

West Nile Virus Prediction

Problem

West Nile virus is widely spread to humans through infected mosquitos and its symptoms are ranged from a persistent fever to serious neurological illnesses. The primary goal is to build a model that predicts outbreaks of West Nile virus in mosquitos for given environmental conditions, such as weather, season, and location. The data has been established by the City of Chicago and the Chicago Department of Public Health (CDPH) for the period between 2007 and 2013, including the test results of the virus in trapped mosquitos.

Client

The accurate prediction of the virus outbreaks will help the City of Chicago and CDPH more efficiently and effectively allocate resources towards preventing transmission of the deadly virus.

Data

The data is collected from the Kaggle competition and can be found at <https://www.kaggle.com/competitions/predict-west-nile-virus/data>

The data is composed of the main dataset, spray data, weather data, and map data. The main dataset includes the test results of the trapped mosquitos.

Approach

- clean the dataset by imputing the missing data.
- explore the raw data that includes the information on the infected mosquitos spraying over the city and find some basic patterns of the distributions.
- build basic prediction models (such as linear regression) and measure the performance.
- add more features of interest such as the weather condition (e.g., dry vs. wet, hot vs. cold) to improve the model.
- perform the modeling with various methods (such as random forest, Boosting, SVM, etc) to find a model that gives a better performance
- use advanced Machine Learning models on the increased feature space to narrow down the most important features to use.
- interpret the correlation between the predicted outbreak and the given environmental features that the model suggests.

Deliverables

- Jupyter notebook that contains exploratory data analysis, feature engineering, and modeling.
- Final report that includes the entire process from data cleaning to model solutions.
- Slide deck for presentation.