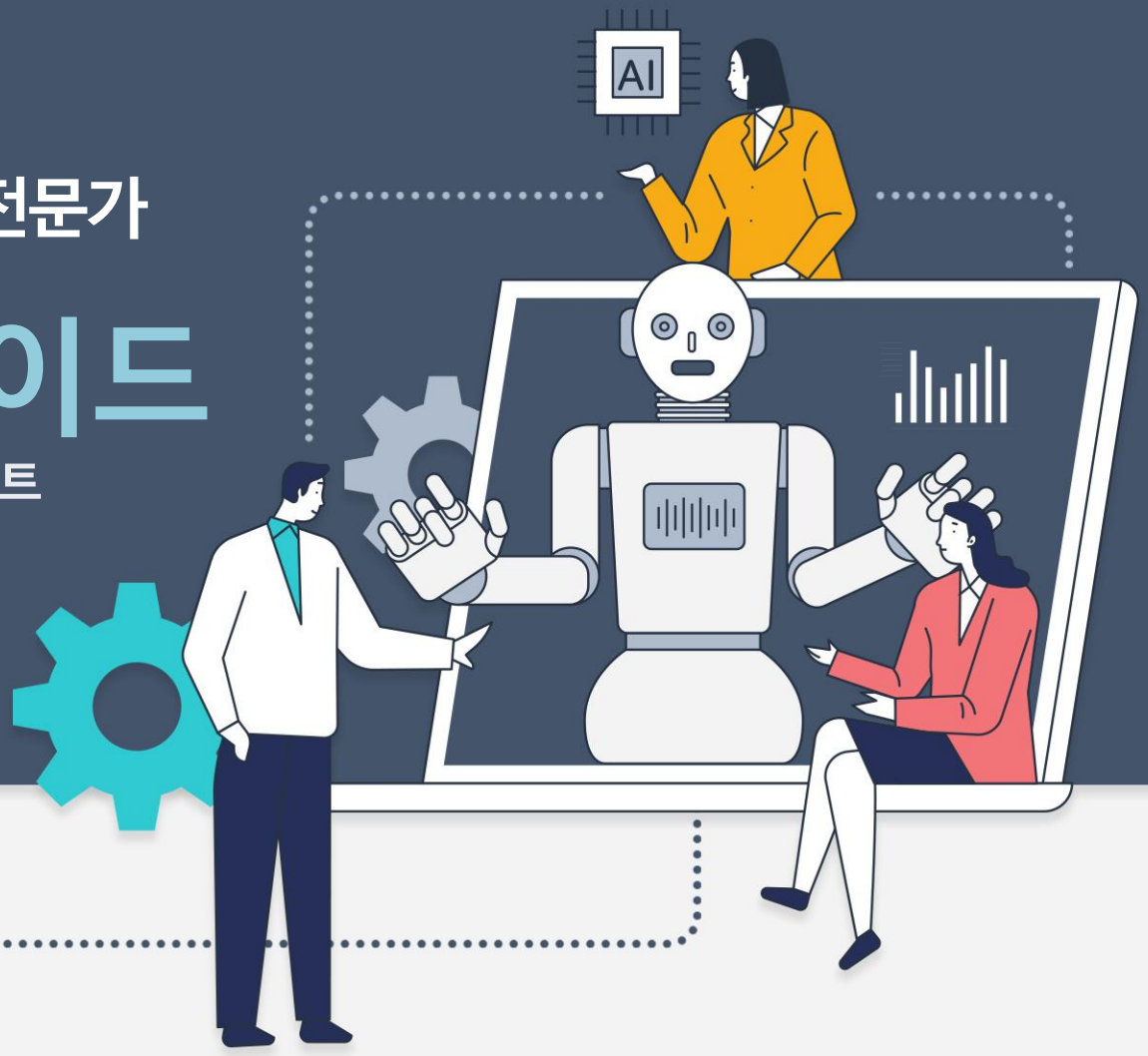
데이터 사이언스/엔지니어링 전문가

포트폴리오 가이드

- 기업 요구사항 기반의 문제해결 프로젝트



포트폴리오 가이드

21KDT

1. 포트폴리오는 **팀별로 작성하여 제출**
   (LC 과제제출에는 개별 업로드)

2. 제공된 목차 및 작성요령을 참고하여 작성하되,
   템플릿/디자인/구성은 변경 및 추가 가능

3. 포트폴리오를 바탕으로 프로젝트 발표 진행
   - 20~30 슬라이드
   - 발표는 결과 위주로
   - Q&A 포함 30분 이내

포트폴리오 목차

21KDT

01. 프로젝트 배경

✎ 작 성 요 령 ✎

❖ [프로젝트 배경]은 아래와 같은 내용 등으로 구성하여 작성한다.

