



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

기후변화 대응 IoT 응용 인공지능  
모델링을 통한 스마트공원 구현

- 서울숲공원과 소래생태공원 사례를 중심으로 -

2023년 2월

서울대학교 대학원

협동과정 조경학 전공

스마트시티 글로벌 융합 전공

권태경

기후변화 대응 IoT 응용 인공지능  
모델링을 통한 스마트공원 구현  
- 서울숲공원과 소래생태공원 사례를 중심으로 -

지도교수 조 경 진

이 논문을 공학박사 학위논문으로 제출함  
2023년 1월

서울대학교 대학원  
협동과정 조경학  
스마트시티 글로벌 융합  
권 태 경

권태경의 공학박사 학위논문을 인준함  
2023년 2월

위 원 장           손 용 훈           (인)

부위원장           강 준 석           (인)

위 원           박 창 우           (인)

위 원           신 승 우           (인)

위 원           조 경 진           (인)

## 국문 초록

스마트공원은 환경 관련 디지털 첨단기술을 사용하여 일련의 가치를 달성하는 공원으로 정의된 새로운 개념이다. 즉, 스마트공원의 목적은 기존의 도시공원에 ICT 기술 등을 접목하여, 효과적인 운영·유지를 도모하고 환경 개선을 통한 사용자의 편의를 증진하는 데 있다. 본 연구에서는 기후변화와 관련, 대기환경을 개선하는 공간으로 스마트공원의 기능을 상정하고, 이를 위해 IoT 기술 기반 인공지능 모델을 비교하여, 향후 적정하게 적용 가능한 방향을 모색하고자 하였다.

한편 현 시대는 탄소저감, 이상기후, 지구온난화 등의 문제로 이른바 기후변화 시대로 지칭되고 인식되고 있다. 이에 따라 그 원인으로 지목되는 탄소 배출량의 저감이 중요한 과제로 인식되고 있으며, 전 세계적으로 탄소 저감을 위한 다양한 시도들이 이어지고 있다. 기후 위기 시대 도시공원의 여러 가지 기능 중에서, 기후변화대응 중 폭염방재 및 이산화탄소와 미세먼지 저감 기능은 매우 중요한 요소로 자리잡고 있다. 따라서, 도시공원의 기능 중 기후변화대응에 대한 부분을 고찰하기 위하여 선행연구를 살펴보고, 효과를 분석하기 위하여 실제 IoT 장비를 제작하여 설치하고, 데이터를 특정 기간에 걸쳐 취득하였다.

본 연구에서는 다음과 같은 두 가지 연구목적을 가지고 실험과 분석을 하여 결론을 도출하였다.

첫째, 서울숲 공원의 기온 및 공기질 등의 환경 요소와 보행량의 데이터를 약 9개월의 시계열 데이터로 수집하고, 보행량을 목표(target)로 한 후, 각 환경 요소의 상관관계를 인공지능을 활용하여 다수의 모델링 알고리즘으로 회귀분석을 수행하였다. 그리고 모델링 방법 중 가장 우수한 성능을 보인 Voting Ensemble을 통해서 분석된 보행량과 상관관계가 높은 환경 요소와 그의 패턴을

살펴보았다. 그 결과, 온도(temperature), 시간(hour), 습도(humidity), 월(month), 평일(working day), 미세먼지(pm)의 6개 환경 데이터 인자의 순서로 특징중요도(feature importance)가 보행량 예측에 가장 큰 영향을 주는 인자로 측정이 되었다. 평가지표로는 결정계수인  $R^2$ 이 1에 가까울수록 인공지능 모델이 우수한 성능을 보인다고 평가되는데, 가장 모델 평가지수가 높은 것으로 평가된 Voting Ensemble의 경우 결정계수인  $R^2$ 이 0.72로 가장 우수한 것을 알 수 있으며, XGBoostRegressor가 0.70, LightGBM이 0.69, RandomForestRegressor가 0.53 순으로 도출되어 인공지능 학습모델에 대한 평가를 통해서 Voting Ensemble 모델을 최적의 알고리즘으로 결론을 내릴 수 있었다. 이에 따라, 향후 스마트공원의 관리기능으로서 환경 데이터를 활용한 보행량의 예측에, 본 연구에서 확인된 Voting Ensemble 머신러닝 모델을 사용하여 정량적인 예측과 응용에 사용할 수 있을 것이다.

둘째, 기후변화 위기 방재역할로서, 인천 소래생태공원의 갈대숲과 염습지에서 폭염기 환경 데이터를 수집 이용, 도심지인 포구사거리에서 수집한 온·습도 등의 환경 데이터와 비교하면서, 이산화탄소의 발생량에 영향을 주는 요소들과 사이트별 차이를 분석하였다. 그 결과 도로 부인 포구 사거리의 이산화탄소 발생량은 소래생태공원 염습지에 비해서 평균 40ppm, 최고 60ppm 정도 높게 측정되었고, 기온은 포구 사거리가 시간과 날짜와 관계없이 일정하게 약 1.5 °C 정도 높게 유지되었다. 초미세먼지(pm2.5)의 비교에서, 도심부에 위치한 포구 사거리는 매일 평균 갈대숲보다는 약  $3\mu\text{g}/\text{m}^3$  정도 높게 측정되었고, 염습지 대비  $4\mu\text{g}/\text{m}^3$  높게 나타났다. 미세먼지의 비교에서는 도심부에 위치한 포구 사거리는 매일 평균 갈대숲보다는 약  $3.5\mu\text{g}/\text{m}^3$  정도 높게 유지하였고, 염습지 대비는  $4\mu\text{g}/\text{m}^3$ 의 차이를 유지하였다.

이어서 기후 변화 대응을 위하여 환경 데이터와 이산화탄소 발생의 상관관계를 분석하기 위해, Auto ML 머신러닝 모델을 이용하였다. 인공지능분석을 통한 머신러닝 알고리즘 모델별 예측 성능을 비교하고 그중에 가장 우수한 결과를 보인 Stacking Ensemble을 통해서 분석된 상관관계가 높은 환경 요소를 파악하고, 그의 패턴을 염습지, 갈대숲, 그리고 포구 사거리로 나누어 Regression 모델링의 비교 분석을 수행하였다.

그 결과 염습지에서는 온도(sr\_salty\_temp), 돌풍(sr\_salt\_wind\_gust), 습도(sr\_salty\_humi), 기압 (sr\_salty\_pressure), 초미세먼지(sr\_salty\_pm2.5), 시간(hour)의 6개 인자의 순서로 이산화탄소 발생량에 가장 큰 영향을 주는 것을 인지하고, 적용된 Stacking Ensemble의 결정계수인  $R^2$  이 0.90으로 가장 우수한 것을 알 수 있으며, XGBoostRegressor 이 0.88, LightGBM 이 0.88, RandomForestRegressor 가 0.66의 순으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘에 대한 결론을 내릴 수 있었다.

갈대숲에서는 습도(sr\_galdae\_humi), 돌풍(sr\_galdae\_Wind Gust), 기압 (sr\_galdae\_pressure), 풍향(sr\_galdae\_Wind Direction), 시간(hour), 온도 (sr\_galdae\_temp)의 6개 인자가 이산화탄소 발생량에 가장 큰 영향을 주는 것을 인지하고, 적용된 Stacking Ensemble의  $R^2$  이 0.918로 가장 우수한 것을 알 수 있으며, XGBoostRegressor이 0.915, LightGBM이 0.89, RandomForestRegressor가 0.80의 순으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘에 대한 결론을 내릴 수 있었다.

도심부인 포구 사거리에서는 도심부내의 환경센서 설치의 제약으로, 일부 갈대숲의 데이터를 차용하여 사용한 결과, 갈대숲 풍향(sr\_galdae\_wind direction), 온도(sr\_pg4\_temp), 포구 사거리 미세먼지(sr\_pg4\_pm10), 갈대숲 태양조도(sr\_galdae\_solar), 포

구 사거리 습도(sr\_pg4\_humi), 갈대숲 풍속(sr\_galdae\_wind speed)의 순서로 6개 인자가 이산화탄소 발생량에 가장 큰 영향을 주는 것으로 확인되었다. 적용된 Stacking Ensemble의  $R^2$ 이 0.85로 가장 우수한 것을 알 수 있으며, XGBoostRegressor 이 0.83, LightGBM이 0.79, RandomForestRegressor가 0.52의 순으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘을 구할 수 있었다.

따라서 향후 스마트공원의 방재기능으로서 환경데이터를 활용한 이산화탄소 발생과 저감 예측에 연구에서 확인된 최적의 Stacking Ensemble 머신러닝 모델을 사용하여 정량적인 예측과 응용에 사용할 수 있을 것으로 기대된다.

본 연구는 경제적인 IoT 환경 데이터 센싱 개발 방법을 제안하고 최적의 인공지능 머신러닝 모델을 설계하여, 기존의 공원을 스마트공원화하는 데 있어 그 수월성을 제고하는 데 기여하고자 하였다. 이에 따른 기대 효과는 다음과 같다. 첫째, 스마트공원의 관리기능으로서 환경변수에 따른 방문자 보행량에 대한 정확한 예측을 통해서 최적의 서비스를 제공할 수 있을 것이다. 둘째, 전 세계적으로 해결해야만 하는 기후변화 위기 대응에 스마트공원의 탄소저감, 폭염 저감과 같은 방재 기능을 강화함으로써 도시공원의 새로운 가치를 부여할 것으로 기대된다.

**주요어 :** 스마트공원, IoT센싱, 환경데이터, 인공지능, 기후변화 적응

**학 번 :** 2019-31781

# 목 차

I. 서론 .....	1
1. 연구의 배경 .....	1
2. 연구의 목적 .....	3
3. 연구의 내용 및 방법 .....	6
II. 이론적 배경 .....	9
1. 스마트공원 정의 .....	9
2. 스마트공원 사례 .....	15
2.1 한국의 스마트공원 .....	15
2.2 해외의 스마트공원 .....	17
2.3 스마트공원 관련 인터뷰 .....	19
3 스마트공원과 기후변화 대응 .....	25
3.1 스마트공원의 기후변화 대응 사례 .....	25
3.2 공원의 그린 인프라를 통한 기후변화 대응 사례 .....	27
4 스마트공원 구현을 위한 IoT, 인공지능 기술 .....	31
4.1 IoT 환경 데이터 장비 연구 동향 .....	31
4.2 기후변화 대응 정량평가 환경 데이터 수집용 센서 .....	35
4.3 인공지능 선형회귀분석을 통한 데이터 분석 .....	39
4.4 인공지능 모델링 분석 고찰 .....	41
III. 서울숲 공원을 대상으로 한 스마트공원 구현 ..	46
1. 서울숲 공원: IoT와 인공지능을 이용한 공원 관리계획 ..	46
1.1 IoT 환경 데이터 수집 장비 개발과 고려사항 .....	54
1.2 인공지능을 통한 환경 데이터 보행량 영향 요소 평가 .....	76
1.2.1 Auto ML을 사용한 전체 연구 flow .....	78
1.2.2 머신러닝 알고리즘의 종류 .....	80
1.2.3 데이터 전처리 방법 .....	88

1.2.4 모델 평가 방법	92
<b>2. 인공지능 모델링을 통한 서울숲 스마트공원 구현</b>	<b>94</b>
2.1 인공지능을 이용한 공기질과 보행량의 상관관계	94
2.2 머신러닝 분석	110
2.3 소결	121
<b>IV. 소래생태공원을 대상으로 한 스마트공원 구현</b>	<b>122</b>
<b>1. 소래생태공원: 인공지능을 이용한 기후변화 대응</b>	<b>122</b>
1.1 AWS를 통한 스마트공원 환경 데이터 수집	126
1.1.1 자동기상관측소(AWS) 정의	126
1.1.2. 스마트공원을 위한 환경 데이터 수집 장비 구현	127
1.2 인공지능을 통한 탄소 발생 요소 분석 방안	134
<b>2. 인공지능 모델링을 통한 소래 생태 스마트공원 구현</b>	<b>140</b>
2.1 인공지능을 이용한 습지와 CO <sub>2</sub> 상관관계	140
2.2 머신러닝 분석	161
2.3 소결	188
<b>V. 결론</b>	<b>190</b>
<b>1. 연구의 요약</b>	<b>190</b>
1.1 연구의 요약	190
1.2 연구의 한계 및 의의	194
<b>참고문헌</b>	<b>196</b>
<b>부록</b>	<b>204</b>
<b>Abstract</b>	<b>207</b>

## 표 목 차

표 1	버추얼 리얼리티 와 네츄럴 리얼리티 비교	13
표 2	스마트공원의 적용 사례와 클래스 분류	15
표 3	스마트공원을 통해 해결해야 할 일반공원의 문제점과 해결방안	18
표 4	서울숲 관리 서울그린트러스트 사무총장 인터뷰 정리	23
표 5	스마트공원 정책 입안 서울시 시의회 환경 입법조사관 인터뷰 정리	24
표 6	스마트공원과 기후변화 대응 관련 선행연구	27
표 7	공원의 온도 저감 효과에 관한 선행연구	32
표 8	IoT기반 실내외 공기질 모니터링시스템에 관한 선행연구	35
표 9	다양한 온도 측정 장치	39
표 10	선형회귀분석을 통한 데이터 분석에 관한 선행연구	42
표 11	서울숲 IoT 장비 모니터링 수집내용	50
표 12	서울숲공원의 IoT를 사용한 초기 환경 데이터 수집 후보 대상지	51
표 13	서울숲 보행량 계수기와 보행량 계수기 설치 위치	53
표 14	People Counter 제품 사양서	53
표 15	스마트공원 모니터링 계획 세부 항목	56
표 16	임베디드 시스템과 마이컴 비교 IP STB(좌), ESP-WROOM32(우)	58
표 17	연구에 사용된 태양광 패널들	66
표 18	연구에 사용된 태양광 전원 사용 방법	66
표 19	연구에 사용된 와이파이 와 모뎀 서비스 종류	68
표 20	온습도 센서 DHT 11 과 DHT 22비교	69
표 21	온습도 센서 DHT22 (왼쪽) , 일산화탄소 센서 MQ7 (오른쪽)	70
표 22	MQ7 ; 센서 특징	71
표 23	미세먼지 예보등급	72
표 24	서울숲에 실제 설치된 IoT 장비 사례	74
표 25	설치 사례에 다른 IoT 제작 형태	75
표 26	서울숲에 실제 설치 적용된 IoT 센서 장비	76
표 27	스마트공원 측정용 IoT 장비와 데이터 로거 장비의 장단점 비교	77
표 28	스케일링의 종류	93
표 29	서울숲 데이터 유의성 검증 테스트 결과	101

표 30	머신 러닝에 사용된 전처리 모듈과 엔진, 본 처리에 사용된 알고리즘과 엔진	113
표 31	각 학습모델당 적용된 하이퍼 파라미터	119
표 32	사용된 학습모델에 대한 평가	121
표 33	소래생태 습지 공원 대상지 AWS IoT 위치(좌) 및 설치현황(우)	126
표 34	소래 습지 생태공원 AWS IoT장비 모니터링 수집내용	126
표 35	설치된 스마트공원 적용 AWS 사례	131
표 36	소래생태공원 연구에 사용된 기상관측장비, 개발된 원격모니터링 AWS	131
표 37	Automatic Weather Station 외관과 실내용 LCD 디스플레이	132
표 38	AWS용 스크린 디스플레이와 표시내용	132
표 39	AWS 사양서	133
표 40	전원, WiFi없는 지역에 설치 방법	135
표 41	소래 습지 공원 유의성 test 아노바 테스트 결과	148
표 42	염습지 CO2 Regression 실험에 적용된 ML 알고리즘과 모듈	164
표 43	하이퍼 파라미터 정리	169
표 44	알고리즘별 평가 결과	171
표 45	갈대숲 CO2 Regression 실험에 적용된 ML 알고리즘과 모듈	173
표 46	풍향각도 일람표	177
표 47	하이퍼 파라미터 정리	179
표 48	알고리즘별 평가 결과	179
표 49	포구사거리 CO2Regression 실험에 적용된 ML 알고리즘과 모듈	181
표 50	하이퍼 파라미터 정리	187
표 51	알고리즘별 평가 결과	188
표 52	서울숲 관리 서울그린트러스트 사무총장 인터뷰 원본	204
표 53	스마트공원 정책 입안 서울시 시의회 환경 입법조사관 인터뷰 원본	205

## 그림 목 차

그림 1 연구의 전체흐름도	9
그림 2 스마트공원의 탄생 배경	14
그림 3 국내 스마트공원에서 활용되는 다양한 스마트공원 기술 사례	17
그림 4 국내 최초 스마트공원 :IoT See Park	18
그림 5 스마트 파크에서 사용되는 다양한 기술 사례	19
그림 6 세계 최초 AI 공원 베이징 하이덴 공원에 스마트 파크 시스템 적용	19
그림 7 T&D TR-71wF	37
그림 8 현장에 적용된 데이터 로거 사례	37
그림 9 열화상 카메라 E6	37
그림 10 공공데이터와 합성하여 전처리 된 데이터 출력	55
그림 11 HD grade, PTZ IP camera (Solar power)	60
그림 12 마이크로SD 메모리	61
그림 13 2020년 10월 25일 ~ 2021년 5월20일 MicroSD를 이용한 환경 데이터 수집	62
그림 14 Thingspeak 홈페이지	64
그림 15 Web, App 방식 frontend 비교	64
그림 16 태양광으로 충전 테스트 중인 장비	67
그림 17 미세먼지 센서와 사양	72
그림 18 IoT 스마트공원 공기질 센서의 시스템 블록도	73
그림 19 제작된 IoT 디바이스 제품 1	73
그림 20 제작된 IoT 디바이스 제품2	74
그림 21 인공지능, Machine learning and Deep learning 카테고리	79
그림 22 서울숲 공기질과 방문자 상관관계분석 Machine Learning 전체 학습도	80
그림 23 Auto ML (머신러닝) 파이프라인	82
그림 24 Decision Tree	85
그림 25 Ensemble Algorithm	87
그림 26 Stacking Ensemble	89
그림 27 서울시 평균 데이터와 IoT장비를 활용하여 측정한 온도 트렌드 비교	97
그림 28 서울시 평균 데이터와 IoT장비를 활용하여 측정한 습도 트렌드 비교	97
그림 29 파이선으로 읽어들이는 데이터의 헤드 정보	98

그림 30	파이선으로 읽어들이는 온도 특성	98
그림 31	파이선으로 읽어들이는 데이터 특성	99
그림 32	Null 값을 갖는 데이터 수	99
그림 33	데이터 중 Null 값 분포	100
그림 34	파이선 분석을 통한 전체 연간, 월간, 보행량 수 트렌드 분석	102
그림 35	파이선 분석을 통한 전체 일간, 시간당 보행량 수 트렌드 분석	102
그림 36	2021년 5월 ~12월 10일 성동구 코로나 일간 확진자 수 추이	103
그림 37	연간 단위 매월 평균 시간당 평균 보행량 수	104
그림 38	월간 단위 매일 시간당 평균 보행량 수	104
그림 39	1일 보행량 수 박스 플롯	105
그림 40	서울숲공원 계절별 평균 시간당 보행량 상관관계분석 변화량	106
그림 41	서울숲공원 24시간 평균 보행량 상관관계분석 변화량	107
그림 42	서울숲공원 평일 보행량 상관관계분석 변화량	107
그림 43	서울숲공원 수치, 범주 데이터 조건별 보행량 상관관계분석 변화량	109
그림 44	히트 맵을 통한 서울숲 보행량 상관관계분석	110
그림 45	선형회귀를 사용한 보행량 상관관계분석	111
그림 46	선형회귀를 사용한 보행량 상관관계분석	111
그림 47	Voting Algorithm 다이어그램	114
그림 48	Top6 features by their importance	115
그림 49	Voting Ensemble 머신러닝으로 적용된 온도와 방문자와의 상관관계	115
그림 50	Voting Ensemble 머신러닝으로 적용된 시간대와 방문자와의 상관관계	116
그림 51	Voting Ensemble 머신러닝으로 적용된 습도와 방문자와의 상관관계	117
그림 52	Voting Ensemble 머신러닝으로 적용된 월간 대와 방문자와의 상관관계	117
그림 53	Voting Ensemble 머신러닝으로 적용된 미세먼지와 방문자와의 상관관계	118
그림 54	예측값과 실제값의 비교 평가 결과	121
그림 55	알고리즘 잔차 검토 결과	122
그림 56	AWS 사례 ; 출처 : Campbell Scientific site	128
그림 57	미세먼지, 이산화탄소 측정용 센서	134
그림 58	AWS 장비로 수집된 공기질 web display 예	135
그림 59	소래포구사거리 Auto ML Regression수행을 위한 전체 플로우 차트	139
그림 60	소래 공원 갈대숲 Auto ML Regression수행을 위한 전체 플로우 차트	140

그림 61	소래 공원 염습지 Auto ML Regression수행을 위한 전체 플로우 차트	140
그림 62	소래공원에 설치된 AWS 센서의 위치와 인천 환경원 논현지점 센서 위치 비교	143
그림 63	인천 환경원 논현지점의 PM10 데이터와 포구 사거리 환경 데이터 비교	143
그림 64	인천 환경원 논현지점의 PM2.5 데이터와 포구 사거리환경 데이터 비교	144
그림 65	파이선으로 읽어들이는 데이터의 헤드 정보	144
그림 66	파이선으로 읽어들이는 온도, 데이터 특성	145
그림 67	파이선으로 읽어들이는 이산화탄소 데이터 특성	145
그림 68	파이선으로 읽어들이는 전체 데이터 특성	146
그림 69	데이터 중 Null 값 분포	147
그림 70	소래생태공원 각 사이트의 일자별, 시간별 이산화탄소 변화량	150
그림 71	사이트별, 이산화탄소 일별, 시간별 비교 차트	151
그림 72	소래생태공원 각 사이트의 일자별, 시간별 온도 변화량	151
그림 73	사이트별, 온도 일별, 시간별 비교 차트	152
그림 74	소래생태공원 각 사이트의 일자별, 시간별 초미세먼지(PM2.5) 습도 변화량	153
그림 75	사이트별, 초미세먼지 일별, 시간별 비교 차트	154
그림 76	소래생태공원 각 사이트의 일자별, 시간별 미세먼지(PM10) 변화량	155
그림 77	사이트별, 미세먼지 일별, 시간별 비교 차트	155
그림 78	소래생태공원 각 사이트의 일자별, 시간별 습도 변화량	156
그림 79	사이트별, 습도 일별, 시간별 비교 차트	156
그림 80	소래생태공원 각 사이트의 일자별 이산화탄소 변화량	157
그림 81	소래포구 도로에서의 상관관계분석	158
그림 82	히트 맵을 통한 소래포구 환경 요소 상관관계분석	158
그림 83	포구 4거리 와 갈대 지역의 합성 데이터 상관관계	159
그림 84	염습지와 합성 데이터 상관관계	160
그림 85	갈대숲 합성 데이터 상관관계	161
그림 86	이산화탄소 QQ 분포 분석	162
그림 87	분석에 사용된 Stacking Algorithm 다이어그램	164
그림 88	Top6 features by their importance	165
그림 89	Stacking Ensemble 머신러닝으로 적용된 온도와 CO2와의 상관관계	165
그림 90	Stacking Ensemble 머신러닝으로 적용된 돌풍과 CO2와의 상관관계	

166	
그림 91	Stacking Ensemble 머신러닝으로 적용된 습도와 CO2와의 상관관계 167
그림 92	Stacking Ensemble 머신러닝으로 적용된 기압과 CO2와의 상관관계 167
그림 93	Stacking Ensemble 머신러닝으로 적용된 초미세먼지와 CO2와의 상관관계 168
그림 94	Stacking Ensemble 머신러닝으로 적용된 시간대와 CO2와의 상관관계 168
그림 95	예측값과 실제값의 비교 평가 결과 .....171
그림 96	알고리즘 잔차 검토 결과 .....172
그림 97	분석에 사용된 Voting Algorithm 다이어그램 .....173
그림 98	Top6 features by their importance .....174
그림 99	Stacking Ensemble 머신러닝으로 적용된 습도와 CO2와의 상관관계 174
그림 100	Stacking Ensemble 머신러닝으로 적용된 돌풍과 CO2와의 상관관계 175
그림 101	Stacking Ensemble 머신러닝으로 적용된 시간대와 CO2와의 상관관계 175
그림 102	Stacking Ensemble 머신러닝으로 적용된 풍향과 CO2와의 상관관계 176
그림 103	Stacking Ensemble 머신러닝으로 적용된 기압과 CO2와의 상관관계 177
그림 104	Stacking Ensemble 머신러닝으로 적용된 온도와 CO2와의 상관관계 178
그림 105	예측값과 실제값의 비교 평가 결과 .....180
그림 106	알고리즘 잔차 검토 결과 .....180
그림 107	분석에 사용된 Voting Algorithm 다이어그램 .....182
그림 108	Top6 features by their importance .....182
그림 109	Stacking Ensemble 머신러닝으로 적용된 풍향과 CO2와의 상관관계 183
그림 110	Stacking Ensemble 머신러닝으로 적용된 기온과 CO2와의 상관관계 183
그림 111	Stacking Ensemble 머신러닝으로 적용된 미세먼지와 CO2와의 상관관계

184	
그림 112	Stacking Ensemble 머신러닝으로 적용된 조도와 CO2와의 상관관계
185	
그림 113	Stacking Ensemble 머신러닝으로 적용된 습도와 CO2와의 상관관계
185	
그림 114	Stacking Ensemble 머신러닝으로 적용된 풍속과 CO2와의 상관관계
186	
그림 115	예측값과 실제값의 비교 평가 결과 .....189
그림 116	알고리즘 잔차 검토 결과 .....189

# I. 서론

## 1. 연구의 배경

스마트공원은 환경 관련 디지털 첨단기술을 사용하여 일련의 가치를 달성하는 공원으로 정의된 새로운 개념이다<sup>1)</sup>. 즉, 스마트공원의 목적은 기존의 도시공원에 ICT 기술 등을 접목하여, 효과적인 운영 및 유지보수 및 사용자의 편의를 증진 시키는 데 있다. 이것은 버스 및 정류장에 BIS(Bus Information System) 시스템을 탑재하여 자신이 타고자 하는 버스의 도착 예정 시간, 좌석 수, 도착지까지의 예상 소요 시간 등을 확인하여 교통 이용의 편리함을 제고하는 것과 유사하다<sup>2)</sup>. 사람들은 기존과 동일한 교통수단인 버스를 이용하지만, 정량화된 수치와 데이터를 이용한 정보를 알 수 있으므로, 내가 타려는 버스가 몇 분 후에 도착하는지, 좌석은 충분한지, 또 내가 가려는 위치까지 이동하는 데는 시간을 얼마나 걸릴지 미리 알 수 있어 매우 편리하고 불편하지 않다.

이 같은 맥락에서 기존 도시공원에도 ICT 기술을 접목하기 위한 시도가 많이 나타나고 있다. 특히 4차산업의 발전과 더불어, 사물인터넷 (IoT), 빅데이터, 인공지능(AI), 무인 자동차 및 드론 등의 첨단기술이 공원에 접목되어, 공원 이용자와 시설환경이 상호작용을 할 수 있도록 한다<sup>3)</sup>.

스마트도시란 ‘스마트도시 구현 및 산업진흥 등에 법률’ 에서도 제2조 1항에 의하면<sup>4)</sup>, “도시의 경쟁력과 삶의 질이 향상을 위해 건설, 정보, 통신 기술 등을 융복합해 건설된 도시 기반 시설을 바탕으로 다양한 도시서비스를 제공하는 지속 가능한 도시” 로 정의하고 있다. 이 같은 차원에서 스마트공원 또한 최첨단기술을 융복합하여 새로운 차원의 사용자 중심 서비스를 제공하는 것이 가능하다.

최근 코로나19 바이러스 확산에 따라 공원의 사용량이 많이 증가하였다. 일상에

---

1) Loukaitou-Sideris, et al. (2018), *Smart Parks: A Toolkit*, UCLA: Luskin Center, USA.

2) 이영우(2013), 「버스정보시스템(BIS) 수집자료를 이용한 경로통행시간 추정」, 『대한토목학회』, 33(3), 1115 - 1122.

3) 김태경 외(2018), 「4차 산업혁명 시대의 스마트시티 전략」, 『정책연구』, 경기연구원, (), 1-235.

4) 김성원, 이환수(2020), 「스마트도시 조성 및 산업진흥 등에 관한 법률」의 개선방안에 관한 연구, 『법학논총』, 44(2), 59-92.

서 당연시 이용하고 방문했던 카페나 운동을 하는 헬스센터 등도 사람들이 모이는 장소로 위험 요소가 있는 공간으로 인식 되고 있다. 이러한 공간의 제약을 해결하고 이용자의 불안감을 낮추고자 도시의 오픈스페이스 중 하나인 공원으로 사람들이 많이 몰리고 있다<sup>5)</sup>. 도시공원은 일반적으로 도시지역에서 도시 자연경관을 보호하고 시민의 건강·휴양 및 정서 생활을 향상하게 시키는 데 이바지하기 위하여 설치 또는 지정된 공원을 말한다. 도시공원은 시민들에게 쾌적한 도시 환경을 구현하고 후생과 공공의 복리를 증진시키기 위한 역할을 하여 도시공원에 대한 수요가 높아지고 있다<sup>6)</sup>. 도시민의 삶의 질에 관한 관심은 여가에 대한 인식과 기대를 더욱 높이고 있으며, 외부활동에 대한 욕구를 증가시키고 있다. 특히, 도시공간의 대표적인 오픈스페이스인 도시 근린공원은 도시민들에게 운동, 휴식, 이벤트 등과 같은 다양한 외부활동을 제공함으로써 삶의 질에 대한 욕구를 충족시키고 있다<sup>7)</sup>. 하지만 제한된 리소스를 가진 공원에서의 이용량을 나타내주는, 보행량 데이터가 주는 의미는 다양하다. 기후, 계절, 공기질 같은 대기환경 수치 데이터부터 주말이나 공휴일과 같은 범주 데이터에 이를 뿐만 아니라, 코로나 환자의 증감 등 다양한 사회적 요소가 발견된다. 만약 이들 요소 간의 상관관계를 인공지능을 활용하여 모델링하고 예측할 수 있다면, 공원 관리에 많은 부분을 대처하여 사용자에게 이롭게 할 수 있을 것이다.

한편 도시공원의 5가지 주요 기능은 환경성, 자연 접촉, 경관성, 건강증진, 사회 통합으로 나누어진다. 이 중에서 환경성은 크게 기후 변화 대응, 지능형 공원 관리로 나뉘게 된다<sup>8)</sup>. 또한 기후변화대응은 지능형 생태자원관리, 폭염, 미세먼지 등 대응 녹화 기술, 대피시설확보 등이 있다.

이처럼, 급격한 기후변화에 의한 물 부족, 연평균기온 상승, 도시 열섬 등을 동반한, 전 세계적 기후 위기시대에 공원은 도시 공간 내에서 기후 변화에 대응하는 역할을 수행할 수 있다. 기후 변화 대응의 여러 영역 중 폭염 방재 및 미세먼지

5) 신해민 외(2021), 「포스트 코로나 시대에 적합한 공원카페 계획 및 제안」, 『한국색채학회 학술대회』, 140-143.

6) 안소영, 원두환(2015), 「메타분석을 이용한 도시공원의 경제적 가치 함수 추정」, 『한국환경정책학회 학술대회논문집』, 72-73.

7) 임은나 외(2013), 「도시근린공원의 열환경 개선을 위한 열쾌적성 평가」, 『한국지리정보학회지』, 16(4), 153-170.

8) 이은영(2018), 「스마트 도시공원 조성 전략 및 제도 개선에 관한 연구」, 한양대학교 대학원 박사학위논문.

저감은 도시 기능의 매우 중요한 요소라고 할 수 있다. 또한, 대기 환경의 질을 판단하는 지표인 미세먼지와 이산화탄소에 대한 도시공원의 저감효과에 대해서도 활발한 연구가 진행되고 있다. 이러한 부분에도 인공지능 기술을 활용하여 이와 같은 개선 효과를 모델링 하는 것이 스마트공원의 방재 기능을 확인하는데 효과적일 것으로 보인다.

21세기의 도시공원은 자연풍치(風致)의 환경 구현을 통하여 도시민에게 휴식 공간 및 레크리에이션의 장소를 제공하여 삶의 질 향상에 결정적인 기여를 하고, 또한 도시공간 속의 대기를 정화하여 대기오염의 심화를 방지해 주며, 야생조류나 곤충 등의 생물을 보호함으로써 자연생태계가 상호 공생할 수 있는 터전을 건설하는 중요한 역할까지도 수행하는 사회적 심미적 생태적 기능을 모두 갖추고 있다<sup>9)</sup>. 여기에 더해, 본 연구에서는 스마트공원을 통해서 현대 공원 관리의 용이성과 사용자의 편의성을 개선하고자 한다.

## 2. 연구의 목적

본 연구의 첫 번째 목표는, 21세기형 스마트공원 관리에 필요한 저가형 IoT 기반 클라우드 데이터 수집 장비 개발 방법을 제안하고, 실제 대·중·소공원에 활용하여 직접 대기 정보를 수집하여, 이를 정량화하는 것에 활용하는 것이다.

스마트공원의 여러 가지 기능에서, 서울시 도심 속 그린 인프라를 통한 기후변화 대응 효과를 객관적으로 확보하기 위하여<sup>10)</sup>, 도시공원을 중심으로, IoT기반 온도, 습도, 일산화탄소 및 미세먼지 농도 측정시스템을 직접 제작·설치 운영한다. 그리하여 객관적으로 평가하고 가시화 가능한 정량적인 환경 빅데이터를 확보할 수 있도록 한다. 현재 각 지자체에서 기존공원에 IoT 응용 기술을 더하여 스마트공원을 구현하고 도시공원 특성에 맞추어 설계된 기상 장비를 사용하여 환경 데이터 수집을 수행하고자 하거나 사용하려고 할 때 기술적인 부분과 비용적인 어려움이 있다. 이를 해결하기 위하여 경제적으로 사용할 수 있도록 고려하여, 저가격 범용성의 IoT와 클라우드 장비를 쉽게 개발하는 방법을 제공한다<sup>11)</sup>.

9) 문창현(2001), 「이용 만족도 결정요인의 분석을 통한 서울시 도시공원의 공간적 특성에 관한 연구」, 『한국지리환경교육학회지』, 9(1), 97-118.

10) 이동근, 김호걸(2014), 「그린 인프라를 통한 기후변화 문제 해결」. 『환경논총』, 54.(.), 27-36.

첫 번째 연구 대상지인 서울숲은 서울특별시 성동구 성수동 1가 685번지 일대에 있는 공원이다. 서울 대도시의 중심부에 위치하였지만, 넓은 잔디 광장과, 나무 숲, 그리고 수변부를 갖춘 전형적인 도시공원으로서, 새벽부터 늦은 밤까지 꾸준히 많은 방문자의 보행량이 있다. 이러한 이유로, 방문자의 보행량과 도심지 내 공원의 환경데이터를 측정하기 위한 스마트공원 구현의 대상지로 선정하였다.

제작된 환경 센서 장비의 실증을 위하여, 연구 첫해인 2021년 5월부터 2022년 5월까지 서울숲공원 내, 녹지지역인 ‘가족마당’ 과 내부 호수에 근접하게 위치한 ‘수변부’ 부근, 공원 중심부로부터 약 400m 떨어져 있지만, 공원 내에서 최대 도심지에 접하고 있는 ‘서울숲 방문자센터’ 이렇게 3곳을 기준으로, 제작한 IoT 장비와 클라우드 서버를 이용하여 환경 자료를 수집한다. 이렇게 수집된 환경 데이터를 이용하여, 같은 기간에 수집된 서울시의 기상 공공데이터와 검증하고 비교하여 환경 데이터를 유효화할 수 있도록 한다,

취득한 데이터를 중심으로, 공기질을 포함하는 환경데이터와 보행량 사이의 상관관계가 어느 정도 있는지 파악하고, 이 상관관계를 밝히기 위해 인공지능 모델링 해석 및 이를 위한 장치의 도입이 필요하다. 이를 위해 보행량 센서를 기상 데이터 센서 설치 위치 주변에 설치하여 보행량 데이터를 수집한다.

이렇게 인공지능 모델링기법을 활용하여 12)공원 내 온·습도, 미세먼지, 일산화탄소 등의 공기질 데이터와 같은 환경 계수와 그리고, 코로나바이러스 감염자 수, 요일, 계절별, 휴일별, 요일별 보행량 등의 사회적 계수를 활용하여, 서울숲공원의 보행량과 시설의 운영에 미치는 상관관계 모델링 효과와 보행량 예측을 하고 이를 다시 검증하여 공원 운영 설계 시 고려할 사항 등을 살펴보고자 한다.

특히 보행량과 기상 데이터의 상관관계는 인공지능의 알고리즘 중 Voting Ensemble, XGBoostRegressor, LightGBM, RandomForestRegressor 을 모델링 기법을 활용하여 다중회귀분석을 수행하고, 이에 따른 데이터 분석과 머신러닝을 통하여 모델별의 효과를 예측·검증한다.

본 연구의 두 번째 목표는, 폭염과 같은 기후변화 위기에서, 도시공원의 환경 그린 인프라를 통한 환경 개선 효과를 정량적으로 측정·증명하는 데 있다. 인천

---

11) 이정훈 외(2018), 「4차 산업혁명 경기도 모델 구축 및 실행계획 연구」, 경기연구원, 1-399.

12) 김용국 외(2020), 「스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향」. 『한국조경학회지』, 48(6), 1-15.

광역시 남동구에 있는 소래 생태습지공원은 바다가 접하는 지역을 사이에 두고 형성된 염생 습지 지역으로서, 서울과 인천과 같은 대도시 생활권과 근접해있다. 기후조절 기능에 대하여, 해양환경공단·한국해양과학기술원·서울대·부산대 등 10개 기관의 ‘블루카본’ 연구 결과를 보면 염습지의 탄소흡수량이 갯벌보다 최소 1.7배에서 최대 4.7배 많은 것으로 조사되어 탄소의 저감효과를 측정하기 위한 스마트공원 구현의 연구 대상지로 선정한 배경이 있다.

온도저감효과와 이산화탄소저감 효과를 객관적으로 입증하기 위해 도시공원인 소래습지 생태 공원에서 연구를 진행한다<sup>13)</sup>. 2022년 7월~8월 폭염 기간 동안, 인천 소래 습지 생태공원과 도로 부에 연구를 위한 장비로 상용 자동기상측정장치 (Automatic Weather Station, AWS)를 스마트공원에 적합하게 응용 개발하여 설치하고 환경 자료를 수집한다. IoT장비의 경우에는, 서울숲공원에서 사용하였던 자체 개발한 IoT 장비와의 실증 효용성을 비교 증빙하기 위하여, 자동기상측정장치를 응용 개발하여 온도, 습도, 이산화탄소, PM2.5, PM10 미세먼지 농도, 풍향, 풍속, 돌풍, 기압, 조도, 자외선 등의 측정시스템을 도로 부와 병행하여 설치한다. 먼저 이렇게 수집된 데이터를 사이트별로 비교·평가하여 온도, 이산화탄소, 미세먼지 저감 및 방재 효과를 측정한다. 또한 이와 같이 확보된 환경 빅데이터를 인공지능 모델링을 비교 활용하여 이산화탄소 발생량 상관관계를 예측 도시공원의 기후 방재 효과와 인공지능 모형화를 통한 스마트공원 구현을 하는 것에 있다. 즉, 폭염이 있는 여름철 도시공원의 방재 효과를 측정하기 위하여, 두 번째 연구 기간인 2022년 여름철인 7~8월 6주 동안 자동 기상 측정장치(AWS)를 사용하였다, 도시공원과 근거리에 떨어져 있는 도심부 포구 사거리 도로 지역의 공기질 데이터를 비교하여, 온습도, 미세먼지, 이산화탄소 등의 방재 효과를 비교한다.

비교 대상인 대표적인 생태공원인 소래 습지 생태공원에, 동일한 환경 데이터 센싱 장비를 탑재하여 갈대숲과 염습지에 설치한다. 갈대숲은 주변에 매우 많은 갈대가 우거져 있어서 활발한 광합성이 일어나는 대상으로 선정하였고, 염습지는 앞에서 언급한 바와 같이 블루카본(blue carbon)을 통한 탄소의 흡수가 매우 활발한 곳 이므로 대상으로 선정하게 되었다. 블루카본이란 연안에 서식하는 염생 식물이 광합성을 통해 흡수한 탄소와 조석·파도 등 물리적 작용에 의해 갯벌(진

13) 이양주 외(2020), 「경기도 생태(生態) 분야 그린뉴딜 정책연구」, 경기연구원, 1-151.

흡) 사이사이 공간에서 포집된 탄소를 이르는 말이다<sup>14)</sup>. 본 연구에서는, 염습지의 이산화탄소 감소 효과를 같은 시기에 측정하여, 동일하게 기후변화 대응 효과를 예측하고, 영향 인자를 발견하여 방제 효과를 정량화할 수 있게 하는 것을 목표로 한다. 이산화탄소를 목표변수(target)으로 한, 기상 데이터의 상관관계는 인공지능의 알고리즘 중 Stacking ensemble, XGBoostRegressor, LightGBM, RandomForestRegressor 을 활용하여 다중회귀분석을 수행하고, 이에 따른 데이터 분석과 머신러닝을 통해 예측 검증한다.

### 3. 연구의 내용 및 방법

본 장 이후, 연구의 흐름은 다음과 같다.

제2장에서는 먼저 스마트공원을 정의하고 국내·외 스마트공원 사례를 통해서, 기후변화와 관련된 스마트공원의 역할, 특히 그 중에서도 환경 개선 기능에 대한 선행연구 및 적용사례를 고찰한다. 이와 관련하여 첫째, 국내외 스마트공원의 사례와 설치되어 있는 IT 기기들을 살펴본다. 둘째, 스마트공원의 기후 변화 방제 사례, 공원의 그린인프라를 통한 기후변화 대응사례 연구를 검토한다, 셋째, 스마트공원 구현을 위한, IoT, 인공지능 기술의 선행사례를 탐구한다. 이를 통해, 예시되었던 도시공원의 기후변화 대응 열 저감 효과를 정량화하기 위하여 반드시 필요한 IoT 환경 장비들을 분석한다. 그리고 이어서 인공지능 선형회귀분석을 통한 모델링 분석고찰 사례와 인공지능 모델링 분석 고찰을 연구한다.

제3장에서는 스마트공원의 시범 구현을 위해서, 서울숲공원에 IoT 기기를 제작하여 인공지능을 적용하여, 보행량과 환경변수의 상관관계 등을 고찰한다. 연구진행을 위하여, IoT 모니터링 방법과 적용계획에서는, 3.1.1에서 IoT 모니터링 기기 개발 방법과 적용사례 중 장비 개발과 고려사항에 대한 부분을 먼저 다루고, 3.1.2 에서 AutoML을 사용한 전체 연구 flow 등을 해석하고 관련된 머신러닝 알고리즘, 데이터 전처리 방법, 모델 평가 방법에 대한 선행 연구를 진행한다. 서울숲의 경우 온도, 습도, 미세먼지의 효과를 실제 데이터를 통하여 검증하고 비교할 수 있도록 1년 이상의 빅데이터를 수집한다.

---

14) 임민수. 「블루카본(Blue Carbon)」, 『에코타임스』, 2022. 07. 21.

외부기관의 보행량 측정 IoT 장비를 동시계열에 사용하여, 서울숲의 보행 데이터를 측정하고, 인공지능 모델링을 통한 실제 공기질, 공원사용량과의 관계성을 도출하도록 인공지능 관련 선행연구를 고찰한다. 이어서, 인공지능을 이용한 상관관계 모니터링 계획 및 방법에서, 인공지능을 활용한 전체 연구 flow와 함께 머신러닝에 대한 내용과 주로 사용되는 인공모델링 알고리즘의 종류와 데이터 전처리, 하이퍼 파라미터 세팅, 최대 최소 정규화 그리고 모델 평가에 대한 방법을 제시한다.

서울숲 스마트공원 구현에서는 도시공원을 중심으로, 수집된 기상 데이터와 보행량의 상관관계를 인공지능 모델링을 통한 종합적인 비교를 통해서, 스마트공원 관리의 가능성에 대해 결과를 살펴본다. IoT 데이터와 보행량의 상관관계를 파이선과 관련 라이브러리를 사용하여 비교 분석하고, Auto ML 머신러닝 기법을 사용하여, 선형회귀를 4가지 모델링 방법으로 분석하고 효용성을 분석한다.

제4장에서는 소래생태습지 공원을 대상으로 인공지능 모델링을 이용한 기후변화 대응에 대한 고찰한다. 이 곳의 연구에서는, 스마트공원의 환경데이터 수집을 위한 방법으로 자동 기상 관측시스템(AWS)를 스마트공원 중 전원, 무선 와이파이 무공급 지역에 구현할 수 있도록 개량하여, 데이터를 수집함으로써, 기존 데이터 수집요소를 공정화한다. 아울러, 도시공원의 방재 효과 정량화를 위한, 자동 기상 관측시스템(AWS)의 온라인 설치를 제안하고, 데이터를 수집하는 등의 모니터링 방법 및 계획을 논한다. 그리고, 인공지능 모델링 비교를 통한 탄소 발생요소를 분석하여 이산화 탄소 발생량을 예측하고, Stacking Ensemble, Voting Ensemble, XG Boost Regressor, Random Forest, Gradient Boost Regression, Light GBM, Multi Layer Perceptron 등의 알고리즘을 사용하여, 스마트공원의 인공지능 모델 성능평가를 진행하여 차후 탄소저감 및 방재효과 연구에 도움이 될 수 있도록 한다.

그리고 도시공원의 기후변화 대응 그린 인프라로서의 이산화탄소, 온도, 미세먼지 등 방재 효과성과 저감효과의 유효성에 대해서 고찰한다. 파이선과 판다스, 넘파이 등의 라이브러리를 이용하여 습지와 도심지의 IoT 데이터를 비교하고 상관관계 등을 분석하여 이를 다양한 도식으로 나타냈다. 머신러닝을 통하여 이산화탄소 발생에 대한 환경 데이터와의 유의미한 연관성을 학습시킨 뒤에, 다양한 모델링을 통해 염습지, 갈대숲, 그리고 포구 사거리에 적용한다. 사용된 알고리즘의 대표적인 앙상블 알고리즘인, Stacking ensemble, XGBoostRegressor,

LightGBM, RandomForestRegressor로서, 각 모델링 별로 평가 결과를 도출한다.

제5장에서는 연구의 결론 및 한계, 향후 연구 방향을 제시한다. 지금까지 논의한 연구의 흐름은 그림 1 과 같다.

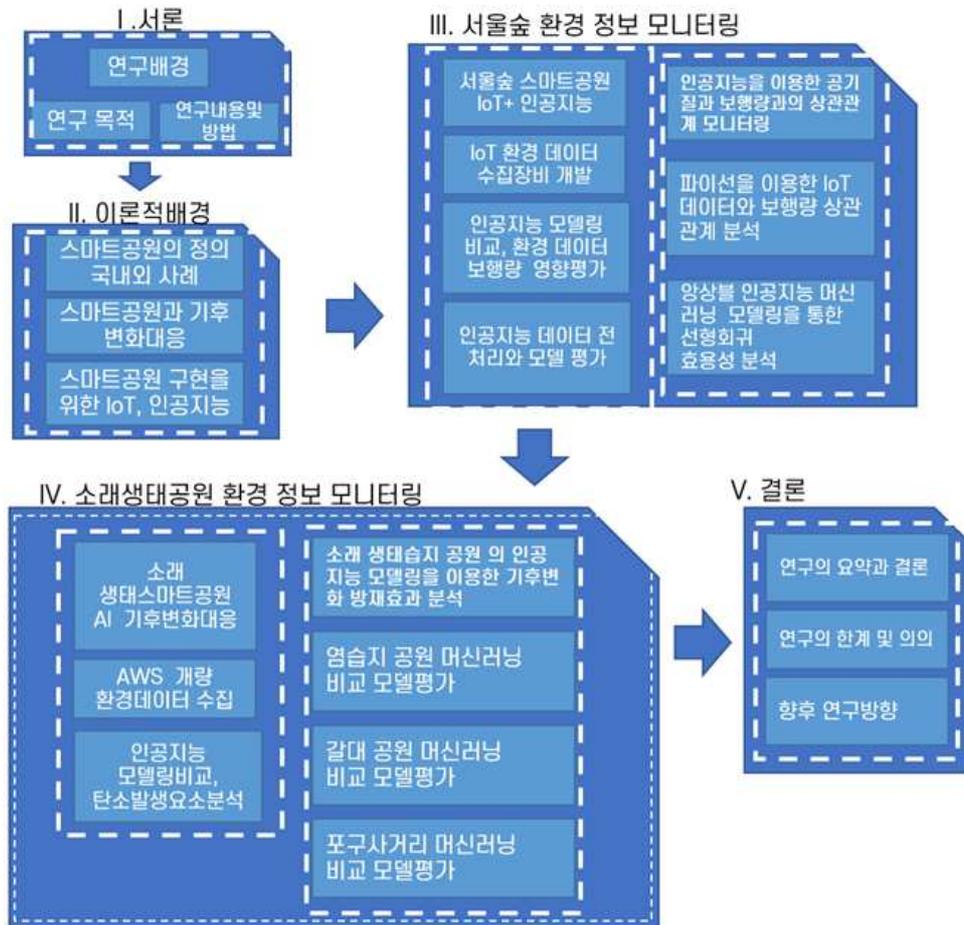


그림 1 연구의 전체 흐름도

## II. 이론적 배경

### 1. 스마트공원 정의

18세기에서 19세기, 독일의 정원이론가 히르시펠트가 옹호한 구불구불한 선을 이용한 풍경식정원은 현대 도시공원과 경관을 탄생시켰다<sup>15)</sup>.

19세기 중엽에 영국에서 태동한 도시공원은 전 세계의 도시에 필요 불가결한 도시 시설이자 도시문화의 중요한 축을 이루고 있다. 초기의 도시공원과 현대의 도시공원은 그 내용과 기능·외관에 있어서 많은 차이를 보인다. 또한 앞서서 특히 센트럴 파크를 설계하고 실행한 움스테드의 공원의 목표는 19세기 산업 도시에서 공중 보건, 알코올 중독, 폭력 및 계급 갈등과 같은 도시 문제를 완화하는 것이었다. 그 당시 도시공원의 목표는 부분적으로 달성되었다. 도시공원에 대한 그의 비전과 아이디어는 조경이 단지 기술적인 작업이 아니라 문화 제품인 공원을 발명함으로써 새로운 문화적 실천을 위한 장소를 만들 수 있음을 보여준다<sup>16)</sup>. 이제 21세기의 도시공원은 자연풍치(自然風致)의 환경 구현을 통하여 도시민에게 휴식 공간 및 레크리에이션의 장소를 제공하여 삶의 질 향상에 결정적인 기여를 한다. 또한 도시공간 속의 대기를 정화하여 대기오염의 심화를 방지해주며, 야생조류나 곤충 등의 생물을 보호함으로써 자연생태계가 상호 공생할 수 있는 터전을 건설하는 중요한 역할까지도 수행하는 사회적 경관적 생태적 기능을 모두 갖추고 있다<sup>17)</sup>. 여기에 더해서, 스마트공원을 통해서 현대 공원 관리의 용이성과 사용자의 편의성을 개선하고자 한다.

한편 4차 산업혁명 기술 중 가상현실 기술이 도시공원에 미친 영향은 가상의 공간에 최소한의 비용과 시간을 투자하여 가상의 오픈스페이스에 문화공원을 구축하도록 하였다. 기존공원의 틀에서 벗어나 가상공원 방문객의 시각 감각기관인 ‘눈’ 뿐 아니라 청각 감각기관인 ‘귀’와 ‘영혼’을 울리는 감성까지 고려한

---

15) 조경진(2014), 「히르시펠트(C.C.L.Hirschfeld) 정원예술론의 의미와 가치의 현대적 해석」, 『한국전통조경학회지』, 32(3), 58-68.

16) 조경진(2003), 「프레데릭 로 움스테드의 도시공원관에 대한 재해석」, 『한국조경학회지』, 30(6), 26-37.

17) 문창현(2001), 「이용 만족도 결정요인의 분석을 통한 서울시 도시공원의 공간적 특성에 관한 연구」, 『한국지리환경교육학회지』, 9(1), 97-118.

가상의 오픈 스페이스에서 추구하는 다양한 소리와 영상을 설치하려는 시도가 있었다. 가상의 현장에 나무 심기, 공원 구현 및 전시 등 가상공간에서 사용자에게 상호작용을 제공하고 다양한 경험을 할 수 있도록 하였다. 하지만 코로나 사태 등으로 많은 사람이 실내에서만 머물게 되면서, 실제로 인간이 즐겨하는 것은 가상의 공간이 아닌 실제의 공원을 걷고 숨을 쉬며 자연의 바람을 맞는 것이라는 것을 깊이 깨닫게 되었다. 우리는 그것을 "버추얼 리얼리티(Virtual Reality)"이라는 용어에 반하여 "네추럴 리얼리티(Natural Reality)"이라고 정의하고자 한다<sup>18)</sup>. 그런 면에서 모든 것이 디지털화되고 암호화된 가상의 공간을 벗어나 현실의 자연 세계에 많은 것들을 현실과 메타버스 디지털 세계에서 동시에 활용될 수 있는 신호로서 변환하고, 디지털라이제이션(Digitization) 할 수 있는 것을 기존 도시공원에 적용하여 추구하고자 하는 새로운 기술과 시도가 있다.

이처럼 스마트공원의 배경에는 수많은 아날로그적인 감성과 디지털 기술의 축적, 이를 활용한 관리와 사용자의 편의에 대하여 많은 논의가 있었다. 그러나 무엇보다 가장 큰 배경으로는 제한된 공원의 자원을 여러모로 매우 유익하게 사용하고자 하는 데 있다. 다양한 집단의 방문객들이 활용할 수 있는 제한된 공간의 효율성과 투자 대비 가치 극대화 즉, ROI(return of the investment)를 위하여 하나의 자원을 매우 다양하게 기획하여 사용하는 원소스 멀티유즈(one source multi use case)의 논리로 공원 보행량을 측정하고 상황에 맞는 서비스를 제공하는 것은 지금까지의 수동적인 공원 관리에서 능동적인 공원 관리로의 변화가 시작된다고 할 수 있다. 전 세계의 기후변화 문제에 따른 방재 효과를 가진 공원의 온도 저감, 탄소 저감, 미세먼지 감소와 같은 환경 개선의 효과가 더욱더 드러남에 따라서 이에 대한 마땅한 해석과 정량화 예측 등에 대한 당위와 필요가 앞으로 더욱더 대두될 것이 확실하다. 이해를 돕기위해서 그림 2 와 같이 스마트공원의 탄생배경을 도식화 하였다.

이와 같은 근대 공원의 현대화와 스마트화는 마침내 스마트공원을 탄생시켰고 기술적 입장, 관리적 입장, 사용자 입장, 기후변화 대응 입장에서 다음과 같이 정의될 수 있다.

첫째, 기술적 입장으로 스마트공원은 야외공간 전체에 네트워크 기술이 통합된 커뮤니티 공원이다. 무료 WiFi가 제공되는 IoT 환경에서 친환경 태양열 스마트

---

18) 권태경(1996), 「디지털클럽 - 가상현실용 휴먼인터페이스」, 대우전자 영상연구소 기술 보고서.

벤치를 통해 방문자는 스마트폰을 통한 인터넷 세계와 연결 상태를 유지하고 기기를 무선으로 충전할 수 있다. 일부 스마트공원에는 운동기구를 사용 시 청정에너지를 생성하는 운동기구가 설치되어 있다<sup>19)</sup>. 수집된 데이터는 실시간으로 온라인 서버에 지속해서 저장되어 24시간 공원의 환경을 장기간 관리하고 예측하는 목적으로 사용될 수 있다.

둘째, 관리적 입장에서 스마트공원은 공원 일상의 데이터 수집을 통해 방문자로부터 많은 것의 정보를 수집할 수 있다. 수집된 데이터는 요일별, 계절별, 휴일에 얼마나 많은 사람이 공원을 방문하는지 알려준다. 또한 방문자가 얼마나 오래 머물고 있는지 추적하여 방문자가 오고 가는 주요 시간을 제공한다. 이 데이터를 통해 공원 리크레이션 프로그램 관리자는 공원을 시설과 운영시간 면에서 개선하여 지역사회에 더 나은 서비스를 제공할 수 있다. 스마트공원 데이터 수집을 통해서 방문자들이 언제 어느 곳에 머물러 있고, 앞으로 공원에 올 가능성이 있는지 더 잘 이해할 수 있어서 다양한 모임, 수업 및 스포츠 행사와 같은 공원 이벤트 일정을 잡기가 더 쉬워진다. 스마트공원은 또한 공원에서 가장 많이 방문한 지점을 기록 할 수 있다. 이것은 공원 내 시설 중 어린이 놀이터에 몇 개의 그네를 더 추가해야 하는지 또는 운동시설 주변에 더 많은 휴식용 벤치를 추가해야 하는지와 같은 것들에 대한 통찰력을 줄 수 있다. 날씨, 미세먼지 및 우적량과 같은 환경 데이터를 기반으로 공원 방문객으로부터 수집한 보행량 데이터와 비교하여 공원 방문을 예측하고 향후 계획을 지원할 수도 있다. 스마트공원은 앞으로도 우리의 요구가 변화하고 기술이 발전함에 따라 더욱 스마트해질 것이다.

셋째, 사용자 관점에서 스마트공원은 버추얼 리얼리티(Virtual Reality) 즉 가상환경이 아닌, 네츄럴 리얼리티(Natural Reality)인 자연 안에서 평화로운 시간을 보내는 부분에서 수많은 이점이 있다. 예를 들어, 자연에서 보내는 시간은 스트레스를 줄이고 우울증과 불안을 완화하며 기억력을 향상시키고 동기를 높일 수 있다. 스마트공원은 건강하고 활동적인 라이프 스타일을 장려하며, 이는 현대 라이프 스타일이 점점 더 책상 앞에 앉아서 머물러 있는 시간이 길어짐에 따라 더욱 중요해졌다. 그리고 건강한 일과 삶의 균형을 유지하려는 욕구가 커짐에 따라 사람들은 무료 와이파이 인터넷을 갖춘 스마트공원에서 자연공간을 바라보며 일을 할 수 있는 장소를 찾을 수 있다. 스마트공원은 이처럼 같은 공원 공간

---

19) “Smart Parks are here and to stay,” <https://www.bos.com/inspired/smart-parks-are-here-and-here-to-stay> (검색일: 2022. 11. 27.).

안에서 방문자가 원격으로 회사 일을 할 수 있는 환경을 제공한다. 코로나 이후 외부생활에 목마른 사람들은 실내공간인 인터넷 카페나 홈 오피스에 머무르는 대신 오픈스페이스인 스마트공원에서 자연을 즐기고 야외에서 생산적인 일에 좀 더 시간을 보낼 수 있다. 또한 스마트공원은 야간 공원 조명과 IP 카메라를 곳곳에 설치하여, 실시간 모니터링을 통하여 이용자의 안전성을 높여 범죄율을 낮추는데 많은 기여를 했다<sup>20)</sup>.

버추얼 리얼리티	네츄럴 리얼리티
	
권태경 작품 (1994,1995,1996)	배곧생명공원 2022년 11월 방문

표 1 버추얼 리얼리티와 네츄럴 리얼리티 비교

넷째, 기후변화 대응 입장에서 스마트공원은 기후변화로 인한 환경문제가 점점 더 드러나는 시점에 유일하게 대중적인 해결책이 될 수 있다. 도심부에 녹지로 조성된 넓은 공간인 공원이 제공하는 온도 저감, 탄소 저감, 미세먼지 저감과 같은 많은 방재 효과가 눈에 직접적으로 보이지는 않지만, 효과적이라는 것이 알려져 있다. 스마트공원은 보이지 않는 데이터를 수집하고 분석하여 정량화된 공기 질 데이터와 다양한 환경 요소를 결합하여 인공지능이 활용 가능한 빅데이터를 제공한다. 이후 인공지능을 이용하여 빅데이터 기계학습을 진행하고, 다양한 인공지능 모델링 중 최적의 알고리즘으로 학습된 데이터를 응용하여 공원 관리 예측에 활용할 수 있다.

표 2는 지금까지 논한, 네 가지 관점에서 스마트공원 기술의 사례에 대해 해석한 바를 보여준다.

20) “How Can Smart Parks Help Get More People Outside?,” <https://www.youtube.com/watch?v=rG2AisQqvug> (검색일: 2022. 11. 24.).

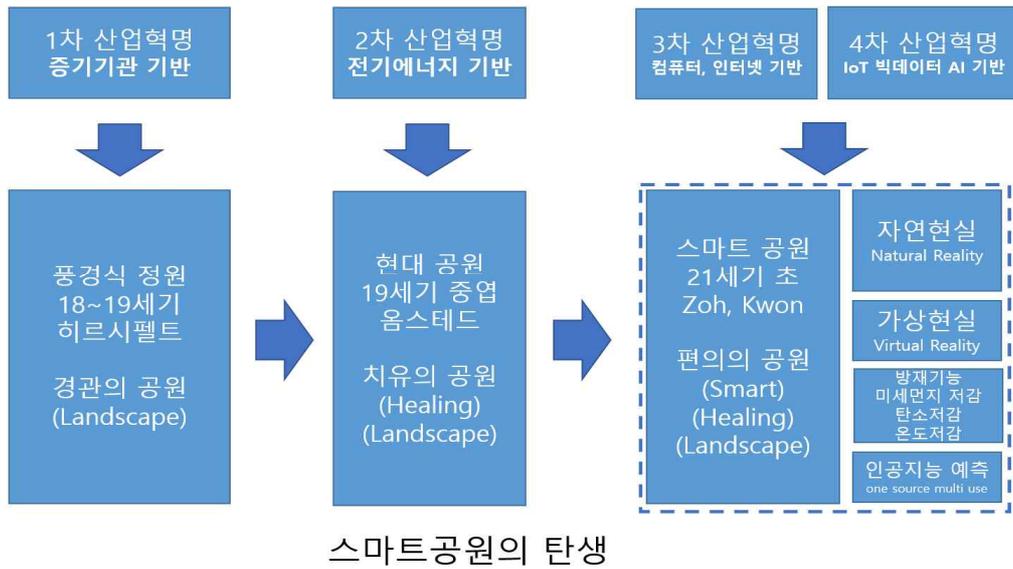
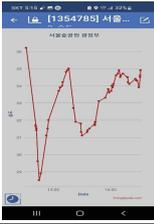


그림 2 스마트공원의 탄생 배경

표 2 스마트공원의 적용 사례와 클래스 분류

Number	사진	Classification of subjects				Keywords
		기술적 관점	관리자 관점	사용자 관점	기후변화 관점	
1		IoT	데이터 수집	초연결사회 우울감과 불안을 완화	환경변수 측정 도구	IoT, 사물인터넷, 연결사회
2		보행량 계수기 (people counter)	일자별 시간별 이벤트별 방문자 수 분석	걷기는 기억력을 향상	공원 사용량 정량화	보행량 측정, 걷기
3		환경센싱	이벤트 일정 계획에 도움	방문 동기 강화	환경변수 정량화	환경센서, 날씨, 공기 질
4		WiFi 자원관리	시설물 추가에 대한 통찰력 제공	네트워크 원격작업	공원 자원 절약	공원 관리, 공원 오피스
5		인공지능 예측	방문 예측 가능	편리한 공원 사용	기후변화 공원가치 정량화 가치추정	AI, 공원 기획
6		스마트 IP카메라	공원 내 이상행동 감지	안전보장	영상 데이터 수집	시큐리티, 안전 관리
7		스마트 가로등	전원 절약	야간 안전보장	리소스 절감 환경보호	스마트 센싱, 에너지 절감
8		해충 관리기	건강 유지	편의성 제공	새로운 유충 탐색	방재

## 2. 스마트공원 사례

### 2.1 한국의 스마트공원

권태경은 2021년 6월 한국조경학회 ‘스마트공원’ 주제의 웨비나에서 “스마트 파크는 사회 물리적 환경이 잘 반영되고 쉽게 접근할 수 있다. 또한 기후변화에 탄력적이며 물과 에너지 효율이 높고, 유지관리가 쉬우며 건강과 안전을 증진하는데 도움이 된다” 며 기상·보행량 측정 관련 빅데이터와 시계열 데이터를 사용한 AI예측 사례를 보여줬다. 그는 관련 연구가 더욱 진행되면 모바일 앱을 이용한 환경 데이터 모니터링과 메타버스 네츄럴 리얼리티를 구현해 더 현실감이 있는 사이버 파크를 제공할 수 있게 되고, 이를 통해 공원 이용자들이 보다 향상된 공원녹지 서비스를 받게 될 것으로 전망했다<sup>21)</sup>. 김용국 외 2인(2020)에 의하면, 스마트공원을 디지털·환경·재료 기술 등을 활용하여 시민들의 안전하고 쾌적한 공원 이용과 관리·운영의 효율 개선을 지원하여, 도시 및 지역사회의 사회·경제 및 환경적 지속가능성 확보에 기여하는 공원<sup>22)</sup>으로 정의하였다. 국내 스마트공원 사례에 사용된 기술은 그림 3과 같이 커뮤니티 운동, 급수시설, 자전거와 보행자 카운터를 포함하여, 안전 관리용 CCTV 시스템, 방문객 안내판, 디지털 사이니지(게시판), 블루투스 스피커, 공기질 측정기, 무료 인터넷, 아쿠아 놀이터, 점자 안내판, 전동 휠체어 급속 충전기 등이 있다.

이와 같은 스마트공원에 관한 관심을 증명하는 바와 같이, 한국수자원공사는 2021년 9월 부산 에코 델타 스마트시티 내 미래형 스마트공원 구현을 위한 ‘부산 에코델타시티 Smart+ 공원시설물’을 공모하였다. 공모 분야는 ‘공원’ 과 ‘철새서식처’ 2개 부분이었다. ‘공원’ 부문은 도시 열섬, 탄소, 미세먼지 등 환경 이슈에 대응하고 안전하고 편리하며 쾌적한 공원 환경을 제공할 수 있는 시설물이다. ‘철새서식처’ 부문은 철새 등 조류의 생태적 특성을 고려해 새와 인간이 공생하는 환경을 제공할 수 있는 시설물로, 탐조대 및 둥지, 먹이 공급시설 등 시설물 종류가 해당되었다<sup>23)</sup>.

21) 장재훈(2021), 「‘스마트공원’, 정보 계측·시뮬레이션 중요」, 『환경과 조경』, 398호.

22) 김용국 외(2020), 「스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향」, 『한국조경학회지』, 48(6), 1-15.

2022년 LH 세종특별본부에서는 세종국가지범 도시 미래형 스마트 실증공원 마스터플랜 수립 전문가 공모를 하여서, 스마트시티 혁신기술인 메타버스, 헬스케어 등을 실증·전시하고 스마트 인프라를 체험할 수 있는 공간을 제공하여, 시민 문화 융합 복합공원) 환경친화적인 오픈스페이스 구현으로 쾌적한 도시 환경 및 시민휴식·놀이공간을 제공하고자 하였다<sup>24)</sup>.



그림 3 국내 스마트공원에서 활용되는 다양한 스마트공원 기술 사례

2017년, 우리나라 최초의 스마트공원인 대구 국채보상 공원은, IoT See Park 이라는 명칭으로 탄생<sup>25)</sup>하였다. 제공하고 있는 서비스들은 그림 4 와 같이 위험 요소를 스스로 감지하는 인공지능 CCTV, 위치 확인 가능 대화형 비상벨, 자동 밝기 조절 스마트 가로등, 공공 WiFi, AR을 활용한 역사 교육 콘텐츠, 스마트

23) “부산 스마트시티, ‘에코-스마트’ 공원시설물 공모”, <https://www.smartcitytoday.co.kr/news/articleView.html?idxno=21386> (검색일: 2022. 10. 17.).

24) “세종국가지범도시 미래형 스마트 실증공원 마스터플랜 수립 전문가 선정”, [https://www.kila.or.kr/board/view?b\\_name=bo\\_news&bo\\_id=4310&per\\_page](https://www.kila.or.kr/board/view?b_name=bo_news&bo_id=4310&per_page)(검색일: 2022. 10. 17.).

25) 이형숙 외(2019), 「스마트공원 개념 정립 및 공원 이용자 인식에 관한 연구-대구 IoT See 시범사업 공원 이용자를 대상으로」, 『한국조경학회지』, 47(5), 41-48.



스마트가로등



스마트 안내판



태양광 벤치



스마트 쓰레기통

그림 4 국내 최초 스마트공원: IoT See Park (출처 : 랜드스케이프 타임즈)

방향 표지판, 환경 및 미세먼지 센서, 휴대폰 충전이 가능한 태양광 벤치, 화재 감지, 자동 쓰레기 적재량 체크, 압축 및 알람기능 스마트 쓰레기통 등이 있다.

이처럼 기존 도시공원을 다양한 기술과 통신 등을 융합하여 스마트공원으로 탈바꿈한 사례는 4차 산업혁명 시대에 공원이 가지는 새로운 가치를 발견하기에 매우 충분하다.

## 2.2 해외의 스마트공원

UCLA 루스킨(Luskin) 센터의 스마트 파크 툴킷<sup>26)</sup> 책에 의하면, 공원 관리에서 현안 문제점들로는 공원 유지관리를 위한 리소스 부족, 사용자 요구에 적합하지 않음, 접근성의 어려움, 노후화 등이 있다(표 3). 이를 해결하기 위한 해결책 역시 ‘기술’을 활용한 리소스 해소, 유저 데이터를 활용한 신규 공급 프로그램 공급, 무료 WiFi, 모바일앱을 통한 젊은 세대와의 소통, 그리고, 지역 커뮤니티 멤버들과 소통하여, 공원의 계획, 디자인 그리고 신기술 도입, 참여를 통한 주인의식 고취 등이 있다.

26) Loukaitou-Sideris, et al. (2018), *Smart Parks: A Toolkit*, UCLA: Luskin Center, USA.

표 3 스마트공원을 통해 해결해야 할 일반공원의 문제점과 해결방안

문제점	해결책	사례
자본을 위한 자원 부족 개선 및 유지관리	기술을 활용한 효율적 관리	스마트 컨트롤러가 있는 관개 시스템 자동 잔디 깎기, 태양광 IoT 쓰레기 압축기
사용자 요구에 적합성 부족	사용자 데이터를 활용한 신규 프로그램 공급	인터랙티브 웹사이트 / 디지털 사이니지 커뮤니티 그룹과 의사소통하고 공원설계 및 프로그래밍에 사용
불편한 접근성	무료 WiFi 및 모바일 앱을 통한 젊은 세대와의 소통	도로포장, 모바일 앱, 디지털 공원 지도 등
노후화	사용자 참여 및 공원 유지 의사소통	공원계획, 설계, 신기술 도입 및 참여를 통한 주인의식 고취를 위해 지역사회 구성원과 소통

출처 : Loukaitou-Sideris et al. (2018) pp14~16



그림 5 스마트 파크에서 사용되는 다양한 기술 사례 (출처 : Loukaitou-Sideris et al. (2018))



그림 6 세계 최초 AI 공원 베이징 하이덴 공원에 스마트 파크 시스템 적용 (출처 : 이데일리 2020-1-23)

현재 적용되고 있는 해외 스마트공원 애플리케이션은, 그림 4와 같이 자전거와 보행자 카운터, 모션 인식 라이트닝- 에너지 절감, 방문객용 모바일 앱, 주차장 유희공간 IoT 센서, 공기질 측정기 탑재형 가로등, 공원 사용을 위한 GIS 정보, 무료 개방형 인터넷, 유동 인구 측정용 가로등이 있다. 이 기술들은 일반적인 방문자들에게 편리한 기능을 제공은 하지만 공원 관리자에게 공원의 방재효과나 보행량 예측을 제공하지는 않는다.

한편 중국의 스마트공원이면서, 세계 최초 AI 공원인 베이징 하이덴 공원에는, 그림 6과 같이 자율주행 버스, AI 얼굴 인식형 개인 운동량 측정시스템, AI 대화형 음악감상 정자, 증강 현실을 활용한 태극권 트레이닝 시스템 등이 적용되었다. 이 중에서도 AI와 얼굴인식 기능을 결합한 ‘스마트 보행로’는 산책을 시작하기 전에 카메라로 얼굴을 인식시켜 놓으면, 보행로를 걸거나 뛰어서 이동한 뒤, 종착지 모니터에 얼굴을 다시 인식시키면, 운동 거리, 소요 열량, 평균 속도 등을 보여주고, QR코드를 통해서, 스마트폰에 저장해두면, 이용할 때마다 누적기록을 저장하여 알려주게 되어 공원을 찾는 이용자들에게 많은 호응을 얻고 있다<sup>27)</sup>. 하지만, 공원 방문자 중심의 편의 설계가 아닌, 과도한 기술로 인한 기술 스트레스와 여기 저기에 설치되어 있는 수많은 카메라 사용으로 인한 사생활 침해의 소지가 있다.

### 2.3 스마트공원 관련 인터뷰

스마트공원 관리와 정책을 위한 사례연구를 위하여 도시공원 관련 전문가인

27) 신정은, 「淸 황실 정원에 AI를 더하다…자율주행버스 달리고, 얼굴인식 운동량 측정」, 『이데일리』, 2020. 01. 23.

서울 그린 트러스트 이한아 사무처장, 서울특별시 시의회 사무처 유시범 입법조사관과 인터뷰를 진행하여 의견을 정리하였다.

먼저 도시공원 관리에 대한 스마트공원의 전문가의 의견을 듣기 위해 이한아 사무처장과 2022년 11월 10일에 인터뷰를 진행하였다. 서울그린트러스트는 본 연구 대상지인 서울숲 공원의 관리를 2016년~2021년 5년간 진행해왔으며 19년 전인 2003년 서울숲 공원의 기획 단계에서부터 운영에 참여하였다. 한편 향후 스마트공원 정책 입안 방향에 대한 서울시 의회의 측면에서 보는 관점을 위하여 2022년 11월 24일에 서울시의회 사무처 유시범 입법조사관과 인터뷰를 시행하였다. 유시범 조사관은 지난 10여 년간 공원 관련 정책 입안을 위한 입법 조사 등을 진행해 온 전문가이다.

먼저 서울그린트러스트 서울숲 운영기관으로서의 스마트공원에 대한 인식은 비용 제약으로 인한 운영상의 어려움은 다소 있으나 반드시 필요한 인프라로 여기고 있다. 마치 대중교통인 버스의 [버스정보안내시스템]의 최대 수혜자가 사용자인 버스 승객인 것 과같이, 스마트공원의 최대 수혜자는 공원 방문자가 될 것이다. 또한 축적된 데이터는 이용자에게 다양한 공공데이터로서 공개가 된다면 공원 활용도가 더 높아질 것으로 예상된다. 서울숲의 방문자 편의를 위한 적용된 ICT 기술의 대표사례는 서울시에서 제공한 무료 무선 인터넷 WiFi 서비스이다.

한편 무료 무선 인터넷은 대중교통에서도 제공되는 범용적인 서비스임에도 불구하고 원인을 알 수 없는 장애 등이 발생할 경우가 많았다. 공원 측에서 자체 해결을 할 수 없어서 며칠씩 개선이 안 된 경우 수많은 민원을 일으킨 서비스인 만큼 사용 빈도가 매우 높고 만족도·효용성이 높은 서비스라고 할 수 있다.

IoT 환경 센싱 기술과 관련해서는 공원 자체적인 서비스를 시작한 하지 못하였으나 몇몇 대학 연구기관 등에서 최근 몇 년간 미세먼지 공기 질 측정을 위한 실험을 공원 내에서 많이 수행하고 있다. 이처럼 공기 질 관리에 대한 지속적인 관심은 조만간 공원 내에 요소 곳곳에 지자체 산하 공공데이터를 기준으로 한 매크로 단위의 공기 질이 아닌, 공원 내 마이크로 단위의 공기 질 표시 장치가 곳곳에 설치가 되어 사용자의 편의를 도울 것으로 예상된다. 인공지능 기술 관련하여는 지금까지 공원 관리에 적용된 적은 없으나 증강 현실(augmented reality, AR), 스토리텔링(storytelling) 등에 관한 기술 활용 선행 조사가 존재하여 인공지능 또한 향후 많은 부분에 활용될 것으로 전망하였다. 한편, 공원과 건강관리에 대한 부분이 최근 부각되고 있는 상황이므로 공원 내에서 인공지능

기반 건강검진센터 등의 운영을 통한 정신, 육체 건강에 대한 원격 의료 등이 이루어진다면 많은 공원 방문객의 호응이 예상된다. 따라서 규제 유예제도로써 시도할만한 충분히 가치가 있을 것으로 판단된다. 보행량 측정으로 인한 공원 관리 운영에 대한 효과는 공유자전거를 주중, 주말에 따라서 사용량의 변동으로 적정 장소에 배치할 수 있다. 그리고 방문객의 혼잡밀집 예상 지역에서 사전에 안전요원을 배치하는 등의 실효성이 매우 커서 도움이 될 것으로 보인다. 공원이 지니는 기후변화 방재 효과와 관련해서는 기존 보고서, 논문 등에서 그 효과를 입증하고 있으며 사이트별 비교를 통한 개선 효과를 보여주고 있다. 앞으로 스마트공원을 통한 공원자원 사용의 극대화과 운영관리의 효율성 개선에 많은 도움이 될 것으로 기대하는 것으로 응답하며 인터뷰가 마무리되었다.

둘째, 정책 입안 방향에 관한 전문가의 의견을 바탕으로 진행한 인터뷰에서는 아직 스마트공원에 대한 인식은 아직은 부족한 편인 것을 알 수 있었다. 상대적으로 스마트 정원, 스마트 팜 등은 인식이 잘 되고 있으며 작은 규모의 건물 내 수직 녹화(Vertical Garden)사업 등에 적용되고 있다. 현재 적용된 공원을 위한 ICT 기술은 이동통신 서비스 제공업체와 협력하여 모바일 실시간 사용자 정보를 이용한 공원을 포함하는 주변의 보행량 정보를 사용하는 부분에 대한 정책적인 고려가 되고 있다. 그러나 개인 위치 정보 사용에 대한 부분 프라이버시 이슈 등이 아직 존재하여 집행에 어려움이 있다. 미세먼지와 같은 환경 데이터에 대해서는 사회적으로 많은 관심이 집중되고 있다. 특히 산림청 등에서 최근 미세먼지 저감과 관련한 해결 방법에 많은 관심과 예산 배분이 이루어지고 있다. 지자체 등 정책기관에서 인지하는 보행량의 요소는 주로 공원 관리 쪽에 많은 부분을 할애하는 것으로 응답하였다. 예를 들어 경의선 숲길 같은 선형 공원의 특성상 상업지구와 많은 부분이 상존 하는데, 공원 내 흡연이나 음주에 대한 민원이 잦다. 따라서 공원 관리 측면에서 보행량 분석을 통한 혼잡상태를 예측하여 사전 안전, 보안 요원 등 투입, 안전 가이드 제공이 가능할 것으로 보인다. 인공지능을 통한 보행량 예측의 경우에 최근 서울 중심부 외곽에 위치한 한 공원에 미디어 파사드 설치 민원에 따라서 관련된 사전 검토를 진행하였는데, 보행량 예측이 된다면, 고액의 설치와 운영 비용이 지출되는 사업인 만큼 충분한 사전 타당성 검토에 활용이 될 것으로 예상된다. 도시공원이 가지는 공기 질 개선 효과, 기후 변화 대응에 대하여는 자그마한 나무가 20년이 지나면서 크게 자라서 그늘을 통한 온도 저감과 미세먼지 제거, 미관 요소 등에 큰 효과가 있는 것을 인지하고 있다.

이처럼 가로수, 숲, 산림이 주는 효과에 대한 정량화 인식이 필요하다. 왜냐하면 도시공원의 숲, 호숫가가 주는 미세먼지, 이산화탄소 저감효과 등에 대한 정량화가 된다면 좀 더 예산 편성과 집행에서 확실한 근거가 될 수 있기 때문이다. 소래습지공원과 같이 정량화된 데이터로 이산화탄소, 미세먼지, 온도 저감에 도움이 되는 공원 등에 관하여는 시 조례 등을 통해서 부지를 그대로 보존하여 역사공원 근린공원을 형성하고 보존하는 것이 가능하다. 결론적으로, 미래 공원의 과제는 데이터의 수집과 정제, 데이터에 기반한 합리적 정책 입안을 통한 정책 입안을 바탕으로 한 이용 편의 및 환경 차원에서의 기능 증진에 있고, 이를 위해 스마트공원화 작업이 요청된다. 스마트공원의 핵심은 데이터 수집·정제 그리고 데이터에 기반한 합리적인 정책 입안이 될 수 있을 것으로 기대된다.

이상에서 다룬 두 건의 인터뷰 내용을 정리하면, 각각 다음의 표 4 및 표 5와 같다.

표 4 서울숲 관리 서울그린트러스트 사무총장 인터뷰 정리

번호	질문	목적	답변
1	스마트공원의 인식	관리자 입장에서 스마트공원 필요성	1) 스마트공원의 최대 수혜자는 사용자 2) 유저인터페이스가 편리하면, 관리시스템에 도입필요 3) 데이터 축적후 : 이용자에게 공개
2	현재 적용된 공원 내 ICT 기술	공원 현황 및 적용 기술 status 파악	1) 무료인터넷 서비스의 제일 효과가 뛰어남 2) 디지털 사이니지(digital signage) 3) 장애인용 시설 - 급속충전기, 점자 안내판 등
3	현재 적용된 공원 내 IoT 환경센싱 기술		1) 미세먼지 측정 (공기 질 측정기)를 각대학연구소에 진행 중
4	현재 적용된 공원 내 인공지능 기술		1)스토리 텔링 정보를 모바일 앱으로 해서 제공하는 서비스 연구 경험 2) 헬스케어 서비스를 2008년 한양대 병원과 공원방문객의 힐링 효과를 측정하기도 하였음 3) 향후 공원에서의 정신, 신체 건강에 대한 건강검진을 인공지능을 통해서 서비스한다면 좋겠음
5	보행량 측정 공원 관리 개선 효과		1) 코로나 전후 오픈스페이스인 공원에 많은 사람이 방문하고 있음. 2) 보행량관련 사전 예방 조치를 취하여 사고미연방지
6	유동 인구 및 밀집 지역 관리 가능성	보행량측정과 예측이 주는 공원 관리의 효용성	현재 공유자전거 배치를 주일, 주말에 따라 다르게 적용중
7	인공지능 보행량 예측 기술 공원 관리 효과		방문자가 집중되는 시기, 시간, 공간 등을 사전에 체크할 수 있어 도움이 될 수 있음
8	도시공원이 미치는 공기 질 및 기후변화 방재 효과 인식	스마트 도시공원 방재 효과에 대한 정량화 필요성 인식	1) 현재까지 서울숲을 대상으로 하는 논문, 보고서 등을 통해 충분히 인지하고 있음 2) 도시 열섬현상에서, 온도 저감효과 등을 사이트별 비교, 도시공원의 개선 효과를 살펴봄

표 5 스마트공원 정책 입안 서울시 시의회 환경 입법조사관 인터뷰 정리

번호	질문	목적	답변
1	스마트공원의 인식	관리자 입장에서 스마트공원의 필요성	1) 스마트정원, 스마트 팜 대비 스마트공원은 아직 낮은 상황, 스마트라는 단어는 첨단 의미로 다가옴 2) 수직정원에 App 등으로 활용하는 스마트정원은 최근 미세먼지 저감, 실내환경 개선 등의 영향으로 확장 중
2	현재 적용된 공원 내 ICT 기술	공원 현황 및 적용 기술 status 파악	1) 모바일 기반 실시간 사용자 정보로 공원 방문자수 체크 시도 2) 하지만 데이터 커버리지 단위가 매우 넓어, 사용자가 공원을 실제 방문하는지에 대한 여부는 불투명
3	현재 적용된 공원 내 IoT 환경센싱 기술		1) 미세먼지 측정 (공기 질 측정기) 는 매우 필요한 기술로 여겨짐 2) 산림청 등에서 최근 미세먼지 저감과 관련한 많은 관심과 예산 배분이 이루어지고 있음
4	현재 적용된 공원 내 인공지능 기술		1) 공원 내 범죄방지를 위해 방법용 AI 카메라 등의 적용
5	보행량 측정 공원 관리 개선 효과		1) 연령대에 따라 대상자 카테고리를 노인, 부모, 청소년, 사회적약자 등으로 분류 2) 어린이 공원, 가족 화장실, 선형 공원의 경우 보행량 인식에 서비스 운영과 관리에 적용
6	유동 인구 및 밀집 지역 관리 가능성	보행량측정과 예측이 주는 공원 관리의 효용성	1) 혼잡상태에서 안전, 보안 요원 등 투입, 안전 가이드 제공 가능
7	인공지능 보행량 예측 기술 공원 관리 효과		1) 사용자가 적은 도시공원의 경우, 사전 시설 설치의 효용성 분석, 이를 통해 의사결정을 진행
8	도시공원이 미치는 공기 질 및 기후변화 방재 효과 인식	스마트 도시공원 방재 효과에 대한 정량화 필요성 인식	1) 미세먼지 저감 2) 오랜 시간이 지나 작은 나무가 큰 숲이 되어 그늘막, 탄소저감등 기후변화 방재 효과가 있음을 인지
9	인공지능으로 이산화탄소 발생량 환경변수상관 관계가 주는 효과		1) 산림이 주는 효과에 대한 인식 필요 2) 도시공원의 숲이 주는 미세먼지, 이산화탄소 저감효과 등에 대한 정량화 필요
10	습지 보전과 같은, 도시공원 설계 입법 및 관리 기대 효과	도심지 유희용지 중 기후관리를 위한 노력	1) 근린공원, 역사공원 등 관련하여 조례를 통해 도시공원을 만들 수 있음
11	스마트공원 관련 Free discussion		1) 스마트공원 안에서의 통계가 중요 2) 데이터에 기준 한 정책 입안을 펼 수가 있음 3) 따라서 데이터 확보가 우선 필요

### 3. 스마트공원과 기후변화 대응

#### 3.1 스마트공원의 기후변화 대응 사례

스마트공원은 한국형 디지털 뉴딜과 친환경 뉴딜의 공통적 요소를 갖고 있다. 2020년 6월 기획재정부가 발표한 ‘친환경 뉴딜’ 부문 사업에 도시 숲 조성 등 ‘스마트 그린 도시 조성을 위한 선도 프로젝트 100개 추진’이 포함되어 있다. 따라서 2020년을 기준으로 15,975개소의 도시공원 가운데, 시민의 이용률이 높은 공원에 방재 거점, 미세먼지 저감, 태양광 등 신재생에너지 생산 등의 새로운 기능을 추가하여 스마트공원으로 새롭게 거듭날 수 있게 되었다. 이로써 김용국 외 2인(2020)은 스마트공원에 대해 시민의 건강과 안전 증진은 물론, 기후변화 적응과 신규 일자리 창출로 지역사회 발전을 도모할 수 있을 것으로 판단하였다<sup>28)</sup>.

이주용 외 1인(2021)은, 공원의 기능과 서비스에 대한 문헌을 검토하고, 조경 및 IT 분야 전문가들과의 인터뷰를 통해 서비스 항목을 도출하였다. 계층 분석법(analytic hierarchy process, AHP)은 경제성 정책성 등을 계량화하는 수치로 경제성에서 평가할 수 없는 사회적 가치도 반영한다. AHP 모델을 4가지 값으로 구성하였으며, 15개의 공원 서비스를 구축하여 38명의 조경전문가를 대상으로 설문조사를 실시하여 공원서비스의 중요도를 산정하였다. 연구 결과, 공해, 폭염 개선 등 기후변화 대응에 대한 환경 및 안전 관련 서비스 중요성이 높았다. 사회통합 서비스의 중요성은 상대적으로 낮았으나 지역사회 참여를 제고할 때 중요한 서비스로 여겨졌다. 미래의 스마트공원은 과도한 인공 시설과 기술을 구현하기보다 자연과 접촉하여 건강상의 이점을 향상 시켜야 한다고 주장하였다.

조현민 외 1인(2022)에 의하면 광장부와 수변부가 주변의 도로부와 각각 2.7℃ 및 2.9℃ 낮은 것으로 나타났으며, 온도가 높은 10시에서 16시 사이 시간대를 기준으로 비교하였을 때는 각각 5.5℃ 및 7.4℃로 매우 큰 차이가 있는 것을 알 수 있었다<sup>29)</sup>. 소래생태공원의 경우에는 본 연구를 통해서 탄소저감, 온도저감, 미세먼지 저감 등의 효과를 확실히 알 수 있었으며 인공지능을 통한 예측의 가능

28) 김용국(2020), 「스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향」. 『한국조경학회지』, 48(6), 1-15.

29) 조현민, 권태경(2022). 「여름철 도시공원의 열환경 개선 효과 - 서울숲 미기상 관측자료 분석을 중심으로」, 『한국조경학회지』, 50(6), 30-41.

성 또한 증명할 수 있었다.

이와 같이 도시공원이 스마트공원으로 점차 탈바꿈하면서, 기후 변화 대응에 대한 연구가 조금씩 늘어가고 있다. 그렇지만 IoT와 인공지능을 실질적으로 활용하여 공원 리소스 관리와 열저감, 탄소저감효과의 정량화를 수행하는 스마트공원을 실제 구현한 연구사례는 찾을 수 없었다.

표 6 스마트공원과 기후변화 대응 관련 선행연구

논문명	연구자	연구내용 및 방법
이용자 중심의 스마트공원 조성을 위한 공원 서비스의 상대적 중요도 분석 <sup>30)</sup>	이주용 이형숙 (2021)	공원의 기능과 서비스에 대한 문헌을 검토하고, 조경 및 IT 분야 전문가들과의 인터뷰를 통해 서비스 항목을 도출하였음. AHP 모델은 4가지 값으로 구성되었으며, 15개의 공원 서비스를 구축하여 38명의 조경 전문가를 대상으로 설문조사를 실시하여 공원서비스의 중요도를 산정하였음. 그 결과 공해, 폭염 개선 등 기후변화 대응에 대한 환경 및 안전 관련 서비스 중요도가 높음. 사회통합 서비스의 중요성은 상대적으로 낮았으나 지역사회 참여를 제고할 때 중요한 서비스로 여겨짐. 미래의 스마트 파크는 과도한 인공 시설과 기술을 구현하기보다 자연과 접촉하여 건강상의 이점을 향상 시켜야 한다고 주장함.
31) 스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향	김용국 송유미 조상규 (2020)	첫째, 스마트공원의 개념을 “디지털·환경·재료기술등을 활용하여시민들의안전하고쾌적한공원이용과관리·운영의효율개선을지원하여, 도시및지역사회의 사회경제 및 환경적 지속가능성 확보에 기여하는 공원”으로 조작적 정의함. 둘째, 스마트공원의 세 가지 원칙으로 공원의 본질적 가치개선, 도시문제해결을 위한 혁신적 기능향상, 설계와구현·관리과정의 스마트화를제시함. 셋째, 신도시와 기성 도시에 구현된 스마트공원 구현 사례 분석을 통해 개선 시사점을 도출함. 넷째, 녹지 영역, 수경시설 영역, 도로 및 광장영역, 조경 시설물 영역, 공원 설계 방법 등 다섯 가지 측면에서의 스마트공원 설계·관리 방향을 제시함. 다섯째, 스마트공원 구현·관리 활성화를 위한 정책적 시사점으로 도시 성장 단계와 공원 유형별 스마트공원 정책사업 모델 발굴 및 시범사업 추진, 한국형 뉴딜 정책과 연계한 스마트공원 사업 추진, 조경전문가가 주도하는 스마트공원 정책 추진의 필요성 등을 제시함.

30) 이주용, 이형숙(2021), 「이용자 중심의 스마트공원 조성을 위한 공원 서비스의 상대적 중요도 분석」. 『한국조경학회지』, 49(4), 55-63.

31) 김용국(2020), 「스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향」. 『한국조경학회지』, 48(6), 1-15.

### 3.2 공원의 그린 인프라를 통한 기후변화 대응 사례

지구온난화로 인한 기후변화 문제는 현대 사회가 당면한 가장 큰 현안 가운데 하나로, 이에 따른 도시열섬현상 심화, 홍수 등 도시 및 도시생태계의 생태적 안정성에 문제가 나타나고 있다. 더불어 인구증가와 인간을 위한 토지이용 요구의 증가에 따른 급속하고 과밀한 도시화의 결과로 도시의 생태적 문제가 양적인 측면과 질적인 측면에서 모두 크게 악화되고 있다<sup>32)</sup>.

무엇보다도, 기상재해 중에서, 사람들의 건강에 직결된 것은 폭염이다. 이러한 폭염에 대한 다양한 시도가 진행되고 있는 상황에서 도시 인프라로서, 도시공원의 온도 저감 효과는 새롭게 조명받을 수밖에 없다. 기후변화에 따른 폭염으로 전세계적으로 화상, 자외선 노출, 열 스트레스, 온열질환 및 호흡기 질병 등의 피해가 급증하여, 건강을 위협하고 있기 때문이다. 이러한 점에 대비하여 많은 도시에서 거리, 공공시설, 공원 등에 열 환경 개선을 위한 그늘막, 쿨루프, 쿨링포그 등의 시설을 적용하는 사례들이 늘어나고 있다<sup>33)</sup>.

이에 따라, 도시 내 폭염 문제에 대한 해결책으로, 자연지반으로 투수 지역이 확보되고 그늘을 위한 수목이 식재된 공간으로 도시의 냉섬 역할을 하는 녹지인 도시공원 녹지 면적의 확보가 가장 중요한 사안으로 떠오르고 있다.

따라서 본 연구에서는 국내 스마트공원과 폭염 방재, 탄소저감, 그린 인프라에 관련된 연구 방법을 중심으로 선행연구를 검토하였고, 또한 도시공원의 기후변화와 같은 자연재해로부터 방재 기능에 관한 연구사례도 함께 검토하였다.

본 연구의 연구방법에 적용하기 위해 관련 선행연구를 검토하였다.

박종화 외 1인(2016)은 전통적으로 공원 및 녹지조성은 시민에게 휴식, 산책, 운동 등 다양한 위락 공간 제공하는 데 초점이 맞추어져 있었으나, 최근에는 열섬 문제 완화를 위한 효과적 대응책으로 주목받게 된 것을 강조하였다<sup>34)</sup>. 해당 연구에서는 일산 신도시를 대상지로 선정하고 온도 저감에 영향을 미칠 수 있는 공원의 최소크기와 공원 크기에 따른 온도 저감의 거리 조락 효과 차이를 파악하고

---

32) 김명자(2021). 「기조발제: 기후·팬데믹 복합위기와 4차 산업혁명 시대, 지속 가능한 발전과 법률가의 사명」. 『저스티스』, 182(2), 23-86.

33) 김진목, 김세용(2015). 「[도시재생 특별법]에서 기후변화 대응 전략 반영의 필요성」, 『한국주거학회 학술대회논문집』, 141-145.

34) 박종화, 조기혁(2016), 「공원 크기에 따른 공원의 온도 저감 효과 분석: 일산 신도시를 대상으로」, 『국토계획』, 51(5), 247-261.

도시 열섬 완화를 위한 공원계획 시사점과 가이드라인을 제시하였다.

윤민호 외 1인(2009)은 위성영상 데이터 Landsat 5 TM Band 6에서 표면온도를 추출하고, 이를 자동기상관측소의 실측 기온을 이용하여 기온으로 보정하였다<sup>35)</sup>. 위와 같은 방법으로 위성영상으로부터 지표 온도를 추출 보정한 후 녹지가 도시 기온에 미치는 영향을 살펴보았다. 기온이 거리에 따라 급격하게 변화하는 지역을 기온 완화구간이라고 정의하고, 100m당 0.1℃ 이상의 기온저감 기준으로 기온 완화 효과 영역을 설정하였다. 위와 같은 방법으로 수혜면적을 도출한 후 해당 행정구역 내 인구가 동일하게 분포한다고 가정하여 기온 저감 수혜인구를 도출. 녹지의 기온 저감 영향이 녹지로부터 반경 500m까지 미친다는 선행연구를 토대로 기준을 설정하면 시가지 중 100m당 0.1℃ 이상의 기온 저감 영향을 받는 면적은 67.33km<sup>2</sup>로 전체 서울시 면적의 11.12%, 시가지 면적을 기준으로 하면 18.09%를 차지한다는 결과를 도출하였다.

김원주 외 3인(2010)은 서울의 공원녹지 정책이 공원녹지의 양적 확충에 주안점을 두고 수립됐으나, 기후변화에 대응하고 탄소를 적게 쓰는 사회에 적응해 나아가기 위해서는 녹지의 미기후 조절 기능, 수면의 열섬 감소 효과 등 직·간접적 탄소저감과 탄소흡수 기능의 확대할 수 있도록<sup>36)</sup> 공원녹지의 질적 향상을 추구해야 한다고 주장하였다. 이와 관련, 해당 연구에서는 공원녹지 분야에서의 탄소흡수원 확보 및 탄소저감방안을 제시하고 있다. 우선 기초자료 분석을 통해 서울시 공원녹지 자원의 탄소저장량을 산정하고 서울시 기후변화 경향과 열섬효과를 분석하고 있으며, 또한 공원녹지 분야의 탄소 흡수원 확보 및 탄소저감방안에 관한 국내외 사례를 조사한 후, 각각에 대한 시사점을 도출하였다.

2010년 저탄소 녹색성장 기본법 이 시행된 후 녹색과 관련된 법률 검토를 통해 내용을 분석한 결과 우리나라 법률이 규정하고 있는 녹색의 개념은 ① 환경성, ② 환경 에너지 활용, ③ 국민의 신체적 복지, ④ 지속가능성이라는 네 가지 핵심 개념으로 구성되어있다<sup>37)</sup>.

Konijnendijk et al.(2013)은 2000년 이후 출판된 225개의 문헌 연구를

---

35) 윤민호, 안동만(2009), 「위성영상을 이용한 도시녹지의 기온 저감 효과 분석」, 『한국조경학회지』, 37(3), 46-53.

36) 김원주 외(2010), 「공원녹지 분야의 탄소흡수원 확보 및 탄소저감방안」, 서울연구원 정책과제연구보고서, 1-155.

37) 윤순진(2009). 「‘저탄소 녹색성장’의 이념적 기초와 실재」, 『환경사회학연구』, 13(1), 219-266.

통해 현대 도시공원의 편익을 크게 생물다양성, 주택가격, 건강 및 웰빙, 냉각 효과, 대기질 및 탄소격리, 물 관리, 관광, 사회적 통합 등 8개 분야로 구분했다<sup>38)</sup>.

한편 기후변화는 도시민들의 건강에 다양한 영향을 미치는데 이에 대한 도시공원 및 녹지의 편익은 공기정화를 통한 지역 기후의 안정화<sup>39)</sup>(Jim and Chen, 2008), 도시열섬현상의 완화전략으로써 중요한 그늘 제공을 통한 냉각효과(Bowler et al. 2010)를 포함한다. 주변의 도시지역보다 시원한 도시공원의 냉각 효과는 공원 규모에 따라 차이를 나타내는데 3만m<sup>2</sup> 이상의 공원으로 조사되었으나, 3만m<sup>2</sup> 미만의 공원은 변동이 심한 것으로 나타났다<sup>40)</sup>.

Bowler et al.(2010)는 문헌 고찰 연구를 통해 도시공원은 하루 평균 0.94℃의 주변 지역 온도를 낮춘다는 결과를 도출하였다<sup>41)</sup>. 도시공원을 통한 냉각 효과는 식물 중, 차폐율(canopy cover)<sup>42)</sup>, 공원 규모 및 형태가 주요한 영향을 미치는 것으로 나타나므로 기후변화에 따른 열섬현상을 완화하기 위한 도시공원 계획 및 설계과정에서 고려되어야 할 필요성이 높다(Feyisa et al., 2014)

또한 기후변화의 맥락에서 도시공원은 미세먼지를 제거하는데 실질적인 기여를 하는 것으로 밝혀졌고, 그 밖에도 질소산화물 및 황산화물 등의 수준을 낮추는데 효과가 있는 것으로 분석 되었다<sup>43)44)</sup>(재인용 김용국, 2014)(Yin et al., 2011).

---

38) Konijnendijk, C. C., Annerstedt, M., Nielsen, A. B., Maruthaveeran, S.(2013), "Benefits of urban parks. A systematic review," *A Report for IFPRA*, Copenhagen & Alnarp.

39) Jim, C. Y., Chen, W. Y.(2008), "Assessing the ecosystem service of air pollutant removal by urban trees in Guangzhou (China)," *Journal of environmental management*, 88(4), 665-676.

40) 김용국(2014). 「녹색 복지 관점에서 서울시 생활권 도시공원의 분배적 형평성 분석」, 『한국조경학회지』, 42(3), 76-89.

41) Bowler, D. E., Buyung-Ali, L., Knight, T. M., Pullin, A. S.(2010), "Urban greening to cool towns and cities: A systematic review of the empirical evidence," *Landscape and urban planning*, 97(3), 147-155.

42) Feyisa, G. L., Dons, K., Meilby, H.(2014), "Efficiency of parks in mitigating urban heat island effect: An example from Addis Ababa," *Landscape and urban planning*, 123, 87-95.

43) Yin, S., Z. Shen, X. Zou, S. Che and W. Wang(2011), "Quantifying air pollution attenuation within urban parks: An experimental approach in Shanghai, China," *Environmental Pollution*, 158(8-9), 2155-2163.

44) 김용국(2014), 「녹색 복지 관점에서 서울시 생활권 도시공원의 분배적 형평성 분석」, 『한국조경학회지』, 42(3), 76-89.

주현수 외 1인(2009)은 도시 근린공원 규모의 녹지가 오존농도 상승의 주 원인인 질소산화물과 황산화물 등을 흡수제거하는 기능은 대류 확산에 의한 대기 질 영향보다는 작지만, 상대적으로 친환경적인 토지이용 형태를 유지함으로써 대기환경 악화를 억제하는 효과가 탁월하므로 도시지역에서의 오염원 추가건설 제어를 위한 녹지 보호 정책이 필요하다고 주장한다<sup>45)</sup>(재인용 김용국, 2014).

기후변화 외에도, 일본의 방재공원 관련 법률 및 제도 또한 살펴볼 필요가 있다. 일본은 1923년 관동대지진 당시 도쿄시 인구의 약 70%가 대피처로 공원 등을 포함한 오픈스페이스를 이용한 것으로 확인되었으며, 공원녹지를 포함한 오픈스페이스가 지진 직후 발생한 대화재 확산을 방지하는 효과가 있던 것으로 보고되었다. 이후 1978년부터 도시공원 정비사업 중점사항으로 방재공원 조성이 추진되었다. 1993년 「도시공원법 시행령」을 개정하며 재해응급 대책에 필요한 시설로 공원시설이 포함되었으며, 재해 발생 시 대피장소 또는 대피로 등이 되는 도시공원을 ‘방재공원’으로 표현하기 시작하였다<sup>46)</sup>. 이처럼 도시공원은 방재공원의 역할을 충분히 기대하고 있다.

이애란(2018)은 최근 도시 공원이 자연 속 여가 기능과 더불어 기후변화에 대응하기 위한 재해 예방형 시스템을 도입하고 있음을 논하였다. 해당 연구에서는 구체적으로 도시공원이 자연형 물순환과 미기후 조절 기능을 제공하며, 지역경관 및 생활의 질 향상과 환경정화기능을 더하고 있음에 주목하였다<sup>47)</sup>.

「국토의 계획 및 이용에 관한 법률」 시행령은 도시공원을 제2조 기반 시설에서 공간시설로 구분하고 있으며, 방재시설은 하천·유수지·방화설비·방풍설비 등으로 명시하고 있다<sup>48)</sup>. 또한 같은 법에 따른 광역도시계획수립지침과 도시·군의 계획수립지침에서도 방재계획에 관한 규정을 명시하고 있다.

이처럼, 기존 도시공원의 기후변화대응에 대한 방재 등의 기능을 갖는 스마트공원으로의 변모는 앞으로 공원이 지속해서 발전해 나가야 할 방향이라고 생각한다.

---

45) 주현수, 김석철(2009), 「도심지역 녹지의 국지적 대기환경영향에 관한 연구」. 『화학공학』, 47(3), 386-393.

46) 장미홍, 박창열(2019), 「일본 도시공원의 방재 기능 특징과 시사점 연구」 『서울 도시연구』, 20(2), 77-89.

47) 이애란(2018). 「민간공원 특례사업의 기후변화대응 방재계획 분석: 청주시 사례를 대상으로」, 『도시연구』, 14, 119-141.

48) 장미홍, 박창열(2019), 「일본 도시공원의 방재 기능 특징과 시사점 연구」, 『서울 도시연구』, 20(2), 77-89.

표 7 공원의 온도 저감 효과에 관한 선행연구

논문명	연구자	연구내용 및 방법
공원 크기에 따른 공원의 온도 저감 효과 분석 - 일산 신도시를 대상으로 49)-	박종화 조기혁 (2016)	전통적으로 공원 및 녹지 조성은 시민에게 휴식, 산책, 운동 등 다양한 위락 공간 제공하는데 초점이 맞추어져 있었으나, 최근에는 열섬 문제 완화를 위한 효과적 대응책으로 주목받게 된 것. 본 연구는 일산 신도시를 대상으로 선정하고 (1) 온도 저감에 영향을 미칠 수 있는 공원의 최소 크기와 (2) 공원 크기에 따른 온도 저감의 거리 조락 효과 차이를 파악하고 도시 열섬 완화를 위한 공원계획 시사점과 가이드라인을 제안
위성영상을 이용한 도시녹지의 기온 저감 효과 분석 <sup>50)</sup>	윤민호 안동만 (2009)	Landsat 5 TM Band 6에서 표면온도를 추출하고, 이를 자동기상관측소의 실측 기온을 이용하여 기온으로 보정하였음. 위와 같은 방법으로 위성영상으로부터 지표 온도를 추출 보정한 후 녹지가 도시 기온에 미치는 영향을 살펴보기, 기온이 거리에 따라 급격하게 변화하는 지역을 기온 완화공간이라고 정의하고, 100m당 0.1℃ 이상의 기온 저감을 기준으로 기온 완화 효과 영역을 설정하였음. 위와 같은 방법으로 수혜면적을 도출한 후 해당 행정구역 내 인구가 동일하게 분포한다고 가정하여 기온 저감 수혜인구를 도출. 녹지의 기온 저감 영향이 녹지로부터 반경 500m까지 미친다는 선행연구를 토대로 기준을 설정하면 시가지 중 100m당 0.1℃ 이상의 기온 저감 영향을 받는 면적은 67.33km <sup>2</sup> 로 전체 서울시 면적의 11.12%, 시가지 면적을 기준으로 할 경우 18.09%를 차지
공원녹지 분야의 탄소흡수원 확보 및 탄소저감방안 <sup>51)</sup>	김원주 김운수 (2010)	서울의 공원녹지 정책은 공원녹지의 양적 확충에 주안점을 두고 수립되어 왔으나, 기후변화에 대응하고 저탄소사회에 적응해 나아가기 위해서는 녹지의 미기후 조절 기능, 수면의 열섬저감효과등 직·간접적 탄소저감과 탄소흡수기능의 확대할 수 있도록 공원녹지의 질적 향상을 꾀함. 공원녹지 분야에서의 탄소흡수원 확보 및 탄소저감방안을 제시하고 있다. 우선 기초자료 분석을 통해 서울시 공원녹지 자원의 탄소저장량을 산정하고 서울시 기후변화 경향과 열섬효과를 분석하고 있으며, 또한 공원녹지 분야의 탄소흡수원 확보 및 탄소저감방안에 관한 국내외 사례를 조사한 후, 각각에 대한 시사점을 도출

## 4. 스마트공원 구현을 위한 IoT, 인공지능 기술

### 4.1 IoT 환경 데이터 장비 연구 동향

기존에 온도계, 습도계로만 가버이 여겨졌었던 환경 센서는 IoT 기술을 바탕으로, 매 순간 매초의 공기질 관련 환경 데이터들을 빅데이터로 정리하여 클라우드

49) 박종화, 조기혁(2016), 「공원 크기에 따른 공원의 온도 저감 효과 분석: 일산 신도시를 대상으로」, 『국토계획』, 51(5), 247-261.

50) 윤민호, 안동만(2009), 「위성영상을 이용한 도시녹지의 기온 저감 효과 분석」, 『한국조경학회지』, 37(3), 46-53.

51) 김원주 외(2010), 「공원녹지 분야의 탄소흡수원 확보 및 탄소저감방안」, 서울연구원 정책과제연구보고서, 1-155.

서버에 전송한다. IoT 환경 데이터 수집 장비가 이처럼 발전하게 된 배경에는 최근 공기질 중에서 가장 많은 이슈가 있었던 미세먼지의 급격한 증가가 있다. 미세먼지는 세계보건기구(WHO)에서도 1군 발암물질로 분류하고 있으며, 미국 유럽 일본 등 선진 각 국에서도 주거용 건축물 등에 실내 가이드라인을 설정해 운영 중이다<sup>52)</sup>. 우리나라는 미세먼지에 의한 건강 위협이 국제적으로 심각한 수준이다. 1년 중 대기질 기준( $25\mu\text{g}/\text{m}^3$ )을 초과하는 날을 기준으로 평가할 때, 우리나라는 23.83일(OECD 평균 12.35일)로, 중 대기질이 가장 나쁜 상황이다. 대기오염에 대한 노출된 인구 비율 기준으로 대기질 기준( $25\mu\text{g}/\text{m}^3$ )을 초과하는 농도에 노출된 인구 비율이 47%로 OECD 국가(평균: 3%) 중 가장 높다. 우리나라는 대기오염을 획기적으로 줄이지 못하면 조기 사망률이 향후 40년 이후에는 현 수준보다 3.1배 급증할 것이라는 경고를 받은 바 있어 대기오염 물질 배출에 대한 대책 마련이 요구된다<sup>53)</sup>.

본 연구는 이 같은 문제를 도시 공원의 스마트화 차원에서 다루고자 하며, 이에 따라 본 연구에 적용할 수 있는 국내 IoT 공기질 측정 장비 개발 관련된 연구 방법을 중심으로 선행연구를 검토하였다. 검토 과정에서는 선행 연구들의 연구 방법을 집중적으로 검토하였으며, 이를 바탕으로 본 연구의 방법을 마련하는 데 활용하고자 하였다.

오창석 외 5인(2015)은 서울시의 ‘2011~2014년 7월 다중이용시설 실내 공기질 측정 결과에서 대부분의 다중이용시설이 실내공기오염도 기준을 위반했다는 점을 검토하고, 이러한 실내에서 발생하는 대기오염은 폐 질환, 천식, 기관지염 등의 질병을 유발한다는 점을 지적하였다. 이에 환경 센서를 이용하여 실내공기질을 측정하고, 측정된 데이터를 분석하여 환기 개선에 필요한 액추에이터 신호를 생성하고 실시간으로 실내 공기질 측정 모니터링 시스템을 구현하여, 생활에서 대기오염을 자율적으로 관리하는 방안을 제시하였다.

김종현 외 3인(2018)은 공기오염의 위험성 경고를 위한 실내 공기 측정 및 시각화 시스템을 제안하였다<sup>54)</sup>. 실내 활동이 잦은 계절에는 히터 사용과 환기의

52) 이은택 외(2019), 「공동주택의 실내 미세먼지 저감을 위한 기술개발」, 『대한설비공학회 학술발표대회논문집』, 111-112.

53) 임상택 외(2019), 「지능형 AI기반의 미세먼지 저감 제어 시스템」, 『한국컴퓨터정보학회 학술발표논문집』, 27(1), 53-56.

54) 김종현 외(2018), 「실내 공기 개선을 위한 공기 측정 및 시각화 시스템 설계」, 『In Proceedings of KIIT Conference』, 128-131.

부족으로 인해 실내 공기 오염도가 심각하다. 또한 미세먼지와 같은 대기환경 오염의 원인이 증가하여 실외뿐 아니라 실내에도 영향을 끼친다. 그러므로 실내 상태를 주기적으로 확인하여 환경을 개선할 필요성이 요구된다. 해당 연구에서는 아두이노, 라즈베리파이 및 다양한 센서를 이용하여 실내 환경의 자료를 수집하여 시각화를 통해 대기환경 오염의 위험성을 경고해주는 시스템을 제안하였다.

김명진 외 4인(2019)은 미세먼지가 사회적으로 심각하게 문제시 되는 상황에서, 실내·외에 구애받지 않고 미세먼지 측정기를 활용하여, 미세먼지 수치를 확인하고 이에 대처할 수 있는 시스템을 제안하였다<sup>55)</sup>. 해당 시스템 구성은 크게 세 가지 단계로 구성된다. 첫째, 실내 공기질 수치를 파악하기 위해 개발한 미세먼지 측정기에 부착된 미세먼지 센서와 온도 센서를 이용하여 데이터를 수집한다. 둘째, 수집된 미세먼지 데이터에 세계보건기구의 미세먼지 기준치를 적용하여 네가지색의 LED 등과 OLED Display Module 의 텍스트로 시각화하여 실내 공기질 정보를 제공한다. 셋째, 누적된 데이터를 기반으로 안드로이드 앱상에서 실내 공기질의 변화를 그래프로 확인할 수 있다. 또한 사용자가 공기질의 급격한 악화 구간에 대한 피드백을 직접 기록하고 공유할 수 있도록 하였다.

김우형 외 2인(2021)은 아두이노 기반의 미세먼지 농도측정기를 소형화하여 제작하고 블록체인 서버 방식을 통해 이더리움 블록체인에서 제공해주는 프로그램을 사용하여, 웹 화면과 LCD 화면에 동시에 디스플레이하도록 하였다<sup>56)</sup>.

손용훈 외 5인(2021)은 환경부 기후변화 사업으로 경기도 의왕시, 인천광역시 계양구, 경상북도 예천군의 어린이 놀이공원을 대상으로 하여 기후변화 저감 시설인 쿨루프, 쿨페인트, 쿨링포그에 대한 모니터링을 시행하고 이에 대한 효과성을 분석하였다. IoT장비의 연구 방법 및 사례를 소개하고 대상지에 IoT장비 및 데이터 로거를 설치하여 폭염저감 시설에 대한 모니터링을 통해 데이터를 수집하고 분석했다. 또한, 수집한 데이터를 바탕으로 대상지의 열쾌적성 평가 지수인 UTCI 를 통해 효과성을 도출하여 대상지에 적용된 폭염저감 시설인 쿨루프, 쿨페인트, 쿨링포그에 대하여 평가하고 효과적인 폭염 대응 방안을 제시하였다.

지금까지 살펴본 기존연구에서의 IoT와 환경데이터를 이용한 사례를 이용하여

---

55) 김명진 외(2019). 「실내 공기질 파악 및 관리에 관한 연구」, 『한국정보과학회 학술발표논문집』, (), 1673-1675.

56) 김우형 외(2021), 「블록체인 기반 지능형 미세먼지 감지 시스템」, 『한국정보통신학회 여성 ICT 학술대회 논문집』, 86-89.

연구방법을 수립하는데 활용하였다. 하지만 아쉽게도 선행 연구들에서 도시공원을 중심으로 실외공기질을 정량화 하는 부분에 대한 기술을 가지고 IoT장비와 인공지능을 적용한 사례는 없었다.

표 8 IoT기반 실내외 공기질 모니터링시스템에 관한 선행연구

논문명	연구자	연구내용 및 방법
IoT 기반 실내 공기질 모니터링 시스템 <sup>57)</sup>	오창세, 서민석, 이정혁 외 (2015)	서울시의 '2011~2014년 7월 다중이용시설 실내 공기질 측정 결과' 에 의하면 대부분의 다중이용시설에서 실내공기 오염도 기준을 위반한 것으로 조사되었다. 이러한 실내에서 발생하는 대기오염은 폐 질환, 천식, 기관지염 등의 질병을 유발하며 본 논문에서는 이와 관련하여 환경 센서를 이용하여 실내공기질을 측정하고, 측정된 데이터를 분석하여 환기 개선에 필요한 액추에이터 신호를 생성하고 실시간으로 실내 공기질 측정 모니터링 시스템을 구현하여, 생활에서 대기오염을 자율적으로 관리.
실내 공기 개선을 위한 공기 측정 및 시각화 시스템 설계 <sup>58)</sup>	김종현, 한진주, 김장원 외 (2018)	공기오염의 위험성 경고를 위한 실내 공기 측정 및 시각화 시스템을 제안. 실내 활동이 잦은 계절에는 히터 사용과 환기의 부족으로 인해 실내 공기 오염도가 심각함. 또한 미세먼지와 같은 대기환경 오염의 원인이 증가하여 실외뿐 아니라 실내에도 영향을 끼침. 그러므로 실내 상태를 주기적으로 확인하여 환경을 개선할 필요성이 요구됨. 이 논문에서는 아두이노, 라즈베리파이 및 다양한 센서를 이용하여 실내 환경의 자료를 수집하여 시각화를 통해 대기환경 오염의 위험성을 경고해주는 시스템을 제안..
KEI 기후친화형 어린이 쉼터 모니터링 보고서 <sup>59)</sup>	손용훈, 권태경, 김도은 외 (2021)	환경부 기후변화 사업으로 경기도 의왕시, 인천광역시 계양구, 경상북도 예천군의 어린이 놀이공원을 대상으로 하여 기후변화 저감 시설인 쿨루프, 쿨페인트, 쿨링포그에 대한 모니터링을 시행하고 이에 대한 효과성을 분석함. IoT장비의 연구 방법 및 사례를 소개하고 대상지에 IoT장비 및 데이터 로거를 설치하여 폭염저감 시설에 대한 모니터링을 통해 데이터를 수집하고 분석했음. 또한, 수집한 데이터를 바탕으로 대상지의 열쾌적성 평가 지수인 UTCI를 통해 효과성을 도출하여 대상지에 적용된 폭염저감 시설인 쿨루프, 쿨페인트, 쿨링포그에 대하여 평가하고 효과적인 폭염 대응 방안을 제시함

57) 오창세 외(2015), 「IoT 기반 실내 공기질 모니터링 시스템」, 『한국통신학회논문지』, 40(5), 886-891.

58) 김종현 외(2018), 「실내 공기 개선을 위한 공기 측정 및 시각화 시스템 설계」, 『In Proceedings of KIIT Conference』, 128-131.

59) 손용훈 외(2021), 「KEI 기후친화형 어린이 쉼터 모니터링 보고서」, 서울대학교.

## 4.2 기후변화 대응 정량평가 환경 데이터 수집용 센서

최근 스마트공원 관리뿐만 아니라, 기후변화 대응 도시 공원 내에서의 그늘막과 열 환경 개선 시설의 효과를 평가하기 위해서 다양한 연구들이 진행되고 있다. 하기 국내 선행 연구에서는 주로 데이터 로거를 사용한 연 관류량과 온도측정을 통해서, 쿨루프 열적 특성 측정 방법을 제안하였으며, 폭염 모니터링 시뮬레이션을 위한 방법으로 열적외선 카메라를 탑재한 드론 등의 장비를 활용하여 원격으로 지붕 표면온도 분포자료를 측정하였다.

조진희 외 3인 (2018) 은 「미시적 도시공간의 폭염 취약 지역분석을 통한 정책적 대응 방안」 에서 청주시 도심지역을 중심으로 공원 등 도시 계획시설의 배치 및 구현전략을 통해서 미시적 시뮬레이션을 수행하였다<sup>60)</sup>.

한편 이근상 외 1인 (2017) 은 UAV 열적외선 영상을 이용한 지붕 재질별 온도분석에서 열적외선 영상을 이용하여, 전도율에 따른 재질과 쿨페인트 적용에 따른 영향을 시뮬레이션하여 효과를 입증하였다<sup>61)</sup>.

또한 해외의 연구에서는 차열 포장의 효과를 측정하기 위하여 동일한 온도 조건에 5가지 포장재의 온도감소 효과를 알베도 측정기를 사용하여 측정하였다. 광학적 특성과 열전특성의 두 가지 기준으로 코팅의 분석을 가시광선과 적외선 사진을 촬영하여 실험 등이 존재한다<sup>62)</sup>. 이와 같은 선행 연구들은 크게 데이터 로거와 온도 측정기를 활용하여 실외 기온, 실내 기온, 표면온도 등을 측정한 연구, 혹은 드론이나 시뮬레이터 등의 복합적인 장비를 활용하여 온도 등을 측정한 연구로 나누어서 살펴볼 수 있다.

그러나, 이처럼 효과 측정과 관련해서는 다양한 연구와 방법이 시도되었으나 여전히 각 스마트공원의 기후변화 대비 대기 측정 요소 기술에 대한 체계적 관리와 효율성에 대한 검증은 부재한 상황이다. 그리하여 본 논문에서는 현재 각 지자체에서 스마트공원을 구성하고 기상 장비를 사용하여 수행하고자 하거나 사용하

---

60) 조진희, 이채연, 배민기, 오후. (2018). 미시적 도시공간의 폭염취약지역분석을 통한 정책적 대응방안. 충북 FOCUS, (), 1-25.

61) 이근상, 이종조(2017), 「UAV 기반 열적외선 센서를 이용한 태양광 셀의 발열 검출」, 『대한공간정보학회지』, 25(1), 71-78.

62) Wang, K., Yang, L., & Kucharek, M. (2020). Investigation of the effect of thermal insulation materials on packaging performance. Packaging Technology and Science, 33(6), 227-236.

려고 할 때 경제적인 부분 등을 고려하여, 저가격 범용성의 IoT와 클라우드 장비를 개발·운영하는 방법을 제공하고자 한다<sup>63)</sup>.

(1) 데이터 로거

TR-71wf(Wireless LAN)  
TR-71nw(Wired LAN)



그림 7 T&D TR-71WF

T&D사의 TR-71WF 데이터 로거를 통한 온도측정의 경우 이전의 국가 및 지자체에서 주로 사용되었다. 사용방법 및 절차는 진행된 그늘막, 쿨링포그, 차열 페인트 등의 폭염 대응시설에 대한 효과를 측정할 때와 기존의 연구들에서 주로 활용되었던 방법으로 온도를 측정하고자 하는 위치에 데이터 로거와 표면 온도계를 설치하여 해당 지점의

온도를 측정하고 장비 내부 혹은 클라우드 서버에 저장하는 방식으로 이루어졌다. 실제 논문사례에서도 사용사례가 빈번히 공개되어있다.

데이터 로거를 통한 효과 측정은 다소 저렴한 비용으로 지점 온도를 측정할 수 있다는 장점이 있으나, 스마트공원 내 다양한 온도 저감 효과를 측정하는 것에 적합하지 않은 부분들이 많았다. 첫째, 데이터 로거와 표면 온도계의 경우 바닥면에 부착하는 형태로 설치할 수밖에 없으므로 표면 온도를 측정하는 것에는 적합하나 실온의 온도를 측정하기는 어렵다. 특히 표면온도와 일정 높이에서의 온도는 큰 차이가 있어서 데이터 로거를 통해서 실제 공원에 제공되는 그늘막을 통해 사람들이 체감하는 위치인 바닥에서의 1.5~1.7m 높이의 온도를 측정하기가 어렵다. 둘째, 장비의 유지나 관리가 쉽지 않다. 데이터 로거의 경우 AA 배터리로 전원을 공급하고 또한 24시간 동안 장비가 운용되다 보니 장기간의 데이터를 수집하기 위해서는 전원을 직접 교체해주어야 한다. 또한 장비의 크기가 작고 외부에 노출된 형태로 설치되다 보니 폭우나 강풍 등의 기상 조건에 위치가 바뀌거나 장비가 고장이 나는 등의 문제가 발생할 수 있다. 그림 8 참조. 이러한 문제점은 야외에 장비를 놓고 측정하여야 하는 도시 그늘막 사업에 있어 적합하지 않음

63) 이정훈 외 (2018), 「4차 산업혁명 경기도 모델 구축 및 실행계획 연구: 실행계획」. 경기연구원, 1-399.

며, 기존의 유사 사업들에서도 이와 같은 문제들이 지적된 바 있다.



그림 8 현장에 적용된 데이터 로거 사례

## (2) 열화상 카메라



그림 9 열화상 카메라 E6  
출처 :FLIR사

열화상 카메라는 시중에서 널리 활용되는 FLIR사의 E6 열화상 카메라를 사용하였다. 이전의 국가 및 지자체에서 진행된 그늘막, 쿨링포그, 차열페인트 등의 폭염 대응시설에 대한 효과를 측정할 때 측정 대상이 되었던 일반 직사 조영부, 그늘막 시뮬레이터, 그리고 그늘막 음영 부에 설치 후에 비교를 진행하였다. 사용된 적외선 열화상 카메라는  $-20^{\circ}\text{C}\sim 250^{\circ}\text{C}$  범위에서 표면 온도 측정이 가능하며 0.1~1.0의 방사율 선택 및 보정 기능이 있다. 이미지 분해능은  $240\times 180$  픽셀로 총 43,200픽셀의 고해상도 컬러화면으로 표면 온도를 측정한다. 영상을 활용하여 광범위한 부분의 온도를 일시에 측정할 수 있어서 매우 많은 분야에 활용되고 있다. 하지만, 가격이 일반적으로 \$2,500 내외로 고가인 점, 온도 이외에 다른 습도와 공기질에 대한 부분은 동시에 처리되지 못하는 부분과 실시간으로 영상을 업로드하여 처리하지 못하는 단점 등이 있다. 표 9는 일반적으로 사용되는 범용 측정장비를 보여준다.

표 9 다양한 온도 측정 장치

	Device		Model	Number
Measuring Devices	surface thermometer		TR-0206	1
	data logger		TR-71WF	1
	thermo-hygrometer		T006	1
	Thermal Imaging Camera		FLIR E6 WiFi	1

### 4.3 인공지능 선형회귀분석을 통한 데이터분석

최근 빅데이터의 개념은 대용량 데이터를 활용·분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술을 포괄하는 용어로 변화하고 있다. 빅데이터를 수집, 저장, 처리, 분석하는 것뿐만 아니라 이를 통한 새로운 가치 창출의 전 과정을 포괄하는 독립된 영역으로 확대되고 있다<sup>64)</sup>. 스마트공원 안에서 취득되는 다양한 환경 데이터는 인공지능기술을 활용하여, 공원 관리에 많은 도움이 될 수 있다. 가령, 평소에 공공 자전거 활용도의 빅데이터를 이용하여, 평일 및 공휴일의 수요를 인공지능으로 예측하여 사전에, 공원에 공공 자전거를 효율적으로 배치하는 것도 가능하고, 보행량을 기반으로, 공기질 또는 날씨에 상관관계에 기반하여, 공원 유동 인구에 대응하는 행사를 제공하거나, 사용자에게 쾌적한 공원 관리에 더욱 힘쓸 수 있다. 빅데이터와 인공지능은 지식기반사회에서의 다양한 사회현상을 다룸으로써 분석 주체를 비록 이해관계자들에게 통찰력, 대응력, 경쟁력, 그리고 창조력 등의 가치를 제공하는 역할을 한다<sup>65)</sup>.

이와 관련된 일례로, 뉴욕시는 세계 최고의 과학기술 중심지로 만들기 위해 대학들의 컨소시엄으로 구성된 도시지능연구소(UIL)을 통해 NYC311 데이터뿐만 아니라 뉴욕시 빌딩 데이터, 에너지 소비 데이터, 대기질 데이터, 와이파이 데이터, 신용카드 데이터, 교통 데이터, 통계 데이터 등을 스마트화하여 지속가능한 도시로의 구축을 진행하고 있다<sup>66)</sup>. 이렇게 수집된 빅데이터를 분석함으로써 도시 운영 및 관리에 있어 필요한 다양한 예측을 수행할 수 있다. 이렇듯 빅데이터의 활용 영역은 학술적 연구는 물론 산업 및 비즈니스, 공공분야에 이르기까지 그 활용의 범위는 매우 광범위하게 이루어지고 있음을 알 수 있다<sup>67)</sup>.

서경덕(2020)은 인공지능을 사용하여 저가형 센서 데이터가(PM2.5) 고가형 센서와 근접한 성능을 내는 방법을 제안하였다<sup>68)</sup>. 또한 고가형 센서의 데이터를

64) 장혜정(2020), 「시민 니즈와 참여 기반의 스마트시티 문제 해결을 위한 빅데이터 활용 절차에 관한 연구」, 『한국정보전자통신기술학회 논문지』, 13(2), 102-112

65) 이현구, 오재인(2017), 「스마트시티 서비스 유형분류」, 『한국경영과학회지』, 42(4), 147-161.

66) 김익희(2018), 「뉴욕, 데이터 중심 스마트시티」, 『국토』, 443, 98-105.

67) 서창수 외(2020), 「지속가능한 스마트시티 서비스 활성화를 위한 빅데이터 우선순위에 관한 연구」, 『도시행정학보』, 33(4), 61.

68) 서경덕(2020), 「저가형 초미세먼지 센서 정확도 향상을 위한 인공지능 기반 센

학습하여, 저가형 센서의 데이터를 보정해주는 여러 가지 인공지능 기법을 소개하였다. 성능평가 방법은 Root Mean Squared Error를 이용하였으며, 사용한 머신러닝 방법으로는 선형 회귀 기법, 다항식 회귀 기법, 랜덤 포레스트 기법, Gradient Boost 기법과 딥러닝인 Long Short-Term Memory(LSTM) 방법을 이용하였다.

김규석 외 1인은, 인공지능 모형중 다중회귀분석을 통하여, 공공 자전거 이용자 수를 예측하는 데 사용하였다. LSTM은 학습 데이터의 지속성을 갖는 RNN(Recurrent Neural Network) 기법의 단점인 입력과 출력 사이의 거리가 멀수록 연관 관계가 적어지는 장기 의존성 문제를 개선한 기법이다. 연구 결과, LSTM 모형은 다른 모형들에 비해 상대적으로 정확도가 높아 서울시 공공 자전거 이용자 수요예측을 위해 적합한 것으로 판단하였다<sup>69)</sup>.

이렇게 인공지능 모델링에 적용되는 다양한 알고리즘이 소개되면서, 인공지능 모델링의 성능을 비교하고 최적의 솔루션을 찾고자 하는 선행연구도 많이 시도되었다.

표 10 선형회귀분석을 통한 데이터 분석에 관한 선행연구

논문명	연구자	연구내용 및 방법
저가형 초미세먼지 센서 정확도 향상을 위한 인공지능 기반 센서 데이터 예측 기법 <sup>70)</sup> .	서경덕 (2020)	인공지능을 사용하여 저가형 센서 데이터가(PM2.5) 고가형 센서와 근접한 성능을 내는 방법에 대해 제안함. 고가형 센서의 데이터를 학습하여, 저가형 센서의 데이터를 보정해주는 여러 가지 인공지능 기법을 소개함. 성능평가 방법은 Root Mean Squared Error(이하 RMSE)를 이용하였으며, 사용한 머신러닝 방법으로는 선형회귀 기법, 다항식 회귀 기법, 랜덤 포레스트 기법, Gradient Boost 기법과 딥러닝인 Long Short-Term Memory(LSTM) 방법을 이용하였음
코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안 <sup>71)</sup>	김규석, 서예진 (2021)	인공지능 모형중 다중회귀분석을 통하여, 공공 자전거 이용자 수를 예측하는 데 사용하였음. LSTM은 학습 데이터의 지속성을 갖는 RNN(Recurrent Neural Network) 기법의 단점인 입력과 출력 사이의 거리가 멀수록 연관 관계가 적어지는 장기 의존성 문제를 개선한 기법. 연구 결과, LSTM 모형은 다른 모형들에 비해 상대적으로 정확도가 높아 서울시 공공 자전거 이용자 수 예측을 위해 적합한 것으로 판단함

서 데이터 예측 기법」, 『한국정보기술학회 종합학술발표논문집』, 179-183.  
 69) 김규석, 서예진(2021), 「코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안」, 『한국정보기술학회논문지』, 19(10), 11-18.  
 70) 서경덕(2020), 「저가형 초미세먼지 센서 정확도 향상을 위한 인공지능 기반 센서 데이터 예측 기법」, 『In Proceedings of KIIT Conference』, 179-183.  
 71) 김규석, 서예진(2021), 「코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안」, 『한국정보기술학회논문지』, 19(10), 11-18.

## 4.4 인공지능 모델링 분석 고찰

### 4.4.1 수요 예측기법

일반적으로, 예측이란 불가피하며 이를 바탕으로 대부분의 의사결정이 이루어진다. 또한 예측은 개인의 사사로운 결정뿐 아니라 가구나 기업의 수·전력, 교통수단, 관광객, 기존 제품 및 신제품의 수요 등을 결정하는 데 중요한 역할을 한다.

인공지능을 활용한 전통적으로 쓰이고 있는 수요 예측기법에는 시계열분석법과 회귀분석법이 있다. 시계열분석법은 일정한 시간 간격을 두고 기록된 데이터를 바탕으로 미래를 예측하며 대표적으로 Auto Regressive(AR), Moving Average(MA), Auto Regressive Integrated Moving Average(ARIMA), Exponential Smoothing(지수평활법) 등이 있다.

회귀분석법은 단일 또는 다수의 독립변수와 종속변수 간의 관계를 찾기 위해 관계식을 추정하고, 추정된 관계식으로 독립변수들을 통해 미래를 예측하는 방법이다<sup>72)</sup>.

인공지능 기반 수요 예측기법의 가장 큰 장점은 예측의 정확도와 계산량의 효율성이라 할 수 있다. 기존 수요 예측기법에 비해 상대적으로 높은 정확도를 보이며 예측실행 시간도 덜 소요되기 때문이다. 또 독립변수와 종속변수 간의 관계를 나타내는 모델을 설정할 필요 없이 알고리즘을 구축할 수 있다<sup>73)</sup>.

머신러닝 기반의 수요 예측 기법으로는 Support Vector Regression(SVR), 부스팅(Boosting), 가우시안 프로세스(Gaussian Process(GP))이 있다. 딥러닝 기반의 수요 예측 기법으로는 Recurrent Neural Networks(RNN) 및 long short-term memory models(LSTM), Restricted Boltzmann Machine(RBM) 등을 기반으로 변형 및 발전된 다양한 기법들이 있다.

### 4.4.2 선행 연구사례

---

72) 정혜린, 임창원(2019), 「인공지능 기반 수요예측 기법의 리뷰」, 『응용통계연구』, 32(6), 795-835.

73) 정혜린, 임창원(2019), 「인공지능 기반 수요예측 기법의 리뷰」, 『응용통계연구』, 32(6), 795-835.

머신러닝은 다양한 연구 분야에서 예측력 및 속도가 뛰어나 활용되고 있다. 특히, 머신러닝은 시계열 기법에 비해 다변수 데이터를 적용될 시 비약적인 성능 향상을 보여주며, 정부 정책 측면에도 활발히 활용되고 있다<sup>74)</sup>. 허경호 외 1인(2020)은 중력 모델을 이용한 전주시 근린공원의 수요 예측을 하였다. 하지만, 인공지능을 이용한 스마트공원의 보행량 예측모형과 가장 유사한 모델은 공유자전거 수요 예측 모델이다. 이와 관련하여 선행된 연구들이 다수 존재한다.<sup>75)</sup>.

공유자전거 수요 예측에는 다른 분야에서의 수요 예측과 마찬가지로 시계열 방법론이나 머신러닝 방법론이 많이 활용되고 있다<sup>76)</sup>. Kaltenbrunner et. al(2010)에서는 Auto-Regressive Moving Average(ARMA)를 활용해 대여소의 비치 자전거 대수를 예측하는 방법을 제안했다<sup>77)</sup>. 하지만 시간에 따라 비치 자전거 대수의 편차는 변동이 커서 시계열이 안정적(stationary)이라는 ARMA의 가정이 위배되는 문제점이 있다. 이에 Yoon et al, (2012)에서는 ARMA를 대신에 Auto-Regressive Integrated Moving Average(ARIMA) 모델을 기반으로 한 방법을 제안하였다<sup>78)</sup>. 이 논문에서도 예측 정확도를 높이기 위해 주변 대여소의 정보를 사용하여 모델링하였다.

최승한 외 1인(2020)은 Recurrent Neural Networks(RNNs)는 시계열 데이터에서 자동으로 학습하기 위해 이전 단계의 네트워크 출력을 입력으로 사용하는 신경 네트워크의 유형이다<sup>79)</sup>. SimpleRNN은 이론적으로 모든 이전 타임스텝 정보를 t 시간 내에 유지할 수 있지만 실제로는 장기 의존성이 학습되지 않은 문제를 만든다. 즉, 피드포워드 자연망의 깊숙한 층일수록 배우기가 더

74) 황지환, 박홍석, 김재동(2022), 「양상블 학습 기반의 동시조달 수리부속 예측모형 연구」, 『한국산학기술학회 논문지』, 23(8), 431-438.

75) 허경호, 박종민(2020), 「중력 모델을 이용한 전주시 근린공원의 수요 예측」, 『한국환경생태학회 학술발표논문집』, 2020(1), 35-36.

76) 김경옥, 이창환(2021), 「공유자전거 시스템의 이용 예측을 위한 K-Means 기반의 군집 알고리즘」, 『정보처리학회논문지』, 10(5), 5.

77) Kaltenbrunner, et al.(2010), “Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system,” *Pervasive and Mobile Computing*, 6(4), 455-466.

78) Yoon, J. W et al.(2012). “Cityride: a predictive bike sharing journey advisor,” *IEEE 13th International Conference on Mobile Data Management*, IEEE, 306-311.

79) 최승한, 한미경(2020), 「어텐션 적용 LSTM을 사용한 자전거 수요 예측 모델의 구현」, 『한국통신학회 학술대회논문집』, 342-343.

어렵다고 하여, Long Short-Term Memory (LSTM)는 장기 의존성 문제를 피하기 위해 설계된 아키텍처 기반 RNN이며 LSTM의 핵심은 네트워크가 무엇을 기억하고, 무엇을 삭제하며, 무엇을 장기적 상태로 읽어야 하는지 학습한다고 하였다. 특히 LSTM에서 LSTM 2개 계층 모델의 오차보다 LSTM 1개 계층에 어텐션 계층을 결합한 모델이 평균 MSE(mean square error)와 MAE(mean absolute error) 오차가 모두 적다는 것을 발견하였다.

민소아 외 1인(2021)은 다변량의 시계열 데이터로서 시계열 데이터 군집 분석(time series clustering), 서포트 벡터 회귀(support vector regression), LSTM 네트워크(long short-term memory networks)의 분석 방법론을 통해 서울시 공공 자전거 수요 예측을 하였으며, LSTM은 학습 데이터의 지속성을 갖는 RNN(Recurrent Neural Network) 기법의 단점인 입력과 출력 사이의 거리가 멀수록 연관 관계가 적어지는 장기 의존성 문제를 개선한 기법이라 하였다<sup>80)</sup>.

김규석 외 1인(2021)은 코로나 상황에 따른 PM(personal mobility) 수요에 계절, 강우량, 요일 등을 활용하였을 뿐 아니라, 다중 선형회귀 분석을 수행하여, 변수들이 인과관계를 측정하고, MAPE 값을 구하여, 해당 머신러닝 모형의 예측 정확도를 측정하였다<sup>81)</sup>.

김예슬 외 2인(2020)은 공공 자전거 신규대여소 입지선정을 위하여, 선형회귀 모형(linear regression model)과 랜덤 포레스트(random forest)를 고려하였다. 선형회귀 모형은 종속변수와 독립변수 간의 선형성 가정하에 계산이 쉽고 빠르며, 결과에 대한 설명이 용이하다는 장점이 있는 반면 랜덤 포레스트는 해석이 어렵다는 단점은 있지만, 예측력에서 우수한 성능을 보인다. 특히 랜덤 포레스트는 종속변수와 독립변수 간의 선형성 가정이 필요하지 않은 점을 장점으로 가지고 있다고 하였으며, 데이터 전처리에 대한 부분을 강조하였다<sup>82)</sup>.

김세형 외 3인(2020)은 기존의 연구들과 차별화를 두어 영상 이미지 기반의 CNN(convolutional neural network) 기반의 BikeNet 과 Night Light를 활

---

80) 민소아, 정윤서(2021), 「서울시 공공자전거 수요예측 모형 비교 연구」, 『한국 데이터정보과학회지』, 32(3), 585-592.

81) 김규석, 서예진(2021), 「코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안」, 『한국정보기술학회논문지』, 19(10), 11-18.

82) 김예슬 외(2020), 「서울 공공자전거 신규 대여소를 위한 수요량 예측 분석」, 『응용통계연구』, 33(6), 739-751.

용한 Transfer Learning을 통해 예측하는 방법론을 제안하였다<sup>83)</sup>.

한편, 최근 인공지능을 위한 인공지능<sup>84)</sup>이라 불리는 Auto ML을 이용한 회귀 분석과 예측이 많이 이용되고 있다. 이와 관련하여, 성기우 외 2인(2021)은 Auto ML을 이용하여 자동차의 필드에서의 고장에 원인과 인과관계가 있는 다변량 요소들을 대입하여, 원인 인자들과의 영향도를 분석하고, 모형을 통해 빠른 예측과 원인분석을 하기 위한 방법을 제안하였다<sup>85)</sup>.

또한 최근에는 Auto ML에서 가장 최적의 성능을 보여주는 Ensemble 알고리즘을 활용하여, 다양한 하이퍼 파라미터와 자동화된 튜닝시스템을 활용한 다양한 분야의 연구가 진행되고 있다. 이에 힘입어서 2022년부터 앙상블 알고리즘에 대한 논문이 급속도로 증가하고 있다. 김한성(2022)은 앙상블 기법을 활용하여 호흡기 질환과 증상 분류를 호흡음에 기반한 CNN 기반 앙상블 모델을 사용하여, 호흡 질환의 연관성에 대하여 분석하였다<sup>86)</sup>. 김규태(2022)는 제조 산업에서 센서를 통해 수집되는 다중 데이터를 통해, 제조공장의 결함 데이터 감지에 대한 예측 모델로 다중계층 스택킹 앙상블(Stacking Ensemble)모델을 제안하였다<sup>87)</sup>.

김윤조 외 2인(2022)은 소아 흉부 영상에서 정상인지와 폐렴인 자를 구분하는 방법이 딥러닝 프로세싱에서 처리가 어려웠으나, 영상에서 밝기 값 정규화 및 다중 스케일이 가능한 앙상블 딥러닝 기법을 활용하여 폐렴을 자동 추출하는 기법을 제안하였다<sup>88)</sup>.

본 연구에서는 첫 번째 연구 대상지인 ‘서울숲공원’의 보행량과 공원 요소의 상관관계를 측정하기 위하여, 미세먼지(pm), 습도(humi), 온도(temp), 일산화

---

83) 김세형 외(2020), 「인공위성 이미지를 활용한 공유자전거 대여 수요 예측」, 『한국통신학회 학술대회논문집』, 182-183.

84) “삼성SDS. “인공지능을 위한 인공지능, AutoML”, [https://www.samsungsds.com/kr/insights/ai\\_automl.html](https://www.samsungsds.com/kr/insights/ai_automl.html)(검색일: 2022. 10. 15.)

85) 성기우 외(2021), 「Auto ML을 활용한 필드 클레임 예측 및 고장 주요 인자 분석」, 『한국신뢰성학회 학술대회논문집』, 132-132.

86) 김한성(2022), 「호흡 질환과 증상 분류용 CNN 기반 앙상블 모델 및 연관성에 따른 호흡 증상 분류 모델 연구」, 강원대학교 대학원 석사학위논문.

87) 김규태(2022), 「다중계층 스택킹 앙상블을 이용한 제조데이터의 결함 분류 모델에 관한 연구」, 성균관대학교 대학원 석사학위논문.

88) 김윤조 외(2022), 「흉부 X-선 영상에서 밝기 값 정규화 및 다중 스케일 폐-집중 패치를 사용한 앙상블 딥러닝 모델 기반의 폐렴 자동 분류」, 『정보과학회논문지』, 49(9), 677-685.

탄소(co)와 직접 센싱한 환경 데이터와 공공 기상 데이터를 활용하여 매 시간당 적운량(cloud), 풍속(windspeed), 강우량(rainfall) 등의 추가 기상 데이터를 시계열 데이터에 합성하였다. 여기에 보행량을 목표(target)로 한 후, 각 환경 요소의 상관관계를 인공지능을 활용하여 다수의 모델링 알고리즘으로 회귀분석을 수행하고, 모델링 방법 중 가장 우수한 성능을 보인 Voting Ensemble을 통해서 분석된 보행량과 상관관계가 높은 환경 요소와 그의 패턴을 살펴보도록 한다.

이어서 두 번째 연구 대상지인 ‘소래 습지 생태공원’의 온도 저감, 이산화탄소저감 등 기후변화에 대한 효과를 다양한 인공지능 모델링 분석 기법을 사용하여 분석하도록 한다. 특히 환경 데이터와 이산화탄소 발생의 상관관계를 분석하기 위해, 최근 각광받고 있는, Auto ML 머신러닝 모델을 이용하였다. 인공지능분석을 통한 머신러닝 알고리즘 모델별 예측 성능을 비교하고 그중에 가장 우수한 성능을 보인 Stacking Ensemble을 통해서 분석된 상관관계가 높은 환경 요소를 파악하고, 그의 패턴을 염습지, 갈대숲, 그리고 포구 사거리로 나누어 Regression 모델링의 비교 분석을 수행하도록 한다.

### III. 서울숲 공원을 대상으로 한 스마트공원 구현

#### 1. 서울숲공원: IoT와 인공지능을 이용한 공원 관리계획

서울숲은 서울특별시 성동구 성수동 1가 685번지 일대에 있는 공원이다. 본래 골프장과 경마장이 있던 곳으로, 시민들에게 휴식 공간을 제공하기 위해 2005년 6월에 완성하였다. 면적은 1,156,498m<sup>2</sup> 이며, 테마공원 5개로 구성되어있다<sup>89)</sup>. 한강과 증랑천이 공원 근처를 흐른다. 마포구 월드컵공원 (100만 평)과 송파구 올림픽공원 (50만 평)에 이어 서울에서 세 번째로 큰 공원이다. 관련한 보행자와 환경 데이터에 관한 연구로는 2021년 6월 한국 조경학회 제6차 웨비나를 통해서 ‘스마트공원 관리의 가능성 : IoT, Big Data와 AI를 활용한 서울숲 관리’ 를 통해서 소개된 적이 있다<sup>90)</sup>. 웨비나에서는 서울숲을 대상으로 보행자의 행동 특성을 파악하고, 향후 개선 및 학술의 기초자료로 활용하고자 분석한 내용을 공유했다. 코로나19 이후 사회적 거리두기로 관광지나 산업지역에 사람이 언제, 어디서, 얼마나 출입했는지 실시간으로 파악해 신속한 대응체계를 갖추기 위함이었다. 하지만 단기 조사로는 정밀한 분석의 한계성이 있다는 점을 언급하며, 다른 도시공원에서도 계측 체계를 마련해야 한다는 의견을 개진했다. 이렇게 서울숲공원을 스마트공원으로 탈바꿈하여 다양한 환경 데이터와 보행 자료를 수집하기 위한 계획 및 방법은 다음과 같다.

##### (1) 인공지능을 활용한 서울숲 빅데이터와 보행량 예측 분석

인공지능을 이용한 공기질과 보행량의 연구의 공간적 범위는 서울숲공원의 IoT 장비가 설치된 야외가족마당, 수변부, 방문자센터 후보 대상지 중에서, 전원과 통신 등의 문제가 없었던 가족마당의 데이터를 중심으로 스마트공원 모니터링 최종 대상지로 하였다.

서울숲의 효과적인 모니터링을 위해 몇 가지의 요소들을 검토하여 IoT 장비의

89) 이제이(2013), 「1인 미디어에 나타난 서울숲의 미시담론 연구-장소감을 중심으로」, 서울대학교 대학원 박사학위 논문.

90) 장재훈(2021), 「스마트공원, 정보 계측·시물레이션 중요」, 『환경과 조경』, 398,

위치를 선정하였다. 먼저 측정된 데이터 중에서 서울숲의 대표적인 가족마당의 공기질 정보와 보행 데이터를 이용하여 실제 공원 사용 수치인 보행량의 관계성을 찾는 것에 다음과 같은 목표를 두었다.

첫째, IoT 기반 탄소측정 장치를 통한 서울숲의 공기질 빅데이터와 보행량 빅데이터를 확보하고, 둘째, 빅데이터를 활용하여, 머신러닝을 이용한 공기질 평가 및 방문자 수 예측을 할 수 있는 최적의 AI 알고리즘을 확보하고, 셋째, 이와 같은 알고리즘의 최적화를 통하여, 인공지능의 예측 성능을 평가한다. 이렇게 수집된 데이터는 인공지능 모델링 기법을 활용하여<sup>91)</sup> 공원 내 온·습도, 미세먼지, 일산화탄소 등의 공기질 데이터와 같은 환경 계수와 그리고 코로나바이러스 감염자 수, 요일, 계절별 보행량 등의 사회적 계수를 활용하여, 사이트별 보행량과 시설의 운영에 미치는 상관관계 모델링 효과와 보행량 예측을 하고 이를 다시 검증하여 공원 운영 설계 시 고려할 사항 등을 살펴보고자 한다.

## (2) 서울숲 IoT 장비 공간적 범위와 시간적 범위

본 연구는 인공지능을 활용한, 방문자 예측을 위한 장기간의 연구를 위한 IoT 장비를 2021년 5월 21일 ~ 2022년 5월 26일까지 약 1년간 서울숲에 설치하여 모니터링하였다. 후보 대상지 내 IoT 장비는 초기 예정 모니터링 장소인 수변부, 가족마당, 방문자센터에 각각 설치하여 모니터링을 통해 온·습도, 미세먼지, 일산화탄소의 데이터를 수집하였다.

사전 연구 기간인 2020년 11월 ~ 2021 5월까지의 클라우드는 방식이 아닌, 외부 저장장치인 MicroSD에 저장하도록 하였으나 데이터 저장이 쉽지 않아, 데이터를 모두 정리하였다. 다시 장비를 재설계하고 정비하여 2021.5월부터 ~ 2022년 5월까지 클라우드 서비스에 접속하여 실시간 저장하였다.

본 연구에서는 서울숲공원 환경 데이터 수집 후보 대상지 3곳의 데이터를 다 사용하지는 못했는데, 전원이 수변부 특성상 누설전류로 인하여 잦은 전원의 차단으로 데이터의 손실이 컸던 수변부와 공원 외곽에 설치되어 다소 공원 내부의 환경 데이터로 활용하기 어려운 방문자센터의 데이터는 다음 연구에서 활용하는 것으로 하고, 측정 기간 안정적인 상태로 데이터의 수집이 된 야외가족마당의

---

91) 김용국 외(2020), 「스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향」, 『한국조경학회지』, 48(6), 1-15.

데이터만을 사용하기로 하였다.

표 11 서울숲 IoT 장비 모니터링 수집내용

데이터명	수집내용	수집 기간	수집목적
Temp	온/습도/미세먼지/일산화탄소 수집	2021년 5월 21일 ~ 2022년 5월 26일	수변부 주변 공기질 측정
Humi			
PM			
CO			
Temp	온/습도/미세먼지/일산화탄소 수집	2021년 5월 21일 ~ 2022년 5월 26일	가족마당 공기질 측정
Humi			
PM			
CO			
Temp	온/습도/미세먼지/일산화탄소 수집	2021년 5월 21일 ~ 2022년 5월 26일	방문자센터 공기질 측정
Humi			
PM			
CO			

표 12 서울숲공원의 IoT를 사용한 초기 환경 데이터 수집 후보 대상지

가족마당 water	가족마당 plaza	방문자센터 Visitor
		
		
 <p>많은 수목과 탁 트인 넓은 가족마당 위치 미세먼지 저감효과와 도시열섬현상 제거로 도시 내 온도를 낮추어주는 효과가 매우 클 것으로 기대</p>	 <p>수변부 위치 음이온으로 인한 미세먼지 저감 효과와 복사열 온도 저감 효과 등 기대</p>	 <p>수변부, 가족마당과 각각 약 400m 거리 도로 부에 바로 접해 있어, 일산화탄소, 미세먼지, 온도 등 비교대상지</p>

### (3) 보행량 데이터 수집

최학희 외 3인(2016)은 IoT 기반의 공원 방문자 수와 비례하게 VM(Virtual Machine)을 유동적으로 설계 Q-learning 기법의 자원 예측 모델을 이용, 기존 IoT공원의 VM자원 할당 모델에 적용하였다<sup>92)</sup>. 하지만 이 부분은 보행량을 직접적으로 수집하는 것이 아닌 가상환경에서 시뮬레이션하여 공원자원의 할당량을 효율성 있게 하는 부분에 초점이 맞추어 있다.

물리적으로 기존 보행량 측정은 기존 카메라나, 적외선 방식 등이 있으나, 각각 사생활 침해 이슈와, 정확도 등이 다소 떨어지는 등 직접적인 사용에는 애로가 있다. 본 연구에 사용된 보행 데이터 센서는 아이데이터 주식회사의 협력으로 서울숲공원 안에 IoT 공기질 측정기가 설치된 위치와 가장 가까운 지역에 설치가 되었고, 감지 방식은 인체 체온 감지 방식을 사용하였다. 본 장비는 10분 간격이나 60분 간격으로 보행자 정보를 수집하였다가 LTE-M망을 통해서 자체 클라우드 서버에 업로드하도록 되어있다.

한편 보행량 데이터는 표 13 위치인 포레 아파트 입구, 수변부, 방문자센터와 에서 측정하였다. 이렇게 측정된 3개의 보행 데이터 중, 본연구에서는 IoT 센서 설치 위치 중 공원 가족마당와 가장 가까운 포레 아파트 측의 보행 데이터만을 매칭하여 사용하였고, 측정 기간에 시간당 보행량의 유효데이터가 수집된 2021년 5월 21일 ~ 2022년 1월 26일까지 적용된 점을 감안하여, 가족마당의 IoT 데이터 또한 시계열성을 동일하게 유지하기 위하여, 2021년 5월 21일 ~ 2022년 1월 26일까지 데이터만 같이 병행하여 사용하게 되었다.

---

92) 최학희 외(2016), 「사물인터넷(IoT) 기반의 공원을 위한 자원 예측 모델 설계 및 구현」, 『한국통신학회 학술대회논문집』, 391-392.

표 13 서울숲 보행량 계수기와 보행량 계수기 설치 위치



93) 사진 : 아이데이터 (주) 제공

표 14 People Counter 제품 사양서

Product spec			
감지 방식	인체 체온 감지	데이터수집	10분또는 60분간격으로 데이터 기록
커버리지	최대 10M	데이터 전송	6시간 데이터 전송
방향 감지	양방향/단방향	데이터 저장	12개월
설치조건	가로등, 가로수 등에 고정	배터리저장/전송	5년/2년

(출처 : 아이데이터 (주))

#### (4) 데이터 합성

머신러닝은 학습 데이터로부터 목적함수를 구성하고 테스트 데이터를 통해 목적함수를 확인함으로써 발생하는 데이터에 대한 예측을 수행한다. 머신러닝에서 입력 데이터는 전처리 과정을 통해 정규화 과정을 거친다<sup>94)</sup>.

IoT 데이터를 통해서 수집된 데이터는 수치형 데이터인 미세먼지(pm), 습도(humi), 온도(temp), 일산화탄소(co)에 해당한다. 이 데이터만을 이용해서 서

93) 아이데이터(주). <https://i-data.kr/products.php> (검색일: 2022. 10. 09.)

94) 황치곤, 윤창표(2020), 「머신러닝을 위한 온톨로지 기반의 Raw Data 전처리 기법」, 『한국정보통신학회논문지』, 24(5), 600-608.

울숲의 보행량의 Regression을 수행하기에는, 사실상 좀 더 많은 다변형 독립변수인 환경 데이터가 필요하다. 추가적인 환경 데이터에 관련하여서는 공공 기상 데이터를 활용하여 2021년 5월 21일부터 2022년 1월 26일까지의 시간당 적운량(cloud), 풍속(windspeed), 강우량(rainfall) 등의 추가 기상 데이터를 시계열 데이터에 합성하였다. 그리고 시계열 데이터의 기본인 일시 (data & time) 데이터를 이용하여, 년, 월, 일, 시간, 주간과 같은 주기 데이터를 별도로 유도한다. 여기에, 범주형 데이터인 계절 변수(season)와 휴일, 평일 여부를 표시해주는 데이터 칼럼을 추가하였다. 아울러 계절 더미변수의 변수 카테고리는 다음과 같이 정의하였다. 우리나라의 계절과 월간 절기는, 한국민족문화대백과사전에 정의된 것과 같이 3~5월은 봄은 '1', 6~8월은 여름은 '2', 9~11월은 가을은 '3', 12월~2월은 겨울은 '4' 로 설정하였다<sup>95)</sup>. 휴일 (hoiday)와 평일 (working day)의 상황에 해당하는 날을 '1' 로 설정하여 머신러닝에 사용할 수 있도록 하였다. 매주 요일을 데이터에서 활용할 수 있도록, 월요일 '0', 화요일 '1', 수요일 '2', 목요일 '3', 금요일 '4', 토요일 '5', 일요일 '6' 을 각각 변수로 지정하였다. 측정된 보행자 수도, 기존 데이터에 합성한 후에는 다음과 같은 데이터 포맷을 가질 수 있다. 따라서 그림 10 와 같이 전처리된 후에 데이터는 다음과 같은 형태로 된다. 추가로 머신러닝을 수행함에 있어서 원-핫인코딩(one-hot encoding)을 사용하여, 계절을 봄, 여름, 가을, 겨울로 구분하여 전처리한다.

---

95) 김규석, 서예진(2021), 「코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안」, 『한국정보기술학회논문지』, 19(10), 11-18.

```
train.head()

datetime season holiday workingday rainfall windspeed cloud pm humi temp co count year month day hour minute second
0 2021-05-21 11:00:00 1 0 1 0.0 3.0 10.0 23.97 46.37 27.93 6.04 113 2021 5 21 11 0 0
1 2021-05-21 12:00:00 1 0 1 0.0 2.2 10.0 25.44 44.49 28.97 6.21 78 2021 5 21 12 0 0
2 2021-05-21 13:00:00 1 0 1 0.0 1.6 10.0 21.50 41.40 30.65 6.58 94 2021 5 21 13 0 0
3 2021-05-21 14:00:00 1 0 1 0.0 2.0 10.0 19.38 43.61 27.45 5.95 58 2021 5 21 14 0 0
4 2021-05-21 15:00:00 1 0 1 2.0 1.4 10.0 14.42 49.28 24.87 5.88 78 2021 5 21 15 0 0
```

그림 10 공공데이터와 합성하여 전처리 된 데이터 출력

## 1.1 IoT 환경 데이터 수집 장비 개발과 고려사항

인공지능을 이용한 공기질과 보행량의 연구의 공간적 범위는 서울숲공원의 야외가족마당의 환경 데이터와 보행량을 도시 스마트공원 모니터링 대상지로 하였다. 본 연구에서는 하기와 같은 환경 데이터 모니터링 항목을 방법을 목표로, 실제 필요한 스마트공원의 구현을 위한 IoT 장비를 제작하고, 이를 각 공원에 적용하여 스마트공원의 관리를 도모하고자 한다. 표 15는 데이터 모니터링 항목 및 방법이다. 그리고, 이어서 데이터 모니터링 장비 개발에 따른 요구사항이 기술되어 있다. 아울러 각 지자체가 가지고 있는 저가의 예산과 함께 클라우드 서버를 이용한 관리를 용이하기 위하여, 현장에서 진행되었던 실제 측정 사례를 기준으로 최적의 솔루션을 제작할 수 있게 하여, 그 가능성을 함께 살펴보았다.

표 15 스마트공원 모니터링 계획 세부 항목

구분	상세계획 내용
모니터링 대상	• 스마트공원 대상지 온도, 습도, 공기질 등 환경 데이터
측정 항목	• 실외 온도, 습도, 일산화탄소, 미세먼지
측정 방법	• 측정도구, 측정 기간, 측정일시, 측정 간격(주기) 등에 따름

### ○ 환경 데이터 모니터링 장비 개발 요구사항

- ▶ 전원 : 측정지역의 전원이 공급 가능한 지역 선정, 전원 소비량 문제
- ▶ 통신 : 측정지역의 인터넷 환경 등이 열악하여, LTE 무선 인터넷 지원 필요
- ▶ 데이터 형식 : 빅데이터로 실시간 저장된 데이터는 인공지능에서 활용할 수 있는 CSV 포맷
- ▶ 측정 주기 : 측정 간격을 10초, 20초 ~ 수분 등으로 프로그램으로 변경할 수 있도록 하여 측정
- ▶ 데이터 저장방식 : 데이터의 저장 공간은 MicroSD 또는, 클라우드 서버에 저장
- ▶ 하드웨어 : 마이컴 IoT 하드웨어, DHT-22 온습도 센서
- ▶ 측정 범위 : 온도, 습도, 일산화탄소, 미세먼지 등
- ▶ 온도측정 범위 : -40~ 80C
- ▶ 메모리 저장 : offline data logger 및 외장 16Gbyte Micro SD 적용

- ▶ 통신 옵션 : 무선 인터넷 WiFi, LTE 방식(선택)
- ▶ 전원 옵션 : 태양광 모듈을 통한 무전원 방식으로 야외에 설치 용이
- ▶ 데이터 수집 주기 : 매 10초 이상 설정 주기로 측정
- ▶ 데이터 저장 : 클라우드 서버에 실시간 측정데이터 저장
- ▶ 기타 : 방수가 가능하여 외부 설치에 용이함

### 1.1.1 IoT 하드웨어: 임베디드 시스템 또는 마이컴

임베디드 시스템(embedded system)은 개인용 컴퓨터와는 달리, 미리 정해진 특정 기능을 수행하기 위해 컴퓨터의 하드웨어와 소프트웨어가 조합된 전자 제어 시스템을 말하며, 필요에 따라서는 일부 기계장치가 포함될 수 있다<sup>96)</sup>. 임베디드 시스템의 장점은, 무엇보다 여러 가지 프로그램을 동시에 활용할 수 있으며 멀티태스킹에 있으며, 이를 위해서는 메인 프로세서에 탑재되어 돌아가는 실시간 운영체제(Real Time Operating System)가 주로 사용된다<sup>97)</sup>.

대부분의 가전제품은, 임베디드시스템을 기본으로 장착하며 복잡한 영상처리와 AI 처리에도 활용되고 있다. 이로써 같이 사용되는 메모리의 크기와 소비되는 전력 또한 증가하는 단점이 있다<sup>98)</sup>.

한편 마이크로프로세서는 대표적으로 잘 알려진 범용 키트인 아두이노에 사용되는 Atmel의 ATmega328P 프로세서가 있다<sup>99)</sup>. 마이크로프로세서의 장점은, 주로 센서를 연결하여 단순한 센싱으로 데이터를 읽어내고, 이를 후단(Backend)에 보내는 데이터 프로세싱에 매우 용이하여 저전력으로, 1년 이상 장시간 사용되어도 소프트웨어가 단순하여 시스템이 정지되는 사례가 적고, 가격이 저렴

96) 홍철호 외 (2005), 「임베디드 형태의 VOD 클라이언트 시스템의 개발」, 『한국산학기술학회 논문지』, 6(4), 315-324.

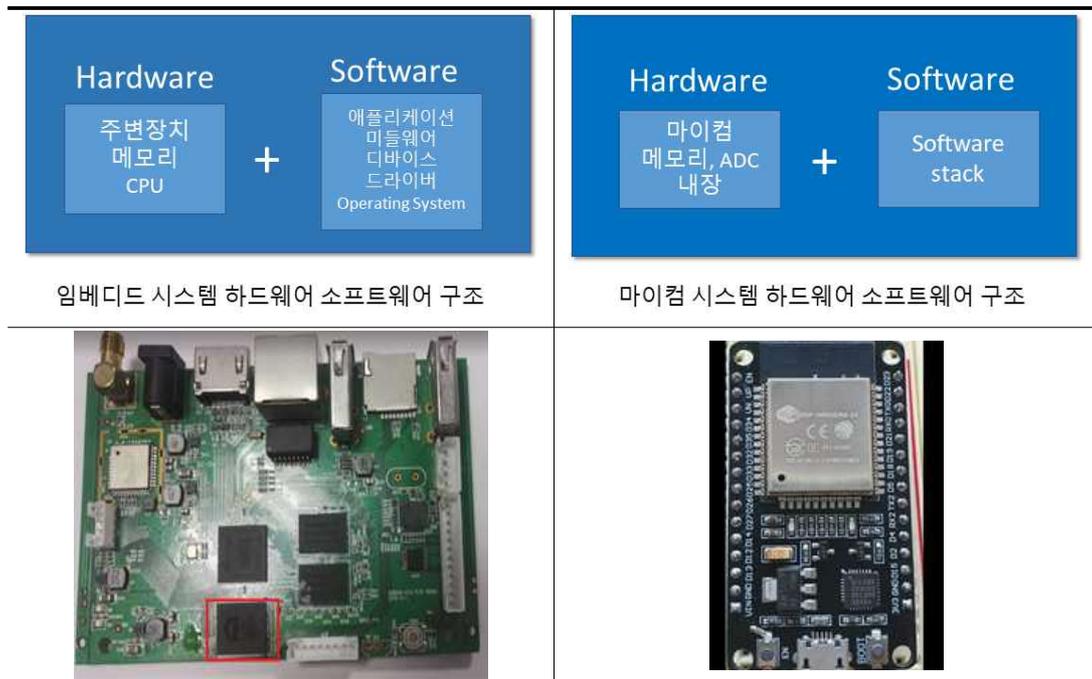
97) Lee, J. C., et al.(2004), "Setup of the development tools for a small-sized controller built in a robot using Linux," Korea Atomic Energy Research Institute.

98) 김문희 외(2004), 「임베디드 시스템과 소프트웨어공학」, 『정보과학회지』, 22(6), 5-13.

99) 박세연 외(2015), 「사물인터넷 (IoT) 환경에서 스마트폰 알람을 이용한 아두이노 (Arduino) 기반 가전기기 대기전력을 최소화한 ON 시스템 구현」, 『한국전자통신학회 논문지』, 10(10), 1175-1182.

한 부분이 장점이다<sup>100)</sup>. 단점은 여러 가지 태스크(Task)를 수행하기에 어렵고, 고도의 프로그램을 처리하기에는 CPU의 성능이 부족<sup>101)</sup>하여 AI나 영상처리와 같이 무거운 작업에는 사용할 수가 없다. 따라서 본 IoT 모니터링 장비 개발에서는, 경제적이면서도 적은 전원으로도 장기간 사용에 용이한 마이컴을 하드웨어로 사용하게 되었다.

표 16 임베디드 시스템과 마이컴 비교 IP STB(좌), ESP-WROOM32(우)



개발된 하드웨어는 표 16 블록도와 같이 전형적인 폰 노이만 방식의 하드웨어이다<sup>102)</sup>. CPU, 메모리, I/O장치 등으로 구성된다. CPU는 다양한 성능의 임베디드 프로세서가 현존하지만 가장 경제적이면서 AD, DA 컨버터를 내장한 범용 8bit 마이컴을 사용하였다. 14개의 디지털 IO를 갖추고 있고, 6개의 아날로그 입력,

100) Lee, H.(2001), 「차세대 Embedded 마이크로프로세서 기술 동향」, 『The Magazine of the IEIE』, 28(7), 49-55.  
 101) 반효경(2008), 「운영 체제와 정보기술의 원리」, 『이화여대 출판사』.  
 102) 김정환 외(1998), 「폰 노이만 구조를 채택한 데이터플로우 다중스레드 컴퓨터」, 『정보과학회논문지』 (A), 25(9), 917-925.

16MHz 주파수로 동작한다. I2C, PWM 등을 지원하여, 추후 다른 기능 등을 추가로 확장할 수 있다. 입력전압 레인지는 7V~ 12V이며 오퍼레이션 전압은 5V TTL 전압이다. max power는 6~20V이고, IO pin 전류는 40mA, 3.3V max 전류는 50mA이고, 내장 플래시 메모리는 32KB yet이며, SRAM 2k byte, EEPROM 1k byte로 본 애플리케이션에 최적으로 사용될 수 있다.

### 1.1.2 영상 처리 장치 또는 AD 컨버터

최근 도시공원에서 영상을 인식하기 위한 장비로 가로등이나 전신주에 CCTV를 설치하여 많이 활용되고 있다. 이런 배경에는 공원 이용객의 안전과 편의를 보호하고자 하는 목적이 있기에 대부분 사용자는 사생활 침해의 이슈를 간과하고 수용하고 있다<sup>103)</sup>. 하지만, 좀 더 사적인 공간에서의 영상 정보 획득 장치인 카메라는 매우 불편한 존재로 여겨진다. 여기에, 최근 AI 스피커 같은 음성 인식 서비스 또한, 간혹 해킹을 통해서 타인의 대화를 엿듣는 것으로 인식되어 불편함을 초래할 수 있다.

본 장비에는 리모트 모니터링 시스템으로서, 태양광을 사용하고 원격으로 PTZ(Pan, Tilt, Zoom) 기능이 되는 IP camera를 적용하는 것이 선택사항으로 가능하다. 이것은 기존 온습도와 같은 기상정보와 공기질 정보를 획득하면서, 여러 가지 발생할 수 있는 현장의 상황을 점검하고 태양광으로 충분히 전원이 공급되지 않는 상황 등을 모니터링하기에 매우 용이한 기능이다.

하지만 일반적으로는, 사생활 침해의 소지가 있거나 카메라가 필요하지 않은 경우에는 마이컴에 탑재된 회로를 기반으로 한 아날로그신호를 디지털로 전환해주는 AD 컨버터만을 활용하여, 기상 데이터 측정 장비만을 사용하여 데이터를 계속해서 수집할 수 있도록 하였다.

---

103) 한재경(2018), 「범죄예방용 CCTV의 설치 및 운영에 관한 소고」, 『공공사회연구』, 8(4), 109-137.



그림 11 HD grade, PTZ IP camera  
(Solar power)

### 1.1.3 MicroSD 또는 클라우드 서버

환경 데이터를 저장하기 위한 데이터 저장 장치 중 하나인 MicroSD 메모리는, 소형사이즈에 다양한 용량의 외장 메모리 장치로서, 외부 어댑터를 사용하면, 노트북 등의 USB 포트에 쉽게 사용할 수가 있다. 따라서 IoT 장비에서도 네트워크 지원 안 되는 경우에 데이터의 백업이나, 차량용 블랙박스의 경우와 같이 대용량 영상 데이터를 계속해서 저장하는 용도로 많이 활용된다<sup>104)</sup>. 개발 초기에는 무선 인터넷 통신이 지원되지 않았기 때문에, Micro SD 메모리에 획득된 데이터를 날자, 시간 등의 타임 스탬프 정보와 함께 메모리에 저장해두었다. 하지만 추후 데이터 획득 중에 발생한 여러 가지 사례로 보았을 때, 데이터 저장 중 전원이 차단이 될 때 데이터 분실(Data missing)이나 데이터 손실(Data loss)이 쉽게 발생하는 부분이 무척 큰 단점이었다. 특히 전원이 차단된 부분은, 반드시 현장에 방문해서 동작 여부를 모니터링할 경우에만 알 수 있어서, 매우 불편한 부분이다. 따라서 IoT 기기를 와이파이 같은 통신서비스에 접속하여, 클라우드 서버에 데이

104) 이금선, 박만곤(2010), 「유비쿼터스 러닝에서 마이크로SD 카드를 활용한 간편 교육 매체 전달 시스템의 설계」, 『한국멀티미디어학회 학술발표논문집』, 205-207.

터를 실시간에 저장할 수 있도록 하여, 어렵게 취득한 데이터를 안전하게 보관할 수 있으며, 동작 여부도 클라우드 서버에 접속하여 바로 확인할 수 있다<sup>105)</sup>. 클라우드 서버의 종류에는 자체 서버 구축, 아마존 웹서버(Amazon), 매틀랩(Matlab)사의 Thing Speak 등이 있다. 자체 서버의 경우에는 대용량의 데이터를 모으거나 추가 비용이 없는 장점이 있다. 기타 서비스는 일정한 사용료를 내고, 서비스에서 제공하는 방식을 사용하여 데이터를 수집하고 읽어낼 수 있다.



그림 12 마이크로SD 메모리  
(제공 : 주식회사 홈엔카, 마이크로마이크로칩스)

실제, 초기에 서울숲에 장비를 설치한 상황에서, 네트워크가 없는 상황으로 MicroSD로 데이터를 저장하였고, 그림 13 와 같이 2020년 10월25일부터 2021년 5월 20일까지 지속적인 방문을 통해서 데이터를 획득했다. 하지만 수집된 데이터를 확인한 결과 전원의 부재, 기타 이상으로 데이터 손실이 많이 발생한 것을 확인할 수 있었다. 사실, 이처럼 IoT 장비의 가장 큰 장점은 이처럼 현장을 방문하지 않고서도, 데이터를 24시간 아무 때나, 클라우드 서버에서 바로 읽거나 사용할 수 있는 부분에 있다.

105) 박춘식(2011), 「클라우드 컴퓨팅에서의 보안 고려사항에 관한 연구」 『한국 산학기술학회 논문지』, 12(3), 1408-1416.

2020년

- 10월 25일
- 설치

2021년

- 11월 1일
- 11월 21일
- 1월 22일
- 1월 25일
- 2월 16일
- 3월 39일
- 5월 5일
- 5월 20일
- 1차 데이터 수집
- 2차 데이터 수집
- 3차 데이터 수집
- 4차 데이터 수집
- 5차 데이터 수집
- 6차 데이터 수집
- 7차 데이터 수집
- 8차 데이터 수집



그림 13 2020년 10월 25일 ~ 2021년 5월20일 기간동안 수행된 MicroSD를 이용한 환경 데이터 수집

#### 1.1.4 Web based 또는 App based

IoT 기기에서 클라우드 서버에 업로드되어, 저장된 데이터를 실시간으로 확인하거나, 빅데이터에 액세스할 수 있도록 하는 부분은, 프로그램에서 프론트엔드(Front End)라 한다<sup>106)</sup>. 반대로, 서버 쪽에서 애플리케이션 프로그램이나 웹페이지에서 원하는 시간, 간격의 데이터 등을 처리하고 자 할 때, 데이터를 서버에서 가공 처리하거나 보관하는 일을 담당하는 것을 백엔드 (Back End)라 한다.

현재 제작된 하드웨어에서 사용하는 클라우드 서비스는 ‘매틀랩’ (Matlab)사에서 제공하는 ‘Thing Speak’ 이라는 IoT 클라우드 서비스로서, 여기에 저장된 데이터를 가공하여 이용자가 원하는 형태로 보여주는 서비스에 접근하는 방식은 Web 방식의 기반<sup>107)</sup>을 사용하게 된다.

이 프로그램을 사용하여, 현재 수집된 데이터 하루치, 월 단위 또는 년 단위의 변동 주기를 쉽게 파악할 수 있다<sup>108)</sup>. 여기에 thing show라는 안드로이드 앱 형태의 프로그램을 사용하면, 컴퓨터를 사용하지 않고도, 즉시 현장에서 데이터의 주기나 변동을 바로 확인할 수 있어서, 현장에서 작업할 때 매우 편리하다.

---

106) 차시호(2022), 「사물 웹(WoT) 환경에서 네트워크 모니터링 애플리케이션 개발을 위한 웹 프론트엔드 프레임워크의 적용 현황 및 트렌드」, 『디지털산업정보학회 논문지』, 18(1), 47-54.

107) Maureira, M. A. G., et al.(2011). “ThingSpeak - an API and Web Service for the Internet of Things,” World Wide Web..

108) Pasha, S.(2016), “ThingSpeak based sensing and monitoring system for IoT with Matlab Analysis,” *International Journal of New Technology and Research*, 2(6), 263-492.



그림 14 Thingspeak 홈페이지 (출처 : thingspeak.com)



그림 15 Web, App 방식 frontend 비교

### 1.1.5 전원 처리

실외에서 사용하는 IoT 장비를 구동하기 위한, 안정적인 전원과 무선 네트워크는 안정적인 데이터 수집에 필요한 매우 중요한 요소이다. 초기에 서울숲에 설치할 때는 반드시, 전원 콘센트가 제공되는 곳에만 IoT 장비를 설치하여 환경 데이터를 습득하였다. 하지만 이후에 다양한 장소에 IoT 장비를 사용하기 위해서, 여러 가지 종류의 전원을 고안하게 되었다.

첫째로, 태양광 패널을 사용하여, 장비에 전원 공급<sup>109)</sup>을 할 수 있도록 하였다. 하지만, 첫 번째 태양광 패널에서 직렬로 장비에 연결하여 출력되는 power 용량이 10Watt로 부족하여, IoT 디바이스와 LTE 통신모뎀을 구동하는 데 매우 역부족이었다. 그래서 좀 더 용량이 큰 20Watt 태양광 패널을 장착해보았으나, 실패하고, 충전지가 달린 5V 2A 출력의 8,000mAh 충전 배터리를 연결하여 해결되었다. 하지만, 우기가 계속되는 상황에서는, 태양광의 한계상 충전에 어려움을 겪을 수 있다.

두 번째 방법은, 외부 PTZ 카메라에 장착되어있는 태양광 모듈에서 전원을 분기하여 사용하는 방법 또한, 일반적인 맑은 날에서는 동작이 전혀 문제가 없었지만, 우기가 계속될 때 전원이 충전되지 못한 상황에서 지속해서 IoT 장비에서 사용하므로, 전원이 차단되지 못하고 방전되는 현상이 빈번하게 발생하였다.

따라서 우기와 같은 날씨와 상관없이 측정을 계속하기 위해서는, 대용량의 배터리를 시스템에 연결한 후 일정 시간 후에 교체하여 주는 방법으로, 전원을 계속 유지하도록 하는 것이 가장 좋은 방법이 될 것이다.

---

109) 안인수(2012), 「태양에너지를 이용한 일체형의 전원공급시스템과 DC 램프 모듈 개발」, 『한국컴퓨터정보학회논문지』, 17(2), 23-29.

표 17 연구에 사용된 태양광 패널들

	
<p>리오 링크 태양광 패널 (10Watt, 5V)</p>	<p>접이식 휴대용 태양광 패널 (20Watt, 5V )</p>
	
<p>태양광 충전 보조 배터리 8,000mAh, 5V 2A</p>	<p>휴대용 보조 배터리 26,800mAh, 5V 2A</p>

표 18 연구에 사용된 태양광 전원 사용 방법

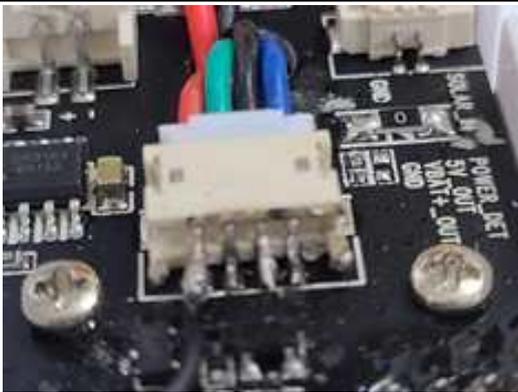
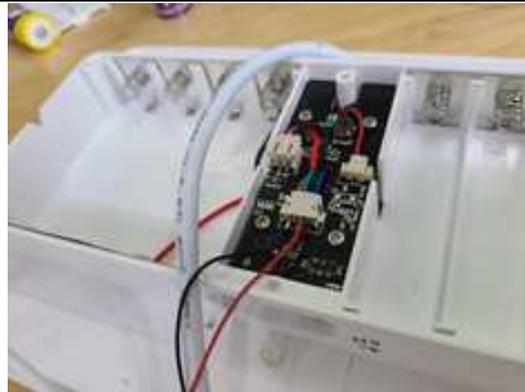
	
<p>PTZ 카메라 태양광 전원 모듈에서 전원 분기</p>	<p>PTZ 태양광 전원 모듈의 후면 개방 사진</p>



그림 16 태양광으로 충전 테스트 중인 장비

### 1.1.6 통신 방법

태양광 IP 카메라와 IoT 장비의 경우에는 유선 인터넷이 아닌, 무선 WiFi(와이파이)를 사용하게 되어있다. 그런데 아쉽게도 공원에서 현재 제공 중인 공용 WiFi는, 스마트폰으로 인증 절차를 밟아야만 비로소 암호화된 네트워크를 사용할 수 있게 되어있다. 안드로이드 핸드폰 같은 장치에서 WiFi접속을 한 후 별도의 웹서비스를 거쳐야 한다. 따라서 웹페이지 접속 등의 절차가 제공되지 않는 IoT 장비는 단순히 아이디와 패스워드만 가지고서 인터넷에 접속할 수 있는 인터넷 공유기나 LTE 접속용 모뎀이 필요하다.

첫 번째 방법은, 표 19 에서 보는 바와 같이 LTE 모뎀은 통신사에서 제공하는 유심 카드(U sim)를 꼽아서 데이터 공유 기능 등으로, 와이파이 신호를 IoT 장비에 제공한다. 두 번째는, 일반적으로 가장 많이들 사용하는 EGG 장비로, LTE 신호를 와이파이 신호로 바꾸어 주는 역할을 한다.

표 19 연구에 사용된 와이파이 와 모뎀 서비스 종류

	
<p>LTE 모뎀</p>	<p>LTE to 와이파이 예그</p>
	
<p>와이파이 접속용 WSP32 마이칩</p>	<p>공공 와이파이 안내 표지판</p>

### 1.1.7 센서의 선정

#### 1.1.7.1 센싱 타겟

##### (1) 온습도 센서

DHT22 센서는 온도를 감지하는 부분은 반도체의 세라믹으로 이루어져 있다<sup>110)</sup>. 온도에 따라서 물질의 저항값이 변하는 소재의 특성을 이용하였으며 값의

110) Mihai, B. O. G. D. A. N.(2016). "How to use the DHT22 sensor for measuring temperature and humidity with the arduino board," *Acta*

변화를 감지해 온도를 출력하고 있다.

습도는 두 전극 사이의 저항 변화를 측정함으로써 공기 중의 습도 변화를 알아낼 수 있다<sup>111)</sup>. 실생활에서는 기상예보나 에어컨에서 활용되고 있다.

연구에서 적용 시에, 간혹 제품들이 품질상 성능에 문제가 되어, 데이터 미싱이 발생하는 경우가 빈번하게 발생하였다.

그리고, 외부 생김새는 비슷하지만, DHT11, DHT22 센서는 측정할 수 있는 온도 대역이 다르므로, 이점을 주의해야 한다.

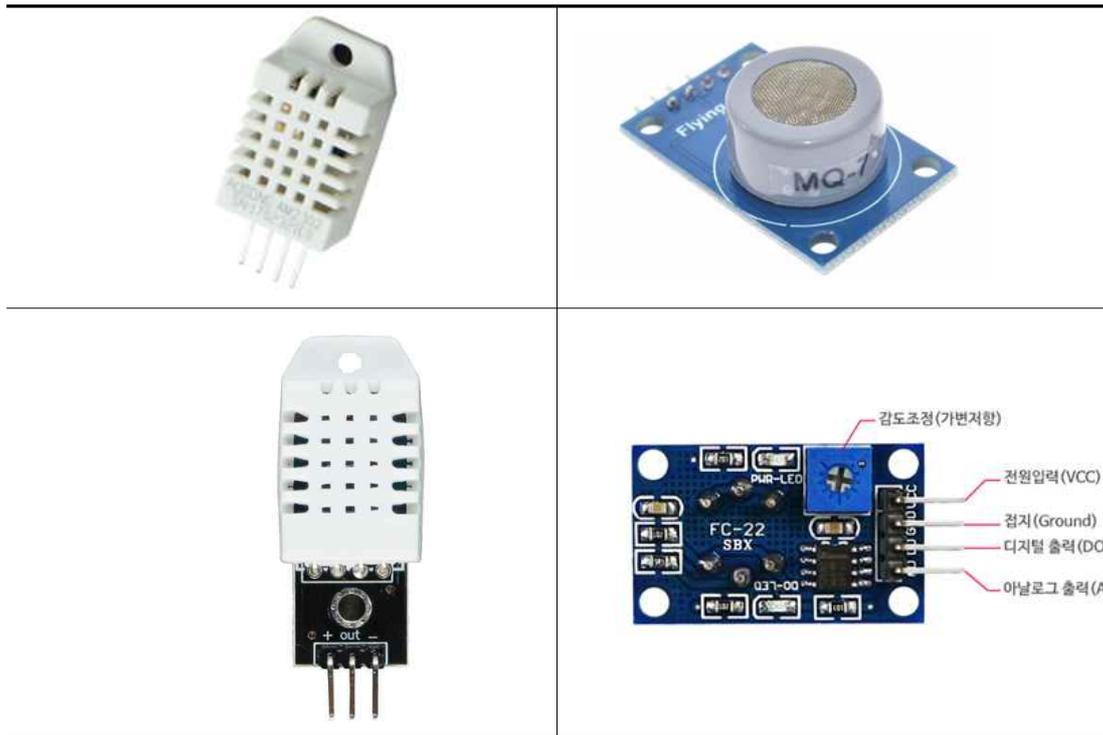
표 20 온습도 센서 DHT 11 과 DHT 22비교

Sensor Model name	DHT-11	DHT-22
온도 측정범위	0~50	-40~ 80
온도 측정 오차	2	0.5
습도 측정 범위	20~80%	0~100%
습도 측정 오차	5%	2%
측정간격	1초	2초

*Uiversitatis Cibiniensis - Technical Series*, 68, 22-25.

111) Liu, T.(2013), “Digital-output relative humidity & temperature sensor/module DHT22 (DHT22 also named as AM2302),” Aosong Electronics.

표 21 온습도 센서 DHT22 (왼쪽) , 일산화탄소 센서 MQ7 (오른쪽)



## (2) 일산화탄소 센서

일산화탄소는 탄소와 산소로 구성되며, 무색, 무취, 무미로 사람이 인지하기 어려운 유독성 기체다<sup>112)</sup>. 일산화탄소는 연료 속의 탄소 성분이 불완전 연소하였을 때 발생하며 대부분 대기 중으로 배출된다.

한국환경공단 에어코리아에서 기준은 다음과 같이 4단계로 나뉘어 있다. 좋음 0 ~ 2ppm, 보통은 2.01~ 9ppm, 나쁨은 9.01~15ppm, 매우 나쁨은 15.01ppm 이상을 말한다. 사용되는 센서는 MQ 7으로 특성은 다음 표 22 와 같다.

112) Sai, K. B. K. at al.(2019), "IOT based air quality monitoring system using MQ135 and MQ7 with machine learning analysis," *Scalable Computing: Practice and Experience*, 20(4), 599-606.

표 22 MQ7 ; 센서 특징

동작전압	5V
소비전류	150mA
센서 D0 출력	0 또는 1
센서 A0 출력	0 ~ 1023
크기	3.2 * 2.0 * 2.1cm
핀구성	4핀(Vcc, GND, D0, A0)
특성	천연가스 (LPG, 부탄, 에탄, 프로판등 ) 등을 검출하는 가스 센서이다 부착된 가변저항을 이용해 감도를 조정한다. 매우 높은 민감도와 빠른 반응 시 가능 가진다.

### (3) 미세먼지 센서

미세먼지(Particulate Matter)는 세계보건기구(WHO)가 지정한 1급 발암물질이다. 미세먼지 PM10은 10 $\mu$ m의 크기로 대부분 인후 또는 기관지 점막으로 흡입된 후, 가래와 함께 배출되거나 식도를 통해 위 속으로 넘어간 후 체외로 배출된다. 하지만 중금속과 화학물질이 매우 많아 흡입하면 암세포를 유발한다<sup>113)</sup>.

Shinyei 사(社)의 PPD42NS 센서는 광학 감지 방식을 사용하여 미세먼지 농도를 측정한다<sup>114)</sup>.

특징은 다음과 같다.

- ▶ 1 $\mu$ m 및 2.5 $\mu$ m 먼지 검출
- ▶ PM10, PM2.5 모드중 선택 가능
- ▶ 저항 히터로 공기를 자동 흡입하여 측정
- ▶ PWM 출력 기능 제공
- ▶ 입자 탐지 범위 :최대 8000pcs/283ml(1 $\mu$ m입자이상)

동작원리는, 적외선 센서로 공간을 비추고 반대편에 있는 포토다이오드 감지기가 포커싱렌즈를 통해 먼지를 측정하는 방식이다. 이때 적외선 LED가 켜고 끄는 것이 아니라 계속 켜있는 상태로 측정한다<sup>115)</sup>. 센서는 먼지가 얼마나 있는지를

113) 로니킴(2021), 『모두의 아두이노 환경 센서』, 길벗.

114) Austin, E., et al.(2015). "Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor," *PLoS one*, 10(9), e0137789.

115) 김형래 외(2022), 「광산란 방식 간이형 미세먼지 센서의 입경 분리능 평가기

측정하는 것이 아니라, 일정 시간 동안 먼지가 얼마나 발생했는지를 측정한다.

표 23 미세먼지 예보등급

PM ( $\mu\text{g}/\text{m}^3$ , per day)	Good	Normal	Bad	Very bad
PM10	0~30	31~80	81~150	151 ~
PM2.5	0~15	16~35	36~75	76 ~



SIZE[mm]	58.5*45*17
WEIGHT[gram]	20
SUPPLY VOLTAGE[volt]	5
SIGNAL OUTPUT	P1(small particle) P2(smaller particle) PWM, LPO(%)
DUST CONCENTRATION[pcs/ft <sup>3</sup> ]	50~800,000
UNIFORMITY[%]	± 15
CHARACTERISTIC	Heart beat , MICOM CAL. High Sensitivity IR LED

그림 17 미세먼지 센서와 사양  
(출처 : Shinyei 사)

### 1.1.8 구성된 하드웨어 전체 블록다이어그램

그림 18 은 전체 하드웨어의 블록다이어그램이고, 개발된 소프트웨어는, 리얼타임 OS 등은 적용하지 않았고, 전력 소모량이 적으면서, 경제적으로 마이컴 내장형 메모리에 바로 내장하여 사용할 수 있도록, 마이컴 개발환경에서 가장 범용이고, 라이브러리 사용에 용이한 C언어로 작성되었다. ADC 컨버팅 후 저장되는 파일타입은 CSV 파일 포맷으로 하여, 추후 인공지능 프로그래밍 때 판다스(pandas) 라이브러리 등을 사용하여 쉽게 읽어들이, 머신러닝에 활용할 수 있다.

코딩방식은 폴링 방식으로, 10초 간격으로, 4개의 센서값을 읽어들이어서, 외부 메모리에 RTC(Real Time Clock) 값과 함께 저장되는 형식이다. 와이파이 모듈과의 이식성을 염두에 두어서, 차후에는 클라우드 서버에, 빅데이터를 저장할

법 개발」, 『한국대기환경학회지』, 38(1), 159-168.

수 있도록 고려하여, 프로그래밍하였다.

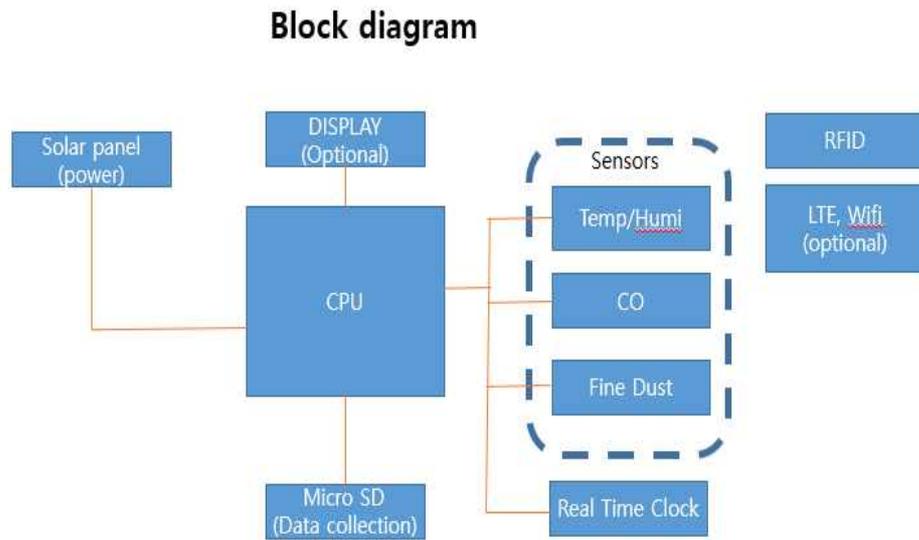


그림 18 IoT 스마트공원 공기질 센서의 시스템 블록도

Project plans  
Introduction to Device



그림 19 제작된 IoT 디바이스 제품 1



그림 20 제작된 IoT 디바이스 제품2

### 1.1.9 외부 케이스와 스탠드

#### 1.1.9.1 우천, 설치 높이, 바람의 영향

실내외부에서 기상 데이터를 습득하고, 클라우드 서버에 보내는 역할을 하는데 있어서 측정장소와 측정방법, 측정하는 높이는 매우 중요하다. 서울숲에 실제 설치된 IoT 장비들은 이러한 부분을 반영하여, 함체 안에 제작된 키트와 함께 LTE 모듈이 같이 적용되었다. 외부환경에 지속해서 노출됨으로써, 태양, 바람 등 냉기 등에 강해야 하며, 특히 비나 눈 같은 습기에 강하면서도 전원 공급에 문제가 없는 강건한 구조이어야 한다.

표 24 서울숲에 실제 설치된 IoT 장비 사례



### 1.1.9.2 실내 사용 또는 실외 사용

표 25 설치 사례에 따른 IoT 제작 형태

		
		
<p>실내 천장부, 중간부 측정 / 고정형</p>	<p>실내 및 실외 거치 형태</p>	<p>스탠드 타입/ 높이 조정 가능</p>

### 1.1.10 최종 디자인

본 연구를 위하여, IoT 환경 데이터 센서 장치를 자체 개발 및 적용하여 자료를 수집하였다. 표 26에서처럼 내, 외부 형태로 개발되었고, 장비를 서울숲 가족마당, 수변부와 방문자센터에 각 설치하였다. 내부에는 센싱하는 정보인, 온도, 습도, 일산화탄소, 미세먼지 등의 센서와 통신을 위한 LTE 모듈을 내장하고 있다.

개발된 소프트웨어는, 리얼타임 OS 등은 적용하지 않았고, 전력 소모량이 적으면서, 경제적으로 마이컴 내장형 메모리에 바로 내장하여 사용할 수 있도록, 마이컴 개발환경에서 가장 범용이고, 라이브러리 사용에 용이한, C언어로 작성되었다. 체크 후 저장되는 파일타입은 CSV 파일 포맷으로 하여, 추후 AI 프로그래밍 때 판다스(pandas) 등으로 쉽게 읽어 들여, 머신러닝에 활용할 수 있다.

앞서서 설명한 사항대로, 코딩방식은 폴링 방식으로, 10초 간격으로, 4개의 센서값을 읽어들이어서, 클라우드 서버에 저장되는 형식이다.

데이터는 추가적으로 전처리되어, 1시간, 10분 또는 1분 단위로, 단위로 평균값을 내어, 별도의 정형화된 데이터로 정리되었다<sup>116)</sup>. 제작한 스마트공원 측정용 IoT 장비와 데이터 로거 장비의 장단점을 비교한 내용은 표 27과 같다<sup>117)</sup>.

표 26 서울숲에 실제 설치 적용된 IoT 센서 장비



116) 권태경 외(2022), 「IoT 모니터링 시스템을 통한 기후변화 취약계층 지원 차열 사업의 효과 분석-김해시 진영읍 주거지역을 사례로」, 『한국기후변화학회지』, 13(3), 383-398.

117) 권태경, 조현민, 임영신(2022), 「도시 그늘막 효과 측정을 위한 시뮬레이션 장비 개발에 관한 연구」, 『한국환경기술학회지』, 23(3), 169-176.

표 27 스마트공원 측정용 IoT 장비와 데이터 로거 장비의 장단점 비교

항목	스마트공원 측정용 IoT 장비 (서울대 자체 개발)	데이터로거(TR-71wf)
WiFi 실시간 데이터 전송	가능	불가능 (옵션)
실시간 전송 에러 확인	가능	불가능
전원방식	전원 5V 3A 태양광 (optional)	3V/ AA x2 battery
태양광 전원사용	가능 (날씨 영향 있음)	불가능
time interval	1sec~ 1 hour/ variable	1min, 5min, 10min fixed
클라우드 서버	가능	옵션
CSV 포맷	가능	가능
복합센서기능	온도, 습도, 미세먼지, 일산화탄소 등 동시 측정 / 동시 시계열 데이터 분석 가능	온도 센서
측정 높이	0~ 200cm (자체 고정스탠드)	부착형
온도 동시 측정 채널 수	최대 5	최대 2
생활 방수기능	가능	불가능
Big Data storage sustainability	Cloud server Max 1년 이상 (3초 단위 측정 빅데이터)	Max 3개월(10분 단위) 기기파손시 복구 불가
모바일 전용 실시간 모니터링 App	가능 (모니터 주기, day, week, month)	불가능
인공지능 데이터 상관분석	가능	불가능

출처 : 권태경 외 2인(2022)<sup>118)</sup>

118) 권태경, 조현민, 임영신(2022), 「도시 그늘막 효과 측정을 위한 시뮬레이션 장비 개발에 관한 연구」, 『한국환경기술학회지』, 23(3), 169-176.

## 1.2 인공지능을 통한 환경 데이터 보행량 영향 요소 평가

인공지능(AI)은 컴퓨터가 인간의 지능을 모방할 수 있도록 하는 기술이다. 인공지능의 기법 중 딥러닝, 머신러닝, AI의 카테고리를 이해하려면 다음 정의를 고려한다.

딥러닝은 인공신경망을 기반으로 하는 머신러닝의 최하위 집합이다. 인공신경망의 구조는 다중 입력, 출력 및 은닉 계층으로 구성되어 있기 때문에 학습 과정이 깊다. 각 계층에는 입력 데이터를 다음 계층이 특정 예측 작업에 사용할 수 있는 정보로 변환하는 단위가 포함되어 있다. 이 구조 덕분에 기계는 자체 데이터 처리를 통해 학습할 수 있다.

머신러닝은 기계가 경험을 사용하여 작업을 개선할 수 있도록 하는 기술(예: 딥러닝)을 사용하는 인공지능의 하위 집합이다<sup>119)</sup>.

머신러닝 작업은 통상 다음과 같은 3가지의 부류로 구분된다.

- 지도 학습 (Supervised Learning) :시스템이 레이블이 지정된 훈련 데이터에서 기능을 추론하는 지도 학습
- 비지도 학습(Unsupervised Learning) :학습 시스템이 레이블이 지정되지 않은 데이터의 구조를 추론하려고 시도하는 비지도 학습
- 강화학습(Enforcement learning)의 세 가지 광범위한 범주로 분류된다.

그리고, 일반적으로 학습 프로세스는 다음 단계를 기반으로 한다.

- ▶ 알고리즘에 데이터를 입력한다. (이 단계에서는 예를 들어 특징 추출을 수행하여 모델에 추가 정보를 제공할 수 있다.)
- ▶ 이 데이터를 사용하여 모델을 훈련한다.
- ▶ 모델을 테스트하고 배포한다.
- ▶ 배포된 모델을 사용하여 자동화된 예측 작업을 수행한다. (즉, 배포된 모델을 호출하고 사용하여 모델에서 반환된 예측을 수신한다.)

본 연구에서는 이와 같은 머신러닝을 사용하여, 수집된 환경 데이터를 훈련하여

---

119) “Deep learning vs. machine learning in Azure Machine Learning”, <https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?source=recommendations> (검색일: 2022. 10. 15.)

서울숲공원의 보행량과 소래 습지 생태공원의 이산화탄소 산출량 등을 예측하고 이를 평가하고자 한다.

## Deep learning, machine learning, and AI

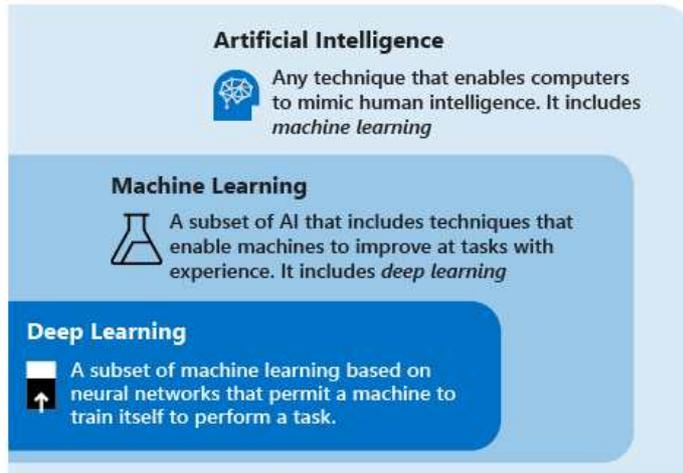


그림 21 인공지능, Machine learning and Deep learning 카테고리

출처 : Microsoft 홈페이지 120)

120) “Machine Learning Algorithm Cheat Sheet for Azure Machine Learning Designer”, [https://learn.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet?WT.mc\\_id=docs-article-lazzeri](https://learn.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet?WT.mc_id=docs-article-lazzeri) (검색일: 2022. 10. 15.)

### 1.2.1 Auto ML을 사용한 전체 연구 flow

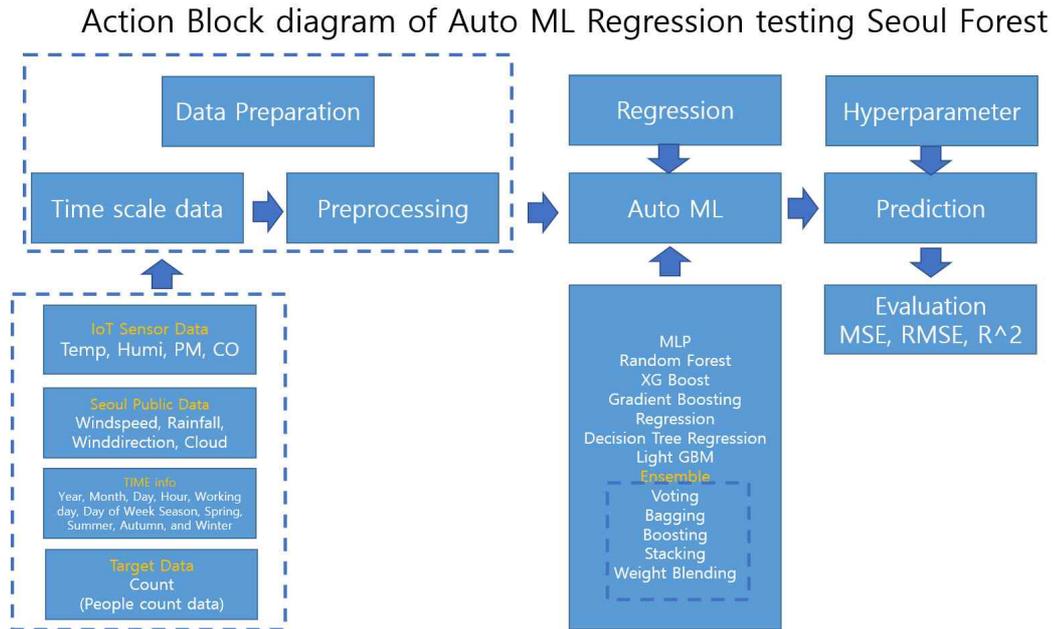


그림 22 서울숲 공기질과 방문자 상관관계분석 Machine Learning 전체 학습도

그림 22 에서 같이 머신러닝을 사용하여, 서울숲의 공기질과 보행량의 상관관계를 분석 계획을 준비하였다. 먼저 시계열 데이터로 수집된 데이터들을 데이터의 전처리 및 강화 후, 필터링된 기상 데이터와 계절, 날짜, 월 같은 범주 데이터는 서울숲의 보행량을 예측을 수행하기 위해 선형회귀 관련 다양한 알고리즘을 적용하여 최적의 성능을 갖고 있는 알고리즘을 비교 평가하게 된다. 여기에서 일반적인 데이터의 분석과 그래프 등은 구글 코랩과 파이썬을 사용하여 분석 및 평가하였다. 그리고 머신러닝 효과 검증을 위하여는, Stacking ensemble, Voting Ensemble, XG Boost Regressor, Random Forest, Gradient Boost Regression, Decision Tree, Light GBM, Multi Layer Perceptron 등의 머신러닝 알고리즘을 사용하여 Azure Auto ML에서 구현 평가되었다.

전체 워크플로를 보기 위해 서울숲 환경 데이터파일인 'selforest\_final.csv' 를 로드하고 필요한 열을 선택하고 target(목표변수)를 보행량의 칼럼인 "count" 로 선택했다. 학습 데이터와 테스트 데이터를 분리하기 위해 분할 데이

터의 비율을 70%, 30%로 설정하고 알고리즘에 따라 하이퍼 파라미터 변경과 분할된 테스트 데이터를 사용한 교차검증은, 보행량 예측에 사용할 수 있는 최상의 머신러닝 모델을 선택하는 데 사용된다<sup>121)</sup>.

Auto ML로 선정된 알고리즘 각각에 트레이닝 데이터를 사용하였고, 테스트 데이터는 위에서 나눈 분할 데이터를 기반으로 값을 넣어 학습시켰다. Auto ML은 시간 소모적이고 반복적인 머신러닝 모델 개발 작업을 자동화하는 최신의 인공지능 분석 및 개발 프로세스로 본 연구에서는 지속적으로 반복적으로 분석해야 하는 분석업무를 효율적으로 사용할 수 있도록 개발하였다<sup>122)</sup>. 현재 Auto ML의 제공사는 다음과 같이 많이 있다. IBM AutoAI, Microsoft AzureML, Google AutoML, Amazon SageMaker을 개발하였으며, 삼성SDS도 Brightics AI를 개발하였다<sup>123)</sup>.

본 연구에서는 Microsoft에서 제공하는 Azure ML을 사용한다<sup>124)</sup>. AutoML은 인공지능을 위한 인공지능<sup>125)</sup>이라고 불릴 정도로, 데이터 과학자, 분석가나 개발자는 모델 품질을 유지하면서 확장성, 효율성 및 생산성이 높은 ML 모델을 빌드할 수 있다. 기존의 머신러닝 모델 개발은 리소스를 많이 사용하므로 수십 개의 모델을 생성하고 비교하는 데 상당한 도메인 지식과 시간이 필요하였다<sup>126)</sup>. 그러나 Auto Machine Learning을 사용하면 프로덕션 준비 ML 모델을 매우 쉽고 효율적으로 얻는 데 걸리는 시간을 단축할 수 있다.

그림 23 는 이와 같은 Auto ML의 파이프라인의 일례이다.

먼저 데이터 세트가 공급되면, 이에 대한 전처리를 자동으로 수행하여, 손실 데이터(Missing data), 아웃라이어등을 모두 해결해주고, 하이퍼 파라미터를

---

121) 조성래 외(2015), 「의사결정트리와 인공 신경망 기법을 이용한 침입탐지 효율성 비교 연구」, 『디지털산업정보학회논문지』, 11(4), 33-45.

122) 성기우 외(2021), 「Auto ML을 활용한 필드 클레임 예측 및 고장 주요 인자 분석」, 『한국신뢰성학회 학술대회논문집』, 132-132.

123) “인공지능을 위한 인공지능, AutoML“, 삼성SDS, [https://www.samsungsds.com/kr/insights/ai\\_automl.html](https://www.samsungsds.com/kr/insights/ai_automl.html) (검색일: 2022. 10. 15.)

124) Yu-Jin NAM, Won-Ji SHIN(2019), 「A Study on Comparison of Lung Cancer Prediction Using Ensemble Machine Learning」, 『인공지능학회지』, 7(2), 19-24.

125) “인공지능을 위한 인공지능, AutoML“, 삼성SDS, [https://www.samsungsds.com/kr/insights/ai\\_automl.html](https://www.samsungsds.com/kr/insights/ai_automl.html) (검색일: 2022. 10. 15.)

126) “자동화된 Machine Learning(AutoML)이란”, <http://learn.microsoft.com/ko-kr/azure/machine-learning/concept-automated-ml> (검색일 2022. 10. 15.)

유틸리티를 높여 최적의 ML 알고리즘 모델링을 추출해준다.

