MACHINE LEARNING CASO DE USO

INFORME 6

Diego Martin Olea

ÍNDICE

1. Descripción de caso de uso

Detección automática de phising en trabajadores de una empresa. A día de hoy los delitos están cambiando o se está produciendo en distintos ámbitos al que consideraríamos tradicional por parte de la sociedad en las que hasta ahora era necesaria la participación física para realizar el hecho delictivo o el delito.

Las tecnologías han provocado un gran avance en cuanto a comunicación, accesibilidad a diversas formas ya sea en algunos casos de forma gratuita y otra de pago. El principal medio de difusión y de información dejó de ser los medios tradicionales como la prensa, televisión, radio para pasar a ser internet, redes sociales, páginas web. Este cambio se produjo por el avance tecnológico en diversos campos y la creación de internet.

Lo que se podría considerar el mayor avance en nuestra sociedad en nuestros últimos 20/30 años no quiere decir que este exento de peligros, ya que al igual que existe la buena fe y el buen hacer de las personas en la sociedad y en el día a día, existen personas que se van a aprovechar de esta buena fe para actuar de forma contraria y en algunos casos infringiendo la ley.

Internet no es la excepción. Es por ello que actualmente nos encontramos en un momento de desarrollo de muchas tecnologías y aplicaciones que nos facilitan el día a día, nuestro trabajo, etc. Pero que sirven para personas más especializadas en una puerta para poder acceder a nuestros datos o información que tenemos en internet y que puede provocar un grave perjuicio tanto a nivel personal como a nivel empresarial para la empresa si se filtran datos o información confidencial.

Por lo que vamos a proponer un modelo de detección automática de phising para poder evitar esta filtración de información.

El problema: El phishing es uno de los ciberataques más usados en España y en todo el mundo y que se puede realizar de forma muy sencilla. Consiste en la simulación prácticamente idéntica de una página web (bancos, empresas de telefonía u organismos oficiales o un ente de nuestra confianza) con el objetivo de engañar a los usuarios y conseguir ilegalmente sus datos personales.
Los ataques de phishing están creciendo tanto que a día de hoy hasta el 90% de las empresas se encuentra en peligro de sufrir uno este año. En cuanto a la posición de España en el ranking mundial de países más afectados, en la actualidad ocupa nada más y nada menos que el tercer puesto, según informa la consultora Deloitte. Aquí podemos ver las empresas más atacadas por este delito.

LAS EMPRESAS MÁS SUPLANTADAS

La empresa especializada en ciberseguridad Check Point Research ha elaborado un listado con **las 10 empresas que más suplantaciones han sufrido** por parte de los ciberdelincuentes. Se han duplicado e imitado sus páginas web o sus correos electrónicos para hacer creer a las víctimas que era la empresa la que les contactaba, cuando en realidad se trataba de **piratas informáticos**. Entre sus objetivos, el más claro era el de **robar información confidencial y credenciales de pago**, para apropiarse indebidamente del dinero de sus víctimas.

En el último trimestre del pasado año, las empresas más suplantadas fueron, por este orden:

1. El portal de Internet y servicio de correo electrónico **Yahoo (20%)**, a través de emails que anunciaban premios falsos.
2. La empresa de logística internacional **DHL (16%)**, cuya suplantación se produjo sobre todo alrededor del Black Friday.
3. La compañía tecnológica multicanal conocida por todos, **Microsoft (11%)**.
4. El buscador más famoso de Internet, **Google (5,8 %)**.
5. La red social profesional **LinkedIn (5,7%)**.
6. El servicio de transferencia de archivos **WeTransfer (5,3%)**.
7. La plataforma de streaming **Netflix (4,4%)**.
8. Otra empresa de logística que opera a nivel mundial, **FedEx (2,5 %)**,
9. La firma de servicios financieros **HSBC (2,3 %)**.
10. El servicio de mensajería instantánea **WhatsApp (2,2%)**.

Memoriza los nombres de estas empresas, porque si recibes un email o un mensaje de una de ellas es muy probable que, si no eres cliente, se trate de un **caso de phishing**.