- 프로젝트 주제 및 선정배경
- 가상 시나리오
- 프로젝트 개요
  - 컨셉, 훈련 내용과의 관련성, 개발 환경 등
- 프로젝트 구조

02. 팀 구성 및 역할

✎ 작 성 요 령 ✎

❖ [팀 구성 및 역할]은 훈련생 별로 해당 프로젝트를 진행하면서 주도적으로 참여한 부분을 중심으로 작성한다.

- 표, 도식 활용하여 팀원 각각의 역할을 작성

03. 수행절차 및 방법

✎ 작 성 요 령 ✎

❖ [수행절차 및 방법]은 프로젝트 수행절차 및 방법을 제시한다.
(구성요소를 포함하여 예시 표와 다르게 수정하여 작성 가능함)

03-1. work-flow

21KDT

구분		기간	활동	도구
사전 기획		• O/O(월) ~ O/O(목)	• 프로젝트 기획 및 주제 선정 • 기획안 작성	
		• O/O(금)	• 프로젝트 주제 & 아이디어 발표	
데이터 구축 및 분석	데이터 수집 및 학습 시스템 구축	• O/O(월) ~ O/O(목)	• 필요 데이터 및 수집 절차 정의 • 외부 데이터 수집	
		• O/O(월) ~ O/O(목)	• 데이터 시스템 구축	
	데이터 전 처리 및 분석	• O/O(월) ~ O/O(목)	• 데이터 선정 및 데이터 전 처리 /탐색	
		• O/O(월) ~ O/O(목)	• 모델링(모형구현)	
서비스 구축 및 최적화		• O/O(월) ~ O/O(금)	• 서비스 시스템 설계 • 서비스 플랫폼 구현 • 최적화 및 오류수정	
프로젝트 발표		• O/O(월)		

✎ 작 성 요 령 ✎

❖ [프로젝트 수행 결과]는 프로젝트 결과물이 도출된 과정을 세부적으로 기록

- 예시(8 ~18p)는 하나의 사례로 간단하게 제시한 것이므로 프로젝트의 성격에 따라 보다 자세하게 기록하며, 결과를 서술하는 과정에서는 논리성, 창의성, 완결성이 잘 드러나도록 한다.

- 프로젝트의 결과는 그 과정이 잘 드러날 수 있도록 데이터 수집 가공 과정부터 분석 활용까지 전체적인 프로세스를 확인할 수 있도록 단계별로 작성

04-1. 데이터명세

21KDT

출처	데이터 이름	제공 형태	요약
서울 열린데이터 광장	서울 생활인구 내국인	csv	기준일, 시간대, 자치구코드, 총생활인구, 성별/나이별
서울 열린데이터 광장	자치구 코드	xlsx	시도, 시군구, 구 이름, 전체 이름
공공데이터포털	서울시 코 확진자	penApi	연번, 확진일, 국적, 지역, 여행력, 접촉력, 조치사항, 상태
공공데이터포털	서울시 코로나19 예방접종 현황	openApi	접종일, 접종대상자, 당일 1차접종, 1차누계, 1차접종률, 당일 2차접종, 2차누계, 2차접종률

