

# MLOps 프로젝트

영화 평점 예측 서비스 개발 및 MLOps 파이프라인 구축

# 목차

01

주제 선정

02

전체 아키텍처

03

기술적 내용 및  
해결방법

04

소감 및 향후 계획

# 주제 선정

TMDB API 기반

영화 평점 예측 시스템  
구축

01

로컬 개발 모델을 API로 배포하고 MLOps 파이프라인 일부 구현.  
모델 코드 모듈화, 컨테이너 기반 학습 환경 및 API배포 구현

---

02

인기 영화 데이터를 자동으로 수집하고 학습하는 파이프라인  
경험.

데이터 수집(Airflow) - 서빙(FastAPI) - 사용자UI(Streamlit)  
까지의 End-to-End 시스템 구현

---

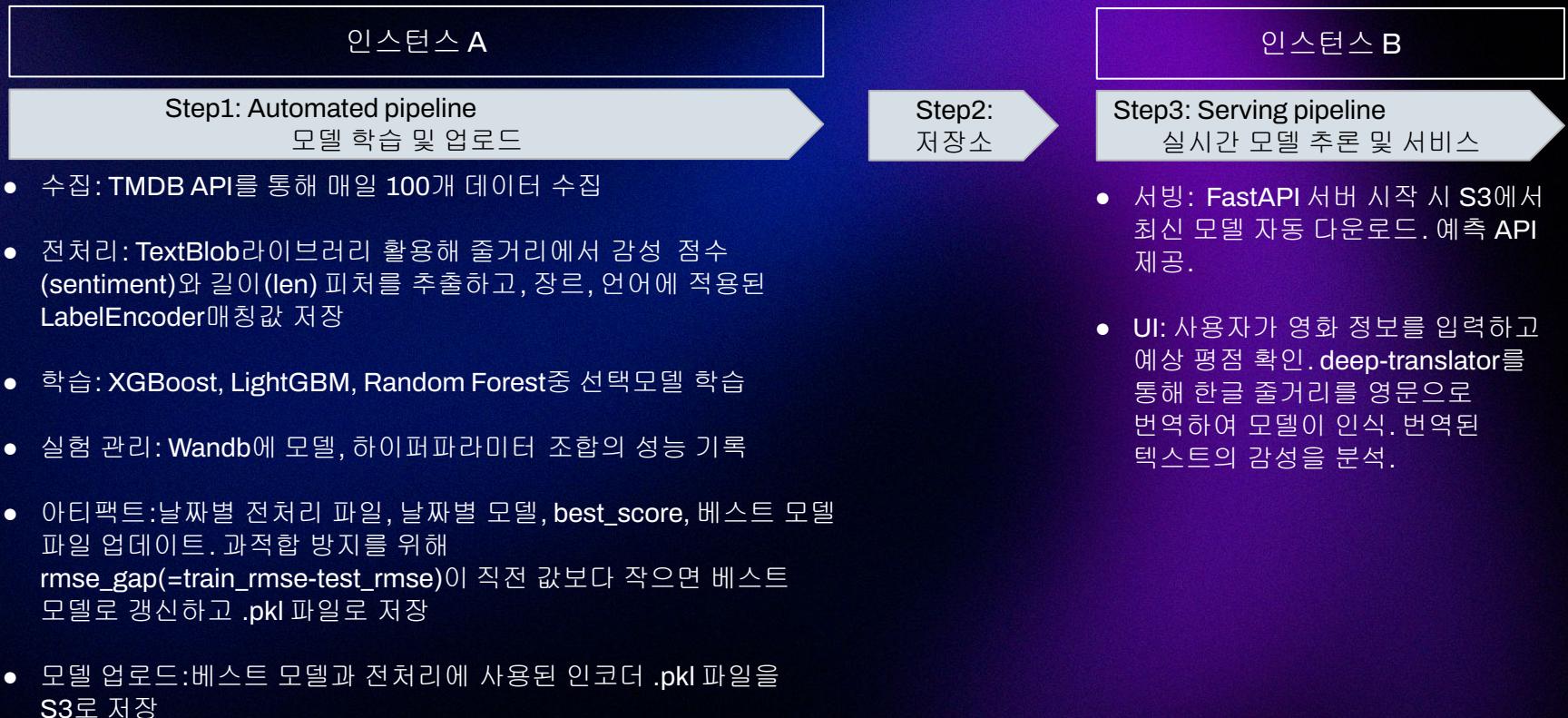
03

개봉 전 영화의 줄거리 감성 분석을 통해 평점을 예측하는  
비즈니스 가치 창출

# 아키텍처



# 아키텍처



# 아키텍처

app/	
└── model/	# S3에서 다운로드 받은 모델/인코더 임시저장
└── main.py	# FastAPI 서빙 및 S3 모델 자동 다운로드 로직
└── dags/	
└── movie_pipeline.py	# Airflow DAG파일
└── data-prepare/	
└── movies_cleaned_{ds}.csv	# 데이터, 모델 아티팩트 관리
└── main_genre_encoder.pkl	# 전처리된 학습용 데이터 (날짜별)
└── original_language_encoder.pkl	# 전처리에 사용된 장르 인코더
└── model_{ds}.pkl	# 전처리에 사용된 언어 인코더
└── best_model.pkl	# 모델 백업본 (날짜별)
└── best_score.txt	# 베스트 모델
└── scripts/	
└── collect.py	# 데이터 수집
└── preprocess.py	# 데이터 전처리
└── train.py	# WandB 연동 모델 학습 및 베스트 모델 선별
└── parameters.py	# 모델 하이퍼파라미터 및 설정값 관리
└── deploy.py	# S3로 베스트모델 업로드
.env	# 환경 변수
app.py	# Streamlit 웹UI
Dockerfile	
docker-compose-airflow.yaml	# 학습환경 실행 설정
docker-compose-fastapi.yaml	# 서빙환경 실행 설정

# 기술적 내용 및 해결방법 #1

## 1. 시스템 가용성 (Scale-out)

2. 데이터 신뢰성 (Data Leakage)

3. 실험 관리 (Wandb)

4. 아키텍처 고도화 (S3)

5. 데이터 일관성 (Encoding)

### 문제

단일 t3.small 인스턴스에서 Airflow와 Docker 컨테이너들을  
동시 가동 시 RAM 부족으로 인한 시스템 다운 발생

### 해결

Scale-up 대신 Scale-out 방식을 채택하여 학습(Airflow)와  
서빙(FastAPI) 인스턴스를 분리하고 부하를 분산함.

리소스 간섭을 방지하고 분산 환경에서의 안정적인 시스템  
