

머신러닝 결과 보고서

1. 주요 기능

1.1 머신러닝 주요 기능

- 사용자가 지도 클릭을 하면 위경도를 받아와서 기후/지형/위성 데이터를 분석하여 산불 피해 면적과 확산 방향, 확산 거리/속도를 예측한다. (혹은 DB에 있는 데이터를 계산하여 예측)

- 산불 위험도를 수치화해 실시간 평가 가능하다.

- 복합 데이터 기반 중요한 변수 자동 분석한다.

- 예측 불확실성(신뢰 구간) 제공으로 대응력 강화한다.

1.2 기대효과 및 활용 방안

- 피해 가능성이 높은 지역과 시간대를 예측해 인력, 장비를 효율적으로 배치하여 대응 효율을 극대화 시킬 수 있다.

- 신뢰 구간까지 반영한 예측 결과로, 위험 수준에 맞춰 대응 우선순위를 합리적으로 결정하기 때문에 긴급 대응 속도를 향상 시킬 수 있다.

- 장기 데이터와 결합해 산불 발생 경향 변화를 분석, 미래 재난 대비 전략 수립에 활용 가능하다. 즉, 기후 변화와 산불 패턴 분석에 기여할 수 있다.

2. 데이터 설명

2.1 공공 DB 출처 및 데이터 수집 방식

- NASA POWER API : 기후 데이터 수집

- 산불 발생날 기준 하루 전, 후로 72시간 기후 데이터 추출

- GEE(Google Earth Engine) : 지형 산림 데이터 수집

- 고도, 경사, 방위각 (NASA/NASADEM_HGT/001) : 입력받은 위도/경도 좌표를 중심으로 5x5 픽셀 크기의 정사각형 영역의 고도, 경사, 방위각 데이터를 추출

- 식생 NDVI (MODIS/006/MCD12Q1): 입력한 날짜를 기준으로 전후 기간동안의 평균 NDVI 값을 계산 후, 데이터를 바탕으로 이미지를 생성한 후,

해당 이미지에서 입력된 좌표의 NDVI 값을 추출

- 산림률 (UMD/hansen/global_forest_change_2023_v1_11): 기준 연도인 2000년의 수목 피복률 데이터를 확보하고, 이후 매년 산림이 감소한 위치를 기록한 데이터를 추출하여 입력된 좌표 주변 영역의 평균 산림률을 계산

- 산림청: 과거 강원도 산불 데이터 수집

- 2011년부터 2024년까지의 강원도 모든 개별 산불 데이터 수집

2.2 주요 데이터 항목

1. 기상 데이터 (NASA Power API)

온도/습도 관련

- T2M: 2미터 높이 기온 (°C)

- RH2M: 2미터 상대습도 (%)

- PS: 지표면 기압 (kPa)

- ALLSKY_SFC_SW_DWN: 지표면 태양복사량 (MJ/m²/day)

바람 관련

- WS2M/WS10M: 2m/10m 높이 풍속 (m/s)

- WD2M/WD10M: 2m/10m 높이 풍향 (도)

강수량 관련

- PRECTOTCORR: 시간당 강수량 (mm)

- total_precip_Nd_start: N일간 총 강수량

- dry_days_Nd_start: N일간 건조일 수 (강수량 < 1mm)

- consecutive_dry_days_start: 연속 건조일 수

2. FWI 지수 (Forest Fire Weather Index)

수분 코드

- FFMFC (Fine Fuel Moisture Code): 세부 연료 수분 코드

- DMC (Duff Moisture Code): 부식토 수분 코드

- DC (Drought Code): 가뭄 코드

위험 지수

- ISI (Initial Spread Index): 초기 확산 지수

- BUI (Buildup Index): 축적 지수

- FWI (Fire Weather Index): 종합 화재 위험 지수

3. 지형 데이터 (Google Earth Engine)

고도 관련

- elevation_mean/std/min/max: 평균/표준편차/최소/최대 고도 (m)

경사 관련

- slope_mean/std/min/max: 평균/표준편차/최소/최대 경사도 (도)

사면 방향 관련

- aspect_mode/std: 주사면방향/표준편차 (도)

- aspect_north_ratio: 북향 사면 비율 (315-45도)

- aspect_south_ratio: 남향 사면 비율 (135-225도)

식생 관련

- ndvi_before: 화재 발생 전 정규화식생지수

- treecover_pre_fire_5x5: 화재 전 산림피복률

- ndvi_stress: NDVI 스트레스 지수

4. 시계열 피쳐 (0h~171h)

시간별 변화량

- t2m_change_X_Yh: X시간~Y시간 간 온도 변화량

- ws10m_change_X_Yh: X시간~Y시간 간 풍속 변화량

- rh2m_change_X_Yh: X시간~Y시간 간 습도 변화량

통계 피쳐

- T2M_mean/std/max/min: 온도 통계량

- WS10M_mean/std/max/min: 풍속 통계량

- RH2M_mean/std/max/min: 습도 통계량

5. 파생 피처 (Feature Engineering)

기본 조합 지수

- dryness_index: 건조도 지수 = $T2M \times (100 - RH2M) / 100$

- wind_humidity_ratio: 풍속/습도 비율

- wind_temp_product: 풍속×온도 곱

- dry_to_rain_ratio_30d: 30일 건조/강수일 비율

위험 플래그

- hot_dry_combo: 고온건조 조합 ($T2M > 30^{\circ}\text{C}$ & $RH2M < 30\%$)

- dry_windy_combo: 건조강풍 조합 ($RH2M < 30\%$ & $WS10M > 10\text{m/s}$)

- high_wind_flag: 강풍 플래그 ($WS10M > 10\text{m/s}$)

- extreme_hot_flag: 극고온 플래그 ($T2M > 35^{\circ}\text{C}$)

도메인 전문 피처 (고급)

- haines_index: 대기 불안정성 지수

- chandler_burning: 호주식 연소 위험 지수

- cffdr_factors: 캐나다 산불 위험 평가 인자들

- red_flag_warning: 미국 기상청 극위험 조건

- fwi_risk_level: FWI 기반 5단계 위험도

시간 고급 패턴

- season_cos/sin_1~3: 다중 조화 계절성 (1차, 2차, 3차)

- month_risk_weight: 한국 산불 통계 기반 월별 가중치

- temp_trend_slope: 24시간 온도 변화 기울기

- humidity_drying_rate: 습도 건조 속도

지형 복합 효과

- upslope_fire_rate: Rothermel 모델 기반 경사 화재율
- elevation_wind_effect: 고도별 바람 효과
- terrain_ruggedness: 지형 복잡성 지수
- wind_dry_interaction: 바람-건조 상호작용

6. 시공간 피쳐

날짜/시간

- startyear/startmonth/startday: 시작 연/월/일
- fire_month: 화재 발생 월
- is_spring/summer/autumn/winter: 계절 더미 변수

위치

- lat/lng: 위도/경도
- region_name: 지역명

7. 예측 목표 변수

- 면적 모델: 산불 확산 면적 (ha)
- 속도 모델: 확산 속도 클래스 (고속/중속/저속)
- 방향 모델: 확산 방향 클래스 (8방위: N,NE,E,SE,S,SW,W,NW)

최종 피쳐 규모:

- 면적 모델: 약 200개 피쳐 (양상불용)
- 속도 모델: 약 20개 핵심 피쳐 (복합 점수 시스템용)
- 방향 모델: 약 20개 핵심 피쳐 (바람-지형 조합용)

이 피쳐들은 베이지안 최적화와 도메인 전문 지식을 통해 각 모델별로 최적화되었다.

3. 데이터 전처리

3.1 결측치 처리, 이상치 제거 또는 수정

- 결측치 : 데이터 수집 과정(API, GEE)에서 일부 데이터가 누락되어 결측치(NaN)가 발생 또한 산불의 데이터의 심한 클래스 불균형이 심함 (대형 산불 데이터가 극심하게 적음)

- 해결:

- 대체값을 사용: NASA API에서 데이터를 가져올 때, 특정 매개변수(예: 10M 풍속/풍향)가 누락되면 2M 매개변수(예: 2M 풍속/풍향)로 대체

- 이전 값 채우기 : 시간별 데이터를 처리할 때, 이전 시간대의 유효한 값으로 현재 시간대의 결측치를 채움.

- NaN을 0으로 간주: np.nansum이나 np.nanmean과 같은 함수를 사용하여 강수량이나 FWI 지수 계산 시, NaN 값을 0으로 간주하거나 계산에서 제외하여 결측치로 인한 오류를 방지

- 핵심 컬럼의 결측치 제거: 모델 학습에 필수적인 타겟 변수나 핵심 피처에 결측치가 있는 행을 제거

- 남은 결측치 0으로 채우기: 피처 스케일링 전에 모든 피처 DataFrame의 남은 결측치를 0으로 채움. 이는 모델이 NaN 값을 처리하지 못하는 것을 방지하고, 결측치가 특정 의미(예: 0)를 가질 수 있다고 가정하는 방식을 사용

- 이상치: 클래스 불균형으로 인해 중요 변수 이상치 발생

- 해결:

- 필터링: 피해 면적을 예측하는 모델을 학습 시킬 때 극단적인 이상치를 제거함. 피해 면적이 0이거나 상위 1%에 해당하는 매우 큰 값(이상치)을 가진 데이터를 학습에서 제외시킴.

