



WEST NILE VIRUS

Group 4

Darion, Joseph,
Rebecca, Riche

CONTENTS

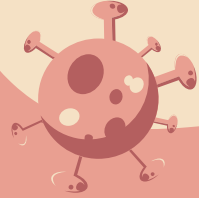
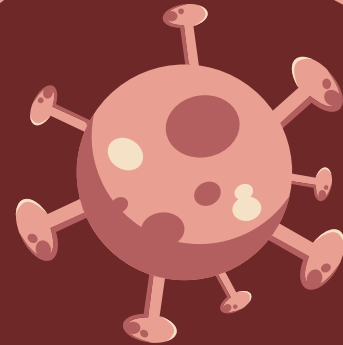
- Introduction
- Problem Statement
- Methodology
- Exploratory Data Analysis
- Feature Engineering
- Feature Importance
- Model Selection & Evaluation
- Cost-Benefit Analysis
- Conclusion & Recommendations





INTRODUCTION

Background: Due to the recent epidemic of West Nile Virus in Chicago, the Department of Public Health has set up a surveillance and control system to collect data on mosquito population over time



PROBLEM STATEMENT

Using the mosquito data collected over time, derive an effective plan to deploy pesticides throughout the city to control the mosquito population

Approach: Using ML/NN with Binary Classification, discover the important features for predicting the presence of WNV and develop a model to make predictions that the city of Chicago can use when it decides where to spray pesticides

WHAT IS WEST NILE VIRUS?

Transmission

WNV is transmitted to humans via the bite of infected mosquitoes.



01



02

Effects

While most people infected with WNV do not feel sick, about 1 in 5 people develop a fever and flu-like symptoms. Age group over 60 is the most vulnerable.

03

Incurable

Currently, there are no medication for the virus, prevention is the only way to stop the virus.

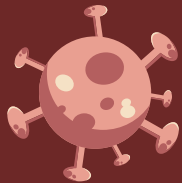


1,000

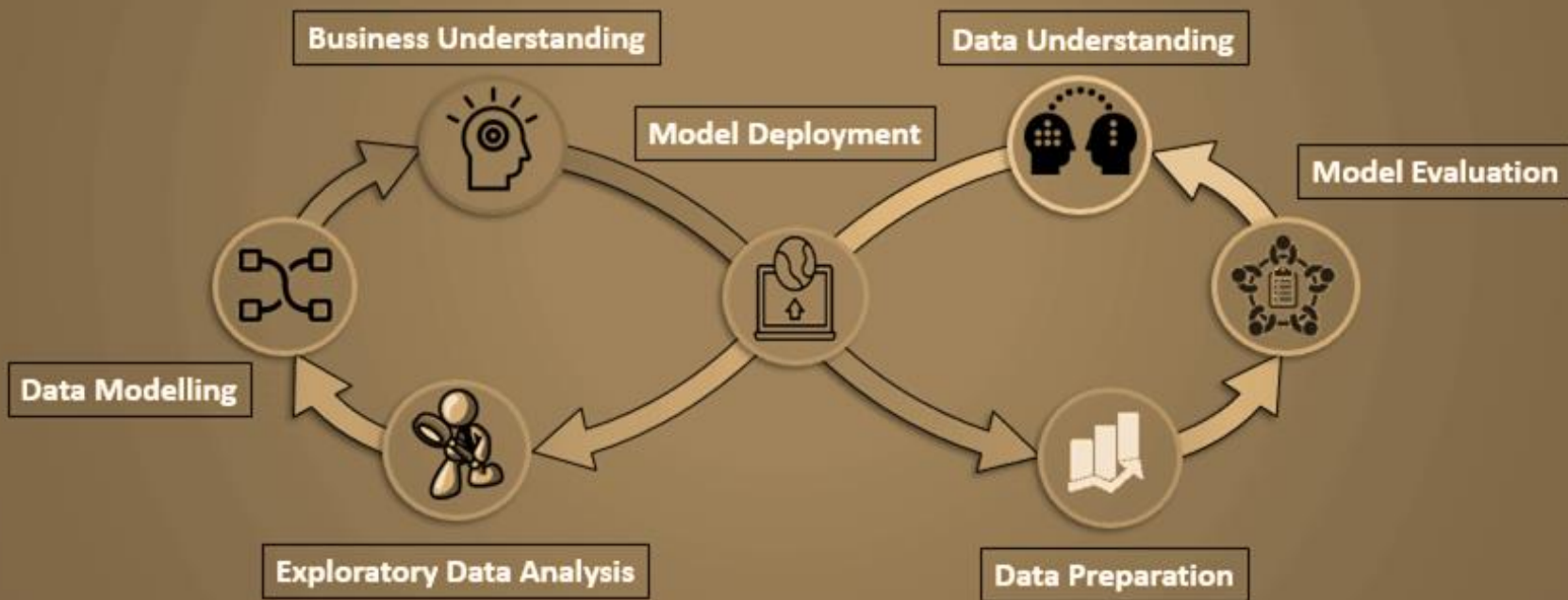
Cases detected in US 2019

60

Deaths
caused

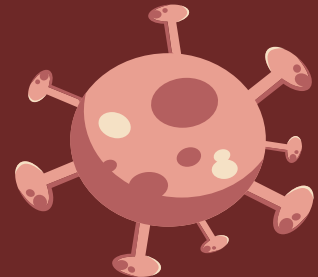


METHODOLOGY



WHAT IS IN THE DATA?

- Data used for the prediction model is from a Kaggle competition for West Nile Virus Prediction for the city of Chicago ([link](#))
- Data provided:
 - Main dataset contains testing results of mosquitos found in traps placed in Chicago. WNV test data has been split into a train and a test dataset.
 - Spray dataset contains the spray dates and locations of the spray.
 - Weather dataset contains the weather conditions of each day at the two weather stations.

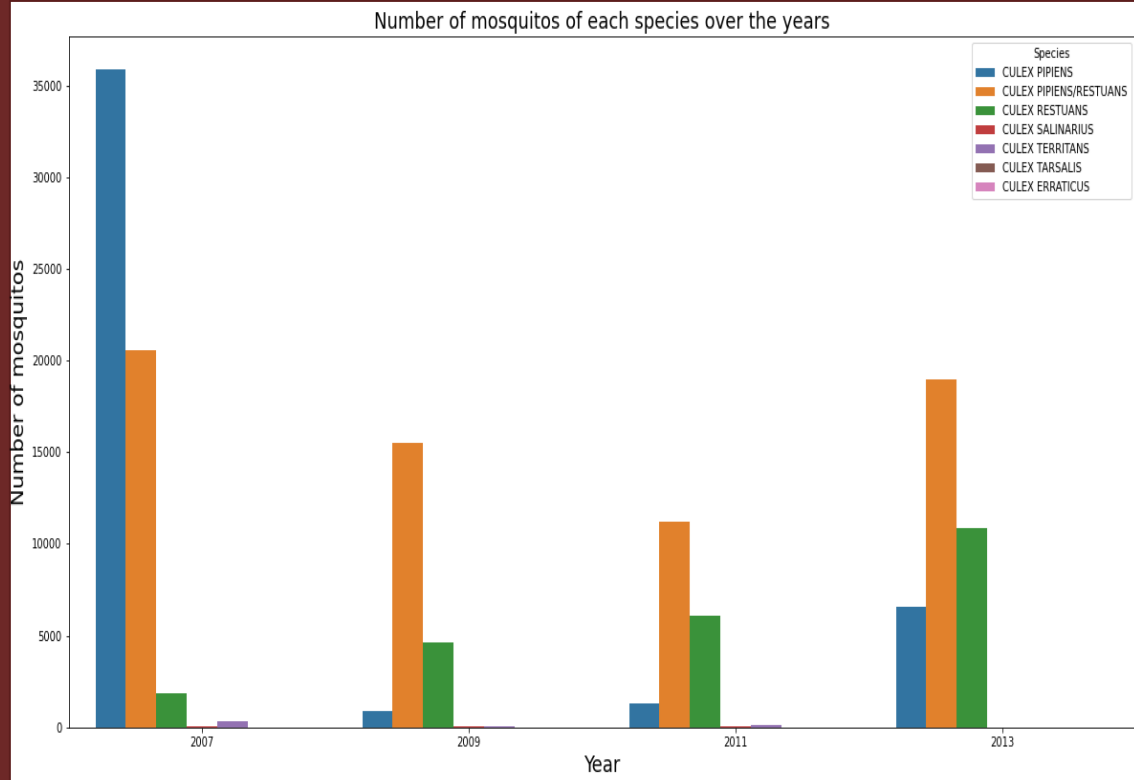


EDA: TRAIN DATA

Total number of mosquitoes dropped over the years, especially the Culex Pipiens species. It experienced a tremendous drop from 2007 to 2009.

