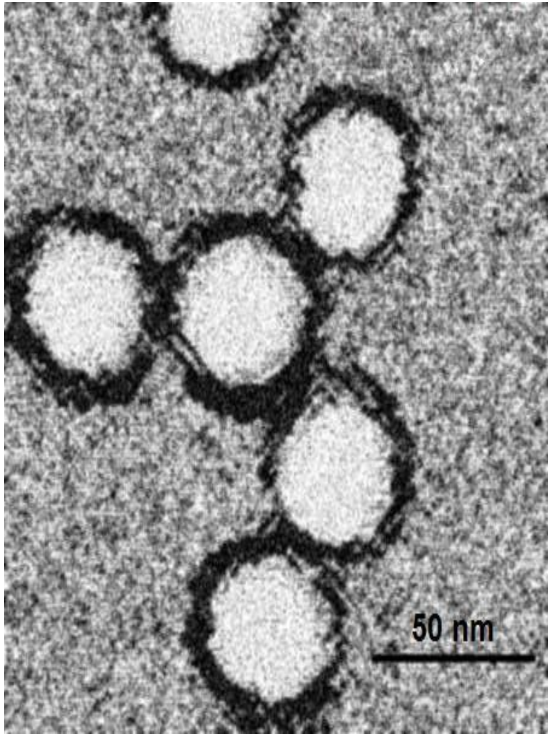
Predicting West Nile Virus

Anna Fenner

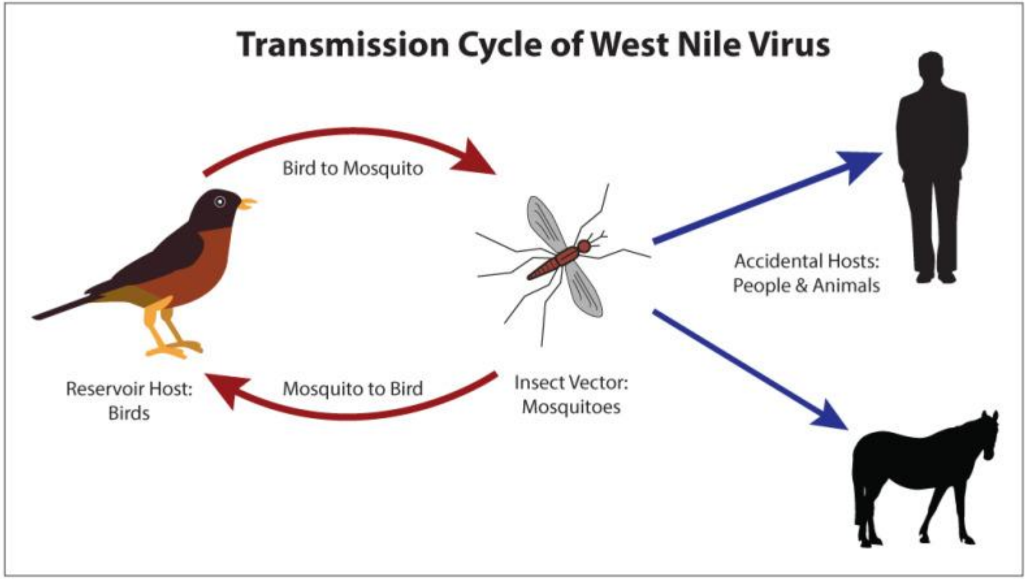
Why



In 1999, the West Nile virus was first detected in the US (1).

Mosquitoes are infected with the West Nile virus by infected birds and typically infect humans during late summer and early fall, leading to potentially severe illnesses and, in some cases, fatalities (2).

