DengAI – Disease Spread Prediction

Problem Statement:

Dengue fever is a mosquito-borne disease that occurs in tropical and sub-tropical parts of the world. In mild cases, symptoms are similar to the flu: fever, rash, and muscle and joint pain. In severe cases, dengue fever can cause severe bleeding, low blood pressure, and even death.Because it is carried by mosquitoes, the transmission dynamics of dengue are related to climate variables such as temperature and precipitation. Although the relationship to climate is complex, a growing number of scientists argue that climate change is likely to produce distributional shifts that will have significant public health implications worldwide.In recent years dengue fever has been spreading. Historically, the disease has been most prevalent in Southeast Asia and the Pacific islands. These days many of the nearly half billion cases per year are occurring in Latin America:Using environmental data collected by various U.S. Federal Government agencies—from the Centers for Disease Control and Prevention to the National Oceanic and Atmospheric Administration in the U.S. Department of Commerce we have to predict the number of dengue fever cases reported each week in San Juan, Puerto Rico and Iquitos, Peru.

Background:

The task is to predict the number of dengue cases each week (in each location) based on environmental variables describing changes in temperature, precipitation, vegetation, and more. An understanding of the relationship between climate and dengue dynamics can improve research initiatives and resource allocation to help fight life-threatening pandemics. The models proposed for this project are ARIMA, ARIMAX, SARIMA, SARIMAX AND Random Forest.

Methodology:

The models to tackle this problem statement is:

1)Statistical Models

- ARIMA
- ARIMAX
- SARIMA
- SARIMAX

2)Machine Learning Models

- Random Forest

Dataset:

The data has been obtained from drivendata.org, which was released as a part of its competition. It consists of 1456 entries of training data and 416 entries of testing data for two cities: San Juan and Iquitos. Climate influenced variables like maximum and minimum temperature, humidity, precipitation, etc. were considered.

Evaluation Measures:

The metric which is being used to evaluate our model in this competition is based on mean absolute error.

Software and Hardware Requirements:

Python based stats and forecasting libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such prophet,auto_arima, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs, Kaggle notebooks and colab notebooks for training.