

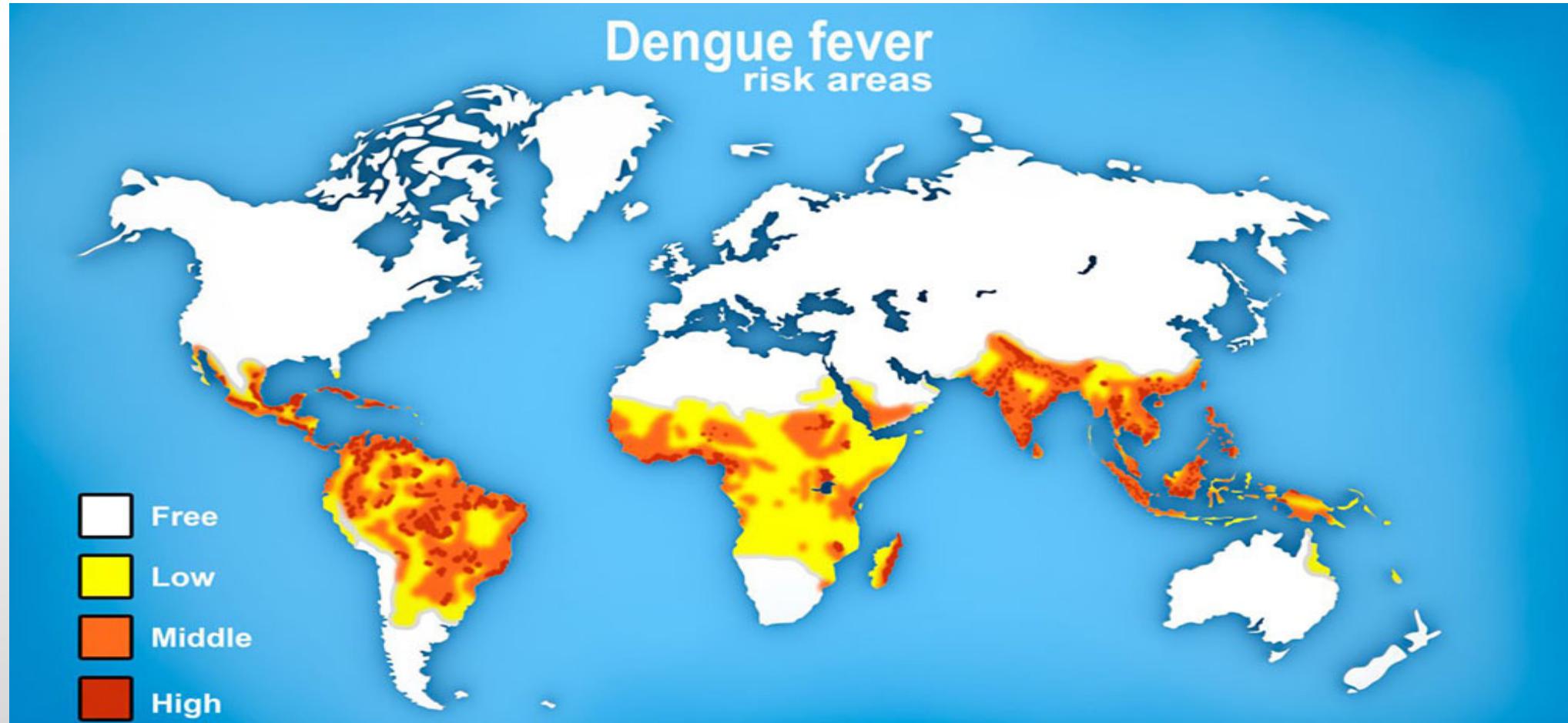
# Dengue Fever Outbreak Prediction:

## Using Climate Variability



Aysu Erdemir

April 2023



- Dengue is a **viral infection** that spreads from mosquitoes to people.
- It is more common in **tropical and subtropical climates**.
- It causes flu-like symptoms such as fever, headache, body aches, nausea and rash; but sometimes it worsens to **severe dengue**, a life-threatening complication.



Stakeholder:

CDC, Center for Disease Control and Prevention



Business Problem:

Can we **predict** when the **next dengue outbreak** will happen, using the relationship between climate and dengue dynamics?

Project Value:

- Improve research initiatives to better understand the complex relationship between climate and dengue.
- Resource allocation to help fight life-threatening dengue epidemics.

