DengAI: Predicting Disease Spread

Team: Hyper-Parameters

UMLSC 2021

Parameters:
Janvi Bhagiya | Sai Rithvik | Saili Myana

Motivation

- Dengue fever has been spreading
- 0.5 billion cases per year in Latin America
- A complex task to model the relation between climate and dengue dynamics
- Can improve research initiatives and resource allocation to help fight life-threatening pandemics.

Objective

- To predict the number of dengue cases each week (in locations like San Juan, Puerto Rico and Iquitos, Peru) based on environmental variables describing changes in temperature, precipitation, vegetation, and more.

Existing Literature

- Anuranjan M B et al. [2] in 2019 aimed at building a random forest regression and Gradient Boosting regression models and also perform data manipulation techniques along with dimensionality reduction.
- P.Muhilthini et al. [3] in 2018 proposed a dengue forecasting model based on the number of dengue cases observed every week for several years in many countries and weather conditions. This used GBR to find the pattern and dependencies in the given training data set to predict the number of dengue cases.
- In 2014, a study [4] in Sri Lanka, using past weather patterns and past dengue cases as inputs, and introduced an ANN architecture to predict the dengue outbreak in Kandy district.

[2]https://www.researchgate.net/publication/339473958_Predicting_Dengue_Spread_in_San_Juan_and_Iquitos_using_Machine_Learning
[3] Muhilthini, P., Meenakshi, B. S., Lekha, S. L., Santhanalakshmi, S. T. (2018). Dengue Possibility Forecasting Model using Machine Learning Algorithms.
[4] P. H. M. N. Herath, A. A. I. Perera, and H. P. Wijekoon, “Prediction of dengue outbreaks in Sri Lanka using artificial neural networks.” International Journal of Computer Applications, vol. 101, no. 15, pp. 1–5, 2014.

Dataset

Data Features :

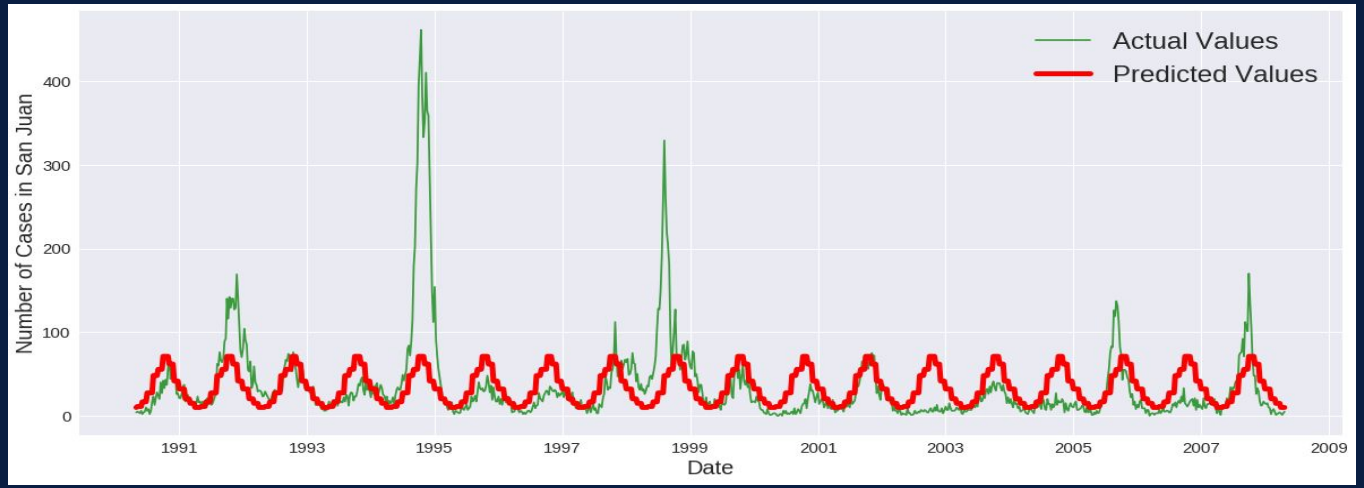
- Temperature
  - Max, min, average, diurnal range
  - dew point
- Precipitation
  - Total rainfall
- Humidity
  - Mean relative and mean specific
- Vegetation
  - Level of vegetation in NW, NE, SW and SE quadrants of city as measured by the satellite imagery

Data is taken from the Hackathon “DengAI: Predicting Disease Spread” on drivendata.org [1]

[1] https://www.drivendata.org/competitions/44/dengai-predicting-disease-spread/

Methodology

- We require models that use encoding and preserve the time series nature of the data because the distribution of the number of dengue cases depends on the month/season of the year too.
- Model 0 : Simple Regressor to predict the “Trend of the year”
- MAE for Model 0: 26.5



Methodology

- To add the variation in cases due to environmental data and vegetation index, we model the difference between number of cases and Model 0.
- Model 1: ANN to capture the variations in weather and predict the residual of the actual cases and the ones predicted by Model 0.

Final Model = No of cases from Model 0 + Model 1

Future Work

- Feature Engineering to decide the final features for predicting the outbreak by Model 1
- Rolling Window Statistics could also be considered
- Sequential Models can be considered to take into account the previous data