$Lse | 79.78061 | 9.384802 | 81.92274 | 189.19378 | |
| test\$Tus | 46.21512 | 41.525599 | 58.66201 | 135.20163 | |
| test\$Dca | 41.38491 | 39.264985 | 54.88595 | 89.16996 | |
| test\$Hre | 42.39026 | 13.506215 | 46.22878 | 122.29808 | |
| test\$Lsa | 42.05669 | 26.318761 | 50.05771 | 83.23415 | |



MDA: MeanDecreaseAccuracy
MDG: MeanDecreaseGini

3. Problematic Smartphone Use level of Support Vector Machine Technique

스마트폰 과의존에 대한 SVM기법 적용을 위해 데이터를 학습용과 검증용 데이터로 구분하였다. 전체 데이터는 25,465이며, 학습용 데이터는 70%(14,817), 검증용 데이터는 30%(6,233)로 구분하였다.

SVM 분석결과는 Table 3과 같다. 학습용 데이터는 스마트폰 과의존에 대한 정분류율은 87.87% (0.8787204)로 나타났고, 검증용 데이터의 정분류율은 87.77%(0.8777475)로 나타났다. 학습용 데이터와 검증용 데이터의 분류율의 차이가 유사하게 나타나 과적합 문제는 없었다.

4. Comparison of Classification Analysis for Problematic Smartphone Use Level

DT, RF, SVM 기법에 따라 학습용 데이터와 검증용 데이터로 구분하여 정분류율을 확인한 결과 Table 4. 와 같이 나타났다.

3가지 기법 모두 정분류율과 과적합의 문제가 발생하지 않았으며 86%이상 분류율을 보였다. 그중 SVM 기법은 학습용 데이터로 검증용 데이터 사이에 0.1로 가장 작게 차이를 나타냈고, 분류율도 87.87%로 가장 높았다.

Table 3. Analysis Result of Support Vector Machine

| Train data | | | |
|---|-----------|------|-------|
| table(pred, train\$problematic smartphone use1) | | | |
| pred | low | high | |
| low | 13009 | 1796 | |
| high | 1 | 11 | |
| > # Classification rate (%) | | | |
| > tr.svm <- predict(svm.model) | | | |
| > svm2 <- table(tr.svm, train\$problematic smartphone use1) | | | |
| > addmargins(svm2) | | | |
| tr.svm | low | high | Sum |
| low | 13009 | 1796 | 14805 |
| high | 1 | 11 | 12 |
| Sum | 13010 | 1807 | 14817 |
| > sum(svm2) | | | |
| [1] | 14817 | | |
| > diag(svm2) | | | |
| low | high | | |
| 13009 | 11 | | |
| > sum(diag(svm2))/sum(svm2) | | | |
| [1] | 0.8787204 | | |
| Test data | | | |
| te.svm | low | high | Sum |
| low | 5471 | 762 | 6233 |
| high | 0 | 0 | 0 |
| Sum | 5471 | 762 | 6233 |
| > sum(svm3) | | | |
| [1] | 6233 | | |
| > diag(svm3) | | | |
| low | high | | |
| 5471 | 0 | | |
| > sum(diag(svm3))/sum(svm3) | | | |
| [1] | 0.8777475 | | |

Table 4. Comparison of Classification Analysis Results

| Classification Analysis Technique | DT | | RF | | SVM | |
|-----------------------------------|-------|-------|-------|-------|-------|-------|
| | train | test | train | test | train | test |
| Classification rate (%) | 87.63 | 88.00 | 86.62 | 85.87 | 87.87 | 87.77 |

V. Conclusions

본 연구는 스마트폰 과의존을 진단하고 예측하기 위하여 인공지능 방법인 의사결정트리, 랜덤포레스트, 서포트 벡터머신 기법 중 가장 정확한 분류를 나타내는 분석방법이 무엇인지 파악하고자 시도되었다. 또한 스마트폰 이용 형태 및 개인적 특성변수 중에서 분류율에 영향을 미치는 중요 변수가 무엇인지를 규명하였다.

분석에 사용한 데이터는 한국정보화진흥원에서 제공한 ‘2018 스마트폰 과의존 실태조사’에 응답한 총 25,465명의 자료였다.

연구 결과, 3가지 분류분석 기법은 정분류율이 유사하게 나타났으며, 모델에 대한 과적합 문제가 발생되지 않았다. 3가지 분류분석 방법 중 서포트벡터머신의 분류율이 가장 높게 나타났고, 다음으로 의사결정트리 기법, 랜덤포레스트 기법 순이었다. 스마트폰 이용 유형 중 분류율에 영향을 미치는 상위 3개 변수는 생활서비스형, 정보검색형, 여가추구형이었다.

