

Estimating Heatwave-attributed Mortalities using Machine Learning

David Guan

Mentors: Jingjing Peng, Peng Yu

Objectives

- Design a machine learning pipeline to predict monthly, provincial-level heatwave mortalities in Spain
- Evaluate feature importances

Methodology

- 80 features were extracted from weather and population data
- The provinces and months were one-hot encoded into features
- Data covered years 2015-2024
- Training dataset made of 6000 data points from 50 provinces * 120 months
- Trained and tested 35 different regression models

Features

Weather

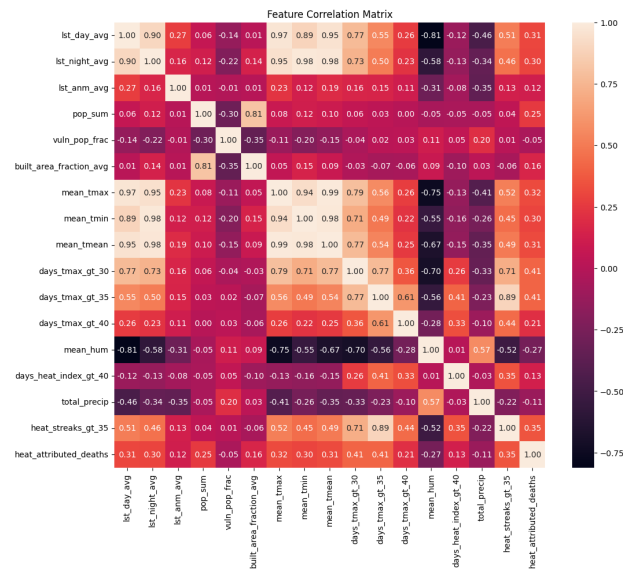
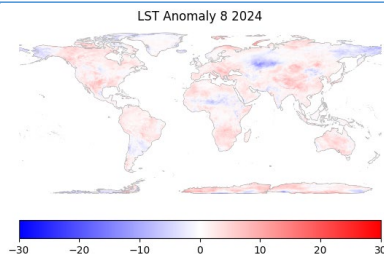
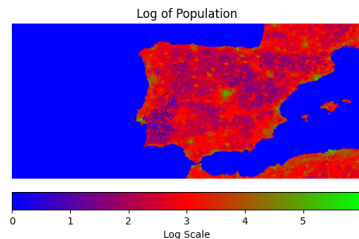
- Average daily max, min, mean temperature
- No. days with max temp > 30, 35, 40 (°C)
- No. days with heat index > 40 (°C)
- No. 3-day streaks with max temp > 35 (°C)
- Day and Night LST
- LST anomaly
- Total precipitation
- Average humidity

Human factors

- Population
- Vulnerable population
- Built area fraction

Categorical (one hot encoded)

- Month
- Province



Estimating Heatwave-attributed Mortalities using Machine Learning

David Guan

Mentors: Jingjing Peng, Peng Yu

Results

Using K-fold cross validation to test 35 different models, The Histogram-based Gradient Boosting Regressor performed most accurately

In order to reduce skew, a log transformation was applied to the target variable before training, then reversed after predictions

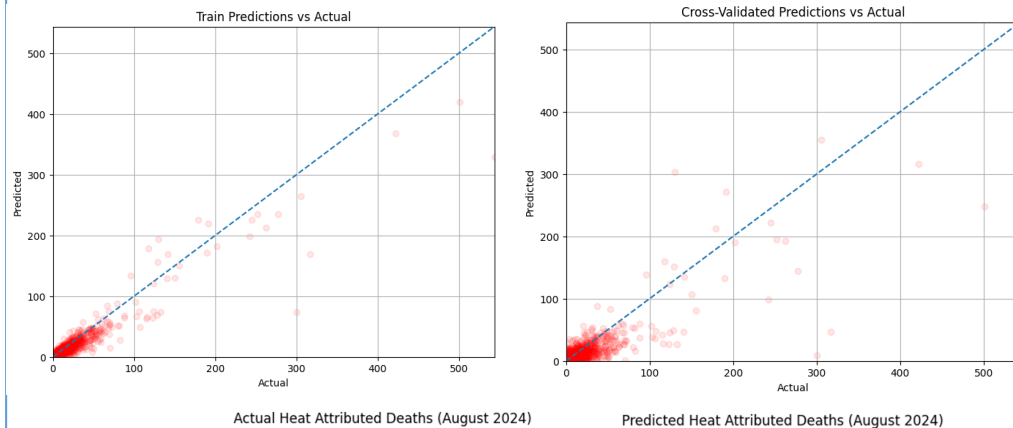
Model hyperparameters were tuned to optimize learning

Performance Metrics

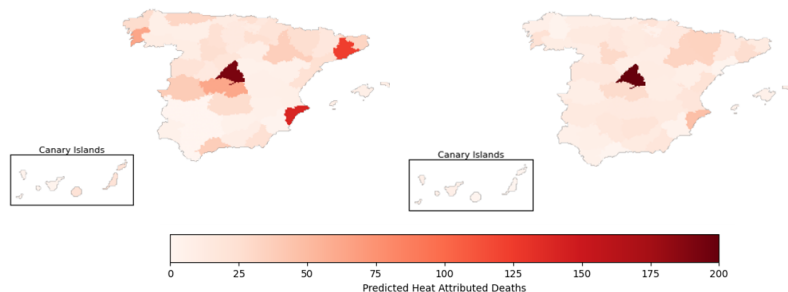
Using 10-fold cross validation on
Hist Gradient Boosting Regressor

	MAE	RMSE	R ²
Train	1.29	6.37	0.888
Test	2.16	10.65	0.688

Train and test performance



Example
from a single
month



Estimating Heatwave-attributed Mortalities using Machine Learning

David Guan

Mentors: Jingjing Peng, Peng Yu

Future Directions

- Expand model to other countries
- Use feature importances to create a heatwave severity index
- Find finer temporal resolution data to predict on a weekly or even daily scale

Data sources

Data	Frequency	Source
Provincial heat wave mortality in Spain	Daily	Carlos III Health Institute's Mortality Monitoring system
Weather	Daily	European Centre for Medium-Range Weather Forecasts
LST and LST anomaly	Monthly	JPSS VIIRS Enterprise
Population		Gridded Population of the World (GPWv4.11)
Built Area fraction		Gridded Population of the World (GPWv4.11)

Feature Importances

Permutation importance was used to determine feature importances by randomly shuffling each feature's values and observing how much accuracy decreases

