**Methodology Intro:**

In order to achieve higher overall accuracy and recall rate, we carefully select features for training and testing inputs. After the feature selection we divide them into four different data arrays: training input, training output, testing input and testing output; First we feed the training input and output to train our Random Forest classifier; Once the classifier is trained for accept one set of specific features, we feed in the test input into the trained classifier; after the classifier make its prediction based on the test input, we evaluate the trained classifier using four standard metrics: Accuracy, precision, recall and F-score;

Each time we will select our features based on different factors and compare them side by side. Our goal was try to achieve highest overall metrics while eliminate as much “unimportant” features from original feature set.

**Results Explanation:**

Fig.1 shows a side-by-side performance comparison of five feature sets based on Precision, Recall, F-Score and Accuracy. These metrics will be our standard to measure the performance of each classifier. As figure shows, set A shows significantly lower precision and recall rate compare to other sets; Set B contains more features than set A however these features still lack importance for determine the phishing sites; Set C, D and E are very close on performance however we think Set E performs best because its contain least numbers of features while still yield best performance.

Fig. 2 shows the feature counts affect the accuracy. According to this chart we cannot gave a concrete correlation between these two; However, we do observe three behaviors: First, small amount of feature sets does not always result in lower accuracy; second, for set that contains 21 features underperformed on all the metrics compare to sets have 8 and 16 features; Finally, the importance of features the does not contributes much when overall performance is higher than 0.94; (Check)

**Discussion:**

1. Why Set C performed so well?
2. Why Set B underperformed?
3. Comment on feature importance and Set E outperforming the other sets.