Problem Statement

- Phishing is when attackers send malicious emails designed to trick people into falling for a scam. Typically, the intent is to get users to reveal financial information, system credentials or other sensitive data.
- Phishing continually evolves to bypass security and human detection, so organisations must continually train staff to recognise the latest phishing strategies. It only takes one person to fall for phishing to incite a severe data breach. That’s why it’s one of the most critical threats to mitigate and the most difficult since it requires human defences.
- Cyber criminals use phishing emails because it’s easy, cheap and effective. Email addresses are easy to obtain, and emails are virtually free to send. With little effort and cost, attackers can quickly gain access to valuable data. Those who fall for phishing scams may end up with malware infections (including ransomware), identity theft and data loss.

- Detecting and preventing phishing offenses is a significant challenge for researchers due to the way phishers carry out the attack to bypass the existing anti-phishing techniques.
- Moreover, the phisher can even target some educated and experienced users by using new phishing scams. Thus, software-based phishing detection techniques are preferred for fighting against the phishing attack.

- Mostly available methods for detecting phishing attacks are blacklists/whitelists, natural language processing, visual similarity, rules, machine learning techniques , etc.
- Techniques based on blacklists/whitelists fail to detect unlisted phishing sites as well as these methods fail when blacklisted URL is encountered with minor changes.
- In the machine learning based techniques, a classification model is trained using various heuristic features (i.e., URL, webpage content, website traffic, search engine, WHOIS record, and Page Rank) in order to improve detection efficiency.