

한기호 HAN KIH0

mail : eqfq1@naver.com

github: https://github.com/kihohan/BigData_Project

Educations

- **한국경제신문**
 - 청년취업아카데미 빅데이터 소셜마케팅 전문인력양성과정 2018.12 ~ 2019.02
- **상명대학교**
 - 주전공: 게임학과, 연계전공: 빅데이터학과 2013.03 ~ 2019.08

Works

- **(주) 이팝콘 데이터사이언스팀 (2019.07 ~ 2020.11)**
 - Regex, 클라이언트에 특화된 Tokenizer 개발, Rule-Based, Deeplearning 등을 이용하여 카테고리, 상품을 분류하는 업무를 진행하였습니다. 또한 카테고리, 제조사, 브랜드, 서브브랜드, 라인 5개 체계를 만들어 사내 마스터링 체계를 정립하였습니다.
 - 이커머스 데이터의 판매량을 점검하고 이슈가 있을시 수집팀과 커뮤니케이션하여 데이터 수집 간격을 변경하거나, Parameter를 수정하는 트러블 슈팅을 진행하였습니다.
 - 오픈소스 Airflow를 이용하여 프로세스를 자동화 하였고, Redash를 통해 데이터를 시각화하여 기획팀과 커뮤니케이션 할 수 있는 시스템을 구축하였습니다.
 - Sql과 Nosql를 이용하여 Client에 맞는 데이터를 Migration하는 업무를 수행하였습니다.

Projects

- 자연어 처리를 이용한 상품 마스터링 프로젝트 (2019.11 ~ 2019.01)
- 소셜 빅데이터를 통한 삼성 에어드레서 마케팅 방안 (2019.01 ~ 2019.02)
- 해파리 출몰 예측모델 개발 (2018.04 ~ 2018.05)
- 아주대 CDM데이터를 활용한 COPD 악화 예측모델 개발 (2018.09)
- 산업단지 화재 발생 시 피해 최소화를 위한 소방 관제시스템 모델 구축 (2018.10 ~ 2018.11)
- 빅데이터 분석을 통한 GS 편의점 발주량 추천 시스템 (2019.06)
- 김해시 화재 발생 예측 모델 개발 (2019.12)
- 공공데이터를 이용한 전국 유치원 실태 파악 및 개선 방안 제시 (2018.05)

Skills

- Programming Language : Python, R, Sql, Nosql
- Framework: Tensorflow, Keras, Scikit-learn

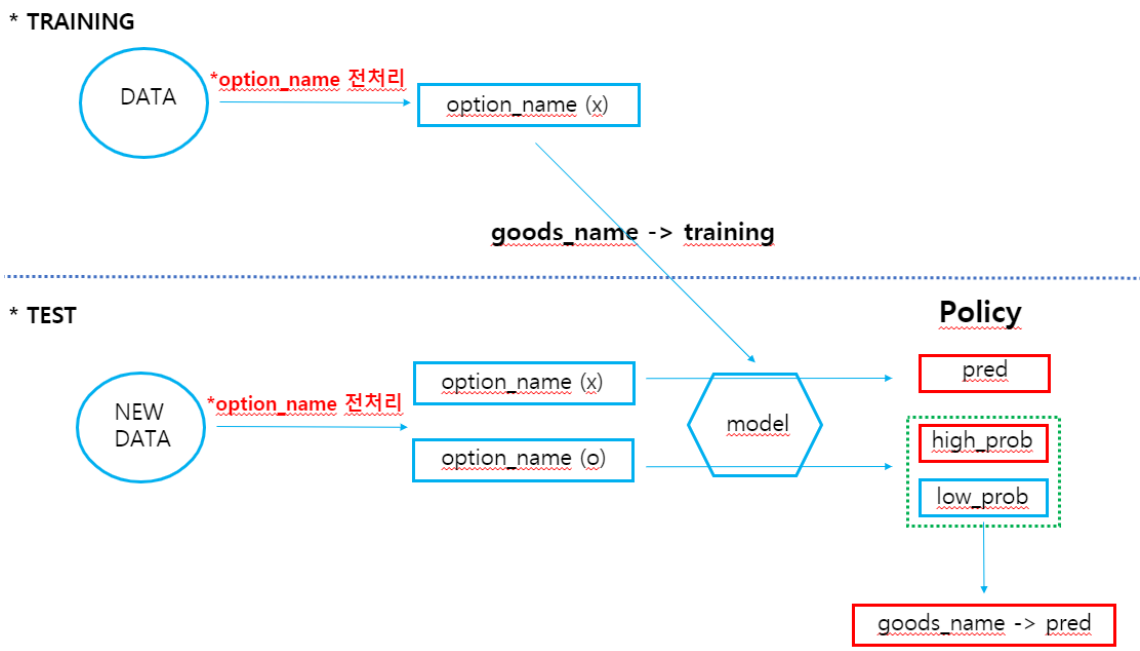
Awards

- 빅데이터 소셜마케팅 분석보고서 공모전 – 대상 (한국경제신문, 한국정보화진흥원)
- 산업단지 빅데이터 아이디어 공모전 - 장려상 (한국산업단지공단)
- 제5회 해양수산 공공, 빅데이터 활용 경진대회 – 우수상 (해양수산부)

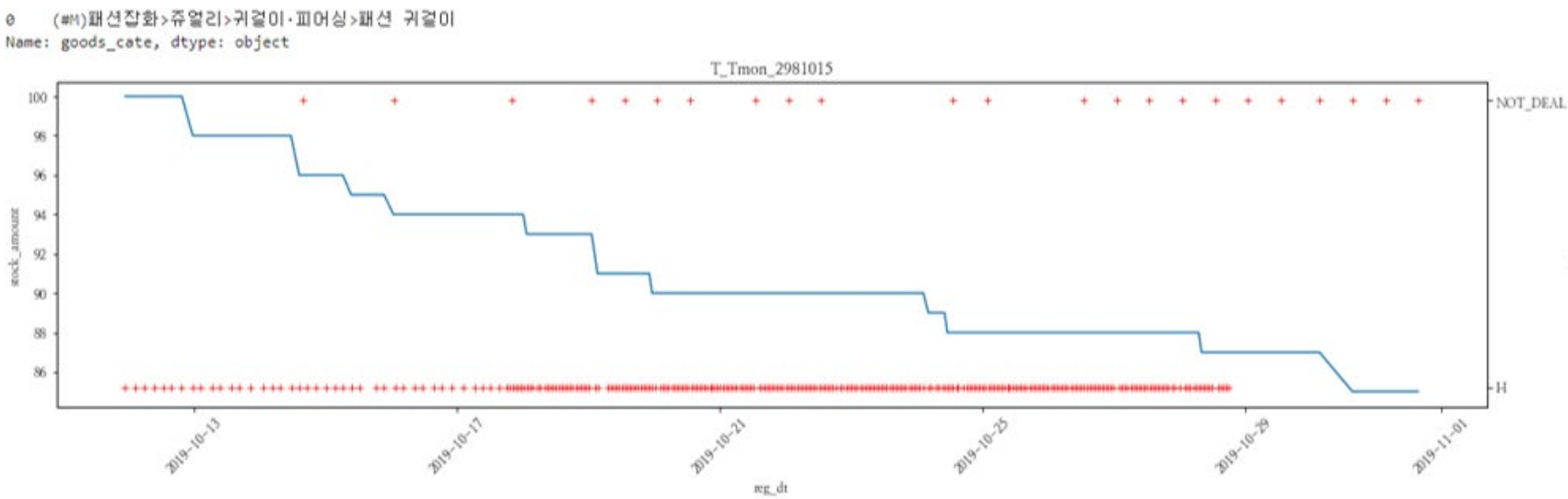
Works

(주) 이팝콘 데이터사이언스팀 (2019.07 ~ 2020.11)