Why



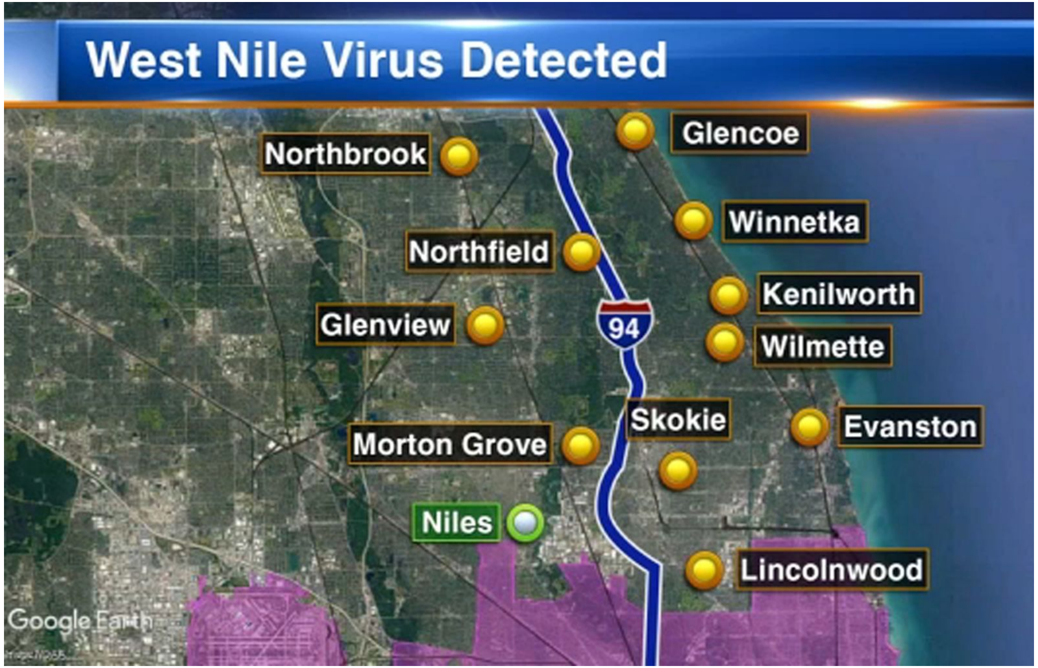
Using a more accurate machine learning model, we aim to predict and prevent West Nile virus outbreaks by understanding mosquito transmission conditions.

The Data

The City of Chicago and the Chicago Department of Public Health have a comprehensive program to monitor and track the West Nile virus (3).

1) The training dataset contains dates, trap names, locations, number of mosquitoes, if West Nile virus was present or not, etc.

(2) Weather dataset has two weather stations situated at two airports. It includes dates, temperatures, sunrise and sunset times, and other relevant weather-related metrics.



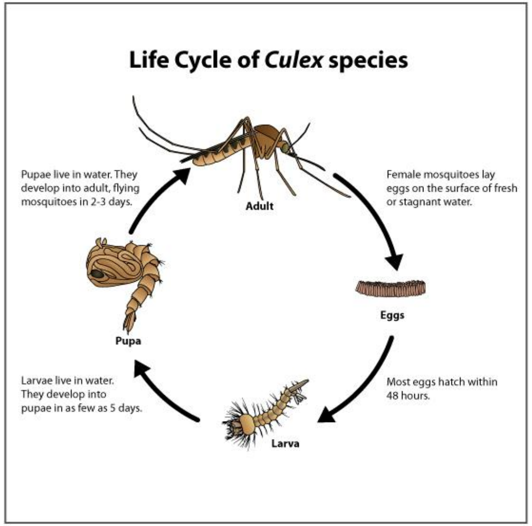
https://abc7chicago.com/west-nile-virus-illinois-niles-symptoms/12099586/

