

텍스트 분석을 통한 사이트

범죄 위험도 분석

TEAM 5 박준수 신우현 박소영 이은강

2027.03.15

Contents

1. Background

2. Topic

3. Datasets

4. Pipeline

5. Prototype

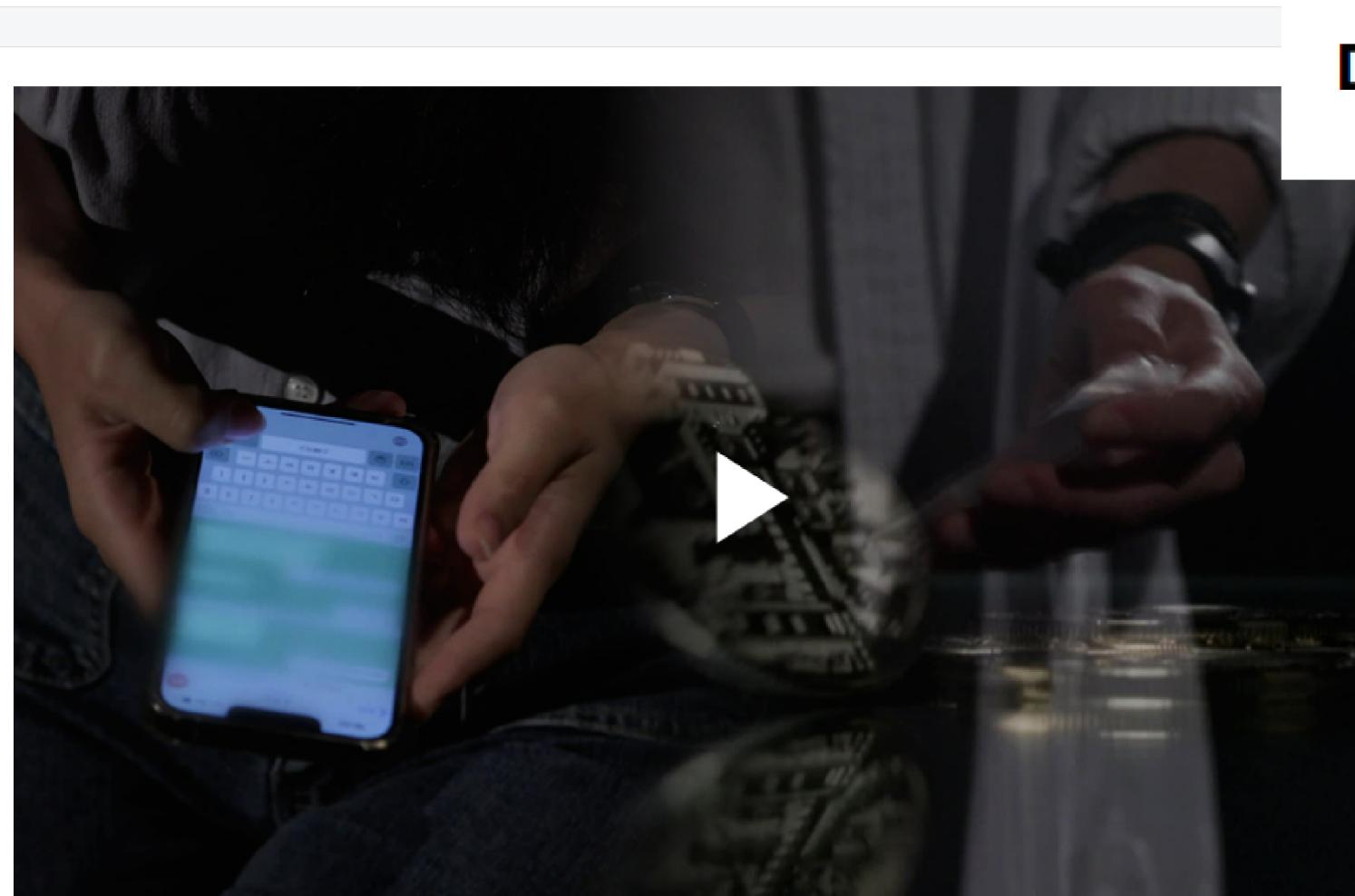
1. Background

뉴스 7

다크웹에서 텔레그램으로 숨어드는 마약…대부분은 20~30대

입력 2024.02.28 (19:20) | 수정 2024.02.28 (19:46)