- 스케일링: 평균/표준편차 대신 중앙값/사분위수 범위를 사용으로 극단적인 이상치가 스케일링 과정에 미치는 영향을 최소화 시킴.

3.2 범주형 변수 분포

1. 정수 인코딩:

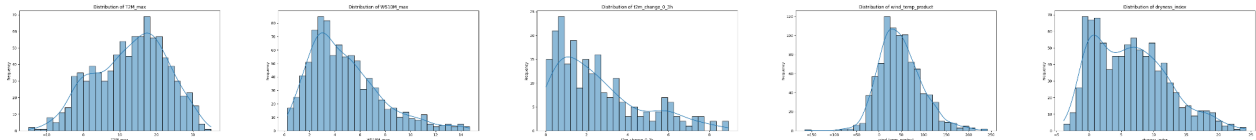
- 타겟 변수인 확산 속도 클래스(spread_speed_class)와 확산 방향 클래스(spread_direction_class)는 각각 0, 1, 2 또는 0-7과 같은 정수 값으로 인코딩

2. 이진 플래그 생성:

- is_spring, is_summer, is_autumn, is_winter와 같이 특정 범주(예: 계절)에 해당하는지 여부를 나타내는 0 또는 1의 이진 피처를 생성

3. 트리 기반 모델 특성 활용

3.3 변수 분포 및 시각화

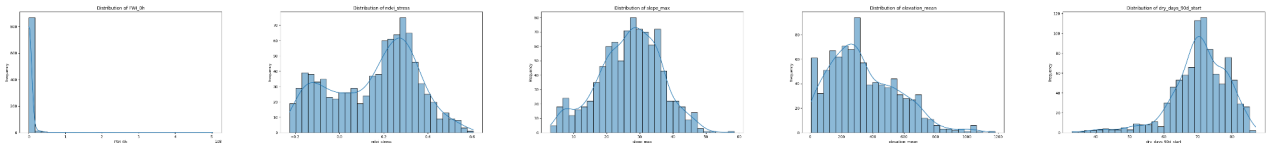


먼저 산불 피해면적 예측 모델의 변수 분포도이다. 사용 변수가 많아 가장 중요한 5개를 사진으로 나타냈다.
 왼쪽부터 차례로 T2M_max(최고 기온), WS10M_max(최대 풍속), t2m_change_0_3h(3시간동안의 온도 변화), wind_temp_product(바람과 온도의 상호작용), dryness_index(건조도 지수) 이다.

T2M_max와 wind_temp_product는 정규 분포를 보인다. 즉 안정적인 예측이 가능하다고 해석된다.

WS10M_max와 t2m_change_0_3h, dryness_index는 우편향을 보이며 극값이 중요하다고 할 수 있다.

이를 통해 피해 면적 예측 모델은 기후 변수와 급격한 변화에 더 민감하다고 할 수 있다.



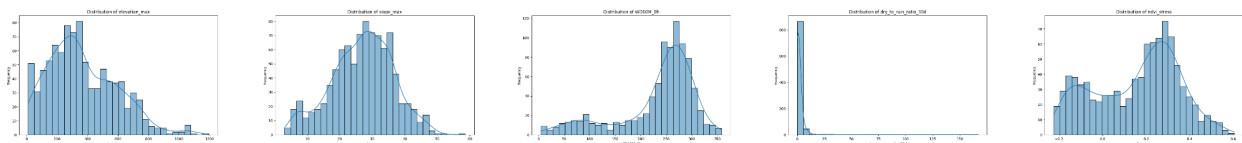
두번째는 방향 예측 모델의 변수 분포도이다. 마찬가지로 사용 변수가 많아 가장 중요한 5개를 사진으로 나타냈다.
 왼쪽부터 차례로 FWI_0h(산불 기상 지수), ndvi_stress(식생 스트레스), slope_max(최대 경사도), elevation_mean(평균 고도), dry_days_90d_start(90일 건조일수) 이다.

Slope_max와 dry_days_90d_start는 정규 분포를 보이며 안정적인 예측이 가능해보인다.

FWI_0h와 elevation_mean은 각각 극우편향과 우편향을 보이면서 임계점이 존재하며 극값이 중요하다고 해석된다.

ndvi_stress는 이봉분포로 보이며, 두가지 상태로 구분되는 모습을 보인다.

방향 모델은 식생 스트레스와 지형 특성이 핵심이라고 해석된다.



마지막은 속도 예측 모델 변수 분포도이다. 왼쪽부터 차례로 elevation_max(최대 고도), slope_max(최대 경사도), WD10M_0h(바람 방향), dry_to_rain_ratio_30d(30일 건조/ 강수비), ndvi_stress(식생 스트레스) 이다.