운영 역량을 확보함

# 기술적 내용 및 해결방법 #1

## 분산 환경 (Scale-out) : Airflow, FastAPI

인스턴스 (3) 정보

인스턴스 (3) 정보		인스턴스 ID	인스턴스 상태	인스턴스 유형	상태 검사	경보 상태	가용 영역
<input type="checkbox"/>	mlops-user2	[REDACTED]	실행 중	t3.micro	3/3개 검사 통과	경보 보기	+ ap-northeast-2c
<input type="checkbox"/>	mlops-user1	[REDACTED]	실행 중	t3.small	3/3개 검사 통과	경보 보기	+ ap-northeast-2c

```

✓ Network airflow_mlops_default      Removed
  0.1s
ubuntu@ip-172-31-44-5:~/airflow_MLOps$ docker compose -f docker-compose-airflow.yaml up -d
[+] up 3/3
✓ Image airflow_mlops-fastapi      Built
✓ Network airflow_mlops_default     Created
✓ Container airflow_mlops-fastapi-1 Created
ubuntu@ip-172-31-32-66:~/airflow_MLOps$ ls
Dockerfile app dags docker-compose-airflow.yaml mlops-user2.pem scripts
pycache app.py data-prepare docker-compose-fastapi.yaml requirements.txt wandb
ubuntu@ip-172-31-32-66:~/airflow_MLOps$ cd app
-bash: cd: ap: No such file or directory
ubuntu@ip-172-31-32-66:~/airflow_MLOps$ cd app
ubuntu@ip-172-31-32-66:~/airflow_MLOps$ app$ ls
_pycache_ main.py model
ubuntu@ip-172-31-32-66:~/airflow_MLOps/app$ cd model
ubuntu@ip-172-31-32-66:~/airflow_MLOps/app/model$ ls
ubuntu@ip-172-31-32-66:~/airflow_MLOps/app/model$ cd ..
ubuntu@ip-172-31-32-66:~/airflow_MLOps/app$ docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS
NAMES
55cb61d39d58 airflow_mlops-fastapi "/usr/bin/dumb-init ..." 3 minutes ago Up 3 minutes 0.0.0.0
:8000->8000/tcp, [::]:8000->8000/tcp, 8080/tcp airflow_mlops-fastapi-1
ubuntu@ip-172-31-32-66:~/airflow_MLOps$ 
)0 △ 0

