

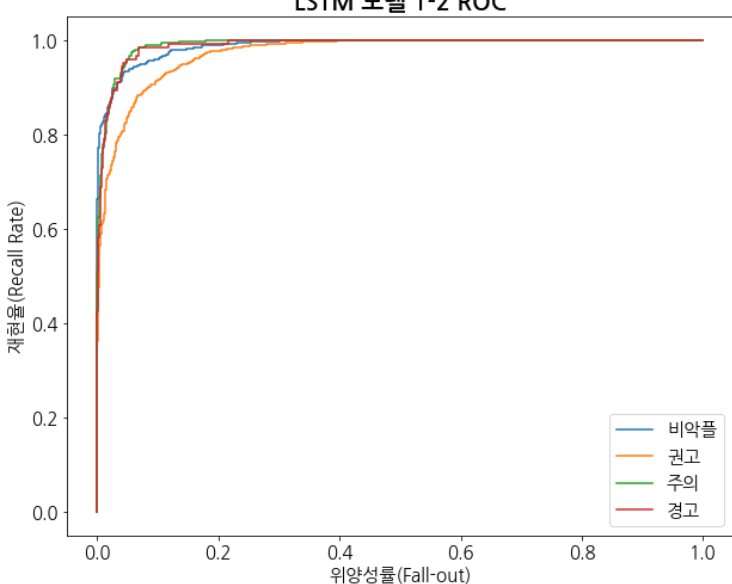
[2월 2주차 프로젝트 수행 일지]

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	<div>1. 데이터 탐색</div> <div>2. 형태소 분석</div> <div>3. 혐오사전 구축</div> <div>4. 모델(LSTM)</div> <div>5. HTML / SQ Lalchemy 보조</div>	< 데이터 정제 >

김은지	<div>1. 데이터 탐색</div> <div>2. 혐오사전 구축</div> <div>3. 모델(robert a)</div>	<div><div><div>비난</div><div>혐오</div><div>차별</div><div>선정</div><div>욕설</div><div>폭력</div><div>범죄</div></div><div><table><tr><td>비난</td><td>6.6</td><td>7.4</td><td>7.1</td><td>8.7</td><td>7.7</td><td>8.6</td><td>8.3</td></tr><tr><td>혐오</td><td></td><td>6.8</td><td>7.5</td><td>9.2</td><td>8.0</td><td>8.7</td><td>8.7</td></tr><tr><td>차별</td><td></td><td></td><td>6.8</td><td>7.8</td><td>8.2</td><td>8.2</td><td>8.8</td></tr><tr><td>선정</td><td></td><td></td><td></td><td>8.5</td><td>9.7</td><td>9.5</td><td>9.4</td></tr><tr><td>욕설</td><td></td><td></td><td></td><td></td><td>7.8</td><td>9.4</td><td>8.8</td></tr><tr><td>폭력</td><td></td><td></td><td></td><td></td><td></td><td>8.3</td><td>9.1</td></tr><tr><td>범죄</td><td></td><td></td><td></td><td></td><td></td><td></td><td>7.6</td></tr></table></div><div><div>9.5</div><div>9.0</div><div>8.5</div><div>8.0</div><div>7.5</div><div>7.0</div></div></div> <div><div>●</div><div>지난주 단일 분류된 문장들의 비윤리 강도 점수 평균으로 3개의 카테고리를 분류를 했지만, 다중 레벨 되어있는 문장이 대부분인데 반영하지 않은 것 같다는 피드백을 받음</div></div> <div><div>●</div><div>이를 해결하고자 2개의 조합에 대한 비윤리 강도 점수를 matrix로 시각화. 단독일 때와 2개의 조합일 때의 변화량을 그 분류가 받는 영향이라 생각함.</div></div> <div><div>●</div><div>단일 분류의 순위 / 다중 분류의 순위 / 영향력을 순위에 따라 점수를 매겨 합산 후 순서대로 3개의 카테고리를 나눔.</div></div> <div><div>●</div><div>결론 : 우리가 나누고자 한 결과와 같았다. (권고 : 비난 / 혐오 / 차별, 주의 : 욕설 / 범죄, 경고 : 폭력, 선정)</div></div>	비난	6.6	7.4	7.1	8.7	7.7	8.6	8.3	혐오		6.8	7.5	9.2	8.0	8.7	8.7	차별			6.8	7.8	8.2	8.2	8.8	선정				8.5	9.7	9.5	9.4	욕설					7.8	9.4	8.8	폭력						8.3	9.1	범죄							7.6
비난	6.6	7.4	7.1	8.7	7.7	8.6	8.3																																																			
혐오		6.8	7.5	9.2	8.0	8.7	8.7																																																			
차별			6.8	7.8	8.2	8.2	8.8																																																			
선정				8.5	9.7	9.5	9.4																																																			
욕설					7.8	9.4	8.8																																																			
폭력						8.3	9.1																																																			
범죄							7.6																																																			
김성훈	<div>1. 데이터 탐색</div> <div>2. 형태소 분석</div> <div>3. 모델(LSTM, KoBERT)</div>	<div>< 신조어 등록 ></div> <div><div>●</div><div>기존에 있던 vocab과 unsmile data를 형태소 분석한 결과와 비교하여 새로 추가 될 단어들을 kiwi 형태소분석기의 user_dic에 추가</div></div> <div>< 혐오사전 ></div> <div><div>●</div><div>혐오사전 구축 과정</div><div><div>-</div><div>각 분류에서만 쓰였던 형태소를 추출하여 유사도 높은 키워드들 검색</div></div><div><div>-</div><div>각 검색된 키워드들이 포함된 문장을 찾고 그 문장의 분류를 카운트</div></div><div><div>-</div><div>카운트 정보를 기반으로 하여 각 키워드들을 세 개의 혐오사전에 할당</div></div></div> <div><div>●</div><div>새로 분류 전의 혐오사전을 기반으로 재분류화 작업</div></div>																																																								