Feature Engineering

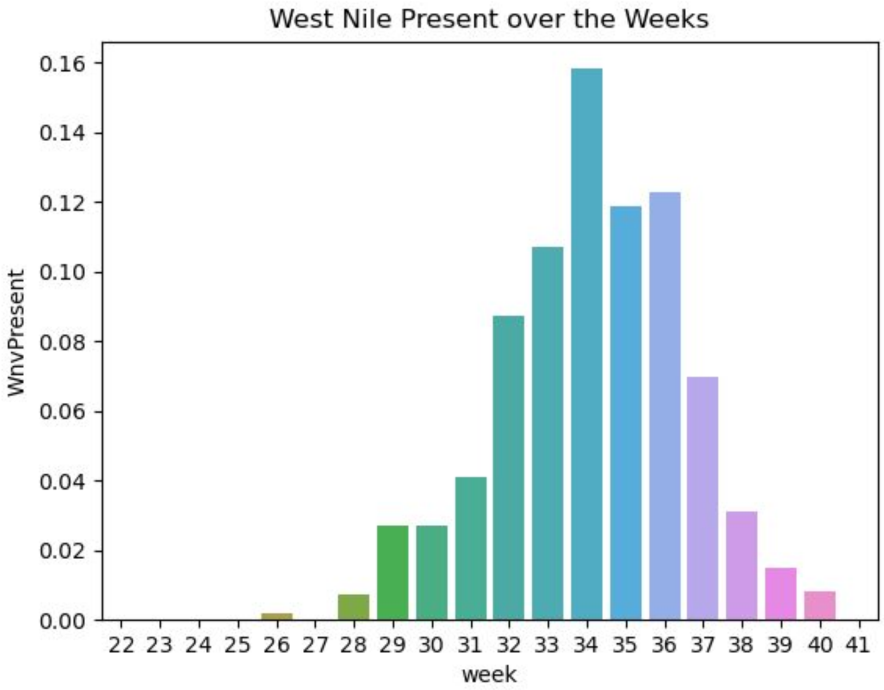
Mosquitoes thrive in drought-like weather (2). Once eggs hatched, it takes 7-8 days to become adults (4). Mosquitoes are active at night (7).

New columns added to the dataset based upon those facts:

1. Created a binary column if it had not rained that day or if it had rained that day.
2. Created a relative humidity column.
3. Shifted three columns (average temperature in Celsius, precipitation, and relative humidity) by 7, 14, 21 days.
4. Two columns were added based upon how much daylight and nighttime there were for each day.

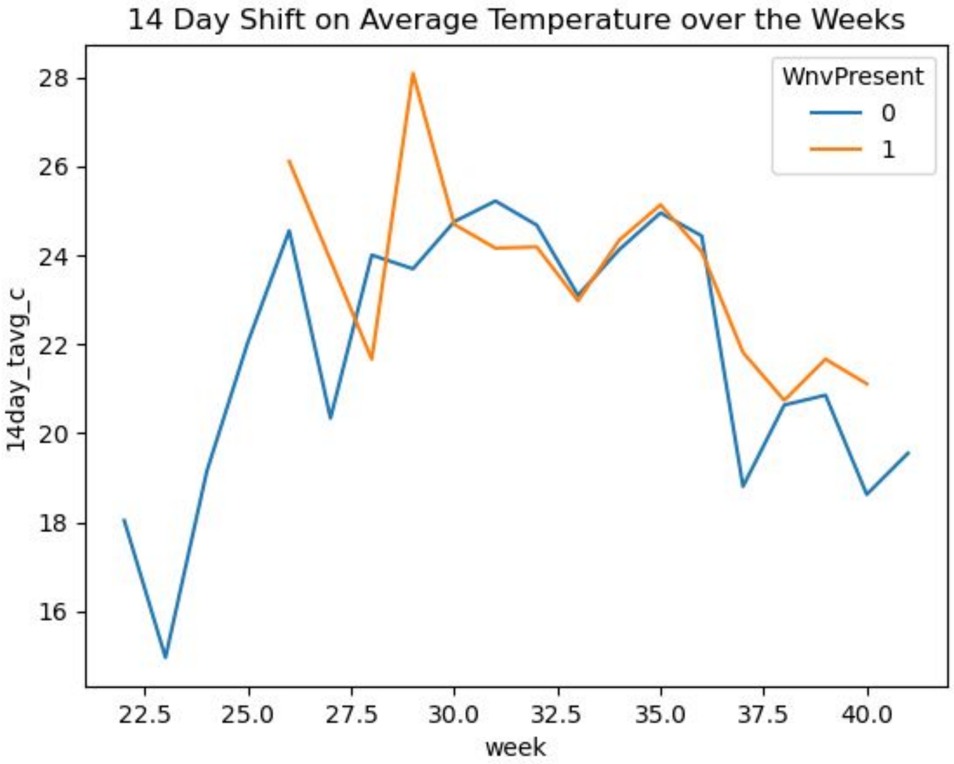


Insights from Exploring the Data



Observed a trend of mosquitoes carrying the West Nile virus exhibited an upsurge beginning around week 32 and reaching their peak between weeks 33-36

