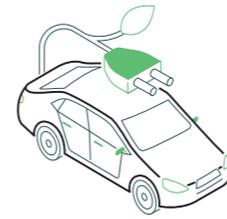




환경 데이터 우수 분석 사례집  
PDF 다운로드



# ENVdata

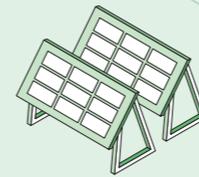
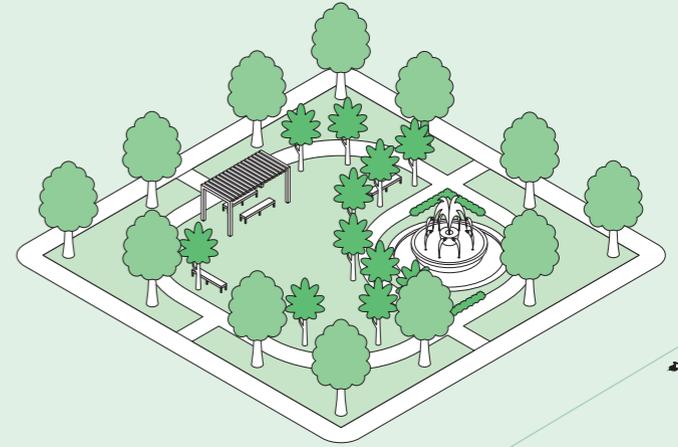
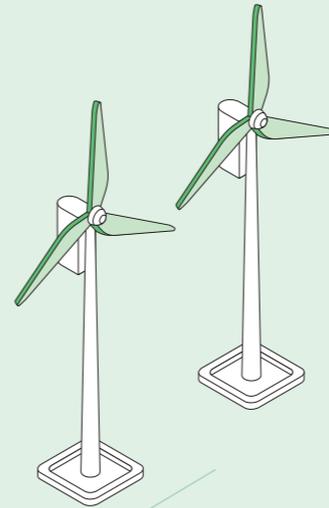
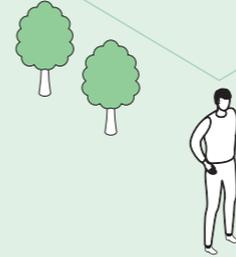
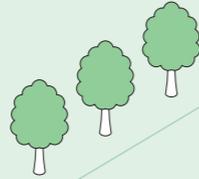
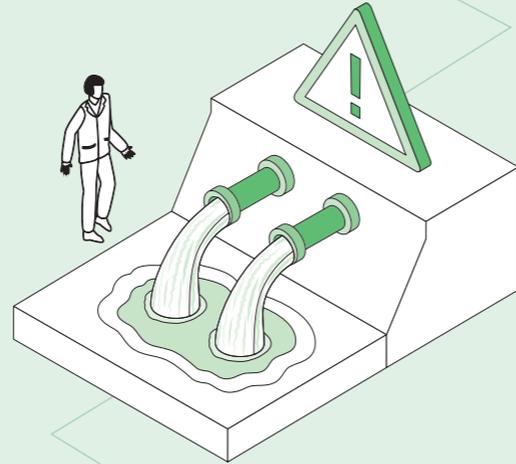
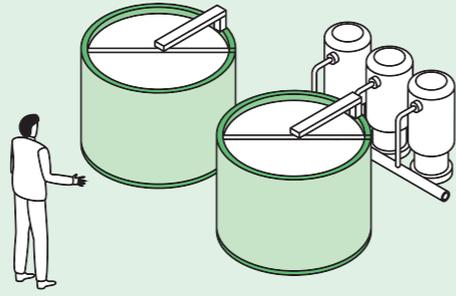
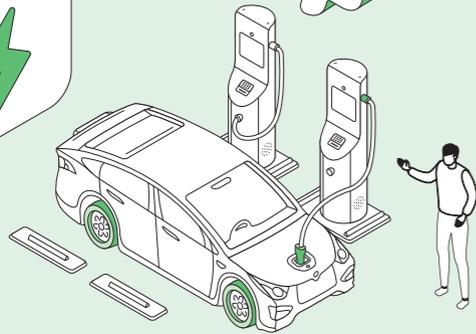
# 환경 데이터 우수 분석 사례집



# 디지털 전환

모든 것이 사라진다  
다만 데이터는 사라지지 않는다

# 환경 데이터



## 환경 데이터 우수 분석과제

비표적 분석기법과 머신러닝을 이용한 수질오염원 추적	6
머신러닝 기반 항공기소음 인식·분류모델 개발	8
도시생태축 공간계획을 위한 산림 연결성의 시각화	10
하수도데이터 머신러닝을 통한 하수도 지반침하 요인 분석	12
데이터 활용 지능형 공동주택 폐기물 관리 프로젝트	14

전기차 환경 빅데이터 분석·활용	16
야생조류 시(조류독감) 발생 분석	18
수질오염사고 위험도 분석	20
공공하수처리시설 슬러지 발생량 예측	22
수소차 충전소 입지 선정 분석 모델	24

01

# 비표적 분석기법과 머신러닝을 이용한 수질오염원 추적



## 목표

- ▶ 수질사고 원인 규명 및 수질오염원 추적을 위한 분류-예측 도구 개발
- 유입 하천별 오염물질의 분류 특성을 구분하여 오염 추적자를 발굴하고 오염원 원인자를 규명
- 수계별 부영양화 및 녹조 대증식(bloom) 현상을 일으키는 주요 물질들을 파악하여 수계별 맞춤형 관리대책을 마련

## 활용데이터

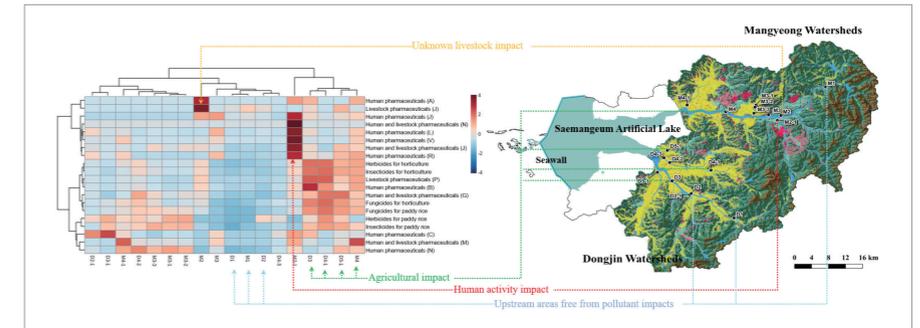
- ▶ 새만금수계 수질측정망(18개 지점) 월별 데이터('21. 3 ~ '22, 2 / 12개월)
- 만경강 본류(3지점) 및 유입 하천(전주천, 익산천, 탑천 등 6지점) 등 9개 지점
- 동진강 본류(3지점) 및 유입 하천(원평천, 신평천, 정읍천, 고부천 6지점) 등 9개 지점
- 분해능 70,000 이상의 초정밀 질량분석기를 이용한 화학물질분석 결과(약 1.3만 종)

