

포트폴리오

AI 개발자 이루오

2021.09 ~ 2023.02

[DL 연구개발] zero-shot classification

2023.01 ~ 2023.02

기업 평가 데이터를 zero-shot 모델을 사용하여 unseen labels로 분류하는 웹앱

[개발 환경]

- ubuntu 22.04
- Python 3.9
- k8s, docker
- Git
- TeamCity

[사용 기술]

1. DB API
 - pymongo
2. 데이터 전처리 및 NLP
 - pandas
 - kiwipiepy
3. NLI model
 - PyTorch, transformers
4. 시각화
 - matplotlib, plotly, wordcloud, squarify
5. 웹앱
 - streamlit

[주요 업무 및 성과]

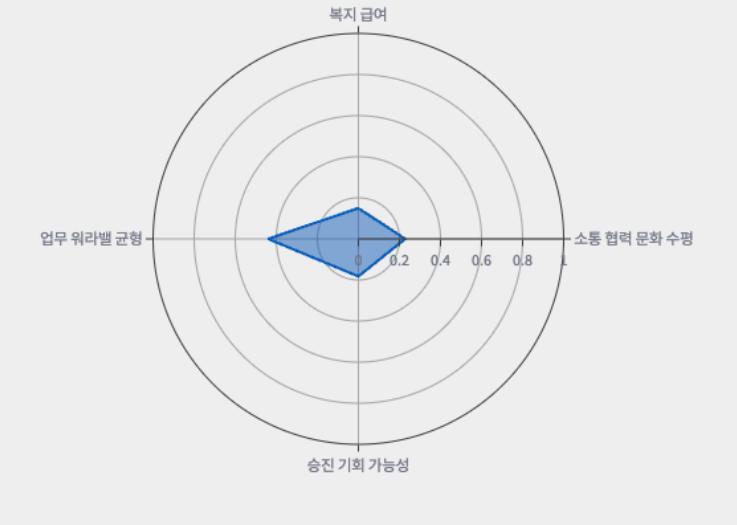
- 모델 성능 평가를 위한 10,000건의 테스트 데이터 구성
- zero-shot classification과 gpt-3 모델 fine-tuning
- 타 부서 업무 협업 웹앱 제공
(streamlit 사용하여 DL 모델 서빙)



2022년 한국농수산식품유통공사-장점 샘플 결과

sequence	class	labels	scores
0	연차 사용 자유롭고, 코로나 시 재택근무 사용 가능유연근무제 사용 가능	업무 위라밸 균형	0.5440
1	공사는 농수산식품 분야에서 인지도가 있는 기업으로 다양한 사업을 경험할 수 있으며, 해외 근	승진 기회 가능성	0.7742
2	무역, 마케팅 등 직무에 필요한 교육을 시켜준다. 사후 관리도 해준다.	소통 협력 문화 수평	0.4656
3	- 몇몇 혼다들을 제외하고 대체로 수평적인 문화 - 몇몇 부서들을 제외하고 대체로 지켜지는 유	업무 위라밸 균형	0.7619
4	적당적당 괜찮은 회사 식품쪽이어서 타 회사 대비 엄청 빠져거나 지루하거나 하지 않음	업무 위라밸 균형	0.8372
5	연차 사용 눈치 없음, 항복항공권, 숙박비, 체제비, 월급	복지 급여	0.6653
6	연차를 눈치보지 않고 쓸 수 있음., 성과급, 복지포인트 지급됩니다. 계약직까지 청규직과 동등	업무 위라밸 균형	0.4973
7	연차를 눈치 보지 않고 사용할 수 있습니다 전시를 구현할 수 있는 다양한 아이디어 발현 가능	승진 기회 가능성	0.4149
8	다른 사람들은 뻐른 칠판 및 편안한 분위기를 즐기는 중 그치만 나는 아님. 해외지사에 근무할 때는 소통 협력 문화 수평	소통 협력 문화 수평	0.6896
9	업무량은 많지 않으며, 사내 분위기 좋고 30분 단위로 연차 사용 가능함. 명절 선물 잘 나오고	업무 위라밸 균형	0.6064

2022년 한국농수산식품유통공사-장점 샘플 결과



2022년 한국농수산식품유통공사 각 레이블 평균 추론 스코어

샘플 텍스트 테스트 리뷰 데이터 테스트 연도별 트렌드 결과 비교 레이블 키워드 관련 빈출 어휘

? 자세히 보기 : 사용한 DL model - [mDeBERTa-v3-base-xnli-multilingual-nli-2mil7]

- 이 다국어 모델은 100개 언어에 대해 자연어 추론(NLI)을 수행할 수 있으므로 다국어 제로샷 분류에도 적합합니다. 기본 mDeBERTa-v3-base 모델은 100개 언어로 구성된 CC100 다국어 데이터 세트에서 Microsoft에 의해 사전 훈련되었습니다. 그런 다음 모델은 XNLI 데이터 세트와 다국어 NLI-26lang-2mil7 데이터 세트에서 fine-tune되었습니다. 두 데이터 세트 모두 40억 명이 넘는 사람들이 사용하는 27개 언어로 된 270만 개 이상의 가설-전제 쌍을 포함합니다.

✓ 분류하고자 하는 샘플 텍스트를 입력하세요.

연 베네핏카드 300만원과 휴양시설 이용비 지원 그리고 3년 재직자에게 1개월의 리프
레쉬 휴가 등 여러 복지가 잘 되어 있으며 신규 프로젝트 진행시 대부분의 경험상 바텀
업 형식으로 일정 조율이 되어 야근할 일이 거의 없음. 출석체육아휴직 매우 자유롭
고 눈치 안보이고 특히 자녀가 있을 경우 사내어린이집이 매우 유용함. 업무 진행시 동
료들 성격이 대부분 둥글둥글해서 사람 스트레스가 거의 없는 편이다

scores
복지 급여
업무 위라밸 균형
소통 협력 문화 수평
승진 기회 가능성

[AI 연구] NLP 분석

2022.12 ~ 2023.01

기업 평가 데이터에 대한 토픽모델링 분석과 차원축소를 통한 시각화 결과 등을 실시간으로 보여주는 웹앱

[개발 환경]

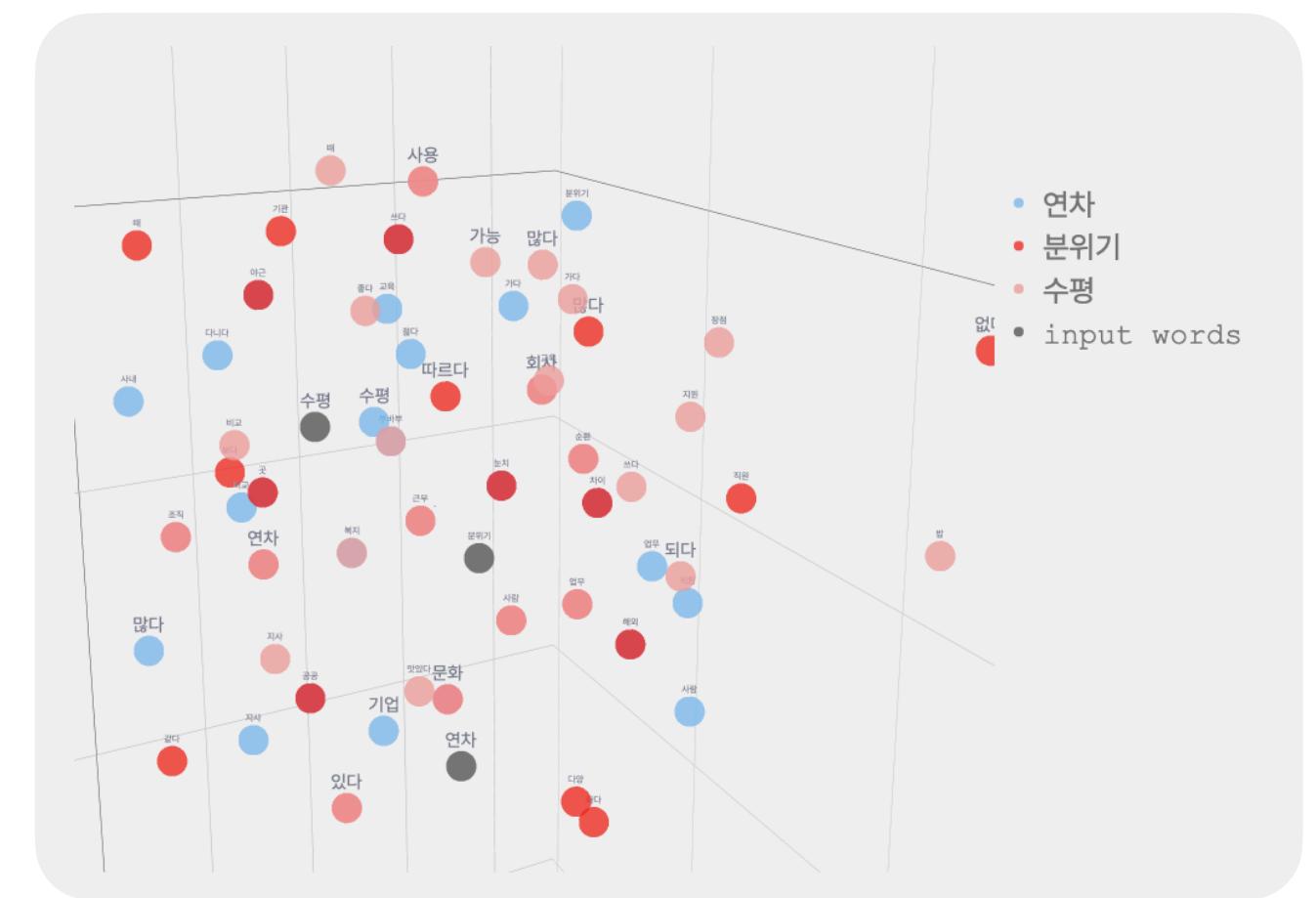
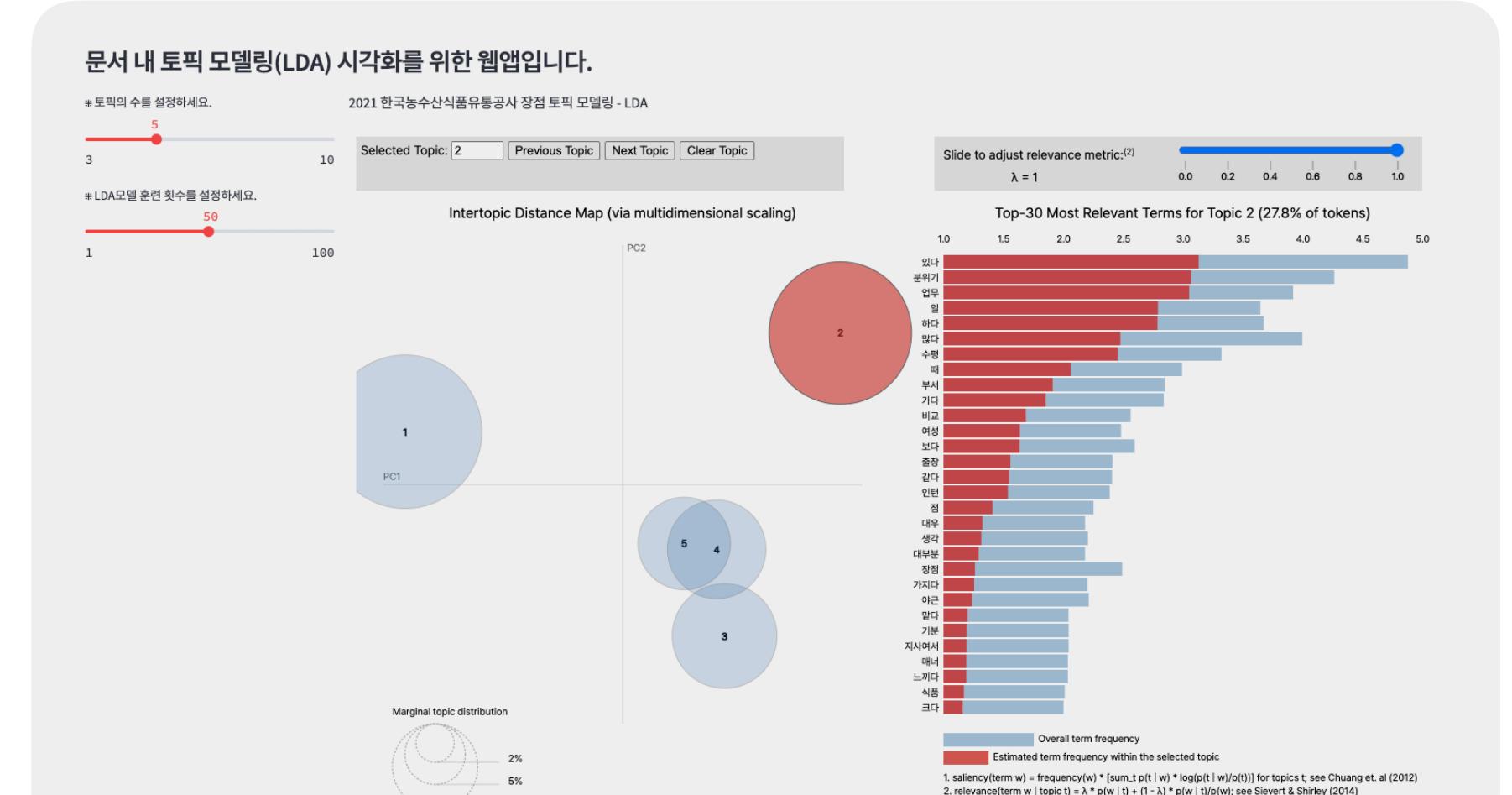
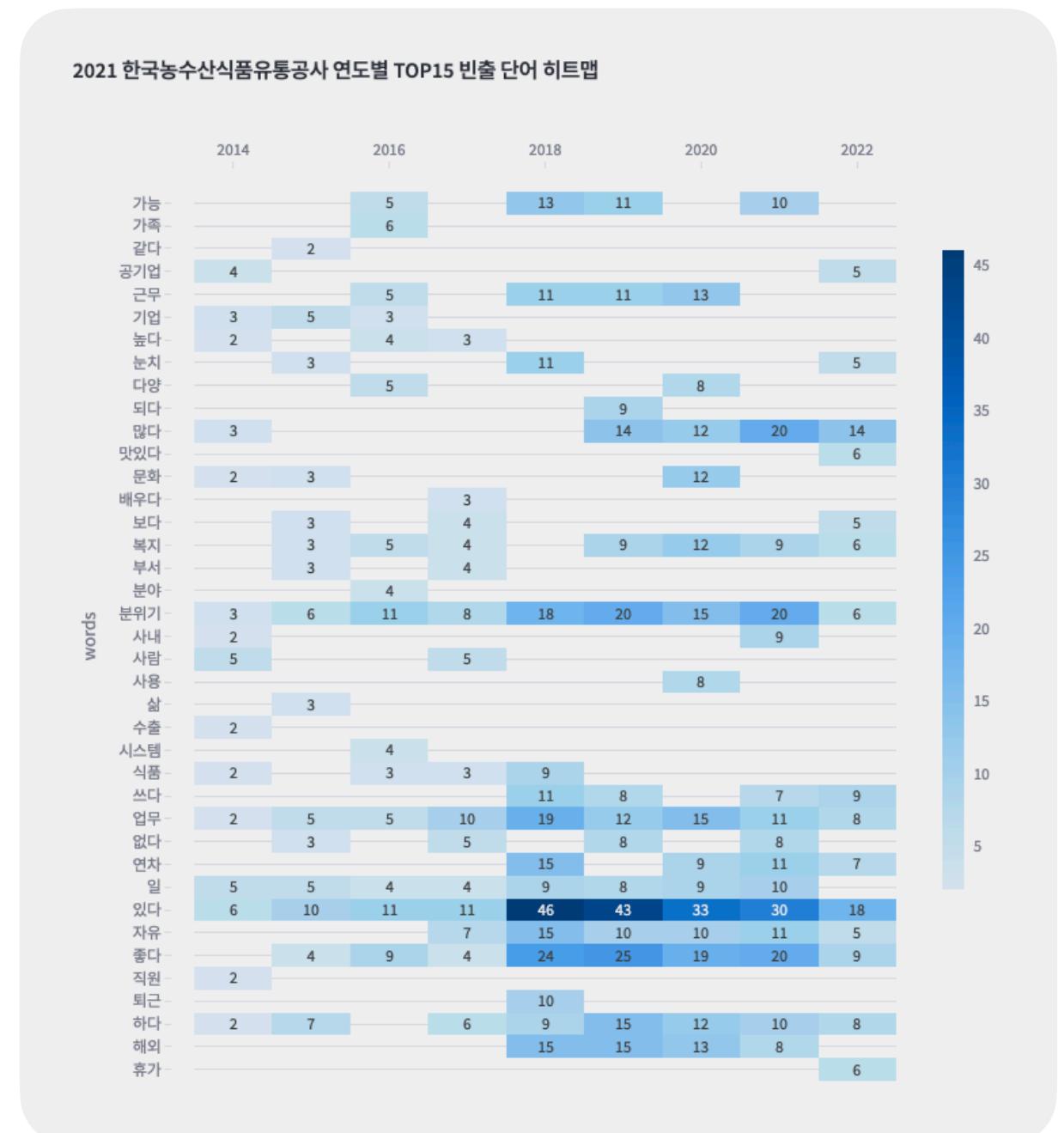
- ubuntu 22.04
- Python 3.9
- k8s, docker
- Git
- TeamCity

[사용 기술]

1. DB API
 - pymongo
2. 데이터 전처리 및 NLP
 - pandas, numpy
 - kiwipiepy
3. 토픽모델링 분석
 - gensim, pyLDAvis
4. 차원축소
 - scikit-learn, PCA, TSNE
5. 시각화
 - matplotlib, seaborn, plotly
6. 웹앱
 - streamlit

[주요 업무 및 성과]

- 문서 간 토픽모델링을 위한 LDA 분석 및 시각화
- 텍스트 간 유사 범주 어휘 분석을 위해 PCA/t-SNE 분석 적용 및 시각화
- Heatmap 시각화를 통한 연도별 분석
- 연도별 평점 트렌드 분석 시각화 제공



[팀 프로젝트] 리트리버: 웹스크래퍼 개발

2022.11 ~ 2023.01

python기반 웹스크래퍼, 데이터 ETL 파이프라인 및 REST API 개발 및 리딩

[개발 환경]

- ubuntu 22.04
- Python 3.9
- k8s, docker
- Git
- TeamCity

[사용 기술]

1. DB
 - sqlalchemy, psycopg2-binary, pymongo
 - PostgreSQL, MongoDB
2. 데이터 전처리
 - pandas, numpy
3. 데이터 수집
 - requests, beautifulsoup
4. REST API
 - fastAPI
5. 스케줄링
 - apscheduler

[주요 업무 및 성과]

