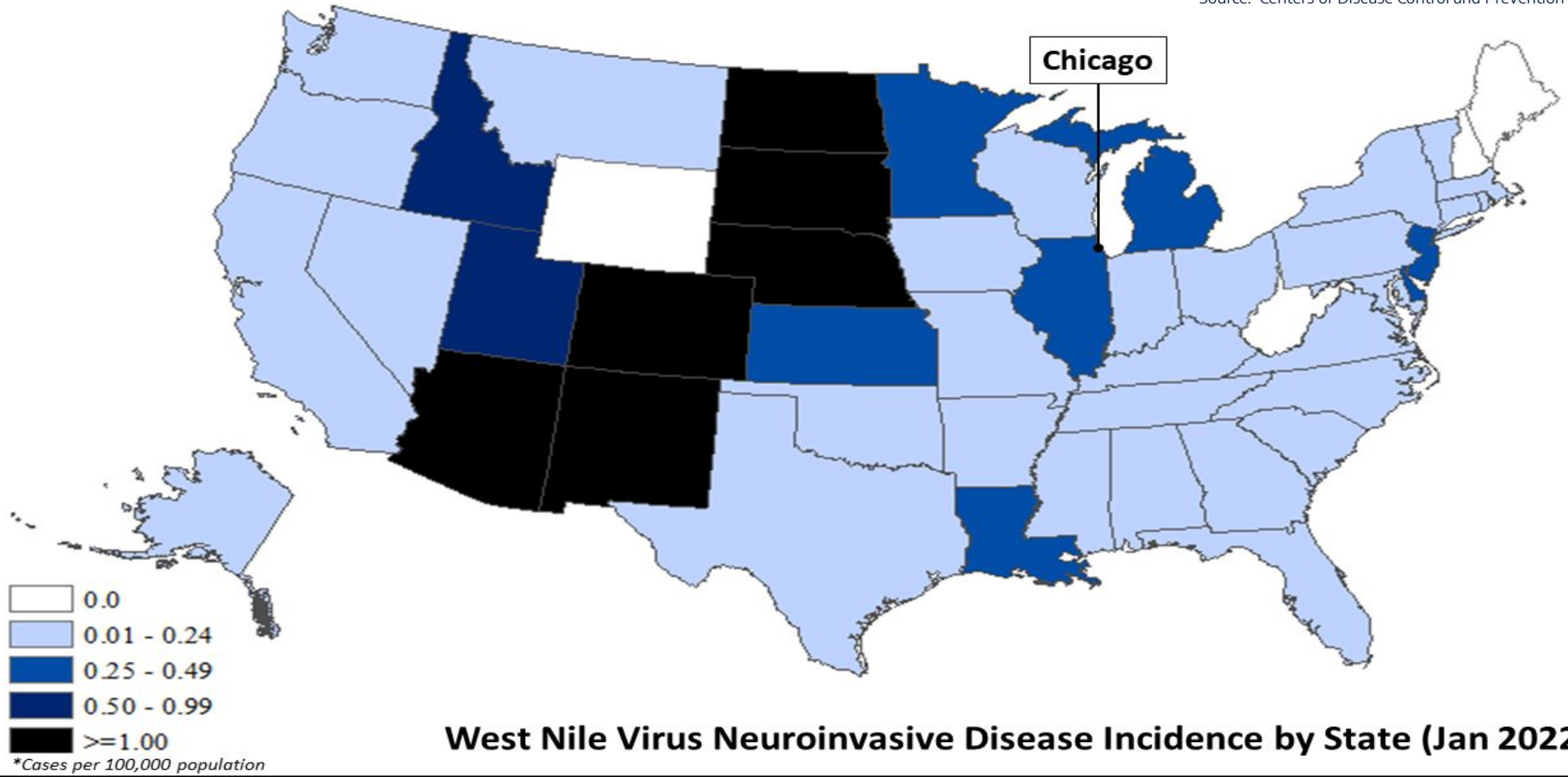
West the Mozzies Nile?

DSI Project 4

Brenda, Nicholas, Yong

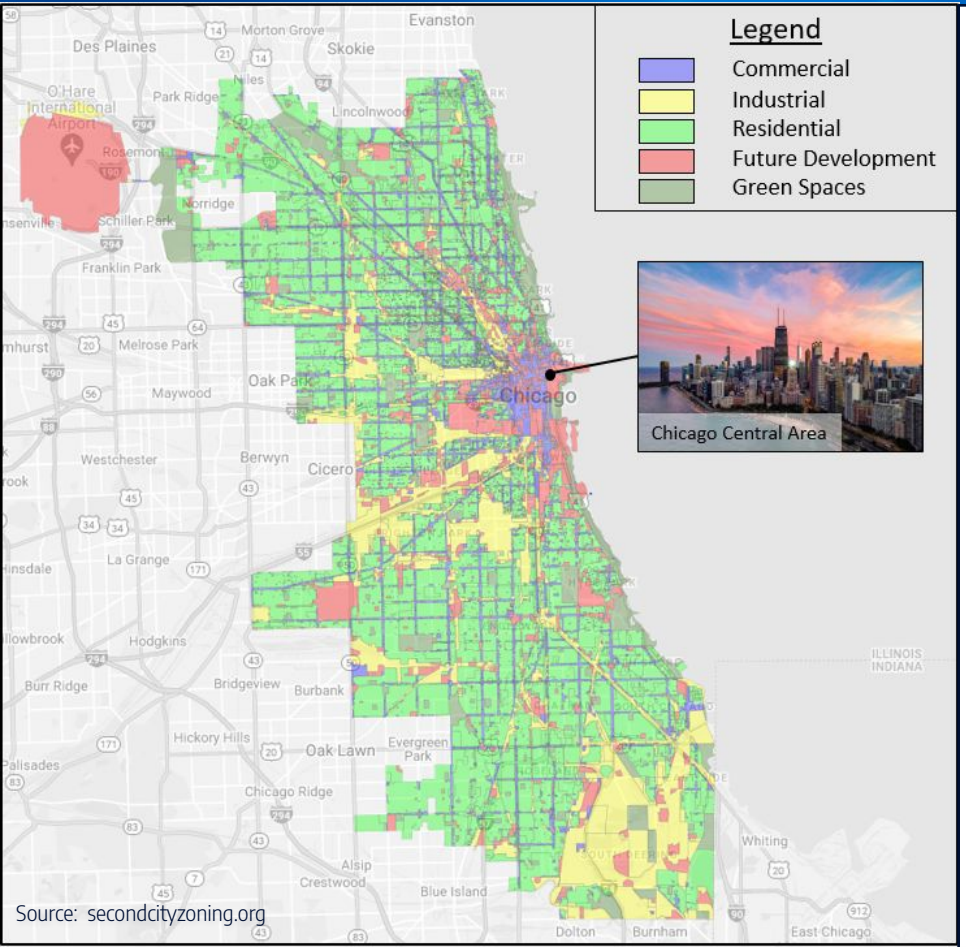
Introduction: West Nile Virus (WNV)

