



강원도 산불 예측 및 피해 최소화 프로젝트

- 머신러닝과 딥러닝을 활용한 모델 개발 -

성열민, 김범모, 배상일, 오성준

CONTENTS

1 프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

2 데이터 탐색

FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

3 통계 분석

분석 개요
분석 내용
분석 결과 해석

4 ML 모델링

산불 예측 모델 생성 및 학습
모델 성능 검증
산불 위험 예측

5 DL 모델링

산불 분류 모델 생성 및 학습
모델 성능 검증
산불 이미지 분류 예측

6 서비스 및 문서

서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고문헌 / 부록

Chapter 1.

프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

데이터 탐색
FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

통계 분석
분석 개요
분석 내용
분석 결과 해석

ML 모델링
산불 예측 모델 생성 및 학습
모델 성능 검증
산불 위험 예측

DL 모델링
산불 분류 모델 생성 및 학습
모델 성능 검증
산불 이미지 분류 예측

서비스 및 문서
서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고 문헌 / 부록



프로젝트 개요

데이터 탐색

통계 분석

ML 모델링

DL 모델링

서비스 및 문서

팀 구성



성열민

- 데이터 전처리
- EDA / 통계
- Machine Learning & Deep Learning



김범모

- 데이터 전처리
- 통계
- Machine Learning & Deep Learning



배상일

- 데이터 전처리
- EDA
- 지도시각화 (Python)



오성준

- 데이터 전처리
- EDA
- 지도시각화 (Python)

주제 선정 배경

제안요청서

국립산림과학원 기술용역 입찰 공고

- 입찰공고번호 : 20230522-00
- 공 고 명 : 강원도지역 산불발생예측 및 산불분류모델 개발 용역
- 수 요 기 관 : 국립산림과학원



사업목적

- ✓ 강원도 산림지역 **선제적 산불 대응**
- ✓ 산불피해현황 파악 및 대응체계 분석
- ✓ 산불예방 업무에 적용할 **모델 구축**



제안 개요

- ✓ 과업명 : 강원도지역 산불발생예측 및 산불분류모델 개발
- ✓ 사업기간 : 23.05.22 ~ 23.06.23
- ✓ 사업비 : 84,000,000원



주요과업

- ✓ 산림청 / 기상청 API 공공데이터 활용
- ✓ 강원도 산불피해현황 및 기상요인 파악
- ✓ **공간정보 분석 / 지도시각화**
- ✓ **산불 발생 예측 / 산불 분류모델 개발**
- ✓ **모델 성능평가 / 산불 위험 예측**



기대효과

- ✓ 산불예방 및 **현장 대응체계 고도화**
- ✓ ‘**진화자원 배치 의사결정시스템**’ 활용
- ✓ **선제적 산불 예방체계 마련계획 수립**

주제 선정 배경

강원도 대형산불

▣ 강원도민일보

[강원산불] '양간지풍' 불면 순식간에 초토화...대형산불 원인

지난 4일 오후 발생한 강원 동해안 일대 산불로 여의도 면적(290ha)을 웃도는 산림 360 ha가 잿더미가 됐다. 현재까지 1명이 숨지고, 11명이 부상하는...

2019. 4. 5.

▣ MBC뉴스

울진·삼척 산불 확산 '초비상'. [재난사태 선포] '역대 4번째'

어제 오전 경북 울진에서 시작된 불이 강한 바람을 타고 강원도 삼척까지 번졌습니다. 산림청과 소방당국이 진화작업에 총력을 기울이고 있는데,...

2022. 3. 5.

▣ 연합뉴스

[동해안 산불] 피해면적 [1만4천222ha]..역대 2번째, 2000년 이후로는 최대

(울진=연합뉴스) 김현태 기자 = 울진·삼척산불 사흘째인 6일, 경북 울진군 북면 일대에 불길이 지나간 곳들이 검게 타버린 모습을 보인다.

2022. 3. 6.



2019년 4월 4일 오후 강원 동해안 일대 산불로 360㏊가 잿더미가 되었다. 이처럼 영동지역에 불이 불으면 대형산불로 번지는 이유는 **양간지풍**의 영향이 크다고 한다.

2022년 3월 4일 경북 울진군 북면 두천리에서 시작된 불이 강원도 삼척까지 번져 주민 4,600명이 대피하고 주택 75채가 불에 탄 **대형 산불**이며 피해 규모가 커지자 정부는 강원과 경북에 4번째 **국가 재난 사태**를 선포했다.

2022년 3월 4일 강원 강릉에서 동해까지 번져서 **피해면적이 1만 4222㏊**로 역대 두번째 규모였던 산불이다. 화재 당시 **최대 풍속이 23.7㎧**였고 화재로 인한 이재민 수는 850명으로 집계되었다.

조사 결과 매년 강원도에서 대형 산불이 발생 하였고 발생 빈도도 강원도가 가장 높다는 것을 알 수 있었다.

주제 선정 배경

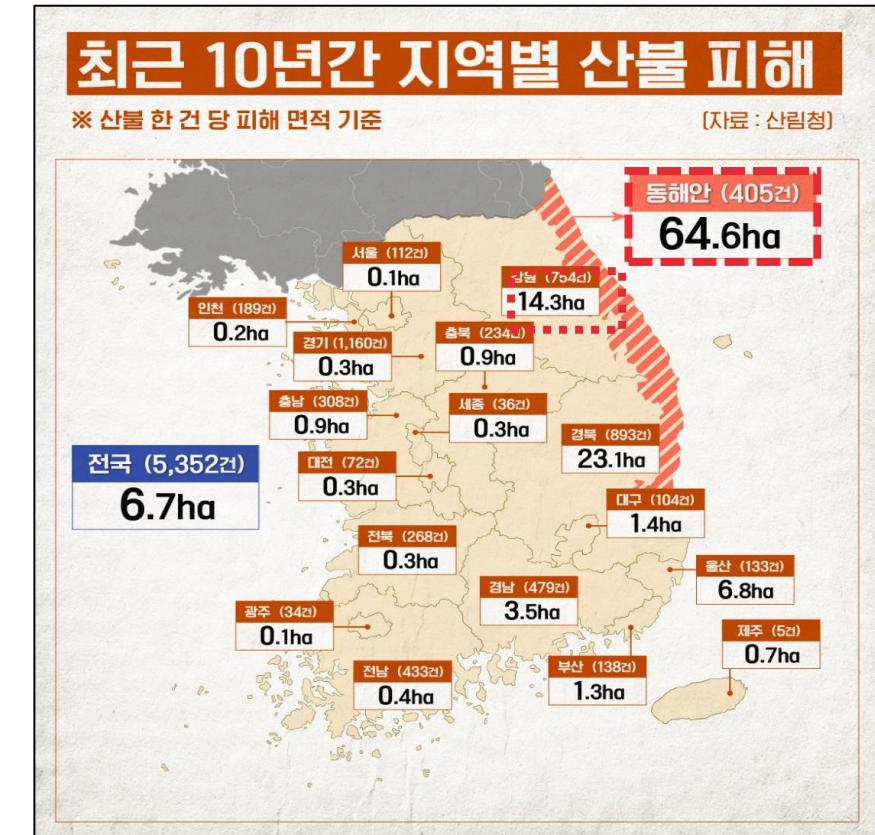
강원도 지역의 지형적 특징

Point. 동해안 지역에 집중된 산불 발생으로 인한 피해 산림 면적 증가

- ✓ 경북 ▶ 강원 순으로 건수별 피해 면적 크기가 크다.
- ✓ 태백산맥을 중심으로 대형 산불은 영동지방에 집중된다.
- ✓ 전국 피해 면적에 비해 동해안 피해 면적이 약 10 배 크다.

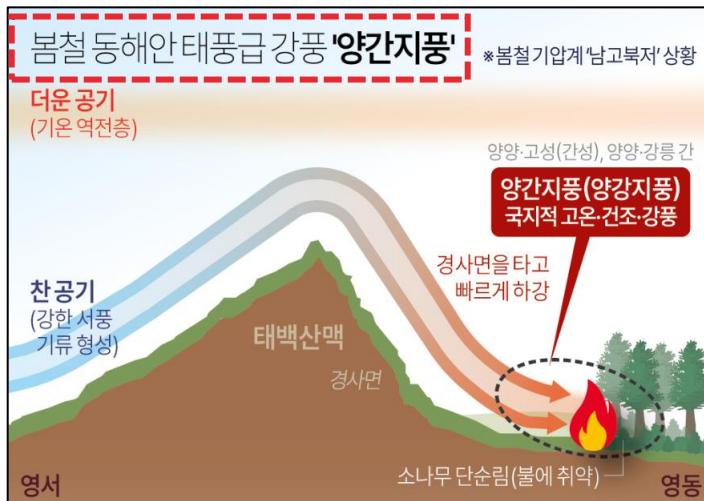
Point. 역대 주요 대형 산불 다수가 강원도에서 발생

- ✓ ‘국립기상연구소’의 발표에 따르면, 강원도 동해안 지역은 한번 불이 붙으면 짧은 시간에 큰 피해가 번지는 지역이다.



주제 선정 배경

양간지풍



실효습도

$$H_e = (1 - r)(H_0 + rH_1 + r^2H_2 + r^3H_3 + r^4H_4)$$

r = 실효습도 계수 (0.7) H_0 = 당일 상대습도
 H_e = 실효습도 H_n = n일 전의 상대습도

Point. 고온, 건조, 강풍의 특성을 띠고 봄철 대형 산불의 원인

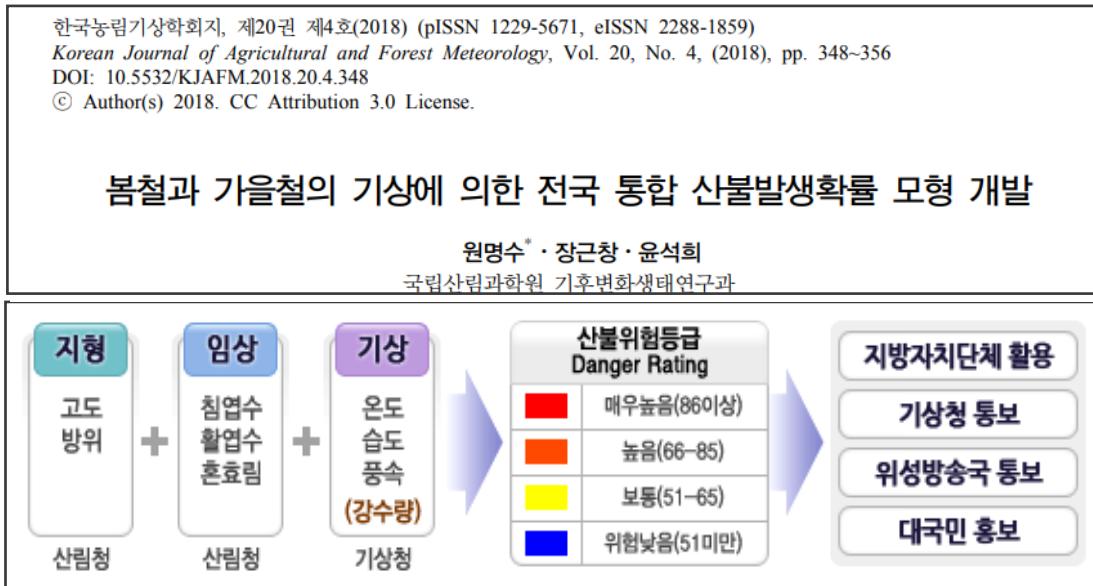
- ✓ 양간지풍 : ‘양양’과 ‘간성’ 사이에 부는 바람
- ✓ 양강지풍 : ‘양양’과 ‘강릉’ 사이에 부는 바람
- ❖ 계절별 산불 비중은 “봄(3월~5월)”에 59.1%으로 집중된다.

Point. ‘실효습도’의 중요성

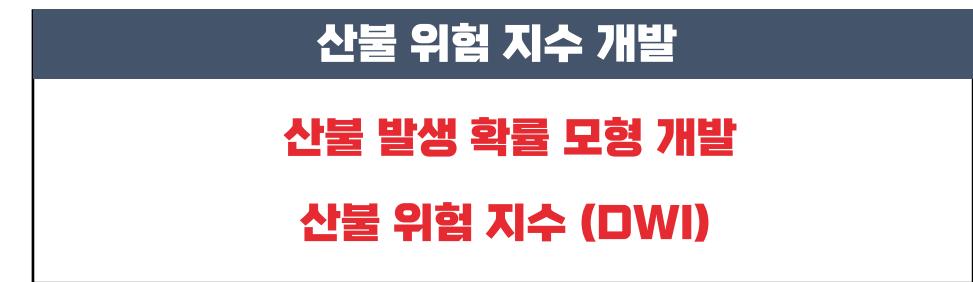
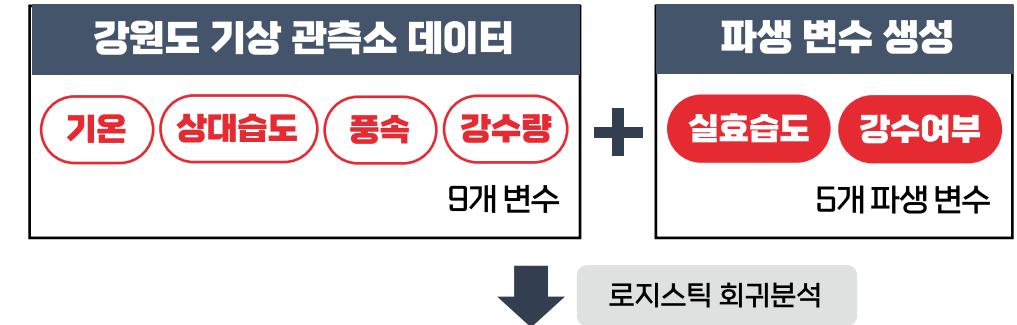
- ✓ 실효습도는 당일과 전일들의 상대습도에 가중치를 적용하여 얻어지는 평균습도로 산출한 목재 등의 건조도를 나타내는 지수이다.
- ✓ 실효습도 50% 이하가 되면 대형화재 위험성 △
- ✓ 당일의 평균상대습도를 포함하여 5일 간의 평균상대습도의 누적치를 이용

지수 설정

산불 위험등급 산출 과정



산불 위험지수 산출 과정



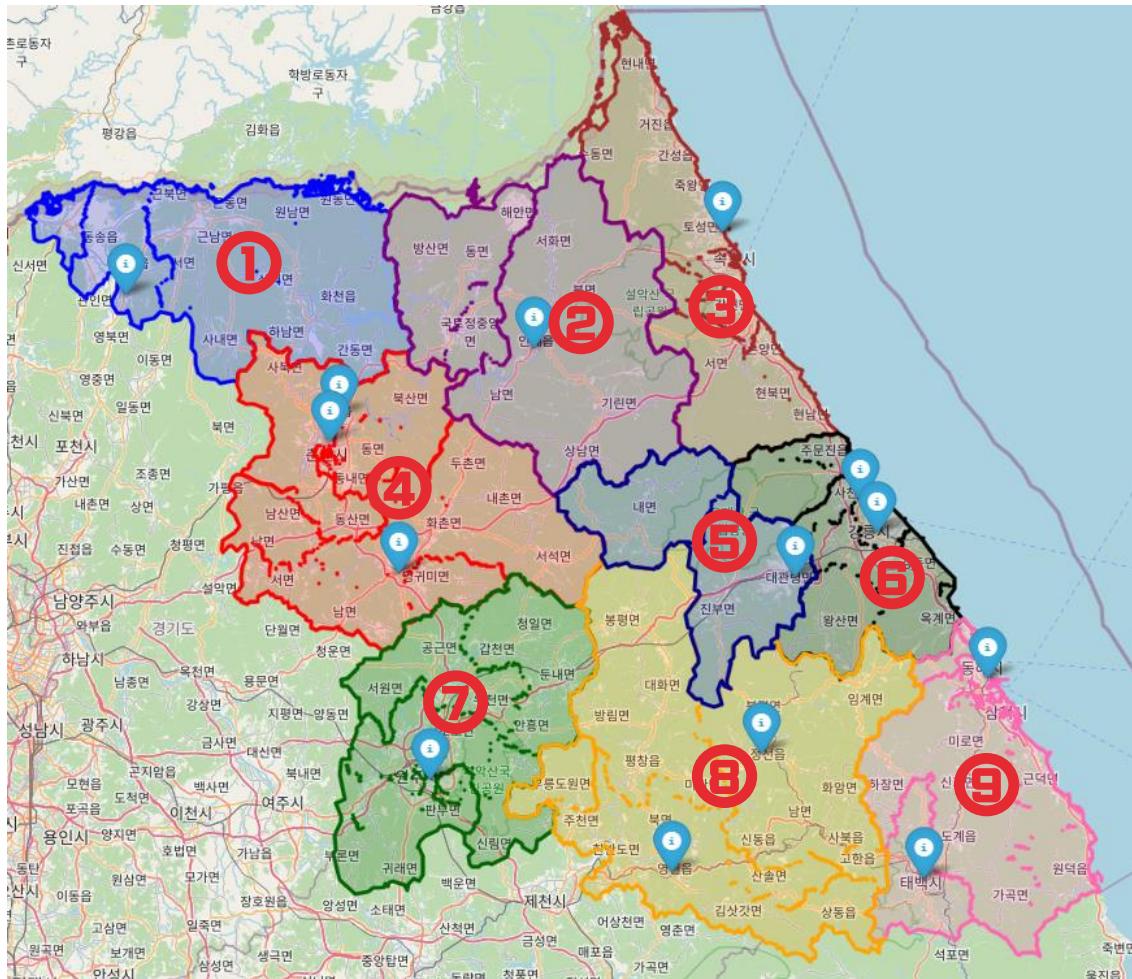
Point. 산불 위험지수 산출에 필요한 정보 파악

- ✓ 기상요인 : 기온, 상대습도, 실효습도*, 풍속, 강수량
- ✓ 기상 요인을 고려하여 1~100 까지의 지수를 산출
- ⌚ 4단계 산불위험등급 (매우높음, 높음, 보통, 낮음)으로 구분



평가 지표

지도시각화



공간정보 분석

데이터 시각화

강원도 행정구역 지도

관측소 위치 데이터

산불 발생 지점 데이터

기상 데이터



강원지방기상청 관할 구역 지도 참고

구역 설정

강원도 기상관측소 위치, 지형 / 기후적 특징을 기준

총 9개의 지역으로 분할

✓ 기대효과 1: ‘산불발생건수 / 산불피해면적’

시각화를 통한 산불발생현황 파악

✓ 기대효과 2: 강원도 지역별 특성을 고려한 모델 개발

평가 지표

통계 분석



평가 지표

Machine Learning

XGBoost



LightGBM



✓ 평가지표 : AUC-ROC, Accuracy

Deep Learning

EfficientNet - B7



✓ 평가지표 : AUC-ROC, Accuracy

산불발생예측
모델 개발

데이터셋 분리

모델 학습 성능 검증

하이퍼 파라미터 조정

성능 평가

✓ 기대효과 : 선제적 산불 예방체계 마련계획 수립

산불분류
모델 개발

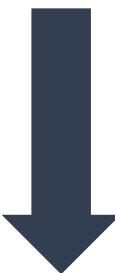
데이터셋 분리

모델 학습 성능 검증

훈련/예측 단계 최적화

성능 평가

✓ 기대효과 : 산불 조기 발견 / 현장 대응체계 고도화



수행 절차

WBS (Work-Breakdown Structure)

구 분		1주차(5/22~27)	2주차(28~6/3)	3주차(6/4~10)	4주차(6/11~18)	5주차(6/19~23)
기획	기획 및 회의					
데이터 수집	API 데이터 수집					
데이터 수집	논문 자료 수집					
데이터 전처리	데이터 수정					
DB 연동	DB 업데이트					
데이터 탐색	탐색적 자료 분석					
데이터 시각화	공간 및 지도 시각화					
통계 분석	로지스틱 회귀분석					
AI 모델링	모델 생성 및 평가					
서버 연결	Streamlit 배포					
서비스 구현	Web 연동 및 배포					
최종	프로젝트 발표					

수행 기간

SUN	MON	TUE	WED	THR	FRI	SAT
	5/22 프로젝트 시작	23	24	25 서비스 기획	26	27
28	29	30	31	6/1 데이터 수집	2	3 DB 업데이트 및 데이터 전처리
4	5 기획 발표	6	7	8 DB (BigQuery) 연결	9	10 통계 분석
11	12	13	14	15 AI 모델링 (ML / DL)	16 1차 배포 및 서비스	17 서비스 구현
18	19	20	21	22 최종 서비스 배포	23 프로젝트 마감	
PPT 작성						
추가 수정 및 검토						

Chapter 2.