# Data:

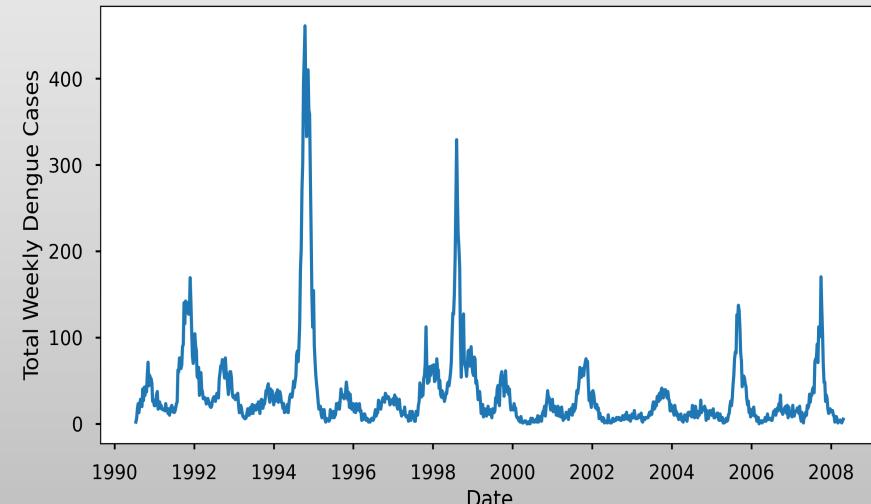
Time series data obtained from [DrivenData](#) including **weekly dengue case counts** in **Puerto Rico** from year 1990 to 2008 along with accompanying climate features such as:

- Humidity,
- Max, min, average temperature
- Precipitation
- Vegetation index

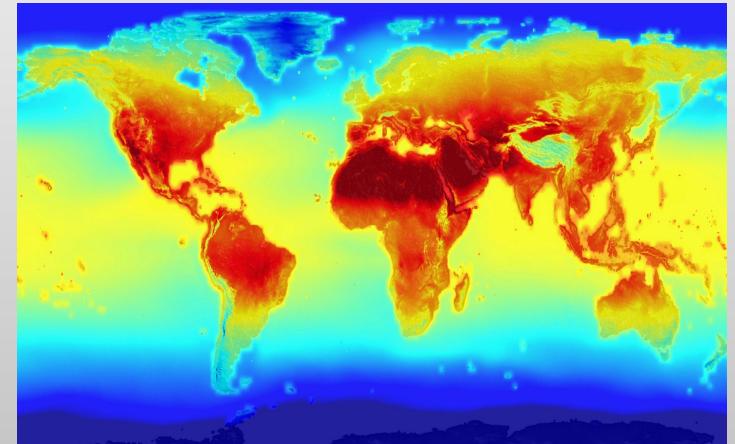
Additional data from 2008–2013 included climate features without case counts.



