



SMU

AI+X 선도인재양성 프로그램

머신러닝 사례

고시마전지 노동 여가



프로젝트 개요

Intro

미니프로젝트 개요

주요 내용은 다음과 같습니다!

구분	내용
미션	[개인 미션] 미세먼지 농도 예측 머신러닝 모델링
사전학습과목	데이터 처리, 머신러닝
데이터 출처	① 에어코리아 – 미세먼지 데이터 ② 기상청 기상자료 개방포털 – 날씨 데이터
데이터 구분	Tabular 데이터 csv 4개 파일

Intro

미니프로젝트 목표

이번 미니 프로젝트의 목표는?

구분

내용

목표

새로운 도메인의 학습데이터를 탐색하고
분석에 용이하도록 **전처리**한 후
머신러닝 모델을 완성해
테스트데이터에 대해 **예측**하고 **평가**한다.

데이터 탐색

데이터 전처리

알고리즘 선택

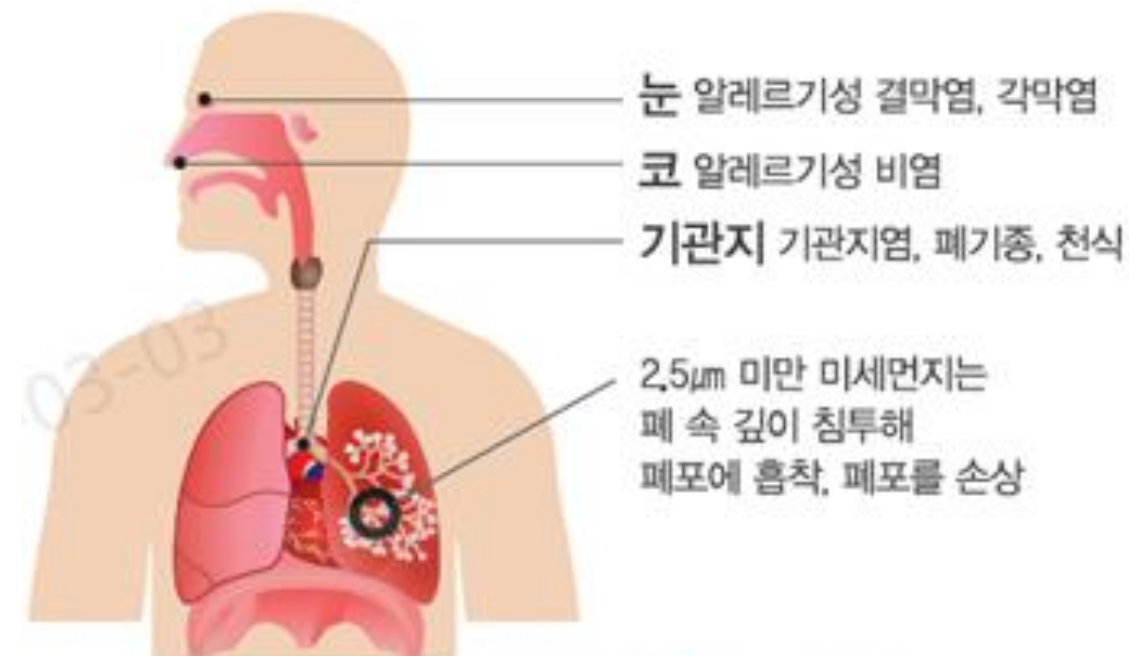
머신러닝 모델
완성

예측 및 평가

Domain Knowledge

도메인 이해

- 대기환경기준물질: 아황산가스, 일산화탄소, 이산화질소, 오존, 미세먼지(PM-10, PM-2.5)
- 미세먼지: 아주 작은 크기의 모든 오염물질
대기 중에 떠다니거나 흩날려 내려오는 입자상물질인 먼지 중 다음의 흡입성 먼지
 - 미세먼지 : 입자의 지름이 10마이크로미터(μm) 이하인 먼지(PM-10)
 - 초미세먼지 : 입자의 지름이 2.5마이크로미터(μm) 이하인 먼지(PM-2.5)



