



# | 유제승 1997. 01. 20.

Back-End / AI / DataAnalyst

□ 010-9014-4310

✉ wptmd1410@gmail.com

⌚ <https://github.com/Yu-JeSeung>

## Project

### 동숲

- 대학교 통합 커뮤니티 및 정보 제공 애플리케이션
- 2025.03 ~

### Easy Trip

- 이용자 성향 기반 데이트 코스 추천 AI
- 2024.03 ~ 2024.12

## Paper

### 한국차세대컴퓨팅학회 춘계학술대회 우수논문상

유제승, 우승원, 이현민, 김경열, 이희진  
"텍스트 마이닝 기반 국내 머신러닝 연구의 동향 분석  
. " 한국차세대컴퓨팅학회 학술대회 2025.05 (2025):  
199-202.

## Works & Skill

### (주) 서울금속 개발팀

- 시제품 개발 및 설계

2017.06 ~ 2021.07

[Python](#)   [MySQL](#)   [PostgresSQL](#)   [FastAPI](#)

[Docker](#)   [Nginx](#)   [AWS](#)   [Deep Learning](#)

## Education

### 동양미래대학교 학사 과정 재학 (4.07 / 4.5)

- 컴퓨터 소프트웨어공학과
- 2025.03 ~ 2026.02

### 동양미래대학교 전문 학사 과정 졸업 (3.98 / 4.5)

- 컴퓨터 소프트웨어공학과
- 2022.03 ~ 2025.02

### 두원공과대학교 전문 학사 과정 졸업 (3.88 / 4.5)

- 기계설계학과
- 2015.03 ~ 2018.02

### 현화고등학교 졸업

- 인문계(자연계열)
- 2012.03 ~ 2015.02

## Activity

### 구름톤 유니브 3기

2024.06 ~ 2024.12

### 학생자치기구 대의원회 의장

2024.01 ~ 2024.12

동양미래대학교 학생을 위한 통합형 커뮤니티 및 정보 제공 애플리케이션

## • 챗봇

: 크롤링한 학교 사이트 정보를 기반으로 RAG를 구성하고 대규모 언어 모델(GPT-OSS:20b)과 결합하여 개발

### 문제점

: 자연어 처리 로직과 교내 데이터를 학습한 모델을 같이 사용함에 있어 특정 단어를 라우팅하는 로직이 길어졌습니다.

### 개선점

: 핵심 쿼리 생성기 로직을 구현하여 Attention Weight를 만들어 라우팅 로직을 단순화하였습니다.



## • 시간표 자동 분석 (이미지 처리)

: Tesseract OCR 엔진을 활용하여 이미지에서 한글을 추출하는 모델을 사용하여 개발

### 문제점

: 시간표 이미지를 분석하다 보니 각 셀마다 확인하는 작업으로 인한 처리 속도가 저하되었습니다.

### 개선점

: 멀티 프로세싱 라이브러리 적용으로 작업을 병렬처리 진행하였습니다.

응답속도를 기준 대비 약 20초 정도 감소할 수 있었습니다.