Source: Centers of Disease Control and Prevention (CDC)

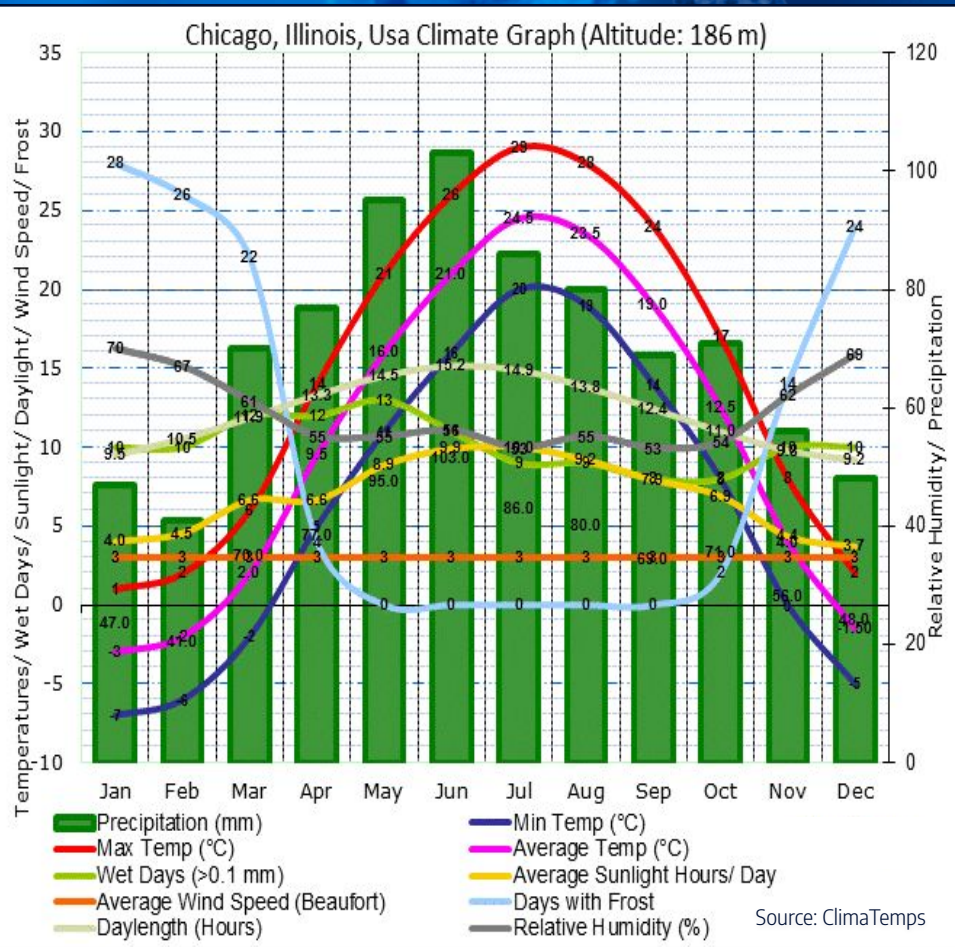


West Nile Virus Neuroinvasive Disease Incidence by State (Jan 2022)

Study Area: Chicago



Source: secondcityzoning.org



Source: ClimaTemps

Problem Statement:

As members of the Disease And Treatment Agency, division of Societal Cures In Epidemiology and New Creative Engineering (DATA-SCIENCE) research team under the University of Chicago, the Chicago Municipal Government Department of Public Health has engaged the team to develop a prediction model to identify areas in Chicago which are more prone to WNV.

Data & Workflow

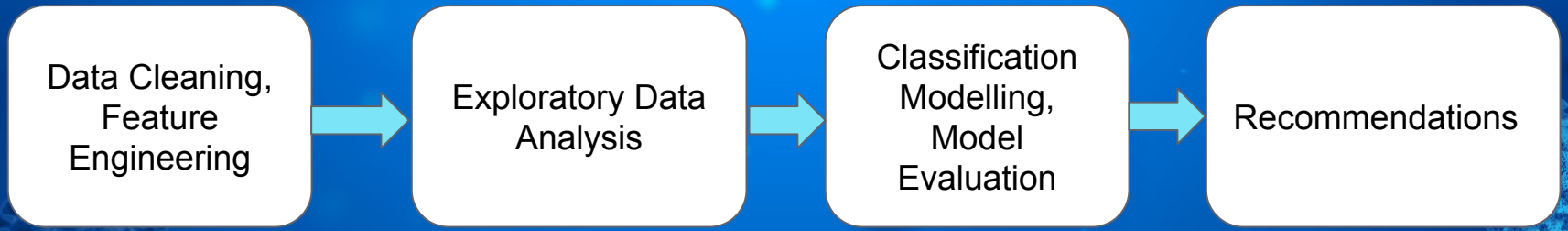
Data Provided

Spatial Features
- Latitude
- Longitude
- Trap Id
- Address
- Block

Mosquito Information
- Species
- Number of Mosquitoes
- WNV Presence

Weather Information
- Temperature
- Precipitation
- Wind
- etc.

