

# 기계학습 캐글 데이터 실습



# CONTENTS

---

**01. 실습**

**02. 결과 및 분석**

# PART 01

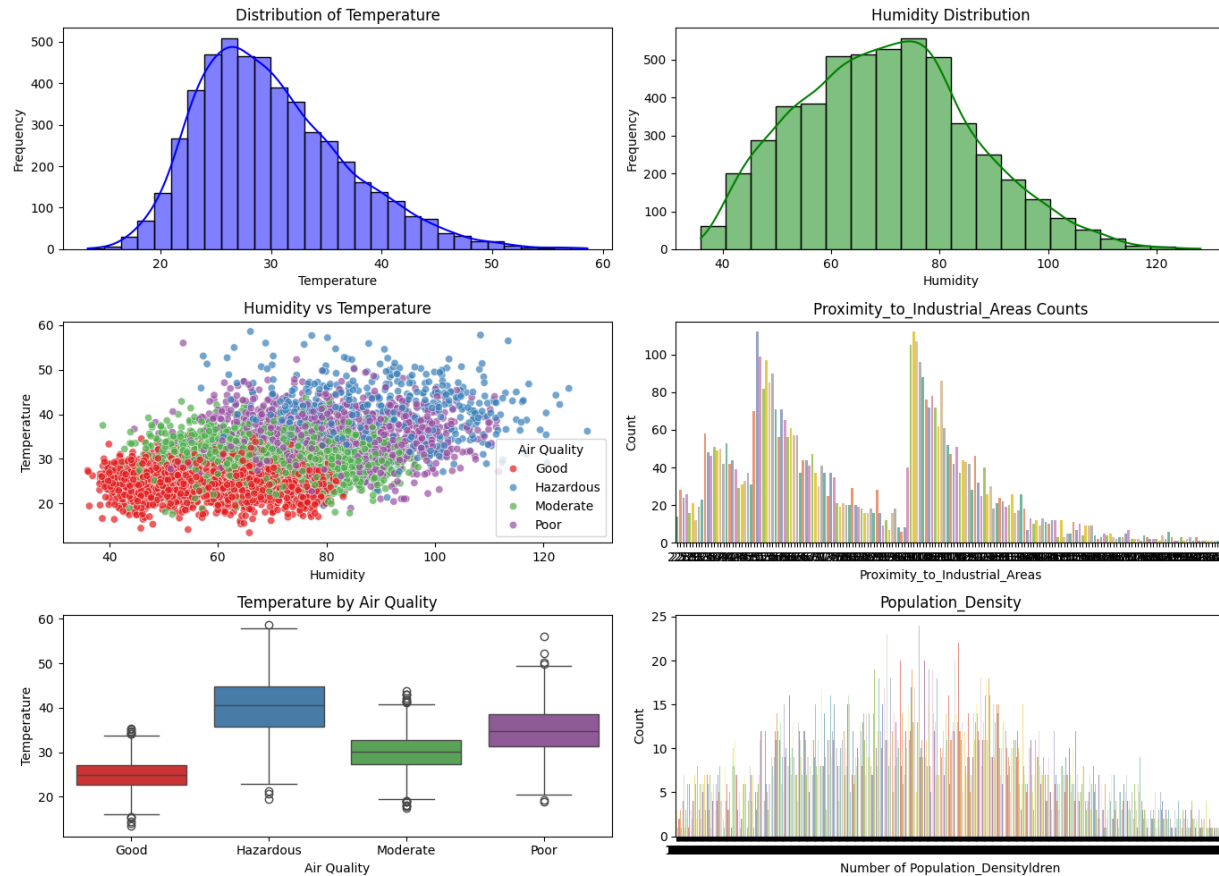
---

실습



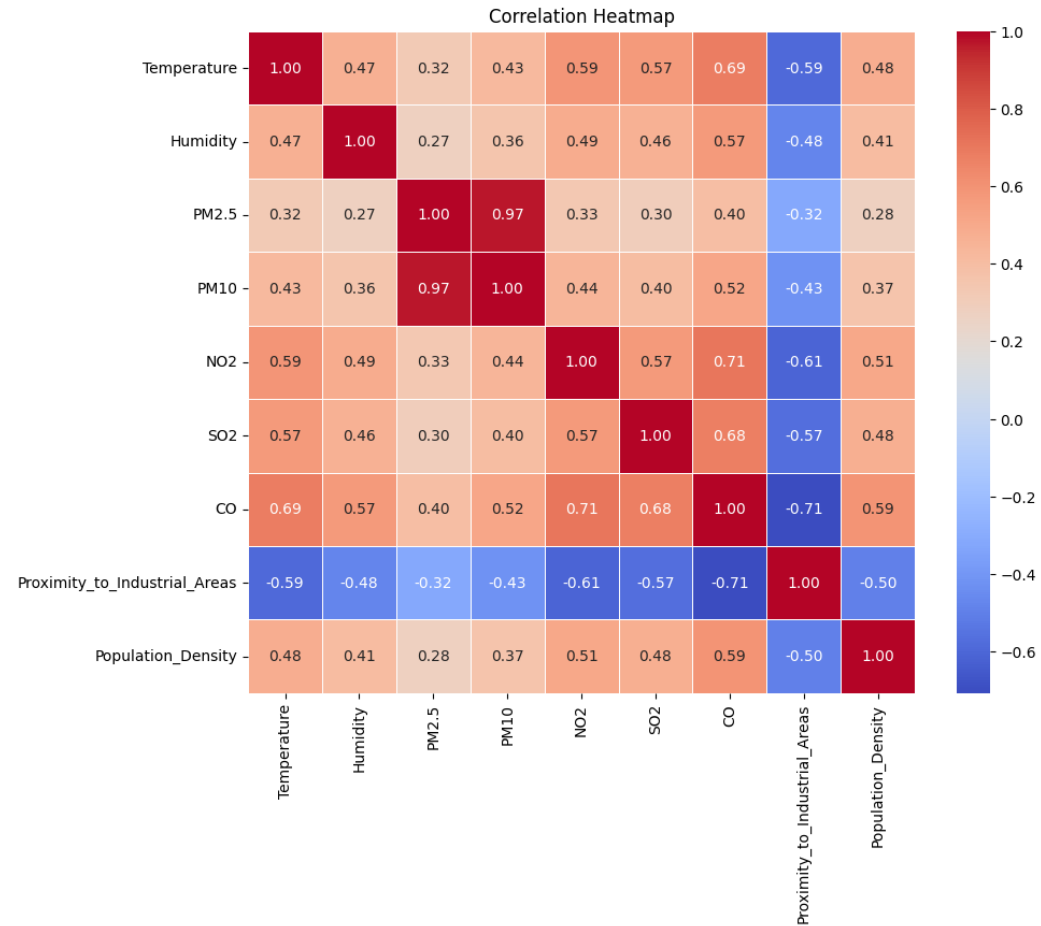


**사용한 데이터셋 :** <https://www.kaggle.com/datasets/mujtabamatin/air-quality-and-pollution-assessment>





**사용한 데이터셋 :** <https://www.kaggle.com/datasets/muitabamatin/air-quality-and-pollution-assessment>





## 평가결과 확인

	Multiple Linear Regression	Decision Tree Regression	Ridge Regression	Lasso Regression	Elastic Net Regression	Random Forest Regression	XGBoost	LightGBM
ME	0.021295	-0.001000	0.079622	0.049755	0.059090	0.003660	0.002010	0.000618
MAE	0.247684	0.053000	0.645696	0.786256	0.721274	0.132520	0.159529	0.143782
MSE	0.113945	0.053000	0.648324	0.757153	0.687685	0.098401	0.120655	0.100648
MSLE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
RMSE	0.337558	0.230217	0.805186	0.870146	0.829268	0.313689	0.347354	0.317250
RMSLE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MPE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
R2	0.292339	0.62554	0.530523	0.451716	0.502020	0.928744	0.912629	0.927117



## 최적화 후 학습재실행

	Decision Tree Regression	최적화 후
ME	-0.001000	0.017482
MAE	0.053000	0.147680
MSE	0.053000	0.183762
MSLE	NaN	NaN
RMSE	0.230217	0.427507
RMSLE	NaN	NaN
MPE	NaN	NaN
R2	0.62554	0.867655

# PART 02

---

## 결과 및 분석







### 모델 선정 및 결과:

- 평가결과를 토대로 제일 높은 성능을 보인 Decision Tree Regression를 그리드서치를 이용해서 예측 정확도가 높은 파라미터로 변경함
- 파라미터 튜닝 후 성능 측정 결과  $R^2$ 값이 0.62554  $\rightarrow$  0.867655가 향상됨

# 감사합니다

