kohate Dataset Report

Train Dataset

Preview

comments	contain_gender_bias	bias	hate
ㅇ정보부에서 조사하나봐요 .. 승리 곧 자살하겠네 아니..경찰이 질소넣고 자살했다고 하겠네..	False	others	hate
어차피 속바지인데 뭔 노출ㅋㅋ	True	gender	hate
손담비 나혼에 고정으로 가자~~매력 넘침~	False	none	none
개듣보네 ㅋㅋㅋㅋㅋㅋ남자 뭔 프로게이머같이생겼네	False	none	hate
누가봐도 한거다. ㅎㅎㅎ	False	none	offensive
처음 본 여자한테 찝적 거렸는데, 여자가 기분 나빴다 ? 이게 죄야 ? 성적 발언을 한 것도 아니고, 상대가 기분이 나쁘면 무조건 피해자로 규정하나 ? 이건 아니지 싶다 ㅋㅋㅋ	True	gender	none
에이핑큰줄알았더니 짝퉁김남주였네	False	none	offensive
개쳐늙으면 꼴보기 싫어지는게 사람의 마음이지 얘네야 광고수익으로 번게 많으니 얼굴에 돈들이고 관리하니 금술이 당연히 좋을수밖에 사람이란 다 그런거임 아무리 이쁘고 잘생겼어도 관리안하면 상대쪽에서 싫어하게되어있음	False	others	hate
미투 최대 피해자김생민‥불쌍해	True	gender	none
솔직히 인성이 종아리에서 나왔음	False	none	offensive

Description

comments	contain_gender_bias	bias	hate
6316	6316	6316	6316
6316	2	3	3
크러시가 정용화,임시완급도 아닌데 부산에서 누가 알아본다고ㅋㅋ	False	none	none
1	5336	4126	2798

Info

Non-Null Count	Null Count	Dtype
6316	0	object
6316	0	category
6316	0	category
6316	0	category

Validation Dataset

Preview

comments	contain_gender_bias	bias	hate
제발~~부끄러운줄알았음..이혼은왜하셨나요??사생활지저분한쇼호스트가 티비나와서할소린아닌줄..	False	others	hate
윌리엄 그동안 스킵했었는데 오늘 이모습하고 서준이에서 눈여겨볼 아이로 성장했움	False	none	none
우익 꺼지셈	False	others	offensive
이뻐보이진 않는데..	False	none	none
와ㅜㅜ 고배우 연기 진짜ㅜㅜ 와 소름돋음	False	none	none
방금 유동근 아재 회장여자 한테 엄포 놓을때 포스 ㅎㄷㄷ간만에 정도전때 포스 나오더만 진작 이렇게 만들지 이상하게 캐릭터를 만들어 놓아서.	False	none	none
백상까지 조작한 sm.문재인들어서고 방송 문화 경제등 비리와 부패가 더 심각하고 망하고있다.조작사기댜 sm비리와 조작으로 인기구걸하는데도 조사없는 문재인끌어내려야지	False	others	hate

당신을 보면 딱 이구절만 생각나네요바보가 신념을 가지면 무섭습니다.결혼한다니 축하드리고 잘사시길 바랍니다. 자기 주제에 맞게 사시길.	False	none	offensive
육아하는 아빠라니 육아는 부모라면 당연한거고요....ㅋ..;;	False	none	none
저런곳에 일하는여자가 저런거 당했다고 하루하루 괴롭게 지낸다고?개가 웃겠다ㅋㅋ돈만주면 2차 나가는거 졸 좋아하는 애들 아니냐? 그리고 성매매 업소도 아니라고? 성매매 업소도 성매매 업소라고 안하고 성매매하거든?어이없네	True	gender	hate

Description

comments	contain_gender_bias	bias	hate
1580	1580	1580	1580
1580	2	3	3
여자들이 죽인 유니. 진짜 한국여자들 설리 구하라도 죽이더니 예쁘고 몸매좋은여자들만보면 못물어뜯어서 안달남. 대단해	False	none	none
1	1328	1022	688

Info

Non-Null Count	Null Count	Dtype
1580	0	object
1580	0	category
1580	0	category
1580	0	category

Test Dataset

Preview

comments	contain_gender_bias	bias	hate
송중기 시대극은 믿고본다. 첫회 신선하고 좋았다.	False	none	none
지현우 나쁜놈	False	none	offensive
알바쓰고많이만들면되지 돈욕심없으면골목식당왜나온겨 기댕기게나하고 산에가서팔어라	False	none	hate
설마 ㅈ 현정 작가 아니지??	True	gender	hate
이미자씨 송혜교씨 돈이 그리 많으면 탈세말고 그돈으로 평소에 불우이웃에게 기부도 좀 하고사시죠.	False	none	offensive
일베충들 ㅂㄷ거리는것봐라 ㅉㅉ	False	others	hate
아이즈원 힘내세요...일본 진출도 했으니 일본서 좋은 모습 보여줘도 팬들은 응원 합니다.	False	none	none
강부자 선생님 전미선 비보에 오열을 하셨다니 눈물이 나네요 힘내세요	False	none	none
알았어 그만	False	none	offensive
이영자씨는 진정성 있는거라면 녹화불참으로 끝내지말고 자진하차해라 시청자는 고려도 안하고 일방적 불참은 아닌듯 엠비씨도 시청율 좋아서 고민하는거 같은데 결방할게 아니고 폐지해라	False	none	offensive