Workflow



Data Cleaning, Feature Engineering

Data Cleaning

Train Dataset
- Merging of rows for mosquitoes > 50

Merger of Train/Test Dataset with Weather Dataset

Weather Dataset
- Imputation of Data for Missing Data

Spray DataSet
- Dropping duplicated data

Feature Engineering

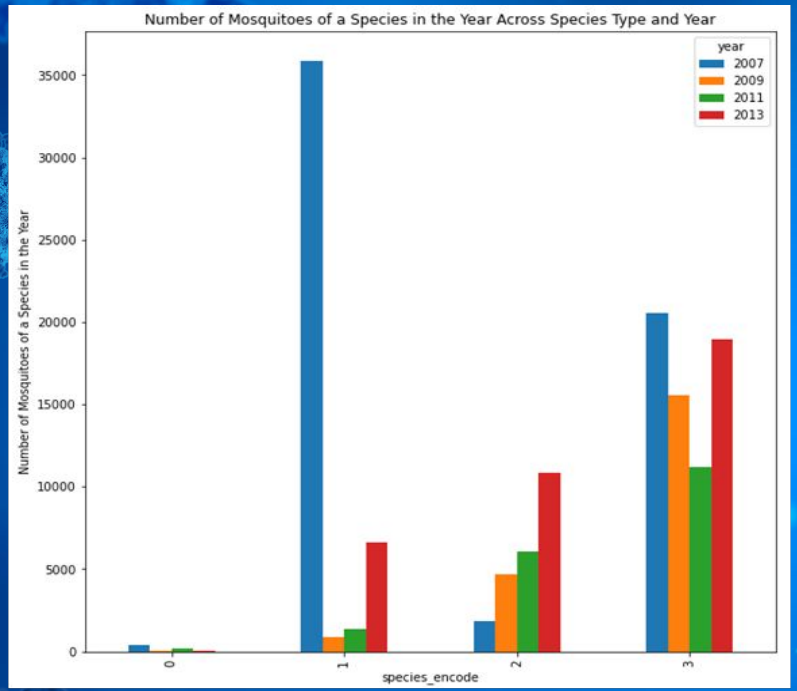
Rolling of Weather

Multicollinearity

Create Humidity Feature

One-Hot Encode

Number of Mosquitos by Species Type and Year



Culex erraticus, Culex salinarius, Culex tarsalis, Culex territans

Culex pipiens

Culex restuans

Culex pipiens/ restuans

Culex erraticus, Culex salinarius, Culex tarsalis, Culex territans

- Each year: 22 – 405
- Average (1 year): 167
- Total (4 years): 669

Culex pipiens

- Each year: 867 - 35871
- Average (1 year): 11168
- Total (4 years): 44671

Culex restuans

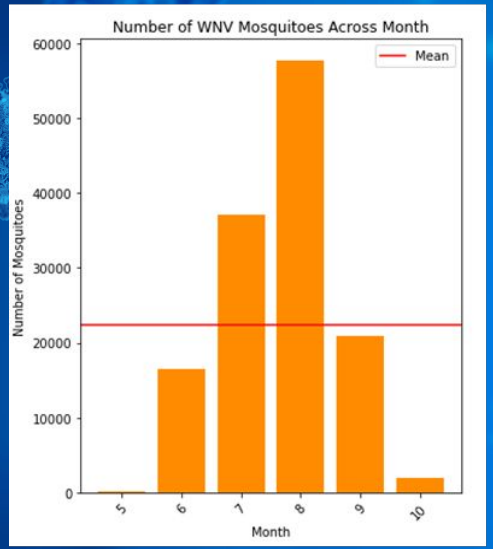
- Each year: 1847 - 10859
- Average (1 year): 5858
- Total (4 years): 23431

Culex pipiens / restuans

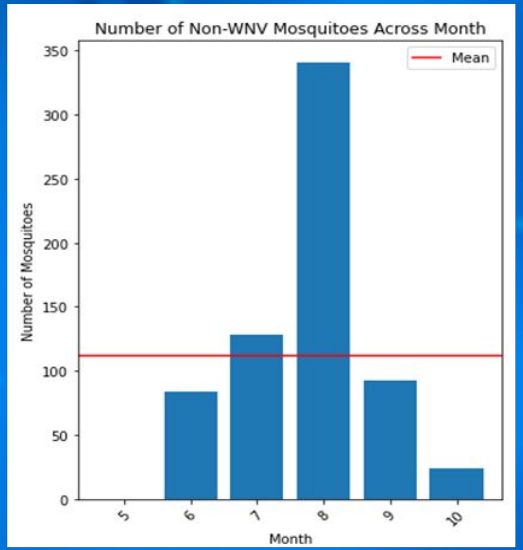
- Each year: 11206 - 20565
- Average (1 year): 16567
- Total (4 years): 66268