〈미세먼지가 인체에 미치는 영향〉

환경부, 법제처

Domain Knowledge

도메인 이해

- 환경부에서는 대기오염으로 인한 국민건강 피해를 최소화하기 위해 미세먼지 농도를 하루 4번 예보
 - 예보 등급을 좋음, 보통, 나쁨, 매우나쁨의 네단계로 구분
 - 하루 4회(오전5시, 오전11시, 오후5시, 오후11시) 예보
 - 전국 10개 권역에 대한 대기질 예보 제공

예보의 정확도가 중요!!!

<미세먼지 농도별 예보 등급>

좋음 PM₁₀ 0~30(μg/m³) PM_{2.5} 0~15(μg/m³)

대기오염 관련 질환자군에서도 영향이 유발되지 않을 수준



보통 PM₁₀ 31~80(μg/m³) PM_{2.5} 16~35(μg/m³)

환자군에게 만성 노출시 경미한 영향이 유발될 수 있는 수준



나쁨 PM₁₀ 81~150(μg/m³) PM_{2.5} 36~75(μg/m³)

환자군 및 민감군(어린이, 노약자 등)에게 유해한 영향 유발, 일반인도 건강상 불쾌감을 경험할 수 있는 수준



매우나쁨 PM₁₀ 151(μg/m³)이상 PM_{2.5} 76(μg/m³)이상

환자군 및 민감군에게 급성 노출시 심각한 영향 유발, 일반인도 악한 영향이 유발될 수 있는 수준



환경부

Example

실제 사례 소개

기상환경데이터와 머신러닝을 활용한 미세먼지농도 예측 모델

- 임준묵. "기상환경데이터와 머신러닝을 활용한
미세먼지농도 예측 모델." 한국 IT 서비스학회지
18.1 (2019): 173-186.

Example

실제 사례 소개

1. 미세먼지 농도 예측의 의미

미세먼지 자체를 없애는 것은 사실상 불가능 하므로 노출 최소화를 위해 미세먼지 농도 예측

2. 미세먼지 농도 측정소

미세먼지 측정소는 특정 지역에 집중, 반면 기상자료(날씨) 측정소는 전국에 골고루 위치 따라서, 기상자료(날씨)를 통해 미세먼지 농도를 예측할 수 있다면 유용

