

THE LOATHSOME, LETHAL MOSQUITO





Problem Statement

- West Nile virus (WNV) is the leading cause of mosquito-borne disease in the United States
- Most commonly spread by infected mosquito
- No vaccines to prevent or medications to treat WNV in people
- From spring to fall, mosquitoes are trapped and tested for WNV
- These tests influence when and where the city will spray airborne pesticides to control adult mosquito populations
- We will build a model to predict outbreaks of WNV to help City of Chicago prevent transmission of WNV.
- Model performance will be guided by ROC AUC score.



Data Cleaning & EDA

3 key sets of data :

- training and test sets of the main dataset

training set consists of data from 2007, 2009, 2011, and 2013

test sets consists of data from 2008, 2010, 2012, and 2014

- Geospatial data of spraying efforts in 2011 and 2013
- Weather data from 2007 to 2014. Contains missing values in several columns.



Data Cleaning & EDA

Observations

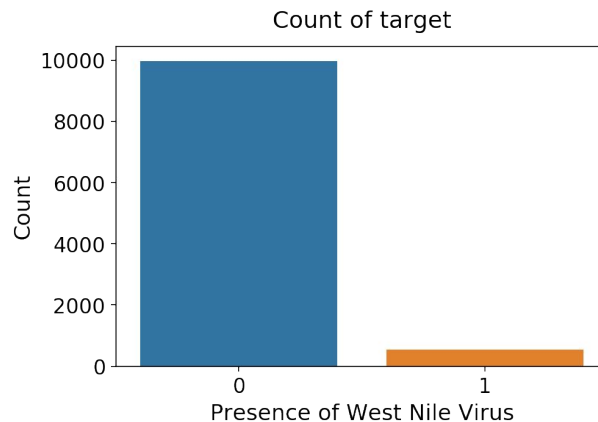
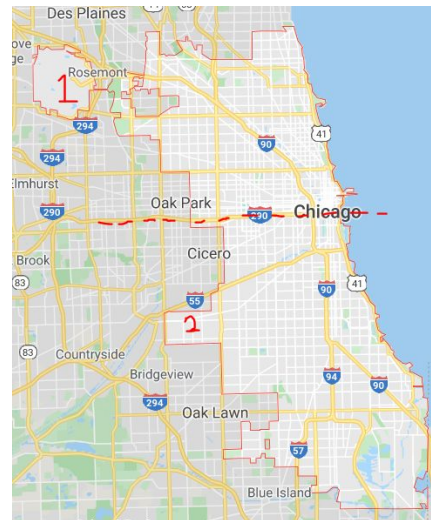
Varying years for train, test and spray datasets

Assigned weather variables based on the demarcation of Lat/Long

Spray data indicated more mosquitoes as we spray - but in reality we spray because of more mosquitoes

Spray data not used because of gaps in year timeline

Imbalanced class



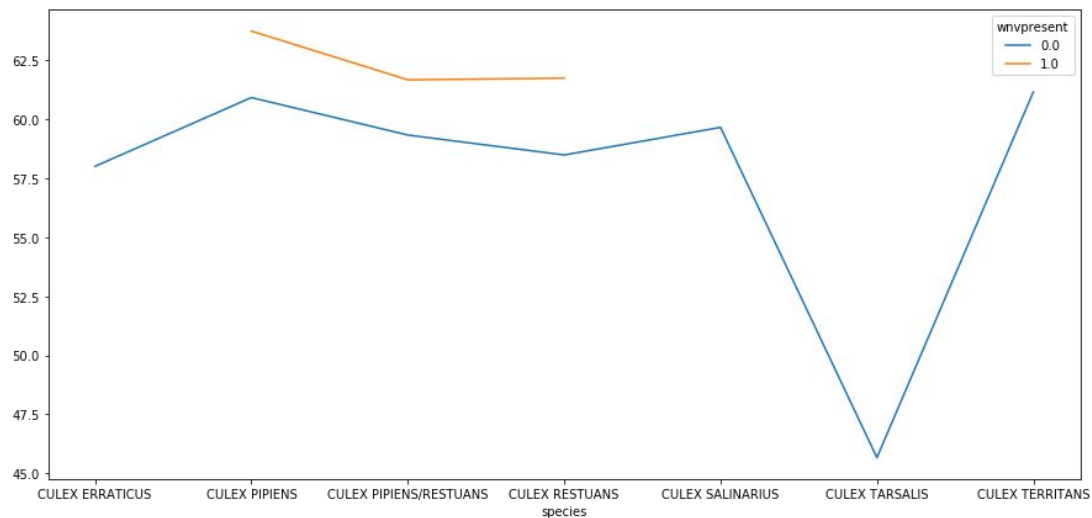


Feature Engineering

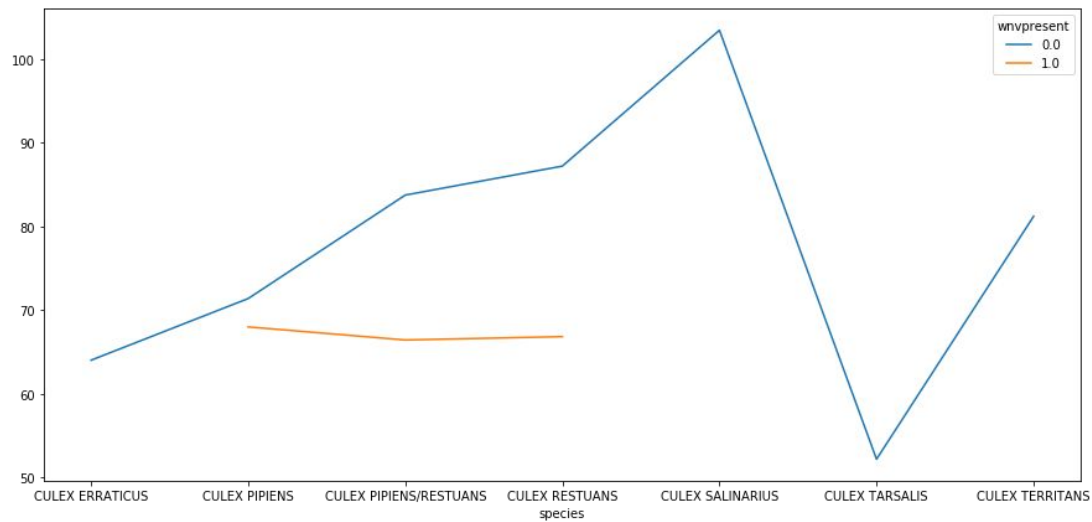
- Improvements can be made to the raw data provided by incorporating feature engineering:
 - Factors contributing to WNV:
 - Weather
 - Humidity
 - Temperature
 - Species



Dew Point



Wet Bulb





Dew Point

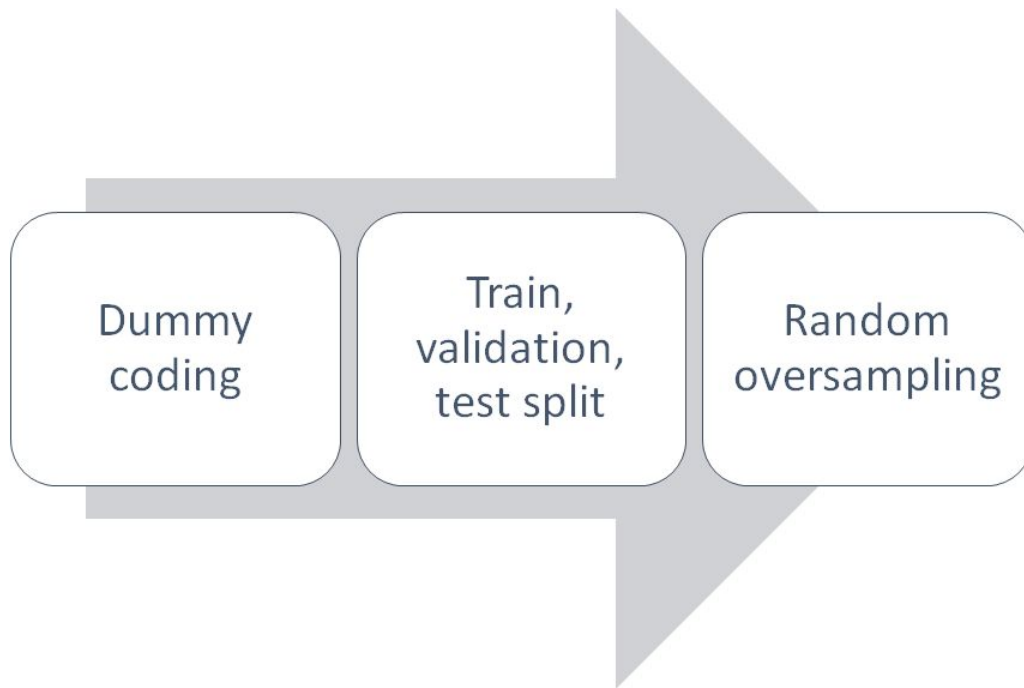
		species_mean_dp	species_max_dp	species_min_dp
species	wnvpresent			
CULEX ERRATICUS	0.0	58.000000	58	58
CULEX PIPIENS	0.0	60.912973	73	38
	1.0	63.729167	73	40
CULEX PIPIENS/RESTUANS	0.0	59.326726	73	38
	1.0	61.664122	73	38
CULEX RESTUANS	0.0	58.482349	73	38
	1.0	61.734694	73	50
CULEX SALINARIUS	0.0	59.651163	73	41
CULEX TARSALIS	0.0	45.666667	61	38
CULEX TERRITANS	0.0	61.148649	73	43

