West Nile Virus Prediction

Predict West Nile virus in mosquitoes across the city of Chicago

DSI03 : 28 Nov 2023

Introduction

West Nile virus (WNV) is the leading cause of mosquito-borne disease in the continental United States. It is most commonly spread to people by the bite of an infected mosquito.

There are no vaccines to prevent or medications to treat WNV in people.

Agenda

01

Problem Declaration

02

Data Process

03

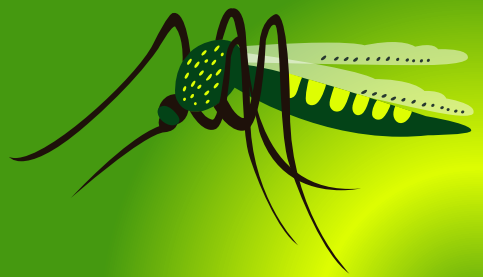
Overview & Data Cleansing

04

Modeling & Evaluation

05

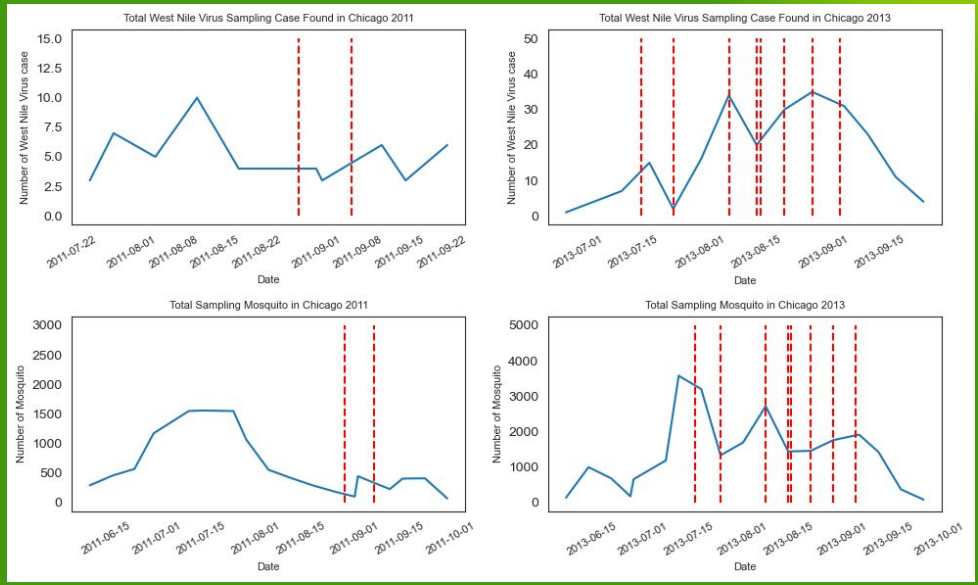
Conclusions & Recommendations



Problem Declaration

"..In 2011 and 2013, in order to prevent West Nile Virus; WNV Infection in Chicago, insecticide sprayings are conducted in some specific areas to reduce the number of mosquitoes.

Anyways, after gaining the number of mosquitoes caught in traps around Chicago city on the spraying days, Government of Illinois are suspected if the Chicago City Council were doing the right way as the number of mosquitoes and the number of West Nile Virus sampling cases still increased even after spraying dates..