프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

데이터 탐색

FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

통계 분석

분석 개요
분석 내용
분석 결과 해석

ML 모델링

산불 예측 모델 생성 및 학습
모델 성능 검증
산불 위험 예측

DL 모델링

산불 분류 모델 생성 및 학습
모델 성능 검증
산불 이미지 분류 예측

서비스 및 문서

서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고 문헌 / 부록



프로젝트 소개

데이터 탐색

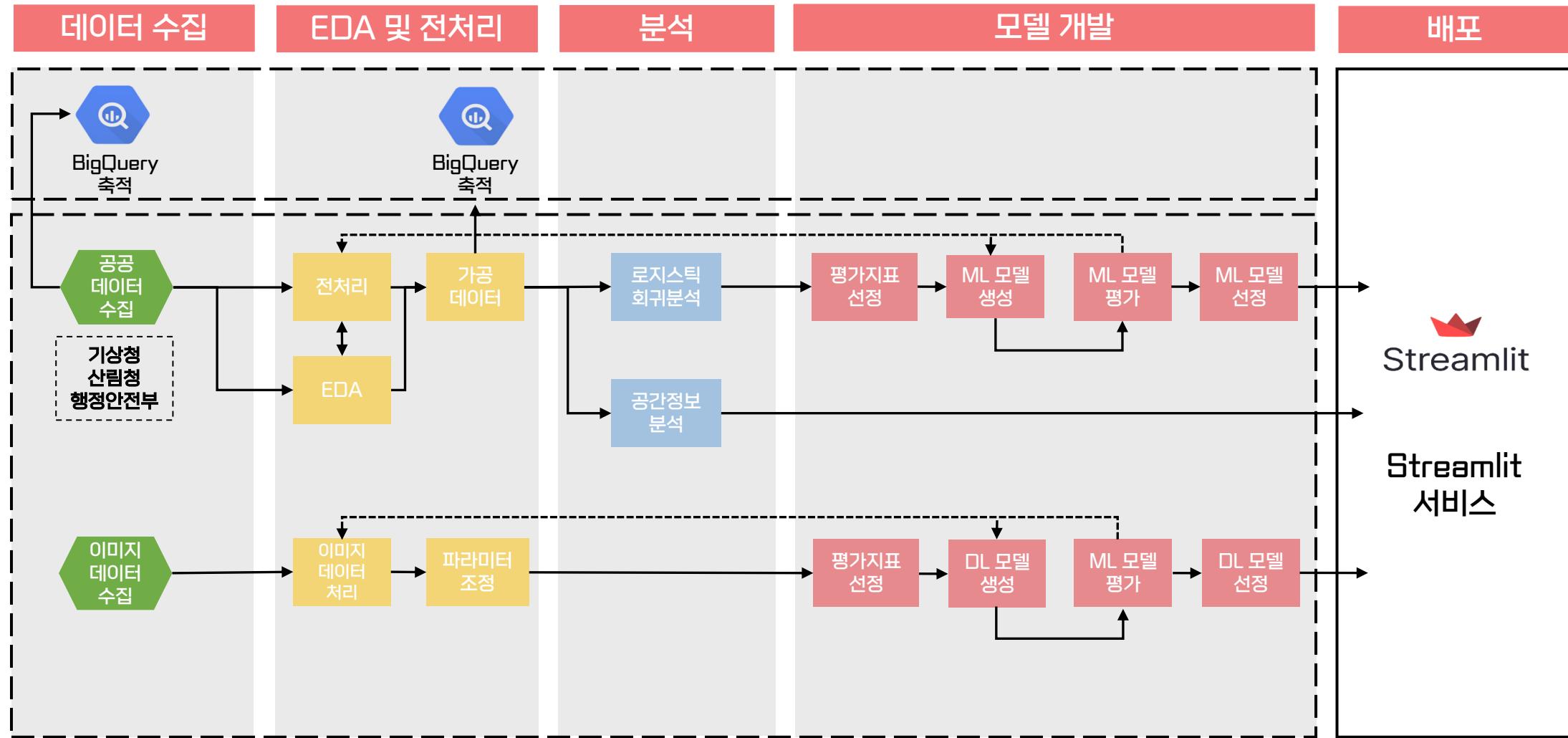
통계 분석

ML 모델링

DL 모델링

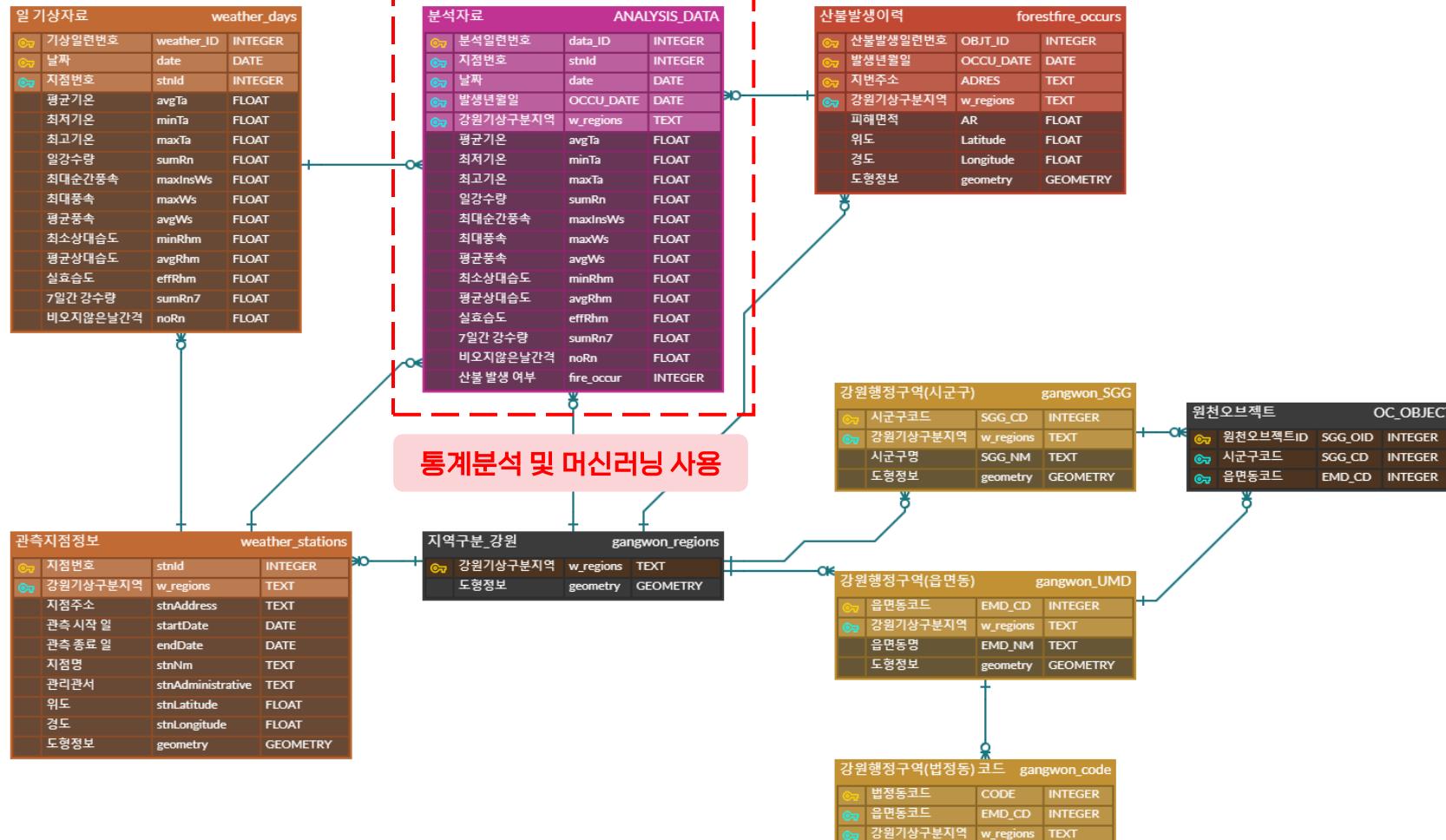
서비스 및 문서

FLOW CHART



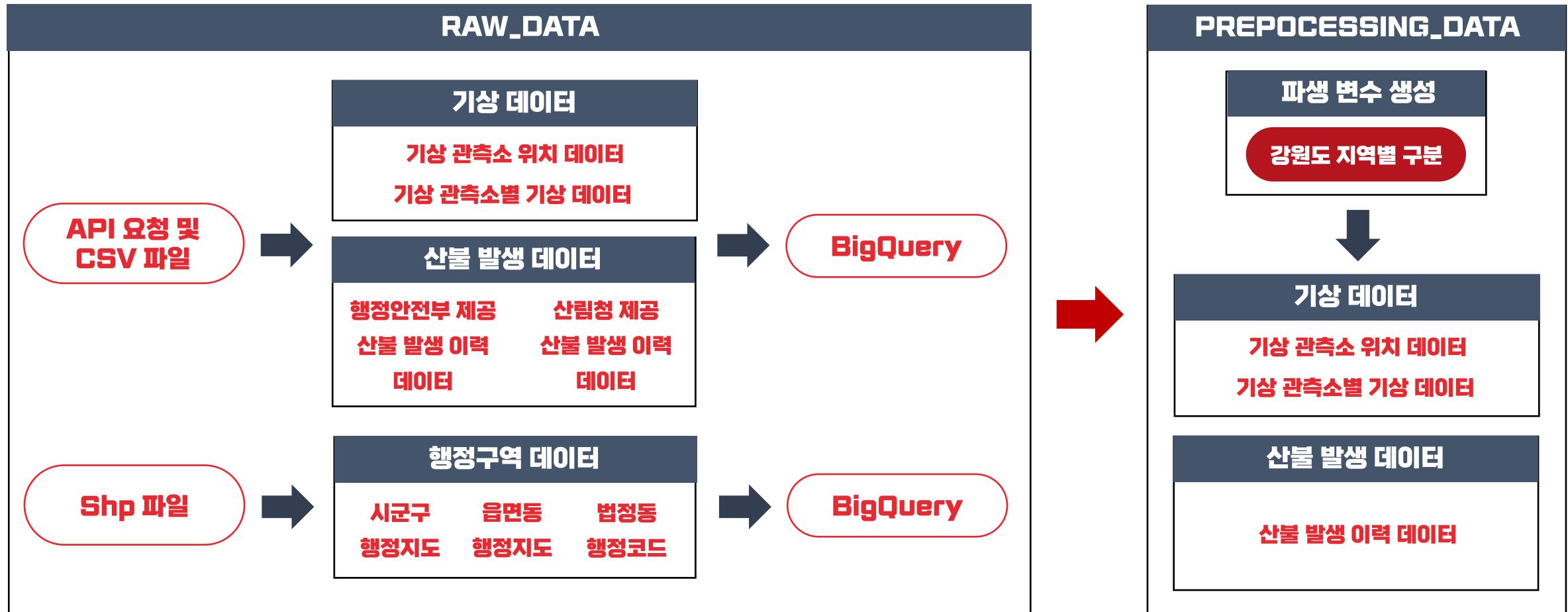
ERD

ERD



데이터 수집

데이터 수집 및 처리 과정



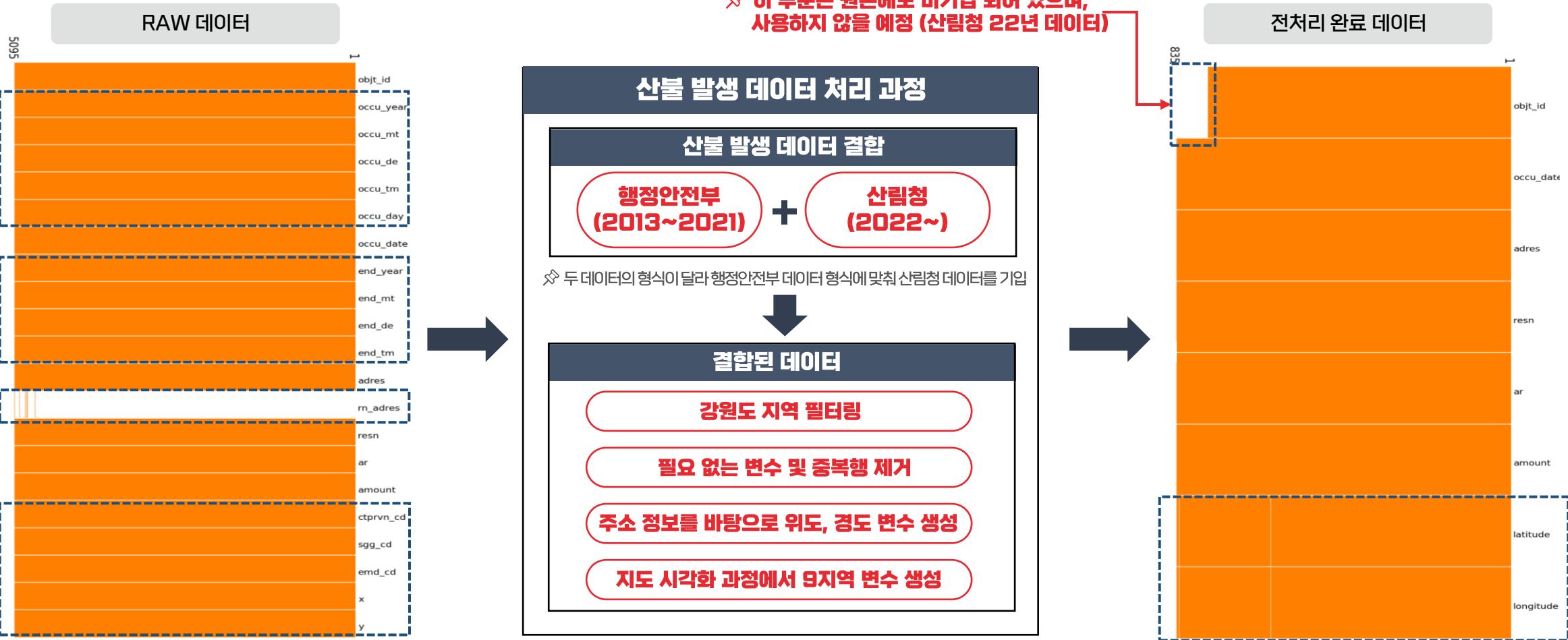
데이터 전처리

기상 데이터 처리



데이터 전처리

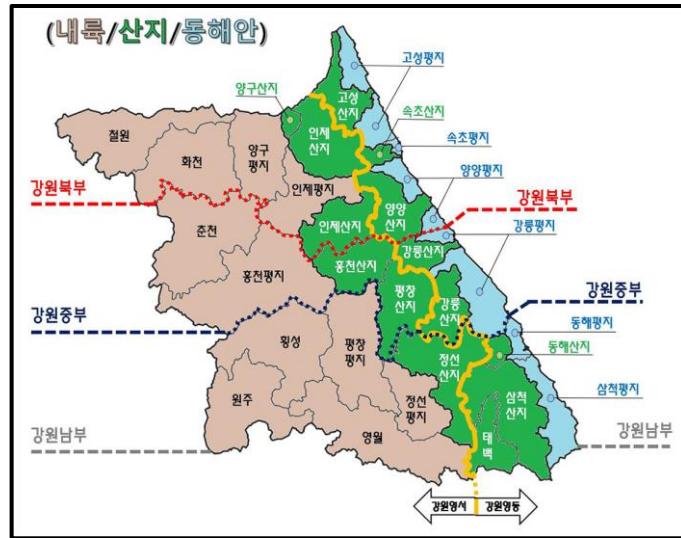
산불 발생 데이터 처리



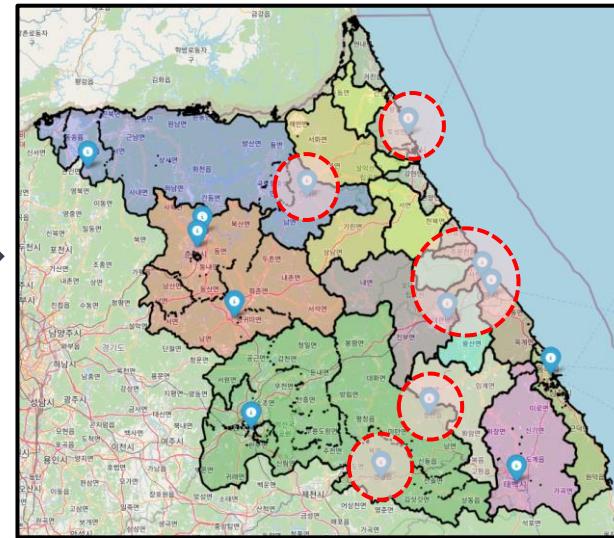
데이터 전처리

강원도 지역 분할

강원지방기상청 관할 구역 자료



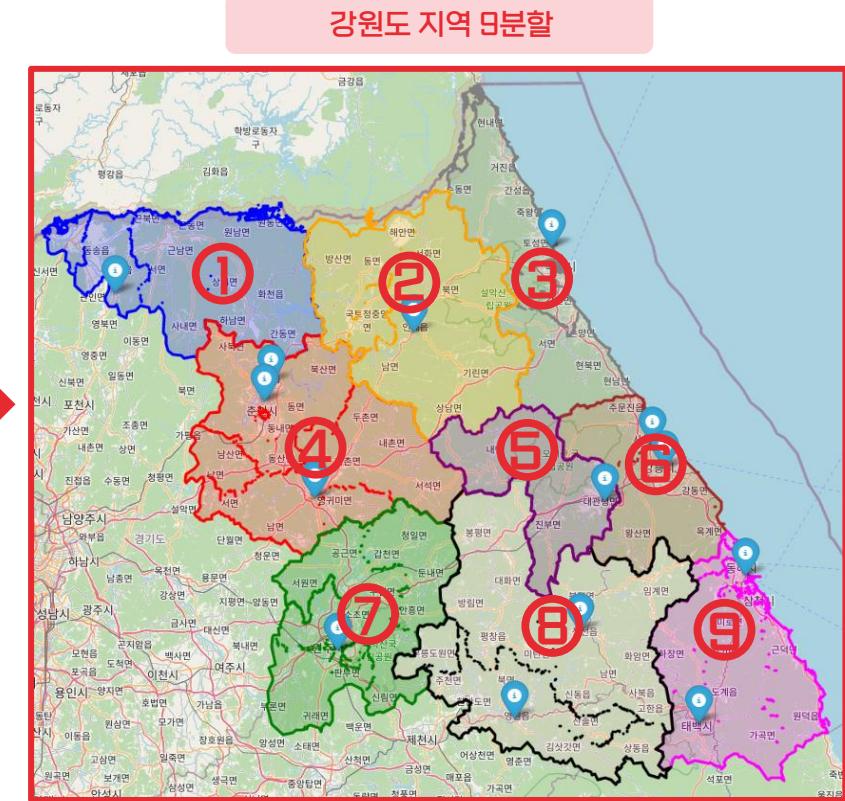
강원도 지역 12분할



강원도 지역 기상 관측소 위치

Point. 강원도 지도 시각화를 위한 구역 분할 과정

- ✓ 강원지방기상청 관할 예, 특보 구역 자료를 참고하여 강원도 지역을 12분할로 진행
- ✓ 지역을 12분할 후, 기상 관측소 위치와 연관이 없는 지역이 발생한 것을 확인
- ❖ 지형적 특성 및 기상 관측소 위치를 고려하여 지역을 9분할하여 진행



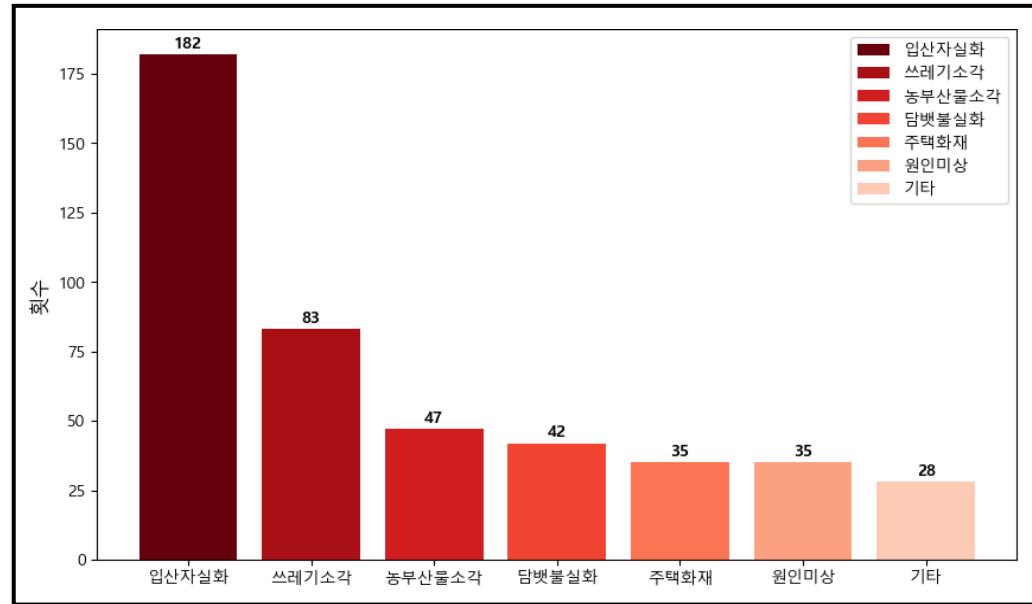
✓ 이후 슬라이드부터 9개 지역에 대해 '지역 1~9'로 표시 예정

- | | | |
|----------|----------|----------|
| ① 강원북부내륙 | ② 강원북부산지 | ③ 강원북부해안 |
| ④ 강원중부내륙 | ⑤ 강원중부산지 | ⑥ 강원중부해안 |
| ⑦ 강원남부내륙 | ⑧ 강원남부산지 | ⑨ 강원남부해안 |

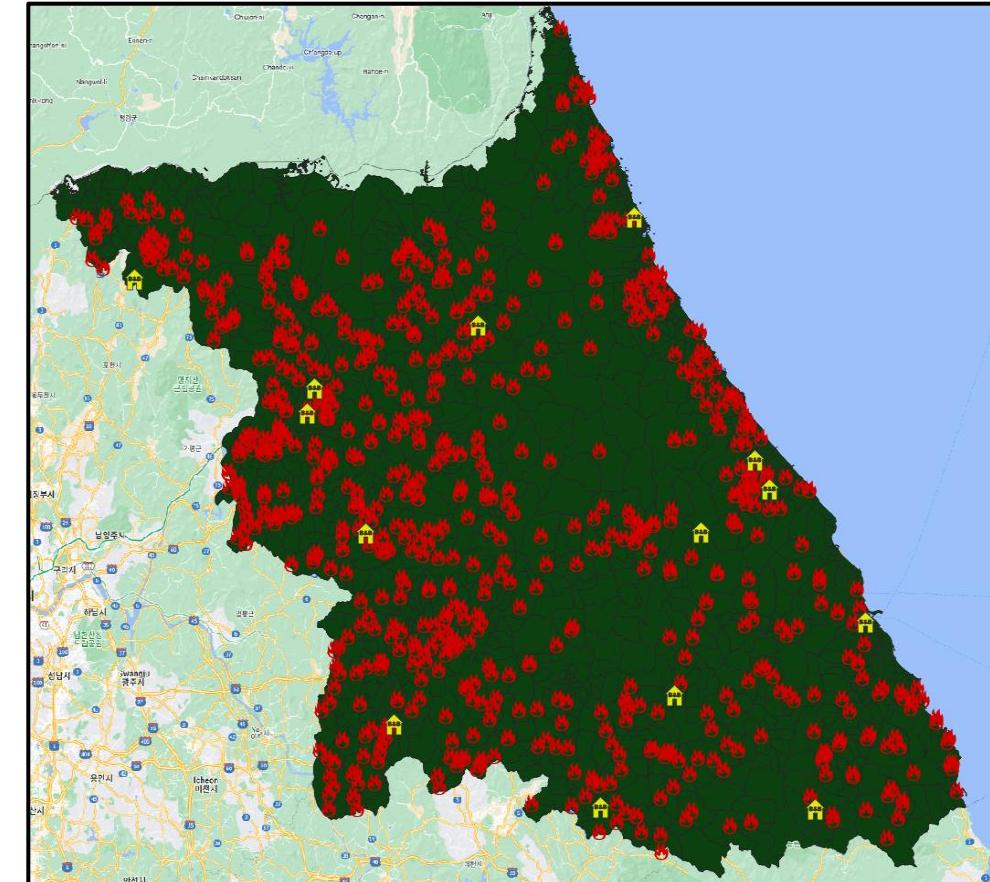
탐색적 데이터 분석 (EDA)

공간 분석 (지도 시각화)

산불 발생 원인별 횟수



산불 발생 위치



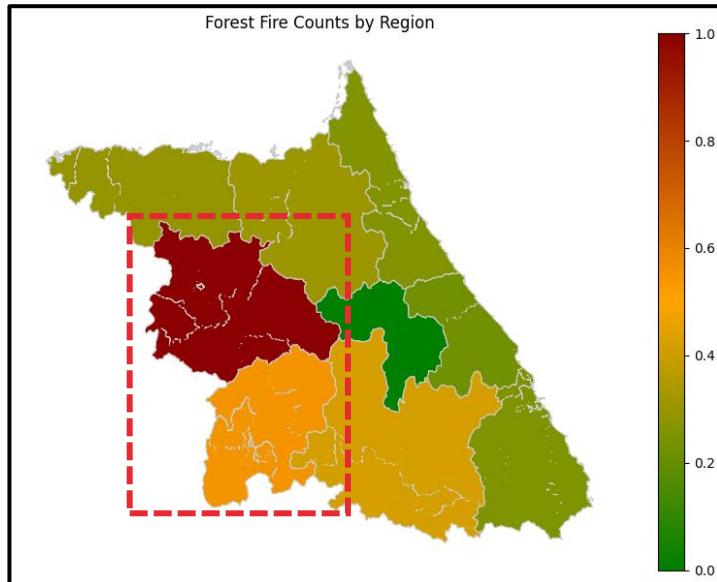
Point. 산불 발생 이력 데이터 시각화

- ✓ 강원도 전역 산불 발생 원인 및 발생 위치 시각화를 진행
- ✓ 산불 발생 원인은 사람들의 부주의가 가장 크고, 이러한 이유 때문에 산불 발생이 유동 인구가 많은 지역에서 높게 나타날 것으로 예상
- ☞ 더 자세한 분석을 위해 발생 건수, 피해 범위, 피해 금액에 대한 시각화 진행

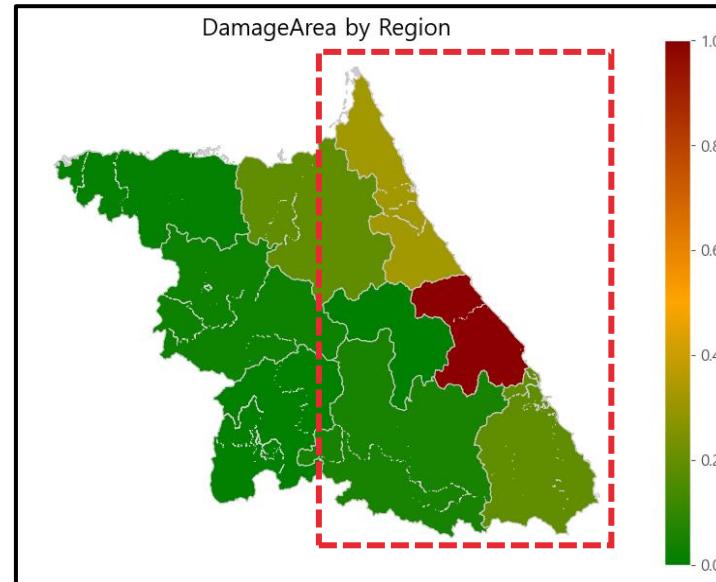
탐색적 데이터 분석 (EDA)

공간 분석 (지도 시각화)

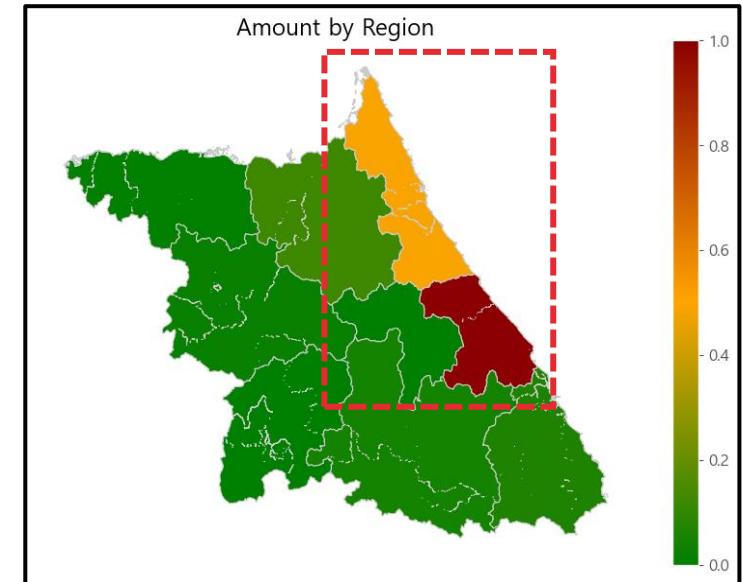
지역별 산불 발생 건수



지역별 산불 피해 범위



지역별 피해 금액



- ✓ 산불 발생 건수는 상대적으로 유동인구가 많은 대도시 지역이 높게 나타남을 확인

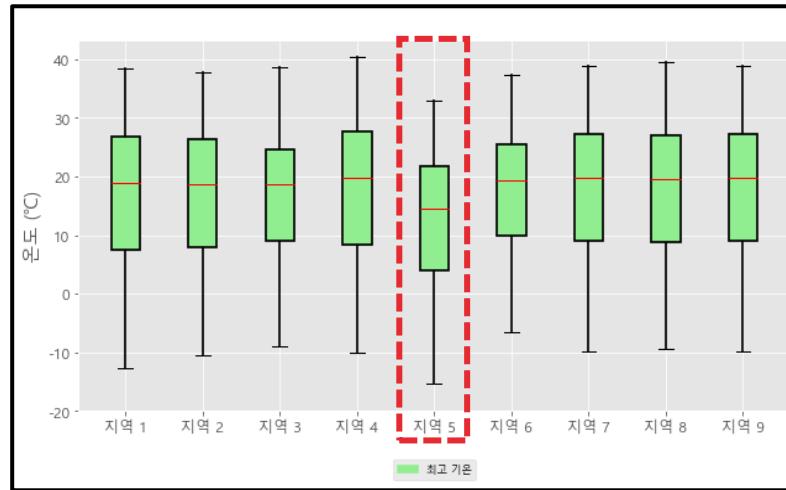
- ✓ 양간지풍 영향을 받는 해안 지역의 산불 피해가 내륙보다 더 크고 심한 것으로 나타남

- ✓ 산불 피해 범위 지도로 알 수 있듯이 피해 범위가 넓은 해안 지역에서 피해 금액도 높게 나타남

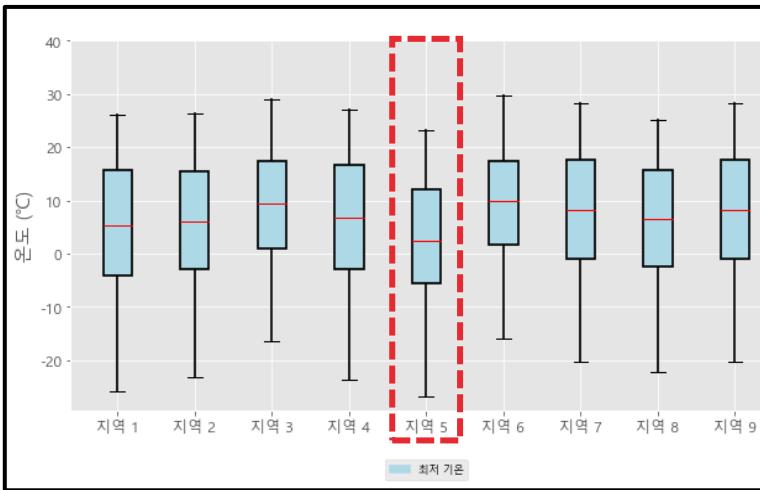
탐색적 데이터 분석 (EDA)

기상청 데이터 시각화 (기온)

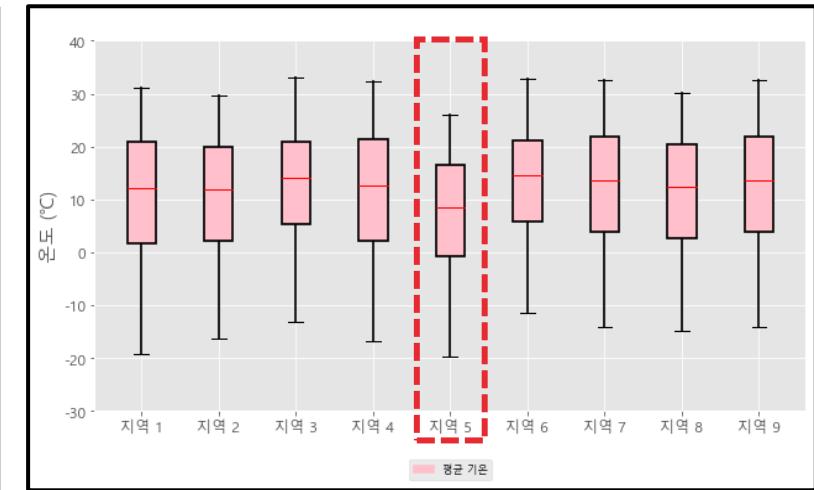
9개 지역별 최고 기온 데이터



9개 지역별 최저 기온 데이터



9개 지역별 평균 기온 데이터



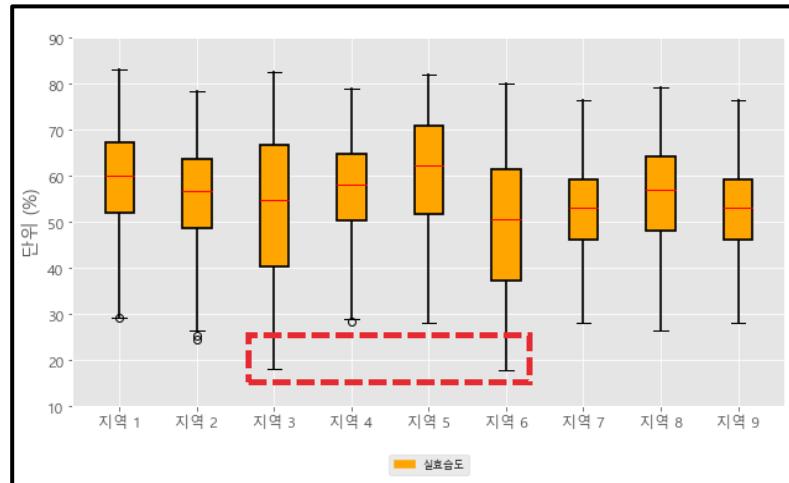
Point. 9개 지역별 기온 시각화

- ✓ 지역별로 기온을 시각화해 본 결과 기상관측소가 고지대에 있는 지역 5를 제외한 나머지 지역의 최고, 최저, 평균 기온 **데이터 모두 차이가 크지 않은 것**을 확인함
- ❖ **기온이 산불에 큰 영향을 주지 않을 것**으로 생각해 볼 수 있음

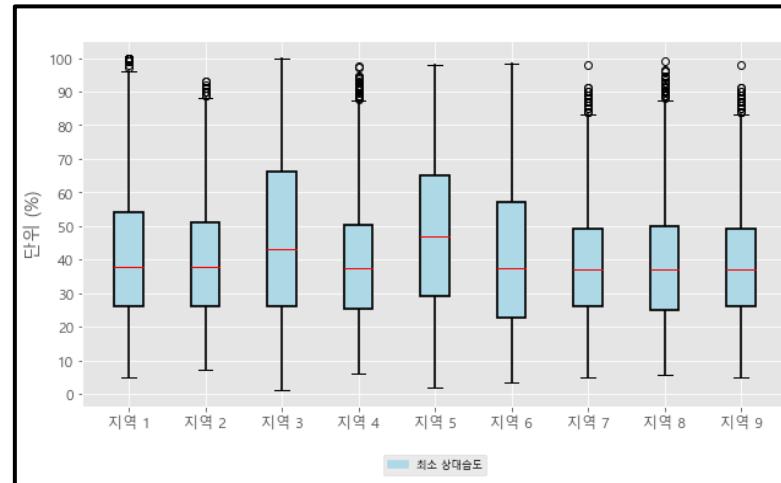
탐색적 데이터 분석 (EDA)

기상청 데이터 시각화 (습도)

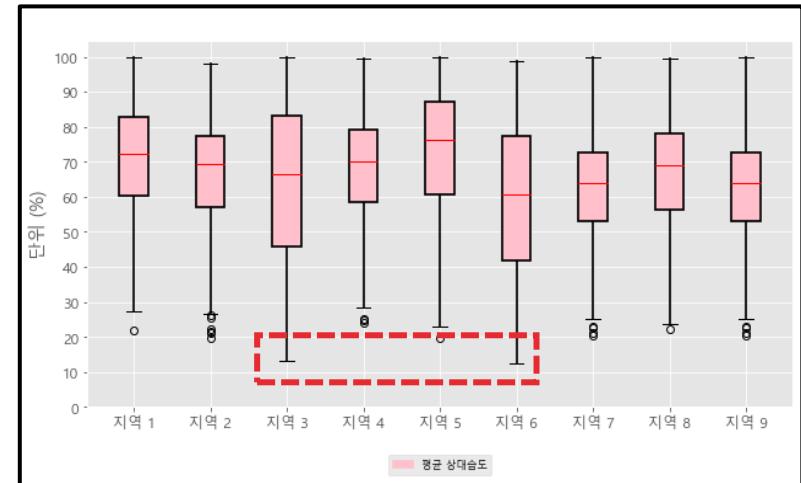
9개 지역별 실효습도 데이터



9개 지역별 최소 상대습도 데이터



9개 지역별 평균 상대습도 데이터



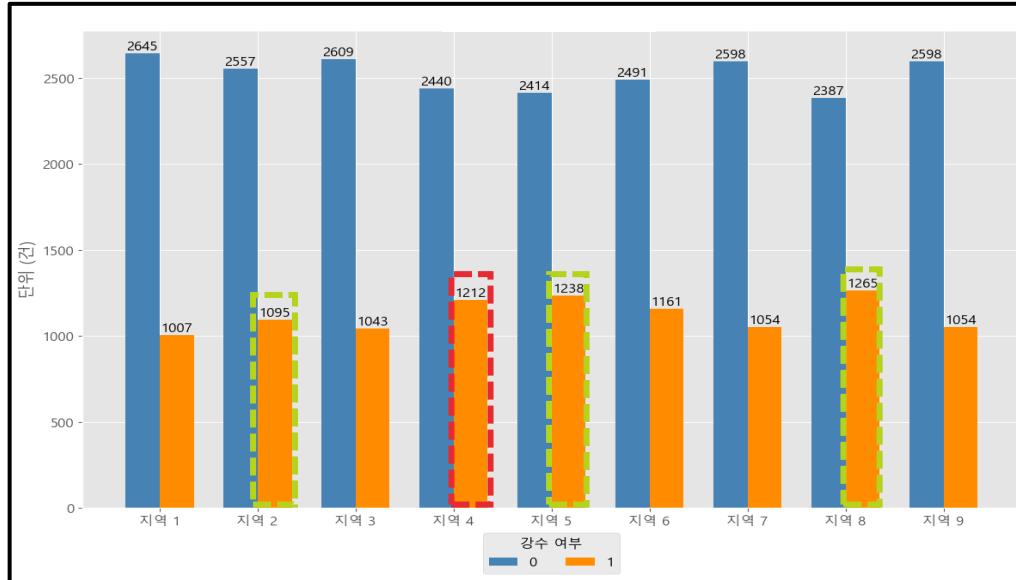
Point. 9개 지역별 습도 시각화

- ✓ 평균 상대습도가 지역 5에서 가장 높게 나타나는 것을 확인
- ✓ 평균 상대습도 및 실효습도 그래프를 통해 지역 3, 6 이 해안 지역임에도 불구하고, **습도가 낮은 수치까지 내려가는 것을 확인**할 수 있음 (양간 지풍 영향 추정)
- ❖ 지도 시각화에서 확인한 바와 같이 **산불 피해 범위는 낮은 상대습도와 연관이 있을 것으로 예상**

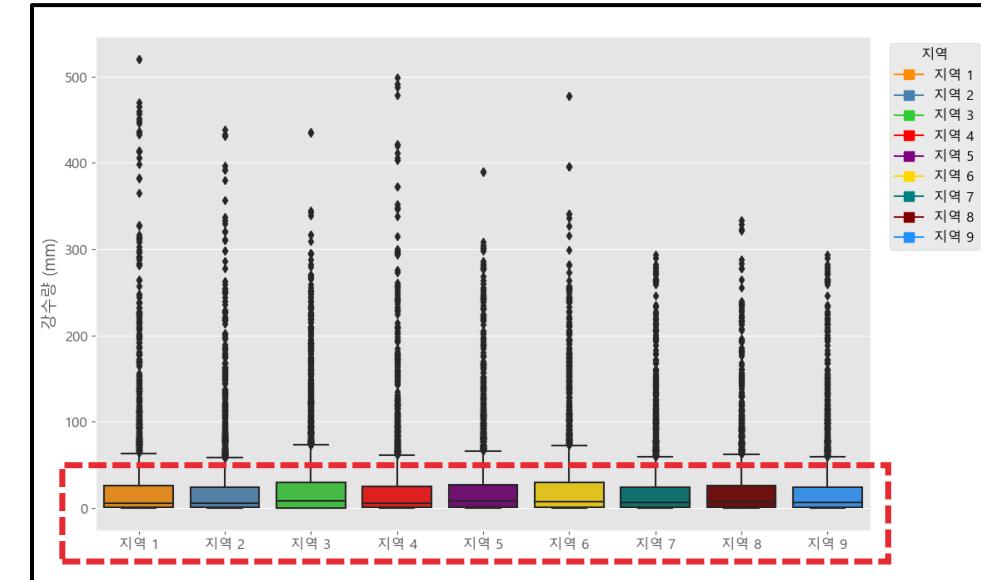
탐색적 데이터 분석 (EDA)

기상청 데이터 시각화 (강수)

9개 지역별 강수 여부 데이터



기준일로부터 6일전까지 7일간 강수량 (9개 지역)



Point. 9개 지역별 강수량 시각화

- ✓ 강원 산지(지역 2,5,8)에서 강수 횟수가 높게 나타나는 것을 확인 할 수 있음
 - 지도시각화 과정에서 산불 발생 건수 가 적었으므로, 강수 와 산불 발생은 반비례 관계 예상
- ✓ 예외적으로, 지역 4는 강수 빈도가 높은 반면에 산불 발생 건수도 높았음
 - 해당 지역은 지도시각화 과정에서 유동 인구가 많은 지역적 특성을 고려 할 필요가 있음

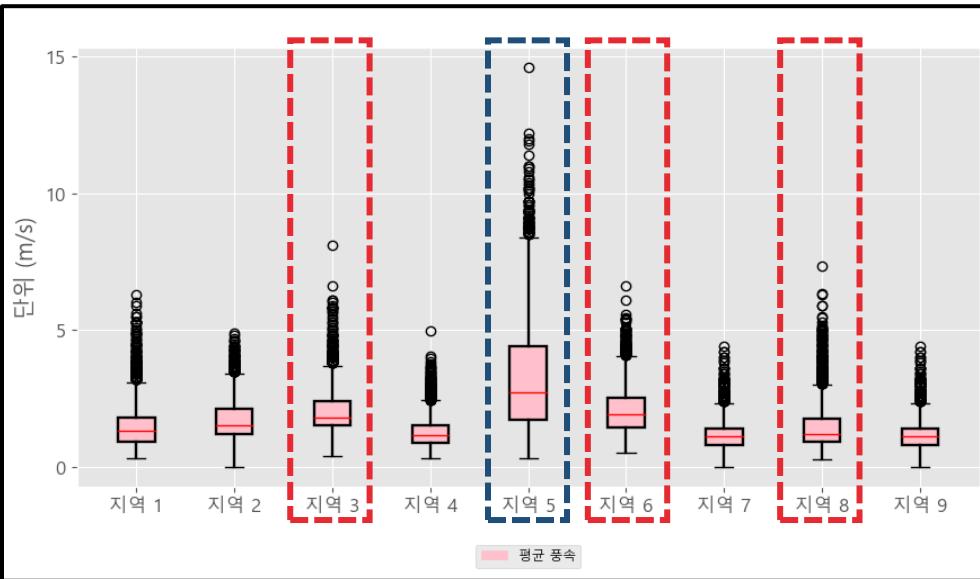
Point. 9개 지역별 7일간 강수량 시각화

- ✓ 강수량 관련 데이터는 비가 오지 않는 날의 값이 모두 0 이기 때문에 평균이 0 으로 치우친 것 을 알 수 있음

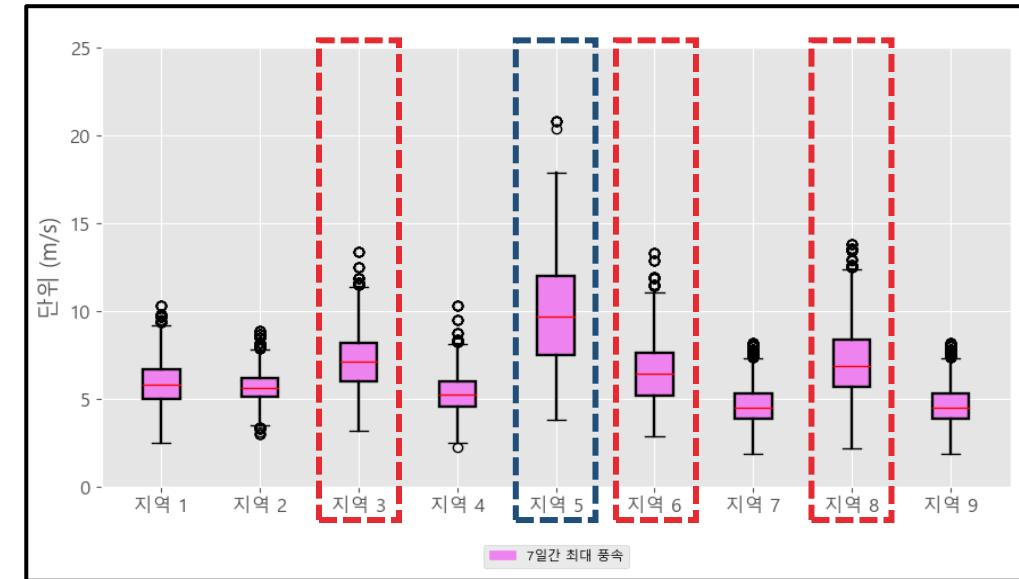
탐색적 데이터 분석 (EDA)

기상청 데이터 시각화 (풍속)

9개 지역별 평균 풍속 데이터



9개 지역별 7일간 최대 풍속 데이터



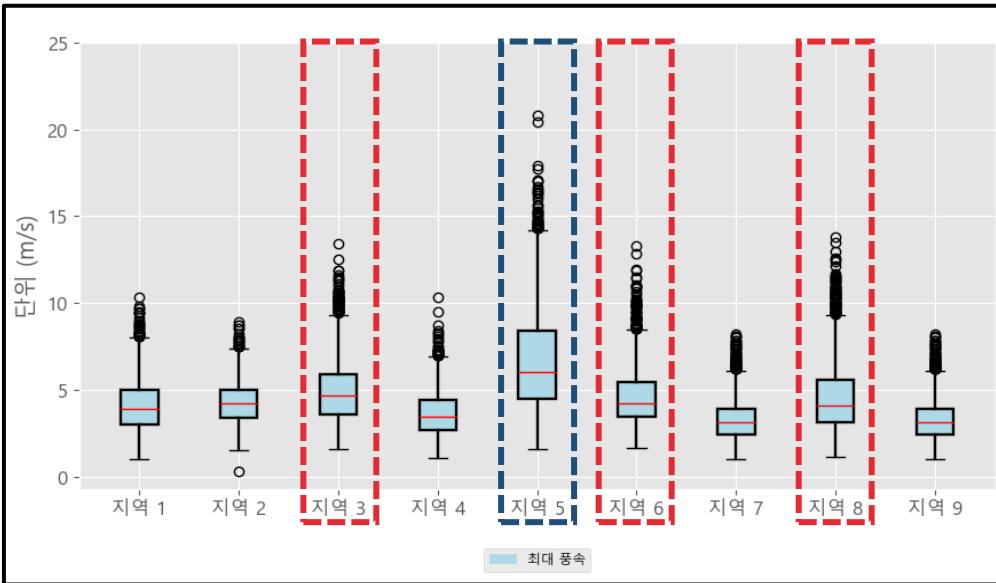
Point. 9개 지역별 풍속 시각화

- ✓ 풍속 데이터 중 **지역 5**는 관측소가 고지대에 위치하고 있기 때문에 **전체적으로 높은 수치를 보임**
- ✓ 지역 5를 제외한 해안(지역 3, 6), 산지(지역 8)에서 **풍속이 높게 나타남**
 - 해당 지역에 대한 지도 시각화를 확인한 결과 해안(지역 3, 6)에서 **산불 피해가 상대적으로 커짐**,
 - 따라서, **습도가 낮은 지역에서의 강풍이 산불 대형화로 이어질 수 있다**는 것을 예상해 볼 수 있음

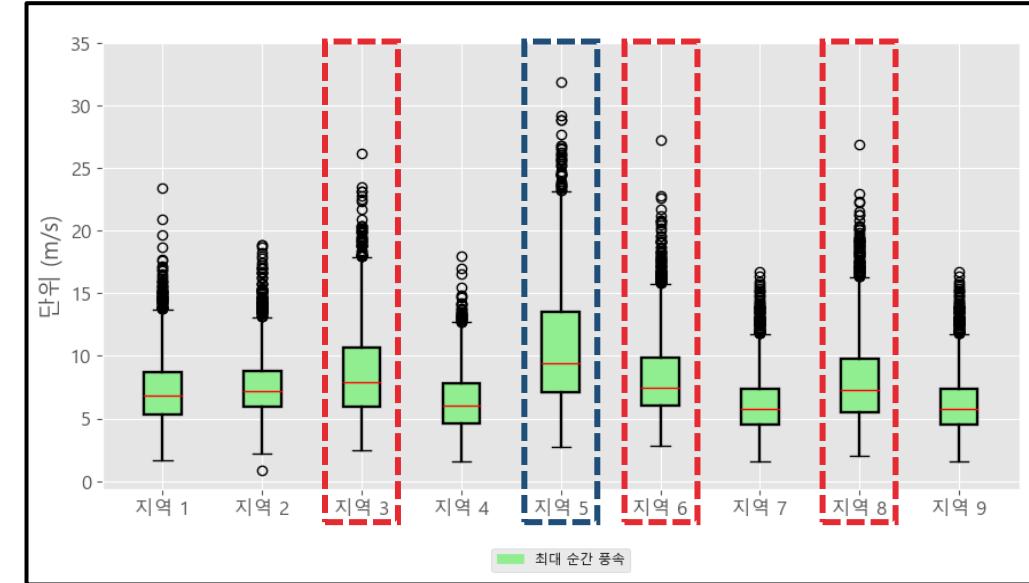
탐색적 데이터 분석 (EDA)

기상청 데이터 시각화 (풍속)

9개 지역별 최대 풍속 데이터



9개 지역별 최대 순간 풍속 데이터



Point. 9개 지역별 풍속 시각화

- ✓ 풍속 데이터 중 **지역 5**는 관측소가 고지대에 위치하고 있기 때문에 **전체적으로 높은 수치를 보임**
- ✓ 지역 5를 제외한 해안(지역 3, 6), 산지(지역 8)에서 **풍속이 높게 나타남**
 - 해당 지역에 대한 지도 시각화를 확인한 결과 해안(지역 3, 6)에서 **산불 피해가 상대적으로 커짐**,
 - 따라서, **습도가 낮은 지역에서의 강풍이 산불 대형화로 이어질 수 있다**는 것을 예상해 볼 수 있음

Chapter 3.

프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

데이터 탐색

FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

통계 분석

분석 개요
분석 내용
분석 결과 해석

ML 모델링

산불 예측 모델 생성 및 학습
모델 성능 검증
산불 위험 예측

DL 모델링

산불 분류 모델 생성 및 학습
모델 성능 검증
산불 이미지 분류 예측

서비스 및 문서

서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고 문헌 / 부록



프로젝트 소개

데이터 탐색

통계 분석

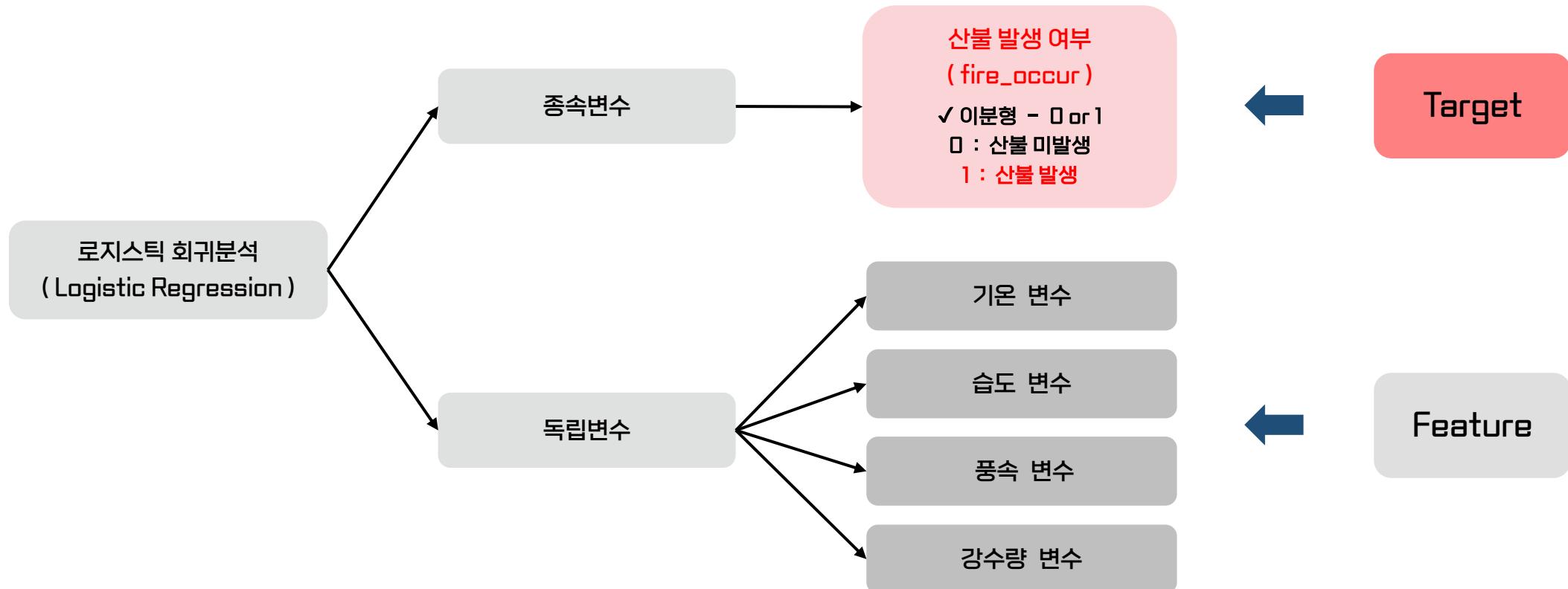
ML 모델링

DL 모델링

서비스 및 문서

분석 개요

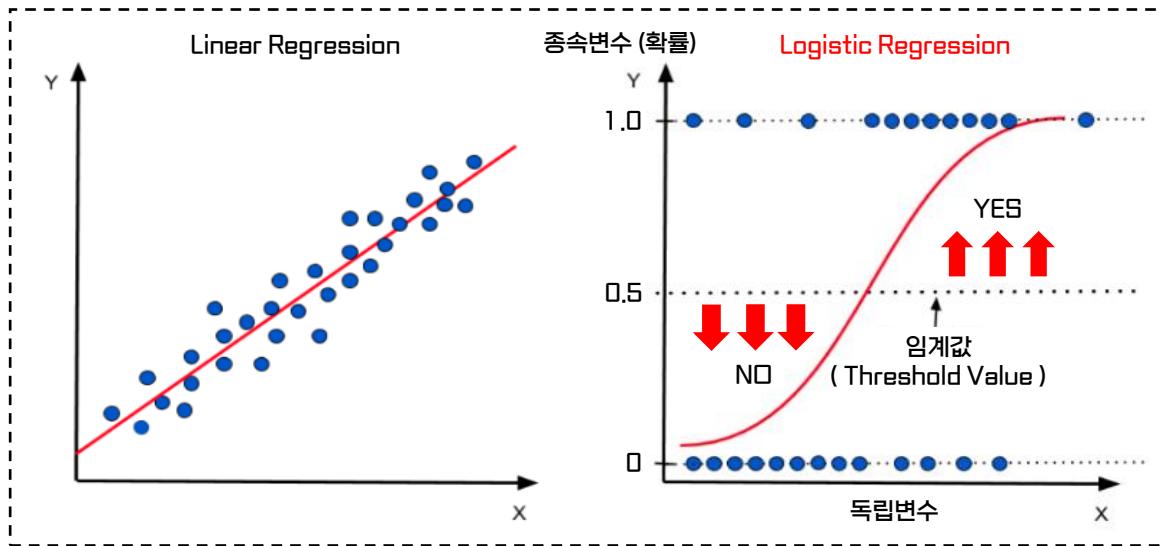
로지스틱 회귀분석



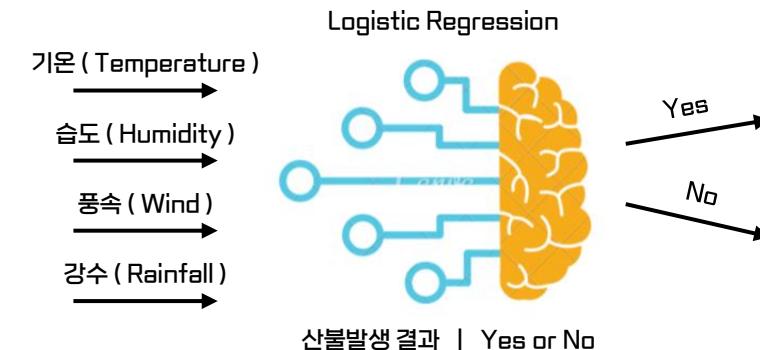
분석 개요

로지스틱 회귀분석

독립변수 / 종속변수 관계



산불발생여부 예측



Point. 로지스틱 회귀분석 (Logistic Regression)

- ✓ 0 또는 1의 값만을 가지는 범주형 종속변수를 예측하는데 사용
- ✓ 확률 (분석결과 종속변수 값)이 0.5 보다 크면 그 사건이 일어나며,
0.5 보다 작으면 사건이 일어나지 않을 것으로 예측

분석 개요

로지스틱 회귀분석

오즈비(Odds Ratio; OR)

$$\checkmark \text{ Odds} = \frac{\text{산불 발생함}}{\text{산불이 발생하지 않음}}$$

$$= \frac{p}{1-p} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_i X_i}$$

$$\text{오즈비 (OR)} = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{AD}{BC}$$

구분	산불발생 (예측)		합계
	YES (1)	NO (2)	
산불발생 (실제)	YES (1) A	NO (2) B	A + B
	C	D	C + D
합계	A + C	B + D	A + B + C + D

Point. 오즈비(Odds Ratio; OR)

- ✓ 확률 p 가 주어졌을 때, 사건이 발생할 확률이 발생하지 않을 확률에 비해 몇 배 높은지를 의미
- ✓ 종속변수인 ‘산불발생여부 (fire_occur)’의 범주가 ‘1’이 산불발생이고, ‘0’이 산불 미발생임을 가정할 때, p 가 0.8 이라면 오즈비는 $(0.8 / (1 - 0.8)) = 4$
- ✓ 이때, 산불이 발생할 확률이 발생하지 않을 확률보다 4배가 높다고 해석

Point. 오즈비(Odds Ratio; OR) 해석 유형

- ✓ $OR = 1$ 이면, 사건이 발생할 확률에 유의미한 영향을 준다고 볼 수 없다고 판단
- ✓ $OR > 1$ 이면, 사건이 발생할 확률이 오즈비만큼 더 크게 나타난다고 해석
- ✓ $OR < 1$ 이면, 사건이 발생할 확률이 오즈비만큼 더 낮게 나타난다고 해석
- ✓ 오즈비를 분석결과 표기할때는 보통 신뢰구간을 같이 표기함 (99% 신뢰구간에서 예측 진행 예정)

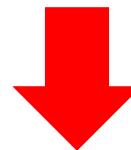
분석 개요

로지스틱 회귀분석

가설 설정

- ✓ 귀무가설 (H_0) : 독립변수 (기상요인) 가 종속변수 (산불발생 여부) 에 영향을 미치지 않는다 ($\beta_1 = 0$)
- ✓ 대립가설 (H_1) : 독립변수 (기상요인) 가 종속변수 (산불발생 여부) 에 영향이 있다 ($\beta_1 \neq 0$)
 - ▶ 독립변수 (기상요인) 중 적어도 하나는 종속변수에 영향이 있다

- ✓ 강원도 9개 지역 중 산불피해범위, 산불피해금액 이 많았던 지역 6 (강원중부해안) 을 예시로 분석한 통계량 소개 예정
- ✓ 나머지 지역은 '부록 - 로지스틱 회귀분석 통계량' 참고



가설 검정

모형의 적합도 검정

- ✓ 카이제곱 값 / 유의확률 p 값
- ✓ Hosmer & Lemeshow 검정
- ✓ AIC (Akaike 정보기준)

모형의 설명력

- ✓ -2 로그 우도
- ✓ Cox & Snell R^2
- ✓ Nagelkerke R^2

계수에 대한 검정

- ✓ Wald 통계량

분석 내용

회귀모형 적합도

모형의 적합도 검정 - 지역 6

케이스 처리 요약

가중되지 않은 케이스 ^a		N	퍼센트
선택 케이스	분석에 포함	3652	100.0
결측 케이스		0	.0
전체		3652	100.0
비선택 케이스		0	.0
전체		3652	100.0

총 3,652 개의 케이스가 분석에 포함되었으며,
결측값은 없는 것으로 나타남

범주형 변수 코딩

	빈도	모수 코딩	
		(1)	(-1)
Rntf	0	2491	.000
	1	1161	1.000

독립변수로 설정된 강수 여부 (Rntf)는 범주형 변수로,
0은 강수 없음, 1은 강수 있음으로 분류 하여 빈도수 나타남

분석 내용

회귀모형 적합도

모형의 적합도 검정 – 지역 6

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

		카이제곱	자유도	유의확률
1 단계	단계	48.099	1	<.001
	블록	48.099	1	<.001
	모형	48.099	1	<.001
2 단계	단계	17.085	1	<.001
	블록	65.184	2	<.001
	모형	65.184	2	<.001
3 단계	단계	5.325	1	.021
	블록	70.509	3	<.001
	모형	70.509	3	<.001

- ✓ 변수 선택 방법 : 전진 (Forward - Wald)
- ✓ 유의한 변수에 따라 3단계로 과정이 나타남

✓ 모형에 해당하는 부분이 독립변수 전체계수에 대한 회귀모형의 적합도를 나타낸다

✓ 대립가설을 선택하여 독립변수들을 이용한 로지스틱 회귀분석 진행하기 위해서는, 유의확률 p 값은 0.01보다 작게 나타나야 함 (99% 신뢰구간 진행)

✓ 표를 토대로 보면, $\chi^2 = 70.509$, $p < 0.001$

✓ 따라서, ‘독립변수 (기상요인)가 종속변수 (산불발생 여부)에 영향을 미치지 않는다’라는 귀무가설을 기각하고 대립가설을 채택한다

분석 내용

회귀모형 적합도

모형의 적합도 검정 - 지역 6

= Hosmer와 Lemeshow 검정 =

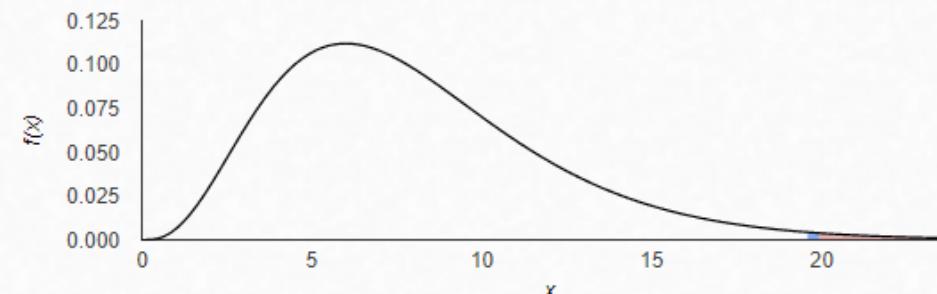
단계	카이제곱	자유도	유의확률
1	3.415	8	.906
2	9.109	8	.333
3	19.756	8	.011

- ✓ 3단계 변수 선택 과정의 모형을 기준으로
유의확률 $p > 0.01$
- ✓ 따라서, 추정된 로지스틱 모형 (3단계)은
적합하며 근사적으로 카이제곱 분포를 따른다

카이제곱 분포

Chi-Square Distribution
 $X \sim \chi^2_{(\nu)}$

$\nu = 8$
 $x = 19.756$ $P(X > x) = 0.0113$



$$\mu = E(X) = 8 \quad \sigma = SD(X) = 4 \quad \sigma^2 = Var(X) = 16$$

분석 내용

회귀모형 설명력

모형의 설명력 – 지역 6

모형 요약			
단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	572.120 ^a	.013	.084
2	555.035 ^b	.018	.113
3	549.709 ^b	.019	.122

- ✓ Cox & Snell R^2 : 값이 높을수록 완벽한 모형에 적합
- ✓ Nagelkerke R^2 : 값이 높을수록 완벽한 모형에 적합

✓ -2 로그 우도 (Likelihood; LL) : 값이 낮을수록 완벽한 모형에 적합
(카이제곱 분포를 따르는 통계량)

✓ AIC (Akaike 정보기준) : 통계 모형의 상대적인 품질 평가
(값이 낮을수록 좋은 품질의 모형)

✓ AIC 함수식: $-2 \text{ Log 우도} + 2k$ (k 는 ‘단계’ 를 의미)

✓ 표를 토대로 보면,

$$\text{2단계 AIC} = 555.035 + 2 \times 2 = 559.035$$

$$\text{3단계 AIC} = 549.709 + 2 \times 3 = 555.709$$

✓ 따라서, AIC 값이 3단계에서 전단계인 2단계 보다 줄어들었으나,
그 차이가 매우 적어 4단계에서는 AIC가 줄어들지 않았음을 추측 가능

분석 결과 해석

회귀모형 유의성 (계수 검정)

Wald 통계량 - 지역 6

✓ 오즈비 (OR) = $\text{Exp}(B)$

방정식의 변수							
	B	S.E.	Wald	자유도	유의확률	$\text{Exp}(B)$	EXP(B)에 대한 99% 신뢰구간 하한 상한
1 단계 ^a	avgRhm	-.048	.008	39.196	1	<.001	.953 .935 .972
	상수항	-1.657	.349	22.492	1	<.001	.191
2 단계 ^b	avgRhm	-.037	.008	20.362	1	<.001	.964 .944 .984
	sumRn7	-.052	.018	8.541	1	.003	.950 .908 .994
	상수항	-1.780	.363	23.986	1	<.001	.169
3 단계 ^c	avgWs	.417	.174	5.760	1	.016	1.517 .970 2.374
	avgRhm	-.023	.010	4.895	1	.027	.978 .952 1.004
	sumRn7	-.056	.018	9.594	1	.002	.945 .902 .991
	상수항	-3.449	.810	18.132	1	<.001	.032

✓ 오즈비 (OR) = $-3.449 + 0.417 \times \text{avgWs} - 0.023 \times \text{avgRhm} - 0.056 \times \text{sumRn7}$

- ✓ **B의 부호가 (+) :** $\text{OR} / \text{Exp}(B)$ 값이 1 이상으로 나타나며, 해당 변수 값이 커질수록 내부값이 1인 집단 (산불 발생)에 분류될 확률 커짐
- ✓ **B의 부호가 (-) :** $\text{OR} / \text{Exp}(B)$ 값이 1 미만으로 나타나며, 해당 변수 값이 커질수록 내부값이 0인 집단 (산불 미발생)에 분류될 확률 커짐

✓ avgWs : $B = (+) 0.417, p > 0.01, \text{OR} = 1.517$

∴ 평균풍속은 산불발생 확률에 있어 크게 영향을 미친다고 볼 수 없다

✓ avgRhm : $B = (-) 0.023, p > 0.01, \text{OR} = 0.978$

∴ 평균습도는 산불발생 확률에 있어 크게 영향을 미친다고 볼 수 없다

✓ sumRn7 : $B = (-) 0.056, p < 0.01, \text{OR} = 0.945$

∴ 7일간 강수량이 1 증가할수록 산불발생 확률 0.945 배 감소한다

독립변수들의 상관행렬 - 지역 6

상관행렬

상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
상수항	1.000	.185	.147	-.354	.013	.054	.001	-.293	-.010	-.198	-.269	.139	.154	-.426	.012

분석 결과 해석

지역별 산불 발생여부에 미치는 영향

✓ 99% 신뢰구간 ($p \leq 0.01$)을 기준으로 유의한 영향을 미치는 요인을 파악

지역 No.	독립 변수	B	S.E.	Wald	p (유의확률)	OR (오즈비)	99% CI (신뢰구간)	
							LLCI (하한)	ULCI (상한)
1	avgTa	-.359	.105	11.664	< .001	.699	.533	.916
	maxTa	.404	.101	15.877	< .001	1.498	1.154	1.946
	maxInsWs	.131	.060	4.882	.027	1.141	.978	1.329
	minRhm	-.058	.023	6.413	.011	.943	.889	1.001
	effRhm	-.068	.022	9.268	.002	.934	.882	.990
	sumRn7	-.045	.019	5.299	.021	.956	.909	1.005
2	minTa	-.169	.042	16.032	< .001	.845	.758	.942
	maxTa	.199	.037	28.471	< .001	1.221	1.109	1.344
	maxWs	.272	.121	5.074	.024	1.313	.962	1.793
	effRhm	-.096	.017	31.576	< .001	.908	.869	.949
	Rntf	.959	.343	7.809	.005	2.610	1.078	6.317

분석 결과 해석

지역별 산불 발생여부에 미치는 영향

✓ 99% 신뢰구간 ($p \leq 0.01$)을 기준으로 유의한 영향을 미치는 요인을 파악

지역 No.	독립 변수	B	S.E.	Wald	p (유의확률)	OR (오즈비)	99% CI (신뢰구간)	
							LLCI (하한)	ULCI (상한)
3	maxTa	.059	.015	14.507	< .001	1.061	1.021	1.103
	avgRhm	-.034	.014	15.803	.014	.967	.934	1.002
	effRhm	-.058	.020	5.985	.004	.943	.895	.994
4	minTa	-.069	.035	3.932	.047	.934	.854	1.021
	maxTa	.105	.031	11.637	< .001	1.110	1.026	1.202
	maxInsWs	.251	.073	11.813	< .001	1.285	1.065	1.551
	avgWs	-.919	.360	6.536	.011	.399	.158	1.007
	minRhm	-.030	.012	6.323	.012	.971	.941	1.001
	effRhm	.049	.015	10.892	< .001	.952	.916	.989
	sumRn7	-.079	.015	26.323	< .001	.924	.888	.962
	maxwind7	.186	.095	3.873	.049	1.205	.944	1.537

분석 결과 해석

지역별 산불 발생여부에 미치는 영향

✓ 99% 신뢰구간 ($p \leq 0.01$)을 기준으로 유의한 영향을 미치는 요인을 파악

지역 No.	독립 변수	B	S.E.	Wald	p (유의확률)	OR (오즈비)	99% CI (신뢰구간)	
							LLCI (하한)	ULCI (상한)
5	maxTa	.066	.031	4.432	.035	1.068	.985	1.158
	avgWs	-.583	.242	5.787	.016	.558	.299	1.042
	effRhm	-.124	.026	22.486	<.001	.883	.825	.945
6	avgWs	.417	.174	5.760	.016	1.517	.970	2.374
	avgRhm	-.023	.010	4.895	.027	.978	.952	1.004
	sumRn7	-.056	.018	9.594	.002	.945	.902	.991
7	avgTa	-.232	.073	10.278	.001	.793	.658	.955
	maxTa	.261	.070	14.083	<.001	1.298	1.085	1.553
	maxInsWs	.112	.046	6.037	.014	1.119	.995	1.259
	effRhm	-.096	.013	50.481	<.001	.909	.878	.941
	sumRn7	-.079	.020	16.092	<.001	.924	.879	.972

분석 결과 해석

지역별 산불 발생여부에 미치는 영향

✓ 99% 신뢰구간 ($p \leq 0.01$)을 기준으로 유의한 영향을 미치는 요인을 파악

지역 No.	독립변수	B	S.E.	Wald	p (유의확률)	OR (오즈비)	99% CI (신뢰구간)	
							LLCI (하한)	ULCI (상한)
8	maxTa	.037	.011	11.280	< .001	1.038	1.009	1.068
	minRhm	-.080	.017	21.894	< .001	.923	.884	.965
	effRhm	-.062	.019	10.559	.001	.940	.895	.987
	sumRn7	-.036	.015	5.925	.015	.965	.929	1.002
9	maxTa	.060	.015	16.353	< .001	1.062	1.022	1.104
	avgRhm	-.038	.016	5.959	.015	.963	.925	1.002
	effRhm	-.059	.023	6.420	.011	.943	.889	1.001

분석 결과 해석

상관행렬

✓ maxTa, effRhm 2가지 변수가 강원도 9개 지역에서 전반적으로 산불발생 여부와 관련하여 유의한 결과를 보임

지역 No.	변수들의 계수 값 (Coefficient)													
	avgTa	minTa	maxTa	sumRn	max InsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf	max wind7	noRn
1	-.046	.341	-.173	.048	-.088	-.032	-.096	-.039	-.041	-.393	.178	.041	-.388	-.044
2	.015	.270	-.218	-.041	-.039	-.049	-.150	-.005	-.092	-.374	.211	.087	-.519	-.079
3	.165	.175	-.325	.005	-.229	.077	-.038	-.087	-.123	-.250	.148	.125	-.511	-.033
4	.089	.194	-.270	-.010	.003	.017	-.282	-.029	-.111	-.352	.221	.024	-.440	-.047
5	.057	.327	-.327	-.004	.016	-.088	-.161	-.053	-.161	-.335	.169	.012	-.428	.014
6	.185	.147	-.354	.013	.054	.001	-.293	-.010	-.198	-.269	.139	.154	-.426	.012
7	-.039	.217	-.112	.011	-.181	.076	-.062	.122	-.167	-.300	.196	.059	-.310	-.032
8	.123	.167	-.293	.007	-.148	.053	-.068	-.077	-.034	-.389	.217	.058	-.437	-.080
9	.156	.157	-.312	.019	-.202	.075	-.070	-.059	-.034	-.373	.232	.166	-.465	-.068

✓ Cell Color : p-value > 0.01 p-value > 0.1

분석 결과 해석

상관행렬

✓ avgTa, sumRn, maxInsWs, maxWs, noRn
 5가지 변수는 강원도 9개 지역별로 부호가 일정하지 않아
 산불 발생에 미치는 영향이 분명하지 않음

지역 No.	변수들의 계수 값 (Coefficient)													
	avgTa	minTa	maxTa	sumRn	max InsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf	max wind7	noRn
1	-.046	.341	-.173	.048	-.088	-.032	-.096	-.039	-.041	-.393	.178	.041	-.388	-.044
2	.015	.270	-.218	-.041	-.039	-.049	-.150	-.005	-.092	-.374	.211	.087	-.519	-.079
3	.165	.175	-.325	.005	-.229	.077	-.038	-.087	-.123	-.250	.148	.125	-.511	-.033
4	.089	.194	-.270	-.010	.003	.017	-.282	-.029	-.111	-.352	.221	.024	-.440	-.047
5	.057	.327	-.327	-.004	.016	-.088	-.161	-.053	-.161	-.335	.169	.012	-.428	.014
6	.185	.147	-.354	.013	.054	.001	-.293	-.010	-.198	-.269	.139	.154	-.426	.012
7	-.039	.217	-.112	.011	-.181	.076	-.062	.122	-.167	-.300	.196	.059	-.310	-.032
8	.123	.167	-.293	.007	-.148	.053	-.068	-.077	-.034	-.389	.217	.058	-.437	-.080
9	.156	.157	-.312	.019	-.202	.075	-.070	-.059	-.034	-.373	.232	.166	-.465	-.068

✓ Cell Color : p-value > 0.01 p-value > 0.1

분석 결과 해석

로지스틱 회귀모델식 산출

구분	산불 발생확률 예측 모델 (Logistic Regression)	모델 설명 변수
지역 1	$[1 + \exp\{-(-3.167 - (0.359 * avgTa) + (0.404 * maxTa) + (0.131 * maxInsWs) - (0.058 * minRhm) - (0.068 * effRhm) - (0.045 * sumRn7))\}^{-1}]^{-1}$	✓ avgTa, maxTa, maxInsWs, minRhm, effRhm, sumRn7
지역 2	$[1 + \exp\{-(-3.412 - (0.169 * minTa) + (0.199 * maxTa) - (0.096 * effRhm) - (0.959 * Rntf))\}^{-1}]^{-1}$	✓ minTa, maxTa, effRhm, Rntf
지역 3	$[1 + \exp\{-(0.651 + (0.059 * maxTa) - (0.034 * avgRhm) - (0.058 * effRhm))\}^{-1}]^{-1}$	✓ maxTa, avgRhm, effRhm
지역 4	$[1 + \exp\{-(2.125 - (0.069 * minTa) + (0.105 * maxTa) + (0.251 * maxInsWs) - (0.919 * avgWs) - (0.030 * minRhm) - (0.049 * effRhm) - (0.079 * sumRn7) + (0.186 * maxwind7))\}^{-1}]^{-1}$	✓ minTa, maxTa, maxInsWs, avgWs minRhm, effRhm, sumRn7, maxwind7
지역 5	$[1 + \exp\{-(1.917 + (0.066 * maxTa) - (0.583 * avgWs) - (0.124 * effRhm))\}^{-1}]^{-1}$	✓ maxTa, avgWs, effRhm
지역 6	$[1 + \exp\{-(3.449 + (0.417 * avgWs) - (0.023 * avgRhm) - (0.056 * sumRn7))\}^{-1}]^{-1}$	✓ avgWs, avgRhm, sumRn7
지역 7	$[1 + \exp\{-(1.150 - (0.232 * avgTa) + (0.261 * maxTa) + (0.112 * maxInsWs) - (0.096 * effRhm) - (0.079 * sumRn7))\}^{-1}]^{-1}$	✓ avgTa, maxTa, maxInsWs, effRhm, sumRn7
지역 8	$[1 + \exp\{-(1.181 + (0.037 * maxTa) - (0.080 * minRhm) - (0.062 * effRhm) - (0.036 * sumRn7))\}^{-1}]^{-1}$	✓ maxTa, minRhm, effRhm, sumRn7
지역 9	$[1 + \exp\{-(0.240 + (0.060 * maxTa) - (0.038 * avgRhm) - (0.059 * effRhm))\}^{-1}]^{-1}$	✓ maxTa, avgRhm, effRhm

Chapter 4.

프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

데이터 탐색

FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

통계 분석

분석 개요
분석 내용
분석 결과 해석

ML 모델링

산불 예측 모델 생성
모델 학습
모델 성능 검증
산불 위험 예측

DL 모델링

산불 분류 모델 생성
모델 학습
모델 성능 검증
산불 이미지 분류 예측

서비스 및 문서

서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고 문헌 / 부록



프로젝트 소개

데이터 탐색

통계 분석

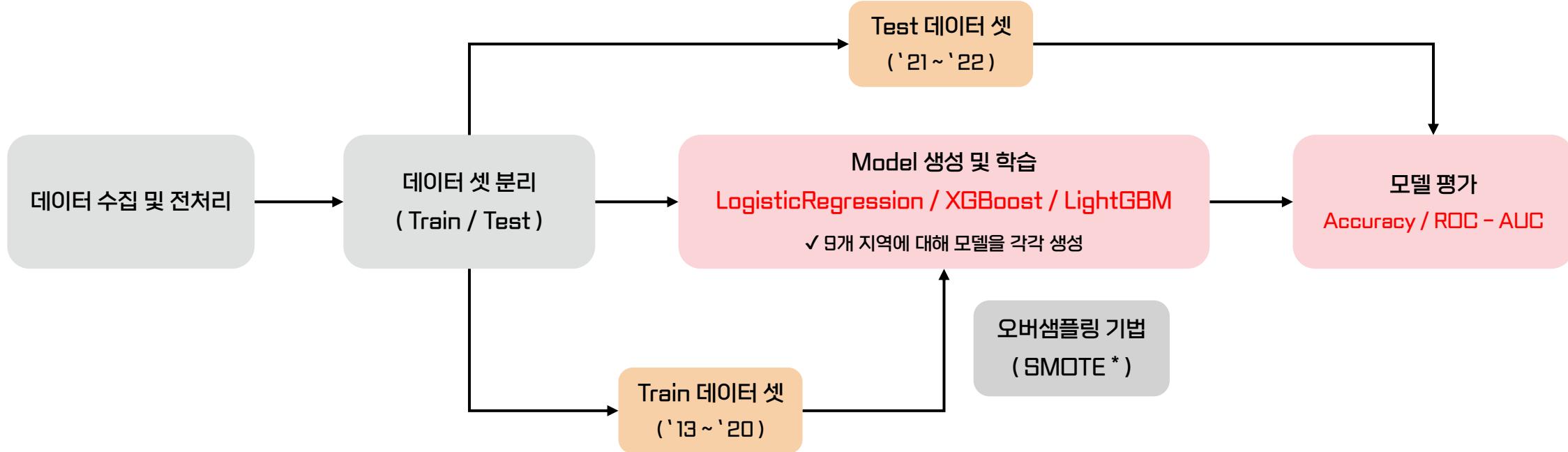
ML 모델링

DL 모델링

서비스 및 문서

산불 예측 모델 생성 및 학습

모델 훈련 및 평가 순서도

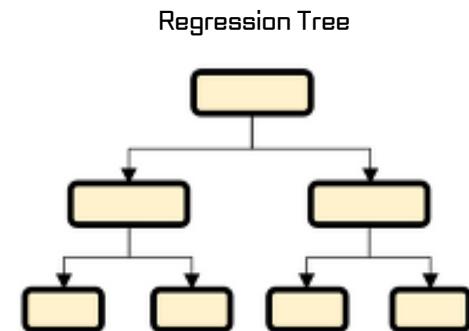
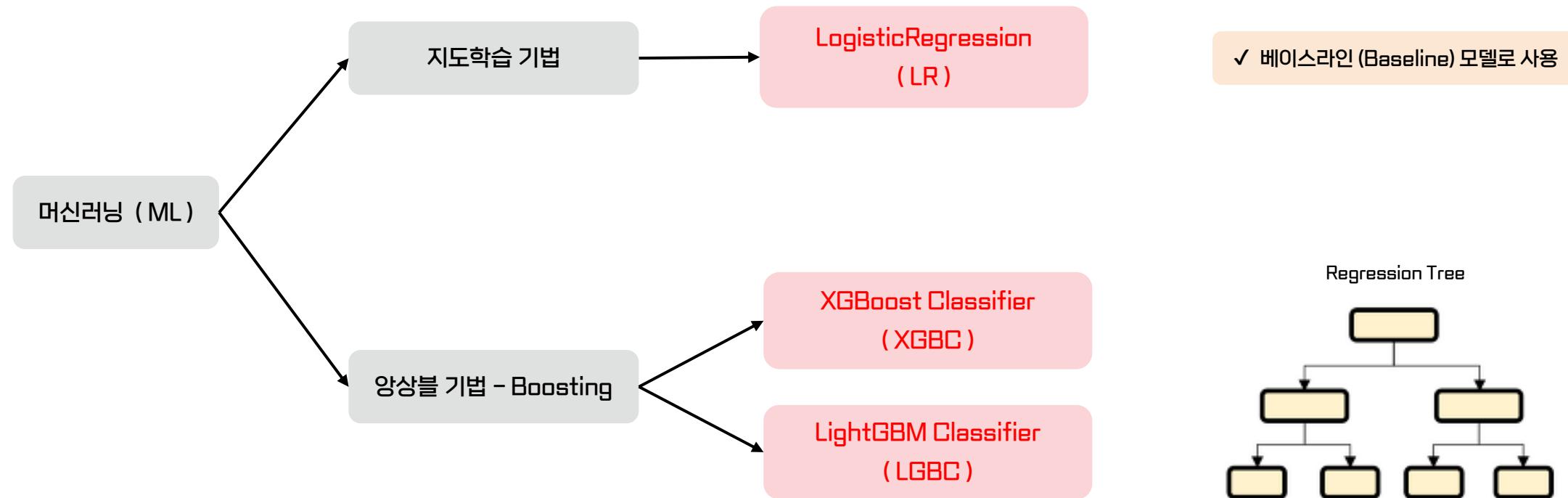


Point. SMOTE (Synthetic Minority Over-sampling Technique)

- ✓ 오버샘플링 기법의 하나로, 산불 발생과 관련된 **더미 종속변수 (fire_occur)** 데이터의 불균형 해결을 위해 사용
- ✓ SMOTE는 오버샘플링의 단점인 **과적합 (Over-Fitting)**을 보완 할 수 있는 방법으로
자료를 복제하는 것이 아니며, **기존의 자료를 적절히 조합하여 새로운 샘플을 만든다.**

산불 예측 모델 생성 및 학습

모델 생성 및 학습



산불 예측 모델 생성 및 학습

모델 생성 및 학습

Logistic Regression (LR)

- ✓ 로지스틱 회귀(Logistic Regression)의 시그모이드(Sigmoid) 함수 최적선을 찾음
- ✓ 임계값(Threshold Value)을 기준으로 이진 분류(산불 발생 / 미발생) 예측 성능을 확인

XGBoost Classifier (XGBC)

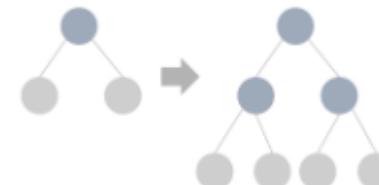
- ✓ 균형 트리 분할(Level Wise) 방식 사용
- ✓ 트리의 균형을 맞추면서, 지속적으로 분할
- ✓ 가중치 부여를 경사하강법을 이용하여 진행하여 과적합(Overfitting)을 방지

LightGBM Classifier (LGBC)

- ✓ 리프 중심 트리 분할(Leaf Wise) 방식 사용
- ✓ 트리의 균형을 맞추지 않고, 최대 손실값을 가지는 리프 노드를 지속적으로 분할
- ✓ 예측 오류 손실을 최소화 / 속도가 빨라 시간 절약

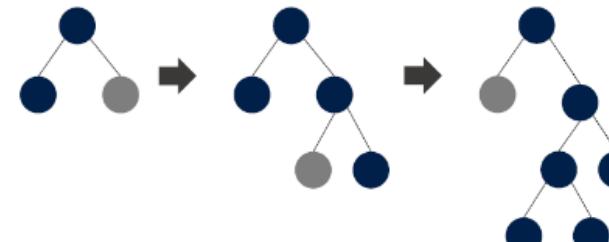
XGBoost Classifier

균형 트리 분할 (Level Wise)



LightGBM Classifier

리프 중심 트리 분할 (Leaf Wise)



산불 예측 모델 생성 및 학습

모델 평가지표

평가지표 : Accuracy

혼동행렬 (Confusion Matrix)		예측값	
		TRUE	FALSE
실제값	TRUE	TP (True Positive)	FN (False Negative)
	FALSE	FP (False Positive)	TN (True Negative)

✓ Accuracy = $\frac{TP + TN}{TP + TN + FP + FN}$

평가지표 : ROC - AUC

혼동행렬 (Confusion Matrix)		예측값	
		TRUE	FALSE
실제값	TRUE	TP (True Positive)	FN (False Negative)
	FALSE	FP (False Positive)	TN (True Negative)

✓ TPR (Sensitivity) = $\frac{TP}{TP + FN}$

✓ FPR (1-Specificity) = $\frac{FP}{FP + TN}$



모델 성능 검증

모델 평가지표 : Accuracy

✓ Accuracy 를 평가지표로 했을 때,
강원도 9개 지역 전반적으로 XGBC 모델이 선택됨

구분	지역 1	지역 2	지역 3	지역 4	지역 5	지역 6	지역 7	지역 8	지역 9
LR	0.859	0.852	0.692	0.773	0.804	0.705	0.779	0.781	0.779
XGBC	0.975	0.963	0.904	0.886	0.973	0.941	0.922	0.901	0.922
LGBM	0.974	0.964	0.859	0.838	0.937	0.882	0.941	0.862	0.941
모델 선택	(XGBC) acc. = 0.98	(XGBC) acc. = 0.96	(XGBC) acc. = 0.90	(XGBC) acc. = 0.89	(XGBC) acc. = 0.97	(XGBC) acc. = 0.94	(LGBM) acc. = 0.94	(XGBC) acc. = 0.90	(LGBM) acc. = 0.94

