Description of the Project:

The focus of the study is on detecting fraudulent URLs using machine learning models, specifically SVM and Random Forests algorithm. The performance of the algorithms in detecting malicious URLs was evaluated, achieving up to 90% precision and 88% recall using only descriptive features. The study aims to contribute to the development of an automated filtration system for detecting phishing URLs.

Functionality Added:

We have used regex (regular expressions) which are used for identifying patterns in python to split the URL into sub parts and trained model with 80% data and tested on remaining 20% of data. Used new machine learning model logistic regression on the new dataset for this project.

Requirements:

1. Install python and jupyter notebook in the system.
2. After completing the first step, please install the below libraries.
3. pip install matplotlib
4. pip install numpy
5. pip install pandas
6. pip install Requests
7. pip install scikit\_learn
8. pip install seaborn
9. You can now open the Jupyter notebook file named “Submission” and execute the project step by step, starting from the top by running each cell systematically.
10. The code works perfectly, the only problem might occur when reading the files. Please provide the specific path of that file if you encounter any errors, rest of the code will work fine.