



인스타그램 좋아요 예측

KDT MiniProject3 1조
김수지, 남승완, 박정은, 정석영, 홍승우

Index

🔖 01. 주제 소개 및 선정이유

🔖 02. 데이터 수집 방법 - 크롤링

🔖 03. 데이터 전처리 및 데이터 설명

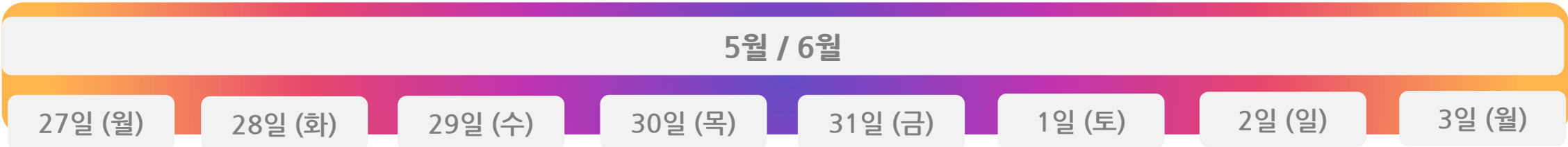
🔖 04. 모델 설정

🔖 05. 모델 성능 평가

🔖 06. 시사점 및 결론



0. 프로젝트 일정



크롤링

데이터 확인 및 방향 정리

모델 만들기 & 모델 학습 및 테스트

모델별 결과 도출 및 정리

PPT 작성 및 발표 준비

발표



0. 조원 소개



김수지
- 데이터 전처리, 모델링



남승완
- 크롤러 제작, 모델링



박정은
- 크롤링, 모델링, 발표자료



정석영
- 크롤링, 모델링, 발표



홍승우
- 크롤링, 모델링



01. 주제 소개 및 선정 이유



SNS 마케팅의 중요성이 증가하면서, 영향력을 평가하고 적절한 광고 비용을 산정하는 것이 필수적임.

좋아요 수와 팔로워 수는 이러한 평가의 지표로 사용됨.

→ 인스타그램 게시글의 좋아요 수를 예측하는 모델의 필요성을 느낌.



2. 데이터 수집 방법 - 크롤링



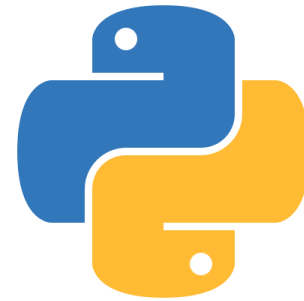
분석 데이터 및 분석 도구

분석 데이터



-
팔로워 수, 좋아요 수,
사진, 글 내용, 해시태그

분석 도구



colab





3. 데이터 전처리 및 데이터 설명



<이미지>

NumPy의 vstack을 사용하여 쌓음

2,580
게시물

8,474만
팔로워

446.7만
좋아요

좋아요 수를 팔로워 수로 나눈 비율
(영향도) 사용

결측치 처리

- 텍스트 데이터: 공백으로 채움
- 수치 데이터: 앞의 값으로 채움



3. 데이터 전처리 및 데이터 설명



 에스파_카리나	⋮	 에스파_지젤	⋮	 에스파_윈터	⋮
 에스파_닝닝	⋮	 블랙핑크_지수	⋮	 블랙핑크_제니	⋮
 블랙핑크_리사	⋮	 블랙핑크_로제	⋮	 레드벨벳_조이	⋮
 레드벨벳_웬디	⋮	 레드벨벳_예리	⋮	 레드벨벳_아이린	⋮
 레드벨벳_슬기	⋮				

