

자살 충동 예측 프로젝트

2025. 12. 1

M Therapy

강혜영, 이기화, 임성현, 소현진(발표자)

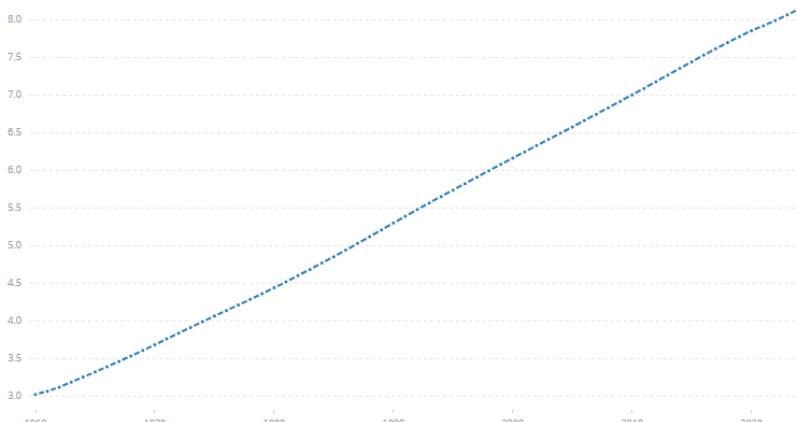
목 차

1. 문제 정의 & 분석 목표
2. 데이터 소개 & 위험 요인 탐색
3. 머신러닝 예측 모델 개발
4. 정책 개입 시나리오 & 보호 효과
5. 결론

1. 글로벌 트랜드

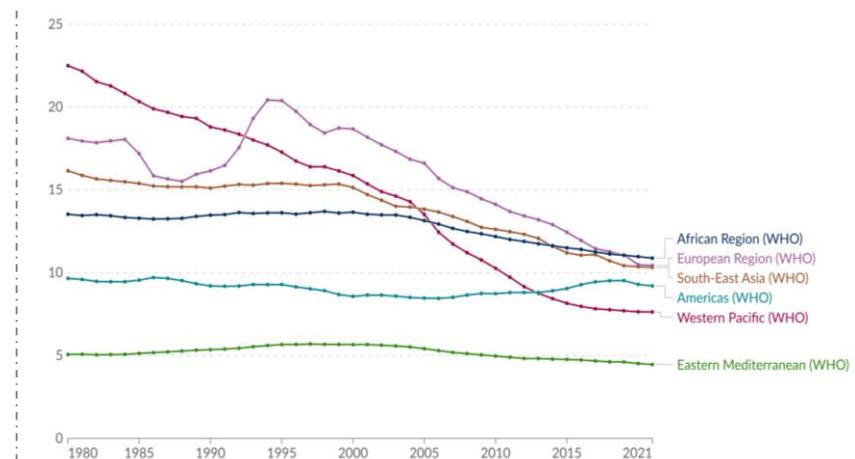
전 세계 자살율은 감소하지만 학생 자살률은 반대로 빠르게 증가

전 세계 인구수 81억



<fig_1> WORLD BANK GROUP | Data, Population, total
(source site : <https://data.worldbank.org/>)

표준 자살율 감소



< Fig. 1 > 자살 통계 (전세계 국가 / 1980 ~ 2021)
출처 "IHME, Global Burden of Disease (2024)"

청소년 자살율 증가

WHO Suicide worldwide in 2021:
global health estimates

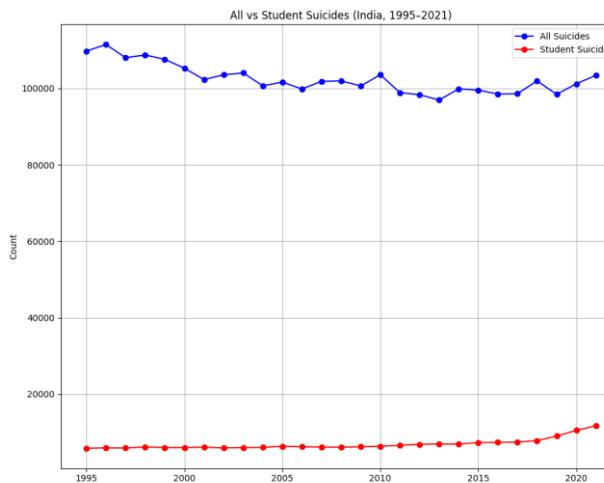
글로벌 트랜드 : 청소년 자살율 증가

- 청소년 자살률이 성인보다 2~3배 빠르게 증가하고 있습니다.
- WHO에 따르면 인도는 전체 자살 수에서 상위 3위를 기록했습니다.
- 전 세계적으로 청소년들의 정신적 스트레스가 증가하고 있습니다.

2_1. 인도에 집중하는 이유

인도는 세계 인구 1위 이자, 학생 자살 위기의 최전선

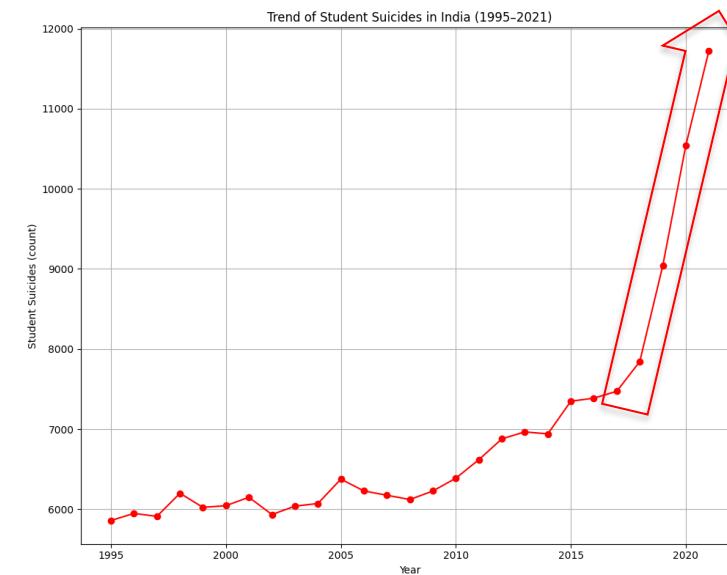
인구 1위, 인도 자살 현황



| # | Country | Population 2025 |
|---|---------|-----------------|
| 1 | India | 1,463,865,525 |

[https://www.worldometers.info/
population/most-populous-countries/](https://www.worldometers.info/population/most-populous-countries/)

인도 학생 자살 현황



프로젝트 목표

우울증 데이터셋 분석,
ML 모델링을 통해

인도 학생 자살 위험
조기 예측 및
개입 시나리오 제시

2_2. 인도에 집중하는 이유 : 신분제도 + 학업구조

신분제도 속 신분 상승의 유일한 기회는 고학력



Caste 기반 사회적 압력 + 경쟁 교육 + 경제 부담

(1) 신분제도와 학업 스트레스의 결합

‘시험 성적 = 사회 이동성(Social Mobility)’
불가촉천민(Dalit), OBC, 기타 하위 계층 학생:
→ 성적이 유일한 탈출 통로
→ 가족·사회로부터 과도한 ‘실패 불가’ 압박

(2) Caste Gap이 만든 정서적 고립감

학교 내 차별, 멘토 부족, 사회적 지지 부족
'Support Level 낮음 → Depression·Suicidal Thoughts 급증'

(3) "교육 도시"와 심리적 부담의 집중

IIT, NIT 진학 준비 핵심 지역(코타, 하이데라바드, 벵갈루루)
→ 학생 자살률이 극단적으로 높음

3. 인도 학생 위험 요인 : 구조적 압박

인도의 학업 압박 + 계급 구조 + 경제적 불평등이 학생을 구조적 고위험군으로 만듦

Caste : Class



Student Suicide Risk

- 학업 압박
IIT-NIT 입시, 코칭 센터 도시(Kota 등)에서의 극단 경쟁.
- 신분제도(Caste) + 계급사회(Class)
성적이 거의 유일한 사회 이동 통로.
- 경제적 불평등 + 교육비 부담
저소득층일수록 학업 실패 비용이 치명적.
- 지지 및 보호 체감의 부족
차별, 멘토 부재, 정신건강 인프라 부족.