3. 미세먼지 농도에 영향을 미치는 요인

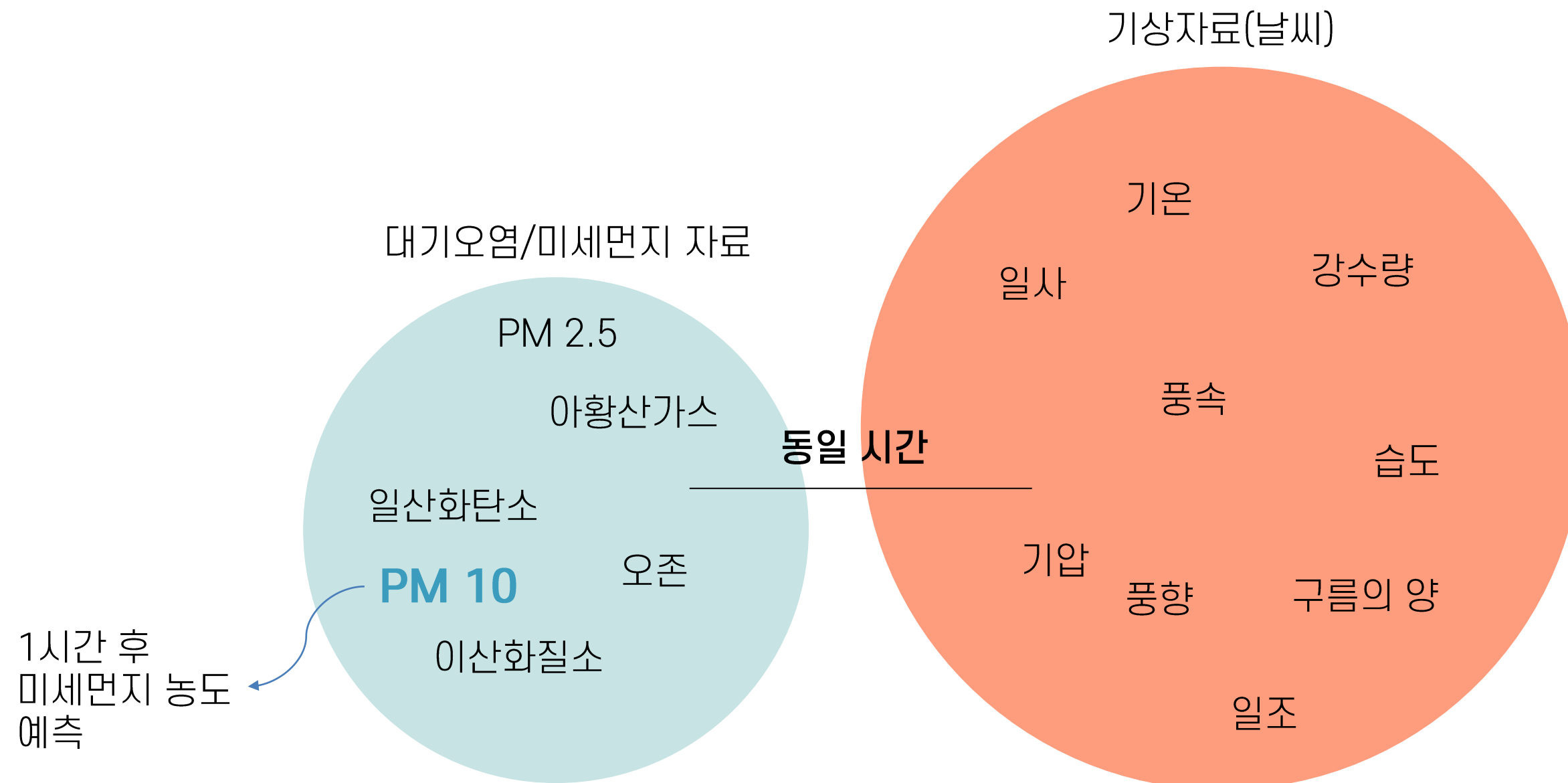
중국 등 국외 요인보다는 국내의 기상상태(날씨)에 따른 2차 미세먼지의 생성이 큰 영향을 미침

Example

실제 사례 소개

4. 미세먼지농도 예측 모델 제안

에어코리아의 대기오염/미세먼지 자료와 기상자료개방포털의 다양한 기상자료를 수집하고 예측에 있어 중요한 시간 데이터를 변수로 포함시켜 머신러닝 모델을 구현



The background features several large, overlapping, organic shapes in shades of orange, teal, and brown. These shapes are positioned in the corners and along the sides of the page, creating a modern, abstract design.

데이터 소개

Dataset

데이터출처

https://www.airkorea.or.kr/web/pastSearch?pMENU_NO=123

The screenshot shows the AirKorea website interface. The top navigation bar includes the AirKorea logo, language options (Korean, English), and search filters. The main menu on the left highlights '통계정보' (Statistics). The central content area is titled '최종확정 측정자료 조회' (Search for Final Confirmed Measurement Data). Below this, there are tabs for '측정소별 측정자료 조회' (Search by Measurement Station), '항목별 측정자료 조회' (Search by Item), and '확정자료 다운로드' (Download Confirmed Data). The '지역명검색' (Search by Area Name) section shows a search for '종로구' (Jongno-gu) with a red box highlighting the selected station '종로구' (Jongno-gu) and its address '서울 종로구 종로352길 19 종로56가동 주민센터' (Seoul Jongno-gu Jongno 352-gil 19 Jongno 56-ga-dong Juminseon). The '데이터 구분' (Data Type) section shows '시간' (Time) selected. The '조회기간' (Search Period) is set to '2021-01-01' to '2021-01-05'. The '측정자료(수치)' (Measurement Data (Value)) section displays a table of air quality data for the selected period and location.

