Detecting Cargo Ships Using Machine Learning

By
Chris Murdter

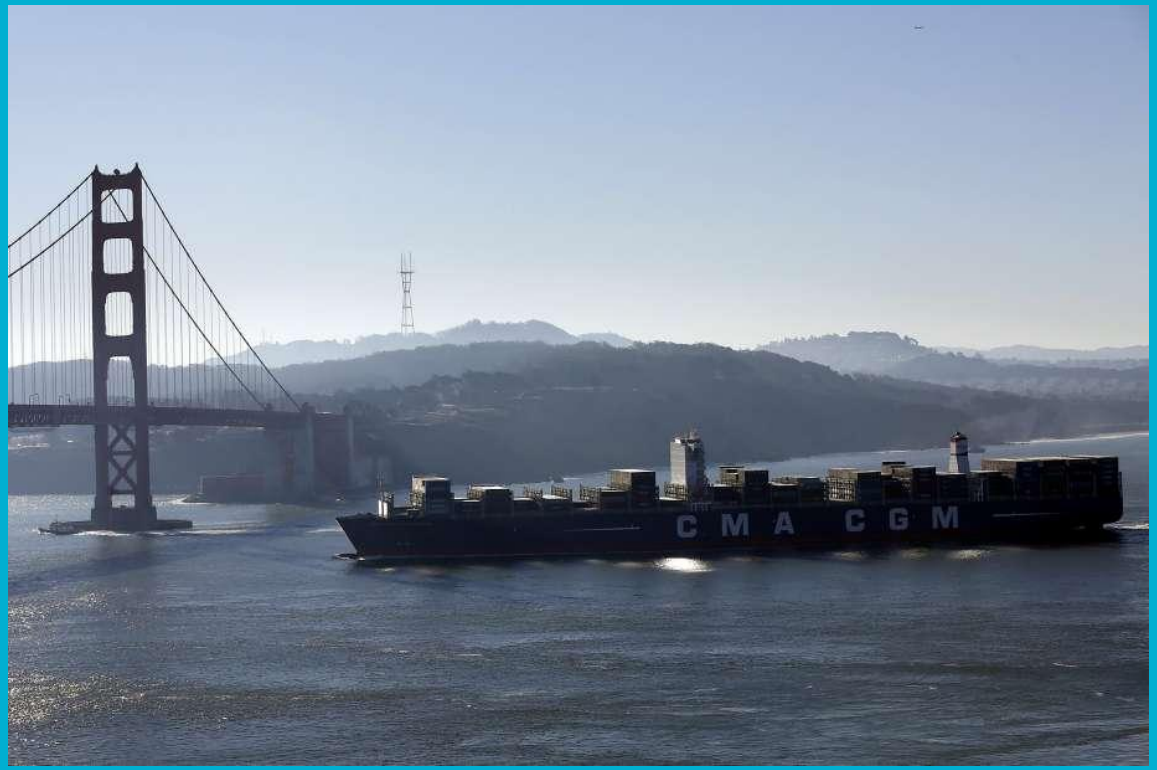
Problem

- How to improve tracking ship behavior?
- Observe ships that have gone dark/turned off AIS tracker
- oil spills, piracy, trafficking