Observed only two species that carry the Wnv virus, namely the Culex Pipiens and Culex Restuans.

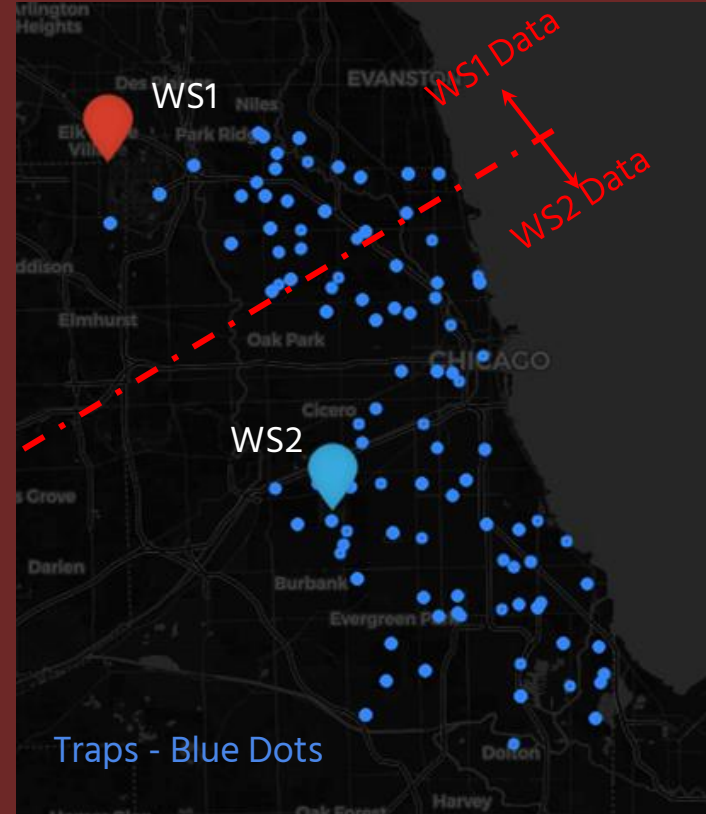
WNV Mosquitoes Species	
CULEX PIPPIENS/RESTUANS	262
CULEX PIPPIENS	240
CULEX RESTUANS	49



EDA: WEATHER STATIONS & TRAPS

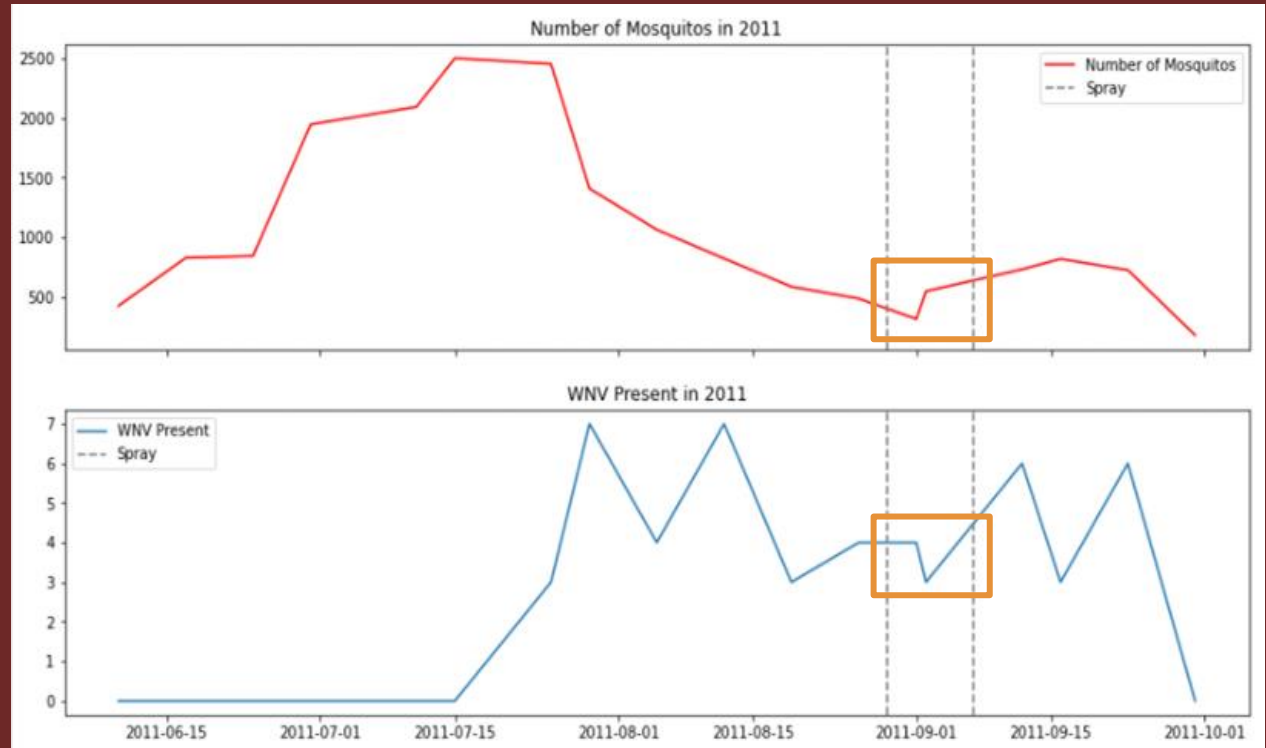
The map (on the right) shows the location of the traps and weather stations.

- Appended the weather data to the traps data.
- Based on the traps' proximity to the weather station.



