



# Predicting West Nile Vectors In Chicago

by Amy Zhou

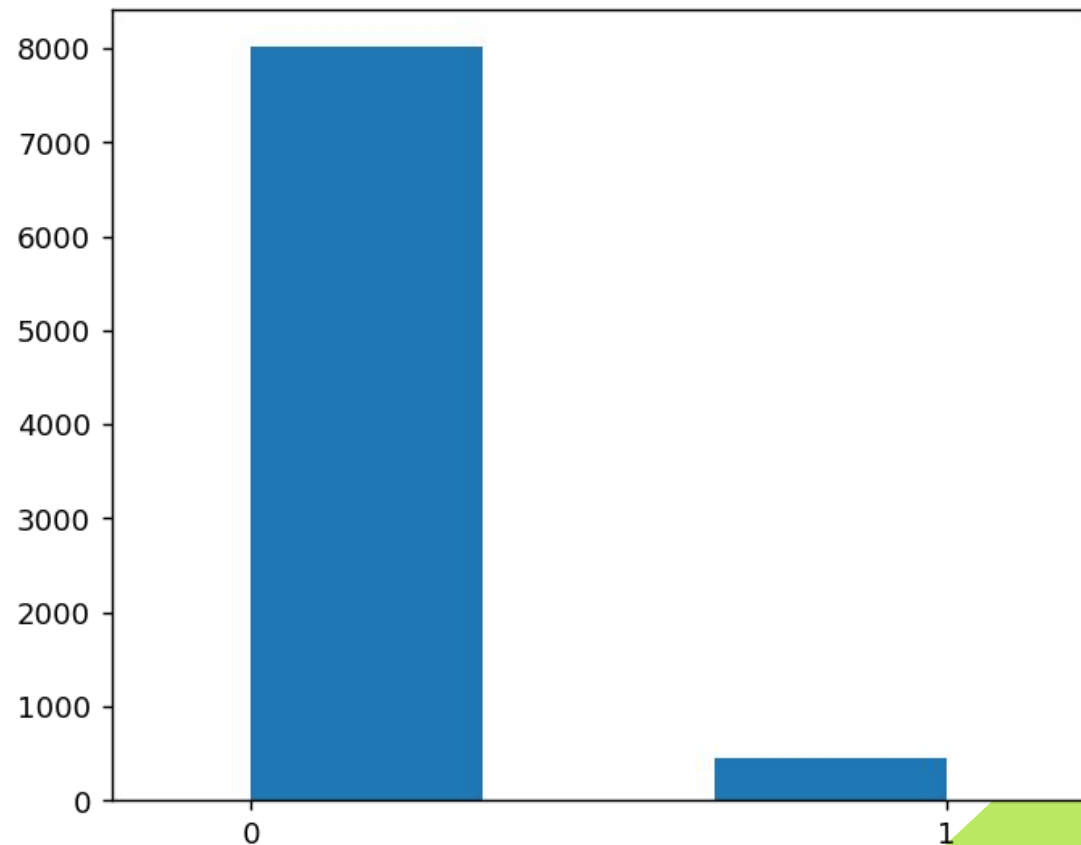
# Problem

- West Nile carrying mosquitos in Chicago
- Need to optimize where to spray to prevent outbreaks