- $50 billion a year

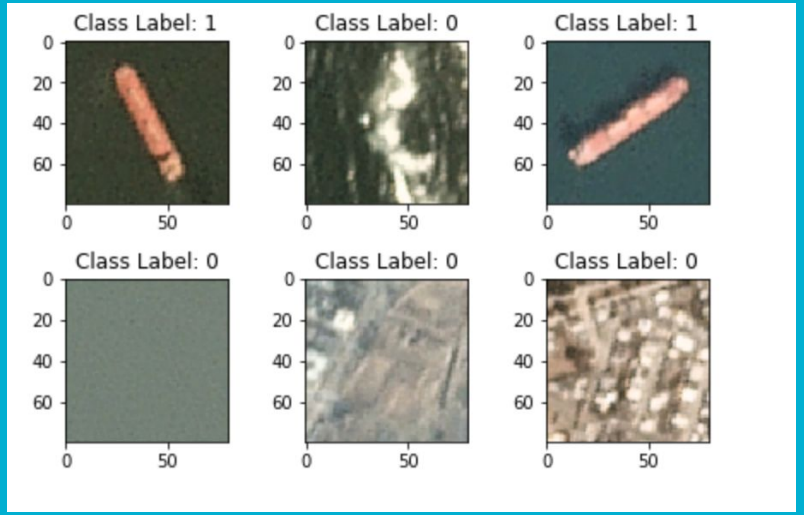
Data

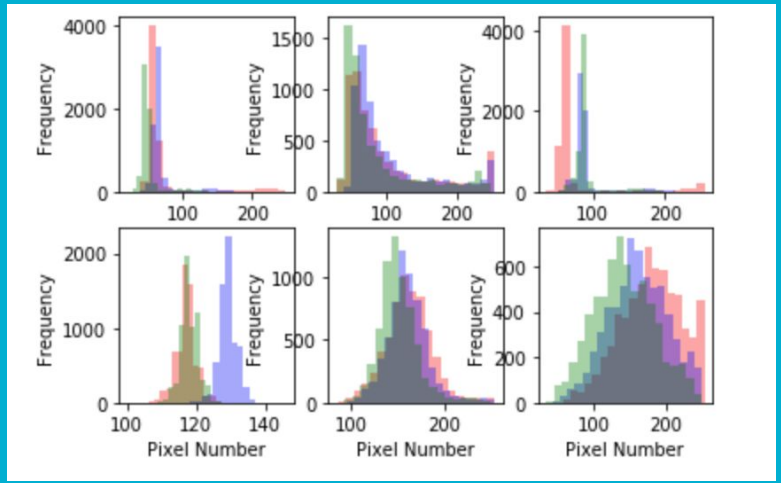
- Data comes from Kaggle
- SF Bay Area satellite images
- 4,000 80x80 RGB images
- Labeled “ship” or “no ship”
- Included longitude/latitude of image centerpoint



EDA

- Data column has 19,200 integers in a list. 6400 each for red, green and blue.
- Looked at RGB distributions to find differences in images.
- Possible False positives or False negatives





Logistic Regression

- logistic regression is a model that can be applied to assign observations to a discrete set of variables
- the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.)

```
1  LR = LogisticRegression()
2  reg = LR.fit(X_train, Y_train)
3  pred_LR = LR.predict(X_test)
4  score = LR.score(X_test, Y_test)
5  print(score)
```