## 분석방법

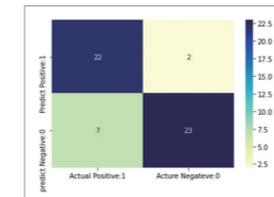
- **(통계적 방법)** 독성이 알려진 216종의 농약류, 의학물질 등에 대하여 통계 툴을 이용한 군집 분석(heatmap 이용)과 차원축소(주성분 분석 및 부분최소제곱판별법 이용) 방식을 이용하여 데이터 특성을 분류(classification)
- **(머신러닝 방법)** 수중에서 분해되는 모든 대사산물을 포함한 1만 3천여 종의 화학물질들을 머신러닝 툴을 이용하여 로지스틱 회귀, SGDClassifier, 결정트리, 랜덤 포레스트를 이용한 이진/다중 분류 지도학습 및 K-means를 이용한 비지도학습 실시
- **(지역적 분류)** 새만금호로 유입되는 오염물질이 어느 강과 하천을 따라 유입되고 있는지 분류해 낼 수 있는지를 분석
- **(계절적 분류)** 새만금호는 유역 면적의 50% 이상을 농경지가 차지하고 있어 농업활동 시기별로 계절적 차이가 뚜렷한 편으로, 새만금호로 유입되는 오염물질의 계절적 차이가 분류 가능한지를 학습
- **(수질영향 인자 탐색)** 새만금 유역의 상류는 2급수 내외로 맑은 수준이나 여러 지천이 합류 되면서 하류로 갈수록 5~6급수까지 수질이 악화하는 현상을 보여 수질별 분류가 가능한지를 학습하고 주요 인자(feature\_importance)를 도출함으로써 오염 추적자를 탐색
- **(조류 제어인자 탐색)** 조류 대번식을 제어할 수 있는 인자로 인(P, phosphorus)이 지목되고 있으나, 인의 유입경로 파악이 어렵고 실질적 관리방안이 미비한 실정으로, 머신러닝 방법으로 인과 상관성이 높은 화합물질들을 탐색
- 18개 지점에서 매월 총 12개월간 채취한 시료에서 검출된 214개에 달하는 화합물질의 농도를 한꺼번에 도시하기 위하여 화합물질을 용도에 따라 그룹핑하고, 이를 다시 개월 수로 묶어 heatmap으로 나타냄

## 분석결과

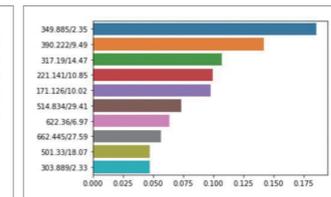
- PLS-DA 분석 결과, 농약류는 원예용(밭)과 수도용(논) 농약이 명확히 구분되는 반면, 의약품은 가축용과 인체용을 뚜렷하게 구분할 수 없었음
- PCA 분석 결과, 상류에서는 오염물질의 수가 적은 반면, 하류로 갈수록 오염물질의 양이 많아지고 있는 경향을 보이고 있음



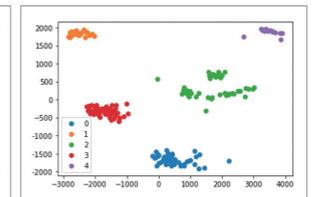
Heat-map of grouped by chemical usage



Binary classification with logistic regression



Importance feature top 10 from each model



Clustering with PCA analysis

## 기대효과

- 현재 낙동강수계에만 설치되어 운영 중인 수질측정센터를 한강·금강·영산강 수계로 확대할 계획으로, 인공지능을 이용한 오염원 추적기법 개발로 공공수역에 대한 체계적인 감시 기능이 강화될 것으로 기대

02

# 머신러닝 기반 항공기소음 인식·분류모델 개발

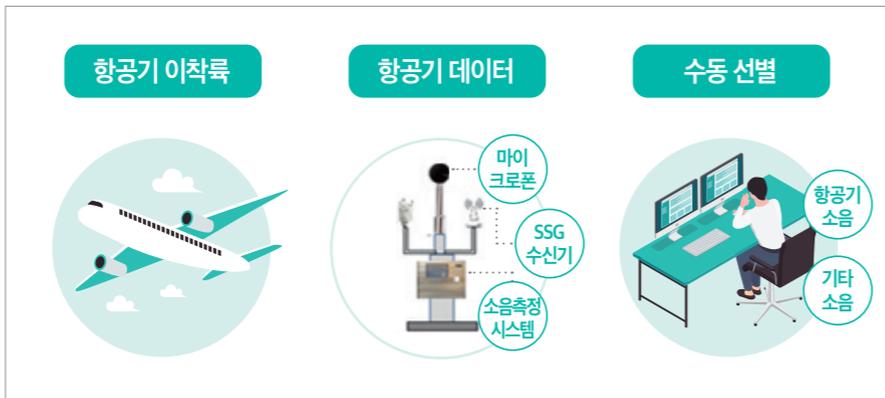


## 목표

- 항공기소음 융합 데이터를 머신러닝 기반 알고리즘을 활용하여 분석하여 선별 정확도가 떨어지는 지점에 적용성 여부를 평가하여 항공기소음 분류 자동화 기반 마련

## 활용데이터

- 국가소음정보시스템, 기상자료포털, 공항소음포털, 관제시스템 등을 통해 수집한 총 330종류의 데이터 19만 개를 융합하여 분석



## 분석방법

**▶ 항공기소음 인식·분류 관련 데이터 수집·검토**

지점	김포공항	광주공항	대구공항	제주공항	합계
데이터셋	59,205	24,928	56,164	50,095	190,392

\* 4개 공항에 대하여 사계절('21년, 1월~12월, '22년, 1월) 데이터 수집

**▶ 데이터 탐색 → 데이터 정제 → 이상치 제거 → 데이터 분할**

- (데이터 정제 및 이상치 제거) 학습 가능한 상태의 적절한 데이터 준비
- (데이터 분할) 학습데이터와 검증데이터를 7:3으로 무작위 추출

**▶ 적용 알고리즘 모두 항공기소음 분류에 효과적으로 구축**

- 4개 공항 모두 3개 알고리즘에 대하여 90% 이상의 정확도
- \* 기존 선별정확도가 낮은 2개 지점에 정확도가 1.5~2배 향상
- Decision Tree ⇒ Random Forest ⇒ XGBoost 순으로 효과적
- \* 효과의 차이가 적어 머신러닝 알고리즘 모두 항공기소음 분류 예측에 효과적

## 분류 알고리즘 성능 평가

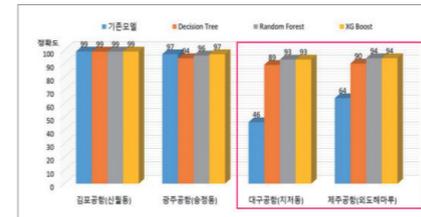
- F-1 Score 검토 결과 구축 모델의 분류 클래스는 균일
- \* F-1 Score 분포 : 전 지점의 모든 알고리즘에서 0.86~0.96 분포

구분	금번 구축 모델(개선 모델)			기존 모델
	Decision Tree	Random Forest	XGBoost	
김포공항 신월동	99%	99%	99%	99%
광주공항 송정동	94%	96%	97%	97%
대구공항 지저동	89%	93%	93%	46%
제주공항 외도해마루	90%	94%	94%	64%

## 통합 모델 구현

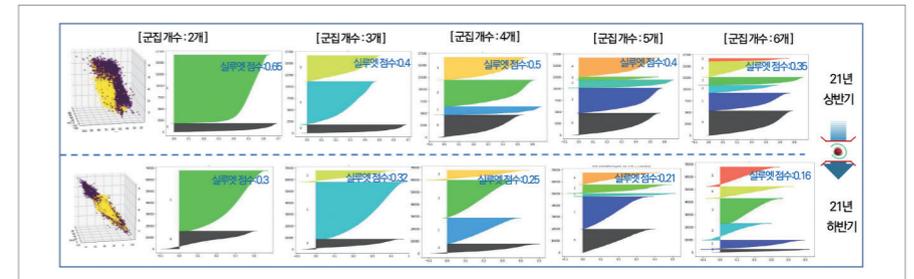
▶ 공항 지점별 통합 모델 구축

## 분석결과



구분	금번 구축 모델(개선 모델)			기존 모델
	Decision Tree	Random Forest	XG Boost	
김포공항 신월동	99%	99%	99%	99%
광주공항 송정동	94%	96%	97%	97%
대구공항 지저동	89% (▲)	93% (▲)	93% (▲)	46%
제주공항 외도해마루	90% (▲)	94% (▲)	94% (▲)	64%

항공기소음 분류 알고리즘별 정확도 비교(Decision Tree → Random Forest → XGBoost)



군집 분석(K-means Clustering)을 통한 모델 변화 감지

## 기대효과

- XGBooster 알고리즘을 통한 항공기소음 인식·분류 모델 정확도 94% 확인 및 항공기소음 자동선별 기반 확보
- 항공기소음 선별업무 자동화를 통한 업무 효율화 및 예산 절감 기대

03

# 도시생태축 공간계획을 위한 산림 연결성의 시각화



### 목표

- 핵심보전지역인 산림 기반 도시생태축 구축의 실효성 높은 계획 수립과 평가 등에 활용할 수 있는 산림 연결성의 정량적 데이터 생성 및 시각화

### 활용데이터

- 환경부 세분류 토지피복도(21년 기준, 17,705도엽)
- 국토지리정보원 시군구 행정경계(기준: 2022.5.25.)
- 국토지리정보원 수치지형도 인덱스(1:250,000, 26개 구역)
- 국토교통부 토지소유정보(기준: 2021.12.28., 39,390,253개 Record)

### 분석방법

$$\text{연결지수} = \frac{1}{A_{\text{total}}} (A_1^2 + A_2^2 + A_3^2 + A_4^2 + \dots + A_n^2)$$

$$\text{연결성(율)} = \frac{\text{연결지수}}{\text{영역면적}} \times 100$$

#### ▶ (분석항목1) 1km 육각형 격자공간 단위의 산림 연결성 분석

- 전국 산림의 단절·파편화 현상 정량화를 위해 102,356개의 육각형 격자망 내부에 존재하는 산림 면적(비율)과 연결성(연결 비율) 도출  
※ 연속 격자를 활용한 공간 분석에서 육각형이 공간적 연결 및 이동 분석에 적합함

#### ▶ (분석항목2) 산림 단절·파편화 복원효과 분석

- 특정 영역(1km 육각 격자) 내 100m 이내로 단절·파편화된 산림이 연결되면 얻을 수 있는 연결성(연결 비율)의 개선 효과 분석

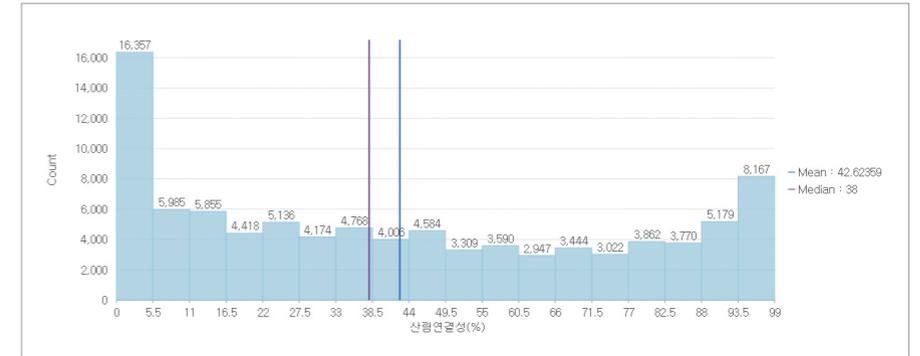
#### ▶ (분석항목3) 복원 가능지역 분석

- 특정 영역(1km 육각 격자) 내 산림이 아닌 지역의 국·공유지 중 생태복원이 용이한 지목이 차지하는 면적(비율) 분석

### 분석결과

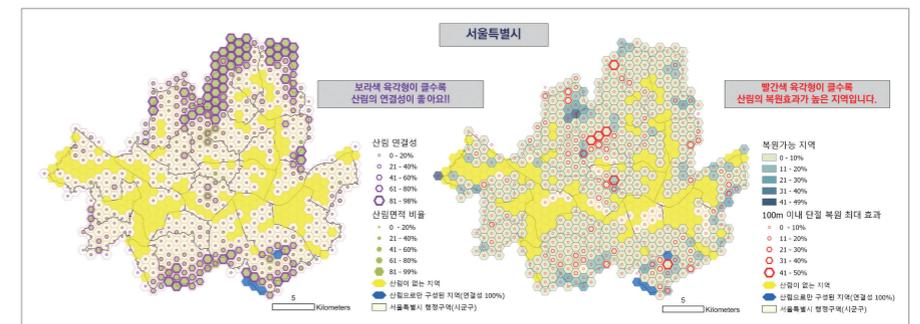
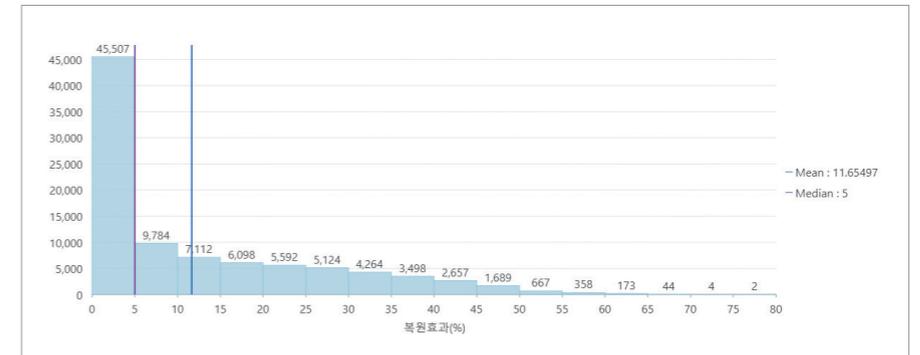
#### ▶ 토지피복 기반 전국 102,356km<sup>2</sup> 영역의 산림 연결성 분석 및 시각화 완료

- 주로 도시화 지역 등이 포함된 산림이 없는 영역은 4,463km<sup>2</sup>(4.4%)
- 단절·파편화 현상 없이 산림으로만 구성된 영역은 5,320km<sup>2</sup>(5.2%)
- 산림 단절·파편화가 있는 영역은 92,573km<sup>2</sup>(90.4%)로 연결성 평균은 약 42.6%이며, 중앙값은 38%로 나타남



#### ▶ 산림 단절·파편화가 있는 92,573km<sup>2</sup> 영역의 복원효과 분석 및 시각화 완료

- 해당 영역의 과반수가 복원효과 5% 이하이며, 50% 이상은 1,248km<sup>2</sup>로 나타남



### 기대효과

- 도시생태축 관련 공간계획 수립 시 지자체 단위 생태공간 연결성의 정량적 목표와 실천계획 제시에 활용 가능
- 실용적이고 일관된 도시생태축 정책 추진을 위한 표준데이터 생성과 제공 기반 마련