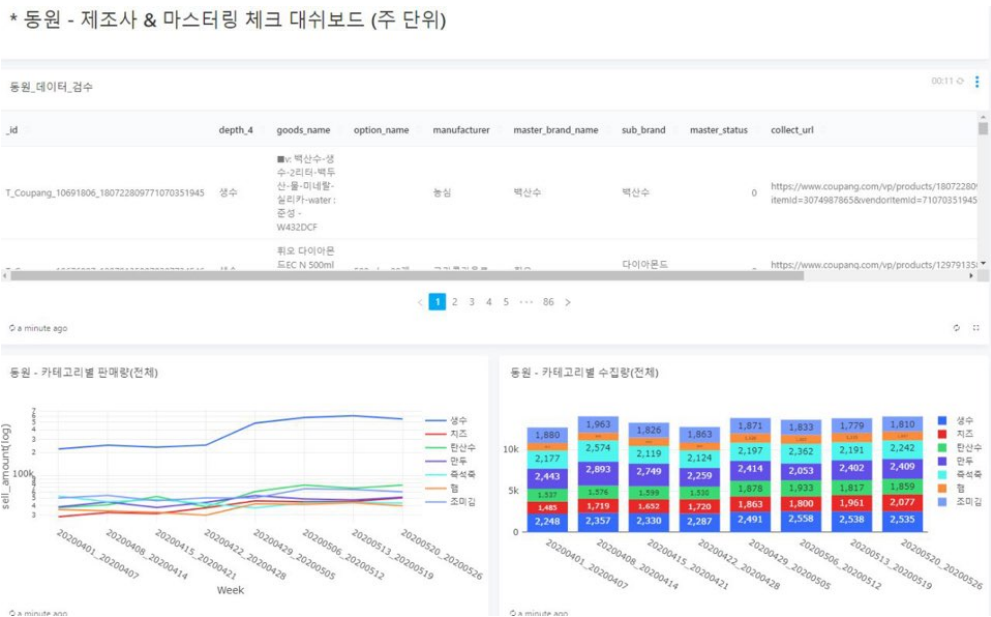
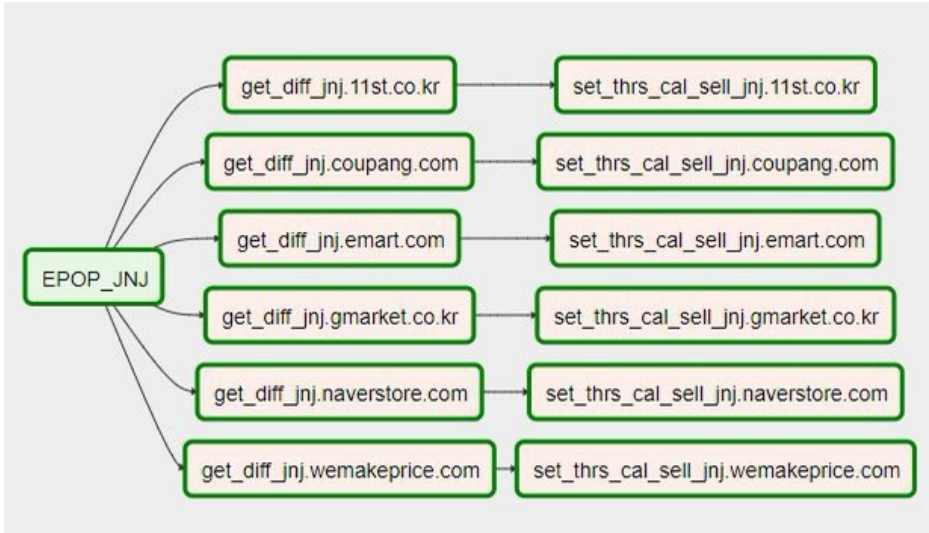
1. Text, image 데이터를 이용하여 카테고리 및 상품을 분류하는 방안을 기획, 개발, 운영을 담당하였습니다.



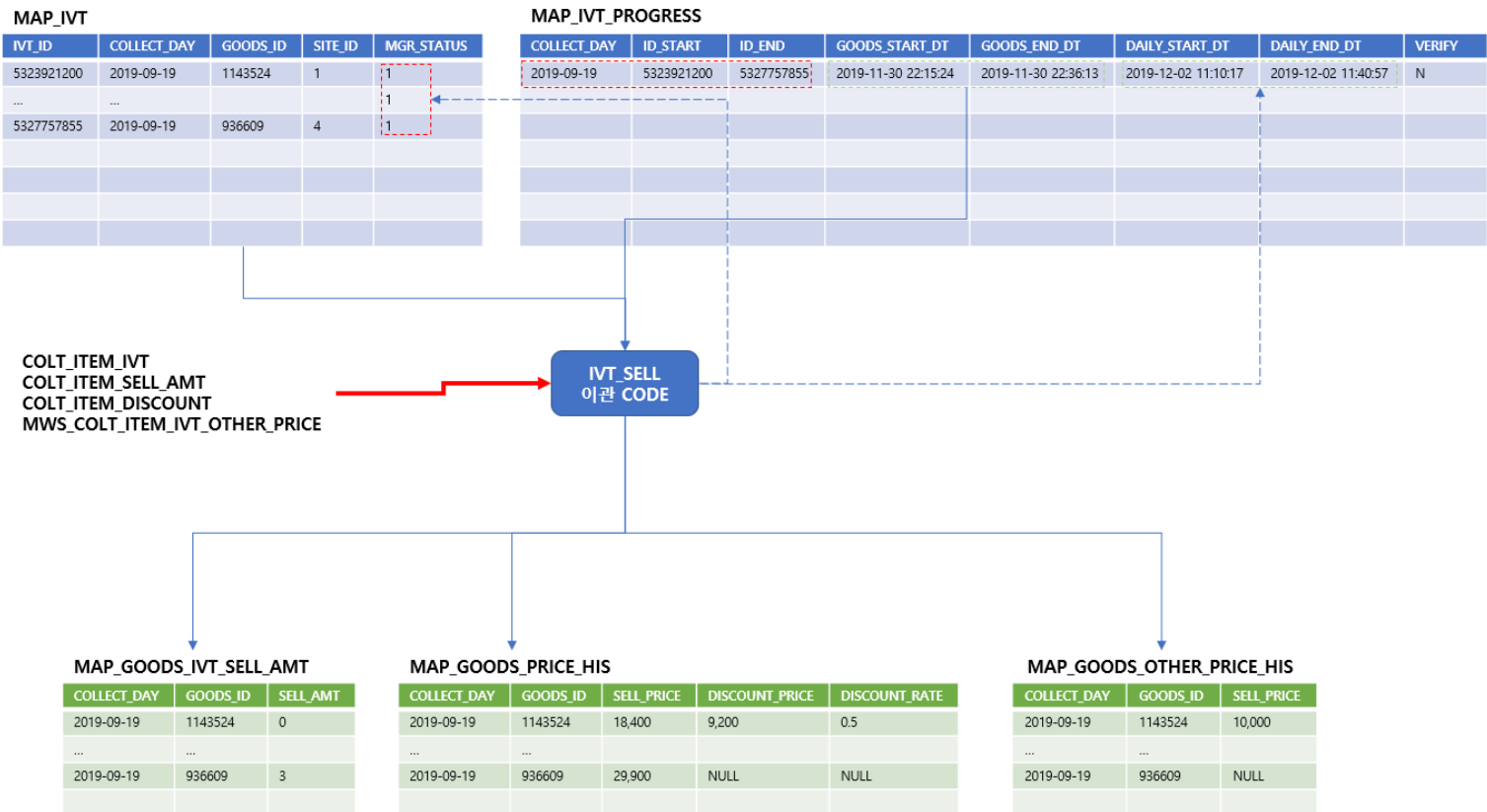
2. 재고 데이터를 이용하여 판매량 데이터를 계산하는 모듈을 관리하였습니다.



3. Airflow를 이용하여 프로세스를 자동화하고, Redash를 이용하여 기획팀과 커뮤니케이션 할 수 있는 시스템을 구축하였습니다.

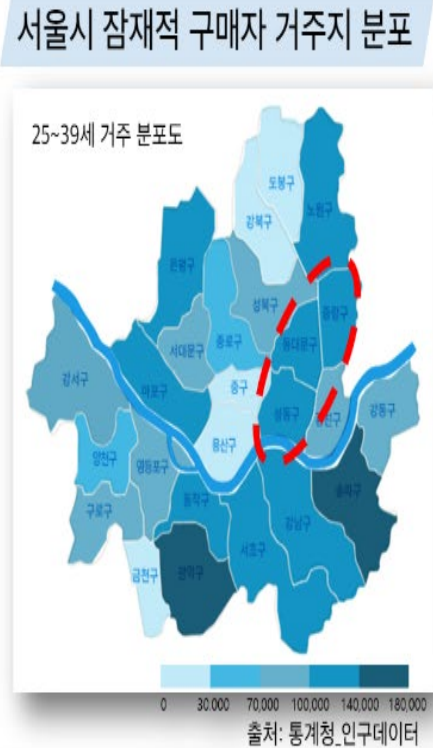


4. Raw_DB에서 데이터를 Migration하는 업무를 진행하였습니다.





설화수 진설 팩트 리필 단일 색상 은은한색 설화수



SAMSUNG

Projects

해파리 출몰 예측모델 개발 (2018.04 ~ 2018.05)

Technical Skills: Python, Scikit-learn

Goal: 기존 해상 기상정보 기구인 스마트부이에 해파리 출몰 예측모델 기능 추가 개발

Result: 해파리 종류별로 출몰에 영향을 주는 Feature를 도출하고, Logistic 알고리즘을 이용하여 해파리 출몰 예측 모델 개발

- 해파리 출몰 정보 (Open API)

JellyFishName	newsPerFreq	newsPerSta	newsSecs	OccurrenceSpot
노무라입깃해파리	20161117	20161111	2016112200	○서해 출현 서남도 출현 해역 - 충남 태안 각 읍면출도 연안
보름달물해파리	20161117	20161111	2016112200	○남해 출현 일부해역 앞진 출현 경남 일부해역 앞진 출현 - 경남 창원 앞도 연안 저밀도 출현 해역 - 경남 거제 앞도 연안 - 경남 사량도, 고성 연안 부산 광안리 연안
노무라입깃해파리	20161110	20161104	2016111000	○동/서/남해 출현 저밀도 출현 해역 인천 수월면 앞도 연안 - 전남 소라도 연안 - 경북 영일만 연안

출몰지(OccurrenceSpot)를 각각 한 지역으로 나누어 구글맵(Google Map)을 활용하여 각각의 지역의 위도와 경도를 저장하여 데이터 셋 구축.

JellyFishName	newsPerSta	newsPerEnddensity	spot	longitude	latitude
노무라입깃해파리	20141003	20141010	소량	인천 소정도	124.7141 37.77241
노무라입깃해파리	20141003	20141010	소량	인천 연평도	125.6549 37.6595
노무라입깃해파리	20141003	20141010	소량	충남 격렬비열도	125.5262 36.62027

- 추가/수정된 열
출몰지(spot): 기존 출몰지를 한 셀로 나누어 표시
위도/경도(longitude/latitude): 출몰지의 위도와 경도

해양성분과 노무라입깃해파리의 로지스틱 회귀분석의 ROC 곡선

- 암모니아성질소(NH3N)과 투명도(transparency)
노무라입깃해파리의 출몰 여부를 올바르게 구분해낼 확률이 옳바르지 못하게 구분해낼 확률보다 크다는 점을 확인할 수 있다.
- 아질산성질소(NH2N)와 용존무기질소(DIN)
노무라입깃해파리의 출몰을 옳바르게 예측하지 못한다는 것을 할 수 있다.

즉, 위의 분석 모델에서 NH3N과 transparency는 노무라입깃 해파리를 출몰 여부를 파악하는데 유의미한 변수이다.

- 어장 환경 모니터링

해역명	연도	장소명	조사일자	조사시간	수온(℃)	염분(ppt)	pH	용존산소(mg/L)	용존산소(%)								
서해충무	가부동	가	2016	8	11	126.2111	36.8905	14.17	14.15	32.03	7.89	7.97	9.70	9.9	1.52	1.29	
영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해	영남해

기존 어장 환경 모니터링과 수정된 해파리 출몰 정보를 합쳐 각 기간에 해파리가 출몰했다면 0(출몰하지 않음), 1(출몰함)으로 변환하여 데이터 셋 구축.

...

용존산소(mg/L)	용존산소(%)	규산규소(mg/L)	규산규소(%)	부유물(mg/L)	부유물(%)	클로로필(mg/L)	클로로필(%)	노무라입깃해파리	보름달물해파리					
0.013	0.014	0.028	0.031	0.079	0.092	28.2	20.2	3.23	1.55	1.5	0	0	1	1

- 추가/수정된 열
노무라입깃해파리: 출몰하지 않았다면 0, 출몰했다면 1.
보름달물해파리: 출몰하지 않았다면 0, 출몰했다면 1.
기타) 해파리는 주로 표층에 분포하므로 저층은 활용하지 않음

해양성분과 보름달물해파리의 로지스틱 회귀분석의 ROC 곡선

- 용존산소량(DO), 용존무기인(DIP), 규산규소(SiO2)
학습된 모델에서 DO, DIP, SiO2가 보름달물해파리 출몰에 영향을 준다는 결과가 나왔지만 옆의 그래프에서는 모델의 성능이 좋지 않다는 결과를 얻을 수 있었다.

분석 모델의 만들기 위해 더 많은 양의 데이터가 존재한다면 의미있는 결과를 도출할 수 있을 것으로 예상된다.

Projects

아주대 CDM데이터를 활용한 COPD 예측모델 제시 (2018.09)

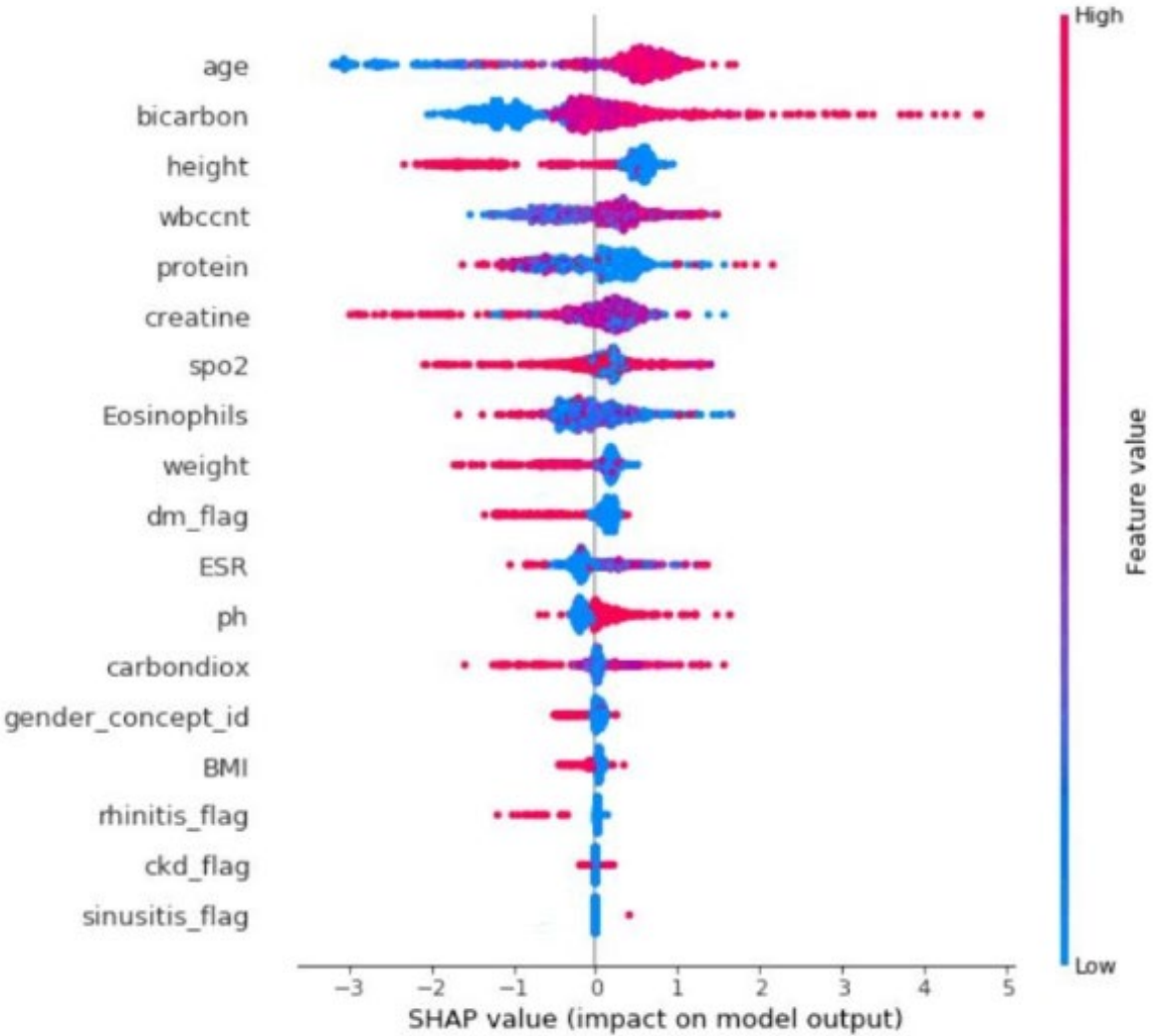
Technical Skills: BigQuery, Python, Scikit-learn