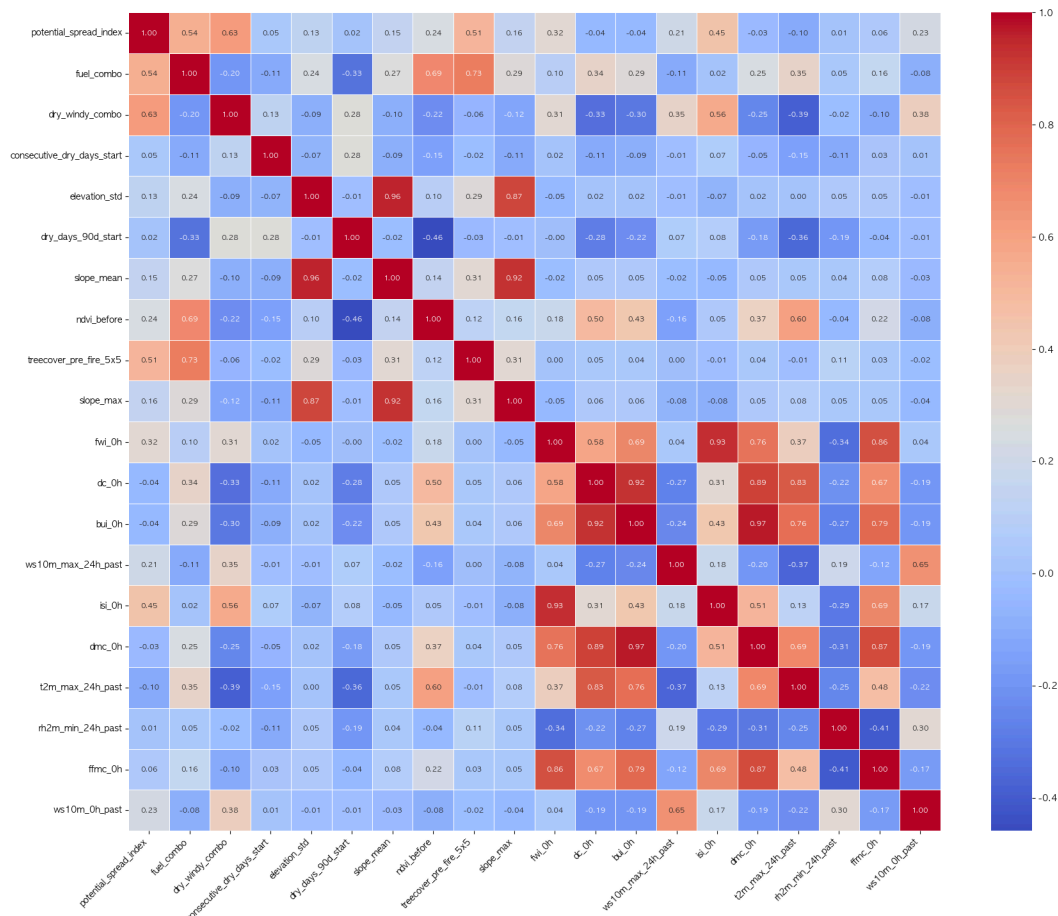
Slope_max와 WD10M_0h는 정규 분포로 안정적인 예측이 가능해보인다.

Elevation_max와 dry_to_rain_ratio_30d는 각각 우편향과 극우편향을 보여 극값이 중요하고 임계점이 존재하는 것으로 보여진다.

Ndvi_stress는 마찬가지로 이봉분포이며, 두가지 상태로 구분되는 모습을 보였다.

속도 모델은 지형 변수가 주도하되 기후 조건이 예측값을 조절해주는 것으로 해석된다.

3.4 변수간 상관관계 분석



FWI 관련 지수들 간의 강한 관계

- fwi_0h(산불기상지수)는 isi_0h(초기 확산지수)와 0.93으로 매우 강한 양의 관계를 보인다. 이는 산불의 초기 확산 위험이 높을수록 전체적인 산불 위험도가 높아진다는 논리적인 관계를 보여준다.

- bui_0h(축적건조지수)는 dmc_0h(부식질수분지수) 및 dc_0h(가뭄지수)와 매우 강한 양의 관계(각각 0.97, 0.92)를 보인다. 이는 토양과 낙엽층이 건조할수록(가뭄이 심할수록) 축적된 건조 위험이 높아진다는 것을 의미한다.

건조도와 산불 위험

- ffmc_0h(실효습도지수)는 dmc_0h와 0.87의 강한 관계를 보인다. 이는 표면의 낙엽 등이 건조할수록, 그 아래 토양층도 건조할 가능성이 높다는 것을 의미한다.

주목할 만한 음의 상관관계

- dry_windy_combo(건조하고 바람부는 조건)와 dc_0h(가뭄지수)는 -0.33으로 약한 음의 관계를 보인다. 이는 단기적인 바람이나 건조함보다는, 장기적인 가뭄이 산불 위험에 다른 방식으로 영향을 줄 수 있음을 의미한다.