04

# 하수도데이터 머신러닝을 통한 하수도 지반침하 요인 분석



**목표**

- 노후하수관, 하수도 수질, 기상, 교통량에 대해 머신러닝을 실시하여 지반침하의 주요 요인을 파악해 선제적·체계적 관리

**활용데이터**

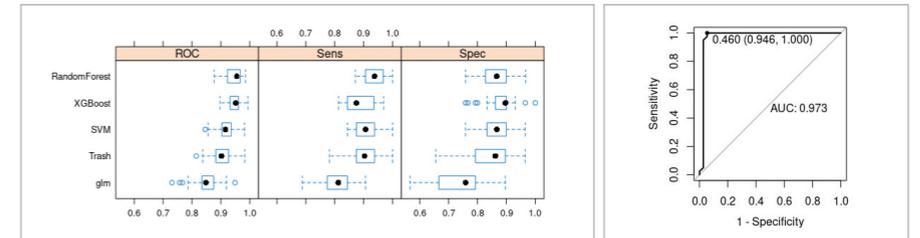
데이터명	세부내용
지반침하사고 발생 신고	사고 발생위치, 발생원인(하수도 침하)
종관기상관측 및 방재기상관측	지점, 평균 온도, 최저 온도, 최고 온도, 평균 상대습도, 평균 해면기압, 강수량
하수관로 데이터	연도별 하수관로 설비량
차량 주행거리 데이터	시도코드, 시군구코드, 행정코드, 전체 추정교통량, 승용차 추정교통량, 버스 추정교통량, 화물차 추정교통량
공공하수처리시설 현황	지역, 하수처리장명, 유입 BOD, 유입 COD, 유입 SS, 유입 TN, 유입 TP, 유입 대장균 수
찌꺼기 발생 현황	하수처리장명, 찌꺼기(톤)

**분석방법**

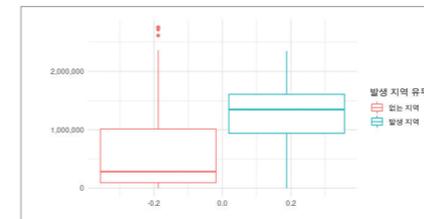
- ▶ **(차이분석) 노후하수관로에 따른 지반침하 발생 유무의 연관 관계를 확인하고자 차이검정 실시**
  - 독립표본 t검정을 실시하기 전, 데이터가 정규분포를 만족하는지와 등분산성을 가정하는지 확인  
※ Shapiro-Wilk 검정을 이용하여 정규성을 확인하고, Levene's Test를 이용해 등분산성 가정 확인
  - 정규성을 만족하지 못하면 비모수 방법인 Wilcoxon Rank-Sum Test를 실시
- ▶ **(머신러닝) 하수도, 기상, 교통 데이터를 전처리하고 데이터를 분류해 학습 데이터로 머신러닝 모형 형성**
  - 기상 데이터(평균 기온, 기온차, 기압, 강수량)로만 머신러닝을 실시해 만든 임의 모형을 예측 성능의 최소 기준으로 설정
- ▶ **(모형선택) 머신러닝 모형에 테스트 데이터를 적용하여 예측력(정확도, 민감도, AUC 등), 분석 시간, 분석 난이도 측면을 고려하여 최적의 모형 선택 및 변수 중요도 확인**
  - 기상 데이터로만 구성된 임의 모형(랜덤 포레스트 기법 사용)보다 예측력이 낮으면 모형 기각
  - 변수 중요도를 확인하여 지반침하 발생 주요 요인 확인

**분석결과**

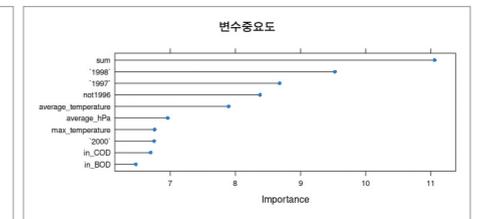
- **(결론1)** 지반침하 발생의 주요 요인인 노후하수관, 기온과 기압, 하수의 BOD 및 COD가 높은 지역을 중심으로 하수관을 점검 및 교체
- **(결론2)** 무분별한 지반침하 예방 사업 추진보다는 해당 분석 모형을 토대로 예측하여, 선제적·체계적 관리 필요



분석 모델 알고리즘별 분석 예측력 비교(Random Forest 알고리즘 모형 선택)



노후하수관로 연장에 따른 지반침하 차이검정 결과



하수도 데이터 머신러닝을 통해 분석한 모형의 변수중요도

**기대효과**

- 지반침하 발생에 영향을 주는 기온, 기압, 하수 수질(BOD, COD)을 지역별로 고려하여 지역 맞춤형 하수관 개발 및 교체
- 지역별 노후하수관에 분석 모형을 적용함으로써, 지반침하 확률을 계산하여 노후하수관 교체 우선순위를 계량적으로 평가 가능
- 유관기관이 제출하는 자료를 실시간으로 적용해 지반침하 발생 확률이 높은 지역에 선제적·체계적 관리 가능
- 분석 모형을 토대로 지반침하 예방사업 예산 효율성 제고, 대국민 안전 강화

# 데이터 활용 지능형 공동주택 폐기물 관리 프로젝트



## 목표

- ▶ 신규 데이터 분석·활용으로 안전한 폐기물 공공관리 체계 적극 구현
- 배출실적 신규 데이터 분석을 통한 오류 데이터 추출로 정확한 통계(전국 폐기물 발생 및 처리 현황 등) 산출에 기여

## 활용데이터

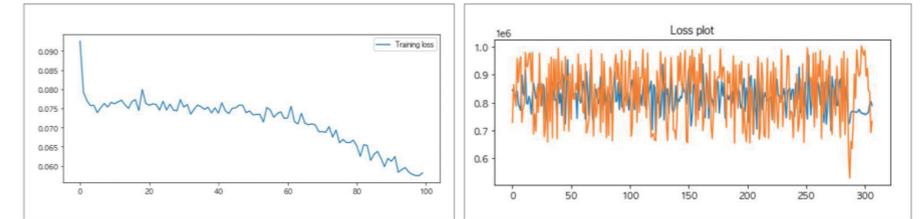
- 공동주택 생활폐기물의 품목별 배출 및 처리 신고 데이터

정보시스템명	순환자원정보센터(www.re.or.kr)		
운영기관	한국환경공단(순환자원정보부)		
시스템 개요	공동주택 등에서 발생하는 재활용폐기물의 전 과정 관리시스템		
	 공동주택	 수거업체	 지자체
	- 재활용폐기물처리 계약사항 신고 품목별 처리량 - 처리방법 신고(월별)	- 재활용폐기물처리 계약사항 확인 - 품목별 수거량, 처리방법 제출(월별)	- 관할구역 내 공동주택 재활용폐기물 실적관리(통계산출) 및 모니터링
분석데이터	<ul style="list-style-type: none"> <li>• 생활폐기물 배출자 처리실적 신고정보 : 월 평균 16만 건, 연간 약 192만 건</li> <li>• 수거업체 수거실적 입력정보 : 월 평균 5천 건, 연간 약 6만 건</li> <li>• 배출자/수거업체별 처리방법 입력정보 : 월 평균 47만 건, 연간 약 569만 건</li> <li>• 공동주택 등 배출자, 수거업체, 지자체 등 회원정보 : 1.5만 명</li> </ul>		

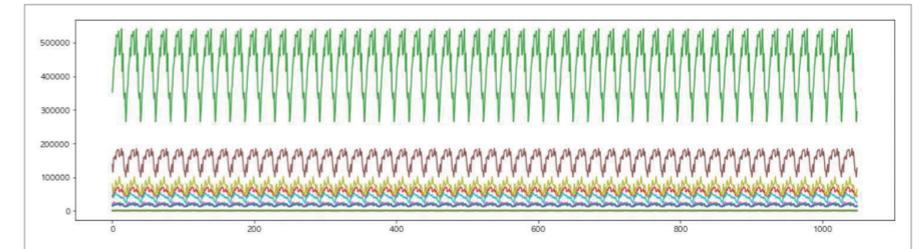
## 분석방법

- ▶ (분석과제1) 공동주택 재활용폐기물 신고 실적 데이터 적정성 분석
  - 지역별·품목별·세대수별 「오차범위 50% 통계자료」 생성
  - 처리실적 신고량이 평균 실적 신고량과 크게 오차가 나는 데이터 추출
- ▶ (분석과제2) 수거업체 실적 분석을 통한 수거위험 감지
  - 전월 수거실적 대비 50% 이하 감소 및 150% 이상 증가 데이터 추출
- ▶ (분석과제3) 공동주택 재활용폐기물 발생량 예측
  - 발생량 예측을 위한 데이터 분리 및 학습, 검증
  - LSTM 모델 생성
- ▶ (분석과제4) 공동주택 재활용폐기물 처리 흐름 분석
  - 재활용폐기물 품목별 관내 처리비율 산출
  - 재활용폐기물 품목별 지역별(광역시·도) 처리비율 시각화

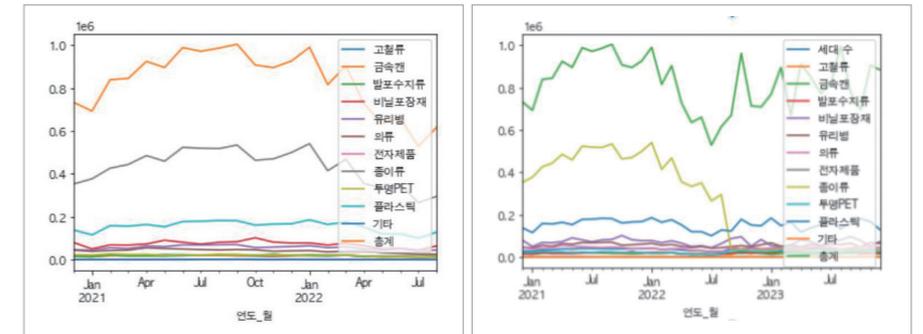
## 분석결과



모델 학습 결과



품목별 예측 모델링



품목별 폐기물 발생량 예측 결과(기준 - 예측)

## 기대효과

- 배출실적 신고 데이터 분석으로 지자체가 적극 대응할 수 있는 관리체계 구축
- 통계 데이터 신뢰도 확보 및 정보관리 사각지대 해소\*로 공공관리 강화 기여
  - \*'21년 이전 통계자료 작성 시 공동주택 재활용폐기물 발생량은 추정치로 관리
- ▶ **전국 약 2,617만명\* 입주민의 안전한 폐기물 관리 체계 마련으로 국민행복 실현**
  - \* 2021년 인구주택총조사 총 2,202만 가구, 아파트 가구비율 51.9%, 평균 가구원수 2.29명 적용

06

# 전기차 환경 빅데이터 분석·활용



## 목표

- 전기차 관련 환경 데이터 분석을 통해 전기차-충전기 보급정책 개선, 배터리 등 기술 발전 유인, 관련 시장 활성화 지원, 폐배터리 재활용 지원 등

## 활용데이터

1)전기차 등록 정보 (VMIS) → 2,541만 건(22.9월 말 전기차 357,912대의 누적 정보)

- 주요 정보내용 | 차종별 코드, 자동차종, 차명, 안전정보 등
- 자동차 등록 대수, 차량 제원 정보 등(46종)

2)전기차 검사 정보 (VIMS) → 823만 건(22.9월 말 전기차 357,912대의 누적 정보)

- 주요 정보내용 | 자동차별 검사 항목에 따른 검사 결과 및 이력 정보

3)배터리 필수 정보 → 2,489만 건(연간 취득 가능 건수, 2022년 기준)

- 주요 정보내용 | 전기차 배터리 기본 정보, 상태 정보, 진단 정보
- 전기차 배터리 기본 정보, 상태 정보, 진단 정보(59종)

4)충전 배터리 데이터 정보 → 1억 466만 건(연간 취득 가능 건수, 2022년 기준)

- 주요 정보내용 | 차량 배터리 SOC, 절연정보, 배터리 전압, 충전 전류, 충전량
- 충전기에 연결된 배터리 정보
- 양방향 통신 가능한 급속 충전기에서 데이터 수집 가능

5)실시간 전기차 주행 데이터 → 10만 건

- 주요 정보내용 | 전기차 동력 계통, 냉각 계통, 전력 계통, 운전 행태
- 모터 회전 수, 전류, 토크, 냉각수 온도, 배터리 셀 온도, 인버터 온도, 배터리 전압, 배터리 셀 전압, 배터리 전류, SOC, 충전 전압 등

6)실시간 전기차 충전 데이터 → 2천 건

- 주요 정보내용 | 전기차 충전 패턴, 배터리 상태
- 충전 일시, 충전량, 충전소, 차량모델, 차량번호 등 36개 항목

## 분석방법

### ▶ (분석항목 1) 전기차 충전 데이터 및 배터리 상태 정보 분석

- 전기차 충전 데이터 및 배터리 상태 정보 분석으로 전기차 성능에 영향을 미치는 인자 도출
- 충전 배터리 데이터 빈도·기술통계 분석으로 전기차의 충전, 방전 데이터 특성 파악
- 전기차 배터리의 환경 변수들에 대한 성능 평가 및 성능 분석

### ▶ (분석항목 2) 전기차 배터리 특성 계수 분석

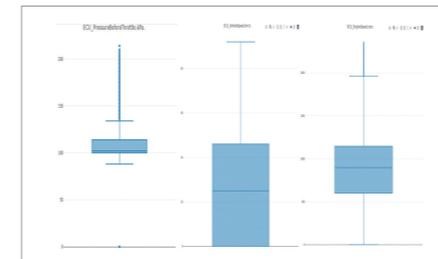
- 배터리 전압, 전류, 온도와 외부 환경정보(지형적 특성·습도·고도·주행패턴·최대 가속·평균 속도) 등으로부터 독립변수 분리, 상관관계 분석으로 배터리 특성 계수 산정

### ▶ (분석항목 3) 전기차 상태 진단 및 고장 원인 분석

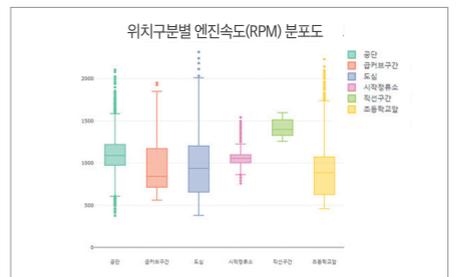
- 실시간 전기차 주행 데이터·충전 배터리 데이터·배터리 필수 정보 분석으로 전기차 상태 진단
- 급가속·급출발·급정지·급감속 등 위험 운전 패턴 분석으로 고장 관련 상관성 분석

## 분석결과

### ▶ 기술통계 및 상관관계 분석



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	100.0	102.0	111.1	114.0	214.0
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	0.00	25.00	25.78	46.00	92.00
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	601.1	897.9	911.3	1145.9	2352.0



<차량 상태변수의 기술통계 및 box plot(변수별, 위치구분별)>

- 분석의 모든 과정에서 데이터 파악을 위해 기술통계를 사용하며, 시점 통일 시 시간 최소 단위 (시, 분, 초, 일 등)에 맞추어 스케일링 시 영향을 주는 이상치 판단 및 제거 기준지표로 이용함
- 고장의 원인을 파악하기 위하여 데이터 통합, 전처리, 시점통일이 끝난 버스 데이터를 예시로 기술통계 및 변수 간 상관관계의 결과를 도출
  - 상기 변수 이외에 차량별/계통별로 관측되는 차량상태 변수와 기상변수, 이상판단 통계량, 차량검사 이력 변수 등 변수간의 상관관계(관련성)를 파악함
- 이상치 탐지를 통해 얻어진 이상치 빈도 데이터를 이용하여 이상치 빈도와 상관성이 높은 변수를 대상으로 발생시점(비정상) 전후의 변수값의 비교분석, 평균차이 검정 수행으로 원인 파악

## 기대효과

- 전기차 충전 시 배터리 등 전기차의 상태 정보를 분석하여 고장 등 전기차 이상 여부를 예측하여 차량 관리에 활용, 주행 중 운전 습관과 충전 습관을 분석하여 경제 운전과 안전운전에 활용

07

# 야생조류 AI(조류독감) 발생 분석



## 목표

- AI 발생과 환경변수(기상정보, 자연환경정보, 조류 밀집도 등) 간의 상관분석을 통해 거점별·연도별 AI 발생 특징을 분석하고 AI에 감염된 조류종의 생태 데이터(서식지, 먹이 등)를 활용하여 서식지 분포 확률모델을 개발

## 활용데이터

데이터명	출처	비고
야생조류 AI 검사내역	국립야생동물질병관리원	2019~2022
국내 철새도래지 위치 데이터	행정안전부	오픈데이터
종관기상관측 데이터	기상청	
해외 AI 보고 데이터	FAO(유엔 농업기구)	
전국 수치표고모델	국토지리정보원	
전국수계망 위치 데이터	국토교통부	
토지피복도 중분류 데이터	환경부	
주거 인구 데이터	(주)biz GIS	

## 분석방법

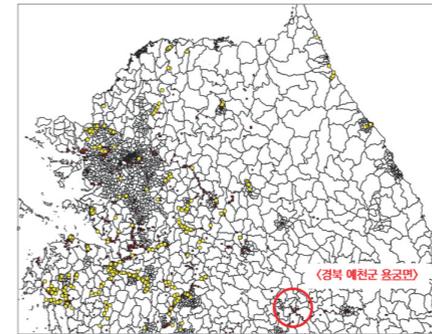
- 과거 데이터를 통해 AI 전파 위험이 큰 야생 조류종 특정
- 해외 AI 보고 데이터와 국내 발생 데이터를 비교하여 AI 발생 주요 영향인자 도출
- 주요 영향인자를 국내 데이터와 비교하여 지정 거점(87개)별 특성 및 고위험군 분류
- 환경요인을 기반으로 Maxent(서식지 예측 모델)를 활용하여 전파 고위험 조류종의 서식지를 예측

## 분석결과

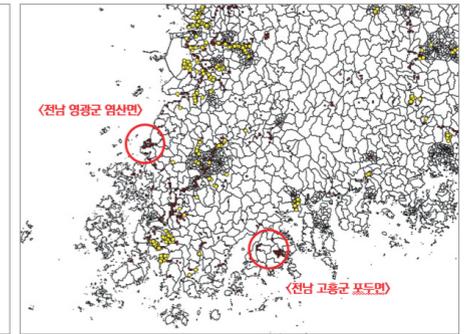
- AI 발생 패턴과 상관이 높은 변수 도출

데이터 분류	변수명	AI 양성 건수와 상관 계수
해외 AI 보고	러시아 AI 보고 건수	0.6065
	30cm 지중온도(3주 지연)	0.4307
종관기상관측	1시간 최대 강수량(2주 지연)	0.5325
	평균 증기압(3주 지연)	0.4522

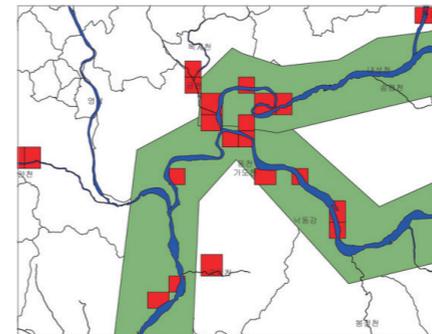
- 예찰지역 외 고병원성 AI 감염 조류 서식의 의심 지역 도출



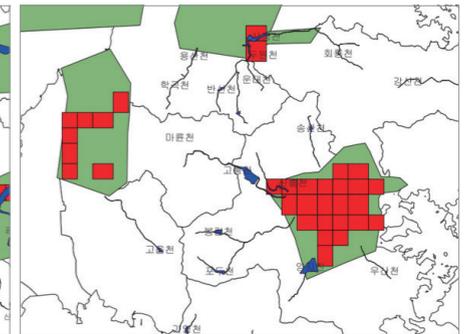
예찰지역 이외 AI 감염 조류 서식 확률 높은 지역 (경북)



예찰지역 이외 AI 감염 조류 서식 확률 높은 지역 (전남)



예찰지역 이외 AI 감염 조류 서식 확률 높은 지역 (용궁면)



예찰지역 이외 AI 감염 조류 서식 확률 높은 지역(포도면)

## 기대효과

- 현재 지정된 거점 이외에 AI 감염 조류 서식지 분석을 통해 도출한 지역을 추가 거점으로 활용하여 감염 대응력 강화
- 분석 결과와 실관측 데이터를 비교하여 결과 검증 및 정확도 개선 기대
- 기존 예찰지역 외 감염 여부를 확인할 수 있는 지점 추가를 위한 후보지 선정에 활용

# 수질오염사고 위험도 분석



## 목표

- 수질오염사고 위험지수 개발, 위험도 분석 및 지도 시각화를 통한 방제 활동 정책의 결정자료로 활용

## 활용데이터

데이터명	출처	시점	공간
수질오염사고 데이터	한국환경공단	2015~2020	상세 주소
전국행정구역지도 데이터	행정안전부	2022	시군구
전국오염원 조사 자료 (생활계, 축산계, 산업계, 토지계, 양식계)	국립환경과학원	2019	읍면동, 리
전국오염원 조사 자료 (기타수질오염원)			
개인하수처리시설관리업 사업장 데이터			
산업단지현황 조사 데이터	한국산업단지공단	2022	시군