인도 정신건강 데이터를 활용해

위험도를 예측하고

정책적 개입 시나리오 설계

4. 정신건강 데이터셋 소개

**약 14만 명을 대상으로 수집한 정신 건강 및 자살 관련 설문 데이터로
인구통계, 학업/직업, 스트레스, 생활 습관/만족도 등의 변수를 활용해
우울증(Depression, 0/1) 여부 예측 가능**

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|----|----|-----------|--------|-----|-------------|---------|---------------------|----------|-----------|------|------------|-----------------------|---------------------|-------------------|-----------------------------|------------|------------|------|---|
| 1 | Id | Name | Gender | Age | City | Working | Professor | Academic | Work_Pret | CGPA | Study_Sati | Job_Sati | Dietary_H | Degree | Have you Work/Stu_Financial | Family_His | Depression | | |
| 2 | 0 | Aaradhy | Female | 26 | Ludhiana | Working | Chef | | | 5 | | | 3 | More than Healthy | BHM | No | 1 | 2 No | 0 |
| 3 | 1 | Vivan | Male | 26 | Varanasi | Working | Teacher | | 4 | | | 3 | Less than Unhealthy | LLB | Yes | 7 | 3 No | 1 | |
| 4 | 2 | Vuvraj | Male | 33 | Vitthalpura | Student | | 5 | 8.97 | 2 | | 5-6 hours | Unhealthy | B.Pharm | Yes | 3 | 1 No | 1 | |
| 5 | 3 | Vuvraj | Male | 22 | Mumbai | Working | Teacher | | 5 | | | 1 | Less than Moderate | BBA | Yes | 10 | 1 Yes | 1 | |
| 6 | 4 | Rheja | Female | 30 | Kanpur | Working | Business Analyst | | 1 | | | 1.5-6 hours | Unhealthy | BBA | Yes | 9 | 4 Yes | 0 | |
| 7 | 5 | Vani | Female | 59 | Ahmedabad | Working | Financial Analyst | | 2 | | | 5.5-6 hours | Healthy | MCA | No | 7 | 5 No | 0 | |
| 8 | 6 | Ritika | Male | 47 | Thane | Working | Chemist | | 5 | | | 2.7-8 hours | Moderate | M | No | 6 | 2 No | 0 | |
| 9 | 7 | Rajveer | Male | 38 | Nashik | Working | Teacher | | 3 | | | 4.7-8 hours | Unhealthy | B.Pharm | Yes | 10 | 3 Yes | 0 | |
| 10 | 8 | Aishwarya | Female | 24 | Bangalore | Student | | 2 | 5.9 | 5 | | 5-6 hours | Moderate | BS | No | 3 | 2 Yes | 0 | |
| 11 | 9 | Rishabh | Female | 26 | Patna | Working | Electrical | | 4 | | | 1.5-2 hours | Unhealthy | BE | Yes | 7 | 2 Yes | 0 | |
| 12 | 10 | Worush | Male | 35 | Delhi | Working | Software Engineer | | 3 | | | 2.7-8 hours | Moderate | BSc | No | 6 | 4 Yes | 0 | |
| 13 | 11 | Aishwara | Female | 51 | Patna | Working | Teacher | | 1 | | | 2.7-8 hours | Moderate | B.Arch | No | 9 | 5 No | 0 | |
| 14 | 12 | Tolani | Male | 39 | Jalpur | Working | Data Scientist | | 2 | | | 5 More than Moderate | BCA | Yes | 1 | 3 No | 0 | | |
| 15 | 13 | Aadiya | Female | 29 | Pune | Working | Chef | | 4 | | | 4.7-8 hours | Unhealthy | BHM | No | 6 | 5 Yes | 0 | |
| 16 | 14 | Kiran | Male | 50 | Thane | Working | Plumber | | 1 | | | 1 More than Moderate | BE | Yes | 8 | 5 Yes | 0 | | |
| 17 | 15 | Aditi | Female | 23 | Patna | Working | Marketing Manager | | 2 | | | 2.5-6 hours | Healthy | M | Yes | 6 | 4 Yes | 1 | |
| 18 | 16 | Suhani | Female | 49 | Lucknow | Working | Teacher | | 5 | | | 3 More than Healthy | B.Ed | No | 2 | 1 No | 0 | | |
| 19 | 17 | Jiya | Female | 56 | Meenut | Working | Accountant | | 5 | | | 3.7-8 hours | Moderate | B.Com | No | 0 | 5 Yes | 0 | |
| 20 | 18 | Bharesh | Male | 50 | Agra | Working | Entrepreneur | | 5 | | | 2.7-8 hours | Moderate | BBA | Yes | 3 | 2 Yes | 0 | |
| 21 | 19 | Armaan | Male | 45 | Rajkot | Working | Teacher | | 2 | | | 4.5-6 hours | Moderate | MBA | Yes | 2 | 3 No | 0 | |
| 22 | 20 | Shahani | Female | 37 | Surat | Working | Teacher | | 3 | | | 1.7-8 hours | Healthy | B.Ed | No | 5 | 2 Yes | 0 | |
| 23 | 21 | Prachi | Female | 37 | Hyderabad | Working | Marketing Manager | | 2 | | | 7.5-8 hours | Unhealthy | BA | Yes | 6 | 4 Yes | 0 | |
| 24 | 22 | Abdullah | Male | 37 | Meenut | Working | Teacher | | 5 | | | 5.5-6 hours | Healthy | B.Arch | No | 9 | 2 Yes | 0 | |
| 25 | 23 | Aishwarya | Female | 46 | Hyderabad | Working | UX/UI Designer | | 2 | | | 3.5-6 hours | Healthy | BCA | No | 7 | 5 No | 0 | |
| 26 | 24 | Abhinav | Male | 39 | Pune | Working | Content Writer | | 5 | | | 2 More than Healthy | M.Com | Yes | 7 | 2 Yes | 0 | | |
| 27 | 25 | Siddhesh | Male | 38 | Thane | Working | Chef | | 2 | | | 1 More than Healthy | MHM | Yes | 2 | 2 No | 0 | | |
| 28 | 26 | Aditya | Male | 31 | Srinagar | Student | | 3 | 7.03 | 5 | | Less than Healthy | BA | No | 9 | 1 Yes | 0 | | |
| 29 | 27 | Aarav | Male | 50 | Ghazababa | Working | FHR Manager | | 5 | | | 3 Less than Healthy | MA | Yes | 12 | 5 Yes | 1 | | |
| 30 | 28 | Asha | Female | 47 | Kolkata | Working | Entrepreneur | | 3 | | | 5.7-8 hours | Moderate | BS | No | 5 | 1 No | 0 | |
| 31 | 29 | Kashish | Female | 19 | Agra | Working | Professional | | 1 | | | 5 More than Healthy | Class 12 | No | 5 | 2 No | 0 | | |
| 32 | 30 | Priyanka | Female | 28 | Varanasi | Student | | 3 | 5.59 | 2 | | 7-8 hours | Moderate | BCA | Yes | 4 | 5 Yes | 1 | |
| 33 | 31 | Aaradhy | Female | 37 | Pune | Working | Chef | | 2 | | | 1.7-8 hours | Unhealthy | BHM | No | 5 | 3 Yes | 0 | |
| 34 | 32 | Anavi | Male | 25 | Jalpur | Working | Teacher | | 4 | 8.13 | 3 | | 5-6 hours | Moderate | ITech | Yes | 1 | 1 No | 0 |
| 35 | 33 | Tammy | Male | 20 | Patna | Student | | 2 | 5.7 | 3 | | 1 More than Healthy | PA | No | 4 | 1 No | 0 | | |
| 36 | 34 | Vihann | Male | 41 | Chennai | Working | Teacher | | 1 | | | 4.7-8 hours | Unhealthy | LLB | Yes | 6 | 5 No | 0 | |
| 37 | 35 | Shiv | Male | 45 | Rajkot | Working | Chef | | 1 | | | 4.7-8 hours | Moderate | B.Pharm | Yes | 9 | 1 Yes | 0 | |
| 38 | 36 | Anvi | Female | 50 | Kalyan | Working | Professional | | 4 | | | 4.7-8 hours | Unhealthy | Class 12 | No | 0 | 2 No | 0 | |
| 39 | 37 | Darsih | Male | 51 | Nagpur | Working | Educational Consult | | 1 | | | 5.7-8 hours | Unhealthy | B.Ed | Yes | 0 | 4 No | 0 | |
| 40 | 38 | Samai | Male | 45 | Mumbai | Working | Civil Engineer | | 3 | | | 2 Less than Healthy | MCA | No | 9 | 5 No | 0 | | |
| 41 | 39 | Raunak | Male | 59 | Jalpur | Working | Manager | | 4 | | | 1 Less than Unhealthy | BHM | Yes | 2 | 4 No | 0 | | |
| 42 | 40 | Armaan | Male | 51 | Nashik | Working | Electrician | | 2 | | | 4.7-8 hours | Healthy | M | No | 1 | 2 Yes | 0 | |
| 43 | 41 | Mahi | Female | 47 | Ludhiana | Working | Pharmacist | | 4 | | | 1 Less than Healthy | B.Pharm | Yes | 9 | 2 Yes | 0 | | |
| 44 | 42 | Shaurya | Male | 60 | Ludhiana | Working | Manager | | 4 | | | 1.5-6 hours | Moderate | MHM | Yes | 3 | 2 No | 0 | |
| 45 | 43 | Vidya | Female | 18 | Vaddodara | Working | Professional | | 5 | | | 4 Less than Moderate | Class 12 | Yes | 10 | 5 No | 1 | | |
| 46 | 44 | Jai | Male | 55 | Jalpur | Working | Civil Engineer | | 1 | | | 1.7-8 hours | Unhealthy | BBA | Yes | 0 | 2 No | 0 | |
| 47 | 45 | Ayoth | Male | 36 | Ludhiana | Working | Marketing Manager | | 1 | | | 2.7-8 hours | Healthy | BA | Yes | 0 | 2 Yes | 0 | |
| 48 | 46 | Arsh | Male | 49 | Meenut | Working | Teacher | | 2 | | | 1.7-8 hours | Moderate | M.Ed | Yes | 0 | 1 Yes | 0 | |

Mental Health & Suicide Survey Data
약 140,700명, 20개 변수(Train.csv)

[주요 컬럼 그룹]

인구통계: Gender, Age, City

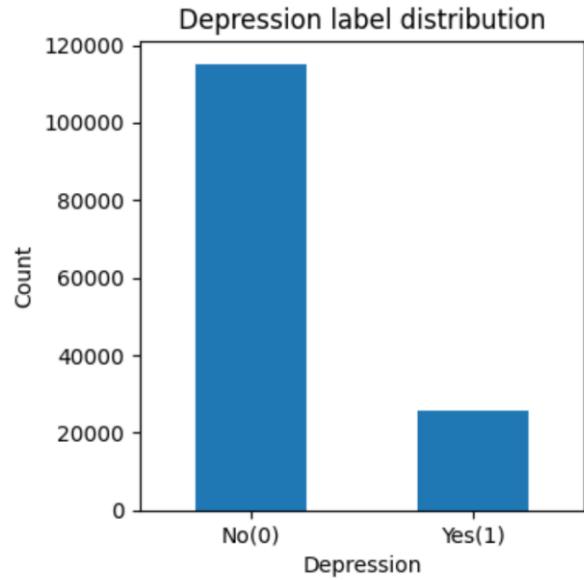
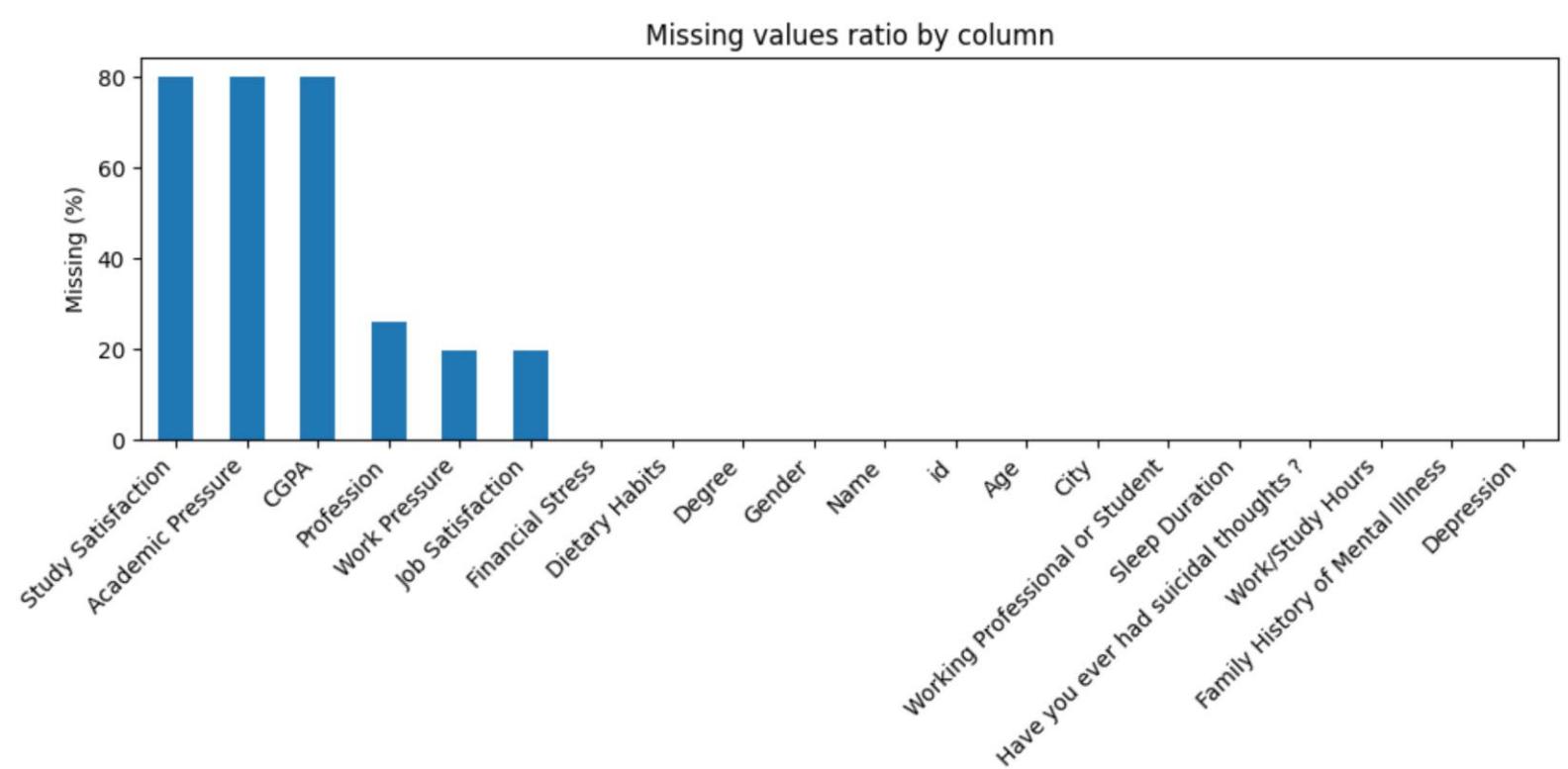
학업·직업: Student/Working, Profession, Degree, CGPA

스트레스: Academic Pressure, Work Pressure, Financial Stress

생활습관·만족도: Sleep Duration, Dietary Habits, Study/Job Satisfaction

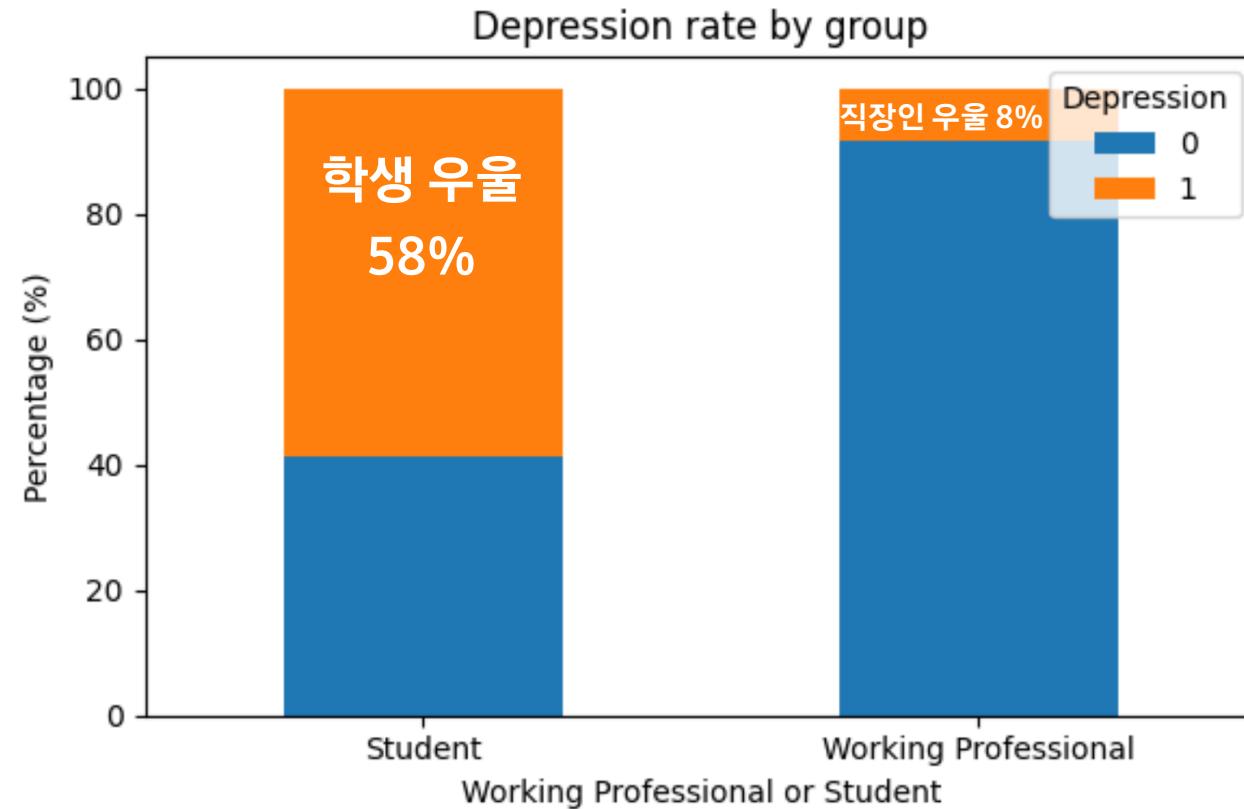
- 타깃 - Depression (0/1)

