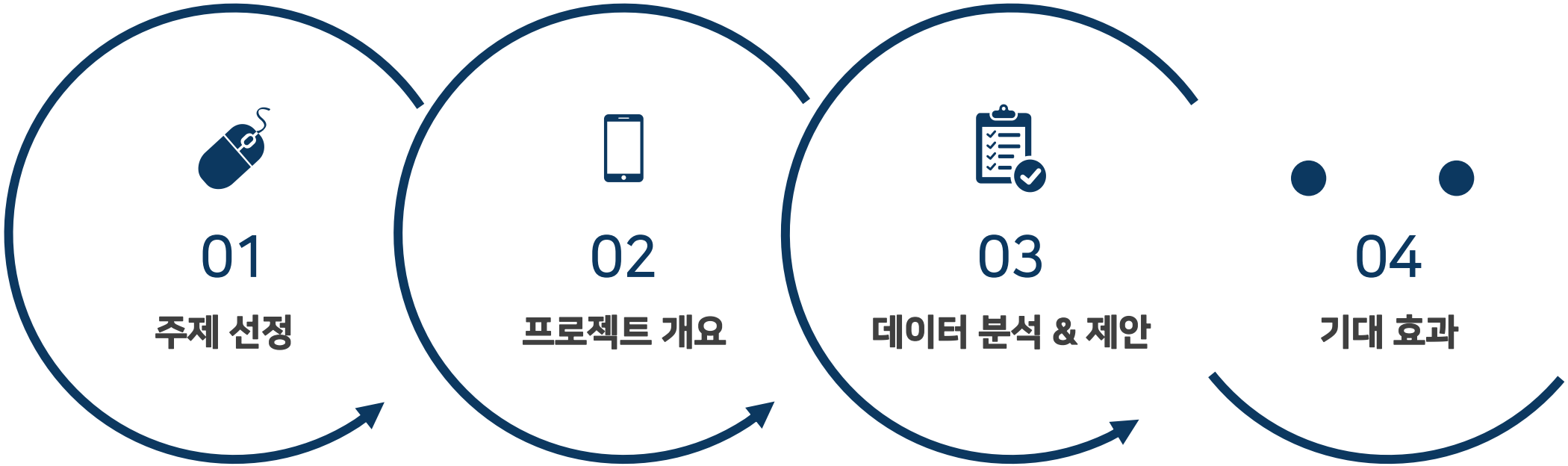
You are not alone.

사이버 폭력 근절을 위한 혁신적 모니터링 제안

박지현 박희원 이원영 위예진

You are not alone.

Table of contents



1

주제 선정

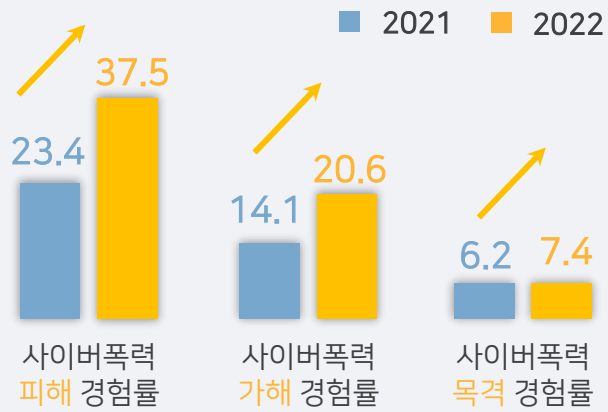
1 (1) 프로젝트 배경

계속되는 청소년 사이버 폭력의 증가

청소년 약 9,700명 대상으로 시행한 실태 조사 결과,
전년도 대비 사이버폭력 경험률이 증가한 것을 알 수 있다.

방통위 "청소년 10명 중 4명 사이버 폭력 경험 있어"

사이버폭력 가해·피해 ·목격 경험률



(출처 : 방송통신위원회)

청소년의 인터넷 사용 시간 증가

■ 사이버 폭력은 익명성을 보장받는 커뮤니티에서 이뤄져 가해자 특정하기 어렵다.

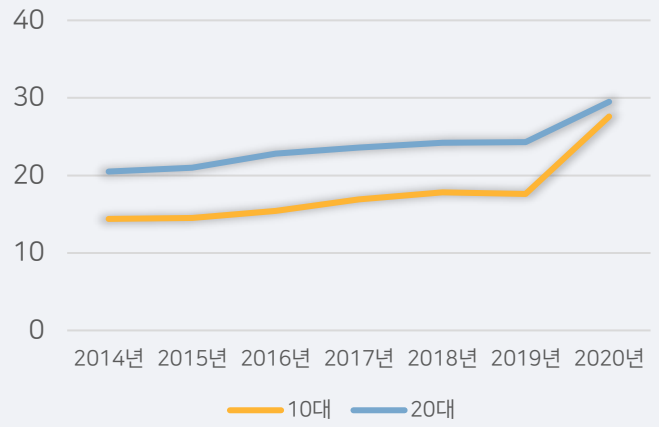
■ 청소년 주 평균 인터넷 이용시간이 해마다 증가하고 있으며,
여가활동 뿐만 아니라 학습, 정보 획득을 위해서도 인터넷을 많이 사용하는 것으로 나타났다.

■ 즉, 청소년들의 사이버 폭력 노출 위험성이 높아지고 있다.

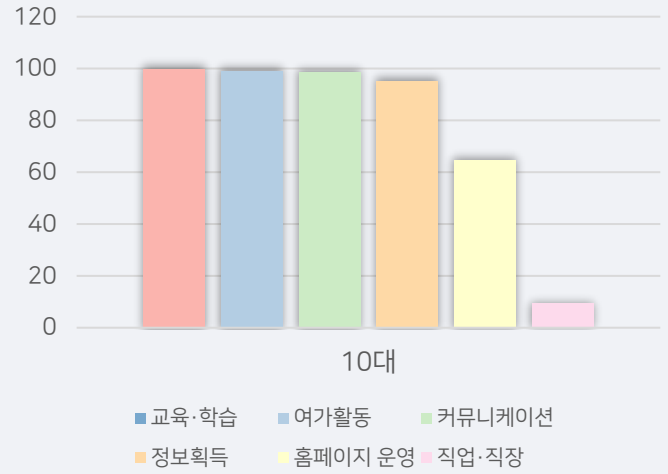
코로나19 이후 '사이버 학교폭력' 증가...가해자 처벌하기 쉽지 않아
사이버 학교폭력 피해자에게 정신적, 심리적인 피해 준다
익명성을 보장받는 커뮤니티에서 이뤄져 가해자 특정하기 어렵다

(출처 : Pax news)

청소년 인터넷 주평균 인터넷 이용시간



청소년 인터넷 이용목적 (복수응답)



(출처 : 한국인터넷진흥원)

1 (2) 프로젝트 목적

정신 건강 예방에 효과적인 Gerald Kaplan의 프레임워크

1차 예방	2차 예방	3차 예방
문제가 발생하기 전에 예방하는 것이 목표 e.g.) 교육, 환경 개선	문제가 초기 단계에 있을 때 발견하고 치료하는 것이 목표 e.g.) 조기 진단, 정기 검진	이미 발생한 문제의 영향을 최소화하고 재발을 방지 하는 것이 목표 e.g.) 상태 개선, 증상 관리

활 용

디지털 폭력 예방 프레임워크

1차 예방	2차 예방	3차 예방
데이터 분석을 통한 사이버 폭력 예방을 위한 교육 전략 - 사이버 폭력 신고 인식 교육 집중 개선 - 유해 컨텐츠 노출 피해 예방 교육 확대	가해 청소년 분류 모델을 이용한 가해 청소년 조기 발굴 및 모니터링 - Catboost를 이용한 예측 모델 활용	사이버폭력 피해 학생들의 멘탈을 케어 해주는 챗봇 적극 활용 - koGPT2를 이용한 챗봇 서비스