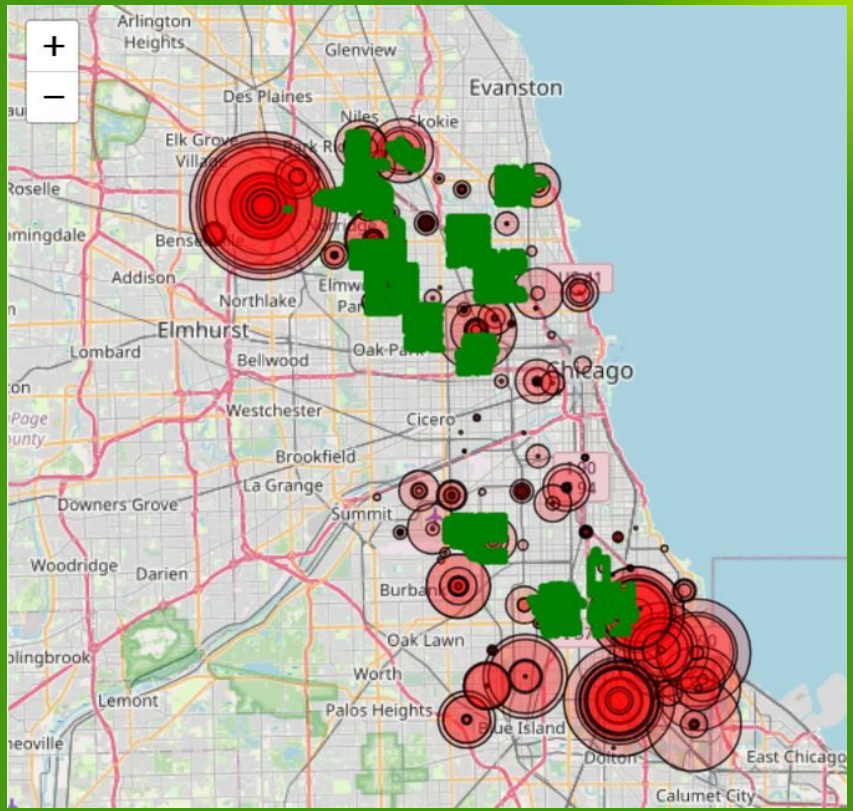


Problem Declaration (Cont.)

Head of Data Science of Government of Illinois found that the spraying location were not matched with the major outbreaking areas so they were curious on how they determined the spraying locations.

After meeting with Chicago City Council, Government of Illinois requested the Head of Data Science and his team to develop a classification model that can notify the risk of West Nile Virus outbreaks in specific geographic regions in Chicago.

In the Data Science team splint planning, the team decided to use daily local weather data to create the classification model after studying the West Nile Virus..”



Data Process

Cleaning Data

Check null and duplicate data , fix missing value

Working on Features

Create and extract columns in need

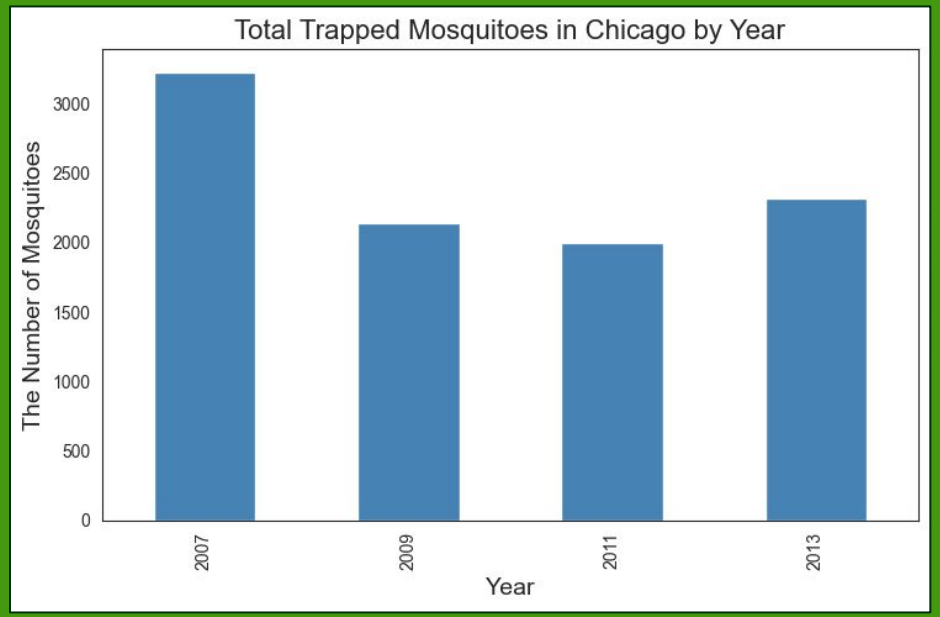
Create Dataset

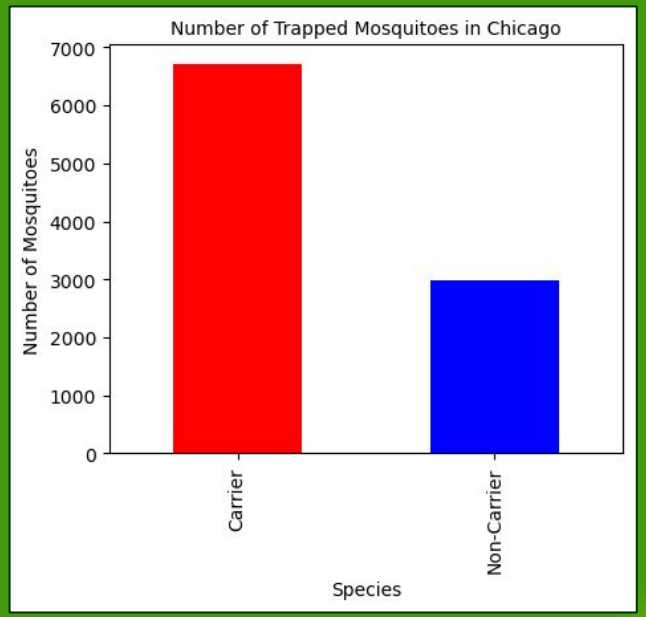
Selected features for for Train & Test data

Overview & Data Cleansing

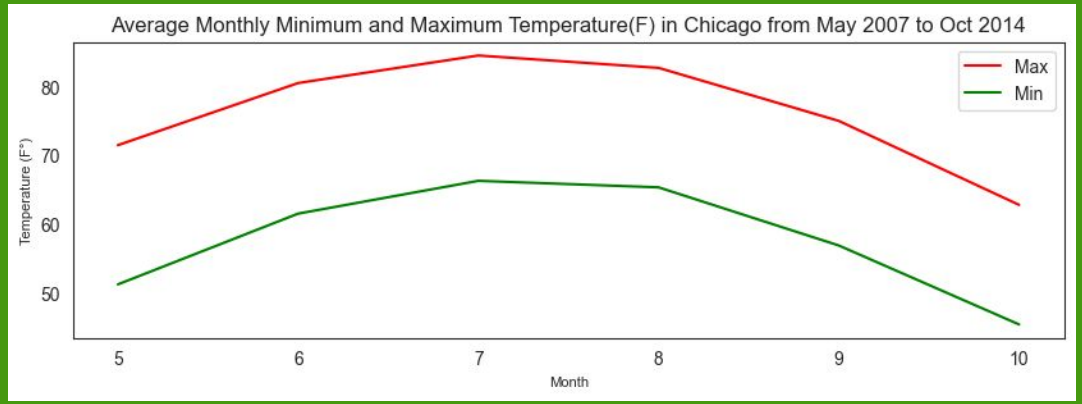
- Observations were collected in 2007, 2009, 2011 and 2013 in aspect of geographical features, presence of West Nile Virus, recording date related to Mosquito trap ID.
- 9,663 as total number of observations after duplicated value removal.
- Merging the above DataFrame with local daily weather dataset by closest climate station and climate recording date as joining keys.
- Re-grouping mosquitoes species by WNV carrier and non- WNV carrier
- SMOTE technique is applied to the dataset which haves imbalance of target variable classes (5% - as NW Virus positive, 95% - as NW Virus negative)

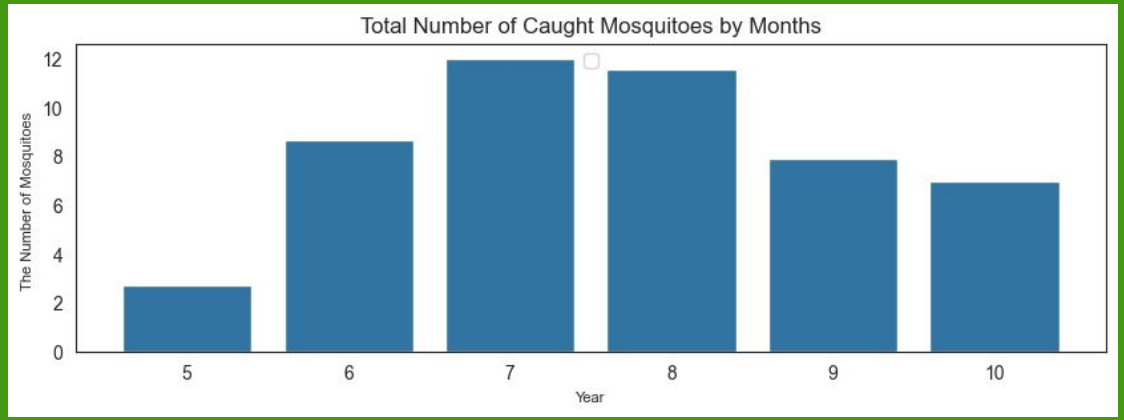
The number of carrier mosquitoes is approximately doubled to the non-carrier mosquitoes