✓ Best Accuracy Score Model

2021 ~ 22년 산불 발생 관측 수

0: 산불 미발생 1: 산불 발생	0: 722 1: 8	0: 719 1: 11	0: 721 1: 9	0: 706 1: 24	0: 728 1: 2	0: 716 1: 14	0: 714 1: 16	0: 707 1: 23	0: 714 1: 16
혼동 행렬	[707 15] [4 4]	[618 101] [7 4]	[496 225] [0 9]	[554 152] [13 11]	[683 45] [1 11]	[506 210] [5 9]	[556 158] [3 13]	[551 156] [4 19]	[556 158] [3 13]

✓ TN (True Negative) : 혼동행렬의 (0, 0) / TP (True Positive) : 혼동행렬의 (1, 1)

모델 성능 검증

모델 평가지표 : ROC - AUC

✓ ROC - AUC 를 평가지표로 했을때,
강원도 9개 지역 전반적으로 LR 모델이 선택됨

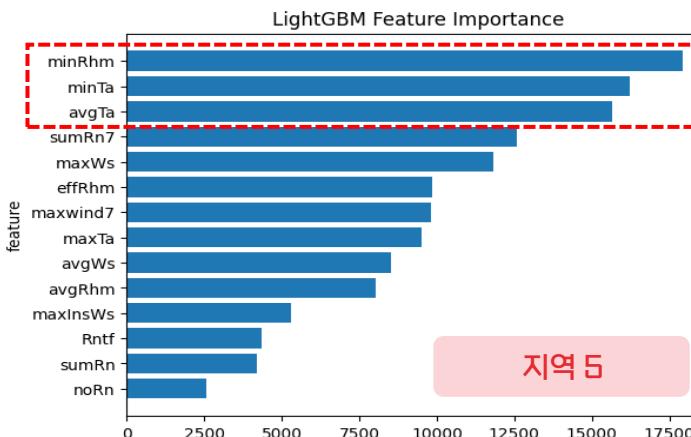
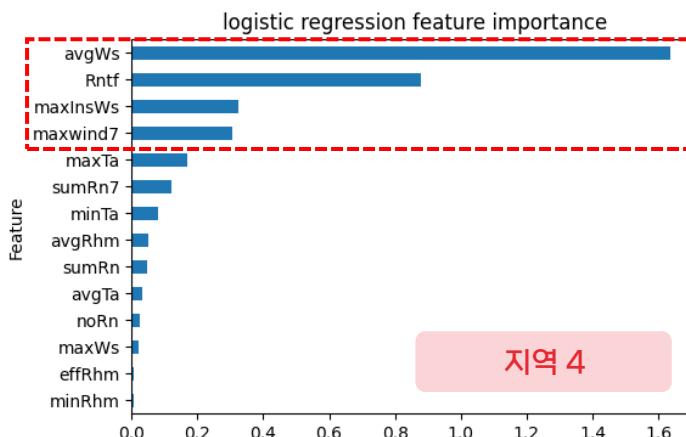
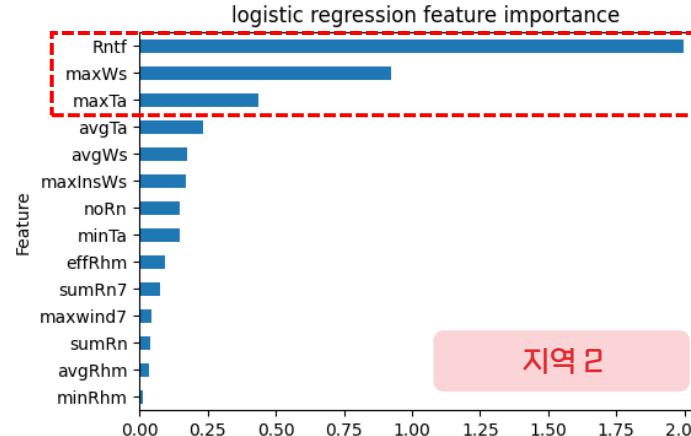
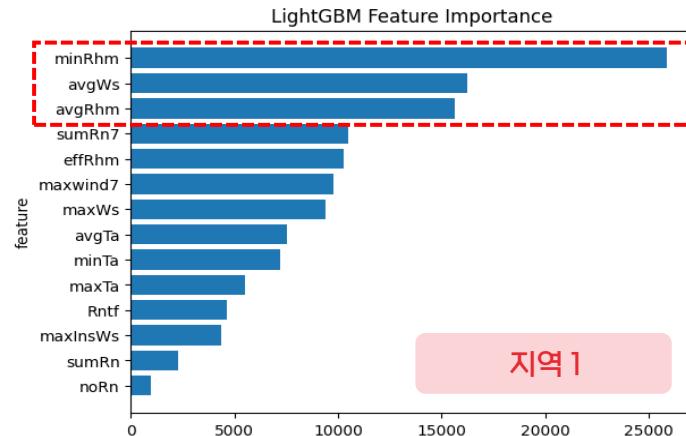
구분	지역 1	지역 2	지역 3	지역 4	지역 5	지역 6	지역 7	지역 8	지역 9
LR	0.892	0.791	0.795	0.752	0.658	0.791	0.837	0.869	0.837
XGBC	0.807	0.686	0.639	0.658	0.659	0.759	0.770	0.747	0.770
LGBC	0.905	0.733	0.742	0.711	0.856	0.790	0.763	0.816	0.763
모델 선택	⟨ LGBC ⟩ AUC = 0.91	⟨ LR ⟩ AUC = 0.79	⟨ LR ⟩ AUC = 0.80	⟨ LR ⟩ AUC = 0.75	⟨ LGBC ⟩ AUC = 0.86	⟨ LR ⟩ AUC = 0.79	⟨ LR ⟩ AUC = 0.84	⟨ LR ⟩ AUC = 0.87	⟨ LR ⟩ AUC = 0.84

✓ Best ROC - AUC Score Model

모델 성능 검증

Feature Importance

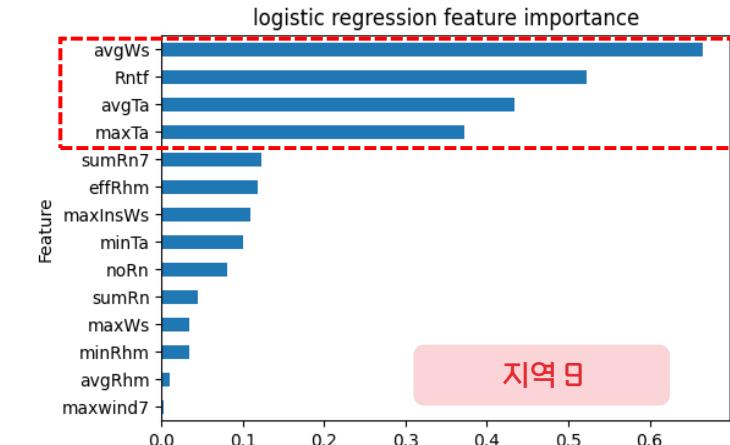
✓ 9개 지역별 Best AUC Score Model 을
기준으로 Feature Importance 비교



모델 성능 검증

Feature importance

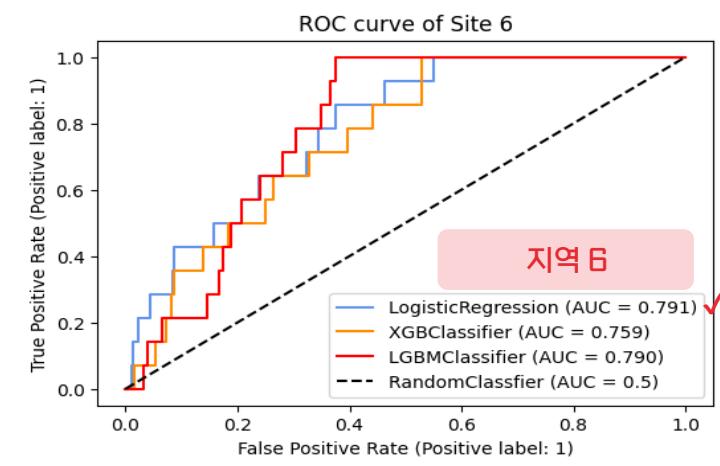
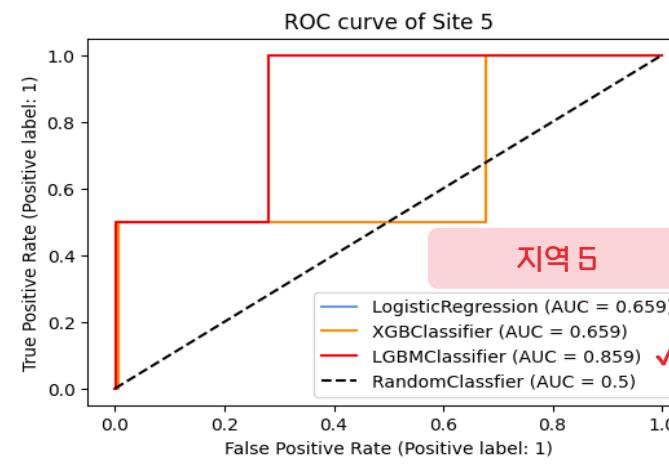
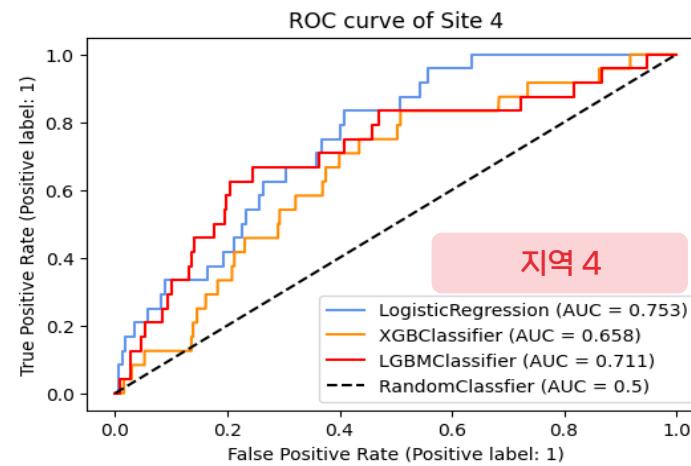
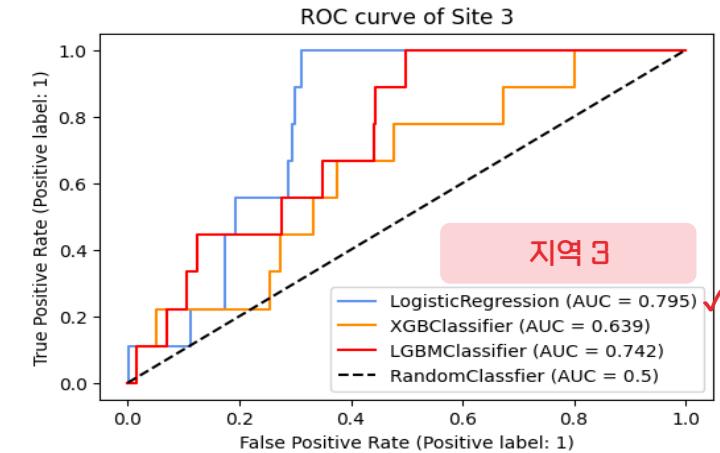
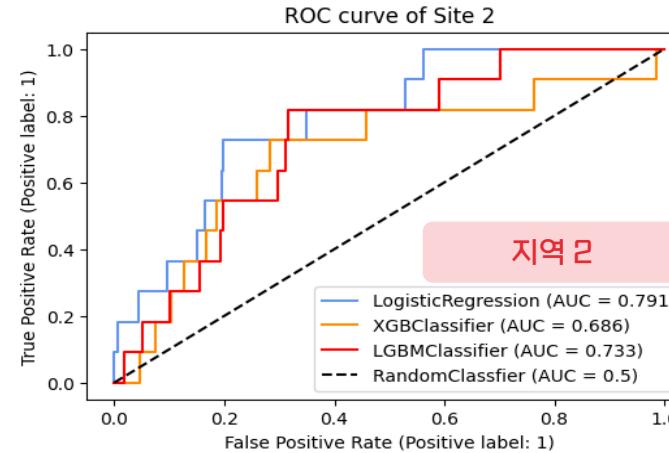
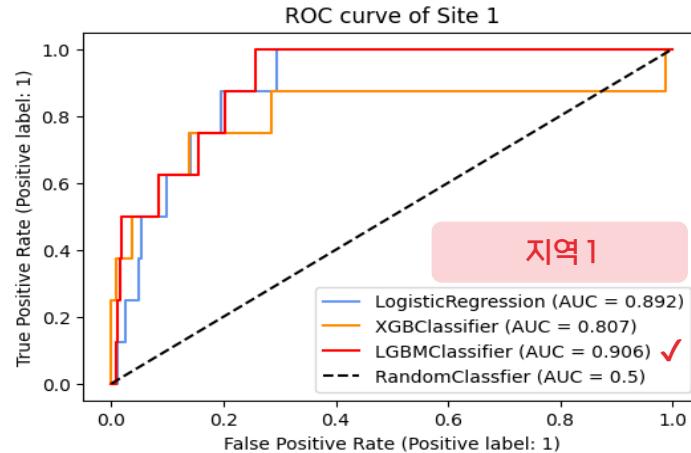
✓ 9개 지역별 Best AUC Score Model 을
기준으로 Feature Importance 비교



모델 성능 검증

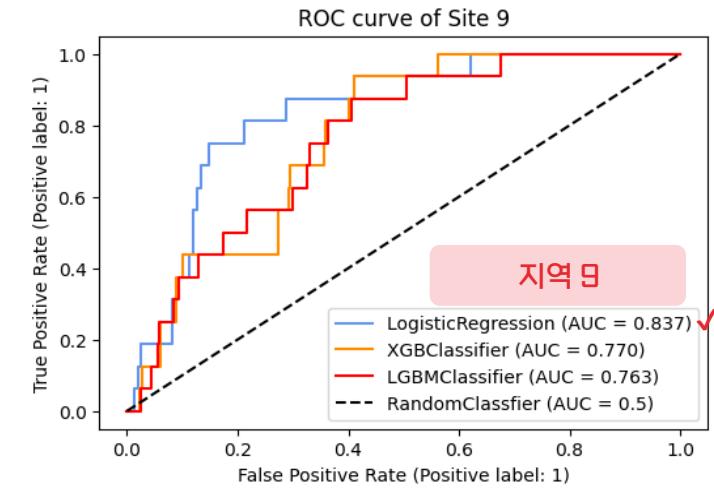
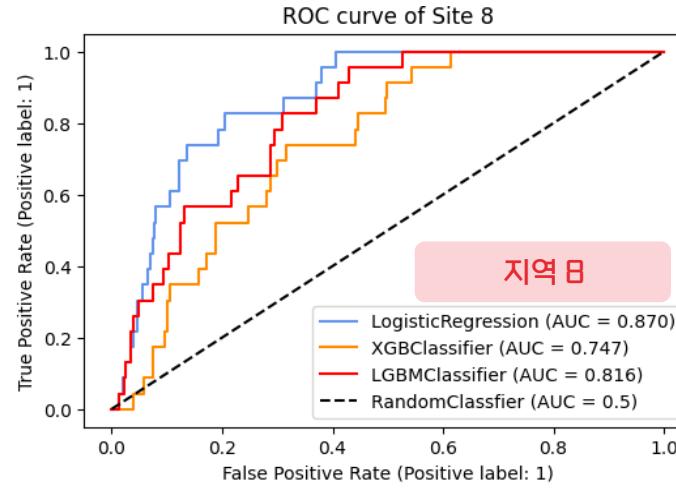
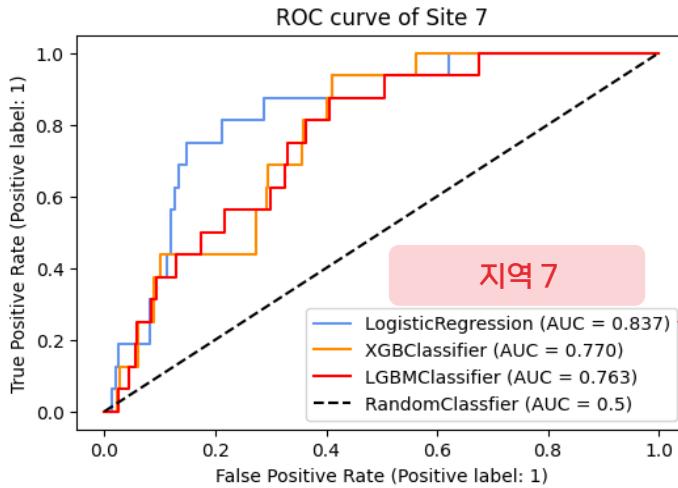
ROC - AUC Curve

✓ 9개 지역별 Best ROC - AUC Curve 를
통해 AUC Score 비교



모델 성능 검증

ROC - AUC Curve



Point. 산불 위험지수 (DWI)

- ✓ Feature Importance를 통해 선택된 변수들을 고려
- ✓ 각 지역별로 ROC-AUC Curve를 통해 얻어지는 AUC Score가 높은 모델을 선택
- ✓ 모델을 통해 얻어진 산불발생확률 예측치를 이용해 산불위험지수 (DWI ; 1~10) 설정

산불 위험 예측

산불위험지수 (DWI) 확률구간

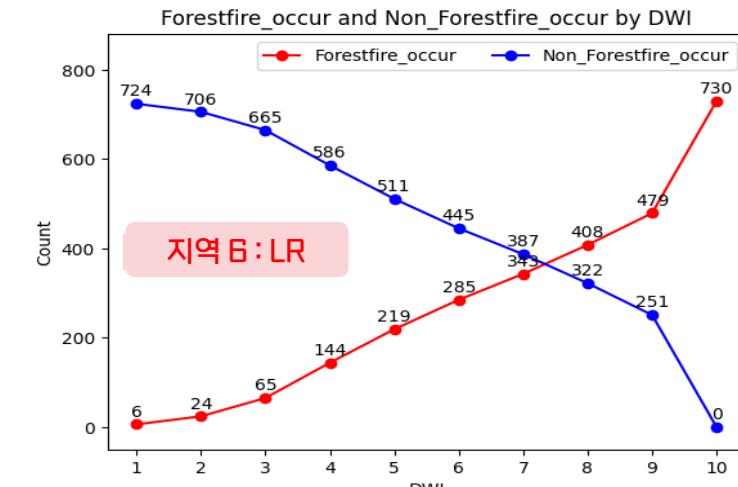
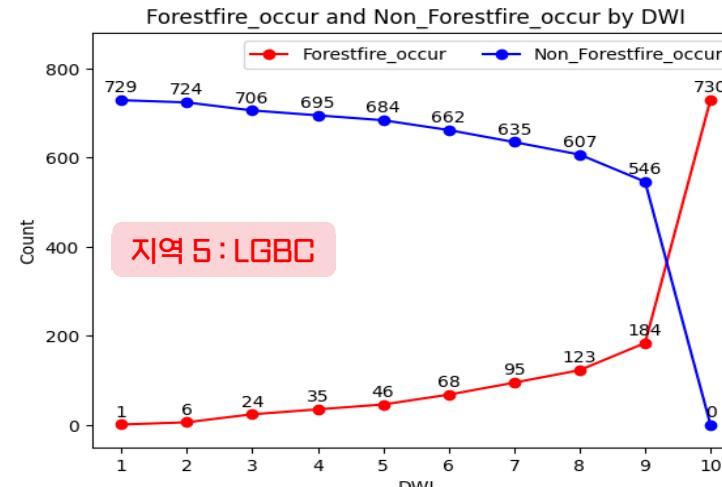
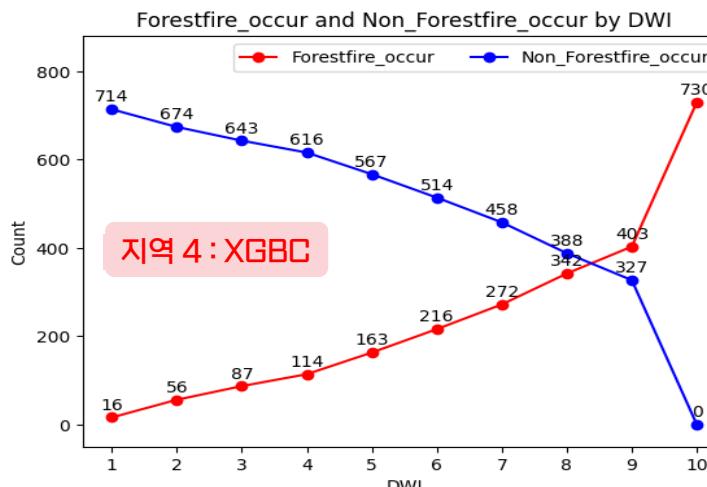
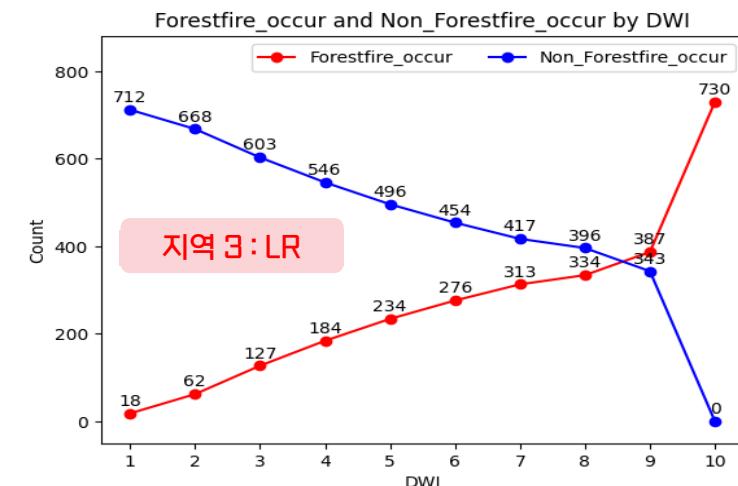
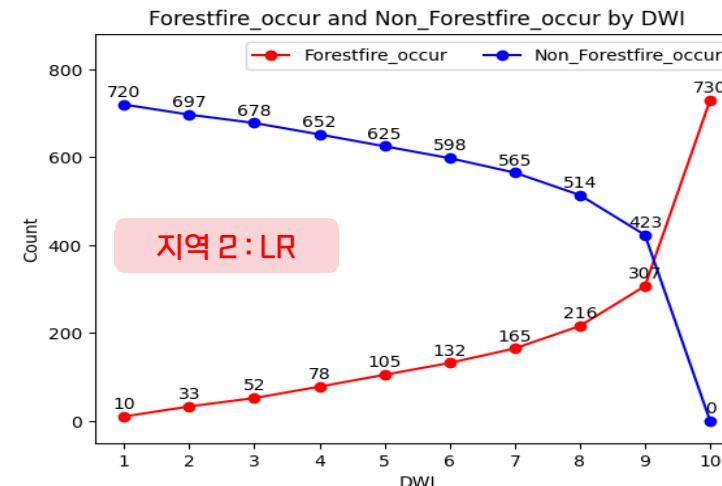
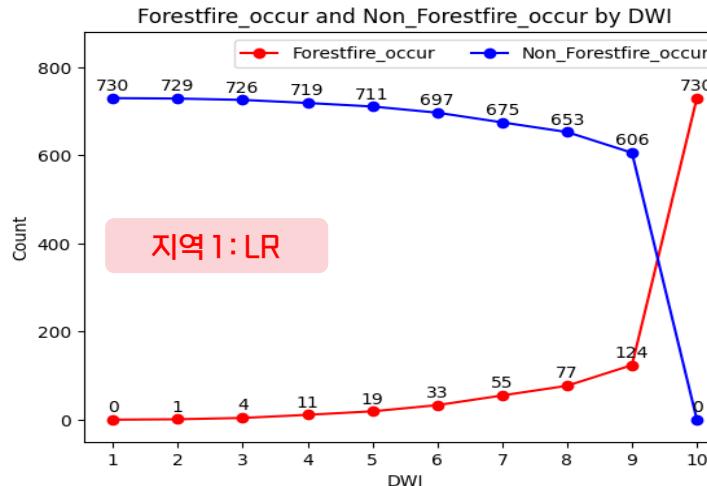
✓ 9개 지역별 Best ROC - AUC Score Model에 데이터를 대입하여 얻어진 확률들의 총 예측치를 이용하여, 10% 구간별 백분위수 추정 및 해당 예측 확률구간 지수화 설정

확률 구간	DWI	예상 확률 구간 (Estimated Probability Interval Range)									
		지역 1 (LGBC)	지역 2 (LR)	지역 3 (LR)	지역 4 (LR)	지역 5 (LGBC)	지역 6 (LR)	지역 7 (LR)	지역 8 (LR)	지역 9 (LR)	
10%	1	[.0000 ~ .0012]	[.0000 ~ .0006]	[.0000 ~ .0000]	[.0000 ~ .0000]	[.0000 ~ .0003]	[.0000 ~ .0000]	[.0000 ~ .0000]	[.0000 ~ .0009]	[.0000 ~ .0000]	
20%	2	[.0012 ~ .0020]	[.0006 ~ .0047]	[.0000 ~ .0009]	[.0000 ~ .0018]	[.0003 ~ .0009]	[.0000 ~ .0053]	[.0000 ~ .0016]	[.0009 ~ .0057]	[.0000 ~ .0016]	
30%	3	[.0020 ~ .0034]	[.0047 ~ .0163]	[.0009 ~ .0142]	[.0016 ~ .0130]	[.0009 ~ .0023]	[.0053 ~ .0614]	[.0016 ~ .0146]	[.0057 ~ .0291]	[.0016 ~ .0146]	
40%	4	[.0034 ~ .0066]	[.0163 ~ .0385]	[.0142 ~ .0530]	[.0130 ~ .0587]	[.0023 ~ .0048]	[.0614 ~ .1442]	[.0146 ~ .0755]	[.0291 ~ .0708]	[.0146 ~ .0755]	
50%	5	[.0066 ~ .0108]	[.0385 ~ .0638]	[.0530 ~ .1351]	[.0587 ~ .1644]	[.0048 ~ .0109]	[.1442 ~ .2727]	[.0755 ~ .1768]	[.0708 ~ .1592]	[.0755 ~ .1768]	
60%	6	[.0108 ~ .0186]	[.0638 ~ .1126]	[.1351 ~ .3605]	[.1644 ~ .2680]	[.0109 ~ .0210]	[.2727 ~ .3877]	[.1768 ~ .2984]	[.1592 ~ .2621]	[.1768 ~ .2984]	
70%	7	[.0186 ~ .0364]	[.1126 ~ .1935]	[.3605 ~ .5306]	[.2680 ~ .3958]	[.0210 ~ .0520]	[.3877 ~ .4996]	[.2984 ~ .4049]	[.2621 ~ .4016]	[.2984 ~ .4049]	
80%	8	[.0364 ~ .0826]	[.1935 ~ .3390]	[.5306 ~ .6671]	[.3958 ~ .5289]	[.0520 ~ .1572]	[.4996 ~ .5908]	[.4049 ~ .5482]	[.4016 ~ .5822]	[.4049 ~ .5482]	
90%	9	[.0826 ~ .02128]	[.3390 ~ .6251]	[.6671 ~ .7907]	[.5289 ~ .7430]	[.1572 ~ .3714]	[.5908 ~ .6930]	[.5482 ~ .7583]	[.5822 ~ .7926]	[.5482 ~ .7583]	
100%	10	[.2128 ~ 1.000]	[.6251 ~ 1.000]	[.7907 ~ 1.000]	[.7430 ~ 1.000]	[.3714 ~ 1.000]	[.6930 ~ 1.000]	[.7583 ~ 1.000]	[.7926 ~ 1.000]	[.7583 ~ 1.000]	

