APPLIED DATA SCIENCE

WEBPHISHING DETECTION

Brainstorm&Idea Priortization

10 minutes to prepare
1 hour to collaborate
2-8 people recommended

DESIGNEDBY;

RAKSHANA.G
PREETHI.D
MONISHA.N
JAYASHREE.K

Share template feedback

Before you collaborate

A little preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

Team gathering
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

Set the goal
Think about the problem you'll be focusing on solving in the brainstorming session.

Learn how to use the facilitation tools
Use the Facilitation Superpowers to run a happy and productive session.

Open article

1 Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

How might we [your problem statement]?

Key rules of brainstorming
To run an smooth and productive session

- Stay in topic.
- Defer judgment.
- Go for volume.
- Encourage wild ideas.
- Listen to others.
- If possible, be visual.

2 Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

RAKSHANA.G — My View

Phishingisatypeofsocial engineeringattackoftenusedtosteal userdata,includinglogincredentials andcreditcardnumbers.Itoccurs whenanattacker,masqueradingasa trustedentity,dupesavictiminto openingaanemail,instantmessage,or textmessage.Therecipientisthen trickedintoclickingamaliciouslink, whichcanleadtotheinstallationof malware,thefreezingofthesystemas partofaransomwareattackorthe revealingofsensitiveinformation.

PREETHI D — My Suggestion

Presence of IP address in URL: If IP address present in URL then the feature is set to 1 else set to 0. Most of the benign sites do not use IP address as an URL to download a webpage. Use of IP address in URL indicates that attacker is trying to steal sensitive information.

MONISHA.N — My point

MACHINE LEARNING ALGORITHM Three machine learning classification model Decision Tree, Random forest and Support vector machine has been selected to detect phishing websites.

JAYASHREE.K — My Turn

This paper aims to enhance detection method to detect phishing websites using machine learning technology. We achieved 97.14% detection accuracy using random forest algorithm with lowest false positive rate. Also result shows that classifiers give better performance when we used more data as training data. In future hybrid technology will be implemented to detect phishing websites more accurately, for which random forest algorithm of machine learning technology and blacklist method will be used.

3 Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

DATA SET

URLs of benign websites were collected from www.alexa.com and The URLs of phishing websites were collected from www.phishtank.com. The data set consists of total 36,711 URLs which include 17058 benign URLs and 19653 phishing URLs. Benign URLs are labelled as "0" and phishing URLs are labelled as "1"

DECISION TREE ALGORITHM

One of the most widely used algorithm is machine learning technology. Decision tree algorithm is easy to understand and also easy to implement. Decision tree begins its work by choosing best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to build tree until it finds the leaf node of the tree. Leaf node is a class label which is used to predict target value or class in tree representation each internal node of the tree belongs to attributes and each leaf node of the tree belongs to class label. In decision tree algorithm, gini index and information gain methods are used to calculate these nodes.

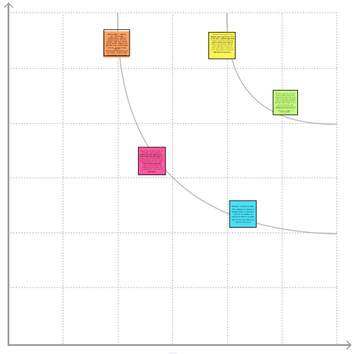
IMPLEMENTATION AND RESULT

IMPLEMENTATION AND RESULT Scikit-learn tool has been used to import Machine learning algorithms. Dataset is divided into training set and testing set in 50:50, 70:30 and 90:10 ratios respectively. Each classifier is trained using training set and testing set is used to evaluate performance of classifiers. Performance of classifiers has been evaluated by calculating classifier's accuracy score, false negative rate and false positive

4 Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

Share the mural
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

Export the mural
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint
Define the components of a new idea or strategy.
Open the template

Customer experience journey map
Understand customer needs, motivations, and obstacles for an experience.
Open the template

Strengths, weaknesses, opportunities & threats
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template

Share template feedback

Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions