Couque d'Asse 쿠크다스

심주안/이지훈/은승연/조규희

Contents

I. 개요

- 팀원소개
- 사업기획배경

Ⅱ. 빅데이터 분석

- 데이터 수집
- 모델링

Ⅲ. 결론

- 상용화
- 한계



Introduction







이지훈 Python/Django 빅데이터 분석 ejihoon6065@naver.com

千到7十至01 领机 堂卫到高01 领此!

Team 4



심주안
Python/R
이미지 딥러닝/기획
tlawndks0202@gmail.com



조규희 Python/R/Hadoop 인공지능 분석 kjo3@jh.edu

SUSTAINABLE GEALS DEVELOPMENT





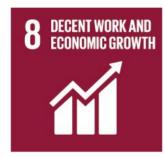
































사업 기획 배경

"사망 직전에도...SNS까지 뒤쫓은 악플"

"[뉴스초점] '얼굴 없는 살인자' 악성 댓글"

"소리없는 흉기 '악플' 댓글문화"

"[악플공화국] "'장난이었는데...' 악플 심각성"

악플이란?

1욕설:일반적인욕설

2 저속한 표현: 타인에게 불쾌감을 주는 표현

3 선정적인 표현: 성적으로 자극적인 표현

4 폭력적인 표현: 신체적 위협에 대한 표현

5 차별적인 표현: 인종/국가/종교 등에 기반한 차별표현

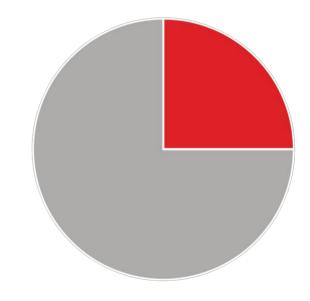
6 비하적인 표현: 상대방에게 모멸감과 수치심을 주는 표현

참조'악성 댓글의 사용 실태와 개선 방안 연구(2012, 구현정)'

시행 제도 및 현황

네이버	클린봇		
카카오	욕설 자동 치환		
구글	퍼스펙티브API		
인스타	댓글 취소하기 / 제한하기 기능 도입		
유투브	악성 댓글 자동 삭제기능 도입		

"4명 중 1명이 악성댓글로 인한 피해를 경험"



출처: 정보통신 평가원

분석 목표

"우리에겐 보다 섬세한 필터기가 필요하다!"

A 단순 욕설이 적힌 댓글이 아닌 불건전 목적을

가진 문맥을 파악

B 비방목적 악플 노출의 원천 차단

C 악성댓글 20% 감소율에 목적

Data analysis

데이터 수집 - YouTube API 이용

수집한 댓글 개수

논란 유튜버

일반 키워드

약 85,000

10

- 한글로 쓰여진 댓글 수집 슈기, 한혜연, 감스트, 공혁준, 장마, 코로나, 이스타항공, 밴쯔, 김민아
 - 유튜버별 최고 조회수 10개 영상의 댓글 추출
- 문재인, 박원순, n번방, 크보, k리그, 맨유, 기성용
- 유튜버별 최고 조회수 10개 영상의 댓글 추출