▶ 요약



마약사범 5명중 1명이 SNS로 거래…10대 사범도 매년 증가

(브뤼셀=연합뉴스) 정빛나 특파원 = 흔적이 남지 않는 이른바 '다크웹' 판매사이트를 활용해 마약을 거래한 관련자들이 다국적 공조 수사를 통해 무더기 검거됐다.

EU 경찰조직인 유로폴은 2일(현지시간) 9개국과 공동으로 다크웹을 기반으로 한 '모노폴리마켓' 명칭의 판매사이트에 대한 수사를 벌여 해당 사이트에서 마약 거래를 한 용의자 288명을 체포했다고 밝혔다.

현금 및 가상화폐 5천80만 유로(약 749억원)와 마약 850kg, 총기 117정 등도 압수했다.

1. Background

사건/사고 ▶ 텔레그램 n번방 성범죄 +

"야한 채팅할 사람"...미성년자 일상으로 파고드는 '그루밍'

CBS노컷뉴스 이은지·서민선·박하얀 기자 ☎ | 2020-06-24 05:05

[조주빈 검거 100일, 우리 사회 달라졌다②]

청소년 10명 중 1명은 온라인서 '성적 유인' 경험

카톡·SNS·게임 통해 접근..."용돈·아이템 주겠다"

'온라인 그루밍'의 시작..."n번방" 성착취로 연결

현행법상 처벌 어려워..."그루밍 방지법 제정해야"

'박사' 조주빈이 검거된 지 100일이 넘었다. 온라인에서 은밀히 자행되던 성착취 범죄에 경악한 우리 사회는 이후 술한 대책을 쏟아냈지만, '실제' 변화가 있었는가를 묻는다면 확신에 찬 답은 어려워 보인다. CBS노컷뉴스는 연속 기획으로 '디지털 성범죄'의 현 상황을 점검하고, 이를 뿌리 뽑기 위한 앞으로의 과제를 짚어봤다.[편집자주]



1. Background

문제상황: 온라인 환경에서의 마약 및 성매매 범죄율 증가

온라인 마약 거래 통계

2023년 총 2만 6천 13건

트위터 (1만4천779건)

텀블러 (578건)

페이스북 (84건)

구글 (42건)

네이버, 카카오 (16건)

청소년 온라인 성매매 통계

최근 5년간 아동·청소년 성 착취 연도별 피해자 현황

(단위: 명, %, 출처: 한국여성인권진흥원디지털성범죄피해자지원센터)