산불 위험 예측

산불위험지수 산출

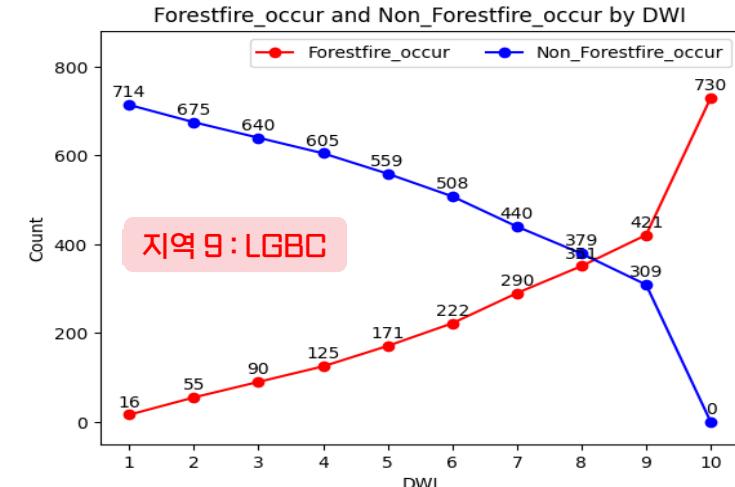
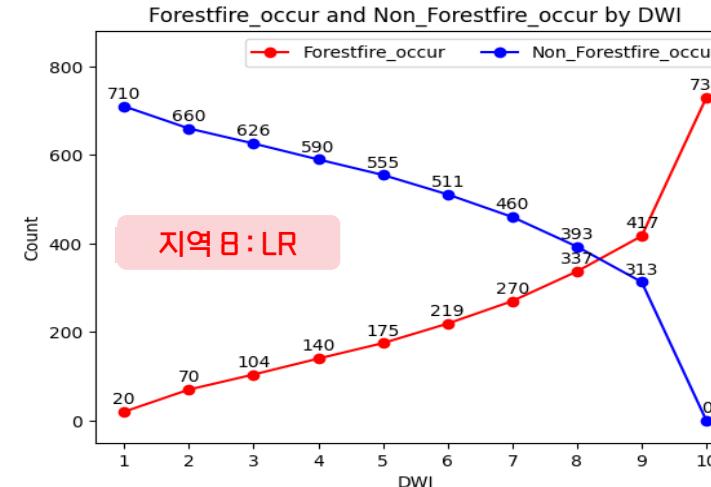
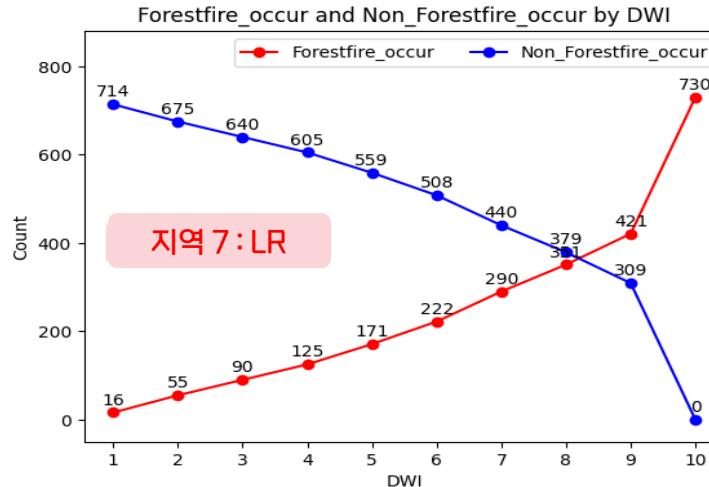
✓ 9개 지역별 **산불위험지수** 를 산출하여
그래프로 비교



산불 위험 예측

산불위험지수 산출

- ✓ 9개 지역별 산불위험지수를 산출하여
그래프로 비교



Point. 산불 위험지수 산출

- ✓ 실제로 산불이 발생한 날들의 DWI가 높고,
산불이 발생하지 않은 날의 DWI는 낮은 것으로 나타남
- ✓ 따라서, DWI 가 산불발생 위험도를 잘 예측할 수 있는 지수라는 것을 의미

Point. 산불 발생예측 모델 기대효과

- ✓ 지역별 산불 주요 요인에 따른 산불발생 예측 가능
- ✓ 산불발생 예측을 이용한 선제적 산불 예방체계 마련계획 수립 가능

Chapter 5.

프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

데이터 탐색

FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

통계 분석

분석 개요
분석 내용
분석 결과 해석

ML 모델링

산불 예측 모델 생성 및 학습
모델 성능 검증
산불 위험 예측

DL 모델링

산불 분류 모델 생성 및 학습
모델 성능 검증
산불 이미지 분류 예측

서비스 및 문서

서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고 문헌 / 부록



프로젝트 개요

데이터 탐색

통계 분석

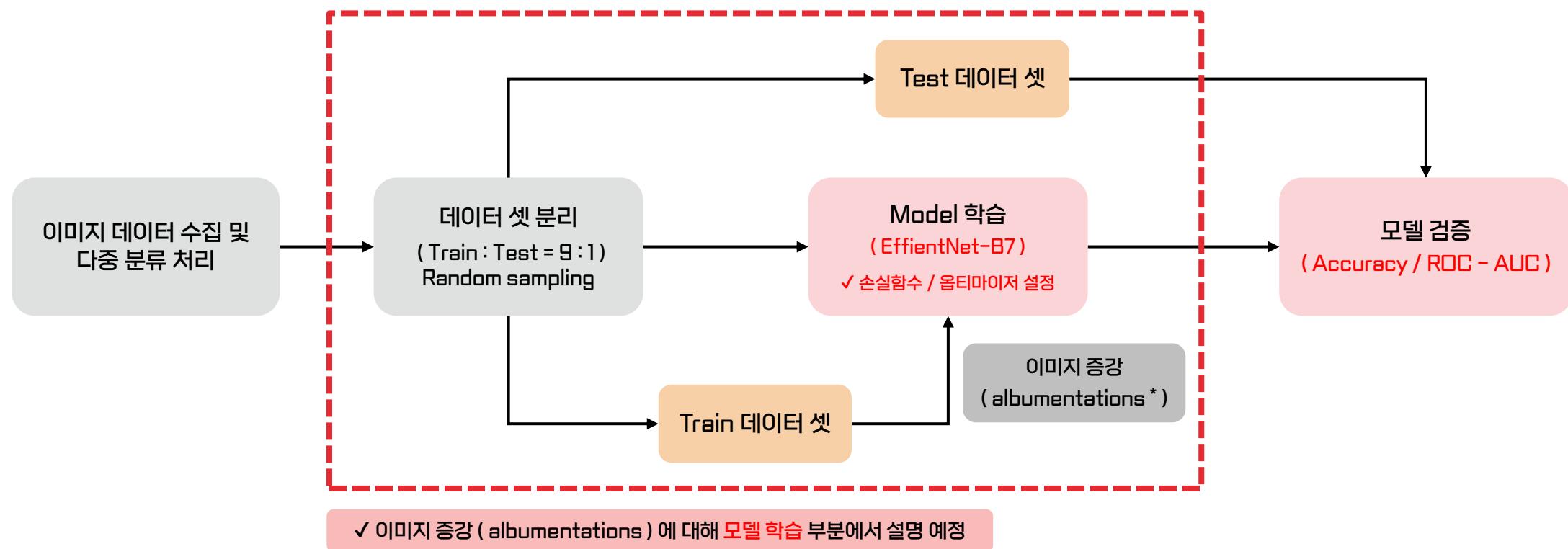
ML 모델링

DL 모델링

서비스 및 문서

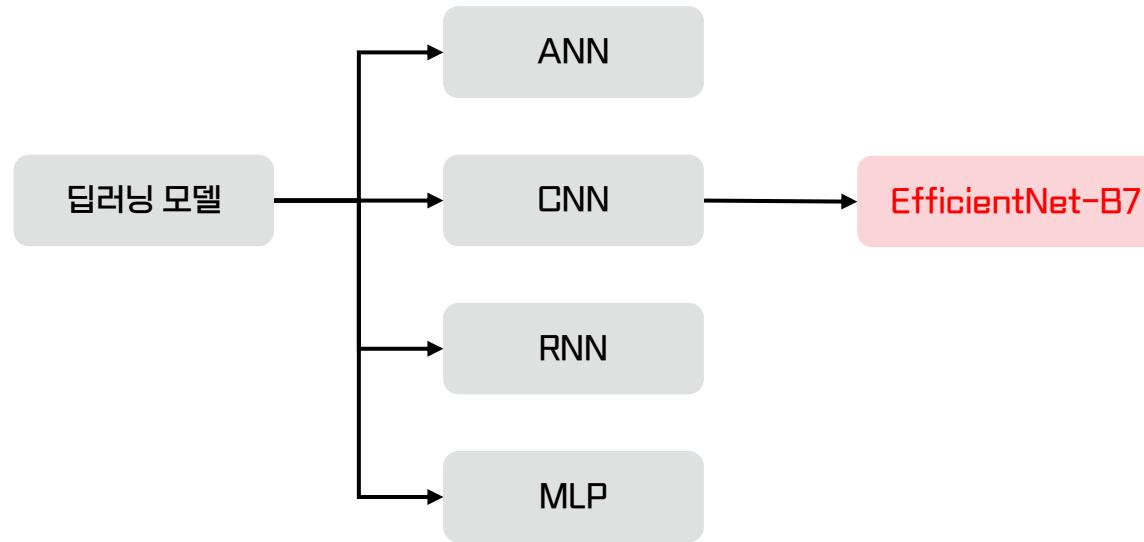
산불 분류 모델 생성 및 학습

모델 훈련 및 평가 순서도



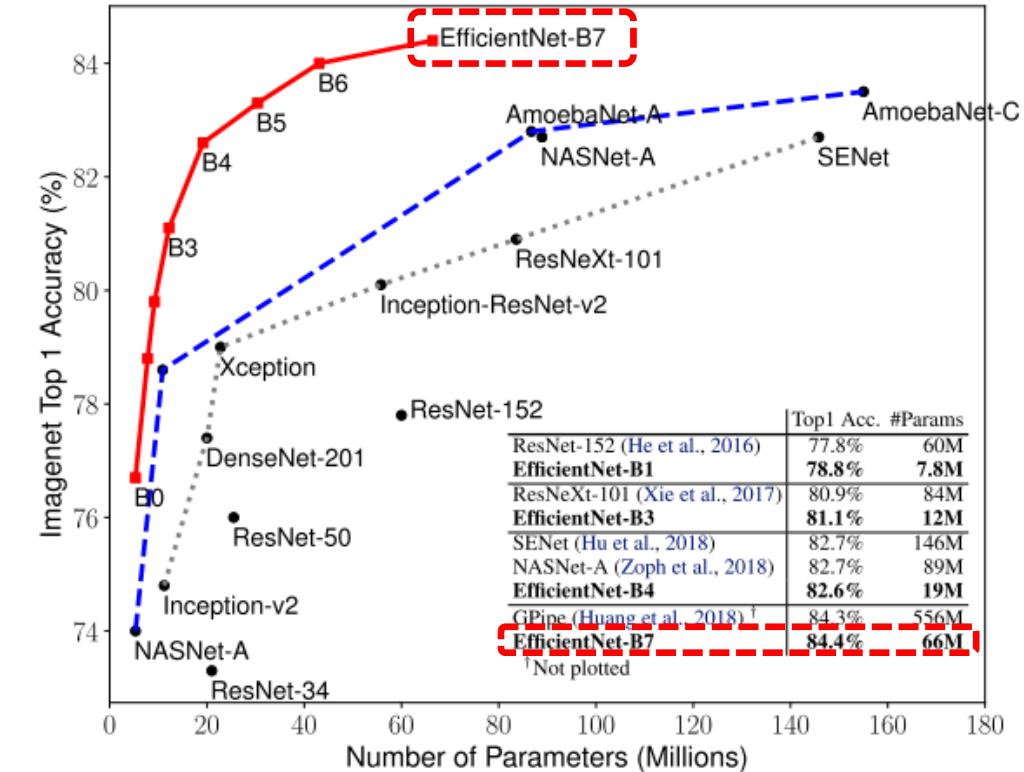
산불 분류 모델 생성 및 학습

EfficientNet



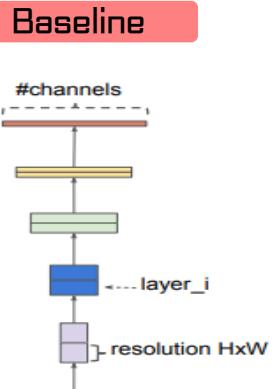
Point. EfficientNet-B7

- ✓ 기존 모델들은 성능 개선을 위해 Width, Depth, Resolution 중 하나만 Scaling 하였다.
- ✓ EfficientNet-B7은 Compound Scaling 을 이용해 적절한 Weight 를 주어 더 적은 parameter 로 더 빠르고 더 정확한 연산 을 한다.

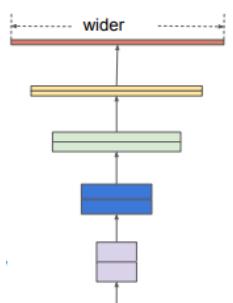


산불 분류 모델 생성 및 학습

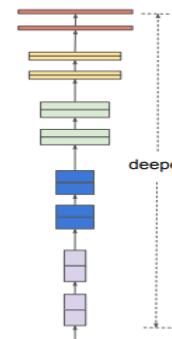
Compound Scaling



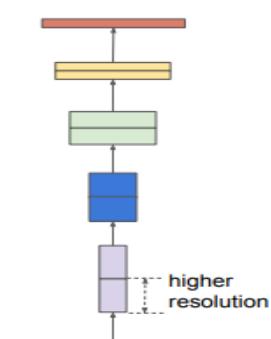
Width



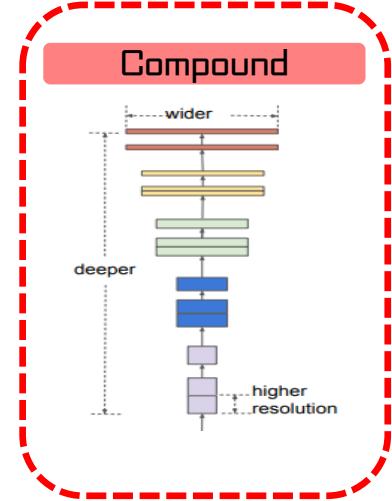
Depth



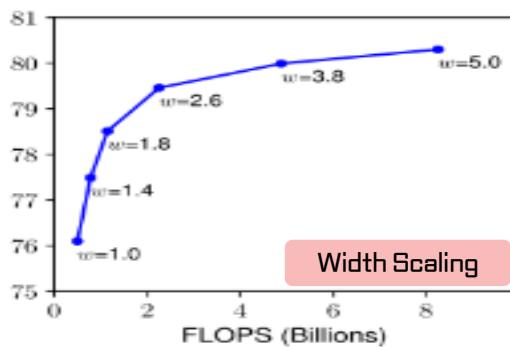
Resolution



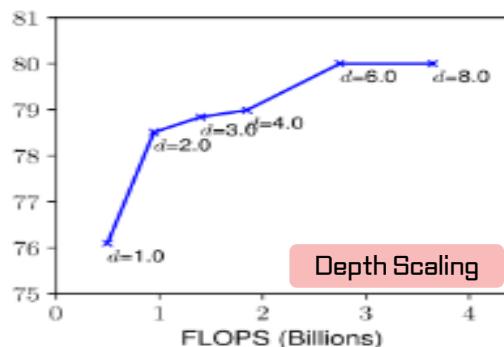
Compound



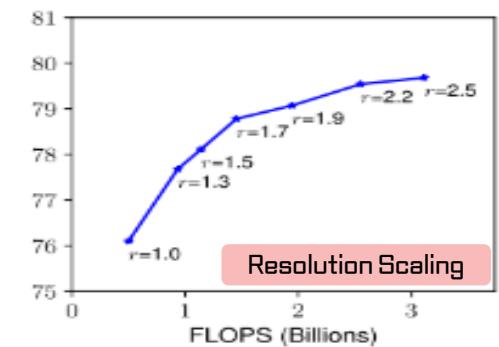
가중치에 따른 정확도와 연산 수 관계



- ✓ Filter의 개수 (Channel)를 늘리는 방법
 - 정확도 80%가 되려면 Filter를 5배 늘려야 한다.



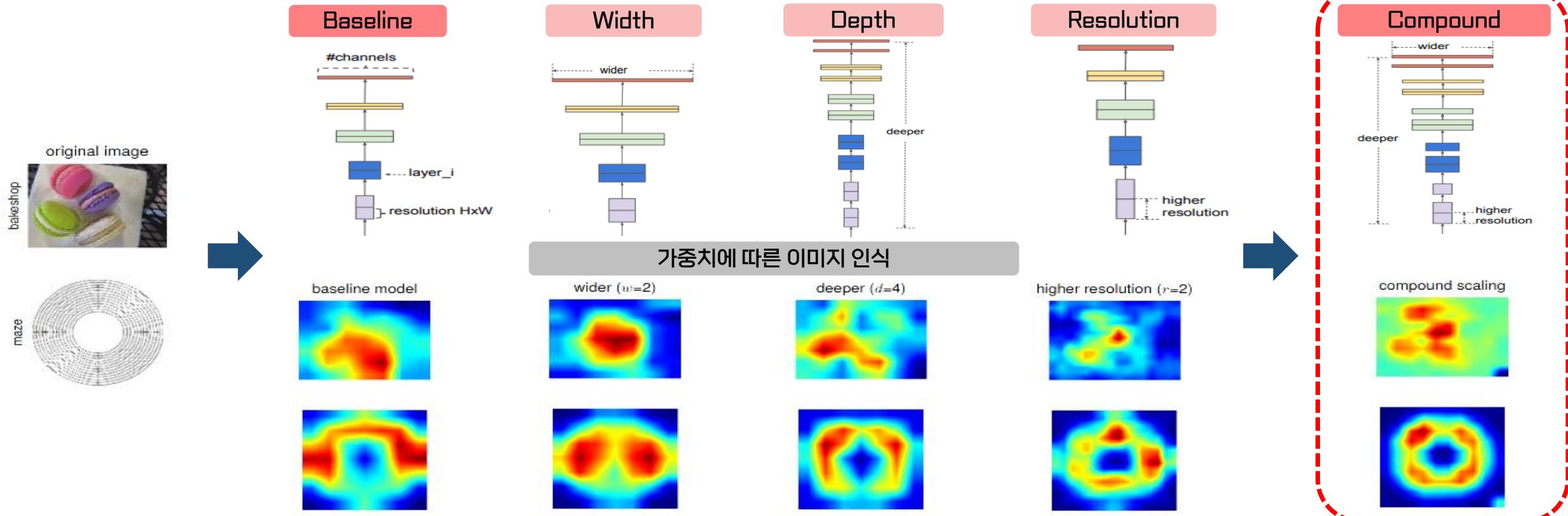
- ✓ Layer의 개수를 늘리는 방법
 - 정확도 80%가 되려면 Layer를 8배 늘려야 한다.



- ✓ 해상도 (Image Size)를 높이는 방법
 - 정확도 80% 정도가 되려면 해상도를 2.5배 높여야 한다.

산불 분류 모델 생성 및 학습

Compound Scaling



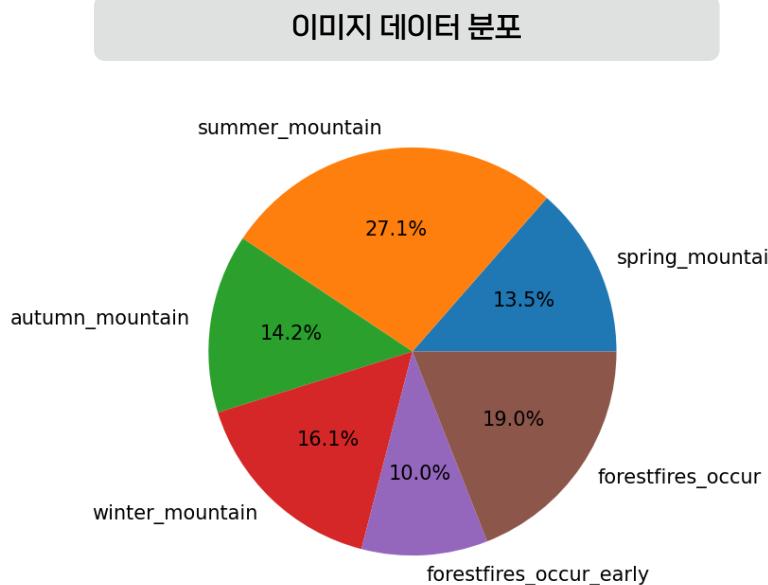
Point. Compound Scaling

- ✓ Width Scaling, Depth Scaling, Resolution Scaling

각각에 적절한 Weight 값을 주어 비슷한 연산 수로 더 좋은 정확도를 나타낸다.

산불 분류 모델 생성 및 학습

이미지 데이터 정의



Point. 다중 분류 처리

- ✓ 이미지 데이터 셋과 이미지를 수집하는 과정에서 **6개의 클래스로 구분** 하여 Train 데이터 셋 생성
- ✓ **원-핫 인코딩 (One-Hot encoding)** 방식을 이용하여,
표현하고 싶은 단어의 인덱스 1의 값을 부여, 그렇지 않은 단어에는 0을 부여하는 방식

산불 분류 모델 생성 및 학습

이미지 증강



Point. 이미지 증강 (albumentations)

- ✓ Image Resize : 이미지 크기 조정
- ✓ Image Brightness : 밝기와 대비 조정
- ✓ Image VerticalFlip : 이미지를 수직으로 뒤집는 작업
- ✓ Image HorizontalFlip : 이미지 수평으로 뒤집는 작업
- ✓ Image ShiftScaleRotate : 이미지 이동, 확대/축소, 회전
- ✓ Image Filter : Emboss / Sharpen / Blur 처리
- ✓ Image PiecewiseAffine : 이미지 왜곡
- ✓ Image Normalize : 이미지 정규화

산불 분류 모델 생성 및 학습

손실 함수와 옵티마이저 설정

```
100% [140/140] [02:15<00:00, 1.34it/s]
에폭 [20/20] - 훈련 데이터 손실값 : 0.4857
에폭 [20/20] - 검증 데이터 손실값 : 1.1469 / 검증 데이터 ROC AUC : 0.9441
=====
최종 모델의 ROC AUC: 0.9441259820426486
```

Point. 손실 함수 (Loss function)

- ✓ 모델 성능이 얼마나 나쁜지 측정하는 함수
- ✓ 최적 파라미터는 손실 함수가 최솟값일 때의 파라미터
- ✓ 예측값과 실제값의 차이를 구하는 함수

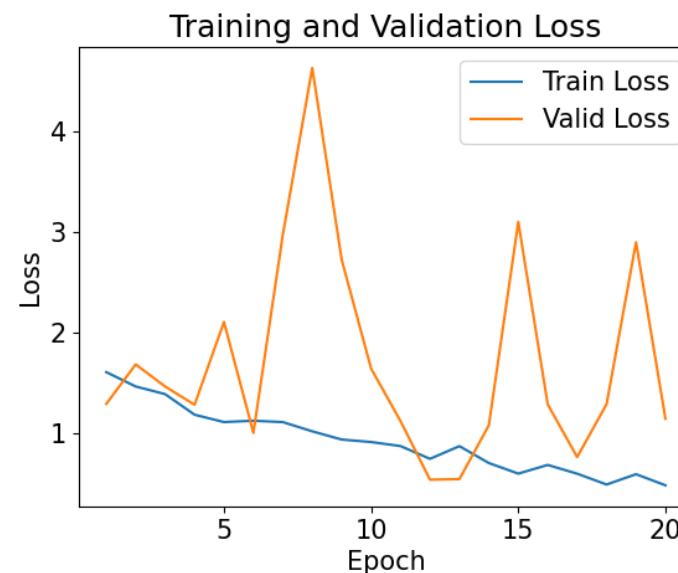
Point. 옵티마이저 (Optimizer)

- ✓ 신경망의 최적 가중치를 찾아주는 알고리즘
- ✓ 학습률과 가중치를 작게 조절하는 규제 기법인 가중치 감소를 적용

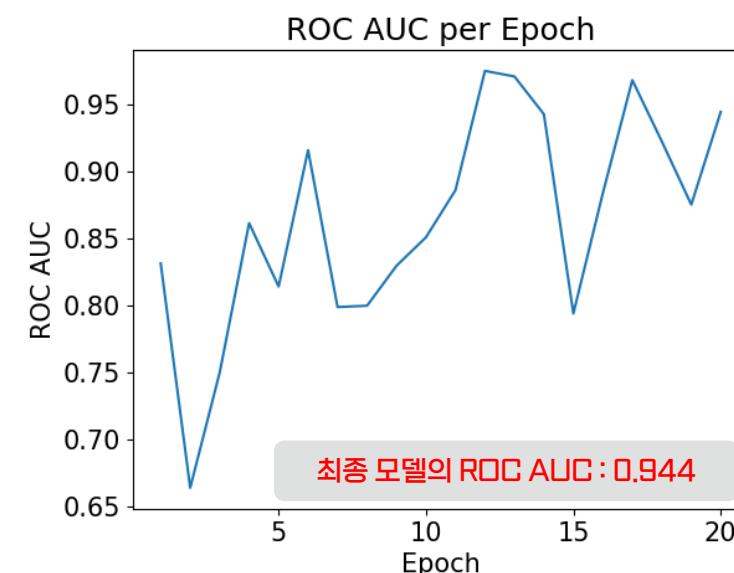
모델 성능 검증

모델 성능 검증

에폭수에 따른 손실값



에폭수에 따른 ROC - AUC



Point. 에폭 (Epoch)

- ✓ 훈련 데이터를 반복적으로 입력하여 모델을 업데이트
- ✓ Epoch의 수가 너무 적으면 과소적합이 일어날 수 있고, 너무 많으면 과대적합이 일어날 수 있다.
- ✓ Epoch이 증가하면 손실 (Loss) 값이 줄어들고, ROC-AUC 값은 증가함을 볼 수 있다.

산불 이미지 분류 예측

산불 이미지 분류 예측

Input image



Output

WARNING: Forest fires occur suspected!

9R29+RW 싱가포르, (1.352083, 103.819836)

Predicted Class and Probabilities:

spring_mountain:	0.02%
summer_mountain:	0.42%
autumn_mountain:	0.09%
winter_mountain:	2.99%
forestfires_occur_early:	58.37%
forestfires_occur:	38.10%

Non Forest fires occur suspected.

9R29+RW 싱가포르, (1.352083, 103.819836)

Predicted Class and Probabilities:

spring_mountain:	1.02%
summer_mountain:	95.97%
autumn_mountain:	0.03%
winter_mountain:	2.52%
forestfires_occur_early:	0.32%
forestfires_occur:	0.15%

But, 실제 작업 환경이 GPS 또는 Wifi 정보를 통해 위도, 경도에 대한 정보를 제공받을 수 없는 환경이라 기능 구현에 집중



Forest fire occur로 판단되면 경고

현재의 주소와 위도, 경도가 출력되게 구현

6개의 클래스에 대한 예측 확률에 대해 표기



Point. 산불 이미지 분류 모델 기대효과

- ✓ 산불 이미지 분류 예측을 통해 산불 조기 발견 가능
- ✓ 또한, 예측을 이용하여 산불 현장 대응체계 고도화

Chapter 6.

