Informe Ejecutivo: Análisis de Sitios Web Argentinos para Detección de Fraudes

Fecha: [fecha]

Destinatarios: [Stakeholders Internos (ej: Equipos de desarrollo) Stakeholders Externos (ej: Clientes, Usuarios, Partners)]

Introducción

Este informe presenta los hallazgos iniciales y las conclusiones clave de un análisis exploratorio de datos enfocado en la detección de sitios web fraudulentos. El objetivo es proporcionar una visión general de los patrones identificados en sitios web legítimos argentinos, sentando las bases para futuros modelos de detección de fraude. Este análisis se realizó sobre una muestra de 100 sitios web argentinos legítimos y se complementó con datos de sitios de phishing conocidos.

Anatomía de una URL:

- **Imagen:** [Aquí se insertaría una imagen que ilustra las diferentes partes de una URL, como protocolo, subdominio, dominio, TLD, etc.]

Hallazgos Clave [Resultados del Análisis (basado en el notebook adjunto)]

1. Dominio y Antigüedad

La gran mayoría de los sitios web argentinos analizados utilizan dominios confiables como `.com.ar`, `.edu.ar` o `.gob.ar`. Solo un 18% de los sitios presentan TLDs (dominios de nivel superior) diferentes a estos, lo que podría ser un indicador de riesgo a investigar más a fondo. Es importante destacar que ninguno de los sitios analizados fue creado en los últimos diez años, lo cual es coherente con su legitimidad y popularidad, ya que los sitios fraudulentos suelen tener un tiempo de vida muy corto.

2. Certificado de Seguridad (SSL)

Un hallazgo significativo es que casi el 80% de los sitios legítimos analizados cuentan con un certificado SSL (Secure Sockets Layer). Este certificado es crucial para asegurar la conexión entre el usuario y el sitio web, protegiendo la información. Aunque la ausencia de un certificado SSL no es una prueba definitiva de fraude, es un indicador

importante a considerar, ya que muchos sitios de phishing no invierten en esta medida de seguridad.

3. Tiempos de Respuesta

La mayoría de los sitios legítimos responden en menos de un segundo. Un tiempo de respuesta inusualmente alto (más de dos segundos) podría ser una señal de alerta, ya que los sitios fraudulentos a menudo tienen infraestructuras menos robustas o están alojados en servidores de baja calidad, lo que resulta en demoras significativas.

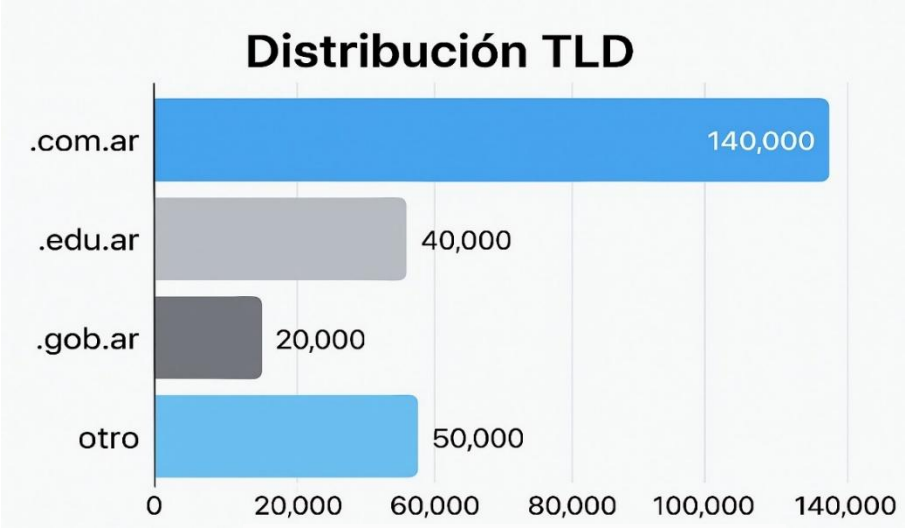
4. Palabras Clave Sospechosas

Se identificó que entre el 3% y el 6% de los sitios analizados contienen palabras clave sospechosas tanto en la URL como en el título. Estas palabras suelen estar asociadas con intentos de fraude, como términos relacionados con banca, seguridad, actualizaciones urgentes o promociones engañosas. Si bien su presencia no confirma un fraude, sí justifica una revisión más detallada.

