

**91.25 Ciencia de Datos para la Toma de Desiciones**

**Trabajo de Aplicacion**

**¨Modelo Predictivo para casos de Phishing¨**

Alumnos:

* Juan Cruz Camacho ()
* Tobías Pucci Romero ()
* Lucas Ken Kiriyama ()
* Brian Emanuel Slavkin (102820)

Cuatrimestre: 1º 2021

Docentes:

* Dr. Ing. Emilio Picasso
* Mag. Ing. Mariano Bonoli Escobar
* Mag. Ing. Xavier González

Resumen Ejecutivo

Indice

Introducción

Para comenzar la búsqueda de nuestro medido predictivo, debemos comenzar entendiendo que es el phishing y cual es la importancia de lograr evitar estas prácticas en el ámbito cibernético.

La palabra phishing quiere decir suplantación de identidad, se utiliza para denominar al conjunto de técnicas de ingeniería social que utilizan los ciberdelincuentes para obtener información confidencial o personal de otros usuarios de la red de manera fraudulenta y así tomar la identidad del estafado.

La practica mas frecuente de estas técnicas se realiza mediante correos electrónicos falsos los cuales aparentan no serlo, como podría ser un servicio de banco, subscripciones de paginas reconocidas por el individuo a estafar, redes sociales, etc. Estos emails derivan en enlaces o links los cuales también aparentan ser reales. Le exigen al usuario sus datos como pueden ser nombre de usuario, contraseñas o cuentas bancarias, mediante encuestas las cuales aparentan ser de los remitentes reales, en los cuales por lo general el estafado confía y así se concreta la obtención de datos del fraude. Otra técnica muy utilizada es mediante archivos adjuntos o botones los cuales generan una descarga la cual puede ser un virus o software del cual el remitente obtiene la información buscada.

Estos enlaces van a ser nuestro principal objeto de estudio, ya que no podemos evitar que el delincuente envíe los correos electrónicos ya que al identificarlos y bloquearlos siempre esta la posibilidad de crear uno nuevo.

Es importante también remarcar que los enlaces pueden llegar no solo por mail, sino que hoy en día estas técnicas mutaron otras vías, como mensajes de texto o redes sociales como WhatsApp, Facebook, Instagram, etc.

En el año 2011 en la Argentina se presento un proyecto de ley en el Senado de la nación en el cual se define el concepto:

*“La presente Ley tiene como objeto tipificar el delito de obtención ilegitima de datos confidenciales, conocido como “phishing” el cual se define como la capacidad de duplicar una página Web para hacer creer al visitante que se encuentra en el sitio Web original, en lugar del falso. Normalmente, se utiliza con fines delictivos enviando SPAM e invitando acceder a la página señuelo. El objetivo del engaño es adquirir información confidencial del usuario como contraseñas, tarjetas de crédito o datos financieros y bancarios.”[[1]](#footnote-1)*

<https://www.argentina.gob.ar/justicia/convosenlaweb/situaciones/phishing>

<https://phishtank.org/what_is_phishing.php>

<https://es.wikipedia.org/wiki/Phishing>

Los enlaces a los que nos estamos refiriendo son direcciones web, también denominada URL (Uniform Resource Locator). Estos son una serie de caracteres, número y símbolos que ayudan a localizar un recurso único en la Web, los cuales pueden ser paginas HTML, documentos CSS, imágenes, etc.

La composición de un URL es un concepto clave para el análisis de nuestro trabajo. Esto se puede observar en la siguiente imagen:



El protocolo (Scheme en inglés) es el primer componente que siempre se encuentra a la izquierda, la que indica como el navegador debe acceder al recurso al cual queremos llegar. Es una cuestión técnica. Los utilizados en las paginas web son http o https, y la diferencia entre estos es que el https impide que otros usuarios puedan interceptar la información confidencial que se transfiere entre el usuario y el servidor.

Luego viene el subdominio y el dominio que indican el servidor web que se solicita. El dominio es el nombre del sitio y el subdominio es una ramificación del dominio que sirve para crear distintas paginas. La extensión es la primera división que se realizó en internet, las cuales se comenzaron creando genéricas como *com*, *net*, *org* y una extensión para cada país del mundo. Como quedan pocos dominios libres en las extensiones fueron surgiendo con el tiempo nuevas.

Luego tenemos componente optativos como lo pueden ser la carpeta y la subcarpeta también conocidos como el directorio el cual muestra al servidor el camino hacia donde esta el recurso. En caso de buscar dentro del dominio específicamente una pagina, imagen o video el URL incorpora el componente recurso.

Por ultimo tenemos los parámetros, los cuales también son opcionales, y aparecen al final del URL separados mediante el ¨ ? ¨. Estos sirven a los programadores para incluir información adicional que utilizan las paginas web para enviar información a los servidores. Consiste, como se puede ver en el ejemplo de una clave y un valor (clave=valor).

<https://superadmin.es/blog/que-es/direccion-web-url/>

<https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_URL>

<https://es.godaddy.com/blog/diferencia-entre-http-y-https/>

Objetivo del análisis

Buscamos obtener un modelo que generalice bien antes nuevas observaciones, permitiendo detectar si una página tiene contenido legítimo o si este es un método de phishing web usando la menor información posible detectar.

Este servicio podría usarse como plugin (complemento) en buscadores web y en servicios de email o mensajería para prevenir a los usuarios el acceso a páginas potencialmente peligrosas. Una posibilidad es informar la probabilidad de que dicha página pueda tratarse de una actividad ilegítima a través de una predicción discreta

1. Proyecto de ley incorporando el art. 157 ter al Código Penal, tipificando el delito de obtención ilegitima de datos confidenciales ("phishing"). N° de expediente: 2257/11 [↑](#footnote-ref-1)