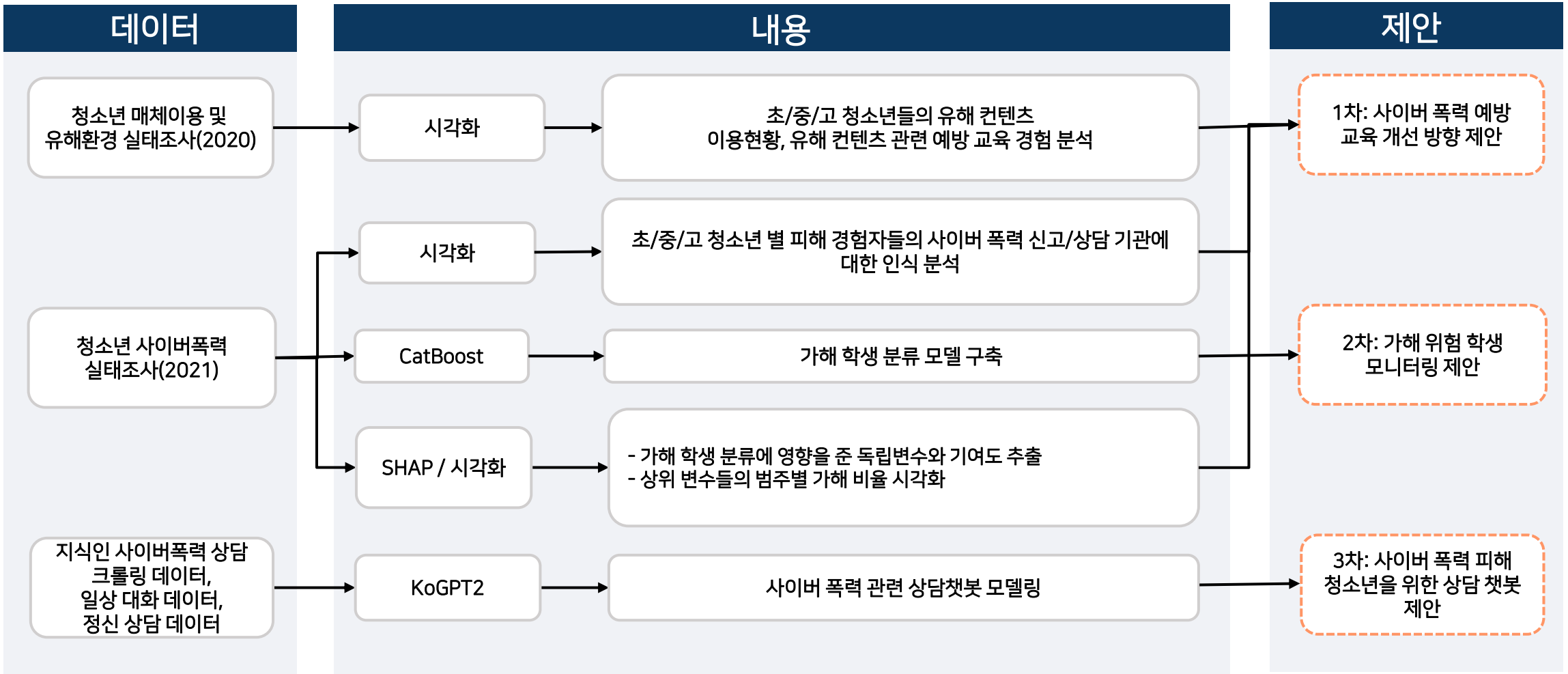
“ You are not alone : 사이버 폭력 근절을 위한 단계적 예방 전략 ” 프로젝트 제안

2

프로젝트 개요

2 (1) 프로젝트 개요

목표 : 청소년의 사이버폭력 피해/가해 실태 및 요인을 분석하여 1차,2차,3차 예방 방안을 제안함



2 (2) 활용 데이터

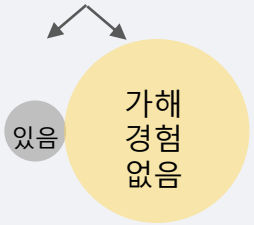
NIA 한국지능정보사회진흥원

2021년 한국지능정보사회진흥원 사이버폭력 실태조사 데이터

1. 사이버 폭력 피해 청소년의 신고 인식 시각화

➡ Q8_8. 괴롭힘을 당하고도 사이버폭력 관련 기관에 도움을 요청하지 않은 가장 큰 이유는 무엇입니까?

2. 사이버 폭력 가해 청소년 분류 모델링

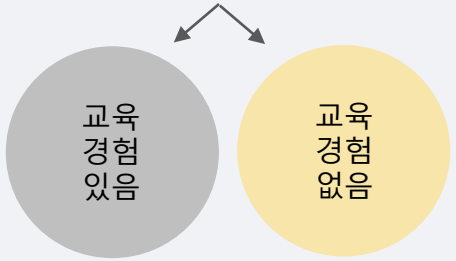


➡ 가해 위험 학생을 예측하는데 가장 영향을 준 요인파악

여성가족부

2020년 청소년 매체 이용 및 유해환경 실태조사 데이터

성인용 컨텐츠 노출로 인한 피해 예방 교육 여부 시각화



NAVER

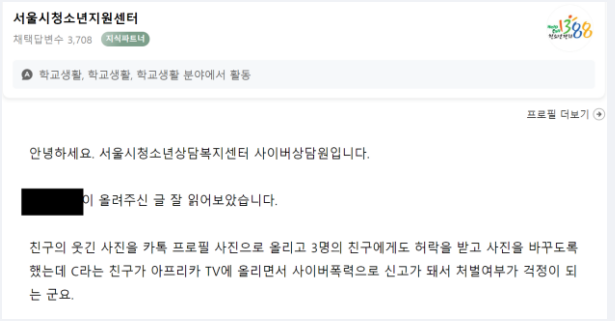
챗봇 학습 데이터

1. AI hub 감성 대화 말뭉치

감정 + 심리 상담 데이터

2. 네이버 지식인 사이버폭력 Q&A 크롤링 데이터

공식 기관의 사이버 상담원 데이터

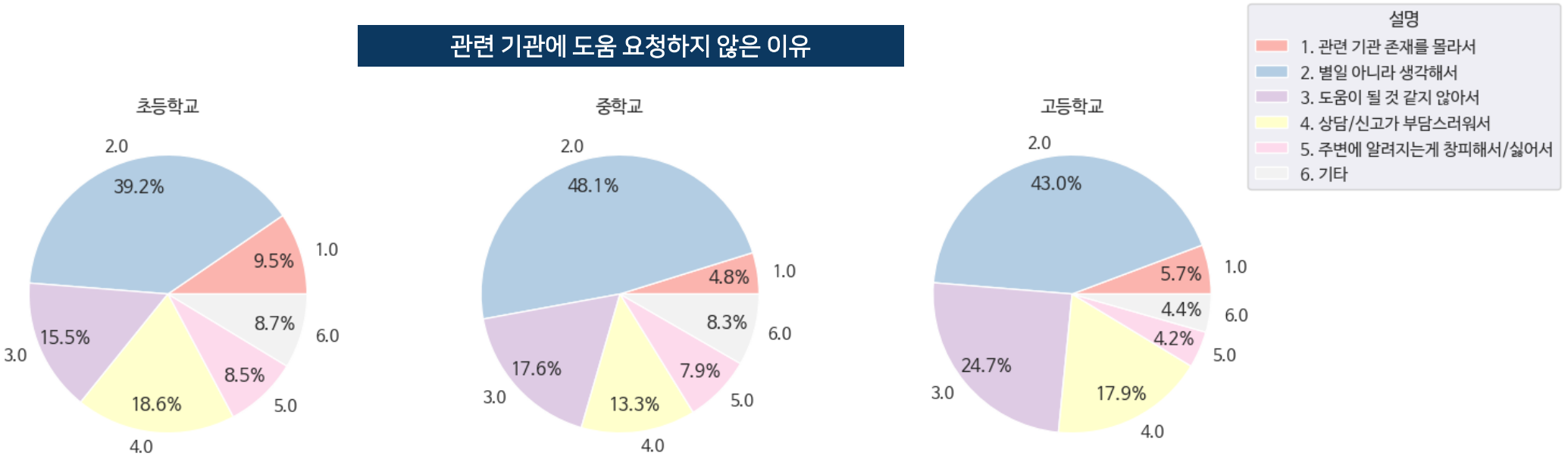


3

데이터 분석 & 제안

3 1차 예방 (1) 피해 경험 학생이 관련 기관에 도움 요청 하지 않은 이유 분석

분석 결과 : 청소년 사이버 폭력 실태 조사



응답별 비율을 시각화한 결과에 따르면,

1. 연령대가 높아질수록 관련 기관이 **도움될 것 같지 않아서** 신고하지 않았다고 응답한 비율이 증가함
2. **상담이나 신고가 부담스럽다**는 응답도 평균 16.6% 로 많은 편에 속함

