

---

# 기상데이터와 안개 발생 관계 분석

---

산업인공지능학과 8조  
문정민  
한재철  
신진영

# A Table of Contents.

## 1 분석 목적

안개 판단 기능의 필요성 / 안개 판단 장치

## 2 사용할 데이터

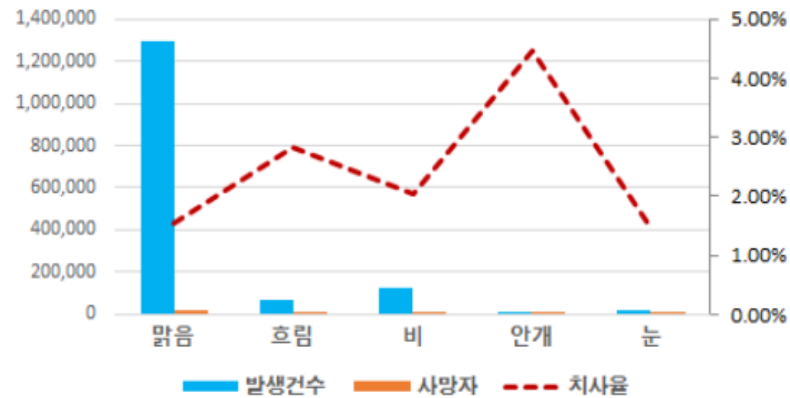
공공데이터의 활용

## 3 진행 계획

분석 방법 / 결과 활용

## ● 안개 판단의 필요성

- 국내 뿐만 아니라 미국, 유럽 등 선진국에서도 안개로 인한 교통사고가 빈번하게 발생하고 있다. 상당수가 대형사고로 확산되었음
- 시야가 확보되지 않는 안개 속 도로에서는 다른 기상 상태에 비해 사고 발생 가능성이 크며 치사율도 강설시보다 42배나 높음(교통사고 100건당 사망자수 : 안개 106명, 비 29명, 눈 25명)



기상 상태에 따른 교통사고 비교  
(출처: 도로교통공단 교통과학연구원)

## ● 안개 판단 방법 및 장치

- 안개란 매우 미세한 물방울이 대기 중에 떠 있어 수평 시정(可視距離)이 1km 미만일 때를 이야기하며, 이를 판단하기 위해서 시정계를 사용함
- 시정계는 매우 고가의 가격으로 시중에 판매되며, 대량 구매에 어려움이 있음



안개 판단 시정계 (NAS KOREA)

## 공공 데이터의 활용 : 기상청 기상자료 개방 포털

- 종관기상관측(ASOS): 종관규모의 날씨를 파악하기 위하여 정해진 시각에 모든 관측소에서 같은 시각에 실시하는 지상관측
- 기온, 습도, 기압, 지면온도, 풍향, 풍속, 일조, 시간, 장소, 날씨현상, 이슬점온도, 시정, 구름 등의 데이터를 제공

지점	지점명	일시	기온(°C)	강수량(mm)	풍속(m/s)	풍향(16방위)	습도(%)	이슬점온도(°C)	현지기압(hPa)	해면기압(hPa)	시정(10m)	현상번호(국내식)	지면온도(°C)
131	청주	2021-01-01 1:00	-7.2		0.4	0	85	-9.2	1020.6	1028.3	806	1905	-0.9
131	청주	2021-01-01 2:00	-7.8		0.9	160	84	-10	1020.6	1028.3	962	1905	-1
131	청주	2021-01-01 3:00	-8.2	0	0.3	0	84	-10.4	1021.1	1028.8	880	1905	-1.2
131	청주	2021-01-01 4:00	-8.9		0	0	84	-11.1	1021	1028.8	827	1905	-1.3
131	청주	2021-01-01 5:00	-9.1		0.8	270	83	-11.4	1020.6	1028.4	903	1905	-1.3
131	청주	2021-01-01 6:00	-9.4	0	0.2	0	85	-11.4	1020.5	1028.3	880	19	-1.3
131	청주	2021-01-01 7:00	-9.9		0.7	180	85	-11.9	1020.3	1028.1	794	19	-1.4
131	청주	2021-01-01 8:00	-9.9		0.5	180	85	-11.9	1020.7	1028.5	580	19	-1.4
131	청주	2021-01-01 9:00	-8.9		0.2	0	83	-11.2	1021	1028.8	528	19	-0.9
131	청주	2021-01-01 10:00	-7.2		0.4	0	79	-10.2	1021	1028.7	528	1905	-0.5
131	청주	2021-01-01 11:00	-5.2		0.6	200	79	-8.2	1020.7	1028.3	596	1905	-0.3
131	청주	2021-01-01 12:00	-4.1	0	0.4	0	74	-8	1019.7	1027.3	715	1905	-0.2
131	청주	2021-01-01 13:00	-3.5		1	290	73	-7.6	1019	1026.6	705	1905	0
131	청주	2021-01-01 14:00	-3.5		0.6	160	77	-6.9	1018.7	1026.3	761	1905	0
131	청주	2021-01-01 15:00	-2.5	0.2	0.4	0	78	-5.8	1018.1	1025.6	1023	1905	0