5. 타겟 분포

모델 학습 시 클래스 불균형 고려**Depression 18.2%****위험군은 상대적으로 소수**

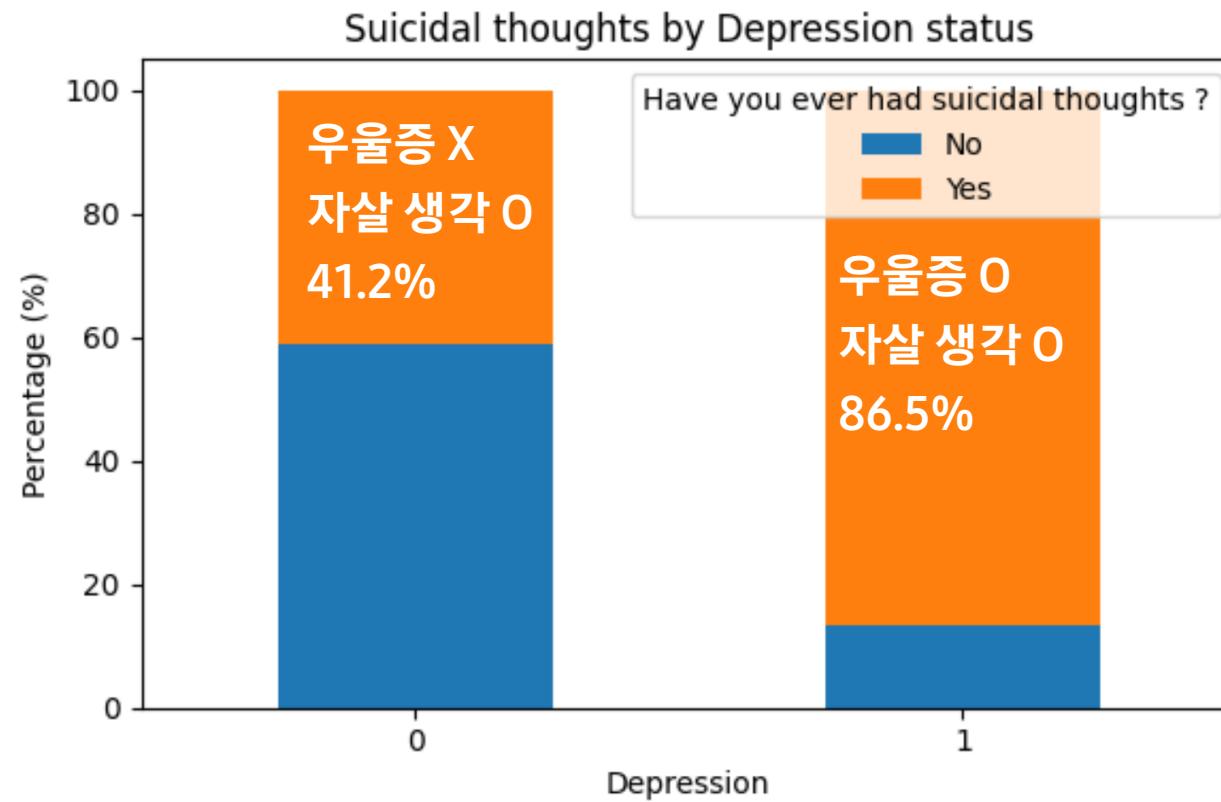
6. 학생 vs 직장인 - 우울

학생 우울 58%, 직장인 8%



7. 우울 vs 자살 생각

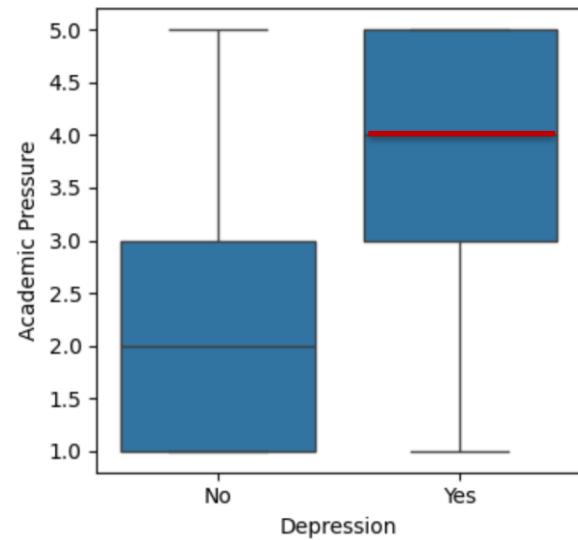
우울 없어도 41% 자살 생각



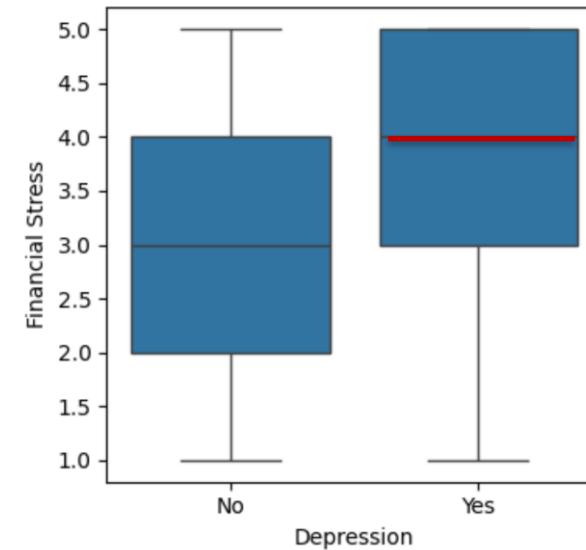
8. 스트레스 지표 분포

학업, 경제 스트레스 모두 압도적으로 높음

학업 스트레스



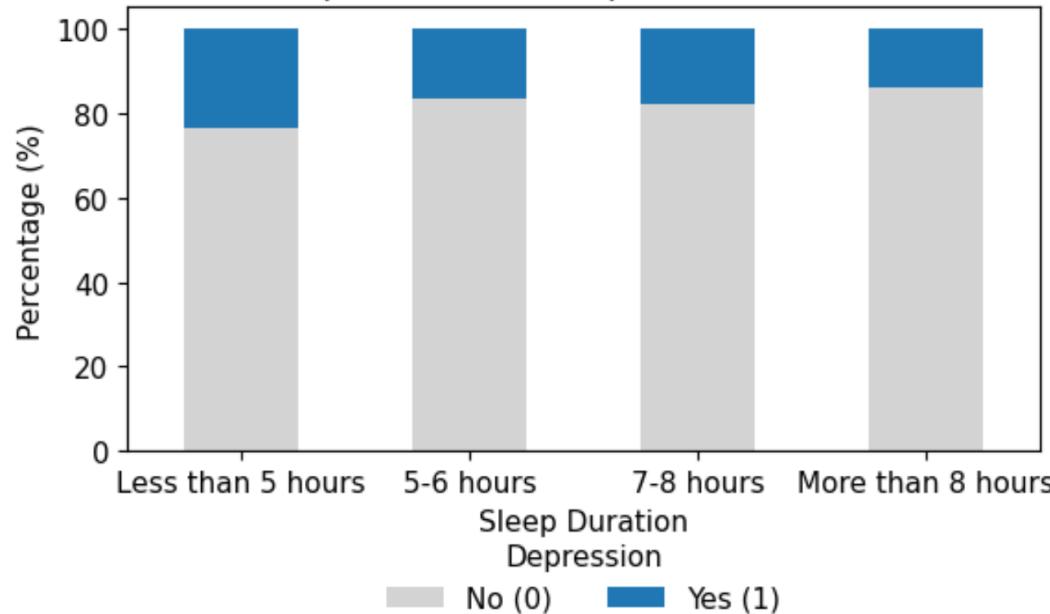
경제 스트레스



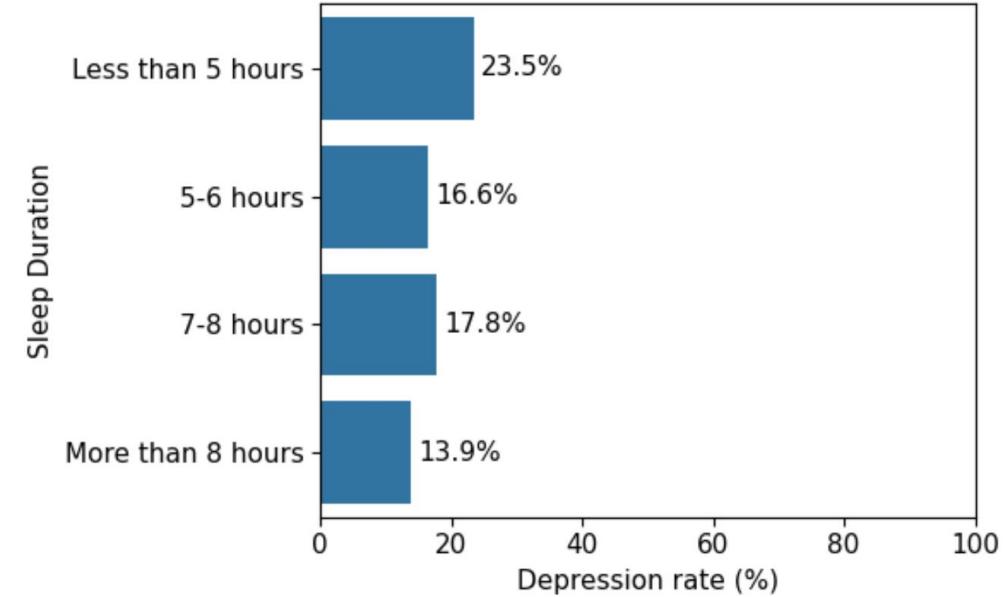
10. 수면과 우울

6시간 미만 학생 위험도가 압도적

수면 시간 vs 스트레스



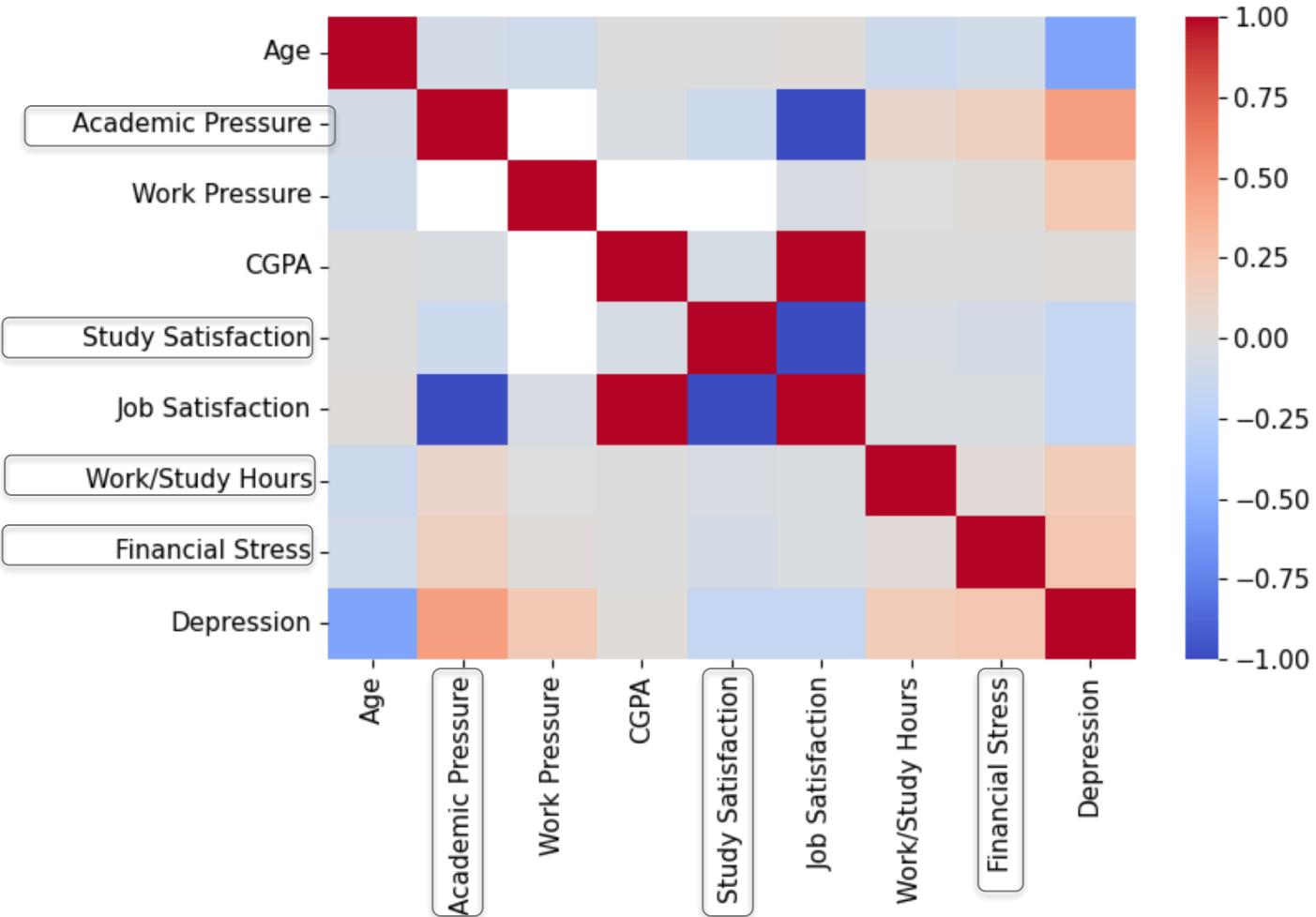
우울 : 수면시간 vs 스트레스



11. 숫자형 변수 상관관계

Academic Pressure, Financial Stress, Study Satisfaction 스트레스 구조 형성

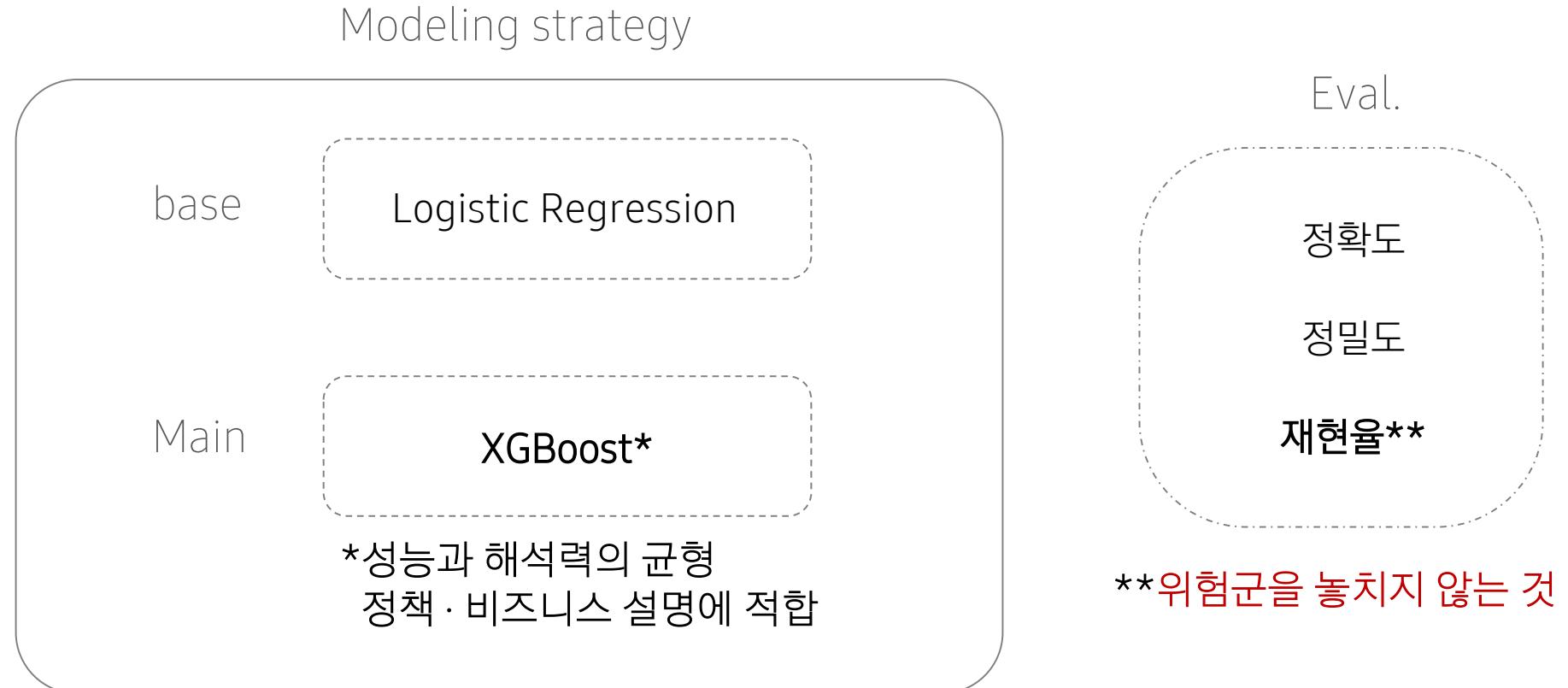
숫자형 변수 상관관계



인도 **학생** 정신건강 데이터를 활용해
핵심 요인 분석 및 예측

12_1. 모델 선정

자살 충동 여부
(Binary Classification)



12_2. 모델 파이프라인

ML 기반 학생 자살 충동 위험 예측 파이프라인 모델 구축

전처리



모델링



평가

Missing value 처리

- 수치형 : Median 대체
- 범주형 : OneHot

OneHotEncoder

(handle_unknown='ignore')

