

기계학습 기반의 도박중독 수준 예측 모델

이서진¹, 최지웅², 홍승현³, 김남재⁴, 김종완^{5*}

^{1,3}삼육대학교 상담심리학과 학부생

²삼육대학교 인공지능 융합학부 학부생

⁴(주)글로벌비즈텍 대표

⁵삼육대학교 SW융합교육원 교수(주)글로벌비즈텍 대표

shine36sfg@naver.com, iostream1204@gmail.com, hsssh02@naver.com,

knj7360@gbt2019.com, kimj@syu.ac.kr

Machine Learning-Based Gambling Addiction Level Prediction Model

Seo-Jin Lee¹, Ji-Woong Choi², Seung-Hyeon Hong³, Nam-Jae Kim⁴, Jong-wan Kim⁵

^{1,3}Dept. of Counseling Psychology, Sahmyook University

²Dept. of Artificial Intelligence Convergence, Sahmyook University

⁴Global Biz Tech Co., Ltd.

⁵Software Convergence Education Center, Sahmyook University

요약

본 논문은 도박중독 수준에 따라 중독자를 선별하는 선형회귀 기반의 기계학습 모델을 제안한다. 도박 인구는 점차 증가하고, 청소년들도 도박을 경험하며 문제가 심각하다. 도박중독 유행률을 조사하는 것은 향후 발생할 도박으로 인한 문제를 해결하는 것에 도움이 된다. 제안 모델은 한국판 도박 문제 선별척도(K-Canadian Problem Gambling Index) 데이터를 기반으로 학습하였고, 개인의 중독 수준에 따라 도박중독자를 분류한다. 본 모델은 도박중독 수준에 따라 신속하고 정확하게 도박중독자를 구분함으로써 도박중독 치료 연계와 함께 도박중독률을 낮추는 데 기여할 것이다.

1. 서론

도박중독이란 알코올 의존과 같은 물질 의존이나 중독 개념을 도박에 적용한 것으로, 도박 행동에 중독이 된 것이다.[1, 2] 한국의 도박중독 추정 인구 비율은 5.5%, 청소년 중 돈내기 경험을 해본 비율은 25.8%이다.[3] 최근에는 스마트폰과 인터넷으로도 도박을 접할 수 있어 쉽게 중독될 수 있다. 또한 도박자로 인해 사기, 폭행 등 다양한 사회적 문제가 발생할 수 있기에 빠른 치료적 개입이 필요하다.

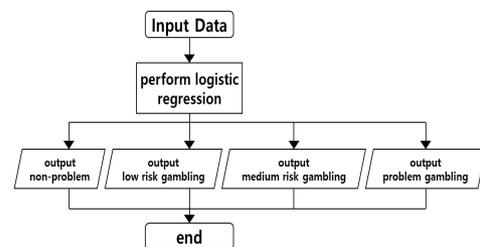
현재 한국은 도박중독에 대처하기 위해 중독관리통합지원 센터를 운영하고 있다. 하지만 중독자보다 센터의 수가 절대적으로 부족하고, 숨어있는 온라인 도박중독까지 대처하는 것에는 무리가 있다.[4] 도박이 지니는 폐해를 예방하고 해결하는 데 필요한 대책을 세우기 위해서 가장 우선으로 해야 할 일은 실태를 체계적으로 평가하는 일이다.[5]

개인의 도박중독 수준은 한국판 도박 문제 선별척도를 사용하여 지역별 중독관리 통합지원 센터, 한국도박문제관리센터에서 평가해 왔다. 그러나 KCPGI는 조사 대상이 많을 때 도박 수준을 분류하는 데 있어 인력과 시간, 비용이 많이 들어 신속한 결과 도출이 늦어진다. 본 논문에서는 KCPGI를 이용한 도박중독 수준을 신속하게 예측하기 위해 기계

학습을 통한 선형회귀 기반의 인공지능 모델을 제안한다. 본 모델은 인공지능을 사용해 도박중독 수준에 따라 기존의 방법보다 신속하고 정확하게 도박중독자를 구분함으로써 도박중독의 위험성을 알리고, 치료에 빠르게 연계할 수 있으며 더 나아가 도박중독률을 낮출 수 있을 것이다.

2. 도박중독 수준 예측 모델

본 논문은 도박 문제 선별척도 설문 데이터를 바탕으로 도박중독 수준을 예측하는 로지스틱 회귀모델을 생성했다. 모델의 순서도는 그림 1과 같다.



(그림 1) 도박중독 수준 예측 모델 순서도

한국판 도박 문제 선별척도[6]의 9개 항목의 설문을 입력 데이터로 사용한다. 각 항목은 Likert 4점 척도로 평가된 점수이며, 총점은 0점부터 27점까지

* 교신저자(Corresponding Author)

생성될 수 있다. 목표 라벨은 개인이 속한 사용자 군을 나타내고, 0, 1, 2, 3의 세 가지 범주이다. 0 라벨은 비 문제 도박(0점)에 해당하며, 1 라벨은 저위험 도박(1~2점), 2 라벨은 중위험 도박(3~7점), 3 라벨은 문제도박(8점 이상)에 해당한다.