날짜 (년-월-일)	PM ₁₀ ($\mu\text{g}/\text{m}^3$)	PM _{2.5} ($\mu\text{g}/\text{m}^3$)	오존 (ppm)	이산화질소 (ppm)	일산화탄소 (ppm)	아황산가스 (ppm)
01-01-01	24	14	0.022	0.016	0.5	0.002
01-01-02	25	14	0.018	0.020	0.6	0.002
01-01-03	27	16	0.013	0.025	0.6	0.002
01-01-04	23	13	0.011	0.027	0.6	0.003
01-01-05	24	14	0.008	0.032	0.6	0.003

Dataset

데이터출처

데이터

기상관측

지상

- 종관기상관측(ASOS)

- 방재기상관측(AWS)

- 농업기상관측(AAOS)

- 북한기상관측

- 공공기관 기상관측

- 계절관측

- 황사관측(PM10)

- 낙뢰관측

해양

고층

항공

세계기상전문(GTS)

기상위성

레이더

기상예보

수치모델

기후

응용기상

지진화산

종관기상관측(ASOS) - 자료

■ 자료설명

자료설명

종관기상관측이란 종관규모의 날씨를 파악하기 위하여 정해진 시각에 모든 관측소에서 같은 시각에 실시하는 지상관측을 말합니다. 종관규모는 일기도에 표현되어 있는 보통의 고기압이나 저기압의 공간적 크기 및 수명을 말하며, 주로 매일의 날씨 현상을 뜻합니다.

자료형태	분, 시간(매정시), 일, 월, 연	제공기간	1904년~(지점별, 요소별 다름)
제공지점	103개 * 원하는 지점이 없는 경우, 방재기상관측(AWS) 메뉴 이용	제공요소	기온, 강수, 바람, 기압, 습도, 일사, 일조, 눈, 구름, 시정, 지면상태, 지면·초상온도, 일기현상, 증발량, 현상번호
유의사항	- 1회 조회 가능 최대 기간: 분 1일, 시간 1년, 일 10년, 월·연 제한 없음(장기간 자료는 '파일셋 조회' 메뉴 이용) - 시간/분 자료에 대해 관측값의 정상 여부를 판단하는 품질검사 플래그(QC FLAG) 정보 제공 * 제공 요소: 기온, 습도, 기압, 지면온도, 풍향, 풍속, 일조 / 플래그 종류(의미): 0(정상), 1(오류), 9(결측) - 전일 자료는 당일 10시 이후 확인 가능		
비고	- 10분 또는 1시간 최다강수시각은 최다강수가 나타난 시각 시간으로, (-) 표기가 있는 경우 전날을 뜻함 - 강수량은 겨울철(11월~익년 3월) 3시간 간격으로 제공		
지침	요소별 관측방법이나 자료 산출방식에 대한 상세 설명은 [지상기상관측지침] 참조		