Visualizaciones Clave

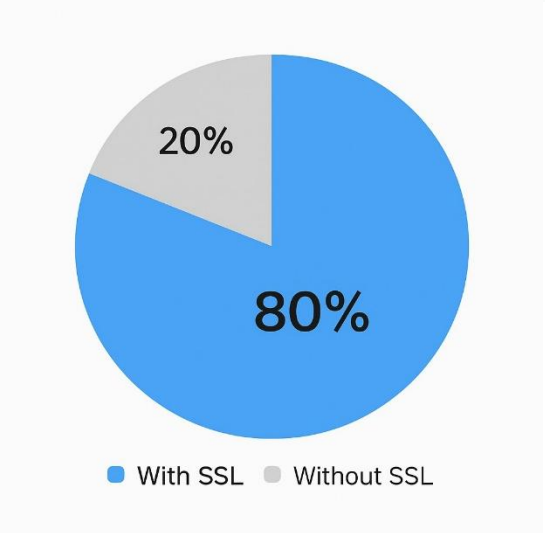
Distribución de Dominios por TLD

La siguiente gráfica muestra la distribución de los dominios analizados según su TLD. Se observa una clara predominancia de los TLDs `.com.ar`, que refuerza la idea de que los sitios legítimos argentinos suelen utilizar estas extensiones.



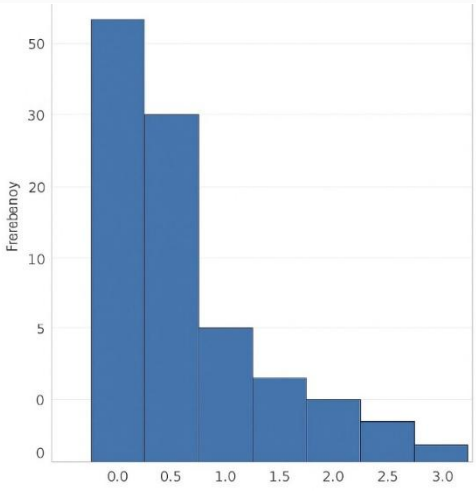
Proporción de Sitios con Certificado SSL

Este gráfico circular ilustra la alta proporción de sitios legítimos que emplean certificados SSL, un indicador clave de seguridad y confianza.



Distribución de Tiempos de Respuesta

Este histograma muestra que la mayoría de los sitios legítimos tienen tiempos de respuesta rápidos, concentrándose en menos de un segundo.



Conclusiones y Próximos Pasos

Este análisis inicial ha permitido identificar patrones y características clave en sitios web legítimos argentinos que pueden ser utilizados para la detección de fraudes. Los indicadores como TLDs inusuales, ausencia de certificado SSL, tiempos de respuesta elevados y la presencia de palabras clave sospechosas son señales de alerta importantes.

Para los próximos pasos, se seguirá ampliando la recopilación de datos para incluir una muestra más diversa de sitios legítimos y fraudulentos, así como también se profundizará en el análisis del contenido de los distintos tipos de url, para una detección más precisa.

Otros agregados que se pueden considerar para profesionalizar el documento:

Logotipo de la organización (en este caso equipo de trabajo!): Incluir el logotipo en el encabezado o pie de página.

Tabla de Contenidos: Si el documento es extenso, incluir una tabla de contenidos para facilitar la navegación.

Glosario de Términos: Si se utilizan términos técnicos, incluir un glosario puede ser útil para algunos destinatarios.

Información de Contacto: Incluir información de contacto para preguntas o comentarios.

Control de Versiones: Si el documento se actualizará, incluir un control de versiones (fecha y descripción de los cambios)