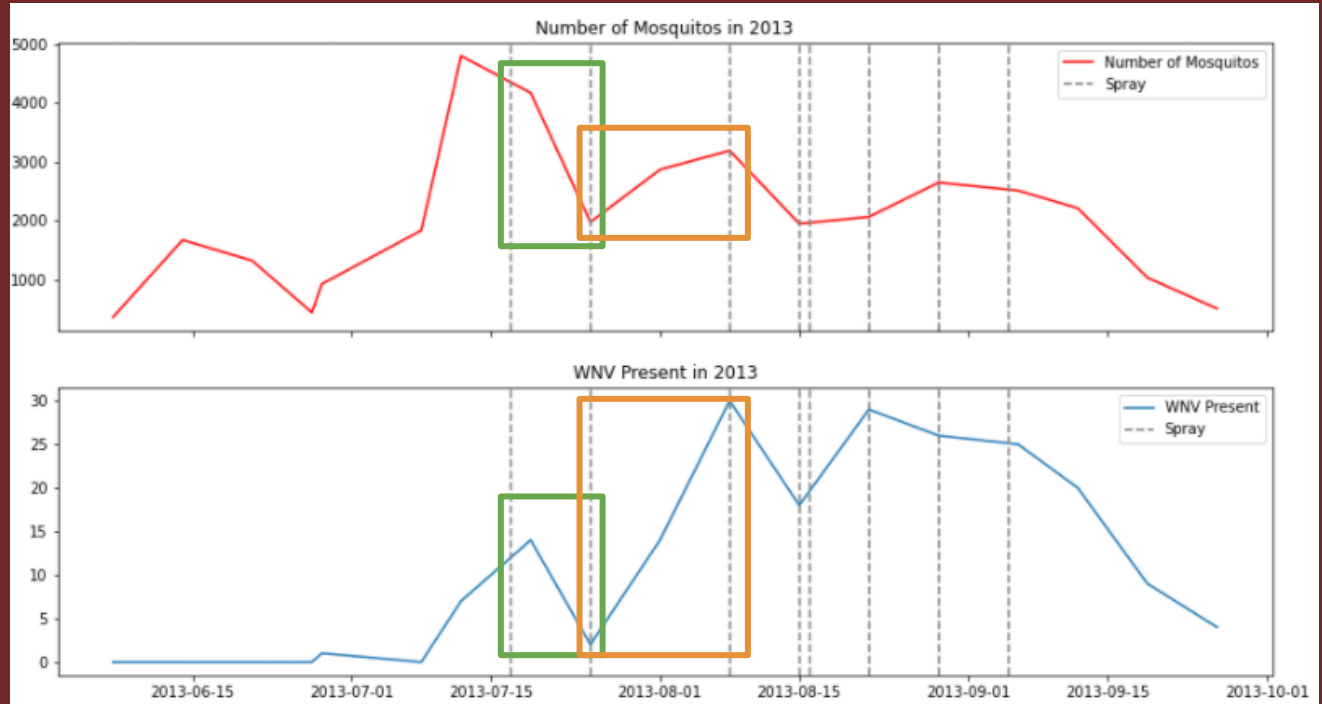
EDA: MOSQUITOES NUMBERS & WNV IN 2011

- Two sprays were done in 2011, with little observable effect on the mosquito population and WNV presence



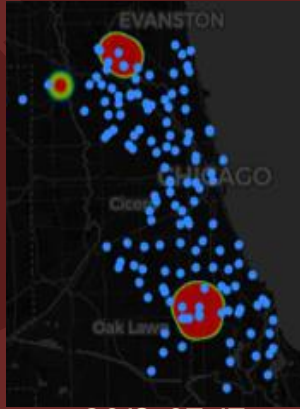
EDA: MOSQUITOES NUMBERS & WNV IN 2013

- 1st spray in 2013 had helped to reduce the number of mosquitos and WNV presence
- 2nd spray did not have the same effect, perhaps location of spray was not ideal