Non-WNV/WNV Mosquito by Month

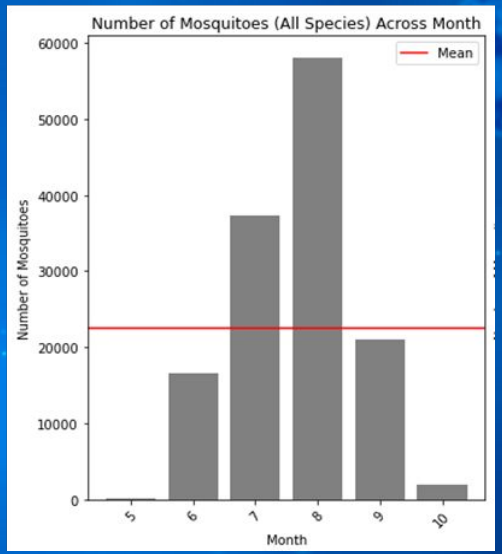
Species that can Carry WNV



Species that do not Carry WNV

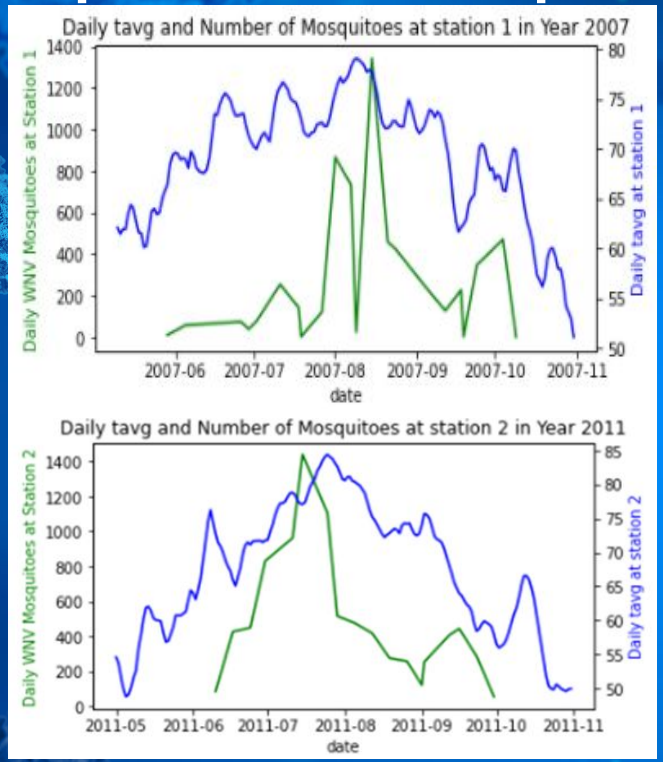


All Species



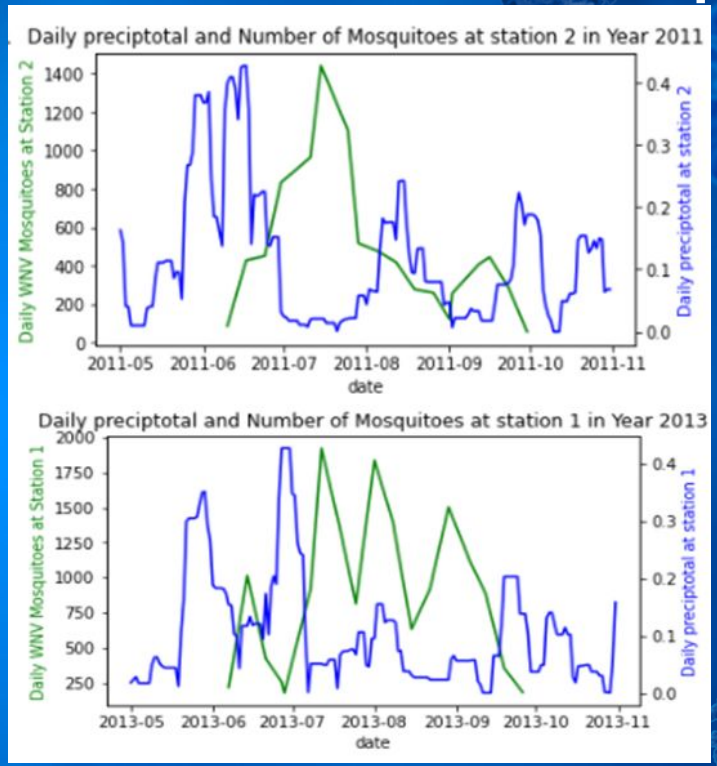
Increases from May, remains high in July and August, decreases from September to October

Temperature, Precipitation & Number of WNV Mosquitos



Higher temperature: July, August

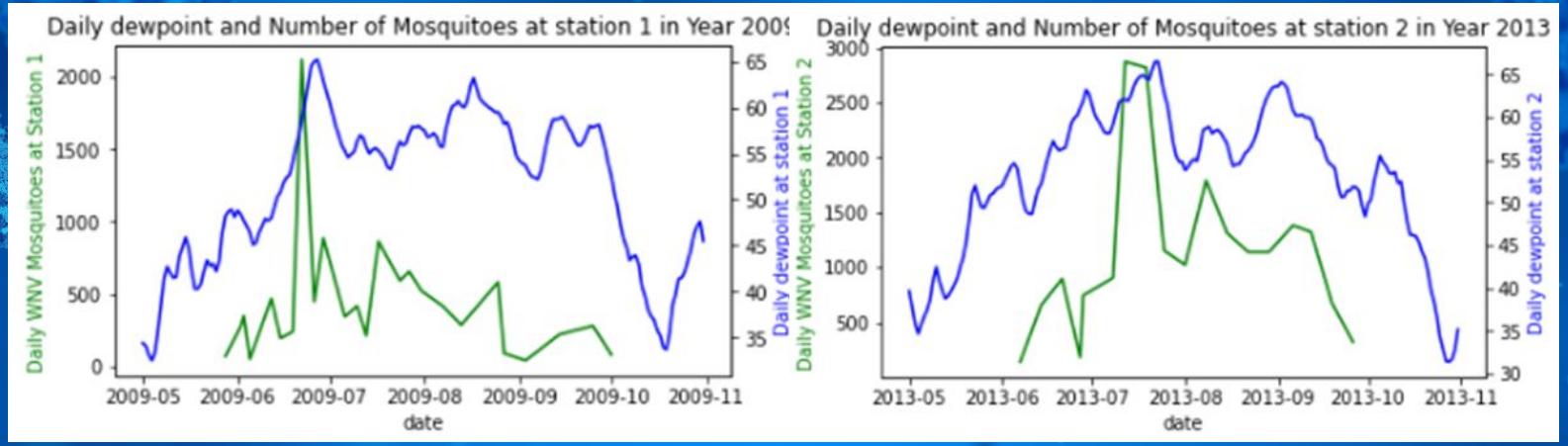
High number of WNV mosquitos: July, August