```

## 기술적 내용 및 해결방법 #2

1. 시스템 가용성 (Scale-out)
2. 데이터 신뢰성 (Data Leakage)
3. 실험 관리 (Wandb)
4. 아키텍처 고도화 (S3)
5. 데이터 일관성 (Encoding)

### 문제

초기 RMSE 0.44라는 극단적인 고성능 기록. EDA 분석 결과 미래 시점 데이터인 revenue, vote\_count 변수가 포함된 Data Leakage 확인

### 해결

Leakage 변수를 식별하여 제거하고, overview(줄거리)의 감성점수(Sentiment), 길이(Length) 파생변수 생성하여 모델 일반화 개선

# 기술적 내용 및 해결방법 #2

## 데이터 신뢰성 확보

	<b>id</b>	<b>title</b>	<b>budget</b>	<b>revenue</b>	<b>runtime</b>	<b>vote_count</b>	<b>vote_average</b>	<b>release_year</b>	<b>genre_encoded</b>	<b>lang_encoded</b>	<b>overview_len</b>	<b>overview_sentiment</b>
0	1306368	The Rip	100000000	0	113	575	7.046	2026	0	1	146	0.000000
1	1043197	Dust Bunny	0	928164	106	74	6.900	2025	0	1	212	0.200000
2	1242898	Predator: Badlands	105000000	184478861	107	1574	7.800	2025	0	1	158	-0.133333
3	83533	Avatar: Fire and Ash	350000000	1320010340	198	1548	7.340	2025	11	1	361	-0.380519
4	1242501	Icefall	0	0	99	89	6.500	2025	0	1	379	-0.233333

1. 수집된 변수 26개:

```
['adult', 'backdrop_path', 'belongs_to_collection', 'budget', 'genres', 'homepage', 'id', 'imdb_id', 'origin_country', 'original_language', 'original_title', 'overview', 'popularity', 'poster_path', 'production_companies', 'production_countries', 'release_date', 'revenue', 'runtime', 'spoken_languages', 'status', 'tagline', 'title', 'video', 'vote_average', 'vote_count']
```

2. 전처리된 변수 12개:

```
['id', 'title', 'budget', 'revenue', 'runtime', 'vote_count', 'vote_average', 'release_year', 'genre_encoded', 'lang_encoded', 'overview_len', 'overview_sentiment']
```

3. X 변수 5개: ['runtime', 'genre\_encoded', 'lang\_encoded', 'overview\_sentiment', 'overview\_len']

# 불필요 변수 제외: vote\_count, revenue, budget, release\_year

4. y 변수: "vote\_average"

# 기술적 내용 및 해결방법 #3

1. 시스템 가용성 (Scale-out)
2. 데이터 신뢰성 (Data Leakage)
- 3. 실험 관리 (Wandb)**
4. 아키텍처 고도화 (S3)
5. 데이터 일관성 (Encoding)

## 문제

반복되는 하이퍼파라미터 튜닝과 재학습 과정에서 성능 지표를 기록하는 것에 한계가 있고 모델간 비교가 어려움

## 해결

Wandb를 연동하여 학습 로그를 시각화하고 실험 이력을 대시보드로 확인하며 최적의 파라미터를 체계적으로 선별 가능.