연예인 13인의 인스타그램 게시물 8181개를 크롤링

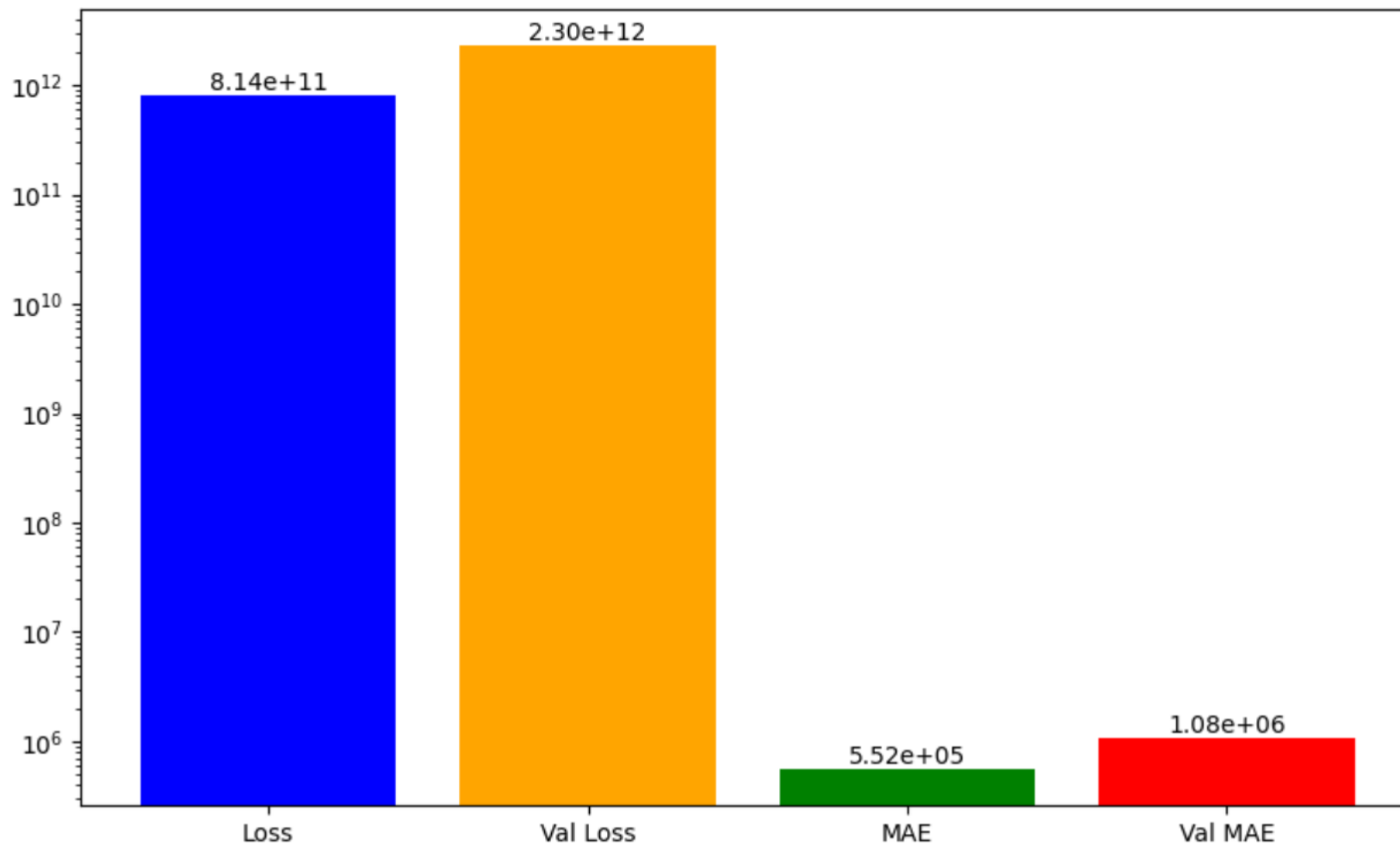
Content (object)
Hashtags (object)
Likes (float64)
Follower (int64)
Engagement_Rate (float64)



4. 모델 설정



이미지: vgg16
텍스트: lstm & 어텐션





5. 모델 성능 평가



모델	이미지 사이즈	배치사이즈	에폭	MAE
VGG16	224,224	32	50	1083961
VGG16	224,224	256	50	1106951
EfficientNetB0	128,128	256	26	962845
EfficientNetB0	128,128	256	34	801863
이미지 vgg16 텍스트 lstm&어텐션 병합모델	224,224	256	24	593231
이미지 vgg16 텍스트 lstm&어텐션 병합모델	224,224	256	50	551744.0625



6. 시사점 및 결론



▶ 한계




모델성능

다른 모델들보다 최종
모델이 우수한 결과



데이터의 한계

실시간으로 반영하지 못함
수집한 데이터로는 좋아요 수
예측이 용이하지 않음



모델 복잡도
및
학습 과정

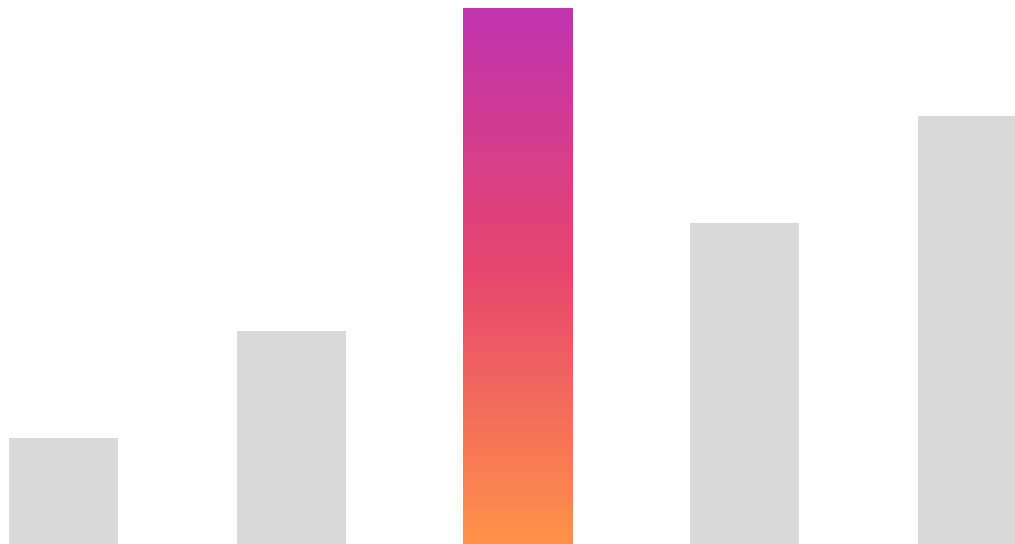
오차를 줄이기 위해
복잡한 모델을 사용함
검증데이터에서는
더 큰 오차 → 과적합



6. 시사점 및 결론



▶ 개선 방향



데이터의 품질을 높이며 더욱 더 다양한 모델을 사용하고,
또 모델의 단순화, 하이퍼파라미터 튜닝 등을 통해서 모델을 개선시켜 나가야 함

좋아요 수에 대한 모델들의 오차값을 고려할 때,
인스타그램의 좋아요 수 예측은 변동성이 크기 때문에 어려움

Thank you

로그아웃