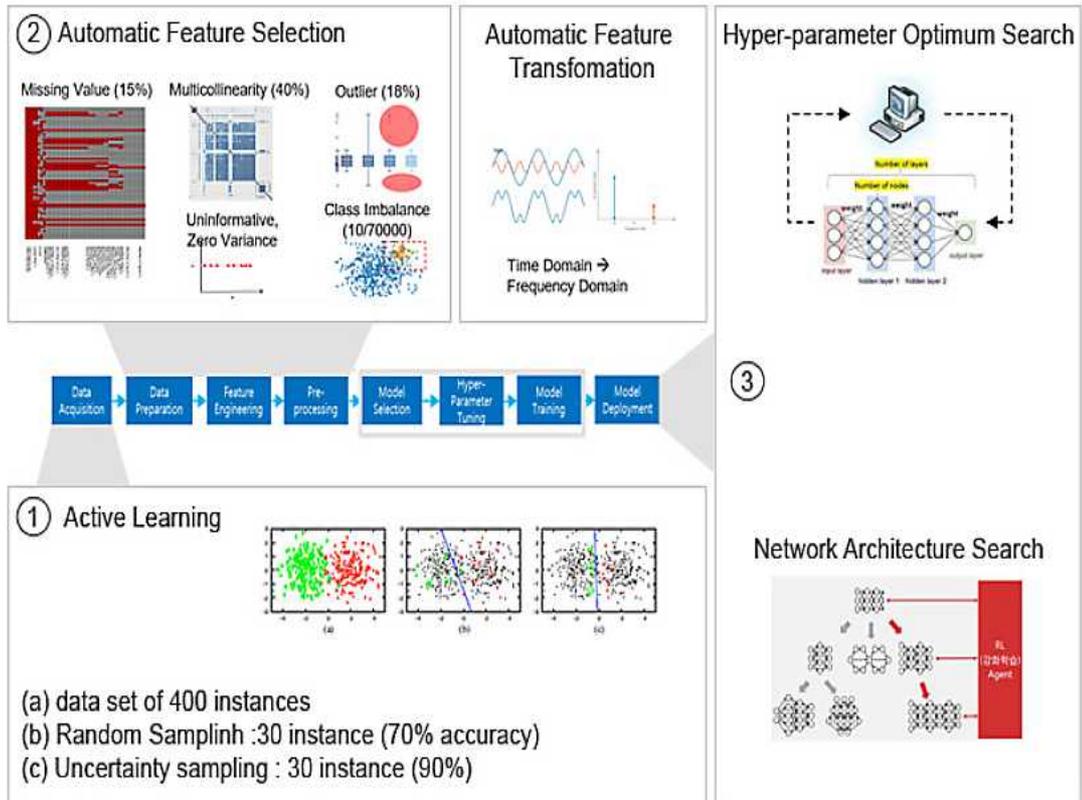


그림 23 Auto ML (머신러닝) 파이프라인  
출처:삼성SDS

Auto ML은 현재 수행평가에 약 50여 개의 알고리즘을 자동으로 적용하며, 각 알고리즘의 하이퍼 파라미터값 또한 자동으로 조정하여, 결과값에 대한 최적의 성능(best performance)을 갖는 순으로 나열하고, 평가 결과 또한 같이 도출하게 된다.

## 1.2.2 머신러닝 알고리즘의 종류

지도 학습의 대표적 학습 알고리즘은 Ridge Regression(릿지 회귀), (Lasso Regression(라쏘 회귀), Logistic Regression(로지스틱 회귀), (Support Vector Machine(서포트 벡터 머신), Decision Tree(의사결정나무), Random Forest(랜덤 포레스트), Artificial Neural Network(인공신경망) 등이 있다<sup>127)</sup>. Auto ML에 사용되는 머신러닝의 알고리즘 중 최적은 대부분 Ensemble(앙상블) 기반의 학습 방법으로 구성되어있다<sup>128)</sup>. 왜냐하면, 기존의 단일 알고리즘보다는 캐글이나 다양한 경진대회를 거치면서 알고리즘의 개선이 있어왔고, 여러 개의 머신 러닝 모델링 결과를 모아서 다시 한번 모델링을 하면서, 그 예측 성능을 매우 높이는 것으로 알려져 왔다. 그 결과 앙상블이 가장 최적의 예측 결과 스코어를 가지고 있는 것으로 알려져 왔기 때문이다. 이 절에서는 앙상블 모델링에 주로 사용되는 알고리즘을 정리한다.

### 1.2.2.1 Multi Layer Perceptron

MLP는 Multi Layer Perceptron(멀티레이어 퍼셉트론)의 약자로 인공신경망(ANN)의 유명한 클래스이다. 입력층과 출력층 사이에 하나 이상의 은닉층(Hidden layer)과 비선형 연산기를 갖는 인공신경망이다. 이때 은닉층의 개수가 2개 이상인 신경망을 Deep Neural Network, DNN(심층 신경망)이라고 한다. 각 계층의 노드를 뉴런이라고 하며 각 뉴런은 Sigmoid, Softmax, ReLU와 같은 활성화 함수를 가지고 있다<sup>129)</sup>. Gradient Decent방식을 통해서 Result-Target의 차의 제곱을 줄여나가는 Back propogation(백프로퍼게이션)을 수행하여 학습이 이루어진다. 단점은 Over fitting(과적합)과 학습 시간이 은닉층의 개수에 따라서 장시간 증가한다.

### 1.2.2.2 Random Forest

---

127) 이제승, 이현희(2019), 「기계학습 알고리즘을 이용한 보행 만족도 예측모형 개발」, 『국토계획』, 54(3), 106-118.

128) 우덕채 외(2019), 「트랜잭션 기반 머신러닝에서 특성 추출 자동화를 위한 딥러닝 응용」, 『한국 IT 서비스학회지』, 18(2), 143-159.

129) 장윤정, 정대원(2019), 「딥러닝 기반의 영상 분석 기술 동향」, 『항공우주산업 기술동향』, 17(1), 113-122.

Random Forest(랜덤 포레스트) 알고리즘은 분류 및 회귀 문제에 가장 많이 사용되는 Ensemble(앙상블) 학습모델이다. 또한 Random Forest는 여러 개의 Decision Tree(결정 트리)를 만들고, 각 트리에 통과시켜 Regression(회귀)의 경우에는 평균화를 수행하고, Segmentation(분류)의 경우에는 Voting을 통해서 좀 더 정확한 결과를 생성한다.

Random Forest 알고리즘의 장점은 분류 및 회귀 작업 모두에 사용할 수 있는 점이 있으며, 이상치에 민감하지 않아<sup>130)</sup>, 변수를 제거하지 않고 바로 실행하는 경우 정확도 높은 점이며, 단점은 모델의 해석이 어렵고 트리 개수가 많아질수록 학습 시간이 오래 걸리는 단점이 있다.

Random Forest로 예측을 수행하는 방식은 다음과 같다.

- ▶ 테스트 기능별로 무작위로 의사 결정 트리를 만들고 규칙을 사용하여 대상을 예측한다.
- ▶ 각 예측 대상에 대한 투표를 계산한다.
- ▶ 최종 예측은 투표율이 높은 예측 대상으로 계산된다.

### 1.2.2.3 XG Boost

XGBoost는 EXtreme Gradient Boosting의 약자로 XGBoost는 Gradient boosting 알고리즘을 기반으로 한다<sup>131)</sup>. Gradient Boosting이란 여러 개의 weak learner(약한 학습기)를 조합하여 사용하는 Ensemble Learning 중 하나인 Boosting 기법에 오류 식을 최소화하는 방향성을 가지고 반복적으로 가중치 값을 갱신하는 방법인 Gradient Decent(경사 하강법)을 접목한 것이다<sup>132)</sup>.

장점은 최적화(Optimization)와 알고리즘을 통해 병렬처리를 지원하여, 학습 속도가 빠르고, Classification And Regression Tree 기반으로 회귀와 분류 둘 다 우수한 기능을 발휘하지만, 트리 개수가 많아질수록 처리시간이 길어진다.

기술적인 측면에서 일반적인 Gradient Boosting Machine (GBM) 은 순차적

---

130) Harishkumar, et al.(2020). Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Computer Science, 171, 2057-2066.

131) 이용준 외(2020), 「XGBoost를 활용한 고속도로 콘크리트 포장 파손 예측」, 『한국건설관리학회 논문집』, 21(6), 46 - 55.

132) 심대천 외(2022), 「XGBoost Regression 기계학습을 이용한 제주도 지하수 염소이온 농도예측 연구 사례」, 『지질학회지』, 58(2), 243-255.

으로 가중치를 증가시켜 학습하지만, XGBoost는 병렬 수행을 통하여 비교적 빠른 속도를 나타내며 과적합 규제(regularization) 기능이 내장되어 과적합에 대한 내구성을 가진다<sup>133)</sup>.

#### 1.2.2.4 Gradient Boosting Regression

최근 가장 널리 사용되는 앙상블 알고리즘 중 하나는 Gradient boosting이며 이는 회귀 및 분류 문제에 사용할 수 있다. 또한 XG Boost는 Gradient boosting Decision Tree를 기반으로 더 빠른 속도와 성능을 제공하도록 설계되었다.

#### 1.2.2.5 Decision Tree Regression

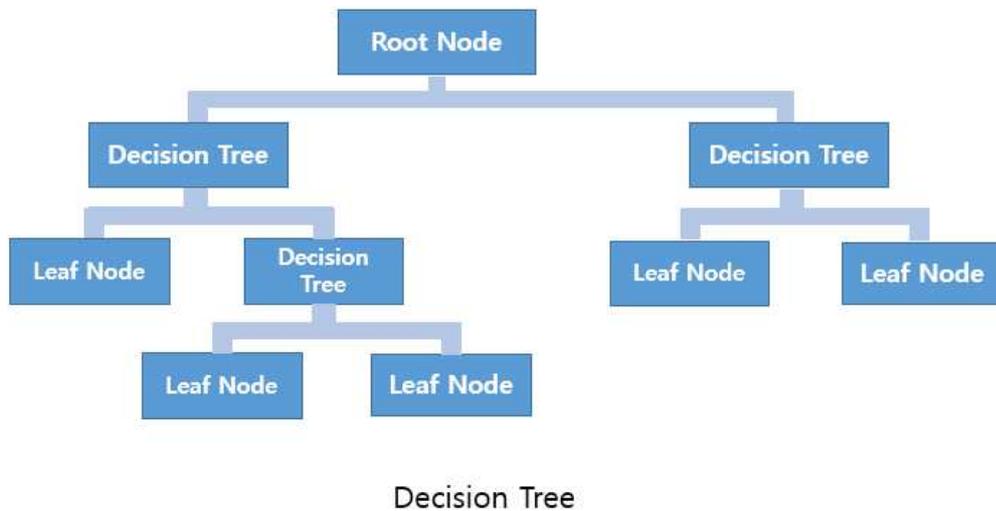


그림 24 Decision Tree

DT(Decision Tree) 알고리즘은 Supervised Learning(지도 학습) 범주에 속하는 잘 알려진 머신러닝 모델이다. 또한 분류 및 예측 작업 모두에 사용할

---

133) Chen, T., Guestrin, C.(2016), "Xgboost: A scalable tree boosting system," *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.

수 있다. DT의 기본 아이디어는 데이터에 있는 규칙 학습을 통해 자동으로 찾아내 Tree(트리)기반의 분류 규칙을 만드는 것이다. 일반적으로 규칙을 가장 쉽게 표현하는 방법은 if/else 기반으로 나타내는 것이다<sup>134)</sup>. 그리하여 데이터의 어떤 기준을 바탕으로 규칙을 만들어야 가장 효율적인 분류가 될 것인가가 알고리즘의 성능을 크게 좌우한다. 그림 24은 Decision Tree의 구조를 간략하게 나타낸 것이다. 규칙 노드로 표시된 노드는 규칙 조건이 되는 것이고, Leaf Node(리프노드)로 표시된 노드는 결정된 클래스 값이다. 그리고 새로운 규칙 조건마다 서브 트리가 생성된다<sup>135)</sup>.

### 1.2.2.6 Light GBM

Light GBM은 Gradient Boosting 프레임 워크로 Tree 기반 학습 알고리즘이다. XGBoost의 단점을 보완한 모델로, 손실이 큰 트리에 지속적인 분할을 통해 깊이는 늘어나지만 불필요한 연산을 줄일 수 있어 학습 시간이 단축되었고, 성능은 더 좋아지거나 비슷한 수준을 나타낸다<sup>136)</sup>. Light GBM은 Tree가 수직적으로 확장되는 반면에 다른 알고리즘은 Tree가 수평적으로 확장된다. 즉 Light GBM은 leaf-wise인 반면, Gradient Boosting 알고리즘은 level-wise이다.

데이터 사이즈는 날이 갈수록 커지고 있고 전통적인 데이터 분석 알고리즘으로 빠른 결과를 얻기란 더욱 어려워졌다. Light GBM의 장점은 가볍게 설계되어 속도가 빠르다. Light GBM은 큰 사이즈의 데이터를 다룰 수 있고 실행시킬 때 적은 메모리를 차지한다. 단점은 깊이가 깊어질수록, 오버피팅이 발생하므로 Max-depth튜닝을 통한 조정이 필요하다.

### 1.2.2.7 Ensemble

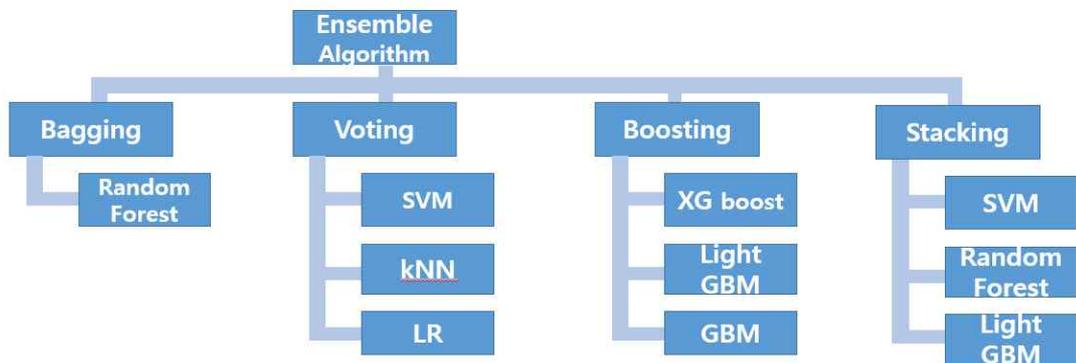
Ensemble(앙상블) 학습이란 여러 개의 분류기를 생성하고 Prediction (예

134) 김민, 김동규(2022), 「Ensemble 알고리즘을 이용한 Machine Learning 하드웨어 Backdoor 탐지 방법」, 『대한전자공학회 학술대회』, 559.

135) 양호원(2008), 「신경망과 의사결정트리를 이용한 Stream Data 예측 시스템 설계 및 구현」, 조선대학교 대학원 박사학위 논문.

136) “LIGHTGBM 이란? 그리고 PARAMETER 튜닝하기”, <https://nurilee.com/2020/04/03/lightgbm-definition-parameter-tuning> (검색일: 2022. 10. 15.)

측)을 결합함으로써 더 정확도가 높은 예측을 도출하는 기법이다<sup>137)</sup>. 대표적인 Classification(분류) 알고리즘으로써 앙상블 학습을 통한 분류는 여러 개의 Classifier를 생성하고 그 예측을 결합함으로써 단일 분류기보다 정확한 최종 예측을 출력하는 기법을 말한다<sup>138)</sup>. Ensemble 학습의 유형은 Voting(보팅), Bagging(배깅), Boosting(부스팅), 그리고 Stacking(스태킹)의 네 가지로 나눌 수 있으며, 이외에도 웨이팅을 포함한 다양한 Ensemble 방법이 있다.



Ensemble Algorithm

그림 25 Ensemble Algorithm

### 1.2.2.7.1 Voting Ensemble

동일한 Datasets을 활용하여, 여러 개 학습모델의 결과를 바탕으로 Voting을 실시하여, 최종예측값을 도출하는 앙상블 기법이다. 즉, 여러 개의 분류기가 최종

137) 심대천 외(2022), 「XGBoost Regression 기계학습을 이용한 제주도 지하수 염 소이온 농도예측 연구 사례」, 『지질학회지』, 58(2), 243-255.

138) 김민, 김동규(2022). 「Ensemble 알고리즘을 이용한 Machine Learning 하드웨어 Backdoor 탐지 방법」, 『대한전자공학회 학술대회』, 559.

예측 결과를 투표를 통해 결정하는 방식이다. Voting과 Bagging의 다른 점은 보팅의 경우 일반적으로 서로 다른 알고리즘을 가진 분류기를 결합하여 학습을 진행해 Voting을 수행하는 것이다. Bagging의 경우 각각의 분류기가 모두 같은 유형의 알고리즘 기반이지만, 데이터 샘플링을 서로 다르게 가져가면서 학습을 수행해 보팅을 수행하는 것이다. Voting(보팅)의 유형은 Hard Voting(하드 보팅)과 Soft Voting(소프트 보팅)으로 나뉘어진다. Hard Voting은, 예측한 결과값 중 다수의 분류기가 결정한 예측 결과값을 최종 Voting 결과값으로 선정한다<sup>139)</sup>. Soft Voting은 각 분류기별 레이블값 결정 확률을 평균 낸 확률이 가장 높은 레이블 값을 최종 Voting 결과값으로 선정, 일반적으로 이 방식이 성능이 더 좋아 자주 사용된다. 대표적인 알고리즘으로 SVM, kNN, LR등이 있다.

#### 1.2.2.7.2 Bagging Ensemble

Bagging Ensemble의 경우, 여러 개의 분류기 중 예측 결과를 투표를 통해 결정하는 방식으로 Voting과 동일하다. 그러나, 앙상블에 사용된 각각의 분류기가 모두 같은 유형의 알고리즘 기반이지만, 데이터 처리방식을 서로 다르게 수행하면서 학습을 진행해 Voting을 수행하는 것이다. 대표적인 Bagging Ensemble 방식으로 Random Forest 알고리즘이 있다.

#### 1.2.2.7.3 Boosting Ensemble

Boosting은 여러 개의 분류기가 순차적으로 학습을 수행하되, 앞에서 학습한 분류기가 예측이 틀린 데이터에 대해서는 올바르게 예측할 수 있도록 다음 분류기에는 가중치(weight)를 부여하면서 학습과 예측을 진행한다<sup>140)</sup>. 계속해서 분류기에 가중치를 부스팅하면서 학습을 진행하기에 부스팅 방식으로 불린다. 예측 성능이 뛰어나 <sup>141)</sup>Ensemble Learning을 주도하고 있으며 대표적으로

---

139) “머신러닝 앙상블 학습(Ensemble Learning) - 보팅(Voting)“, <https://ek-koh.github.io/data%20analysis/voting> (검색일: 2022. 10. 15.)

140) 김민, 김동규(2022). 「Ensemble 알고리즘을 이용한 Machine Learning 하드웨어 Backdoor 탐지 방법」, 『대한전자공학회 학술대회』, 559.

141) 김태환 외(2020), 「설드 TBM 기계 데이터 및 머신러닝 기법을 이용한 암석의

XGBoost(eXtra Gradient Boost), LightGBM(Light Gradient Boost)가 있다.

#### 1.2.2.7.4 Stacking Ensemble

기존 Voting, Boosting 외에 최근 추가된 앙상블 방법이다. 여러 모델을 기반으로 예측된 결과를 Meta 학습기에 입력으로 넣고 최종 예측을 통해서 결과값을 출력하는 앙상블 학습기법을 말한다. 각 개별모델은, 예측한 데이터 출력한 후, 이렇게 각각 예측한 데이터를 다시 학습 데이터로 사용하여, 최종 모델을 선정하고 평가하게 된다.

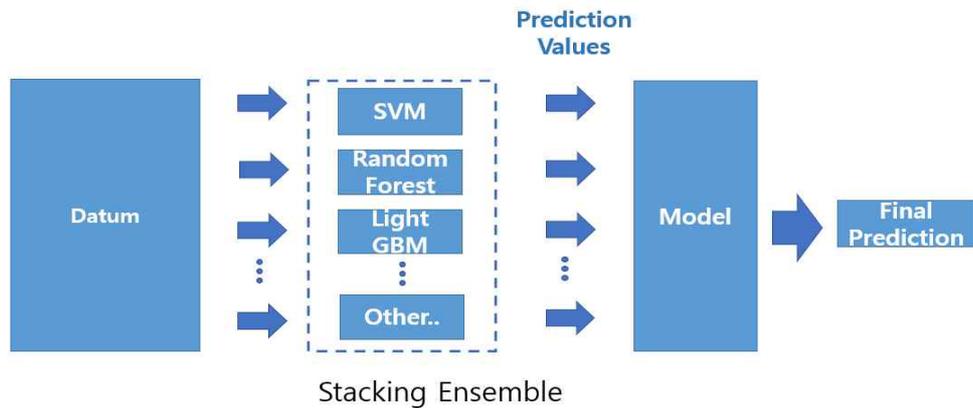


그림 26 Stacking Ensemble

#### 1.2.2.7.5 Weight Blending Ensemble

각 모델의 예측값에 대하여 Weight를 곱하여, 최종 출력을 계산하는 앙상블 방식으로 모델에 대한 가중치 조절을 통해서 최종예측값을 도출한다. 이때 가중치의 합은 1.0이 되도록 한다.

---

일축압축강도 예측」, 『터널과 지하공간』, 30(3), 214-225.

## 1.2.3 데이터 전처리 방법

### 1.2.3.1 하이퍼 파라미터

대부분의 머신러닝(ML) 방법은 적절한 하이퍼 파라미터 집합을 선택하여 특정 작업에 맞게 조정해야 한다. 예를 들어 랜덤 포레스트의 경우 추정기의 수, 추정기당 기능 수 또는 리프 당 최소 샘플 수를 조정해야 한다.

일반적으로, 데이터 세트는 트레인(Train), 검증(Validation) 및 테스트(Test) 세트로 분할된다. 다양한 하이퍼 파라미터 세트가 Train 데이터에 대해 훈련되고 검증 데이터에 대해 테스트 된다. 가장 성능이 좋은 모델은 최종 모델로 사용되며 훈련 및 검증 세트에서 재학습되고 테스트 세트에서 테스트 된다.

하이퍼 파라미터는 머신러닝에서, 모델 훈련 프로세스를 제어할 수 있는 조정 가능한 파라미터이다. 예를 들어, 신경망에서는 은닉층의 수와 각 층의 노드 수를 결정한다. 모델 성능은 하이퍼 파라미터에 크게 의존한다.

하이퍼 파라미터 최적화라고도 하는 하이퍼 파라미터 튜닝은 최상의 성능을 내는 하이퍼 파라미터의 구성을 찾는 프로세스이다<sup>142)</sup>. 이 프로세스는 일반적으로 계산 비용이 많이 드는 수동이다. 하지만 Auto ML을 사용하면 초매개변수인 하이퍼 파라미터 조정을 자동화하고, 실험을 빠른 속도로 병렬로 실행하여 초매개변수를 효율적으로 최적화할 수 있다. 이러한 이유로, Auto ML을 인공지능의 인공지능이라 칭하는 이유이며, 최대장점으로 꼽힌다.

XGBoost의 하이퍼 파라미터인 Colsample\_bytree는 의사결정 나무(DT) 생성에 필요한 변수(feature) 추출에 사용하며 과적합을 조정한다. Max\_depth는 의사결정나무의 최대 깊이를 조정하며 Min\_child\_weight는 Leaf node에 할당되는 최소 데이터의 수를 의미한다.<sup>143)</sup> Learning\_rate ( $\eta$ )는 의사결정나무의 개수를 조정하는 역할을 한다.

Gridsearch CV는 하이퍼 파라미터의 값들을 순차적으로 입력하여 최적의 하이퍼 파라미터를 탐색하는 격자 탐색(grid search)과 학습 데이터와 검증 데이

---

142) “Hyperparameter tuning a model with Azure Machine Learning (v1)”, <https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-tune-hyperparameters-v1?source=recommendations> (검색일: 2022. 10. 15.)

143) 심대천 외(2022), 「XGBoost Regression 기계학습을 이용한 제주도 지하수 염소이온 농도예측 연구 사례」, 『지질학회지』, 58(2), 243-255.

터를 K번 나누어 K번의 성능을 평균값으로 사용하는 교차검증(cross validation)을 합친 API이다<sup>144)</sup>

Stochastic search(확률적 탐색)은 주어진 시간 예산 내에서 사전 정의된 확률 분포에서 각 하이퍼 매개변수에 대한 후보 값을 무작위로 선택하여 하이퍼 매개변수를 최적화한다. 이것은 종종 수동 또는 Grid search보다 "더 크고 덜 유망한 구성 공간을 효과적으로 검색하여 더 나은 모델을 찾는다<sup>145)</sup>.

AutoML(Automated Machine Learning)은 자동화된 매개변수 튜닝에 대한 표준 방식보다 한 단계 더 나아간다<sup>146)</sup>. 개선된 모델을 생성하기 위해 여러 약한 회귀자의 시너지를 활용하는 앙상블 학습기를 구축한다. 즉, 랜덤 포레스트, 지원 벡터 머신(SVM) 같은 모델 클래스에서 모델을 동시에 선택하고 결합하는 동시에 하이퍼 매개변수를 최적화한다.

이를 위해 그리드 또는 확률적 검색에 의존하지 않고, 가우시안 프로세스를 기반으로 하는 효율적인 베이지안 최적화 방법을 활용하여 시간 또는 메모리 사용량과 같은 주어진 계산 예산 내에서 가장 유망한 모델 및 하이퍼 매개변수 조합을 지능적으로 선택한다<sup>147)</sup>. 대표적인 하이퍼 파라미터는 다음과 같이 구성된다.

- ▶ max\_depth : Tree의 최대 깊이를 말한다. 이 파라미터는 모델 과적합을 다룰 때 사용된다. 모델이 과적합 된 경우에는 max\_depth 값을 줄이게 된다.
- ▶ min\_data\_in\_leaf : Leaf가 가지고 있는 최소한의 레코드 수이다. 디폴트값은 20으로 최적값이며, 주로 적합을 해결할 때 사용되는 파라미터이다.
- ▶ lambda : lambda 값은 regularization 정규화를 한다. 일반적인 값의 범위는 0 에서 1 사이이다.
- ▶ learning\_rate : 최종 결과에 대한 각각의 Tree에 영향을 미치는 변수이다. GBM은 초기의

---

144) 이용준 외(2020), 「XGBoost를 활용한 고속도로 콘크리트 포장 파손 예측」, 『한국건설관리학회 논문집』, 21(6), 46 - 55.

145) Bergstra, J., Bengio, Y.(2012). "Random search for hyper-parameter optimization," *Journal of machine learning research*, 13(2).

146) Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F.(2015). "Efficient and robust automated machine learning," *Advances in neural information processing systems*, 28.

147) Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F.(2015). "Efficient and robust automated machine learning," *Advances in neural information processing systems*, 28.

추정값에서 시작하여 각각의 Tree 결과를 사용하여 추정값을 업데이트한다. learning rate는 이러한 추정에서 발생하는 변화의 크기를 컨트롤한다. 일반적인 값은 0.1, 0.001, 0.003 등등이 있다.

- ▶ num\_leaves : 전체 Tree의 leave 수이고, 디폴트값은 31이다.

### 1.2.3.2 최대 최소 정규화

변수 간의 척도가 다른 경우 머신러닝을 진행하면 학습이 제대로 되지 않는 문제가 있다. 이를 해결하기 위한 데이터 전처리 방법은 Min-Max Normalization 방법을 적용하였다. 이 방법은 데이터를 최대 최솟값을 이용하여 다음 식 (1)과 같이 0과 1 사이로 값을 정규화시킨다.

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

이를 통해서 변수 간의 편차를 줄이고 비슷한 형태로 데이터를 만들어서, 학습이 원활하게 하였다. Scikit-Learn에서는 다양한 종류의 스케일러를 제공하고 있다. 그중 대표적인 기법들이다<sup>148)</sup>.

표 28 스케일링의 종류

	종류	설명
1	StandardScaler	기본 스케일. 평균과 표준편차 사용
2	MinMaxScaler	최대/최솟값이 각각 1, 0이 되도록 스케일링

148) “[Python] 어떤 스케일러를 쓸 것인가?”, <https://mkjjo.github.io/python/2019/01/10/scaler.html> (검색일: 2022. 10. 15.)

3	MaxAbsScaler	최대 절댓값과 0이 각각 1, 0이 되도록 스케일링
4	RobustScaler	중앙값(median)과 IQR(interquartile range) 사용. 아웃라이어의 영향을 최소화

#### (1) StandardScaler

평균을 제거하고 데이터를 단위 분산으로 조정한다. 그러나 이상치가 있다면 평균과 표준편차에 영향을 미쳐 변환된 데이터의 확산은 매우 달라진다. 따라서 이상치가 있는 경우 균형 잡힌 척도를 보장할 수 없다.

#### (2) MinMaxScaler

모든 feature 값이 0~1사이에 있도록 데이터를 재조정한다. 다만 이상치가 있는 경우 변환된 값이 매우 좁은 범위로 압축될 수 있다. 즉, MinMaxScaler 역시 아웃라이어의 존재에 매우 민감하다.

#### (3) MaxAbsScaler

절대값이 0~1사이에 매핑되도록 한다. 즉 -1~1 사이로 재조정한다. 양수 데이터로만 구성된 특징 데이터셋에서는 MinMaxScaler와 유사하게 동작하며, 큰 이상치에 민감할 수 있다.

#### (4) RobustScaler

아웃라이어의 영향을 최소화한 기법이다. 중앙값(median)과 IQR(interquartile range)을 사용하기 때문에 StandardScaler와 비교해보면 표준화 후 동일한 값을 더 넓게 분포시키고 있음을 확인할 수 있다.

$IQR = Q3 - Q1$  : 즉, 25퍼센타일과 75퍼센타일의 값들을 다룬다.

## 1.2.4 모델 평가 방법

예측된 결과는 다음과 같은 모델 식을 통해서 평가하게 된다. 식에서  $y$ ,  $\hat{y}$ ,  $\bar{y}$  는 각 실제값과 예측값, 평균값을 의미하며,  $n$ 은 자료의 수를 의미한다 (식 1,2,3,4)

### (1) Mean Absolute Error (MAE)

절대 오차의 평균이다. 오차는 예측된 값과 실제 값의 차이이다<sup>149)</sup>. 실제 값과 예측값의 차이를 절댓값으로 변환 뒤 합산하여 평균으로 구한다.

$$MAE = \frac{\sum_{i=1}^n y - \hat{y}}{n} \quad (2)$$

### (2) Mean Squared Error (MSE)

실제값과 예측값의 차이를 제곱하여 평균으로 구하는 MSE는 테스트 데이터 세트에서 수행된 예측의 제곱 오차 평균이다. MSE가 작을수록 좋지만, MSE를 과도하게 줄이면 과적합의 오류를 범할 가능성이 커진다.

$$MSE = \frac{\sum_{i=1}^n (y - \hat{y})^2}{n} \quad (3)$$

---

149) “Tutorial: Designer - train a no-code regression model”, <https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-designer-auto-mobile-price-train-score> (검색일: 2022. 10. 15.)

(3) Root Mean Squared Error (RMSE)

테스트 데이터 세트에서 수행된 예측의 제공 오차 평균의 제곱근이다. 평균제곱 오차에 제곱근을 씌워 구하는 RMSE

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y})^2}{n}} \quad (4)$$

(4) Mean Absolute Percentage Error (MAPE)

절대 오차 평균의 백분율 값이다<sup>150)</sup>. 0에 가까울수록 회귀모형의 성능이 좋다고 해석할 수 있다.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y - \hat{y}}{y} \right| \quad (5)$$

(5) R<sup>2</sup> Coefficient of Determination (결정계수)

R 제곱 값이라고도 하는 이 통계 메트릭은 모델이 데이터에 얼마나 잘 맞는지를 나타낸다. 통상적으로 0에 근접할수록 쓸모없는 회귀식으로, 1에 가까울수록 모델이 우수한 성능을 보인다고 평가한다. 결정계수는 회귀모델에서 독립변수가 종속변수를 얼마만큼 설명해주는지를 가리키는 지표이며 설명력이라 부르기도 한다. 즉, 결정계수가 높을수록 독립변수가 종속변수를 많이 설명한다는 뜻이다.

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = \frac{\sum_{i=1}^n (y - \hat{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (6)$$

---

150) “Tutorial: Designer - train a no-code regression model”, <https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-designer-auto-mobile-price-train-score> (검색일: 2022. 10. 15.)

## 2. 인공지능 모델링을 통한 서울숲 스마트공원 구현

### 2.1 인공지능을 이용한 공기질과 보행량의 상관관계

#### 2.1.1 데이터 전처리

데이터 세트는 일반적으로 분석 전에 사전 처리가 필요하다. 왜냐하면, 데이터 세트를 검사할 때 일부 누락된 값을 발견했을 수 있다. 모델이 데이터를 올바르게 분석할 수 있도록 이러한 누락된 값을 정리해야 한다. 직접 제작하여 설치한 IoT 장비를 통하여 습득한 데이터의 실효성을 파악하기 위해서, 서울시에서 측정된 공공데이터(서울 열린데이터 광장) 중, 평균온도를 서울숲에서 취득한 데이터와의 트렌드를 비교하여, 데이터의 유효성을 비교하였다. 데이터 비교를 위해서 사용한 소프트웨어 환경은 구글 Colab, 파이썬 언어를 사용하여 작업을 하였다. 설치한 기기의 온도는, 실험한 장비 내부의 표면온도를 측정하여 보니, 직사광선을 쬐이는 장비 특성상, 서울시에서 수집한 대기 온도보다는 최대 12도 정도 높게 측정되었으나, 그림 27의 데이터 그래프상의 월평균 차이와 하루평균 차이 곡선을 비교해보면, 매우 리니어한 것을 파악할 수 있다.

동일한 방식으로 설치한 IoT장비를 통하여 습득한 데이터의 실효성을 파악하기 위해서, 서울시에서 측정된 공공데이터 중, 평균습도를 서울숲에서 취득한 데이터와의 트렌드를 비교하여, 데이터의 유효성을 비교하였다. 온도 데이터에서 추정된 이유와 같은 원리에서, 설치한 기기의 습도는, 실험 측정용 박스 안의 표면온도를 측정한 것으로, 직사광선을 쬐이는 장비 특성상, 서울시에서 수집한 대기습도보다는 최대 30% 정도 낮게 측정되었으나, 그림 28의 데이터 그래프상의 월평균 차이와 하루평균 차이 곡선을 비교해보면, 매우 리니어한 것을 파악할 수 있다.

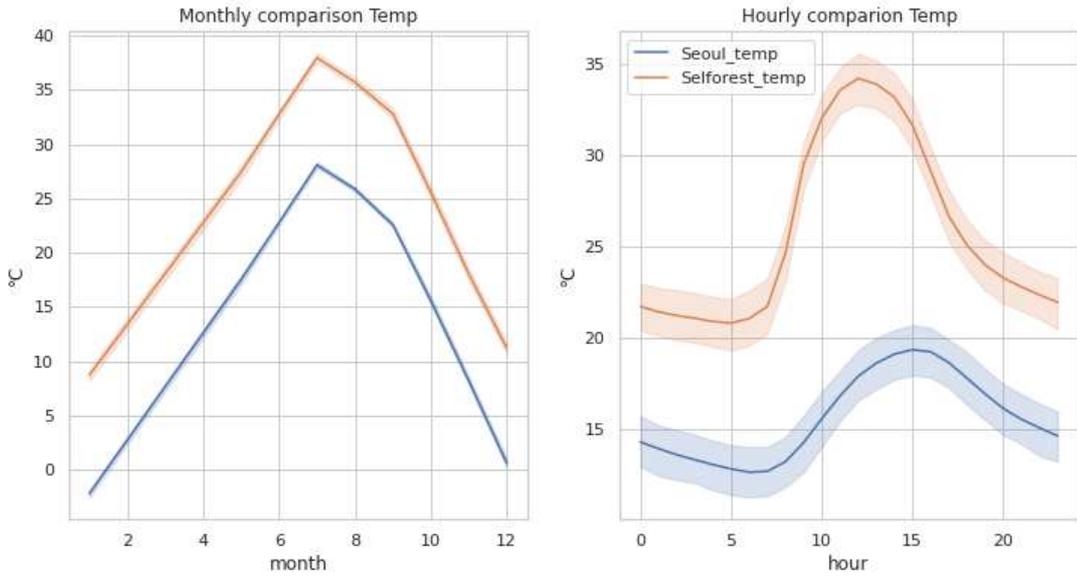


그림 27 서울시 평균 데이터와 IoT장비를 활용하여 측정한 온도 트렌드 비교

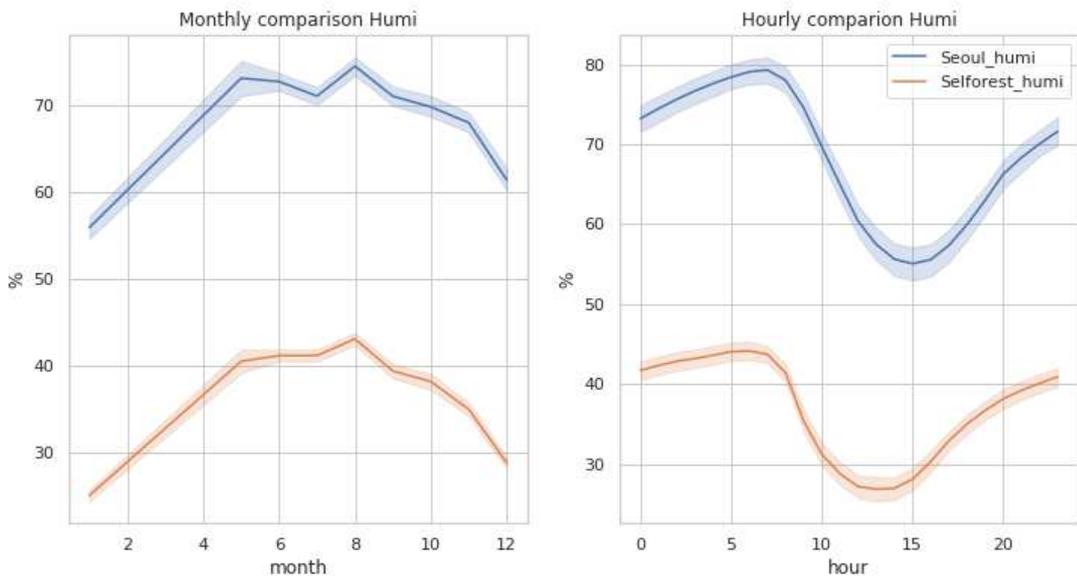


그림 28 서울시 평균 데이터와 IoT장비를 활용하여 측정한 습도 트렌드 비교

연간 보행량 데이터인 2021년도 5월 21일부터 2022년 1월 26일까지의 보행 데이터와 기상 데이터를 기준으로 데이터 분석을 시행한다. 데이터 전처리를 통해서, 기존 수집된 수치형 데이터 외에 다른 범주형 데이터를 추가하였다. 계절 (season)은 4개의 선택변수에서 데이터의 가중치를 조정하기 위하여, 원-핫인코딩(one-hot encoding)을 사용하여 봄(1), 여름(2), 가을(3), 겨울(4)로 배정하고, 휴무일(holiday)은 주말과 공휴일을 1로 정하고, 평일은 반대로 일하는 날을 1로 정하였다. 원-핫 인코딩은 단어 집합의 크기를 벡터의 차원으로 하고, 표현하고 싶은 단어의 인덱스에 1의 값을 부여하고, 다른 인덱스에는 0을 부여하는 단어의 벡터 표현 방식이다. 우적량(rainfall), 풍속(Wind speed), 운량(cloud)은 기상청 기상자료 개방 포털(data.kma.go.kr)에서 공공데이터를 활용하여 시간 단위로 구하였다.

이리하여 그림 29 처럼 파이선으로 전체 데이터를 읽어들이고, 헤드 값을 표시하였다.

Unnamed: 0	datetime	sel_temp	sel_humi	Spring	Summer	Autumn	Winter	season	holiday	...	cloud	pm	humi	temp	co	count	year	month	day	hour
0	2021-05-21 11:00:00	15.7	89	1	0	0	0	1	0	...	10	23.97	46.37	27.93	0.30	113	2021	5	21	11
1	2021-05-21 12:00:00	16.3	88	1	0	0	0	1	0	...	10	25.44	44.49	28.97	0.31	78	2021	5	21	12
2	2021-05-21 13:00:00	16.9	83	1	0	0	0	1	0	...	10	21.50	41.40	30.65	0.33	94	2021	5	21	13
3	2021-05-21 14:00:00	17.7	80	1	0	0	0	1	0	...	10	19.38	43.61	27.45	0.30	58	2021	5	21	14
4	2021-05-21 15:00:00	16.9	87	1	0	0	0	1	0	...	10	14.42	49.28	24.87	0.29	78	2021	5	21	15

그림 29 파이선으로 읽어들이고 표시한 데이터의 헤드 정보

여기에서, 온도 데이터의 특성을 살펴보면, 모두 6,013개의 자료가 수집되었고, 최고온도 52.14℃, 최저 온도 -6.8℃, 평균온도 25.73℃에 해당한다.

```
[14] train.temp.describe()

count      6012.000000
mean       25.737821
std        12.188387
min        -6.860000
25%        15.737500
50%        28.475000
75%        34.310000
max         52.140000
Name: temp, dtype: float64
```

그림 30 파이선으로 읽어들이고 표시한 온도 특성

```

train.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6012 entries, 0 to 6011
Data columns (total 24 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Unnamed: 0             6012 non-null   int64
1   datetime               6012 non-null   datetime64[ns]
2   Sel_temp               6012 non-null   float64
3   Sel_humi               6012 non-null   int64
4   Spring                 6012 non-null   int64
5   Summer                 6012 non-null   int64
6   Autumn                 6012 non-null   int64
7   Winter                 6012 non-null   int64
8   season                 6012 non-null   int64
9   holiday                6012 non-null   int64
10  workingday             6012 non-null   int64
11  rainfall               6012 non-null   float64
12  windspeed              6012 non-null   float64
13  winddirection          6012 non-null   int64
14  cloud                  6012 non-null   int64
15  pm                     6012 non-null   float64
16  humi                   6012 non-null   float64
17  temp                   6012 non-null   float64
18  co                     6012 non-null   float64
19  count                  6012 non-null   int64
20  year                   6012 non-null   int64
21  month                  6012 non-null   int64
22  day                    6012 non-null   int64
23  hour                   6012 non-null   int64
dtypes: datetime64[ns](1), float64(7), int64(16)
memory usage: 1.1 MB

```

그림 31 파이선으로 읽어들이는 데이터 특성

```

train.isnull().sum()
Unnamed: 0      0
datetime        0
Sel_temp        0
Sel_humi        0
Spring          0
Summer          0
Autumn          0
Winter          0
season          0
holiday         0
workingday      0
rainfall        0
windspeed       0
winddirection   0
cloud           0
pm              0
humi            0
temp            0
co              0
count           0
year            0
month           0
day             0
hour            0
dtype: int64

```

그림 32 Null 값을 갖는 데이터 수

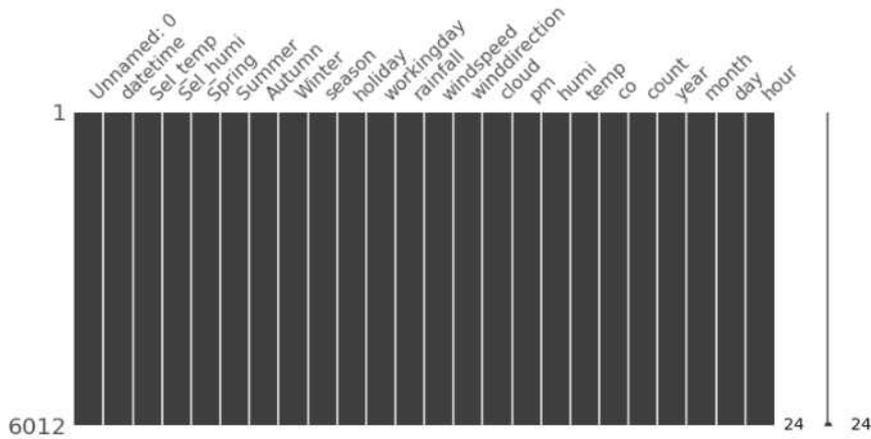


그림 33 데이터 중 Null 값 분포

그림 33처럼 각각 데이터의 Null 값을 제외한 일반 데이터의 구성을 표시하였으며 데이터에 Null 값 없이 모두 만족한 것을 확인할 수 있다. 그리고, 수집된 데이터의 유의성을 테스트하기 위하여, IBM사의 SPSS 소프트웨어를 사용하여, 아노바테스트를 실시하였다. 보행량 변수인 'count' 를 target으로 설정하고 수행하였다. 측정된 데이터 대부분이 P Value 가 0.05 또는 0.01 이하로, 유의성 검증에 문제가 없음을 확인할 수 있었다. 다만, 강우량(rainfall), 일산화탄소(co), 풍향(winddirection), 연도(year)는 유의성이 기각되었다.

표 29 서울숲 데이터 유의성 검증 테스트 결과

		ANOVA				
		제곱합	자유도	평균제곱	F	유의확률
Spring	집단-간	18.858	315	.060	1.526	.000
	집단-내	223.496	5696	.039		
	전체	242.353	6011			
Summer	집단-간	234.963	315	.746	3.656	.000
	집단-내	1162.115	5696	.204		
	전체	1397.078	6011			
Autumn	집단-간	86.883	315	.276	1.205	.009
	집단-내	1303.728	5696	.229		
	전체	1390.611	6011			
Winter	집단-간	292.773	315	.929	6.935	.000
	집단-내	763.400	5696	.134		
	전체	1056.173	6011			
season	집단-간	1196.888	315	3.800	7.008	.000
	집단-내	3088.314	5696	.542		
	전체	4285.202	6011			
holiday	집단-간	111.050	315	.353	1.667	.000
	집단-내	1204.351	5696	.211		
	전체	1315.401	6011			
workingday	집단-간	107.865	315	.342	1.651	.000
	집단-내	1181.237	5696	.207		
	전체	1289.102	6011			
rainfall	집단-간	230.806	315	.733	.426	1.000
	집단-내	9793.908	5696	1.719		
	전체	10024.713	6011			
windspeed	집단-간	385.577	315	1.224	1.155	.034
	집단-내	6038.577	5696	1.060		
	전체	6424.154	6011			
winddirection	집단-간	3827712.379	315	12151.468	1.010	.444
	집단-내	68552497.385	5696	12035.200		
	전체	72380209.764	6011			
cloud	집단-간	5411.430	315	17.179	1.176	.020
	집단-내	83199.839	5696	14.607		
	전체	88611.269	6011			
pm	집단-간	229335.825	315	728.050	1.483	.000
	집단-내	2795743.579	5696	490.826		
	전체	3025079.404	6011			
humi	집단-간	107454.395	315	341.125	2.681	.000
	집단-내	724843.474	5696	127.255		
	전체	832297.869	6011			
temp	집단-간	156774.999	315	497.698	3.851	.000
	집단-내	736199.721	5696	129.249		
	전체	892974.719	6011			
co	집단-간	7.233	315	.023	.888	.919
	집단-내	147.237	5696	.026		
	전체	154.470	6011			
year	집단-간	26.466	315	.084	.900	.894
	집단-내	531.974	5696	.093		
	전체	558.441	6011			
month	집단-간	4487.217	315	14.245	1.502	.000
	집단-내	54020.517	5696	9.484		
	전체	58507.734	6011			
day	집단-간	27898.340	315	88.566	1.146	.042
	집단-내	440205.957	5696	77.283		
	전체	468104.297	6011			
hour	집단-간	60090.379	315	190.763	4.769	.000
	집단-내	227852.022	5696	40.002		
	전체	287942.401	6011			

## 2.1.2 파이선을 이용한 IoT 데이터와 보행량 상관관계분석

수집된 데이터를 기반으로, 본격적인 머신러닝을 수행하고, 데이터 분석을 한다. 먼저, 시간 분석을 통한, 연간 방문객 수, 월별 방문자 수, 일간 방문자 수, 매일 시간별 방문자 수(보행량)의 통계를 분석하면 다음과 같다. 실험에 사용된 환경은, 구글(Google)사의 클라우드 기반 AI 테스트환경인, 코랩(Colab), 파이선(Python) 언어와 판다스(Pandas), 넘파이(numpy) 씨본(seaborn), 사이킷 런(Psykit learn) 등의 라이브러리를 활용하였다.

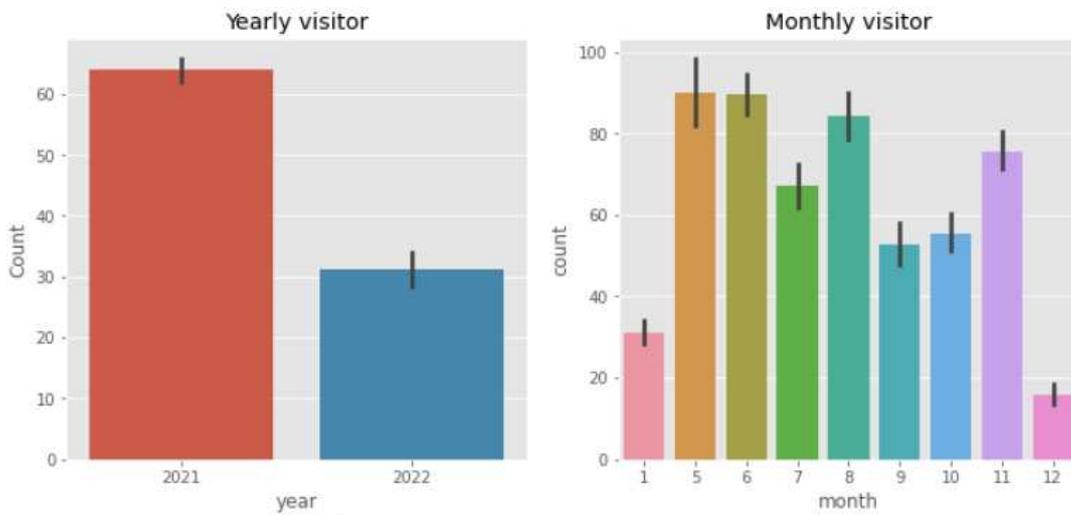


그림 34 파이선 분석을 통한 전체 연간, 월간, 보행량 수 트렌드 분석

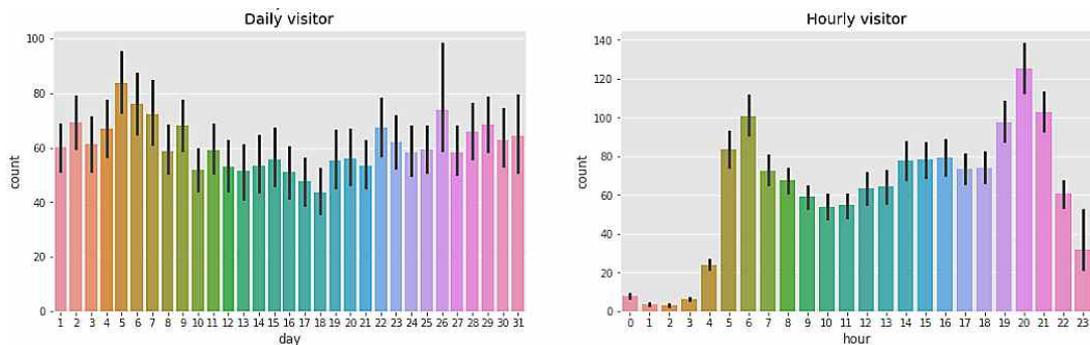


그림 35 파이선 분석을 통한 전체 일간, 시간당 보행량 수 트렌드 분석

일단 데이터수집량이 2021에 비해 2022년이 적었기 때문에, 연간보행량의 수는 2021년도가 약 2배 정도 높은 것을 알 수 있었다. 연간 보행량의 수를 월로 표현하였을 때, 시간당 보행량을 비교해본다면, 봄철에는 보행량이 늘어나며, 여름에는 줄고, 가을에는 2021년 상황에서 코로나 등의 영향으로 줄어드는 것을 알 수 있다. 막대그래프에 의하면, 2021년 5, 6월에 보행자는 증가하는 추세를 보였으나, 겨울철에 급감하는 것과 폭염인 여름철인 7월에도 감소하는 것을 확인할 수 있었다. 한편, 코로나 감염자 수가 증가한 9월과 10월에는 계절적인 요소와 관계없이 급감하는 것을 확인할 수 있었다.

여기에서 성동구의 코로나 감염자 수(서울시 공공데이터 참조)는 다른 환경 데이터와는 별개로, 시계열 데이터상에서 매시간별 추이를 알 수 없고, 일간 데이터만 공급된 이유로 기존 데이터 사용을 하지 않고, 별개의 일간 데이터로만 살펴 보았다.

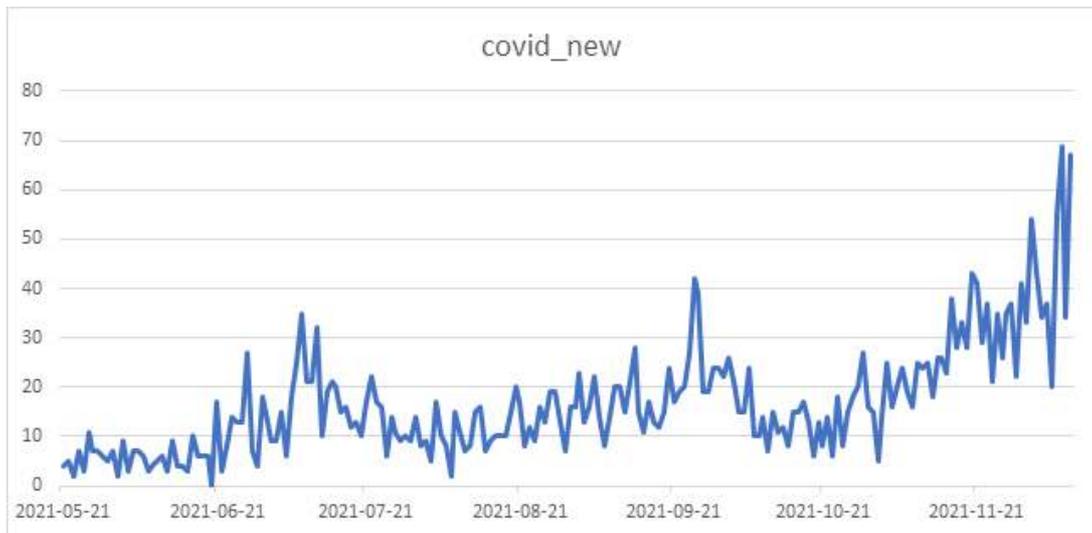


그림 36 2021년 5월 ~12월 10일 성동구 코로나 일간 확진자 수 추이  
(자료 : 서울시 공공데이터)

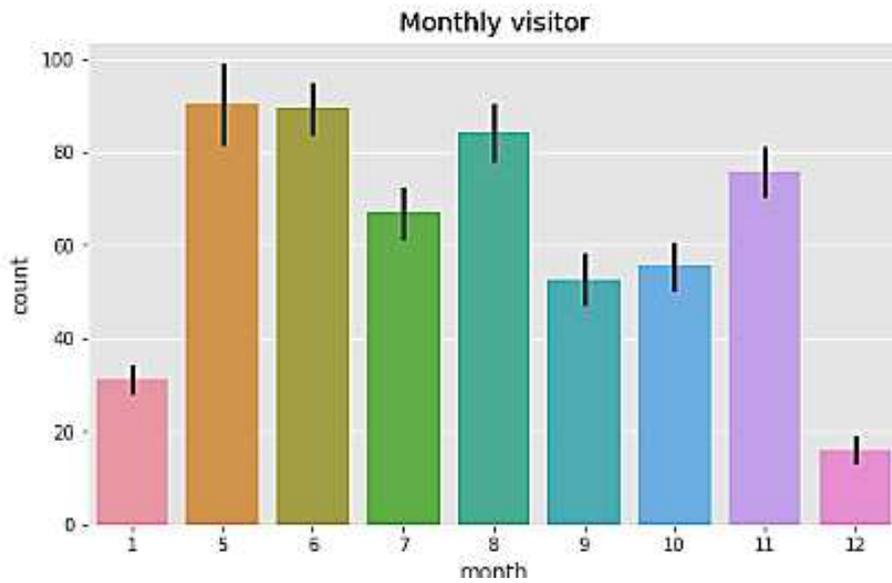


그림 37 연간 단위 매월 평균 시간당 평균 보행량 수  
Daily visitor

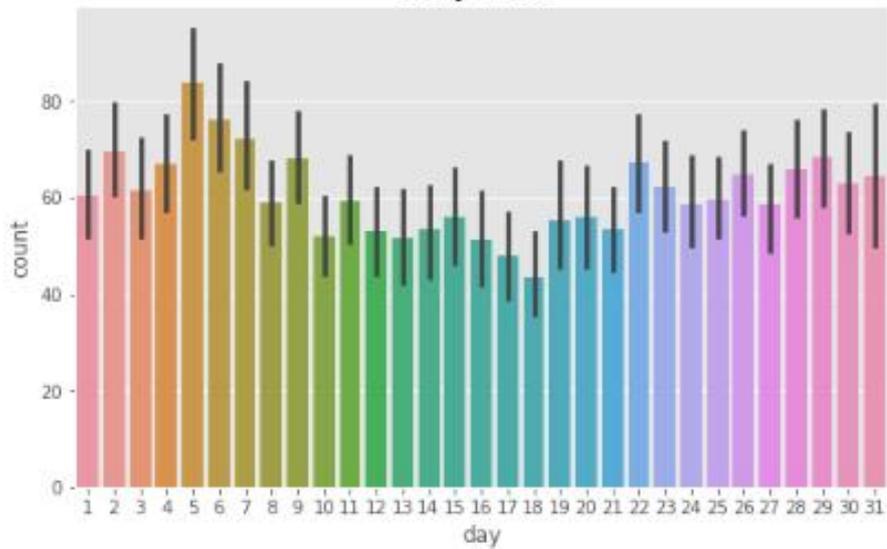


그림 38 월간 단위 매일 시간당 평균 보행량 수

그림 38에서 월간 단위 시간당 매일 평균 보행량 수를 기준으로 살펴본다면, 월 초순 또는, 하순에 포진이 되어있는 것을 확인할 수 있다. 이것은 월 기준으로 어떤 연유에서 월 중순보다는 상순과 하순에 집중되어있는지에 대한 부분을 추후 검증해볼 필요가 있다. 일자별, 24시간 방문자를 비교해보면, 오전 5~7시 새벽 운동시간과 저녁 시간인 19시 ~ 21까지 저녁 산책 시간에 보행량이 증가하는 것을 알 수 있다.

### 2.1.3 박스 플롯을 통한 데이터 비교

박스 플롯(상자그림) 기술 통계학에서 '상자 수염 그림' 또는 '상자 그림'은 수치적 자료를 표현하는 그래프이다<sup>151)</sup>. 이 그래프는 가공하지 않은 자료 그대로를 이용하여 그린 것이 아니라, 자료로부터 얻어낸 통계량인 5가지 요약 수치를 가지고 그린다. 이때 5가지 요약 수치란 최솟값, 제 1사 분위 Q1, 제 2사 분위 Q2, 제 3사 분위 Q3, 최댓값을 일컫는 말이다. 히스토그램과는 다르게 집단이 여러 개인 경우에도 한 공간에 수월하게 나타낼 수 있다. 여기서 4분위는 데이터를 사 등분 한 것이며, 중위수는 데이터의 한가운데 순위에 해당하는 값, 제 3사 분위 수 (Q3) :중앙값 기준으로 상위 50% 중의 중앙값, 전체 데이터 중 상위 25%에 해당하는 값 제 1사 분위 수 (Q1):중앙값 기준으로 하위 50% 중의 중앙값, 전체 데이터 중 하위 25%에 해당하는 값이다. 본 박스 플롯에서는 하루 평균 보행자가 대략 3천 명 내외인 것을 알 수 있다.

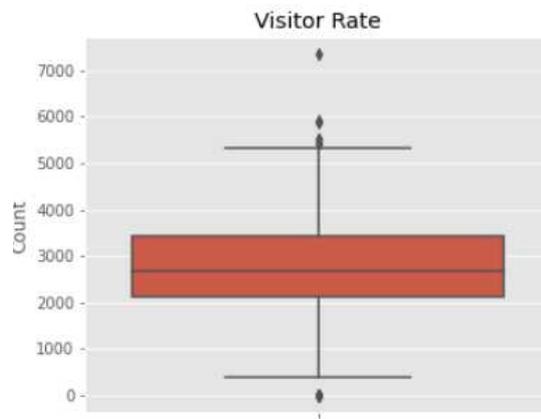


그림 39 1일 보행량 수 박스 플롯

151) “상자 수염 그림”, 위키피디아, [https://ko.wikipedia.org/wiki/%EC%83%81%EC%9E%90\\_%EC%88%98%EC%97%BC\\_%EA%B7%B8%EB%A6%BC](https://ko.wikipedia.org/wiki/%EC%83%81%EC%9E%90_%EC%88%98%EC%97%BC_%EA%B7%B8%EB%A6%BC) (검색일: 2022년 10월 15일)

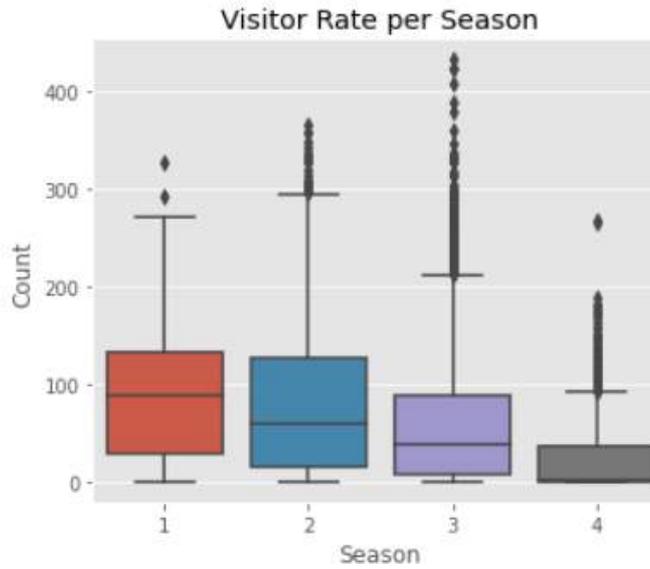


그림 40 서울숲공원 계절별 평균 시간당 보행량 상관계분석 변화량

그림 40은 봄·여름·가을·겨울의 평균 시간당 보행자 수를 비교한 박스 플롯 차트이다. 연간 보행량의 수를 계절로 표현하였을 때 봄철에는 보행량이 늘어나며, 겨울철에 확실히 방문자 수가 적은 것을 확인할 수 있다. 한편, 앞서서 설명한 바와 같이 코로나 확진자 수가 증가한 9월과 10월에는 계절적인 요소와 관계없이 급감하는 것을 확인할 수 있었다. 그림 42와 같이 휴무일과 일하는 날을 비교하였을 때는, 휴무일의 보행량이 높은 것을 박스 플롯으로 확인할 수 있다.

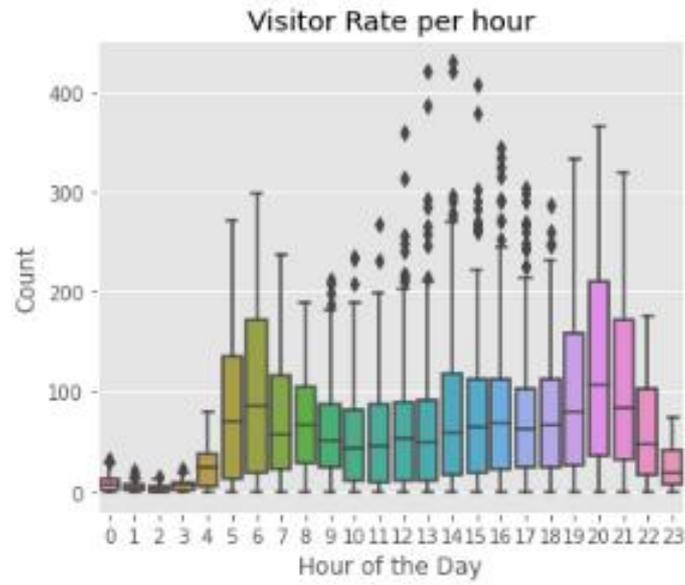


그림 41 서울숲공원 24시간 평균 보행량 상관관계 분석 변화량

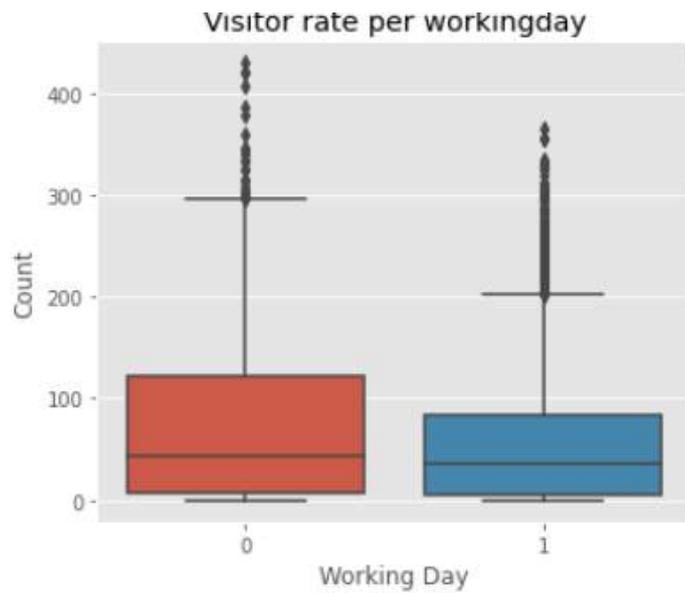


그림 42 서울숲공원 평일 보행량 상관관계분석 변화량

## 2.1.4 수치, 범주 데이터 조건별 보행량 상관관계분석 변화량

꺾은 선 그래프를 통한 일간 데이터 분석을 살펴보면, 휴무일 방문 시간과 일하는 평일의 경우, 휴일의 방문 시간 주기가 골고루 분포된 것과 달리, 평일에는 주로 밤과 오전 시간에 집중된 것을 확인할 수 있었다. 또한, 주일과 주말의 보행량을 비교하여 본다면, 평일, 휴일과 동일한 패턴으로, 낮 시간의 보행량이 주말에 훨씬 높은 것을 알 수 있었다. 계절별 시간대별 보행량 또한, 봄철과 가을철에는 골고루 분포된 것과 달리, 폭염기인 여름철에는 대낮에는 보행량이 확연히 줄어드는 것을, 겨울철에는 따스한 오후 시간에 주로 보행량이 증가하는 것을 확연히 알 수 있다.

히트 맵 히트 맵(heat map)은 열을 뜻하는 히트(heat)와 지도를 뜻하는 맵(map)을 결합한 단어로, 색상으로 표현할 수 있는 다양한 정보를 일정한 이미지 위에 열 분포 형태의 비주얼한 그래픽으로 출력하는 것이 특징이다<sup>152)</sup>.

주로 웹사이트의 방문자를 분석하는 웹로그 분석에 많이 사용하는 분석 기법으로 웹페이지에서 발생하는 방문자의 마우스 클릭을 열 분포 형태의 이미지로 변환하여 사이트 이미지 위에 겹쳐서 보여주며, 클릭이 많이 발생하는 영역은 빨간색으로 클릭이 적게 발생하는 영역은 파란색으로 표현한다.