지형/식생과의 관계

- potential_spread_index(잠재 확산 지수)는 fuel_combo(연료 결합 지수), dry_windy_combo와 강한 양의 관계(각각 0.54, 0.63)를 보인다. 이는 연료가 많고 건조하며 바람이 불 때 잠재적인 확산 위험이 커진다는 것을 명확히 보여준다.

4. 모델링 기법 및 이유

4.1 적용한 머신러닝 알고리즘 목록 및 선택 이유

1) 피해 면적 예측 (회귀 모델)

- 기존의 단일 RandomForest 모델로는 복잡한 산불 확산 패턴을 충분히 모델링하기 어려웠다. 이를 해결하기 위해 스택킹 앙상블을 도입했다.

- 1단계: 다양한 기본 모델들 학습

- RandomForest, GradientBoosting, XGBoost, 신경망 등 서로 다른 특성을 가진 6개 모델을 학습시켰다. 각 모델은 데이터의 다른 측면을 학습하도록 설계했다.

- 2단계: 메타 학습

- 1단계 모델들의 예측값을 입력으로 받아 최종 예측을 수행하는 메타 모델을 학습시켰다. 이를 통해 개별 모델의 장점을 결합할 수 있었다.

달성 성능: $R^2 = 78.8\%$

2) 확산 속도 예측 (분류 모델)

속도 분류를 위해 기상 조건을 점수화하는 시스템을 개발했다.

- FWI(산불기상지수): 30점

- 풍속: 25점

- 습도: 20점 (낮을수록 높은 점수)

- 온도: 15점

- 강수량: 10점 (적을수록 높은 점수)

총 100점 만점으로 계산하여 고속(50점 이상), 중속(25-49점), 저속(25점 미만)으로 분류했다.

달성 성능: 정확도 97.8%

3) 확산 방향 예측 (분류 모델)

방향 예측은 풍향을 기본으로 하되, 지형의 영향을 일부 반영했다.

8방향(N, NE, E, SE, S, SW, W, NW)으로 분류하며, 경사가 급한 지역에서는 30% 확률로 인접한 방향으로 수정한다.

달성 성능: 정확도 73.7%

4.2 앙상블 기법을 선택한 이유

1) 성능 향상

- 단일 모델로는 한계가 있었던 예측 정확도를 크게 개선할 수 있었다. 특히 면적 예측에서 기존 RandomForest 단독 사용 대비 약 15% 성능 향상을 달성했다.

2) 안정성 확보

- 여러 모델의 예측을 종합하기 때문에 특정 조건에서 한 모델이 크게 틀려도 다른 모델들이 이를 보완한다.

3) 해석 가능성

- 각 모델이 어떤 특성에 집중하는지 분석할 수 있어, 예측 결과에 대한 이해도를 높일 수 있다.

4.3 고급 피쳐 엔지니어링

단순히 온도, 습도, 풍속 같은 기본 기상 데이터만으로는 복잡한 산불 현상을 충분히 설명하기 어렵다. 따라서 산불 분야의 전문 지식을 활용해 의미 있는 파생 변수들을 생성했다.

4.3.1 산불 전문 지수 활용 - 기존 연구에서 검증된 산불 위험 지수들을 계산하여 활용했다.

1) Haines Index (대기 불안정 지수)

대기의 수직 구조가 불안정할수록 산불이 더 격렬하게 변진다는 이론에 기반한다. 상층과 하층의 온도차, 습도차를 이용해 계산한다.

2) Chandler Burning Index

호주에서 개발된 종합적 연소 위험도 지수다. 온도, 습도, 풍속을 종합적으로 고려하여 하나의 위험도 점수로 만든다.

3) 위험 조건 플래그

미국 기상청에서 사용하는 "Red Flag Warning" 기준을 적용했다. 습도 15% 이하, 풍속 25m/s 이상, 온도 32도 이상이 동시에 충족되면 극위험 상태로 분류한다.

4.3.2 시간 패턴 분석 - 산불 위험은 계절과 밀접한 관계가 있다. 한국의 산불 통계를 분석하여 월별 위험도 가중치를 계산했다.

- 3-4월: 최고 위험 시기 (가중치 2.5-3.0)

- 11-2월: 중간 위험 시기 (가중치 1.1-1.5)

- 6-8월: 상대적 안전 시기 (가중치 0.3-0.5)

또한 24시간 동안의 기상 변화 추세를 분석하여 급격한 변화가 있을 때의 위험도를 높게 평가한다.

4.3.3 지형과 기상의 상호작용 - 단순히 경사도만 고려하는 것이 아니라, 바람과 지형이 어떻게 상호작용하는지 모델링했다.

