

MLOPS

오류를 넘어, 두 번째 월드에서 부딪히다.

에러 폭탄 같은 고난 앞에서도 꺾이지 않는 마음을 가진 팀, 2조 오뚝이입니다. 지치고 쓰러지더라도 일어나 도전하며 성장하는 것을 목표로 하고 있습니다. 에러와 삽질로 인해 좌절의 순간이 와도, 중요한 건 꺾이지 않는 마음! 마치 3분 요리처럼 빠르게 회복하며 다시금 프로젝트에 몰두하고 나아갑니다.

TEAM2, 프로젝트에 몰입하다

최진호

리더, Airflow, EC2,

김남섭

Docker, Container, 가상환경

박지은

가상환경, Streamlit

송주은

모델 실험, Streamlit, Docker

김현진

Docker, Container, 가상환경, 모델실험

프로젝트 협업 방식

소회의실 : 매일 실시간 질문과 토의

자료: 깃헙 및 공유 드라이브 활용

업무 분담: ① 모델학습/MLflow와 Docker 구축 전원 연습
② 인프라 구축/모델 고도화/서비스 구현 업무 분담하여 진행

영화 감성 분석 프로젝트 개요

MLOps Project | 목표 및 주요 작업



MLOps Project | 기술 스택 및 아키텍처

영화리뷰 기반 머신러닝 모델 운영 및 파이프라인

일반댓글 / 평점



Training.csv

실제 영화평
-API 연동

Feature: 댓글, 영화명, 장르, 설명, 평점, 감독, 출연진 등



데이터 전처리



데이터 학습



모델 평가



모델 저장



Amazon
EC2

NGINX



Streamlit



FastAPI



모델등록 crawl.csv

영화 감성 분석 수행 절차 및 방법

수행절차 | 데이터셋 및 데이터 처리

CGV, 메가박스 크롤링

영화정보, 댓글 전처리

데이터 처리

Kobert 기반

한국어 Tokenizer

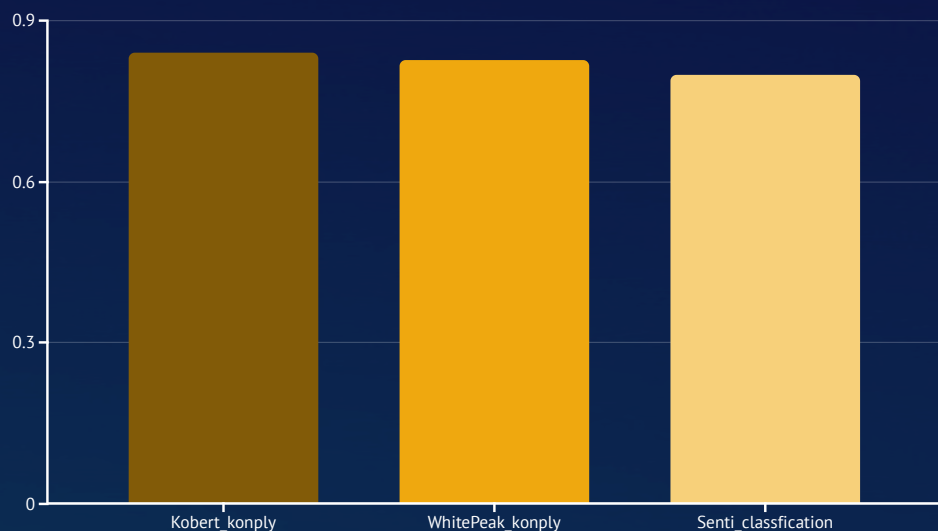
Konply (명사, 형태소 등)

데이터 결과

긍정, 중립, 부정

수행절차 | 모델 선택 및 이유 (훈련 및 성능 지표)

