

기계학습을 이용한 AUDIT-K 기반의 알코올 사용 장애 분류

김두언¹, 권기혁², 김남재³, 이동학⁴, 김종완^{5*}

¹삼육대학교 상담심리학과 학부생

²삼육대학교 인공지능융합학부 학부생

³(주)글로벌비즈텍 대표

^{4,5}삼육대학교 SW융합교육원 교수

endjs0623@gmail.com, 7lhyuck@gmail.com, knj7360@gbt2019.com, adhlee@syu.ac.kr, kimj@syu.ac.kr

Classification of alcohol use disorders based on AUDIT-K using machine learning

Du-Eon Kim¹, Ki-hyeok Kwon², Nam Jae Kim³, Dong-Hak Lee⁴, Jongwan Kim⁵

¹Dept. of Counseling Psychology, Sahmyook University

²Dept. of Artificial Intelligence Convergence, Sahmyook University

³Global Biz Tech Co., Ltd.

^{4,5}Software Convergence Education Center, Sahmyook University

요 약

본 연구는 알코올 사용 장애 식별 도구인 AUDIT-K를 기반으로 위험 음주자를 식별하는 인공지능 모델을 제안한다. 음주로 인한 사고는 점차 줄어들었지만, 재범의 수는 과거에 비해 큰 폭으로 증가했다. 위험 음주자가 자신의 음주 상태를 알 수 있게 해준다면, 음주 수준에 따라 대상자를 신속하게 분류하는 인공지능 모델을 통해 위험 음주자나 알코올 중독자에 대한 개입이나 치료를 더 신속히 하여 위험 음주로 인한 문제 발생을 예방하는 것에 도움이 된다. 이 모델은 사용자의 점수에 따라 위험 음주자를 분류하는 목적으로 제작되었다. 본 모델은 위험 음주자를 구분함으로써 자신의 음주 습관 위험성을 인지하고 알코올 중독 예방에 도움이 될 것이다.

1. 서론

알코올 사용은 전 세계적으로 중요한 공중 보건 문제로 자리 잡고 있으며, 과도한 음주는 다양한 신체적, 정신적 건강 문제를 초래할 수 있다. 위험 음주는 200개 이상의 질병과 손상의 주요 원인이며, 2016년 기준 전세계 사망의 5.3%, 질병 부담의 5.1%를 차지하며 인류 건강에 큰 부담을 주는 위험 요인인 것으로 확인되었다.[1] 이는 개인의 건강을 심각하게 해치고, 사회적, 경제적 비용을 증가 시키며, 가족 및 지역사회에 부정적인 영향을 미친다. 음주는 폭력성에도 영향을 미치는데, 한국 사회에서는 음주가 반 사회적 행동을 할 수 있게 하는 도구로 사용될 수도 있다.[2] 이에 따라 알코올 사용 장애의 조기 발견과 효과적인 중재가 중요한 과제로 대두 되고 있다. 하지만 알코올 사용의 중재를 위해서는 음주자가 스스로 자신의 음주 행위가 어느 수준인지 아는 것이 중요하다. 그리고 음주자 스스로가 자신의 음주 패턴을 스스로 깨닫게 돕는 것이 필요하다. 머신 러닝은 대규모 데이터 분석과 패턴 인식에 강점을 지니고 있어, 알코올 사용 장애의 자동화된 선별 및 예측에 유용하게 활용될 수 있다. 본 연구에서는 로지스틱 회귀 방식의 머신 러닝 알고리즘의

성능을 분석하여, 가장 효과적인 방법을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 먼저, AUDIT-K 기계학습 모델을 설명한다. 이후, 머신러닝 알고리즘의 적용 결과를 분석하고, 이를 바탕으로 논의 및 결론을 제시한다. 본 연구를 통해 AUDIT-K 척도를 기반으로 한 자동화된 위험 음주자 선별 시스템의 가능성을 모색하고, 실무 적용 방안을 제안하고자 한다.

