레이블링 분류 기준

1. 욕설

* 단어 기준

- 시발, 병신, 지랄, 염병, 좆까, 존나, 지랄,, 뒤질래, 꺼져, 닥쳐, 니미, 뻐큐, 씨댕, 엠창

(자음, 또는 글자 기준 조합으로 유추할 수 있으면 단어 기준에 충족함.

Ex. ㅅㅂ, ㅄ, 시1발, ㄲㅈ

* 조합 기준

- 개, 미친, 쓰레기, 걸레 같은 단어와 같이 쓰여서 욕이 되는 경우

개쓰레기, 미친년, 쓰레기놈, 개소리

2월 21일 / 2월 23일 / 2월 28일 / 3월 3일

1. 혐오, 비하, 악플

대상을 **심하게 비난하거나 깎아내려서 정신적인 고통과 같은 감정상태를 야기할 수 있는 표현**

**대상에게 적대감을 드러내는 표현.**

* 단어 기준

- 특정 사상, 집단, 인종, 종교를 지칭하는 단어가 단독으로 쓰일 경우는 제외

일베, 메갈, 오유, 루리웹, 엠팍, 특정사이트, 여야, 국힘, 민주,정의, 허경영, 흑형,

- 특정 대상을 혐오, 비하하기 위해 생성된 단어나 신조어는 포함하여 부정적 의미로 쓰인 경우

Ex. 더불어만진당 대깨문 니거 짱개 한녀 한남 ( 추가 정보가 필요하다 )

Ex. 더불어만진당이라 누가 만들었냐?? ( 비하의 의미로는 정보 부족 )

Ex. 더불어만진당 수준 여윽시 떨어지죠?? ㅋㅋㅋ ( 특정 집단을 비하 )

* 조합 기준

- 위의 단어에 추가로 쓰인 단어나 문장이 욕설 또는 비하의 의미가 될 때

Ex. 일베충, 민주당새끼들, 정의당 병신들,

* 문장 기준

- 단어의 조합이나 문장의 맥락으로만 파악이 가능한 것은 제외.

Ex. 혹시나 했더니 역시 그 나라.. K5는 역시 사이언스… 과학이다 여윽시…

- 비꼬거나 반어법으로 애매하게 표현한 것은 제외

Ex. 니가 그럼 그렇지..ㅋㅋ.. 아이구 정말 대단들하시네요~~ㅋㅋ

Ex. 끼리끼리 모여서 잘하는 짓이다. 축구는 박지성.. 인생은 박주영처럼~~

Ex. 역시는 역시다. 문재인은 과학!

1. 섹슈얼

대상을 **성적인 표현으로 모욕하거나 희롱하고, 성적 대상화**하는 경우

* 단어 기준

- 성희롱, 성차별, 성폭행, 성행위를 직접적으로 연상시키는 단어

Ex. 섹스, 씹질, 빠구리, 떡치다, 강간, 고자

- 단어 자체에서 맥락에 따라 성적 표현 유무가 결정되는 단어는 제외.

Ex. 자지, 보지, 백마, 대물,

* 문장 기준

- 맥락에서 판단할 수 있는 문장은 제외.

Ex. 아 저 여자랑 (정보)하고 싶다. 한 번 빼고 싶다.

- 문장 자체에서 성희롱, 성차별, 성폭행, 성행위를 직접적으로 연상할 경우 포함.

Ex. 저 여자랑 떡치고 싶다. 저 남자는 거시기가 크다.

\*\*\*\* 너무 긴 문장은 짤라주거나.. 패스한다.

가급적 데이터 변형을 금하고, 중복표기 하려하자.