West Nile Virus Prevention

Group members:
Elise, Fadhil, Prerak, Shandeep

01

INTRODUCTION

02

EDA & MODELS

03

COST - BENEFIT ANALYSIS

04

DEPLOYMENT

05

LIMITATIONS AND FUTURE STEPS

01

INTRODUCTION

West Nile Virus

West Nile Virus is a mosquito-borne disease that has plagued continental United States since 1999. About 1 in 150 infected people develop a serious illness, which could lead to fatality. Currently, there is no vaccine available to treat WNV.

Introduction

PROBLEM STATEMENT

- Predict locations with potential of being infested with mosquitoes carrying WNV

- Explore factors contributing to the growth and spread of the virus through mosquitoes

CONTEXT

The Chicago Department of Public Health (CDPH) has set up a surveillance and control system to trap mosquitoes and test for the presence of WNV. Current research highlights that increases in WNV infection rates in mosquito populations could serve as an indicator of a potential outbreak in humans. Aggressive and timely interventions could therefore curb the spread of the virus.

SCOPE

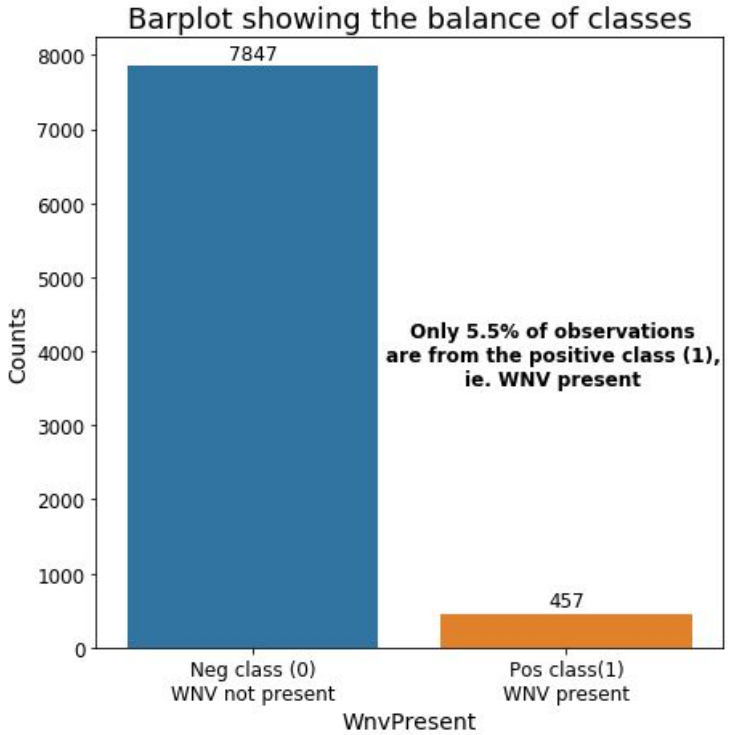
- Derive a plan to deploy pesticides throughout the city of Chicago

- Employ various modelling techniques to help guide decisions in order to maximise pesticide effectiveness and minimise costs

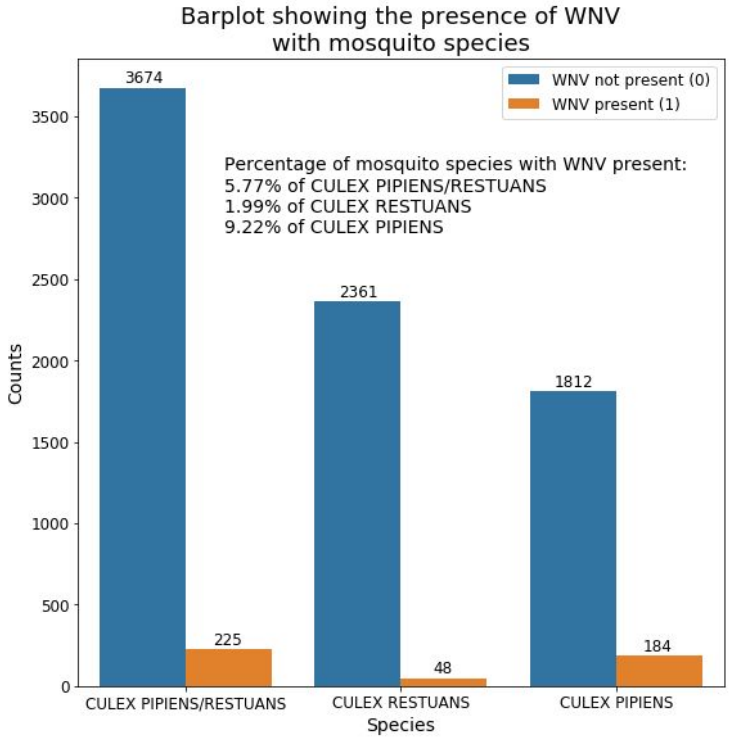
02

EDA & MODELS

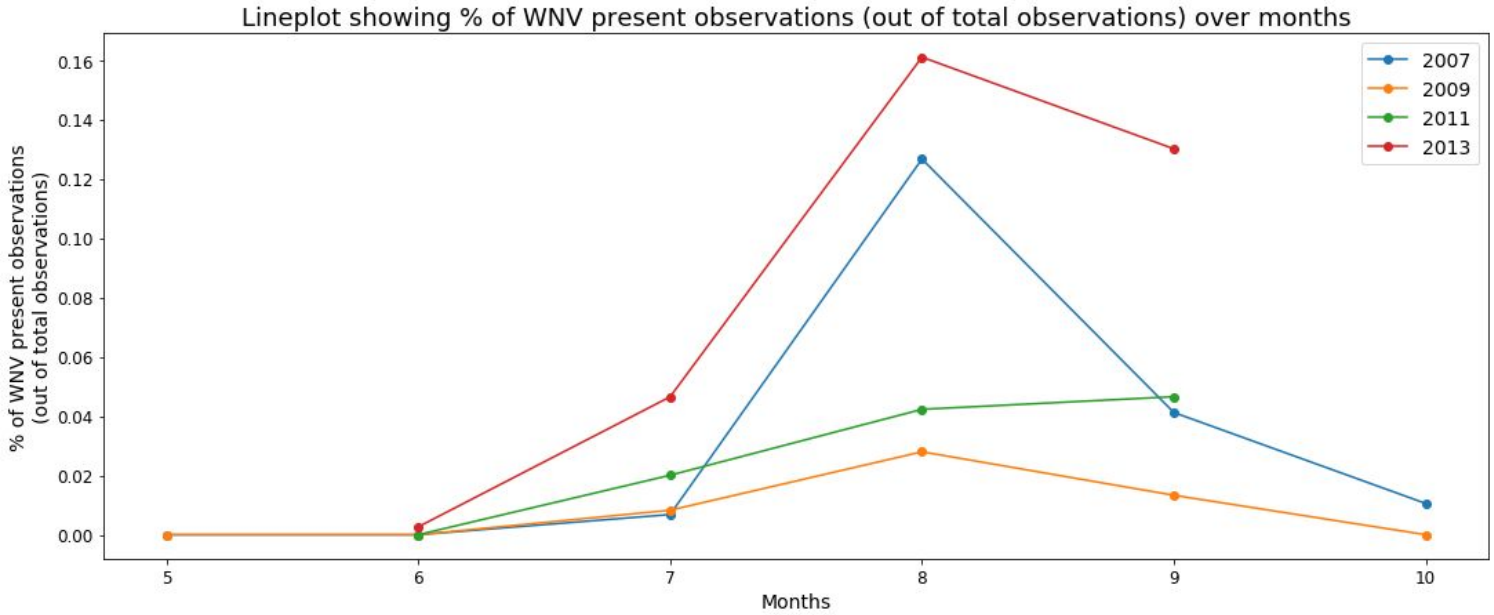
IMBALANCED CLASSES



CULPRIT MOSQUITO SPECIES



WNV BY MONTHS



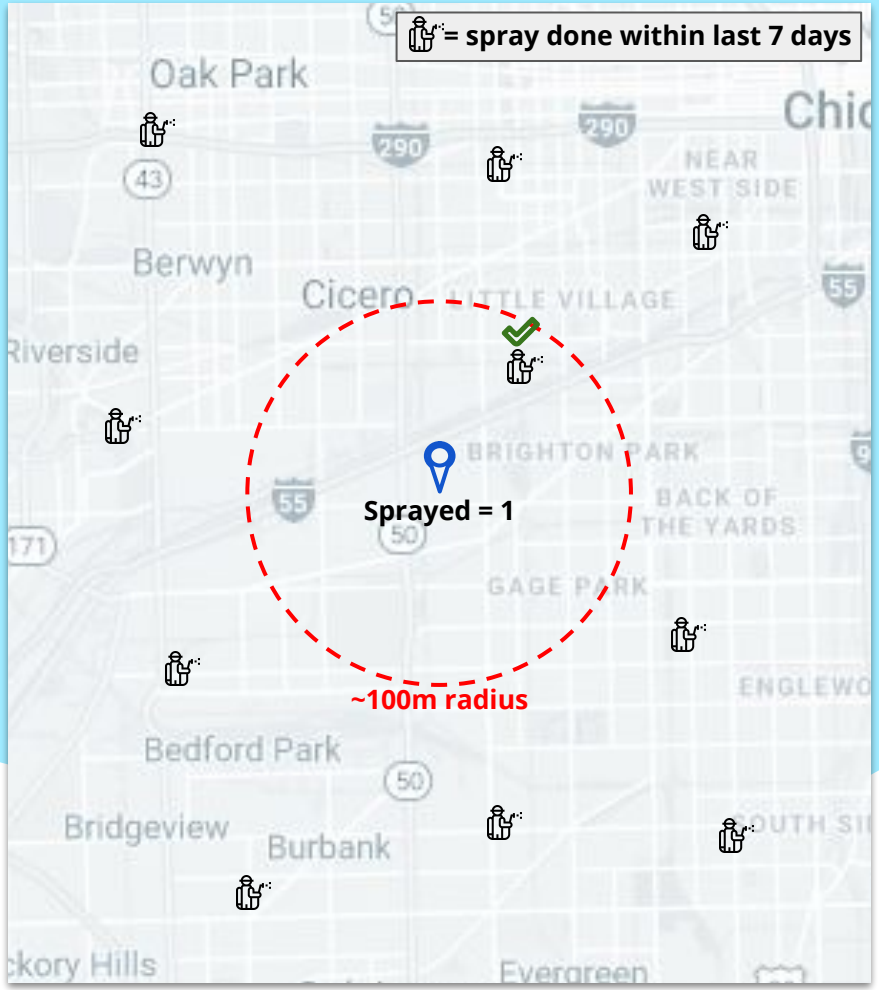
FEATURE ENGINEERING

Sprayed

Assumptions made:

1. Spraying is effective for up to 7 days.
2. Spraying is effective in a ~100m radius from the location of spraying.

Sprayed = 1 if assumptions are satisfied, 0 otherwise.