실험의 재현성을 확보하고 데이터에 기반한 모델 개선 근거 정립.

효율적인 실험 관리 및 모델 최적화 가능

# 기술적 내용 및 해결방법 #3

## Wandb 도입

Runs (6)

Search runs

Filter Group Sort New sweep

Name	State	Created	Runtime	Model Type	Params.Learnr	Params.Max_Dt	Params.N_Est	Params.Rando	Params.Subsai	Rmse_Gap	Test_Rmse	Train_Rmse
xgb..._2026-01-27_depth4	Finished	2d ago	4s	xgboost	0.05	4	200	42	0.7	0.9054	0.90007	0.074075
xgb..._2026-01-27_depth4	Finished	2d ago	2s	xgboost	0.05	4	200	42	0.7	0.86418	0.94488	0.080299
xgb..._2026-01-27_depth4	Finished	2d ago	3s	xgboost	0.05	4	200	42	0.7	0.87507	0.95378	0.078715
xgb..._2026-01-27_depth4	Finished	2d ago	4s	xgboost	0.05	4	200	42	0.7	0.85509	0.93081	0.075723
xgb..._2026-01-26_depth4	Finished	2d ago	4s	xgboost	0.05	4	200	42	0.7	0.85509	0.93081	0.075723
xgb..._2026-01-21_depth4	Finished	4d ago	2s	xgboost	0.05	4	200	42	0.7	0.95973	1.04952	0.089794
xgb..._2026-01-21_depth4	Finished	4d ago	44s	xgboost	0.05	4	200	42	0.7	0.73805	0.82114	0.083285
xgb..._2026-01-24_depth4	Finished	4d ago	4s	xgboost	0.05	4	200	42	0.7	0.73805	0.82114	0.083285
xgb..._2026-01-21_depth4	Finished	6d ago	3s	xgboost	0.05	4	200	42	0.7	0.89036	0.87997	0.079811
xgb..._2026-01-23_depth4	Finished	6d ago	1s	xgboost	0.05	4	200	42	0.7	0.88314	0.95751	0.074372
xgb..._2026-01-20_depth4	Finished	6d ago	3s	xgboost	0.05	4	200	42	0.7	-	0.97415	-
xgb..._2026-01-23_depth4	Finished	6d ago	2s	xgboost	0.05	4	200	42	0.7	0.8924	0.90770	0.074953
xgb..._2026-01-04	Finished	7d ago	3s	xgboost	0.05	4	200	42	0.7	-	0.82486	-

yoon0125 - Projects > movie-rating-predictor > Workspace

Yoon Chung  
yoon0125.org

Project Yoon0125's workspace Personal workspace

Runs 61

Search panels with regex

Settings + New report + Add panels

Charts 3

train\_rmse

xgb...\_2026-01-28\_depth4  
xgb...\_2026-01-28\_depth4  
xgb...\_2026-01-27\_depth4  
vhv...\_2026-01-27\_depth4  
vhv...\_2026-01-27\_depth4

test\_rmse

xgb...\_2026-01-28\_depth4  
xgb...\_2026-01-28\_depth4  
xgb...\_2026-01-27\_depth4  
vhv...\_2026-01-27\_depth4  
vhv...\_2026-01-27\_depth4

rmse\_gap

xgb...\_2026-01-28\_depth4  
xgb...\_2026-01-28\_depth4  
xgb...\_2026-01-27\_depth4  
vhv...\_2026-01-27\_depth4  
vhv...\_2026-01-27\_depth4

System 10

# 기술적 내용 및 해결방법 #4~5

1. 시스템 가용성 (Scale-out)
2. 데이터 신뢰성 (Data Leakage)
3. 실험 관리 (Wandb)

## 4. 아키텍처 고도화 (S3)

## 5. 데이터 일관성 (Encoding)

### 문제

- 1) 실무 환경에서 인스턴스간 직접 전송(SCP)방식 위험성 인지
- 2) 학습과 서빙 시의 인코딩 매핑값이 달라져 예측 결과가 왜곡

### 해결

- 1) 인스턴스간 직접 전송 방식 폐기 → AWS S3 모델 레지스트리 활용한 모델 아티팩트 관리 방식 도입. 서빙 인스턴스와의 의존성 제거
- 2) LabelEncoder를 .pkl 파일로 저장하여 배포단계에 포함.  
학습-검증-서빙 전 과정에서 데이터 일관성 확보