---

152) “히트맵”, 위키피디아, [https://ko.wikipedia.org/wiki/%ED%9E%88%ED%8A%B8\\_%EB%A7%B5](https://ko.wikipedia.org/wiki/%ED%9E%88%ED%8A%B8_%EB%A7%B5) (검색일 2022년 10월15일)

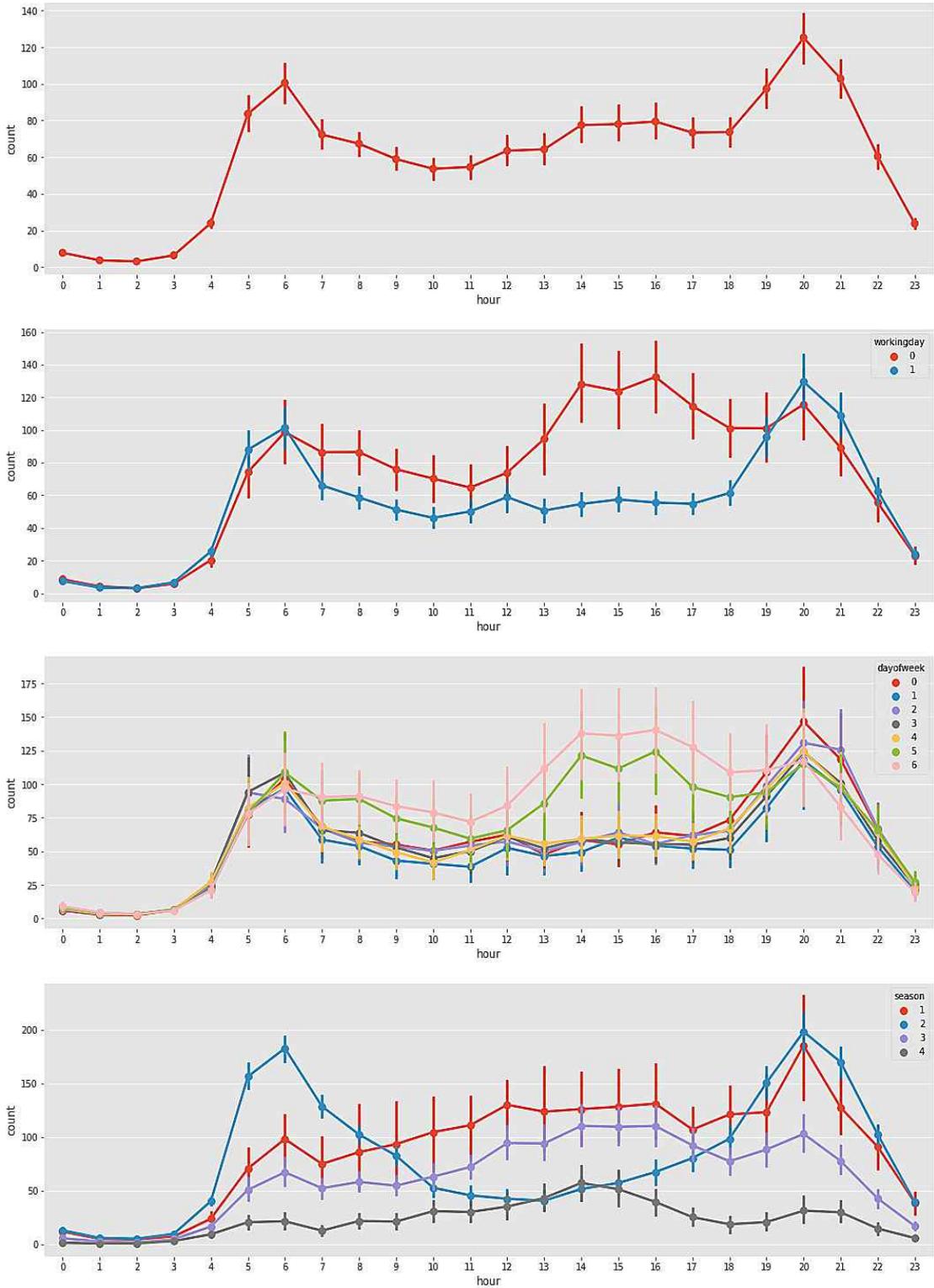


그림 43 서울숲공원 수치, 범주 데이터 조건별 보행량 상관관계분석 변화량  
 범례) 요일 : 월요일0, 화요일1, 수요일2, 목요일3, 금요일4, 토요일5, 일요일6 / 계절 : 봄1, 여름2, 가을3, 겨울4

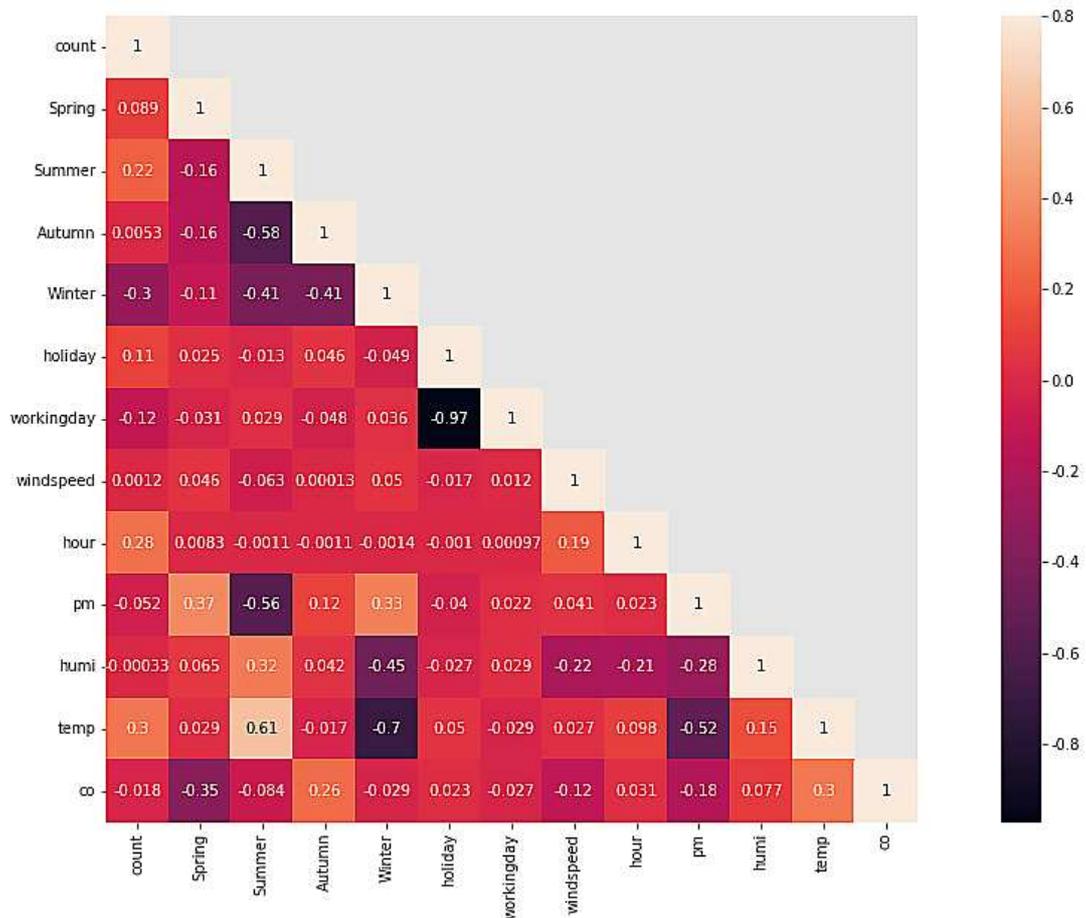


그림 44 히트 맵을 통한 서울숲 보행량 상관관계분석

히트 맵을 분석하면, 공원의 보행량 수는 온도와 상관관계가 0.3으로 가장 높고, 그리고, 시간 데이터와도 상관관계가 0.28, 겨울 계절과는 -0.3, 여름에는 0.22 순으로 높게 측정되었다. 아울러 휴일에는 0.11, 평일에는 -0.12 등으로 영향이 큰 것을 확인할 수 있었다. 그리고, 온도와 겨울철 계절은 -0.7로 역 상관관계가 매우 높은 것을 온도와 미세먼지와도 역 상관관계가 -0.52로 높다.

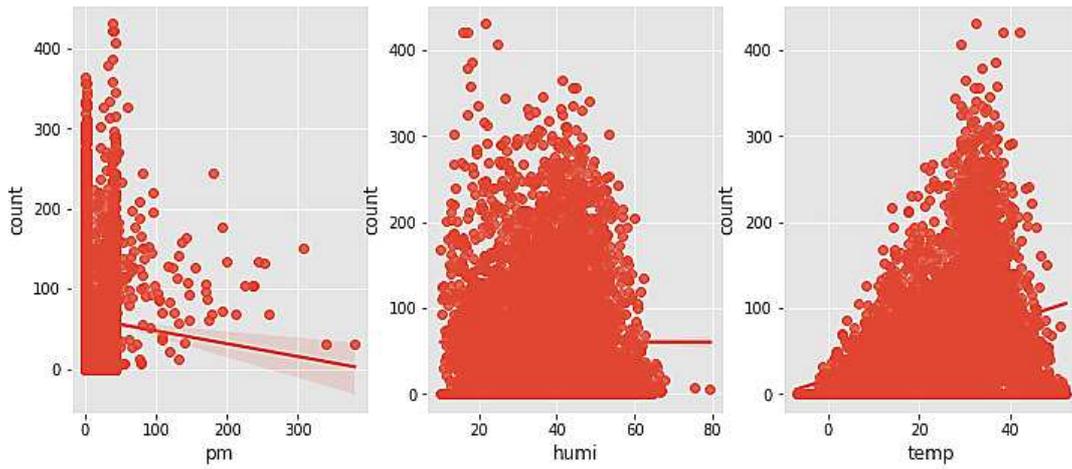


그림 45 선형회귀를 사용한 보행량 상관관계분석

온도와 습도와 보행량의 선형회귀 상관관계에서, 쾌적한 날씨에서 보행량의 수가 많은 것을 확인, 미세먼지 수치 또한 보행량에 부정적인 영향을 주는 것을 확인하였다. 풍속, 강우량, 운적량 중에서, 선형회귀를 통해서 살펴보았을 때는, 강우량이 가장 크게 보행자의 수에 반비례하는 것을 확인할 수 있다.

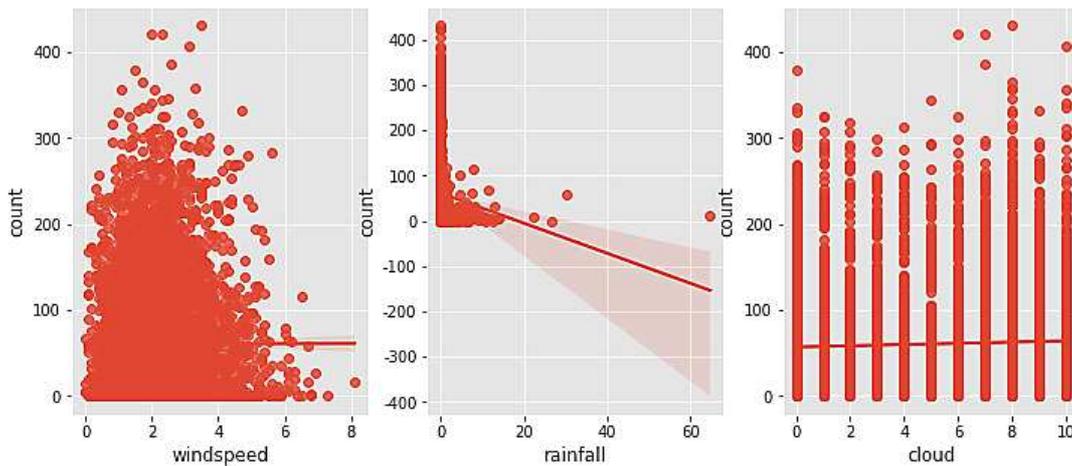


그림 46 선형회귀를 사용한 보행량 상관관계분석

## 2.2 머신러닝 분석

앞장 일반적인 선형회귀(Regression) 데이터를 단순한 수치 분석학 결과에 비교하여, 좀 더 세밀하게 보행량 예측에 크게 영향을 주는 요소를 머신러닝을 활용하여, 살펴본다. 또, 사용된 머신러닝 알고리즘에 따른, 평가도에 따라서, 최고의 효용성을 갖춘 알고리즘을 살펴보고자 한다. 사용된 데이터는, 앞절에서 파이선으로 데이터 분석되어 보인 환경 빅데이터와 같으며, 70:30의 비율로 트레이닝 데이터(Training Data)와 밸리테이션 데이터(Validation Data)로 사용되었다.

### 2.2.1 머신러닝 Best model

본 회귀분석에 사용된 머신러닝은 Microsoft사의 AutoML을 사용하였으며, 그중에서 Ensemble 모델, XG Boost regressor, LGBM, Random forest의 4가지 모델만을 사용하여, 알고리즘에 대한 정확성 등을 평가하고자 한다.

각각의 알고리즘에 사용된 전처리 모듈과 엔진, 본 처리에 사용된 알고리즘과 엔진은 표 30 와 같다 먼저, 첫 번째 Voting Algorithm은 XGboostRegressor 3개와 ExtraTree regressor, LGBM 의 5가지 알고리즘을 사용하였으며, 모듈은 Scikit Learn과 AutoML을 각각 사용하였다. 전처리에 사용된 스케일러는 각각, Standard scaler와 MaxAbsScaler등이 활용되었다.

두 번째로, XGboost regressor는 StandardScaler를 전처리로 사용한다. 세 번째로, Light GBM은 MaxAbsScaler를 사용하여 전처리를 진행한다. 끝으로, RandomForestRegressor는 StandardScaler를 사용하여 전처리를 진행한다.

표 30 머신 러닝에 사용된 전처리 모듈과 엔진, 본 처리에 사용된 알고리즘과 엔진

Algorithm Name	pre processing	Preprocessing	Algorithms	module
Voting Ensemble	Standard Scaler	sklearn	XGBoostRegressor	automl
	StandardScaler	sklearn	XGBoostRegressor	automl
	MaxAbsScaler	sklearn	ExtraTreesRegressor	sklearn
	MaxAbsScaler	sklearn	XGBoostRegressor	automl
	MaxAbsScaler	sklearn	LightGBMRegressor	automl
XGBoostRegressor	StandardScaler	sklearn	XGBoostRegressor	automl
LightGBM	MaxAbsScaler	sklearn	LightGBM	sklearn
RandomForestRegressor	StandardScaler	sklearn	RandomForestRegressor	sklearn

## 2.2.2 앙상블 인공지능 머신러닝 모델링을 통한 선형회귀 효 용성 분석

그림47 다이어그램은 이 특정 모델을 생성하기 위해 Automated ML이 적용한 데이터 사전 처리, 기능 엔지니어링, 확장 기술 및 머신러닝 알고리즘을 보여준다.

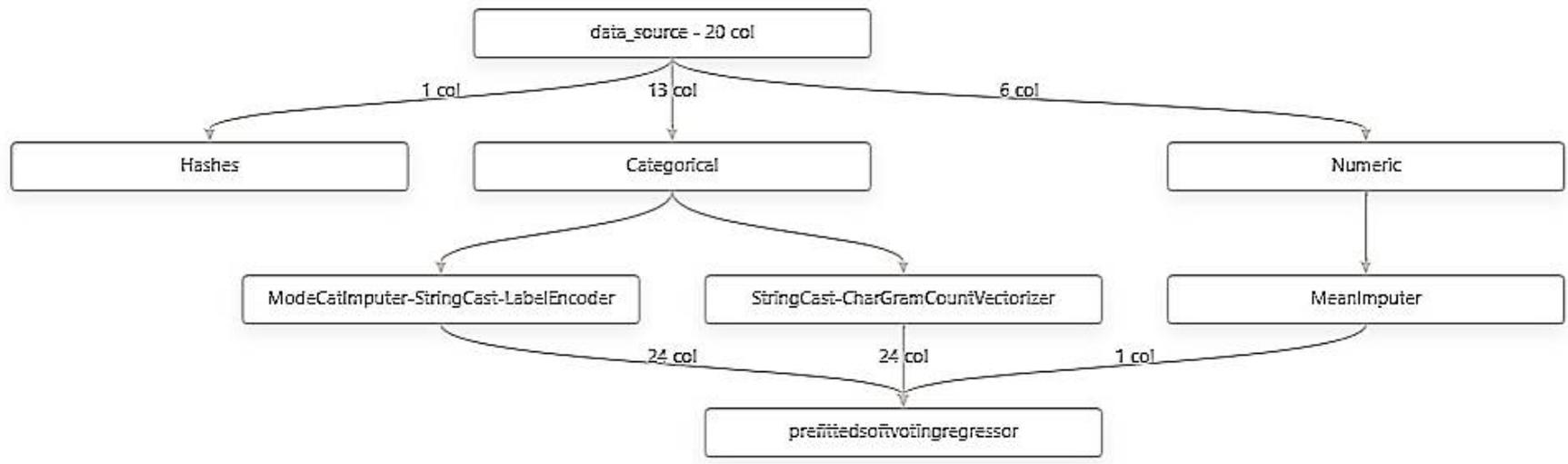


그림 47 Voting Algorithm 다이어그램

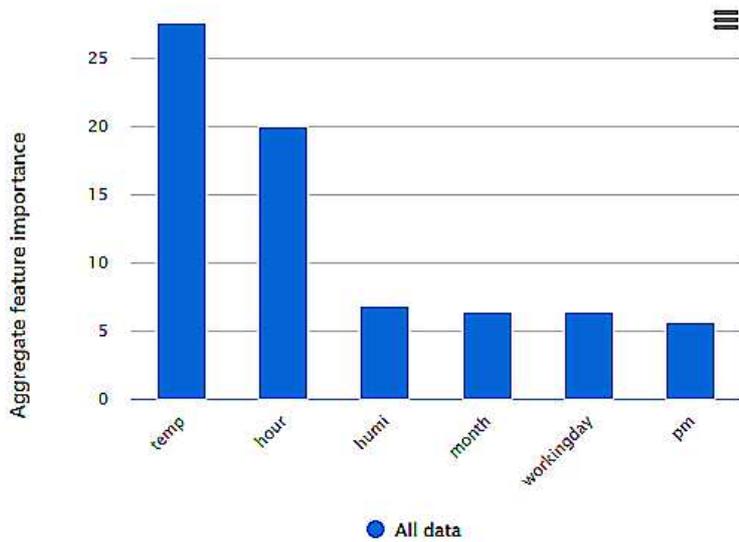


그림 48 Top6 features by their importance

적용된 머신러닝 중에서, 가장 높은  $R^2$  값을 갖는, 앙상블 알고리즘을 분석해보면, 타겟 데이터를 보행량으로 두고, 적용된 데이터에서 상위 6개의 요소를 출력하면, 온도(temp), 시간(hour), 습도(humi), 월(month), 평일(working day), 미세먼지(pm)의 순으로 영향을 주는 것을 알 수 있다.

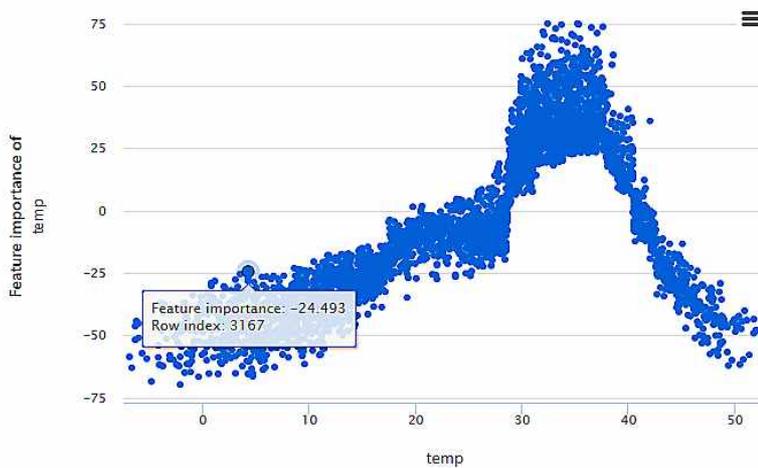


그림 49 Voting Ensemble 머신러닝으로 적용된 온도와 방문자와의 상관관계

상위 6개의 독립변수의 값과 방문자와의 상관관계를 그래프로 살펴보면, 온도의 경우에는, 약 30~40℃ 기준에서 가장 많은 보행량이 있는 것을 알 수 있다. 이제 실제 상온의 온도와 측정 장비와의 잔차인 8~10℃ 를 고려하였을 때, 쾌적한 기온인 20~30℃ 구간임을 데이터 분석으로 알 수 있다.

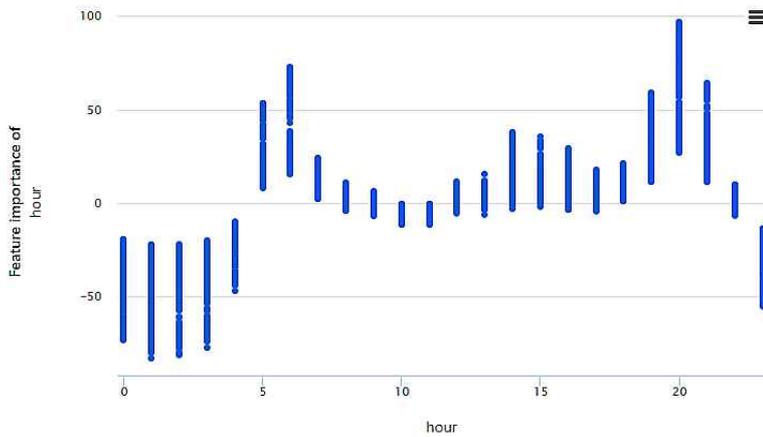


그림 50 Voting Ensemble 머신러닝으로 적용된 시간대와 방문자와의 상관관계

그림 50에서 시간대의 경우에는, 오전 5~7시, 그리고, 밤 8시 내외에서 가장 많은 보행량이 있는 것을 알 수 있다. Training 데이터를 훈련하는 가중치에서 가장 많은 영향력이 있는 것을 알 수 있다.

그림 51에서 습도의 경우에는, 기존 서울시 습도와 비교하였을 때, 잔차가 평균 30% 정도가 낮게 측정되는 것을 고려하였을 때, 습도가 높은 80% 이상인 우기에는 보행량에 부정적인 영향을 주는 것을 알 수 있고, 쾌적한 날씨에서는 매우 긍정적인 효과를 주는 것을 확인할 수 있다.

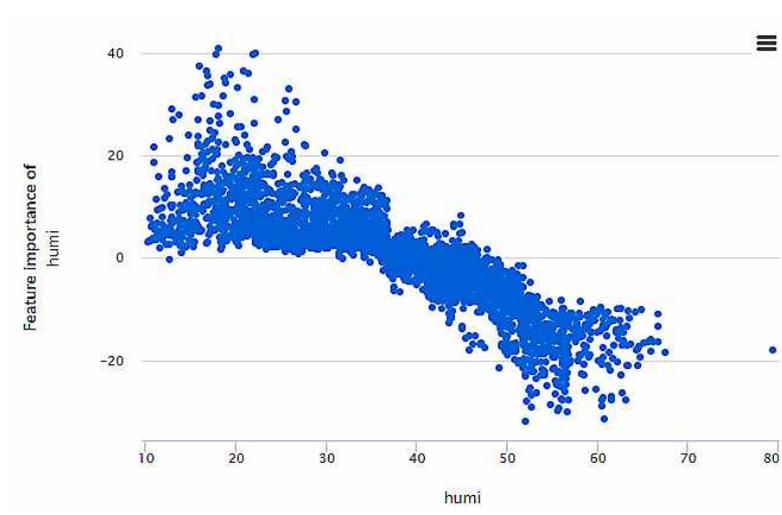


그림 51 Voting Ensemble 머신러닝으로 적용된 습도와 방문자와의 상관관계

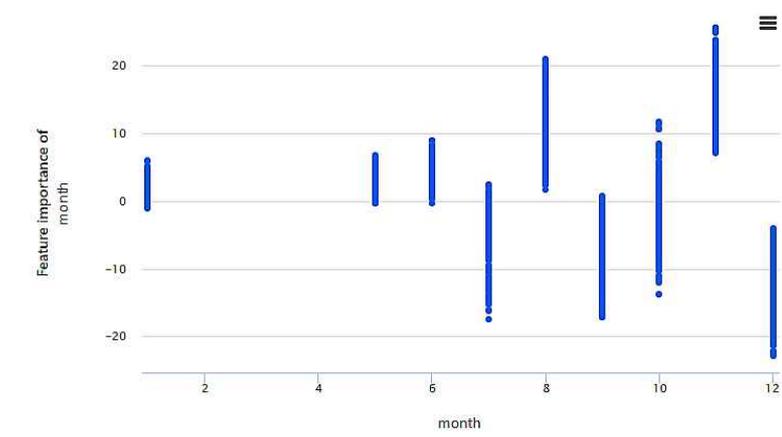


그림 52 Voting Ensemble 머신러닝으로 적용된 월간 대와 방문자와의 상관관계

그림 52에서 월간 특성의 경우에는, 봄철인 5월과 가을철인 10월 내외에 많은 보행량이 있는 것을 알 수 있다. 폭염기인 7월과 겨울인 12월에는 보행량에 부정적인 영향을 주는 것을 확인할 수 있다.

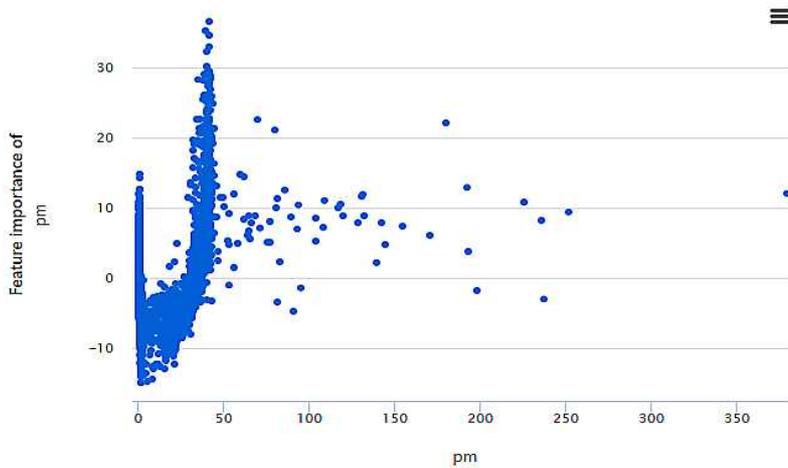


그림 53 Voting Ensemble 머신러닝으로 적용된 미세먼지와 방문자와의 상관관계

그림 53에서 미세먼지의 경우에는, 봄철에 주로 발생하는 경향으로 인해서, 보행량에 부정적인 요소임에도 불구하고 수치가 50 이상으로 증가하여도, 보행량이 증가하는 것으로 해석되었다. 따라서 머신러닝을 통한 환경 데이터 해석의 경우, 유의하지 않으면, 실수로, 부정적인 요소도 긍정적인 요소로 받아들일 수 있어서, 주의해야 하는 현상을 보여준다.

### 2.2.3 하이퍼 파라미터

하이퍼 파라미터는 앞에서 언급했듯이 학습을 시키는 컴퓨터의 하드웨어의 성능과 밀접한 연관이 있으므로 컴퓨터 환경에 맞게 설정해주어야 한다<sup>153)</sup>. 본 연구에서 사용된 머신러닝(ML)에 적용되었던 Ensemble 모델, XG Boost regressor, LGBM, Randomforest 알고리즘에서, 각각 데이터 셋을 학습할 때, 하이퍼 파라미터를 자동으로 튜닝하고 적용하여 사용된 최적의 값을 정리하면 다음 표 30과 같다.

153) 구형일(2018), 「인공지능 및 딥러닝 동향」, 『전기의세계』, 67(7), 7-12.

표 31 각 학습모델당 적용된 하이퍼 파라미터

Algorithm Name	pre processing	Pre processing	Algorithms	module	Hyperparameter values	
Voting Ensemble	Standard Scaler	sklearn	XGBoostRegressor	automl	eta	0.5
					gamma	0.01
					max_depth	8
					max_leaves	0
					num_leaves	0
					n_estimators	50
					reg_alpha	1.354166667
	reg_lambda	1.666666667				
	StandardScaler	sklearn	XGBoostRegressor	automl	eta	0.2
					gamma	0
					max_depth	8
					max_leaves	0
					num_leaves	400
					n_estimators	0
					reg_alpha	0.104166667
	reg_lambda	1.666666667				
	MaxAbsScaler	sklearn	ExtraTreesRegressor	sklearn	max_features	0.8
					min_samples_leaf	0.002364682
					min_samples_split	0.005285389
					n_estimators	10
	MaxAbsScaler	sklearn	XGBoostRegressor	automl	tree_method	auto
MaxAbsScaler	sklearn	LightGBMRegressor	automl	min_data_in_leaf	20	
XGBoostRegressor	StandardScaler	sklearn	XGBoostRegressor	automl	tree_method	auto
LightGBM	MaxAbsScaler	sklearn	LightGBM	sklearn	min_data_in_leaf	20

	r	arn			n_leaf	
RandomForestRegressor	StandardScaler	sklearn	RandomForestRegressor	sklearn	max_features	0.3
					min_samples_leaf	0.009017208
					min_samples_split	0.052853886
					n_estimators	25

## 2.2.4 모델 평가

인공지능 모델 평가에 사용될 지표는 실제 값과 예측값의 차이를 절댓값으로 변환 뒤 합산하여 평균으로 구하는 MAE, 평균제곱오차에 제곱근을 씌워 구하는 RMSE, 실제값과 예측값의 차이를 제곱한 값의 합(Sum of Squared Residual, SSR)을 실제값과 평균의 차이를 제곱한 것의 합(Sum of Squared Total, SST)으로 나눈 값을 말하며 통상적으로 0에 근접할수록 쓸모없는 회귀식으로 1에 가까울수록 모델이 우수한 성능을 보인다고 평가하는  $R^2$  을 사용하였다.

4가지의 학습모델 중 Voting Ensemble의  $R^2$  0.72로 가장 우수한 것을 알 수 있으며, 그 뒤로, XGBoostRegressor이 0.70, LightGBM이 0.69, RandomForestRegressor가 0.53의 순으로 결과값을 알 수 있으며, MAE와 RMSE도 같은 순서로 표 32 에 기술하여, Voting Ensemble 모델이 여러 학습 모형 중에 가장 정확도가 높은 것을 알 수 있었다.

표 32 사용된 학습모델에 대한 평가

Algorithm Name	MAE	RMSE	$R^2$
Voting Ensemble	23.65045	36.29419	0.7222403
XGBoostRegressor	24.79629	37.58314	0.7022615
LightGBM	25.05527	37.87813	0.6975094
RandomForestRegressor	33.88134	46.7844	0.5388957

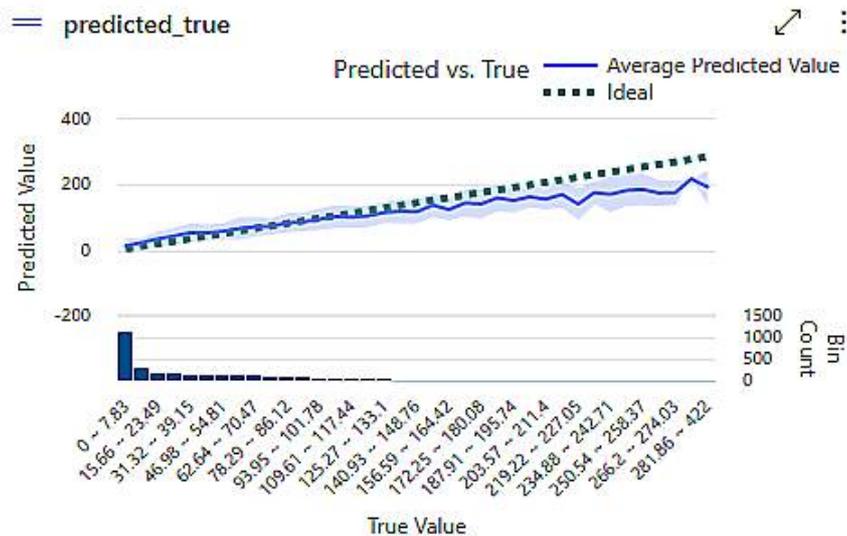


그림 54 예측값과 실제값의 비교 평가 결과

또한 예측값과 실제 값을 비교해보았을 때의 그래프와, 잔차는 그림 55와 같이 학습이 잘 적용된 것을 알 수 있다. 시각화 측면에서 Auto ML이 출력하는 가장 좋은 차트는 그림 56 Predicted vs. True 차트이다. 이 차트는 주석에 표시된 대로 x축에 실제 값을 표시하고 y축에 예측값을 표시한다<sup>154)</sup>.

녹색 점선은  $x = y$ 일 때 이상을 나타내기 위해 표시된다. 파란색 선이 녹색 점선에 가까울수록  $R^2$ 값이 높고 모델이 더 좋은 것이다. 이 시각화의 장점은 사용자가 예측의 정확성과 부정확성과 예측 위치를 빠르게 볼 수 있다. 예를 들어, 실제 값이 더 높을 때 예측이 덜 정확해진다. 파란색 선 주변의 연한 파란색 음영은 오차 한계를 나타낸다. 상한과 하한 사이의 범위가 넓을수록 예측의 신뢰도가 떨어진다. 본 ML 학습에서는 예측값과 실제값이 이상 곡선과 잘 매치되는 것을 확인할 수 있다.

잔차(Residuals)는 예측값과 실제값의 차이로 실제값 0에서 예측값을 뺀 값이다. 이는 잔차가 양수 또는 음수 값이 될 수 있음을 의미한다. 양수 값은 예측이

154) “Evaluating Models in Azure Machine Learning (Part 2: Regression)”. <https://adatis.co.uk/evaluating-models-in-azure-machine-learning-part-2-regression> (검색일: 2022. 10. 15.).

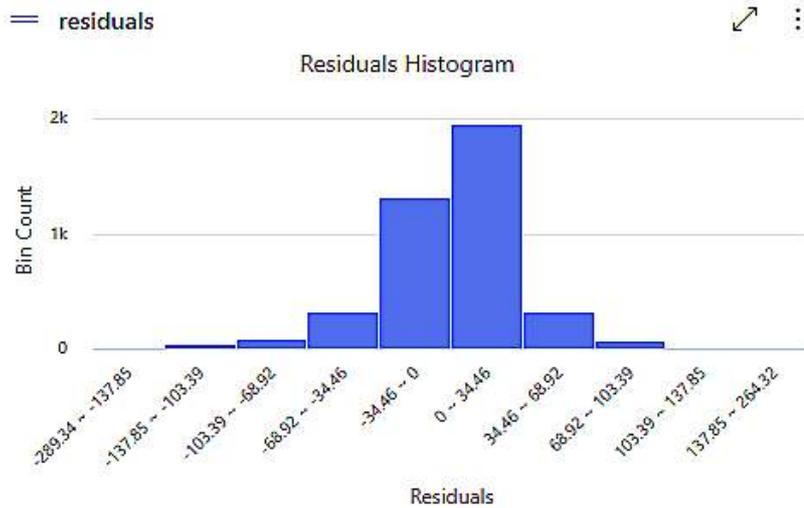


그림 55 알고리즘 잔차 검토 결과

너무 낮음을 의미한다. 음수 값은 예측이 너무 높다는 것을 의미한다. 회귀와 관련하여 예측이 실제와 같을 것으로 기대하는 것은 비현실적이므로 대신 예측이 실제에 최대한 근접하도록 목표를 세워야 한다. Azure AutoML은 잔차 분포를 표시하는 히스토그램을 출력한다. 이상적으로는 잔차가 가능한 0에 가까워야 하므로 잔차는 0을 중심으로 하는 종형 곡선으로 정규 분포되어야 한다. 히스토그램이 음으로 치우쳤으면 이는 예측 값이 너무 높다는 편향이 있음을 나타냅니다. 마찬가지로 히스토그램이 양으로 치우쳤으면 너무 낮은 값을 예측하는 편향이 있다. 본 머신러닝에서는 0을 중심으로 잔차가 적당한 종의 형태로 분포되어있음을 확인할 수 있으므로, 예측값과 실측값이 균형이 있게 배치된 것을 알 수 있다.

## 2.3 소결

이번 연구에서는 서울숲공원의 환경 데이터가 공원 사용 빈도의 근거인 보행량의 상관관계를 분석하기 위해, Auto ML 머신러닝 모델을 이용하였다. 특히, 광범위한 시, 군, 구 단위의 매크로 기상 데이터가 아닌, 서울숲 안의 마이크로 환경 데이터인 온도, 습도, 일산화탄소, 미세먼지 등을 직접 개발한 IoT 장비와 그곳에 보행량을 측정하는 IoT 장비를 같이 설치하여, 데이터를 장기간에 걸쳐 수집하였다. 수집된 클라우드 데이터는 전처리를 거쳐, 공공데이터와 합성을 하여 유의미한 데이터로 구성하였다. 일반적인 데이터와 범주 데이터, 시계열 데이터와의 분석은 파이선과 구글 코랩을 사용하여, 각종 그래프로 도식화하여, 현상을 쉽게 살펴볼 수 있도록 하였으며, 다시 인공지능분석을 통한 머신러닝 알고리즘별 예측 성능을 비교하고 그중에 가장 우수한 성능을 보인 Voting Ensemble을 통해서 분석된 보행량과 상관관계가 높은 환경 요소와 그의 패턴을 살펴보았다.

그 결과, 온도(temperature), 시간(hour), 습도(humidity), 월(month), 평일(working day), 미세먼지(pm)의 6개 환경 데이터 인자의 순서로 특징중요도(feature importance)가 보행량 예측에 가장 큰 영향을 주는 것으로 측정되었다. 적용된 Voting Ensemble의 결정계수인  $R^2$ 이 0.72로 가장 우수한 것을 알 수 있으며, XGBoostRegressor가 0.70, LightGBM이 0.69, RandomForestRegressor가 0.53 으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘에 대한 결론을 내릴 수 있었다.

## IV. 소래생태공원을 대상으로 한 스마트공원 구현

### 1. 소래생태공원: 인공지능을 이용한 기후변화 대응

인천광역시 남동구에 있는 소래 생태습지공원은 바다가 접하는 지역을 사이에 두고 형성된 염생 습지 지역으로서, 서울과 인천과 같은 대도시 생활권과 근접해 있다. 이 지역은 1933년 소금생산을 목적으로 염전을 만든 이후 1997년 이르기 까지 직접 소금을 생산해오던 곳이었다. 그러나 1990년대 들어 염전의 경쟁력을 상실한 이곳은 버려진 땅으로 사람들의 기억 속에서 잊혀 가는 곳이었다<sup>155)</sup>. 그런던중 버려진 폐염전 부지와 갯골을 중심으로 염생 습지들이 대규모로 형성되었고, 전국에서 가장 많은 염생식물이 자생하고 있고 멸종위기 조류와 도요물떼새가 도래하며 장아산과 연결되어 있어 육상식물과 해양식물이 어우러져 있다. 이와 같은 연안습지는 다음과 같은 기능과 가치가 있다.

- ▶ 수산자원의 보고
- ▶ 생물자원의 보고
- ▶ 오염물의 정화
- ▶ 바다와 육지의 완충
- ▶ 기후조절 기능
- ▶ 심미적 기능
- ▶ 경제적 가치

이중에 기후조절 기능에 대하여, 해양환경공단·한국해양과학기술원·서울대·부산대 등 10개 기관의 ‘블루카본’ 연구 결과를 보면 염습지의 탄소흡수량이 갯벌보다 최소 1.7배에서 최대 4.7배 많은 것으로 조사됐다<sup>156)</sup>. 염습지의 단위 면적(km<sup>2</sup>)당 연간 탄소흡수량은 염생식물이 흡수한 생체축적량 2t, 갯벌 속

---

155) 안경수(2001), 「수도권 해양 생태공원 조성사례-생태계 현황조사에 의한 습지 조성 계획 중심으로」, 『물과 미래: 한국수자원학회지』, 34(5), 40-67.

156) 안광호, 「탄소 흡수능력 뛰어난 염습지 ‘조립 경제효과’ 일반 땅의 5배」, 『경향신문』 (2021. 04. 20.)

으로 흡수된 토양격리량 91t 등 총 93t이다. 이에 반해 비 식생 깃벌의 연간 탄소흡수량은 20t에서 최대 54t (이산화탄소 환산 시 최대 198.0t)으로 상대적으로 낮았다. 따라서, 인천의 염습지와 갈대숲 그리고 도시의 공기질 데이터를 비교한다면, 염습지의 이산화탄소, 미세먼지 등의 감소 효과를 측정할 수 있을 것으로 기대한다.

본 연구는 소래생태 습지의 염습지와 갈대숲의 주변 1km 근접해있는 도심 도로 대비 이산화탄소, 온도, 미세먼지, 초미세먼지 저감 및 폭염 방재 효과를 측정하여 분석한다. 측정 기간은 AWS를 활용하여, 7월9일 ~8월 17일까지, AWS IoT 장비를 표 33과 같이, 염습지, 갈대숲, 그리고 포구로 사거리에 설치하였다.

표 33 소래생태 습지 공원 대상지 AWS IoT 위치(좌) 및 설치현황(우)



표 34 소래 습지 생태공원 AWS IoT장비 모니터링 수집내용

IoT 위치	데이터명	수집내용	수집 기간	수집목적
소래포구 사거리	sr_pg4_temp	온도/습도/이산화탄소/미세먼지/초미세먼지/기압	22/07/09 ~ 8/17	혹서기 염습지 대비 도로 부의 공기질 데이터를 이용한 기상 데이터 저감 효과 측정
	sr_pg4_humi			
	sr_pg4_co2			
	sr_pg4_pm2.5			
	sr_pg4_pm10			
sr_pg4_pressure				
염습지	sr_salty_temp	온도/습도/이산화탄소/미세먼지/초미세먼지/태양광조도/자외선/	22/07/09 ~ 8/17	혹서기 염습지 공기질 데이터를
	sr_salty_humi			

	sr_salty_co2	우적량/풍속/돌풍/풍향/기압		이용한 기상 데이터 저감 효과 측정
	sr_salty_pm2.5			
	sr_salty_pm10			
	sr_salty_solar			
	sr_salty_UVI			
	sr_salty_rainfall			
	sr_salty_Wind_Speed			
	sr_salty_Wind_Gust			
	sr_salty_Wind_Direction			
	sr_salty_pressure			
갈대숲	sr_galdae_temp	온도/습도/이산화탄소/미세먼지/ 초미세먼지/태양광조도/자외선/ 우적량/풍속/돌풍/풍향/기압	22/07/09 ~ 8/17	혹서기 염습지 대비 갈대숲의 공기질 데이터를 이용한 기상 데이터 저감 효과 측정
	sr_galdae_humi			
	sr_galdae_co2			
	sr_galdae_pm2.5			
	sr_galdae_pm10			
	sr_galdae_solar			
	sr_galdae_UVI			
	sr_galdae_Rainfall			
	sr_galdae_Wind_Speed			
	sr_galdae_Wind_Gust			
	sr_galdae_Wind_Direction			
	sr_galdae_pressure			

## 1.1 AWS를 통한 스마트공원 환경 데이터 수집

앞 절인 서울숲 연구사례에서 설명한 자체 개발한 IoT 장비와 별도로, 수집된 이후에 공신력이 있는 장비를 데이터수집에 대한 검증을 다시 하기 위하여 자동 기상관측시스템인 상용 AWS 환경을 스마트공원에 맞추어 재설계하여, 데이터 수집 장비를 사용하도록 한다. 환경 데이터 정량화를 위하여 서울숲에 설치하여 온도, 습도, 일산화탄소, 미세먼지 등을 측정하였던 IoT 장비의 장단점을 배경으로 신규 장비를 선정하였다. 기상과 공기질 데이터 중에서 온도, 습도, 이산화탄소, 미세먼지, 초미세먼지, 기압, 우적량, 풍속, 풍향, 돌풍, 태양조도, 자외선 등의 다양한 데이터를 획득하고자 한다.

### 1.1.1 자동기상관측소(AWS) 정의

자동 기상관측장비(Automated Weather Stations) 날씨와 기후를 측정, 모니터링 및 연구하는 데 사용되는 구성 요소의 통합 시스템을 제공한다. 이러한 관측소는 일반적으로 날씨, 기상학 응용 분야에 사용된다.

#### (1) 자동기상관측소의 요소들

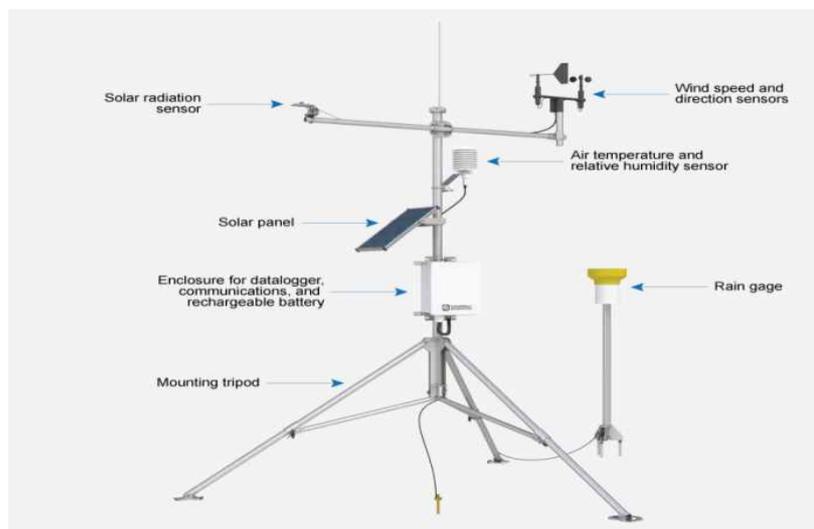


그림 56 AWS 사례 ; 출처 : Campbell Scientific site

AWS(자동 기상 관측소)는 함께 작동하여 기상 매개변수를 측정, 저장 및 전달하는 다양한 구성 요소로 구성된다. 자동화된 기상 관측소는 설계가 크게 다를 수 있다. 예를 들어, 제한된 개인 용도로 설계된 단순한 자동화 기상 관측소는 독립형 장치일 수 있다. 반면에 보다 정교한 연구 수준의 자동화 기상 관측소에는 필요에 맞는 맞춤형 솔루션을 만들기 위해 개별적으로 선택하는 여러 구성 요소가 있다. 맞춤형 자동화 기상 관측소의 구성 요소에는 다음과 같다.

- ▶ 센서들
- ▶ 데이터 로거
- ▶ 통신장비 (데이터수집 및 재생)
- ▶ 소프트웨어
- ▶ 전력 공급장치
- ▶ 장비 고정 및 설치 대

### 1.1.2. 스마트공원을 위한 환경 데이터 수집 장비 구현

#### 1) 원격 모니터링 기능을 갖는 자동기상관측소(AWS) 설계 방향

본 연구에 추가 투입되어 활용한 AWS(자동 기상 관측소)는 AWS 기능과 함께 태양광 기반 IP Camera가 내장되어 원격모니터링 장비(Remote Monitoring System)를 내장한 것이 특징이다. 또한, 인터넷이 접속되지 않는 외부상황을 고려하여 LTE 모듈과 함께 대형 배터리를 함께 함체 내에 탑재할 수 있게 설치되어, 일반적으로 장기간에 걸쳐 진행하지 못했던 환경 데이터 수집을 24시간 매 5분 단위로 약 7일에 걸쳐서 외부 기상이나 기온과 관계없이 진행할 수 있는 장점이 있다. 외부환경에 따라서, 표 35와 같이 함체를 자체 스탠드에 탑재(mount)하거나, 별도 외부에 있는 전주, 가로등에도 쉽게 부착할 수 있도록 설계되어, 다양한 도시 환경과 자연환경에서 다양한 환경 자료를 수집하는 것에 최적의 상태를 유지할 수 있다.

AWS에서 측정된 데이터와 촬영된 영상은 각각 제공되는 클라우드 서버에 저장되어, 원격으로 접근할 수 있으며, 영상은 한 달 이내에 로컬 드라이브에 저장할

수 있으며, 환경 데이터는 1년 단위, 월 단위, 주 단위, 일 단위 등으로 각각 로컬 드라이브에 저장할 수 있어 데이터가 통신력이 매우 우수하고 사용이 용이하다.

추가 센서 모듈을 탑재하여, 온도, 습도, 조도 와 같은 외부의 기본 환경 데이터 이외에도, 이산화탄소, 미세먼지, PM10, PM2.5, 함체 내 온도, 습도 등을 같이 측정할 수 있어서 다양한 환경 데이터의 수집이 가능하다. 수집된 데이터는 정형화되어 5분 단위로, 클라우드 서버에 저장되어, 시계열 데이터로 손쉽게 활용됨으로써 인공지능 Machine Learning (머신러닝) 빅데이터로서의 가치가 매우 높다.

표 35 설치된 스마트공원 적용 AWS 사례

	
<p>스마트공원용으로 적용 후, 실제 현장에 적용된 스마트공원용 AWS 사례,</p>	<p>국립재난 안전연구원 사이트 소하천변 환경 자료수집목적으로 설치 중인 모습 (경북 양산시)</p>

출처 : 에어플 탄소 환경 연구소, 국립재난 안전연구원 (2022년 10월)

표 36 소래생태공원 연구에 사용된 기상관측장비 (우측)과 기존 개발된 원격모니터링 AWS 사례들(좌측)

	
<p>ULPD Lab 에서 개발된 기상관측장비(AWS) 들</p>	<p>본 연구에 사용된 기상관측장비 AWS</p>

2) 관측용 장비 사양서

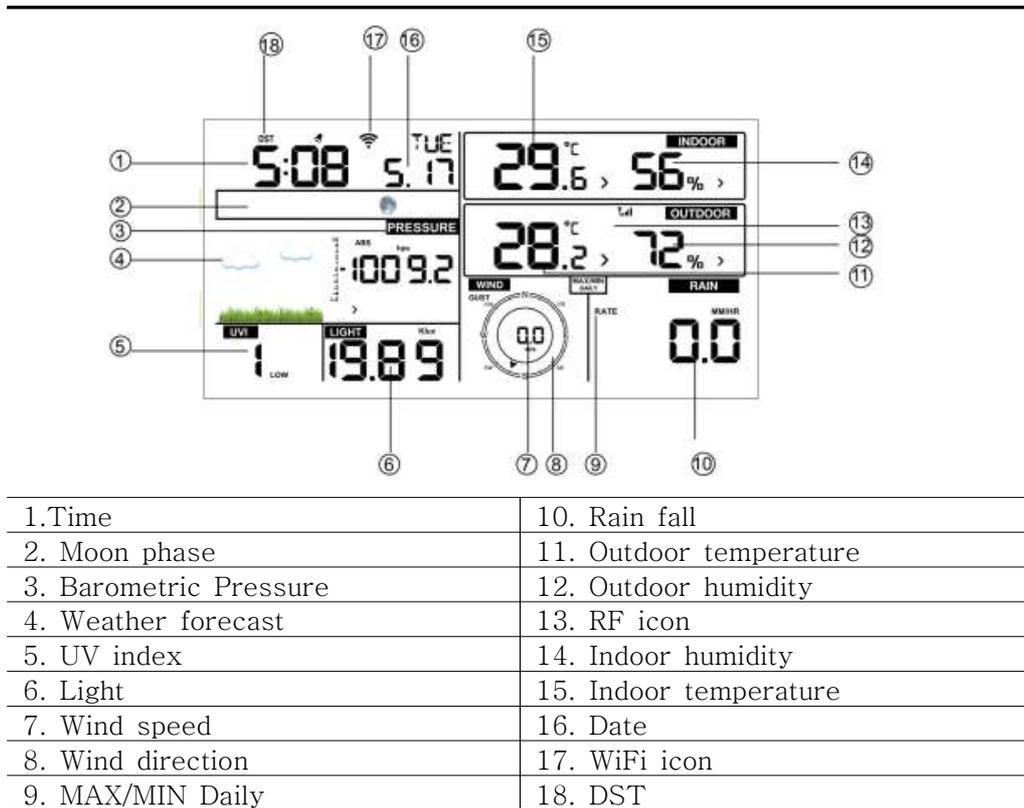
(1) Automatic Weather Station

표 37 Automatic Weather Station 외관(우측)과 실내용 LCD 디스플레이(좌측)



제공 : 에어플 탄소 환경 연구소

표 38 AWS용 스크린 디스플레이와 표시내용



### 3) Weather Station 수집 데이터

Weather Station에 수집하는 데이터 사양은 다음과 같다.

표 39 AWS 사양서

외부 환경 데이터 (Outdoor data)	실내환경데이터 (Indoor data)	전원 사양 (Power consumption)	미세먼지, 초미세먼지, 이산화탄소환경센서
Transmission distance in open field : 100m (300 feet) Frequency : 433MHz Temperature range : -40° C to 60° C (-40° F to +140° F) Accuracy : ±1° C Resolution : 0.1° C Measuring range rel. humidity : 10% to 99% Accuracy : ±5% Rain volume display : 0mm to 6000mm (show --- if outside range) Accuracy : ±10% Resolution : 0.1mm (if rain volume < 1000mm) 1mm (if rain volume > 1000mm) Wind speed : 0m/s to 50m/s (0mph to 100mph) (show --- if outside range) Accuracy : ±1m/s (wind speed < 5m/s) +/-10% (wind speed > 5m/s) Light : 0k Lux to 200k Lux Accuracy : ±15% Measuring interval thermo-hygro sensor : 16 s	Indoor temperature range : 0° C to 50° C (32° F to +122° F) (show --- if outside range) Resolution : 0.1° C Measuring range rel. Humidity : 10% to 99% Resolution : 1% Measuring range air pressure : 700hPa to 1100hPa (20.67inHg to 32.5inHg) Accuracy : ±3hpa Resolution : 0.1hPa (0.01inHg) Alarm duration : 120 sec Measuring interval indoor data : 60s	Base station : 5V DC adaptor (included), Power Consumption configuration mode) : 0.5 Watts (1.25 Watts during WiFi) Base station : 3 × AAA batteries (not included) Remote sensor : 2 × AA batteries (not included) The primary power source is the solar panel. The batteries provide backup power when there is limited solar energy	측정 실내 PM2.5, PM10, CO <sub>2</sub> 농도, 온도 및 습도 매 60 초. 실내 사용 전용. 장치에 유효한 이산화탄소 가치 구경측정 기능. 무선 범위, 열린 지역에서 300 피트 (100 미터) 까지  명세: 무선 전송 거리 (공중에서): 최대. 100 미터, (장애물이 있는 경우에, 또는 신호 방해는, 거리 단축될 것입니다. 주파수: 433 MHz 감지기보고 간격: 1 분  전력 소비: 입자 감지 센서: AC 전원 (5V 1A) -포함되지 않음 백업을위한 2 x AA 1.5V 배터리-포함되지 않음

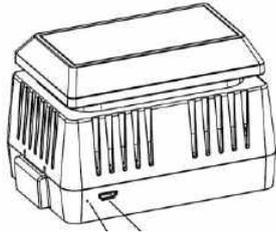


그림 57 미세먼지, 이산화탄소 측정용 센서  
(제공 : 에어플 탄소 환경 연구소)

#### 4) 전원, 무선 와이파이 무공급 지역 설비

지자체나 연구소에서 야외지역에서 대기환경 또는, 공기질에 관련 실험을 하고자 하면, 클라우드 서버에 접속하기 위한 와이파이(WiFi)와 데이터 접속용 모뎀과 센서 장비 등에 전원을 공급하기 위해서는 전원이 필요하다. 기존 방법으로는 전원선을 별도로 공사하여 유입하거나, 태양광 소자, 또는 발전기 등을 활용하여 실험하였다. 그러나, 태양광의 경우에는 우천 시나, 구름이 많아서 제대로 전원 공급을 하지 못하는 것을 많은 실험을 통해서 알 수 있었다. 또한, 발전기를 사용할 때는, 석유등의 연소로 인한 공기질의 저하와 소음 및 연료의 불충분으로 장기간동안 측정이 불가능하였다. 따라서, 소래 습지 생태공원 연구에서 사용한 방법은 대용량 충전 배터리를 장착하여 사용하게 되면, 무선 모뎀, 각종 센서류를 구동하면서, 약 1주일간의 사용이 무난하였다.

전원의 사용량의 경우 기존 6~7일 유지되던 것을 클라우드 서버와의 통신 간격을 5분에서 1분 단위 주기로 하는 경우에는, 약 5~6일 정도로 하루 정도의 사용 시간이 줄어들게 된다. 사용된 충전 배터리는 60,000mAh 용량의 배터리로, 본 사항은 앞서 언급한 지자체나, 연구소에서 연구목적으로 간이용도로 사용할 경우에 매우 유익할 것으로 사료된다. 표 40 과 같이 하이박스 안에 기기들을 설치하면, 장기간의 사용 시에 우기, 건기, 구름의 적운량 등에 상관없이 지속적으로 데이터를 습득할 수 있다.

표 40 전원, WiFi없는 지역에 설치 방법

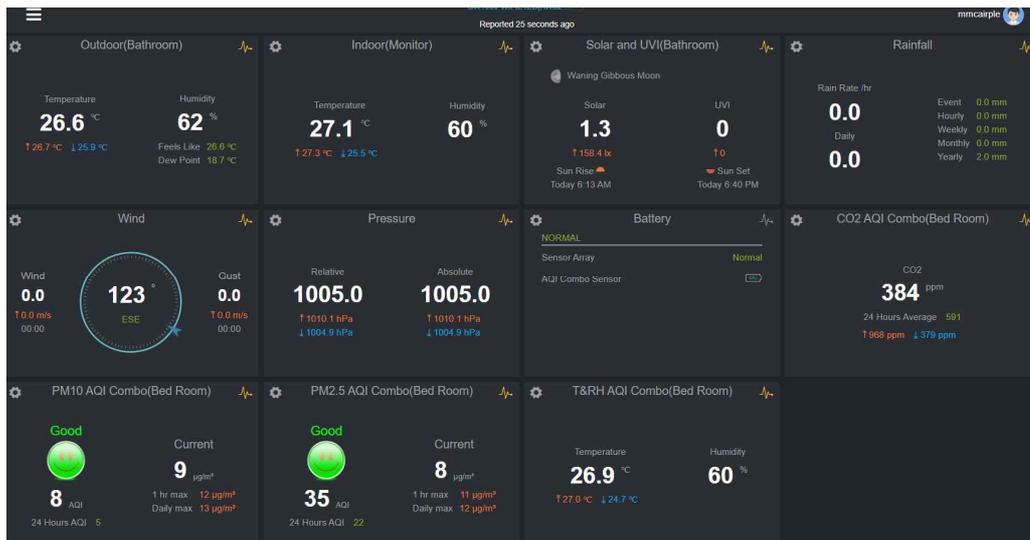
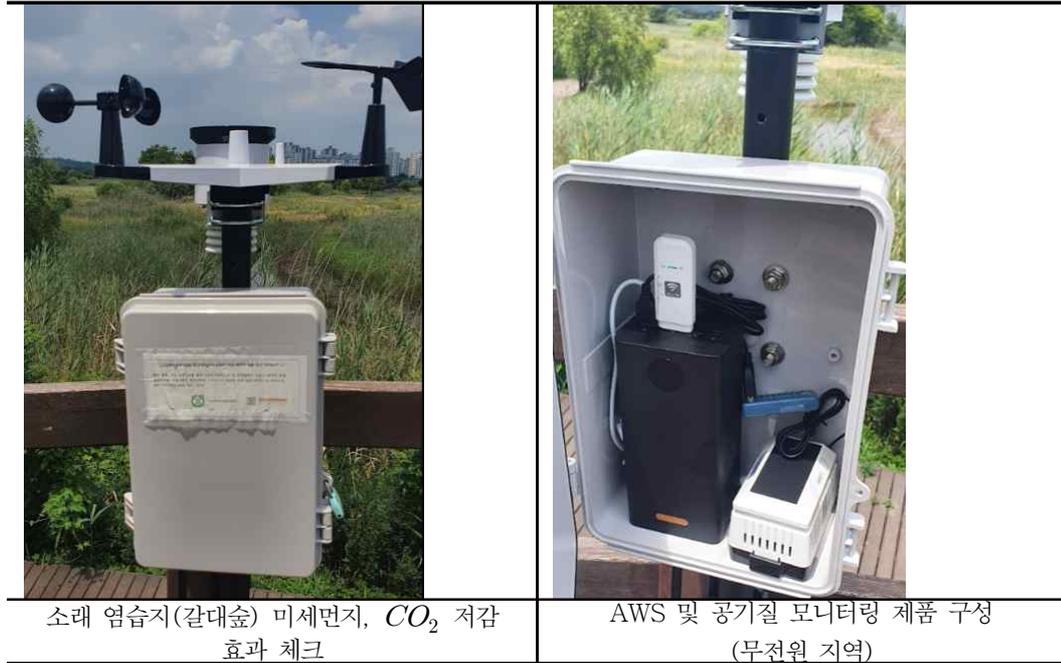


그림 58 AWS 장비로 수집된 공기질 web display 예 (제공 : 에어플 탄소 환경 연구소)

## 1.2 인공지능을 통한 탄소 발생 요소 분석 방안

본 연구에서 인공지능 모델링은 데이터 전처리 과정을 거쳐 Auto ML을 적용하고, 앙상블 알고리즘을 사용하여, 최적화된 Regression 회귀 관계를 파악하고, 이를 실험하여 결과값을 도출한다.

### 1.2.1 인공지능과 이산화탄소 발생량 기계학습 예측 사례

최근 이산화탄소(CO<sub>2</sub>) 등 온실가스로 인한 기후위기가 심화되고 있다. 실제로 글로벌 기후변화에 관한 정부 간 협의체(IPCC)에 따르면 최근 온실가스 배출량을 유지할 경우, 2040년 안에 지구 기온 상승폭이 1.5°C를 넘을 것으로 예상된다. 이 경우 극단적 폭염 발생 빈도는 과거 산업화 이전보다 8.6배 증가하게 된다<sup>157)</sup>.

이 문제를 해결하기 위해 한국을 포함한 12개 국가는 '탄소 중립'을 선언하며 저탄소를 넘어 온실가스 배출량을 제로화하겠다고 발표했다. 탄소 배출 제로화에서 중요한 요소 중 하나는 이산화탄소 산출량 분석이다. 현재 이산화탄소가 얼마나 배출되고 있고 어느 정도 감축을 했는지를 수치화해야 객관적인 분석이 가능하기 때문이다. 국내기업 SIA는 인공위성 영상분석을 통해 탄소 방출량을 정확하게 검출할 수 있는 딥러닝 알고리즘을 연구하고 있다<sup>158)</sup>.

기후변화를 예측, 효과적으로 대응하는 한편, 주된 원인으로 꼽히는 이산화탄소를 인공지능(AI) 기술로 통제하려는 노력이 각국 정부와 기업 차원에서 추진 중이다.<sup>159)</sup> 글로벌 컨설팅사인 캡제미니(Capgemini)에서는 AI를 활용해 지난 2년간 온실가스 배출이 12.9% 감소했고, 전력 효율성은 10.9% 향상됐다고 발표했다. 캡제미니는 2030년까지 AI를 통해 기존 시나리오 대비 4~17% 가량 온실가스 배출을 낮출 것으로 전망하고 있다<sup>160)</sup>.

---

157) 박설민, 「똑똑하게 줄여요... AI, 탄소중립 '게임체인저' 될까」, 『시사위크』 (2022. 04. 11.)

158) 김동원, 「에스아이에이, 이산화탄소 배출량 AI로 검출하는 기술 가능성 제기」, 『AI타임즈』 (2022. 03. 18.)

159) 최아름, 「인공지능도 친환경 시대...그린 AI '주목」, 『정보통신신문』 (2022. 04. 02.)

160) 박설민, 「똑똑하게 줄여요... AI, 탄소중립 '게임체인저' 될까」, 『시사위크』 (2022. 04. 11.)

유럽우주국(ESA)의 ‘AI4EO(Artificial Intelligence for Earth Observation)’ 이니셔티브는 최근 AI 기술을 위성을 통한 지구관측 분야에 적용해 전 지구적 당면 문제에 대한 해결책을 모색하기 위한 챌린지를 열었다. 이를 통해 미세먼지(PM2.5) 및 이산화질소(NO<sub>2</sub>)의 표면 농도에 대한 AI 기반 분석 및 예측모델 개발을 목표로 3개월간 진행됐다. 참가자들은 위성 데이터를 활용해 고해상도의 PM2.5 및 이산화질소의 표면 농도 시각화와 대기질 측정을 위한 AI 모델을 개발했다<sup>161)</sup>.

한화시스템이 개발한 ‘맹그로브’는 IoT 플랫폼을 기반으로 데이터를 수집하고 모니터링 시스템을 구축하는 동시에 클라우드, AI, 머신러닝 플랫폼을 기반으로 예측 모델을 구현해 실현하여, 기업의 직접 배출 탄소와 간접 배출 탄소 영역의 완전 자동화를 지원하고, 모든 가치사슬을 통해 배출되는 탄소 영역은 사용자가 등록해 관리할 수 있는 기능을 제공한다. 이를 통해 실시간으로 에너지 사용량 및 탄소배출량을 측정하고 데이터 분석으로 예측한 수치를 비교 분석해 효율적인 관리가 이루어진다.<sup>162)</sup>

이처럼, 전문가들 사이에서 AI가 탄소중립의 ‘게임체인저’가 될 것으로 기대 받는 이유는 뛰어난 ‘분석능력’ 때문이다. 많은 양의 데이터를 실시간으로 감지·분석할 수 있는 능력을 지닌 AI가 산업 현장이나 실생활에서 발생하는 이산화탄소의 양을 측정하고, 이를 기반으로 산업 현장에서는 탄소 배출량을 관리할 수 있다는 것이다.

실제로 글로벌 컨설팅기업 보스턴컨설팅그룹(BCG)은 ‘탄소와 비용을 줄이는 AI역량(2021)’ 리포트에서 “기업의 탄소 발자국의 여러 측면 및 단기 비용 절감 분야에 대해 심층적인 통찰력을 도출하는 AI의 능력은 지속 가능한 혁신의 가속화와 어려운 시기 속 비용 절감을 위한 희망적인 방안을 제시한다”고 분석했다<sup>163)</sup>.

홍태훈 외 4인(2018)은 딥러닝 알고리즘을 사용한 실내 이산화탄소 농도의

---

161) 최아름, 「인공지능도 친환경 시대…그린 AI ‘주목」, 『정보통신신문』(2022. 04. 02.)

162) “한화의 기술력과 AI가 관리하는 탄소배출관리 시스템, 맹그로브”, <https://www.hanwha.co.kr/media/discover/view.do?seq=5046> (검색일: 2022. 10. 15.)

163) 최아름, 「인공지능도 친환경 시대…그린 AI ‘주목」, 『정보통신신문』(2022. 04. 02.)

예측모델을 변수를 재실자 수, 환기량, 가전기기, 온도, 습도, 실내이산화탄소 농도 등으로 하고, LSTM(Long Short Term Memory) 예측모델을 활용하였다<sup>164)</sup>.

권재민 외 4인(2019)은 다중선형회귀 분석 기반의 건설장비 이산화탄소 배출량 예측모델을 제안하였는데, 환경적 요인변수는 기온, 온도, 토양을 선정하고, 기계적 요인 분석을 장비 타입, 모델, 버킷용량, 연식, 장비 중량, 친환경 엔진 유무 등으로 작업별 CO2 배출량 회귀모형을 도출하였다<sup>165)</sup>.

김효준 외 2인(2020)은 기계학습을 이용하여, 실내 이산화탄소 농도 예측모델을 검증하였다. 알고리즘은, 인공신경망(ANN), 서포트 벡터 머신(SVM), 랜덤 포레스트(Random Forest, RF), KNN(K Nearest Neighbor)를 사용하였다. 입력변수로는 공기조화기 공급 팬 스피드, 터미널 유닛 댐퍼 개도율, 공급 이산화탄소 농도와 관계가 있는 외기 댐퍼 개도율, 재실 인원수를 선정하였다. 각 알고리즘에 따른 결정계수  $R^2$  은 각각, ANN 0.97, SVM 0.96, RF 0.95, KNN 0.93 순이었다<sup>166)</sup>.

원동준 외 3인(2021)은 기계학습을 활용하여 경기도 산업단지의 실외 미세먼지를 예측하는 연구를 하였다. 입력변수는 초미세먼지(PM2.5), 미세먼지(PM2.5), 이산화황(SO2), 이산화질소(NO2), 일산화탄소(CO), 오존(O3), 온도, 습도, 풍향, 풍속, 강수량, 도로 구간별 차량 속도 변수 간의 상관관계분석 및 회귀분석을 통해, 변수의 유의성을 파악하고, 시간대별 PM2.5를 예측하였다. 사용된 알고리즘은 Random Forest, XGBoost, LightGBM, Deep neural network, Voting Ensemble 모델을 사용하였으며, RMSE는 각각 6.27, 6.41, 6.22, 6.64, 6.12로 에어코리아의 예측모델의 10.77에 비해 성능이 높게 평가되었다<sup>167)</sup>.

전수정 외 2인(2020)은 머신러닝을 활용하여, 서울지역 어린이집을 대상으로

---

164) 홍태훈 외(2018), 「딥러닝 알고리즘을 사용한 실내 이산화탄소 농도의 예측모델 개발」, 『한국에너지기후변화학회 학술대회』, 186.

165) 권재민, 이재학, 조민도, 최영준, 한승우(2019), 「다중선형회귀분석 기반 건설장비 이산화탄소 배출량 예측모델 개발」, 『한국건축시공학회 학술발표대회 논문집』, 38-39.

166) 김효준 외(2020), 「실내 이산화탄소 농도 예측을 위한 기계학습 모델 검증」, 『한국건축친환경설비학회 논문집』, 14(6), 699-706.

167) 원동준 외(2021), 「기계학습을 활용한 경기도 산업단지 미세먼지 예측」, 『정보과학회논문지』, 48(7), 764-773.

이산화탄소를 포함하는 실내공기질 예측을 하였다. 2019년 12월부터 2020년 6월까지 7개월간 30초 간격의 데이터를 클라우드에 입수하여, 외부, 급기, 배기 부에서, 온도, 습도, 미세먼지, 초미세먼지를 측정하고, 실내공기질은 Python을 사용하여 30분 단위 평균값을 산출하였다. 예측모델로는, SVM, Random Forest, GBM, XGBoost 의 모델링을 사용하여, 실내 이산화탄소 농도 등을 예측하였다.  $R^2$ 을 이산화탄소기준으로 각 모델링 별로 비교한 결과 SVM의 경우 : Training : 0.89, Testing : 0.75, RF 의 경우 Training : 0.96, Testing : 0.91, GBM의 경우 Training : 0.96, Testing : 0.91, XGB의 경우 Training : 0.99, Testing : 0.91로 예측모델 결과가 나왔다<sup>168)</sup>.

### 1.2.2 Auto ML을 사용한 전체 연구 flow

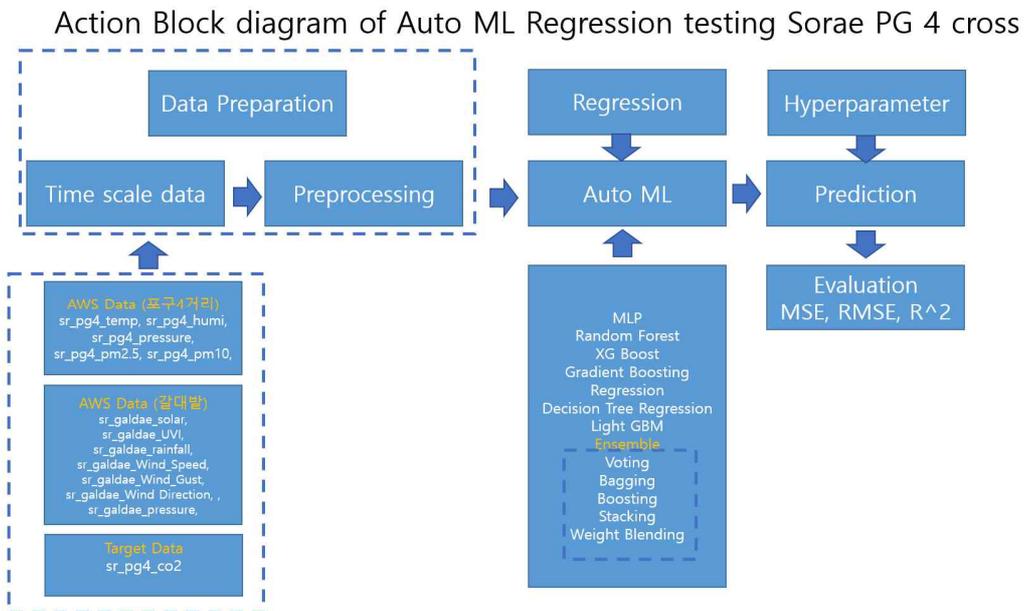


그림 59 소래포구사거리 Auto ML Regression수행을 위한 전체 플로우 차트

168) 전수정 외(2020), 「머신러닝 기법을 이용한 서울지역 어린이집 실내공기질 예측」, 『한국생태환경건축학회 학술발표대회 논문집』, 20(2), 112-113.

