



애플리케이션 기획 배경 및 목표

현재 사람들은 좌측의 사이트와 같이 대기의 미세먼지는 알 수 있는 사이트는 알고 이를 이용하는 사람들도 있으나 우리는 그 다음날, 그 다음주, 의 미세먼지 농도를 예측해주면서 비교적 농도가 낮은 구 단위를 추천해주면서 그 추천 장소 주변의 좋은 음식점 카페들을 평점과 같은 지표에 따라서 추천해주는 앱을 만드는 것이 목표이다.



1. API 접근 -> 현재 위치의 미세먼지 + 대기정보 값 출력
2. RANDOM FOREST 알고리즘을 이용해 데이터를 바탕으로 미세먼지 값 예측
3. 예측한 미세먼지 데이터를 바탕으로 비교적 대기오염 정도가 낮은 인근 장소를 추천

API에 접근(DB) + 현재 위치(GPS)의 미세먼지 정보 표시

데이터명	한국환경공단_대기오염정보	상세설명	
서비스유형	REST	심의여부	자동승인
신청유형	개발계정 활용신청	처리상태	승인
활용기간	2020-11-25 ~ 2022-11-25		



Sw 개요 및 구조 -1

공공 데이터 기관에서 미세먼지 측정 결과 API를 받아오는 서비스를 이용

동작구		미세먼지 보기
이름	동구	
대기정보가 업데이트 되었습니다.		
측정시각	2020-12-02 17:00	
아황산가스 농도	0.008ppm	
일산화탄소 농도	0.4ppm	
오존 농도	0.061ppm	
이산화질소 농도	0.032ppm	
미세먼지 농도	87 $\mu\text{g}/\text{m}^3$	
통합대기환경 수치	85	
통합대기환경 지수	2등급	
아황산가스 지수	1등급	
일산화탄소 지수	1등급	
오존 지수	2등급	
이산화질소 지수	2등급	
미세먼지 지수	2등급	

-인터넷을 통해 정보를 받아오므로 AndroidManifest.xml에 퍼미션 추가

-라이브러리 retrofit(http통신을 수월하게 해주는 라이브러리) 추가

-위치를 입력하고 미세먼지 보기 버튼을 누르면 위치 값을 통해서 측정소에서 정보를 가져와 텍스트 뷰에 정보 값들이 보여지도록 구성

<인터페이스 제작>



Sw 개요 및 구조 -2

미세먼지 예측방법 – RANDOM FORREST 위주로

먼저 데이터들을 각 변수들을 모아서 데이터셋을 만든다. 년도, 월, 날짜순으로 SO₂, CO, O₃, NO₂, PM₁₀, 기온, 강수량, 풍향, 풍속, 상대습도등을 가지고 하나의 엑셀파일, CSV파일을 만든다.

2010~ 2017의 자료를 모아서
아래와 같은 데이터 셋을 만들었다.

location	year	month	day	SO2	CO	O3	NO2	PM10	temp	prec	cbwd	ws	humi	pres	gasoline	kerosene	bunker	LPG	oil
강남구	2010	1	1	0.006	0.6	0.007	0.037	36	-7.6		70	1.9	48	1010.6	794	224	32	880	7452
강남구	2010	1	2	0.008	0.6	0.008	0.037	56	-3.6	1.4	290	2.8	76.4	1006.2	794	224	32	880	7452
강남구	2010	1	3	0.007	0.7	0.012	0.041	50	-6.8		290	2.2	49.3	1010.7	794	224	32	880	7452
강남구	2010	1	4	0.005	0.6	0.009	0.042	38	-5.9	14.2	320	3	84.1	1002.4	794	224	32	880	7452



SW 개요 및 구조 -3

```
DATA <- read.csv("C:/Users/MJ/Desktop/Data.csv")
install.packages("pls")
library(pls)
```

```
str(DATA)
set.seed(3)
train=sort(sample(1:nrow(DATA), 2338))
test=setdiff(1:585, train)
install.packages("randomForest")
library(randomForest)
rf.wine=randomForest(PM10~., data=DATA, subset=train, mtry=3, importance=TRUE)
yhat.rf = predict(rf, DATA, newdata=DATA[test,])
mean((yhat.rf-DATA.test)^2)
importance(rf, DATA)
varImpPlot(rf, DATA)
rf, DATA
```

Call:

```
randomForest(formula = PM10 ~ ., data = Wine, mtry = 3, importance = TRUE,
subset = train)
```

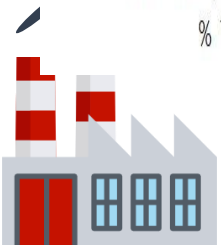
Type of random forest: regression

Number of trees: 500

No. of variables tried at each split: 3

Mean of squared residuals: 0.3712579

% Var explained: 89.55



미세먼지 예측방법 - RANDOM FORREST 위주로

- 위에서 만든 데이터셋을 이용해서 일종의 회귀 모형을 만든다.
- 데이터셋을 불러내서 RandomForest패키지를 이용하여 데이터를 training data, test data로 8:2의 비율로 분류하고 각각의 데이터를 바탕으로 분류하는 분류도구를 만든다.
- 분류를 해 보면 대략 90퍼센트의 확률로 맞춤으로 나오는데 이 정도면 꽤나 좋은 예측력을 보이므로 그대로 사용하거나 아니면 좀 더 수정해서 좋은 모형을 만들고 최종 분류를 한다.
- 위에서 나온 모형을 가지고 최종 분류를 하여 나온 값들을 이용해서 각각 예측을 한다.
ex)강남구의 21년 1월의 PM10은 60이다.



예시:맛집추천 어플리케이션 다이닝코드



SW 개요 및 구조 -4

대기 오염 정보가 낮을 것으로 예측되는 장소를 추천

-selenium이라는 모듈을 이용하여 네이버 플레이스,카카오 맵 등 음식점, 데이트 장소 등의 장소를 평점 매기는 사이트/앱에 평점,리뷰,날짜,위치 등을 크롤링한다

-크롤링한 데이터로 우선 시/군/구 등 장소별로 나누고, 리뷰 날짜별로 구별한뒤, 마지막으로 장소마다 평균평점을 낸 후 평점이 3.5이상인 경우만 저장하여 순위를 매긴다. 그렇게 분류한 데이터를 DB를 생성해 저장한다

-앞에서 예측한 미세먼지 정보를 DB에 저장하여 미세먼지 테이블을 생성한다

-구글 맵 서비스를 이용하여 메인화면에 지도를 생성하고 미세먼지 수치가 낮을 것으로 예측 되는 장소를 추천