**Weekly case counts**

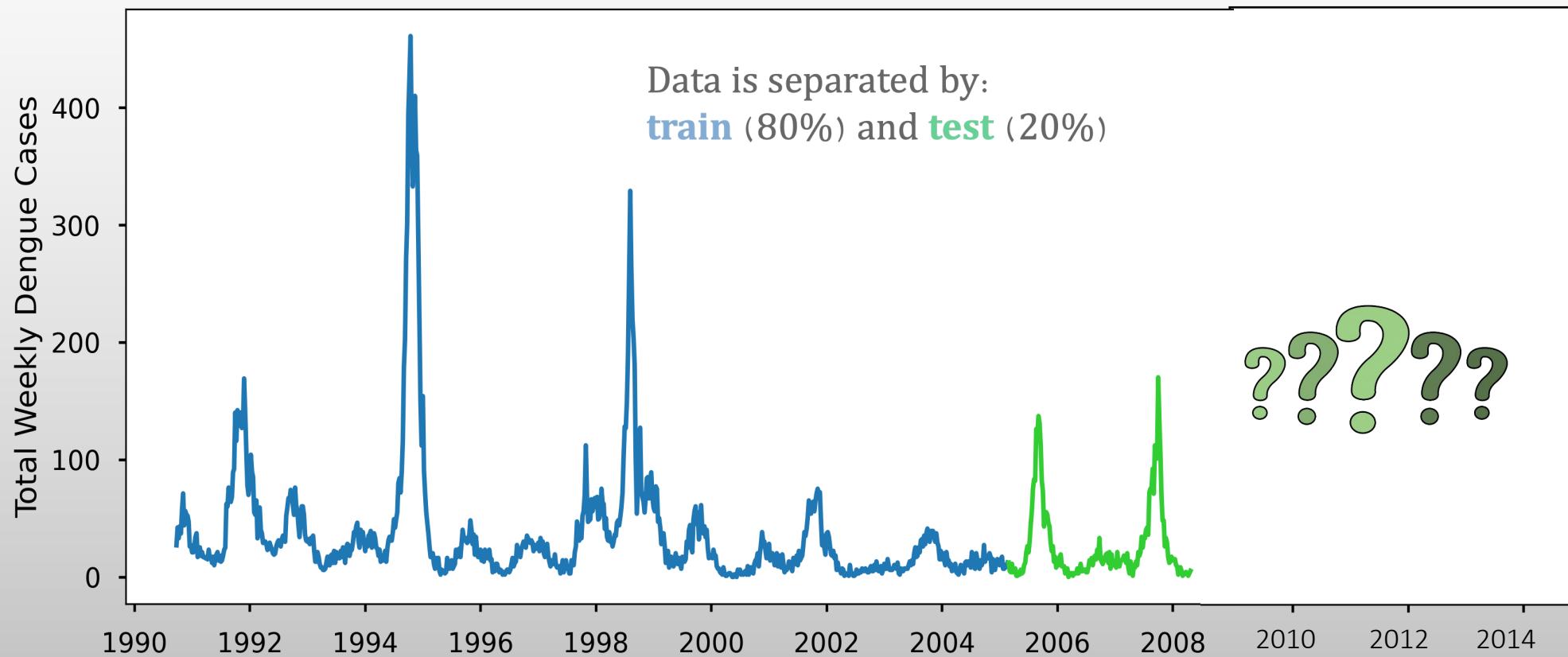


**Climate variables**

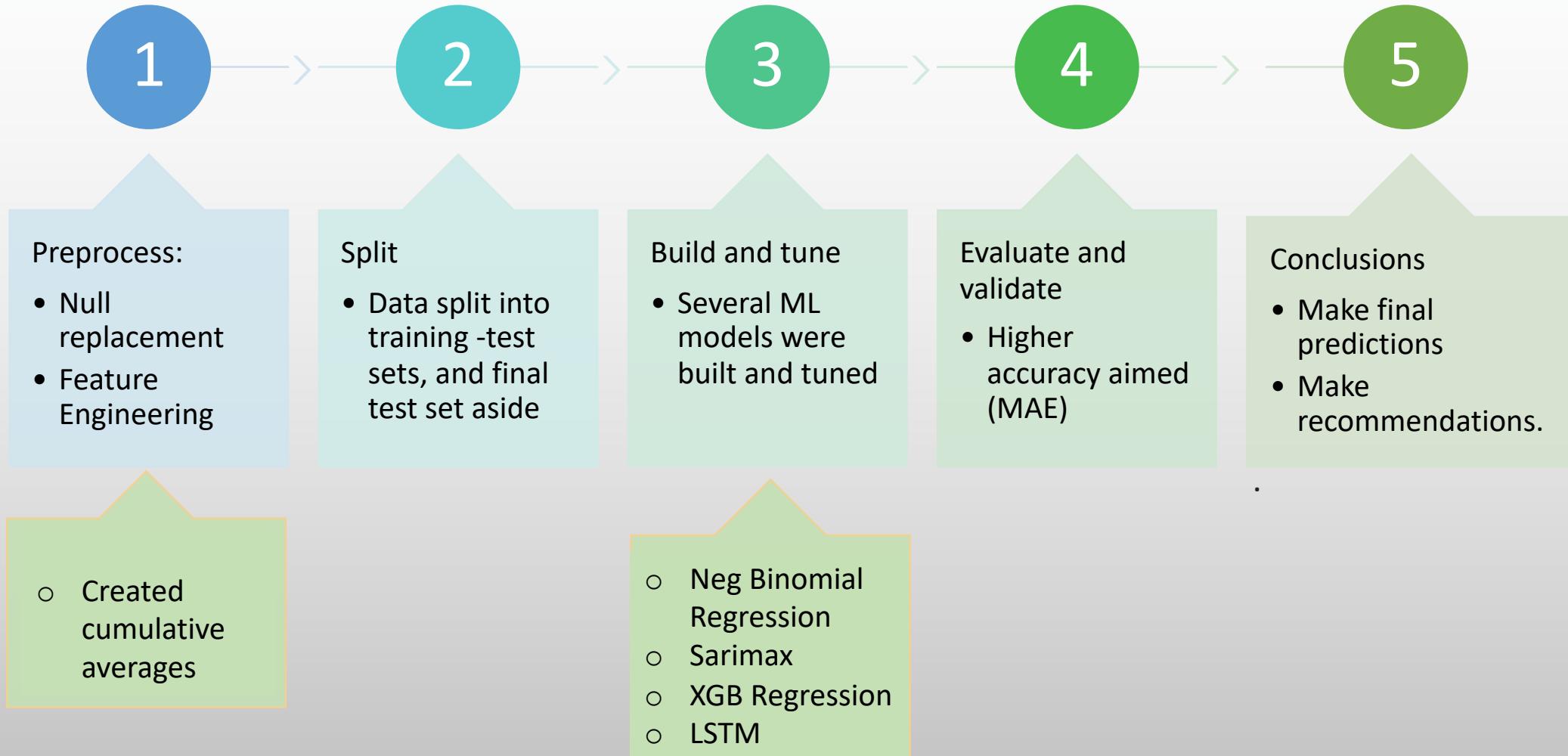


# Goal:

- Build several machine learning models to **forecast** upcoming weekly dengue cases as accurately as possible.
- Performance is evaluated according to the **Mean Absolute Error** (absolute difference between actual and predicted case counts).