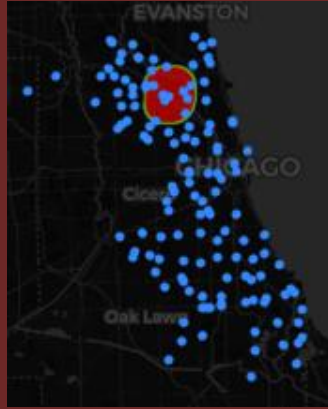


EDA: SPRAY & SPREAD OF WNV

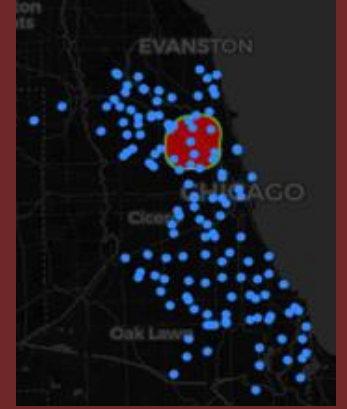
Spray Timeline



2013-07-17

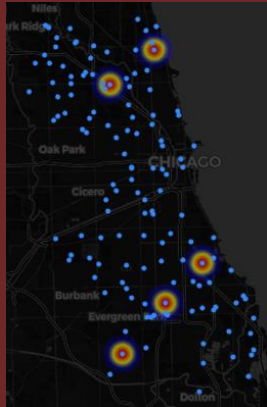


2013-07-25

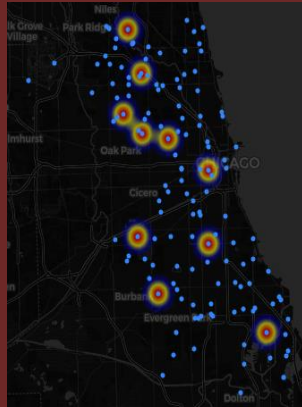


2013-08-08

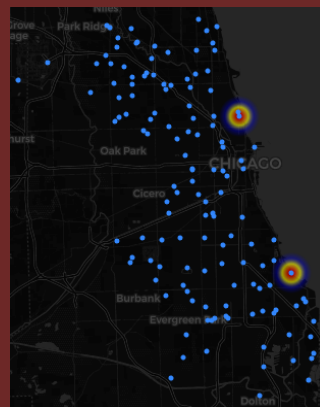
Mosquito Spread Timeline



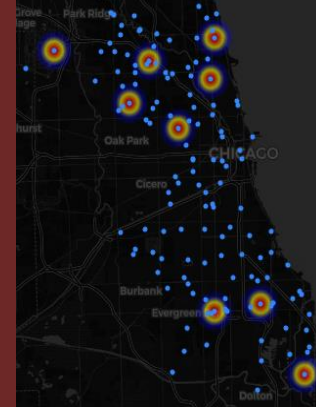
2013-07-12



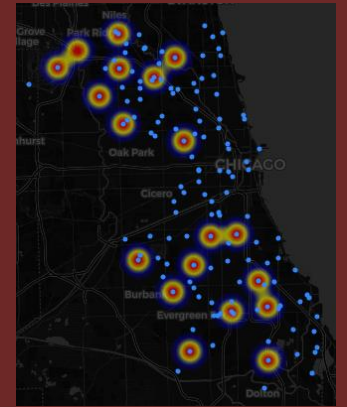
2013-07-19



2013-07-25



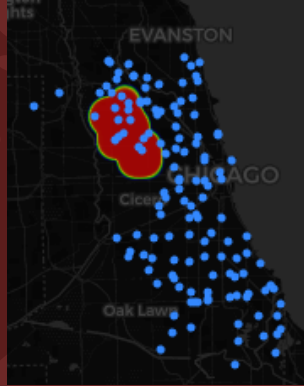
2013-08-01



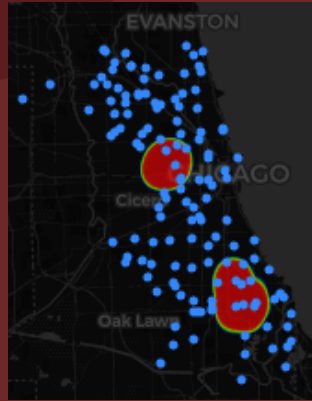
2013-08-08

EDA: SPRAY & SPREAD OF WNV

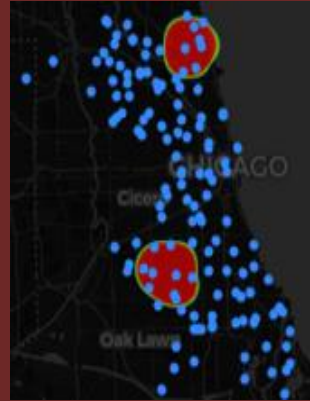
Spray Timeline



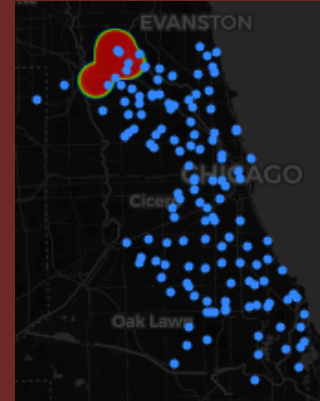
2013-08-15



2013-08-22

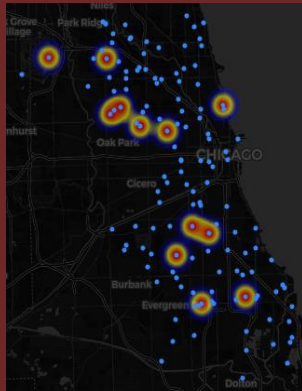


2013-08-29

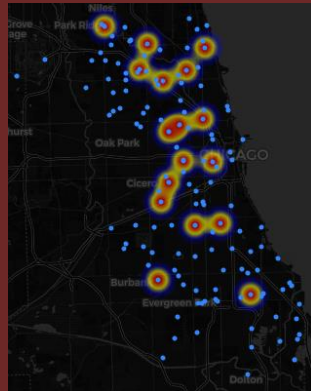


2013-09-05

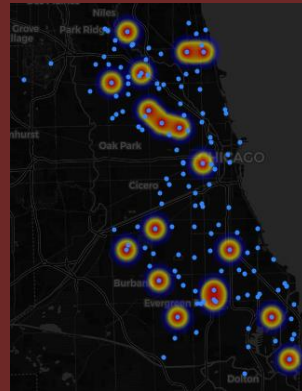
Mosquito Spread Timeline



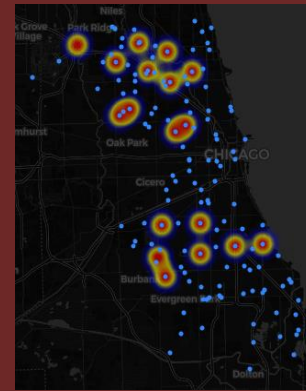
2013-08-15



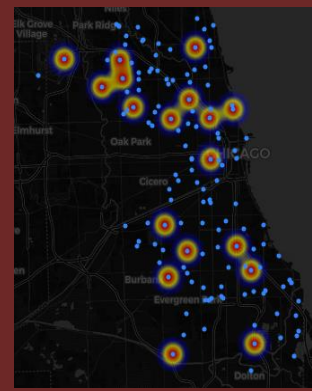
2013-08-22



2013-08-29



2013-09-06



2013-09-12

EDA: SUNSET & SUNRISE

This dataset contains the weather conditions data for Station 1 and 2. However, there are some missing data for some of the features.

Two of the features affected are Sunset and Sunrise timings. The missing data were from Station Two.

Using the Sun, Timezone, Latitude and Longitude, we managed to fill in the missing data for Station Two.

Another feature affected is Temperature Average. For that, we took the Max and Min Temperature's Mean to fill in the missing data.

```
# Extract sunrise/sunset values for Station 2
# This is the location of Station 2
latitude = 41.786
longitude = -87.752

sun = Sun(latitude, longitude)

central = timezone('US/Central')

sunrise_list = []
sunset_list = []
for value in weather['Date']:
    abd = value
    abd_sr = sun.get_local_sunrise_time(value, local_time_zone=central)
    abd_sr = abd_sr - abd_sr.dst()
    abd_ss = sun.get_local_sunset_time(value, local_time_zone=central)
    abd_ss = abd_ss - abd_ss.dst()
    sunrise_list.append(abd_sr.strftime('%H%M'))
    sunset_list.append(abd_ss.strftime('%H%M'))
```

Above code is for getting the Sunset and Sunrise timing for Station Two.

FEATURE ENGINEERING: WEATHER

01

High Humidity conditions favor mosquito activity, while low humidity suppresses activity and may even cause mortality

Humidity Feature

03

As the mosquitoes' breeding cycle might vary from 1 to 8 weeks, time-lagged weather features were created

Time-Lag Features

Code Sums

This feature contains the different weather conditions code. As not all weather conditions are present, we grouped the remaining conditions to 'Dry' or 'Wet'

02

Sun Hours

Mosquitos are most active from dusk to dawn. The number of hours with sunlight might help in the modelling

04

WEATHER FEATURES



Weather Condition Features

- 1) Average Temperature
- 2) Wet Bulb Temperature
- 3) Precipitation
- 4) Station Pressure
- 5) Sea Level
- 6) Resultant Wind Speed
- 7) Resultant Wind Direction
- 8) Average Wind Speed

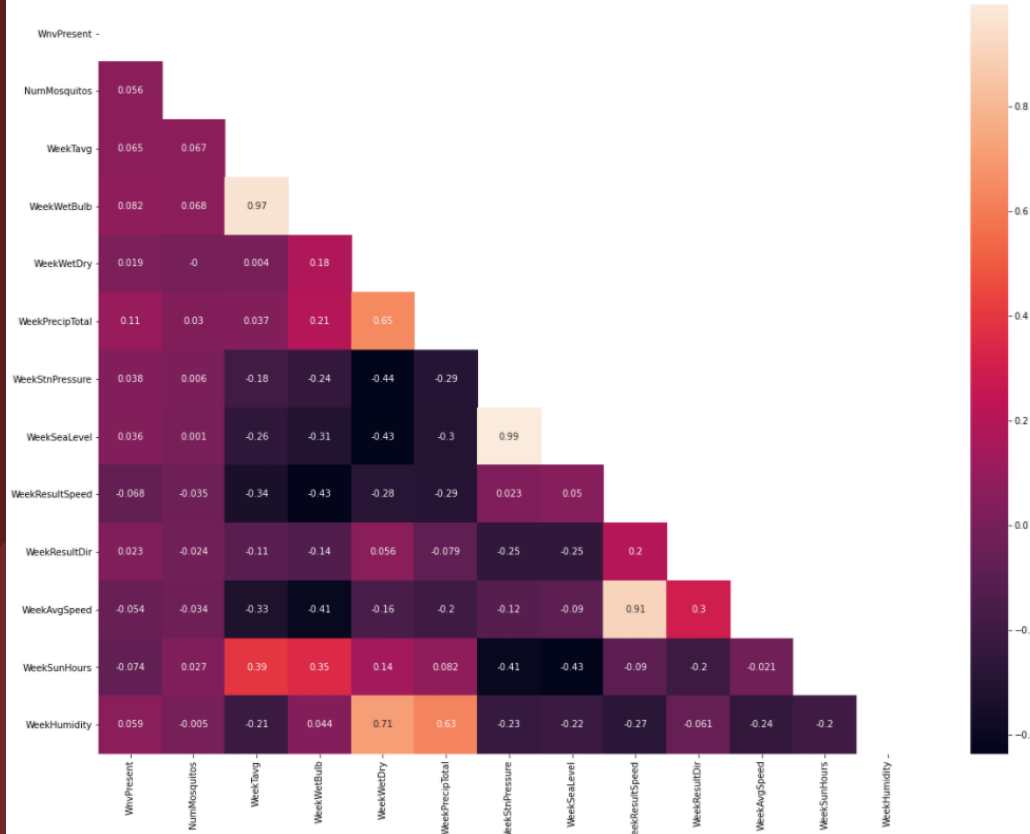
Engineered Weather Features

- 1) Sun Hours
- 2) Wet/Dry Day
- 3) Humidity

Time-Lag Features

As mosquitoes require one to eight weeks duration to be incubated and grow, we created 4 weekly time-lag average variable for each of the weather features.