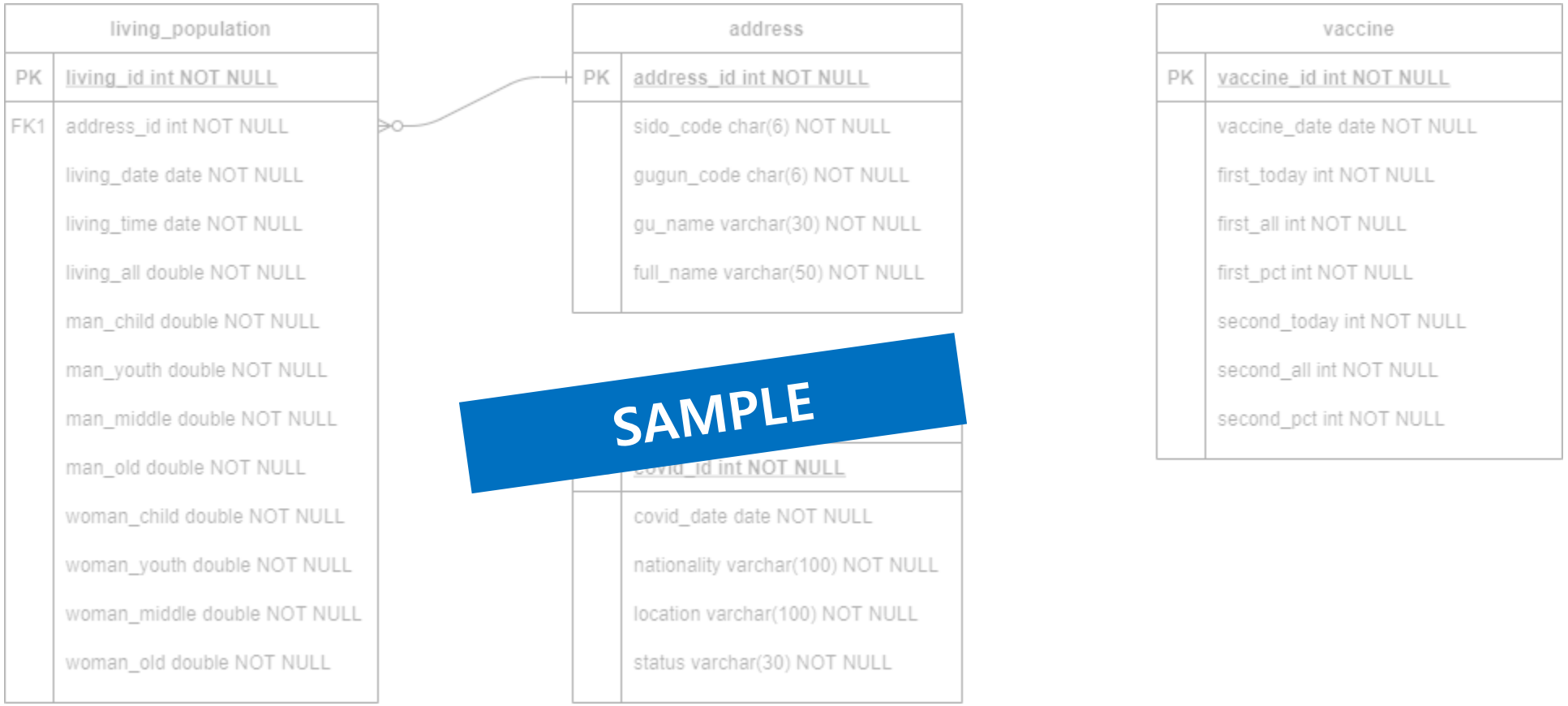
04-2. 데이터 파이프라인

21KDT

SAMPLE

04-3. ERD

21KDT



SAMPLE

04_4. 데이터 전 처리 및 탐색, 시각화

21KDT

	강수량	적설	눈비
0	0.0	0.0	0
1	0.0	0.0	0
2	0.0	0.0	0
3	0.0	0.0	0
4	0.0	0.0	0
...	...	...	...
2646186	0.0	0.0	0
2646187	0.0	0.0	0
2646188	0.0	0.0	0
2646189	0.0	0.0	0
2646190	0.0	0.0	0

	운량	날씨
0	4.0	2
1	6.0	3
2	9.0	4
3	7.0	3
4	6.0	3
...	...	
2646186	7.0	
2646187	8.0	
2646188	8.0	3
2646189	7.0	3
2646190	6.0	3

