

『MINI AIFFELTHON 보고서』

팀 명		우리 쏘카 타
제안 프로젝트	명칭	카셰어링에서의 데이터 기반 수요예측
	소개	<p>경기도 지역의 쏘카 사용자 연령 중 20,30 대의 연령층과 행복주택 주거지를 데이터 기반 수요예측 모델링</p> <p>전개 방향은 신규수요지를 예측 할 수 있는 모델 구축을 목표로 하며 ,</p> <ol style="list-style-type: none"> 1) 각종 논문 및 카셰어링 비즈니스 분석을 통한 수요예측에 필요한 Feature 를 도출하고 2) 도출 된 Feature 기반 학습에 필요한 내,외부 Data 를 수집한다 3) 기존 Data 와 수집 Data 를 Classification 하여(K-means 등) 수요지에 대한 탐색 및 예측 가능한 모델을 만들어본다 4) 탐색된 수요지에 대해 수요예측모델을 적용하여(딥러닝 모델등) 신규 쏘카존 설치와 기존 쏘카존의 조정과 변경을 실시한다. <p>시간대 및 차종, 요일 예측치등 제공가능한 예측치를 생성하여 카셰어링 비즈니스를 제안한다</p>

1.수행 및 분석

1] Feature 도출 및 Data 확보

1) 논문 및 카셰어링 비즈니스 분석을 통한 수요지 예측에 필요한 Feature 도출

- 미래형 공동주택의 청년계층 카셰어링 이용수요 분석
(토지주택연구원 국토도시연구실)
- 서울시의 카셰어링 이용도에 대한 지역적 요인특성분석
(한국철도학회논문집)
- 군집화에 의한 예측모델 성능 고도화 방법(경영정보관련학회)
- Feature 선정

의존지역		양의상관	음의상관	양의상관	양의상관
	자가용 업무중심지역	역세권지역	종합개발지역	버스편이지역	녹색교통 편이지역
	Commercial centric	Station sphere,		Bus centric,	Green mode
Num of parking lot	957	-152	65	83	-3
Registered private car	-918	-189	-110	207	139
Registered car	-907	-170	-110	230	-181
Tax revenue	899	-359	25	-12	-73
Num of employees	811	530	72	57	-16
Num of financial compar	771	565	119	157	-4
Diffusion ratio of house	502	83	485	182	377
Density of employees	128	-876	6	-103	272
Density of company	209	-859	78	-156	261
Num of company	335	858	207	257	-50
Ratio of employ (employs	350	824	207	298	-37
Num of railway station ar	85	44	847	342	9
Num of bus station	126	216	818	219	-138
Street ratio -	-380	66	-716	-268	-298
Num of entrance of statio	18	13	632	-48	-422
Deg of public developme	92	156	46	-861	-118
Deg of complex develop	3	371	250	850	104
Deg of commercial devel	87	440	408	650	-31
Interval of bus	-53	-134	26	11	924
Num of long distance ro	-40	-239	-24	19	878
Num of business vehicles	325	-11	-114	475	-531
Accessibility of pedestrian	205	-50	-205	172	153
Accessibility of bicycle -	-145	-6	264	113	176
Interval of subway -	-138	-11	196	-44	-127

2) Feature 에 적절한 내,외부 Data 확보

- 참고 Site

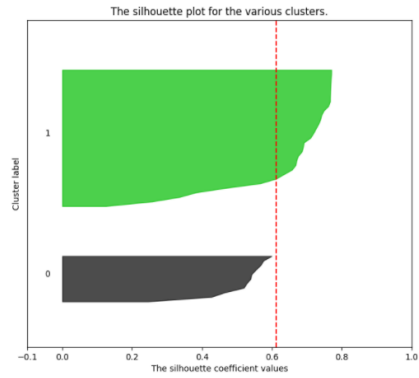
- . 공공데이터포털(<https://www.data.go.kr/>)
- . 국가통계포털(<https://kosis.kr/>)
- . 지표누리(<https://www.index.go.kr/>)
- . 경기통계(<https://stat.gg.go.kr/>)

- 수집 Data

- . 인구현황 : 관련문서 19 종
- . 소득현황 : 관련문서 14 종
- . 대학교현황 : 관련문서 5 종
- . 아파트현황 : 관련문서 2 종
- . 대중교통(지하철,버스) : 관련문서 5 종
- . 쏘카존 현황 : 관련문서 1 종
- . 토지용도 : 1 종

2] 미니 아이펠톤 수행 결과 : 1 차 수요지 선정 및 모델링

- 문제정의 : 경기도 42 개 지역(시, 군, 구)에 대한 군집화 수행
- 데이터 수집 (6 개 feature)
 - . 쏘카존 수, 토지용도, 인구, 대학교 수, 아파트 수, 지하철역 수
- 모델링
 - . 사용모델 : k 평균 군집화 (k-means clustering)
 - . 수행절차
 - . 군집개수(K)를 달리하여 k 평균 군집화를 진행
 - . 실루엣 계수가 가장 높은 K 값을 선정
 - . 수행결과
 - . K=2 또는 K=3 으로 선정
 - . 42 개 지역에 대한 군집화 작업 완료