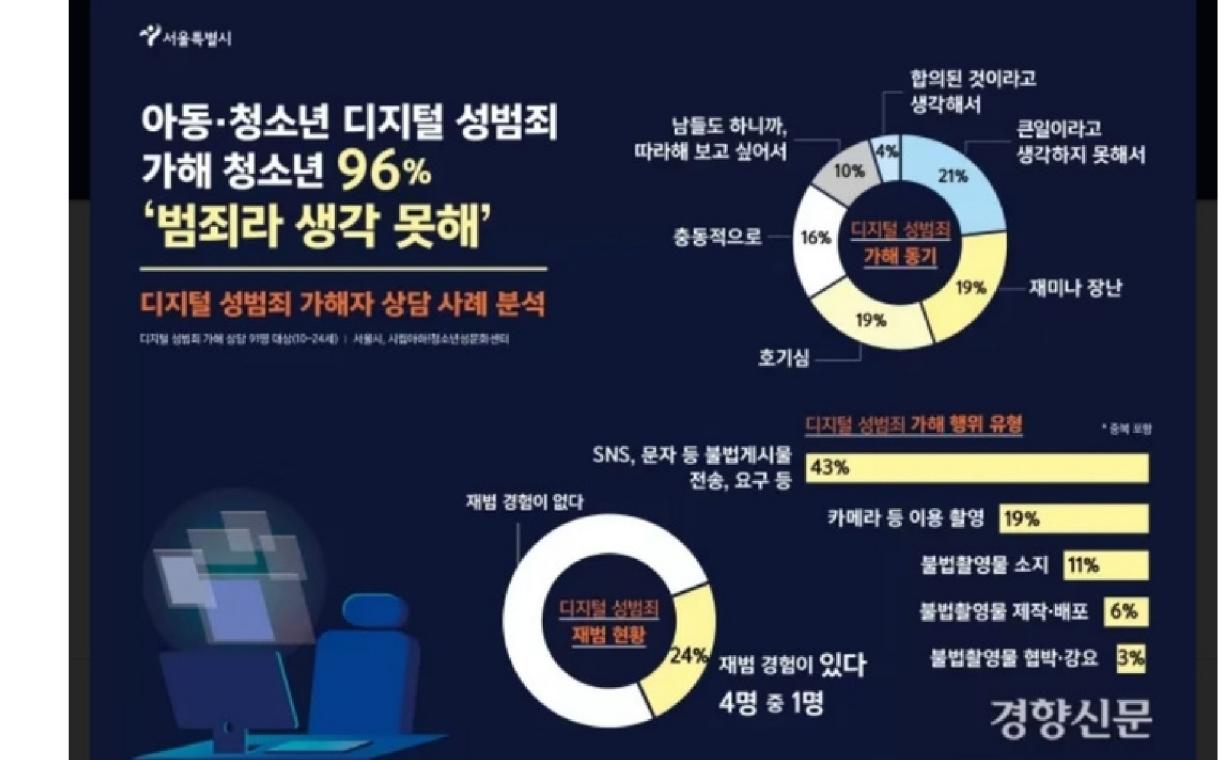
구분	2018년	2019년	2020년	2021년	2022년	합계
계	59(2.9)	99(4.8)	541(26.4)	721(35.2)	629(30.7)	2049(100.0)
14세 미만	14(3.4)	11(2.6)	91(21.7)	134(32.0)	169(40.3)	419(100.0)
14세~19세 미만	45(2.8)	88(5.4)	450(27.6)	587(36.0)	460(28.2)	1630(100.0)

그래픽 문자원

아동청소년에 대한 온라인 그루밍 성착취 피해

2020년 이후 14세 미만 아동 피해자 급증

2020년 이후 14세 ~19세 미만 피해자 400건대 돌파



2. Topic : 사이버 보안에 최적화된 언어 모델 학습

- 미숙한 청소년들을 대상으로 한 온라인 마약거래, 성범죄를 사전 방지하여 피해자를 줄이기 위함.
- 미숙한 청소년 가해자들에게 범죄라는 사실을 각인시킬수 있음

주제 선정 이유

목표

- 온라인 범죄와 관련된 게시물, 채팅에 노출시 자동으로 경고해주는 서비스

활용방안

차별성

- 미숙한 청소년들을 대상으로 한 온라인 마약거래, 성범죄 관련 게시물, 채팅을 분석하여 유의한 범죄 위험도를 나타내주는 것이 목표

온라인 그루밍 피해자 영상, 22초만에
찾아…서울시, AI로 새벽에도 감시

-
- 텍스트 마이닝 기법을 통해 마약, 성범죄 관련 게시물, 채팅을 사전에 차단, 경고 할 수 있음

3.Datasets

마약 통계 데이터

-다크웹 사이트 내 마약 관련 출현 단어를 월별로 보여줌

순번	포럼명	단어명	게시시작년월일	게시종료년월일	단어빈도통계값	문서빈도통계값	단어역문서빈도통계값
1	highkorea	대마초	20141001	20141031	8	7	22
2	highkorea	중독	20141001	20141031	5	4	16
3	highkorea	대마초	20141101	20141130	50	2	203
4	highkorea	대마초	20141201	20141231	32	15	60
5	highkorea	수꽃	20141201	20141231	10	2	35