Te recordamos que la mejor manera de defenderse de estos ciberataques, además de proteger adecuadamente tus equipos, consiste en tener precaución. Desconfía de los regalos y premios inesperados o desproporcionados y recuerda que **ningún banco**

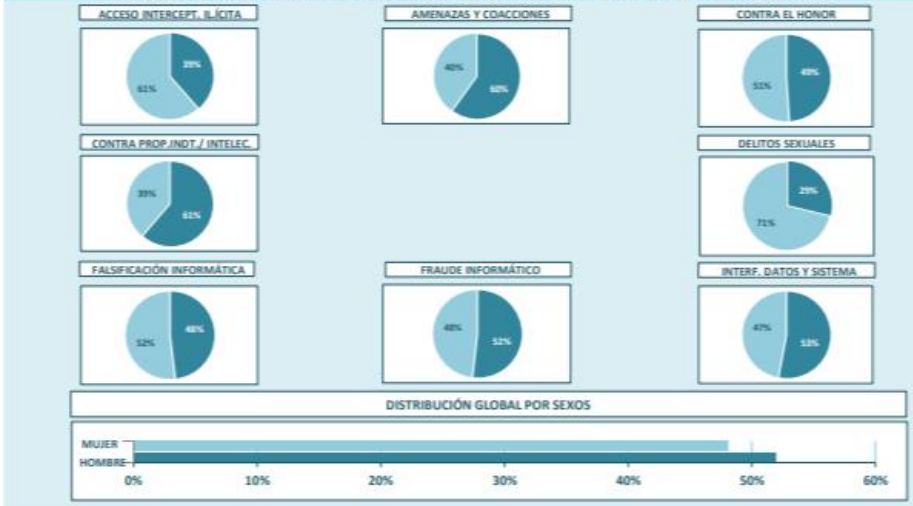
Analizando el Informe sobre la Cibercriminalidad en España 2021 que podemos encontrar en *https://www.interior.gob.es/* vemos las estadísticas de los tipos de delitos realizados en ciberseguridad donde se encuentra el fraude informático como el más realizado y que afecta sobre todo a la población activa y trabajando como podemos ver en esta imagen.

>> 4.6. Victimizaciones registradas según grupo penal y sexo. Año 2021

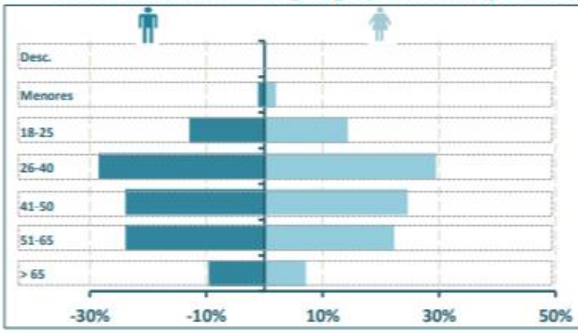
VICTIMIZACIONES	Hombre	Mujer	Desconocido	Total
ACCESO E INTERCEPTACIÓN ILÍCITA	1.755	2.799	1	4.555
AMENAZAS Y COACCIONES	10.530	7.031	25	17.586
CONTRA EL HONOR	714	745	2	1.461
CONTRA LA PROPIEDAD INDUSTRIAL/INTELECTUAL	38	24	0	62
DELITOS SEXUALES (*)	340	842	16	1.198
FALSIFICACIÓN INFORMÁTICA	3.926	4.242	7	8.175
FRAUDE INFORMÁTICO	106.279	98.913	88	205.280
INTERFERENCIA EN LOS DATOS Y EN EL SISTEMA	944	838	1	1.783
Total VICTIMIZACIONES	**124.526**	**115.434**	**140**	**240.100**

(*)Excluidos los agresiones sexuales con/sin penetración y los abusos sexuales con penetración