ColumnTransformer

XGBoostClassifier

- scale_pos_weight
= (neg/pos)
- 비선형 관계 강함

정확도

정밀도

재현율

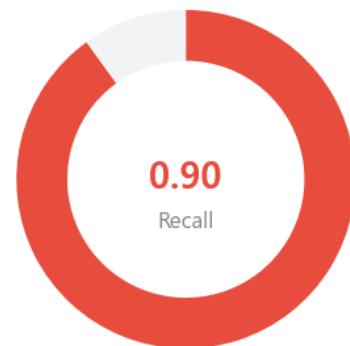
← 위험군 탐지 최우선

12_3. 모델 성능

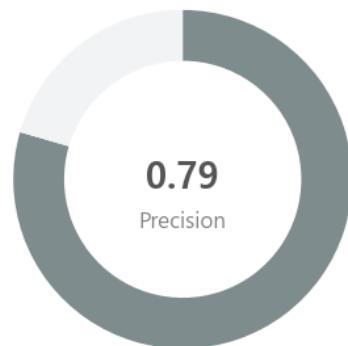
Recall 0.90 → 위험군 대부분 찾아냈다

모델 성능 지표 (Donut Chart)

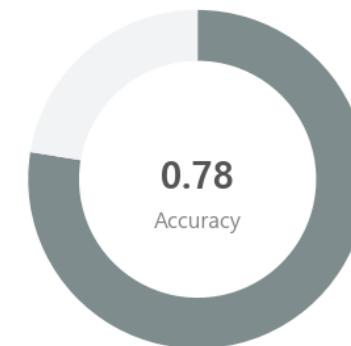
Recall



Precision



Accuracy



모델 성능 지표 (Donut Chart)