종관기상관측(ASOS) 데이터

## 분석 방법

1

### 데이터 탐색

- 결측치 파악
- 날씨데이터 시각화
- 시간에 따른 안개 발생 빈도 시각화

2

### 데이터 처리

- 데이터 정규화  
(이상 값 처리)

3

### 모델 구축

- 데이터 셋 분할  
(train : test 비율 설정)
- GBM 모델을 사용

4

### 모델 검증

- 변수 중요도 파악
- 최종 모델 선택
- 모델 성능 파악

## 결과 활용 : 안개 감지 장치 Proto type 설계

- 안개 발생 판단에 영향이 높은 변수를 측정하는 센서 선정
- 센서 데이터를 취합 처리 및 안개 여부를 확인 할 수 있는 모니터링 시스템 설계
- 모델의 정확도를 시스템을 이용해 검증

## 데이터 전처리

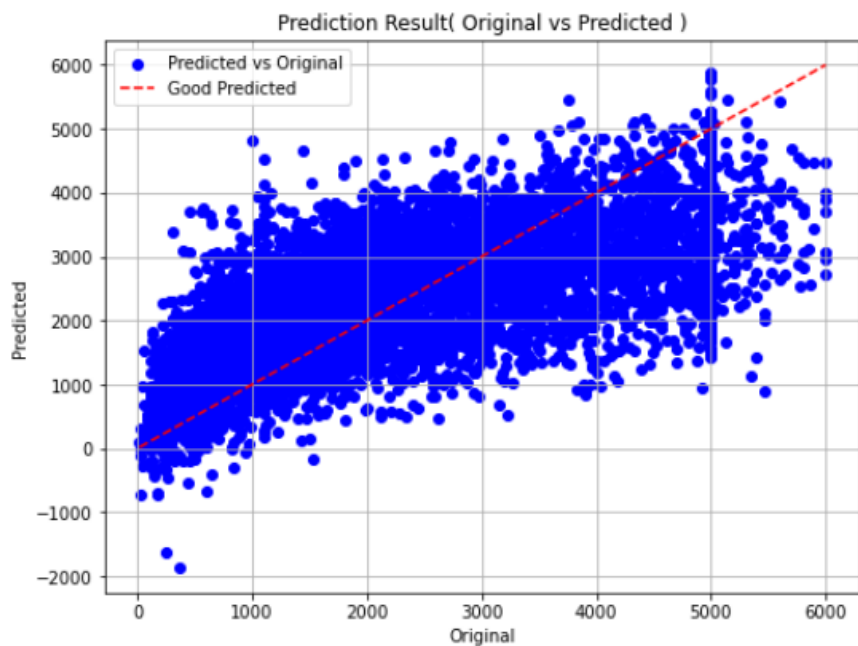
- 안개 발생과 관계없는 데이터 제거(지점, 지점명, 현상번호 등등)
- 결측치 처리(야간의 일조, 일사는 0으로 채움, 강수량 결측 데이터는 0으로 채움)
- 시간 데이터를 월, 일, 시간으로 나눔
- 2020-2023 데이터 통합