	날짜	계절
0	2019-07-17	여름
1	2019-07-17	여름
2	2019-07-17	여름
3	2019-07-17	여름
2646187	2020-09-29	가을
2646188	2020-09-29	가을
2646189	2020-09-29	가을

눈비

강수량 적설 결합 ➔ 값의 존재 유무에 따라
1, 0에 해당하는 범주형 변수 생성

날씨

운량 값을 기준으로 순서형 변수(1, 2, 3, 4) 생성

운량값	0 – 1	3 – 5	6 – 8	9 – 10
날씨	맑음	구름조금	구름많이	흐림
변수	1	2	3	4

※ 기상청 하늘상태 표현을 참고하여 세분화하였음

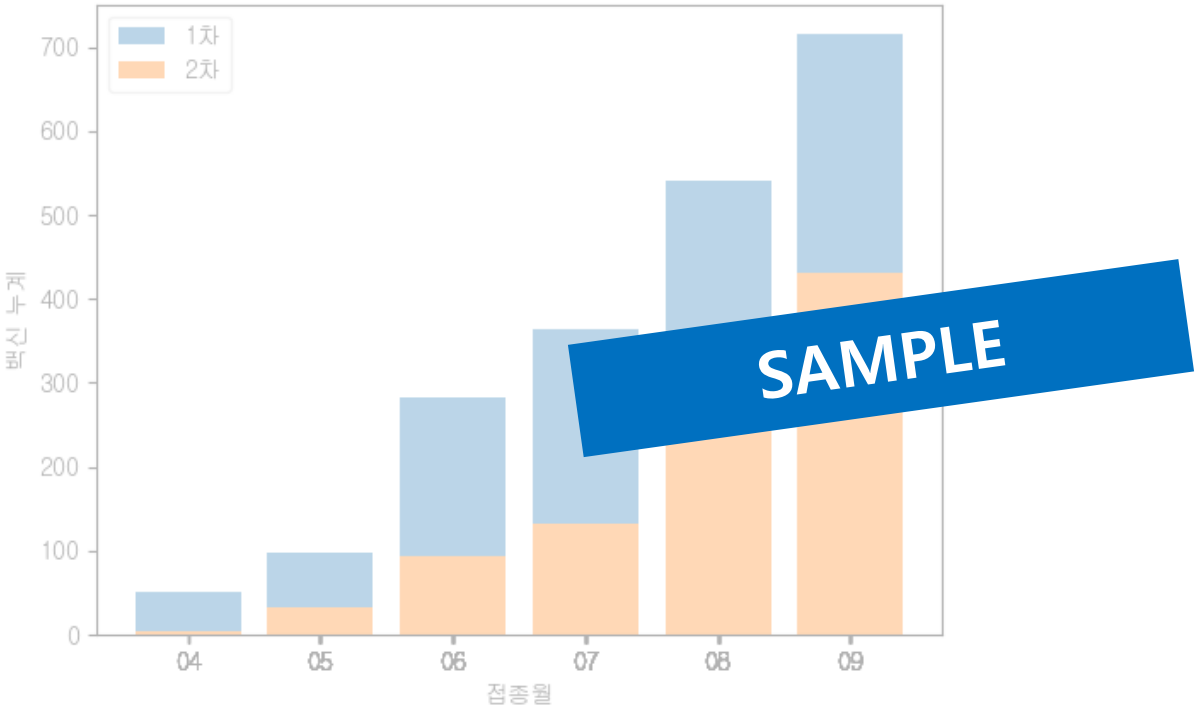
계절

날짜	3 – 5월	6 – 8월	9 – 11월	12 – 2월
계절	봄	여름	가을	겨울

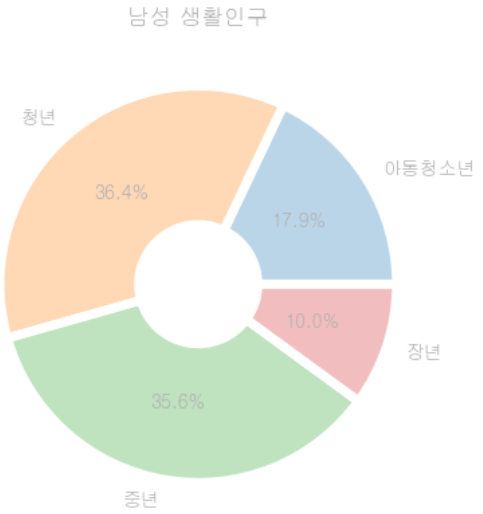
※ 한국민족대백과사전 기준으로 세분화하였음

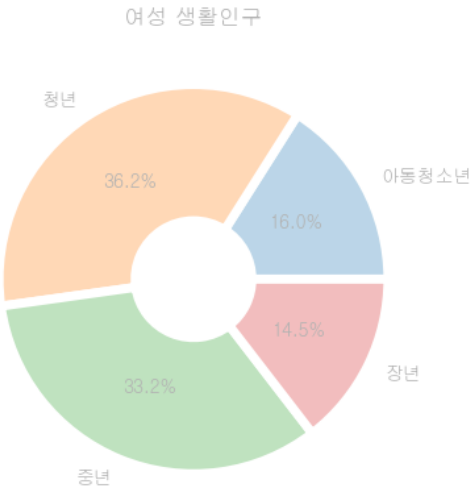
04-4. 데이터 시각화

21KDT



SAMPLE





04_5 Modeling(모델 개요 및 분석)

21KDT

1) 지역별/업종별 데이터로 분리

Data Set
Train 8 : Test 2

SAMPLE

Test 2 으로 데이터 분할

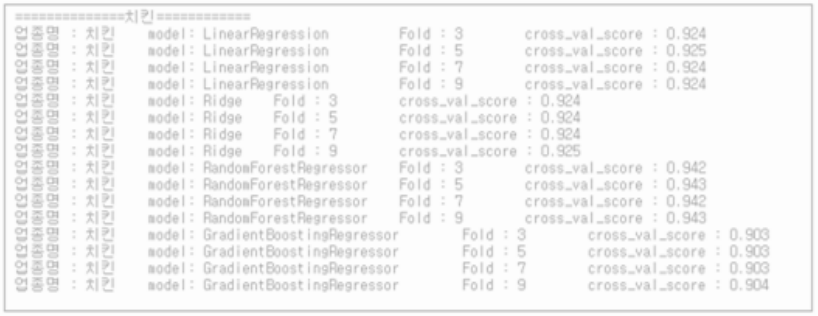
Ridge Regression

Linear Regression

RandomForestRegressor

GradientBoostingRegressor

3) 회귀모형에 적합



4) 모델 평가와 예측력 확인

모델명	스코어
RandomForestRegressor	0.876
LinearRegression	0.832