## 분석방법

1. 데이터 수집	2. 데이터 전처리	3. 데이터 분석	4. 결과 도출
<b>현황 데이터</b> <ul style="list-style-type: none"> <li>• 수질오염사고 데이터</li> <li>• 사고 방제조치 데이터</li> <li>• 사고 원인 데이터</li> <li>• 산업단지현황 조사 데이터</li> </ul>	<b>공간정보 데이터 전처리</b> <ul style="list-style-type: none"> <li>• 시군구 행정구역 공간벡터를 활용한 사고데이터 매핑</li> <li>• 수질오염사고 주소 정제 및 지오코딩</li> </ul>	<b>수질오염사고 현황 분석</b> <ul style="list-style-type: none"> <li>• 수질오염사고 유형별/원인별 GIS 기반 발생지 차이 분석</li> <li>• 사고유형 및 사고지점, 사고피해 연관 분석</li> <li>• 수질오염사고 및 인근 환경 데이터 간 상관관계 분석</li> </ul>	<b>사고현황 및 위험도 시각화</b> <ul style="list-style-type: none"> <li>• 사고현황 시각화(예시)</li> <li>• 사고 위험도 시각화(예시)</li> </ul>
<b>오염원 데이터</b> <ul style="list-style-type: none"> <li>• 하수처리시설 데이터 (개인, 공공)</li> <li>• 전국오염원 조사 자료 (2019)</li> </ul>	<b>분석 데이터셋 구축</b> <ul style="list-style-type: none"> <li>• 시군구별 사고횟수 생성</li> <li>• 수질오염 위험시설 데이터 매핑</li> <li>• 인근 하천 데이터 매핑</li> <li>• 기타 수질오염원 데이터 매핑</li> </ul>	<b>수질오염사고 위험지수 산출</b> <ul style="list-style-type: none"> <li>• 시군구별 사고횟수 및 유형 연계분석, 위험도 산출</li> <li>• 산출 위험지수 타당성 분석을 통한 오염사고 인근 환경 기준치 조정</li> </ul>	
<b>공간정보 데이터</b> <ul style="list-style-type: none"> <li>• 시군구 행정구역 지도 데이터</li> </ul>			

### ▶ 수질오염 사고 현황 대시보드 제작

- 한강수계 수질오염사고 방제조치 데이터를 통한 지역별, 유형별 현황을 분석하여 그래프로 시각화

### ▶ 수질오염사고 발생 위험지수 개발

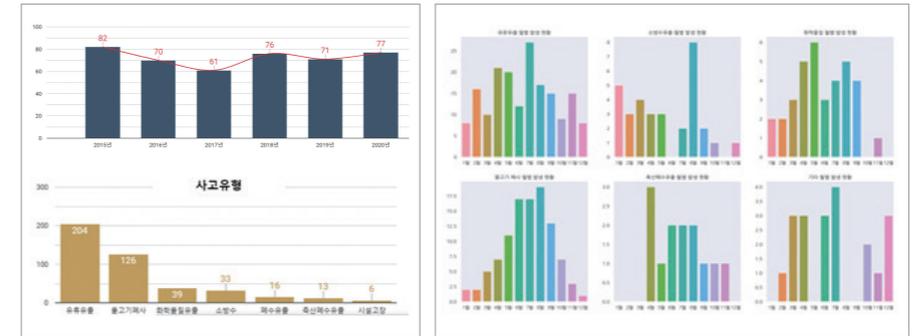
- 화학물질안전원의 위험물질 관리 위험도 산출 방법론을 인용하여 수질오염사고 횟수 기반 위험지수 및 환경 기반 위험지수 개발
- 지수에 따른 위험도 분류 및 최종 위험지수 산출

### ▶ 수질오염사고 위험도 지도 시각화

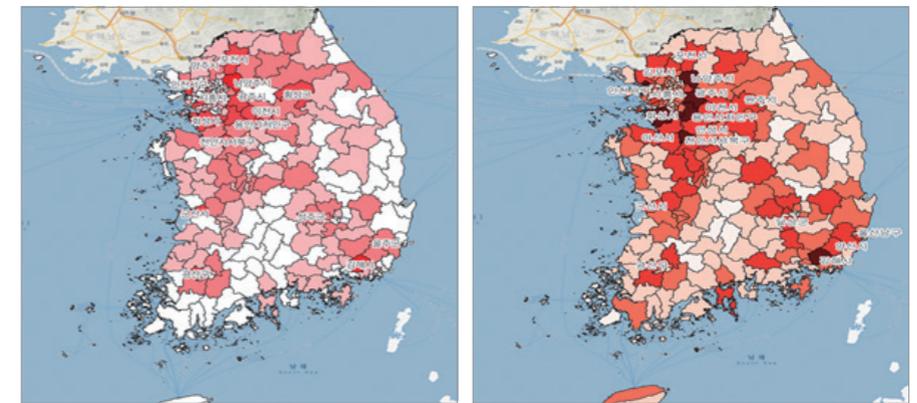
- QGIS를 활용한 시군구 행정구역 단위 수질오염사고 위험도 지도 시각화

## 분석결과

### • 지역별·유형별 수질오염사고 현황 및 요인 분석



### • 전국 행정구역 지도를 기반으로 최종 위험도 5 이상 행정구역 지도 시각화 및 실제 사고 다발지역 비교



## 기대효과

- 사고빈도와 지역별 환경요인을 분석한 사고발생 가능성을 통합한 위험도를 산출하여 예방대책 마련 업무에 활용
- 위험도 기반의 방제활동 관리체계를 확립하여 효율적인 방제활동 지원을 위한 지표로 활용

# 공공하수처리시설 슬러지 발생량 예측



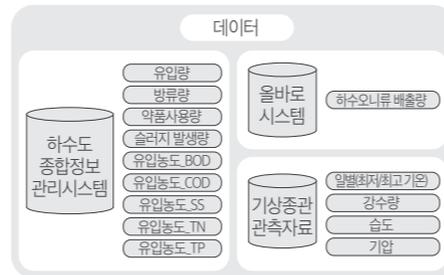
## 목표

- 하수도종합정보관리시스템 등의 사업장 빅데이터와 기상청 기상관측자료 등 외부 데이터를 활용한 머신러닝 알고리즘 기반의 공공하수처리시설 슬러지 발생량 예측을 통해 비정상 운영 위험군 분류
- 과년도 데이터 세트를 구축하고 시계열 분석을 통해 모델을 학습한 후, 실제 슬러지 발생량과 예측된 슬러지 발생량의 오차율(절댓값)을 이용해 위험일자 분류

## 활용데이터

- 일일 처리 유량 700톤 이상의 시설용량과 일단위 분석가능한 90여 개 공공하수처리시설 데이터

시스템	세부내용
올바로시스템	하수 오니류 배출량, 사업장 기초정보
하수도종합정보관리시스템 (Soosiro)	유입량, 방류량, 유입농도 (BOD, COD, TN, TP, SS), 약품 사용량, 슬러지 발생량
기상청 데이터포털 (KMA)	일별(최저/최고 기온) 시간별(강수량, 습도, 기압)



[그림] 슬러지 발생량 예측 활용데이터 및 출처

## ▶ 모델 설계

- 인자별 변수 중요도 확인 결과 유입량(약 25%)과 약품 사용량(약 21%)이 슬러지 발생량 예측에 중요한 변수임을 확인
- 알고리즘별 성능 비교 결과 Random Forest Regressor 알고리즘의 MAPE와 RMSE의 성능 가장 우수 → 슬러지 발생량 예측 알고리즘으로 선정