<p>장유림</p>	<ol style="list-style-type: none"> 1. 데이터 탐색 2. 혐오사전 구축 3. 모델(kobert, roberta) 	<ul style="list-style-type: none"> ● Unsmile에 대한 신조어 탐색 및 혐오사전 추가(진행 중) <p>< 모형 - LSTM ></p> <ul style="list-style-type: none"> ● 점수모델과 분류모델을 LSTM으로 분석 ● 발견된 이슈 : 하이퍼파라미터를 바꾸며 여러 모형을 시도 <ul style="list-style-type: none"> ○ 과적합(Overfitting) 발생 <ul style="list-style-type: none"> ■ Dropout 삽입 -> 효과 ■ BatchNormalization 삽입 -> 효과 미비 ■ 파라미터 수 축소(hidden size) -> 효과 ■ 다른 모형 도입(Bi-LSTM) -> 약간 효과 ○ 이유 : <ul style="list-style-type: none"> ■ 데이터의 불균형 ■ 점수모델의 경우 연속형으로 회귀모델을 만들었을 때 zero-inflated data이므로 문제가 될 위험성 큼 ● 결과 <ul style="list-style-type: none"> ○ Drop down과 hidden size를 줄인 모형으로 선택 ○ 점수의 경우 비악플 점수 0이 모델에 큰 영향을 미치는 것으로 파악. 비악플을 제거한 후 학습 ○ 성능 파악을 위해 F1 Score / ROC / AUC 등 성능 지표 출력 <p>Unsmile 데이터를 추가하여 재학습 후 성능 개선을 보임.</p>
<p>이동근</p>	<p>웹 어플리케이션 작업 마무리</p> <p>시연 영상 제작 시작</p>	 <p>< 모형 - RoBERTa ></p> <ul style="list-style-type: none"> ● 배경 : pre-trained 된 데이터가 더 많다는 멘토님의 추천으로 적용해보기로 함. <ul style="list-style-type: none"> ○ BERT기반의 언어모델로 단어 한개가 누락된 문장 입력으로 받아, 해당위치에 들어가기에 적합한 단어가 무엇인지 맞추는 과정을 학습

- 과정: 1) NSP제거
2) Dynamic Masking
3) 학습 데이터 증가
- dataset을 이용한 시각화, 토큰나이저 까지 실행
-> 모델 학습부분에서 에러 발생, 정확한 원인 파악 못함

< 모형 - KOBERT >

- KOBERT를 활용하여 분류모델을 실행
- KOBERT 재학습을 통한 unsmile data 학습
- 재학습 결과
 - 설정
 - epoch = 15, batch_size = 64
 - train acc = 99%
 - test acc = 86.7%
 - 1000개의 문장을 추출하여 예측한 결과 975개를 맞춤
- LSTM보다 KoBERT가 더 학습이 잘 된 것으로 파악. 최종 모델 선정
- 악플에 대한 최종 Score = 점수예측값 * 분류가중치 + 혐오사전점수
- Score >= 20 : 경고
Score >= 50 : 10초 Ban
Score >= 125 영구 Ban

<웹 어플리케이션>

<Node.js>

:::: 스크립트 관련 개선사항

- 웹 서버가 정상적으로 실행됩니다.
- Fast API 서버와의 정상적인 통신이 가능합니다.

<MySQL>

:::: DB 설계

- 스트리머 당 DB가 생성되며 DB 안에 viewerinfo table과 stream session 을 이름으로 하는 채팅로그 table이 들어갑니다.

<Python Server>

:::: 파이썬 관련 개선사항

- 더이상 모듈에서 분류작업을 진행하지 않습니다.
- FastAPI를 통해 파이썬 서버 (port 8000)를 구축 완료해 해당 서버로 메시지를 주고받으며 분류합니다.
- KoBERT모델 환경 구축이 끝나 분류작업이 정상적으로 진행됩니다.
- LSTM 회귀 모델이 정상적으로 점수를 매깁니다.
- FastAPI 서버와 웹 서버간 정상적인 통신이 구축 완료되었습니다.

<시연 영상>

:::: 영상 제작 관련

		<ul style="list-style-type: none"> ● 초안이 완성되었습니다. <p><차주 예정사항> 혐오사전 작업이 완료될 시 혐오사전에서 단어를 검색해 점수를 추가로 세부조정하는 기능을 구현합니다. 추가적인 토론을 통해 HTML 페이지의 추가기능을 탑재하고 시연 영상을 가독성 높게 편집할 예정입니다.</p> <p>< PPT 제작 ></p> <ul style="list-style-type: none"> ● ppt 제작을 위한 내용 총정리 후, 멘토링 발표용 ppt를 제작했습니다.
--	--	---

2. 강사님 피드백

XX 반 XXX 강사님	
---------------------	--