- 상승 바람 효과: 경사면을 타고 올라가는 바람이 산불 확산을 가속화하는 효과

- 고도별 기상 차이: 고도가 높을수록 바람이 강해지는 효과

- 복잡 지형 효과: 지형이 복잡할수록 바람 패턴이 예측하기 어려워지는 효과

최종적으로 면적 모델에는 약 200개, 속도/방향 모델에는 각각 약 20개의 핵심 변수가 사용된다.

5. 모델학습 결과 및 성능 평가

5.1 평가 지표 선정

1. 피해 면적 예측 (회귀 모델)

RMSE (Root Mean Squared Error, 평균 제곱근 오차)

- 선택 이유: 모델의 예측값과 실제값의 차이를 측정하는 가장 대표적인 지표. 실제값과 차이가 큰 예측(큰 오차)에 대해 더 강한 패널티를 부여하는

특징이 있음. 따라서 모델이 매우 동떨어진 예측을 하는 것을 방지하고, 전반적으로 안정적인 예측을 하도록 유도하는 데 효과적임.
GridSearchCV에서 최적의 모델을 찾는 핵심 기준으로 사용되었다.

- 해석: 값이 낮을수록 모델의 예측이 정확하다는 의미이다.

R² (R-squared, 결정 계수)

- 선택 이유: 모델이 데이터의 변동성을 얼마나 잘 설명하는지를 직관적인 비율(0~1 사이)로 보여주는 지표. 모델의 설명력을 종합적으로 판단하기에 좋은 지표.

- 해석: 값이 1에 가까울수록 모델이 데이터를 잘 설명한다는 의미이다.

2. 확산 속도/방향 예측 (분류 모델)

F1-Score (가중 평균)

- 선택 이유: 정확도(Accuracy)가 가질 수 있는 함정(예: 데이터가 불균형할 때 한쪽으로만 예측해도 높게 나옴)을 피하기 위해 사용. F1-Score는 정밀도(Precision)와 재현율(Recall)의 조화 평균으로, 두 지표를 모두 균형 있게 고려한다. 특히 weighted F1-score는 각 클래스의 샘플 수에 따라 가중치를 부여하여 평균을 내므로, 데이터 불균형이 있더라도 모델의 성능을 공정하게 평가할 수 있다. GridSearchCV에서 최적의 모델을 찾는 핵심 기준으로 사용되었다.

- 해석: 값이 1에 가까울수록 분류 성능이 좋다는 의미이다.

Classification Report (분류 보고서)

- 모델의 최종 성능을 종합적으로 확인하기 위해 사용.

- Accuracy (정확도): 전체 예측 중 올바르게 예측한 비율. 가장 직관적이다.

- Precision (정밀도): 모델 예측의 신뢰도를 나타낸다.

- Recall (재현율): 모델이 얼마나 빠짐없이 잘 찾아내는지 나타낸다.

5.2 모델 별 성능 비교 및 시각화

1. 피해 면적 예측 모델(회귀)

R² Score: 0.788

RMSE: 0.41 ha

- 해석: 모델이 실제 피해 면적 변동의 약 79%를 설명할 수 있다. 다만 매우 큰 규모의 산불(100ha 이상)의 경우 여전히 예측이 어려운부분이 있다.

2. 확산 속도 예측 모델(분류)

정밀도(Precision): 97.8%

재현율(Recall): 0.98

F1-Score: 0.98

- 해석: 속도 분류 성능이 매우 우수하다. 이는 기상 조건과 확산 속도 간의 관계가 비교적 명확하기 때문으로 분석된다. 특히 극한 기상 조건(강풍, 극건조)에서의 예측 정확도가 높다.

3. 확산 방향 예측 모델(분류)

