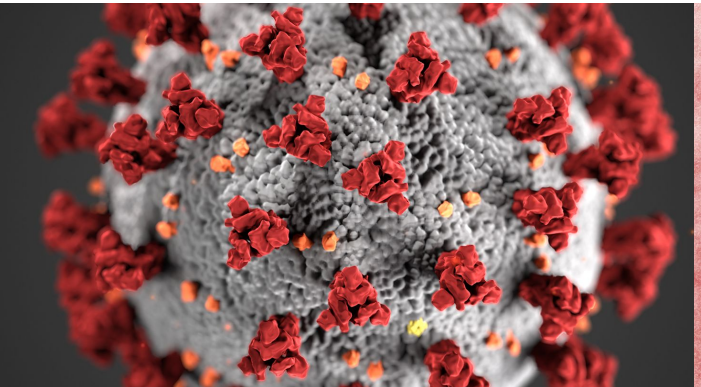
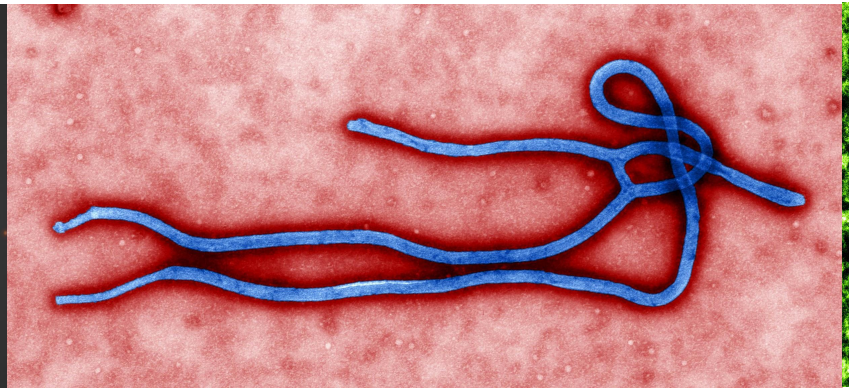
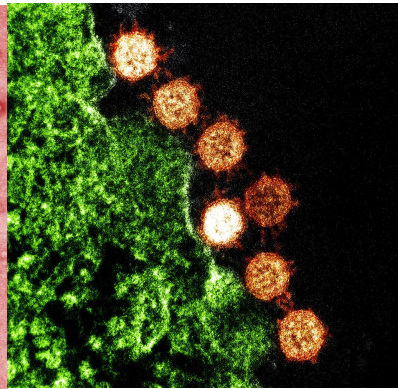
Encouraging Vaccination Compliance: A Model to Target Non-Vaccinators

Amy Sillman
Metis Data Science Bootcamp
Project 3 Pressentation
February 10, 2021

Problem and Approach

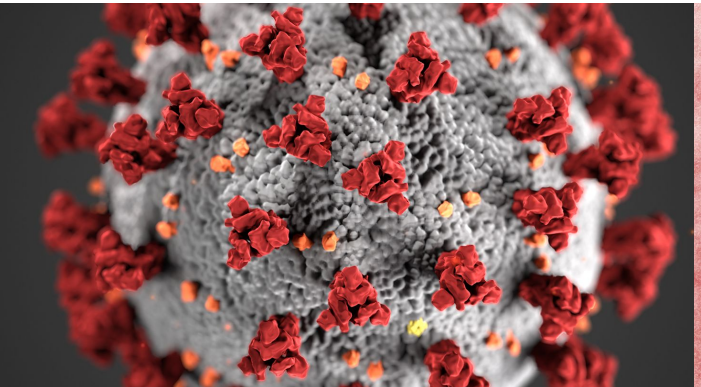
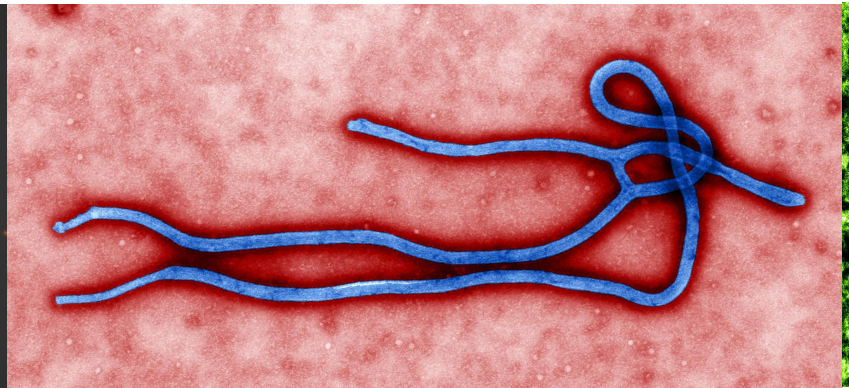
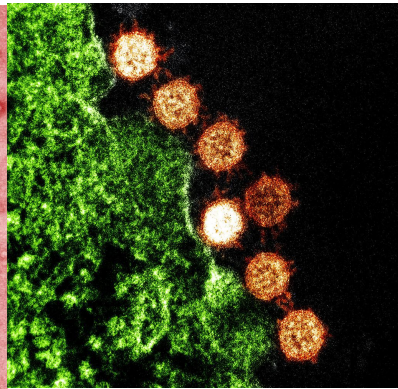
- New infectious diseases arise periodically
- Mass vaccination is crucial for controlling the spread of infection
- Improved public health campaign necessary to encourage vaccination

Problem and Approach

- New infectious diseases arise periodically
- Mass vaccination is crucial for controlling the spread of infection
- Improved public health campaign necessary to encourage vaccination

