

머신러닝 기반 청년 우울 분류 모델

이가영¹, 배준한², 이재린³, 김시영⁴, 김종완⁵^{1,4}삼육대학교 보건관리학과 학부생²삼육대학교 인공지능융합학부 학부생³삼육대학교 사회복지학과 학부생⁵평택대학교 스마트콘텐츠학과 교수

ga6772@naver.com, baezu2zu2@gmail.com, jaeriny@naver.com, sykim12277@naver.com, kimj@ptu.ac.kr

Machine Learning-Based Youth Depression Classification Model

Ga-young Lee¹, Jun-han Bae², Jae-Rin Lee³, Si-Young Kim⁴, Jongwan Kim⁵^{1,4}Dept. of Public Health, Sahmyook University²Division of Artificial Intelligence Convergence, Sahmyook University³Dept. of social welfare, Sahmyook University⁵Dept. of Smart Contents, Pyeongtaek University

요약

현대 사회에서 청년층의 우울 문제는 심각한 사회적 과제로 대두되고 있으며, 조기 선별과 관리의 필요성이 점점 커지고 있다. 그러나 기존의 설문조사에 대한 해석은 전문가 의존도가 높아 대규모 집단에 신속히 적용하기 어렵다는 한계가 있다. 본 연구는 이러한 한계를 극복하기 위해 한국형 우울 척도(national depression scale, NDS)와 개인적 특성 및 건강 이력 등 다차원적 변수를 통합하여 기계학습 기반 분류 모델을 제안하였다. 특히 k-최근접 이웃(KNN) 알고리즘을 적용하여 다양한 변수 간의 복합적 관계를 분석하고, 청년 우울 수준을 정상, 경증, 중등도, 중증으로 분류하였다. 실험 결과, 제안된 모델은 정밀도 98%의 분류 성능을 나타냈다. 본 연구는 다차원적 요인을 반영한 우울 분류 접근을 통해 정신건강 증진 및 조기 개입 전략 마련에 기여할 수 있을 것으로 기대된다.

1. 서론

현대 사회에서 청년층의 정신건강 문제는 지속적으로 대두되고 있다. 특히 정신건강 문제 중 우울 증상을 겪는 청년의 수는 계속해서 증가하는 추세이며 이는 우울증이 청년층의 주요 건강 문제로 자리 잡고 있음을 시사한다[1].

정신질환은 조기 발견 및 예방을 통해 만성 질환 예방, 사회적 기능 회복, 삶의 질 향상 및 정신증 발병 예방이 가능하므로 우울증 초기에 징후를 신속히 파악하고 사전에 관리하는 것은 보건학적으로 중요한 과제이다. 그러나 전통적인 설문 결과 해석은 전문가의 판단에 의존하는 경우가 많아, 시간과 비용이 소요되고 대규모 집단을 대상으로 신속한 분석을 수행하는 데 한계가 존재한다.

기계학습은 이러한 한계를 극복할 수 있는 효과적인 방법으로, 설문 데이터의 패턴을 학습하여 우울 수준을 자동으로 분류함으로써 기존의 수작업 해석 방식보다 신속하고 정확한 결과를 제공한다. 따라서 본 연구는 기계학습 기법을 활용하여 청년들의 우울 수준을 효율적으로 분류하는 방법을 제안한다. 특히 k-최근접 이웃(KNN) 알고리즘을 적용하여 다차원적 설문 데이터를 효과적으로 분석하고 그 유용성을 검증하고자 한다.

2. 관련 연구

2.1 우울 수준 분류

정신건강 설문 데이터를 활용한 기계학습 기반 우울 분류 연구는 꾸준히 이루어져 왔으나, 다차원적 변수를 동시에 반영한 시도는 드물다. 기존 연구들은 주로 BDI(Beck Depression Inventory), PHQ-9(Patient Health Questionnaire-9), HADS(Hospital Anxiety and Depression Scale)와 같은 단일 척도에 의존하거나 설문 점수만을 활용해 분류 성능을 검증하였다[2][3]. 일부에서는 여러 척도를 결합하여 성능 향상을 시도했으나, 여전히 개인적·사회적 요인을 충분히 반영하지 못하는 한계가 있었다. 특히 청년층 연구에서도 사회적 환경과 관계망이 우울에 중요한 영향을 미친다는 결과가 있음에도, 이를 통합한 인공지능의 분류 모델은 찾아보기 어렵다.

2.2 한국인 우울 척도 (NDS)

본 연구에서는 표 1과 같이 참여자들의 우울 수준을 국립정신건강센터에서 개발한 한국형 우울 척도(national depression scale, NDS)를 활용하여 측정한다[4]. NDS는 대한신경정신의학회의 인증을 받은 국립정신건강센터에서 공인된 도구로써 신뢰할 수 있는 정신건강 평가도구이다. 기계학습을 활용한

자동 판정 시스템은 결과 분석에 대한 시간을 단축하고, 인력 의존도를 낮추며, 대규모 사용자에게 효율적인 관리가 가능하다는 장점이 있다.

<표 1> NDS 설문지

번호	문항 내용	0	1	2	3
		전혀 그렇지 않다 (없음)	가끔 그렇다 (2일 이상)	자주 그렇다 (1주 이상)	거의 매일 그렇다 (거의 2주)
1	하루 종일 우울하다	0	1	2	3
...	(이하 생략)	...	...	...	...
12	하루를 생활하기가 버겁다	0	1	2	3

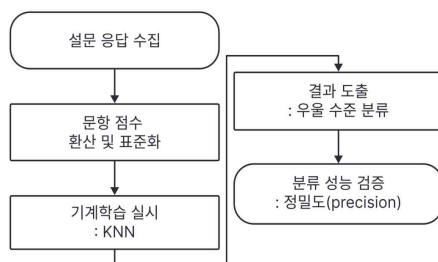
3. 우울 측정 모델

본 연구는 참여자의 설문조사 결과를 기반으로 기계학습 기술을 적용하여 우울 수준을 정상, 경증, 중등도, 중증의 네 가지 범주로 분류하고자 한다. 이를 위해 그림 1에 제시한 바와 같이 NDS 설문을 활용하여 청년 참여자들의 우울 관련 데이터를 수집하였다. 수집된 데이터는 각 문항 점수를 합산하여 절단점(cut-off score)에 따라 네 가지 클래스의 우울장애로 최종 판별한다.

본 연구는 이러한 우울 데이터 분석을 위한 주요 기법으로 KNN 알고리즘을 활용하였다. KNN은 데이터 간의 유사성(거리)을 기반으로 분류를 수행하여 높은 직관성을 지니며, 가장 가까운 이웃들의 다수결 원칙만으로 다중 클래스를 구분할 수 있어 구현 용이성이 높다. 더욱이, NDS 설문 점수 외에 정신질환 경험 유무, 약물 복용 경험 유무 등 성격이 상이한 다차원적 변수들을 통합 분석하는 데 유연하게 적용될 수 있다는 장점을 고려하여 본 연구의 분석 기법으로 최종 선정되었다.

모델의 성능 평가 지표로는 KNN이 데이터 불균형 상황에서 정상인을 우울장애로 판단할 경우 발생하는 불필요한 검사 비용을 최소화하기 위해 수식 (1)과 같이 정밀도(precision)를 통해 성능 검증을 실시하였다. TP(true positive)는 클래스로 분류한 것 중 올바른 것을 나타내고, FP(false positive)는 클래스로 분류한 것 중 올바르지 않은 것을 나타낸다. 본 연구에서는 각각의 클래스에 대해 정밀도를 구해 평균을 구했다.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$



(그림 1) 분류 모델

4. 실험

4.1 실험 환경

본 연구에서는 NDS 설문지를 활용하여 청년층의 우울 수준을 측정 및 분류하였다. 실제 데이터는 총 32개의 설문 응답을 온라인으로 수집하였다. 설문은 익명성을 보장한 상태에서 진행되었으며, 연구 목적에 동의한 응답자만 참여하였다. 모집은 온라인 커뮤니티 및 다양한 웹사이트를 통해 공개적으로 이루어졌다. 수집된 32개의 실제 데이터를 기반으로, 실제 설문 데이터의 평균을 중앙값으로 하는 표준 정규분포 기반 난수를 1,073개 생성하여 총 1,105개의 데이터를 사용하였다.

수집된 데이터는 각 설문 문항의 점수를 합산한 후, 절단점을 기준으로 참여자들의 우울 수준을 분류하였다.

4.2 KNN 기반 우울 분류 결과

본 연구에서는 NDS 응답 점수와 개인적 건강 이력 변수를 통합하여 KNN 알고리즘을 적용한 결과, 정밀도 98%를 기록하였다. 이는 KNN이 분류한 긍정 사례의 대부분이 실제 우울장애 여부와 일치함을 보여주는 것이다. 이러한 성과는 청년층의 우울 위험군을 조기에 식별하고 현장에서 신속히 적용할 수 있는 자동화된 평가도구로서의 활용 가능성을 제시한다. 특히 기존 연구들이 단일 척도나 제한된 문항에 의존했던 것과 달리, 본 연구는 다차원적 요인을 반영함으로써 우울 수준을 보다 정밀하게 분류할 수 있음을 입증하였다.

5. 결론

분석 결과, KNN은 98%의 정밀도를 기록하여 우울장애 집단을 높은 정확성으로 분류하는 것을 확인하였다. 이는 기계학습 기법이 청년 우울 데이터를 분류하는 과정에서 신뢰성 있는 방법으로 활용될 수 있음을 시사한다. 또한 불필요한 오판으로 인한 사회적·경제적 비용을 최소화하며, 신뢰성 높은 간이 선별 도구의 개발을 위한 기반을 마련했다는 실용적 의의를 지닌다.

본 연구는 기존의 단일 척도 기반 분석에서 벗어나 정신건강 경험, 약물 복용 경험 등 다차원적 변수를 KNN 알고리즘에 통합하여 우울 위험군 분류의 새로운 방법론적 가능성을 제시하였다.

참고문헌

- [1] 조재근, 강은빈, 안선영, 원영주, "지역사회요소와 청년의 우울 경험 간의 연관성", 대한보건연구, 제51권, 1호, 165~177, 2025.
- [2] Sheikh Afaan Farooq, Iqra Altaf Gillani, "Depression Prediction: A Clinical Questionnaire and Machine Learning Based Approach", 2022.
- [3] 박희승, 김경원, 문은수, 임현주, 서화규, 김경은, 강태우, "유방암 환자의 우울증 감지를 위한 자기평가 설문지를 이용한 기계 학습 모델 탐색", 2024.
- [4] 박주연, "한국인 정신건강 척도 사용자 지침서", 국립정신건강센터, 05~17, 2024.