- DDD(domain driven development) 설계
- ERD설계 및 DM 구축
- python기반 웹 스크래퍼 개발
- fastAPI를 사용하여 REST API 개발
- OAuth2.0 리소스 서버 구축
- k8s와 컨테이너 환경에서 배포 및 서비스
- teamcity를 통한 CI/CD

rtrv_host	
GET	/retriever/v1/hosts Get All Hosts
POST	/retriever/v1/hosts Create Host
GET	/retriever/v1/hosts/{host_id} Get Host By Id
PUT	/retriever/v1/hosts/{host_id} Update Host By Id
PUT	/retriever/v1/hosts/{host_id}/login Update Host Login

rtrv_path(sample page)	
GET	/retriever/v1/hosts/{host_id}/paths Get All Paths By Host Id
POST	/retriever/v1/hosts/{host_id}/paths Create Path
GET	/retriever/v1/paths Get All Paths
GET	/retriever/v1/paths/{path_id} Get Path By Id
PUT	/retriever/v1/hosts/{host_id}/paths/{path_id} Update Path

rtrv_path_rule	
GET	/retriever/v1/paths/{path_id}/rules/test Get Sample Page
POST	/retriever/v1/paths/{path_id}/rules/test-html Html Test Rule
POST	/retriever/v1/paths/{path_id}/rules/test-json Json Test Rule
POST	/retriever/v1/hosts/{host_id}/paths/{path_id}/rules/html Create Html Rule
POST	/retriever/v1/hosts/{host_id}/paths/{path_id}/rules/json Create Json Rule

GET	/retriever/v1/rules Get All Rules
GET	/retriever/v1/hosts/{host_id}/rules Get All Rules By Host Id
GET	/retriever/v1/paths/{path_id}/rules Get All Rules By Path Id
GET	/retriever/v1/rules/{rule_id} Get Rule By Id
PUT	/retriever/v1/rules/html/{rule_id} Update Html Rule
PUT	/retriever/v1/rules/json/{rule_id} Update Json Rule

[B2C] AI 추천 시스템

2022.11 ~ 2023.01

“취업자 개인 맞춤형 플랫폼, GRABBER” 의 “AI진단 서비스” 추천시스템 알고리즘 개발

[개발 환경]

- ubuntu 22.04
 - Python 3.7
 - docker
 - GitLab

[사용 기술]

1. DB API

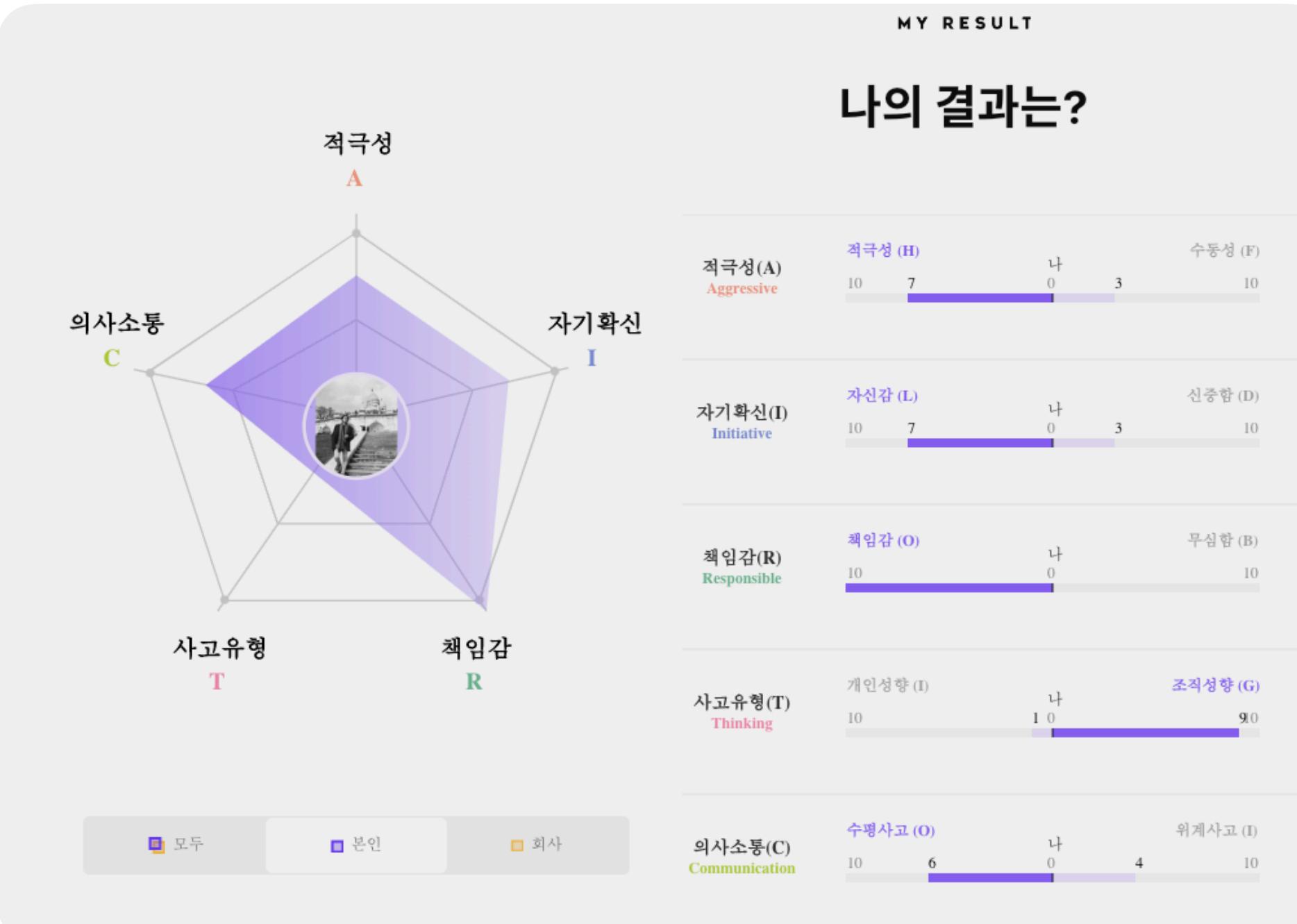
- ## - pymongo

2. 형태소 분석 & NLP

- pandas, numpy
 - mecab, kiwi
 - gensim, text2distance
 - scikit-learn

주요 업무 및 성과

- 98만건 텍스트 데이터 형태소 분석 및 자연어처리
 - 문서/토큰 간 유사도 매트릭스 생성
 - 유사 기업 클러스터링 엔진 연구 개발 및 테스트
 - 구직자-기업 간 매칭률 산출 알고리즘 개발
 - 1000명의 베타 테스터를 통한 B2C 플랫폼, GRABBER "AI진단" 서비스 론칭
 - “기계학습을 이용한 구직자-구인자 컬쳐핏 매칭 방법”에 대한 발명자로서 특허출원 (제10-2022-0109802호)



나의 관심정보

나의 관심직종	IT, 개발	나의 연봉 및 복지 성향
나의 관심기업	코드스테이츠	연봉에 대한 니즈
나의 관심직무	개발자	높음
	AI/ML	7/10
	데이터분석가	낮음
		3/10

나와 적합한 기업 유형

대기업
스타트업
글로벌기업
공기업

*기업유형 설명
어떤 타입의 기업을 선호하는지를 측정한 결과입니다.
크게 대기업(그룹사) 유형, 공기업(공공기관) 유형, 스타트업 유형, 글로벌기업 유형이 있습니다.