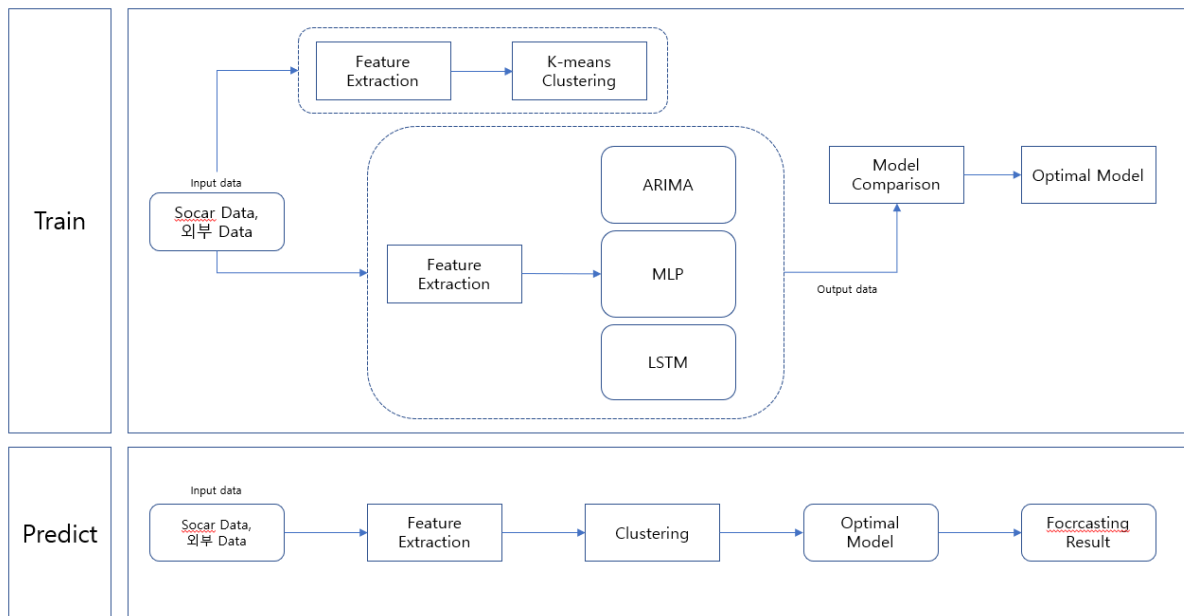
```
regions = np.array(collected_data.index)
print("군집1 : ", regions[np.where(cluster_labels==0)])
print()
print("군집2 : ", regions[np.where(cluster_labels==1)])
```

군집1 : ['고양시 덕양구' '김포시' '남양주시' '부천시' '성남시 분당구' '시흥시' '용인시 기흥구' '의정부시' '파주시' '평택시' '화성시']

군집2 : ['가평군' '고양시 일산동구' '고양시 일산서구' '과천시' '광명시' '광주시' '구리시' '군포시' '동두천시' '성남시 수정구' '성남시 중원구' '수원시 권선구' '수원시 영통구' '수원시 장안구' '수원시 팔달구' '안산시 단원구' '안산시 상록구' '안성시' '안양시 동안구' '안양시 만안구' '양주시' '양평군' '여주시' '연천군' '오산시' '용인시 수지구' '용인시 처인구' '의왕시' '이천시' '포천시' '하남시']

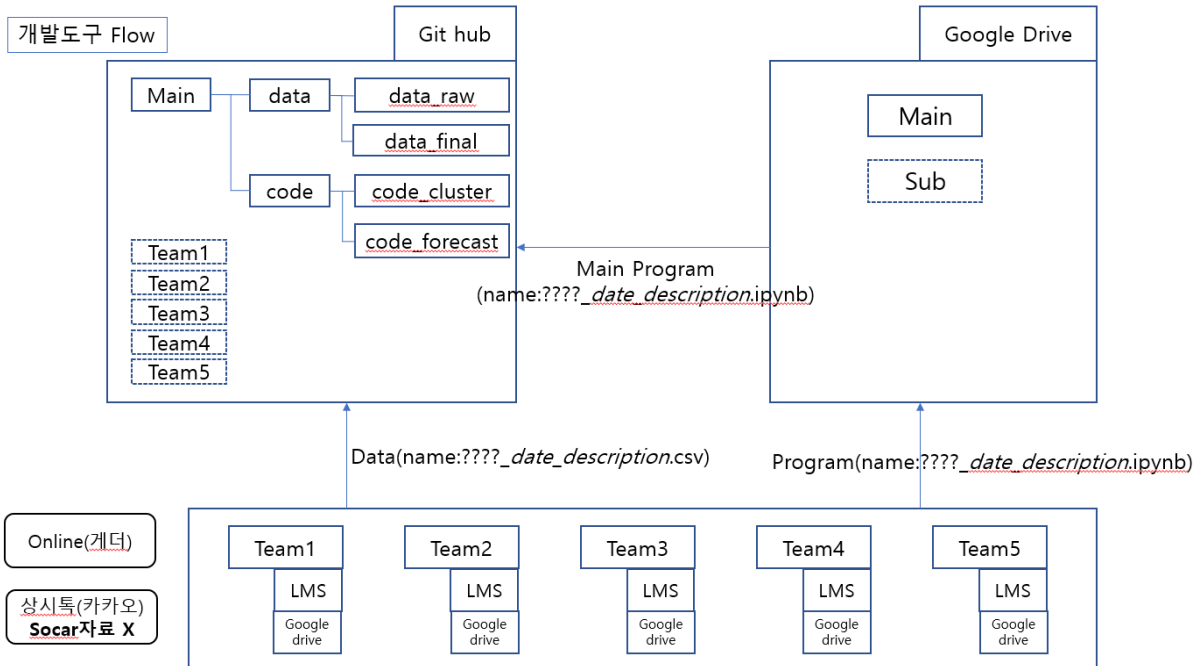
3] 시스템 구성 및 아키텍처

- 시스템 구성도



<Figure> 군집화-수요 예측 모델 구조

- 개발 Kit



2.결과 및 계획

1] 변경 계획안

Task		목표기간	세부내용
팀빌딩	- 팀빌딩 및 계획서 제출 · <u>제출 : 2022.12.26 18:00</u>	2022.12.26 ~ 2022.12.26	- 팀장 및 역할, 운영 정립
	- 운영관리체계 정립	2022.12.27 ~ 2022.12.27	- 개발관리체계 정립 · 프로그램 운영 체계 · Naming Rule 정립 · 운영체 소통관리(on,offline)
	- 미니 프로젝트	2022.12.27 ~ 2023.01.02	
Mini Aiffelthon	· Data 선정 및 EDA	2022.12.28 ~ 2023.12.30	- 데이터 확보 및 관리체계 수립 - Data 분석 및 관계도 작성
	· 모델 학습 및 선정	2022.12.28 ~ 2023.12.30	- 2 ~3 모델 학습 · Hyper parameter 검토
	· 프로젝트 계획서 수정 · <u>제출 : 2023.01.03 18:00</u>	2022.12.27 ~ 2023.01.03	- 프로젝트 계획서 작성 · 진도관리 체계 포함
Aiffelthon	- 카세어링비즈니스 분석	2022.01.02 ~ 2023.01.04	- 카세어링 Flow 검토
	- Feature 검토 및 재구성	2022.01.02 ~ 2023.01.06	- Feature 정리표
	- Data 확보 및 전처리	2022.01.04 ~ 2023.01.13	- Data 분석 및 관계도 재검토 · 산출물 : Data set 구성
	- 모델 검토 및 모델링 · Clustering 모델 · 수요예측 모델	2022.01.09 ~ 2023.01.20	- 모델 학습 : 이해도 향상 - 실 Data 로 모델링 · hyper parameter 선정
	- 모델 평가 및 재구성 · Clustering 모델 · 수요예측 모델	2022.01.16 ~ 2023.01.31	- Optimizer , 손실함수
	- 프로젝트 완료 보고서	2023.01.23 ~ 2023.02.03	- 프로젝트 완료 보고서 - 소스코드
	- 개발프로그램 취합 및 정리	2023.01.30 ~ 2023.02.03	
	- 발표 자료 최종 검토 · <u>제출 : 2023.02.07 18:00</u>	2023.02.06 ~ 2023.02.07	

2] 발생문제 및 개선 사항

1) 수요지 예측 Feature 적합성 검토가 필요

- 논문 및 비즈니스 분석을 통해 선정된 Feature 가 적절한지 분석 및 검토가 필요
 - . 군집화모델을 통한 수요지 Grouping : 기존과 Data 수집된 지역의 비교 분석
 - . 멘토를 통한 자문을 구함

2) 선정된 Feature 에 대한 Data 확보

- 외부 Data 의 소스 및 부족한 경우가 발생
 - . 1 월 첫째주까지 Data 확보를 마치도록 하고, 부족한 Data 는 대체 Feature 를 선정하거나 Data 생성모델로 극복 예정

3) 배포까지 구성하는 것에 대한 고민

- 새로운 수요지에 대한 Feature 구성 및 외부 Data 의 확보의 어려움으로 우선 신규 수요지 선정 및 선정된 지역의 수요 예측 모델의 완성을 목표로 함.