EDA: WNV MOSQUITOES & WEATHER CONDITIONS

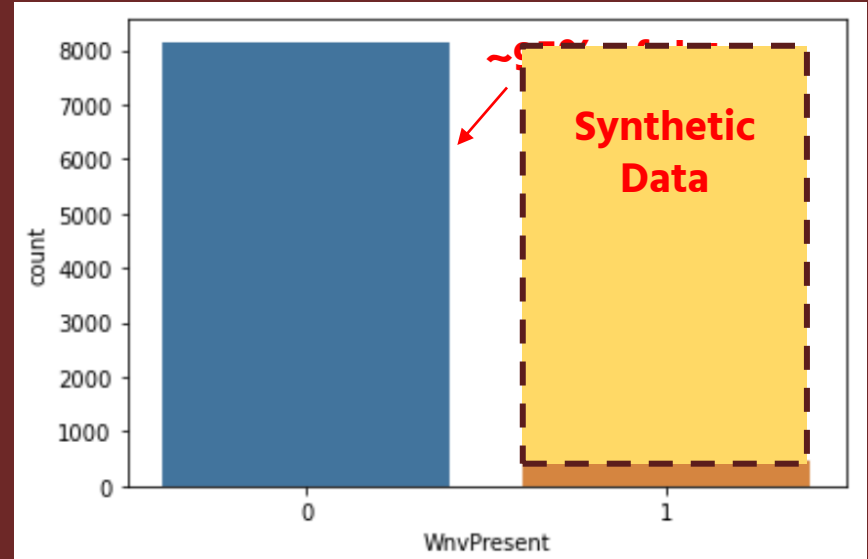


Weather Conditions

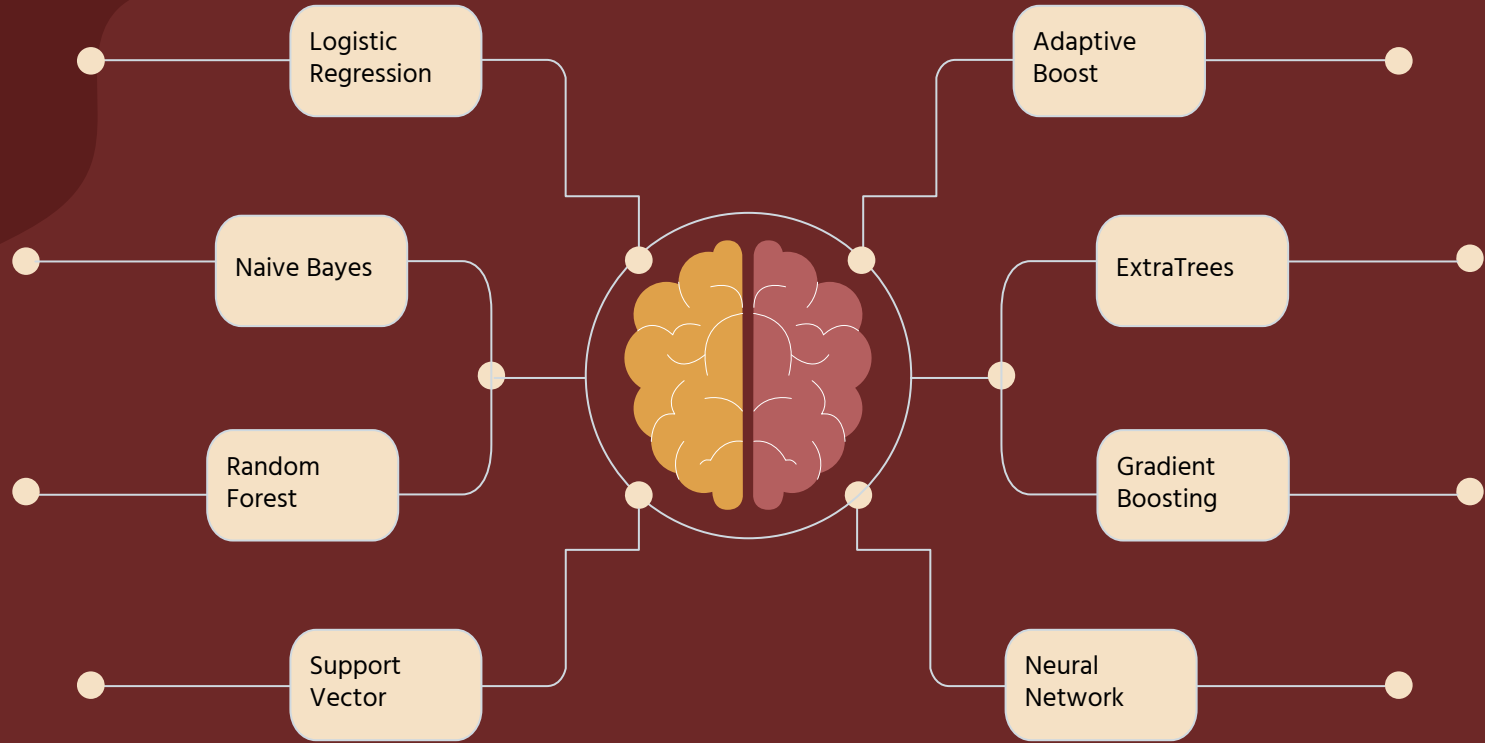
- As we study the different weather conditions, we will look specifically on five weather conditions which will affect the growth of mosquitoes.
- The five conditions are:
 - Wind
 - Temperature
 - Precipitation
 - Humidity
 - Sunlight Duration

PREPROCESSING

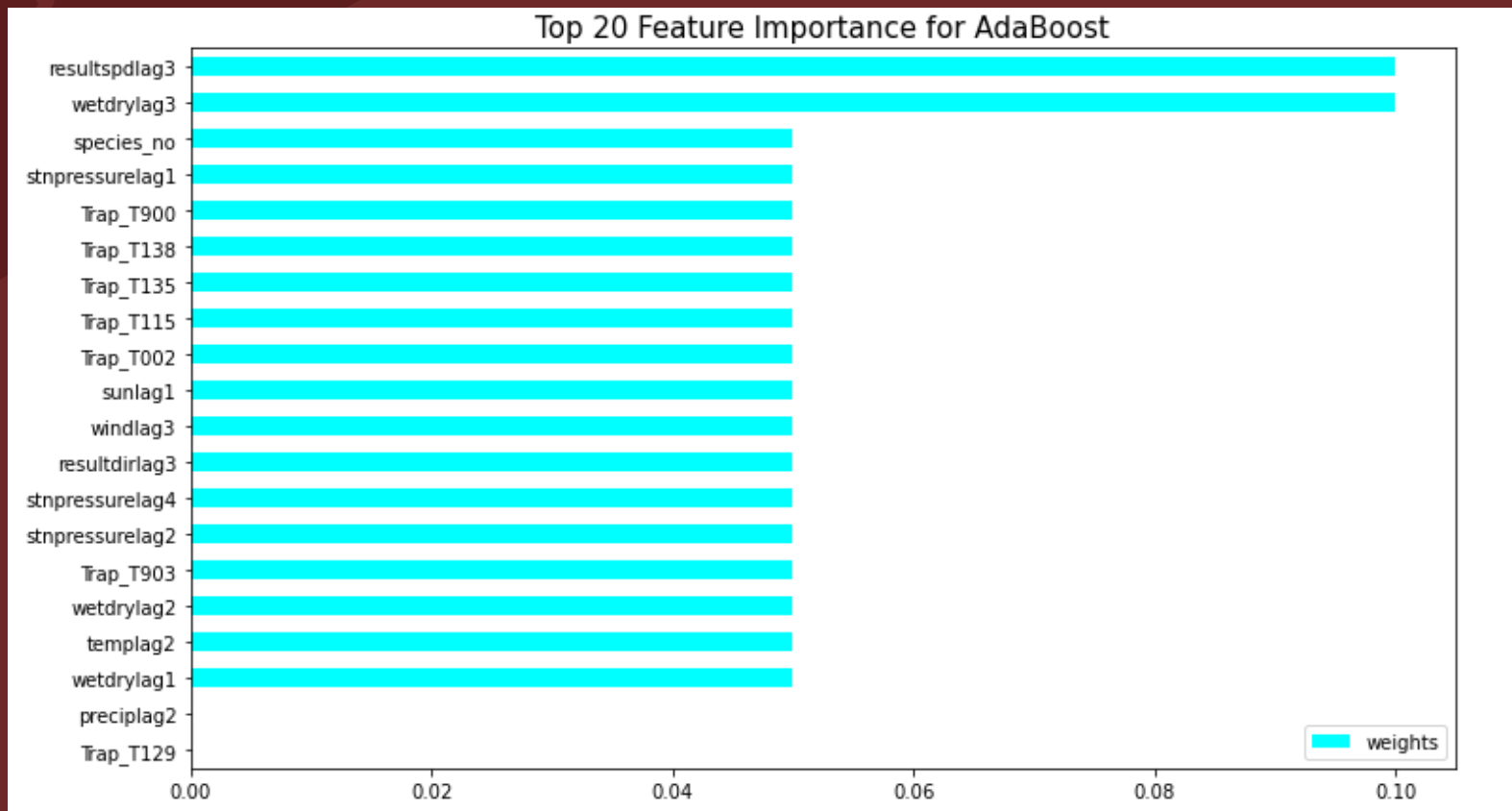
- Heavily imbalanced data in training set.
- Baseline prediction: 95% Wnv Not Present.
- Optimize AUC score instead of accuracy.
- Adopt over-sampling technique such as SMOTE.