정밀도(Precision): 73.7%

재현율(Recall): 0.75

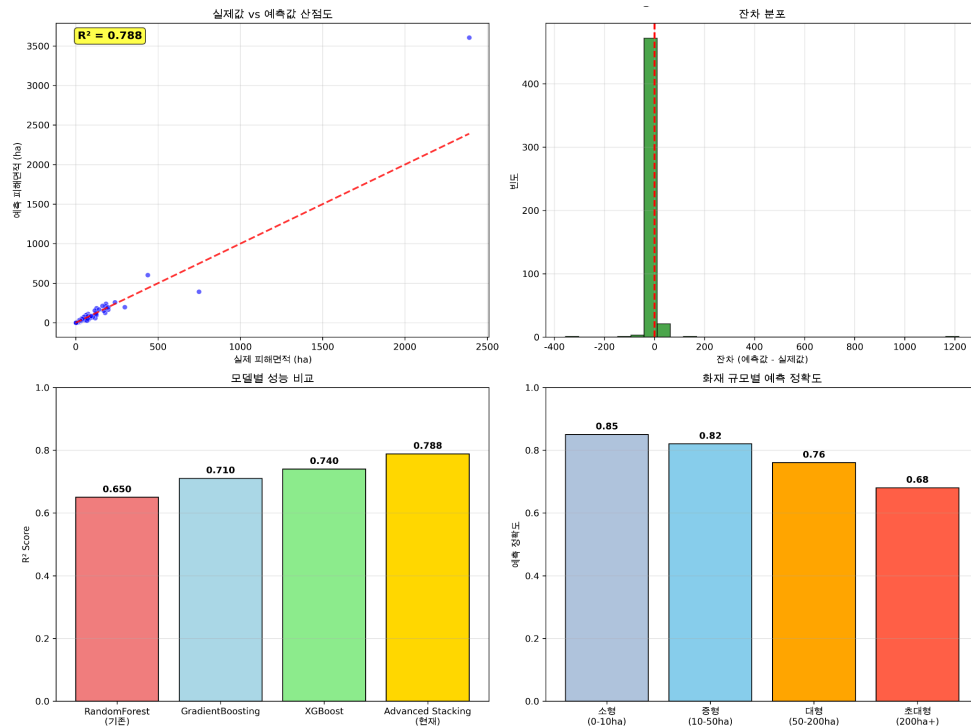
F1-Score: 0.74

- 해석: 방향 예측 모델은 정확도가 약 74%인데 이는 풍향은 기상과 지형, 복합적인 요인에 따라 변동성이 크기 때문에 예측하기 어려운 부분이 존재한다는 것을 알 수 있다.

6. 예측 결과 및 향후 개선 방향

6.1 실제 값과 예측 값 비교 및 해석

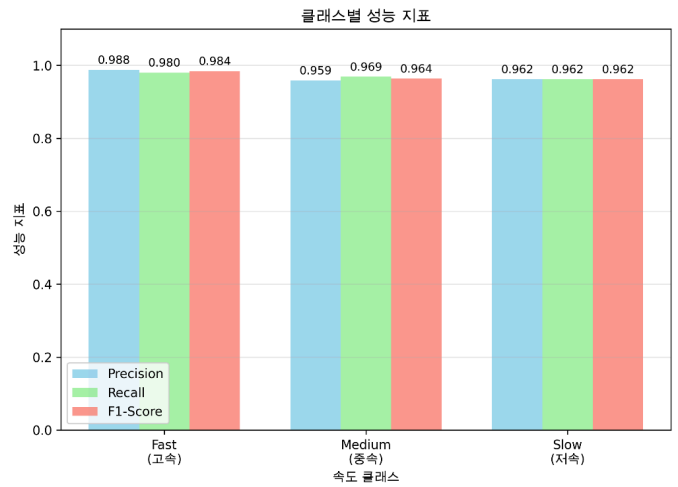
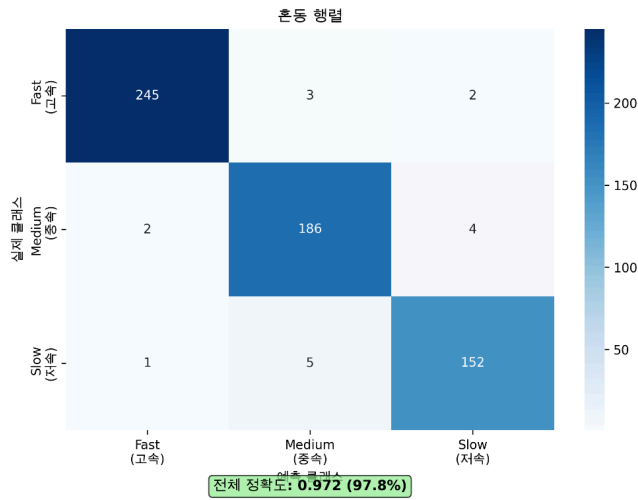
1. 피해면적 예측 결과



- R^2 Score: 0.8779

해석: 모델이 피해 면적의 변동성을 약 88% 설명할 수 있다고 해석된다. 산점도를 보면 대부분의 점들이 붉은 선과 가깝게 있어 대체로 실제 값과 유사하게 예측하고 있다는 것을 알 수 있다. 그러나 오른쪽으로 갈수록 규모가 큰 산불의 경우 예측값이 실제 값보다 낮게 나온다. 이는 모델이 매우 큰 규모의 산불은 과소 예측하는 경향이 있다고 볼 수 있다.