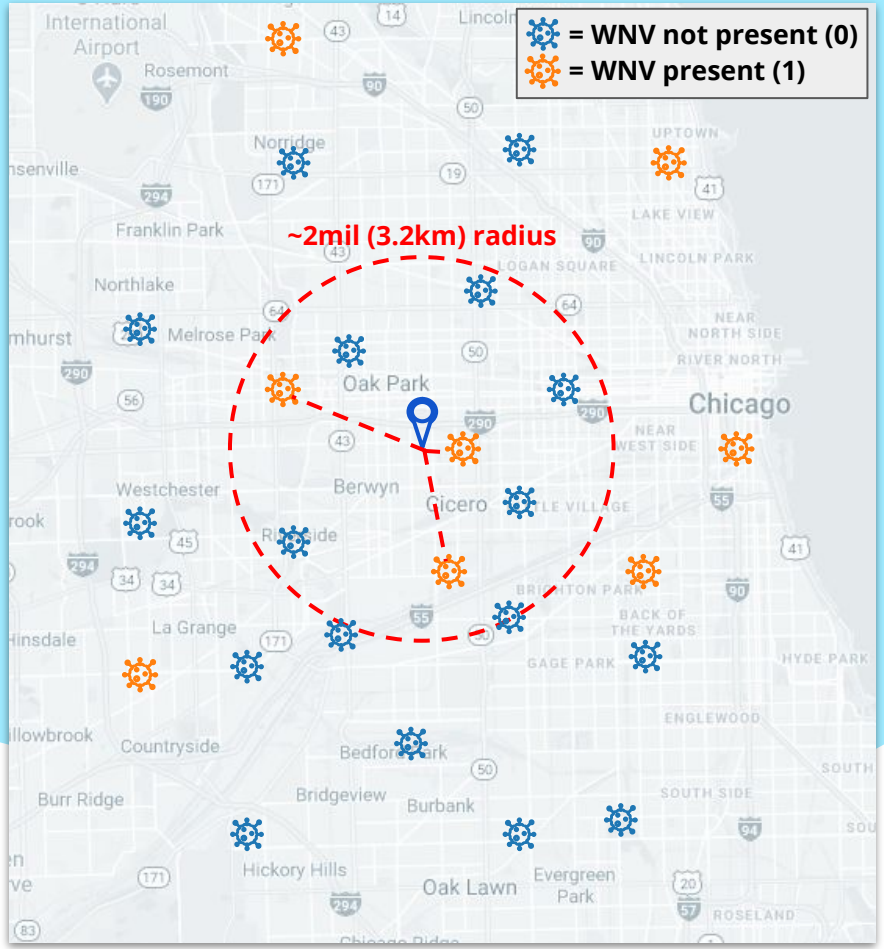


FEATURE ENGINEERING

Average Positive Distance

- Average distance of previously positively tested mosquito traps from each individual trap being tested.
- Accounts for spatial correlation between mosquito traps.
- The closer these previously positively tested traps are, the higher the chances that mosquitoes could have flown over and brought the virus with them (and vice-versa).

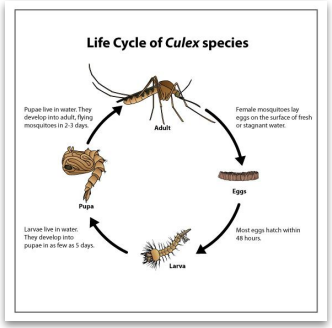
- Assumption: mosquitoes fly ~2mil (3.2km) on average.