성능 지표



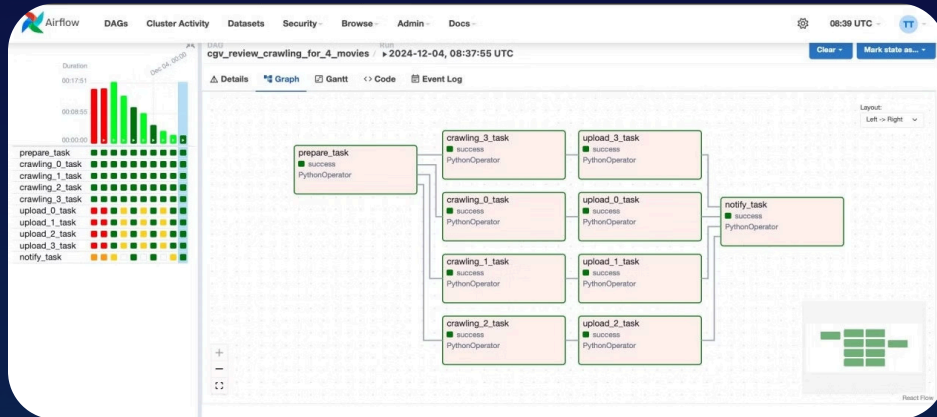
Kobert_konply 정확도 : 0.84

WhitePeak_konply 정확도 : 0.83

Senti_classification 정확도 : 0.80

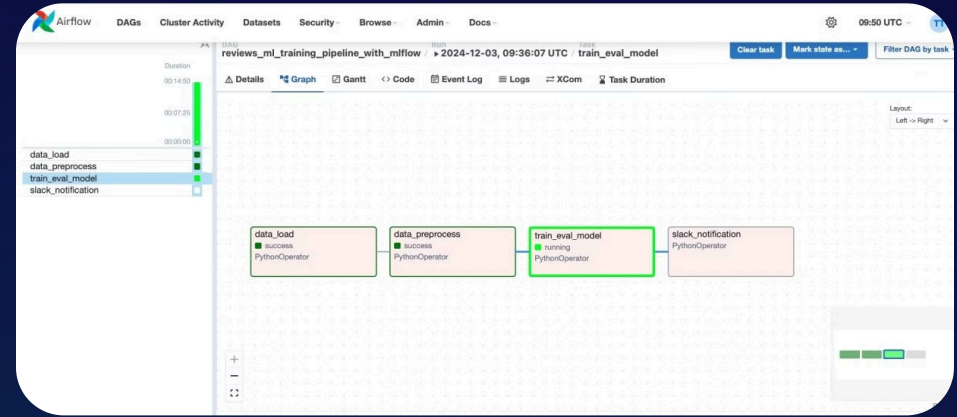
*senti_classification: konply 없이

수행절차 | 모니터링



Airflow dag : 데이터 크롤링

영화정보 크롤링 완료 후 슬랙 알림



Airflow dag : 모델 학습

여러 모델을 한번에 돌리기에는 서버가 다운

한 번에 하나의 모델을 돌리고 결과 비교

MODEL_TEST_BOT 앱 오후 2:55

Dataset 정보 *

데이터 경로: ratings.txt

데이터 샘플링 개수: 10000

학습 데이터 개수: 8000

평가 데이터 개수: 2000

데이터 전처리 시간: 54.73 seconds

Model 정보 *

모델명: skt/kobert-base-v1

저장 경로: kobert_1202_1450

학습 결과 *

평가 손실: 0.3874

평가 정확도: 0.8465

검측 정확도: 0.8465

학습 시간: 213.79 seconds

성능 포먼스 *

평가 실행 시간: 9.09 seconds

당 샘플 개수: 10.74

MLFlow 정보 *

실행 경로: mlflow-artifacts:/624224875204442792/

수행절차 : 슬랙 알림


데이터 크롤링 → 모델학습 → 슬랙 알림








수행절차 : Docker, Docker-compose

Containers [Give feedback](#)

Container CPU usage (i)
No containers are running.

Container memory usage (i)
No containers are running.