# Data Collection

- 2007-2013, May-Oct
- Mosquito trap data
- Weather data
- Spray effort data

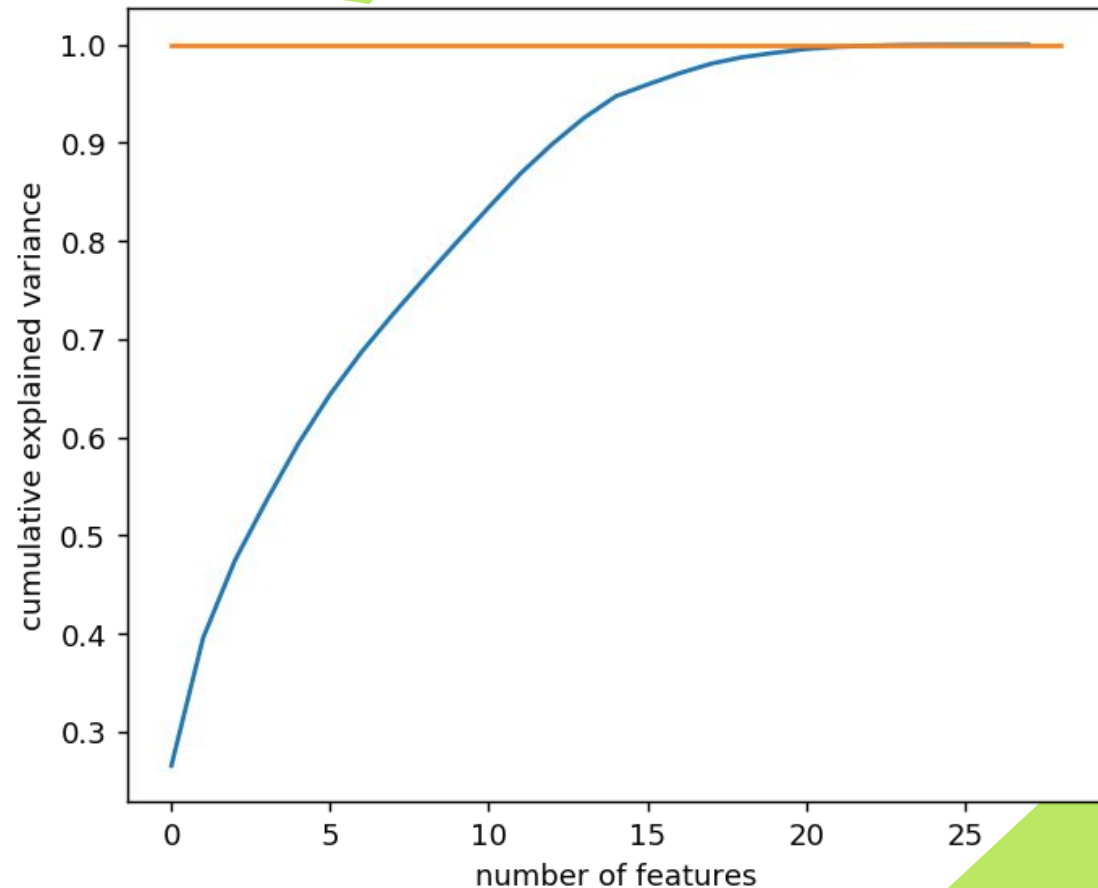
Distribution of Positive (1), Negative (0) Labels



# Approach

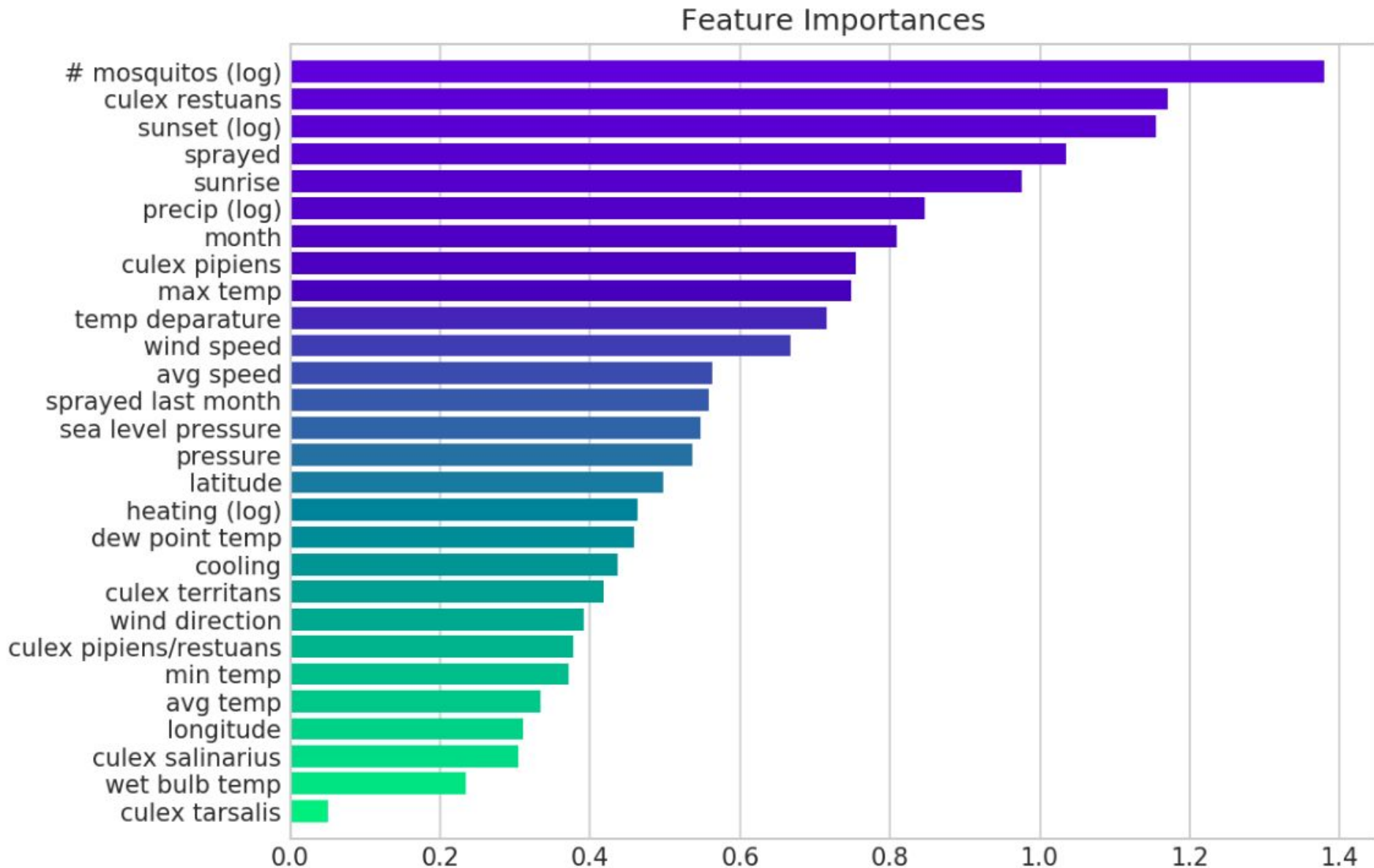
- Feature engineering
- Resample due to imbalanced data
- PCA to address multicollinearity
- Model/param selection by highest recall score