DISTRIBUCIÓN PORCENTUAL DE LAS VÍCTIMAS POR GRUPO PENAL SEGÚN SEXO



>> 4.7. Victimizaciones según grupo de edad y sexo. Año 2021



Grupo de edad	Hombre	Mujer	Descon.
Edad descon.	132	118	19
Menores de edad	1.396	2.320	17
De 18 a 25 años	16.091	16.509	4
De 26 a 40 años	35.437	34.096	11
De 41 a 50 años	29.762	28.448	29
De 51 a 65 años	29.775	25.785	42
Mayores 65 años	11.933	8.158	18
TOTAL	**124.526**	**115.434**	**140**

46

GOBIERNO DE ESPAÑA | MINISTERIO DEL INTERIOR | DE SEGURIDAD | DIRECCIÓN GENERAL DE COORDINACIÓN Y ESTUDIOS

seo Sistema Estadístico de Criminalidad 2021

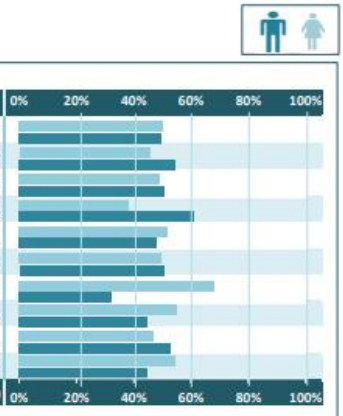
INFORME SOBRE LA CIBERCRIMINALIDAD EN ESPAÑA

4.- DATOS ESTADÍSTICOS DE CIBERCRIMINALIDAD - Perfil de la VÍCTIMA

(Fuente de datos: Sistema Estadístico de Criminalidad: Datos de los cuerpos policiales, excepto Ertzaintza)

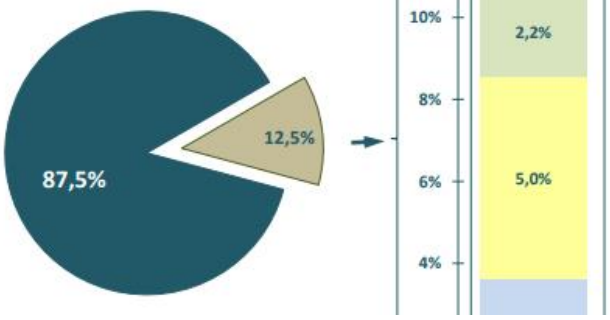
>> 4.8. Victimizaciones por tipología penal y sexo. Año 2021

TIPO DE HECHO	Hombres	Mujeres	Desconoc.	TOTAL
ESTAFAS TARJ CRÉD, DÉB. Y CHEQUES DE VIAJE	45.796	46.394	9	92.199
OTRAS ESTAFAS	45.114	37.797	65	82.976
ESTAFA BANCARIA	14.687	14.167	9	28.863
AMENAZAS	9.342	5.868	25	15.235
USURPACIÓN DE ESTADO CIVIL	3.911	4.230	7	8.148
COACCIONES	1.184	1.161	0	2.345
DESCUBRIMIENTO/REVELACIÓN DE SECRETO	734	1.554	1	2.289
ACCESO ILEGAL INFORMÁTICO	1.020	1.245	0	2.265
ATAQUES INFORMÁTICOS	923	823	0	1.746
RESTO	1.815	2.195	24	4.034
Total VICTIMIZACIONES	**124.526**	**115.434**	**140**	**240.100**



>> 4.9. Nacionalidad de la víctima. Año 2021

NACIONALIDAD	Víctimas	%
1.- ESPAÑOLES	**210.198**	**87,5%**
2.- EXTRANJEROS	**29.902**	**12,5%**
2.1.- UNIÓN EUROPEA	**8.297**	**3,5%**
RUMANÍA	2.939	1,2%
ITALIA	1.560	0,6%
FRANCIA	764	0,3%
BULGARIA	695	0,3%
PORTUGAL	560	0,2%
Resto	1.779	0,7%
2.2.- AMÉRICA	**11.894**	**5,0%**
COLOMBIA	1.868	0,8%
VENEZUELA	1.731	0,7%
ECUADOR	1.321	0,6%
Resto	6.974	2,9%
2.3.- ÁFRICA	**5.241**	**2,2%**
MARRUECOS	3220	1,3%
SENEGAL	460	0,2%
Resto	1.561	0,7%
2.4.- ASIA	**1.897**	**0,8%**



Vemos una evolución desde los 18-25 años a las etapas de estar trabajando donde incrementan exponencialmente y se vuelve a reducir este delito en edades más avanzadas en población por encima de los 65 años.

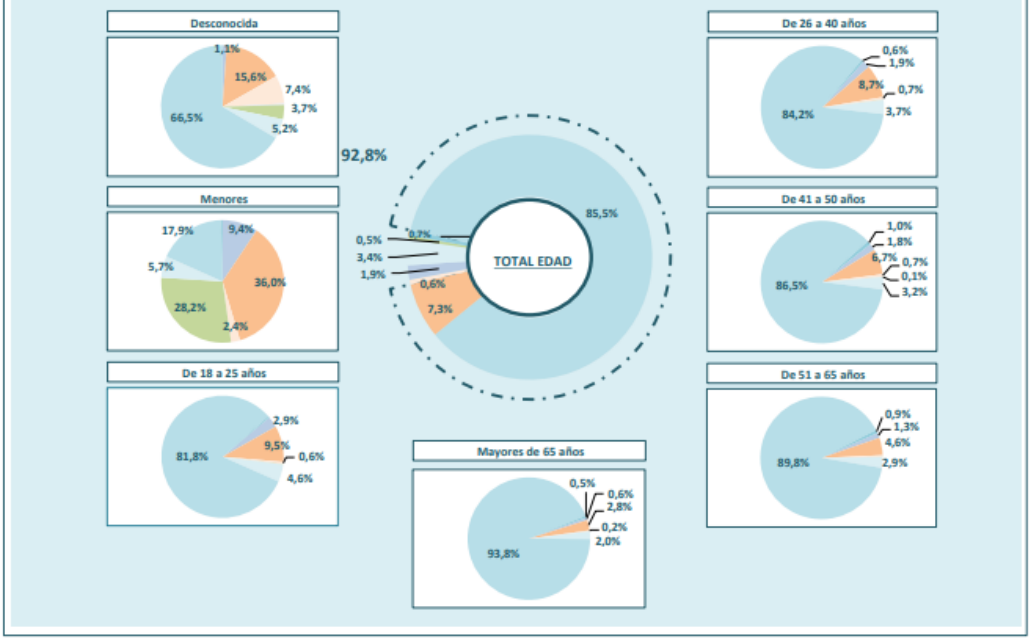
4.- DATOS ESTADÍSTICOS DE CIBERCRIMINALIDAD - Perfil de la VÍCTIMA

(Fuente de datos: Sistema Estadístico de Criminalidad: Datos de los cuerpos policiales, excepto Ertzaintza)

>> 4.10. Victimizaciones registradas según grupo penal y edad. Año 2021

GRUPO PENAL	Rango de edad de la víctima						
	Descon.	Menores	18-25	26-40	41-50	51-65	> 65
ACCESO E INTERCEPTACIÓN ILÍCITA	3	350	944	1.341	1.039	748	130
AMENAZAS Y COACCIONES	42	1.345	3.111	6.072	3.907	2.555	554
CONTRA EL HONOR	20	89	201	470	394	244	43
CONTRA PROPIEDAD INDUST./INTELEC.	1	0	2	14	26	17	2
DELITOS SEXUALES(*)	10	1.053	36	50	30	16	3
FALSIFICACIÓN INFORMÁTICA	14	212	1.494	2.561	1.892	1.590	412
FRAUDE INFORMÁTICO	179	668	26.684	58.587	50.357	49.949	18.856
INTERFERENCIA EN DATOS Y EN SISTEMA	0	16	132	449	594	483	109
Total VICTIMIZACIONES	**269**	**3.733**	**32.604**	**69.544**	**58.239**	**55.602**	**20.109**

(*)Excluidos las agresiones sexuales con/sin penetración y los abusos sexuales con penetración



>> 4.11. Edad de la víctima. Año 2021

1.1 Problema:

La empresa X se enfrenta a un creciente riesgo de ataques de phishing dirigidos a sus empleados. Estos ataques comprometen la seguridad de la información confidencial y pueden resultar en pérdidas financieras, filtración de datos personales, contraseñas para acceder a información de la empresa. Actualmente en la empresa solo existe un control de detección de correos electrónicos de phishing basado en una serie de reglas predefinidas aplicada de forma general para los correos y fiándose en la conciencia de la propia persona/trabajador a la hora de controlar y abrir estos correos. Estas medidas son insuficientes para tener un control y una protección de la información tanto de la empresa como personal ya que las tácticas cada vez son más difíciles de detectar por esta serie de normas y de eso se valen los atacantes para conseguir engañar al usuario.

1.2 ¿Cómo se está afrontando ahora?:

Actualmente, la empresa utiliza filtros de correo electrónico y concienciación del usuario para identificar posibles intentos de phishing. Esto no es suficiente para detener todos los ataques ya que dejas un margen de actuación dependiente de la propia persona que recibe el correo por lo que algunos correos electrónicos fraudulentos pasan desapercibidos y pueden conseguir su objetivo y comprometer la información.

Acción que buscamos poder hacer para solucionar el problema:
Implementar un modelo de machine learning para analizar automáticamente los correos electrónicos entrantes y clasificarlos como legítimos o de phishing. Con esta diferencia el usuario estará más seguro a la hora de abrir un correo o dar una información personal. Se busca mejorar la precisión y la capacidad de adaptación del sistema a nuevas tácticas de phishing y así evitar las filtraciones de información de la empresa.

1.3 KPIs – Indicadores de negocio:

Para ello usaremos los siguientes KPIs suponiendo que la empresa tiene 200 empleados y cada empleado recibe 10 correos diarios.

Directo: La tasa de detección de phishing por el modelo de machine learning.
No ambiguo y comprensible: Número absoluto de correos electrónicos de phishing detectados correctamente y la reducción de falsos positivos.

Tasa de Detección de Phishing:

Cálculo: (Número de correos electrónicos de phishing detectados correctamente / Total de correos electrónicos de phishing) * 100.

Suposición: Si asumimos que el 1% de los correos electrónicos son de phishing y el modelo detecta correctamente el 90% de ellos, la tasa de detección sería:

$$\text{Tasa de Detección} = (0.01 \times 0.9) \times 100 = 0.9\%$$

Falsos Positivos:

Cálculo: Número de falsos positivos.

Suposición: Si el modelo clasifica incorrectamente el 1% de los correos electrónicos legítimos como phishing, la cantidad de falsos positivos sería:

$$Falsos\ Positivos = 0.01 \times Total\ de\ correos\ electrónicos\ legítimos$$

Precisión y Recall:

Cálculo:

Precisión:

$$Precisión = \frac{Verdaderos\ Positivos}{Verdaderos\ positivos + Falsos\ positivos}$$

Recall:

$$Precisión = \frac{Verdaderos\ Positivos}{Verdaderos\ positivos + Falsos\ negativos}$$

Suposición: Usando las cifras anteriores, se calcularían precisión y recall.

Tiempo de Respuesta:

Cálculo: Promedio del tiempo de procesamiento de correos electrónicos.

Suposición: Si el tiempo promedio de procesamiento es de 1 segundo por correo electrónico, el tiempo de respuesta sería de 1 segundo.

Porcentaje de Mejora:

$$\boldsymbol{Cálculo} = \frac{Tasa\ de\ detección\ actual - Tasa\ de\ detección\ anterior}{Tasa\ de\ detección\ anterior}\ X\ 100$$

Suposición: Si la tasa de detección anterior era del 80%, la mejora sería calculada.

Nivel de Confiabilidad:

$$Cálculo = \frac{Número\ de\ correos\ electrónicos\ clasificados\ con\ alta\ confiabilidad}{Total\ de\ correos\ electrónicos\ analizados}\ X\ 100$$

Suposición: Si el modelo clasifica el 95% de los correos electrónicos con alta confiabilidad, la métrica sería calculada.

1.4 ¿Cuáles son los mínimos que se esperan?:

Se espera una mejora del 20% en la detección de phishing en comparación con las medidas actuales. El objetivo es reducir los casos no detectados y minimizar los falsos positivos a menos del 5%.

1.5 Validación: ¿Qué criterio se va a usar para decidir si la solución es aceptable?:

Como hemos explicado antes, el objetivo es reducir el número de casos no detectados y a su vez minimizar el número de falsos positivos para así garantizar a la persona que usa este sistema que los correos que recibe son seguros y tiene la garantía de poder abrirlos y responder, dar información sin riesgo a que sea una estafa o que comprometa la información dada.

Para ello la solución se considerará aceptable si logra una mejora sustancial en la detección de phishing, manteniendo los falsos positivos dentro de límites aceptables. Ya que en este caso es mejor tener un falso positivo y no abrir ese correo ni dar ninguna información y en caso de ser falso positivo pedir un segundo correo a alguien de confianza dentro de la empresa para así asegurarnos que quiere obtener esa información y esa demanda es legítima.

1.6 Experimentación: ¿Cómo vamos a corroborar el funcionamiento?:

Se realizarán pruebas piloto utilizando un conjunto de datos de correos electrónicos correspondientes a la empresa. La experimentación se llevará a cabo durante un período de dos meses, con evaluaciones periódicas para ajustar el modelo según sea necesario.

1.7 Productivización: ¿Qué salida debe tener la solución que se desarrolle?:

La solución deberá integrarse en el sistema de correo electrónico existente de la empresa y analizar continuamente los correos electrónicos entrantes. Se proporcionará una interfaz para monitorear el rendimiento y realizar ajustes según sea necesario.

Cuando el modelo detecta un correo electrónico que tiene indicios de phishing, en lugar de activar una alarma inmediata, puedes implementar un sistema de alerta gradual que evalúa la gravedad de la situación. Para garantizar la seguridad de la empresa se establecerán una serie de niveles de gravedad e intervención por parte del sistema para así en caso de ser necesario impedir y bloquear ese correo.

Nivel de confianza:

Cuando el modelo identifica un correo electrónico sospechoso, asigna un nivel de confianza a la predicción.

Este nivel de confianza se basa en la certeza del modelo sobre si el correo electrónico es un intento de phishing.

Para niveles de confianza bajos o moderados se implementarán una serie de medidas como marcar el correo electrónico como sospechoso y notificarlo al usuario. El sistema ante este tipo de casos sugerirá que el usuario receptor de este correo verifique quien es el remitente y que tipo de información le está pidiendo o dando.

Para niveles de confianza más altos se realizarán acciones más activas y directas por parte del sistema como bloquear el acceso a enlaces o archivos adjuntos para así evitar que las personas descarguen e instalen archivos maliciosos que roben o accedan a información del usuario tanto personal como relacionada con la empresa y notificar a los equipos de seguridad para que bloqueen al remitente y lo añadan a una lista de correos maliciosos.

Si el modelo tiene un nivel de confianza extremadamente alto y detecta tácticas avanzadas de phishing se activa una alarma de emergencia. Esta alarma activara un protocolo de actuación en el que los equipos de seguridad y monitorización de la empresa intervendrán incluso llegando a la desconexión temporal de la cuenta de correo, o de la red que esté usando el usuario para así evitar que ese correo consiga el objetivo.

2. Equipo de trabajo

Se designará un equipo compuesto por distintos especialistas:

Experto en Seguridad Informática (Líder del Equipo):
Coordinará el equipo.
Evaluará las amenazas de phishing y tendrá conocimiento avanzado de las distintas tácticas utilizadas por los atacantes.