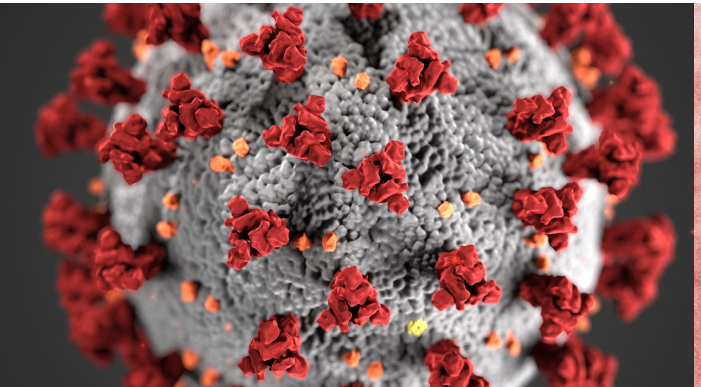
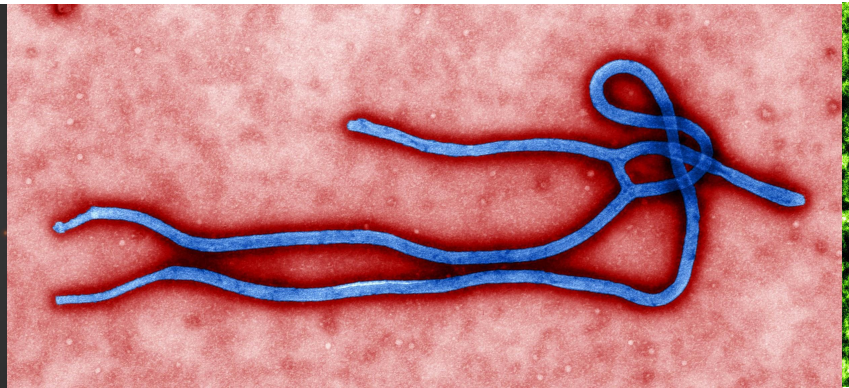
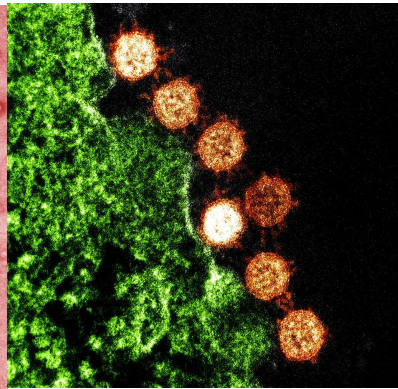
Data-driven approach to target people less likely to vaccinate

Description of Data

Source:

- Data originally from US DHHS National 2009 H1N1 Flu Survey,
- Accessed from DrivenData competition website

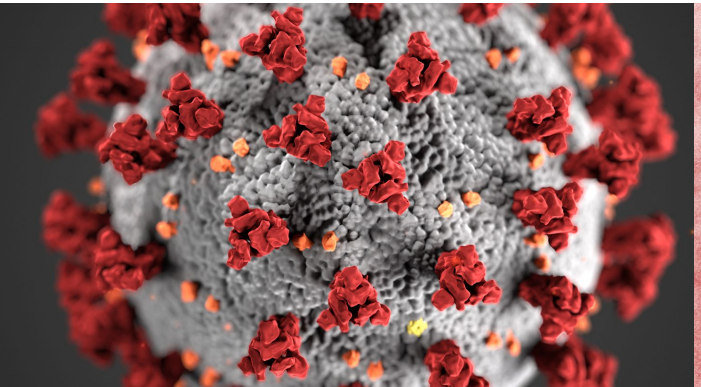
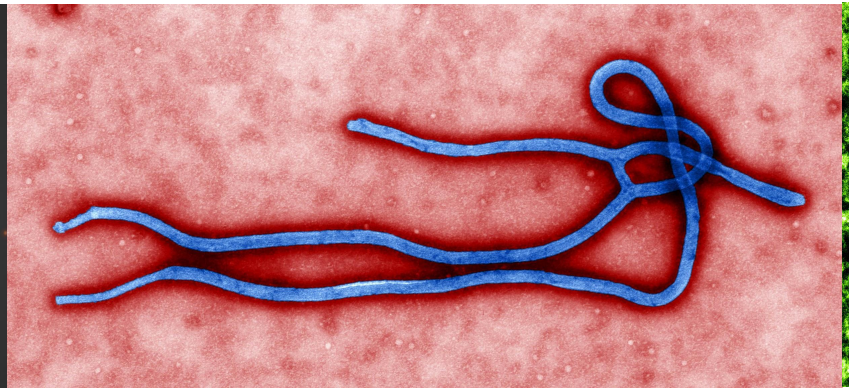
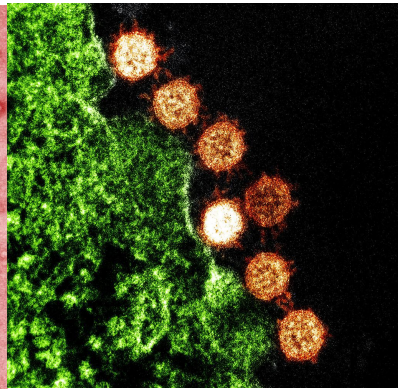
Description of Data

Source:

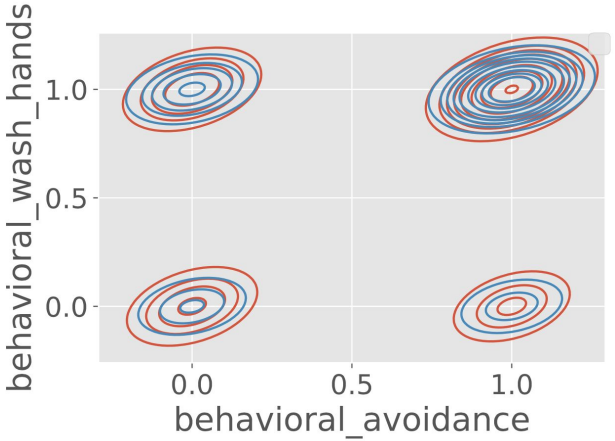
- Data originally from US DHHS National 2009 H1N1 Flu Survey,
- Accessed from DrivenData competition website

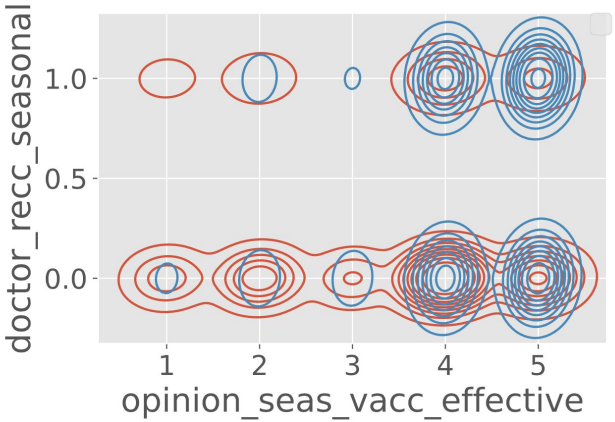
~26K survey respondents answered the y/n target vaccine questions (Training Set)

- 35 features: demographic, health, behavioral factors
- ~12-14K respondents did not answer the Health Insurance or Employment questions; removed these features
- Removed records with NaN values, leaving ~**20K** records in the Training Set
- Categorical features were encoded using dummy variables