Principal Component Analysis



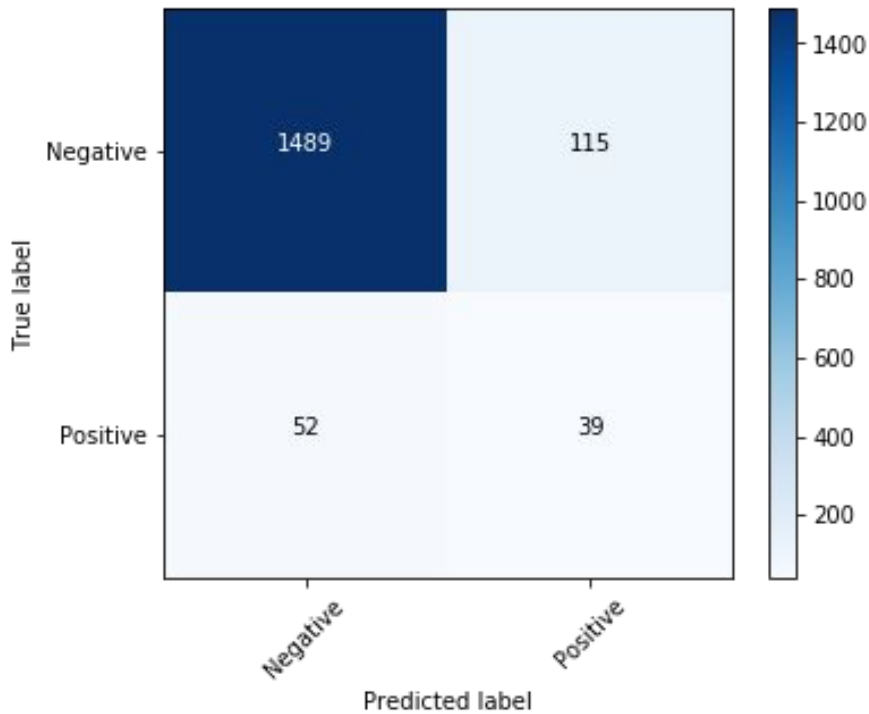
# PCA + Random Forest Model Feature Importances

Components: 20, Estimators: 60, Max depth: 100, Min samples/leaf: 1, Min samples to split: 2

