

# Predict Local Epidemics of Dengue Fever Proposal

Group Members: Lu Cao, Elva Gao, Kinya Wang

Dataset Link: <https://www.drivendata.org/competitions/44/dengai-predicting-disease-spread/>

Dear Managers,

We would like to propose to analyze the question of “how we can predict local epidemics of dengue fever” in order to bring more social impact in terms of public health. We will discuss the rationale of the problem and more specifics below throughout the memo.

## I. Problem Specification

Dengue fever is a mosquito-borne disease that occurs in tropical and subtropical regions of the world. Due to the transmission characteristics of dengue fever, more and more scientists believe that climate factors have a complex and non-neglectable relationship with dengue fever's spread. In our project, we want to get a better understanding of the relationship between climate and dengue dynamics to improve research initiatives and resource allocation to help fight this life-threatening pandemic.

The purpose of this project is to train a model to predict the number of dengue fever cases reported each week in regions including Puerto Rico, Peru, and San Juan based on environmental variables including temperature, precipitation, vegetation, etc. Being able to predict the infected cases would be impactful with providing a deeper understanding of the root and causes of dengue fever spread, along with helping with research and resource allocation to prevent and fight severe pandemics.

## II. Data Description & Solution Feasibility

Our dataset consists of information of one of the two cities, date of the week, temperature distribution and precipitation measured by different climate forecast systems, and vegetation index measured on a weekly basis over the time span of around 20 years. There are a total of 22 features, which include 1 ordinal variable, 1 discrete variable, and 20 continuous variables.

We could explore the relationship between the number of dengue fever cases and features in this dataset using different graphing and statistical measure methods. Those underlying correlations will enable us to model the number of increased cases of dengue fever every week in Puerto Rico and San Juan.

As we mentioned above, the climate is a major factor that affects the spread of this disease. Since this dataset gives us comprehensive information about climate through a long timestamp, we are confident that we have enough data to assess the relationship between climate and disease from various perspectives, which will result in a good predictor.

Thanks,

Lu Cao, Elva Gao, Kinya Wang