-> 사이버폭력 관련기관에 대한 인식 개선 필요

정책 제안 1

사이버 어울림 프로그램 - 초등 기본

역량	전체제목	구분		1차시	2차시	3차시
사이버 폭력 인식 및 대처	사이버 폭력, 로그 아웃!	차시명		사이버 세상에도 '폭력'이 일어난다고요?	이것도 사이버 폭력이 될 수 있나요? 아하 그랬구나!	무엇부터 할 수 있나요? 함께해요!
		학습주제		사이버폭력의 특징과 심각성 알기	사이버폭력의 개념과 생활 속 사이버폭력의 유형을 알고, 우리가 할 수 있는 일 생각하기	생활 속 사이버폭력의 대처 방안과 예방법을 알고 실천하기
		도입		사이버폭력 영상 살펴보기	학교생활 속 사이버폭력 사례 살펴보기	사이버폭력의 잘못된 대처 사례 살펴보기
		전개	활동1	▶ 역할극을 통한 사이버폭력 간접 체험하기	▶ 생활 속 사이버폭력 상황 및 유형 살펴보기	▶ 생활 속 사이버폭력의 올바른 대처방안 알아보고 연습하기
			활동2	▶ 사이버폭력의 특징과 심각성 알아보기	▶ 사이버폭력 실제 사례 살펴보기	▶ 사이버폭력 예방을 위한 방법을 알고 실천하기
		정리		사이버폭력도 '심각한 학교폭력'임을 알기	사이버폭력 상황에서 친구를 위해 할 수 있는 일 정리하기	사이버 공간에서 올바른 행동과 언어생활 실천 다짐하기

사이버 어울림 프로그램 - 중등 기본

역량	전체제목	구분		사이버세상에서 폭력이란?	내가 올린 글이 폭력이라고요?	내가 어떻게 대처해야 하나요?
사이버 폭력 인식 및 대처	듣지도 보지도 못한 현명한 사이버폭력 대처!	차시명		사이버세상에서 폭력이란?	내가 올린 글이 폭력이라고요?	내가 어떻게 대처해야 하나요?
		학습주제		사이버폭력 민감하게 인식하기	사이버폭력의 심각성을 이해하고 예방수칙을 익혀 실천하기	사이버폭력 발생 시 단계별 대처 행동을 알고 실천하기
		도입		사이버폭력 관련 뉴스 시청하기	사이버폭력 심각성 이해하기	카드뉴스를 통해 사이버폭력의 대처 행동의 필요성 확인하기
		전개	활동1	▶ 사이버폭력 자가 진단하기	▶ 나의 사이버 생활 살펴보기	▶ 사이버폭력 단계별 대처행동 알아보기
			활동2	▶ 학교에서 발생하는 사이버 폭력 조사하기	▶ 사이버폭력 예방 수칙 조사하기	▶ 사이버폭력 대처 시나리오 제작하기
		정리		사이버폭력 알아채기	사이버폭력 금지 서약서 작성하기	영상 시청 및 활동 소감 발표하기

AS - IS

기존의 사이버 폭력 예방 교육인 어울림 프로그램(기본)에서는 사이버 폭력의 심각성 및 대처 방법에 대해 교육하고 있으나, **관련 기관이 어떤 도움을 주는지에 대한 내용이 미흡**함

제 안

TO - BE

신고/상담 요청을 하면 **각각의 관련 기관에서 어떤 절차가 진행이 되는지, 얼마나 도움이 되는지를 강조**하여 교육내용에 포함할 것을 제안

데이터 전처리 : 청소년 사이버 폭력 실태 조사

	city	district	school_level	is_coed	Q6_1	Q6_2	Q6_3	Q6_4	Q6_5	violence_exp
0	서울	마포구	초등학교	남여공학	2	2	2	2	2	0
1	서울	마포구	초등학교	남여공학	0	0	0	0	0	0
2	서울	마포구	초등학교	남여공학	0	0	0	0	0	0
3	서울	마포구	초등학교	남여공학	1	1	2	2	2	0
4	서울	마포구	초등학교	남여공학	2	2	0	2	2	0
...	...	...	...	...	...	...	...	...	...	...
9012	경북	구미시	고등학교	남여공학	1	0	2	1	2	0
9013	경북	구미시	고등학교	남여공학	2	1	2	2	2	0
9014	경북	구미시	고등학교	남여공학	2	2	0	2	2	0
9015	경북	구미시	고등학교	남여공학	0	0	0	0	0	0
9016	경북	구미시	고등학교	남여공학	2	2	2	1	2	0

9017 rows × 10 columns

가해경험 여부 - {'없다':0, '있다':1}

사이버 언어폭력 / 사이버 명예훼손 / 사이버 스토킹 / 사이버 성폭력 / 신상정보 유출 / 사이버 따돌림 / 사이버 갈취 / 사이버 강요

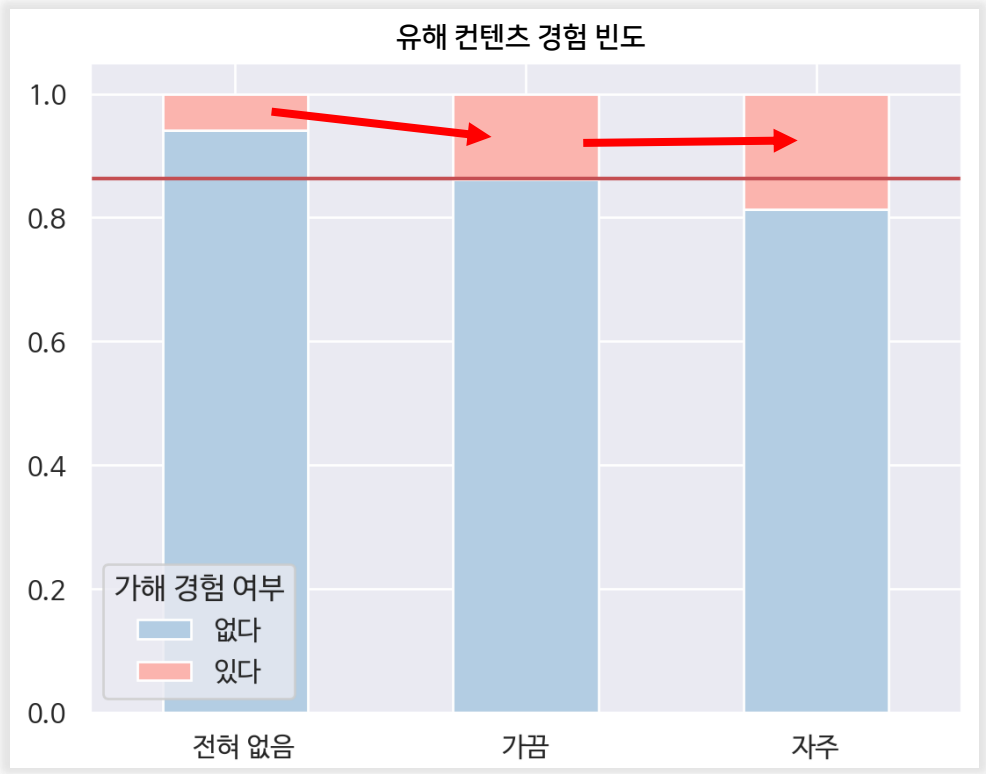
위와 같은 경험을 묻는 질문 (Q9a_1~Q9a_8)에 하나라도 경험이 있다고 답변하면 가해 경험을 1, 아니라면 0으로 컬럼 통합

유해컨텐츠 경험 빈도 – {'전혀없음':0, '가끔':1, '자주':2}

폭력적이고 잔인한 내용의 온라인 콘텐츠를 본 적 있다는 질문 (Q6_1)에
1년에 한두번 / 6개월에 한두번 ⇒ '가끔':1
한 달에 한두번 / 일주일에 한두번 / 매일 ⇒ '자주':2