본 연구에서는 기계학습 기법을 사용해 음주 위험 수준에 따라 위험 음주자 혹은 알코올 중독자를 빠르게 선별하여 중독의 예방이나 치료의 조기 개입을 위한 모델을 구축하고자 한다.

2. 위험 음주자 선별 척도 (AUDIT-K)

AUDIT(Alcohol Use Disorders Identification Test)는 세계보건기구(WHO)에서 1989년에 개발한 도구로, 알코올 사용 장애를 선별하는 데 널리 사용되고 있다. AUDIT-K는 AUDIT(Alcohol Use Disorder Identification Test) 척도(Babor et al., 2001)를 최은진 외(2006)가 사용한 것으로 문제 음주와 알코올 의존에 관한 사항에 관한 10개의 문항으로 구성되어 있다. 한국어로 번안하여 신뢰성과 타당성을 확보한 척도로, 국내에서도 알코올 사용 패턴을 평가하는 데 효과적으로 사용되고 있다.[3]

* 교신저자(Corresponding Author)

<표1> AUDIT-K 척도 문항

문항	0점	1점	2점	3점	4점
1 얼마나 자주 술을 마십니까?	전혀 안 마심	월 1회 미만	월 2~4회	주 2~3회	주 4회
2 술을 마시는 날은 보통 몇 잔을 마십니까?	1~2잔	3~4잔	5~6잔	7~9잔	10잔 이상
3 한번 술과식에서 6잔 (또는 맥주 2천cc) 이상을 마시는 횟수는 몇 번입니까?	없음	월 1회 미만	월 1회	주 1회	거의 매일
...	...	...	...	...	...
9 당신의 음주로 인해 자신이나 다른 사람이 다친 적이 있습니까?	없음	-	있지만 지난 1년 동안 없었다.	-	지난 1년 동안 있었다.
10 친척이나 친구, 의사가 당신의 음주에 대해 걱정을 하거나 술을 끊으라고 권고를 한 적이 있습니까?	없음	-	있지만 지난 1년 동안 없었다.	-	지난 1년 동안 있었다.

3. 실험 환경

학습모델은 python 3.11에서 Scikit-learn의 LogisticRegression이다. Scikit-learn은 파이썬에서 사용되는 오픈소스 머신러닝 라이브러리다. 본 연구는 Scikit-learn에서 제공하는 LogisticRegression() 클래스의 fit() 메서드와 predict() 메서드를 사용해 모델을 학습시킨다.

4. 데이터 분류

본 논문은 사용자의 음주 상태를 식별하기 위해 로지스틱 회귀(Logistic Regression)를 사용하는 예측 모델을 제안한다. 입력 데이터는 10개 항목의 설문지다. 1번부터 8번까지의 항목은 Likert 5점 척도로 0점부터 4점까지 배점하고 9번과 10번 항목은 Likert 3점 척도를 사용하고 각 0점, 2점, 4점으로 배점 한다. label은 0, 1, 2로 각각 정상 음주, 위험 음주, 알코올 사용 장애로 구분된다. 설문지의 라벨은 총점수로 결정되는데 남성은 수식 (1), 여성은 수식 (2)를 적용한다.

<표2> AUDIT-K 음주 수준 분류표

점수		내 용
남성	여성	
0~9점	0~5점	정상음주
10~19점	6~9점	위험음주 (단기개입 및 모니터링 필요)
20점 이상	10점 이상	알코올 사용장애 (사용장애 추정군, 치료적 개입 필요)