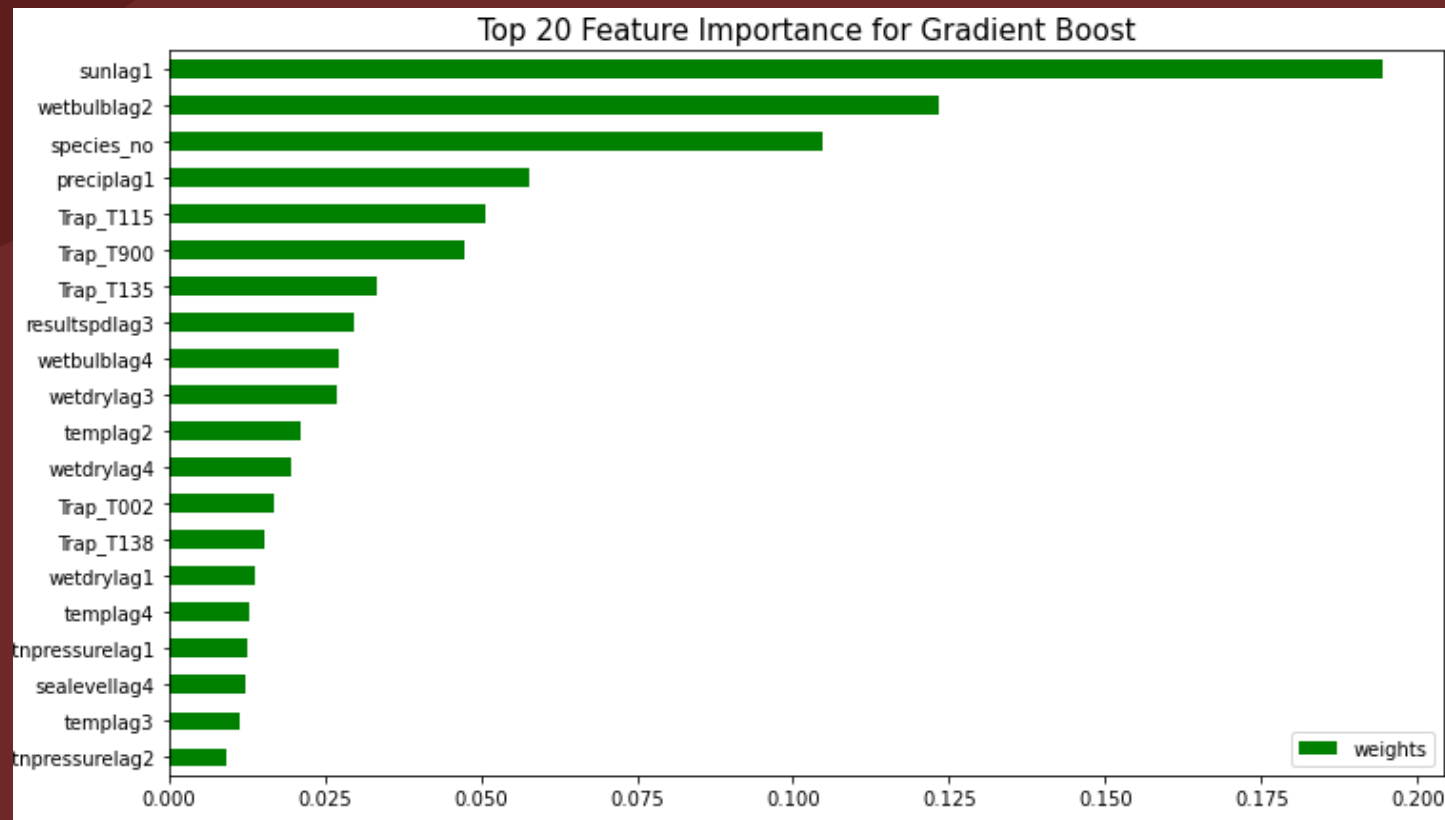
MODELLING



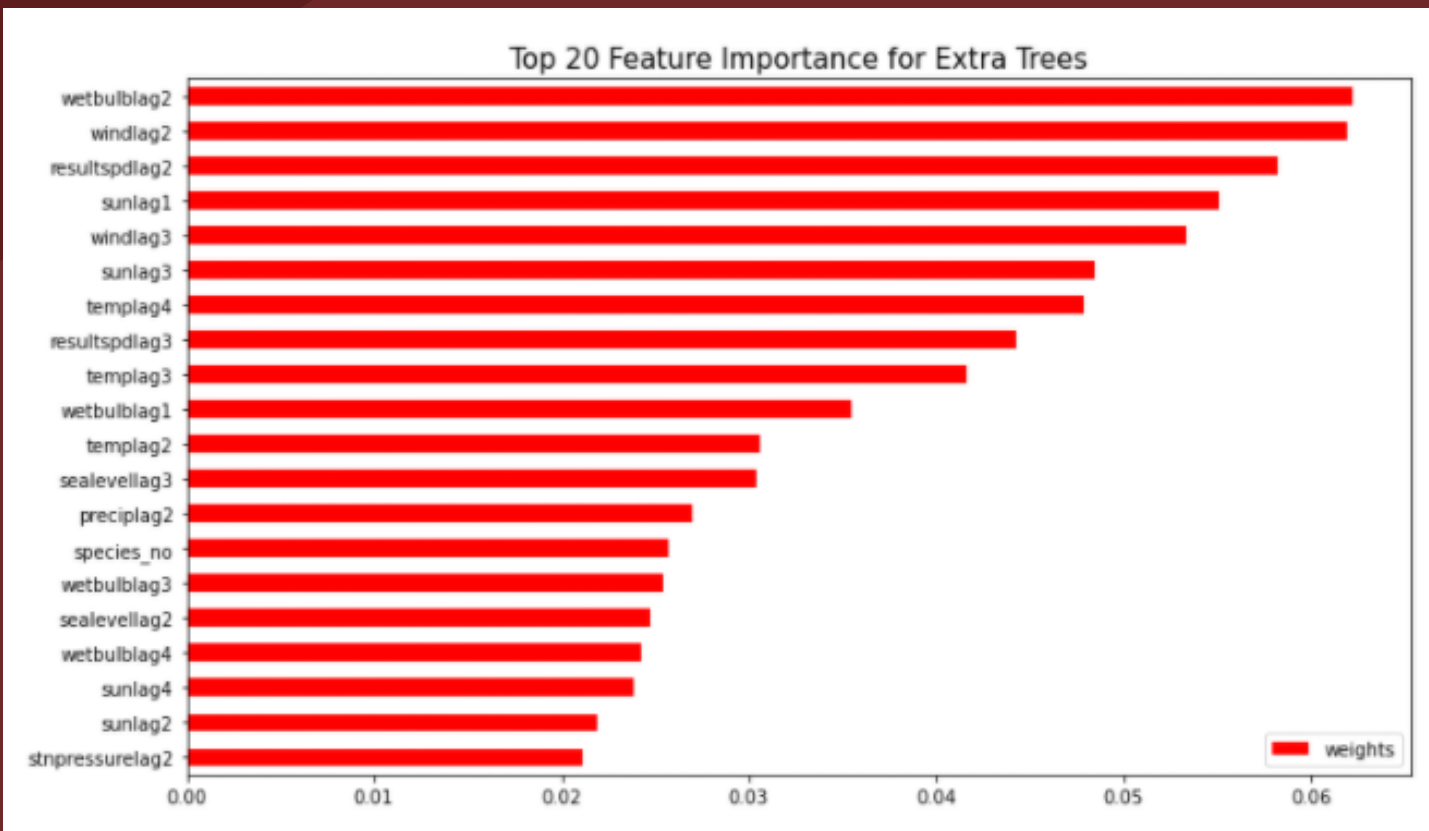
FEATURE IMPORTANCE: ADABOOST



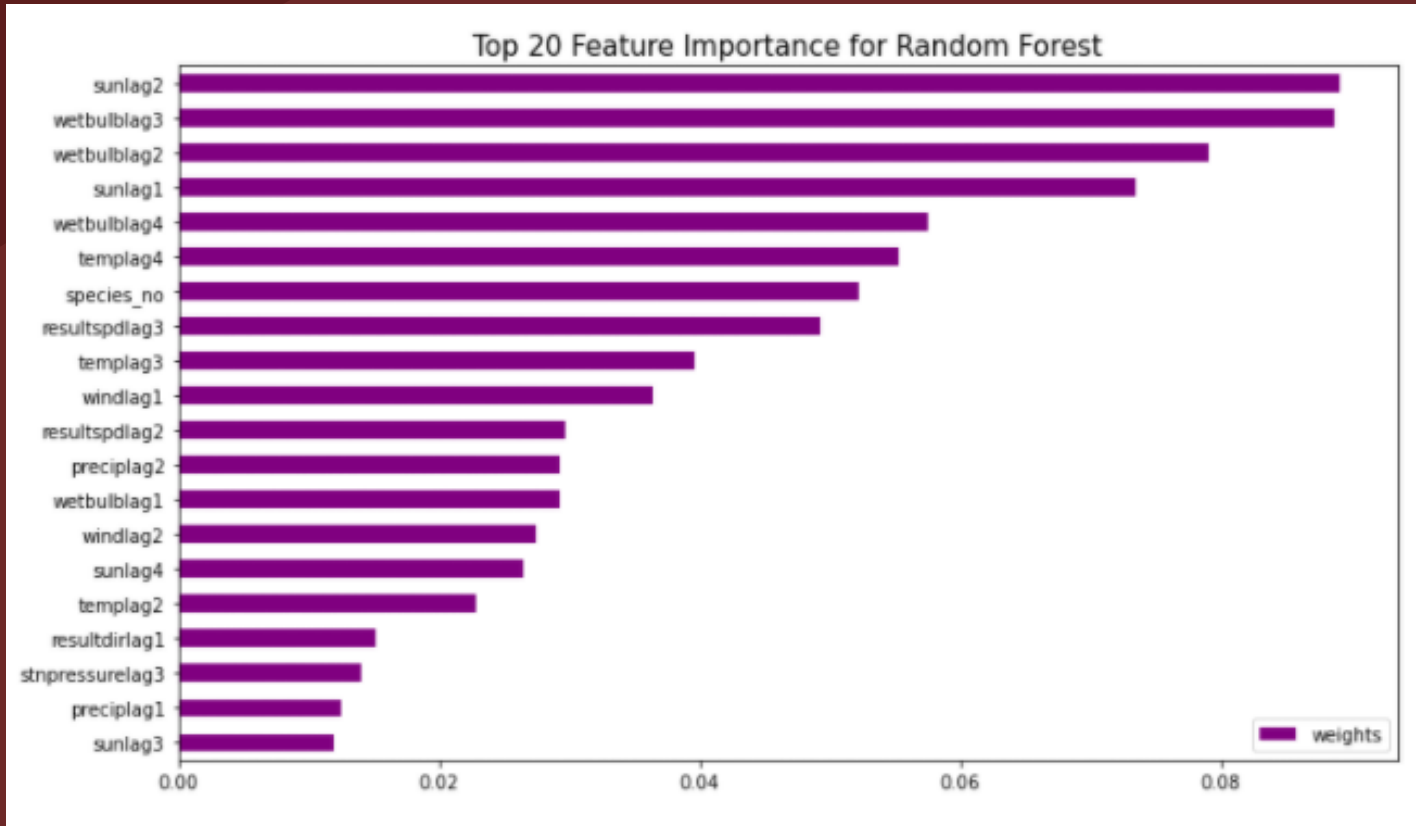
FEATURE IMPORTANCE: GRADIENT BOOSTING



FEATURE IMPORTANCE: EXTRA TREES

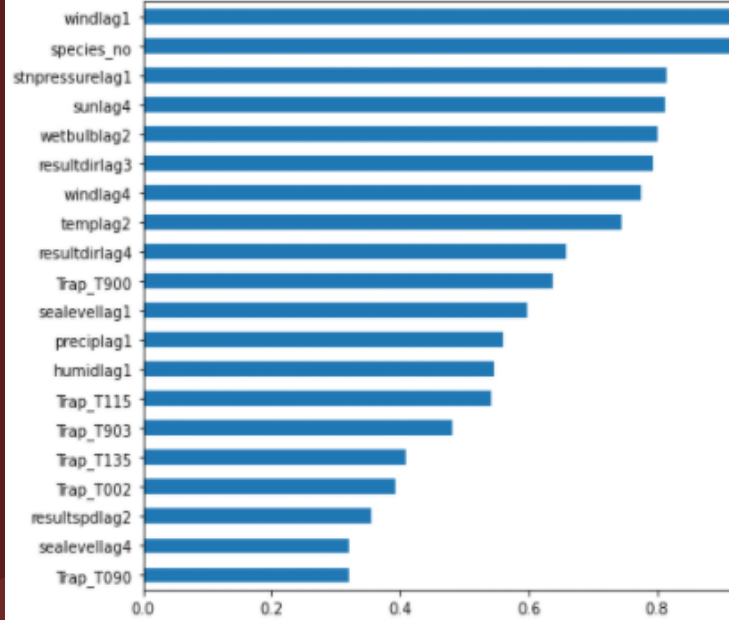


FEATURE IMPORTANCE: RANDOM FOREST

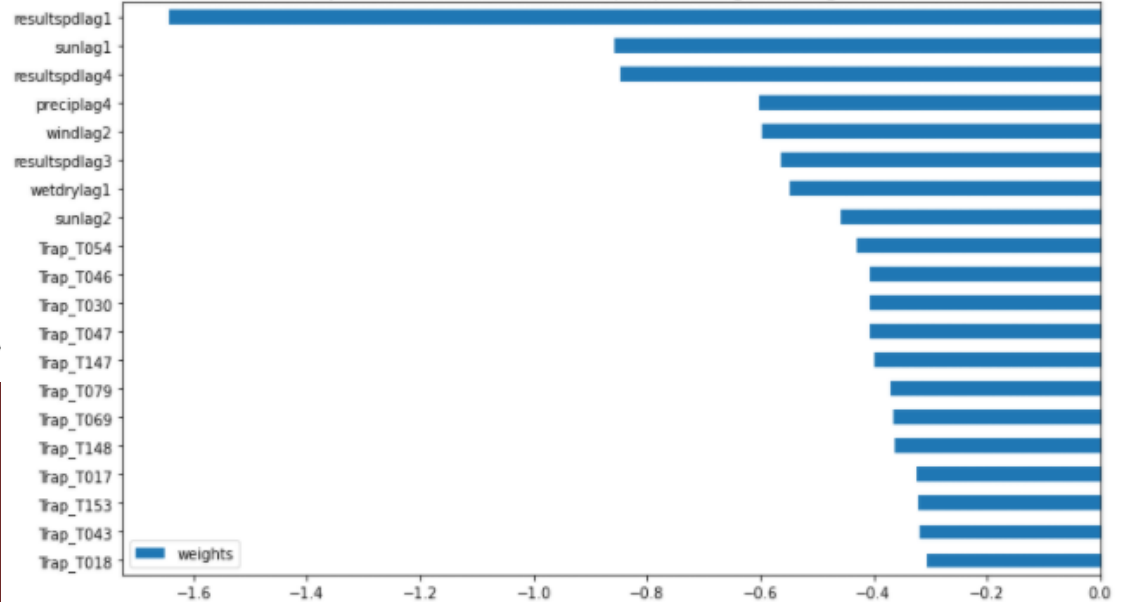


FEATURE IMPORTANCE: LOGISTIC REGRESSION

Top 20 Feature Coefficients (+ve) for Logistic Regression



Top 20 Feature Coefficients (-ve) for Logistic Regression

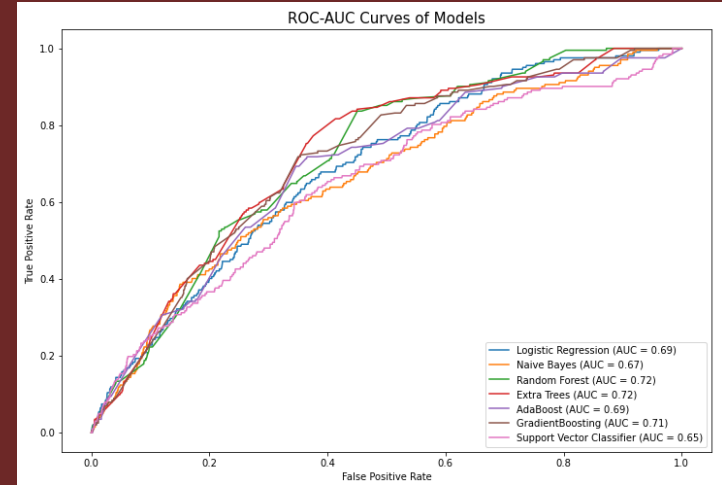


MODEL SELECTION

Model Type	Train AUC	Test AUC	F1-Score	Recall	Precision
AdaBoost	0.768642	0.584945	0.555765	0.584945	0.550529
Logistic Regression	0.832545	0.584520	0.556905	0.584520	0.551264
Support Vector Classifier	0.860891	0.574051	0.572417	0.574051	0.570921
GradientBoosting	0.676681	0.520733	0.520463	0.520733	0.582550
Random Forest	0.752317	0.508107	0.493737	0.508107	0.635600
Extra Trees	0.755211	0.501450	0.479825	0.501450	0.553212
Naive Bayes	0.749913	0.500000	0.475390	0.500000	0.453089

AdaBoost Model

- Highest Test AUC score.
- Highest Recall score.
- Lowest number of False Negatives.
- Some overfitting, but acceptable.



	Predicted Negative	Predicted Positive
Actual Negative	0.76	0.14
Actual Positive	0.06	0.03

Normalized Confusion Matrix – Positive = Wnv Present

ADDITIONAL MODEL: NEURAL NETWORK

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 182)	33306
dropout (Dropout)	(None, 182)	0
dense_1 (Dense)	(None, 128)	23424
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8256
dropout_2 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 32)	2080
dropout_3 (Dropout)	(None, 32)	0