Goal: 아주대학교 CDM데이터를 활용하여 COPD 예측모델 제시

Result: 2박 3일간, COPD 악화 인자 중 Smoking, Age, Bicarbon, Protein 등이 리스크 요인이라는 것을 데이터를 통해 증명하고, 예측모델을 개발

FEATURE SELECTION in aggravation of COPD

Demographic factor	Age	Lab	WBC
	sex		Eosinophil
	BMI		Cr
	weight		ESR
	height		CRP
COMORBIDITY	DM	ABGA	lactate
	CKD		Ig E
	Acute sinusitis		HbA1c
	Allergic rhinitis		pH
PFT	FEV1		SpO2
	FEV1/FVC		HCO3
			pCO2



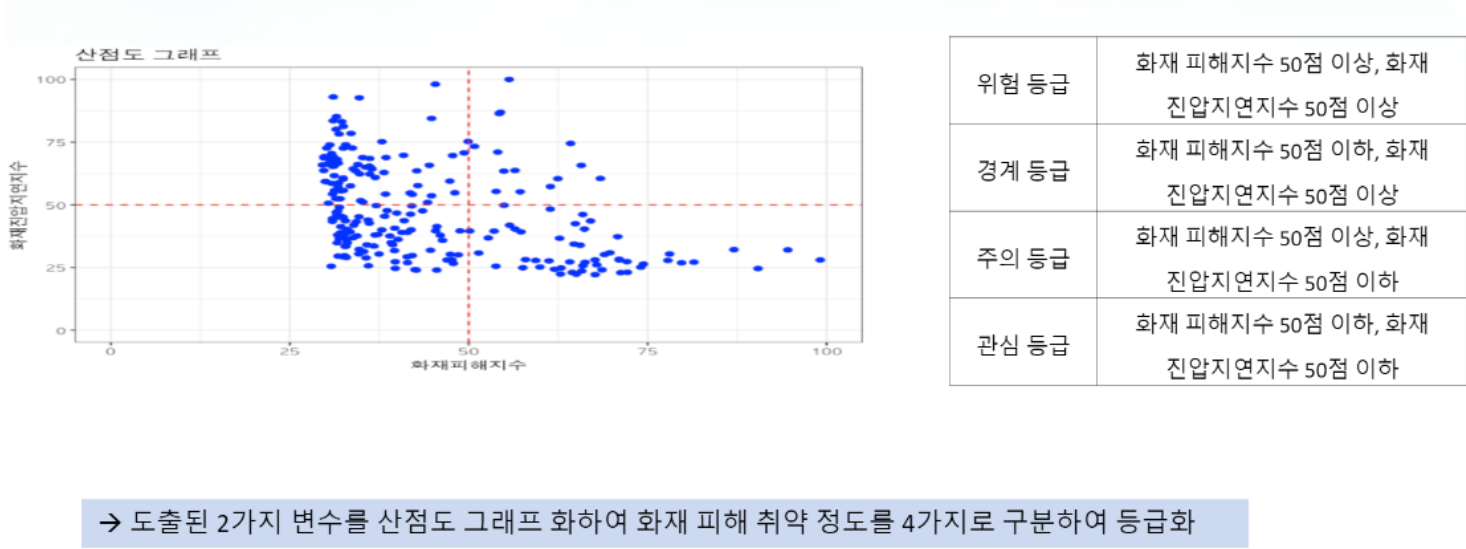
Projects

산업단지 화재 발생 시 피해 최소화를 위한 소방 관제시스템 모델 구축 (2018.10 ~ 2018.11)