프로젝트 개요

팀 구성
주제 선정 배경
지수 설정
평가 지표
수행 절차 / 기간

데이터 탐색

FLOW CHART
ERD
데이터 수집
데이터 전처리
탐색적 데이터 분석 (EDA)

통계 분석

분석 개요
분석 내용
분석 결과 해석

ML 모델링

산불 예측 모델 생성
모델 학습
모델 성능 검증
산불 위험 예측

DL 모델링

산불 분류 모델 생성
모델 학습
모델 성능 검증
산불 이미지 분류 예측

서비스 및 문서

서비스 (Streamlit)
한계점 / 개선사항
프로젝트 문서
개발환경 / 참고 문헌 / 부록



프로젝트 개요

데이터 탐색

통계 분석

ML 모델링

DL 모델링

서비스 및 문서

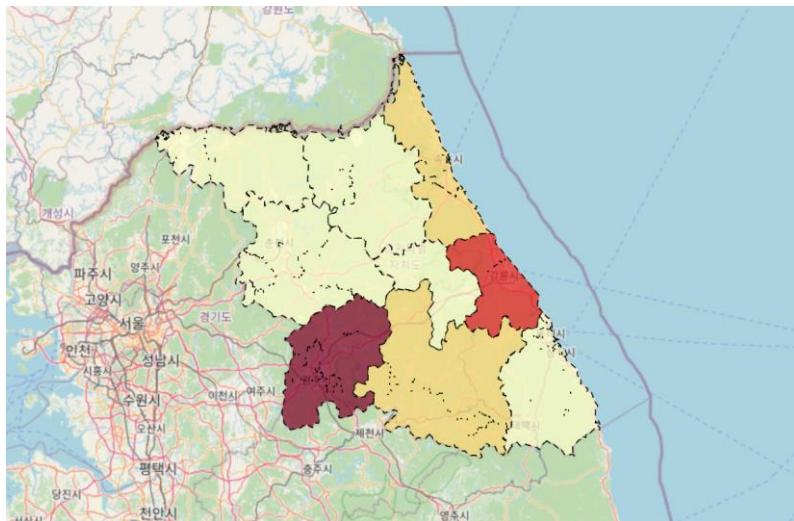
서비스 (Streamlit)

Streamlit 서비스 구현

강원도 산불 예측 및 피해 최소화 프로젝트

🔥🌳 산불위험지수(DWI) 산출 🌳🔥

강원도 9개 지역별 설정한 ML Model에 입력하여, 얻어진 확률들의 예측치를 이용하여 산불위험지수(DWI) 지도시각화



Point. 실시간 강원도 산불위험지수(DWI) 지도시각화

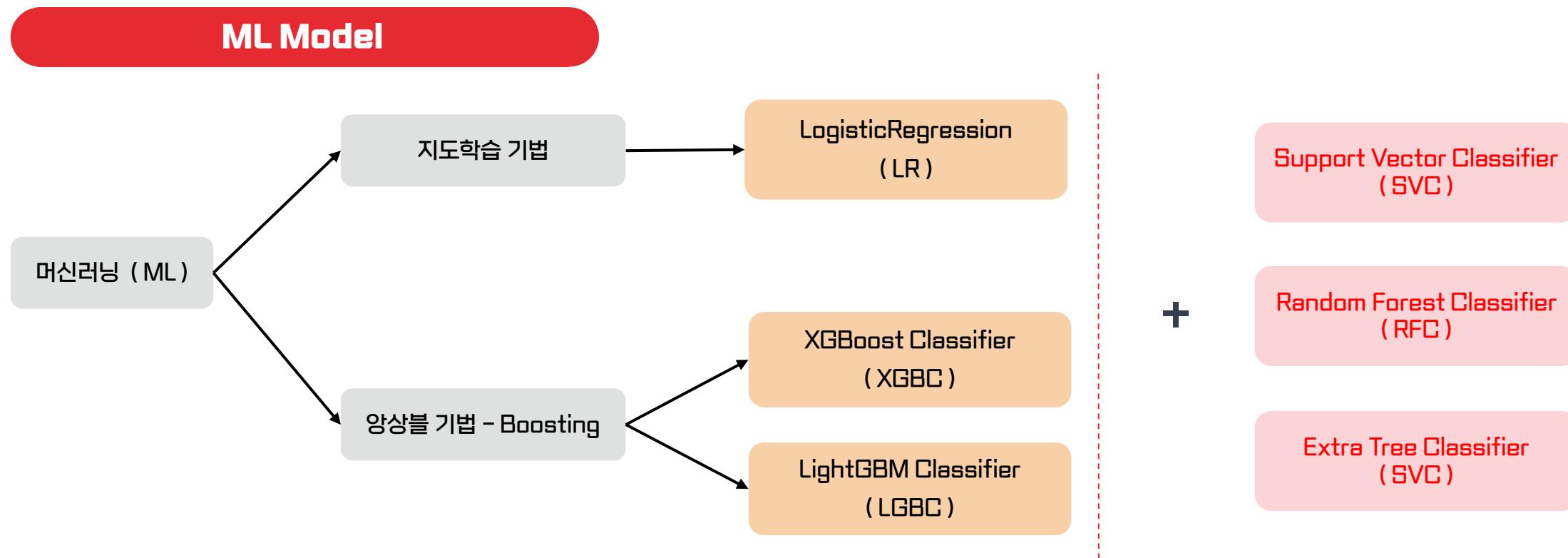
- ✓ 9 분할한 강원도 지역에 대해 각 지역별로 최적의 모델 적용
- ✓ 기상청 API 요청을 통해 실시간 강원도 지역 산불위험지수 산출
- ✓ 산불위험지수 구간에 따라 10분위로 지도시각화 표현



실시간 기상정보 처리 부분은 구현에 집중

기상청 API 서비스는 실시간으로 기상정보를 제공해주지 않고 전일 11시 기준 데이터를 제공
현 시점에서 여름 및 최근 강수여부 등의 요소로 산불위험지수가 보통 낮게 나옴
발표를 위해 인위적으로 날짜를 산불위험지수가 높게 나오는 봄철로 수정한 사진으로 첨부

한계점 / 개선사항

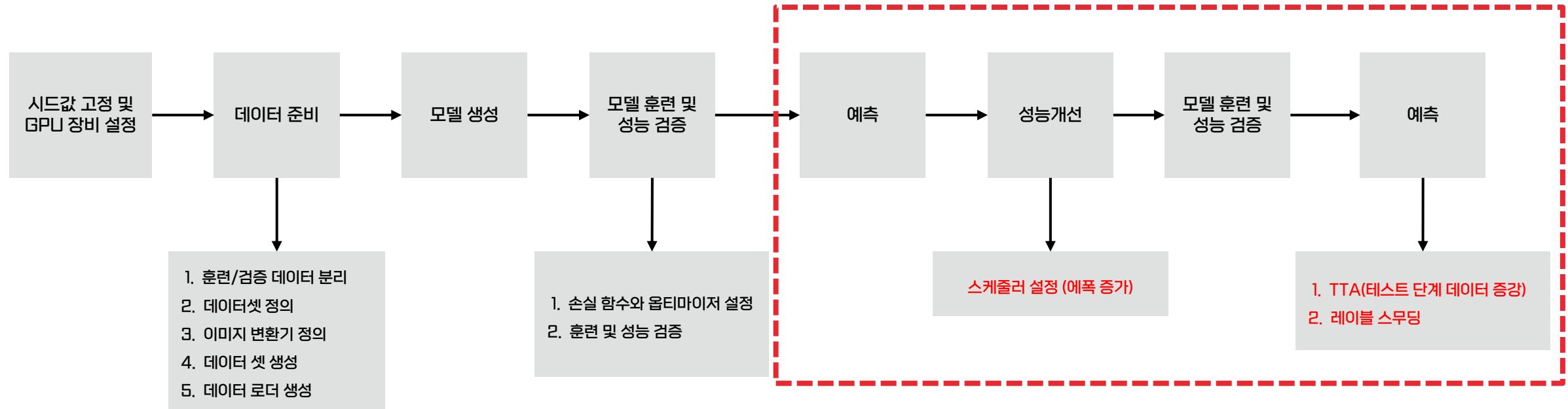


Point. 다양한 머신러닝 모델 적용하여 성능 개선 및 평가

- ✓ 제한적인 머신러닝 모델(LR, XGBC, LGBC)만을 사용하여 모델 평가 진행
- ✓ 다양한 머신러닝 모델을 적용하여 성능 개선 및 평가 진행하여 작업해 볼 예정
ex) SVC (Support Vector Classifier), RFC (Random Forest Classifier), ETC (Extra Tree Classifier) 등

한계점 / 개선사항

DL Model



Point. 성능 개선 및 예측 단계 적용

- ✓ EfficientNet Model 생성시 epoch 조절, 스케줄러 추가, TTA 기법, 레이블 스무딩 적용 등이 있으나,
Colab 개발 환경에서 모델 생성시 GPU 사용량에 대한 한계로 인해 잘 적용되지 않음
- ✓ GPU 사용량이 증가된 개발 환경에서 추후 작업해 볼 예정

프로젝트 문서

■ 요구사항 정의서

서비스 구분		구현 방안		확인		
요구 분류	요구 사항	요구 내용	비고	수용 여부	우선 순위	적용 일자
신규	데이터 수집	API 공공 데이터 수집		수용	상	23.05.22
신규	대형 산불 예측	기상 데이터를 수집하여 대형 산불 예측		수용	중	23.05.22
신규	웹 페이지	Streamlit 앱 구현		수용	중	23.05.22
신규	DB 연동	Google BigQuery를 이용한 DB서버 구축		수용	상	23.05.24
신규	데이터 조회	기상 및 지역 데이터 조회 가능		수용	상	23.05.26
수정	데이터 수정	결측치 데이터 제거		수용	중	23.05.26
수정	데이터 수정	데이터 전처리를 통한 PreProcessing_Data 생성		수용	상	23.05.30
수정	데이터 수정	Analysis_Data 생성		수용	상	23.06.01
신규	통계 분석	로지스틱 회귀분석을 통한 강원도 지역별 요인 파악		수용	상	23.06.01
신규	모델링	머신러닝을 통한 산불발생예측 모델 생성 및 성능평가		수용	상	23.06.07
신규	모델링	딥러닝을 통한 산불분류 모델 생성 및 성능평가		수용	상	23.06.07
신규	모델링	산불위험지수를 이용한 산불위험 등급화		수용	상	23.06.12

프로젝트 문서

트러블 슈팅 (Troubleshooting)

2023.5.25 [기상청_지상(종관, ASOS) 시간자료 조회서비스] 데이터 수집 과정 중 SERVICE ERROR 메시지 에 대해서 문의

문의

```
예외 발생: Expecting value: line 1 column 1 (char 0)
params : {'serviceKey': 'redacted', 'pageNo': '1', 'numOfRows': '999', 'dataType': 'json', 'dataCd': 'ASOS', 'dateCd': 'HR', 's
tartDt': '20130101', 'startHh': '01', 'endDt': '20230524', 'endHh': '23', 'stnIds': '98'} response : <OpenAPI_ServiceResponse>
<cmmMsgHeader>
<errMsg>SERVICE ERROR</errMsg>
<returnAuthMsg>HTTP ROUTING ERROR</returnAuthMsg>
<returnReasonCode>04</returnReasonCode>
</cmmMsgHeader>
</OpenAPI_ServiceResponse>
예외 발생: Expecting value: line 1 column 1 (char 0)
params : {'serviceKey': 'redacted', 'pageNo': '1', 'numOfRows': '999', 'dataType': 'json', 'dataCd': 'ASOS', 'dateCd': 'HR', 's
tartDt': '20130101', 'startHh': '01', 'endDt': '20230524', 'endHh': '23', 'stnIds': '93'} response : <OpenAPI_ServiceResponse>
<cmmMsgHeader>
<errMsg>SERVICE ERROR</errMsg>
<returnAuthMsg>HTTP ROUTING ERROR</returnAuthMsg>
<returnReasonCode>04</returnReasonCode>
</cmmMsgHeader>
</OpenAPI_ServiceResponse>
```

오류신고 및 담당자 문의



이름 김범모

이메일 goodbox92@naver.com

오류 데이터명 기상청_지상(종관, ASOS) 시간자료 조회서비스

* 내용

전국 산불 예측 모형 공부 중에 있습니다.
 '기상청_지상(종관, ASOS) 시간자료 조회서비스' 신청 후
 기간 : 20130101 - 전날
 대상 : 전 관측소
 에 대해서 API 데이터 요청에 있어
 <errMsg>SERVICE ERROR</errMsg>
 <returnAuthMsg>HTTP ROUTING ERROR</returnAuthMsg>

316 / 4000 bytes

오류내용 입력

답변

해당 부서 담당자와 답변

활용지원센터 답변

등록자 공공데이터활용지원센터

등록일 2023-05-26

** 안녕하세요 공공데이터활용지원센터입니다.

먼저 이용에 불편드린점 양해부탁드립니다.

HTTP ROUTING ERROR 오류는 공공데이터포털의 요청에 대하여 제공기관에서 정상적이지 않은 응답을 보낼경우 보여지는 오류입니다.

공공데이터포털에 제공중인 데이터의 DB 및 관리는 각 제공기관에서 관리하고 있기에,

추가문의는 오류신고 및 문의로 해주시기 바랍니다.

기상청_지상(종관, ASOS) 시간자료 조회서비스 오픈 API 상세

<https://www.data.go.kr/data/15057210/openapi.do>

처리일자

2023-06-07

답변내용

안녕하세요? 아래와 같이 문의하신 부분에 대해 답변을 드립니다. 기관명 : 기상청 국가기후데이터센터 담당자명 : 박주희 연구원 주가문의 연락처 : 042-481-7468 오류 및 문의 내용 : 전국 산불 예측 모형 공부 중에 있습니다. '기상청_지상(종관, ASOS) 시간자료 조회서비스' 신청 후 기간 : 20130101 - 전날 대상 : 전 관측소 에 대해서 API 데이터 요청에 있어 <errMsg>SERVICE ERROR</errMsg> <returnAuthMsg>HTTP ROUTING ERROR</returnAuthMsg> <returnReasonCode>04</returnReasonCode> 에러가 발생하고 있습니다. 답변: 안녕하세요? 해당 API서비스는 현재 정상 운영 중으로 확인됩니다. 다만, 문의하신 HTTP ERROR 의 경우 통신의 문제로 인해 발생될수 있습니다. 통신 문제가 없고 지속 오류 발생 시 상단에 연락처로 연락바랍니다. 감사합니다.

대안

[기상청_지상(종관, ASOS) 일자료 조회서비스] 데이터로 수정하는 방향으로 진행 완료

프로젝트 문서

트러블 슈팅 (Troubleshooting)

2023.6.1 [행정안전부_산불발생이력] 데이터 서비스 제공이 2021년으로 2022년 데이터 제공에 대해서 문의

문의

The screenshot shows the 'OpenAPI Data Detail View' for the 'Forest Fire Occurrence Information' dataset. The page has a sidebar on the left with links like 'OpenAPI 소개', 'OpenAPI Data', 'OpenAPI 인증키 발급', 'OpenAPI My Data', 'API 활용통계', 'OpenAPI 자료실', and 'OpenAPI Q&A'. The main content area displays the dataset details:

- Dataset Name:** 산불발생이력
- Formats:** XML, WMS
- Category:** 제난안전
- Provider:** 산림청
- Contact:** 관리부서명: 산불방지과, 관리부서 전화번호: 042-181-4258
- API Type:** REST
- Data Source:** 데이터 포맷: XML + WMS
- Usage Statistics:** 활용 신청: 28, 갱신주기: 1년
- Collection Period:** 수집일: 2021년, 수정일: 2023-02-17
- Key Features:** 주요항목: 지역/원인/시기별 산불 발생 및 진화 현황 통계
- API URLs:** XML 호출 URL: safemap.go.kr/open ApiService/data/getFrfireSttusData.do (아래 예제 참고), WMS 호출 URL: safemap.go.kr/open ApiService/wms/getLayerData.do (아래 예제 참고)

At the bottom, there are tabs: 그리드(XML), OPEN API(XML), OPEN API(WMS), MAP(WMS). The '그리드(XML)' tab is selected.

답변

해당 부서 담당자와 통화

- ✓ 산림청에서 행정안전부로 데이터를 제공하면 **행정안전부에서 API 서비스를 제공하는 방식**으로 업무 프로세서가 되어있음.
- ✓ 산림청에서는 해당 데이터를 **연간 한 번 (7 ~ 8월)** 공문형식으로 전달
- ✓ 행정안전부는 받은 데이터를 해당 **형식에 맞게 재정리** 이후 API 서비스로 제공
- ✓ **프로세스의 한계로 인해 간접주기**가 생길 수 밖에 없으며, 현행 방안으로는 실시간 데이터 수집 및 행정안전부에서 2022년 자료를 제공 받을 수 없음

대안

행정안전부에서 제공해준 프로세스에 따라 2022년 데이터에 대해 [산림청_산불발생통계] 정보를 **수기로 재편집하여 진행 완료**

개발 환경

개발 환경



사용 데이터

데이터 명	기간	활용 목적	출처
기상청_지상(종관,ASOS) 일자료 조회 서비스	2013.01 ~ 2022.12	EDA, 머신러닝 및 딥러닝 Feature 활용	기상청
기상청_산불발생통계	2022.01 ~ 2022.12	지도 시각화 및 EDA	산림청
기상청_관측지점정보	2023.05.24 기준	지도 시각화 및 지역 분할	기상청
행정안전부_산불발생이력	2013.01 ~ 2021.12	지도 시각화 및 지역 분할	공공데이터포털
행정구역_읍면동(법정동)	2023.05.24 기준	지도시각화	국가공간정보포털
행정구역시군구_경계	2023.05.24 기준	지도시각화 및 지역 분할	국가공간정보포털
행정표준코드(강원)	2023.05.24 기준	지역분할	행정표준코드관리시스템

참고 문헌

뉴스 기사

- 1) 이상학. (2019. 04. 05). [강원산불] '양간지풍' 불면 순식간에 초토화..대형산불 원인, 강원도민일보, <https://www.yna.co.kr/view/AKR20190405040800062>
- 2) 김민형. (2022. 03. 05). 울진·삼척 산불 확산 '초비상'…재난사태 선포 '역대 4번째', MBC 뉴스, https://imnews.imbc.com/replay/2022/nwtoday/article/6347197_35752.html
- 3) 김병규. (2022. 03. 06). [동해안 산불] 피해면적 1만4천222㏊…역대 2번째, 2000년 이후로는 최대, 연합뉴스, <https://www.yna.co.kr/view/AKR20220306050300530>
- 4) 이경원. (2023. 04. 12). [사실은] 동해안의 지독한 봄철 산불…“피해 규모 13배”, SBS 뉴스, https://news.sbs.co.kr/news/endPage.do?news_id=N1007151270&plink=COPYPASTE&cooper=SBSNEWSEND
- 5) 이재윤. (2023. 04. 11). 봄철 동해안 태풍급 강풍 '양간지풍', 연합뉴스, <https://www.yna.co.kr/view/GYH20230411001400044>
- 6) Barış Özmen. (2019. 03. 28). AutoML for Data Augmentation, <https://blog.insightdatascience.com/automl-for-data-augmentation-e87cf692c366>
- 7) 강원지방기상청. 예보업무, <https://www.kma.go.kr/gangwon/html/info/business02.jsp>
- 8) 기상자료개방포털. 실효습도, <https://data.kma.go.kr/climate/ehum/selectEhumChart.do?pgmNo=110>

논문

- [1] Won, M., Jang, K., & Yoon, S. (2018). Development of the National Integrated Daily Weather Index (DWI) Model to Calculate Forest Fire Danger Rating in the Spring and Fall. *Korean Journal of Agricultural and Forest Meteorology*, 20(4), 348–356.
- [2] Chae, K., Park, J., YuRi-Lee, & Cho, Y. S. (2018). Development of a Gangwon Province Forest Fire Prediction Model using Machine Learning and Sampling. *한국빅데이터학회지*, 3(2), 71–78.
- [3] Kim, S., & Ahn, J. (2020). 기상 데이터를 이용한 데이터 마이닝 기반의 산불 예측 모델. *The Korea Academia-Industrial Cooperation Society*. Vol. 21, No. 8 pp. 521-529.
- [4] Lee Chanjung, Lim Mooyoung, Lee Yohan, (2023). Machine Learning for Big Data Analytics in Development of Wildfire Prediction Models, *J. Korean Soc. Hazard Mitig.*
- [5] Jinmu, C., & Yurp, P. S. (2020). 우리나라 계절적 강수와 건조도가 지역적 산불 발생에 미치는 영향. *한국지역지리학회지 제26권 제4호(2020)* 307–319.
- [6] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning* (pp. 6105–6114).

부록

로지스틱 회귀분석 통계량 - 지역 1(강원북부내륙)

지역 1 Forward-Wald 단계선택을 통해 6개의 변수 선택

케이스 처리 요약

가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스 분석에 포함	3652	96.0
결측 케이스	153	4.0
전체	3805	100.0
비선택 케이스	0	.0
전체	3805	100.0

범주형 변수 코딩		
	모수 코딩	빈도
Rntf	0	2645 .000
	1	1007 1.000

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률	단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	6.981	8	.539	1	506.792 ^a	.033	.207
2	8.678	8	.370	2	481.957 ^b	.039	.249
3	2.582	8	.958	3	465.854 ^b	.044	.275
4	4.085	8	.849	4	454.259 ^b	.047	.295
5	2.674	8	.953	5	450.287 ^b	.048	.301
6	2.776	8	.948	6	441.910 ^c	.050	.315

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

카이제곱	자유도	유의확률
1 단계 단계 121.560	1	<.001
블록 121.560	1	<.001
모형 121.560	1	<.001
2 단계 단계 24.835	1	<.001
블록 146.395	2	<.001
모형 146.395	2	<.001
3 단계 단계 16.103	1	<.001
블록 162.499	3	<.001
모형 162.499	3	<.001
4 단계 단계 11.595	1	<.001
블록 174.093	4	<.001
모형 174.093	4	<.001
5 단계 단계 3.972	1	.046
블록 178.066	5	<.001
모형 178.066	5	<.001
6 단계 단계 8.377	1	.004
블록 186.442	6	<.001
모형 186.442	6	<.001

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn
1 단계 상수항	1.000	-.046	.341	-.173	.048	-.088	-.032	-.096	-.039	-.041	-.393	.178	.041	-.388	-.044

	B	S.E.	Wald	자유도	유의확률	Exp(B)	EXP(B)에 대한 99% 신뢰구간 하한	상한
1 단계 ^a effRhm	-.145	.015	92.891	1	<.001	.865	.832	.899
상수항	3.479	.698	24.833	1	<.001	32.435		
2 단계 ^b minRhm	-.092	.021	20.228	1	<.001	.912	.865	.961
effRhm	-.072	.021	12.367	1	<.001	.930	.883	.981
상수항	2.177	.762	8.167	1	.004	8.822		
3 단계 ^c maxTa	.057	.015	14.311	1	<.001	1.059	1.018	1.101
minRhm	-.085	.020	18.168	1	<.001	.918	.872	.967
effRhm	-.082	.021	15.160	1	<.001	.922	.873	.973
상수항	1.529	.786	3.779	1	.052	4.613		
4 단계 ^d avgTa	-.332	.100	11.050	1	<.001	.718	.555	.928
maxTa	.373	.098	14.617	1	<.001	1.452	1.129	1.867
minRhm	-.053	.022	5.673	1	.017	.948	.895	1.004
effRhm	-.093	.021	19.122	1	<.001	.911	.862	.962
상수항	-.979	1.098	.794	1	.373	.376		
5 단계 ^e avgTa	-.404	.105	14.714	1	<.001	.667	.509	.876
maxTa	.440	.102	18.590	1	<.001	1.553	1.194	2.020
maxInsWs	.120	.059	4.212	1	.040	1.128	.970	1.311
minRhm	-.049	.022	4.718	1	.030	.953	.899	1.009
effRhm	-.087	.022	16.156	1	<.001	.917	.867	.969
상수항	-.2819	1.415	3.966	1	.046	.060		
6 단계 ^f avgTa	-.359	.105	11.664	1	<.001	.699	.533	.916
maxTa	.404	.101	15.877	1	<.001	1.498	1.154	1.946
maxInsWs	.131	.060	4.882	1	.027	1.141	.978	1.329
minRhm	-.058	.023	6.413	1	.011	.943	.889	1.001
effRhm	-.068	.022	9.268	1	.002	.934	.882	.990
sumRn7	-.045	.019	5.299	1	.021	.956	.909	1.005
상수항	-.3167	1.424	4.949	1	.026	.042		

부록

로지스틱 회귀분석 통계량 - 지역 2 (강원북부산지)

지역 2 Forward-Wald 단계선택을 통해 5개의 변수 선택

케이스 처리 요약		
가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스 분석에 포함	3652	96.0
결측 케이스	153	4.0
전체	3805	100.0
비선택 케이스	0	.0
전체	3805	100.0

범주형 변수 코딩		
	모수 코딩	
	빈도	(1)
Rntf	0	2557 .000
	1	1095 1.000

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률
1	6.430	8	.599
2	2.560	8	.959
3	6.180	8	.627
4	4.159	8	.843
5	5.557	8	.697

단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	592.042 ^a	.021	.124
2	570.480 ^a	.026	.158
3	561.663 ^b	.029	.172
4	552.513 ^a	.031	.187
5	547.542 ^a	.033	.195

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

	카이제곱	자유도	유의확률
1 단계	76.498	1	<.001
단계			
블록	76.498	1	<.001
모형	76.498	1	<.001
2 단계	21.563	1	<.001
단계			
블록	98.061	2	<.001
모형	98.061	2	<.001
3 단계	8.816	1	.003
단계			
블록	106.877	3	<.001
모형	106.877	3	<.001
4 단계	9.150	1	.002
단계			
블록	116.027	4	<.001
모형	116.027	4	<.001
5 단계	4.971	1	.026
단계			
블록	120.999	5	<.001
모형	120.999	5	<.001

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
1 단계	상수항	1.000	.015	.270	-.218	-.041	-.039	-.049	-.150	-.005	-.092	-.374	.211	.087	-.519	-.079

부록

로지스틱 회귀분석 통계량 - 지역 3 (강원북부해안)

지역 3 Forward-Wald 단계선택을 통해 3개의 변수 선택

케이스 처리 요약		
가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스	분석에 포함	3652 96.0
결측 케이스	153 4.0	
전체	3805 100.0	
비선택 케이스	0 .0	
전체	3805 100.0	

범주형 변수 코딩		
	모수 코딩	
	빈도	(1)
Rntf	0 2609 .000	
	1 1043 1.000	

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정		
	카이제곱	자유도
1 단계	단계 72.009	1 <.001
	블록 72.009	1 <.001
	모형 72.009	1 <.001
2 단계	단계 9.212	1 .002
	블록 81.221	2 <.001
	모형 81.221	2 <.001
3 단계	단계 8.690	1 .003
	블록 89.912	3 <.001
	모형 89.912	3 <.001

방정식의 변수						
	B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계 ^a	avgRhm	-.055	.008	53.940	1 <.001	.946
	상수항	-1.173	.341	11.845	1 <.001	.309
2 단계 ^b	maxTa	.042	.014	9.459	1 .002	1.043
	avgRhm	-.065	.008	62.665	1 <.001	.937
	상수항	-1.363	.358	14.507	1 <.001	.256
3 단계 ^c	maxTa	.059	.015	15.803	1 <.001	1.061
	avgRhm	-.034	.014	5.985	1 .014	.967
	effRhm	-.058	.020	8.275	1 .004	.943
	상수항	-.651	.436	2.232	1 .135	.522

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률	단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	5.133	8	.743	1	556.344 ^a	.020	.124
2	11.416	8	.179	2	547.131 ^a	.022	.139
3	3.658	8	.887	3	538.441 ^a	.024	.154