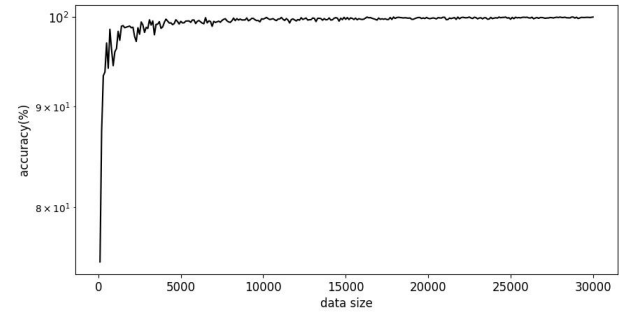
$$label = \begin{cases} 0 & \text{if } \text{sum} \leq 9 \\ 1 & \text{if } 10 \leq \text{sum} \leq 19 \\ 2 & \text{if } \text{sum} \geq 20 \end{cases} \quad (1)$$

$$label = \begin{cases} 0 & \text{if } \text{sum} \leq 9 \\ 1 & \text{if } 10 \leq \text{sum} \leq 19 \\ 2 & \text{if } \text{sum} \geq 20 \end{cases} \quad (2)$$

5. 모델 평가

본 모델의 정확도를 평가하기 위해 데이터를 학습시키고 정확도를 평가했다. 평가를 하기 위해서 데이터의 수를 1,000개 단위씩 증가시키며 30,000개까지 학습시키며 정확도를 평가했다. 데이터의 양이 5,000개 이하일 때는 정확도가 불안정하지만 5,000개 이상일 때

부터 정확도가 100%로 나온다. 정확도는 측정하거나 계산된 양이 실제값과 얼마나 가까운지 나타내는 기준이다. 데이터의 양에 따라 이 모델이 얼마나 정확한지 나타내는 것이다. 본 논문에서 제시한 모델은 (그림 1)과 같이 데이터 개수를 5,000개 이상 설정하였을 시, 100% 성능을 나타내며, 오차 없이 설문 응답자를 올바른 이용자 군으로 분류한다.



(그림 1) 데이터 개수에 따른 테스트 정확도

6. 결론

본 연구에서는 AUDIT-K 척도를 기반으로 위험 음주자를 효과적으로 분류하기 위한 인공지능 분류 모델을 개발하였다. 제안된 모델은 높은 정확도와 예측 성능을 보여, 기존의 수동 평가 방식에 비해 신속하고 객관적인 평가 도구로서의 가능성을 입증했다. 특히, 모델은 공간의 제약을 받지 않고 시행하여 사용자의 음주 상태 식별이 가능하다는 점에서 임상적 활용도가 높다. 이는 음주 문제의 조기 발견과 중재에 있어 중요한 도구가 될 수 있음을 시사한다. 따라서, 본 연구의 모델은 음주 문제를 효과적으로 관리하기 위한 보조 수단으로 사용될 수 있으며, 특히 임상 환경에서 신속한 스크리닝 도구로 활용될 수 있는 잠재력이 있다. 추후 연구에서는 실제 임상 환경에서의 적용 가능성을 평가하기 위해 추가적인 검증 연구가 요구된다. 이러한 노력을 통해, 인공지능 기반 음주 위험 예측 모델이 실제 건강 관리와 중재에 중요한 역할을 시행할 수 있을 것으로 기대된다.

감사의 글

본 연구는 2021년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (2021-0-01440).

참고문헌

- [1] 이은숙, and 조혜정. "성인의 연령층별 음주 상태, 폭음 및 위험음주와 주관적 건강인지와와의 연관성." 대한보건연구 45.2 (2019): 69-82.
- [2] 서경현. "음주 동기, 문제음주 및 음주 가족력과 대학생의 데이트 폭력." 한국심리학회지: 문화 및 사회문제 9.2 (2003): 61-78.
- [3] Park, Byung-Sun, & Sun-young LEE. (2023). Validating the Factor Structure and Measurement Invariance of the AUDIT-K Scale: Focusing on the Use of the RMSEAD Index for Nested Model Comparison. Health and Social Welfare Review, 43(4), 211 - 226. <https://doi.org/10.15709/HSWR.2023.43.4.211>