Insights from Exploring the Data



Discernible pattern emerged when plotting a 14-day average temperature shift across the weeks. Showed a pattern for the rise of mosquitoes with the West Nile virus when temperatures ranged between 23 to 26 degrees Celsius.

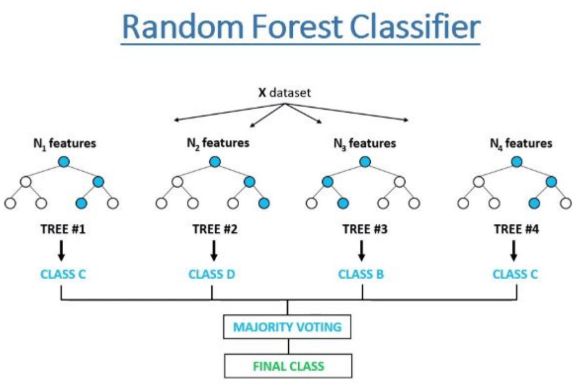
Modeling

Machine Learning Algorithms used:

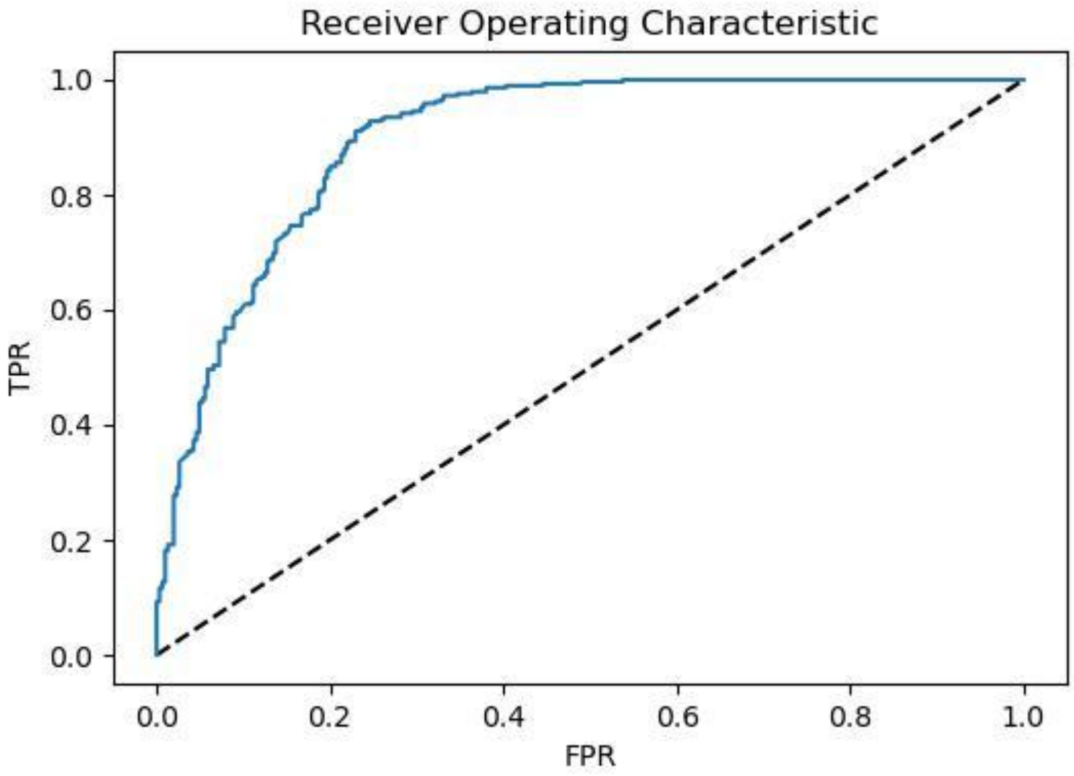
1. Logistic Regression
2. XGBoost Classification
3. Random Forest Classification.

Hyperparamter Tuning used:

1. GridSearchCV from Scikit Learn package
2. Hyperopt package.



Best Model - XGBoost



ROC AUC score is a single scalar value to quantify the overall performance the model. Scores range from 0 to 1. Closer to 1 indicates a better model.

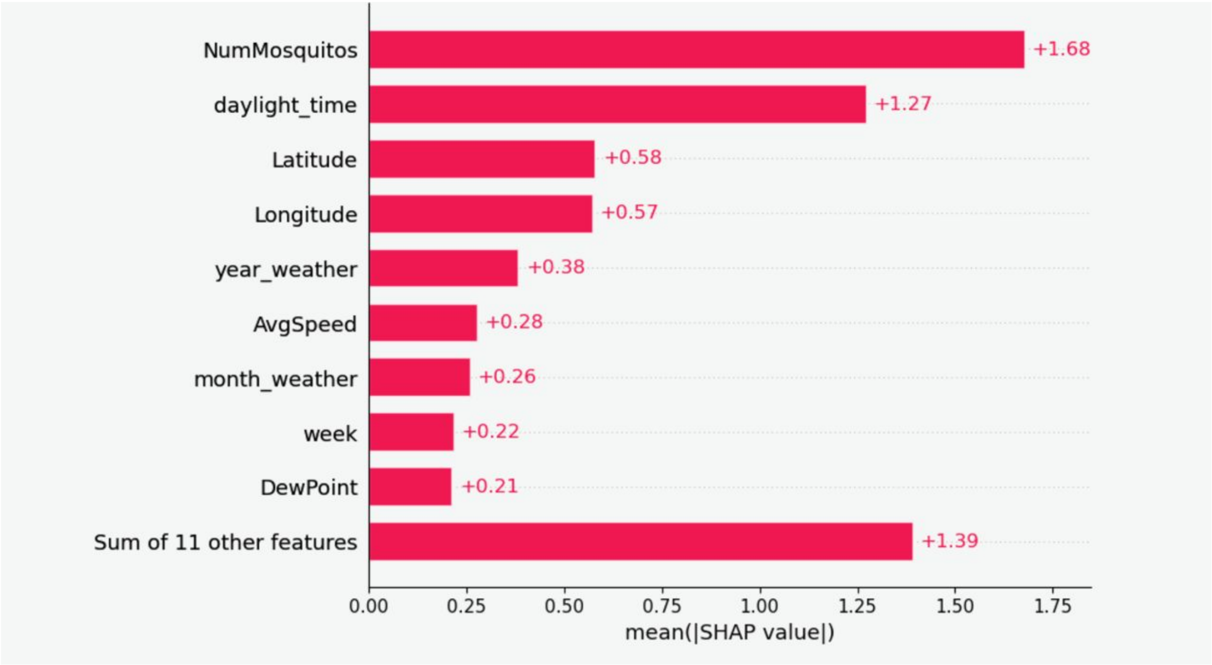
Model ROC AUC score : 0.8417

ROC curve is the graphical representation of the performance of model at various thresholds.

Key Features

Used the SHAP package on the best model to extract what were most important features.