데이터 전처리

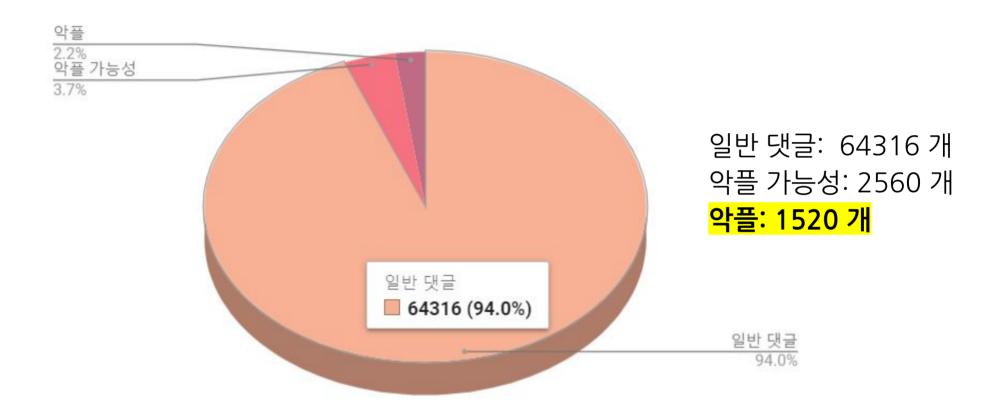
악플 레이블링

	0 (악플 아닌것)	1 (모호함)	2 (악플)
비속어	-	О	О
공격성	-	-	О
표현의 자유	0	0	-

레이블 예시

- ① "이부부는 시간이 갈수록 더 애정이 커지는거 같음 ㄹㅇ"
- 1 "세상이 어떻게 돌아가는지 원...저런 새끼 무서워서 밖에 못나갈듯;;"
- 2 "그냥 나가 죽어라"

데이터 전처리



데이터 불균형이 두드러진다.

데이터 전처리

Konlpy

댓글 정제

불용어 제거

한국어는 쉽다 >'는' 제거 토큰화

한국어는 쉽다 >한국어 Noun, 는 Josa, 쉽다 Adjective 정수인코딩

한국어는 쉽다 >[25,1,3] 패딩

한국어는 쉽다 >[0,0,0,25,1,3,0]

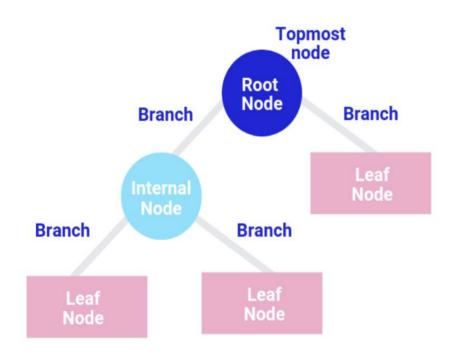
Test & Train set 분류 Sklearn.model

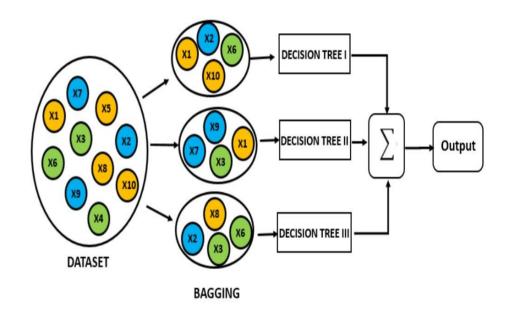
Machine Learning

- **1. KNN**
- 2. Decision Tree
- 3. Random Forest

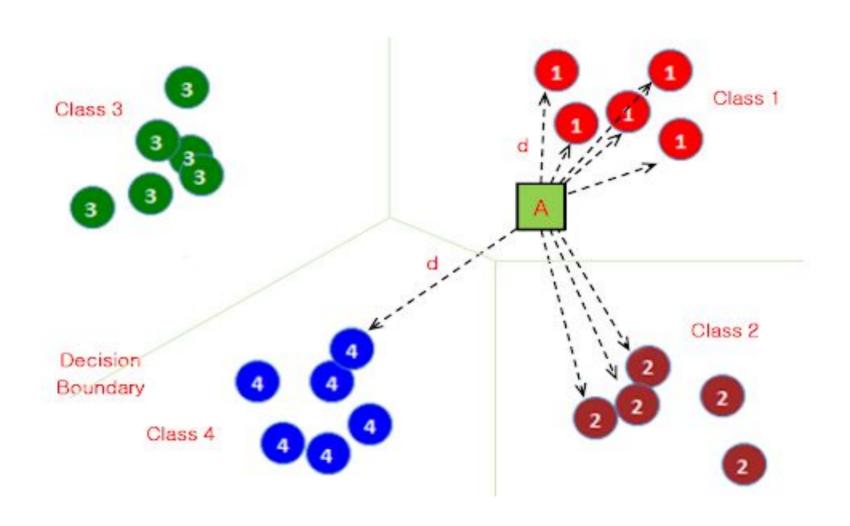
Decision Tree

Random Forest





KNeighbors



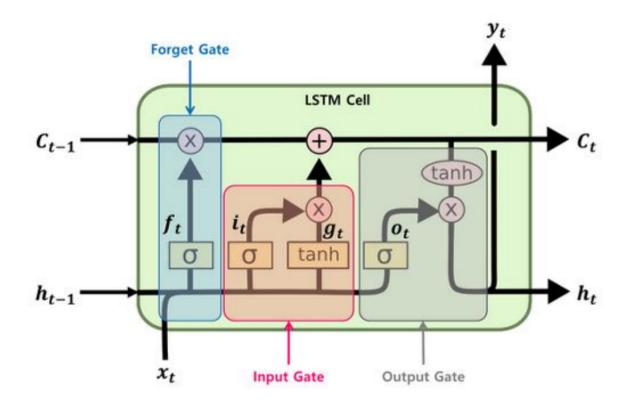
모델별 정확도

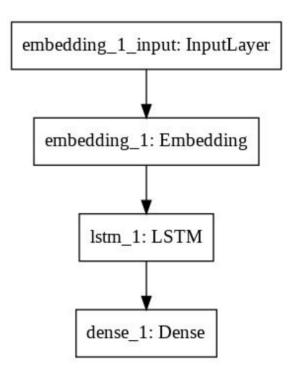
	KNN	Decision Tree	Random Forest
accuracy	0.93	0.88	0.94

Deep Learning Model

- 1. LSTM
- 2. Data Augmentation
- 3. Results

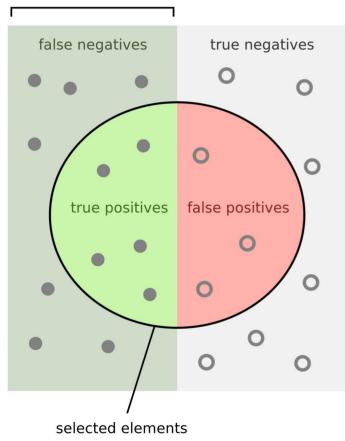
LSTM 모델



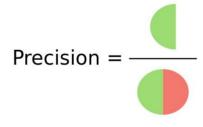


Metrics

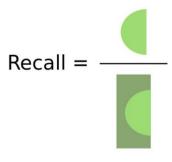
relevant elements



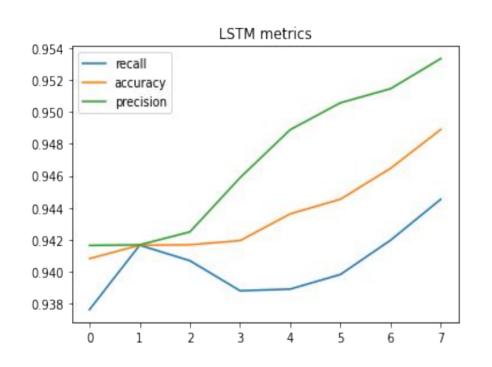
How many selected items are relevant?

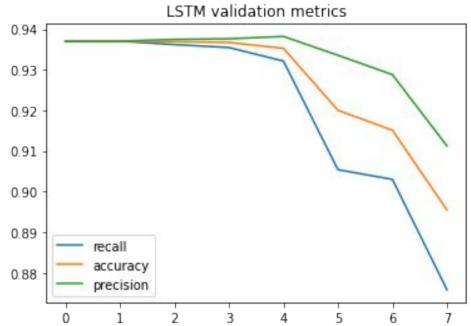


How many relevant items are selected?