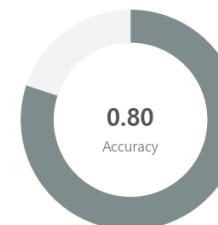
Recall



Precision



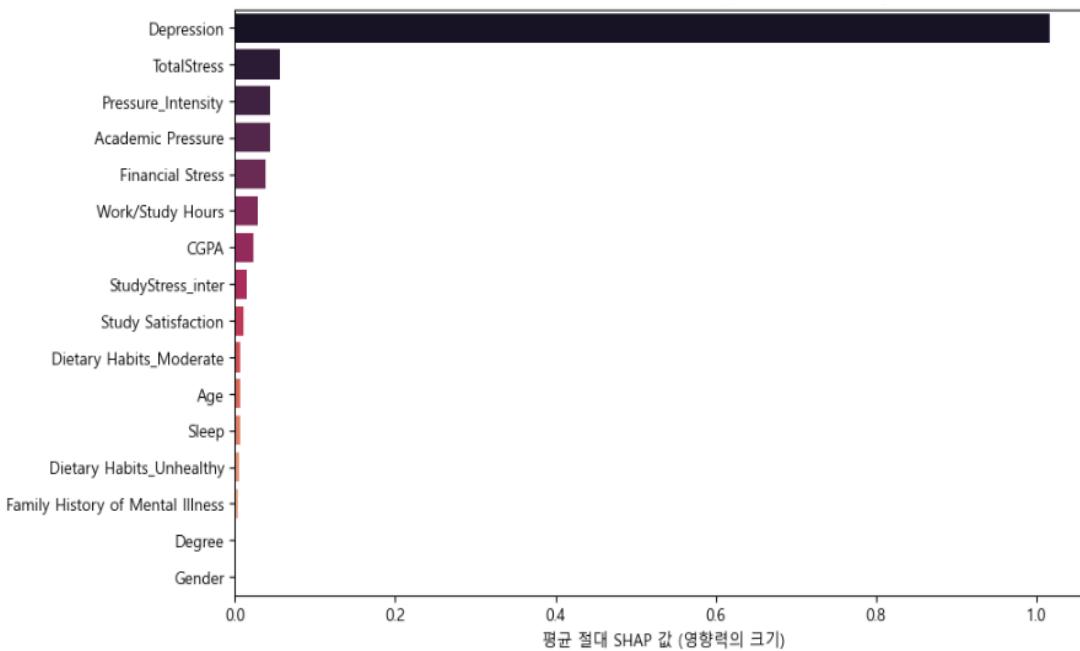
Accuracy



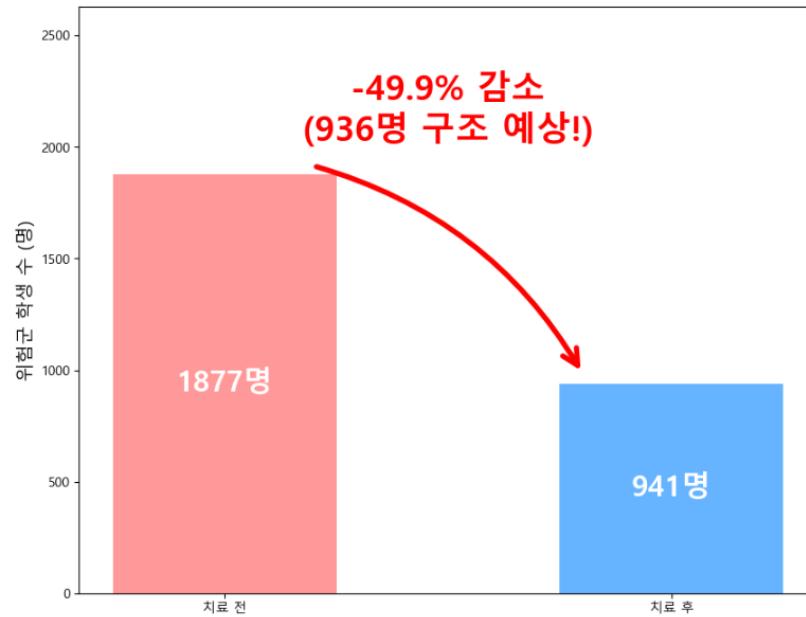
12_4. SHAP

학생 우울 핵심 요인 Depression 줄이면 자살 충동이 절반으로 감소

학생 자살 충동 위험 요인



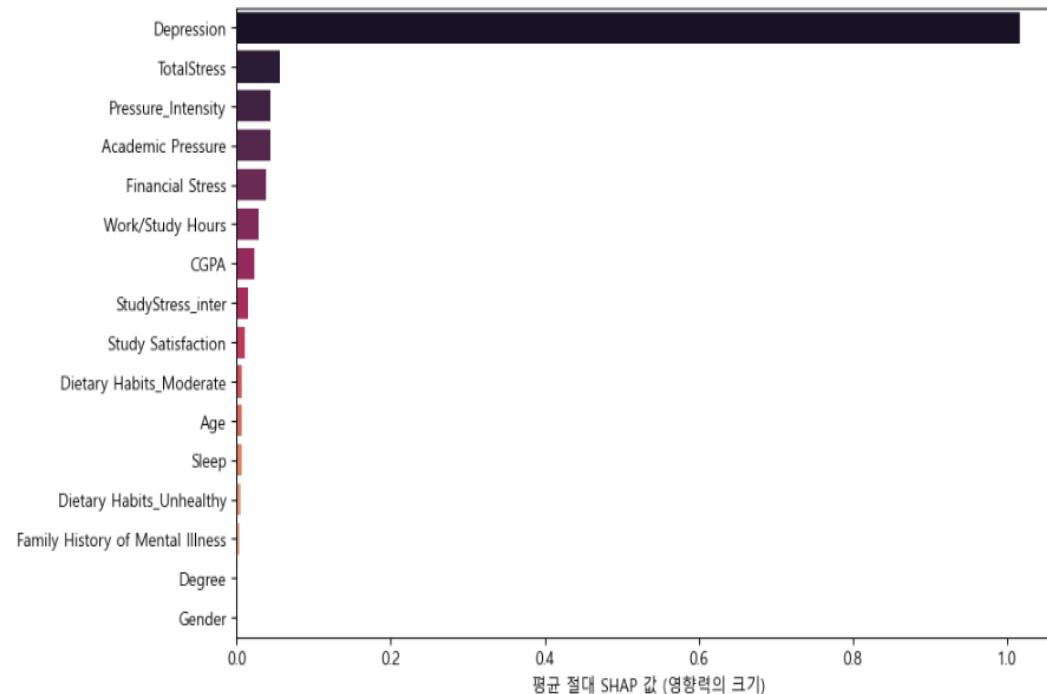
Depression 감소 시뮬레이션



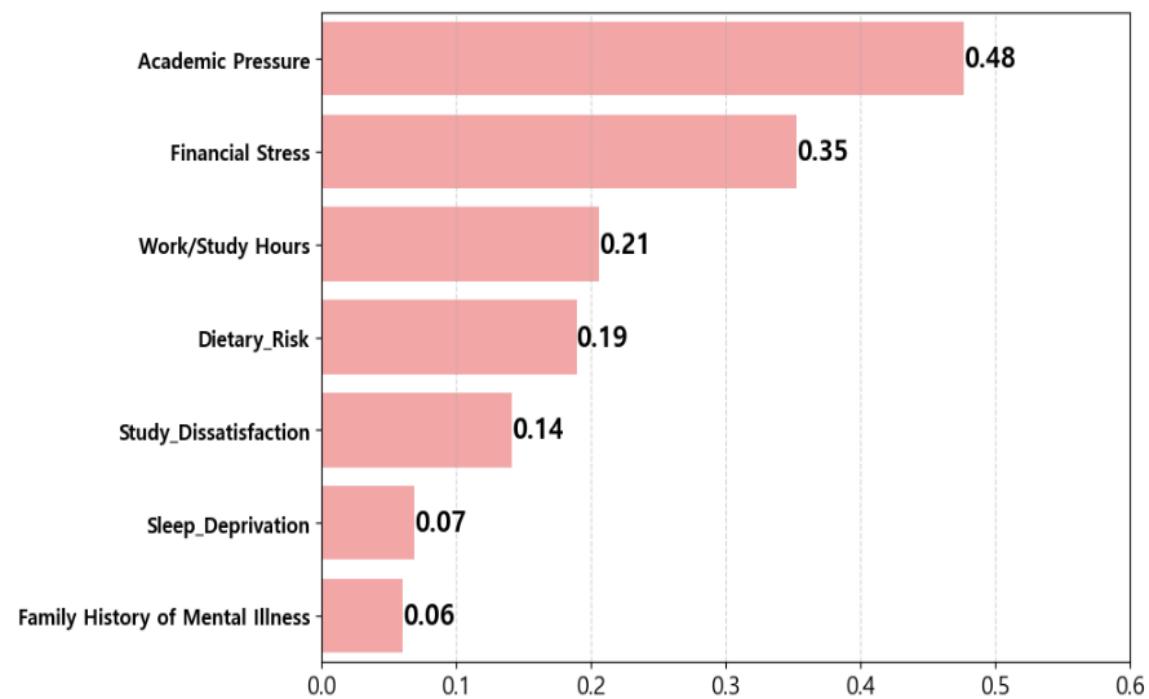
12_4. Feature Importance

Academic Pressure → Financial Stress 핵심 요인

학생 자살 충동 위험 요인



학생 우울 핵심 요인



13. 시나리오

재정지원, 수면 교육, 상담, 영양 급식을 통한 개입 시나리오

| Scenario | 조정 Feature | 수정 방식(시뮬레이션) |
|-----------|-------------------------------------|----------------------------|
| 1 장학·재정지원 | Financial Stress | 상위 30% 학생 대상 스트레스 값 × 0.7↓ |
| 2 수면 교육 | Sleep Duration | 6h 미만 → 7h 일괄 조정 |
| 3 상담 | Academic Pressure, Work/Study Hours | AP 20% 감소, Hours 10% 감소 |
| 4 영양 · 급식 | Dietary Habits | Dietary Habits → Healthy |

보호효과 분석

14_1. 보호 효과

학생 위험 확률을 0~1로 표현하는 ML 기반 조기경보 지표

보호 효과

개입 시나리오를 적용했을 때
평균 위험도가 얼마나 줄었는가

$$\text{PE Index} = \text{Risk Reduction}(\%) \times 10$$

XGBoost 모델로 학생 우울 위험도를 0~1
확률로 예측 → Risk(%) = 평균 확률 × 100

평균 위험도 감소율(%)에 10을 곱한 지수형
스코어 적용

평균 위험도 감소율

'그 집단 전체가 얼마나 덜 위험해 졌는지를 %로 표현한 값'

$$\text{Risk Reduction (\%)} = \frac{\text{Baseline Risk} - \text{Intervention Risk}}{\text{Baseline Risk}} \times 100$$

Baseline Risk: 개입 전 평균 위험도(%)

Intervention Risk: 해당 시나리오 적용 후 평균 위험도(%)

'진짜 위태로운 사람들(고위험군)이 얼마나 줄었는지 보여주는 지표'

고위험군 감소율

$$\text{High Risk Reduction (\%)} = \frac{\text{Before} - \text{After}}{\text{Before}} \times 100$$

High Risk Count (Before): 개입 전 고위험군 인원 수

High Risk Count (After): 개입 후 고위험군 인원 수

14_2. 보호 효과

정책 적용 후 위험도가 얼마나 % 감소하는지 측정

순서

해석

1. 기준선 위험도 계산

`predict_proba`의 1번 클래스 확률 = 우울 위험도(baseline Risk)

양수(+) → 위험도가 줄어든 것 (보호효과 있음)

음수(-) → 위험도가 오히려 오른 것 (역효과)

2. 개입된 데이터 넣어 다시 예측

Risk Reduction(%)에 10을 곱한 스코어형 지수

0.5 이상인 사람 수 → 개입 후 고위험군 수

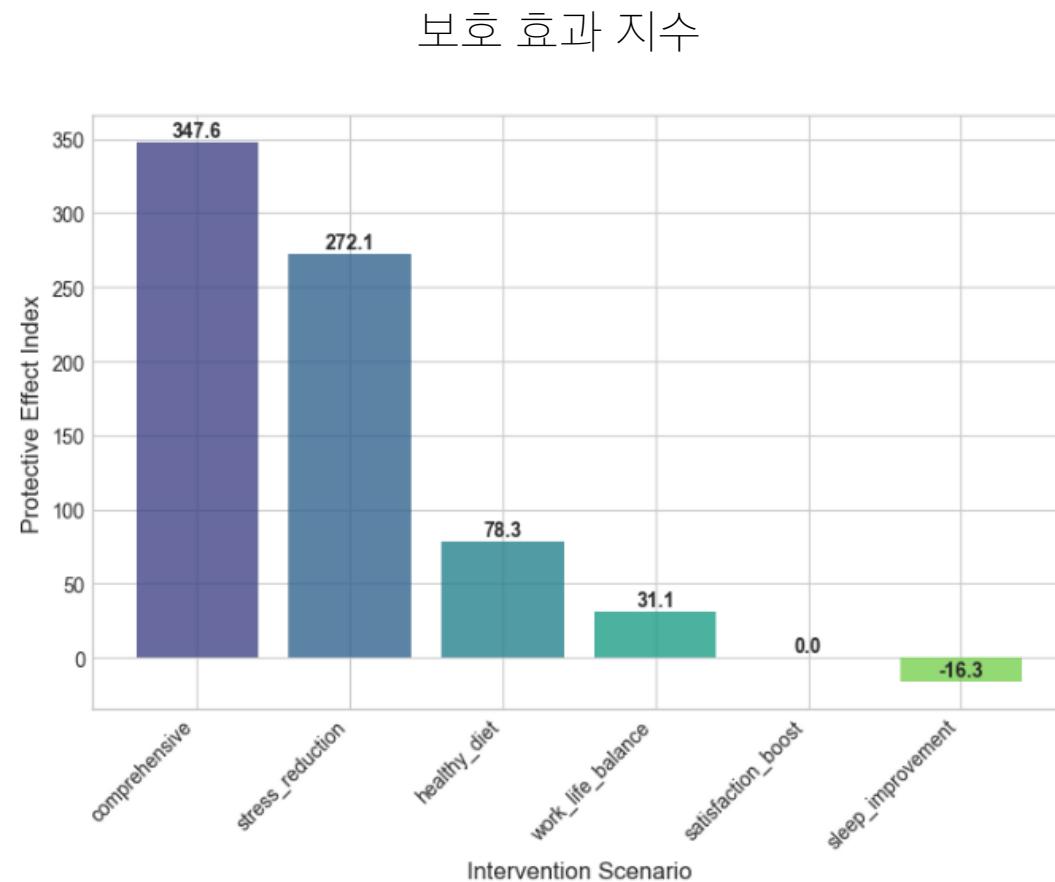
값이 클수록 “정책이 강한 보호효과”

3. 보호효과(위험도 감소율) 계산

(고위험군 : 위험도 0.5 이상)

14_3. 보호 효과

종합 개입과 스트레스 완화로 압도적 개선 가능성 확인



종합 개입 347점, 스트레스 완화 272점

→ 다른 정책과 비교해 압도적

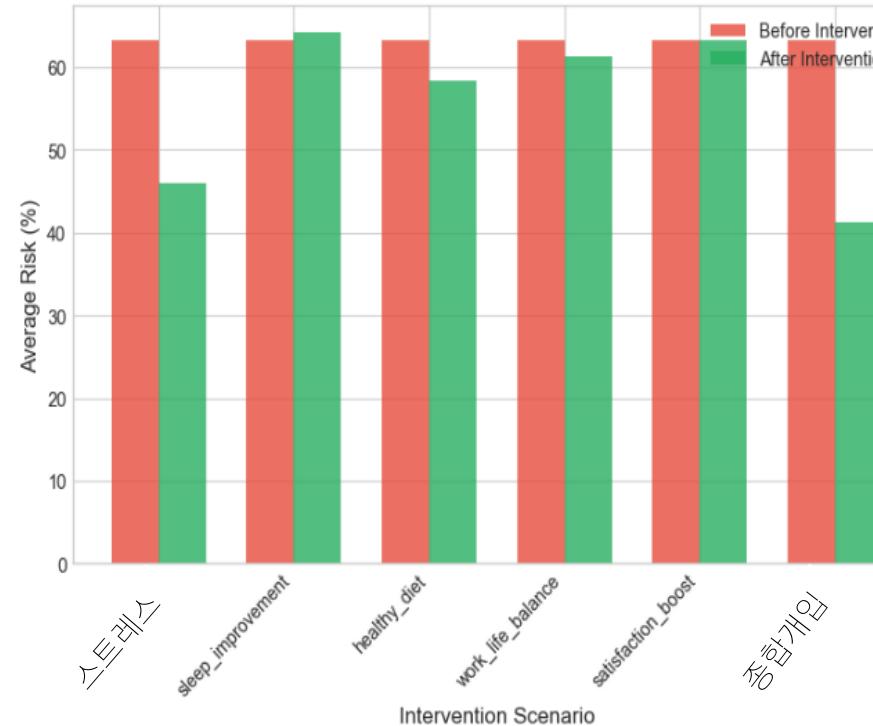
식습관 개선(healthy_diet)도 일정 효과 존재 (78점)

수면 개선은 음수(-16) → 정책 단독 적용은 비효율적

14_4. 보호 효과

우울 위험을 크게 낮추는 구조적 개입

Before vs After



개인별 위험도 변화

| | Original_Risk | Intervention_Risk | Risk_Reduction | Risk_Reduction_Pct |
|------|---------------|-------------------|----------------|--------------------|
| 1313 | 0.757193 | 0.163334 | 0.593859 | 78.429016 |
| 2483 | 0.783857 | 0.190740 | 0.593117 | 75.666527 |
| 1935 | 0.754143 | 0.162523 | 0.591620 | 78.449341 |
| 253 | 0.773970 | 0.183605 | 0.590365 | 76.277542 |
| 1846 | 0.804962 | 0.221652 | 0.583311 | 72.464355 |
| 2287 | 0.788250 | 0.205605 | 0.582645 | 73.916298 |
| 488 | 0.782503 | 0.201789 | 0.580714 | 74.212357 |
| 455 | 0.761814 | 0.181861 | 0.579953 | 76.127922 |
| 1861 | 0.743522 | 0.164944 | 0.578578 | 77.815887 |
| 1791 | 0.776067 | 0.200738 | 0.575329 | 74.133888 |
| 2647 | 0.761794 | 0.186912 | 0.574882 | 75.464233 |
| 2524 | 0.772444 | 0.197608 | 0.574836 | 74.417831 |
| 521 | 0.747327 | 0.172985 | 0.574341 | 76.852783 |
| 2772 | 0.769873 | 0.196301 | 0.573572 | 74.502167 |
| 2324 | 0.749342 | 0.176106 | 0.573235 | 76.498512 |
| 92 | 0.744782 | 0.171573 | 0.573209 | 76.963295 |

15. 요약 및 인사이트

데이터 기반 활용한 정책 방향 제안

요약

학생 집단을 분류하여
XGBoost로 스트레스 핵심 요인을 예측하고,

개입 시나리오를 적용
(스트레스 완화, 수면, 식습관, 워라밸, 종합 개입)

“우리 학생에게 어떤 정책이 가장 효과적인지” 분석

인사이트

1. 고위험 학생 조기 선별·집중 지원 체계 구축
2. 학업·재정 스트레스 완화를 1순위 정책
3. AI 기반 학생 정신건강 조기 경보 시스템 도입
4. 데이터 기반 정책의 선순환 구조 정착

16. 정책적 제안

재정 지원 및 환경 개입을 통한 학생 보호



스트레스 완화 (Stress Reduction)

- ✓ 상담 시스템(전문교사)
- ✓ 스포츠 프로그램
- ✓ 정기 야외 체험 활동
- ✓ 진로 · 진학 상담 (공정)



수면 개선 (Sleep Improvement)

- ✓ 건강 수면 캠페인
- ✓ 학교 건강 수면 교육
- ✓ 야간 학습 가이드라인



양질 급식·영양 (Healthy Diet / Nutrition)

- ✓ 학교 건강 식습관 교육
- ✓ 양질 급식(영양 교육 포함)

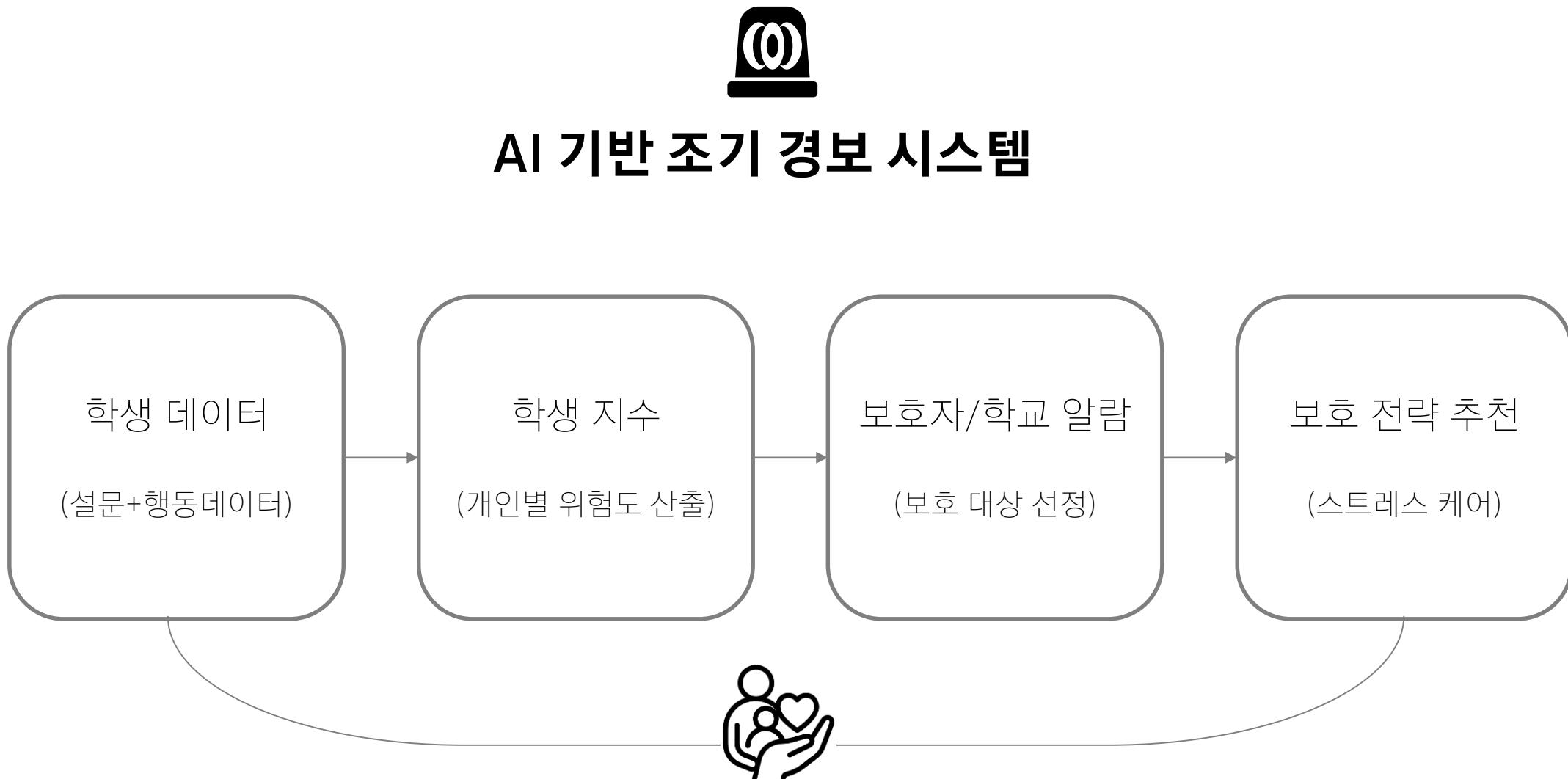


장학·공정성 (Financial Aid / Fairness)

- ✓ 소득수준별 장학금
- ✓ 근로장학금
- ✓ 의무교육(중등) 강화
- ✓ 중도 탈락 방지 프로그램
- ✓ 기업 장학금
- ✓ 학업 차별 금지 및 맞춤형

학습 지원

17. 비즈니스 모델



18. AI 기반 예측 모델과의 상생

우리는 데이터사이언스로 생명을 지킬 수 있습니다.

자살은 '예측 가능한 사건'

학생의 일상 데이터는

모두에게 위험 신호를 보여주고 있습니다.



위험군을 빠르게 찾아내고

효과적인 개입 전략으로

학교·지역사회·기업의 삶의 질 향상