본 모델의 알고리즘은 표 1과 같다.

<표 1> 도박중독 수준 예측 모델의 알고리즘

Algorithm 1 Training and Evaluating Logistic Regression Model

```

1: Q = 9
2: N = 10000
3: Output: Accuracy score
4: Initialize random seed to 42
5: Generate random integer matrix X of size (10000, 9)
   with values in [1, 4]
6: Compute total scores val = sum(X, axis=1)
7: Convert total scores to labels y using Equation (1)
8: Split X and y into training and testing sets (80% train,
   20% test)
9: Train Logistic Regression model on training set
10: Predict on test set
11: Compute accuracy of predictions
12: return accuracy score

```

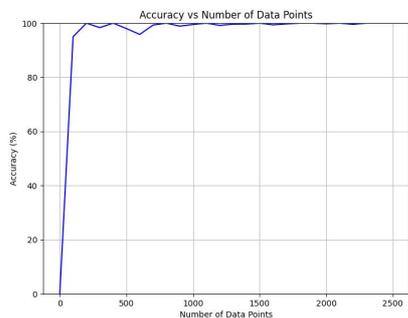
4. 실험

4.1 실험 환경

제안한 도박중독 수준 예측 모델의 실험환경은 다음과 같다. Software/Framework는 python 3.11, scikit-learn 1.5.0을 사용하였고, 모델은 Logistic Regression을 사용하였다. 모델의 데이터 크기는 100 ~ 3,000개이다.

4.2 실험 결과

본 논문에서는 학습 및 검증에 사용된 데이터의 수가 200개 이상으로 증가함에 따라, 테스트 데이터의 양 역시 비례하여 증가하였음을 관찰하였다. 정확도는 전체 예측 중에서 올바르게 예측된 사례의 비율을 나타내며, 이는 모델의 전반적인 성능을 평가하는 중요한 지표이다.



(그림 2) 데이터 개수에 따른 모델의 정확도

본 논문에서 제안한 모델의 정확도는 그림 2와 같다. 모델의 정확도는 600개에서 최저점을 기록한 후, 모델이 패턴을 완벽하게 이해하게 되었다. 2,500개

이상의 데이터로부터 안정적인 성능을 나타내고 100%의 정확도를 기록하였다. 이러한 결과는 검증 데이터가 적을 경우, 모델 학습의 완성도를 판별하기 어려운 점을 뒷받침한다.

본 논문에서 제시한 모델은 표 2와 같이 데이터 개수를 2,500개로 설정하였을 시, 100% 성능을 나타내며, 오차 없이 설문 응답자를 올바른 사용자 군으로 분류한다. 따라서 본 모델은 사용자의 도박중독 수준을 분류하는 데 적절하다.

<표 2> 모델 성능

Metric	Accuracy
Test Accuracy	100%
Precision	100%
Recall	100%
F1 Score	100%

5. 결론

본 논문에서는 로지스틱 회귀를 사용한 도박중독 수준 예측 모델을 제안하였다. 제안된 방법으로 도박중독 수준을 분류한 결과 100%의 정확도를 확인할 수 있었다. 기계학습을 이용해 높은 정확도를 보이기에 기존의 방법보다 도박중독의 유병률을 빠르게 조사할 수 있고, 도박중독 문제를 감소시키기 위해 사용할 수 있다. 현재 수많은 도박중독자를 선별하고 관리하는 것에는 무리가 많아 이를 자동화하고 빠르게 선별할 수 있는 도박중독 수준 예측 모델을 사용하면 도박중독자의 치료율도 증가할 수 있다.

본 논문에서 제시한 모델은 다량의 도박중독 수준 선별 데이터를 빠르고 정확하게 분류하는 장점이 있기에, 향후 다양한 중독 분야로 확장할 것이다.

감사의 글

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

참고문헌

- [1] 김교현. 도박행동의 자기조절 모형: 상식모형의 확장. 한국심리학회지: 건강, 11, 243-274. (2006).
- [2] 김교현. 한국 도박중독 문제의 책임과 대처. 한국심리학회지: 건강, 14, 27-39. (2009).
- [3] 전주영. (2024.3.1). 국민 5.5% 도박중독 청소년 등 예방 집중, 동아일보.
- [4] 차명희. 한국과 일본의 도박중독 대처에 대한 비교 연구. 교정담론, 16(3), 281-308. (2022)
- [5] 강성균, 김교현, 이민규, 임지향. 도박중독의 측정: KNODS, KCPGI 및 KSOGS의 비교. 한국심리학회지: 건강, 15(3), 569-581. (2010).
- [6] 김아영, 차정은, 권선중, 이순목. CPGI의 한국판 제작 및 타당화. 한국심리학회지:일반, 30(4), 1011-1038. (2011).