컬럼 정의서

순서	컬럼영문명	컬럼한글명
1	SN	순번
2	FORUM_NM	포럼명
3	WRD_NM	단어명
4	NTCE_STRT_YMD	게시시작년월일
5	NTCE_END_YMD	게시종료년월일
6	WRD_FQ_STTST_VL	단어빈도통계값
7	DOC_FQ_STTST_VL	문서빈도통계값
8	DFIDF_FQ_STTST_VL	단어역문서빈도통계값

1. 순번
2. 포럼명
3. 단어명
4. 게시시작년월일일
5. 게시종료년월일
6. 단어빈도 통계값
7. 문서빈도 통계값
8. 단어역문서빈도 통계값

3.Datasets

성범죄 관련 단어 데이터

-다크웹 사이트 내 성범죄 단어가 포함된 게시물을 보여줌

순번	단어명	예시문1	예시문2	예시문3	예시문4	예시문5		
1	강간	rixtxri745dfw.onion/neo/uploads/200420/MATRIX_0945 ... #736 2019-2-2 오후 3:50 2576 연예인 성폭행 강간 그런거 나오면 여자는왜 ... 안 보고 금발 로리 + 동아시아 로리 선호. 화질 구리면 아무리 내용이 좋아도 칼삭제함. ... 구매했거나 Tor브라우저로 만들었고, 그 계정 무조건 Tor브라우저로 로그인/로그아웃?	r15933 2019-9-23 오전 5:28 44455 베이비 박스에 아기 버리는 김치년들 보면 아까워 죽 ... 결리면 한 4년 정도 뜨겠네. 한국은 이상하게 신생아 목숨은 죽도 아닌걸로 보고 2019-9-22 오전 11:08 납치할꺼면 조선족을 납치해야함 이년들 불체자 ... 이는 아닌것같은 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ	#15933 2019-9-23 오전 5:28 44455 베이비 박스에 아기 버리는 김치년들 보면 아까워 죽 ... 결리면 한 4년 정도 뜨겠네. 한국은 이상하게 신생아 목숨은 죽도 아닌걸로 보고 2019-9-22 오전 11:08 납치할꺼면 조선족을 납치해야함 이년들 불체자 ... 이는 아닌것같은 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ	019-2-2 오후 5:28 44455 베이비 박스에 아기 버리는 김치년들 보면 아까워 죽 ... 결리면 한 4년 정도 뜨겠네. 한국은 이상하게 신생아 목숨은 죽도 아닌걸로 보고 2019-9-22 오전 11:08 납치할꺼면 조선족을 납치해야함 이년들 불체자 ... 이는 아닌것같은 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ	... #28842 2020-4-29 오전 10:13 세븐나이츠라는 찾아보고있는데 라노그 마	#21995 2020-2-9 오전 12:58 영상 #9883 2019-2-2 오후 5:54 49482 레이프를 모아놓은곳 없나 작성	이나 먹어라. 2 너 성범죄자알림E 0 올라온 놈 이지? 뚱꼬털 많이 있게
2	납치	으로 있는 애들 강간하고 돈 뺐고 사진 촛나 촬영해 놓으면 신고 못하지 않을까? 우리나라에 ... #15933 2019-9-23 오전 5:28 44455 베이비 박스에 아기 버리는 김치년들 보면 아까워 죽 ... 결리면 한 4년 정도 뜨겠네. 한국은 이상하게 신생아 목숨은 죽도 아닌걸로 보고 2019-9-22 오전 11:08 납치할꺼면 조선족을 납치해야함 이년들 불체자 ... 이는 아닌것같은 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ	#15933 2019-9-23 오전 5:28 44455 베이비 박스에 아기 버리는 김치년들 보면 아까워 죽 ... 결리면 한 4년 정도 뜨겠네. 한국은 이상하게 신생아 목숨은 죽도 아닌걸로 보고 2019-9-22 오전 11:08 납치할꺼면 조선족을 납치해야함 이년들 불체자 ... 이는 아닌것같은 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ	019-2-2 오후 5:28 44455 베이비 박스에 아기 버리는 김치년들 보면 아까워 죽 ... 결리면 한 4년 정도 뜨겠네. 한국은 이상하게 신생아 목숨은 죽도 아닌걸로 보고 2019-9-22 오전 11:08 납치할꺼면 조선족을 납치해야함 이년들 불체자 ... 이는 아닌것같은 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ	... #28842 2020-4-29 오전 10:13 세븐나이츠라는 찾아보고있는데 라노그 마	#21995 2020-2-9 오전 12:58 영상 #9883 2019-2-2 오후 5:54 49482 레이프를 모아놓은곳 없나 작성	이나 먹어라. 2 너 성범죄자알림E 0 올라온 놈 이지? 뚱꼬털 많이 있게	