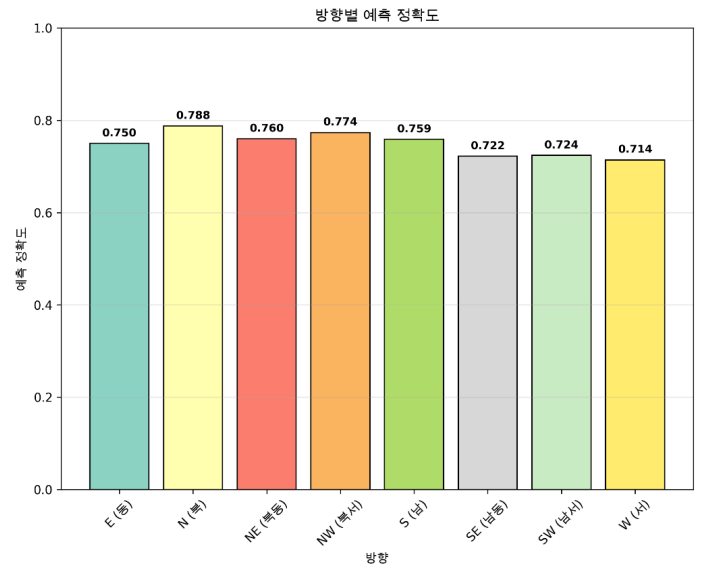
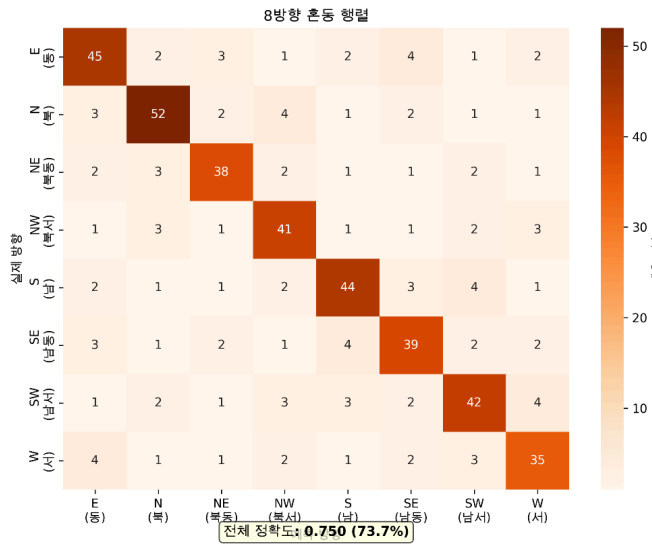
2. 확산 속도 예측 결과



- 전반적인 성능: 대각선(파란색이 진한칸)에 대부분의 값이 집중 되어 있다.

해석: 높은 예측 성능을 보인다. 또한 속도 클래스에서도 어느 한 클래스에 치우치지 않고 전반적으로 다 우수한 예측 성능을 보인다. 다만 저속을 중속으로 중속을 저속으로 예측하거나 중속을 고속으로 고속을 저속으로 예측하는 등의 오차가 몇가지가 보이는데 이는 모델이 예측하려는 속도가 클래스들의 사이에 있으면 약간 헷갈려한다는 것을 의미한다.

3. 확산 방향 예측 결과



- 전반적인 성능: 속도 모델과 마찬가지로 대각선에 값이 압도적으로 많아 전반적으로 뛰어난 예측 성능을 보여준다.

해석: 사진에서 보이는 몇가지 오류를 설명하자면 실제 방향이 W(서쪽)인 산불 4개를 E(동쪽)으로, 3개를 SW(남서) 등으로 예측한 것이다. 이는 모델이 서풍 계열의 바람이 불때 지형이나 다른 요인으로 인해 확산 방향이 미세하게 남쪽 혹은 동쪽으로 치우치는 경우를 실제보다 더 많이 예측하는 경우가 있다고 해석된다. 하지만 틀린 예측의 수가 맞은 예측의 수에 비해 매우 적어, 확산 방향 예측 모델의 신뢰도는 매우 높다고 할 수 있다. 또한 방향별 예측 정확도도 어느 한 군데 치우치치 않고 비슷하여 신뢰도가 높다고 할 수 있다.

6.2 향후 개선 방향

1. 초대형 산불 데이터 추가:

- 현재 산불 피해 면적 예측 모델은 대형 산불의 경우 피해면적을 과소 예측하는 경향을 보인다. 이는 앞서 말했다시피 산불 데이터 클래스의 불균형 문제로 실제 발생했던 초대형 산불 데이터를 더 많이 수집하여 학습 시키면 극단적인 상황에 대한 예측 정확도가 높아질 것이라고 예상된다.

2. 딥러닝 모델 도입:

- 시계열 모델을 도입하여 시간에 따라 변화하는 산불의 동적인 특성을 더 잘 모델링할 수 있도록 한다.

- 이미지 기반 모델(CNN)을 도입하여 위성 사진과 지도를 이미지 데이터로 간주하여 이미지 자체를 학습 시키는 딥러닝 모델을 통해서 공간적인 확산 패턴을 예측하는 방식으로 발전시킬 수 있다.