기온(°C), 강수량(mm), 풍속(m/s), 습도(%), 이슬점온도(°C), 일조(hr), 일사(MJ/m2), 시정(10m), 지면온도(°C), 월, 일, 시간

-6.6	0.0	1.3	64.0	-12.2	0.0	0.0	2042	-3.5	1	1	1
-6.9	0.0	1.5	68.0	-11.8	0.0	0.0	1871	-3.0	1	1	2
-6.5	0.0	1.4	67.0	-11.6	0.0	0.0	1825	-2.6	1	1	3
-5.7	0.0	0.8	67.0	-10.8	0.0	0.0	1962	-2.3	1	1	4
-5.4	0.0	0.8	66.0	-10.7	0.0	0.0	2000	-2.0	1	1	5
-5.0	0.0	0.3	67.0	-10.1	0.0	0.0	2000	-2.0	1	1	6
-4.9	0.0	1.1	68.0	-9.9	0.0	0.0	1760	-2.1	1	1	7
-4.5	0.0	0.9	69.0	-9.3	0.0	0.0	1658	-1.8	1	1	8

## 모델 학습 : 선형 회귀

- Z-Score 정규화
- 선형 회귀 모델 학습
- 결정 계수 값이 0.3405로 낮게 나옴

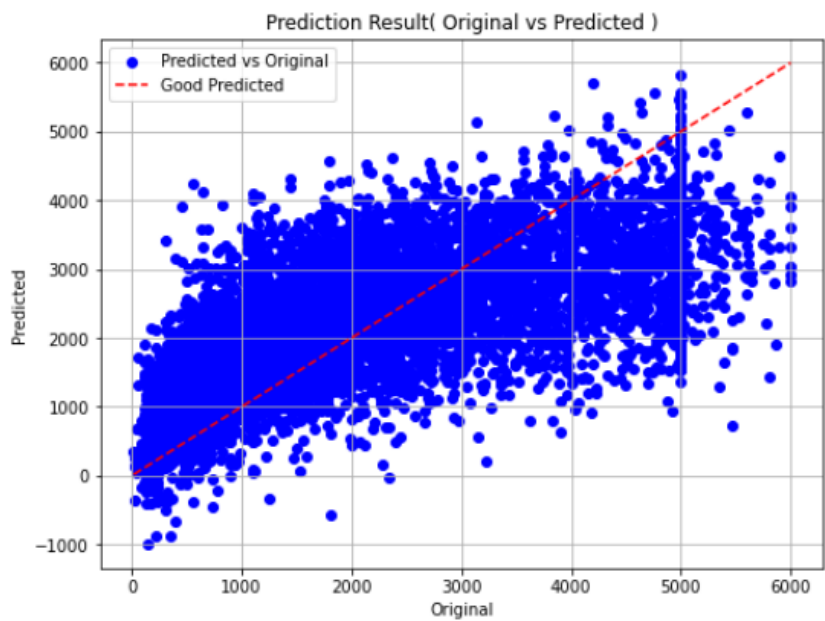


결정 계수 ( $R^2$ ): 0.3405  
평균 제곱 오차 (MSE): 1525203.4322



## 모델 학습 : 다항 회귀

- 선형 회귀 모델의 성능 개선을 위해 사용
- 다항 회귀의 차수를 올렸을 경우 성능 개선 됨
- 과적합 방지를 위해 Ridge 사용

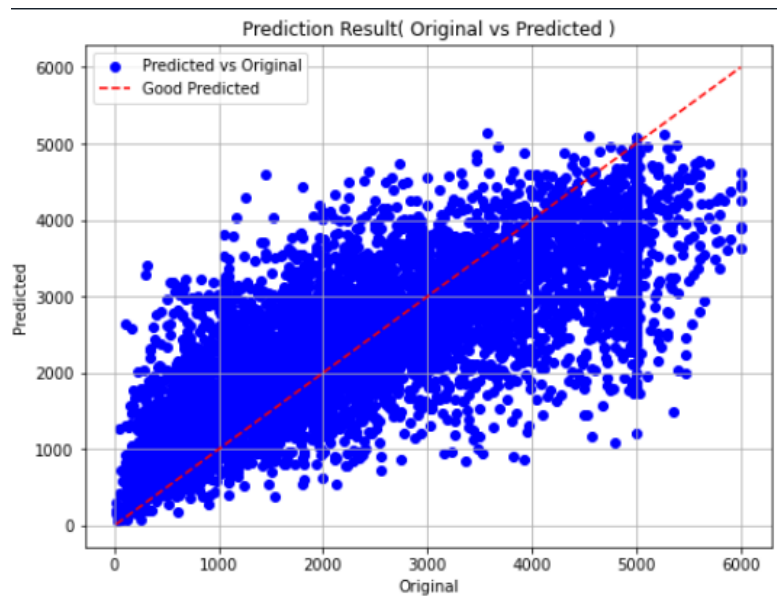


=== 2차 다항 회귀 ===  