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 1)	17
Total params: 67,611		
Trainable params: 67,611		
Non-trainable params: 0		

Results for NN Model

AUC: 0.5386
F1-Score: 0.1593
Recall: 0.1337
Precision: 0.1971

- Lower Test AUC score compared to AdaBoost.
- Low complexity neural network.
- Not easily interpretable.

COST-BENEFIT ANALYSIS

	City Level Spraying	Targeted Spraying
Direct Cost (\$0.92/acre) - Spray(truck/backpack)	\$1,653,467.04	\$175,086.72
Indirect Cost: - Hospitalization cost - Loss of production cost	-	\$459,617.3 (using yr 2020 as a benchmark due to falsely predicted negative)
Benefits	\$634,704.05 (If the control of mosquito is done regularly on a city level spraying, the chances of a west nile virus case is effectively 0. The amount saved from both the indirect and direct costs associated if we were to use the targeted spray can be used to fund the city level spraying)	\$1,018,762.99 (amount saved from excessive spraying which could be used to increase catch basins and maintenance of them, leading to increase in data for better analysis to increase accuracy and decrease west nile virus cases efficiently)
Benefit-cost ratio	0.38	1.61

CONCLUSIONS & RECOMMENDATIONS

Important features:

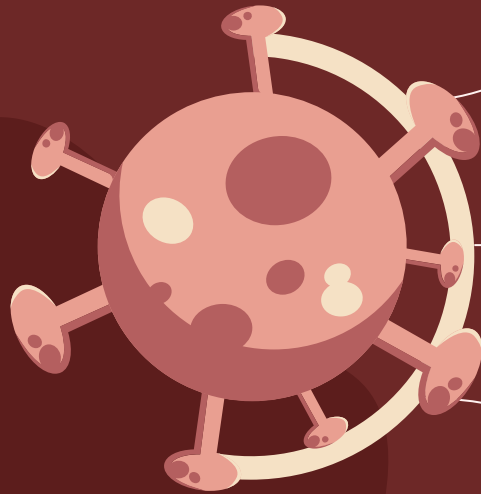
- Mosquito species - some species are more competent vectors of Wnv
- Duration of the day - longer the Sun is out, the lower the mosquito activity
- Temperature - lower the temperature, lower the mosquito activity
- Wind - higher the wind speed, lower the mosquito activity
- Humidity - lower the humidity, lower the mosquito activity

Do

1. Use insect repellent that contains DEET, Picaridin, IR3535 or Oil of Lemon Eucalyptus
