Final Project Progress Update: Phishing Detection

Team Members: Jyotshna Nallabothula
Sai Surendra Kommineni
Venkata Satyanarayana Pulaparthi
Vinod Kumar Bodeppagiri

Progress so far:

- Dataset: https://archive.ics.uci.edu/dataset/327/phishing+websites
- So far, we have:
  - Cleaned the data by checking for null values and datatypes
  - Did exploratory analysis like checking correlation with target class, plotting correlation matrix, checking class imbalance.
  - Done some model evaluation on Support vector machines, Naïve Bayes and computed scores such as Accuracy, precision, recall, roc & f1

Challenges:

- We initially thought of using the library, https://pypi.org/project/elm/ which is not working well with scikit, panda's library & Jupiter notebook. We found another library on GitHub https://github.com/masaponto/Python-ELM that has integration with scikit library.
- We had some conceptual misunderstandings while explaining the performance evaluation in the proposal. We mentioned using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) for evaluation, but we realized that they are for regression-based problems, but we are doing classification. So, we used accuracy, precision, recall & f1 scores.

Collaboration:

- Our team meets frequently every day, and everyone is contributing meaningfully.

Next steps:

- Properly training ELM model and computing the scores and evaluating the best model for the given problem of phishing detection.
- Our plan for completion is to basically go through the documentation and implement the code. Use the scoring library from scikit to compute various scores.
- We don’t see any major challenges.