Action Block diagram of Auto ML Regression testing Sorae galdae area

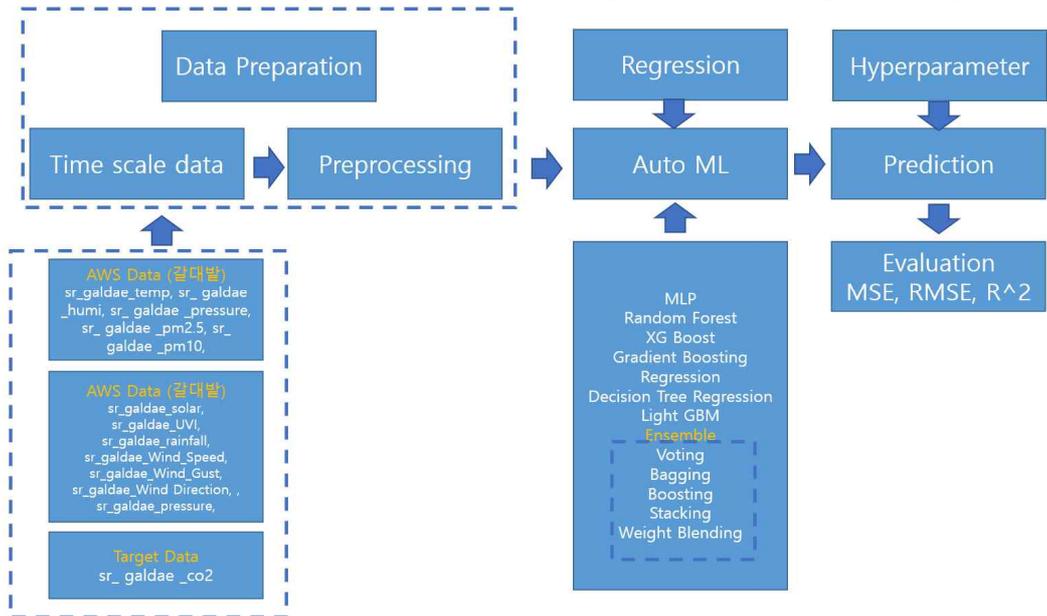


그림 60 소래 공원 갈대숲 Auto ML Regression수행을 위한 전체 플로우차트  
Action Block diagram of Auto ML Regression testing Sorae salty area

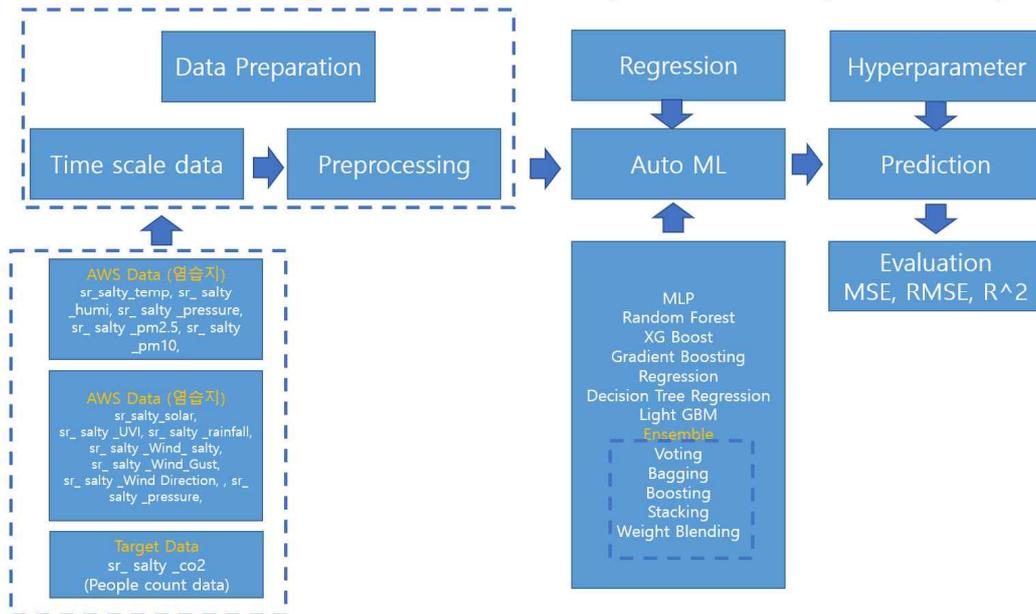


그림 61 소래 공원 염습지 Auto ML Regression수행을 위한 전체 플로우차트

인공지능을 통한 소래 공원의 탄소 발생 요소 분석은 그림 59, 60, 61에서와 같이 머신러닝을 위해서, 인천 소래생태공원의 갈대숲과 염습지 폭염기 환경데이

터와, 도심지인 포구 사거리에서 수집한 환경 데이터를 각각 개체 비교하여, 이산화탄소의 발생량에 영향을 주는 요소들과 사이트별 차이를 분석한다. 먼저 시계열 데이터로 수집된 데이터들을 데이터의 전처리 및 강화 후, 필터링된 기상 데이터와 계절, 날짜, 시간 같은 범주 데이터는 전체 워크플로를 보기 위해 'sorae\_echopark\_final.csv'를 로드하고 필요한 열을 선택하고 각 사이트별 ML을 진행하면서 Target을 'co2' 로 선택했다.

학습 데이터와 테스트 데이터를 분리하기 위해 분할 데이터의 비율을 7:3으로 설정하고 알고리즘<sup>169)</sup>에 따라 하이퍼 파라미터 변경과 분할된 테스트 데이터를 사용한 교차검증은, 이산화탄소 발생량 예측에 사용할 수 있는 최상의 머신러닝 모델을 선택하는 데 사용된다. Auto ML로 선정된 알고리즘 각각에 트레이닝 데이터를 사용하였고, 테스트 데이터는 위에서 나눈 분할 데이터를 기반으로 값을 넣어 학습시킨다. 이 연구의 모델은 Stacking Ensemble, Voting Ensemble, XG Boost Regressor, Random Forest, Gradient Boost Regression, Light GBM, Multi Layer Perceptron 등의 알고리즘을 사용하여 Azure Auto ML에서 구축한다.

---

169) Chung, D. H. et al.(2021). "Machine Learning for Predicting Entrepreneurial Innovativeness," *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 16(3), 73-86.

## 2. 인공지능 모델링을 통한 소래 생태 스마트공원 구현

### 2.1 인공지능을 이용한 습지와 CO<sub>2</sub> 상관관계

본 연구의 목적은, 소래 생태습지공원의 가지고 있는 이산화탄소 저감 효과를, 특히, 1년 중 폭염기인 7~8월에 측정하여 살펴봄으로써, 기후변화 방재 효과로서의 스마트공원의 역할을 살펴보고자 한다. 지구온난화와 이상기후는 전 세계적인 문제이며, 탄소 Net Zero를 이룩하기 위하여 이산화탄소를 줄이기 위한 많은 노력이 기울여지고 있는 상황에서, 공원의 이산화탄소 저감 효과는 매우 큰 의미가 있다고 볼 수 있다. 특히 인공지능을 활용하여, 수집된 환경 데이터를 이산화탄소 발생과 관련지어 Regression 모델링 분석하여, 데이터 수집에 이용된 사이트 별로 이산화탄소 발생 예측에 영향을 주는 요소들을 살펴봄으로써, 향후 탄소 저감에 조금이라도 도움이 되고자 한다.

#### 2.1.1 데이터 전처리

IoT 장비를 통하여 습득한 환경 데이터의 실효성을 파악하기 위해서, 인천시에서 측정한 공공데이터 중, 인천시 보건환경연구원에서 운영 중인 환경정보 공유시스템에서 소래생태공원과 가장 근접한 위치인 논현점의 PM2.5, PM10 정보를 이용하여, 포구 사거리에 설치한 센서에서 취득한 데이터와의 트렌드를 비교하여, 데이터의 유효성을 비교하였다. 그림 62에 논현 공공데이터 취득지점과의 위치 정보를 표시하였다. 공공데이터 특성상, 일일 평균 데이터를 기준으로, 취득한 기간인 7월 9일부터 8월 17일까지의 폭염기 동안의 매일 평균 데이터를 비교해 보면, 다소 차이가 있으나, 매우 리니어한 현상을 쉽게 발견할 수 있어서, 데이터가 유효한 것을 알 수 있다.



그림 62 소래 습지 생태공원에 설치된 AWS 센서의 위치와 인천 보건 환경원 논현지점 센서 위치 비교

(포구 사거리와 갈대 습지 1,000m, 갈대 습지와 염습지 사이 거리 500m)

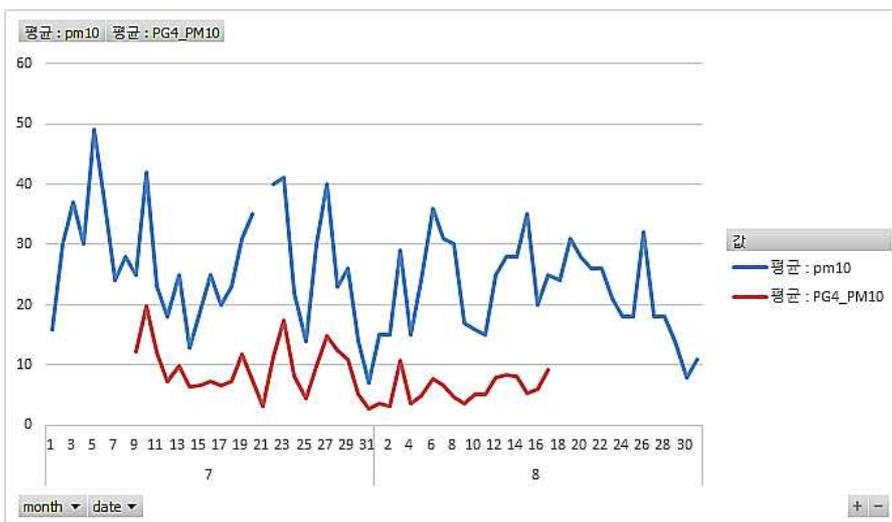


그림 63 인천 보건 환경원 논현지점의 PM10 데이터와 소래포구 사거리의 취득한 환경 데이터 비교

(파란색 : 논현지점. 빨간색 : 소래포구사거리 PM10 평균값)

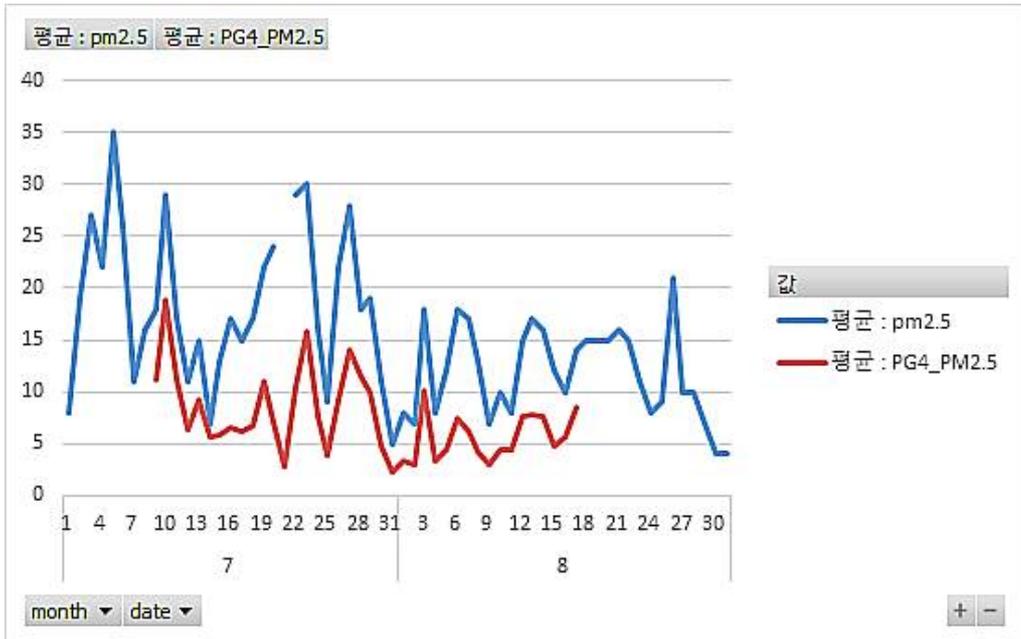


그림 64 인천 보전 환경원 논현지점의 PM2.5 데이터와 소래포구사거리의 취득한 환경 데이터 비교

(파란색 : 논현지점, 빨간색 : 소래포구사거리 PM10 평균값)

그림 66는 수집된 AWS 정보를 파이선을 통해서 불러 읽어들이는 데이터의 헤드 정보이다.

```

ftrain.head()
#train2.head()

Unnamed: 0  datetime  sr_pg4_temp  sr_pg4_humi  sr_pg4_pressure  sr_pg4_co2  sr_pg4_pm2.5  sr_pg4_pm10  sr_salty_temp  sr_salty_humi  ...  sr_galdae_pressure  sr_galdae_co2  sr_galdae_pm2.5  sr_galdae_pm10  year
0  0.0  2022-07-09 00:00:00  24.6  87.0  1002.8  388.0  7.0  8.0  23.9  92.0  ...  1001.4  400.0  6.0  7.0  2022.0
1  1.0  2022-07-09 00:05:00  24.5  87.0  1002.8  384.0  8.0  9.0  23.9  92.0  ...  1001.4  399.0  6.0  6.0  2022.0
2  2.0  2022-07-09 00:10:00  24.5  88.0  1002.8  385.0  9.0  10.0  23.9  93.0  ...  1001.4  399.0  6.0  7.0  2022.0
3  3.0  2022-07-09 00:15:00  24.6  87.0  1002.7  386.0  9.0  10.0  23.9  93.0  ...  1001.3  397.0  6.0  7.0  2022.0
4  4.0  2022-07-09 00:20:00  24.6  87.0  1002.8  389.0  9.0  11.0  23.9  93.0  ...  1001.4  398.0  6.0  7.0  2022.0

5 rows x 38 columns

```

그림 65 파이선으로 읽어들이는 데이터의 헤드 정보

```
▶ train.sr_pg4_temp.describe()
count    11182.000000
mean      27.622277
std       3.105700
min       20.800000
25%       25.300000
50%       27.300000
75%       29.600000
max       38.300000
Name: sr_pg4_temp, dtype: float64
```

그림 66 파이선으로 읽어들이온 온도, 데이터 특성

파이선으로 포구 사거리 빅데이터의 온도 특성을 읽어 들이면, 전체 데이터 수는 11,182개이고, 평균은 27.6°C, 최솟값은 20.8°C, 최댓값은 38.3°C로 확인된다.

```
[12] train.sr_pg4_co2.describe()
count    11182.000000
mean     401.137990
std      20.205992
min      357.000000
25%      387.000000
50%      397.000000
75%      412.000000
max      487.000000
Name: sr_pg4_co2, dtype: float64
```

그림 67 파이선으로 읽어들이온 이산화탄소 데이터 특성

파이선으로 포구 사거리 빅데이터의 이산화탄소 특성을 읽어 들이면, 전체 데이터 수는 11,182개이고, 평균은 401, 최솟값은 357, 최댓값은 487로 확인된다.

