

Predicting West Nile Virus Rates in Chicago

Problem Statement - A Capstone Element for Springboard's Data Science Program

Problem Statement

Using data provided by the Chicago Department of Public Health, can a model be created to accurately predict when and where an instance of the virus may occur?

Context

West Nile virus is most commonly spread to humans through infected mosquitos. Around 20% of people who become infected with the virus develop symptoms ranging from a persistent fever to serious neurological illnesses that can result in death.

In 2002, the first human cases of West Nile virus were reported in Chicago. By 2004 the City of Chicago and the Chicago Department of Public Health (CDPH) had established a comprehensive surveillance and control program that is still in effect today.

Every week from late spring through the fall, mosquitos in traps across the city are tested for the virus. The results of these tests influence when and where the city will spray airborne pesticides to control adult mosquito populations.

Criteria for Success

Build and train a model that accurately predicts when and where an instance of the West Nile Virus will occur amongst several species of Mosquitos within the City of Chicago.

Scope of Solution Space

Precisely predict when and where the virus will occur; identify the species with the highest positive rate; identify time of year, weather and geographic region with highest rates.

Constraints

Mosquitos are only collected weekly from late spring through fall, providing data on odd years from 2007 to 2013.

Stakeholders

The City of Chicago City Management, the Department of Public Health

Data Sources

The data used in this Capstone was provided through Kaggle at:
<https://www.kaggle.com/c/predict-west-nile-virus>