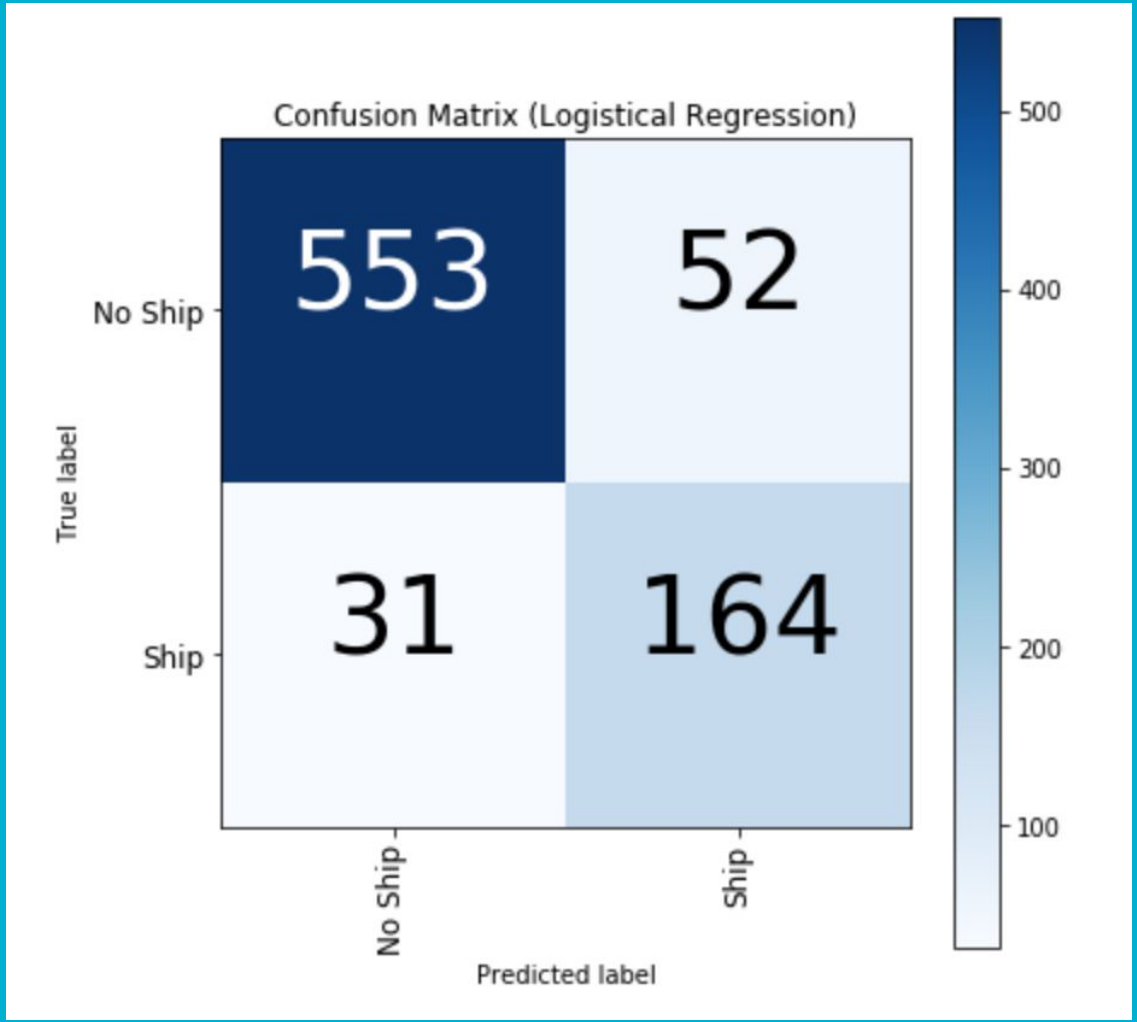
XGB Classifier

- Implements gradient boosting
- predict the residuals or errors of prior models and then added together to make the final prediction.
- It is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models.

```
1  XGB = XGBClassifier()
2  reg = XGB.fit(X_train, Y_train)
3  pred_XGB = XGB.predict(X_test)
4  score = XGB.score(X_test, Y_test)
5  print(score)
```

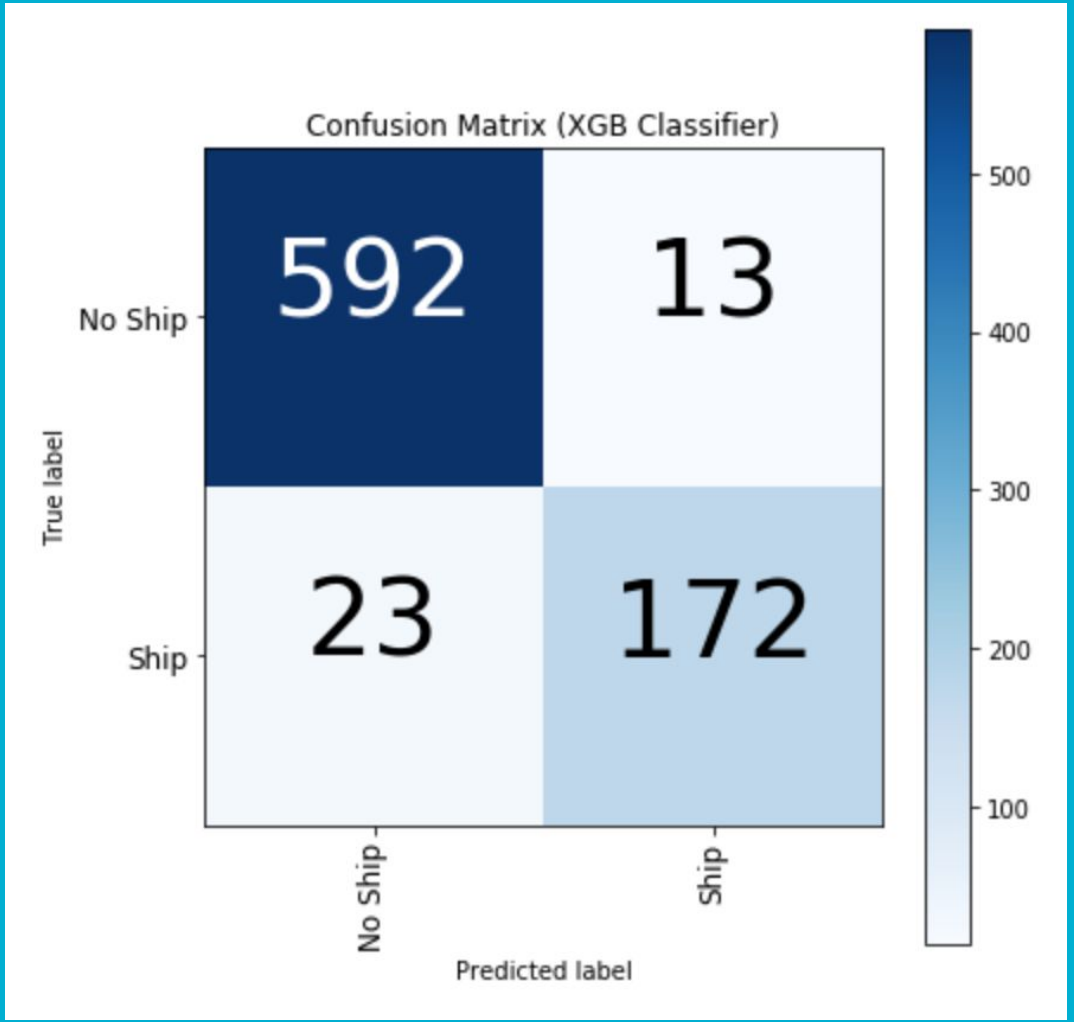
Logistic Regression Results

- Mean accuracy of 89.6%
- Made a confusion matrix to see how many images were classified correctly and incorrectly



XGB Classifier Results

- Mean accuracy of 95.5%
- Noticeable difference from the Logistic Regression results
- Large decrease in false positives



Final Visualization

Next Steps

- Apply Image segmentation
- Segmenting ships in images could increase accuracy and reduce false negatives/positives
- Nice highlighted visualization

Questions?

Comments?

concerns?