Ridge	0.832
GradientBoostingRegressor	0.828

Linear 보다 계수가 안정적

5) 각 모델별 Score 합계 산출하여 스코어 상위 2개 모델을 선택

Ridge, RandomForestRegressor 모델 선택

04_5. Modeling(모델 평가 및 개선)

21KDT

	광역시도명	업종명	모델명	fold 수	스코어
4	서울	기타	Ridge	3	0.663
5	서울	기타	Ridge	5	0.663
6	서울	기타	Ridge	7	0.663
7	서울	기타	Ridge		
8	서울	기타	RandomForestRegressor		
...	...	...	...		
343	경기도	한식	Ridge	9	0.932
344	경기도	한식	RandomForestRegressor	3	0.952
345	경기도	한식	RandomForestRegressor	5	0.953
346	경기도	한식	RandomForestRegressor	7	0.953
347	경기도	한식	RandomForestRegressor	9	0.954

SAMPLE

	fold 수	스코어
0	3	0.852591
1	5	0.854182
2	7	0.854591
3	9	0.854977

1) 지역 / 업종별 Ridge & RF Model Fold 값, Score 값 추출

2) 전체 Ridge & RF Model Fold 별 Score 값의 평균

Fold 수가 9인 경우 Model의 성능이 좋음

04-5.Modeling(모델 선정 결과)

21KDT

모델 비교

모델명	Ridge	Random Forest Regressor
Train R^2	0.835	0.915
Test RMSE		0.388
Test R^2		0.862
학습시간(초)	0.050	7.629