Technical Skills: Python, Scikit-learn, R

Goal: 전국 소방서 데이터와 전국 공장 데이터를 이용하여 산업단지 화재 발생 시 피해 최소화를 위한 소방 관제시스템 모델 구축

Result: 기존 데이터를 이용하여 인명피해 지수, 재산피해 지수, 골든타임 지수, 공장규모 지수와 같은 파생변수를 개발하여 화재 피해 취약 정도를 파악



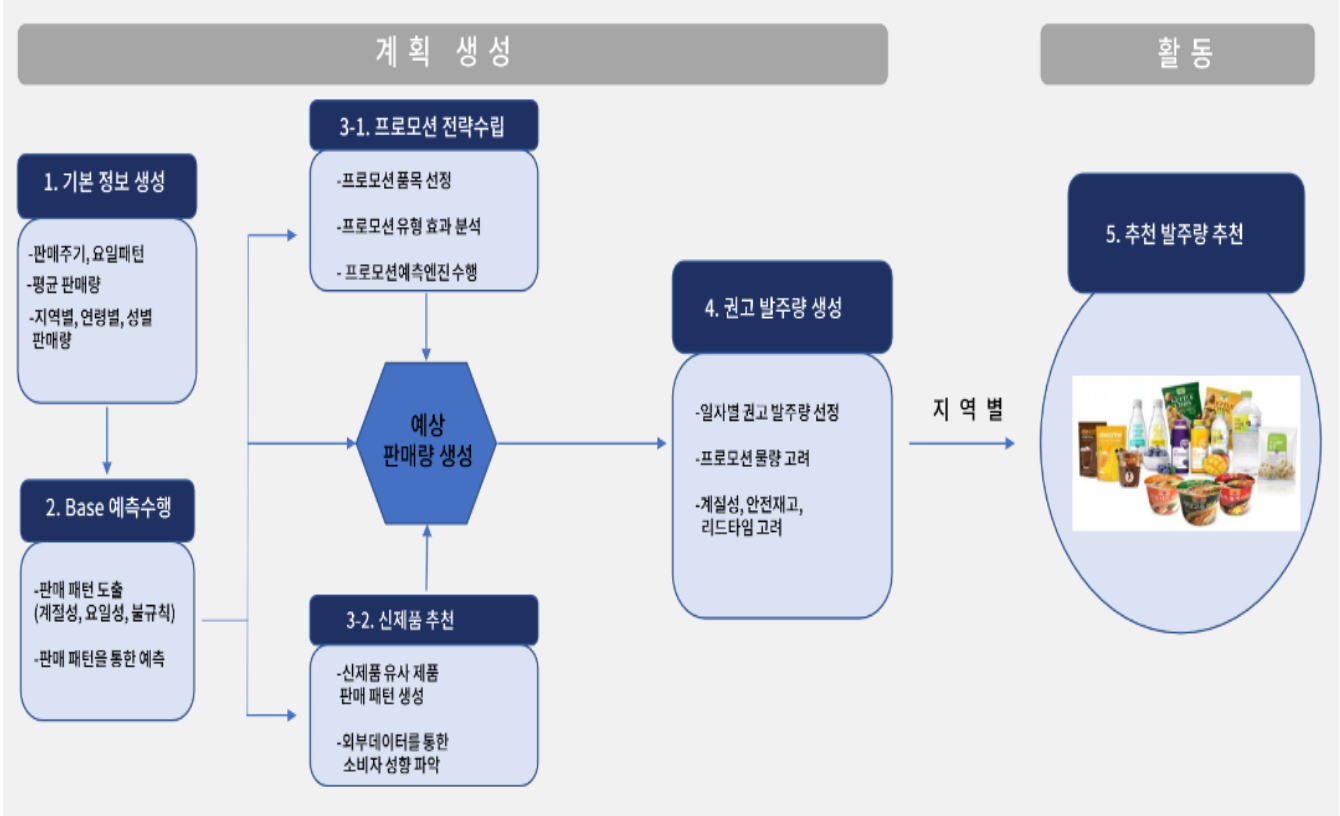
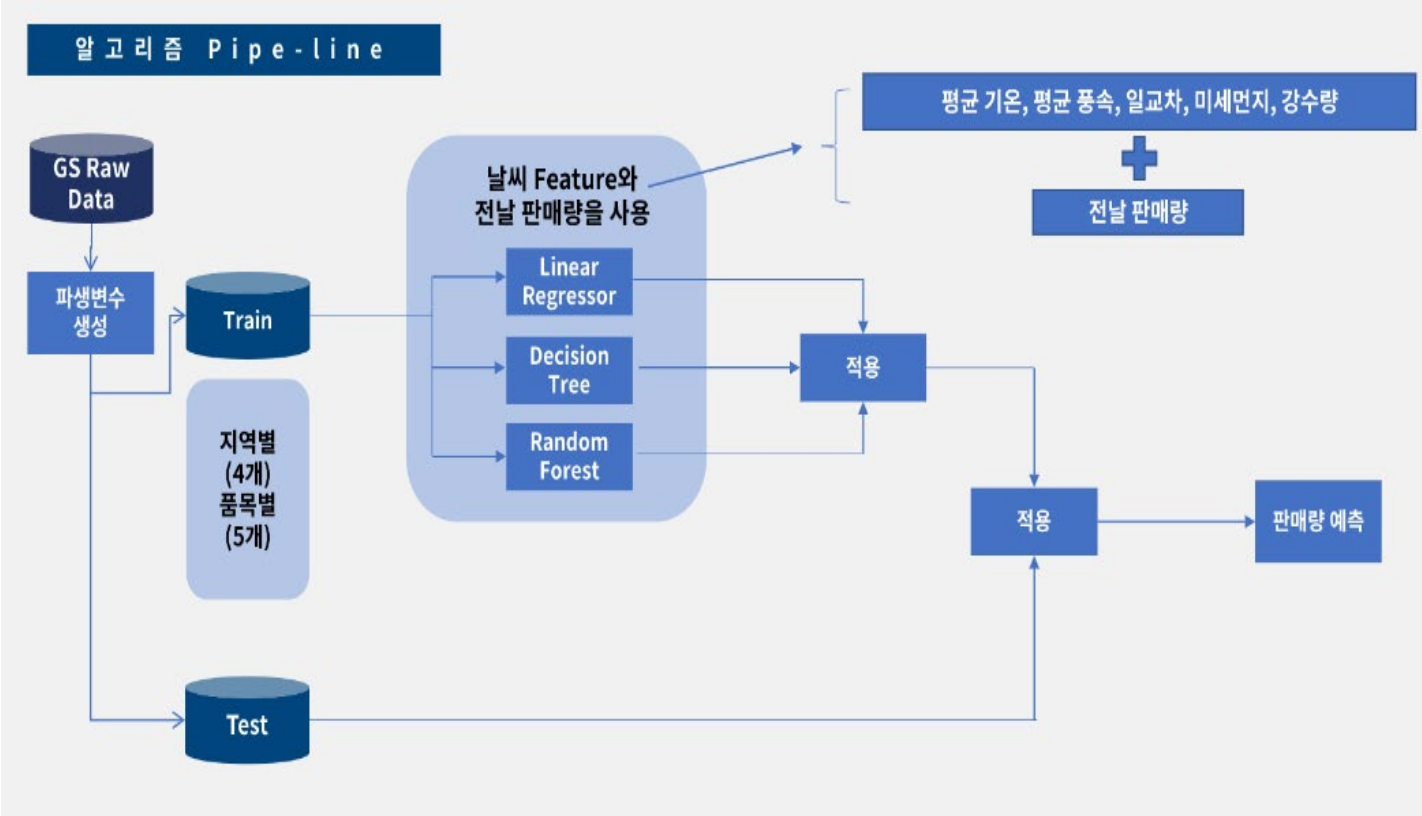
Projects