## • 비속어 필터링

: KoELECTRA 기반 문장 분류 모델을 파인튜닝하여 개발

### 문제점

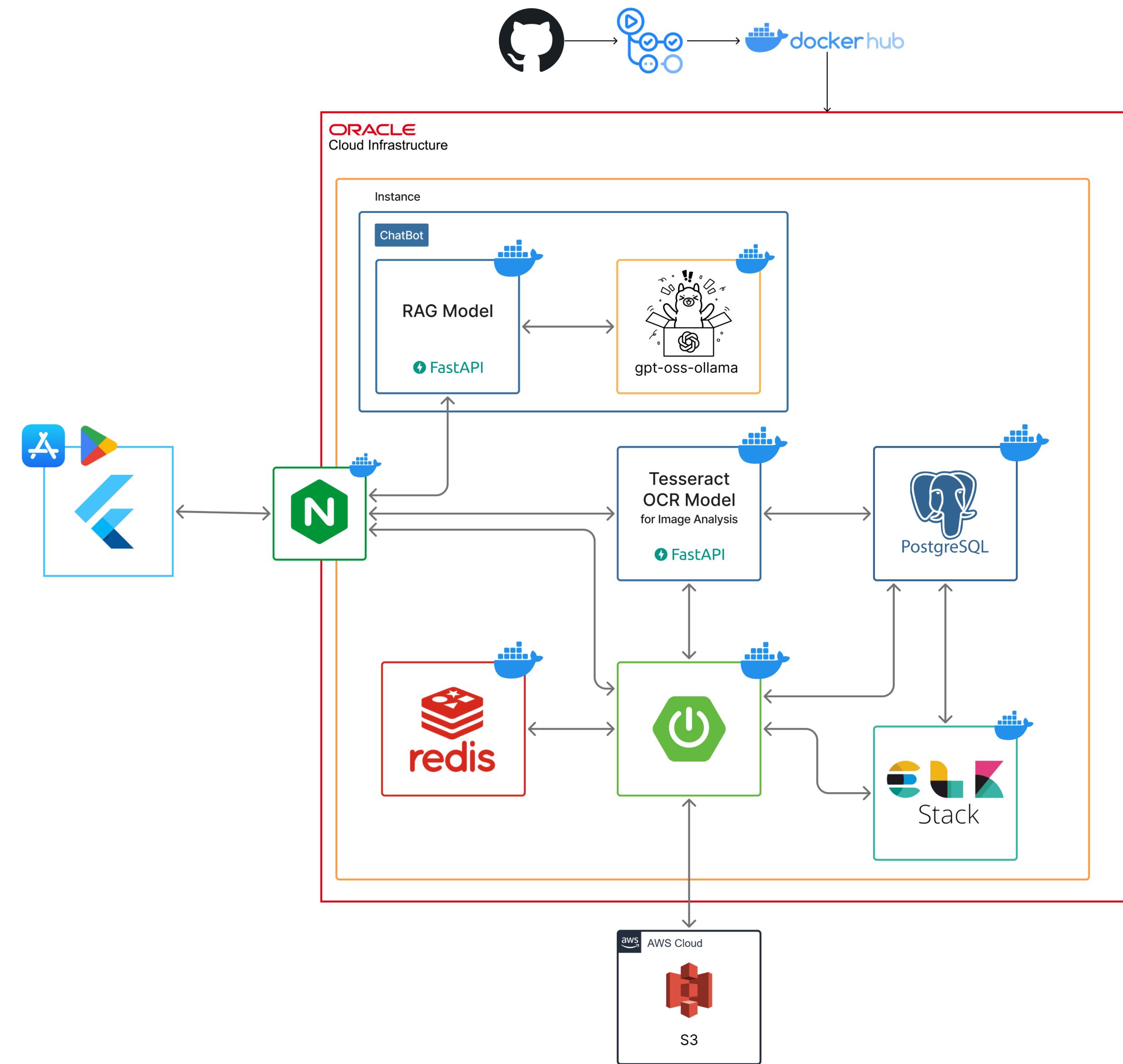
: 새롭게 등장하는 비속어를 필터링 하기에 지속적인 모델 개선이 필요했습니다.

### 개선점

: 클라이언트의 입력 데이터를 수집 및 라벨링 후 재학습하여 모델의 안정성을 높였습니다.

# Architecture

- 챗봇과 이미지 분석을 별도의 Docker 컨테이너로 분리하여 장애가 시스템 전체로 확산되지 않도록 하였습니다.
- 배포 시 신규 컨테이너를 생성하고 Health Check로 검증한 뒤 Blue-Green 방식을 사용하여 안전하게 트래픽을 전환하는 무중단 배포 환경을 구성하였습니다.
- Nginx를 도입하여 외부에서 접근할 수 없도록 설정하였고 모든 컨테이너는 내부 포트만 사용할 수 있도록 보안을 높였습니다.



# Eazy Trip

2024.03 ~ 2024.12

Github : <https://github.com/Ez-Trip>

이용자 성향 기반 데이트 코스 추천 AI

Python    Flask    LSTM    AWS    MySQL

- 이용자 성향 분석

#### 문제점

: 성향에 대한 선택 시 연속적인 단어로 문장을 구성하여 사용자 선택의 대한 값이 불분명함.

#### 개선점

: 선택한 값의 대한 단어를 저장하여 문장을 구성하지 않고,

특정 알파벳으로 매핑하여 순서를 더 간결하고 적은 문장으로 변환 가능.