※ 학습시간은 치킨(업종)을 기준으로 산출하였음

Train Score의 경우 RF가 우수한 성능

웹 구현까지 고려해봤을 때, 준수한 성능과 빠른 시간을 보이는

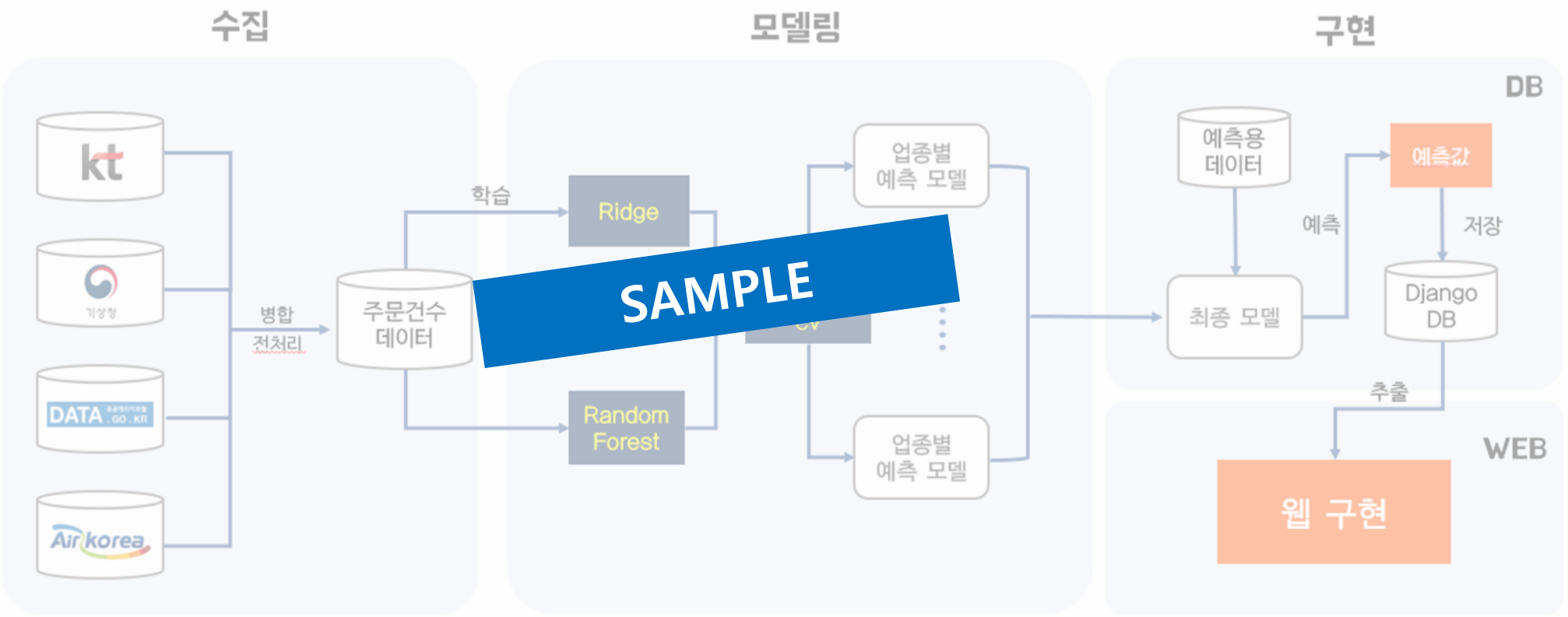
Ridge 모델을 적용하는 것이 현실적으로 타당

04_6 서비스 개요

21KDT

04_6 서비스 work-flow

21KDT



04_7 구현 서비스 시연

05. 결론 및 향후 과제

21KDT

✎ 작 성 요 령 ✎

❖ [결론 및 향후과제] 프로젝트에 대한 결론 및 향후 현 프로젝트 결과를 발전시킬 수 있는 방향을 제시한다

07. 느낀점

21KDT

✎ 작 성 요 령 ✎

❖ [느낀 점]은 프로젝트 수행에서 개인, 우리 팀이 잘한 부분과 아쉬운 점을 작성한다. 또한 프로젝트를 수행하면서 느낀 수행 상 어려움, 갈등요소 등을 작성하고 이를 해결한 방법을 작성한다.