 ☒ Only show running containers

<input type="checkbox"/>	Name	Container ID	Image	Port(s)
<input type="checkbox"/>	<input type="checkbox"/>   team2container	-	-	-
<input type="checkbox"/>	 airflow-1	690371319e45	airflow-image	8080:8080
<input type="checkbox"/>	 mlflow-1	11edfab38a85	mlflow-image	5001:5000
<input type="checkbox"/>	 fastapi-1	ab9bf8ca9960	fastapi-image	8000:8000
<input type="checkbox"/>	 nginx-1	be7d923c4b2c	nginx:latest	80:80
<input type="checkbox"/>	 streamlit-1	ca8c2f44a180	streamlit-image	8501:8501

수행절차 : EC2 배포

대시보드

EC2 글로벌 보기

이벤트

인스턴스

인스턴스 유형

시작 템플릿

스팟 요청

Savings Plans

예약 인스턴스

전용 호스트

용량 예약

인스턴스 (1) 정보

최종 업데이트 날짜
less than a minute 전

연결

인스턴스 상태

작업

인스턴스 시작

인스턴스를 속성 또는 (case-sensitive) 태그로 찾기

모든 상태

인스턴스 상태 = running

필터 지우기

	Name	인스턴스 ID	인스턴스 상태	인스턴스 유형	상태 검사	경보 상태	가용 영역	퍼블릭 IPv4
<input type="checkbox"/>	Fastcampus-M...	i-0ff18f6f53876335a	실행 중	t4g.xlarge	3/3개 검사 통과	경보 보기	ap-northeast-2a	ec2-3-36-5

Amazon S3

범용 버킷

디렉터리 버킷

테이블 버킷

Access Grants

액세스 지정

객체 Lambda 액세스 지정

다중 리전 액세스 지정

배치 작업

S3용 IAM Access Analyzer

이 계정의 퍼블릭 액세스 차단 설정

Storage Lens

대시보드

Storage Lens 그룹

AWS Organizations 설정

fastcampus-ml-p2-bucket

객체

속성

권한

지표

관리

액세스 지정

객체 (4) 정보

S3 URI 복사

URL 복사

다운로드

열기

삭제

작업

폴더 만들기

업로드

객체는 Amazon S3에 저장되어 있는 기본 엔티티입니다. Amazon S3 인벤토리 를 사용하여 버킷에 있는 모든 객체의 목록을 얻을 수 있습니다. 다른 사용자가 객체에 액세스할 수 있게 하려면 명시적으로 권한을 부여해야 합니다. 자세히 알아보기

집두사로 객체 찾기

	이름	유형	마지막 수정	크기	스토리지 클래스
<input type="checkbox"/>	cinema/	폴더	-	-	-
<input type="checkbox"/>	image/	폴더	-	-	-
<input type="checkbox"/>	model/	폴더	-	-	-
<input type="checkbox"/>	reviews/	폴더	-	-	-

영화 감성 분석 회고



결과

- MLops 전반적인 flow 구현
- 시간 부족으로 인한 Streamlit 을 이용한 시각화 미흡



성장

- 수많은 시행착오를 통한 MLops 데 대한 이해
- 환경 설정의 중요성과 실무에서의 적용에 대한 고민
- 비용 관리에 대한 고민



진행 중 문제점

- 아나콘다, venv 등 가상 환경 설정 및 라이브러리 버전 충돌
- train 한국어 데이터 긍정, 부정 라벨링 문제
- 한국어 모델 train에 많은 시간 소요
- 언어 모델 실험에 컴퓨터 리소스 중요. 서버 다운
- EC2 프리티어 - 용량, 컴퓨터 리소스 문제 발생
- Docker, 컨테이너, 리눅스, EC2 사용의 어려움
- 흥미로운 서비스들 시도했지만 구현이 어렵거나 시간 부족

영화 클릭 : 댓글 속 이야기가 펼쳐집니다.

댓글 감성 분석

영화별 댓글 감성 분석 결과를 확인합니다.

인사이트 제공

긍정/부정 댓글 비율을 시각적으로 제공합니다.