



[1월 2주차 프로젝트 수행 일지]

프로젝트 주제	악성댓글 감정분석		
프로젝트 팀명	Writer's Warden	프로젝트 팀원	김나래, 김성훈, 김은지, 이동근, 장유림

※ 프로젝트 수행 일지는 매 주 금요일 17:00 전까지 최종본을 저장해주세요

1. 프로젝트 수행 계획 및 현황

김나래	<p>데이터 라벨 축소</p> <p>데이터 탐색</p> <p>형태소 분석</p> <p>혐오사전 구축</p>	<p>댓글 점수 계산 알고리즘</p>
김은지	<p>데이터 라벨 축소</p> <p>데이터 탐색</p> <p>형태소 분석</p> <p>PPT 제작</p>	<p><현재까지 진행단계></p> <ol style="list-style-type: none"> 유사한 카테고리를 묶어 비악플 포함 8개의 카테고리를 4개(비악플, 선정, 혐오, 폭력)으로 축소함. <ol style="list-style-type: none"> 클러스터링으로 분류하고자 했지만, 결과적으로 잘 되지 않았음. 악플의 종류를 3개로 분리했을 때 각 카테고리의 특징이 두드러짐을 확인함. 각 카테고리별 혐오사전을 구축

<p>김성훈</p>	<p>데이터 라벨 축소</p> <p>데이터 탐색</p> <p>형태소 분석</p> <p>혐오사전 구축</p>	<ol style="list-style-type: none"> 각 카테고리별 비윤리 점수 높은 구간의 워드클라우드 확인 혐오 관련 단어의 유사 단어 탐색 2gram으로 합성어('-충', '-녀' 등) 규칙성 있는 단어 추출 
<p>이동근</p>	<p>회귀모델 (선택나무)</p> <p>분류모델 (CNN)</p> <p>분류모델 (LSTM)</p> <p>생성</p> <p>Talkset Train, Unsmile Data에 대한 예측자료 생성</p>	
<p>장유림</p>	<p>데이터 라벨 축소</p> <p>데이터 탐색</p> <p>형태소 분석</p> <p>PPT 제작</p>	<ol style="list-style-type: none"> 점수 모델에서 가중치 및 알고리즘 순서 정의 분류 모델에서 CNN이외 LSTM 모델 생성 Decision Tree Regression 모델 생성 전체 데이터셋에 대해 모델 수준 검증

단어	CNN	LSTM
비약플	109538	109173
혐오	97008	97570
선정	10800	10801
폭력	4899	4701

▲ 모든 데이터에 대해 점수, 분류의 원본 데이터와 직접 비교했다.

분류	CNN	LSTM
비약플	109538	109173
혐오	97008	97570
선정	10800	10801
폭력	4899	4701

▲ CNN, LSTM의 분류 count 차이를 볼 수 있다.

7. TalkSet data로 학습된 모델에 Unsmile data 예측

단어	CNN	LSTM
비약플	109538	109173
혐오	97008	97570
선정	10800	10801
폭력	4899	4701

▲ Unsmile의 모든 데이터에 대해 점수와 분류를 예측했다.

< 아이디어 회의 상세 내용 >

<라벨 분류>

1. 클러스터링

주피터 노트북 만져서 비지도학습의 클러스터링 구현
근데 해보니까 너어어어어무 오래 걸림.

		<p>2. Talkset Data 범주를 축소</p> <p>a. 어떻게(어떤 기준으로 무엇을) 축소할지 회의 필요</p> <p>b. 코딩/수작업으로 라벨을 수정 =></p> <p>0. 비윤리 문장 아님['IMMORAL_NONE']</p> <p>1. 선정['SEXUAL'] : 선정['SEXUAL']</p> <p>2. 혐오['HATE'] : 혐오['HATE'], 차별['DISCRIMINATION'], 비난['CENSURE']</p> <p>3. 폭력['VIOLENCE'] : 폭력['VIOLENCE'], 범죄['CRIME'], 욕설['ABUSE']</p> <p>c. 비식별화된 부분을 새로운 태그로 정리할 필요가 있음 ex. 정치인-> AA, @라고만 되어있는경우 비속어일 가능성 판단 후 -> BB</p> <p>d. 위의 작업이 완료 되면 ex. AA + 혐오표현의 규칙성 정의</p> <p>e. 규칙성에 따른 점수(가중치 부여)</p> <p>3. 우리가 직접 정의</p> <p>a. 어떤 분류로 나눌지 생각</p> <p><진행 계획></p> <p>1. Talkset Train 모델 생성 (CNN, LSTM)</p> <p>2. Unsmile을 모델에 돌리기 - 교정 작업 (단어사전: 혐오단어사전, 오분류된것)</p> <p>3. (2) 로 나온 결과 -> 눈으로 보고 CNN/LSTM 결과 선택해서 정제 + Talkset Train으로 다시 모델 생성 (CNN, LSTM)</p> <p>4. 유튜브 라이브 채팅 긁어온 것 -> 이걸 돌려서 평가 -> 여기까지 오면 더이상 모델 생성은 무리. -> 실시간 채팅에 적용 여부 확인 및 적용</p>
--	--	--

2. 강사님 피드백	
<p>H 반</p> <p>전정호 강사님</p>	