자료

파일셋

캘린더

OPEN-API

■ 검색조건

■ 자료형태

시간 자료

■ 기간

20210101

01

~

20210131

00

■ 지점

지도로 선택

전체

강원도

경기도

경상남도

경상북도

광주광역시

대구광역시

대전광역시

부산광역시

서울특별시

세종특별자치시

울산광역시

전체

지면·초상온도

지면온도

지면온도 QC플래그

5cm 지중온도

10cm 지중온도

20cm 지중온도

30cm 지중온도

현상번호

현상번호

기온

기온

Dataset

데이터 소개

출처	내용	장소	Row/Column	기간	파일명	사용
에어코리아	미세먼지 및 오염물질 정보	종로구 35가길 19	365일X24시간 =8760개	2021년 1월~12월 시간대 별	air_2021.csv	학습
				2022년 1월~12월 시간대 별	air_2022.csv	테스트
기상청	날씨 정보	서울시	365일X24시간 =8760개	2021년 1월~12월 시간대 별	weather_2021.csv	학습
				2022년 1월~12월 시간대 별	weather_2022.csv	테스트

Dataset

데이터 소개 – air_2021.csv

	지역	망	측정소코드	측정소명	측정일시	SO2 아황산가스 (ppm)	CO 일산화탄소 (ppm)	O3 오존 (ppm)	NO2 이산화질소 (ppm)	PM10 미세먼지 ($\mu\text{g}/\text{m}^3$)	PM25 초미세먼지 ($\mu\text{g}/\text{m}^3$)	주소
0	서울 종로구	도시대기	111123	종로구	2021100101	0.003	0.6	0.002	0.039	31	18	서울 종로구 종로35가길 19
1	서울 종로구	도시대기	111123	종로구	2021100102	0.003	0.6	0.002	0.035	27	16	서울 종로구 종로35가길 19
2	서울 종로구	도시대기	111123	종로구	2021100103	0.003	0.6	0.002	0.033	28	18	서울 종로구 종로35가길 19
3	서울 종로구	도시대기	111123	종로구	2021100104	0.003	0.6	0.002	0.03	26	16	서울 종로구 종로35가길 19
4	서울 종로구	도시대기	111123	종로구	2021100105	0.003	0.5	0.003	0.026	26	16	서울 종로구 종로35가길 19
5	서울 종로구	도시대기	111123	종로구	2021100106	0.003	0.5	0.002	0.025	22	14	서울 종로구 종로35가길 19
6	서울 종로구	도시대기	111123	종로구	2021100107	0.003	0.5	0.002	0.024	25	14	서울 종로구 종로35가길 19
7	서울 종로구	도시대기	111123	종로구	2021100108	0.003	0.5	0.005	0.023	24	14	서울 종로구 종로35가길 19
8	서울 종로구	도시대기	111123	종로구	2021100109	0.003	0.5	0.011	0.02	23	12	서울 종로구 종로35가길 19
9	서울 종로구	도시대기	111123	종로구	2021100110	0.003	0.6	0.026	0.016	40	22	서울 종로구 종로35가길 19
10	서울 종로구	도시대기	111123	종로구	2021100111	0.003	0.5	0.031	0.014	49	29	서울 종로구 종로35가길 19
11	서울 종로구	도시대기	111123	종로구	2021100112	0.003	0.5	0.035	0.017	43	24	서울 종로구 종로35가길 19
12	서울 종로구	도시대기	111123	종로구	2021100113	0.003	0.6	0.04	0.021	43	27	서울 종로구 종로35가길 19
13	서울 종로구	도시대기	111123	종로구	2021100114	0.003	0.5	0.051	0.02	42	24	서울 종로구 종로35가길 19
14	서울 종로구	도시대기	111123	종로구	2021100115	0.003	0.5	0.059	0.017	40	26	서울 종로구 종로35가길 19
15	서울 종로구	도시대기	111123	종로구	2021100116	0.003	0.4	0.064	0.014	39	20	서울 종로구 종로35가길 19
16	서울 종로구	도시대기	111123	종로구	2021100117	0.003	0.5	0.07	0.016	42	22	서울 종로구 종로35가길 19
17	서울 종로구	도시대기	111123	종로구	2021100118	0.003	0.5	0.06	0.017	44	25	서울 종로구 종로35가길 19
18	서울 종로구	도시대기	111123	종로구	2021100119	0.004	0.5	0.057	0.023	43	25	서울 종로구 종로35가길 19
19	서울 종로구	도시대기	111123	종로구	2021100120	0.005	0.6	0.059	0.022	48	28	서울 종로구 종로35가길 19
20	서울 종로구	도시대기	111123	종로구	2021100121	0.004	0.5	0.062	0.016	51	33	서울 종로구 종로35가길 19
21	서울 종로구	도시대기	111123	종로구	2021100122	0.004	0.5	0.053	0.019	45	28	서울 종로구 종로35가길 19
22	서울 종로구	도시대기	111123	종로구	2021100123	0.003	0.3	0.054	0.01	33	24	서울 종로구 종로35가길 19
23	서울 종로구	도시대기	111123	종로구	2021100124	0.003	0.3	0.046	0.015	11	6	서울 종로구 종로35가길 19