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn
1 단계	상수항 1.000	.165	.175	-.325	.005	-.229	.077	-.038	-.087	-.123	-.250	.148	.125	-.511	-.033

부록

로지스틱 회귀분석 통계량 - 지역 4 (강원중부내륙)

지역 4 Forward-Wald 단계선택을 통해 8개의 변수 선택

케이스 처리 요약

가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스	분석에 포함	3652 96.0
결측 케이스	153 4.0	
전체	3805 100.0	
비선택 케이스	0 .0	
전체	3805 100.0	

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률
1	2.803	8 .946	
2	4.041	8 .853	
3	3.459	8 .902	
4	5.682	8 .683	
5	7.006	8 .536	
6	2.166	8 .976	
7	5.558	8 .697	
8	3.973	8 .860	

범주형 변수 코딩

Rntf	빈도		(1)
	0	2440 .000	
1	1212	1.000	

모형 요약

단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	1188.277 ^a	.059	.184
2	1168.611 ^a	.064	.200
3	1132.983 ^b	.073	.229
4	1109.228 ^b	.079	.247
5	1101.645 ^b	.081	.253
6	1093.291 ^c	.083	.260
7	1087.226 ^c	.085	.265
8	1083.422 ^c	.086	.268

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

카이제곱	자유도	유의확률
1 단계		
블록	222.557	<.001
모형	222.557	<.001
2 단계		
단계	19.666	<.001
블록	242.224	<.001
모형	242.224	<.001
3 단계		
단계	35.628	<.001
블록	277.852	<.001
모형	277.852	<.001
4 단계		
단계	23.755	<.001
블록	301.607	<.001
모형	301.607	<.001
5 단계		
단계	7.582	.006
블록	309.189	<.001
모형	309.189	<.001
6 단계		
단계	8.354	.004
블록	317.544	<.001
모형	317.544	<.001
7 단계		
단계	6.065	.014
블록	323.609	<.001
모형	323.609	<.001
8 단계		
단계	3.803	.051
블록	327.412	<.001
모형	327.412	<.001

B	S.E.	Wald	자유도	유의확률	방정식의 변수		EXP(B)에 대한 99% 신뢰구간 하한	상한
					Exp(B)			
minTa	.069	.035	3.932	1	.047	.934	.854	1.021
maxTa	.105	.031	11.637	1	<.001	1.110	1.026	1.202
maxInsWs	.251	.073	11.813	1	<.001	1.285	1.065	1.551
avgWs	-.919	.360	6.536	1	.011	.399	.158	1.007
minRhm	-.030	.012	6.323	1	.012	.971	.941	1.001
effRhm	-.049	.015	10.892	1	<.001	.952	.916	.989
sumRn7	-.079	.015	26.323	1	<.001	.924	.888	.962
maxwind7	.186	.095	3.873	1	.049	1.205	.944	1.537
상수항	-2.125	1.194	3.169	1	.075	.119		

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
1 단계	상수항	1.000	.089	.194	-.270	-.010	.003	.019	-.282	-.029	-.111	-.352	.221	.024	-.440	-.047

부록

로지스틱 회귀분석 통계량 - 지역 5 (강원중부산지)

지역 5 Forward-Wald 단계선택을 통해 3개의 변수 선택

블록 1: 방법 = 전진 단계선택 (Wald)

케이스 처리 요약		
가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스	분석에 포함	3652 96.0
결측 케이스	153 4.0	
전체	3805 100.0	
비선택 케이스	0 .0	
전체	3805 100.0	

범주형 변수 코딩		
	빈도	모수 코딩 (1)
Rntf	0 2414 .000	
	1 1238 1.000	

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률	모형 요약		
				단계	-2 로그 우도	Cox와 Snell의 R-제곱
1	4.487	8	.811	1	159.623 ^a	.004
2	6.334	8	.610	2	149.786 ^b	.006
3	5.475	8	.706	3	144.928 ^b	.008

모형 계수의 총괄 검정

카이제곱	자유도	유의확률	방정식의 변수		
			B	S.E.	Wald
1 단계	12.920	<.001	effRhm	.026	11.419
블록	12.920	<.001	상수항	.1302	.420
모형	12.920	<.001	avgWs	.230	6.989
			effRhm	.027	18.094
2 단계	9.837	.002	상수항	2.531	2.287
블록	22.758	<.001	1 단계 ^a	.087	.917
모형	22.758	<.001	상수항	.843	.517
			avgWs	.130	.430
3 단계	4.858	.028	2 단계 ^b	.068	.545
블록	27.616	<.001	effRhm	.117	.001
모형	27.616	<.001	상수항	2.531	.890

	B	S.E.	Wald	자유도	유의확률	Exp(B)	EXP(B)에 대한 99% 신뢰구간
						하한	상한
1 단계 ^a	effRhm	.026	11.419	1	<.001	.917	.858 .979
	상수항	.1302	.420	1	.517	.430	
2 단계 ^b	avgWs	.230	6.989	1	.008	.545	.301 .985
	effRhm	.027	18.094	1	<.001	.890	.829 .955
상수항	2.531	1.674	2.287	1	.130	12.570	
3 단계 ^b	maxTa	.031	4.432	1	.035	1.068	.985 1.158
	avgWs	.242	5.787	1	.016	.558	.299 1.042
	effRhm	.026	22.486	1	<.001	.883	.825 .945
상수항	1.917	1.683	1.297	1	.255	6.800	

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
1 단계	상수항	1.000	.057	.327	-.327	-.004	.016	-.088	-.161	-.053	-.161	-.335	.169	.012	-.428	.014

부록

로지스틱 회귀분석 통계량 - 지역 7 (강원남부내륙)

지역 7 Forward-Wald 단계선택을 통해 5개의 변수 선택

케이스 처리 요약		
가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스 분석에 포함	3652	96.0
결측 케이스	153	4.0
전체	3805	100.0
비선택 케이스	0	.0
전체	3805	100.0

범주형 변수 코딩			
		모수 코딩	빈도
		(1)	
Rntf	0	2598	.000
	1	1054	1.000

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률	모형 요약		Nagelkerke R-제곱
				단계	-2 로그 우도	
1	14.066	8	.080	1	818.568 ^a	.040
2	5.690	8	.682	2	806.124 ^a	.043
3	7.254	8	.509	3	780.467 ^b	.050
4	8.069	8	.427	4	771.688 ^b	.052
5	5.325	8	.722	5	765.975 ^b	.053

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

카이제곱	자유도	유의확률
1 단계	147.729	1 <.001
블록	147.729	1 <.001
모형	147.729	1 <.001
2 단계	12.444	1 <.001
블록	160.173	2 <.001
모형	160.173	2 <.001
3 단계	25.657	1 <.001
블록	185.830	3 <.001
모형	185.830	3 <.001
4 단계	8.779	1 .003
블록	194.609	4 <.001
모형	194.609	4 <.001
5 단계	5.713	1 .017
블록	200.322	5 <.001
모형	200.322	5 <.001

방정식의 변수

B	S.E.	Wald	자유도	유의확률	Exp(B)	EXP(B)에 대한 99% 신뢰구간 하한	상한
1 단계 ^a effRhm	-.133	.012	122.754	1 <.001	.875	.848	.903
상수항	2.776	.518	28.708	1 <.001	16.051		
2 단계 ^b maxTa	.035	.010	11.774	1 <.001	1.036	1.009	1.063
effRhm	-.139	.012	134.298	1 <.001	.870	.844	.897
상수항	2.414	.524	21.260	1 <.001	11.178		
3 단계 ^c maxTa	.041	.010	17.002	1 <.001	1.042	1.016	1.069
effRhm	-.115	.013	81.066	1 <.001	.891	.862	.921
sumRn7	-.070	.019	14.166	1 <.001	.933	.889	.978
상수항	1.623	.556	8.529	1 .003	5.068		
4 단계 ^d avgTa	-.215	.073	8.738	1 .003	.806	.669	.973
maxTa	.245	.070	12.298	1 <.001	1.278	1.067	1.531
effRhm	-.104	.013	63.748	1 <.001	.901	.871	.932
sumRn7	-.070	.019	14.193	1 <.001	.932	.889	.978
상수항	.005	.777	.000	1 .995	1.005		
5 단계 ^e avgTa	-.232	.073	10.278	1 .001	.793	.658	.955
maxTa	.261	.070	14.083	1 <.001	1.298	1.085	1.553
maxInsWs	.112	.046	6.037	1 .014	1.119	.995	1.259
effRhm	-.096	.013	50.481	1 <.001	.909	.878	.941
sumRn7	-.079	.020	16.092	1 <.001	.924	.879	.972
상수항	-.150	.902	1.624	1 .202	.317		

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
1 단계	상수항	1.000	-.039	.217	-.112	.011	-.181	.076	-.062	.122	-.167	-.300	.196	.059	-.310	-.032

부록

로지스틱 회귀분석 통계량 - 지역 B (강원남부산지)

지역 B Forward-Wald 단계선택을 통해 4개의 변수 선택

케이스 처리 요약

가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스	분석에 포함	3652 96.0
결측 케이스	153 4.0	
전체	3805 100.0	
비선택 케이스	0 .0	
전체	3805 100.0	

= Hosmer와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률
1	2.433	8 .965	
2	4.169	8 .842	
3	8.533	8 .383	
4	4.482	8 .811	

범주형 변수 코딩

Rntf	모수 코딩 (1)		
	빈도	0	2387 .000
	1	1265 1.000	

모형 요약

단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	708.210 ^a	.039	.185
2	676.103 ^a	.047	.226
3	666.396 ^a	.049	.238
4	658.453 ^a	.052	.248

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

카이제곱	자유도	유의확률
1 단계	143.479	1 <.001
블록	143.479	1 <.001
모형	143.479	1 <.001
2 단계	32.106	1 <.001
블록	175.585	2 <.001
모형	175.585	2 <.001
3 단계	9.708	1 .002
블록	185.293	3 <.001
모형	185.293	3 <.001
4 단계	7.943	1 .005
블록	193.236	4 <.001
모형	193.236	4 <.001

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
1 단계	상수항	1.000	.123	.167	-.293	.007	-.148	.053	-.068	-.077	-.034	-.389	.217	.058	-.437	-.080

부록

로지스틱 회귀분석 통계량 - 지역 9 (강원남부해안)

지역 9 Forward-Wald 단계선택을 통해 3개의 변수 선택

케이스 처리 요약

가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스	분석에 포함	3652 96.0
결측 케이스	153 4.0	
전체	3805 100.0	
비선택 케이스	0 .0	
전체	3805 100.0	

범주형 변수 코딩

Rntf	빈도		모수 코딩 (1)
	0	2332 .000	
1	1320 1.000		

= Hosmer와 Lemeshow 검정 =

단계	모형 요약			Cox와 Snell의 R-제곱	Nagelkerke R-제곱
	카이제곱	자유도	유의확률		
1	6.834	8	.555	522.037 ^a	.018 .119
2	6.641	8	.576	509.695 ^a	.021 .142
3	17.799	8	.023	503.099 ^a	.023 .153

블록 1: 방법 = 전진 단계선택 (Wald)

모형 계수의 총괄 검정

1 단계	단계	카이제곱	자유도	유의확률
	블록	65.306	1	<.001
	모형	65.306	1	<.001
2 단계	단계	12.342	1	<.001
	블록	77.648	2	<.001
	모형	77.648	2	<.001
3 단계	단계	6.596	1	.010
	블록	84.244	3	<.001
	모형	84.244	3	<.001

방정식의 변수

	B	S.E.	Wald	자유도	유의확률	Exp(B)	EXP(B)에 대한 99% 신뢰구간 하한	상한
1 단계 ^a	avgRhm	-.060	.008	51.501	1 <.001	.942	.922	.962
	상수항	-.839	.405	4.287	1 .038	.432		
2 단계 ^b	maxTa	.051	.015	12.298	1 <.001	1.053	1.014	1.093
	avgRhm	-.071	.009	65.457	1 <.001	.932	.911	.953
	상수항	-1.067	.418	6.522	1 .011	.344		
3 단계 ^c	maxTa	.060	.015	16.353	1 <.001	1.062	1.022	1.104
	avgRhm	-.038	.016	5.959	1 .015	.963	.925	1.002
	effRhm	-.059	.023	6.420	1 .011	.943	.889	1.001
	상수항	-.240	.536	.200	1 .655	.787		

상관행렬

	상수항	avgTa	minTa	maxTa	sumRn	maxInsWs	maxWs	avgWs	minRhm	avgRhm	effRhm	sumRn7	Rntf(1)	maxwind7	noRn	
1 단계	상수항	1.000	.156	.157	-.312	.019	-.202	.075	-.070	-.059	-.034	-.373	.232	.166	-.465	-.068

부록 - 데이터 정의서

데이터 셋	Raw_Data			프로젝트 명	Forest_Fire
테이블 명	weather_stations (기상청 관측지점 정보)			작성일	2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고	
1	stnId	지점 번호	INTEGER		
2	startDate	시작일	STRING		
3	endDate	종료일	STRING		
4	stnNm	지점명	STRING		
5	stnAddress	지점주소	STRING		
6	stnAdministrative	관리관서	STRING		
7	stnLatitude	위도	FLOAT	종관기상관측 지점	
8	stnLongitude	경도	FLOAT		
9	elevation	노장해발고도(m)	FLOAT		
10	barometer	기압계(관측장비지상높이(m))	FLOAT		
11	thermometer	기온계(관측장비지상높이(m))	FLOAT		
12	anemometer	풍속계(관측장비지상높이(m))	FLOAT		
13	raingauge	강우계(관측장비지상높이(m))	FLOAT		

부록 - 데이터 정의서

데이터셋 테이블명		Raw_Data wether_days (기상청 지상(종관,ASOS) 일 자료)		프로젝트명 작성일	Forest_Fire 2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고	
1	stnId	지점 번호	INTEGER	종관기상관측 지점 번호(첨부)	
2	stnNm	지점명	STRING	종관기상관측 지점명(첨부)	
3	tm	시간	STRING		
4	avgTa	평균 기온(°C)	FLOAT		
5	minTa	최저 기온(°C)	FLOAT		
6	minTaHrmt	최저 기온 시각(hhmi)	FLOAT		
7	maxTa	최고 기온(°C)	FLOAT		
8	maxTaHrmt	최고 기온 시각(hhmi)	FLOAT		
9	sumRnDur	강수 계속시간(hr)	FLOAT		
10	mi10MaxRn	10분 최대강수량(mm)	FLOAT		
11	mi10MaxRnHrmt	10분 최대강수량 시각(hhmi)	FLOAT		
12	hr1MaxRn	1시간 최대강수량(mm)	FLOAT		
13	hr1MaxRnHrmt	1시간 최대 강수량 시각(hhmi)	FLOAT		
14	sumRn	일강수량(mm)	FLOAT		

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
15	maxInsWs	최대 순간풍속(m/s)	FLOAT	
16	maxInsWsWd	최대 순간 풍속 풍향(16방위)	FLOAT	
17	maxInsWsHrmt	최대 순간풍속 시각(hhmm)	FLOAT	
18	maxWs	최대 풍속(m/s)	FLOAT	
19	maxWsWd	최대 풍속 풍향(16방위)	FLOAT	
20	maxWsHrmt	최대 풍속 시각(hhmm)	FLOAT	
21	avgWs	평균 풍속(m/s)	FLOAT	
22	hr24SumRws	풍정합(100m)	FLOAT	-
23	maxWd	최대 풍향(16방위)	FLOAT	
24	avgTd	평균 이슬점온도(°C)	FLOAT	
25	minRhm	최소 상대습도(%)	FLOAT	
26	minRhmHrmt	평균 상대습도 시각(hhmm)	FLOAT	
27	avgRhm	평균 상대습도(%)	FLOAT	
28	avgPv	평균 증기압(hPa)	FLOAT	
29	avgPa	평균 현지기압(hPa)	FLOAT	
30	maxPs	최고 해면 기압(hPa)	FLOAT	

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
31	maxPsHrmt	최고 해면기압 시각(hhmi)	FLOAT	
32	minPs	최저 해면기압(hPa)	FLOAT	
33	minPsHrmt	최저 해면기압 시각(hhmi)	FLOAT	
34	avgPs	평균 해면기압(hPa)	FLOAT	
35	ssDur	가조시간(hr)	FLOAT	
36	sumSsHr	합계 일조 시간(hr)	FLOAT	
37	hr1MaxIcsrHrmt	1시간 최다 일사 시각(hhmi)	FLOAT	
38	hr1MaxIcsr	1시간 최다 일사량(MJ/m ²)	FLOAT	
39	sumGsr	합계 일사량(MJ/m ²)	FLOAT	
40	ddMefs	일 최심신적설(cm)	FLOAT	
41	ddMefsHrmt	일 최심신적설 시각(hhmi)	FLOAT	
42	ddMes	일 최심적설(cm)	FLOAT	
43	ddMesHrmt	일 최심적설 시각(hhmi)	FLOAT	
44	sumDpthFhsc	합계 3시간 신적설(cm)	FLOAT	
45	avgTca	평균 전운량(10분위)	FLOAT	
46	avgLmac	평균 중하층운량(10분위)	FLOAT	

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
47	avgTs	평균 지면온도(°C)	FLOAT	
48	minTg	최저 초상온도(°C)	FLOAT	
49	avgCm5Te	평균 5cm 지중온도(°C)	FLOAT	
50	avgCm10Te	평균 10cm 지중온도(°C)	FLOAT	
51	avgCm20Te	평균 20cm지중온도(°C)	FLOAT	
52	avgCm30Te	평균 30cm 지중온도(°C)	FLOAT	
53	avgM05Te	0.5m 지중온도(°C)	FLOAT	
54	avgM10Te	1.0m 지중온도(°C)	FLOAT	
55	avgM15Te	1.5m 지중온도(°C)	FLOAT	
56	avgM30Te	3.0m 지중온도(°C)	FLOAT	
57	avgM50Te	5.0m 지중온도(°C)	FLOAT	
58	sumLrgEv	합계 대형증발량(mm)	FLOAT	
59	sumSmllEv	합계 소형증발량(mm)	FLOAT	
60	n99Rn	9-9강수(mm)	FLOAT	
61	iscs	일기현상	STRING	
62	sumFogDur	안개 계속 시간(hr)	FLOAT	

부록 - 데이터 정의서

데이터셋 테이블명	Raw_Data		프로젝트명	Forest_Fire
	forestfire_occurs (행정안전부 산불발생이력)		작성일	2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
1	OBJT_ID	일련번호	STRING	
2	OCCU_YEAR	발생년도	STRING	
3	OCCU_MT	발생월	STRING	
4	OCCU_DE	발생일	STRING	
5	OCCU_TM	발생시간	STRING	
6	OCCU_DAY	발생요일	STRING	
7	OCCU_DATE	발생년월일	STRING	-
8	END_YEAR	종료년	STRING	
9	END_MT	종료월	STRING	
10	END_DE	종료일	STRING	
11	END_TM	종료시간	STRING	
12	ADRES	지번주소	STRING	
13	RN_ADRES	도로명주소	STRING	
14	RESN	사유	STRING	

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
15	AR	피해면적	STRING	
16	AMOUNT	피해금액	STRING	
17	CTPRVN_CD	시도코드	STRING	
18	SGG_CD	시군구코드	STRING	-
19	EMD_CD	읍면동코드	STRING	
20	X	X좌표	STRING	
21	Y	Y좌표	STRING	

부록 - 데이터 정의서

데이터셋	Raw_Data			프로젝트명	Forest_Fire
테이블명	forestfire_occurs_add (산림청 산불발생현황)			작성일	2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고	
1	damagearea	피해면적합계(ha)	STRING		
2	endday	진화종료 일	STRING		
3	endmonth	진화종료 월	STRING		
4	endtime	진화종료 일시	STRING		
5	endyear	진화종료 연도	STRING		
6	firecause	발생원인	STRING		
7	locbunji	발생장소_지번	STRING		
8	locdong	발생장소_동리	STRING		
9	locgungu	발생장소_시군구	STRING		
10	locmenu	발생장소_읍면	STRING		
11	locsi	발생장소_시도	STRING		
12	startday	발생일	STRING		
13	startdayofweek	발생요일	STRING		
14	startmonth	발생월	STRING		
15	starttime	발생시간	STRING		
16	startyear	발생연도	STRING		

부록 - 데이터 정의서

데이터셋 테이블명	Raw_Data		프로젝트명	Forest_Fire
	gangwon_SGG (국토교통부 행정구역(시군구)_강원)		작성일	2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
1	ADM_SECT_C	행정구역코드	STRING	
2	SGG_NM	시군구명	STRING	
3	SGG_OID	원천오브젝트ID	FLOAT	
4	COL_ADM_SE	원천시군구코드	STRING	-
5	GID	OBJECTID	INTEGER	
6	geometry	도형정보	STRING	

부록 - 데이터 정의서

데이터셋 테이블명	Raw_Data	프로젝트명	Forest_Fire	
	gangwon_UMD(국토교통부 행정구역(읍면동)_강원)	작성일	2023.05.25	
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
1	EMD_CD	읍면동코드	STRING	
2	EMD_NM	읍면동명	STRING	
3	SGG_OID	원천오브젝트ID	INTEGER	-
4	COL_ADM_SE	원천시군구코드	STRING	
5	GID	OBJECTID	INTEGER	
6	geometry	도형정보	STRING	

데이터셋 테이블명	Raw_Data	프로젝트명	Forest_Fire	
	gangwon_code(국토교통부 법정동 코드_강원)	작성일	2023.05.25	
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
1	code	법정동코드	INTERGER	-
2	address	주소	STRING	

부록 - 데이터 정의서

데이터셋 테이블명	PreProcessing_Data weather_stations (관측지점정보(파생))	프로젝트명 작성일	Forest_Fire 2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입 비고
1	stnId	지점 번호	INTEGER
2	stnAddress	지점주소	STRING
3	stnAdministrative	관리관서	STRING
4	stnLatitude	위도	FLOAT
5	stnLongitude	경도	FLOAT

데이터셋 테이블명	PreProcessing_Data gangwon_regions (강원도 9개 지역 구분)	프로젝트명 작성일	Forest_Fire 2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입 비고
1	w_regions	구분 지역 이름	STRING
2	geometry	도형정보	STRING

부록 - 데이터 정의서

데이터셋 테이블명		PreProcessing_Data weather_days (일간 기상 자료(파생))		프로젝트명 작성일	Forest_Fire 2023.05.25
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고	
1	stnId	지점 번호	INTEGER	종관기상관측 지점	
2	tm	일시	DATE		
3	avgTa	평균 기온(°C)	FLOAT		
4	minTa	최저 기온(°C)	FLOAT		
5	maxTa	최고 기온(°C)	FLOAT		
6	sumRn	일강수량(mm)	FLOAT		
7	maxInsWs	최대 순간풍속(m/s)	FLOAT		
8	maxWs	최대 풍속(m/s)	FLOAT		
9	avgWs	평균 풍속(m/s)	FLOAT		
10	minRhm	최소 상대습도(%)	FLOAT		
11	avgRhm	평균 상대습도(%)	FLOAT		
12	stnAddress	지점주소	STRING		
13	stnAdministrative	관리관서	STRING	종관기상관측 지점	
14	stnLatitude	위도	FLOAT		

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
15	stnLongitude	경도	FLOAT	종관기상관측 지점
16	w_regions	강원도 기상 구분 지역	STRING	9개로 구분
17	h1	1일전 상대습도	FLOAT	
18	h2	2일전 상대습도	FLOAT	
19	h3	3일전 상대습도	FLOAT	
20	h4	4일전 상대습도	FLOAT	
21	effRhm	실효습도	FLOAT	
22	sumRn7	7일간 강수량	FLOAT	
23	Rntf	강수 여부	INTEGER	비가 옴 1 / 비가 오지 않음 0
24	maxwind7	7일간 최대 풍속	FLOAT	
25	noRn	비가 오지 않은 날 간격	INTEGER	

부록 - 데이터 정의서

데이터셋 테이블명	PreProcessing_Data forestfire_occurs (산불발생이력 (파생))	프로젝트명	Forest_Fire	
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
1	OBJT_ID	일련번호	STRING	
2	OCCU_DATE	발생년월일	DATE	
3	ADRES	지번주소	STRING	-
4	AR	피해면적	STRING	
5	AMOUNT	피해금액	STRING	
6	latitude	위도	FLOAT	주소 기반
7	longgitude	경도	FLOAT	주소 기반
8	w_regions	강원도 기상 구분 지역	STRING	9개로 구분

부록 - 데이터 정의서

데이터셋 테이블명	Analysis_Data			프로젝트명	Forest_Fire
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	작성일	비고
1	w_regions	강원도 기상 구분 지역	STRING		9개로 구분
2	tm	일시	DATE		
3	avgTa	평균 기온(°C)	FLOAT		
4	minTa	최저 기온(°C)	FLOAT		
5	maxTa	최고 기온(°C)	FLOAT		
6	sumRn	일강수량(mm)	FLOAT		
7	maxInsWs	최대 순간풍속(m/s)	FLOAT		
8	maxWs	최대 풍속(m/s)	FLOAT		-
9	avgWs	평균 풍속(m/s)	FLOAT		
10	minRhm	최소 상대습도(%)	FLOAT		
11	avgRhm	평균 상대습도(%)	FLOAT		
12	h1	1일전 상대습도	FLOAT		
13	h2	2일전 상대습도	FLOAT		
14	h3	3일전 상대습도	FLOAT		

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
15	h4	4일전 상대습도	FLOAT	
16	effRhm	실효습도	FLOAT	-
17	sumRn7	7일간 강수량	FLOAT	
18	Rntf	강수 여부	INTEGER	비가 옴 1 / 비가 오지 않음 0
19	maxwind7	7일간 최대 풍속	FLOAT	-
20	noRn	비가 오지 않은 날 간격	FLOAT	
21	fire_occur	산불 발생 여부	INTEGER	발생 1 / 미발생 0

부록 - 데이터 정의서

데이터 셋	Analysis_Data	프로젝트 명	Forest_Fire	
테이블 명	GangwonNorthInland (강원북부내륙) – 지역 1 GangwonNorthMount (강원북부산지) – 지역 2 GangwonNorthCoast (강원북부해안) – 지역 3 GangwonCentralInland (강원중부내륙) – 지역 4 GangwonCentralMount (강원중부산지) – 지역 5 GangwonCentralCoast (강원중부해안) – 지역 6 GangwonSouthInland (강원남부내륙) – 지역 7 GangwonSouthMount (강원남부산지) – 지역 8 GangwonSouthCoast (강원남부해안) – 지역 9	작성일	2023.05.25	
번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
1	w_regions	강원도 기상 구분 지역	STRING	9개로 구분
2	tm	일시	DATE	
3	avgTa	평균 기온(°C)	FLOAT	
4	minTa	최저 기온(°C)	FLOAT	
5	maxTa	최고 기온(°C)	FLOAT	
6	sumRn	일강수량(mm)	FLOAT	
7	maxInsWs	최대 순간풍속(m/s)	FLOAT	

부록 - 데이터 정의서

번호	열(Column) 이름(Eng.)	열(Column) 이름(Eng.)	데이터 타입	비고
8	maxWs	최대 풍속(m/s)	FLOAT	
9	avgWs	평균 풍속(m/s)	FLOAT	
10	minRhm	최소 상대습도(%)	FLOAT	-
11	avgRhm	평균 상대습도(%)	FLOAT	
12	effRhm	실효습도	FLOAT	
13	sumRn7	7일간 강수량	FLOAT	
14	Rntf	강수 여부	INTEGER	비가 옴 1 / 비가 오지 않음 0
15	maxwind7	7일간 최대 풍속	FLOAT	-
16	noRn	비가 오지 않은 날 간격	FLOAT	
17	fire_occur	산불 발생 여부	INTEGER	발생 1 / 미발생 0