

Heat? Hit!

(히트-hit)

일일 열사병 환자 수 예측 프로그램

현대건설 기술교육원 SMART안전 1기 <3조> 김인영 이성준 정유진 최미선

I. 프로젝트 배경 및 목적

[프로젝트 배경]

- 최근 산업현장의 열사병 재해 발생량 증가
- 열사병으로 인한 근로자의 심각한 건강 문제 우려

[프로젝트 목적]

- 열사병을 사전에 예방할 수 있는 시스템 구축
- 기본적인 기후 데이터 기반의 열사병 환자 발생 가능성 예측 모델 개발

II. 개발 과정

① ————— ② ————— ③ ————— ④ ————— ⑤

상관관계
분석

학습모델
선정 및 훈련

교차검증

성능평가

예측결과
시각화

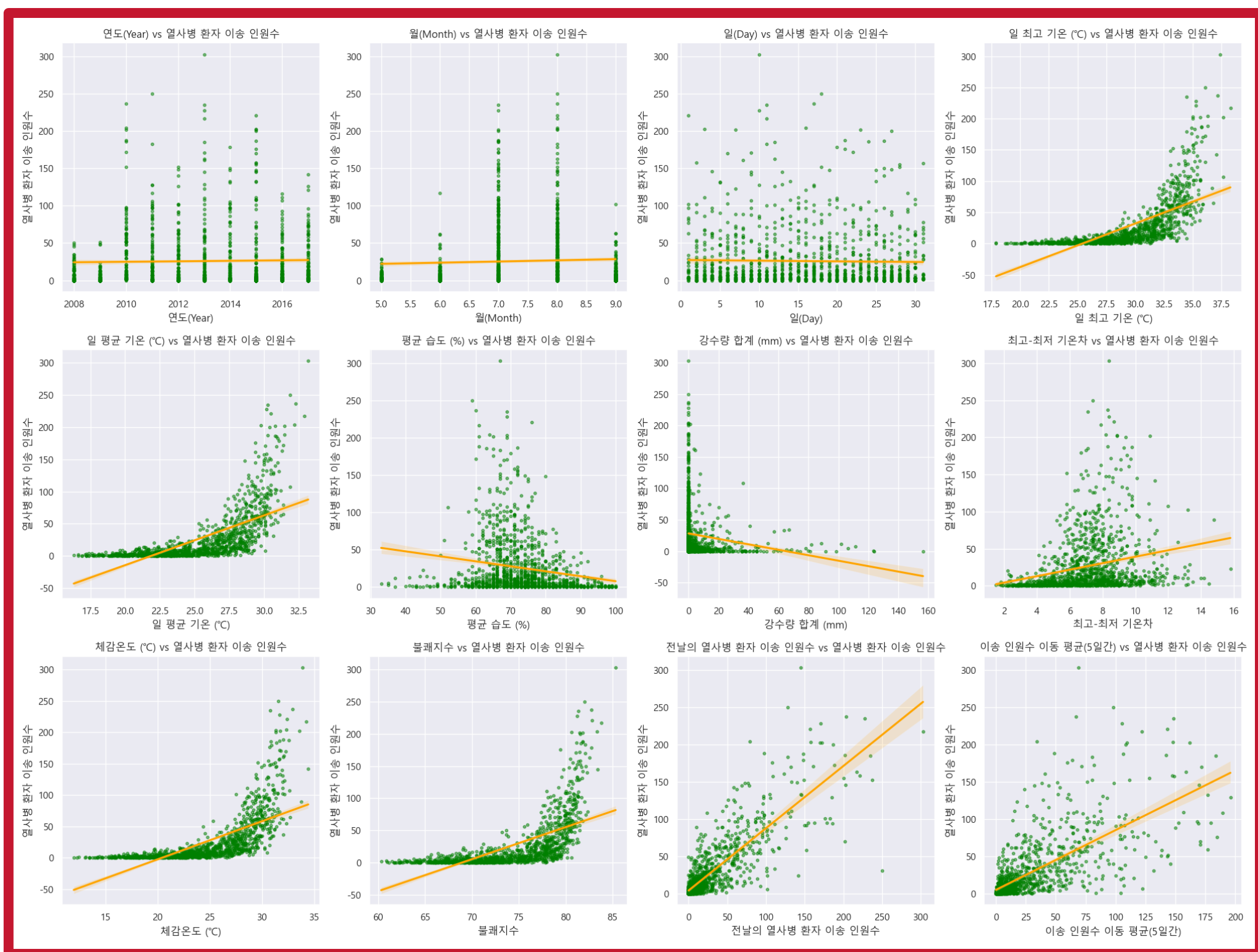
- 훈련을 위한 train데이터, 예측을 위한 test데이터 로 구분
- 데이터 구성: 날씨 데이터, 날씨 관련 데이터 등 총 41가지

III. 상관관계 분석

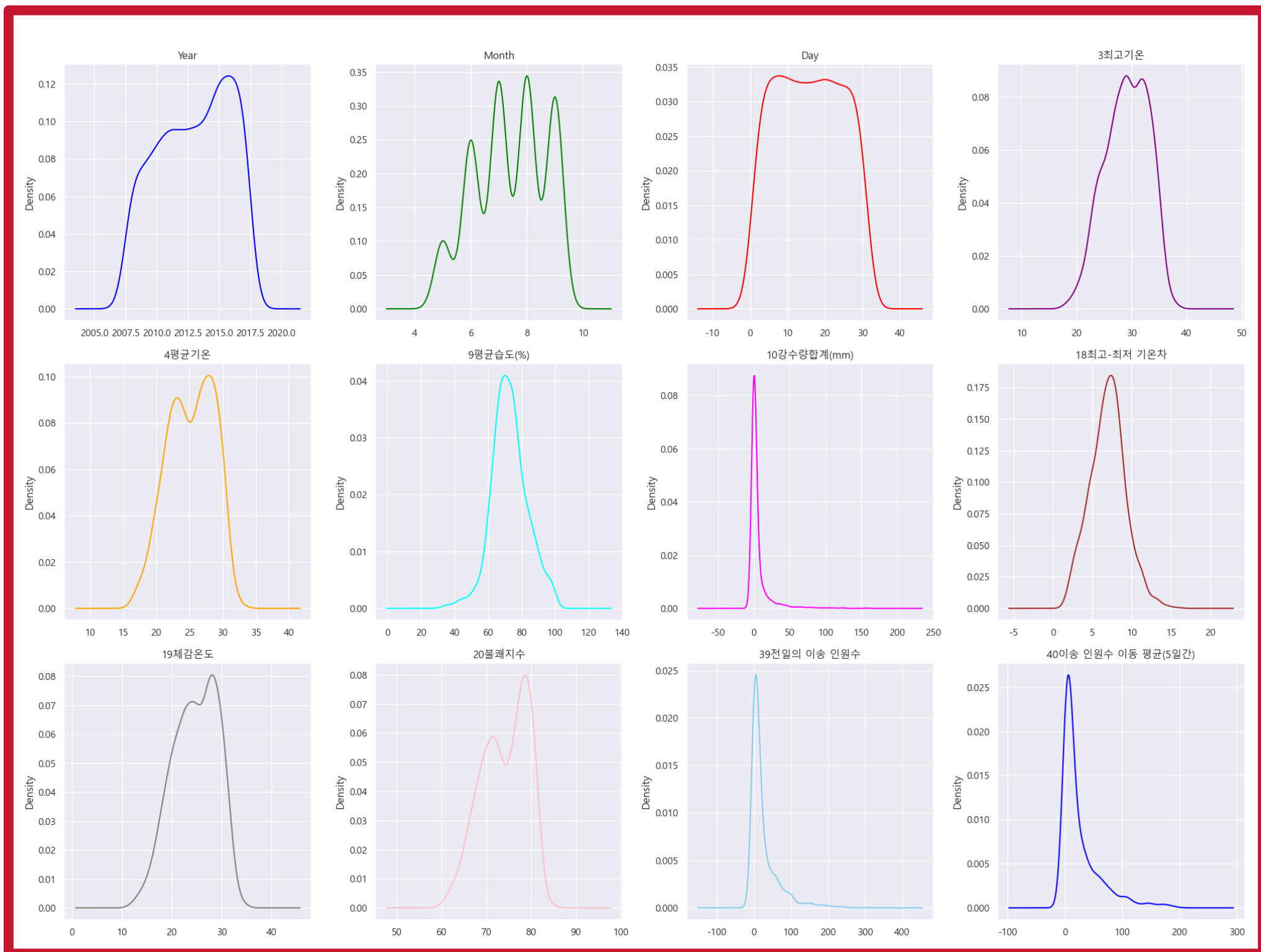
- 종속변수: 열사병 환자 이송 인원수
- 독립변수: 상관관계가 있는 12가지 변수

상관관계가 있는 독립변수

| 연도(Year) | 월(Month) | 일(Day) | 최고기온 | 평균기온 | 평균습도 |
|----------|----------|--------|------|---------------|--------------|
| 강수량 | 일교차 | 체감온도 | 불쾌지수 | 전일의 이송 인원수 | 이송 인원수 평균 |



▲ 종속변수와 각 독립변수 간의 상관관계 그래프(산점도 및 회귀선)



▲ 종속변수와 각 독립변수 간의 상관관계 그래프(밀도 그래프)

IV. 학습모델 선정 및 시각화

[Gradient Boosting 모델 선정]

- 결정 트리 기반, 독립변수와 종속변수 간의 비선형 관계에 특화
- 많은 데이터, 복잡한 독립변수에 대해 효과적인 모델

[Gradient Boosting 모델 예측 결과]

- 산점도: 대각선에 가까운 점일수록 정확도 높은 예측
- 실제 값-예측 값 히스토그램: 실제 값과 예측 값의 분포 평가
- 선 그래프: 고점과 저점의 일치에 따른 정확도 평가
- 오차 분포 히스토그램: 분포가 중심에 모여 있을수록 좋은 예측



V. 교차검증 및 성능평가

[KFold 교차검증 및 성능평가 지표]

- KFold 교차검증: 모델의 전반적인 성능 평가
- 성능평가 지표: 모델의 세부적 성능을 측정해 강점과 약점 분석

[성능평가를 위한 비교모델]

- 비교모델로 Random Forest 모델 선정
- 결정 트리 기반의 비선형 회귀 모델

Gradient Boosting vs Random Forest

| 학습 모델 | KFold MSE | KFold RMSE | KFold MAE | KFold R^2 | MSE | RMSE | MAE | R^2 |
|----------|--------------|---------------|--------------|--------------|--------|-------|-------|------|
| G.B | 177.74 | 13.22 | 7.42 | 0.89 | 1405.8 | 37.49 | 16.62 | 0.81 |
| R.F | 186.28 | 13.48 | 7.47 | 0.88 | 1481.2 | 38.49 | 17.09 | 0.80 |

VI. 결론

- 평가:** 우수한 성능 지표로 신뢰할 수 있는 결과를 제공하는 모델
- 한계:** 기후 변화 갑작스러운 날씨 변동으로 예측에 어려움 존재
- 개선:** 장기적 기후 트렌드를 모델에 통합하는 업데이트 필요
- 의의:** 높은 수준의 안전관리로 작업 효율성 증대, 건설안전수준 향상
- 결론:** 과학적 분석으로 재해예방률 높이고 안전공학분야 발전에 기여