# 기술적 내용 및 해결방법 #4~5

## S3 기반 모델 레지스트리

The screenshot shows the AWS S3 console interface. The left sidebar has sections for 'Amazon S3', 'Bucket' (with 'Default Bucket', 'Directory Bucket', 'Table Bucket', 'Lambda Bucket'), 'Access Management' (with 'AWS IAM Access Management', 'AWS Lambda', 'AWS Lambda Layers'), and 'Storage Management' (with 'AWS Storage Lens', 'AWS CloudWatch Metrics'). The main content area shows the 'latest/' folder of the 'mlops-bucket-user1' bucket. It lists three objects: 'best\_model.pkl', 'main\_genre\_encoder.pkl', and 'original\_language\_encoder.pkl'. Each object is shown with its file type (pkl), last modified date (2026. 1. 28. pm), size (279.1KB, 730.0B, 663.0B), and storage class (Standard). There are checkboxes for selecting multiple objects.

선택	이름	유형	마지막 수정	크기	스토리지 클래스
<input type="checkbox"/>	<a href="#">best_model.pkl</a>	pkl	2026. 1. 28. pm 5:29:54 PM KST	279.1KB	Standard
<input type="checkbox"/>	<a href="#">main_genre_encoder.pkl</a>	pkl	2026. 1. 28. pm 5:29:54 PM KST	730.0B	Standard
<input type="checkbox"/>	<a href="#">original_language_encoder.pkl</a>	pkl	2026. 1. 28. pm 5:29:54 PM KST	663.0B	Standard

# 프로젝트 산출물

## Airflow

The screenshot shows the Airflow web interface with the following details:

- Header:** Airflow, DAGs, Cluster Activity, Datasets, Security, Browse, Admin, Docs.
- Date:** 2026-01-28 오후 02:06:38
- Run ID:** 25
- Filter Buttons:** All Run Types, All Run States, Clear Filters.
- Graph View:** DAG tmdb\_movie\_rating\_pipeline, Run 2026-01-28, 12:59:08 KST / train\_and\_save\_best\_model. The graph shows tasks: extract\_tmdb\_data, preprocess\_movie\_data, train\_and\_save\_best\_model, and deploy\_model\_to\_fastapi. The 'train\_and\_save\_best\_model' task is failed.
- Logs View:** Task: deploy\_model\_to\_fastapi, Run: manual\_2026-01-27T03:59:39.346433+00:00. The logs show an upstream failed error.

The screenshot shows the Airflow web interface with the following details:

- Header:** Airflow, DAGs, Cluster Activity, Datasets, Security, Browse, Admin, Docs.
- Date:** 2026-01-28 오후 02:06:38
- Run ID:** 25
- Filter Buttons:** All Run Types, All Run States, Clear Filters.
- Graph View:** DAG tmdb\_movie\_rating\_pipeline, Run 2026-01-28, 12:59:08 KST / train\_and\_save\_best\_model. The graph shows tasks: extract\_tmdb\_data, preprocess\_movie\_data, train\_and\_save\_best\_model, and deploy\_model\_to\_fastapi. The 'train\_and\_save\_best\_model' task is failed.
- Logs View:** Task: deploy\_model\_to\_fastapi, Run: manual\_2026-01-27T03:59:39.346433+00:00. The logs show an upstream failed error.