# Methodology:



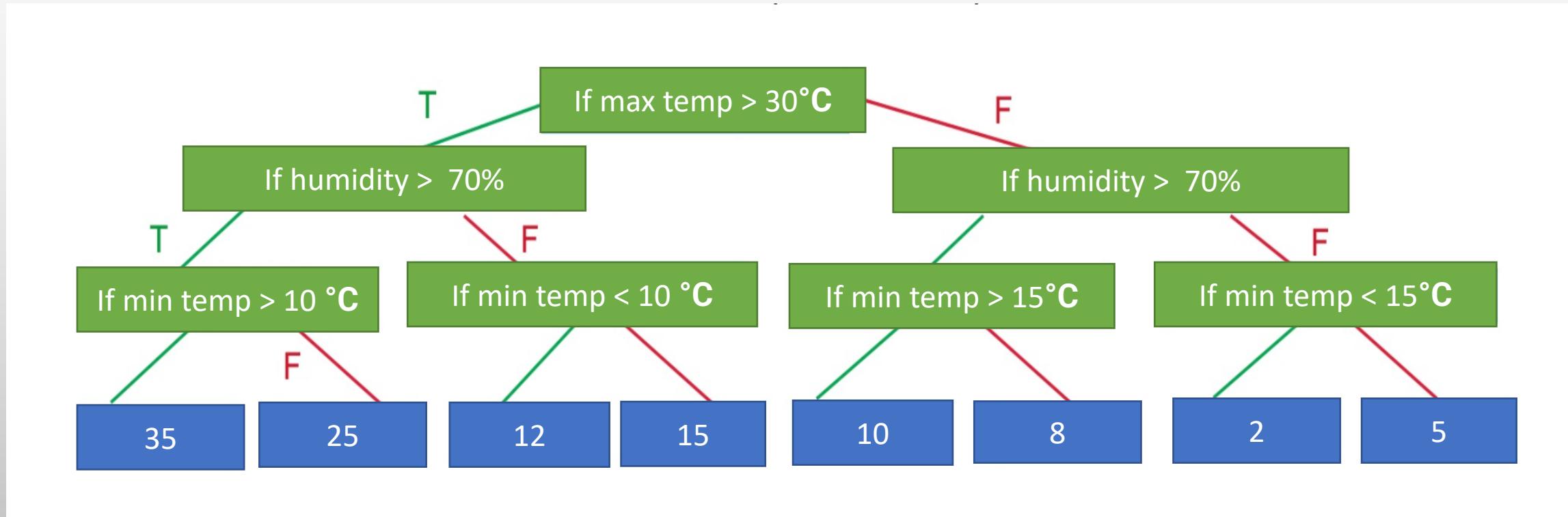
# Preprocess:

- Null values were imputed using **interpolation** and **predictive modeling** (KNN).
- Main climate variables were:
  - **Shifted** by 2 weeks (to account for mosquito lifecycle and incubation period of the virus).
  - **Feature-engineered** by taking the mostly correlated **cumulative averages** at various lags ranging from 2-4 months.