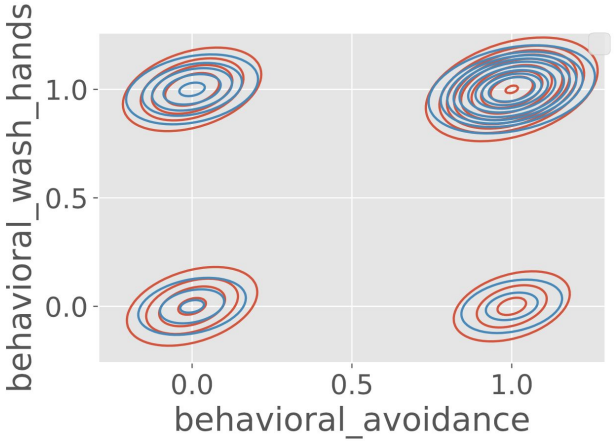
Insights from Exploratory Data Analysis

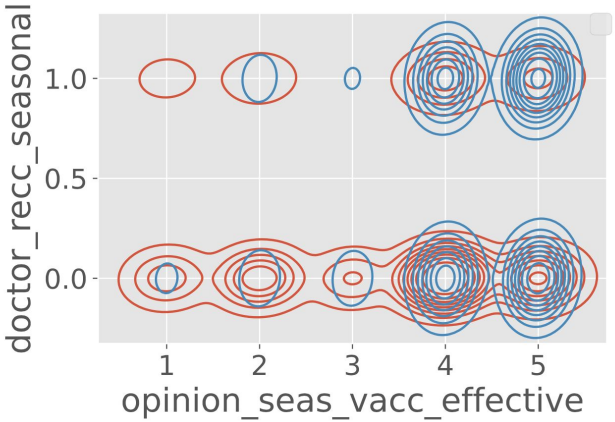




Blue: Seasonal Flu Vaccine Compliant

Insights from Exploratory Data Analysis





Blue: Seasonal Flu Vaccine Compliant

Features associated with seasonal flu vaccine compliance:

- Frequent hand washing
- Avoiding sick people
- Doctor recommended the seasonal vaccine
- Opinion that the seasonal flu vaccine is effective

Approach

Campaign will target:

- **Non-vaccinators (true negatives)**
- **People “on the fence” (false negatives)**

Approach

Campaign will target:

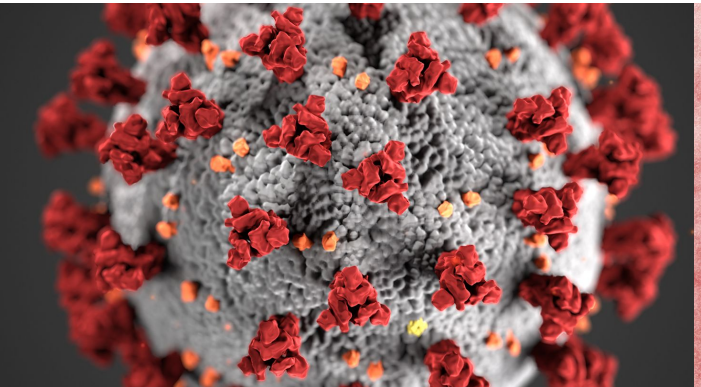
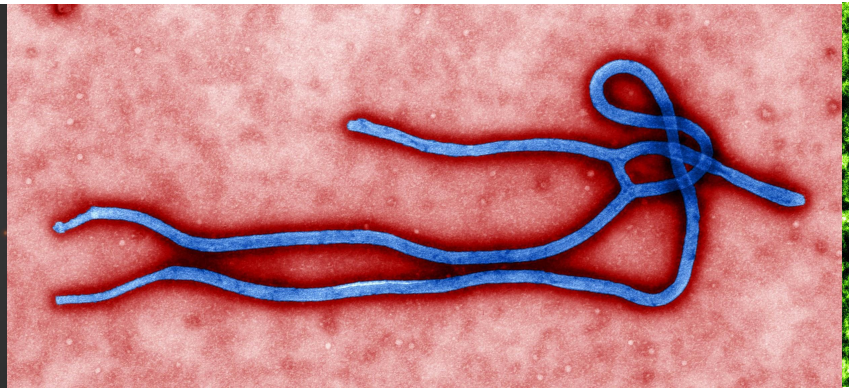
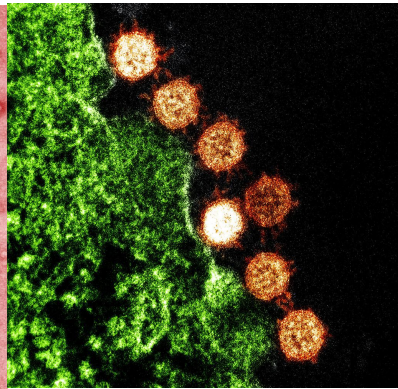
- Non-vaccinators (true negatives)
- People “on the fence” (false negatives)

Therefore:

- Increase precision, so that vaccinators are not targeted
- Predicting false negatives is fine-- may have characteristics of non-vaccinators, targeting may be beneficial
- False positives should be minimized, so these people are properly identified and targeted

Model Exploration

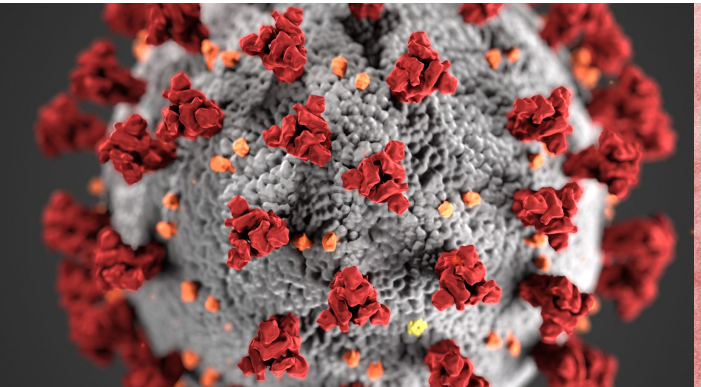
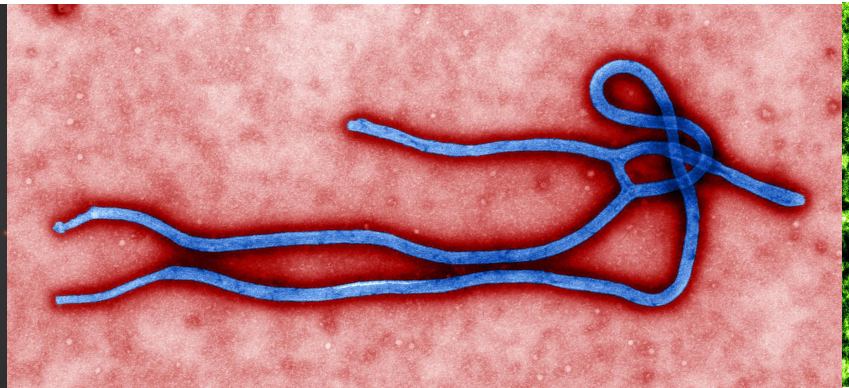
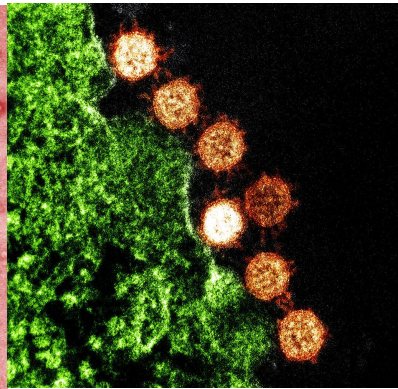
- Decision Tree
- Random Forest
- Naive Bayes Bernoulli
- Logistic Regression

Model Exploration

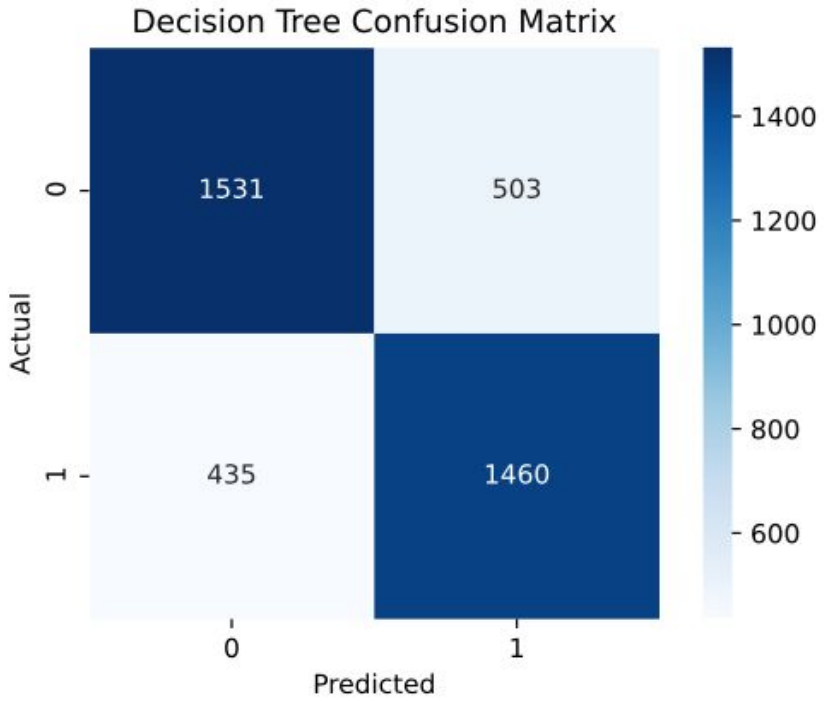
- Decision Tree
- Random Forest
- Naive Bayes Bernoulli
- Logistic Regression

