

[1월 1주차 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>라이브 채팅 댓글 수집</p> <p>형태소 분석</p> <p>학습 데이터셋 탐색</p> <p>형태소 세부 수정</p> <p>신조어 단어 1차 선별작업</p>	<p>워크 플로우차트</p> <p><작업 방향></p> <p>1-A. Talkset Train 데이터를 활용한 점수 평가 모델 생성</p> <ul style="list-style-type: none"> ● Talkset Train 데이터로 해당 댓글에 대한 점수를 매긴다. (점수에 대해서는 별도 기준 마련 가능하지 않을까? or 점수가 매겨진걸 활용) <p>1-B. Talkset Train 데이터 활용 or 별도 활용 통한 댓글 분류 모델 생성 (LSTM, CNN?)</p> <p>1-B-i. 라벨을 별도로 생성 or Talkset Train에서 제공하는 라벨을 활용</p>
김성훈	<p>신조어 단어 1차 선별작업</p> <p>형태소 분석</p> <p>라이브 채팅 댓글 수집</p> <p>학습 데이터셋 탐색</p>	

<p>김은지</p>	<p>형태소 분석</p> <p>학습 데이터셋 탐색</p> <p>형태소 세부 수정</p> <p>신조어 단어 1차 선별작업</p>	<p>2. Youtube API 활용해서 댓글을 가져와 실시간으로 분류해 처리하는 기능을 실험</p> <p>3. DB에 id와 점수를 연동해 관리하는 기능을 실험</p> <p><회의 내용></p> <p>주제 배경</p> <ul style="list-style-type: none">● 현재 라이브 방송 관련한 악성 댓글 감지 시스템<ul style="list-style-type: none">○ 커뮤니티 : 주먹(관리자가 요청받고 처리하는 시스템), 타 환경 대비 악플 감지 정도 덜 함.○ 스트리머 : 나이트봇(단어 기준) + 관리자○ 유튜브 : 실시간 댓글 감지● 어디에 가장 악성 댓글 감지 시스템이 필요할까?<ul style="list-style-type: none">○ 인터넷 방송, 라이브 방송 -> 악성 댓글로 인해 분위기 흐려지는 상황 다수 발생-> 즉각적으로 댓글을 차단하는 속도가 중요할 것.<주의> 실시간으로 효과 있다는 것을 어떻게 증명할 것인가?○ 유튜브 -> 아이들의 유튜버화 높아짐. 즉, 악성 댓글의 노출도가 높아짐○ => 추가 필요한 자료 : 현재 라이브 시스템은 정말 악성댓글을 못 거르는가? 이유는? 단순 속도 때문?인스타그램의 악성 댓글 탐지 어떻게 이뤄지고 있는가? <p>[형태소 분석 문제 및 해결]</p> <ul style="list-style-type: none">● userdic 부분을 보완하고 일부 용언을 제외하면 필요한 형태소만
------------	--	---

추출할 수 있을 것 같다. <명사>

- 아예 **어간까지 싹다 긁어와서** 형식 형태소로 감정을 나타내는 부분도 감안해볼까 한다.
- 대부분의 **키워드가 명사로** 이루어져 있기때문에 명사를 추출하는게 가장 효율적이라고 생각이 듬.
- **띄어쓰기**를 이용한 경우, 과정을 좀 더 상세히 확인해보면서 해야할 것 같다.
- 프로그램명, 신조어 등 긴 단어에 대해서는 **soynlp의 extract noun**을 사용하는게 적절하다고 생각.
- 위의 과정이 완료된후 **user_dic에 넣고 komoran 형태소분석을 한 후 또 걸러내고 이 과정의 반복**이 필요하다고 생각함.
- 명사에 있어 유사도 분석을 진행하면 라벨링으로 쓸 수 있을것같다.
- 형태소 분석기를 Okt로 해보는것도 좋지 않을까.

신조어 찾기



이동근

분류 모델 생성

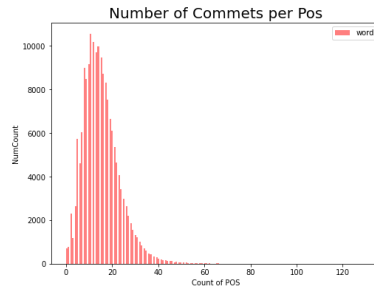
학습 데이터셋
탐색

신조어 단어 1차
선별작업

TalksetTrain 데이터를 형태소 추출 후 제공된 라벨을 활용해 분석

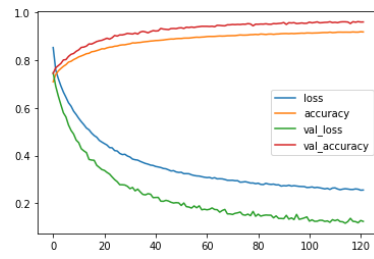


◀ 형태소 분석 워드 클라우드



◀ 형태소에 따른 데이터 개수

형태소 추출한 데이터와 라벨을 이용해 모델 생성 후 평가



◀ 모델 평가 지표

```
1/1 [=====] - 0s 66ms/step
문장:
남자가 말하는 이유는 동서고금을 막론하고 계집때문이다

-----
예측값: ['HATE']
실제값: ['HATE']
1/1 [=====] - 0s 74ms/step
문장:
출간소름보다 오토바이 소름이 더 싫어. 그 꺾을 때문에 귀 쫓겨질 것 같아

-----
예측값: ['IMMORAL_NONE']
실제값: ['IMMORAL_NONE']
1/1 [=====] - 0s 74ms/step
문장:
결민이를 이상한 말 할 때 마다 머리카락 잡아 당기고싶다

-----
예측값: ['VIOLENCE']
실제값: ['VIOLENCE']
1/1 [=====] - 0s 49ms/step
문장:
왕토리를 소모해서 그렇다더라고

-----
예측값: ['IMMORAL_NONE']
실제값: ['IMMORAL_NONE']
1/1 [=====] - 0s 74ms/step
문장:
저런 애들 보면 부모는 무슨 생각을 할까?

-----
예측값: ['CENSURE']
실제값: ['CENSURE']
1/1 [=====] - 0s 56ms/step
문장:
ㅇㅇ 짹거리네 ㅋㅋㅋ

-----
예측값: ['CENSURE']
실제값: ['CENSURE']
```

◀ 테스트 결과

<div>장유림</div> <div> <div>신조어 단어 1차</div> <div>선별작업</div> <div>형태소 분석</div> <div>라이브채팅</div> <div>댓글수집</div> </div>	<div>< 점수화 관련 회의 내용 ></div> <div> <div>1. 범죄성립하는 기준으로 점수화</div> <div>2. 표현의 자유는 보장되어야 함. 비판은 필요 근거없는 비난을 거르는게 필요하다</div> <div>3. 라이브. 몰타기 마녀사냥화되는게 문제 연속적으로 나오는 말들 탐지</div> <div>4. 금지어 설정 좀 더 스마트화</div> </div>
---	--

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