Wet Bulb

		species_mean_wp	species_max_wp	species_min_wp
species	wnvpresent			
CULEX ERRATICUS	0.0	64.000000	64.0	64.0
CULEX PIPIENS	0.0	71.370069	3386.5	47.0
	1.0	67.979167	76.0	47.0
CULEX PIPIENS/RESTUANS	0.0	83.739644	3386.5	46.0
	1.0	66.416031	76.0	47.0
CULEX RESTUANS	0.0	87.202713	3386.5	46.0
	1.0	66.816327	76.0	58.0
CULEX SALINARIUS	0.0	103.447674	3386.5	47.0
CULEX TARSALIS	0.0	52.166667	65.0	47.0
CULEX TERRITANS	0.0	81.204955	3386.5	49.0

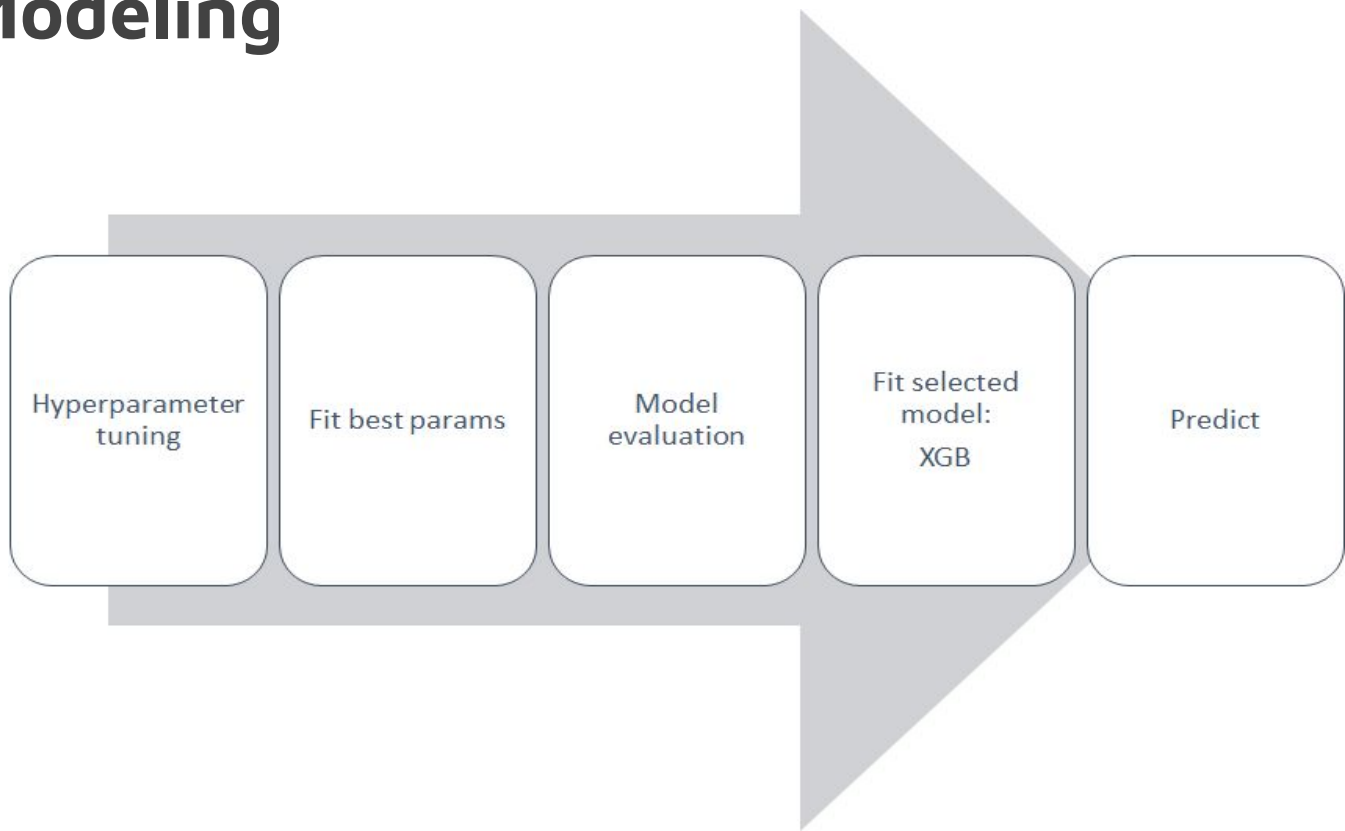


