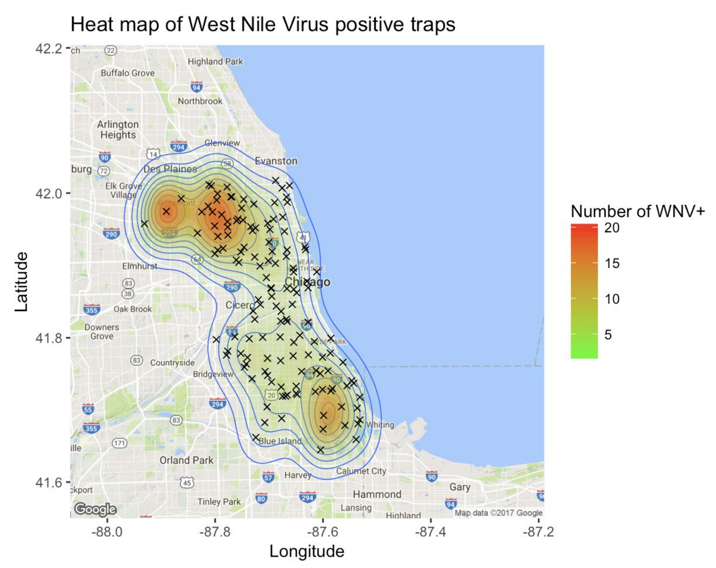
Fighting the West Nile Virus

Cameron Cruz: camcruz
Aayush Agrawal: aayush2k

Motivation

- Members of team have personally been afflicted with mosquito-borne diseases
- Improving disease detection methods contributes to improved public health
- Prediction of illness patterns allows for preventative measures to be taken



Problem

- Chicago has placed various mosquito traps throughout the city in order to test mosquitoes for the presence of WNV
- However, manually checking a trap and waiting for lab test results takes about a week, significantly delaying availability of actionable information

Goal: Reduce detection-time by making same-day predictions on where WNV will be present using readily available data.

Data

- Data from Chicago WNV Surveillance Reports recorded between 2007 and 2014
  - ~100 traps throughout Chicago
  - 10,000 total observations
  - Unbalanced: 95% WNVPresent=0, 5% WNVPresent=1
- Daily weather data from NOAA for 2007-2014
  - Weather stations: O'Hare International Airport and Midway International Airport
  - 13 total weather features

Mosquito Trap Data (location, time, mosquito species) + Closest Weather Station Data (temperature, dew, precipitation) ➡ WNV Present or Not

Approach

Models

- Logistic Regression with L1 Lasso Penalty (LR)
- Decision Tree (DT)
- Random Forest (RF)
- Support Vector Machine (SVM)
- Neural Network (NN)

During training, we adjusted weights using inverse proportions of class frequency to account for class imbalance.

For example, balanced binary cross entropy (alpha=9.5, beta=0.5)