Training data was split into training and validation sets for model exploration

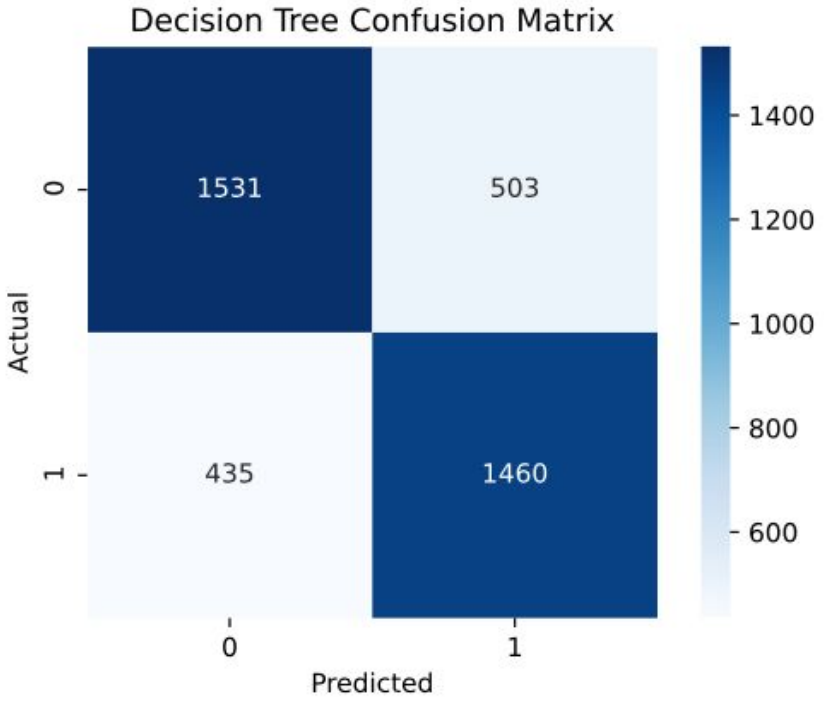




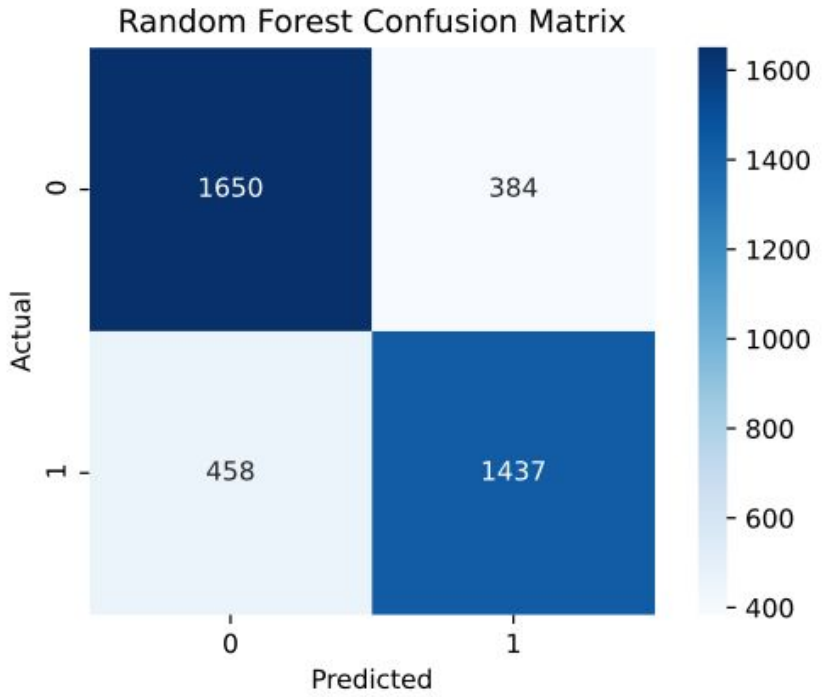
Decision Tree and Random Forest

Decision Tree and Random Forest

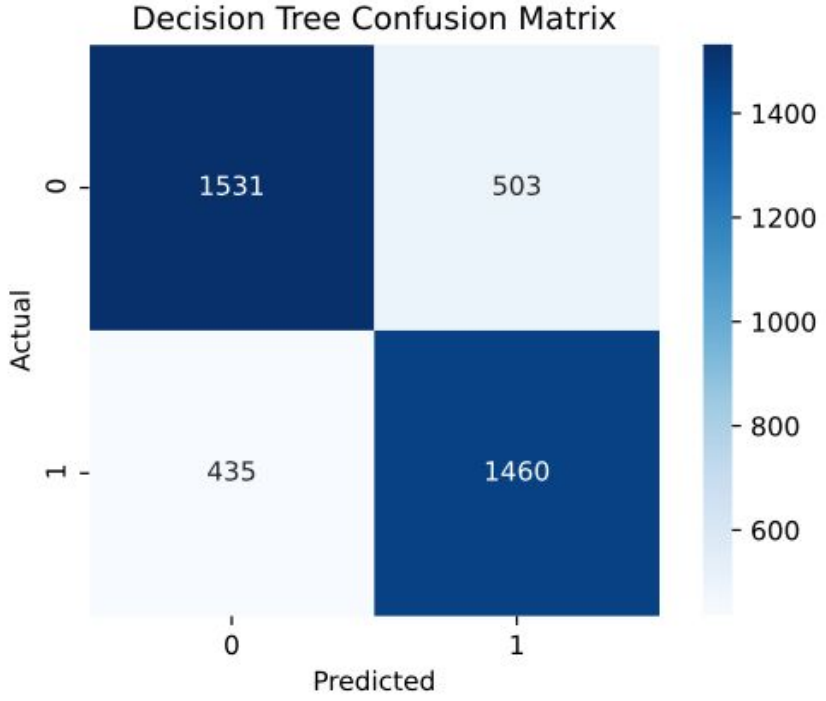


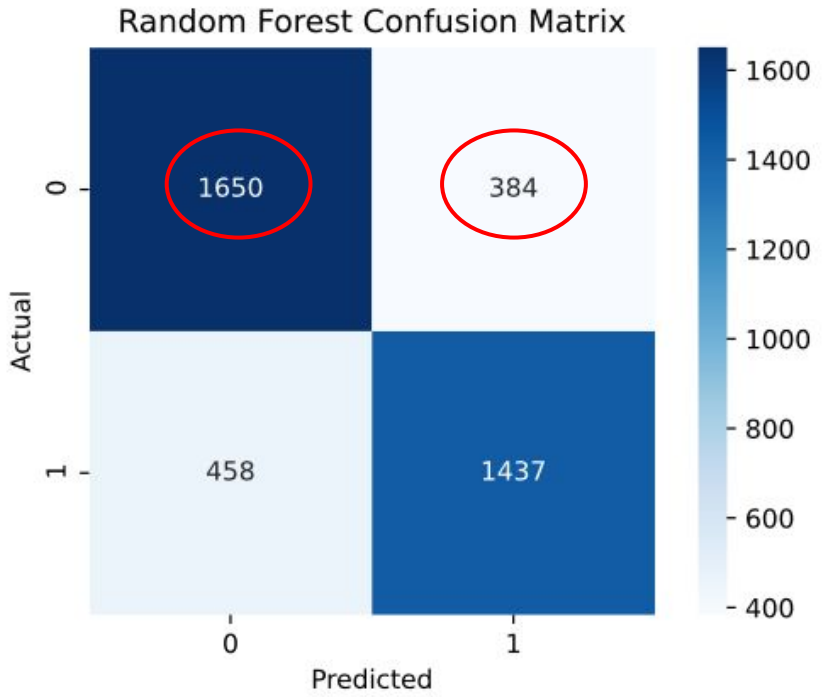
Decision Tree and Random Forest





Decision Tree and Random Forest

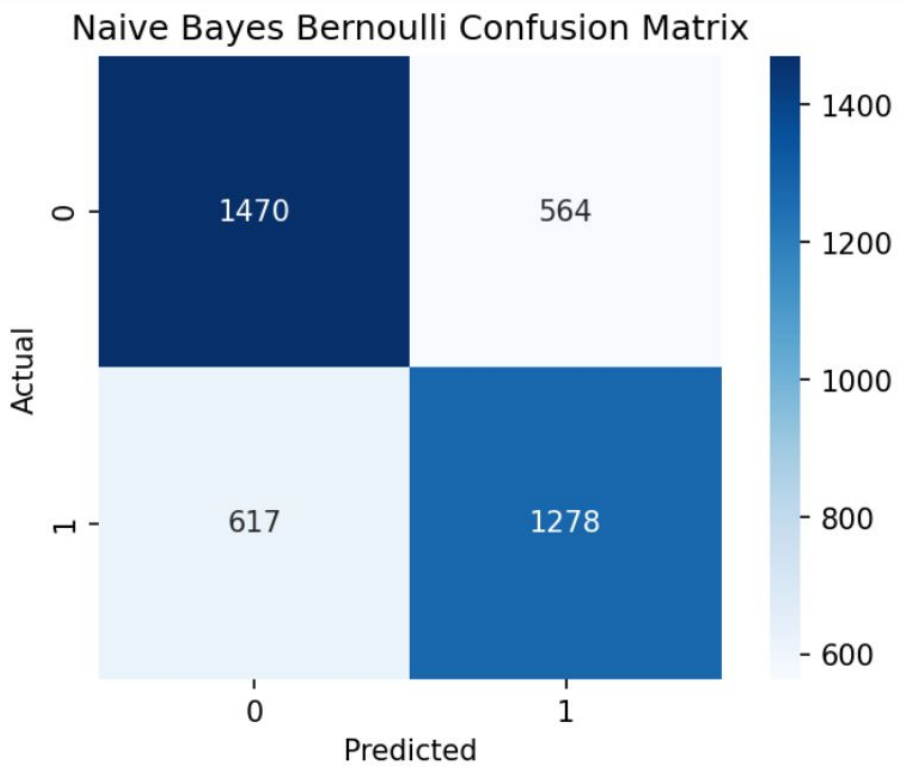