나를 위한 기업 추천		
아토세이프	아이도트	마케팅지음랩
기업정보	기업정보	기업정보
기업형태 일반,중소기업	기업형태 일반,중소기업	기업형태 일반,중소기업
대표자 조창희/최종국	대표자 정재훈	대표자 김지연
설립일자 2010.12.02	설립일자 2014.06.05	설립일자 2012.06.18
상장여부 비상장	상장여부 비상장	상장여부 비상장
주소 서울 금천구 가산디지털*로 ***, ***호	주소 서울 송파구 법원로 ***, **층 씨****호, 씨****호	주소 서울 서초구 효령로 **, ***호
산업명 (C20422)치약, 비누 및 기타 세제 제조업	산업명 (J58222)응용 소프트웨어 개발 및 공급	산업명 (M71310)광고 대행업
연봉정보	연봉정보	연봉정보
올해 입사자 평균연봉 3,000만원 이상 ~ 5,000만원 미만	올해 입사자 평균연봉 3,000만원 이상 ~ 5,000만원 미만	올해 입사자 평균연봉 ~ 3,000만원 미만
종업원 수 1명 이상 ~ 20명 미만	종업원 수 20명 이상 ~ 40명 미만	종업원 수 1명 이상 ~ 20명 미만
<input type="button" value="전체 숨김"/>	<input type="button" value="전체 숨김"/>	<input type="button" value="전체 숨김"/>

[외주 프로젝트] 서울시 IoT 도시데이터 분석

2022.08 ~ 2022.09

서울시 금천구/서대문구 장애인 주차구역 IoT 데이터 분석 및 시각화

[분석 환경]

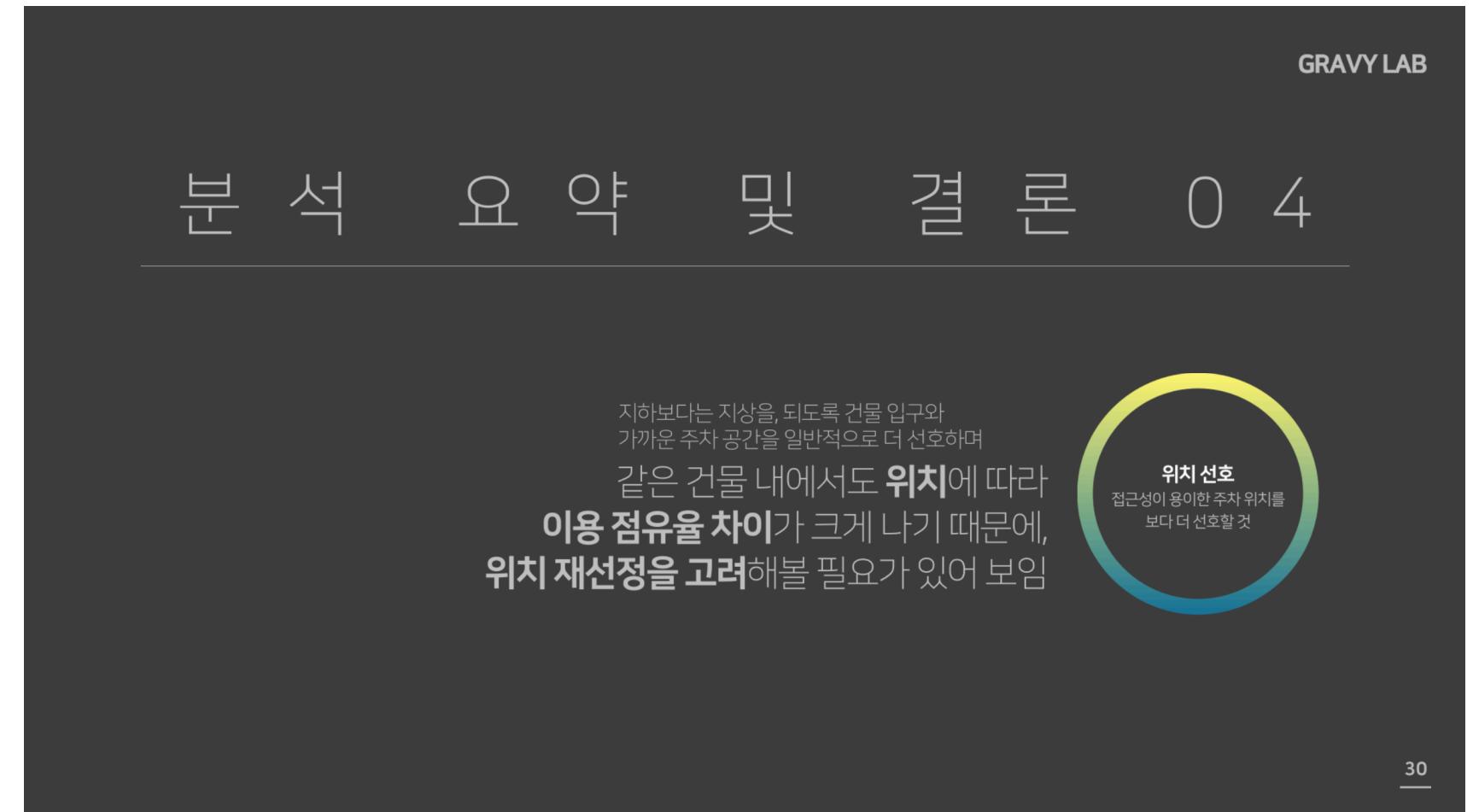
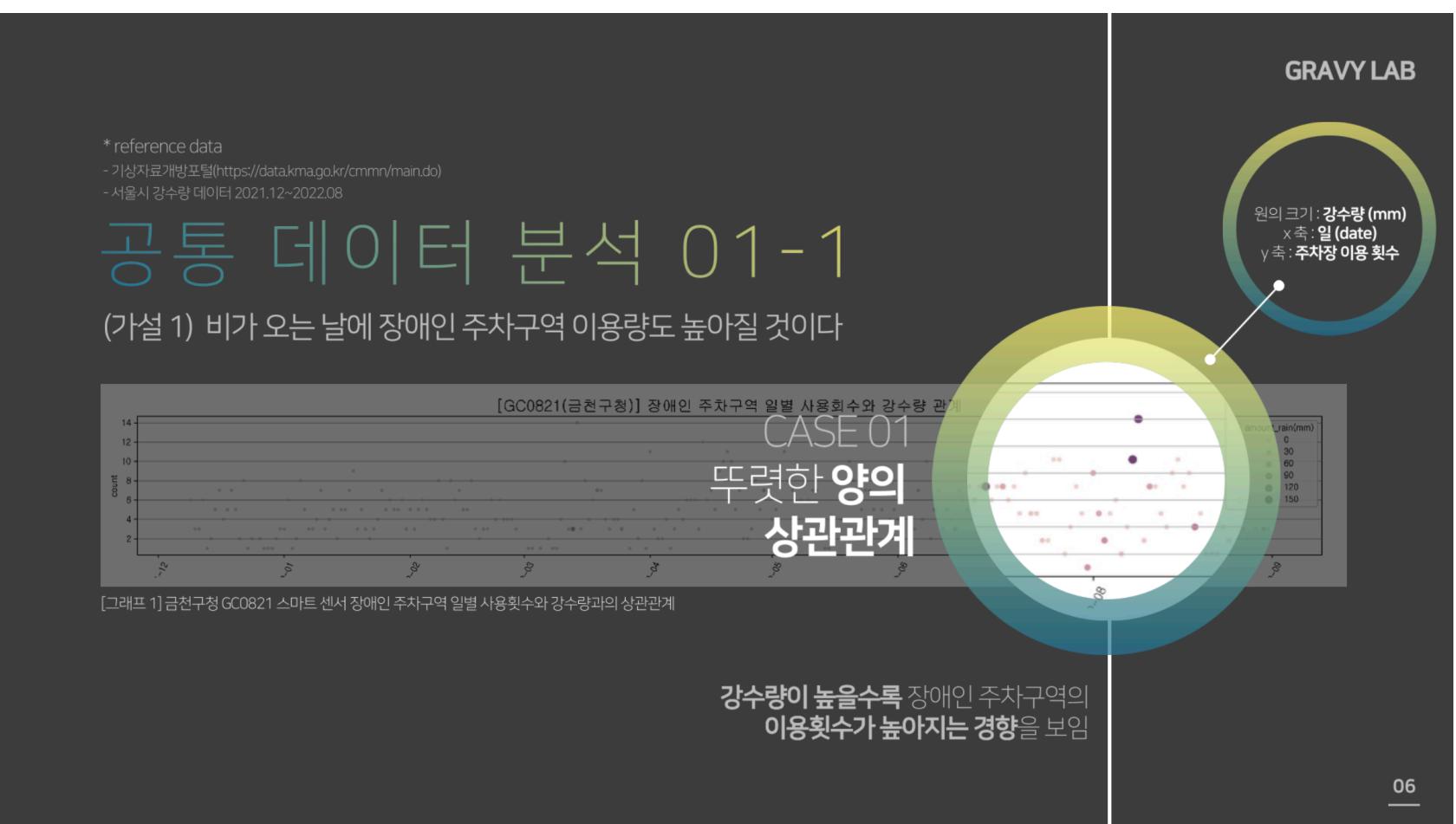
- Python 3.7
- JupyterLab