컬럼 정의서

- 순번
- 단어명
- 예시문1
- 예시문2
- 예시문3
- 예시문4
- 예시문5
- 예시문6
- 예시문7
- 예시문8
- 예시문9
- 예시문10
- 예시문11
- 예시문12
- 예시문13
- 예시문14
- 예시문15
- 예시문16
- 예시문17
- 예시문18
- 예시문19
- 예시문20

순서	컬럼영문명	컬럼한글명
1	SN	순번
2	WRD_NM	단어명
3	EXSTC_1	예시문1
4	EXSTC_2	예시문2
5	EXSTC_3	예시문3
6	EXSTC_4	예시문4
7	EXSTC_5	예시문5
8	EXSTC_6	예시문6
9	EXSTC_7	예시문7
10	EXSTC_8	예시문8
11	EXSTC_9	예시문9
12	EXSTC_10	예시문10
13	EXSTC_11	예시문11
14	EXSTC_12	예시문12
15	EXSTC_13	예시문13
16	EXSTC_14	예시문14
17	EXSTC_15	예시문15
18	EXSTC_16	예시문16
19	EXSTC_17	예시문17
20	EXSTC_18	예시문18
21	EXSTC_19	예시문19
22	EXSTC_20	예시문20

4. Pipeline

데이터 수집 및 전처리

모델 선택

fine-tuning

모델 평가

사이버치안/범죄 2021-01-28 최초등록 / 2022-12-02 업데이트됨

마약 통계 데이터

#치안 #보안

사이버치안/범죄 2021-01-28 최초등록 / 2022-12-02 업데이트됨

성범죄 관련 단어사전

#치안 #보안

- 데이터 수집

봉달이, 떨	마약
ㅍㅌ닐, 코x인	마약
성x행	성범죄

- 카테고리 별로 라벨링
ex) 마약, 성범죄 등

4. Pipeline

데이터 수집 및 전처리

모델 선택

fine-tuning

모델 평가

- 각 단어의 파생어 여럿 존재
 - 마약 -> 떨, 드럭
- 일상적 단어가 은어가 되는 경우도 존재
 - 아이스, 캔디



Klue-BERT, KoBERT, KorBERT