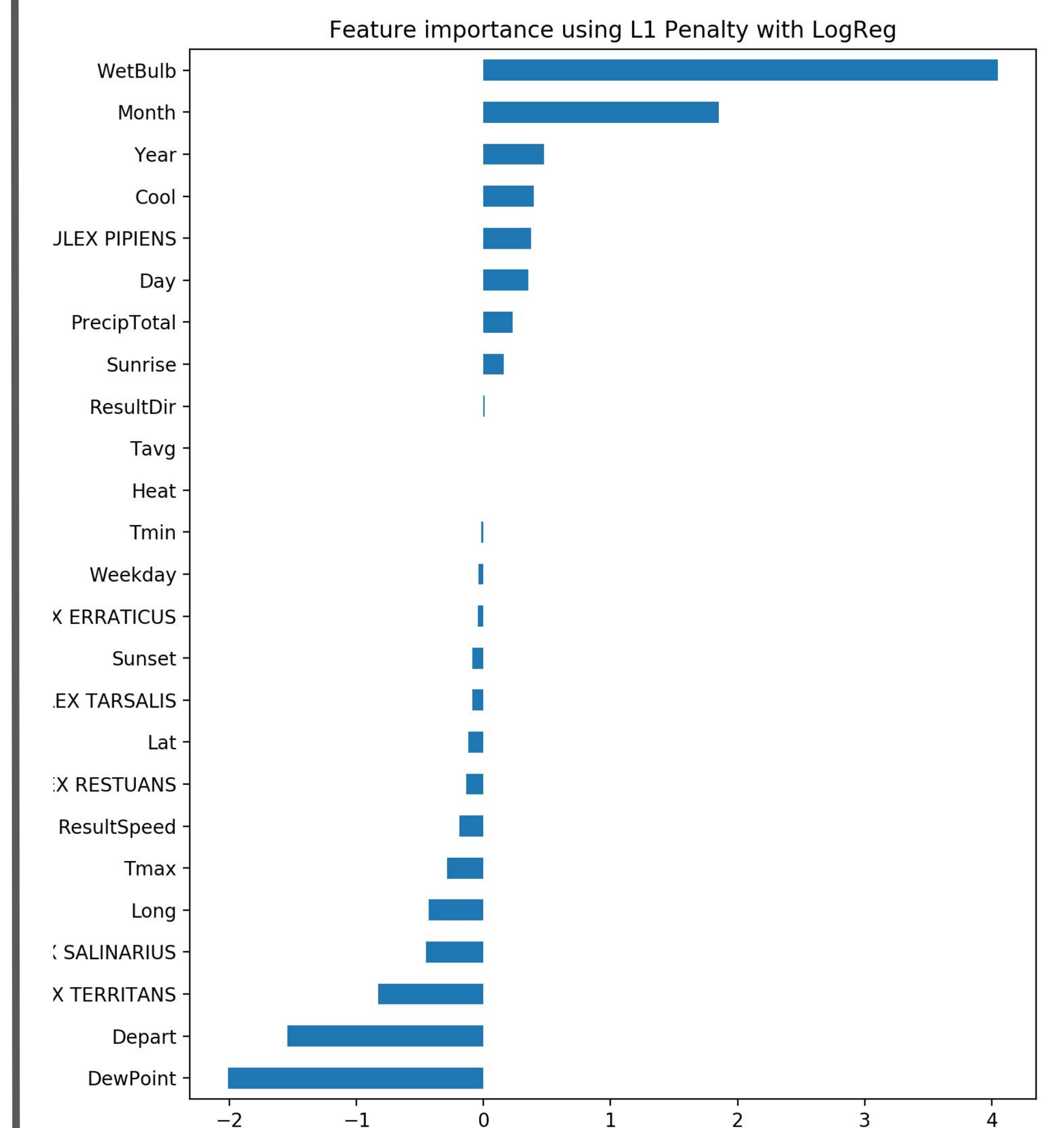
$$\alpha\,[labels \;\times - log(\hat{p})] + \beta\,[(1 - labels) \;\times - log(1 - \hat{p})]$$

Results

	LR	DT	RF	SVM	NN
ROC	0.79	0.80	0.83	0.83	**0.85**
Precision	0.13	0.15	**0.16**	0.15	0.14
Recall	0.79	0.67	0.74	0.79	**0.84**
F1-score	0.22	0.24	**0.27**	0.25	0.24

Ablation Study with NN

	With All	Without Lat/Lon	Without Species	Without Weather	Without Date
ROC	0.85	0.84	0.84	0.84	0.85
Precision	**0.14**	0.10	0.11	0.12	0.12
Recall	0.84	0.93	**0.95**	0.90	0.91
F1-score	**0.24**	0.17	0.19	0.21	0.21



Top 5 Features for RF

Feature	Significance
Sunrise	0.137
Sunset	0.129
CULEX PIPIENS	0.087
Month	0.078
Dew Point	0.072

Analysis

- Precision is generally low for all methods – likely due to traps sharing weather features but not all having WNV
- Location and species are more important to NN than weather and date
- Sunrise/sunset calculated times may be indicative of season
- CULEX PIPIENS is most likely to be carrying WNV.

Conclusions

- We value recall over precision since false negatives have more consequences than false positives in disease detection, therefore NN is preferred.
- Additional features on transportation, imports (which country goods are from), and fine-grained weather would help differentiate observations better.

Works Cited

Ruiz, M.O., Chaves, L.F., Hamer, G.L. *et al.* Local impact of temperature and precipitation on West Nile virus infection in *Culex* species mosquitoes in northeast Illinois, USA. *Parasites Vectors* 3, 19 (2010) doi:10.1186/1756-3305-3-19

"West Nile Virus." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 12 Nov. 2019, https://www.cdc.gov/westnile/index.html.

"West Nile Virus (WNV) Mosquito Test Results." *City of Chicago Data Portal*, City of Chicago, 26 Sept. 2019, https://data.cityofchicago.org/widgets/jqe8-8r6s.