[사용 기술]

- 데이터 전처리
 - pandas, numpy
- 데이터 시각화
 - matplotlib, seaborn, plotly

[주요 업무 및 성과]

- 시계열 데이터 분석 및 정제
- 시계열 데이터 분석 시각화 리포트 작성



[기업 협업] 사용자 추천시스템

2022.02 ~ 2022.03

아카이드로우 기업 협업 프로젝트 | 텍스트 유사도 알고리즘을 사용하여 개인화 가구 추천시스템 구현

[개발 환경]

- Python 3.8
- Jupyter
- Github

[사용 기술]

1. 데이터 전처리

- pandas, numpy
- mecab

2. 추천/검색 알고리즘

- gensim
- text2distance

3. 데이터 시각화

- matplotlib, seaborn, plotly

[주요 업무 및 성과]

- 가구 스타일 정보 데이터 형태소 분석 및 전처리
- 문서/토큰 간 유사도 매트릭스 생성
- 콘텐츠 기반 추천/검색 모델 알고리즘 개발
- 모델 최적화 및 검색 필터 고도화

```
검색 시간 : 0.7665초   검색 알고리즘 : sorensen

!!사용자가 이미 선택한 아이템, 같은 카테고리 아이템들은 추천목록에서 제거!!
사용자가 선택한 아이템 : 멜 쿠션 고급 인테리어 에펠의자 7colors (화이트)
사용자가 이미 선택한 아이템 목록
멜 쿠션 고급 인테리어 에펠의자 7colors (화이트)

=====
추천 아이템 Top10
=====

추천 1순위 : 플레인 철제 드레스룸 시스템 옷장 행거 코너장 2colors (화이트) - 드레스룸/행거
추천 2순위 : 세탁 장스탠드 - 조명
추천 3순위 : 컬러스토리 와이드 서랍장 3단 (화이트) - 수납가구
추천 4순위 : 플레인 철제 드레스룸 시스템 옷장 행거 400 5단 선반장 2colors (화이트) - 드레스룸/행거
추천 5순위 : 유아 의자 화이트 - 유아동가구
추천 6순위 : 모던보이스퀘어 단스탠드 (화이트) - 조명
추천 7순위 : 3단 분리수거함 - 생활/운동
추천 8순위 : 피카 얼라이트 장스탠드 조명 - 조명
추천 9순위 : 골드니스 확장형 수납 화장대 세트 (화이트) - 화장대/거울
추천 10순위 : 인테리어 타공판 1000*600 (화이트) - 홈데코
선택한 제품명을 입력하세요.
(랜덤 선택은 입력 없이 enter, 프로그램 종료는 x)x
추천 시스템을 종료합니다.
감사합니다!
```

Top 5 해당하는 비율

- Top 1: 선택한 프로젝트 아이템이 들어있는 비율 : 0.571 다른 프로젝트 아이템이 들어있는 비율 : 0.429
- Top 2: 선택한 프로젝트 아이템이 들어있는 비율 : 0.357 다른 프로젝트 아이템이 들어있는 비율 : 0.643
- Top 3: 선택한 프로젝트 아이템이 들어있는 비율 : 0.286 다른 프로젝트 아이템이 들어있는 비율 : 0.714
- Top 4: 선택한 프로젝트 아이템이 들어있는 비율 : 0.214 다른 프로젝트 아이템이 들어있는 비율 : 0.786
- Top 5: 선택한 프로젝트 아이템이 들어있는 비율 : 0.214 다른 프로젝트 아이템이 들어있는 비율 : 0.786