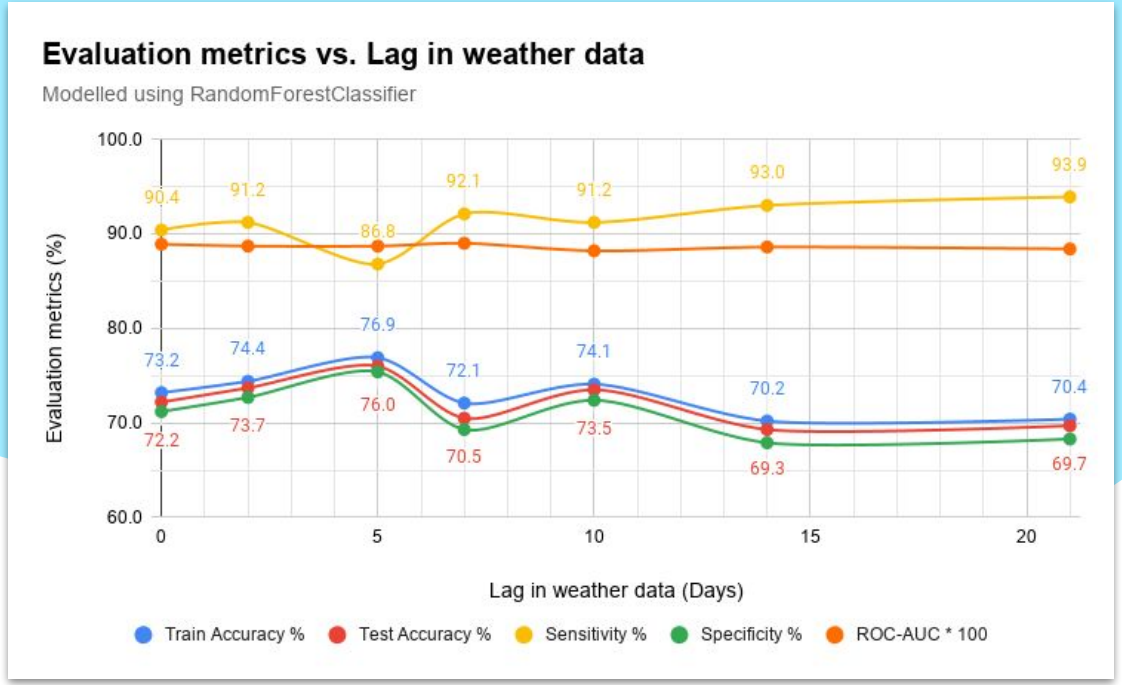


WEATHER DATA LAG

Lag = 10 days

- If today's weather is ideal for mosquito breeding, they lay more eggs than usual.
- Eggs will grow into adult mosquitoes over 10 days.
- Only after 10 days will we see an increase in the presence of WNV in the mosquito traps.





CLASSIFICATION MODELLING SUMMARY

Classifier	Class Balancing Technique	Accuracy on training set	Accuracy on testing set	Sensitivity (Recall)	Specificity	ROC-AUC	Kaggle ROC-AUC
LogisticRegression	SMOTE	70.7%	71.2%	85.1%	70.4%	0.858	0.661
RandomForestClassifier	SMOTE	77.8%	77.5%	85.1%	77.1%	0.872	0.706
SVC	SMOTE	70.3%	69.4%	86.0%	68.4%	0.844	0.675
GradientBoostingClassifier	SMOTE	78.0%	76.5%	83.3%	76.1%	0.871	0.705

CLASSIFICATION MODELLING SUMMARY

Classifier	Class Balancing Technique	Accuracy on training set	Accuracy on testing set	Sensitivity (Recall)	Specificity	ROC-AUC	Kaggle ROC-AUC
LogisticRegression	SMOTE	70.7%	71.2%	85.1%	70.4%	0.858	0.661
RandomForestClassifier	SMOTE	77.8%	77.5%	85.1%	77.1%	0.872	0.706
SVC	SMOTE	70.3%	69.4%	86.0%	68.4%	0.844	0.675
GradientBoostingClassifier	SMOTE	78.0%	76.5%	83.3%	76.1%	0.871	0.705