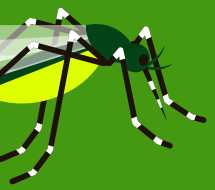




Temperatures seems to have positive relationship with the number of Mosquitoes



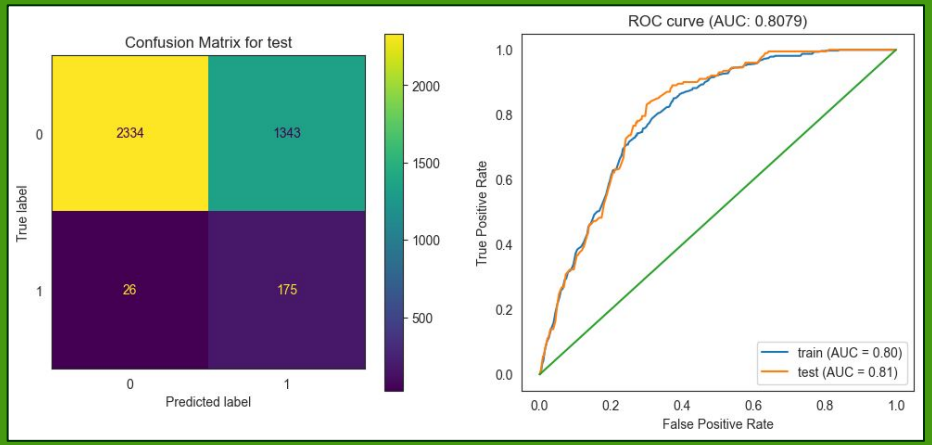




Modeling & Evaluation

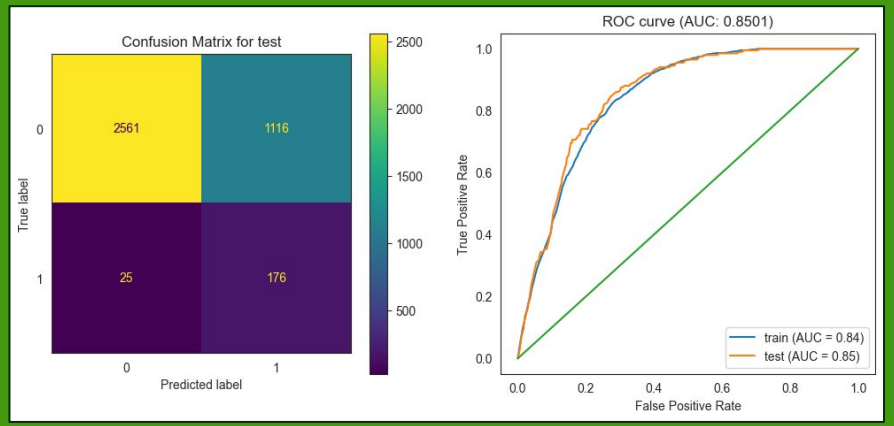
1. Logistic Regression
2. Random Forest
3. XGBoost
4. SVM
5. ADA

Logistic Regression with GridSearch



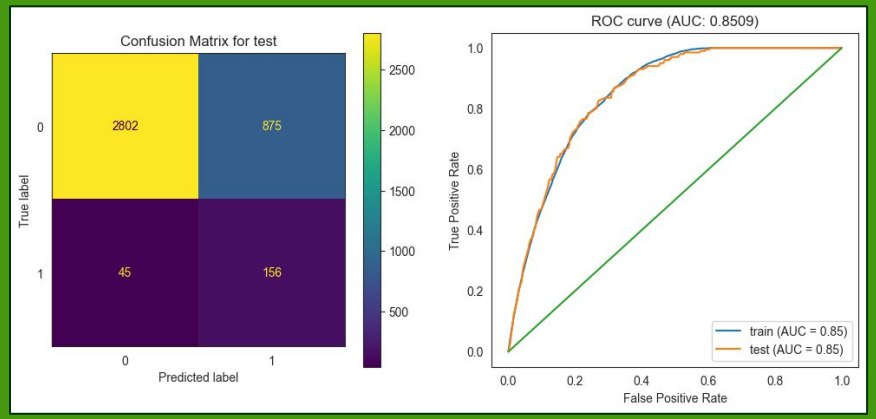
Dataset	Accuracy	Precision	F1	Recall
Train	0.734	0.695	0.758	0.834
Test	0.644	0.116	0.206	0.890