전문가 셋 데이터 비교평가

precision recall f1-score support

0	0.71	0.83	0.77	6
1	0.86	0.75	0.80	8

accuracy	0.79	14
macro avg	0.79	0.79
weighted avg	0.80	0.79

Average_Precision : 0.7857142857142856

[ML 개인프로젝트] 집값 예측 웹앱 개발

2021.11 ~ 2022.12

광명시 아파트 값 예측을 위한 ML 서비스를 웹앱으로 개발

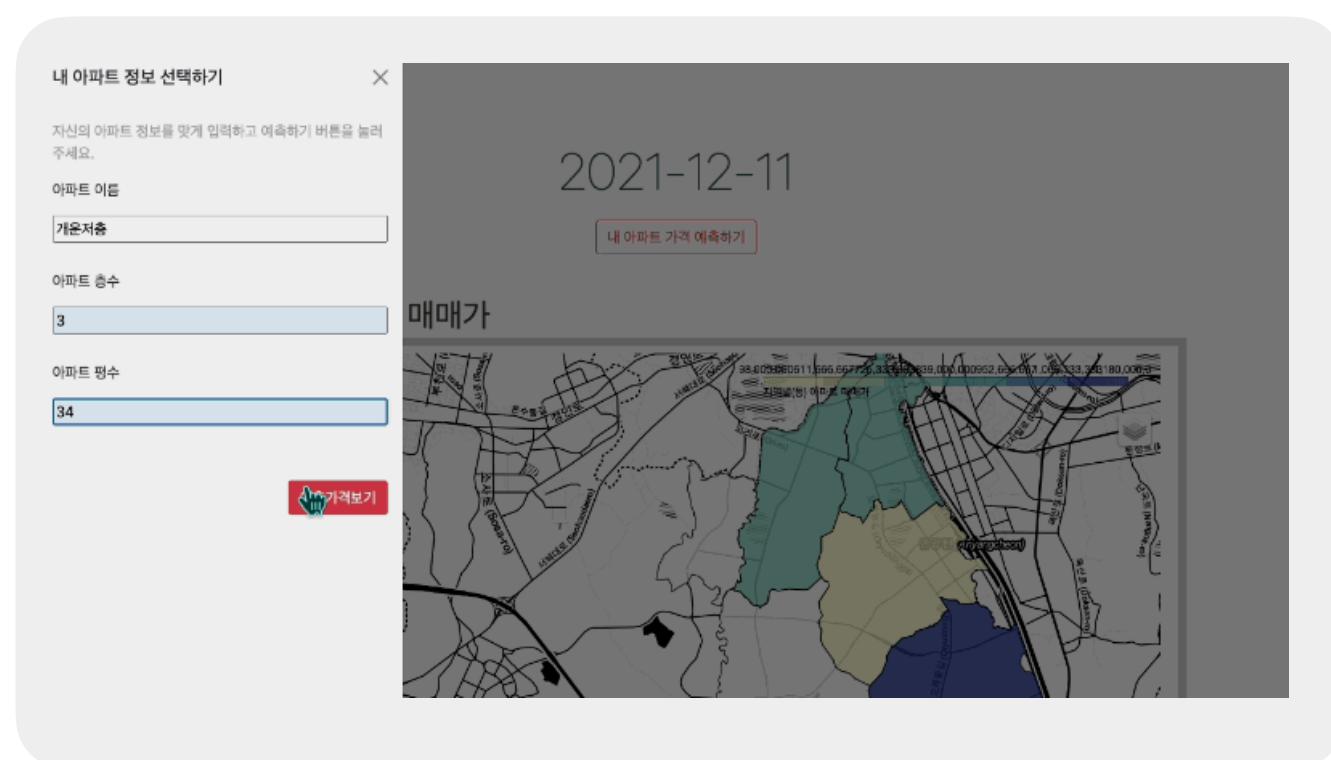
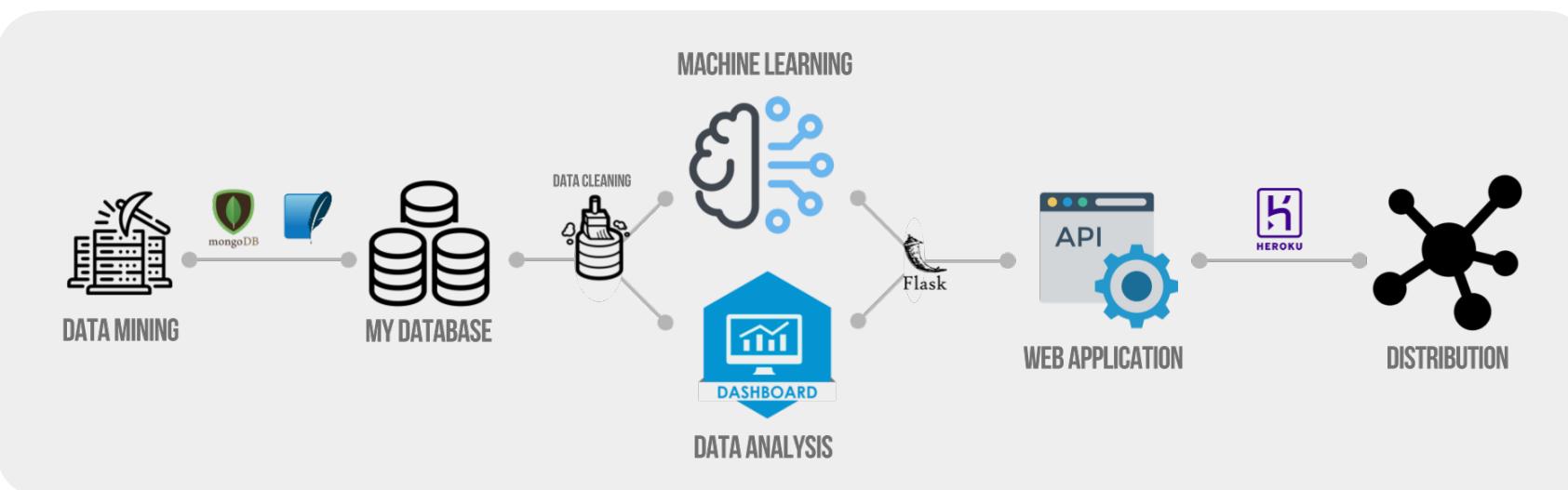
시스템 구성도

[개발 환경]

- Python 3.8

[사용 기술]