# Best performing model:

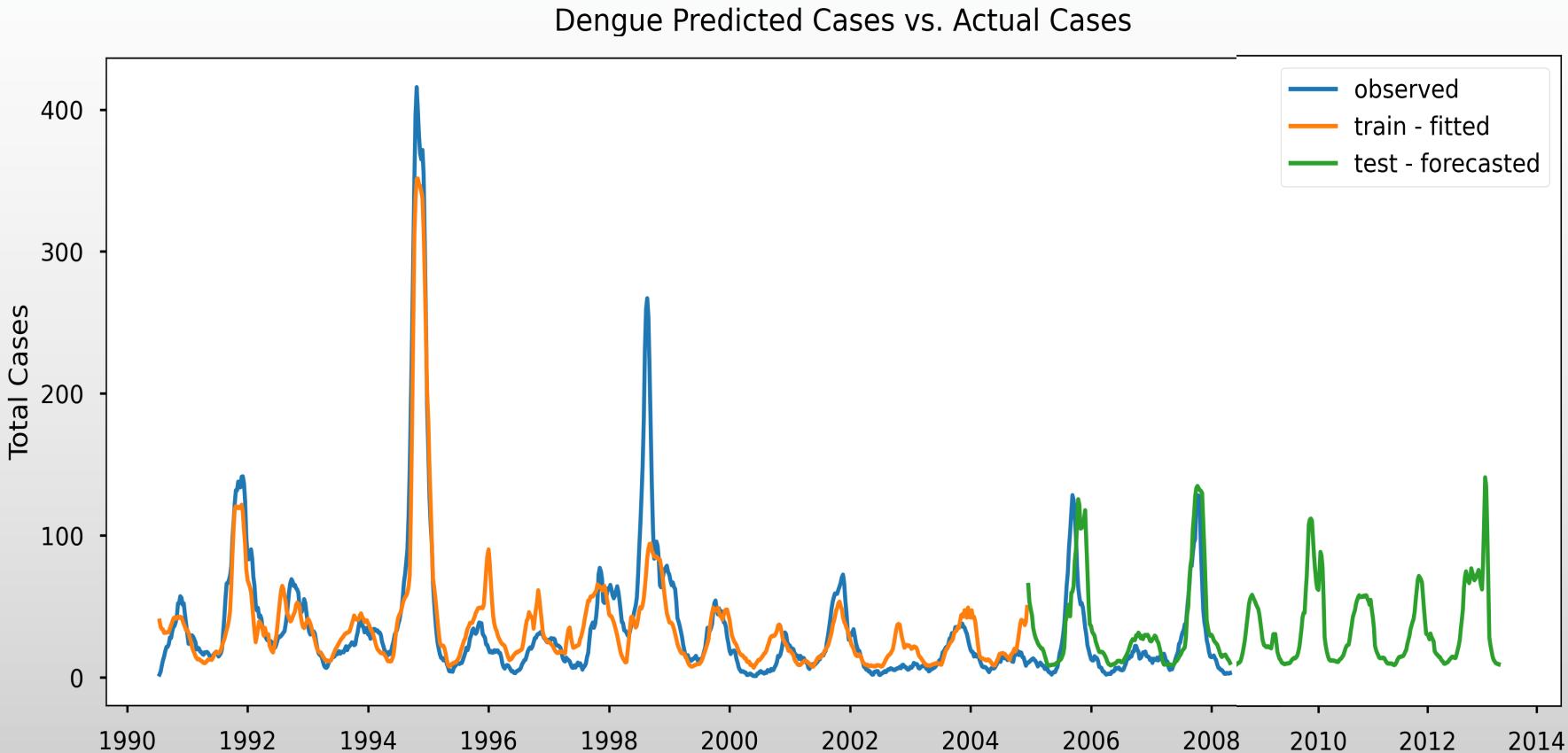
## XGBoost Regression (decision tree + gradient boosting)

In the simple example below, a decision tree is used to estimate dengue case numbers (the label) based on max temperature, min temperature and humidity (the features).