컬럼의 값을 위와 같은 규칙으로 범주 재설정

분석 결과 : 청소년 사이버 폭력 실태 조사



시각화

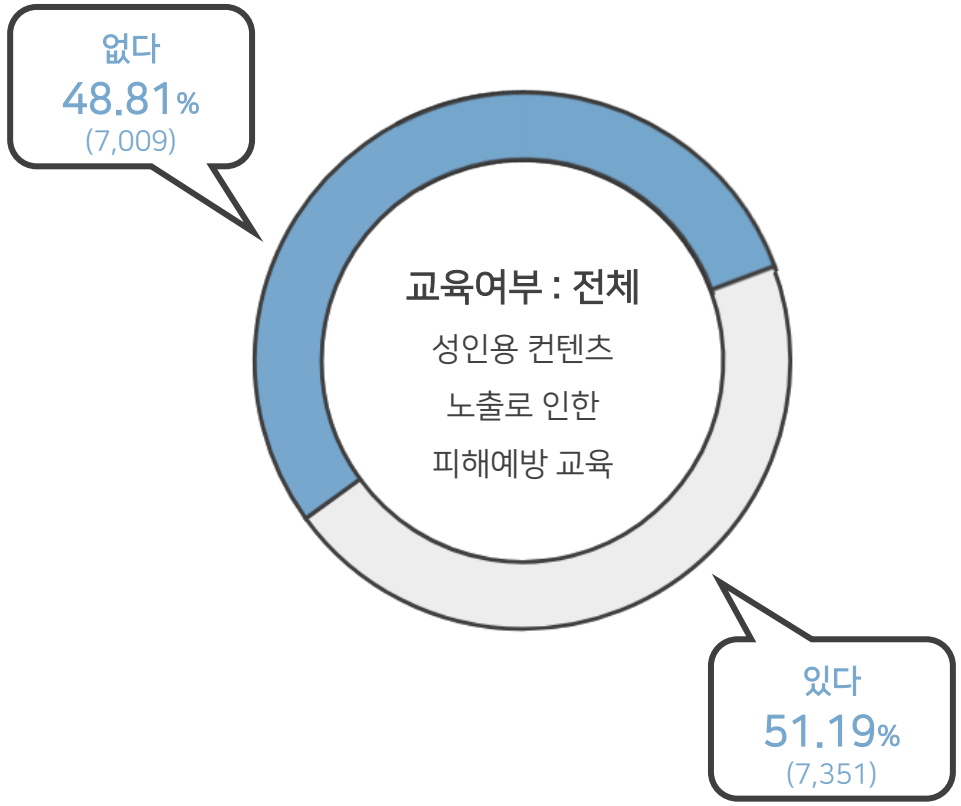
유해컨텐츠 경험의 빈도가 '전혀없음'→'가끔'→'자주'로 빈도가 잦아짐에 따라,
가해경험 여부에 'O(있다)' 에 응답한 청소년의 수가 늘어나는 것을 알 수 있다.

통계검정

카이제곱 통계량이 214.24621, p-value < 0.001 이므로
두 범주형 변수 사이에는 통계적으로 유의미한 관계가 있다고 볼 수 있다.

시각화와 통계검정 결과에 따르면,
유해컨텐츠 빈도와 가해경험 여부 사이에는 **유의미한 관계**가 있다.

분석 결과 : 청소년 매체 이용 및 유해환경 실태조사 데이터



정책 제안 2

데이터 분석 결과,

유해컨텐츠 빈도와 가해경험 여부 사이에는 **유의미한 관계**가 있음

청소년 매체이용 및 유해환경 실태조사 데이터에 따르면,

성인용 컨텐츠 노출로 인한 피해 예방 교육을 듣지 않은 청소년이 약 49%로 절반 가까이 됨

제 안

성인용 컨텐츠 노출로 인한 피해 예방 교육 혹은 유해 컨텐츠 관련 교육을 더 많은 청소년이 수강할 수 있도록 교육을 확대

3 2차 예방 가해 학생 분류 모델링

데이터 전처리 : 청소년 사이버 폭력 실태 조사

사용 컬럼

사이버 폭력 목격/가해/피해 경험과 관련된 컬럼, 종속성 있는 컬럼 제외

→ 사이버폭력 목격/가해/피해 경험과 직접적인 관련이 없는 컬럼들은 최대한 모두 사용하여 가해 경험 학생을 분류 (총 79칼럼)

1. 시도
2. 행정구
3. 학교급
4. 학교 세부 유형
5. 남녀공학구분
6. 지역규모
7. 성별
8. 하루 평균 인터넷 사용 시간
9. 인터넷으로 주로 이용하는 활동
10. 지인과 의사소통을 위해 가장 많이 활용하는 커뮤니케이션 방식

⋮

70. 온라인에서 유해매체 시청 빈도
71. 인터넷 사용에 대한 부모님의 제재
72. 인터넷 사용에 대한 부모님의 교육
73. 인터넷 사용에 대한 학교의 제재
74. 인터넷 사용에 대한 학교의 교육
75. 현재 함께 살고 있는 가족
76. 아버지의 최종 학력
77. 어머니의 최종 학력
78. 가정의 경제수준
79. 학업 성적

is_coed(남여공학) - 결측치 처리

초등학생만 남여공학컬럼에 대해 결측치 존재
-> '남여공학'으로 대체

Q13 , Q14, A - 결측치 처리

결측치가 선택 안함, 해당사항 없음을 의미
-> 0으로 대체

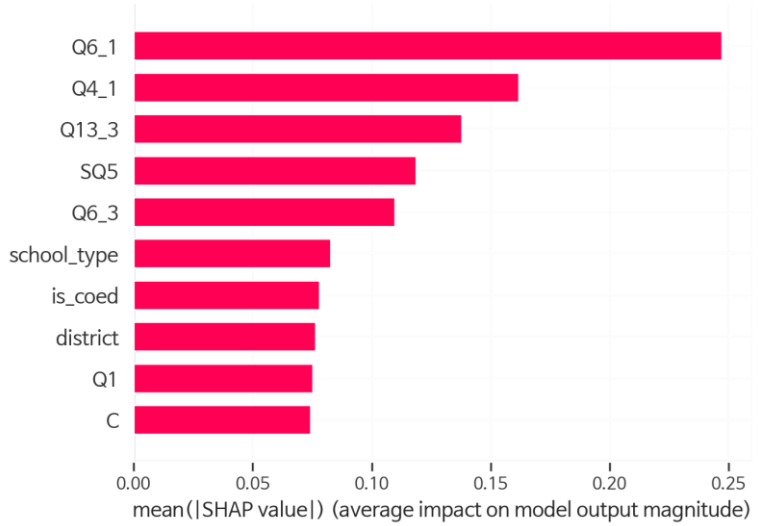
Q6, Q20, B - 값 변경

9 or 99 값이 전혀 없음을 의미함
-> 0으로 대체

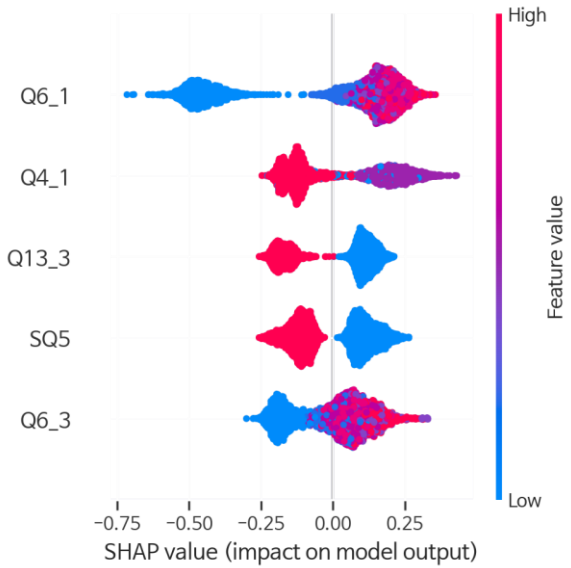
3 2차 예방 가해 학생 분류 모델링

모델 생성 및 중요도 시각화

[모델 생성 및 영향을 주는 변수: Top 10]



[변수들의 영향도: Top 5]



다음과 같을 때, 가해 행동 경험 있음에 영향

- Q6_1: 온라인에서 **폭력적이고 잔인한 컨텐츠를 자주 볼수록** (빨간색)
- Q4_1: **상대방을 욕하거나 감정을 상하게 하는 행동이 문제가 되지 않는다고 생각할수록** (파란색)
- Q13_3: **부모님이 온라인에서 개인 정보 보호 및 관리에 대해 알려준 적 없을 때** (파란색)
- SQ5: **남성**일 때 (파란색)
- Q6_3: **온라인에서 유명인을 헐뜯는 컨텐츠를 자주 볼수록** (빨간색)