Dataset

데이터 소개 – weather_2021.csv

플래그 종류(의미): 0(정상), 1(오류), 9(결측)

지점	지점명	일시	기온(°C)	기온 QC플래그	강수량 (mm)	강수량 QC플래그	풍속(m/s)	풍속 QC플래그	풍향 (16방위)	풍향 QC플래그	습도(%)	습도 QC플래그
108	서울	2021-01-02 0:00	-3.4				2.5		320		55	
108	서울	2021-01-02 1:00	-4.1				2.3		320		39	
108	서울	2021-01-02 2:00	-5				2.7		270		42	
108	서울	2021-01-02 3:00	-5.8				1.9		270		45	
108	서울	2021-01-02 4:00	-6.3				1.7		290		45	
108	서울	2021-01-02 5:00	-6.6				1.1		320		40	
108	서울	2021-01-02 6:00	-6.6				2.7		320		35	
108	서울	2021-01-02 7:00	-7.4				1.5		290		38	
108	서울	2021-01-02 8:00	-8.3				1.3		270		41	
108	서울	2021-01-02 9:00	-7.6				2		270		39	
108	서울	2021-01-02 10:00	-6				2.3		320		35	
108	서울	2021-01-02 11:00	-5.3				3.2		290		32	
108	서울	2021-01-02 12:00	-3.2				3		270		33	
108	서울	2021-01-02 13:00	-2.3				2.5		290		31	
108	서울	2021-01-02 14:00	-1.5				4.4		250		31	
108	서울	2021-01-02 15:00	-2.3				5.2		290		33	
108	서울	2021-01-02 16:00	-1.7				3.4		290		34	
108	서울	2021-01-02 17:00	-2.3				4.5		270		35	
108	서울	2021-01-02 18:00	-3.2				2.7		250		39	
108	서울	2021-01-02 19:00	-3.8				2.8		270		50	
108	서울	2021-01-02 20:00	-4.2				2.9		320		41	
108	서울	2021-01-02 21:00	-5				2.3		290		39	
108	서울	2021-01-02 22:00	-5.5				2.8		320		42	
108	서울	2021-01-02 23:00	-6.1				1.9		270		45	