빅데이터 분석을 통한 GS 편의점 발주량 추천 시스템 (2019.06)

Technical Skills: Python, Scikit-learn, Tensorflow, Keras

Goal: 미세먼지, 날씨, 편의점 내부데이터를 이용하여 발주량 추천 시스템 개발

Result: 지역, 카테고리를 군집화하고, 날씨, 미세먼지, 판매량 등 다양한 feature을 이용하여 발주량을 추천해주는 머신러닝 모델 개발



Projects

김해시 화재 발생 예측 모델 개발 (2019.12)

Technical Skills: Python, Scikit-learn

Goal: 경남지역의 공공데이터를 이용하여, 김해시 화재 발생 예측 모델 개발

Result: 소방시설, 날씨, 건물 특징 등의 feature를 통해 LightGBM 모델을 이용하여 화재 발생 예측 모델 개발



Projects

공공데이터를 이용한 전국 유치원 실태 파악 및 개선 방안 제시 (2018.05)

Technical Skills: Python, Scikit-learn

Goal: 유치원 현황 시설을 데이터 기반으로 파악하여, 유치원 시설 개선방안 제시

Result: 데이터셋을 구축하고 부모님 안심지수, 취원율, 유아 환경지수, 놀이터 비율, 스쿨버스 운영 비율와 같은 파생변수를 개발하여 현황 실태를 파악