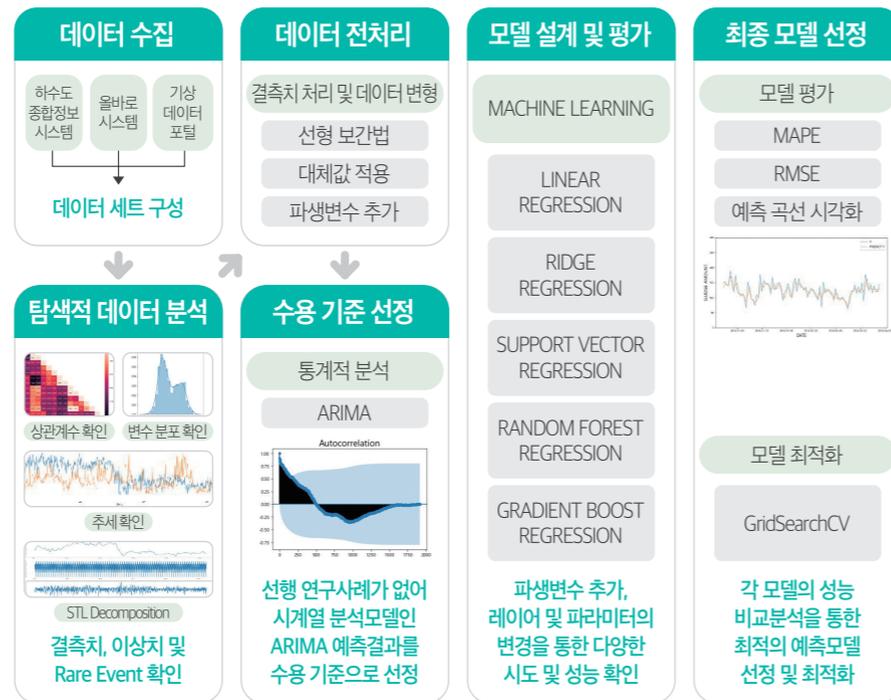
## ▶ 발생량 예측

- 시설별 과거 약품 사용량 대비 슬러지 발생량을 학습하여 슬러지 발생량 예측
- 공공하수처리시설별 과거 데이터('14~'17) 학습 후, 2018년도 이후를 기준으로 일별, 월별 슬러지 발생량 예측

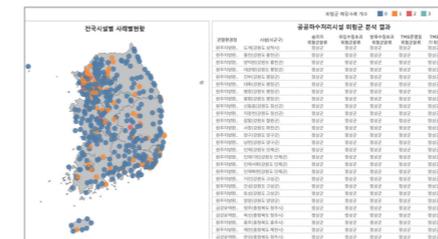
## ▶ 예측값을 이용한 위험일자 분류

- 예측구간\*을 이용한 통계적 검정을 통해 위험일자 분류
- \* 예측구간 : 하나의 예측값에 대하여 일정한 확률범위(확률적인 근거 하) 내에서 정해지는 상한과 하한에 의한 범위를 의미, 모델의 정확도가 충분히 검증되지 않은 상황에서는 예측구간을 사용하는 것이 바람직함
- 위험으로 분류된 일수가 한 분기에 35일 이상인 경우 해당 시설을 위험군으로 판정

## 분석방법



## 분석결과



## 기대효과

- 데이터 기반의 과학적 지도점검으로 환경행정 투명성 및 효율성 제고  
※ 비정상 운영이 이루어질 가능성이 있다는 분석으로 확정적인 것이 아니며 점검 대상 선정을 위한 보조자료로 활용
- 시스템별 산재된 사업장 데이터의 교차검증 및 신뢰성 확보

# 수소차 충전소 입지 선정 분석 모델



## 목표

- 수도권(서울, 경기, 인천)을 대상으로 법적 규제(교육환경법, 철도안전법 등)에 따른 입지제한 기준과 개발/환경적 기준을 모두 고려한 입지선정 기법을 개발하여 수소충전소 설치 후보지역 평가

## 활용데이터

- 서울, 경기, 인천지역 데이터

데이터	구분	기준
연속주제도_국토계획/도시지역 중 "UQA111", "UQA112" 레이어 추출	용도지역 건축제한 기준	제1종, 제2종 전용 주거지역과 교차하는 부지 제외
POI 데이터 구매	주택건설 기준	공동주택, 어린이 놀이터, 의료시설, 유치원, 어린이집 및 경로당으로부터 50m 이내 교차하는 부지 제외
연속지적도상 지목이 '학인' 데이터 추출	교육환경 보호 기준	학교경계(학교 예정부지 포함)로부터 직선거리 200m 이내 교차하는 부지 제외
연속수치지형도 '철도경계' 레이어 추출	철도안전법 기준	철도경계선(바깥 궤도 끝 선)으로부터 30m 이내 교차하는 부지 제외
한국보호지역 통합 DB(KDPA) 추출	국가 보호구역	보호구역과 교차하는 부지 제외
연속지적도상 Shape_Area 추출	여유부지 필요 면적	2,000㎡ 이내

## 분석방법



## ▶ 후보지 검토

자연보전 측면	생태자연도 1등급·별도관리구역 - 개발적성평가 등급과 관계없이 생태자연도 1등급 또는 별도관리구역에 해당하는 지역은 원천적으로 배제
개발비용 측면	LPG 충전소 - 충전소 부지 확보 및 미래차 복합충전소 구축 확대를 위해 LPG 충전소가 기 설치된 부지 적극 활용
개발수요 측면	인구 밀도 - 수소차 등록/운용이 높은 지역 또는 인구 밀도가 높은 지역을 대상으로 우선 설치 - 공공시설물이 인접한 지역에 우선 설치

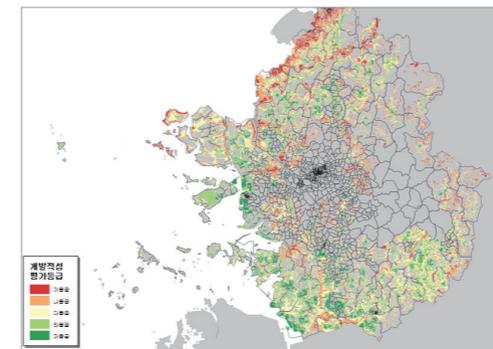
## ▶ 입지선정 모델 제작

- 친환경 수소차 충전소 부지 선정을 위하여 자연보전, 소요비용, 수요 등을 고려한 빅데이터 분석
- 수도권(서울, 경기, 인천)을 대상으로 통합입지평가 및 개발적성평가 수행

## ▶ 부지 선정 및 시각화

- 30만여 개 부지를 대상으로 개발적성평가 등급 산정
- 수요인원 및 설치(건설) 비용을 반영한 개발적성평가 등급을 통해 최종 후보지 GIS 시각화

## 분석결과



수도권의 수소차 충전소 입지 선정 결과

## 기대효과

- 수소충전소 설치 후보지역 평가를 통해 수소충전소 보급을 위한 先계획 後개발 관리체계 구축 지원



# 데이터 기반의 스마트한 환경행정을 실현합니다

환경부와 한국환경공단은 과학적이고 합리적인 환경행정 실현을 위해  
「환경정보 융합 빅데이터 플랫폼」을 구축하고, 환경 데이터 포털 서비스(ENVDATA)를 통해  
데이터 기반의 환경행정 혁신과 서비스 구현을 위한 양질의 환경정보를 제공하고 있습니다.

## 주요 사업



환경 데이터 제공,  
활용을 위한 사용자  
분석환경 제공



환경정보 융합  
빅데이터 플랫폼  
구축·운영



환경 데이터 활용  
촉진을 위한  
정책지원 및 역량교육



환경 데이터 플랫폼  
협이기구 구성·운영



환경 데이터 연계·수집·  
표준화 및 품질관리



환경 데이터 활용과제  
발굴·수행