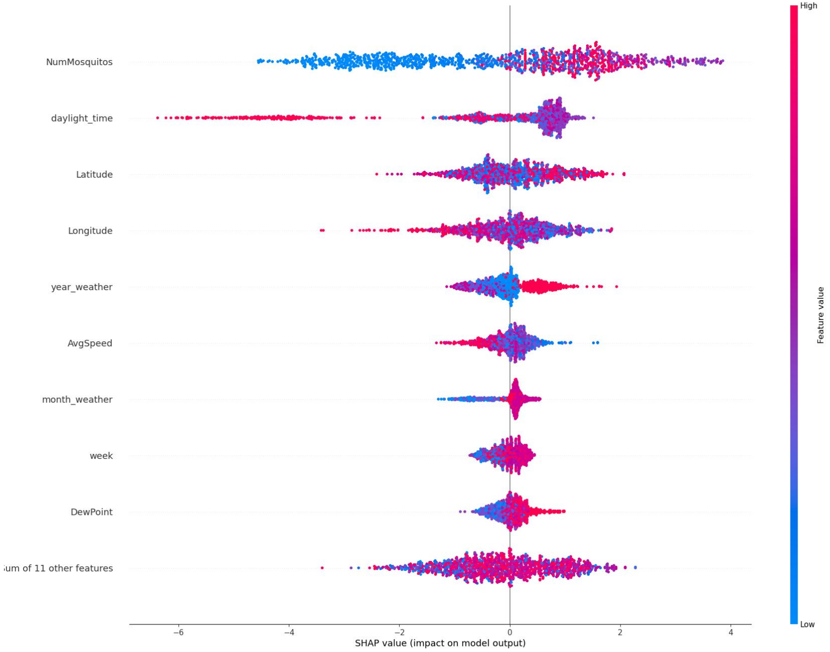
The top features were the number of mosquitoes, amount of daylight in a day, and location (latitude and longitude).



Further Analysis

Beeswarm plot is a plot that visualizes the distribution of each SHAP value for each feature. The points on the graph reveal the relationships of what features have an impact on the model's predictions.

For the number of mosquitoes, this plot reveals low negative SHAP values have low feature values whereas higher SHAP values have higher feature values.



Conclusion

1. Number of mosquitoes is the most important feature in the model for predicting the West Nile virus.
2. The daylight duration corresponds with mosquito activity periods and is the second most important feature.
3. Latitude and longitude ended up being key features for the model. These features can be used to focus on where West Nile virus is more likely to be present and curb transmission.

Future Work

- Remove the feature for “number of mosquitoes” as could significantly streamline West Nile virus prediction.
- Manipulate existing dataset features by shifting them over various days like sunrise, sunset, and daylight duration.
- Include additional data from other years to see how how existing years play a role into next year’s prediction.

References

1. Colpitts, T. M., Conway, M. J., Montgomery, R. R., & Fikrig, E. (2012). West Nile Virus: biology, transmission, and human infection. *Clinical Microbiology Reviews*, *25*(4), 635–648. https://doi.org/10.1128/cmr.00045-12
2. D'Amore, C., Grimaldi, P., Ascione, T., Conti, V., Sellitto, C., Franci, G., Kafil, S. H., & Pagliano, P. (2023). West Nile Virus diffusion in temperate regions and climate change. A systematic review. *Le Infezioni in Medicina : Rivista Periodica Di Eziologia, Epidemiologia, Diagnostica, Clinica E Terapia Delle Patologie Infettive*, *31*(1). https://doi.org/10.53854/liim-3101-4
3. Wendy Kan. (2015). West Nile Virus Prediction. Kaggle. https://kaggle.com/competitions/predict-west-nile-virus
4. *Culex Mosquito Life Cycle | CDC*. (2022, July 12). Centers for Disease Control and Prevention. https://www.cdc.gov/mosquitoes/about/life-cycles/culex.html
5. Bergstra, J., Yamins, D., Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. TProc. of the 30th International Conference on Machine Learning (ICML 2013), June 2013, pp. I-115 to I-23.
6. Shocket, M. S., Verwillow, A. B., Numazu, M. G., Slamani, H., Cohen, J. M., Moustaid, F. E., Rohr, J. R., Johnson, L. R., & Mordecai, E. A. (2020). Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23°C and 26°C. *eLife*, *9*. https://doi.org/10.7554/elife.58511
7. Vector Disease Control International. (2023, May 3). *West Nile virus: Education, public health, mosquito management*. https://www.vdci.net/vector-borne-diseases/west-nile-virus-education-and-mosquito-management-to-protect-public-health/#:~:text=West%20Nile%20virus%20is%20spread,feed%20from%20evening%20to%20morning