# Validate and Evaluate:

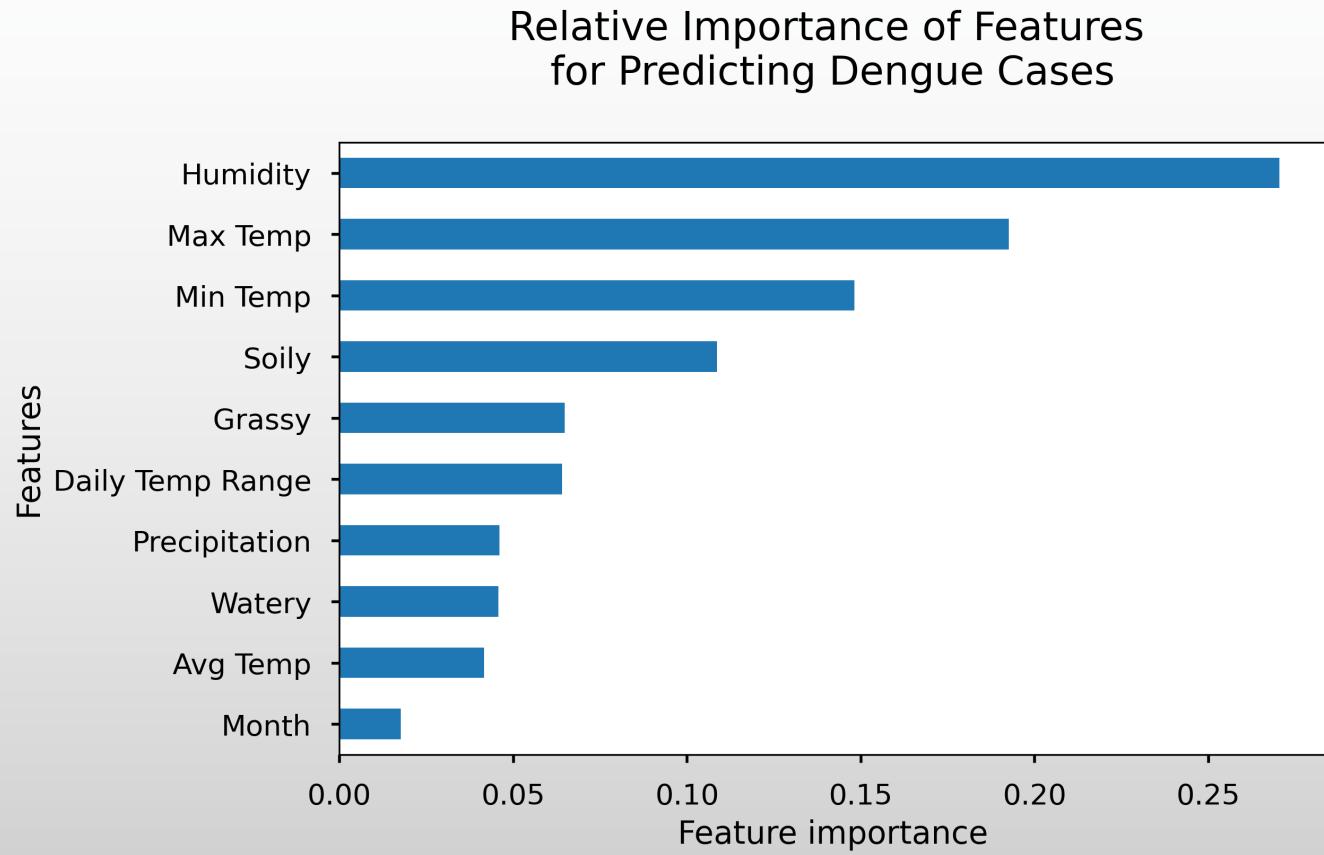
## Model Performance and forecasting



	Train	Test
Mean Absolute Error in weekly case counts	14.6	17.7

- Final model detects majority of the **individual peaks / outbreaks**.
- 2 more moderate size outbreaks are expected to happen at the **end of 2009 and 2012**

# Feature Importance



Sustained humidity and sustained maximum temperature are the two most important features in predicting dengue cases.



# Conclusions / Recommendations:

- Dengue cases rely on climate variables, but the relationship is complex.
  - Further models should take into consideration cumulative computations of climate features over a period rather than isolated numbers.
- Climate change and global warming may make dengue outbreaks and similar mosquito born illnesses more deadly in the future.
  - Knowing the next outbreak would help countries to allocate more resources to the health care system prior to the outbreak for timely intervention.



# Limitations and Improvements

- More recent data needs to be collected to achieve more accurate predictions for future.
- Since the relationship between dengue and climate is complex:
  - **Nonlinear relationships** need to be taken into account with more complex models.
  - More meaningful climate related features need to be engineered.





**Email:** erdemiraysu@gmail.com

**GitHub:** @erdemiraysu

**LinkedIn:**

[linkedin.com/in/aysuerdemir/](https://www.linkedin.com/in/aysuerdemir/)