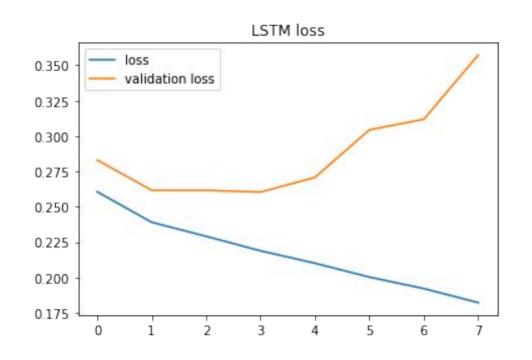


LSTM 모델 검증 및 결과





LSTM 모델 검증 및 결과



loss: 0.2702

acc: 0.9394

recall: 0.9394

precision: 0.9394

Train loss 는 감소하는 반면 Validation loss 가 모두 증가하는 양상을 보여, Overfitting 이 우려된다.

Data Augmentation

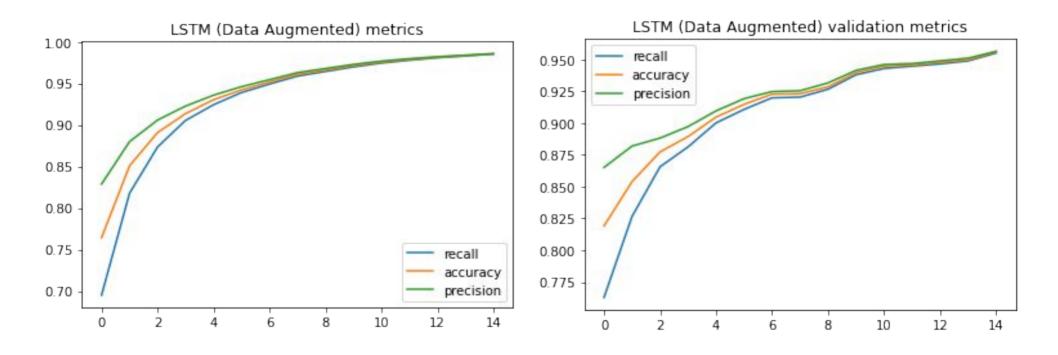
일반 댓글	64311	x 1	64311
악플 가능성	2560	x 5	12800
악플	1520	x 10	15200

68,391

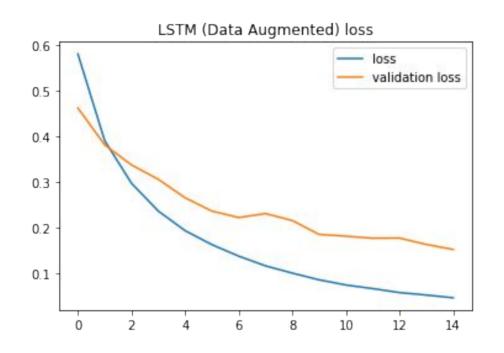


92,311

LSTM 모델 검증 및 결과



LSTM 모델 검증 및 결과



loss: 0.1566

acc: 0.9538

recall: 0.9530

precision: 0.9544

Train loss 와 Validation loss 가 모두 감소하는 양상을 보이며, Test 정확도가 소폭 상승하였다.

Conclusion



하루종일 악플만 보고 있으니까 정신 나갈 것 같아!





크리에이터 멘탈 지킴이,

쿠크다스

- 1. 叶星对京型毛纪
- 2. 叶型 准装计过
- 3. 正全等别社 叶莹 甜树

한계 및 시사점

- 1. 수집 데이터 부족 (다양성 및 범용성)
 - 2. 악플 판별의 기준이 자의적
- 3. 악플과 댓글 수가 편향적 (데이터 불균형)



한글 악성 단어 사전을 데이터를 체계적으로 증강한다