Project 4

Personal Private M

To Do

- Push this plan to github (an inceptive to do).

Doing

Done

- Create a shared repository.
- Basic EDA (.info, .head, etc.)
- Geographical visualization utilizing Tableau.
- Create a baseline model.
- Submit baseline model to Kaggle.
- Merge weather document onto the training data.
- Clean up the test data: change object types, merge weather document onto the testing data.
- Create a file that outputs information to CSV file.
- Change the trap designations to integers.
- Create West Nile Analysis document.
- Dummy the mosquito names.
- Put all variables in a Random Forest model and check score.
- Consider the confusion matrix based on the original Random Forest model.

Oversample train data in order to get more West Nile virus positives.

Re-run random forest based on oversampled train data.

Divvy up presentation responsibilities.

Make second submission based on random forest model.

Create for loop to iterate through longitude and latitude in the train data and determine the minimum distance from any given one latitude and longitude to another given latitude and longitude.

Run through this for loop for a very long time.

Update model based on predictive minimum distance feature.

Separate the distances by year so that the minimum distance is calculated in relation to the other distances within that year.

Create interactive terms for our distance and type of mosquito.

Iterate through different thresholds on the way to maximizing sensitivity.

Add weather variables to our models.

Consider the issue of "leakage" -- the spillover of identical data onto multiple rows due to greater than fifty mosquitoes at that location, trap, etc.

Put data from both stations into model and run feature importance on the groupings of data so that we can determine which groupings to eliminate from the model.

Consider how and whether to calculate spray data into our predictions.

Look at features that were misclassified in order to determine why our training and testing data are performing so differently.

Comment out code.

Markdown for code.

Create structure for presentation.

Practice presentation.