- Catboost classifier 모델 사용

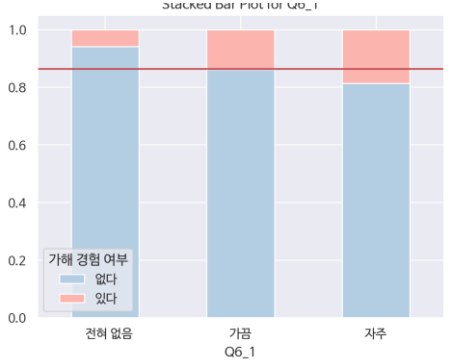
  → 가해 청소년 10명 중 약 6.5명을 예측할 수 있었음

3 2차 예방 가해 학생 분류 모델링

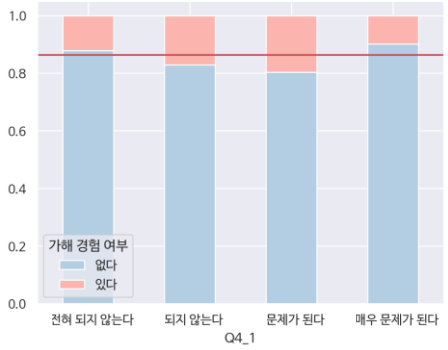
분석 결과

[시각화: stack bar plot]

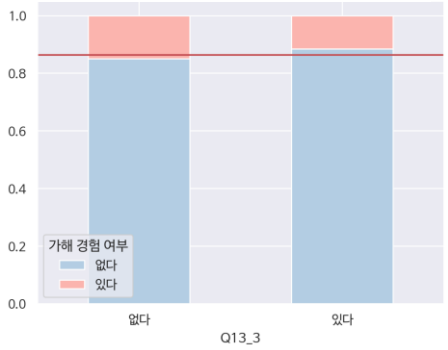
온라인에서 폭력적이고 잔인한 컨텐츠 시청 빈도



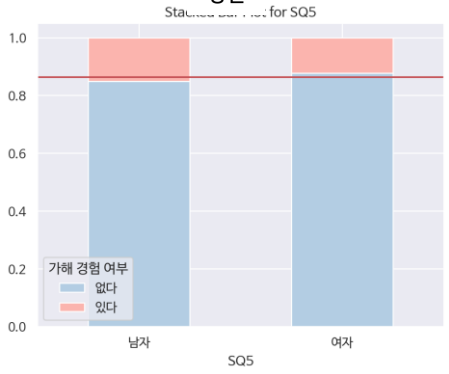
상대방을 욕하거나 감정을 상하게 하는 행동이 문제가 되는지



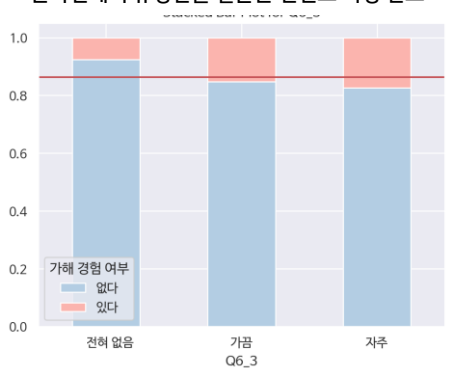
온라인에서 개인 정보 보호 및 관리에 대한 부모님의 교육 여부



성별



온라인에서 유명인을 헐뜯는 컨텐츠 시청 빈도



[카이제곱 통계 검정 결과]

feature	chi2_val	p_value	dof
Q6_1	231.6947	0.0000	2
Q4_1	169.6199	0.0000	3
Q13_3	21.0764	0.0000	1
SQ5	16.0489	0.0001	1
Q6_3	150.2465	0.0000	2

통계 검정

- 앞선 중요 변수들이 실제로 가해 경험 여부에 영향을 주는지 통계 검정(Chi-squared) 실시
- 앞선 shap value의 결과와 비슷한 양상을 보임

시각화와 통계검정 결과에 따르면,
다음 특성과 **가해 경험 여부** 사이에는 **유의미한 관계**가 있다.

(p-value < 0.001)

- **폭력적이고 잔인한 컨텐츠 시청 빈도**
- **상대방을 욕하거나 감정을 상하게 하는 행동에 대한 인식**
- **온라인 개인 정보 보호 및 관리에 대한 부모님의 교육 여부**
- **성별**
- **온라인에서 유명인을 헐뜯는 컨텐츠 시청 빈도**

3 2차 예방 가해 학생 분류 모델링

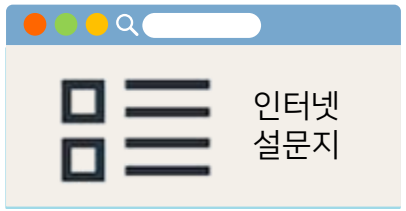
정책 제안

- 학교 폭력을 연상할 수 없는 질문지로 가해 행동 위험군 예측 가능
- 예측 모델의 중요 변수들은 통계적으로 유의미한 결과를 가짐

→ 위험군의 특성 파악 가능, 지도 방향 제시

제 안

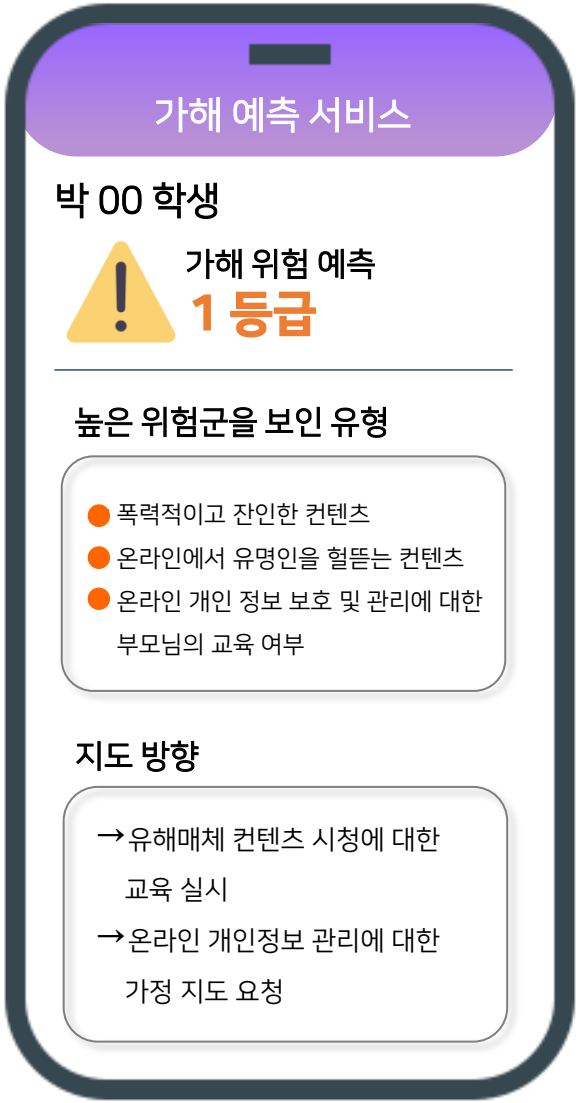
“ 주기적인 설문조사를 통한 가해 학생 모니터링 ”



청소년 사이버폭력
교내/학급내 설문조사 진행

잠재 가해 예측 모델
(Catboost)