CLASSIFICATION MODELLING SUMMARY

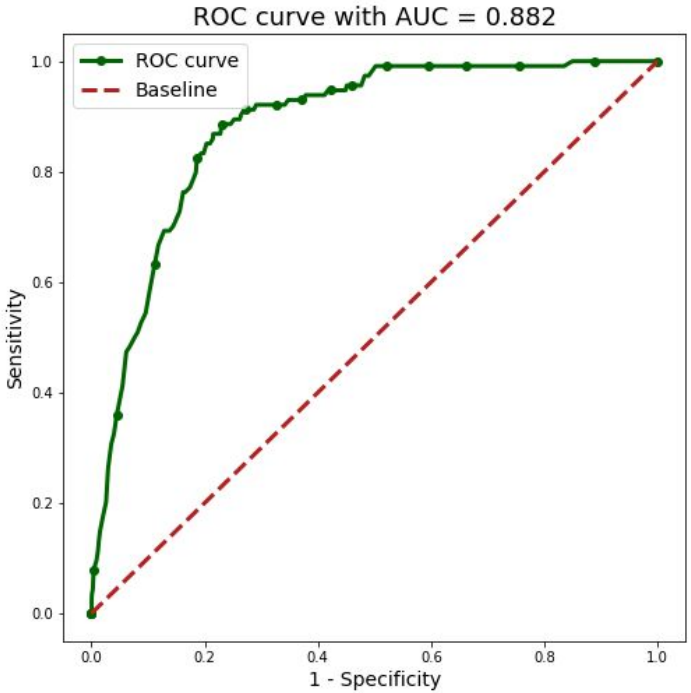
Classifier	Class Balancing Technique	Accuracy on training set	Accuracy on testing set	Sensitivity (Recall)	Specificity	ROC-AUC	Kaggle ROC-AUC
LogisticRegression	SMOTE	70.7%	71.2%	85.1%	70.4%	0.858	0.661
RandomForestClassifier	SMOTE	77.8%	77.5%	85.1%	77.1%	0.872	0.706
SVC	SMOTE	70.3%	69.4%	86.0%	68.4%	0.844	0.675
GradientBoostingClassifier	SMOTE	78.0%	76.5%	83.3%	76.1%	0.871	0.705
RandomForestClassifier	ADASYN	77.1%	76.2%	84.2%	75.7%	0.872	0.713
RandomForestClassifier	ClusterCentroids	62.7%	62.4%	92.1%	60.7%	0.848	0.707
RandomForestClassifier	class_weight = ‘balanced_subsample’	74.1%	73.5%	91.2%	72.4%	0.882	0.717

CLASSIFICATION MODELLING SUMMARY

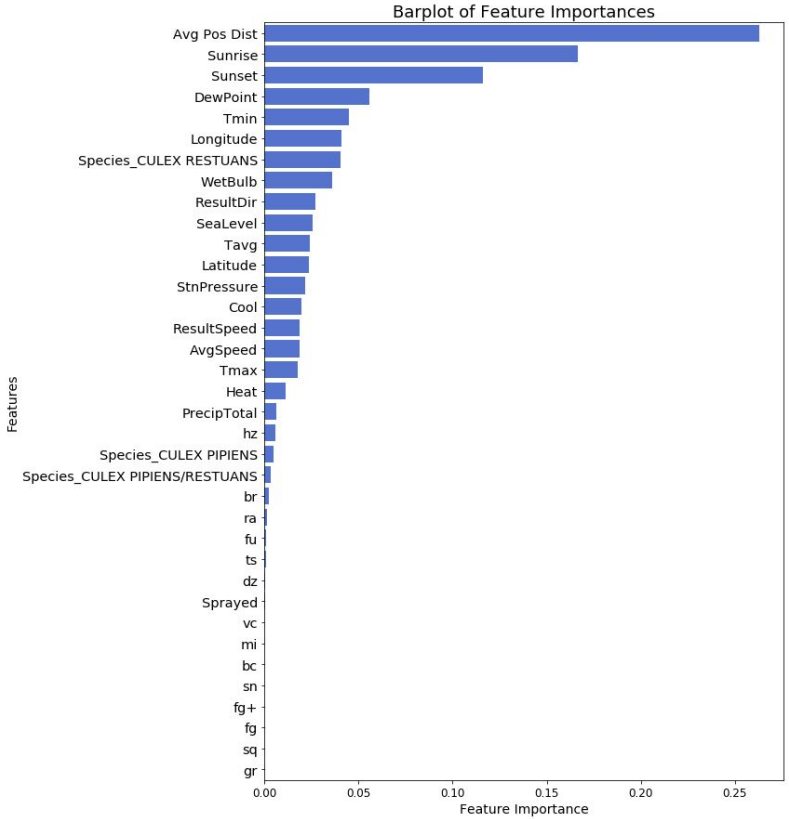
Classifier	Class Balancing Technique	Accuracy on training set	Accuracy on testing set	Sensitivity (Recall)	Specificity	ROC-AUC	Kaggle ROC-AUC
LogisticRegression	SMOTE	70.7%	71.2%	85.1%	70.4%	0.858	0.661
RandomForestClassifier	SMOTE	77.8%	77.5%	85.1%	77.1%	0.872	0.706
SVC	SMOTE	70.3%	69.4%	86.0%	68.4%	0.844	0.675
GradientBoostingClassifier	SMOTE	78.0%	76.5%	83.3%	76.1%	0.871	0.705
RandomForestClassifier	ADASYN	77.1%	76.2%	84.2%	75.7%	0.872	0.713
RandomForestClassifier	ClusterCentroids	62.7%	62.4%	92.1%	60.7%	0.848	0.707
RandomForestClassifier	class_weight = 'balanced_subsample'	74.1%	73.5%	91.2%	72.4%	0.882	0.717