1. DB
 - sqlite, MongoDB, sqlalchemy
2. 데이터 수집
 - requests
3. 데이터 전처리
 - pandas, numpy
4. ML model
 - scikit-learn, xgboost, lightgbm
5. 데이터 시각화 및 대시보드
 - matplotlib, plotly, folium
 - MetaBase
6. 웹 프레임워크 및 배포
 - Flask
 - Heroku

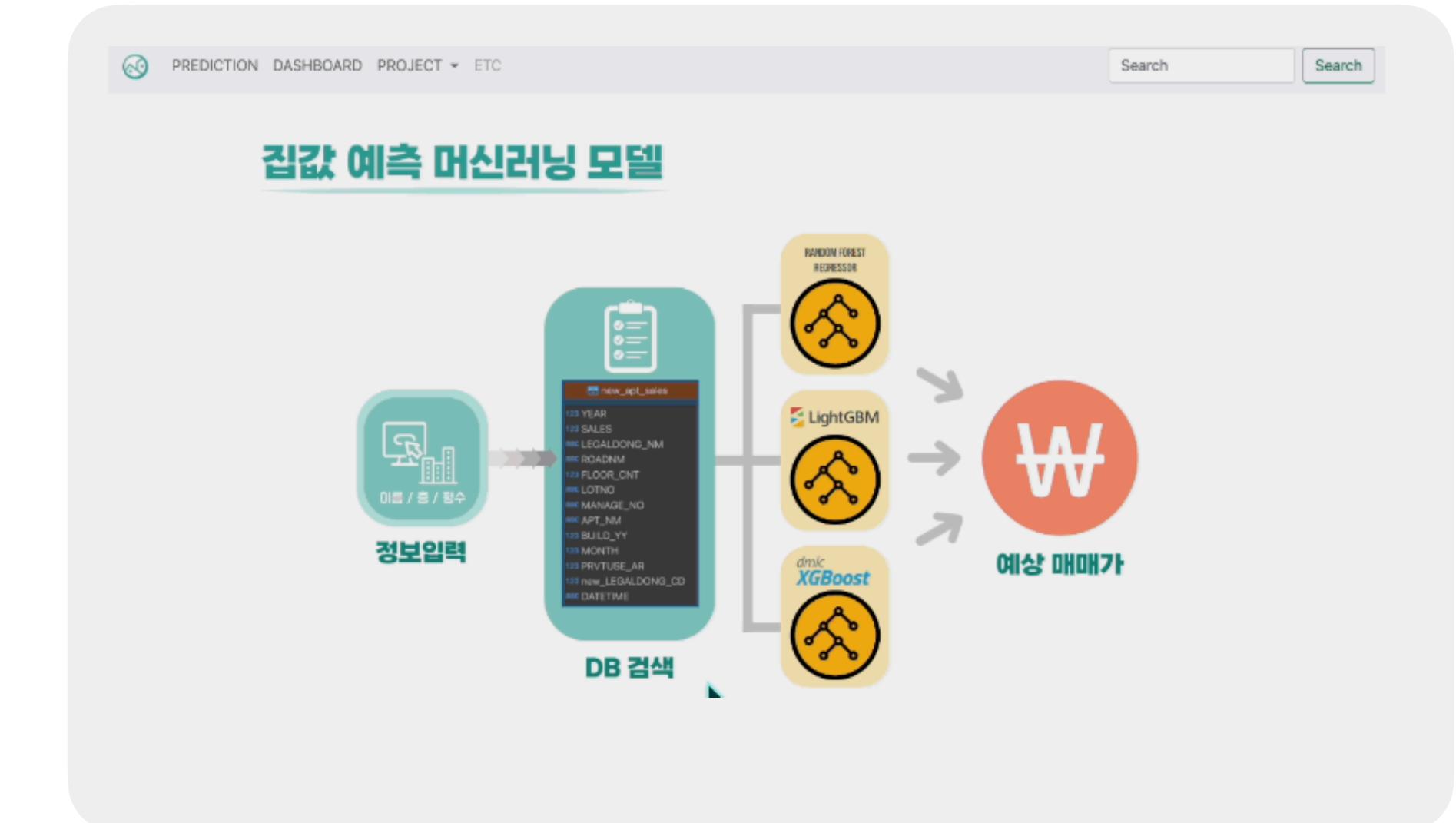


[주요 업무 및 성과]

- 광명시 공공 데이터 open api를 통해 RDB 형태로 수집
- 데이터 전처리
- 양상별 모델 최적화 및 모델 서빙
- 데이터 시각화 및 대시보드 연동
- 웹앱 개발 및 배포

Evaluation

MODELS	MAE	RMSE	R2_SCORE
LinearRegression	174718437.51 KRW	238319972.99 KRW	0.08
ElasticNetCV	274784163.17 KRW	360266710.18 KRW	-1.1
RandomForestRegressor	63981882.48 KRW	113676925.82 KRW	0.9
XGB Regressor	149348580.28 KRW	180378806.28 KRW	0.47
LGBM Regressor	154223908.87 KRW	186687372.47 KRW	0.44



[NLP 개인프로젝트] Bert 악성댓글 탐지기

2022.03 ~ 2022.03

Bert기반 악성 댓글, 성차별/혐오 댓글 탐지 모델

[개발 환경]

- Python 3.8
- Github

[사용 기술]

- 데이터 전처리
 - pandas, numpy
 - mecab
- DL model
 - transformers
 - tensorflow
 - scikit-learn

[주요 업무 및 성과]

- Koco Korean-hate-speech 데이터 전처리
- 형태소 분석기를 통해 텍스트 데이터 분석 및 NLP 수행
- fine-tuning 및 pre-trained 모델 최적화

performance & evaluation

• 악성 댓글 탐지 모델 평가 비교

	Bert-base-multilingual	kyKim/Bert-kor
Accuracy	0.73	0.78
F1-Score	0.80	0.82
Recall	0.79	0.75
Precision	0.80	0.90

• 성차별/혐오 댓글 탐지 모델 평가 비교

	Bert-base-multilingual	kyKim/Bert-kor
Accuracy	0.92	0.94
F1-Score	0.69	0.75
Recall	0.61	0.66
Precision	0.80	0.88

• 악성 댓글 탐지 with Bert-base-multilingual

```
sentence = "푸틴한테는 한마디도 못하는 역겨운 거지나라 합정공주년"  
print(detect.predict(sentence))  
# 1 (악성댓글)
```

```
sentence_ = "강아지와 개새끼의 차이가 뭔가요"  
print(detect.predict(sentence_))  
# 0 (악성댓글 X)
```