4. Pipeline

데이터 수집 및 전처리

모델 선택

fine-tuning

모델 평가

- task
 - 해당 게시글, 대화의 주제 분류
 - 위험 단어 출력
 - 전체적인 위험도 분석

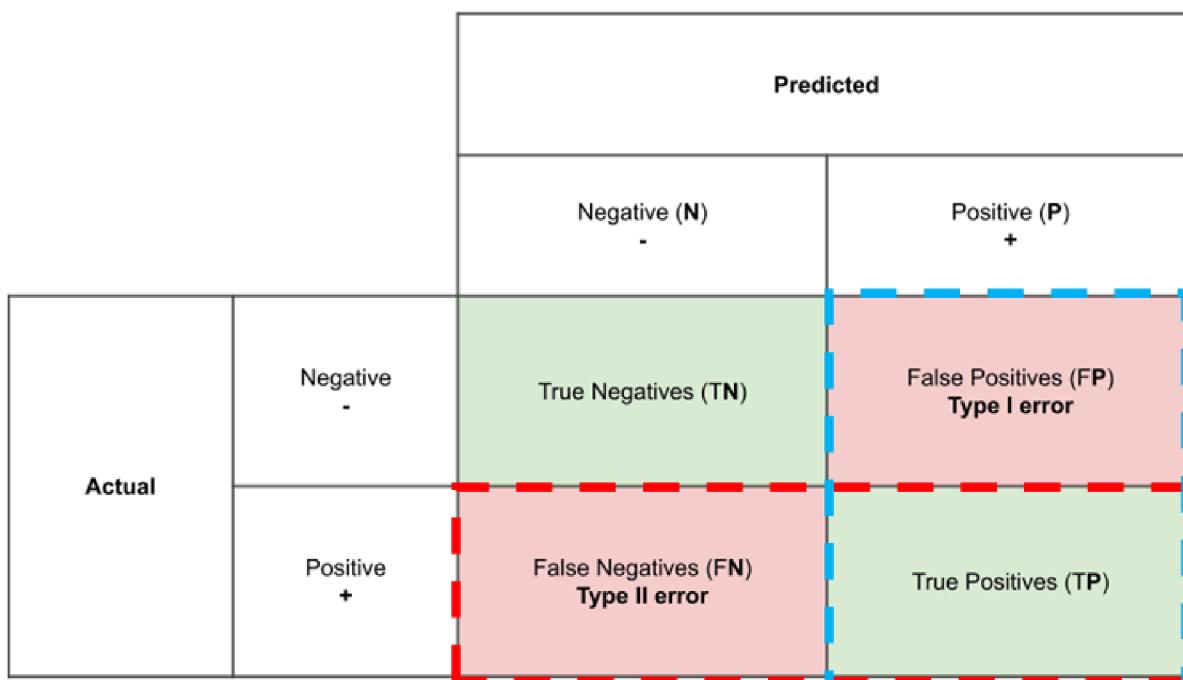
4. Pipeline

데이터 수집 및 전처리

모델 선택

fine-tuning

모델 평가



	our model	ChatGPT
주제 분석	주제 분석	주제 분석
위험 단어	위험 단어	위험 단어

- 평가지표 활용 평가

- ChatGPT 와 직접 비교

5. Prototype

게시물 분석

The screenshot shows two search results pages from the Ilbe website. The top page is for '일간베스트' (Best of the Day) and the bottom page is for '정치' (Politics). Both pages display a list of news items with dates, titles, and small thumbnail images. The news items are mostly in Korean and discuss various topics like politics, society, and culture.



범죄 예측 레포트

성범죄

마약거래

위험도 : 위험

토픽 분류

성범죄

마약 거래

거래

만남

...

범죄 예측 레포트

성범죄

마약거래

위험도 : 위험

토픽 분류

성범죄

마약 거래

거래

만남

...

5. Prototype

은어 표현 파악

차가운 술, 보리 캔디, 샘플 부터 주문해주세요.

차가운 술, 샘플 부터 주문해주세요.

유사도 비슷한 단어들로
은어의 뜻을 유추

Top 10

차가운 술	스몰	머핀	농구공	오키
크리스탈	뿔방	앵두	얼굴	방두

Thank You