2. Eliminate stagnant water
3. Wear long pants, long-sleeved shirts at night



MOVING FORWARD



01

Demographics

Human demographics within the area affects probability of Wnv.

02

Geography

Landscapes affect the mosquitoes' livelihoods and activity.

03

Mosquito Species

Study other species that could possibly spread the virus.

THANKS

Do you have any questions?



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**



Sources

- 1) Prevention Measures - [\[https://www.chicago.gov/city/en/depts/cdph/provdrs/health_protection_and_response/news/2020/july/cdph-conducts-a-comprehensive-mosquito-surveillance-and-control.html\]](https://www.chicago.gov/city/en/depts/cdph/provdrs/health_protection_and_response/news/2020/july/cdph-conducts-a-comprehensive-mosquito-surveillance-and-control.html)
- 2) Life Cycle of Mosquitoes - [\[https://www.mosquito.org/page/lifecycle\]](https://www.mosquito.org/page/lifecycle)
- 3) Mosquito Treatment Duration - [\[https://www.callnorthwest.com/2019/05/how-long-does-a-mosquito-treatment-last/\]](https://www.callnorthwest.com/2019/05/how-long-does-a-mosquito-treatment-last/)
- 4) Mosquito Spray Information - <https://www.montcopa.org/DocumentCenter/View/8932/FAQMosquitoSpraying?bidId=>
- 5) Calculation for Sunrise/Sunset for Station two - [\[https://stackoverflow.com/questions/38986527/sunrise-and-sunset-time-in-python/38986561\]](https://stackoverflow.com/questions/38986527/sunrise-and-sunset-time-in-python/38986561)
- 6) Daylight Savings For Calculations - [\[\(https://www.timeanddate.com/time/zone/usa/chicago\)\]](https://www.timeanddate.com/time/zone/usa/chicago)