연구결과를 토대로 제언 및 정책적 시사점을 제시하고자 한다.

첫째, 가장 높은 분류율을 나타낸 서포트벡터머신을 활용하여 스마트폰 과의존 예측모형을 개발하는 연구를 제안한다.

둘째, 스마트폰 과의존 예방교육을 실시할 때 단순 강의식 집단교육은 지양하고, 스마트폰 과의존 성향이 있는 사용자를 예측하여 맞춤형 교육프로그램을 적용하기를 권장한다.

REFERENCES

[1] Yonhap News, "Korea's smartphone holding rate is the world's number one". <https://www.yna.co.kr/view>, 2019.5.17.

[2] Bian, M. & Leung, L., "Linking loneliness, shyness, smartphone addiction symptoms, and patterns of smartphone use to social capital", *Social Science Computer Review*, Vol 33, No 1, pp. 61-79, Dec. 2015. DOI: <https://doi.org/10.1177/0894439314528779>

[3] Newman, D. B., Tay, L., & Diener, Ed, "Leisure and Subjective Well-Being", *Journal of Happiness Studies*, Vol 15, No 3, pp. 555-578, April, 2014. DOI: <https://link.springer.com/article/10.1007/s10902-013-9435-x>

[4] M. Kim, Y. Kang, D. Jung, & G. Lee, "Older Adults' Smart Phone Use and Access to Health Information", *Journal of Qualitative Research*, Vol. 14, No. 1, pp.13 -22, April, 2013. DOI: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3836587/>

[5] S. Y. Han, E. J. Ma, S. G. Choi & D. S. Hong, "Analysis on the Impact of Using Smartphone to Subjective Quality of Life", *The Korean Association for Information Society*, Vol. 20, pp. 49-84, Dec. 2011. DOI: <https://www.kci.go.kr/kciportal/ci/sereSearch/ciSereArtiView.kci?sereArticleSearchBean.artId=ART001595366>

[6] Hribernik, J. & Mussap, A. J.. "Research Note: Leisure Satisfaction and Subjective Wellbeing", *Annals of Leisure Research*, Vol 13, No 4, pp. 701-708, Mar. 2010. <https://doi.org/10.1080/11745398.2010.9686871>

[7] National Information Society Agency, "2018 The survey on internet overdependence", <http://www.nia.or.kr>

[8] <http://www.yeslaw.com/>

[9] S. M. Bae, & Y. S. Koh, "An Exploratory Study for Influential Factors of Smartphone Over-dependence in Elder and Older Group", *Journal of Digital Convert- gence*, Vol. 15, No. 7, pp. 111-121, Jul. 2017. <https://doi.org/10.14400/JDC.2017.15.7.111>

[10] S. M, Bae, "Influence of Smartphone usage Types and Excessive Expectation for Smartphone on Adolescents' Smartphone Overdependence," *Information Poli -cy*, Vol. 25, No. 4, pp. 65 - 83, Dec. 2018. <https://doi.org/10.22693/NIAIP.2018.25.4.065>

[11] Hyun-Jung Lee & Hyun-Soon Yu, "The Impact on University Students' Smartphone Addiction Towards the Intention and the Degree of Sport Participation, Leisure Satisfaction, and Life Satisfaction", *Journal of Korean Physical Education Association for Girls and Women*, Vol. 33, No. 2, pp. 57-70, June 2019. DOI: http://www.riss.or.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&control_no=59a84a77b51b1e954884a65323211ff0

[12] D. S. Lee, "Comprehension of Bigdata and R" *Free Academy-Press*, pp. 157-235. 2019.

Authors



Yu Jeong Kim received the B.S., M.S. and Ph.D. degrees in Nursing Science from Cheonam University, Korea, in 1995, 1998 and 2006, respectively. Dr. Kim worked for 8 years as a nurse at Chonnam National University

Hospital. She is a Professor in the Department of Nursing Science at Chosun Nursing College from 2015. She is interested in Big data, AI and Serious Game.



Dong Su Lee received the Ph.D. degrees in Administration from Chosun University, Korea, in 2005. respectively. Dr. Lee is representing the Korea Big Data Agency. He is currently working in the Department of computer science

and statistics at Chosun University. He is interested public big data, artificial intelligence, and deep learning.