ex) 숙박시설 - 음식점 - 선호하는 데이트순서

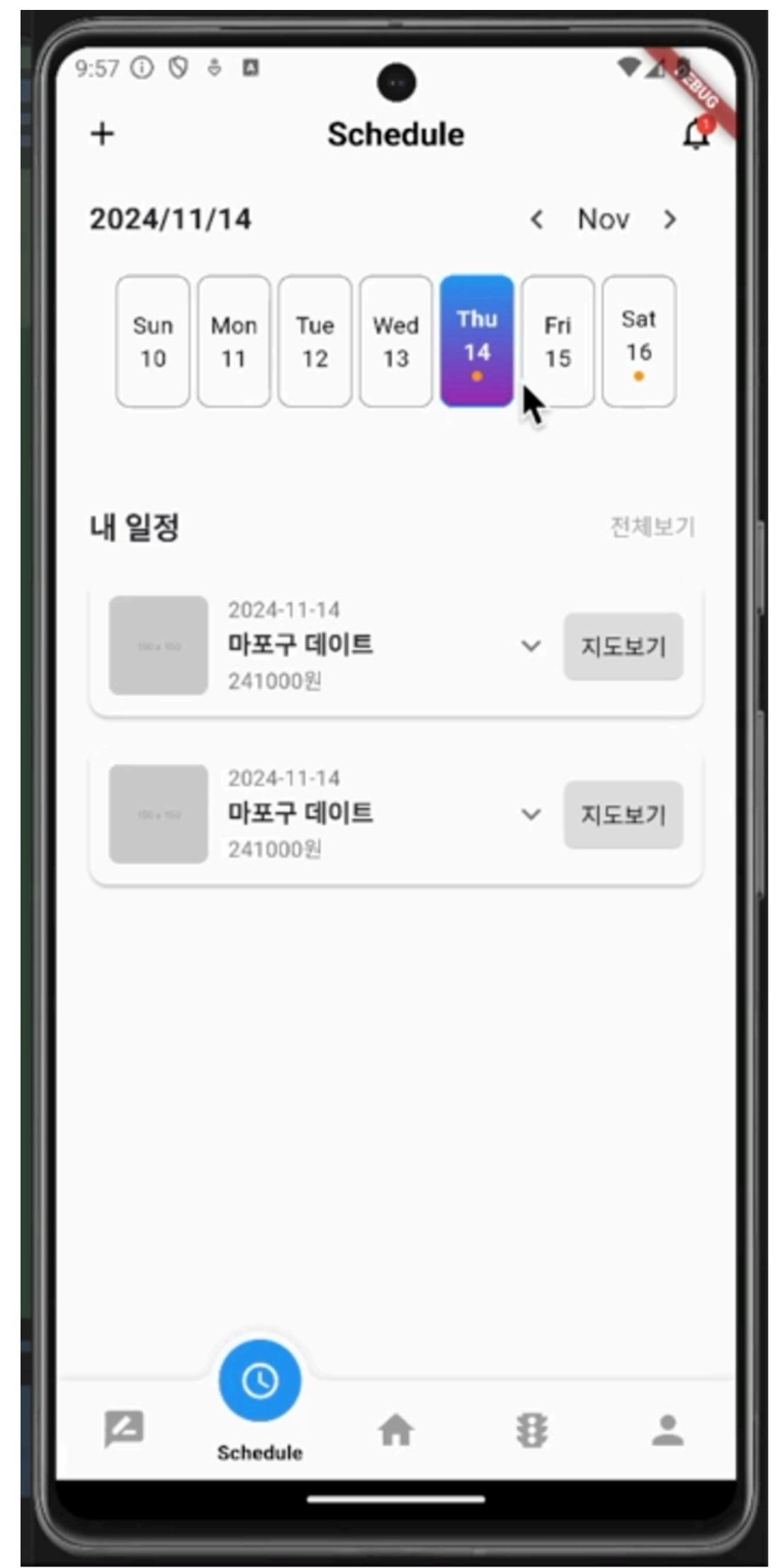
- 학습 과적합

#### 문제점

: 순환신경망(RNN)의 고질적인 문제점인 과적합으로 처음에 학습했던 데이터가 소실되고, 늦게 학습한 데이터의 가중치가 더 높아지는 문제점이 발생.

#### 개선점

: 순환신경망의 종류인 LSTM(Long Short-Term Memory model)을 사용하여 기울기 폭파로 인한 장기의존성 문제를 해결했습니다.