High precipitation: June, July, August

High number of WNV mosquitos: July, August

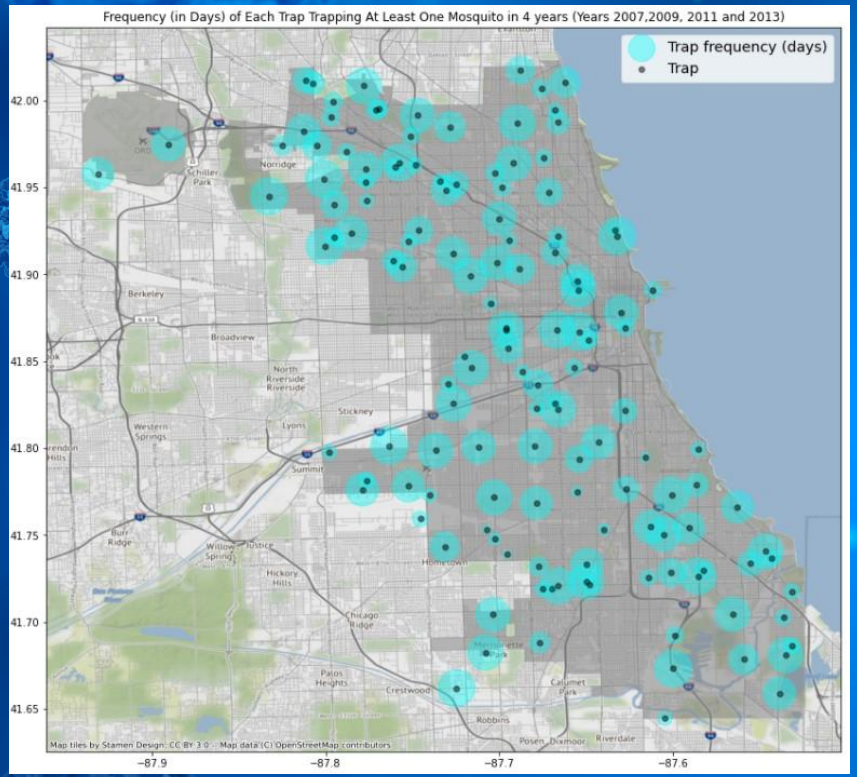
Dewpoint & Number of WNV Mosquitos



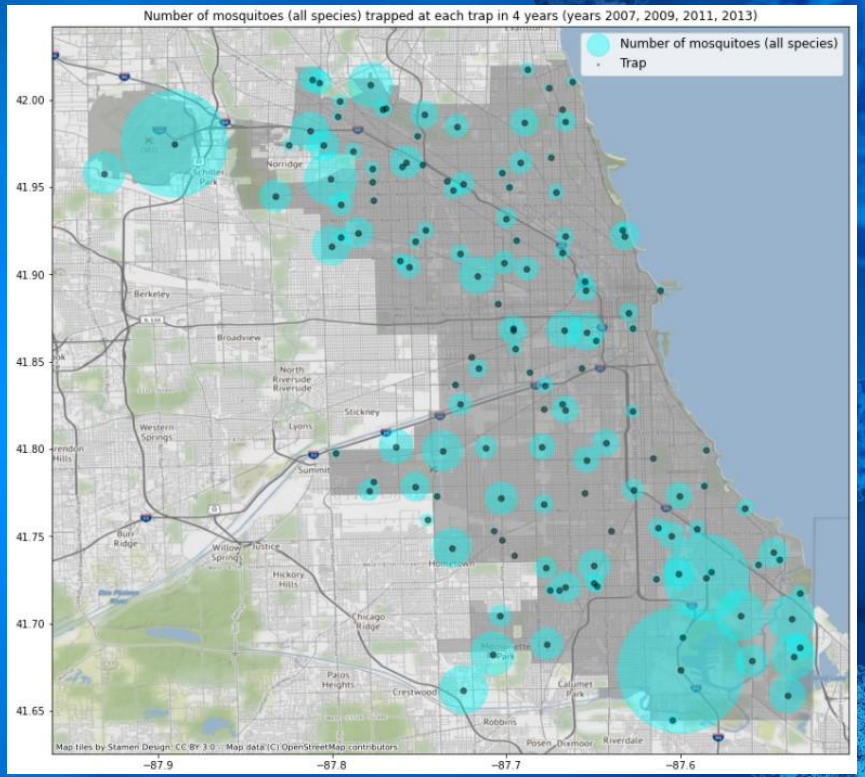
Higher dewpoint: July, August

High number of WNV mosquitos: July, August

Trap Frequency (Days) & Number of WNV Mosquitos Trapped



Frequency (Days) of Each Trap Trapping at least 1 Mosquito over 4 Years

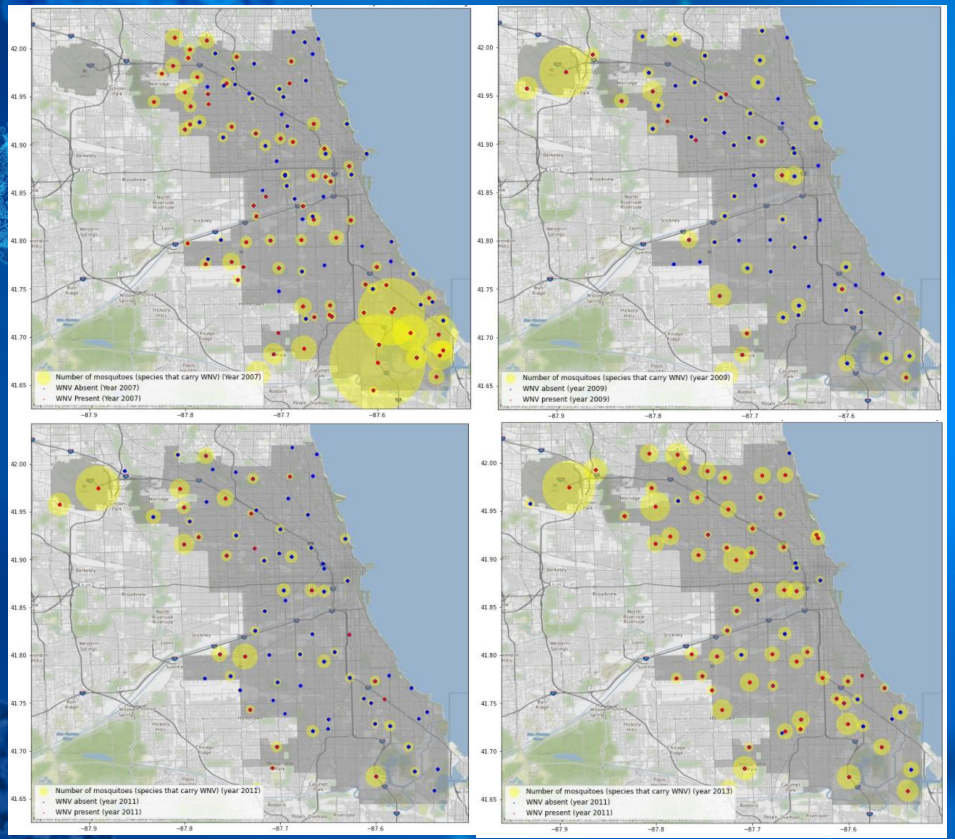


Number of Mosquitoes (All Species) Trapped by Each Traps over 4 Years

Number of WNV Mosquitoes Trapped by Each Trap

2007

2009



2011

2013

More concentrations of high numbers of WNV mosquitos (2007, 2009, 2011, 2013):

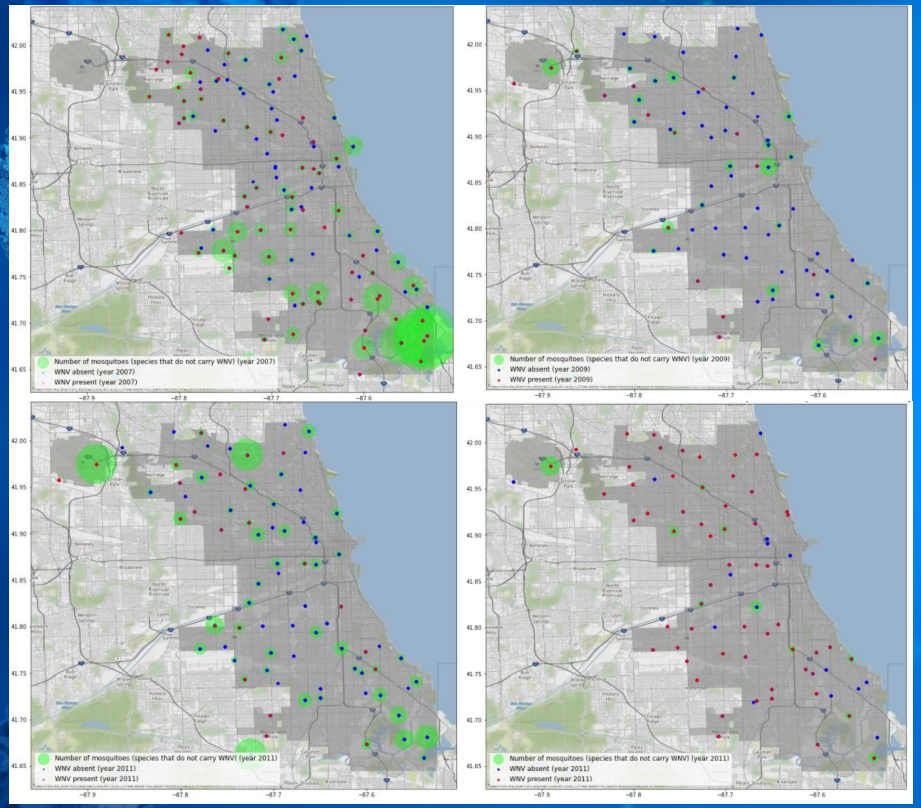
- North
- South-east
- Central

Traps with top 5 highest number of WNV mosquitos in each year:

- North: T900, T903, T002, T151, T100, T002, T030, T008
- South-east: T151, T138, T103, T128, T212, T135, T225, T158
- Central: T114

Number of Non-WNV Mosquitoes Trapped by Each Trap

2007 2009



2011 2013

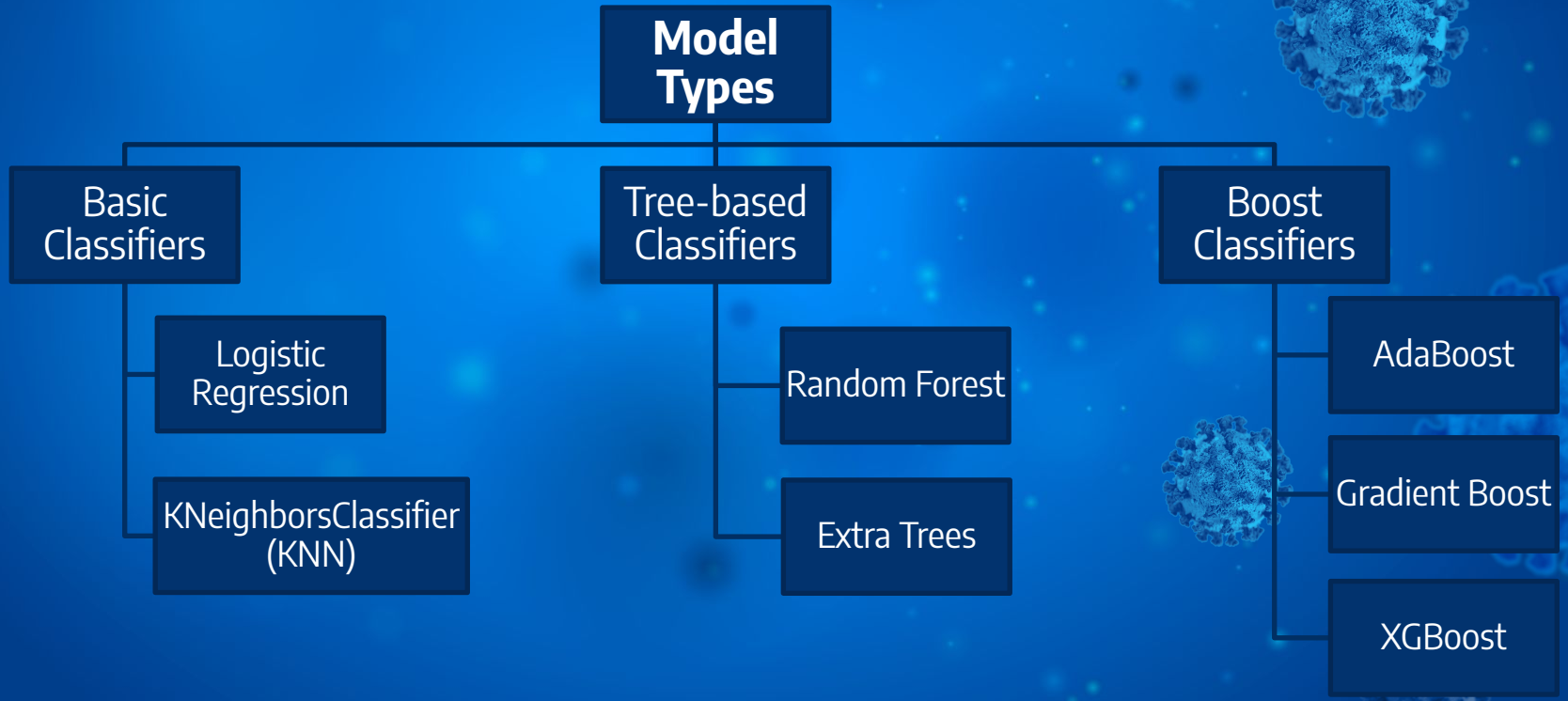
More concentrations of high numbers of Non-WNV mosquitos (2007, 2009, 2011, 2013):

- South-east
- Central
- North

Traps with top 5 highest number of non-WNV mosquitos in each year:

- South-east: T212, T200, T215, T138, T103, T159, T135, T221
- Central: T048, T074, T235
- North: T900, T003, T230

Model Framework

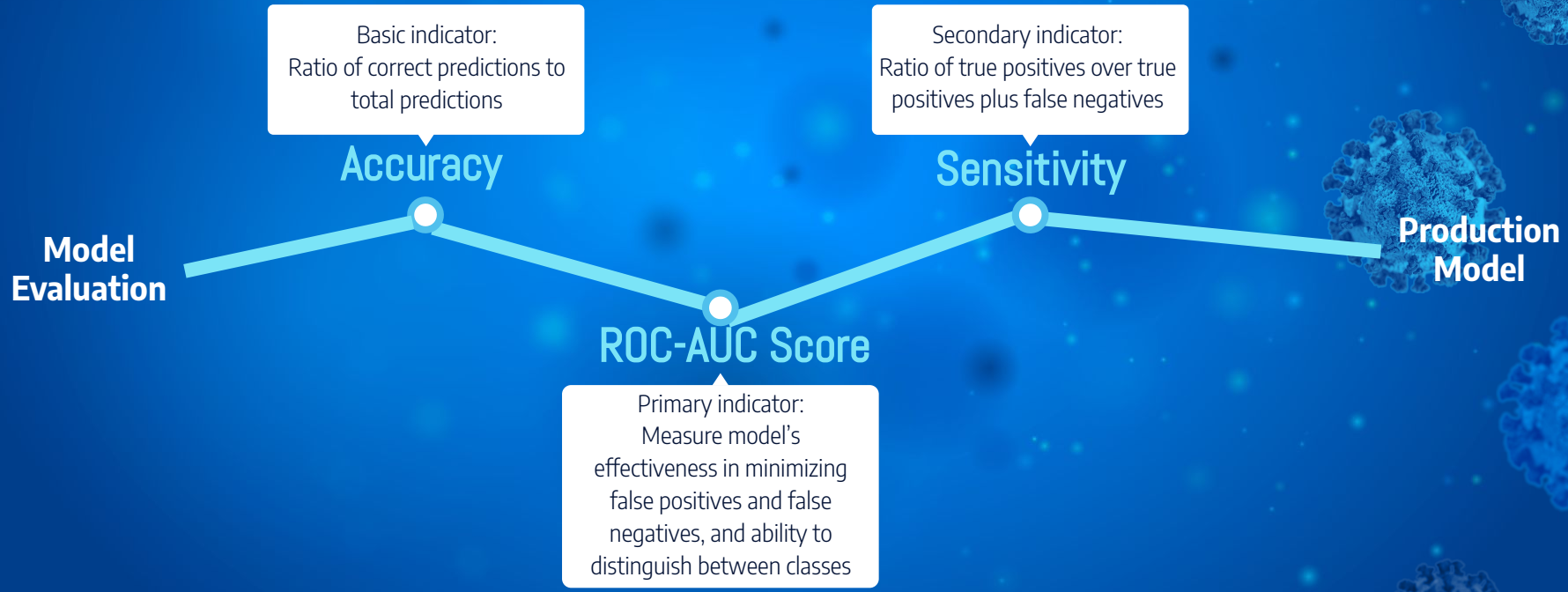


Modelling

- Using GridSearchCV, various parameters are used to tune the above models.
- The result are as follows:

	model	GridSearch_score	train_score	validation_score	accuracy_valid	recall_valid	roc_auc_score	f1score_test
0	LogReg	0.7302	0.7823	0.7884	0.8931	0.1758	0.5545	0.1481
1	KNN	0.8013	0.8713	0.8256	0.6574	0.8901	0.7672	0.2154
2	RandomForest	0.8130	0.8752	0.8709	0.7677	0.8352	0.7996	0.2754
3	ExtraTrees	0.7917	0.9016	0.8544	0.8264	0.7143	0.7735	0.3030
4	AdaBoost	0.8031	0.8131	0.8563	0.6754	0.8681	0.7664	0.2204
5	GradientBoost	0.8056	0.8378	0.8642	0.7195	0.8352	0.7741	0.2394
6	XGBoost	0.8106	0.8702	0.8671	0.7828	0.7912	0.7868	0.2780

Scoring Metrics



Final Model

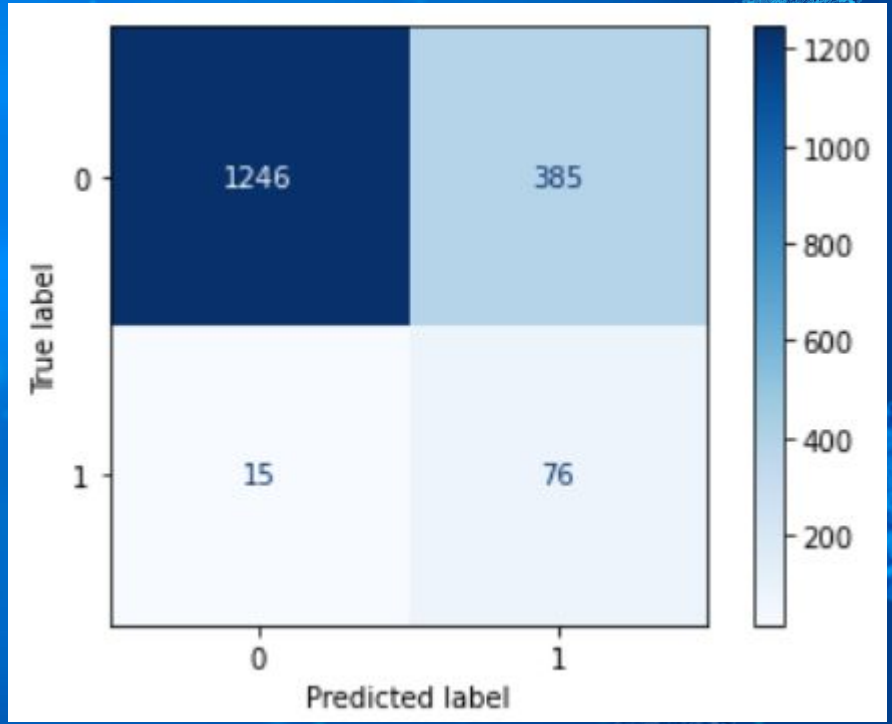
- The model chosen is the Random Forest Classifier.
- The best parameters are:
  - Class Weight: Balanced Sample
  - Max Depth: 10
  - Max Features: Auto
  - Min Samples Leaf: 5
  - N Estimators: 200

Evaluation

- Accuracy:

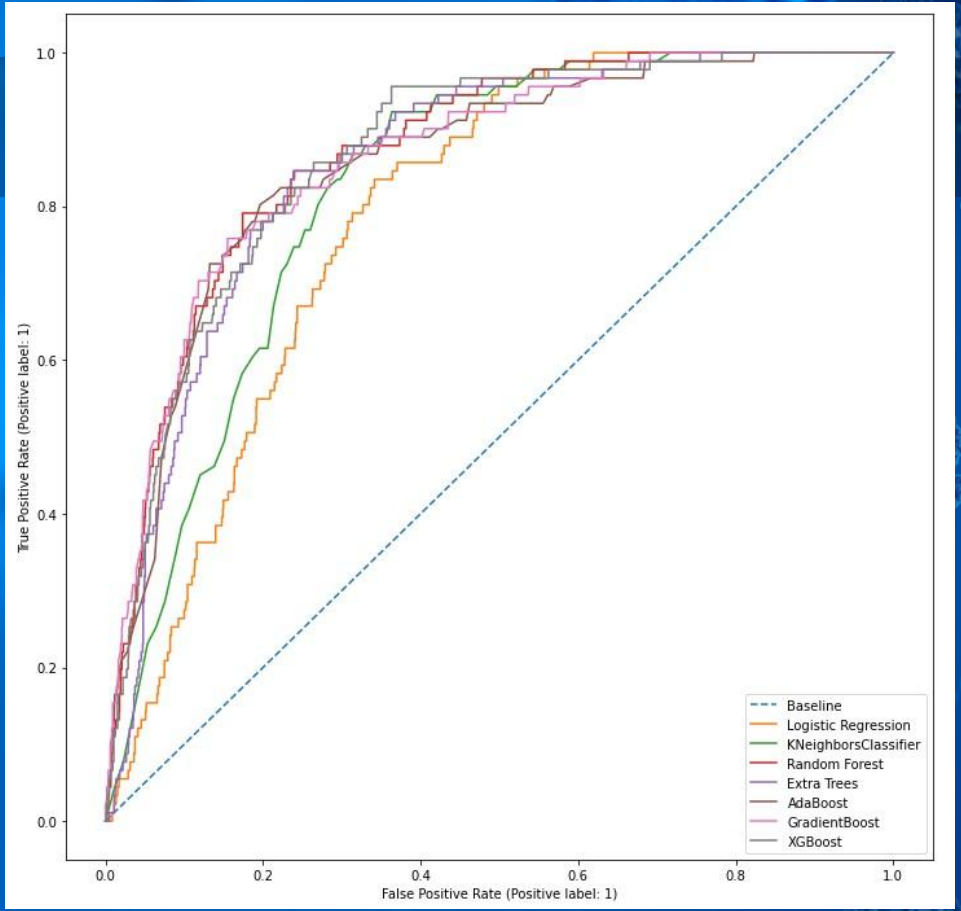
  (TN+ TP) / (TN + FP+ TP + FN)

