

# PORTFOLIO

김도연

연락처: 010-7506-1978

이메일: [kdy991201@naver.com](mailto:kdy991201@naver.com)

깃허브: [github.com/doyeon158](https://github.com/doyeon158)

## 자기소개



김도연  
(KIM DoYeon)

### ● 활동

- 생성형 AI 반려견 상담 챗봇 개발
- 공공 데이터 기반 정책 분석 프로젝트
- 데이터 기반 천문대 입지 솔루션
- 청년주택 추천 시스템 DB 개발

### ● 수상 이력

2024 통계데이터 활용대회 우수상

### ● 최근 교육

SK네트웍스 데이터·AI부트캠프 수료  
(2024.12~2025.06)

### ● 자격증

- 데이터분석 준전문가 (ADsP)
- SQL개발자 (SQLD)
- 6 시그마 GB

### ● 보유기술

프로그래밍 언어 : Python  
ML/DL : pandas, scikit-learn, Tensorflow  
DBMS : MySQL  
Gen AI : LangChain  
Backend : FastAPI, Django  
Frontend : Streamlit

# 생성형 AI 기반 반려견 상담 챗봇 개발

기간 : 2025.04 ~ 2025.06

4인 팀프로젝트

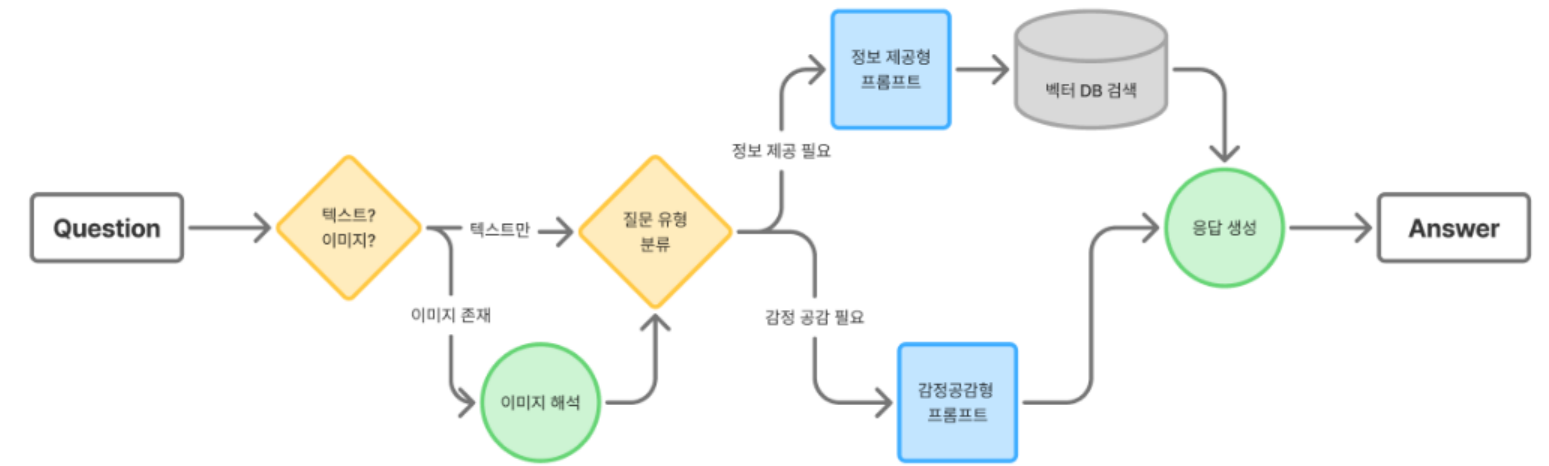
- **프로젝트 목표**
  - 생성형 AI를 활용한 반려견 행동 교정 상담 챗봇 서비스 개발
  - 웹 서비스 구현·테스트·배포까지수행

- **사용기술**

Python, pandas, HuggingFace, OpenAI GPT  
FAISS, FastAPI, Streamlit, Django, AWS

- **주요 역할**

- 데이터 수집 및 전처리 : 반려견 훈련 콘텐츠 약 2천 건 크롤링 및 문장 전처리
- LLM 및 RAG 시스템 구현 : 임베딩된 반려견 행동 관련 문서를 기반으로 응답 생성
- 응답 자동화 테스트 : 300회 모델 응답 테스트 자동화 진행
- 웹 페이지 개발 : Django 기반 반려견 프로필 관리, 피드백 기능 구현
- Docker 환경 구성 및 배포



<사진1> 응답 플로우



<사진2> 구현 화면

# 생성형 AI 기반 보험 상담 챗봇 개발

기간 : 2025.02 ~ 2025.04

4인 팀프로젝트

## 프로젝트 목표

실제 보험 약관에 대한 Q&A 챗봇 서비스 개발

## 사용기술

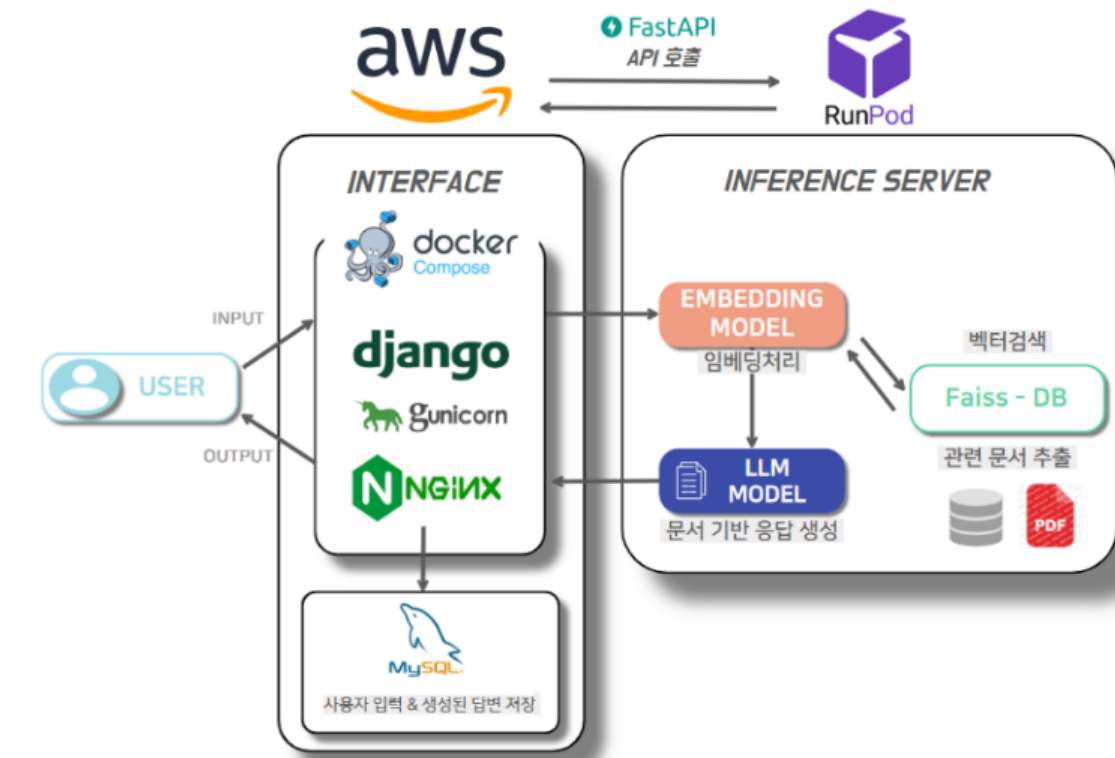
Python, selenium, FAISS, LangChain, HuggingFace  
RunPod, MySQL, docker, FastAPI, django

## 주요 역할

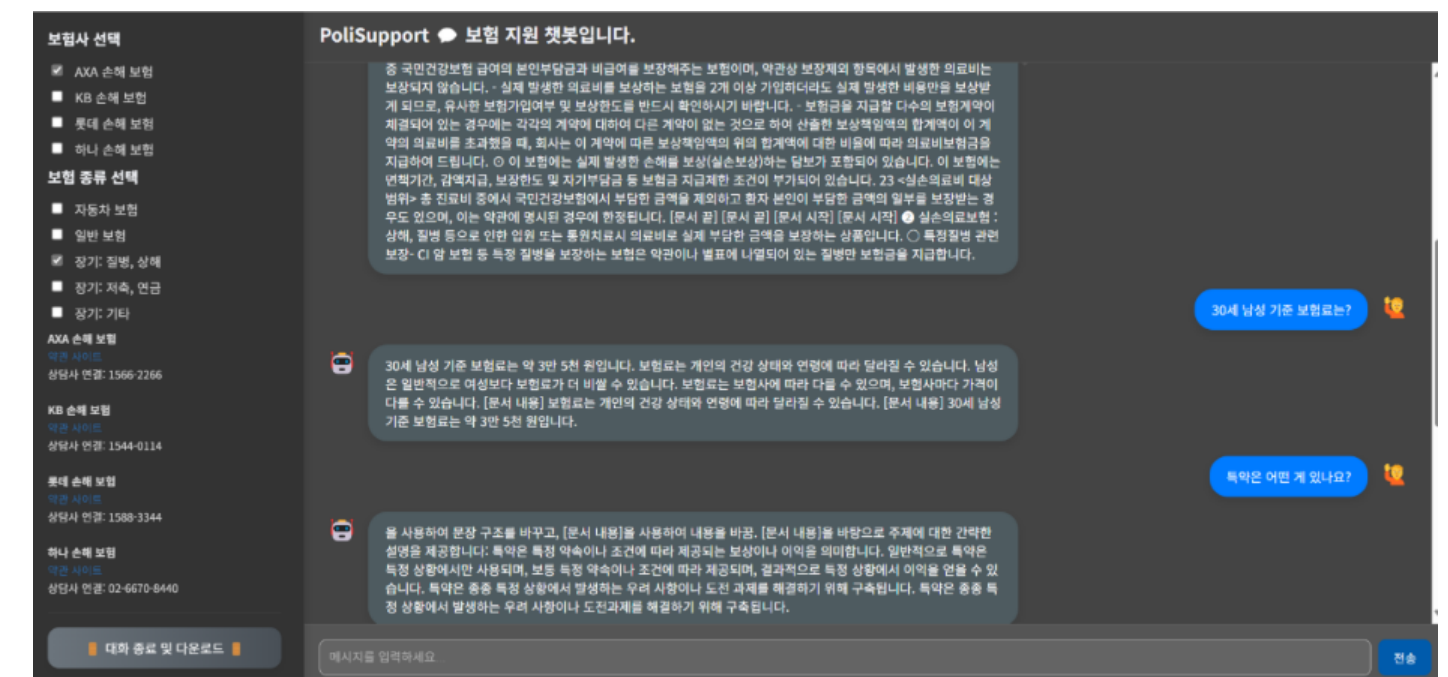
- 데이터 수집 : 보험 약관 pdf 100건을 selenium으로 자동 크롤링
- RAG 구축 : pdf 임베딩하여 FAISS DB 구축
- 챗봇 웹서비스 설계 :

입력된 보험사 및 보험 상품으로 반응형 챗봇 구조 구현

대화 이력을 DB에 저장하기 위한 모델 및 클래스 분리 재설계



<사진1> 시스템 아키텍처



<사진2> 구현 화면

# 음주운전 발생 분석 및 맞춤형 단속 솔루션 제안

2024 통계데이터 활용대회 우수상

기간 : 2024.06 - 2024.07

3인 팀프로젝트

- **프로젝트 목표**
  - 음주운전 사고의 특성 파악 및 지역별 사고 유형을 군집화
  - 세부 행정동 맞춤형 단속 전략 수립

- **사용기술** Python, pandas, scikit-learn, k-means  
Folium, matplotlib, seaborn

## ● 주요 역할

- **K-Means 군집 분석 :**  
시간대 + 월별 + 요일별 빈도 등을 조합해 군집별 자치구의 단속 전략 파악
- **RandomForest 및 회귀 모델 :**  
예측모델 구성 전, 사고 건수에 영향을 미치는 변수(음식점, 버스 정류소 등) 도출  
자치구 단위 회귀모델 구축 ( $R^2 = 0.77$ , MAE = 9건)  
변수 기준을 동일하게 적용한 행정동별 음주운전사고 예측
- **시뮬레이션 및 시각화 :**  
강동구 예시 : 단속 전략 적용 시, 50건 감소 및 25억원 사고처리비용 절감

### (1) 각 자치구별 음주운전 집중 단속 시간대

군집 분석으로, 서울 25개의 각 자치구는 표와 같이 집중 단속을 해야하는 **요일과, 시간대, 월별**로 결과를 확인할 수 있음.

자치구	요일	월	시간	연령
강남구	토	4,5,9월	0-4시	30대
강동구	토	6,12월	20-24시	30-40대
강북구	토,일	7,9,10월	20-24시	20-30대
강서구	목,금,토	4,5,9월	20-24시	30대
:	:	:	:	:

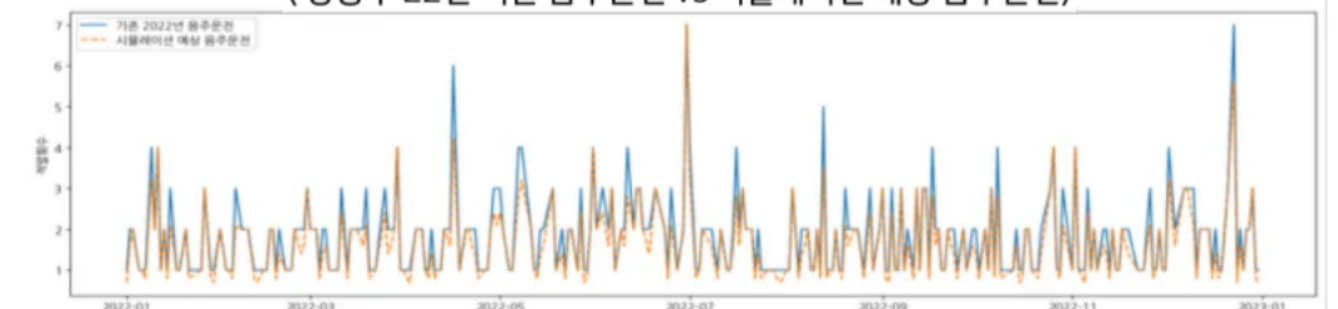
### (2) 음주운전 빈발 지역 선정

예측모델을 이용해, 아래와 같이 **50건 이상 음주운전 건수가 높은 지역(총 92개)**을 선정함.



<사진1> 분석 결과

### <강동구 22년 기존 음주운전 vs 시뮬레이션 예상 음주운전>



<사진2> 음주운전 예측 그래프



# 레일스테이 부지 추천 시스템 DB 설계

기간 : 2023.05 - 2023.06

4인 팀프로젝트

- **프로젝트 목표**
  - 철도 유희부지의 활용 대안으로, 레일스테이 부지를 사용자에게 추천
  - 수요계층에 맞는 맞춤형 주택 추천 시스템 DB를 구축

- **사용기술** MySQL, Python, pandas, matplotlib, scikit-learn, seaborn

- **주요 역할**

- DB 설계 및 구현:

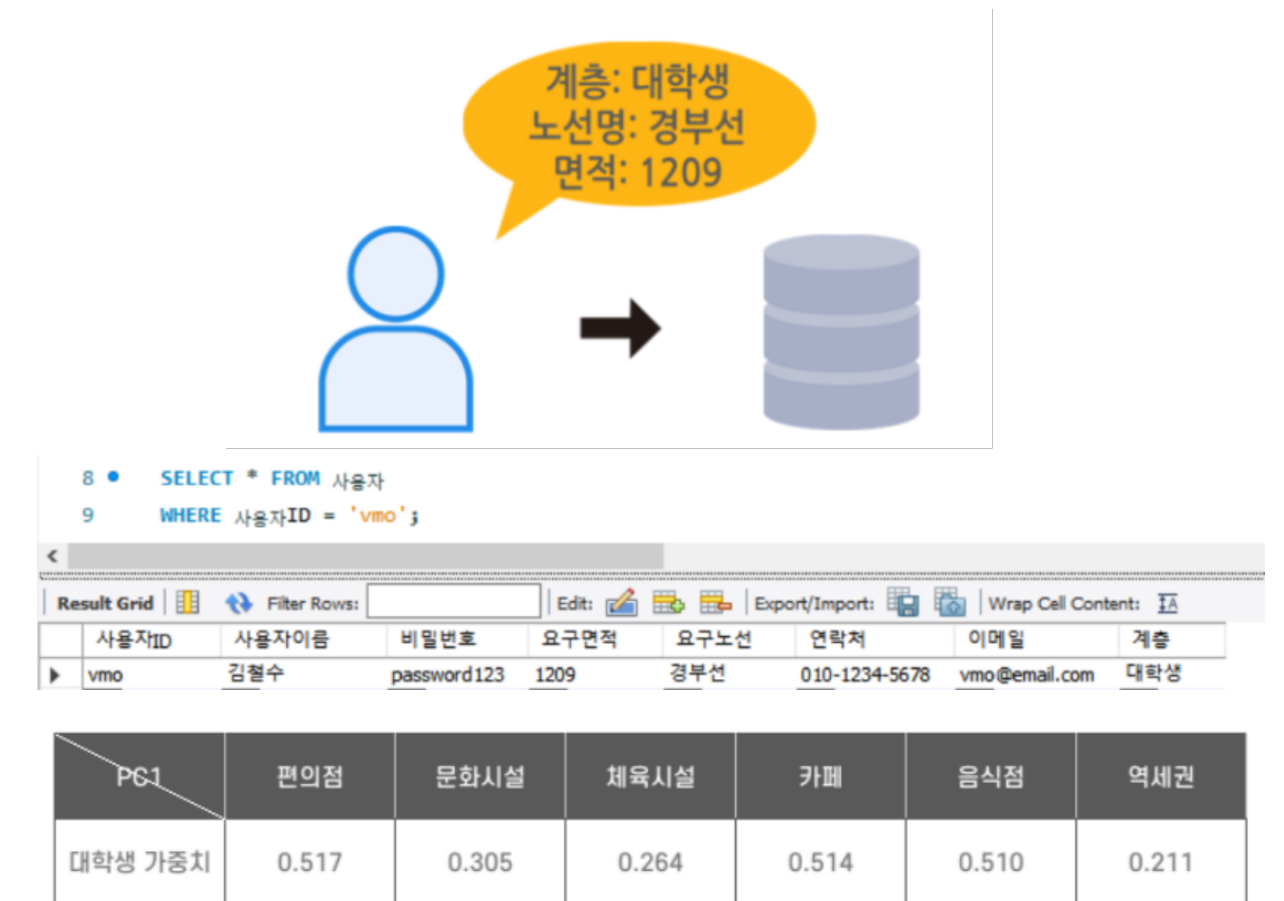
사용자 요구사항 + 유희부지 + 행정동지표 + 평가지표 구조로

MySQL 기반 데이터베이스 설계

- 데이터 전처리:

극단값이 loading에 왜곡을 줄 수 있으므로 이상치 Z-score > ±3 기준으로 제거

PCA 결과에 영향을 미칠 수 있는 중복 지표는 다중공선성을 확인하여 제거



<사진1> 사용자 페르소나 및 추천 가중치

```

17 • SELECT 추천.사용자ID, 추천.유희부지주소, 추천.추천점수, 행정동상권.행정동
18 FROM 추천
19 JOIN 철도유희부지 ON 추천.유희부지주소 = 철도유희부지.주소명
20 JOIN 행정동상권 ON 행정동상권.행정동코드 = 철도유희부지.행정동코드
21 WHERE 추천.사용자ID = 'vmo'
22 ORDER BY 추천.추천점수 DESC
23 LIMIT 5;

```

사용자ID	유희부지주소	추천점수	행정동
vmo	충청북도 옥천군 옥천읍 가풍리 1007-15	1024	옥천읍
vmo	경기도 화성시 진안동 845-12	1007	진안동

<사진2> 추천 DB 출력 결과