Description

comments	contain_gender_bias	bias	hate
471	471	471	471
471	2	3	3
남자가 잘못한거라면... 반성도 없다면...나였다면 ... 여자처럼 아주 못되게 할것같다왜??? 나를 배신한거니까	False	none	offensive
1	404	342	189

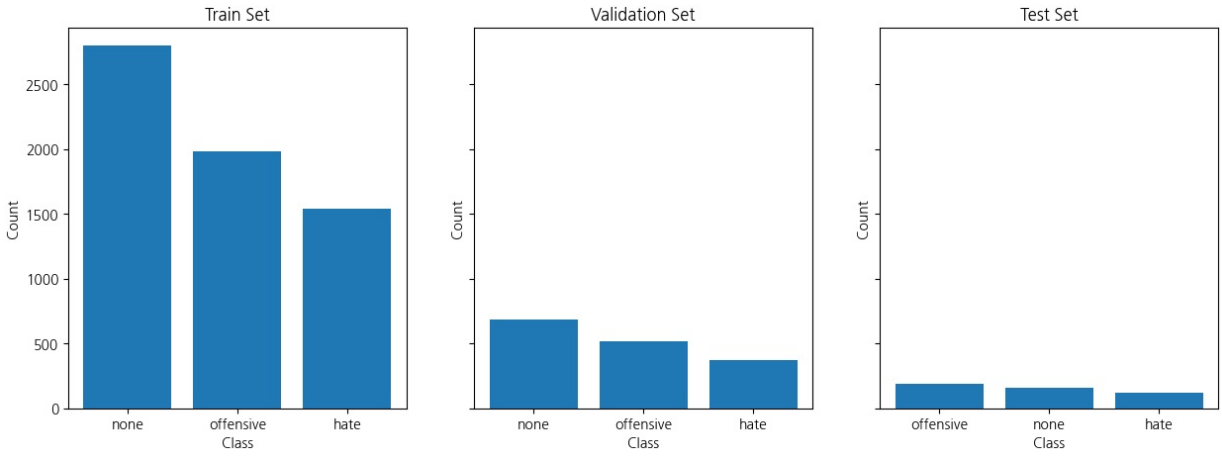
Info

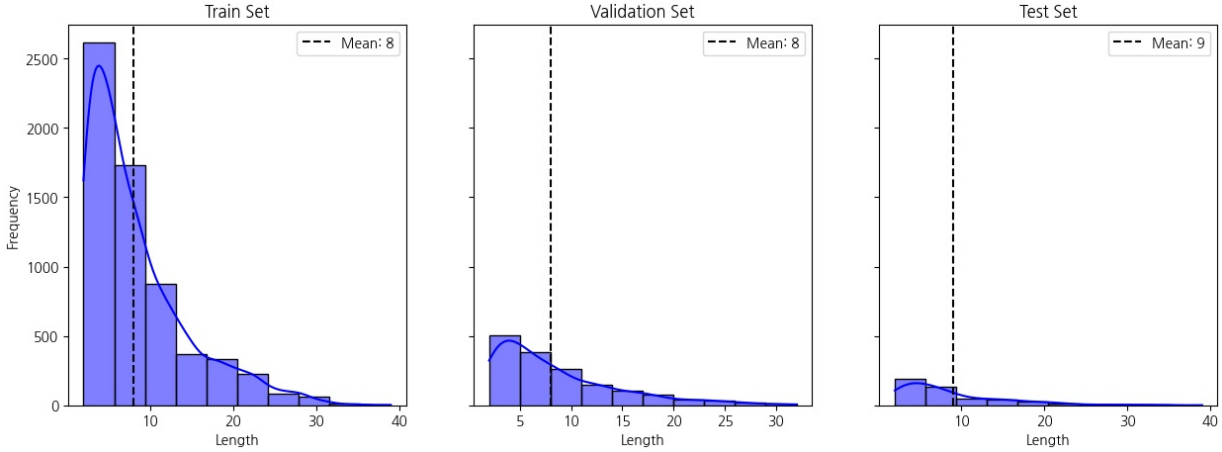
Non-Null Count	Null Count	Dtype
471	0	object

471	0	category
471	0	category
471	0	category

Visualizations

Class Distribution





Dataset	Longest Sentence	Shortest Sentence	Mean Sentence Length	Sum of Sentences
Train	39	2	8	6316
Validation	32	2	8	1580
Test	39	2	9	471

Word Cloud

통계 요약 코멘트:

- 데이터의 전반적인 특성 요약:

- 이 데이터셋은 총 세 개의 주요 부분으로 나뉘며 (Train, Validation, Test), 각각의 문서 수는 6316, 1580, 그리고 471개입니다.
- 모든 세트에서 평균 문장 길이는 8단어로 동일하게 유지되었습니다.
- 키워드와 길이 등으로부터 데이터 성격이나 주제의 변화 가능성을 추론:
- '진짜', '뭐'가 반복적으로 등장하는 "Train"과 "Test"는 긍정적이거나 중립적인 내용을 암시합니다.
- 반면, "Validation"은 보다 긍정적인 단어들을 사용하여 '긍정적'인 데이터를 나타냅니다.
- train, validation, test 데이터셋 간의 차이 분석:
- 세 개의 데이터셋 모두 평균 문장 길이와 주요 키워드 구성에서는 큰 차이가 없습니다.
- 중요한 점은 각 데이터셋의 문서 수와 주요 키워드의 차이로 인해 데이터 성격이 약간씩 변화했다는 것입니다.
- "Train"과 "Test"는 긍정적인 어조를 보이지만, "Validation"은 그보다 더 긍정적입니다.