가해 위험군 주시

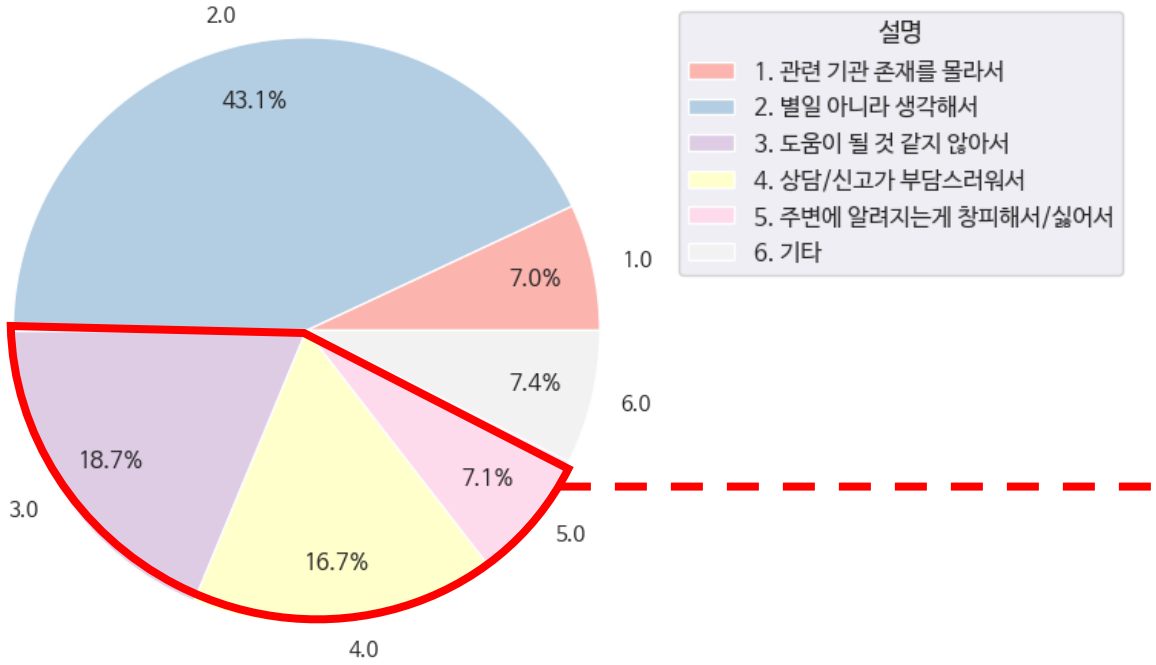


3차 예방

피해 경험 학생이 관련 기관에 도움 요청 하지 않은 이유 분석

분석 결과 : 청소년 사이버 폭력 실태 조사

피해 경험자 중 관련 기관에 도움 요청하지 않은 가장 큰 이유



응답별 비율을 시각화한 결과에 따르면,

도움이 될 것 같지 않아서라는 응답이 18.7%,
상담이나 신고가 부담스럽다는 응답이 16.7%,
주변에 알려지는 게 창피해서/싫어서라는 응답이 7.1%

제 안

부담스러운 정보 노출을 꺼려하는 청소년들의 부담을 덜어주기 위한 챗봇 상담 서비스 제안

3차 예방

챗봇 상담 서비스 구현 과정

1. 데이터 수집 및 전처리

네이버 지식인 Q & A 데이터

AI hub 감성 대화 말뭉치

2. 챗봇 학습

Ko-gpt2
: 질문에 대한 답변 생성

SentenceBERT
: 매끄러운 답변을 위해
유사한 문장 제외

3. 서비스 구현

Streamlit

Amazon EC2

Streamlit과
AWS EC2를 이용한
챗봇 웹사이트 구현

3 3차 예방 (1) 지식인 Q&A 데이터

1 Selenium을 이용해 '사이버 폭력 ' 을 키워드로 검색하여 질문, 답변 데이터를 크롤링

2 피해자 질문 선별 후 질문, 답변 내용 요약 작업

3 띄어쓰기 검사 & 맞춤법 검사

다음 기관의 답변 수집

- **교육부 학교폭력/성폭력 상담**
- **청소년폭력예방재단**
- **서울시청소년지원센터,**
- **한국청소년상담복지개발원**

	제목	질문	답변
2	보이스 피	보이스 피	안녕하세요. 청소년사이버상담센터입니다.청소년 사이버 폭력 관련해서 소개할 게시물이 있어서 글
3	학교폭력	친구를 뒷	안녕하세요. 청소년사이버상담센터입니다.청소년 사이버 폭력 관련해서 소개할 게시물이 있어서 글
4	사이버폭	안녕하세	안녕하세요. 한국청소년상담원입니다. 비공개님의 친구가 사이버폭력을 당했는데학교에서는 별다른
5	사이버 폭	이거 사이	안녕하세요. 청소년사이버상담센터 입니다.친구와의 카톡 대화가 사이버폭력에 해당하는지 여부에 대
6	사이버폭	a 라는 친	안녕하세요. 청소년사이버상담센터 입니다.비공개님의 글 잘 보았습니다.사이버폭력 신고와 처벌에 관
7	사이버 폭	제가 오늘	안녕하세요. 청소년사이버상담센터입니다.올려주신 글 잘 읽어 보았어요.질문자님이 히메컷을 한 것
8	사이버 폭	제가 최근	안녕하세요, 청소년사이버상담센터입니다[redacted]이 적어주신 글 잘 읽어보았습니다. 단톡방에 초
9	이거 사이	모르는 사	안녕하세요. 청소년사이버상담센터 입니다.사이버폭력에 해당되는지 여부에 관해 질문글 올려 주셨네
10	이것도 사	제가 큰 실	안녕하세요. 청소년사이버상담센터 입니다. 만나서 반가워요~^^친구랑 싸우고 나서, 답답한 마음을
11	학교폭력	저는 고등	안녕하세요. 청소년사이버상담센터 입니다.비공개님의 질문이 적힌 글을 잘 보았습니다.학교폭력과 관
12	카톡 상의	제가 6학년	안녕하세요. 청소년사이버상담센터 입니다.비공개님의 글 잘 보았습니다.사이버폭력 신고와 처벌에 관
13	사이버폭	안녕하세	안녕하세요.청소년사이버상담센터입니다. oo님께서 올려주신 고민 잘 읽어보았어요.카카오 톡으로 욕
14	사이버폭	제가 뿌린	안녕하세요. 청소년사이버상담센터 입니다.비공개님의 글 잘 보았습니다.사이버폭력 처벌에 관해 질문
15	사이버 폭	카톡상으로	안녕하세요. 청소년사이버상담센터 입니다.친구와의 카톡 대화가 사이버폭력에 해당하는지 여부에 대

총 1049개 데이터 수집

- 사이버 폭력 피해자를 위한 챗봇
  -> 피해자 상담 데이터만 추출
- 질문, 답변 내용을 적절한 길이로 정리

	Q
수정 전	사이버 폭력 가해자가 됨 제가 15살인데 이제 중학교 올라오는 14살 애가 단체 메신저 방에서 메신저에 질문 하나가 올라왔는데 그게 선배 하나 테 ㅈ. 나 대네 이거였었는데 그걸 06일 걸이 라면서 저희를 불특정 다수로 정해놓고 찍혀도 상관없어 어차피 인생 좆 돼서? 그럼 식으로 말해서 제가 거기 있는 다른 후배한테 초대해 달라 해서 뭐라고 했거든요? 찍혀도 상관없어? 뭔 다 듣고야 나 많으면 다 백이라 나대는 거야? 싸가지 챙겨 영 아 네가 07이미지 다 깎아내리잖아 이 는 식으로 했거든요 근데 걔네 어머님 보셨는지 집단 폭행이라고 위화감과 불이익과 무서움을 줬고 괴로움을 줬고 뭐 그 대상도 아니면서 상처 준 너희들 가만 볼 수 없다며 내일 경찰서와 학교에 신고하겠다는 대 처벌을 받는 디면 어떤 처벌을 받을 것이고 어떻게 처리가 될까요? 부모님 소환이나 그런 사항 자세하게 적어주세요
수정 후	사이버 폭력 가해자가 됨 제가 15살인데 메신저로 14살 후배에게 찍혀도 상관없어? 싸가지 챙겨 병신아 네가 07이미지 다 깎아내리잖아 이랬거든요. 근데 걔네 어머님이 보셨는지 집단 폭행이라고 위화감과 불이익과 무서움을 줬고 괴로움을 줬다고 내일 경찰서와 학교에 신고하겠다는데 어떤 처벌을 받을 것이고 어떻게 처리가 될까요?