Preprocessing

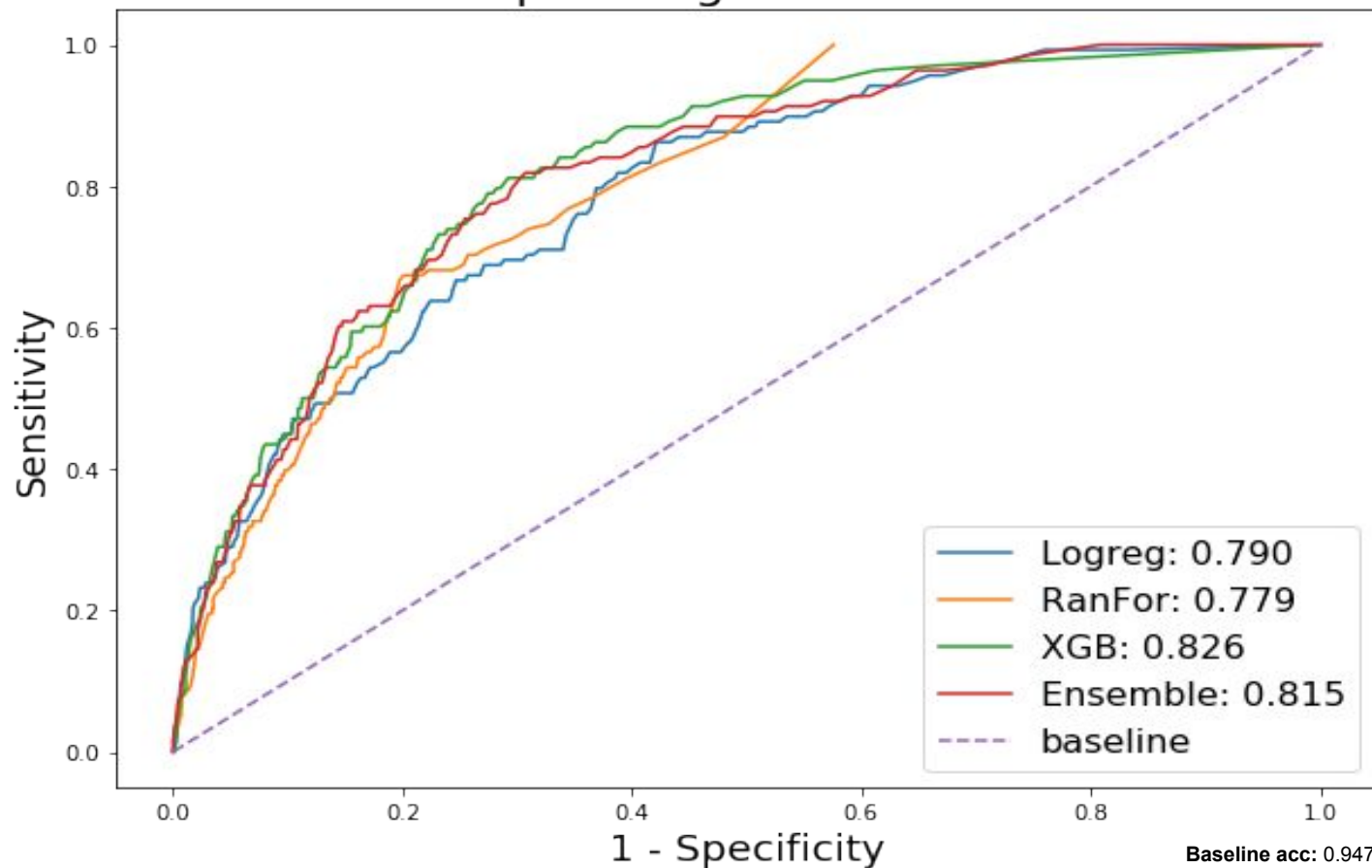




Modeling



Receiver Operating Characteristic Curve

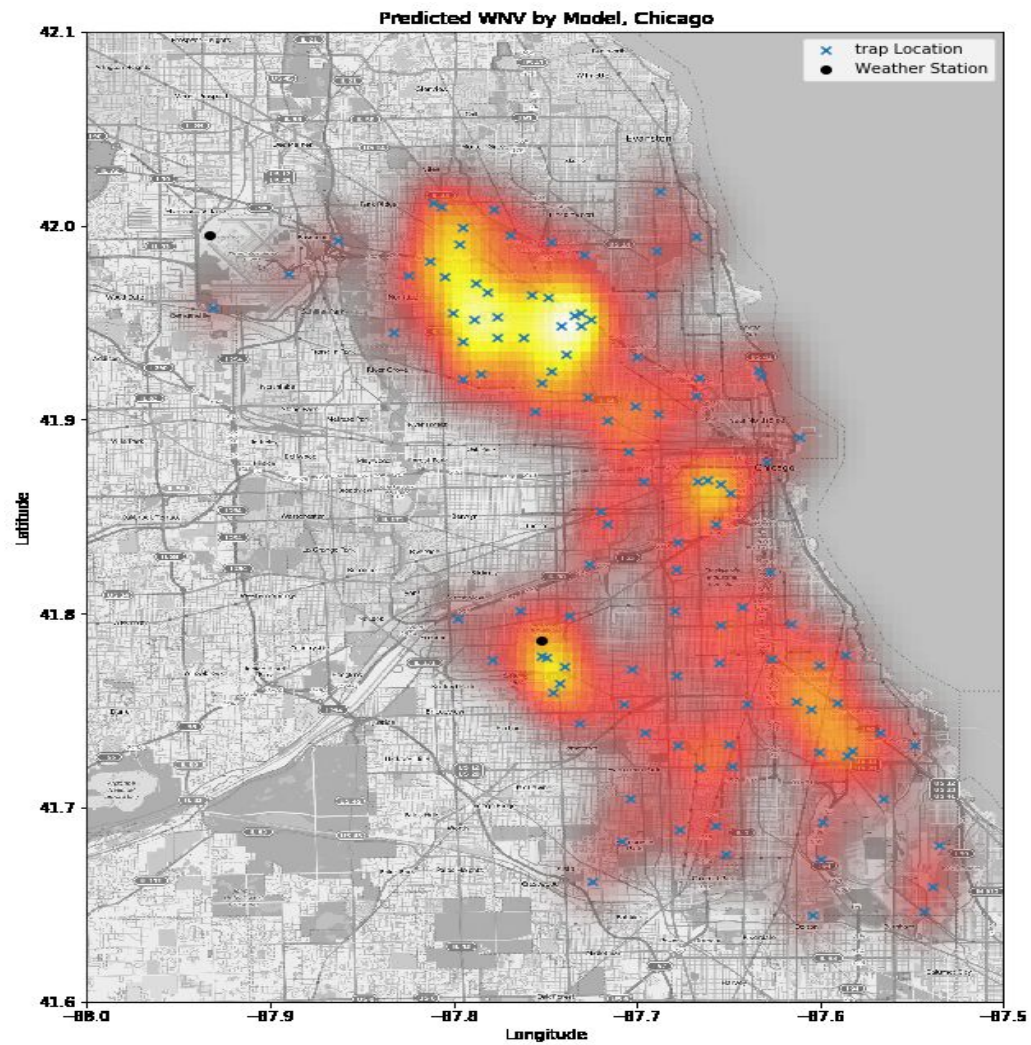


Baseline acc: 0.947

XGB acc: 0.836

Specificity : 0.854

Sensitivity : 0.558





Conclusion & Recommendations

- Preemptive targeted spraying
- Launch education campaigns at hotspot neighbourhoods to nip larvae growth in the bud
- Inculcate proactive behaviour and habits to remove stagnant water
- Broadcast information to self-protect during days/ months where mosquito count is expected to peak
- Increase efforts in preventing the proliferation of Culex Pipiens being the main WNV carrier by using environmentally friendly measures such as "targeted biological mosquito control".



Directions for future research

- Collect more data
- Link Trap ID to weather station to track mosquitoes population more accurately
- More engineered features
- Explore time & spatial analysis