  0.7677



Evaluation

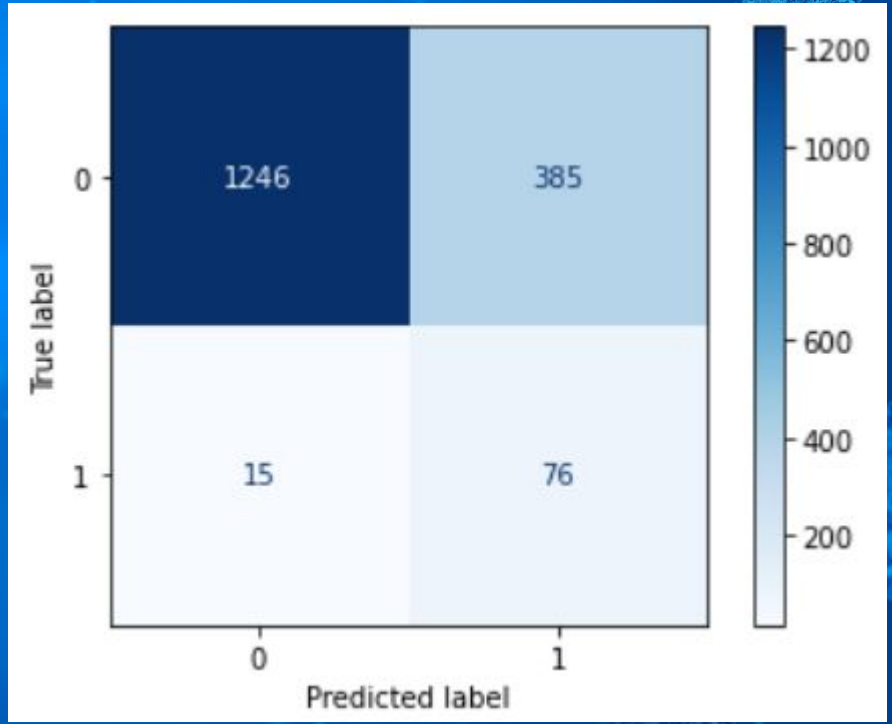
- ROC AUC Score:
  0.7996



Evaluation

- False Negatives: 15
- Sensitivity: TP / (TP + FN)

  0.8352



“Spraying is a safe and effective way to control disease carrying mosquitoes. By reducing the number of mosquitoes in high-risk areas we expect that this will help minimize the number of human cases of West Nile Virus in Chicago

-Cort Lohff (CDPH Environmental Health Medical Director, 2013)

West Nile virus disease cases reported to CDC by state of residence, 1999–2020

State	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total
Alabama	0	0	2	49	37	16	10	8	24	18	0	3	5	62	9	2	9	19	60	28	5	9	375
Alaska	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2
Arizona	0	0	0	0	13	391	113	150	97	114	20	167	69	133	62	107	103	78	111	26	174	12	1,940
Arkansas	0	0	0	43	25	28	28	29	20	9	6	7	1	64	18	11	18	9	18	8	9	1	352
California	0	0	0	1	3	779	880	278	380	445	112	111	158	479	379	801	783	442	553	217	225	235	7,261
Colorado	0	0	0	14	2,947	291	106	345	576	71	103	81	7	131	322	118	101	149	68	96	122	35	5,683
Connecticut	0	1	6	17	17	1	6	9	4	8	0	11	9	21	4	6	10	1	3	23	1	8	166
Delaware	0	0	0	1	17	0	2	0	1	1	0	0	1	9	3	0	6	0	1	10	0	0	52
Dist. of Columbia	0	0	0	34	3	2	5	2	0	8	2	6	15	10	1	3	5	1	4	13	11	0	125
Florida	0	0	12	28	94	41	21	3	3	3	3	12	24	73	7	17	13	8	5	35	2	51	455
Georgia	0	0	6	44	50	21	20	8	50	8	4	13	22	99	10	13	15	6	48	36	14	8	495
Hawaii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
Idaho	0	0	0	0	1	3	13	996	132	39	38	1	3	17	40	19	13	9	25	16	13	6	1,384
Illinois	0	0	0	884	54	60	252	215	101	20	5	61	34	290	117	44	77	154	90	176	28	42	2,704
Indiana	0	0	0	293	47	13	23	80	24	4	4	13	9	77	23	10	21	18	26	35	4	3	727
Iowa	0	0	0	54	147	23	37	37	30	6	5	9	9	31	44	15	14	37	12	104	5	3	622
Kansas	0	0	0	22	91	43	25	30	40	31	13	19	4	56	91	54	34	37	27	47	13	8	685
Kentucky	0	0	0	75	14	7	5	6	4	3	3	3	5	23	3	1	2	8	10	12	10	0	194
Louisiana	0	0	1	329	124	109	171	180	40	49	21	27	10	335	54	125	51	59	53	83	20	18	1,859
Maine	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	2	0	1	5
Maryland	0	0	6	36	73	16	5	11	10	14	1	23	19	47	16	6	45	6	6	45	7	1	393
Massachusetts	0	0	3	23	17	0	6	3	6	1	0	7	6	33	8	6	10	17	6	49	5	12	218
Michigan	0	0	0	614	19	16	62	55	17	17	1	29	34	202	36	1	18	43	40	102	12	32	1,350
Minnesota	0	0	0	48	148	34	45	65	101	10	4	8	2	70	79	21	9	83	30	63	5	0	825
Mississippi	0	0	0	192	87	51	70	183	136	65	53	8	52	247	45	43	38	43	63	50	15	7	1,448

To spray or not to spray?

Cost-Benefit Analysis

Costs

- Cost of spraying agent (Zenivex E4)
- Spraying operational cost (manpower, transport, logistics)
- Trap installation, surveillance and maintenance
- Other negative externalities

Benefits

- Savings on medical treatment fees
- Avoid loss in economic productivity
- Reduce pain and suffering
- Alleviate burden on caregivers

USD4,635,0000

Annual Cost of Spraying

USD111,383

Annual Benefits of Spraying

-USD4,523,518

Net Cost/Benefit

Conclusion & Recommendations

Recommendation

1) Targeted locational spraying
2) Context-specific study on landuses
3) Identify more predictor variables
4) Educational initiatives

Conclusion

Although the production model slightly falls short of the scoring benchmarks, it partially addresses the problem statement as the study shows the time and location where WNV is projected to escalate.

Thank you.

Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash
- COVID-19 virus image from CDC by Alissa Eckert & Dan Higgins · Public domain and thus free of any copyright restrictions