# HX6001 - IBM NALAIYA THIRAN PROJECT

WEBSITE PHISHING DETECTION USING MACHINE LEARNING TECHNIQUES

USE CASE DIAGRAM AND ABSTRACT


TEAM MEMBERS:

HEMANTH N (2019503519)

RITIKA EC (2019503551)

KEERTHIKA M (2019503023)

SAI SOUNDHARYA LAKSHMI (2019503554)

UNDER THE GUIDANCE OF

DR. V.P. JAYACHITRA

ABSTRACT

In the last few decades, the web and online services have revolutionized the modern world. However, by increasing our dependence on online services, online security threats have also increased rapidly. Phishing websites are amongst the biggest threats Internet users face today, and many existing methods often fail to keep up with the increasing number of threats. Phishing refers to the practice of trying to obtain sensitive information like passwords and credit card details by mimicking a trust worthy entity. These websites have high level of visual similarities to the legitimate ones in an attempt to defraud honest internet-users.

The numbers of phishing websites are expected to increase over time. Thus, smart solutions are needed to keep pace with the continuous evolution of this problem. Smart solutions are the subject of our interest in our approach. The problem of detecting phishing websites has been addressed many times using various methodologies from conventional classifiers to more complex hybrid methods. The proposal is to combine the heuristic-based approach with the historical dataset about URL that exists. The accuracy of our proposal depends mainly on a set of discriminative features extracted from the website. Hence, the way in which those features are processed plays an extensive role in accurately classifying websites.

We hence aim to utilise different properties of a website URL, and use a machine learning model to detect website phishing. The main objective of our proposal is to maintain a safe and user friendly environment and eventually eradicate any threats from website phishing by detecting them beforehand. The proposed model planned to be deployed using IBM Watson could also be implemented as a classifier in real time, protecting users from phishing by parsing websites and feeding the necessary inputs into the model for immediate predictions. As a tool, the proposed model would also be helpful for web hosting services, cyber security firms, and other interested organizations who would require accurate classifiers for detecting phishing websites.

USE CASE DIAGRAM: