

AI 전체 계획서

당신은 혹시 .. 인터넷 상에서 무지성 키보드 워리어?

데이터

데이터 레이블링 작업 자체는 매우 시간이 많이 걸리므로 외부 데이터셋을 활용하여 만들 예정이다.

데이터셋은 아래와 같이 한국어 혐오 표현 데이터셋을 사용할 예정이다. 아래 데이터는 총 9,381개이며, 레이블은 편견, 혐오, 그 외로 구성되어 있다. 편견은 성편견 여부, 혐오는 혐오, 불쾌한 발언, 그 외로 이뤄져있다.

데이터셋

<https://github.com/kocohub/korean-hate-speech>

레이블 관련 추후 계획

레이블이 과도하게 많을 경우 정확도가 떨어지는 일이 발생할 수 있으므로, 초기 모델에 입력 후 정확도를 비교해 {편견, 혐오, 그 외}로 할지 {편견, 성편견, 혐오, 불쾌한 발언, 그 외}로 할지 결정할 예정이다.

또한 전반적인 모델이 만들어졌다면 긍정 데이터셋을 추가할 예정이다.

모델

데이터셋 분할

데이터 70, 검증 20, 테스트 10

적용 예정 모델

- 로지스틱 회귀
- xgboost, lightgbm

- RNN
- Transformer

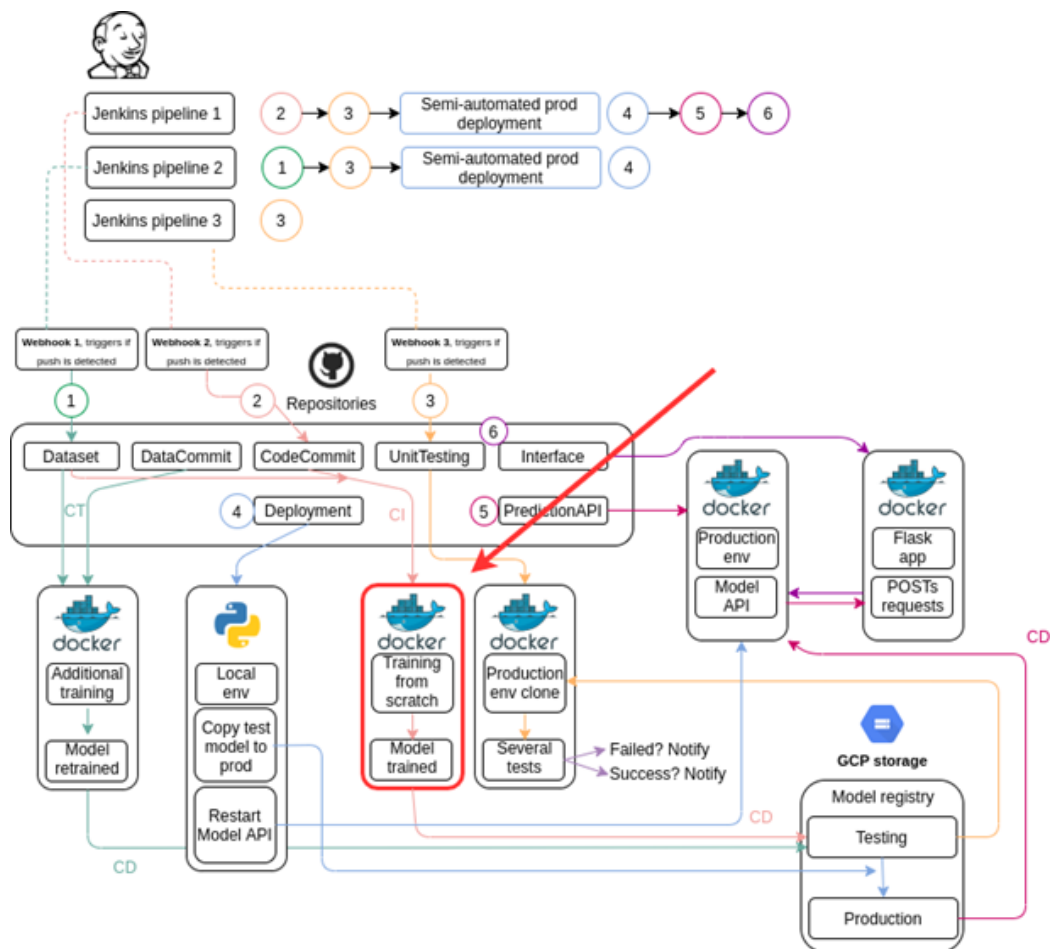
성능 평가 지표

- 정확도, 정밀도, 재현율, f1
- 오차행렬
- ROC

그 외 모델 평가시 고려사항

구현하기 쉬운 모델(3)	추론시간(5)	훈련 시간(5)
---------------	---------	----------

최종 계획



데이터 파이프라인을 제작해 Model API 서버 작업까지 할 예정이며, 프레임워크는 Flask, FastAPI 중 하나를 사용할 예정이다.

Jeckins는 CI 툴로 많이 쓰이긴하는데 구현시 번거러움과 너무나 많은 플러그인으로 인해 GitLab 사용이 더 좋을 것 같다. ([참고](#))

그 외 자료

- spring에서 Flask 연결

<https://velog.io/@dogyun-k/RestTemplate로-다른-서버와-통신하기#개요>

- CI/CD 개념

<https://artist-developer.tistory.com/24>