# 프로젝트 산출물

## FastAPI

FastAPI 0.1.0 OAS 3.1 /openapi.json

**default**

**GET /predict Predict**

Parameters

Name	Description
runtime * required	int<0> (max=2)
genre_encoded * required	int<0> (max=2)
lang_encoded * required	int<0> (max=2)
sentiment * required	float (max=1.0)
overview_len * required	int<0> (max=2)

Cancel

Execute Clear

Responses

Code curl -X POST '\

Request URL [https://13.13.86.36:8000/predict?runtime=100&genre\\_encoded=1&lang\\_encoded=0&sentiment=-0.2&overview\\_len=22](https://13.13.86.36:8000/predict?runtime=100&genre_encoded=1&lang_encoded=0&sentiment=-0.2&overview_len=22)

Server response

Code	Details
200	<p>Response body</p> <pre>{   "predicted_rating": 7.385941501433129 }</pre> <p>Response headers</p> <pre>content-type: application/json date: Wed, 28 Dec 2022 08:43:31 UTC server: uvicorn</pre>

Responses

Code	Description	Links
200	Successful Response	No links
422	Validation Error	No links

Media type application/json

Controls Accept header.

Example Value Schema

"string"

Media type application/json

Example Value Schema

```
{
  "detail": [
    {
      "loc": [
        "string",
        0,
        "string"
      ],
      "msg": "string",
      "type": "string"
    }
  ]
}
```

# 프로젝트 산출물

## Streamlit 기반 UI 설계

### AI 영화 평점 예측기 (다국어 지원)

입력하신 영화 정보를 바탕으로 약 70-80% 정확도로 예상 평점을 분석합니다.

줄거리의 분위기(감성)까지 점수에 반영됩니다!

상영시간(분)

- +

언어

▼

주요 장르

▼

**영화 줄거리**

이곳에 영화의 줄거리를 입력하세요. 길이에 따라 예측값이 달라집니다.  
한글로 입력하셔도 AI가 번역하여 분석합니다.

**예상 평점 확인**

### AI 영화 평점 예측기 (다국어 지원)

입력하신 영화 정보를 바탕으로 약 70-80% 정확도로 예상 평점을 분석합니다.

줄거리의 분위기(감성)까지 점수에 반영됩니다!

상영시간(분)

- +

언어

▼

주요 장르

▼

**영화 줄거리**

살아있는 기적이라 불리는 시각장애인 전각 장인 '임영규'의 아들 '임동환'이, 40년 전 실종된 줄 알았던 어머니의 백골 시신 발견 후, 그 죽음 뒤의 진실을 파헤치는 이야기

**예상 평점 확인**

**분석 결과**

예상 평점 **6.47 / 10**      줄거리 분위기  
어둡고/부정적 ⚪ (점수: -0.37)

💡 이 시뮬레이션은 'Drama' 장르와 'ko' 언어의 특성을 반영한 결과입니다.

# 프로젝트 산출물

## 프로젝트 관리 (Notion, Git)

**MLOps 프로젝트**

TMDB 데이터를 활용한 영화 평점 예측 서비스 개발 및 MLOps 파이프라인 구축

모든 작업 담당자별 업무

Ao 작업	상태	담당자	마감일	우선순위	설명
■ 프로젝트 기초 공부	완료		01/21/2026	중간	프로젝트 진행 전 MLOps 공부, 참고자료
▶ ■ 주제 및 방향 설정	완료		01/21/2026	중간	프로젝트 주제 및 업무 분장
■ Data 수집	완료		01/21/2026	높음	데이터 크롤링 진행 # Data source : <a href="https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset/data#ratings_small.csv">https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset/data#ratings_small.csv</a> . The subset of 100,000 ratings from 700 users on 9,000 movies
○ 1차 엔터프	완료		01/22/2026		설계 점검
■ 전체 아키텍처 설계	진행 중		01/23/2026	높음	엔터프 전, 시전 아키텍처 설계
■ Data 전처리	완료		01/23/2026	높음	이상치, 결측치, 인코딩, 데이터 누수, 출거리분석
■ 모델링	완료		01/23/2026	높음	LightGBM, RandomForest, XGBoost 등 모델 선택, 파리미터 선택
■ 모델 배포	진행 중		01/26/2026	중간	웹앱, API 배포
■ Automated pipeline	진행 중		01/26/2026	높음	성능 경기 방식에 따른 베스트 모델 선정, 저장
○ 2차 엔터프	시작 전		01/26/2026		구현 중간 점검
■ Serving pipeline 구축	진행 중		01/27/2026	높음	모델 추론에 따른 결과값 서빙
■ HTML 구축	시작 전		01/27/2026	중간	결과값 출력을 HTML로 구축
■ EC2 서버 구축 및 환경 설정	시작 전		01/28/2026	낮음	
■ EC2 서버에 모델 파일 업로드/테스팅	시작 전		01/28/2026	높음	테스팅
○ 3차 엔터프	시작 전		01/28/2026		완성도 점검
■ PPT 작성	시작 전		01/29/2026	중간	발표자료 작성