결정 계수 ( $R^2$ ): 0.4368  
평균 제곱 오차 (MSE): 1302373.8943

=== 3차 다항 회귀 ===  
결정 계수 ( $R^2$ ): 0.4840  
평균 제곱 오차 (MSE): 1193262.3492

## 모델 학습 : 랜덤 포레스트

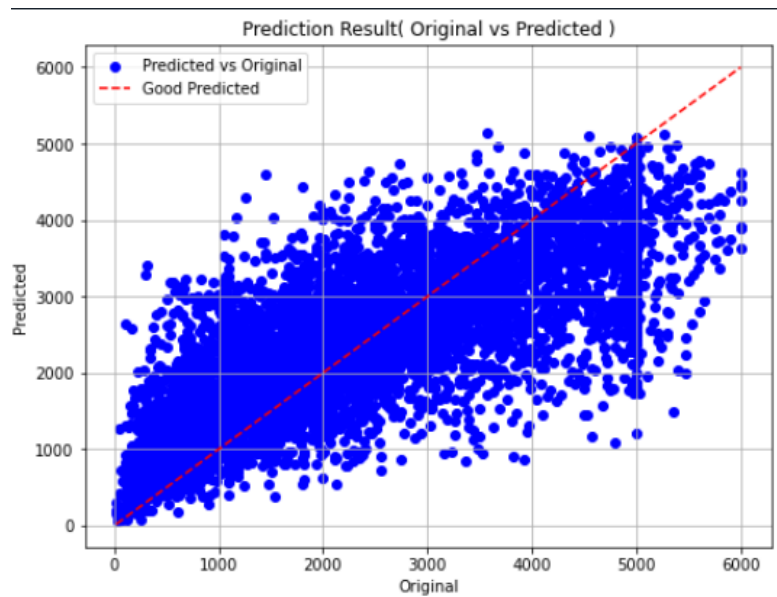
- 선형 회귀 모델의 성능 개선을 위해 사용
- 결정 계수 값이 0.6398로 가장 좋은 성능을 보임
- 특징 중요도 분석을 했을 때 습도, 기온이 가장 큰 영향을 미치는 것으로 분석 됨



```
시간: 0.0599  
일: 0.1175  
월: 0.0957  
일조(hr): 0.0177  
일사(MJ/m2): 0.0431  
지면온도(°C): 0.0881  
습도(%): 0.2985  
풍속(m/s): 0.0826  
기온(°C): 0.1970  
결정 계수 (R²): 0.6398  
평균 제곱 오차 (MSE): 833081.7617
```

## 모델 학습 : 랜덤 포레스트

- 선형 회귀 모델의 성능 개선을 위해 사용
- 결정 계수 값이 0.6398로 가장 좋은 성능을 보임
- 특징 중요도 분석을 했을 때 습도, 기온이 가장 큰 영향을 미치는 것으로 분석 됨



```
시간: 0.0599  
일: 0.1175  
월: 0.0957  
일조(hr): 0.0177  
일사(MJ/m2): 0.0431  
지면온도(°C): 0.0881  
습도(%): 0.2985  
풍속(m/s): 0.0826  
기온(°C): 0.1970  
결정 계수 (R²): 0.6398  
평균 제곱 오차 (MSE): 833081.7617
```