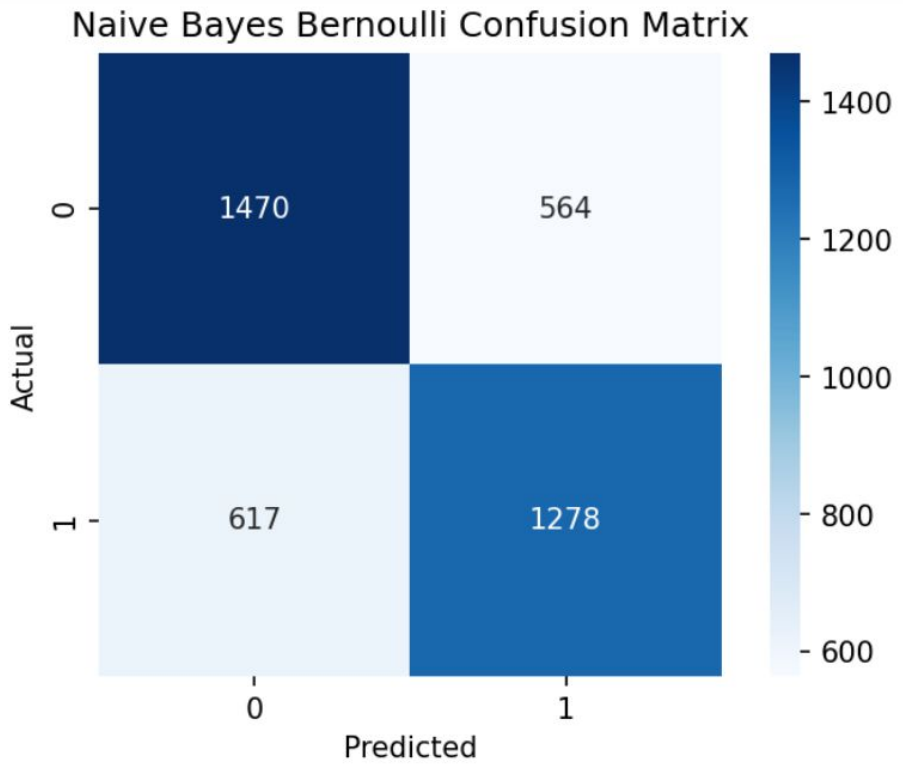
Random Forest performs slightly better than Decision Tree
(fewer false positives, more true negatives)

Naive Bayes (Bernoulli)

Naive Bayes (Bernoulli)

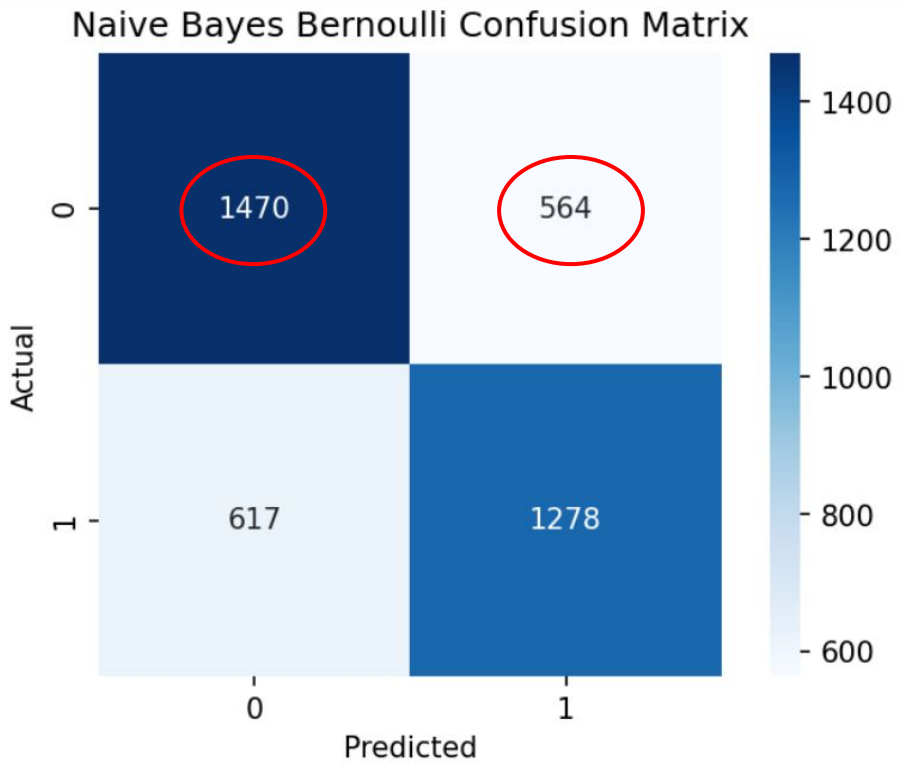


Naive Bayes (Bernoulli)



Naive Bayes (Bernoulli) performs slightly worse than Random Forest

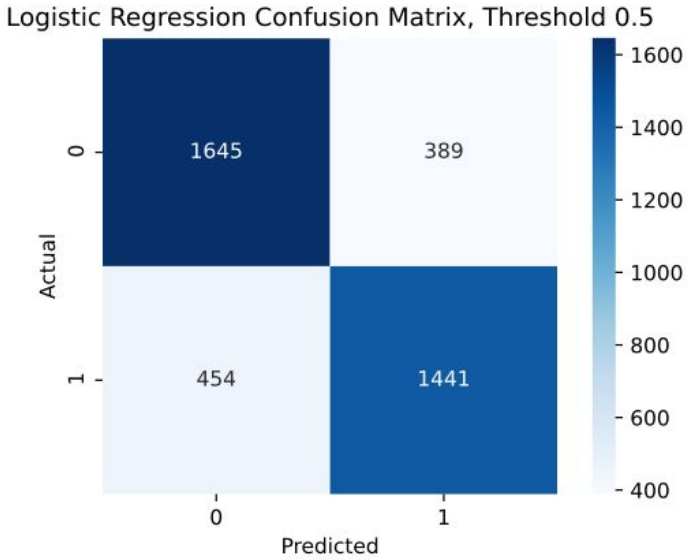
Naive Bayes (Bernoulli)



