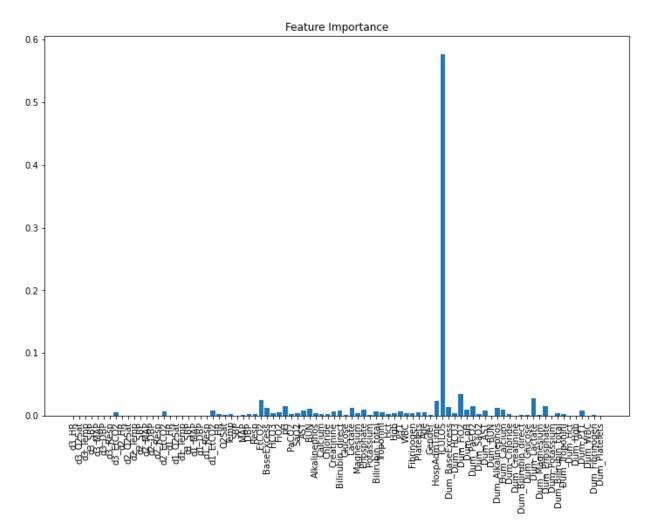Ultimate Technologies Challenge Report

Final Model Metrics

By Aisling Casey – June 6th, 2021

Model Parameters

Gradient Boosting Model Parameters:

```
{'ccp_alpha': 0.0, 'criterion': 'friedman_mse','init': None,
 'learning_rate': 0.01,'loss': 'deviance', 'max_depth': 5,
 'max_features': 'sqrt','max_leaf_nodes': None,
 'min_impurity_decrease': 0.0,'min_impurity_split': None,
 'min_samples_leaf': 1,'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,'n_estimators': 300,
 'n_iter_no_change': None,'random_state': 42,'subsample': 1.0,
 'tol': 0.0001,'validation_fraction': 0.1,'verbose': 0,
'warm_start': False}
```


Feature Importance

Results

```
             precision    recall  f1-score   support

        0.0       0.94      0.97      0.96    420053
        1.0       0.60      0.43      0.50     43136

   accuracy                           0.92    463189
  macro avg       0.77      0.70      0.73    463189
weighted avg       0.91      0.92      0.91    463189
```

Note: Here, the probability threshold has been set to 20.5%, which yields the best f1 score on the test set.

Confusion Matrix – Testing Data

	Predicted Inactive	Predicted Active
Actual Inactive	407893	12160
Actual Active	24783	18353

Note: The probability threshold has been set to 20.5%.

ROC Curve - Gradient Boost - Hourly

Training Data (AUC = 0.86)
Testing Data (AUC = 0.84)
No Skill

Precison- Recall - Gradient Boost - Hourly