# Phishing Detection

Approach, Tools and Techniques

In this assignment, the supervised learning approach has been used, since the dataset is labeled.

Hence, I have used Classification algorithms like Decision Tree and Random Forest has been used.

In this assignment, I have used Jupyter Notebook, Python Libraries like:

* Numpy
* Pandas
* Matplotlib
* Seaborn
* Scikit Learn

Steps followed

1. Loading the dataset onto the Jupyter Notebook.
2. Preparing and Analyzing the Data.

* Shuffling/Randomizing the data to make data evenly distributed.
* Cleaning the data (checking for missing values, rows or columns)
* Visualizing the data(correlation matrix plot , histogram).
* Splitting the dataset into training set and testing set (80% of the dataset for training and 20% of the dataset for testing).

1. Choosing the machine learning model ( Decision Tree , Random Forest, Naïve Bayes, KNN )
2. Training the model.
3. Evaluating the model and hyper parameter tuning

* Confusion Matrix
* Accuracy of training and testing data.
* Recall
* Precision
* F1 score

Conclusion

I have used 4 classification algorithms in the given dataset and have got an accuracy of 97.7% when using K nearest neighbors algorithm.

In the case of phishing detection, Precision is an important metric. More the precision, more accurate the result and model.

While using KNN algorithm, I have got more precision (97.8%) and accuracy making KNN the better model for this particular dataset.