# Eazy Trip

2024.03 ~ 2024.12

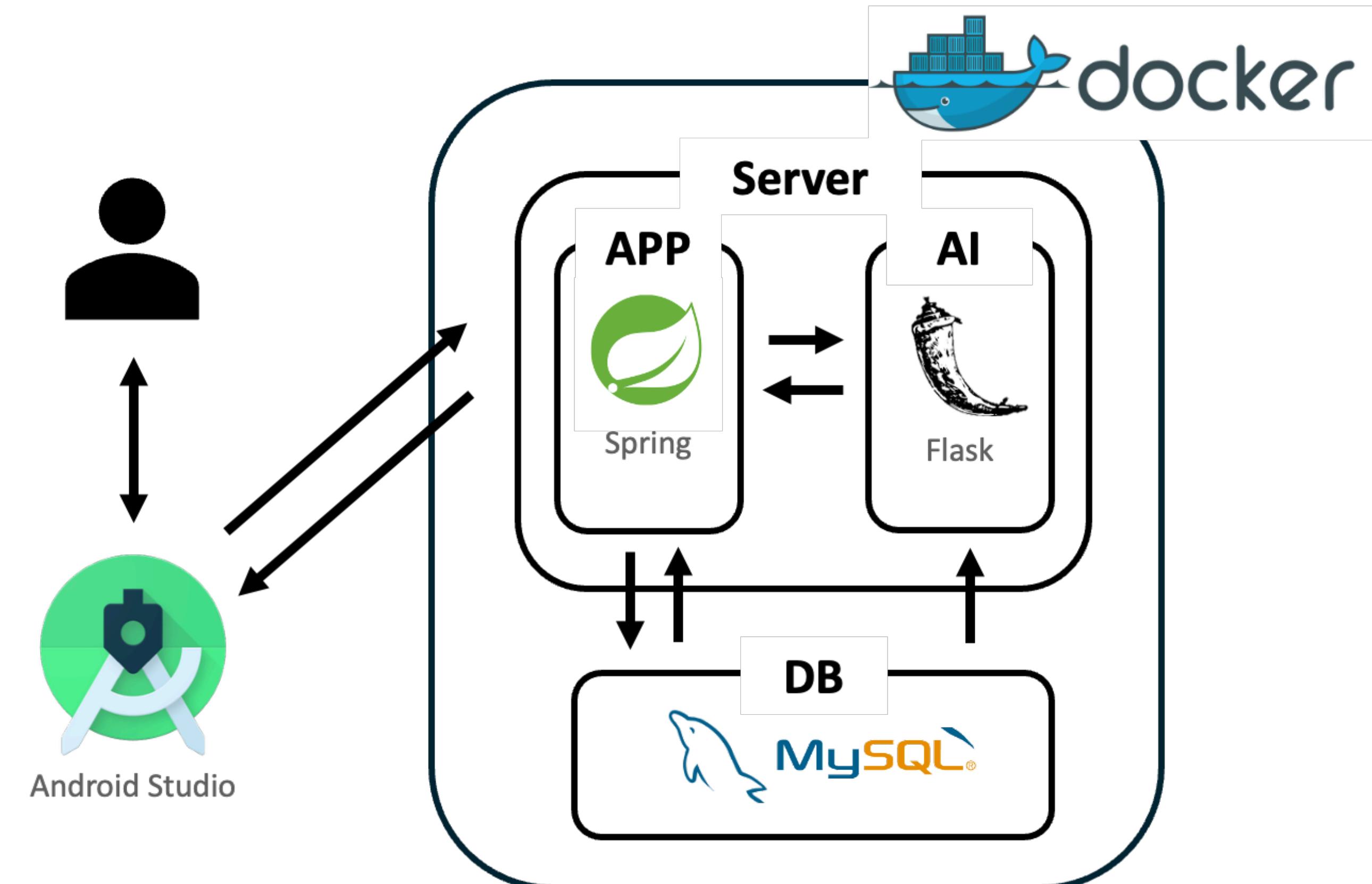
Github : <https://github.com/Ez-Trip>

이용자 성향 기반 데이트 코스 추천 AI

Python    Flask    LSTM    Spring Boot    MySQL

## Architecture

- 서버와 AI모델 서버를 별도의 Docker 컨테이너로 구성하여 서버별 이미지 분기점을 별도로 가지게 하였습니다.
- LSTM 알고리즘을 도입하더라도 업체의 평점을 기반하여 데이트코스를 추천하여 같은 데이트코스를 추천하는 문제점이 있었습니다. 한번 추천한 데이트코스는 별도의 로직을 통해 최소 10번 이후 같은 코스가 나오도록 설정하였습니다.
- AWS Ubuntu 서버를 활용하여 모델 추론을 진행하려했으나, 모델의 크기 및 추론속도를 감안하여 Flask 웹서버를 사용하였습니다.



# 논문

2025.04 ~ 2025.05

Python Tokenizer 텍스트 마이닝 데이터 시각화

"텍스트 마이닝 기반 국내 머신러닝 연구의 동향 분석 ." 한국차세대컴퓨팅학회 학술대회 2025.05 (2025): 199-202.  
유제승, 우승원, 이현민, 김경열, 이희진 Link : <https://www.riss.kr/>

## • 전처리 과정

### 문제점

: 논문의 초록을 수집하여 전처리를 진행하는 도중 한글 및 영어가 같이 나오는 문제점을 발견  
또한, 토큰화 과정 중 분석에 차질을 유발하는 불용어들이 대거 발견되었습니다.

### 개선점

: 전체 데이터를 한글과 영어로 토큰화 및 불용어 제거를 하고  
같은 의미의 단어를 하나로 합치는 과정을 통해 분석에 차질이 생기는 노이즈를 제거하였습니다.

## • Word2Vec

### 문제점

: TF-IDF의 결과값을 기준으로 유사도가 높은 단어를 사용하였지만  
중심단어를 기준으로 포괄적인 범위가 측정되어 범용단어가  
더 많이 노출되는 문제점이 발생하였다.

### 개선점

: Word2Vec의 Skip-gram 방식을 사용하여 특정 범위와 최소빈도수를 제한하여  
주요 알고리즘을 도출하였습니다.