CLASSIFICATION MODELLING SUMMARY

Confusion Matrix	Actual WNV present (1)	Actual WNV not present (0)
Predicted WNV present (1)	104 (TP)	541 (FP)
Predicted WNV not present (0)	10 (FN)	1421 (TN)



FEATURE IMPORTANCE



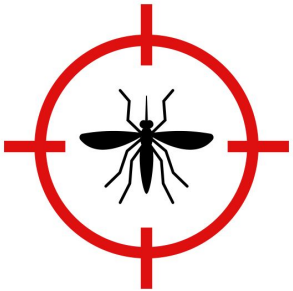
	RFC without Avg Pos Dist	RFC with Avg Pos Dist
Accuracy on training set	66.2%	73.2%
Accuracy on testing set	65.8%	72.2%
Sensitivity (Recall)	90.4%	90.4%
Specificity	64.3%	71.2%
ROC-AUC	0.855	0.889
Kaggle ROC-AUC	0.710	0.722

03

COST - BENEFIT ANALYSIS

Cost of Spraying

The insecticide that we recommend is Zenivex(tested and proven pesticide).



- Price/fluid ounce is $22,948.75 / 35200 = $0.65

- 1.5 oz cost/acre = $0.65 * 1.5 = $0.97/acre

- spray location in sqm = 3.14 * (100m**2) = 31,400sqm
- 1 acre = 4046.86sqm

1 spray location = $0.97 * (31400/4046.86) = **$7.52**

On average, based on our model prediction

Total Spraying Cost:

$52,767 per year

(excl. overhead costs)

*The calculations above are based on our model built on 8304 readings.

Benefit of Spraying

Based on Sacramento data, 163 reported cases incurred ≈ USD2.28 million in medical treatment.
Assuming the same infection rate and population density in Chicago,

There would be **511** reported cases of WNV.

This would incur **$573,342** in medical treatment

*given the best case scenario of all reported cases being WNF

$573,342 > $52,767

Medical Costs per Person:

- WNND:
  - Mild Cases: **$76,31**
  - Severe Cases: **$33,143**
  - Productivity Loss: **$10,800**
- WNF:
  - Medical Cost: **$167**
  - Productivity Loss: **$955**

(Source: NCBI
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3322011/)

Prevention is better than Cure!

Steps that can be taken to curb the virus:

- Education and Programs on West Nile Virus
- Campaigns to reduce and curb mosquito population
- Monetary enforcement (i.e fines) against mosquito breeding in households
- Utilise the model country-wide for spray locations 

04

DEPLOYMENT

DEMO TIME!

05

LIMITATION AND FUTURE STEPS

LIMITATIONS

Insufficient information on the positive class resulted in low accuracy

IMBALANCE DATA

Had to compromise accuracy to minimise false negatives

LOW ACCURACY

Not easily accessible by users and limited information

DEPLOYMENT

Unable to get a good estimate of overhead costs

OVERHEAD COSTS

FUTURE STEPS

DATA COLLECTION

- Compile data from other states
- Collect more future data

DEPLOYMENT

- Deploy our model to cloud for convenient access
- Provide more information
- Improve the interface

Thank you!

Elise, Fadhil, Prerak, Shandeep