아울러 그림 68과 같이 수집된 데이터는 사이트별로 온도, 습도, 미세먼지, 초미세먼지, 기압, 이산화탄소 등과 함께, 우적량, 풍향, 풍속, 조도, 자외선 등의 다양한 정보를 사이트별로 가지고 있다.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11520 entries, 0 to 11519
Data columns (total 38 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Unnamed: 0                            11467 non-null  float64
1   datet ime                             11467 non-null  datet ime64[ns]
2   sr_pg4_t emp                          11182 non-null  float64
3   sr_pg4_humi                          11182 non-null  float64
4   sr_pg4_p ressure                      11182 non-null  float64
5   sr_pg4_co2                           11182 non-null  float64
6   sr_pg4_pm2.5                         11182 non-null  float64
7   sr_pg4_pm10                          11182 non-null  float64
8   sr_sal ty_t emp                       11206 non-null  float64
9   sr_sal ty_humi                       11206 non-null  float64
10  sr_sal ty_solar                       11206 non-null  float64
11  sr_sal ty_UVI                         11206 non-null  float64
12  sr_sal ty_rainfall                   11206 non-null  float64
13  sr_sal ty_Wind_Speed                 11206 non-null  float64
14  sr_sal ty_Wind_Gust                  11206 non-null  float64
15  sr_sal ty_Wind_Direction             11206 non-null  float64
16  sr_sal ty_p ressure                  11206 non-null  float64
17  sr_sal ty_co2                       11206 non-null  float64
18  sr_sal ty_pm2.5                     11206 non-null  float64
19  sr_sal ty_pm10                      11206 non-null  float64
20  sr_galdae_t emp                      11358 non-null  float64
21  sr_galdae_humi                      11358 non-null  float64
22  sr_galdae_solar                     11358 non-null  float64
23  sr_galdae_UVI                       11358 non-null  float64
24  sr_galdae_Rainfall                  11358 non-null  float64
25  sr_galdae_Wind_Speed                11358 non-null  float64
26  sr_galdae_Wind_Gust                 11358 non-null  float64
27  sr_galdae_Wind_Direction             11358 non-null  float64
28  sr_galdae_p ressure                  11358 non-null  float64
29  sr_galdae_co2                       11358 non-null  float64
30  sr_galdae_pm2.5                     11358 non-null  float64
31  sr_galdae_pm10                      11358 non-null  float64
32  year                                 11467 non-null  float64
33  month                                11467 non-null  float64
34  day                                  11467 non-null  float64
35  hour                                 11467 non-null  float64
36  minute                               11467 non-null  float64
37  second                               11467 non-null  float64
dtypes: datet ime64[ns](1), float64(37)
memory usage: 3.3 MB

```

그림 68 파이선으로 읽어들이는 전체 데이터 특성

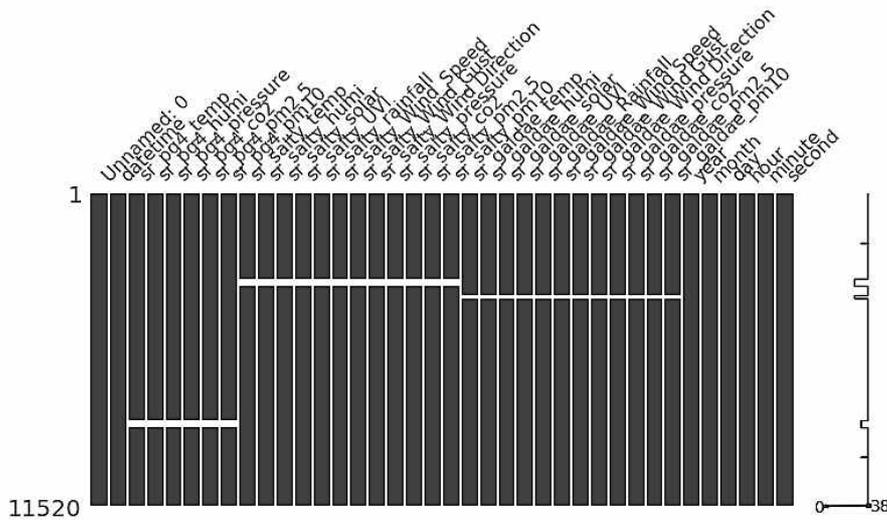


그림 69 데이터 중 Null 값 분포

환경 데이터의 측정데이터의 간격은 매 5분 간격으로 평균값으로 클라우드에서 저장되고 있다. 개별 행의 결측치를 직접 엑셀과 같은 표로 확인하는 것은 매우 어려운 일이므로 이에 결측치를 확인하고 패턴을 감지하기 위하여 Python package 중 하나인 Missingno matrix를 사용하였다<sup>170)</sup> 손실구간은 대부분 데이터 수신부 Gateway에서 네트워크 또는 전원부 접속 불량으로 생겼으며, 외부 이산화탄소 센서 부는, 자체 배터리를 이용한 파워 공급이 외부전원이 공급이 원활하지 않은 경우나, 전원 부족으로 기기 작동이 차단되어, 미싱(missing)된 경우로 추정된다.

결측치를 잘못된 값으로 채운다면 편향된 추정치와 왜곡된 통계량 및 결론을 얻을 수 있다<sup>171)</sup><sup>172)</sup>. 결측치를 처리하기 위해서는 누락된 데이터를 제거하거나 결측치를 특정 값으로 보간하는 방법 등 다양한 방법이 존재한다. 본 연구에서는, 평균으로 보간하여 결측치를 갖는 해당 열을 보충하는 방법을 사용하였다.

한편 데이터 이상치란 관측된 데이터의 범위에서 벗어난 아주 작거나 큰 값을

170) Bilogur, A.(2018), “Missingno: a missing data visualization suite,” *Journal of Open Source Software*, 3(22), 547.

171) 심대천 외(2022), 「Prediction of chloride concentration in groundwater on Jeju Island using XGBoost regression machine learning」, 『지질학회지』, 58(2), 243-255.

172) Acock, A. C.(2005), “Working with missing values,” *Journal of Marriage and family*, 67(4), 1012-1028.

말한다. 이상치 데이터는 모델의 예측 성능을 하락시키는 불필요한 요소로써 제거가 필수적이다<sup>173)</sup>. 이상치 제거를 위하여 먼저 사분위법(Inter Quantile Range, IQR)을 사용하였다. 사분위법은 데이터를 오름차순으로 정렬하였을 때 중간 50%의 해당하는 데이터들을 말하며  $Q_1$ ,  $Q_2$ ,  $Q_3$  각 0.25, 중앙값, 0.75에 위치하는 데이터를 기준으로 네 구간으로 나눈다. 사분위법(IQR)을 이용한 이상치는  $Q_1$ ,  $Q_3$ 에 있는 데이터 값에 1.5배 하여 중앙값을 빼거나 더해 구한 값의 이하, 이상 값을 말한다(식 7). 그러나 큰 값을 가지는 데이터가 상당수 제거되어 사분위법을 이용한 이상치 제거는 사용하지 못하였다.

$$IQR = Q_3 - Q_1$$

$$Q_1 - 1.5 \times IQR \leq X \leq Q_3 + 1.5 \times IQR \quad (7)$$

그리고 수집된 데이터의 유의성을 테스트하기 위하여, IBM사의 SPSS 소프트웨어를 사용하여, 아노바테스트를 실시하였다. 공기질 변수 중 이산화탄소인 'CO<sub>2</sub>'를 타겟으로 선정하고, 실시하였을 때, 측정된 데이터 대부분이 P Value가 0.05 또는 0.01 이하로, 유의성 검증에 문제가 없음을 확인할 수 있었다. 다만, 갈대숲 우적량(sr\_galdae\_Rainfall), 분 정보(minute)는 유의성이 기각되었다.

표 41 소래 습지 공원 유의성 test 아노바 테스트 결과

		ANOVA			F	유의확률
		제곱합	자유도	평균제곱		
sr_pg4_temp	집단-간	42274.263	168	251.633	42.281	.000
	집단-내	63989.697	10752	5.951		
	전체	106263.961	10920			
sr_pg4_humi	집단-간	561780.103	168	3343.929	23.270	.000
	집단-내	1545100.170	10752	143.704		
	전체	2106880.273	10920			
sr_pg4_pressure	집단-간	15228.948	168	90.649	8.066	.000
	집단-내	120836.263	10752	11.238		
	전체	136065.212	10920			
sr_pg4_co2	집단-간	1952487.121	168	11621.947	49.173	.000
	집단-내	2541219.607	10752	236.349		
	전체	4493706.728	10920			
sr_pg4_pm2.5	집단-간	21985.132	168	130.864	5.006	.000
	집단-내	281053.277	10752	26.140		
	전체	303038.408	10920			

173) 심대천 외(2022), 「XGBoost Regression 기계학습을 이용한 제주도 지하수 염소이온 농도예측 연구 사례」, 『지질학회지』, 58(2), 243-255.

sr_pg4_pm10	집단-간	26153.029	168	155.673	5.326	.000
	집단-내	314242.590	10752	29.226		
	전체	340395.618	10920			
sr_salty_temp	집단-간	47518.279	168	282.847	59.113	.000
	집단-내	52810.623	11037	4.785		
	전체	100328.902	11205			
sr_salty_humi	집단-간	645596.354	168	3842.835	38.207	.000
	집단-내	1110085.021	11037	100.579		
	전체	1755681.374	11205			
sr_salty_solar	집단-간	114197341.996	168	679746.083	17.901	.000
	집단-내	419107453.041	11037	37972.950		
	전체	533304795.038	11205			
sr_salty_UVI	집단-간	8718.469	168	51.896	16.198	.000
	집단-내	35361.724	11037	3.204		
	전체	44080.194	11205			
sr_salty_rainfall	집단-간	4088.362	168	24.335	1.203	.038
	집단-내	223241.754	11037	20.227		
	전체	227330.116	11205			
sr_salty_Wind_Speed	집단-간	1726.758	168	10.278	12.469	.000
	집단-내	9098.022	11037	.824		
	전체	10824.780	11205			
sr_salty_Wind_Gust	집단-간	6219.645	168	37.022	18.464	.000
	집단-내	22129.834	11037	2.005		
	전체	28349.478	11205			
sr_salty_WindDirection	집단-간	2195578.170	168	13068.918	3.416	.000
	집단-내	42229597.074	11037	3826.184		
	전체	44425175.243	11205			
sr_salty_pressure	집단-간	13955.186	168	83.067	7.633	.000
	집단-내	120110.266	11037	10.883		
	전체	134065.452	11205			
sr_salty_pm2.5	집단-간	1686.674	168	10.040	1.739	.000
	집단-내	63703.786	11037	5.772		
	전체	65390.460	11205			
sr_salty_pm10	집단-간	1689.695	168	10.058	1.605	.000
	집단-내	69174.844	11037	6.268		
	전체	70864.539	11205			
sr_galdae_temp	집단-간	49419.201	168	294.162	59.737	.000
	집단-내	53812.335	10928	4.924		
	전체	103231.536	11096			
sr_galdae_humi	집단-간	683680.610	168	4069.527	38.051	.000
	집단-내	1168738.674	10928	106.949		
	전체	1852419.284	11096			
sr_galdae_solar	집단-간	127860499.813	168	761074.404	16.631	.000
	집단-내	500100306.854	10928	45763.205		
	전체	627960806.667	11096			
sr_galdae_UVI	집단-간	10040.779	168	59.767	15.308	.000
	집단-내	42666.026	10928	3.904		
	전체	52706.806	11096			
sr_galdae_Rainfall	집단-간	4521.545	168	26.914	1.259	.014
	집단-내	233612.896	10928	21.377		
	전체	238134.441	11096			
sr_galdae_WindSpeed	집단-간	2024.971	168	12.053	24.383	.000
	집단-내	5402.112	10928	.494		
	전체	7427.083	11096			
sr_galdae_WindGust	집단-간	6332.503	168	37.693	25.458	.000
	집단-내	16180.162	10928	1.481		
	전체	22512.665	11096			
sr_galdae_WindDirection	집단-간	16182017.614	168	96321.533	8.722	.000
	집단-내	120686789.452	10928	11043.813		
	전체	136868807.066	11096			
sr_galdae_pressure	집단-간	13835.719	168	82.355	7.457	.000
	집단-내	120696.844	10928	11.045		
	전체	134532.564	11096			
sr_galdae_co2	집단-간	8662907.842	168	51564.928	155.049	.000
	집단-내	3634342.156	10928	332.572		
	전체	12297249.998	11096			
sr_galdae_pm2.5	집단-간	4002.005	168	23.821	2.464	.000
	집단-내	105648.163	10928	9.668		
	전체	109650.167	11096			
sr_galdae_pm10	집단-간	4112.669	168	24.480	2.330	.000
	집단-내	114805.688	10928	10.506		
	전체	118918.357	11096			
vear	집단-간	.000	168	.000		

month	집단-내	.000	11037	.000		
	전체	.000	11205			
	집단-간	82.593	168	.492	2.034	.000
day	집단-내	2667.363	11037	.242		
	전체	2749.956	11205			
	집단-간	61496.338	168	366.050	5.923	.000
hour	집단-내	682096.950	11037	61.801		
	전체	743593.288	11205			
	집단-간	74215.921	168	441.761	10.553	.000
minute	집단-내	462019.061	11037	41.861		
	전체	536234.982	11205			
	집단-간	54320.942	168	323.339	1.087	.212
second	집단-내	3284516.522	11037	297.591		
	전체	3338837.464	11205			
	집단-간	.000	168	.000		
	집단-내	.000	11037	.000		
	전체	.000	11205			

## 2.1.2 파이선을 이용한 IoT 데이터 상관관계분석

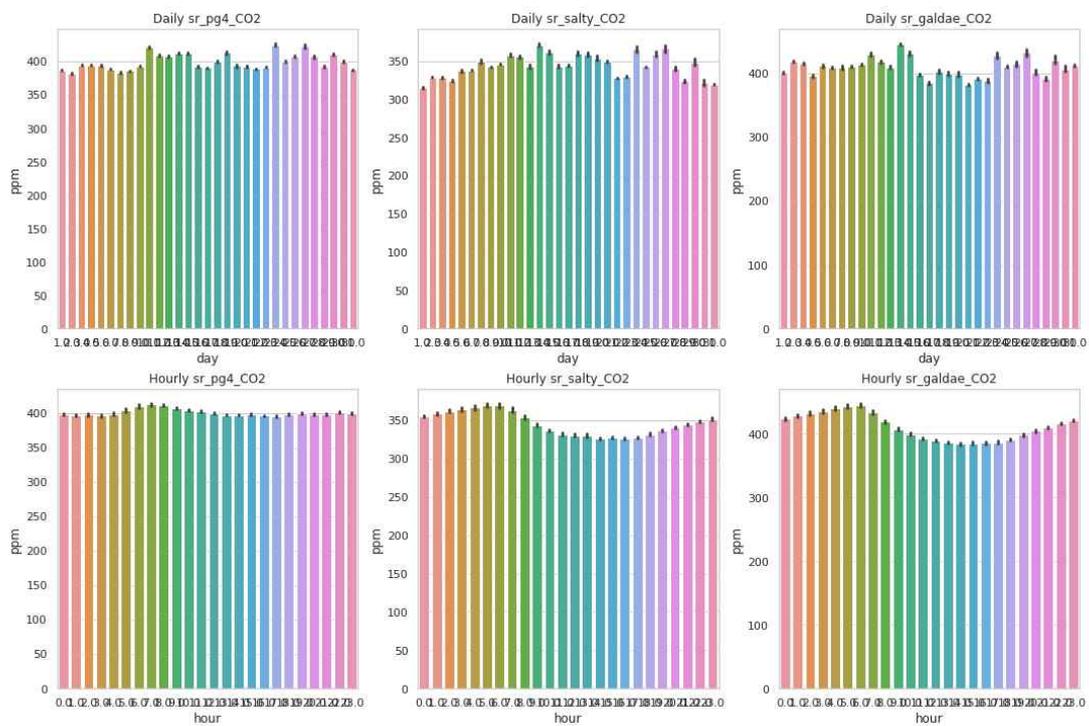


그림 70 소래생태공원 각 사이트의 일자별, 시간별 이산화탄소 변화량

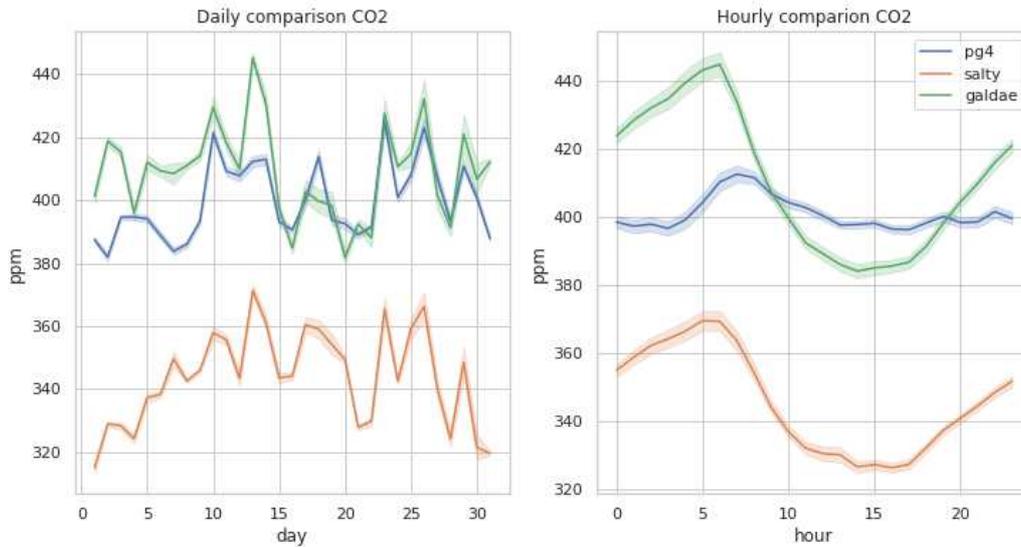


그림 71 사이트별, 이산화탄소 일별, 시간별 비교 차트

막대그래프를 통한 매일 일자별 소래생태공원의 이산화탄소 발생량의 변화를 비교하였다(그림 70). 그림 71 이산화탄소 비교에서 도시에 위치한 포구 사거리 는 시간과 날짜와 관계없이 일정하게 유지되는 것을 확인할 수 있다. 한편 갈대숲 은, 주간에 태양광으로 인한 광합성이 일어나게 되어, 오전 7시부터, 이산화탄소 량이 감소하고, 오전 10~ 19시까지 는 포구 사거리에 비해서 이산화탄소 량이 적게 발생하는 것을 확인할 수 있다. 하지만, 광합성을 하지 않은 밤이나, 태양이 약한 날은, 도로부인 포구 사거리에 비해서 이산화탄소 발생 량이 급격히 증가하는 것을 확인할 수 있었다. 한편, 염습지도 갈대숲과 유사하게, 태양광을 통한 광합성 이 일어나는 시간에는, 이산화탄소의 양이 감소함을 확인할 수 있었고, 평소 수치 또한 도로부인 포구 사거리에 비해서도 평균 40ppm, 최고 60ppm 정도 적게 측정되는 것을 확인할 수 있다.

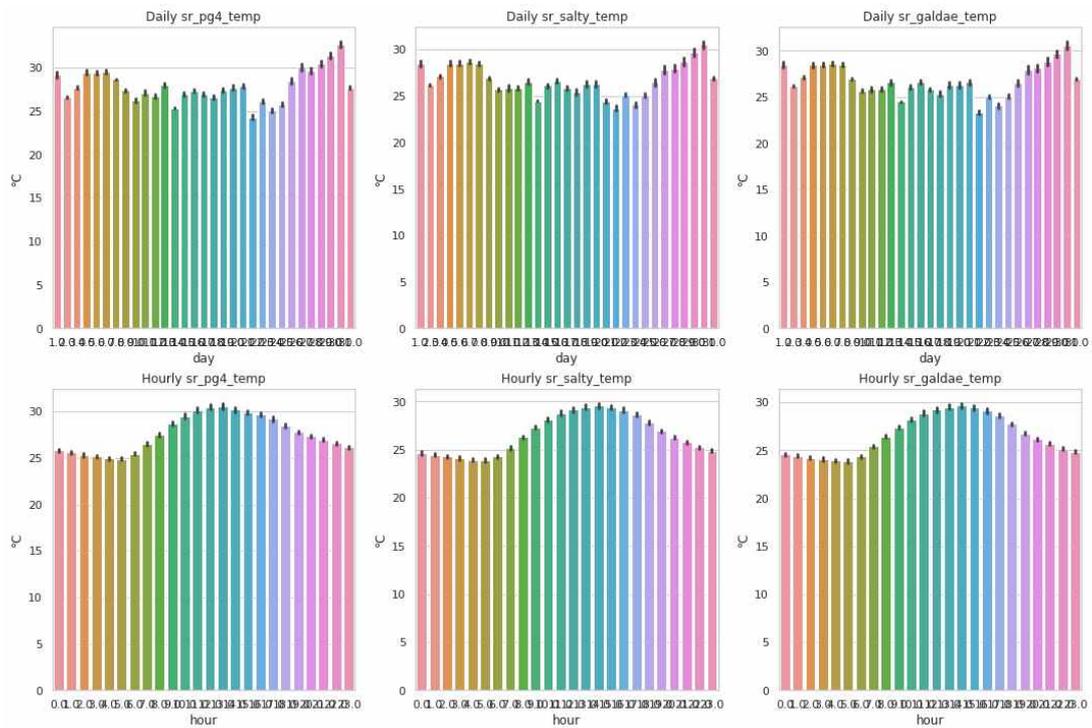


그림 72 소래생태공원 각 사이트의 일자별, 시간별 온도 변화량

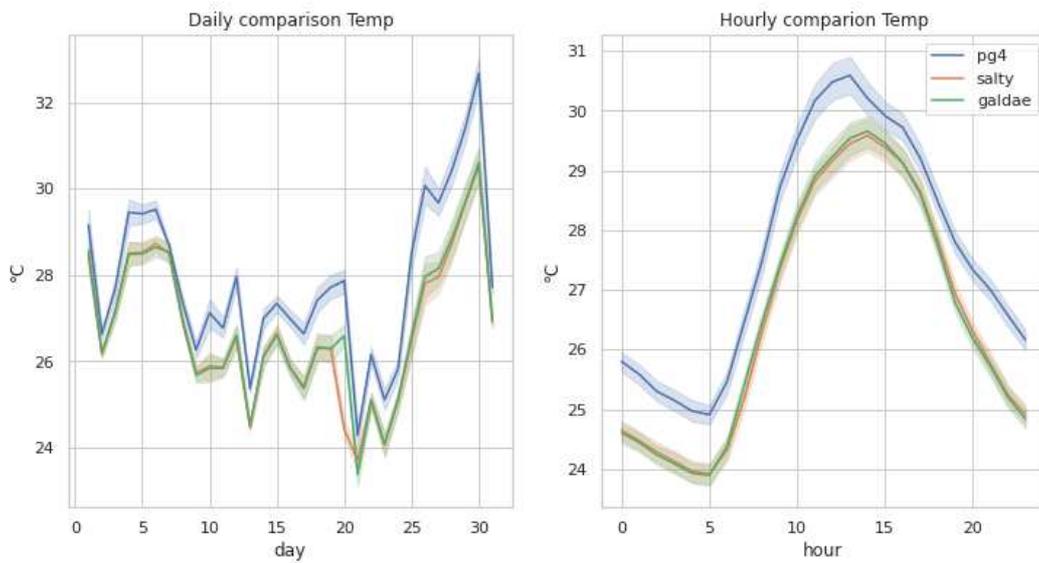


그림 73 사이트별, 온도 일별, 시간별 비교 차트

그림 72의 막대그래프를 통한 매일 일자별 소래생태공원의 온도변화를 비교하였다. 그림 73의 온도 비교에서, 도시에 있는 포구 사거리의 시간과 날짜와 관계 없이 일정하게 약 1.5 °C 정도 높게 유지되는 것을 확인할 수 있다. 이와 같은 온도변화는 새벽 5시까지 꾸준히 감소하는 것으로 유지되다가, 다시 상승하기 시작하여 약 14시를 기점으로 감소세로 접어든다. 하지만, 도시에 있는 포구 사거리는 야간에도 온도가 더 이상 떨어지지 않고, 갈대숲, 염습지와는 그대로 1.5°C 차이를 지속적으로 유지한다.

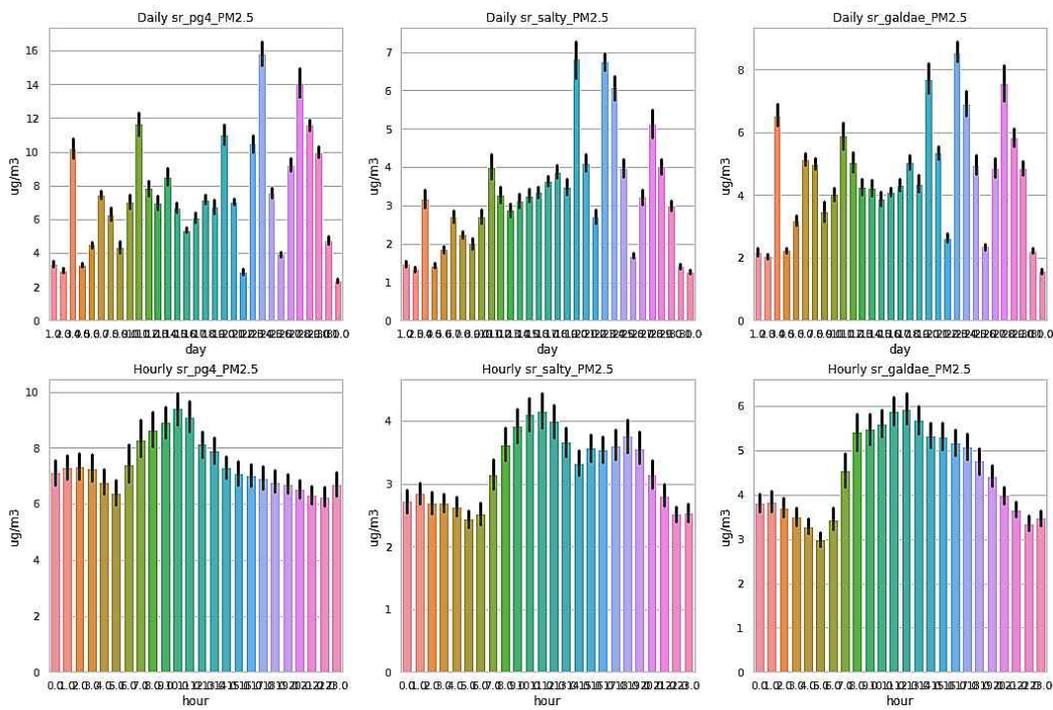


그림 74 소래생태공원 각 사이트의 일자별, 시간별 초미세먼지(PM2.5) 습도 변화량

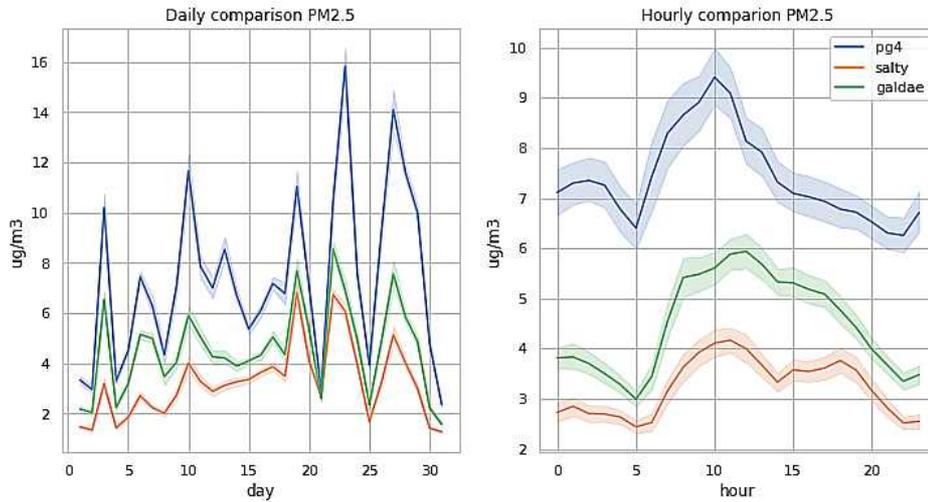


그림 75 사이트별, 초미세먼지 일별, 시간별 비교 차트

그림 74 막대그래프를 통한 매일 일자별 소래생태공원의 초미세먼지(PM2.5)의 변화를 비교하였다. 그림 75의 초미세먼지의 비교에서, 도시에 있는 포구 사거리의 매일 평균 갈대숲보다는 약  $3\mu\text{g}/\text{m}^3$  정도를 높게 유지하였고, 염습지 대비는  $4\mu\text{g}/\text{m}^3$ 의 높은 차이를 유지하였다. 한편, 염습지와 갈대숲의 초미세먼지를 비교하여 보았을 때는, 정오 무렵에 가장 미세먼지의 발생량이 많고, 두 곳의 차가 컸으며, 오전 5시경이  $0.5\mu\text{g}/\text{m}^3$ 로 가장 적은 것을 확인할 수 있었다.

그림 76 막대그래프를 통한 매일 일자별 소래생태공원의 미세먼지(PM10)의 변화를 비교하였다. 그리고 그림 77의 미세먼지의 비교에서, 도시에 있는 포구 사거리는 매일 평균 갈대숲보다는 약  $3.5\mu\text{g}/\text{m}^3$  정도를 유지하였고, 염습지 대비는  $4\mu\text{g}/\text{m}^3$ 의 차이를 유지하였다. 한편, 염습지와 갈대숲의 미세먼지를 비교하여 보았을 때는, 정오 무렵에 가장 미세먼지의 발생량이 많고, 두 곳의 차가 컸으며, 오전 5시경이  $0.5\mu\text{g}/\text{m}^3$ 로 가장 적은 것을 확인할 수 있었다.

그림 76 막대그래프를 통한 매일 일자별 소래생태공원 습도의 변화를 비교하였다. 그림 77의 습도 비교에서, 도시에 있는 포구 사거리는 매일 평균 갈대숲, 염습지보다는 약 5% 정도를 낮게 유지되었다. 한편, 염습지와 갈대숲의 습도를 비교하여 보았을 때는, 정오 무렵에 약 3% 정도 갈대밭이 적게 나타나고, 그 외에는 두 곳의 습도 차가 가벼운 것을 확인할 수 있었다.

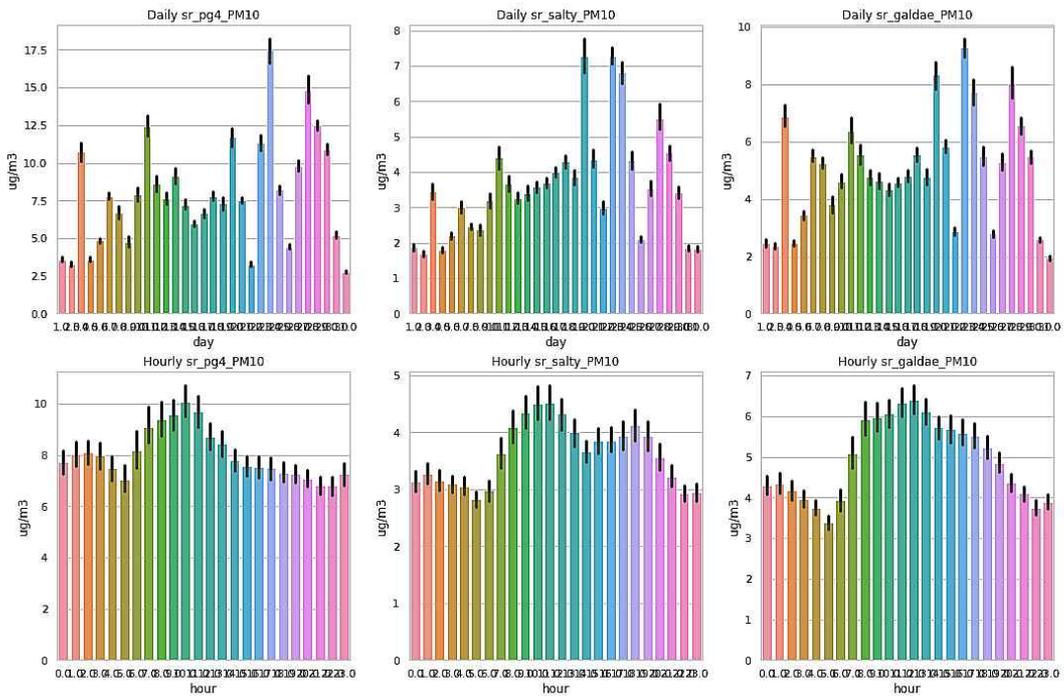


그림 76 소래생태공원 각 사이트의 일자별, 시간별 미세먼지(PM10) 변화량

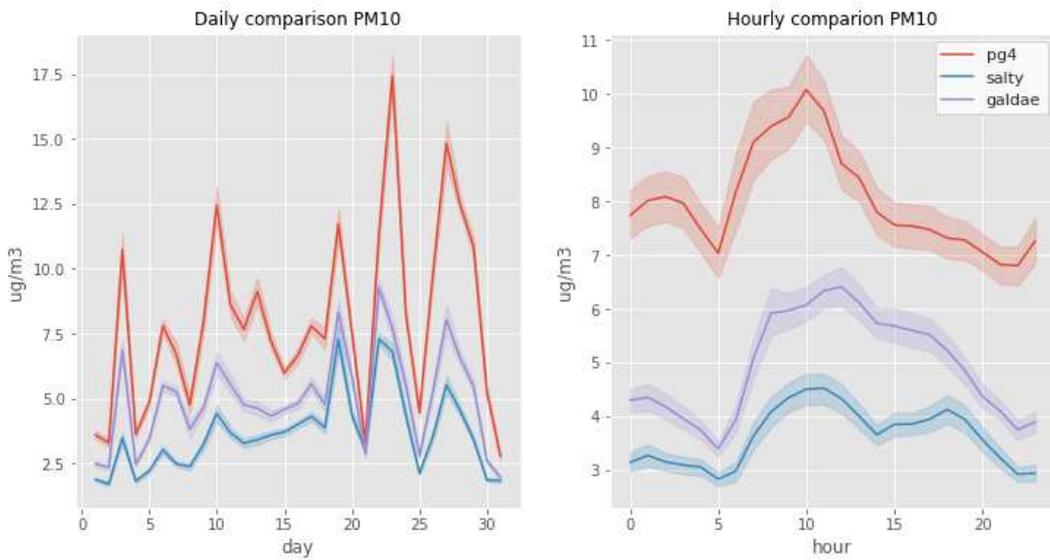


그림 77 사이트별, 미세먼지 일별, 시간별 비교 차트

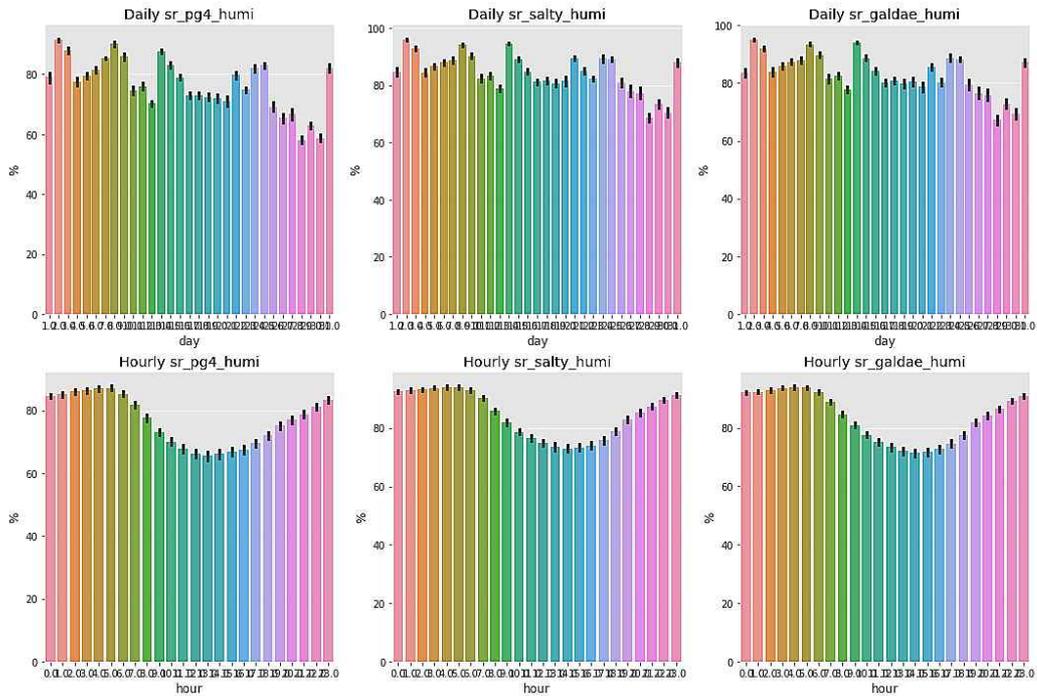


그림 78 소래생태공원 각 사이트의 일자별, 시간별 습도 변화량

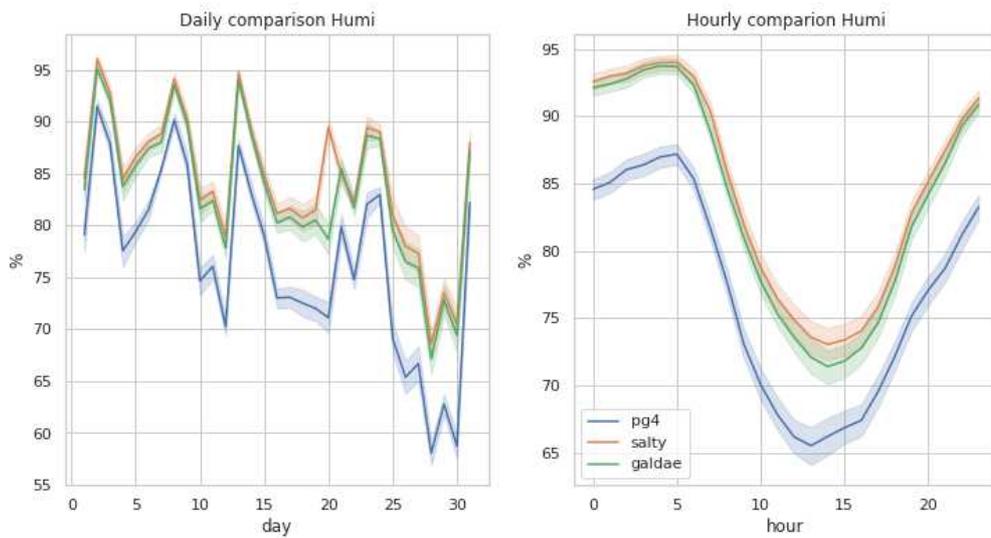


그림 79 사이트별, 습도 일별, 시간별 비교 차트

### 2.1.3 박스 플롯을 통한 데이터 비교

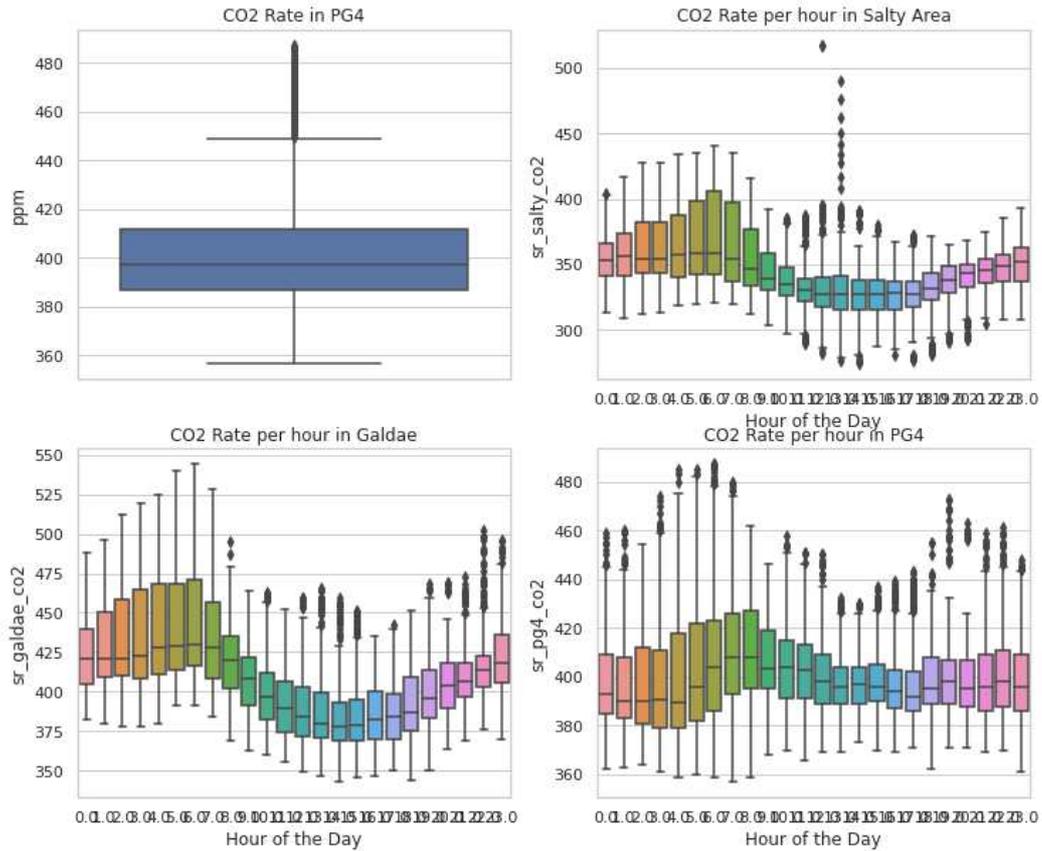


그림 80 소래생태공원 각 사이트의 일자별 이산화탄소 변화량

그림 80 박스 플롯에서는 하루 평균 이산화탄소 발생량이 대략 400ppm 내외인 것을 알 수 있다. 또한, 사이트별로 매일 시간별 통계를 살펴보았을 때, 염습지와 갈대밭의 이산화탄소 발생량이 매일 기상에 따라서 변동이 큰 것을 알 수 있으며, 그와 대조적으로 도심지에 있는 포구 사거리는 유동량이 많지 않으면 확인할 수 있다.

## 2.1.4 수치 범주 데이터 조건별 CO<sub>2</sub> 상관관계분석 변화량

본 연구에서, 소래포구 도로를 기준으로, 이산화탄소 발생량에 영향을 주는 상관관계분석을 파이선과 히트 맵을 통해서 먼저 살펴보았다.

	sr_pg4_temp	sr_pg4_humi	sr_pg4_co2	sr_pg4_pm2.5	sr_pg4_pm10
sr_pg4_temp	1.000000	-0.919477	0.089197	0.273673	0.252389
sr_pg4_humi	-0.919477	1.000000	-0.022775	-0.105941	-0.088788
sr_pg4_co2	0.089197	-0.022775	1.000000	0.367769	0.369326
sr_pg4_pm2.5	0.273673	-0.105941	0.367769	1.000000	0.995462
sr_pg4_pm10	0.252389	-0.088788	0.369326	0.995462	1.000000

그림 81 소래포구 도로에서의 상관관계분석

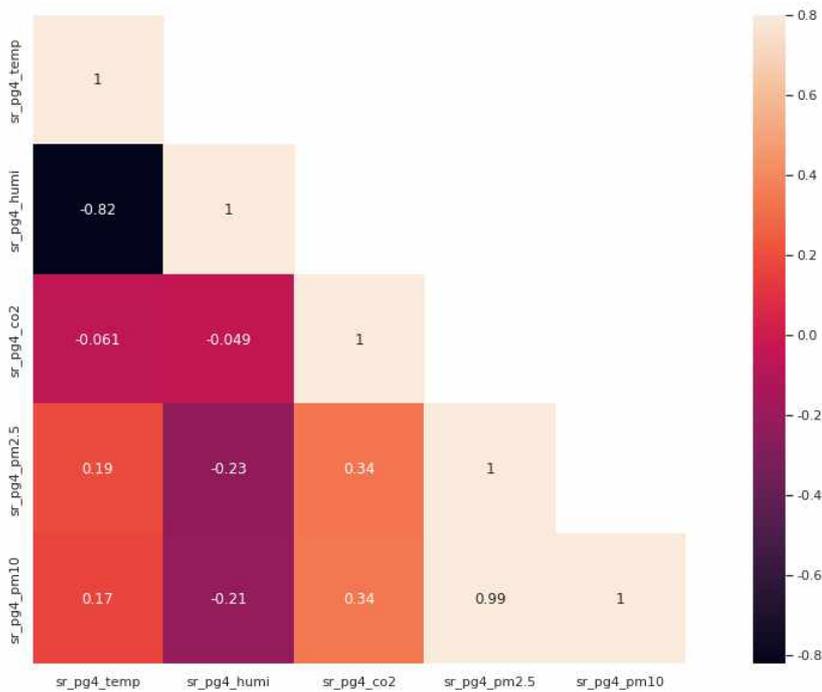


그림 82 히트 맵을 통한 소래포구 환경 요소 상관관계분석

소래포구사거리의 경우에 온도, 습도, 미세먼지, 초미세먼지, 이산화탄소만 두고 상관관계를 분석해보면, 초미세먼지 0.34, 미세먼지 0.34로 가장 높고, 온도는 -0.061, 습도는 -0.049로 그렇게 상관관계는 높지 않음을 확인할 수 있었다. 따라서 뒤에 히트 맵 분석과, ML을 통한 Regression 수행에서는 포구 사거리와 가까운 갈대숲의 풍향, 풍량, 풍속 데이터, 조도 등의 환경 데이터를 사용하여,

이산화탄소 발생량에 영향을 주는 요소를 살펴보고자 하였다. 그리고, 이어서 각 측정 사이트별로, 다른 가능한 모든 환경 요소를 추가적으로 고려하여 이산화탄소 발생량에 영향을 주는 요소의 가중치를 살펴보았다.

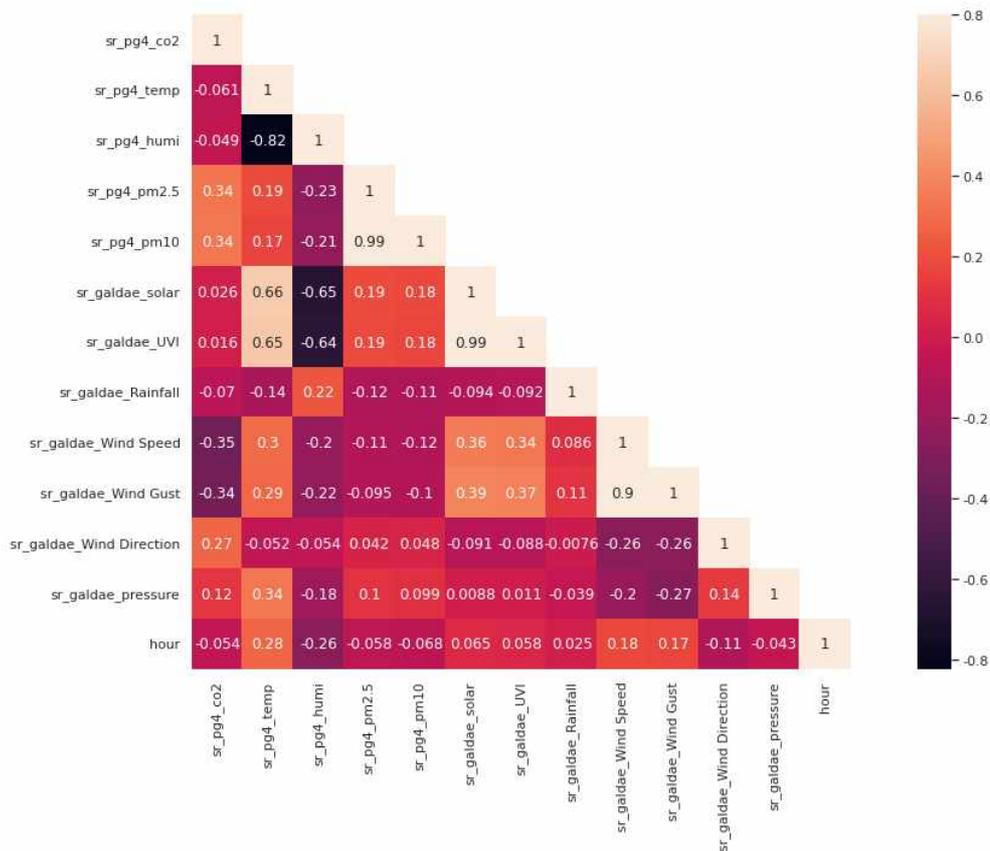


그림 83 포구 4거리 와 갈대 지역의 합성 데이터 상관관계

그림 83 소래포구 사거리 도로에서의 환경 요소는 도시 도로 여건상, 풍향 풍속 계 등의 추가 센서 설치가 어려워서, 위치적으로 가까운 갈대밭의 환경 데이터를 같이 사용하였다. 히트 맵을 분석하면, 이산화탄소는 미세먼지, 초미세먼지의 상관관계는 동일하게 0.34로 높은 수치임을 확인할 수 있었다. 또한, 풍속과 돌풍과의 이산화탄소 발생량의 관계는 -0.35, -0.34로 바람은 이산화탄소에 부정적인 영향을 주는 것을 알 수 있다.

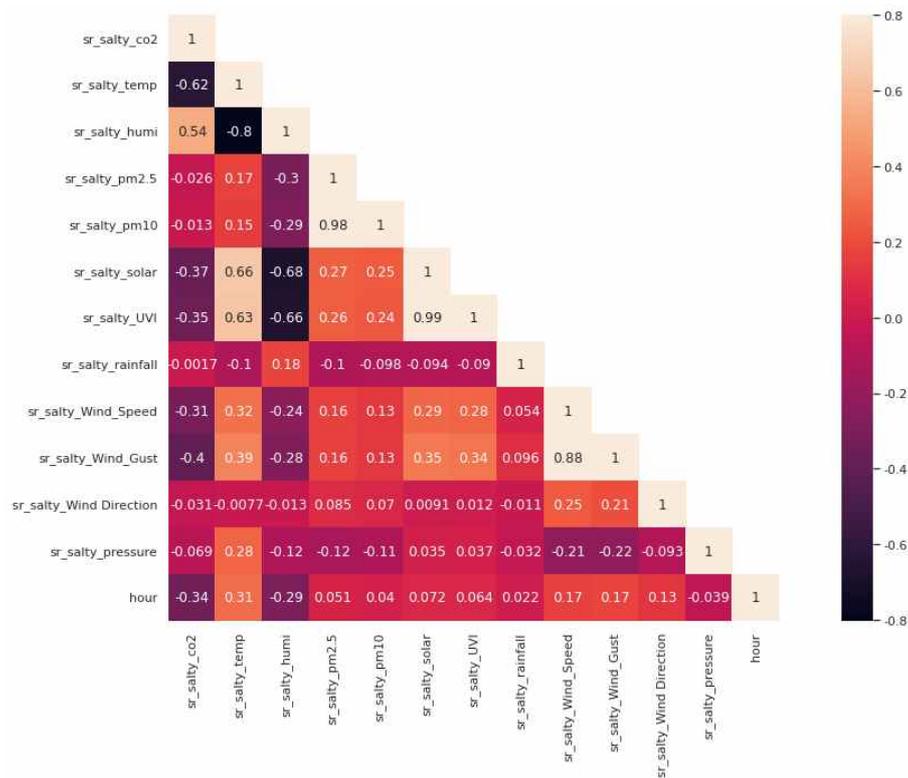


그림 84 염습지와 합성 데이터 상관관계

그림 84 소래생태공원의 염습지에서의 이산화탄소 발생 관련 환경 요소 히트 맵을 분석하면, 이산화탄소는 온도 -0.62, 습도 0.54, 태양 -0.37, 자외선 -0.35로 태양과 온도로 인한 감소 효과가 큰 것이 판단된다. 또한 돌풍 -0.4, 풍속 -0.31와 같이 돌풍의 속도와 바람의 강약에 많은 영향을 받는 것을 확인할 수 있다. 따라서, 도시 안에 공기 순환이 안 될 경우, 이산화탄소가 많이 발생하여 분산되지 못하고, 양이 증가하는 것을 예측할 수 있다.

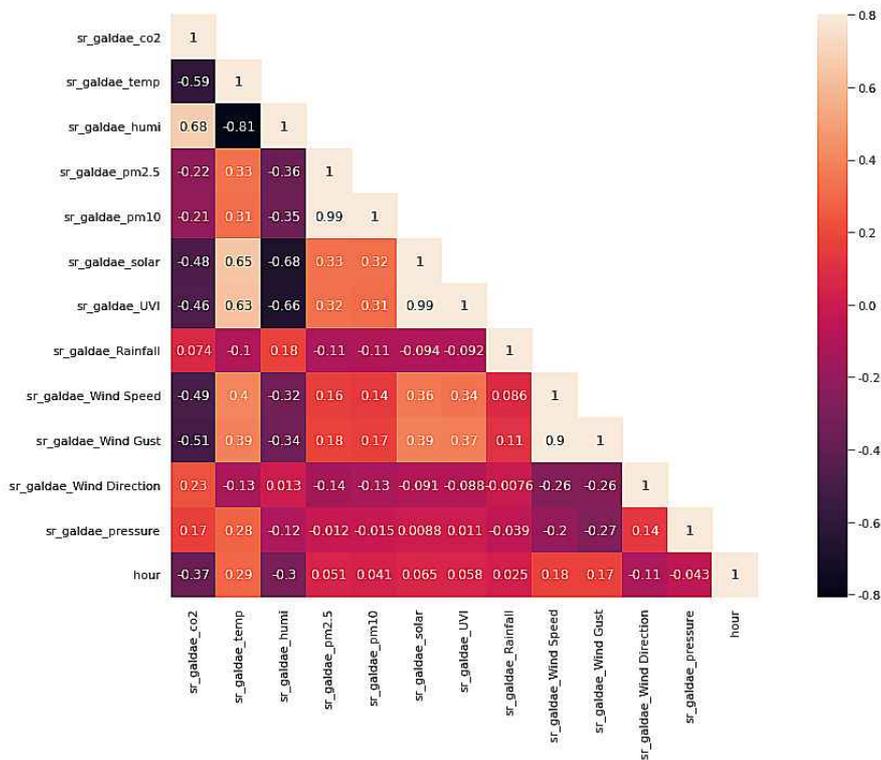


그림 85 갈대숲 합성 데이터 상관관계

그림 85 소래포구 갈대숲에서의 이산화탄소 발생 관련 환경 요소 히트 맵을 분석하면, 염습지와 매우 유사한 환경 상관관계임을 발견할 수 있다. 이산화탄소는 온도 -0.59, 습도 0.68, 태양 -0.48, 자외선 -0.46으로 태양과 온도로 인한 감소 효과가 큰 것이 판단된다. 또한 돌풍 -0.51, 풍속 -0.49와 같이 돌풍의 속도와 바람의 강약에 많은 영향을 받는 것을 확인할 수 있다. 따라서, 이산화탄소를 줄이는 요인은, 강한 햇볕과 바람을 통한 식물의 광합성 작용이 가장 큰 요소이다.

포구 사거리의 이산화탄소 측정데이터의 분포를 객관적으로 살펴보기 위해서, QQ plot을 시행하였다. QQ plot(Quantile Quantile plot)은 두 변수 간의 분포를 비교하기 위해서 사용되는 그래프이다<sup>174)</sup>. 그리고 각 예측값들과 실측값 간에 발생한 잔차의 정규성 검토를 위해 Q-Q Plot(Quantile-Quantile Plot)을 사

174) “[파이썬 시각화] QQ플롯(Quantile-Quantile plot)”, <https://blog.naver.com/breezhome50/222339337268> (검색일: 2022. 10. 15.)

용하였다. Q-Q Plot은 정규성을 평가하기 위한 일반적인 시각적 기법이며, 수집 데이터를 표준 정규분포의 분 위수와 비교하여 그리는 그래프이다. 모집단이 정규성을 보일수록 선형성을 보이며 잔차가 대각선 방향의 직선을 따르면 정규성이 있다고 해석하며, 잔차가 정규성일 보일수록 회귀분석이 잘 되었다고 평가할 수 있다.

QQ Plot은 수치로 평가하는 정량적 분석(Quantitative Analysis)이 아닌 시각적으로 평가하는 정성적 분석(Qualitative Analysis)이기 때문에 시각적인 그래프를 통해 정규성을 확인하였으며 파이선 프로그램을 이용해 그래프를 그렸다. 예측값 A와 예측값 B의 깊이별 실측값과의 Q-Q Plot 그래프는 그림 86와 같다.

먼저 outlier를 포함한 이산화탄소 정보의 정규분포와 두 번째는 outlier를 포함하지 않은 온도의 log 값을 취한 후 정규분포를 살펴본다. 포구 사거리 이산화탄소의 분포가 다소 왼쪽에 치우쳐져 있다. 대부분의 머신러닝은 종속변수가 normal이어야 하기에 정규분포를 갖는 것이 바람직하다. 대안으로 아웃라이어 데이터(outlier data)를 제거하고 "co2" 변수에 로그를 씌워 변경해 봐도 정규분포를 따르지는 않지만, 이전 그래프보다는 좀 더 우측으로 중심이 옮겨 있다.

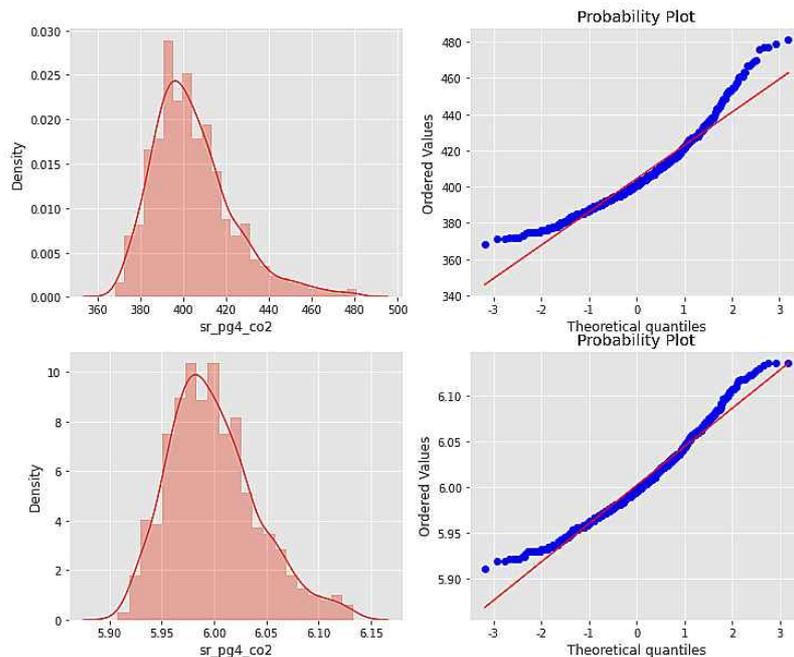


그림 86 이산화탄소 QQ 분포 분석

## 2.2 머신러닝 분석

AutoML을 이용하여, 각 환경 데이터를 학습한 후, 예측을 수행하기 위해 Stacking Ensemble, XGBoostRegressor, RandomForestRegressor, LightGBM 등의 4가지 AI 모델을 중심으로 분석을 통하여, 각 머신러닝의 효율성을 MSE, RMSE,  $R^2$ 값을 이용하여 평가한다. 일반적인 선형회귀(Regression) 데이터를, 단순한 수치 분석학 결과에 비교하여, 좀 더 세밀하게 이산화탄소에 크게 영향을 주는 요소를 머신러닝을 활용하여 살펴본다. 사용된 데이터는, 앞절에서 파이선으로 데이터 분석되어 보인 환경 빅데이터를 사이트별로 처리하여 적용하였으며, 각각 70:30의 비율로 트레이닝데이터(Training Data)와 밸리데이션 데이터(Validation Data)로 사용되었다. 데이터의 전처리 및 강화 후, 필터링된 기상 데이터와 환경 데이터는 이산화탄소 수준을 예측하는 데 사용된다. 마지막으로 교차검증은 예측에 사용할 수 있는 최상의 모델을 선택<sup>175)</sup>하는 데 사용된다.

### 2.2.1 염습지 공원 머신러닝 모델링 비교

본 회귀분석에 사용된 머신러닝은 AutoML을 사용하였으며, 그중에서 Stacking Ensemble 모델, XG Boost regressor, LGBM, Randomforest의 4가지 모델만을 사용하여, 알고리즘에 대한 정확성 등을 평가하고자 한다. 각각의 알고리즘에 사용된 전처리 모듈과 엔진, 본 처리에 사용된 알고리즘과 엔진은 표 42와 같다.

먼저, 첫 번째 Stacking Ensemble Algorithm은 XGboostRegressor 4개와 LGBM의 모두 5가지 알고리즘을 사용하였으며, 모듈은 Scikit Learn과 AutoML을 각각 사용하였다. 전처리에 사용된 스케일러는 각각, Standard scaler와 MaxAbsScaler등이 활용되었다.

두 번째로, XGboost regressor는 StandardScaler를 전처리로 사용한다. 세 번째로, Light GBM은 StandardScaler를 사용하여 전처리를 진행한다. 끝으로, RandomForestRegressor는 MaxAbsScaler를 사용하여 전처리를

---

175) 조남훈(2006), 「교차검증을 이용한 SVM 전력수요예측」, 『전기학회논문지』 (A), 55(11), 485-491.

진행한다.

표 42 연습지 CO<sub>2</sub> Regression 실험에 적용된 ML 알고리즘과 모듈

Algorithm Name	pre processing	Preprocessing	Algorithms	module
Stacking Ensemble	StandardScaler	sklearn	XGBoostRegressor	automl
	StandardScaler	sklearn	XGBoostRegressor	automl
	StandardScaler	sklearn	LightGBMRegressor	automl
	StandardScaler	sklearn	XGBoostRegressor	automl
	MaxAbsScaler	sklearn	XGBoostRegressor	automl
XGBoostRegressor	StandardScaler	sklearn	XGBoostRegressor	automl
LightGBM	StandardScaler	sklearn	LightGBMRegressor	automl
RandomForest	MaxAbsScaler	sklearn	RandomForestRegressor	sklearn

### 2.2.1.1 연습지 공원 효용성 분석

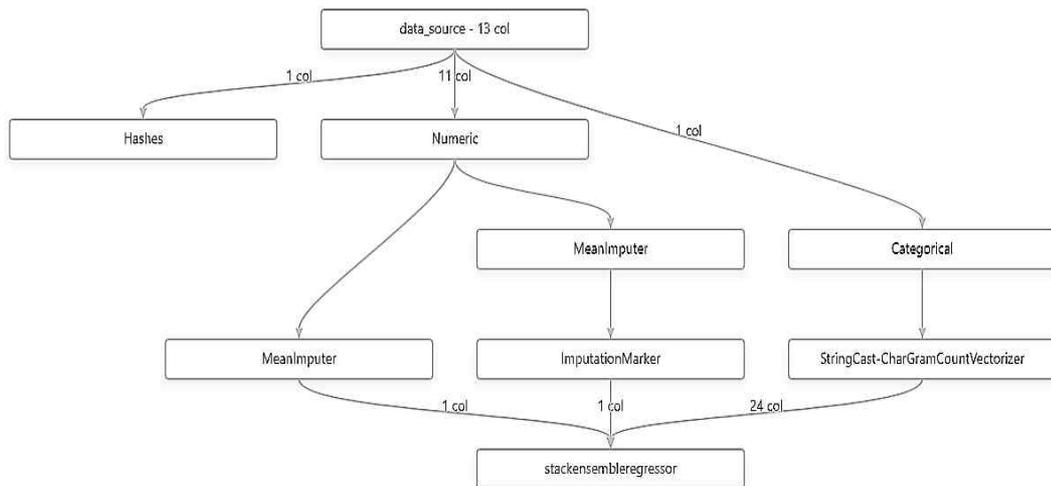


그림 87 분석에 사용된 Stacking Algorithm 다이어그램

본 다이어그램은 이 특정 모델을 생성하기 위해 Automated ML이 적용한 데이터 사전 처리, 기능 엔지니어링, 확장 기술 및 머신러닝 알고리즘을 보여준다.

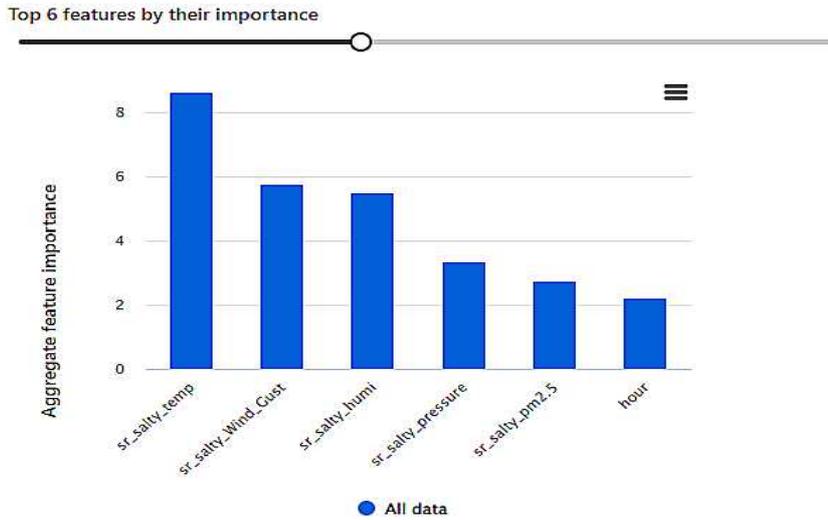


그림 88 Top6 features by their importance

적용된 머신러닝 중에서, 가장 높은  $R^2$  값을 갖는, 앙상블 알고리즘을 분석해보면, 타겟 데이터를 보행량으로 두고, 적용된 데이터에서 상위 6개의 요소를 출력하면, 온도(sr\_salty\_temp), 돌풍(sr\_salty\_Wind\_Gust), 습도(sr\_salty\_humi), 기압(sr\_salty\_pressure), 초미세먼지(sr\_salty\_pm2.5), 시간(hour)의 순으로 영향을 주는 것을 알 수 있다.

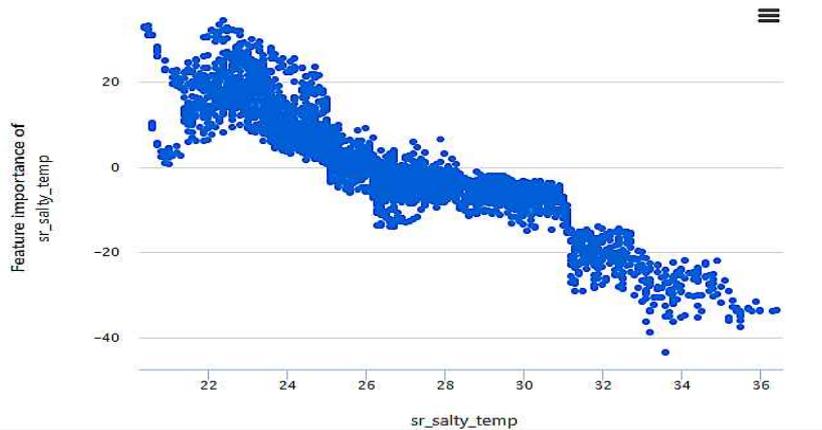


그림 89 Stacking Ensemble 머신러닝으로 적용된 온도와  $CO_2$ 와의 상관관계

그림 89 염습지에서 기온과 이산화탄소와의 상관관계를 그래프로 살펴보면, 온도의 경우에는, 음의 기울기를 가지고 있어서, 이산화탄소 저감에 효과가 큰 것을 알 수 있다. 이것은 야외에서 태양광에 노출로 인한 온도가 상승함에 따라서 야외식물의 광합성이 활발히 일어나는 것을 이산화탄소 저감이 되는 것으로 유추된다.

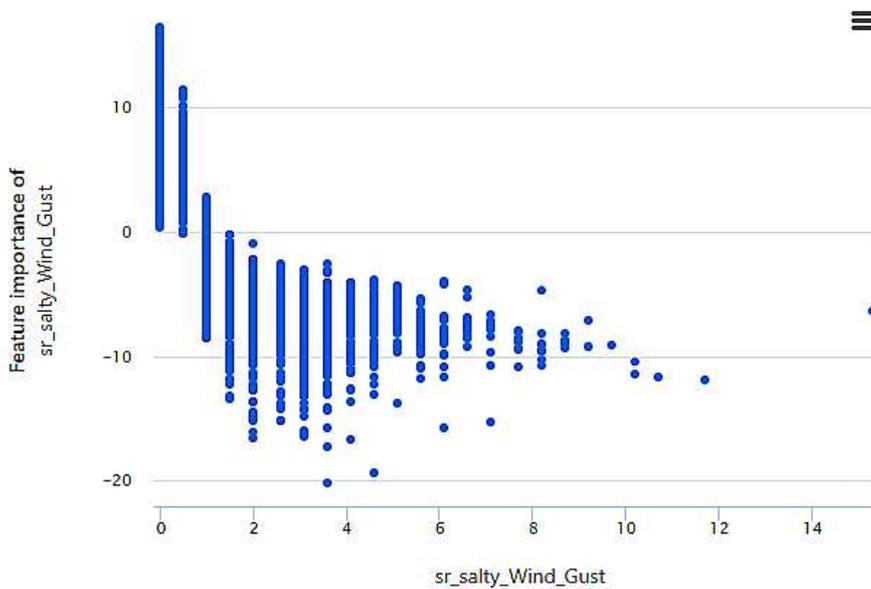


그림 90 Stacking Ensemble 머신러닝으로 적용된 돌풍과 CO<sub>2</sub>와의 상관관계

그림 90 강한 돌풍은, 이산화탄소의 감소 효과가 매우 큰 것으로 보인다. 상관관계 곡선에서, 이산화탄소는 바람이 강하게 불수록, 음의 기울기로 감소하는 것을 알 수 있다. 따라서, 도시 지역에서 바람길이나 공기 순환을 통해서 이산화탄소 저감이 가능한 것을 알 수 있다.

그림 91 습도의 경우에는, 이산화탄소와의 Regression 그래프가 양의 기울기를 가지는 것을 알 수 있다. 일반적으로 습도가 높은 경우에는, 비가 오거나 구름에 태양이 가려서, 식물이 광합성을 하지 못할 때이다. 따라서, 이산화탄소는 습도가 높은 경우에 상승요인으로 작용하는 것을 확인할 수 있다.

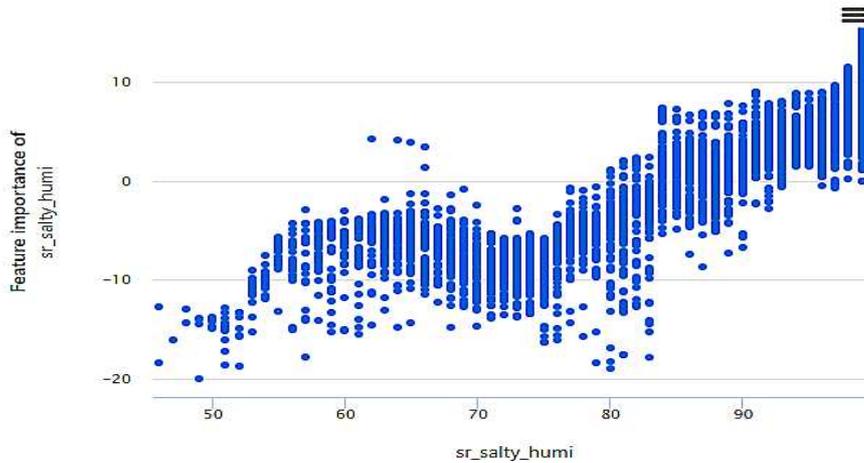


그림 91 Stacking Ensemble 머신러닝으로 적용된 습도와 CO<sub>2</sub>와의 상관관계

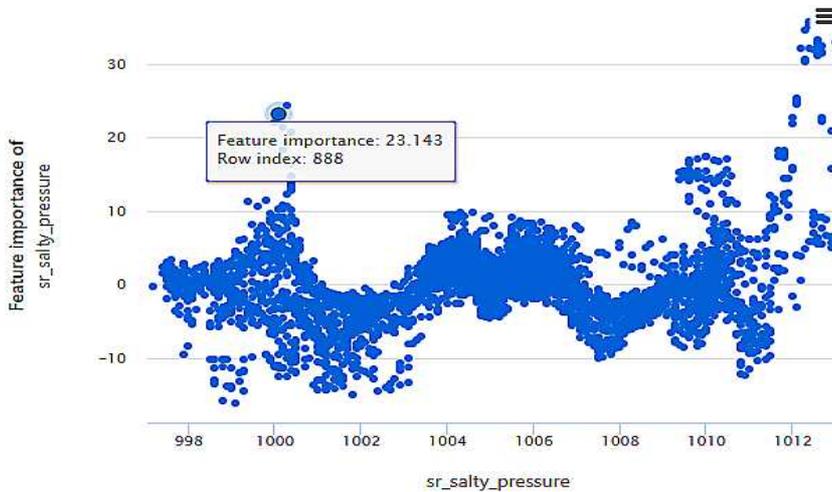


그림 92 Stacking Ensemble 머신러닝으로 적용된 기압과 CO<sub>2</sub>와의 상관관계

그림 92 기압 특성의 경우에는, 다소 높을수록 이산화탄소 발생량에 영향을 많이 주는 것을 알 수 있다. 이와 같은 원인은 고기압은 열돔 현상 등의 원인으로 해석할 수 있는데, 이것은 온도를 상승시켜 폭염 현상을 가져오는 한편, 바람의 정체를 불러일으키고, 결국 주변의 이산화탄소와 공기가 순환하는 것이 차단되어, 탄소의 양이 심하게 증가하는 것으로 추정된다. 또한 북태평양 고기압은, 온도가 높은 것이 아니고 습도가 매우 높은 고온다습형으로, 높은 습도는 이산화탄소 저감에 음의 요소로 작용한다.

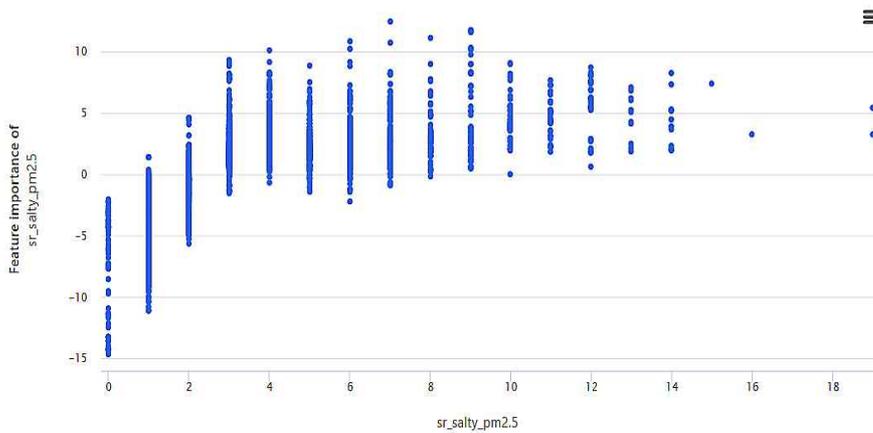


그림 93 Stacking Ensemble 머신러닝으로 적용된 초미세먼지와 CO<sub>2</sub>와의 상관관계

그림 93 초미세먼지는, 이산화탄소의 발생량과 매우 양의 상관관계를 가지고 있다. 즉, 초미세먼지가 많이 발생할수록, 이산화탄소 또한, 같이 증가하는 요소이다. 일반적으로, 실내에서는, 미세먼지 제거를 위해서, 문을 모두 닫고 공기청정기를 가동한다. 그래서, 밀폐된 실내 공간에서 미세먼지는 감소하나, 이산화탄소는 증가하는 역효과가 있다. 하지만, 야외공간의 경우에는, 미세먼지가 광합성에 필요한 태양광선을 차단하여, 식물의 산소 발생을 줄이는 원인으로 해석할 수 있다.

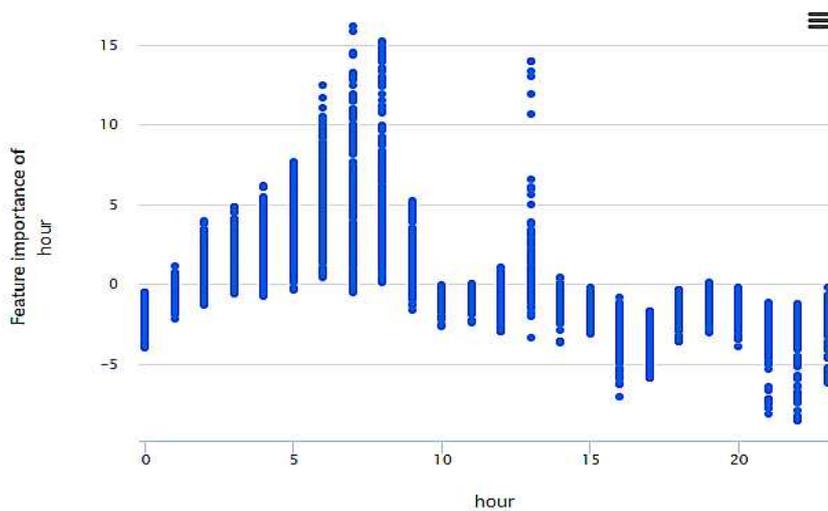


그림 94 Stacking Ensemble 머신러닝으로 적용된 시간대와 CO<sub>2</sub>와의 상관관계

그림 94 시간대 특성의 경우에는, 태양광과 직접적인 영향이 가장 크다. 특히 해가 지는 오후 20시 이후부터 오전 8~9시까지는, 광합성을 멈추게 되어 많은 양의 이산화탄소가 발생하는 원인으로 추정된다. 한편, 태양 직사광선이 한참 작열하는 오전 10시부터 오후 17시까지는 이산화탄소의 양이 적어지는 것을 확인할 수 있다.

### 2.2.1.2 하이퍼 파라미터

본 연구에서 사용된 머신러닝(ML)에 적용되었던 Ensemble 모델, XG Boost regressor, LGBM, RandomForest 알고리즘에서, 각각 데이터 세트를 학습할 때, 하이퍼 파라미터를 자동으로 튜닝하고 적용하여 사용된 최적의 값을 정리하면 다음 표 43과 같다.

표 43 하이퍼 파라미터 정리

Algorithm Name	pre processing	Preprocessing	Algorithms	module	Hyperparameter values	
Stacking Ensemble	StandardScaler	sklearn	XGBoostRegressor	automl	eta	0.4
					gamma	0.1
					max_depth	10
					max_leaves	31
					n_estimators	50
					reg_alpha	0
					reg_lambda	0.625
	StandardScaler	sklearn	XGBoostRegressor	automl	tree_method	auto
					subsample	0.8
					eta	0.5
					gamma	0.01
					max_depth	8
					max_leaves	0
					n_estimators	50
	StandardScaler	sklearn	LightGBMRegressor	automl	reg_alpha	1.354167
					reg_lambda	1.666667
					subsample	1
					tree_method	auto
					learning_rate	0.168423
					max_bin	1023
					max_depth	9
StandardScaler	sklearn	LightGBMRegressor	automl	min_data_in_leaf	0.003226	
				min_split_gain	0.736842	
				n_estimators	100	
				num_leaves	255	
StandardScaler	sklearn	LightGBMRegressor	automl	reg_alpha	0	

					reg_lambda	0.75
					subsample	0.9
					subsample_freq	6
					eta	0.3
					gamma	0
					max_depth	6
					max_leaves	0
					n_estimators	800
					reg_alpha	0
					reg_lambda	2.395833
subsample	0.8					
tree_method	auto					
	MaxAbsScaler	sklearn	XGBoostRegressor	automl	tree_method	auto
XGBoostRegressor	StandardScaler	sklearn	XGBoostRegressor	automl	eta	0.3
					gamma	0
					max_depth	6
					max_leaves	0
					n_estimators	800
					reg_alpha	0
					reg_lambda	2.395833
					subsample	0.8
tree_method	auto					
LightGBM	StandardScaler	sklearn	LightGBMRegressor	automl	learning_rate	0.168423
					max_bin	1023
					max_depth	9
					min_data_in_leaf	0.003226
					min_split_gain	0.736842
					n_estimators	100
					num_leaves	255
					reg_alpha	0
					reg_lambda	0.75
					subsample	0.9
subsample_freq	6					
max_features	0.5					
RandomForest	MaxAbsScaler	sklearn	RandomForestRegressor	sklearn	min_samples_leaf	0.009017
					min_samples_split	0.000899
					n_estimators	10

### 2.2.1.3 모델 평가

인공지능 모델 평가에 사용될 지표는 실제 값과 예측값의 차이를 절댓값으로 변환 뒤 합산하여 평균으로 구하는 MAE, 평균제곱오차에 제곱근을 씌워 구하는 RMSE, 실제값과 예측값의 차이를 제곱한 값의 합(Sum of Squared Residual, SSR)을 실제값과 평균의 차이를 제곱한 것의 합(Sum of Squared Total, SST)으로 나눈 값을 말하며 통상적으로 0에 근접할수록 쓸모없는 회귀식으로

1에 가까울수록 모델이 우수한 성능을 보인다고 평가하는  $R^2$  을 사용하였다.

4가지의 학습모델 중 Stacking Ensemble의  $R^2$  이 0.90으로 가장 우수한 것을 알 수 있으며, 그 뒤로, XGBoostRegressor가 0.88, LightGBM이 0.88, RandomForestRegressor가 0.66의 순으로 결과값을 알 수 있으며, MAE와 RMSE도 같은 순서로 표 44에 기술하여, Ensemble 모델이 여러 학습모형 중에 가장 정확도가 높은 것을 알 수 있었다.

표 44 알고리즘별 평가 결과

Algorithm Name	MAE	RMSE	$R^2$
Stacking Ensemble	5.613527	8.332838	0.9032055
XGBoostRegressor	6.125941	8.980431	0.8877488
LightGBM	6.428201	9.22858	0.8813046
RandomForest	11.39422	15.43259	0.6692627

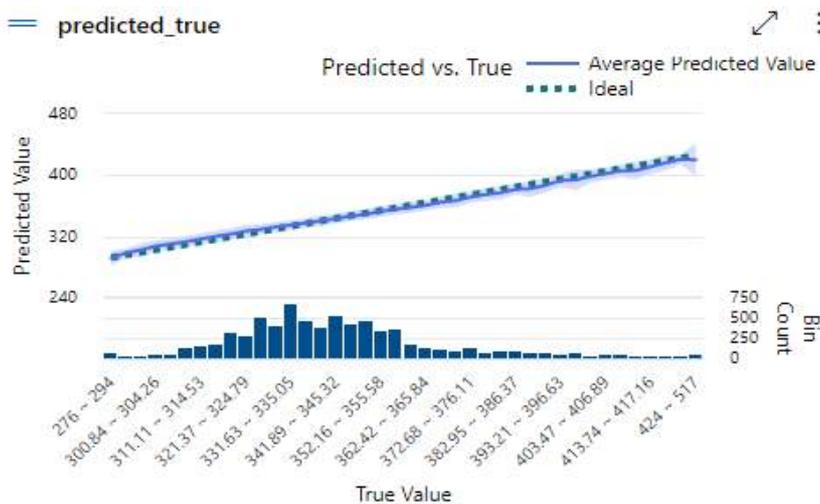


그림 95 예측값과 실제값의 비교 평가 결과

또한 예측값과 실제 값을 비교해보았을 때의 예측과 실제 그래프와 잔차는 그림 95 와 같이 학습이 잘 적용된 것을 알 수 있다. 본 ML 학습에서는 예측값과 실제값이 이상 곡선과 잘 매치되는 것을 확인할 수 있다.

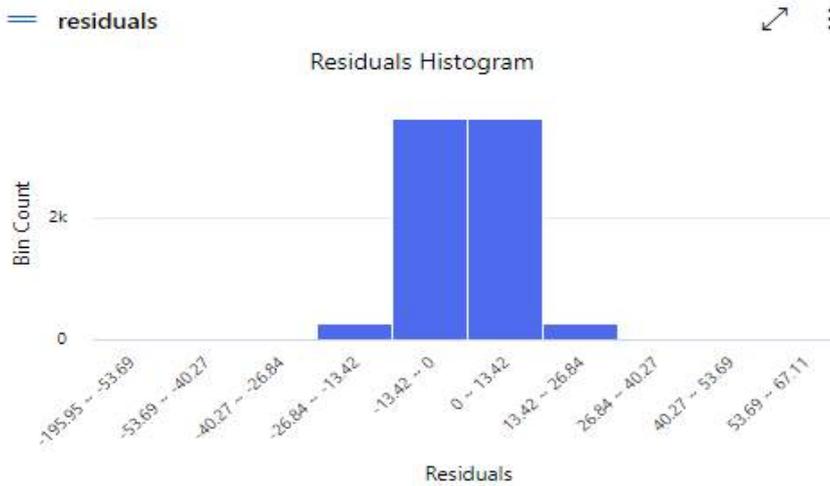


그림 96 알고리즘 잔차 검토 결과

잔차(Residuals)는 예측값과 실제값의 차이로 실제값에서 예측값을 뺀 값이다. 본 머신러닝에서는 0을 중심으로 잔차가 적당한 종의 형태로 분포되어있음을 확인할 수 있으므로, 예측값과 실측값이 균형이 있게 배치된 것을 알 수 있다.

## 2.2.2 갈대숲 머신러닝 모델링 비교

본 회귀분석에 사용된 머신러닝은 AutoML을 사용하였으며, 그중에서 Stacking Ensemble 모델, XG Boost regressor, LGBM, Randomforest의 4가지 모델만을 사용하여, 알고리즘에 대한 정확성 등을 평가하고자 한다. 각각의 알고리즘에 사용된 전처리 모듈과 엔진, 본 처리에 사용된 알고리즘과 엔진은 표 45와 같다

먼저, 첫 번째 Stacking Ensemble Algorithm은 XGboostRegressor 1개와 LGBM 1개, 모두 2개의 알고리즘을 사용하였으며, 모듈은 Scikit Learn과 AutoML을 각각 사용하였다. 전처리에 사용된 스케일러는 각각, Standard

scaler와 MaxAbsScaler등이 활용되었다. 두 번째로, XGboost regressor는 MaxAbsScaler를 전처리로 사용한다. 세 번째로, Light GBM은 MaxAbsScaler를 사용하여 전처리를 진행한다. 끝으로, RandomForestRegressor는 StandardScaler를 사용하여 전처리를 진행한다.

표 45 갈대숲 CO<sub>2</sub> Regression 실험에 적용된 ML 알고리즘과 모듈

Algorithm Name	pre processing	Preprocessing	Algorithms	module
Stacking Ensemble	MaxAbsScaler	sklearn	XGBoostRegressor	automl
	MaxAbsScaler	sklearn	LightGBMRegressor	automl
XGBoostRegressor	MaxAbsScaler	sklearn	XGBoostRegressor	automl
LightGBM	MaxAbsScaler	sklearn	LightGBMRegressor	automl
RandomForestRegressor	StandardScaler	sklearn	RandomForestRegressor	sklearn

### 2.2.2.1 갈대숲 효용성 분석

그림 97 다이어그램은 이 특정 모델을 생성하기 위해 Automated ML이 적용한 데이터 사전 처리, 기능 엔지니어링, 확장 기술 및 머신러닝 알고리즘을 보여준다.

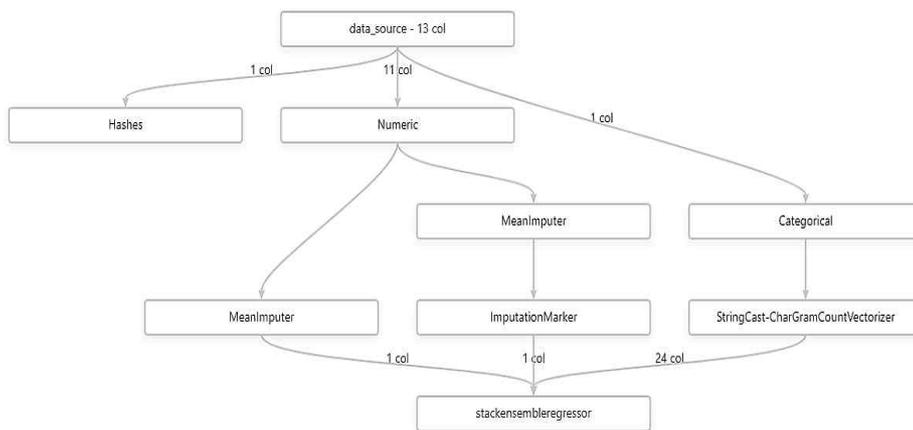


그림 97 분석에 사용된 Voting Algorithm 다이어그램

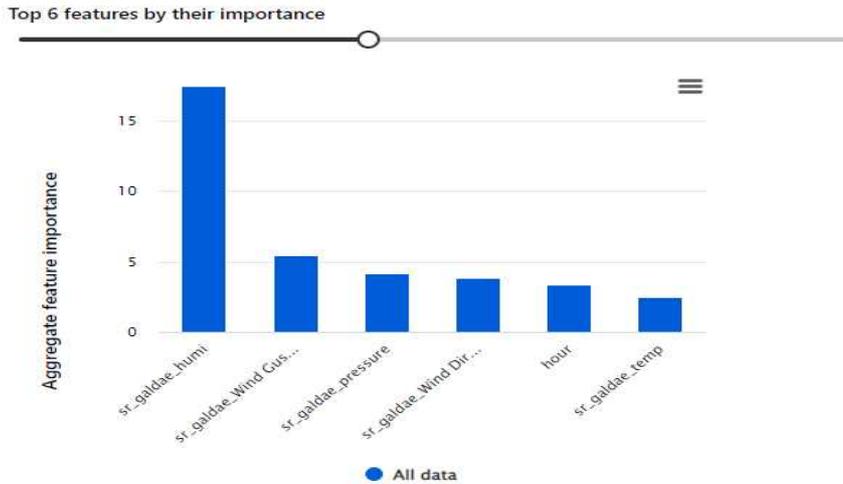


그림 98 Top6 features by their importance

적용된 머신러닝 중에서 가장 높은  $R^2$  값을 갖는, Stacking Ensemble 알고리즘을 분석해보면, 타겟 데이터를 보행량으로 두고, 적용된 데이터에서 상위 6개의 요소를 출력하면, 습도(sr\_galdae\_humi), 돌풍(sr\_galdae\_Wind Gust), 기압(sr\_galdae\_pressure), 풍향(sr\_galdae\_Wind Direction), 시간(hour), 온도(sr\_galdae\_temp)의 순으로 영향을 주는 것을 알 수 있다.

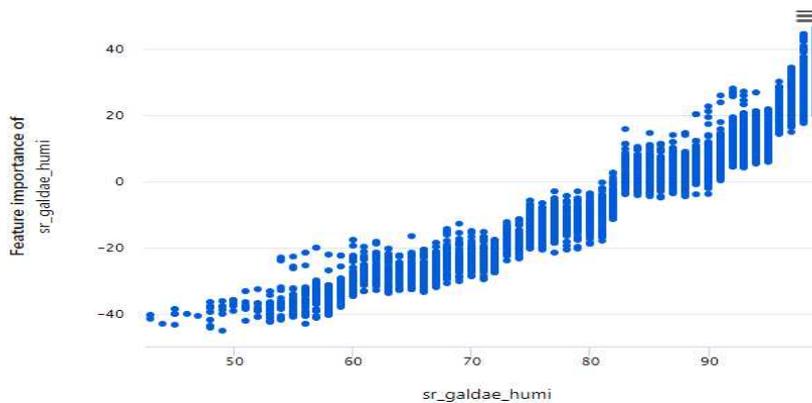


그림 99 Stacking Ensemble 머신러닝으로 적용된 습도와  $CO_2$ 와의 상관관계

습도의 경우에는 이산화탄소와의 Regression 그래프가 양의 기울기를 가지는 것을 알 수 있다. 일반적으로 습도가 높은 경우에는, 비가 오거나 구름에 태양이 가려서, 식물이 광합성을 하지 못할 때이다. 따라서, 이산화탄소는 습도가 높은

경우에 상승요인으로 작용하는 것을 확인할 수 있다. 본 그래프에서 습도와 CO<sub>2</sub>의 Regression은 매우 유의미하게 동작한 것을 알 수 있다.

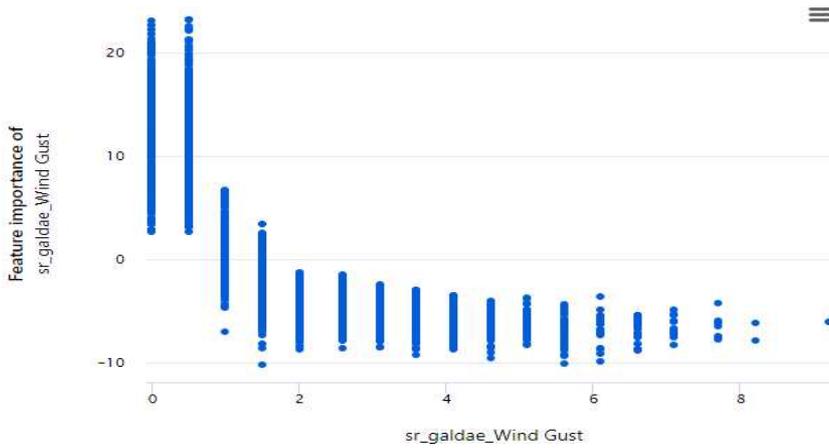


그림 100 Stacking Ensemble 머신러닝으로 적용된 돌풍과 CO<sub>2</sub>와의 상관관계

그림 100에서 나오는 바와 같이 강한 돌풍은 이산화탄소 저감 효과가 매우 큰 것으로 보인다. 상관관계 곡선에서, 이산화탄소는 바람이 강하게 불수록, 음의 기울기로 줄어드는 것을 알 수 있다. 따라서, 도시지역에서 바람길이나 공기 순환을 통해서 이산화탄소 저감이 가능한 것을 알 수 있다.

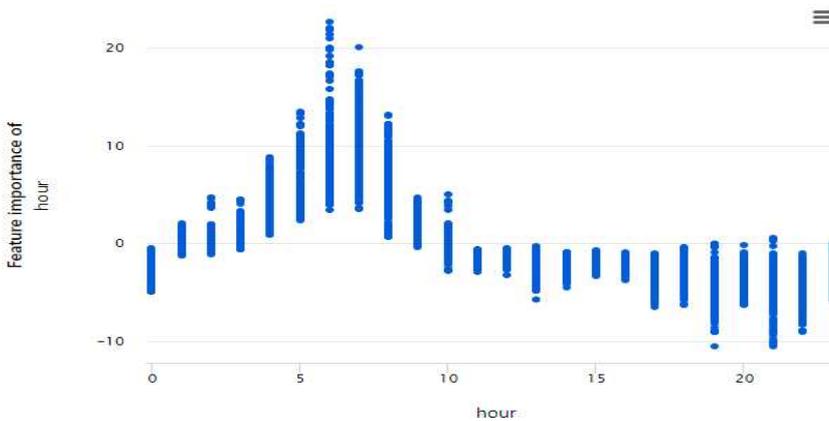


그림 101 Stacking Ensemble 머신러닝으로 적용된 시간대와 CO<sub>2</sub>와의 상관관계

그림 101을 통한 시간대 특성의 경우에는 태양광과 직접적인 영향이 가장 크며 생태공원의 특성상, 차량의 통행이 많은 다른 도시공원과는 매우 다른 형태를 보이는 것을 확인하였다. 특히, 생태습지공원의 특징으로, 해가 진 이후에도, 20시를 넘어 새벽 0시까지도 지속적으로 이산화탄소의 감쇠 효과를 살펴볼 수 있다. 이것은, 도시의 활동이 줄어드는 밤에 많은 경제활동이 멈추면서 이산화탄소의 배출이 줄어드는 것이 원인으로 추정된다. 한편, 남동공단 공장 주변의 생산활동이 시작과 광합성이 되지 않은 새벽 5~7시 시간에는 이산화탄소의 발생량이 매우 많아지는 것을 확인할 수 있다.

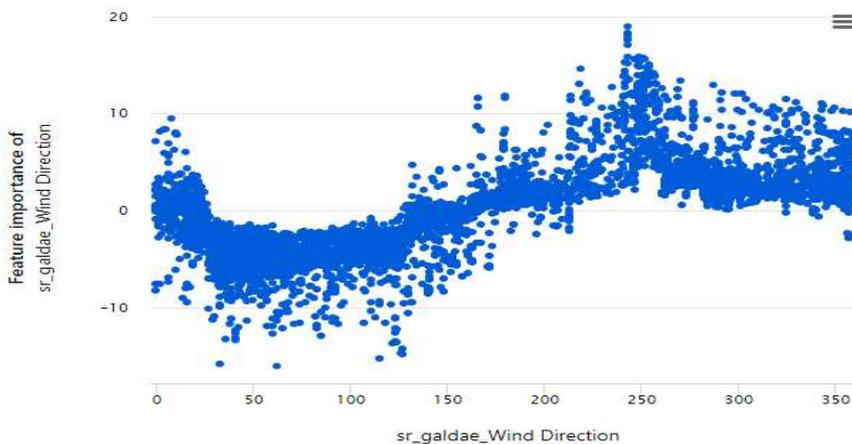


그림 102 Stacking Ensemble 머신러닝으로 적용된 풍향과 CO<sub>2</sub>와의 상관관계

풍향의 경우에는 각도에 따른 풍향을 다음과 같이 정의를 하고 있다. 표 46에서 보는 바와 같이<sup>176)</sup> NNE, NE, ENE, E, ESE, SE, SSE, S etc는 각각 22.5, 45, 67.5, 90, 112.5, 135, 157.5, 180 를 동쪽에서 불어오는 방향, 즉 동풍에 해당한다. 그림 102의 그래프상에서 Feature 중요도(Feature Importance)에서, 서풍일 때 증가 하고, 동풍 방향에서 오는 경우, 수치가 적어지는 것을 알 수 있다. 이는 포구 사거리 기준으로 동풍의 경우에 생태습지공원 방향에서 바람이 불어오는 경우이고, 서풍은 남동공단 등의 도시지역에서 불어오는 바람의 경로로, 공원이 가져오는 탄소 저감 효과에 대한 긍정적인 효과를 반영하는 매우 중요한 요소라 볼 수 있다.

176) 소선섭 외(2000), 『대기관측법』, 교문사, 126.

표 46 풍향각도 일람표

한글	영문약어	방위각 <sup>o</sup>	밀리각법
		(도각법)	(Mils)
북	N	0/360	0/6400
북북동	NNE	22.5	400
북동	NE	45	800
동북동	ENE	67.5	1200
동	E	90	1600
동남동	ESE	112.5	2000
남동	SE	135	2400
남남동	SSE	157.5	2800
남	S	180	3200
남남서	SSW	202.5	3600
남서	SW	225	4000
서남서	WSW	247.5	4400
서	W	270	4800
서북서	WNW	292.5	5200
북서	NW	315	5600
북북서	NNW	337.5	6000

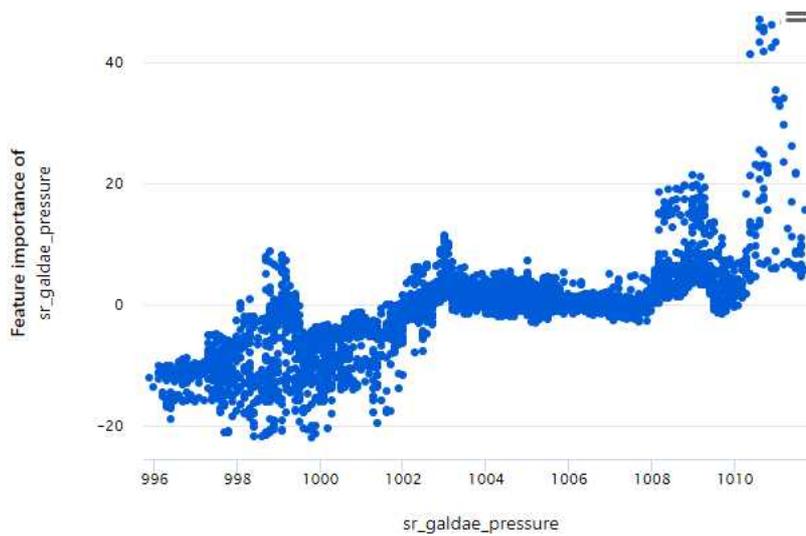


그림 103 Stacking Ensemble 머신러닝으로 적용된 기압과 CO<sub>2</sub>와의 상관관계

그림 103 기압 특성의 경우에는 다소 높을수록 이산화탄소 발생량에 영향을 많이 주는 것을 알 수 있다. 이와 같은 원인은 고기압은 열돔 현상 등의 원인으로

해석할 수 있는데, 이것은 온도를 상승시켜 폭염 현상을 가져오는 한편, 바람의 정체를 불러일으키고, 결국 주변의 이산화탄소와 공기가 순환하는 그것이 차단되어, 탄소의 양이 심하게 증가하는 것으로 추정된다. 또한 북태평양 고기압은, 온도가 높은 것이 아니고 습도가 매우 높은 고온다습형으로, 높은 습도는 이산화탄소 저감에 음의 요소로 작용한다.

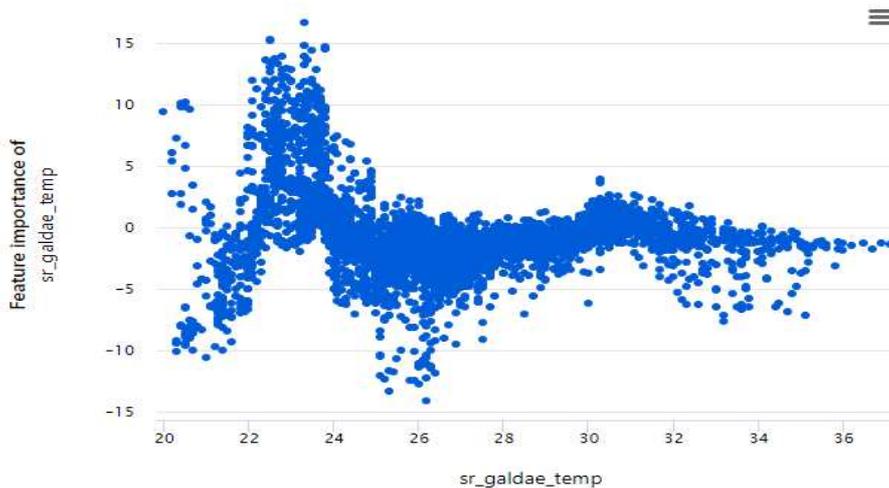


그림 104 Stacking Ensemble 머신러닝으로 적용된 온도와 CO<sub>2</sub>와의 상관관계

그림 104의 기온과 이산화탄소와의 상관관계를 그래프로 살펴보면, 온도의 경우에는, 22°C에서 23°C까지는 양의 기울기를 가지고 있어서, 이산화탄소 발생이 증가하나 23°C에서 26°C까지는 이산화탄소 저감에 효과가 큰 것을 알 수 있다. 이것은 야외에서 태양광에 노출로 인한 온도가 상승함에 따라서 갈대 식물의 광합성이 활발해서 이산화탄소 저감이 되는 것으로 유추된다.

### 2.2.2.2 하이퍼 파라미터

하이퍼 파라미터는 학습을 시키는 컴퓨터의 하드웨어의 성능과 밀접한 연관이 있으므로 컴퓨터 환경에 맞게 설정해주어야 한다.<sup>177)</sup>. 본 연구에서 사용된 머신러닝(ML)에 적용되었던 Ensemble 모델, XG Boost regressor, LGBM,

177) 손채준 외(2021), 「상태 표현 방식에 따른 심층 강화 학습 기반 캐릭터 제어기의 학습 성능 비교」, 『한국컴퓨터그래픽스학회 논문지』, 27(5), 55-61.

Randomforest 알고리즘에서, 각각 데이터 세트를 학습할 때, 하이퍼 파라미터를 자동으로 튜닝하고 적용하여 사용된 최적의 값을 정리하면 다음 표 47과 같다.

표 47 하이퍼 파라미터 정리

Algorithm Name	pre processing	Preprocessing	Algorithms	module	Hyperparameter values	
Stacking Ensemble	MaxAbsScaler	sklearn	XGBoostRegressor	automl	tree_method	auto
	MaxAbsScaler	sklearn	LightGBMRegressor	automl	min_data_in_leaf	20
XGBoostRegressor	MaxAbsScaler	sklearn	XGBoostRegressor	automl	tree_method	auto
LightGBM	MaxAbsScaler	sklearn	LightGBMRegressor	automl	min_data_in_leaf	20
RandomForestRegressor	StandardScaler	sklearn	RandomForestRegressor	sklearn	max_features	0.5
					min_samples_leaf	0.002863
					min_samples_split	0.001073
					n_estimators	25

### 2.2.2.3 모델 평가

인공지능 머신러닝 4가지의 학습모델 중 Stacking Ensemble의  $R^2$  0.918로 가장 우수한 것을 알 수 있으며, 그 뒤로, XGBoostRegressor가 0.915, LightGBM이 0.89, RandomForestRegressor가 0.80의 순으로 결과값을 알 수 있으며, MAE와 RMSE도 같은 순서로 표 48에 기술하여, Ensemble 모델이 여러 학습모형 중에 가장 정확도가 높은 것을 알 수 있었다.

표 48 알고리즘별 평가 결과

Algorithm Name	MAE	RMSE	$R^2$
Stacking Ensemble	6.931462	9.492528	0.9180315
XGBoostRegressor	6.995154	9.627101	0.9157135
LightGBM	7.777669	10.49587	0.8998283
RandomForestRegressor	10.24548	14.66115	0.8044016

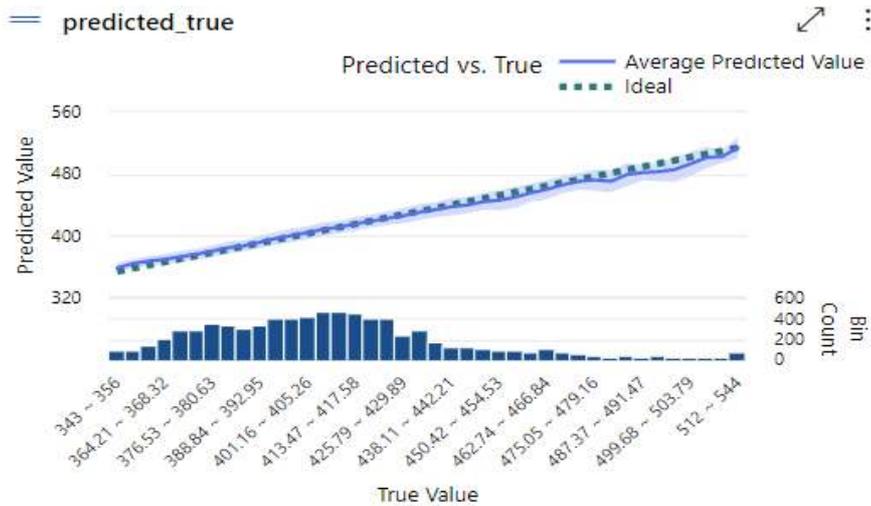


그림 105 예측값과 실제값의 비교 평가 결과

또한, 예측값과 실제 값을 비교해보았을 때의 그래프와 잔차는 그림 105, 106 과 같이 학습이 잘 적용된 것을 알 수 있다. 본 ML 학습에서는 예측값과 실제값이 이상 곡선과 잘 매치되는 것을 확인할 수 있다.

본 머신러닝에서는 0을 중심으로 잔차가 적당한 종의 형태로 분포되어있음을 확인할 수 있으므로, 예측값과 실측값이 균형이 있게 배치된 것을 알 수 있다.

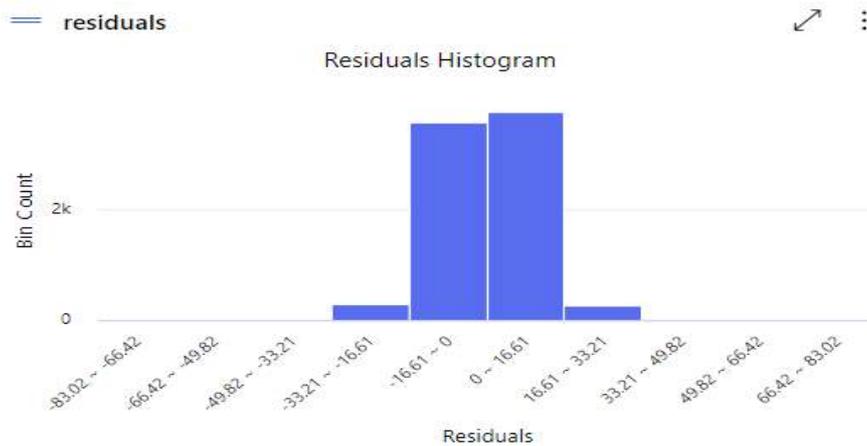


그림 106 알고리즘 잔차 검토 결과

### 2.2.3 포구 사거리 머신러닝 모델링 비교

소래 습지 생태공원의 환경데이터와 도심지 내 대조군으로서, 소래포구사거리를 선택하였으며, 소래포구 도로에서의 환경 요소는 도시 도로 여건상, 풍향 풍속 계 등의 추가 센서 설치가 어려워서, 근거리에서 측정된 갈대밭의 풍향, 풍속, 조도, 우적량, 돌풍 등의 환경 데이터를 차용하여 사용하였다. 본 회귀분석에 사용된 머신러닝은 AutoML을 사용하였으며, 그중에서 Stacking Ensemble 모델, LGBM, XG Boost regressor, Randomforest의 4가지 모델만을 사용하여, 알고리즘에 대한 정확성 등을 평가하고자 한다. 각각의 알고리즘에 사용된 전처리 모듈과 엔진, 본 처리에 사용된 알고리즘과 엔진은 표 49와 같다

먼저, 첫 번째 Stacking Ensemble Algorithm은 XGboostRegressor 3개와 LGBM 1개의 전체 4가지 알고리즘을 사용하였으며, 모듈은 Scikit Learn과 AutoML을 각각 사용하였다. 전처리에 사용된 스케일러는 각각, Standard scaler와 MaxAbsScaler등이 활용되었다.

두 번째로, LightGBMRegressor는 StandardScaler를 전처리로 사용한다.

세 번째로, XGBoostRegressor는 StandardScaler를 사용하여 전처리를 진행한다.

끝으로, RandomForestRegressor는 MaxAbsScaler를 사용하여 전처리를 진행한다.

표 49 포구사거리 CO<sub>2</sub>Regression 실험에 적용된 ML 알고리즘과 모듈

Algorithm Name	pre processing	Preprocessing module	Algorithms	module
Stacking Ensemble	StandardScaler	sklearn	XGBoostRegressor	automl
	StandardScaler	sklearn	XGBoostRegressor	automl
	StandardScaler	sklearn	LightGBMRegressor	automl
	MaxAbsScaler	sklearn	XGBoostRegressor	automl
LightGBM	StandardScaler	sklearn	LightGBMRegressor	automl
XGBoostRegressor	StandardScaler	sklearn	XGBoostRegressor	automl
RandomForestRegressor	MaxAbsScaler	sklearn	RandomForestRegressor	sklearn

### 2.2.3.1 포구 사거리 효용성 분석

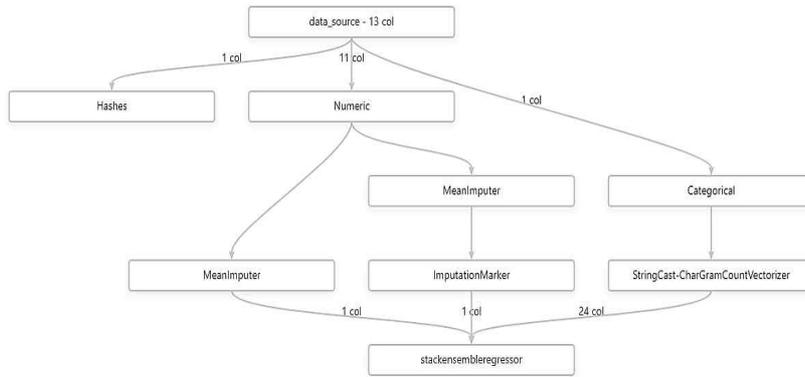


그림 107 분석에 사용된 Voting Algorithm 다이어그램

그림 107 다이어그램은 AI 모델을 생성하기 위해 Automated ML이 적용한 데이터 사전 처리, 기능 엔지니어링, 확장 기술 및 머신러닝에 관한 것이다.

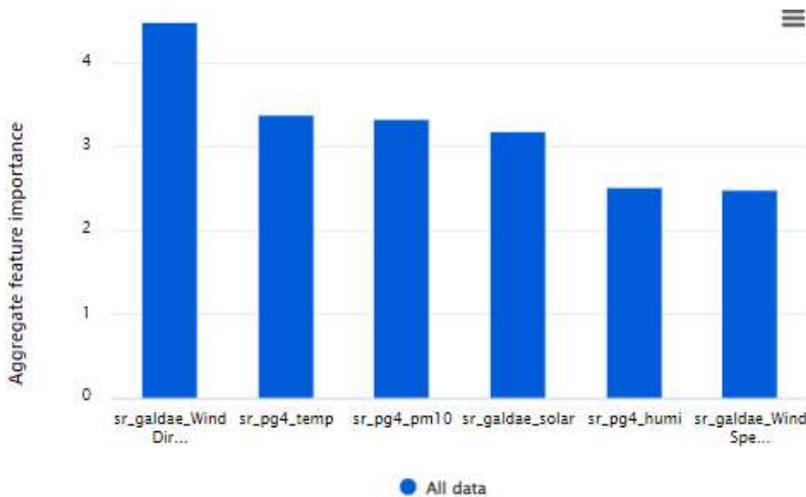


그림 108 Top6 features by their importance

그림 108에서 적용된 머신러닝 중에서 가장 높은  $R^2$ 값을 갖는, 앙상블 알고리즘을 분석해보면, 타겟 데이터를 보행량으로 두고, 적용된 데이터에서 상위 6개의 요소를 출력하면, 갈대숲 풍향(sr\_galdae\_Wind Direction), 온도(sr\_pg4\_temp), 포구 사거리 미세먼지(sr\_pg4\_pm10), 갈대숲 태양조도

(sr\_galdae\_solar), 포구 사거리 습도(sr\_pg4\_humi), 갈대숲 풍속 (sr\_galdae\_Wind Speed)의 순으로 영향을 주는 것을 알 수 있다.

포구 사거리의 이산화탄소와 갈대숲의 풍향의 regression 중요도의 특징은 기존 갈대숲과 매우 유사한 패턴을 보인다(그림 109). 풍향이 서풍이면 이산화탄소가 증가하고, 동풍 방향에서 오는 경우, 수치가 적어지는 것을 알 수 있다. 이는 포구 사거리 기준으로 동풍의 경우에 생태습지공원 방향에서 바람이 불어오는 경우이고, 서풍은 남동공단 등의 도시지역에서 불어오는 바람의 경로로, 공원이 가져오는 탄소 저감 효과에 대한 긍정적인 효과를 반영하는 매우 중요한 요소라 볼 수 있다.

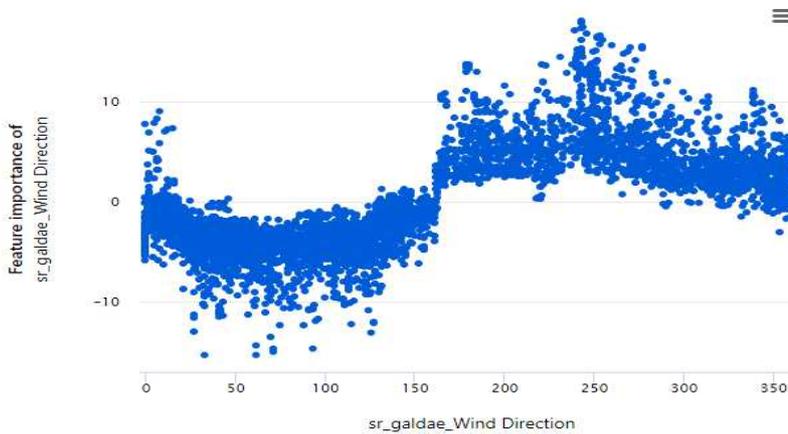


그림 109 Stacking Ensemble 머신러닝으로 적용된 풍향과 CO<sub>2</sub>와의 상관관계

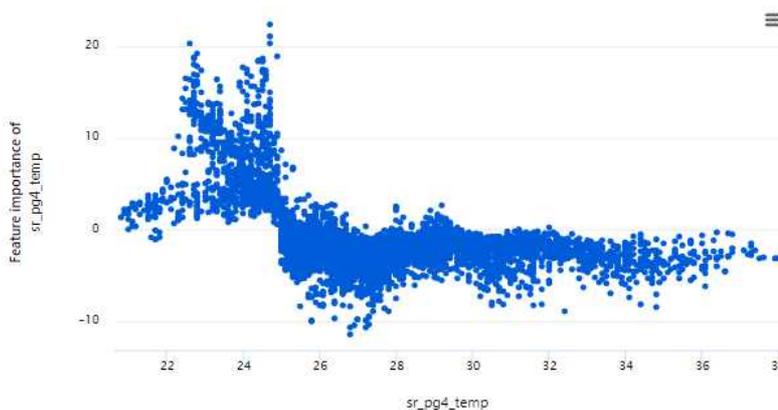


그림 110 Stacking Ensemble 머신러닝으로 적용된 기온과 CO<sub>2</sub>와의 상관관계

포구 사거리에서, 기온과 이산화탄소와의 상관관계를 그래프로 살펴보면, 온도의 경우에는, 상온의 경우 상대적으로 저온인 날씨에는 이산화탄소의 양이 증가하지만, 온도가 상승함에 따라서 음의 기울기를 가지고 있어서, 이산화탄소 저감에 효과가 있는 것을 알 수 있다(그림 110). 이것은 소래생태공원보다는 기울기가 적은 편이기는 하지만, 이산화탄소가 저감에는 도움이 되는 것으로 확인된다.

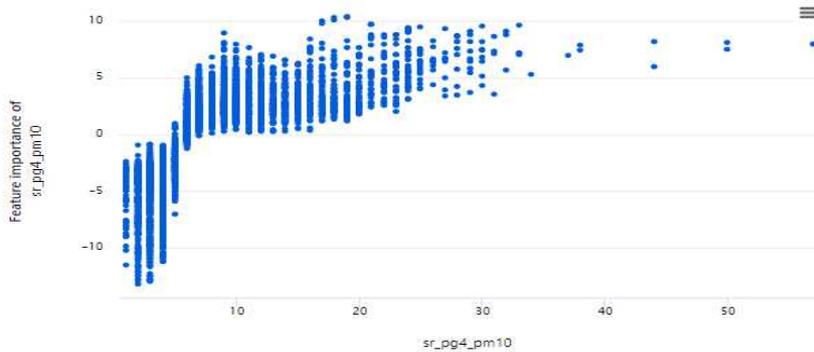


그림 111 Stacking Ensemble 머신러닝으로 적용된 미세먼지와 CO<sub>2</sub>와의 상관관계

미세먼지는, 이산화탄소의 발생량과 매우 양의 상관관계를 가지고 있다(그림 111). 미세먼지가 많이 발생할수록, 이산화탄소 또한, 같이 증가하는 요소이다. 야외공간의 경우에는, 미세먼지가 광합성에 필요한 태양광선을 차단하여, 식물의 산소 발생을 줄이는 원인으로 해석할 수 있다.

그림 112는 태양광의 밝기인 조도와 이산화탄소의 관계를 보여준다. 태양광과 포구 사거리 이산화탄소와의 상관관계는, 특히 조도가 0인 밤인 경우보다는, 흐릿한 날씨, 또는 새벽녘 저녁 같은 태양광이 약한 상황에서 발생량의 상관관계가 매우 높은 것을 알 수 있다. 조도가 높아짐에 따라서, 점차 이산화탄소가 저감되는 음의 요소로 작용할 만한 영향력이 크다.

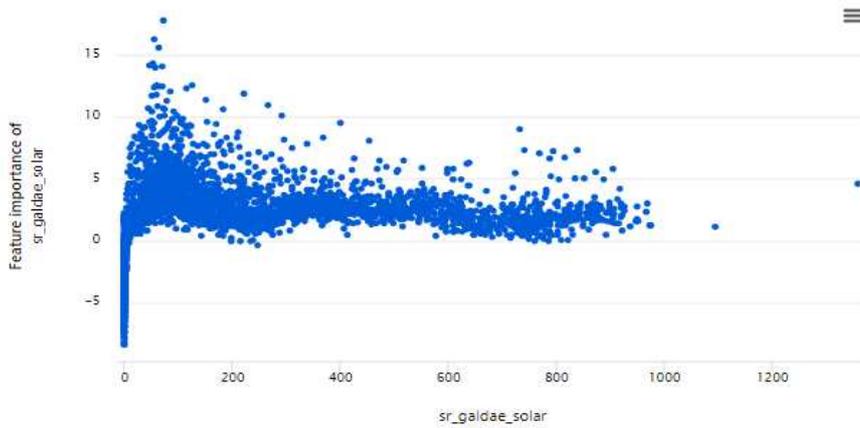


그림 112 Stacking Ensemble 머신러닝으로 적용된 조도와 CO<sub>2</sub>와의 상관관계

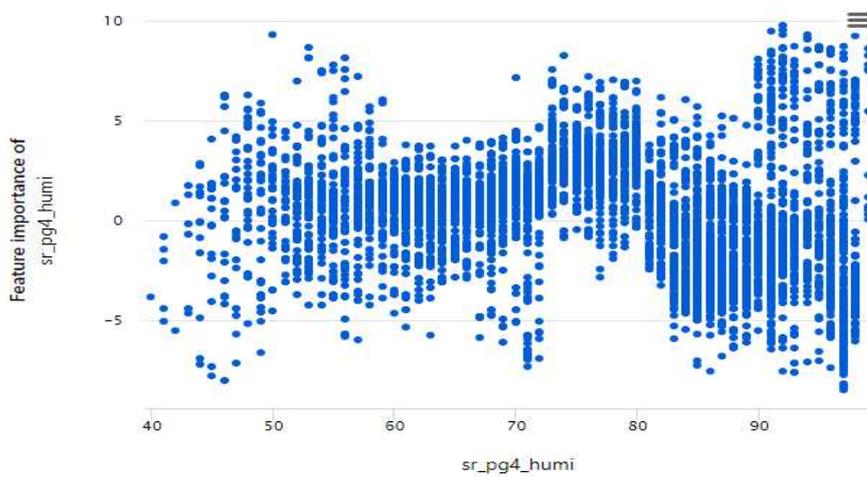


그림 113 Stacking Ensemble 머신러닝으로 적용된 습도와 CO<sub>2</sub>와의 상관관계

포구 사거리의 습도의 경우에는 생태공원과 비교하면 매우ダイナミック한 영향력을 볼 수 있다(그림 113). 이산화탄소 발생 상관관계의 레인지가 상당히 광범히 하여, 습도에 따른 이산화탄소 발생량에 대한 regression 경향을 찾기는 어려우나, 많은 부분에서 영향력이 있음을 명백히 알 수 있다.

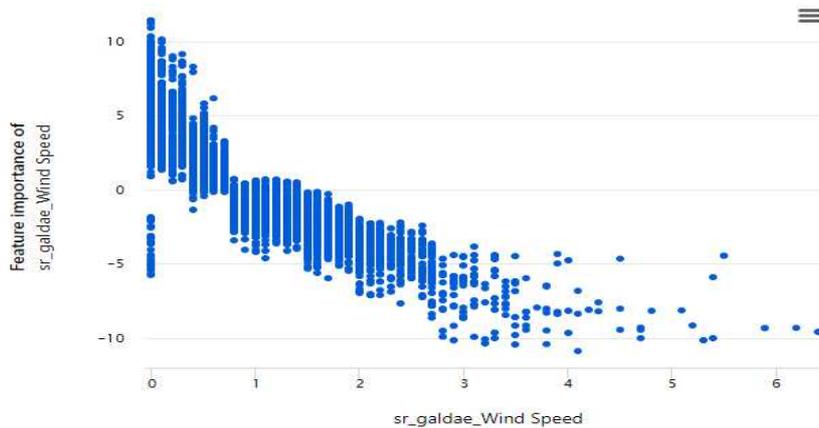


그림 114 Stacking Ensemble 머신러닝으로 적용된 풍속과 CO<sub>2</sub>와의 상관관계

포구 사거리의 이산화탄소에 대한 갈대숲의 풍속 영향력은 매우 큰 음의 기울기를 갖는다(그림 114). 따라서 도시 기후 변화에서 공원이 갖는 이산화탄소에 대한 방재 효과는 매우 크다는 것을 여실히 증명해 보이는 사례이다.

### 2.2.3.2 하이퍼 파라미터

하이퍼 파라미터는 앞에서 언급했듯이 학습을 시키는 컴퓨터의 하드웨어의 성능과 밀접한 연관이 있으므로 컴퓨터 환경에 맞게 설정해주어야 한다<sup>178)</sup>. 본 연구에서 사용된 머신러닝(ML)에 적용되었던 Ensemble 모델, LGBM, XG Boost regressor, Randomforest 알고리즘에서, 각각 데이터 세트를 학습할 때, 하이퍼 파라미터를 자동으로 튜닝하고 적용하여 사용된 최적의 값을 정리하면 다음 표 50과 같다.

178) 손채준 외(2021), 「상태 표현 방식에 따른 심층 강화 학습 기반 캐릭터 제어기의 학습 성능 비교」, 『한국컴퓨터그래픽스학회 논문지』, 27(5), 55-61.

표 50 하이퍼 파라미터 정리

Algorithm Name	pre processing	Preprocessing	Algorithms	module	Hyperparameter values	
Stacking Ensemble	StandardScaler	sklearn	XGBoostRegressor	automl	eta	0.3
					gamma	0
					max_depth	6
					max_leaves	0
					n_estimators	100
					reg_alpha	1.666666667
					reg_lambda	1.979166667
	subsample	0.5				
	StandardScaler	sklearn	XGBoostRegressor	automl	eta	0.5
					gamma	0.01
					max_depth	8
					max_leaves	0
					n_estimators	50
reg_alpha					1.354166667	
reg_lambda					1.666666667	
subsample	1					
StandardScaler	sklearn	LightGBMRegressor	automl	learning_rate	0.168422632	
				max_bin	1023	
				max_depth	9	
				min_data_in_leaf	0.003226393	
				min_split_gain	0.736842105	
				n_estimators	100	
				num_leaves	255	
				reg_alpha	0	
reg_lambda	0.75					
MaxAbsScaler	sklearn	XGBoostRegressor	automl	tree_method	auto	
LightGBM	StandardScaler	sklearn	LightGBMRegressor	automl	learning_rate	0.168422632
					max_bin	1023
					max_depth	9
					min_data_in_leaf	0.003226393
					min_split_gain	0.736842105
					n_estimators	100
					num_leaves	255
					reg_alpha	0
					reg_lambda	0.75
					subsample	0.9
subsample_freq	6					
XGBoost Regressor	StandardScaler	sklearn	XGBoostRegressor	automl	colsample_bytree	0.9
					eta	0.5
					gamma	0.01
					max_depth	8
					max_leaves	0
					n_estimators	50
					reg_alpha	1.354166667
					reg_lambda	1.666666667

					subsample	1
					tree_method	auto
RandomForestRegressor	MaxAbsScaler	sklearn	RandomForestRegressor	sklearn	max_features	sqrt
					min_samples_leaf	0.006151579
					min_samples_split	0.000753222
					n_estimators	10

### 2.2.3.3 머신러닝 모델 평가

인공지능 모델 평가를 진행하여 4가지의 학습모델 중 Stacking Ensemble의  $R^2$  는 0.85로 가장 우수한 것을 알 수 있으며, 그 뒤로, XGBoostRegressor이 0.83, LightGBM이 0.79, RandomForestRegressor가 0.52의 순으로 결과 값을 알 수 있다. MAE와 RMSE도 같은 순서로 표 51에 기술하여, Ensemble 모델이 여러 학습모형 중에 가장 정확도가 높은 것을 알 수 있었다.

표 51 알고리즘별 평가 결과

Algorithm Name	MAE	RMSE	$R^2$
Stacking Ensemble	5.644986	7.761528	0.8522429
LightGBM	6.007524	8.303173	0.8309693
XGBoostRegressor	6.622258	9.069654	0.7979506
RandomForestRegressor	10.56714	13.97395	0.5213627



그림 115 예측값과 실제값의 비교 평가 결과

그림 115와 같이 예측값과 실제 값을 비교해보았을 때의 그래프와 잔차는 학습이 잘 적용된 것을 알 수 있다. 본 ML 학습에서는 예측값과 실제값이 이상 곡선과 잘 매치되는 것을 확인할 수 있다.

잔차(Residuals)는 예측값과 실제값의 차이로 실제값에서 예측값을 뺀 값이다. 본 머신러닝에서는 0을 중심으로 잔차가 적당한 종의 형태로 분포되어있음을 확인할 수 있으므로, 예측값과 실제값이 균형이 있게 배치된 것을 알 수 있다(그림 116).

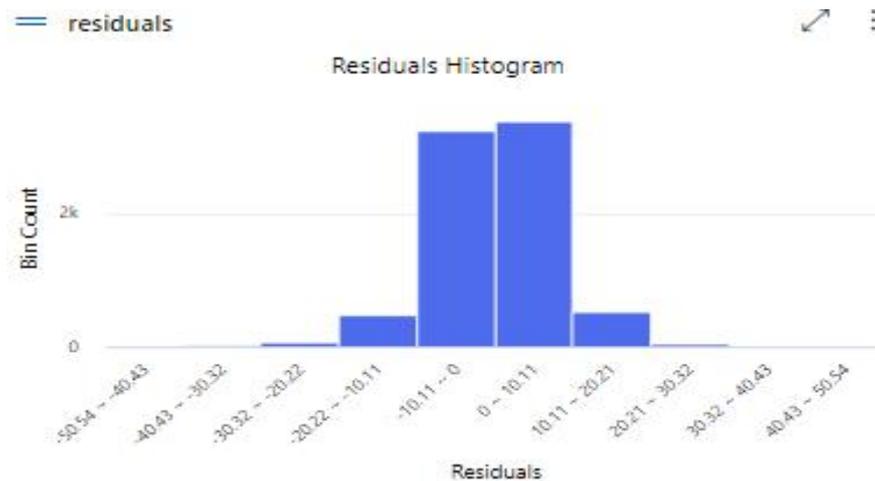


그림 116 알고리즘 잔차 검토 결과

## 2.3 소결

이 연구에서는 인천 소래생태공원의 ‘갈대숲’ 과 ‘염습지’ 에서 폭염기 환경 데이터를 수집·이용, 도심지인 ‘포구 사거리’ 에서 수집한 온·습도 등의 환경 데이터와 비교하면서, 이산화탄소의 발생량에 영향을 주는 요소들과 사이트별 차이를 분석하였다. 일반적인 데이터와 범주 데이터, 시계열 데이터와의 분석은 파이선과 구글 코랩을 사용하여, 각종 그래프로 도식화하여, 현상을 쉽게 살펴볼 수 있도록 하였다. 사이트별로 온, 습도, 미세먼지, 초미세먼지 그리고 이산화탄소의 차이를 비교하여, 염습지의 온도감소 효과, 탄소 저감 효과 등을 살펴보았다.

그 결과 도로부인 ‘포구 사거리’ 의 이산화탄소 발생량은 소래생태공원 ‘염습지’ 에 비해서 평균 40ppm, 최고 60ppm 정도 높게 측정되었고, 기온은 ‘포구 사거리’ 가 시간과 날짜와 관계없이 일정하게 약 1.5 °C 정도 높게 유지되었다. 초미세먼지(pm2.5)의 비교에서, 도시에 위치한 ‘포구 사거리’ 는 매일 평균 ‘갈대숲’ 보다는 약  $3\mu\text{g}/\text{m}^3$  정도를 높게 유지하였고, ‘염습지’ 대비  $4\mu\text{g}/\text{m}^3$  높게 나타났다. 미세먼지의 비교에서는 도시에 위치한 ‘포구 사거리’ 는 매일 평균 ‘갈대숲’ 보다는 약  $3.5\mu\text{g}/\text{m}^3$  정도 높게 유지하였고, ‘염습지’ 대비는  $4\mu\text{g}/\text{m}^3$ 의 차이를 유지하였다.

습도 비교에서 도시에 있는 포구 사거리는 매일 평균 갈대숲, 염습지보다는 약 5% 정도를 낮게 유지되었다. 이어서, 기후 변화 대응을 위하여 환경데이터와 이산화탄소 발생의 상관관계를 분석하기 위해, Auto ML 머신러닝 모델을 이용하였다. 여름철 폭염 기간에 수집된 환경 빅데이터는 전처리를 거쳐, 사이트별 유의미한 데이터로 구성하였다. 인공지능분석을 통한 머신러닝 알고리즘별 예측 성능을 비교하고 그중에 가장 우수한 성능을 보인 Stacking Ensemble을 통해서 분석된 상관관계가 높은 환경 요소와 그의 패턴을 염습지, 갈대숲, 그리고 포구 사거리로 나누어 Regression 분석을 수행하였다.

그 결과 ‘염습지’ 에서는 온도(sr\_salty\_temp), 돌풍(sr\_salty\_Wind\_Gust), 습도(sr\_salty\_humi), 기압(sr\_salty\_pressure), 초미세먼지(sr\_salty\_pm2.5), 시간(hour)의 6개 인자의 순서로 이산화탄소 발생량에 가장 큰 영향을 주는 것을 인지하고, 적용된 Stacking Ensemble의 결정계수인  $R^2$  이 0.90으로 가장 우수한 것을 알 수 있으며, XGBoostRegressor이 0.88,

LightGBM이 0.88, RandomForestRegressor가 0.66의 순으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘에 대한 결론을 내릴 수 있었다.