개수 18

feat/add language\_encoder #4

Merged Jun-Yoont merged 1 commit into main from scripts 3 hours ago

Conversation 0 Commits 1 Checks 0 Files changed 1

yoon-chung commented yesterday

무엇을 했나요?

-original\_language 피쳐 추가, language\_encoder 추가

왜 필요한가요?

-모델 성능 개선

테스트 방법

- 

체크리스트

- [o] 코드 실행 확인
- [o] 브랜드 및 코드 체크

No reviews

No one—assign yourself

No labels

No project

No milestone

Development

Successfully merging this pull request may close these issues.

No yet

feat/add language\_encoder #4

Merged Jun-Yoont merged 1 commit into main from scripts 3 hours ago

Conversation 0 Commits 1 Checks 0 Files changed 1

Changes from all commits • File filter • Conversations • Jump to •

feat/add language\_encoder

main [4]

yoon-chung commented yesterday

```

+ 2 14 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 279 280 281 282 283 284 285 286 287 288 289 289 290 291 292 293 294 295 296 297 298 299 299 300 301 302 303 304 305 306 307 308 309 309 310 311 312 313 314 315 316 317 318 319 319 320 321 322 323 324 325 326 327 328 329 329 330 331 332 333 334 335 336 337 338 339 339 340 341 342 343 344 345 346 347 348 349 349 350 351 352 353 354 355 356 357 358 359 359 360 361 362 363 364 365 366 367 368 369 369 370 371 372 373 374 375 376 377 378 379 379 380 381 382 383 384 385 386 387 388 389 389 390 391 392 393 394 395 396 397 398 399 399 400 401 402 403 404 405 406 407 408 409 409 410 411 412 413 414 415 416 417 418 419 419 420 421 422 423 424 425 426 427 428 429 429 430 431 432 433 434 435 436 437 438 439 439 440 441 442 443 444 445 446 447 448 449 449 450 451 452 453 454 455 456 457 458 459 459 460 461 462 463 464 465 466 467 468 469 469 470 471 472 473 474 475 476 477 478 479 479 480 481 482 483 484 485 486 487 488 489 489 490 491 492 493 494 495 496 497 498 499 499 500 501 502 503 504 505 506 507 508 509 509 510 511 512 513 514 515 516 517 518 519 519 520 521 522 523 524 525 526 527 528 529 529 530 531 532 533 534 535 536 537 538 539 539 540 541 542 543 544 545 546 547 548 549 549 550 551 552 553 554 555 556 557 558 559 559 560 561 562 563 564 565 566 567 568 569 569 570 571 572 573 574 575 576 577 578 579 579 580 581 582 583 584 585 586 587 588 589 589 590 591 592 593 594 595 596 597 598 599 599 600 601 602 603 604 605 606 607 608 609 609 610 611 612 613 614 615 616 617 617 618 619 619 620 621 622 623 624 625 626 627 628 629 629 630 631 632 633 634 635 636 637 638 639 639 640 641 642 643 644 645 646 647 648 649 649 650 651 652 653 654 655 656 657 658 659 659 660 661 662 663 664 665 666 667 668 669 669 670 671 672 673 674 675 676 677 678 679 679 680 681 682 683 684 685 686 687 688 689 689 690 691 692 693 694 695 696 697 698 699 699 700 701 702 703 704 705 706 707 708 709 709 710 711 712 713 714 715 716 717 717 718 719 719 720 721 722 723 724 725 726 727 728 729 729 730 731 732 733 734 735 736 737 738 739 739 740 741 742 743 744 745 746 747 748 749 749 750 751 752 753 754 755 756 757 758 759 759 760 761 762 763 764 765 766 767 768 769 769 770 771 772 773 774 775 776 777 778 779 779 780 781 782 783 784 785 786 787 788 789 789 790 791 792 793 794 795 796 797 798 799 799 800 801 802 803 804 805 806 807 808 809 809 810 811 812 813 814 815 816 817 817 818 819 819 820 821 822 823 824 825 826 827 828 829 829 830 831 832 833 834 835 836 837 838 839 839 840 841 842 843 844 845 846 847 848 849 849 850 851 852 