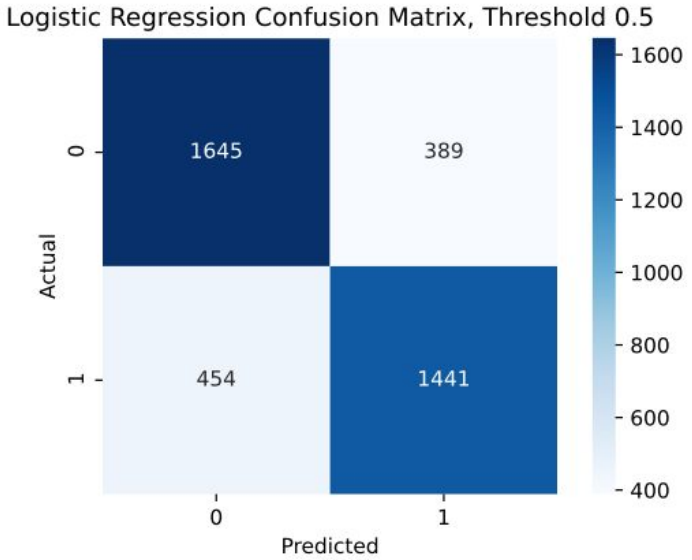
Naive Bayes (Bernoulli) performs slightly worse than Random Forest (more false positives, fewer true negatives)