‘갈대숲’에서는, 습도(sr\_galdae\_humi), 돌풍(sr\_galdae\_Wind Gust), 기압 (sr\_galdae\_pressure), 풍향(sr\_galdae\_Wind Direction), 시간 (hour), 온도 (sr\_galdae\_temp)의 6개 인자가 이산화탄소 발생량에 가장 큰 영향을 주는 것을 인지하고, 적용된 Stacking Ensemble의  $R^2$ 이 0.918로 가장 우수한 것을 알 수 있으며, XGBoostRegressor이 0.915, LightGBM이 0.89, RandomForestRegressor가 0.80의 순으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘에 대한 결론을 내릴 수 있었다.

도심부인 ‘포구 사거리’에서는, 도심부내의 환경센서 설치의 제약으로, 일부 갈대숲의 데이터를 차용하여 사용한 결과, 갈대숲 풍향(sr\_galdae\_Wind Direction), 온도(sr\_pg4\_temp), 포구 사거리 미세먼지(sr\_pg4\_pm10), 갈대숲 태양조도(sr\_galdae\_solar), 포구 사거리 습도(sr\_pg4\_humi), 갈대숲 풍속(sr\_galdae\_Wind Speed)의 순서로 6개 인자가 이산화탄소 발생량에 가장 큰 영향을 주는 것으로 확인되었다. 적용된 Stacking Ensemble의  $R^2$ 이 0.85로 가장 우수한 것을 알 수 있다. 뒤이어 XGBoostRegressor이 0.83, LightGBM이 0.79, RandomForestRegressor가 0.52의 순으로 인공지능 학습모델에 대한 평가를 통해서 최적의 알고리즘을 구할 수 있었다. 따라서 향후 스마트공원의 방재기능으로서 환경데이터를 활용한 이산화탄소 발생과 저감 예측에 연구에서 확인된 최적의 Stacking Ensemble 머신러닝 모델을 사용하여 정량적인 예측과 응용에 사용할 수 있을 것으로 기대된다.

한편 본 연구를 통해서 이산화탄소에 영향을 주는 주요 요소들을 살펴보았을 때 대한 태양광, 온도 등 식물의 광합성을 주는 인자들과 돌풍, 풍향, 풍속 등에 의한 영향이 큰 것으로 확인되었다. 특히 풍향의 경우에는 포구 사거리를 기준으로 보았을 때, 서쪽에 있는 남동 공단 측에서부터 불어오는 서풍은 이산화탄소를 증가시키고, 소래생태공원이 위치한 동쪽에서 불어오는 동풍의 경우에는 이산화탄소를 줄이는 효과가 큰 것으로, 상관 관계 분석이 되어, 향후 추가적인 연구가 더 필요할 것으로 생각된다.

## V. 결론

### 1 연구의 요약

#### 1.1 연구의 요약과 결론

지금까지의 논의를 제2장에서 제시한, 스마트 공원의 정의에 바탕을 두고 살펴보면 다음과 같다. 첫째 기술적 차원 차원에서 서울숲 스마트공원, 소래생태 스마트공원은 본 연구를 중심으로 IoT 환경을 위한 무선 네트워크의 적용방식을 제안하여 환경 데이터 측정 기기를 개발하고 24시간 공원의 환경을 장기간 관리하고 예측하였다. 둘째 관리적 차원에서 서울숲 스마트공원의 경우 요일별, 계절별, 휴일 등 시계열과 관련하여 공원의 사용 시간, 장소 활용도 등을 자세히 파악할 수 있었다. 셋째 사용자 차원에서는 코로나 시기에 실내에만 머물러 야외활동을 자제하던 공원 방문객들이 오픈스페이스인 공원에서 건강한 일과 삶의 균형을 찾을 수 있음을 사회적 요소인 코로나 발생자 등의 사회적 공공데이터와 비교하여 서울숲 스마트공원 이용에 대한 빅데이터 수집을 통해서 확인할 수 있었다. 넷째 기후변화 대응 차원에서, 서울숲 스마트공원의 온도, 습도, 미세먼지, 일산화탄소 등의 계절, 요일, 시간별 환경 변화와 보행량의 상관관계를 분석하여 인공지능 머신러닝을 통한 기후변화 대응 데이터 학습과 예측의 모델링을 성능 비교할 수 있었다. 소래생태 스마트공원의 경우에는 탄소 저감, 온도 저감, 미세먼지 저감 등의 효과를 확실히 알 수 있었으며 인공지능을 통한 예측의 가능성 또한 증명할 수 있었다.

본 연구는 스마트공원의 기후변화에 대한 대응의 일환으로 첫 번째 스마트공원 구현인 서울숲공원에 2021년 5월 21일부터 2022년 1월 26일까지 직접 제작한 IoT 센서를 적용하여, 연간 대기 요소와 공원 방문자에 대한 자료를 수집하였다. 이에 따른 상관관계를 분석하고자, 관련 모니터링을 시행하였다. 입수된 스마트공원의 빅데이터를 처리하고, 서울숲공원의 환경 데이터가 공원 이용 빈도의 근거인 보행량의 상관관계를 분석하기 위해, 최신 인공지능 기술인, Auto ML 머신러닝 모델을 이용하였다. 특히 기존에 공공데이터로서 공급되고 있는 시, 군, 구 단위의 대기 정보와 차별화를 두기 위해 서울숲에서도, 특히 가족마당의 마이크로 환경

데이터인 온도, 습도, 이산화탄소, 미세먼지 등을 직접 개발한 IoT 장비와 그곳에 보행량을 측정하는 IoT 장비를 같이 설치하여, 데이터를 장기간에 걸쳐 수집하였다. 일반적인 데이터와 범주 데이터, 시계열 데이터와의 분석은 파이선과 구글 코랩을 사용하여, 각종 그래프로 도식화하고, 대상지의 상태를 쉽게 살펴볼 수 있도록 하였다. 이후 다시 인공지능분석을 통한 머신러닝 알고리즘별 예측 성능을 비교하고 그중에 가장 우수한 성능을 보인 Voting Ensemble을 통해서 분석된 상관관계가 높은 환경 요소와 그 패턴을 살펴보았다.

그 결과 온도(temperature), 시간(hour), 습도(humidity), 월(month), 평일(working day), 미세먼지(pm)의 6개 환경 데이터 인자의 순서로 특징중요도(feature importance)가 보행량 예측에 가장 큰 영향을 주는 인자로 측정이 되었다. 평가지표로는 결정계수인  $R^2$ 이 1에 가까울수록 인공지능 모델이 우수한 성능을 보인다고 평가되는데, 가장 모델 평가지수가 높은 것으로 평가된 Voting Ensemble의 경우 결정계수  $R^2$ 이 0.72로 가장 우수한 것을 알 수 있으며, XGBoostRegressor이 0.70, LightGBM이 0.69, RandomForestRegressor가 0.53으로 평가되었다. 향후, 스마트공원의 관리기능으로서 환경 데이터를 활용한 보행량의 예측에, 본 연구에서 확인된 Voting Ensemble 머신러닝 모델을 사용하여 정량적인 예측과 응용에 사용할 수 있을 것으로 기대된다.

두 번째 스마트공원 구현에서는, 기후변화 위기 방재역할로서, 인천 소래생태공원의 ‘갈대숲’ 과 ‘염습지’ 에서 폭염기 환경 데이터를 수집 이용, 도심지인 ‘포구 사거리’ 에서 수집한 온·습도 등의 환경 데이터와 비교하면서, 이산화탄소의 발생량에 영향을 주는 요소들과 사이트별 차이를 분석하였다.

2022년 7월 9일 ~ 8월 17일 기간 동안, 인천 소래생태공원과 주변 세 곳의 사이트의 폭염기 환경 자료를 수집 이용하였다. ‘갈대숲’ 과 ‘염습지’ 의 도심지인 ‘포구 사거리’ 에서 수집한 환경 데이터와 비교하여 이산화탄소의 발생량에 영향을 주는 요소들을 분석하였다. 일반적인 데이터와 범주 데이터, 시계열 데이터와의 분석은 파이선과 구글 코랩을 사용, 각종 그래프로 도식화하여 대상지의 상태를 쉽게 살펴볼 수 있도록 하였다. 사이트별로 온·습도, 미세먼지, 초미세먼지 그리고 이산화탄소의 차이를 비교하여, 염습지의 온도감소 효과, 탄소 저감 효과 등을 살펴보았다.

그 결과 도로부인 ‘포구 사거리’ 의 이산화탄소 발생량은 소래생태공원 ‘염

습지'에 비해서 평균 40ppm, 최고 60ppm 정도 높게 측정되었고, 기온은 '포구 사거리'가 시간과 날짜와 관계없이 일정하게 약 1.5 °C 정도 높게 유지되었다. 초미세먼지(pm2.5)의 비교에서, 도시에 위치한 '포구 사거리'는 매일 평균 '갈대숲' 보다는 약  $3\mu\text{g}/\text{m}^3$  정도 높게 유지되었고, '염습지' 대비  $4\mu\text{g}/\text{m}^3$  높게 나타났다. 미세먼지(pm10)의 비교에서는 도시에 위치한 '포구 사거리'는 매일 평균 '갈대숲' 보다는 약  $3.5\mu\text{g}/\text{m}^3$  정도 높게 유지되었고, '염습지' 대비는  $4\mu\text{g}/\text{m}^3$ 의 차이를 유지하였다.

이어서, 기후 변화 대응을 위하여 환경데이터와 이산화탄소 발생의 상관관계를 분석하기 위해, Auto ML 머신러닝 모델을 이용하였다. 여름철 폭염 기간에 수집된 환경 빅데이터는 전처리를 거쳐, 사이트별 유의미한 데이터로 구성하였다.

인공지능분석을 통한 머신러닝 알고리즘 모델별 예측 성능을 비교하고 그중에 가장 우수한 결과를 보인 Stacking Ensemble을 통해서 분석된 상관관계가 높은 환경 요소를 파악하고, 그 패턴을 '염습지', '갈대숲', 그리고 '포구 사거리'의 세 구역에 따라 나누어 Regression 모델링의 비교 분석을 수행하였다.

그 결과, 첫 번째 사이트인 '염습지'에서는 온도(sr\_salty\_temp), 돌풍(sr\_salty\_Wind\_Gust), 습도(sr\_salty\_humi), 기압(sr\_salty\_pressure), 초미세먼지(sr\_salty\_pm2.5), 시간(hour)의 6개 인자의 순서로 이산화탄소 발생량에 가장 큰 영향을 주는 것을 알 수 있었다. 이때 적용된 Stacking Ensemble의 결정계수인  $R^2$ 이 0.90으로 가장 우수한 것을 알 수 있었으며, XGBoostRegressor이 0.88, LightGBM이 0.88, RandomForestRegressor가 0.66의 순으로 나타났다. 이와 같이 인공지능 학습 모델 평가를 수행함으로써 최적의 알고리즘을 선정할 수 있었다.

두 번째 사이트인 '갈대숲'에서는, 습도(sr\_galdae\_humi), 돌풍(sr\_galdae\_Wind Gust), 기압(sr\_galdae\_pressure), 풍향(sr\_galdae\_Wind Direction), 시간(hour), 기온(sr\_galdae\_temp)의 6개 인자가 이산화탄소 발생량 가장 큰 영향을 주는 것을 알 수 있었다. Stacking Ensemble의  $R^2$ 이 0.918로 가장 우수한 것을 알 수 있었으며, XGBoostRegressor가 0.915, LightGBM이 0.89, RandomForestRegressor가 0.80의 순으로 인공지능 학습모델로 평가되었다. 이에 따라 인공지능 학습모델 평가를 수행함으로써 최적의 알고리즘을 선정할 수 있었다.

세 번째 사이트인 도심부 ‘포구 사거리’에서는, 도심부 내의 환경센서 설치의 제약으로, 일부 갈대숲의 데이터를 차용하여 사용한 결과, 갈대숲 풍향(sr\_galdae\_Wind Direction), 온도(sr\_pg4\_temp), 포구 사거리 미세먼지(sr\_pg4\_pm10), 갈대숲 태양조도(sr\_galdae\_solar), 포구 사거리 습도(sr\_pg4\_humi), 갈대숲 풍속(sr\_galdae\_Wind Speed)의 순서로 6개 인자가 이산화탄소 발생량에 가장 큰 영향을 주는 것으로 확인되었다. 이때 Auto ML에서 적용된 Stacking Ensemble의  $R^2$ 이 0.85로 가장 우수한 것을 알 수 있으며, XGBoostRegressor가 0.83, LightGBM이 0.79, RandomForestRegressor가 0.52의 순으로 평가되었다. 이에 따라 인공지능 학습모델 평가를 수행함으로써 최적의 알고리즘을 선정할 수 있었다.

지금까지 논의한 바는 향후 스마트공원의 방재 기능으로서 환경데이터를 활용한 이산화탄소 발생과 저감 예측 연구에서 확인된 최적의 Stacking Ensemble 머신러닝 모델의 정량적인 예측과 응용에 사용할 수 있을 것으로 기대된다.

이와 같이 소래습지 생태공원 대상 연구를 통해서 이산화탄소의 저감에 영향을 주는 주요 요소들을 살펴보았을 때 태양광, 온도 등 식물의 광합성을 주는 인자들과 돌풍, 풍향, 풍속 등에 의해서 영향이 큰 것을 알 수 있었다. 특히 풍향의 경우에는 포구 사거리를 기준으로 서쪽에 있는 남동 공단 측에서부터 불어오는 서풍은 이산화탄소를 증가시키고, 소래생태공원이 위치한 동쪽에서 불어오는 동풍의 경우에는 이산화탄소를 줄이는 효과가 큰 것으로 상관관계분석이 되어, 향후 추가적인 연구가 더 필요할 것으로 보인다.

본 연구에서는 기후변화 대응의 정량화된 평가시스템을 도입한 서울숲 스마트공원구현, 인천 소래 생태습지 스마트공원구현을 각기 통해 경제적인 IoT 환경 데이터 센싱 개발 방법을 제안하고 최적의 인공지능 머신러닝 모델을 설계하여, 기존의 공원을 스마트공원화 하는 데 있어 그 수월성을 제고하는 데 기여하고자 하였다. 이에 따른 기대 효과는 다음과 같다. 첫째, 스마트공원의 관리기능으로서 환경변수에 따른 방문자 보행량에 대한 정확한 예측을 통해서 최적의 서비스를 제공할 수 있을 것이다. 둘째, 전 세계적으로 해결해야만 하는 기후변화 위기 대응에 스마트공원의 탄소저감, 폭염 저감과 같은 방재 기능을 강화하는 데 참고할 연구 사례로서 향후 도시공원의 가치 제고에 기여할 것으로 기대된다.

## 1.2 연구의 한계 및 의의

본 연구에서는 기존 모니터링 온습도 측정 방법과 달리, 멀티채널의 실시간 입력방식의 IoT 장비를 제작하여 서울숲공원에 설치, 2021년 5월 21일부터 2022년 1월 26일까지 서울숲 보행량 수집기간에 맞추어 마이크로 기상 자료를 수집하였다. 그렇지만 공원의 최초 설계 시에 IoT 장비에 반드시 필요한 전원공급과 인터넷망 연결에 대한 부분이 전혀 고려되지 않아 초기에 태양광 등 다른 대체 방법을 찾기 위한 애로사항이 많았다. 따라서 앞으로 스마트공원을 설계한다면 초기에 전력과 인터넷이 원활하게 공급될 수 있도록 요소지 내 시설 준비 및 배치에 유의해야 할 것이다.

그리고 자체개발 장비성능의 공인인증성을 고려하여, 인천의 소래생태공원에는, 서울숲의 IoT장비와는 차별화를 두어 2022년 7월 9일 ~ 8월 17일 기간 동안에 상업용 AWS 장비를 설치하여, 스마트공원의 기후변화 대응 효과 분석을 위한 인공지능 모델링 실험을 각각 비교 진행하였으며, 그 결과 자체 개발 장비의 성능상의 문제는 발견되지 않았다.

이처럼 본 연구에서는 제작된 IoT 센서를 활용하여, 장시간에 걸쳐 매 20초마다 클라우드 서버에, 온도, 습도, 일산화탄소, 미세먼지 데이터를 저장하고, 데이터 분석에 활용하였다. 하지만 제작된 장비가 일반 고가의 양질 장비보다는 아직은 미흡한 부분과 성찰되고 공신력이 있는 장비로서 인증을 받지 못한 한계가 있다. 특히 설치된 위치가 공간의 제약으로 인하여 전통적인 백엽상의 설치 위치인 햇빛을 직접 받지 않고 바람이 잘 통하는 통기성 구조로 지상에서 1.2~1.5m 높이에 설치되지 아니하고, 가족마당 중심에 있는 기둥에 설치되어 태양광에 노출되도록 설치되었다. 그 결과, 서울시 평균온도보다 최대 12도 높게, 습도는 최대 30% 정도 낮게 측정되는 되었는데, 이 점은, 연구에 대한 한계점이라고 할 수 있다.

그럼에도 불구하고 본연구는 스마트공원의 데이터 정량화를 통하여, 기후변화의 대응 효과를 인공지능을 통해서 살펴보는 최초의 시도라고 할 수 있고, 또 향후 지자체에서 저렴하게 활용할 수 있는 IoT 장비를 공급하는 차원에서의 개발 방법을 공개하였다는 점에서 그 의의가 크다고 할 수 있다.

또한, 향후 스마트공원에 투입되는 많은 IoT 장비를 위해서는 지속적인 전원과 무선 통신(WiFi)이 필요한 점을 사전에 인식하고 양지하여, 스마트공원설계에

필요한 요소별 위치에 사전에 적절한 전원, 통신 등을 갖춰 놓는다면, 다양한 방면에서 활용이 가능할 것이다.

한편 인천 소래습지 공원의 경우에는 도심부 현장 상황으로 인하여 추가적인 외부 센싱 장비를 설치하지 못하는 한계가 있었다. 본 연구에서는 사이트 연구 대상지 중 도심지에 풍향, 풍속, 우적량계를 설치하지 않고, 추가 2개 지점의 사이트 연구 대상지 중 포구사거리에 조금 더 가까운 갈대숲으로 설정하여, 해당지점의 데이터를 이용하였다. 해당 데이터에서 이산화탄소 발생에 영향을 주는 인자를 인공지능 영향 평가 모델을 통해 살펴보았다. 그러나 도심지 안에서의 직접적인 풍향, 풍속, 우적량 등의 마이크로(micro) 데이터를 활용하지 못한 점이 본 연구의 한계점이라고 할 수 있다.

추후 연구를 통해 이 부분에 대한 개선이 이루어질 경우 본 연구에서 개발한 시뮬레이션 장비와 인공지능 엔진 등을 활용, 스마트공원의 관리적인 부분뿐만 아니라 향후 스마트공원의 폭염, 미세먼지, 이산화탄소 저감 등 기후변화에 따른 방재 효과로서의 도심지 스마트공원의 수요와 기술 확장이 매우 클 것으로 사료된다.

본 연구의 의의와 한계점을 바탕으로, 향후에는 전국 도시공원에 적절한 IoT 환경 센서와 여기에 인공지능기술을 결합한 엣지컴퓨팅 기술을 도입하여, 스마트공원을 확산시키고, 정량적인 도시공원의 기후 위기 방재 효과를 재고할 수 있는 방향을 모색하고자 한다.

## 참고 문헌

### ■ 단행본

1. 로니킵(2021), 『모두의 아두이노 환경 센서』, 길벗.
2. 소전섭 외(2000), 『대기관측법』, 교문사.
3. Loukaitou-Sideris, et al. (2018), Smart Parks: A Toolkit, UCLA: Luskin Center, USA.

### ■ 국내 학술지 및 학술발표대회 논문

4. 구형일(2018), 「인공지능 및 딥러닝 동향」, 『전기의세계』, 67(7), 7-12.
5. 권재민, 이재학, 조민도, 최영준, 한승우(2019), 「다중선형회귀분석 기반 건설장비 이산화탄소 배출량 예측모델 개발」, 『한국건축시공학회 학술발표대회 논문집』, 38-39.
6. 권태경 김이경, 박예진, 유다연, 조현민, 임영신(2022), 「IoT 모니터링 시스템을 통한 기후변화 취약계층 지원 차열 사업의 효과 분석-김해시 진영읍 주거지역을 사례로」, 『한국기후변화학회지』, 13(3), 383-398.
7. 권태경, 조현민, 임영신(2022), 「도시 그늘막 효과 측정을 위한 시뮬레이션 장비 개발에 관한 연구」, 『한국환경기술학회지』, 23(3), 169-176.
8. 김규석, 서예진(2021), 「코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안」, 『한국정보기술학회논문지』, 19(10), 11-18.
9. 김규석, 서예진(2021), 「코로나19 확산에 따른 공공 자전거 이용자 수 예측을 위한 인공지능 모형의 제안」, 『한국정보기술학회논문지』, 19(10), 11-18.
10. 김명자(2021). 「기조발제: 기후·팬데믹 복합위기와 4차 산업혁명 시대, 지속 가능한 발전과 법률가의 사명」. 『저스티스』, 182(2), 23-86.
11. 김명진, 박대영, 서명희, 이정하, 우종우. (2019). 「실내 공기질 파악 및 관리에 관한 연구」, 『한국정보과학회 학술발표논문집』, (), 1673-1675.
12. 김문희, 김두현(2004), 「임베디드 시스템과 소프트웨어공학」, 『정보과학회지』, 22(6), 5-13.
13. 김민, 김동규(2022), 「Ensemble 알고리즘을 이용한 Machine Learning 하드웨어 Backdoor 탐지 방법」, 『대한전자공학회 학술대회』, 559.
14. 김성원, 이환수(2020), 「「스마트도시 조성 및 산업진흥 등에 관한 법률」의 개선 방안에 관한 연구」, 『법학논총』, 44(2), 59-92.
15. 김세형, 박재형, 안상호, & 강주영(2020), 「인공위성 이미지를 활용한 공유자전거 대여 수요 예측」, 『한국통신학회 학술대회논문집』, 182-183.
16. 김수이(2021), 「탄소중립의 개념과 당위성」, 『전기저널』, (531), 18-24.
17. 김예슬, 박시온, 박건웅(2020), 「서울 공공자전거 신규 대여소를 위한 수요량 예측 분석」, 『응용통계연구』, 33(6), 739-751.
18. 김용국, 송유미, 조상규(2020), 「스마트 그린시티 구현을 위한 스마트 공원설계·관리 방향」, 『한국조경학회지』, 48(6), 1-15.
19. 김용국(2014). 「녹색 복지 관점에서 서울시 생활권 도시공원의 분배적 형평성 분석」, 『한국조경학회지』, 42(3), 76-89.

20. 김우형, 안정식, 김남우(2021), 「블록체인 기반 지능형 미세먼지 감지 시스템, 『한국정보통신학회 여성 ICT 학술대회 논문집』, 86-89.
21. 김윤조 (2022), 「흉부 X-선 영상에서 밝기 값 정규화 및 다중 스케일 폐-집중 패치를 사용한 앙상블 딥러닝 모델 기반의 폐렴 자동 분류」, 『정보과학회논문지』, 49(9), 677-685.
22. 김익희(2018), 「뉴욕, 데이터 중심 스마트시티」, 『국토』, 443, 98-105.
23. 김정환, 안진서, 홍헬렌(1998), 「폰 노이만 구조를 채택한 데이터플로우 다중스레드 컴퓨터」, 『정보과학회논문지』 (A), 25(9), 917-925.
24. 김종현 한진주, 김장원, 온병원(2018), 「실내 공기 개선을 위한 공기 측정 및 시각화 시스템 설계」, 『In Proceedings of KIIT Conference』, 128-131.
25. 김진목, 김세용(2015). 「도시재생 특별법」에서 기후변화 대응 전략 반영의 필요성」, 『한국주거학회 학술대회논문집』, 141-145.
26. 김태경, 봉인식, 이상대, 이성룡, 황금희, 김성하, 정효진(2018), 「4차 산업혁명 시대의 스마트시티 전략」, 『정책연구』, 경기연구원, 1-235. (), 1-235.
27. 김태환 외(2020), 「셸드 TBM 기계 데이터 및 머신러닝 기법을 이용한 암석의 일축압축강도 예측」, 『터널과 지하공간』, 30(3), 214-225.
28. 김형래 외(2022), 「광산란 방식 간이형 미세먼지 센서의 입경 분리능 평가기법 개발」, 『한국대기환경학회지』, 38(1), 159-168.
29. 김효준, 조영흠, 류성룡(2020), 「실내 이산화탄소 농도 예측을 위한 기계학습 모델 검증」, 『한국건축환경설비학회 논문집』, 14(6), 699-706.
30. 문창현(2001), 「이용 만족도 결정요인의 분석을 통한 서울시 도시공원의 공간적 특성에 관한 연구」, 『한국지리환경교육학회지』, 9(1), 97-118.
31. 민소아, 정윤서(2021), 「서울시 공공자전거 수요예측 모형 비교 연구」, 『한국데이터정보과학회지』, 32(3), 585-592.
32. 박세연, 황찬규, 박동철(2015), 「사물인터넷 (IoT) 환경에서 스마트폰 알람을 이용한 아두이노 (Arduino) 기반 가전기기 대기전력을 최소화한 ON 시스템 구현」, 『한국전자통신학회 논문지』, 10(10), 1175-1182.
33. 박종화, 조기혁(2016), 「공원 크기에 따른 공원의 온도 저감 효과 분석: 일산 신도시를 대상으로」, 『국토계획』, 51(5), 247-261.
34. 박춘식(2011), 「클라우드 컴퓨팅에서의 보안 고려사항에 관한 연구」 『한국산학기술학회 논문지』, 12(3), 1408-1416.
35. 반효경(2008), 「운영 체제와 정보기술의 원리」, 『이화여대 출판사』.
36. 서경덕(2020), 「저가형 초미세먼지 센서 정확도 향상을 위한 인공지능 기반 센서 데이터 예측 기법」, 『한국정보기술학회 종합학술발표논문집』, 179-183.
37. 서경덕(2020), 「저가형 초미세먼지 센서 정확도 향상을 위한 인공지능 기반 센서 데이터 예측 기법」, 『한국정보기술학회 종합학술발표논문집』, 179-183.
38. 서창수, 김연태, 김성희(2020), 「지속가능한 스마트시티 서비스 활성화를 위한 빅데이터 우선순위에 관한 연구」, 『도시행정학보』, 33(4), 61.
39. 성기우, 김시범, 강영빈(2021), 「Auto ML을 활용한 필드 클레임 예측 및 고장 주요 인자 분석」, 『한국신뢰성학회 학술대회논문집』, 132-132.
40. 손채준, 권태수, 이윤상(2021), 「상태 표현 방식에 따른 심층 강화 학습 기반 캐릭터 제어기의 학습 성능 비교」, 『한국컴퓨터그래픽스학회 논문지』, 27(5), 55-61.

41. 신해민, 최주현, 길수연, 김유선(2021), 「포스트 코로나 시대에 적합한 공원카페 계획 및 제안」, 『한국색채학회 학술대회』, 140-143.
42. 심대천, 이진용, 장지옥, 이민욱(2022), 「XGBoost Regression 기계학습을 이용한 제주도 지하수 염소이온 농도예측 연구 사례」, 『지질학회지』, 58(2), 243-255.
43. 안경수(2001), 「수도권 해양 생태공원 조성사례-생태계 현황조사에 의한 습지 조성 계획 중심으로」, 『물과 미래: 한국수자원학회지』, 34(5), 40-67.
44. 안소영, 원두환(2015), 「메타분석을 이용한 도시공원의 경제적 가치 함수 추정」, 『한국환경정책학회 학술대회논문집』, 72-73.
45. 안인수(2012), 「태양에너지를 이용한 일체형의 전원공급시스템과 DC 램프 모듈 개발」, 『한국컴퓨터정보학회논문지』, 17(2), 23-29.
46. 오창세, 서민석, 이정혁, 김상현, 김영돈, 박현주(2015), 「IoT 기반 실내 공기질 모니터링 시스템」, 『한국통신학회논문지』, 40(5), 886-891.
47. 우덕채, 문현실, 권순범, 조운호(2019), 「트랜잭션 기반 머신러닝에서 특성 추출 자동화를 위한 딥러닝 응용」, 『한국 IT 서비스학회지』, 18(2), 143-159.
48. 원동준, 김선겸, 김영훈, 송규원(2021), 「기계학습을 활용한 경기도 산업단지 미세먼지 예측」, 『정보과학회논문지』, 48(7), 764-773.
49. 윤민호, 안동만(2009), 「위성영상을 이용한 도시녹지의 기온 저감 효과 분석」, 『한국조경학회지』, 37(3), 46-53.
50. 윤순진(2009). 「저탄소 녹색성장의 이념적 기초와 실재」, 『환경사회학연구』, 13(1), 219-266.
51. 이근상, 이종조(2017), 「UAV 기반 열적외선 센서를 이용한 태양광 셀의 발열 검출」, 『대한공간정보학회지』, 25(1), 71-78.
52. 이금선, 박만곤(2010), 「유비쿼터스 러닝에서 마이크로SD 카드를 활용한 간편 교육 매체 전달 시스템의 설계」, 『한국멀티미디어학회 학술발표논문집』, 205-207.
53. 이동근, 김호걸(2014), 「그린 인프라를 통한 기후변화 문제 해결」. 『환경논총』, 54. (), 27-36.
54. 이애란(2018). 「민간공원 특례사업의 기후변화대응 방재계획 분석: 청주시 사례를 대상으로」, 『도시연구』, 14, 119-141.
55. 이영우(2013), 「버스정보시스템(BIS) 수집자료를 이용한 경로통행시간 추정」, 『대한토목학회』, 33(3), 1115-1122.
56. 이용준 외(2020), 「XGBoost를 활용한 고속도로 콘크리트 포장 파손 예측」, 『한국건설관리학회 논문집』, 21(6), 46-55.
57. 이은택, 이성재, 김수주, 윤해수(2019), 「공동주택의 실내 미세먼지 저감을 위한 기술개발」, 『대한설비공학회 학술발표대회논문집』, 111-112.
58. 이제승, 이현희(2019), 「기계학습 알고리즘을 이용한 보행 만족도 예측모형 개발」, 『국토계획』, 54(3), 106-118.
59. 이주용, 이형숙(2021), 「이용자 중심의 스마트공원 조성을 위한 공원 서비스의 상대적 중요도 분석」. 『한국조경학회지』, 49(4), 55-63.
60. 이현구, 오재인(2017), 「스마트시티 서비스 유형분류」, 『한국경영과학회지』, 42(4), 147-161.
61. 이형숙, 민병욱, 양태진, 엄정희, 김권, 이주용(2019), 「스마트공원 개념 정립 및 공원 이용자

- 인식에 관한 연구-대구 IoT See 시범사업 공원 이용자를 대상으로」, 『한국조경학회지』, 47(5), 41-48.
62. 임상택, 백순창, 송용준, 백영태, 최차봉, 송승인(2019), 「지능형 AI기반의 미세먼지 저감 제어 시스템」, 『한국컴퓨터정보학회 학술발표논문집』, 27(1), 53-56.
  63. 임은나, 이우성, 최철현, 송봉근, 정성관(2013), 「도시근린공원의 열환경 개선을 위한 열쾌적성 평가」, 『한국지리정보학회지』, 16(4), 153-170.
  64. 장미홍, 박창열(2019), 「일본 도시공원의 방재 기능 특징과 시사점 연구」 『서울 도시연구』, 20(2), 77-89.
  65. 장윤정, 정대원(2019), 「딥러닝 기반의 영상 분석 기술 동향」, 『항공우주산업기술동향』, 17(1), 113-122.
  66. 장재훈(2021), 「스마트공원, 정보 계측·시물레이션 중요」, 『환경과 조경』, 398호.
  67. 장혜정(2020), 「시민 니즈와 참여 기반의 스마트시티 문제 해결을 위한 빅데이터 활용 절차에 관한 연구」, 『한국정보전자통신기술학회 논문지』, 13(2), 102-112
  68. 전수정, 패트릭, 윤근영(2020), 「머신러닝 기법을 이용한 서울지역 어린이집 실내공기질 예측」, 『한국생태환경건축학회 학술발표대회 논문집』, 20(2), 112-113.
  69. 조정진(2003), 「프레데릭 로 옴스테드의 도시공원관에 대한 재해석」, 『한국조경학회지』, 30(6), 26-37.
  70. 조정진(2014), 「히르시펠트(C.C.L.Hirschfeld) 정원예술론의 의미와 가치의 현대적 해석」, 『한국전통조경학회지』, 32(3), 58-68.
  71. 조남훈(2006), 「교차검증을 이용한 SVM 전력수요예측」, 『전기학회논문지』 (A), 55(11), 485-491.
  72. 조성래, 성행남, 안병혁(2015), 「의사결정트리와 인공 신경망 기법을 이용한 침입탐지 효율성 비교 연구」, 『디지털산업정보학회논문지』, 11(4), 33-45.
  73. 조진희, 이채연, 배민기, 오후. (2018). 미시적 도시공간의 폭염취약지역분석을 통한 정책적 대응방안. 충북 FOCUS, (), 1-25.
  74. 조현민, 권태경(2022). 「여름철 도시공원의 열환경 개선 효과 - 서울숲 미기상 관측자료 분석을 중심으로」, 『한국조경학회지』, 50(6), 30-41.
  75. 주현수, 김석철(2009), 「도심지역 녹지의 국지적 대기환경영향에 관한 연구」. 『화학공학』, 47(3), 386-393.
  76. 차시호(2022), 「사물 웹(WoT) 환경에서 네트워크 모니터링 애플리케이션 개발을 위한 웹 프론트엔드 프레임워크의 적용 현황 및 트렌드」, 『디지털산업정보학회 논문지』, 18(1), 47-54.
  77. 최승한, 한미경(2020), 「어텐션 적용 LSTM을 사용한 자전거 수요 예측 모델의 구현」, 『한국통신학회 학술대회논문집』, 342-343.
  78. 최학회, 응위엔, 바롱, 김준혁, 김동성(2016), 「사물인터넷(IoT) 기반의 공원을 위한 자원 예측 모델 설계 및 구현」, 『한국통신학회 학술대회논문집』, 391-392.
  79. 한재경(2018), 「범죄예방용 CCTV의 설치 및 운영에 관한 소고」, 『공공사회연구』, 8(4), 109-137.
  80. 허경호, 박종민(2020), 「중력 모델을 이용한 전주시 근린공원의 수요 예측」, 『한국환경생태학회 학술발표논문집』, 2020(1), 35-36.
  81. 홍철호, 김동진, 정영창, 김정도. (2005), 「임베디드 형태의 VOD 클라이언트 시스템의 개

- 발」, 『한국산학기술학회 논문지』, 6(4), 315-324.
82. 홍태훈, 안중백, 김지민, 김학평, 이민현, 정우진, 공민진(2018), 「딥러닝 알고리즘을 사용한 실내 이산화탄소 농도의 예측모델 개발」, 『한국에너지기후변화학회 학술대회』, 186.
83. 황지환, 박홍석, 김재동(2022), 「양상블 학습 기반의 동시조달 수리부속 예측모형 연구」, 『한국산학기술학회 논문지』, 23(8), 431-438.
84. 황치곤, 윤창표(2020), 「머신러닝을 위한 온톨로지 기반의 Raw Data 전처리 기법」, 『한국정보통신학회논문지』, 24(5), 600-608.
85. Lee, H.(2001), 「차세대 Embedded 마이크로프로세서 기술 동향」, 『The Magazine of the IEIE』, 28(7), 49-55.
86. Yu-Jin NAM, Won-Ji SHIN(2019), 「A Study on Comparison of Lung Cancer Prediction Using Ensemble Machine Learning」, 『인공지능학회지』, 7(2), 19-24.

## ■ 국외 학술지 및 학술발표대회 논문

87. Acock, A. C.(2005), “Working with missing values,” *Journal of Marriage and family*, 67(4), 1012-1028.
88. Austin, E., Novosselov, I., Seto, E., Yost, M. G.(2015). “Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor,” *PloS one*, 10(9), e0137789.
89. Bergstra, J., Bengio, Y.(2012). “Random search for hyper-parameter optimization,” *Journal of machine learning research*, 13(2).
90. Bilogur, A.(2018), “Missingno: a missing data visualization suite,” *Journal of Open Source Software*, 3(22), 547.
91. Bowler, D. E., Buyung-Ali, L., Knight, T. M., Pullin, A. S.(2010), “Urban greening to cool towns and cities: A systematic review of the empirical evidence,” *Landscape and urban planning*, 97(3), 147-155.
92. Chen, T., Guestrin, C.(2016), “Xgboost: A scalable tree boosting system,” *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
93. Chung, D. H., Yun, J. S., Yang, S. M.(2021). “Machine Learning for Predicting Entrepreneurial Innovativeness,” *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 16(3), 73-86.
94. Feurer, M., Klein, A., Eggenberger, K., Springenberg, J., Blum, M., Hutter, F.(2015). “Efficient and robust automated machine learning,” *Advances in neural information processing systems*, 28.
95. Feyisa, G. L., Dons, K., Meilby, H.(2014), “Efficiency of parks in mitigating urban heat island effect: An example from Addis Ababa,” *Landscape and urban planning*, 123, 87-95.
96. Harishkumar, K. S., Yogesh, K. M., Gad, I.(2020). Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models. *Procedia Computer Science*, 171, 2057-2066.
97. Jim, C. Y., Chen, W. Y.(2008), “Assessing the ecosystem service of air pollutant removal by urban trees in Guangzhou (China),” *Journal of environmental management*, 88(4), 665-676.
98. Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R.(2010), “Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system”, *Pervasive and Mobile Computing*, 6(4), 455-466.
99. Konijnendijk, C. C., Annerstedt, M., Nielsen, A. B., Maruthaveeran, S.(2013), “Benefits of urban

- parks. A systematic review,” *A Report for IFPRA*, Copenhagen & Alnarp.
100. Lee, J. C., Jun, H. S., Choi, Y. R., Kim, J. H. (2004). Setup of the development tools for a small-sized controller built in a robot using Linux (No. KAERI/TR--2710/2004). Korea Atomic Energy Research Institute.
  101. Maureira, M. A. G., Oldenhof, D., & Teernstra, L. (2011). ThingSpeak—an API and Web Service for the Internet of Things. World Wide Web.
  102. Mihai, B. O. G. D. A. N.(2016). “How to use the DHT22 sensor for measuring temperature and humidity with the arduino board,” *Acta Universitatis Cibiniensis–Technical Series*, 68, 22-25.
  103. Pasha, S.(2016), “ThingSpeak based sensing and monitoring system for IoT with Matlab Analysis,” *International Journal of New Technology and Research*, 2(6), 263-492.
  104. Sai, K. B. K., Subbareddy, S. R., Luhach, A. K.(2019), “IOT based air quality monitoring system using MQ135 and MQ7 with machine learning analysis,” *Scalable Computing: Practice and Experience*, 20(4), 599-606.
  105. Wang, K., Yang, L., & Kucharek, M. (2020). Investigation of the effect of thermal insulation materials on packaging performance. *Packaging Technology and Science*, 33(6), 227-236.
  106. Yin, S., Z. Shen, X. Zou, S. Che, W. Wang(2011), “Quantifying air pollution attenuation within urban parks: An experimental approach in Shanghai, China,” *Environmental Pollution*, 158(8-9), 2155-2163.
  107. Yoon, J. W., Pinelli, F., Calabrese, F.(2012). “Cityride: a predictive bike sharing journey advisor,” *IEEE 13th International Conference on Mobile Data Management*, IEEE, 306-311.

## ■ 학위 논문

108. 김규태(2022), 「다중계층 스택킹 앙상블을 이용한 제조데이터의 결함 분류 모델에 관한 연구」, 성균관대학교 대학원 석사학위논문.
109. 김한성(2022), 「호흡 질환과 증상 분류용 CNN 기반 앙상블 모델 및 연관성에 따른 호흡 증상 분류 모델 연구」, 강원대학교 대학원 석사학위논문.
110. 양호원(2008), 「신경망과 의사결정트리를 이용한 Stream Data 예측 시스템 설계 및 구현」, 조선대학교 대학원 박사학위 논문.
111. 이성주(2006), 「오토존데를 이용한 고충관측 자동화 평가와 활용에 관한 연구」, 조선대학교 대학원 박사학위 논문.
112. 이은영(2018), 「스마트 도시공원 조성 전략 및 제도 개선에 관한 연구」, 한양대학교 대학원 박사학위논문.
113. 이제이(2013), 「1인 미디어에 나타난 서울숲의 미시담론 연구-장소감을 중심으로」, 서울대학교 대학원 박사학위 논문.

## ■ 신문 및 잡지 자료

114. 김동원, 「에스아이에이, 이산화탄소 배출량 AI로 검출하는 기술 가능성 제기」, 『AI타임즈』 (2022. 03. 18.)
115. 박설민, 「똑똑하게 줄여요... AI, 탄소중립 ‘게임체인저’ 될까」, 『시사위크』 (2022. 04. 11.)
116. 신정은, 「淸황실 정원에 AI를 더하다...자율주행버스 달리고, 얼굴인식 운동량 측정」,

- 『이데일리』, 2020. 01. 23.
117. 안광호, 「탄소 흡수능력 뛰어난 염습지 ‘조림 경제효과’ 일반 땅의 5배」, 『경향신문』 (2021. 04. 20.)
118. 임민수. 「블루카본(Blue Carbon)」, 『에코타임스』, 2022. 07. 21.
119. 최아름, 「인공지능도 친환경 시대...그린 AI ‘주목’」, 『정보통신신문』 (2022. 04. 02.)

## ■ 보고서

120. 권태경(1996), 「디지털글러브-가상현실용 휴먼인터페이스」, 대우전자 영상연구소 기술 보고서.
121. 김원주 외(2010), 「공원녹지 분야의 탄소흡수원 확보 및 탄소저감방안」, 서울연구원 정책 과제연구보고서, 1-155.
122. 손용훈 외(2021), 「KEI 기후친화형 어린이 컴퓨터 모니터링 보고서」, 서울대학교.
123. 이정훈 외(2018), 「4차 산업혁명 경기도 모델 구축 및 실행계획 연구」, 경기연구원, 1-399.
124. 이양주 외(2020), 「경기도 생태(生態) 분야 그린뉴딜 정책연구」, 경기연구원, 1-151.

## ■ 웹페이지 자료

125. “[Python] 어떤 스케일러를 쓸 것인가?”, <https://mkjjo.github.io/python/2019/01/10/scaler.html> (검색일: 2022. 10. 15.)
126. “[파이썬 시각화] QQ플롯(Quantile-Quantile plot)”, <https://blog.naver.com/breezehome50/222339337268> (검색일: 2022. 10. 15.)
127. “머신러닝 앙상블 학습(Ensemble Learning) - 보팅(Voting)“, <https://ek-koh.github.io/data%20analysis/voting> (검색일: 2022. 10. 15.)
128. “부산 스마트시티, ‘에코-스마트’ 공원시설물 공모”, <https://www.smartcitytoday.co.kr/news/articleView.html?idxno=21386> (검색일: 2022. 10. 17.).
129. “삼성SDS. “인공지능을 위한 인공지능, AutoML”, [https://www.samsungsds.com/kr/insights/ai\\_automl.html](https://www.samsungsds.com/kr/insights/ai_automl.html)(검색일: 2022. 10. 15.)
130. “상자 수염 그림”, 위키피디아, [https://ko.wikipedia.org/wiki/%EC%83%81%EC%9E%90\\_%EC%88%98%EC%97%BC\\_%EA%B7%B8%EB%A6%BC](https://ko.wikipedia.org/wiki/%EC%83%81%EC%9E%90_%EC%88%98%EC%97%BC_%EA%B7%B8%EB%A6%BC) (검색일: 2022년 10월 15일)
131. “세종국가지법도시 미래형 스마트 실증공원 마스터플랜 수립 전문가 선정”, [https://www.kila.or.kr/board/view?b\\_name=bo\\_news&bo\\_id=4310&per\\_page](https://www.kila.or.kr/board/view?b_name=bo_news&bo_id=4310&per_page)(검색일: 2022. 10. 17).
132. “인공지능을 위한 인공지능, AutoML“, 삼성SDS, [https://www.samsungsds.com/kr/insights/ai\\_automl.html](https://www.samsungsds.com/kr/insights/ai_automl.html) (검색일: 2022. 10. 15.)
133. “자동화된 Machine Learning(AutoML)이란”, <http://learn.microsoft.com/ko-kr/azure/machine-learning/concept-automated-ml> (검색일: 2022. 10. 15.)
134. “한화의 기술력과 AI가 관리하는 탄소배출관리 시스템, 맵그로브”, <https://www.hanwha.co.kr/media/discover/view.do?seq=5046> (검색일: 2022. 10. 15.)
135. “히트맵”, 위키피디아, [https://ko.wikipedia.org/wiki/%ED%9E%88%ED%8A%B8\\_%EB%A7%B5](https://ko.wikipedia.org/wiki/%ED%9E%88%ED%8A%B8_%EB%A7%B5) (검색일 2022년 10

- 월15일)
136. 아이데이터(주). <https://i-data.kr/products.php> (검색일: 2022. 10. 09.)
  137. “Deep learning vs. machine learning in Azure Machine Learning”, <https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?source=recommendations> (검색일: 2022. 10. 15.)
  138. “Evaluating Models in Azure Machine Learning (Part 2: Regression)”, <https://adatis.co.uk/evaluating-models-in-azure-machine-learning-part-2-regression> (검색일: 2022. 10. 15.)
  139. “How Can Smart Parks Help Get More People Outside?,” <https://www.youtube.com/watch?v=rG2AisQqvug> (검색일: 2022. 11. 24.)
  140. “Hyperparameter tuning a model with Azure Machine Learning (v1)”, <https://learn.microsoft.com/en-us/azure/machine-learning/v1/how-to-tune-hyperparameters-v1?source=recommendations> (검색일: 2022. 10. 15.)
  141. “LIGHTGBM 이란? 그리고 PARAMETER 튜닝하기”, <https://nurilee.com/2020/04/03/lightgbm-definition-parameter-tuning> (검색일: 2022. 10. 15.)
  142. “Machine Learning Algorithm Cheat Sheet for Azure Machine Learning Designer”, [https://learn.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet?WT.mc\\_id=docs-article-lazzeri](https://learn.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet?WT.mc_id=docs-article-lazzeri) (검색일: 2022. 10. 15.)
  143. “Smart Parks are here and to stay,” <https://www.bos.com/inspired/smart-parks-are-here-and-here-to-stay> (검색일: 2022년 11월 27일).
  144. “Tutorial: Designer - train a no-code regression model”, <https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-train-score> (검색일: 2022. 10. 15.)
  145. “Tutorial: Designer - train a no-code regression model”, <https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-train-score> (검색일: 2022. 10. 15.)

## ■ 데이터북

146. Liu, T.(2013), “Digital-output relative humidity & temperature sensor/module DHT22 (DHT22 also named as AM2302),” Aosong Electronics.

## 부 록

표 52 서울숲 관리 서울그린트러스트 사무총장 인터뷰 원본

번호	질문	목적	답변
1	스마트공원의 인식	관리자 입장에서 스마트공원의 현장 필요성 등	<ol style="list-style-type: none"> <li>1) 운영상의 어려움은 다소 있으나 반드시 필요함. 마차 대중교통인 버스의 [버스정보안내시스템]의 최대 수혜자가 사용자인 버스 승객인 것 과같이, 스마트공원의 최대 수혜자는 공원 방문자가 될 것임</li> <li>2) 공원 방문자는 편의성을 제공 받을 수 있음</li> <li>3) 관리시스템 도입 - 데이터입력 , 기록하는 과정, 편리하게 입력하는 과정</li> <li>4) 데이터 축적 : 이용자에게 공개</li> </ol>
2	현재 적용된 공원 내 ICT 기술	공원 현황 및 적용 기술 status 파악	<ol style="list-style-type: none"> <li>1) 무료인터넷 서비스의 경우에 원인을 알 수 없이 며칠씩 먹통이 되곤 해서, 사용자들이 많은 민원과 불편을 불러 일으켜 옴</li> <li>2) 디지털 사이니지(digital signage)</li> <li>3) 장애인용 시설 - 급속충전기, 점자 안내판 등</li> </ol>
3	현재 적용된 공원 내 IoT 환경센싱 기술		<ol style="list-style-type: none"> <li>1) 미세먼지 측정 (공기 질 측정기) -&gt; 성동구, 한양대학교 많은 기관에서 공기 질을 측정하고 이를 토대로 연구에 활용하고 있음</li> </ol>
4	현재 적용된 공원 내 인공지능 기술		<ol style="list-style-type: none"> <li>1) 현재는 적용된 기술은 없으나, 공원의 본 기능은 휴식, 경관 감상이나 코로나 시대 상황으로 인해서, 공간, 건물, 장소 등에 대한 스토리 텔링 정보를 모바일 앱으로 해서 제공하는 서비스에 대하여 진지하게 논의한 적이 있음</li> <li>2) 스마트 서울이라는 헬스케어 서비스를 2008년 한양대 병원과 추진하여 공원 방문객들의 산책 후 힐링 효과를 측정하기도 하였음</li> <li>3) 향후 공원에서의 정신, 신체 건강에 대하여 매일 건강검진을 통해서 헬스케어 기능에 대한 인공지능 부분이 고려되면 좋겠음</li> </ol>
5	보행량 측정으로 이끌어낼 수 있는 공원 관리 개선 효과		<p>보행량측정과 예측이 주는 공원 관리의 효용성</p>
6	유동 인구 및 밀집 지역 관리 가능성	<p>효용성</p>	<p>코로나 전후 오픈스페이스인 공원에 많은 사람이 방문하고 있음. 밀집상황에 대한 처리를 위하여 동선을 바꾸기도 하여 문제가 될 만한 상황을 되도록 만들지 않음</p> <p>현재 공유자전거 배치가 목, 금, 토, 일에 공원 쪽에 집중되고, 월~목까지는 도시 측에 집중되도록 배치하는 것을 알고 있다. 이러한 상황에 필요한 기술로 보임</p>

7	인공지능 보행량 예측 기술이 공원 관리에 미치는 기대 효과		방문자가 집중되는 시기, 시간, 공간 등을 사전에 체크할 수 있어 도움이 될 수 있음
8	도시공원이 미치는 공기 질 및 기후변화 방재 효과 인식	스마트 도시공원 방재 효과에 대한 정량화 필요성 인식	현재까지 기존 서울숲을 대상으로 연구하시는 논문, 보고서 등을 통해서 충분히 인지하고 있음 특히 도시 열섬현상에서, 온도 저감효과 등을 사이트별로 비교하여 도시공원의 개선 효과를 본 적이 있음

표 53 스마트공원 정책 입안 서울시 시의회 환경 입법조사관 인터뷰 원본

번호	질문	목적	답변
1	스마트공원의 인식	관리자 입장에서 스마트공원의 현장 필요성 등	<ol style="list-style-type: none"> <li>1) 현재 스마트정원, 스마트 팜이라는 용어를 많이 접하고 사용하고 있음. 하지만 스마트공원은 아직 낯선 상황</li> <li>2) 스마트라는 단어는 첨단 의미로 다가옴</li> <li>3) 기존 정원을 수직으로 세우고 관수등을 모바일 App 등으로 활용하는 스마트정원은 최근 미세먼지 저감, 실내환경 개선 등의 영향으로 확장 중</li> </ol>
2	현재 적용된 공원 내 ICT 기술	공원 현황 및 적용 기술 status 파악	<ol style="list-style-type: none"> <li>1) 일부 이동통신 서비스 제공업체와 협력하여 모바일 실시간 사용자 정보를 이용한 공원을 포함하는 주변의 보행량 정보를 사용하는 부분에 대한 정책 고려 중</li> <li>2) 개인정보 이슈 사용에 대한 우려 존재</li> <li>3) 성별, 나이 등 세부적인 정보에 따른 장단점 존재</li> <li>4) 하지만 데이터 송수신 기지국의 커버리지 단위가 매우 넓어, 사용자가 공원을 실제 방문하는지에 대한 여부는 불투명</li> </ol>
3	현재 적용된 공원 내 IoT 환경센싱 기술		<ol style="list-style-type: none"> <li>1) 미세먼지 측정 (공기 질 측정기) 는 매우 필요한 기술로 여겨짐</li> <li>2) 산림청 등에서 최근 미세먼지 저감과 관련한 많은 관심과 예산 배분이 이루어지고 있음</li> </ol>

4	현재 적용된 공원 내 인공지능 기술		1) 공원 내 범죄방지 및 사용자 보호를 위한 방문자의 이동을 쫓아서(Human tracking) 촬영하는 을 하는 AI 카메라 등의 적용
5	보행량 측정으로 이끌어낼 수 있는 공원 관리 개선 효과	보행량측정과 예측이 주는 공원 관리의 효용성	1) 연령대에 따라 대상자 카테고리를 노인, 부모, 청소년, 사회적약자 등으로 분류 2) 어린이 공원은 어린이 중심의 설계와 운영 3) 화장실은 가족 단위의 사용이 늘면서 가족 화장실을 사용 4) 선형 공원의 경우 상업지구를 끼고 있어, 음주 흡연 등의 일반인 사용자의 관리가 어려운 부분을 보행량 인식에 따른 개선할 수 있는 효과 기대
6	유동 인구 및 밀집 지역 관리 가능성		1) 혼잡상태에서 안전, 보안 요원 등 투입, 안전 가이드 제공 가능
7	인공지능 보행량 예측 기술이 공원 관리에 미치는 기대 효과		1) 사용자가 적은 도시공원의 경우, 미디어 파사드의설치여부가 이용자 분석 등을 통한 효용성의 정량화된 사전분석을 통해 의사결정을 하게 되면 많은 예산을 절감하거나 효율적으로 활용 가능
8	도시공원이 미치는 공기 질 및 기후변화 방재 효과 인식	스마트 도시공원 방재 효과에 대한 정량화 필요성 인식	1) 미세먼지 저감 2) 오랜 시간이 지나 작은 나무가 큰 숲이 되어 기후변화 방재 효과가 있음을 인지 3) 그늘막 설치로 인한 온도 저감 효과
9	스마트공원 기술 중 인공지능으로 이산화탄소 발생량과 환경변수와의 상관관계가 주는 효과 파악		1) 산림이 주는 효과에 대한 인식 필요 2) 도시공원의 숲이 주는 미세먼지, 이산화탄소 저감효과 등에 대한 정량화 필요
10	습지 보전과 같은, 도시공원 설계 입법 및 관리 기대 효과	도심지 유희용지 중 기후관리를 위한 노력	1) 근린공원, 역사공원 등 관련하여 조례를 통해 도시공원을 만들 수 있음
11	스마트공원 관련 Free discussion		1) 스마트공원 안에서의 통계가 중요 2) 데이터에 기준 한 정책 입안을 펼 수가 있음 3) 따라서 데이터 확보가 우선 필요

## Abstract

# Smart Park Implementation through AI modeling and IoT for Climate Change Response

– Focusing in the case of Seoul Forest Park  
and Sorae Ecological Park –

Ph.D. Dissertation

Kwon, Tae Kyung

Interdisciplinary Program in Landscape Architecture

Smart City Global Convergence

Graduate School

Seoul National University

A smart park is a new concept defined as a park that achieves substantial value using environment-related advanced digital technologies. In other words, the purpose of smart parks is to promote effective operation and maintenance by grafting ICT technology and to enhance user convenience through environmental improvement. In this study, the smart park is assumed to be a space for improving the environment in relation to climate change. For this purpose, the Internet of Things (IoT) artificial intelligence model was tested to find a suitable application direction in the future.

Global climate change is leading various efforts to reduce carbon emissions around the world. Among the various functions of

urban parks in this era of the climate crisis, heat disaster prevention and reduction of carbon dioxide and fine dust are particularly important. Therefore, to consider the response to climate change as one of the functions of urban parks, previous studies were reviewed. To analyze the effects, IoT equipment was manufactured and installed, and data were acquired over a specific period. Finally, conclusions were drawn through experiments and analyses with the following two research objectives.

First, data such as temperature, air quality, and the amount of walking by visitors in Seoul Forest Park were collected for about 9 months. The number of people who visited the park was counted and set as a target. The correlation between each environmental factor was analyzed using artificial intelligence. Regression analysis was performed with various modeling algorithms. We focused on the environmental factors (and their patterns) that showed a high correlation with the number of people counted through a voting ensemble, which exhibited the best performance among the modeling methods. As a result, feature importance was determined for six environmental data factors: temperature, hour, humidity, month, working day, and fine dust (PM). Feature importance was measured as the factor that had the greatest influence on the prediction of the number of people counted. As an evaluation index, the closer the coefficient of determination ( $R^2$ ) is to 1, the better the evaluation of the AI model. The voting ensemble, which had the highest model evaluation index, yielded the highest  $R^2$  (0.72). XGBoostRegressor, LightGBM, and RandomForestRegressor produced  $R^2$  values of 0.70, 0.69, and 0.53, respectively. Therefore, the voting ensemble model could be considered the optimal algorithm through the evaluation of the AI -learning

model. In the future, it is expected that the voting ensemble machine-learning model identified in this study will be used for quantitative prediction of the number of people counted by using environmental data as a management function of the smart park. Second, this study collected and used heatwave environmental data from the “reed forest” and “salt wetland” of Sorae Ecological Park in Incheon. These data were compared with environmental data such as temperature and humidity collected from the “POGURO crossroad,” an urban area. In doing so, the factors affecting the amount of carbon dioxide generated and differences according to the site were analyzed. As a result, the carbon dioxide emissions at the “POGURO crossroad” were measured to be 40 ppm on average and were up to 60 ppm higher than those at the “salt wetland” of Sorae Ecological Park. The temperature at the “POGURO crossroad” was constant and was about 1.5 °C higher regardless of the time and date. Regarding ultrafine dust (PM 2.5), the “POGURO crossroad,” located in the city center, measured a value about 3  $\mu\text{g}/\text{m}^3$  higher than that of the average “reed forest” and 4  $\mu\text{g}/\text{m}^3$  higher than that of the “salt wetland.” Likewise, for fine dust, the “POGURO crossroad” showed a value about 3.5  $\mu\text{g}/\text{m}^3$  higher than that of the average “reed forest,” and the 4  $\mu\text{g}/\text{m}^3$  difference compared to the “salt wetland” remained.

To analyze the correlation between environmental data and carbon dioxide generation in response to climate change, an AutoML machine learning model was used. By comparing the prediction performance of each machine learning algorithm through artificial intelligence analysis and identifying the highly correlated environmental factors analyzed through the stacking ensemble, which showed the optimal results. Comparative analysis by regression modeling was performed by dividing the

obtained data by those of the the patterns of “salt marsh” “reed forest” and POGURO crossroad” .

As a result, in the “salt marsh,” the amount of carbon dioxide generated was estimated by six factors in descending order: temperature (sr\_salty\_temp), gusts (sr\_salty\_wind\_gust), humidity (sr\_salty\_humi), atmospheric pressure (sr\_salty\_pressure), ultrafine dust (sr\_salty\_pm2.5), and time (hour). The  $R^2$  corresponding to the stacking ensemble was the highest (0.90). XGBoostRegressor, LightGBM, and RandomForestRegressor yielded  $R^2$  values of 0.88, 0.88, and 0.66, respectively. Through evaluation of the AI learning model, it was possible to identify the optimal algorithm.

In the “reed forest,” the six factors of humidity (sr\_galdae\_humi), gust (sr\_galdae\_wind gust), atmospheric pressure (sr\_galdae\_pressure), wind direction (sr\_galdae\_wind direction), time (hour), and temperature (sr\_galdae\_temp) had the greatest effect on carbon dioxide generation in descending order. The  $R^2$  of the applied stacking ensemble was the highest (0.918), followed by those of the XGBoostRegressor (0.915), LightGBM (0.89), and RandomForestRegressor (0.80).

Because of the limitations for installing environmental sensors in the downtown area ( “POGURO crossroad” ), data from some reed forests were used instead. As a result, the reed forest wind direction (sr\_galdae\_wind direction), temperature (sr\_pg4\_temp), fine dust at the POGURO crossroad (sr\_pg4\_pm10), solar illumination in the reed forest (sr\_galdae\_solar), humidity at the POGURO crossroad (sr\_pg4\_humi), and wind speed in the reed forest (sr\_galdae\_wind speed) were the six factors that had the greatest effect on carbon dioxide generation. Again, the  $R^2$  of the applied stacking ensemble was the highest (0.85), followed

by the XGBoostRegressor (0.83), LightGBM (0.79), and RandomForestRegressor (0.52).

Therefore, we expect that the optimal stacking ensemble machine-learning model identified in this study can be used for the quantitative prediction of carbon dioxide generation. This in turn could be useful for emission reduction measures based on environmental data, which could become a disaster prevention function for smart parks in the future.

Through this study, the economic benefit of an IoT method based on sensing environmental data was proposed. An optimal artificial intelligence machine-learning model was designed so that existing parks could be easily transformed into smart parks, which will make it possible to provide optimal service through the accurate prediction of the number of visitors, counted according to environmental variables. This is expected to give new value to urban parks by strengthening the disaster prevention functions of smart parks, including carbon reduction and heatwave prevention.

**Keywords :** Smart Park, IoT sensing, Environmental Data,  
Artificial Intelligence, Climate Change Adaptation

***Student Number :*** 2019-31781