띄어쓰기 검사

Pykospacing 패키지

맞춤법검사

Request를 이용한
부산대 맞춤법 검사기

```
def busan_correct(text):
    text = text.replace('\n',' ')
    response = requests.post('http://164.125.7.61/speller/results', data={'text1': text})
    data = response.text.split('data = [', 1)[-1].rsplit('];', 1)[0]
    if '맞춤법과 문법 오류를 찾지' in data or '결과를 받지 못했습니다.' in data:
        return text
    data = json.loads(data)

    flag = 0

    for err in data['errInfo']:
        start=err['start']
        end=err['end']
        before = text[start+flag : end+flag]
        if before.isalpha(): #영어는 교정하지 않음
            continue
        after = err['candWord'].split('|')[0]

        #print(f'{before}->{after}_')

        text = text[:start+flag] + after + text[flag+end:]

        if len(before) < len(after): #바뀐 후의 자릿수가 더 많으면 그만큼 더해줘야함
            flag += (len(after) - len(before))
        else: #바뀌기 전의 자릿수가 더 많으면 flag를 빼줘야함
            flag -= (len(before) - len(after))

    return text
```

3차 예방 (2) AI hub 감성 대화 말뭉치

1. 데이터 수집 및 전처리

AI hub 감성 대화 말뭉치

일반인 1,500명을 대상으로 하여
음성 15,700문장 및 코퍼스 27만 문장 구축 및
세대별 감성 대화 텍스트 구축을 통해
감성 대화 엔진을 개발하여 세대별 감성 대화 서비스 제공

<데이터 예시>

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		연령	성별	상황키워드	신체질환	감정_대분류	감정_소분류	사람문장1	시스템문장	사람문장2	시스템문장	사람문장3	시스템문장3	
2	1	청년	남성	진로,취업,	해당없음	불안	두려운	이번 프로	실수하시다	내 능력이	능력을 올	퇴근 후 여	꼭 좋은 결과 있길 바라요.	
3	2	청년	남성	진로,취업,	해당없음	불안	두려운	회사에서	큰 프로젝	나에게 너	프로젝트를	동료 직원(	동료 직원에게 도움을 요청	
4	3	청년	남성	진로,취업,	해당없음	불안	두려운	상사가 너	직장 상사	무섭게 생	상사분과	먼저 다가	직장 상사와 친해지시면 좋	
5	4	청년	남성	진로,취업,	해당없음	불안	두려운	이번에 힘	첫 직장이	첫 직장이(	잘 적응 하	직장 동료	직장에 잘 적응하시길 바라	
6	5	청년	남성	진로,취업,	해당없음	불안	두려운	직장에서	직장 사람	내가 낯가	직장 사람(	서로 같은	직장 사람들과 좋은 관계를	
7	6	청년	남성	진로,취업,	해당없음	불안	두려운	내가 평소	팀장님 앞(	팀장님 앞(	말실수 안	평소에 말	팀장님 앞에서 실수를 안 ㅎ	
8	7	청년	남성	진로,취업,	해당없음	불안	두려운	내 직급이	의견을 무	무엇 때문(	의견을 제	집에서 미	회의 시간에 의견을 내셔서	
9	8	청년	남성	진로,취업,	해당없음	불안	두려운	부장님께	실수를 하	내가 한 실	큰 피해가	실수한 부	좋은 결과가 나오길 바라요	
10	9	청년	남성	진로,취업,	해당없음	불안	두려운	내일 회사	틱 증상 때	내가 틱 증	틱 증상이	솔직하게 (	자신감 있게 임하셔서 좋은	
11	10	청년	남성	진로,취업,	해당없음	불안	두려운	친구랑 같(	친구분이랑	난 그 친구	안 멀어지	앞에서 회	친구분도 취업하셔서 사이	
12	11	청년	남성	진로,취업,	해당없음	불안	두려운	내가 새로	회사 적응(	전 회사에/	회사에 잘	회사 동료	회사에 잘 적응하시길 바라	
13	12	청년	남성	진로,취업,	해당없음	불안	두려운	평소에 날	팀장님과으	왜 평소에	팀장님에게	좋은 아이(	좋은 아이디어로 인정받으	
14	13	청소년	남성	학교폭력/	해당없음	슬픔	마비된	학교에서 (	그 상황을	아이들이 (	이유를 모	아이들 중	대화가 잘 풀려서 이 상황이	

1 연령 : 청소년
상황 키워드 : 학교폭력 / 따돌림
데이터만 필터링하여 사용

2 질문 내용을 반복하여 답하는 데이터는 제외하였음

<제외 데이터 예시>

Q	A
반 애들이 나만 따돌려서 너무 화가 나.	친구들이 따돌려서 화가 나시는군요.
친구들이 내 말을 무시하는 게 화가 나.	친구들이 말을 무시해서 정말 화가 나겠어요.
친구들이 내 이름을 가지고 놀리는 게 짜증 나.	친구들이 이름을 가지고 놀려서 짜증 나시는군요.
애들이 우리 부모님까지 욕하는 게 너무 화가 나.	아이들이 부모님을 욕해서 화가 나셨군요.

**'지식인 데이터 ' 와 'AI hub 감성 대화 말뭉치'
데이터를 병합하여 총 6,614개 데이터 수집**

```
Chatbot_Data
```

	Q	A
0	사이버 폭력이 얼굴도 이름도 모르는 사람이 저에게 욕하고 기분 나쁜 말을 쓰고 그런...	사이버폭력에 대해 자세하게 알고 싶어 하시는군요. 사이버 따돌림 이란 인터넷 휴대전...
1	특정 연예인을 욕하는 글을 올리는 것도 사이버 폭력이 맞나요?	특정 연예인에 대한 비방 욕설 글의 게시가 사이버폭력에 해당하는지 궁금하시군요. 질...
2	제가 사이버 폭력을 당했는데요. 아이들이 저를 욕하고 제 허락도 없이 제 이름을 인...	학교폭력으로 신고된 사안에 대하여 이후 사과가 이루어진것이 아니라면 즉 학교폭력으로...
3	사이버 폭력 신고할 수 있나요 제가 사이버 폭력을 당했는데요 아이들이 저를 욕하고 ...	신고는 가능합니다. 다만 학교폭력 행위에 가담하고 방관했다는 명확한 증거가 없어 사...
4	사이버 폭력 신고할 수 있나요 제가 사이버 폭력을 당했는데요. 아이들이 저를 욕하고...	학교폭력으로 신고하게되면 의무적으로 관련 학생들의 보호자에게 통보하게 되어 있습니다...
...	...	...

3차 예방 (2) AI hub 감성 대화 말뭉치

2. 챗봇 학습

질문에 대한 답변 생성 : KoGPT2

- 주어진 텍스트의 다음 단어를 잘 예측할 수 있도록 학습된 언어모델으로 문장 생성에 최적화 되어 있음
- 부족한 한국어 성능을 극복하기 위해 40GB 이상의 텍스트로 학습된 한국어 디코더(decoder) 언어모델

전처리한 데이터 학습

출력

답변</s>

Ko-GPT2

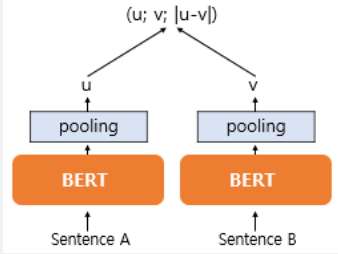
입력

<usr>질문<sent> <sys>답변</s>

- Maxlen(최대 단어 수) : 300
- Epoch : 8

<special token>
- usr : 질문의 시작을 알림
- sent : 질문이 끝났음을 알림
- sys : 답변의 시작을 알림
- /s : 문장이 끝났음을 알림

매끄러운 답변을 위해 유사한 문장 제외 : Ko-SentenceBERT




```
>>> from sentence_transformers import SentenceTransformer, util
>>> model = SentenceTransformer('snunlp/KR-SBERT-V40K-klueNLI-augSTS')
>>> sentences = ['잠이 옵니다', '졸음이 옵니다', '기차가 옵니다']
>>> vectors = model.encode(sentences) # encode sentences into vectors
>>> similarities = util.cos_sim(vectors, vectors) # compute similarity
>>> print(similarities)
tensor([[1.0000, 0.6577, 0.2732],
        [0.6577, 1.0000, 0.2730],
        [0.2732, 0.2730, 1.0000]])
```

- BERT의 문장 임베딩 성능을 개선
- KLUE-NLI와 KorSTS dataset으로 Fine-tuning된 모델 사용
- 문장들 간의 코사인 유사도를 계산하여 의미가 비슷한 문장을 제외하고 출력할 것임

유사도가 0.7이상인 문장은 제외하고 출력

Before	After
기분이 나쁘셨군요. **기분이 안 좋으셨군요.** 어떻게 하면 기분 전환이 될까요?	기분이 나쁘셨군요. 어떻게 하면 기분 전환이 될까요?