853 854 855 856 857 858 859 859 860 861 862 863 864 865 866 867 868 869 869 870 871 872 873 874 875 876 877 878 879 879 880 881 882 883 884 885 886 887 888 889 889 890 891 892 893 894 895 896 897 898 899 899 900 901 902 903 904 905 906 907 908 909 909 910 911 912 913 914 915 916 917 917 918 919 919 920 921 922 923 924 925 926 927 928 929 929 930 931 932 933 934 935 936 937 938 939 939 940 941 942 943 944 945 946 947 948 949 949 950 951 952 953 954 955 956 957 958 959 959 960 961 962 963 964 965 966 967 968 969 969 970 971 972 973 974 975 976 977 978 979 979 980 981 982 983 984 985 986 987 988 989 989 990 991 992 993 994 995 996 997 998 999 999 1000 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1009 1010 1011 1012 1013 1014 1015 1016 1017 1017 1018 1019 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1109 1110 1111 1112 1113 1114 1115 1116 1117 1117 1118 1119 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1198 1199 1199 1200 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1209 1210 1211 1212 1213 1214 1215 1216 1217 1217 1218 1219 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1289 1290 1291 1292 1293 1294 1295 1296 1297 1297 1298 1299 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1309 1310 1311 1312 1313 1314 1315 1316 1317 1317 1318 1319 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1389 1390 1391 1392 1393 1394 1395 1396 1397 1397 1398 1399 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1409 1410 1411 1412 1413 1414 1415 1416 1417 1417 1418 1419 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1498 1499 1499 1500 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1509 1510 1511 1512 1513 1514 1515 1516 1517 1517 1518 1519 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1598 1599 1599 1600 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1609 1610 1611 1612 1613 1614 1615 1616 1617 1617 1618 1619 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1698 1699 1699 1700 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1709 1710 1711 1712 1713 1714 1715 1716 1717 1717 1718 1719 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1798 1799 1799 1800 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1809 1810 1811 1812 1813 1814 1815 1816 1817 1817 1818 1819 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1898 1899 1899 1900 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1909 1910 1911 1912 1913 1914 1915 1916 1917 1917 1918 1919 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1998 1999 1999 2000 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2009 2010 2011 2012 2013 2014 2015 2016 2017 2017 2018 2019 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2098 2099 2099 2100 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2109 2110 2111 2112 2113 2114 2115 2116 2117 2117 2118 2119 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2
```

# 소감 및 향후 계획

## 성장한 내용

모델 코드 모듈화와 컨테이너 기반 환경 구축, AWS, Docker를 활용하여 로컬 개발 환경을 클라우드에 구현하는 역량을 쌓았습니다.

특히 리소스 간섭 등 여러 기술적 문제를 해결하는 과정에서, DAG파일 편집, Docker 컨테이너 기반 기술, airflow workflow 구성, S3 연동 등 엔지니어링 기술에 조금 친숙해진 시간이었습니다.

또한 Wandb (Weights&Biases)를 활용한 실험 기록 관리를 통해 효율적인 하이퍼파라미터 튜닝 프로세스를 경험하였습니다.

## 향후 계획

본 프로젝트에서는 AWS S3기반 모델 레지스트리를 구축하고 배포 파이프라인을 구현하며 MLOps 기틀을 마련했습니다. 향후 모델의 버전별 기록과 배포가 자동으로 관리되는 수준 높은 MLOps 환경을 경험해보고 싶습니다.

소스코드 <https://github.com/yoon-chung/MLOps.git>

감사합니다