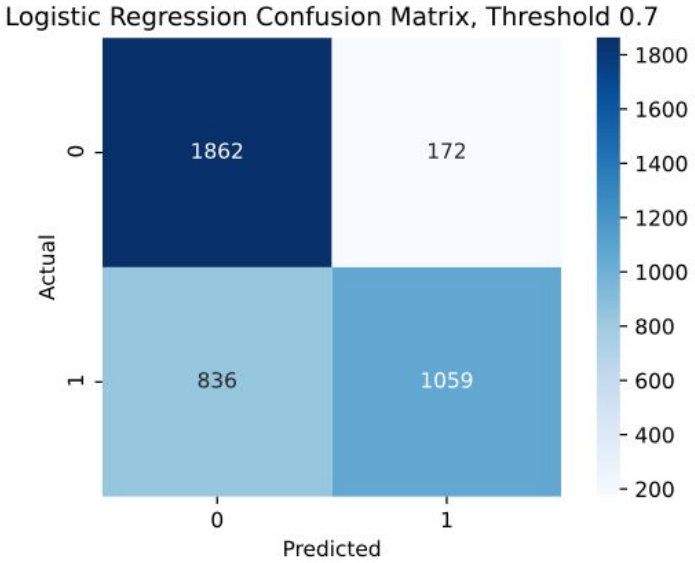
Logistic Regression

Logistic Regression

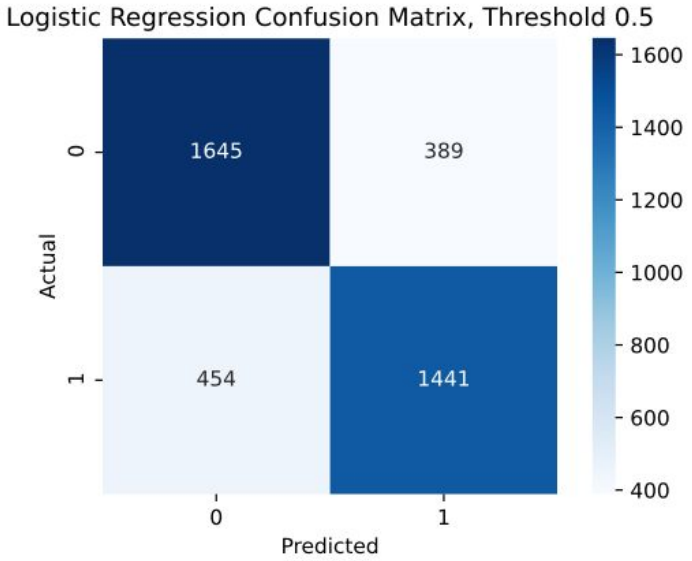


Logistic Regression

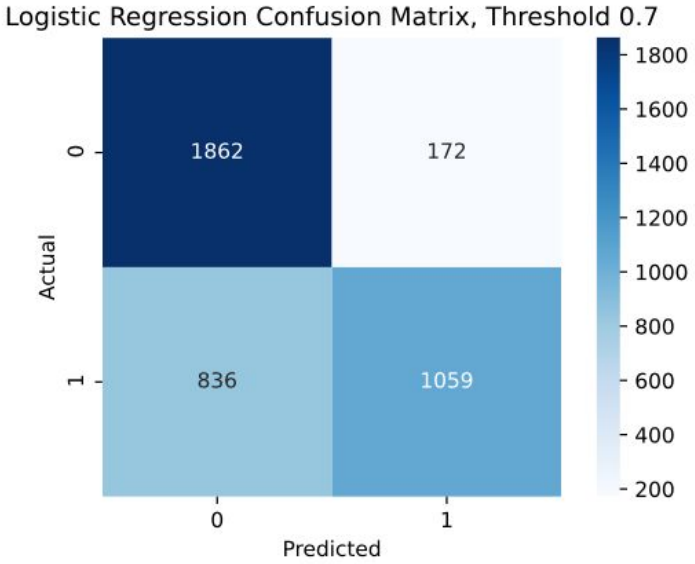




Logistic Regression

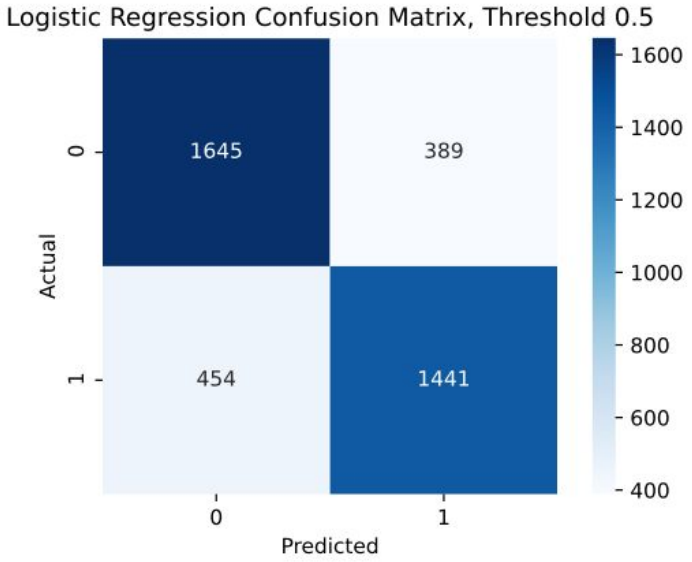


Precision: 0.7874

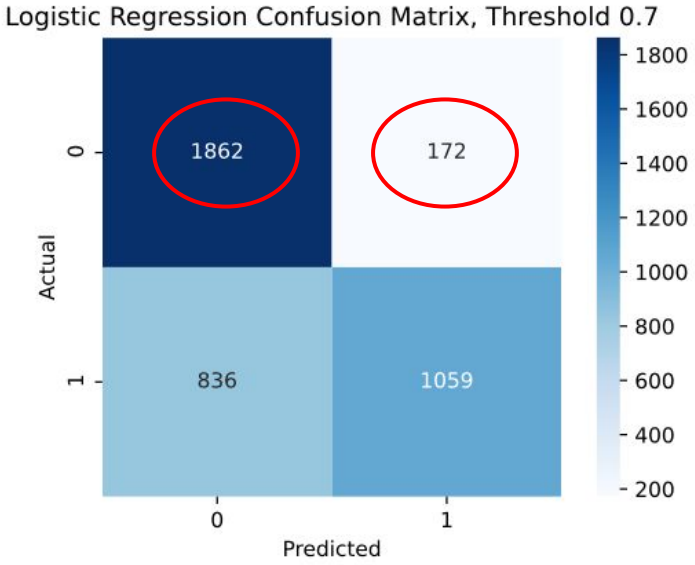


Precision: 0.8603

Logistic Regression



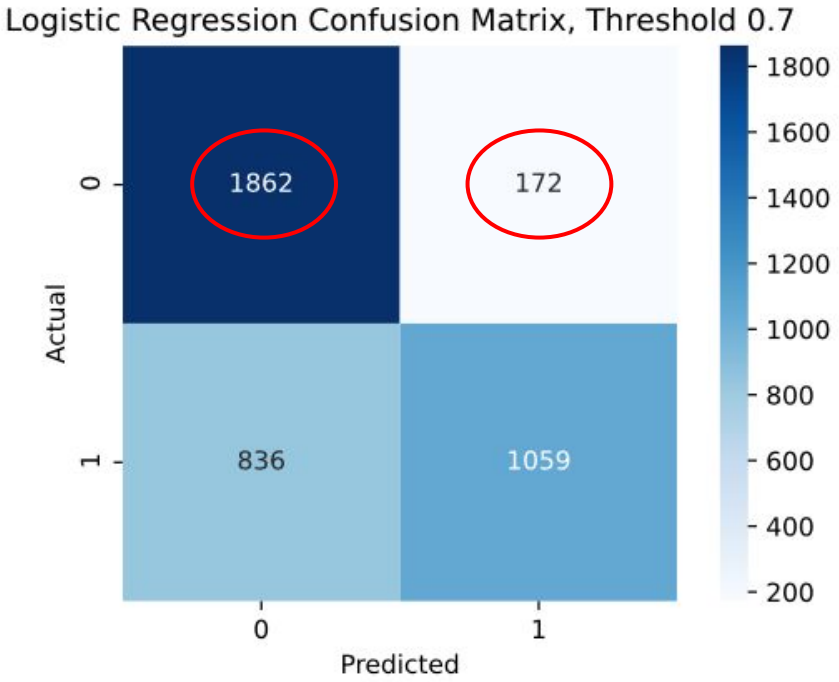
Precision: 0.7874



Precision: 0.8603

Logistic Regression Model, Threshold 0.7
(fewest false positives, more true negatives)

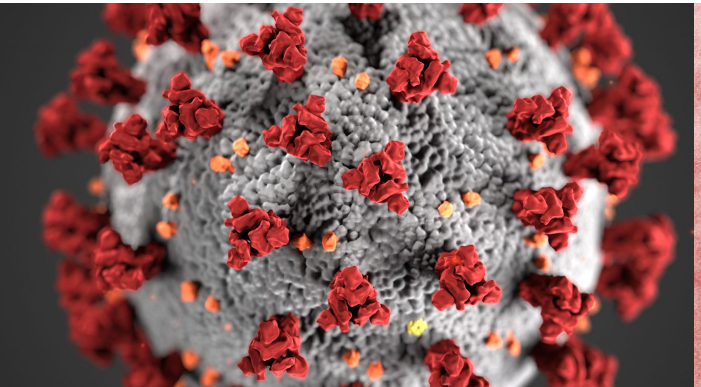
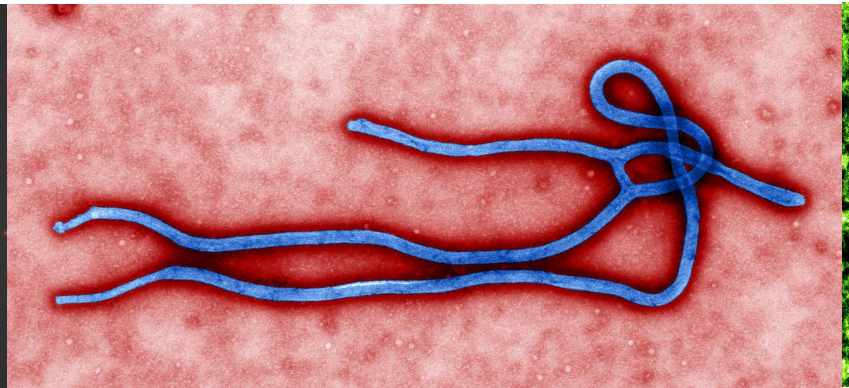
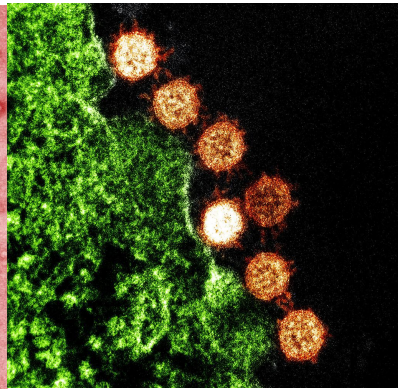
Top Candidate: Logistic Regression, Threshold 0.7



Predicts more true negatives to focus campaign on, minimizes false positives

Next Steps

- Further refine model; decrease false positives even more?
- Apply refined model to test data set

Thank you!
(and please wash your hands!)