Analista de Datos:
Preparará y limpiará los datos para el entrenamiento del modelo.
Desarrollará y ajustará modelos de machine learning para la detección de phishing.
Evaluará el rendimiento del modelo y realizará mejoras continuas para así conseguir una mayor eficiencia del sistema.

Desarrollador de Machine Learning / Ingeniero de Software:
Implementará el modelo de detección de phishing en la infraestructura de la empresa.
Integrará el modelo en el sistema de correo electrónico existente.
Desarrollará y mantendrá interfaces para el monitoreo y ajuste del modelo.

Analista de Incidentes de Seguridad:
Investigará y responderá a los incidentes de seguridad relacionados con phishing.
Colaborará con el equipo para mejorar las tácticas de respuesta a incidentes.

3. Detalle del caso de uso

3.1 Detalle Funcional:

Se analizarán las características clave de los correos electrónicos, como la dirección del remitente, el contenido del correo, la presencia de enlaces sospechosos y patrones de redacción. Con este análisis podemos encontrar símbolos, detalles, longitud del correo que nos puedan ayudar a la detección del phising.

3.2 Identificación de Orígenes de Datos:

En el sistema que usaremos añadiremos una serie de datos e información que usará el propio sistema para detectar y analizar los correos, se incorporarán los siguientes datos:

Correos Electrónicos Históricos que contendrán:
Contenido del correo electrónico.
Dirección del remitente y destinatario.
Fecha y hora de envío.
Archivos adjuntos y enlaces.

Datos de Actividad del Usuario:
Frecuencia de interacción con correos electrónicos.
Historial de clics en enlaces.
Respuestas a correos electrónicos.
Horarios en los que abre los correos.

Datos de Red:
Direcciones IP involucradas.
Protocolos de red utilizados.
Tráfico de datos asociado a correos electrónicos.
Para así controlar el remitente, desde donde manda el correo y valorar la seguridad y protocolos que usa.

Metadatos del Sistema de Correo Electrónico:
Registro de eventos de correo electrónico.
Información del servidor de correo.

Metadatos:
Etiquetas que indican si un correo electrónico es legítimo o sospechoso.
Historial de correos electrónicos enviados por el remitente.
Reputación del dominio del remitente.

Características del Contenido analizando metadatos sobre el contenido del correo electrónico, como frecuencia de palabras clave y estructura del texto.

4. Desarrollo del caso de uso

4.1 Puntos Intermedios o Seguimiento:

Se realizarán pruebas para validar la efectividad del modelo como:

Evaluación de Métricas de Rendimiento:
Revisar las métricas clave, como la tasa de detección, falsos positivos, precisión, recall y tiempo de respuesta y comparar las métricas con los objetivos establecidos para así poder verificar si el sistema implantado está dando resultado o si hay algo que hay que corregir ya que no está solucionando ni mejorando el sistema de detección de phising.
Análisis de Falsos Positivos y Falsos Negativos:
Analizar los casos de falsos positivos y falsos negativos para entender las razones detrás de las clasificaciones incorrectas y modificar el método de detección para así evitar una etiqueta errónea en un correo.

Evaluación del usuario:
Recopilar comentarios de los usuarios sobre la precisión de las clasificaciones del modelo.

Pruebas de ataque:
Realizar pruebas de ataque para así evaluar la capacidad del modelo ante posibles ataques dirigidos a engañar o evadir la detección y ver las vulnerabilidades y debilidades del sistema para ir corrigiendo y mejorando el modelo.

4.2. Aporte Esperado por Big Data:

El equipo de Big Data se encargará de incorporar al sistema grandes bases de datos para que así el sistema tenga la máxima información posible y eso ayude a la detección de patrones y poder identificar mejor los casos de phising.

5. Conclusión

Este caso de uso busca mejorar significativamente la capacidad de la empresa para identificar y mitigar los riesgos asociados con los ataques de phishing, utilizando machine learning como una herramienta proactiva en su estrategia de seguridad informática.