Dataset

데이터 소개 – weather_2021.csv

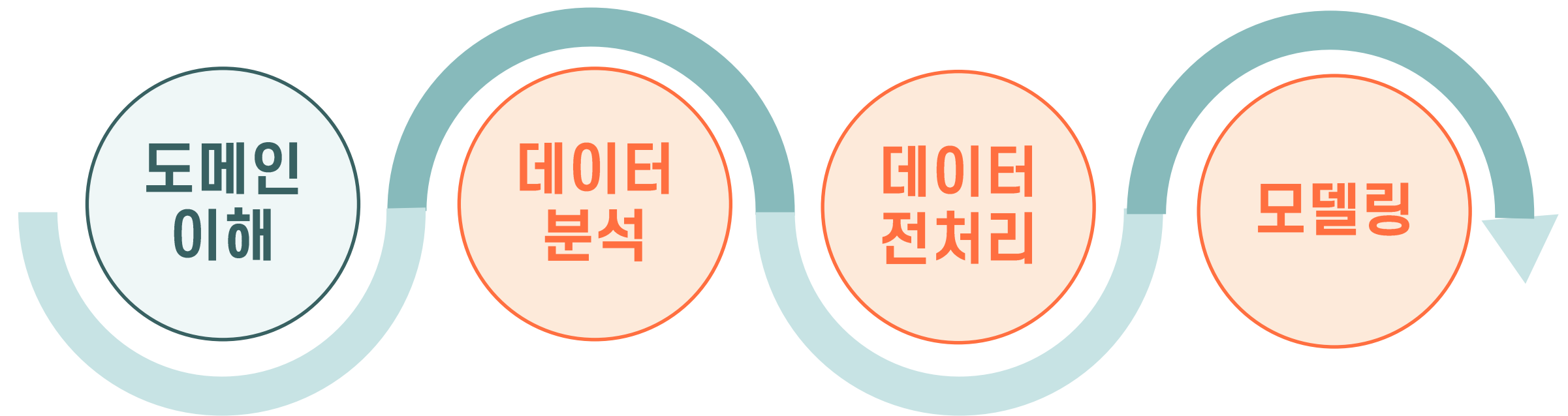
- **증기압**: 증기가 고체 또는 액체와 동적 평형 상태에 있을 때 증기의 압력 (증기가 되려는 힘)
- **이슬점 온도**: 불포화 상태의 공기가 냉각될 때, 포화 상태에 도달하여 수증기의 응결이 시작되는 온도
- **일조**: 일정한 물체나 땅의 겉면에 태양 광선이 비치는 시간 (1시간 중 비율)
- **일사(량)**: 태양으로부터 오는 태양 복사 에너지가 지표에 닿는 양 (면적당 에너지 량)
- **전운량**: 하늘을 육안으로 관측하여 전부 구름일 때 10, 구름이 덮고 있는 하늘의 비율에 따라 0~10
- **중하층운량**: 중층과 하층에 있는 구름의 분포량(중하층 구름이 날씨에 영향 주므로 따로 표기)
- **운형(운형약어)**: 구름의 종류. 약어 코드로 기재됨
- **최저운고**: 가장 낮은 구름의 높이
- **현상번호(국내식)**: 비, 소낙비, 싸락눈, 눈보라 등의 기상현상을 나타낸 코드번호
- **지면온도**: 지면 0cm 온도
- **지중온도**: 땅 속 온도

프로젝트 진행방법 안내

우리가 풀어야 할 문제는?

서울 지역의 미세먼지 데이터와 날씨 데이터를 활용,
미세먼지 예측에 관련 있는 데이터 항목으로 데이터를 구성, 전처리 하여
미세먼지 농도를 예측하는 머신러닝 모델 구현

머신러닝을 활용한 예측의 'A~Z' 경험



[Step 1] 탐색적 데이터 분석
Solution 파일 제공

[Step 2] 데이터 전처리
시계열 데이터 처리 포함

[Step 3] 머신러닝 모델링1
[Step 4] 머신러닝 모델링2

미니프로젝트 진행과정

Step 2 데이터 전처리

[illegible]

Document

참고 자료

- [논문] 미세먼지 머신러닝 예측모델에 대한 고찰
- [논문] 기상환경데이터와 머신러닝을 활용한 미세먼지 농도 예측 모델
- [논문] 시계열 분석을 활용한 서울시 미세먼지 예측
- [논문] 미세먼지 예측을 위한 기계학습 알고리즘의 적합성 평가
- [자료] 시계열 이론 학습 자료: [Forecasting: Principles and Practice\(번역본, 설명\)](#)
[Time_series lecture_notes\(영문, 수학\)](#)

참고하세요 😊

Summary

요약

- ✓ 데이터 분석
- ✓ 데이터 전처리
- ✓ 머신러닝 모델링

Thank you !

감사합니다.