Project Planning Phase

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	18 October 2022
Team ID	PNT2022TMID44414
Project Name	Web Phishing Detection
Marks	8 Marks

Sprint	Milestone
Sprint1	**URL detector** URL is the first thing to analyses a website to decide whether it is a phishing or not **Some of URL-Based Features are** • Digit count in the URL • Total length of URL • Checking whether the URL is typo squatted or not • Checking whether it includes a legitimate brand name or not • Number of subdomains in URL • TLD is one of the commonly used one
Sprint 2	**Domain Detection** The purpose of Phishing Domain Detection is detecting phishing domain names. Therefore, passive queries related to the domain name, which we want to classify as phishing or not, provide useful information to us. Some useful Domain-Based Features are • Its domain name or its IP address in blacklists of well-known reputation services? • How many days passed sincethe domain was registered? • Is the registrant name hidden?

Sprint 3	**Page Based Features and Content Based Features** • Page-Based Features are using information about pages which are calculated reputation ranking services. • Obtaining these types of features requires active scan to target domain. Page contents are processed for us to detect whether target domain is used for phishing or not • Global page rank • Country page rank • Position at the Alexa top 1 million site • Some processed information about pages are • Page titles • Meta tags • Hidden text • Text in the body • Images etc.
Sprint 4	**Detection process** Detecting Phishing Domains is a classification problem, so it means we need labeled data which has samples as phish domains and legitimate domains in the training phase