DEPARTMENT OF SOFTWARE ENGINEERING

EMPIRICAL SOFTWARE ENGINEERING (SE-404)

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Project Proposal

Detection of malicious URLs using Machine learning algorithms

SUBMITTED TO:
MS. KISHWAR KHAN

TEAM MEMBERS:
AMAN BHATIA – 2K18/SE/019
ASHISH KUMAR – 2K18/SE/041

- **Topic:** Text Mining

- **Type of the Project**

  We will be detecting malicious URLs using machine learning algorithms.

- **Objective of the Study**

  Machine learning algorithms have been widely used to solve various real-world problems. From accelerating businesses to providing better healthcare, the use of machine learning algorithms is well explored. With the technologies evolving day by day, there have been some dark sides associated with it. The fraudulent activities are happening on a very large scale which needs to be stopped. Detecting malicious URLs is one of the most important steps towards preventing fraud over the internet. Thus, to make the internet a safer place, there is a great need for systems that help in identifying phishing URLs. So we present different ways in which malicious URLs can be detected using different machine learning algorithms.

- **Idea and Methodology**

  Phishing attacks are one of the most widespread problems over the internet. A lot of internet users fall into the hands of attackers every day which accounts into millions of dollars of fraud around the globe every day. Thus, there is a need to employ intelligent algorithms to solve these serious problems and this is the motivation behind this project. In this project, we present different ways in which malicious URLs can be detected using different machine learning algorithms.

  The process goes through five major steps. The first and foremost step is to collect data for the study. The data collection step is followed by feature engineering where we take the features that are relevant and of our interest. After that, data preprocessing is done before feeding the data into machine learning algorithms. The performance of each classifier is evaluated as the final step. The overview of the methodology used in this project is as shown in fig. 1.

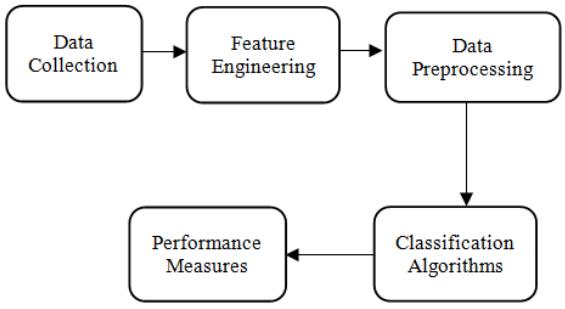

Fig. 1. Methodology of the experiment

- **Project Timeline**