3차 예방 (1) 웹 애플리케이션 구축

✔ 챗봇 서비스 구현 과정 : ⓐ Streamlit을 이용한 웹 애플리케이션 구축 → ⓑ 웹 상에 챗봇을 배포하기 위한 서버 구축

ⓐ 웹 애플리케이션 구축

1 Hugging Face에 학습시킨 모델 업로드

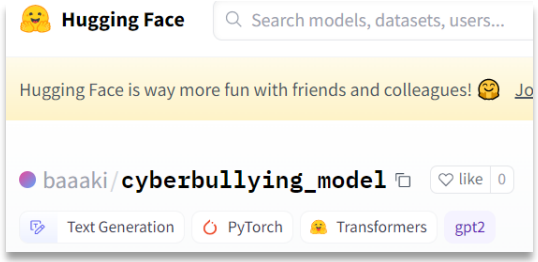
Hugging Face :
Transformers 모듈을 사용하여 자신만의 모델을 만들고,
훈련 및 평가할 수 있는 환경을 제공

업로드

```python
# Repository 생성 & model upload
REPO_NAME = 'cyberbullying_model'
AUTH_TOKEN = [redacted]

## Upload to Huggingface Hub
model.push_to_hub(
    REPO_NAME,
    use_temp_dir=True,
    use_auth_token=AUTH_TOKEN
)
koGPT2_TOKENIZER.push_to_hub(
    REPO_NAME,
    use_temp_dir=True,
    use_auth_token=AUTH_TOKEN
)
```

업로드 확인



2 Streamlit을 이용해 웹페이지 만들기

Streamlit

파이썬으로 애플리케이션을 쉽게 구축할 수 있도록 도와주는 오픈소스 웹 애플리케이션 프레임워크

1. Streamlit 을 실행할 python 코드 작성

Hugging Face에 업로드했던 모델을
Transformers 패키지를 이용해 불러온다

```python
from transformers import AutoTokenizer, AutoModelForCausalLM, pipeline
tokenizer = AutoTokenizer.from_pretrained("baaaki/230506_8")
model = AutoModelForCausalLM.from_pretrained("baaaki/230506_8")

from sentence_transformers import SentenceTransformer
similarity = SentenceTransformer('snunlp/KR-SBERT-V40K-klueNLI-augSTS')
return model, tokenizer, similarity
```

명령어를 통해 홈페이지를 구성한다

```python
st.header("😈CyberBullying Chatbot (Demo)")

with st.spinner("loading model..."):
    model, tokenizer, similarity = get_pipe()

if 'message_history' not in st.session_state:
    st.session_state.message_history = []
history = st.session_state.message_history
```

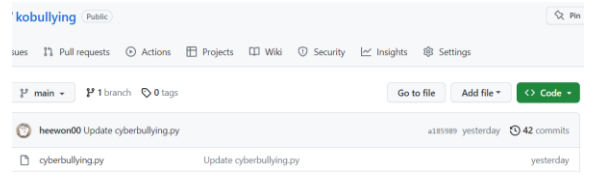
2. Localtunnel 을 통해 임시 웹페이지를 만들어 테스트

```
!streamlit run cyberbullying.py & npx localtunnel --port 8502
```

```
your url is: https://eager-moles-exist-35-204-19-22.loca.lt
```

Localtunnel을 통해 내부망 PC의 Localhost를
외부 인터넷 망에 노출시킬 수 있다.

3. 이후 서버 구축을 위해 작성한 python 코드 파일을 github에 업로드



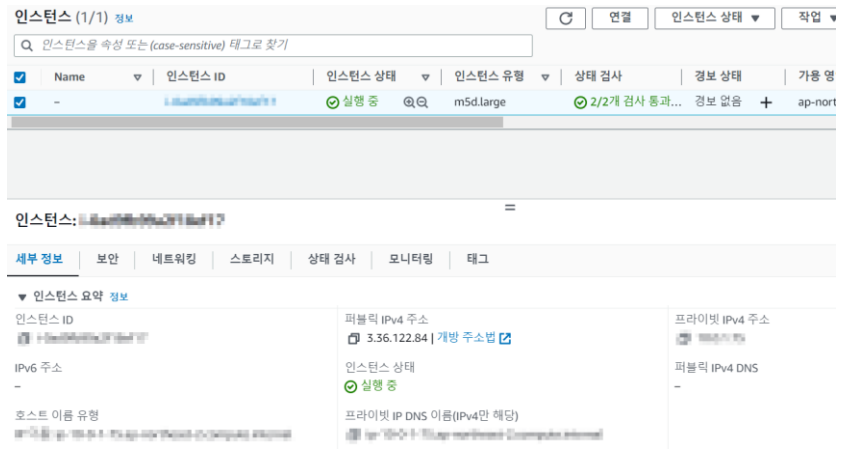
3차 예방 (2) 웹 상에 챗봇을 배포하기 위한 서버 구축

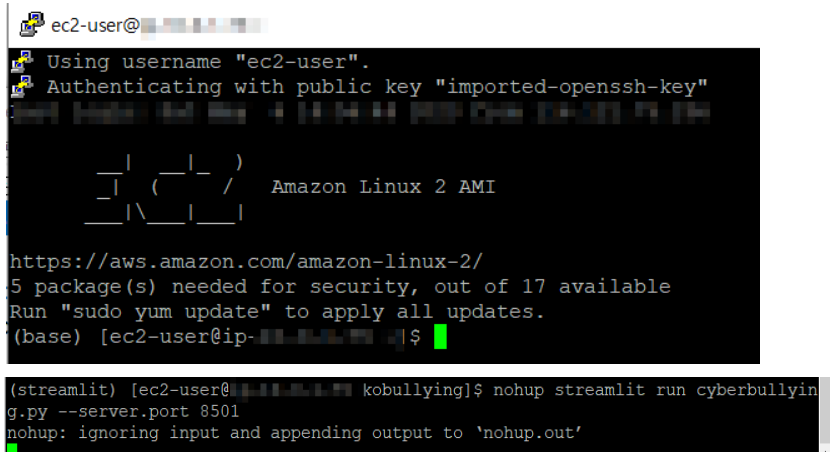
ⓑ 웹 상에 챗봇을 배포하기 위한 서버 구축

챗봇이 정상적으로 24시간 작동하기 위해 항상 실행되고 있는 연동된 서버가 필요함

Amazon EC2

클라우드에서 가상 서버를 제공하는 웹 서비스. 서버 구축을 위해 사용

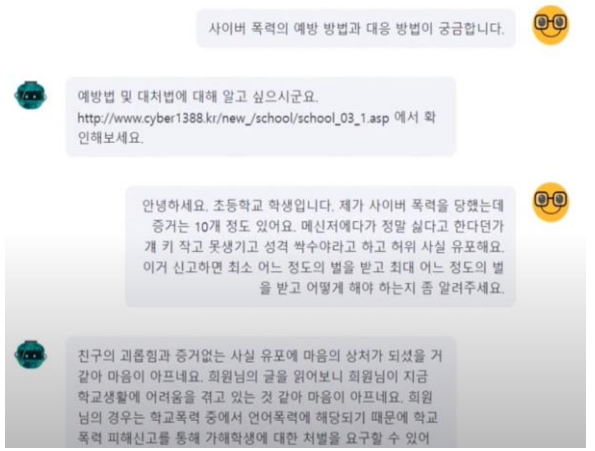




챗봇 서버용 AWS EC2 개설 후 Streamlit 호스팅

시연 영상

간단한 문장들과 함께
실제 지식인에 올라왔던 질문들을
입력해보았다.



https://youtu.be/sifodF_AUy8

4

기대효과

4 기대효과 1, 2, 3차 예방 측면

1차 예방

✔ 제안

데이터 분석을 통한 사이버 폭력 예방을 위한 교육 전략
- 사이버 폭력 신고 인식 교육 집중 개선
- 유해 컨텐츠 노출 피해 예방 교육 확대

✔ 기대효과

- 사이버 폭력 예방에 효과적인 주제의 교육을 집중 개선 · 확대함으로써 효율적인 예방 교육 구성 가능
- 청소년들이 사이버 공간에서의 안전한 활동 가능

2차 예방

✔ 제안

가해 청소년 분류 모델을 이용한 조기 발굴 및 모니터링

✔ 기대효과

- 효과적으로 학생들의 사이버 폭력 현황 모니터링 가능
- 사이버 폭력 위험군 조기 발견 및 지도

3차 예방

✔ 제안

피해자들을 위한 사이버 폭력 상담 챗봇

✔ 기대효과

- 상담/신고에 대한 부담 감소
- 피해자들이 더욱 다양한 방법으로 도움 요청 가능
- 상담사 인력 보충 : 초보적인 상담은 인공지능 상담챗봇에 맡기고, 보다 상세하고 전문적인 상담에 집중할 수 있음

▶▶ 청소년 사이버폭력 관련 상담의 질적 제고 기대

You are not alone.

감사합니다.