Random Forest



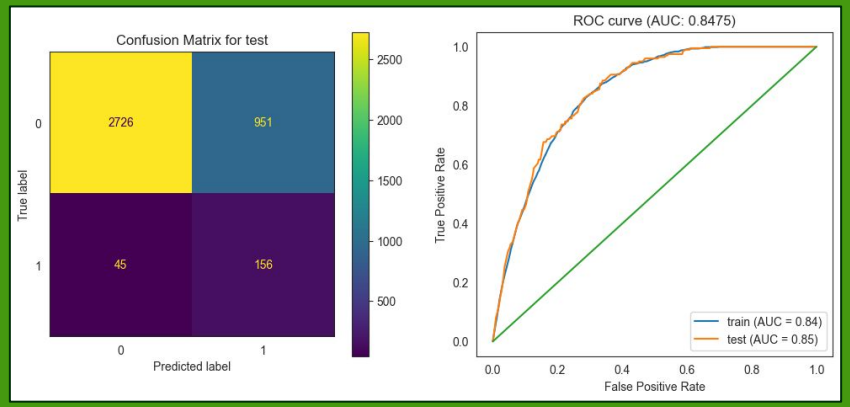
Dataset	Accuracy	Precision	F1	Recall
Train	0.796	0.746	0.815	0.898
Test	0.705	0.136	0.235	0.875

XGBoost



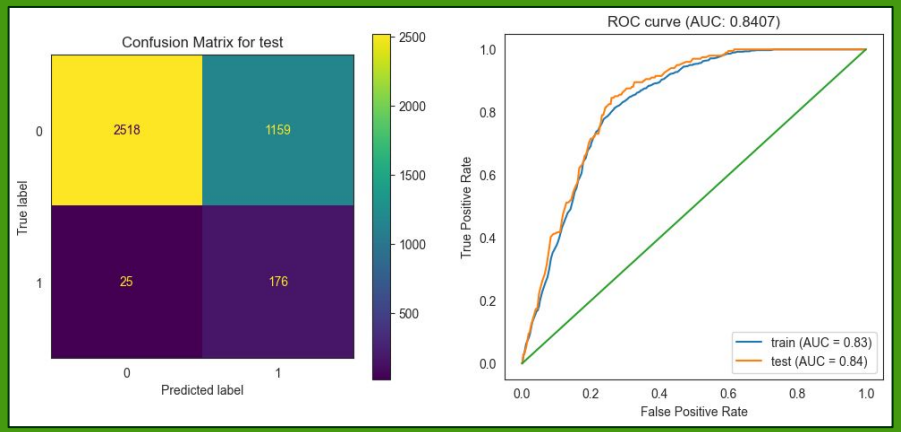
Dataset	Accuracy	Precision	F1	Recall
Train	0.833	0.792	0.843	0.902
Test	0.762	0.151	0.253	0.776

ADA



Dataset	Accuracy	Precision	F1	Recall
Train	0.810	0.773	0.823	0.879
Test	0.743	0.140	0.238	0.776

SVM



Dataset	Accuracy	Precision	F1	Recall
Train	0.771	0.730	0.789	0.859
Test	0.694	0.131	0.229	0.876

Model Evaluation Comparison

Model	Recall score	Precision score
Logistic Regression with GridSearchCV	0.890	0.114
Random Forest	0.875	0.235
XGBoost	0.776	0.253
ADA	0.776	0.238
SVM	0.875	0.229

Conclusions & Recommendations

After creating 6 models, the data science team decided to deliver Logistic Regression with GridSearchCV Model, as a first proposal to Government of Illinois, which is able to detect 89% of total West Nile Virus presence on unseen dataset. In addition, from the predicted results of the model, only 11% of the predicted presence of West Nile Virus is actually existed, which might lead to over-budgeting on hospitality expense in Chicago. If Government of Illinois approved to use this model, Chicago City Council also needs to find countermeasures for operation cost-saving.

As for spray that not effective might change to use Aerial spraying that is process by Airplanes and helicopters to treat very large areas with larvicides that kills mosquito larvae that hatch from eggs or adulticides to quickly kills flying mosquitoes.Both larvicides and adulticides can temporarily help reduce the number of mosquitoes in an area.

Cost - Benefit

Early intervention can help prevent the spread of the virus, reducing the overall cost of responding to a widespread outbreak and lowers the burden on healthcare systems.This can lead to sustained cost savings in public health expenditures related to WNV surveillance, treatment, and response in the future.

Data Team at Government of Illinois :

Ms.Plaii
Analytics & Visualisation

Mr.Mhor
Code Hacker

Mr.Win
Scrum Master