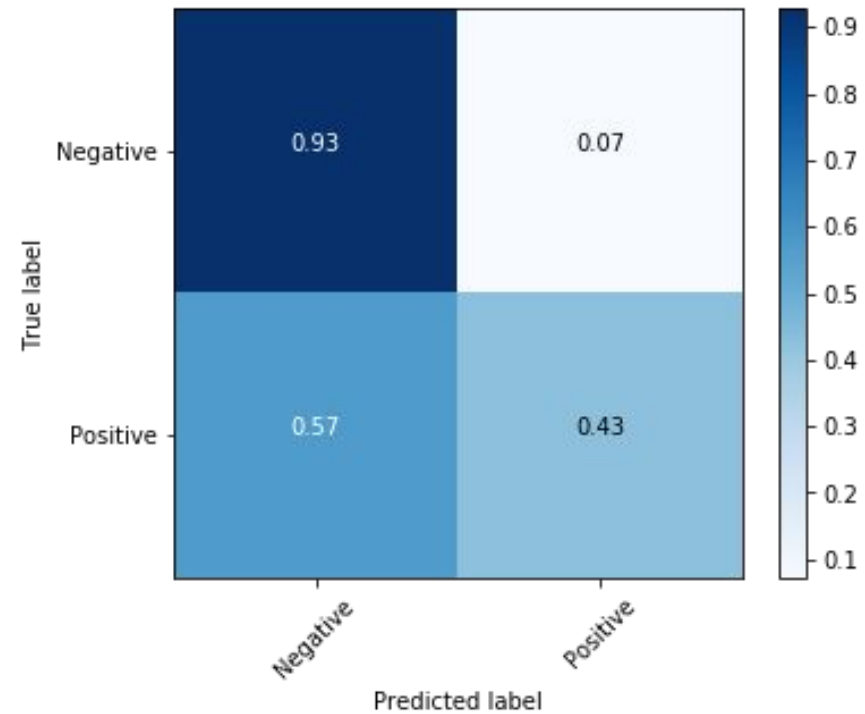


# Outcome

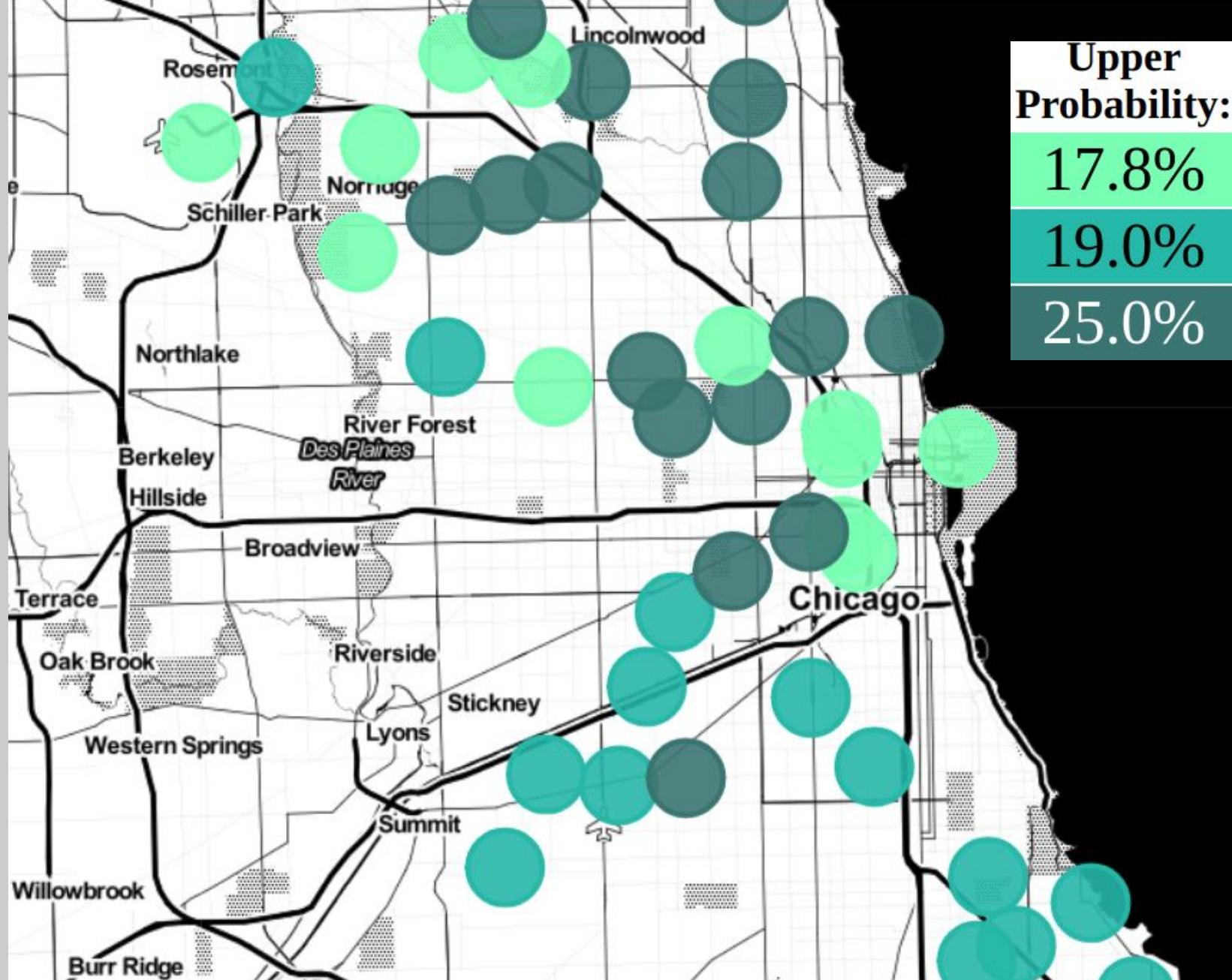
Confusion matrix



Normalized confusion matrix

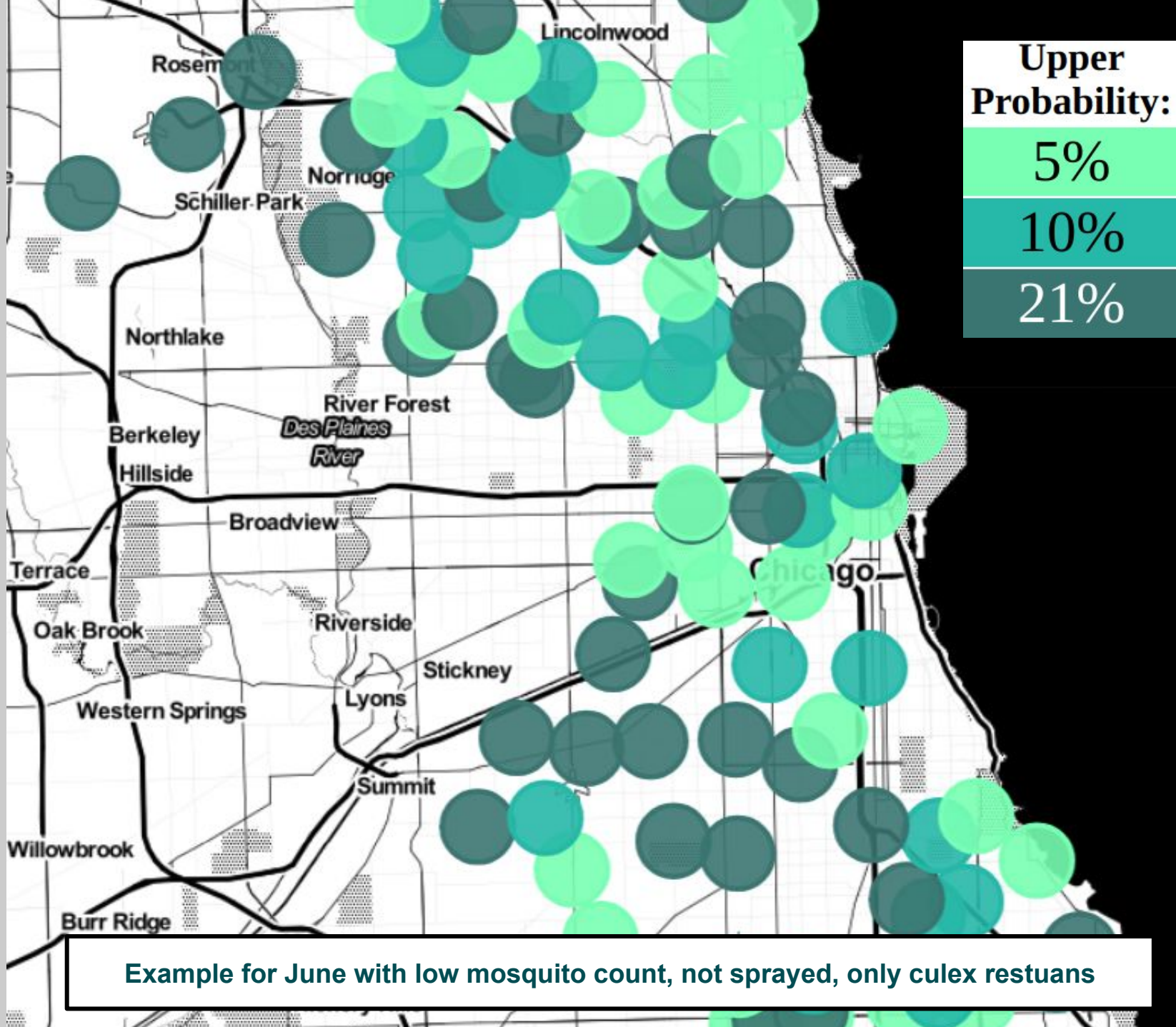


Resampling + PCA + random forest model performance on test data



Example for May with average mosquito count, sprayed, all mosquito species types







## Further Steps

- Adjust resample ratio, PCA params
- Decrease features, optimal pruning
- Include label from previous time period
- Represent species as percent of count
- Include additional weather conditions